

Albrecht Kurze

Modellierung des QoS-QoE-Zusammenhangs für mobile Dienste
und empirische Bestimmung in einem Netzemulations-Testbed

Wissenschaftliche Schriftenreihe
Dissertationen der Medieninformatik
Band 4

Prof. Dr. Maximilian Eibl (Hrsg.)

Albrecht Kurze

**Modellierung des QoS-QoE-Zusammenhangs für
mobile Dienste und empirische Bestimmung in
einem Netzemulations-Testbed**



TECHNISCHE UNIVERSITÄT
CHEMNITZ

**Universitätsverlag Chemnitz
2016**

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Angaben sind im Internet über <http://dnb.d-nb.de> abrufbar.

Cover-Foto: iStock.com/NI QIN (real444)

Technische Universität Chemnitz/Universitätsbibliothek
Universitätsverlag Chemnitz
09107 Chemnitz
<http://www.tu-chemnitz.de/ub/univerlag>

Herstellung und Auslieferung
Verlagshaus Monsenstein und Vannerdat OHG
Am Hawerkamp 31
48155 Münster
<http://www.mv-verlag.de>

ISSN 2195-2574 print; ISSN 2195-2582 online
ISBN 978-3-944640-60-0

<http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-195066>



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Modellierung des
QoS-QoE-Zusammenhangs
für mobile Dienste
und empirische Bestimmung
in einem Netzemulations-Testbed

Dissertation
zur Erlangung des akademischen Grades

Dr.-Ing.

Herr Dipl.-Inf. Albrecht Kurze
geboren am 29. Mai 1978 in Leisnig

Fakultät für Informatik
an der Technischen Universität Chemnitz

Gutachter:

Prof. Dr. Maximilian Eibl,
Technische Universität Chemnitz
Prof. Dr. Josef F. Krems
Technische Universität Chemnitz

Tag der Verteidigung:

16.07.2015

Überblick

Danksagung	IX
Kurzfassung	XI
Inhaltsverzeichnis	XIII
Tabellenverzeichnis	XXIII
Abbildungsverzeichnis	XXV
1 Die Einleitung	
Von der Motivation über die Zielsetzung zu Methodik und Rahmen der Arbeit	1
2 Die Theorie	
Begriffe und Konzepte zu Quality of Service und Quality of Experience	35
3 Das Modell	
Der dienstspezifische Zusammenhang von QoS und QoE.....	113
4 Die Empirie	
QoE-Evaluation, Testplanung und Testrealisierung der Studie	219
5 Das Testbed	
Netz und Endgeräte in der Studie	259
6 Die Dienste	
Vorüberlegungen und Voruntersuchungen zu den Diensten in der Studie	351
7 Die Probanden	
Der Mensch in der Studie	387
8 Die Auswertung	
Bestandteile und Prinzipien	423
9 Der Testparcours	
Ergebnisse und Diskussion der Voruntersuchungen und des Nutzertests.....	503
10 Das Fazit	
Zusammenfassung und Ausblick	591
Literaturverzeichnis	601
A Anhang	627

Danksagung

Der eigentlichen Arbeit möchte ich einige Worte des Dankes voranstellen, die denen gelten, die durch ihre Unterstützung zum Gelingen der Arbeit beigetragen haben. Zuerst gilt dieser Dank meinem Doktorvater Professor Dr. Maximilian Eibl, der mich an der Professur Medieninformatik (MI) stets forderte und somit auch förderte. Durch seine Ermutigungen und fachlichen Ratschläge zur Verbesserung wurde diese Arbeit maßgeblich geformt.

Weiterhin möchte ich Professor Dr. Josef F. Krems von der Professur Allgemeine Psychologie und Arbeitspsychologie (AAP) und Professor Dr.-Ing. Ulrich Heinkel von der Professur Schaltkreis- und Systementwurf (SSE) an der Technischen Universität Chemnitz danken. Ohne ihre Beteiligung an der Konzeption und Durchführung des zugrundeliegenden Projektes wäre die Arbeit nicht in dieser Form möglich gewesen. Professor Dr. Krems gilt dabei zusätzlicher Dank für die Übernahme der Zweitbegutachtung.

Ich möchte ebenfalls die jeweils beteiligten Mitarbeiterinnen und Mitarbeiter der genannten Professuren mit in den Dank einbeziehen. Vielen Dank daher an Cynthia Heller (AAP) in der Testrunde 2 und Claudia Mair (AAP) in der Testrunde 1 sowie Dr.-Ing. Daniel Kriesten (SSE). Hinzu kommen die wissenschaftlichen und studentischen Hilfskräfte, namentlich Josephine Halama, Marie Kappen und Daniel Richter, die sich an der Organisation und Durchführung der Nutzerstudie beteiligt haben. Katharina Einert gebührt ein ganz besonders großes Dankeschön für die Übernahme dieser Aufgaben in der ersten Testrunde. Auch der Seite des Projektgebers soll mein Dank gelten, besonders den dort beteiligten Mitarbeitern, die mit ihrem Vertrauen dieses Projekt erst ermöglichten und durch ihr stetiges Engagement das Erreichen der gesteckten Ziele unterstützten. Vielen Dank möchte ich auch an die zahlreichen freiwilligen Probanden für die Teilnahme an der Studie aussprechen.

Ganz herzlich möchte ich mich bei meinen gegenwärtigen und ehemaligen Kollegen bedanken, die mich ein langes Stück des Wegs an der Professur Medieninformatik begleitet haben. Vielen Dank für die schöne Zeit mit euch! Besonderer Dank gilt denen, die schon vor mir promovierten und mich so mit ihren Arbeiten motivierten.

Über den gesamten Entstehungsprozess der Arbeit hinweg habe ich den fachlichen Austausch im Rahmen des gemeinsamen Doktorandensymposiums in Kooperation mit der Universität Passau geschätzt. Vielen Dank dafür an Professor Dr. Harald Kosch und Dr. Britta Meixner.

Ein ganz besonderer Dank gilt abschließend meiner Familie, die mich stets unterstützte und bekräftigte. In gewisser Art musste dabei auch die Erstellung dieser Arbeit mit durchlitten werden – vielen Dank für euer Verständnis und eure Rücksichtnahme.

Kurzfassung

In der vorliegenden Arbeit wird der Zusammenhang von Quality of Service (QoS) und Quality of Experience (QoE) für mobile Internet-Dienste thematisiert. Dazu wurde der Einfluss von leistungsrelevanten Netzparametern (QoS, z. B. Durchsatz und Latenz) zusammen mit weiteren Nebenfaktoren (z. B. mobiles Endgerät oder App) auf das Nutzererlebnis (QoE, z. B. Zufriedenheit) erforscht. Es sollten grundsätzliche Wirkungen zwischen „schneller“ (QoS) und „besser“ (QoE) aufgeklärt und für ausgewählte Dienste quantifiziert werden, um von einem angestrebten Grad der Nutzerzufriedenheit auf die notwendigen Werte der QoS-Netzparameter zu schließen. Dabei stand die Frage „Wie schnell ist gut genug?“ im Vordergrund, um zwischen den Interessen der Netzanbieter („so schnell wie nötig“) und den Interessen der Nutzer („so gut wie möglich“) zu vermitteln.

Sowohl QoS als auch QoE sind hochkomplexe Konzepte. Insbesondere QoE bedarf einer multi- bzw. interdisziplinären Betrachtung zwischen Ingenieurs- und Humanwissenschaften, da neben Technik und Kontext der Mensch ein wichtiger Faktor ist. Zur Berücksichtigung der relevanten Einflussfaktoren und internen Zusammenhänge zwischen QoS und QoE wurde ein eigenes Modell entworfen. Dieses integriert sowohl die technische Netz- als auch die Nutzersicht auf die mobilen Dienste in der Art, dass plausible Vorhersagen zum QoS-QoE-Zusammenhang möglich sind. Durch den weichen und unscharfen (nichtdeterministischen) Einflussfaktor Mensch sind diese Zusammenhänge aber nicht rein rational aufzuklären und zu quantifizieren. Daraus entsteht die Notwendigkeit einer empirischen Untersuchung in Form einer QoE-Evaluation.

Es wurde ein umfangreiches psychophysikalisches Laborexperiment mit echten Nutzern, echten Endgeräten und echten Diensten konzipiert, um planmäßig unterschiedliche Netz- und Nutzungssituationen zusammenzubringen. Dafür war zu klären, was typische mobile Netz- und Nutzungssituationen überhaupt sind – und wie sich diese sinnvoll in die empirische Untersuchung integrieren lassen. Aus der Analyse mobiler Netzsituationen entstanden sinnvolle Kombi-

nationen der primär leistungsrelevanten QoS-Parameter. Innerhalb eines entwickelten WLAN-Testbeds zur drahtlosen Einbindung der mobilen Endgeräte können durch eine Netzemulation diese Netzsituationen flexibel und gezielt nachgebildet werden. Für die Einbeziehung der mobilen Dienste in die empirische Untersuchung wurden Prinzipien zur Testrelevanz, Testeignung und Testeffizienz definiert und umgesetzt, welche die Besonderheiten des Testaufbaus und des Testdesigns angemessen berücksichtigen. Der entwickelte Testparcours der Nutzungssituationen umfasst typische Nutzungsszenarien für sechs ausgewählte mobile Dienste. Die Stichprobe der ausgeführten Studie von 216 Probanden mit Vorerfahrungen in der Nutzung des mobilen Internets ist überwiegend jung und gut gebildet.

In der ausführlichen Auswertung der erhobenen Daten konnten vielfältige Erkenntnisse zum QoS-QoE-Zusammenhang gewonnen werden. Die vorhergesagten QoS-QoE-Charakteristiken zu den verschiedenen Diensten im Testparcours wurden als entweder kontinuierlich-elastisch oder sprunghaft-fest bestätigt. In der statistischen Analyse konnte für die untersuchten Dienste jeweils ein Zufriedenheitskorridor zwischen einem unteren und oberen Schwellwert erfolgreich ermittelt werden. Dabei sind teilweise die im Modell berücksichtigten QoS-unabhängigen Einflussfaktoren auf die QoE, z. B. die Art der Präsentation der Stimuli in der App, als ebenso relevant zu erkennen wie die QoS-Netzparameter selbst.

Inhaltsverzeichnis

1	Die Einleitung	1
1.1	Motivation	2
1.1.1	Multidisziplinäres Problem und interdisziplinäre Betrachtung	2
1.1.2	Maschine, Mensch, Markt: schneller, besser, billiger	3
1.1.3	Mensch und Maschine: Daten-Tsunami und „gut genug“	5
1.1.4	Netzevolution und Daten-Tsunami: Höher, weiter, schneller im mobilen Internet	6
1.1.5	Das rechte Maß	10
1.1.6	Maschine und Markt: Knappe Güter und weiße Flecken	11
1.1.7	Markt und Mensch: Kundenzufriedenheit = Nutzerzufriedenheit / Preis?	12
1.1.8	Zusammenfassung zur Motivation	13
1.2	Zielsetzung	14
1.3	Methodik und Rahmen	15
1.3.1	Methodik: Theorie, Modell und Empirie	15
1.3.2	Theorie: Grundannahmen und Ausgangspunkt	17
1.3.3	Modell: Von der Black Box zur White Box	18
1.3.4	Empirie: Realismus und Kontrolle	21
1.3.5	Vorarbeiten	25
1.3.6	Überblick zu den ausgeführten Studien	26
1.4	Aufbau der Arbeit	28
2	Die Theorie	35
2.1	Begriffswelten: Von QoS zu QoE	35
2.1.1	Anwendungsbereiche zur QoS und QoE: Netz, Nutzer, Multimedia	36
2.1.2	Problematik der gemeinsamen Betrachtung von QoS und QoE	38
2.1.3	Die Qualität: Zwischen Beschaffenheit und Eignung	39
2.2	Quality of Service und das Netz	41
2.2.1	Definitionen zur QoS	42
2.2.2	Quality of Something: Bedeutungsunterschiede und Bedeutungsveränderung der QoS	46
2.2.3	Netzqualität und Key Performance Indicators im Mobilfunk	48
2.2.4	Die Netzleistungsfähigkeit: Network Performance (NP)	49
2.2.5	Das Netz in einer Ende-zu-Ende-Sicht	50
2.2.6	Netzparameter zur Beschreibung der netzwerkabhängigen Quality of Service	53
2.2.7	QoS-Spezifikation und Best Effort	56
2.2.8	Over-Provisioning, Garantien, QoS-Mechanismen und Netzverkehrsklassen	57
2.3	Quality of Experience und die Nutzer	60
2.3.1	Definitionen zur QoE	60
2.3.2	Quality of Everything: Die QoE zwischen QoS und Qo*	66
2.3.3	Quality of Experience und der Zusammenhang zur User Experience	68
2.3.4	QoE-Dimensionen/Einflussfaktoren und -größen	69
2.3.5	Überblick zu den menschlichen Faktoren	71
2.3.6	Experience aus Wahrnehmung und Interpretation	74
2.3.7	Grundsätzliches zu den Qualitätsaspekten der QoE	77
2.3.8	Ausgewählte Qualitätsaspekte der QoE	78
2.3.9	Erlebnis und Erwartung im Beurteilungsprozess	81

2.3.10	Arten von Zufriedenheitsfaktoren	84
2.3.11	Nutzungsprozess und Dynamik der QoE.....	86
2.4	Kontext und Nutzungssituationen	88
2.4.1	Arten von Kontext	88
2.4.2	Nutzungsziel und Aufgabe	91
2.4.3	Besonderheiten der mobilen Nutzung	93
2.4.4	Marktbezogene Faktoren und nichttechnische Angebotseigenschaften	94
2.5	Konzeptueller Zusammenhang zwischen QoS und QoE.....	96
2.5.1	Schichtung der Qualitäten	96
2.5.2	Aus Ende-zu-Ende-QoS wird Ende-zu-Ende-QoE	100
2.5.3	Verbindung von Netz und Nutzer über die QoS	101
2.5.4	Wahrnehmung der Netzeigenschaften durch die Nutzer	103
2.5.5	Von der Beschaffenheit zur Bewertung als Eignung zum Zweck	108
2.6	Zusammenfassung.....	110
3	Das Modell	113
3.1	Die Modellierung der QoX-Matrix	114
3.1.1	Andere Modelle und eigener Ansatz zur Modellbildung	114
3.1.2	Qualitätsbegriffe: Von Elementen, Merkmalen und Co.....	116
3.1.3	Die Elemente des Modells	117
3.1.4	Die Schichten des Modells.....	119
3.1.5	Das Prinzip der Mediatoren und Moderatoren	124
3.1.6	QoS-QoE-Ursache-Wirkungs-Kette zwischen den Schichten	125
3.1.7	Von der Kette zur Matrix	127
3.1.8	Zur weiteren Diskussion des Modells	130
3.2	Das mobile Endgerät	131
3.2.1	Hardware: Geräteklassen und Leistungsklassen.....	132
3.2.2	Software: Mobile Plattformen und Herstelleranpassungen	133
3.2.3	Einfluss auf die Kommunikationsleistung	134
3.2.4	Einfluss auf die Applikationsleistung	134
3.2.5	Einfluss auf Präsentation und Interaktion	135
3.2.6	Das Endgerät als Erlebnis- und Erwartungsfaktor.....	136
3.3	Die Dienste – Teil 1: Begriff und Elemente	137
3.3.1	Begriffsbestimmung	137
3.3.2	Anwendung/App	141
3.3.3	Gegenseite und Kommunikationsszenario	144
3.3.4	Inhalt	146
3.4	Die Dienste – Teil 2: Netzsicht	148
3.4.1	Von QoS zu QoD mit der Kommunikationsleistung	149
3.4.2	Protokollwelten: TCP/IP-Stack und Co.....	150
3.4.3	Entstehen und Zusammenhang der Effektivwerte	153
3.4.4	Mobilfunk und TCP-Blues: Congestion, Bufferbloat und Queues	155
3.4.5	Rolle der TCP-Implementierungen und -Parametrisierungen.....	157
3.4.6	Übertragungscharakteristik	159
3.5	Die Dienste – Teil 3: Nutzersicht	162
3.5.1	Software-Ergonomie und Usability der Nutzerschnittstelle	163
3.5.2	Applikationsleistung und Nutzerstimuli.....	164
3.5.3	Funktion und Verhalten.....	167
3.5.4	Qualität und Quantität	168
3.5.5	Zeiten und Dauern.....	170

3.5.6	Entstehen und Zusammenhang der Nutzerstimuli	174
3.5.7	Moderation von QoA zur QoE	176
3.5.8	Präsentation der Nutzerstimuli	177
3.5.9	Psychologie der Leistung	181
3.5.10	Complexity of Experience	182
3.5.11	Interaktivität/Interaktion von Nutzer und Dienst	184
3.6	Die Dienste – Teil 4: Gesamtsicht und Klassen	186
3.6.1	Von Anforderungen und Einschränkungen zum dienstspezifischen QoS-QoE-Zusammenhang	187
3.6.2	Terminologie zur Beschreibung des QoS-QoE-Zusammenhangs	187
3.6.3	Charakterisierung der QoS-QoE-Dienstkurven	189
3.6.4	Bildung von Klassen	194
3.6.5	Mögliche Taxonomien der Dienste	195
3.6.6	Klassen aus Nutzerstimuli	197
3.6.7	Dienstklassen zum QoS-QoE-Zusammenhang	199
3.6.8	Klassifizierung nach zeitlichen Anforderungen	202
3.6.9	Echtzeit: Wann, wenn nicht jetzt?	204
3.6.10	Streaming: Zwischen Skylla und Charybdis ... und den Sirenen der Qualität ..	207
3.6.11	Finish Time: Ist es schon fertig?	208
3.6.12	Interaktive Dienste	209
3.6.13	Die Relevanz der einzelnen QoS-Netzparameter	213
3.7	Zusammenfassung	215
4	Die Empirie	219
4.1	QoE-Evaluation	219
4.1.1	Untersuchungsziele	219
4.1.2	Grundsätzliche Methoden und Techniken zur QoE-Evaluation	220
4.1.3	QoE-Evaluation: Wer, Was, Wie?	222
4.1.4	QoE: Quality of Experience vs. Quality of Expectation	224
4.1.5	Erfassung der Qualitätsaspekte	225
4.1.6	MOS, MOS-Skala und Wahrnehmungsqualität	226
4.1.7	Festlegungen zur Testrunde 2 der eigenen Studie	227
4.1.8	Untersuchungsort und Methode	228
4.1.9	Quantifizierung des Zusammenhangs zwischen QoS und QoE	230
4.2	Unabhängige Variablen und Bedingungsvariation	231
4.2.1	Durch das QoS-QoE-Modell induzierte Einflussvariablen	234
4.2.2	Durch das Testdesign induzierte Variablen und Besonderheiten	234
4.2.3	Bedingungen	235
4.2.4	Art der Zuweisung bzw. Variation der Variablen	236
4.2.5	Rahmenbedingungen	238
4.2.6	Testplanung und Testplan	238
4.2.7	Randomisierung	239
4.2.8	Shuffling von Aufgabenblöcken	239
4.2.9	Testläufe und Testreihen	241
4.2.10	Testprozedere und Testzuweisung	241
4.2.11	Testorganisation und Identifikationsattribute	242
4.3	Abhängige Variablen und Ergebniserfassung	242
4.3.1	Nutzerbewertungen	244
4.3.2	Bearbeitungserfolg	246
4.3.3	Nutzerkommentare	246

4.4	Weitere Ergebnisdaten	247
4.4.1	Testlogbuch	247
4.4.2	Bearbeitungszeiten	248
4.4.3	Protokollierung der Parametrisierung	248
4.4.4	Überwachung und Aufzeichnung des Netzverkehrs	248
4.4.5	Überwachung und Aufzeichnung der Endgeräte per Video	249
4.4.6	Weitere mögliche Datenquellen	250
4.5	Durchführung der Studie	251
4.5.1	Interne Vorabtests und Pretests	251
4.5.2	Testlogistik	252
4.5.3	Testumgebung	253
4.5.4	Technischer Testablauf	254
4.6	Zusammenfassung	256
5	Das Testbed	259
5.1	Mobilfunknetz und mobile Netzsituationen bzw. Netzbedingungen	260
5.1.1	Mobilfunknetze und Netzelemente des Mobilfunks	261
5.1.2	Technologische Ausgangsbasis: Mobilfunkstandards	264
5.1.3	Das Entstehen der Netzbedingungen und Abgrenzung zur E-Technik	266
5.1.4	Notwendige Vereinfachungen zu den Netzsituationen	268
5.1.5	Praxis: QoS-Erhebungen im Mobilfunk	271
5.1.6	Struktur und Festsetzung der Matrix technischer Parameter	275
5.2	Die Einbeziehung des Netzes als Qualitätselement in der Studie	278
5.2.1	Definition der Anforderungen an das Netz	279
5.2.2	Simulation	280
5.2.3	Emulation	281
5.2.4	Echter Mobilfunk	281
5.2.5	Live-Netz und drahtlose Konnektivität	282
5.2.6	Eigener Ansatz	282
5.3	Theorie und verfügbare Lösungen zur Netzemulation	286
5.3.1	Abgrenzung im Stack	286
5.3.2	Perfektes Netz mit kontrolliertem Flaschenhals	288
5.3.3	Globale vs. individuelle Sicht	292
5.3.4	Verfügbare Ansätze und Lösungen zur Netzemulation	293
5.4	Grundlegende technische Realisierung	296
5.4.1	Grundprinzipien der Realisierung	296
5.4.2	Auswahl der Hardware- und Software-Basis	298
5.4.3	Basiskonfiguration	300
5.5	Netzemulation mit tc+netem	302
5.5.1	Egress vs. Ingress und UP vs. DOWN	302
5.5.2	Positionierung	303
5.5.3	Konzept und Komponenten	304
5.5.4	Hierarchical Token Bucket und Klassenbaum	305
5.5.5	netem	308
5.6	Framework: APIs, Web-GUI und Co.	309
5.6.1	Geräte-Cloning	309
5.6.2	Datenhaltung und Konfiguration	310
5.6.3	Netzemulations-Web-API	311
5.6.4	Master Control API	312
5.6.5	Master Control Web-GUI	312

5.6.6	Verknüpfung von Testbed und Fragebogensystem	315
5.6.7	Elektronisches Testlogbuch	318
5.7	Grenzen der realisierten Netzemulation im Testbed	319
5.7.1	Perfekte Netzumgebung und kontrollierter Flaschenhals	319
5.7.2	Hard- und Software der Plattform zur Netzemulation	321
5.7.3	Delay und Jitter	321
5.7.4	Bandbreite und Burstiness	323
5.7.5	Paketverlust.....	324
5.7.6	Umschaltgeschwindigkeit der Parametrisierung	326
5.7.7	Effektivwerte	327
5.8	Die mobilen Endgeräte im Testbed	328
5.8.1	Auswahlkriterien zu den mobilen Endgeräten.....	328
5.8.2	Zusammenstellung und Auswahl der Endgeräte für das Testbed	329
5.8.3	TCP-Implementierungen der Endgeräte	332
5.8.4	Nutzung der Endgeräte im WLAN-Testbed.....	333
5.8.5	Netz Wahrnehmung auf den Endgeräten im Testbed.....	334
5.9	Überwachung und Aufzeichnung des Netzverkehrs	336
5.9.1	Verkehrsausleitung zur Remote-Überwachung.....	336
5.9.2	Deep-Monitoring – der tiefe Blick	339
5.9.3	Monitoring auf den Endgeräten	340
5.10	Überwachung und Aufzeichnung der Endgeräte per Video.....	341
5.10.1	Anforderungen, prinzipielle Möglichkeiten und Einschränkungen	341
5.10.2	Realisierung der Videoüberwachung unter Android	344
5.10.3	Realisierung der Videoüberwachung unter iOS.....	347
5.11	Zusammenfassung.....	349
6	Die Dienste	351
6.1	Die Einbeziehung der Dienste in die Studie	352
6.1.1	Echte vs. unechte Dienste	352
6.1.2	Festlegungen zur Einbeziehung der Elemente der Dienste	353
6.2	Testrelevanz: Die Auswahl relevanter mobiler Dienste	355
6.2.1	Theoretische Relevanz und Überlegungen	355
6.2.2	Praktische Relevanz.....	356
6.2.3	Nutzersicht	357
6.2.4	Netzsicht und Netzbetreibersicht.....	358
6.2.5	Prognosen	359
6.3	Testeignung: Das Prinzip von Testbarkeit und Vergleichbarkeit.....	360
6.3.1	Echtheit	363
6.3.2	Kontrollierbarkeit	364
6.3.3	Beobachtbarkeit.....	366
6.3.4	Benutzbarkeit	367
6.3.5	Wiederholbarkeit	369
6.3.6	Gleichheit	373
6.4	Testeffizienz: Das Prinzip der Bereiche und Abstufungen in der Parametrisierung ...	373
6.4.1	Bildung von Testreihen	374
6.4.2	Festlegung des Parameterbereichs und der Abstufungen	375
6.4.3	Festsetzung und Nutzung der Presets	379
6.5	Art und Reihenfolge in der Darstellung des Testparcours	382
6.5.1	Nutzungsszenario	382
6.5.2	Technische Sicht und Netznutzung – der Blick von innen.....	383

6.5.3	Nutzersicht – der Blick von außen	384
6.5.4	Parametrisierung und Erwartungswerte	385
6.6	Zusammenfassung.....	385
7	Die Probanden	387
7.1	Der Mensch als Proband.....	387
7.1.1	Urteilsfehler und psychologische Effekte	388
7.1.2	Aufgabenausführung.....	389
7.1.3	Kontext und Fokus.....	390
7.2	Allgemeine Punkte zu den Probanden.....	393
7.2.1	Probandenwissen	393
7.2.2	Schriftliche Information, juristische Absicherung und Studienethik	394
7.2.3	Instruktionen und Hilfestellungen	394
7.2.4	Incentive.....	395
7.2.5	Anwerbung, Screening und Auswahl der Probanden	395
7.3	Beschreibung der Stichprobe	396
7.3.1	Annahmen zur Grundgesamtheit	397
7.3.2	Geschlechterverhältnis und Altersstruktur.....	398
7.3.3	Bildung und beruflicher Status	401
7.3.4	Nutzung mobiler Endgeräte	403
7.3.5	Nutzung des mobilen Internets	405
7.4	Nutzung ausgewählter mobiler Dienste durch die Probanden	407
7.4.1	Einführung zur Auswertemethodik.....	407
7.4.2	Aussagen zu den Diensten.....	409
7.4.3	Aussagen zu den Probanden	413
7.5	Weitere Erkenntnisse zu den Probanden	417
7.5.1	Gesamtüberblick Rohergebnisse	417
7.5.2	Inter-Rater-Vergleich.....	418
7.6	Zusammenfassung.....	421
8	Die Auswertung.....	423
8.1	Methodik der Auswertung zum QoS-QoE-Zusammenhang	426
8.1.1	Tiefenauswertung.....	426
8.1.2	Signal-basierte Auswertung.....	430
8.1.3	Parameter-basierte Auswertung	431
8.1.4	Signal-basiert vs. Parameter-basiert.....	432
8.2	Validierung, Aggregation und Partitionierung	433
8.2.1	Validierung und Datenbereinigung	434
8.2.2	Validierung	435
8.2.3	Aggregation vs. Partitionierung	437
8.3	Bearbeitungserfolg	440
8.3.1	Auswertung der Angabe zum Bearbeitungserfolg	441
8.3.2	Auswertung des Grundes für Abbruch oder Wiederholung	442
8.4	Nutzerkommentare	443
8.4.1	Word-Clouds	443
8.4.2	Inhaltliche Analyse der Nutzerkommentare	445
8.5	Bewertungsverhalten	447
8.5.1	Teilbewertungen vs. Gesamtzufriedenheit.....	447
8.5.2	Intra-Rater-Vergleich: Wiederholungsanalyse und Bewertungskonformität.....	449
8.5.3	Zur Visualisierung und Interpretation des Intra-Rater-Vergleichs	452

8.5.4	Allgemeine Erkenntnisse aus dem Intra-Rater-Vergleich	454
8.6	Bedingte Verteilungen der Zufriedenheit	455
8.6.1	Ausgangspunkt der statistischen Beschreibung und Analyse	456
8.6.2	Feinheiten und Grenzen der Statistik in der Auswertung	457
8.6.3	Skalen als Zahlen: Skala und Annahmen zu den Eigenschaften der Bewertungen	459
8.6.4	Aussagen zur Grundgesamtheit – Interferenzstatistik	463
8.6.5	Vom Schätzen, Testen und Analysieren.....	464
8.6.6	Metriken in der Auswertung von Nutzertests	466
8.6.7	Statistische Maßzahlen: Mittelwert & Co.	467
8.6.8	Anteilswerte, statistisches Modell und Annahmen zur Verteilung	469
8.6.9	QoE-Metriken: Mean Opinion Score (MOS)	472
8.6.10	Mittelwerte oder Anteilswerte? Beides!	474
8.7	Ergebnisdarstellung und Visualisierung der bedingten Verteilungen	478
8.7.1	Parameterraum und Ergebnisraum	478
8.7.2	Symbole und Abkürzungen.....	479
8.7.3	Visuelles Arrangement und Ergebnisgegenüberstellung	479
8.7.4	Anteilswerte der Bewertungen	481
8.7.5	Mittelwertsäulen der Bewertungen	483
8.7.6	Mittelwertkurven der Bewertungen.....	483
8.8	Prinzip der Schwellwerte	486
8.8.1	Motivation zur Suche der Schwellwerte	486
8.8.2	Wahl der Schwellen	488
8.8.3	Bezugsmetrik und Wertebereiche der Schwellen	489
8.8.4	Ermittlung auf Basis der Stichprobe oder Grundgesamtheit	493
8.8.5	Berücksichtigung der QoS-QoE-Charakteristik	493
8.8.6	Bestimmung der Schwellwerte	494
8.8.7	Interpretation der Schwellwerte	496
8.9	Güte- und Fehlerabschätzung.....	496
8.9.1	Objektivität.....	497
8.9.2	Reliabilität	498
8.9.3	Validität	498
8.9.4	Einschätzung des empirischen Untersuchungsdesigns.....	499
8.9.5	Generelle Fehlerabschätzung	500
8.10	Zusammenfassung.....	500
9	Der Testparcours	503
9.1	Drive	504
9.1.1	Nutzungsszenario	504
9.1.2	Technische Sicht und Netznutzung.....	505
9.1.3	Nutzersicht	506
9.1.4	Parametrisierung und Erwartungswerte	508
9.1.5	Bearbeitungserfolg	508
9.1.6	Nutzerkommentare	509
9.1.7	Bewertungsverhalten	510
9.1.8	Bedingte Verteilungen der Zufriedenheit	513
9.2	Facebook	514
9.2.1	Nutzungsszenario	514
9.2.2	Technische Sicht und Netznutzung.....	514
9.2.3	Nutzersicht	516

9.2.4	Parametrisierung und Erwartungswerte	518
9.2.5	Bearbeitungserfolg	519
9.2.6	Nutzerkommentare	520
9.2.7	Bewertungsverhalten	521
9.2.8	Bedingte Verteilungen der Zufriedenheit	521
9.3	Google Maps	524
9.3.1	Nutzungsszenario	524
9.3.2	Technische Sicht und Netznutzung	524
9.3.3	Nutzersicht	526
9.3.4	Parametrisierung und Erwartungswerte	527
9.3.5	Bearbeitungserfolg	529
9.3.6	Nutzerkommentare	530
9.3.7	Bewertungsverhalten	531
9.3.8	Bedingte Verteilungen der Zufriedenheit	535
9.4	MTV-Music	536
9.4.1	Nutzungsszenario	536
9.4.2	Technische Sicht und Netznutzung	536
9.4.3	Nutzersicht	537
9.4.4	Parametrisierung und Erwartungswerte	539
9.4.5	Bearbeitungserfolg	540
9.4.6	Nutzerkommentare	540
9.4.7	Bewertungsverhalten	542
9.4.8	Bedingte Verteilungen der Zufriedenheit	547
9.5	Spiegel.de	547
9.5.1	Nutzungsszenario	547
9.5.2	Technische Sicht und Netznutzung	548
9.5.3	Nutzersicht	553
9.5.4	Parametrisierung und Erwartungswerte	554
9.5.5	Bearbeitungserfolg	554
9.5.6	Nutzerkommentare	555
9.5.7	Bewertungsverhalten	556
9.5.8	Bedingte Verteilungen der Zufriedenheit	556
9.6	Youtube	559
9.6.1	Nutzungsszenario	559
9.6.2	Technische Sicht und Netznutzung	560
9.6.3	Nutzersicht	566
9.6.4	Parametrisierung und Erwartungswerte	567
9.6.5	Bearbeitungserfolg	568
9.6.6	Nutzerkommentare	570
9.6.7	Bewertungsverhalten	570
9.6.8	Bedingte Verteilungen der Zufriedenheit	572
9.6.9	Visuelle Qualität des Videoinhalts	576
9.7	Zusammenfassung und Ergebnisüberblick	578
9.7.1	Auswahl	578
9.7.2	Technische Sicht und Netznutzung	579
9.7.3	Nutzersicht und QoS-QoE-Charakteristik	580
9.7.4	Parametrisierung	581
9.7.5	Besonderheiten der Szenarien	582
9.7.6	Versionen der mobilen Apps	582
9.7.7	Übersicht zum Bearbeitungserfolg	584

9.7.8 Nebenerkenntnisse	584
9.7.9 Überblicksdarstellung der Ergebnisse	586
9.7.10 Schwellwerte	588
10 Das Fazit	591
10.1 Der Blick zurück: Zusammenfassung und Zielerreichung	591
10.1.1 Methodischer Ansatz	591
10.1.2 Theorie und Modellierung	592
10.1.3 Testbed und Technik	593
10.1.4 Voruntersuchungen und Vorüberlegungen zu den Diensten	594
10.1.5 Empirie und QoE-Evaluation	595
10.1.6 Auswertung des Nutzertests	596
10.1.7 Erzielte Ergebnisse des Nutzertests	596
10.1.8 Was bleibt?	597
10.2 Der Blick voraus: Ausblick und offene Punkte	598
Literaturverzeichnis	601
A Anhang	627
A.1 QoS und QoE	627
A.2 Ende-zu-Ende-Messungen im Testbed	628
A.3 Testparcours	631
A.4 Fragebogen und Sicht der Nutzer	634
A.4.1 Screening-Fragebogen	635
A.4.2 Startseite des Fragebogens	636
A.4.3 Aufbau des Fragebogens	636
A.4.4 Bearbeitungserfolg	637
A.4.5 Bewertung Gesamtzufriedenheit und Anmerkungen	638
A.4.6 Drive	639
A.4.7 Facebook	642
A.4.8 Google Maps	645
A.4.9 MTV-Music	648
A.4.10 Spiegel.de	651
A.4.11 Youtube	654
A.4.12 Nachbefragung	657
A.5 Rohergebnisse	660
A.5.1 Übersichtstafeln	660
A.5.2 Anzahl Bearbeitungen, Bearbeiter und Bearbeitungserfolg	670
A.5.3 Zusatzanmerkungen zur Validierung	672
A.6 Zusatzkenntnisse	673
A.6.1 Netzwahrnehmung durch die Probanden	673
A.6.2 Wahrnehmung der Qualität des Inhalts	674
A.7 Credits und Disclaimer	676

Tabellenverzeichnis

Tabelle 1: Eckdaten der ausgeführten Studien im Überblick.....	27
Tabelle 2: Zusammengefasste Darstellung möglicher QoS-Klassen in 3G-Mobilfunknetzen	59
Tabelle 3: Arten von Zufriedenheitsfaktoren und ihre Auswirkungen	85
Tabelle 4: Nutzungsklassen aus Nutzungszielen und Nutzungsarten	92
Tabelle 5: Charakterisierung von QoS und QoE zum Verständnis innerhalb der Arbeit	111
Tabelle 6: Charakterisierung der Schichten des QoS-QoE-Zusammenhangs	121
Tabelle 7: Zusammenfassung zum möglichen Einfluss der Hardware und Software der mobilen Endgeräte im Modell	131
Tabelle 8: Zusammenfassung zum möglichen Einfluss der Anwendung/App im Modell.....	143
Tabelle 9: Stufen der zusammengehörigen Netznutzungen.....	159
Tabelle 10: Kategorien der Nutzerstimuli mit Beispielen	165
Tabelle 11: Qualität und Qualität unterschiedlicher Medien des Inhalts - objektiv vs. subjektiv	168
Tabelle 12: Zeiträume und Sequenzialität von Übertragung und Nutzung in der Nutzersicht ..	171
Tabelle 13: Symbolik zur Charakterisierung von Einflussrelevanz und Einflussrichtung.....	189
Tabelle 14: Charakterisierung der genannten Grundformen zum QoS-QoE-Zusammenhang ..	192
Tabelle 15: Klassenbildung aus Elastizität und Interaktivität mit Beispieldiensten	195
Tabelle 16: Dienstklassen und relevante Unterscheidungskriterien	201
Tabelle 17: 5-Punkte-MOS-Skala	227
Tabelle 18: Klassifizierung des Zusammenhangs zwischen QoS (Parameterraum) und QoE (Bewertungsraum)	230
Tabelle 19: Variablenplan der Einflussgrößen und Identifikationsmerkmale im Test	232
Tabelle 20: Blockbildung in der Zusammenstellung der Aufgabenabfolgen und ihrer Varianten	240
Tabelle 21: Variablenplan der Zielgrößen und Ergebnisraumdaten	243
Tabelle 22: Charakterisierung und Klassifizierung der Studie.....	257
Tabelle 23: Netzgenerationen und theoretische QoS-Netzparameter im Mobilfunk	265
Tabelle 24: Netztests in Mobilfunknetzen in Deutschland	272
Tabelle 25: Art und Richtung der Beeinflussung von Netzparametern in der Netzemulation (Auszug).....	291
Tabelle 26: Übersicht zu einer Auswahl möglicher Open-Source-Lösungen zur Netzemulation	293
Tabelle 27: Technologien und max. Bruttogeswindigkeiten der Netzwerkschnittstellen	300
Tabelle 28: Verwendung der Netzschnittstellen und Richtungen des Netzwerkverkehrs in der Netzemulation	301
Tabelle 29: Die mobilen Endgeräte des Testbeds der Testrunde 2 mit Testplatzzuordnung und technischen Eckdaten	330
Tabelle 30: Versionen der führenden mobilen Plattformen.....	331
Tabelle 31: Ort und Art der Sichtbarkeit der Anzeigen zur Netztechnologie und zum Betreiber	336
Tabelle 32: Geschlechterverhältnis und Altersstruktur der Probanden der Testrunde 2	398
Tabelle 33: Generationsstruktur der Probanden in Testrunde 2	400
Tabelle 34: Höchster Bildungsabschluss der Probanden	401
Tabelle 35: Beruflicher Status der Probanden.....	402
Tabelle 36: Absolute Häufigkeit der von den Probanden privat genutzten Geräte nach Plattform und Formfaktor.....	404

Tabelle 37: Anteile der mobilen Plattformen der Probanden im Vergleich zum Markt in Deutschland	404
Tabelle 38: Vergleichsdaten zur Nutzung mobiler Endgeräte in Deutschland	405
Tabelle 39: Nutzung des mobilen Internets nach Formfaktor der Endgeräte	406
Tabelle 40: Von den Probanden genutzte Mobilfunknetze im Vergleich zu den Marktzahlen ..	406
Tabelle 41: Erfahrung der Probanden in der Nutzung des mobilen Internets (nach Formfaktor)	407
Tabelle 42: Werttransformation der Häufigkeitsskala zur Auswertung	408
Tabelle 43: Nutzungshäufigkeit ausgewählter mobiler Dienste in Testrunde 2	410
Tabelle 44: Veränderung der Nutzungsintensität ausgewählter mobiler Dienste zwischen 2012 und 2013	412
Tabelle 45: Verteilung der Modalwerte und Mediane der Nutzungshäufigkeit mobiler Dienste zwischen den Probanden	413
Tabelle 46: Nutzertypen nach Nutzungsintensität und jeweilige Anteile der Probanden	415
Tabelle 47: Übersicht zu den Probanden mit Häufigkeit des Auftretens der Bewertungen	418
Tabelle 48: Erwartungskonformität der Bewertungsänderung beim Wiederholungsvergleich ..	452
Tabelle 49: Notwendige Annahmen und Möglichkeiten zur statistischen Beschreibung und Analyse	458
Tabelle 50: Vergleich der Darstellungen der Mittelwertkurven mit unterschiedlicher Skalierung der x-Achse	485
Tabelle 51: Schwellen, Kriterien, Erläuterungen, Bezugsmetriken und sinnvolle Wertbereiche	492
Tabelle 52: Parametrisierung und Erwartungswerte für Drive	507
Tabelle 53: Sentiment-Analyse in Drive	509
Tabelle 54: Parametrisierung und Erwartungswerte für Facebook	519
Tabelle 55: Datenvolumen je Aufgabenausführung in Abhängigkeit von App und Gerät in Google Maps	525
Tabelle 56: Parametrisierung und Erwartungswerte für Google Maps	529
Tabelle 57: Parametrisierung und Erwartungswerte für MTV-Music	539
Tabelle 58: Parametrisierung und Erwartungswerte für Spiegel.de	553
Tabelle 59: Qualitätsstufen und technische Kenndaten des Inhalts für Youtube (mobil)	563
Tabelle 60: Parametrisierung und Erwartungswerte für Youtube	568
Tabelle 61: Übersicht zur Kategorie und Auswahl der mobilen Dienste im Testparcours der Testrunde 2	578
Tabelle 62: Netznutzung der mobilen Dienste im Testparcours	579
Tabelle 63: Klassifizierung und relevante Stimuli der mobilen Dienste im Testparcours	580
Tabelle 64: Besonderheiten zu den Szenarien der mobilen Dienste im Testparcours	583
Tabelle 65: Übersicht zu den genutzten Versionen der mobilen Apps im Probandentest	583
Tabelle 66: Schwellwerte der untersuchten Dienstszenarien im Test (bezogen auf die Grundgesamtheit)	589
Tabelle 67: Endgerätekategorien und maximale Datenraten	627
Tabelle 68: Ende-zu-Ende-Messungen für die ausgewählten Netzparameterkombinationen im Testbed unter Android	628
Tabelle 69: Ende zu Ende Messungen für die ausgewählten Netzparameterkombinationen im Testbed unter iOS	629
Tabelle 70: QoS-Netzparameterwerte der Presets	632
Tabelle 71: Effektivwerte Datenrate Download in Drive	633
Tabelle 72: Effektivwerte Datenrate Upload in Drive	633
Tabelle 73: Übersicht zum Bearbeitungserfolg, Problemen, Anzahl Bearbeitungen, Bearbeitern und validen Bewertungen	671

Abbildungsverzeichnis

Abbildung 1:	Der Zusammenhang von Mensch, Maschine und Markt in den Betrachtungen	4
Abbildung 2:	Innovationszyklus von Netzleistung und Nutzung sowie „Fenster der knappen Betriebsmittel“	9
Abbildung 3:	Modellierung des Zusammenhangs – Von der Black Box zur White Box bzw. Grey Box	19
Abbildung 4:	Vergleich verschiedener Ansätze, unterschieden nach Realismus und benötigtem Aufwand	23
Abbildung 5:	Schematische Übersicht zum Aufbau der Arbeit und zum Zusammenhang der einzelnen Abschnitte	29
Abbildung 6:	Zusammensetzung der QoS aus Network Performance und Non Network Performance	47
Abbildung 7:	KPI-Hierarchieebenen im Mobilfunk	49
Abbildung 8:	Schematische Darstellung des Netzes und unterschiedliche Sichten auf „das Netz“	51
Abbildung 9:	Kategorien von QoE-Einflussfaktoren	71
Abbildung 10:	Mögliche Zusammenhänge der Qualitätsaspekte zur QoE	79
Abbildung 11:	QoE-Teilmodell zur Zufriedenheit als Waage zum Abgleich von Erlebnis und Erwartung	83
Abbildung 12:	Klassische Sichtweise und 2-Faktoren-Sichtweise zur Zufriedenheit	85
Abbildung 13:	Verschiedene Arten von Kontext	89
Abbildung 14:	QoS-Schichtmodelle/Klassifikation im Vergleich (Orientierung am Stack)	97
Abbildung 15:	QoS-Schichtung (indirekte Stack-Orientierung)	99
Abbildung 16:	Von Ende-zu-Ende QoS zu Ende-zu-Ende-QoE	101
Abbildung 17:	Zusammenhang von Network Performance, Quality of Service und Quality of Experience	102
Abbildung 18:	Zusammenstellung zur Nutzersicht auf das Netz und die Netzeigenschaften	105
Abbildung 19:	Qualität als Beschaffenheit, Beurteilung und Eignung zum Zweck im Zusammenhang von QoS und QoE	109
Abbildung 20:	Das Modell zum Zusammenwirken der Elemente, aufgetrennt in die gebildeten Schichten	122
Abbildung 21:	Erzeugung des QoS-QoE-Zusammenhangs	126
Abbildung 22:	QoX-Matrix mit Mediatoren und orthogonalen Moderatoren	129
Abbildung 23:	Die Gegenseite als komplexes System und als Abstraktion	145
Abbildung 24:	QoD – von Stack zu Stack durch das Netz	149
Abbildung 25:	Populäre Dienste und dafür typische Protokolle auf den Protokollschichten des TCP/IP-Stacks	151
Abbildung 26:	Zusammenhang und gegenseitige Beeinflussung ausgewählter leistungsbezogener Parameter zur QoD	153
Abbildung 27:	Zusammenhang der Stimuli und sich ergebender Zielkonflikt am Beispiel von Streaming	175
Abbildung 28:	Zusammenstellung von Beispielen zur Präsentation	179
Abbildung 29:	Complexity of Experience und interpersonelle Bewertungsunterschiede der Dienste	183
Abbildung 30:	Von Anforderungen und Einschränkungen zum dienstspezifischen QoS-QoE-Zusammenhang	186

Abbildung 31: Angenommene Dienstkurven zum QoS-QoE-Zusammenhang.....	191
Abbildung 32: Die Größenordnungen relevanter Zeitmarken und die Rolle der Zeit in verschiedenen Sichten	203
Abbildung 33: Aufschlüsselung zu den Methoden und Techniken zur QoE-Erfassung und QoE-Bewertung	220
Abbildung 34: Beispiel der genutzten Skala zur Erfassung der Zufriedenheitsbewertungen .	246
Abbildung 35: Bilder aus der Testumgebung	253
Abbildung 36: Netzsegmente und Netzelemente des Mobilfunks zur Datenkommunikation und vereinfachte Sicht	263
Abbildung 37: Steigerung der Kapazität durch Sektorisierung und Netzverdichtung	267
Abbildung 38: Schematische Darstellung der Netze und unterschiedliche Sichten auf „das Netz“	269
Abbildung 39: Netzparameterstufen als 3-Tupel in der Matrix technischer Parameter.....	277
Abbildung 40: Skizze zur Einordnung der Netzemulation in das technische Grundkonzept ..	283
Abbildung 41: Einordnung des Testbeds mit Netzemulation in das QoX-Matrix-Modell	284
Abbildung 42: Modell zur Netzemulation als kontrollierter Flaschenhals zwischen externem und internem Netz	289
Abbildung 43: Schnittstellen der Netzkonfiguration im Blockschaltbild	301
Abbildung 44: Kombinierte Netzemulation für UP und DOWN.....	304
Abbildung 45: Klassenbasierter Baum zur Verkehrsdifferenzierung und -manipulation in der Netzemulation	307
Abbildung 46: Screenshot der Web-GUI (Ausschnitt)	313
Abbildung 47: Ausschnitt Web-GUI mit Warnungen zu detektierten Fehlern im Testplan	313
Abbildung 48: Anbindung des externen Fragebogensystems zur Testlaufsteuerung an das Testbed.....	314
Abbildung 49: Erfassung eines Eintrags für das elektronische Testlogbuch über ein Online-Formular.....	318
Abbildung 50: Die mobilen Endgeräte des Testbeds.....	331
Abbildung 51: Wirkung mit Dauer-Ping im WLAN-Testbed zur Verbesserung der Einhaltung der Zielwerte.....	335
Abbildung 52: Wahrnehmbare Netzigenschaften in der System- bzw. Statusleiste im Testbed mit WLAN	335
Abbildung 53: Hinweise und Anzeigen zur fehlenden SIM-Karte bei den Android-Geräten im Benachrichtigungsbereich.....	335
Abbildung 54: Positionierung der optionalen Verkehrsausleitung an den Netzschnittstellen	337
Abbildung 55: Spiegelung der Bildschirmausgaben unter Android, notwendige zusätzliche Hardware und Software	346
Abbildung 56: Übersicht zur Ermittlung der Relevanz mobiler Anwendungen und Dienste ..	354
Abbildung 57: Von den Facetten der Bewertung der Testbarkeit und Vergleichbarkeit zur Feststellung der Testeignung	361
Abbildung 58: Reihenfolgenabhängigkeit durch Veränderung der Vorbedingungen	371
Abbildung 59: Mögliche Ansätze zur Reihenbildung in der Netzparametrisierung	374
Abbildung 60: Beispiele für die Auswahl von 5 Presets aus der Matrix technischer Parameter.....	376
Abbildung 61: Bildung der spezifischen Preset-Bezeichnungen	380
Abbildung 62: Bildung neuer Preset-Stufen bei einer möglichen Erweiterung oder Verfeinerung des Testbereiches	380
Abbildung 63: Höchster Schulabschluss der Probanden im Vergleich	401

Abbildung 64:	Dichte und Verteilung der Mittelwerte der Häufigkeit der Nutzung mobiler Dienste je Proband	415
Abbildung 65:	Summierte Nutzungshäufigkeit mobiler Dienste je Proband mit weiteren Attributen	416
Abbildung 66:	Inter-Rater-Vergleich als Rangverteilung mit zugehörigen Bewertungstendenzen.....	420
Abbildung 67:	Übersicht zu den Bestandteilen und Schritten der Auswertung	424
Abbildung 68:	Mögliches Vorgehen bei der Tiefenauswertung der Verkehrsdaten ex post ..	428
Abbildung 69:	Analyse der Netzverkehrsdaten für den gesamten Testlauf eines Probanden	429
Abbildung 70:	Prinzip der Verarbeitung der Rohergebnisse durch Validierung, Aggregation und Partitionierung.....	434
Abbildung 71:	Visualisierung der Nutzerkommentare als Word-Cloud, Beispiel Drive	444
Abbildung 72:	Vergleich der Gesamtzufriedenheit zu den Teilbewertungen am Beispiel von Drive.....	448
Abbildung 73:	QoS-QoE-Differenzmatrix am Beispiel der Gesamtzufriedenheit für Drive.....	453
Abbildung 74:	5-stufige Skala als Bewertungsskala für die Zufriedenheit mit zugeordneten Scores und Kennfarben.....	462
Abbildung 75:	5-stufige Skala und die vier gebildeten gerichteten dichotomen Aggregationen	471
Abbildung 76:	Beispiel für Varianten unterschiedlicher Anteilswerte und zugehörige wertgleiche Mittelwerte	475
Abbildung 77:	Zusammenhang von Mittelwert und Standardabweichung auf einer 5-stufigen diskreten Skala.....	477
Abbildung 78:	Standardabweichung zu Mittelwert, Beispieldarstellung für zwei ausgewählte Dienste.....	477
Abbildung 79:	Beispiel für die Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung	480
Abbildung 80:	Beispiel zu den Darstellungen der Mittelwertkurven mit unterschiedlicher Skalierung der x-Achse.....	484
Abbildung 81:	Relevante Bereiche und Lagebeziehung der Schwellen im Bewertungsraum und Parameterraum	487
Abbildung 82:	Bereichsschwellen (MOS und Anteile) aus den synthetischen R-Values des E-Modells.....	488
Abbildung 83:	Schwellwerte im Bewertungsraum und Parameterraum.....	495
Abbildung 84:	Word-Cloud aus den Nutzerkommentaren zu Drive	509
Abbildung 85:	Zusammenstellung zum Bewertungsverhalten in Drive.....	511
Abbildung 86:	Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Drive	512
Abbildung 87:	Meldungen zu neueren Versionen der Facebook-App unter Android im Verlauf der Testrunde 2	516
Abbildung 88:	Word-Cloud aus den Nutzerkommentaren zu Facebook	520
Abbildung 89:	Zusammenstellung zum Bewertungsverhalten in Facebook.....	522
Abbildung 90:	Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Facebook.....	523
Abbildung 91:	Word-Cloud aus den Nutzerkommentaren zu Google Maps	530
Abbildung 92:	Zusammenstellung zum Bewertungsverhalten in Google Maps	532
Abbildung 93:	Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Google Maps (Teil 1).....	533
Abbildung 94:	Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Google Maps (Teil 2).....	534

Abbildung 95: Word-Clouds aus den Nutzerkommentaren zu MTV-Music.....	541
Abbildung 96: Zusammenstellung zum Bewertungsverhalten in MTV-Music.....	543
Abbildung 97: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 1).....	544
Abbildung 98: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 2).....	545
Abbildung 99: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 3).....	546
Abbildung 100: Verschiedene Arten des Zugangs und der Inhaltsversionen von Spiegel.de im Vergleich.....	548
Abbildung 101: Waterfall-Diagramm zum Download der einzelnen Seiten und Elemente in Spiegel.de.....	549
Abbildung 102: Zeitliche Bestandteile beim Browsing (Downloads und Darstellung eines Inhaltselements im Browser).....	549
Abbildung 103: Word-Cloud aus den Nutzerkommentaren zu Spiegel.de.....	555
Abbildung 104: Zusammenstellung zum Bewertungsverhalten in Spiegel.de.....	557
Abbildung 105: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Spiegel.de.....	558
Abbildung 106: Aufschlüsselung zur Vielfalt der technischen Möglichkeiten bei der mobilen Nutzung von Youtube.....	560
Abbildung 107: Zeitlicher Verlauf und Durchschnittswerte der Inhaltsdatenraten der ausgewählten Inhalte in Youtube.....	564
Abbildung 108: Word-Clouds aus den Nutzerkommentaren zu Youtube.....	569
Abbildung 109: Zusammenstellung zum Bewertungsverhalten in Youtube.....	571
Abbildung 110: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Youtube (Teil 1).....	573
Abbildung 111: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Youtube (Teil 2).....	574
Abbildung 112: Zuordnung der Presets zu ausgewählten Netzparameterstufen.....	581
Abbildung 113: Zusammenstellung der Bewertungen der Zufriedenheit für alle Dienste im Testparcours.....	587
Abbildung 114: Screenshots der Apps für die Ende-zu-Ende-Messungen im Testbed (iOS).....	630
Abbildung 115: Startansicht mit Position der Android-Systemleiste und iOS-Statusleiste auf den Endgeräten.....	631
Abbildung 116: Startseite zum Fragebogen.....	636
Abbildung 117: Schematischer Aufbau einer Fragebogenseite.....	637
Abbildung 118: Mögliche Angaben zum Bearbeitungserfolg und Grund im Fragebogen.....	637
Abbildung 119: Gemeinsamer Fragenbogenteil zu jeder Aufgabe.....	638
Abbildung 120: Spezifischer Fragenbogenteil zum Szenario Drive.....	639
Abbildung 121: Zusammenstellung zur Nutzersicht im Szenario Drive.....	640
Abbildung 122: Spezifischer Fragenbogenteil zum Szenario Facebook.....	642
Abbildung 123: Zusammenstellung zur Nutzersicht im Szenario Facebook.....	643
Abbildung 124: Spezifischer Fragenbogenteil zum Szenario Google Maps.....	645
Abbildung 125: Zusammenstellung zur Nutzersicht im Szenario Google Maps.....	646
Abbildung 126: Spezifischer Fragenbogenteil zum Szenario MTV-Music.....	648
Abbildung 127: Zusammenstellung zur Nutzersicht im Szenario MTV-Music.....	649
Abbildung 128: Spezifischer Fragenbogenteil zum Szenario Spiegel.de.....	651
Abbildung 129: Zusammenstellung zur Nutzersicht im Szenario Spiegel.de.....	652
Abbildung 130: Spezifischer Fragenbogenteil zum Szenario Youtube.....	654
Abbildung 131: Zusammenstellung zur Nutzersicht im Szenario Youtube.....	655

Abbildung 132: Nachbefragung im Fragebogen Teil 1	657
Abbildung 133: Nachbefragung im Fragebogen Teil 2	658
Abbildung 134: Nachbefragung im Fragebogen Teil 3	659
Abbildung 135: Rohergebnisüberblick zu allen Bewertungen der Zufriedenheit insgesamt je Aufgabenausführung	662
Abbildung 136: Rohergebnisüberblick zu allen Bewertungen der Teilbewertung 1 je Aufgabenausführung	664
Abbildung 137: Rohergebnisüberblick zu allen Bewertungen der Teilbewertung 2 je Aufgabenausführung	666
Abbildung 138: Rohergebnisüberblick zu allen Bewertungen der Teilbewertung 3 je Aufgabenausführung	668

1 Die Einleitung

Von der Motivation über die Zielsetzung zu Methodik und Rahmen der Arbeit

In der vorliegenden Arbeit wird das Nutzererlebnis im mobilen Internet thematisiert. Dabei existiert ein Zusammenhang zwischen Quality of Service (QoS) und Quality of Experience (QoE), der sich sowohl konzeptuell als auch kausal beschreiben lässt. Dieser Zusammenhang ist in seiner Komplexität nur schwierig in allen relevanten Belangen zu verstehen und nur aufwendig in konkreten Zahlen zu fassen. An dieser Stelle setzt die Arbeit durch eine passende Modellierung und empirische Untersuchung an.

Das Kapitel Einleitung gliedert sich in folgende Abschnitte, die jeweils einen Aspekt zur Thematik behandeln:

- **Motivation:** Wie ist die Arbeit einzuordnen, worum geht es genau und warum beschäftigt sich diese Arbeit mit dem Thema? Wen kann und sollte das Thema interessieren?
- **Zielstellung:** Welche prinzipiellen und konkreten Ziele sollten mit dieser Arbeit erreicht werden?
- **Methodik und Rahmen:** Wie war der Weg, um ausgehend von der Motivation die gestellten Ziele zu erreichen? Wie war der Ausgangspunkt?
- **Aufbau der Arbeit:** Erläuterungen zu den nachfolgenden Kapiteln

Zu Anfang ist ggf. auch hilfreich klarzustellen, was die Arbeit vorrangig nicht ist (Diskussion um Netzneutralität) bzw. welche Aspekte nur soweit gestreift werden, wie es die Einordnung zum Gesamtverständnis erfordert:

- Behandlung von QoS-Mechanismen
- Erklärung der Funktion komplexer IP-Netze/Mobilfunknetze
- Erklärung des Zustandekommens typischer Netzsituationen im Mobilfunk

1.1 Motivation

Die Nutzung des mobilen Internets hat in den letzten Jahren an Umfang und Bedeutung gewonnen. Das Verständnis von QoS und QoE ist Voraussetzung um die Nutzer mobiler Netze und Dienste zufriedenzustellen oder gar zu begeistern sowie Voraussetzung für eine effiziente Nutzung der Netzressourcen (nach [1] S. 1). In der Motivation soll entsprechend dargelegt werden, warum dem so ist und weshalb dazu noch immer Forschungsbedarf besteht. Dazu wird erörtert, wieso die Betrachtung zu QoS und QoE interdisziplinär erfolgen sollte und wie Mensch, Maschine und Markt sich in das Gesamtbild einfügen. Aus der Diskussion der Zusammenhänge zwischen diesen drei M ergeben sich jeweils einige Problemfelder, die in einer gemeinsamen Betrachtung die Motivation zur Thematik formen. Neben der Beantwortung der Frage zum Warum ergibt sich auch die Frage, für Wen die Beschäftigung mit der Thematik interessant ist.

Für QoS und QoE gibt es dabei verschiedene Interessenten (Stakeholder) und Interessen (in Anlehnung an [2] S. 20):

- Nutzer/Kunde: Nutzung mobiler Netze und Dienste
- Netzbetreiber: Planung, Errichtung und Betrieb von (mobilen) Netzen
- Dienstanbieter¹: Anbieter (mobiler) Web-Seiten, Apps und Inhalte
- Hersteller von Netzwerkausrüstungen (Hardware und zugehöriger Software)
- Hersteller von (mobilen) Endgeräten

1.1.1 Multidisziplinäres Problem und interdisziplinäre Betrachtung

Für QoS wie auch QoE gilt: Es gibt nicht nur einen einzigen Erklärungsansatz, sondern viele unterschiedliche Sichtweisen, Aspekte und Konzepte. Bei den recherchierten Erklärungsansätzen zur QoS und QoE ist deutlich zu erkennen, aus welcher Denkrichtung sie kommen, bzw. an welchem Ende sie ansetzen. Besonders QoE ist dabei ein multidisziplinäres Problem. Um die Unterschiede zu verdeutlichen, sind die nachfolgenden Punkte absichtlich überspitzt:

¹ entweder als Anbieter von *managed Services*, wenn der Serviceanbieter gleichzeitig auch als Betreiber des Netzes agiert, über das der Service erbracht wird, oder als Anbieter von *Over-the-Top-Services* über fremde Netze bzw. das Internet

- aus der Tiefe der Technik, mit dem Nutzer als Anhängsel am oberen Ende des Netzwerk-Stacks: von den Netztechnikern & Co.
- aus der Tiefe der menschlichen Psyche, das Innere des menschlichen Wesens, seine Fähigkeiten und Bedürfnisse im Kontakt mit der Technik von außen: von den Psychologen & Co.
- aus der Absicht durch das Betreiben von Netzen und das Anbieten und Verkaufen von Diensten an Kunden als Geschäftsmodell Geld zu verdienen: von den (scharf kalkulierenden) Betriebswirtschaftlern & Co

Und wo ordnet sich die Informatik ein, speziell die Medieninformatik? Sie ist irgendwo in der Mitte des Zusammenwirkens von Mensch und Maschine in Form von Nutzern mobiler Netze, mobiler Geräte und Apps mit Mensch-Maschine-Interaktion zu finden. Diese Position in der Mitte mit einem ausgewogenen Blick sowohl auf die Nutzerseite als auch auf die Technikseite ist dabei sicherlich nicht die schlechteste. Bei näherer Beschäftigung mit der Thematik wird schnell klar, dass die Auseinandersetzung mit QoS und QoE nicht nur ein multidisziplinäres Problem ist, sondern einer interdisziplinären Betrachtung bedarf.

Akademische Kooperation und eigener Anteil: Die Arbeit entstand im Rahmen eines privatwirtschaftlich geförderten interdisziplinären Projektes, das an der Professur Medieninformatik (MI) der Technischen Universität Chemnitz in Kooperation mit weiteren Partnern der TU Chemnitz, der Professur Allgemeine Psychologie und Arbeitspsychologie (AAP) sowie der Professur Schaltkreis- und Systementwurf (SSE), ausgeführt wurde. Verweise zu Art und Umfang der Zusammenarbeit finden sich in den jeweiligen Abschnitten, die dies betrifft, ansonsten erfolgte die Bearbeitung durch den Verfasser.

1.1.2 Maschine, Mensch, Markt: schneller, besser, billiger

Die Motivation zur Beschäftigung mit dem Nutzererlebnis im mobilen Internet entsteht aus dem Spannungsfeld im Dreiecksverhältnis zwischen Mensch, Maschine und Markt, kurz den drei Ms, woraus sich verschiedene Perspektiven und Problemstellungen ableiten lassen, siehe Abbildung 1. Jede der Ecken kann dabei mit unterschiedlichen Maßen und Einheiten charakterisiert und bewertet werden – als Qualität – und verfolgt ein eigenes Ziel, die zusammen zu „schneller, besser, billiger“ führen. Auf Definitionen dazu wird an dieser

Stelle explizit verzichtet, diese werden im Kapitel 2 präsentiert. Für das weitere Verständnis im Rahmen der Einleitung sollten die folgenden Erläuterungen aber genügen.

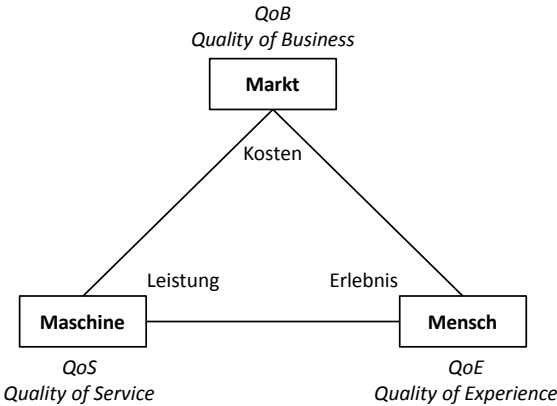


Abbildung 1: Der Zusammenhang von Mensch, Maschine und Markt in den Betrachtungen zusammengefasst, erweitert und ergänzt in Anlehnung an [3] S. 30 und [4] S. 59

Maschine soll nachfolgend als Gesamtheit aller technischen Aspekte und Komponenten verstanden werden. Eine sinnvolle Zergliederung dazu wird im Weiteren erarbeitet. Das mobile Netz sowie die Netzsituationen und weitere technische Faktoren sind hier einzuordnen. Eine alternative Bezeichnung wäre System, da die Erklärungen dazu oftmals auf komplexen technischen Systemen aus der Kombination von Hardware und Software basieren. *QoS* bezieht sich auf Syntax und Mechanismen zur Beschreibung und Verbesserung der Leistung aus der Netzperspektive: schnell vs. langsam mit dem Ziel **schneller**.

Mensch steht in dieser Betrachtung für die Vereinigung von zwei Rollen: dem Nutzer (hin zur Maschine) und dem Kunden (hin zum Markt) mit seinen Zielen, Anforderungen und weiteren menschlichen Faktoren. *QoE* bezieht sich auf das Erlebnis bei der Nutzung (multimedialer) Dienste aus der Nutzerperspektive: gut vs. schlecht mit dem Ziel **besser**.

Markt steht an dieser Stelle stellvertretend für das Anbieten und Nachfragen von Diensten, was (mobile) Kommunikationsnetze, die Anbieter entsprechender Dienstleistungen und nichttechnische Angebotsaspekte (Preis, Service, ...)

sowie den Wettbewerb an sich einschließt. Quality of Business (*QoB*) steht hier einerseits für die Kosten des Anbieters zur Errichtung und zum Betrieb eines Netzes mit einer bestimmten Netzleistung und andererseits für die Kosten des Kunden, die einem bestimmten Erlebnis gegenüberstehen: teuer vs. billig mit dem Ziel **billiger**.

Dabei wird von folgenden grundsätzlichen Zusammenhängen der drei M ausgegangen:

- Maschine und Mensch: schneller ist besser
- Maschine und Markt: schneller ist teurer
- Markt und Mensch: billiger ist besser

Das Dreieck der drei M führt damit zu einem Zielkonflikt in „schneller, besser, billiger“. Die einzelnen Beziehungen zwischen Mensch, Maschine und Markt sollen dazu etwas genauer betrachtet werden, was aus der Motivation heraus zur Zielstellung hinleitet.

1.1.3 Mensch und Maschine: Daten-Tsunami und „gut genug“

Dass Probleme in der Nutzung des mobilen Internets nicht nur theoretischer Natur sind, sondern auch praktische Bedeutung haben, wird jeder Smartphone-Besitzer höchstwahrscheinlich schon einmal selbst festgestellt haben, aber auch Studien belegen dies. Je nach regionalem Markt zeigen Befragungen, dass im Schnitt 60 % der Smartphone-Nutzer wenigsten wöchentlich Probleme mit der Nutzung des mobilen Internets beklagen, z. B. dass Downloads zu langsam sind, in Westeuropa immerhin fast jeder fünfte Nutzer sogar täglich [5]² S. 5. Ausgehend von diesen Zahlen ist festzuhalten: Die Befriedigung der Erwartungen der Nutzer ist offenbar ein bekanntes aber auch ungelöstes Problem. Intuitiv wäre zu vermuten, dass sich das Problem selbst löst, schließlich entwickelt sich die Netztechnologie ständig weiter und die Netzbetreiber investieren kontinuierlich in den Netzausbau, was aus einem „schneller“ heraus zu einem „besser“ führen sollte. Entsprechend lohnt ein Blick zurück und nach vorn (Abschnitt „Daten-Tsunami“), um die Dynamik dabei zu erklären. Danach soll noch kurz nachgefragt werden, ob es denn auch ein bisher möglicherweise falsch angegangenes Problem ist (Abschnitt „Das rechte Maß“).

² internationale Studie aus 2013, in 12 Länder ausgeführt, konkrete Werte mindestens wöchentlichere Probleme zwischen 38 % und 79 %

1.1.4 Netzevolution und Daten-Tsunami: Höher, weiter, schneller im mobilen Internet

*„Es ist, als ob das mobile Web im Schneewittchensarg im Handy schlummerte und nur darauf wartete, dass es endlich wachgeküsst wird.“
[6] S. 70*

Nach jahrelangem eher mäßigem Erfolg für das mobile Internet gab es für dieses „endlich wachgeküsst“ o. g. Zitates des Jahres 2008 vorrangig drei Gründe:

- bessere Netze (EDGE und UMTS sind verfügbar)
- bessere Endgeräte (das erste iPhone wurde 2007 verfügbar)
- bessere Apps und Dienste (optimiert für die mobile Nutzung)

Seitdem boomt das mobile Internet. Der Daten-Tsunami ist dabei ein gern gebrauchtes Bild für das ungebremsste, exponentielle Wachstum mit annähernder Verdopplung des transportierten Datenvolumens pro Jahr. Genaue Zahlen liefert rückblickend beispielsweise Akamai im Report „State of the Internet“, siehe [7], mit Datenmaterial für die mobile Nutzung von Ericsson oder als Prognose für die jeweils nächsten fünf Jahre der Cisco Visual Network Index (VNI), siehe [8], der in Verbindung mit dem Daten-Tsunami eigentlich immer³ genannt wird. Die offiziellen Zahlen zum Datenvolumen im Mobilfunk hält für Deutschland der Jahresbericht der Bundesnetzagentur bereit, der zwischen 2012 und 2013 einen Anstieg von ca. 71 % verzeichnet [9] S. 77. Für den Daten-Tsunami gibt es dabei verschiedene Motoren:

- **immer mehr Nutzer im mobilen Internet:** Alleine für Deutschland wurde von mehr als 34 Mio. Nutzern des mobilen Internets im Laufe des Jahres 2014 ausgegangen [10]⁴, womit sich das jährliche Wachstum von ca. 25 % über die letzten Jahre bestätigt (2013: ca. 27 Mio., 2012: ca. 21 Mio., 2011: ca. 17 Mio., nach [11]⁵). Dies entspricht 51 % der privaten Internet-Nutzer im Jahr 2013 nach zuvor 37 % im Jahr 2012 (nach [12]⁶).

³ Das „immer“ lässt sich nicht beweisen, aber nach der Literaturrecherche zu dieser Arbeit ist dem wohl so...

⁴ Arbeitsgemeinschaft Online Forschung (AGOF), 2. Quartal 2014

⁵ auf Basis der Zahlen der Arbeitsgemeinschaft Online Forschung (AGOF)

⁶ mit Verweis auf Zahlen des Statistischen Bundesamtes, siehe [13], die Angaben aus den Erhebungen differieren geringfügig

- **immer intensivere Nutzung:** Es ist insgesamt neben dem Zuwachs der Anzahl von Nutzern auch von einer Intensivierung der Nutzung mobiler Apps auszugehen (siehe Erhebungsergebnisse in [14]⁷, bezogen auf sämtliche App-Nutzung). Diese Intensivierung ergibt sich vor allem aus einer immer häufigeren und längeren Nutzung.
- **immer anspruchsvollere, netzintensive Dienste:** Die Nutzer verbrauchen auch deshalb immer mehr Datenvolumen, weil neue anspruchsvolle Dienste hinzukommen und die bereits genutzten netzintensiver werden. Dabei ist die Nutzung von Video (z. B. Youtube) der größte Treiber ([8] und [15] S. 4), mit immer höheren Qualitätsstufen (von SD, über HQ zu HD siehe [16]). Hinzu kommen die multimediale Nutzung Sozialer Netze (z. B. Facebook), die Verlagerung der Datenspeicherung in die Cloud und IP-basierte Echtzeitkommunikation, z. B. Skype oder Videotelefonie [15] S. 4.
- **immer leistungsfähigere mobile Endgeräte:** Sogenannte Smart-Devices (Smartphones und Tablets) verursachen den weitaus größten Anteil des mobilen Datenvolumens [8]. Dabei ist der Trend zu größeren Bildschirmen mit höheren Auflösungen und leistungsfähigen Prozessoren festzustellen, die z. B. wiederum bessere Videoqualitäten sinnvoll nutzbar machen.
- **immer mehr Datenvolumen pro Nutzer:** Als Ergebnis der Intensivierung der Nutzung und anspruchsvollerer Dienste steigt das durchschnittliche Datenvolumen pro Nutzer immer weiter an, von durchschnittlich 358 MB (2013/H2) auf 394 MB (2014/H2) [17]⁸ S. 14. Dabei ist aber zu beachten, dass sich der Datenverkehr sehr ungleichmäßig zwischen den Kunden verteilt, was die Mittelwerte verschleiern: 1 % der Mobilfunknutzer erzeugen fast 31 % des gesamten Verkehrsvolumens, wohingegen 50 % der Nutzer faktisch nichts beitragen [17]⁹ S. 15. In anderen Märkten sind die Verhältnisse zum Teil noch extremer¹⁰ mit 1 % der Nutzer als Ursache für 50 % des Download-Volumens bzw. eine gute Bestätigung der 90/10-Regel, bei der 90 % des gesamten Volumens von nur 10 % der

⁷ internationale Erhebung

⁸ Zahlen aus europäischen Mobilnetzen, die etwas über den zurückliegenden Zahlen aus Deutschland liegen, siehe [18] S. 27f (nur Postpaid-Kunden)

⁹ Messungen im mobilen Internet, Europa, 2. Halbjahr 2013

¹⁰ entsprechend wird in [19] S. 11 auch scherzhaft LTE als *Long Term Extreme* statt als *Long Term Evolution* titliert

Nutzer stammt [19] S. 11. Wehe der Intensivierung der Nutzung im Mainstream...

- **immer leistungsfähigere Netze:** Der Zuwachs an Netzleistung verläuft einerseits in Generationenschüben durch die Einführung neuer Technologien sowie schrittweise und punktuell durch Netzumrüstungen und Netzverdichtung. Daraus ergibt sich ein kontinuierlicher Leistungszuwachs, vor allem hinsichtlich von Spitzendatenraten und verfügbarer Kapazität.

Insbesondere der letzte Punkt mag verwirren, ist er doch Anfangs- und Endpunkt der Betrachtungen. Dabei ist festzustellen, dass technologische Entwicklung und Nutzung sich gegenseitig treiben. Neue Möglichkeiten durch bessere Technik werden genutzt und erzeugen selbst wiederum neue Anforderungen an die Technik. Dabei entsteht aus den sich verändernden Anforderungen eine Eigendynamik, die durch höher-weiter-schneller zu einem Innovations- und Wachstumszyklus führt, womit der Endpunkt wieder der Anfangspunkt ist, siehe Abbildung 2a. Diese Koevolution von Netz, Mobilgeräten und Diensten ist damit ermöglichende Voraussetzung der sich intensivierenden Nutzung und notwendige Konsequenz zugleich.

Ausgehend von diesen Entwicklungen ist festzuhalten: Der technologisch bedingte Leistungszuwachs in den Netzen kann für „anspruchlose“ Dienste für eine Lösung „von selbst“ sorgen (prinzipiell ausreichend Ressourcen), gleichzeitig kommen aber durch den Innovationszyklus neue „anspruchsvolle“ Dienste hinzu. Das Problem löst sich also nicht grundsätzlich von selbst, sondern es verschiebt sich nur. In der Informatik ist dieses Phänomen als „Fenster der knappen Betriebsmittel“ [21] S. 246 für verschiedene Ressourcen grundsätzlich bekannt, siehe Abbildung 2b. Neben dem beschriebenen Automatismus zur Verknappung zwischen Technik und Nutzung gibt es auch wirtschaftliche Aspekte zu beachten, denn bekanntlich sind knappe Güter auch wirtschaftliche Güter. Diese Überlegungen bilden die Überleitung zu den Beziehungen von Mensch und Maschine zum Markt und werden nachfolgend noch weiter erläutert. Doch zunächst soll noch bei Mensch und Maschine geblieben werden, um kurz zu erörtern, wie sich denn ein rechtes Maß darstellen könnte, wenn eine Entwicklung im Sinne von „höher, weiter, schneller“ nicht ausreicht.

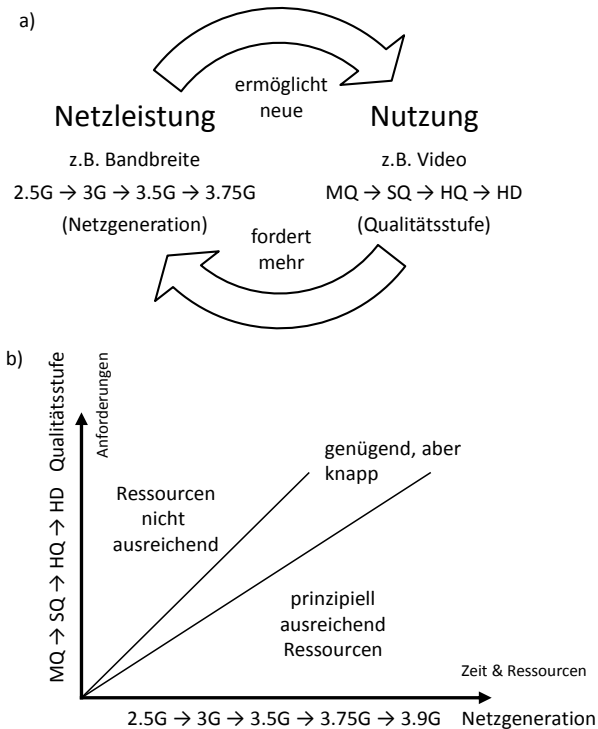


Abbildung 2: Innovationszyklus von Netzleistung und Nutzung sowie „Fenster der knappen Betriebsmittel“

mobiles Video als ausgewählte Nutzung zur Illustration, auch durch jeweils andere, anspruchsvollere Dienste ersetzbar

a) Innovationszyklus inspiriert von [20] S. 8 (dort bezogen auf die Entwicklung der Ethernet-Technologie), prinzipiell sind auf der linken Seite auch Merkmale der Geräte wie Displaygröße und allgemeine Leistungsfähigkeit einsetzbar

b) „Fenster der knappen Betriebsmittel“ nach [21] S. 246, Beispiel zur Illustration, nicht direkt auf konkrete Zahlen bezogen

1.1.5 Das rechte Maß

Bislang werden überwiegend die nackten Zahlen der Megabits und Millisekunden als Testkriterien für den mobilen Internet-Zugang herangezogen, z. B. in Zeitschriften wie *connect* und *Chip* mit den „größten“ und „härtesten“ Netztests des Landes. Dies entspricht der Messung von „höher, weiter, schneller“. Dabei wäre eigentlich bei der Ermittlung der „Qualität“ zwischen Feststellung der technischen Beschaffenheit und der nutzerrelevanten Güte zu unterscheiden. Denn die Technik und die reine Feststellung des nackten Messwertes durch die Technik können nicht die Frage beantworten, ob z. B. 2 MBit/s *besser* als 1 MBit/s sind. Rein mathematisch betrachtet sind 2 MBit/s zwar *mehr* (im Sinne von *schneller*) als 1 MBit/s – aber ob 2 MBit/s auch tatsächlich *besser* als 1 MBit/s sind, das kann nur der Mensch aus seiner Nutzersicht beurteilen. Dabei werden die Interaktion der Nutzer mit dem Netz über mobile Geräte und Apps bei der Nutzung bestimmter Dienste, der Kontext und die damit verbundenen Erwartungen, was die QoE grob umreißt, erst langsam ein Teil der Untersuchungen.

Dazu muss man natürlich wissen: Was ist dafür wichtig – und wie viel ist gut bzw. wie viel ist „gut genug“? Somit ist die Frage nach dem rechten Maß bewusst doppeldeutig, zum einen als nutzerrelevante Größe der QoE und zum anderen als konkreter Zahlenwert der technischen Bezugsgrößen der QoS, die zusammenhängen. Gleichzeitig impliziert die Nutzerrelevanz, dass der Mensch das rechte Maß vorgibt. Von ihm kommt das „gut genug“ für Technik und Netz. Entsprechend kann der Blick allein auf die Technik nicht ausreichend sein. Befragungen zeigen sogar eher, dass vielen Nutzern die verborgene Technik des Netzes grundsätzlich egal ist, zumindest solange es gut (genug) ist [22] S. 21.

QoE ist nur schwierig zu fassen und zu quantifizieren, sowohl als Messung auch als Prognose, vor allem aus dem Interesse der Anbieter heraus. QoS ist hingegen einfacher zu beziffern. Daher ist es interessant, den Zusammenhang zu kennen, um die Größen in nutzerrelevanten Messungen und Planungen sinnvoll aufeinander abbilden zu können. An genau dieser Stelle ist noch weiterer Forschungsbedarf auszumachen. Dabei ist aber eben nicht nur die Sicht des Netzes als Maschine mit „höher, weiter, schneller“ einzunehmen, wie in

vielen Forschungsarbeiten der Vergangenheit, sondern die Einbeziehung des Menschen als Nutzer notwendig.

1.1.6 Maschine und Markt: Knappe Güter und weiße Flecken

“Regardless of the network operator type (mobile, fixed-broadband, etc.), reducing costs is always a common factor.” Light Reading¹¹

Abseits der Labors ist Mobilfunk und damit auch das mobile Internet vorrangig ein Geschäft – einer der Gründe, warum überhaupt die Frage nach dem „gut genug“ gestellt wird. Auch die Motivation der eigenen Studie entstammt letztlich den Überlegungen zum Zusammenwirken der drei Ms, inklusive des Marktes. Das ökonomische Prinzip ist die Triebfeder von Verbesserungsbestrebungen, es wird ständig nach einer Kosten-Nutzen-Optimierung gesucht.

Damit liefert der Markt die Begründung der praktischen und konkreten Relevanz zur Beschäftigung mit dem Thema – in Ergänzung zum grundsätzlichen erkenntnistheoretischen Interesse. Aspekte zum Markt werden nachfolgend nicht weiter intensiv betrachtet, wären aber eine nette Ergänzung, vor allem aus betriebswirtschaftlicher Sicht. Die weitere Verwendung und Verwertung der ausgearbeiteten Überlegungen und Ergebnisse sind mit Rücksicht darauf ebenso ausgeklammert wie Diskussionen zur Netzneutralität etc., auf die die Anwendung der Erkenntnisse hinauslaufen könnte.

Für die Anbieter bedeuten Daten-Tsunami und Flatrate-Angebote ein Dilemma. Es werden immer mehr Daten übertragen, die Kunden zahlen tendenziell aber immer weniger. Der Betrieb der Netze muss also günstiger werden, d. h. die Kosten (€/Bit) müssen zwangsläufig fallen. Aus marktwirtschaftlicher Sicht der Anbieter sind daher gezielte Planung und bedarfsgerechte Ressourcenverwaltung sinnvoll, die ein „gut genug“ der Beziehung Mensch und Maschine berücksichtigten und ein Nachsteuern durch Netzausbau und Modernisierungen einschließen.

¹¹ In der Ankündigung zu einer Expertendiskussionsrunde des Fachportals LightReading, siehe http://www.lightreading.com/radio.asp?webinar_id=217, abgerufen am 04.06.2014.

Bandbreite im Funkspektrum und Kapazität in den mobilen Netzen waren und sind knappe und damit wirtschaftliche Güter – und werden es trotz technologischer Fortschritte vermutlich auch zukünftig bleiben. Der Netzausbau schreitet zwar kontinuierlich voran, aber nicht überall gleich schnell, so verbleiben „weiße Flecken“ vor allem abseits der Ballungsräume, die aus wirtschaftlichen Gründen nicht mit der jeweils letzten Mobilfunkgeneration versorgt sind und eine vergleichsweise geringe Netzleistung aufweisen. Daraus ergibt sich, dass die Fenster der knappen Ressourcen ganz unterschiedlich ausfallen können. In diesen unterversorgten Gebieten sind selbst die anderenorts unproblematischen (anspruchlosen) Dienste noch eine Herausforderung und damit grundsätzlich auch einer Betrachtung wert.

1.1.7 Markt und Mensch: Kundenzufriedenheit = Nutzerzufriedenheit / Preis?

“The telecoms industry is lagging behind other industries when it comes to user satisfaction.” [5] S. 3

Die Betrachtung Markt und Mensch schließt das Dreieck. Der Mensch ist hierbei in der Rolle des Kunden zu sehen. Aus Sicht der Anbieter ist die Zufriedenheit der Kunden die kritischste Kenngröße für einen erfolgreichen Geschäftsbetrieb [23]¹² S. 7. Für eben diese Kunden ist die Netzleistung wiederum der weitaus wichtigste Einzelfaktor (19 %) zur Kundenloyalität als Ausdruck langfristiger Zufriedenheit, gefolgt von dem PreisLeistungsverhältnis (16 %) und deutlich vor weiteren vorwiegend nichttechnischen Angebotseigenschaften (mit ≤ 10 %), siehe [5] S. 6. Unter der Annahme, dass Netzabdeckung, Verfügbarkeit sowie Zuverlässigkeit des Netzes und vielleicht sogar die grundlegende Sprachtelefonie Hygienefaktoren¹³ sind, kann die mobile Datennutzung möglicherweise sogar zum dominierenden Leistungsfaktor für die Wahrnehmung der Netzqualität und des Gesamtangebots avancieren.

¹² für 85% der Befragten, Umfrage unter 70 Netzanbietern global (nicht nur mobil), Heavy Reading, 2013

¹³ Nein, das hat nichts mit der Sauberkeit oder Pflege des Smartphones zu tun... Eine Unterscheidung der verschiedenen Arten von Zufriedenheitsfaktoren wird später noch geliefert. Kurz: die erwartete Mindestleistung.

Markt bedeutet auch Wettbewerb zwischen verschiedenen Anbietern. Unter der Annahme, dass dem Nutzer die technologische Grundlage und theoretische Leistungsfähigkeit des Netzes egal sind, so bleibt als Differenzierungsmerkmal im Wettbewerb die Kundenzufriedenheit mit der erlebten Leistung und den damit verbundenen Kosten. Für einen Anbieter kann es entsprechend interessant sein, einen Sweet Spot zwischen notwendiger Leistung (eigenen Kosten) und akzeptablem Preis (Kosten der Kunden) für möglichst viele zufriedene Kunden eingrenzen zu können. Dabei ist zu beachten, dass sich Netzleistung und Dienstgüte aus Sicht des Anbieters bzw. des Nutzers unterschiedlich darstellen können.

1.1.8 Zusammenfassung zur Motivation

In der mobilen Netznutzung gewinnt die Datennutzung immer weiter an Bedeutung für Nutzer und Anbieter. Die technologische Entwicklung der Netzleistung treibt die Nutzung und diese wiederum die technologische Entwicklung in einem Innovationszyklus. Das Problem der knappen Ressourcen löst sich somit nicht von alleine, im besten Fall verschiebt es sich nur. Einfache Metriken zur QoS als Feststellung der Beschaffenheit sind nicht unbedingt nutzerrelevant und damit nicht zwangsläufig aussagekräftig zur QoE. Die Netzleistung bleibt auch deshalb „knapp“, weil sie ein wirtschaftliches Gut ist. Für den Markt gilt es dabei abzuwägen zwischen einem „zu viel“ durch die Kosten für Errichtung und Betrieb und einem „zu wenig“ vor allem aus Nutzersicht, was auf die Kundenzufriedenheit durchschlägt. Die Kenntnis der Anforderungen der Nutzer ist die Voraussetzung für die bedarfsgerechte Bereitstellung passender Angebote. Die gesamten Überlegungen führen zu einer Suche nach dem „gut genug“. Die Beziehung Mensch-Maschine mit dem Zusammenhang von QoS und QoE steht im Zentrum der weiteren Betrachtungen. Der Markt liefert dabei einen Gutteil der Begründung der praktischen Relevanz des Themas.

1.2 Zielsetzung

Wie in der Motivation erläutert, sind prinzipiell unterschiedliche Aspekte Mensch, Maschine und Markt für weitere Untersuchungen interessant. Die vorliegende Arbeit konzentriert sich auf Maschine und Mensch, bzw. auf deren Zusammenwirken in der Nutzung des mobilen Internets und somit auf den Zusammenhang von QoS und QoE. Auch bei Maschine und Mensch treffen verschiedene Sichten und Welten aufeinander, wie später noch diskutiert wird. Die eigene Annäherung soll dabei technische Sicht und Sicht der Nutzer sinnvoll integrieren.

Das Verständnis der Zusammenhänge von QoS und QoE ist Voraussetzung für eine adäquate Betrachtung und mögliche konkrete Wertbestimmungen. Dies erfordert ein plausibles Modell, das entweder zu ermitteln und ggf. zu adaptieren oder zu entwickeln ist. Das Modell soll als Untersuchungs- und Erklärungsmodell fungieren.

Für die Beschreibung des **QoS-QoE-Zusammenhangs** sind geeignete und relevante Größen zu ermitteln. Dies betrifft sowohl die einflussnehmenden Größen als auch die beeinflussten Größen. Bei den gesamten Überlegungen soll das **Nutzererlebnis** im Mittelpunkt stehen. Das **Netz** kann grundsätzlich als kritischer Faktor für das Nutzererlebnis angesehen werden. Für ein tiefgehendes Verständnis soll geklärt werden, wie das Erlebnis im mobilen Netz über davon abhängige Größen wahrgenommen und beurteilt wird. Es ist davon auszugehen, dass neben dem Netz **weitere Faktoren** eine Rolle spielen. Diese sind ebenfalls in geeigneter Art zu berücksichtigen.

Für ausgesuchte relevante und typischerweise mobil genutzte **Dienste** sollen der **spezifische QoS-QoE-Zusammenhang** und dazu relevante Einflussgrößen ermittelt werden. Dabei soll geprüft werden, ob sich grundlegend unterschiedliche Arten des QoS-QoE-Zusammenhangs bestätigen bzw. finden lassen.

Die Motivation hat gezeigt: Ein Anbieten von „zu viel“ ist aus Anbietersicht nicht wünschenswert, ebenso wie ein „zu wenig“ aus Kundensicht. Neben den prinzipiellen Erkenntnissen zur Theorie des QoS-QoE-Zusammenhangs soll

daher auch die Quantifizierung in Form konkreter Werte angegangen werden. Es ist zu klären, ob es für die sinnvoll ausgewählten Dienste ein „gut genug“ gibt und sich praktikable **Schwellwerte** bzw. ein Zielwert als Kompromiss der verschiedenen Interessen bestimmen lassen: Wie viel/schnell ist gut genug? Dazu bietet sich ein Parameterbereichstest an, über den sich die notwendige Quality of Services (durch QoS-Netzparameter charakterisiert) für angestrebte QoE-Qualitätsziele bestimmen lässt. Zur konkreten Bestimmung sind geeignete Metriken und quantifizierte Schwellen zu ermitteln oder festzulegen. Es sollen dabei typische mobile Nutzungs- und Netzsituationen betrachtet werden.

1.3 Methodik und Rahmen

Nach *Warum* (Motivation) und *Was* (Zielsetzung) sollen in diesem Abschnitt das *Wie*, also die grundsätzliche Methodik und der Rahmen der vorliegenden Arbeit erläutert werden. Dieser Abschnitt liefert damit die Hintergründe und Begründung für Inhalt und Aufbau der Arbeit. Dazu werden die Besonderheiten des eigenen Weges und Lösungsansatzes skizziert. Die Methodik begründet kurz die drei notwendigen Hauptbausteine der Arbeit: Theorie, Modell und Empirie. Als Rahmen werden der Ausgangspunkt der Vorarbeiten mit einer Vorstudie und eine Übersicht der ausgeführten Arbeiten präsentiert.

1.3.1 Methodik: Theorie, Modell und Empirie

Mit den unterschiedlichen Sichtweisen und Denkrichtungen der multi-/interdisziplinären Betrachtungen zu QoS und QoE sind ebenfalls unterschiedliche methodische Ansätze verknüpft: die Psychologie als hauptsächlich empirische Wissenschaft und die Ingenieurwissenschaft mit stark rationalen Ansätzen und Erklärungsmodellen. Der Psychologe als Empiriker hat sich damit arrangiert, mit geschicktem Experimentieren durch den Blick von außen auf interessierende latente, also verdeckte Größen im Inneren, zu schlussfolgern. Der Ingenieur wiederum will das Gesamtsystem solange zerlegen, bis er die Einzelteile in Händen hält, Rolle und Funktion jedes einzelnen Teils versteht und daraus das gemeinsame Funktionieren des Ganzen aus dem Inneren heraus erklären kann.

Die Arbeit versucht die Verbindung von Theorie und Empirie, als den zwei Wegen, wie in der Wissenschaft zur Erkenntnis gelangt wird:

- Theorie: rationale Erklärung und Modellierung, von der Ursache zur Wirkung, vorhersagbar und nachvollziehbar
- Empirie: auf der Basis von Erfahrung bzw. der Realität, Bestimmung durch methodische Erhebung von Daten

Dabei sind Theorie und Empirie zu einem Untersuchungsgegenstand nicht getrennt, sondern verzahnt. Je nach Schwerpunkt liefert die Theorie einen plausiblen Ausgangspunkt für die empirische Bestimmung konkreter Werte oder die Empirie hilft, die Theorie zu bestätigen bzw. weiterzuentwickeln. Typischerweise steht einer der Punkte im Zentrum des Forschungsinteresses und wird von dem anderen soweit wie nötig getragen (vgl. [24]). Wie aus der Zielstellung zu entnehmen, sind für das „gut genug“ auch konkrete Werte gefordert. Entsprechend nimmt die Empirie innerhalb der Arbeit eine bedeutende Stellung ein, was sich auch schon an den vielen einzelnen Bausteinen mit „Test ... und ... Studie“ im Inhaltsverzeichnis erkennen lässt.

Für die gemeinsame Betrachtung von Theorie und Empirie lassen sich einige Kernfragen formulieren:

- Welche Konzepte gibt es zu QoS und QoE? Was verbirgt sich dahinter?
- Wie hängen die Konzepte von QoS und QoE zusammen (konzeptueller Zusammenhang)?
- Welche relevanten Elemente gibt es bei der Betrachtung von QoS und QoE?
- Wie hängen diese Elemente konkret/kausal zusammen (Ursache und Wirkung im Modell)?
- Wie werden die relevanten Elemente im Test sinnvoll integriert und berücksichtigt?
- Wie sind die gewünschten Erkenntnisse aus den gewonnenen Daten auszuwerten?

Ein geeignetes Modell greift diese Fragen auf und verbindet so Theorie und Empirie. Es wird davon ausgegangen, dass der Leser nicht mit der Testtheorie vertraut ist. Entsprechend geht dies einher mit einer kurzen Einführung der begrifflichen Grundlagen zum Basismodell und zur Empirie.

1.3.2 Theorie: Grundannahmen und Ausgangspunkt

Durch die Zielsetzungen und Anforderungen ergab sich, dass eine rein theoretische Abhandlung oder etwa eine Metastudie nicht zielführend war. Aus den Erkenntnissen der Vorarbeiten wiederum war abzuleiten, dass die empirische Untersuchung für tragfähige Aussagen auch ein passendes und plausibles theoretisches Gerüst bedingt. Echte Systeme in der realen Welt sind typischerweise hochkomplex, viele verschiedene Elemente interagieren, die Ursache führt nur mittelbar zur Wirkung etc. Für die Theorie sind aber Abstraktionen zu bevorzugen. Aus den theoretischen Überlegungen entsteht deshalb ein formales Modell, das die Rolle und Zusammenhänge der einzelnen Elemente erfasst und beschreibt.

Sowohl Theorie als auch Empirie brauchen gewisse Grundannahmen als Ausgangspunkt. Als Grundannahmen für die Untersuchungen zum QoS-QoE-Zusammenhang können gelten:

- Netzcharakteristik (QoS) und Nutzererlebnis (QoE) im mobilen Internet hängen zusammen.
- Der Nutzer hat Erwartungen bei der Nutzung mobiler Dienste in mobilen Netzen; er stellt Anforderungen. Beispiel: Der Nutzer erwartet, dass ein Video ohne Unterbrechungen wiedergegeben wird.
- Die Anforderungen des Nutzers sind in bestimmte technische Anforderungen zu übersetzen.
- Das Netz erfüllt diese Anforderungen, oder der Nutzer registriert Einschränkungen.

Daher wird ein Ursache-Wirkungs-Zusammenhang zwischen QoS und QoE angenommen. Das Nutzererleben hängt dabei von dem Netz ab. Dabei ist davon auszugehen, dass neben dem Netz auch weitere technische ebenso wie nichttechnische und menschliche Faktoren einen Einfluss ausüben. Ein Nutzer nimmt das Netz nicht unmittelbar wahr, sondern durch die genutzten Dienste sowie sein Endgerät und interagiert über diese mit dem Netz. Dabei sollte verstanden werden, was der typische (mobile) Nutzer überhaupt für grundsätzliche Anforderungen hat. Intuitiv einleuchtend ist, dass es aus Sicht der QoS anspruchsvolle aber auch anspruchlose Dienste gibt. Dabei können die genutzten Apps ebenso wie die Inhalte gut oder schlecht an die mobile Nutzung

angepasst sein. Durch diesen mobilen Aspekt ist zu prüfen, inwieweit Erkenntnisse aus der Desktop-Welt überhaupt auf Smartphone und Co. übertragen werden können. Ebenso ist anzunehmen, dass Wahrnehmung und Erlebnis aus Nutzersicht sehr individuell sind. Für konkrete Dienste sind unter Berücksichtigung o. g. Punkte sinnvolle Annahmen zum QoS-QoE-Zusammenhang möglich, eine Überprüfung oder Bestimmung kann jedoch nur empirisch erfolgen.

1.3.3 Modell: Von der Black Box zur White Box

Modelle sind geschaffene externe Abbildungen realer Zusammenhänge. Sie vermitteln damit zwischen Theorie als Abstraktion und Empirie als Erfassung der Realität. Ausgehend von den Kenntnissen über die innere Funktionsweise¹⁴ und den Möglichkeiten, die Wirkzusammenhänge zu untersuchen, werden Modelle und darauf aufbauend Tests unterschieden in:

- Black Box = ohne Kenntnis (ungünstig, Ausgangszustand)
- White Box = mit vollständiger Kenntnis (ideal, angestrebt, auch Glass Box genannt)
- Grey Box = teilweise Kenntnisse (realistisch, noch zu akzeptieren, zwischen Black Box und White Box angesiedelt)

Die beteiligten Größen, je nach Terminologie als Variablen, Faktoren oder Parameter bezeichnet, werden dabei nach ihrer Rolle eingeteilt, siehe Abbildung 3. Aus der Sicht eines Versuchs zur empirischen Aufdeckung von Zusammenhängen ist das kybernetische Modell einer Black Box der Ausgangspunkt (siehe [26] S. 5f.). Dafür sind nur recht wenige Grundannahmen notwendig: die interessierenden (möglichen) Einflussgrößen als Input führen zusammen mit Störfaktoren zu den interessierenden Zielgrößen als Output, bei aber a priori unbekanntem internen Zusammenhängen. Für eine Black Box bleibt nur der Versuch der Aufdeckung der Zusammenhänge und Abhängigkeiten im Nachhinein (ex post). Das Modell entsteht auf Basis der gesammelten Daten in der Auswertung. Dies kann funktionieren, muss es aber nicht, etwa wenn viele Neben- und Störfaktoren zusammenkommen, so dass eine Aufklärung alleine auf Basis der gewonnenen Daten entweder gar nicht oder nur sehr grob möglich ist.

¹⁴ Die Begriffe der Black Box und White Box bzw. Glass Box werden auch im Bereich der Software-Tests in ähnlicher Bedeutung genutzt: ohne Kenntnis bzw. mit Kenntnis der Interna, siehe z. B. [25] S. 81ff., S. 59ff sowie [25] S. 119.

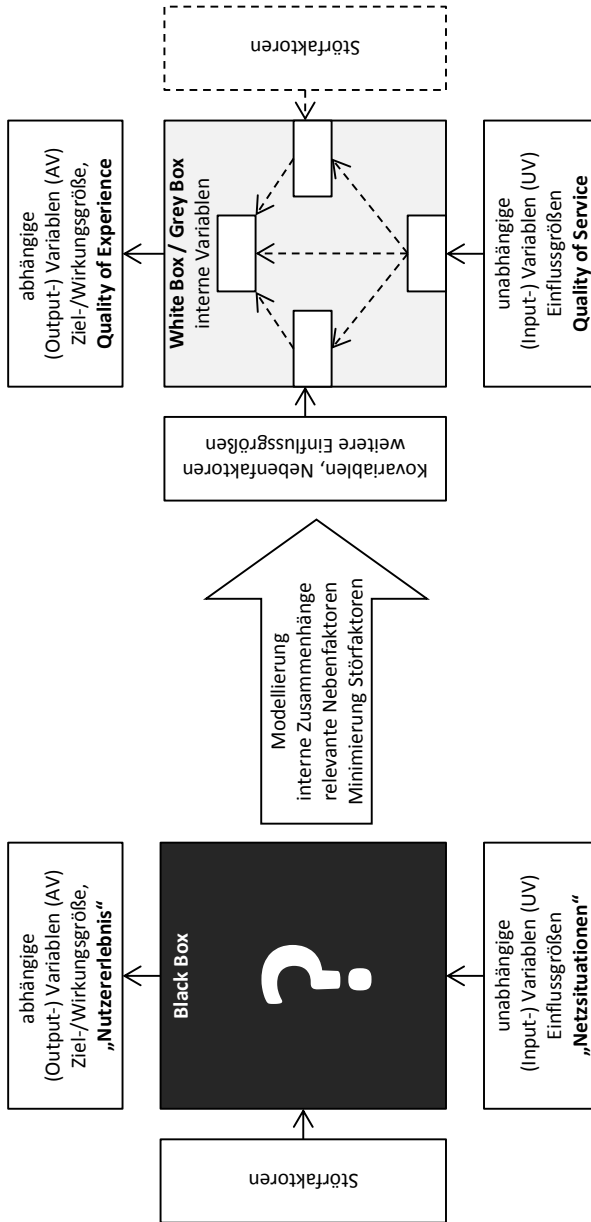


Abbildung 3: Modellierung des Zusammenhangs – Von der Black Box zur White Box bzw. Grey Box

Im Fall einer White Box wird von Vorwissen ausgegangen. Für die Bildung eines sinnvollen Modells ist dazu die Beschäftigung mit der Theorie zu QoS und QoE notwendig. Die prinzipiellen Komponenten des Modells können damit benannt werden. Die Trennung in systematische Nebenfaktoren und ungeplante Störfaktoren hilft das Modell zu verfeinern. Wenn kein direkter und unmittelbarer Zusammenhang zwischen Einflussgröße und Wirkgrößen vorliegt, sind Abstraktionen durch Zwischengrößen in Form von internen Variablen möglich. So ist eine fachlich-sachlich begründete Erklärung von Ursache und Wirkung aus der Erklärung der internen Zusammenhänge unter dem Einfluss von Nebenfaktoren mit minimierten, weil bekannten Störfaktoren, möglich.

Es sollte ein Basismodell für die QoS-QoE-Zusammenhänge entwickelt werden, das als Ausgangspunkt für konkrete Testfälle in Form verschiedener Dienste mit spezifischen Nutzungsszenarien für die empirische Ermittlung genutzt werden kann. Das Ziel der angestrebten Modellierung ist dabei eine White Box bzw. zumindest eine Grey Box. Überlegungen zu und Rückgriffe auf das Modell sind in verschiedenen Phasen der empirischen Studie sinnvoll. Grob ist dabei zwischen *ex ante* (im Vorhinein), also schon bei Planung und Vorbereitung, sowie *ex post* (im Nachhinein), also in der Auswertung und Diskussion der Ergebnisse nach der Testdurchführung zu unterscheiden.

Damit ergeben sich die Vorteile einer White/Grey Box *ex ante* bzw. *ex post* wie folgt.

- ex ante*:
- gezielte Berücksichtigung von Nebenfaktoren
 - Steigerung der Testeffizienz durch sinnvolle Eingrenzung auf relevante Einflussgrößen und Bereiche für konkrete Dienste
 - Bestimmung der Wirkrichtung und Abschätzung der Wirkstärke
 - Bestimmung relevanter Zwischengrößen und Bildung entsprechender Erwartungswerte

- ex post:
- Ergebnisse mit Mehrwert über die getesteten App-Versionen hinaus
 - Rückschluss von konkreten Werten zur Verallgemeinerung auf grundsätzliche Zusammenhänge
 - Überprüfung der Bewertung zur Relevanz/Wirkstärke der Einflussfaktoren: zumindest in Größenordnungen (vorrangig, gleichrangig, nachrangig)

Eine Verfeinerung und Konkretisierung des Grundmodells für verschiedene Dienste ergibt sich aus gesonderten Vorüberlegungen und Voruntersuchungen dazu sowie aus den entwickelten Nutzungsszenarien. Besonders die intensiven Voruntersuchungen führen im Vergleich zu einem Test als Black Box zu einem höheren Aufwand, aber auch zu einem besseren Verständnis dessen, was überhaupt und wie gut getestet wird.

1.3.4 Empirie: Realismus und Kontrolle

Die Theorie kann zwar sinnvolle Grund- und Nebenannahmen liefern, die Überprüfung und quantitative Bestimmung erfordert aber letztlich die Empirie. Warum es einen solchen empirischen Anteil in der Arbeit gibt, lässt sich recht kurz beantworten: Der interessierende Zusammenhang zwischen QoS und QoE ist nicht ausschließlich durch technische Messungen zu bestimmen, weil der Mensch als „unscharfes und weiches“ aber relevantes Element beteiligt ist. Daher ist auch nicht von einem deterministischen Zusammenhang, sondern von einem wahrscheinlichkeitstheoretischen Zusammenhang auszugehen, der sich über Verteilungen charakterisieren lässt, was wiederum für sinnvolle Aussagen eine bestimmte statistische Masse erfordert. Die Einbeziehung des Menschen in die Untersuchungen bedeutet Probanden für eine Nutzerstudie. Aus den Ergebnissen der empirischen Nutzerstudie sollten sich ergeben:

- Bestätigung der Annahmen zu relevanten Einflussgrößen und Anforderungen
- Quantifizierungen der gesuchten Schwellwerte
- Rückschlüsse für die Verbesserung von Modell und Theorie

Prinzipiell waren die grundsätzlichen Zusammenhänge und Einflussrichtungen klar. Für die Wirkung von Nebenfaktoren, insbesondere deren Stärke, konnten nur ungefähre Vorabannahmen getroffen werden. Diese ergaben sich vor allem aus den Voruntersuchungen zu den konkreten Szenarien im Lauf der Testplanung und Testvorbereitung.

In der Empirie wird grob zwischen der einfachen Erfassung unter gegebenen Bedingungen (Beobachtungen) und der planmäßigen und kontrollierten Herstellung der Bedingungen (Experimente) unterschieden. Für die angestrebten Erkenntnisse und Nebenbedingungen wurde ein Laborexperiment zur Durchführung in einer kontrollierten Testumgebung als geeignet angesehen und ausgewählt. Die Nutzerbewertungen sollen sich gezielt bestimmten technischen Parametern zuordnen lassen.

Aus den Grundannahmen und der Zielstellung ergab sich damit auch der Rahmen der empirischen Untersuchungen:

- Input (unabhängige Variablen, UV): QoS (Quality of Service) typisch für mobile Nutzung
- Output (abhängige Variablen, AV): QoE (Quality of Experience) auf Basis echter Nutzerbewertungen

Die Ermittlung des Zusammenhangs erfolgt dabei als empirisches Messen, d. h. Testen durch Variation des Inputs und Messen des Outputs mit qualitativer und quantitativer Erfassung des Zusammenhangs. Für die Umsetzung der empirischen Untersuchungen lässt sich demnach als entscheidende Frage formulieren: Wie bringt man die Probanden dazu, eine QoE-Bewertung (AV) passend zur QoS-Parametrisierung (UV) abzugeben, so dass der interessierende Zusammenhang korrekt erfasst wird? Aus dieser Frage ergibt sich die Bedeutung der Aspekte *Realismus* und *Kontrolle* für die Empirie. Die relevanten Einflussgrößen sind unter Berücksichtigung gewisser Anforderungen und Einschränkungen in geeigneter Art und Weise in die Untersuchungen zu integrieren. Dafür sind prinzipiell verschiedene Methoden denkbar, ausgehend vom formalen Modell über Simulation und Emulation bis hin zur Untersuchung echter Elemente und Systeme, siehe Abbildung 4 zur Einordnung und zu den gewählten Bestandteilen der eigenen Betrachtungen und Realisierung.

Realismus: Wie echt sind die beteiligten Elemente und damit die untersuchten Bedingungen? Neben einem allgemeinen Erkenntnisgewinn sollten die Untersuchungen vorrangig praktisch verwertbare, konkrete Ergebnisse liefern. Deshalb wurde eine hohe Realitätsnähe angestrebt, die sich darin äußert, möglichst viele „echte“ Elemente zu integrieren. Die Empirie ist bei einem solchen Ansatz im Zweifel ausschlaggebend, jenseits aller „müsste, sollte, könnte“ einer noch so schönen Theorie. „Echt“ bezieht sich in einem Nutzertest natürlich zuerst auf echte Menschen als Probanden. Auch für die weiteren Faktoren sollte ein hoher Grad an Realismus angestrebt werden. Dies gilt besonders für die Elemente, mit denen der „echte Mensch“ unmittelbar interagiert, d. h. die mobilen Endgeräte und Apps. Für die Probandenstudie bedurfte es weiterhin eines geeigneten Testbeds, eines wohlüberlegt zusammengestellten Testparcours, eines passenden Testdesigns und einer umfassenden Auswertung. Diese Auflistung entspricht damit grob den Bestandteilen und der Struktur der Arbeit neben der Theorie. Den theoretischen Beschreibungen ist jeweils eine Erläuterung zur Berücksichtigung in der empirischen Studie gegenübergestellt, die aufklärt, welche Elemente wie und warum integriert wurden.

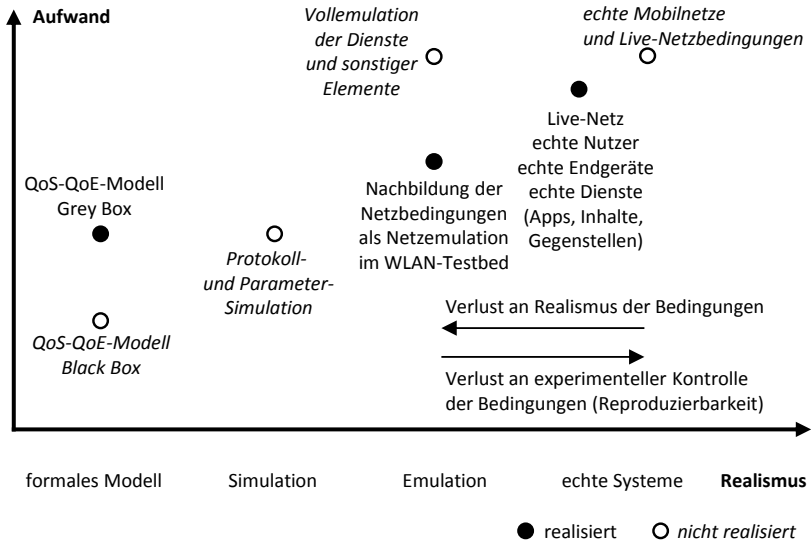


Abbildung 4: Vergleich verschiedener Ansätze, unterschieden nach Realismus und benötigtem Aufwand in Anlehnung an [27] S. 4, erweitert und angepasst an die eigenen Anforderungen und Bestandteile; Skalierung der Achsen im Original logarithmisch, hier vorrangig als aufsteigende Kategorien zu verstehen

Im Vorgriff auf die detaillierten Erläuterungen lässt sich die Berücksichtigung der einzelnen Elemente in der Studie wie folgt zusammenfassen:

- geeignete Nachbildung der Netzsituationen durch Emulation in einem Testbed
- Nutzung echter mobiler Endgeräte
- Nutzung echter Dienste, d. h. echte Apps, echte Inhalte und echte Gegenseiten (Server)
- was wiederum eine Einbeziehung des echten Internets erfordert
- echte Nutzer als Probanden

Aus der gewählten Herangehensweise für die Studie als Laborexperiment und das Testdesign ergeben sich für die einzelnen Elemente weitere Aspekte wie Kontrolle und Skalierbarkeit. Die vertiefende Diskussion zu den technischen bzw. inhaltlichen Gesichtspunkten der Kontrolle erfolgt in den zugehörigen Kapiteln der Elemente.

Skalierbarkeit: Wie groß ist der Schritt von 1 Testergebnis zu 10, zu 100, zu 1000? Empirie und Statistik gehen Hand in Hand. Die Statistik gibt dabei Orientierungswerte zum benötigten Umfang vor. Ansatz und technische Lösung müssen daher geeignet sein, um potenziell hunderte Probanden innerhalb überschaubarer Zeit testen zu können. Dies bedeutet ggf. auch einen hohen Grad an Parallelität, um den personellen Aufwand zur Betreuung der Studie im Rahmen zu halten. Aus technischer Sicht steigen mit der Parallelität die Herausforderungen, ebenso wie mit starker zeitlicher Verdichtung als auch mit starker zeitlicher Ausdehnung der Tests (externe Drift, dazu später mehr).

Aufwand und Kosten: Sowohl technisch als auch organisatorisch/logistisch muss der Aufwand beherrschbar bleiben. Dabei ergibt eine Abschätzung, dass mit steigendem Grad an gefordertem Realismus der Aufwand insgesamt ebenfalls wächst. Der Aufwand bezieht sich an dieser Stelle auf die eingesetzten finanziellen, personellen und zeitlichen Ressourcen. Die Fixierung dieser Größen erfordert ggf. Kompromisse in den anderen Punkten und klare Abgrenzungen, bis wohin überhaupt eine Umsetzung sinnvoll ist.

1.3.5 Vorarbeiten

Der in der vorliegenden Arbeit beschriebenen Studie ging eine kleinere Studie voraus. Entsprechend der Abfolge werden diese Studien nachfolgend Testrunde 1 (Vorstudie) bzw. Testrunde 2 (Hauptstudie) genannt. Der Hauptteil der vorliegenden Arbeit bezieht sich auf die Testrunde 2 mit dem genannten Rahmen und Partnern. Abgesehen vom aktuellen Abschnitt und einigen kleineren punktuellen Einschüben wird nicht weiter auf Testrunde 1 eingegangen, da im Vergleich sowohl die theoretischen Überlegungen wesentlich weiterentwickelt wurden als auch die konkreten empirischen Ergebnisse der Testrunde 2 deutlich belastbarer sind.

Auch wenn auf die Testrunde 1 nicht im Detail eingegangen wird, so ist sie doch erwähnenswert. Bei einer derartigen Komplexität wie in den beschriebenen Tests ist es unwahrscheinlich, dass in dem genannten Zeitrahmen, die erzielten Ergebnisse sofort erreichbar gewesen wären. Pragmatisch formuliert: Die Testrunde 1 diente als Vorstudie für die Testrunde 2. Vorstudien dienen üblicherweise der Überprüfung der Machbarkeit, wofür die statistische Masse nicht zu groß gewählt wird. Erste Erkenntnisse in Form konkreter Zahlen sind zwar ebenfalls aus der Testrunde 1 entstanden, retrospektiv sind aber die gewonnenen allgemeinen Erkenntnisse zum Ansatz und zur Umsetzung wichtiger. So zeigte sich, dass es mit dem gewählten Ansatz und relativ einfachen Mitteln prinzipiell möglich ist, die interessierenden QoS-QoE-Zusammenhänge zu untersuchen und brauchbare konkrete Werte zu ermitteln. Gleichzeitig ergab sich aber auch, dass noch erhebliches Verbesserungspotenzial bestand. Als kritische Punkte zeigten sich die Testeffizienz und die Beachtung von relevanten Nebenfaktoren neben der Variation der primär interessierenden Größen. Das genaue Zusammenwirken und die Rolle einzelner Faktoren waren auf dem Stand der Testrunde 1 zu vermuten, aber nicht im Detail zu berücksichtigen oder gar aufzuklären. Daher wurden die Nebenfaktoren in der Testrunde 2 in die Untersuchung explizit eingebunden. Dies erforderte die Weiterentwicklung von Modellen zu den internen Zusammenhängen: weg vom Ansatz einer Black Box mit Beschränkung auf die Interpretation von Input und Output hin zu einer White Box mit sinnvollen Annahmen zu internen Zusammenhängen, Erwartungswerten und der gezielten Überprüfung in der Auswertung.

Die Verfahrensweise beim Übergang von Testrunde 1 zu Testrunde 2 ergab sich damit wie folgt: Bewährtes und Erprobtes wurden übernommen und fortgeführt, aus gemachten Fehlern und Unzulänglichkeiten gelernt, an kritischen Punkten nachgebessert sowie zwischenzeitlich neue Erkenntnisse und Ideen integriert. Von diesem Vorgehen profitierte die Planung, Durchführung und Auswertung der Testrunde 2 generell. Entsprechend zeigen sich die Unterschiede zwischen der Vorstudie und der Hauptstudie:

- Erweiterung der Zielstellung und des Untersuchungsgegenstandes
- deutlich größerer Testumfang (Faktor 3,5 bezogen auf den Umfang der Stichprobe von Probanden und die erzielten Einzelergebnisse)
- Vertiefung der theoretischen Überlegungen zum Modell des QoS-QoE-Zusammenhangs und zum Konzept zur Ermittlung als QoE-Evaluation
- Optimierung der Testplanung, z. B. durch bessere Eingrenzung interessanter Bereiche
- verfeinerter Testansatz und Testmethodik
- verbesserte Technik und weitere Automatisierungen zur Testdurchführung
- weitergehende technische Dokumentation der Testläufe
- deutlich erweiterter Umfang der Auswertung der erhobenen Daten

Ein Vergleich der Ergebnisse (Längsschnitt) zwischen den Testrunden ist nicht direkt möglich, da insgesamt zu viele Unterschiede existieren. Diese sind durch die vorgenommenen Verbesserungen als auch durch den stetigen externen Wandel begründet, da die relevanten Elemente – neben den Nutzern – selbst schon ständigen Veränderungen unterworfen sind.

1.3.6 Überblick zu den ausgeführten Studien

Einen Überblick zu den Eckdaten der ausgeführten Studien gibt Tabelle 1. Dabei werden die wesentlichen Fakten in Hauptblöcken zusammengefasst dargestellt, die zugehörigen Kapitel geben zur Testrunde 2 dazu jeweils detailliert Auskunft.

Tabelle 1: Eckdaten der ausgeführten Studien im Überblick

Allgemein		Vorstudie (Testrunde 1)	Hauptstudie (Testrunde 2)
Testzweck		Überprüfung der Realisierbarkeit und Erzeugung erster Ergebnisse im kleinen Rahmen	deutliche Erweiterung des Testumfangs und Ausweitung des Untersuchungsgegenstandes, Gewinnung belastbarer Aussagen
Zeitraum der Testdurchführung		Oktober 2012	Juni und August 2013
Testumfang			
Pretest		1 Testslot mit 4 Probanden	1 Testslot mit 4 Probanden + 1 Testslot mit 3 Experten
Testblöcke im Haupttest		1	2 Testblock I im Juni, Testblock II im August
Testslots im Haupttest		13 Testslots	42 Testslots Testslots 1 bis 21 im Testblock I Testslots 22 bis 42 im Testblock II
Probanden im Haupttest, davon in der Auswertung		60 Probanden 60 Probanden	216 Probanden 210 Probanden
Testdesign		<i>UV: QoS, technische Parametrisierung</i>	
Netzparameterkombinationen (Preset)		13	30 insgesamt (5 für jeden der 6 Dienste)
davon disjunkt		13	13
		<i>AV: QoE, Nutzerbewertungen</i>	
erfasste Ergebnisse, generell		Erfolg, Nutzen, Zufriedenheit (gesamt), optionale Freitextanmerkungen	Erfolg, Zufriedenheit (gesamt), optionale Freitextanmerkungen
dienstspezifische Items		Zufriedenheit: 2 bis 6 Teilbewertungen	Zufriedenheit: 2 oder 3 dienstspezifische Teilbewertungen, ggf. Freitextanmerkung (Pflicht)
Testbed			
drahtlose Basistechnologie		WLAN	WLAN
Netz		Netzemulation	Netzemulation
Endgeräte (konkret)		3 unterschiedliche Modelle	5 unterschiedliche Modelle
mobile Plattformen		2 (Android und iOS)	2 (Android und iOS)
Endgeräteformate (Formfaktor)		1 (Smartphone)	2 (Smartphone und Tablet)
Testplätze insgesamt (Probanden parallel)		5	6

Fortsetzung der Tabelle auf der nächsten Seite

Fortsetzung der Tabelle 1

Testparcours	Vorstudie (Testrunde 1)	Hauptstudie (Testrunde 2)
Dienste und Szenarien (Nutzung im Browser oder als App)	7: Amazon.de (Browser), Facebook (App), Google Maps (App), Informationssuche (Browser), Spiegel.de (Browser), WhatsApp (App), Youtube (Browser)	6: Drive (App), Facebook (App), Google Maps (App), Browser), MTV-Music (App), Spiegel.de (Browser), Youtube (App)
Aufgaben-Wiederholungen Aufgabenausführungen	1 bis 3x je Proband 12 je Proband 720 insgesamt	2x je Proband 12 je Proband 2520 insgesamt
Stichprobe (Probanden) Generierung	eigene Gewinnung (MI) innerhalb und außerhalb der Universität	eigene Gewinnung (AAP) innerhalb und außerhalb der Universität sowie über Dienstleister außerhalb der Universität
Auswahl der Probanden	nach Angaben der Vorbefragung mit angestrebter Quotierung	nach Angaben der Vorbefragung teilweise Quotierung erreicht
Vorbefragung/Nachbefragung demografische Angaben	ja	nach Smartphone und Tablet getrennt ja
Plattformerfahrung (eigenes Gerät)	ja	ja
Netzerfahrung (eigener Netzanbieter)	ja	ja
Nutzung ausgewählter mobiler Dienste	nur in der Nachbefragung	nur in der Nachbefragung

1.4 Aufbau der Arbeit

Nachfolgend wird ein kurzer Überblick über den Aufbau der Arbeit, grundsätzliche Konventionen und den Inhalt der einzelnen Kapitel vermittelt. Abbildung 5 skizziert schematisch den Aufbau der Arbeit mit den einzelnen Bestandteilen in Form der Kapitel und den inhaltlichen Zusammenhang der einzelnen Abschnitte untereinander.

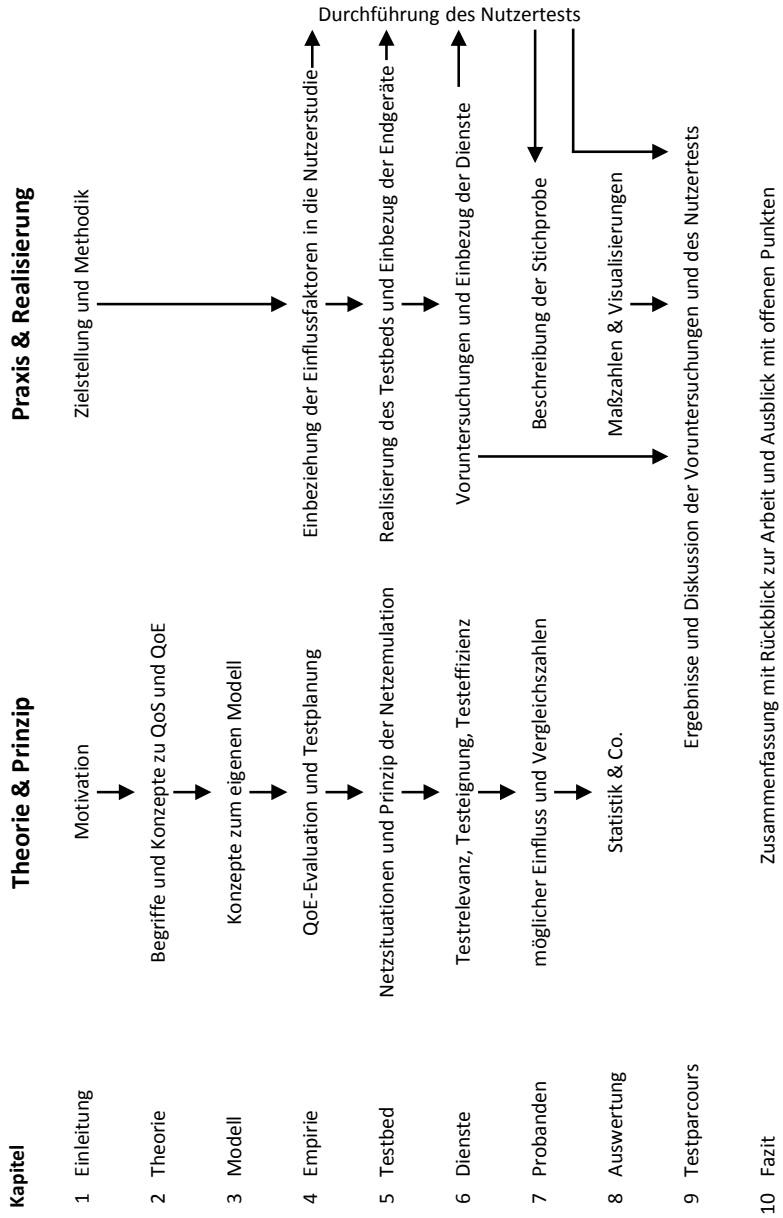


Abbildung 5: Schematische Übersicht zum Aufbau der Arbeit und zum Zusammenhang der einzelnen Abschnitte

Grundsätzliches: Stellenweise erfolgt eine Trennung zwischen dem Prinzip als theoretische Überlegungen und der Praxis als Anwendung dieser Prinzipien zur Realisierung der Studie. Dies wird in der Beschreibung immer dann so aufgetrennt, wenn das entwickelte Prinzip in gleicher Art und Weise mehrfach anzuwenden war, z. B. bei den Prinzipien der Voruntersuchungen und der Auswertung, die für die Darstellung des Testparcours genutzt wurden.

Konventionen: Zur Vermeidung von Verwechslungen oder Doppeldeutigkeiten sollen vorrangig die etablierten englischen Begriffe bzw. ihre auf dem Englischen basierenden Abkürzungen genutzt werden. Für Definitionen und andere wörtliche Übernahmen wird weitgehend in gleicher Art verfahren, um in der ohnehin komplizierten Thematik zusätzliche sprachliche Unschärfe durch Spielraum in der Übersetzung zu vermeiden.

Illustrationen: Die schematischen Darstellungen vom Netz zum Nutzer orientieren sich überwiegend entsprechend der Schreibrichtung in der westlichen Welt von links nach rechts bzw. von unten nach oben, so wie ein Haus vom Fundament zur Spitze aufgebaut ist. In den Illustrationen werden für einzelne Elemente wiederkehrende stilisierte Symbole genutzt, die jeweils immer gleich interpretiert werden können (Nutzer, Endgeräte, Funkantennen, Netzwerke, App-Icons, Server etc.).

Kapitel 2 behandelt die grundsätzliche Theorie zu QoS und QoE. Dafür werden die Begriffe über Definitionen eingeführt und die dazugehörigen Konzepte vorgestellt. Für QoS werden das Netz und die QoS-Netzparameter dabei soweit aufgeschlüsselt, dass die Ende-zu-Ende-Betrachtung und der Zusammenhang zur Network Performance verständlich werden. Für die QoE werden die Einflussfaktoren, der Beurteilungsprozess und relevante Aspekte diskutiert. Dies umfasst neben den technischen Einflussfaktoren (Netz) und dem Menschen als Nutzer auch den Kontext, aus dem sich die Nutzungssituationen ergeben. Abschließend wird der Zusammenhang von QoS und QoE auf konzeptueller Ebene skizziert.

Kapitel 3 umfasst die Modellierung des kausalen Zusammenhangs von QoS und QoE. Dazu wird ein Modell vorgestellt, das die verschiedenen Einfluss-

faktoren und ihr Zusammenwirken auf verschiedenen Schichten beachtet. Dabei wurde besonderer Wert daraufgelegt, dass die relevanten Komponenten aus dem Test entsprechend integriert sind, um auf dieses Modell in den Voruntersuchungen zu den Diensten zurückzukommen und letztlich auch plausible Erklärungen zu den erzielten Ergebnissen liefern zu können. Voruntersuchungen und Testergebnisse selbst lieferten wiederum Einsichten und Erkenntnisse, die in theoretischen Unterbau des Modells zurückgefließen sind. Dabei sind neben dem Einfluss des Endgerätes vor allem die Betrachtungen zu den Diensten von hoher Relevanz. Für die Dienste werden jeweils Netzsicht und Nutzersicht erläutert, die in der gemeinsamen Gesamtsicht zur Beschreibung eines dienstspezifischen QoS-QoE-Zusammenhangs geeignet sind und schließlich zu den Dienstklassen führen, aus denen sich die Rolle der QoS-Netzparameter ableiten lässt.

Kapitel 4 thematisiert die Empirik innerhalb der Arbeit, worauf sich alle weiteren Kapitel beziehen. Das Kapitel beginnt mit einem Exkurs zur QoE-Evaluation und konkretisiert nachfolgend die Testplanung und Vorbereitung durch die Beschreibung der Operationalisierung und der gewählten Testrealisierung. Dabei werden die relevanten Einflussgrößen im Test, die gewählten Ansätze zur Variation und die gemessenen Wirkgrößen zusammengefasst vorgestellt. Betonung verdienen dabei die Besonderheiten des entwickelten Testdesigns. Abgerundet wird das Kapitel durch Anmerkungen zur eigentlichen Durchführung der Studie, die angesichts des Umfangs und notwendigen Aufwands entsprechender Organisation bedurfte.

Kapitel 5 widmet sich der Einbeziehung des Netzes und der mobilen Endgeräte in die Studie. Das Kapitel beginnt mit der Diskussion mobiler Netze (Mobilfunk) und der Analyse, wie sich mobilfunktypische Netzsituationen darstellen. Ausgehend von den gestellten Anforderungen an das technische Konzept wird ein Kombinationsansatz mit einer Netzemulation als geeignet für die Nutzung in den eigenen Untersuchungen eingeführt. Dazu werden Ansatz und Grenzen einer Netzemulation detailliert erläutert. Danach wird die in der Studie genutzte Technik in Form des entwickelten Testbeds und Frameworks vorgestellt. Die Realisierung mit etablierter Hardware und Software bildet dabei die Basis für das entwickelte verteilte Framework, das mit verschiedenen APIs die benötigten Funktionalitäten in einem konkreten Testbed mit WLAN-

Technologie als drahtloser Netzzugang zu mobilen Endgeräten bereitstellt. Die Kopplung zum elektronischen Fragebogensystem zur automatischen Testlaufsteuerung sowie flexible Mechanismen zur Netzverkehrs- und Geräteüberwachung runden die technischen Möglichkeiten des Testbeds ab.

Kapitel 6 näherten sich der Untersuchung mobiler Dienste von einer theoretischen Sichtweise aus an. Es werden die entwickelten Prinzipien für Vorüberlegungen und Voruntersuchungen zu den Diensten vorgestellt. Die dabei identifizierten Aspekte wie Testrelevanz, Testeignung und Testeffizienz sind neben der Technik des Testbeds eine wichtige Grundlage für eine erfolgreiche Realisierung der Studie und ein bedeutender Schritt auf dem Weg von der Black Box zur White Box bei der Untersuchung der Dienste. Aus der Anwendung dieser Prinzipien konnten wertvolle Erkenntnisse für die weitere Testplanung, für eine sinnvolle Parametrisierung und letztlich auch zur Interpretation der erzeugten Ergebnisse zu den Diensten gewonnen werden.

Kapitel 7 beschreibt die Probanden, also die Stichprobe im Test. Dazu wird kurz auf die Rolle des Menschen als Proband und mögliche Implikationen für den Test eingegangen. An die allgemeinen Aussagen zur Auswahl und Generierung der Stichprobe schließt sich die Vorstellung der erfassten Attribute der Nachbefragung wie Alter, Geschlecht, Bildung etc. zur Charakterisierung der Stichprobe an. Dabei werden die ermittelten eigenen Werte mit Vergleichszahlen einer möglichen Grundgesamtheit ergänzt und die mögliche Bedeutung der abgefragten Merkmale diskutiert. Auf Basis der Nachbefragung können verallgemeinerbare Aussagen zur Popularität von mobilen Diensten unter den Probanden generiert werden. Gleichzeitig erfolgt eine Charakterisierung der Probanden (Nutzertypen) über die erhobenen Daten. Dies schließt Analysen zur Verteilung der Nutzungsintensität der Dienste im mobilen Internet und des allgemeinen Anspruchsniveaus ein.

Kapitel 8 beschäftigt sich damit, wie die Auswertung der gewonnenen Daten des Tests angegangen wurde. Dazu werden die Prinzipien und einzelnen Bestandteile der eigenen Auswertung vorgestellt. Dies umfasst auch, welche statistischen Grundannahmen dafür notwendig und sinnvoll sind. Neben den theoretischen Erörterungen zu den möglichen Aussagen und praktikablen Metriken zur Bezifferung der Ergebnisse werden erste allgemeine Ergebnisse

präsentiert. Weitergehende Ergebnisse zu den untersuchten Diensten werden hingegen erst später unter Anwendung der erläuterten Prinzipien und entwickelten Visualisierungen im nächsten Kapitel gezeigt.

Kapitel 9 beinhaltet die Beschreibung der ausgewählten und untersuchten Dienste der eigenen Studie, den Testparcours. Dafür werden die Ergebnisse der Voruntersuchungen zu diesen Diensten unter Anwendung der Prinzipien aus Kapitel 5 und die gewonnen empirischen Daten, aufbereitet nach den Prinzipien und Bestandteilen der Auswertung aus Kapitel 8, präsentiert und diskutiert. Die Dienste werden dabei einzeln für sich in gleicher Systematik jeweils in einem Unterkapitel vorgestellt und abschließend zur Verdeutlichung der Gemeinsamkeiten und Unterschiede zusammengefasst.

Kapitel 10 bildet den Abschluss der Arbeit. Dazu werden in einem Blick zurück die erreichten Ziele und Erkenntnisse zusammengefasst und weitere, auch neu entstandene Fragestellungen und Ideen, in einem Blick nach vorn angeschnitten.

Im **Anhang** finden sich ergänzende Abbildungen und Tabellen zu den einzelnen Kapiteln. Besonders hervorzuheben sind der genutzte Fragebogen mit den Illustrationen zur Sicht der Probanden auf den Testparcours und die erhobenen Rohergebnisse zur Dokumentation.

2 Die Theorie

Begriffe und Konzepte zu Quality of Service und Quality of Experience

In diesem Kapitel werden die Begriffe und Konzepte zum Verständnis von QoS und QoE vermittelt. Dem vorangestellt ist eine Diskussion zu den Begriffswelten und zur Qualität, um zu verdeutlichen, dass diese mehrere Bedeutungen haben kann und im Fall von QoS und QoE auch hat.

QoS und QoE sind als komplexe Konzepte selbst nur schwierig in einzelnen Definitionen zu fassen. Daher wird auch jeweils für QoS und QoE auf die Bestandteile und Prozesse zur Formierung eingegangen. Dabei sind das Netz für QoS und der Mensch als Nutzer für QoE als die jeweiligen Kernbestandteile nicht von der grundsätzlichen Diskussion der Konzepte zu trennen und werden gemeinsam mit diesen in jeweils einem Abschnitt erörtert. Daran schließen sich die Erläuterungen zum Kontext an, der Mensch und Maschine in der Diskussion miteinander verbindet und zu den Nutzungssituationen führt.

Nach den isolierten Diskussionen zu QoS und QoE sollen die beiden Konzepte gemeinsam betrachtet werden. Ausgehend von ihrem konzeptuellen Zusammenhang wird daher verdeutlicht, warum es die mittelbare Sicht der Nutzer über die Dienste auf das Netz ist, um den gesuchten Zusammenhang zu formalisieren. Die Überlegungen führen zu den konkreten Untersuchungen im Sinne des interessierenden Ursache-Wirkungs-Zusammenhangs, der dann im nächsten Kapitel modelliert wird.

2.1 Begriffswelten: Von QoS zu QoE

Für den Einstieg zu den relevanten Begriffen und Konzepten soll zunächst einmal auf die Anwendungsbereiche der QoS und QoE eingegangen und die interdisziplinären Anknüpfungspunkte aufgezeigt werden. Den weiteren Ausführungen zu den Konzepten rund um QoS und QoE wird die Diskussion zum grundlegenden Begriff der Qualität vorangestellt, um Missverständnissen

durch oftmals unbewusste, weil umgangssprachlich falsche Anwendung vorzubeugen.

2.1.1 Anwendungsbereiche zur QoS und QoE: Netz, Nutzer, Multimedia

Typischerweise wird dem Leser am Anfang eine Einführung auf der Basis etablierter Definitionen geboten. So soll versucht werden, durch eine begriffliche Annäherung die relevanten Aspekte der Thematik ein- und voneinander abzugrenzen. Dabei sind Begriffe und Bedeutung spezifisch für einen bestimmten Anwendungsbereich bzw. ein Themengebiet. Die nachfolgenden Überlegungen vereinen aber eigentlich sogar unterschiedliche (Begriffs-) Welten: Netze, Nutzer und Multimedia.

Netzwelt: Innerhalb der Netzwelt ist eine historisch bedingte Differenzierung festzustellen, sodass es eigentlich mehrere Netzwelten sind, die zu betrachten sind, nämlich Telkonetze (ursprünglich zur Telekommunikation, d. h. im Ursprung Sprachtelefonie zwischen Menschen) vs. Datennetze (zum Austausch von Daten):

- die „klassische“ Telkowitz der leitungsvermittelten zunächst analogen¹⁵ Telekommunikationsnetze und der dann später digitalen¹⁶ Telefonnetze und ihre mobile Pendant (Mobilfunknetze)
- paketbasierte Netze auf Basis des Internet Protocols (IP) als geschlossene Netze und unter einheitlicher Kontrolle, z. B. die IP-basierten Netze (All-IP) der Next Generation Networks (NGN) in der Telkowitz oder unternehmensinterne Netzwerke in Welt der Datennetze
- das „offene“ Internet, als IP-basiertes Netz „zwischen“ den IP-Netzen unterschiedlicher administrativer Kontrolle

Insgesamt interessiert das (mobile) Internet. Im Detail ist dabei aber oftmals doch die Sicht einer anderen Netzwelt einzunehmen, eben gerade weil „das“ Internet viele ganz unterschiedliche IP-Netze vereint und die jeweiligen Besonderheiten zu beachten sind. Insbesondere der Bereich der nutzerseitigen

¹⁵ Public Switched Telephone Networks (PSTN) und die scherzhafte Umschreibung als Plain Old Telephone System (POTS), das zu kaum mehr als dem „normalen“ Telefonieren taugte

¹⁶ Integrated Services Digital Network (ISDN)

Zugangsnetze zum Internet hängt historisch bedingt eng mit klassischen Telekommunikationsnetzen und deren Strukturen zusammen. Teilweise ist dabei gerade der Blick auf einzelne Punkte (Flaschenhals) entscheidend. Dabei bewegen sich die Betrachtungen zur Netzwelt zwischen Elektrotechnik und Informatik, wobei die Auseinandersetzung mit den Eigenschaften des Funkkanals etc. klassischerweise das Arbeitsgebiet der Elektrotechnik ist. Ab der Ebene der Bits und Bytes, der Protokolle, Schnittstellen, Datenstrukturen und Algorithmen ist die Netzwelt auch Teil der Forschung zur Informatik. Wie angedeutet, kann innerhalb der Netzwelten eine weitere Differenzierung zwischen drahtlosen (mobilen) Netzen und drahtgebundenen (Fest-) Netzen erfolgen. Die weitere Betrachtung soll dabei zunächst möglichst abstrakt bleiben, bevor die Besonderheiten der mobilen Netze erörtert werden.

Nutzerwelt: Die Auseinandersetzung mit der Wahrnehmung und den Denkprozessen des Menschen ist eigentlich das Metier der Psychologie und Kognitionswissenschaft. An der Schnittstelle von Mensch und Maschine, besonders in Bezug auf interaktive Computersysteme und Anwendungen sind Überlappungen zur Usability-Forschung und User Experience festzustellen, die sowohl von psychologischer Seite wie auch von der Informatik, hier konkret der Medieninformatik, bearbeitet werden. Dabei spielt die Forschung zur Mensch-Maschine-Interaktion (HCI¹⁷) eine besondere Rolle.

Multimediawelt: Bei der modernen Netznutzung geht es nicht mehr vorrangig um klassische Telefondienste, sondern um die Nutzung multimedialer Dienste in Apps auf smarten mobilen Endgeräten im mobilen Internet. Neben der grundsätzlichen Umsetzung (Software-Technik) spielen für die mobile Nutzung die Nutzerschnittstellen und der passende Inhalt in Form statischer und dynamischer Medien wichtige Rollen. In der Forschungslandschaft sind diese Aspekte im Bereich der Informatik und dabei speziell in der Medieninformatik angesiedelt.

¹⁷ Human Computer Interaction (HCI) als gebräuchlicher Begriff, mit Computer stellvertretend für digitale Maschinen, der in seiner englischen Abkürzung nachfolgend genutzt wird

Die nachfolgenden Definitionen und Erläuterungen sind damit bezogen auf Netze und Systeme zur Telekommunikation und deren Nutzung für multimediale Inhalte, Anwendungen bzw. Dienste im mobilen Internet durch menschliche Nutzer.

2.1.2 Problematik der gemeinsamen Betrachtung von QoS und QoE

Im Fall von QoS und QoE ist allein schon das begriffliche Einfangen der Thematik ein schwieriges Unterfangen. Beim Versuch der Zuweisung von Bedeutung zu den Termen QoS und QoE ist schnell festzustellen, dass sich einerseits die Begriffswelten von QoS und QoE überschneiden und andererseits oftmals noch eine begriffliche Ausweitung zur Darstellung von Gemeinsamkeiten und Unterschieden notwendig ist. So ist beispielsweise neben der Betrachtung zu QoS und QoE auch das Verhältnis von QoS zur Network Performance und von QoE zur User Experience zu beachten.

Zusätzlich existieren jeweils eine Vielzahl von Definitionen, für die sich teilweise das Verständnis sowie die Interpretation dazu über die Zeit verschoben haben (QoS), bzw. bei genauer Betrachtung festzustellen ist, dass sich eigentlich noch gar keine allgemein akzeptierte State-of-the-Art-Definition herausgebildet hat (QoE). Erschwerend kommt hinzu, dass die Begriffe von QoS und QoE nur in einer multidisziplinären Betrachtung der zuvor diskutierten Welten in den gewünschten Zusammenhang zu bringen sind. Allein daraus ergeben sich unterschiedliche Sichten aber auch Begriffsverständnisse. Entsprechend ist es nachfolgend kaum möglich, einem Begriff genau eine Definition gegenüberzustellen, die jeweils allen Ansprüchen gerecht wird. QoS und QoE sind vielmehr komplexe Konzepte, die nachfolgend entsprechend dargelegt werden sollen, um sowohl die Reichhaltigkeit der damit verbundenen Ansätze als auch die weitere begriffliche Verwendung von QoS und QoE innerhalb der vorliegenden Arbeit zu vermitteln. Dabei ergibt sich allerdings, dass teilweise auch konkurrierende Konzepte und Definitionen festzustellen und nicht zu vermeiden sind. In diesen Fällen wird für das eigene Verständnis innerhalb der vorliegenden Arbeit eine zielführende Diskussion angestrebt, die zur gemeinsamen Gesamtbetrachtung von QoS und QoE beiträgt. Daher sollen trotz der teilweise erschlagenden Vielzahl von Definitionen zur QoS und QoE diese nicht ausgelassen werden, da sie ein schöner Spiegel dessen sind, was alles zu QoS und QoE zählt.

2.1.3 Die Qualität: Zwischen Beschaffenheit und Eignung

Die nachfolgenden Überlegungen erscheinen vielleicht zunächst etwas überflüssig, weil zu philosophisch – aber sie sind überaus wichtig. Schon so ein vermeintlich einfacher Begriff wie die Qualität bringt einige Probleme mit sich, besonders in einer eher umgangssprachlichen Nutzung, z. B. als besonders hohe Güte/Vortrefflichkeit (positiv konnotiert). Für den allgemeinen Begriff der Qualität gibt es grundsätzlich verschiedene Bedeutungen, wie sie z. B. die DIN¹⁸ 55350-11 auflistet und kurz einordnet (siehe nichtnormativer Anhang in [28] S. 10f).

Im Weiteren wird auf die Qualität in verschiedenen Bedeutungen zurückgekommen, die unbedingt auseinander zu halten sind (nach [28] S. 10f, [29], [30] S. 5ff und [31] S. 14 zusammengestellt, erweitert und ergänzt):

- Qualität als reine **Feststellung** oder Wahrnehmung der Gesamtheit aller Eigenschaften, z. B. eines Objektes oder Systems: Dies entspricht damit der *neutralen Beschaffenheit* bzw. **Charakteristik** (z. B. als *wie viel*). Damit verbunden sind die **Qualitätskriterien**.
- Qualität als **Abgleich** zwischen einem **Ist** (realisierte Beschaffenheit aus inhärenten Merkmalen) und einem **Soll** (geforderte Beschaffenheit). Die Bewertungsmöglichkeiten können dabei stetig oder diskret sein, z. B. von *sehr schlecht* zu *sehr gut*.
- Qualität als **Bewertung** der Beschaffenheit, also ein Abgleich der „Eigenschaften über ihre **Eignung** zu einem bestimmten **Zweck**“ [29] S. 353. Die Bewertung erfolgt dabei graduiert auf einer Skala, von *gut* bis *schlecht*, *hoch* bis *niedrig* bzw. mit anderen Ausprägungen von *geeignet* zu *ungeeignet* zum bestimmten Zweck, d. h. als **Qualitätsurteile** [29] S. 353.

Auf diese unterschiedlichen Bedeutungen können die Definitionen zur Qualität wie in QoS und QoE sowie ihr konzeptueller Zusammenhang zurückgeführt werden. Werden diese Bedeutungen vermischt, oder geht aus Definitionen nicht klar hervor, welche Art von Qualität gemeint ist, sind Fehlinterpretationen und Missverständnisse möglich. Über die Worte im Fettdruck sollte sich meist die jeweilige Bedeutung unterscheiden lassen. Für die

¹⁸ DIN: Deutsches Institut für Normung e.V., siehe <http://www.din.de> abgerufen am 05.02.2015

Synonyme der Qualität gelten die Ausführungen zu den verschiedenen Bedeutungen entsprechend ebenfalls. Es ist jeweils zu überlegen, welche der Bedeutungen greift.

Bei den Qualitätsurteilen (für die Eignung zu einem Zweck) kann zwischen relativen und absoluten Bewertungen unterschieden werden [29] S. 353. Relative Qualitätsbewertungen vergleichen, in dem sie eine Rangreihe bilden (besser als). Absolute Qualitätsbewertungen beziehen sich auf ein Ideal, also auf einen bestmöglichen Wert. Dieses Ideal kann entweder global oder auch lokal definiert werden, z. B. durch Eingrenzung der Bezugsgröße [29] S. 353. Auf diese Punkte wird in der Diskussion der QoE-Maße als subjektive bzw. objektive Qualität noch einmal zurückgekommen.

Als Ausgangspunkt der weiteren Ausführungen, vor allem mit Blick auf die Definitionen zur QoS und QoE, lohnt die Betrachtung zur Entwicklung allgemeiner Qualitätsdefinitionen. [30] S. 5 verweist dabei auf den Bedeutungswandel der Qualität, der auch in Standards (ISO¹⁹ bzw. DIN) nachzuvollziehen ist.

Eine der Definitionen ergibt demnach die Qualität als:

„Gesamtheit aller Charakteristika einer Einheit bezüglich ihrer Eignung, festgelegte und vorausgesetzte Bedürfnisse zu erfüllen“ [30] S. 5 (mit Verweis auf die DIN 55350 Teil 11 in der älteren Version von 1987 bzw. die ISO 8402).

[32] S. 2 verweist auf die ISO 9000 (Qualitätsmanagement), die den graduellen Abgleich zwischen einem Soll und einem Ist zum Qualitätsbegriff hinzufügt, der sich auch in weiteren Definitionen zur Qualität wiederfindet. Die äquivalente deutsche DIN EN ISO 9000:2005 definiert die Qualität als:

¹⁹ International Organization for Standardization, siehe <http://www.iso.org> abgerufen am 05.02.2015

*„Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt.“²⁰
[33] S. 25*

Noch weiter geht der „modernere menschbezogene Qualitätsbegriff“ [30] S. 5, mit einer Qualität, die explizit auf den Menschen als Beurteiler und Bezugsgröße im Beurteilungsprozess abzielt [30] S. 5. Dabei wird dem Abgleich bzw. Vergleich auch der elementare Bestandteil des Verständnisses einer Beurteilung hin zu einer Eignungsbewertung aufgegriffen:

„Ergebnis eines Wahrnehmungs- und eines Beurteilungsprozesses, bei dem der/die Beurteilende die wahrgenommenen Merkmale einer Einheit vergleicht mit individuellen Erwartungen, sachgerechten Erfordernissen oder gesellschaftlichen Forderungen.“ [30] S. 5 mit weiteren Verweisen

Im Verlauf der weiteren Ausführungen wird sich zeigen, dass zumindest in einigen der Definitionen zu QoS und QoE jeweils eine der Definitionen zur Qualität wiederzuerkennen ist.

2.2 Quality of Service und das Netz

Überlegungen zur Quality of Service sind typisch für Netzwerke mit Paketorientierung und der Nutzung des Internet Protocols (IP). Es erfolgt eine begriffliche Annäherung ausgehend von Definitionen zur QoS. Dabei sind deutliche Unterschiede zu erkennen, die eine kritische Auseinandersetzung zu Begriff und Konzept erfordern. In dieser Diskussion wird klargestellt, wie QoS, Netz und Netzleistung zueinander zu sehen sind. Nach der Diskussion der KPI-Hierarchie-Ebenen der Netzqualität kann die weitere Betrachtung auf die leistungsbezogenen QoS-Parameter konzentriert werden. Nachfolgend sollen stellvertretend für das Netz zelluläre mobile Netze zur Datenübertragung, kurz Mobilfunk, berücksichtigt werden. Dabei soll aber noch nicht in technische Details der Funkübertragung oder internen Funktion der einzelnen Netzelemente eingetaucht werden. Daher erfolgt stattdessen zunächst eine sinnvolle Abstraktion der beteiligten Netzsegmente. Dazu wird erörtert, wie aus den jeweiligen Netzleistungen die Ende-zu-Ende-QoS des Netzes entsteht. Diese

²⁰ Ergänzend die englische Version der ISO 9000 für spätere Vergleiche der Definition von QoS und QoE: „degree to which a set of inherent characteristics fulfils requirements“ [33] S. 25

lässt sich in geeigneter Art durch QoS-Parameter beschreiben, aus denen die für die weitere Diskussion relevanten Leistungshauptparameter ausgewählt werden. Die Bedeutung der Werte ergibt sich in engem Zusammenhang mit Garantien, QoS-Mechanismen und Netzverkehrsklassen bzw. dem Fehlen eben dieser Möglichkeiten. Auf die weiteren Aspekte eines Mobilfunknetzes, die zu den mobilen Netzsituationen führen, wird in 5.1 zusammen mit der Berücksichtigung des Netzes im Test weiter eingegangen.

2.2.1 Definitionen zur QoS

Im Bereich QoS existieren zahlreiche Definitionen dazu, was QoS eigentlich ist, was QoS beinhaltet, wie QoS beschrieben oder erfasst werden kann. Neben vielen Ad-hoc-Definitionen in zahllosen Publikationen existieren auch Definitionen in Spezifikationen und Standards von ITU²¹, ETSI²², 3GPP²³ und IETF²⁴, die zumindest eine gewisse Grundlage für das allgemeine Verständnis liefern können. Zunächst sollen einige Ad-hoc-Definitionen herangezogen werden, danach folgen solche aus den o. g. Organisationen.

Eine gebräuchliche Übersetzung von Quality of Service ins Deutsche ist *Dienstgüte*. Steinmetz liefert eine allgemeine Definition für die technische Dienstgüte, die schon einige wichtige Anhaltspunkte liefert:

„Dienstgüte kennzeichnet das definierte, kontrollierbare Verhalten eines Systems bezüglich quantitativ messbarer Parameter.“ [21] S. 238

²¹ ITU: International Telecommunication Union, die Unterorganisation der Vereinten Nationen für Informationstechnik und Telekommunikation, siehe <http://www.itu.int> abgerufen am 10.10.2013

²² ETSI: European Telecommunications Standards Institute, ein europäisches Standardisierungsgremium für Informations- und Kommunikationstechnik, siehe <http://www.etsi.org> abgerufen am 10.10.2013

²³ 3GPP: 3rd Generation Partnership Project, verantwortlich für die Entwicklung der Mobilfunksysteme der dritten und weiterer Generationen, siehe <http://www.3gpp.org/> abgerufen am 11.03.2014

²⁴ IETF: Internet Engineering Task Force, siehe <http://www.ietf.org/> abgerufen am 10.10.2013

Für die unterschiedlichen Netzwelten ergeben sich jeweils auch spezifische Definitionen zur QoS, die sich innerhalb nur eines einzigen Werkes finden lassen, z. B. in [34]:

- in der ISO/OSI-Welt²⁵, bezogen auf die Transportschicht: *„Die Dienstgüte (QoS Quality of Service) beschreibt die Qualitätsmerkmale, die ein Dienstbringer seinem Dienstanutzer zur Verfügung stellen kann.“* [34] S. 117
- in der Welt der Rechnernetze: *„QoS (Quality of Service, Dienstgüte) beschreibt die Eigenschaften eines Kommunikationssystems bezüglich der für einen bestimmten Dienst erbrachten Leistung. Dienstgüteattribute sind Leistung und Leistungsschwankungen, Zuverlässigkeit und Sicherheit.“* [34] S. 163
- in der Internet-Welt: *„Unter Dienstgüte (QoS: Quality of Service) versteht man das Verhalten eines Netzes, das bestimmte Arten von Verkehr bevorzugt gegenüber anderen Arten behandelt. Damit ist zunächst nichts über die verwendeten Mechanismen oder die Art und quantitative Festlegung der erreichbaren Güte gesagt.“* [34] S. 389, ergänzend wird auf die Definition von Steinmetz bezuggenommen.

Die ITU-T-Empfehlung E.800²⁶ [35] definiert die wichtigsten Begriffe zum QoS-Umfeld in der Telekommunikation und wird als der kleinste gemeinsame Nenner zur Begriffsbestimmung oft zitiert. Insbesondere bei der ITU darf man den langen historischen Hintergrund dabei aber nicht unbeachtet lassen. Die Study Group 12²⁷, bereits 1957 gegründet und in der POTS-Netzwelt verankert, beschäftigt sich seit 1990 mit allem rund um (Network) Performance, QoS und QoE, was neben den Netzen auch Dienste und Endgeräte umfasst [36] S. 4. Die Definition der ITU ist damit ein gutes Beispiel, was QoS im Verständnis der Telkowitz ist, bzw. sein sollte:

“Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.”
[35] S. 3

²⁵ ISO: International Organization for Standardization, OSI: Open Systems Interconnection

²⁶ ITU-T Recommendation E.800: Definitions of terms related to quality of service

²⁷ ITU-T SG12: Performance, QoS and QoE <http://www.itu.int/ITU-T/studygroups/com12> abgerufen am 07.12.2014

QoS ist demnach eine Charakteristik, also eine Menge zusammengehöriger Qualitätskriterien, bei der Nutzung eines technischen (Produktes, Systems oder) Dienstes. Die Bedürfnisse des Nutzers werden dabei als Bezugsgröße angeführt, der Nutzer steht selbst aber nicht im Mittelpunkt.

Die ETSI, neben der ITU eine der Größen im Bereich Standards und Normen in Europa, folgt in Bezug auf die Definition der QoS weitgehend der ITU. Innerhalb der ETSI werden verschiedenen QoS-Sichtweisen auch in unterschiedlichen technischen Komitees (TC) bearbeitet. Dabei widmet sich ETSI TISPAN²⁸ der Network-QoS mit zugeordneten QoS-Methoden und QoS-Mechanismen, während sich ETSI TC STQ²⁹ um Ende-zu-Ende-QoS und entsprechende Leistungsmerkmale kümmert. Die Terminologie der ETSI und ihrer Unterarbeitsgruppen, z. B. der STF360³⁰, zur QoS ist leicht abgewandelt aber prinzipiell vergleichbar zur ITU-T.

Die ETSI übernimmt dabei auch die Aufgaben zur formellen Spezifikation der 3GPP-Standards der Mobilfunkwelt. Innerhalb des Vokabulars der 3GPP ist QoS demnach:

“The collective effect of service performances which determine the degree of satisfaction of a user of a service. It is characterised by the combined aspects of performance factors applicable to all services, such as; service operability performance; service accessibility performance; service retainability performance; service integrity performance; and other factors specific to each service.” [37] S. 25

Der erste Teil der Definition entspricht damit der QoS-Definition der ITU-T E.800 in einer früheren Version aus dem Jahr 1994 (vergleiche [38] S. 3), ergänzt um die Auflistung relevanter Leistungsfaktoren (Key Performance Indicators, KPI). Grundsätzlich ist die Definition der Qualität als Charakteristik zu

²⁸ Telecommunications and Internet converged Services and Protocols for Advanced Networking, <http://www.etsi.org/tispan/> abgerufen am 06.12.2014

²⁹ Technical Committee (TC) Speech and multimedia Transmission Quality (STQ), <https://portal.etsi.org/tb.aspx?tbid=275&SubTB=275,634> abgerufen am 06.12.2014

³⁰ Specialists Task Force 360, http://portal.etsi.org/stfs/stf_homepages/STF360/STF360.asp abgerufen am 06.12.2014

erkennen, die selbst wiederum den Grad der Zufriedenheit der Nutzer beeinflussen kann, aber nicht mit dieser gleichgesetzt ist.

Für die IETF, treibende Kraft für viele erfolgreiche Quasistandards im Internet (Requests for Comments, RFC), stehen mehr die praktische Anwendung von Verfahren und die dazugehörigen Protokolle zur Nutzung innerhalb der IP-Netze des Internets im Vordergrund. Sie lässt den Nutzer konsequent aus der Definition für QoS heraus und fokussiert auf die Sicht des Netzes zum Datentransport, womit die QoS definiert ist als:

“A set of service requirements to be met by the network while transporting a flow.” [39] S. 2

Die Verfahren sowie Funktionen im Netz und Protokolle wie IntServ oder DiffServ werden unter dem Begriff QoS-Mechanismen zusammengefasst. Diese stehen, wie in der Einleitung bereits geschrieben, nicht im Zentrum der eigenen Betrachtungen, trotzdem sind auch die Definitionen der IETF zu einzelnen Aspekten und QoS-Parametern durchaus interessant.

Die Definition des DSL-Forums³¹ betont explizit die Perspektive des Netzes als Leistung auf der Paketebene in der QoS und geht daneben auf die QoS-Mechanismen ein:

“Quality of Service (QoS): - is a measure of performance at the packet level from the network perspective. Quality of Service (QoS) also refers to a set of technologies (QoS mechanisms) that enable the network operator to manage the effects of congestion on application performance as well as providing differentiated service to selected network traffic flows or to selected users” [40] S. 7

Im Gegensatz zur ersten sehr allgemeinen Definition zu dem, was Dienstgüte auf einer höheren konzeptuellen Ebene ist, werden in den technisch geprägten Definitionen bereits wichtige Bestandteile genannt, die nachfolgend noch detailliert erörtert werden.

³¹ eine Industrievereinigung, firmiert ab 2008 als Broadband-Forum, siehe <http://www.broadband-forum.org> abgerufen am 25.08.2013

Aus den unterschiedlichen Definitionen ergibt sich der Bedarf für eine kritische Auseinandersetzung dazu, was im Weiteren unter QoS zu verstehen ist, bzw. was darunter verstanden werden soll. Dieser Gedanke führt zu Quality of Something – als etwas andere Deutung des Akronyms QoS.

2.2.2 Quality of Something: Bedeutungsunterschiede und Bedeutungsveränderung der QoS

Beim Vergleich der Konzepte zur QoS hinter den Definitionen der verschiedenen Netzwelten zeigt sich ein deutlicher Unterschied im Umfang dessen, was alles zur QoS zählt. Dies reicht von der holistischen Sicht der ITU-T in der Telkowitz bis zum eher knappen Bedeutungsumfang in der Internet-Welt. Prinzipiell ergibt sich daraus ein Konflikt in der weiteren Darstellung. Wie unter den Begriffswelten erläutert, ist der Anwendungsrahmen der gesamten Erläuterungen auf Dienste im mobilen Internet bezogen, was im Fall von Mobilfunk selbst wiederum aber über Netze der Telkowitz realisiert wird. Daher soll kurz erläutert werden, wie das weitere Verständnis zur QoS einzuordnen ist.

Im ursprünglichen Verständnis der ITU umfasst die QoS einen Teil Network Performance und einen Teil Non Network Performance, siehe Abbildung 6. In den Bereich der Network Performance ist der genutzte Dienst (ursprüngliche Telkowiedienste) sowie das Netz und damit die auch Verbindungsqualität einzuordnen. Auf der Seite der Non Network Performance sind die nichttechnischen Aspekte angesiedelt. Nach ITU-T Richtlinie E.800 zählen Eigenschaften des Angebots (z. B. Umfang, Struktur, Tarifierung) und des Kundenservices (z. B. Bereitstellung, Fehlerbehebung, Hotline und Rechnungswesen) zu diesen nichttechnischen QoS-Leistungsfaktoren, siehe [41] S. 42ff. Welche konkreten Fragestellungen für die Kundenzufriedenheit dabei relevant sein können, zeigen beispielsweise branchenspezifische Studien, wie. z. B. zur Mobilfunkbranche in [42]. Diese Art der Fragestellungen ist im Weiteren nicht im Zentrum des Interesses.

In beiden Teilen, also sowohl Network Performance als auch Non Network Performance, soll im ITU-T-Verständnis zur QoS die Befriedigung der Nutzer- bzw. Kundenbedürfnisse erreicht werden. Insgesamt wurde die QoS also mit einem hohen Anspruch einer ganzheitlichen Sicht definiert und sollte auch so verstanden werden. Der ebenfalls damit verbundene Service-Begriff wird an späterer Stelle noch einmal aufgegriffen und zum eigenen Begriff der Dienste abgegrenzt.

Wie so oft klaffen formale Definition, ursprüngliche Intention und allgemeines Verständnis als freie Interpretation meist weit auseinander. [43] S. 15 weist mit Recht darauf hin, dass das Verständnis von QoS in Wissenschaft und Technik oftmals ganz anders war als die skizzierte ganzheitliche Sicht, bzw. sich im Lauf der Zeit anders entwickelt hat. Überwiegend wird als QoS dabei nur der technische Anteil gesehen, meist ausschließlich bezogen auf das Netzwerk bzw. dessen Leistung, wobei der Nutzer und seine Anforderungen bzw. Bedürfnisse überwiegend komplett ausgeklammert bleiben. Implizit wird QoS als Begriff somit statt einer umfassenden Dienstgüte in einer ganzheitlichen Sicht meist eher mit Netzcharakteristik gleichzusetzen sein. Diese führte zu einer notwendigen Bedeutungsveränderung der QoS. Entsprechend ist eine Reduzierung der QoS auf die Beschreibung der Eigenschaften einer Sicht Ende-zu-Ende auf „das Netz“ und die Netzleistung üblich und möglich. Über die QoS werden demnach die technischen Anforderungen zur Nutzung eines Dienstes ans Netz definiert, genauso wie über die QoS die Eigenschaften des Netzes bei der Nutzung eines Dienstes vermittelt werden.

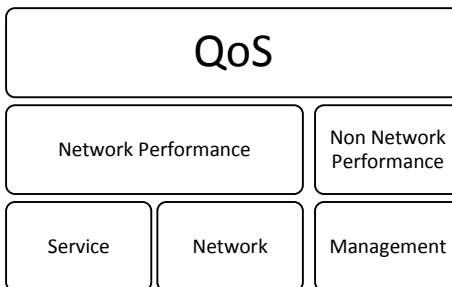


Abbildung 6: Zusammensetzung der QoS aus Network Performance und Non Network Performance

im ganzheitlichen Verständnis der ITU-T, in Anlehnung an [35] S. 2

Aus den unterschiedlichen Definitionen der Netzwelten, Sichten darauf und der beschriebenen Bedeutungsveränderung kann man dazu kommen, QoS als Abkürzung für *Quality of Something* zu sehen, also von Etwas. Hier soll dieses Etwas konkret das Netz sein. Mit der Begrenzung auf den Teil des Netzes und der Netzleistung soll zunächst die Diskussion zu QoS fortgesetzt werden, bevor am Ende des Kapitels versucht wird, diese Aspekte zusammen mit der QoE wieder in einen größeren Zusammenhang zu setzen.

2.2.3 Netzqualität und Key Performance Indicators im Mobilfunk

Die für Mobilfunknetze relevanten Leistungsmerkmale, bzw. Key Performance Indicators (KPIs), lassen sich als aufeinander aufsetzende Ebenen, ähnlich einer Bedürfnispyramide, hierarchisch strukturiert definieren [41] S. 43, siehe Abbildung 7. Auf unterster Ebene steht die prinzipielle Verfügbarkeit, auf oberster Ebene die Integrität einer Dienstnutzung im Netz, aus der sich die interessierende und aus Anbietersicht wichtige Nutzer-/Kundenzufriedenheit ableiten lässt.

Die Erfüllung der Anforderungen an die tieferliegenden Ebenen ist Voraussetzung für höhere Ebenen. Die nachfolgenden Überlegungen beziehen sich darauf, dass die Anforderungen der darunterliegenden Ebenen erfüllt sind, womit davon auszugehen ist, dass zumindest in irgendeiner Weise eine vollständige Nutzung prinzipiell möglich ist. In der KPI-Gesamtsicht ist „Kein Netz“ zwar auch eine typische Mobilfunksituation, aber keine, die weiter betrachtet wird. Ebenso werden netzimmanente Ausfälle und Unterbrechungen etc., die eine vollständige Nutzung verhindern, in den Betrachtungen ausgeklammert.

Eine Ergänzung zur KPI-Hierarchie bilden die 3 C aus: Coverage, Capacity und Customer Experience. Eine ausreichende Netzabdeckung (Coverage) ist Grundvoraussetzung dafür, dass überhaupt ein Dienst angeboten oder genutzt werden kann. Ist eine Netzabdeckung gewährleistet, so muss sichergestellt sein, dass auch ausreichend Netzkapazität (Capacity) vorhanden ist. Die Aspekte von Abdeckung und Kapazität hängen eng miteinander sowie mit der Struktur und Technologie des Netzes zusammen. Aus Coverage und Capacity, Netzauslastung, Empfangsbedingungen und Eigenschaften des Endgerätes ergeben sich die möglichen QoS-Parameterwerte als Netzsituation, die selbst wiederum im untersuchten QoS-QoE-Zusammenhang in Nutzerzufriedenheit

und weiter in Kundenzufriedenheit (Customer Experience) übersetzt werden. Mehr dazu wird in 5.1 erläutert, zunächst sollte dies für das Verständnis ausreichend sein. Die nachfolgenden Erläuterungen konzentrieren sich damit auf Integrity bzw. die Ebene der QoS-Parameter.

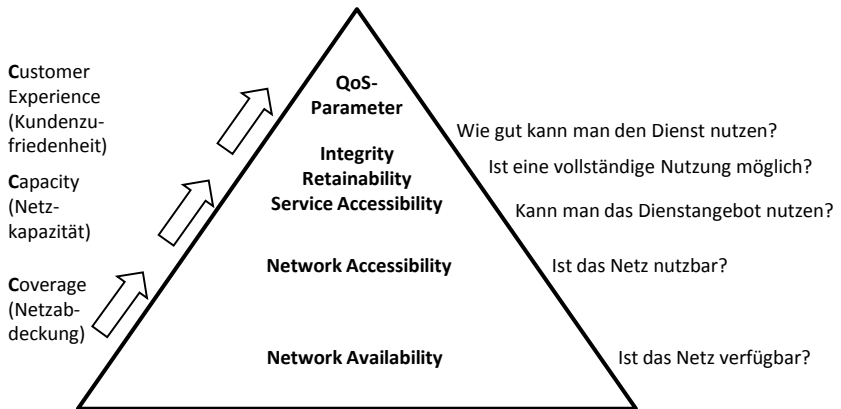


Abbildung 7: KPI-Hierarchieebenen im Mobilfunk

in Anlehnung an [41] S. 43 (Pyramide), erweitert nach [44] S. 29f (4 Ebenen) und ergänzt (3 C)

2.2.4 Die Netzleistungsfähigkeit: Network Performance (NP)

Wie bei Quality of Something diskutiert, wird die QoS teilweise auf die Network Performance reduziert betrachtet. Die ITU-T E.800 definiert die Network Performance wie folgt:

“The ability of a network or network portion to provide the functions related to communications between users.

NOTE 1 – Network performance applies to the network provider’s planning, development, operations and maintenance and is the detailed technical part of QoSO.

NOTE 2 – Network performance parameters are meaningful to network providers and are quantifiable at the part of the network which they apply.” [35] S. 4f

Im ITU-T-Verständnis ist die QoS nicht mit der Netzleistung gleichzusetzen, die Netzleistung aber ein wichtiger technischer Teil in der vom Anbieter geplanten und angebotenen Quality of Service. Ein Beispiel soll dies verdeutlichen:

- Network Performance: Aus der Sicht des Netzes/Netzanbieters ergibt sich, dass 10 Nutzer gleichzeitig je maximal 1 MBit/s Datenrate erhalten können.
- (Network-) QoS: Aus der Sicht des Nutzers bzw. des nutzerseitigen Endpunktes (hier aus Sicht des mobilen Endgerätes) ergibt sich, dass er maximal 1 MBit/s erhalten kann.

Die Network Performance beschreibt dabei die Netzleistung aus der Sicht des Netzes, ggf. auch aus dem Netzzinneren heraus für einzelne Netzsegmente oder Netzelemente, die (Network-) QoS³² hingegen die Ende-zu-Ende-Sicht von den Netzzugangspunkten aus (Unterschied siehe Abbildung 8). Es ist eigentlich besser bei der Network Performance von der Netzleistungsfähigkeit zu sprechen. Beides hängt natürlich zusammen, vor allem in Netzen wie den Mobilfunknetzen, die durch ihre Technologie nicht ausschließlich dedizierte Ressourcen aufweisen, sondern auf dem Prinzip der skalierten und gemeinsam genutzten Ressourcen aufbauen (zelluläre und segmentierte Netze sowie Funk als gemeinsames Medium). Da in den weiteren Betrachtungen jeweils der einzelne Nutzer im Fokus des Interesses liegt, sind diese Zusammenhänge außerhalb der weiteren Diskussion angesiedelt. Die Unterscheidung und der konzeptuelle Zusammenhang von NP über QoS zur QoE, der sich daraus ergibt, werden am Ende des Kapitels (in 2.5.3) noch einmal aufgegriffen.

2.2.5 Das Netz in einer Ende-zu-Ende-Sicht

Rein formal sind Kommunikationsnetze Graphen mit Knoten, den Netzelementen (Router, Switches, Gateways, ...) und Kanten, den Verbindungen dazwischen (Links). Diese dienen dazu die Daten zwischen den Endpunkten (Endgeräten) einer Kommunikation weiterzuleiten.

³² in Anlehnung an die Unterscheidung zu Network Performance und N-QoS siehe auch [45] S. 6f und [45] S. 14f.

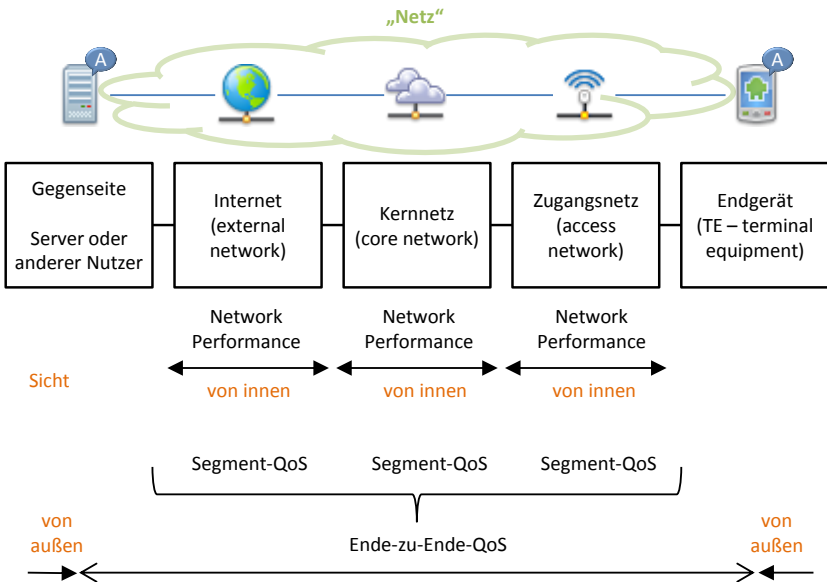


Abbildung 8: Schematische Darstellung des Netzes und unterschiedliche Sichten auf „das Netz“

Ende-zu-Ende-QoS durch verschiedene Netzsegmente, zusammengestellt nach [35] S. 1 und [46] S. 8

Typischerweise wird die lokale Seite (TE) links dargestellt, zur Beibehaltung der Konsistenz innerhalb der Arbeit, mit dem Nutzer in der Darstellung oben oder auf der rechten Seite, wird davon bewusst abgewichen.

Im Weiteren wird zunächst allgemein von IP-fähigen Netzen ausgegangen, noch ohne Einengung auf eine bestimmte Technologie, wenngleich wie angedeutet drahtlose Netze für drahtlose Endgeräte noch detaillierter diskutiert werden. Konkrete Ausprägungen möglicher Netzsituationen im Mobilfunk werden im Abschnitt 5.1 zusammen mit der Einbeziehung des Netzes in die empirischen Untersuchungen diskutiert.

Eine detaillierte Sicht auf die Netzbestandteile ist zunächst nicht notwendig, nur die schematische Sicht der wichtigsten Netzsegmente, siehe Abbildung 8. Aus einer übergeordneten Sicht ergibt sich die skizzierte Dreiteilung in Zugangnetz, Kernnetz eines Providers und dem aus Sicht eines Providers externen Netz, das in der Gesamtheit das Internet als Netz zwischen den Netzen spannt. Jedes der Netzsegmente kann eine andere innere Struktur, Topologie,

technische Realisierungen, Netztechnologien und Medien (Funk, Glasfaser, Kupferkabel) etc. beinhalten. Jedes Teilnetz kann entsprechend eine unterschiedliche Network Performance aufweisen. Zur Vereinfachung wird in der weiteren Darstellung und Diskussion das Netz nur einseitig betrachtet, wäre aber vom (zentralen) Internet aus nach beiden Seiten der Darstellung als vergleichbar strukturiert anzusetzen.

Aus der Gesamtsicht auf das Netz, von einem Endpunkt zum anderen Endpunkt der Kommunikation, ergibt sich die Ende-zu-Ende-Sicht auf das Netz. Als Konsequenz dieser Ende-zu-Ende-Sicht sind die Endpunkte der Betrachtung, also mobiles Endgerät und Gegenseite/Server, auch implizite Teile des Netzes. Wird abstrakt von „dem Netz“ oder „dem Netzwerk“ gesprochen, soll das Zusammenwirken der einzelnen Netzbestandteile, d. h. der beteiligten Segmente und Elemente, als Gesamtsystem zur Ermöglichung der Kommunikation bzw. Datenübertragung verstanden werden. Aus Sicht der Endpunkte sind Interna des Netzes dabei nicht sichtbar. Für eine echte Ende-zu-Ende-Betrachtung definiert das Zusammenwirken in dieser Ende-zu-Ende-Sicht, was als Ende-zu-Ende-Leistung prinzipiell erreichbar ist, bzw. wie die Gesamtcharakteristik als Beschaffenheit des Netzes Ende-zu-Ende zu beschreiben ist. Die zugrundeliegenden Netze können in ihrer Leistungsfähigkeit asymmetrisch sein. Entsprechend kann auch die Ende-zu-Ende-Sicht asymmetrisch ausfallen.

Die Betrachtung der Ende-zu-Ende-Leistung des Netzes führt zur Ende-zu-Ende QoS (E2E QoS), die nachfolgend als netzwerkabhängige QoS bzw. Network-QoS (N-QoS, siehe [45] S. 14f.) auch als solche aus Sicht der Endpunkte verstanden sein soll. Diese netzwerkabhängige QoS ist der Anfangspunkt der Betrachtungen für die spätere Modellierung des interessierenden Zusammenhangs. Weitere Arten von QoS bzw. andere QoS-abhängige Größen sind damit ebenfalls automatisch netzwerkabhängig. Im Weiteren wird mit QoS implizit die netzwerkabhängige QoS in einer Ende-zu-Ende-Sicht gemeint, sofern nichts anderes spezifiziert wird oder zwischen mehreren Formen von QoS zu unterscheiden ist.

2.2.6 Netzparameter zur Beschreibung der netzwerkabhängigen Quality of Service

Die Netzparameter sind technische Merkmale des Netzes, die die technische Netzleistung charakterisieren. Mit den Netzparametern kann auch die netzwerkabhängige QoS als Netzwerkvorgaben oder in Form von Anwendungsanforderungen beschrieben werden. Die QoS-Parameter bilden damit die QoS-Syntax (Begriff nach [45] S. 7) als benennbare und messbare Größen. Prinzipiell werden zumindest leistungsbezogene und funktionalitätsbezogene Beschreibungen unterschieden. Die leistungsbezogenen Aspekte lassen sich meist in Form quantitativer Parameter als Kombination von Wert und Einheit beschreiben, die funktionalen als qualitative Parameter. Die quantitativen Parameter beziehen sich dabei typischerweise auf ein Wertekontinuum, die qualitativen auf binäre Aussagen (z. B. ja/nein, vorhanden/nicht vorhanden) bzw. auf kategoriale Werte (vgl. [45] S. 17).

In den Standardisierungsorganisationen wie ITU-T und ETSI werden die zuvor genannten Parameterkategorien teilweise erweitert und verfeinert. Die ITU-T-Empfehlung I.350 (siehe [47]) unterscheidet demnach ganz allgemein für Netze zur Datenkommunikation in folgende drei Kategorien für zugeordnete Parameter (nach [45] S. 48):

- Geschwindigkeit
- Genauigkeit
- Zuverlässigkeit

Die Zuordnung der Parameter erfolgt dabei in jeweils in einer der drei Phasen einer (Netz-) Kommunikation, bzw. den zu geordneten Funktionen (nach [45] S. 48):

1. Aufbau: notwendige Schritte zum Etablieren einer Datenkommunikation
2. Nutzdatenübertragung: die eigentliche Übertragung der Nutzdaten über das Datennetz
3. Abbau: notwendige Schritte zum ordnungsgemäßen Beenden einer Datenkommunikation

So ergeben sich neun Kombinationen (3 x 3) aus Kategorie und Phase bzw. Funktion. Meist beschränken sich die Betrachtungen in ihrer expliziten Sicht auf die Phase der eigentlichen Nutzdatenübertragung, beziehen aber jeweils

eine Phase vor und nach der eigentlichen Nutzdatenübertragung implizit zumindest mit ein. Im Weiteren ist so die Konzentration auf die Phase der Nutzdatenübertragung aus Sicht von oberen QoS-Schichten auch ausreichend, unter der Annahme, dass die tieferen Schichten die Phasen und Funktionen zum Aufbau und Abbau einer Kommunikationsbeziehung („Verbindung“) bereitstellen und möglichst transparent übernehmen.

Zur Klassifizierung der netzbezogenen QoS-Parameter stellt [45] S. 27 verschiedene Modelle vor. Dabei lassen sich mehrere relevante leistungsbezogene Kategorien identifizieren, die sich jeweils auf unterschiedliche Grundgrößen beziehen:

- Zeiten:
 - Delay (Verzögerung in der Datenübertragung zwischen den Endpunkten in je eine Richtung, u. a. Zeiten durch Warten, zur Zwischenspeicherung und Verarbeitung sowie zur Überbrückung der Wegstrecken bei endlicher Geschwindigkeit)
 - Latenz³³ (Round Trip Time, Umlaufzeit mit Berücksichtigung des Delays beider Richtungen, d. h. von einem Endpunkt zum anderen Endpunkt und wieder zurück)
 - Jitter als Schwankungen der Verzögerungszeit bzw. Latenz
 - Synchronität (relevant, wenn beispielsweise Audio- und Videodatenströme getrennt voneinander übertragen werden)
- Kapazität: als Fähigkeit eine bestimmte Datenmenge pro Zeiteinheit zu transportieren
 - je Übertragungsrichtung
 - Capacity, Speed, Throughput (Durchsatz) → Datenraten
- Verfügbarkeit (Availability): Konnektivität
- Verlässlichkeit (Reliability):
 - Datenverlust (Packet Loss)
 - Datenverfälschung (Bit Error Rate)
 - Reihenfolgentreue

³³ Leider ist die Latenz teilweise als One Way Latency oder als Two Way Latency üblich. In der eigenen Verwendung ist die Latenz wie angegeben immer die Zweibege-Zeit (als Summe), Delay hingegen nur die Einweg-Zeit, wobei die Delay-Werte je Richtung unterschiedlich ausfallen können.

Weitere Kategorien wie Security oder Simplicity, wie sie die ETSI vorsieht (siehe [48]), bleiben nachfolgend unberücksichtigt. Für die weiteren Betrachtungen wird weiterhin davon ausgegangen, dass die Verfügbarkeit gegeben ist, d. h. Netzkonnektivität vorliegt und Verlässlichkeit durch Mechanismen im Netz (Link-Level) bei Bedarf grundsätzlich gegeben ist.

Es interessieren in den weiteren Betrachtungen vorrangig die Parameter zur Beschreibung des paketbasierten Verkehrs auf Basis des Internet Protokolls auf der Vermittlungsschicht bzw. Internet-Schicht (siehe Abbildung 14 zur Schichtung und zum Stack-Modell). Die ITU gibt Empfehlungen zu den relevanten Parametern der Network Performance und QoS (in ITU-T Y.1540, [49]) inklusive Hinweisen zu Netzverkehrsklassen (in ITU-T Y.1541, [50]) und zu den Ende-zu-Ende-Betrachtungen zum Erreichen der Vorgaben (in ITU-T Y.1542, [51]). Daraus ergeben sich die wichtigsten leistungsbezogenen Parameter der weiteren Ausführungen auf der Internet-Schicht. Die folgenden quantitativen Charakteristiken sollen den Kern der relevanten netzbezogenen Beschaffenheit aus Sicht der Ende-zu-Ende-Dienstnutzung (QoS) beschreiben:

- **Durchsatz:** Datenmenge pro Zeiteinheit nach Richtung
 - als Datenrate (kBit/s) in Downstream-Richtung (von der Gegenseite zum Endgerät)
 - als Datenrate (kBit/s) in Upstream-Richtung (vom Endgerät zur Gegenseite)
- **Latenzzeit:** Ende-zu-Ende-Verzögerung (ms), vom Endgerät zur Gegenseite und wieder zurück (Round Trip Time, RTT) als Summe aus den Werten von IP-Packet Transfer Delay für beide Richtungen
- **Jitter:** Schwankung (ms) der Latenz (als IP-Packet Delay Variation)
- **Paketverlustrate:** als Rate (%) verlorener, beschädigter oder zu spät eintrreffender IP-Pakete

Der Durchsatz soll sich dabei auf die Gesamtheit aller übertragenen Daten beziehen, unabhängig von den Kommunikationsprotokollen höherer Schichten und den dort üblichen Betrachtungen von Flows, wie bei TCP (in Anlehnung an die Festlegungen der ITU-T in [49] S. 21).

In frühen Arbeiten zur QoS der neunziger Jahre (wie in [52]) sind neben den leistungsbezogenen Parametern des Netzes zur Datenübertragung noch weitere Parameterkategorien definiert, z. B. die subjektive Bild- oder Tonqualität, die heute eher zur QoE gezählt werden oder kostenorientierte Parameter im Bereich QoB.

2.2.7 QoS-Spezifikation und Best Effort

Durch Angabe entsprechender Beschränkungen oder Bezugswerte erhalten die Werte der einzelnen QoS-Parameter die gewünschte Aussagekraft/-stärke, z. B. als Maximalwerte, Minimalwerte oder auch als Durchschnittswerte pro Zeiteinheit. Die Parameter können als geplante, angebotene bzw. prinzipiell mögliche Werte angegeben werden. Wenn die QoS-Parameter die QoS-Syntax bilden, dann entspricht die Angabe zur Bedeutung bzw. Stärke der Werte einer QoS-Semantik (vergleiche [45] S. 41 zum Begriff und zu möglichen Wertdefinitionen).

Eine formale Unterscheidung wäre dabei (in Anlehnung an [34] S. 390):

- ohne Garantien, als bestmöglicher Wert (Best Effort), der sich aus der theoretischen Netzleistungsfähigkeit selbst ergibt (netzimmanent)
- ohne Garantien, als vorhersagbarer Wert auf Basis von Schätzungen aus vergangenen Verhalten (historische Werte)
- mit Garantie, als garantierter Wert durch QoS-Mechanismen bzw. Methoden zum Traffic-Management

Die Überlegungen zu den ersten beiden Punkten führen zur Definition und Nachbildung der Netzsituationen in den empirischen Untersuchungen, siehe 5.2. Die eigenen Überlegungen zu den gesuchten Schwellwerten als Werte mit besonderer Bedeutung für den QoS-QoE-Zusammenhang setzen ebenfalls hier an (siehe 8.8.6).

Ohne Garantien durch spezielle QoS-Mechanismen ist in den einzelnen Netzsegmenten und somit auch im Netz in der Ende-zu-Ende-Sicht nur **Best Effort** anzunehmen. Dies bedeutet für die einzelnen Parameter eine Angabe und Abschätzung hin zu theoretisch bestmöglichen Werten, also maximale Werte für den Durchsatz und minimale Werte für Latenz, Jitter und Paketverlustrate. Mangels etablierter und allgemein anerkannter QoS-Mechanismen entspricht

dies der allgemeinen Situation bei der Einbeziehung des „offenen Internets“ als Bestandteil des Netzes. Derartige Wertangaben können somit nur als (unverbindliche) Richtwerte gesehen werden.

2.2.8 Over-Provisioning, Garantien, QoS-Mechanismen und Netzverkehrsklassen

Die bisherigen Überlegungen haben einen Aspekt außer Acht gelassen: Die alleinige Kenntnis von Anforderungen an das Netz bedeutet noch lange nicht ihre Erfüllung. Dies wirft natürlich in Verbindung mit der Aussage zu Best Effort bei der Nutzung über das offene Internet eine Sinnfrage auf: Ist die Angabe von QoS-Parametern zum Beschreiben „des Netzes“ (oder wie dann später der Anforderungen der Dienste an das Netz) überhaupt sinnvoll? Ja, ein üblicher Weg auch ohne dedizierte QoS-Mechanismen auszukommen ist **Over-Provisioning**. Dies bedeutet die Vermeidung von Überlastsituationen durch Überdimensionierung des Netzes [34] S. 116. Im Grunde entspricht dies einer einfachen Form einer statistischen Annäherung zur Lösung, so dass mit dem Vorhalten von genügend Ressourcen die Anforderungen im Mittel erfüllt werden können, auch ohne dass spezielle QoS-Mechanismen eingesetzt werden. Praktisch bedeutet dies den Netzausbau, also die Aufrüstung des Netzes mit der Technologie, die prinzipiell in der Lage ist, die gestellten Anforderungen überhaupt erfüllen zu können. Hinzu kommt die Netzverdichtung, also die Bereitstellung von Kapazität, so dass in der Mehrzahl der betrachteten Fälle die Anforderungen auch tatsächlich erfüllt werden können. Werden solche Annahmen getroffen, dann ist nicht mehr ein einzelner Nutzer isoliert zu betrachten, sondern es wird eine statistische Betrachtung über eine Vielzahl gleichzeitiger Nutzer angesetzt. Diese Überlegungen führen zur Kapazitätsplanung als Teil des Netzmanagements und damit über den Rahmen der Arbeit hinaus.

Der aufmerksame Leser wird erkennen, dass mit Over-Provisioning gewissermaßen der Grundstein für den Innovationszyklus von Netz und Nutzung aus der Einleitung gelegt ist. Vereinfacht ausgedrückt lässt sich zur anfänglichen Sinnfrage damit resümieren: Auch ohne QoS-Mechanismen, nur mit Best Effort und Over-Provisioning, ist die Beschreibung und Kenntnis der Einschränkungen des Netzes und der Anforderung der Dienste trotzdem sinnvoll.

Ohne QoS-Mechanismen verlagern sich die Überlegungen zu Verbesserungen der nutzerrelevanten Qualität weg vom Netz zu höheren Protokollschichten und algorithmischen Mechanismen in den endseitigen Applikationen. Dieser Punkt wird nachfolgend in der Diskussion höhere Schichten bis hin zur Applikationsschicht nochmals aufgegriffen (wie z. B. beim Adaptive Bitrate Streaming zur Anpassung des Inhalts an das Netz).

Sind Garantien für kritische Parameter notwendig, so bedeutet dies in paketbasierten Netzen auch die Notwendigkeit von **QoS-Mechanismen** zur Erfüllung dieser Garantien. Im Fall der Garantien ist zu unterscheiden zwischen ([45] S. 27):

- statistische (weiche) Garantie: im zeitlichen Mittel
- deterministische (harte) Garantie: zu jedem Zeitpunkt

Experten mit einem Hintergrund in Betrieb und Verwalten von Rechnernetzen dürften bei QoS wohl zuerst an DiffServ³⁴ und IntServ³⁵ als QoS-Mechanismen im Internet gedacht haben (siehe [55] für ausführliche Erläuterungen). Mögliche technische Ansätze für QoS-Mechanismen bzw. allgemeiner zum Traffic-Management durch gezielte Verkehrsformung sind aber deutlich vielfältiger:

- Reservierung, z. B. wie IntServ
- Priorisierung (Class of Services), z. B. DiffServ
- Policing (harte Beschränkungen mit Verwerfen von Paketen)
- Traffic Shaping (Beschränkungen mit zeitlichem Ausgleich)
- Anpassen des Kommunikationspfades (siehe [21] S. 286f), z. B. durch Handover zu anderen Netzen bzw. Netztechnologien
- Anpassung des Datenstroms (siehe [21] S. 287f), z. B. durch Reduktion der Medienqualität oder WAN-Optimierungen (TCP)

Auf diese QoS-Mechanismen kann im Detail nicht weiter eingegangen werden. Eine plausible Annahme ist allerdings auch, dass diese ohnehin nicht flächendeckend und Ende-zu-Ende zur Verfügung stehen. Daher soll der Blick lediglich auf die prinzipiellen Möglichkeiten innerhalb der interessierenden Mobilfunknetze gelenkt werden. Das Konzept der Netzverkehrsklassen als

³⁴ RFC 2475 der IETF, siehe [53] und weitere damit zusammenhängende RFC

³⁵ RFC 2210 der IETF, siehe [54] und weitere damit zusammenhängende RFC

möglicher QoS-Mechanismus soll dazu kurz angerissen werden, um später bei der Betrachtung der Dienstklassen darauf zurückzukommen. Auf technischer Ebene der Übertragung in modernen Mobilfunksystemen ist eine Unterscheidung entsprechender QoS-Klassen in den Komponenten des Mobilfunknetzes durch die Spezifikationen der 3GPP vorgesehen, siehe Tabelle 2. Dabei werden die Klassen schon unterhalb der Vermittlungsschicht bzw. des Internet Protokolls unterschieden und damit außerhalb des Rahmens der weiteren Betrachtungen. Ausführliche Erläuterung dazu sind beispielsweise in [56] S. 300ff zu finden. Auch für die 3G-Nachfolgetechnologien sind äquivalente Netzverkehrsklassen vorgesehen.

Prinzipiell sehen die Verkehrsklassen meist eine graduelle Besserstellung entsprechend eingeordneten Verkehrs gegenüber allgemeinem bzw. nicht klassifiziertem Verkehr vor, der als Best Effort behandelt wird. Diese Ansätze nach „besser als Best Effort“ haben damit meist ein Problem der korrekten Einordnung. Ohne strikte administrative Kontrolle besteht die Gefahr, dass plötzlich jeder Verkehr wichtiger als sämtlicher anderer Verkehr ist. In wieweit die QoS-Netzverkehrsklassen tatsächlich zur Unterscheidung von Verkehrsströmen in produktiven Mobilfunkwirknetzen genutzt werden, konnte nicht ermittelt werden. Doch selbst wenn sie eingesetzt werden, so ist von einer Beschränkung ihrer Anwendung auf Provider-eigene Dienste auszugehen (siehe dazu die spätere Diskussion zu Mangaged Services vs. OTT-Services in 3.3.1).

Tabelle 2: Zusammengefasste Darstellung möglicher QoS-Klassen in 3G-Mobilfunknetzen

QoS-Klasse	Bandbreite	Delay	Jitter/Variation	Loss	Richtung
Conversational	kritisch, ggf. Garantien	kritisch, sollte niedrig sein	kritisch, sollte niedrig sein	kritisch, sollte niedrig sein	bidirektional
Streaming	kritisch, ggf. Garantien, Anforderung kann teils hoch sein	weniger kritisch	weniger kritisch	kritisch, sollte niedrig sein	unidirektional
Interactive	variabel, typisch stoßweise	moderat	moderat	moderat	bidirektional
Background	variabel	tolerant	tolerant	tolerant	egal

zusammengestellt aus [57] S. 386 und S. 401 für 3GPP und 3GPP2

2.3 Quality of Experience und die Nutzer

Die allgemeine Bedeutungsveränderung von einer holistischen zur technischen QoS lieferte letztlich erst den Anlass, dass die Notwendigkeit für die QoE überhaupt erkannt wurde – in klarer Abgrenzung zur technischen QoS [43] S. 15. Einige der Definitionen zur QoS in der Telkowitz bringen zwar teilweise formal Netz, Dienst und Nutzer zusammen, liefern aber keine Anhaltspunkte dazu, was aus Nutzersicht dabei relevant ist, noch wie die Befriedigung der Nutzerbedürfnisse zustande kommt. Wird die QoS durch die beschriebene Bedeutungsveränderung bzw. -verknapfung weitgehend auf die Netzleistung reduziert, muss die Sicht der Nutzer in Form der QoE zur Beantwortung dieser Fragen herangezogen werden. Vor allem Arbeiten der letzten Jahre zur QoE-Forschung zeigen immer mehr in die Richtung einer umfassenden und integrativen Betrachtung des Nutzererlebnisses bei der Nutzung interaktiver Systeme und Dienste, wozu sich Ingenieurwissenschaften und Humanwissenschaften weiter annähern müssen.

Ausgehend von einer Vielfalt an Definitionen soll in Analogie zur kritischen Auseinandersetzung mit dem Begriff der QoS auch die QoE hinterfragt werden. Nicht zuletzt ist QoE vielfach das neue Buzzword, welches QoS einfach ersetzt hat. Danach wird die Verbindung der QoE zum verwandten Begriff der User Experience erörtert. Mit der Diskussion möglicher Einflussfaktoren zur QoE werden auch die menschlichen Faktoren auf ihre Wirkung zur QoE untersucht. Die begriffliche Aufschlüsselung von QoE in Q und E führt zur Diskussion des Nutzererlebnisses (E) und den Qualitätsaspekten mit dem zugehörigen Beurteilungsprozess (Q) inklusive einer Unterscheidung von Zufriedenheitsfaktoren. Der Abschnitt schließt mit der Auseinandersetzung zu Nutzungsprozess und Dynamik der QoE.

2.3.1 Definitionen zur QoE

Gerade die QoE ist ein vielschichtiges Problem, angesiedelt in verschiedenen Wissenschaftsdisziplinen mit unterschiedlichen Ausgangspunkten und Interessenlagen sowie einem ganzen Bündel von relevanten Einflussfaktoren und Maßen. Es ist also nicht so, dass keine Definitionen existieren – im Gegenteil, es gibt eine Vielzahl von Ansätzen zur Beschreibung von QoE, woraus ebenso viele Definitionen bzw. Versuche zur Definition hervorgehen. Entsprechend

fehlt aber eine einzige, umfassend befriedigende und allgemeingültige Definition von QoE. Nachfolgend sollen einige QoE-Definitionen zur Verdeutlichung herangezogen werden. Dabei soll zunächst von den Definitionen ausgegangen werden, die ihre Wurzeln in der Netzwelt haben und somit Anknüpfungspunkte oder gar Schnittmengen mit der QoS aufweisen.

Wie bei QoS existieren auch für QoE zahlreiche Ad-hoc-Erklärungen/Definitionen, z. B.:

“QoE is a subjective metric that quantifies the perceived quality of a service by the viewers.” [58] S. 4

“QoE is how a user perceives the usability of a service when in use – how satisfied he or she is with a service. The term QoE refers to the perception of the user about the quality of a particular service or network.” [59] S. 4

Ad-hoc-Erklärungen wie diese können zwar innerhalb einer Arbeit eine sinnvolle Bedeutungsklärung für den Begriff der QoE liefern, eine richtige Definition wird aber zusätzlich noch allgemein akzeptiert und gewinnt dadurch ihren Wert für ein ganzes Forschungsgebiet.

Die ITU-T-Recommendation P.10/G.100 (Amendment 2), bezogen auf die Telekommunikations-/Telefonwelt, definiert die QoE für die Nutzung von IP-basierten Diensten wie folgt:

“The overall acceptability of an application or service, as perceived subjectively by the end user.

Note 1: Quality of Experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc).

Note 2: Overall acceptability may be influenced by user expectations and context”[60] S. 2

Die Definition liefert damit einige wichtige Anhaltspunkte, die an späterer Stelle noch einmal aufgegriffen und gesondert diskutiert werden (subjektive Wahrnehmung, Erwartung, Kontext, Ende-zu-Ende-Sicht) und nennt mit der

Akzeptanz auch ein mögliches Maß für die QoE zur Bestimmung des Nutzererlebnisses.

Die 3GPP-Definition zur QoE entspricht dem Wortlaut der genannten ITU-T Definition (siehe [61] S. 9). Durch nachfolgende Ausführungen wird die QoE dabei auf die Parameter begrenzt, die der Kontrolle des Netzbetreibers unterliegen, d. h. Aspekte des Endgeräts etc. sind ausgeklammert. Demnach ist je Dienst ein spezifischer Satz Kenngrößen aus Ende-zu-Ende-Messungen für die Beschreibung der QoE geeignet. Die QoE beschreibt somit nur die Leistungsmetriken aus der Nutzersicht. (zusammengefasst nach [61] S. 9)

Die ETSI TR 102274 definiert QoE als Nutzererlebnis, das neben Kommunikationsdiensten auch auf deren Nutzerschnittstellen (z. B. bei mobilen Applikationen) abzielt, womit sich die QoE ergibt als:

*“user’s perceived experience of what is being presented by a communication service or application user interface
NOTE: It takes into account the individual Quality of Services and measures the acceptability of a service or application by including factors such as usability, utility, fidelity and level of support from the application or service provider (e.g. sales, delivery, error corrections).”*
[62] S. 10

Besonders die Auflistung nichttechnischer Angebotsfaktoren erinnert an die holistische QoS-Sicht der ITU-T, wobei hier die individuelle QoS nur Einflussfaktor zur QoE ist. Nicht einmal innerhalb einer Standardisierungsorganisation sind die Definitionen jedoch immer deckungsgleich bzw. ist auch ein Wandel zu beobachten. In der ETSI EG 202534 ist die QoE bei ansonsten gleichem Wortlaut (und gleicher Zusatzanmerkung) nicht als Nutzererlebnis, sondern als Nutzerleistung (im Sinne eines Nutzungsergebnisses) definiert, womit sich die QoE ergibt als:

“The performance of a user when using what is presented by a communication service or application user interface.” [63] S. 9 ergänzt um die Notiz aus der o. g. ETSI TR 102274

Die ETSI TR 102643 liefert eine leicht abweichende Definition zur QoE, ebenfalls auf Basis der Nutzerleistung (Nutzungsergebnis), die vor allem ausgehend vom angegebenen Beispiel in die angestrebte Richtung zur Herstellung des gesuchten Zusammenhangs zwischen QoS und QoE geht und die Brücke zu psychologischen Messungen, also Messungen am Menschen, schlägt. Demnach ist die QoE:

“measure of user performance based on both objective and subjective psychological measures of using an ICT service or product

NOTE 1: It takes into account technical parameters (e.g. QoS) and usage context variables (e.g. communication task) and measures both the process and outcomes of communication (e.g. user effectiveness, efficiency, satisfaction and enjoyment).

NOTE 2: The appropriate psychological measures will be dependent on the communication context. Objective psychological measures do not rely on the opinion of the user (e.g. task completion time measured in seconds, task accuracy measured in number of errors). Subjective psychological measures are based on the opinion of the user (e.g. perceived quality of medium, satisfaction with a service).

EXAMPLE: A service provider may conclude that a service with a certain level of QoS used for a particular communication situation offers users excellent QoE, whilst with a different level of QoS provides poor QoE.” [2, S. 64] S. 9f

Die Definition selbst ist recht kurz, weist aber mit objektiven und subjektiven Maßen zur QoE-Bestimmung den Weg zur empirischen Überprüfung. Richtig interessant wird die Definition erst durch die Zusatzanmerkungen mit Nennung weiterer Einflüsse wie dem Kontext und einigen konkreten Maßen wie der Zufriedenheit. Das formulierte Beispiel geht exakt in die angestrebte Richtung zur Herstellung eines Zusammenhangs zwischen variabler QoS und QoE, wie er im Weiteren untersucht werden soll.

Das DSL-Forum definiert nur die direkt messbaren (technischen) Größen zur Beschreibung der Leistung eines Systems als QoE, wenngleich bezogen auf die Perspektive eines Nutzers:

“Quality of Experience (QoE): is the overall performance of a system from the point of view of the users. QoE is a measure of end-to-end performance at the services level from the user perspective and an indication of how well the system meets the user’s needs.” [40] S. 7

Diese Definition kontrastiert damit die vorherigen Definitionen deutlich. Zwar kommt jeweils der Nutzer darin vor, zum einen aber als Quelle der QoE (Nutzererlebnis bzw. -ergebnis) und zum anderen als reiner Bezugspunkt für die Leistungscharakteristik des genutzten technischen Systems.

In Ergänzung der zwangsläufig technischen Ausgangslage der Definitionen mit Wurzeln in der Netzwelt (Netz/Dienst) könnten auch Definitionen zum Begriff der QoE von außerhalb ergänzt werden. Dabei wird aber zwangsläufig der Anwendungsrahmen (Nutzung multimedialer Dienste über Netzwerke) verlassen, was entsprechend zu beachten ist. Eine solche Verbindung zur Nutzerwelt/HCI-Welt wird mit dem Konzept der User Experience nachfolgend noch separat diskutiert, wobei auch der Begriff der Quality of Experience auftaucht. Aber auch ohne den genannten Anwendungsrahmen zu verlassen, sollte klargestellt sein, dass die QoE-Definitionen schon so sehr weit auseinandergehen. Dies wurde auch in der Forschergemeinschaft erkannt, was schließlich zur Arbeit an einer neuen Definition von QoE führte.

Im Rahmen des Dagstuhl Seminars 09192³⁶ wurde 2009 bei einem Treffen von QoE-Experten eine Definition zur QoE erarbeitet, die die Nutzersicht und Einflussfaktoren darauf stärker betonen sollte. Demnach ist die QoE:

“Degree of delight of the user of a service. In the context of communication services, it is influenced by content, network, device, application, user expectations and goals, and context of use.” [30] S. 8

³⁶ <http://www.dagstuhl.de/de/programm/kalender/semhp/?semnr=09192> abgerufen am 28.02.2014; In den zum Seminar direkt veröffentlichten Materialien ist die Definition nicht zu finden, aber in nachfolgenden Arbeiten mehrerer Teilnehmer, z. B. der angegebenen Quelle.

Ein Qualinet³⁷-White-Paper entwickelte den Entwurf der Dagstuhl-Definition weiter und fasste 2013 das Ringen der wissenschaftlichen Gemeinschaft zur Definition von QoE zusammen. Dabei entstand durch zahlreiche beteiligte Experten aus verschiedenen Fachbereichen eine weiterentwickelte (Arbeits-)Version:

“QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user’s personality and current state.” [43] S. 6

Diese Definition zur QoE spiegelt den aktuellen Forschungsstand zur QoE in einer umfassenden interdisziplinären Betrachtungsweise recht gut wider. Es wird klar, worum es geht, dass technische und im Nutzer selbst begründete Einflüsse wichtig sind, dass QoE eine Qualität im Sinne eines graduellen Vergleichs zwischen Erlebnis und Erwartung ist und wie sich die zugehörige Bewertung der nutzerrelevanten Größen ergibt. Dabei wird bewusst von positiven (Verzückung) wie auch negativen Empfindungen (Verdross) ausgegangen, die mit der Beurteilung zusammenhängen. Eine solche Fokussierung auf das Nutzererlebnis in einer holistischen Nutzersicht, die weit über rein technisch fassbare Aspekte hinausgeht, kann als State-of-the-Art zum Verständnis des Konzepts der QoE angesehen werden. Ausgehend von der genannten Qualinet-Arbeitsversion offeriert [31] eine verallgemeinerte Definition von QoE, die prinzipiell auch über den Anwendungsbereich der Nutzung multimedialer Dienste über Netzwerke hinaus anwendbar ist [31] S. 19:

“QoE is the degree of delight or annoyance of a person whose experiencing involves an application, service, or system. It results from the person’s evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person’s context, personality and current state.” [31] S. 19

³⁷ ein europäisches Forschungsprojekt, siehe <http://www.qualinet.eu> abgerufen am 28.02.2014

2.3.2 Quality of Everything: Die QoE zwischen QoS und Qo*

Ähnlich wie bei QoS die kritische Auseinandersetzung mit der Auslegung des Begriffs zu *Quality of Something* führte, ist auch für QoE eine derartige Diskussion notwendig.

Verwirrende Vielfalt von Konzepten und Definitionen: Weitere Definitionen zur QoE listet [64] S. 56 auf und identifizierte dabei bereits schon sechs unterschiedliche Ansätze zur Definition von QoE. In Überarbeitung dieser ursprünglichen Auflistung dort ergibt sich QoE als:

- Erweiterung der QoS (um Nutzeraspekte)
- Leistungsmetriken aus Sicht der Nutzer
- subjektives menschliches Erleben als Erlebnis
- objektives Ergebnis einer Nutzung / Nutzerleistung unter Berücksichtigung menschlicher Faktoren
- Grad der Verzückung bzw. des Verdrusses eines Nutzers / einer Person in Bezug auf ein Produkt (im weitesten Sinn)
- dazugehöriges Qualitätsurteil
- eine Form der Kundenzufriedenheit (aus Sicht des Marktes)
- eine Form von Usability bzw. User Experience

Allein schon diese Vielfalt rechtfertigt eine alternative Erklärung des Akronymes QoE als *Quality of Everything*. Trotzdem sollen noch einige weitere Punkte besonders diskutiert werden.

Alter Wein in neuen Schläuchen: Ein recht zutreffendes aber auch etwas ernüchterndes Zitat dazu, das schon die relevanten Punkte aufgreift:

“QoE has somehow become what QoS was supposed to be once up on a time when it was defined by ITU-T: a set of parameters reflecting user satisfaction with a service.” [65] S. 6

Die ITU-T-Sicht auf QoS ist, wie dargelegt, sehr weitreichend. Zusammen mit der bei QoS festgestellten Bedeutungsveränderung, hin zu einer Beschränkung auf die technischen Parameter des Netzes, verlockt natürlich das ursprüngliche Konzept einer holistischen QoS mit dem Begriff der QoE zu besetzen, wie beispielsweise in der (veralteten) QoE-Sicht des TeleManagement Forums

(vergleiche [66] S. 1 wobei der Nutzer vorrangig als Kunde gesehen wird). So besteht allerdings das Risiko, dass QoS und QoE als zwei Begriffe für ein Konzept gebraucht werden, obwohl die Trennung der Konzepte und gleichzeitige Klarstellung des konzeptuellen Zusammenhangs eigentlich besser sind.

Rein technisch erfasste QoE: In einigen Definitionen und Konzepten ist die Gleichsetzung der QoE mit der Beschreibung der nutzerrelevanten Größen zu sehen (wie in der Definition des DSL-Forums). Im Grunde sind dies aber ebenfalls technische Werte wie eine QoS-Charakteristik auch, wenngleich näher an einer Nutzersicht. Konzepte auf einer solchen Basis rein technischer Definitionen werden einer ganzheitlichen und auf den Nutzer ausgerichteten Sicht nicht gerecht und müssen entsprechend als unvollständig angesehen werden. Sie beschreiben unter Umständen zwar direkt QoE-relevante Einflussgrößen, nicht aber die QoE selbst. Mit Blick auf die Schichtung der Qualitäten (siehe 2.5.1) wäre eine solche Charakteristik nur eine Applikations-spezifische QoS bzw. Systemleistung aus Nutzersicht. Die diskutierten Definitionen der ETSI zur QoE mit den genannten Maßen wie Akzeptanz, Nutzen und Zufriedenheit helfen aber bei der Unterscheidung. Das Konzept zur Beschreibung der direkt nutzerrelevanten Einflussgrößen als Nutzerstimuli wird an späterer Stelle in der Arbeit noch einmal aufgegriffen, um daraus die besondere Rolle der Dienste zu begründen.

Qo*-Inflation: Wie bei jedem populären Konzept besteht auch für QoE die Gefahr, dass das Konzept überdehnt wird und plötzlich eine Vielzahl ähnlicher Begriffe und Konzepte auftauchen. Aus der inflationären Nutzung von QoX/Qo*³⁸, mit einem Platzhalter für quasi jeden anderen Begriff, ergeben sich besonders im Bereich von QoE weitere Begriffe wie Quality of Interaction, Quality of Feedback, Quality of Learning, etc. (z. B. in [58]), die irgendwie alle mit QoE zusammenhängen, aber entweder auf spezielle Anwendungsgebiete beschränkt sind, z. B. das E-Learning, oder wiederum selbst die Notwendigkeit zur Abgrenzung von der QoE aufwerfen. [67] S. 2 kritisiert dazu, dass so das eigentliche Anliegen von QoE unscharf wird. Wird QoE zu

³⁸ In der eigenen Arbeit wird QoX mit dem X als Platzhalter noch einmal aufgegriffen, allerdings nicht als kritisierendes universelles Begriffskonstrukt, sondern für die wohldefinierte Ursache-Wirkungskette im eigenen Modell, von QoS zu QoE. Daher ist an dieser Stelle Qo* als Entsprechung zu QoX der genannten Quelle zu verstehen.

einem Buzzword unter vielen, so besteht schnell die Gefahr, dass schon bald der nächste Begriff mit gleichem Anliegen zum neuen Hype wird.

Nach der kritischen Interpretation von QoS als *Quality of Something* ist die Deutung von QoE in einer ebenso kritischen Auseinandersetzung hinsichtlich begrifflicher Unschärfe und Überblähung als *Quality of Everything*, also der Qualität von Allem (irgendwie), damit eigentlich nur konsequent.

2.3.3 Quality of Experience und der Zusammenhang zur User Experience

Aspekte zur QoE haben nicht zwangsläufig ihre Wurzeln innerhalb der Netzwelten. Deshalb kann es interessant und lohnend sein, auch außerhalb der Netzwelten zu schauen, ob Quality als Experience als Begriff auftaucht und was in anderen Welten damit assoziiert ist. In der Recherche stößt man dabei quasi zwangsläufig auf einen weiteren verwandten Begriff, die User Experience (UX).

Dem Bereich der HCI-Forschung entstammt die folgende recht frühe Überlegung (1996) zum Erlebnis der Nutzer bei der Nutzung interaktiver Systeme:

“By ‘experience’ we mean all the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they’re using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it. If these experiences are successful and engaging, then they are valuable to users [...] We call this ‘quality of experience’.”
[68] S. 12

In der HCI-Welt etablierte sich nachfolgend aber nicht Quality of Experience als Begriff, sondern User Experience (UX) für das skizzierte Konzept. In der vergleichenden Betrachtung von QoE und UX kommt [69] S. 37 zu dem Schluss, dass QoE und User Experience das gleiche zugrundeliegende Konzept teilen. In beiden Konzepten steht das Nutzererlebnis im Vordergrund der Betrachtungen, lediglich der konzeptuelle Ursprung und die Verankerung in der Forschergemeinschaft unterscheiden sich demnach. QoE und UX sind somit wie zwei Seiten einer Medaille, der gleiche Kern, aber eben nicht exakt

gleich. Die Seite der QoE ist dabei von der Netzwelt bzw. genauer der Telewelt geprägt und hat sich aus Überlegungen zur QoS entwickelt. Die UX-Seite wiederum wird durch Ansätze der HCI-Welt geprägt und stellt letztlich eine Weiterentwicklung der älteren Ideen zur Usability dar. (zusammengefasst nach [69] S. 37)

Wie auch bei der QoE kann man für die UX eine begriffliche Annäherung über eine Definition versuchen, kommt dabei aber zu einem ähnlichen Ergebnis, nämlich einer Vielzahl von Definitionen³⁹ mit jeweils leichten Unterschieden. Wie auch bei QoS und QoE bieten Standards und Spezifikationen auch für die UX zumindest einen Ansatzpunkt für einen kleinsten gemeinsamen Nenner. So definiert die DIN EN ISO 9241-210 die User Experience als Benutzererlebnis wie folgt:

„Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produkts, eines Systems oder einer Dienstleistung resultieren“ [70] S. 7

Zwischen QoE und UX sind mindestens folgende Anknüpfungspunkte auszumachen, die innerhalb der vorliegenden Arbeit in einer gemeinsamen Diskussion nachfolgend noch weiter ausgearbeitet werden:

- Einflussfaktoren (insbesondere menschliche Faktoren)
- Erlebnis und Qualitätsaspekte (pragmatisch vs. hedonistisch)
- Nutzungsprozess und Dynamik
- Kontext mit Nutzungssituation und Nutzungszielen

2.3.4 QoE-Dimensionen/Einflussfaktoren und -größen

Die ITU-T-Recommendation G.1080 unterscheidet bei den Einflussfaktoren zur QoE lediglich in objektive Einflussfaktoren, die durch die Quality of Service gebildet werden, und subjektive Einflussfaktoren, welche durch die menschlichen Komponenten beschrieben werden [71] S. 5, siehe Abbildung 9a. Dabei kann jede der beiden Seiten weiter in einzelne Faktoren aufgeschlüsselt werden. Diese Sicht greift etwas zu kurz, da die situativen Aspekte nicht explizit bzw. umfassend berücksichtigt werden. Im QoE-Verständnis von

³⁹ Eine Sammlung von weit mehr als 20 Definitionen zur User Experience findet sich unter <http://www.allaboutux.org/ux-definitions> (abgerufen am 12.10.2014).

Qualinet werden daher System (Technik) und Mensch explizit um den Kontext erweitert [43] S. 11f. Diese Aufteilung entspricht auch der Sicht auf die Einflussfaktoren der User Experience in der ISO 9241-210, womit klar wird, dass auch die Experience vergleichbar definiert ist (siehe [67] S. 4). Es ist von Überlappung zwischen den jeweiligen Kategorien auszugehen [43] S. 11f. Es wird teilweise Einflüsse auf die QoE geben, die nicht klar nur in eine der Kategorie einzuteilen sind, siehe Abbildung 9b. Marktbezogene Faktoren könnten eine eigene Kategorie bilden, die Betrachtung innerhalb des Kontextes als eine besondere Form der kontextuellen Einflussfaktoren ist aber äquivalent, zumal diese Aspekte ohnehin nicht im Fokus der eigenen Überlegungen stehen.

Die jeweiligen Kategorien System, Mensch und Kontext lassen sich weiter in verschiedene Typen von Einflussfaktoren zerlegen, wie in der Abbildung angedeutet. Weitere Erläuterungen zur Qualinet-Sicht dazu sind in [72] S. 57ff kompakt aufbereitet. Die Erläuterung zum Menschen als Nutzer und die damit verbundenen menschlichen Faktoren werden unmittelbar im Anschluss diskutiert, zum Kontext und der Technik des Systems erfolgt die Diskussion jeweils separat. Die eigenen Ausführungen unterscheiden sich dabei in der Aufteilung der einzelnen Elemente und dem Detailgrad, sind aber prinzipiell State-of-the-Art. Die Rolle der Dienste wird in der eigenen Diskussion stärker ausgearbeitet. Hinzu kommen in der eigenen Betrachtung nachfolgende Überlegungen, zu weiteren möglichen Unterscheidungspunkten, die sich auch später in der Modellierung des QoS-QoE-Zusammenhangs wiederfinden:

- nach grundsätzlicher Art und Herkunft: Mensch, System, Kontext (wie zuvor angesprochen)
- bei technischen Faktoren nach ihrer Ursache
 - direkter vs. indirekter Einfluss auf die QoE
 - QoS-abhängig vs. QoS-unabhängig
- nach ihrer Einflusseite bei der Beurteilung: Erlebnisfaktoren vs. Erwartungsfaktoren
- Wirkung aus Sicht der Zufriedenheit: Hygienefaktor, Leistungsfaktor, Begeisterungsfaktor

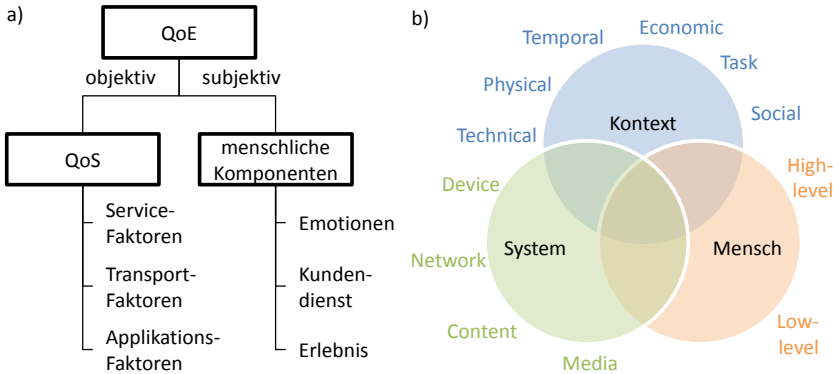


Abbildung 9: Kategorien von QoE-Einflussfaktoren

a) zweiseitige Sicht der ITU-T G.1080 weiter aufgeschlüsselt in einzelne Faktoren nach [71] S. 5 (übersetzt)

b) dreiteilige Qualinet-QoE-Sicht in Anlehnung an [72] S. 57 mit Bezug auf die Einflussfaktoren der User Experience in der ISO 9241-210, um verschiedene Typen je Kategorie ergänzt (nach [72] S. 57ff)

2.3.5 Überblick zu den menschlichen Faktoren

Menschliche Faktoren sind wichtige Einflussfaktoren zur QoE. Im Grunde genommen steht der Mensch als Nutzer und somit das menschliche Erlebnis sogar im Mittelpunkt des Interesses der Betrachtungen zur QoE. Die Beschäftigung mit dem „weichen“ Faktor Mensch ist eigentlich die Spielwiese der Kognitionswissenschaftler und Psychologen, weniger die der Techniker oder Ingenieure. Entsprechend dieser Einschränkung mögen die folgenden Ausführungen verstanden werden: von einem Techniker für Techniker, keinesfalls umfassend und vollständig, sondern stark vereinfacht (soweit der Techniker dies einschätzen kann) aber gut genug, um die Prinzipien zu verstehen und weiteren Anknüpfungspunkte für die QoS-QoE-Gesamtsicht herzustellen.

Als etwas ernüchternder Hinweis vorab: [72] S. 57 weist darauf hin, dass bedingt durch die Komplexität und einen Mangel an empirischen Ergebnissen bei genauer Betrachtung oftmals nur schlecht verstanden ist, wie die menschlichen Faktoren auf die QoE einwirken. Der Mensch als komplexes System ist in gewisser Art somit selbst weitgehend eine Black Box. Einige Zerlegungen erscheinen dazu trotzdem sinnvoll und plausibel, um die Komplexität aufzubrechen und letztlich dadurch die weiteren Betrachtungen so weit zu vereinfachen.

chen, wie es für das Verständnis der menschlichen Faktoren zur QoE notwendig ist. Dafür lassen sich in der Literatur verschiedene Teilmodelle ermitteln. Einige der grundlegenden Konzepte dazu sollen kurz vorgestellt werden.

Für eine grundlegende Erklärung der menschlichen Faktoren zur QoE aus Sicht der Informatik kann auf Erkenntnisse aus der HCI-Forschung zurückgegriffen werden. Zur Physiologie und Psychologie der menschlichen Informationsverarbeitung in der Mensch-Maschine-Interaktion vermitteln beispielsweise [73] S. 45ff und S. 63ff einen kompakten Einstieg. Wie in jeder Wissenschaftsdisziplin wird mit dem verglichen, was bekannt ist. [73] S. 47f zieht entsprechend Parallelen zwischen der menschlichen Informationsverarbeitung und der Datenverarbeitung in Computersystemen, worauf sich die nachfolgenden Erläuterungen beziehen. Demnach ist eine grobe Unterscheidung in Wahrnehmung (Input), interne Verarbeitung (Prozesse) sowie Motorik (Output) möglich. Einen Teil der internen Verarbeitung bilden ein (Langzeit-) Speicher (Gedächtnis) und ein innerer Zustand (Hunger, Durst, Unlust, Spannung, ...). Entsprechend des veränderlichen momentanen inneren Zustands kann die Verarbeitung unveränderter äußerer Informationen bei ein und derselben Person zu unterschiedlichen Ergebnissen führen. Unterschiedliche Personen wiederum können vielfältige Vorerfahrungen, Erinnerungen oder Anspruchsniveaus etc. mitbringen.

Wesentliche Teile der menschlichen Informationsverarbeitung laufen unterhalb der Bewusstseinsschwelle ab (Erläuterungen nach [73] S. 47f mit Verweis auf das Rasmussen-Modell [74] zur menschlichen Kognition). So sind beispielsweise die Wahrnehmung äußerer Reize, der Abgleich mit einem dynamischen Modell der Außenwelt und die motorische Koordination unbewusst. Werden in der unbewussten Verarbeitung etwaige Unstimmigkeiten (Inkongruenzen) entdeckt, wird die bewusste Verarbeitungsebene eingeschaltet. Auf der Ebene der bewussten Verarbeitung wird die Wahrnehmung durch selektive Aufmerksamkeit hinsichtlich der angestrebten Ziele gesteuert und Handlungen werden geplant.

Das Konzept zur Unterteilung in zwei Ebene der Verarbeitung kann auch für QoE-Betrachtungen eine Rolle spielen (vgl. [75] S. 25):

- kognitiv-rational: analytisch, logisch, abstrakt, aktive Kontrolle, regelbasiert, langsam
- emotional-intuitiv: kontextspezifisch, automatisch, vorstellungsverknüpft, anstrengungslos, emotionsgetrieben

Das emotional-intuitive System gibt die Präferenzen und Bewertungen vor, solange das kognitiv-rationale System nicht interveniert. Dabei spielt der Kontext eine wichtige Rolle.

[72] S. 57f unterscheidet zwischen veränderlichen (variant) und unveränderlichen (invariant) menschlichen Einflussfaktoren zur QoE. Dabei wird die Informationsverarbeitung auf zwei Ebenen mit den Kategorien Lower-Level und Higher-Level aufgeteilt ([72] S. 58ff mit Verweis auf zahlreiche weiterführende Quellen):

- Lower-Level: physische, emotionale und mentale Verfassung (Wahrnehmung, Emotionen, Stimmung, Aufmerksamkeitsniveau)
- Higher-Level: Verstehen der Stimuli, Interpretieren und Bewerten

Bei der Informationsverarbeitung auf höherer Ebene werden Vorwissen und Vorerfahrungen mit einbezogen, die als relativ stabile Merkmale wirken [72] S. 58. An dieser Stelle knüpfen die Diskussionen zur Erwartung im Beurteilungsprozess, zum soziokulturellen Kontext der Nutzer und zur Zahlungsbereitschaft (Marktaspekte) an. Ebenfalls auf höherer kognitiver Ebene sind Aspekte zum (Nutzungs-) Ziel, inneren (Wert-) Einstellungen und zur Motivation einzuordnen [72] S. 59. Es sind unterschiedliche Nutzerpersönlichkeiten und Nutzertypen innerhalb einer größeren Nutzergemeinschaft erkennbar. Zur Nutzerpersönlichkeit und damit verbunden zum soziokulturellen Kontext können verschiedene Attribute der Nutzer einen Hinweis geben, z. B. Geschlecht, Alter, Bildung und beruflicher Status. Anmerkungen dazu werden direkt mit den erhobenen Daten aus der empirischen Untersuchung verknüpft erörtert (siehe Kapitel 7). Mit den Nutzertypen wird ähnlich verfahren, die entsprechende Erläuterung erfolgt ebenfalls direkt zusammen mit den gesammelten Daten (siehe Kapitel 7).

Zu den eher schnell veränderlichen Einflussfaktoren können die Stimmung und Emotionen zählen, die sich u. U. während der Nutzung oder von Nutzung zu Nutzung ändern [72] S. 60. Dazu ist die Diskussion zur Dynamik der QoE am Ende dieses Abschnitts der passende Anknüpfungspunkt inklusive Erklärungen zu möglichen Veränderungen der Erwartung.

Zusammengefasst sind relevante menschliche Faktoren:

- innere Verfassung und Gemütszustand, Gefühle, Stimmung
- Aufmerksamkeit: Konzentration vs. Ablenkung
- Intention/Ziel und Motivation/Antrieb
- Vorerfahrung und Vorwissen
- Erwartung (aus zuvor erlebter Leistung, die sich ggf. mit der Zeit verändert)

Für die nachfolgenden Erläuterungen unter Einbeziehung des Menschen soll dieser als Nutzer zur Vereinfachung zerfallen in:

- **Auge**, stellvertretend für die *Wahrnehmung* äußerer Reize (hier von Endgerät und App)
- **Hand**, stellvertretend für die *Interaktion* mit der Außenwelt (hier zu Endgerät und App)
- **Hirn**, stellvertretend die *Kognitionsvorgänge* im Inneren, Gedächtnis, innere Einstellung und Zustand sowie die QoE-Beurteilung selbst

2.3.6 Experience aus Wahrnehmung und Interpretation

Intuitiv wird wohl jedem Leser klar sein, was mit dem Erlebnis der Nutzer im mobilen Internet grundsätzlich gemeint ist, sofern er selbst Nutzer entsprechender Angebote ist. Das E aus QoE, also was unter „Experience“ konkret zu verstehen ist, soll kurz andiskutiert werden, bevor nachfolgend auf den Beurteilungsprozess, als das Q in QoE und mögliche Maße als konkrete Größen der QoE eingegangen wird.

Mit der Wahrnehmung realer physikalischer Größen beschäftigt sich die Psychophysik. Die physikalischen Größen lassen sich wie folgt unterscheiden (nach [30] S. 19): räumlich, zeitlich, eigenschaftlich (nach ihrer Art). Demnach kann eine bestimmte Leistung (Reiz, Stimulus) als physikalisches Ereignis

nis begriffen werden und die Leistungswahrnehmung als Wahrnehmungsergebnis ([30] S. 19). Die Wahrnehmung des Ereignisses ist dabei stets subjektiv. [30] S. 19 weist in diesem Zusammenhang darauf hin, dass subjektiv nicht „ungenau“ etc. bedeuten muss, sondern lediglich, dass der Mensch als Messorgan eingebunden ist. Zu den weiteren Prinzipien der psychophysikalischen Messung, wie der Verknüpfung der verschiedenen Arten von Ereignissen und möglichen Fehlerquellen, gibt [30] S. 23ff Auskunft.

Die menschliche Wahrnehmung vieler Reizgrößen ist nicht linear, sondern z. B. logarithmisch oder exponentiell. Das im 19. Jahrhundert nach seinen Entdeckern Weber und Fechner benannte Weber-Fechner-Gesetz beschreibt einen logarithmischen Zusammenhang zwischen Reizgröße und Reizwahrnehmung. Weitere Verweise und ein mathematisches Beschreibungsmodell zur Anwendung in der mathematischen Psychologie liefert [76]. Neuere Arbeiten im QoE-Umfeld gehen auf Anwendung und Bedeutung des Weber-Fechner-Gesetzes sowie weitere nichtlineare Zusammenhänge ein (siehe [77]), woraus folgt, dass diese Prinzipien der Nichtlinearität auch für die QoE Anwendung finden können. In der empirischen Untersuchung wird bei der Parametrisierung der Stimuli darauf zurückgekommen.

Wahrnehmung und nachgelagerte Interpretation führen dazu, dass eine objektive Bestimmung der Leistung aus Reizgrößen nicht dem subjektiven Erleben bzw. dem Nutzererlebnis gleichzusetzen ist. Es soll davon ausgegangen werden, dass es sinnvoll ist, zwischen objektiven Eigenschaften und im Fall der QoE der subjektiven Wahrnehmung der jeweiligen Merkmale zu differenzieren. Die objektiven Eigenschaften sind direkt messbar und als quantitative Metrik beschreibbar, für die Wahrnehmung lassen sich hingegen nur plausible Annahmen zur qualitativen Wirkung treffen. Anknüpfungspunkte zur Wahrnehmung (Perception) dieser objektiv messbaren Eigenschaften einer Dienstenutzung sind in der weiteren Betrachtung die Nutzerstimuli der Dienste als die relevanten Reizgrößen der wahrgenommenen Leistung (Performance) und ihre Präsentation (Presentation). Diese werden dienstspezifisch im Kapitel 3 zur Nutzersicht auf die Dienste erörtert (siehe 3.5).

Wie zuvor beschrieben, werden Ereignisse der äußeren Welt durch die Sinnesorgane wahrgenommen. Aus dem Strom der Wahrnehmung und Interpretation dieser äußeren Ereignisse entsteht das individuelle Erleben bzw. „das (Nutzer-) Erlebnis“. [43] S. 4. Qualinet definiert entsprechend dazu:

“An experience is an individual’s stream of perception and interpretation of one or multiple events.” [43] S. 4

Für das Erlebnis in dieser Definition als E in QoE ist noch nicht ein Beurteilungsprozess mit anschließender Bewertung notwendig [43] S. 4. Diesen Aspekt der Beurteilung des erlebten Leistungseindrucks erklärt das Zufriedenheitsmodell als Q in QoE separat. Grundsätzlich sollte mit dieser Erläuterung aber klarwerden, dass das E in QoE nicht im Sinne von etwas wie Vorerfahrungen zu verstehen ist. Daher ist Erleben oder Erlebnis als Übersetzung für Experience weniger leicht falsch interpretierbar als Erfahrung, was ggf. implizit eine gewisse Erwartungshaltung einschließt. Auf die Erwartung wird gesondert eingegangen.

Im UX-Umfeld wird teilweise die Trennung in neutrale Wahrnehmung und Interpretation von der anschließenden Bewertung aufgehoben und beide Schritte zusammen zum Erlebnis verbunden:

„Erlebnisse sind emotional bewertete Ereignisse. [...]“ [78] S. 5

So ist das Erlebnis je nach Sicht also entweder eine der Eingaben für die Bewertung oder mit dem Ergebnis der Bewertung verknüpft. Unter der Annahme, dass die Bewertung ein impliziter Bestandteil der Gesamtbetrachtung sowohl zur QoE als auch zur UX ist, ergibt sich nur ein formaler Unterschied hinsichtlich des Detailgrades der Betrachtung in einzelne Kognitionsschritte.

Die Interaktion der Nutzer mit einem Dienst/System/Produkt, die Wahrnehmung der Stimuli und das daraus entstehende Erlebnis sind miteinander im Nutzungsprozess verknüpft. Diese gemeinsame Betrachtung wird in der Nutzersicht auf Dienste noch weiter vertieft (siehe 3.5).

2.3.7 Grundsätzliches zu den Qualitätsaspekten der QoE

Bei den Qualitätsaspekten sind Parallelen zwischen den Konzepten zur QoE und zur User Experience (UX) erkennbar. Für die UX sind besonders die mit positiven Emotionen verbundenen Erlebnisse wie Freude, Vergnügen und Spaß wichtige Bausteine zu einer umfassenden Betrachtung des Erlebnisses. Die UX geht so über die oftmals nüchternen und fast ausschließlich nutzenorientierten Aspekte der Usability hinaus, die sich weitgehend auf eine Problemvermeidung im Umgang mit der Technik reduzieren lassen (siehe [69] S. 38f zur Zusammenfassung weiterführender Quellen dazu). Damit ergibt sich in der User Experience die Unterscheidung zwischen pragmatischer und hedonistischer Qualität ([79] S. 199 mit Verweis auf [80]⁴⁰). Die pragmatische Qualität bezieht sich dabei auf die wahrgenommene Nützlichkeit (Nutzen und Gebrauchstauglichkeit), konkret also Usability-Merkmale, die hedonistische Qualität ist hingegen nicht zielorientiert, sondern basiert eher auf Spaß, Neugier und sozialem Vergleich (nach der Zusammenfassung von [80] in [79] S. 199). Dabei wird davon ausgegangen, dass pragmatische und hedonistische Qualität unabhängig voneinander wahrgenommen werden und ungefähr gleich zur wahrgenommenen Attraktivität beitragen (nach der Zusammenfassung von [80] in [79] S. 200). Eine gewisse Usability in Sinn einer pragmatischen Gebrauchstauglichkeit bzw. Benutzbarkeit ist aber möglicherweise Voraussetzung für weitere hedonistische Qualitäten. Die UX-Forschung ist sich an dieser Stelle nicht einig, ob es tatsächlich eine Art interne Abhängigkeit in Richtung und Reihenfolge gibt, siehe Ausführungen und weitere Verweise unter [79] S. 201.

In der Aufschlüsselung möglicher interner Zusammenhänge der Qualitätsaspekte zur QoE, siehe Abbildung 10, sind gewisse Ähnlichkeiten zur Erklärung der System Acceptability und Usability nach Nielsen zu erkennen (siehe [81] S. 74ff). Diese finden sich auch in den Konzepten der User Experience wieder, siehe auch die zusammengefasste Erläuterung dazu in [79] S. 195. Dabei ist plausibel, dass für ausgewählte Aspekte jeweils andere Aspekte mitberücksichtigt werden und in ihrer Gesamtwirkung zur individuellen Beantwortung der Frage zur Akzeptanz führen.

⁴⁰ dort im Englischen zwischen ursprünglich „ergonomic“ und „hedonic“, die Umwidmung in pragmatisch und hedonistisch passt allerdings besser

2.3.8 Ausgewählte Qualitätsaspekte der QoE

[82] S. 13 führt eine Reihe von QoE-Maßen bei der Nutzung von Netzdiensten auf, die über das hinausgehen, was intuitiv oftmals als „die Qualität“ verstanden wird. Es gibt sowohl universelle Größen, als auch spezifische, die sich sinnvoll nur für einzelne Anwendungsfälle begründen lassen. Die „Qualität“ als solche im allgemeinen Sprachgebrauch bezieht sich meist auf den Inhalt, z. B. als „Audioqualität“, „Bildqualität“ oder „Videoqualität“. Sie ist somit bezogen auf die objektiven Eigenschaften des Inhalts als auch an deren subjektive Wahrnehmung als Wahrnehmungsqualität (z. B. Sprachverständlichkeit) gebunden. So ist beispielsweise die Audioqualität und damit die Sprachverständlichkeit nur in Verbindung mit Diensten interessant, die überhaupt eine Komponente Audio aufzuweisen haben.

Die **Effektivität** wird im Zusammenhang mit QoE als ein Maß der Genauigkeit und des *Zielerreichungsgrades* (Vollständigkeit) bei der Bearbeitung einer Aufgabe definiert [82] S. 10. Die **Effizienz** hingegen wird im Zusammenhang mit QoE als ein Maß der *eingesetzten Ressourcen* definiert, z. B. die benötigte Zeit [82] S. 10. Diese Maße sind damit äquivalent zu dem, was die ISO 9421 mit Effektivität und Effizienz erweitert um die Zufriedenheit unter dem Aspekt der Funktionalität zur Usability zusammenfasst, siehe Erläuterungen dazu in [83] S. 196f. und S. 207 (allerdings in der allgemeinen Anwendung der Mensch-Maschine-Schnittstelle).

Die **Usability** beschreibt die Gebrauchstauglichkeit, also eine Qualität im Sinne einer Eignung zu einem bestimmten Zweck. Effektivität und Effizienz sind zunächst generisch. Für eine konkrete Anwendung bzw. *Nutzungssituation* können die jeweils speziellen objektiven Maße zur QoE unterschiedlich sein, [82] S. 10 nennt dazu Beispiele:

- benötigte Anzahl Klicks
- benötigte Zeit (z. B. zur Erreichung eines gegebenen Ziels)
- Zeit zwischen Aktionen
- Anzahl Unterbrechungen oder Abbrüche der Bearbeitung
- korrekt bzw. falsch verstandene Worte (bei Konversationstest)

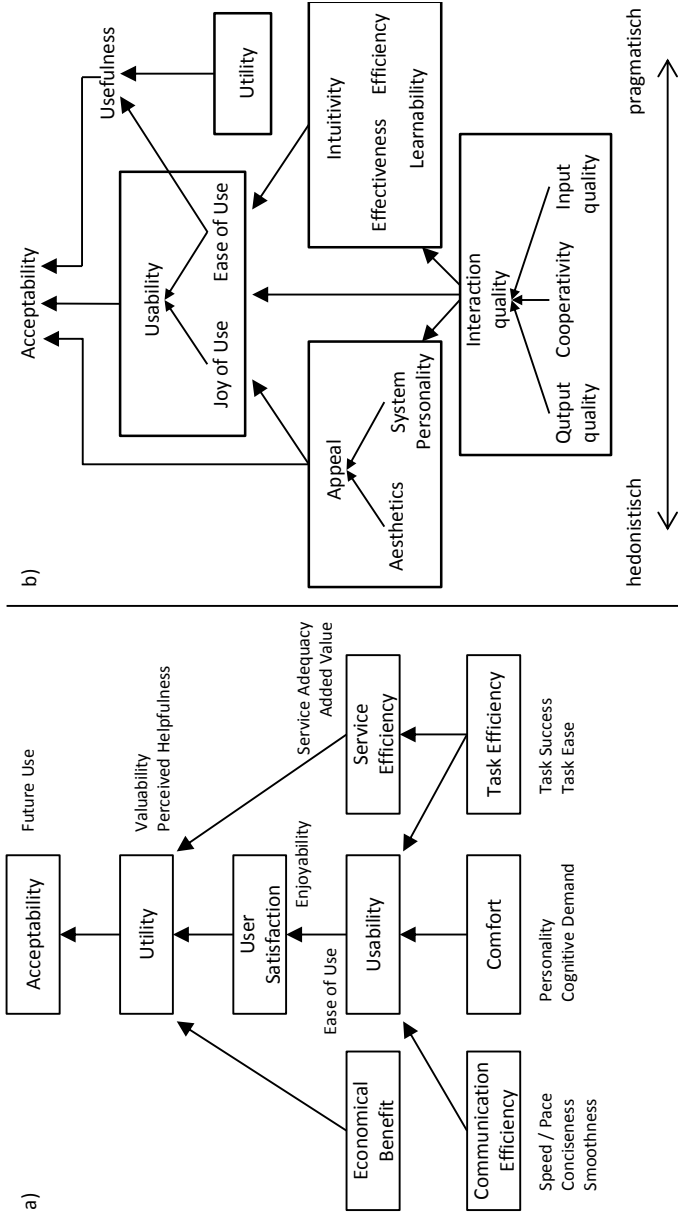


Abbildung 10: Mögliche Zusammenhänge der Qualitätsaspekte zur QoE

a) nach [30] S. 15, Ausschnitt, dort bezogen auf die Sprachkommunikation aber ähnlich zu verallgemeinern; b) nach [30] S. 17, für multimodale Interaktion, bei In- und Output-Quality sowie bei der allgemeinen Erscheinung sind Anknüpfungspunkte zum Endgerät gegeben

Die **Bequemlichkeit** der Nutzung eines Dienstes sagt etwas drüber aus, wie komfortabel die Nutzung eines Dienstes ist. Dies ist von den zugeordneten (*Un-*) *Annehmlichkeiten* abhängig, z. B. von einer einfachen und klaren Bedienung. Wie viel **Freude** (Enjoyment) es bereitet, einen bestimmten Dienst zu nutzen, kann als Maß für die *Motivation* gelten [82] S. 10. Prinzipiell wäre auch eine objektive Messung möglich, z. B. im Fall von Video-Streaming wie lange, oder wie oft ein Nutzer ein Video schaut oder im Fall von Web-Browsing die Wiederkehrtrate zu bzw. die Verweildauern auf bestimmten Webseiten.

Die Nutzung mobiler Services und Netze durch einen Nutzer erfolgt normalerweise zur *Erreichung* eines bestimmten *Zieles* oder zur *Befriedigung* von *Bedürfnissen* (Kommunikationsbedürfnis, Wunsch nach Unterhaltung, Informationsbedürfnis). Aus der Benutzung ergeben sich damit normalerweise ein **Nutzen** und somit auch ein gewisser Wert bzw. *Vorteil*. Eine mögliche weitere Aufschlüsselung der vor allem aus Marketingsicht wichtigen Begriffsdefinitionen für den Nutzen (eines Kunden) liefert das Nutzenschema/Nutzenleiter nach Vershofen (siehe [84] S. 79 aus einer Zeit lange vor Betrachtungen zu QoE oder UX). Demnach lässt sich zwischen einem eher funktionalen Grundnutzen und Zusatznutzen mit weiteren Formen, z. B. emotionalem oder sozialem Nutzen, differenzieren. In der Diskussion der Auswahl der QoE-Maße zur QoE-Evaluation in Testrunde 1 und 2 wird noch einmal kurz darauf zurückgekommen.

Die **Zufriedenheit** entsteht im *Abgleich* der Gesamtheit der wahrgenommenen Leistung als *Erlebnis* mit den individuellen *Erwartungen*. Die Aspekte der Zufriedenheit werden nachfolgend noch genauer zusammen mit ihrem Entstehen und dem Beurteilungsprozess erläutert.

[85] S. 2 definiert die **Akzeptanz** in der QoE-Domäne als die minimal akzeptable Qualität, die die Erwartungen der Nutzer an einen bestimmten Dienst erfüllt. Diese Überlegungen sind interessant, da sie ein Optimierungsziel vorgeben, was dem eigenen Konzept mit der Existenz von Schwellwerten entspricht. Im Unterschied zu einer graduellen und ggf. feingranularen Erfassung der Zufriedenheit wird bei der Akzeptanz von einer einfachen *binären*

Ja/Nein-Bewertung der Nutzer ausgegangen, die sich auf die weitere, d. h. zukünftige Nutzung bezieht: *Würden Sie (als Nutzer) den Dienst bei der erlebten Qualität/Leistung weaternutzen?*

Ob ein Nutzer möglichen Einschränkungen auch *Verständnis* entgegenbringt, sie also akzeptiert, kann auch von weiteren Faktoren abhängen. Die Akzeptanz kann mit den *Kosten* zur Nutzung verbunden sein und so zur Frage führen: *Wie viel wären die Nutzer bereit für eine bestimmte Qualität zu zahlen?* [86] S. 163 verweist auf Basis experimenteller Untersuchungen auf einen solchen möglichen Zusammenhang von Kosten und Akzeptanz. Demnach kann eine höhere Erwartungshaltung von Gold-Nutzern (Premium) im Vergleich zu Silber-/Bronze-Nutzern vermutet werden, die sich in niedrigeren Akzeptanzwerten für die gleiche Leistung niederschlägt.

2.3.9 Erlebnis und Erwartung im Beurteilungsprozess

Dieser Abschnitt liefert eine mögliche Erklärung, wie das *Q* in QoE als Bewertung in einem Beurteilungsprozess entsteht. Für die weiteren Überlegungen wird davon ausgegangen, dass die Zufriedenheit ein wichtiges und weitgehend umfassendes QoE-Maß ist, entsprechend konzentrieren sich die Ausführungen auf diesen Qualitätsaspekt der QoE.

Für die Zufriedenheit allgemein bzw. genauer für die Kundenzufriedenheit existieren verschiedene Modelle, siehe z. B. [87] S. 37. Einem Teil der vorgestellten Modelle gemeinsam ist die Überlegung, dass Zufriedenheit aus dem Abgleich einer Erwartungshaltung (Soll) und der Wahrnehmung der tatsächlich erlebten Leistung (Ist) entsteht. [88] S. 8 verweist bezüglich dieses „Konfirmations-/Diskonfirmations-Paradigma[s]“ auf weitere Arbeiten, ebenso wie [89] S. 128f. Demnach werden durch das Erlebnis die Erwartungen entweder nicht erfüllt, erfüllt oder übererfüllt. Als weitere allgemein anerkannte Erkenntnis wird davon ausgegangen, dass es keinen Punkt gibt, an dem Unzufriedenheit schlagartig in Zufriedenheit umschlägt, sondern dass von einer Toleranzzone auszugehen ist, die um einen Punkt der Indifferenz herum existiert (weder zufrieden noch unzufrieden). Diese Auffassungen zum Entstehen der Zufriedenheit werden sowohl in der eigenen Modellierung für die Zufriedenheit als relevantes QoE-Maß als auch in der später vorgestellten empirischen Erfassung (Skala) so genutzt: als relativer Abgleich zwischen dem subjektiven

Erlebnis (Experience) aus der wahrgenommenen Leistung und der individuellen Erwartung (Expectation) als Vergleichsgrundlage. Dabei spielt der Kontext ebenfalls eine Rolle (Interaktionskontext, situativer Kontext, soziokultureller Kontext), da er sowohl Erlebnis als auch Erwartung moderieren kann.

Die Erwartung selbst kann von innen oder von außen geprägt sein. Speziell für die mobile Internetnutzung können folgende Punkte⁴¹ für das erwartete *Soll* von Bedeutung sein:

- eigener Anspruch bzw. Niveau (Basic vs. Premium, Bronze vs. Gold, ...)
- Vorerfahrung/Vertrautheit (Anfänger vs. Experte)
- Vergleich zu vorherigen Leistungserfahrungen (allgemein bzw. in vergleichbaren Situationen)
- Orientierungsgrößen/Referenzen (z. B. Indikatoren zu den Netzeigenschaften wie Mobilfunktechnologie und Signalstärke)
- Wissen um Alternativen (z. B. andere Netztechnologien oder Anbieter)
- (Werbe-) Versprechungen/extern geweckte Erwartungen zu den jeweiligen Angeboten (Versorgung, Stabilität, Verfügbarkeit, Eignung für bestimmte Zwecke und Leistungsversprechen)
- Image (einer Marke oder eines Anbieters, z. B. des Netzanbieters)

Aus einigen der genannten Punkte wird deutlich, dass der Nutzer kommerzieller Angebote (Netzzugang und Dienste) auch immer ein Kunde ist. So können auch Aspekte aus der Welt der Betriebswirtschaftslehre, wie z. B. dem Marketing, aus nichttechnischer Richtung eine Rolle spielen, was in gewisser Art wieder zur ganzheitlichen (QoS-) Sicht führt.

Es ist plausibel davon auszugehen, dass es individuelle Unterschiede in der Erwartung gibt, die in der Konsequenz auch für objektiv gleiche Leistungen und wahrnehmbare Ereignisse/Stimuli sowie ggf. auch für subjektiv gleiche Erlebnisse zu unterschiedlichen QoE-Bewertungen führen. Gleichzeitig ist auch anzunehmen, dass es Gemeinsamkeiten, wie z. B. bestimmte Grenzen, zur Erwartung gibt, aus denen sich die Annahmen zu sinnvollen Anforderungen bei der Nutzung der Dienste ableiten lassen.

⁴¹ Zusammenstellung der Punkte in Anlehnung an [90] (nicht eingesehen) nach [91], angepasst an die eigenen Überlegungen für Mobilfunk und mobile Internetnutzung

Als bildliche Darstellung für den relativen Abgleich im Beurteilungsprozess zur Qualität kann eine Waage⁴² dienen, siehe Abbildung 11. Auf die Seite der wahrgenommenen Leistung als Erlebnis und auf die andere Seite der Erwartung wirken jeweils die entsprechenden Einflussfaktoren ein. Als wichtiger Aspekt sei betont, dass nicht nur die Erwartungsseite mit erwartetem „Soll“ subjektiv ist, sondern auch die Seite der erlebten Leistung mit dem „Ist“, da wie zuvor diskutiert, auch die Wahrnehmung subjektiv ist. Die Abbildung des Ausschlags der Waage als innerer Wert auf einen externen Wert erfolgt durch Encoding, also durch Zuordnung zu einem Wert einer Skala, entsprechend zum erfassten QoE-Maß, hier der Zufriedenheit. In der Diskussion zur Empirie in der vorliegenden Arbeit wird die Problematik dieser Abbildung auf die Skala bei der Testoperationalisierung (4.3.1) und in der Auswertung (8.6.3) noch einmal aufgegriffen.

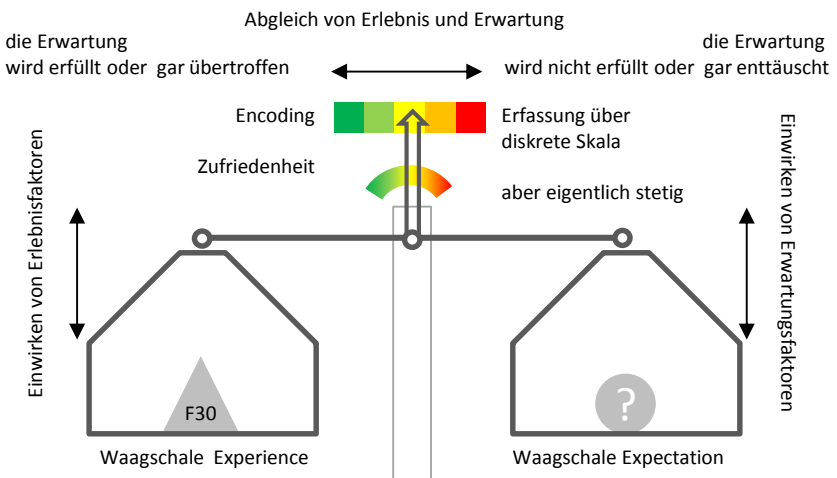


Abbildung 11: QoE-Teilmodell zur Zufriedenheit als Waage zum Abgleich von Erlebnis und Erwartung

Abgleich: Überwiegt das Erlebnis, d. h. die erlebte Leistung, dann neigt sich die Waage zur Seite Erlebnis (und umgekehrt).

Encoding: Das Ergebnis wird einem externen Wert auf einer (diskreten) Skala zugeordnet. Darstellung inspiriert von [91] mit Verweis auf [90] (nicht eingesehen) mit Bezug auf die Mitarbeiter-/Kundenzufriedenheit, eigene Darstellung auf den Beurteilungsprozess zur QoE angepasst und um Abbildung auf Bewertungsskalen und Auswirkungen des Abgleichs erweitert.

⁴² siehe vorherige Fußnote

Die skizzierte Waage soll als Sinnbild eines einfachen confirmatorischen Modells begriffen werden. Ein deutlich komplexeres und detaillierteres Modell zum Prozess der Qualitätsformung zur QoE ist in [31] S. 22f beschrieben (mit Verweis auf weiterführende Literatur dazu). Dabei kommt dem Kontext, bzw. genauer der Wahrnehmung der kontextuellen Informationen neben den physikalischen Reizgrößen der interessierenden Signale eine große Bedeutung zu. Insbesondere das interne Zusammenwirken der menschlichen Einflussfaktoren (Wahrnehmung, Interpretation, innerer Zustand, Erwartungen, ...) ist dabei sehr komplex.

Für die weitere Diskussion sollte die Unterscheidung in Erlebnisfaktoren und Erwartungsfaktoren ausreichen, wie in Abbildung 11 skizziert, je nachdem an welcher Seite der Waage der Einfluss ansetzt. Erlebnisfaktoren führen zu Unterschieden in der QoE-Bewertung durch eine Beeinflussung der Leistung (objektiv) bzw. der Wahrnehmung der Leistung (subjektiv) als Erlebnis. Erwartungsfaktoren führen zu Unterschieden in der QoE-Bewertung durch eine Beeinflussung der Erwartung (subjektiv). Die dazugehörigen Überlegungen zur Unterscheidung der Wirkseite und Wirkrichtung sind jedoch nicht immer trivial, bzw. teilweise a priori gar nicht plausibel zu beantworten, wie sich in der weiteren Diskussion zeigen wird.

2.3.10 Arten von Zufriedenheitsfaktoren

Für die weiteren Überlegungen zur Bewertung von Nutzungs- und Netzsituationen und zur späteren Auswertung ist die Unterscheidung zwischen gängiger Sichtweise und 2-Faktoren-Theorie zur Zufriedenheit nach [89] S. 122 interessant, siehe Abbildung 12. In der gängigen Sichtweise werden Zufriedenheit und Unzufriedenheit als zwei entgegengesetzten Ausprägungen eines Merkmals aufgefasst. In der 2-Faktoren-Sichtweise hingegen sind Zufriedenheit und Unzufriedenheit zwei verschiedene Größen, die von unterschiedlichen Faktoren beeinflusst werden. [89] S. 122

Formalisiert wird zwischen drei Arten von Einflussfaktoren auf die Zufriedenheit und dem jeweiligen Erfüllungsgrad der Erwartungen unterschieden: Basisfaktoren, Leistungsfaktoren und Begeisterungsfaktoren (siehe [89] S. 122 nach Herzberg). Die Auswirkungen der Nichterfüllung, Erfüllung und Übererfüllung sind Tabelle 3 zu entnehmen. Das Zusammenspiel der einzelnen

Faktoren beeinflusst die Gesamtzufriedenheit. Für die Betrachtungen zum Zusammenhang von QoS und QoE ist davon auszugehen, dass sich für alle drei Arten von Zufriedenheitsfaktoren für die mobile Nutzung des Internets entsprechende Beispiele finden lassen.

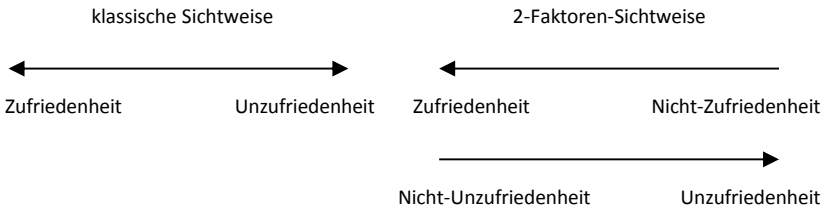


Abbildung 12: Klassische Sichtweise und 2-Faktoren-Sichtweise zur Zufriedenheit
nach [89] S. 122

Tabelle 3: Arten von Zufriedenheitsfaktoren und ihre Auswirkungen

Art und Erläuterung	Nichterfüllung	Erfüllung	Übererfüllung
Begeisterungsfaktoren (Motivatoren) werden nicht als selbstverständlich vorausgesetzt, führen zu positiven Emotionen; ggf. als Wettbewerbsvorteil oder Differenzierungsmerkmal nutzbar	ohne Auswirkung bzw. Nicht-Zufriedenheit	starke Zufriedenheit	sehr starke Zufriedenheit
Leistungsfaktoren werden erwarten und explizit verlangt, führen entweder zu negativen oder positiven Emotionen	Unzufriedenheit	moderate Zufriedenheit	Zufriedenheit
Basisfaktoren (Hygienefaktoren) werden als Kernleistung bzw. grundlegende Eigenschaft selbstverständlich vorausgesetzt, Ziel ist die Vermeidung von negativen Emotionen	Unzufriedenheit	neutraler Zustand der Nicht-Unzufriedenheit	ohne Auswirkung, keine besondere Wertschätzung, nur Nicht-Unzufriedenheit

auf Basis der Motivator-Hygiene-Theorie von Herzberg zusammengestellt nach [89] S. 122 und [92]

[78] S. 6f verweist in Bezug auf die Differenzierung des Einflusses von Produktmerkmalen auf das Nutzererlebnis in der User Experience auf das ähnliche Kano-Modell von Zufriedenheitsfaktoren auf die Kundenzufriedenheit (siehe [93], nicht eingesehen da auf Japanisch, Erläuterungen dazu in [94]). [78] S. 6f gibt dazu ein anschauliches Beispiel, wie sich Zufriedenheitsfaktoren bei den mobilen Endgeräten im Laufe eines Jahrzehnts verschoben haben. Waren Ende der 90er Jahre integrierte Antennen und die Möglichkeit zur Nutzung einfacher monochromer Bildnachrichten noch Begeisterungsmerkmale, so sie diese mittlerweile selbstverständliche Basisfaktoren [78] S. 6f. Die neuen Begeisterungsfaktoren sind bei modernen Smartphones hochauflösende Bildschirme, edle Materialien und innovatives Design [78] S. 6f. Diese Entwicklung bezieht sich dabei nicht nur auf die Merkmale der Endgeräte wie im genannten Beispiel, sondern auch auf die eigentlichen Merkmale der Nutzung von Telekommunikationsdiensten und Multimediainhalten im Allgemeinen: gestern noch Begeisterungsfaktor, heute Leistungsfaktor und morgen schon Basisfaktor. [31] S. 28 gibt dazu in Bezug auf die QoE weitere Beispiele. Dieser Wandel der Zufriedenheitsfaktoren vollzieht sich mit dem Wandel der Seite der Erwartung, die als dynamisch anzusehen ist, was zum Nutzungsprozess und der Dynamik der QoE überleitet.

2.3.11 Nutzungsprozess und Dynamik der QoE

Verschiedene Arbeiten (z. B. [95]) verweisen darauf, dass QoE kein singuläres oder statisches Ereignis ist, sondern besser als dynamischer Prozess begriffen werden sollte. Die Nutzung eines Dienstes beinhaltet entsprechend immer auch eine Zeitkomponente. Diese bezieht sich entweder auf einen kurzen Zeitraum (micro-temporal), z. B. einen einzelnen Nutzungsprozess (Session), ggf. beeinträchtigt von kurzzeitigen Störungen, oder auf den Eindruck über einen längeren Zeitraum mehrerer Nutzungen, über Tage oder auch Wochen hinweg (macro-temporal) [95] S. 161f. Der Nutzungsprozess verläuft dabei in mehreren Phasen des Erkundens und Reflektierens, in denen sich das Nutzungserlebnis entwickelt [95] S. 161f. Micro-temporale Aspekte sind bei der QoE-Messung als menschlicher Faktor auf die Nutzerzufriedenheit ggf. zu berücksichtigen. Von macro-temporalen Aspekten ist eine Auswirkung auf die langfristige Kundenzufriedenheit zu erwarten (Quality of Customer Experience, QoCE wie in [96] diskutiert).

[97] weist daraufhin hin, dass die Nutzer die Erlebnisse einzelner Nutzungen vergleichen und daraus eine bestimmte Erwartungshaltung entwickeln. Die Mechanismen zur Dynamik der „Zone-of-Tolerance“, abseits der QoE im Sinne einer allgemeinen Servicequalität, werden anschaulich in [98] beschrieben. Der Zufriedenheitskorridor, definiert durch Erwartungen, gefühlter Qualität und endgültiger Bewertung sowie dem Einfluss von Nichterfüllung, Erfüllung und Übererfüllung von Erwartungen, verschiebt sich durch die Nutzung und Nutzungserfahrung dynamisch. In Ergänzung zu den vorherigen Ausführungen zum Entstehen von Zufriedenheit wird besonders herausgearbeitet, wie die zeitliche Abfolge verschiedener Leistungserfahrungen (gut/schlecht) die Toleranzzone positiv wie auch negativ beeinflussen kann.

Die Betrachtung der zeitlichen Dynamik ist ein weiterer Anknüpfungspunkt zur User Experience, wobei in Phasen *vor*, *während* und *nach* der Nutzung unterschieden wird ([73] S. 33 und [99] mit Bezug auf die ISO 9241-210):

- vor: gewisse Erwartung und Annahmen noch vor der Nutzung
- während: Usability mit effektiver und effizienter Aufgabenerledigung ohne Einschränkungen
- nach: Reflexion zur erlebten Nutzung, ggf. Bildung einer emotionalen Bindung

Die Qualitätsaspekte (zumindest der User Experience) können sich über die Zeit unterschiedlich entwickeln [100]. Dazu resümiert [100] S. 78, dass die hedonistische Qualität mit der Zeit verblassen kann, z. B. weil der Reiz des Neuen verloren geht; die pragmatische Qualität hingegen bleibt stabiler bzw. gewinnt sogar leicht, wofür die mit der Zeit zunehmende Vertrautheit eine plausible Begründung ist [100] S. 78.

Die temporale Komponente der Nutzung kann auch in QoE-Episoden beschrieben werden, was vor allem mit Blick auf eine QoE-Evaluation zu beachten ist (nach [101] S. 134ff):

- momentan: kurze Stimuli-Eindrücke werden quasi direkt zusammen mit dem Erleben bewertet, prinzipiell sind auch kurzfristige oder kontinuierliche Änderungen möglich

- episodisch: in Erinnerung an eine Episode der Nutzung, direkt nach dem Ende eines episodischen Erlebens (z. B. nach dem Schauen eines Video-clips)
- kumulativ: in Erinnerung an mehrere Episoden als Ergebnis längerfristiger Nutzung und zukünftiger Nutzungsannahmen

Die Erkenntnisse zur möglichen Dynamik der QoE sind für eine vollständige Betrachtung zum QoS-QoE-Zusammenhang zu berücksichtigen, insbesondere wenn auch von einer Dynamik der QoS auszugehen ist (wechselnde Netzsituationen).

2.4 Kontext und Nutzungssituationen

Der Kontext bildet eine der drei überlappenden Kategorien von Einflussfaktoren auf die QoE. Für die weitere Berücksichtigung wird der Kontext in verschiedene Arten zerlegt. Für den Kontext ist von einem engen Zusammenhang mit Erwartungen und Zielen der Nutzer auszugehen, woraus die Nutzungssituationen abzuleiten sind. Dazu werden die Besonderheiten einer mobilen Nutzung erörtert. Der Abschnitt schließt mit der Diskussion marktbezogener Faktoren, die sich ebenfalls im Kontext einordnen lassen.

2.4.1 Arten von Kontext

Intuitiv ist der Kontext vermutlich die schwammigste Einflusskategorie, unter der man sich alles irgendwie, nichts aber konkret vorstellen kann. Dies mag dazu verleiten, den Kontext als Sammelbecken für „alle übrigen“ Einflüsse zu nutzen, die sich nicht direkt den menschlichen Faktoren oder Systemfaktoren der Technik zuordnen lassen. Deshalb sind zur Konkretisierung die Klassifikationen aus der Literatur für diesen Aspekt besonders wichtig. [95] S. 160 verweist für den Kontext in Bezug auf QoE auf das HCI-Modell zum Kontext von Mantovani [102]. Demnach gibt es verschiedene Arten von Kontext, die einen Nutzer beeinflussen können.

Für die eigenen Überlegungen passt das vereinfachte Modell mit 3 Schalen recht gut, siehe Abbildung 13a (nach [95] S. 160):

- soziokultureller Kontext (Herkunft, Bildung, persönliche Prägung, Werte und Vorstellungen)
- situativer Kontext (zeitlich und örtlich, z. B. unterwegs, zu Hause, Labor-situation, Anwesenheit anderer Menschen, Aufgabe oder Nutzungsziel)
- Interaktionskontext (z. B. das genutzte Gerät, die Gerätekategorie, App, mit technischen und menschlichen Faktoren)

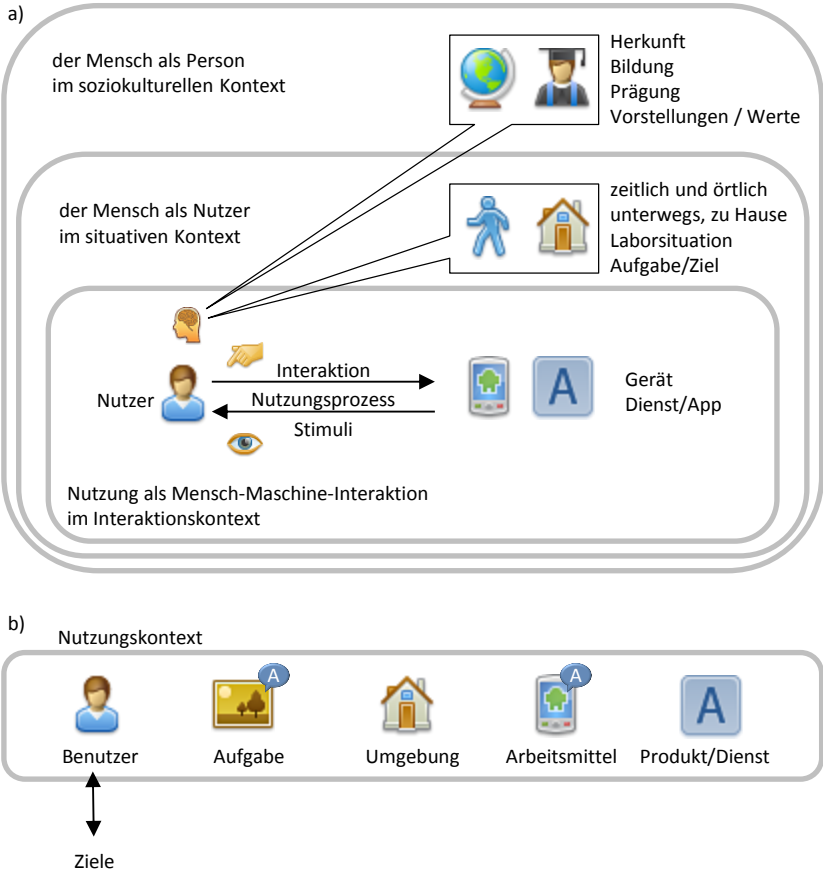


Abbildung 13: Verschiedene Arten von Kontext

a) ineinander geschachtelt, angelehnt an das QoE-Kontext-Modell aus [95] S. 160, vereinfacht
 b) Nutzungskontext aus dem Anwendungsrahmen für Gebrauchstauglichkeit nach ÖNORM EN ISO 9242-11:1995, angelehnt an [79] S. 197

Eine etwas feingranulare Aufteilung mit weiteren Kategorien des Kontexts sieht [72] S. 64ff vor (wie in Abbildung 9b skizziert). Prinzipiell reichen aber die genannten 3 Kategorien für die weiteren Betrachtungen aus. Im Anwendungsrahmen der Definitionen zur Gebrauchstauglichkeit nach [79] S. 197, siehe Abbildung 13b, stellt sich der Nutzungskontext als Schnitt durch die o. g. Schalen dar, der aber weitgehend die gleichen Elemente bzw. Komponenten umfasst. Entsprechend ist es eher eine Frage der gewünschten oder notwendigen Granularität zur Beschreibung und Einbeziehung in eine Modellbildung.

Der Kontext kann auch für eine „Wahrnehmungs- und Beurteilungssituation“ [30] S. 6 für die Formierung der QoE stehen, Einflüsse des Kontexts wirken dabei moderierend. Es wird allgemein von einer Wirkung der Situation auf die Erwartung an ein Produkt ausgegangen [78] S. 8, was im erweiterten Sinn auch die Dienste bzw. eine mobile Netznutzung als genutztes „Produkt“ umfasst. Im situativen Kontext einer mobilen Nutzung können dabei aber sogar zwei Situationen verknüpft sein: eine spezifische Netzsituation (aus Sicht der Technik/des Netzes) und eine spezifische Nutzungssituation (aus Sicht des Nutzers). Ein Beispiel soll die verdeutlichen: [5] S. 8 gibt auf Basis einer Befragung einen Hinweis auf Besonderheiten durch den situativen Kontext bei der Nutzung mobiler Datendienste. Demnach sind je nach Umgebung deutliche Unterschiede in der Zufriedenheit festzustellen, z. B. am Flughafen höher oder im Bus niedriger. Dabei wird allerdings nicht ausgeführt, worauf diese Unterschiede zurückzuführen sind. Intuitiv wären Leistungsunterschiede, also unterschiedliche Netzsituationen als Ursache plausibel, genauso sind aber auch Erwartungsunterschiede durch die Situation möglich.

Ein Hinweis auf den tatsächlichen situativen Kontext, in dem eine Nutzung mobiler Apps stattfindet, gibt [103] auf Basis einer Befragung. Demnach erfolgt die Nutzung der Apps, wenn die Nutzer allein sind (70 %), zum Zeittot schlagen (69 %), oder während auf jemanden oder etwas gewartet wird (61 %) (nur Top-3-Nennungen aus [103]⁴³ für Smartphones, Tablets leicht abweichend). Auf den besonderen situativen Kontext einer Laborsituation und dessen Wirkung wird bei der Analyse der Testbedingungen zur eigenen Studie noch gesondert eingegangen.

⁴³ Studie US-Markt, Q3 2014

Der Interaktionskontext ergibt sich aus der Beschreibung der Elemente Endgerät sowie App und Inhalt, die den genutzten Dienst (als Produkt) repräsentieren. Die Diskussion zur Wirkung erfolgt aufgeschlüsselt bei den genannten Elementen. Wie beim situativen Kontext ist auch beim Interaktionskontext ein Einfluss sowohl auf das Erlebnis als auch die Erwartung plausibel. Diese mögliche beidseitige Beeinflussung erschwert Abschätzungen zur Wirkung im Beurteilungsprozess deutlich. Zwischen situativem Kontext und Interaktionskontext sind die Indikatoren und Hinweisen zur Netzsituation einzuordnen (Netztechnologie und Empfangsstärke), siehe Beispiele und Anmerkungen in 2.5.4.

Der soziokulturelle Kontext kann durch einen Blick auf Nutzerpersönlichkeit und Nutzertyp näher beschrieben werden. Die vertiefende Diskussion dazu erfolgt zusammen mit der Darstellung der Ausprägungen der Probanden im Kapitel 7.

2.4.2 Nutzungsziel und Aufgabe

Eine spezifische Nutzungssituation beinhaltet über den Kontext auch ein Nutzungsziel bzw. einen angestrebten Nutzen. Dies erfolgt entweder implizit über die Intention des Nutzers, z. B. aktuelle Informationsbeschaffung oder Unterhaltung, oder explizit, wie im Test, über ein Szenario und die Aufgabenstellung, die relevante Elemente und Schritte vorgeben. Die Nutzungsziele lassen sich dabei in Kategorien bzw. Nutzungsklassen einteilen. Diese können völlig nichttechnisch beschrieben werden. Eine mögliche Aufteilung ist in Tabelle 4 vorgestellt.

Auch wenn die Kategorien prinzipiell beliebig und frei definierbar sind, so sollte doch eine einstellige Anzahl ausreichen, die meisten Nutzungsziele passend zusammenzufassen. Die Nutzungsziele und Nutzungsklassen sind aber nicht nur beliebig, sie sind auch sehr abstrakt bzw. generisch und kaum in einer technischen Beschreibung geeignet zu fassen. Die Nutzungsklassen sollen daher weiter aufgeschlüsselt werden, in Nutzungsarten. Diese sind schon konkreter und besser geeignet, über mögliche Anforderungen und Einschränkungen eine sinnvolle Beschreibung zu verfassen. Ein sehr grober Ansatz dazu wäre eine Unterscheidung des vordergründig genutzten Mediums (Text, Bild,

Audio, Video ... Multimedia) und über das Kommunikationsszenario die Unterscheidung der beteiligten Partner (Person-to-Content oder Person-to-Person). Mit Nutzungsziel und Nutzungsart sind Anknüpfungspunkte zu den Diensten und zur weiteren dienstspezifischen Diskussion des QoS-QoE-Zusammenhangs gegeben (ab 3.3).

Bestimmte Nutzungsarten, wie z. B. Video-Streaming, können sich unter verschiedenen Nutzungszielen wiederfinden. Dieser Punkt ist wichtig, da er eine Begründung liefern kann, warum beispielsweise ein Video-Stream eines Musikclips zur Unterhaltung und ein ansonsten (technisch) vergleichbarer Video-Stream aus einer Bilanzpressekonferenz zu geschäftlichen Zwecken doch zu unterscheiden sind – zumindest aus Sicht der Anforderungen eines potenziellen Nutzers.

Tabelle 4: Nutzungsklassen aus Nutzungszielen und Nutzungsarten

Nutzungsklasse (aus Nutzungsziel)	Nutzungsart (Unterklasse)
Arbeit / Geschäft	Tunneling (VPN), Remote-Arbeit (Außendienst), Terminplanung/Kalender ...
Bildung	e-Vorlesungen, Video (-Streaming), interaktive Lernmaterialien ...
Dienstleistung	Shopping, Buchungsdienste, Online-Banking, App-Stores, Speicherdienste, Backup-Dienste ...
Informationssuche	Suchmaschine nutzen, Web-Browsing (allgemein und gezielt), Nachschlagewerke (z. B. Wikipedia) ...
Kommunikation	Messaging, E-Mail, Soziale Netze, Telefonie (Internet, VoIP), Videotelefonie (Internet), e-Meetings, Whiteboard-Anwendungen ...
Ortsbezogene Dienste (Location Based Services)	Kartendienste (Orientierung), Navigation (Routen), Points of Interest (lokale Angebote, Fahrpläne) ...
Unterhaltung	Video (-Streaming), Audio/Musik (-Streaming), Spiele ...

nichterschöpfende Zusammenstellung, angelehnt an [104] S. 12 und eigene Überlegungen im Kreis der Projektpartner

Grundsätzlich ist in den Nutzungszielen eine Ähnlichkeit zu den Überlegungen der User Experience zu erkennen. [105] führt dazu sechs „User Experience Spaces“ auf, die untereinander gewisse Überschneidungen aufweisen und jeweils durch ganz spezifische Fragestellungen zu beschreiben sind:

- I-ware: die Verwaltung der eigenen Daten
- You-ware: Kommunikation und Beziehungen zu anderen Nutzern, z. B. Nachrichtenaustausch oder auch Soziale Netze
- Know-ware: Informationssuche, z. B. per Google etc.

- Be-ware: Selbstverbesserungen, Bildung, Assistenzsysteme, Gesundheits-Apps etc.
- Fun-ware: Unterhaltung wie Spiele oder Medienkonsum
- Buy-ware: E-Commerce, Online-Shopping

2.4.3 Besonderheiten der mobilen Nutzung

“Consumers expect robust mobile experiences that are more than a squeezed PC experience, meaning the ebusiness professionals must design for mobile to take advantage of its highly contextual and task-oriented nature.” [106] zitiert nach [107]

Es stellt sich grundsätzlich die Frage: Ist die mobile Nutzung am Smartphone oder Tablet anders als die stationäre Nutzung an einem PC? Aus technischer Sicht ist diese Frage eindeutig positiv zu beantworten, wie die Beschreibungen zum Netz oder auch zum Endgerät im Vergleich einer mobilen zu einer stationären Nutzung zeigen:

- Netz: höhere Latenz, niedrigere (und zwischen den Nutzern geteilte) Bandbreite und typischerweise schwankende Netzcharakteristik
- Endgerät: geringere Hardware-Leistung, Einschränkungen bei Eingabe und Ausgabe durch Abmessungen des Gerätes

Aber auch in der Nutzersicht sind deutliche Unterschiede zu erkennen. Besonders in der mobilen Nutzung ist von einer stark kontext- und aufgabenorientierten Nutzung auszugehen. Dies kann die Existenz der vielen spezialisierten Apps erklären – obwohl eine ähnliche Nutzung ebenfalls im (mobilen) Browser möglich wäre – und in der Desktop-Nutzung auch im Browser üblich ist (Beispiel Google Maps). Schon alleine aus dem Interaktionskontext mit o. g. begrenzten Ein- und Ausgabemöglichkeiten leiten sich besondere Erfordernisse für die gebrauchstaugliche Gestaltung der Nutzeroberflächen und Inhalte der mobilen Apps und Dienste allgemein ab. Gerade diese Faktoren können einen wesentlichen Einfluss auf die Seite Experience aus Sicht der Nutzer haben. Die mobile Nutzung ist demnach nicht der kleine Bruder der PC-Nutzung auf dem Leistungsstand von vor 10 Jahren, sondern erfordert eine besondere Berücksichtigung.

Mit Blick auf die QoE-Formierung ist natürlich auch die Seite der Erwartung interessant. Intuitiv wäre einerseits plausibel anzunehmen, dass die Nutzer ohnehin nur geringere Anforderungen an die mobile Nutzung stellen und so beispielsweise kein HD-Videostreaming erwarten, andererseits wäre es auch plausibel, von teilweise sogar höheren Anforderungen auszugehen, z. B. bei Ladedauern mobiler Websites, der Kartendarstellung der aktuellen Position zur Orientierung, der Fahrplanauskunft zum nächsten Bus, der Buchung der zugehörigen Fahrkarte etc. An den Beispielen wird deutlich, dass dies Nutzungssituationen sind, die eben gerade durch den mobilen Charakter und u. U. einen gewissen äußeren Zeitdruck geprägt sein können. Letztlich sind Fragestellungen dazu nur empirisch zu beantworten.

Studienergebnisse zur Desktop-/PC-Nutzung und dort genannte Werte für maximale Seitenladedauern etc. sind entsprechend kritisch zu hinterfragen, ob sie auch auf die mobile Nutzung Anwendung finden können, oder eher nur als grober Richtwert der Größenordnung gesehen werden sollten. Für Ergebnisse aus Befragungen ist entsprechend zu verfahren, insbesondere, wenn sie außerhalb eines entsprechenden Kontexts erhoben worden sind.

2.4.4 Marktbezogene Faktoren und nichttechnische Angebotseigenschaften

Überwiegend werden die marktbezogenen Faktoren und nichttechnische Angebotseigenschaften (meist auf den Preis konzentriert) nicht als eigenständige Kategorie von Einflussfaktoren geführt, sondern den kontextuellen Faktoren zugeordnet (siehe Abbildung 9 und [72] S. 67).

Für die marktbezogenen Faktoren nimmt der Mensch die Rolle eines Kunden ein. Grundsätzlich ist von einer Beeinflussung der Erwartungsseite durch die nichttechnischen Faktoren des Angebotes auszugehen. In einem Realweltkontext können so die Kosten der Nutzung eines Dienstes die Zufriedenheit beeinflussen. In Befragungen zeigte sich eine direkte Korrelation zwischen der Zufriedenheit mit dem PreisLeistungsverhältnis und der Zufriedenheit mit der Netzleistung (siehe [5] S. 6). [108] S. 115 verweist in diesem Zusammenhang

auf eine entsprechende experimentelle Studie [109] (stationärer Internet-Zugang, Ende der 90er Jahre). In dem beschriebenen INDEX-Feldversuch⁴⁴ wurden Untersuchungen von QoS-Konfigurationen mit unterschiedlichen Preissetzungen und variabler Netzparametrisierung durchgeführt, um die Zahlungsbereitschaft für bessere Dienstgüte zu ermitteln. In diesem Ansatz erweitert damit der Preis den Parameterraum. Die Annahme war dabei, dass der sich ein Nutzer bei einem kleinen Preis mit weniger Dienstgüte zufriedengibt als bei einem hohen Preis, was sich ausgehend von den Ergebnissen (in gewissen Bereichen) auch bestätigte [108] S. 115. Ergänzend sei auf 2.3.8 mit weiteren Anmerkungen zu den Kosten und Studienergebnissen zur Akzeptanz verwiesen.

Dabei ist grundsätzlich zu hinterfragen, ob die Kunden nur dafür bezahlen, dass es funktioniert – oder dafür, dass es (besonders) gut funktioniert – und wenn dem so ist, wie viel ihnen dies wert ist. Neben den Kosten können auch Werbeversprechen oder entsprechend gepflegte Markenimages als nichttechnische Angebotseigenschaften zählen, die die Erwartung und damit die Zufriedenheit beeinflussen. Diese Aspekte sind nicht Gegenstand der weiteren detaillierten Betrachtungen, gehören aber zu einer ganzheitlichen und umfassenden Darstellung von QoS und QoE und sollten daher nicht unerwähnt bleiben. Für eine vollständige Betrachtung der gesamten Problematik zur QoE sind daher auch betriebswirtschaftliche Erwägungen einzubeziehen. Auf den Markt, Customer Experience, Quality of Customer Experience und Quality of Business kann nicht weiter detailliert eingegangen werden, da dies außerhalb der Kompetenz des Autors liegt. Eine intensivere Beschäftigung zur geschäftlichen Perspektive auf QoE findet der interessierte Leser in [110].

⁴⁴ INDEX: Internet Demand Experiment

2.5 Konzeptueller Zusammenhang zwischen QoS und QoE

Nachdem die Konzepte zu QoS und QoE jeweils einzeln dargestellt wurden, soll nun der Zusammenhang zwischen QoS und QoE, zunächst auf konzeptueller Ebene, diskutiert werden. Dazu wurde in der Literatur recherchiert, wobei einige Ansätze mehrfach auftauchten, von denen wiederum einige in die weiteren eigenen Überlegungen einfließen. Diese Konzepte werden zum konzeptuellen Gesamtzusammenhang verknüpft und dazu nachfolgend noch vertieft diskutiert. Aus den Betrachtungen der unterschiedlichen Konzepte ergibt sich, dass es keine Beschreibung des „einen“ QoS-QoE-Zusammenhangs geben kann, sondern derartige Betrachtungen dienstspezifisch erfolgen müssen. Aus den Ansätzen zum konzeptuellen Zusammenhang zwischen QoS und QoE wird daher im nachfolgenden Kapitel ein formalisiertes Modell zum dienstspezifischen Ursache-Wirkungs-Zusammenhang konstruiert.

2.5.1 Schichtung der Qualitäten

Die Unterteilung in verschiedene Schichten, bzw. die Bildung und Betrachtung unterschiedlicher Qualitäten ist eines der wiederkehrenden Konzepte, die die Recherchen zur (frühen/holistischen) QoS wie auch zum Zusammenhang von QoS und QoE ergeben haben. Die Sichtung der Literatur gab einige Hinweise für eine sinnvolle Strukturierung in Form von aufeinander aufbauenden Schichten für die Modellierung des QoS-QoE-Zusammenhangs. Einen Überblick zu verschiedenen Ansätzen liefern [4] und [111]. Die betrachteten Ansätze sehen meist 4 ± 1 Schichten bzw. Hierarchiestufen der Qualität vor. Für die Schichten, Leistungen oder Qualitäten sind unterschiedliche Abgrenzungen und Bezeichnungen möglich, die je nach Arbeiten ggf. noch leicht in ihren Bedeutungen variieren. Die grundsätzlichen Überlegungen zu Unterscheidung und Bezeichnung sind insgesamt aber nachvollziehbar und vereinbar. Jede der Schichten kann durch unterschiedliche Parameter bzw. Maße als Qualitätsmerkmale beschrieben werden.

Mit Blick auf die Schichtbildung sind zwei Ansätze erkennbar: mit unmittelbarem Bezug zum Netzwerk-Stack bzw. direkt daran orientiert oder relativ abstrakt mit ggf. nur mit indirektem Bezug zum Stack.

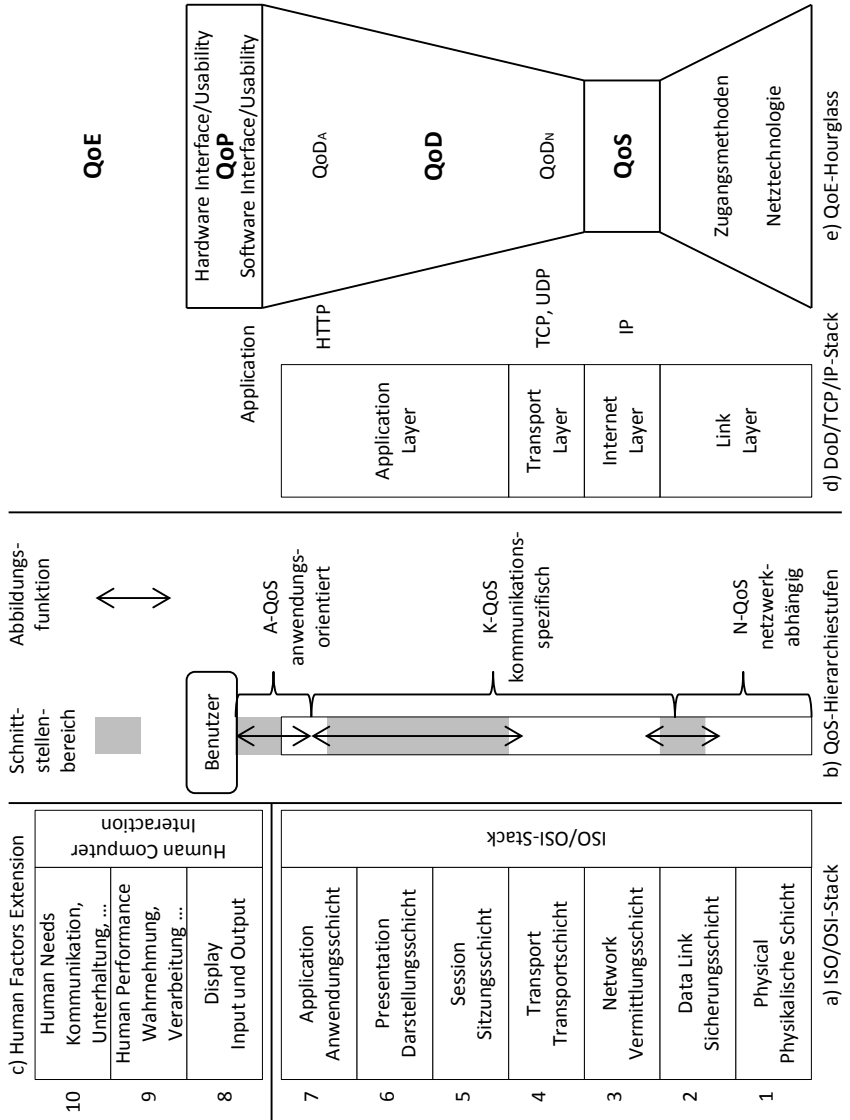


Abbildung 14: QoS-Schichtmodelle/Klassifikation im Vergleich (Orientierung am Stack)

- a) ISO/OSI-Stack zusammengestellt nach [112] S. 77, b) QoS-Hierarchiestufen nach [45] S. 15
- c) Human Factors Extension nach [113], d) DoD/TCP/IP-Stack in Anlehnung an [34] S. 351
- e) QoE-Hourglass-Modell nach [114] S. 27 (auf den TCP/IP-Stack bezogen, in der Form einer angedeuteten Eieruhr, mit IP respektive QoS an der dünnsten Stelle, siehe All-over-IP unter 3.4.2 zur Orientierung)

Bei Angaben zum Netzwerk-Stack erfolgt meist eine Orientierung am ISO/OSI-Schichtreferenzmodell⁴⁵ mit 7 Schichten, siehe Abbildung 14a und Erläuterungen dazu in [112] S. 76f. Eine QoS-Hierarchiestufe bzw. Schicht kann eine oder mehrere Schichten im OSI-Modell umfassen, siehe Abbildung 14b. Diese Einteilung und Zuordnung zu den Schichten des OSI-Modells ist aber nur eine mögliche, die hier exemplarisch ausgewählt wurde. Es wurden weitere ähnliche Aufteilungen gefunden, die sich in der Schichtbildung leicht durch Bezeichnung und Zuordnung unterscheiden. In der praktischen Anwendung abseits der Theorie des OSI-Modells sind diese Schichten aber ohnehin anders. Prinzipiell sind die Kernaspekte aber ähnlich erfasst (nach [45] S. 14ff zu den Schichten):

- oberste Schicht (anwendungsorientierte QoS): Beschreibung aus Sicht der Anwendung, die transportierten Nutzdaten erhalten einen Zweck und Bedeutung
- mittlere Schicht (kommunikationsspezifische QoS): transportorientierte Aspekte eines Dienstes (Ende-zu-Ende), Übertragung der Nutzdaten
- unterste Schicht (netzwerkabhängige QoS): Beschreibung der zugrundeliegenden Netzeigenschaften, der Topologie und des Transportmediums, also des Netzes

Ausgehend von diesen am Stack orientierten Ansätzen wird ersichtlich, warum die Orientierung vom Netz zum Nutzer von unten nach oben erfolgt, auch wenn der Nutzer im ISO/OSI-Modell nur als Element außerhalb der Beschreibung auftritt. Die Erweiterung des ISO/OSI-Modells nach oben⁴⁶ ist eine Variante, um nutzerrelevante Aspekte neben netzrelevanten Aspekten einzubeziehen. Für die Brücke zwischen QoS und QoE schlug [113] eine Ergänzung des OSI-Stacks um die menschlichen Faktoren vor, als Human Factors Extension, siehe Abbildung 14c. Die nutzerbezogenen Aspekte der HCI werden darin in drei zusätzlichen Schichten abgebildet. Noch recht techniknah sind dabei die Aspekte der Eingabe und Ausgabe an der Schnittstelle zwischen Mensch

⁴⁵ International Organization for Standardization (ISO) Standard 7498 Open Systems Interconnection (OSI) Model, kurz OSI-Modell

⁴⁶ Andere scherzhafte „Erweiterungen“ sehen oberhalb der Anwendungsschicht: financial, political und religious layer... (Anekdote aus einer Vorlesung)

und Maschine. Darüber angesiedelt sind die Wahrnehmungs- und Verarbeitungsprozesse im Menschen. An oberster Stelle sind die menschlichen Bedürfnisse und Anforderungen bei der Nutzung einzuordnen. (nach [113])

Das „QoE Hourglass Model“ [114] S. 25ff beschreibt einen ähnlichen Ansatz, allerdings mit Orientierung am DoD-Stack⁴⁷ und Erweiterung nach oben, siehe Abbildung 14d. Die Quality of Presentation (QoP) umfasst dabei die Darstellung der Leistungsaspekte und berücksichtigt den möglichen Einfluss der Software (Anwendung/App/Plattform) als auch der genutzten Hardware auf die wahrgenommene Leistung. Die Quality of Delivery (QoD) umfasst die transportorientierten Aspekte und weitere Verarbeitung auf der Applikationsschicht. Dabei berücksichtigt die QoD_A den möglichen Einfluss der Applikationsschicht mit Aspekten der Anwendung, die selbst nicht QoS-abhängig sind, die QoD_N hingegen umfasst den möglichen Einfluss des Netzes auf die Datenübertragung. (nach [114] S. 29f zusammengefasst)

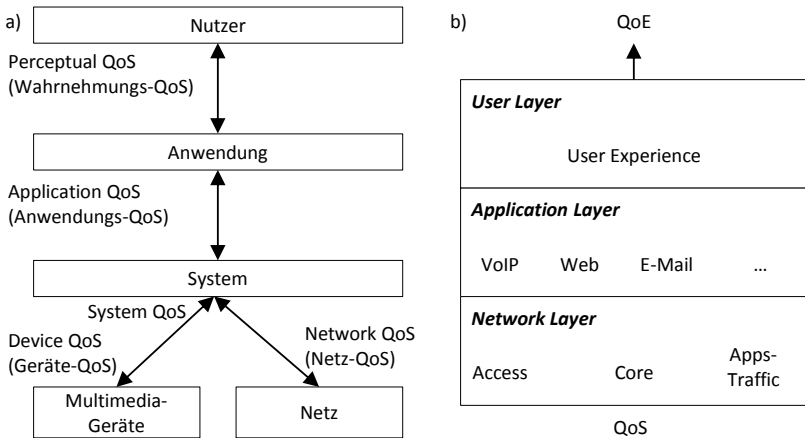


Abbildung 15: QoS-Schichtung (indirekte Stack-Orientierung)

a) QoS-Schichtmodell für vernetzte Multimedia-Systeme, nach [21] S. 238 bzw. [115] S. 17

b) zur Schichtung für QoE-Evaluation von Netzanwendungen, nach [116] S. 1353, vereinfachte Darstellung

⁴⁷ DoD: Department of Defense (Finanzierung der frühen Internet-Entwicklung in den USA). TCP/IP-Stack als äquivalenter Begriff in Orientierung an die wichtigsten Protokolle ist ebenso möglich, vergleiche 3.4.2 und Abbildung 25.

Nach der Orientierung am Netzwerk-Stack sollen auch noch Beispiele mit indirekter Orientierung bzw. abstrakter Schichtbildung kurz in die Diskussion eingebracht werden. Aus der QoS-Schichtung für vernetzte Multimedia-Systeme ergibt sich auch die Unterscheidung in Nutzer, Anwendung, System, Gerät und Netz [21] S. 238f, siehe Abbildung 15a, die über die gebildeten Schichten wechselwirken. Über die Schichtung von der Netzschicht über die Applikationsschicht zur Nutzerschicht, siehe Abbildung 15b, kann auch ein Framework zur QoE-Evaluation gebildet werden, bei der die einzelnen Schichten entsprechend isoliert bzw. im Zusammenwirken evaluiert werden können (siehe [117] und [116] sowie [118] für eine ausgewählte Anwendung). Dabei setzt QoS auf der Netzschicht an und die QoE entsteht aus der Nutzerschicht heraus.

2.5.2 Aus Ende-zu-Ende-QoS wird Ende-zu-Ende-QoE

Die Ende-zu-Ende-Betrachtung zur QoS durch das Netz reicht von einem Netzzugangspunkt bzw. Endgerät zum Endpunkt auf der anderen Seite. Endgerät und Gegenseite (als Server oder anderes Endgeräte je nach Kommunikationsszenario) können so auch mit in die Ende-zu-Ende-Betrachtung eingeschlossen werden, um die objektive Qualität eines mobilen Dienstes auf Geräteebene zu erfassen. Durch die Einbeziehung des Nutzers erweitert sich die Betrachtung über die Geräte hinaus. Die ITU-T E.800 bezeichnet dies noch als Ende-zu-Ende-QoS (im Sinne der holistischen QoS, siehe [35] S. 1). In der ITU-T E.804 deutlich jüngeren Datums ist die Ende-zu-Ende-QoS hingegen tatsächlich auf die Endgeräte beschränkt. Dabei ergibt die erweiterte Ende-zu-Ende-Sicht inklusive des Nutzers schließlich die QoE [44] S. 16. Eigentlich müsste diese erweiterte Betrachtung dann konsequenterweise auch als Ende-zu-Ende-QoE bezeichnet werden, siehe Abbildung 16. In Kommunikationsszenarien von Mensch zu Mensch (Person-to-Person) ist dies leicht zu verstehen, ist aber auch für den Zugriff auf Inhalte (Person-to-Content) sinnvoll. Durch die Ende-zu-Ende-Sicht wird dabei noch einmal verdeutlicht, dass eine isolierte Betrachtung eines einzelnen Endes der Kommunikation weder in der Theorie sinnvoll noch in der späteren empirischen Untersuchung möglich ist, sondern alle beteiligten Elemente in ihrem gemeinsamen Zusammenwirken zu berücksichtigen sind. Dieses Konzept ist damit vielleicht noch am anschaulichsten, aber zugleich auch entsprechend bedeutsam.

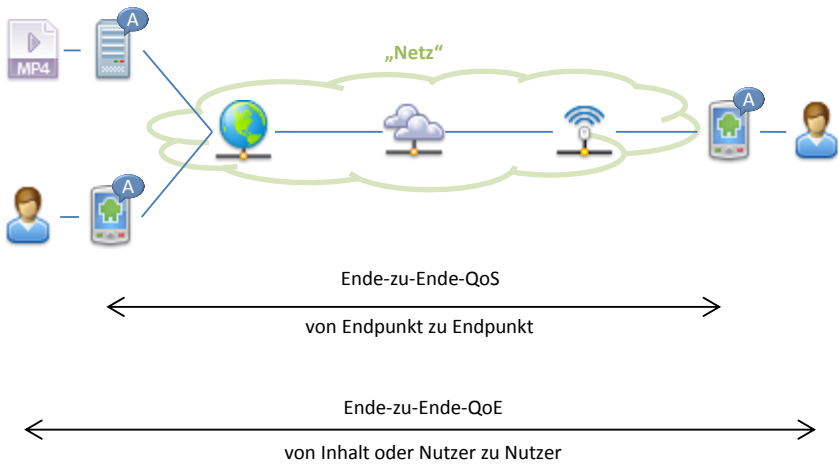


Abbildung 16: Von Ende-zu-Ende QoS zu Ende-zu-Ende-QoE

Erweiterung des Konzepts der Ende-zu-Ende-QoS von Endpunkt zu Endpunkt durch das Netz zur Ende-zu-Ende-QoE auf beiden Seiten der Kommunikationsbeziehung in Anlehnung an [44] S. 16, angepasst und erweitert um „Ende-zu-Ende“ für die QoE und unterschiedliche Kommunikationsszenarien

2.5.3 Verbindung von Netz und Nutzer über die QoS

Im Verständnis der ITU-T kann innerhalb der QoS zwischen der Seite der Interessen der Anbieter und der Seite der Nutzer unterschieden werden [35] S. 2f. Daraus ergeben sich mit *QoS offered*, *achieved*, *perceived* und *QoS requirements* vier Aspekte der QoS zum Zusammenwirken zwischen den Seiten [32] S. 4f, siehe Abbildung 17 Mitte. Wie bei der QoS als Quality of Something beschrieben, kann abweichend vom ITU-T-Verständnis oftmals davon ausgegangen werden, dass die QoS im Rahmen der Dienstleistung auf Aspekte der Netzleistung fokussiert. Somit sind Network Performance und QoS zusammen zu betrachten. Die Nutzerseite wiederum expandiert in einem erweiterten Verständnis über die rein technischen und objektiven Aspekte der netzleistungsabhängigen QoS-Wahrnehmung hinaus zur Quality of Experience.

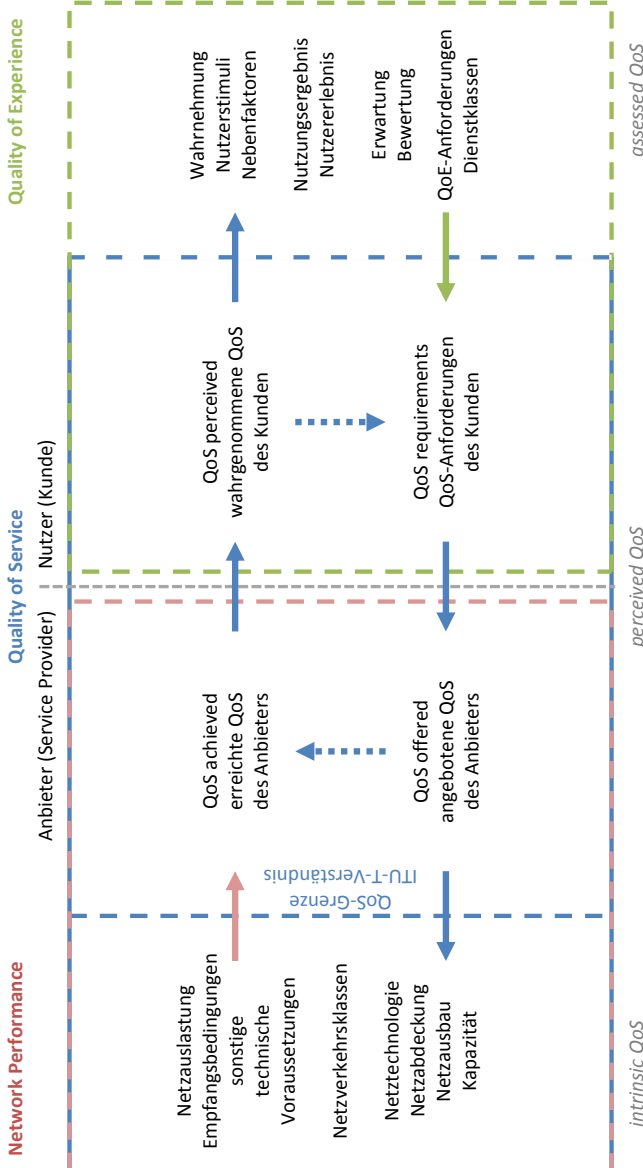


Abbildung 17: Zusammenhang von Network Performance, Quality of Service und Quality of Experience in Anlehnung an [32] S. 4 und [119] S. 149, ausgehend vom QoS-Verständnis der ITU-T G.1000 mit den vier Sichtweise zur QoS (in der Mitte), erweitert und ergänzt um Network Performance und Quality of Experience

Aus dieser auf beiden Seiten erweiterten Sichtweise ergibt sich die Basis für den konzeptuellen Zusammenhang zwischen dem Anbieter und Netz mit der QoS auf der einen Seite sowie dem Nutzer mit der QoE auf der anderen Seite. Dabei ist der konzeptuelle Zusammenhang grundsätzlich in beide Richtungen zu beschreiben:

- von QoS/Netz zu QoE/Nutzer: über die Wahrnehmung der (erreichten) Netzleistung durch die Nutzer, ggf. geprägt durch mögliche Einschränkungen im Nutzererlebnis
- von QoE/Nutzer zu QoS/Netz: über QoS-Anforderungen zur (notwendigen) Netzleistung, die sich aus QoE-Nutzeranforderungen ableiten

Ausgehend von einem generellen QoS-Verständnis werden teilweise die unterschiedlichen Auslegungen der QoS mit entsprechenden Adjektiven konkretisiert (siehe [120] S. 5ff aus einer Zeit vor der Prägung des QoE-Begriffs bzw. Zusammenfassung dazu in [119] S. 149). Dabei wird klar, dass die Konzepte der unterschiedlichen Qualitäten bei einer Dienstnutzung schlicht mit unterschiedlichen Begriffen belegt werden können, prinzipiell inhaltlich aber vergleichbar sind:

- assessed QoS: vom Nutzer bewertet (wie in der dargelegten Sicht zur QoE)
- perceived QoS: wahrgenommene QoS (wie bei der ITU-T Sicht zur QoS)
- intrinsic QoS: wesensmäßige QoS des Netzes (wie bei der ITU-T-Sicht zur Network Performance)

2.5.4 Wahrnehmung der Netzeigenschaften durch die Nutzer

Was die Nutzer genau erleben, also die Beschreibung des Nutzererlebnisses, wurde bislang ausgeklammert, ist aber für die weiteren Überlegungen essenziell. Wie zuvor dargelegt, wird über das Nutzererlebnis der Zusammenhang zwischen QoS und QoE hergestellt, sowohl konzeptuell als auch später in der Ursache-Wirkungs-Betrachtung. Für die Richtung von QoS zu QoE kann lohnen, noch einmal genauer zu betrachten, wie sich dabei die Netzleistung in der Wahrnehmung der Nutzer darstellt. Bei den Netzeigenschaften bzw. von ihnen abhängigen veränderlichen Größen ist eine Unterscheidung der unmittelbar wahrnehmbaren Eigenschaften und den nur mittelbar wahrnehmbaren Eigen-

schaften sinnvoll, um sich der Beantwortung der folgenden Fragen aus Nutzersicht zu nähern: *Wie sieht langsames Netz aus? Wie fühlt es sich an?* Dazu soll die Diskussion zur Wahrnehmung der Netz- bzw. Verbindungseigenschaften durch die Nutzer in drei Teilen erfolgen: unmittelbare Wahrnehmung, quantitative Ermittlung und mittelbare Wahrnehmung.

Unmittelbare Wahrnehmung: Nutzer können die Netzeigenschaften nicht unmittelbar und vollständig erfassen, sondern nur Indikatoren/Symbole auf ihren mobilen Geräten als Hinweise auf die Netzeigenschaften deuten, siehe Abbildung 18a. Diese Anzeigen sollen im Weiteren als Netzindikatoren bezeichnet werden. Aus der alleinigen Darstellung dieser Informationen ist kein direkter Rückschluss auf die genaue Netzleistung und damit die Netzqualität möglich.

"The signal bars don't always mean that much, and there's no standardized way to display that info." Gabriel Brown⁴⁸

Für die Veranschaulichung der unmittelbaren Wahrnehmung eignet sich eine Analogie mit dem mobilen Internet als mobiler Datenautobahn. Die Netzindikatoren sind ein bisschen wie der Blick auf den eingelegten Gang und das Gaspedal. Ein hoher Gang oder ein weit durchgedrücktes Gaspedal können eine hohe Geschwindigkeit bedeuten, müssen es aber nicht, z. B. wenn Anhängelast und Stellung der Bremse unbekannt sind – in Analogie zur Auslastung des Netzes bzw. dem langsamsten Element der gesamten Kette. Aus technischer Sicht sind die Netzindikatoren mit den interessierenden QoS-Parametern verknüpft, allerdings in unbekannter genauer Ausprägung. Aus der Netztechnologiestufe im Zugangnetz des Mobilfunknetzes, der (vermutlichen) Signalstärke und den Eigenschaften des genutzten Gerätes sind lediglich Abschätzungen zu theoretisch erreichbaren Bestwerten möglich. Diese Werte dürften den Nutzern aber überwiegend unbekannt sein – außer vielleicht in Teilen aus markigen Werbeaussagen.

⁴⁸ Senior Analyst bei Heavy Reading für mobile Kommunikationssysteme, Meinungsäußerung in http://www.lightreading.com/messages.asp?pidl_msgth-readid=56898&_mc=em_db_reg_lrelauchlivechat abgerufen am 11.09.2013



Abbildung 18: Zusammenstellung zur Nutzersicht auf das Netz und die Netzeigenschaften

a) Netzindikatoren in der Nutzersicht am Beispiel ausgewählter populärer mobiler Endgeräte (Screenshots von Android und iOS, Ausschnitte zur Verdeutlichung), mögliche Situation und Sicht zur Darstellung einer aktiven Datenverbindung im Mobilfunk (Beispiele)

Ausschnitte aus Screenshots vom oberen Bildschirmrand der genannten Mobilgeräte

b) Speedtest.net App als Beispiel eines „Netztachos“, Screenshot aus der zugehörigen App auf einem mobilen Gerät erstellt (Ausschnitt zur Verdeutlichung, zufällige Messung als Beispiel)

Es ist anzunehmen, dass zumindest größenordnungsmäßig deutliche Unterschiede (z. B. zwischen „E“ (EDGE) und „H+“ (HSPA+) oder 1 und 5 „Signalbalken“)⁴⁹ durch die Nutzer wahrgenommen und registriert werden. Plausibel ist ebenfalls, dass die Nutzer auf Basis ihres eigenen Erlebens früherer vergleichbarer Situationen auf eine bestimmte Netzleistung schließen. Diese früheren Erlebnisse (und damit auch die Erwartung) können aber natürlich individuell unterschiedlich sein. Aus Nutzersicht können die Netzindikatoren damit wie Erwartungsfaktoren wirken. Sie sind damit eine Art Prädiktor zur zu erwartenden Leistung und damit verbunden zum Nutzererlebnis. Beispiel: Funktioniert mit „3G“ das Videostreaming nicht richtig, so war das ja zu erwarten..., bei „LTE“ wäre die Enttäuschung im Fall von Problemen aber möglicherweise deutlich größer.

“...from the user's perspective, when they consider quality, for them latency, data rates, coverage, outages, degradations are difficult or impossible to distinguish...”

Patrick Donegan in [121]⁵⁰ zitiert nach [122] S. 18

Quantitative Ermittlung: Wieder in der Analogie der mobilen Datenautobahn kann die Ermittlung konkreter Werte aus Nutzersicht interpretiert werden. Die Effektivwerte der Kommunikationsleistung entsprechen den Anzeigen der Instrumente auf dem Armaturenbrett – nur statt in km/h z. B. in kBit/s, siehe Abbildung 18b. Im Unterschied zum Automobil ist eine Art Netztacho auf den mobilen Endgeräten aber nicht direkt eingebaut bzw. permanent sichtbar. Er kann aber in Form von Apps und Services (wie z. B. Speedtest⁵¹) nachgerüstet werden. Der Nutzer muss sich dann allerdings entscheiden, ob er zur Geschwindigkeitsmessung auf die Rennstrecke will, um QoS-Messungen auszuführen, oder lieber seiner Route zum Ziel folgt, und die Apps und Dienste nutzt, die ihn wirklich interessieren bzw. unterhalten und dabei sein Vorankommen begutachtet. Dies lenkt die Diskussion auf die mittelbare Wahrnehmung der Netzleistung als Applikationsleistung. Mit den nackten Zahlen aus

⁴⁹ Erläuterungen und Angaben zu den Netzgenerationen inklusive möglicher/typischer Werte in 5.1.2

⁵⁰ kostenpflichtiger Report, die relevanten Aussagen sind in der zitierten Kurzversion ebenfalls enthalten

⁵¹ <http://www.speedtest.net/mobile/> abgerufen am 06.01.2014, als ein Beispiel einer Vielzahl ähnlicher Services und Apps

den ermittelten Verbindungseigenschaften im Abgleich mit zeitnahen Beobachtungen zur Nutzung eines Dienstes kann sich der Nutzer höchstens eine grobe Orientierung schaffen. So lässt er ggf. von der Nutzung bestimmter Dienste in ungünstigen Netzsituationen ganz ab.

“User Experience: Perception is Reality”

Steven P. Shalita in [123] S. 28 zur QoE

Mittelbare Wahrnehmung: Über die unmittelbaren Anzeigen hinaus ist es die mittelbare Wahrnehmung der Applikationsleistung über verschiedene Arten von Stimuli, die einen indirekten Schluss auf die momentanen Netz- bzw. Verbindungseigenschaften zulassen. Dabei ist die individuelle Erfahrung wichtiger als Messwerte. In der Analogie der mobilen Datenautobahn sind die Nutzerstimuli so etwas wie der Fahrtwind, das Motorengeräusch und das mehr oder weniger schnelle Vorbeihuschen der Landschaft – sowie gelegentliche technische Probleme in Form von eigenartigem Verhalten (Fehlermeldungen). Analog zum schlimmsten Fall eines Unfalls sind dabei Verbindungsabbrüche und abgebrochene Nutzungsversuche zu sehen. Eine vereinfachte aber durchaus plausible Annahme aus Nutzersicht dazu ist: *Es fühlt sich langsam an? Es gibt Probleme? Dann ist wohl das Netz schuld ...* Entsprechend dieser Analogie ist auch ein möglicher Einfluss der unmittelbaren Wahrnehmung über die Erwartungshaltung auf die mittelbare Wahrnehmung plausibel (siehe o. g. Beispiel 3G vs. LTE beim Videostreaming).

Fazit: Die Wahrnehmung der Netzeigenschaften entsteht aus dem Zusammenspiel der direkt sichtbaren Eigenschaften wie Indikatoren, weiterer Nebenfaktoren und den Nutzerstimuli, welche als QoS-abhängig angenommen werden. Entsprechend ist es in dieser Sicht/Schicht nicht sinnvoll von technischen Kenngrößen tieferer Ebenen zu sprechen, sondern sich auf die unmittelbar wahrnehmbaren Qualitätsmerkmale zu konzentrieren, die tatsächlich nutzerrelevant sind. Diese können je nach Dienst bzw. ganzen Klassen von Diensten ganz unterschiedlich sein. Entsprechend ist dazu auch eine Unterscheidung sinnvoll bzw. die Nutzersicht auf das Netz generell mit der Nutzung der Dienste zu verbinden. Für die Gesamtbetrachtung spielt die mittelbare Wahrnehmung der Netz- und Verbindungseigenschaften über die Dienste damit die

entscheidende Rolle, was schließlich zum angestrebten dienstspezifischen Qualitätsabgleich von QoS und QoE führt, der im Anschluss diskutiert wird.

2.5.5 Von der Beschaffenheit zur Bewertung als Eignung zum Zweck

Wie eingangs angedeutet und im Laufe der Diskussionen zu QoS und QoE verdeutlicht, bezieht sich das Q der Qualität in QoS und QoE bei genauer Betrachtung auf unterschiedliche Bedeutungen. In QoS ist es oftmals nur die Feststellung einer Beschaffenheit mit bestimmten Qualitätskriterien und in QoE eine Bewertung als graduelles Qualitäturteil auf Basis einer Abwägung zwischen Erlebnis und Erwartung. Im Sinne der Qualität ist also nicht die Einschätzung gut/schlecht direkt mit QoS verbunden als vielmehr die Beschreibung dessen als wie viel, was dafür verantwortlich ist, dass das Erlebnis der Nutzer gute oder schlechte QoE ist. Für die gemeinsame Betrachtung der Qualitäten, zur Verbindung von Beschaffenheit und Bewertung, ist die Eignung zum angestrebten Zweck ein sinnvoller Ansatz. Ein und dieselbe Beschaffenheit kann je nach Zweck, aus dem sich die Eignung ableitet, zu verschiedenen Bewertungen führen (nach [29] S. 353 für die Qualität allgemein), siehe Abbildung 19 für ein Beispiel zur Qualität in QoS und QoE. Die Frage nach dem jeweiligen Zweck erlaubt die Einbeziehung der Dienste in die Diskussion und leitet zu den Überlegungen zur Rolle der Dienste im nächsten Kapitel über. Auch in der Literatur findet sich eine ähnliche Überlegung, ausgehend von der Person-to-Person-Kommunikation (Video) als „fitness-for-purpose“, siehe [124] S. 2ff mit Verweis auf [125], woraus sich der der IF-USING-WITH-THEN-Ansatz der ETSI zum QoS-QoE-Zusammenhang entwickelt hat (siehe [63] S. 45).

Die Verbindung der Bedeutungen der Qualität und ihre Anwendung in QoS und QoE führen in der gemeinsamen Betrachtung genau dahin, worauf die Zielstellung der Arbeit ausgerichtet ist. Die durch die QoS charakterisierte Beschaffenheit (Netz) gewinnt aus der QoE-Eignungsbewertung (Nutzer) für bestimmte Zwecke (Dienst) die gesuchte Bedeutung. Die Zuordnung von Wertbedeutung durch Angabe von Richtung (bis) und Stärke (mindestens) für die Messung/Bestimmung der Qualitätskriterien durch zugeordnete Qualitäturteile von gut (Anforderungen erfüllt) bis schlecht (Anforderungen nicht er-

füllt) ist dabei sinnvoll zur Unterscheidung bestimmter Wertbereiche der Beschaffenheit. Dies führt letztlich über den quantitativen Zusammenhang zu den gesuchten QoS-QoE-Schwellwerten.

Die gemeinsame Betrachtung von QoE und QoS kann somit auch als das Zusammentreffen von Anforderungen und Einschränkungen gesehen werden. Alternative Begriffe wären dabei „bedarfsorientierte Merkmale“ für die Anforderungen und schlicht „technische Merkmale“ für die Einschränkungen (z. B. in [126] S. 7 als Begriffe so genutzt). Für unterschiedliche Dienste entstehen aus der Abbildung der spezifischen Anforderungen aufeinander, über die verschiedenen Qualitätsschichten hinweg, unterschiedliche und somit dienstspezifische QoS-QoE-Zusammenhänge, je nach Rolle und konkreten Werte der Qualitätsmerkmale als Parameter. Das Ziel der Abbildung der Anforderungen ist die Feststellung bzw. Festlegung der Bedeutung der einzelnen QoS-Parameter für einen bestimmten Dienst. Die QoS-Anforderungen ergeben sich dabei aus den QoE-Anforderungen.

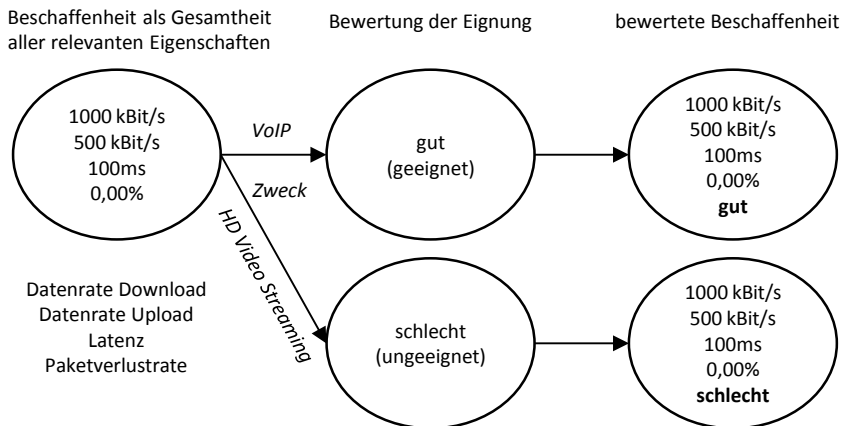


Abbildung 19: Qualität als Beschaffenheit, Beurteilung und Eignung zum Zweck im Zusammenhang von QoS und QoE

Der angestrebte Zweck bestimmt die Eignung, womit eine Beurteilung (QoE) der Beschaffenheit (QoS) möglich wird.

2.6 Zusammenfassung

QoS und QoE sind komplexe Konzepte im Anwendungsbereich der Nutzung multimedialer Dienste über mobile Kommunikationsnetze mit Verankerung in unterschiedlichen Begriffswelten. Bei einer Vielzahl schon existierender, jeweils leicht unterschiedlicher Definitionen soll zur Zusammenfassung nicht noch jeweils eine Ad-hoc-Definition hinzugefügt werden, sondern die Beschreibung in Form einer Charakteristik erfolgen, siehe Tabelle 5. In jedem Fall geht die QoE in der Betrachtung der Qualität über die rein transportorientierten Netzwerkeaspekte hinaus, während die QoS, abweichend von einstigen ganzheitlichen Ansätzen der Telkowitz, oftmals darauf reduziert werden kann. Dabei sind die unterschiedlichen Bedeutungen der Qualität zu beachten. Die QoS beschreibt als Ende-zu-Ende-QoS vorrangig die Beschaffenheit des Netzes und damit einen Teil der Qualitätskriterien zur QoE. Der QoE wiederum entstammen die zugehörigen Qualitätsurteile, die sich auf das Nutzererlebnis oder Nutzungsergebnis beziehen. Für die QoE spielen neben den technischen Einflussfaktoren (inklusive des Netzes) die diskutierten menschlichen und kontextuellen Einflussfaktoren eine entscheidende Rolle.

Der konzeptuelle Zusammenhang zwischen QoS und QoE lässt sich über die Schichtung der Qualität sowie eine Ende-zu-Ende-Sicht über das Netz hinaus beschreiben. Dabei spielt die Wahrnehmung der Netz- bzw. genauer der Verbindungseigenschaften eine wichtige Rolle. In der Sicht der Nutzer sind die Eigenschaften des Netzes aber nicht unmittelbar sinnvoll zu erfassen. Die mittelbare Wahrnehmung der Netzeigenschaften als wichtigster Teil des Nutzererlebnisses ist dienstspezifisch am besten zu vermitteln. Dabei kann der Zusammenhang von QoS und QoE als Abgleich zwischen der Qualität in QoS als Netzcharakteristik und der Qualität in QoE als Zufriedenheits-/Eignungsbeurteilung für die Nutzung eines Dienstes (Zweck) verstanden werden. Über die Diskussion der Dienste wird im nächsten Kapitel der QoS-QoE-Zusammenhang so formalisiert, dass von diesem Ausgangspunkt ein Ursache-Wirkungs-Zusammenhang konstruiert werden kann.

Tabelle 5: Charakterisierung von QoS und QoE zum Verständnis innerhalb der Arbeit

Kriterium	QoS	QoE
Anwendungsbereich	Netzwelten	Netzwelten/Nutzerwelt/Multimedialwelt
Perspektive	technische Sicht (System)	Nutzersicht
Ausgangspunkt	netzentriert	nutzerzentriert
Fokus auf	Diensterbringung	Dienstnutzung
wichtig für	Beschreibung der Anforderungen an das Netz und Feststellung der Einschränkungen durch das Netz	Beschreibung der Anforderungen der Nutzer und Auswirkungen von Einschränkungen
Erklärungsansatz	technikorientiert (Maschine)	ganzheitlicher Erklärungsansatz, interdisziplinär (Mensch, Maschine, Markt)
beeinflusst von	Netzleistung/-qualität	System (mit Netz), Mensch (Nutzer), Kontext
Qualität als	Beschaffenheit	Bewertung zu einer Eignung im kontextabhängigen Abgleich von Erlebnis und Erwartung
Beschreibung von	objektiver Leistung	subjektives Nutzererlebnis und objektives Nutzungsergebnis
Beschreibung mit	QoS-Netzparametern, z. B. Durchsatz und Latenz	QoE-Bewertungen, z. B. Zufriedenheit und Akzeptanz
erfasste Größen	pragmatisch (nutzenorientiert)	pragmatisch und hedonistisch

in Anlehnung an [43] S. 15f und [127] S. 87, erweitert und ergänzt

3 Das Modell

Der dienstspezifische Zusammenhang von QoS und QoE

In diesem Kapitel wird der Zusammenhang von QoS und QoE als Ursache-Wirkungs-Zusammenhang modelliert. Aus der Sicht der Nutzer auf das Netz aus dem vorherigen Kapitel sollte klargeworden sein, dass QoS und QoE nicht direkt und unmittelbar zusammenhängen, sondern nur mittelbar, bzw. in der Sprache des Modells über interne Variablen. Daher ist ein passendes Modell notwendig, das den inneren Zusammenhang von Ursache und Wirkung berücksichtigt und darüber auch quantitative Voraussagen erlaubt.

Dazu werden zunächst die Ansätze, Begriffe und Konzepte zur Modellierung vorgestellt. Dabei werden die relevanten Elemente und Zwischenschichten/-stufen eingeführt. Die Trennung in QoS-veränderliche interne Größen und weitere Einflussgrößen neben dem Netz spielt eine entscheidende Rolle. Aus den Überlegungen, *Was Wo Wie* zusammenwirkt, entsteht das Modell der QoX-Matrix.

Daran schließt sich die Diskussion der einzelnen Elemente an, die neben Netz und Nutzern (aus dem vorherigen Kapitel) einen entscheidenden Anteil haben, den vorab identifizierten konzeptuellen Zusammenhang in einen konkreten Ursache-Wirkungs-Zusammenhang zu verwandeln. Dazu wird das mobile Endgerät in die Diskussion eingebracht. Besonders intensiv werden die Dienste vorgestellt, getrennt in Netzsicht und Nutzersicht. Die gemeinsame Gesamtsicht auf die Dienste führt schließlich zu den Überlegungen zum dienstspezifischen QoS-QoE-Zusammenhang, der sich aus Anforderungen und Einschränkungen vorhersagen bzw. erklären lassen sollte. Dazu wird erläutert, wie sich die Rolle von Einflussfaktoren und die dienstspezifischen QoS-QoE-Zusammenhänge geeignet beschreiben lassen. Dies umfasst die Vorstellung üblicher Taxonomien für die Dienste und eine systematische Betrachtung zur Bildung sinnvoller Dienstklassen auf Basis der angenommenen QoS-QoE-Zusammenhänge. Für die Dienstklassen lässt sich jeweils eine spezifische Charakteristik des QoS-QoE-Zusammenhangs vorhersagen.

3.1 Die Modellierung der QoX-Matrix

Ein gutes Modell sollte einem bestimmten Zweck dienen und dafür Vereinfachungen komplexer Sachverhalte und Zusammenhänge bieten (zur Qualität von Modellen siehe [128] S. 224⁵²). So erfüllt auch das eigene Modell einen spezifischen Zweck. Es soll helfen, den QoS-QoE-Zusammenhang für die genutzten Dienste in den Vorüberlegungen zum Test zu berücksichtigen, Parameterbereiche einzugrenzen, Erwartungswerte zu bilden und die erzielten Ergebnisse erklären können.

Durch den angestrebten Zweck der Betrachtungen stehen Anfang als QoS (Netzparameter) und Ende als QoE (Bewertungen) fest. Entsprechend orientiert sich die Bezeichnung der Variablen als unabhängig oder abhängig daran, wie sie in der empirischen Überprüfung im Test wirken. Die Modellierung sieht die Nutzung der identifizierten Konzepte und Einflussgrößen zur Herstellung eines Ursache-Wirkungs-Zusammenhangs zwischen QoS-Netzparametrisierung und QoE-Bewertungen für unterschiedliche mobile Dienste vor. Dazu werden der gewählte Ansatz und die einzelnen Bestandteile diskutiert.

3.1.1 Andere Modelle und eigener Ansatz zur Modellbildung

Einen umfassenden und systematischen Vergleich von Modellen bzw. Erklärungsansätzen zur QoE unter Einbeziehung von QoS etc. liefert [111], ergänzt um [4]. Darin werden relevante Arbeiten der letzten Jahre hinsichtlich der Vollständigkeit und Berücksichtigung besonderer Aspekte verglichen. [111] integriert dazu auch drahtlose Verbindungen und unterschiedliche Diensttypen. [4] stellt einem kurzen Überblick zu anderen Modellen ein eigenes Modell eines holistischen QoE-Ökosystems gegenüber, in dem quasi alle wichtigen Bereiche (Mensch, Maschine und Kontext mit dem Markt) erfasst werden. Für den Überblick zum State-of-the-Art sei daher auf diese beiden Arbeiten verwiesen.

Das eigene Modell darf nicht zu kompliziert oder zu abstrakt sein. Es muss sich für konkrete Szenarien des Tests entsprechend operationalisieren lassen,

⁵² dort bezogen auf das modellbasierte Testen von Software, aber auch verallgemeinerbar

darf aber auch keinen relevanten Punkt ausschließen. Daher ist eine zu feingliedrige Struktur ebenso hinderlich wie ein zu diffuser Ansatz. Zu einfache oder zu spezielle Modelle scheiden demnach aus, die z. B. nur zur Modellierung der Videoqualität in Live-Streaming-Szenarien geeignet sind. Holistische Modelle sind ebenfalls nur bedingt als Basis für o. g. Zwecke geeignet. Ist der interessierende Zusammenhang nur ein Teilaspekt eines Welt-Erklär-Modells, ist er womöglich nur schwer operationalisierbar. Der Aspekt der QoS als Einflussgröße und damit verbunden des Netzes auf die QoE ist dabei in einer holistischen Gesamtbetrachtung eben nur einer von vielen (abstrakten) Einflussfaktoren. Für das eigene Modell als Erklärungsansatz für die eigenen Untersuchungen standen im Prinzip aber einige Punkte für die Überlegungen bereits fest. Für den QoS-QoE-Zusammenhang soll die QoS an den Anfang von Ursache-Wirkungs-Betrachtungen gestellt werden, die QoE ans Ende. Die relevanten Bestandteile zur Untersuchung waren ebenso vorausgewählt (Netz, Endgerät, unterschiedliche Dienste und die Nutzer).

In der eigenen Recherche sind zu Arbeiten zur expliziten Entwicklung von QoS/QoE-Modellen auch Ad-hoc-Modelle aufgetaucht, die jeweils einzelne Aspekte zur Erklärung ganz spezieller Probleme oder Szenarien fokussieren und dabei teilweise aber auch gute und verallgemeinerbare Einzelideen hervorbringen. Mit Einzelaspekten aus etwas mehr als einem Dutzend einzelner Quellen ist die Recherche dabei sicherlich noch nicht einmal ansatzweise erschöpfend. Es sollte allerdings auch klar sein, dass nicht ein Dutzend völlig unterschiedlicher Ansätze und Modelle zum Vorschein kamen, sondern dass sich einige Punkte herauskristallisierten. Ein Teil davon wurde schon auf die grundlegenden Konzepte zum QoS-QoE-Zusammenhang zurückgeführt und dort vorgestellt.

Das eigene Modell ist keine völlige Neuschöpfung, sondern greift die jeweils passenden Konzepte und Einzelaspekte heraus und kombiniert diese für den o. g. Zweck. Die Modellierung schließt damit an die vorgestellten Konzepte aus dem Abschluss des vorherigen Kapitels an. Dafür werden die einzelnen Konzepte so miteinander verknüpft, dass letztlich das eigene QoS-Matrix-Modell entsteht. In der Diskussion der Bestandteile des eigenen Modells erfolgen dann detaillierte Beschreibungen zu den als relevant erachteten Aspekten.

3.1.2 Qualitätsbegriffe: Von Elementen, Merkmalen und Co.

Für die weiteren Überlegungen ist eine genaue sprachliche Abgrenzung notwendig, was mit einem bestimmten Begriff zur Qualität gemeint ist. Die nachfolgende Anwendung dieser Begriffe dürfte diese anschaulicher erscheinen lassen. Die dazu gewählten Begriffe zur Qualität orientieren sich an den Definitionen in [129] S.1002 (sowie [30] S. 7) und an [30] S. 12 mit leichten Anpassungen, weshalb sie nicht genau äquivalent⁵³ sind:

- Qualitätselement: physikalischer oder algorithmischer Bestandteil, der zur Qualität beiträgt (z. B. Endgerät oder App)
- Qualitätsmerkmal: messbare oder wahrnehmbare bzw. zumindest benennbare Eigenschaft der Qualität
- Qualitätsaspekt: Kategorie der Qualität, die ein oder mehrere Merkmale umfasst
- Qualitätsfaktor: Kategorie von qualitätsrelevanten Eigenschaften, die ein oder mehrere Elemente umfasst (z. B. die Präsentation als gebildeter Faktor oder System, Kontext und Mensch als vorgegebene Faktoren)

Unterschiedliche Schichten von Qualität sind ebenso sinnvoll wie Sichten auf das Modell. Entsprechend werden die folgenden Begriffe im eigenen Modell hinzugefügt:

- Qualitätsschicht: abstrahiert zwischen Ursache und Wirkung, entsteht aus dem Zusammenwirken der beteiligten Elemente und ggf. einer weiteren Qualitätsschicht
- Qualitätssicht: beschränkt die Betrachtung auf die aus der gewählten Perspektive relevanten Aspekte und Faktoren (z. B. Netzsicht und Nutzersicht)

Die Qualitätsschichten sind in der Folge besondere Qualitätsaspekte, die jeweils voneinander abhängen.

Beispiel: Die Qualitätsaspekte der Schicht *QoS* können mit den QoS-Parametern als Qualitätsmerkmale charakterisiert werden und ergeben sich aus dem Qualitätselement *Netz* (Ende-zu-Ende) als Anfang der Betrachtungen. Die

⁵³ dort ausgehend von der Qualität zur Sprachkommunikation definiert, hier allgemein für mobile die Nutzung mobiler Dienste in mobilen Netzen

Qualitätsaspekte der Schicht *QoA* können mit den Nutzerstimuli als Qualitätsmerkmale charakterisiert werden und ergeben sich aus der Schicht *QoD* und dem Einwirken der Elemente Endgerät, App und Inhalt. (Erläuterung zu *QoA* und *QoD* finden sich nachfolgend im Abschnitt 3.1.4.)

Die Sichten können hingegen im Grunde beliebig gewählt werden. Für die eigene Betrachtung sollen die Sichten so gewählt werden, dass zwischen Netzsicht (bzw. technischer Sicht, von unten bzw. von innen) und Nutzersicht (von oben bzw. außen) zu unterscheiden ist.

Die Qualitätsfaktoren fassen die Qualitätselemente nach bestimmten Gesichtspunkten zusammen und bilden so die Kategorien bzw. Bereiche der Qualitätsmerkmale. Dazu können die Kategorien prinzipiell beliebig definiert werden. Prinzipiell lassen sich auch einzelne Elemente als separate Faktoren definieren. Bei geeigneter Kombination lassen sich Hierarchien bilden, ebenso ist aber auch die Nutzung als Facetten möglich (wie in 2.3.4 gezeigt). Dienst, Endgerät und Netz wiederum bilden gemeinsam die Kategorie der technischen bzw. systemischen Faktoren (System), diese werden um Kontext und Nutzer als weitere Qualitätsfaktoren ergänzt.

3.1.3 Die Elemente des Modells

Das Modell entsteht aus dem Zusammenwirken der beteiligten Elemente. Die Qualitätselemente weisen bestimmte Eigenschaften (Merkmale) auf, die für ihre Wirkung im Modell relevant sind. Die Wechselwirkung der Elemente und ihrer Merkmale erklärt die innere Funktion des Modells.

Für die Nutzung des mobilen Internets sind als relevante Qualitätselemente zu berücksichtigen:

- **Nutzer** (vereinfacht bestehend aus Auge, Hand und Hirn)
- **Endgerät** (bestehend aus Hardware (physikalisch) und Betriebs-Software (algorithmisch))
- **Dienst** (als Konstrukt aus App, Inhalt und Gegenseite)
 - **App** (Software (lokal) zur Nutzung eines Dienstes mit einer bestimmten Usability etc.)
 - **Inhalt** (die Nutzdaten des Dienstes mit einer bestimmten Qualität und Quantität)

- **Gegenseite** (zur Ende-zu-Ende-Kommunikation: ein anderer Nutzer oder eine Maschine)
- **Netz** (zur Übertragung des Inhalts zwischen Endgerät/App und Gegenseite)
- **Kontext** (in verschiedenen Ausprägungen)

Dienst und Kontext sind spezielle Konzepte bzw. Konstrukte, die sich selbst wieder auf andere Elemente beziehen und jeweils eine Art Metaelement in den Betrachtungen bilden. Im eigenen Modell verbindet der Kontext die Seite System/Maschine mit Netz, Endgerät und Dienst mit der Seite des Menschen als Nutzer.

Die einzelnen Elemente können selbst wiederum aus Einzelementen bestehen, z. B. beim Netz aus Netzelementen oder Netzsegmenten, oder beim Endgerät aus Hardware und zugehöriger Software. Sie sind damit zweckmäßig und zielführend entweder abstrakt oder konkret zu nutzen (z. B. „das Endgerät“ vs. „Apple iPhone 5 32 GB mit iOS 6.1.3“).

Die Auflistung der Elemente ergibt sich im Prinzip schon aus praktischen Gesichtspunkten und welche Bestandteile bei der Nutzung mobiler Dienste im mobilen Internet tatsächlich variieren können. Dabei decken sich die Überlegungen weitgehend zu den als relevant erachteten Bestandteilen, wie bei den Konzepten zu QoS und QoE erläutert (siehe z. B. Abbildung 9 und Abbildung 15a, wengleich dort in unterschiedlichen Schichten). [130] unternimmt eine Unterscheidung in verschiedene Parameterräume, von denen jeder wiederum mehrere Dimensionen aufweisen kann, die zusammen als ARCU definiert sind (siehe [130] S. 665):

- Application Space (Eigenschaften der genutzten Anwendungen, was auch Aspekte des Inhalts umfasst)
- Ressource Space (hier sind als Ressourcen die QoS-Parameter des Netzes, Endgerät und Gegenseite als Server einzuordnen)
- Context Space (räumlich, zeitlich, sozial, Kosten, Aufgabe etc.)
- User Space (Merkmale des Nutzers, Nutzertyp, Anforderungen, Erwartungen, Motivation)

Diese Aufteilung enthält im Prinzip die gleichen Einzelbestandteile, wenn auch etwas anders zusammengefasst. Für die eigene Modellbildung erschien die Bündelung von App, Inhalt und Gegenstelle als Dienst vorteilhaft. Dies soll lediglich verdeutlichen, dass die gewählte Aufteilung in Bestandteile bzw. Elemente nur eine von vielen möglichen Varianten ist.

3.1.4 Die Schichten des Modells

Das eigene Modell nutzt das Konzept der Schichtung der Qualität von QoS zu QoE wie im konzeptuellen Zusammenhang vorgestellt. Die Schichten abstrahieren zwischen Ursache und Wirkung und umfassen dabei das Zusammenwirken der beteiligten Qualitätselemente (App, Inhalt, Netz, Nutzer, ...). Für das eigene Modell sollen die Schichten durch die zu unterscheidenden Qualitätsaspekte bei der Nutzung eines Dienstes im mobilen Internet definiert und abgegrenzt werden. Diese bedeutet eine Orientierung am Stack-Modell aber keine 1-zu-1-Abbildung oder eindimensionale Erweiterung.

Am Anfang, auf der untersten Schicht, sollen die interessierenden QoS-Netzparameter wiederzufinden sein, von denen die Schichten darüber abhängen. Am Ende, auf der obersten Schicht, sind die subjektiven und objektiven QoE-Maße angesiedelt. Für die weitere Betrachtung werden dazu nur die subjektiven QoE-Bewertungen des Nutzererlebnisses berücksichtigt. Die Größen dieser Schichten stehen in keinen direkten Zusammenhang zwischen QoS und QoE, es bedarf also der Zwischenschichten für die Herstellung des dienstspezifischen QoS-QoE-Zusammenhangs.

Für die Festlegung der Schichten soll gelten:

- Abgrenzung/Übergang nach unten, der Input: auf diese Schicht bezogen unabhängige Variablen (Merkmale), die aber selbst von der Schicht unterhalb abhängig sind
- zwischen den Schichtübergängen, der interne Zusammenhang: Zusammenwirken relevanter Elemente, die aus dem Input der unabhängigen Variablen von unten und weiterer Variablen auf dieser Schicht die abhängigen Variablen nach oben bilden
- Abgrenzung/Übergang nach oben, der Output: die von dieser Schicht abhängigen Variablen als Qualitätsmerkmale

Für höhere Schichten soll damit gelten, dass zumindest einige Qualitätsmerkmale von anderen Qualitätsmerkmalen der darunterliegenden Schicht abhängig sind. Deshalb ist es sinnvoll, ein Modell als Ursache-Wirkungs-Kette zu konstruieren, das diesen Zusammenhang zwischen den genannten Endpunkten herstellt. Dabei ist ein möglichst einfaches aber doch universelles Modell vorteilhaft, d. h. nur so viele Zwischenschichten wie nötig.

Für das eigene Modell werden vier Schichten im QoS-QoE-Zusammenhang als Kategorien der Qualität vorgeschlagen, siehe Tabelle 6. Die Quality of Application (Performance), QoA, entspricht ungefähr der Definition des DSL-Forums zur (technischen) „QoE“ [40] S. 7, die Quality of Delivery, QoD, orientiert sich am entsprechenden Begriff des QoE-Hourglass-Modells [114] S. 25ff, ist allerdings keine direkte Entsprechung. Auf Unterschiede in der eigenen Nutzung der Begriffe wird mit der nachfolgenden Diskussion zu Mediatoren und Moderatoren noch einmal zurückgekommen.

Erst aus der Quality of Experience als Beurteilung heraus wird für die anderen Schichten eine Eignung zum Zweck festgestellt. Ohne diese Bewertung steht das *Q* in den anderen Schichten lediglich für die instrumentell und objektiv messbare Beschaffenheit im Sinne der Charakterisierung der geplanten oder erreichten Parameterwerte.

Aus dem Zusammenfügen der Schichten ergibt sich die Gesamtsicht auf das Zusammenwirken der Elemente und den Zusammenhang der Schichten untereinander, siehe Abbildung 20. QoS und QoE wurden bereits im Kapitel 2 thematisiert. Die Beschreibung der Zwischenschichten QoD und QoA ergibt sich aus den Anmerkungen zum Einfluss der Qualitätsmerkmale auf diesen Schichten und aus den weiteren Besreibungen zur Netzsicht (QoD) bzw. Nutzersicht (QoA).

Tabelle 6: Charakterisierung der Schichten des QoS-QoE-Zusammenhangs

Schicht	Bedeutung / Interpretation	bezogen auf: Leistung / Ergebnis	Beschreibung mit: Maße / Merkmale	alternative Zuordnung⁵⁴
QoE	nutzerzentrierte bewertete Qualität	beurteiltes Nutzererlebnis und Nutzungsergebnis (als eine Art Nutzerleistung und somit auch des Netzes)	QoE-Maße (Bewertungen) subjektiv: z. B. Zufriedenheit objektiv: z. B. Effizienz, Effektivität	UX in der HCI
QoA	dienst- und nutzerrelevante wahrnehmbare Qualität	Applikationsleistung des genutzten Dienstes (und somit auch des Netzes)	Nutzerstimuli: Qualität und Quantität Zeiten und Dauern Funktion und Verhalten	A-QoS technische "QoE" QoS-perceived/experienced
QoD	dienstrelevante transportorientierte Qualität	Kommunikationsleistung Ende-zu-Ende des Dienstes (und somit auch des Netzes)	erreichte Effektivwerte auf verschiedenen Schichten des Netzwerk-Stacks, z. B. mit QoS-Parameter	K-QoS Ende-zu-Ende-QoS QoS-delivered
QoS	angebotene bzw. verfügbare netzwerkabhängige Qualität	Ende-zu-Ende-Sicht der nutzerrelevanten Leistung des Netzes am nutzerseitigen Zugangspunkt	QoS-Netzparameter als Leistungsparameter mit theoretischen Bestwerten (Best Effort) oder als garantierte Werte mit QoS-Mechanismen	N-QoS QoS-offered

(U-QoS: User-QoS, A-QoS: anwendungsorientierte QoS, K-QoS: kommunikationsspezifische QoS, N-QoS: netzwerkabhängige QoS)

⁵⁴ z. B. Begriffe nach [45] S. 14ff

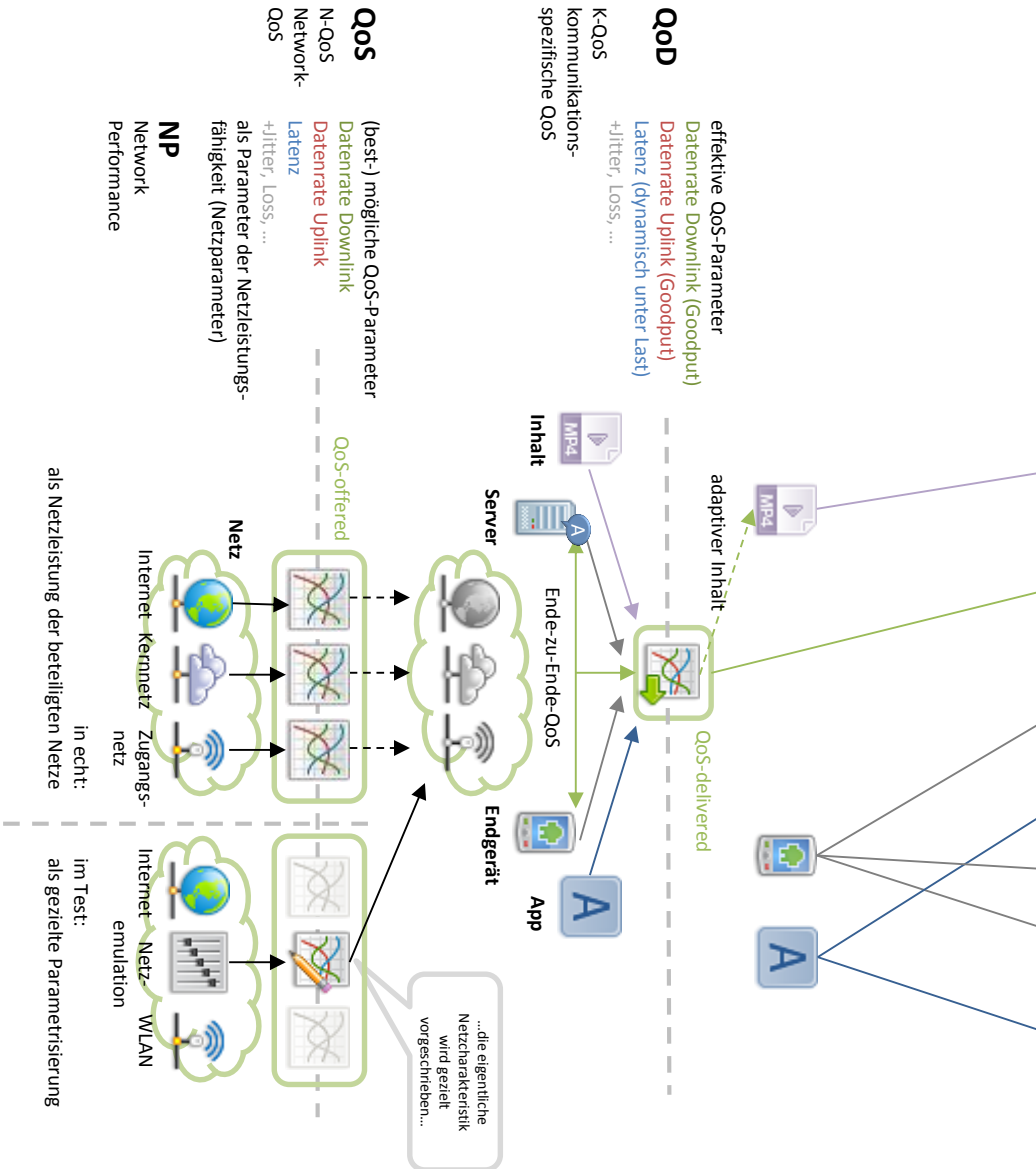
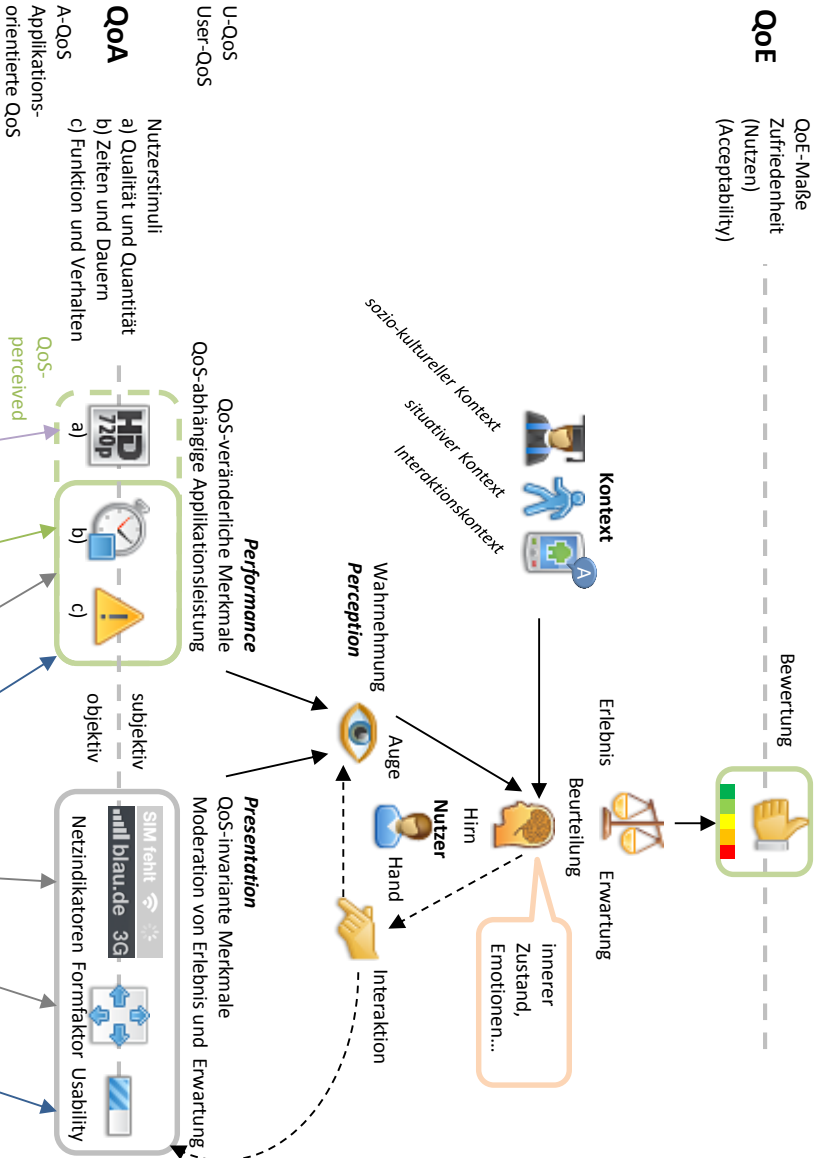


Abbildung 20: Das Modell zum Zusammenwirken der Elemente, aufgetrennt in die gebildeten Schichten Durch die Auftrennung der Darstellung in einzelne Sichten entsprechend der Schichten wiederholen sich einige Elemente. Die Schichten sind horizontal übereinander angeordnet, in der Ursache-Wirkungs-Beziehung von unten (QoS) nach oben (QoE) zu lesen. Fortsetzung →



Fortsetzung der Abbildung 20, Abbildungsbeschriftung nebenstehend

3.1.5 Das Prinzip der Mediatoren und Moderatoren

Zur gezielten Untersuchung der Wirkung der QoS-Charakteristik (Netzparameter) auf die QoE-Bewertungen (Nutzerzufriedenheit) ist es sinnvoll, bei der weiteren Betrachtung in Mediatoren und Moderatoren zu unterscheiden. Durch die Trennung in Mediatoren und Moderatoren wird die Modellierung des grundlegenden Zusammenhangs zwischen QoS-Netzparametern und QoE-Bewertungen verschlankt und vereinfacht, ist aber konkret genug, um die interessierende Wirkung zu erklären und flexibel genug um weitere Einflussfaktoren und Zusammenhänge im Modell gezielt berücksichtigen zu können.

Die Mediatoren vermitteln QoS-Änderungen als interne Variablen mit eigenen Wertänderungen weiter. Ein Mediator ist dabei kein rein theoretisches Konstrukt, sondern manifest, er kann mit geeigneten (technischen) Mitteln und Methoden erfasst werden. Die erreichten Effektivwerte der Kommunikationsleistung (QoD) sind auf dem Endgerät oder im Netzwerk aus den Verkehrsdaten erfassbar, die Nutzerstimuli der Applikationsleistung (QoA) aus den Ausgaben (Bild/Ton) der Endgeräte. Die Mediatoren als interne Variablen sind abhängige Variablen ihrer eigenen unabhängigen Variablen und selbst wieder unabhängige Variablen für die von ihnen abhängigen Variablen.

Die Moderatoren ergeben sich aus den beteiligten Elementen neben dem Netz. Sie wirken also als (Neben-) Faktoren in der Form von externen Variablen auf den Zusammenhang der QoS-veränderlichen Größen ein. Sie können dabei den QoS-QoE-Zusammenhang moderieren, ohne selbst verändert zu werden bzw. von den QoS-Netzparametern abhängig sein zu müssen. Die Benutzerschnittstelle einer App, die Sichtbarkeit bestimmter Nutzerstimuli oder auch die Hardwareeigenschaften wie Displaygröße etc. ändern sich ja nicht, nur weil die QoS-Parameter sich ändern. Für andere Einflussgrößen ist anzunehmen, dass diese QoS-adaptiv sein können, aber nicht müssen (z. B. statischer Inhalt vs. adaptiver Inhalt). So ist die Einordnung einzelner Faktoren mitunter nicht immer ganz eindeutig. Unter der Annahme, dass es grundsätzlich unterschiedliche Charakteristiken des QoS-QoE-Zusammenhangs gibt, wäre es möglich, dass die Nebenfaktoren diese Charakteristik nur auf eine bestimmte Art ändern, sie z. B. um einen konstanten Wert verschieben oder strecken. Die Nebenfaktoren sind damit auch ein Ansatzpunkt für eine Partitionierung in der Auswertung, siehe entsprechender Abschnitt in den

zugehörigen Überlegungen. Für eine gezielte Untersuchung des QoS-QoE-Zusammenhangs ist es vorteilhaft, möglichst gut zwischen diesen Faktoren zu trennen.

Zur Notation:

$A \rightarrow B$ A beeinflusst B; B ist also von A abhängig;
A ist die unabhängige Variable,
B ist die abhängige Variable

$A \rightarrow C \rightarrow B$ C ist Mediator für die Wirkung von A auf B

$\begin{matrix} D \\ \downarrow \\ A \end{matrix} \rightarrow B$ D ist Moderator für die Wirkung von A auf B

Die Nutzung von Mediatoren und Moderatoren kann beliebig erweitert und kombiniert werden, woraus im nächsten Schritt die QoS-QoE-Kette konstruiert wird.

Die Idee zu einem ähnlichen Konzept mit Mediation zwischen verschiedenen Einflussfaktoren und einer Moderation der Kausalbeziehung zur QoE sieht ebenfalls [4]⁵⁵ S. 61 vor. Im Unterschied zum eigenen Modell allerdings weniger stark formalisiert, d. h. es wird nicht von einer (mehrstufigen) gerichteten Ursache-Wirkungs-Kette ausgegangen und die Moderatoren beziehen sich hauptsächlich auf Kontext und Nutzer (z. B. das Alter etc.).

3.1.6 QoS-QoE-Ursache-Wirkungs-Kette zwischen den Schichten

Die Betrachtung auf den einzelnen Schichten ist in sich stimmig und noch gut überschaubar. In der Betrachtung über die Schichten hinweg wird es eher unübersichtlich, da es Qualitätselemente gibt, die auf mehreren Schichten einwirken. Die Nutzung des vorgestellten Prinzips mit Mediator und Moderator erscheint daher vorteilhaft. Aus dem Gesamtzusammenhang soll zunächst der vereinfachte Zusammenhang der veränderlichen Größen isoliert werden.

⁵⁵ wie auch die zugehörige Dissertation des Hauptautors [64]

Für die Beziehung von QoS und QoE bietet sich die Konstruktion einer Ursache-Wirkungs-Kette zur gezielten Erklärung des Einflusses der QoS-Netzparameter auf die QoE-Nutzerzufriedenheit an (Top-Down-Betrachtung vom Nutzer aus). Durch das Einfügen der Mediatoren wird der QoS-QoE-Zusammenhang zur QoX-Kette erweitert (mit dem X als Platzhalter der verschiedenen Qualitätsschichten), siehe Abbildung 21 zu den Quadranten:

- 1) Die von den Nutzern erzeugten QoE-Bewertungen hängen von den QoS-Netzparametern ab, allerdings nicht direkt. Dies ist der gesuchte Zusammenhang, der nachfolgend hergeleitet wird.
- 2) Die von den Nutzern erzeugten QoE-Bewertungen hängen direkt von der erreichten Applikationsleistung des genutzten Dienstes als QoA-Nutzerstimuli und weiteren Einflussfaktoren ab.
- 3) Die erreichte dienstspezifische Applikationsleistung der QoA hängt direkt von der Kommunikationsleistung als QoD-Effektivwerte und weiteren Einflussfaktoren ab.
- 4) Die erreichten QoD-Effektivwerte hängen direkt von den QoS-Netzparametern und weiteren Einflussfaktoren ab.

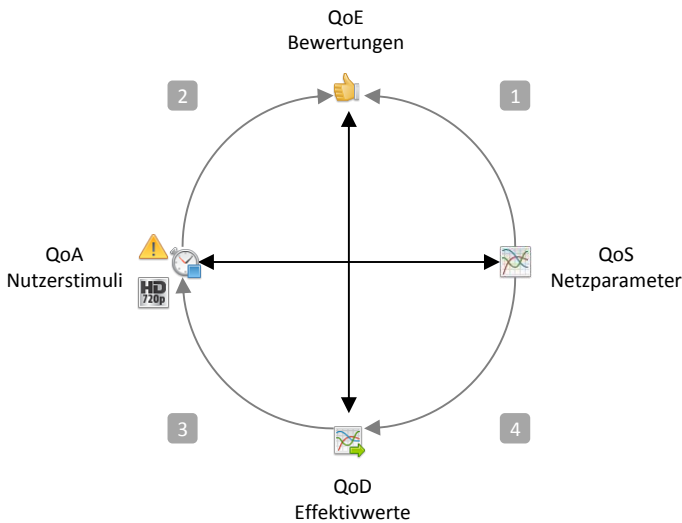


Abbildung 21: Erzeugung des QoS-QoE-Zusammenhangs
zur besseren Anschaulichkeit der Transitivität sind die vier Achsen orthogonal dargestellt

Sind die Zusammenhänge in 2) bis 4) korrekt, dann gilt zusammengefasst auch der Zusammenhang in 1) als hergestellt (Transitivität). Die von den Nutzern erzeugten QoE-Bewertungen hängen somit indirekt, also mittelbar, von den QoS-Netzparametern und weiteren Nebenfaktoren ab. Damit eignet sich dieser Ansatz zur Untersuchung des interessierenden Zusammenhangs zwischen QoS und QoE.

Aus $QoS \xrightarrow{\downarrow} QoD \xrightarrow{\downarrow} QoA \xrightarrow{\downarrow} QoE$ wird kurz $QoS \rightarrow QoD \rightarrow QoA \rightarrow QoE$ woraus $QoS \rightarrow QoE$ folgt.

Prinzipiell führen auch verkettete Funktionsterme wie beispielsweise „ $QoE = h(f(g(QoS)|_{QoDa})|_{QoPo})$ “ (aus der Modellierung in [114] S. 32) zu einer ähnlichen Beschreibung des Zusammenhangs zwischen QoS und QoE, sind aber weniger leicht handhabbar. Im Unterschied zum Verständnis des QoE-Hourglass-Modells in [114] definieren die eigenen Betrachtungen die QoD als Mediator. Die Quality of Presentation ist im eigenen Modell keine eigene Schicht, da kein Mediator, sondern ein Moderator. Die Trennung von Mediatoren und Moderatoren ist im eigenen Modell konsequent und so in der weiteren Betrachtung eine deutliche Vereinfachung.

3.1.7 Von der Kette zur Matrix

Nach den QoS-veränderlichen Größen sollen Qualitätselemente mit ihrem Einfluss auf die verschiedenen Schichten bzw. Glieder der Kette wieder in die Betrachtungen einbezogen werden. Dabei wird eine Systematisierung und Vereinfachung in der Darstellung der Zusammenhänge und möglicher Einflüsse angestrebt.

Die QoS-QoE-Kette ist schlank und vermittelt nur den grundlegenden kausalen Zusammenhang über wenige QoS-abhängige Mediatoren. Orthogonal zur QoS-QoE-Kette stehen die Nebenfaktoren als Moderatoren. So kann das Zusammenwirken der beteiligten Elemente entlang der QoS-QoE-Kette genau zugeordnet werden. Damit wird aus der eindimensionalen QoS-QoE-Kette quasi die zweidimensionale QoX-Matrix, die erläutert, *Was Wo Wie* einwirkt, als eine Art Adressierung in einer Einflussmatrix, siehe auch Abbildung 22.

Die Darstellung als Matrix bzw. Gitter ist damit eine Vereinfachung, in der die beteiligten Elemente jeweils nur noch einmal auftauchen und mit den gebildeten Schichten gezielt in Zusammenhang gebracht werden können. In der interner Erklärung und Funktion ist eine Äquivalenz zur Schichtdarstellung gegeben, die sich auf einzelne Teilaspekte oder den Übergang von einer Schicht zur nächsten konzentrieren kann.

Bildlich gesprochen sind die Sichten mögliche und sinnvolle Schnitte durch das Modell. So können auch verschiedene Sichten parallel genutzt werden. Technische Netzsicht und Nutzersicht sind leicht nachzuvollziehen. So geht [95] beispielsweise von einer Unterscheidung möglicher Einflüsse auf die QoE von direkt (z. B. App und Endgerät) und indirekt (z. B. Netz) aus. Auch diese Unterscheidung ist im eigenen Modell recht einfach abzubilden. Das Netz wirkt entlang der QoX-Kette indirekt, das Gerät beispielsweise mit einigen Merkmalen (Display etc. als QoS-unveränderliche Größe) hingegen auch direkt auf die QoE.

Wie in Abbildung 22 zum QoX-Matrix-Modell skizziert, sind Überlappungen in den Betrachtungen zwischen Interaktionskontext und System sowie zwischen soziokulturellen Kontext und Mensch festzuhalten. Die Dienste bestehen im eigenen Modell aus technischen Subelementen und werden auch so eingeordnet. Im Kontext sind über die Nutzung der Dienste mit Nutzungsprozess und Nutzungszielen die Anknüpfungspunkte zwischen Technik und Nutzer zu finden.

Erläuterungen zur Abbildung 22:

Die QoX-Matrix bildet das Ausgangs- und Erklärungsmodell für die theoretische Diskussion und empirische Untersuchung.

- unten: Ausgangsbasis der QoS-QoE-Kette mit den Zwischenschichten (Mediatoren)
- darüber: Einflussmatrix der verschiedenen Faktoren
- Sichten: von links aus Sicht des Systems (technische Faktoren), von rechts aus Sicht des Menschen, von oben aus Sicht des Kontexts, der Technik und Mensch in der Nutzung einbettet
- Zeilen der Matrix: beteiligte Qualitätselemente (am linken oder rechten Rand)
- horizontale Pfeile in den Zellen: relevante Eigenschaften der beteiligten Elemente
- Spalten der Matrix (senkrechte Pfeile): Einfluss auf die QoS-QoE-Kette als Moderatoren

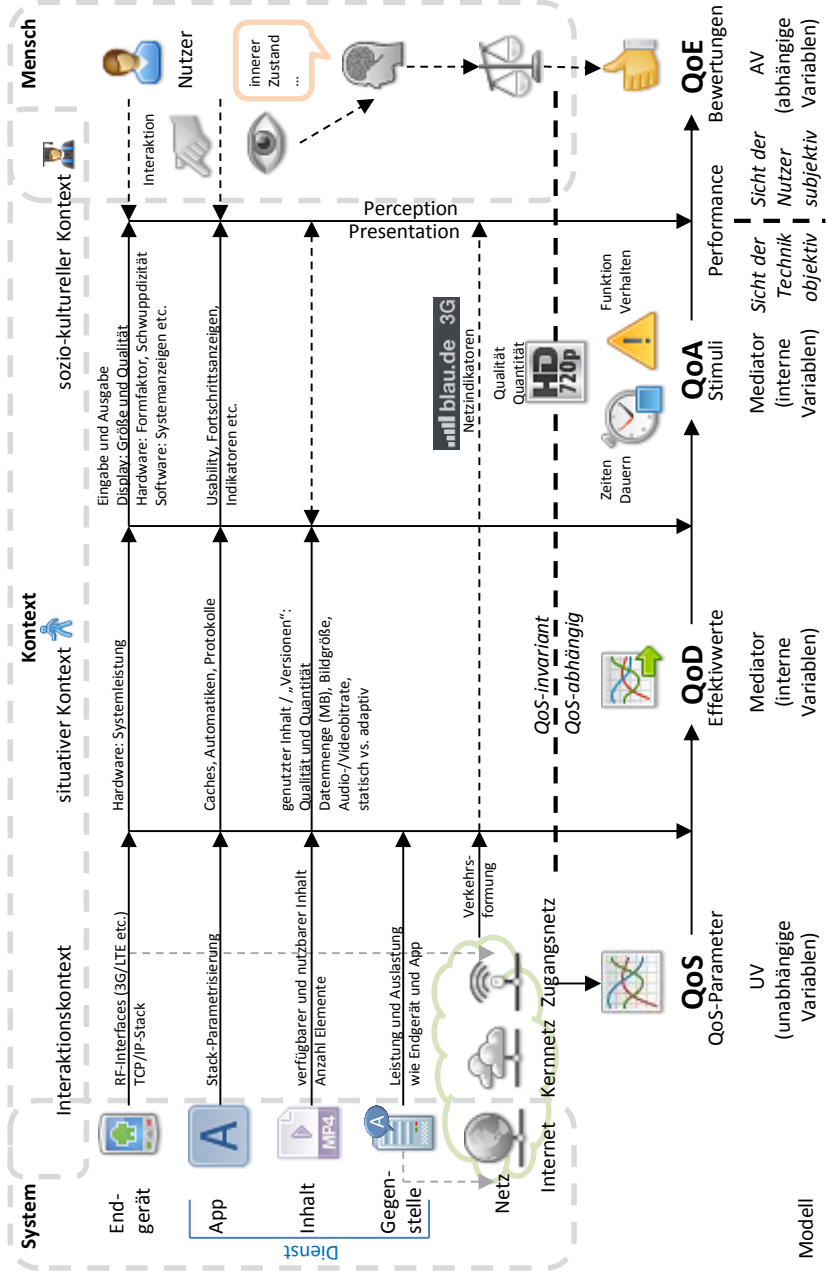


Abbildung 22: QoX-Matrix mit Mediatoren und orthogonalen Moderatoren

3.1.8 Zur weiteren Diskussion des Modells

Eine vollständige und zweckmäßige Diskussion des Zusammenwirkens im Modell ist eine anspruchsvolle Aufgabe, über die der Autor nicht nur einmal sinniert hat. Anders als eine rein lineare Darstellung sind die internen Zusammenhänge komplex und nur schwierig ohne Vorgriffe und Mehrfachbeschreibungen zu erörtern. Eine reine Diskussion Top-Down vom Nutzer zum Netz entfiel damit ebenso wie Bottom-Up vom Netz zum Nutzer. Netz und Nutzer wurden aber ohnehin zusammen mit den Konzepten zur QoS und QoE bereits im vorherigen Kapitel beschrieben. Die weitere Diskussion konzentriert sich damit auf die noch nicht eingeschlossenen Elemente mit den Diensten im Fokus und den zwei als relevant identifizierten Sichten und den beiden Zwischenschichten *QoD* und *QoA* dazu. Das mobile Endgerät als Einflussfaktor wird der Diskussion der Dienste vorangestellt. Für die Beschreibung der Dienste wird eine mehrteilige Beschreibung gewählt. Die beiden Sichten *technische Sicht/Netzsicht* und *Nutzersicht* werden in der Folge in den Kapiteln zur Empirie immer wieder aufgegriffen.

Damit die Betrachtungen zum Netz im Modell noch ohne die konkreten Eigenschaften der Endpunkte auskommen, können diese zunächst als hypothetische Endpunkte eingeführt werden, ähnlich wie die Hypothetical Reference Endpoints bei der ITU-T (siehe z. B. [50] S. 44). Für diese ist von solchen Eigenschaften auszugehen, dass die technologisch vorgegebenen Bestwerte des Netzes erreicht werden können, zusammen mit angenommenen Eigenschaften und Implementierungen auf höheren Protokollschichten. Für die nachfolgenden Betrachtungen konkreter Endgeräte und Gegenseiten sind hingegen die jeweils relevanten Beschränkungen zu berücksichtigen. Im Modell ist dies in gewisser Art ein Kniff, um die eigentliche Betrachtung der tatsächlich erreichbaren Leistung (mit Einbeziehung der echten Endpunkte) von der prinzipiell erreichbaren Leistung (mit den hypothetischen Endpunkten) zur tatsächlich erreichten Leistung unter Einbeziehung aller relevanten Faktoren zu verlagern. Im Matrixmodell ist dies durch die gestrichelten Pfeile der Elemente Endgerät und Gegenseite an die Enden des Netzes angedeutet.

3.2 Das mobile Endgerät

Das mobile Endgerät⁵⁶ kann als Qualitätselement auf verschiedene Art und an unterschiedlichen Stellen in der QoS-QoE-Matrix einen Einfluss ausüben. Der Einfluss kann über die Leistung (objektiv) wie auch über Präsentation und Wahrnehmung inklusive haptischer Aspekte (subjektiv) erfolgen. Die Eigenschaften der mobilen Endgeräte können getrennt nach Hardware und Software beschrieben werden. Tabelle 7 fasst die Einflussmöglichkeiten der Endgeräte entsprechend zusammen, während die nachfolgenden Unterabschnitte auf die erwähnten Punkte näher eingehen. Die Eigenschaften eines Endgerätes können mit anderen Faktoren bzw. Elementen im Modell wechselwirken, z. B. dem Inhalt. So ist es beispielsweise möglich, dass auf einem Gerät aufgrund fehlender Leistungsfähigkeit oder mangels geeigneter Darstellungsmöglichkeit (unzureichende Bildschirmauflösung) die Auswahl des nutzbaren Inhalts beschränkt ist.

Tabelle 7: Zusammenfassung zum möglichen Einfluss der Hardware und Software der mobilen Endgeräte im Modell

Einflussebene	Hardware	Software
QoA → QoE Präsentation und Interaktion („von außen zu erleben“)	Ergonomie der Hardware quantitative und qualitative Eigenschaften der Eingabe und Ausgabe Bildschirm (als Ausgabe): Größe / Auflösung / Brillanz / Schärfe / Farbtreue Tastatur bzw. Touchscreen (als Eingabe): Genauigkeit, Empfindlichkeit Formfaktor haptische und ästhetische Kriterien	Ergonomie der Software Darstellung der Netzindikatoren Bedienphilosophie Bildschirmtastaturen und Eingabemethoden Statusanzeigen und Benachrichtigungen
QoD → QoA Anwendungsleistung („innere Werte“ und Messwerte)	CPU, RAM, Flash-Speicher CPU: Anzahl Kerne, Taktrate, Architektur RAM: Größe (Caches, Parallelbetrieb von Apps) interner Speicher (Flash) : Größe und Geschwindigkeit (Start von Apps)	Schwuppdizität (siehe Fußnote 61) System-Caches, z. B. max. Zwischenspeichergrößen Protokollunterstützung und APIs
QoS → QoD Kommunikationsleistung (Spezifikation und Messwerte)	UE-Kategorien, nutzbare Netztechnologie und nutzbares Netz Kommunikationsschnittstellen und unterstützte Standards (z. B. 3G, LTE, WLAN) theoretisch maximale Geschwindigkeit	TCP/IP-Netzwerk-Stack mit verschiedenen Implementierungen und Parametrisierungen

⁵⁶ in gleicher Bedeutung an anderer Stelle teilweise auch Terminal oder User Equipment (UE) genannt

3.2.1 Hardware: Geräteklassen und Leistungsklassen

Eine grundlegende Unterscheidung der Endgeräte bezieht sich auf Leistungsfähigkeit, Funktionsumfang und die Erweiterbarkeit der vorinstallierten Software:

- „Handy“ (einfaches Mobiltelefon): Telefonieren und SMS
- Feature Phone: zusätzlich einfache Apps und Internet-Browser, Tastatursteuerung oder einfache Touchscreens
- „smarte“ Geräte: zusätzlich erweiterbar durch Apps, weitere Unterscheidung nach Formfaktor in Smartphone, Phablet und Tablet (mobile Geräte), Steuerung über einen großflächigen Touchscreen, eingebaute Sensoren
- Datenstick/Datenkarte: zur Nutzung am Laptop (tragbare Geräte)

Im Weiteren sollen mit Endgerät nur noch „smarte“ Geräte gemeint sein, die von ihrer Größe her wirklich mobil sind, d. h. üblicherweise für die Nutzung unterwegs geeignet. Die portablen Geräte sind hingegen zwar von einem Ort zum anderen transportierbar, im Unterschied dazu aber nur eingeschränkt „on the go“ nutzbar, schon weil sie meist einen Platz zum Sitzen voraussetzen und andere Bedienphilosophien mitbringen. Trotzdem kann ein Blick in Richtung dieser Geräte interessant sein, weil sie hinsichtlich möglicher Kommunikationsleistung und Applikationsleistung sowie daraus resultierender möglicher Nutzungen (z. B. HD-Video) den mobilen Geräten einige Schritte voraus sein können (siehe Innovationszyklus aus verfügbarer Leistung, Nutzung und geforderter Leistung in der Einleitung).

Die mobilen Endgeräte lassen sich grob in Geräteklassen zwischen Low End und High End einteilen (Einstieger-, Mittel-, Spitzenklasse), die sich in Ausstattung und Leistungsfähigkeit der Hardware in ihren Merkmalen (RAM, interner Speicher, CPU, GPU, Displaygröße und Displayauflösung, ...) unterscheiden. Beispiele zu diesen Merkmalen in den unterschiedlichen Leistungsklassen auf dem Stand der Technik zum Zeitpunkt der Testplanung gibt Tabelle 29 mit der Beschreibung der Endgeräte in der eigenen Studie. Die kurzen Innovationszyklen der Branche mit einer jährlich neuen Spitzenklasse verschieben ständig die Abgrenzungsmerkmale: heute Spitzenklasse, morgen Mittelklasse, ...

3.2.2 Software: Mobile Plattformen und Herstelleranpassungen

Der Markt der mobilen „smarten“ Endgeräte wurde zum Testzeitpunkt durch zwei Software-Plattformen dominiert: Android und iOS (siehe Zahlen in [131]). Aus diesen Plattformen ergeben sich weitgehend die Eigenschaften für eine ganze Familie von Geräten. Trotzdem sind grundsätzlich Unterschiede innerhalb einer Plattform und zwischen Versionen möglich und auch festzustellen. Diese sind teilweise durch Anpassungen der Hersteller begründet, wie z. B. bei Android üblich. So passen einzelne Hersteller wie u. a. Samsung, Sony oder HTC die Android-Versionen ihrer eigenen Bedienphilosophie an und fügen kleine Software-Beigaben⁵⁷ hinzu und wählen ggf. auch technisch relevante Parametrisierungen anders (siehe nachfolgende Diskussion zur Stack-Parametrisierung).

Im Unterschied zu den Hardware-Eigenschaften können sich die Software und damit das Gesamtverhalten eines Gerätes über den Nutzungszeitraum (Monate/Jahre) durchaus deutlich ändern, z. B. durch größere System-Updates. So sind rückblickend „Verbesserungen“ der Bedienphilosophie aber auch der gefühlten Geschwindigkeit oftmals Hauptpunkte der Entwicklungsbemühungen gewesen. Vor allem ab der Android-Version 4.1 wurde Wert darauf gelegt, die gefühlte Geschwindigkeit beim Umgang mit der Benutzeroberfläche zu steigern [132]. Gleichzeitig gehen aber auch teilweise „Verschlechterungen“ mit Updates einher, insbesondere wenn eine neuere Version der Systemsoftware die ehemals flinken Geräte überfordert, weil z. B. neue leistungshungrige Assistenzfunktionen hinzugekommen sind oder der Arbeitsspeicher noch nur knapp ausreicht. Entsprechend sollten Betrachtungen bzw. konkrete Erkenntnisse zu einem Gerät nicht nur unter Angabe der Hardware, sondern auch zum genutzten Stand der System-Software erfolgen.

Prinzipiell ist es möglich, dass die Nutzer verschiedener mobiler Plattformen grundsätzlich unterschiedlich zufrieden sind (z. B. Android vs. iOS). Derartige Ergebnisse besonders aus älteren Studien⁵⁸ sind jedoch genau zu prüfen und kritisch zu hinterfragen, da sich durch die schnellen Innovationszyklen in

⁵⁷ Beispiele siehe z. B. <http://www.connect.de/ratgeber/touchwiz-sense-android-uis-im-vergleich-1489525.html> abgerufen am 17.04.2013

⁵⁸ z. B. <http://de.statista.com/statistik/daten/studie/13031/umfrage/nutzerzufriedenheit-von-smartphone-nutzern-im-vergleich-usa/> abgerufen am 29.05.2014, US-Markt, 2009, mit Vorteilen für iOS im Vergleich zur anderen Smartphones

Hardware und Software die Voraussetzungen ständig verschieben. Gegebenenfalls sind die Unterschiede innerhalb einer Plattform größer als zwischen den jeweiligen Spitzenmodellen der einzelnen Plattformen.

3.2.3 Einfluss auf die Kommunikationsleistung

Das mobile Endgerät ist der nutzerseitige physikalische Endpunkt für die Ende-zu-Ende-Kommunikation und somit für die Betrachtungen zur Ende-zu-Ende-QoS und zur erreichbaren Kommunikationsleistung von Relevanz. Die theoretischen Bestwerte (brutto) dazu ergeben sich aus den unterstützten Mobilfunkstandards und UE-Kategorien des Endgerätes, siehe Tabelle 67 im Anhang. Die Endgeräte unterscheiden sich auch bei vergleichbaren nominellen Eigenschaften in ihren tatsächlichen technischen Eigenschaften, wie z. B. dem Funkmodem, der Antennenkonstruktion (Hardware) und der TCP-IP-Stack-Parametrisierung (Software), die sich auf die effektive Kommunikationsleistung unter Realbedingungen auswirken können (siehe Testergebnisse in [133] zu unterschiedlichen Endgeräten unter vergleichbaren realen Netzbedingungen im Mobilfunk).

3.2.4 Einfluss auf die Applikationsleistung

Inwiefern sich die Merkmale der Hardware auf die Applikationsleistung verschiedener Dienste tatsächlich relevant auswirken, ist abseits synthetischer Benchmarks⁵⁹ schwierig zu beantworten. Für einige Dienste ergibt sich eine Abschätzung aus der standardisierten Ermittlung einzelner Vergleichswerte (Benchmarks), allerdings meist ohne Abgleich zur Relevanz, z. B. Ermittlung der JavaScript-Leistung⁶⁰ im Browser. Solche Vergleiche beziehen zudem zwangsläufig immer mehrere Elemente mit ein: Hard- und Software des Endgerätes sowie die eigentliche App. Für bestimmte Dienste sind zur Beurteilung der Applikationsleistung besondere Merkmale zu berücksichtigen, z. B. eine Hardware-Beschleunigung für unterstützte Video-Codecs. Einen Einblick zu konkreten Zahlen wie die Endgeräte direkt oder indirekt auf die Kommunikationsleistung und die Applikationsleistung einwirken können, gibt [134], wenn

⁵⁹ z. B. Antutu Benchmark <http://www.antutu.com> abgerufen am 22.01.2015

⁶⁰ z. B. SunSpider JavaScript Benchmark <https://www.webkit.org/perf/sunspider/sunspider.html> abgerufen am 22.01.2015

auch durch Fortschreiten der technischen Entwicklung nicht mehr auf aktuellem Stand. Die gefühlte Geschwindigkeit des Gerätes, die *Schwuppdizität*⁶¹, z. B. als Zeit für App-Start, Flüssigkeit der Bedienung und Flüssigkeit der Anzeige, wird durch die Hardware und Software geprägt und steht zwischen Applikationsleistung und Wahrnehmung der Nutzer.

3.2.5 Einfluss auf Präsentation und Interaktion

Über die Präsentation der erzielten Leistung und Interaktion mit dem Endgerät erfolgt die eigentliche Wahrnehmung der Endgeräte durch die Nutzer. Quantitative und qualitative Eigenschaften der Eingabe und Ausgabe spielen dabei eine hervorgehobene Rolle für die Hardware-Ergonomie. Der Bildschirm als Touchscreen agiert dabei oftmals in einer Doppelfunktion mit jeweils relevanten Merkmalen, als Gerät zur Eingabe über eine Bildschirmtastatur bzw. als Point-and-Click-Eingabe (Genauigkeit, Empfindlichkeit) und zur Ausgabe (Größe, Auflösung, Brillanz, Schärfe, Farbtreue, Blickwinkelabhängigkeit). Die Eingabe und Ausgabe von Audio komplettiert die visuelle Komponente entsprechend zur audiovisuellen Schnittstelle zwischen Mensch und Maschine, teils ergänzt um haptische Komponenten (Vibration).

Auch die System-Software kann mit Bedienphilosophie und Usability (Bildschirmtastaturen, Eingabemethoden), den Statusanzeigen und Benachrichtigungen (Hintergrund/Vordergrund) sowie der Darstellung und Sichtbarkeit der Netzindikatoren (Netztechnologie und Signalstärke) einen gerätespezifischen Einfluss auf das Nutzererlebnis und die Erwartungshaltung der Nutzer ausüben. Beispiele für Anzeigeelemente dieser Art sind in Abbildung 18a und Abbildung 52 gezeigt.

⁶¹ Begriff aus der PC-Zeitschrift *c't*, der zur Abgrenzung der messbaren (Hardware-) Leistung in Benchmarks etc. von der gefühlten Arbeitsgeschwindigkeit in der täglichen Nutzung, der „Schwuppdizität“, dient. Die erstmalig publizierte Nutzung des Begriffs in dieser Bedeutung konnte leider nicht mehr nachvollzogen werden.

3.2.6 Das Endgerät als Erlebnis- und Erwartungsfaktor

Das mobile Endgerät kann als Erlebnis- und Erwartungsfaktor wirken. Sind trotz objektiv gleicher Leistung und Präsentation doch Unterschiede in der Bewertung festzustellen, dann sind diese Unterschiede subjektiv zu begründen, z. B. durch eine andere Wahrnehmung und/oder Erwartung. Dabei sind mögliche Einflussfaktoren, von denen kein direkter Einfluss auf die Leistung ausgeht:

- Formfaktor: verschiedene Hauptnutzungszwecke und Anwendungsbereiche woraus unterschiedliche Anforderungen abgeleitet werden (z. B. Smartphones zur Kommunikation und Tablets zur Unterhaltung)
- Beschaffenheit und äußere Erscheinung: haptische und ästhetische Aspekte („Handschmeichler“, Design)
- Anpassung der Erwartung an die vermeintliche Leistung (z. B. High-End-Gerät und High-End-Erwartungen) oder externe Meinungen sowie Image der Marke und des Gerätes

[135] stellt die Ergebnisse einer Studie vor, in der Unterschiede in den QoE-Bewertungen für heterogene Geräte (Bildschirmgröße, Leistung) bei gleichen Netzbedingungen gefunden werden konnten, allerdings ohne Klärung der eigentlichen Zusammenhänge (Black Box). Es blieb unklar, ob die Unterschiede durch Leistung, Leitungswahrnehmung oder Erwartung zu begründen waren. Trotzdem ist dies ein wichtiger Hinweis auf den möglichen Einfluss der Endgeräte.

Fragt man die Nutzer direkt, was ihnen in der alltäglichen Nutzung bei einem mobilen Endgerät wichtig ist, so ist die häufigste Antwort „die Batterie“, noch vor Aspekten, die dem Design der Nutzerschnittstelle, der Applikationsleistung und erst dann der QoS zuzuordnen sind (Nennung in absteigender Abfolge, siehe [136]). Punkte wie die Batterieleistung bleiben in der eigenen Betrachtung und Berücksichtigung im Modell außen vor. Sie sind zwar auf das mobile Endgerät bezogen, gewinnen ihre Relevanz aber eher aus dem Aspekt des situativen Kontexts einer mobilen Nutzung fernab der nächsten Steckdose.

Die Eigenschaften der Endgeräte stehen offenbar auch in Beziehung zur eigentlichen Nutzung, zumindest weisen entsprechende Ergebnisse in diese

Richtung, siehe [137]⁶². Nutzer von Smartphones mit größeren Displays (ab 4,5“) verursachen demnach deutlich mehr Datenvolumen (44%, Wi-Fi und mobile Datennetze aggregiert) und nutzen ihre Geräte dabei in allen relevanten Kategorien auch intensiver als Nutzer von Geräten mit kleinerem Display [137].

3.3 Die Dienste – Teil 1: Begriff und Elemente

In diesem Abschnitt werden die grundsätzlichen Überlegungen zu den Diensten bzw. Services eingeführt. Die Dienste verbinden auf vielfältige Art und Weise Technik und Nutzer, entsprechend kommt ihnen eine große Bedeutung zu, da sie den eigentlichen Zweck einer Netznutzung bilden.

Es wird zunächst auf den Begriff und danach kurz auf die einzelnen Elemente eingegangen, die im eigenen Verständnis einen Dienst formen. Grundsätzlich erscheint eine Aufteilung der Diskussion zu den Diensten zweckmäßig: in Netzsicht, Nutzersicht und Gesamtsicht des QoS-QoE-Zusammenhangs, welche die Dienstklassen sowie die formale Beschreibung des dienstspezifischen Zusammenhangs enthält. Entsprechend dieser Aufteilung erfolgt eine detaillierte Betrachtung zu den Diensten in den nachfolgenden Abschnitten.

3.3.1 Begriffsbestimmung

Es ist sinnvoll, den Begriff des Dienstes bzw. des Service näher zu beleuchten und eine passende Begriffsbestimmung zu suchen. Dies ist bei genauer Betrachtung nicht so trivial, wie es zunächst erscheinen mag.

Service kann im Folgenden als englische Entsprechung für Dienst genutzt werden. Service im Sinne von gutem Kundendienst oder Kundenbetreuung ist hingegen im Weiteren ausgeklammert (siehe dazu ganzheitliche QoS-Sicht in 2.2.2 und nichttechnische Angebotseigenschaften in 2.4.4).

Art des Dienstes und Anbieterrollen: Für telefonbasierte Dienste ist der Netzanbieter/Zuganganbieter meist zugleich auch der Anbieter der eigentlichen Dienste wie Telefonie als Audiokommunikation, Videotelefonie und ggf.

⁶² Studie im US-Markt des Jahres 2013

von Mehrwertdiensten wie SMS⁶³ und MMS⁶⁴. Für diese Nutzungen ist im Grunde sofort ersichtlich „Wer ist Anbieter?“ und „Was ist der Dienst?“. Entsprechend ist keine weitere Unterscheidung hinsichtlich des Dienstes notwendig, weder in der Sicht des Anbieters, noch des Nutzers/Kunden, noch der Spezifikation. Die ITU-T E.800 spezifiziert für einen solchen telefonbasierten Dienst:

“A set of functions offered to a user by an organization constitutes a service.” [35] S. 3

Für die Nutzung von Diensten im Internet sind diese Überlegungen komplizierter. Je nach Sicht oder Zielsetzung ist eine Unterscheidung angebracht, meist aber nicht spezifiziert. Daher ist genau zu prüfen bzw. zu definieren: Was ist die angebotene Funktionalität, worauf sich die Ausführungen beziehen?

- ... der Zugang zum (mobilen) Internet als Datendienst?
- ... oder ein einzelner, ganz bestimmter Dienst im (mobilen) Internet?
- ... oder die Gesamtheit „aller Dienste“ im (mobilen) Internet, also „das“ mobile Internet?

Entsprechend dieser Auflistung gibt es auch unterschiedliche Anbieter (Wer bietet an?):

- Der Netzanbieter realisiert mit der Ermöglichung des Zugangs zum Internet einen Datendienst (und somit mit dem Netz eine Kommunikationsleistung).
- Der Dienstanbieter realisiert einen anwendungsbezogenen bzw. inhaltsbezogenen Dienst (z. B. Facebook).

Prinzipiell kann auch der Netzanbieter noch in einen Netzbetreiber und einen Serviceanbieter unterschieden werden. Der Netzbetreiber errichtet und betreibt ein Mobilfunknetz und verkauft diese Leistungen an einen Serviceanbieter weiter, der selbst wiederum den Netzzugang (und „Kunden-Service“ im

⁶³ Short Message Service

⁶⁴ Multimedia Message Service

ausgeklammerten Sinn) an Kunden weiterverkauft. Für die Diskussion an dieser Stelle reicht die zusammengefasste Betrachtung als Netzanbieter ohne eine solche Unterscheidung aus.

Neben der formellen Unterscheidung ist noch die Sicht des Nutzers bzw. Kunden zu berücksichtigen. In dieser Sicht mag es eine Rolle spielen, zwischen den Anbietern zu unterscheiden oder auch nicht – ohne Netz ergibt sich kein Nutzen der Dienste im Internet und ohne die Dienste im Internet verbleibt nur ein Netz ohne Nutzen. Es kann daher interessant sein, wen der Nutzer in der Verantwortung sieht, falls bestimmte Dienste einmal nicht so funktionieren, wie er es erwartet.

Versuch der abstrakten Begriffsbestimmung: Die Qualinet-Definitionen zu QoS und QoE widmen sich auch Applikation und Service, die in Zusammenhang zu sehen sind. Die Definition der Applikation in Qualinet erfolgt in nachvollziehbarer Art:

“A software and/or hardware that enables usage and interaction by a user for a given purpose. Such purpose may include entertainment or information retrieval, or other.” [43] S. 6

Der Dienst wird in Qualinet hingegen stark abstrakt definiert, so können prinzipiell alle Arten von Diensten gemeint sein:

“An episode in which an entity takes the responsibility that something desirable happens on the behalf of another entity.” [43] S. 6 mit Bezugnahme auf das Dagstuhl Seminar 09192

Erst zusätzliche Anmerkungen beziehen den Dienst auf den Bereich der Kommunikation, allerdings ohne weitere Konkretisierung. Eine solch abstrakte Definition ist leider, auch wenn sie formal korrekt ist, wenig hilfreich für das Verständnis. Deshalb soll eine Analogie zur Verdeutlichung des Dienstbegriffes herangezogen werden.

Analogie paketbasierter Dienste: Der Vergleich der Dienstnutzung im mobilen Internet zu klassischen Telekommunikationsnetzen wie dem Plain Old Telephone System (POTS) und Telefondiensten ist eigentlich nicht der beste Vergleich. Zu eng sind dort Netzbetrieb und Diensterbringung miteinander und dem Paradigma der leitungsbasierten Vermittlung verwoben. Sucht man eine eingängige Entsprechung des (mobilen) Internetzugangs aus Nutzersicht in der analogen Welt, dann ist der Postverkehr bzw. genauer der Paketverkehr eigentlich besser geeignet. Es lässt sich in dieser Analogie gedanklich besser trennen, was die Leistung welches Anbieters ist – und Pakete spielen entsprechend des Paradigmas der paketbasierten Vermittlung im Internet auch eine Rolle. Dabei ist es sinnvoll und wichtig zwischen den Diensten *Transport* der (Daten-) Pakete und interessierendem *Inhalt* der (Daten-) Pakete zu trennen. Der *Transportdienst* ist Aufgabe des Netzes, aber aus Nutzersicht nur Mittel zum Zweck, nämlich zur Ermöglichung der Nutzung der *Inhaltsdienste* – wie DHL für Amazon. Bei der Betrachtung des QoS-QoE-Zusammenhangs nimmt somit der Transportdienst des Netzes Einfluss auf das spezifische Nutzererlebnis der einzelnen Inhaltsdienste.

Managed Services vs. OTT-Services: Aus der Unterscheidung der Anbieter ergeben sich diese beiden Begriffe, die sinnvollerweise im Umfeld der Dienste zu unterscheiden sind.

Managed Services werden durch und innerhalb eigener Subsysteme des Netz-anbieters, unter seiner Kontrolle, angeboten und erbracht. Im Bereich der stationären Nutzung in Next Generation Networks (NGN) sind die gemanagten IPTV-Systeme, wie z. B. *Entertain* der Telekom, neben der IP-basierten Sprachtelefonie (VoIP) die bekanntesten Managed Services auf IP-Basis. In mobilen Datennetzen können Rich Communication Services (RCS) wie JOYN⁶⁵ zu den Managed Services auf IP-Basis gezählt werden. Für derartige Services werden unter Umständen die netzimmanenten QoS-Mechanismen genutzt, d. h., QoS-Garantien sind ggf. möglich.

Over the Top Services (OTT) bezeichnen Dienste, die im offenen Internet angeboten und darüber erbracht werden. Ihre Nutzung läuft über die Köpfe (und

⁶⁵ <http://www.joynus.com/de/> abgerufen am 18.01.2015

Portemonnaies) der Netzanbieter hinweg. QoS-Mechanismen stehen typischerweise nicht zur Verfügung, entsprechend ist bei der Erbringung bzw. Nutzung derartiger Dienste nur von Best Effort auszugehen. In den nachfolgenden Betrachtungen, wie auch den praktischen empirischen Untersuchungen, stehen OTT-Dienste aus dem offenen Internet im Mittelpunkt.

Eigene Festlegung und Dienstbestandteile: Der Begriff des Dienstes soll sich nachfolgend auf die Erbringung bzw. Nutzung von Multimediadiensten bzw. von multimedialen Inhalten und Telekommunikationsdiensten im mobilen Internet als IP-basierte Inhaltsdienste beziehen. Der Zugang zum mobilen Internet an sich, als technische Möglichkeit der Endgeräte und Bestandteil des Dienstangebotes eines Netzanbieters, ist dafür als Transportdienst in Form des Netzzugangs eine technische Voraussetzung und als impliziter Bestandteil im Element *Netz* in Form der netzbezogenen QoS erfasst.

Bezogen auf die definierten Qualitätselemente ist ein Dienst im Weiteren ein geschaffenes Konstrukt, charakterisiert durch den Austausch eines *Inhalts* zwischen einer konkreten Anwendung (*App*) auf lokaler Seite und der *Gegenseite* durch das Netz. Je nach Auslegung kann so der Austausch eines der Elemente oder auch nur die Veränderung einer relevanten Eigenschaft eines der Elemente zu einem anderen Dienst führen. (So sind zumindest in der begrifflichen Behandlung der Dienste diese voneinander zu trennen: beispielsweise Youtube HQ vs. Youtube HD oder Google Maps als Web-App im Browser vs. mobile App.)

3.3.2 Anwendung/App

Die Anwendung bzw. mobile Applikation (kurz App) beschreibt ein algorithmisches Qualitätselement als genutzte Software auf dem Endgerät der Nutzer und kann als Teilelement eines Dienstes auf mehreren Schichten des QoX-Modells Einfluss nehmen, siehe Zusammenstellung in Tabelle 8. Dabei kann die App in der technischen Sicht in QoD und QoA Auswirkungen haben, als auch über die Nutzersicht die QoE direkt beeinflussen. Entsprechende Anknüpfungspunkte zwischen App und beiden Sichten sind vorhanden. Um Dopplungen in der Diskussion zu reduzieren, soll an dieser Stelle nur ein Überblick über die möglichen (isolierten) Einflüsse gegeben werden. Das Zu-

sammenwirken über die Schichten hinweg gemeinsam mit den anderen beteiligten Qualitätselementen wird in der Netzsicht und Nutzersicht noch einmal ausgehend von folgenden Punkten aufgegriffen:

- Das Element App kann den eigentlichen Ende-zu-Ende-Datentransport und damit die Kommunikationsleistung beeinflussen.
- Je nach App können sich auch bei gleicher Kommunikationsleistung unterschiedliche Nutzerstimuli ergeben, da auch die Applikationsleistung beeinflusst werden kann.
- Die App kann über die Präsentation in der Nutzersicht auf den Dienst sowohl das Erlebnis (Wahrnehmung der Leistung) als auch die Erwartung beeinflussen.

Weitere Hinweise zu konkreten Wirkungen der ausgewählten Apps im Test geben die Erläuterungen des Testparcours, die sich teilweise auch wieder verallgemeinern lassen.

Mobile Apps können über System-APIs die Art des Netzzugangs ermitteln und ihr Verhalten daraufhin anpassen (Beispiel in [138] S. 37f). Aus den Erkenntnissen zu den Netzbedingungen der realen Welt jenseits optimaler Entwicklungsumgebungen sollten die App-Entwickler Konsequenzen ziehen. Eine clevere App kann versuchen, ungünstige Netzbedingungen weniger schlecht erscheinen zu lassen, z. B. durch die Wahl des passenden Inhalts (geringere Datenmenge) oder durch die Verlagerung netzintensiver Phasen, die ohne Nutzerinteraktion auskommen, in den Hintergrund. Die Apps sind (zumindest in ihrem Verhalten) als Element im Modell damit nicht nur ein Einflussfaktor auf QoD, QoA und QoE, sondern ggf. selbst QoS-adaptiv, z. B. bei der Inhaltsauswahl oder in der QoA mit Funktion und Verhalten, die sich meist in der App ausdrücken. Dies wird durch die entsprechenden Qualitätsmerkmale im Modell berücksichtigt.

Eine besondere Diskussion verdient der direkte Einfluss der Apps auf die QoE in der Nutzersicht. Losgelöst von der Diskussion der zumindest in Teilen QoS-abhängigen Applikationsleistung führt dies zu Überlegungen rund um die Usability und User Experience interaktiver Systeme, der Benutzeroberfläche und der Ergonomie der Software im Allgemeinen. Wie beim Endgerät, so ist auch

bei der App von einem Einfluss zum QoS-unabhängigen Grundgeschwindigkeitseindruck auszugehen (z. B. durch die Framerate des UI-Threads).

Tabelle 8: Zusammenfassung zum möglichen Einfluss der Anwendung/App im Modell

Einflussebene	Möglicher Einfluss
<p>QoA → QoE Präsentation und Interaktion („von außen zu erleben“)</p>	<p>Ergonomie, Benutzerschnittstelle, Usability und User Experience</p> <ul style="list-style-type: none"> • Benutzereingaben • Statusanzeigen und Benachrichtigungen • Fehlermeldungen <p>Präsentation der Nutzerstimuli</p> <ul style="list-style-type: none"> • Art der Darstellung • direkte oder indirekte Wahrnehmbarkeit (z. B. verdeckter Start einer Datenübertragung) • Einsatz und Positionierung von Indikatoren, z. B. zur genutzten Inhaltsversion (z. B. HQ vs. HD) <p>qualitätsverbessernde Maßnahmen in der Präsentation</p> <ul style="list-style-type: none"> • Funktion und Verhalten: Kaschieren von Fehlern • Zeiten und Dauern: „vorläufige“ Antworten und Platzhalter (Dummy-Inhalt) • Qualität und Quantität: Ausgleich von Qualitätsmängeln des Inhaltes • Berücksichtigung der Eigenschaften der menschlichen Wahrnehmung und Informationsverarbeitung (z. B. Farbwirkung) • Umwandlung von (negativer) passiver Wartezeit in aktive Nutzungszeit <p>allgemeines Applikationsverhalten und Schwuppdizität</p> <ul style="list-style-type: none"> • Startzeit der App • häufige Abstürze • Fehlermeldungen • Hänger und ähnliche Auffälligkeiten bei der Navigation innerhalb der App
<p>QoD → QoA Anwendungsleistung („innere Werte“ und Messwerte)</p>	<p>Einfluss durch unterschiedliche Strategien zur Verbesserung der Stimuli in der Applikationsleistung</p> <ul style="list-style-type: none"> • weniger Daten übertragen: Nutzung lokaler Caches und Datenreduktion mittels Nutzdatenkompression • Daten übertragen, bevor sie gebraucht werden: Preloading, Prefetching oder Precaching zur Übertragung von Inhalten in einen lokalen Zwischenspeicher noch vor einer ersten aktiven Nutzung durch den Nutzer, so dass dieser Inhalt im Moment des Zugriffs durch den Nutzer nicht erst durch das Netz transportiert werden muss • Mechanismen zur automatischen Auswahl des Inhaltes passend zu den Netzbedingungen oder der Leistungsfähigkeit des Endgerätes, z. B. bei Adaptive Bitrate Streaming • intelligente Fallback-Mechanismen

Fortsetzung der Tabelle auf der nächsten Seite

Fortsetzung der Tabelle 8

Einflussebene	Möglicher Einfluss
QoS → QoD Kommunikationsleistung (Spezifikation und Messwerte)	Protokoll und Parametrisierung der Applikationsschicht (Overhead, Turns) <ul style="list-style-type: none"> • Verbindungspersistenz, z. B. HTTP mit oder ohne keep-alive beim Web-Browsing • Header Compression, optionale Protokollbestandteile • möglicher zusätzlicher Overhead durch optionale Verschlüsselung (SSL/TLS) Wahl des Transportprotokolls und der damit verbundenen Vor- und Nachteile (TCP/UDP) <ul style="list-style-type: none"> • Latenz-empfindliches Request-Response vs. verlustempfindliche Einwegkommunikation • TCP-Streams vs. UDP-Datagramme Übertragungsstrategie und Transportparametrisierung <ul style="list-style-type: none"> • intelligente Wahl der besten (schnellsten) Gegenseite (Server) • Parametrisierung des TCP/IP-Stacks, z. B. durch Setzen von Socket-Optionen • Gleichzeitigkeit und Sequenz der Netznutzung, z. B. beim Web-Browsing, zur Übertragung einzelner Inhaltselemente

3.3.3 Gegenseite und Kommunikationsszenario

Die Gegenseite lässt sich aus objektiver Sicht äquivalent zur lokalen Seite als Qualitätselement beschreiben (Leistungsfähigkeit der Hardware und Software, Auslastung, Netzanbindung, ...). Entsprechend hat die Gegenseite ebenso Einfluss auf die Ende-zu-Ende-Betrachtungen wie die lokale Seite, sowohl auf die Leistung auf der Netzwerkschicht wie auch auf der Applikationsschicht. Moderne serverseitige Anwendungen zur Realisierung der Dienste sind selbst oftmals komplexe und verteilte Systeme, siehe Abbildung 23. Die inneren Elemente und ihr Zusammenwirken sind selbstverständlich wichtig, können aber nicht weiter im Detail betrachtet werden. Die Merkmale der Gegenseite sind für den mobilen Nutzer nur indirekt durch ihre Auswirkungen auf die Qualitätsmerkmale des Modells wahrnehmbar. Die Gegenseite soll daher auch lediglich als abstraktes Element berücksichtigt werden.

Je nach der Natur der Kommunikationspartner, hier konkret der Gegenseite, wird in Bezug auf die Kommunikationsszenarien zwischen Person-to-Person, Person-to-Machine und Machine-to-Machine unterschieden. Eine Variante von Person-to-Machine ist Person-to-Content: Die Nutzer sind an dem Inhalt

auf der anderen Seite des Netzes interessiert, bzw. ihren Inhalt auf die andere Seite zu bewegen. Damit lassen sich die meisten Szenarien im mobilen Web beschreiben. Alternative Bezeichnungen zur Gegenseite für das entfernte Ende der Kommunikation durchs Netz wären damit *Server* oder *Backend*.

Im direkten Vergleich unterschiedlicher Dienste kann die Performance der Gegenseite vor allem auf die Antwortzeit einen Einfluss haben. Dabei ist allerdings zu berücksichtigen, was die Ursache ist, was als Latenz interpretiert wird – und auf welcher Schicht des Stacks die Verzögerungen entstehen. Auf der Anwendungsschicht bedeutet die Beantwortung komplexer Anfragen wie HTTP-Requests, die im skizzierten Backend auf Datenbackabfragen abgebildet werden, teilweise erhebliche Antwortzeiten, die durch die algorithmischen und physikalischen Bestandteile der Gegenseite dominiert sind und zu dem eigentlichen Datentransport durch die Netze hinzuzurechnen sind. (Mobil-) Optimierungen auf der Gegenseite können zu QoE-Verbesserungen führen, besonders wenn sie an Optimierungen des Inhalts und der Übertragung anschließen (z. B. Reduzierung der Gesamtdatenmenge, der Anzahl Elemente und somit der Anzahl Turns und Protokollnutzung, z. B. Unterstützung von keep-alive für HTTP).

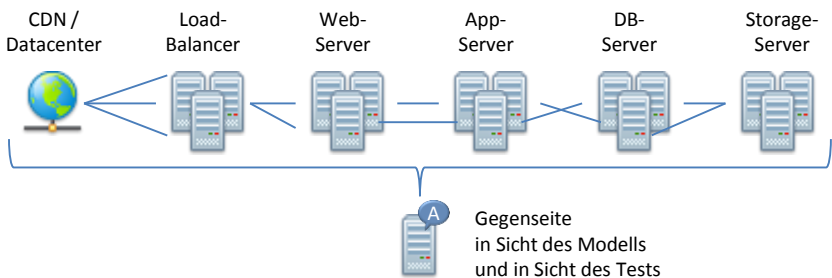


Abbildung 23: Die Gegenseite als komplexes System und als Abstraktion

schematische Darstellung zur Verdeutlichung, CDN: Content Distribution Network zur Verteilung über mehrere Standorte; in Anlehnung an [139], vereinfacht und abstrahiert zur zusammengefassten Sicht auf die Gegenseite

3.3.4 Inhalt

Der Inhalt ist ein eigenes Qualitätselement mit den Merkmalen technische Qualität und Quantität. Weitere nichttechnische also intellektuelle, künstlerische oder emotionale Komponenten des Inhalts sollen nachfolgend explizit ausgeklammert werden. Dies bezieht sich beispielsweise auf die Formulierung eines Textinhalts, gezeigte Bildmotive oder die Handlung eines Filmes sowie beteiligte Personen etc.

Als Inhalt sollen nachfolgend alle Medien verstanden werden, die ein Nutzer liest, ansieht, anhört oder anderweitig konsumiert (Texte, Bilder, Videos, Töne, Musik usw.) sowie im erweiterten Sinn alle Nutzdaten und programmatisch genutzte Daten (z. B. API-Nutzung⁶⁶). Es wird dabei davon ausgegangen, dass dabei zumindest ein Teil der Daten von einer entfernten Gegenseite durch ein Netzwerk (Internet, Mobilfunknetz) zur Nutzung zum mobilen Endgerät des Nutzers übertragen werden muss (bzw. umgedreht).

Es sind unterschiedliche Ausprägungen des Inhalts möglich, z. B. verschiedene Inhaltsversionen (Web-Browsing mit mobiler Version vs. Desktop-Seite) oder Qualitätsstufen (z. B. Video in HQ und HD), die sich in ihren objektiven Eigenschaften und ihrer subjektiven Wahrnehmung unterscheiden können.

Es existieren Anknüpfungspunkte zum Zusammenwirken mit den anderen Qualitätselementen sowohl in der Netzsicht (Übertragung des Inhalts durch das Netz) als auch in der Nutzersicht (Wahrnehmung des Inhalts als Qualität und Quantität). Der Inhalt kann im Modell auf QoD und QoA einwirken und selbst auch adaptiert werden. Für den Inhalt ist es daher sinnvoll, in der weiteren Diskussion zwischen dem Einfluss durch den Inhalt und den Einfluss auf den Inhalt zu unterscheiden.

Einfluss durch den Inhalt: Es ist denkbar, dass in Abhängigkeit des Zusammentreffens der Fähigkeiten und Einschränkungen von entfernter Seite (Anbieter) und lokaler Seite (Endgerät/App) nicht jeder Inhalt bzw. nicht alle Inhaltsteile in gleicher Art und Weise genutzt werden können. Der tatsächlich

⁶⁶ API: Application Programming Interface, Schnittstelle zur programmatischen Nutzung bzw. zum Datenaustausch

genutzte Inhalt ergibt sich aus der Schnittmenge aus dem verfügbaren Inhalt auf entfernter Seite und dem nutzbaren Inhalt auf lokaler Seite.

Der Inhalt kann das Erlebnis als auch die Erwartung beeinflussen. Ein (fiktives) Beispiel HD-Video-Streaming soll dies verdeutlichen:

- Der Nutzer stellt etwas längere Ladezeiten des HD-Videos und eine kurze Unterbrechung fest, akzeptiert dies aber, weil er bei „HD“ schon öfter ähnliche Erfahrungen gemacht hat.
- Die subjektive Wahrnehmung ist von der guten Bildschärfe und Lebendigkeit der Farben etc. geprägt. Genau dies hat der Nutzer durch den eingeblendeten Indikator „HD“ aber auch erwartet, was die etwas längere Wartezeit und eine kurze Unterbrechung nicht überschatten.

Konkrete Annahmen zur Beurteilung bzw. Einbeziehung der Inhaltseigenschaften in QoE-Bewertungen sind schwierig, da nicht klar ist, ob die Probanden die Eigenschaften überhaupt wahrnehmen (können), bzw. wenn sie wahrgenommen worden sind, wie sie dann in ihre QoE-Bewertung einfließen. In einigen Fällen mag es offensichtlich sein, z. B. bei Video (SD vs. HD), der Bildanzahl oder Größe (Vorschaubildchen vs. Vollformatfoto). In anderen Fällen ist dies weniger offensichtlich, z. B. bei der Anzeige eines Bildes ohne ins Bild zu zoomen, so dass Bilder mit einer Auflösung größer als die Bildschirmauflösung ohne Vorteile bleiben. Eventuell können Nutzer die Unterschiede nur im direkten Vergleich wahrnehmen, z. B. in einem A-B-Vergleich (mit Variante A ist es insgesamt gut, mit Variante B ist es insgesamt auch gut, im direkten Vergleich ist aber Variante B besser als Variante A). Unter Umständen ist den Nutzern vielleicht gar nicht einmal klar, dass es verschiedene Versionen gibt (Wissen um Alternativen), was ansonsten zu unterschiedlichen QoE-Bewertungen führen könnte.

Einfluss auf den Inhalt: Durch den Inhalt beeinflusste Qualitätsmerkmale können statisch oder adaptiv sein, d. h. entweder von der Netzleistung unabhängig oder abhängig, je nachdem, ob sich die Eigenschaften in Abhängigkeit der QoS-Eigenschaften ändern. Für adaptiven Inhalt gilt: Passt sich das Netz nicht dem Inhalt an (Anpassung an die Anforderungen), dann passt die Gegenseite oder die App den Inhalt dem Netz an (Anpassung an die Einschränkungen).

Die Auswahl/Anpassung des Inhalts kann entweder manuell (ggf. bewusst) oder automatisch (ggf. unbewusst) geschehen. Manuell bedeutet beispielsweise den gezielten Aufruf der „kleineren und schnelleren“ mobilen Webseite eines Dienstes statt der „großen“ Desktop-Seite oder die Nutzung lediglich von HQ-Videos, obwohl HD nutzbar wäre. Die automatische Auswahl und Anpassung des Inhalts erfolgt entweder ohne oder mit Berücksichtigung der tatsächlichen Kommunikationsleistung. Ohne die Berücksichtigung der tatsächlichen Kommunikationsleistung erfolgt die Anpassung beispielsweise nur aufgrund des erkannten Netzes, z. B. 3G und somit der vermutlichen Kommunikationsleistung oder auf Basis der Fähigkeiten und Einschränkungen⁶⁷ der Endgeräte bzw. aufgrund des mobilen Charakters des Gerätes, z. B. über eine Erkennung eines mobilen Browsers. Beispiele dazu sind die Auswahl einer Inhaltsversion, z. B. mobil vs. Desktop beim Web-Browsing, die Einschränkung der verfügbaren Qualitätsstufen und die Auswahl einer Qualitätsstufe (z. B. Audio / Video) beim Streaming. Mit Berücksichtigung der tatsächlichen Kommunikationsleistung passt sich der Inhalt ggf. sogar dynamisch mehrfach über die Zeit an wechselnde Bedingungen an. Beispiele dafür sind Qualitätsstufenwechsel (z. B. permanent wirksames Fallback von HD auf HQ) oder Adaptive Bitrate Streaming (ABR) mit dynamischem Step-Up und Step-Down der Qualitätsstufe in Abhängigkeit der aktuellen Netzcharakteristik.

Beispiele zur Rolle des Inhalts, inklusive der Anpassung an die Netzbedingungen, finden sich in der Beschreibung des Testparcours im Kapitel 9.

3.4 Die Dienste – Teil 2: Netzsicht

Die Dienste können aus der Netzsicht (Datentransport) unter technischen Gesichtspunkten charakterisiert werden. Unterschiedliche Dienste weisen dabei unterschiedliche Netznutzungsmuster und Verkehrsmodelle auf. Zur Beschreibung der Quality of Delivery im eigenen Modell soll dazu die Kommunikationsleistung eingeführt werden. Diese wird maßgeblich auch von den genutzten Protokollen im TCP/IP-Stack und der Stack-Parametrisierung bestimmt, die nachfolgend kurz diskutiert werden sollen. Aus der spezifischen

⁶⁷ Fähigkeiten und Einschränkungen können hier als Capabilities und Constraints verstanden werden wie in MPEG-21 beschrieben, siehe z. B. [140] S. 37ff zur Digital Item Adaptation Usage Environment Description.

Netznutzung eines Dienstes ergibt sich die Übertragungscharakteristik, die abschließend thematisiert wird.

3.4.1 Von QoS zu QoD mit der Kommunikationsleistung

Die Kommunikationsleistung beschreibt die für einen bestimmten Dienst Ende-zu-Ende erreichte kommunikationsorientierte Leistung in Abhängigkeit aller beteiligten Elemente zum Transport des Inhalts. Es sind damit technische Effektivwerte, die sich aus dem Zusammenspiel von mobilem Endgerät, Anwendung, Inhalt, Netz und Gegenseite ergeben. Die Betrachtungen sollen dabei unter Last erfolgen, also dynamisch bezogen auf die tatsächliche Nutzung (im Gegensatz zu bestmöglichen (statischen) Werten ohne Last). Die Beschreibung der zugehörigen Qualitätsmerkmale erfolgt durch Abwandlung der benannten QoS-Parameter. Wird nachfolgend verkürzt von Effektivwerten gesprochen, so sind die Qualitätsmerkmale der Quality of Delivery (QoD) als Kommunikationsleistung gemeint. Die Unterscheidung QoS und QoD mag zunächst vielleicht nicht ganz einsichtig sein, schließlich sollen beide ja mit ähnlichen Parametern beschrieben werden. Im Fall der QoS repräsentieren die genannten Parameter für Best Effort im Modell angenommene bestmögliche Werte. Für die QoD interessieren nicht prinzipiell erreichbare, sondern die tatsächlich erreichten Werte, also die Effektivwerte. Während bei der Betrachtung des Netzes mehr die Konnektivität im Fokus der Ende-zu-Ende-Betrachtungen stand, wäre eine treffende Beschreibung für die nachfolgenden Erläuterungen zur QoD „von Stack zu Stack durch das Netz“, was auch höhere Protokollschichten einschließt, siehe Abbildung 24.

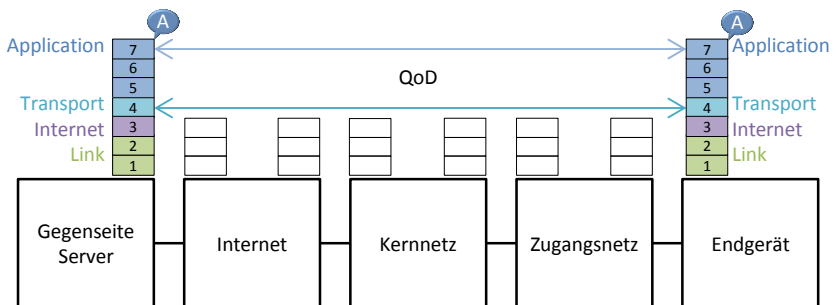


Abbildung 24: QoD – von Stack zu Stack durch das Netz
in Anlehnung an [35] S. 1 und [46] S. 8, erweitert und ergänzt

Dabei soll im Modell auch nicht mehr von hypothetischen Endpunkten, sondern von den Eigenschaften der tatsächlichen genutzten Hard- und Software ausgegangen werden, die sich zusammen mit dem Inhalt auf einzelne Datenübertragungen im Rahmen der Nutzung konkreter Dienste beziehen (z. B. einzelne TCP-Flows). Begriffliche werden die Betrachtungen zur QoS so auf die kommunikationsspezifische QoS (K-QoS) erweitert (wie in [45] S. 15f genutzt). Zur besseren Unterscheidung wird im eigenen Modell dazu QoD als Begriff genutzt (als Merkhilfe: D wie Delivery oder D wie Datenbewegung⁶⁸).

3.4.2 Protokollwelten: TCP/IP-Stack und Co.

Jedem Dienst sind Protokolle aus den jeweiligen Schichten des OSI-Modells bzw. TCP/IP-Stacks zugeordnet. Abbildung 25 gibt einen Überblick zu einigen ausgewählten (populären) Diensten bzw. Klassen und den dafür typischen Protokollen auf verschiedenen Schichten des TCP/IP-Stacks im DoD-Modell, das im Gegensatz zum bereits zuvor erwähnten formalen ISO-OSI-Referenzmodell hohe praktische Relevanz besitzt. Weitere Erläuterungen dazu findet der interessierte Leser in [34] S. 350ff in recht kompakter Form – oder extrem detailliert im Standardwerk *TCP/IP Illustrated* [141]. An dieser Stelle sollen nachfolgend nur kurz die Aufgaben der jeweiligen Schichten erörtert werden, um für die Erläuterungen in den Kapiteln 5 (zur Netzemulation) und 9 (Dienste im Testparcours) daran anknüpfen zu können.

In der **Applikationsschicht** finden sich sowohl dienstspezifische Protokolle als auch einige universell genutzte Protokolle. Aus einem All-over-IP, bezogen auf die Vermittlungsschicht (Internet), ist für viele populäre Dienste inzwischen schon ein All-over-HTTP geworden. Egal ob Cloud-Nutzung, Pseudo-Streaming oder Web-Browsing, das Hyper Text Transfer Protocol (HTTP) kann als das populärste L7-Protokoll⁶⁹ für eine Vielzahl verschiedener Dienste angesehen werden. Oftmals wird HTTP in Kombination mit Verschlüsselung (SSL/TLS) als HTTPS genutzt.

⁶⁸ Datenbewegung als Begriff wie in [21] S. 237, der den Transportaspekt durch das Netz hervorhebt

⁶⁹ Layer 7, übliche Abkürzung für die Applikationsschicht in der Nummerierung des OSI-Stack-Modells

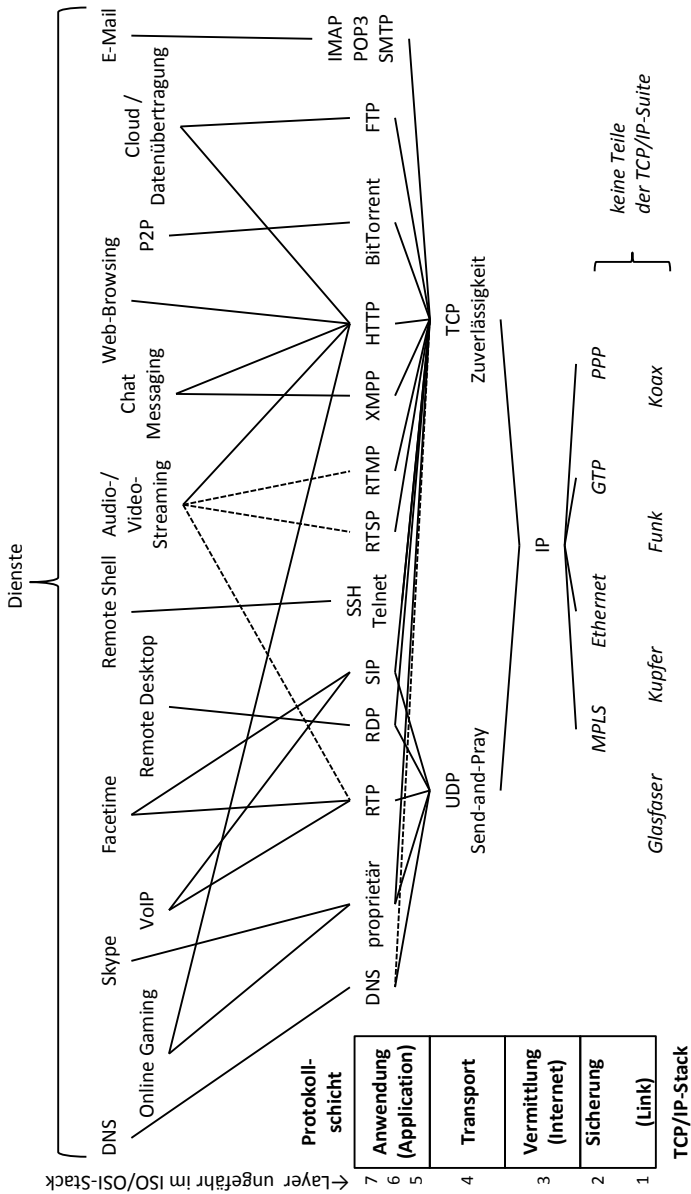


Abbildung 25: Populäre Dienste und dafür typische Protokolle auf den Protokollschichten des TCP/IP-Stacks

beispielhafte Zusammenstellung ohne Anspruch auf Vollständigkeit

Daraus ergeben sich teilweise einige Besonderheiten bzw. zu berücksichtigende Aspekte:

- zusätzliche Turns bei Initiierung des Datenaustauschs, die besonders bei vielen sequenziellen Übertragungen mit jeweils nur kleinen Datenmengen schnell zu einer Dominanz der Latenz für die effektive Leistung führen
- mehr Overhead durch mehr Daten

Die **Transportschicht** verbindet die Kommunikationspunkte transparent Ende-zu-Ende. Dabei nutzt die Masse der Dienste:

- User Datagram Protocol (UDP), vor allem wenn auf Zuverlässigkeit⁷⁰ verzichtet werden kann, oder zugunsten von Echtzeit-Anforderungen verzichtet werden muss
- Transport Control Protocol (TCP), wenn die Vorzüge der Zuverlässigkeit für höhere Protokollschichten gewünscht oder notwendig sind

In gleicher Art wie HTTP in der Applikationsschicht ist das Transport Control Protocol (TCP) in der Transportschicht von hoher Relevanz für viele populäre Dienste – leider aber auch Wurzel unterschiedlicher Probleme, besonders in Netzen mit hoher Latenz und Paketverlusten. Darauf wird im nachfolgenden Abschnitt noch weiter eingegangen.

Die **Internet-/Vermittlungsschicht** vermittelt die adressierten Datenpakete mittels Routing durch verschiedene untereinander verbunden Netze („das Internet“) von einer Quelle zum Ziel. Das Internet Protocol (IP) ist, wie der Name schon vermuten lässt, auf dieser Schicht von zentraler Bedeutung.

Die **Sicherungs-/Netzzugangsschicht** (Link) wird durch die jeweilige Technik und netzspezifische Protokolle im genutzten Netzsegment bestimmt (z. B. ARP, PPP, ...). Im Fall des nutzerseitigen Endpunktes bildet das Endgerät mit zugehöriger Hardware (z. B. Mobilfunkmodem oder WLAN-Schnittstellen), Gerätetreibern und Protokollimplementierungen den Übergang zum Netz. Die Netzzugangsschicht verkapselt die Spezifika des mobilen Netzzugangs. Be-

⁷⁰ UDP wird durch die fehlende Zuverlässigkeit scherzhaft auch als „send and pray“ bezeichnet [142] S. 207.

trachtungen dazu liegen außerhalb der weiteren Diskussion in der Arbeit, vertiefende Informationen dazu sind beispielsweise in [143], [56] und [144] zu finden.

3.4.3 Entstehen und Zusammenhang der Effektivwerte

Die Qualitätsmerkmale der Quality of Delivery (QoD) leiten sich aus den QoS-Parametern ab. Der effektive Wert jedes einzelnen Parameters der QoD hängt eng mit den anderen Parametern zusammen. Besonders für TCP als zuverlässiges verbindungsorientiertes Protokoll trifft dies zu. Dies bedeutet auch, dass durch die Mechanismen des TCP/IP-Stacks die Netzparameter in ihrer Wirkung auf der Schicht der QoD interagieren und im Weiteren in der Form der Effektivwerte als interne Variablen wirksam werden, siehe Abbildung 26.

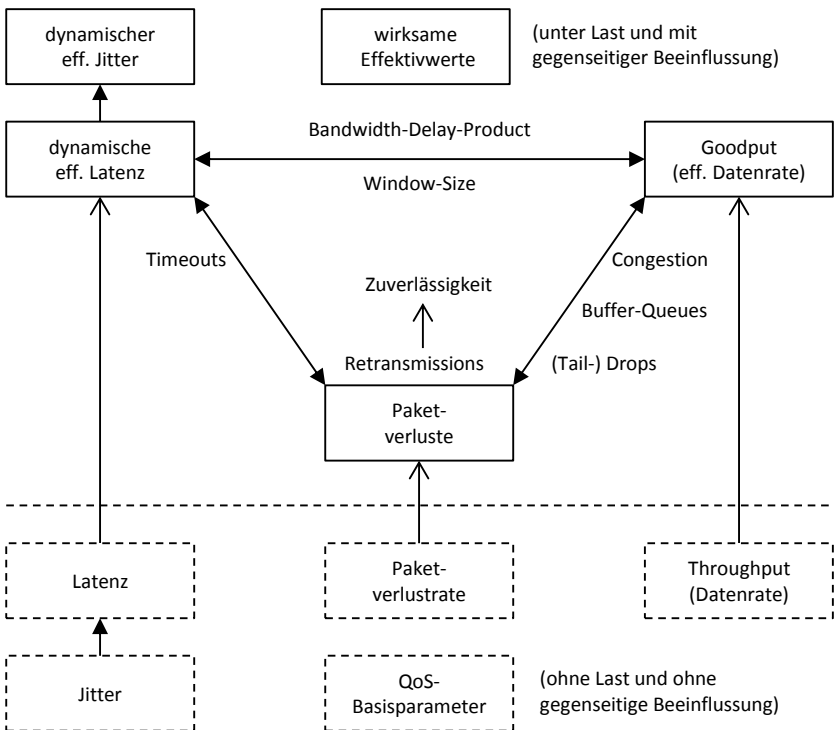


Abbildung 26: Zusammenhang und gegenseitige Beeinflussung ausgewählter leistungsbezogener Parameter zur QoD

Beispiel für TCP mit der Verortung möglicher Ursachen/Wirkungen

Die Kenntnis der Protokolle und Implementierungen erlaubt eine Abschätzung des TCP-Durchsatzes als Kommunikationsleistung, in Abhängigkeit der Paketverlustrate oder der Latenz, z. B. durch Simulation. Appendix IX der ITU-T-Empfehlung Y.1541 ([50] S. 43ff) widmet sich dem Zusammenhang von erreichbarer Bandbreite für verschiedene TCP-Window-Size-Werte zusammen mit unterschiedlichen Paketverlustklassen sowie Latenzwerten und gibt entsprechende Erwartungswerte an. Weitere Erläuterungen und Näherungsformeln finden sich u. a. in [145], wobei allerdings zu beachten ist, dass die Betrachtungen von den TCP-Implementierungen, als Eigenschaften der Endpunkte (also auch des Endgerätes) und der Stack-Parametrisierung abhängen. In der Realität behindert die Vielzahl möglicher Einflussgrößen damit eine tatsächliche Abschätzung. Auf die internen Zusammenhänge zwischen den QoS-Parameter auf die Gesamtleistung, auch in Verbindung mit weiteren Faktoren wie Puffergrößen und Implementierungen der TCP/IP-Stacks, sowie auf mögliche Verbesserungen der Performance geht [146] ein bzw. verweist auf entsprechende weiterführende Quellen. Die weitere Berücksichtigung des Overheads von Verschlüsselung und Applikationsprotokoll ergibt schließlich den Goodput (den „guten Durchsatz“) als effektive Bandbreite für einen Dienst mit entsprechendem Protokoll-Stack und konkreter Parametrisierung.

Der Goodput als effektiver Durchsatz wird u. a. beeinflusst durch:

- Bandwidth-Delay-Product
- Round Trip Time
- TCP-Window-Size
- TCP-Implementierungen auf Empfänger- und Senderseite
- Paketverluste und dadurch notwendige Retransmissionen
- verfügbare Bandbreite in Gegenrichtung (ACK-Pakete)
- Puffergrößen in zwischenliegenden Netzelementen

Die effektiv erreichte Latenz wird u. a. beeinflusst durch:

- Serialisierungs- und Bitübertagungszeit
- Wartezeiten in Queues: Warteschlangenlänge (Puffergrößen in zwischenliegenden Netzelementen) durch Congestion in Folge fehlender Bandbreite
- Prioritätsmanagement (Shaping)
- Paketverluste und dadurch notwendige Retransmissionen

Der Paketverlust wird u. a. beeinflusst durch:

- Warteschlangenüberlauf (Tail Drop) durch Congestion in Folge fehlender Bandbreite
- verspätet eintreffende Pakete (Timeouts), durch hohe Latenz
- in tieferen Schichten nicht ausgeglichene Übertragungsfehler (selten)

Paketverluste durch Congestion im Überlastfall sind mit den Timeout-bedingten Quasiverlusten in engem Zusammenhang zu sehen. Paketverluste auf höheren Stack-Schichten (IP) in Mobilfunknetzen aufgrund nicht ausgeglichener Übertragungsfehler sind hingegen trotz der drahtlosen Natur der Übertragung recht selten ([147] S. 425 und [148] S. 5). Die Mechanismen zur Korrektur und automatischen Übertragungswiederholung auf unteren Schichten sind sehr effektiv, selbst wiederum aber Quelle für zusätzliche Latenz und Jitter. Die Mechanismen zum Ausgleich von Paketverlusten auf höheren Schichten (Retransmissionen bei TCP auch aufgrund der anderen genannten Gründe) führen letztlich zur erwünschten Zuverlässigkeit für viele Dienste, allerdings auf Kosten der übrigen Parameter.

3.4.4 Mobilfunk und TCP-Blues: Congestion, Bufferbloat und Queues

Das Transportprotokoll TCP ist für eine Vielzahl von Diensten die wichtigste Einflussgröße für die Ende-zu-Ende-Leistung nach dem Netz selbst und den Eigenschaften der Endpunkte. Im Vergleich zu kabelgebundenen Verbindungen gewinnen mobilfunkspezifische, technologisch bedingte Aspekte besondere Bedeutung (nach [147] S. 424f):

- vergleichsweise niedrigere Durchsatzraten treffen auf vergleichsweise hohe Grundlatenz
- gemeinsam genutzte Kapazität (shared medium), die zu Engpässen führen kann
- zeitliche Variabilität der Charakteristiken
- u. U. große Bandbreitenunterschiede zwischen den Netzsegmenten (mit Paketstau (Congestion) am Flaschenhals von schnell zu langsam)

TCP wird dabei schnell zum „Problemfall“. Die Zusammenhänge sind äußerst komplex und können deshalb hier nur kurz angerissen werden, weitergehend Informationen zur mobilen Transportschicht sind in [147] S. 405ff zu finden, mit speziellen Hinweisen zu TCP für Mobilfunksystem ab [147] S. 424ff. Die

nachfolgenden Erläuterungen beziehen sich auf die genannten Quellen. Prinzipiell reagiert TCP empfindlich auf jegliche Art QoS-Problem, was in zusätzlicher Latenz, niedrigerem effektivem Durchsatz und höheren, wiederum auszugleichenden, Paketverlusten resultieren kann.

TCP dient als Transportprotokoll nicht nur der Sicherstellung der angestrebten Zuverlässigkeit, sondern auch zur Flusssteuerung und Überlastkontrolle. TCP ist empfindlich gegenüber Congestion und den damit verbundenen Effekten, bzw. TCP ist *congestion controlled*. TCP-Algorithmen reagieren auf Verzögerungen und Paketverluste aufgrund „verstopfter“ Pufferwarteschlangen an Flaschenhälsen durch Anpassung verschiedener Parameter, u. a. der Senderdatenrate. Congestion, also die „Verstopfung“ in Form von Paketstaus infolge mangelnder Bandbreite oder Verarbeitungskapazität, d. h. von Sättigung, kann selbst wiederum Quelle zusätzlicher Verzögerungen oder Paketverluste sein. Wie im Abschnitt zu den Netzsituationen im Mobilfunknetz (5.1) noch erläutert wird, kann vereinfacht davon ausgegangen werden, dass dieser Flaschenhals innerhalb des Mobilfunksystems im drahtlosen Zugangsnetz liegt. Die TCP-Flusssteuerung hat somit maßgeblichen Einfluss auf die effektiv erreichbaren Nutzdatenraten und die dynamische Latenz unter Last, die hauptsächlich durch die Queuing-Effekte in den Pufferwarteschlangen dynamisch vergrößert wird.

Mit zunehmender Sättigung der verfügbaren Bandbreite der Verbindung steigt die Latenz sehr stark an. Das Phänomen sehr langer Pufferwarteschlangen am Übergang von einem Netzsegment mit hohem Durchsatz zu einem Netzsegment mit vergleichsweise niedrigem Durchsatz bei gleichzeitig exzessivem Zwischenspeichern von Paketen ist als *Bufferbloat* bekannt, siehe [149]. Unter diesen Voraussetzungen sind Mobilfunknetze geradezu für Bufferbloat prädestiniert. [148] geht auf die Problematik Bufferbloat in (echten) Mobilfunknetzen im Detail ein, während sich [150] der Simulation angepasster Algorithmen und Warteschlangen mit Berücksichtigung der Delay-Summe und ggf. auch schwankender Bandbreite widmet. Als Fazit der genannten Quellen und auch eigener Vergleichsmessungen in echten Mobilfunknetzen zur Vorbereitung der Tests ist festzuhalten, dass durch Bufferbloat leicht effektive Paketumlaufzeiten von knapp unter 1 Sekunde (3G+) bis zu mehreren Sekunden (bei niedrigen Bandbreiten wie bei 2G/EDGE) erreicht werden können.

[147] S. 425f gibt dazu weitere Auskunft, welche Maßnahmen und Parameter für TCP in drahtlosen Umgebungen mit typischer Mobilcharakteristik (2,5G / 3G) vorteilhaft sein können. Für neuere Mobilfunkgenerationen sind trotz höherem Durchsatz und niedrigerer Grundlatenz die Probleme noch immer vorhanden, wie die Daten in [148] zeigen. Die dort genannten Anti-Bufferbloat-Ansätze wie Active Queue Management (AQM) oder die Reduzierung der Puffergrößen gehen über die Betrachtung der Endpunkte der Kommunikation hinaus. Sie verlagern eine mögliche Lösung in die beteiligten Netzelemente (Router) des Netzes. Dadurch erweitert sich die Rolle des Netzes in den Modellbetrachtungen, von der QoS auch hin zur QoD, was in der Modelldarstellung durch den Pfeil zur Moderationswirkung zwischen QoS und QoD angedeutet ist (siehe Abbildung 22).

3.4.5 Rolle der TCP-Implementierungen und -Parametrisierungen

Die TCP-Implementierungen und -Parametrisierungen der Endgeräte müssen im Zusammenspiel mit denjenigen der Gegenseite als relevante Leistungsfaktoren zur QoD berücksichtigt werden. Bedingt durch dieses Zusammenwirken ist eine gemeinsame Erläuterung an dieser Stelle für mobiles Endgerät und Gegenseite sinnvoll.

Es existieren zahlreiche TCP-Implementierungen, die sich in ihrem Verhalten und internen Parametern unterscheiden. Prinzipiell lassen sich *delay-based* und *non-delay-based* Implementierungen unterscheiden [151] S. 1. Zur Minimierung der Auswirkungen von hoher Latenz und Paketverlusten wurden im Laufe der Zeit zahlreiche Verbesserungen zum ursprünglichen TCP entwickelt. Diese Veränderungen sind in unterschiedlichen TCP-Implementierungen zusammengefasst, die teils klangvolle Namen wie Tahoe, Reno, Westwood, Veno, Cubic oder Fast-TCP tragen (siehe [150] S. 1299 für weitere Verweise dazu). Nach [148] sind Delay-basierte TCP-Implementierungen (wie TCP Vegas) prinzipiell resistenter gegen Bufferbloat und daher im Vergleich eventuell vorteilhaft für den Einsatz unter mobilfunktypischen Bedingungen.

Für die (stationären) Implementierungen der Gegenseite sind etwas einfachere Angaben und konkrete Zahlen zu den Implementierungen zu ermitteln als für die mobilen Endgeräte. Eine Übersicht zu den zahlreichen, alleine von Linux

unterstützten TCP-Implementierungen und TCP-Congestion-Algorithmen findet sich u. a. in [152]. Eine Art Zensus der Server der beliebtesten Internetziele ([153] und [154] sowie Update in [155]) zeigt, dass einige wenige TCP-Implementierungen⁷¹ eine weite Verbreitung aufweisen, insgesamt aber eine erstaunliche Vielfalt eingesetzter TCP-Implementierungen für die Gegenseiten festzustellen ist.

Auch auf mobiler Seite sind Betrachtungen zu den TCP-Implementierungen höchst interessant, Details aber schwierig zu ermitteln. Festzuhalten ist: Alleine aus Sicht der erreichbaren Effektivwerte ist es möglich, dass sich die Endgeräte relevant unterscheiden. Dabei ist nicht nur der Vergleich der Systemplattformen sinnvoll bzw. ausreichend, sondern ggf. für jedes einzelne Modell notwendig, da die entsprechenden Vorgabewerte modellspezifisch sein können [148]. Als Teil der Systemsoftware ist davon auszugehen, dass sich diese Eigenschaften mit neuen oder veränderten Versionen der Systemsoftware ebenfalls ändern können.

Auf die Feinheiten der TCP-Implementierungen kann an dieser Stelle nicht weiter eingegangen werden. In den späteren praktischen Untersuchungen mit den mobilen Geräten und echten Gegenseiten zeigten sich aber teilweise eklatante Differenzen in der erzielten Leistung, die auf die andiskutierten Unterschiede in den genutzten TCP-Implementierungen bzw. Parametrisierungen zurückzuführen sind (siehe 9.1.2). Hinweise zu den TCP-Implementierungen der ausgewählten Endgeräte sind im Abschnitt des Testbeds zu finden (5.8.3).

Letztlich trifft die Vielfalt der Implementierungen und Parametrisierungen der mobilen Seite auf die der Gegenseite, was in der Kombination noch mehr Vielfalt ergibt. Für eine abschließende Aussage zu den in einer Ende-zu-Ende-Kommunikation tatsächlich genutzten TCP-Implementierungen/Eigenschaften/-Parametern wäre jeweils eine eigene individuelle Untersuchung und Bewertung notwendig. Prinzipiell lassen sich die dafür notwendigen Daten aus der Überwachung bzw. der Aufzeichnung des Netzwerkverkehrs extrahieren, bzw. auf dieser Datenbasis auf die eingesetzten TCP-Implementierungen

⁷¹ BIC und die CUBIC-Abwandlungen (Linux-Server-Standard), CTCP-Abwandlungen (Windows-Server-Standard) und schließlich noch RENO, nach [155] S. 1321f

schließen. Um solche Aussagen automatisiert zu erzeugen, wird ein entsprechendes Werkzeug benötigt. [153] und [151] beschreiben ein solches Werkzeug⁷² und die Überlegungen für Entscheidungsregeln zur Feststellung der TCP-Eigenschaften. Die dort vorgeschlagene Softwarelösung ist allerdings nicht für die passive Beobachtung bzw. nachträgliche Auswertung der Kommunikation zweier unabhängiger Endpunkte konzipiert, z. B. zwischen mobilem Gerät und Server im Internet, sondern agiert selbst als ein Endpunkt um Aussagen zur jeweiligen Gegenseite zu liefern (active probing).

3.4.6 Übertragungscharakteristik

Die Übertragungscharakteristik kann in mehrere Unterkriterien aufgeschlüsselt werden, die eine sehr feingliedrige technische Beschreibung der Netznutzung eines Dienstes erlauben. Zusammen mit den bereits erwähnten Protokollen können die folgenden Punkte mittels geeigneter Werkzeuge auch zur Analyse der Netznutzung genutzt werden. Davon wurde im Rahmen der Voruntersuchung der Dienste im eigenen Test Gebrauch gemacht.

Tabelle 9: Stufen der zusammengehörigen Netznutzungen

Einheit	Stufe	Anmerkung und Beispiel zur Erläuterung
Session / Task	Sitzungsstufe	alle zusammengehörigen Transactions, die in der Nutzersicht zu einer Nutzung bzw. zu einem Nutzungsziel/Aufgabe gehören z. B. alle Seitenaufrufe in einer Web-Browsing-Sitzung (Hauptseite und Unterseiten)
Transaction	Seitenstufe	alle zusammengehörigen Netzaktivitäten in der Anwendung und Nutzersicht, die zu einer entsprechenden Nutzerwahrnehmung führen (im Sinne eines singulären Nutzererlebnisses) z. B. jeweils eine neue Webseite inklusive aller zugehörigen Elemente wie einzelner Bilder, Stylesheets, JavaScripts etc.
Turn / Element	Elementstufe	Richtungswechsel im Datenaustausch, z. B. bestehend aus einer Anfrage (Request) an einen Server und die zugehörige Antwort (Response) auf Ebene des Applikationsprotokolls z. B. ein einzelnes Bild GET/Bild.jpg HTTP/1.1 → HTTP/1.1 200 OK
Flow	Datenflussstufe	alle zusammengehörigen Pakete (von der Quelle zum Ziel), z. B. Beschreibung über Quintupel aus Quell-IP, Ziel-IP, Transportprotokoll, Quell-Port, Ziel-Port auf Ebene des Transportprotokolls
Packet	Paketstufe	Datenpakete als Träger der Nutz- und Steuerdaten auf Ebene des Internetprotokolls

Systematik angelehnt an die Beschreibung verschiedener Verkehrsmodelle in [126] S. 109ff, vom Modell der WWW-Nutzung verallgemeinert, angepasst und erweitert um den eigentlichen Datenfluss und die Nutzersicht, Beispiel für Web-Browsing

⁷² TCP Congestion Avoidance Algorithm Identification (CAAI), siehe auch [154] zum Download.

Zusammengehörige Netznutzungen: Die Datenübertragung erfolgt in IP-Netzen in Form einzelner IP-Pakete. Es ist aber oftmals sinnvoll, eine Menge zusammengehöriger Pakete als Netznutzung gemeinsam zu betrachten. Dies kann auf unterschiedlichen Stufen bzw. Abstraktionsebenen erfolgen (siehe Tabelle 9).

Die Stufen/Abstraktionsebenen lassen sich über verschiedene Parameter charakterisieren, entweder als Einzelwerte oder mittels statistischer Verteilungen und zugehöriger Maßzahlen: z. B. als mittlere oder typische Werte für Zeitdauer, Datenmenge, (Paket-) Anzahl und Größe, Paketankunftszeit bzw. mittlerer Abstand (Packet Inter Arrival Time) und hinsichtlich Gleichzeitigkeit/Parallelität bzw. Abfolge/Sequenz.

Ein entsprechendes Beispiel zur geeigneten Charakterisierung der zusammengehörigen Netznutzungen der mobilen Nutzung einer mobilen Webseite (Spiegel.de) inklusive passender Visualisierung als Waterfall-Diagramm findet sich in der Beschreibung des Testparcours weiter hinten in der Arbeit, siehe Abbildung 101. Die zusammengehörigen Netznutzungen auf Ebene der Nutzersicht werden in der Diskussion des Betrachtungsrahmens der Dienstklassen noch einmal aufgegriffen (3.6.7).

Lebensdauer und Datenmengen: In Abhängigkeit der typischen Lebensdauer und übertragenen Datenmenge eines Flows wird in Analogie zur Tierwelt zwischen Maus und Elefant unterschieden (nach Mice&Elephant in [156] S. 15ff):

- Maus: viele kurzlebige TCP-Flows, nur jeweils wenige Daten
- Elefant: wenige langlebige TCP-Flows, jeweils große Datenmengen

Einige Dienste weisen fast nur viele kurzlebige TCP-Flows auf (z. B. Web-Browsing mit jeweils nur wenigen kB), andere Dienste können Gigabytes mit nur einem einzigen Flow übertragen (z. B. Dateitransfers oder Streaming). Besonders für „Maus-Dienste“ mit vielen konsekutiven TCP-Flows steigt die Rolle der Latenz für die effektive Kommunikationsleistung stark an.

Zeitlicher Verlauf und Zustellungstyp: Die Charakterisierung der Netznutzung im zeitlichen Verlauf erlaubt eine Unterscheidung der Dienste über den

Zustellungstyp⁷³, die auch zur Klassenbildung eingesetzt werden kann (nach [157] S. 158ff):

- Isochron (kontinuierliche und zeitlich konstante Übertragung), z. B. VoIP
- Streaming (kontinuierliche Übertragung, zeitliche Schwankungen werden durch Puffer ausgeglichen), z. B. Video-Streaming
- Burst (stoßartig, kammartig, unregelmäßig, in einem Schub, schnellstmöglich), z. B. Web-Browsing

Neben diesen starren Mustern ist im zeitlichen Verlauf auch eine adaptive Nutzung möglich, z. B. wenn von höheren Protokollschichten (in den Anwendungen) eine Anpassung der übertragenen Inhalte an die festgestellte, aktuell mögliche Übertragungsleistung erfolgt. Wenn Dienste kontinuierlich über längere Zeit viele Turns verursachen, z. B. VPN und Online-Games etc., dann sind sie „geschwätzig“ (chatty, Begriff aus [158] S. 24).

Richtung und Symmetrie: Die Asymmetrie zwischen Empfangen und Senden von Daten, respektive unterschiedliche Anforderungen an den Datenkanal zum/vom Endgerät sind zu beachten. Die Betrachtung der Symmetrie und Hauptrichtung ist von besonderer Bedeutung, da die technischen Charakteristiken der zugrundeliegenden Netze (wie im Mobilfunk) oftmals auch asymmetrisch sind, mit höherer Leistung in Down-Richtung als in Up-Richtung. Die Richtung zum betrachteten Endpunkt/Endgerät hin soll dabei als Down (-stream/-link/-load) und die Richtung vom betrachteten Endpunkt/Endgerät weg als Up (-stream/-link/-load) bezeichnet werden.

Ein Dienst nutzt das Netz entweder bevorzugt mit einer Hauptrichtung (unidirektional) oder in beide Richtungen (bidirektional). Sind die Datenmengen in beide Richtungen annähernd ausgeglichen, dann ist die Netznutzung als symmetrisch zu bezeichnen, ansonsten als asymmetrisch. Die Asymmetrie kann, bei ansonsten technischer Gleichheit, ein Unterscheidungsmerkmal zwischen „Kommunikation“ und „Streaming“ sein:

- Kommunikation: meist annähernd symmetrisch
- Streaming: meist asymmetrisch mit Hauptrichtung üblicherweise im Downstream

⁷³ Begriff aus [45] S. 62 entnommen

Anzahl Kommunikationspartner: Ein weiteres Unterscheidungskriterium ist die Anzahl der Kommunikationspartner. Dabei ist in der Nutzung des mobilen Internets (über Mobilfunktechnik) zu unterscheiden zwischen:

- Unicast mit 1:1-Beziehung, d. h. ein Absender und ein Empfänger
- Multicast mit 1:n-Beziehung, d. h. ein Absender und eine Gruppe von mehreren Empfängern

Multicast ist prinzipiell für Verteildienste in Content-to-Person-Szenarien interessant, besonders in einer Beziehung 1:n, vom Netz zu mehreren Nutzern parallel, z. B. für Live-Mobile-TV mittels Multimedia Broadcast Multicast Service (MBMS) [56] S. 376ff, aber in Produktivnetzen (bislang) ohne Bedeutung. Im Weiteren wird daher nur von Unicast ausgegangen, was auch einfacher in die skizzierte Ende-zu-Ende-Betrachtung des Netzes passt.

3.5 Die Dienste – Teil 3: Nutzersicht

Die Nutzersicht auf die Dienste bildet die Brücke zur interdisziplinären Betrachtung des QoS-QoE-Zusammenhangs. Dabei treffen der technisch geprägte Teil der Netzwelt, die Mensch-Maschine-Schnittstelle der HCI-Welt, der Inhalt der Multimedia-Welt und die internen Prozesse des Menschen (Psychologie und Kognitionswissenschaft) zusammen. In der Nutzersicht spielen Komponenten der nutzerseitigen Hardware und Software eine wichtige Rolle. Einige der nachfolgenden Punkte sind daher an anderer Stelle in der Arbeit schon einmal thematisiert worden, z. B. beim Endgerät oder den Apps. Im Gegensatz zur Diskussion dort, wo möglichst dienstunspezifisch und isoliert von einem Element als Ursache des Einflusses auf die QoX-Matrix ausgegangen wird, soll die Nutzersicht die dienstspezifischen Einflüsse aller Elemente zur Wirkung auf den Menschen als Nutzer zusammenfassen. Ergänzend sind die vorherigen Ausführungen zum Kontext mit Nutzungsziel und Besonderheiten der mobilen Nutzung zur Vervollständigung der mobilen Nutzersicht zu beachten.

Die bei der Nutzung von Diensten über die Applikation wahrgenommene Leistung der QoA ist gemäß dem Modell von der QoD und somit auch von der QoS abhängig. Für diese Applikationsleistung werden drei Kategorien von unterscheidbaren Größen zur Beschreibung eingeführt, die Nutzerstimuli. Neben

den Nutzerstimuli ist auch ihre Präsentation wichtig, d. h. neben dem *Was* ist auch das *Wie* der Darstellung zur Wahrnehmung relevant, z. B. in Form von Anzeigen zur Wartezeit oder Indikatoren. Die Interaktivität der Dienste in der Nutzersicht wird explizit angesprochen, weil diese quasi als eine Art Auslöser ein wichtiger Teil des dienstspezifischen Nutzererlebnisses sein kann. Zunächst soll mit dem QoS-unabhängigen Teil der Nutzersicht auf die mobilen Apps begonnen werden, mit der Ergonomie und Usability der Benutzerschnittstelle.

3.5.1 Software-Ergonomie und Usability der Nutzerschnittstelle

Es wird im Weiteren davon ausgegangen, dass zur Nutzung der Dienste eine App oder eine (mobile) Webseite im Browser etc. zum Einsatz kommt. Besonders mit Blick auf diesen anwendernahen Teil der Dienstnutzung sind Fragen nach der Ergonomie und dabei besonders hinsichtlich der Usability und Gestaltung der Nutzeroberfläche grafischer Dialogsysteme unausweichlich. Diese sind zweifelsohne wichtig und in der HCI-Forschung auch schon gut untersucht, allerdings gleichzeitig auch so umfangreich, dass auf eine umfassende Darstellung an dieser Stelle verzichtet werden muss. Ausführliche Erläuterungen zu den Grundlagen der Software-Ergonomie sind beispielsweise in [159] S. 27ff. und [73] S. 25ff. zu finden.

Aspekte der ergonomischen, d. h. menschengerechten Gestaltung und Gebrauchstauglichkeit (Usability) sollen im Weiteren aber als eine Art Grundanforderungen angesehen werden (Pflicht). Ist eine App quasi „unbenutzbar“, weil schlecht konzipiert, kompliziert zu bedienen, gegen alle Gestaltgesetze und mentalen Modelle der Nutzer entworfen, dann ist dies kein Problem mit einem Bezug zur QoS bzw. möglichen QoS-abhängigen Einflüssen vorgelagert. In diesen Fällen dürften die weiteren Betrachtungen aber ohnehin überflüssig sein.

Für die Umsetzung der Erkenntnisse zur Software-Ergonomie existieren zahlreiche Normen, Gesetze, Richtlinien und Best-Practice-Empfehlungen (siehe [159] S. 130ff und S. 150ff). Diese sollten bei Entwurf und Realisierung von Apps bzw. allgemeiner von mobilen Diensten eingehalten werden. Dabei stellt der mobile Nutzungskontext besondere Herausforderungen durch (relativ)

kleine Bildschirme zur Ausgabe und eingeschränkte Möglichkeiten zur Eingabe bei der Interaktion. Die ETSI TR 102 972 gibt Hinweise zur Gestaltung von Elementen in den Nutzeroberflächen mobiler Geräte, Apps und Dienste im Allgemeinen [160] (kurz zusammengefasst in [161]). Es wird darin erläutert, wie mobile Dienste/Apps mit den Einschränkungen mobiler Netze umgehen sollten und wie dies dem Nutzer zu kommunizieren ist, allerdings auf dem Stand der Anfangszeit der 3G-Evolution. Für die Gestaltung moderner mobiler Apps geben eher die Richtlinien der großen Plattformen wie Android⁷⁴ oder iOS⁷⁵ praktisch anwendbare Hinweise zum Design und zur konkreten Umsetzung der Benutzerschnittstelle.

Für die weitere Erörterung interessieren indes einige Facetten der Gestaltung der Nutzerschnittstelle ganz besonders: Wie werden in der Nutzersicht die QoS-abhängigen Stimuli dargestellt, d. h. wie werden sie präsentiert bzw. wie sind sie in der Nutzersicht zu beobachten? Wie sind die QoS-abhängigen Nutzerstimuli dabei in die Interaktionen der Nutzersicht eingebettet? Diese Punkte schließen an die Diskussion der Nutzerstimuli gesondert an (in 3.5.8). In der Voruntersuchung echter Dienste wird an späterer Stelle zur Beurteilung der Teseignung entsprechend in die Facetten der Benutzbarkeit (App/Dienst allgemein und losgelöst von der QoS) und der Beobachtbarkeit der QoS-abhängigen Leistungsänderungen (Stimuli) unterschieden (siehe Diskussion in 6.3.3 und 6.3.4.).

3.5.2 Applikationsleistung und Nutzerstimuli

Nachfolgend sollen die Nutzerstimuli als eigener Begriff der Qualitätsmerkmale für die Charakterisierung der Applikationsleistung der QoA genutzt werden. Sie sind mess- und wahrnehmbare Reizgrößen (Signale), aus denen sich maßgeblich das Nutzerlebnis (Experience) ergibt. Die weitere Betrachtung soll sich dabei auf audiovisuelle Stimuli konzentrieren.

⁷⁴ Android Design Guidelines <http://developer.android.com/design/index.html> abgerufen am 06.01.2015

und Android UI Guidelines <http://developer.android.com/guide/topics/ui/index.html> abgerufen am 06.01.2015

⁷⁵ iOS Human Interface Guidelines <https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/index.html> abgerufen am 06.01.2015

Zur Ursache: Die Nutzerstimuli sind dienstspezifisch und stehen somit zwischen Netz (Teil der Ursache) und Nutzer (Wirkung). Mindestens ein Teil der Nutzerstimuli ändert sich QoS-QoD-abhängig. Sie sind somit „der spürbare Unterschied“ in der Nutzersicht zwischen unterschiedlichen Netzsituationen bezogen auf eine spezifische Nutzungssituation. Das Entstehen und der Zusammenhang der Nutzerstimuli untereinander wird nachfolgend noch erläutert, zunächst sollen aber die Konzepte zur Beschreibung und Wirkung dargelegt werden, damit diese richtig eingeordnet werden können.

Zur Beschreibung: In der vorliegenden Arbeit wird davon ausgegangen, dass sich mit drei Hauptkategorien die QoE-relevanten Stimuli bei der Nutzung mobiler Dienste im mobilen Internet charakterisieren lassen, siehe Tabelle 10. Dass dabei jeweils immer zwei Begriffe kombiniert auftreten, ist gewünscht, so wie die zwei Seiten einer Medaille zusammengehören. Die Auflistung erhebt nicht den Anspruch auf Vollständigkeit, ist aber so abstrakt angelegt, dass die in der eigenen Voruntersuchung der Dienste identifizierten und nachfolgend als relevant erachteten Reizgrößen einzuordnen und zu beschreiben sind. Eine ausführliche Diskussion zu jeder der drei Kategorien folgt nach.

Tabelle 10: Kategorien der Nutzerstimuli mit Beispielen

Kategorie	Beispielmetrik	objektiv/ instrumentell (Messung)	subjektiv/ menschlich (Wahrnehmung)	QoE-Einfluss- charakteristik
F&V Funktion und Verhalten	Stalls (Buffer- Underruns), Ausset- zer, Abbrüche, Fehlermeldungen	(Strich-) Liste: Anzahl, Verteilung (Distanz)	Zählen des Auftre- tens/Abschätzen, Beurteilung der Schwere von Beein- trächtigungen	sprunghaft, hart
Q&Q Qualität und Quantität	Videoqualität Videodatenrate (Quantität)	Analyse / technische Metadaten: Auflösung (Pixel x Pi- xel), PSNR (dB), SSIM kBit/s	Sinnesorgane: Bildein- druck, Schärfe, Arte- fakte als Qualität Mengen, Werte, Schätzungen als Quantität	stetig oder in Stufen
Z&D Zeiten und Dauern	Ladedauer, Reakti- onszeit	(Stopp-) Uhr: Sekunden	innere Uhr: Zeitgefühl und Schät- zen	Wartedauern: stetig und kontinuier- lich, weich Deadlines: hart

PSNR: Peak Signal Noise Ratio; dB: Dezibel; SSIM: Structural SIMilarity

Je nach Dienst sind unterschiedliche Ausprägungen, Kombinationen und Relevanz (Gewichtung) konkreter Größen und Metriken aus den drei Kategorien möglich, mit denen sich die nutzerrelevante Applikationsleistung erfassen und modellieren lässt. Die Nutzerstimuli lassen sich sowohl objektiv als auch subjektiv erfassen und beschreiben:

- objektiv/instrumentell als technische Messung, als nutzerrelevanter Erwartungswert bzw. Prädiktor für den QoS-QoE-Zusammenhang
- subjektiv als wahrgenommene Leistung, die zum Nutzererlebnis beiträgt und von der die gesuchte subjektive Bewertung (direkt) abhängt

[118] S. 6 beschreibt einen Ansatz in ähnlicher Art zur Gewichtung der direkt auf die QoE einwirkenden Einflussfaktoren, der Key Quality Indicators (KQI).

Zur Wirkung: Für die Nutzerstimuli können verschiedene Charakteristiken angenommen werden, wie sie sich selbst QoS-abhängig ändern und wie sie auf die QoE einwirken:

- stetig und kontinuierlich in einem Kontinuum, z. B. bei Wartedauern, woraus eine weiche QoE-Änderung resultieren sollte (elastisch)
- diskret, in festen Stufen, z. B. Unterbrechungen beim Abspielverhalten oder unterschiedliche Videoqualitätsstufen, wofür sprunghafte QoE-Änderungen zu erwarten sind (unelastisch)

Die Nutzerstimuli der Applikationsleistung sind direkte Einflussfaktoren für die QoE. Die unterschiedlichen Arten der Änderungscharakteristik und der QoE-Einflusscharakteristik der Nutzerstimuli können in ihrer Überlagerung die unterschiedlichen Charakteristiken des QoS-QoE-Gesamtzusammenhangs vorhersagen und erklären. Dominiert dabei eine sprunghafte Änderung der Stimuli, so ist auch eine sprunghafte QoE-Änderung plausibel. Wird die sprunghafte Änderung der Stimuli selbst wiederum durch eine nur geringfügige QoS-Änderung verursacht, so ist ein ebenfalls enger und sprunghafter QoS-QoE-Übergang plausibel (weitere Diskussion dazu in der Gesamtsicht unter 3.6.3).

3.5.3 Funktion und Verhalten

Es wird davon ausgegangen, dass es aus Nutzersicht für einen Dienst eine bestimmte erwartete Funktion bzw. ein entsprechend erwartetes Verhalten gibt. Sind diese nicht beeinträchtigt, so sollen sie als normale Funktion und Verhalten (der Anwendung, des Systems) angesehen werden. Ausgehend davon ist bei Verschlechterung der QoS-Netzparameter eine Beeinträchtigung von Funktion und Verhalten möglich. Als implizite Aspekte sind Zuverlässigkeit bzw. die Toleranz gegenüber dem Auftreten von Fehlern und die Fehlerfreiheit in der Kategorie Funktion und Verhalten enthalten. Prinzipiell ist davon auszugehen, dass die Nutzer an Zuverlässigkeit und Fehlerfreiheit interessiert sind und diese auch aus technischer Sicht für eine Vielzahl von Diensten notwendig sind. Diese Kategorie ist damit eigentlich schon am Rande der Betrachtungen, da ja grundsätzlich von der Ebene der Integrity (siehe 2.2.3) ausgegangen werden sollte, was bedeutet, dass die Möglichkeit zur vollständigen Dienstnutzung, wenn auch mit Einschränkungen, anzunehmen ist.

Beispiele für Nutzerstimuli der Kategorie Funktion und Verhalten sind:

- fehlende Funktion generell oder eingeschränkte Zuverlässigkeit
- Abspielverhalten: Aussetzer, Hänger, Nachpuffern, Buffer-Underruns, Stalls
- Fehler und Fehlermeldungen (besonders zu Netz oder Verbindung)

Diese Beeinträchtigungen sind mit bestimmten Ereignissen verbunden, z. B. ist ein Puffer leergelaufen oder eine Fehlermeldung erschienen. Aus der Sicht eines Nutzers passieren sie eher von selbst oder im Eindruck der Wirkung eines (vermeintlich plötzlichen) externen Einflusses – z. B. schlechtes Netz / „Funkloch“. Bei geeigneter Modellierung kann das Auftreten bzw. der Zeitpunkt des Ereignisses aber teilweise recht genau berechenbar sein, besonders wenn sie mit festen technischen Anforderungen, z. B. an den Durchsatz als notwendige minimale Datenrate, einhergehen. Das Leerlaufen des Puffers beim Videostreaming kann so beispielsweise bei bekannten Parametern wie Inhaltsdatenrate und Pufferinhalt im Abgleich zur Übertragungsdatenrate vorhergesagt werden (eigene Untersuchungen dazu siehe Anmerkung in 8.1.1).

Nutzerstimuli der Kategorien Funktion und Verhalten sind entweder binär (z. B. hat funktioniert vs. hat nicht funktioniert), diskret quantifizierbar (abzählbar), als Verhältniszahlen oder als Verteilungsmuster zu charakterisieren. Es ist vom Wirkprinzip her eine besondere Schwere der Kategorie Funktion und Verhalten zu vermuten, woraus sich ein großer Einfluss auf die QoE-Bewertung erwarten lässt. Für besonders schwerwiegende Merkmale ist daher ein eher harter und abrupter Übergang zwischen Zufriedenheit und Unzufriedenheit innerhalb eines engen Bereiches anzunehmen.

3.5.4 Qualität und Quantität

Objektive Eigenschaften und subjektive Wahrnehmung des Inhalts lassen sich in Qualität und Quantität unterscheiden, die je nach Anwendung oder Dienst in verschiedener Form auftreten. Die Inhaltseigenschaften zur Qualität und Quantität können mit passenden objektiven Maßen beschrieben werden. Dabei ist eine Unterscheidung je nach Medium (statisch vs. dynamisch bzw. Text, Bild, Audio, Video) hinsichtlich der relevanten Metriken möglich und sinnvoll, siehe Tabelle 11. Die Maße zur subjektiven Wahrnehmung korrespondieren mit den objektiven Größen. Quantitative und qualitative Eigenschaften hängen oftmals direkt miteinander zusammen, z. B. eine bessere Bildqualität bei einer größeren Dateigröße oder Datenrate. An anderer Stelle wird die QoE teilweise nicht auf das Nutzungserlebnis allgemein bezogen, sondern tatsächlich die Qualität des Mediums (Media Quality), siehe die Anmerkungen zur Qualitätswahrnehmung und Wahrnehmungsqualität unter 4.1.6.

Tabelle 11: Qualität und Quantität unterschiedlicher Medien des Inhalts - objektiv vs. subjektiv

	objektiv	subjektiv
<i>statische Medien, z. B. Bilder und Texte</i>		
Qualität	spatiale Auflösung (Pixel), Farbtiefe, Kompressionsfaktor	Bildschärfe (Detailauflösung), Farbtreue, Kontrastumfang, Artefakte
Quantität	Datenmenge (für die einzelnen Elemente und insgesamt), Anzahl Elemente, z. B. Anzahl eingebundene Bilder, Textmenge	(sichtbare) Bildgröße, Textmenge und Menge weiterer Elemente, Anzeigen
<i>dynamische Medien, z. B. Audio und Video</i>		
Qualität	zeitliche Auflösung (Bildrate)	Video z. B. Bewegungswiedergabe Audio z. B. Verständlichkeit, Wiedergabetreue
Quantität	Datenrate (Datenmenge pro Zeiteinheit), Abspielänge	Anzeigen, Indikatoren

Objektive Qualitätsmetriken wie Peak Signal to Noise Ratio (PSNR), Structural SIMilarity (SSIM), Moving Pictures Quality Metric (MPQM) oder Media Delivery Index (MDI), liefern zwar Werte für automatisierte Tests, siehe [162], ihre eigentliche Nutzerrelevanz im Sinne einer Qualität ist allerdings nicht immer ganz klar. Im besten Fall liefern sie eine Aussage zur Qualität als gradueller Abgleich zwischen Soll (Charakterisierung der relevanten Merkmale) und Ist (Bezugswerte zur Einordnung), die in einem direkten Zusammenhang zur nutzerrelevanten Qualität stehen. Ohne externen Maßstab liefern sie ansonsten nur eine technische Kennzahl.

Die objektiven Eigenschaften führen, in Abhängigkeit weiterer Faktoren, z. B. der Hardware des Endgerätes (Bildschirmgröße, Displaytechnologie etc.), der Systemsoftware und der App, zur subjektiven Wahrnehmung des Inhalts. Neben den unmittelbar sichtbaren Ausprägungen können auch Hinweise auf Qualität und Quantität sowohl das Nutzererlebnis als auch die Erwartung beeinflussen:

- Indikatoren zur Qualität (z. B. Einblendung „HD“)
- Hinweise zur Inhaltsversion, z. B. „Mobilversion“
- Indikatoren/Anzeigen zur Quantität (z. B. Dateigröße oder Zähler zur Datenmenge etc.)

Durch diese Doppelwirkung lassen sich a priori nur schlecht Abschätzungen zur genauen Wirkung bzw. Relevanz in der Nutzersicht treffen. Besonders die Unterscheidung in Begeisterungs-, Leistungs- und Basisfaktoren zur Zufriedenheit ist dabei jeweils zu beachten. Für die Charakterisierung der Ausprägungen von Qualität und Quantität und die weitere Wirkung zur Formierung der QoE-Bewertung soll gelten:

- *kontinuierlich* in unendlicher Anzahl Varianten (z. B. bei Dateigrößen):
Es ist von einem eher weichen Übergang zwischen Zufriedenheit und Unzufriedenheit in der Beurteilung zur QoE auszugehen. Für die weitere Betrachtung kann eine Bildung von Klassen sinnvoll sein (z. B. sehr kleine Dateien <10 kB, kleine Dateien <100 kB, mittelgroße Dateien <1000 kB, große Dateien, sehr große Dateien (Zahlen als Beispiel)).

- *diskret* mit überschaubarer Anzahl Varianten (z. B. bei Inhaltsversionen, mobile Version vs. Desktop-Version, HQ vs. HD): Es sind harte Übergänge bzw. große Unterschiede in der QoE-Bewertung plausibel, sofern die Abstände in der Wahrnehmung groß genug sind.

3.5.5 Zeiten und Dauern

"Time is measured objectively, but perceived subjectively [...]"

Ilya Grigorik⁷⁶ in [164] Chapter 10

Die Bestimmung von Zeiten und Dauern ist gekennzeichnet durch verschiedene Zeitpunkte: Anfang und Ende, Anfrage und Antwort, Aktion und Reaktion. Der letzte Punkt führt zum Zusammenhang mit der Interaktivität der Dienste und der Interaktion der Nutzer mit diesen.

Mögliche Arten von Zeiten und Dauern sind beispielsweise:

- Dauer bis etwas startet (time to start), z. B. Wiedergabestartzeiten
- Dauer bis etwas endet (time to finish), z. B. Ladedauern
- Antwortzeiten/Reaktionszeiten/Verzögerungszeiten

Die objektiven Werte sind prinzipiell kontinuierlich quantifizierbar. Für die subjektive Wahrnehmung ist entscheidend, ob die jeweiligen Zeitpunkte für die Nutzer zu differenzieren sind.

Zur angenommenen Wirkung: Für längere Dauern, z. B. Ladedauern, ist von einem eher weichen und kontinuierlichen Übergang zwischen Zufriedenheit und Unzufriedenheit über einen weiten Bereich auszugehen. Bei kurzen Reaktionszeiten im Echtzeitbereich sind abrupte Änderungen in einem engen Bereich möglich.

Die Zeitwahrnehmung und ihre Beurteilung spielen auch außerhalb der QoE-Domäne in der Psychologie als Psychophysik und bei der Mensch-Maschine-Interaktion (HCI) eine große Rolle. Zu den wichtigsten Erkenntnissen zur Zeit

⁷⁶ mit Verweis auf Jakob Nielsen: "Usability Engineering" [81] und Steven Seow "Designing and Engineering Time" [163] für weiterführende Diskussionen zur Zeit und Zeitwahrnehmung

aus der Psychologie und zugehörigen Annahmen zur subjektiven Wahrnehmung zählen (siehe z. B. [165]):

- Menschen haben in ihrer subjektiven Zeitwahrnehmung nur eine begrenzte Genauigkeit (Absolutwert) und Auflösung (Differenzen).
- Zeit vergeht relativ, in Abhängigkeit weiterer Reize und des inneren Zustands (interessant vs. langweilig).

Daneben sind die allgemeinen Anmerkungen zur Nichtlinearität menschlicher Wahrnehmung externer bzw. physikalischer Reizgrößen zu berücksichtigen, die ggf. ebenfalls, zumindest in bestimmten Bereichen, zum Tragen kommen können. Diese Punkte beziehen sich hauptsächlich auf die „gefühlte“ Zeit als die Abschätzung verstrichener Zeit auf Basis der inneren Uhr. Auf bedeutsame Zeitmarken und sinnvolle Abschätzungen zu Ober- und Untergrenzen (aus der HCI) wird in der Diskussion der Dienstklassen in 3.6.8 noch einmal zurückgekommen.

Übertragungszeit vs. Nutzungszeit: Ein mögliches Unterscheidungskriterium für Dienste, das zwischen Nutzerorientierung und Transportorientierung einzustufen ist, bzw. diese miteinander verbindet, ist das zeitliche Verhältnis von Übertragung zu Nutzung. Tabelle 12 schlüsselt einige der Aspekte auf, die nachfolgend noch weiter vertieft werden.

Tabelle 12: Zeiträume und Sequenzialität von Übertragung und Nutzung in der Nutzersicht

Übertragung zu Nutzung	Übertragungszeitraum zu Nutzungszeitraum		
	<< oder < z. B. Web-Browsing	=	> oder >> z. B. Download eines Fotos
gekoppelt vor	eher akzeptabel	interaktiv, mit aktiver Wartezeit	eher störend
während	-	Echtzeit / Streaming	-
entkoppelt vor oder nach	Hintergrund ohne aktives Warten		

Wahrnehmung von Wartezeiten: Grundsätzlich ist zu klären, ob und wie der Nutzer die Datenübertragung überhaupt erlebt. Für die Relevanz dieses Punktes ist deshalb zu beachten, ob die Übertragungszeit aus Sicht der Nutzer ein Teil der Kernnutzung eines Dienstes ist, d. h. vom Nutzer aktiv wahrgenommen wird, oder ob er gedanklich die eigentliche Nutzung des Dienstes von der Datenübertragung als Nutzung des Netzes entkoppelt. Dies steht damit auch im Zusammenhang mit der Unterscheidung von Vordergrund zu Hintergrund. Erst so wird klar, ob überhaupt Wartezeiten entstehen, die der Nutzer ggf. auch tatsächlich als störend empfindet. Dabei ist wiederum der Kontext einer mobilen Nutzung zu berücksichtigen. Eine intuitiv einleuchtende und kurze Zusammenfassung zur Problematik der Wartezeiten dazu ist:

“Mobile experiences fill the gaps while we wait. Nobody wants to wait while they wait.”

Mike Krieger (Mitgründer von Instagram) in [166] S. 16f

Es ist plausibel, davon auszugehen, dass nicht nur absolute Werte eine Rolle spielen, sondern auch Verhältnisse, wie von Wartezeit zur eigentlichen Nutzung, z. B.:

- 30 s Wartezeit für Download eines Bildes um dieses dann maximal 5 s zu betrachten → schlechtes Verhältnis
- 10 s Wartezeit vor Beginn der Wiedergabe eines Videos um dieses dann für 3 min zu nutzen → schon eher angemessenes Verhältnis

Die Überlegungen lassen sich auch auf die Sequenz und Parallelität bzw. verallgemeinert auf die Kopplung von Übertragung und Nutzung ausdehnen:

- gekoppelt, erst Übertragung, dann Nutzung: ggf. mit aktivem Warten im Vordergrund: Beeinträchtigungen durch lange Wartedauern sind möglich.
- gekoppelt, Nutzung noch während der Übertragung im Vordergrund: Echtzeit, „Live“: Beeinträchtigungen durch Aussetzer und Unterbrechungen oder der Verlust der zeitlichen Integrität sind möglich.
- entkoppelt: Übertragung im Hintergrund, die Nutzung ist davon losgelöst und erfolgt zuvor oder danach, d. h. ohne bewusstes Warten: Beeinträchtigungen entstehen lediglich bei fehlender Funktion und Übertragungsabbrüchen etc.

Vordergrund vs. Hintergrund (Fokus): Eine grobe Unterscheidung zwischen (inter-) aktiver Nutzung im Vordergrund und einer eher passiven Nutzung im Hintergrund erscheint sinnvoll. Passiv schließt in diesem Fall auch automatisiert ablaufende Datenübertragungen ein. Es ist plausibel, für eine ohnehin unbemerkte Nutzung im Hintergrund von relativ wenigen und vermutlich unkritischen Anforderungen auszugehen. Daher kann auch eine Strategie beim Entwurf von Diensten sein, die Übertragung von der eigentlichen Nutzung technisch zu entkoppeln – oder anders formuliert: *“move bits when no-one’s watching”* [166] S. 68. Wie dies in der Praxis aussehen kann, illustriert [166] S. 69 am Beispiel einer App zum Teilen von Fotos (Instagram). Im Rahmen des Testparcours der eigenen Studie war ein solcher Ansatz zur Optimierung der Nutzersicht auch im Szenario Facebook festzustellen, siehe Erläuterungen in 9.2.3.

Entsprechend stellt sich noch eine interessante Frage: Was machen die Nutzer (normalerweise), falls doch einmal längere Wartezeiten (im Vordergrund) auftreten? Zumindest für den Fall des Web-Browsings, allerdings am PC, gibt es Hinweise aus der Literatur (nach [167] S. 5, Auflistung in absteigender Reihenfolge der Nennungen). Die Nutzer...:

- warten tatsächlich „aktiv“ im Vordergrund
- schicken die Seite in den Hintergrund und lenken sich mit anderen Anwendungen ab, bzw. arbeiten an anderer Stelle weiter
- checken schnell mal E-Mails
- öffnen im Browser einen anderen Tab oder ein anderes Fenster
- suchen sich eine andere Beschäftigung abseits des Geräts
- brechen das Warten ab (Schließen der Seite)

Es ist also festzustellen, dass die Nutzer durchaus versuchen, lange Wartezeiten zumindest gefühlt zu verkürzen, in dem sie die Datenübertragungen in den Hintergrund verlagern und sich anderen Vordergrundaktivitäten widmen. Moderne Smartphones sind zwar mittlerweile Multitasking-fähig, d. h. mehrere Anwendungen sind (zumindest scheinbar) parallel auszuführen, ob sich diese Erkenntnisse allerdings direkt auf einen mobilen Nutzungskontext übertragen lassen, konnte auf vergleichbarer Basis nicht ermittelt werden. Die Anmerkungen zum mobilen Kontext und zum Warten während des Wartens lassen daran aber Zweifel aufkommen.

3.5.6 Entstehen und Zusammenhang der Nutzerstimuli

Grundsätzlich ist ein dienstspezifischer Ursache-Wirkungs-Zusammenhang zur Erklärung der Entstehung der Nutzerstimuli nutzbar, ausgehend von der QoD zur QoA, moderiert durch die Qualitätselemente Endgerät, App und Inhalt. Neben den QoS-veränderlichen Größen ist so auch ein QoS-unabhängiger Anteil anzunehmen. Dieser QoS-unabhängige Teil der Applikationsleistung ist dabei in der Leistung des mobilen Endgerätes (Hardware/Software) oder der App (algorithmische Umsetzung) zum Dienst begründet bzw. im Fall des Inhalts ggf. auch als konstant anzusehen.

Bei der Erklärung zur Entstehung der Nutzerstimuli können die einzelnen Kategorien eigentlich nicht isoliert voneinander betrachtet werden. Sie stehen miteinander und weiteren technischen Faktoren in Zusammenhang. Für jede der Kategorien sind die jeweils anderen beiden als Bezugsgrößen anzunehmen:

- **Zeiten und Dauern:** Die Ladedauer ergibt sich aus der Datenmenge des Inhalts und der effektiven Datenrate in die jeweilige Übertragungsrichtung als wirksame Netzeigenschaft. Antwortzeiten sind dabei durch die Latenz nach unten begrenzt.
- **Funktion und Verhalten** stehen in Zusammenhang mit Qualität und Quantität des Inhalts (Anforderungen, z. B. Inhaltsdatenraten) und zeitlichen Aspekten (Vorpufferzeit, Deadline / harte Echtzeit) zusammen mit den wirksamen Netzeigenschaften, die z. B. zu Funktionseffekten wie Fehlermeldungen oder gestörtem Abspielverhalten führen. Die Qualitätswahrnehmung kann durch diese funktionsbedingten Effekte beeinträchtigt werden.
- **Qualität und Quantität:** Eine hohe Qualität ist wiederum oft verbunden mit entsprechender Quantität (Datenmenge bzw. Datenrate, z. B. pro Sekunde Video), was aber ggf. auf Kosten der Zeit zum Vorpuffern und die Anzahl der Unterbrechungen bei der Wiedergabe (Streaming) oder die benötigte Dauer zur vollständigen Datenübertragung geht.

Beim Zusammenhang der Nutzerstimuli der QoA ist aus Nutzersicht vordergründig nicht ihr technisches Entstehen, sondern ihre weitere Wirkung zu berücksichtigen, da die internen Zusammenhänge einem naiven Nutzer verborgen bleiben. Im Fall von Einschränkungen, so dass nicht alle Anforderungen

vollumfänglich erfüllt werden können, führt der Zusammenhang der Nutzerstimuli zur generellen Abwägung aller relevanten Kriterien gegeneinander.

Entsprechend ist für jeden Dienst einzeln abzuwägen, welche Kriterien wichtiger als andere sind:

- Verletzung der zeitlichen Kriterien (Zeiten und Dauern)
- Verletzung der Fehlerfreiheit (Funktion und Verhalten)
- Verminderung der Qualität (Qualität und Quantität)

Letztlich wird dann auf Qualität und Fehlerfreiheit verzichtet, wenn dies verkraftbar ist und die zeitliche Integrität im Vordergrund steht, z. B. bei Voice over IP (VoIP) und Live-Videostreaming. Ansonsten wird angenommen, dass Funktion und inhaltliche Integrität wichtiger sind. Abbildung 27 illustriert diesen Zusammenhang der Nutzerstimuli für ein ausgewähltes Beispiel. Ist unklar, welche Relevanz die einzelnen Stimuli haben, so kann versucht werden, durch geschicktes Testen eine Hierarchie zu ermitteln. So kann sich beispielsweise beim Streaming ergeben: Unterbrechungsfreiheit vor Vorpufferzeit vor Qualitätsstufe.

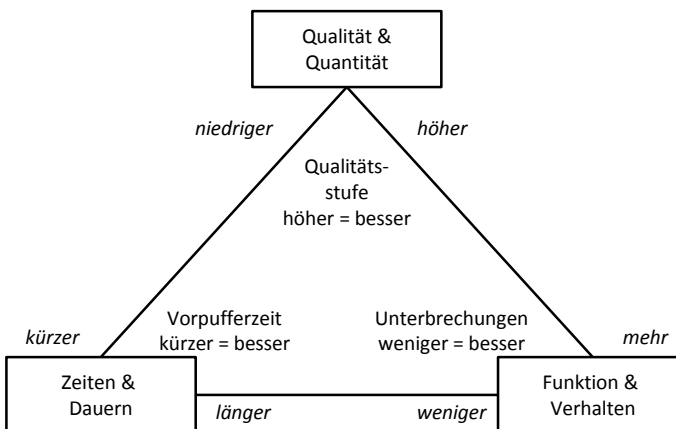


Abbildung 27: Zusammenhang der Stimuli und sich ergebender Zielkonflikt am Beispiel von Streaming

{weniger, kürzer, höher} sollte zu jeweils besseren Bewertungen führen,
 {mehr, länger, weniger} sollte zu jeweils schlechteren Bewertungen führen

Beim Entwurf und Anbieten von Diensten sind diese Überlegungen zur Nutzersicht explizit zu berücksichtigen, als eine Art QoE-Engineering bzw. QoE-Management. In der tatsächlichen Sicht eines Nutzers wird diese Abwägung (sofern überhaupt bewusst) eher weniger formal und möglicherweise auch unter anderen individuellen Gewichtungen ablaufen (siehe Erläuterungen zur menschlichen Kognition in 2.3.5). In der eigenen Studie wird versucht, sich diesem Punkt mit der Auswertung der Nutzerkommentare und des Bewertungsverhaltens zu nähern (siehe 8.4 und 8.5).

3.5.7 Moderation von QoA zur QoE

Die Wirkung der QoA (als Nutzerstimuli) auf die QoE-Bewertungen der Nutzer wird moderiert durch:

- den Kontext als Nutzungssituation (mit Nutzungsziel) und Interaktionskontext (siehe App, Endgerät und Inhalt), dies führt zur weiteren Zerlegung in...
- QoS-unabhängige technische Merkmale der Darstellung und Sichtbarwerdung der Nutzerstimuli in der Nutzersicht (nachfolgend Präsentation genannt) mit der App (Benutzerschnittstelle) und dem Endgerät (Formfaktor, Displaygröße und -qualität, gefühlte Geschwindigkeit) als wichtigen Qualitätselementen, siehe 3.2.6 für die mögliche dienstunspezifische Wirkung der Endgeräte
- Nutzer-immanente Eigenschaften der Wahrnehmung (Physiologie und Psychologie)
- Interaktivität/Interaktion von Dienst und Nutzer

So könnte die Schnittstelle zwischen Mensch und Maschine im QoS-QoE-Modell auch als das Zusammentreffen der drei P beschrieben werden: Auf der Seite der Technik stehen die Leistung (*Performance*) der Anwendung und des Endgerätes mit den Nutzerstimuli und ihre Präsentation (*Presentation*) der Wahrnehmung (*Perception*) auf der Seite des Nutzers gegenüber, aus denen sich letztlich das Nutzererlebnis ergibt.

3.5.8 Präsentation der Nutzerstimuli

Die Gesamtheit aus Software und Hardware, die die Nutzersicht prägen können, wird in [114] S. 30f als Quality of Presentation bezeichnet. Letztlich ist es eine Charakteristik aus verschiedenen Faktoren, so dass diese Bezeichnung prinzipiell korrekt ist. Um die QoS-Unabhängigkeit zu betonen, soll im Weiteren in der eigenen Begriffswelt zur Abgrenzung von den QoX-Schichten nur kurz von *Presentation* bzw. der Präsentation der Stimuli gesprochen werden. In der Sprache des QoX-Matrix-Modells ist die Präsentation ein wichtiger Teil der Moderation zwischen QoA und QoE.

Die Präsentation bezieht sich auf die Darstellung der relevanten Nutzerstimuli in der Benutzeroberfläche der Anwendung/App (z. B. in Form der Rückmeldung zum Übertragungsfortschritt) und des Endgerätes (z. B. zur Bildschirmgröße und -auflösung). Entsprechend sind die dazu diskutierten Punkte dieser Elemente nachfolgend einzubeziehen. Diese Elemente sind weitgehend QoS-invariant, also von der tatsächlichen Netzsituation unabhängig. Für Elemente des Inhalts (z. B. bei mobilen Webseiten) gelten die Erkenntnisse entsprechend. Die Möglichkeiten zur Darstellung sollten zusammen mit den physiologischen Grenzen der Nutzer zur Wahrnehmung daher auch bei der Auswahl der Inhalte berücksichtigt werden.

Um auf den Zusammenhang von QoS und QoE über die Wahrnehmung der Nutzer in der Analogie der mobilen Datenautobahn zurückzukommen: Es sind die Nutzerstimuli der QoA und ihre Präsentation, die vom Nutzer tatsächlich direkt wahrgenommen werden. Die Präsentation beeinflusst dabei lediglich die subjektive Wahrnehmung der Leistung als Erlebnis und/oder die Erwartung, nicht aber die objektive Leistung selbst. Die Anmerkungen zur Beeinflussung sowohl der Erlebnisseite als auch der Erwartungsseite im QoE-Bewertungsprozess sind daher zu beachten und werden nachfolgend (Psychologie der Leistung in 3.5.9) noch einmal aufgegriffen.

Für jede Kategorie der Nutzerstimuli ist von der Forderung einer passenden Präsentation auszugehen, z. B. Anzeigen zur Wartezeit, Indikatoren zu Qualität und Quantität sowie Meldungen etc. zu Funktion und Verhalten. Zur „richtigen“ Art und Weise der Präsentation kann die HCI-Forschung Hinweise geben. Die Erläuterungen erfolgen mit der Diskussion der Dienstklassen mit

jeweils unterschiedlicher Rolle der Nutzerstimuli und somit auch ihrer Präsentation. Für die Gestaltung mobiler Apps gibt es Design- bzw. Entwicklerempfehlungen für die aktuellen Plattformen (Android⁷⁷ und iOS⁷⁸). Diese beinhalten ebenfalls Hinweise zu den möglichen Anzeigeelementen in der Nutzersicht, z. B. für Fortschrittsbalken, Aktivitätsanzeigen und zur Netzaktivität. Zahlreiche Beispiele dazu aus echten Apps finden sich in den Abbildungen zum Testparcours mit der Sicht der Nutzer im Anhang (ab A.4.6), sowie zusammengefasst in der Abbildung 28. In ähnlicher Art sind die allgemeinen HCI-Empfehlungen und genutzte Elemente neben nativen Apps auch für mobile Web-Apps und Webseiten zu nutzen. Dabei spielt auch eine Rolle *wann*, *wie* und *wo* einzelne Elemente des Inhalts (*was*) sichtbar werden.

Die gezeigten Beispiele beziehen sich im Grunde auf die Darstellung der Nutzerstimuli in der einen oder anderen Form, sind also in gewisser Art funktional. Auch die geschickte Nutzung von weiteren dekorativen Elementen kann die Präsentation und somit die QoE verbessern, z. B. durch die Aufteilung des Bildschirms in Bereiche oder die Umlagerung des eingebundenen Videoplayer mit einem Rahmen, so dass die reduzierte Größe des Videobildes in der Startansicht nicht unmittelbar auffällt etc.

Neben der Art der Darstellung über Indikatoren oder textuelle Hinweise kann auch das *Wo* der Präsentation für die Sichtbarkeit und weitere Wirkung eine Rolle spielen. Dabei kann es möglicherweise einen Unterschied geben, ob die Darstellung prominent und unübersehbar im Vordergrund erfolgt oder eher versteckt und unauffällig im Hintergrund (z. B. im Benachrichtigungsbereich bei Android).

Einige Beispiele sollen noch einmal verdeutlichen, was konkret mit der Präsentation gemeint ist, bzw. wie diese wirken kann.

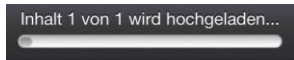
⁷⁷ Ladebalken und Co. unter Android:

<http://developer.android.com/design/building-blocks/progress.html>, abgerufen am 06.01.2015

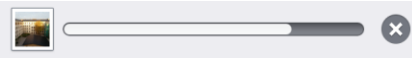
⁷⁸ Activity Indicator, Network Activity Indicator, Progress View unter iOS: <https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/Controls.html> abgerufen am 06.01.2015

Fortschrittsanzeigen

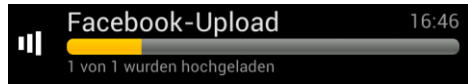
ohne Zusatzinformationen



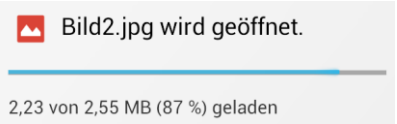
mit Abbruchmöglichkeit



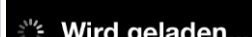
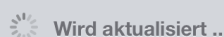
mit Aktivitätsindikator und verwirrender Beschriftung



mit relevanten Zusatzinformationen(Quantität)



Aktivitätsindikatoren



Laden von Inhaltelementen z.B. eingebettete Bilder

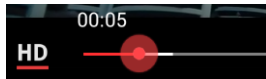
normal (plötzliches Erscheinen des fertigen Bildes nach vollständiger Übertragung)



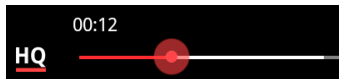
progressiv (zunächst unscharf und Verfeinerung beim Laden)



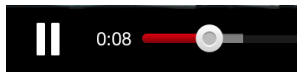
Qualitätsindikatoren



mit (Indikator vorhanden)



ohne (Indikator fehlt)



Rück- und Fehlermeldungen

im Hintergrund



im Vordergrund

Verzögerung bei der Wiedergabe

Aufgrund von einer langsamen Internetverbindung wird dein Titel noch geladen. Für beste Wiedergabe kannst du Playlists über WLAN herunterladen.

Möchtest du weiterhin versuchen den Titel abzuspielen?

Nicht weiter versuchen	Nicht anzeigen	Weiter versuchen
------------------------	----------------	------------------

Abbildung 28: Zusammenstellung von Beispielen zur Präsentation

Screenshots zusammengestellt aus einigen Apps der Dienste im Testparcours (Google Drive, Facebook, MTV-Music, Youtube), Android und iOS gemischt, Ausschnitte zur Verdeutlichung, die vollständigen Screenshots in vergleichbarer Art sind im Anhang zu finden (A.4.6), ergänzend sei auf die dienstunabhängigen Anzeigen zur Netz Wahrnehmung verwiesen (2.5.4)

Beispiel 1: Eine Optimierung nur aus technischer Sicht würde lediglich die Zeit optimieren, eine Optimierung in der Nutzersicht versucht hingegen die Zeitwahrnehmung der Nutzer zu optimieren. Ergibt sich in der QoA eine Ladedauer von 10 s, dann ändert sich dieser objektive Wert durch eine geschickte bzw. ungeschickte Präsentation selbst nicht, sondern nur die subjektive Wahrnehmung als eher kürzer oder als eher länger.

Beispiel 2: Die Präsentation von HD-Video kann mit einem direkt sichtbaren Indikator *HD* erfolgen. So ist einerseits die höhere Qualität des Inhalts (Stimulikategorie Qualität und Quantität) als schärferes Bild sichtbar, gleichzeitig ist die Erwartung aber ggf. auch hoch, bezogen auf diese Kategorie. Für die Kategorie Zeiten und Dauern wird die Erwartung dann eventuell angepasst und liegt nicht so hoch. Der Nutzer berücksichtigt dabei möglicherweise die HD-Qualität in der Beurteilung der notwendigen Vorpufferzeit – er wird ja schließlich durch den Indikator explizit darauf hingewiesen. Fehlen zur Qualität entsprechende Anhaltspunkte, z. B. in Form dieser Indikatoren, dann kann dies der Nutzer bei Bildung seiner Erwartungshaltung auch nicht einbeziehen und entsprechend auch nicht in der QoE-Bewertung honorieren.

Beispiel 3: Auf einem schlechten oder sehr kleinen Bildschirm kann es schwer sein, die HD-Qualität rein visuell über den Inhalt zu erkennen. So ist HD-Video nur dann sinnvoll, wenn das Display des Endgerätes eine adäquate Auflösung und Größe bietet, so dass ein normaler Nutzer aufgrund physiologischer Gegebenheiten überhaupt einen Unterschied erkennen kann.

Für die Wirkung der Präsentation lässt sich so zusammenfassen: QoE-Verbesserungen bei objektiv gleicher QoA sollten durch passende Maßnahmen in der Präsentation erreicht werden können. Möglichkeiten dazu sind die adäquate Präsentation der Nutzerstimuli, die Umwandlung von (negativer) passiver Wartezeit in aktive Nutzungszeit und die Berücksichtigung der menschlichen Wahrnehmung und Informationsverarbeitung sowie des Interaktionskontextes mit dem Endgerät. Durch die Präsentation lässt sich das Erlebnis als auch die Erwartung bewusst gestalten, sowohl zum Positiven wie auch zum Negativen. Erklärungsansätze zur QoE, die dies nicht ausreichend berücksichtigen und sich ausschließlich auf die dem Erlebnis vorgelagerte Leistung konzentrieren,

müssen damit zwangsläufig zu kurz fassen, weil sie einen außerordentlich wichtigen Aspekt der Nutzersicht ausklammern.

3.5.9 Psychologie der Leistung

Bei der Moderation durch die Präsentation wird eine Wirkung sowohl auf Erlebnis- als auch auf Erwartungsseite angenommen. Zur Wirkrichtung auf jeweils einer Seite sind ggf. noch plausible Annahmen möglich, zum Zusammenwirken auf beiden Seiten hingegen kaum. Daher ist es möglich, dass im Rahmen der Moderation verstärkende als auch abschwächende Effekte zu verzeichnen sind, über die aber keine plausiblen Annahmen zu treffen sind. An dieser Stelle kann daher nur eine geschickte empirische Ermittlung weitere Erkenntnisse generieren. Für die Aufklärung des Einflusses der Präsentation und Wahrnehmung soll davon ausgegangen werden, dass sich die Änderungen der QoE-Bewertungen der Nutzer *ceteris paribus* (unter sonst gleichen Umständen, hier der Nebenfaktoren wie auch der Präsentation) aus den QoS-veränderlichen Stimuligrößen erklären lassen. Ebenso gilt der Umkehrschluss, dass gleiche Stimuli mit unterschiedlichen QoE-Bewertungen den Einfluss der Präsentation aufdecken können (bei Beachtung der statistischen Voraussetzungen).

Die Präsentation ist zwar ein technischer Einfluss, dieser wirkt allerdings ausschließlich subjektiv in der Nutzersicht. Entsprechend ist Vorsicht geboten: Wenn etwas plausibel klingt, so verhält es sich in der Nutzersicht noch lange nicht genauso. Dabei kann es auch so sein, dass nicht alle empirisch belegten Erkenntnisse intuitiv einleuchten oder einfach erklärbar sind. Dies soll durch einige Beispiele verdeutlicht werden.

[75] verweist darauf, dass entgegen der üblichen Meinung das progressive Laden von Bildern (relativ schnell schon ein unscharfes Bild zu sehen, das dann sukzessive scharf dargestellt wird) als schlechter empfunden wird als die gleiche Wartezeit und prompte Darstellung des finalen Bildes. Als mögliche Begründung wird in [75] S. 26 ein erhöhter Wahrnehmungsaufwand der Nutzer genannt, die erst das unscharfe Bild und dann noch einmal das vollständige Bild wahrnehmen und verarbeiten müssen.

Die QoS-abhängigen Nutzerstimuli sind bei komplexeren Nutzungen, wie z. B. beim Web-Browsing die Ladedauer einer Seite, nicht alleinig wirksam. Dabei sind Hinweise zu auftretenden Wartezeiten (z. B. Spinner) möglicherweise nicht immer vorteilhaft, weil man den Nutzer diese Zeit eben als Wartezeit erleben lässt. Ebenso wenig kann es für das Nutzererleben positiv sein, die Nutzer erst Sekunden lang auf komplett leere Seite starren zu lassen und dann schlagartig die fertige Seite zu präsentieren. Dabei sind Platzhalter in einem Webseiten-Skelett ggf. besser (wenn sich die Struktur nicht mehr stark ändert), um dem Nutzer während des Aufbaus der Seite schon die Struktur einer Seite erfassen zu lassen – und dann erst die eigentlichen Inhalte einzusetzen (siehe Beispiel in [168]). So kann u. U. die passive Wartezeit aus Nutzersicht schon in aktive Nutzungszeit zur Orientierung umgewandelt werden. (zusammengefasst nach [168]). Weitere Beispiele sind in [169] bzw. [170] zu finden, woraus auch der Titel dieses Unterabschnitts entlehnt ist.

3.5.10 Complexity of Experience

Die Dienste können sich in der Nutzersicht hinsichtlich der Komplexität des Nutzererlebnisses unterscheiden, [158] S. 43f bezeichnet dies als Complexity of Experience. Dabei ist davon auszugehen, dass es Dienste gibt, die nur eine recht geringe Komplexität des Nutzererlebnisses aufweisen. Dies deckt sich mit den eigenen Überlegungen zu den gebildeten Kategorien der Nutzerstimuli und ihrem Zusammenwirken. Eine Unterscheidung in *funktioniert* – oder *funktioniert nicht* für Funktion und Verhalten ist recht einfach und auch in der Nutzersicht recht eindeutig. Das Gegenteil wären Dienste, die ganz unterschiedliche Nutzerstimuli aufweisen und diese ggf. in wechselseitig abhängigen Ausprägungen kombinieren, z. B. Onlinespiele oder Videotelefonie, bei denen also zwischen allen drei Kategorien Kompromisse auftreten können.

Besonders interessant wird diese Überlegung, wenn orthogonal zur Complexity of Experience der interpersonelle Unterschied in der Bewertung – als Hinweis auf unterschiedliche Wahrnehmung und Beurteilung – betrachtet wird, siehe Abbildung 29. Unterschiedliche Wahrnehmung und Gewichtung der Stimuli sind besonders bei Diensten mit hoher Erlebniskomplexität zu erwarten. Unterschiedliche Erwartungshaltungen wiederum dürften sich am ehesten bei Diensten mit geringer Complexity of Experience zeigen.

Die folgenden Ausführungen beziehen sich auf [158] S. 43f und eigene Überlegungen in der Theorie wie auch auf Erkenntnisse aus der eigenen Studie. Dienste mit geringer Komplexität und geringen Bewertungsunterschieden (Quadrant III) sind objektiv gut prognostizierbar. Ein Beispiel wäre Video-streaming (wie Youtube in der eigenen Studie). Das Auftreten oder Ausbleiben von Unterbrechungen der Wiedergabe wird von allen Nutzern ähnlich wahrgenommen und ähnlich bewertet. Generell dürften Dienste mit sprunghafter QoS-QoE-Charakteristik in II oder III anzusiedeln sein, Dienste mit einer elastischen QoS-QoE-Änderungscharakteristik in den Quadranten I und IV. Dienste in den Quadranten I, II und IV sind eher subjektiv geprägt, sowohl in Wahrnehmung als auch Erwartung. Im Quadrant IV sind Dienste zu vermuten, die z. B. von Wartedauern für längere Downloads oder Uploads geprägt sind. Dabei spielen Unterschiede in der Erwartung eine größere Rolle als Wahrnehmungsunterschiede. Im Quadrant I kommen solche Unterschiede in der Wahrnehmung hinzu, z. B. durch die Qualitätswahrnehmung (Audio) in der Internettelefonie und weiche Echtzeitkriterien. Interaktive Dienste sollten demnach im Bereich mittlerer Komplexität und Bewertungsunterschiede zu finden sein.

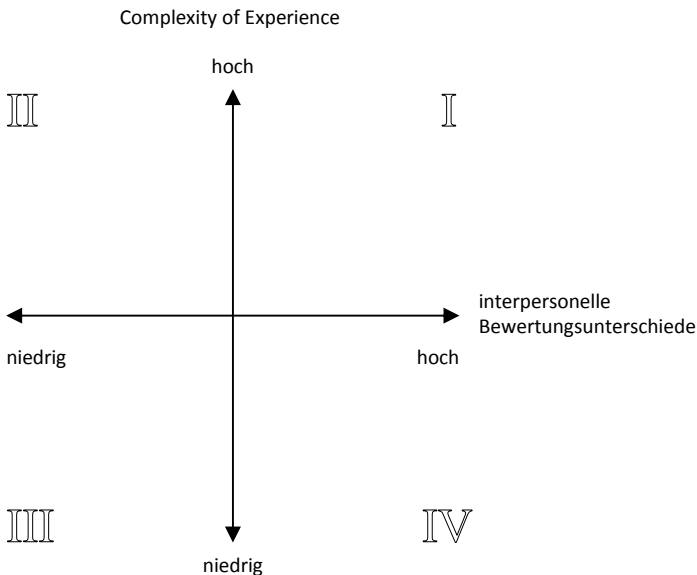


Abbildung 29: Complexity of Experience und interpersonelle Bewertungsunterschiede der Dienste in Anlehnung an [158] S. 43

3.5.11 Interaktivität/Interaktion von Nutzer und Dienst

Die Nutzung der Dienste erfolgt als Interaktion der Nutzer mit dem Gerät und den Apps. Die Hand repräsentiert das Verhalten des Nutzers bzw. seine Aktionen in der Interaktion mit dem restlichen System. Die Hand wirkt im Modell also als Rückkopplung vom Beobachter auf das Beobachtungsobjekt. Der Nutzer wirkt auf das System ein und beeinflusst damit die übrigen Faktoren ggf. selbst (Selektion und Modifikation). Gleichzeitig ist dies ihm aber ggf. bewusst, und er berücksichtigt dies wiederum (Interpretation). Die genutzten Begriffe und Konzepte zur Interaktivität und zur Interaktion sind in [171] zusammengefasst beschrieben.

Ein Beispiel: Wählt ein Nutzer bewusst ein großes Bild zur Übertragung aus (Selektion), so ist seine Erwartung ggf. darauf angepasst, dass es etwas länger dauern könnte (Interpretation).

Daher ist es plausibel anzunehmen, dass die Art der Interaktivität grundsätzliche eine Rolle spielen kann, wie eine bestimmte App oder ein Dienst an sich wahrgenommen und bewertet wird. Die Nutzerstimuli sind dabei in die Interaktion eingebettet. Bei einem Dienst mit niedriger Interaktivität kommt dem Nutzer nicht viel Aktion zu. Für unterschiedliche Dienste kann entsprechend des Grades der Interaktivität die Bedeutung der Komponente Hand variieren. Aus Sicht der Dienste im QoS-QoE-Zusammenhang interessieren vorrangig die Interaktionen, die (direkt) mit einer zugehörigen Netznutzung gekoppelt sind.

Die Rückkopplung von Auge, Hand und Hirn ist auch die Erklärung, warum interaktive Nutzertests notwendig sind und warum das Zeigen von Videos aufgezeichneter, beispielhafter und garantiert korrekt ausgeführter Aufgabenbearbeitungen mit den entsprechenden Nutzerstimuli nicht ausreicht, die QoE zu bestimmen. Die Selbstwirkungserfahrung würde dabei fehlen. Dabei ist es plausibel, dass gerade diese Selbstwirkungserfahrung ein wichtiger Bestandteil des Nutzungserlebnisses ist.

Die Nutzer sollen dabei ihre eigene Handlung als Ursache der nachfolgenden Wirkungen möglichst gut und direkt erleben können:

- Laden der Webseite nach Berühren des Icons zur verlinkten Seite
- Start der Wiedergabe nach Druck auf Start
- Start des Uploads nach Druck auf Senden
- interaktives Verschieben und Zoomen einer Kartendarstellung etc.

Grundlagen zur allgemeinen Geräteinteraktion bzw. der Interaktion mit grafischen Benutzeroberflächen aus der HCI-Welt sind beispielsweise in [172] S. 133ff bzw. 243ff zu finden. Vertiefende Erläuterungen zur Rolle der Interaktivität in Bezug auf QoE, auch in verschiedenen Szenarien der Nutzung (Mensch-zu-Maschine und Mensch-zu-Mensch) findet der Leser in [173]. Die Autoren teilen dabei die Auffassung, dass die Interaktivität in der Nutzung der Dienste eine entscheidende Rolle für die QoE einnimmt.

Für einen interessanten Ansatz zur Visualisierung der Interaktivität bzw. der Interaktion der Nutzer mit App und Endgerät verweist [174] S. 114 auf [175]. Dabei dienen die Fingerspuren auf dem Touchscreen eines Tablets, quasi als Überbleibsel der Nutzung, zur Visualisierung der typischen Interaktionsmuster, die sich je nach Dienst mehr oder weniger deutlich unterscheiden. So ist im Nachhinein einer Nutzung u. U. gut zu erkennen, wie oft, in welcher Art und an welcher Position bei einem bestimmten Dienst mit dem Bildschirm in Aktion getreten wurde. Dabei ist leicht zwischen eingabeintensiven Diensten, z. B. E-Mail, Scroll-Dienste wie Web-Browsing, Diensten mit Wischgesten (Darstellung von Bildalben etc.) oder der Videonutzung (Start-Stopp-Knopf) zu trennen. Die gezeigten Beispiele erinnern den Verfasser der vorliegenden Arbeit dabei an die Vorbereitungshandlungen für die mobilen Geräte zu den einzelnen Testslots mit den Probanden, bei denen ein Mikrofasertuch jeweils zur Beseitigung eben dieser (dann ungewollten) Interaktionsspuren auf den Displays diente.

3.6 Die Dienste – Teil 4: Gesamtsicht und Klassen

Für die Gesamtsicht zum dienstspezifischen QoS-QoE-Zusammenhang soll der Bogen zurück zum Beginn des Kapitels und zur Theorie geschlagen werden: Wie sind aus Anforderungen und Einschränkungen die Ursachen und Wirkungen im Gesamtzusammenhang zu erklären? Derartige Erklärungen sind prinzipiell dienstspezifisch, es bietet sich aber eine Klassenbildung nach unterschiedlichsten Kriterien an, auf die kurz eingegangen wird. Dies führt zu den Taxonomien für Dienste und den Dienstklassen, welche die Netzverkehrsklassen und Nutzungsklassen miteinander verbinden. Nach einer Übersicht zu den relevanten Kriterien mit Gemeinsamkeiten und Unterschieden werden die identifizierten Grundklassen nacheinander vorgestellt. Die Diskussion schließt mit der Ergründung der Relevanz einzelner QoS-Netzparameter für die Dienste im QoS-QoE-Zusammenhang.

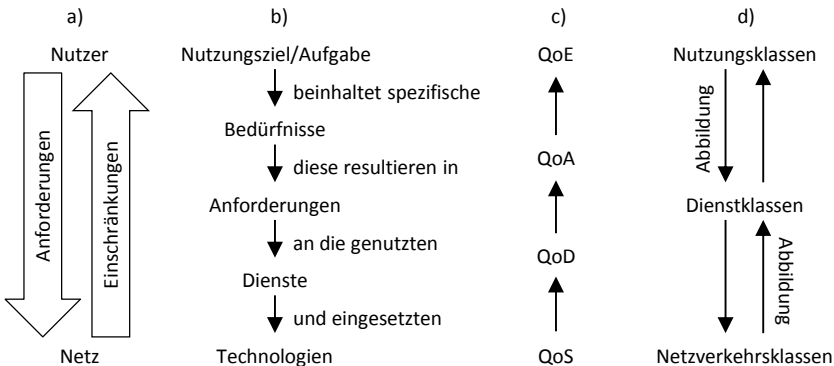


Abbildung 30: Von Anforderungen und Einschränkungen zum dienstspezifischen QoS-QoE-Zusammenhang

- a) konzeptueller Zusammenhang von QoS und QoE durch Anforderungen und Einschränkungen zwischen Nutzer und Netz; vereinfachte Sicht, Anforderungen gehen im Modell vom Nutzer aus und resultieren in Anforderungen an das Netz, Einschränkungen gehen im Modell vom Netz aus und resultieren ggf. in Einschränkungen für den Nutzer
- b) vom Nutzungsziel über dienstspezifische Anforderungen zu Anforderung an das Netz, nach [104] S. 13, erweitert und ergänzt
- c) konstruierte Ursache-Wirkungskette von der Technik mit dem Netz zur QoE in der Nutzersicht
- d) die Dienstklassen zwischen der Sicht der Nutzer und zugeordneten Nutzungsklassen (oben) und der Sicht des Netzes (unten) und den zugeordneten Netzverkehrsklassen, Anforderungen und Einschränkungen werden dabei über die jeweiligen Qualitätsmerkmale beschrieben

3.6.1 Von Anforderungen und Einschränkungen zum dienstspezifischen QoS-QoE-Zusammenhang

An dieser Stelle sollte das Zusammenwirken der einzelnen Elemente und Qualitäten soweit erklärt sein, dass sich aus Netzsicht und Nutzersicht eine Gesamtsicht auf den Zusammenhang zwischen QoS und QoE erkennen lässt. Es ist jeweils klargestellt, welche einzelnen Dimensionen (Input) und Maße (Output) in den Schichten aufeinandertreffen. Davon ausgehend lässt sich jetzt ableiten, wie aus Nutzeranforderung die Netzanforderungen entstehen und wie sich Einschränkungen vom Netz bis hin zu Einschränkungen des Nutzererlebnisses und damit auf die erzielbaren QoE-Bewertungen auswirken, siehe Abbildung 30a. Dabei führen bestimmte Nutzungsziele letztlich zu Anforderungen an die genutzte Technologie und damit an das Netz, siehe Abbildung 30b, was somit so etwas wie die Umkehrung der Ursache-Wirkungskette des Modells (Abbildung 30c) beschreibt.

Zur Erfüllung ihrer Erwartungen (QoE) stellen Nutzer Anforderungen an die Dienste und damit an die Applikationsleistung (QoA). Die Dienste stellen wiederum Anforderungen an die effektive Kommunikationsleistung (QoD). Aus diesen Anforderungen ergeben sich die Anforderungen an das Netz (QoS) und damit die Netzleistung. Im Grunde kann überhaupt nicht überbetont werden, dass die Nutzeranforderungen für sämtliche Betrachtungen zu QoE und QoS die zentrale Rolle spielen. Ein nutzerrelevantes Kriterium kann dabei von ein oder mehreren netzrelevanten Kriterien bzw. ihrem Zusammenwirken abhängen.

Da die Anforderungen ebenso dienstspezifisch sind wie die Wirkungen der Einschränkungen, ist grundsätzlich die Beschreibung des QoS-QoE-Zusammenhangs nur dienstspezifisch sinnvoll – bzw. verallgemeinert für vergleichbare Dienste in Dienstklassen. Dazu wird diskutiert, wie diese Dienstklassen sinnvoll gebildet werden können.

3.6.2 Terminologie zur Beschreibung des QoS-QoE-Zusammenhangs

Die Beschreibung des dienstspezifischen QoS-QoE-Zusammenhangs für einzelne Parameter und die Charakteristik des Zusammenhangs an sich sollten jeweils in vergleichbarer Art und Weise erfolgen. Dafür wird eine sinnvolle

Terminologie für verschiedene Aspekte des QoS-QoE-Zusammenhangs benötigt. Dies umfasst die Formulierung der Anforderungen und Einschränkungen, die Rolle einzelner Merkmale (z. B. der QoA-Nutzerstimuli oder der QoS-Netzparameter) und die Charakterisierung des QoS-QoE-Zusammenhangs. Leider ist dies scheinbar nicht so einfach, wie der Blick in typische Auflistungen zeigt (z. B. in Verbindung mit Taxonomien, dazu später noch mehr). Die Angaben sind geprägt von einer Mischung aus festen und diffusen bzw. absoluten und relativen Angaben: einfache Zahlenwerte, Zahlenbereiche, implizite Annahmen oder Angaben zur Begrenzung (min./max.), Klassifizierungen und verbale Beschreibungen. Die Formulierungen sind dabei oft wenig formal, so dass nur intuitiv zu erfassen ist, was eigentlich gemeint ist.

Dabei kann die prinzipielle Beschreibung der Einflussgrößen (in der QoX-Matrix) auf einige wenige Punkte zurückgeführt werden:

- Relevanz im Vergleich zu anderen Einflussgrößen: (sehr) niedrig, mittel, (sehr) hoch, neutral (d. h. ohne Einfluss)
- Einflussrichtung:
 - *positiv*: die Werte der abhängigen Variablen (AV) verändern sich in die gleiche Richtung wie die unabhängigen Variablen (UV)
 - *negativ*: entsprechend umgekehrt
- Änderungsbereich (Breite des Bereichs, in dem Änderungen der UV zu Änderungen der AV führen)
 - *eng* (fest, hart, sprunghaft): Änderungen in einem relativ kleinen Bereich der UV führen zu großen Änderungen der AV
 - *weit* (weich, fließend, kontinuierlich): Änderungen in einem relativ großen Bereich der UV führen zu Änderungen der AV

Die Änderung der AV in Abhängigkeit einer Änderung der UV kann bezogen auf eine technische Messung als Empfindlichkeit (Sensitivität) beschrieben werden. Bei der begrifflichen Nutzung als Elastizität ist allerdings Vorsicht geboten, da von jeweils unterschiedlichen Wirkungen bestimmter Parameterbereichen auszugehen ist. Zudem sollte beachtet werden, ob sich die Elastizität auf die UV oder AV bezieht. Für Betrachtungen zum Zusammenhang von QoS und QoE wird ein weiterer Änderungsbereich als elastischer Zusammenhang festgelegt, ein enger als unelastisch (siehe [176] S. 1183 bzw. Beispieldarstellungen in Abbildung 31 im nachfolgenden Abschnitt).

Tabelle 13: Symbolik zur Charakterisierung von Einflussrelevanz und Einflussrichtung

Einfluss- richtung	negativ größere Werte → schlechtere Bewertungen			ohne	positiv größere Werte → bessere Bewertungen		
	<i>hoch</i>	<i>mittel</i>	<i>niedrig</i>	<i>neutral</i>	<i>niedrig</i>	<i>mittel</i>	<i>hoch</i>
Symbolik	---	--	-	o	+	++	+++

Sofern nicht nur zwei sich ausschließende Gegenpole existieren, ist es besser, eine Graduierung vorzusehen. Die Abstufung der Relevanz kann mit der Einflussrichtung sinnvoll kombiniert werden, siehe Tabelle 13. Für Merkmale ohne explizite Angabe ist zu schließen, dass diese keinen relevanten Einfluss ausüben, ihre konkrete Ausprägung oder Änderung also keiner besonderen Beachtung bedarf.

Zusätzlich kann es möglich oder sinnvoll sein, auch absolute Werte anzugeben, gerade mit Blick auf die gesuchten Schwellwerte:

- absolute Höhe eines Wertes oder Bereiches: (sehr) niedrig, mittel, (sehr) hoch
- Art eines Wertes oder Bereiches mit Richtungsbegrenzungen: Minimalwert, Maximalwert, Mittelwert bzw. mindestens, höchstens, durchschnittlich etc.
- ggf. zugehöriger Bezugsrahmen, z. B. zeitlicher Rahmen für die Dynamik (Beispiel: „Ein Mittelwert von nicht weniger als 2,4 MBit/s in 10 s.“)

Die vorgestellte Terminologie wird auch zur Beschreibung der QoS-QoE-Zusammenhänge der untersuchten Dienste in der eigenen Studie genutzt, siehe entsprechende Ausführungen zu den Diensten des Testparcours im Kapitel 9.

3.6.3 Charakterisierung der QoS-QoE-Dienstkurven

Zur Natur des QoS-QoE-Zusammenhangs sind auf Basis der dargelegten Modellierung mit Relevanz, Einflussrichtung und Änderungsbereich einige Grundannahmen möglich und als plausibel anzusehen. Damit lässt sich der QoS-QoE-Zusammenhang als eine mathematische Abbildung zwischen der Menge der Funktionsargumente und einer Menge von Funktionswerten formulieren. Den Parameterraum bilden die QoS-Parameter und Nebenfaktoren in der QoX-Matrix, die Funktionswerte entstammen dem Bewertungsraum als

Teil des Ergebnisraums. Der QoS-QoE-Zusammenhang ist dabei grundsätzlich anzunehmen als nichtabnehmend stetig für *bessere* Parameter, d. h. monoton, aber nicht streng monoton. Was *bessere/schnellere* Parameter sind, ergibt sich wiederum aus sinnvollen Annahmen zur Einflussrichtung und sollte in den meisten Fällen leicht fassbar sein (z. B. für die QoS-Netzparameter: höhere Datenraten und niedrigere Latenz, Jitter, Loss). Nichtlinearitäten sind möglich und zu beachten. Was *bessere* Bewertungen bzw. QoE-Werte sind, ergibt sich aus den betrachteten QoE-Maßen, der genutzten Skala und zugehöriger Metrik. Auf diese Punkte wird in der Diskussion zur Hinführung auf die empirische Untersuchung und in der Auswertung der erhobenen Daten später noch intensiv eingegangen. An dieser Stelle soll die abstrakte Festsetzung von *gut* und *schlecht* bzw. von *besser* kontrastiert mit *schlechter* zum Verständnis ausreichen. Sollten zu den genannten Grundannahmen des Zusammenhangs tatsächlich Abweichungen in empirischen Daten festzustellen sein, sind Erklärungen außerhalb der Theorie zu suchen oder ggf. weitere relevante Einflussfaktoren in der Modellierung zu berücksichtigen.

Es ist zu erwarten, dass sich für die Abbildung eines QoS-QoE-Zusammenhangs charakteristische Bereiche ergeben:

- im Bereich der Unzufriedenheit am unteren QoE-Ende und unterhalb bestimmter QoS-Parameterwerte: keine oder nur geringe Änderung der QoE-Bewertung (insgesamt schlecht) in Abhängigkeit der QoS-Parametrisierung als *Bodeneffekt*
- Änderungsbereich zwischen bestimmten QoS-Parameterwerten: Änderung der QoE-Bewertungen in Abhängigkeit der QoS-Parametrisierung (QoS-Änderung führt zu QoE-Änderung)
- im Bereich der Zufriedenheit am oberen QoE-Ende: keine oder nur geringe Änderung der QoE-Bewertung (insgesamt gut) in Abhängigkeit der QoS-Parametrisierung (selbst große QoS-Änderung führen nur noch zu kleinen QoE-Änderungen) als *Deckeneffekt*
- Ob sich die genannten Bereiche unveränderter QoE-Werte unterhalb oder oberhalb bestimmter QoS-Parameterwerte zeigen, hängt von der Einflussrichtung der betrachteten Parameter ab.

Wie erläutert, sind die QoS-QoE-Zusammenhänge dienstspezifisch. Entsprechend soll die Abbildung des Zusammenhangs fortan als *Dienstkurve* bezeichnet werden. Die Charakteristik der Dienstkurve ergibt sich aus den Anforderungen und Auswirkungen von Einschränkungen. In Abhängigkeit der betrachteten Parameterraumgröße (x-Achse) in Wirkung auf die QoE-Metrik des Bewertungsraums (y-Achse) sind für unterschiedliche Dienstypen auch unterschiedliche Grundformen der Dienstkurven anzunehmen, siehe Abbildung 31 und Tabelle 14 zur Erläuterung der Charakteristik. In der QoS-QoE-Domäne findet auch eine Utility-Funktion als QoE-Maß Anwendung, siehe z. B. [118]. Dabei wird „Utility“ aber nicht auf den Nutzen beschränkt, sondern ist eher im Sinne eines generischen (QoE-) Wertes der Nutzung zu verstehen. Die für diese Utility-Funktion vorausgesagten Kurvenverläufe in [176] S. 1183 entsprechen den o. g. Überlegungen und fanden ihre Anwendung auch in anderen Arbeiten zur QoE, siehe z. B. [77] S. 3, oder [118] S. 2.

Neben den extremen Ausprägungen der konstruierten Grundtypen sind Zwischenformen denkbar. Möglicherweise ergeben sich aus dem Zusammentreffen ganz verschiedener Anforderungen und Einschränkungen auch noch weitere Formen, z. B. mit einzelnen „kleinen“ Sprüngen für Qualitätsstufen (z. B. SQ, HQ, HD wie beim Inhalt mit Qualität und Quantität diskutiert), die mit Verbesserungen im Parameterraum nach und nach verfügbar werden.

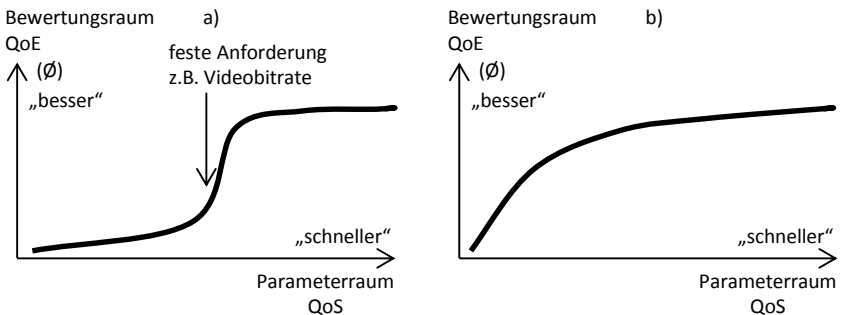


Abbildung 31: Angenommene Dienstkurven zum QoS-QoE-Zusammenhang

a) sprunghafte Änderung (feste Anforderung, „nicht elastisch“)

b) allmähliche Änderungen (weiche Anforderung, „elastisch“)

Beispiele zur Illustration, x-Achse Datenrate Download (kBit/s) als QoS-Netzparameter und auf der y-Achse die durchschnittliche QoE-Bewertung (Zufriedenheit) ermittelt auf einer wertbegrenzten Intervallskala, ähnlich zu [118] S. 2

Tabelle 14: Charakterisierung der genannten Grundformen zum QoS-QoE-Zusammenhang

Charakteristik	„hart“, sprunghafte Änderung in einem vergleichsweise engen Parameterbereich	„weich“, allmähliche Änderung über einem vergleichsweise weiten Parameterbereich
Erklärung des Entstehens	eine große Zahl Nutzer ändert quasi schlagartig die Bewertung	die Nutzer ändern nach und nach ihre Bewertungen, was zu einer recht konstanten Änderungsquote der Bewertungen führt
Einordnung	eher objektiv begründet	eher subjektiv begründet
Anforderungen	feste Anforderungen	flexible Anforderungen
Elastizität (QoS-QoE)	unelastisch	elastisch
vermutete Ursachen	entstehend aus Auffälligkeiten bei Funktion und Verhalten oder besonderen zeitlichen Anforderungen	entstehen eher aus den Erwartungen, Vorkenntnissen und Ansprüchen der Nutzer
Beispiele für Ursachen	<p>technisch oder psychologisch/neurowissenschaftlich (Hirn als Maschine) messbar begründet</p> <ul style="list-style-type: none"> • Inhalt: Datenraten (z. B. Videostreaming) • Timeouts/Deadline, Interaktivität, Synchronität • Fehlerraten • wie lange eine Konversation noch als live/synchron empfunden werden kann 	<p>alle Arten von Wartezeiten (breite Verteilung der Geduld) und ggf. stufenlose Änderungen von Qualität und Quantität</p> <ul style="list-style-type: none"> • Ladedauern • Bildgrößen • Anzahl einzelner Elemente
Abschätzung zum interessanten Parameterbereich	recht gute Eingrenzung rund um erwartete spezifische „Auffälligkeiten“	Anwendung der HCI-Zeitmarken und Abschätzung zum oberen und unteren Ende
Näherung durch Funktion	Näherung des Sprungs im Änderungsbereich mittels Sigmoidfunktion (logistische Funktion beidseitig)	je nach Ausprägung mit annähernd logarithmischem Funktionszusammenhang (logistische Funktion einseitig)

Es sei an dieser Stelle auf die Ausführungen zur Nutzersicht mit der Complexity of Experience im Abgleich mit den interpersonellen Bewertungsunterschieden verwiesen, die durch die Erläuterungen hier noch weiter unterstrichen werden.

Die genaue Art und Form der Dienstkurve ergibt sich aus dem mathematischen Modell, wie die QoE-Metrik zustande kommt. Im Fall einer üblichen, wertbegrenzten Bewertungsskala (z. B. von 1 bis 5), ergibt sich entsprechend der Verschiebung der Anteilswerte der Bewertungen eine logistische Funktion, weil mit besseren Eingabeparametern ein immer größerer Anteil der Messungen (Bewertungen durch die Probanden) zunächst einen immer besseren Ausgabewert erzeugt und letztlich quasi alle Bewertungen auf den Skalenendwert entfallen sollten. Je nach gefundener Ausprägung kann die logistische Funktion unterschiedlich parametrisiert und damit im Parameterraum verschoben bzw. unterschiedlich steil im Bezug zum Bewertungsraum sein.

Sind die Grundannahmen und die eigenen Annahmen zur Wirkung der verschiedenen Nutzerstimuli auf die Nutzerzufriedenheit als QoE-Maß in den konkreten Nutzungsszenarien im eigenen Test korrekt, so sollten sich in ähnlicher Form wie in Abbildung 31 auch die eigenen Ergebniskurven mit geeigneter Metrik darstellen⁷⁹. Die Kenntnisse der Art des Zusammenhangs, bzw. zur Wirkung von Anforderungen und Einschränkungen erlauben aber nicht nur die Erklärung der Ergebnisse, sondern auch sinnvolle Propositionen zur Charakteristik der Dienstkurve. So lassen sich die QoE-Änderungen aus QoS-Änderungen durch Variationen der Netzparameter zumindest angenähert voraussagen, was die Eingrenzung auf interessante Parameterbereiche und die Wahl sinnvoller Abstufungen bei der Evaluation in einem Nutzertest erlaubt.

Eines der gesteckten Ziele der Arbeit war die Ermittlung, ob es so etwas wie ein „gut genug“ für die mobilen Dienste gibt. Dies kann als ein QoS-QoE-Optimierungsziel angesehen werden, das ausgehend von den Charakteristiken des QoS-QoE-Zusammenhangs in den Dienstkurven nicht nur möglich, sondern sogar höchst sinnvoll erscheint. So kann aus der Suche nach „gut genug“ die Suche nach konkreten Schwellwerten werden. Zur Konkretisierung sind weitere Festlegungen für Parameterraum und Metrik im Bewertungsraum notwendig, die in der Auswertung der eigenen Daten noch einmal aufgegriffen werden. Die Diskussion dazu schließt sich im Abschnitt 8.8 an, nachdem der Leser sich in den dazwischenliegenden Kapiteln mit den Hintergründen und Details der empirischen Untersuchung vertraut machen konnte.

⁷⁹ Die elastische wie auch die unelastische Form der Dienstkurven konnte im empirischen Teil bestätigt werden, siehe Abbildung 113.

3.6.4 Bildung von Klassen

Es ist nicht praktikabel, jeden mobilen Dienst einzeln zu betrachten. Stattdessen bietet sich die Bildung von Kategorien bzw. Klassen an. Dies kann auf Basis gemeinsamer bzw. vergleichbarer Kriterien erfolgen:

- in der Netzsicht: mit vergleichbaren technischen Anforderungen an das Netz (QoS) oder mit vergleichbaren Netznutzungen (QoD)
- in der Nutzersicht: mit vergleichbaren Nutzungszielen, Bedürfnissen und Anforderungen (QoE) oder mit vergleichbaren Beschreibungen durch die Nutzerstimuli (QoA)
- in der Gesamtsicht des Zusammenhangs: mit vergleichbaren QoS-QoE-Zusammenhängen

Für die Sicht des Netzes wurden bereits zusammen mit den QoS-Mechanismen die QoS-Netzverkehrsklassen vorgestellt (siehe 2.2.8), die für diese Sicht eine gute Orientierung liefern, wenngleich andere Ansätze zur Bildung von Netzverkehrsklassen u. U. mehr Klassen vorsehen. Für die Sicht der Nutzer in einer spezifischen Nutzungssituation wurden die Nutzungsklassen eingeführt (siehe 2.4.2).

Eine unpassende Zusammenfassung zu Klassen ist ebenso ungünstig wie eine überbordende Vielzahl an Klassen. Die Heterogenität von Diensten in willkürlich gebildeten oder abstrakten Kategorien kann ein Problem sein, wenn von diesen verallgemeinert auf die Anforderungen an (mobile) IP-Netze geschlossen werden soll. Deshalb sollten die entsprechende Zuordnung und auch die Verallgemeinerung gut überlegt sein. Die Rückführung auf einige wenige Zwischenklassen, die grundlegende Mechanismen der Anforderungen und Einschränkungen berücksichtigen und aufeinander abbilden, ist zur Vereinfachung und Verbindung der Nutzersicht mit der Netzsicht hingegen sinnvoll. Diese gemeinsame Betrachtung führt schließlich zu den *Dienstklassen*, siehe Abbildung 30d. Die Dienstklassen geben Hinweise auf die Ursache der Anforderungen, die relevanten Merkmale der entsprechenden QoS-QoE-Zusammenhänge und mögliche technisch bedingte Einschränkungen. Im Detail ist für jeden einzelnen Dienst zu prüfen, welche Anpassungen ausgehend von einigen wenigen Basisklassen sinnvoll sind.

3.6.5 Mögliche Taxonomien der Dienste

Zur Zusammenfassung von Diensten ist eine Taxonomie üblich, also die Einordnung in ein Klassifikationsschema nach bestimmten Kriterien, ggf. auch hierarchisch gegliedert. Die Recherchen ergaben dabei nicht nur ein Klassifikationsschema, sondern eine ganze Vielzahl, mit ganz verschiedenartigen Kriterien. Prinzipiell lassen sich die Ansätze zur Unterscheidung der Dienste kombinieren und so eine feinere Klassifizierung erreichen. In vielen der im Rahmen der Recherche gefundenen Taxonomien zu Diensten findet sich dieses Prinzip wieder, auch wenn es oftmals kaum so explizit beschrieben ist. Nachfolgend sollen einige dieser Taxonomien vorgestellt werden. Diese entstammen dabei teilweise aus unterschiedlichen Netzwelten (Telko/Internet), lassen sich aber sinngemäß adaptieren.

Eine mögliche Klassenbildung ergibt sich aus der Unterteilung der Elastizität und der Interaktivität (nach [119] S. 150). Entsprechend der eigenen Systematik entstehen die Dienstklassen damit aus der Betrachtung der Charakteristik des QoS-QoE-Zusammenhangs (Elastizität) und der Nutzersicht (Interaktivität). Bei geeigneter Darstellung in einer Matrix, siehe Tabelle 15, können in den einzelnen Zellen die Dienstklassen aus weiteren Klassifizierungsschemata zugeordnet werden.

Tabelle 15: Klassenbildung aus Elastizität und Interaktivität mit Beispieldiensten

Elastizität	Interaktivität	
	interaktiv	nicht-interaktiv
elastisch	Responsive / Retrieval ⁸⁰ <i>Web-Browsing etc.</i> <i>Google Maps</i>	Store-and-Forward Messaging Background <i>E-Mail</i> <i>P2P</i> <i>allgemeine Downloads</i>
nicht-elastisch	Realtime / Conversational <i>VoIP, Videotelefonie</i> <i>Online-Spiele</i> <i>Telekonferenzen</i>	Live-Streaming Broadcast-ähnliche Dienste <i>Mobile-TV</i>
sowohl als auch	„direct“ Services	Streaming (allgemein)

in Anlehnung an [119] S. 150, ergänzt um die Einordnung der Klassen aus alternativen Ansätzen und *Beispieldienste* (kursiv), sowie „sowohl als auch“

⁸⁰ allgemeine Abrufdienste

Die ITU-T Y.1241 ergänzt noch Distributionsdienste mit bzw. ohne individuelle Kontrolle für die Verteilung von Medieninhalten [46] S. 6. Distributionsdienste ohne individuelle Kontrolle, z. B. auf Basis von Multicast für Mobile-TV etc., spielen in produktiven Mobilnetzen (noch) keine relevante Rolle. Die Dienste mit individueller Kontrolle (on-demand) können je nach Medium auch in (interaktive) Abrufdienste und (Audio-/Video-) Streaming unterschieden werden.

[34] S. 144ff verweist auf die ITU-T I.211⁸¹ als Ausgangspunkt für die Definition von Breitbanddiensten in Rechnernetzen. Dabei werden die Klassen nach dem Grad der Interaktivität in interaktive Dienste und zuvor genannte Verteildienste unterschieden und innerhalb dieser Kategorien weiter entsprechend dem Kommunikationsschema aufgeschlüsselt, hier nur für die interaktiven Dienste (nach [34] S. 144f):

- Conversational (Dialogdienste im weiteren Sinne): gekennzeichnet durch einen direkten und bidirektionalen Datenaustausch (synchron), z. B. Echtzeitdienste wie VoIP
- Messaging (Nachrichtenübermittlung im weiteren Sinne): asynchrone Datenübermittlung, die eine Zwischenspeicherung ermöglicht, z. B. E-Mail
- Retrieval (Abrufdienste im weiteren Sinne): Zugriff auf (zentrale) Server zur Datennutzung (was im Umkehrfall auch Uploads einschließen sollte), z. B. Web-Browsing

Die ITU-T G.1010 bildet Klassen aus der Unterscheidung zur Auswirkung von Fehlern und Verlusten bei der Übertragung als Fehlertoleranz (tolerant vs. intolerant) orthogonal zum zeitlichen Rahmen der Nutzung ([178] S. 5f und siehe auch [179] S. 328): interactive: << 1s, responsive: ~2 s, timely: ~10 s, non-critical: >> 10 s. Es ist leicht ersichtlich, dass ein Begriff wie „interactive“ schnell falsch interpretiert werden kann, wenn nicht klargestellt ist, ob die Interaktivität in der Nutzsicht oder wie hier der zeitliche Rahmen einer Klasse gemeint ist.

⁸¹ noch von Integrated Digital Service Network (ISDN) ausgehend, also von Datendiensten in der Telkowitz, siehe [177]

Eine weitere Möglichkeit ist die Klassifizierung auf Basis der Charakteristik des QoS-QoE-Zusammenhangs, ausgehend von der Sensibilität auf die Überschreitung festgelegter Grenzen⁸² bestimmter Netzparameter. Ein Beispiel für Paketverlust und Latenz in drei Klassen⁸³ findet sich in [158] S. 46:

- QoS tolerant: auch unter ungünstigen Bedingungen mit erhöhter Paketverlustrate und Latenz noch gut nutzbar
- QoS sensitiv: mit moderat erhöhter Paketverlustrate und Latenz noch nutzbar, aber schon Probleme festzustellen
- QoS kritisch: erhöhte Paketverlustrate und Latenz führen zu QoE-Einschränkungen

3.6.6 Klassen aus Nutzerstimuli

Für eine Nutzungsart sollte von jeweils vergleichbaren aufgabenorientierten Nutzeranforderungen auszugehen sein. Für die Betrachtungen aus der Nutzersicht sollte dabei klar sein, dass sich ein naiver Nutzer keine Gedanken darum macht, wie sein Wunsch „hier und jetzt“ einen lustigen Youtube-Clip zu schauen, technisch eingeordnet wird, noch welche Netznutzung oder Inhaltsdatenrate dahinterstehen. Für den Nutzer zählen nur seine (meist impliziten) Anforderungen, geprägt durch seine Erwartung, aus *seiner eigenen* Nutzersicht und damit korrespondierende Einschränkungen in *seiner eigenen* Nutzersicht. Entsprechend kann die Beschreibung dieser Größen auch nur so erfolgen, dass sie ein Nutzer prinzipiell verstehen kann. Dies bedeutet beispielsweise maximale Start-, Lade- oder Übertragungsdauern statt Übertragungsraten und Latenz. Die Anforderungen und Einschränkungen sind in der Nutzersicht entsprechend mit den Nutzerstimuli zu beschreiben. Dabei sollten vergleichbaren Anforderungen an die relevanten Leistungsmerkmale, ihre Präsentation und die QoS-unabhängigen Leistungsbestandteile zu identifizieren sein, die sich in Klassen zusammenfassen lassen. Nachfolgend sind einige der Überlegungen zu den jeweils relevanten Ausprägungen der Nutzerstimuli dazu aufgeführt.

⁸² genaue Werte hier ausgelassen, da in genannter Quelle nicht auf die mobile Nutzung fokussiert

⁸³ angepasste Übersetzungen der Bezeichnungen

Für (Bulk-) **Daten- / Dateitransfers** (im Hintergrund) ohne direkte Interaktion, z. B. App-Downloads, Senden und Empfangen von E-Mails, etc.:

- Zeiten und Dauern: Dauer bis Down- oder Upload vollständig ausgeführt ist (eher indirekt)
- Funktion und Verhalten: korrekter Abschluss
- Qualität und Quantität: ggf. Dateigröße

Für die **interaktive Nutzung**, z. B. Web-Browsing, Informationsbeschaffung und die Nutzung der meisten Apps:

- Zeiten und Dauern: Dauer, bis die gewünschte Aktion komplett ausgeführt ist, z. B. bis die Webseite im Browser inkl. aller (sichtbaren) Elemente (vollständig) geladen ist
- Funktion und Verhalten: (noch) nicht oder nur unvollständig geladene Elemente
- Qualität und Quantität: Anzahl Elemente, wie z. B. Bilder und deren visuelle Qualität

Für **Messaging**, z. B. WhatsApp, etc.:

- Zeiten und Dauern: Dauer zwischen Absetzen der eigenen Nachricht und Eintreffen der Bestätigung der Zustellung (falls vorhanden) bzw. der erwarteten Antwort auf die Nachricht
- Funktion und Verhalten: verlorene Nachrichten, verspätete Zustellung
- Qualität und Quantität: Umfang, z. B. angehangene Elemente wie Bilder etc.

Für (Multimedia-) **Streaming**, z. B. Audio-/Video-Streaming:

- Zeiten und Dauern: Vorpufferzeit, Anfangsverzögerung (Puffer füllen), Synchronität
- Funktion und Verhalten: Anzahl, Verteilung und Dauer von Unterbrechungen oder Sprüngen in der Wiedergabe, Abbruch der Wiedergabe
- Qualität und Quantität: Audio- / Videoqualität (in angemessener Art, also keine Briefmarkengröße mit Minidatenrate etc.)

Für **Echtzeitanwendungen**, z. B. Echtzeitkommunikation, Echtzeit-Multimedia, Video-Telefonie, Chat, etc.:

- Zeiten und Dauern: Antwortzeit bzw. Reaktionszeit (innerhalb enger Grenzen)
- Funktion und Verhalten: Effekte in Form von Bild/Ton-Artefakten, Sprüngen, Aussetzern, Klötzchen und Hängern
- Qualität und Quantität: Audio- / Videoqualität (siehe Streaming)

Wie bei der Diskussion von Funktion und Verhalten schon einmal andiskutiert, ist die Anforderung *Fehlerfreiheit* implizit meist schon enthalten. Für Echtzeit und Streaming ist sie grundsätzlich wünschenswert für die restlichen Klassen erforderlich.

[180] S. 9 definiert einige Service-Typen und zugehörige relevante QoE-Parameter, die sich mit den eigenen Überlegungen weitgehend decken. Weitere QoE- und QoS-Anforderungen (in der stationären Anwendung) für verschiedenste Dienstklassen sind beispielsweise im Technical Report TR-126 des DSL-Forums [40] enthalten.

3.6.7 Dienstklassen zum QoS-QoE-Zusammenhang

Die Abbildung gleichartiger QoS-QoE-Zusammenhänge in Klassen kann sinnvoll sein. Dabei ist eine möglichst geringe Anzahl Klassen zu bevorzugen. Das Konzept der Quality of Transactions (QoT) bietet ein einfaches Klassifizierungsschema für drei Grundklassen. Darin lassen sich die QoE-Anforderungen unterschiedlichster Dienste einordnen. Dabei wird eine Erklärung versucht, wie die netzbedingten Ursachen von Einschränkungen einzuordnen sind [118] S. 5f. In den eigenen Überlegungen wurde dieser Ansatz aufgegriffen und weiterentwickelt, siehe Tabelle 16. Dazu wurden zusätzliche Kriterien hinzugefügt, die den Zusammenhang über Anforderungen und Einschränkungen besser verdeutlichen und dabei die spezifischen Ursachen sowie mögliche extern fixierte Größen berücksichtigen. Aus diesen Überlegungen kann auf den Betrachtungsrahmen sowohl in der Nutzersicht wie auch in der Netzsicht geschlussfolgert werden. Daraus ergeben sich die Relevanz der betrachteten Netzparameter und die Annahmen zur Charakteristik des QoS-QoE-Zusammenhangs für Dienste der jeweiligen Klasse.

Für den Betrachtungsrahmen spielt das Konzept der Transaction eine wichtige Rolle. Als Transaction werden dabei alle zusammengehörigen Netzaktivitäten bezeichnet, die zu einer entsprechenden Nutzerwahrnehmung führen (siehe 3.4.6 zum Zusammenhang mit der Netzsicht). Die Anzahl zusammengehöriger Transaktionen kann für die Dienste mit dem Grad der Interaktion zusammenhängen:

- einzelne Transaktion, z. B. Download einer Datei (wie in Drive oder bei Youtube)
- viele Transaktionen, die als inkrementeller Fortschritt wahrnehmbar werden können, z. B. bei sukzessiver Aktualisierung der Kartenkacheln in Google Maps, Laden mehrerer Seiten beim Web-Browsing
- kontinuierliche Transaktion, z. B. beim Streaming oder Echtzeitdiensten (wie bei VoIP)

Auch die jeweilige Netzsicht mit Punkten zu Netznutzung und Übertragungscharakteristik sind für die Wirkung möglicher Einschränkungen zu beachten. [181] weist diesbezüglich auf den jeweiligen Betrachtungsrahmen in der Netzsicht hin. Für Echtzeit zählt quasi jedes Paket, womit auch eine Betrachtung hinsichtlich Anforderungen und Einschränkungen auf Paketebene sinnvoll ist. Für Dienste mit elastischem QoS-QoE-Zusammenhang ist hingegen eine Betrachtung auf Flow-Ebene besser geeignet (siehe 3.4.6 für Erläuterungen).

Für konkrete Dienste ist die jeweilige Basisklasse der Ausgangspunkt weiterer Überlegungen. Bei genauer Betrachtung zeigt sich, dass die Dreiteilung von QoT recht grob ist. Durch Unterschiede in der technischen Sicht und Netznutzung wie auch in der Nutzersicht können aus diesen Grundklassen weitere nachgelagerte Klassen abgeleitet werden (Hierarchie). Dabei sind jeweils spezifische Besonderheiten möglich. Entsprechend ist immer genau zu überlegen und zu überprüfen, inwieweit verallgemeinerte Annahmen möglich bzw. konkrete Anpassungen notwendig sind.

Ausgehend von der eingeführten Klassifizierung werden nachfolgend noch die drei Grundklassen der Dienstklassifizierung einzeln vorgestellt (Finish Time dabei mit weiterer Auftrennung in Hintergrund und interaktive Dienste).

Tabelle 16: Dienstklassen und relevante Unterscheidungskriterien

Kriterium	Basisklasse		
	Echtzeit	Streaming	Finish-Time
Was wird negativ bemerkt? (gemeinsame Auswirkung in Nutzersicht)	Beeinträchtigungen bei Funktion und Verhalten Zeiten und Dauer: - Verlust von Kontrolle (Antwort-/Reaktionszeit) - Verlust von Synchronität Verlust von Qualität	Unterbrechungen der Wiedergabe Wartezeit zum Wiedergabestart Zeiten zum Nachpuffern niedrige Qualität (unscharf, ...)	benötigte Dauer
Was ist die Ursache? (gemeinsames technisches Kriterium für Einschränkungen)	Verletzung der Deadline (Ausbleiben eines Paketes bis zu einem bestimmten Zeitpunkt, durch Verlust oder verspätetes Eintreffen)	Leerlaufen des Puffers (Knick in Abspielkurve) Wartezeiten zum Vorpuffern (Füllen des Puffers) Qualitätsstufe (Datenrate)	Zeitpunkt (Finish-Time) des Eintreffens des letzten zugehörigen Paketes der Transaction
extern fixierte relevante Größe (mit Quelle der Anforderung)	max. erlaubte Verzögerung (durch Nutzer oder Quelle) <i>und/oder</i> min. notwendige Datenrate als notwendige Datenmenge bei fixierter Zeit (durch Inhalt)	min. Datenrate (Qualitätsstufe) als notwendige Datenmenge pro Zeiteinheit gemittelt (durch Inhalt)	ohne
zeitlicher Aspekt der Anforderungen	zu jedem Zeitpunkt (kontinuierlich und konstant)	kontinuierlich oder gepuffert in einzelnen Bursts	so schnell wie möglich (Burst)
zeitlicher Ausgleich von Einschränkungen	unmöglich	durch Puffer	unnötig
Relevanz der Fehlerfreiheit (zielt auf Paketverlustrate und Korrekturmechanismen)	wünschenswert	wünschenswert oder erforderlich in Abhängigkeit der Technik	erforderlich

Fortsetzung der Tabelle auf der nächsten Seite

Fortsetzung der Tabelle 16

Kriterium	Basisklasse					
	Echtzeit		Streaming		Finish-Time	
Durchsatz	min. Datenrate innerhalb des gegebenen engen zeitlichen Fensters		effektive Datenrate im zeitlichen Mittel des gegebenen Puffers		effektive Datenrate über gesamte Dauer bis zum Abschluss der Übertragung	
Latenz und Jitter	sehr kritisch		mittel		kritisch bis unkritisch in Abhängigkeit der notwendigen Anzahl Turns (durch Inhalt)	
Betrachtungsrahmen der Übertragung (Netzsicht)	Pakete		Pakete oder Flow		Flow	
effektiver Änderungsbereich	eng		variable Charakteristik (typisch relativ eng)		weit	
mögliche Unterklassen durch Nutzersicht oder Technik/Netzsicht	harte Echtzeit	weiche Echtzeit	Live-Streaming	On-Demand Streaming	interaktiv Time-to-first-Byte	Hintergrund Time-to-last-Byte

eigene Weiterentwicklung, Abwandlung und deutliche Erweiterung ausgehend von [182] S. 6 (QoS-Klassen)

3.6.8 Klassifizierung nach zeitlichen Anforderungen

Der zeitliche Rahmen der Betrachtungen ist für viele Klassifikationsansätze ein Hauptunterscheidungskriterium, sowohl ausgehend von den Anforderungen der Nutzer als auch den Einschränkungen des Netzes. Auch in der Rückführung auf einige wenige Dienstklassen zum grundsätzlichen QoS-QoE-Zusammenhang ist die Zeit entscheidend. Dabei entsteht die besondere Bedeutung dieser Überlegungen aus der Berücksichtigung des zeitlichen Rahmens der Anforderungen je Transaktion an die einzelnen Parameter. Wie in der Diskussion zum Entstehen und zum Zusammenwirken der Nutzerstimuli diskutiert, hängt die Zeit, als eine der Kategorien von Nutzerstimuli, dabei eng mit den beiden anderen Kategorien zusammen. Eine Zusammenstellung zum zeitlichen Rahmen mit Hinweisen zur adäquaten Präsentation findet sich in Abbildung 32, inklusiver bedeutsamer Zeitmarken.

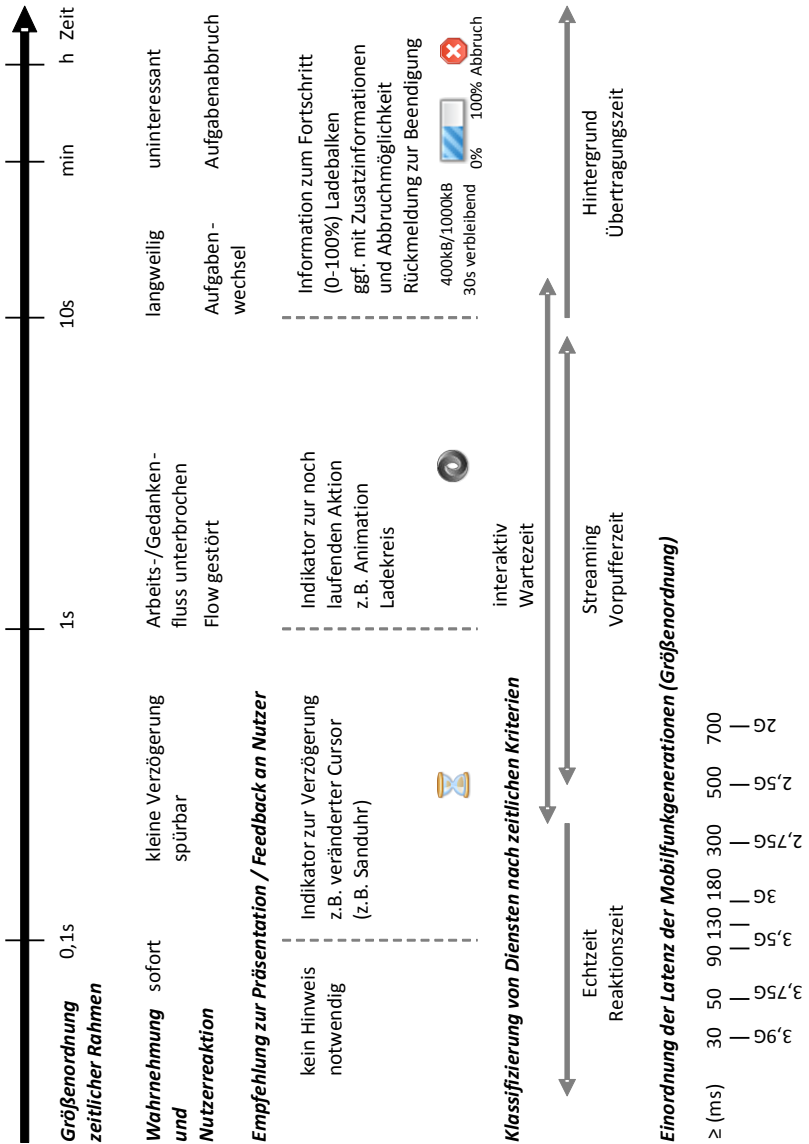


Abbildung 32: Die Größenordnungen relevanter Zeitmarken und die Rolle der Zeit in verschiedenen Sichten
 zusammengestellt in Anlehnung an [81] S. 135ff, [183] S. 19 und [182] S. 3, erweitert und ergänzt

3.6.9 Echtzeit: Wann, wenn nicht jetzt?

Die Unterscheidung von Echtzeit (Realtime) und Nicht-Echtzeit (Non-Realtime) ist für viele Klassifikationsansätze der Ausgangspunkt von Nutzungsklassen und Netzklassen, entweder ausgehend von den Anforderungen der Nutzer oder den Anforderungen an bzw. Einschränkungen durch das Netz. In der weiteren Diskussion ist es sinnvoll die Quelle, also Art und Herkunft der Anforderungen und die Auswirkung der Verletzung einzubeziehen, um so die Bedeutung des Begriffs Echtzeit zu erläutern.

Als Quelle der Anforderungen Echtzeit sind wenigstens zwei Ursachen auszumachen, auf die nachfolgend noch weiter eingegangen wird:

- interne zeitliche Koordination: vom Nutzer selbst ausgehend (enge Kontrollschleife)
- externe zeitliche Synchronisation: Vorgabe von außen (Synchronität zwischen Quelle und Senke)

Die Auswirkungen von Einschränkungen in der Nutzersicht können in Anlehnung an weiche und harte Echtzeit beschrieben werden, je nachdem wie stark sich die Verletzung einer Deadline auswirkt, bzw. wie weit diese zu tolerieren bzw. auszugleichen ist (Erfüllung zu jeder Zeit vs. Erfüllung im zeitlichen Mittel):

- weich: Einschränkungen sind wahrnehmbar, z. B. erhöhte Konzentration notwendig, Qualitätseinbußen sind feststellbar etc., Dienst aber noch bedingt nutzbar
- hart: Funktion und Nutzen sind stark eingeschränkt oder ggf. gar nicht mehr gegeben

Typischerweise zeichnen sich die Dienste, bei denen die Echtzeitanforderung vom Nutzer selbst ausgeht, durch eine enge Kontrollschleife auf Nutzerseite aus. Die akzeptable Zeit zwischen Aktion und erwarteter bzw. notwendiger Reaktion ist dabei verhältnismäßig kurz. Der Nutzer erwartet im Prinzip eine sofortige Reaktion auf eine eigene Aktion. Daher ist der sinnvolle Bereich für eine entsprechende Reaktions- bzw. Antwortzeit mit einer Obergrenze von unter 100 ms bis 400 ms anzunehmen (siehe nachfolgende Beispiele mit zugehörigen Referenzen).

Beispiele für typische Anwendungsszenarien mit Echtzeitanforderungen der Nutzer wären:

- Entertainment: Online Gaming mit <100 ms für Action-Spiele, z. B. Multiplayer First Person Shooter [184]
- Kommunikation: Internettelefonie (Voice over IP) bzw. Videotelefonie (z. B. *Skype*⁸⁴ oder *Apple Facetime*⁸⁵) von max. 25 ms Mund-zu-Ohr-Verzögerung ohne Echokompensation (störende Rückkopplung) bzw. von noch guten 150 ms hinauf zu gerade noch akzeptablen max. 400 ms mit Echokompensation [185] S. 180
- Business: Remote Desktop und Remote Shell Nutzung (Arbeiten auf entfernten PCs und Servern) mit ca. 200 ms, die zur „normalen“ Verzögerung der lokalen interaktiven Nutzung noch hinzukommen [116] S. 1356

Eine andere Art von Echtzeitanforderungen entsteht durch die Synchronisation mit einer externen Quelle oder Senke. Dabei kann weiter unterschieden werden zwischen „echter Echtzeit“ und Echtzeit-Streaming:

- „echte Echtzeit“: Erzeugung, Transport und Konsum der Daten in Echtzeit
- Echtzeit-Streaming: Transport und Konsum erfolgen in Echtzeit, die Erzeugung nicht zwangsläufig – außer beim Live-Streaming

Ohne Zwischenspeicherung (Puffer) müssen anfallende Daten genauso schnell transportiert und konsumiert werden können, wie sie produziert werden. Echtzeit-Streaming ist dabei eigentlich ein Spezialfall, der Echtzeit und Streaming zusammenbringt, z. B. bei IPTV und Live-Video-Streaming, bedingt durch eine externe Synchronisationsquelle.

Besonders hohe Anforderungen aus Netzsicht stellt Cloud-Gaming, siehe [186]. Dabei trifft die Forderung nach niedrigen Verzögerungszeiten für die Interaktion mit der Spielwelt auf gleichzeitig hohe Datenraten (mit bis zu 10 MBit/s) für das Streaming der hochauflösenden Ansichten der Spielwelt in Echtzeit, die von leistungsstarken Servern (in der Cloud) berechnet werden.

⁸⁴ <http://www.skype.com/de/> abgerufen am 03.03.2015

⁸⁵ <https://www.apple.com/de/ios/facetime/> abgerufen am 03.03.2015

Die Relevanz solcher Dienste hinsichtlich einer mobilen Nutzung ist allerdings noch unklar, das Beispiel soll aber illustrieren, dass für einzelne Dienste die Anforderungen kombiniert gesehen werden müssen.

Gibt es feste und enge zeitliche Anforderungen auf Nutzerebene, die an eine entsprechende Netznutzung gebunden sind, so entstehen aus diesen Anforderungen auch Echtzeitanforderungen auf Netzebene für den Zustellungstyp einzelner Pakete. Im Fall von Echtzeit-Anforderungen an das Netz ist zu unterscheiden, ob diese zu jedem Zeitpunkt (kontinuierlich) oder im zeitlichen Mittel gelten und durch Puffer ausgeglichen werden können [157] S. 156. Im Zustellungstyp Echtzeit müssen die Daten innerhalb gewisser Zeitschranken (vor der Deadline) ihr Ziel erreichen. Dabei ist die durch das Netz vorgegebene Verzögerung (Einweg) bzw. Latenz (Zweiwege-Kontrollschleife) im mobilen Bereich oftmals schon so hoch, dass sehr anspruchsvolle Nutzungen, die geringe Verzögerungszeiten voraussetzen, a priori ausgeschlossen sind (siehe Abbildung 32 unten zum Vergleich der technologisch bedingten Verzögerung im Vergleich zu den Anforderungen). Entsprechend spielen die Annahmen zu den psychologisch bedingten Echtzeit-Anforderungen in der Entwicklung der Netze der übernächsten Generation (5G) eine große Rolle, um zukünftig auch Dienste mit extrem kurzer Kontrollschleife (wie die motorische Kontrolle, z. B. bei Teleoperationen) ermöglichen zu können, siehe [187].

Die benötigte Datenrate des genutzten Inhalts (Quantität) bestimmt den notwendigen Durchsatz für die Datenübertragung. Entsprechend scheiden Dienste mit notwendigen Datenraten oberhalb der technischen Möglichkeiten des Netzes zur Nutzung weitgehend aus, da zwangsläufig die Anforderungen der Echtzeit durch Queuing und Paketstaus oder Fehlerfreiheit durch das Verwerfen von Paketen verletzt werden müssen. Anforderungen an die Fehlerfreiheit wirken verschärfend auf die Echtzeitanforderungen, da eventuell notwendige Neuübertragungen zum Ausgleich von Paketverlusten innerhalb des zeitlichen Budgets der gestellten Anforderungen erfolgen müssen.

Bei gelegentlich anzutreffenden Angaben von Zeitmarken zur Echtzeit ist zu beachten, ob sich die Angaben zum zeitlichen Rahmen auf die Nutzeranforderungen oder die Netzanforderungen (Latenz) beziehen. Bei Angaben zu Nutzeranforderungen sind ggf. noch zeitaufwendige lokale Verarbeitungsschritte

(z. B. Audio-Sampling und Encoding) vom genannten Zeitbudget abzuziehen. Das Zeitfenster für die eigentliche Datenübertragung und damit die maximale Latenz kann sich so weiter verringern (eine beispielhafte Auflistung der Zeitkomponenten für VoIP ist in [188] S. 37ff zu finden, eine Beispielrechnung in [189] S. 56ff).

Echtzeit-Services wurden in den bisher durchgeführten Testrunden nicht betrachtet. Nachfolgend wird deshalb auf weitere Ausführungen dazu verzichtet. Am nächsten noch in den Bereich der (weichen) Echtzeit einzuordnen ist die Server-basierte Autovervollständigung in Such- und Eingabefeldern. Werden dabei die Vorschläge erst mit merklicher Verzögerung eingeblendet, dann kann dies weniger nützlich, überflüssig oder gar störend sein.

3.6.10 Streaming: Zwischen Skylla und Charybdis ... und den Sirenen der Qualität

Beim Streaming kommt der Wartezeit während des Vorpufferns eine entscheidende Rolle zu. Je größer der Puffer desto kleiner sind die Auswirkungen zwischenzeitlicher Schwankungen der Datenrate des Inhalts oder der Übertragung nach dem Start der Wiedergabe. Die Wahl des richtigen Pufferfüllstands zum Wiedergabestart ist damit ein Kompromiss zwischen der Gefahr des Leerlaufens während der Wiedergabe und der Wartezeit zum Vorpuffern, also quasi die Wahl zwischen zwei Übeln – wie bei Skylla und Charybdis in der griechischen Mythologie. Für die Nutzergeduld zum Vorpuffern sind dabei gewisse Obergrenzen anzusetzen. Die technischen Anforderungen ergeben sich hauptsächlich aus der genutzten Qualitätsstufe, also aus dem Inhalt, konkret aus der Datenrate. Dabei wirken die hohen Qualitätsstufen wie die Sirenen der griechischen Mythologie: sehr verlockend. Stehen mehrere Qualitätsstufen zur Verfügung, dann kann durch passende Indikatoren, z. B. durch ein HD-Logo, ein Feedback zur visuellen Qualität des Inhalts und damit verbunden der Quantität an die Nutzer erfolgen (Moderation der Erwartung).

Ob Streaming als echtes Streaming (ohne Erzeugung einer vollständigen lokalen Kopie) oder als Pseudo-Streaming (Download) erfolgt, ist in dieser Betrachtungsweise nicht vordergründig entscheidend, sowohl aus Netz- wie auch aus Nutzersicht. Der verbindende Aspekt ist der Puffer als Zwischenspeicher zum kurz- oder mittelfristigen Ausgleich von Schwankungen der effektiven

Datenrate. Ideal wäre eine Strategie, die nur solange vorpuffert, dass es nach dem Wiedergabestart nicht zum Leerlaufen des Puffers und damit zu einer Unterbrechung kommt, dabei aber nicht über eine maximale Wartezeit hinausgeht. Der Zusammenhang der beteiligten Nutzerstimuli dazu diene als Beispiel in der bereits zuvor gezeigten Abbildung 27.

Bei modernen Streaming-Technologien entfällt die Notwendigkeit einer vollständigen Übertragung des Inhalts vor der Nutzung (dies wäre ein vom Endzeitpunkt dominierter Download und damit in der Klasse Finish-Time). Wie bei den Echtzeitdiensten ist daher eine kontinuierliche Netznutzung zur Übertragung des Inhalts parallel zur Nutzung des übertragenen Inhalts als charakteristisches Merkmal für Streaming auszumachen. Im Unterschied zur Klasse Echtzeit kann der Puffer aber so groß sein, dass eine Übertragung des Inhalts in einzelnen Burst möglich und sinnvoll sein kann. Dieser Aspekt leitet zum Adaptive Bitrate Streaming (ABR) über.

Beim Adaptive Bitrate Streaming wird der Inhalt (Qualitätsstufe und damit die Datenrate) dynamisch passend zu den Einschränkungen des Netzes gewählt, ausgehend von den Anforderungen zu Unterbrechungsfreiheit und kurzer Startdauer. Im Grunde war die Entwicklung der Techniken zum Adaptive Bitrate Streaming eine konsequente Reaktion auf fehlende Ende-zu-Ende-QoS-Mechanismen bei gleichzeitig veränderlichen Netzbedingungen im Internet, speziell in der mobilen Nutzung.

Für die Erfüllung der Anforderungen beim Streaming lässt sich zusammenfassen, dass die benötigte Bandbreite (vom Inhalt vorgegeben) mit kleinen Puffern im Mittel stabil gewährleistet werden muss. Mit Blick auf die genutzten Protokolle und Technologie soll auf die Anmerkung zu All-over-HTTP (siehe 3.4.2 und Abbildung 25) und die weiteren Ausführungen zum untersuchten Dienst Youtube (siehe 9.6.2) verwiesen sein.

3.6.11 Finish Time: Ist es schon fertig?

Die von Wartedauern dominierten Dienste können in der Klasse Finish Time zusammengefasst werden. Der Name leitet sich dabei von der benötigten Zeit bzw. dem Zeitpunkt zum Abschluss der zugeordneten Netznutzung an. Finish Time soll zur weiteren Diskussion in interaktive Anwendungen und eine nicht-

interaktive Klasse Hintergrund unterschieden werden. Die Unterscheidung kann sich direkt aus der Nutzersicht in Vordergrund und Hintergrund ergeben, je nachdem ob die verstrichene Zeit der Übertragung wahrgenommen wird oder nicht.

Für **interaktive Dienste** spielt die Dauer der Übertragung eine Rolle, kürzer ist demnach prinzipiell auch besser. Auf diese Klasse wird aufgrund der hohen Relevanz im nachfolgenden Abschnitt noch gesondert eingegangen. In anderen Taxonomien sind vergleichbare Bezeichnungen und Klassen zu finden, z. B. Responsive, Interactive,

Für **Hintergrunddienste** ist die genaue Dauer ohne Bedeutung, wichtig ist nur, dass die Übertragung jemals oder zumindest zu einem bestimmten Zeitpunkt abgeschlossen ist. Beispiel: Die Synchronisation sollte immer zur vollen Stunde abgeschlossen sein, dann ist es egal, ob es 5 min, 10 min oder 45 min dauert, solange nicht mehr als 60 min benötigt werden, da dann die Aktion noch nicht abgeschlossen wäre, wenn bereits die nächste starten sollte. Alternative Bezeichnungen mit vergleichbaren Konzepten ergeben sich aus anderen Taxonomien: Background, Dateitransfers, Bulktransfers, Store-and-Forward (z. B. von Textnachrichten und E-Mails) und Non-critical-Traffic.

Aus der gemeinsamen Betrachtung von zeitlichem Rahmen und Datenmenge ergibt sich für Finish Time die Abschätzung zur notwendigen Datenrate (als Datenmenge pro Zeit):

- eine kleine Datenmenge ist in Bezug auf die Zeit unkritisch, da bereits geringe Datenraten ausreichen
- eine große Datenmenge ist bei ausreichend Zeit ebenfalls unkritisch
- eine große Datenmenge innerhalb kurzer Zeit ist hingegen herausfordernd, da hohen Spitzendatenraten notwendig sind

3.6.12 Interaktive Dienste

Das charakteristische Merkmal der interaktiven Dienste aus Nutzersicht ist die Interaktion mit der App bzw. dem Inhalt, was zu einzelnen Paaren aus Aktion und Reaktion führt. Diese Interaktionen der Nutzersicht werden in der Netzsicht typischerweise ebenfalls in Aktion und Reaktion als Request-Response-Paare (Turns bei TCP/HTTP) umgesetzt. Damit gilt für die Dienste in dieser

Klasse weitgehend: je höher die Nutzerinteraktivität, desto höher die Netzin-teraktivität. Der Großteil der typischen Dienste im mobilen Internet ist in diese Klasse einzuordnen, z. B. Web-Browsing, bzw. alle Dienste, bei denen die Netznutzung als Antwortzeit ohne Echtzeitanforderungen oder als Wartezeit im Vordergrund wahrgenommen wird.

In der Mensch-Maschine-Interaktion (HCI) beschäftigte man sich schon früh beim Umgang mit Dialogsystemen mit der Zeit, vor allem in Bezug auf die Reaktionszeit interaktiver Systeme und stieß dabei auf mögliche Probleme (siehe [83] S.110f):

- zu langsame Reaktion (z. B. Eingabe nicht sinnvoll möglich, weil Feed-back stark verzögert)
- zu schnelle Reaktion (Ausgabe nicht sinnvoll wahrnehmbar)
- variable Reaktionszeit, die den Nutzer keinen Erwartungswert bilden lässt, weshalb ggf. auch künstlich erzeugte Mindestzeiten sinnvoll sein können

Erhöhte Antwortzeiten stören die Interaktion mit dem Nutzer, besonders dann, wenn sie nicht erkennbar sind (Mangel an Feedback), siehe [83] S.112ff. Wenn zusätzlich noch die Antwortzeit unvorhersehbar ist, also stark schwankt, verschärft sich das Problem, da der Nutzer keine passende Erwartungshaltung aufbauen kann. Dies veranlasst den Nutzer u. U. seine Aktion zu wiederholen, in der Annahme selbst einen Fehler gemacht zu haben, oder dass das System seinen Befehl ignoriert hat, bzw. dieser verloren gegangen ist (nach [83] S.113f). So entsteht leicht Frustration beim Anwender, der ggf. sogar auf Fehler in Funktion und Verhalten schlussfolgert. Eine ausführliche Auseinandersetzung zum Thema Zeitwahrnehmung in der Mensch-Maschine-Interaktion findet der interessierte Leser in [163]. Die Erkenntnisse zu interaktiven Systemen sollten sich in angemessener Art auch für interaktive Dienste anwenden lassen.

[190] S. 7f fasst einige Arbeiten zur (Antwort-) Zeit im Bereich der Anwendungs-Performance-Bewertung zusammen. Die darin genannten Quellen und Erkenntnisse zu den immer wieder auftauchenden grundlegenden Zeitmarken sind nicht neu. Sie gehen bis auf Studien vom Ende der 60er Jahre von Robert

B. Miller zurück (siehe [191]). Auf dessen Arbeiten basieren quasi alle weiteren Empfehlungen etc. zu Antwortzeiten und Zeitverhalten zur User Experience beim Zusammenwirken von Mensch und Maschine. Auch die berühmte Arbeit von Nielsen zur Usability interaktive Systeme von Anfang der 90er Jahre, siehe [81] S. 135ff und Update in [196], nennt drei Zeitmarken auf dieser Basis zur Orientierung: 0,1 s; 1 s und 10 s, siehe Abbildung 32. Nielsen gibt dazu auch Einschätzungen zum menschlichen Empfinden und Empfehlungen zur Rückmeldung an den Nutzer an, siehe [81] S. 135ff. Neben diesen drei Zeitmarken finden sich noch leicht abweichende Angaben etwa in der ITU-T G.1010 zur Klassifizierung von Diensten (siehe Angaben in 3.6.5) oder für Antwortzeiten interaktiver Systeme (in der HCI) in [159] S. 96:

- bis ca. 1 s: Antwort wird als unmittelbar empfunden
- bis ca. 5 s: verzögert
- bis ca. 10 s: stark verzögert
- über 10 s: Nutzer geht davon aus, dass keine Antwort mehr erfolgt ... und reagiert selbst wiederum darauf: löst die Aktion noch einmal aus, oder gibt auf, oder ...

In Orientierung an die genannten drei Zeitmarken aus [81] S. 135ff ergeben sich daraus auch Empfehlungen für eine adäquate Präsentation der Zeit und des Feedbacks zum Fortschritts in der Bearbeitung, worauf sich die nachfolgenden Erläuterungen beziehen. Der Wert von 1 s gilt als Marke, bis wohin der Nutzer noch im Flow bleibt, selbst wenn er schon Verzögerungen wahrnehmen kann. Diese Marke kann daher für Anwendungen als wichtig gelten, die mehrere konsekutive Interaktionen erfordern (z. B. bei Google Maps zum gezielten Navigieren auf der Karte, weitere Erläuterungen dazu unter 9.3). Beim mobilen Web-Browsing kann der Wert von 1000 ms für die Darstellung des unmittelbar sichtbaren Bereiches als eine Art magische Zielmarke gelten (siehe [192] für eine ausführliche Erläuterung). Für derartige Verzögerungen reicht eine kurze Rückmeldung an den Nutzer, dass das System umgehend zur weiteren Nutzung wieder bereitsteht. Für Aktionen, die etwas länger dauern, sollte eine Rückmeldung an den Nutzer erfolgen, dass das System noch aktiv beschäftigt ist, was z. B. in Form von Animationen erfolgen kann. Dazu können auch die Endlosanimationen der Aktivitätsindikatoren mit Ladekreis (Spinner) etc. gezählt werden.

Für noch länger dauernde Aktionen (über 10 s) sollte als Empfehlung aus Usability-Sicht ein Fortschrittsbalken eingesetzt werden, der eine prozentuelle Fortschrittsanzeige enthält, möglichst zusammen mit Zusatzinformation wie der Gesamtdatenmenge und einer Abschätzung zur verbleibenden Dauer bis zum Abschluss (siehe [81] S. 136 mit Beispiel). Gleichzeitig sollte für den Nutzer die Möglichkeit bestehen, die Aktion über eine Schaltfläche etc. abubrechen. Besteht aus Nutzersicht keine Veranlassung zum aktiven Warten, kann die Übertragung auch verdeckt im Hintergrund erfolgen. Die Darstellung des Fortschritts kann ebenfalls verdeckt erfolgen. Dann reicht ggf. ein kurzer Hinweis zum erfolgreichen Abschluss in der direkten Nutzersicht aus.

Die als allgemeingültig angesehenen Erkenntnisse aus den vorherigen Referenzen geben damit den Rahmen an, in dem sich Antwortzeiten und Ladedauern aus Nutzersicht bewegen dürfen. Insgesamt sollte ein von der Erwartung, d. h. der Geduld der Nutzer, geprägter elastischer QoS-QoE-Zusammenhang für Dienste dieser Klasse anzunehmen sein. [193] führt aus, dass bei von der Zeit dominierten Diensten (vornehmlich Wartezeit) sich ein Grundmuster erkennen lässt, wie weit Toleranzschwelle (keine Beeinträchtigung festzustellen) und Frustrationsschwelle (nicht mehr akzeptable Beeinträchtigungen) auseinanderliegen. Je nach Kategorien von Interaktion und Dienst sind unterschiedliche Grundwerte anzusetzen. Es ist von der Zufriedenheit zur Frustration aber immer annähernd Faktor 4 in der Zeit festzustellen, z. B. von 1 s zu 4 s, von 2 s zu 8 s, von 10 s zu 40 s usw. Der Faktor 4 kann auch als zweimaliger Verdoppelungsschritt oder 4 Halbschritte mit Faktor $\sqrt{2}$ beschrieben werden. In der Festlegung der Parametrisierung für die Tests der ausgewählten Dienste wird noch einmal auf diese Erkenntnis zurückgekommen. Ergänzend zu [193] sei auf [194] verwiesen, wie sich daraus ein Application Performance Index (APDEX) bei bekannter Toleranzschwelle und tatsächlich benötigter Zeit konstruieren lässt. Weitere Hinweise zur Problematik der Wahrnehmung von Wartezeiten im QoE-Kontext gibt [195].

Die Erkenntnisse zu den relevanten Zeitmarken und zur passenden Präsentation sind, wie dargelegt, recht einfach und können mittlerweile als universell gelten [196]. Leider sind diese sinnvollen Empfehlungen aber auch nach über 30 Jahren HCI-Forschung noch immer überall (in den Köpfen der Entwickler/Designer) angekommen – bzw. berücksichtigen sie dies in ihren goldenen

Käfigen der Entwicklungsumgebungen nicht angemessen. Unter Idealbedingungen sind dabei oftmals nämlich keine besonderen Maßnahmen notwendig, da keine längeren Antwort- oder Wartezeiten auftreten. In Kombination mit Netzbedingungen, die nicht optimal sind, können dann allerdings aus der Nichtbeachtung mögliche Probleme entstehen, die noch immer relevant für aktuelle mobile Apps sind. Dies hat sich in den Voruntersuchungen zu den Diensten des Testparcours der eigenen Studie leider bestätigt, dazu später mehr (siehe z. B. in 9.4).

3.6.13 Die Relevanz der einzelnen QoS-Netzparameter

Die Relevanz der einzelnen Netzparameter ergibt sich aus den definierten Anforderungen und den Auswirkungen von Einschränkungen der jeweiligen Dienste. Dabei ist die Relevanz teilweise leicht zu erkennen, etwa bei Mindestdatenraten zur Übertragung von Live-Streaming-Inhalten mit festen Inhaltsdatenraten, teilweise aber auch nur im komplexen Zusammenwirken mehrerer Einflussgrößen zu begründen.

Die Relevanz der **Latenz** für Echtzeitdienste ist leicht zu begründen: je höher die Latenz, desto schwieriger wird es, die engen zeitlichen Vorgaben einzuhalten. Für andere Dienstklassen ist die Rolle der Latenz nicht ganz so vordergründig zu erfassen. Dies führte im Rahmen der Überlegungen zur Auswahl der Dienste, zur Festsetzung der Szenarien und Netzsituationen sowie letztlich auch in der Interpretation der eigenen Ergebnisse mit Partnern und der Fachöffentlichkeit immer wieder zu einer kritischen Auseinandersetzung mit der Latenz. Die Anmerkungen zur Äquivalenz der Parameterwirkung über die internen Schichten des Modells auf die tatsächlich relevanten Nutzerstimuli in der Nutzersicht sollten dabei im Hinterkopf behalten werden. In Abhängigkeit von Inhalt und Netznutzung kann die Latenz bei stark interaktiven Diensten nach der Bandbreite schnell zum „neuen Flaschenhals“ im mobilen Web werden (siehe [197] für Vergleichszahlen mit verschiedenen Werten für Bandbreite und Latenz zum Laden einer Webseite).

Die Relevanz der Latenz steigt in Abhängigkeit des Zusammenwirkens mehrerer Elemente:

- Eigenschaften des Inhalts
 - Anzahl einzelner Elemente und damit der Anzahl Turns
 - Anzahl Gegenseiten und damit der Anzahl DNS-Lookups, TCP-Streams und HTTP-Connections
- Eigenschaften des Inhalts und des Netzes
 - Verhältnis von theoretischer Übertragungszeit für die Datenmenge der einzelnen Elemente bei gegebener Datenrate zur Latenz
- Eigenschaften des Inhalts und der App
 - Parallelität und Sequenzialität der Turns
 - Cache-Nutzung und Revalidierung von Inhaltselementen
- Eigenschaften der App und deren Protokollnutzung
 - HTTP: keep-alive
 - TCP: Anzahl der notwendigen Three-Way-Handshakes

Die Latenz lässt sich technisch oft nur schwierig vermindern, besonders in einer Ende-zu-Ende Betrachtung, wenn nicht nur die Technologie eines Netzsegments eine wichtige Rolle spielt, sondern auch die Entfernung zwischen den Kommunikationspartnern (Propagation Delay). Steigt diese Entfernung, so steigt für bestimmte Dienste auch die Relevanz einer niedrigen Grundlatenz der Netztechnologie. Diese Überlegungen wiederum sind nicht neu, sondern gehen noch auf die „gute alte“ Modem- und ISDN-Zeit der frühen 90er Jahre zurück, siehe z. B. [198]. Der Spitzendurchsatz selbst im mobilen Bereich ist seitdem ca. um Faktor 3000 größer geworden, die Größenordnung der Latenz hingegen hat deutlich weniger stark abgenommen (siehe Tabelle 23 zur Orientierung).

Die Relevanz des **Jitters** steigt zusammen mit der Relevanz der Latenz, im Fall besonders kleiner Latenzwerte und besonders kleiner oder nicht vorhandener Puffer (Streaming/Echtzeit) zum Ausgleich umso stärker.

Die Relevanz des **Durchsatzes** (Datenraten) ergibt sich aus dem angesetzten zeitlichen Rahmen in der Nutzersicht und zu transportierender Datenmenge des Inhalts. Die Elastizität leitet sich aus den Möglichkeiten zum zeitlichen Ausgleich von Einschränkungen ab (Echtzeit: ohne bzw. klein, Streaming:

mittel, Finish Time: groß). Im Fall von Diensten mit festgesetzten zeitlichen Rahmen (Echtzeit und Streaming) ergeben sich so bei fester Datenmenge bzw. Datenrate des Inhalts auch feste Anforderungen an die minimal erforderliche Datenrate der Übertragung.

Ist **Fehlerfreiheit** erforderlich, so steigen die Anforderungen daran zusammen mit den zeitlichen Anforderungen. Sind die zeitlichen Anforderungen ohnehin sehr eng gestaltet, so verbleibt kaum Zeit für einen Fehlerausgleich, z. B. durch Retransmissionen. In diesen Fällen sind (auf Kosten des Durchsatzes) alternative Mechanismen zum Fehlerausgleich eine mögliche Lösung, z. B. die Nutzung von Redundanzen durch Zusatzdaten, z. B. in Verfahren zur Vorwärtsfehlerkorrektur (Forward Error Correction, FEC).

Die Angabe konkreter Werte zu den einzelnen Parametern ist schwierig, aber es sind gute Abschätzungen zu interessanten Bereichen möglich, sofern die richtigen Heuristiken mit sinnvollen Annahmen zu den abhängigen nutzerrelevanten Einflussgrößen eingesetzt werden. Auf diese Überlegungen wird später bei der Festlegung der jeweils sinnvollen Bereiche und Abstufungen der QoS-Netzparametrisierung für die Dienste im Test zurückgekommen. Konkrete Einschätzungen zur Relevanz der einzelnen Parameter für die ausgewählten Dienste sind im Kapitel 9 zur Diskussion des Testparcours enthalten.

3.7 Zusammenfassung

Die QoX-Matrix wurde als geeigneter Modellrahmen für die Untersuchung des QoS-QoE-Zusammenhangs eingeführt. Die Kernpunkte des eigenen Modells können wie folgt zusammengefasst werden:

- Einbeziehung aller relevanten Qualitätselemente aus den Kategorien System, Kontext, Mensch
- Bildung von Qualitätsschichten: mit QoS als unterster Schicht und QoE als oberster Schicht sowie QoD und QoA als Zwischenschichten
- Konstruktion einer Ursache-Wirkungs-Kette zwischen den Schichten zur gezielten Erklärung des Einflusses der QoS (Anfang) mit den variierten unabhängigen Variablen auf die QoE (Ende) mit den interessierenden abhängigen Variablen

- Die Qualitätsmerkmale der Zwischenschichten bilden als Mediatoren die internen Glieder der QoS-QoE-Kette. Sie vermitteln als veränderliche QoS-abhängige Größen die Wirkung der QoS-Netzparameter auf die QoE-Qualitätsaspekte.
- Externe (Neben-) Faktoren wirken orthogonal entlang der QoS-QoE-Kette als Moderatoren auf den internen Zusammenhang ein, wodurch die QoX-Matrix aufgespannt wird.

Die Qualitätselemente Netz, Nutzer, Kontext, Endgerät und Dienst (App, Inhalt, Gegenseite) wurden jeweils einzeln erläutert. Den Diensten (in eigener Definition und Umfang) kommt sowohl aus Netzsicht wie auch aus Nutzersicht eine zentrale Rolle zu. Zur Netzsicht wie auch Nutzersicht auf die Dienste wurden jeweils die relevanten Zusammenhänge herausgearbeitet.

In der Netzsicht erfolgte die Diskussion zu den technischen Einflussfaktoren auf die Transportleistung der QoD. Dies umfasst die Protokolle auf unterschiedlichen Schichten des Netzwerk-Stacks bis zu den TCP-Implementierungen der beteiligten Endpunkte der Kommunikation. Das dynamische Zusammenwirken der Ausgangsgrößen der QoS führt dabei zu den Effektivwerten der QoD. Kriterien zur Beschreibung der Übertragungscharakteristik und Netznutzung von Diensten rundeten die Netzsicht ab.

QoS-veränderliche und QoS-invariante Merkmale der Dienste prägen im gemeinsamen Zusammenspiel die Nutzersicht und damit entscheidend den dienstspezifischen QoS-QoE-Zusammenhang. Dabei erfolgt eine Moderation der Wahrnehmung durch die QoS-unabhängige Präsentation der QoS-abhängigen Applikationsleistung. Zur Beschreibung der Applikationsleistung als Qualitätsmerkmale der QoA wurden drei Kategorien von Nutzerstimuli eingeführt (Funktion und Verhalten, Zeiten und Dauern, Qualität und Quantität) und detailliert erörtert. Diese Größen hängen selbst wieder dienstspezifisch miteinander zusammen. In der Nutzersicht auf die Dienste sind enge Parallelen zur HCI-Forschung mit Usability, User Experience, der Gestaltung von Benutzeroberflächen und der Mensch-Maschine-Interaktion gegeben. Durch den subjektiven Charakter der Nutzersicht (Psychologie der Leistung) sind nicht alle Effekte und Wirkungen der Einflussfaktoren a priori plausibel abzuschätzen,

was eine empirische Untersuchung erforderlich macht. Dabei ist die Interaktivität ein wichtiger Aspekt des Nutzungsprozesses zur QoE und damit auch zur QoE-Evaluation.

Die gemeinsame Betrachtung von Netzsicht und Nutzersicht auf die Dienste führt zur Bildung dienstspezifischer QoS-QoE-Zusammenhänge. Dabei treffen die Anforderungen der Nutzer an die Dienste auf die Einschränkungen des Netzes zur Realisierung der Dienste. Für die Beschreibung der Zusammenhänge und Rolle einzelner Parameter/Qualitätsmerkmale wurde eine einfache Terminologie vorgestellt. Aus den Überlegungen zur Elastizität zwischen unabhängigen und abhängigen Variablen erklären sich Annahmen zu den unterschiedlichen Charakteristiken des dienstspezifischen QoS-QoE-Zusammenhangs, je nachdem wie Änderungen der QoS-Netzparameter auf die QoE-Bewertungen einwirken – von elastisch bis sprunghaft. Zur weiteren Diskussion wurden Ansätze zur Bildung von Klassen vorgestellt. Es lassen sich jeweils Klassen in den verschiedenen Sichten bilden (Nutzungsklassen und Netzverkehrsklassen). Die Dienstklassen verbinden diese über vergleichbare QoS-QoE-Zusammenhänge. Dazu wurden einige Grundklassen und die dafür relevanten Kriterien der QoS-QoE-Zusammenhänge erläutert. Betrachtungen zum zeitlichen Rahmen von Anforderungen und Einschränkungen nehmen dabei eine zentrale Rolle ein. Ausgehend von diesen Überlegungen wurden die Dienstklassen (Echtzeit, Streaming und Finish Time mit Hintergrunddiensten sowie interaktiven Diensten) nachfolgend detailliert diskutiert. Aus den dazu formulierten Überlegungen kann auf interne Zusammenhänge und die Relevanz der einzelnen QoS-Netzparameter für eine Dienstklasse bzw. einen Dienst geschlussfolgert werden.

4 Die Empirie

QoE-Evaluation, Testplanung und Testrealisierung der Studie

In diesem Kapitel wird die konkrete Realisierung der Studie zur empirischen Ermittlung der Untersuchungsziele erläutert. Ausgehend von einem Exkurs zur QoE-Evaluation werden die Besonderheiten der eigenen Studie vorgestellt. Durch die Konkretisierung der Methode und weiterer Kriterien wie dem Testdesign wird eine solche Studie erst durchführbar. Dazu werden die unabhängigen und abhängigen Variablen systematisiert, operationalisiert und nacheinander vorgestellt. Abschließend werden die relevanten Aspekte der Studiendurchführung dargelegt.

Konzeption, Design und Planung des Tests erfolgten in gemeinsamer Abstimmung aller beteiligten Partner. Die Konkretisierung des Testdesigns und Erstellung der Testpläne wurden in Testrunde 2 durch AAP übernommen, entsprechend finden sich in diesem Kapitel mehrere Hinweise auf diese Anteile.

4.1 QoE-Evaluation

Vor den weiteren Ausführungen zu allgemeinen Punkten des empirischen Teils der Arbeit, wie sie auch zu anderen Zielsetzungen und Fragestellungen passen würden, soll noch auf spezifische Punkte einer QoE-Evaluation eingegangen werden. Neben der Klärung allgemein üblicher Punkte zu Nutzerstudien, was Untersuchungsort und Methode umfasst, schließt der Abschnitt mit einer kurzen Diskussion zum quantitativen Zusammenhang von QoS und QoE.

4.1.1 Untersuchungsziele

Hauptziel der Studie war die empirische Ermittlung des QoS-QoE-Zusammenhangs für ausgewählte mobile Dienste. Dazu sollte die Wirkung der QoS-Netzparameter auf die Nutzerzufriedenheit ermittelt werden. Die Untersuchungen sollten verschiedene (mobiltypische) Netzsituationen bzw. Parameterkombinationen in typischen Bereichen abdecken. Dabei sollten vermutete Zusammenhänge in Art und Lage im Parameterraum bestätigt oder durch die

Exploration neue Abhängigkeiten und interessante Bereiche entdeckt werden. Dies sollte vor allem im Bereich des vermuteten Übergangs zur Zufriedenheit erfolgen (Zielwertsuche). So sollten die Bereiche der notwendigen QoS-Netzparameter für angestrebte QoE-Schwellen eingegrenzt werden. Aus den Ergebnissen sollte auf die Anforderungen/Erwartungen der Nutzer geschlussfolgert werden, um so zu ermitteln, wie sich die Nutzeransprüche verteilen. In der empirischen Ermittlung wird daher nicht nur die Wirkung der QoS-Netzparameter, sondern auch die Relevanz „echter“ Nebenfaktoren in der Nutzung mobiler Dienste berücksichtigt. Dabei sollte ein möglicher Einfluss bestätigt oder aufgeklärt werden.

4.1.2 Grundsätzliche Methoden und Techniken zur QoE-Evaluation

Die QoE-Evaluation dient zur empirischen Erfassung und Bewertung der Quality of Experience. Die nachfolgende Diskussion dazu bezieht sich weitgehend auf die Zusammenfassung in [64] S. 90ff. Ziel der QoE-Evaluation ist oftmals die Quantifizierung der QoE um Nutzererlebnis und Nutzungsergebnis in (statistisch) interpretierbare Zahlen zu fassen [64] S. 90f.

Die Methoden und Techniken zur QoE-Evaluation lassen sich systematisieren, siehe Abbildung 33. Dabei ist eine weitgehende Übereinstimmung mit der Nutzerforschung, z. B. in der HCI, festzustellen. Die nutzbaren Methoden können grundsätzlich in subjektive und objektive Methoden unterschieden werden. (nach [64] S. 91 zusammengefasst)

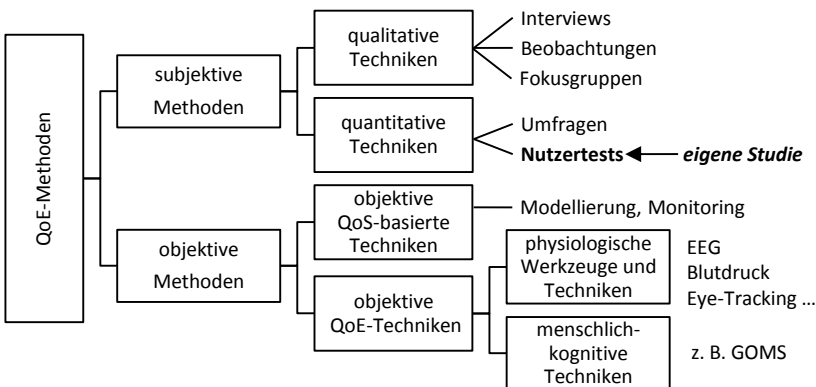


Abbildung 33: Aufschlüsselung zu den Methoden und Techniken zur QoE-Erfassung und QoE-Bewertung in Anlehnung an [64] S. 91, erweitert und ergänzt

Die subjektiven Methoden können weiter in qualitative und quantitative Techniken zerlegt werden. Zu den qualitativen Techniken zählen Beobachtungen (direkt oder per Video), Interviews, Fokusgruppengespräche und Fragebögen mit offenen Fragen. Quantitative Techniken basieren auf Umfragen und Nutzertests mit Fragebögen, die auf geschlossenen Fragen und Bewertungsskalen aufbauen (nach [64] S. 91f zusammengefasst). Die subjektive QoE-Evaluation steht im Mittelpunkt der weiteren Überlegungen und der Ausführungen zur eigenen Studie, wobei in den dazugehörigen nachfolgenden Abschnitten detailliert darauf eingegangen wird.

Die objektiven Methoden basieren entweder auf der objektiven QoS-Erfassung aus der Technik heraus oder der objektiven QoE-Erfassung der Nutzer. Die objektive QoS-Erfassung auf technischer Basis kommt dabei prinzipiell ohne menschlichen Nutzer aus und ist im Grunde schon am Übergang zur QoS-Messung zu sehen, siehe Diskussion zur Quality of Everything. Dabei erfolgt die Erfassung der relevanten Stimuli entweder aus dem Netz bzw. der Datenübertragung heraus oder von den Endgeräten aus (Monitoring). Daher wäre es eigentlich besser und korrekter, dabei von der Evaluation der objektiv direkt QoE-relevanten Größen zu sprechen. Eine weitere Form wäre die Modellierung auf Basis erstellter Zusammenhänge und Einflussgrößen. Dabei kann ein Modell mit Ursache-Wirkungs-Zusammenhang konkreter Parametergrößen weiterhelfen.

Als besondere Form der objektiven Evaluation könnten Expertenevaluationen gesehen werden, bei der Experten eine weitgehend objektive Beurteilung versuchen, z. B. auf Basis fester Heuristiken, Anforderungskatalogen und Bereichsgrenzen zu relevanten Stimuli und ihrer Präsentation. Die Überlegungen dazu können an die Ausführungen zur Quality of Expectation bzw. Expectation of Quality (siehe 4.1.4) und die Erkenntnisse aus der HCI-Forschung anknüpfen. [64] S. 103f verweist mit Blick auf die menschlich-kognitiven Techniken zur QoE-Evaluation auf das GOMS-Modell (Goals, Operators, Methods, Selection Rules). Weitere Erläuterungen zum GOMS-Modell sind zusammengefasst in [159] S. 104ff zu finden. Demnach baut das GOMS-Modell „auf einer eher mechanistischen Sicht der Funktion des menschlichen

Gehirns“ auf [159] S. 104. Für bestimmte Operatoren lassen sich durchschnittliche Zeiten ansetzen, z. B. für Eingaben für Warten etc., die mit gewissen heuristischen Regeln kombiniert werden [159] S. 104.

Die objektiven QoE-Techniken unter Einbeziehung des Menschen integrieren den Nutzer als eine Art zwischengeschalteten Sensor, der selbst ein Signal zur weiteren Messung erzeugt, das sich technisch erfassen lässt. Im einfachsten Fall sind dies Zeitmessungen, z. B. die benötigte Zeit zur Bearbeitung einer genau definierten Aufgabe, oder auch hochkomplexe Messungen am Probanden mit hohem instrumentellen Aufwand. Auf die prinzipiellen Möglichkeiten der objektiven QoE-Techniken wird unter 4.4.6 weiter eingegangen.

Neben dieser hierarchisch strukturierten Beschreibung der Methoden und Techniken bietet sich auch eine andere Strukturierung mit einer etwas einfacher verständlichen Beschreibung nach *Wer Was Wie* an, die prinzipiell auf die entsprechenden Methoden und Techniken abgebildet werden kann.

4.1.3 QoE-Evaluation: Wer, Was, Wie?

Für die Ermittlung der interessierenden Maße in einer QoE-Evaluation gibt es unterschiedliche Herangehensweisen, je nachdem *Wer Was Wie* misst bzw. bewertet.

Wer: Das *Wer* lässt sich grob trennen in:

- klassischer Nutzertest im Labor oder im Feld: Evaluation durch ausgewählte (naive⁸⁶) Nutzer, die im Idealfall repräsentativ sind (Stichprobe aus einer Grundgesamtheit)
- Expertenevaluation: Begutachtung und Bewertung durch ein Expertenpanel (z. B. im Ofcom⁸⁷-Report [158] mit 2 oder 4 Experten)
- Crowdsourcing: meist massenhafte Datenerhebung durch (freiwillige) Nutzer, z. B. übers Internet (siehe z. B. [199] und [200] zur Erläuterung und zu möglichen Problemen)

⁸⁶ im Sinne von „ohne (besondere) Sachkenntnis“ (im Gegensatz zu Experten)

⁸⁷ Organisation zur Regulierung des Telekommunikationsmarktes in UK, vergleichbar mit der Bundesnetzagentur, siehe <http://ofcom.org.uk/> abgerufen am 01.12.2014

Als Gemeinsamkeit dieser Ansätze ist die Einbeziehung des Menschen zu einer Beurteilung der QoE zu erkennen. Eine QoE-Erfassung ohne die Einbeziehung des Menschen ist im Grunde eher ein Monitoring der QoE-relevanten Einflussgrößen (z. B. der Nutzerstimuli der Dienste), was eng mit dem dritten Punkt der nachfolgenden Auflistung zusammenhängt. Dabei wird die Erfassung aber im Sinn der Qualität auf die Feststellung der Charakteristik reduziert. Der Bewertungsmaßstab muss daher anderweitig definiert sein.

Die Einbeziehung des Menschen in die eigene Studie wurde als zwingend erachtet. Damit ist die eigene Studie hauptsächlich ein klassischer Nutzertest im Labor. Teile der Voruntersuchungen können als eine Art Mini-Expertenevaluation eingeschätzt werden. Ein Crowdsourcing wurde nicht erwogen, wäre aber auch nicht mit dem Ansatz und der geplanten technischen Realisierung zu vereinen gewesen.

Was: Bei der Erfassung kann im Allgemeinen nach drei Ansätzen unterschieden werden, je nachdem *Was* bewertet oder gemessen werden soll (siehe [82] S. 8 bzw. [201] S. 55f):

- Testing user-perceived QoS: Tests zur subjektiven Bewertung der durch Nutzer wahrgenommenen Dienstgüte (meist bezogen auf die Qualitätswahrnehmung)
- Surveying subjective QoE: subjektive Erfassung (Bewertung) der QoE unter Einziehung weiterer Einflussfaktoren
- Modeling Media Quality: objektive technische Messung und Modellierung zur Vorhersage beispielsweise der Sprachqualität (wobei vorherige QoE-Evaluationen zur Modellierung notwendig oder zumindest sinnvoll sind)

Die Unterscheidung zwischen erstem und zweitem Punkt ist u. U. nicht sofort eingängig. Die Qualitätswahrnehmung (z. B. Sprache) ist mit dem MOS als Maß der Medienqualität (z. B. Audio) in der Telkowitz verwurzelt. Die Qualitätswahrnehmung ist dabei aber nur ein Aspekt der Gesamtbetrachtungen zur QoE. Die eigene Studie und die zugehörigen Überlegungen sind im zweiten Punkt (subjektive QoE-Erhebung) einzuordnen. Die weiteren Einflussfaktoren zur ganzheitlichen Sicht sind entsprechend dargelegt.

Für das **Wie** ist die Präsentationsform der Stimuli entscheidend. Eine Variante dabei ist, zu einer Zeit jeweils nur eine Ausprägung der Einflussgrößen (Single Stimulus) zu präsentieren und per Absolute Category Rating (ACR) auf einer ACR-Skala bewerten zu lassen. In der eigenen Studie wurde diese Variante gewählt, die eigene Skala dazu wird später zusammen mit der Erfassung der abhängigen Variablen noch diskutiert. In der eigenen QoE-Evaluation werden den Probanden die Stimuli nicht nur passiv präsentiert, sondern die Probanden erzeugen sich innerhalb einer Aufgabenausführung jeweils eine Ausprägung der interessierenden Stimuli aktiv selbst. Unter diesem Gesichtspunkt sind weitere Ansätze zu alternativen Präsentationsformen ohnehin nur schwer vorstellbar. Zur Vollständigkeit sei auf weitere Ansätze wie Degredation Category Rating ([202] S. 22f) und Comparison Category Rating ([202] S. 23f) für den direkten paarweise Vergleich hingewiesen. Dabei ist auch eine relative Bewertung besser/schlechter bezogen auf einen anderen Stimulus bzw. eine Referenzausprägung möglich. Weitere Erläuterungen zur Präsentationsform (am Beispiel der Bewertung visueller Sequenzen) liefert [30] S. 109ff.

4.1.4 QoE: Quality of Experience vs. Quality of Expectation

Mit der Unterscheidung der *Zufriedenheitswaage* in eine Waagschale Experience (erlebte Leistung) und eine Waagschale Expectation (erwartete Leistung) wird verständlich, warum teilweise bei QoE auch von der Quality of Expectation gesprochen wird. [203] S. 11 verweist in diesbezüglich auf die Gleichsetzung von Quality of Experience (QoE) und Expectation of Quality (EoQ) als $QoE = EoQ$.

Aus einer bestimmten QoE-Bewertung lassen sich verschiedene Schlüsse ziehen:

- sind weder erlebte Leistung noch die Erwartung dazu bekannt, sind keine weiteren Aussagen möglich
- bei bekannter Erwartung lässt sich die unbekannte erlebte Leistung abschätzen
- bei bekannter erlebter Leistung lässt sich die unbekannte Erwartung abschätzen

Für die systematische Untersuchung des QoS-QoE-Zusammenhangs muss also entweder die erlebte Leistung oder die dazugehörige Erwartung bekannt sein, um etwas über die andere Seite aussagen zu können. Für die eigene Studie wird der Ansatz der bekannten (und vorab bestimmten) Leistung gewählt, die ein gewisses zugehöriges Nutzererlebnis erwarten lassen. Es wird also auf die konkreten Anforderungen und Erwartungen der Nutzer durch das systematische Austesten gezielter Einschränkungen des Netzes geschlossen. Die Messung des QoS-QoE-Zusammenhangs dient damit eigentlich der Ermittlung der Seite der Erwartung: Wie verteilt sich die Erwartung zu bestimmten Leistungen in der untersuchten Stichprobe von Probanden (bzw. von diesen interferenzstatistisch geschlussfolgert in der Grundgesamtheit der mobilen Nutzer). Dies entspricht den üblichen Ansätzen zur QoE-Messung. Somit ist Quality of Expectation als alternativer Begriff eigentlich gar nicht so schlecht, zumindest bezogen auf die übliche Testmethodik.

Zu den Einflussfaktoren auf die Erwartungsseite lagen vorab keine konkreten Erkenntnisse zu den Probanden vor. Entsprechend sind die Größen im Modell, die auf Erwartungsseite einwirken, recht unsicher zu handhaben. So sind Einflussstärke und Einflussrichtung meist unbekannt. Einige Erkenntnisse zu Erwartungsfaktoren können sich durch die ausgeführten Messungen ergeben. In der Auswertung wäre entsprechend zu prüfen, ob über die Leistungsseite hinaus Aussagen zu reinen Erwartungsfaktoren möglich sind. In der eigenen Auswertung wird auf das Bewertungsverhalten und die Einteilung in Nutzertypen noch einmal im Intra-Rater-Vergleich sowie im Inter-Rater-Vergleich eingegangen, um die eigenen Probanden diesbezüglich zu charakterisieren.

4.1.5 Erfassung der Qualitätsaspekte

In engem Zusammenhang mit der Unterscheidung der QoE-Methoden stehen auch die Maße/Zielgrößen (en. Measures), mit denen die Qualitätsaspekte der QoE erfasst werden können: subjektiv und objektiv [82]. Für ein und dasselbe Messobjekt sind prinzipiell Maße in beiden Kategorien möglich. Die **subjektiven** Maße werden durch eine **Bewertung** durch den Nutzer selbst zugeordnet, sie sind also durch eine Meinung gekennzeichnet und werden explizit abgegeben. Die **objektiven** Maße ergeben sich aus der **Messung** objektiv erfassbarer Größen. Sie sind nicht von der Meinung des Nutzers abhängig und ergeben sich implizit. Die objektiven Maße beschreiben dabei so etwas wie

ein Nutzungsergebnis bzw. eine Art Nutzerleistung (User Performance), die subjektiven Maße hingegen das bewertete Nutzererlebnis [82]. Die Autoren von [82] plädieren außerordentlich dafür, objektive Maße zu nutzen und auch die subjektiven Maße quantitativ zu erfassen.

Eine subjektive Qualitätsbewertung eines Nutzers kann als relative Qualitätsbewertung bezogen auf sein eigenes Ideal interpretiert werden. Den individuellen Maßstab zu dieser Gütebewertung bringt jeder Nutzer schließlich selbst mit. Für objektive Maße kann von relativen und absoluten Bewertungen ausgegangen werden. Dafür muss ein sinnvoller (einheitlicher) externer Maßstab zur Beurteilung aber erst definiert werden, z. B. verschiedene Bereiche für die benötigte Zeit zur Aufgabebearbeitung etc. Ein Ideal für eine absolute Qualität kann beispielsweise „<1 s“ oder „0 unnötige Klicks“ sein, was *sehr gut* bewertet werden würde.

4.1.6 MOS, MOS-Skala und Wahrnehmungsqualität

Der Mean Opinion Score (MOS) bzw. die MOS-Skala wird als Bewertungsskala für unterschiedliche QoE-Maße eingesetzt. Ursprünglich bezieht sich der MOS auf die subjektive Bewertung der Wahrnehmungsqualität, z. B. bei der Audioqualität in Form der Sprachverständlichkeit, siehe [202]. Besonders oft findet die 5-Punkte-MOS-Skala Verwendung, siehe Tabelle 17. Die ITU-T P.800 geht von einer ACR-Skala⁸⁸ aus, siehe [202] S. 18 und S. 11. Dabei bekommen die Probanden einen einzelnen Stimulus präsentiert und wählen die Kategorie der Skala als Bewertung aus, die ihrer Meinung nach am ehesten zutrifft. Auf die damit verbundenen Implikationen wird noch gesondert eingegangen. Zunächst ist die MOS-Skala aber eine recht gute Orientierung dazu, was im Allgemeinen als QoE-Bewertung verstanden wird. Auf die MOS-Skala wird in der Festlegung der eigenen Skala noch einmal eingegangen (4.3.1), der MOS als gebräuchliche QoE-Metrik zur Darstellung der erzielten Ergebnisse wird in der Auswertung weiter thematisiert (8.6.9).

Den „einen“ MOS gibt es eigentlich nicht. Der MOS wird teilweise mit Indexwert genutzt, um einen speziellen Anwendungsfall oder eine abweichende Benutzung zu symbolisieren. So steht MOS_p für einen MOS auf Basis aus einer

⁸⁸ ACR: absolute category rating

Vielzahl Einzelbewertungen zur Nutzerwahrnehmung (perceived), wie zuvor beschrieben, oder als MOS_{CQE} (Communication Quality Estimated) auf Basis von Modellrechnungen, was in die dritte Kategorie von QoE-Evaluationen fällt (Modeling Media Quality). Aus dem synthetischen E-Modell [205] zur Berechnung der Nutzerzufriedenheit im Bereich QoE-Bewertungen der Sprachverständlichkeit lassen sich unter Berücksichtigung der relevanten objektiven Einflussgrößen zugeordnete R-Values berechnen, die wiederum auf MOS_{CQE}-Werte umgerechnet werden können, siehe [205] S. 14.

Tabelle 17: 5-Punkte-MOS-Skala

Score	Opinion ACR-Skala	(Übersetzung)	Beeinträchtigung DCR-Skala
5	Excellent	hervorragend	nicht wahrnehmbar
4	Good	gut	wahrnehmbar, aber nicht störend
3	Fair	zufriedenstellend	leicht störend
2	Poor	mangelhaft	störend
1	Bad	schlecht	sehr störend

nach ITU-T P.800 ([202] S. 18) erstellt, ergänzt um gängige Übersetzungen und Beschreibung der Beeinträchtigungen nach 5-Punkte-Impairment-Scale der ITU-R BT.500-13 ([204] S. 18) bei der Single Stimulus Bewertung

4.1.7 Festlegungen zur Testrunde 2 der eigenen Studie

Die hauptsächliche Untersuchung des QoS-QoE-Zusammenhangs kann im Bereich der subjektiven Erhebungen als quantitative Nutzerstudie mit Bewertung durch echte Nutzer eingestuft werden. Unter der qualitativen Erfassung und Erhebung sind in der eigenen Studie die Nutzerkommentare im Fragebogen, Interviews zu den (Pre-) Pre-Tests und die Videoerfassung der Endgeräte (mit Fokus auf die Interaktion) einzuordnen. Aussagen auf Basis der objektiven QoS sind durch die Erfassung und Auswertung der Netzverkehrsdaten und Videodaten (mit Fokus auf die Stimuli) prinzipiell möglich.

Die objektiven QoE-Maße mögen theoretisch überlegen sein, gebräuchlich sind noch immer die subjektiven Maße, z. B. in der Art des Mean Opinion Score, der zahlreiche einzelne Individual Opinion Scores statistisch zusammenfasst. Wie so oft wird die richtige Antwort nicht *entweder oder* sondern *sowohl als auch* sein. Insbesondere wenn die Meinung des Nutzers als Abgleich zum Gefühl, als innere Einstellung oder Erwartung relevant ist, dann

kann dies auch subjektiv ermittelt werden. Wie unter 2.3.9 beschrieben, konzentrieren sich die eigenen Überlegungen in der Theorie auf das Nutzererlebnis, weshalb auch in der Empirie dieser Ansatz weiter verfolgt wurde. Für die Erhebung sollte eine eigene, angepasste Skala zum Einsatz kommen, deren Aufbau in gewissem Umfang auch eine quantitative Auswertung zulässt (mehr dazu in 4.3.1 und 8.6.3).

In der Testrunde 1 wurden neben der Zufriedenheit auch noch die Funktion (bzw. funktionale Qualität) und der (grundlegende) Nutzen erfasst. Tendenziell zeigte sich, dass für Funktion und Nutzen einfacher und früher gute Bewertungen zu erreichen waren als für die Zufriedenheit. Im Rahmen einer Laborstudie sind Aspekte wie der Nutzen schlecht zu ermitteln, ein echtes Bedürfnis zur Nutzung lässt sich kaum erzeugen. Durch eine geschickte Aufgabenstellung lassen sich höchstens die Schritte nachvollziehbar beschreiben, die normalerweise zur Generierung eines Nutzens notwendig sind. Aussagen zur grundsätzlichen Funktion lassen sich auch (und einfacher) ohne Probanden ermitteln, wobei davon auszugehen ist, dass die Erfüllung der angestrebten Funktion eine Voraussetzung für die anderen Bewertungen ist. Auch in Bezug auf die Kosten ist offen, ob diese innerhalb einer künstlichen Situation eines Labortests sinnvoll integriert werden können. Wenn primär Interesse an der Zufriedenheit besteht und die Voraussetzungen für die Erfassung anderer QoE-Maße eingeschränkt sind, dann sollten sich die Fragen an die Nutzer auch auf die Zufriedenheit konzentrieren. Die Beschränkung auf die Erfassung der Zufriedenheit als relevantes QoE-Maß in der Testrunde 2 war daher sinnvoll. Je nach weiterer Notwendigkeit wäre die Akzeptanz ungeachtet dessen, noch eine sinnvolle Ergänzung gewesen.

4.1.8 Untersuchungsort und Methode

Typischerweise erfolgt eine QoE-Evaluation mit Nutzerbeteiligung im Labor, allerdings mit möglicherweise erheblichen Nachteilen bezogen auf die Aussagekraft. [206] plädiert daher für Felduntersuchungen zur QoE, also in der natürlichen Nutzungsumgebung der Nutzer. Dafür wird die Kombination aus quantitativer und qualitativer Erhebung vorgeschlagen, z. B. durch Bewertung des Nutzererlebnisses direkt nach einer Nutzung und das Führen eines Nutzungstagebuchs in Verbindung mit Nutzerinterviews. Für die ganzheitliche Sicht auf die QoE ist dies sicherlich ein interessanter Ansatz, wäre aber auch

eine andere Zielstellung als die Konzentration auf den QoS-QoE-Zusammenhang.

Eine reine Befragung der Probanden zur Nutzerzufriedenheit, losgelöst von der Erfassung des Zusammenhangs zwischen QoS und QoE, war für die eigene Studie grundsätzlich ausgeschlossen. Auch eine reine Beobachtungsstudie, bei der nur eine Erfassung von QoS-Situation und dazugehöriger QoE-Bewertung möglich ist, schied mangels Eingriffsmöglichkeiten aus. Die Studie konnte damit nur als Experiment bzw. als Quasi-Experiment ausgeführt werden, wofür die Vorteile einer Laborumgebung Voraussetzung sind, siehe [207] S. 7:

- weitgehend kontrollierte Bedingungen
- planmäßige und gezielte Eingriffe (Variationen) zur Aufdeckung der gesuchten Abhängigkeiten
- was nicht vollständig planbar bzw. geplant variierbar ist, kann zumindest erfasst („kontrolliert“) werden

In der Laborumgebung erlauben die Möglichkeiten des geschaffenen Testbeds die Durchführung der Studie als Experiment. Laborstudien sind ihrer Anlage nach aber auch mit einigen Nachteilen verbunden, siehe [207] S. 10: Künstlichkeit der Untersuchungsbedingungen und u. U. geringe Relevanz für die Teilnehmer. Dies lässt sich nur bedingt kompensieren, besonders technisch sind diesbezüglich Grenzen gesteckt, wenn es hauptsächlich um psychologische Faktoren geht. Die Vorteile der Kombination aus Laborstudie und Experiment als **Laborexperiment** überwiegen aber die genannten möglichen Nachteile.

Je nach Sichtweise auf die unabhängigen Variablen ist eine Klassifizierung als Quasi-Experiment sinnvoll. Im Gegensatz zum Experiment, das von einer völlig freien Zuweisung der Probanden zu den Faktorstufen im Testplan ausgeht, sind im Quasi-Experiment Einschränkungen bei der Zuweisung vorhanden. Für die Sicht des Quasi-Experiments spricht, dass die Probanden nicht völlig frei allen möglichen Bedingungen zugewiesen wurden. Die Zuweisung von Probanden zu den einzelnen Endgeräten erfolgte auf Basis ihres Attributs der mitgebrachten Vorerfahrungen mit den jeweiligen mobilen Plattformen.

4.1.9 Quantifizierung des Zusammenhangs zwischen QoS und QoE

Eine mögliche Art des QoS-QoE-Zusammenhangs ist an den Dienstkurven bereits am Beispiel eines QoS-Parameters (Datenrate) gezeigt worden, der mit höherer Quantität (QoS) zu besserer Qualität (QoE) führt. Die QoE-Literatur legt nahe, dass sich QoS-QoE-Zusammenhänge nach Art der QoS-Änderung und zugehöriger QoE-Änderung quantitativ klassifizieren und beschreiben lassen [208], siehe Zusammenfassung in Tabelle 18. Prinzipiell sollten sich je nach Parameterwirkung auch andere Formen des quantitativen Zusammenhangs als die bereits zuvor diskutierten logistischen Grundtypen vermuten lassen. Die angegebene Art der Veränderungen der QoS-Parameter oder von ihnen abhängiger interner Variablen bewirken die jeweiligen Veränderungen der QoE-Bewertungen im Bewertungsraum. Über Regressionsmodelle lässt sich für empirisch ermittelte Daten das Maß der Übereinstimmung zwischen angenommenem und tatsächlichem Zusammenhang abgleichen. Dabei ist grundsätzlich zu beachten, dass die Zusammenhänge ggf. nur in bestimmten Abschnitten der Dienstkurve gelten (Näherung). Ein wahrscheinlichkeitstheoretischer Zusammenhang, der die Veränderungen über den Parameterraum mit der statistischen Erklärung der Ergebnisse zusammenbringt, ist plausibel (siehe 8.6.8).

Tabelle 18: Klassifizierung des Zusammenhangs zwischen QoS (Parameterraum) und QoE (Bewertungsraum)

QoS-Veränderung (Ursache)	QoE-Veränderung (Auswirkung)	
	<i>additiv</i>	<i>multiplikativ</i>
<i>additiv</i>	linear: $QoE \sim QoS$	exponentiell: $\log(QoE) \sim QoS$ IQX-Hypothese, siehe [209] und [210] z. B. MOS zur benötigten Zeit
<i>multiplikativ</i>	logarithmisch: $QoE \sim \log(QoS)$ Weber-Fechner-Gesetz, siehe [211] z. B. MOS zur Datenrate (Download)	potenziell: $\log(QoE) \sim \log(QoS)$

zusammengestellt aus [212] S. 6ff und [213] S. 13 sowie den in der Tabelle selbst genannten Quellen

Die Angaben sind jeweils um konkrete Parameter zu ergänzen, ggf. ist das Vorzeichen für die Richtung der Änderung zu beachten.

Grundsätzlich sind in einer empirischen QoE-Evaluation die möglichen Nicht-linearitäten in den Betrachtungen zu den Stimuli, bei der Auswahl der Bereiche und Abstufungen der Parameter sowie in der Diskussion der erzielten Ergebnisse zu beachten. Die Kenntnis der grundlegenden Art des Zusammenhangs zwischen QoS und QoE und zusätzlich der quantitativen Beziehung erlaubt eine Nutzung der Einflussgrößen als Prädiktoren im Qualitätsmodell:

- für gegebene QoS-Werte, QoD-Werte, bzw. QoA-Werte ist eine QoE-Voraussage möglich, z. B. als wahrscheinlicher Mindestwert, typischer oder mittlerer Wert
- für angestrebte QoE-Werte ist eine Abschätzung der notwendigen QoS-Werte möglich

4.2 Unabhängige Variablen und Bedingungsvariation

Für die weiteren Ausführungen ist es notwendig, genau zu spezifizieren, welche Variablen es gibt und welche wie variiert oder wie gemessen werden können. Prinzipiell ist zu unterscheiden zwischen einflussnehmenden Variablen und beeinflussten Variablen.

Die einflussnehmenden Variablen werden auch als Einflussgrößen, Faktoren oder Input-Variablen bezeichnet. Je nach ihrer Art kann unterschieden werden zwischen:

- unabhängige Variablen (UV): aktiv kontrollierbar, Teil der Untersuchungshypothese
- Kovariablen bzw. Kovariate: über ebenfalls einen Einfluss aus, können miterfasst werden, aber keine aktive Variation
- Störvariablen: unkontrollierter bzw. ungeplanter Einfluss, nicht interessierende Einflussgröße
- Kontrollvariablen: Störvariablen, die im Testplan berücksichtigt werden

Die beeinflussten Variablen sind die abhängigen Variablen (AV), die auch als Ziel-/Wirkungsgrößen oder Output-Variablen bezeichnet werden. Diese werden später separat erörtert. Die Einflussgrößen wirken auf diese direkt oder indirekt über interne Variablen ein. In der Diskussion ist zu unterscheiden,

woher die unabhängigen Variablen kommen, bzw. wodurch sie induziert werden. Im Variablenplan sind die Einflussgrößen zusammen mit ihrer Art, Anzahl Ausprägungen im Test und der im Test genutzten Variation aufgeschlüsselt, siehe Tabelle 19. Detaillierte Erläuterungen zu den Unterscheidungen nach Netzparametern, Gerät und Dienst vermitteln die Kapitel 5 bzw. 6.

Tabelle 19: Variablenplan der Einflussgrößen und Identifikationsmerkmale im Test

Variable / Faktor	Anzahl und Art der Ausprägungen sowie Variation im Test	Anmerkung
<i>Testparcours: primär Unterscheidung nach Dienst (QoS-QoE-Zusammenhang)</i>		
Dienst	6 aus 6 Kategorien, within	{D, F, G, M, S, Y}, Anfangsbuchstaben des Dienstszenarios als Kürzel (Drive, Facebook, Google Maps, MTV-Music, Spiegel.de, Youtube)
Preset	2 pro Proband je Dienst, within	dienstspezifische Netzparameterstufe: Dienst x Netzparameterstufe, die Presets sind die aktiv variierte Kombination (Tupel) aus Netzparametern
Netzparameterstufe	5 Stufen für jeden der 6 Dienste	{10, 20, 30, 40, 50}, monoton geordnet, größer = schneller/besser
max. Download	stetig, [80; 7200] (kBit/s)	Darstellung als 3-Tupel aus den QoS-Netzparametern
max. Upload	stetig, [48; 5760] (kBit/s)	
min. Latenz	stetig, [50; 300] (ms)	
App	1 pro Proband aus n pro Dienst, between	unterschiedlich, Abhängigkeit zu Dienst und Gerät
Inhalt	m pro Proband aus n je Dienst (between oder within)	unterschiedlich, Abhängigkeit zu Dienst, App, Gerät, Aufruf und Variante möglich m und n je nach Szenario
<i>Testbed: Unterscheidung nach Gerät mit zugeordneten Attributen (Heterogenität der Technik)</i>		
Gerät	1 aus 5 Kategorien, between	{Tab 2 10.1, Ace 2, S3, iPad 4, iPhone 5}
Plattform	1 aus 2 Kategorien, between	{Android, iOS} als Attribut an das Gerät gebunden
(weitere Attribute)	optional	z. B. Formfaktor, Leistungsklasse, ... (an Gerät gebunden, ohne weitere Betrachtung)

Fortsetzung der Tabelle auf der nächsten Seite

Fortsetzung der Tabelle 19

Variable / Faktor	Anzahl und Art der Ausprägungen sowie Variation im Test	Anmerkung
<i>Testdesign: Unterscheidung bedingt durch die dienstspezifische Aufgabenwiederholung je Proband (Wirkung über den Inhalt)</i>		
Aufruf	2 aus 2 Stufen, within	{1, 2}, erster oder zweiter Aufruf innerhalb eines Testlaufs (Reihenfolge)
Variante	2 aus 2 Kategorien, within	{a, b}, Unterscheidung der beiden Aufgabenausführung in einem Testlauf
<i>Testorganisation: IDs zur Zuordnung und weitere nichttechnische Faktoren</i>		
(Testrunde)	bisher 2 Testrunden	nur nähere Betrachtung der Testrunde 2
Testblock	2 IDs (fortlaufend)	{1, 2} (in der Testrunde 2)
(Zeitraum)	optional	Unterscheidung großer zeitlicher Abstände (Monat)
Testslot	42 IDs (fortlaufend)	fortlaufende Nummerierung
(Zeitstempel)	optional	Wann hat der Testslot stattgefunden (Datum, Zeit)?
(Betreuung)	optional	Wer hat den Test vorbereitet und betreut?
Testplatz	1 aus 6 IDs	6 Testplätze mit 5 Gerätetypen, Besetzung nach Vorkenntnissen
Testlauf	1 aus n IDs (fortlaufend) je Testplatz	12 Aufgabenausführungen verbunden über einen Probanden auf einem Testplatz in einem Fragebogen, geplante Zuweisung von Presets zu den Aufgaben aus Testplan
Fragebogen	1 aus 6 IDs je Testlauf	Abfolge der Dienste und Varianten, zyklische Wiederholung der Fragebögen auf den Testplätzen
Aufgabennummer	12 aus 12 IDs je Fragebogen	Dienst, Variante und Aufruf fortlaufend durch Fragebogen festgelegt
Proband	fortlaufende IDs	Nutzung zur Pseudonymisierung, die Darstellung der Erhebung der Organismusvariablen (Attribute) der Probanden erfolgt gesondert
<i>Störvariablen: Ausmittlung oder Kontrolle ex post und Validierung</i>		
	intern bedingt	Abweichungen bei den Aufgabenausführungen durch die Probanden
	extern bedingt	Probleme mit externem Netz oder Gegenseite, Inhaltsänderungen

Die Art bezieht sich auf das Skalenniveau:

Kategorien=Nominalskala

Stufen=Ordinalskala

stetig=metrische Skala (min. Intervallskala)

Angaben zu den Variablen in Klammern: ohne weitere Betrachtung

4.2.1 Durch das QoS-QoE-Modell induzierte Einflussvariablen

Die Qualitätselemente des QoS-QoE-Modells müssen sich in den einflussnehmenden Variablen widerspiegeln und waren Ausgangspunkt der Überlegungen zur Manifestation. Dabei ist eine grundlegende Überlegung, ob die QoS-QoE-Qualitätselemente unabhängige Variablen oder Störvariablen sind. Sofern sich die konkreten Realisierungen der einzelnen Elemente aus gezielten Vorgaben ergeben, sind sie unabhängige Variablen. Ergeben sie sich aus einem zufälligen äußeren Einfluss, sind sie eigentlich Störvariablen. Für einige der Größen ist keine eindeutige Abgrenzung möglich. Sie ergeben sich aus dem Zusammenspiel interner wie externer Größen, z. B. die Netzgesamtkarakteristik und der Inhalt (sofern nicht explizit fixiert). Für die Gegenseite ist nur sehr begrenzt eine aktive Vorgabe möglich, entsprechend ist von einer potenziellen Störvariable auszugehen.

Die QoS-Netzparameter sind entsprechend des QoS-QoE-Modells die primär interessierenden Einflussgrößen für die verschiedenen Dienste. Neben den Diensten und Netzparametern als Hauptfaktoren ergeben sich die weiteren Einflussvariablen aus den weiteren Elementen des Modells zum QoS-QoE-Zusammenhang und den Voruntersuchungen dazu. Dabei sind die Facetten der Testeignung zu beachten, z. B. die Heterogenität des Testbeds. Diese weiteren Faktoren sind nur bedingt wirklich voneinander unabhängig, d. h. Ausprägung und Wirkung der einen Variable hängt mit der Ausprägung einer anderen Variable zusammenhängen. Ein Beispiel: Eine App für einen bestimmten Dienst, die für Android programmiert ist, kann nur auf den Geräten im Test eingesetzt werden, die Android nutzen. Ein bestimmter Inhalt wiederum kann an eine bestimmte App gebunden sein etc. Durch die Nebenfaktoren können Gruppen bzw. Partitionen in unterschiedlicher Anzahl je Dienst definiert sein.

4.2.2 Durch das Testdesign induzierte Variablen und Besonderheiten

Die unterschiedlichen Varianten einer Aufgabe stehen in engem Zusammenhang mit den Aufgabenwiederholungen zu einem Dienst bzw. sind notwendig, um diese überhaupt sinnvoll testen zu können und primär technisch bedingt. Die Hintergründe, wie z. B. die Abhängigkeit vom Inhalt (Caches), und die erwartenden Auswirkungen sind spezifisch je Dienst und in den Betrachtungen zur Einbeziehung der Dienste in die Studie detailliert erläutert (6.3.5). Es ergibt sich eine Unterscheidung je Dienst in ersten und zweiten Aufruf. Die

Variablen Variante und Aufruf sind damit durch die Art des Testdesigns induziert. Sie sind in der Auswertung zu berücksichtigen, z. B. im Intra-Rater-Vergleich, bei der Aufdeckung der Bewertungsstrategien der Probanden in der Auswertung der Nutzerkommentare und bei den Besonderheiten für die Auswertung der bedingten Verteilungen.

4.2.3 Bedingungen

Die Bedingungen beschreiben die Gesamtheit und Anzahl aller zusammen einflussnehmenden Größen. In einem Experiment sollten die aktiv variierten, unabhängigen Variablen und zumindest potenziell relevante Störvariablen berücksichtigt werden.

Die maximale Zahl anzunehmender Bedingungen ergibt sich aus dem Produkt aller Faktorstufen aller betrachteten Faktoren. Die Bedingungsanzahl kann daher bei ungeschickter Planung sehr hoch ausfallen, was schnell zu praktisch nicht mehr handhabbar hohem Testumfang für statistisch sinnvolle Aussagen führen würde (Dienst x Netzparameterstufe x Reihenfolge x Variante x App x Endgerät x ...). Aus den bisherigen Ausführungen sollte klargeworden sein, dass es zwar eine hohe Anzahl Faktoren und Stufen gibt, diese sich aber nicht in allen Kombinationen im Test manifestieren können. Die Eingrenzung der Anzahl Bedingungen durch dienstspezifische Überprüfungen ist daher essenziell. Es ist ausreichend, nur die Faktoren in die Betrachtungen einfließen zu lassen, für die sich die Bedingungen relevant unterscheiden. Dies unterstreicht die hohe Bedeutung intensiver Voruntersuchungen nochmals. In den meisten Fällen ist so beispielsweise gegeben, dass sich die 5 Gerätetypen auf die Unterscheidung von 2 Plattformen reduzieren lassen.

Die Grundsätze von *ceteris paribus*, d. h. „bei gleichen sonstigen [Umständen]“, oder „One-Factor-at-a-Time“ (siehe [26] S. 2ff), bei dem sich eine Variable ändert, während die anderen konstant gehalten werden, sind jeweils zu prüfen. Je nach Überlegung ist dieser Ansatz so zu erweitern, dass sich aus der gemeinsamen Änderung einzelner Variablen jeweils voneinander abzugrenzende Bedingungen ergeben, die meist durch die Abhängigkeiten zwischen den Variablen bestimmt werden.

Mit Blick auf den Testumfang ist auch verständlich, warum die Netzparameter in nur 5 Stufen und als Wertkombination in Form dienstspezifischer Vorgabewerte, der Presets, variiert werden (vertiefende Diskussion dazu in 6.4). Das Preset ist damit die primäre Bedingungsvariable, die planmäßig und kontrolliert variiert wird. Durch die Variation als Parameterkombination sind allerdings gesicherten Einzelaussagen zu den einzelnen Parametern nicht direkt möglich. Es lassen sich aber sachlich/fachlich begründete Aussagen auf Basis der Zusammenhänge und Abhängigkeiten im QoS-QoE-Modell ableiten, die die Relevanz der einzelnen QoS-Parameter berücksichtigen (z. B. die Download-Datenrate für die Download-Zeit). Die Presets sind prinzipiell ordinal, d. h. durch die Netzparameterstufe in eine natürliche Ordnung zu bringen (größer = besser). Als n -Tupel betrachtet sind die Preset eigentlich sogar n -dimensional metrisch. Dabei sind die Presets selbst diskret, ihre Komponenten aber Werte aus einem stetigen Parameterraum. Entsprechend sind zumindest Interpolationen theoretisch möglich. Zur eindeutigen Benennung der Bedingungen sind somit zumindest Dienst und Netzparameterstufe als Preset und ggf. dienstabhängig weitere Angaben zu den jeweils relevanten Nebenfaktoren notwendig.

Es wird für die Bedingungen angenommen, dass die relevanten Parameter für die Darbietungsdauer einer Bedingung als konstant anzusehen sind (Konstanzmethode, siehe [30] S. 33 zur Unterscheidung). Als Darbietungsdauer wird jeweils die gesamte Zeit der Bearbeitung einer Aufgabe angesetzt. Die geschaffenen technischen Möglichkeiten des Testbeds stellen sicher, dass jeweils die Parametrisierung zwischen den Aufgabenbearbeitungen korrekt und verzögerungsfrei gewechselt wird. Damit sind die sich ergebenden Stimuli vorab (ex ante) als Erwartungswert abschätzbar, was den Weg zur parameterbasierten Auswertung öffnet.

4.2.4 Art der Zuweisung bzw. Variation der Variablen

Für die Betrachtungen zur Variation der Variablen ist von grundsätzlicher Bedeutung, für welche der Variablen in der Studie eine intrapersonelle oder interpersonelle Variation genutzt werden sollte. Bei einer Variation als **Within-Faktor** (Innersubjekt bzw. intrapersonell) werden durch alle Subjekte, hier die Probanden, alle (oder zumindest mehrere) Ausprägungen eines Merkmals ge-

testet. Bei einer Variation als **Between**-Faktor (Zwischensubjekt bzw. interpersonell) werden die Ausprägungen zwischen den Subjekten aufgeteilt und nur jeweils eine Ausprägung eines Merkmals pro Proband getestet (Terminologie siehe [207] S. 67).

Für den Test der Dienste und Netzparameterstufen ergaben sich einige gesonderte Überlegungen. Die ersten Ansätze der beteiligten Partner zur Testrunde 1 präferierten eine intrapersonelle Variation für die QoS-Netzparameterstufen, also das Testen vieler (konkret aller) Stufen von technischen Parametern in einer entsprechend hohen Anzahl Wiederholungen durch eine Person. Theoretisch ist diese intrapersonelle Variation vorteilhaft, da so pro Proband komplette Datensätze für alle Ausprägungen eines Merkmals entstehen (hier die technischen Netzparameterkombinationen in Form der Presets). Im Pretest der Testrunde 1 wurde entsprechend dieser Vorüberlegungen die intrapersonelle Variation eingesetzt. So hatten die Probanden nur Aufgaben zu 2 oder 3 verschiedenen Diensten zu bearbeiten, jeweils abwechselnd, diese allerdings in bis zu 7 Wiederholungen für alle jeweiligen Stufen. Dieser Ansatz führte durch die Monotonie leider zu einer zu schnellen Ermüdung der Probanden bis hin offenen Unmutsbekundungen. Für die Testrunde 1 wurde entsprechend des Feedbacks eine Änderung im Testverfahren zur interpersonellen Variation umgesetzt. Statt für eine Auswahl von Diensten alle Stufen zu testen, wurden alle Dienste mit jeweils einer gewissen Auswahl von Stufen getestet. Ein Ansatz völlig ohne Messwiederholung, also mit nur einer Stufe pro Dienst schied aus, da dann wesentlich mehr Probanden benötigt worden wären. Die mehrfache Aufgabenbearbeitung eines Dienstes mit verschiedenen, aber nicht allen Netzparameterstufen ist damit ein Kompromiss zur Vergrößerung der Datenbasis gegenüber einer einfachen Bearbeitung und ohne die Probanden übermäßig zu ermüden oder zu langweilen. Diese Veränderungen gegenüber dem Pretest erzielten den gewünschten Erfolg hinsichtlich der Motivation der Probanden. Der erprobte und bewährte Ansatz wurde in Testrunde 2 grundsätzlich übernommen und im Kreis der Partner weiterentwickelt. Es liegt somit ein Mischdesign vor, das die Variation der unabhängigen Variablen sowohl Within-Subject als auch Between-Subject vorsieht. Die Art der Variation je Variable ist dem Variablenplan in Tabelle 19 zu entnehmen.

4.2.5 Rahmenbedingungen

Bei der Umsetzung und Ausführung der Studie waren einige Rahmenbedingungen einzuhalten. Bedingt durch die Vorgaben des Testbeds und des Testraums erfolgte eine Begrenzung auf sechs Testplätze, was damit auch der Maximalzahl von sechs Probanden gleichzeitig in einem Testslot entsprach.

Als Grenze zur Ermüdung der Probanden wurde in Testrunde 1 eine Dauer eines Tests von 60 Minuten festgesetzt. Innerhalb dieser Zeit sollten möglichst viele Daten in Form von Aufgabenausführungen gesammelt werden. Zusammen mit den darauf abgestimmten Szenarien des Testparcours von circa 5 Minuten pro Aufgabe ergab sich eine Zahl von 12 Aufgabenausführungen pro Proband. [214] empfiehlt im Zusammenhang mit QoE-Evaluationen (subjektive Qualitätsbewertungen) eine Testdauer von 90 Minuten nicht zu überschreiten, um eine Verzerrung der Bewertungen durch die Ermüdung der Probanden zu vermeiden. Bei dem angesetzten Zeitrahmen von geplanten 60 Minuten war daher anzunehmen, dass auch langsame Probanden unter dieser 90-Minuten-Grenze bleiben sollten.

4.2.6 Testplanung und Testplan

Für Testplanung und Testplan zeichnet der Partner AAP auf Basis gemeinsamer Absprachen verantwortlich, der Autor brachte dabei die Erfahrungen aus der Testrunde 1 ein.

Der Testplan ist ein wichtiges und zentrales Element für die Qualität eines Testdesigns. In einem Testplan werden die Faktorstufen (interessierende Ausprägungen eines Merkmals) einem Faktor (einer interessierenden unabhängigen Variable) systematisch zugewiesen. Die Kombination der unterschiedlichen technischen Parameter und die Abbildung auf verschiedene Aufgaben führen schnell zu einer großen Anzahl Möglichkeiten. Diese Komplexität lässt sich nur beherrschen, wenn die experimentelle Überprüfung planvoll erfolgt. Deshalb werden aus den manipulierbaren technischen Parametern und den ausgewählten Applikationen sinnvolle Testpläne erstellt. Durch diese vorherige Planung sollte die systematische und ausgewogene Variation aller relevanten Parameter gewährleistet sein.

Im Testplan wurden Randomisierung und Shuffling von Aufgabenblöcken genutzt. Dadurch ergaben sich in der Testplanung folgende Festlegungen:

- Auswahl/Zusammenstellung der Aufgaben: durch Testdesign fixiert (6 mal 2)
- Durchmischung/Reihenfolge der Aufgaben (3 Blöcke in 6 verschiedenen Anordnungen)
- Variation des technischen Parameters Preset in Testreihen (2 aus 5)

4.2.7 Randomisierung

Die Randomisierung ist ein Versuch, mögliche Störfaktoren statistisch auszuschalten. Dafür erfolgt die Zuweisung der Versuchspersonen zu den Versuchsbedingungen zufällig. Im Testplan wird dies durch die Reihenfolge von Aufgabe, Variante, Aufruf und die Auswahl der Presets berücksichtigt. Die Randomisierung erfolgte durch den Partner AAP.

Eine vollständige Randomisierung als Berücksichtigung aller möglichen Kombinationen der Ausprägungen der testrelevanten Variablen wurde ausgeschlossen, da im genutzten Fragebogenwerkzeug nicht wie gewünscht automatisiert realisierbar. Die verbleibende Möglichkeit einer manuellen Erstellung individueller Fragebögen wurde in Testrunde 1 genutzt. Allerdings ist diese Art der Erstellung für eine große Anzahl Probanden zu aufwendig und entfiel deshalb in der Testrunde 2. Stattdessen wurde eine (Pseudo-) Teil-Randomisierung zusammen mit einer Blockbildung als Kompromiss gewählt. Lediglich die Zuweisung der Ausprägung der Netzparameterstufen erfolgte dabei manuell. Die Testpläne sind durch die manuelle Randomisierung nicht ganz balanciert, d. h. nicht alle Varianten und Reihenfolgen je Presets bzw. Preset-Übergänge treten gleichhäufig auf.

4.2.8 Shuffling von Aufgabenblöcken

Der Entwurf des Partners AAP sah statt einer vollständigen Reihenfolgenvariation einzelner Aufgaben eine Zusammenfassung in Blöcken gleicher Länge und deren Permutation vor. Jeweils vier Aufgaben werden dafür in einem von drei Blöcken zusammengefasst. Ausgehend von einem ersten Entwurf erfolgte in interner Abstimmung eine Modifikation. Dabei wird ein direktes Aufeinanderfolgen der beiden Varianten einer Aufgabe vermieden, damit die Proban-

den die Aufgaben als in sich geschlossene Einheiten wahrnehmen und bewerten. Die Variation der Aufgabenreihenfolge (Nr.) in den Fragebögen ergibt sich aus der Anordnung der Blöcke: $3! = 6$ Anordnungsvarianten. Darauf abgestimmt wurden nur sechs Fragebögen mit entsprechender Fragenreihenfolge benötigt, was zu einer erheblichen Vereinfachung der Testdurchführung im Vergleich zur Testrunde 1 führte.

Die genaue Zusammenstellung der Blöcke mit Abfolge der Aufgaben und ihrer Varianten in den Fragebögen ist der Tabelle 20 zu entnehmen. Durch dieses Shuffling sind bestimmte Kombinationen von Dienst, Variante und Aufruf nur an festen Positionen im Testablauf (Aufgabennummer) zu finden. Für die Reihenfolge der Aufgabenzusammenstellung (Aufgabennummer) wird angenommen, dass sie keinen grundsätzlichen und unmittelbaren technischen Einfluss ausübt. Aus nichttechnischer Sicht ist ein Einfluss möglich, siehe psychologische Effekte der Probanden (7.1.1). Falls vorhanden, sollte ein möglicher Effekt durch das Block-Shuffling weitgehend ausgemittelt werden.

Tabelle 20: Blockbildung in der Zusammenstellung der Aufgabenabfolgen und ihrer Varianten

Aufgaben-Nr.	Fragebogen																							
	1				2				3				4				5				6			
	Block	Dienst	Variante	Aufruf	Block	Dienst	Variante	Aufruf	Block	Dienst	Variante	Aufruf	Block	Dienst	Variante	Aufruf	Block	Dienst	Variante	Aufruf	Block	Dienst	Variante	Aufruf
1	I	Y	a	1	I	Y	a	1	II	G	a	1	II	G	a	1	III	S	b	1	III	S	b	1
2	I	G	b	1	I	G	b	1	II	M	b	1	II	M	b	1	III	Y	b	1	III	Y	b	1
3	I	D	b	1	I	D	b	1	II	S	a	1	II	S	a	1	III	F	b	1	III	F	b	1
4	I	M	a	1	I	M	a	1	II	F	a	1	II	F	a	1	III	D	a	1	III	D	a	1
5	II	G	a	2	III	S	b	1	I	Y	a	1	III	S	b	2	I	Y	a	2	II	G	a	1
6	II	M	b	2	III	Y	b	2	I	G	b	2	III	Y	b	1	I	G	b	1	II	M	b	1
7	II	S	a	1	III	F	b	1	I	D	b	1	III	F	b	2	I	D	b	2	II	S	a	2
8	II	F	a	1	III	D	a	2	I	M	a	2	III	D	a	1	I	M	a	1	II	F	a	2
9	III	S	b	2	II	G	a	2	III	S	b	2	I	Y	a	2	II	G	a	2	I	Y	a	2
10	III	Y	b	2	II	M	b	2	III	Y	b	2	I	G	b	2	II	M	b	2	I	G	b	2
11	III	F	b	2	II	S	a	2	III	F	b	2	I	D	b	2	II	S	a	2	I	D	b	2
12	III	D	a	2	II	F	a	2	III	D	a	2	I	M	a	2	II	F	a	2	I	M	a	2

Numerrierung der gebildeten Aufgabenblöcke mit römischen Ziffern {I, II, III}
 Dienstkürzel nach Anfangsbuchstaben der getesteten Dienste {Facebook, Drive, Google Maps, MTV-Music, Spiegel.de, Youtube}
 nach dem Entwurf von AAP, mit Vermeidung der direkten Wiederholung eines Dienstes

4.2.9 Testläufe und Testreihen

Die sechs Fragebögen wurden nacheinander wiederholt auf allen Testplätzen des Testbeds eingesetzt. Die so erzeugten Abfolgen der Fragebögen und damit der Aufgaben, Varianten und Reihenfolgen dienten als Basis für die Testpläne, in denen nachfolgend lediglich nur noch die Zuordnung der Netzparameterstufen in Form der Presets variiert werden musste.

Für die technische Realisierung und Datenhaltung wurde eine Datenstruktur entwickelt:

- einfach genug, um leicht manuell gehandhabt zu werden
- mächtig genug, um sie im Testsystem automatisiert zu verarbeiten, inklusive der notwendigen Zuordnungsinformationen für die zielgenaue Verknüpfung von Testbed und Fragebogen-Tool (siehe 5.6.6)
- flexibel genug, um zukünftige Erweiterungen zu unterstützen, z. B. zur Variation weiterer Variablen neben dem Preset (optional) und der Berücksichtigung von Testplatz oder Endgerät für Sub-Presets etc.

Innerhalb des Testbeds werden diese Datenstrukturen an verschiedenen Stellen genutzt und sind eine wichtige Voraussetzung für die weitgehende Automatisierung der Tests und damit der Skalierung des Testumfangs:

- Testmanagement: Testlaufverwaltung, Zuordnen und Laden der Fragebögen
- Testlaufsteuerung: Verknüpfung von Fragebogen und Netzparametrisierung
- Testlaufdokumentation: Zuordnung zu (versteckten technisch genutzten) Fragebogenfeldern zur späteren Auswertung des Ergebnisdatensatzes

4.2.10 Testprozedere und Testzuweisung

Zusammen mit den Überlegungen zur Art der Variation der einzelnen Faktoren ergaben sich das Testprozedere und die Zuweisung der Probanden zu den Bedingungen damit wie folgt:

- Ein Proband wird nach Vorkenntnissen einem der 6 Testplätze zugewiesen. Damit erfolgt die Zuweisung zu den Between-Faktoren im Test.
- 12 Aufgabenausführungen sind in einem Testlauf über einen Probanden verbunden.

- Es gibt 6 Dienste/Szenarien als Aufgaben, die jeder Proband zweifach ausführt.
- Die Ausführungen je Dienst unterscheiden sich ggf. jeweils leicht durch Variante und Aufrufreihenfolge.
- Die Aufgabenreihenfolge wird pseudo-zufällig durch die Reihenfolge der Testplatzbelegung und zyklische Wiederholung der Fragebögen (1 aus 6) vorgegeben.
- Je Aufgabenausführung wird eines aus 5 dienstspezifischen Presets als Netzsituation quasi-zufällig per Testplan zugewiesen. Die Wahl des Presets für einen zweiten Aufruf ist von der ersten Bearbeitung unabhängig (dies entspricht Ziehen mit Zurücklegen).

4.2.11 Testorganisation und Identifikationsattribute

Aus dem Testprozedere ergeben sich auf natürliche Art (Abzählen bzw. zugewiesene Nummerierung) die Identifikationsattribute der einzelnen Testelemente für die Testorganisation. Von einer gesamten Testrunde mit mehreren Testblöcken, welche wiederum eine Vielzahl Testslots zusammenfassen, über die Besetzung der Testplätze mit Probanden und ausgeführten Testläufen bis hin zur einzelnen Aufgabenausführung kann so alles eindeutig adressiert und zugeordnet werden, siehe Tabelle 19. Eine sinnvolle Systematik und Kombination der einzelnen Werte vereinfacht das Management der erfassten Daten und erlaubt die unkomplizierte Nutzung in der Auswertung. Neben den im Fragebogen erfassten Ergebnissen können über die gleichen Identifikationsattribute auch die zusätzlich erhobenen Daten mit gewünschter bzw. erforderlicher Granularität zugeordnet werden. Da einige Identifikationsattribute fest mit bestimmten Ausprägungen interessierender Einflussgrößen verbunden sind (z. B. Testplatz mit Gerät und somit Formfaktor und Betriebssystem) können die Attribute auch äquivalent in der Auswertung eingesetzt werden.

4.3 Abhängige Variablen und Ergebniserfassung

In diesem Abschnitt wird dargelegt, welche abhängigen Variablen (AV), also beeinflusste Ziel- bzw. Wirkgrößen als Messgrößen zum QoS-QoE-Zusammenhang erfasst wurden. Zunächst sollen die direkt von den Probanden erzeugten Ergebnisdaten diskutiert werden, bevor nachfolgend weitere mögliche

Ergebnisdaten aufgeführt werden. Tabelle 21 fasst die erhobenen Zielgrößen und Ergebnisraumdaten zusammen.

Tabelle 21: Variablenplan der Zielgrößen und Ergebnisraumdaten

Element	Art	Erfassung über	Anmerkung
<i>subjektive Ergebnisse durch Bewertung und Selbstauskunft der Probanden</i>			
Zufriedenheits- bewertungen	quantitativ (qualitativ)	Skala	primäre AV (Skalenniveau je nach Interpretation)
2 oder 3 Teilbewertungen			dienstspezifische Items, auf relevante Stimuli je Dienst be- zogen
Gesamtzufriedenheit			wichtigste AV
Bearbeitungserfolg	qualitativ	Multiple-Choice-Ka- tegorien	sekundäre AV
Problemangabe	qualitativ	Multiple-Choice-Ka- tegorien oder sonstiger Grund (Freitext)	
Nutzerkommentare allgemein	qualitativ optional	Freitext	sekundäre AV
Abspielverhalten	verbindlich		nur Aufgaben MTV- Music und Youtube
<i>weitere erfasste Daten</i>			
Testlogbucheintrag	strukturiert qualitativ	intellektuell durch Testleitung im Test- logbuch	Erfassung besonde- rer Vorkommnisse hoher Aufwand zur Auswertung
Bearbeitungszeiten	objektiv, quantitativ	Protokollierung im Fragebogen und Logdateien	Auswertung nur ein- geschränkt sinnvoll
Parametrisierung	objektiv	Protokollierung im Fragebogen und Logdateien	gemeinsame Daten- haltung von UV und AV zur Auswertung
Netzwerk- Monitoring	Netzverkehrsdaten	Testbed	Erfassung als pcap- Datei sehr hoher Aufwand zur Auswertung
Endgeräte- Monitoring	Videodaten	Testbed	Erfassung als Video- datei sehr hoher Aufwand zur Auswertung

4.3.1 Nutzerbewertungen

Die durch die Probanden selbst erzeugten Ergebnisdaten wurden als subjektive Angaben durch Abfrage und Selbstauskunft (Bewertung) über ein elektronisches Web-Fragebogen-System erfasst. Der Fragebogen entstand in Kooperation der Projektpartner unter Federführung des Partners AAP. An dieser Stelle sollen die Intention und der Ansatz zur Erfassung einzelner abgefragter Daten im Vordergrund der Diskussion stehen. Die kompletten Formulierungen und zugehörige Darstellungsbeispiele zum Fragebogen sind im Anhang A.4 zu finden.

Es wurde die **Zufriedenheit** als relevantes QoE-Maß ausgewählt. Sie ist die eigentliche primäre Zielgröße zur Untersuchung der Zusammenhänge zur QoS-Netzparametrisierung und den einwirkenden Nebenfaktoren.

Bewertungen als Messungen: Die Zufriedenheit sollte stellvertretend für ein relevantes Maß der QoE als abhängige Variable gemessen werden. Bei Messungen wird grundsätzlich unterschieden zwischen (nach [215] S. 285):

- Messungen direkt: unmittelbare Ermittlung der interessierenden Eigenschaft am Messobjekt
- Messungen indirekt: Die interessierende Eigenschaft kann nicht direkt ermittelt werden, also muss von anderen (direkt) messbaren Eigenschaften auf diese geschlussfolgert werden (bei bekanntem Zusammenhang). Dieser Zusammenhang wird normalerweise durch Modelle ausgedrückt. „Aussagen, die durch indirekte Messungen gewonnen werden, sind zunächst nur im Rahmen der Modellannahmen gültig. Die Qualität der Rückübertragung bzw. Interpretation solcher Aussagen aus der formalen Modellwelt in die Realität hängt von der Qualität des Modells ab.“ ([215] S. 285)

Die interessierende Eigenschaft ist die Zufriedenheit. Diese lässt sich nicht direkt von außen messen. Es sind lediglich subjektzentrierte Schätzurteile der Probanden als Bewertungen möglich (Terminologie siehe [216] S. 144). Ihr Zustandekommen wird im QoS-QoE-Modell beschrieben.

Der eigentlich gewünschten, aber nicht direkt möglichen QoE-Messung stehen also QoE-Bewertungen der Probanden gegenüber. Entsprechend ist sinnvoll kurz zu überlegen, ob dies adäquat ist. Bewerten bedeutet, Merkmalen (nach bestimmten Regeln) graduierte Werte zuzuordnen. Die Probanden sind dabei die Bewerter ihres eigenen Persönlichkeitsmerkmals Zufriedenheit. Mit der graduierten Erfassung beschäftigt sich nachfolgend die Skala. Lassen sich den Graduierungen selbst wieder Zahlen zuordnen, dann sind Bewertungen eine besondere Form der Messung. Bewertungen auf geeigneten Skalen sind entsprechend subjektive Messungen.

Items: Die Auswahl und Formulierung von bis zu drei abgefragten Aspekten (Items) je Dienst richteten sich nach den Erkenntnissen zu den nutzerrelevanten Stimuli aus den Voruntersuchungen. Jede Aufgabenbearbeitung schließt mit der Bewertung der Gesamtzufriedenheit mit dem mobilen Internet in der jeweiligen Situation ab. Prinzipiell besteht bei diesem Vorgehen die Gefahr, dass die Aufmerksamkeit der Probanden durch die gezielte Abfrage erst auf diese Aspekte gelenkt wird. Retrospektiv ist festzuhalten, dass eine umgekehrte Bewertungsreihenfolge u. U. besser gewesen wäre, also erst die Gesamtzufriedenheit und dann die Einzelaspekte.

Grundsätzlich sollten nur Aspekte abfragt werden, die eine QoS-abhängige Veränderung der Bewertung erwarten ließen. Items ohne Trennschärfe, d. h. ohne deutliche Veränderung der Bewertungen, sind in ihrer Aussage weniger wertvoll. Um dennoch keine relevanten Aussagen bzw. Zusammenhänge auszulassen, blieb im Zweifel eher eine Frage mehr im Fragebogen als unbedingt notwendig, da die Eingrenzung der relevanten Stimuli ja auch nur als subjektive Einschätzung durch den Autor und die beteiligten Projektpartner erfolgte.

Skala: Es wurde eine 5-stufige Skala⁸⁹ als Rating-Skala genutzt, siehe Abbildung 34, auf deren Basis die Probanden ihr Nutzererleben bewerteten und die sich wie folgt charakterisieren lässt:

- beschriftete ACR-Skala: die Bedeutung jedes einzelnen diskreten Skalenspunktes ist beschrieben
- zur subjektiven Selbsteinschätzung

⁸⁹ Likert-Skalen-ähnlich, da keine eindeutig positiven oder negativen Items zu bewerten waren

- in der verbalen Beschreibung wird eine von der mittleren neutralen Bewertung symmetrische Graduierung beschrieben
- in der optischen Gestaltung wird ein gleichmäßiger Abstand der Graduierung der Skala suggeriert
- die Beschreibungen suggerieren ungefähr gleiche Abstände

Auf Basis dieser Beschreibung ist die Skala daher nicht nur als qualitative, sondern auch als quantitative Skala interpretierbar. Im Kapitel 8 zur Auswertung wird die Problematik des Skalenniveaus und der statistischen Auswertbarkeit der Skala noch ausführlich diskutiert. Die genaue optische Darstellung kann den Screenshots zum Fragenbogen im Anhang entnommen werden.

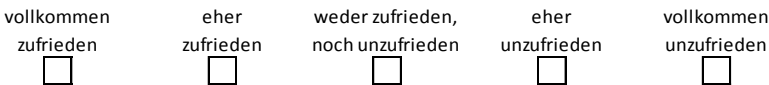


Abbildung 34: Beispiel der genutzten Skala zur Erfassung der Zufriedenheitsbewertungen

4.3.2 Bearbeitungserfolg

Der Bearbeitungserfolg zu jeder Aufgabenbearbeitung wird inklusive möglicher Ursachen für Wiederholungen oder Abbrüche erfasst. Die konkreten Antwortoptionen sind den Beispieldarstellungen im Anhang zu entnehmen. Die Diskussion zur Auswertung des Bearbeitungserfolgs als QoS-QoE-abhängige Größe findet sich im Abschnitt zur Auswertung der erfassten Daten (8.3).

4.3.3 Nutzerkommentare

Die Probanden erhielten die Möglichkeit, kurze Anmerkungen zu jeder Aufgabenbearbeitung zusammen mit ihrer Bewertung zu verfassen. Die Erfassung erfolgte im Web-Fragebogen als freiwillig zu füllendes Freitextfeld für allgemeine Kommentare. Ein zusätzliches Pflichtfeld wurde bei den Diensten MTV-Music und Youtube eingeblendet, falls dabei eine von „vollkommen zufrieden“ abweichende Bewertung des Abspielverhaltens ausgewählt wurde.

4.4 Weitere Ergebnisdaten

Neben den von den Probanden im Fragebogen erfassten Ergebnisdaten sind eine Reihe weiterer Daten aus unterschiedlichen Quellen in verschiedener Art und Weise erfasst worden, die ebenfalls einen Teil des Ergebnisraums bilden:

- von der Testleitung mit Erfassung im Testlogbuch
- von der technischen Protokollierung mit Erfassung in Logdateien
- aus dem Netzwerk mit Erfassung als Verkehrsdatenaufzeichnungen
- von den Endgeräten mit Erfassung als Videoaufzeichnungen der Aufgabenausführungen und Stimuli

Für die Auswertung sollten die unabhängigen Variablen, die abhängigen Variablen in Form der Nutzerbewertungen und die zusätzlich erhobenen Daten möglichst einfach zusammenzubringen sein. Dafür wurden zum einen die o. g. Identifikations- und Adressierungsmerkmale genutzt (Testslot, Testplatz, Aufgabennummer), zum anderen ist ein Abgleich über mitgeführte Zeitstempel möglich.

4.4.1 Testlogbuch

Im Testablauf festgestellte Auffälligkeiten sollten nachvollziehbar dokumentiert werden, um z. B. Ausreißer oder unsinnige Bewertungen und Kommentare ggf. später aufklären zu können. Die Erfassung erfolgte in einem Testlogbuch:

- während der Testdurchführung: besondere Vorkommnisse im Testbetrieb selbst, direkt festgestellte Abweichungen oder das Einwirken der Testleitung
- nachträglich zur Testausführung: festgestellte Abweichungen, z. B. beim Vorbereiten der Geräte für den nächsten Testslot

Das Testlogbuch wurde in der Testrunde 2 direkt mit der Testmanagement-GUI des Testbeds verknüpft und durch die Testleitung elektronisch geführt. So konnte die strukturierte Erfassung und Auswertung deutlich erleichtert werden. Weitere Anmerkungen und ein Screenshot zur Realisierung finden sich im entsprechenden Abschnitt des Kapitels zum Testbed, siehe 5.6.7.

4.4.2 Bearbeitungszeiten

Im Fragenbogen-Tool und der zentralen Testlaufsteuerung können automatisch die Umschaltzeitpunkte zwischen den Aufgaben des Fragebogens im Webbrowser der Testplatz-PCs der Probanden erfasst werden. Eine Interpretation als objektives QoE-Maß *benötigte Bearbeitungszeit* ist allerdings fraglich, da nicht zu trennen ist, wie groß die Zeitanteile für das Lesen der Aufgabenstellung, die tatsächliche Bearbeitung (eingesetzte Ressource Zeit), das Ausfüllen des Fragebogens und sonstige Ablenkungen jeweils sind. Entsprechend wird nachfolgend auch auf eine Auswertung der erfassten Zeitmarken verzichtet.

4.4.3 Protokollierung der Parametrisierung

Im Fragenbogen wurden zusätzlich zu jeder Aufgabenausführung automatisch die relevanten Identifikations- und Parameterraumdaten gespeichert: Testslot, Testplatz, Preset etc. Dafür wurden die geschaffenen Möglichkeiten zur Verknüpfung von Fragebogen-Tool und Testbed genutzt. So sind alle primär interessierenden Daten (UV und AV) im Fragenbogensystem zusammengefasst und konnten von da aus direkt in die Auswertung transferiert werden. Die korrekte technische Funktion des Testbeds kann zusätzlich anhand von Logdateien nachvollzogen werden, die die jeweilige Parametrisierung, Umschaltzeitpunkte und ggf. aufgetretene technische Probleme innerhalb des Testbeds beinhalten.

4.4.4 Überwachung und Aufzeichnung des Netzverkehrs

Die während der Versuchsdurchführungen erzeugten Verkehrsdatenströme innerhalb der Testumgebung sollten möglichst erfasst und aufgezeichnet werden. Erklärungen zur technischen Realisierung finden sich im Kapitel zum Testbed (5.9). Die Verkehrsdatenaufzeichnungen können für eine optionale Tiefenauswertung genutzt werden, um:

- die tatsächlich erreichten QoS-Parameterwerte zu ermitteln
- externe Störgrößen zu ermitteln bzw. auszublenden
- die Korrektheit der Aufgabenausführungen zu prüfen

In der Diskussion der Bestandteile der Auswertung wird auf die tatsächliche Verwendung der erfassten Verkehrsdaten detailliert eingegangen, siehe 8.1.1.

4.4.5 Überwachung und Aufzeichnung der Endgeräte per Video

Die Überwachung der Geräteausgaben in Form des sichtbaren Bildschirm-inhalts (und des Audiosignals) sowie der Geräteeingaben erscheint zunächst vielleicht wie eine Spielerei, ist aber eine technische Besonderheit, die sich bei genauer Betrachtung durchaus lohnen kann. Der unbemerkte, nicht störende Blick quasi über die Schulter der Probanden ist von großem Vorteil für:

- Erkennung wiederkehrender Bedienprobleme und Stolpersteine in der Bearbeitung der Aufgaben durch die Probanden
- Einschätzung der Korrektheit der einzelnen Aufgabenausführungen, Schrittfolgen und Zeitpunkte
- Erfassung der tatsächlich erreichten Nutzstimuli als relevante QoE-Einflussgrößen

Es wurde eine praktikable technische Lösung geschaffen und zum Einsatz gebracht, siehe Erläuterungen dazu im Kapitel Testbed (5.10). Im Rahmen der Laborstudie wurden einige reguläre Testläufe der Probanden überwacht und aufgezeichnet. Als Proof of Concept hat sich die erarbeitete Lösung bewährt, die erwarteten Vorteile und zusätzlichen Erkenntnisse bestätigten sich. Eine Überwachung und Auswertung aller Testläufe war mit den geplanten und eingesetzten Ressourcen allerdings nicht zu realisieren. Stattdessen wurden für alle Szenarien beispielhafte Aufgabenausführungen und beobachtbare Nutzerstimuli bei unterschiedlicher Netzparametrisierung aller Presets zusammen mit den zugehörigen Netzverkehrsdaten zur späteren Nachvollziehbarkeit dokumentiert. Auf Basis der synchronisierten Aufzeichnungen von Endgeräte-Monitoring und Netzwerk-Monitoring sind so auch nachträgliche Analysen zur gleichzeitigen Sicht von *außen* (Videodaten zur Nutzersicht) und *innen* (Netzverkehrsdaten zur Netzsicht) möglich, siehe Erläuterungen im Abschnitt zu den Voruntersuchungen.

Ein solcher Ansatz sollte gerade in Pretests zur Beseitigung letzter Probleme mit den Aufgabenstellungen und zur Dokumentation ausgewählter Testläufe (spätere Demonstration) grundsätzlich zum Einsatz kommen. Die umfassende Überwachung und Auswertung aller Testläufe ist ein netter Bonus, wenn ausreichend (personelle) Kapazitäten vorhanden sind.

4.4.6 Weitere mögliche Datenquellen

Im Test geben die Probanden eine subjektive Bewertung ab. Die so erhobenen Daten sind ggf. verzerrt oder verfälscht. Dies geschieht zum Teil unbewusst durch interne Filterung und Vorbewertung, eventuell aber auch vorsätzlich als unehrliche Antwort. Die nachfolgend beschriebenen Ansätze wurden **nicht realisiert**, sollen aber andeuten, was prinzipiell möglich ist, um an verlässlichere Daten zu gelangen (zumindest vermeintlich), siehe Grundlagen zum physiologischen Messen allgemein in [216] S. 251ff.

Objektive Bewertungen in Nutzertest können aus der Messung der Aktionen bzw. (unbewusster) körperlicher Reaktionen der Probanden gewonnen werden, fast wie in einem Lügendetektor (zusammengestellt aus [217] S. 32ff):

- Videoaufzeichnung der Probanden (Mimik, Gestik, Ablenkungen)
- Eye-Tracking zum Verfolgen der Aufmerksamkeit
- Auswertung der Emotionen, ggf. auch automatische Analyse, z. B. mit FaceReader⁹⁰-Software
- Auswertung von Herzfrequenz, Blutdruck, Gehirnaktivität, Hautleitfähigkeit, Muskeltonus

Bei gesicherter Kenntnis entsprechender Zusammenhänge zwischen objektiver Messung und gesuchtem QoE-Maß kann der Unsicherheitsfaktor Mensch minimiert werden. Der Proband agiert dann nur noch als Aktor und Sensor, ein externes System interpretiert auf Basis der gewonnenen Daten das Ergebnis. Grundlagen und eine Übersicht zum Forschungsstand des „Psychological Computing“ (mit Fokus auf HCI) liefert [218]. Für ausgewählte Anwendungen in der QoE-Domäne konnten in der Literatur sogar schon Ansätze und Ergebnisse ermittelt werden. Beispiele finden sich für die Bewertung der Medienqualität von Videos [219] oder auch beim mobilen Web-Browsing [220] per EEG, wobei sogar die neurologischen Auswirkungen messbar werden, die zu Stress und schließlich Frustration führen sollten. Damit ist eine weitere Anknüpfungsmöglichkeit der Forschung rund um QoE hin zu den Neurowissenschaften genannt, vertiefende Informationen dazu finden sich in [221].

⁹⁰ <http://www.vicarvision.nl/FaceReader.html> und <http://www.noldus.com/facereader/facial-expression-analysis> abgerufen am 22.08.2013

Mit der Einbeziehung und der erweiterten Erhebung von Daten aus den beschriebenen zusätzlichen Quellen würden nicht nur technische Aspekte der Machbarkeit und Interpretation einhergehen, sondern ggf. auch ethische und datenschutzrechtliche Bedenken zu prüfen sein. Sofern in Videoaufnahmen Probanden zu erkennen bzw. zu identifizieren wären, erhöhen sich ggf. auch die Anforderungen an den Datenschutz, da die Daten selbst nicht mehr grundsätzlich anonymisiert oder pseudonymisiert wären. Mit Erfassung und Verarbeitung ggf. auch medizinisch bzw. gesundheitlich relevanter Daten verschärfen sich die Ansprüche an Ethik und Datenschutz noch weiter.

4.5 Durchführung der Studie

Nachdem alle relevanten Vorüberlegungen des Testdesigns und der Testplanung dargelegt sind, folgt in diesem Abschnitt eine kurze Beschreibung der handwerklichen Aspekte der eigentlichen Studiendurchführung als der Phase, in der unmittelbar Probanden beteiligt waren.

4.5.1 Interne Vorabtests und Pretests

Begleitend zu den Vorüberlegungen und Voruntersuchungen, vor allem zu denen im Testparcours, wurden interne Vorabtests durchgeführt. Diese Vorabtests erfolgten in der Phase der fortgeschrittenen Planung als Selbsttests bzw. im kleinen Kreise der beteiligten Partner der TU Chemnitz (drei eingeweihte Nutzer). Eine kommentierte Nutzung zu möglichen kritischen Punkten und erste Vermutungen zu den Bewertungsbereichen standen dabei im Vordergrund. Auf Basis der so ermittelten groben Einschätzungen konnten zusammen mit weiteren plausiblen Propositionen die interessanten Parameterbereiche sinnvoll eingegrenzt und schließlich festgelegt werden.

In Vorbereitung der Testrunde 2 wurden Kollegen mit Kenntnissen zum Studieninhalt aber ohne Beteiligung an der Erarbeitung der Details als interne Vortester eingesetzt. Dabei erfolgte eine direkte Beobachtung der selbstständigen Ausführung mit sofortiger Befragung der Kollegen nach jeder Aufgabe, wenn Verunsicherungen oder Probleme zu beobachten waren. Diese neutrale Erprobung diente primär der Verbesserung der Formulierungen von Aufgaben- und Fragestellungen im Fragebogen. In die gleiche Richtung zielte ein

Pretest, mit dem der entwickelte Fragebogen und die festgelegten Parameterbereiche mit echten Probanden (normale Nutzer ohne weitere Vorkenntnisse) getestet wurden. Dabei bearbeiteten die Probanden die Aufgaben wie geplant ohne Unterbrechungen und bewerteten auch entsprechend. Für die Gewinnung abschließender Erkenntnisse zur Verbesserung hauptsächlich der Aufgabenstellungen aus Nutzersicht diente eine kurze individuelle mündliche Befragung nach der selbstständigen Ausführung aller Aufgaben, in der die Probanden ihre Anmerkungen und festgestellten Problemen äußerten.

Den Abschluss der Vorbereitungen bildete eine Erprobung des erarbeiteten Konzepts mit drei eingeweihten Nutzern aus dem Kreis der beteiligten Partner in einer Art Endabnahme. Die gesammelten Erkenntnisse wurden noch in Form geringfügiger Änderungen in die finalen Fragebögen integriert. Daran anschließend konnte die eigentliche Durchführung der Studie starten. Die Ergebnisse der internen Vorabtests und des Pretests sind nicht Teil der Ergebnismenge, entsprechend werden die darin erhobenen Daten in der Auswertung nicht berücksichtigt.

4.5.2 Testlogistik

Eine Studie mit mehreren Hundert Probanden ist aus Sicht der Durchführung deutlich anspruchsvoller als eine kleine Studie mit nur wenigen Teilnehmern. Die notwendige Testorganisation und Testlogistik ist daher von hoher Bedeutung und sollte keinesfalls unterschätzt werden. Neben der grundlegenden technischen Realisierung, um überhaupt eine solche Studie durchführen zu können, sind diese nichttechnischen Aspekte vielleicht das Haupthindernis für QoE-Evaluationen in großem Umfang. In der Testrunde 1 erfolgte die begleitende Organisation noch an der Professur MI, in der Testrunde 2 dann durch den Partner AAP und einen externen Dienstleister. Die Testorganisation und Testbetreuung durch AAP bzw. den externen Dienstleister umfasste:

- die Anwerbung und Kontaktabbauung mit den Probanden, siehe Abschnitt 7.2.5
- Vorbefragung und Auswahl der Probanden passend zu den gewünschten Vorkenntnissen und zur angepeilten Zusammensetzung der Stichprobe
- Terminfindung mit den Probanden und Einladung der Probanden, passend zu den Auswahlkriterien und den Ressourcen im Testbed



Abbildung 35: Bilder aus der Testumgebung

a) Testplatz mit zugeordnetem mobilen Gerät und Kopfhörer vor PC-Bildschirm

b) Testplätze für die Probanden und Bereich der Testleitung (Web-GUI, Verkehrsaufzeichnung, Geräteüberwachung)

4.5.3 Testumgebung

Die Nutzertests fanden in einer ruhigen Testumgebung eines Computer-Pool-Raums der Professur Medieninformatik statt. Darin standen mehrere komplett ausgestattete moderne Computer-Arbeitsplätze mit Zugang zum Campusnetz bereit. Der Raum wurde während der Testdurchführung exklusiv für die Studie genutzt. Aufbau und Einrichtung des Testbeds zur Netzemulation erfolgten vor Beginn der Testdurchführung innerhalb dieses Raums und speziell angepasst an diesen Raum. Die Auswahl der Arbeitsplätze im Testraum erfolgte so, dass die Probanden sich nicht unmittelbar gegenseitig störten, mit ausreichend seitlichem Abstand und Rücken zu Rücken zueinander orientiert. Die

ausgewählten Arbeitsplätze erhielten eine feste und eindeutige Kennzeichnung mit einer Nummer (fortan als Testplatz bezeichnet), siehe Abbildung 35a. An jedem Testplatz lagen Kopfhörer zum Anschluss an die mobilen Geräte bereit, die für die Tests mit Audioausgabe zu nutzen waren. Die PCs an den Testplätzen wurden während der Tests ausschließlich für die Online-Fragebögen im Web-Browser genutzt. Zusätzlich waren Arbeitsplätze für die Testleitung reserviert, auf denen der korrekte Testablauf überwacht werden konnte. Die Monitore der Testleitung waren für die Probanden während der Tests nicht einsehbar, siehe Abbildung 35b. Durch die Anordnung der Tischreihen und Monitore zueinander entstand gleichzeitig eine gewisse räumliche Trennung im Testraum zwischen dem Bereich der Probanden und dem Bereich der Testleitung.

4.5.4 Technischer Testablauf

Neben den organisatorischen Aspekten waren auch routinemäßige aber wesentliche technische Abläufe zu realisieren. Aus Gründen der Vereinfachung werden diese Durchführungsschritte für die Technik und den sonstigen Ablauf gemeinsam dargestellt, da sie letztlich eine Einheit bilden und ineinander verzahnt sind (siehe Auflistung der Phasen). In den Phasen 2 und 3 wurde die technische Testleitung (MI) zusätzlich zur Probandenbetreuung vor Ort verstärkt (AAP), um die Aufklärung der Probanden inklusive nichttechnischer Einweisung und das Management der Incentives zu übernehmen. Die nicht unmittelbar auf die Probanden bezogenen Tätigkeiten der Testleitung während der Testdurchführung erfolgten verdeckt, für die Probanden nicht wahrnehmbar, siehe Erläuterungen zur Testumgebung. Der eigentliche Testablauf zu jedem Testslot bestand aus einer Wiederholung der immer gleichen vier Phasen:

- 1) Vorbereitung (vor Teststart und Eintreffen der Probanden)
 - a) Inbetriebnahme des Emulations-Testbeds
 - b) Aktivieren der aktuellen Testpläne und zugehöriger Fragebögen in der Testlaufsteuerung
 - c) Start der Fragebögen auf den PCs der Probandenarbeitsplätze
 - d) ggf. letzte Einstellungen auf den Endgeräten kurz vor Teststart
 - e) Vorbereiten der Aufzeichnung der Verkehrsdaten
 - f) optional: Vorbereiten der Videoaufzeichnung der Geräteausgaben
- 2) Probandenbegrüßung und -einweisung
 - a) Zuweisung der Probanden zu den geplanten Testplätzen

- b) schriftliche Formalien: Probandeninformation, Einverständnis- und Datenschutzerklärung
 - c) technische Kurzeinweisung zum Gerät (Lage der Tasten, Schließen der Apps): mündlich und individuell
 - d) Probandeninstruktion: letzte Hinweise zur Studie und zum Testablauf, mündlich, gemeinsam für alle Testteilnehmer, in Ausnahmen individuell bei Nachzögler
- 3) Testdurchführung
- a) Start der Aufzeichnungen der Verkehrsdaten und optional der Geräteausgaben
 - b) Start der Bearbeitung der Aufgaben für alle Probanden gemeinsam, mit Ausnahme einzelner Nachzügler
 - c) Überwachung des Testfortschritts in Testleiter-GUI
 - d) Ad-hoc-Hilfestellung und Problemlösung für die Probanden
 - e) Erfassung besonderer Vorkommnisse und Anmerkungen im Testlogbuch
 - f) Überwachung der Verkehrsaufzeichnung
 - g) optional: Kontrolle der überwachten Geräteausgaben
 - h) individuelles Bearbeitungsende der Probanden, Übergabe des Incentives und Verabschiedung
- 4) Nachbereitung (nach Abschluss der Testdurchführung des letzten Probanden)
- a) Stopp der Aufzeichnungen
 - b) Sicherung der Testdaten (Testpläne, Testlogs) und Aufzeichnungen
 - c) Überprüfung der Endgeräte und Aufladen der Akkus
 - d) Rücksetzen der Testumgebung und Endgeräte auf definierten Ausgangszustand inklusive schriftlicher Protokollierung (Checkliste) nach fixierter Schritt-für-Schritt-Anleitung
 - e) für längere Unterbrechungen: Außerbetriebnahme des Testbeds

4.6 Zusammenfassung

Die empirische Ermittlung des QoS-QoE-Zusammenhangs erfolgte in einem Laborexperiment, also als gezielte und geplante Bedingungsvariation innerhalb einer geschlossenen Testumgebung. Die Studie ist multifaktoriell und multivariat angelegt, es werden mehrere unabhängige und mehrere abhängige Variablen berücksichtigt. Zur Beachtung der zahlreichen Nebenfaktoren wurde ein recht kompliziertes gemischtes Testdesign (Within-Subject und Between-Subject) erforderlich. Dabei sind Wechselwirkungen der Nebenfaktoren und die davon ausgehenden Auswirkungen zu berücksichtigen, die je nach Dienst in der Auswertung zu verschiedenen Ergebnispartitionen führen können. In der Testplanung wurden sinnvolle Festlegungen zur planvollen Bedingungsvariation genutzt, die mögliche Störeinflüsse durch die Aufgabenzusammenstellung ausmitteln sollten, bei gleichzeitig überschaubarem Aufwand und praktikablen Ansatz eines Block-Shufflings.

Die primär interessierenden (subjektiven) Bewertungen der Probanden zur Zufriedenheit als relevantes QoE-Maß wurden in einem Fragebogen über eine 5-stufige Skala erfasst. Bearbeitungserfolg und Nutzerkommentare ergänzten diese Angaben. Daneben wurden weitere Ergebnisraumdaten in Form eines Testlogbuches, der Aufzeichnungen der Netzwerkverkehrsdaten und teilweise auch der Eingaben und Ausgaben der Endgeräte als Videoaufzeichnungen erzeugt, die zur weiteren Auswertung zur Verfügung standen. Dem Haupttest gingen Pretests zur Eingrenzung der interessanten Parameterbereiche und zur Verbesserung der Fragebogeninhalte voraus. Die Testdurchführung in der geschaffenen Testumgebung erfolgte in standardisierter Art und Weise, ebenso wie die Testvorbereitungen. Es waren besondere logistische und testorganisatorische Herausforderungen in Zusammenarbeit mit AAP zu meistern.

Tabelle 22 beschreibt die Empirie zur Studie zusammengefasst in einer Art Steckbrief.

Tabelle 22: Charakterisierung und Klassifizierung der Studie

Kriterium	Realisierung
Untersuchungsziel	Überprüfung der Hypothesen zum QoS-QoE-Zusammenhang als Parameterbereichstests zur Wirkung der QoS-Netzparameter/ QoS-Netzparameterstufen auf die QoE-Bewertungen Überprüfung der Wirkung weiterer Faktoren gemäß der Überlegungen dazu bzw. zu ihrer Erkundung und Verbreiterung der Wissensbasis zum Einfluss der Nebenfaktoren
Untersuchungsart	Nutzertest mit naiven Probanden zur subjektiven QoE-Erhebung (Single Stimulus)
Methode	psychophysikalisches (Quasi-) Experiment zur empirischen Ermittlung: menschliches Erleben und Verhalten (psycho) in Abhängigkeit externer Reizgrößen (physikalisch)
Untersuchungsort/ Umgebung	Labor: Durchführung in kontrollierter Testumgebung eines Testraums mittels des speziell geschaffenen technischen Testbeds
Zahl der unabhängigen Variablen (UV)	multifaktoriell: mehrere unabhängige Variablen
UV: Versuchsplanung	planmäßige Variation der unabhängigen Variablen und gezielte Herstellung der Bedingungen, Randomisierung bei der Zuweisung der Probanden zu den Versuchsbedingungen
Zahl der abhängigen Variablen (AV)	multivariat: mehrere abhängige Variablen
AV: Art der Ergebniserfassung	subjektive Angaben durch Abfrage und Selbstauskunft der Probanden (Bewertung und Kommentare)

Zusammenstellung der Kriterien in Anlehnung an [83] S.423ff (allgemeine Evaluation) und [222] (Experimentarten), erweitert und ergänzt

5 Das Testbed

Netz und Endgeräte in der Studie

Die Technik des Testbeds schafft den notwendigen Rahmen zur empirischen Ermittlung des QoS-QoE-Zusammenhangs für mobile Dienste mit variablen Netzbedingungen. Dazu wird gezeigt, wie „das Netz“ und die Endgeräte in den Nutzertest integriert wurden. Dieses Kapitel beschreibt neben den zugehörigen theoretischen Abwägungen zum mobilen Datennetz und den Netzsituationen vorrangig die im Rahmen der Arbeit entstandene technische Realisierung dazu.

Zu Beginn wird „das Netz“ weiter konkretisiert. Die Nutzung mobiler Dienste im mobilen Internet bedeutet die Nutzung von Mobilfunk. Daher wird kurz die Technik des Mobilfunks diskutiert. Daraus können die Netzsituationen in Form von mobilfunktypischen QoS-Parameterbereichen und QoS-Parameterkombinationen abgeleitet werden. Daran schließt sich die Diskussion der Anforderungen an das Testbed an, die auch eine Analyse der möglichen technischen Konzepte zur Integration des Netzes mit den identifizierten Netzsituationen in die Studie beinhaltet. Ausgehend von diesen Überlegungen wird das Prinzip einer Netzemulation in Kombination mit der Nutzung weiterer echter Elemente als geeignet vorgestellt.

Die gewählte Realisierung des Testbeds und Emulations-Frameworks auf Basis etablierter Open-Source-Software wird erläutert. Dies umfasst die eigentliche Netzemulation mit den Basisfunktionen zur Nachbildung der Netzsituationen und zugehörige Erweiterungen zur praktikablen Nutzung im regulären Testbetrieb. Die Vorzüge und Besonderheiten der eigenen Lösung werden erörtert, die die entwickelte Lösung für einen Einsatz in einer Probandenstudie tauglich machen. Dazu zählen Skalierbarkeit, konsistente Datenhaltung, (Web-) APIs zur Steuerung, eine Web-GUI und die Kopplung mit dem Web-Fragebogensystem zur automatischen Testlaufsteuerung. Durch diese Lösungen konnte mit geringem personellem Aufwand, zielgerichtet in kurzer Zeit eine große Anzahl Probanden getestet werden. Es werden aber nicht nur die

Vorteile dargelegt, sondern es wird auch ein kritischer Blick auf Einschränkungen und Grenzen geworfen. Dies beinhaltet den Ansatz zur Netzemulation an sich, die eigene Realisierung in Hardware und Software sowie auch die Gegebenheiten der Test- bzw. Netzumgebung.

Ein weiterer Teil des Testbeds sind die mobilen Endgeräte. Auswahl und Zusammenstellung werden zusammen mit festgestellten Besonderheiten zum Betrieb innerhalb des Testbeds diskutiert. Abschließend werden die Möglichkeiten des Testbeds erörtert, die über eine reine Netzemulation hinausgehen. Dies sind Überwachung und Aufzeichnung sowohl der Netznutzung als auch der direkten Eingaben und Ausgaben der Endgeräte aus Nutzersicht.

5.1 Mobilfunknetz und mobile Netzsituationen bzw. Netzbedingungen

Die Netzsituationen sind die unterschiedlichen Ausprägungen des Merkmals *Netzcharakteristik* bzw. *Netzleistung/Netzqualität* des Netzes. Im Test sollten diese gezielt variiert werden. Dabei interessieren vor allem Netzsituationen, die mobiltypisch sind und wie eine typische Netzcharakteristik in einem Mobilfunknetz überhaupt aussieht. Dafür soll zunächst kurz auf die Technik der Mobilfunknetze sowie ihre relevanten Netzsegmente und Netzelemente eingegangen werden. Nachfolgend wird geklärt, welche QoS-Charakteristiken Mobilfunknetze aufweisen.

Bei der Auswahl relevanter Parameter und der Festlegung interessanter Bereiche und Abstufungen sind mehrere Wege möglich, die nachfolgend genauer betrachtet werden.

In der Theorie: aus den zugrundeliegenden Technologien und Spezifikationen → Was sind Technik und Netzbetreiber prinzipiell (theoretisch) in der Lage an QoS-Werten/Netzleistung anzubieten (QoS-offered)? Technische Standards spezifizieren dabei technologisch bedingt bestmögliche, also zumindest theoretisch erreichbare Werte.

In der Praxis: aus der Messung „der Netzqualität“ als Ende-zu-Ende-Messung mit Rückschluss auf einzelne Netzabschnitte und Auslastung
→ Welche Leistung (Ende-zu-Ende-QoS) steht Kunden in Mobilfunknetzen typischerweise in der Praxis zur Verfügung? Orientierung gibt dabei die statistische Auswertung einer Vielzahl von Messungen, die als Richtwerte für erreichte relevante Werte (QoS-achieved) dienen können.

Darauf aufbauend wird gezeigt, wie aus diesen Überlegungen sinnvolle Parameterbereiche und Parameterkombinationen für die weitere Nutzung in den Tests in einer Matrix technischer Parameterkombinationen festgelegt wurden. Diese dienen später als Vorgabewerte für die dienstspezifische Parametrisierung der Netzemulation zur gezielten Herstellung der repräsentierten Netzsituationen passend zu den untersuchten Nutzungssituationen.

5.1.1 Mobilfunknetze und Netzelemente des Mobilfunks

Moderne Mobilfunknetze sind komplexe, hierarchisch organisierte Infrastrukturnetze⁹¹ bestehend aus Komponenten des drahtlosen Zugangsnetzes (Radio Access Network, RAN), dem internen Kernnetz (Core Network, CN) des Providers und dem Übergang zum externen Netz mit dem Internet. Abbildung 36 zeigt die wichtigsten Netzabschnitte und Netzkomponenten zur Datenkommunikation im Mobilfunk dazu schematisch, die für die Realisierung eines Zugangs ins mobile Internet notwendig sind. Die Beschreibung weiterer Bestandteile, wie sie in produktiven kommerziellen Netzen üblich sind, geht über den notwendigen Rahmen für die weiteren Erläuterungen hinaus und wurde deshalb ausgelassen. Weitergehende Informationen zu Netzelementen, Protokollen und Prozeduren zur paketbasierten Datenkommunikation im Mobilfunk für GPRS/UMTS hält [56] S. 157ff bereit. Noch ausführlichere Erläuterungen, inklusive Evolved Packet Core und Long Term Evolution (LTE) bieten [223] und [224], die auch die innere Funktion detailliert beschreiben.

Zugangsnetz: Mobilfunknetze bilden zellulare Netze, jede einzelne Basisstation versorgt jeweils drahtlos per Funk eine gewisse Anzahl Endgeräte in einem bestimmten Versorgungsgebiet. In einer modellhaften Vereinfachung werden dabei meist Hexagone (Zellen) zur Darstellung genutzt. Die aktiven

⁹¹ in der Systematik nach [34] S. 149ff

Endgeräte in einer solchen Zelle teilen sich dabei die verfügbaren Ressourcen. Die spektrale Bandbreite für die Nutzung im Mobilfunk ist eng begrenzt und streng reguliert (und damit eine knappe Ressource).

Kernnetz: Der Nomenklatur folgend sind die interessanten Netzelemente des Kernnetzes Nodes, Gateways oder Functions im Sinne von Elementen zur Erfüllung definierter Funktionen. Zu diesen Funktionalitäten zählen die Verwaltung der Nutzerdatenbasis, Mobilitätsaspekte und die Erbringung Provider-eigener Dienste. Prinzipiell bieten moderne Mobilfunknetze die Möglichkeiten zur gezielten Beeinflussung relevanter QoS-Parameter durch QoS-Mechanismen. Über die Policy and Charging Rules Function (PCRF) können dabei auch künstliche Beschränkungen leistungsrelevanter Parameter realisiert werden, z. B. des Durchsatzes, in Übereinstimmung mit vertraglich festgelegten Vereinbarungen. Für die Realisierung paketbasierter Datendienste werden zusätzliche Netzelemente in Form von Gateways benötigt, die auch den Übergang zu externen Netzen und somit ins Internet realisieren.

Externes Netz/Internet: In der abstrakten Sicht ist das (öffentliche) Internet eine Reihe von Netzwerken, betrieben und kontrolliert von unterschiedlichen Anbietern. Von außen erscheint das öffentliche Internet als Black Box ohne direkte (zumindest individuelle) Eingriffsmöglichkeit. Es ist von Best Effort auszugehen, es stehen also keine QoS-Mechanismen, zumindest nicht Ende-zu-Ende, zur Verfügung (Erläuterungen siehe 2.2.8). Die Charakteristik entlang einer Route durch das Netz lässt sich daher nur abschätzen. Die Netzanbindung der Gegenseite (Server oder andere Nutzer) an das Internet kann wiederum in unterschiedliche Netzsegmente und Technologie zerfallen, ist aber für die weiteren Betrachtungen nicht von Bedeutung, alle externen Netzsegmente sind als das Internet zusammengefasst zu betrachten.

Erläuterungen zur Abbildung 36: Beschränkung der Darstellung auf die Mindestkomponenten zur paketbasierten Datenkommunikation unter Auslassung zusätzlicher Elemente und Verzicht der Darstellung des Datendienstes des GSM-Systems (2G), da nicht mehr relevant vertikale Dreiteilung der Darstellung in GPRS/UMTS mit UTRAN oben und EPC mit eUTRAN (LTE) in der Mitte und vereinfachte Sicht unten horizontale Dreiteilung der Darstellung in externes Netz, Kernnetz und Zugangsnetz; in Anlehnung an [144] S. 131 (UMTS vereinfacht), [144] S. 232 (LTE vereinfacht), [111] S. 12 (und [225] S. 132 (erweitert und ergänzt)

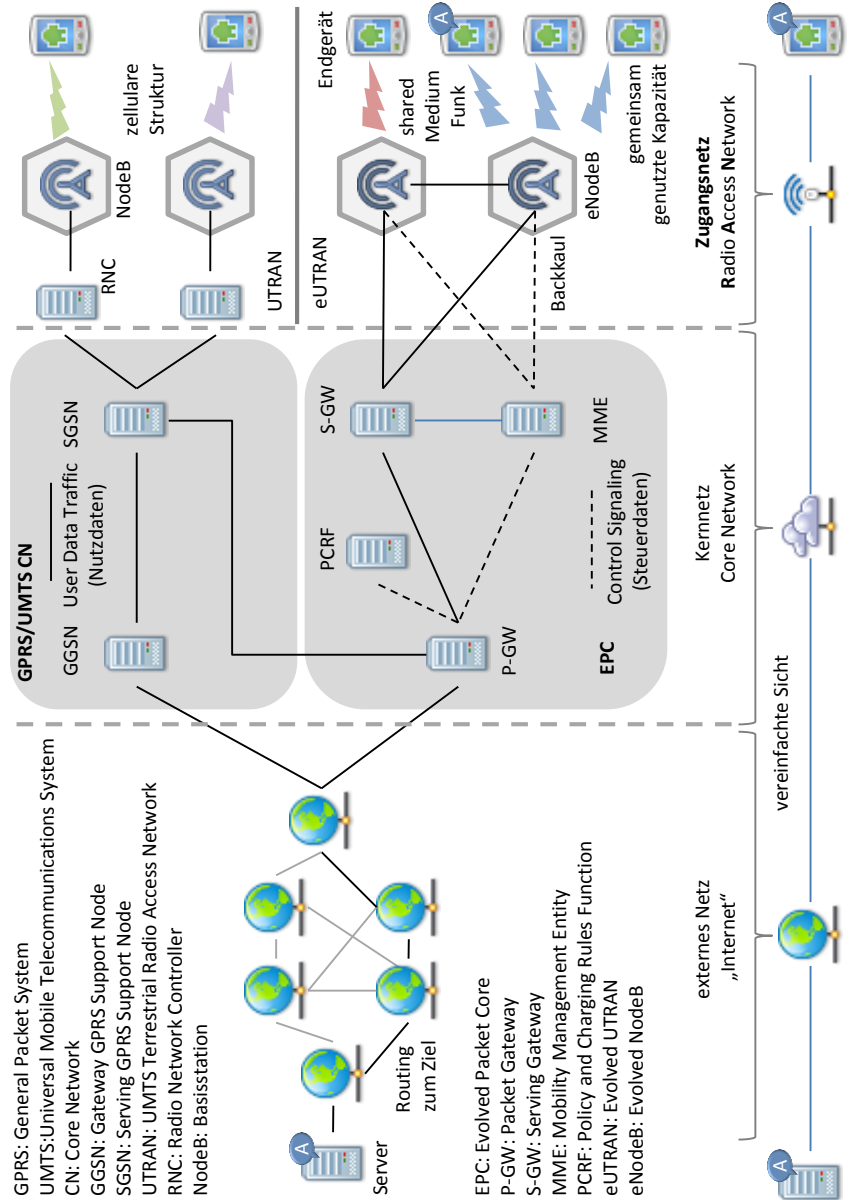


Abbildung 36: Netzsegmente und Netzelemente des Mobilfunks zur Datenkommunikation und vereinfachte Sicht

5.1.2 Technologische Ausgangsbasis: Mobilfunkstandards

Es ist sinnvoll, die technologisch vorgegebenen Eckdaten als eine Abschätzung zur prinzipiell bestmöglichen Leistung zu betrachten. Die Etappen der Entwicklung der QoS-Netzparameter gehen mit der Evolution der Mobilfunkstandards einher. Die Spezifikationen des zuständigen 3rd Generation Partnership Projects (3GPP) sind damit die erste Anlaufstelle für Anhaltspunkte zu technischen Rahmendaten. Einen geschichtlichen Abriss zur Entwicklung der Mobilfunkstandards und den zugehörigen Neuerungen jeder Mobilfunkgeneration mit den jeweiligen Releases liefert [226]. Jeder Generationsschritt bedeutete eine Verbesserung entweder in der Luftschnittstelle zwischen mobilem Endgerät und Mobilfunkinfrastruktur (Radio Access Network, RAN) oder im Kernnetz des Anbieters, die zu besseren Leistungswerten wie höheren Bandbreiten und niedrigerer Latenz führten, siehe Tabelle 23 inklusive Mini-Glossar für die Akronyme der Netzgenerationen. Dabei sind insbesondere die Stufen in der HSPA-Evolution (technische Erläuterungen in [227]) dafür verantwortlich, dass der einmal in Schwung gekommene Innovationszyklus aus Netz und Nutzung immer weiter angetrieben wurde. So konnte innerhalb weniger Jahre der Spitzendurchsatz vervielfacht werden (Faktor > 100). Ausgehend von den UMTS-Parametern wurde dabei zunächst die Downlink-Richtung durch HSDPA und nachfolgend auch die Uplink-Richtung durch HSUPA beschleunigt.

Fortschritte zwischen den Generationen sind vor allem durch Verbesserungen in der Luftschnittstelle auszumachen. Dabei erfolgt durch die Technik eine Annäherung der spektralen Effizienz (in Bit/s/Hz, abhängig vom Signal-Rausch-Verhältnis) immer weiter an das theoretische Optimum, die Shannon-Grenze, siehe [17] S. 15. Damit sind die Bruttodatenraten pro Sektor einer Zelle bei bekannter spektraler Bandbreite grundsätzlich nach oben abzuschätzen. Zukünftige Erweiterungen der Kapazität bedeuten damit größeren Bedarf an Spektrum, kleinzellige Netze, die kombinierte Nutzung von Ressourcen (z. B. Zellbündelung bei der Dual-Cell-Nutzung) und/oder Mehrantennensysteme. Die grundsätzlichen Beschränkungen daraus ergeben sich aus der Regulation (Aufteilung des Frequenzspektrums), den eingesetzten finanziellen Ressourcen zum Erwerb von Frequenzspektrum und zum Ausbau des Netzes (Generationsfortschritt und Verdichtung).

Über die Generationen hinweg wurde die Latenz zwar ebenfalls verbessert, aber bei Weitem nicht um den gleichen Faktor wie der Spitzendurchsatz. Technologisch bedingt (u. a. Koordinierung des Zugriffs auf das gemeinsame Medium Funkkanal) ist die mobilfunktypische Latenz noch immer deutlich höher als in lokalen Netzen und meist auch als in der stationären Internet-Nutzung.

Tabelle 23: Netzgenerationen und theoretische QoS-Netzparameter im Mobilfunk

Generation	Release	Release-Jahr	Downlink (ca. bis zu)	Uplink (ca. bis zu)	Latenz (RTT) (ca. min.) ms	Erläuterung Akronym / Mini-Glossar
2	Phase 1	1992	9,6 bis 14,4 kBit/s	9,6 bis 14,4 kBit/s	> 700	GSM: Global System for Mobile Communications
2.5	97	1998	50 kBit/s	13 bis 26 kBit/s	> 500	GPRS: General Packet Radio Service
2.75	98	1999	236 kBit/s	118 bis 236 kBit/s	300 bis 400	EDGE: Enhanced Data rates for Global Evolution
3	99	2000	384 kBit/s	128 bis 384 kBit/s	170 bis 200	UMTS: Universal Mobile Telecommunication System
	4	2001				All-IP Kernnetz IMS: IP Multimedia Subsystem
3.5	5	2002	7.2 MBit/s (14 MBit/s)	384 kBit/s 5.8 MBit/s	130 90	HSDPA: High Speed Downlink Packet Access HSUPA: High Speed Uplink Packet Access (gemeinsam HSPA: High Speed Packet Access)
	6	2004				
3.75	7	2007	42 MBit/s	23 MBit/s	50 bis 100	HSPA+: HSPA Evolution oder HSPA evolved
3.9	8	2008	100 MBit/s bis 300 MBit/s	50 MBit/s bis 75 MBit/s	30 bis 50	LTE: Long Term Evolution
	9	2009				Verbesserungen SAE: Service Architecture Evolution (im Kernnetz)
4	10	2010	1000 MBit/s	500 MBit/s	-	LTE-Advanced

noch ohne Erkenntnisse zur Latenz für LTE-Advanced (noch keine marktreife Technologie), u. a. zusammengestellt aus [229], [230] S. 1ff und [164] Tab. 7-4 3, erweitert und ergänzt, zusätzlich sei auf die Tabelle 67 im Anhang A.1 mit weiteren Abstufungen der einzelnen Releases verwiesen, je nach Markt erfolgte der Rollout (Einsatz tatsächlicher Netztechnik in den Produktivnetzen) jeweils mit einigen Jahren Verzögerung zum Release-Jahr der Spezifikation, weitere Release nach 10 ohne Darstellung

[228] S. 164ff erläutert die Verbesserungen beim Schritt von GPRS zu UMTS, die in 3G-Netzen immerhin schon zu einer deutlichen Senkung der Latenz führten. In Netzen der übernächsten Generation (5G) sind weitere Senkungen der Latenz zur Ermöglichung neuartiger Dienste, vor allem mit engen Echtzeitanforderungen, vorrangiges Entwicklungsziel, siehe [187].

Die durch Technik oder Anbieter vorgegebenen Werte ordnen sich im Bereich QoS-offered ein. Die aufgeführten Datenraten sind entsprechende Bruttowerte unter Idealbedingungen („mit bis zu...“), die üblicherweise auch so von den Netzanbietern angegeben werden.

Ebenso relevant wie die Unterstützung im Netz, bzw. genauer in der Netzinfrastruktur, ist natürlich auch die Unterstützung im mobilen Endgerät (siehe dazu entsprechende Erläuterungen in 3.2.3 und Tabelle 67 im Anhang). Im Mobilfunk ist im drahtlosen Zugangsnetz eine Asymmetrie der Leistungsfähigkeit festzuhalten. Durch aufwendigere stationäre Antennensysteme und höhere Sendeleistung ist eine höhere Leistungsfähigkeit in Richtung von der stationären Infrastruktur zum mobilen Endgerät (Down) als in Umkehrrichtung (Up) festzustellen. Entsprechend asymmetrisch fällt typischerweise auch die Ende-zu-Ende-Sicht aus.

5.1.3 Das Entstehen der Netzbedingungen und Abgrenzung zur E-Technik

In der Beschreibung des Mobilfunks sind neben den allgemeinen Merkmalen von Netzen auch die mobilen Aspekte von Bedeutung. Dabei ist zwischen Mobilität und drahtloser Charakteristik der Funkübertragung zu trennen. Mobilität bedeutet Handover-Prozeduren zwischen verschiedenen Zellen, also der Infrastruktur des Netzes als auch zwischen verschiedenen Technologien, z. B. Handover zwischen 2G und 3G. Eine detaillierte Erweiterung der Betrachtungen auf die drahtlose Charakteristik wäre prinzipiell möglich, um z. B. die Entstehung der relevanten Netzparameter selbst wiederum zu erklären, z. B. beim Mobilfunk für die Luftschnittstelle aus den Funk- und Verbindungseigenschaften.

Zusammengefasst sind insbesondere im zellulären Mobilfunk für die durch einen Nutzer erreichten Werte relevant (ohne Ausschluss weiterer Faktoren, u. a. nach [231], [232], [233]):

- Technologieausbau des genutzten Mobilfunknetzes, ganz konkret die realisierten Releases der genutzten Infrastrukturkomponenten des drahtlosen Zugangsnetzes (Basisstation), woraus sich die bestmöglichen Parameterwerte ergeben, siehe vorheriger Abschnitt
- Endgerät selbst (UE-Kategorie des Endgerätes, konkrete Ingenieursleistung des Herstellers beim Senden und Empfangen, sowohl der Hardware (Antennen) als auch der Software-Parametrisierung über den Stack)
- Sende- und Empfangsbedingungen des mobilen Endgerätes, die zur Wahl der Modulations- und Codier-Schemata führen (Entfernung zur Basisstation, Bebauung die zu Abschattungen oder Reflexionen führt, Interferenzniveau, Signal-Rausch-Abstand, Relativbewegung mit höheren Geschwindigkeiten in Auto oder Bahn)
- Shared Medium Funk: Anzahl gleichzeitiger Nutzer und Nutzungen, die zur nichtlinearen Aufteilung der Ressourcen führen (lokale Auslastung der Zelle in engem Zusammenhang mit der Zellgröße bzw. Netzverdichtung)
- Netzausbau im Netzzinneren, von der Anbindung der Basisstationen an das Kernnetz bis hin zu den Übergabepunkten aus dem Kernnetz ins Internet

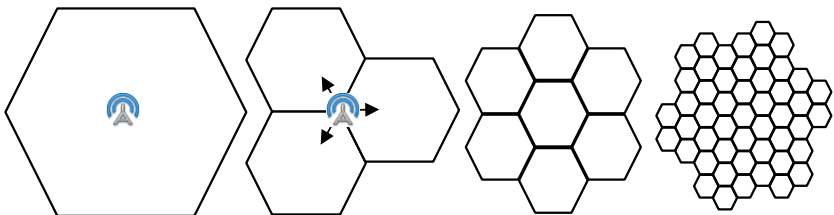


Abbildung 37: Steigerung der Kapazität durch Sektorisierung und Netzverdichtung

Zur Kapazitätssteigerung des Gesamtnetzes können Maßnahmen zur Netzverdichtung genutzt werden. Dies kann über eine Erhöhung der Anzahl Basisstationen und/oder Sektorisierung erreicht werden, siehe Abbildung 37. Jeder einzelne Standort kann dabei weiter in einzelne Sektoren untergliedert sein (z. B. 3 Sektoren mit je 120°-Versorgung). Durch die Netzverdichtung nimmt

die versorgte Fläche je Zelle ab. Damit sinkt potenziell die Zahl der mobilen Endgeräte pro Basisstation. Dies bedeutet, dass durchschnittlich mehr Ressourcen und damit Netzleistung innerhalb einer Zelle je Endgerät bereitstehen, während die durchschnittliche Entfernung zwischen Basisstation und Endgerät ebenfalls sinkt (potenzielle Verbesserung der Sende- und Empfangsbedingungen).

Auf eine detaillierte Darlegung, wie tatsächlich erreichte Werte der Praxis aus den theoretischen Werten der Technik entstehen, sowie auf Erklärungen zum Physical Layer wird nicht weiter eingegangen, da nicht im Fokus der Arbeit. Derartige Beschreibungen sind im Bereich der Elektrotechnik/Nachrichtentechnik besser aufgehoben. An dieser Stelle erfolgt damit die Abgrenzung, nach unten, hin zu den Tiefen der Technik. Es interessieren im Weiteren weniger das genaue Zustandekommen und die Ursache dessen, warum drahtlose mobile Netze so sind, wie sie sind, sondern nur die Feststellung und weitere Wirkung. Vertiefende Informationen für den interessierten Leser liefern beispielsweise [231] mit Erläuterungen zur „user data rate“ in Abhängigkeit der Empfangsbedingungen und Zellauslastung sowie [232] mit einer Brutto-Netto-Betrachtung auf den untersten Stack-Schichten am Beispiel von LTE (kurz zusammengefasst in [233]).

5.1.4 Notwendige Vereinfachungen zu den Netzsituationen

Für die weiteren Betrachtungen sind einige Vereinfachungen notwendig, die nachfolgend noch einmal kurz zusammengefasst erörtert werden sollen. Im genutzten Modell werden zur Verdeutlichung die einzelnen Netzabschnitte als Pfad durch das Netz durch aneinandergereihte Röhren⁹² symbolisiert, siehe Abbildung 38. Der Durchsatz als Analogie zur Bandbreite ergibt sich aus dem Querschnitt der Röhre, die Verzögerung aus der Länge der Röhre und die Paketverluste durch ein Leck. Der maximale Durchsatz wird so von dem Abschnitt mit dem geringsten Querschnitt vorgegeben und die minimale Paketlaufzeit aus der Summe der Längen der einzelnen Abschnitte gebildet.

⁹² In der Literatur werden Long/Fat/Smart/Dumb Pipes gerne als ein eingängiges Mittel zum Illustrieren von Netzen oder Netzabschnitten oder für einen Pfad durch ein Netz genutzt, siehe [21] S. 240.

Es wird im Weiteren davon ausgegangen, dass die Charakteristik der Netzsituationen hauptsächlich durch die Eigenschaften der Luftschnittstelle bzw. des drahtlosen Zugangnetzes (Netztechnologiestufe und Netzausbau) bestimmt und insbesondere limitiert ist. Dabei gibt die Technik, d. h. die Spezifikationen der Mobilfunkgenerationen/Generierungen, die bestmöglichen Werte vor. Abseits dieser vereinfachenden Annahme kann in der Realität der Flaschenhals auch an beliebiger anderer Stelle liegen: im Backhaul, also der Anbindung der Basisstationen an das Kernnetz, in einem Router im Kernnetz, in einem Übergangspunkt ins Internet oder an anderer Stelle im Internet bzw. sogar in den Endpunkten selbst.

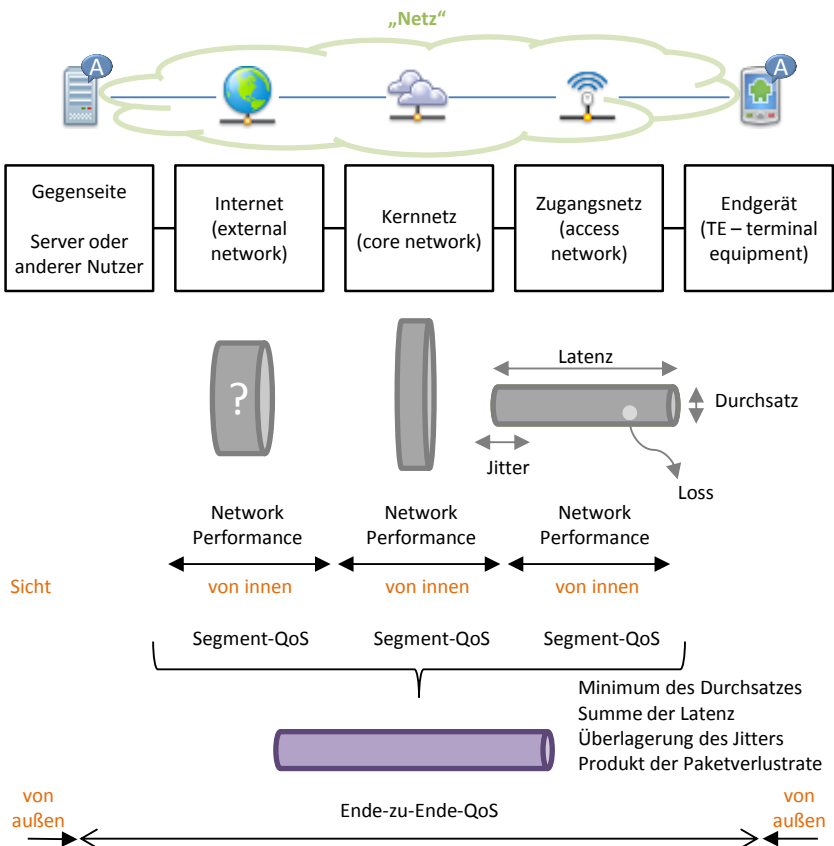


Abbildung 38: Schematische Darstellung der Netze und unterschiedliche Sichten auf „das Netz“ Ende-zu-Ende-QoS durch verschiedene Netzsegmente, zusammengestellt nach [35] S. 1, [46] S. 8 und [21] S. 240

Es wird die Sicht eines einzelnen Nutzers eingenommen. Die Netzsituation ist somit individuell und selbst nur Teil einer Gesamtnetzsituation (z. B. in der Zelle), die nicht weiter betrachtet wird (Netzkapazitätsplanung bzw. -messung).

Die Schritte zum Aufbau und Abbau einer Datenkommunikation sind im Mobilfunk sehr bedeutsam, weil relativ häufig, bedingt durch die Notwendigkeit zum Einsparen von Energie im inaktiven Status. Dabei sind die Phasen zum Aufbau vergleichsweise zeitintensiv. Bei 3G liegt die benötigte Zeit je nach Zustandswechsel im Bereich von Sekunden (ca. 2) und bei HSPA und LTE immerhin noch bei einigen Hundert Millisekunden ([164] Kapitel 7). Die Mechanismen inklusive der Statuswechsel der Funkschnittstellen der Endgeräte und die je Schritt benötigte Zeit beschreibt [164] Kapitel 7 sehr ausführlich und anschaulich. Es wird nachfolgend davon ausgegangen, dass sich das mobile Endgerät bereits mit dem Mobilfunknetz verbunden hat (attach) und in einem sende- bzw. empfangsbereiten Zustand (ready) befindet. Somit ist keine zusätzliche Verzögerung zum Aufbau der Mobilfunkverbindung zu betrachten (siehe Zustandsmodell in [228] S. 108ff). Ebenso entfällt die Phase des Abbaus in der weiteren Diskussion. Nachfolgend soll vorrangig auf die leistungsbezogenen Parameter der Nutzdatenübertragung Bezug genommen werden (vgl. Diskussion der Phasen in 2.2.6). Wie bereits erwähnt, bleiben die Mobilitätsaspekte (Handover etc.) ebenfalls außerhalb der Betrachtungen.

Weiterhin typisch für drahtlose Technologien - und so auch für Mobilnetze - sind dynamische Übertragungskanäle mit räumlich und zeitlich veränderlicher Charakteristik und die gemeinsame Nutzung durch mehrere Teilnehmer gleichzeitig in einer Zelle. Im Test hingegen wird zur Vereinfachung eine statische Betrachtung jeweils bestmöglicher Werte angesetzt, aus der sich die jeweiligen dynamischen QoS-Werte ergeben. QoS und QoD in der eigenen Terminologie können so vereinfacht auch als statischer Soll- und als dynamischer Istwert (Nutzerdatenrate unter Last) gesehen werden, für die durch entsprechende Abstufungen geeignete Wertkombinationen zu wählen sind. Dafür wiederum geben QoS-Erhebungen die passenden Hinweise.

5.1.5 Praxis: QoS-Erhebungen im Mobilfunk

Es ist selbst unter Idealbedingungen für die ermittelten Ende-zu-Ende-Sichten jeweils von leicht schlechteren Werten als den theoretischen Bestwerten auszugehen. Dies ist durch die Eigenschaften der weiteren Netzabschnitte und den Overhead der Protokolle auf den verschiedenen Schichten des Stacks sowie die Eigenheiten der Protokolle (wie bei TCP) zu begründen. Dabei sind reale Netzbedingungen selten ideal. Systematische Messungen in Mobilfunknetzen geben beste Hinweise dazu, was diese Netze wirklich leisten und welche technischen Parameter und Bereiche für weitere Untersuchungen interessant sind. Dabei ist nicht nur relevant, was dem Kunden als theoretische Leistung angeboten wird, sondern was tatsächlich als Ende-zu-Ende-QoS erreicht wird. Die Betonung liegt hierbei auf Ende-zu-Ende. Die erreichte Leistung wird dabei durch das Zusammenwirken aller beteiligten Elemente determiniert. Dafür gilt, dass die QoS-Erhebungen auf Basis von Endgeräten und entsprechenden Gegenseiten (Servern) eigentlich Ende-zu-Ende-Messungen auf der Applikationsschicht sind. Über die gemessene Kommunikationsleistung wird die verfügbare Netzleistung daher nur approximiert. Die Ergebnisse solcher Messungen lassen dann eine Aussage über die typische Leistungsfähigkeit der Luftschnittstelle als angenommener Flaschenhals zu, wenn sich dieser nicht an einer anderen Stelle der Ende-zu-Ende-Betrachtung befindet, z. B. zum lokalen Endgerät oder in überlasteten Servern, und der durch die Messung selbst verursachte Overhead berücksichtigt wird.

Mit fortschreitendem Netzausbau, Netzverdichtung, der Einführung neuer Mobilfunkgenerationen und moderneren Geräten haben sich die in der Praxis erreichten Werte tendenzmäßig immer weiter verbessert. Jede konkrete Wertangabe einer QoS-Erhebung ist daher immer nur eine Momentaufnahme. Die Dynamik räumlicher und zeitlicher Änderungen ist schwierig zu erfassen. Auf die ausführliche Angabe jeweils konkreter Werte im Detail wird an dieser Stelle verzichtet, sondern lediglich auf entsprechende Methoden zur Datensammlung und die entsprechenden Quellen verwiesen, die jeweils aktuelle Ergebnisse bereithalten. Für die eigenen Überlegungen waren diese Erhebungswerte nützliche Richtgrößen und Bestätigung der Überlegungen zu möglichen interessanten Bereichen.

Tabelle 24: Netztests in Mobilfunknetzen in Deutschland

Test	Methode	Umfang (Messungen)*	betrachtete Jahre / Veröffentlichungen**
Bundesnetzagentur	mobile Erfassung Messfahrten (Datensticks)	4 bis 5-stellige Anzahl Einzelmessungen (mobil)	2013/04 [234] (Zeitraum Mobilanteil: 2012/10-2012/11)
Chip***	mobile Erfassung Messfahrten/Begehungen (Messfahrzeug + Rucksacksystem zu Fuß/Zug)	2012: 3650 km Autostrecke 2013: 100000 Datenmessungen, 8400 km Autostrecke und 6000 km Zugstrecke	Ausgabe 9/2012 [235] Ausgabe 10/2013 [236]
Computer-Bild***	mobile Erfassung Crowd Sourcing Mess-App	2012: ca. 40000 Testteilnehmer 2013: ca. 83000 Testteilnehmer mit ca. 12 Millionen Einzelwerten	Ausgabe 25/2012 [237] Ausgabe 23/2013 [238]
Connect***	mobile Erfassung Messfahrten (Messfahrzeug)	2012: ca. 17000 km Rundstrecke 2013: ca. 40000 Dateiübertragungen	Ausgabe 12/2012 [239] Ausgabe 01/2014 [240]

Beschränkung auf Netztests mit Betrachtung in der Vor- und Nachbereitung der eigenen Studie
*) keine einheitlichen Angaben zum Umfang, daher lediglich zur Schätzung der Größenordnung geeignet

**) zum Zeitpunkt der Testplanung bzw. Erstellung der Arbeit vorliegende aktuelle Daten

***) Aktualisierungen ungefähr im Jahresrhythmus

Zur Ermittlung dieser Werte sind verschiedene Ansätze geeignet, auf die nachfolgend noch weiter eingegangen wird: Messkampagnen, Crowdsourcing-Datensammlungen und Real User Monitoring. Das Problem aller QoS-Erhebungen im Mobilfunk ist die Begründung der Repräsentativität als Umfang der Messungen und der Relevanz der erfassten Metriken. Deshalb kann nur eine Kombination aus Ergebnissen der vorgestellten Verfahren eine umfassende Einsicht gewähren. Eine zu einfache Betrachtung bzw. zu wenig hinterfragte Messungen können hingegen leicht fehlinterpretiert werden.

Messkampagnen: Die oft angeführten „Netztests“ sind eine Bestandsaufnahme dazu, was Mobilfunknetze verschiedener Anbieter leisten können. Die organisierten Messkampagnen in großem Umfang werden typischerweise

durch professionelle Dienstleister ausgeführt, entweder im Auftrag der Netzbetreiber selbst oder beauftragt durch Regulierungsbehörden bzw. Fachpresse. Dabei kommen entweder mobile Systeme oder eine Reihe von stationären Messanlagen zum Einsatz. Mobile Systeme in Fahrzeuge zur Durchführung von Messfahrten erlauben die Erfassung entlang von Verkehrsstrecken. Portable Systeme in Rucksäcken ermöglichen Messungen in öffentlichen Verkehrsmitteln, in Fußgängerzonen oder anderen urbanen Anlaufpunkten. Die mobilen Systeme liefern nur räumliche und zeitliche Ausschnitte. Entsprechend sind die Aussagekraft und Verallgemeinerbarkeit vom Umfang sowie der Zusammenstellung und Gewichtung der Einzelwerte abhängig. Stationäre Systeme liefern Daten für längere Zeiträume und erfassen tageszeitliche Schwankungen, liefern dafür aber nur räumlich punktuelle Ergebnisse. Relevante Netztestkampagnen, vorrangig auf Deutschland bezogen, sind in Tabelle 24 aufgeführt. Der Abschlussbericht der Messkampagne im Auftrag der Bundesnetzagentur zur Messung mobiler Breitbandleistung schränkt die Aussagekraft zu kleiner Messkampagnen ein (siehe [234] S. 107), d. h. nur mit großem Aufwand sind Aussagen zu generieren, die einen räumlich und zeitlich verallgemeinerbaren Rückschluss zulassen.

Crowdsourcing-Datensammlungen: Beim Crowdsourcing tragen nicht wenige professionelle Anbieter die Daten zusammen, sondern eine Vielzahl von (freiwilligen) Nutzern. Diese führen (sporadische) Messungen mit ihren üblichen mobilen Endgeräten und speziellen Mess-Apps an verschiedenen Orten zu unterschiedlichen Zeiten aus. Bei ausreichend großer Beteiligung entstehen so riesige Mengen an einzelnen Messungen.

Beispiele für Crowdsourcing-Netzmessungen mit Mess-Apps und zumindest teilweise (aggregiert) zugänglicher Datenbasis oder veröffentlichten Endergebnissen sind beispielsweise:

- ComputerBild Netztest App⁹³ (in Kooperation mit RadioOpt⁹⁴)
- OpenSignal⁹⁵ (vormals OpenSignalMap)

⁹³ <http://www.computerbild.de/artikel/cb-Tests-Handy-Mobilfunknetztest-8926120.html> abgerufen am 30.07.2014

⁹⁴ <https://www.radioopt.com/> abgerufen am 26.02.2015

⁹⁵ <http://opensignal.com/> abgerufen am 02.02.2014

- Ookla Netindex⁹⁶ (Speedtest.net)
- Cisco Visual Network Index (VNI) GIST⁹⁷
- Floq⁹⁸ (Kooperation mit Chip als Chip-Netztest-App⁹⁹)

Real User Monitoring (RUM): Im Gegensatz zum zuvor beschriebenen Ansatz liefern beim Real User Monitoring zwar auch normale Nutzer die einzelnen Datensätze, dafür sind aber keine gesonderten Messungen notwendig. Die Daten ergeben sich aus der Überwachung der regulären Nutzung echter Dienste durch echte Nutzer. Dabei sind verschiedene technische Realisierungen denkbar:

- entweder auf dem mobilen Endgerät, z. B. im mobilen Browser des Nutzers durch JavaScript-Code (mit Messung der Ladezeit bestimmter Elemente oder zum vollständigen Laden der Seite)
- Erzeugung von Protokoll-Aufzeichnungen auf Server-Seite
- Überwachung in Netzelementen

Die großen Betreiber von Content Distribution Networks (CDN) haben genügend Daten für Abrufe von Inhalten von ihren Servern, die sie zumindest grob einer mobilen Nutzung zuordnen können. Ein Beispiel für Real User Monitoring Daten ist der Akamai-Report *The State of the Internet*¹⁰⁰. Im Gegensatz zu sämtlichen bisher vorgestellten Ansätzen, die von einer Seite der Ende-zu-Ende-Übertragung zur anderen messen, ist auch eine Beobachtung im Netz möglich, z. B. durch Deep Packet Inspection und Performance Monitoring der beteiligten Netzelemente und Netzsegmente. Da diese Art der Messung aber quasi von innen heraus erfolgt, bleibt sie entsprechend denjenigen vorbehalten, die solche Messpunkte auch nutzen können, d. h. den Netzbetreibern. Diese geben die Daten nur teilweise über Partner nach außen. Dazu gehört beispielsweise der Sandvine-Report *Global Internet Phenomena*¹⁰¹ für den mobilen und stationären Netzzugang.

⁹⁶ <http://www.netindex.com/> abgerufen am 17.02.2014

⁹⁷ <http://gistdata.ciscovni.com/> abgerufen am 02.02.2014

⁹⁸ <http://www.floq.net/> abgerufen am 02.02.2014

⁹⁹ http://www.chip.de/artikel/CHIP-Netztest-App-fuer-Android-So-testen-Sie-Ihr-Handy-Netz_68324572.html abgerufen am 12.03.2014

¹⁰⁰ <http://www.akamai.com/stateoftheinternet/index.html> abgerufen am 02.02.2014, regelmäßig aktualisiert

¹⁰¹ http://www.sandvine.com/news/global_broadband_trends.asp abgerufen am 24.03.2013

5.1.6 Struktur und Festsetzung der Matrix technischer Parameter

Für die Nutzung im Labortest wurden sinnvolle Bereiche und Kombinationen der drei interessierenden QoS-Netzparameter in einer Matrix der QoS-Parameter festgelegt, siehe Abbildung 39. Für die Datenraten wurde eine exponentielle Abstufung mit annähernd konstantem Stufenfaktor von ungefähr $\sqrt{2}$ gewählt (abgesehen von den Anfangswerten). Dies bedeutet, dass jeweils 2 Stufen einer Verdoppelung des Ausgangswertes entsprechen. Die Art der Abstufung kommt der später beschriebenen Auswahl von Stufen zur Parametrisierung des Netzes für konkrete Dienste im Test entgegen, für die jeweils ausreichend große Stufenabstände und gleiche Faktoren sinnvoll sind. Die Hauptstufen orientieren sich an der angegebenen Technologie und der technologisch bedingten Bestleistung. Die weiteren Zwischenstufen ergeben sich aus Abstufungen von Fehlerkorrekturdaten zu Nutzdaten und Modulationsparametern aus den Spezifikationen, den RAN-Kategorien der Endgeräte und aus den typischerweise erreichten Zwischenwerten in den QoS-Erhebungen.

Aus den Abstufungen der einzelnen Parameter wurden sinnvolle Wertkombinationen aus Download-Datenrate, Upload-Datenrate und Latenz eingegrenzt. Als unteres Ende wurde GPRS gewählt, als oberes Ende HSPA. Die ausgewählten Bereiche ergeben sich dabei aus dem als sinnvoll angesehen Verhältnis der Datenraten im Downstream zum Upstream (knapp unter 20 % bis ca. 80 %, siehe [241] S. 97). Für das Verhältnis von Upload-Wert zu Download-Wert sollen zur Vereinfachung die Annahmen gelten, dass die Upload-Datenrate die Download-Datenrate nicht übersteigt und nach unten ein weitgehend konstanter Minimalwert erreicht wird. Damit ergeben sich die eingezeichneten dreiecksförmigen Bereiche: für Technologien bis UMTS Release 99 ein Bereich links unten und für HS(D/U)PA ein Bereich rechts oben. Die Latenz wird als typische (technologisch begründete) Grundlatenz (ohne Last) für die jeweilige Ausbaustufe festgelegt und ist in der Matrix als Zellwert abzulesen. Andere QoS-Netzparameter wie Paketverlustrate und Jitter blieben unberücksichtigt, da für die angestrebten Untersuchungen zur Parametervariation nur von untergeordneter Relevanz.

Neben den technologisch bedingten Wertkombinationen ist auch ein absichtlich leistungsmäßig gedrosselter Fall vorgesehen. Diese Drosselung, im Fachjargon „Speed Step Down“, ist in fast allen (Flatrate-) Tarifangeboten üblich.

In diesen Angeboten erfolgt die künstliche Drosselung durch Begrenzung der maximalen Übertragungsgeschwindigkeit bei Überschreiten eines bestimmten Datenvolumens. Die genaue Wertkombination ist abhängig vom jeweiligen Anbieter und den Vertragsbedingungen. Typischerweise erfolgt eine Begrenzung (Policing) „auf GPRS-Geschwindigkeit“, also zwischen 32 bis 64 kBit/s im Downstream und 16 bis 64 kBit/s im Upstream. Für die Grundlatenz können weiter die üblichen Werte der jeweils genutzten Mobilfunktechnologie angenommen werden (hier HSUPA-Ausbau). Ansonsten war schon im Jahr 2012 für die Mobilfunknetze in Deutschland mindestens von einem Ausbau mit EDGE auszugehen [230] S. 4. Die üblicherweise erreichten Werte deuten sogar auf einen weiträumigen HSPA-Ausbau hin, siehe eingezeichnete Werte in Abbildung 39.

Ein Blick zum Abgleich zwischen den gebildeten Stufen und typischen Messwerten der Jahre 2012 und 2013 zeigt, dass eine Orientierung am jeweils aktuellen Stand der Netzentwicklung sinnvoll ist. Lediglich sehr hohe Upload-Raten (>2 MBit/s) waren zum Zeitpunkt der Testplanung noch nicht als typisch anzusehen. Für die Parametrisierung waren entsprechend vorrangig die Wertkombinationen interessant, die die obere Grenze der eingezeichneten Bereiche bilden, bzw. ein Plateau knapp unter 2 MBit/s Upload. Die Auswahl und Zuweisung konkreter Wertkombinationen erfolgte dienstspezifisch. Weitere Ausführungen zum Prinzip der Auswahl der Stufen finden sich in der Diskussion der Dienste (6.4), die Begründung zur konkreten Parametrisierung der einzelnen Dienste im Nutzertest im entsprechenden Abschnitt des Testparcours.

Erläuterungen zur Abbildung 39:

Zusammenstellung entsprechend der repräsentierten Mobilfunktechnologien in der Studie max. Datenrate Download (x-Achse), max. Datenrate Upload (y-Achse), min. Latenz (in den einzelnen Zellen)

a) + b) Vergleich der technologieorientierten Netzparameterstufen mit tatsächlichen Erhebungswerten aus jeweils zwei verschiedenen Studien der Jahre 2012 und 2013 aus Mobilfunknetzen zusammengestellt aus [237], [238], [239], [240].

Zu beachten ist, dass die eingetragenen Datenpunkte verschiedene Kennziffern mit unterschiedlichen Aussagen zeigen:

links a) die 10. Perzentile, d. h. 90 % schneller als...; rechts b) Position: Mittelwerte, numerischer Wert: „Ping“

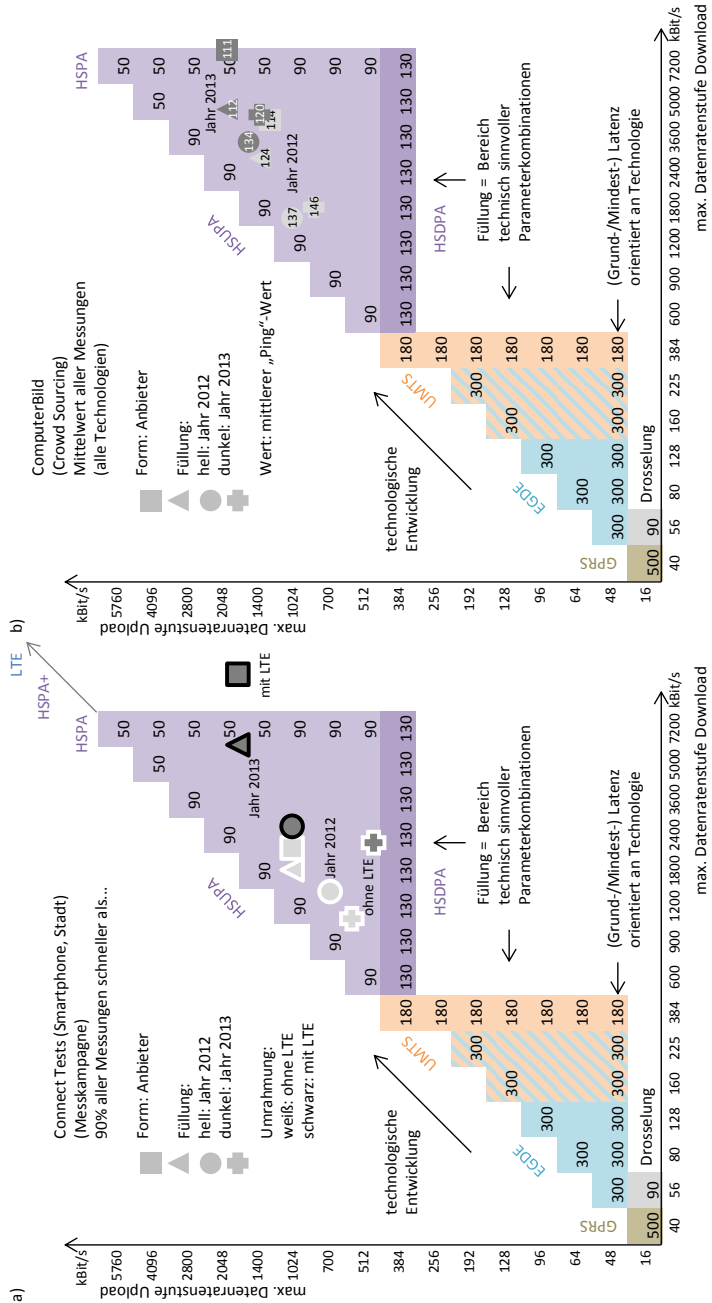


Abbildung 39: Netzparameterstufen als 3-Tupel in der Matrix technischer Parameter

Die Festlegungen zu den einschätzten Bereichen, Abstufungen und Wertkombinationen können dann als adäquat angesehen werden, wenn sie weder *zu schlecht* noch *zu gut* sind.

zu schlecht: Zwar nimmt die theoretische Spitzenleistung mit jeder Mobilfunkgeneration zu, gleichzeitig steigt aber auch die Zellauslastung immer weiter an, so dass die mittleren und typischen Werte weit weniger schnell steigen. Lediglich unter Beachtung von LTE lagen realistisch erreichbare Werte (im Download) leicht außerhalb des betrachteten Parameterbereiches. Die Mehrzahl der mobilen Endgeräte war zum Zeitpunkt der Testplanung aber noch nicht LTE-fähig, was die tatsächliche Relevanz der Werte außerhalb der Untersuchungsgrenzen wieder relativiert.

zu gut: Die genannten Netztests belegen gerade vom Jahr 2012 zu 2013 erhebliche Verbesserungen durch Netzaufrüstungen. Die Richtung der Entwicklung ist klar erkennbar, daher ist davon auszugehen, dass der Ansatz der angenommenen Werte zur Charakterisierung der Leistungsfähigkeit gut mit den praktisch erreichbaren Werten korrespondierte. Hinsichtlich der Latenz wird noch erläutert, dass sich die Werte durch die externen Netzabschnitte des Testbeds und die Wirkung der dynamischen Latenz während der Messung den mittleren Erhebungswerten annähern. Die Entwicklung der tatsächlich erreichten Upload-Datenraten verläuft im Vergleich zum Download deutlich langsamer. Damit sind aber lediglich die besonders hohen Parameterwerte für den Upload in der festgelegten Bereichsmatrix etwas zu gut.

5.2 Die Einbeziehung des Netzes als Qualitätselement in der Studie

Aus den Anforderungen heraus sollen die möglichen technischen Konzepte zur Einbeziehung des Netzes als Qualitätselement in der Studie zunächst an diskutiert werden, bevor der gewählte Kombinationsansatz kurz umrissen wird. Dabei wird erläutert, warum eine Netzemulation für den angestrebten Zweck als geeignet angesehen wird und warum eine Kombination mit einem Live-Netz notwendig ist.

5.2.1 Definition der Anforderungen an das Netz

Aus dem Konzept zur Arbeit ergaben sich die folgenden Anforderungen, die für eine empirische Untersuchung in der technischen Realisierung zu berücksichtigen waren. Dies sind einerseits Anforderungen an „das Netz“ aber auch das Testbed im Ganzen. Dabei sind insbesondere die eingangs beschriebenen allgemeinen Anforderungen wie Echtheit und Skalierbarkeit zu beachten.

Echtheit/Realismus: Das Netz sollte in seinen Eigenschaften möglichst echt sein. Die effektiven Netzsituationen sollten möglichst so sein, wie sie sich auch in einem echten Mobilfunknetz ergeben könnten. Dazu ist eine genaue Kontrolle der relevanten Netzparameter notwendig. Entsprechend sind auch Abschätzungen dazu sinnvoll.

Kontrollierbarkeit/Funktionalität: Hauptzweck des Testbeds ist die Ermöglichung der Untersuchung der Wirkung verschiedener Netzsituationen (QoS) auf die Nutzerzufriedenheit (QoE). Dafür sind grundsätzlich folgende Ansätze möglich:

- die Netzsituationen ergeben sich unbeeinflusst „irgendwie“, ihre Parameterausprägungen werden lediglich festgestellt und den sich ergebenden Nutzerbewertungen gegenübergestellt (passive Kontrolle)
- die Netzsituationen werden gezielt hergestellt, ihre Parameterausprägungen werden geplant und systematisch variiert (aktive Kontrolle)

Die Beschränkung auf eine Erfassung des Zustands der Netzsituation und der dazugehörigen Nutzerbewertung ist nicht ausreichend. Für eine effiziente Testausführung als Experiment wird eine gezielte aktive Kontrolle über die Netzparameter als unabhängige Variablen angestrebt. Für die Durchführung der Studie sollen sich die Netzbedingungen daher möglichst flexibel, schnell und genau herstellen und variieren lassen. Das Testbed muss den Netzzugang geeignet realisieren.

Skalierbarkeit: Der Ansatz und die technische Lösung zur Einbeziehung des Netzes in der Studie müssen geeignet sein, um potenziell hunderte Probanden mit tausenden Aufgabenausführungen mit jeweils wechselnder Parametrisierung innerhalb überschaubarer Zeit testen zu können. Dies bedeutet, dass möglichst viele Probanden parallel getestet werden sollten.

Praxistauglichkeit: Eine Lösung der Art „ja ist möglich und funktioniert prinzipiell“ ist für einen Testumfang ab einer gewissen Größe nicht mehr ausreichend. Es bedarf eines pragmatischen Ansatzes und einer praktikablen Lösung, die manuelle Eingriffe minimiert. Im laufenden Testbetrieb muss das System auch von Nicht-Entwicklern zu bedienen sein. Dies bedeutet neben der Erfüllung der generellen Anforderungen auch ein gewisses Grundniveau an Benutzerfreundlichkeit und Robustheit sowie eine weitgehende Automatisierung des Testbetriebs.

5.2.2 Simulation

Die Unterscheidung von Simulation und Emulation ist auch oftmals auch eine Betrachtung zwischen (rein) Software-basierten Ansätzen in der Simulation und Hardware-integrierenden Ansätzen in der Emulation. Mangels Erfüllung der Anforderungen an die Echtheit der beteiligten übrigen Elemente ergibt sich, dass eine Simulation als alleinige Lösung nicht ausreicht. Die Stärken einer Simulation liegen in anderen Bereichen: als zusätzlicher oder unterstützender Baustein, um ausgewählte Teilzusammenhänge zu modellieren und zu untersuchen. Einsatzmöglichkeiten dafür wären Erklärungen zur Entstehung der Stimuli aus den Effektivwerten heraus sowie zum Zustandekommen der Effektivwerte der Netzsituationen aus dem Zusammenspiel verschiedener Netzelemente und Protokollschichten. Für eine Simulation, also die Nachbildung verschiedener Netzcharakteristiken ausschließlich in Software, spricht die schnelle und gezielte programmatische Beeinflussung relevanter Parameter. Typischerweise genutzte Simulatoren für Netze und Protokolle sind z. B. ns2/ns3¹⁰², Omnet++¹⁰³ und NetSim¹⁰⁴. Im Rahmen der weiteren Ausführungen in der vorliegenden Arbeit spielt die Simulation keine Rolle. Der interessierte Leser findet umfassende Informationen zur Netzsimulation inklusiver drahtloser und mobiler Aspekte (sowie zur Simulation auf der Applikationsschicht) in [242].

¹⁰² <http://www.isi.edu/nsnam/ns/> abgerufen am 10.04.2013

¹⁰³ <http://www.omnetpp.org/> abgerufen am 10.04.2013

¹⁰⁴ <http://tetcos.com/> abgerufen am 10.04.2013

5.2.3 Emulation

In einer Emulation lassen sich verschiedene Elemente durch andere Elemente in Teilen ihrer Eigenschaften nachbilden und mit anderen echten Teilen zusammen nutzen. Mit Blick auf verschiedene Emulationsansätze ist es sinnvoll zu unterscheiden, welche Elemente wie nachgebildet werden und welche echt bleiben. Eine Vollemulation bildet alle relevanten Elemente innerhalb einer geschlossenen Testumgebung nach. Dies bedeutet einen hohen Grad an Kontrolle und Ausschaltung von möglichen Störfaktoren, allerdings teilweise auch erheblichen Aufwand in Bezug auf die realistische Nachbildung der Elemente, siehe entsprechender Unterschnitt in der Diskussion der Einbeziehung der Dienste in die Studie. Aus den Anforderungen ergibt sich, dass es bei einer Emulation des Netzes bzw. der mobilen Leistungscharakteristik bleiben sollte, d. h. nur das Element *Netz* wird (in Teilen) emuliert. In einer Netzemulation lassen sich die mobilen Netzsituationen geeignet mit ausreichendem Realismus und Kontrolle nachbilden.

5.2.4 Echter Mobilfunk

Die Nutzung echter Mobilfunktechnik scheint für Tests mit mobilen Services und mobilen Endgeräten in verschiedenen (Mobilfunk-) Netzsituationen mit Blick auf die Realitätsnähe die erste Wahl zu sein. Prinzipiell bieten moderne Mobilfunknetze die Möglichkeiten zur gezielten Beeinflussung relevanter Netzparameter. Dies erfordert aber den Zugriff auf verschiedenste Komponenten, ggf. auch des Kernnetzes (PCRF), um z. B. Bandbreitenbegrenzungen zu erzwingen. Dies ist im Normalfall nicht möglich. Prinzipiell wird zwischen produktiven Wirknetzen und Testnetzen auf Basis verschiedener Mobilfunkgenerationen unterschieden. Ein geeignetes Testnetz mit entsprechenden Zugriffs- und Eingriffsmöglichkeiten stand leider nicht zur Verfügung. Im Rahmen der Nutzung von Wirknetzen sind die benötigten Möglichkeiten der aktiven Einflussnahme und damit der flexiblen Parametrisierung einzelner Nutzer, wie sie für eine effiziente Testdurchführung notwendig sind, praktisch nicht gegeben. Die Generierung von Aussagen zwischen technischen Parametern und Nutzerbewertungen ist dann nur über beobachtende Messungen von sich zufällig ergebenden Parameterkombinationen (Fluktuationen) möglich. Dies ist ggf. für eine Feldstudie akzeptabel, erfüllt aber nicht die Anforderungen eines Experiments (gezielte Kontrolle der Bedingungen). Entsprechend

mussten Konzepte auf Basis von Mobilfunktechnologie als ungeeignet zurückgewiesen werden. Trotzdem spielt die Nutzung echter Mobilfunknetze im Rahmen der Studie eine wichtige Rolle. Das Verhalten der Endgeräte und Apps im Mobilfunknetz wurde in den Voruntersuchungen analysiert und mit der Nutzung von WLAN verglichen. Für das Verständnis der Netznutzung im Mobilfunk und die Beobachtung der erzeugten Effekte, ausgehend von den erreichten Effektivwerten, ist eine passive Kontrolle (Monitoring) dabei ausreichend.

5.2.5 Live-Netz und drahtlose Konnektivität

Live-Netz bedeutet in diesem Zusammenhang die Einbeziehung externer Ressourcen außerhalb des Testbeds, konkret im Internet. Die Nutzung des „echten“ Internets erlaubt in der Studie echte Apps, echte Inhalte und echte Gegenseiten zu integrieren, birgt aber natürlich auch einige Risiken, z. B. die Wirkung als Störfaktor, da sich diese Elemente einer aktiven Kontrolle entziehen. Als Alternative verbleibt damit nur die passive Kontrolle, um extreme Abweichungen von geplanten Erwartungswerten detektieren und isolieren zu können. Dies erschien als gangbarer Weg.

WLAN-Hardware nach IEEE 802.11a/b/g/n (Wi-Fi) ist einfach verfügbar und preiswert. Moderne Smartphones integrieren ohnehin leistungsfähige WLAN-Schnittstellen. Kleine Testaufbauten für Laborumgebungen sind mit WLAN-Technik leicht realisierbar. Daher wurde WLAN als Technologie für die drahtlose Anbindung der mobilen Endgeräte ausgewählt. Das WLAN selbst ist dabei aber nicht mit der Netzemulation gleichzusetzen, sondern tatsächlich nur die Verbindung hin zu den Endgeräten.

5.2.6 Eigener Ansatz

Kombinationsansätze können die Vorteile der zuvor kurz vorgestellten Ansätze vereinen. Dabei sind unterschiedliche Kombinationen möglich. Der eigene Kombinationsansatz ergab ein Testbed mit folgenden Eigenschaften:

- Anbindung an das Internet zur Ermöglichung echter Dienste (echtes Live-Netz)
- gezielte aktive Beeinflussung durch Netzemulation (aktive Kontrolle)
- drahtloser Konnektivität auf Basis von WLAN zur Nutzung echter mobiler Endgeräte

Die zusätzliche passive Kontrolle (Monitoring) gibt die Möglichkeit, im Nachhinein die Beobachtungen des Nutzerverhaltens und der tatsächlichen technischen Parameter abzugleichen, bzw. den Erfolg der beabsichtigten Parametrisierung zu kontrollieren. Zudem erlaubt diese Funktionalität, die Untersuchung der Zusammenhänge und Modellbildung zu den Einflussgrößen der Nebenfaktoren durch die „echten“ Elemente im Test zu erfassen und nachfolgend zu berücksichtigen. Der gewählte Kombinationsansatz sollte gut genug sein, um den Zusammenhang ausgewählter QoS-Parameter und den QoE-Auswirkungen bzw. Nutzerbewertungen zu untersuchen. Die Entwicklung der Theorie und des Testbeds im zeitlichen Verlauf der Projektbearbeitung, sowie die Verbindung beider Bestandteile zeigt sich im direkten Vergleich eines ersten Konzeptentwurfs dazu (siehe Abbildung 40) und der letztlich realisierten Einbeziehung des Netzes zusammen mit den anderen Elementen im entwickelten QoX-Gesamtmodell (siehe Abbildung 41).

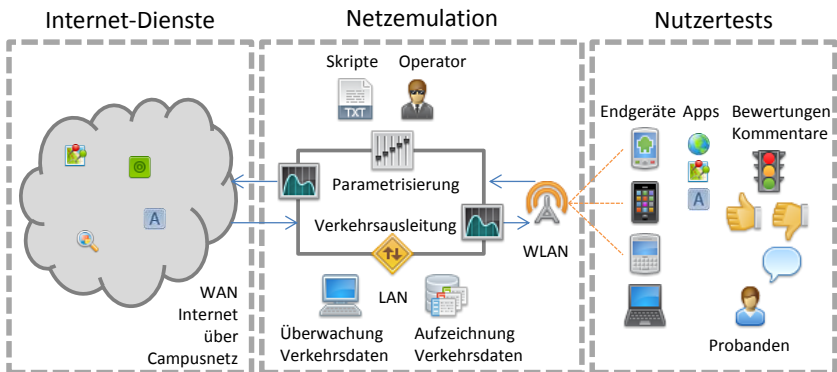


Abbildung 40: Skizze zur Einordnung der Netzemulation in das technische Grundkonzept

Ausgehend von dieser ersten Skizze entstand die technische Realisierung und Testplanung beginnend mit der Testrunde 1.

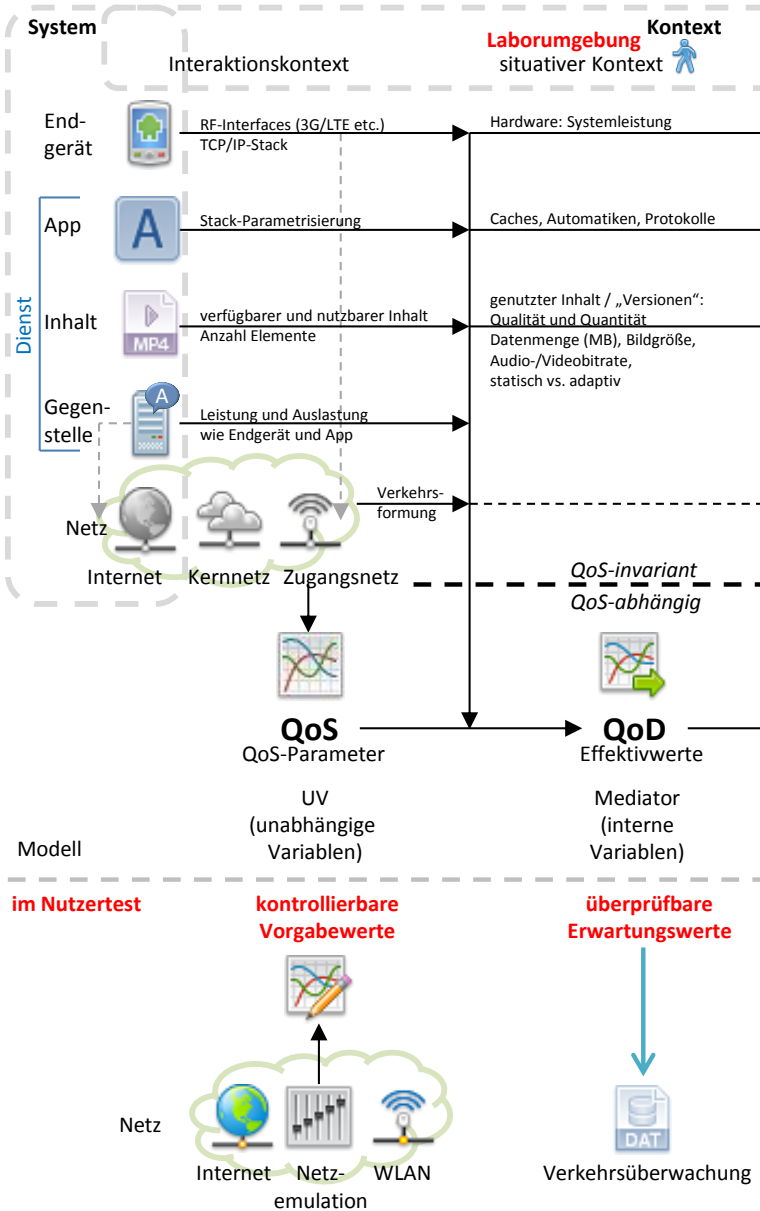
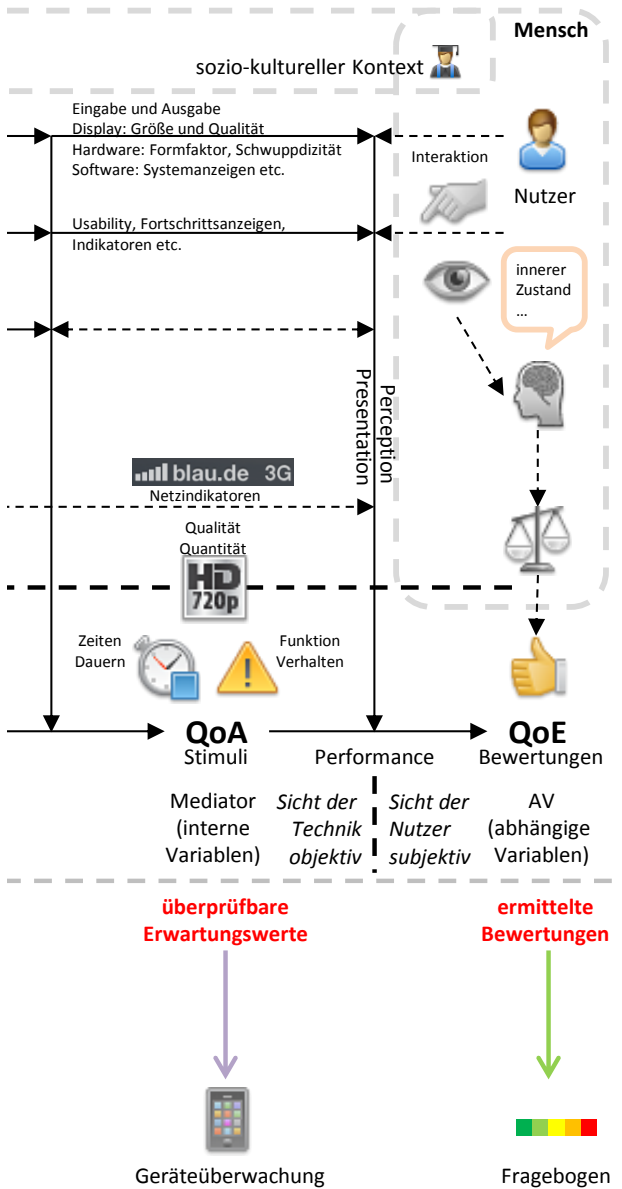


Abbildung 41: Einordnung des Testbeds mit Netzemulation in das QoX-Matrix-Modell
 in Rot: Einordnung der empirischen Untersuchung ins QoX-Matrix-Modell,
 der Rest bleibt „echt“
 Fortsetzung der Abbildung auf der nächsten Seite



Fortsetzung der Abbildung 41

oben: Kontext einer Laborumgebung und unmittelbare Sicht auf das Netz

unten: von der Parametrisierung des Netzes mit der Netzemulation bis hin zur Ergebniserfassung im Fragebogen

5.3 Theorie und verfügbare Lösungen zur Netzemulation

Für die weiteren Betrachtungen ist festzulegen, wie und wo die Netzemulation erfolgen soll. Davon ausgehend wird der Ansatz eines perfekten Netzes mit gezielten Verschlechterungen beschrieben. Es wird abschließend diskutiert, warum die Netzemulation eigentlich eine Link-Emulation ist und wo dabei die theoretischen Grenzen liegen.

Nach der Diskussion zu den theoretischen Ansätzen folgt eine kurze Vorstellung der Ergebnisse zur Recherche tatsächlich verfügbarer Ansätze und Lösungen zur praktischen Realisierung einer Netzemulation, wie sie innerhalb des Testbeds zum Einsatz kommen sollte.

5.3.1 Abgrenzung im Stack

Zur Abgrenzung der Netzemulation ist eine Orientierung am OSI/ISO-Schichtenmodell sinnvoll (siehe Abbildung 14a und [112] S. 76ff für Erläuterungen dazu). Zur Nomenklatur sei an die üblichen Bezeichnungen erinnert: Im Schichtmodell wird von Schicht 1, der physikalischen Bitübertragung (Layer 1, kurz L1), bis zur Schicht 7 der Anwendungen (Layer 7, kurz L7) unterschieden und nummeriert.

Abgrenzung nach oben im Stack: Einfache Ansätze für eine Netzemulation setzen auf Anwendungsschicht L7 an. Diese Systeme sind oftmals selbst als Anwendungsprogramme in Form von Proxy-Lösungen realisiert – im Unterschied zu komplexeren Emulatoren, die zum Teil tief in die Kerne der jeweiligen (Betriebs-) Systeme verankert sind. Die L7-Lösungen sind sehr einfach. Allerdings ist dies auch gleichzeitig ihr Hauptnachteil, da mit erheblichen Einschränkungen hinsichtlich des angestrebten Einsatzes verbunden:

- Beschränkung auf nur ein oder wenige L7-Protokolle (z. B. nur HTTP)
- nicht uneingeschränkt transparent aus Sicht der Anwendungen und des eigentlichen Datentransports (Ende-zu-Ende)
- für einige Protokolle gar nicht geeignet (proprietär oder verschlüsselt)
- zusätzliche Konfiguration auf Endgeräten notwendig (z. B. Proxy-Konfiguration, Zertifikate für verschlüsselte Kommunikation etc.)
- die Differenzierung was und wie emuliert werden soll, ist u. U. nur sehr eingeschränkt möglich

Für einzelne Anwendungen, die ausschließlich auf einem populären Protokoll wie HTTP basieren, kann eine solche Lösung ggf. durchaus genügen. Ein Beispiel ist die Throttling-Funktion im Charles Proxy¹⁰⁵ zum Abbremsen der HTTP-Kommunikation, mit der sich Web-basierte Anwendungen einfach testen lassen. Andere Protokolle werden hingegen nicht ausreichend berücksichtigt, z. B. DNS¹⁰⁶-Abfragen oder komplexe Anwendungen, die mehrere Protokolle (ggf. auch gleichzeitig) nutzen. Für die gewünschte generische Nachbildung der Netzsituationen für eine Vielzahl verschiedener Dienste ist ein solcher Ansatz auf Ebene der Anwendungsprotokolle daher nicht geeignet, da dienst- und protokollabhängig.

Abgrenzung nach unten im Stack: Bei der Nachbildung drahtloser mobiler Netze kann es sinnvoll sein, nach zwei Hauptpunkten zu unterscheiden: der drahtlosen Charakteristik und der Mobilität. Die Mobilitätsaspekte sind für die weiteren Überlegungen nicht von Relevanz und bleiben in der eigenen angestrebten Lösung unberücksichtigt. Auch das eigentliche Zustandekommen der drahtlosen Charakteristik eines Funkkanals wird abstrahiert. Eine Nachbildung solcher Parameter wird als Channel Emulation bezeichnet, für die spezielle (kommerzielle) Laborgeräte verfügbar sind, siehe z. B. [243]. Ein solcher Ansatz wäre deutlich zu komplex bzw. der Zielstellung nicht angemessen.

Für die zu untersuchenden Services sind aus Netz- und Nutzersicht die konkreten Eigenschaften der Schicht L1 gar nicht unmittelbar relevant. Interessant ist für diese Sichten nur, wie sich dies auf die höheren Schichten zur Realisierung von IP-basierten mobilen Datendiensten auswirkt. Es sollen also nicht wirklich echte Netzsituationen im physikalischen Sinne einer drahtlosen Charakteristik erreicht werden, noch soll versucht werden, mit einer Substitutionstechnik eine vergleichbare physikalische Parametrisierung zu emulieren. Die Emulation umfasst daher nicht die Nachbildung der Bestandteile des mobilen Netzes selbst bzw. seiner Technologien. Die Netzsituationen des Mobilfunks werden lediglich als theoretische Bestwerte mit praktisch erreichten Effektivwerten berücksichtigt. Dabei wird entsprechend des Schichtmodells davon

¹⁰⁵ <http://www.charlesproxy.com/> und <http://www.charlesproxy.com/documentation/proxying/throttling>
abgerufen am 28.11.2012

¹⁰⁶ Domain Name System zur Auflösung von Host-Namen in IP-Adressen

ausgegangen, dass die Effektivwerte für die Erzeugung der Nutzerstimuli entscheidend sind. Dynamische und kurzzeitige Änderungen der Netzcharakteristik des Mobilfunks können zwar relevant sein, bleiben aber in den weiteren Ausführungen zunächst unberücksichtigt.

L2/L3-Netzemulation: Für die Nachbildung der relevanten Charakteristiken der IP-Konnektivität ist ein Ansatz auf Schicht L3 (IP-Pakete) prinzipiell ausreichend für die Nachbildung der relevanten Ende-zu-Ende-QoS und der damit verbundenen Netzeigenschaften. Einige der nachfolgend betrachteten konkreten Systeme können sogar auf Schicht L2 (Ethernet-Frames) arbeiten, gehen also noch etwas über die Anforderungen hinaus. Egal ob L3 oder L2, die Parametrisierung erfolgt oberhalb der Bitübertragungsschicht (Physical Layer, L1), also unabhängig von der tatsächlich genutzten Übertragungstechnologie bzw. deren Parametrierung und unterhalb der Protokolle der Transport- und Anwendungsschicht, also unabhängig von konkreten Diensten und deren Protokollen. Damit kann die Betrachtung zur Netzemulation auf den Transport von entsprechenden Frames durch Netzsegmente bzw. die Weiterleitung von IP-Paketen zwischen den Netzsegmenten reduziert werden.

5.3.2 Perfektes Netz mit kontrolliertem Flaschenhals

Der ausgewählte Ansatz zur Netzemulation funktioniert als eine Art kontrollierter Flaschenhals. Die Netzemulation entspricht damit im Pfadmodell der aneinandergereihten Röhren quasi einem Rohrabschnitt „aus Gummi“, als eine Art Zwischenstück, das sich beliebig in Länge und Durchmesser parametrisieren lässt, siehe Abbildung 42.

Damit die Netzemulation tatsächlich der künstliche Flaschenhals ist, müssen die anderen Netzabschnitte entsprechend eine (deutlich) höhere Bandbreite (Querschnitt) und niedrige Verzögerungen (Länge) aufweisen. Für WAN und WLAN im Beispiel des Testbeds sind diese Voraussetzungen einer idealen Testumgebung für die Netzemulation gegeben. Für das „offene“ Internet bzw. die Netzanbindung und Netzauslastung auf Seite der Dienstanbieter im Internet lassen sich a priori keine solchen Aussagen treffen. Für die angestrebten Realweltszenarien ist dies aber hinzunehmen und im Rahmen von Voruntersuchungen abzuklären.

Aus der Positionierung des Flaschenhals ergeben sich einige Unterschiede zwischen echtem Mobilfunk und der Emulation, die zu beachten sind. Im Mobilfunknetz limitiert normalerweise das drahtlose Zugangnetz. Das Netzinterface der Endgeräte kann so gar nicht schneller senden als in der Gesamtsicht möglich. Der Sendepuffer der Endgeräte leert sich entsprechend nur mit der echten Senderate. Der TCP/IP-Stack kann sich besser darauf einstellen. In der skizzierten Link-/Netzemulation sitzt der Flaschenhals hingegen erst nach dem drahtlosen Zugangnetz, also nach dem WLAN, im Emulator selbst. Der Sendepuffer der Endgeräte kann sich dabei (stoßweise) mit der viel höheren WLAN-Senderate leeren, die Ausmittlung der Datenrate durch Drosselung erfolgt ja erst nachfolgend beim Weiterleiten der Pakete. Je nach TCP-Implementierung der Endpunkte der Kommunikation kann dies zu ungewünschten Effekten führen (Bufferbloat, siehe 3.4.4). Daher sind zielgerichtete Begrenzungen der Warteschlangen möglicherweise sinnvoll und wurden entsprechend berücksichtigt (siehe 5.5.5).

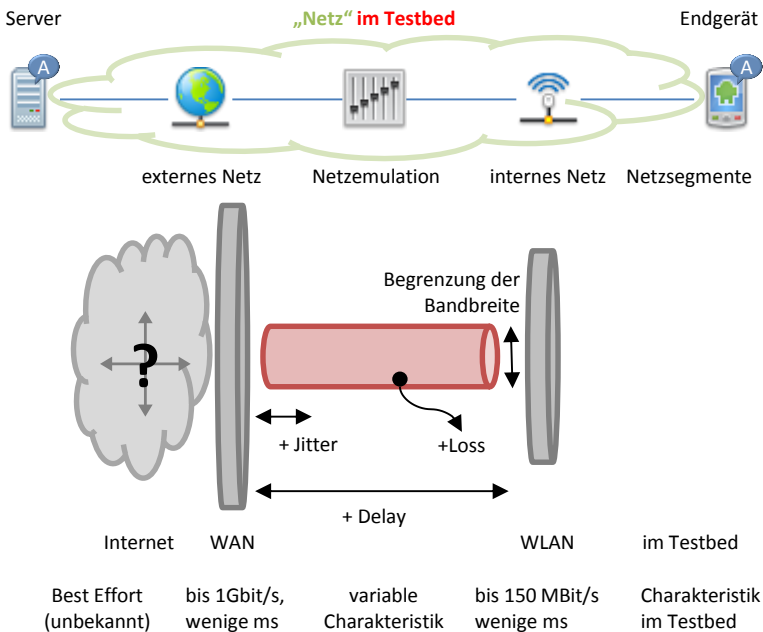


Abbildung 42: Modell zur Netzemulation als kontrollierter Flaschenhals zwischen externem und internem Netz
 in Anlehnung an [21] S. 240 („Röhrenmodell“), erweitert und angepasst

Die Gesamtsicht auf das System als Zusammenwirken der verschiedenen Netzabschnitte ergibt ein Netzscenario. Die Netzemulation präsentiert sich in diesem Modell als ein variabler Netzabschnitt, dessen Hauptcharakteristiken in Form von maximalen Bandbreiten und zusätzlicher Verzögerung (delay) sich leicht, flexibel und gezielt für verschiedene Szenarien anpassen lassen. Die äußeren Bedingungen sollten dabei möglichst konstant bleiben. Die gezielte Variation der Parameter der Netzemulation sollte die Hauptgröße in der Variabilität der Gesamtsicht sein. Eine solch saubere Trennung zwischen fixierter Ausgangssituation und gezielten Auswirkungen der Netzemulation ist zur Vermeidung unerwünschter Effekte empfehlenswert. So hat es wenig Sinn beispielsweise bei unbeeinflussten Schwankungen der Latenz (Jitter) von 10 ms im externen Netz gleichzeitig Effekte im Bereich von 1 ms, verursacht in der Netzemulation, untersuchen zu wollen. Deshalb sollte im Testbed eine möglichst perfekte Außenwelt gegeben sein.

Für das Verständnis der Möglichkeiten und Grenzen einer Netzemulation als künstlicher Flaschenhals zur gezielten Verschlechterung einer perfekten Außenwelt mit ansonsten perfektem Netz ist es essenziell, die Art und Richtung der Beeinflussung einzelner technischer Parameter zu vergegenwärtigen, siehe auch Tabelle 25. Die Bandbreite lässt sich nur unter den Wert senken, den die externe (WAN) oder interne Anbindung (LAN) vorgeben, jedoch nicht künstlich darüber hinaus steigern¹⁰⁷. Die Latenz lässt sich nur über die Ausgangswerte hinaus erhöhen, bereits verstrichene Zeit lässt sich nicht aufholen. Dies gilt isoliert auch für die Erhöhung des Jitters als technischer Parameter. In Kombination mit Veränderungen der Delay-Werte ist durch die Zwischenpufferung prinzipiell auch ein zeitlicher Ausgleich möglich. Die Paketverlustrate lässt sich ebenfalls nur erhöhen, verlorene oder beschädigte Pakete lassen sich nicht in der Netzemulation zurückholen.

Die perfekte Außenwelt ist bei der Anbindung und Nutzung eines real existierenden Netzes wie dem Internet in Kombination mit dem Emulations-Testbed nur ein idealisiertes Modell. In echten Netzen ohne QoS-Mechanismen gibt es Engpässe bei Bandbreiten, zusätzliche ungeplante Latenzzeiten sowie nicht

¹⁰⁷ Zusätzliche Maßnahmen, die wie eine Erhöhung der effektiven Bandbreite wirken, z. B. durch die Reduktion von Protokoll-Overhead oder Datenkompression, sollen hier explizit ausgeschlossen sein.

vorhersagbare Schwankungen dieser Werte. Entsprechend muss dies berücksichtigt werden, wenn die Werte für einzelne Parameter im Ausgangszustand (ohne aktive Netzemulation) und Zielzustand mit Beeinflussung (Netzemulation) größenordnungsmäßig nahe beieinanderliegen. Wenn gezielte innere Variationen und (zufällige) äußere Faktoren sich so gegenüberstehen, muss ggf. eine getrennte Betrachtung der einzelnen Ursachen oder eine gemeinsame Betrachtung der kumulierten Effekte erwogen werden, z. B. auf Basis nachvollziehbarer Messwerte. Auf die praktisch relevanten Grenzen der eigenen Realisierung in der genutzten Netzumgebung wird in 5.7 noch detailliert eingegangen.

In einer Netzemulation nach dem beschriebenen Prinzip lassen sich keine Garantien in der Gegenrichtung einer Beeinflussung erreichen, etwa für eine minimale Bandbreite oder maximale Latenz, wie sie oftmals wünschenswert wären. Grundsätzlich können aber die mobilfunktypischen Datenraten und Latenzwerte in einer Netzemulation realisiert werden. Die Richtung der Begrenzung entspricht dabei auch dem Stand in aktuellen Mobilfunknetzen ohne aktive QoS-Mechanismen für OTT-Dienste: Begrenzung der minimalen Latenz und der maximalen Datenrate durch die Technologie, bzw. bezogen auf die Datenrate auch durch den Vertrag („... mit bis zu ... kBit/s ...“).

Tabelle 25: Art und Richtung der Beeinflussung von Netzparametern in der Netzemulation (Auszug)

Parameter	Art der Beeinflussung	Beeinflussung durch
Bandbreite	Maximalwertbegrenzung	Begrenzung des maximalen Wertes der durchschnittlichen Datenmenge pro Zeiteinheit durch Verzögerung oder Verwerfen von Paketen
Latenz	Minimalwertbegrenzung	zusätzliche Verzögerung von Paketen je Richtung (Delay)
Jitter	Minimalwertbegrenzung	zusätzliche Schwankung des Delays
Paketverlustrate	Minimalwertbegrenzung	(zufälliges) Verwerfen kompletter Pakete

5.3.3 Globale vs. individuelle Sicht

Die in der Testumgebung darstellbaren technischen Parameter beziehen sich auf einzelne, isolierte Nutzer bzw. Endgeräte. Bandbreite, Latenz und die restlichen technischen QoS-Parameter etc. werden dabei für jeden Nutzer unabhängig von den Einstellungen der übrigen Nutzer variiert. Eine globale Sicht, d. h. wie sich das Netz im Ganzen, bzw. auch wie sich eine einzelne Mobilfunkzelle in Summe verhält, wird mit dem gewählten Ansatz nicht abgebildet. In der Realität wird die durch andere Nutzer verursachte Netzlast in einem Shared Medium, wie es Mobilfunk bzw. eine Mobilfunkzelle darstellt, zumindest indirekten Einfluss haben. Diese Effekte, z. B. in Form von Überbuchung bei gleichzeitig hohem Datenratenbedarf vieler Nutzer, müssen bei Interesse mit dem gewählten Ansatz in separaten technischen Konfigurationen abgebildet werden. Entscheidend ist letztlich aus Nutzersicht, wie sich in bestimmten Nutzungssituationen die relevanten technischen Parameter darstellen, unabhängig von den eigentlichen Ursachen, die zu diesem Zeitpunkt wirken.

Einige KPI sind unter Umständen noch weitergehend, da konkrete Zellauslastungen in bestimmten Konstellationen etc. mit zu berücksichtigen sind (Verfügbarkeit und Nutzbarkeit des Netzes). Dies kann eine einfache Emulation nicht nachstellen – dies wäre aber auch eine andere Zielsetzung. So gesehen ist der gewählte Ansatz eigentlich keine Netzemulation mit der Emulation eines ganzen Netzes in einer globalen Sicht mit Berücksichtigung aller möglichen Elemente, sondern eine Emulation von einzelnen mobilfunktypischen Datenverbindungen. Daher wäre der treffendere Begriff einer *Verbindungsemulation* (Link-Emulation) angebracht. Der Begriff Netzemulation ist aber auch entsprechend nutzbar, wenn mit „Netz“ die Netzcharakteristik hinsichtlich relevanter Parameter verstanden wird. Im gewählten Ansatz des kontrollierten Flaschenhalses wird durch die parametrisierte Verbindung der Netzsegmente untereinander die QoS-Ende-zu-Ende-Gesamtcharakteristik des Netzes maßgeblich geprägt. Entsprechend soll der Begriff einer Netzemulation deshalb weiterhin so genutzt werden, unter Berücksichtigung der genannten einschränkenden Aspekte hinsichtlich der tatsächlich möglichen Aussagen.

5.3.4 Verfügbare Ansätze und Lösungen zur Netzemulation

Der eigenen Realisierung ging eine Recherche nach vorhandenen Werkzeugen zur Netzemulation voraus. Dabei sollten geeignete fertige Lösungen für die Nutzung mit WLAN oder zumindest Ansätze als Basis für eine eigene Realisierung eines Testbed gefunden und ausgewählt werden. Recherche- und Auswahlkriterien waren o. g. Anforderungen für die Nutzbarkeit im definierten technischen Konzept und ein gewisser Reifegrad, da der enge zeitliche Rahmen keine Entwicklung von Grund auf und keine größeren Maßnahmen zur Anpassung und Fehlerbeseitigung zuließ. Es sollten möglichst Vorerfahrungen mit realistischen Tests und belastbare Messungen zur technischen Evaluation verfügbar sein.

Prinzipiell ist zwischen Emulatoren auf Basis von Software oder Hardware zu unterscheiden. Einen kurzen Einblick auf dem Stand von 2010 dazu gibt [244]. Demnach sind Hardware-Systeme in der Genauigkeit besser, vor allem dank der höheren zeitlichen Auflösung, allerdings nur kommerziell erhältlich. Software-Lösungen bieten die bessere Flexibilität und Erweiterbarkeit. Eine umfangreiche vergleichende Darstellung verfügbarer Lösungen und eine grobe Bewertung finden sich in [245].

Tabelle 26: Übersicht zu einer Auswahl möglicher Open-Source-Lösungen zur Netzemulation

Produkt	OS-Plattform	Reifegrad	Aktualität	Bemerkung
Seawind ¹⁰⁸	Unix-artige	unbekannt	veraltet, ca. 2001	Realisierung im Userspace
NistNET ¹⁰⁹	Linux ab 2.0.x bis 2.6.x	unbekannt	veraltet, 2005	ursprüngliche Implementierung für Linux
tc ¹¹⁰ (in ip-route2)	Linux ab 2.2.x	ausgereift	aktuell, gepflegt	tc als Basis für netem
netem ¹¹¹	Linux ab 2.4.28 / 2.6.8	ausgereift	aktuell, gepflegt	Erweiterung für tc zur Link Emulation
dummynet ¹¹²	FreeBSD, Mac OSX, Linux, Windows	ausgereift	aktuell, gepflegt	bevorzugt auf FreeBSD und Derivaten
KauNet ¹¹³	FreeBSD ab 7.3, Linux ab 2.6	ausgereift	aktuell, gepflegt	Add-on für dummynet

¹⁰⁸ <http://www.cs.helsinki.fi/research/iwtcp/seawind/> abgerufen am 27.11.2012

¹⁰⁹ <http://snad.ncsl.nist.gov/itg/nistnet/> abgerufen am 28.11.2012

¹¹⁰ <http://lartc.org/> abgerufen am 28.11.2012

¹¹¹ <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem> abgerufen am 28.11.2012

¹¹² <http://info.iet.unipi.it/~luigi/dummynet/> abgerufen am 28.11.2012

¹¹³ <http://www.kau.se/en/kaunet> abgerufen am 25.08.2013

Auf eine detaillierte Betrachtung kommerzieller Systeme wurde verzichtet, da schnell zu erkennen war, dass die frei verfügbaren Software-Lösungen, siehe Tabelle 26, alle Anforderungen befriedigen. Diese Lösungen bieten die allgemeinen Vorteile der Open-Source-Welt wie Flexibilität, Erweiterbarkeit und vor allem Anpassbarkeit an die speziellen Anforderungen, zur Not auch durch den Blick in den Quelltext. Die ermittelten Lösungen genießen in der wissenschaftlichen Gemeinschaft einen guten Ruf und werden in zahlreichen renommierten Projekten genutzt, gepflegt und teils auch aktiv weiterentwickelt. Die Lösungen arbeiten dabei grundsätzlich auf Basis des zuvor vorgestellten theoretischen Prinzips.

Die hauptsächlichen Anwendungsbereiche der Netzemulatoren liegen in den folgenden Forschungs- und Entwicklungsbereichen (mit einigen ausgewählten Beispielprojekten):

- Future Internet Forschung (G-LAB¹¹⁴ (German Lab) / G-Lab-Deep¹¹⁵, z. B. mit Tomato¹¹⁶)
- Nutzertests und Evaluationen zur QoE, z. B. beim Web-Browsing und Video-Konferenzen (ACE¹¹⁷ und NuPEX¹¹⁸)
- Anwendungstest und anwendungsbezogene Tests zur QoS
- Protokollentwicklung, Protocol-Testing, Robustheitstest
- Link Emulation allgemein, besonders für Satelliten- und Mobilkommunikation
- Unterstützung bei der Erprobung mobiler Apps unter annähernd realistischen Bedingungen (jenseits des goldenen Käfigs perfekter lokaler Netzbedingungen)

Die Literaturrecherche gab wichtige Hinweise zur Nutzbarkeit und Leistungsfähigkeit der verschiedenen Lösungen. In [246] findet sich eine Zusammenfassung zum Vergleich der Genauigkeit hinsichtlich Delay zwischen *NISTnet*, *netem* und *KauNet* und in [247] ein Vergleich zum Durchsatz in *netem* und *KauNet*.

¹¹⁴ <http://www.german-lab.de> abgerufen am 25.02.2015

¹¹⁵ <http://www.g-lab-deep.de> abgerufen am 25.02.2015

¹¹⁶ <http://tomato3.german-lab.de> abgerufen am 27.11.2012 bzw. <http://tomato-lab.org/> abgerufen am 17.08.2014

¹¹⁷ <http://ace.ftw.at/> abgerufen am 29.11.2012

¹¹⁸ <http://www.nupex.de> abgerufen am 19.08.2013

[248] resümiert dazu zum Vergleich der verschiedenen Lösungen:

- *netem* bietet eine gute Delay-Genauigkeit, die konfigurierten Minimalwerte werden nicht unterschritten und die Maximalwerte weichen von den 3 betrachteten Lösungen am wenigsten ab.
- Die betrachteten Lösungen halten beim Durchsatz die Zielwerte insgesamt gut ein. *netem* hat geringe Vorteile.
- Die Parametrisierung des Durchsatzes funktioniert für große Nutzdateneinheiten (Pakete) besser, d. h. genauer.
- Die Eigenschaften der Hardware-Plattform können einen (geringen) Einfluss auf die erzielbaren Ergebnisse haben (im Vergleich waren x86-Architekturen). Eine einheitliche Hardware-Plattform war daher für den eigenen Test anzustreben.

In die engere Auswahl haben es zwei Systeme geschafft: *dummysnet* und *tc+netem*. Der Term *tc+netem* umfasst dabei *tc* (Traffic Control) und *netem* (Network Emulator), die gemeinsam genutzt zur angestrebten Netzemulation geeignet sind. Diese Systeme bieten den benötigten Funktionsumfang, um das Testbed wie skizziert zu realisieren, sind gut dokumentiert und in ihrer Funktionsweise und Leistungsfähigkeit detailliert untersucht, siehe auch [245], [249] und [250]. Eine weitere interessante und aktuelle Lösung ist mit KauNet als Weiterentwicklung bzw. Add-on zu *dummysnet* [251] erst relativ spät in den Recherchen aufgetaucht und blieb deshalb weitgehend unberücksichtigt. Für die praktische Umsetzung mit der gegebenen Zielsetzung war dies aber nicht nachteilig.

Für *dummysnet* spricht in erster Linie die einfache Nutzung und Syntax, ein hoher Reifegrad und weite Verbreitung. Der Funktionsumfang von *dummysnet* ist bezogen auf die Anforderungen vollständig (eine Beschreibung zum letzten Entwicklungsstand findet sich in [244]). Ursprünglich wurde *dummysnet* nur zur Netzemulation auf OSI-Schicht 3 (Network) entwickelt. Zwischenzeitlich wurde die Fähigkeit zur Arbeit auf OSI-Schicht 2 (Link Layer) hinzugefügt [249], inklusive der Nachbildung spezifischer Effekte auf dieser Schicht [244]. Für eine akkurate Funktion von *dummysnet* ist eine hohe zeitliche Granularität erforderlich. Für die Betriebssysteme auf denen *dummysnet* aufsetzt, bedeutet dies, dass eine hohe Kernel-Frequenz vorteilhaft ist. Diese Voraussetzungen

gibt es hauptsächlich unter BSD und Derivaten, für die *dummynet* ursprünglich entwickelt wurde. Für die Portierungen auf andere Betriebssysteme, wie z. B. Linux, sind entsprechende Optimierungen notwendig.

Für das im Testbed bevorzugte Betriebssystem Linux ist *dummynet* zwar auch verfügbar, dort steht aber mit *tc+netem* eine vergleichbare Lösung, optimiert auf den Linux-Kernel, unmittelbar bereit. Neben diesem wichtigen praktischen Gesichtspunkt konnte in den o. g. Vergleichen *tc+netem* auch als führende State-of-the-Art-Lösung überzeugen.

Keine der gefundenen Lösungen war praktisch sofort für den angestrebten Zweck einsetzbar. Für die deshalb notwendige eigene Realisierung eines Gesamtsystems fiel die Wahl auf *tc+netem* als Basis zur Netzemulation. Die detaillierte Beschreibung dazu, inklusive der genutzten Möglichkeiten und Grenzen, erfolgt nach der Diskussion der genutzten technischen Plattform, um die Beispiele und Illustrationen mit den korrekten Netzschnittstellen und der tatsächlichen Konfiguration zeigen zu können.

5.4 Grundlegende technische Realisierung

Die Realisierung der technischen Plattform wird anhand festgelegter Grundprinzipien sowie der tatsächlich zum Einsatz gebrachten Komponenten der Hardware- und Software-Seite dargestellt. Dazu wird die Basiskonfiguration als Ausgangspunkt für nachfolgende Erläuterungen zur Netzemulation und den zusätzlichen Überwachungsfunktionen im Testbed eingeführt.

5.4.1 Grundprinzipien der Realisierung

Für die Realisierung der Netzemulation sollten einige Grundprinzipien erfüllt werden. Durch großzügige Planungen sollten ungewollte Engpässe vermieden werden, die dem Ansatz eines kontrollierten Flaschenhalses entgegenstehen. Probate Ansätze dafür sind die Bereitstellung von Überkapazitäten (Over-Provisioning), Lastverteilung und explizite Zuweisung von Ressourcen:

- **Over-Provisioning:** Die externe Schnittstelle sollte die verfügbare GBit-Anbindung der Netzumgebung ausnutzen können. Entsprechend waren GBit-fähige Ethernet-Schnittstellen vorzusehen, selbst wenn im Rahmen der Netzemulation nur mit deutlich niedrigeren Datenraten gearbeitet

werden sollte. Gleiches gilt auch für die interne drahtlose Schnittstelle (WLAN), die auch mit deutlich höherer Geschwindigkeit operieren sollte, als eigentlich notwendig.

- **Lastverteilung:** Die Skalierung der Netzemulation sollte einfach möglich sein, um nicht nur ein Netzszenario für ein mobiles Endgerät, sondern verschiedene Netzsituationen für eine Vielzahl mobiler Endgeräte gleichzeitig testen zu können. Dabei sollte eine einzelne Instanz nicht überlastet werden, was auch für die sonstigen Ressourcen wie CPU-Leistung und Hauptspeicher gilt.
- **Explizite Zuweisung:** Die Endgeräte werden den vorhandenen Ressourcen, wie z. B. den Instanzen zur Netzemulation oder aufgespannten drahtlosen Netzen, explizit und fest zugewiesen. Damit ergibt sich eine gute Vorhersagbarkeit des Gesamtsystems.

Entsprechend des Prinzips der Lastverteilung und der expliziten Zuweisung sollen sich problemlos mehrere Instanzen gleichzeitig und unabhängig voneinander nutzen lassen. Deshalb wurde für die eigentliche Netzemulation ein Ansatz als Appliance gewählt. Diese bedeutet, dass die integrierte Lösung aus Hardware und Software genau die benötigte Funktionalität bei Einhaltung der Grundprinzipien bietet und einfach zu skalieren ist. Die einzelnen Instanzen zur Netzemulation sollen sich zu einem Gesamtsystem zusammenfassen lassen und eine gemeinsame Steuerung und Verwaltung ermöglichen. Diese Aufgaben übernehmen zentralisierte Komponenten auf einem Server. Durch diese funktionelle Kapselung blieb die Komplexität der Gesamtlösung überschaubar.

Das Testbed sollte nicht nur technisch korrekt funktionieren, sondern auch für die Bedienung durch Testpersonal geeignet sein, das mit dem Systementwickler keine Personalunion bildet. Bei einer großen Anzahl Testslots ist dies ein wichtiges Skalierungskriterium. In- und Außerbetriebnahme sollten sich als „Strom Einschalten und am Ende wieder Ausschalten“ beschreiben lassen. Eine konsistente Bedienung über eine Web-GUI sollte eine einfache und zuverlässige Steuerung zur Konfigurationsanpassung und Überwachung der laufenden Tests erlauben.

5.4.2 Auswahl der Hardware- und Software-Basis

Die Festlegung auf eine kommerziell verfügbare Embedded-Lösung¹¹⁹ (System-on-Chip) reduzierte den Realisierungsaufwand für die Hardware-Seite der Appliances auf ein Minimum, da bei geeigneten Geräten alle Anforderungen durch On-Board-Komponenten erfüllt sind. Zudem erleichterte die einheitliche und baugleiche Plattform den Skalierungsschritt von einem Gerät zu mehreren. Im Markt finden sich solche geeigneten Systeme in Form normaler (d. h. billiger) WLAN-Router für die Heimvernetzung. Mit der standardmäßig vorinstallierten und ausgelieferten Betriebssoftware sind diese Geräte normalerweise nicht für die Aufgaben einer Netzemulation im Szenario eines zuvor beschriebenen Testbeds nutzbar. Deshalb mussten mit Blick auf die favorisierte alternative Firmware als Software-Basis zur Netzemulation neben der Hardware auch die Eignung und der Reifegrad der Software-Unterstützung berücksichtigt werden.

Die Auswahl eines technisch geeigneten, leistungsfähigen und Software-kompatiblen Gerätes für die weitere Arbeit fiel auf den Typ *Netgear WNDR 3800* mit folgenden Eigenschaften¹²⁰:

- GBit-Ethernet für die schnelle externe Anbindung (WAN)
- GBit-Ethernet für die interne Anbindung (LAN) mit integriertem Switch-Chip, VLAN-fähig
- Dual-Band WLAN mit 2,4 GHz und 5 GHz nach IEEE 802.11n (max. 150 bzw. 300 MBit/s)
- angemessen schnelle CPU (680 MHz MIPS 24K)
- RAM groß genug (128 MB)
- Flash-Speicher groß genug (16 MB) für das Basis-Linux und zusätzliche Softwarepakete, z. B. die notwendigen Komponenten zur Netzemulation
- USB-Anschluss zur lokalen Erweiterung (z. B. zusätzlicher Massenspeicher)
- gute Unterstützung für alternative Firmware

¹¹⁹ COTS (Commercial Off-The-Shelf)

¹²⁰ <http://support.netgear.de/product/WNDR3800> abgerufen am 30.11.2012

Wobei nur die Hardware-Eigenschaften relevant sind, da der Betrieb im Testbed mit einer alternativen Firmware erfolgte.

Weitere technische Spezifikationen unter: <http://wiki.openwrt.org/toh/netgear/wndr3800> abgerufen am 30.11.2014

Die vorinstallierte Firmware wurde durch OpenWrt¹²¹, eine alternative Betriebssoftware auf Basis von Linux, ersetzt. Die zum Zeitpunkt der Realisierung aktuelle Version „Attitude Adjustment, 12.09 beta“¹²² von OpenWrt kam dabei zum Einsatz.

OpenWrt ist eine gute Ausgangsbasis für die Embedded-Entwicklung, die sich für die Realisierung der Software-basierten Netzemulation durch folgende Punkte auszeichnete:

- moderner Kernel (Linux 3.3 zur Zeit der Implementierung)
- minimales Grundsystem ohne unnötige Komponenten, angepasst auf begrenzte Ressourcen
- stark modularisierte Software-Pakete und flexibles Software-Paketmanagement
- Verfügbarkeit der relevanten Software-Grundpakete zur Netzemulation in aktuellen Versionen
- strukturierte und vereinheitlichte Konfiguration
- einfache Erweiterbarkeit
- ausreichend gute Dokumentation durch breite Anwender- und Entwicklerbasis

Ergänzt wurden die WLAN-Router um Ethernet-Switch-Technik zur Verbindung der Geräte untereinander: für den Uplink, für die Zusammenfassung der Mirror-Ports und für ein internes drahtgebundenes Netz. Dabei fiel die Wahl auf einen einfachen 8-Port-Gigabit-Switch für den Uplink sowie einen managed 8-Port-Switch für das interne Netz und für die Mirror-Ports, um diese selbst wieder über einen einzelnen dedizierten Mirror-Out-Port zusammenfasst zur Verfügung zu stellen. An jedem WLAN-Router zur Netzemulation waren entsprechend drei Kabel mit unterschiedlichen Funktionen vorgesehen, und zur Unterscheidung farblich gekennzeichnet (z. B. blau für die Mirror-Ports).

Die notwendige Technik zum Testbed passt inklusive der zuvor genannten zusätzlichen Netztechnik, Verkabelung und Stromverteilung prinzipiell in eine Transport-Box. Das Testbed-in-a-Box war so für Vorabversuche, noch außer-

¹²¹ <https://openwrt.org/> abgerufen am 30.11.2012

¹²² http://downloads.openwrt.org/attitude_adjustment/12.09-beta/ abgerufen am 30.11.2012

halb der eigentlichen Testumgebung, portabel und jeweils durch das Anstecken von lediglich zwei Kabeln (Stromkabel und Ethernet-Kabel für den Campus-Netzzugang) schnell in Betrieb zu nehmen. Für den eigentlichen Testbetrieb wurde das Testbed dann schließlich so in die Testumgebung integriert, dass die zusätzliche Technik und Verkabelung unauffällig unter Tischen und in Kabelkanälen verschwand. Die Inbetriebnahme bzw. Außerbetriebnahme des Testbeds reduzierte sich so tatsächlich jeweils auf das Betätigen des Schalters an einem Strommehrfachverteiler.

5.4.3 Basiskonfiguration

Die grundlegende Konfiguration der Netzchnittstellen ist ein wichtiger Schritt zur Vorbereitung der Netzemulation. Auf Basis der physikalischen Schnittstellen der Hardware, siehe Tabelle 27, erfolgte die Konfiguration der logischen Schnittstellen zur geplanten Netzemulation und für weitere Zusatzfunktionen.

Auf der Schnittstelle eth0 wurden mittels des in der Hardware integrierten Switch-Chips zwei VLAN eingerichtet: eth0.1 und eth0.2. Zusätzlich wurde im Switch-Chip das Learning deaktiviert, d. h. die verbundenen Ports in eine Art Hub-Modus gebracht. Eine interne Bridge fasst die Schnittstelle eth0.1 sowie die definierten drahtlosen Netze wlan0 und wlan1 zu br-lan zusammen. Das auf den ersten Blick komplizierte Konstrukt aus VLAN, Bridge, Hub-Modus und Isolation eines einzelnen Switch-Ports, siehe Abbildung 43, ist für die späteren Schritte zur Einrichtung der Netzemulation und Überwachungsfunktionen wichtig, siehe Tabelle 28 zur geplanten Verwendung.

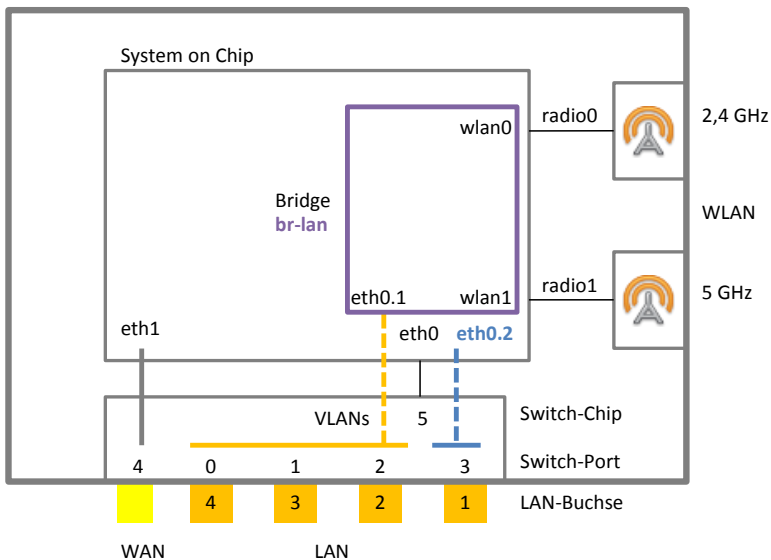
Tabelle 27: Technologien und max. Bruttogeschwindigkeiten der Netzwerkschnittstellen

Netz-schnitt-stelle	Technologie	Speed (Brutto)
eth0	Ethernet	1 GBit/s (switched, akkumuliert)
eth1	Ethernet	1 GBit/s
radio0	WLAN, IEEE 802.11b/g/n	max. 150 MBit/s, bzw. 300 MBit/s (Kanalbündelung)
radio1	WLAN, IEEE 802.11a/n	max. 150 MBit/s, bzw. 300 MBit/s (Kanalbündelung)

Tabelle 28: Verwendung der Netzchnittstellen und Richtungen des Netzwerkverkehrs in der Netzemulation

Netz-schnitt-stelle	Verwendung	Egress	Ingress
eth1	WAN: externes Netz, Anbindung Internet (Uplink)	UP	DOWN
br-lan	Bridge: internes Netz	DOWN	UP
...eth0.1	LAN: Anbindung drahtgebundener Endgeräte (und Videoüberwachung)	DOWN	UP
...wlan0	WLAN: Anbindung drahtloser Endgeräte im 2,4-GHz-Band	DOWN	UP
...wlan1	WLAN: Anbindung drahtloser Endgeräte im 5-GHz-Band	DOWN	UP
eth0.2	Mirror: Ausleitung von Netzwerkverkehr (Überwachung und Aufzeichnung)		

Erläuterungen zu Egress, Ingress, UP und DOWN im Abschnitt 5.5.1

**Abbildung 43: Schnittstellen der Netzkonfiguration im Blockschaltbild**

Die Konnektivität zum externen Netz (Internet) erfolgt über die Schnittstelle eth1, die automatisch per DHCP-Client eine externe Netzkonfiguration (inklusive IP-Adresse) bezieht. Der in OpenWrt integrierte DHCP-Server verteilt an anfragende Endgeräte im internen Netz auf br-lan IP-Adressen aus einem privaten Netzbereich. Ausgewählte DHCP-Clients, die nachfolgend innerhalb der Netzemulation über ihre IP-Adresse zu unterscheiden sein sollen, erhalten über die zugeordnete eindeutige MAC-Adresse eine festgelegte IP-Adresse

aus dem internen Netz. Zur einfachen Nachvollziehbarkeit der Zuordnung sind die Endgeräte 1 bis 6 den IP-Adressen a.b.c.10{1..6} im internen Netz zugeordnet. Per DHCP-Konfiguration werden die externen DNS¹²³-Server als Vorgabe an die internen DHCP-Clients durchgereicht. So können nachfolgend auch DNS-Abfragen von der Netzemulation erfasst werden. Dies ist ein wichtiger Punkt für Dienste mit häufiger DNS-Nutzung (z. B. Web-Browsing). Zwischen externem und internem Netz werden Pakete mit einer 1:n Network Address Translation (NAT) weitergeleitet. Bis zu diesem Realisierungsschritt funktioniert die Konfiguration noch als normaler Zugangspunkt zum Internet via LAN und WLAN.

5.5 Netzemulation mit *tc+netem*

Für Linux existieren mächtige Werkzeuge zur Manipulation von Netzwerkverkehr. Nach der Einführung einiger grundlegender Begriffe (Egress/Ingress) sollen die wichtigsten Konzepte und Komponenten zur Netzemulation mit *tc+netem* vorgestellt werden. Die ausgewählten Komponenten und das grundlegende Konzept (Klassenbaum und Filter) werden soweit erläutert, wie für das weitere Verständnis notwendig, allerdings wird dabei bewusst auf überbordende Beispiele und Erklärungen zur Syntax verzichtet, da diese besser in einer technischen Dokumentation aufgehoben sind. Einen guten Überblick und Einstieg zu Queuing Disciplines (Qdiscs) und zur Nutzung im Traffic Shaping liefert beispielsweise [252] S. 55ff (Grundlagen und theoretische Funktion abseits konkreter Implementierungen).

5.5.1 Egress vs. Ingress und UP vs. DOWN

Die Unterscheidung von Empfangen und Senden sowie der logischen Richtung des Netzwerkverkehrs ist für die weiteren Betrachtungen essenziell. Netzwerkverkehr, der das System über eine Netzwerkschnittstelle erreicht (Empfangen), wird als Ingress definiert, der das System verlässt (Senden) als Egress. Für beide Arten können (logische) Warteschlangen definiert werden, nachfolgend auch Queues genannt. Das Einfügen eines Paketes in eine Queue wird nachfolgend als Enqueue und das Entnehmen als Dequeue bezeichnet. Ein Ingress-

¹²³ Domain Name System

Paket wird nahe¹²⁴ zum Zeitpunkt des Empfangs durch die Hardware in eine Queue eingefügt, ein Egress-Paket nach dem Dequeue durch die Hardware gesendet. Die logische Richtung des Netzverkehrs wird im Weiteren aus Sicht der Endgeräte im internen Netz definiert. Die Richtung UP ist von den Geräten des internen Netzes zum externen Netz (Internet) gerichtet. Die Richtung DOWN ist vom externen Netz (Internet) zu den Geräten des internen Netzes gerichtet. UP als auch DOWN können damit je nach Netzwerkschnittstelle jeweils Ingress bzw. Egress sein, siehe Tabelle 28.

5.5.2 Positionierung

Die *tc*-Konfiguration ist für jede Schnittstelle und Verkehrsrichtung getrennt möglich. Dabei ist es egal, ob echte oder logische Schnittstellen wie VLAN (*eth0.1*) und Brücken (*br-lan*) genutzt werden. Die Mehrzahl der *tc*-Funktionen steht aber nur für Egress-Warteschlangen zur Verfügung. Mit dem Intermediate Queueing Device¹²⁵ (IMQ) bzw. Intermediate Functional Block¹²⁶ (IFB) existieren aber Dummy-Netzschnittstellen, um aus *tc*-Sicht die Ingress-Pakete durch Zwischenpufferung in Egress-Pakete umzuwandeln. Da aber die Netzemulation ausschließlich für durchlaufende Pakete eingesetzt werden sollte, waren diese Zwischenschnittstellen nicht notwendig, da ohnehin in beide Richtungen gesendet wird.

Die Netzemulation in Richtung UP setzt an der Schnittstelle *eth1* an, für die Richtung DOWN ist *br-lan* geeignet, da so über die Bridge alle Endgeräte unabhängig von den tatsächlich genutzten internen Schnittstellen erfasst werden, siehe Abbildung 44. Die Netzemulation ist durch die Positionierung an den ausgewählten Netzschnittstellen so sowohl für drahtlose als auch drahtgebundene Endgeräte (zu Testzwecken und für Vergleichsmessungen) nutzbar.

¹²⁴ Es sind weitere Puffer in den Treiber der Hardware oder der Hardware selbst möglich, die im Weiteren aber nicht betrachtet werden.

¹²⁵ <http://www.linuximq.net/> abgerufen am 02.12.2012

¹²⁶ <http://www.linuxfoundation.org/collaborate/workgroups/networking/ifb> abgerufen am 02.12.2012

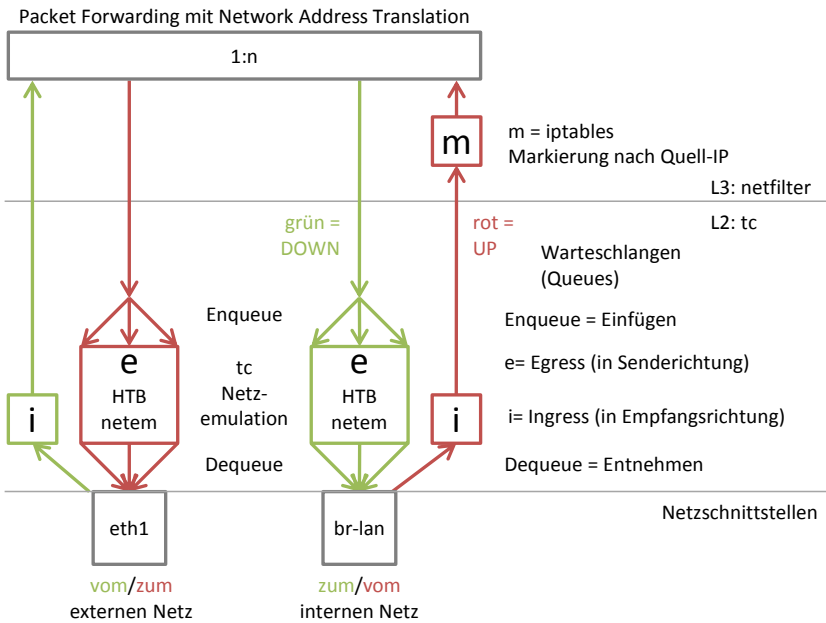


Abbildung 44: Kombinierte Netzemulation für UP und DOWN
 Die zur Netzemulation interessante Paketverarbeitung erfolgt jeweils in Senderichtung (Egress). Das Innenleben der angedeuteten Egress-Bäume wird in Abbildung 45 detaillierter dargestellt.

5.5.3 Konzept und Komponenten

Die relevanten Komponenten teilen sich in Kernel-Ebene und Userspace-Ebene auf. Per Userspace-API lassen sich auf Schicht L2 (Ethernet Frames) *tc* und auf Schicht L3 (IP-Pakete) das *netfilter*-Framework im Kernel konfigurieren. Ausführliche Beschreibungen zur Theorie von *tc* finden sich in [253], zahlreiche Beispiele zur Anwendung und Syntax finden sich im Internet¹²⁷.

Die Paketverarbeitung in *tc* ist als Baum organisiert. Die Wurzel des Baumes bildet die Schnittstelle zum Kernel, über die Pakete gesendet (Egress) oder empfangen werden können (Ingress). Für jede Netzwerkschnittstelle lässt sich eine Egress Root Queuing Discipline (Qdisc) und eine optionale Ingress Qdisc definieren. Es sind sowohl klassenlose als auch klassenbasierte Qdiscs möglich.

¹²⁷ z. B. unter <http://lartc.org/howto> abgerufen am 28.11.2012

Qdiscs sind Algorithmen zum Management der Pufferwarteschlangen mit folgenden Möglichkeiten und Aufgaben:

- Scheduling: Priorisierung und Klassifizierung zur getrennten Verarbeitung mehrerer Verkehrsströme über Filterregeln
- Shaping: Formen des Paketstroms beim Senden, z. B. durch Verzögern von Paketen um Datenratenbeschränkungen einzuhalten
- Policing: Verwerfen von Paketen um Datenratenbeschränkungen einzuhalten
- Link-Emulation durch Paketverluste, Delay, Jitter, Paketfehler etc. über *netem*
- dynamische Änderung der Parametrisierung (Umschaltung, Möglichkeit prinzipiell vorhanden)

Damit sind die notwendigen Funktionen zur Erfüllung der Anforderungen der Netzemulation in *tc+netem* grundsätzlich verfügbar. Die Komponenten von *netfilter*¹²⁸ stellen weitere Hilfs- und Unterstützungsfunktionen für die eigenen Zwecke bereit, die auch im Zusammenhang mit der internen Netzkonfiguration stehen (NAT und Paketmarkierungen).

Durch die hohe Modularität von *tc* sind über die Zeit von vielen Autoren unterschiedliche Ansätze zur Realisierung der jeweiligen Funktionen verwirklicht worden. Es soll nachfolgend nur zu den für eine Realisierung relevanten Komponenten weiter diskutiert werden.

5.5.4 Hierarchical Token Bucket und Klassenbaum

In der Literatur und bei anderen Projekten¹²⁹ in Zusammenhang einer Netzemulation unter Linux mit *tc* findet sich oft der Token Bucket Filter (TBF¹³⁰) als Qdisc zur Beschränkung der Bandbreite, teils auch in Kombination mit *netem* zur Link-Emulation, daher soll kurz auf den TBF eingegangen werden. Aus den Erläuterungen soll auch ersichtlich werden, warum letztlich die Entscheidung gegen TBF ausfiel. Das Prinzip des Token Buckets (siehe [34] S. 392 für ein anschauliches Beispiel und Erläuterung) wird gegenüber anderen Ansätzen, wie z. B. dem Leaky Bucket, für die angestrebten Zwecke

¹²⁸ <http://www.netfilter.org/> abgerufen am 15.08.2014

¹²⁹ z. B. bei <http://tomato-lab.org/> abgerufen am 17.08.2014

¹³⁰ <http://lartc.org/howto/lartc.qdisc.classless.html> abgerufen am 03.12.2012

als vorteilhaft angesehen, da z. B. Bursts als zeitlich wieder ausgeglichene Datenratenspitzen möglich sind (Vergleich der Prinzipien in [252] S. 64ff und S. 66f).

Positiv erscheint bei der TBF-Implementierung im Linux-Kernel die flexible Parametrierung:

- Datenrate und Spitzendatenrate
- maximale Paketwartzeit bzw. Anzahl wartender Pakete zur Begrenzung der Queue-Länge
- kontrolliertes Burst-Verhalten (min./max.)

Leider weist TBF auch einige Nachteile auf:

- Es können keine Klassen zur Verkehrsdifferenzierung eingerichtet werden (klassenlos).
- Es ist keine dynamische Rekonfiguration (`tc change`) möglich. Änderungen der Parameter erfordern ein Löschen (`tc delete`) der TBF-Qdisc, womit auch der komplette (Sub)-Baum verloren geht und erneutes Anfügen (`tc add`) der TBF-Qdisc und eventuell untergeordneter Knoten, z. B. *netem*, erforderlich werden.

Zur Definition eines komplexen Baumes sind klassenbasierte Qdiscs notwendig, um für mehrere Endgeräte unterschiedliche Parametrierungen parallel zu ermöglichen. Klassenbasierte Qdiscs können mit Klassen den Grad¹³¹ der Qdisc erhöhen. Klassen können zur Klassifizierung in einer übergeordneten Qdisc genutzt werden. Klassen besitzen immer eine Qdisc als Kindknoten im Baum. Ist diese Kind-Qdisc nicht spezifiziert, so ist dies standardmäßig eine FIFO-Qdisc (First In First Out).

Die Vorteile der Unterstützung von Klassen und der dynamischen Rekonfiguration bietet Hierarchical Token Bucket (HTB¹³²), bei einem ähnlichen Ansatz wie TBF zur Datenratenbeschränkung. Dies entspricht der eingangs formulierten Forderung, weshalb HTB für den Einsatz im Emulations-Framework ausgewählt wurde. HTB wird als Egress Root Qdisc mit der ID 1:0 und mehreren

¹³¹ Grad im Sinne der theoretischen Informatik als „Breite“ des (Sub-) Baums.

¹³² <http://luxik.cdi.cz/~devik/qos/htb/manual/userg.htm> abgerufen am 02.12.2012

untergeordneten Klassen (1:1x) genutzt. Über diese Klassen lassen sich mehrere voneinander unabhängige Parametrisierungen für einzelne Endgeräte gleichzeitig auf ein derselben Netzwerkschnittstelle realisieren. HTB-Klassen bieten folgenden Parameter zur Beeinflussung (Auszug):

- durchschnittliche maximale Datenrate und Spitzendatenrate (obere Begrenzung) als Shaping, d. h. als mit Verzögerung der Pakete durch (begrenzte) Zwischenpufferung.
- Beeinflussung des Burst-Verhaltens

Die HTB-Egress-Queue wird je Richtung in 6 + 1 unabhängige Sub-Queues aufgespalten (6 Endgeräte und Default), siehe Abbildung 45. Die Klasse 1:10 ist Default-Klasse, der alle nicht explizit klassifizierten Pakete zugeordnet werden.

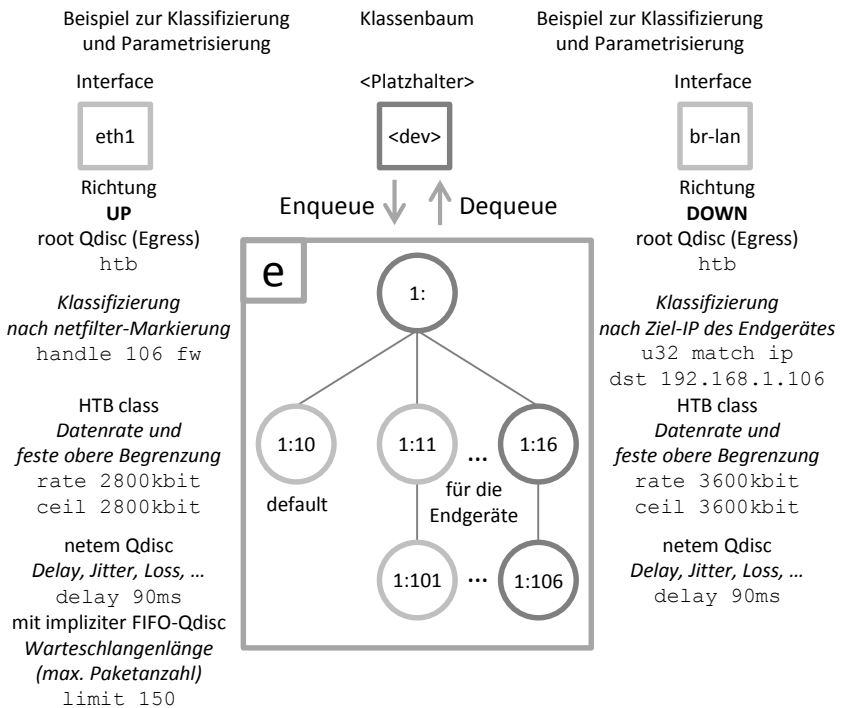


Abbildung 45: Klassenbasierter Baum zur Verkehrsdifferenzierung und -manipulation in der Netzemulation

Darstellung für Egress mit HTB und netem für Richtungen UP und DOWN, Syntax beispielhaft und verkürzt

Im Sinne einer einfachen Nomenklatur und Zuordnung werden die IDs mit den Nummern der Endgeräte (1 bis 6) gebildet, für die HTB-Klassen als 1:1 {1..6}. Im Regelbetrieb sind nur jeweils die Äste des Klassenbaums relevant, für die einer Instanz zur Netzemulation auch tatsächlich Endgeräte zugeordnet sind. Im Normalfall wurden zwei drahtlose Endgeräte je Netzemulator geplant. Für interne Tests und Demonstrationen können aber auch mehr als nur zwei Endgeräte über einen Zugangspunkt betrieben werden, ohne dass eine Veränderung der Grundkonfiguration notwendig ist. Über *tc*-Filter kann die Klassifikation von Paketen erfolgen. Im Netzemulations-Framework werden die Filter genutzt, um IP-Pakete über ihre Quell- bzw. Zieladresse korrekt in den *tc*-Baum einzuordnen. Über die fest zugewiesenen IP-Adressen lassen sich die Endgeräte im internen Netz eindeutig unterscheiden. Für die Richtung DOWN kann die Klassifikation direkt auf Basis der Ziel-IP im internen Netz erfolgen. Für die Richtung UP ist die Unterstützung durch iptables-Regeln zur Markierung der Herkunft der Pakete aus dem internen Netz vor dem 1:n-NAT notwendig.

5.5.5 netem

Zur Link-Emulation über eine Bandbreitenbegrenzung hinaus eignet sich *netem* (Network Emulator), eine spezielle klassenlose Qdisc, die nur als Blattknoten im Baum genutzt werden kann, siehe Abbildung 45. Es lässt sich dabei eine Vielzahl von Parametern nutzen:

- Delay (Paketverzögerung) und Jitter (Schwankung der Packverzögerung): minimale Verzögerung, Jitter-Korrelation und Jitter-Verteilung
- Loss (Paketverlust) und Corruption (Paketfehler): Korrelation und Verteilung
- Duplication (Paketdopplungen) und Reordering (Änderungen in der Paketreihenfolge): Korrelation, Verteilung und Lücken
- Warteschlangenlänge: max. Anzahl wartender Pakete

Implizit beinhaltet jede *netem*-Qdisc eine FIFO-Qdisc als eigentlichen Paketpuffer. Pakete können so vor einer Weiterleitung bis zum geplanten Sendezeitpunkt zwischengepuffert werden, solange der Puffer nicht vollständig gefüllt ist. Ist der Puffer vollständig gefüllt, werden weitere Pakete verworfen. Die Warteschlangenlänge des Puffers kann parametrisiert werden. Dies wurde im

Testbed für die Richtung UP genutzt, um exzessivem Bufferbloat entgegenzuwirken und eine Begrenzung der Latenz bei Sättigung der Uplink-Datenraten durchzusetzen. Die Warteschlangenlänge wird dabei dynamisch in Abhängigkeit der parametrisierten Datenrate und einer angenommenen typischen Paketgröße so gewählt, dass sich eine annähernd konstante maximale Wartezeit im Puffer ergibt. Für die Richtung DOWN wurde der Standardwert der Netzschnittstelle übernommen. Die Minor-IDs der *netem*-Qdiscs formen sich für die Hosts 1 bis 6 zu 1:10{1..6}, was wiederum mit den Nummern der Endgeräte korrespondiert.

5.6 Framework: APIs, Web-GUI und Co.

Im Prinzip sind die WLAN-Router als dezentrale Komponenten in der bis hier beschriebenen Konfiguration mit *tc+netem* schon für ihre eigentliche Aufgabe zur Netzemulation geeignet und können diese auch selbstständig übernehmen. Die (manuelle) Nutzung aus einer Shell heraus zur Initialisierung und (Re-) Parametrisierung ist allerdings äußerst umständlich und für Fehlbedienungen anfällig. An eine parallele Steuerung mehrerer Geräte, wie für die gleichzeitige Testdurchführung mit mehreren Probanden unter dem Aspekt Skalierung notwendig, ist dabei nicht zu denken. Entsprechend werden nachfolgend alle Aspekte eines Frameworks beschrieben, die zu einem praktisch einsetzbaren System führten. Dies umfasst einerseits Erweiterungen direkt auf den WLAN-Router und zusätzliche zentralisierte Komponenten, die auf einem Web-Server gespeichert und betrieben werden. Über die entwickelten APIs¹³³ können externe Systeme mit dem Netzemulations-Framework verbunden werden. So kann die Steuerung der gesamten Netzemulation automatisiert werden. Dies wird am Beispiel der geschaffenen Verknüpfung zwischen dem Testbed zur Netzemulation und dem genutzten Web-Fragebogensystem beschrieben.

5.6.1 Geräte-Cloning

OpenWrt bietet die Möglichkeit ein Overlay-Dateisystem zu nutzen. Dabei ist der Ausgangszustand eines frisch installierten Basissystems in einem Teil des Flash-Speichers der Hardware fixiert und kann nicht überschrieben werden.

¹³³ Application Programming Interface

Änderungen im Dateisystem, z. B. durch Einstellungen, zusätzliche Softwarepakete oder die Dateien der Netzemulation werden in einem gesonderten Bereich des Flash-Speichers abgelegt und transparent, als Overlay, in das Basissystem eingebunden.

Um eine weitere Hardware-Instanz zur Netzemulation zu erzeugen, genügt es, auf einem neuen Gerät das Basissystem zu installieren und direkt über das externe Netz das Overlay-Dateisystem mit sämtlichen Einstellungen etc. zu kopieren. So entsteht innerhalb weniger Minuten ein Klon, der nur in einigen Punkten (Hardware- und Software-spezifische IDs) individuell angepasst werden muss und sofort den gleichen Funktionsumfang bietet wie das geklonte Original. Der Skalierungsschritt von einem Gerät zur Netzemulation zur verteilten Netzemulation mit mehreren Geräten ist so einfach, schnell und elegant zu erledigen.

5.6.2 Datenhaltung und Konfiguration

Die Datenhaltung erfolgt teilweise dezentral auf den einzelnen Netzemulatoren als auch zentral für eine gemeinsame Konfiguration und Protokollierung auf einem Server:

- Zuordnung der mobilen Endgeräte (MAC-Adressen) und PCs der Testplätze (IP-Adressen)
- Preset-Definitionen und Parameterliste (Parameter, Einheiten, zulässige Bereiche)
- Testpläne und Testfortschritte
- Logdateien

Ein Teil der Konfiguration, z. B. die aktuellen Presets, wird nach dem Einschalten der Netzemulatoren automatisch mit dem Server abgeglichen, so dass die einzelnen Instanzen stets mit gleichen Vorgabenwerten arbeiten können, ohne aufwendige manuelle Kopieraktionen zu erfordern. Für die statischen Elemente zur Konfiguration, z. B. vordefinierte Presets und Zusatzinformationen zu den Hosts, wurden bewusst einfache Daten- und Dateiformate gewählt (textbasiert, CSV¹³⁴ und XML). Diese erlauben eine einfache manuelle Erstellung und Bearbeitung, einen problemlosen Import und Export zu externen

¹³⁴ Comma-Separated Values

Werkzeugen wie Excel und für gängige Programmiersprachen sind meist schon integrierte Funktionen zum Lesen und Schreiben vorhanden.

5.6.3 Netzemulations-Web-API

Um die verteilte Netzemulation elegant, einfach und ggf. auch automatisiert von außen zu steuern, wurde für das Emulations-Framework eine Web-API entwickelt. Die Web-API wurde als LuCI¹³⁵-Modul¹³⁶ im OpenWrt realisiert und kann so alle Vorzüge ausschöpfen, die LuCI offeriert:

- direkte Integration in die bestehende LuCI-Umgebung in OpenWrt, die dort z. B. für die grafische Nutzeroberfläche der Geräte eingesetzt wird
- Hosting über Webserver der LuCI-Umgebung
- MVC-Architektur (Model View Controller) zur sauberen Trennung von Funktion und Daten
- einfacher Zugriff auf das Unified Configuration Interface (UCI¹³⁷) zur Datenhaltung
- einfache Erweiterbarkeit über die Skriptsprache Lua¹³⁸

Die Gesamtheit aller (genutzten) Parameter zur Netzemulation und zugewiesenen Werte wird nachfolgend als Setting geführt. Jeder Host hat immer genau ein aktives Setting. Die Hosts können unabhängig voneinander verschiedene Settings haben. Es gibt eine Reihe vordefinierter Settings mit konkreten Werten zu einzelnen oder allen QoS-Netzparametern, die bereits diskutierten Presets. Intern greifen die entwickelten Lua-Komponenten der Web-API auf *tc* zurück, um die Settings in passende Emulationsparameter im Stil der skizzierten *tc*-Beispiele zur User Space API umzusetzen. Die Datenhaltung der dynamischen Elemente, z. B. die momentan aktiven Parameter-Settings der Hosts, nutzt die Möglichkeiten von UCI (Key-Value-Access). Das Datenaustauschformat über das Netz ist JSON¹³⁹, das sich in den genutzten Programmiersprachen zum Senden und Empfangen problemlos lesen und schreiben lässt sowie einfach in interne Datenstrukturen zu überführen ist.

¹³⁵ Framework für Web (User) Interfaces auf Embedded Devices, <http://luci.subsignal.org/trac> abgerufen am 03.12.2012

¹³⁶ <http://luci.subsignal.org/trac/wiki/Documentation/ModulesHowTo> abgerufen am 03.12.2012

¹³⁷ <http://wiki.openwrt.org/doc/techref/uci> abgerufen am 03.12.2012

¹³⁸ <http://www.lua.org/> abgerufen am 03.12.2012

¹³⁹ JavaScript Object Notation, <http://www.json.org/> abgerufen am 03.12.2012

5.6.4 Master Control API

Die Master Control API ist als Web-API (PHP) realisiert und dient zur zentralisierten Steuerung des Gesamtsystems, bestehend aus den zentralen Komponenten und ein oder mehreren verteilten Netzemulatoren. Die Master Control API agiert dabei als transparenter Wrapper. Die API reicht Anfragen gezielt an die Web-APIs der Netzemulatoren durch und leitet von da die Antwort an die Quelle der ursprünglichen Anfrage zurück. Für welches Endgerät die Netzparameter abgefragt oder verändert werden sollen, kann als Ziel explizit angegeben werden (z. B. Endgerät 1) oder implizit über die Quell-IP-Adresse der Anfrage ermittelt werden (z. B. von Testplatz 1, was automatisch ebenfalls zu Endgerät 1 führt). Das Beispiel in der Beschreibung der Verknüpfung zum Fragebogensystem illustriert nachfolgend die gemeinsame Funktion der konzipierten APIs im Testbed.

5.6.5 Master Control Web-GUI

Die zentralen Komponenten beinhalten eine Web-GUI¹⁴⁰, siehe Abbildung 46 und dazugehörige Erläuterungen. Zur Realisierung der Funktionen greift die Web-GUI als Frontend auf die o. g. APIs im Backend zu. In der Phase der Voruntersuchungen konnten so die benötigten Netzcharakteristiken über die Web-GUI schnell und flexibel manuell parametrisiert werden. Im normalen Testbetrieb ist hingegen lediglich eine Initialisierung der automatischen Testlaufsteuerung in der Vorbereitung eines Testslots notwendig. Dabei werden die Testpläne syntaktisch validiert. Im Fall eines festgestellten Fehlers wird der Testbetreuer eindeutig darauf hingewiesen, um noch vor Testbeginn korrigierend einzugreifen, siehe Abbildung 47. Ansonsten ist im laufenden Testbetrieb nur eine Überwachung vorgesehen, um auf unvorhergesehene Probleme zu reagieren. Die Anzeige der Optionen zur Parametrisierung kann per Konfigurationsdatei bequem gesteuert werden, um z. B. Expertenoptionen im normalen Testbetrieb gezielt auszublenden.

Erläuterungen zur Abbildung 46:

links der gestrichelten Linie: Legende; rechts der gestrichelten Linie: der eigentliche Screenshot
Die erste Spalte dient der Beschriftung der Zeilen, die weiteren Spalten repräsentieren jeweils einen Testplatz (1 bis 4 sind gezeigt, ab Testplatz Nr. 5 ist die Darstellung abgeschnitten).

a) Informationen zum Host (Endgerät) bzw. Testplatz (Nr.), Hinweise bei Störungen oder Fehlern im Testbed (hier gezeigt)

b) zugeordnete Ressourcen: ID des Netzemulators und genutztes internes WLAN

¹⁴⁰ Graphical User Interface

- c) automatische Testlaufsteuerung: Setzen und Aktualisieren der Testslot-Daten mit zugeordneten Fragebögen und Testplänen, Statusinformationen und Identifikationsmerkmalen zum Testverlauf
- d) Visualisierung der Testpläne: Tabelle mit Aufgabennummer, Aufgabenvariante und Preset-ID, anfangs mit weißem Hintergrund, Darstellung des Testfortschritts durch Grünfärbung, Verknüpfung zum Testlogbuch über Link in der Aufgabennummer
- e) manuelle Parametrisierung (Setzen neuer Werte oder Reset zum Verwerfen der neuen Werte), Visualisierung des Zustandes der Parameter (laufende Aktualisierung, manuell geändert (neu), unbekannt/Störung)
- f) Preset-Auswahl als Drop-Down-Liste mit Preset-ID und Kurzbeschreibung
- g) Anzeige und freie Eingabe der aktuellen / zu setzenden Werte der einzelnen Parameter: numerisch in Formularfeldern und/oder mit Schiebereglern, die auch den sinnvollen Bereich der Parametrisierung visualisieren, relevante Hauptparameter (oben, dargestellt), weitere Nebenparameter (unten, optional, nicht dargestellt)

SWT-Testbed (Edition 2.0.6, 2013)

Host / Testplatz	1	2	3	4	5
AP/WLAN	mirage1/swt.12	mirage1/swt.12	mirage3/swt.35	mirage2/swt.46	mirage3/
Testlaufsteuerung	Status: aktiv		Testslot: 21	Timestamp: 20130628T1400CEST	
Testlauf	20	20	21	18	18
Fragebogen	655337	655337	981499	464138	464138
Testplan	<ul style="list-style-type: none"> 1 G_A G20 2 M_B M30 3 S_A S10 4 F_A F50 5 Y_A Y50 6 G_B G40 7 D_B D20 8 M_A M10 9 S_B S40 10 Y_B Y40 11 F_B F30 12 D_A D30 	<ul style="list-style-type: none"> 1 G_A G50 2 M_B M10 3 S_A S30 4 F_A F30 5 Y_A Y30 6 G_B G50 7 D_B D40 8 M_A M20 9 S_B S20 10 Y_B Y20 11 F_B F50 12 D_A D40 	<ul style="list-style-type: none"> 1 S_B S10 2 Y_B Y30 3 F_B F30 4 D_A D20 5 G_A G10 6 M_B M10 7 S_A S10 8 F_A F50 9 Y_A Y30 10 G_B G40 11 D_B D30 12 M_A M50 	<ul style="list-style-type: none"> 1 S_B S30 2 Y_B Y30 3 F_B F40 4 D_A D40 5 Y_A Y20 6 G_B G20 7 D_B D20 8 M_A M10 9 G_A G10 10 M_B M40 11 S_A S20 12 F_A F20 	<ul style="list-style-type: none"> 1 S_B S10 2 Y_B Y30 3 F_B F40 4 D_A D40 5 Y_A Y20 6 G_B G20 7 D_B D20 8 M_A M10 9 G_A G10 10 M_B M40 11 S_A S20 12 F_A F20

Parametrisierung

Aktion	1	2	3	4	5
Setzen	Setzen	Setzen	Setzen	Setzen	Setzen
Reset	Reset	Reset	Reset	Reset	Reset
Preset	D30 HSDPA 1.8/1.4	D40 HSUPA 3.6/2.8	M50 HSDPA 1.8/1.4	F20 Edge 128/96	F30 Edg
rate_down kbit	1800	3600	1800	128	225
rate_up kbit	1400	2800	1400	96	192
delay_down ms	45	45	45	150	150
delay_up ms	45	45	45	150	150

Abbildung 46: Screenshot der Web-GUI (Ausschnitt)

1	1095	S_B	S30	1	1095	S_B	S50
2	1096	Y_B	Y40	2	1096	Y_B	Y40
3	1097	F_B	F10	3	1097	F_B	F10
4	1098	D_A	D50	4	1098	D_A	D30
5	1099	V_A	Y50	5	1099	V_A	Y50
6	1100	G_B	G20	6	1100	G_B	G40
7	1101	D_B	D40	7	1101	D_B	D10
8	1102	M_A	M50	8	1102	M_A	M20
9	1103	G_A	G50	9	1103	G_A	G30
10	1104	G_B	G30	10	1104	M_B	M30
11	1105	F_A	F40	11	1105	S_A	S30
12	1106	F_A	S40	12	1106	F_A	F20

Preset-ID nicht definiert

Abbildung 47: Ausschnitt Web-GUI mit Warnungen zu detektierten Fehlern im Testplan

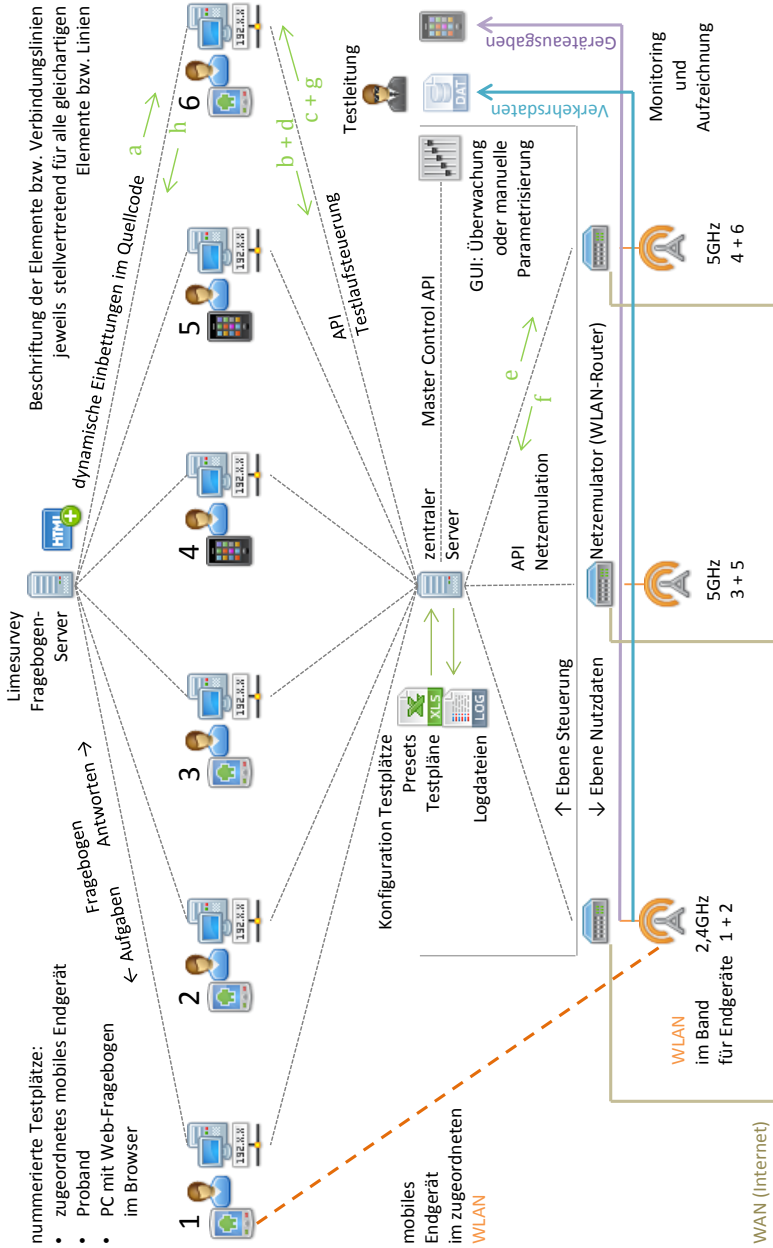


Abbildung 48: Anbindung des externen Fragebogensystems zur Testlaufsteuerung an das Testbed Darstellung der einzelnen Komponenten des Testbeds mit dem Netzemulations-Framework, den Endgeräten und Testplätzen

5.6.6 Verknüpfung von Testbed und Fragebogensystem

Für einen einzelnen Probanden könnte der Testleiter die Parametrisierung mit den gezeigten „Schiebereglern“ eventuell noch manuell durchführen. Für eine größere Zahl Probanden, wie in den Nutzertests mit bis zu sechs Probanden gleichzeitig, ist dies schlichtweg unmöglich. Eine manuelle Parametrisierung erfordert eine Synchronisierung zwischen Proband und Testleitung. Durch Überlastung des menschlichen Bedieners sind Fehler dabei sehr wahrscheinlich bzw. kaum zu vermeiden. Die geplanten Bedingungen und tatsächlich eingestellte Netzparameter würden dann nicht immer zusammenpassen. Dies muss unbedingt verhindert werden. Daher wurde das genutzte elektronische Fragebogensystem Limesurvey¹⁴¹ mit dem Netzemulations-Framework gekoppelt. Die Verknüpfung über die geschaffene API zur Testlaufsteuerung konnte die Testdurchführung stark vereinfachen, den menschlichen Unsicherheitsfaktor durch Automatisierung minimieren und eine konsistente Ergebniserfassung aus Nutzerbewertung und protokollierten technischen Parametern innerhalb einer Datenbasis realisieren. Mit der Automatisierung der Bedingungsvariation wurde eine weitere Voraussetzung für die Skalierung des Gesamttestumfangs erfüllt. Im Prinzip wäre somit eine einzelne, kurz eingewiesene Person als Testleiter vor Ort ausreichend gewesen.

Zur leichteren Nachvollziehbarkeit ist die Schrittfolge zur Verbindung von Testbed und Fragebogen-Tool im Funktionsschema durch Buchstaben und Pfeile gekennzeichnet, siehe Abbildung 48:

- a) Aus Limesurvey wird eine HTML-Seite zu einer Aufgabe vom Server in den Browser auf dem PC am Testplatz des Probanden geladen. Im HTML-Quelltext sind die zuvor eingefügten `<script>`-Einbettungen zu den technischen Zusatzfragen (z. B. zum Preset) enthalten. Limesurvey ergänzt dabei die SGQ¹⁴²-ID zur Identifikation als URL-Parameter für diese JavaScript-Elemente bei der Erzeugung der Aufgabenseite dynamisch.

¹⁴¹ <https://www.limesurvey.org/> abgerufen am 01.12.2012

¹⁴² Survey Group Question: Eine hierarchische ID in Limesurvey, über die sich sämtliche relevante Elemente wie der Fragebogen, über die Fragebogenseite bis hin zur einzelnen Frage bzw. deren Antwortfelder adressieren lassen.

- b) Der Browser rendert die Fragebogenseite und fordert dabei die über `<script>` eingebundenen externen JavaScript-Code-Fragmente vom zentralen Server des Frameworks ab. Die SGQ-ID wird als Abfrageparameter übergeben. Auf dem zentralen Server werden die zugeordneten PHP-Skripte ausgeführt.
- c) Es erfolgt die Rückgabe der dynamisch erzeugten JavaScript-Code-Stücke vom Server an den Browser.
Die JavaScript-Stücke bewirken im Browser des Probanden (für jeweils ein technisches Zusatzfeld, gekennzeichnet durch eine einzelne SGQ-ID):
- Die benannten Elemente werden auf der HTML-Seite ausgeblendet.
Für die Probanden sind die Mechanismen der versteckten Fragefelder zur automatischen Steuerung somit nicht sichtbar.
 - Die statischen Rückgabewerte werden in das zugehörige Antwortfeld eingetragen, z. B. für den Testplatz und Testslot.
 - Für die Preset-Aktualisierung zur Umschaltung der Netzparametrisierung wird ein Ajax-Request an die Framework-API zur Testlaufsteuerung vorbereitet, inklusive zugehöriger JavaScript-Callback-Funktion.
- d) Ein Ajax-Request vom Testplatz-Browser an die Web-API zur Testlaufsteuerung startet. Die SGQ-ID wird als Abfrageparameter übergeben.
Der zentrale Server ermittelt aus der Client-Quell-IP-Adresse den zugeordneten Testplatz, damit den aktuellen Testplan für diesen Testplatz und daraus wiederum das vorgesehene Preset für die im Browser angezeigte Aufgabe mit dieser SGQ-ID sowie den richtigen Netzemulator und das zum Testplatz gehörende mobile Endgerät (Host-ID als vergebene Nummer).
- e) Ein API-Request vom zentralen Server an die Netzemulations-API auf dem richtigen Netzemulation-Appliance für diesen Testplatz wird gestartet. Die Host-ID, Preset-ID und SGQ-ID werden als Abfrageparameter zum Setzen der richtigen Netzparameter für das Endgerät des anfragenden Testplatzes übergeben.

- f) Nach dem Setzen der Netzparameter des Presets für das entsprechende Endgerät auf dem Netzemulator erfolgt eine Antwort mit den tatsächlich eingestellten Parametern an den zentralen Server. Die SGQ-ID wird als aktuelle Zustandsinformation gespeichert, woraus der Testfortschritt abzuleiten ist. Preset und Umschaltzeitpunkt werden dokumentiert. (Eine zentrale Protokollierung dieser Daten ist optional möglich.)
- g) Die Antwort der Netzemulations-API wird unverändert an den anfragenden Browser des Probandenplatzes weitergeleitet. Im Browser wird die Preset-ID aus dieser Antwort in das versteckte Formularfeld der Frage Preset eingetragen. Damit sind Parametrisierung und technische Dokumentation dazu erfolgreich abgeschlossen.
- h) Nach der Aufgabenbearbeitung und Beantwortung der Fragen zur Zufriedenheit durch den Probanden werden die Daten in den versteckten Zusatzfeldern zusammen mit den anderen Feldwerten des Formulars (Bewertungen und Kommentare des Probanden) an das Fragebogensystem geschickt und dort gemeinsam mit diesen Werten in einer Datenbank gespeichert.

Prinzipiell hätten die Schritte b) und d) zur Einbindung und Umschaltung auch gemeinsam ausgeführt werden können, der gewählte Ansatz ist aber schneller beim Ausblenden der Fragen, die der Nutzer nicht zu Gesicht bekommen soll und flexibler konfigurierbar. Die Schritte a) bis g) benötigen in Summe ca. 200 ms. Damit ist quasi sofort mit dem Laden der entsprechenden Aufgaben-seite innerhalb des Fragenbogens im Browser des Testplatz-PCs durch einen Probanden die vorgesehene Netzparametrisierung passend zum Testplan für sein Endgerät und die angezeigte Aufgabe eingestellt und nachvollziehbar dokumentiert.

5.6.7 Elektronisches Testlogbuch

Sollten im laufenden Testbetrieb Probleme auftreten oder ein ungeplantes Eingreifen der Testleitung notwendig sein, z. B. als Hilfestellung an einen Probanden, so sollten diese Vorkommnisse in einem Testlogbuch erfasst werden. Über eine Verknüpfung von Web-GUI und elektronischem Testlogbuch lassen sich die Identifizierungsdaten automatisch korrekt belegen, siehe Abbildung 49. So sind Fehler des Testpersonals bei der Erfassung eines Eintrags effektiv zu verhindern. Die gesammelten Daten können in Tabellenform exportiert und in der Auswertung zur Validierung der Ergebnisse herangezogen werden.

Testlogbucheintrag

* Erforderlich

Testslot *
21 ▾

Testplatz *
4 ▾

Aufgabennummer *
9 ▾

Aufgabencode *
G_A ▾

Preset *
G10 ▾

valid *
Ist die Aufgabenausführung gültig?

Reporter *
Nutzerkürzel / Name des Eintragenden

Kommentar *
kurze Erläuterung

Senden

eindeutige Adressierung zur korrekten Zuordnung des Eintrags

automatische Vorbelegung mit richtigen Werten bei Aufruf über Web-GUI

Einschätzung der Testbetreuung aus der Situation heraus (ja / nein / unklar)

Der Kommentar kann in der späteren Auswertung wertvolle Hinweise geben:

- ob die Einschätzung zur Aufgabenausführung zutrifft
- wie die Tests zu verbessern sind

Abbildung 49: Erfassung eines Eintrags für das elektronische Testlogbuch über ein Online-Formular
Screenshot, Ausschnitt

5.7 Grenzen der realisierten Netzemulation im Testbed

Für die Variation der relevanten Netzparameter sind Aussagen zu folgenden Werten wünschenswert: mögliche Minimal- und Maximalwerte, Granularität der Parametrisierung, Genauigkeit der Parametrisierung und Verhalten bei typischen Werten. Einige der nachfolgenden Aussagen stützen sich auf Angaben aus Veröffentlichungen, die die Eigenschaften von *tc+netem* im Detail untersucht haben (z. B. [250] und [249]), andere auf eigene Feststellungen in Verbindung zur konkreten Hardware und Software im Testbed. Bei der Diskussion zur Variation der einzelnen QoS-Netzparameter in *tc+netem* ist zu beachten, dass sich die Angaben auf den Abschnitt der Netzemulation beziehen und nicht auf das Gesamtsystem. Dazu werden kurz die erreichten Effektivwerte angesprochen.

5.7.1 Perfekte Netzumgebung und kontrollierter Flaschenhals

Das Prinzip eines kontrollierten Flaschenhalses zur Netzemulation in ansonsten perfekter (Netz-) Umgebung war zu überprüfen. Die Bruttogeschwindigkeiten (Speed) der genutzten Netzschnittstellen bzw. der Technologien sind ausreichend hoch, siehe Tabelle 27. Die Internet-Konnektivität auf WAN-Seite wurde im Testbed mit einer 1 GBit/s Ethernet-Anbindung über das Campusnetz der TU Chemnitz an das Wissenschaftsnetz X-WiN¹⁴³ des DFN¹⁴⁴ ausgeführt. Der Anspruch des DFN an das Wissenschaftsnetz ist dabei sehr hoch (nach [254]):

- geringe Latenz von 1 ms pro 100 km
- geringer Jitter von max. 0,1 ms pro 100 km
- kein Paketverlust

Zum „offenen Internet“ jenseits des DFN können keine derartigen Aussagen getroffen werden, es wird also Best Effort angenommen. Daher ist neben der Vorabüberprüfung eine passive Überwachung zur Aufdeckung möglicher externer Flaschenhälse sinnvoll.

¹⁴³ <http://www.dfn.de/xwin/> liefert weitere Informationen zur Anbindung und Topologie, abgerufen am 26.11.2012.

¹⁴⁴ Deutsches Forschungsnetz – DFN-Verein, kurz DFN

Für die drahtlose Verbindung zu den Endgeräten wird im Testbed übliches WLAN genutzt, für das ebenfalls sehr niedrige Verzögerungen im Bereich weniger Millisekunden erreichbar sind. Die zu erwartende relevante Netto-Datenübertragungsleistung für die drahtlosen Schnittstellen ergibt sich im Zusammenspiel mit den genutzten Endgeräten und äußeren Störeinflüssen der Testumgebung. Die Bestimmung der entsprechenden Werte ist nur für die jeweiligen Konstellationen sinnvoll. Die Netto-Werte werden im Allgemeinen deutlich niedriger als die nominelle Brutto-Geschwindigkeit sein, aber noch immer ausreichend für die angestrebten Zwecke und Parameterbereiche.

Um die äußeren Störeinflüsse auf das WLAN abschätzen zu können, wurde die Ausgangssituation in der Testumgebung vor Aufbau des Testbeds erfasst. Die Erfassung erfolgte mit den mobilen Endgeräten und Wi-Fi-Monitoring-Tools, die eine schnelle Beurteilung der Kanalbelegung und ungefähren Signalstärken erlauben. Zum Einsatz kamen Wi-Fi-Analyzer¹⁴⁵ und inSSIDer¹⁴⁶. In den Vorbereitungen zeigte sich, dass die WLAN-Konfiguration sorgfältig zu wählen war. Bedingt durch die äußeren Gegebenheiten stand im 2,4GHz-Band nur ein Kanal zur Verfügung, der 5GHz-Bereich war glücklicherweise nicht belegt. Eine detailliertere Analyse und Überwachung ist nur mit spezieller Hard- und Software gegeben, z. B. einem Spektrumanalysator¹⁴⁷. Damit lassen sich auch potenzielle Störquellen detektieren, die die freien ISM-Bänder in Bereich 2,4 GHz und 5 GHz ebenfalls nutzen, aber nicht den IEEE 802.11-Standards folgen und somit nicht mit einer o. g. einfachen Messung zu erfassen sind. Ein solcher Spektrumanalysator stand allerdings nicht zur Verfügung.

Durch die kurzen Distanzen zwischen den Endgeräten und WLAN-Routern im Testaufbau der Testumgebung waren trotz der überlappungsfreien Kanalraster im 5-GHz-Band Cross-Channel-Interferenzen zu beobachten. Letztlich konnte durch den größtmöglichen Kanalabstand im unteren 5-GHz-Band und eine Reduzierung der Sendeleistung der WLAN-Router eine praktikable Lösung gefunden werden. Dies bedingte wie auch im 2,4-GHz-Band die Nutzung

¹⁴⁵ <https://play.google.com/store/apps/details?id=com.farproc.Wi-Fi.analyzer&hl=de> abgerufen am 17.03.2014

¹⁴⁶ <http://www.metageek.net/products/inssider/> abgerufen am 01.12.2012

¹⁴⁷ wie z. B. <http://www.metageek.net/products/wi-spy-wireless-spectrum-analysis/> abgerufen am 01.12.2012

von zwei Endgeräten pro Kanal. Vergleichsmessungen zeigten keine Nachteile gegenüber einer Einzelnutzung. Die Festlegung der Platzzuordnung und damit der Endgeräte, die Zuordnung zu den Netzemulatoren sowie deren Positionierung innerhalb der Laborumgebung erfolgten unter Berücksichtigung der beschriebenen Nebenbedingungen durch Austesten vor Ort.

5.7.2 Hard- und Software der Plattform zur Netzemulation

Die minimale Verzögerung durch die Paketweiterleitung zwischen der internen WLAN-Schnittstelle (Endgeräteseite) und externen Schnittstelle (zum Internet) wurde inklusive NAT auf kleiner als 1 ms ermittelt. Die Bestimmung erfolgte durch Ausleitung der Pakete direkt an den Interfaces wlan0/wlan1 und eth1 mit *tc mirred* über den isolierten Port auf eth0.2. Die Zeitdifferenz aus Empfang und Senden für das „gleiche“ Paket wurde über Zeitstempel in Wireshark gemessen. Der NAT-Durchsatz wurde in der erstellten internen Konfiguration mit bis zu 440 MBit/s ermittelt, was die prinzipielle Leistungsfähigkeit der gewählten Plattform unterstreicht (interne drahtgebundene Verbindung mit 1 GBit/s, extern 1 GBit/s).

Die in Zusammenarbeit der eingesetzten Hardware und Software erzielbare Leistung ist ausreichend, um die Netzemulation, mit den beschriebenen Grundprinzipien, in den angestrebten Parameterbereichen zu realisieren. Aus den Abschätzungen der Lastentwicklung mit aktiver Netzemulation ist erst im oberen Bereich der maximal effektiv erzielbaren Geschwindigkeiten der drahtlosen Schnittstellen mit CPU-Lastspitzen und ggf. ungeplanten Effekten zu rechnen. In diesem Bereich wurde allerdings nicht getestet, bei einer Ausweitung der Parameterbereiche (maximale Datenrate) wäre dies aber gesondert zu untersuchen.

5.7.3 Delay und Jitter

Der minimal einstellbare Wert für zusätzlichen Delay und Jitter beträgt 0. Der maximale Wert kann durchaus einige Sekunden betragen, auch wenn dies außerhalb eines sinnvollen Bereichs liegt. Für sehr große Werte ist bei gleichzeitig hoher Datenrate zu berücksichtigen, dass für die Zwischenpufferung der Pakete entsprechend viel Speicher (RAM) verfügbar sein muss.

Die zeitliche Granularität der Netzemulation für Delay und Jitter entspricht der verfügbaren Timer-Auflösung im Linux-Kernel [255]. In neueren *netem*-Versionen besteht zudem die Möglichkeit auch feiner auflösende Hardware-Timer (High Precision Event Timer – HPET) zu nutzen [256]. So können im Testbed die Parameter wie Delay und Jitter im Bereich von ≤ 1 ms Granularität gesetzt werden, wie eine kurze Überprüfung bestätigte.

Die Bewertungen der Delay-Genauigkeit unterscheiden sich in der Literatur geringfügig. Die Testmethodik aus [250] erreicht eine sehr gute Genauigkeit von weniger als 1 % Abweichung erst ab einem Zielwert von 50 ms. Diese Bewertung resultiert vermutlich aus einer niedrigen Timer-Granularität von 10 ms ohne HPET-Unterstützung. Nach [249] werden die Vorgabewerte für Delay hingegen sehr genau eingehalten. Im Schnitt sind die erzielten Werte im Rahmen der Granularität, die dort 1 ms beträgt. Stichprobenartige Tests und Kontrollmessungen im eigenen Testbed mit verschiedenen Vorgabewerten bestätigen die sehr hohe Genauigkeit der Delay-Werte.

Die Literatur bemängelt die Jitter-Genauigkeit bei gleichzeitig niedrigen Delay-Werten [250]. Dies resultiert wahrscheinlich wieder aus einer niedrigen Timer-Granularität. Für im Verhältnis ausreichend große Delay-Werte werden auch die angestrebten Varianzen durch die Jitter-Werte gut eingehalten, wenn die zeitliche Granularität fein genug ist [249]. Die Software-Pakete zu *netem* für die eingesetzte OpenWrt-Version enthalten nur die standardmäßige Verteilung „uniform“ und nicht die sonstigen eigentlich mitgelieferten Jitter-Verteilungen (Normalverteilung, Paretoverteilung etc.). Untersuchungen zum Jitter standen nicht im Vordergrund der eigenen Untersuchungen, weshalb auf umfangreiche Messungen zur Jitter-Parametrisierung verzichtet wurde. Da aber im eigenen Testbed eine hohe Delay-Genauigkeit erreicht wird, wären vermutlich auch recht gute Ergebnisse für Jitter zu erwarten.

Für typische Delay-Werte bis zu einigen hundert Millisekunden ohne zusätzlichen Jitter, wie sie im Testbed genutzt werden sollten, waren keine Probleme zu erwarten.

5.7.4 Bandbreite und Burstiness

Die Bandbreite lässt sich sehr flexibel im Bereich weniger Bit/s bis hin zu einigen MBit/s parametrisieren. Die zeitliche Granularität der Kernel-Timer beeinflusst gleichzeitig ebenfalls die Granularität der Datenrate als Quotient von Datenmenge pro Zeiteinheit. Die maximale Datenrate mit idealisierter Gleichmäßigkeit der Paketrage ergibt sich aus der Timer-Auflösung als kleinster Zeiteinheit und maximaler Datenmenge.

Eine Beispielrechnung soll dies verdeutlichen (Timer-Auflösung 1 ms bzw. 1000 Hz, max. Paketgröße von 1500 Bytes):

1000 Hz \rightarrow 1000 Pakete/s zu je 1500 Bytes \rightarrow 1500000 Bytes/s \rightarrow ca. 11700 kBit/s maximal mit idealer Gleichmäßigkeit

Höhere Datenraten und/oder kleinere Paketgrößen sind zwangsläufig immer mit Bursts verbunden, also der stoßweisen Übertragung mehrerer Pakete direkt nacheinander. Typischerweise werden nicht nur sehr große Pakete übertragen, so dass Bursts einiger Pakete auch schon bei deutlich geringeren Datenraten zu beobachten sind. Durch das Burst-Verhalten können einzelne Pakete im Vergleich zum idealen Sendezeitpunkt prinzipiell zeitlich etwas abweichen. Für die Genauigkeit der mittleren Datenrate in einem relevanten Zeitabschnitt (z. B. 1 s) hat dies aber keine entscheidende Bedeutung.

Für den zur Bandbreitenparametrisierung genutzten Hierarchical Token Bucket Filter (HTB) eng verwandten Token Bucket Filter (TBF) finden sich Ergebnisse zur Genauigkeit in [249]. Die Zielwerte werden demnach insgesamt gut eingehalten. Für kleine Pakete beträgt die Abweichung maximal $\pm 0,5\%$, für große Pakete ca. $\pm 0,1\%$. Etwas ältere Ergebnisse bescheinigen HTB eine nicht ganz so hohe aber noch immer sehr gute Genauigkeit, bei Zielwerten um 1 MBit/s ca. 1 %, bei Zielwerten darunter ca. 2 %, siehe [257]. Für typische Bandbreitenwerte, wie sie im Testbed genutzt werden sollten, waren somit keine Probleme zu erwarten, wie auch die Testmessungen bestätigten.

Die Ausprägung der Stoßartigkeit hängt zusätzlich auch mit der Bitgeschwindigkeit (Speed) der eigentlichen Datenübertragung im betrachteten Netzsegment zusammen. Bei einem sehr großen Verhältnis von Speed zu parametrisiertem Durchsatz (Throughput) ist sie sehr ausgeprägt. Bei annähernd gleicher Größenordnung werden die Bursts ausgeglichen. Dabei wäre die Auswirkung auf die Latenz bei der gleichzeitigen Beeinflussung von Datenrate und Delay ggf. zu berücksichtigen.

Delay-Vorgaben und Bandbreitenbeschränkungen können nur den Zeitpunkt des Sendens eines Paketes über ein Netzsegment verzögern, nicht aber die eigentliche Übertragung durch das Netzsegment verlangsamen. Die Bitübertragung in dem entsprechenden Netzsegment und ggf. weiteren Abschnitten danach wird im Normallfall (idealisiertes Modell) mit deutlich höherer Datenrate erfolgen (Flaschenhals in der Netzemulation). Die für die eigentliche Übertragung benötigte Zeit ist demnach kürzer als für die Datenmenge bei der parametrisierten Bandbreite im Grunde anzusetzen ist, siehe auch Latenzmodell. Dies ist beabsichtigt. Wenn solche Übertragungszeiteffekte durch Differenzen aus (Wire-) Speed und konfigurierten Bandbreitenbeschränkungen berücksichtigt werden sollen, sind die Delay-Werte entsprechend anzupassen.

5.7.5 Paketverlust

Der kleinste Wert für den Parameter *loss* in *netem*, der größer als 0 ist, wird mit 0,000000232 %¹⁴⁸ angegeben [256]. Der Maximalwert beträgt 100 %. Die Granularität entspricht dem zuvor genannten kleinsten von 0 verschiedenen Wert. In der Bewertung der Genauigkeit des Einhaltens des Zielwertes sind sich die verschiedenen Untersuchungen nicht einig. Bis 4 % Paketverlustrate wird *netem* in [250] eine Genauigkeit von 99 % bescheinigt, die bei höheren Werten für *loss* abnimmt. Auf eine Datenratenabhängigkeit weisen die Ergebnisse aus [249] hin, wonach mit steigender Datenrate die Abweichung vom Sollwert ansteigt (< 0,5 % bei 1 Mbit/s bis zu 20 % bei 75 Mbit/s).

¹⁴⁸ [255] erläutert, dass die Darstellung für *loss* als 32-Bit-Zahl erfolgt. Damit entspricht 2^{32} also 100 % (alle Pakete werden verworfen). Der angegebene Minimalwert größer Null verwirft also im Mittel 1 von 4294967296 Paketen. (Zur Einschätzung der Größenordnung: Dies entspricht 1 Paket in ungefähr 10h auf einer maximal ausgelasteten Schnittstelle mit einer Geschwindigkeit von 1Gbit/s bei einer durchschnittlichen Paketlänge von 1000 Bytes.)

Bei typischen Werten für loss ($< 5\%$) und mobilfunktypischen Datenraten sind mit *netem* gute Ergebnisse zu erwarten.

Im Testbed werden zwar in den Parameterkombinationen keine Loss-Werte explizit vorgegeben, im Rahmen der Technik bzw. in der Nachbildung der Netzcharakteristik des Mobilfunks treten Paketverluste aber trotzdem auf. Sind die Pufferwarteschlangen an dem künstlichen Flaschenhals zur Datenratenbegrenzung gefüllt, so werden weitere eintreffende Pakete verworfen (tail drops). In der Realität (echter Mobilfunk) sitzt der Flaschenhals normalerweise im drahtlosen Zugangsnetz, also direkt am Gerät. Die mobilen Geräte können daher gar nicht schneller senden, als das Netz es erlaubt. Im Testbed ist das WLAN selbst aber nicht künstlich verlangsamt, sondern der Flaschenhals befindet sich aus Sicht der mobilen Geräte erst hinter dem anderen Ende des WLAN-Segments. Die Warteschlangenlänge (in Upstream-Richtung) wird daher dynamisch in der Netzemulation zur angestrebten Datenrate angepasst, um jeweils die max. Warteschlangenzeit ungefähr konstant zu begrenzen. Dies verhindert für die Endgeräte eine sehr ungleichmäßige Sendecharakteristik mit einem kurzzeitigen Burst mit sehr hoher Datenrate, der dann durch die Netzemulation auf die entsprechende Länge bei vorgegebener Datenrate gedehnt wird. Durch die eingesetzte Begrenzung werden die entsprechenden Regelmechanismen im TCP also gezielt genutzt, um die Sendedatenrate am Endgerät auf die gewünschte Größenordnung (im Mittel) zu begrenzen. Dieser künstliche Paketverlust ist daher eventuell etwas höher als im echten Mobilfunk, wohingegen in diesem die maximalen Datenraten etwas geringer sind (auch aufgrund der notwendigen immanenten Retransmissionen). Dies ist zweifelsohne nicht die beste Lösung, aber eine praktikable, wie die Vergleiche zum Mobilfunk und dort typischen Datenraten mit den im WLAN-Testbed ermittelten möglichen Effektivwerten zeigen. Besser als solche tail drops ist vermutlich ein aktives Management der Queues mit „intelligentem“ Verwerfen von Paketen, nicht nur am Ende der Warteschlange. Eine weitere Alternative wäre die aktive Manipulation der TCP sliding window size im Rückkanal (also den ACK-Paketen im Downstream), um die Senderseite (hier die mobilen Geräte) zu bremsen, ohne Pakete zu verwerfen oder sekundenlange Backlogs in den Pufferwarteschlangen zu provozieren. Diese Alternativen konnten leider nicht realisiert werden.

In Empfangsrichtung (Downstream) konnte in den Vergleichstests im Mobilfunk ebenfalls eine erhebliche Pufferlänge festgestellt werden, bis in den Bereich mehrerer Sekunden. Die Downstream-Warteschlangen sind entsprechend ebenfalls sehr lang, allerdings nicht dynamisch, da anzunehmen ist, dass die Größe der Puffer im Mobilfunk durch die technischen Merkmale der genutzten Ausrüstung vorgegeben ist und damit eine feste Puffergröße existiert. Außerhalb des Testbeds, „im offenen Internet“, können ungeplante Paketverluste auftreten, da dort nur Best Effort ohne Zuverlässigkeitsgarantien genutzt wird.

5.7.6 Umschaltgeschwindigkeit der Parametrisierung

Die Umschaltung der Parameter per Web-API benötigt inklusive Status-Response ca. 200 ms. Für die angestrebten Tests mit Umschaltung zwischen statischen Konfigurationen ist dies ausreichend schnell. Theoretisch wären über die Web-API somit bis zu 5 konsekutive Parameteränderungen pro Sekunde möglich. Die recht lange Zeit ergibt sich aus dem gewählten Ansatz mit einer Implementierung als LUCI-Modul. Dabei wäre auch zu beachten, dass die genutzten „Komfortfunktionen“ von LUCI recht CPU-lastig sind.

Prinzipiell lässt sich mit einer anderen Implementierung, ohne LUCI-Nutzung, der Overhead reduzieren und die Web-API beschleunigen. Noch wesentlich schneller ist die direkte Manipulation der Netzemulationsparameter über die *tc*-Userspace-API, z. B. in einem Shell-Skript. Zusammen mit einer geeigneten zeitlichen Steuerung könnten sich Verläufe für eine einfache Art Dynamik nachbilden lassen. Noch schnellere Wechsel mit nahezu Echtzeitverhalten erfordern Eingriffe direkt in den Modulen zur Netzemulation. Zusammen mit aufgezeichneten dynamischen Charakteristiken echter Mobilfunknetze würde gleichzeitig der Grad an Realismus weiter steigen. [258] erläutert dazu das prinzipielle Vorgehen und die nötigen Änderungen in *tc+netem*. Für die praktische Nutzbarkeit in der Probandenstudie zum QoS-QoE-Zusammenhang müssten aber noch geeignete Mechanismen vorgesehen werden, die innerhalb einer Aufgabenausführung, z. B. ausgelöst durch bestimmte Ereignisse, die Dynamik steuern. Diese Ereignisse selbst könnten auf den Endgeräten oder im Datenstrom detektiert werden. Eine weitere Alternative wäre der Wechsel des Netzemulators, da das bereits erwähnte KauNet sowohl aufgezeichnete Cha-

rakteristiken als auch spezielle Trigger unterstützt, siehe [259]. Derartige Ansätze zu einer Dynamik blieben möglichen zukünftigen Erweiterungen des Untersuchungsgegenstandes vorbehalten.

5.7.7 Effektivwerte

Neben den theoretischen Betrachtungen zur Leistungsfähigkeit erfolgten Ende-zu-Ende-Messungen mit gängigen populären Tools direkt auf den mobilen Endgeräten. Dabei lag der Fokus allerdings nicht auf möglichen Extremwerten. Die Messungen sollten einen Hinweis auf die praktisch erreichbaren Effektiv- bzw. Istwerte geben, bezogen auf die parametrisierten Sollwerte innerhalb der vorab ausgewählten Parametergrenzen. Entsprechend der Begriffsdefinitionen zu Bandbreite und Durchsatz ist zu beachten, dass sich die Parametrisierung des Wertes *bandwidth* in dem genutzten HTB-Filter auf die Bruttodatenrate auf OSI-Schicht L2 bezieht. Entsprechender Overhead durch Ethernet-Protokoll, IP, TCP und HTTP etc. sind bei eigenen Vergleichsmessungen zu berücksichtigen. Die Grundlatenz der Netzsegmente außerhalb der Netzemulation addiert sich mit den parametrisierten Werten für *delay* zur ermittelten Gesamtlatenz. Beispielergebnisse mit den genutzten Apps zur Messung sind im Anhang A.2 aufgeführt. Die erzielten Werte entsprechen dabei dem, was zu erwarten war. Die Netzemulation funktioniert auch in Kombination der Netzparameter wie gewünscht. Eine Übernahme der erzielten Ergebniswerte für andere Dienste ist von verschiedenen Faktoren abhängig. Die Messwerte sind deshalb nur eine grobe Orientierung zu den erreichbaren Effektivwerten in anderen Diensten.

5.8 Die mobilen Endgeräte im Testbed

In der Studie sind die Endgeräte eine potenziell bedeutsame Größe. Daher sollen die Auswahlkriterien wie auch die tatsächliche Zusammenstellung für die Nutzung im Testbed vorgestellt werden. Dabei ergeben sich einige technische Besonderheiten im Zusammenspiel mit der Netzemulation und zur Netz Wahrnehmung auf den Endgeräten aus Nutzersicht. Es wird eingangs jedoch zunächst kurz auf die Hardware und die mobilen Softwareplattformen eingegangen.

5.8.1 Auswahlkriterien zu den mobilen Endgeräten

Die Endgeräte sollten im Test als echte mobile, d. h. drahtlos genutzte, Geräte vertreten sein, um einen möglichst echten Interaktionskontext zu schaffen. Die Nutzung in einem Emulator, z. B. auf einem PC, oder gar eine komplette Substitution durch einen PC wäre nicht zielführend gewesen und schied damit in den Überlegungen frühzeitig aus. Für die Auswahl der Endgeräte waren neben technischen und testplanerischen Aspekten auch Markt- und Relevanzaspekte zu berücksichtigen.

Als **technische Aspekte** ergaben sich folgende Anforderungen an die Endgeräte:

- Unterstützung von WLAN, bevorzugt auch von 5-GHz-WLAN für größere Flexibilität bei der WLAN-Ressourcenzuweisung im Testbed
- aber keine reinen WLAN-Geräte, sondern Geräte mit Mobilfunkunterstützung für etwaige Vergleichstests im Rahmen der Voruntersuchungen
- aktuelle Versionen der Betriebssoftware für eine möglichst breite App-Unterstützung
- Verfügbarkeit aller (möglicherweise) relevanten Apps

Für die **testplanerischen Aspekte** war zu berücksichtigen, dass die Endgeräte als Teil der QoS-QoE-Faktoren auf die QoE-Bewertungen einwirken können. Zur Untersuchung dieses Einflusses sollte die Geräteauswahl bewusst heterogen erfolgen. Der Einfluss der mobilen Plattform, des Formfaktors (Smartphone und Tablet) und der Leistungsklasse sollten gegebenenfalls zu ermitteln sein. Durch die Heterogenität der Zusammenstellung der Endgeräte wird der Parameterraum zum einen direkt vergrößert, zum anderen aber auch indirekt

durch unterschiedliche Apps für die ausgewählten Dienste, die jeweils zu den Endgeräten passen müssen. Damit ergaben sich die folgenden Einschränkungen:

- Beschränkung auf zwei mobile Plattformen (Aufwand in den Voruntersuchungen und Vorbereitungshandlungen, Definition der Aufgaben und Aspekte der statistischen Auswertung)
- Beschränkung auf einen Hersteller je mobiler Plattform, da selbst innerhalb einer Plattform und Geräteklasse Unterschiede existieren können, da die Hersteller teilweise Veränderungen vornehmen oder Grundeinstellungen anders wählen (relevant für Android)
- durch die Kapazität der Testumgebung und den Aufbau des Testbeds ist die Anzahl gleichzeitig nutzbarer Endgeräte auf sechs begrenzt

Auch **Markt- und Relevanzaspekte** sollten in der Auswahl berücksichtigt werden:

- Nutzung der mobilen Plattformen mit den größten Marktanteilen, was zu Android und iOS führte. Andere mobile Plattformen erreichen nur eine deutlich geringere Bedeutung.
- jeweils Auswahl populärer Marken und Modelle von Smartphones und Tablets in drei Leistungsklassen
- durch eine hohe Marktrelevanz sollte eine gewisse Repräsentativität der Endgeräteauswahl erreicht werden

5.8.2 Zusammenstellung und Auswahl der Endgeräte für das Testbed

Die Endgeräte lassen sich grundsätzlich nach Displaygröße in verschiedene Formfaktoren einteilen, wobei Sonderformen, z. B. Convertibles, möglich sind (Smartphone: $\leq 5''$, Phablet (Phone/Tablet): $5''$ bis $7''$, Tablet: $7''$ bis $10''$, Exoten auch größer). Für die Definition und Abgrenzung der Geräteklassen wurden relevante Kriterien und typische Werte recherchiert, die hauptsächlich die Ausstattung und damit die (theoretische) Geräteleistung umfassen, siehe dazu Tabelle 29.

Tabelle 29: Die mobilen Endgeräte des Testbeds der Testrunde 2 mit Testplatzzuordnung und technischen Eckdaten

Gerät Hersteller/ Modell	Test- platz Nr.	Plattform OS/Ver.	Klasse	Format Display- größe	WLAN	CPU Kerne/Takt	RAM MB
Samsung Galaxy Tab 2 10.1	1	Android 4.0.4	Mittel	Tablet 10"	2,4 GHz	2x1 GHz	1024
Samsung Galaxy Ace 2	2	Android 2.3.6	Einstei- ger	Smart- phone 3,8"	2,4 GHz	2x832 MHz	768
Samsung Galaxy S III	3+6	Android 4.1.2	Spitze	Smart- phone 4,8"	2,4 GHz + 5 GHz	4x1,4 GHz	1024
Apple iPad 4	4	iOS 6.1.3	Spitze	Tablet 10"	2,4 GHz + 5 GHz	2x1,4 GHz	1024
Apple iPhone 5	5	iOS 6.1.3	Spitze	Smart- phone 4"	2,4 GHz + 5 GHz	2x1 GHz	1024

Zusammenstellung der technischen Daten nach den zugehörigen Angaben der Hersteller

Die Eingrenzung der Auswahl und Festlegung der mobilen Endgeräte für das Testbed erfolgte in enger Abstimmung der Projektpartner. In der Studie sollten Geräte der Formfaktoren Smartphones und Tablets vertreten sein. Android und iOS wurden aufgrund ihrer Marktrelevanz (vgl. Tabelle 37) als mobile Plattformen gewählt. Neben der Hardware-Ausstattung kann auch die Aktualität der Systemsoftware als (Klassen-) Unterscheidungsmerkmal dienen. So bringen bei Android meist nur neue Spitzenmodelle die neueste Version der Systemsoftware mit. Einige Topmodelle der Vorgängergeneration erhalten vom Hersteller eventuell noch ein Update, teilweise aber stark verzögert. Andere, auch aktuelle Modelle, hinken meist zumindest eine Version hinter der Entwicklung her, bzw. erhalten aufgrund der raschen Produktzyklen überhaupt keine Updates mehr [260]. Ein Überblick zum Versionsstand und Verbreitung der zum Zeitpunkt der Testplanung verfügbaren Android- und iOS-Versionen findet sich in [260], ein Auszug in Tabelle 30. Entsprechend konnte die Auswahl der Geräte zum Zeitpunkt der Testplanung ausgehend vom Software-Stand als recht repräsentativ angesehen werden.

Tabelle 30: Versionen der führenden mobilen Plattformen

Plattform	Version	Verfügbarkeit im Markt
Android	2.3	2010-12
	4.0	2011-11
	4.1	2012-07
	4.2	2012-11, aktuelle Version zum Zeitpunkt der Testplanung, allerdings kaum verbreitet
iOS	5.x	2011-10
	6.x	2012-09, aktuelle Version zum Zeitpunkt der Testplanung mit weiter Verbreitung

Angaben bis zum Beginn der Testplanung, zusammengestellt nach [260] S. 72f



Abbildung 50: Die mobilen Endgeräte des Testbeds

Das iPhone 5 als iOS-Smartphone und das iPad 4 als iOS-Tablet waren als iOS-Spitzengeräte zum Auswahlzeitpunkt direkt gesetzt. Im Gegensatz zu Android werden bei iOS meist auch ältere Gerätegenerationen noch mit aktueller Systemsoftware versorgt, so dass nur die zum Zeitpunkt der Testplanung aktuellste iOS-Version berücksichtigt wurde. Im Vergleich zu iOS ist die Android-Gerätelandschaft extrem fragmentiert [261]. In der Testrunde 2 wurden bei Android ausschließlich Samsung-Geräte ausgewählt. So sollten zu-

mindest die herstellerspezifischen Unterschiede in der Anmutung und Wahrnehmung minimiert werden, bzw. die Beschreibungen zur Aufgabenausführung so weit möglich vereinheitlicht werden. Die Wahl fiel auf Samsung, da Marktführer im Bereich Android-Smartphones im Jahr 2013 [262]. Die ausgewählten Samsung-Geräte waren jeweils ein typischer Vertreter der jeweiligen Klasse im Markt. Damit ergab sich die endgültige Zusammenstellung marktüblicher und populärer Endgeräte für die Nutzung durch die Probanden, siehe Tabelle 29. Die Endgeräte wurden jeweils einem Testplatz fest zugeordnet und erhielten eine eindeutige Kennzeichnung, siehe Abbildung 50.

5.8.3 TCP-Implementierungen der Endgeräte

Angaben zu den TCP-Implementierungen der populären mobilen Plattformen wie Android oder iOS fehlen in den aufgeführten Vergleichen (wie in [155]) leider. Einen Hinweis kann ggf. die Herkunft dieser Plattformen liefern (Linux bei Android¹⁴⁹ bzw. BSD/Nextstep/Darwin/Mac OS X bei iOS¹⁵⁰). [148] geht zwar auf die Android- und iOS-Versionen des Jahres 2012 ein (Vorgänger der Versionen im Test), nennt aber nicht im Detail die mobilen TCP-Implementierungen. Zumindest werden aber ungefähre Richtwerte für den wichtigen Parameter der Receive-Window-Size genannt. Demnach sind für die dort untersuchten Android-Geräte je nach Modell feste obere Grenzen des TCP-Receive-Window in Abhängigkeit der genutzten Netztechnologie vorkonfiguriert (z. B. unterschiedlich für WLAN, UMTS, HSDPA+ oder LTE). Die Autoren von [148] bezeichnen das beobachtete Verhalten der Endgeräte als TCP-Flat, da ein fester oberer Wert bei der Darstellung der Skalierung der TCP-Graphen zu einer flachen Linie führt. Aus den unterschiedlichen Vorgabewerten können sich in typischen Mobilfunksituationen mit vergleichsweise hoher Latenz und Bufferbloat (Long-Fat-Networks¹⁵¹) unterschiedliche Effektivwerte der Übertragung ergeben. Dieses TCP-Flat-Verhalten konnte in stichprobenartig ausgeführten Analysen der aufgezeichneten Verkehrsdaten für die im Test eingesetzten Endgeräte ebenso wie leicht unterschiedliches TCP-Verhalten insgesamt bestätigt werden.

¹⁴⁹ http://de.wikipedia.org/wiki/Android_%28Betriebssystem%29 abgerufen am 11.09.2013

¹⁵⁰ http://de.wikipedia.org/wiki/Apple_iOS abgerufen am 11.09.2013

¹⁵¹ bezogen auf ein großes Bandwidth-Delay-Produkt, im RFC 1072 sind Erläuterungen gegeben in [263] S. 1

5.8.4 Nutzung der Endgeräte im WLAN-Testbed

In den Voruntersuchungen zeigten sich unerwartete Probleme mit iPhone 5 und iPad 4 im WLAN-Testbed im Zusammenhang mit der Netzemulation. Ab ca. 30 ms Delay-Summe in der Parametrisierung sank die effektive Download-Datenrate stark unter den Vorgabewert ab, während die Leerlauf Latenz sprunghaft anstieg. Unterhalb o. g. Delay-Summe erreichten die Geräte die erwarteten Effektivwerte hingegen problemlos. Mit den Android-Geräten und älterer iOS-Hardware (iPhone 4s mit gleicher iOS-Version) traten die Probleme nicht auf.

Mittels systematischer Variation der Parameterwerte bei gleichzeitiger Auswertung der Verkehrsdaten in der Netzemulation (siehe 5.9.1) und mobilem Endgerät (siehe 5.9.3) konnte die Ursache eingegrenzt werden. Offensichtlich wird bei WLAN-Nutzung ein Stromsparmodus aktiviert, wenn mehr als 30 ms lang kein Paket transportiert wurde. Dann können Pakete in Downstream-Richtung jeweils erst nach dem Aufwachen des Funkmoduls im Endgerät zugestellt werden. Dieses Aufwachen erfolgt den Beobachtungen nach aller 90 bis 100 ms. Diese zusätzliche Zeit ist in der Netzemulation nicht berücksichtigt. Beim Aufwachen sind zusätzlich auf der Seite des Endgerätes einzelne Paketverluste zu verzeichnen, d. h. gemäß den Aufzeichnungen werden die Pakete vom WLAN-Router korrekt gesendet aber auf dem Endgerät nicht empfangen. Zusammen mit einem konservativen TCP-Verhalten führt dies zum „WLAN mit angezogener Handbremse“, weil das Endgerät dann ständig zwischen Normalmodus und WLAN-Schlafmodus pendelt. Die zum Vergleich herangezogenen anderen Endgeräte wechseln zwischenzeitlich zwar teilweise auch in einen WLAN-Schlafmodus, dort funktioniert dies aber ohne Probleme.

Um die iOS-Geräte nicht von der Studie im WLAN-Testbed ausschließen zu müssen, wurde nach einer passenden Lösung gesucht. Das beschriebene Verhalten lässt sich über lokale Einstellungen direkt auf den iOS-Geräten etc. nicht verändern. Es musste also von der Netzseite aus eine Gegenmaßnahme gefunden werden. Erprobt wurden eine Verkürzung des Polling-Intervalls des WLAN-Access-Points und eine Art Dauer-Ping. Dies sollte helfen, die Endgeräte in kürzeren Abständen aufwachen zu lassen, bzw. die Geräte am Wechsel in den Stromsparmodus hindern. Die Wahl fiel letztlich auf die Software

fping¹⁵². Diese Variante des bekannten „ping“ erlaubt eine flexibel angepasste Frequenz der Anfragen im Bereich von Millisekunden. Die Software konnte problemlos unter OpenWrt auf den Netzemulatoren zum Einsatz gebracht werden. Sie verhindert den Wechsel der Geräte in den Stromsparmodus zuverlässig. Dabei sind nach eigenen Beobachtungen keine sonstigen Nebeneffekte auf Endgerät, WLAN-Router und im WLAN allgemein zu beobachten. Die dafür zusätzlich notwendigen Pakete im WLAN benötigen nur wenige kBit/s Datenrate je Gerät und Richtung. Diese werden an der Netzemulation vorbeigeleitet. Mit diesem konstanten „Wachhalten“ der Endgeräte liegen die variierten Netzparameter wieder gut im erwarteten Bereich, siehe Abbildung 51 zum Vorher-Nachher-Vergleich einmal ohne bzw. mit Dauer-Ping. Um mögliche Verzerrungen innerhalb der Endgeräteauswahl zu verhindern, wurde die fping-Lösung auch für die Android-Geräte genutzt. Eine möglicherweise schnellere Entleerung der Akkus der Mobilgeräte ist ohne weitere Bedeutung, da die Geräte zwischen den Testslots jeweils wieder nachgeladen wurden.

5.8.5 Netz Wahrnehmung auf den Endgeräten im Testbed

Ort und Art der Sichtbarkeit der Anzeigeelemente zur Netztechnologie und zum Netzbetreiber unterscheiden sich zwischen den mobilen Plattformen im Test, siehe Tabelle 31. Die Geräte sind dabei in ihren Hinweisen entweder eher dezent, z. B. beim Ace 2 und iPad 4, oder sehr direkt und eindeutig, wie beim iPhone 5, das permanent deutlich auf eine fehlende SIM-Karte in der iOS-Statusleiste hinweist, siehe Abbildung 52 und Abbildung 53. Die prominente Darstellung einer bestimmten Mobilfunktechnologie bei gleichzeitiger Nutzung von WLAN im Testbed als Zugang zum Internet ist nicht möglich. Selbst für einen mit dem jeweiligen System nicht vertrauten Nutzer ist damit prinzipiell permanent erkennbar gewesen, dass innerhalb der Testumgebung des Testbeds nicht wirklich per Mobilfunkverbindung auf das mobile Internet zugegriffen wird.

¹⁵² <http://fping.sourceforge.net/> abgerufen am 01.05.2013

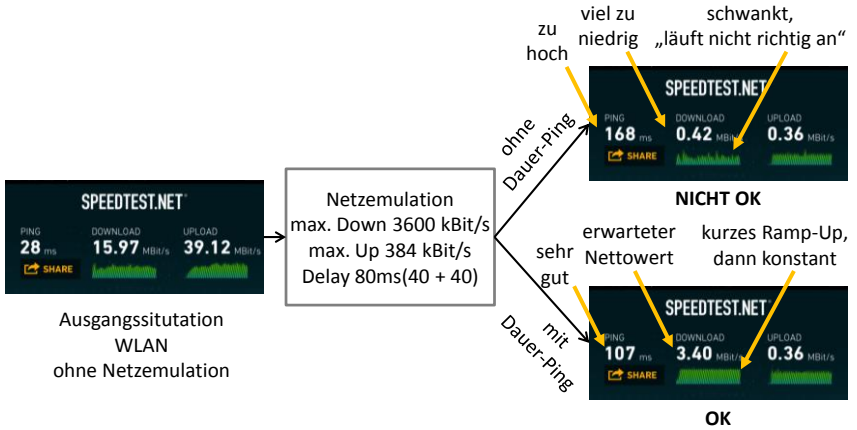


Abbildung 51: Wirkung mit Dauer-Ping im WLAN-Testbed zur Verbesserung der Einhaltung der Zielwerte

Beispielwerte aus der App von Speedtest.net (Screenshots, Ausschnitte zur Verdeutlichung)



Abbildung 52: Wahrnehmbare Netzeigenschaften in der System- bzw. Statusleiste im Testbed mit WLAN

Sicht innerhalb der Testumgebung mit aktiver Datenverbindung über WLAN (Screenshots, Ausschnitte zur Verdeutlichung)

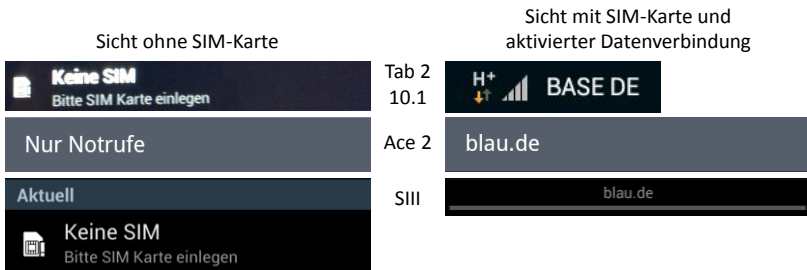


Abbildung 53: Hinweise und Anzeigen zur fehlenden SIM-Karte bei den Android-Geräten im Benachrichtigungsbereich

(Screenshots, Ausschnitte zur Verdeutlichung)

Tabelle 31: Ort und Art der Sichtbarkeit der Anzeigen zur Netztechnologie und zum Betreiber

Ort der Sichtbarkeit	Art der Sichtbarkeit	Sichtbarkeit	Anzeige
Systemleiste (Android) Statusleiste (iOS)	permanent	hoch	Netztechnologie, Netzbetreiber, im Test: WLAN, fehlende SIM-Karte
Benachrichtigungsbereich (Android)	Einblendung bei Bedarf	mittel	WLAN, fehlende SIM-Karte
Sperr-Bildschirm (Android)	beim Entsperrern des Gerätes	mittel	fehlende SIM-Karte
Einstellungsmenü bzw. Unterseiten	Einblendung bei Bedarf	niedrig	fehlende SIM-Karte
Geräteoptionen (Android)	Einblendung bei Bedarf	niedrig	fehlende SIM-Karte

5.9 Überwachung und Aufzeichnung des Netzverkehrs

Zum Verständnis der Netznutzung der verschiedenen Dienste waren Möglichkeiten zur Überwachung (live in Echtzeit) und Aufzeichnung (zur späteren Analyse) notwendig. Darüber hinaus war eine vollständige passive Kontrolle des gesamten Testgeschehens im laufenden Betrieb wünschenswert. Dafür wurden technische Lösungen innerhalb des Testbeds geschaffen und erfolgreich erprobt.

5.9.1 Verkehrsausleitung zur Remote-Überwachung

In *tc* besteht die interessante Möglichkeit bei der Klassifikation per *tc*-Filterregel selektierte Pakete über action¹⁵³-Module einer weiteren Verarbeitung zuzuführen. Eines dieser action-Module ist *mirred*¹⁵⁴. Im Modul *mirred* können Pakete gespiegelt (*mirror*) oder umgeleitet (*redirect*) werden. Da die Pakete noch ihr ursprüngliches Ziel erreichen sollen, ist *mirror* die richtige Wahl, um Kopien zu erzeugen, die an eine separate Schnittstelle zum Traffic-Monitoring/Recording gesendet werden, während die ursprünglichen Pakete unbeeinflusst weiterverarbeitet werden. Über die *tc*-Filterregeln kann der tatsächlich interessante Netzverkehr zwischen den Endgeräten und dem Internet gezielt

¹⁵³ <http://code.metager.de/source/xref/linux/utils/iproute2/doc/actions/actions-general> abgerufen am 01.12.2012

¹⁵⁴ <http://code.metager.de/source/xref/linux/utils/iproute2/doc/actions/mirred-usage> abgerufen am 01.12.2012

ausgeleitet werden, während der uninteressante lokale Netzverkehr nicht gespiegelt wird (o. g. Dauer-Ping und die später noch beschriebene Videoüberwachung).

Die Verkehrsausleitung ist optional und kann flexibel und unabhängig von der restlichen Netzemulation gesteuert werden, erfordert aber eine Anpassung der *tc*-Konfiguration. Die Ausleitungspunkte sind prinzipiell flexibel wählbar: vor oder nach der Netzemulation sowie wiederum für jede Richtung (Senden/Empfangen) und Netzchnittstelle getrennt. Nicht alle Kombinationen sind sinnvoll, aber z. B. zur Überprüfung der internen Funktionsweise durchaus recht interessant. Wird beispielsweise eine Richtung gleichzeitig vor und nach der Netzemulation überwacht, so kann die interne Arbeitsweise der Netzemulation genau überprüft werden. Die aktive Verkehrsausleitung und die zugehörigen *tc*-Filterregeln sind leider nicht ganz umsonst, d. h. mit zusätzlicher CPU-Last verbunden, die mit steigender Paketzahl wächst. Daher sollten die Ausleitungspunkte auch unter diesem Gesichtspunkt mit Bedacht gewählt werden. Für den produktiven Betrieb werden die Ingress- bzw. Egress-Queues der WLAN-Schnittstellen gespiegelt, siehe Abbildung 54.

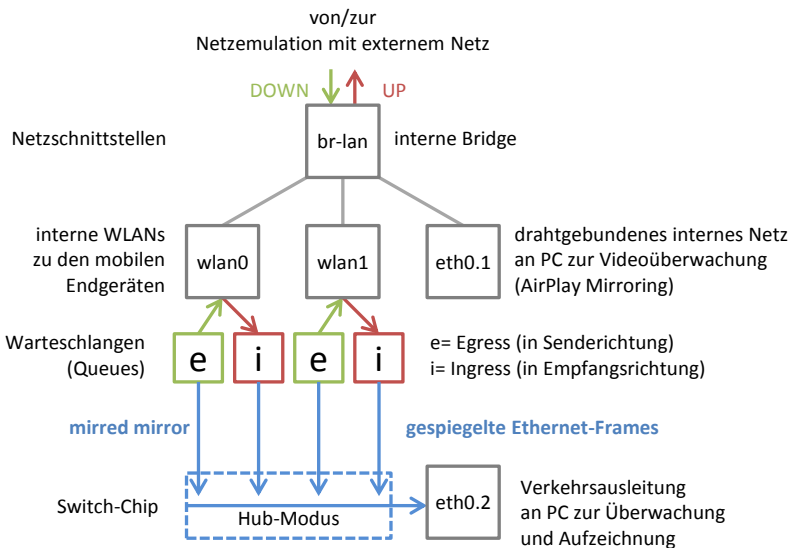


Abbildung 54: Positionierung der optionalen Verkehrsausleitung an den Netzchnittstellen
siehe auch Abbildung 44 für den relativen Bezug zur Netzemulation

In Richtung UP werden so die Pakete vor der Netzemulation ausgeleitet, also quasi vor dem Senden über das variierte Netzsegment, für die Richtung DOWN hingegen werden die Pakete nach der Netzemulation ausgeleitet, also quasi nach dem Empfangen über das variierte Netzsegment. Die gewählte Kombination zur Ausleitung des empfangenen bzw. gesendeten Netzverkehrs entspricht damit der Sicht der Endgeräte auf das Netz aus dem Gesamtmodell mit Ausnahme des Netzabschnitts des lokalen WLAN.

Die dem VLAN eth0.1 zugeordneten Ports des internen Switch-Chips sind über die Netzbrücke br-lan im internen Netz zu erreichen, während der eth0.2 zugeordnete Port nicht Teil des L3-IP-Routings ist. Dieser Port wurde für das Ausleiten des selektierten Netzverkehrs reserviert. Zum Anschluss an diesen Mirror-Port für die Remote-Nutzung zur Live-Überwachung und Aufzeichnung ist ein leistungsfähiger dedizierter PC empfehlenswert. Dieser sollte für diesen Zweck eine isolierte Netzwerkschnittstelle nutzen, um mögliche Paketverluste etc. in der weiteren Verarbeitung zu minimieren und zusätzlichen, selbst verursachten Netzverkehr zu vermeiden. Auf dem PC kann eine Software zur Netzverkehrsanalyse und -Aufzeichnung, wie z. B. *Wireshark*¹⁵⁵, eingesetzt werden. Ergänzende Hinweise zur Analyse der so erfassten Verkehrsdaten sind im Abschnitt 8.1.1 zu finden.

Gegenüber einer lokalen Aufzeichnung direkt auf den Netzschnittstellen der zur Emulation genutzten WLAN-Router, z. B. per *tcpdump*¹⁵⁶, ist die eingesetzte Lösung zur gezielten Ausleitung des Netzverkehrs vor allem im Rahmen der Voruntersuchungen vorteilhaft gewesen. So konnte quasi interaktiv mit den untersuchten Diensten auf den Mobilgeräten experimentiert werden, um parallel dazu live die korrespondierenden Netzaktivitäten zu analysieren.

¹⁵⁵ Wireshark ist eine Software zur Aufzeichnung und Analyse von Netzwerkverkehr, siehe auch <http://www.wireshark.org/> abgerufen am 28.11.2012

¹⁵⁶ <http://www.tcpdump.org/> abgerufen am 30.12.2014

5.9.2 Deep-Monitoring – der tiefe Blick

Viele der beobachteten Datenübertragungen erfolgen verschlüsselt. Dies verhindert den wirklich tiefen Blick in das Geschehen der Netzübertragungen. Teilweise ist dieser tiefe Blick aber wünschenswert, z. B. um bestimmte Funktionen, konkrete Daten oder API-Versionen nachvollziehen zu können. Im Fall von HTTP mit SSL oder TLS als HTTPS bietet sich eine Man-in-the-Middle-Analyse, in Anlehnung an eine gleich strukturierte Angriffsmethode, auf die verschlüsselten Datenverbindungen an. Dabei werden die Ende-zu-Ende verschlüsselten Daten durch Einbeziehung einer dritten Partei analysierbar, da diese zwischen mobilem Gerät und Server jeweils die Rolle des Gegenübers einnimmt und so selbst ver- und entschlüsselt. [264] beschreibt dieses Vorgehen und nutzbare Werkzeuge dazu recht anschaulich.

In gleicher Art wurde im Testbed in einzelnen Fällen im Rahmen der Voruntersuchungen experimentiert. Dazu wurde als einfach zu handhabendes Werkzeug der Charles Web Debugging Proxy¹⁵⁷ eingesetzt, auf den der HTTP-Verkehr aus dem Testbed umgelenkt wurde. Solche Lösungen sind in gewisser Art invasiv, da der Datenverkehr nicht mehr Ende-zu-Ende zu beschreiben ist und sich ggf. sogar andere technische Effektivwerte ergeben. Für den permanenten Einsatz zur Aufzeichnung im regulären Testbetrieb eignen sich diese Lösungen damit nicht, da sie potenziell eine weitere Störquelle sind.

Zusatznutzen und weiterer Erkenntnisgewinn durch das Deep-Monitoring waren in den Voruntersuchungen relativ gering. Die Interpretation des Netzverkehrs, inklusive verschlüsselter Anteile, ist meist ausreichend bzw. zu bevorzugen, weil eben nicht invasiv und somit nicht verzerrt. Im Allgemeinen sind die Verkehrsmuster mit relevanten Datenmengen und zeitlicher Charakteristik interessanter als genaue Request-URL oder einzelne Parameter. Neben der Sicht „von innen“ ist zur Analyse und Bewertung des Netznutzungsverhaltens verschiedener Dienste und Apps also auch der „tiefe Blick“ möglich, aber nicht unbedingt nötig.

¹⁵⁷ <http://www.charlesproxy.com/> abgerufen am 06.01.2014

5.9.3 Monitoring auf den Endgeräten

In den Voruntersuchungen sollte die Netznutzung der Apps per WLAN und per Mobilfunk verglichen werden. Daher war ergänzend zu den schon beschriebenen Möglichkeiten im WLAN-Testbed auch für die Analyse der Mobilfunk-Netznutzung eine Lösung zu finden. Zwischen der Bestimmung der Netznutzung bei WLAN bzw. über echte Mobilfunknetze bestehen deutliche Unterschiede.

WLAN: Eine einfache Erfassung der Verkehrsdaten im WLAN-Testbed ist außerhalb der mobilen Endgeräte von der Netzseite über die genutzte Technik der WLAN-Router möglich, siehe oben. Es sind keine Änderungen bzw. zusätzlichen Programme auf den mobilen Geräten notwendig. (Ungeachtet dessen wäre es ebenso möglich, die bei Mobilfunk genannten Wege auch für WLAN zu nutzen, welche allerdings weniger komfortabel sind.)

Mobilfunk: Es ist keine Erfassung der Verkehrsdaten von außerhalb der mobilen Endgeräte möglich. Die Erfassung muss deshalb direkt auf den Endgeräten bzw. darüber erfolgen. Dazu sind zusätzliche Apps bzw. externe Komponenten (Hardware/Software) notwendig. Für Android wurde mit *Shark for Root*¹⁵⁸ eine App-Lösung auf den Endgeräten gewählt, die *tcpdump* für Android beinhaltet und die Verkehrsdaten direkt auf dem Endgerät in pcap-Dateien speichern kann. Die Dateien wurden nach erfolgreicher Aufzeichnung zur Analyse auf einen PC transferiert. Auf den iOS-Geräten wurde die Möglichkeit eines *Remote Virtual Interface* für das „iOS Packet Tracing“ in Kombination mit einem Apple Macintosh-Computer genutzt, siehe [265]. Dabei werden Netzwerkpakete, die das mobile Gerät über die Mobilfunkschnittstelle empfängt oder versendet, über die USB-Schnittstelle des mobilen Gerätes auf einen angeschlossenen Mac gespiegelt und dort über eine virtuelle Netzschnittstelle per Hilfsprogramm aus der *XCode*¹⁵⁹-Entwicklungsumgebung für die Aufzeichnung, z. B. mit *tcpdump*, bereitgestellt. Die so gewonnenen Aufzeichnungen wurden ebenfalls zur Analyse auf einen PC transferiert.

¹⁵⁸ <https://play.google.com/store/apps/details?id=lv.n3o.shark&hl=de> abgerufen am 12.12.2013

¹⁵⁹ <https://developer.apple.com/xcode/> abgerufen am 12.12.2013

5.10 Überwachung und Aufzeichnung der Endgeräte per Video

Zum Verständnis der QoS-QoE-Zusammenhänge ist neben der technischen Sicht von innen auf die Netznutzung auch die Sicht der Nutzer von außen wichtig. Innerhalb des Testbeds sollte deshalb die Möglichkeit geschaffen werden, die Geräteeingaben und -ausgaben audiovisuell zu überwachen. Die Nutzung sollte im Rahmen der Voruntersuchungen, zur Dokumentation der Studie wie auch im laufenden Live-Testbetrieb möglich sein. Dafür werden zunächst die Anforderungen, technischen Möglichkeiten und kritischen Punkte diskutiert. Danach folgt die Vorstellung einiger erprobter konkreter Lösungen zur Nutzung unter Android bzw. iOS auf den im Testbed zum Einsatz gebrachten Endgeräten.

5.10.1 Anforderungen, prinzipielle Möglichkeiten und Einschränkungen

Es wurden Anforderungen an eine mögliche Lösung gestellt, die sich wie folgt zusammenfassen lassen:

- leichte Integration ins Testbed und einfache Nutzung im Testbed
- ohne Beeinträchtigung der Leistung der mobilen Geräte und ohne Verzerrung des gesuchten QoS-QoE-Zusammenhangs (technische Sicht)
- kein Störfaktor für die Probanden (Nutzersicht)
- Möglichkeit zur Aufzeichnung für eine zeitversetzte Auswertung
- Live-Überwachung als Bonus

Für die technische Realisierung sind prinzipiell mehrere Software- bzw. Hardware-basierte Ansätze möglich, die nachfolgend nur kurz umrissen werden sollen.

- 1) Erfassung von außerhalb und unabhängig von den Geräten, Abfilmen des Bildschirms:
 - Erzeugung mit einer Webcam (oder einer höherwertigen Action-Cam)
 - Vorteile: technisch einfach und nichtinvasiv auf den Endgeräten
 - Nachteile: Hand im Bild, ggf. Schärfepumpen, Probleme mit Reflexionen auf Bildschirm, Kontrast und Überblendungen, ...

- Herausforderung: Positionierung von mobilem Gerät und Kamera darf nicht verändert werden
 - Einschätzung: mit deutlichen Nachteilen verbunden, in der Nutzung mit Probanden nicht praktikabel
- 2) Lokale Erfassung mit Aufzeichnung direkt auf dem Endgerät:
- als Video oder als schnelle Screenshot-Folge (z. B. im Abstand von 1 bis 2 Sekunden)
 - prinzipiell möglich, praktisch aber durch Einschränkungen meist auf Geräte mit Root-Rechten bzw. Jailbreak beschränkt, da das automatisierte Erfassen der Bildinhalte per App ein mögliches Risiko darstellt (Spyware)
 - als reine Software-Lösung ggf. mit Leistungseinbußen verbunden
 - keine Erweiterungen oder Änderungen im Testbed außerhalb der mobilen Geräte notwendig
 - Einschätzung: nicht wirklich attraktiv, da viele Einschränkungen in der praktischen Nutzung
- 3) Lokale Erfassung auf dem Gerät und Ausleitung zur externen Verarbeitung in Echtzeit:
- Erzeugung mit zusätzlicher Software oder mit integrierten Bordmitteln
 - auf den mobilen Geräten direkt oder mit zusätzlich angeschlossener Hardware
 - kabelgebundene oder kabellose Weiterleitung
 - als Videosignal oder Paketdatenstrom
 - Empfang und externe Weiterverarbeitung auf PC oder mittels spezieller Hardware
 - Einschätzung: guter Kompromiss

Der Ansatz 1 wurde in der Testrunde 1 mit einer festen Kamera auf einem Tischstativ zur nachträglichen Dokumentation typischer Testläufe erprobt, konnte dabei allerdings nicht überzeugen. Durch spezielle Haltekonstruktionen könnten Kamera und mobiles Endgerät so zueinander fixiert werden, dass der Bildschirm immer erfasst wird. Solche Lösungen sind allerdings mit Nachteilen für normale Nutzer verbunden (unhandlich, Sichteinschränkungen), die

nicht vertretbar sind. [266] zeigt die Verbindung von Ansatz 2 und 1 mit Software für lokales Screen-Recording und Kameras, die die Probanden um den Hals vor der Brust hängen haben und so mehr oder weniger gut den Sichtbereich des Probanden inklusive mobilem Endgerät erfassen. Als verlockende Option auf ähnlicher Basis erscheint die synchronisierte Erfassung der „echten“ Sicht der Probanden auf das Gerät mit einem Eye-Tracking-Helm, um auch den Fokus der Probanden und den Wechsel des Blicks zwischen App und Fragebogen etc. nachvollziehen zu können. Eine derartige Lösung stand aber leider nicht zur Verfügung und wäre ggf. aus Nutzersicht schon wieder recht invasiv. Am besten geeignet erschien daher auf Basis der gesammelten Erfahrungen der Ansatz 3. Im Abgleich zwischen gestellten Anforderungen und den näher betrachteten technischen Lösungen dazu ergaben sich einige kritische Punkte, die ergänzend zu beachten waren.

drahtgebunden vs. drahtlos: Für die Dokumentation der Testläufe im Rahmen von internen Tests oder Voruntersuchungen mit den Apps sind kabelgebundene Lösungen akzeptabel. Für die Überwachung der tatsächlichen Testläufe von Probanden wären Kabel an den mobilen Geräten aber ein ungewollter Störfaktor, der die Benutzung der Geräte unhandlicher macht und den Nutzungskontext noch weiter in Richtung einer künstlichen Laborsituation verschieben würde. Im praktischen Einsatz mit Probanden ist daher eine drahtlose Lösung vorzuziehen. Bei der drahtlosen Ausleitung (Senden vom mobilen Gerät per WLAN) darf die restliche Netznutzung nicht beeinträchtigt werden, da sonst eine Verzerrung des Zusammenhangs zwischen Netzparametrisierung und Nutzerzufriedenheit möglich wäre.

Zusatz-Apps oder Nutzung eingebauter Funktionalität: Die Geräteüberwachung sollte möglichst wenig bzw. bestenfalls gar nicht auf den mobilen Geräten auffallen, sowohl aus Sicht der Probanden als auch technischer Sicht. In der Bedienung der Geräte und Ausführung der Aufgaben sollten die Probanden ungestört bleiben. Im günstigsten Fall erfolgt die Erfassung unbemerkt im Hintergrund, ohne dass zusätzliche Apps etc. explizit geöffnet bleiben müssen, große Popups oder Hinweismeldungen ins Auge springen oder kryptische Tastenkombinationen zu drücken sind. Eine eingebaute unauffällige Funktionalität möglichst im Hintergrund ist entsprechend zu bevorzugen.

Hardware-Unterstützung: Die Leistung der mobilen Geräte darf durch die Videoüberwachung nicht beeinträchtigt werden. Daher sind Hardware-Lösungen wie eine Ausleitung des Videosignals per Kabel ohne zusätzliche aufwendige lokale Komprimierung prinzipiell im Vorteil. Bei der lokalen Verarbeitung zur Speicherung auf dem Gerät oder zur drahtlosen Ausleitung als Paketdatenstrom ist hingegen zwangsläufig eine Videokomprimierung zur Reduktion der Datenraten notwendig. Echtzeit-Software-Lösungen können sehr rechenintensiv sein und die mobilen Geräte stark beanspruchen. Die Bedienung und eigentliche Nutzung im Rahmen der Aufgaben könnten dadurch ggf. beeinträchtigt und die Zufriedenheitsbewertungen der Probanden verzerrt werden. Dies ist im regulären Testbetrieb unbedingt zu vermeiden. Eine Hardware-gestützte Echtzeit-Videokomprimierung kann hingegen die angestrebte Überwachung ohne Beeinträchtigung der restlichen Nutzung ermöglichen, sofern von den genutzten Geräten geboten.

5.10.2 Realisierung der Videoüberwachung unter Android

Team-Viewer QuickSupport¹⁶⁰ unter Android geht als Lösung über die geforderte Spiegelung der Bildschirmausgaben weit hinaus, da auch eine entfernte Steuerung der Geräte etc. möglich ist:

- rein Software-basiert, allerdings auf bestimmte Handymodelle beschränkt, da entsprechende API-Unterstützung erforderlich, u. a. aber für Samsung geeignet
- funktioniert drahtlos, im Testbed über WLAN
- Paketdatenstrom mit proprietären AV-Codecs
- als Empfänger kann die Software TeamViewer¹⁶¹ auf einem PC dienen, inklusive Möglichkeit zur Aufzeichnung

Die App kann auf den Endgeräten nach dem Etablieren der TeamViewer-Sitzung geschlossen werden, dann weist lediglich ein unscheinbares kleines Symbol in der Systemleiste auf die Verbindung hin. Die gefühlte Geschwindigkeit bei der Bedienung der mobilen Geräte wird durch die zusätzliche CPU-Last beeinträchtigt: auf dem Galaxy S III etwas weniger, auf Tab 2 10.1 und Ace 2

¹⁶⁰ Erprobung mit Version 7.0.990F QS, App ist über Play-Store zu beziehen:

<https://play.google.com/store/apps/details?id=com.teamviewer.quicksupport.market&hl=de> abgerufen am 15.12.2013,

¹⁶¹ <http://www.teamviewer.com> abgerufen am 15.12.2013

deutlicher. Auf weitere Tests zur Auswirkung bei der Nutzung der interessierenden Apps wurde daher verzichtet.

Android-Geräte, vor allem der Spitzenklasse, unterstützen den Standard **Mobile High Definition Link**¹⁶² (MHL), über den die Geräteausgaben über ein Anschlusskabel als digitales AV-Signal zu nutzen sind. Für die weitere externe Verarbeitung ist ein Adapter zur Wandlung auf HDMI notwendig. Im Testbed unterstützte das Galaxy S III den MHL-Standard in HD-Auflösung und konnte erfolgreich erprobt werden. Die Leistung des mobilen Gerätes wurde durch MHL nicht beeinträchtigt. Die Bedienung ist abgesehen vom Faktor Handlichkeit bedingt durch das Kabel nicht eingeschränkt. Diese Variante wurde unter Android zur Videodokumentation der beispielhaften Aufgabenausführungen und Auswirkungen der QoS-Parametrisierung auf die Nutzerstimuli genutzt.

Samsung AllShare Cast bietet bei einigen Samsung-Geräten eine ins System integrierte Möglichkeit, die Bildschirminhalte drahtlos auf einen externen Empfänger zu spiegeln. In der Erprobung mit dem Galaxy S III waren bei Aktivierung der Funktion keine relevanten Beeinträchtigungen in der Leistung zu bemerken. Die Kompression der AV-Daten der Bild- und Tonausgaben erfolgt offenbar unter Nutzung der Hardware-Encoder im Gerät. Die Bedienung wird ebenfalls nicht gestört. Auf die aktivierte AllShare Cast Funktion weist lediglich ein kleines Icon in der Systemleiste hin. Die AV-Paketdatenströme werden drahtlos, unter Umgehung der restlichen Testbed-Technik, per Wi-Fi Direct¹⁶³ zwischen mobilem Endgerät und einem speziellen Empfänger¹⁶⁴ ausgetauscht. Die maximal parametrisierten und genutzten Datenraten der Netzemulation liegen deutlich unterhalb der möglichen Leistung des WLAN-Testbeds, so dass von genug Kapazität für die parallele Nutzung von WLAN für AllShare Cast und den emulierten Mobilnetzzugang auszugehen ist. Die Variante mit AllShare Cast eignet sich unter Android prinzipiell auch zur Videodokumentation der Testläufe von Probanden im regulären Testbetrieb. Eine AllShare Cast ähnliche, aber herstellerübergreifende Weiterentwicklung stellt Miracast¹⁶⁵ dar, das ebenfalls auf eine Wi-Fi-Direct-Verbindung zum

¹⁶² <http://www.mhlconsortium.org> abgerufen am 15.12.2013

¹⁶³ <http://www.wi-fi.org/discover-and-learn/wi-fi-direct> abgerufen am 15.12.2013

¹⁶⁴ Samsung AllShare Cast Dongle

¹⁶⁵ <http://www.wi-fi.org/wi-fi-certified-miracast%E2%84%A2> abgerufen am 15.12.2013

Empfang mit einem externen Dongle setzt. Weitere technische Details zu Miracast-Grundlagen und zum aktuellen Stand der technischen Realisierung finden sich in [267].



Abbildung 55: Spiegelung der Bildschirmausgaben unter Android, notwendige zusätzliche Hardware und Software

Externe Weiterverarbeitung: MHL-Adapter und AllShare Cast Dongle stellen die Ton- und Bildschirmausgaben der Mobilgeräte jeweils über einen HDMI-Ausgang (mit HDCP geschützt) als digitales AV-Signal bereit. Für die weitere Verarbeitung, z. B. Anzeige auf einem PC und Aufzeichnung, sind weitere Signalwandlungen notwendig. Im Testaufbau wurde ein HDMI-VGA-Wandler genutzt, der ein recht brauchbares VGA-Signal und ein analoges Audio-Signal liefert. Eine Redigitalisierung des VGA-Signals erfolgte mittels VGA2USB-Framegrabber¹⁶⁶, der als externe Videoquelle für die Aufzeichnung auf einem PC diente, siehe Abbildung 55. Diese preiswerte Lösung ist hinsichtlich der Bildqualität und vor allem der Bildwiederholrate (max. 10 FPS) limitiert, für dokumentarische Zwecke aber ausreichend. Zur Anzeige, Aufzeichnung und erneuten Videokompression diente der VideoLAN Client¹⁶⁷ (VLC) als nachgeschaltetes universelles Software-Werkzeug im Bereich Videoverarbeitung.

5.10.3 Realisierung der Videoüberwachung unter iOS

Einen guten Überblick zu generellen Möglichkeiten unter iOS den Bildschirminhalt mit Hardware und Software sowohl kabellos als auch kabelgebunden, z. B. auf PC oder Beamer auszuleiten, gibt es in [268]. Bei den iOS-Geräten ist über den Dock Connector (ältere Geräte) bzw. Lightning Connector (ab iPhone 5 und iPad 4) ein Hardware-gestütztes, kabelgebundenes Abgreifen mittels spezieller Adapter als digitales oder analoges Video-Signal möglich. Auf eine praktische Erprobung dieser Möglichkeit wurde verzichtet, da die eingesetzten Geräte das für die beabsichtigten Zwecke besser geeignete AirPlay Mirroring unterstützen.

AirPlay Mirroring¹⁶⁸ ist direkt in iOS integriert und auf Geräten ab dem iPhone 4s verfügbar. Eine Nutzung der vorhandenen Hardware-Ressourcen zur Komprimierung der AV-Datenströme ist anzunehmen, da in der Erprobung keine relevante Leistungsbeeinträchtigung festzustellen war. Die Bedienung wird nicht gestört, auf die aktivierte Spiegelung der Bildschirmausgaben weist lediglich ein unscheinbares Symbol in der Statusleiste hin. Die erzeugten

¹⁶⁶ <http://www.epiphan.com/products/frame-grabbers/vga2usb/> abgerufen am 15.12.2013

¹⁶⁷ <http://www.videolan.org/> abgerufen am 15.12.2013

¹⁶⁸ <http://www.apple.com/airplay/> abgerufen am 15.12.2013

AV-Paketdatenströme werden vom mobilen Gerät drahtlos über das konfigurierte WLAN des Netzemulations-Testbeds zum WLAN-Router gesendet und von da drahtgebunden über die Schnittstelle eth0.2 in das lokale Subnetz an einen Empfänger weitergeleitet¹⁶⁹. Diese zusätzlichen Datenströme sollen nicht durch die Netzemulation beeinflusst werden bzw. diese gar auslasten. Dazu wurde eine Art Bypass auf den WLAN- Routern eingerichtet, der Datenpakete mit Quelle und Ziel innerhalb des konfigurierten lokalen Netzes am künstlichen Flaschenhals der Netzemulation vorbeischleust. Die möglichen WLAN-Datenraten im Testbed zwischen mobilem Endgerät und WLAN-Routern übersteigen die maximal parametrisierten und genutzten Datenraten der Netzemulation deutlich, so dass von genug Kapazität für die parallele Nutzung von AirPlay Mirroring (ca. 4 MBit/s in Senderichtung vom mobilen Gerät) und den emulierten Mobilnetzzugang auszugehen ist. Als Empfänger der Datenströme kann statt eines Apple-TV auch eine Software-Lösung auf Mac oder PC dienen. In der Erprobung wurde die Software *Reflector*¹⁷⁰ auf einem PC genutzt und konnte mit problemloser Inbetriebnahme und integrierter Aufzeichnungsmöglichkeit überzeugen. Auf Technik und weitere Software-Lösungen für AirPlay Mirroring geht [269] im Detail ein.

AirPlay Mirroring spiegelt nicht nur den Bildschirminhalt auf einen externen Empfänger, sondern leitet auch das komplette Audiosignal um, so dass das mobile Gerät bei Tonausgaben selbst stumm bleibt. Dies war im Testbetrieb kritisch, da die Probanden in den Aufgaben teilweise Audio- bzw. Videodienste mit Ton nutzen sollten. Zudem sollten die Probanden von einer „normalen“ Nutzung mit lokalem Kopfhörer ausgehen. Um dies zu suggerieren, lag ein Audiokabel zum Anstecken an ihre Geräte bereit. Aus Sicht der Probanden sollte es erscheinen, als ob dieses Kabel zum Kopfhörer um den Fuß des Monitors vor ihnen gewickelt wäre. Tatsächlich war der Kopfhörer aber unsichtbar hinter dem Monitor über eine Audioverlängerung an den Audioausgang des Empfängers des AirPlay-Mirroring-Datenstroms angeschlossen und wurde von da mit dem Audiosignal gespeist. Das ins Mobilgerät ein-

¹⁶⁹ Vergleiche Abbildung 44 und Abbildung 54 zur Verdeutlichung, womit klar wird, warum die interne Netzkonfiguration recht kompliziert eingerichtet werden musste.

¹⁷⁰ <http://www.airsquirrels.com/reflector/> abgerufen am 15.12.2013, Nutzung der Version 1.2.3 64-Bit unter Windows 7

steckte Audiokabel war somit nur eine funktionslose Attrappe. Diese technische Täuschung hat sich bewährt und blieb von den Probanden unbemerkt. Durch die Art der Audiosignalführung entsteht eine minimale Verzögerung, die allerdings so gering ausfällt, dass eine QoE-Beeinträchtigung unwahrscheinlich ist.

5.11 Zusammenfassung

Die Integration des Netzes bzw. der interessierenden Netzsituationen in die eigenen Untersuchungen bedeutete eine Auseinandersetzung mit der Mobilfunktechnik. Es wurde erläutert, wie aus dem Abgleich theoretisch möglicher Bestwerte aus den Spezifikationen und Erhebungen zu tatsächlich erreichbaren Werten die typischen Netzcharakteristiken ermittelt wurden. Für die weitere Berücksichtigung des Netzes, vor allem der Netzsituationen, waren einige Vereinfachungen zu treffen (keine Mobilität, Annahme der Limitierung im drahtlosen Zugangnetz, keine räumliche/zeitliche Dynamik), die zu beachten sind. Daher ist im Weiteren eher die Wirkung der erreichten Effektivwerte der jeweiligen QoS-Parameter interessant. Die entwickelte Matrix technischer Parameter enthält passende Wertkombinationen aus maximaler Datenrate im Downlink, maximaler Datenrate im Uplink und minimaler Grundlatenz. Dazu wurden sinnvolle Abstufungen gebildet und interessante Bereiche der relevanten Mobilfunktechnologien eingegrenzt.

Aus den verschiedenen Ansätzen zur Untersuchung von Netzsituationen wurde bedingt durch die Anforderungen an die Realitätsnähe eine Netz- bzw. Link-Emulation ausgewählt. Das Prinzip des kontrollierten Flaschenhalses wurde vorgestellt. Die theoretischen Grenzen der Netzemulation wie auch die Grenzen der gewählten Realisierung wurden diskutiert. Dabei wurde eine Tauglichkeit für die angestrebten Zwecke festgestellt. Im Rahmen der geplanten Netzparameter und ihrer Bereiche ist eine gezielte Herstellung der interessierenden Netzsituationen bzw. Netzcharakteristiken möglich.

Zur Realisierung eines eigenen Testbeds wurden handelsübliche WLAN-Hardware und *tc+netem* als bewährte Open-Source-Software zur Netzemulation eingesetzt. Verteilt arbeitende Instanzen sind über Web-API, Web-GUI und zentralisierte Komponenten des eigenen Frameworks so erweiterbar, dass

ein einfach zu bedienendes und skalierbares Testbed entstand, in dem mobile Endgeräte per WLAN genutzt werden können. Die Netzparameter können innerhalb des Testbeds für mehrere Endgeräte gleichzeitig jeweils individuell gesetzt werden, entweder manuell oder automatisiert über eine Testlaufsteuerung durch Kopplung mit dem genutzten Web-Fragebogensystem ohne weitere Eingriffe während der Testdurchführung. So war ein auf das Notwendigste reduzierter Personalaufwand zur Testdurchführung zu gewährleisten.

Neben der stationären Technik wurden auch die mobilen Endgeräte und die damit verbundenen Besonderheiten der Nutzung im Testbed vorgestellt. Die ausgewählten Endgeräte repräsentieren einen Schnitt durch den Markt mobiler Geräte mit den führenden mobilen Plattformen, verschiedenen Klassen und Formfaktoren. Auch wenn die technische Realisierung auf bewährten Prinzipien und Technologie aufsetzt, so ist das geschaffene Testbed als komplexes Gesamtsystem nicht trivial. So mussten Probleme im Zusammenspiel der Netzemulation mit den Endgeräten (iOS) festgestellt werden, konnten aber zufriedenstellend gelöst werden.

Im Testbed wurden auch technische Lösungen zur passiven Kontrolle geschaffen. Die Netznutzung der Endgeräte kann dazu auf verschiedene Arten überwacht und zur weiteren Analyse aufgezeichnet werden. Daneben besteht die Möglichkeit, zur Dokumentation die Geräteeingaben und -ausgaben mit den Nutzeraktionen sowie den erreichten Nutzerstimuli per Video parallel zum Abgleich mit der erreichten Kommunikationsleistung zur späteren Auswertung zu erfassen.

6 Die Dienste

Vorüberlegungen und Voruntersuchungen zu den Diensten in der Studie

In diesem Abschnitt wird diskutiert, was bei der Auswahl und systematischen Untersuchung von (echten) mobilen Diensten zu berücksichtigen ist. Dabei soll verdeutlicht werden, dass es zwar möglich ist, irgendwelche Dienste auszuwählen, eine systematische Zusammenstellung der Auswahl unter verschiedenen Gesichtspunkten (Relevanz) aber deutlich sinnvoller ist. Dafür werden die entwickelten Prinzipien zu notwendigen sowie sinnvollen Vorüberlegungen und Voruntersuchung in der Phase der Studienplanung vorgestellt. Diese Prinzipien können auch für weitere Dienste in ähnlicher Art und Weise angewendet werden.

Die Probleme bei einer „Fast-Alles-Echt“-Nutzung werden verdeutlicht, inklusive möglicher Herangehensweisen, um mit diesem Ansatz überhaupt zu brauchbaren Ergebnissen zu kommen. Dazu wird zunächst das Prinzip der Vorüberlegungen und Voruntersuchungen erörtert:

- Testrelevanz: generell für eine Dienstklasse und/oder für einen bestimmten Dienst
- Testeignung: ganz konkret für einen spezifischen Dienst und zugehöriger App mit allen relevanten Eigenschaften und deren Wahrnehmung, was Testbarkeit (sinnvolle Tests innerhalb des Testbeds) und Vergleichbarkeit (vergleichbare Ergebnisse innerhalb und außerhalb des Testbeds) umfasst
- Testeffizienz: sinnvolle Einschränkung des Parameterraums

Für die Einbeziehung in die Laborstudie waren Testrelevanz und Testeignung jedes einzelnen Dienstes und der zugehörigen Elemente entscheidend, für einen möglichst hohen Erkenntnisgewinn die Testrelevanz und Testeffizienz.

Das Aufbrechen der Komplexität besonders zur Testeignung war neben den theoretischen Gesichtspunkten auch in der praktischen Anwendung im Rahmen der Vorbereitung der Studie mit großem Aufwand verbunden. Die Ausführungen dazu sind deshalb ein wesentlicher Teil sowohl der theoretischen

Abhandlungen, was ein sinnvoller QoS-QoE-Test berücksichtigen sollte, als auch der praktischen Arbeit im Vorfeld der eigentlichen Durchführung der Versuche mit den Probanden.

6.1 Die Einbeziehung der Dienste in die Studie

Die Dienste bzw. die zugehörigen Elemente können auf unterschiedliche Arten in die empirischen Untersuchungen einbezogen werden, siehe Abbildung 4 und die vorangegangenen Erläuterungen zur Emulation beim Netz bzw. den Endgeräten. Prinzipiell möglich sind die Nutzung *echter* Dienste (inklusive echter Apps, echter Gegenseite und echter Inhalte) oder die entsprechende Nachbildung in Form einer Emulation. Beide Ansätze sind mit Vor- und Nachteilen verbunden. Auf die möglichen Probleme bei der Nutzung echter Dienste im empirischen Nutzertest wird in den nachfolgenden Abschnitten noch im Detail eingegangen. Daher soll die Diskussion an dieser Stelle zunächst die *unechten* emulierten Dienste aufgreifen, bevor die Festlegungen zur Einbeziehung der Elemente der Dienste in der eigenen Studie vorgestellt werden.

6.1.1 Echte vs. unechte Dienste

Die ETSI entwickelte mit der „Reference Page“ einen standardisierten (statischen) Inhalt für Web-Browsing-Tests, siehe [270, S. 1]. Dies erlaubt die Nutzung einer Testseite¹⁷¹ mit Text und Bildern auf eigenen Servern, in einer vollständig kontrollierten Umgebung, zur weitgehend exakten Reproduktion der immer gleichen sonstigen Bedingungen, zumindest bezogen auf Gegenseite, Inhalt und lokales Netz. [271] dehnt dieses Konzept auf weitere Dienste und Inhalte aus. So wurden dort eine Nachrichtenseite als Spiegel-Clone und „Fakebook“ (in Anlehnung an das ähnlich klingende soziale Netzwerk) als (nicht mobile) Nachbauten realisiert, zumindest in ihrer Grundfunktion und ungefähren Erscheinung.

Diese Ansätze unterscheiden sich damit fundamental vom gewählten eigenen Weg. Ohne Anspruch an die Echtheit wären die Tests und ggf. auch die Ergebnisse zwar normierter, aber ggf. in ihrer Aussage auch andere – bzw. müsste dann nachgewiesen werden, dass und wie sich die künstlich erzeugten

¹⁷¹ <http://portal.etsi.org/stq/WebReferencePage.asp> abgerufen am 27.02.2014, z. B. im Rahmen der Connect-Netztests genutzt, siehe [240] S. 30

Ergebnisse auch in die echte Welt überführen lassen. Dies bezieht insbesondere alle relevanten Nebenfaktoren ein, die für eine sinnvolle Messung und Aussage zur QoE berücksichtigt werden müssen. Das eigentliche Problem wird damit nur verlagert. Wird Fakebook untersucht, so erhält man auch nur Ergebnisse für Fakebook und nicht für Facebook.

Fakebook statt Facebook ist damit kein besserer oder schlechter Ansatz, sondern lediglich ein anderer. Letztlich wäre auch abzuwägen, welcher Aufwand mit den verschiedenen Wegen verbunden ist. Dies sind einerseits die detaillierten Voruntersuchungen von echten Diensten und ggf. die Berücksichtigung von festgestellten Besonderheiten aus der Echtheit heraus in der Auswertung. Andererseits sind das Nachprogrammieren von u. U. (sehr) komplexen mobilen Apps mit entsprechenden serverseitigen Gegenstücken und zusätzliche Untersuchungen zur Praxisrelevanz der geschaffenen künstlichen Lösung notwendig.

6.1.2 Festlegungen zur Einbeziehung der Elemente der Dienste

Zur App: Es sollten jeweils die echten, aktuellen Apps für die entsprechende Plattform aus den App-Stores der Plattformbetreiber passend zum eingesetzten Gerät genutzt werden. Eine Alternative wäre eine Nutzung im mitgelieferten Web-Browser der mobilen Geräte. Die Auswahl zwischen nativer App und Nutzung im Browser erfolgte dabei unter Abwägung unterschiedlicher Aspekte, wie z. B. Verbreitung zur Nutzung in der App im Vergleich zum Browser, Verfügbarkeit an adäquatem Inhalt und aus Sicht der praktischen Aspekte zur Testgestaltung.

Zu Kommunikationsszenario und Gegenseite: Konkret sind die Dienste und Nutzungsszenarien in der Testrunde 2 auf Person-to-Content beschränkt, mit dem Nutzer auf der lokalen Seite der Nutzung an einem mobilen Endgerät und einer Maschine, die Inhalte empfängt oder sendet, als Server am anderen Ende. Die Gegenseiten sind echt, d. h. als Einflussgröße genauso wirksam wie bei der Nutzung außerhalb der Testumgebung. Entsprechend geht von der Gegenseite ein gewisser zufälliger Einfluss auf die QoD und die davon abhängigen Qualitäten (QoA und QoE) aus.

In den Voruntersuchungen eines Cloud-Speicherdienstes zeigte sich beispielsweise, dass der Flaschenhals in einer Ende-zu-Ende-Betrachtung nicht immer in den mobilen Zugangsnetzen angesiedelt ist, sondern auch an anderer Stelle lokalisiert sein kann. Dabei deuteten im konkreten Fall die Veränderungen in Abhängigkeit der Tageszeit auf mangelnde Lastskalierung der Gegenseite hin, weshalb ein anderer Cloud-Anbieter für die Tests ausgewählt wurde.

Zum Inhalt: Aspekte zum Einfluss unterschiedlicher Inhaltsversionen standen nicht vorrangig im Mittelpunkt des eigenen Tests. So gibt es beispielsweise keinen direkten Vergleich der Audioqualität bei verschiedenen Inhaltsdatenraten im Musik-Streaming-Dienst MTV-Music – entsprechend wurde die Audioqualität auch nicht als Item erfasst. Trotzdem gibt es bei den einzelnen Diensten interessante Hinweise darauf, wie Quantität und Qualität des Inhalts die Wahrnehmung und die QoE-Bewertungen beeinflussen können, z. B. bei Drive die Wahrnehmung der Dateigröße (als Quantität) oder bei Youtube die Qualitätsstufe des Videos (als Qualität).

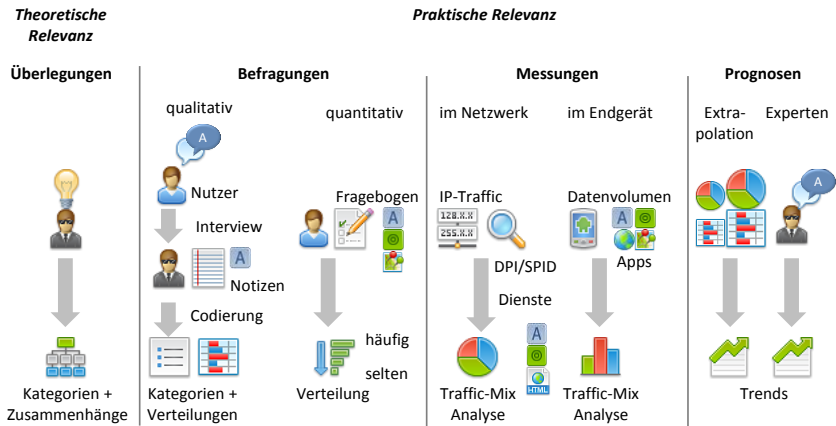


Abbildung 56: Übersicht zur Ermittlung der Relevanz mobiler Anwendungen und Dienste

DPI = Deep Packet Inspection, SPID = Statistical Protocol Identification

6.2 Testrelevanz: Die Auswahl relevanter mobiler Dienste

Generell sollte eine Auswahl von Diensten mit hoher praktischer Relevanz (Nutzer und Netz) angestrebt werden, auch wenn die Relevanz kein zwingendes Kriterium für einen Einschluss oder Ausschluss eines Dienstes im Test war.

Die Beurteilung der Relevanz mobiler Dienste erfordert unterschiedliche Sichtweisen:

- aus theoretischen Überlegungen heraus
- aus der praktischen Nutzersicht
- aus der praktischen Netzsicht (bzw. Netzbetreibersicht)

Jede einzelne dieser Sichtweisen betrachtet andere Aspekte der Nutzung mobiler Anwendungen. Als besonders relevant können Anwendungen gelten, die aus mehreren Sichten heraus interessant erscheinen. Die Relevanz mobiler Dienste in diesen Sichtweisen lässt sich auf etliche Arten ermitteln, siehe Abbildung 56. Die Welt der mobilen Dienste ist nicht statisch. Deshalb sind neben der Betrachtung der gegenwärtigen Situation auch Prognosen für die Einschätzung der zukünftigen Relevanz zu berücksichtigen. Bei den Quellen zur praktischen Relevanz (Studien mit Befragungen, Messungen und Prognosen) ist neben dem berücksichtigten Stand (Jahr) zum Zeitpunkt der Testplanung ein Verweis angegeben, unter dem die jeweils letzte Version mit aktuellen Angaben online zu finden ist (sofern verfügbar).

6.2.1 Theoretische Relevanz und Überlegungen

Aus einem geeigneten Modell lassen sich in einer theoretischen Analyse die Dienste auswählen, die jeweils einer als relevant eingestuften Kategorie oder einem besonderen QoS-QoE-Zusammenhang zuzuordnen sind. Aus theoretischen Überlegungen heraus ist eine Auswahl von Diensten dann relevant, wenn möglichst das gesamte Spektrum an verschiedenen Nutzungs-, Verkehrs- und Dienstklassen abgedeckt wird und nur eine geringe Anzahl an Nebenbedingungen zu beachten ist. So ist es anzustreben, gute Beispiele an Diensten für verschiedene Arten und Charakteristiken von QoS-QoE-Zusammenhängen auszuwählen. Dabei ist für die theoretische Relevanz wiederum nach Nutzersicht und Netzsicht zu differenzieren. In der Nutzersicht sollten

unterschiedlich Interaktionsmodelle, verschiedene Arten Nutzerstimuli und gut parametrisierbare Ausprägungen der Nutzerstimuli angestrebt werden. In der Netzsicht wäre die Abhängigkeit von den QoS-Netzparametern, die im Testbed manipuliert werden können, der Ausgangspunkt der Betrachtungen. Konkret sind damit Dienste interessant, die sensitiv für Änderungen der Datenraten und der Latenz sind.

6.2.2 Praktische Relevanz

In einem Test werden keine hypothetischen und idealen Dienste untersucht, sondern konkrete und echte Dienste mit dazugehörigen Anwendungen, Inhalten und Gegenstellen. Daher sind theoretische Überlegungen zur Auswahl und Relevanz allein nicht ausreichend. Insbesondere unter der Zielsetzung einer praktischen Relevanz der gesamten Untersuchungen gilt dieses Kriterium auch für die Auswahl der Dienste.

Zur Ermittlung dieser praktischen Relevanz sind Befragungen und Messungen geeignete Mittel. Zum Teil konnte dazu auf eigene Ergebnisse aus der Testrunde 1 zurückgegriffen werden (siehe 7.4.2 zur Gegenüberstellung der Ergebnisse zu den Diensten). Zusätzlich konnten zahlreiche weitere Erhebungsergebnisse ermittelt werden. Auch dabei werden typischerweise Klassen bzw. Kategorien genutzt. Bei der Nutzung mehrerer Datenquellen, wie sie nachfolgend aufgeführt werden, ist eine vergleichbare Klassifikation für verschiedene Sichten nicht zu erwarten bzw. festzustellen. So sind auch gemischte Ansätze zur Bildung von Kategorien durchaus üblich. Entsprechend ist eine direkte Abbildung der Kategorien aus verschiedenen Studien/Sichten aufeinander nicht möglich, bzw. auch nicht sinnvoll. Schon innerhalb einer Sicht und Methodik können unterschiedlich feingranulare Aufteilungen einen Vergleich mehrerer Erhebungen und Studienergebnisse deutlich erschweren. Grundsätzlich lieferten die Betrachtungen zu den nachfolgend aufgeführten Erhebungen die Grundlage für die Zuordnung der untersuchten Dienste in der eigenen Studie zu bestimmten Kategorien gemeinsam mit den theoretischen Überlegungen.

6.2.3 Nutzersicht

Auskunft zur Relevanz mobiler Dienste aus Nutzersicht geben typischerweise Befragungen ähnlich zur eigenen Nachbefragung (siehe A.4.12 zum Fragebogen), die mehrere Kriterien berücksichtigen:

- hohe Popularität, hohe Nutzungsfrequenz, hohe Nutzeneinschätzung und Erwartungshaltung
- Breite der Nutzerbasis: Verbreitung vs. Intensität:
 - viele Nutzer, die eine Anwendung zumindest gelegentlich nutzen
 - wenige Nutzer, die eine Anwendung aber sehr intensiv nutzen
- Nutzungsdauer: Womit verbringen die Nutzer die meiste Zeit? (siehe z. B. [14])

Zum Teil werden Dienste und Anwendungen generisch über die Funktion wahrgenommen (E-Mail), zum Teil exemplarisch über ausgewählte Anbieter (Bsp. Youtube für mobiles Video). Eine ausgewählte oder einschränkte Zielgruppe, z. B. nur geschäftliche Nutzer, kann andere Dienste als relevant identifizieren als eine allgemeine oder rein private Nutzerschaft.

Neben den Ergebnissen der Nachbefragung aus der Testrunde 1 dienen die folgenden regelmäßig aktualisierten Studien zur Orientierung für die Auswahl der Dienste der Testrunde 2 aus Nutzersicht:

- Mobilfunkmonitor¹⁷² (2012 [272])
- ComScore MobiLens / Mobile Metrix 2.0 (Frühjahr 2013 [273])
- Mobile Facts¹⁷³ der Arbeitsgemeinschaft Online Forschung (AGOF) (Anfang 2013 [274])

Zusätzlich sind über entsprechende Datenbanken (z. B. Statista¹⁷⁴) zahlreiche weitere Befragungsergebnisse unterschiedlichster Herkunft, Umfangs und Aktualität zu ermitteln.

¹⁷² <http://eplus-gruppe.de/tag/mobilfunkmonitor/> abgerufen am 10.08.2013

¹⁷³ <http://www.agof.de/mobile-facts/> abgerufen am 20.10.2014

¹⁷⁴ <http://de.statista.com/> abgerufen am 22.10.2014, teilweise kostenpflichtige Mitgliedschaft erforderlich

6.2.4 Netzsicht und Netzbetreibersicht

Aus technischer Netzsicht sind vorrangig die Anwendungen interessant, die datenratenintensiv und/oder datenmengenintensiv sind und damit die Netze stark belasten. Aus wirtschaftlicher Sicht der Netzbetreiber sind vor allem die Dienste relevant, bei denen Nutzer als Kunden besonders kritisch sind, wenn es einerseits um Anforderungen bzw. andererseits um Einschränkungen geht.

Die Messung des Anteils des Datenverkehrs, der durch eine Anwendung verursacht wird, kann auch etwas über die Popularität aussagen, ohne explizit die Nutzer befragen zu müssen, wenn entsprechende Modelle zum durchschnittlichen Datenverbrauch etc. vorliegen. Die Messungen können entweder im Netz erfolgen, unterstützt von Deep Packet Inspection (DPI, z. B. wie bei [275]) oder per Statistical Protocol Identification (SPID, siehe [276], [277]), bzw. direkt auf den mobilen Endgeräten der Nutzer (wie bei [278]).

Die Aussagen sind besonders interessant, wenn:

- sie aktuell sind und sich mit Blick auf die Vergangenheit ein Trend ableiten lässt
- sie sich auf mobile Nutzung fokussieren
- nicht nur globale Einschätzungen abgeben, sondern auch zu ausgewählten geografischen Regionen (Europa) bzw. Ländern (Deutschland) aufgeschlüsselt Einblicke bereitstellen
- Mess- und Erhebungsmethodik plausibel sind
- sie frei zugänglich sind
- sie eine offene Datenbasis aufweisen

Zur Situation aus Netzsicht lohnend sind, weil regelmäßig aktualisiert:

- Sandvine Global Internet Phenomena¹⁷⁵ (2H/2012 [275])
- Ericsson Mobility Report¹⁷⁶ (1H/2013 [279], halbjährliche Berichte und quartalsweise Zwischenberichte)
- Allot Mobile Trends Report¹⁷⁷ (Feb. 2013 [280], mit wechselnden Schwerpunkten)

¹⁷⁵ http://www.sandvine.com/news/global_broadband_trends.asp abgerufen am 24.03.2013

¹⁷⁶ <http://www.ericsson.com/mobility-report> abgerufen am 20.10.2014

¹⁷⁷ <http://www.allot.com/> abgerufen am 24.03.2013

Diskrepanzen zwischen den Ergebnissen einer Befragung und dem objektiven Blick ins Netz können bedeuten, dass die Nutzer nicht immer ganz korrekt bzw. ehrlich Auskunft erteilen (z. B. bei VoIP), bzw. selbst scheinbar identische Kategorien doch nicht ganz deckungsgleich sind.

6.2.5 Prognosen

Prognosen extrapolieren die Entwicklung zum Verkehrsvolumen und zum Traffic Mix zumeist für vier oder fünf Jahre. Dabei zeigt sich, dass vereinfacht von einem exponentiellen Wachstum für das Verkehrsvolumen auszugehen ist, das vor allem von datenintensiven Anwendungen, allen voran Mobile Video, getrieben wird.

Als häufig genannte Quellen für technische Prognosen dienen:

- Cisco Visual Network Index (VNI) Forecast¹⁷⁸ (Verkehrsvolumen, 2012 [8])
- Cisco VNI Service Adoption Forecast¹⁷⁹ (Nutzung von Services, 2012 [281])
- Ericsson Mobility Report (siehe 6.2.4)

Neben den technischen Prognosen zur Netznutzung sind auch Expertenbefragungen ein probates Mittel für den Blick in die nähere Zukunft. Dabei ist zu erwähnen:

- Chetan Sharma Consulting: Mobile Industry Predictions Survey¹⁸⁰ (Prognose 2013, [282])

¹⁷⁸ <http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html> abgerufen am 23.04.2014

¹⁷⁹ http://www.cisco.com/c/en/us/solutions/collateral/service-provider/vni-service-adoption-forecast/Cisco_VNI_SA_Forecast_WP.html abgerufen am 23.04.2014

¹⁸⁰ <http://www.chetansharma.com/> abgerufen am 27.03.2013

6.3 Testeignung: Das Prinzip von Testbarkeit und Vergleichbarkeit

Die Testbarkeit beschreibt den Grad, zu dem die interessierenden Zusammenhänge gezielt getestet (kontrolliert und beobachtet) werden können. Generell sollte eine hohe Testbarkeit angestrebt werden. Ein einzelnes Testergebnis macht noch keine Studie - für aussagekräftige Ergebnisse sind viele jeweils gleichartige Testläufe erforderlich. Die Vergleichbarkeit beschreibt, inwieweit die innerhalb des Testbeds gewonnenen Einzelergebnisse tatsächlich vergleichbar sind und inwieweit diese generell mit denen einer realen Nutzung außerhalb des Testbeds übereinstimmen. Generell sollte eine hohe Vergleichbarkeit angestrebt werden.

Die Definition der Testeignung orientiert sich damit nachfolgend an dem Begriff im Sinne der Informatik wie bei Tests von Software-Artefakten (siehe z. B. in [215] und [25]). Die daran angelehnten Begriffe der „*keit“ und „*heit“ (siehe Abbildung 57a) umreißen die wesentlichen Facetten recht gut, auch wenn sie nachfolgend nicht im eigentlichen Sinne reiner Software-Tests genutzt werden sollen. Die Begriffe wurden entsprechend aufgegriffen, (um-)interpretiert, erweitert und in das eigene Konzept zur Voruntersuchung, Testplanung, Durchführung und Auswertung integriert. Die Erläuterungen zu den gewählten Begriffen werden nachfolgend noch geliefert.

Die einzelnen Facetten greifen ineinander bzw. überlappen sich sogar, so dass durch ihre Gesamtheit jeweils Testbarkeit und Vergleichbarkeit definiert werden. Dabei können sich die Facetten auf objektive Eigenschaften als auch ihre Wahrnehmung beziehen. Wenn aufgrund dieser Verschränkungen der QoS-QoE-Zusammenhang nicht mehr sinnvoll in einem Test zu untersuchen ist, dann ist keine Testeignung gegeben. Oder anders formuliert: Für eine Testeignung müssen trotz dieser Verschränkungen die QoE-Änderungen charakteristisch auf die QoS-Änderungen zurückzuführen sein.

Zur Feststellung der Testeignung bedarf es entweder des Blicks in eine magische Kristallkugel – oder des vorgeschlagenen 6-eckigen *Testeignungskristalls*, der die Verschränkungen der relevanten Facetten für Testbarkeit und Vergleichbarkeit berücksichtigt und nachvollziehbar macht, siehe Abbildung 57b.

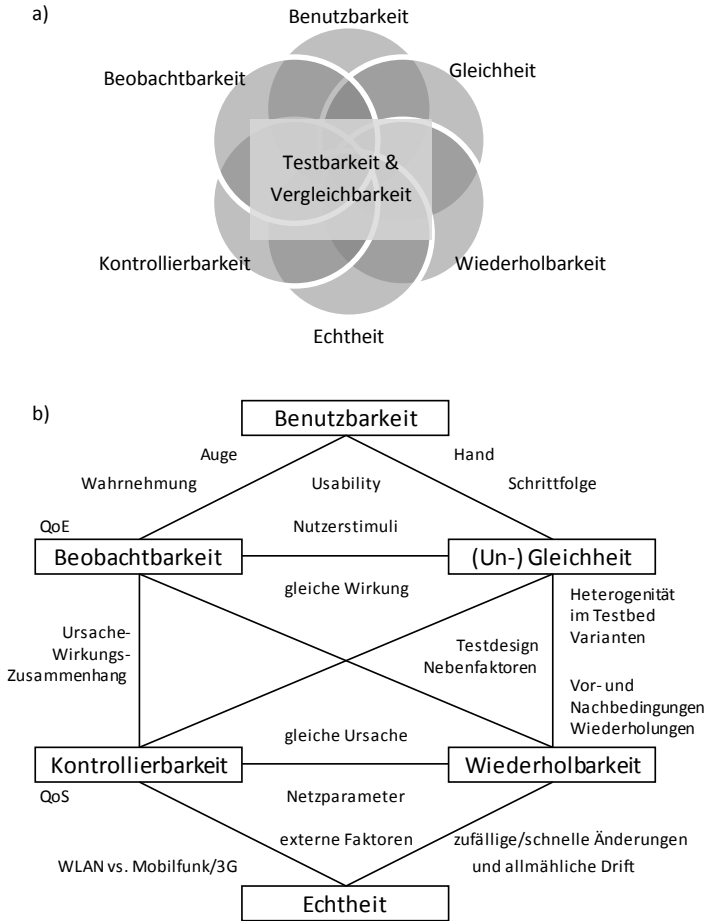


Abbildung 57: Von den Facetten der Bewertung der Testbarkeit und Vergleichbarkeit zur Feststellung der Testeignung

- a) die identifizierten Facetten von Testbarkeit und Vergleichbarkeit: unscharf und ineinander verschränkt
- b) der entwickelte Testeignungskristall mit internen Zusammenhängen

Der Weg des Informatikers sich den Problemstellungen der Tests zu nähern unterscheidet sich daher ggf. von anderen Ansätzen, z. B. in der Psychologie und allgemeinen Methodenlehre zur Versuchsplanung bzw. Systemtheorie (siehe Black-Box-Modell). Dort sind die Mittel der Wahl die Ausschaltung oder Konstanthaltung möglicherweise störender Einflüsse. Hier ist es die Kenntnis und Berücksichtigung der internen Zusammenhänge um unverstandene und womöglich unberücksichtigte Störfaktoren von vornherein in bekannte und berücksichtigte Nebenfaktoren zu überführen.

Die Vorüberlegungen zur Testbarkeit und Vergleichbarkeit sind auch mit Blick auf die Testgüte und die dabei wichtigen Kriterien *Validität* und *Reliabilität* von Bedeutung, da sie überhaupt erst für eine nachvollziehbare Basis des technischen Teils der psychophysikalischen Experimente sorgen.

Teilweise müssen Abstriche bei Testbarkeit und Vergleichbarkeit akzeptiert und in Form von Nebenbedingungen und Ergebnispartitionierungen entsprechend berücksichtigt werden. Ein nicht ganz perfekter Test ist dabei als kleineres Übel im Vergleich zu gar keinem Test zu sehen, solange die getroffenen Einschränkungen bewusst sind.

Scherzhaft könnte man bei der Bewertung von „The Good, the Bad and the Ugly“¹⁸¹ sprechen:

- Good: gute Testbarkeit und hohe Vergleichbarkeit
- Ugly: „unschöne“ Nebenbedingungen, die Testbarkeit und Vergleichbarkeit beeinflussen könnten und daher die Berücksichtigung einzelner Besonderheiten erfordern
- Bad: schlecht, d. h. Testbarkeit und Vergleichbarkeit sind eingeschränkt, es ist selbst bei sonst gleichartigen Tests von unterschiedlichen Ergebnismengen auszugehen

Auch wenn teilweise „The Bad and The Ugly“ zum Vorschein treten, ist es im Rahmen einer aufwendigen Studie gewiss besser, dies vorab zu erkennen und nicht erst in der Auswertung aufzudecken.

¹⁸¹ nach dem gleichnamigen Film, siehe <http://www.imdb.com/title/tt0060196/> abgerufen am 05.12.2013

Das Vorgehen bei der Bewertung der Testbarkeit und Vergleichbarkeit lässt sich in drei Schritte zerlegen (angelehnt an [25] S. 300):

1. Identifizieren: Bestimmung der relevanten Aspekte
2. Analysieren: Bestimmung der Zusammenhänge und wechselseitigen Abhängigkeiten
3. Priorisieren: Abwägen der Facetten gegeneinander

Insbesondere beim Priorisieren ist zwischen den theoretischen Idealen eines Tests, der praktischen Umsetzbarkeit und den angestrebten Testzielen zu vermitteln. In einigen Fällen stehen sich die Anforderungen der einzelnen Facetten entgegen. Diese Widersprüchlichkeiten sind genau abzuwägen und als bestmöglicher Kompromiss zu lösen.

Die Verzahnungen der einzelnen Facetten lassen Anfang und Ende der Voruntersuchungen verschwimmen. Nach einer wenigsten einmaligen Betrachtung jeder Facette sollten zum Ende der Untersuchungen die Propositionen zu den QoS-QoE-Zusammenhängen mit den identifizierten Nutzerstimuli stehen. Da die Voruntersuchungen normalerweise innerhalb eines zeitlich begrenzten Rahmens erfolgen, ist möglicherweise keine vollständige Durchdringung in allen Details zu erreichen, große Überraschungen sollten aber ausgeschlossen sein. Da im Ausgangszustand eine Black Box vorliegt, ist eine vollständige Überdeckung aller Aspekte im Rahmen der Vorerprobungen trotz gründlicher Untersuchungen nicht anzunehmen. Entsprechend ist, bildlich gesprochen, auf dem Weg von der Black Box zur White Box eine ziemlich helle Grey Box schon ein recht gutes Ergebnis.

6.3.1 Echtheit

Der Anspruch der Realitätsnähe ist treibende Kraft für die Tests insgesamt – in vielen Aspekten gleichzeitig aber auch Antagonist zu den restlichen Facetten der Testbarkeit und Vergleichbarkeit. Eine hohe Echtheit und punktgenaues Festzurren der anderen Facetten schließen sich fast aus. Echtheit bedeutet Vielfalt durch echte Dienste, Apps, Gegenseiten, Inhalte und externe Netzsegmente und -elemente, ggf. aber auch echte Probleme. Durch die möglichen externen Störfaktoren bedeutet Echtheit damit auch ggf. Abstriche bei Wiederholbarkeit und Kontrollierbarkeit sowie auch indirekt bei Gleichheit, Beobachtbarkeit und Benutzbarkeit durch (schl-) echte Usability.

Von besonderer Bedeutung ist, ob sich die Apps der Dienste im WLAN-Testbed wie im echten Mobilfunknetz verhalten. Die mobilen Apps können prinzipiell den aktuellen Verbindungstyp (Mobilfunk/WLAN) bzw. die Netztechnologie (3G/LTE) erkennen und nutzen dies teilweise auch (siehe z. B. Youtube-App). Idealerweise sollten sich die Apps unabhängig vom Verbindungstyp vergleichbar verhalten. Dann kann davon ausgegangen werden, dass die Netzparametrisierung im Testbed sich so echt wie bei gleichen Effektivwerten im Mobilfunk auswirkt.

Vorgehen zur Einschätzung: Die ausgewählten Dienste wurden dazu im Rahmen der Voruntersuchungen intensiv im WLAN-Testbed und mittels SIM-Karten verschiedener Netzbetreiber in Mobilfunknetzen (3G) analysiert und verglichen. Dabei wurden sowohl Nutzersicht als auch Netznutzung berücksichtigt. Die Erläuterungen zur technischen Umsetzung der dazu eingesetzten Geräte- und Verkehrsüberwachung sind im Kapitel zum Testbed zu finden (5.9 und 5.10).

6.3.2 Kontrollierbarkeit

Die Kontrollierbarkeit bezieht sich vor allem auf die objektiv messbaren technischen Eigenschaften, die sich als technische Sicht oder *Blick von innen* beschreiben lassen. Teil dieses Blicks von innen ist die Netznutzung und auf dieser Basis eine Ursache-Wirkungs-Kette ausgehend von den grundlegenden QoS-Netzparametern.

Die Kontrollierbarkeit soll zwei im Ansatz grundverschiedene Arten von Kontrolle beschreiben:

- die **aktive** Kontrolle im Sinne einer geplanten **Steuerung**: planmäßig gezielt (*ex ante*), von der kontrollierten Ursache zur erwarteten Wirkung
- die **passive** Kontrolle im Sinne einer (permanenten) **Überwachung**: nachträglich (*ex post*), ggf. auch von der festgestellten Wirkung rückschließend auf die Ursache

Prinzipiell wären beide Arten geeignet, die interessierenden Beziehungen zwischen QoS und QoE zu beleuchten. Für Feldstudien außerhalb einer aktiv kontrollierbaren Umgebung bleibt nur die Überwachung, z. B. durch eine geeignete Art der Erfassung, Aufzeichnung und Auswertung von Netzverkehrsdaten.

Für die Laborstudie muss die Kontrollierbarkeit primär als aktive Kontrolle zu charakterisieren sein. Die Variation der objektiv erfassbaren Auswirkungen sollte daher geplant aus der Variation der interessierenden Netzparameter heraus erfolgen. Dies ist eine wichtige Voraussetzung, da die Laborstudie zur Erzielung der empirischen Ergebnisse als (Quasi-) Experiment angelegt ist und die Auswertung (weitgehend) parameterbasiert möglich sein soll. Die kontrollierte (planmäßige) QoS-Netzparametrisierung sollte daher der (größte) Einflussfaktor sein, der die Veränderungen der Ergebnisse erklärt. Dafür muss ein kontrollierter Ursache-Wirkungs-Zusammenhang zwischen QoS-Parametrisierung und zu erwartender QoE-Bewertung vorhanden sein.

Daneben sollten aus der technischen Sicht Prognosen zu Kovariablen, die nur bedingt zu steuern, sondern ggf. nur zu beobachten sind und ihre Auswirkungen auf Gleichheit und Wiederholbarkeit formuliert werden, sofern möglich. Die Art der Aufklärung der inneren Zusammenhänge dazu bezieht deshalb neben der aktiven Kontrolle (Steuerung) zusätzlich Komponenten zur passiven Kontrolle (Überwachung) ein, die helfen, die jeweiligen Modelle zu konkretisieren.

Vorgehen zur Einschätzung: In internen Testreihen wurden Stufen aus der Matrix technischer Parameter ausgewählt und damit die konzipierten Testszenarien ausgeführt. Dabei wurden die Verkehrsdaten und – soweit möglich – weitere interne Größen erfasst und analysiert, ob sich gezielte Veränderungen zeigten. Daran anschließend erfolgte der Abgleich mit den erzielten Beobachtungen der Nutzerstimuli, woraus sich die weiteren Erwartungswerte ergaben.

6.3.3 Beobachtbarkeit

Die Beobachtbarkeit bezieht sich vorrangig auf die Nutzersicht *von außen* und damit verbunden auch auf die erwarteten Stimuli und Bewertungen. Während sich die Kontrollierbarkeit der Ursache widmet, werden durch die Beobachtbarkeit die beobachtbaren Wirkungen beschrieben. Ein Teil der Beobachtbarkeit ist die Messbarkeit. Kontrollierte Veränderungen müssen messbare, objektive Veränderungen bewirken. Diese sind Voraussetzung für Veränderungen in der Beobachtbarkeit ausgehend von den kontrollierten Veränderungen.

Damit die kontrollierten QoS-Änderungen zu QoE-Änderungen führen können, muss der Nutzer die dienstspezifischen Netzqualitätsunterschiede in Form von konkreten Ausprägungen der Nutzerstimuli wahrnehmen können. Eventuell sind kleine Änderungen schon technisch messbar, aber noch nicht beobachtbar. Für die subjektive Wahrnehmung der Änderungen müssen die QoS-Änderungen daher groß genug sein. QoS-Änderungen und Änderungen der objektiven Eigenschaften können prinzipiell verdeckt sein oder durch kompensatorische Effekte maskiert werden, so dass zwar veränderte objektive Eigenschaften im Inneren des Modells anzunehmen sind, in der für den Nutzer relevanten Sicht von außen sich aber nichts ändert.

Die Nutzerstimuli sind nach dem entwickelten eigenen QoS-QoE-Modell zu identifizieren, zu charakterisieren und mit den Erkenntnissen der kontrollierten QoS-Variation abzugleichen. Für die Nutzerstimuli sollte geklärt sein, ob diese eher diffus und indirekt wahrgenommen (gefühl) werden, oder sich diese klar und direkt ablesen lassen, z. B. als Zahlenwert. Für die Wahrnehmung von Zeiten kann dies beispielsweise bedeuten: Sind Beginn, Fortschritt sowie Ende und damit Dauer der Datenübertragungen klar erkennbar, z. B. durch einen Timer etc., oder schätzt der Proband mit seiner inneren Uhr? Zusammen mit den entsprechenden Überlegungen, auch für die anderen Arten von Stimuli, ergibt sich eine Einschätzung der Complexity of Experience. Ausgehend von den Einschätzungen sind weitergehende Annahmen möglich, ob und in welchem Umfang die verursachten Veränderungen auch die Zielvariablen (Bewertungen) ändern können. Es ist davon auszugehen, dass für eine Veränderung einer Bewertung entsprechend große Unterschiede in den Nutzerstimuli als beobachtbare Größen des QoS-QoE-Zusammenhangs erreicht

werden müssen. Je nach Art der Stimuli, ihrer Präsentation und der Art des Dienstes insgesamt sollte ergänzend eine Abschätzung zur Variabilität der Bewertungen getroffen werden.

Vorgehen zur Einschätzung: In der Vorerprobung zur Nutzersicht von außen werden Nutzerstimuli und kontrollierte Änderungen erfasst und Bewertungen unter Berücksichtigung sinnvoller Parametergrenzen auf Basis o. g. Einschätzungen zugeordnet. Dabei sollten mehrere Experten (d. h. am Test bzw. der Testplanung beteiligte Personen) in internen Tests selbst wie Probanden agieren.

6.3.4 Benutzbarkeit

Die Beobachtbarkeit der Nutzerstimuli und die Benutzbarkeit als Usability nehmen in der Nutzersicht auf die Dienste eine entscheidende Rolle als QoE-Einflussfaktoren ein. Entsprechend ist davon auszugehen, dass die Benutzbarkeit vor allem auch mit Blick auf die (Un-) Gleichheit im Testbed ein wichtiger Unterscheidungsfaktor sein kann. Benutzbarkeit und Beobachtbarkeit sollten sich dabei nicht entgegenstehen.

Beim Blick auf die Benutzbarkeit sind mögliche und interessante Fragen:

- Welche Vorkenntnisse werden benötigt? Im Idealfall ist eine Aufgabe auch ohne besondere Vorkenntnisse lösbar, d. h. auch von Probanden zu bewältigen, die den entsprechenden Dienst zuvor noch nie genutzt haben.
- Wie sieht die Schrittfolge zur Ausgabenausführung aus? Viele Schritte bedeuten viele Möglichkeiten für Abweichungen in der geplanten Ausgabenausführung. Ein langer und im Sinne der angestrebten Erkenntnisse nutzloser Weg durch verschachtelte Menüs sollte ebenso vermieden werden wie ungewollte Abkürzungen zur Aufgabebearbeitung.
- Wie wirken sich potenzielle Fehler der Probanden auf die Ausführung und Bewertung aus? Die Probanden selbst sind durch die Art der Ausgabenausführung ein wesentlicher Einflussfaktor. Entsprechend sollte eine gute Benutzbarkeit möglichst wenige Gelegenheiten bieten, von einer korrekten Ausgabenausführung abzuweichen.
- Sind Nutzungsszenario und Aufgabe sinnvoll beschreibbar? Die Verständlichkeit zur Ausführung einer Aufgabe sollte gegeben sein. Der not-

wendige Text zur Beschreibung der Aufgabe sollte nicht in einem Missverhältnis zur eigentlichen Aufgabenaufführung stehen (z. B. 3 min. Lesen der Aufgabe und nur 10 s zum Ausführen).

- In engem Zusammenhang zur Echtheit: Wie lange wird ein Proband zur Ausführung der Aufgabe benötigen? Die Zeit zur kompletten Bearbeitung inklusive Lesen, Ausführen und Bewerten sollte 5 Minuten im Mittel nicht übersteigen (als interne Vorgabe).

Traurig aber wahr: Hohe Echtheit bedeutet auch echte Apps und damit leider manchmal aber auch schlechte Usability, wie einige der ausgewählten Dienste anschaulich zeigten. Damit war von vornherein davon auszugehen, dass die Benutzbarkeit durch die Nutzung der echten Apps teilweise beeinträchtigt war, was zusammen mit der spezifischen Beobachtbarkeit auch zu unterschiedlichen, aber eben echten, QoS-QoE-Zusammenhängen führen kann. Zudem lassen sich für Apps mit schlechter Usability auch nur einschränkt gute und sinnvolle Nutzungsszenarien konstruieren. Leichte Einschränkungen sind daher zumindest unschön (ugly), erhebliche und grundsätzliche Zweifel sind sogar ein Hinweis auf einen nur schlecht geeigneten Dienst (bad).

Vorgehen zur Einschätzung: Zur Bewertung der Usability können Experten-Evaluationen mit etablierten Heuristiken oder empirische Evaluationen (selbst wieder ein Nutzertest) eingesetzt werden, siehe [30] S. 65ff. In Selbsttests und internen Vorabtests wurden die Ausführungsschritte erprobt, die notwendige Zeit erfasst und die Einzelschritte unter grundsätzlichen Usability-Aspekten (Heuristik) abgeschätzt. Zweifel an einer einfachen und auch für ungeübte Probanden intuitiv sofort korrekten Bedienung bedeuteten, dass die Beschreibung des Szenarios und die zugehörige Aufgabenstellung besonders gut ausgearbeitet sein mussten. Der Blick über die Schulter eines Kollegen beim Ausführen der geplanten Aufgaben war hilfreich, eigentlich offensichtlich kritische Punkte zu erkennen, die zuvor in der eigenen Betriebsblindheit übersehen wurden. Letzte Bedenken und Probleme im Bereich der Benutzbarkeit wurden im Rahmen der Pretests mit Probanden erkannt und soweit möglich abgestellt. Diese Arten der Einschätzung der Benutzbarkeit sollten aber nicht mit einem richtigen Usability Testing verwechselt werden, siehe o. g. Quelle.

6.3.5 Wiederholbarkeit

Bei der Einschätzung der Wiederholbarkeit soll geklärt werden, wie der interne Einfluss des Testdesigns und der Einfluss externer Faktoren (auch über die Zeit) wirken. Die zu verschiedenen Zeitpunkten gewonnenen Einzelergebnisse sollten reproduzierbar und zu aggregieren sein. Dafür sollten sich die externen Ko- bzw. Störvariablen, die sich der aktiven Kontrolle entziehen und die Auswirkungen der kontrollierten Variation möglichst nicht ändern. Gleiche Werte der Input-Variablen sollten zu gleichen Werten der Output-Variablen führen. Dies ist für ganz unterschiedliche **Zeiträume** zu prüfen.

Innerhalb eines Testslots (im Bereich von Minuten), mehrere Ausführungen einer Aufgabe durch einen Probanden als Wiederholung, ausgehend von möglichen internen und externen Änderungen:

- Interdependenzen durch vorgepufferte Inhalte in lokalen Caches der Apps bzw. Nutzung verschiedener Varianten (primär leicht unterschiedliche Inhalte) zur Umgehung von Caches, ggf. offene Verbindungen etc.
- schnelle bzw. zufällige Änderungen externer Inhalte
- kurzzeitige Lastfluktuationen der Gegenseite und externer Netzsegmente

Zwischen den Testslots (im Bereich von Stunden bis Tagen), ausgehend von möglichen externen Änderungen:

- wochentags- oder tageszeitabhängige und zufällige Änderungen externer Inhalte
- (circadiane) Lastschwankungen der Server der Gegenseite und externer Netzsegmente

Über den gesamten Testzeitraum hinweg (im Bereich von Wochen), ausgehend von möglichen externen Änderungen:

- starke Änderungen externer Inhalte
- API-Änderungen beim Dienstanbieter, z. B. Versionsänderungen mit unausweichlichen Updates der Apps zur Nutzung der neuen APIs
- tendenzielle Änderungen der Last der Server der Gegenseite und externer Netzsegmente

Vorgehen zur Einschätzung: Es ist lediglich eine genauere Prüfung der systematischen Einflüsse auf die Wiederholbarkeit möglich. Dafür wurden jeweils mehrere Aufgabenbearbeitungen verglichen und entsprechend auf Änderungen der Vorbedingungen geschlussfolgert. Die Zahl und Größe zufälliger Einflüsse oder zukünftiger Änderungen können bestenfalls auf Basis von Erfahrungen aus der Vergangenheit abgeschätzt werden. Entsprechend sollten starke externe Einflussgrößen im Hinblick auf die Wiederholbarkeit generell vermieden werden. Bei der Wahl der Inhalte und der Aufgabenbeschreibung etc. wurde dies berücksichtigt. Die restlichen Einflussgrößen sollten entsprechend regelmäßig während des Testbetriebs überprüft werden. Durch die technischen Aufzeichnungen innerhalb des Testbeds können die wiederholungsrelevanten Änderungen zudem bei Bedarf in der Auswertung nachvollzogen werden.

Interne Änderungen: Zu Beginn jedes Testslots sollten sich die Geräte in einem definierten Ausgangszustand befinden, also soweit zurückgesetzt, dass die Nutzung durch vorherige Probanden keine ungewollten Spuren hinterlassen hat. Dies bedeutet u. a. die Löschung von Caches, das Laden von Grundeinstellungen für alle genutzten Apps und die Bereitstellung eigener Inhalte, z. B. der Bilder für die Aufgabenausführungen. Für die erste Bearbeitung jeder Aufgabe liegt damit ein frisches Gesamtsystem mit definiertem Startzustand und Vorbedingungen¹⁸² vor. Die Vorbedingungen werden durch die Bearbeitung der Aufgabe in Nachbedingungen überführt. Diese Nachbedingungen (z. B. Datei in einem Cache) können als Vorbedingungen für eine weitere Bearbeitung wirken, siehe Abbildung 58.

In der Wiederholung einer Aufgabe ist das Gesamtsystem damit möglicherweise in deutlich anderen Vorbedingungen, da zwischen den Aufgabenbearbeitungen der Probanden keine externen Eingriffe (durch die Testleitung) zum Löschen von Caches etc. sinnvoll zu realisieren sind. Die Wiederholung der Aufgabenausführung war deshalb grundsätzlich kritisch zu prüfen und ggf. aktiv in der Testplanung und Auswertung zu beachten:

- Testplanung: Nutzung von zwei Aufgabenvarianten, mit z. B. zwei verschiedenen Inhalten

¹⁸² Bedingung soll nachfolgend als angetroffener Zustand und nicht als zwingende Voraussetzung verstanden werden.

- **Auswertung:** Unterscheidung der erzielten Ergebnisse unter verschiedenen Vorbedingungen (Partitionierung), wenn sich keine gleichen Vorbedingungen sinnvoll herstellen lassen oder zur gezielten Untersuchung des Wiederholungseffekts (z. B. bei Spiegel.de)

Das Prinzip des Startzustands mit Vorbedingungen bzw. Endzustands mit Nachbedingungen ist in übertragener Art und Weise wieder dem Prinzip der Software-Tests entlehnt, siehe [128] S. 187.

Die Veränderung der Vorbedingungen ist auch bei Wiederholungen innerhalb einer Aufgabenausführung, z. B. nach einer Fehlbedingung eines Probanden, kritisch zu prüfen. Entsprechend sollte für jedes Szenario abgeschätzt werden, ob andere Angaben als „ja, Aufgabe beim ersten Versuch gelöst“ zum Bearbeitungserfolg möglicherweise problematisch sind. Die Probanden wurden in den Instruktionen darauf hingewiesen, die Apps immer vollständig zu beenden, so dass die Beschreibung der Schrittfolge der Aufgabenstellung in erster und zweiter Bearbeitung von einem einheitlichen Startpunkt ausgehen konnte.

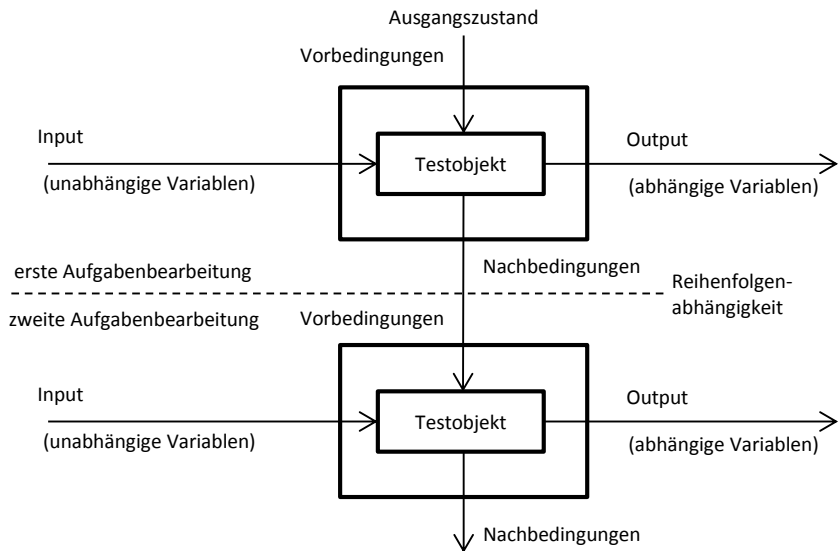


Abbildung 58: Reihenfolgenabhängigkeit durch Veränderung der Vorbedingungen

Die Nachbedingungen der ersten Aufgabenbearbeitung bilden die Vorbedingungen der zweiten Aufgabenbearbeitung. (in Anlehnung an [128] S. 187)

Externe Änderungen: Es wurden echte Dienste mit echten Apps, echten Inhalten und echten Gegenseiten genutzt. Diese können sich in ihren Eigenschaften ändern. Durch eine geschickte Auswahl sollten diese Effekte minimiert werden, bzw. darauf reagiert werden. Solange sich die Auswirkungen in den Nutzerstimuli in Grenzen halten, ist dies als echter externer Störfaktor aber noch zu tolerieren.

Aus den App-Stores kann normalerweise nur die jeweils aktuelle Version einer App bezogen werden. Ist der Testzeitraum lang genug, sind Versionswechsel wahrscheinlich. Die Apps können sich mit neuen Versionen in ihren Eigenschaften prinzipiell ändern. Im Rahmen der Test soll eine genutzte App-Version aber möglichst lange konstant genutzt werden (Prinzip Redaktionschluss). Das Rücksetzen der Geräte auf den definierten Ausgangszustand sollte daher ohne komplette Deinstallation und Neuinstallation der Apps auskommen. Sicherheitshalber wurden Möglichkeiten zum Backup der genutzten App-Versionen auf den Endgeräten vorgesehen, um Fehler des Testpersonals beim Zurücksetzen der Endgeräte oder auch Geräteausfälle notfalls kompensieren zu können (nicht notwendig geworden).

Eine weitere Erkenntnis ergibt sich aus den Voruntersuchungen und Tests über einen Zeitraum von einigen Wochen. In dieser Zeit steht die externe Entwicklung nicht still, es sind zumindest kleine Änderungen feststellbar, z. B. beim Inhalt und bei der Einführung neuer API-Versionen, die nach einer Übergangsfrist verbindlich werden etc. Eine einmal erprobte Testzusammenstellung mit externen (echten) Komponenten funktioniert also nicht ewig, zumindest nicht völlig unverändert. Der eigentliche Zeitraum zur Durchführung der Tests sollte daher nicht überdehnt werden, sonst ist durch die externen, nicht zu beeinflussenden Änderungen zu viel Drift in Form sukzessiver Änderungen zu befürchten. Im Extremfall führt dies bis zum vollständigen Aussperren der dann veralteten App-Versionen durch die Anbieter. Gleichzeitig bedeutet dies aber auch, dass vergleichende Längsschnittstudien über die Zeit möglicherweise grundsätzlich in ihren Aussagen eingeschränkt sind.

6.3.6 Gleichheit

Die Heterogenität der Technik des Testbeds (Endgeräte) und auch das Testdesign (Varianten) stehen der Gleichheit der anderen Facetten der Testeignung entgegen. Die Quellen der Heterogenität stellen in Sinne des Testmodells weitere Einflussfaktoren dar, die neben den Netzparametern zu berücksichtigen sind. Es ist dabei unbedingt sinnvoll, zumindest den Grad der zu erwartenden Heterogenität bzw. Homogenität und damit zusätzlicher Variabilität zu kennen bzw. sogar abschätzen zu können. Letztlich führen die Überlegungen zur (Un-) Gleichheit der Nebenfaktoren zur Kernfrage für die Untersuchungen: Ist der QoS-QoE-Zusammenhang vergleichbar? Zusammen mit den Überlegungen zur Wiederholbarkeit lieferte die Analyse der Gleichheit die entsprechenden Anhaltspunkte, die auf eine sinnvolle Unterscheidung der Ergebnismenge über Dienst und Netzparametrisierung hinaus hinwiesen, z. B. Android vs. iOS, App vs. Browser und verschiedene Inhaltsversionen oder Inhaltsvarianten bis hin zum einzelnen Endgerät.

Vorgehen zur Einschätzung: Die Szenarien wurden hinsichtlich der Gleichheit der anderen Facetten auf den unterschiedlichen Endgeräten des Testbeds und den sich daraus ergebenden Unterschieden geprüft. Sofern deutliche Unterschiede unvermeidbar waren, wurde diese ggf. in der Aufgabenbeschreibung berücksichtigt und als potenzielle Partitionierung für die Auswertung der Ergebnisse vorgemerkt.

6.4 Testeffizienz: Das Prinzip der Bereiche und Abstufungen in der Parametrisierung

Die Wahl der richtigen Parametrisierung ist eine sehr wichtige Entscheidung in der Phase der Testplanung. Nur wenn die Vorabannahmen und Festlegungen dazu korrekt und sinnvoll sind, ist überhaupt mit brauchbaren Testergebnissen zu rechnen. Eine intensive Voruntersuchung und das Aufstellen plausibler Hypothesen für die Parametrisierung zu Art und Größe der Nutzerstimuli sowie eine vorherige Einengung des Parameterraums helfen, die Testeffizienz als Erkenntnisgewinn pro Proband zu steigern. Die gebildete Matrix technischer Parameter mit festen und sinnvollen Kombinationen, orientiert an mobilfunktypischen Werten, sollte als einheitliche Ausgangsbasis

genutzt werden. Für jeden Dienst musste eine Festlegung erfolgen, innerhalb welcher Grenzen und mit welcher Abstufung getestet werden sollte.

6.4.1 Bildung von Testreihen

Prinzipiell ist eine Reihenbildung entsprechend der gebildeten Hypothesen zum QoS-QoE-Zusammenhang sinnvoll, siehe auch Abbildung 59:

- Download-Reihe: Download-Variation (Auswahl in der Matrix parallel zur x-Achse), andere Parameter bleiben weitgehend unverändert
- Upload-Reihe: Upload-Variation (Auswahl in der Matrix parallel zur y-Achse), andere Parameter bleiben weitgehend unverändert
- kombinierte Reihe: Download- und Upload-Reihen kombiniert (diagonal)

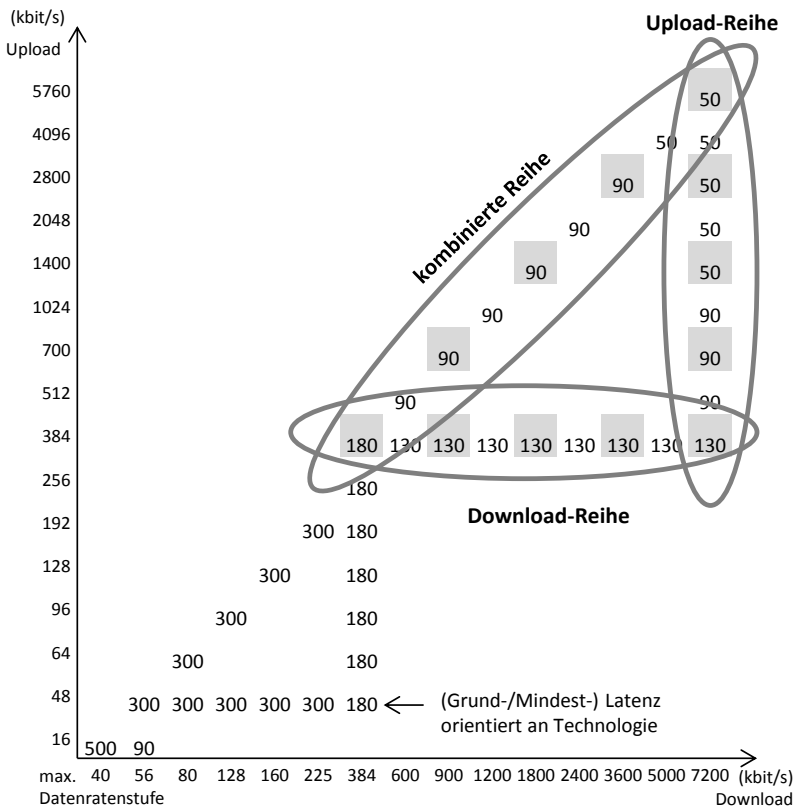


Abbildung 59: Mögliche Ansätze zur Reihenbildung in der Netzparametrisierung
Darstellung jeweils für Beispiele mit 5 ausgewählten Netzparameterstufen (graue Hervorhebung) in der gebildeten Matrix technischer Parameter (vgl. 5.1.6)

Es wurden keine expliziten Latenz-Reihen gebildet. Die Latenz variiert in den Parameterkombinationen passend zu den repräsentierten Technologien und Datenratenstufen automatisch mit. Sofern keine Argumente dagegen sprachen, wurde eine kombinierte Reihe gewählt, in der sich alle drei berücksichtigten QoS-Netzparameter ändern können. Anbetracht der weiteren ohnehin vorhandenen Einflussfaktoren als Nebenfaktoren war bei gegebener Testkonzeption ein Ansatz nach „One-Factor-at-a-Time“ (siehe [26] S. 1 ff für Erläuterungen und ein Beispiel aus der Technik), bei dem sich die Bedingungen immer nur in exakt einem Wert unterscheiden, ohnehin nicht zu realisieren.

6.4.2 Festlegung des Parameterbereichs und der Abstufungen

Die ausgewählten Netzparameterstufen entsprechen den Faktorstufen des Faktors Presets, die in der Testplanung und Auswertung genutzt werden. Eine passende Anzahl Stufen ist prinzipiell schwer abzuschätzen. Bei gegebenem Gesamttestumfang ist zwischen Granularität und Aussagestärke abzuwägen. Eine (zu) große Anzahl Presets bedeuten viele QoS-Varianten und somit eine feinere Auflösung im Parameterraum und vermutlich auch im Bewertungsraum aber gleichzeitig auch nur relativ wenige Ergebnisse je Presets und damit eine niedrige statistische Aussagekraft. Eine (zu) kleine Anzahl Presets hingegen lässt u. U. nur große Sprünge der Bewertungen bei großen Änderungen der QoS-Parameter erkennen, was für die angestrebte Suche nach Schwellwerten zu grob sein kann. Im Umkehrfall, wenn die Stufen zu dicht gewählt sind, lassen sich hingegen keine relevanten Bewertungsänderungen erkennen.

Eine gleiche Anzahl QoS-Faktorstufen für jeden Dienst ist für die Testplanung (gezielte Zuweisung der Faktorstufen an die Probanden) als auch in der Auswertung vorteilhaft. Als Richtwert wurde eine Anzahl von fünf Stufen je Dienst gewählt. Dieser Wert wurde in den Voruntersuchungen als sinnvoll und ausreichend erachtet. So ergaben sich durch die ebenfalls fünf Bewertungsstufen vorteilhafte quadratische Kreuz- bzw. Kontingenztabellen in der Auswertung.

Die Voruntersuchungen zur Kontrollierbarkeit und Beobachtbarkeit lieferten Hinweise dazu, in welchen Grenzen die Parametrisierung sinnvoll sein könnte. Dabei dienten die ermittelten Nutzerstimuli auf Basis der beispielhaften Auf-

gabenausführungen als Anhaltspunkte. Bei großen Unterschieden in der Wiederholbarkeit oder Gleichheit, die eigentlich auf verschiedene QoS-Bereiche und Stufenauswahl hindeuten sollten, wurde ein Kompromiss angestrebt, der möglichst viele der gemeinsam sinnvollen Stufen beinhaltet.

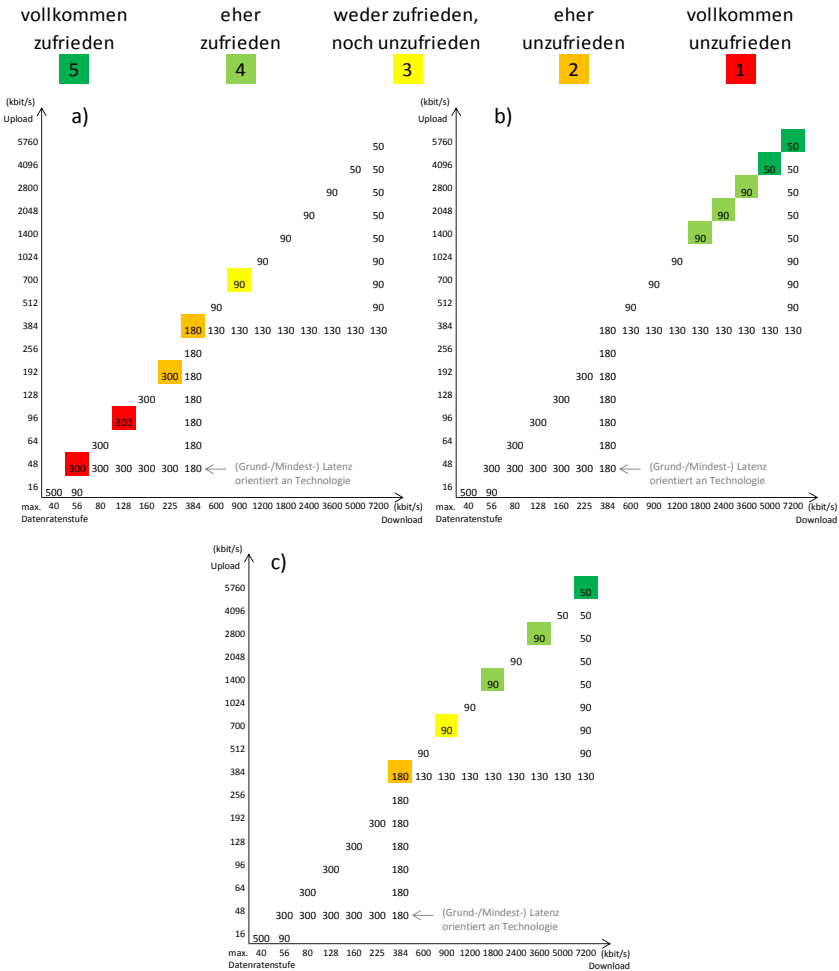


Abbildung 60: Beispiele für die Auswahl von 5 Presets aus der Matrix technischer Parameter
 Darstellung jeweils für Beispiele und die erzielte typische Bewertung (Median) bei einer 5-stufigen Bewertungsskala

- a) Bereich zu niedrig gewählt, keine zufriedene Bewertung
- b) Bereich zu hoch, Abstand zu klein, keine unzufriedene Bewertung
- c) Bereich und Abstand gut gewählt, Übergang zur Zufriedenheit getroffen

Eine Überdeckung des gesamten Bewertungsbereichs von der schlechtesten bis zu zur besten Bewertung, bezogen auf eine typische bzw. gemittelte Antwort, ist nicht zwangsläufig notwendig. Im Rahmen der Tests interessierte hauptsächlich der Bereich des Übergangs von der Unzufriedenheit zur Zufriedenheit. Eine Parametrisierung, die im Mittel¹⁸³ zu „vollkommen unzufrieden“ führt, ist dazu nicht unbedingt erforderlich. Bezogen auf den zuvor genannten interessanten Bereich der Zufriedenheit sollte aber jeweils ein Gegenpol in Form einer nicht zufriedenen Bewertung zu sehen sein. Im Normalfall sollte die Auswahl so erfolgen, dass das mittlere Preset auch ungefähr o. g. interessanten Bewertungsbereich trifft, siehe Abbildung 60.

Ein gleicher Abstand im Sinne fester Intervalle zwischen den ausgewählten Stufen, oder fester Faktoren bezogen auf die Basiswerte, ist nicht zwingend notwendig, kann aber bei der Vermutung entsprechender quantitativer Zusammenhänge durchaus sinnvoll sein (siehe 4.1.9). Da die Auswahl auf Basis der Stufen der Matrix technischer Parameter erfolgte, bedeutet ein gleicher Stufenabstand (z. B. 1 oder 2 Stufen) zwischen zwei benachbarten Presets einen annähernd konstanten Faktor bezogen auf die maximalen Datenraten im Download bzw. Upload.

Bei der Spreizung als dem Abstand zwischen kleinsten und größten Wert ist zu unterscheiden zwischen:

- Parameterraum mit technischen Netzparametern (QoS): z. B. von 384 kBit/s bis 7200 kBit/s
- Stimuli (direkt QoE-relevante Größen): maximal zu erwartende Unterschiede in der Nutzersicht, die sich aus der Spreizung der technischen Parameter ergeben, z. B. von 5 s bis 50 s Wartezeit
- Bewertungsraum mit Nutzerbewertung (QoE): schlechteste und beste erwartete bzw. erzielte Bewertung, skalenabhängig (qualitativ / quantitativ)

Dabei kann es sinnvoll sein, die Spreizung als Differenz oder Vielfaches des Anfangswertes des Bereichs auszudrücken. Eine Überdeckung des gesamten Parameterbereichs der Matrix technischer Parameter mit Faktoren von ca. 300 (bezogen auf die Datenraten im Download und Upload) ist wahrscheinlich für

¹⁸³ Es sei hier vereinfacht von „Mittel“ gesprochen, allerdings ohne die Festlegung auf ein konkretes statistisches Maß wie Median oder arithmetischer Mittelwert.

keinen Dienst sinnvoll, vor allem unter der Maßgabe, die Anzahl ausgewählter Stufen zu begrenzen. Daher ergibt sich, dass die Preset-Auswahl jeweils auf ein *Fenster* für die QoS-Parameter innerhalb der Matrix begrenzt ist.

Die QoS-Spreizung ist passend zur erwarteten und gewünschten QoE-Spreizung zu wählen. Dabei sind plausible Annahmen bzw. Heuristiken zu den einzelnen **Nutzerstimuli** hilfreich, die Bereiche gut abzuschätzen.

Zeiten und Dauern (sofern relevant): maximaler Spreizfaktor ca. 10 bis 20

- Schätzung auf Basis interner Tests (Ausschließen extremer Werte bereits mit wenigen Ergebnissen möglich)
- Annahmen über Boden- und Deckeneffekte, konkret: Dauern für Download- und Upload-Datenübertragung sollten im Bereich ≥ 3 s bis ≤ 60 s liegen, kleinere bzw. größere Werte sollten nicht zu weiteren Bewertungsverbesserungen oder -verschlechterungen führen
- für stark interaktive Nutzungen: ohne merkbliche/störende Verzögerungen (ca. 1 s)

Funktion und Verhalten (sofern relevant): 0 bis ca. 4-malige Auffälligkeiten

- in niedrigen Presets provozieren, mit Funktionseffekten, aber in Relation zum Testumfang, wobei das Verhalten stark bis leicht beeinträchtigt sein sollte
- in hohen Presets vermeiden, keine Funktionseffekte, (Abspiel-) Verhalten etc. ohne Auffälligkeiten

Quantität und Qualität (sofern relevant): maximal 2 Abstufungen oder 2 Varianten

- entweder passend zu den anderen Nutzerstimuli wählen, so dass diese zusammen mit der Netzparameterwahl innerhalb der genannten Bereiche liegen
- oder als zusätzliches Merkmal zur QoE-Differenzierung mit genügend großem Abstand vorsehen, z. B. 2 Qualitätsstufen (HQ vs. HD) oder Inhaltsvarianten (A vs. B)

Zusammengefasst lässt sich das Vorgehen zur Wahl der Parametrisierung wie folgt beschreiben:

- Nutzung der einheitlichen Stufenmatrix technischer Parameter für sinnvolle und typische Netzparameterkombinationen
- unterschiedliche Bereiche und Spreizung je Szenario, die zur Auswahl unterschiedlicher Stufen je Szenario führen können
- 5 Stufen je Szenario mit bevorzugt annähernd gleichem Abstand in der Stufenmatrix, vorrangig in einer kombinierten Reihe
- Eingrenzung und Schätzung auf Basis interner Tests und theoretischer Annahmen, ausgehend von den erwarteten Bewertungen im QoE-Bewertungsraum über die Nutzerstimuli zurück in den QoS-Parameterraum
- Wahl des Bereichs von etwas unzufrieden, über den Übergangsbereich bis hin zur sicheren Zufriedenheit

Für die Zeiten und Dauern können die Ausführungen zum Faktor 4 für die Breite des Toleranzbereiches (siehe 3.6.12) ebenfalls eine plausible Orientierung liefern. Damit ergibt die Verschiebung um eine Stufe (weitere Verdopplung) nach oben *oder* unten den Faktor 8, nach oben *und* unten den Faktor 16.

6.4.3 Festsetzung und Nutzung der Presets

In der Testrunde 1 wurden numerische, fortlaufend vergebene, Preset-IDs genutzt. Die IDs waren global gültig für alle getesteten Dienste. Einziger Vorteil war, dass die gleiche Preset-ID für alle Dienste auch gleiche technische Parameter bedeutete. Dies war aber auch mit Nachteilen verbunden:

- da keine einheitlichen Testreihen gebildet wurden, war die Vergabe nicht so möglich, dass für alle Teilparameter ein höherer ID-Wert auch bessere Parameterwerte bedeutete (monoton steigend für Datenraten bzw. fallend für Latenzwerte)
- unübersichtlich: Lücken und Sprünge in den genutzten Bezeichnungen, da für jeden Dienst nur eine Auswahl der gebildeten Preset-IDs genutzt wurde
- keine gezielte Nachjustierung der repräsentierten Werte der Presets für einen einzelnen Dienst möglich, da jede Änderung sich auch auf die anderen Dienste ausgewirkt hätte, die die gleichen IDs nutzten

In Testrunde 2 wurden die Presets je Dienst definiert. Für Syntax und Semantik und wurden folgende Festlegungen getroffen, siehe auch Abbildung 61:

- 1. Stelle: Präfix aus Dienstkürzel {D, F, G, M, S, Y} entsprechend der Anfangsbuchstaben der Dienste im Testparcours
- 2. und 3. Stelle: numerisch, Netzparameterstufe für diesen Dienst, wird flexibel auf konkrete technische Parameterkombination abgebildet
- die Stufen werden monoton steigend vergeben: größere Werte = bessere Netzparameter = bessere Netzqualität
- anfänglich fünf Netzstufen je Dienst: beginnend mit der niedrigsten ersten Stufe bei 10 bis zur fünften Stufe mit 50

So gebildete Presets können im Rahmen der Voruntersuchungen für jeden Dienst unabhängig von den anderen Diensten noch bis zum eigentlichen Beginn der Probandentests flexibel angepasst werden (z. B. Verschiebung um eine Netzparameterstufe nach oben oder unten), ohne aufwendige Änderungen in der Zuordnung im Testplan zu erfordern.

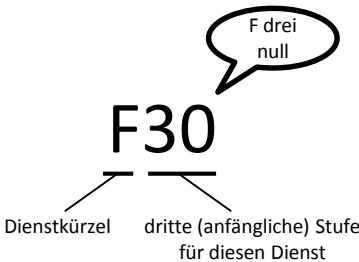


Abbildung 61: Bildung der spezifischen Preset-Bezeichnungen

anfängliche Nutzung
mit **aktiven** Stufen

... 05 ... 10 ... 15 ... 20 ... 25 ... 30 ... 35 ... 40 ... 45 ... 50 ... 55 ...



... 05 ... 10 ... 15 ... 20 ... 25 ... 30 ... 35 ... 40 ... 45 ... 50 ... 55 ...

mögliche Erweiterung zu späterem Zeitpunkt
35 nachträglich eingefügt ohne Bruch der Monotonie

Abbildung 62: Bildung neuer Preset-Stufen bei einer möglichen Erweiterung oder Verfeinerung des Testbereiches

Auch wenn nur eine einstellige Anzahl Netzparameterstufen je Dienst benötigt wird, so ist eine zweistellige numerische Komponente vorteilhaft, um die Eigenschaft der Monotonie in folgenden Fällen zu erhalten, siehe auch Abbildung 62:

- Erweiterung des Testbereiches und Verfeinerung der Abstufung durch Berücksichtigung von Zwischenstufen
- Vergleich bzw. Zusammenführen mehrerer Testrunden aus verschiedenen Presets mit zueinander versetzten Netzstufen

Die abstrakten Preset-Bezeichnungen haben zudem den Vorteil, dass ihnen mehrere Bedeutungen und konkrete Werte oder Tupel entlang der QoS-QoE-Kette des Modells zugeordnet werden können:

- die vorgegebenen QoS-Netzparameterkombinationen
- die typischen QoD-Effektivwerte
- die erwarteten QoA-Werte der Nutzerstimuli
- und letztlich auch die erreichten Bewertungen unter diesen Bedingungen (zusammen mit den wirksamen Nebenfaktoren)

Durch die Entkopplung der konkreten QoS-Werte von den Preset-Bezeichnungen und die dienstspezifische Festlegung ergibt sich als möglicher minimaler Nachteil, dass die Ergebnisse je Preset-Stufe nicht direkt zwischen den Diensten vergleichbar sind, sondern nur über tatsächlich repräsentierte Werte der Netzparameter. Für die Zusammenfassung der Ergebnisse oder deren Darstellung bzw. Visualisierung ist dies allerdings akzeptabel. Die Vorteile der gewählten Systematik überwiegen damit deutlich, sowohl in der Planung als auch in der Auswertung.

6.5 Art und Reihenfolge in der Darstellung des Testparcours

Dieser Abschnitt ist eine Art Muster und Erläuterung zum Lesen der Dokumentation zu den ausgewählten Diensten im Testparcours. Die Darstellung der Vorüberlegungen und Voruntersuchungen zu einem ausgewählten Dienst erfolgt nach einem festgelegten Muster:

- Nutzungsszenario und relevante Aufgabenbestandteile
- technische Sicht/Netzseite und Netznutzung – der Blick von innen
- Nutzersicht – der Blick von außen
- Parametrisierung und Erwartungswerte

Dabei werden die vorherigen Ausführungen zu Testrelevanz sowie Testeignung berücksichtigt und die Ideen zur Testeffizienz durch die gezielte Parametrisierung ergänzt. Insbesondere zur Testeignung können nicht alle Facetten, oder gar die Zusammenhänge der einzelnen Facetten untereinander, für jeden einzelnen Dienst im Detail erläutert werden. Stattdessen werden sie ausgehend von den genannten zwei Sichten des Netzes/Technik bzw. des Nutzers zusammengefasst beschrieben. Dabei werden jeweils die relevanten Nebenfaktoren, vor allem Endgerät, App und Inhalt berücksichtigt. Dies korrespondiert mit der Diskussion der Sichten auf die Dienste und den dienstspezifischen QoS-QoE-Zusammenhang im Modell (Kapitel 3).

6.5.1 Nutzungsszenario

Der Abschnitt Nutzungsszenario liefert die Begründung und Beschreibung eines konkreten Szenarios zu einem Dienst im Test. Die Begründung der Auswahl einer Dienstklasse, bzw. eines konkreten Dienstes oder spezieller Festlegungen im Szenario erfolgen unter dem Gesichtspunkt der Testrelevanz. Mit Blick auf die angestrebte Realitätsnähe sollte das Nutzungsszenario ungefähr einer Realweltnutzung entsprechen – trotz der Künstlichkeit der Laborstudie. Die Probanden sollen die Dienste und Apps möglichst so nutzen, wie es auch im „wirklich wahren Leben“¹⁸⁴ der Fall wäre. Das Nutzungsszenario gibt das Nutzungsziel und damit die relevanten Elemente und Schritte vor. Das Nutzungsszenario sollte dabei so gestaltet werden, dass die gesamte benötigte Zeit

¹⁸⁴ frei nach der Comedy „Dittsche - Das wirklich wahre Leben“

inklusive Lesen und Verstehen der Aufgabe, der eigentlichen Ausführung der Aufgabenschritte und Abgabe der Bewertungen nicht länger als 5 min. dauern sollte.

6.5.2 Technische Sicht und Netznutzung – der Blick von innen

Es ist von plausiblen Annahmen zur Netznutzung auszugehen, die sich u. a. aus der Nutzersicht heraus ableiten lassen. Eine detaillierte Analyse zur Frage „Was geht wann und wie übers Netz?“ ist trotzdem vorteilhaft:

- ex ante – in der Phase der Voruntersuchungen: Der Blick auf die Netznutzung erlaubt das Verständnis der Funktion der Apps hinter der grafischen Benutzerschnittstelle aus Netzsicht – aus der Black Box wird so teilweise eine White Box. Interne Abhängigkeiten können erfasst und qualitativ und quantitativ auf die Auswahl- und Testkriterien hin bewerten werden.
- ex post – in der Phase der Auswertung der Testläufe und Testergebnisse der Probanden: Auf Basis der Verkehrsaufzeichnungen ist teilweise die Validierung der Aufgabenausführung und damit verbunden der Bewertungen der Probanden möglich, bzw. können erreichte Istwerte für die Netzparameter und Nutzerstimuli bestimmt werden (Messraum).

Für die Analyse der Netznutzung wurden die technischen Möglichkeiten des Testbeds genutzt, um Tests in Form der geplanten Aufgabenstellungen zu den Szenarien sowohl ungebremst als auch mit aktiver Netzemulation und Parametrisierung auszuführen. Neben den Verkehrsdaten wurden zusätzlich zugehörige typische Bildschirmausgaben (Screenshots und teilweise auch Videos) sowie die identifizierten Nutzerstimuli erfasst und ausgewertet. Für diesen Blick *von innen* sind die Facetten Kontrollierbarkeit, Echtheit, Wiederholbarkeit und Gleichheit zu berücksichtigen.

Aus der Analyse der Netznutzung, die vorwiegend in Wireshark erfolgte, können typische Muster zur Charakterisierung abgeleitet werden, die den Einzelpunkten der theoretischen Betrachtungen entsprechen, ergänzt um:

- Erkennung der Netznutzung der zugehörigen App, z. B. durch spezifische Merkmale
- Bildung und Betrachtung entsprechender zusammengehöriger Netznutzungen (Pakete, Streams, Sessions ...)

- Trennung verschiedener Phasen in der Netznutzung, die mit der Aufgabebearbeitung korrespondieren
- Identifikation und Isolation der jeweils vermeintlich kritischen Zeitpunkte, Zeitdauern und Zeitabschnitte und zugehöriger Datenmengen in beide Richtungen (Download und Upload)

Im Rahmen der Voruntersuchungen wurden die Dienste isoliert betrachtet. Ausgehend von den so gewonnenen Einzelerkenntnissen und geeigneten Beschreibungsmerkmalen ist in der späteren Auswertung auch eine Analyse der Aufzeichnungen kompletter Testläufe der Probanden als Abfolge der Netznutzungen mehrerer Dienste prinzipiell möglich, siehe 8.1.1.

6.5.3 Nutzersicht – der Blick von außen

Der Blick *von außen* soll so gut wie möglich die relevanten Eindrücke der Probanden bei der schrittweisen Abarbeitung der Aufgaben wiedergeben. Eine umfassende Beschreibung der tatsächlichen Einzelschritte ist der jeweiligen Aufgabenbeschreibung im Anhang zu entnehmen. Einzelne Nutzungsphasen und dazugehörige relevante Schritte der App-Nutzung werden mit Screenshot-Zusammenstellungen rund um die kritischen Nutzerstimuli illustriert. Diese Art der Darstellung im Anhang entspricht somit dem tatsächlichen Blick der Probanden in der Studie: Welche Aktionen waren auszuführen und was konnte wie als Nutzerstimuli inklusive dazugehöriger Präsentation beobachtet werden? Diese Darstellung erfolgt – wie im Test – weitgehend ohne weitere Angaben zu den Zusammenhängen oder Werten, die sich der direkten Beobachtbarkeit entzogen, sondern lediglich um Hinweise zu Unterschieden zwischen Plattformen und Geräten ergänzt. Eine Darstellung aller möglichen Varianten Android, iOS, Smartphone und Tablet, die die (Un-) Gleichheit im Testbed dokumentiert, ist dabei aber nicht zielführend. Stattdessen werden an ausgewählten Beispielen die jeweils typische Nutzersicht dokumentiert und lediglich relevante Unterschiede gesondert erwähnt oder dargestellt. Eine vollständige Wiedergabe aller Schritte ist dabei verständlicherweise ebenso wenig möglich wie eine universelle Darstellung zeitlicher und dynamischer Komponenten. Diese wurden für die Szenarien über die Erfassung der Geräteausgaben mit den geschaffenen Möglichkeiten des Testbeds als Video nachvollziehbar dokumentiert.

Im Hauptteil der Arbeit werden in den Abschnitten zur Nutzersicht die Hintergründe und Zusammenhänge erläutert und bewertet. Dabei sind Beobachtbarkeit, Gleichheit und Benutzbarkeit im Speziellen berücksichtigt. Sofern sinnvoll, wird dabei auf die potenziellen Partitionierungen, z. B. Unterschiede zwischen den Plattformen, eingegangen. Bei der Bewertung der Beobachtbarkeit sind die unterschiedlichen Bedienphilosophien der jeweiligen Plattformen zu beachten, z. B. bei Android die Nutzung des Benachrichtigungsbereiches für Hintergrundaktivitäten.

6.5.4 Parametrisierung und Erwartungswerte

Im Abschnitt Parametrisierung und Erwartungswerte werden die Sicht von außen mit Nutzersicht und Nutzstimuli sowie die Sicht von innen mit technischer Sicht und Netznutzung in einem anwendungsspezifischen Modell zusammengebracht. Entsprechend sind alle einzelnen Facetten dieser Sichten zu berücksichtigen. Mit Blick auf die identifizierten Zusammenhänge und weitere Parameter, z. B. für den Inhalt, erfolgt die Festlegung der QoS-Netzparametrisierung als anwendungsspezifische Presets. Eine Einflussprognose ergänzt die Auflistung der QoS-Parameter mit Richtung, Relevanz und Charakteristik des Zusammenhangs zu den abhängigen Variablen (in der Terminologie und Systematik aus 3.6.2). Die Einflussprognose der QoS-Parameter stützt sich dabei auf die ermittelten Erwartungswerte der Nutzerstimuli für die ausgewählten Netzparameterstufen und deren Einflusscharakter auf die QoE-Bewertungen. Sind aus den Vorüberlegungen und Voruntersuchungen relevante Unterscheidungsfaktoren neben den Presets zu erkennen, dann erfolgt die Darstellung der Erwartungswerte entsprechend dieser Partitionierung.

6.6 Zusammenfassung

In diesem Kapitel wurden die Prinzipien zur Einbeziehung der Dienste in die empirischen Untersuchungen der Studie vorgestellt. Dies umfasst die Begründung der Nutzung echter Dienste, inklusive der Analyse von Vorteilen und Nachteilen im Vergleich zu anderen Ansätzen (Emulation), die Auswahl der Dienste, ihre Voruntersuchung und ihre Parametrisierung im Test.

Dazu wurden die Prinzipien der Testrelevanz, Testeignung und Testeffizienz diskutiert.

Die Testrelevanz als Auswahlkriterium eines mobilen Dienstes kann aus verschiedenen Sichten heraus begründet werden (Nutzer, Netz/Anbieter, theoretische Überlegungen). In der Auswahl der Dienste wurde eine ausgewogene Berücksichtigung aller Sichten angestrebt.

Die Testeignung der Dienste und konkreter Szenarien wurde in den Voruntersuchungen überprüft. Dazu wurden sechs Facetten definiert (Echtheit, Kontrollierbarkeit, Beobachtbarkeit, Benutzbarkeit, Gleichheit, Wiederholbarkeit). Diese Facetten sind verschränkt, was die Überlegungen und Festlegungen erschwert und zu Kompromissen zwingt. Durch Berücksichtigung der relevanten Verschränkungen der Facetten untereinander (Testkristall) wird eine gezielte Überprüfung der Testeignung eines konkreten Dienstes im Rahmen der Voruntersuchungen möglich.

Für eine hohe Testeffizienz wurde das Prinzip der Parametrisierung für die Untersuchungen zum QoS-QoE-Zusammenhang vorgestellt. Aus der entwickelten Matrix technischer Parameter werden je Dienst fünf Netzparameterkombinationen ausgewählt (Presets). Diese bilden eine Testreihe jeweils spezifisch für einen Dienst. Die Bereiche und Abstufungen wurden ausgehend von den Erwartungswerten zur Art des Zusammenhangs, den erreichten Nutzerstimuli und vermutlichen Bewertungen so eingegrenzt, dass möglichst der Übergang von der Unzufriedenheit zur Zufriedenheit empirisch zu ermitteln sein sollte. Die Abstufungen im Parameterraum sind dabei bezogen auf die Datenraten meist mit konstantem Faktor festgelegt (in Erwartung eines annähernd logarithmischen Zusammenhangs zur Bewertung innerhalb der ausgewählten Bereiche).

Die Festlegungen zur Beschreibung der ausgewählten Dienste des Testparcours wurden abschließend vorgestellt. In jeweils vergleichbarer Art und Weise werden dabei Nutzungsszenario, die technische Sicht/Netzseite, die Nutzersicht und die gewählte Parametrisierung mit den Erwartungswerten in der Gesamtsicht des dienstspezifischen QoS-QoE-Zusammenhangs diskutiert.

7 Die Probanden

Der Mensch in der Studie

Der Mensch als Proband spielt eine entscheidende Rolle in der empirischen Bestimmung des QoS-QoE-Zusammenhangs. Entsprechend ist eine intensive Beschäftigung mit den möglichen Besonderheiten des Menschen in dieser Rolle sinnvoll. Daran schließen sich die allgemeinen Punkte zur Beschreibung einer Nutzerstudie an (z. B. Ein- und Ausschlusskriterien, Generierung der Stichprobe etc.). Es folgt die Beschreibung der Stichprobe auf Basis der in einer Nachbefragung erfassten üblichen demografischen Attribute. Diese Attribute sind ebenfalls Variablen. Als Organismusvariablen unterliegen sie einer natürlichen Variiertheit und werden im Test als beigeordnete Einflussgrößen miterfasst.

Ergänzt wurde die Erhebung in der Nachbefragung um speziell interessierende Punkte zur Ausstattung mit mobilen Geräten, der Nutzung des mobilen Internets und der Nutzung populärer mobiler Dienste. Eine Auswertung der Angaben erlaubt Aussagen zur Popularität der Dienste aber auch zu Nutzertypen hinsichtlich der Nutzungsintensität des mobilen Internets. Ein Inter-Rater-Vergleich rundet das Kapitel mit weiteren Aussagen zu den Probanden ab.

Im Anhang finden sich zur Dokumentation die Vor- und Nachbefragung mit der genauen Art der Abfragen und Formulierungen, siehe A.4.1 und A.4.12.

7.1 Der Mensch als Proband

Die Probanden bzw. genauer ihre psychologischen Komponenten bilden einen wichtigen Teil des psychophysikalischen Experiments. Sie werden in der Studie als Bewerter eingesetzt, die die abhängigen Variablen generieren. An dieser Stelle werden dazu die Besonderheiten im Test, z. B. durch das Testdesign oder die Ergebniserfassung, kurz andiskutiert. Die nachfolgenden Punkte beziehen sich hauptsächlich auf die subjektive Sicht der Nutzer aus den Überlegungen zu den relevanten QoE-Faktoren und QoE-Maßen.

7.1.1 Urteilsfehler und psychologische Effekte

Die Probanden sind keine objektiven, unfehlbaren oder gar perfekten QoE-Sensoren, sondern unterliegen in der Bewertung subjektiven Effekten. [216] S. 170f listet eine ganze Reihe möglicher Urteilsfehler und psychologischer Effekte in Nutzertests auf, die ggf. zu beachten sind. Diese beziehen sich hauptsächlich auf Wahrnehmungs- und Beurteilungseffekte als auch auf das Verständnis und die Interpretation der Skala. Zur Skala wurden bereits die Annahmen bei der Festlegung erörtert, weitere Diskussionen speziell zum Skalenniveau folgen noch in der Auswertung.

Im Test sind neben allgemeinen und generischen Effekten wie Langeweile oder Ermüdung auch studienspezifische Effekte möglich. Diese beziehen sich hauptsächlich auf den Prozess der Nutzung der mobilen Dienste im Testparcours wie in der QoE-Theorie erörtert. Im Test ist von der Wirkung der micro- und macro-temporalen Aspekte auszugehen. Die macro-temporale Komponente beinhaltet die gesuchte Erwartungshaltung, aus der sich die Anforderungen ableiten lassen. Die micro-temporale Komponente innerhalb des Tests ist durch die psychologischen Effekte eher eine ungewollte Störgröße auf Erwartungsseite.

Die möglichen psychologischen und zeitlichen Effekte im Nutzungsprozess können sich durch Testdesign und Testparcours erklären:

- Übertragungs-/Ausstrahlungseffekte wie die nachfolgenden
- bei Wiederholung: erste Aufgabenbearbeitung beeinflusst die zweite Bewertung, z. B. erst schnell, dann langsam – oder anders herum (Carry-over-Effekt, [283] S. 10)
- Erinnerung an vorherige Aufgabenausführung und Vergleich beider Aufgabenausführungen (Memory-Effekt, siehe auch [97])
- Verzerrungen durch Abfolge der Parametrisierung und damit der Bewertungen, besonders bei Extremwerten, bei denen die beste oder schlechteste Bewertung bereits vergeben wurde (Primacy-Recency-Effekt, siehe [216] S. 171)
- Nachwirken und Ausstrahlen besonders guter oder schlechter Erlebnisse auf die nachfolgenden Nutzungen (Verschiebung der Zone of Tolerance, siehe [98])

- Zusammenfassung und Verallgemeinerung statt isolierter Betrachtung jeder einzelner Aufgabenausführung und Beurteilung der spezifischen Teilzufriedenheit unter dem Gesamteindruck (Halo-Effekt, siehe [216] S. 170)
- Übertragen der Erfahrungen und Vergleich zwischen den Erlebnissen verschiedener Dienste
- Abfolge und Position der Aufgaben (Positionseffekt, z. B. durch zunehmende Ermüdung oder Langeweile, siehe [283] S. 10)
- Zusammenhang der Bewertung(en) und des Bearbeitungserfolgs (Interaktionseffekt, Bewertung hängt von Bearbeitungserfolg ab)

Aus Testsicht sind die genannten möglichen zusätzlichen Einflussfaktoren entweder Kovariablen oder Störvariablen. Im Modell wirken diese Effekte durch eine Verschiebung auf der Seite der Erwartung. Da diese mögliche Verschiebung sich nicht technisch erfassen oder nachvollziehen lässt, bleibt sie nachfolgend in der Auswertung unbeachtet. Es wird zur Vereinfachung davon ausgegangen, dass solche Effekte auf Erwartungsseite zwar möglicherweise vorhanden sind, im Vergleich zur Veränderung der Leistung auf Erlebnisseite aber nicht relevant sind bzw. in den Ergebnissen ausgemittelt vorliegen (Variation der Reihenfolge und Abfolge der Szenarien und Parametrisierungen durch den Testplan).

Aus dem grundsätzlichen Umstand einer Studie, dass die Nutzer untersucht werden, kann sich der sogenannte Hawthorne-Effekt ergeben (siehe [172] S. 76 mit weiteren Erläuterungen und Verweis auf die ursprüngliche Beschreibung aus dem Jahr 1933 in [284]). Der Hawthorne-Effekt beschreibt ein anderes Verhalten der Probanden in einer Studie im Vergleich zur routinemäßigen Erledigung, das sich durch besondere Aufmerksamkeit und Anstrengung bemerkbar macht. Im ungünstigsten Fall führt dies zu unbrauchbaren Ergebnissen [172] S. 76.

7.1.2 Aufgabenausführung

Eine besondere Art von Urteilsfehler ist im Fall der vorliegenden Laborstudie eher ein „Urteilerfehler“, also ein Fehler, der vom Probanden selbst ausgeht. Konkret geht dieser von der Teilkomponente Hand, also der Interaktion der Probanden mit der Technik des Testbeds und den Diensten des Testparcours

aus. Die Probanden erzeugen sich die zu beurteilenden Stimuli im Rahmen der Aufgabenausführungen selbst. Entsprechend können durch fehlerhafte oder abweichende Aufgabenausführung falsche, also ungeplante bzw. unerwartete Stimuli auf den Probanden einwirken und von diesem bewertet werden. Dies ist dann problematisch, wenn dadurch die Bewertung nicht mehr für den interessierenden QoS-QoE-Zusammenhang charakteristisch ist, die Bewertung also verzerrt wird und die Auswertung dies nicht berücksichtigt.

7.1.3 Kontext und Fokus

In diesem Abschnitt soll, ausgehend von den genannten Grundüberlegungen, die Bedeutung des Kontextes für den eigenen Test abgeschätzt werden. Dies umfasst neben Interaktionskontext, situativem Kontext, soziokulturellem Kontext auch die Netzwahrnehmung durch die Probanden und mögliche Verschiebungen der Aufmerksamkeit.

Interaktionskontext und Netzwahrnehmung: Zwischen Interaktionskontext und situativem Kontext ist die Wahrnehmung der (Mobilfunk-) Netzeigenschaften auf den Endgeräten im Labortest einzuordnen. Die Netzwahrnehmung ist durch die Eigenschaften des Testbeds definiert und wird dort detailliert dargestellt, siehe 5.8.5. Die Nutzer wurden nicht explizit befragt, was sie glaubten oder erwarteten, wie die mobilen Geräte Zugang zum Internet erhielten, noch erhielten sie dazu eine Aussage. Ein Blick auf die Nutzerkommentare hinsichtlich WLAN oder mobilem Internet lieferte 13 Hinweise zur Wahrnehmung durch die Probanden (genaue Auflistung im Anhang A.6.1), was angesichts der Gesamtzahl Probanden und Kommentare recht wenig ist. Dabei waren 8 Äußerungen zur Wahrnehmung als WLAN und 5 Äußerungen zur Wahrnehmung als mobiles Internet festzustellen.

Die Probanden waren sich wohl bewusst, nicht wirklich das mobile Internet über einen Mobilfunkzugang zu nutzen. Entsprechend sollte nicht von einer äquivalenten Erwartungserhaltung der Probanden ausgegangen werden. Sinnvolle Annahmen sind daher entweder eine neutrale Erwartungshaltung hinsichtlich der Zugangstechnologie oder die Erwartungshaltung zu einer WLAN-Umgebung.

Situativer Kontext und Fokus: Im Sinne des QoE-Modells ist davon auszugehen, dass in einer Laborumgebung a priori kein natürlicher Kontext für die Nutzung von Diensten im mobilen Internet vorliegt. Es sind deshalb keine direkten Aussagen möglich, ob die Ergebnisse auf einen anderen situativen Kontext (z. B. Nutzung unterwegs) übertragbar sind, lediglich plausible Annahmen dazu sind möglich. In der Laborsituation werden sowohl eine künstliche Aufmerksamkeit (erzwungener Vordergrund) als auch eine künstliche Ablenkung (Lesen der Aufgabenstellung) erzeugt, die den Fokus der Probanden im Vergleich zu anderen Nutzungssituationen beeinflussen können. So ist es möglich, dass außerhalb einer Laborsituation auch mit einer Verschiebung einzelner bewerteter Aspekte zu rechnen ist. Im Kontext einer Laborsituation ist zunächst von aktivem Warten auszugehen, ggf. abgeschwächt durch Ablenkung außerhalb der mobilen Nutzungssituation, z. B. durch weiteres Lesen der Fragen im Fragebogen, Umschauen im Testraum, Beobachten der hübschen Probandin nebenan etc.

Die Probanden führen die Aufgaben (zumindest teilweise) Schritt für Schritt aus. Dabei wandert der Fokus der Probanden zwischen der Aufgabenausführung auf dem mobilen Gerät und dem Fragebogentext auf dem Bildschirm hin und her. Im Test wurde versucht, durch Hinweise erst die Aufgabenstellungen vollständig zu lesen und dann abzuarbeiten, die Aufmerksamkeit bzw. Ablenkung zumindest etwas zu steuern und Fehler bei der Bearbeitung zu vermeiden. Ob dies wie gewünscht funktioniert hat, kann nicht überprüft werden.

Die Wahrnehmung der Zeit bei Wartedauern kann ggf. durch die Laborsituation beeinflusst sein:

- Eine weniger starke Wahrnehmung ist möglich, z. B. durch das Lesen der Aufgabenstellung Punkt für Punkt und Ablenkungen durch sonstige äußere Reize. Wartedauern könnten generell als weniger unangenehm empfunden werden, es hängt ja nichts davon ab.
- Eine starke Wahrnehmung ist möglich, weil Wechsel der Vordergrundaktivität entfallen. Die Nutzer bleiben bei der Aufgabe und warten das Ende einer Übertragung etc. ab. In anderen Situationen wechseln sie hingegen vielleicht die App und kehren erst nach Beendigung des Downloads etc. zur Fortsetzung der Nutzung zurück.

Für die Beurteilung der Zufriedenheit mit Funktion und Verhalten sowie mit Qualität und Quantität ist von weniger großen Unterschieden als bei der der Zeitwahrnehmung auszugehen, da vermutlich weniger durch eine mögliche Fokusverschiebung beeinflusst.

[85] widmet sich der Frage, wie stark die Abweichungen in den Ergebnissen von QoE-Studien sind, je nachdem ob sie im Kontext einer Laborsituation oder in einer Feldstudie erhoben wurden. Zwar unterscheiden sich die dort betrachteten QoE-Maße (Zufriedenheit über MOS-Skala und Akzeptanz) und der Interaktionskontext (Nutzung am Laptop), die Ergebnisse sind aber recht eindeutig. Demnach ergeben sich bei gleichen Zufriedenheitsbewertungen (MOS) deutliche Unterschiede in der Bewertung der Akzeptanz. In den Feldstudien waren die Probanden gerade bei niedrigen MOS-Bewertungen weit weniger gewillt, die gebotene Leistung als akzeptabel einzustufen, es waren also Unterschiede in der Anspruchshaltung zu erkennen. Auf Basis dieser Erkenntnisse ist es plausibel, außerhalb einer künstlichen Laborumgebung, also in einem echten Nutzungskontext, von einer höheren Anspruchshaltung auszugehen.

Wie auch beim Nutzen gilt: Die Kosten lassen sich schlecht in eine artifizielle Laborstudie integrieren. Gleiches gilt damit für Kosten-Nutzen-Abwägungen (keine echten Kosten, kein echter Nutzen), von denen auszugehen ist, dass sie außerhalb eines Laborkontextes die Zufriedenheit modulieren.

Insgesamt dürfte eine nachgebildete Nutzungssituation eines Laborexperiments nicht 100-prozentig und unmittelbar einer normalen Nutzung außerhalb der Laborsituation entsprechen. Über die dadurch verursachten möglichen Implikationen für die Ergebnisse kann keine weitere Aussage getroffen werden. Da dies aber kein unmittelbar technisches Problem ist, richtet sich der weitere Forschungsbedarf an dieser Stelle an die Humanwissenschaften.

Soziokultureller Kontext: Aspekte zum soziokulturellen Kontext der Probanden, z. B. zur Bildung etc., wurden im Rahmen einer Nachbefragung separat erfasst und werden als Attribute der Probanden in der nachfolgenden Beschreibung der Stichprobe detailliert gesondert dargestellt.

7.2 Allgemeine Punkte zu den Probanden

In diesem Abschnitt werden die allgemeinen Punkte gebündelt, die üblicherweise für empirische Studien mit Beteiligung des Menschen als Probanden zu finden sind. Dies umfasst das Probandenwissen, Incentives und Angaben zur Generierung der Stichprobe von Probanden.

7.2.1 Probandenwissen

Es ist davon auszugehen, dass die angeworbenen Probanden aus dem Kreis der normalen Nutzerschaft keine Experten im Sinne des Beurteilens und Erstellens von QoE-Bewertungen sind, bzw. dazu auch kein spezielles Vorwissen mitbringen (naive Nutzer). Die Probanden erhielten vorab kein Training, weder für die Aufgabenbearbeitung noch zur Bewertung. Während der Testausführung wurde kein spezielles Feedback dazu erteilt. Aus der Konzeption der Studie ergab sich, dass die Probanden möglichst unbeeinflusst an die Bearbeitung der Aufgaben gehen sollten und entsprechend auch kein spezielles Expertenwissen notwendig sein sollte. Die Probanden sollten nicht auf eine ausschließliche Leistungsmessung der Netzparameter fixiert werden, um ein „jetzt nutze ich mal die Stoppuhr“ zu vermeiden. Die Instruktionen lauteten entsprechend, die gesamte Bearbeitung der Aufgabe im Sinne der Nutzung eines mobilen Dienstes zu bewerten. Als eine Art „Blindtest“ ausgelegt, ergibt sich eine mögliche Charakterisierung des Probandenwissens wie folgt (angelehnt an [207] S. 9).

Wussten die Probanden...:

- ... dass sie untersucht werden?
Ja, dies lässt sich in einer Laborstudie nicht vermeiden.
- ... wie die Bedingungsänderungen aussehen?
Nein, die variierten Netzparameter waren durch die Probanden nicht sofort direkt als Messwert etc. zu ermitteln. Ein Proband, der explizit darauf aus ist, wird aber sicher einen Weg finden, z. B. durch Nutzung entsprechender Apps oder Web-Dienste.
- ... was genau untersucht wird? Konnten die Probanden die Untersuchungshypothese?
Nein, die Probanden konnten sie aber aus den Aufgaben, den Fragen und

den beobachteten, teilweise sehr deutlichen Änderungen der Testbedingungen erahnen. Dies legen einige vereinzelte aber recht gezielte Nachfragen der Probanden im Testverlauf nahe.

7.2.2 Schriftliche Information, juristische Absicherung und Studienethik

Vor Beginn der Tests erhielten die Probanden die Gelegenheit, sich den Schriftstücken zur juristischen Absicherung zu widmen und gegenzuzeichnen:

- Probandeninformation zur Studie und den erhobenen Daten
- Einwilligungserklärung zur freiwilligen Teilnahme an der Studie und Möglichkeit zum jederzeitigen Abbruch
- datenschutzrechtliche Einwilligungserklärung zur anonymen bzw. pseudonymisierten Verarbeitung, Speicherung, Weitergabe und Auswertung der Daten sowie der Möglichkeit zum Widerruf

Die Erklärungen orientieren sich an den Ethikempfehlungen der TU Chemnitz und wurden durch den Partner AAP ausgearbeitet. Je ein Exemplar der unterschriebenen Erklärungen erhielten die Probanden, eines verbleibt zur Dokumentation dauerhaft beim Testausrichter. Nach aktuellem Stand der Wissenschaft gehen von der Teilnahme an der Studie für die Probanden keine Gefahren oder Gefährdungen aus.

7.2.3 Instruktionen und Hilfestellungen

Jeder Proband erhielt eine individuelle kurze technische Instruktion zum jeweiligen Endgerät und zum vollständigen Schließen der Apps zwischen den Aufgabenbearbeitungen. Zusätzlich gab es eine kurze gesonderte Anleitung als eine Art Spickzettel an den Arbeitsplätzen. Zusätzlich zur schriftlichen Probandeninformation erhielten die Probanden unmittelbar vor Testbeginn noch letzte Informationen in Form einer mündlichen Instruktion zur Studie. Ein Leitfaden diente dazu, alle wichtigen Punkte für alle Testdurchläufe standardisiert und somit vergleichbar zur Erwähnung zu bringen.

Mit Blick auf das angestrebte Probandenwissen zur Studie und um fehlgeleitete Bewertungen zu vermeiden, wurde hingewiesen, dass diese Punkte nicht *isoliert* im Zentrum der Studie standen:

- Bewertung der Geräte
- Bewertung der Softwareplattformen (besonders im Vergleich untereinander)
- Bewertung der Apps (wie in den App-Stores)
- Bedien- und Verständnisprobleme (Möglichkeit zur Nachfrage und gesonderte Anmerkungsfelder)

Nach dem abschließenden Hinweis, dass bei Problemen und Fehlern die Testleitung jederzeit ansprechbar ist, starteten die Probanden in die selbstständige Bearbeitung der Aufgaben. Im Rahmen gelegentlicher Hilfestellungen war die Testbetreuung hauptsächlich bei kleineren Bedienproblemen gefordert. Sämtliche solcher Eingriffe und andere besondere Vorkommnisse wurden detailliert in einem Testlogbuch erfasst. Auskünfte zur konkreten Netzparametrisierung wurden nicht erteilt. Einige wenige Nachfragen dieser Art zeigen aber, dass zumindest ein Teil der Probanden betreffs der Untersuchungsinhalte und der Testparametrisierung richtig schlussfolgerte.

7.2.4 Incentive

Die Probanden erhielten ein Incentive als finanzielle Aufwandsentschädigung für die Teilnahme an der Studie. Als zusätzlicher Teilnahmeanreiz bestand die Chance auf den Gewinn eines hochwertigen mobilen Endgeräts in einer Verlosung am Ende der Studie. Alternativ bestand die Möglichkeit, die Studienteilnahme als Versuchspersonenstunde¹⁸⁵ werten zu lassen. Das Management der Incentives übernahm der Partner AAP.

7.2.5 Anwerbung, Screening und Auswahl der Probanden

Die Anwerbung der Probanden erfolgte über Mailing-Listen, Handzettel, Mundpropaganda und Digital Signage. Aus den Erfahrungen der Testrunde 1 war von vornherein klar, dass die Stichprobe durch die forcierte Probandenakquise an der TU Chemnitz und in ihrem Umfeld jung und gut gebildet ist. Diese Probanden finden sich hauptsächlich im Panel „i“ (interne Anwerbung

¹⁸⁵ Teil der Studienleistungen für ausgewählte Studiengänge an der TU Chemnitz in den Humanwissenschaften

und Auswahl durch Partner AAP). Zusätzlich wurden noch weitere Probanden durch einen Dienstleister angeworben, das Panel „e“ (externe Anwerbung und Auswahl). Das Verhältnis der Probanden in den Panels i:e beträgt rund 3:1. Ein ausgewogenes Geschlechterverhältnis wurde angestrebt und erreicht.

Die Erfassung und Auswahl der Interessenten erfolgte über einen Screening-Fragebogen, siehe Anhang. Die Hauptauswahlkriterien waren Vorkenntnisse im Umgang mit den mobilen Betriebssystemen im Test (Android oder iOS), die persönliche (private) Nutzung eines entsprechenden Smartphones oder Tablets und Erfahrungen in der Nutzung des mobilen Internets über einen Mobilfunk-Provider.

Die Einladung zu einem konkreten Termin für einen Testslot erfolgte durch Abgleich der Angaben der Probanden zum mobilen Betriebssystem passend zu dem der Testgeräte auf den Testplätzen im Testbed. Es konnten nicht durchgängig alle Testplätze besetzt werden, es verblieben aber auch Interessenten, die nicht in den Test aufgenommen werden konnten.

7.3 Beschreibung der Stichprobe

Dieser Abschnitt der Arbeit widmet sich der Beschreibung der Stichprobe im Test. Dabei werden nur Angaben zu Probanden nach der Vorvalidierung gemacht, siehe 8.2.1 für weitere Erläuterungen. Mit Blick auf die Stichprobe, also die Probanden, die an der Studie teilgenommen haben, soll auf wichtige Fragen besonders eingegangen werden: *Haben die richtigen Nutzer an der Studie teilgenommen? Ist die Auswahl und Zusammensetzung der Stichprobe annähernd repräsentativ?*

Von den Probanden wurde eine Reihe persönlicher Daten erhoben. Diese umfassen das Alter, Geschlecht, Bildung und den beruflichen Status. Darüberhinausgehend wurden von den Probanden studienrelevante Angaben zum Besitz internettauglicher mobiler Endgeräte, ihren mobilen Internetzugang, die Erfahrung mit dem mobilen Internet und die Häufigkeit der Nutzung ausgewählter mobiler Internetdienste abgefragt. Die erfassten Angaben sind Attribute,

d. h. Eigenschaften der Probanden, die nicht aktiv gezielt im Test variiert werden konnten, sondern lediglich zur Auswahl und Stichprobenabgrenzung dienen.

In der nachfolgenden Diskussion soll deshalb geklärt werden:

- Welche Attribute und Ausprägungen weisen die Probanden auf?
- Wie sieht eine zur Stichprobe der Probanden passende Grundgesamtheit aus?
- Welche Ausprägungen weisen die Attribute einer Vergleichsgruppe auf?
- Welche Attribute können ggf. wie auf die QoE-Bewertungen Einfluss nehmen?
- Wo lassen sich durch Zusammenfassung von Ausprägungen sinnvolle Gruppen bilden?

In Bezug auf die Rolle der demografischen Attribute wie Alter oder Geschlecht auf die QoE verweist [64] S. 95 auf die Anwendung des Modells zur Unified Theory of Acceptance and Use of Technology (UTAUT) für einen ausgewählten Online-Dienst (Banking) in [285]. Demnach können die demografischen Attribute als Moderatoren für die QoE wirken. Die Auseinandersetzung mit derartigen psychologischen Modellen geht über den Rahmen der eigenen Betrachtungen hinaus, wäre aber sicherlich ein interessanter Anknüpfungspunkt für eine Annäherung an die Thematik aus Sicht der Psychologie. Statt der ganzheitlichen Sicht soll jeweils nur kurz und isoliert auf mögliche Einflüsse der einzelnen Attribute verwiesen werden.

7.3.1 Annahmen zur Grundgesamtheit

Die Probanden im Test bilden eine Stichprobe aus einer Grundgesamtheit. Zur Grundgesamtheit können nur eingeschränkt Aussagen getroffen werden. Es ist entsprechend abzuwägen, ob prinzipiell von einer größtmöglichen Grundgesamtheit ausgehen ist, was die Gesamtbevölkerung Deutschlands wäre, oder von einer kleineren Teilmenge. So könnte sich die Grundgesamtheit auch auf die Nutzer mobiler Internetdienste auf modernen Endgeräten oder gar die Kundenstruktur eines bestimmten Anbieters beschränken. Daneben sind auch spezielle Zielgruppen denkbar, z. B. Studenten, urbane gut gebildete Erwerbstätige oder Erwachsene im erwerbsfähigen Alter allgemein.

Wo immer möglich und sinnvoll werden zu den eigenen Ergebnissen passende Vergleichszahlen genannt. Dies erfolgt allerdings explizit ohne den Anspruch, damit eine Grundgesamtheit darzustellen, sondern lediglich zur größenordnungsmäßigen Orientierung. Über die Verteilung der im Test untersuchten Merkmale der QoE (Zufriedenheit) in der Grundgesamtheit lassen sich a priori keine Aussagen treffen.

7.3.2 Geschlechterverhältnis und Altersstruktur

Das Mindestalter zur Teilnahme an der Studie betrug 18 Jahre. Durch den Fragebogen wurde das Alter in Jahren erfasst. Für die Darstellung wurden Altersklassen gebildet, für die neben der Gesamtheit der Stichprobe auch nach Geschlecht und Anwerbungskanal (Panel) unterschieden wird, siehe Tabelle 32.

Tabelle 32: Geschlechterverhältnis und Altersstruktur der Probanden der Testrunde 2

Alters- klasse	Gesamt			nach Geschlecht		nach Panel	
	absolut	relativ		absolute	Häufigkeit	absolute	Häufigkeit
	Anzahl	Anteil	akkumul.	männlich	weiblich	e	i
18 - 20	27	12,86%	13%	6	21	6	21
21 - 23	54	25,71%	39%	24	30	6	48
24 - 26	49	23,33%	62%	25	24	4	45
27 - 29	35	16,67%	79%	26	9	7	28
30 - 32	19	9,05%	88%	8	11	8	11
33 - 35	10	4,76%	92%	6	4	7	3
36 - 38	2	0,95%	93%	1	1		2
39 - 41	1	0,48%	94%		1		1
42 - 44	3	1,43%	95%	2	1	3	
45 - 47	1	0,48%	96%	1		1	
48 - 50			96%				
51 - 53	1	0,48%	96%		1	1	
54 - 56	2	0,95%	97%	1	1	1	1
57 - 59	3	1,43%	99%	2	1	3	
60 - 62	2	0,95%	100%	1	1	2	
63 - 65	1	0,48%	100%		1	1	
gesamt	210			103	107	50	160
relativ		100%		49%	51%	24%	76%
Lage- und Verteilungsparameter (Angaben zum Alter in Jahren)							
Modalwert	22			27	20	21	22
Median	25			26	24	30	24
Mittelwert	26,9			27,6	26,2	33,4	24,9
Std.-Abw.	8,3			7,9	8,6	12,9	4,6

Panel zur Unterscheidung des Anwerbungskanals: e = extern, i = intern

In einzelnen Altersklassen existieren größere Unterschiede (18-20) und (27-29), das Geschlechterverhältnis insgesamt ist aber ausgeglichen. Dies entspricht damit näherungsweise dem nahezu ausgeglichenen Geschlechterverhältnis der Smartphone-Nutzer in Deutschland¹⁸⁶. Die Altersstruktur im Test wird durch die sehr jungen Probanden dominiert, die hauptsächlich aus dem universitären Umfeld stammen (Studenten). Das Panel e ist geringfügig älter als das Panel i. Es sind sehr deutliche Unterschiede zur Altersstruktur der Gesamtbevölkerung zu erkennen (vgl. [287]¹⁸⁷). Vergleichszahlen zur Alterszusammensetzung einer möglichen Grundgesamtheit von Nutzern mobiler Datendienste sind schwierig zu ermitteln, da neben der Altersstruktur auch der Besitz geeigneter mobiler Endgeräte sowie deren Nutzung im mobilen Internet zu berücksichtigen wären.

Das Alter ist prinzipiell ein Attribut, das sich auf die QoE-Bewertungen auswirken kann. Durch die unterschiedliche Techniksozialisation und den damit verbundenen Umgang mit der Technik ist dies eine plausible Annahme (siehe o. g. Modelle). Eine zu feingranulare Sicht ist dabei allerdings nicht zielführend. Deshalb kann auch eine Einteilung in größere Alterskohorten bzw. Generationen erfolgen. Dies ist sinnvoll, da das Aufwachsen mit digitaler Technik nur für die *Generation Y* der „Digital Natives“ zutrifft, wohingegen die früheren Generationen damit erst im Erwachsenenalter in Kontakt kommen konnten, die sogenannten „Digital Immigrants“, siehe [288]. Die Jahrgangsgrenzen für die Generationen sind nicht wirklich fixiert und variieren je nach Quelle. [289] S. 8ff. nennt einige weiterführende Quellen mit konkreten Zahlen und Abgrenzungen, resümiert aber richtig, dass die Jahresgrenzen generell nicht fest sind, da die Generationsübergänge ohnehin fließend sind. Je nach Variante der Festlegung der Generationsgrenzen, siehe Tabelle 33, sind zwischen knapp unter 80 % bis knapp über 90 % der Probanden der *Generation Y* zuzuordnen, der Generation der *Baby Boomer* hingegen nur rund 4 %, der *Generation X* der jeweilige Rest. Damit ergibt sich, dass für die weitere Auswertung mit Blick auf das Alter nur eine Unterscheidung zwischen den Generationen sinnvoll möglich wäre. Dabei ist nur für Generation Y ein ausreichend großer Umfang gewährleistet. Die übrigen Probanden könnten nur in einer Resthäufigkeitsklasse zusammengefasst und vernachlässigt werden. Aufgrund

¹⁸⁶ Umfrage im Dezember 2012 in Deutschland: 54% männlich, 46% weiblich, siehe [286] S. 35

¹⁸⁷ Zahlen auf Basis des Zensus 2011 in Deutschland

dieser Struktur wird im Weiteren auf eine systematische Unterscheidung nach Alter oder Generationen verzichtet bzw. auf Generation Y fokussiert.

[289] S. 17f. geht kurz auf die allgemeinen Eigenschaften der Generation Y sowie ihr Verhältnis zur Technik ein und listet weiterführende Quellen auf. Die wichtigsten Punkte zur Charakterisierung der Generation Y sind (nach [289] S. 17f. und [290] S. 6ff.):

- kurze Aufmerksamkeitsspanne
- (scheinbar) ständiger Zeitmangel
- multimediales Multitasking
- permanente Kommunikation auch jenseits von Telefon und E-Mail, z. B. Instant Messages, Facebook & Co.
- Internet, Mobilkommunikation und mobiles Internet sind im Alltag integriert, allgegenwärtig und selbstverständlich
- “always on” und “always connected”
- Das Internet ist ein wichtiger Teil des Lebens: Die vernetzte Welt ist nicht nur normal, sondern notwendig für die Organisation des Alltags und der Arbeit.

Tabelle 33: Generationsstruktur der Probanden in Testrunde 2

Varianten und Einordnung	Generationen		
	BB <i>Baby Boomer</i>	X <i>Digital Immigrants</i>	Y <i>Digital Natives</i>
Variante 1 ¹⁸⁸			
Jahrgänge	1950–1963	1964–1983	1984–2000
Alter im Test	50–63	30–49	18–29
Anteil im Test	4,3 %	17,1 %	78,6 %
Variante 2 ¹⁸⁹			
Jahrgänge	1946–1964	1964–1980	1981–2000
Alter im Test	50–67	33–49	18–32
Anteil im Test	4,3 %	8,1 %	87,6 %
Variante 3 ¹⁹⁰			
Jahrgänge	1950–1964	1965–1977	1978–1997
Alter im Test	49–63	36–48	18–35
Anteil im Test	4,3 %	3,3 %	92,4 %

n=210

¹⁸⁸ Orientierung an Unterscheidung Ü18 bzw. U30 zu Ü30 bzw. Ü50 und Ü65 wie beispielsweise in den zitierten Zensusdaten, setzt den Generationenwechsel X zu Y sehr spät an

¹⁸⁹ Orientierung an Einteilung aus [289] S. 9 und [290] S. 4

¹⁹⁰ Orientierung an Einteilung aus [291], setzt den Generationenwechsel X zu Y sehr zeitig an

Es ist also sinnvoll von möglichen Unterschieden der Generationen auszugehen in:

- Vorerfahrungen und Kenntnissen im Umgang mit der Technik (hier mobile Endgeräte und mobiles Internet)
- Erwartung (hier in der Nutzung und zuvor erlebten Leistung des mobilen Internets)
- Art der Nutzung (hier die genutzten Dienste)
- Usability: Ansprüche und Verständnis (hier Aufbau der Apps und Wahrnehmbarkeit der Eigenschaften)

7.3.3 Bildung und beruflicher Status

Die Probanden haben zu ihrem höchsten erreichten Bildungsabschluss per Fragebogen Auskunft erteilt. Bei der Ergebnisdarstellung wird aufgrund der geringen Unterschiede nicht zwischen den Testblöcken und Panels unterschieden. Es ist ein sehr hoher Anteil von Probanden mit Abitur bzw. sogar (Fach-) Hochschulabschluss festzustellen, siehe Tabelle 34.

Tabelle 34: Höchster Bildungsabschluss der Probanden

höchster Bildungsabschluss	Häufigkeit	
	absolut	Relativ
(Fach-) Hochschulabschluss	83	39,5 %
(Fach-) Hochschulreife	103	49,0 %
Mittlere Reife	20	9,5 %
Hauptschulabschluss	4	1,9 %
ohne Abschluss	0	0 %
gesamt	210	100,0 %

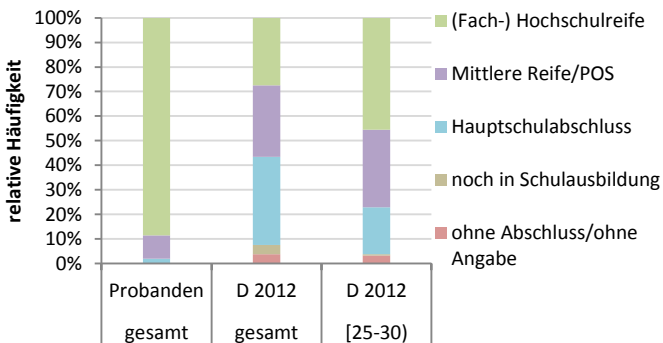


Abbildung 63: Höchster Schulabschluss der Probanden im Vergleich

Probanden n=210, Vergleich zur deutschen Gesamtbevölkerung und einer ausgewählten jüngeren Alterskohorte 25 bis unter 30; Vergleichszahlen aus dem Mikrozensus 2012, [292] S. 26

Abweichend von der üblichen Praxis zwischen allgemeinbildendem Schulabschluss und weitergehender Ausbildung (Beruf/Hochschule) zu trennen, zählen (Fach-) Hochschulabschlüsse an dieser Stelle mit zum höchsten Bildungsabschluss. Um einen Vergleich zum Schulabschluss der Gesamtbevölkerung zu ermöglichen, wurden in den Fragebogendaten die Angabe (Fach-) Hochschulabschluss mit der Angabe (Fach-) Hochschulreife zusammengefasst und einige Vereinfachungen¹⁹¹ zu den Vergleichszahlen genutzt. Es sind deutliche Unterschiede im Bildungsniveau der Stichprobe zur Gesamtbevölkerung und auch zu jüngeren Bevölkerungsgruppen zu erkennen, siehe Abbildung 63. Diese Unterschiede sind vor allem durch ein sehr großes Übergewicht der (Fach-) Hochschulreife in der Stichprobe gekennzeichnet.

Tabelle 35: Beruflicher Status der Probanden

beruflicher Status	Rohangaben		korrigiert	
	Absolut	relativ	absolut	relativ
voll berufstätig	61	29,0 %	64	30,5 %
teilweise berufstätig	6	2,9 %	6	2,9 %
in Hochschulausbildung (Student)	125	59,5 %	126	60,0 %
in Berufsausbildung/Wehrpflicht/Zivildienst	5	2,4 %	6	2,9 %
nicht berufstätig	4	1,9 %	8	3,8 %
Sonstiges	9	4,3 %		
Gesamtergebnis	210	100,0 %	210	100,0 %

korrigierte Werte nach Auswertung der Angabe "Sonstiges"

Die erhobenen Daten zum höchsten Bildungsabschluss und zum beruflichen Status, siehe Tabelle 35, sind nicht unmittelbar mit den üblicherweise erhobenen Daten der Allgemeinbevölkerung zur beruflichen Bildung (wie in [293] oder [292]) in Relation zu setzen, da dort nach schulischem Abschluss, beruflichem Abschluss und Erwerbsstatus getrennt wird. Entsprechend wird auf eine umfassende detaillierte vergleichende Darstellung verzichtet, da mit Ausnahme der Akademiker nicht möglich. Der Anteil von Probanden, die sich noch in Ausbildung an einer (Fach-) Hochschule befinden, ist sehr hoch

¹⁹¹ Es wurden folgende Vereinfachungen angenommen: die Angabe „Polytechnische Oberschule“ (POS, 6,9%) wird als vergleichbarer Abschluss zusammen mit „Realschule“ (22,1%) als „Mittlere Reife/POS“ angegeben, die Punkte „Ohne allgemeinen Schulabschluss“ (3,8%) und „Ohne Angabe zur Art des Abschlusses“ (0,2%) werden zusammengefasst. Die Vergleichszahlen des Statistischen Bundesamtes beziehen sich auf die Gesamtbevölkerung ab 15 Jahren. Für die Teilnahme an der Probandenstudie wurde allerdings ein Mindestalter von 18 Jahren vorausgesetzt. Entsprechend ist eine Verzerrung bei der Angabe „noch in Schulausbildung“ generell nicht zu vermeiden.

(ca. 60 %) ¹⁹². Der Anteil von Probanden mit (Fach-) Hochschulabschluss (ca. 39,5 %) ist deutlich höher als in der Gesamtbevölkerung (14 %) (Vergleichszahlen nach [292] ¹⁹³ S. 28). Ähnlich wie schon bei der schulischen Bildung sind in jüngeren Altersgruppen der Bevölkerung höhere Bildungsabschlüsse jedoch häufiger (17,8 % in der Gruppe 25 bis unter 30 bis hin zu 22,7 % in der Gruppe 30 bis unter 35, nach [292] S. 28). Die festgestellten Abweichungen in Bildung und beruflichem Status zu den Vergleichszahlen der Gesamtbevölkerung erklären sich durch die Rekrutierung der Probanden hauptsächlich an der Universität und in ihrem Umfeld.

Die jungen akademischen Nutzer können als Vorreiter im Bereich der Nutzung mobiler Internet-Dienste gesehen werden [294]. Demnach sind diese Nutzer schon „von morgen“. Sie sind aktiver im mobilen Internet, verursachen deutlich mehr Datenvolumen als der heutige Durchschnittsnutzer und sind dabei besonders sensitiv hinsichtlich Preis- und Qualitätsmerkmalen [294]. Die vorliegende Zusammensetzung der Stichprobe ist unter diesen Gesichtspunkten also sogar ein Vorteil.

7.3.4 Nutzung mobiler Endgeräte

Die Probanden sollten Vorerfahrungen zu den eingesetzten mobilen Plattformen Android bzw. iOS aufweisen. Entsprechend dienten die Angaben der Vorbefragung zum privat genutzten Smartphone/Tablet zur Vorauswahl der Probanden und zur Testplatz-/Gerätezuordnung im Test. Dabei war primär die mobile Plattform ausschlaggebend. Unterschiede zwischen den Angaben zum privat genutzten Gerät in der Vorbefragung und den erfassten Daten in der Nachbefragung traten durch zwischenzeitliche Veränderungen auf, z. B. durch Gerätewechsel. Aus diesem Grund sind Abweichungen in der Zahl der Nennungen der Plattformen und der Zahl der Testläufe auf den einzelnen Plattformen möglich, jedoch als unkritisch anzusehen.

¹⁹² Ein schönes Beispiel für die Gefährlichkeit relativer Zahlen: Die 39,5% der Probanden mit akademischem Abschluss und die 60% der Probanden in Hochschulausbildung addieren sich nicht zu 99,5%, da zwei getrennte Erhebungen vorliegen, d. h. Doppelnennungen bzw. Überschneidungen sind möglich. Beispiel: vorhandener akademischer Abschluss als Bachelor und Fortsetzung der akademischen Ausbildung in einem Masterstudium.

¹⁹³ Angaben Deutschland 2013, Wert akkumuliert für Fachhochschulausbildung, Hochschulausbildung und Promotion

Bis auf eine Ausnahme (nur Tablet) nutzen alle Probanden ein Smartphone (209 von 210). Ein Tablet wird zusätzlich von ca. einem Viertel der Probanden genutzt. Von 54 Besitzern eines Tablets nutzen 53 Probanden ebenfalls ein Smartphone. Bei den Smartphones ist der Anteil von Android deutlich größer als der Anteil von iOS. Bei den Tablets sind die beiden Plattformen dicht zusammen, mit etwas größerer Häufigkeit von iOS. Die genaue Aufschlüsselung der Ergebnisse ist Tabelle 36 zu entnehmen, Vergleichszahlen zum Markt in Deutschland liefert Tabelle 37. Durch die Festlegung auf die beiden führenden mobilen Plattformen (Android und iOS) werden die jeweiligen Anteile im Test (ca. 2,5:1) im Vergleich zum Markt (ca. 3,1:1) zwar leicht verzerrt, bei der Festlegung der Testplätze/Endgeräte im Test aber angenähert (3:1). Entsprechend der Vielzahl unterschiedlicher Hersteller und Modelle im Markt, vor allem bei Android-Geräten, sind auch die Angaben der Probanden zu den Geräten sehr heterogen. Auf eine Differenzierung über die jeweilige Plattform hinaus wird deshalb verzichtet.

Tabelle 36: Absolute Häufigkeit der von den Probanden privat genutzten Geräte nach Plattform und Formfaktor

Status	Smartphone		Tablet	
	Anzahl	Anteil (Probanden)	Anzahl	Anteil (Probanden)
nicht vorhanden	1	0,5 %	156	74,3 %
vorhanden	209	99,5 %	54	25,7 %
gesamt	210	100,0 %	210	100,0 %
Plattform	Anzahl	Anteil (Smartphones)	Anzahl	Anteil (Tablets)
Android	146	69,9 %	23	42,6 %
iOS	59	28,2 %	29	53,7 %
sonstige	4	1,9 %	2	3,7 %
gesamt	209	100,0 %	54	100,0 %

Doppelnennungen möglich

Tabelle 37: Anteile der mobilen Plattformen der Probanden im Vergleich zum Markt in Deutschland

Plattform	Probanden	Markt ¹⁹⁴ D 2013/06
Android	69,9 %	61,2 %
iOS	28,2 %	19,5 %
sonstige	1,9 %	19,3 %
Gesamtergebnis	100 %	100 %

Begrenzung der Darstellung auf Smartphone-Nutzung, n=209

¹⁹⁴ Anteil der Smartphone-Nutzung nach mobilen Betriebssystemen, Umfrage in Deutschland im Juni 2013, siehe [131], andere Angaben als Android und iOS sowie der verbleibende Rest zu 100% unter „sonstige“ zusammengefasst

Tabelle 38: Vergleichsdaten zur Nutzung mobiler Endgeräte in Deutschland

Mobiltelefonbesitzer ¹⁹⁵ in Deutschland		Smartphone-Besitzer in Deutschland	
Altersgruppe	Anteil der Altersgruppe	Altersgruppe	Anteil der Altersgruppe
U30 (14-29)	97,5 %	U30 (14-29)	72 %
Ü30 (30-49)	96,7 %	Ü30 (30-39)	66 %
Ü50 (50-64)	96,5 %	Ü40 (40-49)	46 %
Ü65 (65 und älter)	68,0 %	Ü60 (60 und älter)	10 %

Vergleichszahlen: Umfrage 2013, siehe [22] S. 11, Smartphone: keine Angabe zur Altersgruppe 50 bis 59

Als Vergleichszahlen in der Gesamtbevölkerung kann auch die Verbreitung mobiler Endgeräte herangezogen werden, um die Relevanz der Smartphone-Nutzung zu untermauern, siehe Tabelle 38. Der Anteil der Besitzer eines Smartphones ist in den letzten Jahren in Deutschland kontinuierlich gestiegen, siehe [22] S. 11. Besonders weit sind Smartphones in der Altersgruppe der unter 30-Jährigen verbreitet, was der Struktur der Stichprobe in diesem Aspekt sehr entgegenkommt.

7.3.5 Nutzung des mobilen Internets

Die Vorerfahrung in der Nutzung des mobilen Internets (ggf. auch im Netz eines bestimmten Anbieters) ist ein Attribut, das sich in den QoE-Bewertungen als subjektiver Faktor auf der Seite der Erwartung auswirken kann. Die Probanden sollten daher Vorerfahrungen in der Nutzung des mobilen Internets mitbringen. Diese Angaben der Vorbefragung dienen mit zur Auswahl der Probanden. Unterschiede zwischen den Angaben in der Vorbefragung und den dargestellten Daten der Nachbefragung waren durch zwischenzeitliche Änderungen möglich. Unerwartete Abweichungen wurden entsprechend abgeglichen.

Die Probanden nutzen das mobile Internet überwiegend auf dem Smartphone, entsprechend der Quote zum Smartphone-Besitz unter den Probanden, siehe Tabelle 39. Nur gut ein Viertel der Probanden, die ein Tablet besitzen, hat für dieses einen gesonderten Zugang zum mobilen Internet mit einem zusätzlichen Vertrag. Eine andere mobile Tablet-Nutzung, z. B. in Kombination mit dem Smartphone (WLAN-Tethering), ist nicht auszuschließen, wurde aber auch nicht abgefragt.

¹⁹⁵ Bitkom-Umfrage August 2013, siehe [295], beinhaltet auch Smartphones

Tabelle 39: Nutzung des mobilen Internets nach Formfaktor der Endgeräte

Nutzung des mobilen Internets	Smartphone-Nutzer		Tablet-Nutzer	
	Anzahl	Anteil	Anzahl	Anteil
ja	208	99,5%	13	24,1%
nein	1	0,5%	41	75,9%
gesamt	209	100,0%	54	100,0%

Tabelle 40: Von den Probanden genutzte Mobilfunknetze im Vergleich zu den Marktzahlen

Mobilfunknetz	Probanden im Test						deutsche SIM-Karten 2013/Q2 (BNetzA)	
	Smartphone-Nutzer		Tablet-Nutzer		zusammengefasst		Anzahl	Anteil
	Anzahl	Anteil	Anzahl	Anteil	Anzahl	Anteil		
E-Plus	39	18,7%	2	15,4%	38	18,1%	24.426.000	21,5%
O2	56	26,8%	1	7,7%	55	26,2%	19.411.000	17,1%
T-Mobile	35	16,7%	3	23,1%	34	16,2%	37.492.000	33,0%
Vodafone	77	36,8%	7	53,8%	77	36,7%	32.242.000	28,4%
unbestimmt	2	1,0%	(4)*		2 + 4	2,9%		
gesamt	209	100,0%	13	100,0%	210	100,0%	113.571.000	100,0%

Vergleichszahlen der Bundesnetzagentur, 2. Quartal 2013, siehe [296], gemeldete SIM-Karten ohne Angaben zur tatsächlichen Nutzung

Die Daten zum genutzten Mobilfunknetz wurden getrennt nach Smartphone und Tablet erhoben, siehe Tabelle 40. In Fällen mit abweichenden Angaben wurden die Angaben zum genutzten Netz zusammen mit nicht eindeutig zu bestimmenden Angaben („weiß nicht“ und Service-Provider mit Tarifen in mehreren Mobilnetzen) zu „unbestimmt“ zusammengefasst. Für die weiteren Betrachtungen ist dies unkritisch, da sich daraus keine nennenswerten Verschiebungen ergeben. Es waren Kunden von Anbietern in allen deutschen Mobilfunknetzen im Test vertreten, allerdings ungleichmäßig bzw. auch nicht den Verhältnissen im deutschen Mobilfunkmarkt entsprechend, siehe Tabelle 40. Die Vergleichszahlen zum Markt variieren allerdings auch stark zwischen verschiedenen Erhebungen, vgl. z. B. [296] als Anzahl SIM-Karten nach Geschäftsberichten und [297] aus einer Umfrage zur Nutzung auf dem primären Endgerät.

Die Probanden mit mobiler Internetnutzung auf Smartphone und Tablet weisen bei gemeinsamer Betrachtung alle eine Erfahrung von „mehr als 12 Monaten“ auf. Für weitere Darstellung der Netzerfahrung wird vereinfacht von der jeweils größeren Erfahrung ausgegangen, siehe Tabelle 41.

Tabelle 41: Erfahrung der Probanden in der Nutzung des mobilen Internets (nach Formfaktor)

Erfahrung im mobilen Internet	Formfaktor, Rohergebnisse		aggregiert nach größerem Erfahrungswert	
	Smartphone (absolut)	Tablet (absolut)	Smartphone oder Tablet (absolut)	Smartphone oder Tablet (relativ)
bis zu 3 Monate	21	1	21	10,0 %
bis zu 6 Monate	30	3	30	14,3 %
bis zu 12 Monate	30	5	30	14,3 %
mehr als 12 Monate	128	4	129	61,4 %
Gesamtergebnis	209	13	210	100,0 %

7.4 Nutzung ausgewählter mobiler Dienste durch die Probanden

Auf die Darstellung weiterer Angaben der Tablet-Besitzer, die die Tablets auch tatsächlich für das mobile Internet nutzen, wird verzichtet, da die statistische Masse mit $n=13$ sehr klein ist. In der Nachbefragung der Probanden wurden die Angaben zur Nutzungsintensität von 15 ausgewählten Diensten im mobilen Internet erfasst, siehe nachfolgende Darstellungen und Abfrage dazu im Anhang¹⁹⁶. Die erfassten Angaben dienen zur Charakterisierung der Probanden, geben aber auch ein recht gutes Bild über das allgemeine Interesse an bestimmten mobilen Diensten wieder. Mit Blick auf die ausgewählten Anwendungsszenarien im Test ist natürlich interessant, ob diese tatsächlich für die Probanden eine gewisse Relevanz hatten. Die mobilen Dienste im Test sollten eine breite Nutzung im Markt aufweisen. Es sollte daher eine gute Passung zwischen den realen Nutzungen der Probanden und den Schwerpunkten des Labortests bestehen. Vereinfacht ausgedrückt: Die „richtigen“ Dienste sollten mit den „richtigen“ Probanden getestet worden sein.

7.4.1 Einführung zur Auswertemethodik

Die Probanden konnten auf einer Skala auswählen, wie häufig sie einen bestimmten Dienst nutzen, Auswahloptionen siehe tabellarische Darstellung, 7-stufig in Testrunde 1, 5-stufig in Testrunde 2. Für die weitere Auswertung wurden den Skalenpunkten der Nutzungshäufigkeit die numerischen Werte (Scores) von 0 bis 6 zugeordnet, siehe Tabelle 42.

¹⁹⁶ Wortlaut und Reihenfolge aus dem Fragebogen der Testrunde 2: erarbeitet und abgestimmt mit den beteiligten Partnern, in Testrunde 1 vergleichbar gestaltet

Tabelle 42: Werttransformation der Häufigkeitsskala zur Auswertung

Wortlaut der Skala	Score (x)	Exponentialreihe 2^x
gar nicht	0	1
seltener als einmal im Monat	1	2
(1 bis 2x pro Monat)	2	4
mehrmals pro Monat	3	8
mehrmals pro Woche	4	16
(täglich)	5	32
mehrmals täglich	6	64

Skala mit Scores, 7-Stufen in Testrunde 1, auf 5-stufiger Skala in Testrunde 2 fehlende Werte in Klammern angegeben

Interpretation 1: Die Festlegung der Werte erfolgt als Ordnungskennziffer¹⁹⁷, ist als willkürlich durch den Autor definiert anzusehen und zu den Abständen der Werte wird explizit nichts unterstellt, nur zur Abfolge. Diese Interpretation ist statistisch unkritisch.

Erweiterung der Interpretation 1: Die Scores werden auch zur Mittelwertbildung genutzt, aber lediglich als Trend bzw. Tendenz zum internen Vergleich über den Median hinaus. Eine anderweitige Interpretation des Mittelwertes oder semantische Rücktransformation ist dafür nicht notwendig.

Interpretation 2: Die Festlegung der Werte erfolgt, um die Nutzungshäufigkeit entsprechend der verbalen Beschreibung quantitativ auswerten zu können (kardinales Skalenniveau). Eine solche Interpretation ist im Allgemeinen statistisch als kritisch zu bewerten, siehe z. B. [298] S. 12 ff, für die gewählte Auswertung aber anwendbar, da Aussagen zu den Abständen bzw. sogar zu den Verhältnissen zwischen den Skalenpunkten abgeleitet werden können.

Die Wortlaute der Skala quantifizieren die Nutzung auf konkrete (Mindest- bzw. Maximal-) Werte, allerdings bei unterschiedlicher Bezugsgröße. Die Transformation auf eine gemeinsame Bezugsgröße ist daher sinnvoll. Es sind eigentlich sogar zwei Aussagen in den Antworten der Probanden enthalten: ob ein Dienst überhaupt genutzt wird, und falls er genutzt wird, wie intensiv er genutzt wird. Um diese beiden Aussagen sinnvoll zusammen in einer Kennziffer ausdrücken zu können, ist es günstig, den Wert „gar nicht“ in der transformierten Darstellung auf 0 festzusetzen. Die Häufigkeit einer tatsächlichen

¹⁹⁷ höhere Zahl = häufigere Nutzung

Nutzung ist im Sinne einer absoluten Zahl der Nutzungen in einem bestimmten Zeitrahmen zu sehen. Wählt man als gemeinsame Bezugsgröße einen Monat, lässt sich eine ungefähr passende Verhältnisreihe konstruieren. Es ist günstig diese Reihe mit gleichem Faktor zwischen den Stufen zu definieren – als Exponentialreihe. Zur Erleichterung der weiteren Rechenschritte (z. B. Mittelwertbildung) sollten die Exponenten der Verhältnisstufen als Scores dienen (logarithmische Transformation). Eine mögliche und sinnvolle Basis der Exponentialreihe wäre 2, was bedeutet, dass jede höhere Stufe eine doppelt so hohe Häufigkeit der Nutzung wie die vorherige ergibt (je nach Lesart als näherungsweise, mindestens oder höchstens einzusetzen). Ausgehend von diesen Überlegungen ist eine 7-stufige Skala zur Häufigkeitsbestimmung gut geeignet. Für die vereinfachte 5-stufige Skala werden die Scores der Stufen mit gleichem Wortlaut übernommen, es fehlen aber die Scores 2 und 5. Dies ist nicht ganz ideal, aber durchaus noch brauchbar.

Für den Anfang der Reihe (größer 0) ist die Approximation weniger genau als für das Ende. Für die Exponential-Scores 5 und 6 wären als äquivalente Aussagen zum Wortlaut möglich „ $2^5=32$ ist ungefähr täglich bei 30 Tagen pro Monat“, bzw. „ $2^6=64$ ist ungefähr mindestens zweimal (also mehrfach) täglich bei 30 Tagen pro Monat“. Die bedeutsamen großen Werte der häufigen Nutzung sind also korrekt dargestellt, für die kleinen Werte der seltenen Nutzung (1 und 2) ist die Verzerrung vertretbar. Dies kommt ggf. sogar einer Schätzung (bzw. Selbsteinschätzung) der Probanden entgegen.

7.4.2 Aussagen zu den Diensten

Für die Dienste werden die Verteilungen über alle Probanden ermittelt und durch Maßzahlen beschrieben: häufigste Antwort (Modalwert), Median und Mittelwert. Es wird eine Rangfolge auf Basis des Mittelwerts (absteigend) gebildet.

Die wertäquivalente Interpretation ist eine Möglichkeit zur Bildung einer Rangfolge, anhand der Popularität der Dienste, mit einer Kennziffer, dem (arithmetischen) Mittelwert auf Basis der kardinalen Scores. Dem Autor erscheint eine solche Kennziffer zur Charakterisierung der Bedeutung der Dienste als Popularität aus Nutzung und durchschnittlicher Häufigkeit zusam-

men mit Modalwert und Median sinnvoll. Andere Angaben ähnlicher Befragungsergebnisse, wie sie typisch für Umfragen etc. sind, besitzen oft weniger Aussagekraft, z. B. „x % der Befragten nutzen Dienst XY“ (ohne Angabe einer Häufigkeit) oder „50 % der Befragten nutzen Dienst YX täglich / oft / ...“ (ohne Aussage zu den restlichen 50 %).

Die Probanden¹⁹⁸ konnten auf einer Häufigkeitsskala mit 5 Stufen auswählen. Besonders interessant sind die Muster in den Verteilungen der Anzahl Nennungen je Dienst und Häufigkeit. Dabei zeigt sich, dass für die abgefragten Dienste je nach Popularität sehr unterschiedliche Nutzungsmuster erkennbar sind. Modalwert und Median verschieben sich von „mehrmals täglich“ bei den populären Diensten zu „gar nicht“ am anderen Ende der Rangliste.

Tabelle 43: Nutzungshäufigkeit ausgewählter mobiler Dienste in Testrunde 2

Smartphone

Testrunde 2, Sommer 2013
n=208, jede Zeile addiert sich zu 100%

Rang nach Mittelwert absteigend sortiert
populäre Dienste oben

Rang	Dienst	Score	Nutzungshäufigkeit					Verteilung	Modalwert	Median	Mittelwert
			0	1	3	4	6				
			relative Häufigkeit					0 1 3 4 6			
1	E-Mail	3,4%	4,8%	7,7%	18,8%	65,4%	■■■■■	6	6	5,0	
2	Messaging	12,0%	1,9%	3,8%	9,6%	72,6%	■■■■■	6	6	4,9	
3	Soziale Netze	12,0%	4,3%	6,7%	15,9%	61,1%	■■■■■	6	6	4,5	
4	Informationssuche	1,0%	5,3%	18,3%	46,6%	28,8%	■■■■■	4	4	4,2	
5	Nachrichten lesen	11,1%	13,0%	20,2%	26,4%	29,3%	■■■■■	6	4	3,6	
6	App-Download/Update	7,2%	14,9%	30,3%	37,0%	10,6%	■■■■■	4	3	3,2	
7	Location Based Services	1,9%	13,9%	51,0%	26,9%	6,3%	■■■■■	3	3	3,1	
8	Videostreaming	20,2%	18,8%	32,2%	23,6%	5,3%	■■■■■	3	3	2,4	
9	Fotos teilen	39,4%	19,2%	20,7%	14,9%	5,8%	■■■■■	0	1	1,8	
10	Online-Spiele	45,2%	19,7%	9,1%	13,9%	12,0%	■■■■■	0	1	1,8	
11	Musikstreaming	42,3%	19,7%	19,7%	13,5%	4,8%	■■■■■	0	1	1,6	
12	Online-Shopping	38,5%	25,0%	22,1%	12,5%	1,9%	■■■■■	0	1	1,5	
13	Cloud-Dokumente	52,4%	15,4%	16,3%	13,5%	2,4%	■■■■■	0	0	1,3	
14	Online-Banking	59,6%	10,1%	13,5%	14,4%	2,4%	■■■■■	0	0	1,2	
15	Internettelefonie	54,8%	22,6%	14,4%	5,8%	2,4%	■■■■■	0	0	1,0	

¹⁹⁸ Für einen der 209 Probanden mit Smartphone liegen keine Angaben zu den genutzten Diensten vor, womit sich n=208 für die Smartphone-Nutzung ergibt.

In der Darstellung als Spark Bars (Mini-Säulendiagramm pro Dienst) sind diese Muster in der Verteilung leicht voneinander zu trennen, siehe Tabelle 43.

Smartphones sind Kommunikationsmaschinen, entsprechend führen solche Dienste die Spitzengruppe der sehr populären Dienste an, die fast alle Probanden (sehr) häufig nutzen. Zu diesen gehört neben E-Mail und Sofortnachrichten auch der Austausch in Sozialen Netzen. Für die beiden Letztgenannten findet sich noch ein Gegenpol, der die Probanden in Intensiv- und Nichtnutzer unterscheidet. Der Spitzengruppe folgen Web-Browsing-Dienste, zu denen die allgemeine Informationssuche und das Lesen aktueller Nachrichtenseiten wie Spiegel zählen. Im Mittelfeld landen Dienste, die zwar quasi alle Probanden nutzen, überwiegend aber nur selten bis gelegentlich. Dazu gehören App-Downloads, Location Based Services (LBS) und Videostreaming, deren Nutzung überwiegend gelegentlich stattfindet. App-Downloads und die Nutzung von LBS, wie Navigations- oder Kartendienste, können als Dienste bewertet werden, die je nach Bedarf genutzt werden. Videostreaming als das populärste Online-Entertainment auf dem Smartphone wird vermutlich tatsächlich bei Gelegenheit genutzt. In der unteren Tabellenhälfte landen die unpopulären Dienste, die überwiegend nur selten oder gar nicht genutzt werden. In der Gruppe der Schlusslichter finden sich neben der Cloud-Nutzung weitere solche Dienste, die von mehr als der Hälfte der Probanden überhaupt nicht genutzt werden.

Ein Vergleich der Popularität der Dienste unter den Probanden zwischen den Testrunden 1 und 2 ist interessant, da so Veränderungen über die Zeit (ca. 1 Jahr) bzw. den verschiedenen Stichproben zu beobachten sind. Ein direkter Vergleich der gebildeten Mittelwerte ist durch die unterschiedliche Anzahl Stufen (7 bzw. 5) der Skalen aber nicht sinnvoll. Der Vergleich zwischen den Testrunden bezieht sich deshalb ausschließlich auf die festgestellten Veränderungen in der Platzierung relativ zu den Rängen und die Verteilung innerhalb einer Testrunde, siehe Tabelle 44.

Es sind dabei kleinere Veränderungen zwischen der Erhebung der Testrunde 1 im Herbst 2012 und der Testrunde 2 im Sommer 2013 festzustellen, die sich allerdings ausschließlich innerhalb der jeweiligen Tabellenhälfte auswirken. Zuvor sehr populäre Dienste werden nicht plötzlich sehr unpopulär und umgedreht, sondern es sind mit Ausnahmen eher Platzwechsel zu verzeichnen. Die im Vergleich zur Testrunde 1 noch etwas jüngere Altersstruktur der Testrunde 2 zeigt sich mit den der Generation Y zugeordneten Merkmalen auch in der Nutzungshäufigkeit der mobilen Dienste. Die Probanden sind noch etwas kommunikationsfreudiger, stärker an Entertainment interessiert und vor allem verspielter (+4 Plätze für Online Gaming). Ernsthaftere Anwendungen wie Nachrichten lesen, Cloud-Nutzung und auch Online-Shopping mussten hingegen etwas nachgeben. Internettelefonie ist auf den letzten Platz gefallen. Eine mögliche Erklärung wäre die Zunahme von (preiswerten All-Net-) Sprach-Flatrates im Mobilfunk, so dass die möglichen Kostenvorteile der Internettelefonie wegfallen und diese Dienste die Nutzer in der Mehrzahl nicht ansprechen.

Tabelle 44: Veränderung der Nutzungsintensität ausgewählter mobiler Dienste zwischen 2012 und 2013

Erhebung 2013		Veränderung		Verteilung		Kommentar zur Veränderung	Testrunde
Rang	Dienst	Trend	Plätze	2013	2012		
1	E-Mail	→	± 0	■■■■■	■■■■■	stabile Spitzenposition	
2	Messaging	↑	+ 3	■■■■■	■■■■■	Platzwechsel innerhalb der oberen Tabellenhälfte	1
3	Soziale Netze	↗	+ 1	■■■■■	■■■■■		1
4	Informationssuche	↘	- 1	■■■■■	■■■■■		1+2
5	Nachrichten lesen	↓	- 3	■■■■■	■■■■■		
6	App-Download/Update	→	± 0	■■■■■	■■■■■	stabiles Mittelfeld	
7	Location Based Services	→	± 0	■■■■■	■■■■■		1+2
8	Videostreaming	→	± 0	■■■■■	■■■■■		1+2
9	Fotos teilen	→	± 0	■■■■■	■■■■■	Platzwechsel innerhalb der unteren Tabellenhälfte	
10	Online-Spiele	↑	+ 4	■■■■■	■■■■■		2
11	Musikstreaming	↑	+ 2	■■■■■	■■■■■		1
12	Online-Shopping	↓	- 2	■■■■■	■■■■■		2
13	Cloud-Dokumente	↓	- 2	■■■■■	■■■■■		
14	Online-Banking	↗	+ 1	■■■■■	■■■■■		
15	Internettelefonie	↓	- 3	■■■■■	■■■■■		

2013: 5 Stufen, n=208; 2012: 7 Stufen, n=60

7.4.3 Aussagen zu den Probanden

Nachdem zuvor für die Popularität der einzelnen Dienste Aussagen über alle Probanden hinweg getroffen wurden, sollen nachfolgend die Probanden über die erhobenen Daten zur Dienstenutzung charakterisiert werden. Da die Datenbasis zur Nutzung der Dienste auf Tablets sehr klein ist, bleiben die nachfolgenden Ausführungen auf die Smartphone-Nutzer beschränkt.

Die Aussagen zur Charakterisierung der Probanden hinsichtlich eines Nutzertyps können nur so gut sein, wie die Erfassung der Nutzungsintensität der abgefragten Dienste. Die Zusammenstellung der mobilen Dienste in der Nachbefragung erhebt keinen Anspruch auf Vollständigkeit. Es bleiben sicherlich weitere populäre wie auch unpopuläre Dienste unberücksichtigt. Nach den Vorüberlegungen ist aber davon auszugehen, dass die Auswahl weitgehend relevante Dienste umfasst. Daher sollten brauchbare Aussagen zu den Nutzertypen und zur allgemeinen Zusammensetzung der Stichprobe hinsichtlich der Nutzungsintensität trotzdem möglich sein. Die Erläuterungen zur Transformation in Scores etc. sind dem vorangestellten allgemeinen Abschnitt zur Auswertemethodik zu entnehmen. Für die Probanden werden die Verteilungen über die 15 ausgewählten Dienste ermittelt und durch Maßzahlen beschrieben: häufigste Antwort (Modalwert), Median, Mittelwert und summierter Wert.

Tabelle 45: Verteilung der Modalwerte und Mediane der Nutzungshäufigkeit mobiler Dienste zwischen den Probanden

Häufigkeit der Nutzung		Probanden mit jeweiligem Modalwert			Probanden mit jeweiligem Median		
		Anzahl	Anteil	kumulativ	Anzahl	Anteil	kumulativ
0	gar nicht	79	38,0%	38,0%	24	11,5%	11,5%
1	seltener als einmal im Monat	12	5,8%	43,8%	42	20,2%	31,7%
3	mehrmals pro Monat	39	18,8%	62,5%	92	44,2%	76,0%
4	mehrmals pro Woche	31	14,9%	77,4%	46	22,1%	98,1%
6	mehrmals täglich	47	22,6%	100,0%	4	1,9%	100,0%

Modalwert und Median der Nutzungshäufigkeit aller 15 abgefragten Dienste je Proband, nur Smartphone-Nutzer, n=208

So wie die abgefragten Dienste stark unterschiedlich genutzt werden, so verschieden sind auch die Probanden in ihrer Nutzung der Dienste, siehe Tabelle 45. Die Nutzungshäufigkeit der populären Dienste charakterisiert die breite Masse der Probanden, die Anzahl der genutzten Dienste die Extremwerte:

- Probanden mit einem niedrigen Wert: Ein Selektivnutzer mit insgesamt wenig Nutzung mobiler Dienste nutzt eine große Anzahl Dienste „gar nicht“ und nur eine kleine Anzahl Dienste selten oder gelegentlich.
- Probanden mit einem hohen Wert: Ein Power-User nutzt sehr viele Dienste mehrmals pro Woche oder sogar mehrfach täglich.

Repräsentative Vergleichszahlen zu Nutzertypen sind schwierig zu ermitteln, da unterschiedliche Erhebungsmethoden und Klassifizierungen genutzt werden. Objektive Zahlen ergeben sich aber aus den Analytics-Funktionen¹⁹⁹, die in vielen mobilen Apps integriert sind. Demnach zeigt sich, allerdings auf Basis einer anderen Einteilung der Nutzertypen und Metriken, im Vergleich der Zeiträume 2013/Q1 zu 2014/Q1 eine Intensivierung der Nutzung mobiler Apps. Dabei sind ein überproportionaler Zuwachs der „Super User“ (6 bis 60 App-Nutzungen pro Tag) und ein extrem starker Zuwachs der „Mobil-Abhängigen“ (über 60 App-Nutzungen pro Tag) festzustellen, siehe [14]. Eine weitere alternative Einteilung in fünf verschiedene Nutzergruppen, je nach Aktivitätsprofil vom „Info Seeker“ über den „Social Monitor“ bis hin zum „Digital Mover & Shaker“, hält [280] S. 7f bereit. Aus den ermittelten Ergebnissen können dazu aber keine direkten Vergleichszahlen geliefert werden.

Mit der logarithmisch abgestuften Score-Skala ergibt sich für die Mittelwerte der Nutzungshäufigkeit eine normalverteilungssähnliche Charakteristik, die durch Glockenform der Dichte bzw. S-förmige Verteilungsfunktion gekennzeichnet ist, siehe Abbildung 64. Diese Verteilung ist ungefähr symmetrisch. Ausgehend von der Verteilung ist eine Bildung von Nutzertyp-Gruppen nach Nutzungsintensität möglich (Tabelle 46).

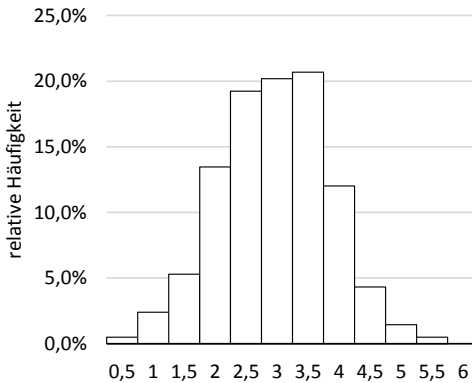
¹⁹⁹ z. B. auf Basis von Flurry Analytics, einem Analyse-Framework zur Integration in mobile Apps, siehe <http://www.flurry.com/solutions/analytics> abgerufen am 24.04.2014

Tabelle 46: Nutzertypen nach Nutzungsintensität und jeweilige Anteile der Probanden

Nutzertyp	Wertebereich	Anteil Probanden im Test
Selektivnutzer	$x < M - 2 \cdot SD$	3 %
Wenignutzer	$M - 2 \cdot SD \leq x < M + SD$	15 %
Durchschnittsnutzer	$M - SD \leq x \leq M + SD$	67 %
Vielnutzer	$M + SD < x \leq M + 2 \cdot SD$	13 %
Intensivnutzer (Power-User)	$x > M + 2 \cdot SD$	2 %

nur Smartphone-Nutzer, Mittelwert $M=2,7$, Standardabweichung $SD=0,9$, $n=208$

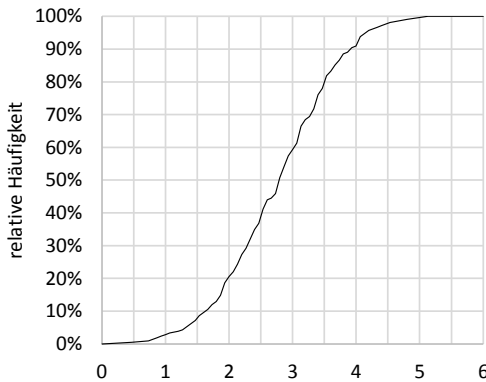
Histogramm



Mittelwert der Dienstenutzung je Proband,
Obergrenze Klasse, Klassenbreite=0,5

a)

empirische Verteilungsfunktion



b)

Mittelwert der Dienstenutzung je Proband

Abbildung 64: Dichte und Verteilung der Mittelwerte der Häufigkeit der Nutzung mobiler Dienste je Proband Mittelwert der Nutzungshäufigkeit aller 15 abgefragten Dienste je Proband, nur Smartphone-Nutzer, $n=208$

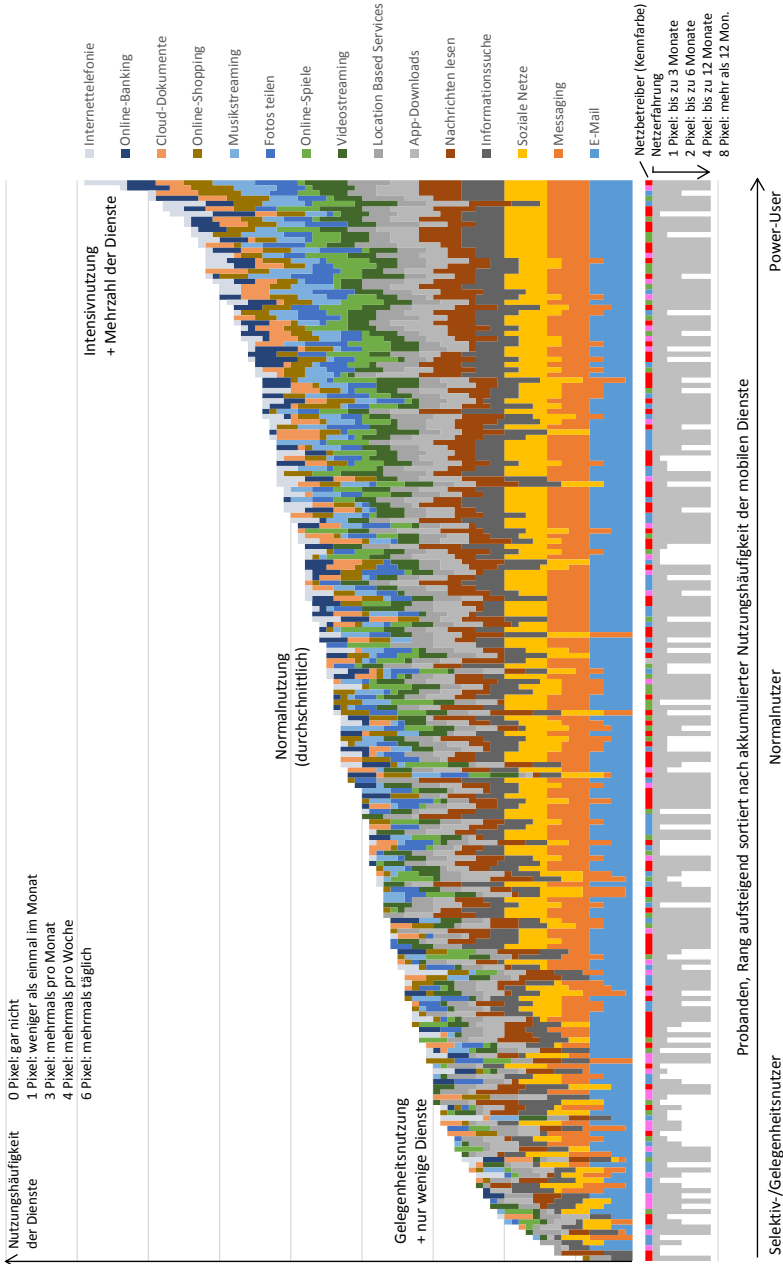


Abbildung 65: Summierte Nutzungshäufigkeit mobiler Dienste je Proband mit weiteren Attributen nur Smartphone-Nutzer, n=208

Zusätzlich zum Mittelwert wurde je Proband die Summe der Nutzungshäufigkeit über alle 15 Dienste gebildet. Es ergibt sich die erwartete Kurvenform der Rangverteilung mit annähernd gleichmäßigem Anstieg im Mittelteil und stark abknickenden Extremwerten an den Enden, siehe Abbildung 65. Diese aggregierte Darstellung ist gut geeignet zur abschließenden Charakterisierung der Probanden und der abgefragten Dienste. Es wird leicht ersichtlich, dass es einige wenige wichtige Dienste sind, die eine breite Basis bilden und von der Mehrheit der Probanden intensiv genutzt werden. Die individuellen Unterschiede der Probanden entstehen eher durch die exotischen Dienste entsprechend der jeweiligen Interessen.

7.5 Weitere Erkenntnisse zu den Probanden

Durch die Auswertung der Bewertungen über alle Szenarien des Testparcours hinweg können weitere Aussagen zu den Probanden generiert werden. Dies sind eine Art Gesamtüberblick und ein Inter-Rater-Vergleich zu Nutzertypen hinsichtlich des Bewertungsverhaltens.

7.5.1 Gesamtüberblick Rohergebnisse

Der Gesamtüberblick zu den Rohergebnissen ist ab Abbildung 135 als Übersichtstafeln im Anhang A.5.1 zu finden, woraus schon einige Erkenntnisse abzuleiten sind. Die Bewertung „vollkommen unzufrieden“ ist insgesamt relativ selten. Wenn Probanden mit „vollkommen unzufrieden“ bewerten, dann in gut der Hälfte der Fälle gleich mehrfach, siehe Tabelle 47. Diese Probanden haben offensichtlich ein allgemein hohes Anspruchsniveau – zumindest im Vergleich zu den anderen Probanden. Sämtliche Probanden, die mit „vollkommen unzufrieden“ bewerten, entstammen der Generation Y. Die Bewertung „vollkommen unzufrieden“ resultiert z. T. aus einer unvollständigen Aufgabenbearbeitung, weil die Probanden die Aufgabe abgebrochen haben oder erst gar nicht starten konnten. Es gibt einerseits (einige) Probanden, die sind grundsätzlich insgesamt nicht „vollkommen zufrieden“, andererseits gibt es auch (einige) Probanden, die quasi mit fast allem „vollkommen zufrieden“ sind, siehe Tabelle 47.

Tabelle 47: Übersicht zu den Probanden mit Häufigkeit des Auftretens der Bewertungen

Gesamtzufriedenheit Bewertung	Anzahl Antworten mit dieser Bewertung (x-mal)												
	0	1	2	3	4	5	6	7	8	9	10	11	12
vollkommen unzufrieden	182	18	7	2	1	0	0	0	0	0	0	0	0
eher unzufrieden	104	57	31	11	6	0	1	0	0	0	0	0	0
weder zufrieden, noch unzufrieden	73	66	46	20	4	1	0	0	0	0	0	0	0
eher zufrieden	4	21	25	52	41	27	17	14	6	2	1	0	0
vollkommen zufrieden	4	6	9	17	30	25	32	21	28	19	10	7	2

n=210 Probanden mit 2520 Bewertungen (nur Gesamtzufriedenheit)

Lesebeispiele: 182 Probanden haben 0-mal (d. h. gar nicht) mit „vollkommen unzufrieden“ bewertet; 2 Probanden haben volle 12-mal (d. h. immer) mit „vollkommen zufrieden“ bewertet

7.5.2 Inter-Rater-Vergleich

In anderen QoE-Evaluationen zeigten sich bei der Untersuchung des QoS-QoE-Zusammenhangs mehrere Nutzertypen, die sich in ihrem Bewertungsverhalten unterscheiden. Die Nutzer lassen sich demnach entsprechend ihrer Erwartung in Kategorien einteilen. [299] S. 132 unterscheidet – auf Basis der Auswertung von QoE-Testergebnissen – in drei Nutzertypen nach Abhängigkeit ihrer Bewertungstendenz und der Bewertungsänderungen:

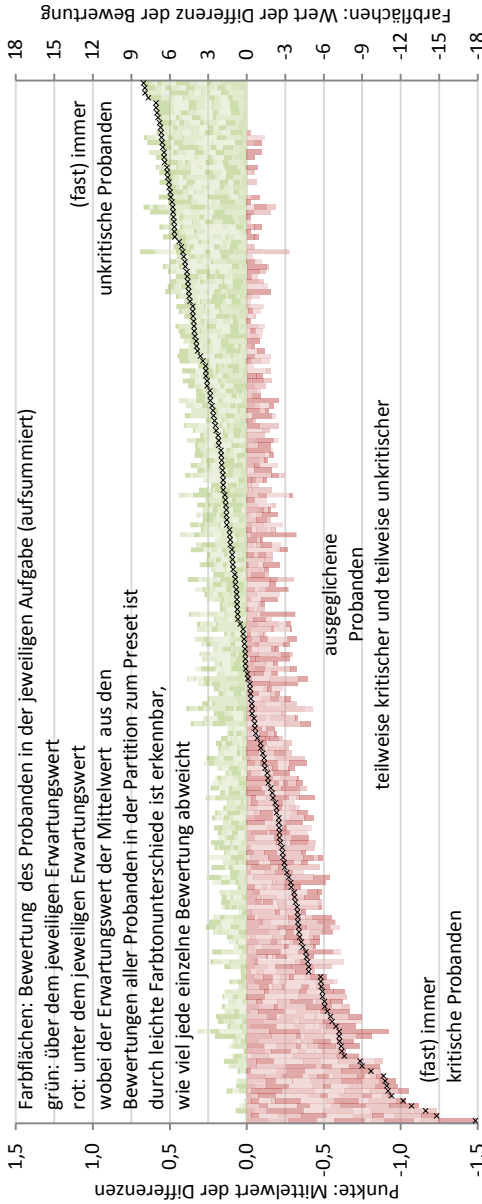
- hektische Nutzer: eher kritisch (hohe Erwartung), schnelle Änderung der Bewertung auch bei unveränderten Bedingungen
- normale Nutzer: bessere Bewertungen als die hektischen Nutzer (normale Erwartung), geringere Varianz der Bewertungsänderungen
- unempfindliche Nutzer: mehr oder minder immer zufrieden (niedrige Erwartung), kaum Veränderungen der immer guten Bewertungen

Entsprechend kann es lohnen, auch für die eigenen Probanden zu untersuchen, ob sich derartige Nutzertypgruppen erkennen lassen. Abweichend zur o. g. Quelle werden die eigenen Daten zweistufig ausgewertet. Im Gegensatz zu den dienstspezifischen bedingten Verteilungen zur Charakterisierung des QoS-QoE-Zusammenhangs interessiert an dieser Stelle eine Charakterisierung der Verteilung der Zufriedenheit zwischen den Probanden insgesamt. Dafür ist es sinnvoll, übergreifend für alle Dienste und Presets zu vergleichen. Eine

Aussage über einen Probanden wird durch den Vergleich zwischen den Probanden generiert. Daher ergibt sich auch die Bezeichnung Inter-Rater-Vergleich. Ergänzt wird dieser um einen Intra-Rater-Vergleich, in dem die Bewertungen der Probanden mit sich selbst verglichen werden. Die Diskussion dazu erfolgt getrennt je nach Dienst auf Basis der Diskussion in 8.5.

Als Vergleichsgruppe im Inter-Rater-Vergleich sollen die Probanden dienen, die unter als gleich angenommenen Bedingungen bewertet haben, d. h. je Aufgabe das gleiche Preset als Netzbedingung und gleiche Partition als technische Voraussetzung. Für jeden Probanden wurde für jede der 12 Aufgabenausführungen eine Differenz zwischen abgegebener Bewertung (als Score) und einem Vergleichswert (als Score) gebildet. Als Vergleichswert wird der arithmetische Mittelwert der Bewertungen der Vergleichsgruppe festgesetzt. Auf eine Unterscheidung valid vs. invalid für die Aufgabenausführungen wurde an dieser Stelle verzichtet, um die Auswertung nicht weiter zu verkomplizieren. Über insgesamt 12 Aufgabenausführungen hinweg ist unwahrscheinlich, dass diese insgesamt alle für einen einzelnen Probanden invalid sind. Die aus einzelnen invaliden Aufgabenausführungen begründete minimale Verzerrung ist für die beabsichtigte Erkenntnis bei dieser Auswertung hinzunehmen. Die so ermittelten 12 Differenzwerte je Testlauf eines Probanden können das Bewertungsverhalten beschreiben, siehe Abbildung 66.

Es sind sowohl (fast) immer kritische wie auch unkritische Probanden zu finden, wie auch insgesamt ungefähr ausgeglichene Probanden, die mal besser und mal schlechter bewerten als der vergleichbare Durchschnitt. Im Bereich der kritischen Probanden ist ein leichtes Übergewicht festzustellen. Dies bestätigt den Blick auf die Rohergebnisse hinsichtlich einiger besonders kritischer Probanden (siehe 7.5.1). Es lässt sich eine Art ansteigender Bewertungskorridor erkennen: von im Mittel der Differenzen knapp unter $-0,5$ bis knapp über $0,5$, also rund einer Bewertungsstufe. Nur die extrem kritischen Probanden (ca. 6 %, am linken unteren Ende) fallen aus diesem Bereich deutlich heraus. Diese Abweichungen könnten mit Problemen in der Aufgabenbearbeitung oder ungünstige Preset-Zusammenstellungen zusammenhängen.



Probanden, n=210
 Rang aufsteigend sortiert nach Mittelwert der Differenzen gebildet zum Erwartungswert der Partition

Erläuterungen zur Abbildung 66: Die x-Achse zeigt die Probanden in angegebener Sortierung, in y-Richtung werden die Bewertungsdifferenzen des Probanden zur jeweiligen Vergleichsgruppe gezeigt.
 Skala links als Mittelwert der Differenzen: generelle Bewertungstendenz und Verteilung Positiv- und Negativbewerter.
 Skala rechts als Wert der Differenz und akkumuliert: Bewertungstendenz und Probandengruppen sind erkennbar, von immer kritisch über ausgeglichen/schwankend zu immer unkritisch. Zusätzlich lässt sich ablesen, wie oft und wie viel jede Bewertung positiver oder negativer als der Vergleichswert ist.

Abbildung 66: Inter-Rater-Vergleich als Rangverteilung mit zugehörigen Bewertungstendenzen

7.6 Zusammenfassung

Es wurden die Besonderheiten des Faktors Mensch in der Rolle des Probanden zur Bestimmung des QoS-QoE-Zusammenhangs beschrieben. Aus dem Kontext einer Laborsituation ergeben sich möglicherweise Verschiebungen im Vergleich zur Nutzung im Feld, die aber nicht weiter aufgeklärt werden können. Es erfolgte die Beschreibung der üblichen allgemeinen Punkte zur Gewinnung der Stichprobe und zum Umgang mit den Probanden.

Ausgehend von den Erhebungen der Nachbefragung erfolgte die Beschreibung der Stichprobe im eigenen Nutzertest. Insgesamt gesehen lässt sich der Großteil der Stichprobe wie folgt charakterisieren:

- Altersstruktur: junge Erwachsene, überwiegend der Generation Y (ca. 90 %), mit mobilen Geräten und dem Internet sozialisiert
- Bildungsniveau: sehr gut gebildet (überdurchschnittlich im Vergleich zur Gesamtbevölkerung bzw. Alterskohorte)
- Geräteausstattung: technikaffin mit Nutzung eines Smartphones und/oder Tablets
- Nutzung des mobilen Internets: regelmäßige Nutzung und überwiegend erfahren
- Nutzung mobiler Dienste: kommunikationsfreudig und interessiert

Aufgrund der Struktur der Stichprobe sind weitere Unterscheidungen hinsichtlich der Merkmale der Stichprobe nicht sinnvoll. Eine Charakterisierung als studentische Stichprobe ist weitgehend zutreffend. Damit ist von Trendsettern in der Nutzung des mobilen Internets und einer insgesamt kritischen Nutzergruppe auszugehen, was als Vorteil für die Studie gesehen werden kann.

Die Analyse verschiedener Nutzertypen im Test zeigte annähernd log-normalverteilte Anteile bezogen auf die Nutzungsintensität mit einem großen Anteil weitgehend durchschnittlicher Nutzer und annähernd gleich großen Anteilen Wenignutzer und Intensivnutzer. Im Inter-Rater-Vergleich ergab sich eine annähernd ausgeglichene Verteilung zwischen kritischen und unkritischen Probanden, mit einem geringen Überschuss sehr kritischer Probanden

Die Nachbefragung zu den genutzten Diensten bestätigt weitgehend die Relevanz der ausgewählten Dienste in den einzelnen Testrunden, wobei sich einige charakteristische Strukturen zur Popularität und Nutzungsintensität der Dienste zeigten.

8 Die Auswertung

Bestandteile und Prinzipien

In diesem Kapitel werden die Herangehensweise zur Auswertung und ein Teil der im Test erzielten Ergebnisse vorgestellt. Dabei soll so viel Zusatzwissen vermittelt werden wie notwendig, um die generierten Daten sinnvoll und statistisch untermauert zu interpretieren. Testaufbau und Testdesign weisen einige Besonderheiten auf, deren Auswirkungen hinsichtlich der möglichen Aussagen zu berücksichtigen sind. Die Auswertung kann nur auf Basis der variierten und erfassten Daten erfolgen, daher sein an dieser Stelle noch einmal auf Kapitel 4 zu den berücksichtigten unabhängigen und abhängigen Variablen verwiesen.

Eine kurze Warnung vorweg: Neben einer Reihe von Ergebnissen und Aussagen, die sofort klar und selbsterklärend sind, gibt es auch solche Auswertungen, die helfen sollen, die Ergebnisse korrekt einzuordnen, den statistischen Puristen aber wohl zumindest ein leises Seufzen entlocken dürften („Skalen als Zahlen“, „normalverteilungähnlich“, ...).

Die Auswertung dient einerseits zur Dokumentation der erzielten Daten und andererseits zur Generierung der gesuchten qualitativen und quantitativen Aussagen durch die empirische Ermittlung im Test:

- Überprüfung und Weiterentwicklung des QoS-QoE-Modells
- Überprüfung der Zusammenhänge und der Nutzerrelevanz der QoS-veränderlichen Reizgrößen und der zugehörigen Metriken
- Abschätzungen zum Einfluss der Nebenfaktoren
- Bestimmung der Charakteristik der dienstspezifischen QoS-QoE-Zusammenhänge
- Bestimmung der gesuchten Schwellwerte

Zunächst wird die grundsätzliche Methodik der Auswertung diskutiert. Die Herangehensweise für die weitere Auswertung sieht eine Zerlegung in einzelne Bestandteile und Schritte vor, siehe Abbildung 67.

Planung und Vorbereitung

- QoS-QoE-Modell:** Ursache-Wirkungs-Zusammenhang, Mediatoren/Moderatoren
- Testdesign:** UV, AV, Zielsetzung, Festlegungen, Einschränkungen, Besonderheiten
- Testbed und Technik:** Netzemulation, Endgeräte, Aufzeichnung und Datenerfassung
- Testparcours (Dienste):** Vorüberlegungen und Vorerprobung, Berücksichtigung Nebenfaktoren

Ausführung

Testergebnisse: aus geplanten Experimenten zur Generierung der Datenbasis

Auswertung

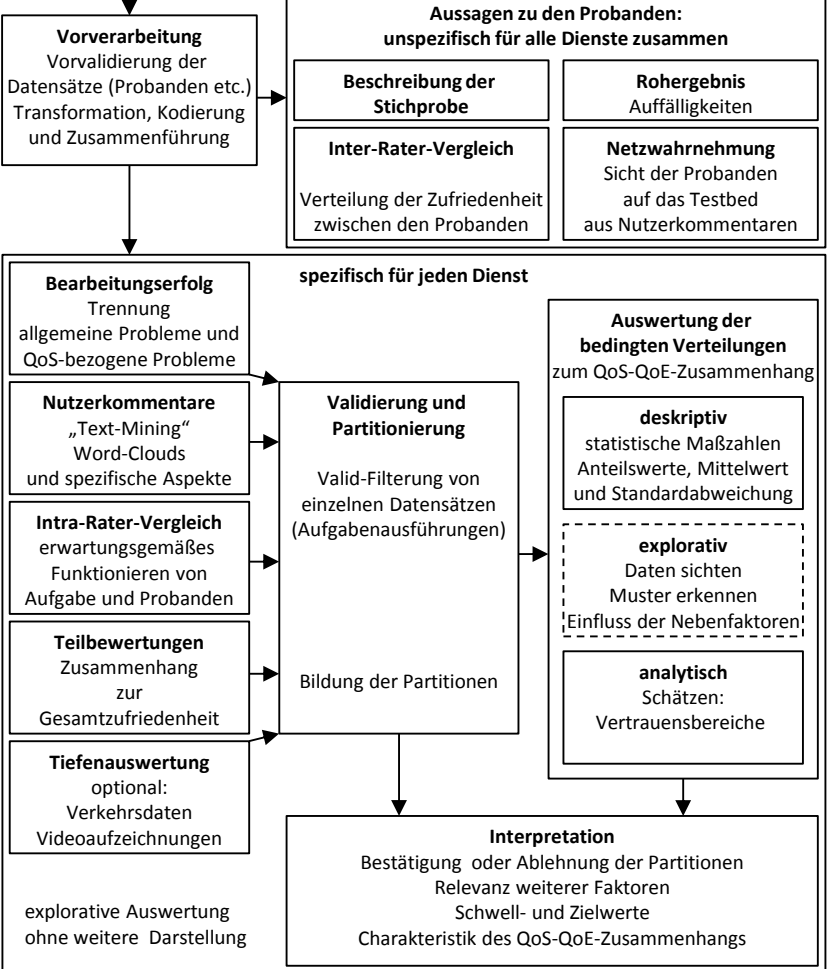


Abbildung 67: Übersicht zu den Bestandteilen und Schritten der Auswertung

Die Erläuterungen zu den einzelnen Teilen in diesem Kapitel sollen als eine Art Leseanleitung und Referenz dienen, warum welche Aspekte in der dienstspezifischen Auswertung enthalten sind. Zur Verdeutlichung wird jeweils ein Beispiel aus den Ergebnissen der untersuchten Szenarien gewählt. Zum Verständnis dieser Überlegungen sei auch auf die zugehörigen spezifischen Vorüberlegungen und Voruntersuchungen verwiesen. Die konkreten Ergebnisse der dienstspezifischen Auswertung des Nutzertests werden im nächsten Kapitel präsentiert, um dort je Dienst die Vorüberlegungen, die Ergebnisse der Voruntersuchungen und die empirischen Ergebnisse aus dem Nutzertest zusammenzubringen. Im vorherigen Kapitel erfolgte bereits die Auswertung zu den Probanden sowie unspezifisch für alle Dienste zusammen (ab 7.3).

Ergebnisse bedeuten nicht automatisch auch Erkenntnisse. Zum besseren Verständnis der Ergebnisse wurden für zahlreiche Bestandteile der Auswertung geeignete Visualisierungen gesucht oder speziell entwickelt. Damit diese richtig interpretiert werden können, sind jeweils Beispiele mit Erläuterungen eingefügt.

Am Ende des Kapitels wird eine Güte- und Fehlerabschätzung versucht. Dies schließt auch Betrachtungen zu Objektivität, Reliabilität und Validität zum eigenen Test mit ein. Die gesamten Überlegungen zur Auswertung insgesamt sind sinnvoll, um Klarheit zu schaffen, welche Erkenntnisse zu gewinnen sind und wie weit man den eigenen Daten trauen kann.

8.1 Methodik der Auswertung zum QoS-QoE-Zusammenhang

Um die empirisch ermittelten Ergebnisse erklären zu können, muss sichergestellt sein, dass der gesuchte Zusammenhang zwischen QoS und QoE tatsächlich in den erhobenen Daten abgebildet und in der Auswertung geeignet zu beschreiben ist. Der interessierende QoS-QoE-Zusammenhang kann entsprechend der Ursache-Wirkungs-Kette hergeleitet und abgebildet werden.

Prinzipiell lassen sich zwei verschiedene Arten der Auswertung unterscheiden, je nachdem worauf sich die interessierenden Bewertungen der Probanden beziehen:

- Die Bewertungen eines Probanden werden auf die tatsächlich erreichten Werte der identifizierten Nutzerstimuli (das „Signal“) zurückgeführt (Signal-basierte Auswertung).
- Die Bewertungen eines Probanden werden auf die variierten Netzparameter zurückgeführt (QoS-Parameter-basierte Auswertung).

Wie in der Planung und Realisierung der Studie ist auch in der Auswertung zu beachten: Aufwand und erwarteter Nutzen müssen in einem vernünftigen Verhältnis bleiben.

8.1.1 Tiefenauswertung

In einer Tiefenauswertung können die tatsächliche Netznutzung auf Basis der Verkehrsdatenaufzeichnungen bzw. die tatsächlich einwirkenden Nutzerstimuli auf Basis der Videoaufzeichnungen für jede einzelne Aufgabenausführung ausgewertet werden. Da dies extrem aufwendig ist, wird dies als optionaler Auswerteschritt vorgestellt, der nur bei Bedarf für einzelne Aspekte ausgewählter Dienste genutzt wird. Die vollständige Tiefenauswertung wäre Voraussetzung für eine Signal-basierte Auswertung. Da die Videoaufzeichnungen aufgrund ihres Charakters des Nachweises der Machbarkeit und Nützlichkeit nur auszugsweise erstellt wurden, wird nachfolgend nur noch auf die Tiefenauswertung auf Basis der Verkehrsdatenaufzeichnungen eingegangen.

Das Ziel der Tiefenauswertung definiert die Gesamtkomplexität, wobei weitere Stufen die vorherigen Stufen bedingen:

1. Validierung der parameterbasierten Aufgabenausführungen zur Filterung invalider Daten aus ungültigen Aufgabenausführungen:
 - Vorverarbeitung: Grobfilterung und Trennung der Verkehrsdaten je Testplatz nach Host (IP-Adresse)
 - zeitliche Segmentierung der Verkehrsdaten passend zu den einzelnen Aufgabenausführungen, dabei Orientierung an der Aufgabenabfolge und zeitlichen Zusatzinformationen (z. B. zur Unterscheidung der Ausführungen in erste und zweite Bearbeitung)
 - Identifikation relevanter Muster und Größen (Abschätzung bzw. Ermittlung der Größenordnung ausreichend)
2. Verfeinerung der Parameter-basierten Auswertung mit Istwert-Ermittlung
 - Ermittlung der Istwerte relevanter Metriken in den verschiedenen QoX-Schichten, z. B. des effektiven Durchsatzes anstatt der max. Download-Datenrate
 - Vergleich Sollwerte (QoS-Parameter) als unabhängige Variablen zu den Istwerten als interne abhängige Variablen
 - Bildung neuer Parameterklassen oder von Unterklassen
3. Signal-basierte Auswertung
 - Berechnung der Istwerte wirksamer Nutzerstimuli, z. B. Download-Dauern, Vorpufferzeiten, Anzahl Unterbrechungen etc. auf Basis der erstellten Modelle
 - korrekte und eindeutige Zuordnung der ermittelten Messwerte als Istwerte zu den Sollwerten und Bewertungen der einzelnen Aufgabenausführungen als zusätzliche Merkmale
 - Auswertung des Zusammenhangs der Istwerte als interne unabhängige Variablen zu den QoE-Bewertungen der Probanden als abhängige Variablen

Diese Schritte sind entweder automatisiert oder manuell/intellektuell möglich, siehe Abbildung 68. Einen Ansatz dafür bieten die konkreten Erkenntnisse zur Netznutzung der einzelnen Anwendungen aus den Voruntersuchungen. Eine manuelle Auswertung ist allerdings extrem aufwendig, entsprechend sind automatische Analysen zu bevorzugen, die große Mengen Verkehrsdaten schnell durchmustern. Dafür sind geeignete Software-Werkzeuge zu wählen bzw. zu

erstellen. In den Voruntersuchungen und einigen manuellen Auswerteschritten konnte sich Wireshark als Werkzeug zur Analyse der Netznutzung mit sinnvollen Filteroptionen und zahlreichen eingebauten Statistiken zu Paketströmen bewähren. Zudem besteht die Möglichkeit zur einfachen Erweiterbarkeit über die Skriptsprache Lua für automatisierte Analysen.

Mögliche Sichten auf die Netznutzung in der Tiefenauswertung sind:

- Ebene einzelner IP-Pakete: Größe, zeitlicher Abstand, ...
- Ebene zusammengehöriger IP-Pakete (UDP/TCP-, „Verbindungen“ als Flows, Sessions bzw. Streams): Datenmengen, Session-Zeiten, durchschnittliche Datenraten, Verhältnis Download zu Upload, ...
- Ebene gesamter Testlauf: Durchsatz (Datenraten) gemittelt und im zeitlichen Verlauf (Graphen), Gesamtdatenmenge und benötigte Gesamtzeiten

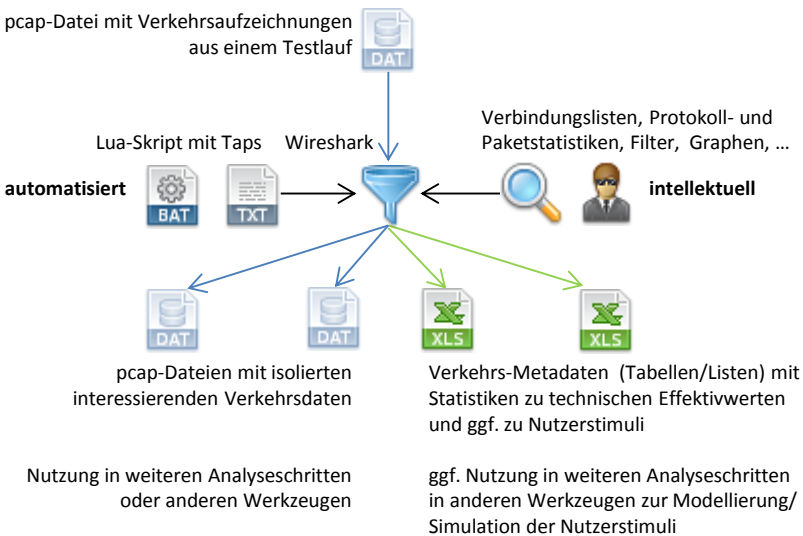


Abbildung 68: Mögliches Vorgehen bei der Tiefenauswertung der Verkehrsdaten ex post
Beispiel mit Wireshark als Analysewerkzeug

Erläuterungen zur Abbildung 69:

Beispiel intellektuell annotiert, begrenzte zeitliche Auflösung auf x-Achse mit 1 Tick/s, daher kann der Linienzug zerreißen. Aufgrund der großen Spreizung der QoS-Parameter über die Presets aller Szenarien hinweg ist eine kombinierte Darstellung mit linearer y-Achse (unten) mit intuitiv korrekter Skalierung zueinander und logarithmisch skaliertes y-Achse (oben) für den Detailvergleich zueinander vorteilhaft. So lassen sich über die Größenordnungen von unter 100 kBit/s bis zu mehreren MBit/s hinweg die jeweiligen charakteristischen Muster besser erkennen.

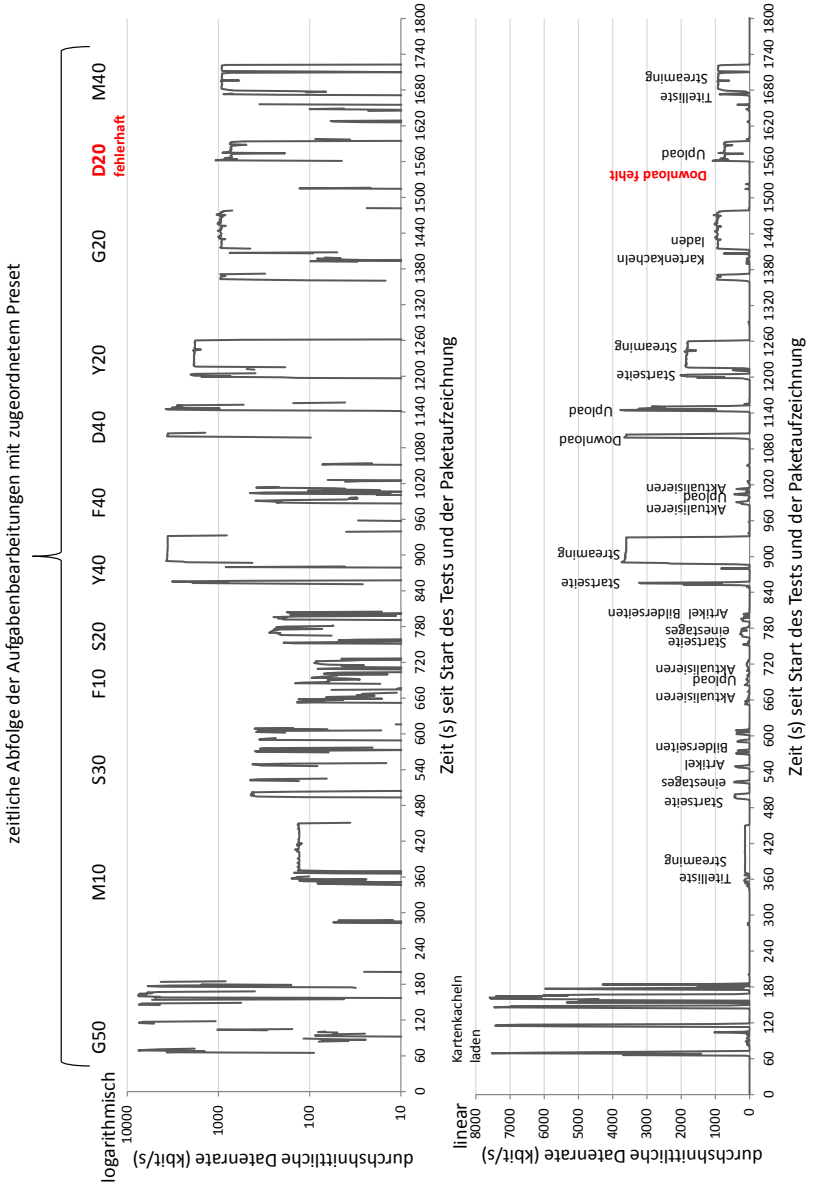


Abbildung 69: Analyse der Netzverkehrsdaten für den gesamten Testlauf eines Probanden

Für die Auswertung erfolgte eine manuelle Tiefenauswertung der Stufe 1 für die Vollständigkeit der Aufgabenausführungen (je ein Download und Upload eines Bildes je Aufruf) in der Aufgabe Drive, da in einer stichprobenartigen Überprüfung vermehrt Abweichungen festgestellt wurden. Ein Beispiel eines nicht ganz einwandfreien echten Testlaufs eines Probanden ist in Abbildung 69 dargestellt. Ebenfalls für das Szenario Drive erfolgte eine manuelle Tiefenauswertung der Stufe 2 für ausgewählte Stichproben zur Überprüfung der erreichten Effektivwerte. Weitere Anmerkungen dazu finden sich unter 9.1.2. Für die restlichen Dienstszenarien ergaben stichprobenartige Überprüfungen auf verschiedenen Stufen keine weiteren Auffälligkeiten oder die unbedingte Notwendigkeit des vollständigen Einsatzes einer Tiefenauswertung.

Für die Auswertung der Verkehrsdatenaufzeichnungen der Testrunde 1 wurde eine automatisierte Tiefenauswertung der Stufe 3 zum Dienstszenario Youtube entwickelt, da ein Preset unglücklich so gewählt wurde (Black Box!), dass ex ante nicht von sicheren Erwartungswerten zum Abspielverhalten auszugehen war. Auf Basis der zeitlichen Charakteristik der Übertragung und extrahierter Metadaten aus dem Datenstrom konnten Aussagen zur benötigten Wiedergabestartzeit und zum tatsächlichen Pufferfüllstand in den Endgeräten im Rahmen einer diskreten Simulation genutzt werden, um das Auftreten und die Dauer von Wiedergabeunterbrechungen ex post nachzuvollziehen. Die eigenen Untersuchungen wurden dabei von [300] und [301] inspiriert. Auch wenn derartig aufwendige Auswertemöglichkeiten ex post technisch prinzipiell möglich sind, so ist es sicher besser sie durch geschickte Testplanung und detaillierte Voruntersuchungen zu vermeiden, was für die Testrunde 2 glücklicherweise weitgehend gelang.

8.1.2 Signal-basierte Auswertung

Die Signal-basierte Auswertung fußt auf einer direkten und unmittelbaren Abhängigkeit der QoE-Bewertungen auf Basis erreichter Istwerte der Nutzerstimuli als relevante Signale (Messraum). Das Signal selbst ist dabei auf die variierten Parameter zurückzuführen (nach dem internen Modell). Eine Signal-basierte Auswertung auf Basis der Istwerte ist a priori als korrekt anzusehen, muss aber zusätzlich noch den gesuchten Zusammenhang zu den QoS-Parametern herstellen. Innerhalb des geschaffenen Testbeds wäre dies prinzipiell möglich, da sowohl die QoS-Effektivwerte (als Verkehrsdaten) wie auch die

wirksamen Nutzerstimuli (als Video) aufgezeichnet und abgeglichen werden können. Die Signal-basierte Tiefenauswertung kann ähnlich wie bei den Voruntersuchungen der Nutzungsszenarien erfolgen. Allerdings werden dann nicht nur einige exemplarische Aufgabenausführungen analysiert, sondern die Testläufe der Probanden aus der Durchführung der Studie – im Fall einer Vollauswertung für jeden Testlauf und sämtliche Aufgabenausführungen aller Probanden.

Im Rahmen der Tests sind große Menge Datenaufzeichnungen entstanden. Eine derartig detaillierte Analyse kann lohnen, ist aber angesichts des immensen Aufwandes als optional zu betrachten und wäre letztlich nur praktikabel durchzuführen, wenn automatische Lösungen die Auswertung der aufgezeichneten Daten zumindest unterstützen.

8.1.3 Parameter-basierte Auswertung

In der Parameter-basierten Auswertung hängen QoS-Parameter und QoE-Bewertungen nicht unmittelbar und direkt zusammen, sondern der Zusammenhang zwischen Parameterraum und Bewertungsraum wird über weitere (interne) Variablen vermittelt. Ist der interessierende QoS-QoE-Zusammenhang im Vorhinein (*ex ante*) zu beschreiben, so dass auch für die internen Variablen sinnvolle Erwartungswerte gebildet werden können, dann ist auch eine Parameter-basierte Auswertung möglich. In dem ausgearbeiteten Modell zum QoS-QoE-Zusammenhang von der Ursache zur Wirkung sind dazu neben dem Hauptfaktor der Netzparametrisierung in der Netzemulation die Neben- und Störfaktoren zu berücksichtigen. Dafür wurden die Zusammenhänge in den Voruntersuchungen mit den Mitteln des Testbeds analysiert. Durch die Grey-Box-Modelle kann in der Parameter-basierten Auswertung von den *ex ante* ermittelten Erwartungswerten für Effektivwerte und Nutzerstimuli ausgegangen werden, getrennt nach den identifizierten Partitionen, die sich durch die Einflüsse der Nebenfaktoren ergeben. Daher sei an dieser Stelle noch einmal auf die Voruntersuchungen der Nutzungsszenarien verwiesen. Dabei wurden die relevanten Überlegungen zu Nebenfaktoren und gegenseitiger Abhängigkeit der einzelnen Faktoren schon in der Phase der Vorbereitung integriert.

Je nach Aufgabe sind die Nebenfaktoren und Störfaktoren mit unterschiedlich starkem Einfluss auf die Bewertung zu sehen. Zum Teil sind dies Effekte, die sich auch in Realweltszenarien außerhalb der Laborumgebung zeigen und ohnehin nicht vollständig vermeidbar sind. Durch diese verbleibenden Neben- und Störfaktoren entsteht eine Art Rauschen. Solange externe oder unbekannte Störfaktoren nicht übermäßig stark einwirken, ist davon auszugehen, dass die interessierenden QoS-QoE-Zusammenhänge auch in einer modellgestützten Parameter-basierten Auswertung gut genug abzubilden sind. Entsprechend wird diese Variante zusammen mit einer Validierung der erzielten Ergebnisse für die eigene Auswertung gewählt. Dabei soll sich die Auswertung der QoE-Bewertungen hauptsächlich auf die gebildeten Presets als Kategorien beziehen, da diese nicht nur eine Netzparameterkombination repräsentieren, sondern auch Klassen von Erwartungswerten der Nutzerstimuli, z. B. bestimmte Ladezeiten.

Bei der Analyse und Bewertung der Ergebnisse sollte stets berücksichtigt werden, dass die parametrisierten Werte als Maximal- bzw. Minimalwerte angegeben sind und auch als solche verstanden werden sollten. Die dazu passende Aussage ist demnach „nicht besser als...“ bzw. „mit bis zu...“.

8.1.4 Signal-basiert vs. Parameter-basiert

Ob eine Signal-basierte Tiefenauswertung als Abgleich zwischen Parameterraum, Messraum und Bewertungsraum zwingend notwendig oder sinnvoll ist, ist letztlich eine Frage der gewünschten bzw. erforderlichen Genauigkeit. Im vorliegenden Fall der Laborstudie war die Ermittlung interessanter Bereiche wichtiger als eine punktgenaue Überprüfung wertexakter Zusammenhänge. Die Abstände der QoS-Parameterwerte in den Presets sind recht groß angelegt, so dass der Parameterraum mit nur fünf Presets je Dienst recht grob aufgelöst ist. Ohne starke externe Störung ist davon auszugehen, dass die Mehrheit aller tatsächlichen Istwerte wie beabsichtigt in die gleiche Preset-Klasse – als den Bereich bis zum nächstniedrigeren Preset definiert – fallen dürften. Umgedreht sollte dies bei der Wahl der Parametrisierung berücksichtigt werden. Eine feine Abstufung und kleine Abstände sind nur in Verbindung mit einer angestrebten und auch realisierbaren Genauigkeit sinnvoll und erfordern dann ggf. tatsächlich eine Signal-basierte Auswertung.

Prinzipiell ist ohnehin bei der Genauigkeit zwischen der Präzision der Parametrisierung und der Größe möglicher Störfaktoren sowie der Genauigkeit der Beobachter abzuwiegen. Solange die Verhältnisse der Größenordnungen dabei stimmen, ist davon auszugehen, dass sich Parameter-basierte und Signal-basierte Auswertung in den relevanten Ergebnispunkten nicht unterscheiden.

Beispiel: Für die Bewertung der Download-Dauer wird angenommen, dass es für die Bewertung nicht relevant ist, ob der Download exakt 10 s, 10,5 s oder 11 s gedauert hat, wenn der gesamte interessierende Bereich fünf Stufen im Bereich von 4 s bis 60 s umfasst.

Für eine statistische Auswertung müssten die vereinzelt Istwerte ohnehin wieder zu Klassen vergleichbarer Werte zusammengefasst werden. Es geht also bei der Bewertung der Genauigkeit letztlich darum, dass sichergestellt ist, dass die Werte richtig aggregiert werden und eine gemeinsame Ergebnismenge bilden, die überlappungsfrei bestimmten QoS-Parameterbereichen zugeordnet werden kann. In der Auswertung werden diese QoS-Parameterbereiche durch die Presets repräsentiert.

8.2 Validierung, Aggregation und Partitionierung

Nicht alle erzielten Ergebnisse sind gültig oder gleich, in dem Sinne, dass sie den gesuchten bzw. gleichen QoS-QoE-Zusammenhang zwischen Netzparametrisierung und Bewertungen abbilden. Entsprechend ist eine Vorverarbeitung der Ergebnisse vor der weiteren Auswertung sinnvoll, siehe auch Abbildung 70:

- Validierung: Begrenzung der zufälligen Abweichungen Soll zu Ist und des Einflusses der Probanden
- Aggregation: gezielte Zusammenfassung zusammengehöriger Ergebnisse
- Partitionierung: Auftrennung der systematischen Unterschiede

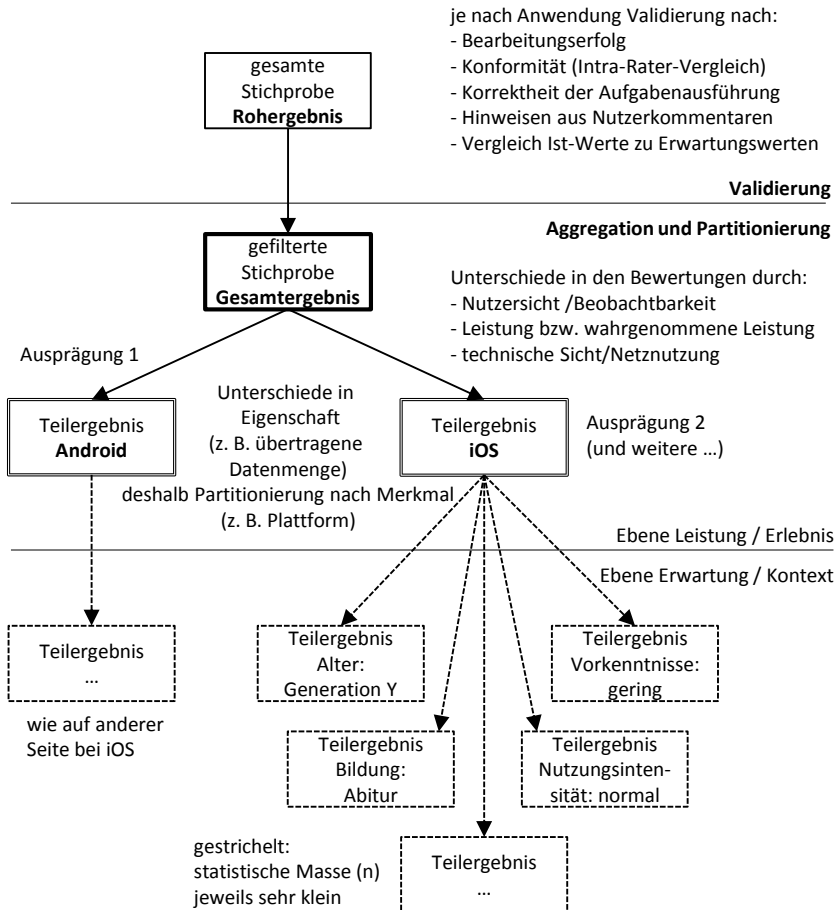


Abbildung 70: Prinzip der Verarbeitung der Rohergebnisse durch Validierung, Aggregation und Partitionierung

8.2.1 Vorvalidierung und Datenbereinigung

Die erhobenen Daten waren vor ihrer Auswertung zunächst in mehreren Schritten zu bereinigen, anzupassen (Kodierung) und zu kombinieren. Unvollständige Ergebnisdatensätze und solche mit unstimmgigen Angaben in der Nachbefragung, z. B. unerklärte Diskrepanzen zur Vorbefragung, wurden im Rahmen der Vorvalidierung der Erhebungsdaten auf Basis der folgenden Ausschlusskriterien vor der weiteren Auswertung ausgesondert (in Klammern Anzahl des Auftretens):

- vorzeitiger Abbruch des Testlaufs: nicht alle Aufgaben bearbeitet bzw. ohne Nachbefragung beendet (1x)
- mangelnde Vorkenntnisse zu den genutzten mobilen Plattformen (2x)
- mangelnde Vorkenntnisse zur Nutzung des mobilen Internets (3x)

Die betroffenen sechs Testläufe und Probanden werden nachfolgend nicht mehr betrachtet und sind nicht Teil der Gesamtergebnismenge. Es verbleiben so von 216 teilnehmenden Probanden noch 210 Probanden zur Auswertung mit insgesamt 2520 Aufgabenausführungen, die sich aus 420 Ausführungen zu jedem der 6 Dienste im Test ergeben.

8.2.2 Validierung

Die Validierung dient der Überprüfung einzelner Aufgabenausführungen hinsichtlich der korrekten Abbildung der interessierenden Zusammenhänge zwischen QoS-Parametrisierung und QoE-Bewertung. Auf dem Weg vom Rohergebnis zum Endergebnis ist sie mit der Zuweisung der einzelnen Ergebnisse zu den korrekten Partitionen kombinierbar. [302] S. 252 geht auf die verschiedenen Arten von möglichen Fehlern ein. Die Überprüfung und Bereinigung erfolgt normalerweise unter zwei Gesichtspunkten: systematische und zufällige Fehler.

Beim Wirken von systematischen Fehlern, z. B. durch Technik / Messung / Testbed verursacht, ist mit einem konstanten Versatz in allen vergleichbaren Aufgabenausführungen zu rechnen, der sich ggf. ermitteln und ggf. abtrennen lässt. Zufällige Fehler sind durch äußere Einflüsse oder die Probanden verursacht, ohne dass Regelmäßigkeiten als Ursache betrachtet werden. Sie sind in konkreten Ausprägungen ggf. auch nur in einzelnen Aufgabenausführungen wirksam. Kleinere zufällige Einflüsse sollten sich durch die Randomisierung im Testplan kompensieren (besonders mit größerem Testumfang). Extreme zufällige Einflüsse sollten ausgeschlossen bzw. ausgefiltert werden.

Fehler soll an dieser Stelle nicht im Sinne von *komplett falsch* verstanden werden, sondern als ein ungeplanter oder vorab unberücksichtigter Nebenfaktor oder Störfaktor. Dieser kann auf die Ergebnisse einwirken – muss es aber nicht. Durch die schon berücksichtigten anderen Faktoren kann er in Richtung und Stärke überlagert und somit in den Ergebnissen möglicherweise ohnehin

maskiert sein. Systematische Fehler sollten durch die Voruntersuchungen weitgehend minimiert sein. Durch die Komplexität der Zusammenhänge und die Vielzahl möglicher Einflussfaktoren sind sie aber grundsätzlich nicht a priori auszuschließen.

Werden in einem ersten Schritt der Identifikation und Bewertung mögliche Abweichungen erkannt, ist in einem zweiten Schritt zu prüfen, welche Gegenmaßnahmen möglich sind, z. B. in Form von Isolation (Invalid-Markierung) oder Korrektur, z. B. durch Kompensation.

Es sei noch einmal auf die Ausführungen zur Entstehung der QoE-Bewertungen verwiesen: Die Probanden bewerten ihr Erlebnis über die wahrgenommene Leistung im Abgleich mit ihrer Erwartung im Kontext. Daher sind prinzipiell nur solche Bewertungen relevant, die sich auf die beabsichtigte Leistungswahrnehmung beziehen. Stark abweichende Aufgabenausführung oder technische Fehlfunktionen führen potenziell zu anderen als den beabsichtigten Leistungswahrnehmungen und sollten daher für die QoS-QoE-Betrachtungen ausgenommen werden, indem sie als invalid gekennzeichnet werden. Die Bewertung bezieht sich zwar auf irgendetwas, misst aber nicht den gesuchten Zusammenhang.

Zur Überprüfung einer korrekten Aufgabenausführung als Voraussetzung für eine valide Bewertung können unterschiedliche Ansätze dienen. Dienstspezifisch ist zu prüfen, welche der Möglichkeiten zur Validierung genutzt werden sollen (nach steigendem Aufwand sortiert):

- aus der Datenerhebung des Fragebogens
 - Bearbeitungserfolg
 - Hinweise aus Nutzerkommentaren
 - Intra-Rater-Vergleich
 - starke Diskrepanzen zwischen Teilbewertungen und Gesamtzufriedenheit
- Auswertung Testlogbuch (während der Tests oder danach festgestellte Besonderheiten)
- Tiefenauswertung der Verkehrsdaten- und Videoaufzeichnungen
 - Korrektheit der Aufgabenausführung
 - Vergleich Istwerte zu Erwartungswerten

Zur Auswertung wurden entsprechende Metadaten zu jedem Datensatz mitgeführt, die ermöglichen, die zugehörigen Bewertungen als invalid zu kennzeichnen und mit einem Auswertekommentar in Form eines Freitextes zu versehen. Einige der aufgeführten Punkte werden in den nachfolgenden Abschnitten noch einmal getrennt angesprochen.

Grundsätzlich ist die Notwendigkeit einer Validierung in allen Szenarien für die Gesamtzufriedenheit und für die einzelnen Teilbewertungen zu prüfen. Dabei können die Bewertungen jeweils für sich isoliert oder im Fall der Gesamtzufriedenheit als logische Verknüpfung betrachtet werden. Dies ist ausgehend von der Analyse zum Zusammenhang der Zufriedenheit in den Teilbewertungen zur Gesamtzufriedenheit abzuschätzen. So kann es sinnvoll sein anzunehmen, dass eine valide Bewertung der Gesamtzufriedenheit nur zustande kommt, wenn alle Teilbewertungen valid sind, weil ihnen ein großer Einfluss auf die Bildung der Gesamtzufriedenheit zugerechnet wird.

Beispiel Drive: Es ist ein eindeutiger und straffer Zusammenhang zwischen Teilbewertungen und Gesamtzufriedenheit zu erkennen. Die Bewertung der Gesamtzufriedenheit kann damit nur valid sein, wenn die Teilbewertungen selbst valid sind.

Die Unterscheidung von validen zu invaliden Testergebnissen ist vor allem bei der Zerlegung der Gesamtergebnismenge in Teilergebnismengen zu beachten. Bei einer relativ kleinen Teilergebnismenge können schon recht wenige invalide Werte das Teilergebnis deutlich verzerren, wenn sie etwa durch eine technisch bedingte Häufung (wie ein Usability-Problem) einen entsprechenden Prozentsatz erreichen.

8.2.3 Aggregation vs. Partitionierung

Grundannahmen: Prinzipiell wird eine Teilergebnismenge durch jeden (Einfluss-) Faktor aufgespannt. Es ist aber nicht sinnvoll „1001“ Teilergebnismengen zu betrachten, von denen jede nur 3 Bewertungen umfasst. Andererseits ist es auch nicht sinnvoll, von einer Ergebnismenge auszugehen, wenn z. B. deutlich 2 oder 3 Teilergebnismengen voneinander zu unterscheiden und zu erklären sind. In der Auswertung sollen für den QoS-QoE-Zusammenhang bedingte Verteilungen erstellt werden. Unterscheidungsmerkmal soll dabei die

jeweilige Testbedingung sein, primär also die Presets als Kombination von Dienst und Netzparametern, bei weiterer Annahme der sonstigen Gleichheit.

Aggregation: In der Auswertung der Ergebnisse wird allgemein ein hoher Verdichtungsgrad angestrebt. Ergebnisse werden so zusammengefasst, dass Aussagen höherer Aussagekraft durch eine größere Zahl von Beobachtungen bzw. allgemeingültige Aussagen entstehen. Dabei ist ein Abwägen zwischen Grobstruktur (Verallgemeinerung) und Feinstruktur (spezielles Ergebnis) notwendig. Eine Verdichtung der Ergebnisse kann dabei in Form von Hierarchien erfolgen, durch Dimensionserniedrigung und Gruppierung.

Dimensionserniedrigung und Faktorisierung (Reduzierung der Merkmale):

- Die Dimensionserniedrigung erfolgt durch Zusammenfassen von Einzelmerkmalen in einer Merkmalskombination. Beispiel: Die Presets bilden eine solche Kombination als n-Tupel aus einzelnen Netzparametern.
- In der Faktorisierung werden zusammengesetzte und redundanzfreie (interne) Einflussgrößen gebildet. Beispiel: Für ein Dienstszenario relevante Effektivwerte sind der Goodput im Download und damit verbunden die Download-Dauer.

Gruppierung und Klassierung (Reduzierung der Ausprägungen):

- Verringerung von Merkmalsausprägungen durch Zusammenfassen mehrerer Werte zu einem neuen Gruppenwert – Beispiel: „Android“ bildet einen solchen Gruppenwert für die Ausprägungen „Android 2.x“ und „Android 4.x“.
- Bildung von Klassen auf Basis quantitativer Merkmale, z. B. durch Zusammenfassung und Intervalleinteilungen – Beispiel: Die Presets bilden solche Klassen für die erreichten Effektivwerte der QoS-Parameter.

Innerhalb der Presets ist Gleichheit bzw. zumindest Ähnlichkeit der Bedingungen anzunehmen, wenn:

- die Nutzersicht mit wahrnehmbarer Leistung und Präsentation vergleichbar ist
- die technische Sicht (Netznutzung) mit objektiver Leistung vergleichbar ist

Partitionierung: Von der Annahme der Gleichheit wird abgewichen, wenn es Hinweise auf weitere potenziell relevante Unterscheidungsfaktoren in Form der Nebenfaktoren durch die Heterogenität des Testbeds, das Testdesign und Zufallsgrößen gibt. Die Bildung von Teilergebnissen, nachfolgend auch Partitionen oder Gruppen genannt, erfolgt anhand eindeutiger Unterscheidungsmerkmale. Dies kann auch hierarchisch über mehrere Faktoren erfolgen (mehrere Ebenen). Die Erkenntnisse zu möglicherweise relevanten Faktoren für unterschiedliche QoS-QoE-Zusammenhänge können in verschiedenen Phasen der Studie entstehen:

- in den Vorüberlegungen und Voruntersuchungen ermittelter Ursache-Wirkungs-Zusammenhang: Die Unterschiede sind schon in Vorfeld aufgefallen und wurden ggf. schon in der Testplanung und Durchführung berücksichtigt (fachlich-sachliche Begründung).
- in der Auswertung: Die Unterschiede treten erst in den Ergebnissen hervor (statistische Ermittlung und Begründung).

Erlebnismfaktoren vs. Erwartungsfaktoren: Auf der QoE-Ebene kann einen Einfluss nehmen, was das Nutzererlebnis über die Leistung und die Wahrnehmung der Leistung moduliert (Erlebnismfaktoren). Prinzipiell gelten die Überlegungen auch für Annahmen zu möglichen Unterschieden in der Erwartung (Erwartungsfaktoren). Unterscheidungen zu den Erwartungsfaktoren sind allerdings nur für jeweils gleiche Leistung und Präsentation sinnvoll. Entsprechend sind alle Partitionierungen und Einflussfaktoren der Ebene Leistung/Erlebnis vorrangig zu berücksichtigen, siehe Abbildung 70. Durch die nachrangige Differenzierung der Erwartungsfaktoren nach den Erlebnismfaktoren ist die jeweilige statistische Masse klein. Dadurch werden statistisch sinnvolle Untersuchungen und Aussagen zu den Erwartungsfaktoren schwierig. Auf Aussagen zur Wirkung der Erwartungsfaktoren, z. B. als Attribute der Probanden wie Alter oder Bildung, wird im Weiteren daher weitgehend verzichtet. Etwaige vorhandene (geringe) Unterschiede verbleiben als Rauschen in den leistungsbegründeten Teilergebnissen und mitteln sich ggf. aus.

Fazit zu Aggregation vs. Partitionierung: Für gleiche Bedingungen ist eine aggregierte Ergebnismenge möglich, für unterschiedliche Bedingungen ist eine Unterscheidung in verschiedene Teilergebnisse (Partitionen) des Gesamtergebnisses anzunehmen. So wird sichergestellt, dass innerhalb der erzielten

Teilergebnisse jeweils der gleiche QoS-QoE-Zusammenhang beschrieben wird, der möglichst beabsichtigt und über das Modell erklärbar ist. Da in den Teilmengen der jeweilige Stichprobenumfang kleiner ist als in der Gesamtmenge, ist für die Teilmengen der statistische Standardfehler größer als in der Gesamtmenge. Das bedeutet allerdings auch, dass es ein Dilemma gibt: Eine an sich formal korrekte Zerlegung wird nicht automatisch besser. Zwar gewinnt sie an Aussagenschärfe, weil fremde Bewertungen ausgeblendet sind, verliert aber gleichzeitig an Aussagekraft, weil es dadurch weniger Bewertungen sind, die zum Teilergebnis beitragen. Daraus folgt: Es werden so viele Einzelergebnisse wie möglich aggregiert und nur so viele Teilergebnisse wie nötig gebildet.

8.3 Bearbeitungserfolg

Die Antwortoptionen zum Bearbeitungserfolg und den möglichen Zusatzangaben sind der Darstellung des Fragebogens im Anhang zu entnehmen, siehe A.4.4. Eine Übersicht zu den erzielten Angaben zum Bearbeitungserfolg findet sich ebenfalls im Anhang, siehe A.5.2.

Die Analyse des Bearbeitungserfolgs kann Hinweise geben auf:

- Qualität der Aufgabenbeschreibung: Waren Aufgabenstellung und Anleitung verständlich?
- Vertrautheit bzw. Vorkenntnisse der Probanden: Eine hohe sofortige Erfolgsquote deutet auf wenige Probleme bzw. entsprechende Vorkenntnisse hin.
- mögliche Usability-Probleme, die zu Aufgabenabbrüchen führen
- Toleranzschwelle, die Probanden brechen die Aufgabenbearbeitung QoS-abhängig entnervt ab

Grundsätzlich gilt: Sind Probleme zu identifizieren, die nicht von der beabsichtigten Parametervariation beeinflusst worden sind, dann sollten diese Aufgabenausführungen als invalid gekennzeichnet werden, damit sie das Ergebnis nicht verzerren.

8.3.1 Auswertung der Angabe zum Bearbeitungserfolg

Nicht sofort gelöst: Anderen Antwortoptionen als „ja, Aufgabe beim ersten Versuch gelöst“ sind als „nicht sofort gelöst“ zu interpretieren. Für einige Aufgaben kann es sinnvoll sein, nur die Aufgabenausführungen als wertvoll zu erachten, die im ersten Anlauf erfolgreich bearbeitet worden sind. Bei einem weiteren Ausführungsversuch könnte durch den vorherigen Versuch eine Verzerrung verursacht werden, wenn z. B. doch noch Datenübertragungen im Hintergrund laufen oder schon Inhalte in einer Cache gespeichert worden sind. Dies ist je nach Szenario zu prüfen. Der angegebene Grund zum Scheitern (im ersten Anlauf) und die Nutzerkommentare können einen Hinweis liefern, ob eine solche Bearbeitung ein vermutlich (in-) valides Ergebnis hervor gebracht hat.

Abbruch der Aufgabenbearbeitung: Sofern mit der beabsichtigten Parametervariation in Verbindung zu bringen („Probleme mit der Internetverbindung“), können diese Abbrüche interessante Hinweise geben, wo bei den Probanden eine rote Linie hinsichtlich der Toleranz in Bezug auf die Wahrnehmung von Netzeigenschaften über Wartezeiten oder Funktionseffekte zu sehen ist. Dann ist zu vermuten, dass auch die Bewertungen entsprechend schlecht ausfallen.

In anderen Evaluationen werden Abbrüche einer Aufgabenausführung als so schwerwiegend erachtet, dass dafür in der quantitativen Auswertung mittels Intervallskala ein zusätzlicher Score-Wert von „0“ zugeordnet wird (wie in [183] S. 25). Entsprechend sinkt in zusammenfassenden Betrachtungen, z. B. mit Maßzahlen wie dem Mittelwert, der berechnete Wert noch stärker als nur mit einer schlechten Bewertung. In der eigenen Auswertung wurde auf diese Möglichkeit verzichtet.

Teilweise scheinen die Probanden im Fall eines Aufgabenabbruchs nicht plausibel zu antworten, z. B. indem die Bewertungen alle „vollkommen zufrieden“ ausfallen. Daher ist bei einer Häufung solcher Bewertungen von einer Verzerrung der Gesamtergebnisse auszugehen. Dies deutet daraufhin, dass irgendetwas als Antwortoption gewählt wurde, nur um die Aufgabe im Fragebogen abschließen zu können. Für den Bearbeitungserfolg kann deshalb sinnvoll sein, bei anderen Angaben als „beim ersten Versuch gelöst“ generell genauer

zu prüfen, ob eine korrekte Aufgabenausführung und eine plausible Bewertung vorliegen.

Keine Bearbeitung der Aufgabe: Für den Fall, dass die Bearbeitung einer Aufgabe gar nicht erst gestartet werden konnte, wurde nur die Gesamtzufriedenheit ohne weitere Teilbewertungen erfasst. Konkret betrifft dies lediglich zwei Aufgabenausführungen in der gesamten Testrunde. Beide Bewertungen wurden als invalid eingestuft, da nicht mit der Netzparametrisierung in Verbindung zu bringen. Bei der weiteren Auswertung wird nicht noch einmal gesondert auf diese Aufgabenbearbeitungen eingegangen.

8.3.2 Auswertung des Grundes für Abbruch oder Wiederholung

Mit Blick auf die Auswertung verschiedener Netzsituationen erscheint die Frage interessant, ob Probanden tatsächlich die Eigenschaften der Internetverbindung als ursächlich ansehen, wenn sie eine Aufgabe nicht sofort bearbeiten konnten – bzw. dazu nicht genug Geduld aufgebracht haben.

Für die Angabe „Sonstiges“ wurde überprüft, ob die Freitextanmerkungen eigentlich einem der vorgegebenen Gründe zuzuordnen sind. In der Auswertung wurden entsprechende Einträge gesichtet und als Angabe mit dem jeweiligen Grund rekodiert. Für alle Angaben außer denen, die auf die Internetverbindung hindeuten, ist dies als unkritisch zu erachten. Es wird im Folgenden für diese Angaben nicht weiter zwischen direkten Angaben der Probanden und der Rekodierung der Auswertephase unterschieden. Diese vereinfachte Betrachtung kann dazu dienen, Bedienprobleme etc. nachträglich besser einschätzen zu können. Alle Angaben, die hingegen auf die Internetverbindung hinweisen, werden in der Auswertung der Szenarien gesondert diskutiert und quantifiziert. Es ist eine interessante Zusatzinformation, ob die Probanden nicht nur erkannt haben, dass es ein vermeintliches Problem gab, sondern ob sie es ursächlich auch richtig (oder ggf. auch falsch) der Netzsituation zuordnen konnten (Schluss von der Wirkung auf die Ursache).

8.4 Nutzerkommentare

Eine stichprobenartige Sichtung der Nutzerkommentare zeigte, dass sich die Analyse durchaus lohnen kann. Auch ohne Nennung des Namens eines bestimmten Dienstes lässt sich bei Kenntnis der getesteten Szenarien von den Kommentaren auf die jeweilige Aufgabe schließen. Prinzipiell würde eine vollständige Erschließung der Nutzerkommentare in Form einer qualitativen Analyse den größten Erkenntnisgewinn bedeuten. Auf eine vollständige intellektuelle qualitative Auswertung inklusive Kodierung der Nutzerkommentare wurde aber verzichtet. Weit weniger aufwendig, aber doch recht nützlich, sind automatische Analysen unter Nutzung geeigneter Werkzeuge zur Textauswertung. Der gewählte Ansatz mit Word-Clouds ist recht einfach, kann aber schon wichtige Punkte und Tendenzen identifizieren. Eine versuchsweise durchgeführte Stimmungsanalyse (Sentiment Analysis in *Rapidminer*²⁰⁰) zum Abgleich der Nutzerkommentare mit den abgegebenen Bewertungen ergab keine relevanten Zusatzkenntnisse (siehe Beispiel zum Szenario Drive in Tabelle 53).

8.4.1 Word-Clouds

Word-Clouds visualisieren die Worte eines Textes (bzw. einer Menge von Texten) als einzelne Elemente innerhalb einer Menge von Worten, der Wortwolke. Dabei werden häufig vorkommende Worte visuell auffälliger kodiert, z. B. größer oder mit stärkerem Kontrast etc. Zum Konzept und zur geeigneten visuellen Kodierung sowie zu den ablesbaren Aussagen gibt [303] einen Einblick. Die Aussagekraft solcher Visualisierungen wird teilweise als nicht allzu hoch eingeschätzt, siehe [304]. Mit der nötigen Vorsicht betrachtet, sind sie aber doch recht nützlich und vor allem mit Blick auf die vorab identifizierten bzw. allgemein vermuteten Problemfelder sehr schön anschaulich.

²⁰⁰ <https://rapidminer.com/solutions/sentiment-analysis/> abgerufen am 05.11.2014

Für die Auswertung der eigenen Daten lassen sich die Verarbeitungsschritte und Tool-Kette wie folgt beschreiben:

Rohtexte (Kommentarfelder) → Tokens (Einzelworte) → optional: Lemmatisierung²⁰¹ (Lemma und Wortart) → Bereinigung (Sonderzeichenfilterung/-konvertierung, optional: vereinheitlichte Kleinschreibung) → Stopp-Wort-Filterung (Deutsch) → Visualisierung (Word-Cloud)

Die Lemmatisierung, also die Rückführung auf die Grundform der einzelnen Worte, erhöht zwar die Häufigkeit der jeweiligen Wortgrundform und damit die visuelle Präsenz, vernichtet aber die wertvollen Zusatzinformationen (z. B. Komparativ), die auf die Art der Bewertungen der Probanden als Vergleiche zwischen den Aufgabenausführungen hinweisen. Im Folgenden wurde deshalb darauf ebenso wie auf durchgängige Kleinschreibung verzichtet. Eine effektive Stopp-Wort-Filterung ist hingegen für eine Analyse auf Basis der Häufigkeitsverteilung der Worte unbedingt notwendig, da ansonsten „ich, es, ist, war“ etc. die Analyse verzerren. Die meisten Visualisierungstools im Umfeld von Word-Clouds sind sehr verspielt (reine Eye-Catcher).



Abbildung 71: Visualisierung der Nutzerkommentare als Word-Cloud, Beispiel Drive

²⁰¹ mit TreeTagger, ursprünglich an der Universität Stuttgart entwickeltes Werkzeug mit einer brauchbaren Lemmatisierung für Deutsch, <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> und GUI <http://www.smo.uhi.ac.uk/~oduibhin/oideasra/interfaces/winttinterface.htm> abgerufen am 09.10.2013

Für die eigenen Darstellungen (Beispiel in Abbildung 71) sollen einfache Grundregeln gelten:

- Für die Berücksichtigung in der Visualisierung ist eine Mindesthäufigkeit von zwei Vorkommen in den Kommentaren notwendig mit einer Begrenzung auf die Top-100-Worte.
- Häufigkeit der Worte: Häufige Worte werden als wichtig erachtet und groß mit starkem Kontrast (dunkel auf hellem Hintergrund) dargestellt. Seltener auftretende Worte hingegen werden kleiner und mit schwächerem Kontrast (ausbleichendes Grau) visualisiert.
- Die absolute Lage bzw. relative Position zueinander ist willkürlich.

8.4.2 Inhaltliche Analyse der Nutzerkommentare

Bei der stichprobenartigen Analyse einzelner Kommentare, aber auch der zusammengefassten automatischen Aufbereitungen lassen sich einige Schwerpunkte der Nutzeräußerungen erkennen:

- Aufgabe und Aufgabenstellung
- App und Usability
- persönliche Haltung, z. B. „Facebook-Verweigerer“ bzw. persönliche Meinung, z. B. zum genutzten Inhalt („Schrottfilm“)
- Beobachtungen und Aussagen zu den aus Nutzersicht relevanten Eigenschaften, teils als Umschreibung, teils auch zusätzlich quantifiziert
- Zeit: z. B. „dauerte lange“, „30 Sekunden“, ...
- Funktion und Verhalten: „stockte“, „Unterbrechungen“, „mehrfach“, „2 mal“, ...
- Quantität des Inhalts: Bild- bzw. Dateigröße
- Qualität des Inhalts: „Top“, „Super“, „HD“, „rauschte“
- Fehler- oder Problembeschreibungen, teils mit Ursachensuche als Vermutungen oder gar zusammen mit vermeintlichen Erklärungen
- verbalisierte Zufriedenheitsbewertungen
- Wahrnehmung der Netzeigenschaften (WLAN/Mobilfunk)

Es lassen sich grundsätzlich positiv bzw. negativ besetzte Äußerungen finden. Bei den Pflichtfeldern für Freitext zum Abspielverhalten bei MTV-Music und Youtube beschränkten sich die Kommentare hauptsächlich auf negative Aussagen, da die Frage auch nur erschien, wenn sich die Nutzer nicht „vollkommen zufrieden“ mit dem Abspielverhalten zeigten.

Bei den Verbalisierungen zur Zufriedenheit ist zu erkennen, dass die Probanden zu Verdikten und Vergleichen neigen. So vergleichen sie die Aufgabenausführungen, was bei der zweiten Ausführung eines Szenarios dann zu Formulierungen wie „langsamer“ bzw. „schneller“ führt. In einzelnen Fällen vergleichen die Probanden auch zwischen den Diensten, so bei Drive und Facebook, weil wohl in beiden Szenarien durch ähnliche Aufgabenstellungen mit Download und Betrachten bzw. Hochladen eines Bildes ähnliche Erwartungen und Stimuli zusammenkommen.

Über eine Analyse der genutzten Worte nach Wortart und Häufigkeit (ohne Stoppwörter) lassen sich schnell isolieren:

- Substantive: wichtige Eigenschaften und Attribute, die die Zufriedenheit bestimmen
- Verben und substantivierte Verben: wichtige (Handlungs-) Bestandteile der Ausführung der Aufgabe bzw. allgemein beim Nutzen des entsprechenden Dienstes
- Adjektive und Adverbien: positiv oder negativ annotierte Charakterisierungen der Eigenschaften bzw. Handlungen (qualitative Aussagen)
- Kardinalwerte: konkrete Dauern oder Häufigkeiten (quantitative Aussagen auf Schätzbasis)

So lässt sich einschätzen:

- Was ist für die Nutzer entscheidend?
- Über welche Stimuli wird es wie wahrgenommen?
- In welcher Richtung und Stärke sind die Äußerungen einzuordnen, z. B. als Tendenz „eher langsamer“ oder als Abgrenzungswert („viel zu langsam“, „extrem langsam“)?

Insgesamt zeigte sich, dass die vorab gebildeten Annahmen zu den relevanten Nutzerstimuli und die erwartete Ausprägung der Nutzerzufriedenheit weitgehend mit den von Nutzern stammenden Kommentaren übereinstimmen. Ausgehend von den Annahmen zu Frustrationsgrenzen wären bei den „schlechten“ Netzparameterkombinationen (jeweils Stufe 10) eigentlich fast noch mehr wirklich negative Kommentare zu erwarten gewesen. Bei sehr deutlichen Unterschieden in der Netzparametrisierung, entweder zeitlich kurz von einer zur

nächsten Aufgabe oder auch zwischen den beiden Aufgabenausführungen eines Szenarios, entsteht teilweise Verwunderung bei den Probanden, wie stark sich die „Internetsituation“ als Gesamtheit der wahrgenommenen Netzeigenschaften verändert hat („erstaunlich“, „irritieren“, „wundersam“, „überraschen“, ...).

8.5 Bewertungsverhalten

Die Analyse des Bewertungsverhaltens ergänzt die nachfolgenden Betrachtungen zu den bedingten Verteilungen der Zufriedenheit. So lässt sich erkennen, ob die Bewertungsunterschiede in den bedingten Verteilungen nur von verschiedenen Probanden mit unveränderlicher, fester Zufriedenheit herrühren, oder weil die Probanden tatsächlich in Abhängigkeit der QoS-Parametrisierung ihre Zufriedenheitsbewertung ändern. Der Teilbewertungsvergleich und die Einzelergebnisse zu den Teilbewertungen können helfen, die Gesamtzufriedenheit zu deuten. Der Vergleich der Teilbewertungen zur Gesamtzufriedenheit wird mit dem Intra-Rater-Wiederholungsvergleich zum Bewertungsverhalten zusammengefasst. Dabei werden auch die Visualisierungen zur Auswertung kombiniert.

8.5.1 Teilbewertungen vs. Gesamtzufriedenheit

Dieser Teil der Auswertung soll klären, wie die dienstspezifischen Einzelbewertungen und die Bewertung der Gesamtzufriedenheit zusammenpassen. Unterschiede zwischen der Bewertung der Gesamtzufriedenheit und den dienstspezifischen Teilbewertungen können unterschiedliche Ursachen haben:

- Die Probanden beziehen weitere Kriterien zur Bildung der Gesamtzufriedenheit ein, die nicht über die abgefragten Einzelzufriedenheiten bzw. das zugrundeliegende Modell der relevanten Stimuli erfasst sind.
- Die Probanden gewichten die Teilbewertungen bei der Bildung der Gesamtzufriedenheit unterschiedlich stark.
- Die Probanden verstehen die Intention der abgefragten Teilbewertungen oder der Gesamtbewertung nicht richtig.
- Die Stimuli und damit die QoE-Teilbewertungen ändern sich bei Variation der QoS-Parameter unterschiedlich stark, ggf. auch nur aus individueller Sicht der Probanden.

- Einzelne Teilbewertungen sind invariant gegenüber den QoS-Parameteränderungen.
- Es sind Fehler oder Probleme in der Aufgabenausführung aufgetreten, wodurch ein Teilaspekt stark abweichend bewertet wird.
- Das Auftreten eines zufälligen externen Einflussfaktors führte zu Fluktuationen und so zur internen Verzerrung eines Teilaspekts.

Beim Versuch dieser Aufschlüsselung kann nachteilig sein, dass im Fragebogen die Gesamtzufriedenheit erst nach den Bewertungen der Teilaspekte abgefragt wurde. Eine möglicherweise bessere Variante wäre gewesen, die Gesamtzufriedenheit vor den dienstspezifischen Teilaspekten abzufragen. So bleibt unsicher, ob die Probanden sich bei der vorrangig interessanten Bewertung der Gesamtzufriedenheit ausschließlich auf die erfassten Einzelaspekte beziehen, oder von diesen losgelöst möglichst umfassend ihr Gesamterlebnis bewerten. Im statistischen Sinn können die Bewertungen der Gesamtzufriedenheit und der Teilbewertungen je Aufgabenausführung als verbunden betrachtet werden (zu einem Datensatz gehörig).

Für die Darstellung der Gesamtheit der Ergebnisse wird die Verteilung der Bewertungsstufendifferenz aus Gesamtzufriedenheit und den einzelnen Teilbewertungen ermittelt, siehe Abbildung 72.

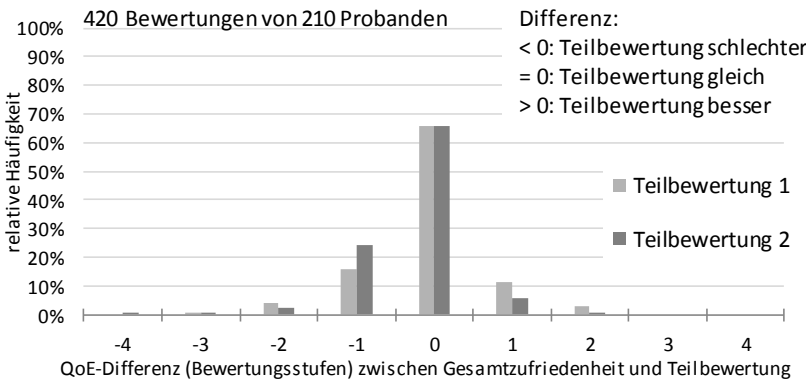


Abbildung 72: Vergleich der Gesamtzufriedenheit zu den Teilbewertungen am Beispiel von Drive

Ein enger Zusammenhang zwischen Gesamtzufriedenheit und einer Teilbewertung zeigt sich in einer betragsmäßig kleinen Differenz zwischen diesen, idealerweise 0. Dann ist davon auszugehen, dass die Teilbewertung zur Gesamtzufriedenheit beiträgt und diese erklärt. Dies bedeutet auch, dass die Teilbewertung selbst zwischen den QoS-Parameterstufen zu differenzieren ist. QoS-Änderungen führen somit nicht nur zu Änderungen der Gesamtzufriedenheit, sondern auch (im gleichen Umfang) zu Änderungen der Teilbewertung.

Negative Differenzen (insbesondere große) weisen auf besonders kritische Teilbewertungen hin. Diese sind vor allem für Stimuli der Kategorie Funktion & Verhalten zu vermuten. Positive Differenzen sind eher für die Teilbewertungen zu den Stimuli zu vermuten, die die Probanden möglicherweise als weniger kritisch empfinden. Dazu können in einigen Aufgabenszenarien auch die Teilbewertungen zählen, die sich auf weniger deutlich starke Stimuliänderungen beziehen und überwiegend ohnehin positiv bewertet werden.

Zur Interpretation des Beispiels in Abbildung 72: Für ca. 66 % der Ergebnisse erklären die dienstspezifischen Zufriedenheitsbewertungen die Gesamtzufriedenheit vollständig (Wertübereinstimmung). Unterschiede von mehr als einer Bewertungsstufe nach oben oder unten sind selten (Download in Summe ca. 7 %, Upload in Summe ca. 3 %). Es existieren auch sehr seltene Extremfälle, bei denen die Bewertungen stark divergieren. Die Dauer des Uploads (Teilbewertung 2) wird im Vergleich etwas schlechter bewertet. (Eine mögliche technische Erklärung liefern die Abweichungen des Upload-Effektivwerts beim Ace 2.)

8.5.2 Intra-Rater-Vergleich: Wiederholungsanalyse und Bewertungskonformität

Im genutzten Testdesign führt jeder Proband für jedes Anwendungsszenario zwei Aufgabenbearbeitungen aus und bewertet diese getrennt. Die Bedingungen wurden den Probanden dabei als Presets zufällig zugewiesen. Jeweils zwei Bewertungsdatensätze zu einem Dienst sind somit über einen Probanden verbunden. Auf dieser Basis erfolgt der Intra-Rater-Vergleich als Vergleich der Probanden mit sich selbst. In klassischen Intra-Rater-Vergleichen ist die Anzahl Rater (die Probanden als Bewerter) meist klein, die Rater bewerten dann

allerdings alle eine Vielzahl Subjects (die Testobjekte, hier Aufgabenbearbeitungen) mehrfach. Das allgemeine Prinzip und die relevanten Fachtermini zum Intra-Rater-Vergleich, vor allem zur Intra-Rater-Reliabilität, erläutert [305]. Im Fall der eigenen Studie war durch das Testdesign die intrapersonelle Variation allerdings auf o. g. zwei Bearbeitungen begrenzt.

Mögliche Aussagen und Erkenntnisse: Der intrapersonelle Vergleich lässt verschiedene interessante Rückschlüsse zur Aufgabe und den Probanden zu:

- Waren QoS-Bereich und QoS-Abstufung gut gewählt, so dass passend zu den Vorüberlegungen auch entsprechende QoE-Änderungen erreicht wurden?
- Sind die Antworten erwartungskonform? Ist zwischen den beiden Bewertungen eines Probanden ein Unterschied entsprechend der Richtung der QoS-Änderungen gegeben und wie stark ist diese Bewertungsänderung?
- Hat die Aufgabenstellung für die Probanden gut funktioniert? Eine große Zahl stark von der erwarteten Richtung der Bewertungsänderung abweichender Probanden kann ein Hinweis sein, dass es Probleme gab.
- Wie präzise bewerten die Probanden? Während die Änderungen der QoE-Bewertungen je nach QoS-Änderung auch unterschiedlich stark ausfallen können, lässt sich über den Vergleich für Ausführungen mit gleichen QoS-Parametern etwas über die Bewertungsgenauigkeit aussagen. Bewerten die Probanden gleiche QoS-Parameter individuell gesehen auch gleich, bzw. wie stark weichen sie von ihrer eigenen ersten Bewertung ab? Starke Unterschiede beim Vergleich dieser Werte zwischen den Szenarien könnten darauf hindeuten, dass es Wiederholungseffekte (technisch oder nichttechnisch bedingt) oder sonstige Einflussfaktoren geben könnte, die zu entsprechenden Fluktuationen führen – oder im schlimmsten Fall, dass die Probanden einfach unmotiviert irgendetwas ggf. auch zufällig bewerten.
- Zeigen sich in der Bewertungsverteilung charakteristische Muster? Sind eventuell bestimmte Nutzertypen erkennbar, z. B. solche, die besonders stark oder besonders schwach auf QoS-Änderungen reagieren?

Wiederholungseffekte: Generell sind Abhängigkeiten zwischen der Variable Preset und den übrigen Variablen der Aufgabenbearbeitung möglich, die die

abhängigen Variablen zur Zufriedenheit beeinflussen, womit Wiederholungseffekte denkbar sind. Bei einem Teil der Szenarien ist ein Unterschied zwischen den Ausführungen durch Reihenfolge und Variante technisch begründet und nicht vermeidbar. Entsprechend ist zu erwarten, dass für diese Szenarien selbst eine wertgleiche technische Parametrisierung durch die veränderten objektiven Eigenschaften zu einem anderen Erlebnis und damit anderer Bewertung führen kann. Deshalb sind für den Wiederholungsvergleich eine dienstspezifische Auswertung vorzusehen und mögliche Verschiebungen zu beachten.

Bewertungskonformität: Können die Ausführungen als prinzipiell vergleichbar angesehen werden, dann lassen sich Aussagen zur Konformität der Bewertungen jedes einzelnen Probanden treffen. Der Proband bzw. seine Bewertungsänderungen werden dazu mit der Preset-Änderung abgeglichen. [305] sieht für die Aspekte der Probanden den Begriff der Selbstkonsistenz (self-consistency) im Intra-Rater-Vergleich vor. Die Konformität hinsichtlich des angenommenen monotonen QoS-QoE-Zusammenhangs ist weitgehend erfüllt, wenn für den intrapersonellen Unterschied zwischen den Ausführungen einer Aufgabe gilt:

- bessere QoS-Werte führen nicht zu schlechteren QoE-Werten
- schlechtere QoS-Werte führen nicht zu besseren QoE-Werten
- gleiche QoS-Werte führen zu (annähernd) gleichen QoE-Werten
- die relative QoE-Änderung zur QoS-Änderung sollte ungefähr vergleichbar sein zu den Ergebnissen der anderen Probanden

Berechnung des Intra-Rater-Vergleichs: Für die Auswertung ist es sinnvoll, die QoS- und QoE-Veränderungen als Differenzen zwischen den Ausführungen je Dienst und Proband zu betrachten:

- QoS: Differenz der QoS-Netzparameterstufen (1 bis 5) aus den Presets
- QoE: Differenz aus den Zufriedenheitsbewertungen (Score 1 bis 5 als Stufe)

Die Differenz ergibt sich jeweils aus (Stufe von Ausführung 2) - (Stufe von Ausführung 1). Negative Differenzwerte bedeuten somit einen höheren Wert in der ersten Ausführung als in der zweiten Ausführung, positive Werte ent-

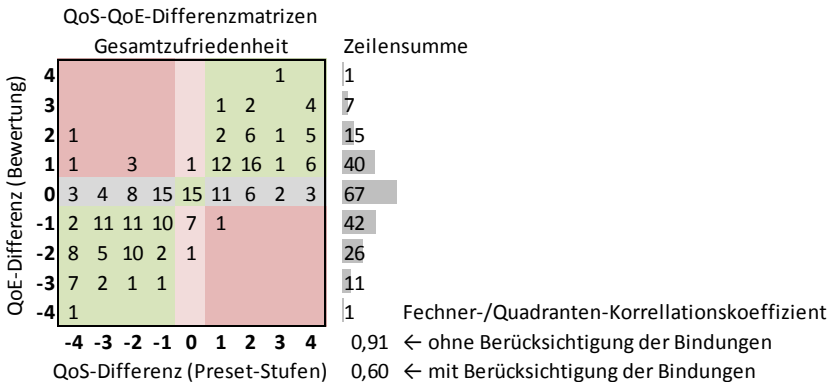
sprechend umgekehrt. Gesamtzufriedenheit und Teilbewertungen werden jeweils getrennt betrachtet. Der Intra-Rater-Vergleich entspricht damit der Bestimmung der Konkordanz als Maß der Übereinstimmung der Änderung. Eine brauchbare Definition zur Terminologie „konkordant“ liefert [298] S. 120. Entsprechend der dort üblichen Terminologie sollen QoS-wertgleiche Wiederholungen oder QoE-wertgleiche Bewertungen als Bindungen bezeichnet werden. Ungeachtet möglicher (feiner) Unterschiede zwischen den Szenarien und den subjektiven Zufriedenheitsniveaus gelten grundsätzlich für den Vergleich zwischen den Ausführungen die Annahmen zur konkordanten Änderung hinsichtlich der Richtung entsprechend der Tabelle 48.

Tabelle 48: Erwartungskonformität der Bewertungsänderung beim Wiederholungsvergleich

Veränderung Bewertungsstufe (QoE)	Veränderung Netzparameterstufe (QoS)		
	niedriger	gleich (Bindung)	höher
höher	unerwartet, gegen den Trend, ggf. zu prüfen	unerwartet/ungenau	OK, konform zum Trend
gleich (Bindung)	ggf. noch OK (Bodeneffekt)	OK, konform (konstant/genau)	ggf. noch OK (Deckeneffekt)
niedriger	OK, konform zum Trend	unerwartet/ungenau	unerwartet, gegen den Trend, ggf. zu prüfen

8.5.3 Zur Visualisierung und Interpretation des Intra-Rater-Vergleichs

Für die sinnvolle Darstellung des Wiederholungsvergleichs wurde eine aggregierte Darstellung gewählt. In einer Kreuztabelle aus Stufenänderungen für Netzparameter und Bewertung werden die entsprechenden Häufigkeiten erfasst. Da die Presets und Bewertungen jeweils 5-stufig sind, ergibt sich eine 9x9-QoS-QoE-Differenzmatrix (jeweils von -4 über 0 bis +4), siehe Abbildung 73. Mit entsprechender Hervorhebung der jeweiligen Bereiche lassen sich schnell die relevanten Aussagen ablesen. Die erwartungskonformen konkordanten Einträge befinden sich in den Quadranten 1 und 3, die stark diskordanten Abweichungen in den Quadranten 2 und 4. Durch die Zeilensumme lässt sich ablesen, wie stark die Probanden überhaupt auf QoS-Änderungen reagieren.



QoS- und QoE-Differenz, je Proband aus zwei verbundenen Ausführungen
absolute Häufigkeit, 210 Probanden mit insgesamt 420 Aufgabenausführungen

Legende zu den Bereichen der Matrix

Quadrant 1	QoE-Änderung erwartungskonform zur QoS-Änderung
Quadrant 3	Bindungen, Wertgleichheiten
Nullpunkt	QoS-QoE-Bindung, QoS und QoE gleich (genau)
x-Achse	QoS-Änderung, aber QoE gleich (neutral, noch konform)
y-Achse	QoS gleich, aber QoE-Änderung (ungenau)
Quadrant 2	QoE-Änderung nicht erwartungskonform zur QoS-Änderung
Quadrant 4	

Abbildung 73: QoS-QoE-Differenzmatrix am Beispiel der Gesamtzufriedenheit für Drive

Zur Interpretation des Beispiels in Abbildung 73: Einige einzelne Ausreißer bestehen, die Mehrzahl der Bewertungen ist aber erwartungskonform. Beim direkten wertgleichen Vergleich bewerten die Probanden in ca. zwei von drei Fällen exakt gleich, und den Rest, mit einer Ausnahme, nur eine Stufe höher oder niedriger, also insgesamt relativ genau.

In der Darstellung zur Auswertung des Testparcours werden jeweils die Gesamtzufriedenheit und zwei bzw. drei Teilbewertungen zusammen angegeben.

Statistische Beschreibung: Durch den paarweisen Vergleich kann ein vereinfachter Fechner-Rangkorrelationskoeffizient berechnet werden, siehe z. B. [298] S. 115ff. In der Darstellung als QoS-QoE-Differenzmatrix entspricht

dieser einer Quadrantenkorrelation der Differenzen und wird als statistische Kenngröße mit angegeben:

- ohne Berücksichtigung der Bindungen: Die Wertgleichheiten (QoE-Differenz 0 und QoS-Differenz 0) bleiben unberücksichtigt.
- mit Berücksichtigung der Bindungen: Die Werte der Achsen und des Ursprungs (0, 0) werden auf die umliegenden Quadranten verteilt.

Der Fechner-Koeffizient ergibt einen normierten Wertebereich $[-1, 1]$. Ausgehend vom QoS-QoE-Zusammenhang wird in den eigenen Ergebnissen generell ein positiver Fechner-Koeffizient erwartet, also dass die Bewertungen konkordant sind und überwiegend dem erwarteten Trend folgen. Hohe Fechner-Koeffizienten bei Gesamtbewertung und relevanten Teilbewertungen sprechen dafür, dass die Probanden die QoS-Veränderungen überwiegend in erwarteter Richtung in QoE-Veränderungen umsetzen. Ein niedriger Koeffizient (vor allem bei Teilbewertungen) weist hingegen auf ein Kriterium hin, welches die Probanden wenig veränderlich bewerten.

8.5.4 Allgemeine Erkenntnisse aus dem Intra-Rater-Vergleich

Es sind insgesamt recht viele QoE-wertgleiche Bewertungen zu verzeichnen, die sich zumindest zum Teil mit Boden- bzw. Deckeneffekten erklären lassen. Von einem Bodeneffekt ist auszugehen, wenn ein Proband bereits mit einer anderen als der niedrigsten Netzparameterstufe „vollkommen unzufrieden“ ist. Dann kann er in seiner Bewertung nicht noch unzufriedener werden, auch wenn die QoS-Stufe sinkt. Von einem Deckeneffekt ist auszugehen, wenn ein Proband bereits mit einer anderen als der höchsten Netzparameterstufe „vollkommen zufrieden“ ist. Dann kann er in seiner Bewertung nicht noch zufriedener werden, auch wenn die QoS-Stufe steigt. Für alle anderen Fälle unveränderter Bewertungen trotz niedriger oder höherer Netzparametrisierung sind die Unterschiede in den Stimuli für den Probanden ggf. noch nicht hoch genug, um eine Bewertungsänderung auszulösen. Dies kann einen Hinweis darauf geben, wie groß die Unterschiede überhaupt sein müssen, um die Wahrnehmung oder die Bewertung der Zufriedenheit in die eine oder andere Richtung verschieben zu können.

Über das Zufriedenheitsniveau selbst oder den absoluten Wert der Stärke der Bewertungsänderungen je QoS-Änderung konnten vorab nur Vermutungen

aufgestellt werden. Dies hängt sowohl vom Ausgangsniveau der technischen Parametrisierung als auch von der zuvor erreichten Zufriedenheitsbewertung und weiteren spezifischen Einflussfaktoren ab (Wiederholung). Es ist dabei auch für die intrapersonellen Änderungen mit nichtlinearen Zusammenhängen und o. g. Boden- und Deckeneffekten zu rechnen. Für Erkenntnisse dazu ist die gewählte Darstellung allerdings nicht geeignet, da sie unterschiedliche Stufenübergänge zusammenfasst. Für die Quadrantenkorrelation wird nur die Richtung als relative Lage zum Zentrum (0, 0) betrachtet, also der Quadrant. Die Lage innerhalb der Quadranten bleibt unberücksichtigt. Im Vergleich zu klassischen Zusammenhangs- bzw. Korrelationsmaßen wie Kovarianz-, Kontingenz- oder Korrelationskoeffizient können so aber auch nichtlineare und insbesondere sprunghafte Zusammenhänge korrekt erfasst werden.

Zu beachten ist, dass im vorliegenden Intra-Rater-Vergleich nicht „der eine“ QoS-QoE-Zusammenhang beschrieben wird, sondern „n“ individuell viele einzelne Zusammenhänge (mit n als der Anzahl Probanden).

Bei individuell stark diskordanten Abweichungen ist zu vermuten, dass ein Ausreißer in zumindest einer der Bewertungen der Aufgabe vorliegt, was auf einen ungeplant starken zufälligen externen Einflussfaktor oder einen Fehler des Probanden bei der Aufgabenbearbeitung hinweist. Diese Bewertungen sind dann oft nicht valid, weil bedingt durch die Abweichungen nicht die beabsichtigte Parametrisierung bewertet wurde.

8.6 Bedingte Verteilungen der Zufriedenheit

Die bedingten Verteilungen beschreiben die ermittelten Einzelbewertungen der Probanden zur Zufriedenheit als relevantes QoE-Maß im Test unter jeweils vergleichbaren (technischen) Bedingungen. Dabei sind die vorherigen Ausführungen zur Bedingungsgleichheit zu beachten. Die Auswertung der bedingten Verteilungen soll die Quantifizierung der Anforderungen an die Erlebnisfaktoren (speziell die QoS-Netzparameter) als auch qualitative Aussagen (speziell zu den Nebenfaktoren) ermöglichen. Daraus ergeben sich die gesuchten QoS-QoE-Zusammenhänge.

Bei der eigenen Auswertung und Ergebnisdarstellung der subjektiven Nutzerbewertungen soll zur Orientierung dienen:

- Was lässt die Statistik generell zu?
- Was ist für subjektive Nutzerbewertungen im Bereich von Studien zur Quality of Experience bzw. für den Zusammenhang von QoS und QoE üblich?
- Wie sehen allgemeine Ansätze zur Auswertung von Befragungen und Studien aus, z. B. aus dem Bereich der User Experience / Usability und auch aus anderen Bereichen wie der Marktforschung, die sich mit Stichproben von Probanden und Nutzerbewertungen beschäftigen (vor allem mit Blick auf die Skala etc.)?

8.6.1 Ausgangspunkt der statistischen Beschreibung und Analyse

Innerhalb einer Aufgabenausführung werden ausgewählte mobile Nutzungssituationen zusammengebracht. Die einzelnen Aufgabenausführungen bilden daher die betrachteten **statistischen Einheiten**.

Ausgehend vom Testdesign mit der Möglichkeit einer Aufgabenwiederholung mit gleichem Preset (QoS-Bindung) ist es sinnvoll zu betrachten, ob zwischen der Anzahl Bewertungen und der Anzahl Bewerter (Probanden) große Diskrepanzen auftreten. Die genauen Werte zur Anzahl Bewertungen im Vergleich zur Anzahl Probanden sind in Tabelle 73 im Anhang aufgeschlüsselt. Basierend auf diesen Daten ist der Unterschied als gering anzusetzen, so dass in der eigenen Auswertung als **statistische Masse** (nachfolgend **n**) die Anzahl Bewertungen angesetzt wird.

Der Umfang der Grundgesamtheit (N) kann so groß angenommen werden, dass etwaige Korrekturfaktoren zur Stichprobenentnahme grundsätzlich entfallen können. Die einzelnen statistischen Einheiten werden damit als stochastisch unabhängig festgesetzt.

Es liegen mehrdimensionale empirische Verteilungen vor. Aus den Modellannahmen folgt, dass eine Abhängigkeit der Variablen der Zufriedenheit von der Variable Preset und weiteren Faktoren existiert. Es interessieren primär die **bedingten Verteilungen** der Zufriedenheiten (Gesamt- und Teilbewertungen)

zum Preset als primäre Bedingungsvariable, ggf. unterschieden nach den weiteren relevanten Nebenfaktoren (den Partitionen).

Als Ausgangspunkt der statistischen Beschreibung ist anzunehmen, dass die genutzte Skala mindestens als Ordinalskala genutzt werden kann und kein statistisches Modell bzw. Annahmen zur Erklärung der ermittelten Verteilungen vorliegen.

8.6.2 Feinheiten und Grenzen der Statistik in der Auswertung

Die **deskriptive** Statistik mit Beschreibung der erzielten Ergebnisse in der Stichprobe ist der klassische Ausgangspunkt der Auswertung von Studienergebnissen und auch bei Nutzertests üblich. Dafür werden die genutzten statistischen Maßzahlen und Verfahren eingeführt und ggf. ihr jeweiliger Einsatz begründet. Grundsätzlich interessieren aber nicht nur die Ergebnisse in der Stichprobe, sondern auch Aussagen zur Grundgesamtheit ausgehend von der Stichprobe. Der sinnvolle und zielgerichtete Einsatz der **analytischen** Statistik in der Auswertung der eigenen Daten, mit ggf. zu beachtenden Einschränkungen, wird erläutert. Es werden, sofern miteinander vereinbar, deskriptiver Ansatz und analytische Erweiterung der Auswertung zusammen vorgestellt, um Dopplungen und Sprünge zu vermeiden. Auf Basis der generierten Maße und Visualisierungen erfolgte eine **explorative** Analyse zu den vermuteten und möglichen weiteren Partitionen in den Daten.

Es gibt zwei kritische Punkte in der Auswertung der Daten, für die weitere Überlegungen notwendig sind, die über den o. g. Ausgangspunkt einer einfachen deskriptiven Statistik hinausgehen. Dies sind Annahmen zur Skala und Annahmen zum statistischen Modell (mit Wahrscheinlichkeitsverteilung, z. B. der Zufriedenheit). In der Recherche zur Aneignung des notwendigen Wissens zum Verständnis der Statistik in der eigenen Auswertung zeigte sich, dass für fast jeden der nachfolgend ausgearbeiteten Punkte unterschiedliche Auffassungen zu finden sind. Die Spanne reicht dabei von statistischen Puristen, die Nutzerbewertungen nicht einmal als Messungen anerkennen, bis zu Praktikern, die ohne Bedenken und Überprüfung der Vorbedingungen jedes nur erdenkliche statistische Verfahren anwenden. Für die eigene Auswertung wurde ein pragmatischer Ansatz gewählt: vorsichtig, aber nicht übertrieben konservativ. Je nachdem, welche Annahmen akzeptiert werden, ist eine mehr oder

weniger detaillierte Auswertung möglich, siehe Tabelle 49. Der geneigte Leser möge selbst entscheiden, welche der Annahmen er auf Basis der gelieferten Begründungen und des eigenen Wissens zur Statistik für plausibel hält und die anderen Ausführungen und Darstellungen geflissentlich ignorieren.

Tabelle 49: Notwendige Annahmen und Möglichkeiten zur statistischen Beschreibung und Analyse

statistisches Modell und Wahrscheinlichkeitsverteilung	Skala	
	qualitativ <i>Nutzung nur als Ordinalskala</i>	quantitativ <i>Nutzung auch als metrische Intervallskala (mit Rangkennziffern als Scores)</i>
<i>ohne Modell (Beschreibung der Stichprobe)</i>	statistische Maßzahlen zur Beschreibung Anteilswerte, Median, Modus, Spanne (rangbasierte Maßzahlen)	wie links + Mittelwert (als Beziehungszahl), Varianz bzw. Standardabweichung
<i>mit Modell (Analyse zur Grundgesamtheit)</i> ohne Annahme zur Verteilung	wie oben + Rangverfahren und nichtparametrische Testverfahren, z. B. U-Test verteilungsfreie statistische Toleranzgrenzen	wie oben und links + Verfahren auf metrischem Messniveau
mit Annahme zur Verteilung	wie oben + statistische Parameter zur Charakterisierung der Verteilung Punkt-/Bereichsschätzung der Anteilswerte (Konfidenzintervall)	wie oben und links + Punkt-/Bereichsschätzung des Mittelwerts (Konfidenzintervall) ggf. bei Erfüllung einer speziellen Verteilung, z. B. Normalverteilung: parametrische Testverfahren, z. B. t-Test Varianzanalyse

8.6.3 Skalen als Zahlen: Skala und Annahmen zu den Eigenschaften der Bewertungen

Die Eigenschaften der genutzten Skala beeinflussen maßgeblich, welche Auswertungen sinnvoll sind und als solide gelten können, bzw. an welcher Stelle besser genau zu prüfen ist, ob eine bestimmte Auslegung noch erlaubt oder vertretbar ist. Die Diskussion mag etwas trocken erscheinen, ist allerdings essenziell, um die grundsätzliche Problematik zu vermitteln, welche Randbedingungen zu beachten sind, wenn die *harte* Ingenieurstechnik der QoS-Werte in Nutzertests auf den *weichen* Faktor Mensch trifft.

Qualitativ vs. quantitativ – Ordinalskala oder Intervallskala: Die eingesetzten Skalen für Bewertungen von Probanden können mindestens mit dem Messniveau einer Ordinalskala angesetzt werden, da sich die einzelnen Ausprägungen in eine natürliche Ordnung bringen lassen (siehe [306]). Die gängige Statistikk-literatur listet neben den Eigenschaften der verschiedenen Skalentypen (siehe z. B. [298] S. 12ff) auf, welche statistischen Maßzahlen und Verfahren für eine qualitative Ordinalskala prinzipiell erlaubt sind (siehe z. B. [298] S. 55ff.). Dies sind Median und Modus für die Lage sowie Spanne und Quartilsabstand für die Verteilung.

Der Median ist unproblematisch aber auch recht eingeschränkt hinsichtlich seiner Aussagekraft. Er gibt nur den typischen Wert der festgestellten Ausprägungen wieder. Es fehlt an Auflösungsvermögen, im Vergleich der Mediane ist es schwer, einen Trend zu verdeutlichen. Auf einer 5-stufigen Skala, wie der genutzten, ist der Median deshalb nicht zielführend, sofern nicht von großen Unterschieden auszugehen ist, die den typischen Wert gleich um eine ganze Stufe verschieben. Umgedreht wiederum kann der Median durch lediglich einen weiteren Messwert gleich um eine ganze Stufe springen.

Angestrebt werden die Vorteile einer quantitativen, metrischen Intervallskala, die zusätzlich weitere statistische Maßzahlen (Mittelwert und Standardabweichung) zulässt und Voraussetzung für viele statistische Verfahren ist. Über den arithmetischen Mittelwert können auch Werte zwischen den definierten Stufen errechnet werden, die eine feinere Unterscheidung bzw. zumindest Trends erkennen lassen. Dazu muss der konkrete Wert selbst nicht rückinterpretiert wer-

den. Der Mittelwert ist dabei nicht zwangsläufig als Lageparameter einer Verteilung zu verstehen, sondern als Verhältniszahl bzw. als gewogener Durchschnitt der Häufigkeit der ermittelten Scores aus den Rangkennziffern, siehe Transformation.

Die zentrale Frage dabei ist, ob die genutzte Skala auch als metrische Skala interpretiert werden darf bzw. ob eine Likert-Skala nicht ohnehin als metrische Skala angelegt ist. Das Problem der Skala und Skalentransformation ist sowohl allgemein bekannt, insbesondere bei der Auswertung von Befragungen oder Nutzertests, als auch bei QoE-Evaluationen. Entsprechend gibt es dazu zahlreiche Literatur, die allerdings gegensätzliche Meinungen aufzeigen. Diese spalten sich in eine kategorische Kontra-Fraktion und eine (vorsichtige) Pro-Fraktion. [307] gibt dazu einen allgemeinen Überblick und verweist auf weitere Arbeiten.

Kontra: Von der Kontra-Fraktion wird als prinzipieller Mangel des Skalenniveaus die fehlende Äquidistanz der einzelnen Skaleneinheiten einer Ordinalskala bemängelt ([308] S. 71f). In die gleiche Kerbe schlägt [306]. Zudem wird bemängelt, dass viele Arbeiten ohne Berücksichtigung der Eigenschaften der Skala bedenkenlos metrische Maßzahlen und Verfahren nutzen. Die Diskussion in [82] führt dahin, dass es im Umfeld der QoE-Messung besser ist, die Überlegungen zur Skala in der Auswertung zu vermeiden und gleich auf eine metrische Skala bei der Erfassung zu setzen, die sowohl eine quantitative Auswertung als auch eine Transformation auf das Niveau einer Ordinalskala erlaubt. Aus theoretischer Sicht ist dies sicherlich richtig, verlagert aber das Problem von der Auswertung zurück auf die Messung bzw. Bewertung. So wird die ohnehin notwendige Interpretation nur verlagert. In [82] wird ausgeführt, wie eine solche quantitative Skala beschaffen sein müsste. ACR-Bewertungsskalen (Absolute Category Rating) mit beschrifteten Skaleneinheiten erfüllen demnach diese Anforderungen nicht. Auch die Mean-Opinion-Score-Skala (MOS) kann damit eigentlich nur als Ordinalskala gewertet werden.

Pro: Die gängige Praxis der QoE-Messung und Auswertung ist allerdings eine andere. Im Allgemeinen wird die 5-Punkte-MOS-Skala als Intervallskala akzeptiert und als solche ausgewertet. Der Name der Skala impliziert es ohnehin

schon, dass eine Mittelwertbildung zur Auswertung angestrebt wird. Die Pro-Fraktion argumentiert, dass bei geeigneter Formulierung und Präsentation der Skala im Fragebogen die Annahme einer Intervallskalierung durchaus möglich ist. Für Likert-Skalen trifft dies bei geschickter Konstruktion zu. [309] S. 88 verweist in diesem Zusammenhang auf eine entsprechende Arbeit ([310] S. 176f), dass „Likert-Skalen hinreichend gut den Anforderungen einer Intervallskalierung [...] entsprechen“ [309] S. 88, dort allerdings um quantitative Angaben in der Beschriftung ergänzt. [311] S. 10 geht mit Verweis auf [312] sogar auf von einer expliziten Definition der Likert-Skalen als Intervallskala aus, was die Skala demnach sogar empfiehlt, wenn quantitative Auswertungsverfahren eingesetzt werden sollen. Mit Bezug auf die zur QoE-Evaluation oft genutzten MOS- bzw. Likert-Skalen verweist [64] S. 95 auf [313]. Demnach sind Likert-Skalen nicht nur unproblematisch, sondern es ist auch die Anwendung parametrischer Verfahren möglich, obwohl die angenommenen Voraussetzungen (z. B. Normalverteilung, Stetigkeit etc.) auf solchen Skalen üblicherweise nicht erfüllt werden, da die Verfahren auch so robust genug sind [313].

In der strikten Theorie gehen die Ansichten also auseinander. Damit ist die die Anwendung als Intervallskala genau so richtig wie auch falsch – je nach Nutzung. Als pragmatischer Ansatz wird in [314] die Anwendung statistischer Methoden und Maßzahlen empfohlen, die die Reichhaltigkeit der Struktur der Antworten entsprechend widerspiegeln und im Zweifel ohne Überlegungen zur Interpretation der Skala auskommen. Bei Abschätzung der Vorteile und Nachteile im konkreten Fall ist oftmals die Auswertung als metrische Skala übliche Praxis. Man geht bei einer nur 5-stufigen Skala von ungefähre Gleichabständigkeit aus und akzeptiert so auch ggf. eine nachträgliche Aufwertung der Skala als zulässig. Für die eigene Auswertung bedeutet dies: Die genutzte Skala ist nicht weniger gleichabständig als die gängige 5-stufige MOS-Skala. Die Voraussetzungen zur Präsentation und Interpretation mit gleichen Abständen zwischen den Skalenpunkten sind grundsätzlich erfüllt. Bei den genutzten fünf Stufen wären sogar gewisse Abweichungen in der Interpretation durch die Probanden zu tolerieren, da diese nicht zu großen Verzerrungen führen können.

Daher ist es plausibel, ein zumindest annähernd metrisches Skalenniveau für die Zufriedenheitsskala anzunehmen und eine entsprechende Transformation zu erlauben. Aussagen auf Basis der Annahme eines metrischen Skalenniveaus werden nicht alleinig, sondern nur ergänzend zu anderen Aussagen angegeben. Für komplexere statistische Verfahren (hier nicht genutzt) ist ggf. gesondert zu prüfen, ob das genäherte metrische Niveau ausreichend ist und ggf. weitere Voraussetzungen zur Verteilung erfüllt sind.

Transformation der Skala: Für die Nutzung als Intervallskala in der Auswertung werden der 5-stufigen Bewertungsskala feste diskrete numerische Werte im Intervall $[1, 5]$ zugeordnet, siehe Abbildung 74. Die Werte, nachfolgend Scores genannt, entsprechen den Rangplätzen der 5-stufigen Skala in einem Punktesystem. Ein höherer Zahlenwert repräsentiert eine bessere Bewertung mit „mehr Zufriedenheit“. Daher sind die Scores intuitiv besser geeignet als das (dt.) Schulnotensystem mit der „1“ als bester Note. Für die visuelle Kodierung der Bewertungen der Zufriedenheit wird der Skala eine erweiterte Ampelfarbskala von dunkelgrün bis rot zugeordnet. Die Farben werden fortan mit dieser Bedeutung genutzt.

vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
5	4	3	2	1

Abbildung 74: 5-stufige Skala als Bewertungsskala für die Zufriedenheit mit zugeordneten Scores und Kennfarben

Eine Interpretation von berechneten Zwischenwerten sollte mit der notwendigen Vorsicht hinsichtlich der damit verbundenen Aussage erfolgen. Prinzipiell ist eine Rücktransformation mittels Rundung zu den definierten festen diskreten Werten möglich aber auf den Sinn der Aussage zu prüfen. Es ist eventuell ausreichend die Zwischenwerte zu interpretieren als „zwischen 4 und 5 mit Tendenz zu eher zufrieden“. Mit Blick auf den MOS und R-Values wird mit den Schwellwerten noch eine andere Interpretation von berechneten Zwischenwerten als Planungsskala präsentiert.

Stetigkeit: Es wird davon ausgegangen, dass die Zufriedenheit als gemessener Wert prinzipiell in unendlich vielen und beliebig feinen Abstufungen vorliegt, also quasi stetig ist. Die Messung selbst, als Bewertung durch die Probanden,

erfolgt aber nur in den o. g. diskreten Abstufungen zwischen festgelegten Minimal- und Maximalwerten. Die Probanden wählen auf der Skala den Wert, der ihrem internen Wert in ihrer Interpretation am nächsten liegt. Durch die diskreten Bewertungsstufen kommt es auf einzelnen Werten zwangsläufig zu Häufungen. Dies ist dann ein Problem, wenn die Stetigkeit der Werte gefordert ist und deshalb angenommen wird, dass es keine oder nur sehr wenige exakt wertgleiche Datensätze gibt. Dies ist zu berücksichtigen, wenn zur Verteilung der Bewertungen Annahmen getroffen oder überprüft werden, z. B. ob eine Normalverteilung vorliegt.

8.6.4 Aussagen zur Grundgesamtheit – Interferenzstatistik

Die Interferenzstatistik hat viele Namen: schließende, analytische, beurteilende, induktive und mathematische Statistik (siehe z. B. [315] S. 111). Das Prinzip der Interferenzstatistik ist der Rückschluss von der Stichprobe auf die Grundgesamtheit auf Basis eines statistischen Modells bzw. eines Wahrscheinlichkeitstheoretischen Zusammenhangs, z. B. unter Annahme bestimmter Eigenschaften ausgewählter Verteilungen. Viele der Auswertungsverfahren der Interferenzstatistik sind an bestimmte Nebenbedingungen geknüpft. Oftmals sind dies metrische Skalen, Stetigkeit, das Vorliegen bestimmter Verteilungen, z. B. Normalverteilung, die Kenntnis der Parameter der Verteilung und ein (ausreichend) großer Stichprobenumfang (n).

Durch die Beschränkung der Untersuchungen auf eine Stichprobe von Probanden statt der vollständigen Untersuchung aller Nutzer in der Grundgesamtheit ergibt sich ein **statistischer Standardfehler** der Stichprobe. Dies bedeutet, dass davon auszugehen ist, dass mit anderen Probanden jeweils ein (leicht) anderes Ergebnis zu erzielen gewesen wäre. Der Standardfehler ist damit die zu erwartende Unsicherheit in den Ergebnissen, die entsteht, wenn nur eine Stichprobe anstelle der Grundgesamtheit vermessen wird. Der gesamte Umfang der Stichprobe ist mit 2520 Aufgabenausführungen von 210 Probanden zwar schon recht ordentlich, aber für den statistischen Standardfehler zählt nur die statistische Masse pro Bedingung. Dies sind in der Studie rund $n=90$ für ein Preset ohne weitere Unterscheidung, bzw. deutlich weniger, wenn einzelne Partitionen betrachtet werden sollen. Dabei ist zu beachten, dass mit immer kleinerem Stichprobenumfang der statistische Standardfehler rapide ansteigt.

Daher ist gut abzuwägen, welche Aussagen für die Ergebnisse des Labortests möglich und nötig sind:

- Bestimmung der tatsächlichen Relevanz der vorab identifizierten weiteren Einflussfaktoren auf Leistungsseite neben den QoS-Parametern
- Bedeutung des statistischen Standardfehlers für die Festlegung der QoS-Schwellwerte

Zur weiteren Auswertung werden nachfolgend das grundlegende statistische Modell und die genutzten analytischen Methoden vorgestellt.

8.6.5 Vom Schätzen, Testen und Analysieren

Nachfolgend soll kurz erläutert werden, welche interferenzstatistischen Auswertungen prinzipiell möglich sind, und welche davon für die angestrebten Aussagen und das vorliegende Datenmaterial sinnvoll sind. In der analytischen Statistik sind drei Hauptbereiche auszumachen: Schätzen, Testen und Analysieren in Form verschiedener Varianzanalysen. Auf Testen und Analysieren soll nur kurz eingegangen werden, da diese in der eigenen Auswertung eher weniger interessieren. Das Schätzen wird aufgrund seiner Bedeutung etwas ausführlicher behandelt.

Der Sinn des **Testens** ist die Überprüfung von Hypothesen (zur Gleichheit der Grundgesamtheit der Stichproben) und ist primär für konfirmatorische Studien interessant. In der eigenen Studie hingegen ist die Abschätzung der Relevanz von Einflussgrößen wichtiger als der statistische Nachweis des signifikanten Unterschieds. Damit sind bereits zwei der wichtigsten Begriffe genannt worden (nach [316] zusammengefasst):

- signifikant = durch die Statistik zu ergründen und begründen (der Begriff als solcher ist an ein 95%-Konfidenzniveau gebunden)
- relevant = bedeutsam, „substanzwissenschaftlich“ fachlich und sachlich zu begründen

Im Fall der eigenen Studie liefern die Vorüberlegungen zu den Einflussfaktoren und zu ihrem Zusammenwirken im Modell die substanzwissenschaftliche Begründung – zumindest auf der Seite der objektiv fassbaren Größen. Eine weitergehende Erklärung des Unterschieds, der teilweise unbewusst falsch genutzten Begriffe und Ansätze bzw. Abgrenzung zum Schätzen finden sich

in [316]. Statistisch sichere (signifikante) Aussagen zu Unterschieden bzw. Gleichheiten bedürfen Tests. Für fachlich und sachlich zu begründende Aussagen zu ggf. relevanten Unterschieden und zur Abschätzung von Effektrichtung und -stärke sind Schätzungen geeignet.

Sind für mehrere Gruppen gleichzeitig Tests auszuführen, dann werden diese Verfahren als **Varianzanalyse** bezeichnet. Weitergehende Erläuterungen liefert die Standardliteratur, z. B. [317] S. 219ff und [318] S. 240ff. Da sich auch mehrere Einflussfaktoren gleichzeitig überprüfen lassen, können Verfahren wie ANOVA und Co. zusätzlich zu signifikanten Unterschieden auch noch die Einflussgrößen einzelner Faktoren quantitativ ermitteln. Mit einer solchen Steigerung der Aussagekraft gehen meist aber gleichzeitig auch höhere Anforderungen an die Voraussetzungen zur Nutzung der statistischen Verfahren einher.

Für das **Schätzen** ist der Ausgangspunkt, dass die interessierenden Größen wie Anteilswerte, Mittelwerte und Standardabweichungen in der Grundgesamtheit unbekannt sind und auf Basis der empirisch ermittelten Werte der Stichprobe geschätzt werden müssen. Eine Form des Schätzens sind Punktschätzungen für die genannten Größen als jeweils ein einzelner Wert. Bereichsschätzungen sind eigentlich immer besser als Punktschätzungen, da sie die statistische Masse einbeziehen und so auch Abschätzungen zulassen, wie genau bzw. verlässlich die mit der abgefragten Stichprobe erzielten Werte als Schätzung sind. Entsprechend sollen für die interessierenden Größen jeweils Bereichsschätzungen in Form von **Vertrauensbereichen**²⁰² angegeben werden. Vertrauensbereiche „sind aus der Stichprobe berechnete Intervalle (oder Bereiche), in denen der wahre, aber unbekannt Parameter der Verteilung der Grundgesamtheit mit großer Wahrscheinlichkeit (Sicherheit) zu erwarten ist“ ([317] S. 156).

Eine große Stichprobe wird, dem Gesetz der großen Zahlen folgend, näher an einem für die Gesamtheit aller Nutzer repräsentativen Wert liegen. Eine kleine Stichprobe wird hingegen wahrscheinlich mehr abweichen, also einen größeren Vertrauensbereich aufweisen. Angaben zum Vertrauensbereich sind daher

²⁰² auch Konfidenzintervalle genannt

nur in Verbindung mit Konfidenzniveau und Stichprobengröße vollständig. Da die statistische Masse in die Berechnung einfließt, lässt sich über die Vertrauensbereiche auch ermitteln, bis zu welchem Grad es sinnvoll ist, eine Ergebnismenge in Untermengen zu zerlegen. Für eine immer kleinere statistische Masse nimmt die Breite des Vertrauensbereiches zu.

Zur Wahl des richtigen Konfidenzniveaus in der Praxis gibt [319] einen Hinweis: 95 % sind in wissenschaftlichen Veröffentlichungen der Standard, für generelle Aussagen in der Industrie (Managemententscheidungen) sind meist 90 % schon ausreichend, für explorative Analysen teilweise sogar schon 80 %. Mit kleinerem Konfidenzniveau verkürzen sich die Konfidenzintervalle, allerdings auf Kosten der Irrtumswahrscheinlichkeit in der Aussage. Für die eigene Auswertung werden die Vertrauensbereiche mit einem Konfidenzniveau von 95 % für ausgewählte Anteilswerte und die Mittelwerte berechnet und angegeben. Zu deren Interpretation, besonders im Vergleich, sei auf [320] und [316] verwiesen.

8.6.6 Metriken in der Auswertung von Nutzertests

Es sei an die Anforderungen für eine gute Metrik erinnert (z. B. [215] S. 289f.). Demnach sollte sie korrekt und nützlich sein. Dies schließt auch eine einfache Erstellung und Berechnung in der Auswertung ein. Die Metrik sollte zuverlässig sein, es wird also tatsächlich erfasst, was als Zahl ausgedrückt werden sollte. Für die erhobenen Werte sollte eine plausible und akzeptierte Interpretation vorliegen. Ausgehend von den wichtigsten statistischen Maßzahlen werden nachfolgend die Überlegungen zu den üblichen und den genutzten Metriken dargelegt.

Eine Kurzzusammenfassung mit weiteren Verweisen zur Vertiefung aus dem Bereich der Auswertung von QoE-Studien liefert [299], was auf die Diskussion des Mean Opinion Score (MOS) als üblicher QoE-Metrik aus Mittelwerten führt (siehe [321] S. 9ff). Einen Überblick zu bewährten Ansätzen zur Auswertung von Befragungen in der praktischen Anwendung abseits von QoE-Evaluationen geben beispielsweise [322] S. 340ff und [83] S. 438ff für Umfragen und Usability-Tests sowie [323] und [324] für Ergebnisse zur Nutzer- bzw. Kundenzufriedenheit, die sich vornehmlich auf Anteilswerte beziehen. Damit sind auch die zwei populärsten Ansätze zur Auswertung von Studien

mit Nutzerbefragungen bzw. Nutzerbewertungen genannt. Gängige Maßzahlen sind Mittelwerte als Verhältniszahlen und Anteilswerte als Gliederungszahlen (Begriffe nach [302] S. 226ff), auf die nachfolgend jeweils noch detaillierter eingegangen wird.

8.6.7 Statistische Maßzahlen: Mittelwert & Co.

Lage- und Verteilungsmaße: Statistische Maßzahlen können zur Beschreibung der erzielten Ergebnisse genutzt werden und Lage und/oder Verteilung der Einzelergebnisse charakterisieren. Sie fassen die Ergebnisse in ein oder zwei Zahlen zusammen und können so den Vergleich verschiedener Bedingungen erleichtern, indem sie einen leicht zu erfassenden Abstand zur Verdeutlichung der Unterschiede definieren. Die Berücksichtigung der Struktur und Zusammensetzung der Antworten sind je nach getroffenen Grundannahmen und der damit verbundenen Berechnung unterschiedlich.

Für die eigene Auswertung kann, wie zuvor ausführlich dargelegt wurde, von einem metrischen Skalenniveau ausgegangen werden, womit der Mittelwert und weitere statistische Maßzahlen mit der Voraussetzung einer metrischen Skala nutzbar sind. Statistisch gesehen ist der arithmetische Mittelwert als Verhältniszahl eine Beziehungszahl, die „auch ohne Kenntnis der Häufigkeitsverteilung“ [302] S. 227 berechnet werden kann. Für eine analytische Betrachtung hingegen wird der Mittelwert typischerweise als Lageparameter zur Charakterisierung der angenommenen Verteilung, z. B. einer Normalverteilung, genutzt.

Anhaltspunkte für Annahmen zur Normalverteilung in Umfrageergebnissen aus der praktischen Anwendung liefert [325]. Für eine ausreichend große Zahl Beobachtung kann die Berechnung von Mittelwert und Standardabweichung auch für nicht normalverteilte Daten, wie die Bewertungen aus Nutzertests, brauchbare Ergebnisse für die Abschätzung der Erwartungswerte in der Grundgesamtheit liefern. Der zentrale Grenzwertsatz wird dafür angewandt (siehe z. B. [317] S. 69). Für genügend großen Stichprobenumfang²⁰³ ($n > 30$) kann demnach der Mittelwert der Zufriedenheitsbewertungen als normalverteilt angenommen werden. Für kleinere Stichprobenumfänge und unbekannte

²⁰³ Dies sollte auch bei der Abschätzung des Erreichens des notwendigen Stichprobenumfangs, z. B. beim Vergleich von Mittelwerten, für die Umfänge der Partitionen beachtet werden.

Varianz in der Grundgesamtheit wird zur Approximation eine t-Verteilung der Mittelwerte angenommen.

Für die eigene Auswertung ist auf Basis des statistischen Modells eine sinnvolle Annahme zur vorliegenden Verteilung gegeben. Die Bedeutung der statistischen Masse sollte bei der Zerlegung und isolierte Betrachtung einzelner Partitionen beachtet werden. Werte um $n=20$ sind grenzwertig, darunter ggf. schon recht schwierig. Neben dem Umfang ist zu berücksichtigen, dass für den Vergleich von Verhältniszahlen wie dem Mittelwert, die repräsentierten statistischen Massen „in ihrer Struktur wenigstens annähernd übereinstimmen [sollten]“ [302] S. 233. Im Fall der Charakterisierung einer Verteilung über den Mittelwert ist es gute Praxis, auch die Standardabweichung als zugehörigen Verteilungsparameter anzugeben.

Im Fall des Schätzens der wahren Werte sollte der Mittelwert nicht nur als Punktschätzung angegeben werden, sondern besser zusammen mit einer Bereichsschätzung, dem Vertrauensbereich des Mittelwerts. Dieser ist von der (unbekannten) Varianz der Grundgesamtheit, der Stichprobengröße und dem Konfidenzniveau abhängig. Das Konfidenzniveau wird in der eigenen Auswertung einheitlich auf 95 % festgelegt (woraus sich $\alpha=5$ % ergibt) und die unbekannt Varianz der Grundgesamtheit aus der empirischen Varianz der Stichprobe geschätzt. Das bedeutet aus Sicht der Statistik, dass der wahre Mittelwert der Grundgesamtheit mit einer Wahrscheinlichkeit von 95 % innerhalb der Grenzen des Vertrauensbereichs liegt. In der eigenen Auswertung erfolgt die Berechnung des Vertrauensbereichs des Mittelwerts grundsätzlich auf Basis der t-Verteilung (als Vereinfachung sowohl für $n \leq 30$ als auch für $n > 30$).

Weitere statistische Maßzahlen: Neben der Lage und Verteilung sind oftmals auch weitere statistische Maßzahlen wie **Kurtosis** und **Schiefe** einer Verteilung relevant. Aus einer starken Konzentration der Bewertungen auf einzelnen Skalenpunkten resultieren zwangsläufig Verzerrungen durch hohe Kurtosis (spitze Wölbung der Verteilung). Insbesondere die Boden- und Deckeneffekte in der Bewertung führen zu einer starken Besetzung der einzelnen Skalenenden und damit zu einer ausgeprägten Schiefe und Steilheit der Verteilung, vor allem zur Bewertung „vollkommen zufrieden“. Schiefe und Extremwertbesetzungen sind bei den Voraussetzungen für statistische Verfahren

zu beachten. Die Struktur der vorliegenden Daten erfordert eine kritische Prüfung zum Einsatz und Bewertung möglicher Ergebnisse solcher Verfahren. Für die eigene Auswertung ist dies ohne Bedeutung, da auf solche Verfahren verzichtet wird.

8.6.8 Anteilswerte, statistisches Modell und Annahmen zur Verteilung

Anteilswerte bzw. relative Häufigkeiten sind eine besondere Art Maßzahl. Sie sind als Gliederungszahlen den Verhältniszahlen untergeordnet, siehe [302] S. 226f: „die Anzahl der Elemente [...] einer Teilmasse [wird] zur Anzahl der Elemente [...] einer übergeordneten Gesamtmasse in Beziehung [ge]setzt“. Die Anteilswerte sind so ohne weitere Voraussetzungen zur deskriptiven Statistik der empirisch erzielten Ergebnisse geeignet.

Für die weiteren Auswerteschritte auf Basis der Anteilswerte sind mehrere Ansätze möglich. Teilweise werden diese Anteile einzeln genutzt, teilweise aggregiert und teilweise aufeinander bezogen (Differenzbildung oder Verhältnisbildung (Odds)), um sie in einer einzigen Zahl auszudrücken. Damit sind die Anteilswerte (einzeln oder aggregierter Ausprägungen der Zufriedenheitsbewertungen) eine Maßzahl für die übergeordnete Gesamtmasse aller Bewertungen (statistische Masse n) eines Presets (für eine bestimmte Partition).

Sofern genügend viele Ergebnisse betrachtet werden, ist davon auszugehen, dass die ermittelten Anteile der Bewertungen die wahre Verteilung in der Grundgesamtheit gut annähern. Für eine plausible analytische Statistik ist ein passendes statistisches Modell notwendig, aus dem sich auch Annahmen zu den relevanten Wahrscheinlichkeitsverteilungen ableiten lassen. Die Verteilung der Bewertungen, also die Bewertungsstruktur je Preset, bzw. über die Presets hinweg im QoS-Parameterraum, ist komplex. Die Überlegungen zu einem formal korrekten statistischen Modell müssen dabei einige wichtige Punkte aufgreifen:

- die Erklärungen zum QoE-Bewertungsprozess und der Abbildung einer eigentlich stetigen latenten Größe auf eine diskrete Skala (Zufriedenheitswaage in 2.3.9)
- die Annahmen für die ganz konkret genutzte Skala (in 4.3.1 und in 8.6.3): Ordnungskriterium der Bewertungen (mindestens Ordinalskala) und Annahmen zu den angenommenen Abständen der Skalenpunkte

Damit ergibt sich, dass eine Erklärung über ein Ordinal-Modell am plausibelsten ist (Erläuterungen in [326] S. 69ff). Eine alternative Bezeichnung dazu wäre das geordnete Multinomial-Modell. Je nach Annahmen über den Abstand zwischen den Bewertungen auf der Skala lässt sich weiter differenzieren (siehe [326] S. 83f für eine Auflistung sinnvoller Varianten). In die nähere Auswahl sinnvoller Erklärungen kommt dabei das geordnet logistische Modell (*ordered multinomial logistic* bzw. *ordinal logit model*), unter der Annahme gleicher Abstände zwischen den Skalenpunkten, weshalb es auch *proportional odds model* genannt wird [326] S. 71. Die Diskussion zur angenommenen Gleichabständigkeit der Skalenpunkte wurde bereits in 8.6.3 geführt, wobei die Argumentation dahin führte, dass dies eine mögliche und plausible Annahme ist.

Andere Beschreibungsansätze, z. B. eine normale Multinomialverteilung, würden nur von Kategorien zur Auswahl bzw. Bewertung ausgehen, was aber weder der Skala noch einer stimmigen Erklärung zum Zustandekommen der Bewertungen entsprechen würde. Zudem kann nicht von einer Unabhängigkeit irrelevanter (Antwort-) Alternativen ausgegangen werden – die Bewertungen werden eben nicht einfach „erwürfelt“.

Über Verteilungsfunktionen des statistischen Modells sollte zu beschreiben sein, wie sich die latente tatsächliche Größe (Zufriedenheit als stetige Größe) aus den theoretischen Überlegungen heraus über den Parameterraum verteilt, wie also der eigentliche stochastische Zusammenhang zwischen QoS und QoE aussieht. Durch die Studie wurden die Anteile der Bewertungen in der angenommenen Grundgesamtheit genähert ermittelt. Entsprechend interessiert, wie gut diese Annäherung ist, vor allem hinsichtlich der Nutzung der Anteilswerte für die Bestimmung der gesuchten Schwellen. Daher sollen in der Auswertung zu den empirisch ermittelten Anteilswerten auch die zugehörigen Vertrauensbereiche der Anteilswerte berechnet und angegeben werden. Die Grundidee dabei ist die Bildung gerichtet aggregierter dichotomer Teilmengen, was bei einer 5-Stufen-Skala vier Zerlegungen ergibt, siehe Abbildung 75 zur Anwendung auf der eigenen Skala. Es ist jeweils die kumulative Wahrscheinlichkeit gegen alle niedrigeren bzw. alle höheren Skalenpunkte zu bilden. Es ist zu beachten, dass die Anteile jeweils nur komplementär im jeweiligen Schnitt sind, nicht aber zwischen den Schnitten. Für die einzelnen

Schnitte erfolgt die Auswertung wie für dichotome Ergebnismengen, z. B. zur Ermittlung der Konfidenzintervalle und Regression.

Der Umfang der Stichprobe ist für die Berechnung der Vertrauensbereiche der Anteilswerte von entscheidender Bedeutung. Die Praxis der Auswertung von kleineren Umfragen, Studien und Nutzertests zeigt, dass es wiederum die Nebenbedingungen wie minimale und maximale Anteilswerte und der Stichprobenumfang sind, die dabei über die Wahl der richtigen Berechnungsmethode entscheiden. [327] vergleicht die Methoden für *kleine* Stichprobenumfänge und dichotome Merkmale. Dabei zeigt sich, dass nicht alle Berechnungsverfahren geeignet sind. Vor allem das gängige einfache Wald-Verfahren schätzt die Vertrauensbereiche zu klein und das exakte Verfahren nach Clopper-Pearson zu konservativ. Bei einfachen Verfahren ergeben sich je nach Anteilswert erst ab einem Umfang von $n=20$ bis $n=100$ je Teilstichprobe sinnvolle Ergebnisse. Dies ist für stark von 50 % abweichende Anteile in den eigenen Daten nicht immer gewährleistet. Die eigenen Berechnungen erfolgen daher mittels angepasster Wald-Methode nach [327] S. 2101. Diese liefert brauchbare Ergebnisse auch für kleinere n und extreme Anteilswerte, wie sie bei Nutzerstudien allgemein üblich sind und auch in der eigenen Studie auftreten.

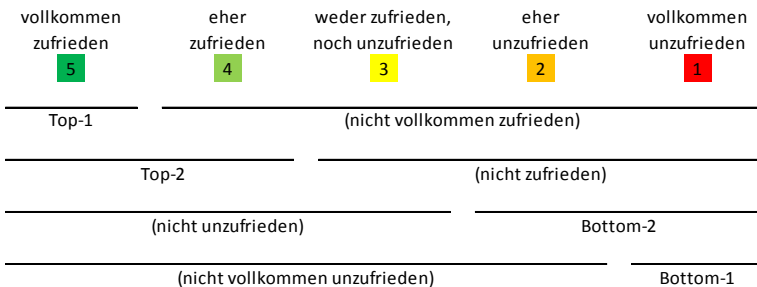


Abbildung 75: 5-stufige Skala und die vier gebildeten gerichteten dichotomen Aggregationen

In der Auswertung interessiert möglicherweise ohnehin mehr die grundsätzliche Tendenz der Bewertungen als die genaue Ausprägung. Genau dafür eignen sich die aggregierten Anteilswerte. Im englischen Sprachgebrauch werden dabei die einzelnen Bewertungsoptionen und damit auch ihre Anteile als „Box“ bezeichnet. Die Top-Box (kurz Top-1) entspricht der bestmöglichen Bewertung, die Bottom-Box (kurz Bottom-1) der schlechtesten Bewertung, Top-2 und Bottom-2 werden in äquivalenter Art bezeichnet, siehe Abbildung 75. Es interessieren im Weiteren davon besonders Top-1, Top-2 und Bottom-2. Dabei ist der Vertrauensbereich allerdings nur in jeweils eine Richtung von Bedeutung: für Top-1 und Top-2 jeweils die Untergrenze (also der wahrscheinlich nicht unterschrittene Mindestwert), für Bottom-2 die Obergrenze (also der wahrscheinlich nicht überschrittene Maximalwert). Die Vertrauensbereiche werden in der Auswertung daher nur für die genannten (aggregierten) Anteilswerte einseitig für ein Konfidenzniveau von 95 % berechnet und angegeben.

8.6.9 QoE-Metriken: Mean Opinion Score (MOS)

Der Quasi-Standard als bekannteste und häufig genutzte QoE-Metrik ist der Mean Opinion Score (MOS). An dieser Stelle soll der MOS als Maßzahl auf Basis der vordefinierten ITU-Skala gesehen werden (siehe [202] zur Skala und Nutzung als Metrik). Der MOS wird als arithmetisches Mittel der zugeordnete Scores, üblich von 1 bis 5 (siehe 4.1.6 zur Skala) aller Einzelbewertungen, den (Individual) Opinion Scores, gebildet. Dafür gelten die entsprechenden Ausführungen zum Mittelwert aus dem vorherigen Abschnitt. Dabei wird davon ausgegangen, dass die Einzelbewertungen stark individuell geprägt sind, der MOS als durchschnittliche Wertung aber den Einfluss der interessierenden technischen Einflussgrößen abbildet [299] S. 131.

Grundsätzliche Kritik zum MOS: In der Arbeit mit dem schönen Titel „SOS – The MOS is not enough“ [299] resümieren die Autoren nach Sichtung verschiedener Veröffentlichungen zu QoE-Nutzertests die gängige Praxis der statistischen Auswertung. Demnach ist dies der MOS als Durchschnittsmaß, teilweise ergänzt um die Standardabweichung zur genäherten Vermittlung der Divergenz der Nutzerbewertungen und zum Teil Vertrauensbereiche für die Mittelwerte.

Dies deckt sich mit den Empfehlungen der ITU zur statistischen Auswertung von „MOS-Studien“ (nach [321] S. 9ff):

- Angabe der Anzahl Schätzer (Probanden)
- Berechnung und Angabe des Mittelwerts (MOS)
- Berechnung und Angabe des 95%-Vertrauensbereichs des Mittelwerts
- Berechnung der Standardabweichung für die weitere interne Verwendung, aber ohne Angabe in den Ergebnissen
- Es wird für die Auswertung angenommen, dass für die einzelnen Scores geeignete Verteilungen vorliegen, eine gesonderte Prüfung der Voraussetzungen wird nicht verlangt.
- Ab 20 Probanden sind brauchbare Ergebnisse zu erwarten.

Aus Sicht der Statistik ist insbesondere zu bemängeln, dass die Standardabweichung als Verteilungsparameter bzw. Streuungsmaß nicht zur Veröffentlichung empfohlen wird. In anderen wissenschaftlichen Gebieten ist über diese Praxis bereits eine Diskussion ausgebrochen, eben weil Standardabweichung und Vertrauensbereich mit dem Standardfehler verschiedene Aussagen generieren, siehe [328]. Die Variabilität der Antworten in der Stichprobe lässt sich unmittelbar nur in der Standardabweichung ablesen, wohingegen bei Betrachtung des Vertrauensbereichs noch immer das Konfidenzniveau und der Stichprobenumfang zu berücksichtigen sind.

Generell ist zu den Mittelwerten bzw. zum MOS zu beachten: In der eigenen Studie wird eine von der (ITU-) MOS-Skala abweichende Skala zur Zufriedenheitsmessung genutzt. Aufgrund der Skalenunterschiede soll nachfolgend begrifflich zwischen MOS und dem empirischen Mittelwert der eigenen Bewertungen getrennt werden, auch wenn beide gleich berechnet werden. Aufgrund der Unterschiede in den Skalen sind die Ergebniswerte der eigenen Studie ggf. nicht direkt mit den MOS-Werten anderer Studien vergleichbar.

8.6.10 Mittelwerte oder Anteilswerte? Beides!

Nachdem zuvor mit der Diskussion von Mittelwert und Co. sowie den (aggregierten) Anteilswerten die gängigen Ansätze zur Auswertung vorgestellt wurden, verbleibt noch die Diskussion, was denn nun *besser* ist. Zu den prinzipiellen Vorteilen und Nachteilen der genannten Ansätze in der praktischen Anwendung geben [329], [323], [330] und [324] Auskunft. Dabei lässt sich zusammenfassen, dass sich aus der Auswertung von Anteilswerten und aus Mittelwerten unterschiedliche Aussagen ableiten lassen (nach den zuvor genannten Quellen).

Mittelwerte: Mittelwerte geben als einzelner Wert teilweise die eigentliche Verteilung der Antworten nur schlecht wieder. Aus dem Mittelwert als reinem Lageparameter kann nicht geschlussfolgert werden, ob die Antworten ungefähr gleichmäßig (bzw. normal) um den ermittelten Wert herum verteilt sind, oder ob sich etwa zwei Schwerpunkte an den entgegengesetzten Enden der Skala (bimodale Verteilung) gegenseitig ausmitteln. Daher ist es gute Praxis, zum Mittelwert zumindest noch einen Verteilungsparameter, üblicherweise die Standardabweichung anzugeben. Mittelwerte eignen sich gut, um Veränderungen als Vergleich über die Zeit zu verdeutlichen (z. B. Leistungsveränderungen) oder zur Untersuchung von Unterschieden, z. B. in der eigenen Studie zwischen verschiedenen Partitionen.

Anteilswerte: Aus der Größe eines einzelnen Anteils, z. B. der Zufriedenheit, können keine Angaben zu anderen kritischen Größen wie der Unzufriedenheit oder der gesamten Verteilung gemacht werden, sofern das untersuchte Merkmal nicht dichotom ist. Es sind für sinnvolle Aussage also ggf. mehrere Werte gleichzeitig zu beachten. Für Aussagen zu einem Anteilswert kann in solchen Fällen ergänzend auch der Mittelwert herangezogen werden. Betrachtungen zum Anteilswert eignen sich für Betrachtungen zu Stärken vs. Schwächen und damit auch zum Identifizieren und Analysieren von kritischen Bereichen, worauf in der Diskussion der Schwellen zurückgekommen wird.

Beides: Oftmals ist es daher sinnvoll und lohnend, sowohl Mittelwert (mit Standardabweichung) als auch Anteilswerte anzugeben. In der Darstellung der Ergebnisse aus QoE-Evaluationen sind die oben aufgeführten Vor- und Nachteile meist leider nicht ausreichend berücksichtigt. Die Bewertungen echter

Nutzer zur Quality of Experience sind oftmals höchst subjektiv, weil durch die individuellen Erwartungen und Erfahrungen der einzelnen Nutzer geprägt. Prinzipiell ist dies mittlerweile erkannt und die Autoren von [299] argumentieren in die gleiche Richtung. Dabei kritisieren sie ausdrücklich, dass der MOS als alleiniges Maß die erzielten Ergebnisse nur unvollständig beschreibt, gerade weil für verschiedene Aspekte die Struktur der Bewertungen entscheidend sein kann. Zusammenfassende Werte, insbesondere reine Lageparameter wie der MOS, berücksichtigen dies nicht ausreichend. Dies deckt sich mit den eigenen Überlegungen. An einem konstruierten Beispiel soll dies verdeutlicht werden, siehe Abbildung 76.

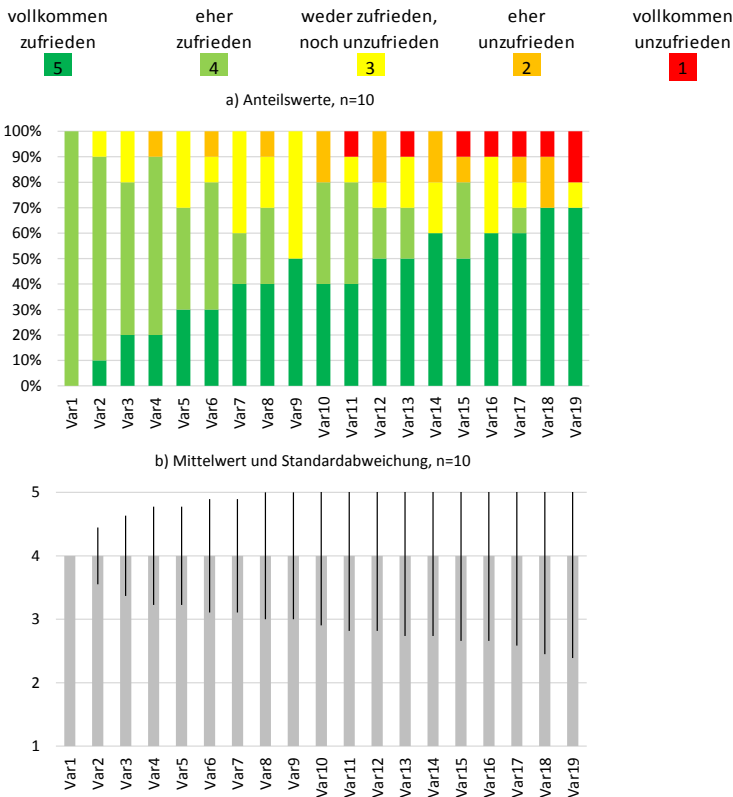


Abbildung 76: Beispiel für Varianten unterschiedlicher Anteilswerte und zugehörige wertgleiche Mittelwerte Antwortverteilungen als relative Häufigkeit (Segmentlänge) bei n=10, Sortierung nach Standardabweichung

Schon bei nur 10 Bewertungen ergeben sich auf einer 5-stufigen Skala mehrere Varianten der Verteilungen, die in ihrer Struktur sehr unterschiedlich sind, aber trotzdem den gleichen Mittelwert von 4,0 erreichen. Daran ist auch abzulesen, warum es sinnvoll ist, den Mittelwert ergänzt um die empirische Standardabweichung anzugeben, die im Beispiel im Bereich von 0 bis 1,6 liegt.

Die im Beispiel dargestellten Varianten sind wohl nur zum Teil wirklich so zu beobachten, vor allem mit Blick auf ein plausibles statistisches Modell und das Zustandekommen von Nutzerbewertungen. Unter realistischen Modellannahmen und in der Praxis (echte Daten) ist nicht von stark bipolaren oder sehr homogenen Werten, die nicht Extremwerte bilden, auszugehen. Der Kernpunkt der Kritik sollte aber klargeworden sein. Diese Überlegungen sind auch mit Blick auf die Festlegungen zur Bestimmung der Schwellen von Bedeutung.

Bei einer diskreten, wertbeschränkten Bewertungsskala hängen erreichter Mittelwert und Streuung der Bewertungen grundsätzlich eng miteinander zusammen, siehe [299] und Abbildung 77. In der Auswertung verschiedener Studienergebnisse zur QoE-Evaluation erkannten die Autoren in [299] einen quadratischen Zusammenhang zwischen erzieltem MOS (Mittelwert) und beobachteter Standardabweichung (SOS) der Einzelbewertungen. Es konnte gezeigt werden, dass sich aus dem Zusammenhang dieser Maße ein spezifischer Parameter berechnen lässt, der für bestimmte Dienstklassen charakteristisch ist. In der eigenen Auswertung deuten die Ergebnisse ebenfalls in diese Richtung. Mit nur je 5 Stützstellen ist der Parameterraum in diesem Fall aber recht grob aufgelöst und nur einseitig besetzt (Mittelwerte meist $> 3,0$), siehe Beispiele in Abbildung 78, so dass auf diese Darstellung im Weiteren verzichtet wird.

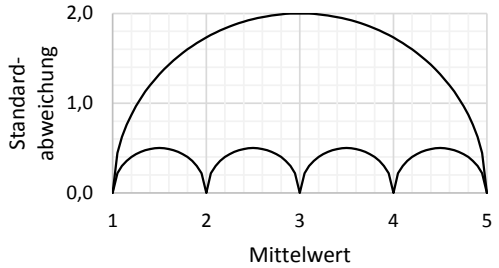


Abbildung 77: Zusammenhang von Mittelwert und Standardabweichung auf einer 5-stufigen diskreten Skala

Darstellung des Bereichs zwischen möglicher minimaler und maximaler Standardabweichung, nach [299] S. 132f berechnet

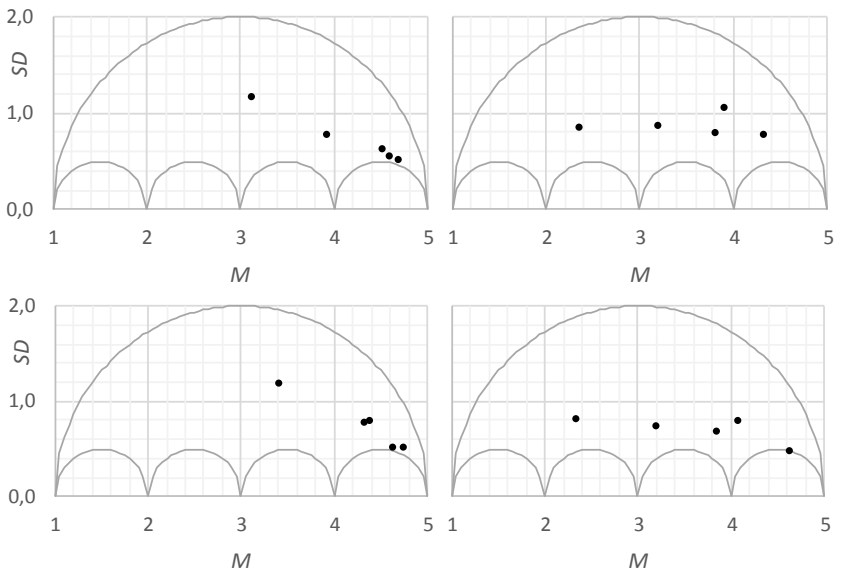


Abbildung 78: Standardabweichung zu Mittelwert, Beispieldarstellung für zwei ausgewählte Dienste jeweils nur Darstellung der Bewertung zur Gesamtzufriedenheit

oben: Drive, links Android, rechts iOS

unten: Google Maps, links Nutzung in der App, rechts Nutzung im Browser auf dem iPad 4

8.7 Ergebnisdarstellung und Visualisierung der bedingten Verteilungen

Eine geeignete Darstellung und Visualisierung der Ergebnisse ist essenziell, um die erhobenen Daten in brauchbarer Form zu dokumentieren und so als Grundlage zur Diskussion sowie zum angestrebten Erkenntnisgewinn zu nutzen. Besonders unter dem Gesichtspunkt zur Beurteilung der Ergebnisse in der Suche nach Schwellwerten und der Einschätzung von Parameterbereichen ist eine gute und korrekte Visualisierung äußerst hilfreich. Damit die Ergebnisdarstellungen des Testparcours nicht jeweils individuell erklärt werden müssen, soll kurz zusammengefasst werden, welche Konventionen und Visualisierungen es gibt.

8.7.1 Parameterraum und Ergebnisraum

Für die eigenen Ergebnisdarstellungen in Diagrammen gilt für die Achsen:

x-Achse: Parameterraum, z. B. QoS-Parameter oder QoS-beeinflussende Parameter (unabhängige Var.)

- kategoriale Darstellungen: Die x-Achse wird in Kategorien unterteilt dargestellt. Dabei bildet ein Preset eine Kategorie. Die Anordnung erfolgt (je Dienst) von links nach rechts aufsteigend. Für diese Darstellungen werden Säulendiagramme genutzt.
- wertproportionale Darstellung: Aus dem Parameterraum wird auf die x-Achse wertproportional abgebildet, z. B. linear oder logarithmisch korrekt skaliert. Für diese Darstellungen werden Liniendiagramme genutzt.

y-Achse: Ergebnis-/Bewertungsraum, QoE-Bewertung (abhängige Var.)

- Anteilswert der Bewertungen: relative Häufigkeit von 0 bis 100 % stetig
- Mittelwert der Bewertungen: Score von 1 bis 5 stetig

Sinnvolle Ausnahmen sind möglich, z. B. bei der Darstellung von Histogrammen und Verteilungskurven, bei denen die charakterisierte Größe generell auf der x-Achse liegt. Bei der Einbeziehung von internen Variablen (Mediatoren), wie z. B. von Effektivwerten, Übertragungszeiten und anderen Nutzerstimuli, ist je nach gewünschter Aussage zu unterscheiden, ob diese als abhängige oder unabhängige Variable auf der x- bzw. y-Achse abgebildet werden. Wird die Ergebnismenge eingeschränkt, z. B. durch Nebenfaktoren (Moderatoren), sind diese Parameter neben oder unter den Diagrammen angegeben.

8.7.2 Symbole und Abkürzungen

Es werden folgende Abkürzungen für statistische Kenngrößen genutzt, die sich an den gängigen Kennzeichnungen²⁰⁴ und englischen Entsprechungen orientieren:

- N : Umfang der Grundgesamtheit (oder je nach Kontext auch die Gesamtstichprobe)
- n : statistische Masse, Umfang der relevanten (Teil-) Stichprobe
- M : arithmetischer Mittelwert der Stichprobe (mean)
- SD : Standardabweichung der Stichprobe (standard deviation)

8.7.3 Visuelles Arrangement und Ergebnisgegenüberstellung

Die Basis zur Auswertung der bedingten Verteilungen bilden Mehrfeldertafeln/Verteilungstabellen (Kontingenz- / Pivot-Tabellen). Damit ist die multivariate Auswertung des Merkmals Zufriedenheitsbewertung zum Merkmal Preset möglich. Da es 6 Anwendungsszenarien mit jeweils einer Gesamtzufriedenheit und 2 bis 3 spezifischen Zufriedenheitsfragen gibt, ist es sinnvoll diese in einzelnen Tabellen bzw. Diagrammen darzustellen, wiederum unterschiedenen nach den sinnvollen Partitionen. Die zusammengehörigen Diagramme sind in Spalten, Zeilen und Blöcken organisiert, siehe auch die Anmerkungen in der Abbildung 79.

Zur weiteren Beschreibung der dargestellten Größen in den Diagrammen gilt:

- Spalten: Gesamtzufriedenheit (1. Spalte) und zwei oder drei spezifische Teilbewertungen
- Zeilen: jeweils eine Kopfzeile beschreibt die dargestellten Größen unterhalb
 - oben: Anteilswerte mit zusätzlichen Angaben
 - unten: Mittelwerte mit zusätzlichen Angaben
- Block: jeweils eine Auswahl von bis zu zwei Partitionen als „links, rechts“ (Kennzeichnung neben oder unter dem zugehörigen Block), wodurch jeweils ein weiterer Nebenfaktor (der jeweils wichtigste) direkt berücksichtigt werden kann

²⁰⁴ Orientierung an den Empfehlungen der American Psychological Association (APA) zur Beschreibung empirischer Ergebnisse, siehe [331] S. 16f mit Verweis auf [332] (nicht eingesehen). Die Nutzung spezieller (diakritischer) Zeichen wie \bar{x} für den empirischen Mittelwert ist in den genutzten Werkzeugen nicht an jeder Stelle möglich, weshalb darauf verzichtet wurde.

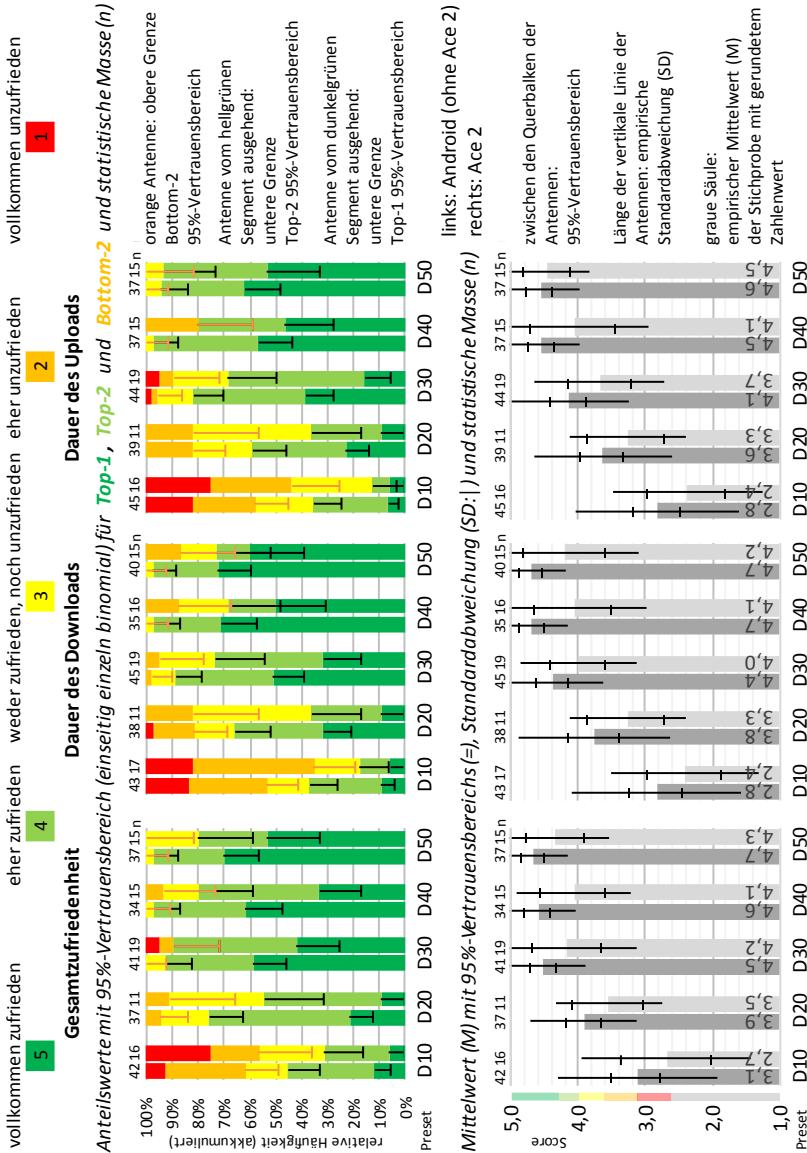


Abbildung 79: Beispiel für die Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung Beispiel aus dem Szenario Drive; obere Reihe (gedreht): deskriptive und analytische Statistik mit Anteilswerten; untere Reihe (gedreht): korrespondierende deskriptive und analytische Statistik mit Mittelwert und Standardabweichung als Lage- und Verteilungsmaßen

8.7.4 Anteilswerte der Bewertungen

Die nachfolgenden Erläuterungen beziehen sich auf Abbildung 79, obere Diagrammreihe. Den verschiedenen Formen der Darstellung der Ergebnisse bedingter Verteilungen von Likert-Skalen widmet sich [333]. Als einzig brauchbare der üblichen Darstellungsformen werden dabei divergierende gestapelte Balkendiagramme identifiziert. Diese bilden von einer neutralen Grundlinie aus je einen Stapelbereich für die negativen Bewertungen zur einen Seite und einen positiven Stapelbereich zur anderen Seite. Für die eigene Darstellung werden die Anforderungen so abgeschwächt, dass zwar eine gestapelte Darstellung genutzt wird, allerdings nicht divergierend.

Die Anteilswerte der Bewertungen (als relative Häufigkeit auf der y-Achse) werden mit gestapelten Säulen je Preset (als Kategorien auf der x-Achse) angegeben. Die prozentualen Anteilswerte ergänzen sich dabei immer zu 100 %. Durch die Stapelung der Anteilswerte beginnend auf der Grundlinie bei 0 % mit der Top-Bewertung (vollkommen zufrieden) wird das Anwachsen der Zufriedenheit von links, dem „schlechteren“ Preset, nach rechts, hin zu den „besseren“ Werten visuell unterstrichen. Der Anteil und Zuwachs des Anteils von „vollkommen zufrieden“ spielen in der Interpretation der Ergebnisse wichtige Rollen und müssen deshalb sofort zu erkennen und intuitiv ohne Rechnen direkt zu vergleichen sein. Die Säulen der Zufriedenheit werden so visuell in gleicher Art größer, wie auch die statistischen Maßzahlen in der Mittelwertdarstellung. Die akkumulierte Darstellung als gestapelte Säulen kommt der Auswertung der aggregierten Werte als Top-2- bzw. Bottom-2-Bewertungen entgegen und hat klare Vorteile gegenüber gruppierten Säulen. Die Top-2-Bewertung der Zufriedenheit, als die Summe von „vollkommen zufrieden“ und „eher zufrieden“, kann direkt von unten nach abgelesen werden. Die Bottom-2-Bewertung, als Summe der beiden schlechtesten Bewertungsoptionen, ist als Differenzwert zu 100 % von oben nach unten leicht abzulesen. Diese Art der Darstellung ist unabhängig von Auslegungen der Skala, der Bildung statistischer Maßzahlen oder Annahmen zur Verteilung der Bewertungen in der Stichprobe oder Grundgesamtheit und eignet sich damit zur Beschreibung der empirischen Ergebnisse der Stichprobe. Auf der 50-%-Linie lässt sich für jedes Preset unmittelbar der Median der Bewertungen ablesen (für weitere interessante Quantilen entsprechend anzuwenden).

Je Preset und Partition wird zusätzlich direkt oberhalb der zugehörigen Säule die statistische Masse n angegeben, die beim Vergleich zwischen den Presets und ggf. verschiedenen Teilergebnismengen zu beachten ist. Bedingt durch die Formatierung ist dabei zu beachten, dass bei einer vergleichenden Darstellung (links, rechts) die Zahlen recht dicht zueinander positioniert sind. Typischerweise sind es maximal zwei zweistellige Zahlen.

Einige der Parameter der Presets (Datenraten) sind so gewählt, dass sich diese jeweils um ungefähr den gleichen Faktor unterscheiden (entweder rund 2 oder $\sqrt{2}$). So nähert schon die gleichabständige kategoriale Darstellung eine entsprechende Abbildung mit logarithmischer Achse zur entsprechenden Basis sehr gut an. Die Abbildung ist damit prinzipiell auch geeignet, Aussagen im Vergleich der Änderungen zwischen den Kategorien zu treffen bzw. über entsprechende visuelle Elemente (Verbindungslinien) Zwischenwerte (grob) zu approximieren. Auf eine interpolierte Darstellung, z. B. als gestapelte Flächen, wird hingegen verzichtet, da für diese weitere Annahmen zu treffen wären.

Die Darstellung der Anteilswerte mit Berücksichtigung des statistischen Standardfehlers ausgewählter (aggregierter) Anteilswerte dient zur Schätzung des Bereichs der wahren Anteilswerte in der Grundgesamtheit ausgehend vom empirisch ermittelten Wert und dem zugehörigen Konfidenzintervall. Es wird jeweils ein einseitiges 95%-Konfidenzintervall als Whisker (Antennen an den Enden der Säulensegmente) für die festgelegte Bezugsgröße dargestellt: untere Vertrauensgrenze für Top-1 und Top-2 sowie die obere Vertrauensgrenze für Bottom-2. Da die Aggregation für die Zufriedenheit (Top-2) visuell von unten nach oben und für die Unzufriedenheit (Bottom-2) von oben nach unten erfolgt, weisen die Antennen zur Visualisierung der Vertrauensbereiche jeweils von den Enden der zugeordneten Balkensegmente nach unten. Für Top-1 und Top-2 wäre entsprechend auch ein jeweils kleiner Anteil möglich (kürzere Säulensegmente), für Bottom-2 ein entsprechend größerer Anteil (längere Säulensegmente).

Die vorgeschlagene Art der Darstellung ist ein guter Kompromiss zur Visualisierung der erzielten Ergebnisse und zu den angestrebten Erkenntnissen. Erst durch die getroffenen Festlegungen lassen sich gestapelte Anteilswerte zur de-

skriptiven Statistik und zugehörige Vertrauensbereiche zur analytischen Statistik für die interessierenden Anteilswerte sinnvoll kombiniert in einem Diagramm mit hoher Informationsdichte darstellen. In den Recherchen taucht diese Art der kombinierten Darstellung erstaunlicherweise nicht auf.

8.7.5 Mittelwertsäulen der Bewertungen

Die nachfolgenden Erläuterungen beziehen sich auf Abbildung 79, untere Diagrammreihe. Ergänzend zu den Anteilswerten werden weitere statistische Kenngrößen zur Beschreibung der Ergebnisse genutzt: der arithmetische Mittelwert als Lageparameter und die Standardabweichung als Verteilungsparameter. Auf der x-Achse werden die Presets als Kategorien genutzt, auf der y-Achse (stetig) die transformierte Score-Skala im Bereich [1, 5]. Zur Orientierung ist am linken Rand die abgeschwächte Bereichsfarbskala angegeben (MOS-Planungsskala, siehe 8.8.2). Zur besseren visuellen Unterscheidung werden für die Links-Rechts-Darstellung in den Säulen leicht unterschiedliche Grautöne genutzt. Die Länge der Säulen entspricht dem erreichten Mittelwert. Neben den genannten Komponenten zur deskriptiven Statistik wird zusätzlich der Vertrauensbereich (Konfidenzintervall) um den empirisch ermittelten Mittelwert mit einem Konfidenzniveau von 95 % zur analytischen Statistik angegeben. In der Darstellung ist zu beachten, dass die Antennen an den Säulenden eine doppelte Funktion übernehmen. Die Länge des vertikalen Strichs der Antennen visualisiert die Standardabweichung, der Bereich zwischen den Querbalken an den Antennen den Vertrauensbereich (welcher im Normalfall kleiner sein sollte als die Standardabweichung).

8.7.6 Mittelwertkurven der Bewertungen

Für die Mittelwertkurven wird auf der y-Achse die transformierte (stetige) Score-Skala wie bei den Mittelwertsäulen genutzt. Auf die x-Achse erfolgt die Projektion aus dem Parameterraum für eine einzelne Komponente, z. B. die maximale Download-Rate (in kBit/s) aus der Parameterkombination Preset. Dafür sind unterschiedliche wertproportionale stetige Skalierungen möglich, siehe Abbildung 80 und Tabelle 50. Die Interpretation ist nur zusammen mit den übrigen Komponenten der Parameterkombination sinnvoll. Für die Darstellung sollte die Komponente gewählt werden, die auf Gesamtzufriedenheit oder jeweilige Teilbewertung den größten Einfluss ausübt. Entsprechend der

Vorüberlegungen ist dies für die betrachteten Szenarien also entweder die maximale Datenrate im Download oder im Upload. Für Szenarien, in denen Teilbewertungen sowohl die maximale Datenrate im Download als auch im Upload als relevante Einflussgrößen aufweisen, können in der Darstellung der Gesamtzufriedenheit beide Komponenten berücksichtigt werden (als zwei Diagramme, optional).

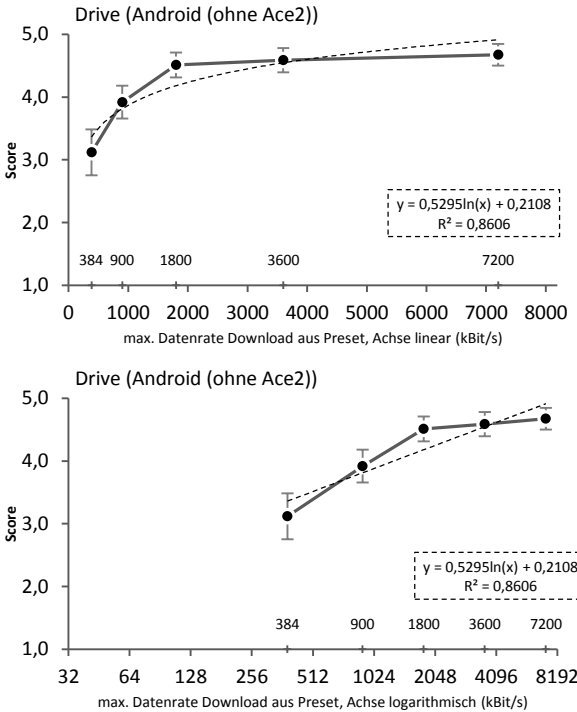


Abbildung 80: Beispiel zu den Darstellungen der Mittelwertkurven mit unterschiedlicher Skalierung der x-Achse

oben: linear
unten: logarithmisch

Tabelle 50: Vergleich der Darstellungen der Mittelwertkurven mit unterschiedlicher Skalierung der x-Achse

linear	logarithmisch zur Basis 2
spezifisch für den jeweiligen Dienst, angepasst an den ausgewählten Parameterbereich (niedrig, mittel, hoch)	einheitlich für alle Dienste, den gesamten genutzten Parameterbereich überdeckend
<ul style="list-style-type: none"> • die Lageverhältnisse sind intuitiv korrekt zu erfassen • zur Verdeutlichung der spezifischen Charakteristik eines Dienstes • Einschätzung ob Wahl der Abstufung im Parameterraum im Vergleich zum Bewertungsraum sinnvoll war • bzw. ob und wo Erweiterungen oder Verfeinerungen des untersuchten Parameterbereichs sinnvoll wären 	<ul style="list-style-type: none"> • für Vergleiche zur relativen Einordnung • Verdeutlichung des angenommenen nicht linearen Zusammenhangs • einzelne Datenpunkte in ungefähr gleicher Distanz • bessere Abschätzung zur Passung der Regressionsfunktion
gemeinsame visuelle Elemente	
<ul style="list-style-type: none"> • x-Achse: Außenskala mit fixem Wertintervall zur Orientierung und Innenskala mit den konkreten x-Werten der gezeigten Datenpunkte • Datenpunkte: empirische Mittelwerte der Gesamtzufriedenheit oder einer Teilbewertung • Verbindung zwischen den Datenpunkten: Geraden als vereinfachte Interpolation und zur visuellen Lenkung: Dies ist keine verlässliche Aussage darüber, dass die Kurve zwischen den ermittelten Datenpunkten tatsächlich so verläuft! • optional: Fehlerindikatoren in y-Richtung: 95%-Vertrauensbereich des Mittelwerts, Erläuterung siehe Mittelwertsäulen • optional: Fehlerindikatoren in x-Richtung: der Bereich des vermutlichen Effektivwerts der Komponente auf der x-Achse (bei Datenraten immer kleiner als der parametrisierte max. Wert, daher nur in negative Richtung anzugeben, meist einheitlich als Prozentsatz des x-Wertes, nicht für alle Dienste sinnvoll bzw. direkt zu bestimmen) • optional: Regressionsfunktion sowie ihre Formel und das zugehörige Bestimmtheitsmaß (R^2) innerhalb der gestrichelten Box angeben. Die Regressionsfunktion nähert die Abhängigkeit zwischen den Größen der x-Achse (QoS) und y-Achse (QoE) in einer mathematischen Funktion auf Basis der erhobenen Datenpunkte und des vermuteten Funktionstyps an. 	

Die Mittelwertkurven werden innerhalb der vorliegenden Arbeit nur zur Zusammenfassung der Ergebnisse aller Dienste in einem einzigen Diagramm, in einer Art Kennlinienfeld genutzt. Dann wird allerdings aus Gründen der Übersichtlichkeit auf die optionalen Elemente wie Fehlerindikatoren oder Regressionsfunktionen verzichtet. Die Zusammenfassung erlaubt den direkten Vergleich von Lage und Verlauf aller Kennlinien und hilft somit die Charakteristiken bestimmter Arten des QoS-QoE-Zusammenhangs zu erkennen und zu unterscheiden.

8.8 Prinzip der Schwellwerte

Die Suche und Ermittlung konkreter Schwellwerte war eines der zentralen Anliegen der Studie. Entsprechend lohnt der Blick auf die ermittelten Daten, ob sich damit die Annahmen zu „gut, besser, gut genug“ aus den Überlegungen zum grundsätzlichen Verlauf der Dienstkurven bestätigen lassen. Es wird nachfolgend dargelegt, welche Schwellen dafür prinzipiell sinnvoll sind, welche Bezugsmetriken möglich sind und wie die konkreten Werte für die jeweiligen Schwellen festgesetzt werden können. Die Bestimmung der Schwellen zu den untersuchten Diensten auf Basis der hier geschilderten Prinzipien erfolgt zusammengefasst im Kapitel zum Testparcours. Es wird im Weiteren davon ausgegangen, dass die Bewertung der Gesamtzufriedenheit eine geeignete Basis zur Bestimmung der gesuchten Schwellen bietet.

8.8.1 Motivation zur Suche der Schwellwerte

Die Motivation der Suche nach den Schwellwerten leitet sich aus einer einfachen Grundannahme ab: Nutzer und Anbieter haben potenziell verschiedene Interessen und Ziele:

- Nutzer: möglichst hohe QoE, Erfüllung der Anforderungen so gut wie möglich
- Anbieter: möglichst niedriger Aufwand (QoS), Erfüllung der Anforderungen so gut wie nötig, weil Aufwand Kosten bedeutet

Ferner wird angenommen, dass es neben dem „so gut wie möglich“ und „so gut wie nötig“ auch „gut genug“ gibt. Allerdings können sich auch hier unterschiedliche Sichten auf das „gut genug“ ergeben:

- Aus Nutzersicht ist „gut genug“ dann erreicht, wenn in den QoE-Bewertungen nur noch unbedeutende Verbesserungen festzustellen sind.
- Aus Anbietersicht ist „gut genug“ schon dann erreicht, wenn die Nutzer prinzipiell überwiegend schon zufrieden sind.
- Als Kompromiss: ein „gut genug“, das geeignet ist, beiderlei Interessen ausgeglichen zu berücksichtigen.

Für die Schwellenwerte bedeutet dies, dass die unterschiedlichen Ziele und Sichten zu beachten sind. Die vorhergesagte (und auch festgestellte) Art der Dienstkurven bekräftigt die Annahmen dazu. Überwiegend ist es sogar sinnvoll, gerade wegen der Nichtlinearitäten in den Dienstkurven. Dabei können verhältnismäßig kleine QoS-Änderungen zu verhältnismäßig großen QoE-Änderungen führen, ebenso wie verhältnismäßig große QoS-Änderungen nur zu verhältnismäßig kleinen QoE-Änderungen führen können.

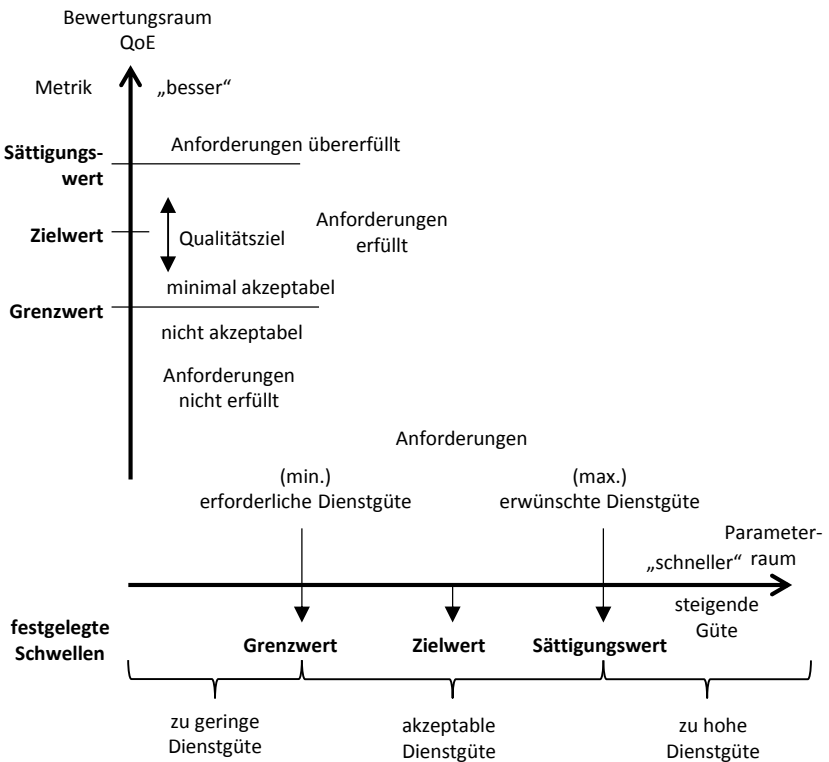


Abbildung 81: Relevante Bereiche und Lagebeziehung der Schwellen im Bewertungsraum und Parameterraum

Beispieldarstellung, inspiriert von [215] S. 145 (Bewertung) und [21] S. 242 (Dienstgüte)

8.8.2 Wahl der Schwellen

Aus den Überlegungen zu den Dienstkurven folgt, dass es einen Übergang von einem Bereich schlechter Bewertung zu einem Bereich guter Bewertung gibt. Diese Bereiche sind durch geeignete Schwellen unterschiedlicher Kennzeichnungen und Werte voneinander zu trennen. [45] S. 41/ S. 52f stellt eine Vielzahl möglicher Schranken und Grenzwerte (für QoS) vor, wie z. B. unterer Grenzwert, Zielwert und oberer Grenzwert, ordnet diese aber nicht hinsichtlich der Bildung bestimmter Bereiche, sondern nach ihrer (Aussagen-) Stärke. Praktischer ist eine einfache Unterscheidung in verschiedene Wertebereiche. Ein Ansatz zur QoS-Unterscheidung in drei Bereiche mit zwei Schwellenwerten findet sich in [21] S. 242ff. Dabei wird davon ausgegangen, dass es einen Wertebereich akzeptabler Dienstgüte gibt, der zwischen erforderlicher und erwünschter Dienstgüte liegt. Im Fall des Akzeptanzbereiches nach ISO 9126 bzw. ISO 14598-4 (siehe Ausführungen dazu in [215] S. 145) wird für eine Qualitätsmetrik noch ein Qualitätsziel eingeführt. Es ergeben sich aus den gemeinsamen Überlegungen dazu drei sinnvolle Bereiche, Schwellen und Bezeichnungen, siehe Abbildung 81.

Sprachqualitäts-				Top-2	Bottom-2
kategorie	R-Value	zu erwartende Nutzerzufriedenheit	MOS	untere Grenze	obere Grenze
beste	100	sehr zufrieden	4,50		
hoch	90	zufrieden	4,34	97%	~0
mittel	80	einige Nutzer unzufrieden	4,03	89%	~0
niedrig	70	viele Nutzer unzufrieden	3,60	73%	6%
schlecht	60	fast alle Nutzer unzufrieden	3,10	50%	17%
	50	nicht empfohlener Bereich	2,58	27%	38%
	0		1,00		

Abbildung 82: Bereichsschwellen (MOS und Anteile) aus den synthetischen R-Values des E-Modells

zusammengestellt nach [205] S. 14 und [334] S. 2f, angepasste Übersetzungen
Nutzung der erweiterten Ampelfarbskala mit reduzierter Sättigung zur Kennzeichnung der Bereiche in eigenen Darstellungen

8.8.3 Bezugsmetrik und Wertebereiche der Schwellen

Bezugsmetrik: Die Definition der Schwellen soll auf Basis geeigneter Metriken für die abhängigen Variablen im Bewertungsraum (QoE) erfolgen. Dafür sind prinzipiell die vorgestellten statistischen Kenngrößen geeignet: Mittelwert und Anteilswerte. Die Bestimmung der Schwellen auf Basis der Anteilswerte der Zufriedenheitsbewertungen wird nachfolgend noch weiter detailliert dargelegt werden. Für Mittelwerte ist entsprechend vorzugehen, bzw. sind die Überlegungen dazu sogar noch etwas einfacher, da nur eine Bezugsmetrik relevant ist. Dazu soll zunächst auf die R-Value-Schwellen eingegangen werden.

R-Value-Schwellen: Im E-Modell der ITU-T (siehe [205]) zur Berechnung der Sprachverständlichkeit finden sich einige Anhaltspunkte zu möglichen Schwellen. Ausgehend von einer linearen Qualitätsskala (siehe [334] S. 2f) mit synthetischen R-Values im E-Modell lassen sich Schwellen mit festen Bedeutungen für zu erwartende Mittelwerte (MOS) und Anteilswerte ableiten, siehe Abbildung 82. Im Unterschied zur Bewertungsskala, auf der die Zufriedenheit der Probanden erfasst wird, können die Abstufungen der berechneten Werte als eine Planungsskala bzw. Vorhersageskala gesehen werden (erwartete Werte ausgehend von den berücksichtigten Einflussgrößen im E-Modell).

Neben dem numerischen Mittelwert einer 5-stufigen ACR-Bewertungsskala (siehe 4.1.6) werden auch Schwellen für die untere bzw. obere Grenze der Anteile von Good-or-Better (GoB) und Poor-or-Worse (PoW) angegeben, also zu den jeweils beiden besten und beiden schlechtesten Bewertungsstufen. Die Verbalisierungen der Bereiche gibt dabei schon einen Hinweis auf den ungefähren Anteil unzufriedener Nutzer.

In der eigenen Studie wurde eine zur MOS-Skala abweichende Skala genutzt, allerdings auch 5-stufig. Entsprechend kann lohnen, die Überlegungen zu den definierten Bereichen zu prüfen. Auf der eigenen Skala und Terminologie zu den gerichtet kumulierten Anteilswerten (siehe 8.6.8) entsprechen die genannten Anteilsmetriken Good-or-Better und Poor-or-Worse den Bewertungen für Top-2 bzw. Bottom-2. Obwohl sich die Bereichsschwellen aus den E-Modell-Annahmen eigentlich auf den zu erwarteten MOS beziehen, wird in [205] nicht weiter erläutert, ob dabei dann vom empirischen Mittelwert oder der Untergrenze des Vertrauensbereiches auszugehen wäre. In einigen der eigenen

Darstellungen zu Mittelwerten (Mittelwertbalken und Kennlinienübersicht) werden die Bereichsgrenzen jeweils am linken Skalenrand farblich abgesetzt zur Einordnung mit angegeben. Dabei kann dann jeweils der Vergleich zum empirisch ermittelten Wert und zur relevanten unteren Grenze des Vertrauensbereichs erfolgen. Bei den Betrachtungen zu den Schwellwerten wird diese Bereichsskala abschließend noch einmal hinzugezogen und vergleichend gegenübergestellt, so dass sich zeigt, ob diese Bereiche und Abstufungen auch für die genutzte eigene Skala und die empirisch überprüften Dienste sinnvolle Ergebnisse liefern. Für die Anteilswerte folgt die Diskussion zu den eigenen Überlegungen und abweichend festgesetzten Anteilsschwellen unmittelbar nach.

Anteilswerte: Die Schwellwertsuche auf Basis der Anteilswerte weist große Ähnlichkeit mit der Suche nach der mittleren effektiven Dosis bei der Erprobung von Medikamenten in der medizinischen Forschung bzw. Bio-Statistik auf, dazu siehe [318] S. 534ff. Es wird dabei davon ausgegangen, dass sich die Untersuchungsobjekte (Probanden) hinsichtlich ihrer Ausprägung bestimmter Merkmale unterscheiden. Aufgrund dieser (von außen nicht unmittelbar ersichtlichen) Unterschiede ist für die Erzielung eines bestimmten Ergebnisses, z. B. Medikaments führt zur gewünschten Wirkung, jeweils eine individuell verschiedene Wirkstoffmenge, auch Dosis genannt, nötig. Diese Untersuchungen dienen damit der Suche eines bestimmten Wertes im Parameterraum, der im Ergebnisraum eine bestimmte Schwelle erreicht, im Allgemeinen bezogen auf ein binäres Ergebnis, z. B. als effektive Dosis ED₅₀, bei der 50% der Untersuchungsobjekte das gewünschte Ergebnis zeigen. Die Dosen des untersuchten Präparats werden oftmals auch exponentiell angesetzt, was ebenfalls eine Parallele zu den eigenen Überlegungen bei der Wahl der Abstufung der QoS-Parametrisierung bildet. Die Gemeinsamkeiten der Gedankenführung sollten deutlich geworden sein. In entsprechender Analogie wurde in der eigenen Studie versucht, für die Probanden die effektive QoS-Dosis zur Erzeugung von Zufriedenheit bei der Nutzung verschiedener mobiler Dienste zu finden.

Entsprechend ist es sinnvoll, in Analogie zur effektiven Dosis eine oder mehrere passende Bezugsmetriken auf Grundlage der Anteilswerte auszuwählen. Im Gegensatz zum Mittelwert ist es mit Anteilswerten möglich, die Schwellen

auf Basis verschiedener Metriken, also bezogen auf unterschiedliche (aggregierte) Anteilswerte zu definieren, wobei die dichotomen Aggregationen des geordneten Multinomialmodells einen guten Ansatzpunkt bieten, siehe 8.6.8. Die Nutzung der aggregierten Bottom-2-Anteilswerte (Unzufriedenheit) für die Bestimmung des unteren Schwellwertes (Grenzwert) erschließt sich ebenso leicht wie die Nutzung der aggregierten Top-2-Anteilswerte (Zufriedenheit) für den oberen Schwellwert (Sättigungswert). Für den mittleren Schwellwert (Zielwert) ist von einem bestimmten Maß an Zufriedenheit auszugehen, allerdings ist kein aggregiertes Maß als Summe zweier Anteilswerte als Metrik sinnvoll. Für die betrachteten Dienste im Testparcours zeigt sich, dass die Top-1-Bewertung ein robustes Maß ist und eine hohe Trennschärfe aufweist. Es zeigen sich die relevanten Veränderungen in der Zufriedenheit hauptsächlich für diese Bewertung. Der Anteil der Bewertung „vollkommen zufrieden“ sollte mit besseren technischen Parametern grundsätzlich monoton wachsend sein (Zugewinn auf Kosten aller anderen Anteilswerte).

Wertebereich: Konkrete Vorgaben für die Werte von Anteilen, z. B. zufriedener bzw. unzufriedener Nutzer, ließen sich in der Recherche – mit Ausnahme obiger R-Value-Vorgaben – nicht ermitteln. Es sind lediglich Anhaltspunkte für sinnvolle Bereiche zur Orientierung möglich. Dabei ist noch zu unterscheiden, ob es sich eher um allgemeine Angaben zur (Kunden-/Nutzer-) Zufriedenheit oder speziell um den Zusammenhang von QoS und QoE handelt. Letztlich sind es freie Festlegungen, die oft auf der Basis spezifischen *Erfahrungswissens* aus der Anwendungspraxis heraus gewählt werden. Durch die Variation der Höhen der Werte für die Schwellen lassen sich die notwendigen Werte im Parameterraum je nach Grundanspruchsniveau verschieben bzw. mehr oder weniger robust gestalten. Dabei sollte auch bedacht werden, dass ein Anteil von 10 % unzufriedener Nutzer, auch wenn es zunächst wenig klingt, bei einer Nutzerbasis im Millionenbereich als absolute Zahl noch eine Menge potenziell unzufriedener Nutzer ist, von denen viele Accounts bei Facebook und Twitter haben, über die sie ggf. ihren Unmut kundtun ...

Tabelle 51 fasst die eigenen Überlegungen zu den sinnvollen Schwellen, möglichen Bezugsmetriken und Wertebereichen mit Erläuterungen zur Intention und Interpretation zusammen.

Tabelle 51: Schwellen, Kriterien, Erläuterungen, Bezugsmetriken und sinnvolle Wertbereiche

Schwellen und Kriterien	Intention und Interpretation	mögliche Bezugsmetrik und Bereich
<p>obere Schwelle: Sättigungswert</p> <p>Ratio</p> <p>Nutzersicht</p> <p>Anbietersicht</p> <p>Bestimmung</p>	<p>Diese Schwelle braucht nicht überschritten werden.</p> <p>Eine Überschreitung ist nicht sinnvoll, da anzunehmen ist, dass Kosten und Aufwand immer weiter steigen, gleichzeitig aber der Zugewinn an Nutzen immer weiter abnimmt²⁰⁵.</p> <p>Für die Mehrheit der Nutzer ergeben sich kaum noch Verbesserungen. Die Zufriedenheit nimmt kaum noch zu.</p> <p>Es ist nicht sinnvoll, wirklich alle Nutzer (vollkommen) zufriedenstellen zu wollen, also auch die mit extrem hohen Ansprüchen.</p> <p>Ein bestimmter Prozentsatz an Zufriedenheit braucht nicht überschritten werden.</p>	<p>Top-2 > 80 – 95 %</p>
<p>mittlere Schwelle: Zielwert</p> <p>Ratio</p> <p>Nutzersicht</p> <p>Anbietersicht</p> <p>Bestimmung</p>	<p>Diese Schwelle sollte erreicht werden.</p> <p>Der Zielwert ist im Bereich zwischen Grenzwert und Sättigungswert als Kompromiss zwischen den Zielen der Nutzer (so gut wie möglich) und des Anbieters (so gut wie nötig) zu sehen.</p> <p>Für die Mehrheit der Nutzer ergeben sich je nach Richtung graduelle Verbesserungen oder Verschlechterungen.</p> <p>Der Zielwert ist das angestrebte variable Qualitätsziel für das eigene Angebot.</p> <p>Ein bestimmter Prozentsatz an sicherer (vollkommener) Zufriedenheit soll erreicht werden.</p>	<p>Top-1 > 30 – 50 %</p>
<p>untere Schwelle: Grenzwert</p> <p>Ratio</p> <p>Nutzersicht</p> <p>Anbietersicht</p> <p>Bestimmung</p>	<p>Diese Schwelle sollte nicht unterschritten werden.</p> <p>Eine Unterschreitung bedeutet eine überproportionale Abnahme der Zufriedenheit und eine überproportionale Zunahme der Unzufriedenheit.</p> <p>Die Schwelle ist gerade gut genug, um die Anforderungen minimal akzeptabel zu erfüllen.</p> <p>Die Schwelle ist der minimal unbedingt erforderliche Wert.</p> <p>Ein bestimmter Prozentsatz unzufriedener Nutzer muss unterschritten werden.</p>	<p>Bottom-2 < 10 – 20 %</p>

Darstellung auf Basis ausgewählter Anteilswerte der Bewertungen, alternative Metriken und jeweilige Bezugswerte können ebenfalls sinnvoll sein

²⁰⁵ In den Wirtschaftswissenschaften ist dies als das „Erste Gossensche Gesetz“ zum abnehmenden Grenznutzen bekannt, siehe [335]. Es beschreibt einen solchen Zusammenhang zwischen (konstant) steigenden Kosten und gleichzeitig immer weiter zurückgehendem Zugewinn an Nutzen.

8.8.4 Ermittlung auf Basis der Stichprobe oder Grundgesamtheit

Bei der Festlegung der Höhe der Schwellwerte ist zu beachten, ob es sich um Werte bezogen auf eine Stichprobe oder auf die Grundgesamtheit aller Nutzer handeln soll. Theoretisch ist sinnvoll, die Schwellen auf Basis der angenommenen Grundgesamtheit anzusetzen und mittels der Vertrauensbereiche zu berücksichtigen, was dies konkret für Stichproben unterschiedlichen Umfangs ergibt. Praktisch bedeutet dies aber, dass ein Unterschreiten sehr niedriger oder Überschreiten sehr hoher Anteilswerte nicht möglich ist, wenn keine genügend große statistische Masse vorliegt, um die Vertrauensbereiche entsprechend klein zu gestalten.

Eine andere Formulierung verdeutlicht die Problematik an einem Beispiel: Um mit 95 % Sicherheit aussagen zu können, dass max. 10 % der Nutzer in der Grundgesamtheit unzufrieden sind, muss man mindestens 30 Probanden in der Stichprobe haben und keiner von diesen darf unzufrieden sein (Bottom-2).

Für die praktische Anwendung ist es daher sinnvoll, nicht nur eine einzelne Schwelle und den zugehörigen empirischen Einzelwert zu betrachten, sondern auch die QoS-QoE-Charakteristik.

8.8.5 Berücksichtigung der QoS-QoE-Charakteristik

Für „nicht schlechtere“ Werte im Bewertungsraum soll als Nebenbedingung der Monotonie des QoS-QoE-Zusammenhangs im Parameterraum gelten, in Abhängigkeit der Richtung „besser“:

- größer = besser: Grenzwert \leq Zielwert \leq Sättigungswert (z. B. Datenraten)
- kleiner = besser: Grenzwert \geq Zielwert \geq Sättigungswert (z. B. Latenz)

Dies bedeutet auch, kann ein Grenzwert nicht explizit bestimmt werden, aber ein Zielwert, dann ist damit auch implizit ein gleich großer Grenzwert bestimmt. Diese Annahmen decken sich auch mit den eigenen Daten. In den erzielten Ergebnissen geht ein Überschreiten von 33% von „vollkommen zufrieden“ (Top-1) meist einher mit dem Unterschreiten von 10% Unzufriedenheit (Bottom-2). Das Überschreiten von 50% von „vollkommen zufrieden“ Top-1 geht meist einher mit dem Erreichen von 90% Zufriedenheit (Top-2).

Bei fixierten QoE-Schwellwerten ergibt sich die Größenordnung für die Distanz der Schwellwerte im Parameterraum, als die relative Lage zueinander, aus der grundsätzlichen Art des QoS-QoE-Zusammenhangs. Die Schwellen hängen damit mit der Dienstkurve zusammen, entsprechend kann zur treffenden Beschreibung auf die dort genutzte Terminologie zurückgegriffen werden: eng vs. weit bzw. hart vs. weich oder sprunghaft vs. elastisch/kontinuierlich. Für einen sprunghaften Verlauf der Dienstkurve sollten die Schwellen eng zusammenliegen, für kontinuierliche QoE-Verbesserungen hingegen weiter auseinanderliegen. Daraus ergibt sich auch, dass die Auswirkungen des Unterschreitens der Schwellen, besonders des Grenzwertes, von der Art des Zusammenhangs abhängen. Bei einem sprunghaften Zusammenhang führt bereits eine geringe Unterschreitung zu sehr deutlichen QoE-Einbußen, bei einem weichen Zusammenhang im Vergleich dazu nur zu einer weiteren graduellen Verschlechterung der QoE. Grundsätzlich sollten die Nichtlinearitäten beachtet werden. Je nach Wahl der Schwellwerthöhe und konkreter Ausprägung des Knickbereichs sollten starke QoE-Änderungen zwischen Grenzwert und Zielwert deshalb vermieden werden.

8.8.6 Bestimmung der Schwellwerte

Ausgehend vom Erreichen bzw. Über- oder Unterschreiten der festgesetzten QoE-Schwellwerte in den empirisch ermittelten QoE-Bewertungen, werden die Schwellwerte im QoS-Parameterraum bestimmt, siehe Abbildung 83.

Für die Bestimmung der Schwellen im Parameterraum und die Charakteristik des Zusammenhangs sieht ein sinnvolles Vorgehen wie folgt aus:

- Identifikation der relevanten Parameter für den Zusammenhang: Ist ein QoS-Parameter oder die Kombination mehrerer QoS-Parameter maßgeblich?
- qualitativ: Klassifizierung der Art des Zusammenhangs (sprunghaft/kontinuierlich)
- quantitativ: Bestimmung der Werte der verschiedenen Schwellen für die als relevant identifizierten Parameter unter Berücksichtigung der Zusammenhangscharakteristik über die empirisch ermittelten Werte mit den genannten Vertrauensbereichen

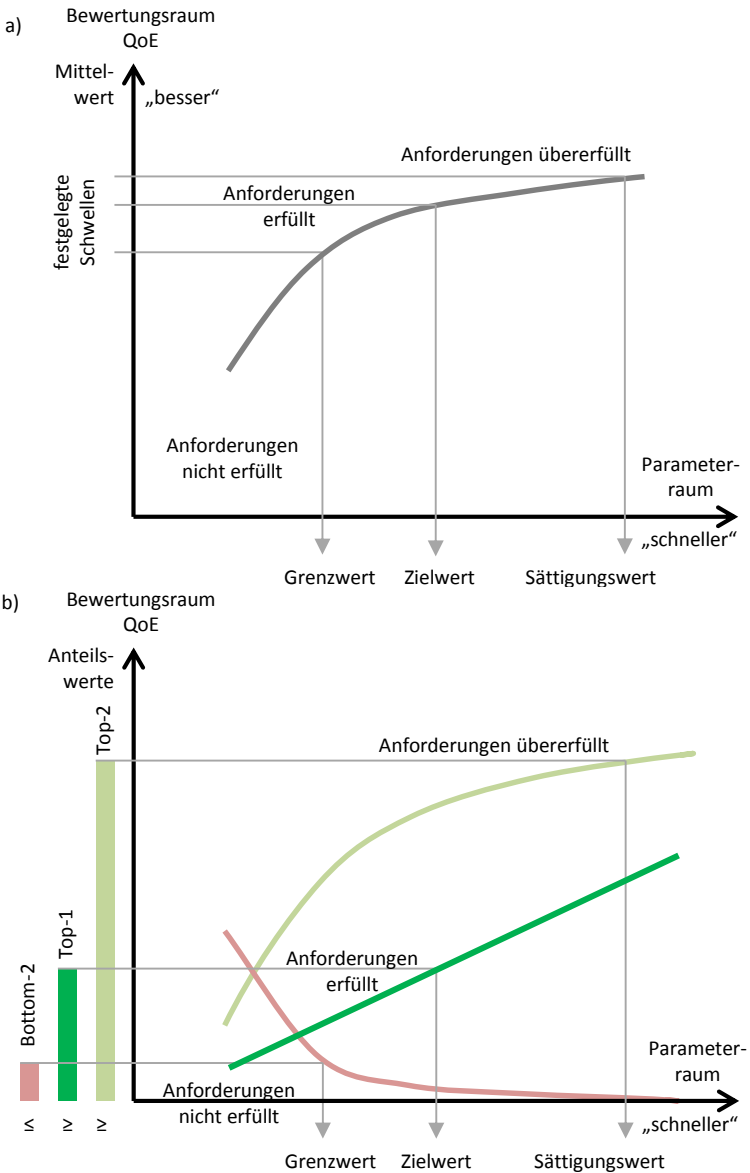


Abbildung 83: Schwellwerte im Bewertungsraum und Parameterraum Illustrationen mit möglichen Kurvenverläufen (interpoliert), die auf Basis der ermittelten Ergebnisse realistisch erscheinen

a) für die Betrachtung des Mittelwerts aus den Bewertungen, eine Metrik für alle Schwellwerte
 b) unterschiedliche Metriken für verschiedene Schwellwerte, hier die Betrachtung der (aggregierten) Anteilswerte

8.8.7 Interpretation der Schwellwerte

Die Richtung (max. Datenraten und min. Latenz) der Begrenzungen der einzelnen Parameter im Testbed ist bei der Interpretation der empirischen Ergebnisse und Schwellen zu beachten. Sollen die QoE-Werte der Schwellen sicher erreicht werden, ist von einer Umkehrung der Begrenzung auszugehen, also von Garantien hinsichtlich minimaler Datenraten und maximaler Latenz. Prinzipiell bestimmt bei einer zusammenfassenden Betrachtung der Dienste, derjenige, der die höchsten Einzelanforderungen hat, die gemeinsamen Anforderungen der Zusammenstellung. Für die Schwellen bzw. deren Werte gilt dies entsprechend. Dabei von einem globalen Wert zu sprechen wäre trotzdem nicht korrekt, da dies implizieren würde, dass alle möglichen Dienste und Anwendungen in einem globalen – oder besser universellen – Wert erfasst sind. Eine solche Aussage ist aber *nicht* möglich, ohne *alle* Dienste und Anwendungen untersucht zu haben, was nicht realistisch ist.

8.9 Güte- und Fehlerabschätzung

Nachfolgend soll eine Güte- und Fehlerabschätzung zur empirischen Ermittlung versucht werden. Die vorsichtige Formulierung als Versuch deshalb, weil nur mehr die Ansätze zur Umsetzung als der Nachweis des Erreichens in Form von statistischen Maßen präsentiert werden können. Für empirisch ermittelte Ergebnisse sind Überlegungen zur Güte- und Fehlerabschätzung von hoher Bedeutung hinsichtlich Aussagekraft und Verallgemeinerbarkeit. Dafür sind drei Testgütekriterien etabliert, für die von einer grundsätzlichen Abhängigkeit auszugehen ist:

Objektivität → Reliabilität → Validität (siehe weitere Erläuterungen in [216] S. 180ff.).

Zuverlässigkeit (Reliabilität) und Richtigkeit (Validität) können zur Genauigkeit zusammengefasst werden [336] S. 187. Diese Diskussionen zur Fehler- und Güteabschätzung werden durch die Überlegungen zu Testbarkeit und Vergleichbarkeit in den Vorüberlegungen ergänzt. Zusätzlich zu den allgemeingültigen Ansätzen zur Güteabschätzung sei auch auf Überlegungen speziell für QoE-Messungen verwiesen (siehe z. B. [337]²⁰⁶). Im Unterschied zu den

²⁰⁶ dort allerdings mit dem Fokus auf Crowdsourcing-Evaluationen

allgemeinen Diskussionen zur Validität psychometrischer Tests in der Literatur ist mit Blick auf die eigene Laborstudie als psychophysikalisches Experiment nicht nur der Faktor Mensch und das Testprozedere kritisch zu prüfen, sondern auch der technische Teil sowie das Zusammenspiel von Mensch und Technik. Eine kurze Einführung zur kritischen Auseinandersetzung mit den Gütekriterien psychophysikalischer Messungen hält [30] S. 20ff bereit.

Letztlich dienen diese Überlegungen zur Abschätzung, wie weit man den erzeugten Daten vertrauen kann und welche Aussagekraft sie beinhalten.

8.9.1 Objektivität

„Die Objektivität eines Tests gibt an, in welchem Ausmaß die Testergebnisse vom Testanwender unabhängig sind.“ ([216] S. 180) Die Objektivität bezieht sich als Gütekriterium einer Messung also auf denjenigen, der die Messung durchführt. Als Messung durch Menschen am Menschen ist dabei in der Laborstudie die durchführende Testleitung adressiert. Die Bewertungen der Probanden sind wie erörtert ja ohnehin subjektiv, zur Generalisierbarkeit der Ergebnisse gibt die Beschreibung der Stichprobe Auskunft.

Die Objektivität bezogen auf die Testleitung wurde durch standardisierte Vorbereitung und Durchführung der Experimente sichergestellt. Dazu gehörten vorgegebene und zu protokollierende Vorbereitungschecklisten zur Technik, ein standardisierter Einweisungs- und Aufklärungsleitfaden, Hinweise zur Beantwortung von Rückfragen und eine möglichst genaue Dokumentation von ungeplanten Abweichungen in einem Testlogbuch. In der eigentlichen Ausführung des Experiments selbst sorgte eine Testlaufautomatik dafür, dass die Netzparametrisierung wie geplant und ohne Einflussnahme durch das Testpersonal erfolgte. Die Objektivität in der Auswertung wird durch definierte Scores zu den Antworten in der quantitativen Auswertung sichergestellt. Zu den qualitativen Aspekten der Auswertung, z. B. zur Interpretation der Nutzerkommentare oder zur Beurteilung der Gültigkeit einzelner Ergebnisse, werden jeweils die entsprechenden Auswerteschritte dokumentiert.

8.9.2 Reliabilität

„Die Reliabilität eines Tests kennzeichnet den Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird.“ ([216] S. 181) Dabei sollte gelten, dass ein gleicher Input zu einem gleichen Output führt. Im vorliegenden Testdesign ist eine Beurteilung der Reliabilität durch einen Vergleich der Probanden mit sich selbst möglich, da eine teilweise Within-Subjekt-Variation für die Netzparameter genutzt wurde. Dafür wurde ein Intra-Rater-Vergleich im Rahmen der Auswertung durchgeführt, der nicht nur parametergleiche Messwiederholungen berücksichtigt, sondern auch die Tendenz der Veränderungen zur Einschätzung der Reliabilität berücksichtigt (Selbstkonsistenz, siehe [305]). Daneben erlaubt der Vergleich der Einzelbewertungen mit der Gesamtzufriedenheit eine Aussage zur internen Konsistenz der Antworten. Beide Ansätze zur Aufklärung des Bewertungsverhaltens sind gesondert dargestellt.

Eine gleiche QoS-Parametrisierung kann durch Einflüsse der Nebenfaktoren wie App und Endgerät oder auch durch das Testdesign zu unterschiedlichen Leistungswahrnehmungen führen. Dazu geben das Modell und die Vorüberlegungen zu den einzelnen Dienstszenarien spezifisch Auskunft. Bei einem relevanten Einfluss werden diese gesondert betrachtet und für jeweils äquivalente Bedingungen gesonderte Ergebnispertitionen vorgeschlagen. Die Ursache und Wirkung externer Störfaktoren (Echtheit und Probanden), die ggf. die Stabilität der Messungen beeinflusst haben könnten, wurden in der Theorie und in der Diskussion der jeweiligen praktischen Bedeutung für jedes Dienstszenario berücksichtigt. Die Anmerkungen zu Parameter-basierter und Signal-basierter Auswertung hinsichtlich der möglichen und notwendigen Genauigkeit sind zu beachten.

8.9.3 Validität

„Die Validität eines Tests gibt an, wie gut der Test in der Lage ist, genau das zu messen, was er zu messen vorgibt.“ ([216] S. 185) Zu den Hauptarten der Validität wie Inhaltsvalidität, Kriteriumsvalidität und Konstruktvalidität (siehe [216] S. 185) können nur beschränkt Aussagen getroffen werden. Diese müssten sich darauf beziehen, ob mit dem entworfenen Test und Testwerkzeugen, wie z. B. dem Testbed und dem genutzten Fragebogen, überhaupt die relevante Messgröße, hier die Zufriedenheit, zu erfassen ist. Grundsätzlich ist

davon auszugehen, da augenscheinlich die relevanten Aspekte (auf Basis des Modells) erfasst sind, die Ergebnisse im Rahmen der prognostizierten Werte als Annahmen zu den getesteten Parameterbereichen liegen und sich in prognostizierter Art und Weise ändern (prognostische Validität).

8.9.4 Einschätzung des empirischen Untersuchungsdesigns

Als eine andere Art der Validität lassen sich die Gütekriterien des gewählten empirischen Untersuchungsdesigns isolieren (siehe [216] S. 52ff. und [338] S. 57):

- interne Validität: Aussagesicherheit (Wirkung der unabhängigen Variablen und Ausschaltung/Einfluss der Störfaktoren)
- externe Validität: Verallgemeinerungsfähigkeit (Situation, Stichprobe, Zeitraum, ...)

Die Wirkungen der unabhängigen Variablen und möglicher technischer Störgrößen wurden in den Vorüberlegungen diskutiert und soweit möglich berücksichtigt. In der Auswertung wurde versucht, Aufgabenausführungen und Bewertungen, die nicht den gesuchten QoS-QoE-Zusammenhang abbilden, durch Überprüfung der Ergebnisse (in unterschiedlicher Komplexität) herauszufiltern. Die entsprechenden Ansätze der Validierung zur Erhöhung der Aussagesicherheit wurden gesondert detailliert dargelegt.

Die externe Validität ist durch das gemischte Testdesign und die Ausführung als Laborexperiment (Situationsvalidität) gegenüber einer Felduntersuchung grundsätzlich als einschränkt einzuschätzen (siehe [216] S. 54ff.). Für die eigenen Ergebnisse liegen keine konkreten Vergleichsdaten dazu vor, es sind aber die Hinweise aus der entsprechenden Diskussion zum Testdesign und zum situativen Kontext zu berücksichtigen, wenn die Ergebnisse entsprechend verallgemeinert werden sollen. Für das Testbed selbst sind die genannten Punkte zu den Vereinfachungen zum Netz und den Netzsituationen sowie zu den Grenzen einer Netzemulation zu berücksichtigen. Für den Faktor Mensch in der Messung des QoS-QoE-Zusammenhangs sind die Annahmen zur Aussagesicherheit weniger detailliert möglich. Mit Blick auf die Laborsituation und die Nutzung der Endgeräte im WLAN-Testbed sind die Anmerkungen zur Netz Wahrnehmung zu beachten. Zur Populationsvalidität gibt die Beschreibung der Stichprobe Auskunft.

8.9.5 Generelle Fehlerabschätzung

Die Messunsicherheit umfasst im Rahmen des Modells zur Messung des QoS-QoE-Zusammenhangs mindestens den Wert des statistischen Standardfehlers, da die empirisch ermittelten Werte nur einer begrenzten Stichprobe entstammen. Daneben sind auch andere Fehlerquellen zu berücksichtigen. Prinzipiell sind Analysen zur Ermittlung von Ausreißern (als einzelne Datenpunkte außerhalb des typischen Bereichs) und Anomalien möglich, z. B. um zuvor nicht berücksichtigte Partitionen zu identifizieren. Dabei wäre aber jeweils nachzuprüfen, ob die vermeintlichen Ausreißer nicht doch durch subjektive Bewertungsunterschiede zu erklären wären und ob vermeintlich neue Partitionen durch das QoS-QoE-Modell erklärbar sind. Die Anzahl Nebenfaktoren im Test sollte nicht unterschätzt werden. Durch deren Einfluss und weitere mögliche externe Einflüsse gewinnt die Rolle der statistischen Masse zum Ausmitteln zufälliger Schwankungen an Bedeutung. Bei verhältnismäßig kleiner statistischer Masse ist entsprechend Vorsicht und ggf. ein kritischer Umgang mit den erzeugten Daten geboten.

8.10 Zusammenfassung

Die erzeugten Daten wurden einer Parameter-basierten Auswertung unterzogen, d. h. unter Annahme einer Klassierung zu den ermittelten Erwartungswerten ausgehend von variierten QoS-Netzparametern. Es erfolgte eine (Vor-)Validierung der gesammelten Daten. Durch Aggregation und Partitionierung wurden zusammengehörige Teilergebnismengen für jeweils gleiche sonstige Bedingungen gebildet (gleicher QoS-QoE-Zusammenhang).

Der Bearbeitungserfolg und Abbruch- bzw. Wiederholungsgründe können über die Bewertung hinweg zusätzliche Hinweise geben, wie die Probanden (vermeintliche) Probleme mit der Internetverbindung einschätzen und mögliche allgemeine Schwierigkeiten in den Dienstszenarien aufdecken.

Durch die Erzeugung der Word-Clouds ist eine einfache Analyse der Nutzerkommentare möglich, die schon recht interessante ergänzende Einblicke gewährt, was für die Nutzer in den Aufgabenbearbeitungen tatsächlich wichtig

war. Die Vermutungen zu möglichen Problempunkten und relevanten Faktoren können auf Basis der Kommentare für die einzelnen Dienste insgesamt bestätigt werden.

Ein Intra-Rater-Vergleich und der Vergleich der Teilbewertungen zur Bewertung der Gesamtzufriedenheit geben Aufschluss zum Bewertungsverhalten. Der Intra-Rater-Vergleich mit passender Visualisierung erlaubt die Einschätzung, ob in den Szenarien wie angestrebt die QoE-Veränderungen durch QoS-Veränderungen bewirkt wurden und ob die Probanden in erwarteter und plausibler Art ihre Bewertungen änderten. Auf Basis der Ergebnisse kann dies bestätigt werden. Stark diskordante Bewertungen hingegen konnten häufig als invalid identifiziert werden.

Der QoS-QoE-Zusammenhang selbst wird über die bedingten Verteilungen von Preset als QoS-Netzparameterkombination zu den QoE-Bewertungen der Nutzerzufriedenheit ermittelt. Es wurde erläutert, warum die genutzte Skala zur QoE-Bewertung als metrische Skala ausgelegt werden kann und wie die festgestellten Verteilungen der Bewertungen statistisch einzuordnen sind. Die Diskussion des Mean Opinion Score als übliche QoE-Metrik offenbarte einige Schwächen, so dass die eigene Auswertung der bedingten Verteilungen zum QoS-QoE-Zusammenhang sowohl Anteilswerte als auch Mittelwerte vorsieht. Auf Basis dieser Metriken sind auch Schätzungen zur Grundgesamtheit in Form von Vertrauensbereichen möglich.

Das Prinzip der Schwellwerte mit unterschiedlichen Schwellen, Bezugsmetriken und Wertebereichen wurde vorgestellt. Auf Basis der gezeigten Überlegungen und Festlegungen konnten in den gesammelten Daten die Charakteristik des QoS-QoE-Zusammenhangs berücksichtigt und die gesuchten Schwellwerte ermittelt werden.

Eine Güte- und Fehlerabschätzung gibt Hinweise zur Einschätzung und Aussagekraft der erzeugten Ergebnisse.

9 Der Testparcours

Ergebnisse und Diskussion der Voruntersuchungen und des Nutzertests

In diesem Kapitel erfolgt die gemeinsame Darstellung der Ergebnisse der Voruntersuchungen und der erzielten empirischen Ergebnisse des eigentlichen Nutzertests zu den entwickelten Dienstszenarien.

Aus relevanten Diensten wurde eine möglichst repräsentative Auswahl zusammengestellt, für die weitere Vorüberlegungen und Voruntersuchungen zur Testeignung erstellt wurden. Es wurde immer ein konkretes Szenario²⁰⁷ mit einem konkreten Dienst als Kombination von echter Anwendung (App), echten Inhalten, echter Gegenseite und sinnvoller Aufgabenstellung betrachtet und getestet. Die Zusammenstellung aller ausgewählten Dienste bildet den Testparcours, den die Probanden im Test zu absolvieren hatten.

Einem festen und einheitlichen Aufbau folgend werden die einzelnen Szenarien vorgestellt:

1. Nutzungsszenario
2. Technische Sicht und Netznutzung
3. Nutzersicht
4. Parametrisierung und Erwartungswerte
5. Bearbeitungserfolg
6. Nutzerkommentare
7. Bewertungsverhalten
8. Bedingte Verteilungen

Dabei entsprechen die Abschnitte Nutzungsszenario bis Parametrisierung den Ergebnissen und Festlegungen der Planungsphase und Voruntersuchungen. Vom Bearbeitungserfolg ausgehend werden die im Laborexperiment erzielten Ergebnisse und Einsichten dargelegt und diskutiert. Im Anhang findet sich zu jedem Szenario der Fragebogen mit der Aufgabenstellung und den zugehörigen Fragen zur Bewertung, ergänzt um die tatsächliche Nutzersicht auf das

²⁰⁷ Anwendungsfall bzw. Use Case können als äquivalente Bezeichnungen verstanden werden

Szenario in Form von Screenshot-Zusammenstellungen direkt von den genutzten mobilen Geräten.

Die Beschreibung und Diskussion gemeinsamer bzw. für alle Dienste einheitlicher Punkte sowie die vergleichende Ergebnisdarstellung werden in einer Übersicht am Ende des Kapitels zusammengefasst. Der ungeduldige Leser kann daher ans Ende des Kapitels springen und bei Unklarheiten oder offenen Fragen die Details im Abschnitt zum entsprechenden Szenario nachlesen.

9.1 Drive

9.1.1 Nutzungsszenario

In Expertenbefragungen wird der schnelle Popularitätsthroughbruch von mobilen Cloud-Diensten erwartet [282]. Die mobile Nutzung von Cloud-Diensten kann dabei sehr vielfältig sein. [339] liefert dazu einen Ansatz zur Klassifikation. Im gewählten Szenario stellt die Cloud die Dienste bereit, die auf den mobilen Geräten genutzt werden (Klassifikation als Mobile as a Service Consumer). Eine Dienstkategorie dabei ist Mobile Cloud Storage. Dazu zählen Dienste wie beispielsweise „Dropbox“, „Box.net“, „Skydrive“ und eben auch „Google Drive“. Diese erlauben die Datenspeicherung in einer zentralen Ablage im Internet, z. B. von Bildern und Dokumenten, die auf unterschiedliche Geräte transferiert und synchronisiert werden können. Im konkreten Fall fiel die Wahl auf Google Drive, da dieser Dienst in den Voruntersuchungen die geforderten Kriterien (mögliche Datenrate beim Download und Upload) vor allem hinsichtlich Kontrollierbarkeit besser erfüllte als eine ebenfalls betrachtete Alternative.

Die Usability der Google-Drive-Apps kann auf beiden Plattformen als durchwachsen angesehen werden. Es sind zur Nutzung relativ viele Zwischenmenüs und Auswahlsschritte notwendig, die sich in Details zwischen den Plattformen bzw. sogar einzelnen Geräten unterscheiden. Zusätzlich wurde mit insgesamt geringen Vorkenntnissen der Probanden zur App gerechnet. Das Anwendungsszenario wurde deshalb bewusst nicht zu kompliziert gewählt.

Im Szenario wird der Cloud-Speicher zur Ablage von Fotos genutzt, die mit dem mobilen Gerät erzeugt wurden (eine Art Online-Galerie):

- Download: Der Nutzer hatte zuvor in dem Cloud-Speicher ein Handyfoto gespeichert, das er nun gerne wieder auf dem mobilen Gerät einem Bekannten zeigen möchte.
- Upload: Der Nutzer hat mit seinem mobilen Gerät ein tolles Foto geschossen, das er nun gerne im Cloud-Speicher sichern möchte.

9.1.2 Technische Sicht und Netznutzung

Aus technischer Sicht ist Drive recht unspektakulär. Eine wie geplant ausgeführte Bearbeitung des Szenarios ist auf Ebene der Verkehrsdaten wie folgt zu beschreiben: Zu zwei disjunkten Zeitpunkten werden je ein kontinuierlicher Download und ein Upload des Bildes schnellstmöglich per HTTPS in einem TCP-Stream ausgeführt.

Die aus Testsicht relevante Eigenschaft des Inhalts ist die Datenmenge (Dateigröße) eines genutzten Bildes. Mit den mobilen Geräten des Testbeds wurden vorab mehrere Fotos aufgenommen. Die Dateigröße schwankte dabei in Abhängigkeit der eingestellten Auflösung, der gewählten Bildqualität und der Komplexität des Motivs. Deshalb erzeugten die Probanden die Fotos im Test nicht selbst, sondern nutzten vorgegebene Dateien. Es wurden vier Motive ausgewählt, mit vergleichbarer und typischer Datenmenge („Handyfoto“) von ca. 2,6 MB pro Datei, zwei für den Download, zwei für den Upload. Bei korrekter Aufgabenausführung ist die transferierte Datenmenge somit vorab recht genau abschätzbar, variiert nur gering und ist daher direkt vergleichbar. Aus der Dateigröße des Bildes und effektiver Datenrate (Download bzw. Upload) ergeben sich die benötigten Übertragungsdauern.

Das Gerät Galaxy Ace 2 nimmt eine technisch begründete Sonderstellung ein. Beim Vergleich der Messdaten ergab sich, dass mit diesem Gerät die effektiven Upload-Datenraten in den Presets D40 und D50 deutlich niedriger ausfielen als beim Rest des Testfeldes. Die Analyse der aufgezeichneten Verkehrsdaten zeigte, dass das Problem auf TCP-Ebene angesiedelt ist. Das signalisierte TCP-Window ist bei Nutzung des Ace 2 sehr klein. Es kann immer nur eine geringe Menge unbestätigter Daten vom Endgerät zum Cloud-

Server unterwegs sein. Die Bestätigungsdauer ist durch die eingestellte mobilfunktypische Latenz vergleichsweise hoch, was letztlich die effektive Upload-Datenrate limitiert. Es ist weder ein allgemeines Problem des Gerätes, wie der Vergleich mit Testmessungen zeigt (siehe erreichbare Effektivwerte im Anhang), noch der App, denn auf den anderen Android-Geräten ist die Datenrate im erwarteten Bereich, noch der Gegenseite (Google-Server) generell, denn andere zeitgleiche Tests waren unauffällig. Der Effekt entsteht in der Kombination aus App, Gerät und Gegenseite mit den damit einhergehenden genannten Nebenfaktoren und der Netzparametrisierung als ungünstige Gesamtkonstellation. Dies zeigt überdeutlich, dass die Effektivwerte ebenso wie die Nutzerstimuli und damit unmittelbar verbunden die QoE-Bewertungen von dem Zusammenwirken einer Vielzahl von Variablen – mit gegenseitiger Beeinflussung – abhängig sind. Da nicht zweifelsfrei zu klären war, ob dies auch auf eine echte Nutzung im Mobilfunk zutrifft (unterschiedliches Stack-Verhalten) oder mit Besonderheiten des Testbeds zu begründen ist, werden für die Auswertung die Ergebnisse des Ace 2 deshalb als gesonderte (atypische) Partition auf Ebene der Effektivwerte behandelt.

9.1.3 Nutzersicht

Die Präsentation der Nutzerstimuli ist auf den genutzten Plattformen in den getesteten Apps sehr unterschiedlich (siehe Abbildung 121 im Anhang). Ausgehend von der Nutzersicht konnte vermutet werden, dass sich die Aspekte der Fortschrittsanzeige und Sichtbarkeit der objektiven Dateigröße als Hinweis auf Quantität und Qualität auf die Bewertungen der beiden mobilen Plattformen im Test auswirken.

Unter *Android* ist in der App ein Fortschrittsbalken erkennbar, der die ungefähr verbleibende Zeit erkennen lässt. Zusätzlich zur Prozentschätzung des Fortschritts ist auch die Dateigröße ersichtlich. So ist für den Probanden erkennbar, dass es sich um eine große Datei (hochauflösendes Foto) handelt. Eventuell honorieren dies die Probanden bei ihrer Bewertung auf Basis eigener Erfahrungen oder erwarteter Dauern aus Überschlagsrechnungen. Beim Download sind die Angaben direkt im Vordergrund sichtbar, beim Upload lediglich im Hintergrund im System-Statusbereich.

Unter *iOS* ist beim Download nur ein animierter Ladekreis sichtbar, der den Nutzern keinerlei Auskunft darüber gibt, wie der aktuelle Fortschritt ist, bzw. wie lange es wahrscheinlich noch dauert. Beim Upload ist in der iPhone-Variante ein nicht beschrifteter Ladebalken sichtbar, beim iPad nur ein Ladekreis ohne Indikation des Fortschritts. Weder beim Download noch beim Upload ist für den Nutzer unter *iOS* ersichtlich, wie groß die zu übertragende Datenmenge ist.

Die Präsentation der Nutzstimuli ist in der getesteten Android-Version demnach wesentlich näher an den HCI-Empfehlungen zur Usability als die getestete *iOS*-Version, insbesondere bei der Rückmeldung des Fortschritts von längeren Dauern über 10 s (siehe Erläuterung in 3.6.12).

Die Datenmenge der übertragenen Bilder bestimmt die Inhaltseigenschaften und die Wahrnehmung in Form von Bildgröße und Bildqualität. Die subjektive Bildqualität kann beim Zoomen in das heruntergeladene Bild als Handyfoto mit hoher Auflösung wahrgenommen werden.

Tabelle 52: Parametrisierung und Erwartungswerte für Drive

QoS-Netzparameter	Einflussprognose <i>Richtung & Stärke, Charakteristik</i>	Presets				
		D10	D20	D30	D40	D50
max. Download (kBit/s)	+++; elastisch, für Download-Phase	384	900	1800	3600	7200
max. Upload (kBit/s)	+++; elastisch, für Upload-Phase	384	700	1400	2800	5760
min. Latenz (ms)	O	180	90	90	90	50
Ergebnispartitionierung und relevante Einflussfaktoren		erwartete Nutzerstimuli und Charakteristik				
<i>Unterscheidung nach Plattform (Präsentation der Stimuli)</i>		<i>oben: ca. Dauer des vollständigen Downloads (s), elastisch unten: ca. Dauer des vollständigen Uploads (s), elastisch</i>				
Android (ohne Ace 2) Ladebalken mit Angaben zum Fortschritt und Datenmenge		65	29	16	9	5
<i>iOS</i> Ladekreis ohne Angaben zum Fortschritt oder Datenmenge (iPhone: Ladebalken beim Upload)		68	40	23	15	10

Dauern für Download und Upload vorgegebener Bilder mit 2,6MB je Bild

9.1.4 Parametrisierung und Erwartungswerte

Die Parametrisierung der Presets wurde so gewählt, dass ein möglichst weiter Bereich der Nutzerstimuli in Form der beabsichtigten Wartedauern von knapp unter 5 s bis knapp über 1 min abgedeckt wird, siehe Tabelle 52. Beim Upload ergibt sich eine kurze zusätzliche Verzögerung von ca. 2 s durch die Verarbeitung auf der Gegenseite, bis das hochgeladene Bild in der Cloud-Ablage erscheint.

9.1.5 Bearbeitungserfolg

Im Vergleich der Szenarien führte Drive in der Aufgabenbearbeitung recht häufig zu Problemen und auch zu einer recht hohen Zahl invalider Ergebnisse, siehe Tabelle 73 im Anhang. Die Probleme können weitgehend auf die Facette Benutzbarkeit zurückgeführt werden. Aufgabenbeschreibung, App- und Gerätebedienung wurden trotz des gewählten einfachen Szenarios überdurchschnittlich oft als Problempunkte angegeben. Diese Punkte können als unabhängig von den Netzparametern angesehen werden.

Das langsame Preset D10 verursachte drei der sieben Angaben zu Verbindungsproblemen. Alle nicht beendeten Aufgabenausführungen waren invalid, da entweder kein Download oder kein Upload ausgeführt wurde (Verifizierung durch Analyse der Verkehrsdaten). Bei der Wiederholung der Aufgabenbearbeitung können bei Drive Verzerrungen in der Bewertung auftreten, falls z. B. versehentlich zwei Bilder statt nur eines Bildes übertragen wurden. Aufgrund der unerwartet hohen Zahl von Problemen bei der Aufgabenbearbeitung wurde zunächst eine Stichprobe der Verkehrsaufzeichnungen auf korrekte Ausführung geprüft. Dabei zeigte sich die Notwendigkeit einer Komplettkontrolle auf Basis der aufgezeichneten Verkehrsdaten ex post (Tiefenauswertung Stufe 1). Es wurde dabei überprüft, ob die Aufgabenausführungen die beabsichtigten Datenübertragungen (Bild-Download und Bild-Upload) umfassten. Alle abweichenden Ausführungen bzw. Teile davon wurden als invalid markiert. Bei Drive ist nur von einer validen Bewertung der Gesamtzufriedenheit auszugehen, wenn beide Teilbewertungen ebenfalls valid sind.

9.1.6 Nutzerkommentare

Die Probanden können die Auswirkungen großer Unterschiede in der Parametrisierung deutlich erkennen und kommentieren diese entsprechend, siehe Abbildung 84. Die Propositionen zu den relevanten Nutzerstimuli können bestätigt werden. Die Übertragungsdauern des Downloads und Uploads treten deutlich hervor und werden in unterschiedlicher Art und Weise verbal bewertet. Die Wahrnehmung der Eigenschaften als (Datei-) „Größe“ und das Fortschreiten der Übertragung als „Ladebalken“ werden mit den Netzeigenschaften der „Internetverbindung“ in Zusammenhang gebracht.

Tabelle 53: Sentiment-Analyse in Drive

Wort	Gewicht
negativ	
langsam	1,00
längere	0,95
lange	0,89
länger	0,79
fb	0,73
dauerten	0,70
Sekunden	0,64
Ladezeit	0,62
langsamer	0,59
viel	0,57
Ladezeiten	0,47
dauert	0,45
lang	0,44
gedauert	0,34
positiv	
schneller	1,00
schnell	0,46
zufrieden	0,39

Sortierung absteigend nach Gewicht

Begrenzung der Liste auf einzelne Worte mit einem Gewicht > 0,3
 Ein hohes Gewicht entsteht im Zusammenhang mit häufiger Nennung und entsprechender negativer bzw. positiver Konnotation ohne konträre Konnotation.



Abbildung 84: Word-Cloud aus den Nutzerkommentaren zu Drive

Als Besonderheit im Vergleich zu den anderen Szenarien wird für Drive noch das Ergebnis der Sentiment-Analyse präsentiert, siehe Tabelle 53. Die Probanden haben 73 von 420 Ausgabeausführungen mit einem Freitext kommentiert. Diese 73 Kommentare wurden in der Auswertung hinsichtlich ihrer Grundtendenz untersucht und als positiv, neutral oder negativ eingestuft und anschließend per *Rapidminer* Sentiment Analysis ausgewertet:

- eindeutig bestimmte Grundstimmung: positiv 12, negativ 36
- der Rest: unbestimmte Grundstimmung (neutral, meist Anmerkungen zur App, zum Gerät oder zur Aufgabe)

Die in der Word-Cloud prominent vertretenen „Upload“ und „Download“ tauchen in dieser Analyse nicht auf, da sie gleichermaßen mit positiven wie auch negativen Äußerungen in Zusammenhang stehen. Adjektive zur Charakterisierung der Übertragungsdauern bzw. deren Komparativ, also im Vergleich der Aufgabeausführung erstellte Beobachtungen zur Netzgeschwindigkeit bzw. den Ladedauern, führen die Negativ-/Positiv-Listen an.

9.1.7 Bewertungsverhalten

Die Aufgabe funktionierte hinsichtlich des Bewertungsverhaltens insgesamt wie geplant und die Mehrzahl der Bewertungen ist erwartungskonform, siehe Abbildung 85a. Einzelne Ausreißer sind zu erkennen. In der Überprüfung konnten diese Abweichungen als invalide Aufgabeausführungen identifiziert werden. Bei wertgleicher Wiederholung bewerten die Probanden überwiegend gleich. Die Aussagen zur Gesamtzufriedenheit können auf die Teilbewertungen übertragen werden. Die Gesamtzufriedenheit fällt leicht besser aus als die Teilbewertungen. Für ca. zwei Drittel der Ergebnisse erklären die Teilbewertungen die Gesamtzufriedenheit vollständig, siehe Abbildung 85b. Unterschiede von mehr als einer Bewertungsstufe nach oben oder unten sind selten. Es existieren auch sehr seltene Extremfälle, bei denen die Bewertungen stark divergieren (invalide Aufgabeausführung). Die Dauer des Uploads wird im Vergleich etwas schlechter bewertet. Eine mögliche technische Erklärung liefern die Abweichungen des Upload-Effektivwerts beim Ace 2. Für beide Teilbewertungen lassen sich QoE-Änderungen durch QoS-Änderungen erklären. Es ist davon auszugehen, dass beide Items zur Gesamtzufriedenheit beitragen.

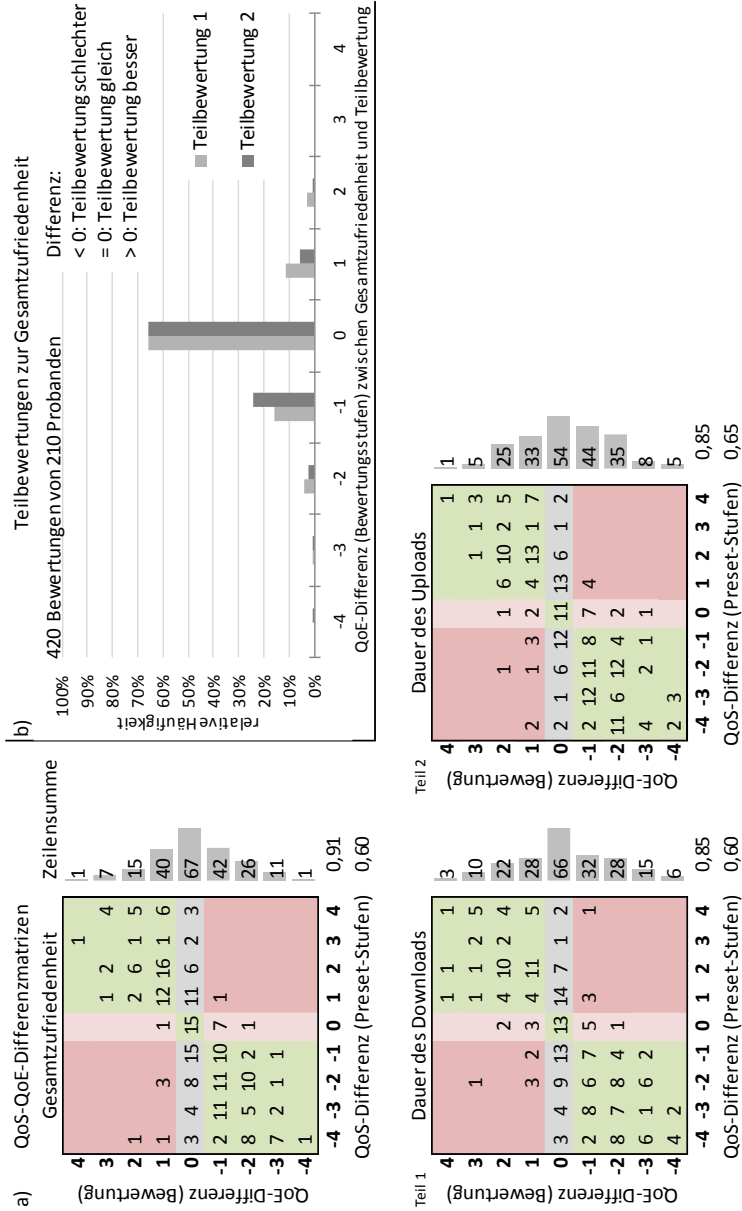


Abbildung 85: Zusammenstellung zum Bewertungsverhalten in Drive

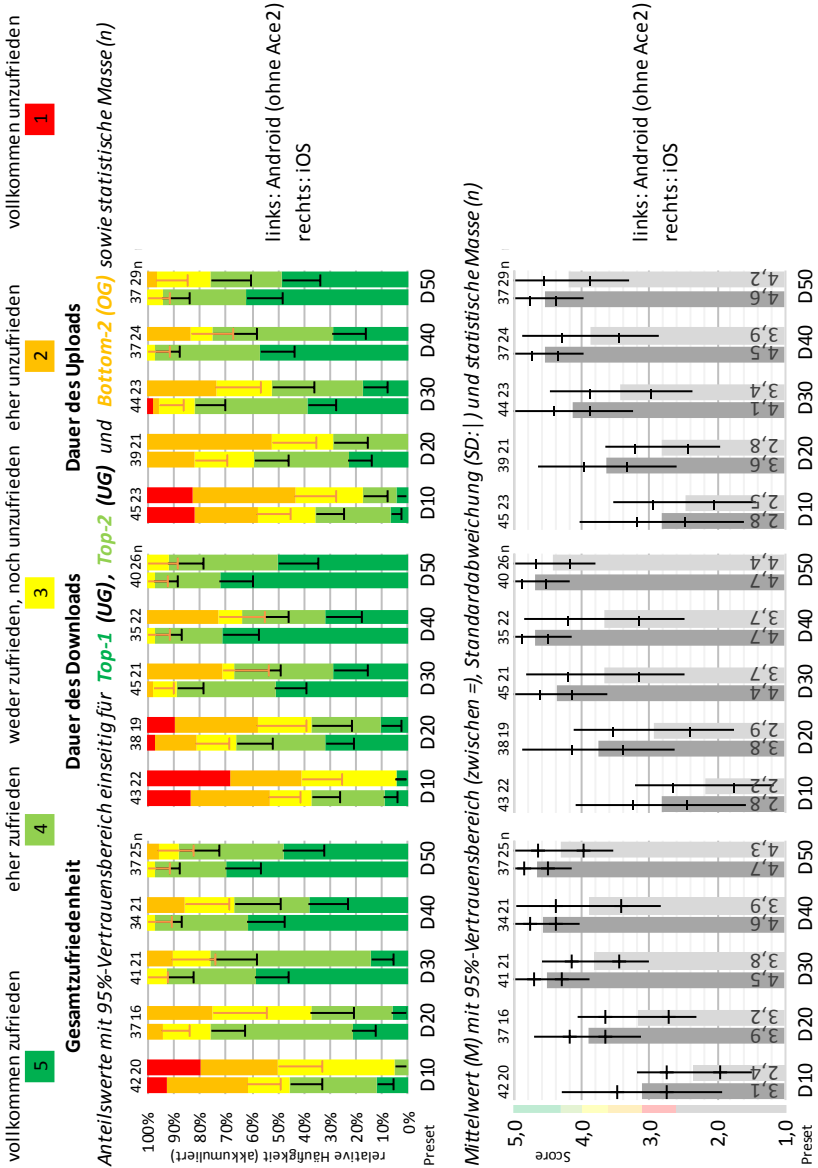


Abbildung 86: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Drive

9.1.8 Bedingte Verteilungen der Zufriedenheit

Die in der Voruntersuchung festgestellten Unterschiede in der Nutzersicht spiegeln sich den Ergebnissen wider, siehe Abbildung 86. Die Ausführungen unter iOS werden teilweise deutlich schlechter beurteilt als unter Android. Die Unterschiede in den Bewertungen zwischen iOS und Android lassen sich nur über Unterschiede in der Wahrnehmung der objektiv gleichen Nutzerstimuli bedingt durch unterschiedliche Präsentationen oder Erwartungshaltungen dazu erklären. Die Nutzerkommentare weisen in die Richtung, dass die Nutzer die Inhaltseigenschaft Bildgröße / Datenmenge mit in ihre Beurteilung einfließen lassen. Bei gleichem Erleben der (Netz-) Leistung modulieren demnach wie vermutet die Wahrnehmung des Fortschritts und die Kenntnis der Dateigröße die Erwartung und damit die Zufriedenheit. Ein „großes“ Bild darf entsprechend auch länger bei der Übertragung dauern, wenn dies dem Nutzer bewusst ist und er eine passende Rückmeldung zum Fortschritt erhält. Fehlen diese Angaben in der Präsentation der Nutzerstimuli, dann sind die Nutzer weniger geduldig und entsprechend deutlich kritischer. Für die Gesamtzufriedenheit bedeutet dies, dass das langsame Preset D10 unter iOS eine der schlechtesten Bewertungen im gesamten Testparcours erhält, während das schnellste Preset D50 nur knapp so gut bewertet wird wie D30 unter Android. Zur Erinnerung: D50 bedeutet im Vergleich zu D40 den doppelten technischen Aufwand hinsichtlich der Datenraten für Download und Upload bzw. zu D30 sogar den vierfachen Wert! Drive ist ein schönes Beispiel, wie bei vergleichbarer technischer Sicht und vergleichbaren Werten der Stimuli (QoD und QoA) die Unterschiede in der Präsentation für die Quality of Experience eine bedeutende Rolle spielen können. Ergänzend sei auf den Vergleich von Android (ohne Ace 2) zum Ace 2 als schon zuvor genutztes Beispiel zur Illustration verwiesen, siehe Abbildung 79.

9.2 Facebook

9.2.1 Nutzungsszenario

Facebook ist eine der populärsten Plattformen für den Upload und das Teilen von Bildern. Schätzungen gehen von ca. 350 Mio. Bild-Uploads pro Tag im Jahr 2013 aus [340] S. 14. Ein Gutteil davon dürfte auf die mobile Nutzung entfallen, weil moderne Smartphones, durch Kamerafunktion und ubiquitärem Internetzugang in einem Gerät vereint, die Nutzung sehr bequem und vor allem sofort ermöglichen. Dafür spricht, dass 68% der gesamten Zugriffe auf Facebook durch mobile Geräte erfolgen [341]²⁰⁸ S. 7. Gemäß aktueller Erhebungen verbringen die Nutzer mobiler Apps die meiste Zeit insgesamt mit der Facebook-App (US-Studie, 2014, siehe [14]).

Im Szenario soll Facebook genutzt werden, um tolle Fotos mit den Freunden zu teilen. Dies geschieht bequem in der mobilen Facebook App. Der Nutzer öffnet die App und checkt auf seiner Pinnwand²⁰⁹ die eigenen Einträge und Fotos sowie die seiner Freunde. Er postet ein neues Handy-Foto auf Facebook und wartet zur Überprüfung anschließend wiederum die Aktualisierung der Pinnwand mit dem soeben hochgeladenen Foto ab. Gegenüber einer Nutzung ohne Upload eines Bildes ist dabei von höheren Anforderungen an die notwendige Netzparametrisierung auszugehen.

9.2.2 Technische Sicht und Netznutzung

Für die detaillierte Betrachtung ist eine Unterscheidung in die Phasen des Downloads und Uploads sowie weiterer Besonderheiten sinnvoll.

Der Download der Bilder ist hauptsächlich verantwortlich für entstehende Wartezeiten beim Aktualisieren der Pinnwand. Die Übertragung erfolgt dabei von Servern in einem Content Distribution Network (CDN) per HTTPS/TCP. Nach den Analysen der Verkehrsaufzeichnungen werden scheinbar zunächst bis zu sechs Bilder geladen. Weitere Bilder werden erst beim Scrollen nach unten zu weiteren Einträgen oder zur Seite (bei Bildserien) dynamisch nachgeladen. Bereits heruntergeladene Bilder verbleiben in einem geräteinternen

²⁰⁸ mit Bezug auf eine Studie von comScore, US, 18+, Dez. 2013

²⁰⁹ Pinnwand an dieser Stelle und nachfolgend als noch häufig genutztes Synonym für die Facebook-Chronik

Cache und müssen nicht erneut übertragen werden. Die beobachtete Bildgröße beträgt zwischen 70 und 110kB. Zwischen die Einträge mit den Bildern der Nutzer können ggf. auch Werbeeinblendungen platziert werden. Der Startzeitpunkt der Bildaktualisierung kann gegenüber dem Start der App leicht verzögert sein (Abgleich der Unterschiede zwischen App und Server).

Die Bildgröße bestimmt die entstehende Wartezeit beim Upload. Die Facebook-App beinhaltet eine lokale Vorverarbeitung der Bilder mit Verkleinerung der Bildabmessungen eines typischen Handyfotos²¹⁰ inklusive einer Kompression zur Verminderung der Datenmenge, die allerdings unterschiedlich auf den beiden genutzten Plattformen ausfällt. iOS komprimiert die Bilddaten deutlich stärker als Android (Faktor 4 bis 6 stärker). So lässt sich im Test trotz vorgegebener vergleichbarer Ausgangsbilder (2,6 MB) ein Unterschied in den übertragenen Datenmengen nicht vermeiden (Android ca. 500 kB, iOS ca. 110 kB je Bild).

Der Bild-Upload selbst erfolgt an einen Facebook-Server per HTTPS in einem TCP-Stream. Die Unterschiede in den Datenmengen sind durch die Apps verursacht, resultieren aber ggf. aus den Vorgaben der genutzten API-Versionen auf den Facebook-Servern. Nach einem erfolgreichen Upload eines Bildes wird automatisch die Pinnwand aktualisiert (ggf. leicht zeitverzögert), was einem Bild-Download entspricht.

Die Facebook-App ist ein schönes Beispiel für schnelle Veränderungen bzw. für den Versuch einer Momentaufnahme eines beweglichen Zieles. Zum Ende der Testrunde 2 hat Facebook wohl auch die Datenmenge beim Upload unter Android weiter reduziert, siehe Abbildung 87. Für die Realisierung der Probandenstudie war dies unkritisch (Redaktionsschluss der Apps vor Testbeginn), wäre aber eventuell bei einer zeitlichen Ausdehnung kritisch geworden, vor allem mit Blick auf eine API-Änderung, die eine Weiternutzung der App-Version im Test beendet hätte.

²¹⁰ Diskussion dazu siehe Drive

- a) 
- b) 
- c) 

Abbildung 87: Meldungen zu neueren Versionen der Facebook-App unter Android im Verlauf der Testrunde 2 Screenshots, Ausschnitte zur Verdeutlichung

9.2.3 Nutzersicht

Für die Nutzerzufriedenheit werden die folgenden Stimuli und ihre Präsentation als relevant angenommen, für die gesonderte Teilbewertungen erhoben werden:

- wahrgenommene Dauer zum Laden/Aktualisieren der Pinnwand (nach dem Start der App und Upload eines Bildes)
- wahrgenommene Dauer für Upload des Bildes (plus Aktualisieren der Pinnwand danach)

In der Nutzersicht treten Unterschiede zwischen den mobilen Plattformen deutlich in Erscheinung, siehe Abbildung 123 im Anhang. Die Apps unterscheiden sich beim Upload hinsichtlich des wahrnehmbaren Starts, Endes und der Dauer der eigentlichen Datenübertragung sowie in der Präsentation des Fortschritts.

Unter Android startet die eigentliche Datenübertragung zum Upload eines Bildes erst nach dem Schritt zur Eingabe einer Bildbeschreibung. Der Fortschrittsbalken erscheint dabei nachfolgend nur im Benachrichtigungsbereich. Für Verwirrung des Nutzers sorgen dabei missverständliche Fortschrittsinfor-

mationen, siehe Abbildung 123.6 im Anhang, die auf ein Übersetzungsproblem hindeuten. So wird mit „1 von 1 wurden hochgeladen“ bereits zum Start der Übertragung ein falscher Status suggeriert. Zudem springt der Fortschrittsbalken nur in relativen groben Stufen von je ca. 20 %, so dass ggf. das (vermeintliche) Ende der Übertragung, allein darauf bezogen, vage bleibt. Die Probanden erhielten deshalb in der Aufgabenbeschreibung einen expliziten Hinweis auf die Darstellung zum wahren Ende des Uploads. Die getestete Version der Facebook-App unter Android ist damit leider ein Beispiel für eine echte aber schlechte Beobachtbarkeit in der Nutzersicht durch Usability-Probleme.

Unter iOS beginnt der Upload bereits nach der Auswahl des Bildes im Hintergrund, noch während der Nutzer mit dem Kommentieren des Bildes beschäftigt ist. So ist zumindest ein Teil des Bildes bereits hochgeladen, wenn der Nutzer vermutlich meint, mit „Posten“ den Upload zu starten. In der nachfolgenden Ansicht springt der direkt sichtbare Fortschrittsbalken dann entsprechend schnell auf den bereits erreichten Wert bzw. ist unter Umständen nur sehr kurz überhaupt sichtbar. Aus Sicht der Nutzer sollte dies positiv sein, da sich die anschließend wahrgenommene Wartedauer (allerdings im Vordergrund) verkürzt darstellt.

Zusammen mit der geringeren Datenmenge bei iOS sind die tatsächliche und die wahrgenommene Wartezeit geringer als bei Android, dort allerdings beim Upload im Hintergrund, so dass Prognosen zur Wirkung insgesamt schwierig sind. Zudem erscheint unter iOS relativ zeitnah nach dem Ende des Uploads das hochgeladene Bild auf der Pinnwand, bei Android etwas verzögert, was an zusätzlichen Verarbeitungsschritten (erneute Kompression) auf den Facebook-Servern liegen dürfte.

Anders als bei Drive erfordert die Nachbildung eines einigermaßen „echten“ Nutzungsszenarios in Facebook mit Größen zu arbeiten, die sich nicht exakt vorhersagen lassen. Diese sind im Modell als Störfaktoren zu sehen. Die Ergebnisse sind als eine Mischung (durchschnittlicher Fall) der verschiedenen Ausprägungen anzusehen. Dies trifft für die Bildanzahl und damit die Wartedauer beim Aktualisieren der Pinnwand zu. Eine jeweils gleiche Bildanzahl wäre zwar planerisch schön ist aber eben nicht „echt“, weil ein echter Nutzer

vorab ja auch nicht beeinflussen kann, wie viele Bilder seine Facebook-Freunde in der Zwischenzeit seit der letzten Aktualisierung neu hinzugefügt haben. Durch die Randomisierung sollte sich der Effekt im Test ausmitteln, also ungefähr gleichverteilt auf die unterschiedlichen Netzqualitätsstufen auftreten.

Die Datenmenge der übertragenen Bilder bestimmt die Inhaltseigenschaften und damit die benötigten Übertragungsdauern und visuelle Qualität. Der Proband hat allerdings keine Möglichkeit die Datenmenge einzusehen oder auf sie einzuwirken. Eine modellkonforme Annahme zu Qualität und Quantität des Inhalts ist, dass die Bildanzahl in die Bewertung einfließt. Dies bedeutet, dass für wenige Bilder eine kürzere Wartezeit akzeptiert wird und für eine größere Bildanzahl auch eine etwas längere Wartezeit, weil für jedes einzelne Bild eine gewisse Wartezeit akzeptiert wird (bis zu einer gewissen Obergrenze je Bild und in Summe).

Dieses Szenario zeigt deutlich, dass es in Labortests schwierig ist, die Balance zu finden zwischen exakter Reproduzierbarkeit und dem Anspruch einer realistischen Nutzung mobiler Services. Eine realistische Nutzung kann immer eine gewisse verbleibende Unsicherheit bzw. Unschärfe mit sich bringen.

9.2.4 Parametrisierung und Erwartungswerte

Die Parametrisierung der Presets wurde so gewählt, dass ein möglichst weiter Bereich der Nutzerstimuli in Form der beabsichtigten Wartedauern abgedeckt wird, siehe Tabelle 54. Zum Download sind keine Angaben zu einzelnen Erwartungswerten möglich, da diese von den o. g. zusätzlichen Faktoren abhängig sind. Daher erfolgt eine Angabe nur je einzeltem Bild. Bedingt durch die Unterschiede in der übertragenen Datenmenge beim Upload sind Android und iOS um ca. 2 Stufen versetzt. Beim Upload ergibt sich eine Verzögerung durch die Verarbeitung auf der Gegenseite und eine weitere Wartezeit, bis das hochgeladene Bild erscheint, die wiederum mindestens der Dauer für einen Bild-Download entspricht. Der Einfluss der Gegenseite erklärt auch, dass die tatsächlich zu erwartenden Werte höher ausfallen als die theoretischen Minimalwerte. Für die Parametrisierung F10 unter Android war zu erwarten, dass einige Probanden beim Upload die Geduld verlieren und dies als Fehler in der Funktion der App oder der Internetverbindung interpretieren.

Tabelle 54: Parametrisierung und Erwartungswerte für Facebook

QoS-Netzparameter	Einflussprognose <i>Richtung & Stärke, Charakteristik</i>	Presets				
		F10	F20	F30	F40	F50
max. Download (kBit/s)	+++; elastisch, für Download-Phase	80	128	225	384	900
max. Upload (kBit/s)	+++; elastisch, für Upload-Phase	48	96	192	384	700
min. Latenz (ms)	0	300	300	300	180	90
Ergebnispartitionierung und relevante Einflussfaktoren		erwartete Nutzerstimuli und Charakteristik				
Download						
<i>keine fixierte Partitionierung (Gegenseite durch Bildanzahl)</i>		<i>ca. Dauer des Downloads je Bild (s), elastisch</i>				
		(11)	(7)	(4,5)	(3)	(1,5)
Upload						
<i>Unterscheidung nach App (Plattform) durch Quantität und Sichtbarkeit</i>		<i>ca. Dauer des Uploads eines Bildes* (s), elastisch</i>				
Android	Quantität Inhalt: ca. 500kB Ladebalken im Hintergrund	87	46	26	15	10
iOS	Quantität Inhalt: ca. 110kB Ladebalken im Vordergrund	29	17	12	9	7

* beobachtet, mit Berücksichtigung der Antwort der Gegenseite (nach interner Verarbeitung)

9.2.5 Bearbeitungserfolg

Wiederholungen und Abbrüche der Bearbeitungen sind im Zusammenhang mit den variierten Netzparametern zu sehen, besonders bei Preset F10 tritt dieser Grund sehr deutlich hervor. Die Probanden gingen von einem Problem mit der Internetverbindung aus, wenn selbst nach längerer Wartezeit noch kein Ergebnis zu erzielen war. Dies entspricht den erwarteten Ergebnissen bei dieser Parametrisierung. Hinzukommen, wie die Angabe „Sonstiges“ offenbarte, dass die Probanden gelegentlich Probleme hatten, den Upload der Bilder generell abzuschließen, besonders bei den langsamen Presets F10 und F20 in Kombination mit Android. Eventuell ist es dabei bedingt durch die langsame Übertragung zu (ungeplanten) Timeouts auf Server-Seite gekommen. Ebenfalls auf Android und die beiden langsamen Presets F10 und F20 beziehen sich Angaben, dass die hochgeladenen Bilder nachfolgend nicht angezeigt werden, was bedeuten könnte, dass die Probanden nicht geduldig genug waren. Eventuell werden von den Probanden die beabsichtigten langen Wartezeiten mit unbeabsichtigten Funktionseffekten in der Wahrnehmung vermischt. Da diese

Wiederholungen und Abbrüche aus Sicht der Probanden ein Problem mit der Internetverbindung sind, sollten auch diese Aufgabenausführungen in der Ergebnisdarstellung berücksichtigt werden.

9.2.6 Nutzerkommentare

Die Annahmen zu den relevanten Stimuli und den identifizierten Problemen (siehe auch Nutzersicht und Bearbeitungserfolg) können bestätigt werden, siehe Abbildung 88. Die Nutzer beziehen zumindest in Einzelfällen die benötigte Dauer auf die erwartete Datenmenge. In Facebook ist diese nicht unmittelbar ersichtlich, daher sind einige Probanden verunsichert:

- „Kann schlecht einschätzen, welches Datenvolumen das Foto gehabt hat. Eine genaue aussage über die Uploadgeschwindigkeit ist schwierig.“
- „Ich weiß ehrlich gesagt nicht, wie zufrieden ich mit der Dauer des Uploads bin, da mir nirgends angezeigt wird, wie groß die Datei war, die ich eben hochgeladen habe.“

Dies dürfte hauptsächlich für unerfahrene Nutzer (bezogen auf Facebook) gültig sein. Erfahrene Nutzer sollten wissen, dass auch in einer echten Nutzung im Mobilfunk keine derartigen Bezugsgrößen verfügbar sind.



Abbildung 88: Word-Cloud aus den Nutzerkommentaren zu Facebook

9.2.7 Bewertungsverhalten

Bei der Korrelation zwischen Variation der Preset-Stufen und Variation der Bewertungsänderung der Teilbewertung 1 zeigt sich, dass diese deutlich niedriger liegt als für die Teilbewertung 2, siehe Abbildung 89a. Insgesamt wird der gewünschte Variationseffekt aber mit der richtigen Tendenz zwischen QoS-Änderung und erwarteter QoE-Änderung trotzdem erzielt. In der Teilbewertung 1 zeigen sich verhältnismäßig viele Bewertungsänderungen, die sich nicht direkt aus der Netzparametrisierung erklären lassen, sondern auf die o. g. Nebenfaktoren zurückzuführen sind. Für beide Teilbewertungen lassen sich QoE-Änderungen durch QoS-Änderungen erklären. Die Übereinstimmung zwischen Gesamtzufriedenheit und Teilbewertungen ist sehr hoch, siehe Abbildung 89b. Es ist davon auszugehen, dass beide Teilbewertungen zur Gesamtzufriedenheit beitragen.

9.2.8 Bedingte Verteilungen der Zufriedenheit

Entgegen der ursprünglichen Annahme ist zwischen den Plattformen bei der Bewertung der Dauer des Foto-Uploads der Unterschied in den Bewertungen, mit Ausnahme von Preset F20, eher schwächer zu verzeichnen als vermutet, siehe Abbildung 90. Eventuell sind die dargelegten Besonderheiten der Präsentation der Nutzerstimuli eine mögliche Ursache, die sich allerdings auf Basis des erhobenen Datenmaterials nicht weiter klären lässt. Es ist daher für das entwickelte Szenario nur sinnvoll, von einer gemeinsamen Ergebnismenge für Facebook auszugehen und nicht weiter nach einzelnen Nebenfaktoren zu differenzieren. Dafür wären weitere Untersuchungen mit verfeinerter Spezifikation des Nutzungsszenarios mit Ausschaltung oder besserer Kontrolle der genannten Nebenfaktoren notwendig. So ist durch den Einfluss der Gegenseite und die Randomisierung der Bildanzahl das gewählte Szenario ein Beispiel eines nicht perfekten Grey-Box-Modells. Dabei konnten nicht alle Faktoren genau fixiert oder spezifiziert werden, weil eine gewisse Realitätsnähe gefordert ist. Die grundlegenden Zusammenhänge können aber bestätigt werden.

Die gemeinsame Darstellung der Mittelwerte als Mittelwertkurven ist der Zusammenfassung am Ende des Kapitels zu entnehmen, auf die Darstellung der gemeinsamen Anteilswerte wird aufgrund der strukturellen Ähnlichkeit der beiden betrachteten Teilmengen verzichtet.

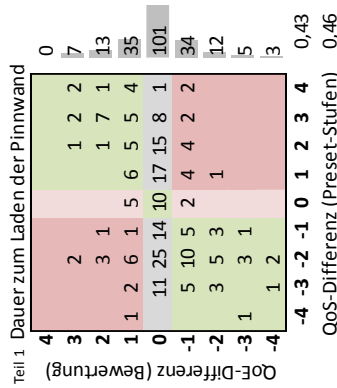
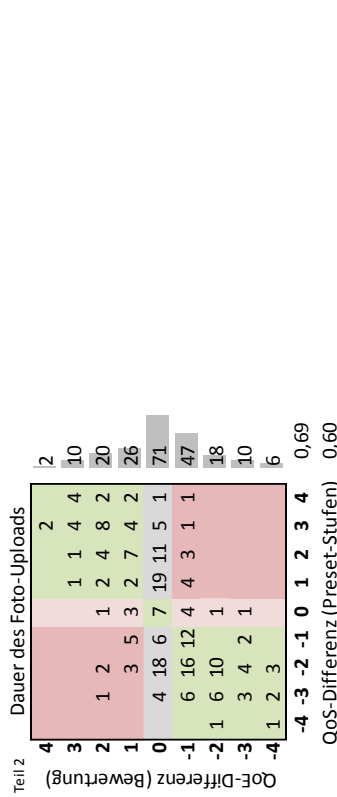
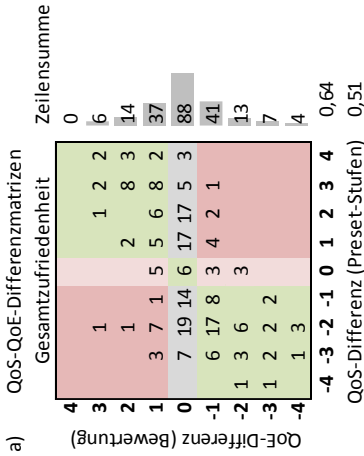
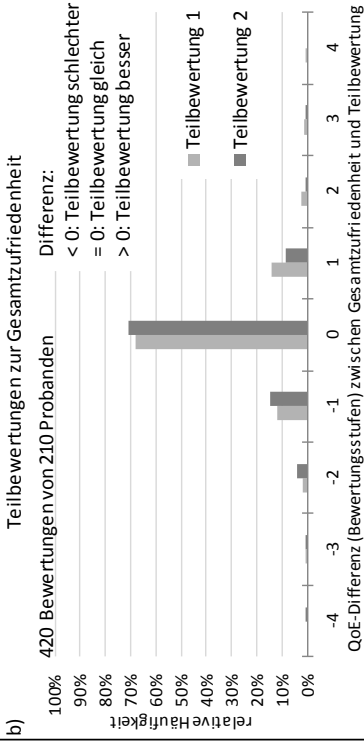


Abbildung 89: Zusammenstellung zum Bewertungsverhalten in Facebook

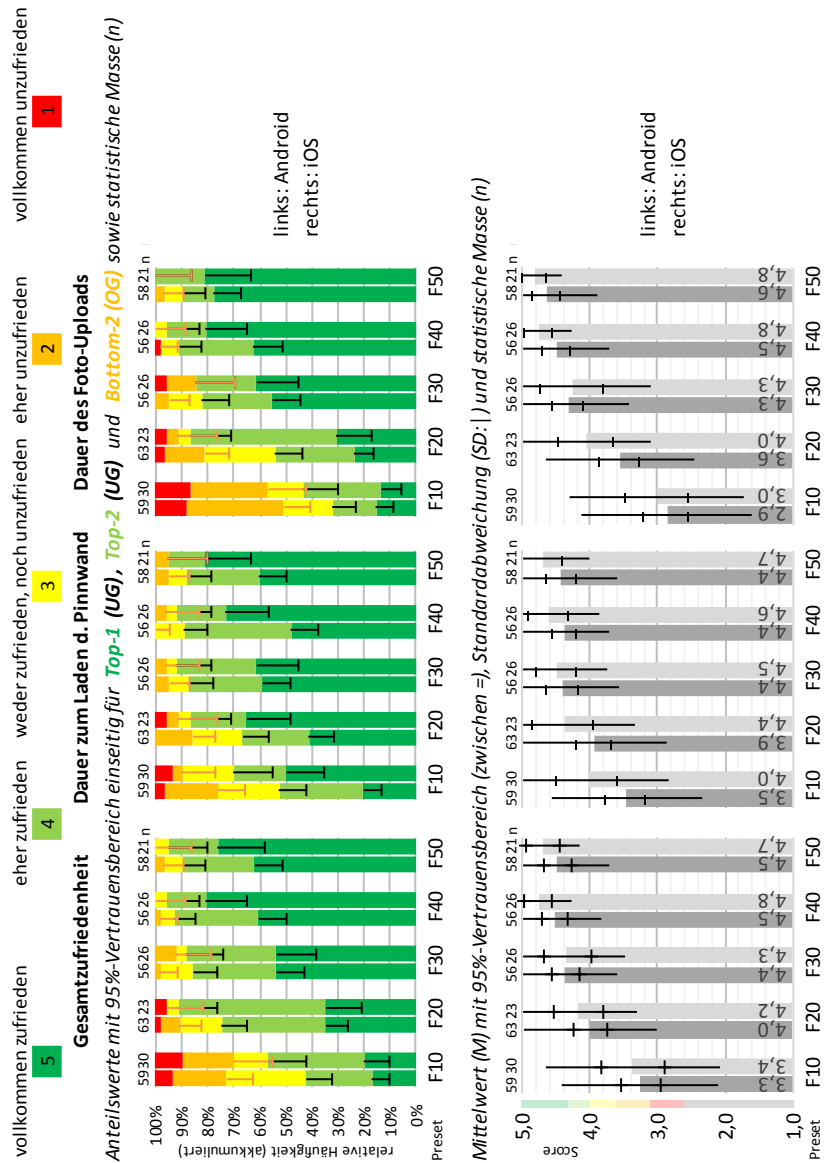


Abbildung 90: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Facebook

9.3 Google Maps

9.3.1 Nutzungsszenario

Karten- und Navigationsdienste spielen für mobile Nutzer eine bedeutende Rolle, sind sie doch Dienste, die mit der Mobilität der Geräte und Nutzer perfekt harmonieren. Entsprechend dieser hohen Relevanz wurde ein mögliches und anspruchsvolles, aber realistisches Szenario entworfen. Als konkreter Dienst wurde Google Maps ausgewählt, da weit verbreitet und zumindest in leicht unterschiedlichen Ausführungen auf den relevanten Plattformen verfügbar.

In Google Maps soll vom aktuellen Standort aus eine Route mit dem Auto zu einem Ausflugsziel geplant werden. Da die Route zuvor noch nicht gefahren wurde, sollen zur späteren Orientierung die Strecke und insbesondere die Anfahrt zum Ziel auf der Karte vorgesichtet werden. Im Zielgebiet wird die nähere Umgebung erkundet, so dass schnell das nächste Restaurant auf der Karte zu lokalisieren ist, um dort vor der geplanten Freizeitaktivität noch zu speisen. Dies ist eine realistische Vorstellung auch für vergleichbare Szenarien. Entweder will man sich da schnell orientieren, wo man gerade ist, z. B. in einer fremden Stadt nach dem Verlassen des Bahnhofs, oder im Detail dort, wo das Ziel liegt. Eine reine Routenplanung und Übersichtsdarstellung sind wenig anspruchsvoll, und wahrscheinlich auch wenig aussagekräftig. Erst der Wunsch einen Kartenabschnitt in hoher Detailfülle darzustellen inklusive schneller Zooms und Kartenbewegungen verspricht, zu deutlichen Unterschieden in Abhängigkeit der QoS-Netzparametrisierung zu gelangen.

9.3.2 Technische Sicht und Netznutzung

In den betrachteten Versionen nutzt Google Maps gekachelte Pixelkarten. Der jeweils sichtbare Kartenausschnitt ist dabei aus mehreren Einzelbildern zusammengesetzt. Es existieren verschiedene Zoomstufen mit unterschiedlichem Detailgrad, entsprechend sind die Kartenkacheln dreidimensional adressiert (geografische Länge und Breite sowie die Zoomstufe).

Im Unterschied zu früheren iOS-Versionen ist Google Maps ab iOS-Version 6 nicht mehr Bestandteil des Lieferumfangs der Apple-Geräte, sondern wurde durch Apple Maps ersetzt. Im Unterschied zu Google Maps nutzt Apple Maps

stufenlose Vektorkarten. Die Anwendungen sind damit nicht vergleichbar. Es wurde daher unter iOS die Google Maps App von Google nachinstalliert. Die iPad-App war zur Zeit der Testplanung und des Testbeginns noch nicht verfügbar. Die für die iPhones konzipierte Version sah auf dem iPad 4 ungewöhnlich aus, entweder als Minidarstellung oder als unscharfe Vergrößerung, wodurch die Nutzersicht (Präsentation) ungewollt negativ beeinträchtigt gewesen wäre. Daher wurde die Nutzung von Google Maps im Browser auf dem iPad 4 als Kompromiss ausgewählt, was für die schnelle Nutzung ad hoc durchaus realistisch erscheint.

Das erzeugte Datenvolumen für vergleichbare Aufgabenbearbeitungen unterscheidet sich zwischen den Endgeräten bzw. Apps im Test, siehe Tabelle 55.

Tabelle 55: Datenvolumen je Aufgabenausführung in Abhängigkeit von App und Gerät in Google Maps

Gerät	Plattform	Bildschirmauflösung	App	Inhalt Kacheln	Datenvolumen (ca.)
Ace 2	Android	800x480	in Google Maps App	min. 8, max. 15 bei 256x256 px	4150 kB
iPhone 5	iOS	1136x640	in Google Maps App	min. 15, max. 24 bei 256x256 px	4500 kB
Galaxy SIII	Android	1280x720	in Google Maps App	min. 15, max. 24 bei 256x256 px	8800 kB
iPad 4	iOS	2048x1536	im Browser (Safari)	min. 12, max. 20 bei 512x512 px	14500 kB

Datenvolumen (Größenordnung) gemittelt für eine geplante Aufgabenausführung in einem der ausgewählten Szenarien; Galaxy Tab 2 10.1 mit vergleichbarer Bildschirmauflösung wie das Galaxy S III; Kacheln: gleichzeitig sichtbar (Anzahl) und Abmessungen

Eine mögliche Erklärung ist die unterschiedliche Bildschirmauflösung der einzelnen Geräte, was bei fixierter Kachelgröße zu einer unterschiedlichen Zahl gleichzeitig dargestellter Kartenkacheln je Zoomstufe führt. Dabei sind auch größere Kartenkacheln auf dem iPad 4 denkbar, um die Zahl geladener Einzelkacheln zu reduzieren, was letztlich aber zum selben Effekt des höheren Datenvolumens führt. Zusätzlich ist zumindest im Browser zu beobachten, dass Abschnitte der Karten auch bei gleicher Zoomstufe durch unterschiedliche Kacheln abgebildet werden können, was wiederum zu zusätzlichen Datenübertragungen führt. Insgesamt ist festzuhalten, dass die Kombination aus Endgerät und Nutzung als App oder im Browser zu Unterschieden in der Netznutzung durch unterschiedliche Datenvolumina führt. Dieser Umstand sollte

bei den weiteren Ausführungen beachtet werden, wenn verkürzt zwischen App als der Nutzung auf den Geräten mit geringerer Bildschirmauflösung – was auch das Galaxy Tab 2 10.1 Android-Tablet einschließt – und Nutzung im Browser, was nur für das hochauflösende iPad 4 zutrifft, gesprochen wird.

Google Maps weist eine starke Download-Orientierung auf. Die Kartenkacheln werden nach Bedarf für den sichtbaren Kartenausschnitt vom Server geladen und bleiben im Cache der App lokal gespeichert. Der Download der Kacheln erfolgt mehrstufig sequenziell und parallel, jeweils ausgelöst durch eine direkte Nutzerinteraktion mit der Kartendarstellung, die neue Kartenkacheln bedingt (Verschieben oder Zoomen). Dabei werden mehrere Kartenkacheln innerhalb eines TCP-Streams übertragen. Durch die genutzte Verschlüsselung (SSL/TLS) war nicht eindeutig zu bestimmen, ob HTTP(S) oder auch SPDY in der Applikationsschicht genutzt wurde, was aber nachfolgend für die Interpretation ohne weitere Bedeutung ist.

Die Nutzung des Caches war in der Aufgabenplanung zu beachten. Es wurden deshalb zwei technisch vergleichbare Varianten mit unterschiedlichen Zielen innerhalb eines Testlaufs genutzt. Zwischen den Testläufen wurde der Cache für jeweils gleiche Ausgangsbedingungen geleert (neues, zuvor noch nicht angesteuertes Ziel). Die Google-Maps-App beinhaltet die Gefahr, dass eine mögliche Vorladefunktion bei WLAN-Nutzung das Ergebnis beeinflusst. In der Voruntersuchung war dieser Aspekt nicht problematisch, ein automatisches Vorladen der Karte wurde trotz WLAN nicht beobachtet.

9.3.3 Nutzersicht

Als wichtigste QoS-veränderliche Reizgröße konnte die Dauer bis zur vollständigen Darstellung des gewünschten Kartenausschnitts in der gewünschten Detailstufe nach jedem Zoom- oder Verschiebeschritt identifiziert werden. Eine Annahme dabei ist, dass einige Stufen und Verschiebungen bei schneller Ausführung wohl nicht wahrgenommen werden, sondern nur die jeweils letzte Aktion. Die vorherigen Nutzeraktionen beeinflussen dabei unter Umständen die jeweils letzte Nutzeraktion in der Art, dass die vorherigen Kartenkacheln noch übertragen werden, auch wenn sie schon nicht mehr im sichtbaren Bildbereich sind.

Google Maps ist im beschriebenen Szenario mit besonders interaktiver Ausprägung ein Sonderfall der Dienste, die durch die Wartezeit definiert sind. Für eine flüssige Bedingung ist zu vermuten, dass der Nutzer eine sehr schnelle bzw. fast unmittelbare Reaktion auf seine Aktionen erwartet – im Prinzip noch während er mit dem Finger auf dem Touchscreen ruht. Daher sind schon fast Anforderungen der Klasse Echtzeit anzunehmen. [342] ordnet Google Maps entsprechend den Diensten mit einer engen Kontrollschleife zu. Google Maps auf Touch-Geräten wie dem iPad & Co. vermittelt demnach den Nutzern, im Vergleich zur gewöhnlichen Nutzung im Desktop-Browser, ein noch stärkeres Gefühl von Interaktivität aus der direkten Kopplung von Eingabe und Ausgabe. Dabei wird der Begriff der Pliability (Flexibilität/Biegsamkeit) als Nutzerwunsch mit in die Quality of Experience einbezogen ([342] mit Verweis auf [343]). Dies ist bei Google Maps sehr anschaulich. Über die eigenen Aktionen kann der Nutzer unmittelbar und direkt die Erscheinung der Karte formen, quasi als Selbstwirkungserfahrung. Google Maps ist damit ein gutes Beispiel für das direkte Zusammenwirken von „Hand“ als Auslöser der Nutzerstimuli, „Auge“ zur Beobachtung und „Hirn“ zum Abgleich von interaktiv erfahrener Leistung als Erlebnis und der individuellen Erwartung zur Bildung der QoE-Bewertungen.

Beim Vergleich der Nutzung als App bzw. im Browser auf dem iPad4 waren Unterschiede in der Nutzersicht beim Nachladen von Kartenkacheln zu erkennen. In der App erfolgt zunächst eine unscharfe Darstellung mit einer Kachel einer anderen Zoomstufe. Im Browser ist ebenfalls zunächst eine unscharfe Darstellung oder (wesentlich häufiger) ein Loch in der Kartendarstellung mit grauem Hintergrund anzutreffen, siehe Abbildung 125 (Schritt 5, iPad 4).

9.3.4 Parametrisierung und Erwartungswerte

Durch die Ausführung der Aufgabe beeinflussen die Probanden selbst die Anzahl paralleler Downloads von Kartenkacheln sowie die Gesamtzahl von Kacheln und somit auch die unmittelbar wahrgenommenen Wartezeiten bis zu einem bestimmten Ausführungsschritt. Entsprechend lässt sich nur schlecht eine quantitative Vorabschätzung erstellen bzw. konkrete Erwartungswerte angeben. Die Nutzer sollten die einzelnen Schritte zügig abarbeiten, besonders von der Routenübersicht zum Zoomen am Zielort. Dabei gilt allgemein: bei

exakt gleicher Aufgabenausführung ist ein Proband mit schneller Netzparametrisierung deutlich „schneller am Ziel“ als wenn die Bedienung immer wieder durch mehr oder minder lange Wartezeiten unterbrochen wird. Bei niedriger Parametrisierung ist möglich, dass die Probanden die Bearbeitung weniger detailliert ausführen, d. h. die Zoom- und Verschiebeschritte auf das absolut notwendige Maß zur Bearbeitung der Aufgabe beschränken. Es sollte davon auszugehen sein, dass sich der Einfluss der Probanden insgesamt ausmittelt (akribisch vs. oberflächlich).

Das Laden der Route und die Routendarstellung vom Start zum Ziel auf der Übersichtskarte sind recht anspruchslos, entsprechend sollte dieser Teil der Aufgabe eher wie ein Kontrollwert agieren. Erst der Wunsch einen Kartenabschnitt in hoher Detailfülle darzustellen, inklusive schneller Zooms und Schieben des Kartenausschnitts, ist deutlicher anspruchsvoller. Dies ist auch eine realistische Nutzung, die eine deutliche QoE-Differenzierung ausgehend von unterschiedlichen QoS-Parametereinstellungen erwarten lässt. Von G10 mit deutlichen Wartezeiten zum Aktualisieren der Kartendarstellung bis zu G50, das eine flüssige Nutzung erlaubt, auch im Browser auf dem hochauflösenden iPad 4 – quasi ohne Wartezeiten oder Absetzen der Finger – wird ein weiterer Bereich für die Wartezeiten abgedeckt, siehe Tabelle 56.

Für das Gefühl der Pliability darf die erlebte Verzögerung nicht zu groß sein. Dies ist nur dann gewährleistet, wenn die Datenmenge der zur Darstellung benötigten Kartenkacheln je Nutzeraktion (Zooming/Panning) innerhalb eines gewissen Zeitfensters transportiert werden kann. Die HCI-Forschung gibt hierfür die Anhaltspunkte (stark interaktive Nutzung: ca. $< 1s$). Dies bedeutet für die QoS-Netzparameter, dass die Größenordnung der effektiven Latenz schon relevant wird. Insgesamt darf die Download-Richtung über die Zeit hinweg gesehen nicht ständig gesättigt sein, um einen Bufferbloat zu vermeiden, der zwangsläufig zu einer höheren effektiven Latenz führt. Mit dieser Heuristik ist eine grobe Abschätzung der notwendigen Download-Datenrate und noch vertretbarer Latenz möglich.

Tabelle 56: Parametrisierung und Erwartungswerte für Google Maps

QoS-Netzparameter	Einflussprognose <i>Richtung & Stärke, Charakteristik</i>	Presets				
		G10	G20	G30	G40	G50
max. Download (kBit/s)	+++ , elastisch	384	900	1800	3600	7200
max. Upload (kBit/s)	o	384	700	1400	2800	5760
min. Latenz (ms)	- bis -- , elastisch	180	90	90	90	50
Ergebnispartitionierung und relevante Einflussfaktoren		erwartete Nutzerstimuli und Charakteristik				
<i>Unterscheidung nach Endgerät und App (Gesamtdatenmenge und Präsentation)</i>		<i>Dauer für Kartendarstellung, elastisch</i>				
Endgerät App vs. Browser		keine Angabe von Erwartungswerten möglich				
		<i>theoretisch minimale Zeit (s) bei vollständiger Datenmenge</i>				
iPhone 5, App		104	45	23	12	6
iPad 4, Browser		336	144	72	36	18

Auswahl der Geräte stellvertretend für das Testbed und zur Verdeutlichung der ungefähren Bereiche

9.3.5 Bearbeitungserfolg

Im Vergleich der Aufgaben verursachte Google Maps insgesamt eine mittlere Anzahl Probleme bei der Bearbeitung. Wiederholungen und Abbrüche der Bearbeitungen sind bei Google Maps in engem Zusammenhang mit den variierten Netzparametern zu sehen. Acht der neun Problemangaben beziehen sich auf das langsamste Preset G10. Wie zu erwarten war, verlässt bei derartig langsamem Netz einige Probanden die Geduld, vor allem bei der Nutzung im Browser auf dem iPad 4. So zeigte sich in einigen der exemplarischen Videoüberwachungen, dass die Nutzer dann schnell die Bearbeitung abgebrochen haben, wenn nach der ersten Verschiebeaktion quasi die gesamte Kartendarstellung nachgeladen werden musste und dies (scheinbar) nicht voranging.

Mit Blick auf die Bewertung sind Wiederholungen einer Aufgabenausführung in Google Maps kritisch zu sehen. Im Wiederholungsschritt ist das Laden von vorgepufferten Kartenkacheln möglich. Dies beschleunigt die Aufgabenausführung u. U. gegenüber einer ununterbrochenen Ausführung. Die Verwertung dieser Aufgabenausführungen in der Ergebnisdarstellung muss entsprechend geprüft werden, z. B. im Abgleich mit der intrapersonellen Bewertungskonformität. In der Videobeobachtung zeigte sich, dass bei der Nutzung im Browser (iPad4) viele Nutzer recht lange brauchten, um von der

Routenliste zur Kartendarstellung zu wechseln, da sie Schwierigkeiten hatten, das richtige Icon sofort zu finden. Dies kann als QoS-unabhängiges Usability-Problem der App gewertet werden, was ggf. zu einigen Aufgabenabbrüchen oder Wiederholungen geführt hat.

9.3.6 Nutzerkommentare

Die Annahmen zu relevanten Nutzerstimuli und Einschränkungen werden durch die Nutzerkommentare bestätigt, siehe Abbildung 91. Dabei treten die Anmerkungen zur interaktiven Arbeit mit der Karte und insbesondere beim Zoomen und (Nach-) Laden der Kartendarstellung wie erwartet deutlich heraus. Neben diesen Kommentaren zur Zufriedenheit mit den Stimuli sind Anmerkungen zur Usability der App zu finden, vor allem von den Nutzern im Browser. Auch Elemente aus der Aufgabenstellung sind zu erkennen, z. B. zum Finden der Point-of-Interests im Zielgebiet, wie die Suche nach einem Restaurant, die die Nutzer zu einem Kommentar veranlassten.



Abbildung 91: Word-Cloud aus den Nutzerkommentaren zu Google Maps

9.3.7 Bewertungsverhalten

Die Mehrzahl der QoE-Änderungen liegt im erwarteten Bereich, siehe Abbildung 92a. Einige wenige vom Trend abweichende Bewertungen sind festzustellen, können aber überwiegend auf die Probanden selbst zurückzuführen sein (unterschiedliche Ausführungsgeschwindigkeit und Akribie bei der Ausführung). Wie erwartet, ist kaum eine deutliche Änderung der Bewertung zur Routenberechnung zu verzeichnen (wahrnehmbare Änderungen bleiben ohne Auswirkungen). Das Laden der Route wird in fast 50 % der Fälle besser bewertet als die Gesamtzufriedenheit, siehe Abbildung 92b, ist also deutlich weniger anspruchsvoll und entsprechend über den gesamten Parameterbereich wenig aussagekräftig. Aus den geringen Differenzen der Reizgröße dazu resultieren auch nur geringe QoE-Änderungen, siehe auch Abbildung 92a Teilbewertung 1, wobei über 73 % der Probanden ihre Bewertung unabhängig von der QoS-Änderung beibehalten haben. Für das Zoomen der Karte kehrt sich die Situation um, es wird auffällig oft schlechter bewertet als die Gesamtzufriedenheit. Die fast symmetrische Ausprägung der Abweichungen der Teilbewertungen nach oben bzw. unten könnte als wechselseitige Kompensation interpretiert werden. Eventuell wird die bessere Teilbewertung 1 zusammen mit der schlechteren Teilbewertung 2 zur Gesamtzufriedenheit gemittelt.

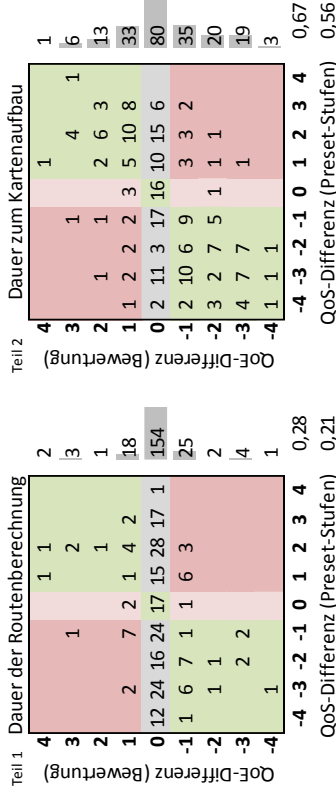
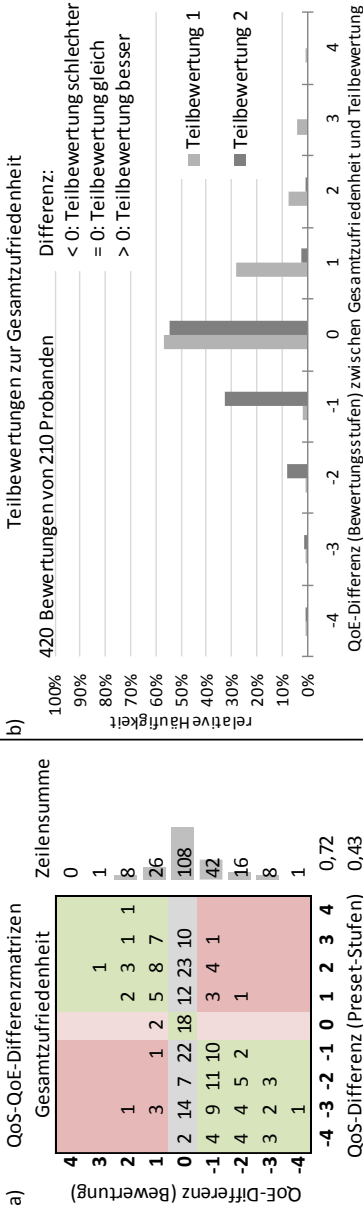


Abbildung 92: Zusammenstellung zum Bewertungsverhalten in Google Maps

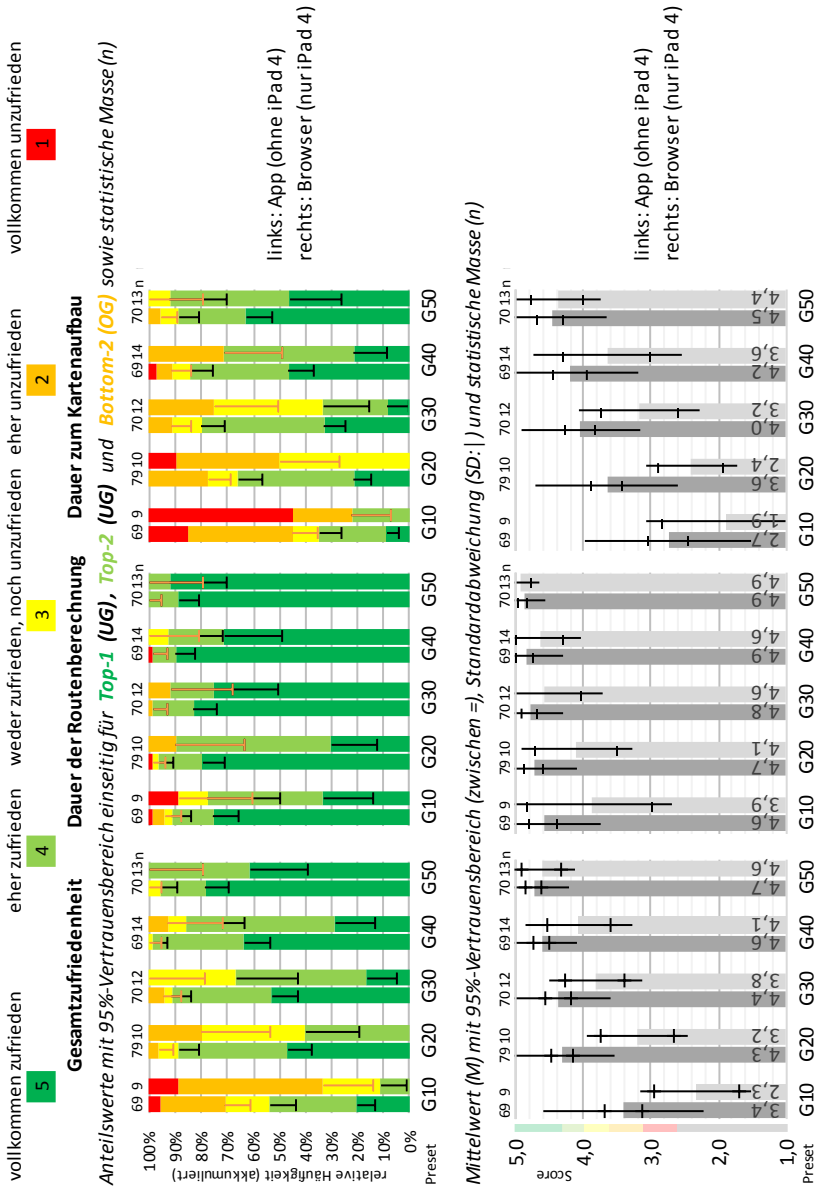


Abbildung 93: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Google Maps (Teil 1) Vergleich der Nutzung von Google Maps als App bzw. im Browser

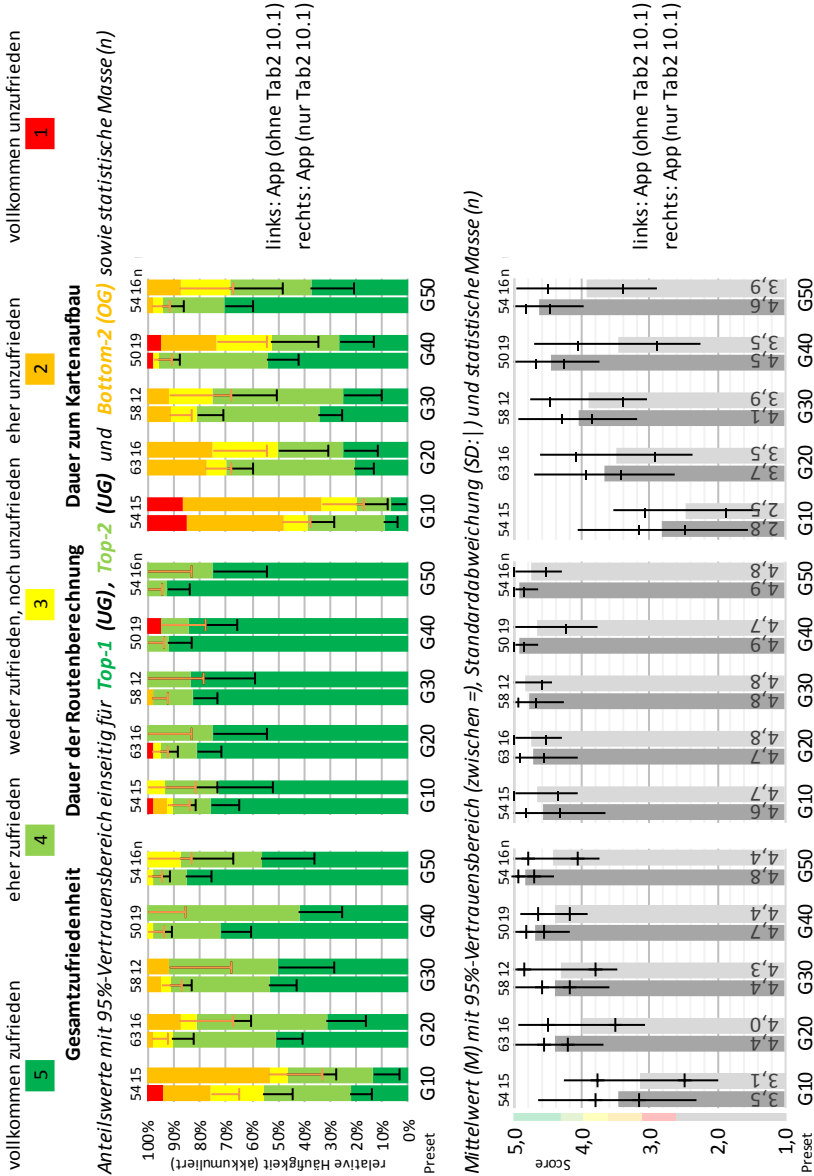


Abbildung 94: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Google Maps (Teil 2) Vergleich der App-Nutzung in Abhängigkeit zum Endgerät

9.3.8 Bedingte Verteilungen der Zufriedenheit

Durch die Beschränkung der Nutzung im Browser auf ein einzelnes Gerät ist diese Teilstichprobe insgesamt recht klein. Trotzdem lässt sich klar erkennen, dass die Probanden die Nutzung auf den restlichen Geräten (App) deutlich besser bewerten als auf dem iPad 4 (Browser). Erst mit G50 nähern sich die Teilergebnismengen auf einem sehr hohen Zufriedenheitsniveau an, die Nutzung im Browser auf dem iPad 4 bleibt trotzdem einen Tick schlechter.

Die Dauer der Routenberechnung ist nur im Browser ein Kriterium, das vom Übergang von niedrigen Presets zum mittleren Preset auf hohem Niveau zulegt. Die Erhöhung der Zufriedenheit beim Kartenaufbau ist hingegen ein mustergültiges Beispiel für einen beinahe gleichmäßigen Zugewinn an (voller) Zufriedenheit.

In Abhängigkeit des jeweiligen Interaktionskontextes (App und Endgerät) sind am oberen Ende des gewählten Parameterbereiches entweder nur noch geringe QoE-Änderungen zu verzeichnen (Nutzung via App im Sättigungsbereich) oder noch deutliche Änderungen zu erkennen (iPad 4 mit Browser).

Beim wichtigen Punkt des Kartenaufbaus bleibt das Android Tablet Tab2 10.1 ab Preset G40 hinter den restlichen Geräten mit App-Nutzung zurück, siehe Abbildung 94. Eine mögliche Erklärung ist, dass die gefühlte Geschwindigkeit bei der Bedienung durch die Faktoren Hardware und Software des Endgerätes unabhängig von der QoS-Parametrisierung beschränkt wird. Dies deckt sich mit den subjektiven Beobachtungen zum genannten Gerät aus den internen Vorabtests. (Als Stichprobe eines einzelnen Gerätes ist die statistische Masse und damit die Aussagekraft allerdings limitiert.) Die unterschiedliche Startansicht der Tablet-App nach der Routenberechnung, welche einige zusätzliche Zoom- und Verschiebeschritte bedingt, ist hingegen keine direkt plausible Erklärung. Würde dies zutreffen, dann sollten sich diese Unterschiede auch und vor allem bei den langsameren Presets G10 bis G30 zeigen, was allerdings so nicht festzustellen ist.

Google Maps ist damit insgesamt ein gutes Beispiel, wie der der Interaktionskontext aus App und Endgerät auf die Quality of Experience einwirken kann und damit den QoS-QoE-Zusammenhang eines Dienstes moderiert.

9.4 MTV-Music

*“Applications are typically developed in a ‘golden’ environment.”
Mike Canney in [344]²¹¹ S. 20*

9.4.1 Nutzungsszenario

Die Nutzung mobiler Endgeräte zur Wiedergabe von Musik unterwegs ist populär. Dabei verlagert sich die Mediennutzung weg vom Kauf von Musikstücken und der lokalen Speicherung hin zu Musik-Streaming-Diensten mit On-Demand-Zugriff und Wiedergabe aus der Cloud (siehe [345] und [341]²¹² S. 8). Entsprechend ist von einer hohen zukünftigen Relevanz auszugehen. Der ausgewählte Dienst MTV-Music gehört zu Rhapsody, einem der großen Anbieter in diesem Bereich am Markt mit weiteren bekannten Marken (u. a. Napster).

Ein mögliches Szenario ergibt sich wie folgt: Der Nutzer möchte zur eigenen Unterhaltung einen Titel eines bestimmten Künstlers anhören. Dazu durchsucht er gezielt das Angebot und startet den gesuchten Titel aus der Titelliste zur sofortigen Wiedergabe als Streaming ohne vorherigen Download des Musikstücks.

9.4.2 Technische Sicht und Netznutzung

Aus technischer Sicht zerfällt das Szenario in zwei Phasen:

- Browsing-Phase vom Start der Anwendung über die Suche des Titels bis zum Laden der Titelliste
- Streaming-Phase des Musiktitels ab beabsichtigtem Start der Wiedergabe

In der Browsing-Phase werden jeweils einige XML-Dateien (wenige KB) und einige JPG-Bilder (je 50 bis 100 kB) aus einem CDN per HTTP/TCP zum Endgerät übertragen. Im letzten Schritt vor dem Streaming sind beim Laden der Titelliste technische Unterschiede im Datenvolumen zu beobachten. Diese ergeben sich aus der Länge der Titelliste und der App. Im Gegensatz zur Android-App lädt die iOS-App in der Titelliste die Cover-Bilder zu den einzelnen

²¹¹ in einem Vortrag zur Analyse von Leistungsproblemen im Netzwerk...

²¹² mit Verweis auf eine IFPI-Studie, die von 2012 zu 2013 einen Zuwachs von 51% zeigt (US-Markt)

Titeln. Unnötigerweise werden dabei teilweise Bilder sogar mehrfach übertragen, was sich bei langen Titellisten negativ auswirkt.

Das Streaming erfolgt als Pseudo-Streaming per Progressive Download von MP4-Audiodateien aus einem CDN per HTTP/TCP. Der Progressive Download erlaubt, dass die Wiedergabe (automatisch) startet, noch bevor die Datenübertragung vollständig abgeschlossen ist, nachdem der Wiedergabepuffer einen bestimmten Füllstand erreicht hat.

Nach dem vollständigen Download eines Titels erfolgt automatisch der Abruf des jeweils nachfolgenden Titels im Hintergrund (Prefetch). Daher sollten die Probanden die Wiedergabe explizit stoppen und die Anwendung vollständig schließen. Die Möglichkeit der App zu Vorab-Download und Offline-Wiedergabe der Musikstücke wurde nicht genutzt.

Die Musikstücke stehen in zwei Qualitätsstufen bereit, die sich in ihrer Inhaltsdatenrate unterscheiden (im Test wurde die Qualitätsstufe LQ genutzt):

- LQ, 64 kBit/s, CBR (constant bit rate), Vorgabewert für Nutzung per Streaming
- HQ, 192 kBit/s, CBR, Vorgabewert für Nutzung per Vorab-Download

Heruntergeladene Inhalte (Titellisten, Cover-Bilder etc. und Musiktitel) bleiben nach einer Übertragung lokal gespeichert. Entsprechend waren zwei verschiedene Inhaltsvarianten innerhalb eines Testslots und das Rücksetzen des Caches zwischen den Testslots notwendig, um vergleichbare Ausgangsbedingungen zu gewährleisten.

9.4.3 Nutzersicht

Als relevante Nutzerstimuli konnten in der Browsing-Phase die wahrgenommenen Dauern in den einzelnen Schritten zur Suche des angegebenen Titels mit Laden der jeweiligen Ansichten und besonders das Erscheinen der eingebundenen Bilder identifiziert werden. In der Streaming-Phase sind die relevanten Nutzerstimuli die wahrgenommene Dauer bis zum Wiedergabestart des Musikstückes (Tap-to-Play, vom Antippen in der Titelliste zum Losspielen) und das Abspielverhalten mit Problemen bei der Wiedergabe (z. B. Unterbrechungen).

Zum Testzeitpunkt stand unter iOS keine Tablet-Version der App bereit. Die Nutzung der iPhone-App auf dem iPad wurde aber als unproblematisch eingeschätzt, da die Präsentation bei dieser Art App und dem gewählten Nutzungsszenario nicht im Mittelpunkt steht.

Bei MTV-Music gab es in der getesteten Version ein Problem, das am ehesten der konkreten Umsetzung der App zuzuordnen ist. Wird für das Vorpuffern zu viel Zeit benötigt, sind Funktionseffekte zu beobachten, die sich je nach Plattform unterschiedlich auswirken, siehe Abbildung 127 (im Anhang) zur Nutzersicht bei M10 in der Variante A:

- Android: Nach ca. 20 s Wartezeit erscheint eine Fehlermeldung „langsameres Netz“, die Wiedergabe startet aber trotzdem wenige Sekunden später.
- iOS: Ist nach ca. 30 s Wartezeit der automatische Wiedergabestart noch nicht erfolgt, so dauert das Laden scheinbar ewig, weil der automatische Start unterbleibt und der Indikator zum Vorpuffern (Ladekreis) dauerhaft angezeigt wird, auch wenn nach mehreren Minuten Wartezeit der Musiktitel längst vollständig übertragen ist.

Dieses Verhalten war immer dann zu beobachten, wenn die Download-Datenrate sehr niedrig parametrisiert war und aus einer langen Titelliste ein Stück zum Streaming ausgewählt wurde, die Browsing-Phase aber noch nicht vollständig abgeschlossen war, was bei zügiger Aufgabenbearbeitung so eigentlich zu erwarten war. Das Eingangszitat zur „goldenen Entwicklerumgebung“ trifft für MTV-Music in der getesteten Version offenbar perfekt zu. Daraus ergibt sich als Nebenerkenntnis die Empfehlung, bei der Entwicklung mobiler Apps auch ungünstige, langsame oder instabile Netzbedingungen einzuplanen und die Apps vorab intensiv daraufhin auszutesten. Für den Test war die Auswahl des Dienstes damit unter dem Gesichtspunkt der Testeignung zwiespältig einzuschätzen.

9.4.4 Parametrisierung und Erwartungswerte

Die zuvor erwähnten Besonderheiten grenzen den sinnvollen Parameterbereich nach unten hart ab. Die untere Grenze wurde dabei so gewählt, dass bei der Inhaltsvariante A die besonderen Effekte zu beobachten waren (lange Titelliste), bei der Variante B hingegen nicht (kurze Titelliste). Bei höheren Presets sollten die erwähnten Funktionseffekte generell nicht auftreten. Die gewählte niedrigste Parameterstufe entspricht im Download schon dem doppelten Wert der Inhaltsdatenrate der Stufe LQ, Unterbrechungen sind nach dem Wiedergabestart entsprechend nicht zu erwarten. Zur gezielten Provokation von Unterbrechungen der Wiedergabe während des Streamings müsste die Download-Datenrate niedriger als die durchschnittliche Inhaltsdatenrate sein. Dann hätten durch die erwähnte Besonderheit die iOS-Geräte aber gar nichts mehr wiedergeben bzw. wären unter Android extrem lange Vorpufferzeiten zu verzeichnen gewesen, die für sich genommen sicherlich schon zu Unzufriedenheit geführt hätten.

Tabelle 57: Parametrisierung und Erwartungswerte für MTV-Music

QoS-Netzparameter	Einflussprognose <i>Richtung & Stärke, Charakteristik</i>	Presets				
		M10	M20	M30	M40	M50
max. Download (kBit/s)	+++ , sprunghaft M10*, ansonsten elastisch	128	225	384	900	1800
max. Upload (kBit/s)	o	96	192	384	700	1400
min. Latenz (ms)	-, elastisch für Browsing-Phase	300	300	180	90	90
Ergebnispartitionierung und relevante Einflussfaktoren		erwartete Nutzerstimuli und Charakteristik				
<i>Unterscheidung nach</i>		<i>oben: ca. Wiedergabestartzeit (s), elastisch</i>				
<i>Plattform/App</i>	<i>Variante</i>	<i>unten: Abspielverhalten, sprunghaft</i>				
Android Meldung „langsameres Internet“ bei M10.A mit Startzeiten > 20 s	A	25 bis 31 Meldung	14	9	4	<3
	B	16	12	6	3	<3
iOS „ewiges“ Puffern bei M10.A ohne automatischen Start	A	teils ∞ kein Start	25	14	5	<3
	B	24	20	14	6	<3

Unterschied in den Varianten (Künstler und Musikstück) durch Datenmenge in der Titelliste (Cover)

Variante A: David Bowie "Heroes" 6:10 min, Variante B: The Lumineers: "Hey Ho" 2:41 min; Qualitätsstufe LQ des Inhalts (64 kBit/s CBR)

* sprunghaft von M10 Variante A, besonders bei iOS

Bessere Datenratenstufen resultieren in insgesamt kürzeren Wartezeiten. Die iOS-App puffert offenbar etwas mehr Inhalt als die Android-Version vor, weshalb die Ladezeiten unter iOS bei niedrigen Presets (deutlich) höher ausfallen, siehe Tabelle 57. Für die Browsing-Phase und die Dauer zur Suche der angegebenen Titel können keine genauen Erwartungswerte angegeben werden, da diese auch von der Bearbeitungsgeschwindigkeit der Probanden (Eingabe der Suche) abhängen.

9.4.5 Bearbeitungserfolg

MTV-Music bereitete den Probanden eine mittlere Anzahl Probleme bei der Bearbeitung der Aufgabe. Aufgabenabbrüche und Wiederholungen ohne Netzeinfluss entstanden dabei durch Usability-Probleme in der Bedienung der App und aus Verständnisproblemen zur Aufgabenstellung. Die Aufgabenabbrüche und Wiederholungen der Bearbeitungen mit der Problembeschreibung „Internetverbindung“ sind im Zusammenhang mit schlechten Netzparametern zu sehen, vor allem bei M10 in Kombination mit iOS (sechs von acht). Dies entspricht der Erwartung zu den Effekten der Parametrisierung. Die Wiederholung einer Aufgabenbearbeitung ist aus technischer Sicht als möglicherweise kritisch einzuschätzen. Einerseits besteht die Gefahr, dass Inhalte aus dem lokalen Cache der Endgeräte genutzt werden, andererseits entspricht dies dem normalen Verhalten: Das Aufgabenziel wird erreicht, es dauert aber insgesamt länger. Da es nur wenige Aufgabenausführungen sind, die dies betrifft und noch von erwartungskonformen Bewertungen auszugehen ist, können diese Ausführungen im Ergebnisdatensatz verbleiben.

9.4.6 Nutzerkommentare

Die Annahmen zu den relevanten Nutzerstimuli können bestätigt werden, besonders hinsichtlich eines verzögerten oder gar ausbleibenden Wiedergabestarts, siehe Abbildung 95. Über die Inhaltsdatenrate bestimmt die Qualitätsstufe maßgeblich den technisch sinnvollen bzw. als gut bewerteten Parameterbereich. Die Audioqualität des wiedergegebenen Inhalts wurde nicht gesondert bewertet. Die Nutzer erachten in einigen Fällen scheinbar trotzdem Qualitätsaspekte des Inhalts als relevant bzw. zumindest einen Kommentar wert, obwohl nicht explizit danach gefragt (Beispiele dazu im Anhang A.6.2). Wenn sich die Probanden zur Qualität äußerten, dann meist negativ. Inwieweit die Wiedergabequalität in die Bewertung der Gesamtzufriedenheit

einfließt, ggf. auch unbewusst, ist durch die Testkonzeption nicht weiter zu ermitteln. Es wurde nur die Qualitätsstufe LQ getestet, die für Streaming die Standardeinstellung ist. Der Testaufbau und die Testumgebung eigneten sich allerdings auch grundsätzlich nicht für eine realistische und zuverlässige Beurteilung der Audioqualität (billige Kopfhörer und Umgebungsgeräusche der anderen Probanden etc.).



Abbildung 95: Word-Clouds aus den Nutzerkommentaren zu MTV-Music

oben links: allgemeine Kommentare; unten rechts: Kommentare zum Abspielverhalten

9.4.7 Bewertungsverhalten

Insgesamt bewerten die Probanden erwartungskonform, in den Teilbewertungen 1 und 3 aber auffällig häufig ohne Bewertungsänderung (geringe QoE-Differenzierung), siehe Abbildung 96a. Die Teilbewertungen bedürfen im Vergleich gesonderter Erläuterung, da deutliche Unterschiede zu erkennen sind, siehe auch Abbildung 96b.

Die Zufriedenheit mit der Dauer der Suche ist für die Probanden kein Differenzierungskriterium, das mit der QoS-Parametrisierung einhergeht. Die Formulierung des Items war rückblickend nicht optimal, da eigentlich auf den gesamten Suchvorgang der Browsing-Phase bis zur Titelliste bezogen. Zudem waren die erreichten Unterschiede dabei zu gering. Die Bewertungen zum Abspielverhalten sind im Vergleich zu Youtube deutlich weniger differenziert, obwohl beide Dienste prinzipiell der Klasse Streaming zuzuordnen sind. Dies entspricht den Erwartungen, da die Parametrisierung so gewählt werden musste, dass die Wiedergabe überhaupt funktionierte (siehe ewiges Puffern). Entsprechend war während der Streaming-Phase nicht unmittelbar mit Problemen zu rechnen, die sich stark auf die Bewertung zum Abspielverhalten hätten auswirken können. Es war aber im Vorhinein nicht klar, ob die Probanden dies auch so zuordnen. In den Bewertungen zeigt sich, dass die Funktionseffekte wie „ewiges Puffern“ oder Meldung „langsames Internet“ von der Mehrzahl der Probanden der Wartezeit bis zum Wiedergabestart zugerechnet werden und nur von einem kleineren Teil dem Abspielverhalten.

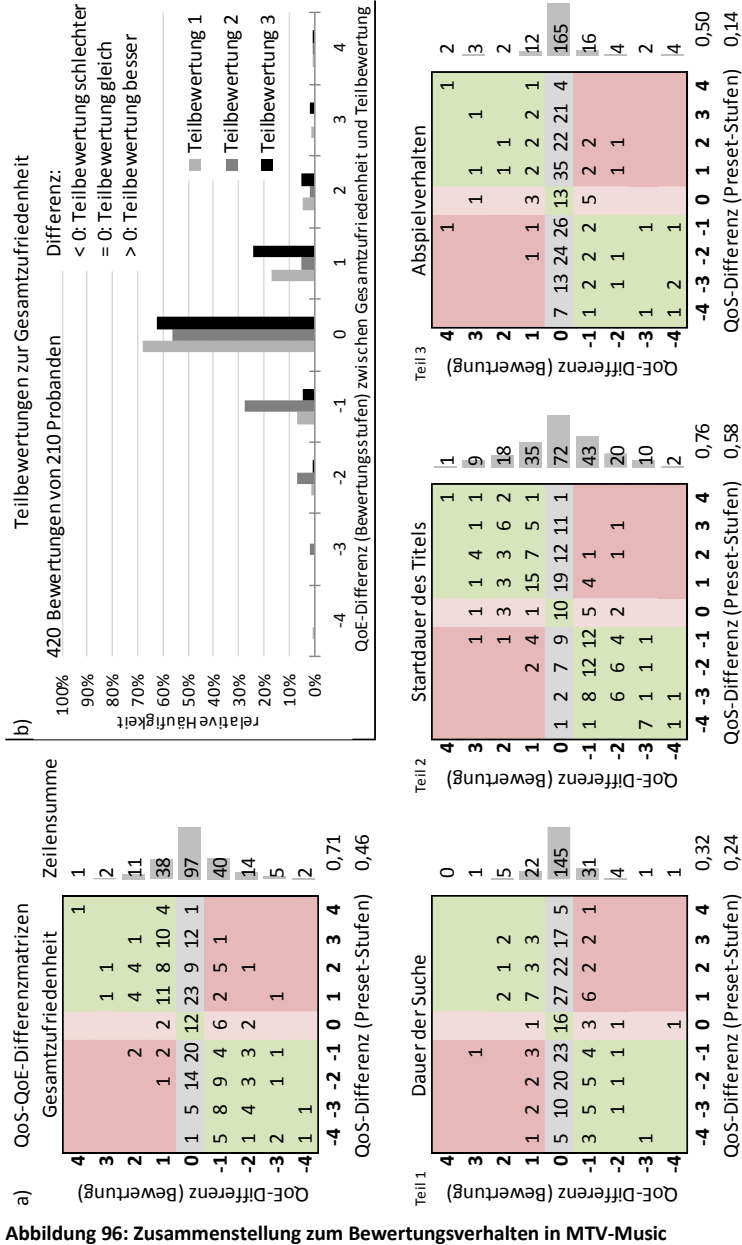


Abbildung 96: Zusammenstellung zum Bewertungsverhalten in MTV-Music

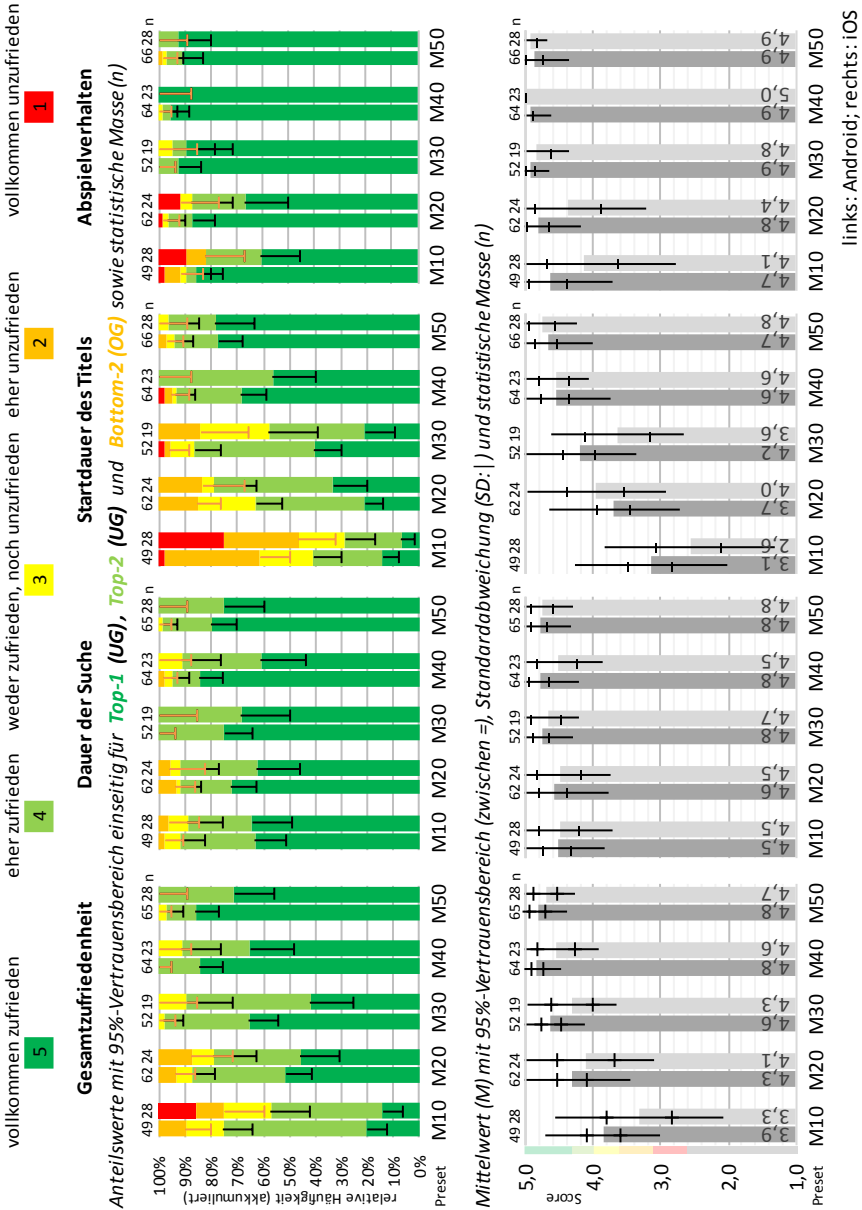


Abbildung 97: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 1) Vergleich der Nutzung unter Android (links) vs. iOS (rechts)

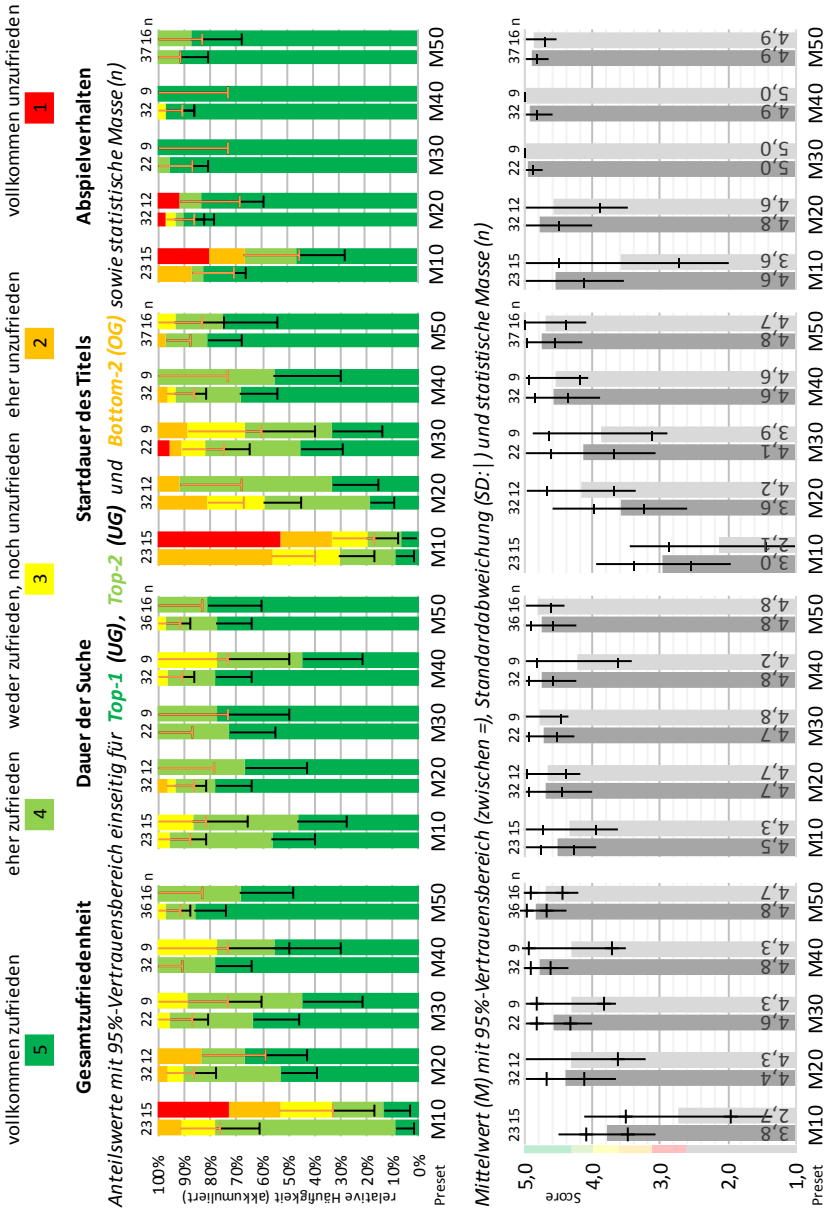


Abbildung 98: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 2) Vergleich der Variante A unter Android (links) vs. iOS (rechts)

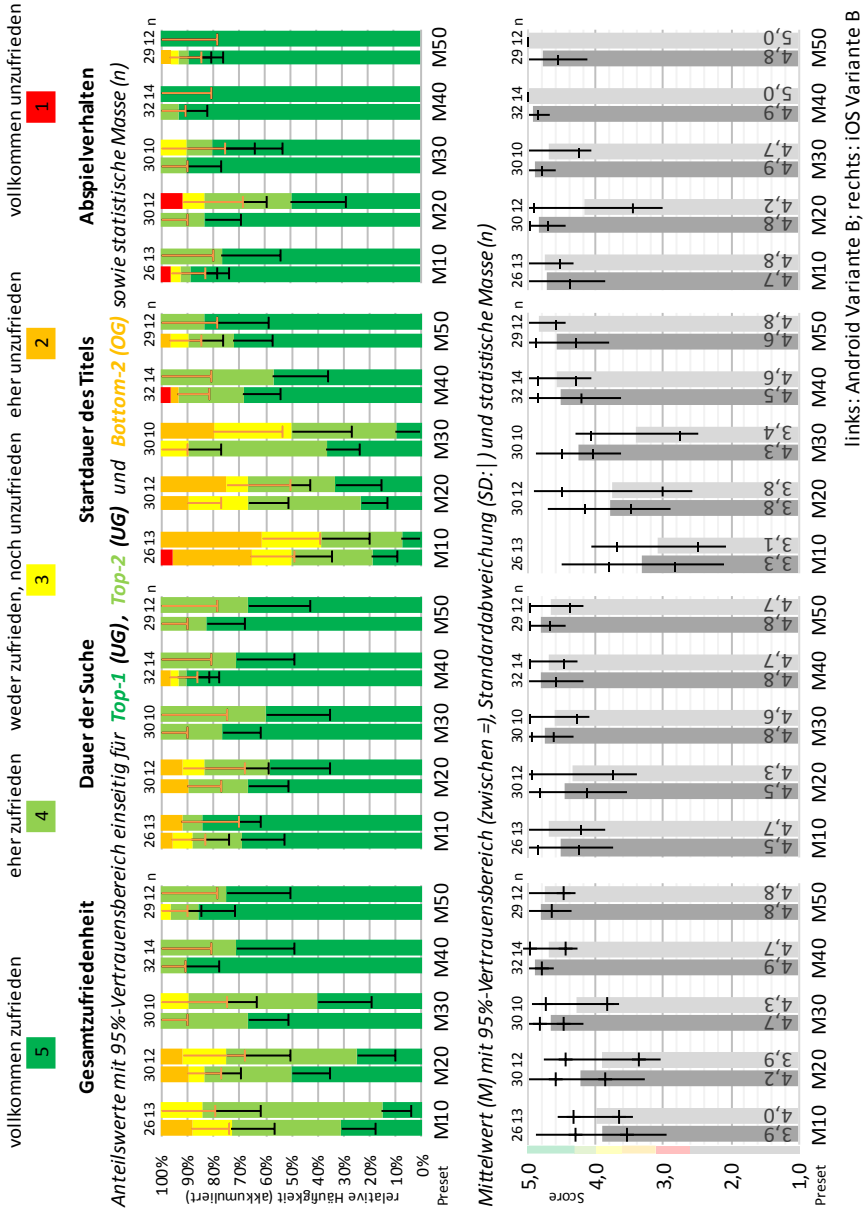


Abbildung 99: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für MTV-Music (Teil 3) Vergleich der Variante B unter Android (links) vs. iOS (rechts)

9.4.8 Bedingte Verteilungen der Zufriedenheit

Die Android-Nutzer sind insbesondere bei den langsameren Presets insgesamt etwas zufriedener als die iOS-Nutzer. Dies passt zur Voraussage, da unter Android die Startdauer des Titels kürzer ist. Die entsprechende Teilbewertung fällt allerdings weniger deutlich besser aus. Beim Vergleich der Plattformen mit Meldung „Internet langsam“ (Android) und dem Ausbleiben des automatischen Wiedergabestarts (iOS) zeigt sich deutlich, dass die Probanden mit beidem unzufrieden sind, siehe Abbildung 98 und Abbildung 99 für M10.A im Vergleich zu M10.B. Die Fehlfunktion wird unter iOS aber noch schlechter bewertet als der kurze Hinweis unter Android, vor allem weil wohl danach dann doch noch die Wiedergabe startet. M30.B ist bei iOS unerklärlich schlechter bewertet. Allerdings ist dabei die statistische Masse auch recht gering, so dass darüber keine weiteren Aussagen möglich sind.

Durch die gewählte Parametrisierung und erzeugten Stimuli wird MTV-Music, obwohl prinzipiell der Klasse Streaming zugehörig, auf die Komponente Wartezeit reduziert. Dies führt folgerichtig zu einer QoS-QoE-Charakteristik, die für die Klasse Finish-Time typisch ist, wie die zusammenfassende Darstellung der Mittelwertkurven zeigt, siehe Abbildung 113.

9.5 Spiegel.de

9.5.1 Nutzungsszenario

Das Lesen von Nachrichtenseiten im mobilen Internet ist beliebt, wie auch die Ergebnisse der eigenen Nachbefragungen zeigen. Spiegel.de ist dabei eine typische und populäre²¹³ Nachrichtenseite. Daraus leitet sich das Nutzungsszenario ab: Der Nutzer möchte sich auf Spiegel.de kurz informieren. Dazu öffnet er die mobile Webseite im Browser und surft ausgehend von der Startseite zu einem ihn interessierenden Artikel und schaut sich kurz einige Bilder der zugehörigen Fotostrecke an. Das Szenario repräsentiert die Nutzung des (mobilen) Internets im Browser mit dem Aufruf entsprechend mobil-optimierter Webseiten recht gut. Browsing im weiteren Sinn schließt dabei auch die Nutzung anderer Dienste im Browser sowie aus technischer Sicht und Nutzersicht vergleichbare Phasen in anderen Anwendungen ein.

²¹³ Top-10 Platzierung im Alexa-Ranking in Deutschland im April 2014:
<http://www.alexa.com/siteinfo/m.spiegel.de> abgerufen am 15.04.2014

9.5.2 Technische Sicht und Netznutzung

Inhalt und Übertragung: Webseiten bestehen typischerweise aus einer Vielzahl von Elementen in Form einzelner Dateien, z. B. einem HTML-Grundgerüst und eingebundenen CSS-Dateien, JavaScript-Dateien, Bildern etc.

Im Fall von Spiegel.de und vielen weiteren Websites kann zwischen verschiedenen Versionen des Inhalts unterschieden werden, entweder für die mobile Nutzung oder die Nutzung im Browser angepasst. Daneben besteht für Spiegel.de auch die Möglichkeit der Nutzung in einer App. In der Voranalyse zeigte sich ein deutlicher Unterschied im Inhalt zwischen mobiler Version und Desktop-Version hinsichtlich der Anzahl einzelner Elemente und der Gesamtdatenmenge, siehe Abbildung 100. Die Nutzung als App oder als mobile Version im Browser ist hingegen als ungefähr vergleichbar einzuschätzen. Für den Test fiel die Wahl daher auf die Nutzung der mobilen Version im Browser, was gut zum Szenario passt.

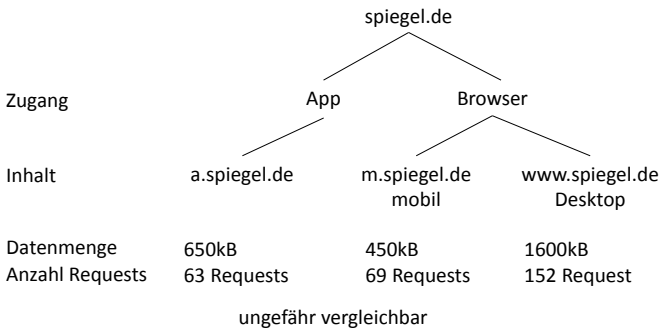


Abbildung 100: Verschiedene Arten des Zugangs und der Inhaltsversionen von Spiegel.de im Vergleich jeweils erster Aufruf der Startseite, Analyse aus eigenen Verkehrsmitteln, Beispiel für Android (iOS vergleichbar)

Erläuterungen zur Abbildung 101: Beispiel einer Analyse mit pcap2har²¹⁴, manuelle Trennung der Einzelseiten und Visualisierung in harviewer²¹⁵; die Angabe der übertragenen Datenmenge und der Events (onload) in den genutzten Auswertewerkzeugen ist nicht immer ganz korrekt

Erläuterungen zur Abbildung 102: Zahlenwerte für Server-Verarbeitung, Übertragung der Nutzdaten und Browser Rendering als Beispiele, farbliche Darstellung in Anlehnung an die Visualisierung in den Waterfall-Charts (siehe Abbildung 101)

²¹⁴ <https://github.com/andrewf/pcap2har> abgerufen am 31.12.2013

²¹⁵ <http://code.google.com/p/harviewer/> und <http://www.softwareishard.com/blog/har-viewer/> abgerufen am 31.12.2013

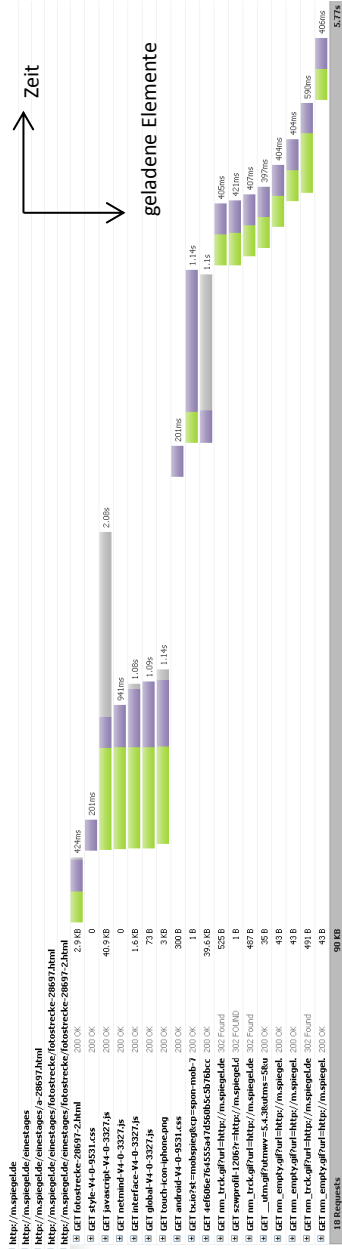


Abbildung 101: Waterfall-Diagramm zum Download der einzelnen Seiten und Elemente in Spiegel.de

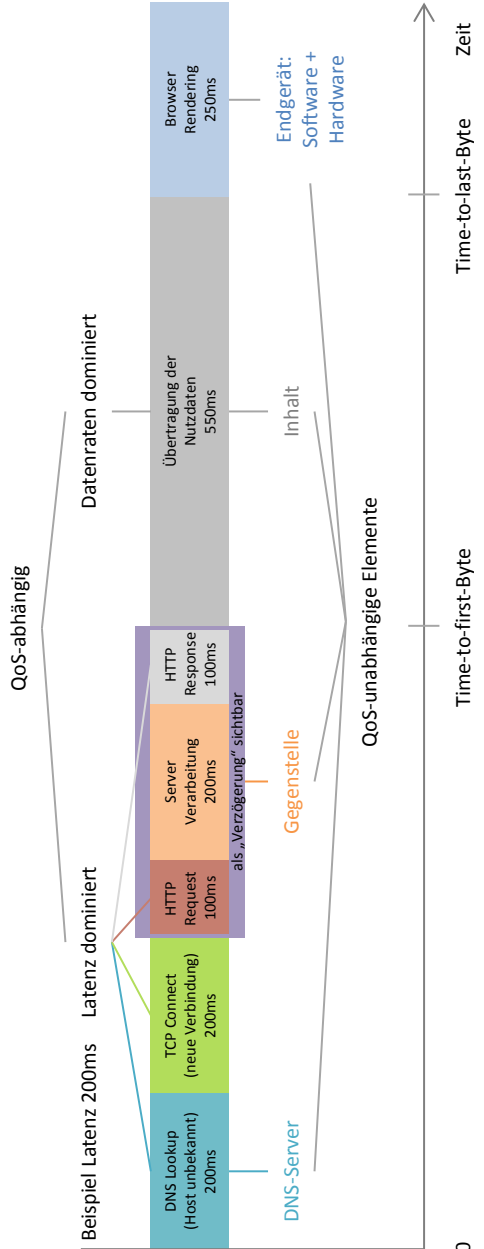


Abbildung 102: Zeitliche Bestandteile beim Browsing (Downloads und Darstellung eines Inhaltselements im Browser)

Hinsichtlich der Art und Anzahl der Elemente sowie der Gesamtdatenmenge ist Spiegel.de ungefähr repräsentativ für eine typische Desktop-Webseite und die dazugehörige mobile Version. Dabei weist die Mobilvariante eine etwas geringere Datenmenge als der Durchschnitt²¹⁶ aller mobilen Websites auf (ca. 700 kB bei ca. 60 Request). Ergänzend sei auf [346] zur Charakterisierung der zeitlichen Entwicklung dieser Größen verwiesen, wobei festzustellen ist, dass sowohl Desktop- wie auch Mobilvariante in der Vergangenheit beständig an Umfang (Elementzahl und Datenmenge) zugelegt haben.

Die einzelnen Elemente werden beim Web-Browsing typischerweise per HTTP/TCP übertragen, teils parallel und teils sequenziell nacheinander, siehe Abbildung 101. Im Unterschied zu den Bulk-Transfers der anderen Szenarien wie Drive, Facebook oder auch Youtube sind die Datenmengen der einzelnen Elemente beim Web-Browsing teilweise sehr klein. Damit ergibt sich, dass die Zeitanteile der Netznutzung, in denen effektiv keine Nutzdaten übertragen werden, gegenüber der eigentlichen Übertragungsdauer für die Nutzdaten an Bedeutung gewinnen, siehe Abbildung 102. Entsprechend steigt die Relevanz der Latenz ab gewissen Datenraten deutlich an. Durch die sequenzielle Netznutzung zum Abruf der einzelnen Elemente erhöht sich die wahrgenommene Wirkung der Latenz dabei weiter. Die Auswirkung unterschiedlicher Latenzwerte und Paketverlustraten auf das mobile Web-Browsing thematisiert [134]. Insgesamt ist für die Mobiloptimierungen also nicht nur die Reduzierung der Datenmenge sinnvoll, sondern auch die Reduzierung der Wirkung der mobil typischerweise höheren Latenz. Mit dem potenziellen Einfluss der Ladestrategie als Reihenfolge der einzelnen Elemente auf die QoE beim Web-Browsing (allerdings nicht mobil) setzt sich [347] auseinander.

Der Inhalt selbst kann beim Browsing als wichtige Einflussgröße gelten, sowohl für die QoA als auch schon für die effektiv erreichbare Download-Datenrate der QoD. Gut aufbereitete mobile Webseiten berücksichtigen die Charakteristik mobiler Netze hinsichtlich begrenzter Bandbreite und erhöhter Latenz und orientieren sich an einigen Grundregeln für den Inhalt (siehe dazu auch Googles PageSpeed-Richtlinien, u.a. zusammengestellt nach [348]):

²¹⁶ fortlaufend aktualisierte Datenerhebung durch <http://mobile.httparchive.org/> (mobile Version) bzw. <http://httparchive.org/> (für Desktop-Seiten) abgerufen am 11.10.2014

- geringere Datenmenge insgesamt für kürzere Übertragungszeiten
- weniger einzelne Elemente, was weniger Requests notwendig macht, was wiederum zu einer reduzierten Anzahl Round Trips führt
- keep-alive (HTTP) für Verbindungen um mehrere Elemente über eine geöffnete TCP-Verbindung zu laden, für die Einsparung des Overheads für den TCP-Handshake (Latenz)
- Elemente nur von wenigen Hosts laden, um DNS-Lookups zu reduzieren und bestehende Verbindungen mehrfach nutzen zu können.
- Cache-Optimierung des Inhalts hinsichtlich der Größe und Wiederverwendung einzelner Elemente zur Zwischenspeicherung, um diese bei einem späteren Aufruf nicht erneut übertragen zu müssen.

Die mobile Version von Spiegel.de kann als recht gut mobil optimiert gelten, auch wenn weiteres Potenzial besteht, wie eine Ermittlung des PageSpeed-Wertes bestätigte (65 von möglichen 100 Punkten auf dem Inhaltsstand zur Zeit der Testdurchführung). Dabei wird die Umsetzung der zuvor genannten Kriterien für eine Optimierung der Ladedauer geprüft und bewertet.

Inhaltsauswahl und Cache: Da ein externer Inhalt genutzt wurde, sind prinzipiell Änderungen in der Zusammensetzung und den Eigenschaften der Inhalte über die Zeit möglich. Der Aufbau der Seiten mit Elementzahl und Datenmenge der einzelnen Elemente zeigte sich aber im Vorbereitungs- und Testzeitraum recht konstant. Damit die Nutzer keine Suche mit umständlicher Eingabe langer Suchbegriffe durchführen mussten, wurde sie von der Startseite aus zu einem Artikel der Rubrik „eines Tages“ gelotst. Diese Artikel sind jeweils weitgehend vergleichbar hinsichtlich Umfang und Aufbau des Inhalts.

Es erfolgte die Nutzung des Dienstes per voreingestelltem System-Browser der jeweiligen Geräte. Diese speichern Teile des Inhaltes in einem Cache zwischen. Dadurch reduziert sich die Datenmenge bei einem weiteren Aufruf um bis zu 90 %, was zu deutlich kürzeren Ladedauern der Seiten führen kann. Dies hatte für den Test Konsequenzen. Der Einfluss der Aufruffreihenfolge (erste Bearbeitung vs. Wiederholung) ist in der Auswertung als wichtiger Nebenfaktor zu berücksichtigen, der durch das Testdesign induziert wird. Für gleiche Voraussetzungen zur jeweils ersten Bearbeitung wurde zwischen den Testslots der Cache der Browser geleert. Bei häufiger Nutzung der gleichen

Websites ist in der Realität davon auszugehen, dass die Nutzer von Cache-Effekten profitieren, wie sie sich in der zweiten Aufgabenausführung zeigen sollten.

Endgerät und App: Neben dem Inhalt und der Aufruffreihenfolge können auch Endgerät und App (in Form der verschiedenen Browser) als relevante Einflussfaktoren in Erscheinung treten. Je nach Komplexität des Inhalts ist die Leistungsfähigkeit des Endgerätes für die benötigte Zeit zum Aufbau und Rendering der Seite möglicherweise von Bedeutung. Die mobilen Browser können sich in verschiedenen netzrelevanten, leistungsbeeinflussenden Kriterien unterscheiden, die z. B. Browserscope²¹⁷ auflistet (nach [349]):

- Max Connections (Zahl gleichzeitiger Verbindungen)
- Connections per Hostname (Zahl gleichzeitiger Verbindungen pro Gegenseite)
- HTTP Simultaneous Request (Zahl simultaner HTTP-Requests)
- Multiple Domains Simultaneity (parallel mehrere Verbindungen zu verschiedenen Domains)
- maximale Größe einzelner Objekte, die noch im lokalen Cache gespeichert werden
- Gesamtgröße des Caches

Der Browser-Einfluss lässt sich auch durch Benchmarks und spezielle Testmethoden wie Mess-Frames oder Skripte, z. B. Loadtimer²¹⁸ und Numion Stopwatch²¹⁹ ermitteln. Eine Übersicht zu den Eigenschaften und Leistungsunterschieden der Browser findet sich in [350] und [351] (allerdings schon etwas ältere Versionen).

²¹⁷ <http://www.browserscope.org/?category=network> abgerufen am 01.09.2013

²¹⁸ <http://www.stevesouders.com/blog/2011/12/01/loadtimer-a-mobile-test-harness/> abgerufen am 15.04.2013

²¹⁹ <http://www.numion.com/Stopwatch/index.html> abgerufen am 15.04.2013

9.5.3 Nutzersicht

Die aus Nutzersicht relevanten Stimuli ergeben sich aus den Ladedauern der einzelnen Seiten: Ladedauer der Startseite, Ladedauer der Artikelseite und Ladedauern innerhalb der Fotostrecke. Dabei ist prinzipiell zu unterscheiden zwischen:

- Sichtbarwerdung erster Elemente
- alle Elemente im sichtbaren Bereich sind geladen (above the fold)
- noch nicht alle Elemente geladen, aber Seite kann genutzt werden (time to interact, TTI²²⁰)
- Seite vollständig geladen

[75]²²¹ S. 22 präsentiert Ergebnisse einer Umfrage dazu, wann die Nutzer tatsächlich beginnen, mit einer geladenen Seite zu interagieren: ca. 47 % sobald schon Text zu sehen ist, ca. 41 % warten die Mehrzahl geladener Bilder ab und nur ca. 10 % warten bis die Seite mit allem Bildern komplett geladen ist. Um unterschiedliche Interpretationen von „geladen“ im Test zu vermeiden, sollten in der Bearbeitung der Aufgaben die Seiten jeweils *vollständig* geladen sein. Sind die Seitenladezeiten aus Nutzersicht zu lang, gewinnt die Abbruchrate als relevantes Maß an Bedeutung.

Tabelle 58: Parametrisierung und Erwartungswerte für Spiegel.de

QoS-Netzparameter	Einflussprognose <i>Richtung & Stärke, Charakteristik</i>	Presets				
		S10	S20	S30	S40	S50
max. Download (kBit/s)	+++ , elastisch	128	225	384	900	1800
max. Upload (kBit/s)	o	96	192	384	700	1400
min. Latenz (ms)	- bis --, elastisch	300	300	180	90	90
Ergebnispartitionierung und relevante Einflussfaktoren		erwartete Nutzerstimuli und Charakteristik				
<i>Unterscheidung nach</i>		<i>Dauer bis Startseite vollständig geladen ist (s), elastisch</i>				
<i>Aufruf (Reihenfolge)</i>	<i>(Plattform)</i>					
erster (ohne Cache)	Android und iOS	34	25	15	8	5
zweiter (mit Cache)	Android	13	10	4,5	3,5	3,5
	iOS	10	4	3,5	3,5	3,5

²²⁰ oder auch time to first interaction

²²¹ keine Beschränkung auf eine mobile Nutzung

9.5.4 Parametrisierung und Erwartungswerte

Für die Wartedauern kann von ungefähr sinnvollen Grenzen der Stimuli entsprechend der Vornahmen von einigen Sekunden bis zu etwas längeren Wartezeiten (max. ca. 30 s) ausgegangen werden, siehe Tabelle 58. Dabei ist die Beschränkung der Betrachtungen auf die Startseite ausreichend, da die anderen Stimuli der Aufgabenbearbeitung sich in vergleichbarer Weise ändern, bzw. entsprechend der geringeren Datenmenge der ausgewählten Unterseiten jeweils immer kleiner sind. Im ersten Aufruf ist eine deutliche Spreizung der Stimuli zu erreichen, im zweiten Aufruf der Startseite fällt die Spreizung durch den Cache weit weniger stark aus. Zwischen den Plattformen und Geräten ist nur ein kleiner Unterschied feststellbar (im zweiten Aufruf), die Unterschiede zwischen Presets sind wie beabsichtigt größer.

9.5.5 Bearbeitungserfolg

Insgesamt führte das Szenario Spiegel.de bei den Probanden zu vergleichsweise wenigen Problemen in der Bearbeitung. Die Nutzung mobiler Webseiten im Browser war den Probanden vertraut, wenn auch einige Kommentare auf weiteres Potenzial in der konkreten Umsetzung der Spiegel.de-Seite schließen lassen (z. B. „Schaltflächen zu klein“). Die Wiederholungen der Bearbeitungen im Zusammenhang mit dem Grund Internetverbindung sind ausschließlich für S10 zu berichten. Dies entspricht der Erwartung hinsichtlich der Abbruchrate bei längeren Wartedauern als zusätzlich relevantes QoE-Maß neben der Zufriedenheit. Die komplette Wiederholung einer Aufgabenbearbeitung bei Spiegel ist aus technischer Sicht sehr kritisch einzuschätzen. Inhalte werden dabei aus dem lokalen Cache der Endgeräte geladen, was zu deutlich kürzeren Seitenladedauern führt (siehe Vergleich Reihenfolge 1 vs. 2). Bei festgestellten 10 Wiederholungen der ersten Aufgabenausführung ist eine Beeinflussung möglich, diese zeigen sich in den Bewertungen aber nicht direkt. Im Fall der zweiten Aufgabenausführung sind die Inhalte der Startseite etc. ohnehin teilweise schon im Cache, in diesen Fällen ist eine Wiederholung zu tolerieren, da von keiner großen Verzerrung der Bewertungen auszugehen ist.

9.5.6 Nutzerkommentare

Der Blick auf die Nutzerkommentare offenbart keine Überraschungen, siehe Abbildung 103. Die erwarteten Stimuli werden bestätigt. In den vergleichenden Kommentaren zur Wartezeit werden überdeutlich die Zufriedenheit und weitere Verbesserungen („schnell“ und „schneller“) betont, die möglicherweise auf die Wirkung des Caches zurückzuführen sind. In den anderen Szenarien, z. B. bei Drive, stehen sich Verbesserungen mit „schneller“ und Verschlechterungen mit „langsamer“ hingegen ungefähr gleich gegenüber.



Abbildung 103: Word-Cloud aus den Nutzerkommentaren zu Spiegel.de

9.5.7 Bewertungsverhalten

Bedingt durch den technischen Einfluss des Caches bei der zweiten Aufgabenausführung ist bei Spiegel.de eine Überprüfung der Antwortkonformität im Wiederholungsvergleich nur bedingt sinnvoll, da selbst bei schlechteren QoS-Parametrisierungen bessere QoE-Bewertungen plausibel sind. Entsprechend ist eine Verschiebung in der Differenzmatrix nach oben zu erkennen. Lediglich Einträge, die zu schlechteren QoE-Bewertungen bei besseren QoS-Parametern führen sind tatsächlich unerwartet. Diese sind allerdings nur in sehr geringer Zahl zu erkennen, siehe Abbildung 104a. Es ist ein sehr hohes Maß an Übereinstimmung der Teilbewertungen mit der Gesamtzufriedenheit festzustellen, siehe Abbildung 104b.

9.5.8 Bedingte Verteilungen der Zufriedenheit

Besonders bei langsamen Netzsituationen profitieren die Bewertungen zur Ladedauer der Startseite von der Wirkung des Caches, siehe Abbildung 105. Für die Gesamtzufriedenheit und die Teilbewertung zur Startseite entspricht der Gewinn durch den Cache ca. einer Netzparameterstufe. Ab S30 gleichen sich die Bewertungen von erstem und zweitem Aufruf weitgehend an. Da unterschiedliche Artikel und Fotostrecken in beiden Aufrufen genutzt werden sollten, fallen für diese Teilbewertungen die Unterschiede zwischen den Aufrufen, im Unterschied zur weitgehend unveränderten Startseite, erwartungsgemäß geringer aus. Zwischen den mobilen Plattformen ist kein relevanter Unterschied festzustellen, allerdings ist die Betrachtung der Plattform nachrangig, so dass besonders bei iOS die statistische Masse sehr klein ist. Auf eine Darstellung wird entsprechend verzichtet. Der Bereich der ausgewählten Netzparameter passt weitgehend. Zum Austesten der Schmerzgrenze der Probanden wäre eventuell sogar noch eine Stufe unter S10 möglich gewesen, um überwiegend unzufriedene Nutzer zu provozieren.

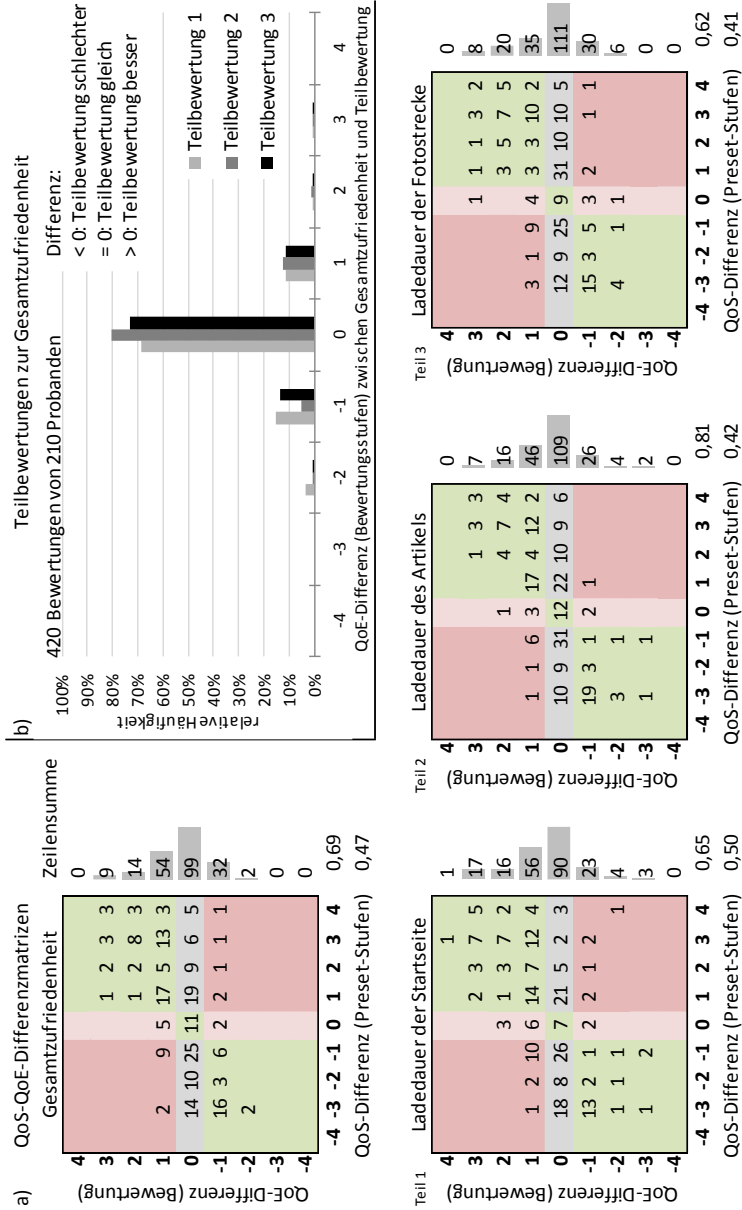


Abbildung 104: Zusammenstellung zum Bewertungsverhalten in Spiegel.de

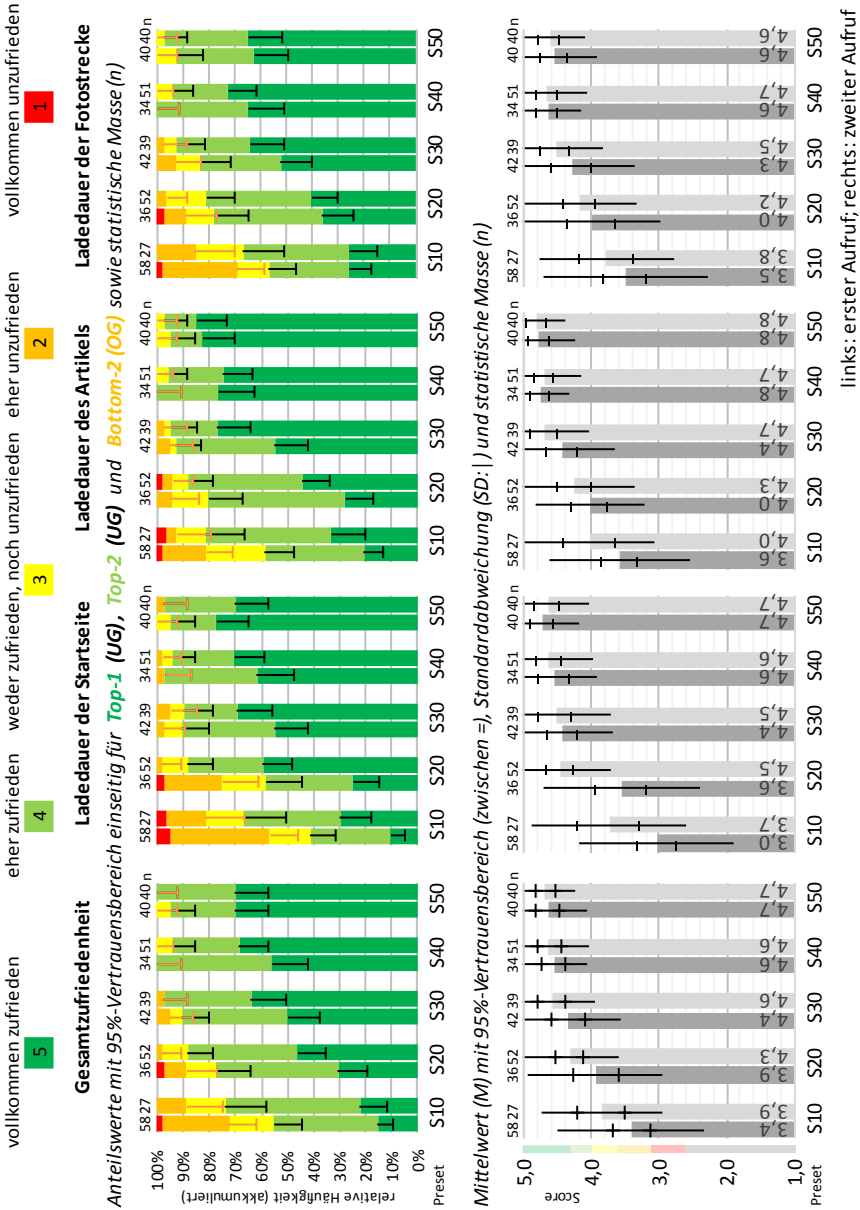


Abbildung 105: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Spiegel.de Vergleich zwischen erstem Aufruf (links) und zweitem Aufruf (rechts)

9.6 Youtube

“The buffering circle is my nemesis.”

ein Youtube-Ingenieur²²² zu seiner Arbeit

9.6.1 Nutzungsszenario

Die Übertragung von IP-basiertem Video verursacht im Internet insgesamt, wie auch im mobilen Internet, einen sehr großen bzw. sogar den größten Einzelanteil am gesamten Verkehrsvolumen, vor allem im Downstream [15] S. 4. Gleichzeitig zeigen Prognosen eine weitere exponentielle Zunahme von Mobile Video im Verkehrsmix von jährlich ca. 69 %, siehe [352] S. 14. Diese Zunahme ist nicht nur dadurch zu erklären, dass immer mehr Nutzer immer mehr Videos anschauen – was sich auf mobilen Geräten mit immer größeren Bildschirmen auch anbietet – sondern auch, dass sich die genutzten Qualitätsstufen immer weiter erhöhen, siehe [16] S. 8f. Dies ist ein schönes Beispiel, wie verfügbare Bandbreite durch die Nutzer innerhalb kürzester Zeit wieder verkonsumiert wird. Wenn die höhere Qualitätsstufe verfügbar ist und prinzipiell auch funktioniert, dann wird sie über kurz oder lang auch genutzt und nachfolgend gefordert (siehe Erläuterungen zum Innovationszyklus in der Einleitung).

Aus den Erhebungen zeigt sich auch, dass Youtube als konkreter Dienst die mobile Videonutzung anteilmäßig deutlich dominiert, mit über 20 % des gesamten Downstream-Volumens ([15] S. 4, Jahr 2013) und über 66 % des Datenaufkommens von mobilen Video ([280] S. 4, Jahr 2012). Auch für die Anbieter der Videoinhalte nimmt die Relevanz der mobilen Nutzung zu. Anfang 2014 wurden bereits fast 40 % der Videoabrufe bei Youtube durch mobile Geräte getätigt [353].

Ein mögliches Szenario könnte damit wie folgt aussehen: Der Nutzer will am Abend ins Kino. Er möchte sich vorab zu den aktuellen Filmen informieren und schaut sich dazu die Trailer (in bestmöglicher Qualität, HD) auf Youtube an.

²²² <http://www.youtube.com/watch?v=OqQk7kLuaK4> abgerufen am 30.11.2013

9.6.2 Technische Sicht und Netznutzung

Die Vielfalt der technischen Möglichkeiten: Aus technischer Sicht ist zunächst festzustellen, dass es nicht das eine mobile Youtube gibt, sondern viele Ausprägungen, je nach Plattform, App, Netz und Inhalt siehe Abbildung 106. Diese Heterogenität wirkt sich in den technischen Anforderungen und auch in der Nutzersicht als Unterschiede in den Startzeiten, beim Abspielverhalten und der wahrgenommenen Qualität des Inhalts aus.

In der Testrunde 1 konzentrierten sich die Untersuchungen auf die Nutzung im Browser in der Qualitätsstufe SQ. In der Testrunde 2 wurde die Nutzung in der Youtube-App favorisiert, mit dem Ziel der bestmöglichen Qualität. Die Qualitätsstufe HD war dabei zum Zeitpunkt der Tests (noch) nicht typisch auf mobilen Endgeräten (Smartphone/Tablets), sondern noch weitgehend auf tragbare Geräte wie Laptops etc. begrenzt. Erste Veränderungen Richtung HD auch auf mobilen Geräten waren aber zu beobachten. Mit Blick auf die Facette Echtheit wurde für die Testrunde 2 daher akzeptiert, dass die Stufe HD mobil noch nicht durchgängig nutzbar war, aber zukünftig voraussichtlich eine größere Rolle spielen könnte.

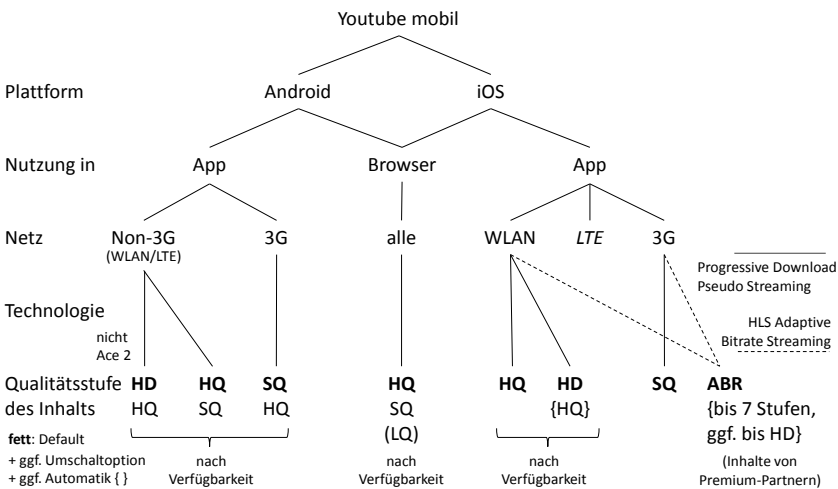


Abbildung 106: Aufschlüsselung zur Vielfalt der technischen Möglichkeiten bei der mobilen Nutzung von Youtube

eigene Tests (außer Kombination Android + LTE, diese durch Partner), für iOS + LTE keine Aussage möglich, Stand 2013/Q2

Erkennt die Youtube-App auf den Mobilgeräten eine 3G-Mobilfunkverbindung, wird dies über entsprechende Parameter bei der Kommunikation mit den Youtube-Servern signalisiert und die Qualitätsstufe HD trotz prinzipieller Verfügbarkeit von der Nutzung ausgeschlossen. Weitere Versuche im Mobilfunk (nicht selbst ausgeführt) zu den nutzbaren Qualitätsstufen ergaben, dass zumindest unter Android bei LTE-Nutzung die Stufe HD zur Verfügung steht. Für iOS in Verbindung mit LTE konnten mangels passender Geräte-Netz-Kombination keine Erkenntnisse gewonnen werden. Per WLAN war HD hingegen prinzipiell nutzbar.

Die Auswahl des Inhalts: Für die Eignung der jeweiligen Videoclips im Test gab es einige Randbedingungen, sowohl für die technischen als auch die nicht-technischen Komponenten. Die Videos sollten:

- von Youtube-Kanälen ohne vorgeschalteten Videoclip (Preroll-Werbung) stammen: Dies erlaubt eine einfachere Vorhersage der Stimuli, da sich die Werbe-Clips in ihren Eigenschaften unterscheiden können und das Player-Verhalten sich womöglich ändert.
- einigermaßen hohe Abrufzahlen aufweisen: So ist von hoher Popularität auszugehen und die Inhalte sollten bereits im Youtube-CDN weit verteilt sein.
- im Bereich einer vorgegebene Länge sein (ca. 1min): Bei kürzeren Clips lassen sich nicht genug Unterbrechungen erzeugen, bei längeren Clips wird hingegen das Zeitbudget der Aufgabe überschritten.
- technisch qualitativ hochwertig sein (Bildqualität).
- möglichst repräsentativ innerhalb des erwarteten Bereichs der Inhaltsbitrate sein (mittlere Datenrate), aber auch technisch anspruchsvoll sein (Bitraten-Peaks, z. B. bei sehr komplexen Szenen mit viel Bewegung und rascher Szenenfolge etc.).
- nicht mit 5 s schwarzer Stille beginnen, weil so der tatsächliche Wieder-gabestartzeitpunkt nur schwer erkennbar ist.

Die Probanden sollten also nicht x-beliebige Videos schauen (keine spezifischen Vorhersagen möglich), sondern explizit vorausgewählte und als geeignet eingeschätzte Videos. Die Auswahl der Videoinhalte kann ein heikler Punkt sein, der die Probanden unnötig polarisiert. Die nichttechnische Kom-

ponente ist dabei nachdrücklich zu berücksichtigen, daher sollten die gewählten Clips einen relativ neutralen Inhalt aufweisen. In der Sichtung der Nutzerkommentare der Testrunde 1 zeigte sich teils deutlich, dass sich über Geschmack oder Humor eben doch streiten lässt, besonders bei „lustigen“ Tiervideos²²³. In der Testrunde 2 wurden populäre deutsche Kinotrailer ausgewählt, die ausgehend von ihrem Inhalt den Massengeschmack treffen sollten und von den Probanden im Rahmen einer Aufgabenausführung vollständig anzuschauen waren: *Star Trek Into Darkness*²²⁴ (Kinostart im Sommer 2013) und *Der Medicus* (Kinostart im Winter 2013). Prinzipiell wäre es möglich gewesen, die Probanden innerhalb der Youtube-App eine Suche nach den speziell ausgewählten Trailern durchführen zu lassen. Dabei war aber nicht sicherzustellen, dass die Probanden dann auch tatsächlich exakt das geplante Video schauten und nicht ein anderes titelgleiches Video. Dies stellt besonders bei Filmtrailern ein Problem dar, da diese typischerweise in vielen Varianten und von vielen Youtube-Kanälen verfügbar sind. Diese Varianten können deutlich andere technische Eigenschaften aufweisen, was im Test unbedingt vermieden werden sollte. Daher wurden die ausgewählten Videos als eine Art Lesezeichen in der „Später ansehen“-Liste der eingerichteten Youtube-Accounts abgelegt, von der aus die Probanden sie schnell und gezielt auswählen konnten. Auf eine Suche und Bewertung zur Browsing-Phase wurde verzichtet.

Der Videoinhalt: Bei den verfügbaren Inhaltsversionen, siehe Tabelle 59, gibt es Überschneidungen zur Desktop-Version von Youtube. Im Bereich der Desktop-Nutzung sind weitere Inhaltsversionen und Streaming-Techniken möglich. Dies gilt auch für die Nutzung auf den mobilen Geräten in geeigneten Browsern mit Flash-Plugin, was allerdings nicht weiter berücksichtigt wurde. Ausgehend vom ermittelten technischen Stand sind weitere Inhaltsversionen und Änderungen über die Zeit möglich und wahrscheinlich.

²²³ „No Animals Were Harmed“ etc. sollte beachtet werden, siehe auch <http://www.humanehollywood.org/> abgerufen am 03.01.2014

²²⁴ Vorgriff auf die Auswertung: Es war nur ein einziger sehr negativer Kommentar zum Inhalt festzustellen („schrottfilm, unzumutbar anzusehen“).

Codecs und Formate sowie die Datenraten der einzelnen Komponenten (Audio und Video) sind für die Gesamtbetrachtung nur zweitrangig. Die Auflösung als qualitatives Differenzierungsmaß der einzelnen Qualitätsstufen ist an dieser Stelle ausreichend. Für technische Anforderungen bei der Übertragung ist die Gesamtdatenrate entscheidend. Limitierungen auf den Endgeräten durch Codecs, Auflösungen oder CPU-Leistung sind prinzipiell möglich. Abgesehen vom Ace 2 (ohne HD) konnte aber beim Rest der Gerätezusammenstellung des Testbeds keine Einschränkung durch die Leistungsfähigkeit der Hardware oder Software festgestellt werden. Sofern die Übertragungsdatenraten ausreichend waren, erfolgte auf den Geräten die Wiedergabe ohne systematische technisch bedingte Beeinträchtigungen. Für die Testgeräte bedeutet die angestrebte bestmögliche Videoqualität innerhalb des WLAN-Testbeds die Stufe HD 720p (itag 22) mit Ausnahme des Galaxy Ace 2, das nicht HD-fähig ist und damit auf die Stufe HQ limitiert war (itag 18), siehe Tabelle 59.

Tabelle 59: Qualitätsstufen und technische Kenndaten des Inhalts für Youtube (mobil)

Qualität (Stufe)	itag ²²⁵	Streaming-Technologie	Auflösung	Bitrate gesamt (kBit/s)	Anmerkung
LQ	17	Progressive Download	176x144	90	
SQ	36	Progressive Download	320x240	200-240	
HQ	18	Progressive Download	480x360	500	ca. Auflösung, je nach Seitenverhältnis, ggf. VBR
HD	22	Progressive Download	1280x720	1200-2400	HD 720p, VBR (Peaks höher)
HD	37	Progressive Download	1920x1080	3200-4500	Full-HD, nicht im Test beobachtet (nur Verweis)
ABR	151	HLS	128x72	60	ABR: Adaptive Bit Rate ABR-Stufe wird automatisch gewählt HLS: HTTP Live Streaming
ABR	132	HLS	426x240	280	
ABR	92	HLS	426x240	390	
ABR	93	HLS	640x360	860	
ABR	94	HLS	854x480	1400	
ABR	95	HLS	1280x720	2800	
ABR	96	HLS	1920x1080	4900	

Werte aus eigenen Versuchen mit Android und iOS, durch Auswertung der Verkehrsdaten, der Videodateien und der zugehörigen Metadaten, z. B. Manifest-Dateien; VBR: Variable Bit Rate; Stand 2013/Q2, nicht erschöpfend

²²⁵ „itag“ ist ein von Youtube intern genutzter Parameter, der die technisch unterschiedlichen Versionen ein und desselben Videos bei Youtube kennzeichnet.

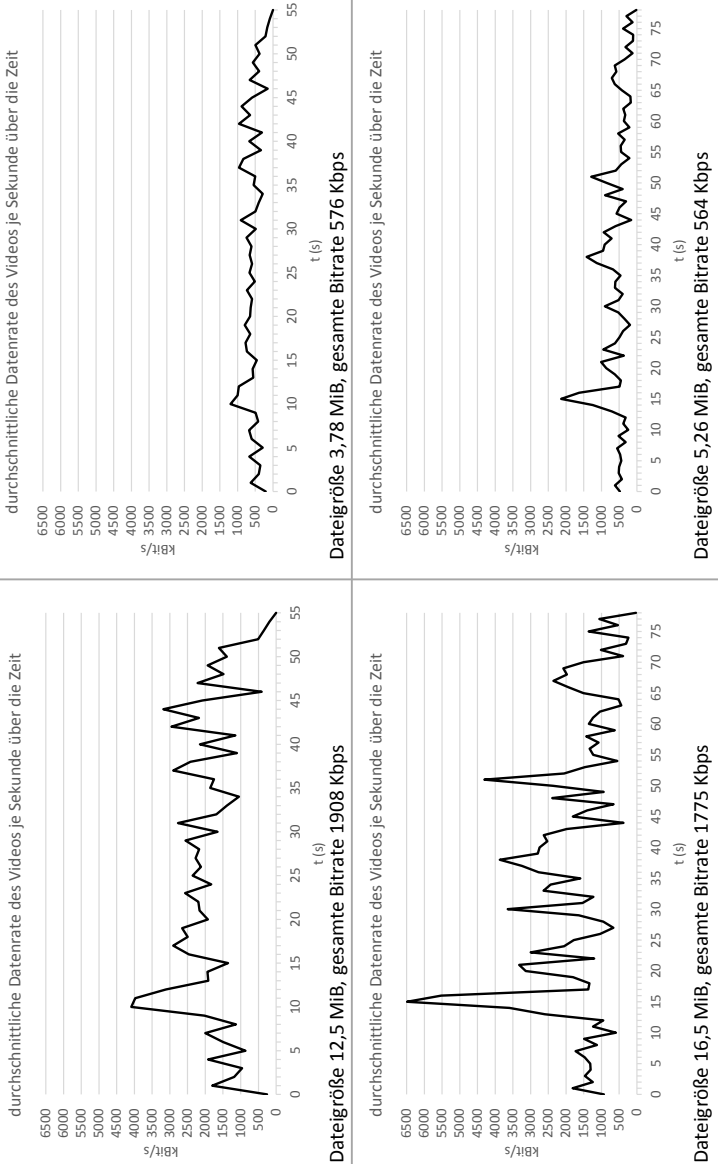


Abbildung 107: Zeitlicher Verlauf und Durchschnittswerte der Inhaltsdatenraten der ausgewählten Inhalte in Youtube oben: Variante A (Der Medicus), unten: Variante B (Star Trek) links: Stufe HD (itag 22), rechts: Stufe HQ (itag 18)

Ermittlung der durchschnittlichen Datenrate pro Sekunde mit ffmpeg-Statistiken zur Byte-Position der einzelnen Frames und weitere Verarbeitung zur Auswertung; Ermittlung der zusammenfassenden Angaben mit MediaInfo ²²⁶

Für die ausgewählten Inhalte wurde neben der durchschnittlichen Datenrate des Inhalts auch der relevante zeitliche Verlauf je Qualitätsstufe vorab ermittelt, da die Inhalte mit variabler Bitrate kodiert vorliegen, siehe Abbildung 107. Dabei sind jeweils einige Spitzen zu erkennen, die deutlich oberhalb der durchschnittlichen Datenrate liegen.

Streaming: Für die Untersuchungen wurden Inhalte ausgewählt, die per Pseudo-Streaming als Progressive Download per HTTP/TCP übertragen werden. Dabei kann zusätzlich noch der genutzte Streaming-Algorithmus in Burst und Throttling unterschieden werden. Bei Burst erfolgt die Datenübertragung ungebremst mit maximaler Geschwindigkeit (typischerweise nur durch das Netz beschränkt, ggf. auch durch maximale Puffergrößen im Client). Beim Throttling ist eine von Server oder Client künstlich gedrosselte, kontinuierliche Übertragung festzustellen, angepasst an die Bitrate des Medieninhalts, so dass nach einem anfänglichen Burst der Pufferfüllstand im Client möglichst konstant gehalten wird. Durch Throttling kann das übertragene Datenvolumen reduziert werden, da die Übertragungszeit ungefähr der übertragenen Videozeit des Inhalts entspricht. Spielt ein Nutzer ein Video nur kurz an, z. B. für 30 s, um es dann abzubrechen, so ist es nicht immer sinnvoll in dieser Zeit 15 min Videoinhalt zu übertragen, auch wenn das Netz dies prinzipiell könnte, weil 14:30 min davon nutzlos wären, wenn sie der Nutzer dann doch nicht anschaut.

In den eigenen Analysen mit den getesteten Datenraten und Videolängen war kein Throttling zu beobachten, die Übertragung erfolgte stets schnellstmöglich als Burst. In anderen Untersuchungen konnte Throttling auch in der mobilen Nutzung von Youtube beobachtet werden, siehe [354]. Dort wurden allerdings andere Nebenfaktoren wirksam (längere Videos und niedrigere Qualitätsstufen, andere Endgeräte sowie 3G-Netze). Dies unterstreicht deutlich, dass die Aussagen zu einem bestimmten Dienst nur bedingt zu generalisieren sind, wenn eine Vielzahl von Nebenfaktoren einwirkt (bzw. dann ein anderer Dienst anzunehmen ist).

Grundsätzlich ist davon auszugehen, dass Adaptive Bitrate Streaming in der praktischen Nutzung zukünftig stark an Bedeutung gewinnt. Dabei wird nicht eine Qualitätsstufe fest ausgewählt, sondern der Videoinhalt zerlegt in zeitlich

kurze Segmente übertragen. Die einzelnen Segmente können in verschiedenen Qualitätsstufen vorliegen und angepasst, z. B. an die Netzbedingungen, automatisch ausgewählt und übertragen werden. Für den eigenen Test blieben diese Möglichkeiten noch unberücksichtigt, da noch nicht durchgängig nutzbar, siehe Abbildung 106.

Abspielverhalten: Bei (echtem Live-) Streaming kann ein „Freeze-and-Jump“²²⁷ auftreten, wenn die Daten nicht schnell genug verarbeitet/übertragen werden können. Dabei bleibt die Wiedergabe stehen, sobald nicht mehr genug Daten verfügbar sind, setzt aber bei einer (automatischen) Fortsetzung der Wiedergabe nicht nahtlos an der Unterbrechungsstelle an, sondern überspringt die entsprechend vergangene Realzeit, um wieder „live“ zu sein. Youtube in der untersuchten mobilen Variante verhält sich hingegen als „Freeze-and-Continue“ (als analoger Begriff). Dabei stoppt die Wiedergabe, das Bild ist eingefroren und nach einiger Zeit (typischerweise einige Sekunden) setzt die Wiedergabe nahtlos fort. Unter iOS ist zudem ein permanentes Fallback von HD auf die niedrigere Qualitätsstufe HQ zu beobachten, wenn die App erkennt, dass die Inhaltsdatenrate die Übertragungsdatenrate so übersteigt, dass weitere Unterbrechungen wahrscheinlich bzw. unausweichlich sind.

9.6.3 Nutzersicht

Als relevante Nutzerstimuli sind die Dauer bis zum Wiedergabestart des Videos (Vorpufferzeit) und das Abspielverhalten des Videos zu identifizieren. Das Abspielverhalten ist dabei gekennzeichnet durch:

- Anzahl der Unterbrechungen und ggf. die zeitliche Verteilung der Unterbrechungen mit Wiedergabezeiten und zum Nachpuffern benötigte Pausenzeiten
- ggf. automatische Qualitätsstufenwechsel (von HD zu HQ bei iOS)
- die visuelle Qualität des Videoinhalts, inklusive etwaiger Indikatoren zur Qualität in der App (Angaben zur maximal möglichen Qualität im Titel der Videos sind üblich, z. B. „Der Medicus [HD]“, allerdings ohne technischen Bezug zur tatsächlichen Qualität, siehe z. B. Abbildung 131.2. Angaben dieser Art sind hier nicht gemeint.)

²²⁷ Begriff aus [355] übernommen.

Das Laden der Startseite, der Videoliste und der einzelnen Seiten zu den Videos entspricht weitgehend dem Szenario Browsing. Diese Aspekte waren nicht vordergründig interessant, da isoliert besser zu untersuchen. Auf die gesonderte Bewertung dieser Nutzerstimuli wurde entsprechend verzichtet. In der Nutzersicht auf das Abspielverhalten müssten eigentlich auch der zeitliche Abstand zwischen feststellbaren Beeinträchtigungen und deren Dauern (Nachpuffern) berücksichtigt werden. Als einfacher Ansatz genügt aber auch eine Prognose/Erfassung ob und wie oft es zum Abreißen der Wiedergabe (Stalls) kommt. Greift unter iOS die Umschaltautomatik ein, so wird die Wiedergabe für etwas längere Zeit unterbrochen. Die Wiedergabe wird in diesen Fällen meist nicht ganz perfekt an der pausierten Stelle im Video fortgesetzt (in Abhängigkeit der I-Frame-Position). Der Wechsel der Qualitätsstufe sollte anhand der visuellen Qualität durch einen aufmerksamen Betrachter zu erkennen sein. Auf die visuelle Qualität wird zusammen mit den Nutzeräußerungen am Ende des Abschnitts gesondert eingegangen.

9.6.4 Parametrisierung und Erwartungswerte

Grundsätzlich ist von Beeinträchtigungen des Abspielverhaltens auszugehen, wenn die parametrisierte maximale Download-Datenrate unterhalb der Inhaltsdatenrate liegt, siehe Tabelle 60. Die Inhaltsdatenrate gibt damit feste technische Anforderungen vor, woraus sich auch nur ein vergleichsweise enger sinnvoller Parameterbereich ergibt. Ausgehend von der Inhaltsdatenrate bei 2000 bis 2400 kBit/s für HD wurden jeweils zwei höhere bzw. niedrigere Stufen für die maximale Download-Datenrate ausgewählt. Im Gegensatz zu MTV-Music sind selbst mit deutlich unzureichenden Datenraten keine deutlich längeren Wiedergabestartzeiten zu provozieren. Für HD ist mit dem Verschwinden der Einschränkungen von Y20 zu Y30 im Abspielverhalten mit einer sprunghaften QoE-Verbesserung zu rechnen. Mit der gewählten Parametrisierung sind bei der Qualitätsstufe HQ keine relevanten QoE-Veränderungen zu erwarten (nur hohe Werte). Trotzdem wurde das Szenario auch auf dem nur HQ-tauglichen Ace 2 getestet. So konnten für alle Geräte einheitliche Testpläne und Aufgabenzusammenstellungen genutzt werden.

Tabelle 60: Parametrisierung und Erwartungswerte für Youtube

QoS-Netzparameter		Einflussprognose	Presets				
		<i>Richtung & Stärke, Charakteristik</i>	Y10	Y20	Y30	Y40	Y50
max. Download (kBit/s)		+++ , sprunghaft von Y20 zu Y30	1200	1800	2400	3600	5000
max. Upload (kBit/s)		o	1024	1400	1400	1400	1400
min. Latenz (ms)		o	90	90	90	90	50
Ergebnispartitionierung und relevante Einflussfaktoren			erwartete Nutzerstimuli und Charakteristik				
<i>Unterscheidung nach</i>			1. Zeile: Startzeit (s), elastisch 2. Zeile: Abspielverhalten mit Anzahl Unterbrechungen (Stalls), sprunghaft 3. Zeile: Qualitätsumschaltung, sprunghaft				
<i>Inhalts- version</i>	<i>Plattform/App</i>	<i>Variante (Videoclip)</i>					
HQ	Android Youtube App mit Qualitätsindikator ohne Umschaltautomatik	A und B vergleichbar					
HD	Android Youtube App mit Umschaltautomatik	A	4,5 13 4 Stalls	3,8 9 1 Stall	2,7 6	2,4 5	2,2 5
		B	12 4 Stalls	8s 2 Stalls	6	4	4
	iOS Youtube App ohne Qualitätsindikator mit Umschaltautomatik	A	11 1 Stall HD→HQ	9 1 Stall (*)	7	5	4
		B	11 1 Stall HD→HQ	9 2 Stalls (*)	6	5	4

Variante A: Der Medicus, B: Star Trek Into Darkness,

HD→HQ = Umschaltautomatik von HD auf HQ beim Nachpuffern,

(*) knapp an der Grenze, so dass ein Eingreifen der Umschaltautomatik bei weiteren externen Einflüssen prinzipiell noch möglich ist

9.6.5 Bearbeitungserfolg

Insgesamt verursachte Youtube den Probanden keine großen Probleme bei der Bearbeitung der Aufgabe. Es ist anzunehmen, dass die Probanden mit der Usability und Funktion der App vertraut waren. Lediglich die genutzte Funktion der Liste „Später ansehen“ im Youtube-Konto bzw. der Weg dahin waren wohl etwas ungewohnt, was sich auch in den Nutzerkommentaren widerspiegelt. Die Nutzer nehmen bei langsamer Netzparametrisierung die Wartezeiten und Funktionseffekte wahr, aber es ist kein Grund für einen Abbruch oder eine Wiederholung. Längere Wartezeiten und häufige Unterbrechungen sind den Nutzern in einem Szenario zur mobilen Videonutzung offenbar nicht fremd. Entsprechend sind lediglich zwei Wiederholungen der Bearbeitungen im Zusammenhang mit dem Grund Internetverbindung für Y10 festzustellen. Eine

Angabe für Y50 im Zusammenhang mit dem Problemgrund Internetverbindung kann nicht erklärt werden (wohl ein externes Problem). Die Wiederholung einer Aufgabenbearbeitung bei Youtube ist aus technischer Sicht unkritisch einzuschätzen, sofern ein Proband nicht zuvor das komplette Video heruntergeladen hat (höchst unwahrscheinlich).



Abbildung 108: Word-Clouds aus den Nutzerkommentaren zu Youtube

oben links: allgemeine Kommentare; unten rechts: Kommentare zum Abspielverhalten

9.6.6 Nutzerkommentare

Die Propositionen zu Nutzerstimuli und Einflussstärke bestätigen sich auf Basis der Nutzerkommentare, siehe Abbildung 108.

Probanden, die mit dem Abspielverhalten nicht vollkommen zufrieden waren, störten sich, wie erwartet, stark an Unterbrechungen und (mehrfachem) Nachladen des Videos. Dies wurde dann als verschiedene Formen von Stocken, Unterbrechungen, Nach- bzw. Zwischenladen sowie als „Ruckler“ beschrieben (auch wenn letzteres eigentlich nicht ganz korrekt ist).

9.6.7 Bewertungsverhalten

Im Vergleich zu den Mustern in den QoS-QoE-Differenzmatrizen, auch den der anderen Szenarien, fällt die höhere Zahl sprunghafter QoE-Verbesserungen beim Abspielverhalten auf, die bei QoE-Differenz von -3 und +3 zu jeweils deutlich größeren Zeilensummen führen, siehe Abbildung 109a. Dies passt zur Annahme, dass die Nutzer in Abhängigkeit deutlicher Änderungen im Abspielverhalten (mit bzw. ohne Beeinträchtigungen) ihre Meinung entsprechend auch deutlich ändern. Für die Bewertung der Startdauer sind überwiegend keine Veränderungen zu erzielen. Es ist ein sehr hohes Maß an Übereinstimmung der Teilbewertungen mit der Gesamtzufriedenheit festzustellen, siehe Abbildung 109b.

Für Youtube konnten die Stimuli sehr genau prognostiziert werden. Daher sind alle unerwarteten Abweichungen in den Bewertungen, insbesondere zum wichtigen Abspielverhalten, gesondert geprüft worden, ob diese mit der QoS-Parametrisierung und den beabsichtigten Erwartungswerten in Zusammenhang stehen.

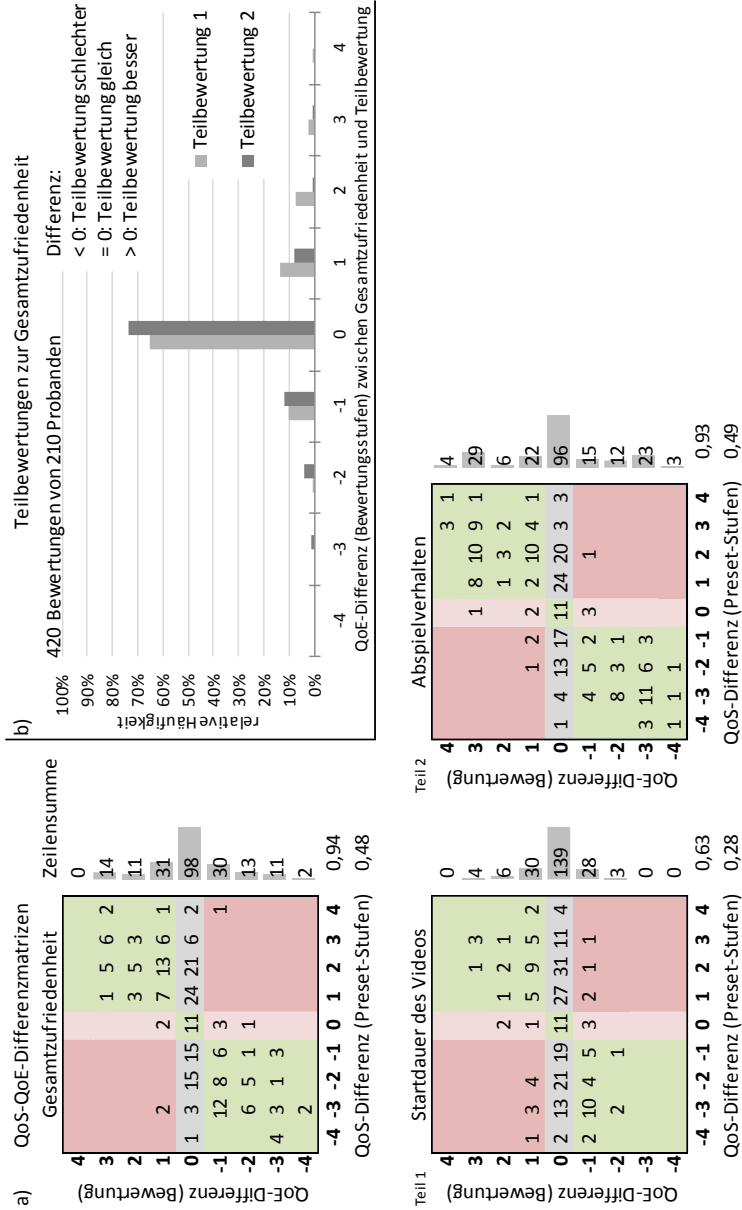


Abbildung 109: Zusammenstellung zum Bewertungsverhalten in Youtube

Für HD bei Y30 sind 7 von 8 unerwarteten Bewertungen durch Unterbrechungen begründet. Dies war ursprünglich nicht geplant, die maximale Download-Datenrate bei Y30 ist aber recht knapp an der tatsächlichen Inhaltsdatenrate der ausgewählten Videos. Schon geringe kurze äußere Einflüsse (Youtube-Server, externes Netz, ...) konnten damit den relevanten Nutzerstimulus der Unterbrechungsfreiheit und damit die Zufriedenheit deutlich ändern (realistisches Verhalten). Für HQ sind 2 von 3 unerwarteten Bewertungen zum Abspielverhalten mit Unterbrechungen kommentiert worden (unerwarteter und ungeplanter externer Einfluss). Die Ursachen dazu waren nicht aufzuklären, sind aber in der Auswertung ohne weitere Bedeutung. Die Gesamtzufriedenheit kann bei den überprüften restlichen Abweichungen, die zu einer invaliden Teilbewertung führten, trotzdem noch als valid gewertet werden, da lediglich von einer Fehlinterpretation der Fragestellung zur Teilbewertung auszugehen ist.

9.6.8 Bedingte Verteilungen der Zufriedenheit

Für die Qualitätsstufe HQ sind, wie erwartet, keine QoE-Änderungen festzustellen, siehe Abbildung 110. Entsprechend bezieht sich die weitere Diskussion ausschließlich auf die Qualitätsstufe HD. Bei Youtube HD ist die für Streaming erwartete sprunghafte QoE-Änderung mit engem Schwellwertbereich sowohl für die Bewertung zum Abspielverhalten als auch in der Gesamtzufriedenheit deutlich zu erkennen:

- Y10 und Y20: Unterhalb der Inhaltsdatenrate wird die Gesamtzufriedenheit durch das erzeugte Abspielverhalten mit Unterbrechungen negativ beeinflusst.
- ab Y30: Oberhalb der Inhaltsdatenrate sind nur geringe Veränderungen der Zufriedenheit festzustellen.

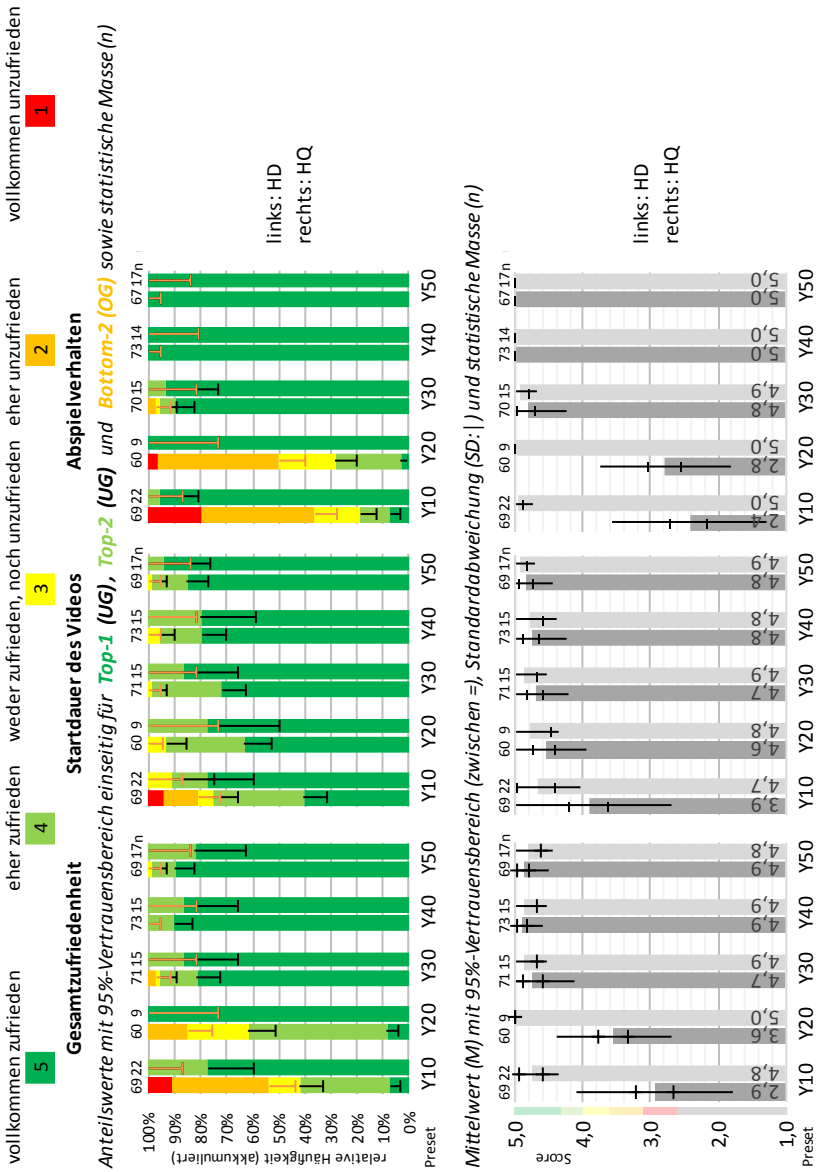


Abbildung 110: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Youtube (Teil 1)

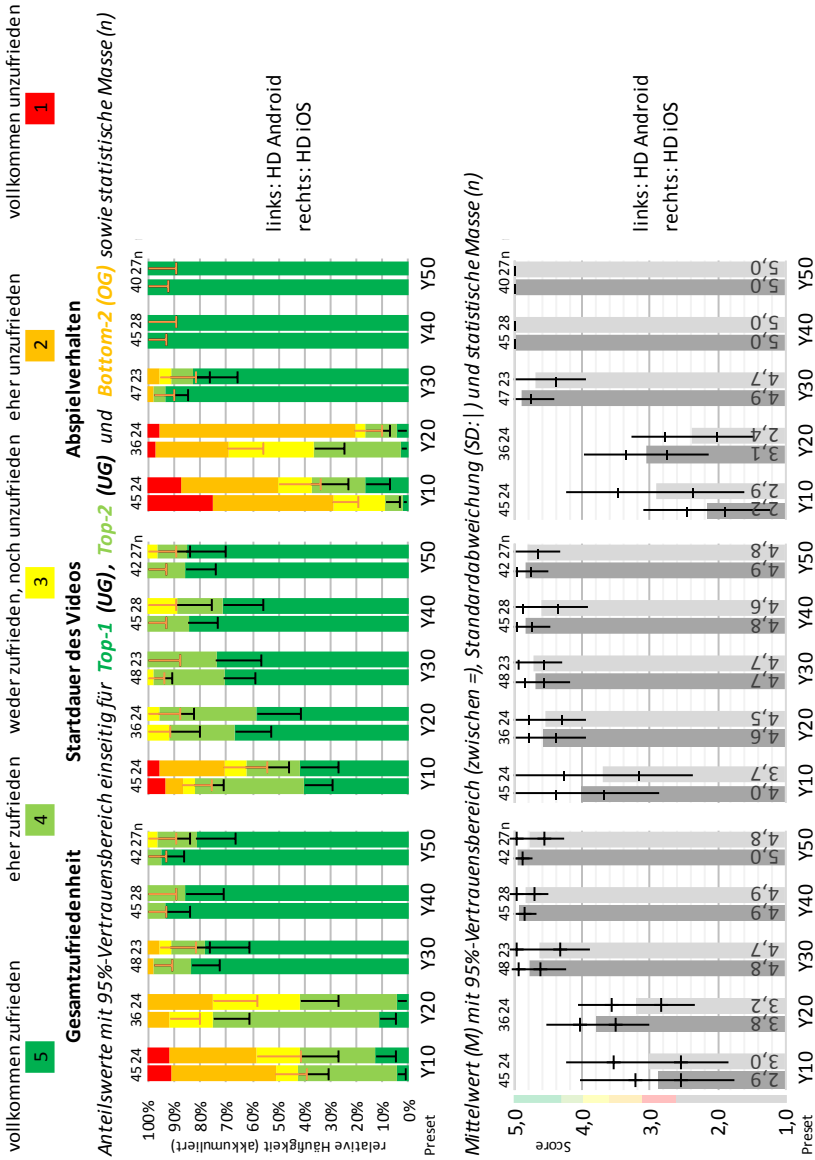


Abbildung 111: Zusammenstellung der bedingten Verteilungen zur Zufriedenheitsbewertung für Youtube (Teil 2)

Mit Blick auf das Abspielverhalten bei Y10 zeigt sich, dass die einmalige Unterbrechung mit automatischem Qualitätsstufenwechsel bei iOS tendenziell etwas besser bewertet wird als die mehrfachen Unterbrechungen bei Android, wobei aber beide Plattformen nicht zufriedenstellend sind, siehe Abbildung 111. Beim Schritt von Y10 zu Y20 würdigen die Probanden bei Android das verbesserte Abspielverhalten bei Y20 (weniger Unterbrechungen). Bei iOS hingegen ist beim Schritt von Y10 zu Y20 mit dem Fehlen bzw. unvorteilhaften Eingreifen der automatischen Qualitätumschaltung tendenziell eher eine Verschlechterung der Bewertungen zum Abspielverhalten festzustellen, wohingegen sich die Gesamtzufriedenheit nur wenig ändert.

Für die Bestimmung der Schwellwerte sollte berücksichtigt werden, dass die Nutzer jede Beeinträchtigung des Abspielverhaltens abstrafen. Daher sollten die Unterbrechungen mit Sicherheit ausgeschlossen werden. Ausgehend von den Erkenntnissen zur Art der Inhalte (variable Inhaltsdatenrate mit deutlichen Spitzen) erscheint eine großzügige Reserve in der Download-Datenrate angeraten, insbesondere wenn keine Dienstgütemechanismen genutzt werden oder diese sich nicht an die jeweiligen Anforderungen der Inhalte anpassen.

Die Beeinflussungen des Nutzererlebnisses durch Wiedergabestartzeit und Beeinträchtigungen des Abspielverhaltens sind ggf. in Relation zur Länge der getesteten Videos zu setzen. Für längere Videos ist u. U. von anderen kritischen Schwellen für die Stimuli auszugehen (2 Unterbrechungen in einem Clip von 60 s sind nicht akzeptabel, 2 Unterbrechungen in einem Clip von 60 min eventuell schon eher ...).

9.6.9 Visuelle Qualität des Videoinhalts

In der Reproduktion innerhalb dieser Arbeit, bzw. als Standbild allgemein, sind die Unterschiede in der visuell erfassbaren Qualität des Inhalts zwischen HD und HQ nicht adäquat darzustellen. Daher soll der **subjektive** Eindruck des Autors zur Vermittlung der visuellen Qualität bzw. der Unterschiede dienen. Die Unterschiede zwischen HD und HQ sind auf ein und demselben Gerät erkennbar, vor allem wenn zuvor schon einmal HD mit den mittlerweile hochauflösenden Displays der mobilen Geräte genutzt wurde. HD hat insgesamt einen (kleinen) Qualitätsvorteil, HQ ist aber auch noch visuell ansprechend. Bei einem direkten Side-by-Side-Vergleich fällt die Zuordnung HD bzw. HQ leichter. Der Vergleich über die Gerätegrenzen hinweg ist hingegen schwieriger, da sich zusätzlich u. a. die Display-Eigenschaften unterscheiden. Fazit: Auch mit HQ kann man zufrieden sein, aber HD ist visuell besser. Wie angekündigt, ist diese Einschätzung sehr subjektiv und kaum überraschend.

Zwischen den Apps der mobilen Plattformen besteht ein Unterschied hinsichtlich der expliziten Signalisierung der genutzten Qualitätsstufe, siehe Abbildung 131 (im Anhang): Android mit Indikator zur Qualitätsstufe (HD bzw. HQ im Test), iOS hingegen gibt keine explizite Indikation der Qualitätsstufe an die Nutzer.

Auf eine explizite Bewertung der visuellen Qualität durch die Nutzer wurde verzichtet. Eine Auswertung der Nutzerkommentare zeigt aber, dass einige Nutzer durchaus die HD-Qualität positiv bemerkten. Die hohe bzw. „Super-“ (Video-) Qualität im Fall von HD wird von den Probanden wahrgenommen und honoriert, siehe Abbildung 108. Die Analyse der Nutzerkommentare hinsichtlich der Wahrnehmung der Qualität des Inhalts ergab einige interessante Aspekte (genaue Auflistung der betreffenden Kommentare im Anhang A.6.2). Bei HQ war nur eine leichte Beschwerde zur technischen Qualität zu verzeichnen. Bei HD ergaben sich 11 positive Anmerkungen zur sehr guten Qualität, davon 7 von Tablet-Nutzer. Eventuell honorieren die Tablet-Nutzer die HD-Qualität häufiger, da aufgrund der größeren Bildschirmdiagonale die visuelle Qualität erst auf diesen Geräten richtig zur Geltung kommt.

Über die Inhaltsdatenrate bestimmt die Qualitätsstufe maßgeblich den technisch sinnvollen bzw. „guten“ Parameterbereich. So entsteht der Eindruck, dass die Qualitätsstufe HQ ungeachtet der schlechteren Inhaltsqualität insgesamt zufriedenerer Nutzer hat als HD, da bei der gewählten Parameterabstufung HQ bereits mit Y10 und Y20 problemlos und ohne Einschränkungen funktioniert, HD hingegen zu Unterbrechungen der Wiedergabe führt. Ein Nutzer mit HQ kann bei Y10 und Y20 zufrieden sein (und darüber auch). Ein Nutzer, der HD nutzt bzw. nutzen will, wird bei Y10 und Y20 wohl nicht zufrieden sein.

Die Nutzer sind also ggf. auch mit der Qualitätsstufe HQ zufrieden? Für eine abschließende Aussage müssten die Qualitätsstufen direkt miteinander verglichen werden und dabei auch die Videoqualität als gesonderte Bewertung erfasst werden. Die Erwartungshaltung und die entsprechenden Bewertungen sind dann eventuell anders, z. B. wenn die Nutzer wissen, dass sie HD haben könnten, sich aber für eine störungsfreie Wiedergabe mit HQ begnügen müssen.

Für die vorliegenden Ergebnisse ist nur festzustellen, dass die Probanden am Ace 2 mit HQ insgesamt zufriedener waren, dabei aber unklar ist, ob die Probanden in die Gesamtzufriedenheitsbewertung Qualitätsaspekte haben einfließen lassen, bzw. sich bewusst waren, dass dies bereits die bestmögliche Qualität auf diesem Gerät ist. Insgesamt bestätigt sich aber, dass HD-Qualität längere Wartezeiten und vor allem Unterbrechungen der Wiedergabe im getesteten Szenario nicht aufwiegt. Damit kann der Qualitätsunterschied HQ zu HD als Beispiel eines Begeisterungsfaktors eingestuft werden (siehe 2.3.10). Es ist davon auszugehen, dass es auch für die Inhaltsqualität Schwellwerte gibt. Für noch niedrige Stufen als HQ ist daher ggf. mit Beeinträchtigungen der Gesamtzufriedenheit zu rechnen. Daneben sollte die Präsentation der visuellen Qualität in Abhängigkeit des Nebenfaktors Gerät bzw. Formfaktor nicht außer Acht bleiben. Dies ist für steigende Größe und Auflösung der Bildschirme anzunehmen (abnehmender Qualitätseindruck mit steigender Bilddiagonale bei gleichbleibender Videoauflösung).

9.7 Zusammenfassung und Ergebnisüberblick

9.7.1 Auswahl

Auf Basis der Erfahrungen der Testrunde 1 und der Überlegungen zur Relevanz entstand eine Zusammenstellung potenziell interessanter Kategorien (in Orientierung an Auflistung in Tabelle 4). Für jede Kategorie wurden selbst wiederum ein Dienst und ein passendes Szenario möglichst repräsentativ und relevant gewählt, siehe Tabelle 61 zur Auswahl und Begründung.

Tabelle 61: Übersicht zur Kategorie und Auswahl der mobilen Dienste im Testparcours der Testrunde 2

Dienst und Kürzel	Kategorie mit Nutzungsklasse und Nutzungsart	Auswahlgrund / Besonderheit
<u>D</u> rive	Dienstleistung Cloud/Storage (Speicherdienste)	Annahme zur zukünftig hohen Relevanz für Nutzer und Netz gute Kontrollierbarkeit und ggf. Verallgemeinerbarkeit
<u>F</u> acebook	Kommunikation Soziale Netze	sehr hohe allgemeine und spezifische Relevanz für Nutzer und Netz
<u>G</u> oogle Maps	Ortsbezogene Dienste (Location Based Services) Kartendienste/Navigation	breite Nutzerbasis und spezifische Charakteristik der Interaktivität
<u>M</u> TV-Music	Unterhaltung/Entertainment Audio-Streaming	Annahme einer zukünftig hohen Relevanz für Nutzer und Anbieter
<u>S</u> piegel.de	Informationssuche Web-Browsing/Nachrichten	sehr hohe allgemeine (Web-Browsing) wie auch spezifische Relevanz für Nutzer und Netz
<u>Y</u> outube	Unterhaltung/Entertainment Video-Streaming	sehr hohe allgemeine (Mobile Video) wie auch spezifische Relevanz für Nutzer und Netz mit Annahme zur zukünftigen Nutzung (HD)

9.7.2 Technische Sicht und Netznutzung

Die einzelnen Dienste sind sich aus technischer Sicht der Netznutzung ähnlich (Burst, HTTP, TCP), siehe Tabelle 62. Entsprechend lassen sich die relevanten Effektivwerte auch auf die jeweiligen Datenraten der genutzten Richtung zurückführen. Deutlichere Unterschiede sind in der Art der transportierten Inhalte festzustellen sowie in der prinzipiell höheren Rolle der Latenz in Abhängigkeit des Inhalts mit vielen Turns für Google Maps und Spiegel.de (sowie in Browsing-Phasen allgemein). Die Punkte zur Charakterisierung der Netznutzung sind im Abschnitt 3.4.6 erläutert.

Tabelle 62: Netznutzung der mobilen Dienste im Testparcours

Dienst	zeitl. Charakteristik	Haupt-richtung	L7 Protokoll	L4 Protokoll	relevante Effektivwerte	Turns bzw. Einzelelemente pro Sessions / Transactions
<u>D</u> rive	Burst	annähernd symmetrisch	HTTPS	TCP	Datenrate Download Datenrate Upload	1 Turn pro Datei (hier Bild)
<u>F</u> acebook	Burst	annähernd symmetrisch	HTTPS	TCP	Datenrate Download Datenrate Upload	1 Turn pro Datei (hier Bild) mehrere sequenziell
<u>G</u> oogle Maps	Burst	Down-load	HTTPS oder SPDY*	TCP	Datenrate Download	mehrere parallel und sequenziell je Zoom-/Verschiebeschritt
<u>M</u> TV-Music	Burst	Down-load	HTTP	TCP	Datenrate Download (Pufferzeitfenster)	1 Turn pro Datei (hier Musiktitel)
<u>S</u> piegel .de	Burst	Down-load	HTTP	TCP	Datenrate Download	mehrere Turns parallel und sequenziell pro Seite
<u>Y</u> outube	Burst	Down-load	HTTP	TCP	Datenrate Download (Pufferzeitfenster)	1 pro Datei (hier Video) oder in mehreren Teilen sequenziell

* bei Google Maps: keine sichere Unterscheidung von „außen“ möglich, da verschlüsselte Kommunikation

L7 und L4 bezogen auf die OSI-Stack-Schicht

Angaben in der Spalte Turns bei MTV-Music und Youtube jeweils auf die Streaming-Phase bezogen, beim Browsing sonst wie Spiegel.de

9.7.3 Nutzersicht und QoS-QoE-Charakteristik

Bei der Einschätzung der Interaktivität der Szenarien zeigen sich große Unterschiede, siehe Tabelle 63. Im Szenario Drive ist die Interaktivität für den Download auf insgesamt dreimaliges Fingertippen reduziert, wovon allein ein einziger Fingertipp die relevante Aktion des Downloads mit anschließender Wartezeit auslöst. Bei Spiegel.de ist zumindest schon eine Nutzung in mehreren Schritten notwendig, in deren Verlauf die verschiedenen Nutzerstimuli präsentiert werden. Bei Google Maps muss sich der Nutzer durch die Interaktion das Ergebnis in mehreren zusammenhängenden Zügen und den damit verbundenen Nutzerstimuli regelrecht erarbeiten. Die ausgewählten Anwendungen des Testparcours lassen sich den Klassen Finish-Time und Streaming zuordnen, siehe Tabelle 63. Anwendungen der Klasse Echtzeit waren nicht im Test vertreten. Entsprechend der Einschätzungen zu den relevanten Stimuli und der Schwellwertcharakteristik erfolgte die Wahl der QoS-Parameterbereiche.

Tabelle 63: Klassifizierung und relevante Stimuli der mobilen Dienste im Testparcours

Dienst	Interaktivität	Klasse	QoS-abhängige Stimuli*	QoS-Parameterbereich	Schwellwertbereich	QoE-Verbesserung
Drive	niedrig	Finish Time	Dauer der Dateiübertragungen	mittel – hoch	weit	kontinuierlich
Facebook	mittel	Finish Time	Ladedauern	niedrig – mittel	weit	kontinuierlich
Google Maps	hoch	Finish Time	Ladedauern im Flow beim Laden der Karte	mittel – hoch	weit	kontinuierlich
MTV-Music	mittel	Streaming / Finish Time	Wiedergabestartzeit (Abspielverhalten)	niedrig – mittel	mittel	sprunghaft / kontinuierlich
Spiegel.de	mittel	Finish Time	Ladedauern der abgerufenen Seiten	niedrig – mittel	weit	kontinuierlich
Youtube	niedrig	Streaming	Abspielverhalten Wiedergabestartzeit (Qualität)	hoch	eng	sprunghaft

* Angaben in Klammern: im Test nur bedingt relevant für die Nutzerzufriedenheit, in jeweils absteigender Bedeutung genannt

Die Voruntersuchungen offenbarten Schwächen in der Präsentation der Nutzerstimuli einiger Anwendungen, die sich so auch in der Auswertung bestätigten. Dabei konnten für jede der drei gebildeten Kategorien entsprechende Negativbeispiele gefunden werden:

- Zeiten und Dauern: Ladekreisel bei Drive unter iOS
- Funktion und Verhalten: Funktionsdefizit ohne Rückmeldung an den Nutzer bei MTV-Music unter iOS
- Qualität und Quantität: keine Angabe zur Datenmenge bei Drive unter iOS, fehlender Qualitätsindikator bei Youtube unter iOS

9.7.4 Parametrisierung

Die zusammengefassten Darstellungen der ausgewählten Netzparameterstufen je Dienst mit Zuordnung der Presets sind der Abbildung 112 sowie der Tabelle 70 (Anhang) zu entnehmen. Die ausgewählten Reihen, Parameterbereiche und Abstufungen wurden auf Basis der Vorüberlegungen jeweils passend gewählt, meist als kombinierte Reihe zur Variation aller drei Hauptparameter.

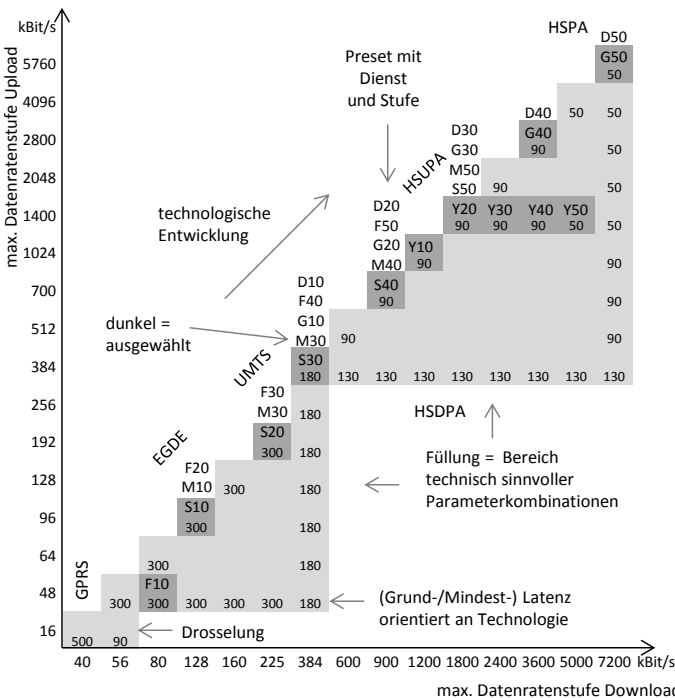


Abbildung 112: Zuordnung der Presets zu ausgewählten Netzparameterstufen

9.7.5 Besonderheiten der Szenarien

Bedingt durch die Heterogenität der Technik, das Testdesign und weitere Aspekte ergaben sich einige Besonderheiten, die je Szenario unterschiedlich zu berücksichtigen waren, siehe Tabelle 64. Diese Besonderheiten bedurften der Beachtung in der Testplanung bzw. Auswertung. Je nachdem wie gut das Verständnis des Zusammenwirkens und die Kontrolle der konkreten Einflussgrößen ausfällt, können die Szenarien zwischen Black Box und White Box eingeordnet werden:

- sehr schlecht/schlecht (schwarz): keine Einsicht in interne Zusammenhänge und relevante Nebenfaktoren, daher auch keine Kontrolle möglich (war durch Vorüberlegungen und Voruntersuchungen ausgeschlossen)
- mittel bis sehr gut (grau): Spanne von grundlegendem Verständnis und Kontrolle bis zu zumindest gutem oder gar sehr gutem Verständnis und Kontrolle, wobei aber trotzdem eine letzte Restunsicherheit verbleibt (Malus der Echtheit)
- perfekt (weiß): vollständiges Verständnis der internen Zusammenhänge und völlige Kontrolle der Nebenfaktoren, faktisch nicht zu erreichen, da dem die Echtheit des Tests entgegensteht

9.7.6 Versionen der mobilen Apps

Die Voruntersuchungen wie auch der Probandentest sind unter Nutzung bestimmter App-Versionen entstanden. Die Erkenntnisse, Aussagen und Abbildungen zu den Apps beziehen sich auf diese Versionen. Im Rahmen der Phase der Voruntersuchungen wurden weitere Updates aus den App-Stores bis zum Start der Probandentests berücksichtigt, sofern sich zeigte, dass diese lediglich Bugfixes etc. enthielten, aber nicht die technische Sicht oder die Nutzersicht deutlich veränderten. Ab dem Startzeitpunkt der Probandentests in Form des Pretests wurden die App-Versionen über den gesamten Testzeitraum nicht mehr verändert – eine Art App-Redaktionsschluss – auch wenn zwischenzeitlich neuere Versionen zu beziehen gewesen wären. Die genauen Versionen der einzelnen Apps sind der Tabelle 65 zu entnehmen.

Tabelle 64: Besonderheiten zu den Szenarien der mobilen Dienste im Testparcours

Dienst	Besonderheit durch Heterogenität der Technik	Besonderheit durch Testdesign in Bezug zum Inhalt	weitere Besonderheit bzw. Sonstiges	Black Box to White Box *
<u>Drive</u>	App: Android vs. iOS durch Nutzersicht			sehr gut
<u>Face-book</u>	App: Android vs. iOS durch Datenmenge		Gegenseite und Inhalt durch Bildanzahl	mittel
<u>Google Maps</u>	App: App vs. Browser durch Datenmenge (+ Endgerät durch gefühlte Leistung)		Probanden durch Aufgabenbearbeitung	mittel
<u>MTV-Music</u>	App: Android vs. iOS bei M10 durch Stimuli (Verhalten)	Variante A vs. B bei M10 durch Datenmenge beim Browsing		sehr gut
<u>Spiegel .de</u>		1. vs. 2. Aufruf durch Cache		gut
<u>Youtube</u>	Inhalt: HD vs. HQ durch Datenrate (Android vs. iOS durch Präsentation)	(Variante A vs. B bei HD Y10 und Y20 durch leichte Unterschiede in den Stimuli)	HD (noch) nicht mobilfunktypisch (3G)	sehr gut

* Verständnis und Kontrolle der Nebenfaktoren im Test und Aufdeckung der internen Zusammenhänge

Tabelle 65: Übersicht zu den genutzten Versionen der mobilen Apps im Probandentest

Dienst	Nutzung in	Android		iOS		Anmerkung
		Version	Datum	Version	Datum	
<u>Drive</u>	App	1.1.592.10	20.02.2013	1.3.1.4684	12.04.2013	
<u>Face-book</u>	App	3.2.1	17.05.2013	6.1.1	13.05.2013	
<u>Google Maps</u>	App	6.14.4	14.05.2013	1.1.6.5479	05.03.2013	iPhone-App (iOS)
	Browser			Safari		nur iPad 4
<u>MTV-Music</u>	App	4.2.1.48	29.05.2013	2.2.19	20.05.2013	iOS: iPhone-App
<u>Spiegel .de</u>	Browser	„Internet“		Safari		System-Browser
<u>Youtube</u>	App	4.4.11	04.04.2013	1.3	15.04.2013	

9.7.7 Übersicht zum Bearbeitungserfolg

Die rekodierten Angaben zum Bearbeitungserfolg für alle Dienste und Presets sind der Tabelle 73 im Anhang zu entnehmen. Einige Aussagen sind verallgemeinernd bzw. vergleichend möglich. Einige Szenarien waren weniger problematisch als andere (Angaben auf je 420 Aufgabenausführungen bezogen):

- niedrige Anzahl Probleme (Anzahl < 20): Youtube und Spiegel
- mittlere Anzahl Probleme ($20 \leq \text{Anzahl} < 40$): Google Maps und MTV-Music
- hohe Anzahl Probleme (Anzahl ≥ 40): Facebook und Drive

Die unterschiedliche Anzahl Probleme ist im Zusammenhang mit der Problemangabe „Internetverbindung“ (besonders bei Facebook) und auch ausgehend von unterschiedlichen Vorkenntnissen sowie Usability-Problemen in den Szenarien zu sehen. Probleme mit der Internetverbindung in den Bearbeitungen der Aufgaben wurden von den Probanden tatsächlich vorwiegend bei langsamen Presets korrekt genannt.

9.7.8 Nebenerkenntnisse

Aus der Zusammenstellung des Testparcours ergaben jeweils interessante Nebenerkenntnisse:

- Drive: deutlich unterschiedliche Zufriedenheit trotz gleicher Leistung allein durch die Präsentation der Nutzerstimuli bedingt
- Facebook: annähernd vergleichbare Bewertungen trotz Unterschieden in den Vorabannahmen zur Leistung, die durch den Einfluss zusätzlicher Nebenfaktoren (Inhalt und Präsentation) im gewählten Szenario allerdings nicht feiner aufzuschlüsseln sind
- Google Maps: deutliche Unterschiede in den Bewertungen, die auf Eigenschaften der Geräte (z. B. Displayauflösung und gefühlt langsames Gerät) sowie der genutzten Apps zurückzuführen sind
- MTV-Music: wie Nutzer durch QoS-bedingte Funktionsprobleme der App richtig frustriert werden – und warum Streaming manchmal doch durch die Wartezeiten dominiert wird
- Spiegel.de: die Nützlichkeit von Caches beim mobilen Browsing, vor allem bei langsamem Netz

- Youtube: Bestätigung der Unterbrechungsfreiheit als wichtigste QoE-Einflussgröße beim Streaming von Videoinhalten, HD eher von Tablet-Nutzern positiv kommentiert, was ein Hinweis zur Präsentation sein kann

Intuitiv einleuchtend und verständlich, fast schon trivial, aber auch empirisch zu belegen: „schlechte“ Apps benötigen mehr Netzleistung als „gute“ Apps. Eine HD-Tauglichkeit und große hochauflösende Displays können in höheren Anforderungen der Nutzer entsprechender Geräte an das Netz resultieren.

Daher ergeben sich an die App-Entwickler einige Empfehlungen, die sicherlich auch schon mehrfach an anderen Stellen zu lesen waren – und hoffentlich meist auch beachtet werden:

- Test der Apps außerhalb des berühmten goldenen Käfigs der eigenen Entwicklungsumgebung: „schlechte“ QoS-Parametern einplanen und austesten
- Wartezeiten lassen sich ggf. bei umfangreichen Datenübertragungen nicht vermeiden, der Nutzer empfindet diese aber zumindest in bestimmten Bereichen als weit weniger schlimm, wenn er dazu vernünftiges Feedback erhält.
- Sinnlose Statusmeldungen zum Fortschritt bzw. Erfolg frustrieren und verwirren die Nutzer.

9.7.9 Überblicksdarstellung der Ergebnisse

Durch die gemeinsame Darstellung der erzielten Ergebnisse für alle Dienste und als relevant eingeschätzter Partitionen werden Unterschiede und Gemeinsamkeiten deutlich, siehe Abbildung 113. Durch die QoS-Netzparametrisierung konnten die angestrebten QoE-Änderungen erreicht werden. Die Parameterbereiche für die Szenarien waren insgesamt sehr gut gewählt. Der interessierende Übergang von der Unzufriedenheit zur Zufriedenheit wurde jeweils passend erfasst. Innerhalb der Parameterraumgrenzen kann, mit Ausnahme von Drive unter iOS, sogar der Sättigungsbereich erreicht werden, in dem weitere QoS-Veränderungen kaum noch zu weiteren QoE-Verbesserungen führen.

Wie erwartet ergaben sich zwei unterschiedliche und unterscheidbare Charakteristiken des QoS-QoE-Zusammenhangs. Mit Ausnahme von Youtube sind die Mittelwertkurven als insgesamt ähnlich und lediglich im Parameterraum der QoS-Netzparameter zueinander verschoben zu erkennen. Die typische Kurvenform (kurvilinear) ist durch einen Bereich des steileren Anstiegs mit Übergang zu einer allmählichen Abflachung geprägt. Diese QoS-QoE-Charakteristik kann als typisch für die Klasse Finish-Time angesehen werden (elastischer Zusammenhang). Youtube HD zeigt hingegen einen wesentlich steileren Anstieg, der als sprunghaft bezeichnet werden kann, mit anschließender völliger Abflachung. Dies entspricht der Erwartung für die QoS-QoE-Charakteristik der Klasse Streaming (unelastisch). Bei Youtube HQ waren erwartungsgemäß keine QoE-Änderungen zu verzeichnen, da die Sprungstelle in ähnlicher Ausprägung wie bei HD erst bei niedrigeren Datenraten zu vermuten ist.

Anmerkungen zur Abbildung 113:

- interpolierte Darstellung (linear → logarithmisch) der Kurvenabschnitte zwischen den empirisch ermittelten Messwerten
- wertproportionale Darstellung der x-Achse mit der Komponente max. Datenrate im Download aus den Presets
- Farbkodierung zur Unterscheidung der Dienste, Kodierung der Partitionen durch Linienstil (durchgängig/gestrichelt)
- Die Farbskala am linken Rand und Bereichsgrenzen dienen zum Vergleich mit der MOS-Planungsskala auf Mittelwertbasis.
- Die Mittelwerte sind zur visuellen Vereinfachung ohne Vertrauensbereiche und ohne Standardabweichung angegeben, diese sind den Einzelergebnissen zu entnehmen.
- Die Darstellung ist auf die Gesamtzufriedenheit reduziert, da die unterschiedlichen Teilbewertungen zwischen den Diensten nicht vergleichbar sind.

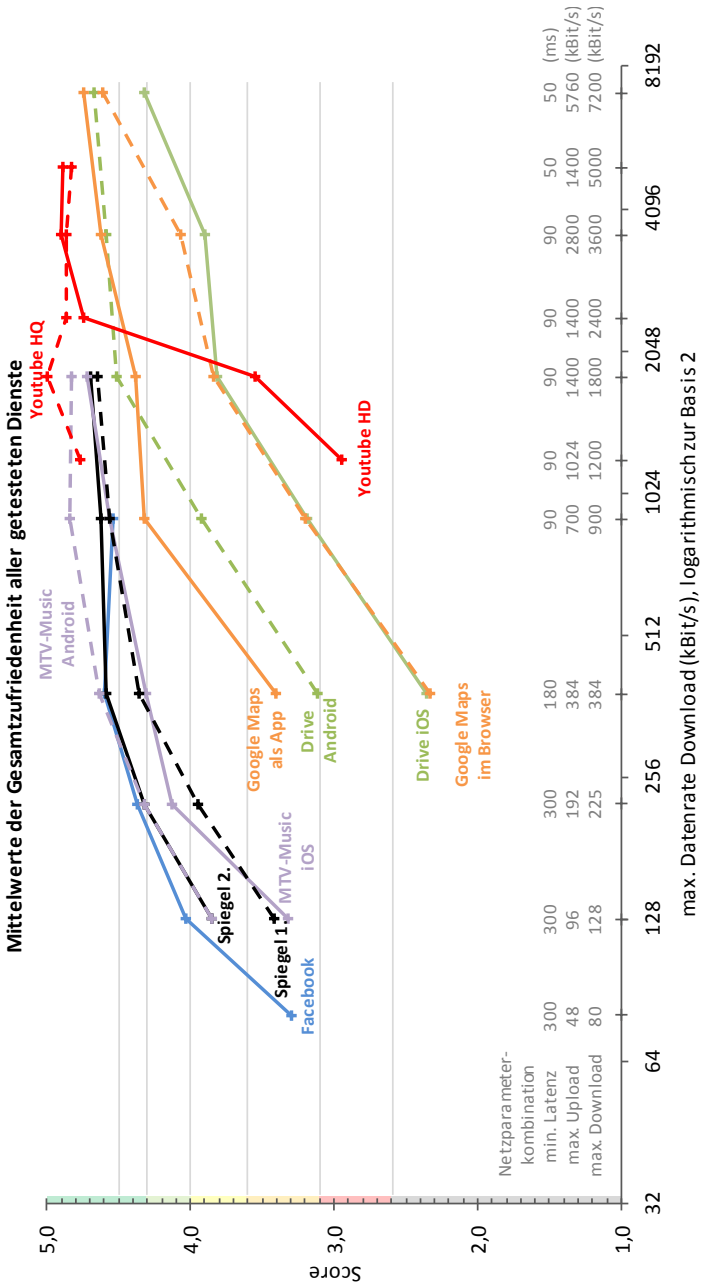


Abbildung 113: Zusammenstellung der Bewertungen der Zufriedenheit für alle Dienste im Testparcours

9.7.10 Schwellwerte

Tabelle 66 listet die Schwellwerte auf Basis der Erläuterungen zu den Anteilswerten auf. Für die meisten Dienste, unterschieden in Untermengen nach relevanten Nebenfaktoren, konnten Abschätzungen zu den Schwellwerten durch die getesteten Presets ermittelt werden. Über die genannten Presets lassen sich für die Schwellwerte ungefähre Richtwerte für die QoS-Netzparameter ableiten und die repräsentierten nutzerrelevanten Stimuli abschätzen (sofern für diese sinnvolle Erwartungswerte zu ermitteln waren).

Wie bei den verschiedenen Typen des QoS-QoE-Zusammenhangs vorhergesagt, liegen die Werte je nach Dienst (siehe Tabelle 63 zur Vorabeinschätzung) unterschiedlich weit auseinander und auf unterschiedlichem Niveau. Wie beabsichtigt hat die angesetzte QoS-Parametrisierung meist zwischen Stufe 2 und 4 (d. h. Preset 20 und 40) zur Zufriedenheit im Bereich des Zielwerts geführt. Die Bereiche und Abstufungen der MOS-Planungsskala passen als schnelle Abschätzung mit den ermittelten Schwellwerten auf Anteilsbasis weitgehend zusammen, siehe Abbildung 113.

Für einige Untermengen ist mangels statistischer Masse keine Aussage möglich, für andere Teilmengen keine direkte Aussage. Die gewählten aggregierten Anteilsschwellen für den Grenzwert bzw. Sättigungswert sind mit 20 % bzw. 80 % noch recht großzügig. Für diese Fälle wurde eine Abschätzung im Parameterraum auf Basis der definierten Lagebeziehungen der Schwellwerte getroffen (Angaben mit \geq). Gleichzeitig entstehen leichte Verschiebungen zwischen deutlich unterschiedlich großen Untermengen durch die Wirkung der statistischen Masse auf die Breite der betrachteten Vertrauensbereiche. Für verschärfte Werte von 10 % bzw. 90 % für die genannten Schwellen wäre eine höhere statistische Masse wünschenswert, andererseits wäre möglicherweise aber auch schon ein niedrigeres Konfidenzniveau ausreichend.

Tabelle 66: Schwellwerte der untersuchten Dienstszenarien im Test (bezogen auf die Grundgesamtheit)

Dienst	Untermenge	Grenzwert (Akzeptanz- grenze)		Zielwert		Sättigungswert (Over- Provisioning)	
		Bottom-2 obere Grenze ≤ 20 % *		Top-1 untere Grenze ≥ 33 % *		Top-2 untere Grenze ≥ 80 % *	
		Preset	Mittel- wert	Preset	Mittel- wert	Preset	Mittel- wert
Drive	Android (ohne Ace 2) iOS	D20	3,9	D30	4,5	D30	4,5
		D50	4,3	≥ D50	.	. (≥ D50)	.
Face- book	gesamt (Android) (iOS)	F20	4,0	F20	4,0	F40	4,6
		F20	4,0	F30	4,4	F40	4,5
		.	.	F30	4,3	F40	4,8
Google Maps	App Browser (iPad 4)	G20	4,3	G20	4,3	G30	4,4
		.	.	G50	4,6	. (≥ G50)	.
MTV- Music	Android iOS	M20	4,3	M20	4,3	M30	4,6
		M30	4,3	M40	4,6	M50	4,7
Spiegel .de	gesamt Aufruf 1 Aufruf 2	S20	4,2	S30	4,5	S30	4,5
		S30	4,4	S30	4,4	S40	4,6
		S20	4,3	S20	4,3	S30	4,6
You- tube	HD HQ	Y30	4,7	Y30	4,7	Y30	4,7

*obere und untere Grenze jeweils bezogen auf 95%-Vertrauensbereich;

. n zu klein zur Bestimmung;

Untermenge in Klammern: Aggregation oder Partitionierung nicht sinnvoll;

Schwellwert in Klammern: keine Bestimmung innerhalb der Parameterraumgrenzen, daher Angabe der Richtung;

Youtube HQ vollständig oberhalb des Änderungsbereichs und jenseits des Sättigungswertes, daher ohne Angabe;

Ergänzend sind zu den Presets der Schwellwerte die gehörigen Mittelwerte der Bewertungen angegeben. Dies erlaubt den Vergleich der Mittelwertkurven in Abbildung 113 zu den eingezeichneten R-Value-Bereichsgrenzen.

10 Das Fazit

Zusammenfassung und Ausblick

Am Ende der Arbeit stehen der zusammenfassende Blick zurück auf den Weg und die erreichten Ziele sowie ein Blick voraus auf offene Punkte.

10.1 Der Blick zurück: Zusammenfassung und Zielerreichung

Im Mittelpunkt der Arbeit stand der Zusammenhang von QoS und QoE als der Einfluss der QoS-Netzparameter zusammen mit weiteren Nebenfaktoren auf das Nutzererlebnis (QoE), konkret die Nutzerzufriedenheit, für ausgewählte mobile Dienste im mobilen Internet.

Es sei an dieser Stelle auch an die ausführlichen Zusammenfassungen zu den Inhalten der einzelnen Kapitel erinnert, die jeweils am Ende der Kapitel 2 bis 9 zu finden sind.

10.1.1 Methodischer Ansatz

Der gewählte methodische Ansatz mit einer Annäherung sowohl von der Theorie als auch von der Empirie aus hat sich weitgehend bewährt. Zur Untersuchung des QoS-QoE-Zusammenhangs wurden rationale Erklärungsansätze, eine passende Modellierung und eine empirische Bestimmung in einem Nutzertest geeignet miteinander verbunden.

Die Arbeit versucht dabei immer den Abgleich zwischen der allgemeinen Beschreibung in Form von Prinzipien und den speziellen Festlegungen und Ergebnissen des eigenen Tests. Für den Nutzertest wurde eine hohe Realitätsnähe mit echten Nutzern, Endgeräten und Diensten angestrebt und umgesetzt. Die Heterogenität dieser Faktoren bedeutet dabei immer ein Abwägen zwischen zusätzlichem Aufwand und zusätzlichem Erkenntnisgewinn. Da die Aussagen zur Empirie nur statistisch zu ermitteln sind, musste die ausgeführte Studie entsprechend skaliert werden, um einen adäquaten Umfang zu erreichen.

10.1.2 Theorie und Modellierung

Neben der praktischen Umsetzung war auch die Schaffung einer soliden theoretischen Grundlage wichtig, um in den Untersuchungen zu besseren Einsichten der internen Zusammenhänge zu kommen. Bei der begrifflichen Einordnung von QoS und QoE zeigten sich bereits die ersten Tücken, da eine Vielzahl von Definitionen für QoS und QoE existiert, sie dabei aber entweder zu kurz greifen oder zu vielfältig ausfallen. Sowohl QoS als auch QoE sind komplexe Konzepte. Dazu wurden jeweils die relevanten Einflussfaktoren und Maße erörtert. Es wurde ausgehend vom Begriff der Qualität gezeigt, wie diese in QoS bzw. QoE zu verstehen ist. Im Fall der QoS ist diese oftmals auf eine Beschreibung der Beschaffenheit (technische Charakteristik) des Netzes in einer Ende-zu-Ende-Sicht reduziert. Für die QoE wurden die subjektive Bewertung des Nutzererlebnisses als Qualität und die zugehörigen Nutzungs- und Beurteilungsprozesse intensiv diskutiert. Neben Maschine/System und Mensch/Nutzer nimmt der (mobile) Kontext eine besondere Rolle ein. In der Diskussion zum konzeptuellen Zusammenhang von QoS und QoE wurde gezeigt, dass der Zusammenhang dienstspezifisch ist und als Eignung (Netzsituation) zum Zweck (Nutzungssituation) gesehen werden kann.

Aufbauend auf den Konzepten zum Zusammenhang von QoS und QoE wurde das QoX-Matrix-Modell entwickelt. Das Modell integriert alle als relevant identifizierten Qualitätselemente: Kontext, Nutzer, Netz, Endgerät und Dienst (bestehend aus App, Inhalt, Gegenseite). Die Ergründung und wertmäßige Bestimmung der Zusammenhänge basiert darauf, in verschiedenen Schichten des Modells die Ursache und Wirkung zu erfassen, zu verstehen und ggf. vorherzusagen zu können. Dabei werden QoS-veränderliche interne Zwischengrößen als Mediatoren und externe Nebenfaktoren als Moderatoren des Zusammenhangs berücksichtigt.

Für die Dienste im Modell wurde eine sinnvolle Abgrenzung zwischen Transportdienst des Netzes und den interessierenden Inhaltsdiensten des offenen Internets erörtert. Zur weiteren Beschreibung der Dienste als Inhaltsdienste im offenen Internet wurden die technische Netzsicht und die Nutzersicht auf die Dienste im Modell eingeführt, die über die Voruntersuchungen bis zur Beschreibung des Testparcours und der Auswertung innerhalb der vorliegenden Arbeit immer wieder aufgegriffen wurden. Über die herausgearbeiteten

Punkte der technischen Sicht/Netzseite kann die Netznutzung der Dienste in unterschiedlicher Granularität umfassend charakterisiert werden. Für die Nutzersicht werden innerhalb der Arbeit drei Kategorien von Nutzerstimuli zur Beschreibung der QoS-abhängigen Leistung in der Wahrnehmung der Nutzer vorgeschlagen. Dabei wird dargelegt, warum die Art der Präsentation in der Nutzersicht eine entscheidende Rolle einnimmt. Ausgehend von der gemeinsamen Betrachtung von technischer Netzseite und Nutzersicht kann für die Dienste eine Bildung von Dienstklassen sinnvoll sein. Für diese sind im Modell über die beschriebene Wirkung von QoS-Parameteränderungen unterschiedliche Grundcharakteristiken zum QoS-QoE-Zusammenhang zu identifizieren und vorherzusagen. Je nach Art und Wirkung der Nutzerstimuli sollte zwischen Diensten mit eher sprunghafter QoS-QoE-Änderungscharakteristik (feste technischen Anforderungen) und Diensten mit eher kontinuierlicher QoS-QoE-Änderungscharakteristik über einen weiten Bereich (von der subjektiven Erwartung geprägt) zu differenzieren sein.

10.1.3 Testbed und Technik

Für die Einbeziehung des Netzes bzw. der interessierenden Netzsituationen in die empirischen Untersuchungen wurde die Realisierung als Netzemulation als tauglich befunden. Zur Abbildung der Netzsituationen wurden typische Wertkombinationen der technischen Leistungsparameter gebildet. Dabei orientieren sich die Werte und Wertbereiche an theoretischen Überlegungen zur Mobilfunktechnologie im Abgleich zu Erhebungen mit typischerweise erzielten Werten in Mobilfunkproduktivnetzen. Die betrachteten Netzsituationen umfassen die maximale Datenrate im Download, die maximale Datenrate im Upload sowie die technologiebedingte Latenz als Minimalwert. Für die Berücksichtigung der Netzsituationen in der Studie waren einige Vereinfachungen im Vergleich zum echten Mobilfunk zu treffen (keine Mobilität, Annahme der Limitierung im drahtlosen Zugangsnetz, keine räumliche/zeitliche Dynamik), die jeweils zu beachten sind. Es wird daher nicht von einer exakten Ursachennachbildung, sondern von der äquivalenten Wirkung der erreichten Effektivwerte im entwickelten Modell ausgegangen.

Das realisierte eigene Testbed zur Netzemulation konnte wie geplant die relevanten Netzparameter schnell und flexibel variieren und so die interessierenden Netzsituationen gezielt herstellen. Die Realisierung basiert dabei auf Open

Source Software und üblicher WLAN-Hardware zur drahtlosen Anbindung der echten mobilen Endgeräte (Smartphones/Tablets). Dabei bietet die realisierte Lösung des Testbeds durch einen verteilten Ansatz eine einfache Skalierung der Technik von einer Instanz zur Netzemulation auf eine Vielzahl gleichartiger Geräte gleichzeitig, um parallel und ohne ungewollte Beeinträchtigungen mehrere mobile Endgeräte in den Test einzubeziehen. Zusätzliche Möglichkeiten wie die Ausleitung und Aufzeichnung der Verkehrs- und Videodaten der Endgeräte haben sich im Testbetrieb bewährt. Die damit erzeugten Aufzeichnungen bilden den Ausgangspunkt für optionale Schritte zur Tiefenauswertung. Die Steuerung und Nutzung des Testbeds wird durch ein Framework aus geschaffenen Web-APIs und eine Web-GUI stark vereinfacht. Die auf dieser Basis realisierte automatische Testlaufsteuerung war eine wesentliche Voraussetzung für die erfolgreiche Skalierung des Laborexperiments auf den erreichten Umfang. Das entwickelte Testbed zur Netzemulation ist für zukünftige Entwicklungen offen und flexibel erweiterbar.

Innerhalb des Testbeds wurde eine Auswahl relevanter, marktüblicher populärer Endgeräte eingesetzt. Dabei waren Smartphones und Tablets der beiden führenden Plattformen Android und iOS in unterschiedlichen Leistungsklassen im Einsatz. Spezielle Anpassungen innerhalb des Testbeds erlauben dabei die problemlose Zusammenarbeit zwischen Netzemulatoren und mobilen Endgeräten per WLAN.

10.1.4 Voruntersuchungen und Vorüberlegungen zu den Diensten

Als ein grundlegendes Ergebnis zur Einbeziehung der mobilen Dienste für die Nutzungssituationen sind aus den Vorüberlegungen und Voruntersuchungen die Prinzipien zur Testrelevanz, Testeignung und Testeffizienz hervorgegangen. Dabei werden jeweils Netzsicht und Nutzersicht auf die Dienste berücksichtigt. Für die Testrelevanz spielen die Relevanz für die Nutzer wie auch die Relevanz für die Netze und Prognosen zur weiteren Entwicklung eine bedeutende Rolle. Die Überprüfung der Testeignung erlaubt die Berücksichtigung der Besonderheiten des Testbeds (z. B. Netzemulation und Heterogenität der Endgeräte) als auch des Testdesigns (z. B. Aufgabenwiederholungen und Aufgabenvarianten). Die Überlegungen zur Testeffizienz dienen für jeden Dienst zur Eingrenzung des Parameterraums, zur Identifikation dienstspezifisch relevanter Nebenfaktoren, zur Modellierung interner Wirkzusammenhänge (Grey

Box) und zur Bildung von Testreihen durch Auswahl geeigneter Netzparameterkombinationen mit zugehörigen Erwartungswerten der Nutzerstimuli sowie der Charakterisierung ihres Änderungsverhaltens.

Es wurden sechs relevante mobile Dienste für die Untersuchung in der Studie ausgewählt, diese bilden den Testparcours. Zu jedem der ausgewählten Dienste entstand ein praktikables Nutzungsszenario, das eine typische Aufgabenstellung zur Erzeugung der beabsichtigten Nutzerstimuli durch die Probanden selbst enthält.

10.1.5 Empirie und QoE-Evaluation

Durch die Beteiligung des Menschen als unscharfem und weichem, d. h. nicht-deterministischem Element im Zusammenhang von QoS und QoE ergibt sich die Notwendigkeit einer empirischen Untersuchung. Diese QoE-Evaluation zur Überprüfung der QoS-QoE-Charakteristiken und zur Quantifizierung des Zusammenhangs erfolgte als Laborexperiment, d. h. unter kontrollierten Bedingungen einer Testumgebung mit gezielter Variation der QoS-Netzparameter als unabhängige Variablen und unter Berücksichtigung der übrigen Nebenfaktoren. Innerhalb der dienstspezifischen Nutzungsszenarien werden so Netz- und Nutzungssituationen planmäßig zusammengebracht. Als abhängige Variablen wurden neben subjektiven Bewertungen der Probanden zur Zufriedenheit (relevantes QoE-Maß) auch der Bearbeitungserfolg und Nutzerkommentare erfasst. Die Beschreibung der Aufgabenszenarios und Bewertung der Aufgabenausführungen erfolgten in einem Web-Fragebogen. Die Probanden bearbeiteten die Aufgaben selbstständig. Durch die entwickelte Kopplung von Web-Fragebogensystem und Netzemulations-Testbed konnte die Variation der Netzbedingungen automatisch gesteuert werden, so dass eine hohe Parallelität mit 6 Testplätzen erreicht wurde.

Es gab Ein- bzw. Ausschlusskriterien für die Teilnahme von Probanden an der Studie: Vorkenntnisse in der Nutzung einer der mobilen Plattformen im Test (Android/iOS) und die aktive Nutzung des mobilen Internets waren erforderlich. Die Probanden können als vornehmlich junge, gut gebildete (studentische) Stichprobe charakterisiert werden. Es erfolgte eine Nutzertypisierung

hinsichtlich der Nutzungsintensität des mobilen Internets (annähernd lognormalverteilt) und Bewertungstendenz im Testparcours (überwiegend ausgeglichen mit wenigen sehr kritischen Probanden).

10.1.6 Auswertung des Nutzertests

Es konnten die Datensätze von 210 Probanden mit insgesamt 2520 Aufgabenausführungen in der Auswertung genutzt werden. Erst die Berücksichtigung aller relevanten Aspekte hilft, aus den gesammelten Daten die interessierenden Aussagen zu generieren und diese richtig einzuschätzen. Die Auswertung wurde in Einzelteile zerlegt und dazu jeweils Vorschläge zur geeigneten visuellen Aufbereitung unterbreitet. Die Auswertung der Nutzerkommentare und des Bewertungsverhaltens mit dem Intra-Rater-Vergleich geben über die quantitative Analyse der bedingten Verteilungen hinaus Hinweise darauf, wie die Probanden die QoS-Änderungen in QoE-Änderungen umgesetzt haben.

Für den QoS-QoE-Zusammenhang mit bedingten Verteilungen wurde dargelegt, warum Mittelwerte und Anteilswerte für ein vollständiges Bild der Daten wichtig sind. Mit dem Schätzen in Form der Berechnung der zugehörigen Vertrauensbereiche wurde die analytische Statistik in die Auswertung einbezogen und die dazu notwendigen statistischen Annahmen zur Skala und angenommenen statistischen Verteilung diskutiert. Es wurde ausgehend von den genutzten Metriken das Prinzip der Schwellwerte mit Grenzwert, Zielwert und Sättigungswert eingeführt. Auf Basis dieser Überlegungen ist eine Bestimmung der interessanten Parameterbereiche je Dienst als Qualitätsabgleich zwischen QoE und QoS möglich.

10.1.7 Erzielte Ergebnisse des Nutzertests

Durch die Variation der QoS-Netzparameter konnten wie beabsichtigt gezielte Unterschiede in den QoE-Bewertungen erreicht werden. Die Propositionen zu den nutzerrelevanten Stimuli und weiteren Nebenfaktoren wurden dabei weitgehend bestätigt. Die erzielten Ergebnisse können durch das Modell und die einbezogenen Einflussfaktoren erklärt werden. Die vorab festgelegten dienstspezifischen QoS-Parameterbereiche waren gut gewählt und zeigten wie gewünscht in den QoE-Bewertungen den Übergang zur Zufriedenheit. Die Frage nach einem „gut genug“ der untersuchten QoS-Parameter bei der Nutzung der ausgewählten Dienste konnte positiv beantwortet werden. Dabei bestätigten

sich die Annahmen zu unterschiedlichen Arten der Dienstkurven der QoS-QoE-Charakteristik für die getesteten Dienste. Besonders im direkten Vergleich war der Unterschied zwischen Youtube HD mit sprunghafter QoE-Änderung und den restlichen Diensten mit eher kontinuierlichen QoE-Veränderungen deutlich zu erkennen. Die gesuchten Schwellwerte waren unter Berücksichtigung der jeweils relevanten statistischen Masse überwiegend zu ermitteln. Die Bedeutung der QoS-invarianten Qualitätselemente zur QoE sollte nicht unterschätzt werden. Die hohe Relevanz einiger dieser Faktoren, wie die passende Präsentation der Nutzerstimuli in der Nutzersicht, die bereits in Voruntersuchungen (als Experteneinschätzung) vermutet wurde, konnte auch in den Ergebnissen des Nutzertests nachgewiesen werden.

10.1.8 Was bleibt?

Die Dienstszenarien als Untersuchungsgegenstände sind bewegliche Ziele. Die Welt des mobilen Internets und der genutzten Dienste ändert sich kontinuierlich. Die präsentierten Ergebnisse der Voruntersuchungen und des Nutzertests stellen damit lediglich eine Momentaufnahme dar. Umso wichtiger sind das Verständnis der ausgearbeiteten relevanten Einflussgrößen auf die QoE, die Rolle der QoS-veränderlichen Größen dabei und der Zusammenhang mit weiteren Faktoren. Mit den grundsätzlichen Überlegungen zum Zusammenhang von QoS und QoE sollte die Aufbereitung der Theorie eine entsprechende Basis für weitere Forschungsarbeiten bieten können. Das entwickelte Modell, die Sichten zur Beschreibung der Dienste, die Prinzipien zur Voruntersuchung und Zusammenstellung eines Testparcours, die Bestandteile der Auswertung und die technische Realisierung des Netzemulations-Testbeds sind nicht auf die vorgestellten Dienste begrenzt, sondern sind in gleicher Art auch für weitere zukünftige QoE-Evaluationen nutzbar. Die ausführliche Diskussion auch zu den praktischen Aspekten der Vorbereitung, Ausführung und Auswertung der Studie sollte außerdem helfen, zukünftige QoE-Evaluationen von dem gesammelten Erfahrungsschatz profitieren zu lassen.

Beim Vergleich des Einflusses der Nebenfaktoren auf die Dienste bestätigte sich, dass genügend Netzleistung notwendig ist, alleine aber nicht für die Nutzerzufriedenheit ausreicht. Daher sollten beim Entwurf mobiler Dienste im Allgemeinen und insbesondere beim Entwickeln der mobilen Apps die Erkenntnisse zur QoS und QoE eine angemessene Berücksichtigung finden.

10.2 Der Blick voraus: Ausblick und offene Punkte

Erweiterung der Untersuchungsgegenstände: Bislang waren keine Dienste im Testparcours vertreten, die im Bereich Echtzeit einzuordnen sind. Entsprechende Kommunikations- oder Unterhaltungsdienste könnten eine interessante Ausdehnung der Untersuchungen bedeuten. Für weitere Untersuchungen wäre eine stärkere und gezieltere Berücksichtigung von Einflussfaktoren zur Erwartung interessant. Im bisherigen Ansatz wurden statische Netzsituationen realisiert, ausgehend von den Überlegungen zur äquivalenten Wirkung auf die QoE. Die dynamische Variation der Netzparameter während der Nutzung wäre eine möglicherweise interessante Weiterentwicklung, um zu ergründen, wie die Nutzer schnellwechselnde Netzbedingungen im Nutzungsprozess empfinden und bewerten. Dazu sind Verbesserungen am Testbed und Anpassungen an der Konzeption der Untersuchungen notwendig. Neben den erfassten subjektiven QoE-Bewertungen zur Zufriedenheit sind weitere QoE-Maße und auch die Einbeziehung objektiver Größen mögliche sinnvolle Erweiterungen in der QoE-Evaluation.

Erhöhung des Testumfangs und der Aussagekraft: Wie in der Auswertung gezeigt, wird trotz des zunächst hoch erscheinenden Umfangs des gesamten Tests beim Blick auf einzelne Ergebnisse unter der Beachtung von Nebenfaktoren wie Gerät oder Inhalt schnell die sinnvolle Grenze der statistischen Masse erreicht. In diesen Fällen könnte eine Erhöhung des Testumfangs helfen, die Ergebnisse weiter zu verbessern.

Verlassen der Laborumgebung: Laborexperimente haben Vorteile aber auch Nachteile. Daher wäre sicherlich die Kombination mit einer Feldstudie eine interessante Erweiterung. Statt der aktiven Kontrolle der Netzparameter als gezielte Variation könnte eine passive Kontrolle als Erfassung, mit der Aufklärung der natürlichen Variation der Parameter ex post kombiniert werden. Eine andere Art der grundsätzlichen Gestaltung der QoE-Evaluation wäre die Verbindung von zumindest teilweise aktiver Kontrolle und einem wesentlich erweiterten Testpanel in einem Crowdsourcing-Ansatz, wofür zu prüfen wäre, ob die entwickelten Testansätze und technischen Lösungen entsprechend angepasst und erweitert werden können.

Konsequenzen aus den Erkenntnissen: Die möglichen Konsequenzen in der praktischen Anwendung der Erkenntnisse blieben in der Arbeit unbeachtet. Dies wären Aspekte, die zu einem ganzheitlichen „Experience Management“ in Sinne der QoE führen. Dies umfasst neben der Bestimmung des QoS-QoE-Zusammenhangs auch Möglichkeiten und Mechanismen zur gezielten Steuerung oder so etwas wie einen QoE-Indexwert populärer mobiler Dienste (ähnlich einem Teuerungsindex für einen repräsentativ zusammengestellten Warenkorb).

Erweiterung der interdisziplinären Forschung: Neben den gesamten technischen Aspekten ist es vor allem der interdisziplinäre Ansatz, der Anknüpfungspunkte für weitere Forschungsarbeiten bietet, um die Nutzer noch besser zu integrieren und noch besser zu verstehen, was für sie wichtig ist, was sie eigentlich wollen und wie dies zu realisieren ist. Dazu finden sich in der Arbeit immer wieder Hinweise dazu, wo es für Elektrotechnik, Humanwissenschaften aber auch die Wirtschaftswissenschaften weitere Betätigungsfelder rund um QoE gibt.

Literaturverzeichnis

Hinweis: Die Angabe überlanger URLs erfolgt ohne zusätzliche Umbruch- oder Trennzeichen.

- [1] MERINO, Pedro; MARTINI, Maria G.; SCHATZ, Raimund; SKORIN-KAPOV, Lea; VARELA, Martin: Improving QoS and QoE for Mobile Communications. In: *Journal of Computer Networks and Communications*. 2013-09-11. Vol. 2013, p. 1–2. DOI 10.1155/2013/645174. ISSN 2090-7141.
- [2] ETSI TR 102 643 V1.0.2 - *Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services* [online]. Technical Report. European Telecommunications Standards Institute (ETSI), 2010.
[Zugriff am 2013-01-11]. Verfügbar unter:
http://www.etsi.org/deliver/etsi_tr/102600_102699/102643/01.00.02_60/tr_102643v010002p.pdf
- [3] VAN MOORSEL, Aad: Metrics for the internet age: Quality of experience and quality of business. In: *Fifth International Workshop on Performability Modeling of Computer and Communication Systems, Arbeitsberichte des Instituts für Informatik, Universität Erlangen-Nürnberg, Germany*. Citeseer, 2001. p. 26–31.
- [4] LAGHARI, K.U.R.; CRESPI, N.; CONNELLY, K.: Toward total quality of experience: A QoE model in a communication ecosystem. In: *IEEE Communications Magazine*. 2012-04. Vol. 50, no. 4, p. 58–65. DOI 10.1109/MCOM.2012.6178834. ISSN 0163-6804.
- [5] ERICSSON: *Keeping Smartphone users loyal* [online]. Firmenschrift. Stockholm : Ericsson, 2013. [Zugriff am 2013-09-11]. Verfügbar unter:
<http://www.ericsson.com/res/docs/2013/consumerlab/keeping-smartphone-users-loyal.pdf>
- [6] ALBY, Tom: *Das mobile Web*. München : Hanser, 2008. ISBN 9783446415072.
- [7] AKAMAI TECHNOLOGIES, INC.; BELSON, David (Hrsg.): *Akamai State of the Internet Report: 3rd Quarter, 2013 Executive Summary* [online]. Firmenschrift. Akamai Technologies, Inc., 2014. [Zugriff am 2014-02-09]. Akamai State of the Internet Report. Verfügbar unter:
<http://www.akamai.com/stateoftheinternet/index.html>
- [8] CISCO SYSTEMS, INC.: *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017 (White Paper)* [online]. White Paper, Firmenschrift. Cisco Systems, Inc., 2013. [Zugriff am 2013-03-27]. Visual Networking Index (VNI). Verfügbar unter:
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html
- [9] BUNDESNETZAGENTUR: *Jahresbericht 2013* [online]. Bonn : Bundesnetzagentur, 2014.
[Zugriff am 2014-10-10]. Verfügbar unter:
http://www.bundesnetzagentur.de/cln_1421/DE/Allgemeines/Presse/Pressemitteilungn/Jahresbericht2013-node.html
- [10] BÖHM, Katharina: Sieben Millionen neue „Mobilisten“!, Pressemitteilung 2014-09-25. *Arbeitsgemeinschaft Online Forschung* [online]. 2014-09-25. [Zugriff am 2014-09-25]. Verfügbar unter:
<http://www.agof.de/pressemitteilung-2014-09-25/>
- [11] ARBEITSGEMEINSCHAFT ONLINE FORSCHUNG: *AGOF mobile facts* [online]. zitiert nach de.statista.com. Frankfurt : Arbeitsgemeinschaft Online Forschung (AGOF), 2014. [Zugriff am 2014-09-24]. Verfügbar unter:
<http://de.statista.com/statistik/daten/studie/221673/umfrage/anzahl-der-mobilien-internetnutzer-in-deutschland/>

- [12] BRANDT, Mathias: Zahl der mobilen Internetnutzer steigt deutlich. *Statista Infografiken* [online]. 2014-03-11. [Zugriff am 2014-09-25]. Verfügbar unter: <http://de.statista.com/infografik/1984/mobile-internetnutzer-in-deutschland/>
- [13] STATISTISCHES BUNDESAMT: *Private Haushalte in der Informationsgesellschaft (IKT)* [online]. Wiesbaden : Statistisches Bundesamt, 2013. [Zugriff am 2014-09-25]. Verfügbar unter: <https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400137004.pdf>
- [14] KHALAF, Simon: Apps Solidify Leadership Six Years into the Mobile Revolution. *Flurry Insights* [online]. 2014-04-01. [Zugriff am 2014-04-18]. Verfügbar unter: <http://blog.flurry.com/bid/109749/Apps-Solidify-Leadership-Six-Years-into-the-Mobile-Revolution>
- [15] SANDVINE INCORPORATED ULC: *Sandvine Global Internet Phenomena Snapshot - 2H 2013, Europe, Mobile Access* [online]. Firmenschrift. Sandvine Incorporated ULC, 2013. [Zugriff am 2013-02-09]. Sandvine Global Internet Phenomena Report. Verfügbar unter: <https://www.sandvine.com/downloads/general/global-internet-phenomena/2013/2h-2013-global-internet-phenomena-snapshot-eur-mobile.pdf>
- [16] ALLOT COMMUNICATIONS: *Allot MobileTrends Report Q1 2014* [online]. Firmenschrift. Allot Communications, 2014. [Zugriff am 2014-03-05]. Allot MobileTrends. Verfügbar unter: http://www.allot.com/mobile_trends_Q1_2014.html?campid=701D0000000n2u4
- [17] SANDVINE INCORPORATED ULC: *Sandvine Global Internet Phenomena Report - 1H 2014* [online]. Firmenschrift. Sandvine Incorporated ULC, 2014. [Zugriff am 2014-09-25]. Sandvine Global Internet Phenomena Report. Verfügbar unter: <https://www.sandvine.com/downloads/general/global-internet-phenomena/2014/1h-2014-global-internet-phenomena-report.pdf>
- [18] VATM UND DIALOG CONSULT: *Marktstudie 2013: 15. TK-Marktanalyse Deutschland 2013* [online]. Berlin : VATM und Dialog Consult, 2013. [Zugriff am 2014-01-16]. VATM-Marktstudien. Verfügbar unter: http://www.vatm.de/uploads/media/2013_TK-Marktstudie.pdf
- [19] FLANAGAN, Michael: *Extreme mobile data usage* [online]. Firmenschrift. Arieso, 2014. [Zugriff am 2014-09-25]. Verfügbar unter: <http://www.arieso.com/extrememobiledatausage>
- [20] D'AMBROSIA, John; MOONE, Paul; NOWELL, Mark: *400 Gb/s Ethernet: Why Now?* [online]. White Paper. ethernet alliance, 2013. [Zugriff am 2013-07-10]. Verfügbar unter: http://www.ethernetalliance.org/wp-content/uploads/2013/05/EthernetAlliance_400GWhyNow_techbrief_FINAL.pdf
- [21] STEINMETZ, Ralf: *Multimedia-Technologie*. 3. Auflage. Berlin; Heidelberg : Springer, 2000. ISBN 3540673326.
- [22] E-PLUS MOBILFUNK GMBH & Co. KG: *Mobilfunkmonitor 2013* [online]. Firmenschrift. E-Plus Mobilfunk GmbH & Co. KG, 2013. [Zugriff am 2013-08-09]. Verfügbar unter: http://eplus-gruppe.de/wp-content/uploads/2013/06/e-plus_mobilfunkstudie_2013_final.pdf
- [23] WALLACE, Sarah; YEADON, Chris; KUMAR, Kishore: Experience Driven CRM – Building Architectures to Provide a Differentiating Customer Experience. [online]. Vortragsfolien. 2013-12-10. [Zugriff am 2013-12-10]. Verfügbar unter: <https://webinar.lightreading.com/15890>
- [24] Exkurs: Theorie und Empirie. *Lehrstuhl für Internationale Politik und Außenpolitik der Universität Köln, Prof. Dr. Thomas Jäger* [online]. [Zugriff am 2014-08-28]. Verfügbar unter: <http://www.jaeger.uni-koeln.de/index.php?id=impressum>
- [25] THALLER, Georg Erwin: *Software-Test: Verifikation und Validation*. Hannover : Heise, 2002. ISBN 3882291982.

- [26] PETERSEN, Harro: *Grundlagen der statistischen Versuchsplanung*. Landsberg/Lech : ecomed, 1991. ISBN 360965340X.
- [27] GAVRAS, Anastasius (Hrsg.): *Experimentally driven research* [online]. White Paper. ICT-Fireworks, 2010. [Zugriff am 2013-08-24]. Verfügbar unter: http://www.ict-fireworks.eu/fileadmin/documents/Experimentally_driven_research_V1.pdf
- [28] *DIN 55350-11:2008-05, Begriffe zum Qualitätsmanagement – Teil 11: Ergänzung zu DIN EN ISO 9000:2005*. Deutsches Institut für Normung e. V. (DIN), 2008.
- [29] KLEMM, Peter: Thesen zum Qualitätsbegriff. In: *Zeitschrift für ärztliche Fortbildung*. 1986. 80. Jahrgang, Nr. 9/1986, S. 353–354.
- [30] MÖLLER, Sebastian: *Quality Engineering Qualität kommunikationstechnischer Systeme*. Berlin, Heidelberg : Springer, 2010. ISBN 9783642115486.
- [31] RAAKE, Alexander; EGGER, Sebastian: Quality and Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 11–33. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [32] *ITU-T Recommendation G.1000 - Communications Quality of Service: A framework and definitions* [online]. International Telecommunication Union (ITU), 2001. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.1000>
- [33] *DIN EN ISO 9000:2005-12, Qualitätsmanagementsysteme – Grundlagen und Begriffe (ISO 9000:2005)*. Deutsches Institut für Normung e. V. (DIN), 2005.
- [34] STEIN, Erich: *Taschenbuch Rechnernetze und Internet*. 3. neu bearb. Auflage. München : Fachbuchverl. Leipzig im Carl-Hanser-Verl., 2008. ISBN 9783446409767.
- [35] *ITU-T Recommendation E.800 - Definitions of terms related to quality of service* [online]. International Telecommunication Union (ITU), 2008. [Zugriff am 2013-08-19]. Recommendations - E Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-E.800-200809-I/en>
- [36] BAAH-ACHEAMFUOR, Kwame: Quality of Service and Quality of Experience in Fixed-Line and Mobile Multimedia Services. *XIII REGULATel-BEREC Summit Meeting* [online]. Vortragsfolien. Buenos Aires. 2014-07-21. [Zugriff am 2014-10-09]. Verfügbar unter: http://www.regulatel.org/wordpress/wp-content/uploads/2014/09/martes/kwame_baah-acheamfuor.pdf
- [37] *ETSI TR 121 905 V12.0.0 - Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UMTS); LTE; Vocabulary for 3GPP Specifications (3GPP TR 21.905 version 12.0.0 Release 12)* [online]. Technical Report. European Telecommunications Standards Institute (ETSI), 2014. [Zugriff am 2014-12-06]. Verfügbar unter: http://www.etsi.org/deliver/etsi_tr/121900_121999/121905/12.00.00_60/tr_121905v120000p.pdf
- [38] *ITU-T Recommendation E.800 - Terms and definitions related to quality of service and network performance including dependability* [online]. International Telecommunication Union (ITU), 1994. [Zugriff am 2013-08-19]. Recommendations - E Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-E.800-199408-S/en>
- [39] CRAWLEY, E.; SANDICK, H.; NAIR, R.; RAJAGOPALAN, B.: RFC 2386: *RFC 2386 - A Framework for QoS-based Routing in the Internet* [online]. Request for Comments. Internet Engineering Task Force (IETF), 1998. [Zugriff am 2014-11-25]. Verfügbar unter: <https://tools.ietf.org/html/rfc2386.html>

- [40] DIGITAL SUBSCRIBER LINE FORUM: TR-126: *TR-126: Triple-play Services, Quality of Experience (QoE) Requirements* [online]. DSL Forum Technical Report. Digital Subscriber Line Forum, Architecture & Transport Working Group, 2006. [Zugriff am 2013-08-25]. Verfügbar unter: <http://www.broadband-forum.org/technical/download/TR-126.pdf>
- [41] BALZER, Wolfgang: Quality of Service in Mobilfunk-Netzen. In: *funkschau*. Nr. 2009/9, S. 42–44. ISSN 0016-2841.
- [42] SERVICEBAROMETER AG: *Serviceprofil: Mobilfunkanbieter 2014* [online]. München : ServiceBarometer AG, 2014. [Zugriff am 2015-02-24]. Kundenmonitor Deutschland. Verfügbar unter: https://www.servicebarometer.net/kundenmonitor/de/serviceprofile.html?file=files/public_docs/sp_serviceprofile/DE2014/KMDE2014_SP_068_de.pdf
- [43] LE CALLET, Patrick; MÖLLER, Sebastian; PERKINS, Andrew (Hrsg.): Qualinet White Paper on Definitions of Quality of Experience (2012) - Version 1.2. In: LE CALLET, Patrick; MÖLLER, Sebastian; PERKINS, Andrew (Hrsg.), *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)* [online]. Lausanne, Switzerland, 2013. [Zugriff am 2014-06-03]. Verfügbar unter: http://www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf
- [44] *ITU-T Recommendation E.804 - Quality of service aspects for popular services in mobile networks* [online]. International Telecommunication Union (ITU), 2014. [Zugriff am 2015-01-04]. Recommendations - E Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-E.804/en>
- [45] STILLER, Burkhard: *Quality of service: Dienstgüte in Hochleistungsnetzen*. Bonn; Albany [u.a.] : Internat. Thomson Publ., 1996. ISBN 3826601718.
- [46] *ITU-T Recommendation Y.1241 - Support of IP-based services using IP transfer capabilities* [online]. International Telecommunication Union (ITU), 2001. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-Y.1241>
- [47] *ITU-T Recommendation I.350 - General aspects of quality of service and network performance in digital networks, including ISDNs* [online]. International Telecommunication Union (ITU), 1993. [Zugriff am 2013-09-02]. Recommendations - I Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-I.350/en>
- [48] *ETSI ETR 003 - Network Aspects (NA); General aspects of Quality of Service (QoS) and Network Performance (NP)* [online]. ETSI Technical Report. European Telecommunications Standards Institute (ETSI), 1994. [Zugriff am 2013-01-11]. Verfügbar unter: http://www.etsi.org/deliver/etsi_etr/001_099/003/02_60/etr_003e02p.pdf
- [49] *ITU-T Recommendation Y.1540 - Internet protocol data communication service – IP packet transfer and availability performance parameters* [online]. International Telecommunication Union (ITU), 2011. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-Y.1540>
- [50] *ITU-T Recommendation Y.1541 - Network performance objectives for IP-based services* [online]. International Telecommunication Union (ITU), 2011. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-Y.1541>
- [51] *ITU-T Recommendation Y.1542 - Framework for achieving end-to-end IP performance objectives* [online]. International Telecommunication Union (ITU), 2010. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-Y.1542-201006-I/en>
- [52] VOGEL, A.; KERHERVE, B.; BOCHMANN, G. von; GECSEI, J.: Distributed multimedia and QoS: a survey. In: *IEEE MultiMedia*. 1995. Vol. 2, no. 2, p. 10–19. DOI 10.1109/93.388195. ISSN 1070-986X.

- [53] DAVIES, Elwyn; CARLSON, Mark A.; WEISS, Walter; BLACK, David; BLAKE, Steven; WANG, Zheng: *RFC 2475 - An Architecture for Differentiated Service* [online]. Request for Comments. Internet Engineering Task Force (IETF), 1998. [Zugriff am 2015-01-16]. Verfügbar unter: <http://tools.ietf.org/html/rfc2475>
- [54] WROCLAWSKI, John: *RFC 2210 - The Use of RSVP with IETF Integrated Services* [online]. Request for Comments. Internet Engineering Task Force (IETF), 1997. [Zugriff am 2015-01-16]. Verfügbar unter: <http://tools.ietf.org/html/rfc2210>
- [55] WANG, Zheng: *Internet QoS: architectures and mechanisms for quality of service*. San Francisco [u.a.] : Kaufmann, 2001. The Morgan Kaufmann series in networking. ISBN 1558606084.
- [56] BANET, Franz-Josef; GÄRTNER, Anke; TERMAR, Gerhard: *UMTS: Netztechnik, Dienstarchitektur, Evolution*. Bonn : Hüthig, 2004. ISBN 3826650344.
- [57] CHEN, Jyh-Cheng; ZHANG, Tao: *IP-based next-generation wireless networks: systems, architectures, and protocols*. Hoboken, NJ : Wiley, 2004. ISBN 0471235261.
- [58] KARADIMCE, Aleksandar; DAVCEV, Danco: QoE Model for Multimedia Content Delivery from mCloud to Mobile Devices. In: *International Journal in Foundations of Computer Science & Technology*. 2013-05-03. Vol. 3, no. 3, p. 1–9. DOI 10.5121/ijfcs.2013.3301. ISSN 1839-7662.
- [59] NOKIA CORPORATION: *Quality of Experience (QoE) of mobile services: Can it be measured and improved?* [online]. Firmenschrift; White Paper. NOKIA CORPORATION, 2004. [Zugriff am 2013-12-12]. Verfügbar unter: http://www.afutt.org/Qostic/qostic1/MOB-GD-MGQ-NOKIA-040129-Nokia-whitepaper_qoe_net-final.pdf
- [60] *ITU-T Recommendation P.10/G.100 - Vocabulary for performance and quality of service, Amendment 2: New definitions for inclusion in Recommendation ITU-T P.10/G.100* [online]. International Telecommunication Union (ITU), 2006. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.100>
- [61] *ETSI TR 126 944 V12.0.0 - Universal Mobile Telecommunications System (UMTS); LTE; End-to-end multimedia services performance metrics (3GPP TR 26.944 version 12.0.0 Release 12)* [online]. Technical Report. European Telecommunications Standards Institute (ETSI), 2014. [Zugriff am 2014-12-27]. Verfügbar unter: http://www.etsi.org/deliver/etsi_tr/126900_126999/126944/12.00.00_60/tr_126944v120000p.pdf
- [62] *ETSI TR 102 274 V1.1.2: Human Factors (HF); Guidelines for real-time person-to-person communication services* [online]. Technical Report. European Telecommunications Standards Institute (ETSI), 2004. [Zugriff am 2015-02-10]. Verfügbar unter: http://www.etsi.org/deliver/etsi_tr/102200_102299/102274/01.01.02_60/tr_102274v010102p.pdf
- [63] *ETSI EG 202 534 V1.1.3 - Human Factors (HF); Guidelines for real-time person-to-person communication services* [online]. ETSI Guide. European Telecommunications Standards Institute (ETSI), 2007. [Zugriff am 2013-01-11]. Verfügbar unter: http://www.etsi.org/deliver/etsi_eg/202500_202599/202534/01.01.03_60/eg_202534v010103p.pdf
- [64] LAGHARI, Khalil ur Rehman: *On quality of experience (QoE) for multimedia services in communication ecosystem*. Dissertation. Évry : Télécom SudParis, Université Pierre et Marie Curie, Institut national des télécommunication, 2012. [Thèse de doctorat dirigée par Crespi, Noël Informatique et Télécommunications Évry, Institut national des télécommunications]

- [65] FIEDLER, Markus: From QoS to QoE – Position Statement. In: FIEDLER, Markus; KILKKI, Kalevi; REICHL, Peter (Hrsg.), *09192 Abstracts Collection – From Quality of Service to Quality of Experience* [online]. Dagstuhl, Germany : Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2009. [Zugriff am 2014-02-14]. p. 6–7. Dagstuhl Seminar Proceedings. Verfügbar unter: <http://drops.dagstuhl.de/opus/volltexte/2009/2236>
- [66] THE OPEN GROUP: *SLA Management Handbook – Volume 4: Enterprise Perspective* [online]. Reading : TeleManagement Forum, 2004. [Zugriff am 2015-01-04]. ISBN 1931624518. Verfügbar unter: <https://www2.opengroup.org/ogsys/catalog/G045>
- [67] DIEPOLD, Klaus: The Quest for a Definition of Quality of Experience. In: *Qualinet Newsletter*. 2012. No. 2/2012, p. 2–8.
- [68] ALBEN, Lauralee: Quality of Experience: Defining the Criteria for Effective Interaction Design. In: *interactions*. 1996-Mai. Vol. 3, no. 3, p. 11–15. DOI 10.1145/235008.235010. ISSN 1072-5520.
- [69] WECHSUNG, Ina; MOOR, Katrien De: Quality of Experience Versus User Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 35–54. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [70] *DIN EN ISO 9241-210:2011-01, Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme*. Deutsches Institut für Normung e. V. (DIN), 2010.
- [71] *ITU-T Recommendation G.1080 - Quality of experience requirements for IPTV services* [online]. International Telecommunication Union (ITU), 2008. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.1080>
- [72] REITER, Ulrich; BRUNNSTRÖM, Kjell; MOOR, Katrien De; LARABI, Mohamed-Chaker; PEREIRA, Manuela; PINHEIRO, Antonio; YOU, Junyong; ZGANK, Andrej: Factors Influencing Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 55–72. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [73] HEINECKE, Andreas M.: *Mensch-Computer-Interaktion* [online]. Berlin, Heidelberg : Springer, 2012. [Zugriff am 2015-01-09]. X.media.press. ISBN 9783642135064. Verfügbar unter: <http://link.springer.com/10.1007/978-3-642-13507-1>
- [74] RASMUSSEN, Jens: *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY, USA : Elsevier Science Inc., 1986. ISBN 0444009876.
- [75] RADWARE, LTD.: *Progressive Image Rendering: Good vs Evil?* [online]. Firmenschrift. Radware, Ltd., 2014. [Zugriff am 2014-09-27]. Insights from Neuroscientific Research. Verfügbar unter: <http://www.radware.com/neurostrata-fall2014/>
- [76] SUN, John Z.; WANG, Grace I.; GOYAL, Vivek K; VARSHNEY, Lav R.: A framework for Bayesian optimality of psychophysical laws. In: *Journal of Mathematical Psychology*. 2012-09. DOI 10.1016/j.jmp.2012.08.002. ISSN 0022-2496.
- [77] REICHL, Peter; TUFFIN, Bruno; SCHATZ, Raimund: Logarithmic laws in service quality perception: where microeconomics meets psychophysics and quality of experience. In: *Telecommunication Systems*. 2011-06-18. Vol. 52, no. 2, p. 587–600. DOI 10.1007/s11235-011-9503-7. ISSN 1018-4864, 1572-9451.
- [78] MOSER, Christian: *User Experience Design*. Berlin, Heidelberg : Springer, 2012. X.media.press. ISBN 9783642133626, 9783642133633.
- [79] GOTTHARTSLEITNER, Hannes; EBERLE, Peter; STARY, Christian: Zur Verschränkung von User Experience und Usability Engineering: Merkmale, Prinzipien und Vorgehensmodelle. In: *Zeitschrift für Arbeitswissenschaft*. 2009. Nr. 2009/3, S. 193–211. ISSN 0340-2444.

- [80] HASSENZAHN, Marc; PLATZ, Axel; BURMESTER, Michael; LEHNER, Katrin: Hedonic and Ergonomic Quality Aspects Determine a Software's Appeal. In: *Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems*. New York, NY, USA : ACM, 2000. p. 201–208. CHI '00. ISBN 1581132166.
- [81] NIELSEN, Jakob: *Usability engineering*. San Francisco : Morgan Kaufmann Publishers, 1994. ISBN 9780080520292.
- [82] BROOKS, P.; HESTNES, B.: User Measures of Quality of Experience: Why Being Objective and Quantitative Is Important. In: *IEEE Network*. 2010-04. Vol. 24, no. 2, p. 8–13. DOI 10.1109/MNET.2010.5430138. ISSN 0890-8044.
- [83] DIX, Alan; FINLAY, Janet; ABOUW, Gregory; BEALE, Russell: *Mensch, Maschine, Methodik*. New York; London; Toronto; Sydney; Tokyo; Singapur; München : Prentice Hall, 1995. ISBN 3930436108.
- [84] VERSHOFEN, Wilhelm: *Handbuch der Verbrauchsforschung. Erster Band: Grundlegung*. Berlin, Heymanns Verlag, 1940.
- [85] SCHATZ, R.; EGGER, S.; PLATZER, A.: Poor, Good Enough or Even Better? Bridging the Gap between Acceptability and QoE of Mobile Broadband Data Services. In: *2011 IEEE International Conference on Communications (ICC)*. 2011-06. p. 1–6.
- [86] REICHL, Peter; TUFFIN, Bruno; MAILLÉ, Patrick: Economics of quality of experience. In: HADJANTONIS, Antonis M.; STILLER, Burkhard (Hrsg.), *Telecommunication Economics* [online]. Springer, 2012. p. 158–166. Lecture Notes in Computer Science, 7216. [Zugriff am 2015-02-17]. ISBN 9783642303814, 9783642303821. Verfügbar unter: http://link.springer.com/chapter/10.1007/978-3-642-30382-1_21
- [87] KAISER, Marc-Oliver: *Kundenzufriedenheit kompakt: Leitfaden für dauerhafte Wettbewerbsvorteile*. Erich Schmidt Verlag GmbH & Co., 2006. ISBN 9783503097302.
- [88] FESTGE, Fabian; MEYER, Matthias; SCHWAIGER, Manfred: 22/2005: *Die Zufriedenheit mit dem Studium der Betriebswirtschaftslehre an der Technischen Universität Dresden* [online]. München : Ludwig-Maximilians-Universität München, 2005. [Zugriff am 2014-02-20]. Schriften zur Empirischen Forschung und Quantitativen Unternehmensplanung, ISSN 1862-9059. Verfügbar unter: http://www.imm.bwl.uni-muenchen.de/forschung/schriftenefo/ap_efoplan_22.pdf
- [89] NERDINGER, Friedemann W.; NEUMANN, Christina: Kundenzufriedenheit und Kundenbindung. In: MOSER, Klaus (Hrsg.), *Wirtschaftspsychologie* [online]. Springer Berlin Heidelberg, 2007. S. 127–146. [Zugriff am 2014-02-20]. ISBN 9783540716365, 9783540716372. Verfügbar unter: http://link.springer.com/chapter/10.1007/978-3-540-71637-2_8
- [90] MEYER, Anton; DORNACH, Frank: *Kundenmonitor Deutschland: Qualität und Kundenorientierung; Jahrbuch der Kundenorientierung in Deutschland 2000*. München : ServiceBarometer, 2000. ISBN 9783980652131.
- [91] EUCUSA-Methode. *Wikipedia* [online]. 2013. [Zugriff am 2013-10-30]. Verfügbar unter: <http://de.wikipedia.org/w/index.php?title=EUCUSA-Methode&oldid=117905242> [Page Version ID: 117905242]
- [92] STEINBEIS-TRANSFERZENTRUM UNTERNEHMENSENTWICKLUNG AN DER HOCHSCHULE PFORZHEIM: Lösungen zur Kundenzufriedenheitsmessung. *Management Monitor* [online]. [Zugriff am 2013-09-18]. Verfügbar unter: <http://www.management-monitor.de/monitor/Loesungen/Kundenzufriedenheit.php>
- [93] KANO, Noriaki; SERAKU, Nobuhiko; TAKASHI, Fumio; TSUJI, Shin-ichi: Attractive Quality and Must-Be Quality. In: *Journal of the Japanese Society for Quality Control*. 1984-04. Vol. 14, no. 2, p. 147–156. ISSN 0386-8230.

- [94] SAUERWEIN, Elmar; BAILOM, Franz; MATZLER, Kurt; HINTERHUBER, Hans H.: The Kano model: How to delight your customers. In: *International Working Seminar on Production Economics*. 1996. p. 313–327.
- [95] GEERTS, D.; DE MOOR, K.; KETYKÓ, I.; JACOBS, A.; BERGH, J. van den; JOSEPH, W.; MARTENS, L.; DE MAREZ, L.: Linking an integrated framework with appropriate methods for measuring QoE. In: *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*. 2010. p. 158–163.
- [96] KILKKI, Kalevi: Quality of Experience in Communications Ecosystem. In: *J. UCS, Journal of Universal Computer Science*. 2008. Vol. 14, no. 5, p. 615–624. ISSN 0948-6968.
- [97] HOßFELD, Tobias; BIEDERMANN, Sebastian; SCHATZ, Raimund; PLATZER, Alexander; EGGER, Sebastian; FIEDLER, Markus: The memory effect and its implications on web QoE modeling. In: *Proceedings of the 23rd International Teletraffic Congress*. 2011. S. 103–110.
- [98] JOHNSTON, Robert: The zone of tolerance: Exploring the relationship between service transactions and satisfaction with the overall service. In: *International Journal of Service Industry Management*. 1995-05-01. Vol. 6, no. 2, p. 46–61. DOI 10.1108/09564239510084941. ISSN 0956-4233.
- [99] GREIS, Thomas: Usability und User Experience unterscheiden. *ProContext Consulting GmbH* [online]. 2010-03-31. [Zugriff am 2014-10-10]. Verfügbar unter: <http://www.procontext.com/aktuelles/2010/03/usability-und-user-experience-unterscheiden.html>
- [100] WILAMOWITZ-MOELLENDORFF, Mageritta von; HASSENZAHN, Marc; PLATZ, Axel: Dynamics of user experience: How the perceived quality of mobile phones changes over time. In: *User Experience - Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction*. 2006. p. 74–78.
- [101] WEISS, Benjamin; GUSE, Dennis; MÖLLER, Sebastian; RAAKE, Alexander; BOROWIAK, Adam; REITER, Ulrich: Temporal Development of Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 133–147. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [102] MANTOVANI, Giuseppe: *Social Context in HCI: A New Framework for Mental Models, Cooperation, and Communication*. In: *Cognitive Science*. 1996. Vol. 20, no. 2, p. 237–269. DOI 10.1207/s15516709cog2002_3. ISSN 1551-6709.
- [103] THE NIELSEN COMPANY: Tech-Or-Treat: Consumers Are Sweet on Mobile Apps. *The Nielsen Company NEWSWIRE* [online]. 2014-10-30. [Zugriff am 2015-02-25]. Verfügbar unter: <http://www.nielsen.com/us/en/insights/news/2014/tech-or-treat-consumers-are-sweet-on-mobile-apps.html>
- [104] PATRICK, Andrew: Communications Psychology: Four Useful Concepts. *Workshop on the Social Aspects of Communication Technologies* [online]. Newcastle-upon-Tyne, UK. 2004-11-08. [Zugriff am 2013-06-05]. Verfügbar unter: <http://www.andrewpatrick.ca/cv/CommunicationsPsychologyConcepts.pdf> [Vortragsfolien]
- [105] MARCUS, Aaron: Six Degrees of Separation. In: *User Experience*. 2004. Vol. 2:6, no. Spring, p. 16.
- [106] ASK, Julie A.; SCHADLER, Ted; HAMMOND, Jeffrey S.; SHELDON, Peter; EVANS FREEMAN, Patti; WILLIS, Kelland; ROBERGE, Douglas: *Why eBusiness Pros Need A Five-Year Plan* [online]. Firmenschrift. Forrester Research, 2012. [Zugriff am 2014-09-19]. Verfügbar unter: <https://www.forrester.com/Why+eBusiness+Pros+Need+A+FiveYear+Plan/fulltext/-/E-RES82181>

- [107] TODE, Chantal: Mobile experiences are quickly falling short of consumer expectations: Forrester. *Mobile Marketer* [online]. 2012-10-17. [Zugriff am 2014-09-19]. Verfügbar unter: <http://www.mobilemarketer.com/cms/news/research/14011.html>
- [108] BODAMER, Stefan: *Verfahren zur relativen Dienstgütedifferenzierung in IP-Netzwerken*. Stuttgart : Dissertation, Universität Stuttgart, Inst. für Kommunikationsnetze und Rechnersysteme, 2004. ISBN 3922403980.
- [109] EDELL, R.; VARAIYA, P.: Providing Internet access: what we learn from INDEX. In: *IEEE Network*. 1999-09. Vol. 13, no. 5, p. 18–25. DOI 10.1109/65.793687. ISSN 0890-8044.
- [110] PERKIS, Andrew; REICHL, Peter; BEKER, Sergio: Business Perspectives on Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 97–108. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [111] BARAKOVIC, Sabina; SKORIN-KAPOV, Lea: Survey and Challenges of QoE Management Issues in Wireless Networks. In: *Journal of Computer Networks and Communications*. 2013. Vol. 2013, p. 1–28. DOI 10.1155/2013/165146. ISSN 2090-7141.
- [112] GERHARDT, Hans-Joachim; BERGMANN, Fridhelm; FROHBERG, Wolfgang: *Taschenbuch der Telekommunikation*. 2. vollständig überarbeitete Auflage. Carl Hanser Verlag GmbH & CO. KG, 2003. ISBN 3446217509.
- [113] BAUER, Ben; PATRICK, Andrew S.: A Human Factors Extension to the Seven-Layer OSI Reference Model. [online]. 2004-01-06. [Zugriff am 2014-02-08]. Verfügbar unter: <http://www.andrewpatrick.ca/OSI/10layer.html>
- [114] MINHAS, Tahir Nawaz: *Network Impact on Quality of Experience of Mobile Video*. Dissertation. Blekinge Tekniska Högskola; Blekinge Institute of Technology, 2012.
- [115] STEINMETZ, Ralf; NAHRSTEDT, Klara: *Multimedia Systems*. Berlin Heidelberg New York : Springer, 2004. ISBN 3540408673.
- [116] CASAS, P.; SEUFERT, M.; EGGER, S.; SCHATZ, R.: Quality of experience in remote virtual desktop services. In: *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. 2013-Mai. p. 1352–1357.
- [117] SCHATZ, Raimund; HOßFELD, Tobias; JANOWSKI, Lucjan; EGGER, Sebastian: From Packets to People: Quality of Experience as a New Measurement Challenge. In: BIRSACK, Ernst; CALLEGARI, Christian; MATIJEVIC, Maja (Hrsg.), *Data Traffic Monitoring and Analysis* [online]. Springer Berlin Heidelberg, 2013. p. 219–263. Lecture Notes in Computer Science, 7754. [Zugriff am 2013-05-21]. ISBN 9783642367830, 9783642367847. Verfügbar unter: http://link.springer.com/chapter/10.1007/978-3-642-36784-7_10
- [118] REICHL, P.; TUFFIN, B.; SCHATZ, R.: Economics of logarithmic Quality-of-Experience in communication networks. In: *2010 9th Conference on Telecommunications Internet and Media Techno Economics (CTTE)*. 2010-06. p. 1–8.
- [119] STANKIEWICZ, R.; CHOLDA, P.; JAJSZCZYK, A.: QoX: What is it really? In: *IEEE Communications Magazine*. 2011-04. Vol. 49, no. 4, p. 148–158. DOI 10.1109/MCOM.2011.5741159. ISSN 0163-6804.
- [120] HARDY, William C.: *QoS: Measurement and Evaluation of Telecommunications Quality of Service*. Chichester ; New York : John Wiley & Sons, 2001. ISBN 9780471499572.
- [121] DONEGAN, Patrick: *Mobile Network Outages & Service Degradations: A Heavy Reading Survey Analysis* [online]. Firmenschrift. Heavy Reading, 2013. [Zugriff am 2015-02-10]. Verfügbar unter: http://www.heavyreading.com/details.asp?sku_id=3103&skuitem_itemid=1524

- [122] DONEGAN, Patrick: *Mobile Network Outages & Service Degradations: A Heavy Reading Survey Analysis (Short Version)* [online]. Firmenschrift. Heavy Reading, 2013. [Zugriff am 2015-02-10]. Verfügbar unter: <http://www.netscout.com/library/Analyst%20Reviews/Mobile%20Network%20Outages%20-%20Short%20Version.pdf>
- [123] SHALITA, Steven: Maximizing the Customer Experience on 3G and 4G LTE. *Heavy Reading Report: Mobile Network Outages & Service Degradations* [online]. 2014-01-23. [Zugriff am 2014-01-23]. Verfügbar unter: <https://webinar.lightreading.com/17445> [Webinar mit Vortragsfolien]
- [124] O'MALLEY, Claire; BRUNDELL, Patrick; LONSDALE, Peter; BROOKS, Peter; SCHLIEMANN, Trond; FROWEIN, Han; HESTNES, Bjorn; AABY, Carl: Eye-2-Eye: Fitness for purpose of person-person communication technologies. In: *International Conference on Communication, Problem-Solving and Learning*. University of Strathclyde, 2001-06. p. 25–29.
- [125] BROOKS, Peter; BRUNDELL, Patrick; HAMNES, Kari; HEIESTAD, Svein; HEIM, Jan; HESTNES, Bjørn; HEYDARI, Bozi; O'MALLEY, Claire; SCHLIEMANN, Trond; u. a.: *Fitness-for-Purpose of Videotelephony in Face-to-Face Situations*. Final Report. ACTS Project AC314 Vis-à-Vis: CEC Deliverable A0314/NSSL/PB/DR/P/005/b1. 1999.
- [126] MAYER, Silvan: *Ein paketorientiertes Mobilfunksystem mit Dienstgüte-Unterstützung für Sprach- und Datendienste*. Stuttgart : Dissertation, Universität Stuttgart, Inst. für Kommunikationsnetze und Rechnersysteme, 2003. ISBN 3922403964. [ISBS 3922403964]
- [127] VARELA, Martín; SKORIN-KAPOV, Lea; EBRAHIMI, Touradj: Quality of Service Versus Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 85–96. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [128] RÖBNER, Thomas: *Basiswissen modellbasierter Test*. Heidelberg : dpunkt-Verl., 2010. ISBN 9783898645898.
- [129] JEKOSCH, Ute: Basic Concepts and Terms of „Quality“, Reconsidered in the Context of Product-Sound Quality. In: *Acta Acustica united with Acustica*. 2004-11-01. Vol. 90, no. 6, p. 999–1006.
- [130] SKORIN-KAPOV, L.; VARELA, M.: A multi-dimensional view of QoE: the ARCU model. In: *2012 Proceedings of the 35th International Convention MIPRO*. 2012-Mai. p. 662–666.
- [131] COMSCORE, INC.: *comScore MobiLens* [online]. zitiert nach de.statista.com. comScore, Inc., 2013. [Zugriff am 2013-10-28]. Verfügbar unter: <http://de.statista.com/statistik/daten/studie/170408/umfrage/marktanteile-der-betriebssysteme-fuer-smartphones-in-deutschland/>
- [132] PORTECK, Stefan: Butterweiche Bohne - Android 4.1 Jelly Bean im Alltagstest. In: *c't*. 2013-07-30. Nr. 17/2012, S. 52–53. ISSN 0724-8679.
- [133] FOCUS INFOCOM: "iPhone beats Android" Chip Online headlines. *Focus Infocom Pressemitteilungen* [online]. 2012-12-19. [Zugriff am 2013-04-09]. Verfügbar unter: http://www.focus-infocom.de/index.php?area=news&action=detail&news_id=58 [geänderte Version ohne Abbildungen aber ansonsten inhaltsgleich, Zugriff am 2014-05-29, verfügbar unter <http://www.focus-infocom.de/index.php/en/news/344-iphone-beats-android-chip-online-headlines>]
- [134] HUANG, Junxian; XU, Qiang; TIWANA, Birjodh; MAO, Z. Morley; ZHANG, Ming; BAHL, Paramvir: Anatomizing application performance differences on smartphones. In: *Proceedings of the 8th international conference on Mobile systems, applications, and services*. New York, NY, USA : ACM, 2010. p. 165–178. MobiSys '10. ISBN 9781605589855.

- [135] SCHATZ, Raimund: On the Impact of Terminal Performance and Screen Size on QoE. *ETSI Workshop on Selected Items on Telecommunication Quality Matters* [online]. Wien. 2012-11-27. [Zugriff am 2013-01-13]. Verfügbar unter: http://docbox.etsi.org/Workshop/2012/201211_STQWORKSHOP/S2_QUALITYofVOICEERVICESandNewFeatures/FTW_SCHATZ_ImpactTerminalQoE.pdf [Vortragsfolien]
- [136] ICKIN, S.; WAC, K.; FIEDLER, M.; JANOWSKI, L.; HONG, Jin-Hyuk; DEY, A.K.: Factors influencing quality of experience of commonly used mobile applications. In: *IEEE Communications Magazine*. 2012-04. Vol. 50, no. 4, p. 48–56. DOI 10.1109/MCOM.2012.6178833. ISSN 0163-6804.
- [137] THE NPD GROUP, INC.: Smartphone Data Consumption is 44 Percent Greater on Larger Screen Phones, According to NPD. *NPD Group Press Releases* [online]. 2013-11-18. [Zugriff am 2014-12-19]. Verfügbar unter: <https://www.npd.com/wps/portal/npd/us/news/press-releases/smartphone-data-consumption-is-44-percent-greater-on-larger-screen-phones-according-to-npd/> [Pressemitteilung]
- [138] LEONG, Jennifer: Optimizing Your App for Real-World Networks. *Velocity* [online]. New York. 2014-09-15. [Zugriff am 2014-09-21]. Verfügbar unter: <http://velocityconf.com/velocityny2014/public/schedule/detail/35644> [Vortrag mit Vortragsfolien]
- [139] EVANS, Shane: Virtualization Application Dependencies for Better Performance. *Velocity* [online]. New York. 2014-09-17. [Zugriff am 2014-09-17]. Verfügbar unter: <http://velocityconf.com/velocityny2014/public/schedule/detail/36960> [Vortrag mit Videomitschnitt]
- [140] TIMMERER, Christian: *Generic adaptation of scalable multimedia resources: interoperable device and coding format independent multimedia communication for universal access*. Saarbrücken : VDMVerlag, 2008. ISBN 9783639003963. [zugleich Dissertation, Alpen-Adria-Universität Klagenfurt, 2006]
- [141] STEVENS, W. Richard: *TCP/IP illustrated*. Reading, Mass : Addison-Wesley Pub. Co, 1994. Addison-Wesley professional computing series. ISBN 0201633469.
- [142] EIDENBERGER, Horst; DIVOTKEY, Roman: *Medienverarbeitung in Java: Audio und Video mit Java Media Framework & Mobile Media API*. Heidelberg : dpunkt-Verl., 2004. ISBN 3898641848.
- [143] WALKE, Bernhard: *Mobilfunknetze und ihre Protokolle 1*. Stuttgart : Teubner, 2001. ISBN 3519264307, 9783519264309.
- [144] SAUTER, Martin: *Grundkurs Mobile Kommunikationssysteme: UMTS, HSPA und LTE, GSM, GPRS, Wireless LAN und Bluetooth* [online]. 5. Auflage. Wiesbaden : Springer Fachmedien, 2013. [Zugriff am 2014-12-31]. ISBN 9783658014605, 9783658014612. Verfügbar unter: <http://link.springer.com/10.1007/978-3-658-01461-2>
- [145] PADHYE, Jitendra; FIROIU, Victor; TOWSLEY, Don; KUROSE, Jim: Modeling TCP Throughput: A Simple Model and Its Empirical Validation. In: *Proceedings of the ACM SIGCOMM '98 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. New York, NY, USA : ACM, 1998. p. 303–314. SIGCOMM '98. ISBN 1581130031.
- [146] CHAN, Mun Choon; RAMJEE, R.: Improving TCP/IP Performance over Third-Generation Wireless Networks. In: *IEEE Transactions on Mobile Computing*. 2008. Vol. 7, no. 4, p. 430–443. DOI 10.1109/TMC.2007.70737. ISSN 1536-1233.
- [147] SCHILLER, Jochen: *Mobilkommunikation*. München [u.a.] : Pearson-Studium, 2003. ISBN 3827370604.

- [148] JIANG, Haiqing; LIU, Zeyu; WANG, Yaogong; LEE, Kyunghan; RHEE, Injong: Understanding bufferbloat in cellular networks. In: *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*. 2012. p. 1–6.
- [149] GETTYS, Jim; NICHOLS, Kathleen: Bufferbloat: Dark Buffers in the Internet. In: *Queue*. 2011-11. Vol. 9, no. 11, p. 40–54. DOI 10.1145/2063166.2071893. ISSN 1542-7730.
- [150] CHAN, S.C.F.; CHAN, K.M.; LIU, K.; LEE, J.Y.B.: On Queue Length and Link Buffer Size Estimation in 3G/4G Mobile Data Networks. In: *IEEE Transactions on Mobile Computing*. 2014-06. Vol. 13, no. 6, p. 1298–1311. DOI 10.1109/TMC.2013.127. ISSN 1536-1233.
- [151] YANG, Peng; XU, Lisong: A survey of deployment information of delay-based TCP congestion avoidance algorithm for transmitting multimedia data. In: *GLOBECOM Workshops (GC Wkshps)*, 2011 IEEE. 2011. p. 18–23.
- [152] PFEIFFER, René: TCP and Linux' Pluggable Congestion Control Algorithms. In: *Linux Gazette*. 2007-02. No. 135. ISSN 1934-371X.
- [153] YANG, Peng; LUO, Wen; XU, Lisong; DEOGUN, Jitender; LU, Ying: TCP Congestion Avoidance Algorithm Identification. In: *Distributed Computing Systems (ICDCS)*, 2011 31st International Conference on. 2011-06. p. 310–321. ISBN 9781612843841.
- [154] XU, Lisong: Internet Congestion Control Census. *University of Nebraska-Lincoln, Department of Computer Science & Engineering* [online]. [Zugriff am 2013-09-02]. Verfügbar unter: <http://cse.unl.edu/~xu/research/TCPcensus.html>
- [155] YANG, Peng; SHAO, Juan; LUO, Wen; XU, Lisong; DEOGUN, J.; LU, Ying: TCP Congestion Avoidance Algorithm Identification. In: *IEEE/ACM Transactions on Networking*. 2014-08. Vol. 22, no. 4, p. 1311–1324. DOI 10.1109/TNET.2013.2278271. ISSN 1063-6692.
- [156] AJMONE MARSAN, Marco; CAROFIGLIO, Giovanna; GARETTO, Michele; GIACCONE, Paolo; LEONARDI, Emilio; SCHIATTARELLA, Enrico; TARELLO, Alessandro: Of Mice and Models. In: *Proceedings of the Third International Conference on Quality of Service in Multiservice IP Networks*. Berlin, Heidelberg : Springer-Verlag, 2005. p. 15–32. QoS-IP'05. ISBN 354024557X.
- [157] KÖHLER, Thomas: Anforderungen an ein TCP/IP-Netzwerk für Voice over IP. In: HEIN, Mathias; REISNER, Michael; VOß, Antje (Hrsg.), *Voice over IP - Sprach-Daten-Konvergenz richtig nutzen*. Poing : Franzis, 2002. S. 155–178. ISBN 3772366864.
- [158] KLEIN, Jeremy; FREEMAN, Jonathan; LESSITER, Jane; FERRARI, Eva; WILDS, Mark: *Assessing Network Quality of Experience* [online]. Sagentia, 2009. [Zugriff am 2014-09-10]. Verfügbar unter: <http://stakeholders.ofcom.org.uk/binaries/research/technology-research/NetworkQoE.pdf>
- [159] DAHM, Markus: *Grundlagen der Mensch-Computer-Interaktion*. München [u.a.] : Pearson-Studium, 2006. ISBN 3827371759.
- [160] ETSI TR 102 972 V1.1.1 - *Human Factors (HF); User Interfaces; Generic user interface elements for 3G/UMTS mobile devices, services and applications* [online]. Technical Report. Sophia Antipolis : European Telecommunications Standards Institute (ETSI), 2009. [Zugriff am 2015-01-16]. Verfügbar unter: https://portal.etsi.org/stfs/STF_HomePages/STF322/ETSI%20TR%20102%20972.pdf
- [161] SCHNEIDER, Matthias; NIMAN, Bruno von; KETOLA, Pekka; WILLIAMS, David: Design Guidelines for Better 3G User Interfaces. In: *The 2nd International Workshop on Mobile Internet User Experience, MIUX'08*. Amsterdam, 2008-09-02.
- [162] VALERDI, J.; GONZALEZ, A.; GARRIDO, F.J.: Automatic Testing and Measurement of QoE in IPTV Using Image and Video Comparison. In: *Fourth International Conference on Digital Telecommunications, 2009. ICDDT '09*. 2009-07. p. 75–81.
- [163] SEOW, Steven C.: *Designing and Engineering Time: The Psychology of Time Perception in Software*. Addison-Wesley Professional, 2008. ISBN 9780132702515.

- [164] GRIGORIK, Ilya: *High Performance Browser Networking* [online]. O'Reilly Media, 2013. [Zugriff am 2013-08-15]. ISBN 9781449364304. Verfügbar unter: <http://chimera.labs.oreilly.com/books/1230000000545>
- [165] LANG, Alfred: *Psychologie der Zeit*. In: ARNOLD, W.; EYSENCK, H. J.; MEILI, R. (Hrsg.), *Lexikon der Psychologie* [online]. Freiburg i.Br. : Herder, 1972. S. 38–43. [Zugriff am 2013-05-21]. Verfügbar unter: http://s272261905.online.de/old/pap1970-79/1972_psych_zeit.htm
- [166] KRIEGER, Mike: *Secrets to Lightning Fast Mobile Design. Warm Gun 2011* [online]. San Francisco, CA. 2011-12-02. [Zugriff am 2015-02-25]. Verfügbar unter: <https://speakerdeck.com/mikeyk/secrets-to-lightning-fast-mobile-design> [Vortragsfolien]
- [167] RHEEM, Carroll: *Consumer response to travel site performance* [online]. Firmenschrift. PhoCusWright Inc., 2010. [Zugriff am 2014-02-09]. Verfügbar unter: http://www.akamai.com/dl/whitepapers/Akamai_PCW_Travel_Perf_Whitepaper.pdf
- [168] WROBLEWSKI, Luke: *Mobile Design Details: Avoid The Spinner. LukeW Ideation & Design* [online]. 2013-09-17. [Zugriff am 2014-10-09]. Verfügbar unter: <http://www.lukew.com/ff/entry.asp?1797>
- [169] STEFANOV, Stoyan: *Psychology of performance. Velocity* [online]. Santa Clara, CA. 2010-06-23. [Zugriff am 2014-09-25]. Verfügbar unter: <http://de.slideshare.net/stoyan/psychology-of-performance> [Vortragsfolien]
- [170] STEFANOV, Stoyan: *Psychology of performance. Stoyan's web dev blog: phpied.com* [online]. 2009-12-04. [Zugriff am 2014-09-25]. Verfügbar unter: <http://www.phpied.com/psychology-of-performance/>
- [171] QUIRING, Oliver; SCHWEIGER, Wolfgang: *Interaktivität – ten years after*. In: *Medien & Kommunikationswissenschaft (M&K)*. 2006. Vol. 54, Nr. 2006/1, S. 5–24. ISSN 1615-634X.
- [172] PREIM, Bernhard; DACHSEL, Raimund: *Interaktive Systeme* [online]. Berlin, Heidelberg : Springer, 2010. [Zugriff am 2014-12-31]. eXamen.press. ISBN 9783642054013, 9783642054020. Verfügbar unter: <http://link.springer.com/10.1007/978-3-642-05402-0>
- [173] EGGER, Sebastian; REICHL, Peter; SCHOENENBERG, Katrin: *Quality of Experience and Interactivity*. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 149–161. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [174] YAU, Nathan: *Data Points: Visualization That Means Something*. Indianapolis, In: John Wiley & Sons, 2013. ISBN 9781118462195.
- [175] KOKKINIDIS, George: *Remnants of a Disappearing UI. Design Language News* [online]. 2010-11-18. [Zugriff am 2015-03-10]. Verfügbar unter: <http://news.designlanguage.com/post/1611663345>
- [176] SHENKER, S.: *Fundamental design issues for the future Internet*. In: *IEEE Journal on Selected Areas in Communications*. 1995-09. Vol. 13, no. 7, p. 1176–1188. DOI 10.1109/49.414637. ISSN 0733-8716.
- [177] *ITU-T Recommendation I.211 - B-ISDN service aspects* [online]. International Telecommunication Union (ITU), 1993. [Zugriff am 2013-09-02]. Recommendations - I Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-I.211-199303-I/en>
- [178] *ITU-T Recommendation G.1010 - End-user multimedia QoS categories* [online]. International Telecommunication Union (ITU), 2001. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.1010>
- [179] SOLDANI, David; CHIAVELLI, Davide; LAIHO, Jaana; LI, Man; MUHAMMAD, Noman; GIAMBIASI, Giovanni; RODRIQUEZ, Carolina: *QoS and QoS Monitoring*. In: SOLDANI, David; LI, Man; CUNY, Renaud (Hrsg.), *QoS and QoS Management in UMTS Cellular Systems*. John Wiley & Sons, Ltd, 2006. p. 315–384. ISBN 9780470034057.

- [180] HÄMMERLE, C.; KASCHWIG, T.; SIVRIKAYA, F.; YAZUK, Baris; RIDEL, Lenny: *Deliverable 7.1 - Assessment & Evaluation: Regional Assessment Framework for testbed experiments* [online]. ICT Project Public Deliverable. 2009. [Zugriff am 2013-08-23]. Perimeter Public Deliverables. Verfügbar unter: http://www.ict-perimeter.eu/component/option,com_jdownloads/Itemid,35/task,finish/cid,21/catid,13/lang,english/
- [181] TRAN, Hung Tuan; ZIEGLER, Thomas: Dimensioning Approaches for an Access Link Assuring Integrated QoS. In: *Proceedings of the Third International Conference on Quality of Service in Multiservice IP Networks*. Berlin, Heidelberg : Springer-Verlag, 2005. p. 221–234. QoS-IP'05. ISBN 354024557X, 9783540245575.
- [182] KASCHUB, Matthias: QoE – User Experience or User Expectation? *ITG 5.2.1/5.2.4 Fachgruppensitzung* [online]. Chemnitz. 2012-05-03. [Zugriff am 2012-06-02]. Verfügbar unter: http://www.tuhh.de/t3resources/itg521/FG_Chemnitz/ITG_Chemnitz_2012_US_Kaschub.pdf [Vortragsfolien]
- [183] FIEDLER, Markus: QoE Concepts and Standards. *Euro-NF Ph. D. Course on QoE* [online]. Karlskrona. 2011-06-20. [Zugriff am 2013-08-21]. Verfügbar unter: [http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L1_Concepts_pdf/\\$file/QoE_Course_L1_Concepts.pdf](http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L1_Concepts_pdf/$file/QoE_Course_L1_Concepts.pdf) [Vortragsfolien]
- [184] DICK, Matthias; WELLNITZ, Oliver; WOLF, Lars: Analysis of factors affecting players' performance and perception in multiplayer games. In: *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*. New York, NY, USA : ACM, 2005. p. 1–7. NetGames '05. ISBN 1595931562.
- [185] TSCHULIK, Peter: Quality of Service. In: HEIN, Mathias; REISNER, Michael; VOB, Antje (Hrsg.), *Voice over IP - Sprach-Daten-Konvergenz richtig nutzen*. Poing : Franzis, 2002. S. 179–211. ISBN 3772366864.
- [186] SHEA, Ryan; LIU, Jiangchuan; NGAI, EC-H.; CUI, Yong: Cloud gaming: architecture and performance. In: *Network, IEEE*. 2013. Vol. 27, no. 4. p. 16–19.
- [187] FETTWEIS, G.; ALAMOUTI, S.: 5G: Personal mobile internet beyond what cellular did to telephony. In: *IEEE Communications Magazine*. 2014-02. Vol. 52, no. 2, p. 140–145. DOI 10.1109/MCOM.2014.6736754. ISSN 0163-6804.
- [188] NÖLLE, Jochen.: *Voice over IP: Grundlagen, Protokolle, Migration*. Berlin; Offenbach : VDE-Verlag, 2003. ISBN 3800727080, 9783800727087.
- [189] TRICK, Ulrich; WEBER, Frank: *SIP, TCP-IP und Telekommunikationsnetze - Next-Generation Networks und VoIP - konkret*. München : Oldenbourg, 2009. ISBN 9783486590005.
- [190] VERHEIJ, Ingmar: *Quantify Perceived Performance: "Quantifying and rating the performance of (virtual) applications"* [online]. Whitepaper. PepperByte, 2011. [Zugriff am 2013-01-18]. Verfügbar unter: <http://www.ingmarverheij.com/wp-content/uploads/downloads/2011/09/Whitepaper-Quantifying-Perceived-Performance-v1.0.pdf>
- [191] MILLER, Robert B.: Response Time in Man-computer Conversational Transactions. In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*. New York, NY, USA : ACM, 1968. p. 267–277. AFIPS '68 (Fall, part I).
- [192] GRIGORIK, Ilya: Breaking the 1000ms Time to Glass Mobile Barrier. *The SF HTML5 User Group* [online]. San Francisco, CA. 2013-03-21. [Zugriff am 2014-09-21]. Verfügbar unter: <http://www.youtube.com/watch?v=Il4swGfTOSM> [Videomitschnitt des Vortrags]
- [193] SEVCIK, Peter: Understanding how users view application performance. In: *Business Communications Review*. 2002. Vol. 32, no. 7, p. 8–9. ISSN 0162-3885.
- [194] SEVCIK, Peter: Defining the application performance index. In: *Business Communications Review*. 2005. No. March 2005, p. 8–10. ISSN 0162-3885.

- [195] EGGER, Sebastian; HOSSFELD, Tobias; SCHATZ, Raimund; FIEDLER, Markus: Waiting times in quality of experience for web based services. In: *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*. 2012. p. 86–96.
- [196] NIELSEN, Jakob: Response Time Limits: The 3 Important Limits. *Nielsen Norman Group* [online]. 2005. [Zugriff am 2014-02-06]. Verfügbar unter: <http://www.nngroup.com/articles/response-times-3-important-limits/>
- [197] GRIGORIK, Ilya: Latency: The New Web Performance Bottleneck. *igvita.com* [online]. 2012-07-19. [Zugriff am 2014-10-07]. Verfügbar unter: <https://www.igvita.com/2012/07/19/latency-the-new-web-performance-bottleneck/>
- [198] CHESHIRE, Stuart: It's the Latency, Stupid. [online]. 1996-05. [Zugriff am 2014-09-09]. Verfügbar unter: <http://rescomp.stanford.edu/~cheshire/rants/Latency.html>
- [199] HOSSFELD, T.; KEIMEL, C.; HIRTH, M.; GARDLO, B.; HABIGT, J.; DIEPOLD, K.; TRAN-GIA, P.: Best Practices for QoE Crowdttesting: QoE Assessment With Crowdsourcing. In: *IEEE Transactions on Multimedia*. 2014-02. Vol. 16, no. 2, p. 541–558. DOI 10.1109/TMM.2013.2291663. ISSN 1520-9210.
- [200] HOSSFELD, Tobias; KEIMEL, Christian: Crowdsourcing in QoE Evaluation. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 315–327. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [201] ALRESHOODI, Mohammed; WOODS, John: Survey on QoE\QoS Correlation Models For Multimedia Services. In: *International Journal of Distributed and Parallel systems*. 2013-05-31. Vol. 4, no. 3, p. 53–72. DOI 10.5121/ijdps.2013.4305. ISSN 2229-3957, 0976-9757.
- [202] *ITU-T Recommendation P.800 - Methods for subjective Determination of Transmission Quality* [online]. International Telecommunication Union (ITU), 1996. [Zugriff am 2013-08-19]. Recommendations - P Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-P.800/en>
- [203] DE MAREZ, Lieven; DE MOOR, Katrien: The Challenge of User-And QoE-Centric Research and Product Development in Today's ICT-Environment. In: *Observatorio (OBS*)*. 2007. Vol. 1, no. 3. ISSN 1646-5954.
- [204] *ITU-R Recommendation BT.500 - Mean Opinion Score (MOS) terminology* [online]. International Telecommunication Union (ITU), 2012. [Zugriff am 2014-02-07]. BT Series, Broadcasting service (television). Verfügbar unter: <http://www.itu.int/rec/R-REC-BT.500/en>
- [205] *ITU-T Recommendation G.107 - The E-model, a computational model for use in transmission planning* [online]. International Telecommunication Union (ITU), 2011. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.107/en>
- [206] KATARZYNA WAC; ANIND K. DEY: Mobile User Experience Beyond the Laboratory: Towards a Methodology for QoE-QoS Evaluation in Natural User Environments (A Position Paper). *Journal of Mobile Media* [online]. 2012-Q3. [Zugriff am 2013-01-18]. Verfügbar unter: <http://wi.mobilities.ca/mobile-user-experience-beyond-the-laboratory-towards-a-methodology-for-qoe-qos-evaluation-in-natural-user-environments-a-position-paper/>
- [207] FAHRENBERG, J.; KLEIN, C.; PEPPER, M.; ZIMMERMANN, P.: *Versuchsplanung* [online]. Übungsbegleitendes Skriptum. Psychologisches Institut der Universität Freiburg, [kein Datum]. [Zugriff am 2013-05-19]. Verfügbar unter: http://www.jochen-fahrenberg.de/uploads/media/Methodenlehre_u._Versuchsplanung_01.pdf
- [208] FIEDLER, M.; HOSSFELD, T.; TRAN-GIA, P.: A generic quantitative relationship between quality of experience and quality of service. In: *IEEE Network*. 2010-03. Vol. 24, no. 2, p. 36–41. DOI 10.1109/MNET.2010.5430142. ISSN 0890-8044.

- [209] HOßFELD, Tobias; TRAN-GIA, Phuoc; FIEDLER, Markus: Quantification of Quality of Experience for Edge-based Applications. In: *Proceedings of the 20th International Teletraffic Conference on Managing Traffic Performance in Converged Networks*. Berlin, Heidelberg : Springer-Verlag, 2007. p. 361–373. ITC20'07. ISBN 9783540729891.
- [210] FIEDLER, Markus; HOßFELD, Tobias: Quality of Experience-related differential equations and provisioning-delivery hysteresis. In: *21st ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE, Miyazaki, Japan*. 2010.
- [211] REICHL, P.; EGGER, S.; SCHATZ, R.; D'ALCONZO, A.: The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment. In: *2010 IEEE International Conference on Communications (ICC)*. 2010-Mai. p. 1–5.
- [212] FIEDLER, Markus: Quantitative Relationships Between QoE and QoS. *Euro-NF Ph. D. Course on QoE* [online]. Karlskrona. 2011-06-21. [Zugriff am 2013-08-01]. Verfügbar unter: [http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L3_Quant_pdf/\\$file/QoE_Course_L3_Quant.pdf](http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L3_Quant_pdf/$file/QoE_Course_L3_Quant.pdf) [Vortragsfolien]
- [213] FIEDLER, Markus: Enabling Seamless Videoconferencing – The PERIMETER Approach. *ETSI Workshop on QoS / QoE / User experience focusing on speech / multimedia conference tools* [online]. Sophia Antipolis. 2010-09-22. [Zugriff am 2013-09-11]. Verfügbar unter: http://docbox.etsi.org/workshop/2010/201009_qosworkshop/report-etsi_qos_qoe_workshop.pdf [Vortragsfolien]
- [214] SCHATZ, Raimund; EGGER, Sebastian; MASUCH, Kathrin: The Impact of Test Duration on User Fatigue and Reliability of Subjective Quality Ratings. In: *Journal of the Audio Engineering Society*. 2012-03-20. Vol. 60, no. 1/2, p. 63–73.
- [215] SPILLNER, Andreas; LINZ, Tilo; ROßNER, Thomas; WINTER, Mario: *Praxiswissen Softwaretest - Testmanagement: Aus- und Weiterbildung zum Certified Tester - Advanced Level nach ISTQB-Standard*. Heidelberg, Neckar : dpunkt-Verl., 2008. ISBN 9783898645577.
- [216] BORTZ, Jürgen; DÖRING, Nicola: *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Berlin : Springer, 1995. ISBN 3540593756.
- [217] DE MOOR, Katrien: Non-technical QoE aspects. *Euro-NF Ph. D. Course on QoE* [online]. Karlskrona. 2011-06-22. [Zugriff am 2013-08-22]. Verfügbar unter: [http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L4_NontechnQoE_pdf/\\$file/QoE_Course_L4_NontechnQoE.pdf](http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L4_NontechnQoE_pdf/$file/QoE_Course_L4_NontechnQoE.pdf) [Vortragsfolien]
- [218] FAIRCLOUGH, Stephen H.: Fundamentals of Physiological Computing. In: *Interact. Comput.* 2009-01. Vol. 21, no. 1-2, p. 133–145. DOI 10.1016/j.intcom.2008.10.011. ISSN 0953-5438.
- [219] ARNDT, S.; ANTONS, J.; SCHLEICHER, R.; MÖLLER, S.; CURIO, G.: Perception of low-quality videos analyzed by means of electroencephalography. In: *2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*. 2012. p. 284–289.
- [220] RADWARE, LTD.: *Mobile Web Stress - The Impact of Network Speed on Emotional Engagement and Brand Perception* [online]. White Paper, Firmenschrift. Radware, Ltd., 2013. [Zugriff am 2014-06-06]. Verfügbar unter: <http://www.strangeloopnetworks.com/resources/research/our-need-for-web-speed/>
- [221] ANTONS, Jan-Niklas; ARNDT, Sebastian; SCHLEICHER, Robert; MÖLLER, Sebastian: Brain Activity Correlates of Quality of Experience. In: MÖLLER, Sebastian; RAAKE, Alexander (Hrsg.), *Quality of Experience*. Springer International Publishing, 2014. p. 109–119. T-Labs Series in Telecommunication Services. ISBN 9783319026800, 9783319026817.
- [222] ERTELT, Denis: *Experimentarten*. *Experimentalpsychologie.de* [online]. [Zugriff am 2014-04-10]. Verfügbar unter: <http://www.experimentalpsychologie.de/page38.html>
- [223] LESCUYER, Pierre; THIERRY, Lucidarme: *Evolved packet system (EPS) : the LTE and SAE evolution of 3G UMTS*. Chichester u.a. : Wiley, 2008. ISBN 0470059761.

- [224] OLSSON, Magnus; ROMMER, Stefan; MULLIGAN, Catherine; SULTANA, Shabnam; FRID, Lars: *SAE and the Evolved Packet Core: Driving the Mobile Broadband Revolution*. Academic Press, 2009. ISBN 9780123748263.
- [225] WANG, Xiaofei; CHEN, Min; TALEB, T.; KSENTINI, A.; LEUNG, V.: Cache in the air: exploiting content caching and delivery techniques for 5G systems. In: *IEEE Communications Magazine*. 2014-02. Vol. 52, no. 2, p. 131–139. DOI 10.1109/MCOM.2014.6736753. ISSN 0163-6804.
- [226] ROSENBROCK, Karl Heinz; HAUSER, Hans; HILLEBRAND, Friedhelm (Hrsg.): *The Creation of Standards for Global Mobile Communication* [online]. Bonn : Hillebrand Consulting Engineers GmbH, 2013. [Zugriff am 2014-05-23]. Verfügbar unter: <http://www.etsi.org/images/files/news/CreationOfStandardsForGlobalMobileCommunication.epub>
- [227] BERGMAN, Johan; GERSTENBERGER, Dirk; GUNNARSSON, Fredrik; STRÖM, Stefan: Continued HSPA Evolution of mobile broadband. In: *Ericsson Review*. 2009. No. 1, p. 7–11.
- [228] SAUTER, Martin: *Grundkurs mobile Kommunikationssysteme: Von UMTS und HSDPA, GSM und GPRS zu Wireless LAN und Bluetooth Piconetzen* [online]. 3. Auflage. Wiesbaden : Friedr. Vieweg & Sohn Verlag GWV Fachverlage GmbH, 2008. [Zugriff am 2014-02-19]. ISBN 9783834803979. Verfügbar unter: <http://dx.doi.org/10.1007/978-3-8348-9445-8>
- [229] SCHNABEL, Patrick: Datenübertragung im Mobilfunknetz. *Elektronik-Kompodium.de* [online]. [Zugriff am 2014-03-12]. Verfügbar unter: <https://www.elektronik-kompodium.de/sites/kom/0910141.htm>
- [230] SCHMALOHR, Martin; VOGL, Aylin; BARTSCH, Eric: B 209/2012: *IRT Mobil-Interop-Tests 2012* [online]. Technischer Bericht. München : Institut für Rundfunktechnik (IRT), 2012. [Zugriff am 2013-01-23]. Verfügbar unter: <http://www.irt.de/webarchiv/showdoc.php?z=NTkxOCMxMDA2MDEwI3BkZg==>
- [231] NGMN ALLIANCE: *User data rates in mobile data networks* [online]. NGMN Alliance, 2010. [Zugriff am 2013-01-24]. Verfügbar unter: http://www.ngmn.org/uploads/media/Explanatory_paper_on_user_data_rates_v1_0_01.pdf
- [232] PETERSEN, Swen: Betrachtung der Netto und Bruttodatenrate bei der LTE-Funktechnologie. In: *FKT*. Nr. 10/2011, S. 540–551. ISSN 1430-9947.
- [233] LIPPERT, Hermann; PETERSEN, Swen: Mehr Netto vom Brutto. In: INSTITUT FÜR RUNDFUNKTECHNIK (Hrsg.), *Jahresbericht 2009*. München : Institut für Rundfunktechnik GmbH (IRT), 2010. S. 12–13. ISSN 1617- 0822.
- [234] LUKAS, Kai; MARX, Almuth; SCHÖTTLER, Bernd Oliver; SUDHUES, Christoph: *Dienstqualität von Breitbandzugängen* [online]. Studie im Auftrag der Bundesnetzagentur, Abschlussbericht. Ismaning : zafaco GmbH, 2013. [Zugriff am 2013-09-02]. Verfügbar unter: http://www.initiative-netzqualitaet.de/fileadmin/user_upload/Abschlussbericht_BNetzA_Studie_Dienstqualitaet_Breitbandzugaenge.pdf [siehe auch URL <http://www.initiative-netzqualitaet.de/abschlussbericht/> Zugriff am 2013-09-02]
- [235] PAULER, Wolfgang: Der härteste Handy-Netztest aller Zeiten - CHIP Handy Welt. *CHIP Online* [online]. 2012-08-28. [Zugriff am 2013-01-14]. Verfügbar unter: http://www.chip.de/artikel/Der-haerteste-Handy-Netztest-aller-Zeiten_57281029.html
- [236] PAULER, Wolfgang; HEINFLING, Benjamin: Das beste Netz aller Zeiten. *CHIP Online* [online]. 2013-09-02. [Zugriff am 2013-09-06]. Verfügbar unter: http://www.chip.de/artikel/Der-haerteste-Handy-Netztest-Deutschlands-Telekom-Vodafone-O2-und-E-Plus-im-Test_63944005.html
- [237] JUST, Christian: Der grosse COMPUTERBILD Netztest. In: *Computer Bild*. 2012-11-17. Nr. 25/2012, S. 82–90. ISSN 1437-6482.

- [238] JUST, Christian; LINK, Michael: Der grosse Mobilfunk-Test 2013/14. In: *Computer Bild*. 2013-10-19. Nr. 23/2013, S. 68–76. ISSN 1437-6482.
- [239] THEISS, Bernd: Der Netztest 2012 - Mobilfunk in Deutschland. In: *connect*. 2012-11. Nr. 12/2012, S. 16–32. ISSN 0944-6141.
- [240] THEISS, Bernd: Der Netztest 2013/2014 - Mobilfunk in Deutschland. In: *connect*. 2013-12-05. Nr. 01/2014, S. 26–43. ISSN 0944-6141.
- [241] CASAIS, Eduardo: Tempo zu Umsatz - Netz-Kennzahlen für Content Provider. In: *ix*. Nr. 2/2014, S. 95–99. ISSN 0935-9680.
- [242] WEHRLE, Klaus; GÜNES, Mesut; GROSS, James: *Modeling and Tools for Network Simulation*. New York : Springer, 2010. ISBN 9783642123306.
- [243] MANFRED HANS SADELER: Fortschrittliche Mobilfunk-Tests. In: *funkschau*. 2010-10-08. Nr. 19/2010, S. 34. ISSN 0016-2841.
- [244] CARBONE, Marta; RIZZO, Luigi: Dummynet revisited. In: *SIGCOMM Comput. Commun. Rev.* 2010-04. Vol. 40, no. 2, p. 12–20. DOI 10.1145/1764873.1764876. ISSN 0146-4833.
- [245] VELÁSQUEZ, Karima; GAMESS, Eric: A comparative analysis of WAN emulators. In: *Proceedings of the 7th Latin American Networking Conference*. New York, NY, USA : ACM, 2012. p. 44–51. LANC '12. ISBN 9781450317504.
- [246] SHAIKH, J.; MINHAS, T.N.; ARLOS, P.; FIEDLER, M.: Evaluation of delay performance of traffic shapers. In: *2010 2nd International Workshop on Security and Communication Networks (IWSCN)*. 2010. p. 1–8.
- [247] MINHAS, T.N.; FIEDLER, M.; SHAIKH, J.; ARLOS, P.: Evaluation of throughput performance of traffic shapers. In: *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*. 2011. p. 1596–1600.
- [248] MINHAS, Tahir; SHAIKH, Junaid; FIEDLER, Markus; ARLOS, Patrik: Evaluation of Traffic Shapers. *Euro-NF Ph. D. Course on QoE* [online]. Karlskrona. 2011-06-23. [Zugriff am 2013-08-21]. Verfügbar unter: [http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L8_Shaper_pdf/\\$file/QoE_Course_L8_Shaper.pdf](http://www.bth.se/com/ccs.nsf/attachments/QoE_Course_L8_Shaper_pdf/$file/QoE_Course_L8_Shaper.pdf) [Vortragsfolien]
- [249] BECKE, Martin; DREIBHOLZ, Thomas; RATHGEB, Erwin P.; FORMANN, Johannes: Link Emulation on the Data Link Layer in a Linux-based Future Internet Testbed Environment. In: *Proceedings of the The Tenth International Conference on Networks*. St. Maarten, 2011. ISBN 9781612080024.
- [250] JURGELIONIS, A.; LAULAJAINEN, J.-P.; HIRVONEN, M.; WANG, A.I.: An Empirical Study of NetEm Network Emulation Functionalities. In: *2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*. 2011-08-31. p. 1–6.
- [251] GARCIA, Johan; CONCHON, Emmanuel; PÉRENNOU, Tanguy; BRUNSTROM, Anna: KauNet: improving reproducibility for wireless and mobile research. In: *Proceedings of the 1st international workshop on System evaluation for mobile platforms*. 2007. p. 21–26.
- [252] FERGUSON, Paul: *Quality of service : delivering QoS on the Internet and in corporate networks*. New York, Weinheim [u.a.] : Wiley, 1998. ISBN 0471243582.
- [253] BROWN, Martin A.: Traffic Control HOWTO. *Guide to IP Layer Network Administration with Linux* [online]. 2006-10. [Zugriff am 2014-08-15]. Verfügbar unter: <http://linux-ip.net/articles/Traffic-Control-HOWTO/>
- [254] J. PATTLACH: IP-Plattform. *Verein zur Förderung eines Deutschen Forschungsnetzes e. V. (Deutsche Forschungsnetz, DFN)* [online]. 2012-04-16. [Zugriff am 2012-11-26]. Verfügbar unter: <http://www.dfn.de/xwin/ipplattform/>
- [255] HEMMINGER, Stephen: Network Emulation with NetEm. In: *linux.conf.au*. Canberra, Australia, 2005-04. [Zugriff am 2015-02-15]. verfügbar unter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.1687&rep=rep1&type=pdf>

- [256] THE LINUX FOUNDATION: netem. *Linux Foundation* [online]. 2009-11-19. [Zugriff am 2013-01-08]. Verfügbar unter:
<http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>
- [257] IVANCIC, D.; HADJINA, N.; BASCH, D.: Analysis of precision of the HTB packet scheduler. In: *18th International Conference on Applied Electromagnetics and Communications, 2005. ICECom 2005*. 2005-10. p. 1–4.
- [258] KELLER, Ariane: *Trace Control for Netem* [online]. Semesterarbeit. Zürich : Eidgenössische Technische Hochschule Zürich, Institut für Technische Informatik und Kommunikationsnetze, 2006. [Zugriff am 2013-01-11]. Verfügbar unter:
<http://tcn.hypert.net/tcn.pdf>
- [259] PÉRENNOU, Tanguy; BRUNSTROM, Anna; HALL, Tomas; GARCIA, Johan; HURTIG, Per: Emulating Opportunistic Networks with KauNet Triggers. In: *EURASIP Journal on Wireless Communications and Networking*. 2011. Vol. 2011, p. 1–14. DOI 10.1155/2011/347107. ISSN 1687-1472, 1687-1499.
- [260] BARCZOK, Achim; WÖLBERT, Christian: Der Update-Frust bleibt. In: *c't*. 2013-04-08. Nr. 9/2013, S. 70–73. ISSN 0724-8679.
- [261] OPENSIGNAL, INC.: *Android Fragmentation Report 2013* [online]. Firmenschrift. OpenSignal, Inc., 2013. [Zugriff am 2013-11-16]. Verfügbar unter:
<http://opensignal.com/reports/fragmentation-2013/fragmentation-2013.pdf>
- [262] GARTNER; IDC: *Smartphone - Marktanteile der Hersteller am Absatz weltweit bis 2013* [online]. zitiert nach de.statista.com. 2014. [Zugriff am 2014-08-12]. Verfügbar unter:
<http://de.statista.com/statistik/daten/studie/173051/umfrage/weltweite-marktanteile-der-fuehrenden-smartphone-hersteller-seit-2009/>
- [263] JACOBSON, V.; BRADEN, R. T.: *RFC 1072 - TCP extensions for long-delay paths* [online]. Request for Comments. Internet Engineering Task Force (IETF), 1988. [Zugriff am 2015-02-26]. Verfügbar unter: <http://tools.ietf.org/html/rfc1072>
- [264] EIKENBERG, Ronald: Gut App-geschaut - Netzwerkverkehr von Smartphones kontrollieren. In: *c't*. Nr. 7/2013, S. 120–124. ISSN 0724-8679.
- [265] APPLE INC.: Technical Q&A QA1176: Getting a Packet Trace. *Mac Developer Library* [online]. 2013-01-31. [Zugriff am 2013-04-12]. Verfügbar unter:
https://developer.apple.com/library/mac/qa/qa1176/_index.html
- [266] BROWN, Barry; MCGREGOR, Moira; LAURIER, Eric: iPhone in vivo: video analysis of mobile device use. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013. p. 1031–1040.
- [267] CZERULLA, Hannes A.; HANSEN, Sven: Miracast-Check - Was die AirPlay-Alternative derzeit leistet. In: *c't*. Nr. 17/2013, S. 148–149. ISSN 0724-8679.
- [268] VINCENT, Tony: 7 Ways to Show Your iPad on a Projector Screen. *Learning in Hand* [online]. 2013-11-21. [Zugriff am 2013-12-15]. Verfügbar unter:
<http://learninginhand.com/blog/5-ways-to-show-your-ipad-on-a-projector-screen.html>
- [269] RADKE, Jeremias: Freiluftbühne - AirPlay mit Android und Windows. In: *c't*. Nr. 8/2013, S. 144–146. ISSN 0724-8679.
- [270] ETSI TR 102 505 V1.3.1: *Speech and multimedia Transmission Quality (STQ); Development of a Reference Web page* [online]. Technical Report. European Telecommunications Standards Institute (ETSI), 2012. [Zugriff am 2014-08-20]. Verfügbar unter:
http://www.etsi.org/deliver/etsi_tr/102500_102599/102505/01.03.01_60/tr_102505v010301p.pdf

- [271] EGGER, Sebastian: Interactive Content for Subjective Studies on Web Browsing QoE: A Kepler derivative. *ETSI Workshop on Selected Items on Telecommunication Quality Matters* [online]. Wien. 2012-11-28. [Zugriff am 2013-01-13]. Verfügbar unter: http://docbox.etsi.org/Workshop/2012/201211_STQWORKSHOP/S8_QoEandFutureConsi deration/FTW_EGGER.pdf [Vortragsfolien]
- [272] E-PLUS MOBILFUNK GMBH & Co. KG: *Mobilfunkmonitor 2012* [online]. Firmenschrift. E-Plus Mobilfunk GmbH & Co. KG, 2012. [Zugriff am 2013-01-22]. Verfügbar unter: http://eplus-gruppe.de/wp-content/uploads/2012/09/NEU_RZ_E-Plus_Broschuere_010912-2.pdf
- [273] COMSCORE, INC.: *2013 Future in Focus - Digitales Deutschland* [online]. Firmenschrift. comScore, Inc., 2013. [Zugriff am 2013-04-02]. Verfügbar unter: http://www.comscore.com/Insights/Presentations_and_Whitepapers/2013/2013_Future_in_Focus_Digitales_Deutschland
- [274] ARBEITSGEMEINSCHAFT ONLINE FORSCHUNG: 2012-II: *mobile facts 2012-II* [online]. Frankfurt : Arbeitsgemeinschaft Online Forschung (AGOF), 2013. [Zugriff am 2013-12-20]. mobile facts. Verfügbar unter: http://www.agof.de/download/Downloads_Mobile_Facts/Downloads_Mobile_Facts_2012/Downloads_Mobile_Facts_2012_II/mf_2012-II_AGOF_mobile_facts_2012-II.pdf
- [275] SANDVINE INCORPORATED ULC: *Sandvine Global Internet Phenomena Snapshot - 2H 2012, Europe, Mobile Access* [online]. Firmenschrift. Sandvine Incorporated ULC, 2012. [Zugriff am 2013-03-27]. Sandvine Global Internet Phenomena Report. Verfügbar unter: <http://www.sandvine.com/general/document.download.asp?docID=71&sourceID=0>
- [276] HJELMVIK, Erik: *The SPID Algorithm - Statistical Protocol IDentification* [online]. Gävle, Sweden : Internet Infrastructure Foundation, 2008. [Zugriff am 2013-03-31]. Verfügbar unter: https://www.iis.se/docs/The_SPID_Algorithm_-_Statistical_Protocol_IDentification.pdf
- [277] ADAMSKY, Florian: *Statistische Analyse zur Identifizierung und Klassifizierung von Protokollen in IP-Strömen*. Bachelorthesis. Friedberg : Fachhochschule Gießen-Friedberg, Fachbereich Informationstechnik, Elektrotechnik und Mechatronik, 2010.
- [278] CISCO SYSTEMS, INC.: Cisco Data Meter. *Cisco VNI* [online]. 2013-03-30. [Zugriff am 2014-06-12]. Verfügbar unter: <http://ciscovni.com/data-meter/index.html>
- [279] ERICSSON: June 2013: *Ericsson Mobility Report 2013* [online]. Firmenschrift. Stockholm : Ericsson, 2013. [Zugriff am 2013-10-20]. Ericsson Mobility Report. Verfügbar unter: <http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-june-2014.pdf>
- [280] ALLOT COMMUNICATIONS: *Allot MobileTrends Report Feb 2013* [online]. Firmenschrift. Allot Communications, 2013. [Zugriff am 2013-03-27]. Allot MobileTrends. Verfügbar unter: <http://www.allot.com/allotmobiletrends-feb-2013.html>
- [281] CISCO SYSTEMS, INC.: Cisco Visual Networking Index (VNI): Global Mobile Data Traffic Forecast, 2012–2017 (Q&A). *Cisco.com* [online]. 2013-02-06. [Zugriff am 2013-03-29]. Verfügbar unter: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI-Forecast_QA.html
- [282] SHARMA, Chetan: 2013 Mobile Industry Predictions Survey. *Chetan Sharma Consulting* [online]. 2013-01-02. [Zugriff am 2013-03-27]. Verfügbar unter: <http://www.chetansharma.com/blog/2013/01/02/2013-mobile-industry-predictions-survey/>
- [283] KITTNER, Ingeborg: Grundbegriffe. In: GUTTMANN, Giselher (Hrsg.), *Allgemeine Psychologie: experimentalpsychologische Grundlagen*. 2. Auflage. Wien : WUV-Universitätsverlag, 1994. S. 1–20. ISBN 9783851140743.

- [284] MAYO, Elton: *The Human Problems of an Industrial Civilization*. New York, NY, USA : Macmillan, 1933.
- [285] AL-QEISI, Kholoud Ibrahim: *Analyzing the use of UTAUT model in explaining an online behaviour: Internet banking adoption*. Dissertation. Brunel University, Brunel Business School, 2009.
- [286] STATISTA: *Smartphones - Statista-Dossier 2013* [online]. Firmenschrift. Statista, 2013. [Zugriff am 2013-10-28]. Verfügbar unter: <http://de.statista.com/statistik/studie/id/3179/dokument/smartphones-statista-dossier/>
- [287] STATISTISCHES BUNDESAMT: *Altersstruktur der Bevölkerung in Deutschland zum 31. Dezember 2011* [online]. zitiert nach de.statista.com. Statistisches Bundesamt, 2012. [Zugriff am 2013-10-28]. Verfügbar unter: <http://de.statista.com/statistik/daten/studie/1351/tab/4/umfrage/altersstruktur-der-bevoelkerung-deutschlands/>
- [288] PRENSKY, Marc: Digital natives, digital immigrants. In: *On the Horizon; MCB University Press*. 2001. Vol. 9, no. 5, October 2001. ISSN 1074-8121.
- [289] ARMUTAT, Sascha: *Zwischen Anspruch und Wirklichkeit: Generation Y finden, fördern und binden* [online]. Düsseldorf : Deutsche Gesellschaft für Personalführung e.V., 2011. [Zugriff am 2013-10-28]. Verfügbar unter: <http://static.dgfp.de/assets/publikationen/2011/GenerationY-finden-foerdern-binden.pdf> [ISSN 1613-2785]
- [290] UNITED NATIONS JOINT STAFF PENSION FUND: *Traditionalists, Baby Boomers, Generation X, Generation Y (and Generation Z) Working Together* [online]. New York : Talent Management Team - Executive Office - United Nations Joint Staff Pension Fund, 2009. [Zugriff am 2013-10-28]. Verfügbar unter: <http://www.un.org/staffdevelopment/pdf/Designing%20Recruitment,%20Selection%20&%20Talent%20Management%20Model%20tailored%20to%20meet%20UNJSPF%27s%20Business%20Development%20Needs.pdf>
- [291] SEEL, Roger: Generationen im Web, Traditionalist, Boomer, Generation X, Generation Y. *myknow Blog* [online]. [Zugriff am 2013-10-28]. Verfügbar unter: <http://blog.myknow.com/generationen-im-web-traditionalist-boomer-generation-x-generation-y/>
- [292] STATISTISCHES BUNDESAMT: *Bildungsstand der Bevölkerung 2013* [online]. Wiesbaden : Statistisches Bundesamt, 2013. [Zugriff am 2013-10-18]. Verfügbar unter: https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Bildungsstand/BildungsstandBevoelkerung5210002137004.pdf?__blob=publicationFile
- [293] STATISTISCHES BUNDESAMT: Bildungsstand - Bevölkerung nach Bildungsabschluss in Deutschland. *Staat & Gesellschaft - Bildungsstand - Bildungsstand - Statistisches Bundesamt (Destatis)* [online]. 2013. [Zugriff am 2013-10-18]. Verfügbar unter: <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Bildungsstand/Tabellen/Bildungsabschluss.html>
- [294] ACTIX INTERNATIONAL LIMITED: University students provide insights into mobile's future. *Actix Press Releases* [online]. 2013-09-10. [Zugriff am 2013-09-11]. Verfügbar unter: <http://www.actix.com/knowledge-centre/news/item/108-university-students-provide-insights-into-mobile-s-future.html> [Pressemitteilung]
- [295] BITKOM: *63 Millionen Handy-Besitzer in Deutschland* [online]. Berlin : Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. (BITKOM), 2013. [Zugriff am 2013-10-18]. Verfügbar unter: http://www.bitkom.org/files/documents/PI_BITKOM_Handy-Verbreitung_in_Deutschland_26_08_2013.pdf

- [296] BUNDESNETZAGENTUR: Teilnehmerentwicklung im Mobilfunk. *Bundesnetzagentur - Mobilfunkteilnehmer* [online]. 2013-04-06. [Zugriff am 2013-10-28]. Verfügbar unter: http://www.bundesnetzagentur.de/cln_1911/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Marktbeobachtung/Deutschland/Mobilfunkteilnehmer/Mobilfunkteilnehmer_node.html
- [297] AXEL SPRINGER; BAUER MEDIA GROUP: *VerbraucherAnalyse 2012* [online]. zitiert nach de.statista.com. Axel Springer, Bauer Media Group, 2012. [Zugriff am 2013-10-28]. Verfügbar unter: <http://de.statista.com/statistik/daten/studie/154159/tab/4/umfrage/verteilung-der-mobilfunkkunden-nach-mobilfunkanbietern-in-deutschland/>
- [298] PINNEKAMP, Heinz-Jürgen; SIEGMANN, Frank: *Deskriptive Statistik*. München : Oldenbourg, 2001. ISBN 9783486256543.
- [299] HOßFELD, T.; SCHATZ, R.; EGGER, S.: SOS: The MOS is not enough! In: *2011 Third International Workshop on Quality of Multimedia Experience (QoMEX)*. 2011. p. 131–136.
- [300] KNOLL, Thomas Martin: A Network-based Method for Video QoE Measurement. *ITG 5.2.1/5.2.4 Fachgruppensitzung* [online]. Chemnitz. 2012-05-04. [Zugriff am 2013-01-10]. Verfügbar unter: http://www.tuhh.de/t3resources/itg521/FG_Chemnitz/ITG_Chemnitz_2012_TUC_Knoll.pdf
- [301] ECKERT, Marcus; KNOLL, Thomas M.: An advanced network based method for Video QoE estimation based on throughput measurement. In: *Proc. EuroView*. Würzburg, 2012-07-24.
- [302] SCHWARZE, Jochen: *Grundlagen der Statistik I*. 9. Auflage. Herne; Berlin : Verl. Neue Wirtschafts-Briefe, 2001. ISBN 9783482564390.
- [303] RIVADENEIRA, A. W.; GRUEN, Daniel M.; MULLER, Michael J.; MILLEN, David R.: Getting our head in the clouds: toward evaluation studies of tagclouds. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California, USA : ACM Press, 2007. p. 995–998. CHI '07. ISBN 9781595935939.
- [304] HARRIS, Jacob: Word clouds considered harmful. *NiemanLab* [online]. 2011-10-13. [Zugriff am 2013-10-09]. Verfügbar unter: <http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/>
- [305] GWET, Kilem L.: Intrarater reliability. In: *Wiley encyclopedia of clinical trials*. Hoboken, N.J. : Wiley, 2008. ISBN 9780471462422.
- [306] JAMIESON, Susan: Likert scales: how to (ab)use them. In: *Medical Education*. 2004-12. Vol. 38, no. 12, p. 1217–1218. DOI 10.1111/j.1365-2929.2004.02012.x. ISSN 0308-0110, 1365-2923.
- [307] MÜLLER, Simon: Können Likert-skalierte Daten als stetige Daten interpretiert werden? *Statistik Stuttgart* [online]. 2013-04-08. [Zugriff am 2014-01-21]. Verfügbar unter: <http://statistik-stuttgart.de/konnen-likert-skalierte-daten-als-stetige-daten-interpretiert-werden/>
- [308] NIELSEN, Jakob; LEVY, Jonathan: Measuring Usability: Preference vs. Performance. In: *Commun. ACM*. 1994-04. Vol. 37, no. 4, p. 66–75. DOI 10.1145/175276.175282. ISSN 0001-0782.
- [309] KUSS, Alfred: *Marktforschung: Grundlagen der Datenerhebung und Datenanalyse*. Springer DE, 2007. ISBN 9783834905918.
- [310] HOYLE, Rick H; HARRIS, Monica J; JUDD, Charles M: *Research methods in social relations*. Fort Worth, TX : Wadsworth, 2002. ISBN 9780155061392.

- [311] LANGER, Wolfgang: *Einführung in sozialwissenschaftliche Skalen-, Index- und Typenkonstruktion* [online]. Begleitmaterial zur Lehrveranstaltung. Halle, 2000. [Zugriff am 2013-11-05]. Verfügbar unter: <http://www.sozioogie.uni-halle.de/langer/pdf/meth1/skalieren2.pdf>
- [312] HOLM, Kurt: Die Gültigkeit sozialwissenschaftlichen Messens. In: HOLM, Kurt (Hrsg.), *Die Befragung 4 – Skalierungsverfahren, Panelanalyse*. München : Francke, 1976. S. 123–133. ISBN 3772011306.
- [313] NORMAN, Geoff: Likert scales, levels of measurement and the “laws” of statistics. In: *Advances in Health Sciences Education*. 2010-02-10. Vol. 15, no. 5, p. 625–632. DOI 10.1007/s10459-010-9222-y. ISSN 1382-4996, 1573-1677.
- [314] CLASON, Dennis L.; DORMODY, Thomas J.: Analyzing data measured by individual Likert-type items. In: *Journal of Agricultural Education*. 1994. Vol. 35, p. 4.
- [315] ZIEGLER, Bernd: *Grundlagen der statistischen Methodenlehre*. Gernsbach : Dt. Betriebswirte-Verl., 1993. ISBN 3886400581, 9783886400584.
- [316] PREL, Jean-Baptist du; HOMMEL, Gerhard; RÖHRIG, Bernd; BLETTNER, Maria: Konfidenzintervall oder p-Wert? Teil 4 der Serie zur Bewertung wissenschaftlicher Publikationen. In: *Dtsch Arztebl International*. 2009. Jg. 106, Heft. 19, S. 335–339. DOI 10.3238/arztebl.2009.0335.
- [317] STORM, Regina: *Wahrscheinlichkeitsrechnung, mathematische Statistik und statistische Qualitätskontrolle*. München : Fachbuchverl. Leipzig im Carl-Hanser-Verl., 2007. ISBN 9783446409064.
- [318] WEBER, Erna: *Grundriss der biologischen Statistik : Anwendungen der mathematischen Statistik in Forschung, Lehre und Praxis*. 9. durchges. Auflage. Jena : Fischer, 1986. ISBN 3334000311.
- [319] SAURO, Jeff: How Confident Do You Need to be in Your Research? *Measuring Usability* [online]. 2015-01-05. [Zugriff am 2015-03-15]. Verfügbar unter: <https://www.measuringu.com/blog/confidence-levels.php>
- [320] CUMMING, Geoff; FINCH, Sue: Inference by Eye: Confidence Intervals and How to Read Pictures of Data. In: *American Psychologist*. 2005. Vol. 60, no. 2, p. 170–180. DOI 10.1037/0003-066X.60.2.170. ISSN 1935-990X, 0003-066X.
- [321] *ITU-R Recommendation BS.1534 - Method for the subjective assessment of intermediate quality levels of coding systems* [online]. International Telecommunication Union (ITU), 2003. [Zugriff am 2014-02-07]. Recommendations - R Series. Verfügbar unter: <http://www.itu.int/rec/R-REC-BS.1534/en>
- [322] KUNIAVSKY, Mike: *Observing the User Experience*. San Francisco, CA : Morgan Kaufmann Publishers, 2003. Morgan Kaufmann series in interactive technologies. ISBN 1558609237.
- [323] SAURO, Jeff: How to interpret survey responses: 5 techniques. *Measuring Usability* [online]. 2011-05-10. [Zugriff am 2013-09-30]. Verfügbar unter: <http://www.measuringusability.com/blog/interpret-responses.php>
- [324] SAMBANDAM, Rajan; HAUSSER, George: An alternative method of reporting customer satisfaction scores. In: *Quirk's Marketing Research Review*. 1998-10. Vol. Volume XII, no. Number 9, p. 22–23 + 73–74. ISSN 0893-7451.
- [325] SAURO, Jeff: Are Net Promoter Scores Normally Distributed? *Measuring Usability* [online]. 2011-01-26. [Zugriff am 2013-09-30]. Verfügbar unter: <http://www.measuringusability.com/blog/normal-nps.php>
- [326] NORUSIS, M. J.: *IBM SPSS Statistics 19 Advanced Statistical Procedures Companion*. Upper Saddle River, NJ : Prentice Hall, 2012. ISBN 9780321748430.
- [327] SAURO, Jeff; LEWIS, James R.: Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations. In: *Proceedings of the human factors and ergonomics society annual meeting*. 2005. p. 2100–2103.

- [328] KOSCHACK, J.: Standardabweichung und Standardfehler: der kleine, aber feine Unterschied. In: *ZFA - Zeitschrift für Allgemeinmedizin*. 2008-06. Vol. 84, Nr. 6, S. 258–260. DOI 10.1055/s-2008-1073146. ISSN 1433-6251, 1439-9229.
- [329] DEVAULT, Gigi: Scoring and Reporting in Surveys Research. *About.com Market Research* [online]. [Zugriff am 2013-10-29]. Verfügbar unter: <http://marketresearch.about.com/od/market.research.surveys/a/Scoring-And-Reporting-In-Surveys-Research.htm>
- [330] SAURO, Jeff: Top-Box Scoring of Rating Scale Data. *Measuring Usability* [online]. 2010-12-14. [Zugriff am 2013-10-30]. Verfügbar unter: <http://www.measuringusability.com/blog/top-box.php>
- [331] GREWE, Kirsten; STADTLER, Marc: *Eine Anleitung zu Formatfragen bei der Erstellung von Praktikumsberichten, Diplomarbeiten und Dissertationen basierend auf dem: Manual zur Manuskripterstellung nach den Richtlinien der American Psychological Association (APA)* [online]. Münster : Universität Münster, 2001. [Zugriff am 2013-11-07]. Verfügbar unter: http://wwwpsy.uni-muenster.de/imperia/md/content/psychologie_institut_3/ae_bromme/pdf/service/manual_zur_manuskripterstellung.pdf
- [332] AMERICAN PSYCHOLOGICAL ASSOCIATION: *Publication manual of the American Psychological Association*. 6th ed. Washington, DC : American Psychological Association, 2010. ISBN 9781433805592.
- [333] ROBBINS, Naomi B.; HEIBERGER, Richard M.: Plotting Likert and other rating scales. In: *Proceedings of the 2011 Joint Statistical Meeting*. 2011.
- [334] *ITU-T Recommendation G.109 - Definition of categories of speech transmission quality* [online]. International Telecommunication Union (ITU), 1999. [Zugriff am 2013-09-02]. Recommendations - G Series. Verfügbar unter: <http://www.itu.int/rec/T-REC-G.109-199909-l/en>
- [335] Gossensche Gesetze. *Gabler Wirtschaftslexikon* [online]. Springer Gabler Verlag, [kein Datum]. [Zugriff am 2015-03-17]. Verfügbar unter: <http://wirtschaftslexikon.gabler.de/Archiv/53994/gossensche-gesetze-v5.html>
- [336] RÖHRIG, Bernd; DU PREL, Jean-Baptist; BLETTNER, Maria: Studiendesign in der medizinischen Forschung. In: *Dt Ärztbl*. 2009. Jg. 106, Heft 11, S. 184–189. ISSN 0012-1207.
- [337] WU, C.-C.; CHEN, K.-T.; CHANG, Y.-C.; LEI, C.-L.: Crowdsourcing Multimedia QoE Evaluation: A Trusted Framework. In: *IEEE Transactions on Multimedia*. 2013. Vol. 15, no. 5, p. 1121–1137. DOI 10.1109/TMM.2013.2241043. ISSN 1520-9210.
- [338] SACHS, Lothar: *Statistische Methoden: Planung und Auswertung 2*. Berlin : Springer, 1990. ISBN 0387520252, 3540520252.
- [339] HUANG, Dijiang; XING, Tianyi; WU, Huijun: Mobile Cloud Computing Service Models: A User-Centric Approach. In: *IEEE Network*. 2013. Vol. 27, no. 5, p. 6–11.
- [340] MEEKER, Mary; WU, Liang: *Internet Trends 2013* [online]. Firmenschrift. KPCB - Kleiner Perkins Caufield Byers, 2013. [Zugriff am 2013-09-11]. Verfügbar unter: <http://www.kpcb.com/insights/2013-internet-trends>
- [341] KHALAF, Simon: The Age of Living Mobile. *Source14* [online]. San Francisco, CA. 2014-04-22. [Zugriff am 2014-05-18]. Verfügbar unter: <http://de.slideshare.net/FlurryMobile/source14-the-age-of> [Vortragsfolien]
- [342] PURGATHOFER, Peter: The many perspectives of Quality of Experience. *QoMEX 2013* [online]. Klagenfurt. 2013-07-05. [Zugriff am 2013-08-21]. Verfügbar unter: http://www.youtube.com/watch?v=57yxr_a6Vdc [Vortrag mit Videomitschnitt]
- [343] LÖWGREN, Jonas: Pliability as an experiential quality: Exploring the aesthetics of interaction design. In: *Artifact*. 2007. Vol. 1, no. 2, p. 85–95. DOI 10.1080/17493460600976165. ISSN 1749-3463.

- [344] CANNEY, Mike: Application Performance Analysis. *Sharkfest '13* [online]. Berkeley, CA. 2013-06-16. [Zugriff am 2013-09-30]. Verfügbar unter: http://sharkfest.wireshark.org/sharkfest.13/presentations/NAP-01_Application-Performance-Analysis_Mike-Canney.pdf [Vortragsfolien]
- [345] MIRZADEH, Manuela: BASE hat den Trend frühzeitig erkannt: Der Boom der Musikstreamingdienste wächst. *E-Plus Gruppe* [online]. 2014-01-14. [Zugriff am 2014-03-19]. Verfügbar unter: <http://eplus-gruppe.de/musikstreamingdienste-boomen/>
- [346] JOHNSON, T.; SEELING, P.: Desktop and mobile web page comparison: characteristics, trends, and implications. In: *IEEE Communications Magazine*. 2014-09. Vol. 52, no. 9, p. 144–151. DOI 10.1109/MCOM.2014.6894465. ISSN 0163-6804.
- [347] STROHMEIER, D.; JUMISKO-PYVKKO, S.; RAAKE, A.: Toward task-dependent evaluation of web-QoE: Free exploration vs. „Who Ate What?“. In: *2012 IEEE Globecom Workshops (GC Wkshps)*. 2012. p. 1309–1313.
- [348] GOOGLE: PageSpeed Insights Rules. *PageSpeed Insights Rules - PageSpeed Insights — Google Developers* [online]. 2014. [Zugriff am 2014-04-21]. Verfügbar unter: <https://developers.google.com/speed/docs/insights/rules>
- [349] What are the Network Tests? *Browserscope* [online]. [Zugriff am 2015-02-24]. Verfügbar unter: <http://www.browserscope.org/network/about>
- [350] CALHOUN, David: Mobile Performance Manifesto. *David Calhoun's Blog* [online]. 2011-10-11. [Zugriff am 2013-04-15]. Verfügbar unter: <http://davidbcalhoun.com/2011/mobile-performance-manifesto>
- [351] SOUDERS, Steve: Silk, iPad, Galaxy comparison. *Steve Souders - High Performance Web Sites* [online]. 2011-12-01. [Zugriff am 2013-04-15]. Verfügbar unter: <http://www.stevesouders.com/blog/2011/12/01/silk-ipad-galaxy-comparison/>
- [352] CISCO SYSTEMS, INC.: *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018* [online]. White Paper, Firmenschrift. Cisco Systems, Inc., 2014. [Zugriff am 2014-03-01]. Visual Networking Index (VNI). Verfügbar unter: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf
- [353] SIGLIN, Tim: Content Delivery Summit: Google Talks YouTube Views, Caching. *Streaming Media Magazine* [online]. 2014-05-12. [Zugriff am 2014-05-12]. Verfügbar unter: <http://www.streamingmedia.com/Articles/ReadArticle.aspx?ArticleID=96946>
- [354] RAMOS-MUNOZ, J.J.; PRADOS-GARZON, J.; AMEIGEIRAS, P.; NAVARRO-ORTIZ, J.; LOPEZ-SOLER, J.M.: Characteristics of mobile youtube traffic. In: *IEEE Wireless Communications*. 2014-02. Vol. 21, no. 1, p. 18–25. DOI 10.1109/MWC.2014.6757893. ISSN 1536-1284.
- [355] MINHAS, Tahir Nawaz; FIEDLER, Markus: Impact of disturbance locations on video quality of experience. In: *Quality of Experience for Multimedia Content Sharing, EuroTV*. 2011.
- [356] ETSI TS 125 306 V11.8.0 - *Universal Mobile Telecommunications System (UMTS); UE Radio Access capabilities (3GPP TS 25.306 version 11.8.0 Release 11)* [online]. Technical Specification. European Telecommunications Standards Institute (ETSI), 2014. [Zugriff am 2014-03-11]. Verfügbar unter: http://www.etsi.org/deliver/etsi_ts/125300_125399/125306/11.08.00_60/ts_125306v110800p.pdf
- [357] MILTON, Michael: *Datenanalyse von Kopf bis Fuß*. Köln : O'Reilly, 2010. ISBN 9783897219595.

A Anhang

A.1 QoS und QoE

Tabelle 67: Endgerätekategorien und maximale Datenraten

Spezifikation	3GPP Release	UE-Kategorie Endgerät	technologische Besonderheit	max. Datenrate Downlink ca. (MBit/s)	max. Datenrate Uplink ca. (MBit/s)	
HSPA / HSPA+	HSDPA	5	1		1.2	
		5	2		1.2	
		5	3		1.8	
		5	4		1.8	
		5	5		3.6	
		5	6		3.6	
		5	7		7.2	
		5	10		14.0	
		7	14		21.1	
	8	20	2x MIMO	42.2		
	8	21	Dual-Cell	23.4		
	8	24	Dual-Cell	42.2		
	10	32	Quad-Cell + MIMO	168.8		
	HSUPA	6	1			0.7
		6	2			1.4
		6	3			1.4
		6	4			2.8
		6	5			2.0
		6	6			5.8
7		7			11.5	
9		8	Dual-Cell		11.5	
9		9	Dual-Cell		23	
LTE	8	1	1x MIMO	10.3	5.2	
	8	2	2x MIMO	51.0	25.5	
	8	3	2x MIMO	102.0	51.0	
	8	4	2x MIMO	150.8	51.0	
	8	5	4x MIMO	299.6	75.4	
LTE Advanced	10	6	2x oder 4x MIMO	301.5	51.0	
	10	7	2x oder 4x MIMO	301.5	102.0	
	10	8	8x MIMO	2998.6	1497.8	

UE: User Equipment, MIMO: Multiple Input Multiple Output, Darstellung soweit relevant verkürzt (zusammengestellt aus [356] S. 46ff und [164] Tab. 7-7, 7-8)

Ausgehend von den Spezifikationen unterstützten zum Zeitpunkt der Testplanung die markt-gängigen mobilen Endgeräte der Spitzenklasse (Smartphones/Tablets) HSPA+ im Downlink bis Kategorie 24 und im Uplink bis Kategorie 6, LTE-fähige Modelle LTE bis Kategorie 3.

A.2 Ende-zu-Ende-Messungen im Testbed

Tabelle 68: Ende-zu-Ende-Messungen für die ausgewählten Netzparameterkombinationen im Testbed unter Android

SWT-WLAN-Testbed				Samsung Galaxy Ace 2				Samsung Galaxy SIII				Samsung Galaxy SIII					
Parameterkombinationen der Presets		cnlab Speedtest				cnlab Speedtest				iperf				overall			
max. DL kbit/s	min. UL kbit/s	ms	down kbit/s	up kbit/s	"Ping" ms	down kbit/s	up kbit/s	Antwortzeit ms	down kbit/s	up kbit/s	down %	up %	down %	up %			
80	48	300	78	98%	49	102%	323	72	90%	48	100%	330	76	95%	44	91%	98%
128	96	300	126	98%	93	97%	324	109	85%	93	97%	366	121	95%	90	93%	96%
225	192	300	215	96%	178	93%	321	199	88%	189	98%	377	208	92%	171	89%	93%
384	384	180	366	95%	355	92%	203	337	88%	359	93%	300	360	94%	364	95%	94%
900	700	90	849	94%	652	93%	112	733	81%	654	93%	180	845	94%	664	95%	94%
1200	1024	90	1127	94%	945	92%	112	1126	94%	937	92%	192	1137	95%	968	95%	93%
1800	1400	90	1647	92%	1281	92%	111	1741	97%	1331	95%	117	1608	89%	1352	97%	94%
2400	1400	90	2212	92%	1280	91%	112	2048	85%	1331	95%	114	2273	95%	1352	97%	94%
3600	1400	90	3290	91%	1281	92%	111	3379	94%	1331	95%	187	3379	94%	1352	97%	94%
3600	2800	90	3288	91%	2506	90%	111	3379	94%	2662	95%	117	3379	94%	2550	91%	93%
5000	1400	50	4663	93%	1305	93%	71	4710	94%	1331	95%	81	4782	96%	1362	97%	95%
7200	5760	50	6644	92%	4338	75%	72	6758	94%	5530	96%	77	6871	95%	5540	96%	89%

(teils umgerechnet auf kbit/s)

Beispielmessungen, Ausreißer sind möglich

Die min. Latenz der Presets ist als Summe der Delay-Werte in Downstream- und Upstream-Richtung angegeben.

Tabelle 69: Ende zu Ende Messungen für die ausgewählten Netzparameterkombinationen im Testbed unter iOS

SWT-WLAN-Testbed		iPhone 5				iPhone 5				iPhone 5							
Parameterkombinationen der Presets		Ookla Speedtest				cnlab Speedtest				iperf				overall			
max. DL	min. UL	max. DL	min. UL	Latenz	"ping"	down	up	down	up	Antwortzeit	down	up	down	up	down	up	
kbit/s	kbit/s	kbit/s	kbit/s	ms	ms	kbit/s	kbit/s	kbit/s	kbit/s	ms	kbit/s	kbit/s	kbit/s	kbit/s	%	%	
80	48	300	72	90%	41	85%	322	76	95%	47	98%	n. A.	76	95%	46	95%	93%
128	96	300	123	96%	92	96%	316	119	93%	93	97%	322	121	95%	91	95%	96%
225	192	300	215	96%	174	91%	315	206	92%	180	94%	322	209	93%	181	94%	93%
384	384	180	369	96%	369	96%	197	357	93%	358	93%	201	359	93%	362	94%	94%
900	700	90	870	97%	676	97%	106	842	94%	656	94%	112	844	94%	663	95%	95%
1200	1024	90	1157	96%	983	96%	105	1112	93%	952	93%	111	1147	96%	968	95%	95%
1800	1400	90	1741	97%	1352	97%	105	1646	91%	1294	92%	111	1710	95%	1352	97%	95%
2400	1400	90	2324	97%	1352	97%	105	2172	91%	1295	93%	112	2263	94%	1352	97%	95%
3600	1400	90	3482	97%	1280	91%	106	3222	90%	1268	91%	111	3379	94%	1352	97%	93%
3600	2800	90	3451	96%	2806	100%	106	3232	90%	2548	91%	115	3379	94%	2693	96%	96%
5000	1400	50	4792	96%	1393	99%	66	4581	92%	1308	93%	72	4782	96%	1352	97%	96%
7200	5760	50	6963	97%	5550	96%	66	6571	91%	5289	92%	71	6871	95%	5560	97%	94%

(umgerechnet auf kbit/s) (teilis umgerechnet auf kbit/s)

n. A. = keine vernünftige Messung möglich, da Latenzmessung erfolgt, noch bevor der Upload abgeschlossen ist

Beispielmessungen, Ausreißer sind möglich

Die min. Latenz der Presets ist als Summe der Delay-Werte in Downstream- und Upstream-Richtung angegeben.

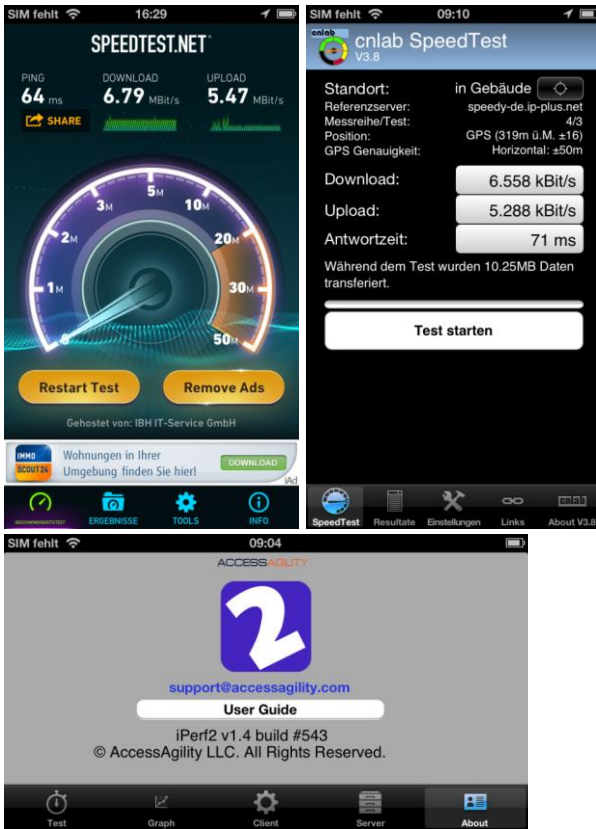


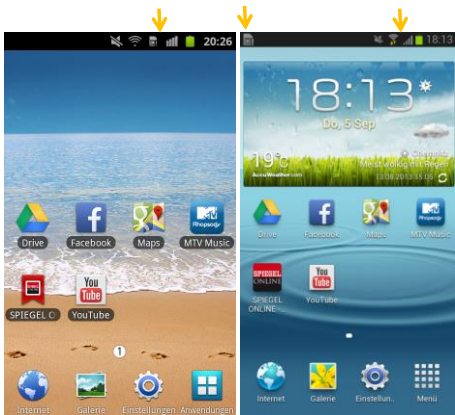
Abbildung 114: Screenshots der Apps für die Ende-zu-Ende-Messungen im Testbed (iOS)

A.3 Testparcours



Samsung Galaxy Tab 2 10.1

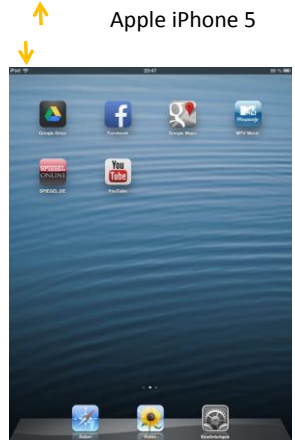
Apple iPhone 5



Samsung Galaxy Ace 2

Samsung Galaxy SIII

Android (Samsung Galaxy Reihe)



Apple iPad 4

iOS (6)

Abbildung 115: Startansicht mit Position der Android-Systemleiste und iOS-Statusleiste auf den Endgeräten

eigentliche Bildschirmgrößen unterschiedlich, zur Darstellung auf einheitliche Höhe skaliert, Pfeil jeweils auf Netzindikatoren

Tabelle 70: QoS-Netzparameterwerte der Presets

Preset	max. Datenrate Download (kBit/s)	max. Datenrate Upload (kBit/s)	min. Latenz (Delay-Summe) (ms)
D10	384	384	180
D20	900	700	90
D30	1800	1400	90
D40	3600	2800	90
D50	7200	5760	50
F10	80	48	300
F20	128	96	300
F30	225	192	300
F40	384	384	180
F50	900	700	90
G10	384	384	180
G20	900	700	90
G30	1800	1400	90
G40	3600	2800	90
G50	7200	5760	50
M10	128	96	300
M20	225	192	300
M30	384	384	180
M40	900	700	90
M50	1800	1400	90
S10	128	96	300
S20	225	192	300
S30	384	384	180
S40	900	700	90
S50	1800	1400	90
Y10	1200	1024	90
Y20	1800	1400	90
Y30	2400	1400	90
Y40	3600	1400	90
Y50	5000	1400	50

Bildung der min. Latenz als Summe der jeweils halben Werte für Delay in Downstream und Upstream (z. B. 25ms + 25ms = 50ms)

Tabelle 71: Effektivwerte Datenrate Download in Drive

Preset	Datenrate Download (alle Geräte)							
	Max.- Wert	Ist-Wert (absolut und prozentual zum Max.-Wert)						
		Minimum		Mittelwert		Maximum		n
	(kbit/s)	(kbit/s)	(%)	(kbit/s)	(%)	(kbit/s)	(%)	
D10	384	362	94%	368	96%	371	97%	18
D20	900	833	93%	851	95%	871	97%	16
D30	1800	1596	89%	1647	91%	1743	97%	19
D40	3600	2517	70%	2994	83%	3104	86%	22
D50	7200	5035	70%	5516	77%	5809	81%	21

Stichproben aus Verkehrsdatenaufzeichnungen, nur valide Aufgabenausführungen

Tabelle 72: Effektivwerte Datenrate Upload in Drive

Preset	Datenrate Upload (ohne Ace 2)							
	Max.- Wert	Ist-Wert (absolut und prozentual zum Max.-Wert)						
		Minimum		Mittelwert		Maximum		n
	(kbit/s)	(kbit/s)	(%)	(kbit/s)	(%)	(kbit/s)	(%)	
D10	384	220	57%	334	87%	368	96%	13
D20	700	302	43%	605	86%	670	96%	13
D30	1400	629	45%	1177	84%	1333	95%	16
D40	2800	1563	56%	2053	73%	2289	82%	16
D50	5760	2730	47%	3937	68%	4820	84%	18
Preset	Datenrate Upload (nur Ace 2)							
	Max.- Wert	Ist-Wert (absolut und prozentual zum Max.-Wert)						
		Minimum		Mittelwert		Maximum		n
	(kbit/s)	(kbit/s)	(%)	(kbit/s)	(%)	(kbit/s)	(%)	
D10	384	334	87%	336	87%	338	88%	5
D20	700	617	88%	619	88%	621	89%	3
D30	1400	956	68%	1108	79%	1180	84%	4
D40	2800	708	25%	1240	44%	1425	51%	5
D50	5760	1691	29%	2155	37%	2604	45%	4

Stichproben aus Verkehrsdatenaufzeichnungen, nur valide Aufgabenausführungen

A.4 Fragebogen und Sicht der Nutzer

Für die Fragenbögen gilt: ©2013 Professuren AAP und MI, TU Chemnitz. Wiedergabe und Erläuterung dienen lediglich zur Dokumentation (Beleg). Die Rechte am Fragebogen liegen bei den genannten Urhebern bzw. beteiligten Projektpartnern. Eine weitere Nutzung ist nicht gestattet. Die Fragenbögen sind in enger Kooperation der Projektpartner entstanden. Dem Anteil von AAP sind zuzurechnen: Dimensionsfestlegung der Bewertungen (Zufriedenheit), Formulierung der allgemeinen Aufgabenstellung und Fragen, grundlegende Realisierung des Fragebogens in Limesurvey, Inhalt der Vorbefragung und Nachbefragung. Durch MI (den Autor) wurde dabei vorrangig beigetragen:

- inhaltliche Aspekte: Vorüberlegungen und Voruntersuchungen, aus denen Anwendungsszenario, Aufgabenstellung, Schrittfolge und relevante Teilbewertungen abgeleitet wurden; Mitwirkung bei konkreten Formulierungen, Erstellung und Bereitstellung der eingebundenen Illustrationen
- technische Aspekte: Erstellung des Programmcodes im Web-Fragebogen-System zur automatischen Fragebogenanpassung, Testbed-Steuerung und Testlaufdokumentation

Die Formulierung der Aufgabenstellungen erfolgte auf Basis beispielhafter Nutzungsszenarien aus den Voruntersuchungen zu den Anwendungen.

Die beschriebene Schrittfolge wird jeweils in der Nutzersicht anhand von Screenshots²²⁸ dokumentiert (Belegfunktion), vor allem mit den beabsichtigten Nutzerstimuli. So soll versucht werden, mit statischen Abbildungen bestmöglich das Nutzererleben in dynamischen Nutzungsprozessen zu dokumentieren. Die Beschriftung und zugehörige Nummerierung der einzelnen Abbildungsbestandteile korrespondiert dabei mit der beschriebenen Schrittfolge des Fragebogens zum jeweiligen Nutzungsszenario.

²²⁸ Die gezeigten Screenshots stammen zum Teil aus der Phase der Testvorbereitung, zum Teil aber auch aus der Phase nach den Tests, um jeweils repräsentative Schritte und konsistente Bildserien der relevanten Phasen für die Dokumentation zeigen zu können. Die Nutzersicht und Nutzerstimuli entsprechen dabei denen der eigentlichen Testdurchführung. Die Screenshots zu den einzelnen mobilen Diensten, von Anwendungen und darin gezeigten Inhalten dienen lediglich zu Dokumentations- und Belegzwecken. Die Bildrechte und sonstigen Rechte wie Urheberrechte etc. der darauf gezeigten Inhalte liegen bei den jeweiligen Rechteinhabern.

A.4.1 Screening-Fragebogen

Die Vorabbefragung mittels Screening-Fragebogen diente zur Auswahl der Probanden. Die Konkretisierung der Fragen entspricht der Nachbefragung mit jeweils vergleichbarer Erfassung, weshalb an dieser Stelle nur eine verkürzte Darstellung erfolgt.

Nutzung internetfähiger mobiler Endgeräte (Smartphone oder Tablet als Voraussetzung):

- Welche der folgenden Geräte nutzen Sie im privaten Gebrauch?
 - Smartphone (ein internetfähiges Gerät mit Touchscreen)
 - Tablet
 - Laptop
 - nichtinternetfähiges Handy

Gerätespezifische Erfassung für Smartphone und Tablet je nach Angabe, hier Beispiel für Smartphone:

- Über welches Betriebssystem verfügt Ihr Smartphone?
- Welches Smartphone-Modell nutzen Sie?
- Haben Sie bei Ihrem Mobilfunkanbieter auch einen Datenvertrag oder eine Datenflatrate abgeschlossen (Prepaid-Karte oder Mobilfunkvertrag), mit der Sie unterwegs auf dem Smartphone das mobile Internet (kein W-LAN) nutzen?
- Seit wann nutzen Sie das mobile Internet auf dem Smartphone?
- Welches Netz nutzen Sie für das mobile Internet auf dem Smartphone?
- Bei welchem Anbieter nutzen Sie das mobile Internet auf dem Smartphone?

Allgemeine Angaben:

- Bitte tragen Sie ihren Beruf oder ihren Studiengang ein.
- Bitte geben Sie ihr Alter an.
- Bitte geben Sie ihr Geschlecht an.

A.4.2 Startseite des Fragebogens

Die Nutzer wurden durch eine Startseite des Fragebogensystems begrüßt (siehe Abbildung 116). Die eingeblendeten Hinweise deckten sich dabei mit den mündlichen Instruktionen.

Vielen Dank, dass Sie an unserer **Studie zur mobilen Internetnutzung** teilnehmen!

Sie werden in den kommenden 60 Minuten einige **Aufgaben** bearbeiten. Bitte **lesen** Sie sich die jeweiligen **Aufgabenstellungen** vor der Bearbeitung immer genau durch. Am **Ende jeder Aufgabe** beantworten Sie bitte die **Fragen zur jeweiligen Aufgabe**.

Beachten Sie bitte: Es gibt keine richtigen oder falschen Antworten auf die Fragen. Wir sind vielmehr an Ihrer ganz persönlichen Meinung interessiert. Unter Umständen kann es zu kurzen Wartezeiten kommen. Bitte schließen Sie während der Bearbeitung der verschiedenen Aufgaben **NICHT** den PC-Browser.

Ihre Daten werden von uns selbstverständlich **vertraulich** und **anonym** behandelt.

Wenn Sie noch weitere **Hintergrundinformationen** zu dieser Studie wünschen, stehen wir Ihnen gerne im Anschluss an den Versuch zur Verfügung.

Abbildung 116: Startseite zum Fragebogen

A.4.3 Aufbau des Fragebogens

Der Aufbau der einzelnen Seiten des Fragebogens zum Testparcours ist in Abbildung 117a skizziert. In der weiteren Darstellung werden vorab die sich wiederholenden und einheitlichen Bestandteile behandelt, während die Darstellung der Szenarien sich jeweils auf Hinweise, Aufgabenstellung und die aufgabenspezifischen Bewertungen beschränkt. Die anderen Bestandteile möge sich der geneigte Leser entsprechend des skizzierten Aufbaus jeweils hinzudenken.

Im Kopfbereich der Fragebogenseiten erfolgte eine Einblendung zum Bearbeitungsfortschritt des gesamten Fragebogens zusammen mit der Aufgabenbezeichnung, siehe Abbildung 117b. Am Ende der Fragebogenseiten befand sich eine Schaltfläche „Weiter“ zum Wechsel auf die nächste Fragebogenseite, siehe Abbildung 117c. Andere standardmäßige Navigationselemente von Limesurvey wurden durch die eingebundenen JavaScript-Snippets der Testlaufsteuerung automatisiert ausgeblendet, um versehentliche Fehlbedienung (Aufbruch des Fragebogens) zu verhindern.

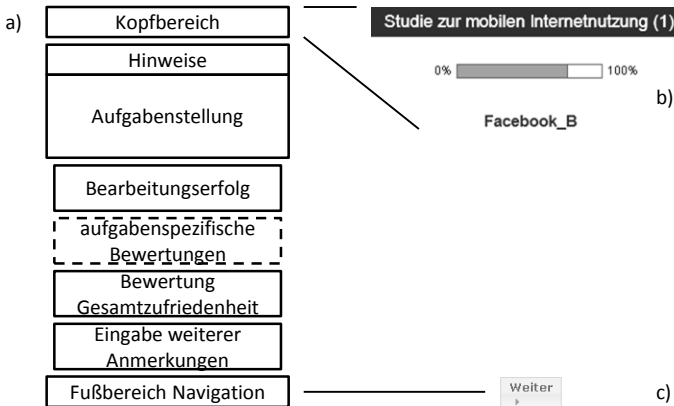


Abbildung 117: Schematischer Aufbau einer Fragebogenseite

- a) Überblick zum Aufbau: Die Einblendung der aufgabenspezifischen Bewertungen erfolgte jeweils erst, wenn ein Proband beim Bearbeitungserfolg eine andere Option als „nein, Aufgabe konnte gar nicht erst gestartet werden“ wählte.
- b) Beispiel aus dem Kopfbereich mit Titel, Fortschritt im Fragebogen sowie Aufgabekürzel mit Szenario und Variante
- c) Navigation über Schaltfläche zur nächsten Fragebogenseite, andere Navigationselemente waren ausgeblendet

A.4.4 Bearbeitungserfolg

Wenn ein Proband angab, nicht beim ersten Versuch der Aufgabenausführung erfolgreich gewesen zu sein, dann erfolgte die Einblendung der zusätzlichen Frage zum Grund, siehe Abbildung 118.

* Können Sie die Aufgabe **erfolgreich** durchführen?
Bitte wählen Sie eine der folgenden Antworten:

- ja, Aufgabe beim ersten Versuch erfolgreich durchgeführt
- ja, Aufgabe nach mehreren Versuchen erfolgreich durchgeführt
- nein, Aufgabe nicht beendet/ abgebrochen
- nein, Aufgabe konnte gar nicht erst gestartet werden

* Aus welchem **Grund** konnte die Aufgabe **nicht beendet**, **nicht gestartet** oder **nicht beim ersten Versuch beendet** werden?
Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

- Probleme mit der Internetverbindung (z. B. Seitenaufbau hat nicht funktioniert, Fehlermeldung, Internetverbindung abgebrochen)
- Schwierigkeiten beim Bedienen der Anwendung (z. B. unklare Navigation, wusste nicht was ich drücken soll)
- Schwierigkeiten beim Bedienen des Geräts
- Aufgabenstellung unklar/nicht verstanden
- sonstiges (bitte ergänzen)

Abbildung 118: Mögliche Angaben zum Bearbeitungserfolg und Grund im Fragebogen

A.4.5 Bewertung Gesamtzufriedenheit und Anmerkungen

Aus Platzgründen wird bei den nachfolgenden Darstellungen der einzelnen Szenarien auf den für alle Aufgaben einheitlichen Teil am Ende der jeweiligen Fragebogenseite verzichtet. Der gemeinsame einheitliche Teil besteht aus (siehe Abbildung 119):

- Bewertung der Gesamtzufriedenheit (nach den einzelnen spezifischen Teilbewertungen)
- einem Kommentarfeld (Freitext) für weitere Anmerkungen, Eingabe freiwillig

Wie **zufrieden** waren Sie **insgesamt mit dem mobilen Internet** in dieser Situation?

Bitte wählen Sie eine der folgenden Antworten:

- vollkommen zufrieden
- eher zufrieden
- weder zufrieden, noch unzufrieden
- eher unzufrieden
- vollkommen unzufrieden

Weitere Anmerkungen zu der Aufgabenbearbeitung können Sie gerne hier ergänzen:

Abbildung 119: Gemeinsamer Fragebogenteil zu jeder Aufgabe

A.4.6 Drive

Hinweise

Sie nutzen einen von uns eingerichteten Account, über den Sie automatisch nach Öffnung der Drive-App angemeldet sind. Sie sollten sich nicht über Ihren privaten Account anmelden.

Aufgabenstellung

- Öffnen Sie bitte die auf der Oberfläche des mobilen Endgeräts installierte **Applikation „Drive“** (bei iOS **„Google Drive“**).
- Sie nutzen einen von uns eingerichteten Account, über den Sie automatisch nach Öffnung der Drive-App angemeldet sind. Sie sollten sich **nicht** über Ihren privaten Account anmelden.

Teil 1: Download

- Wählen Sie auf dem Startbildschirm den **Ordner „Meine Ablage“**. Darin sehen Sie nun eine Reihe von **Dateien**. **Wählen Sie die Datei Bild1.jpg aus den vorhandenen Dateien aus, indem Sie den Dateinamen in der Mitte der Zeile antippen und somit öffnen.**
- Bitte **warten** Sie, bis der Ladevorgang abgeschlossen ist. Wiederholen Sie das Laden bitte nicht! **Bitte schauen Sie sich die Datei nach Beendigung des Ladevorgangs kurz an.**
- Gehen Sie zurück zum Bildschirm **„Meine Ablage“**:
 - Bei Android-Geräten (Plätze Nr. 1, 2, 3, 6)**: „zurück“ (Taste oder Softkey).
 - Bei iOS-Geräten (Plätze Nr. 4, 5)**: Schaltfläche oben links „close“.

Teil 2: Upload

- Bitte **laden** Sie nun eine **weitere Datei** in „Meine Ablage“ **hoch**:
 - Bei Android-Geräten (Plätze Nr. 1, 2, 3, 6)**:
 - Ausgangspunkt ist der Ordner „Meine Ablage“
 - Platz 1**: Wählen Sie das Hochlade-Icon  (= unterstrichener Pfeil nach oben) aus dem Menü-Band aus.
 - Plätze 2, 3, 6**: Wählen Sie die Menü-Taste und dann den Menüpunkt „Hochladen“.
 - Wählen Sie den Ordner „DCIM“ bzw. „Kamera“ in der Albenübersicht aus.
 - Es erscheint nun eine Galerieansicht mit Vorschau Bildern und ggf. die Überschrift „Wählen Sie ein Video aus“ (lassen Sie sich davon nicht irritieren).
 - Wählen Sie eines der Vorschau Bilder aus. Die Datei wird nun hochgeladen.
 - Der Ladebalken läuft verdeckt im System-Statusbereich. Dieser öffnet sich mit einer Wischbewegung vom oberen (Smartphone) bzw. unteren rechten Rand (Tablet) her.
 - Bei iOS-Geräten (Plätze Nr. 4, 5)**:
 - Ausgangspunkt ist der Ordner „Meine Ablage“.
 - Wählen Sie das „+“-Icon oben rechts  und anschließend „Fotos oder Videos hochladen“ aus.
 - Wählen Sie in der Alben-Übersicht „Aufnahmen“ aus.
 - Es erscheint nun eine Galerieansicht mit Vorschau Bildern.
 - Wählen Sie eines der Vorschau Bilder aus. Es erscheint ein Häkchen. Nun wählen Sie die Schaltfläche „Hochladen“ oben rechts aus.
 - Sie sehen nun einen Ladebalken ist am unteren Bildrand (**Platz Nr. 5**) bzw. einen Ladekreis links bei „Uploads“ (**Platz Nr. 4**).
- Bitte **warten** Sie, bis der **Ladevorgang abgeschlossen** ist. Wiederholen Sie das Laden bitte **nicht!**
- Nach Beendigung des Ladevorgangs sollte die Datei in **„Meine Ablage“** aufgelistet sein.
- Schließen** Sie bitte danach die **Drive-Applikation vollständig** und beantworten Sie die unten stehenden **Fragen zur Nutzung von Drive**.

Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

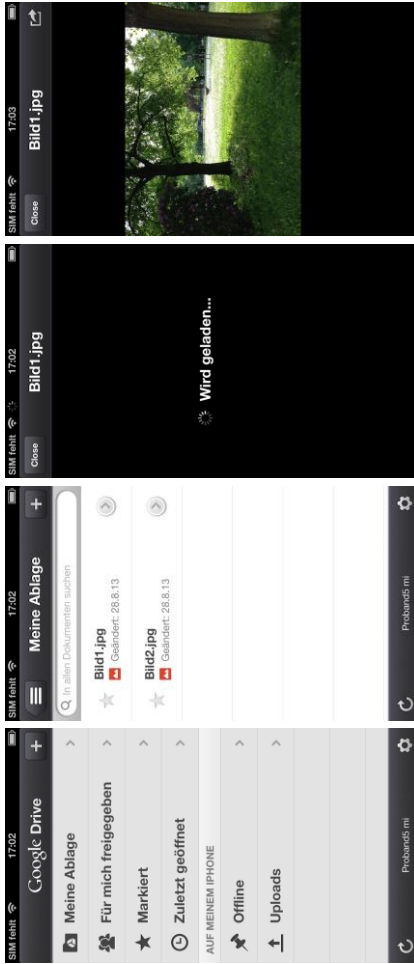
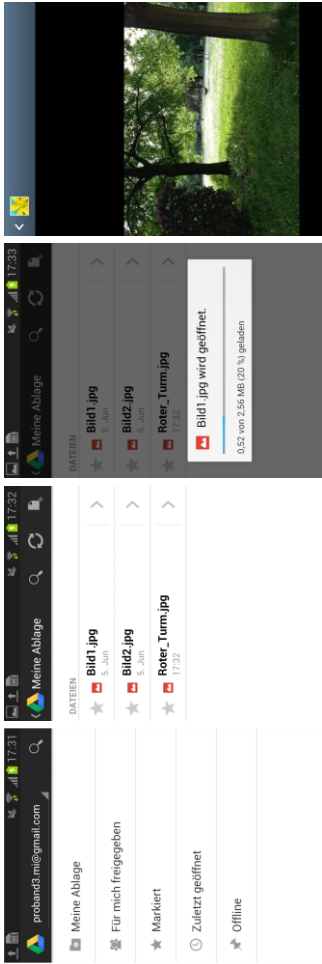
	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer des vollständigen Downloads der Datei	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer des vollständigen Uploads der Datei	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 120: Spezifischer Fragebogenteil zum Szenario Drive

Download

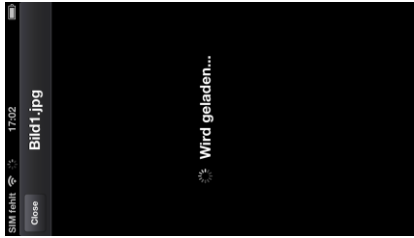
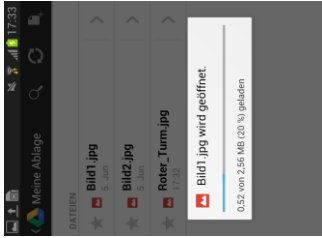
Android
(Darstellung
beschnitten)

iOS



3.1) Startbildschirm der App

3.2) Auswahl des Bildes zum
Download aus der Cloud



4.1) Download des Bildes

4.2) Download beendet
Anzeige des Bildes

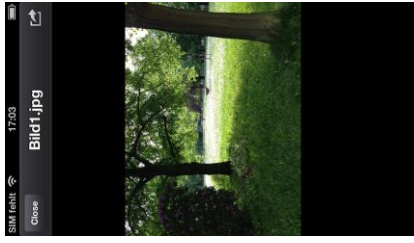
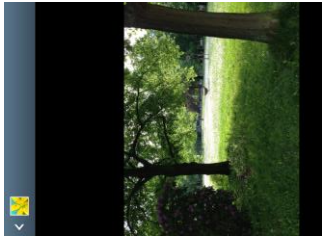


Abbildung 121: Zusammenstellung zur Nutzersicht im Szenario Drive (Fortsetzung →)
Screenshots aus der zugehörigen App

Upload

Android
(Darstellung beschnitten)

1 Datei wird hochgeladen.
770,0 KB
Keine SIM
Bitte SIM-Karte einlegen

Ladebalken im Benachrichtigungsbereich

6.1 Auswahl des Bildes
(Darstellung verkürzt)

iOS

1 Datei wird hochgeladen.
Meine Ablage

Ladekreis

6.2 / 7.1 Upload des Bildes
(iOS: iPhone)

iOS

1 Datei wird hochgeladen.
Meine Ablage

Ladekreis

8) Upload beendet
Bild erscheint in Liste

iOS

1 Datei wird hochgeladen.
Meine Ablage

Ladekreis

7.2 Upload des Bildes
(iOS: iPad, Ausschnitt)

Fortsetzung der Abbildung 121



Kennzeichnung durch Hinweisblasen und zugefügte textuelle Beschreibung zur Verdeutlichung

A.4.7 Facebook

Hinweise

Sie nutzen einen von uns eingerichteten Account, über den Sie automatisch nach Öffnung der Facebook-Applikation angemeldet sind. Sie sollten sich nicht über Ihren privaten Account anmelden.

Aufgabenstellung

1. Öffnen Sie bitte die auf der Oberfläche des mobilen Endgeräts installierte **Applikation „Facebook“**.
2. Sie nutzen einen von uns eingerichteten Account, über den Sie automatisch nach Öffnung der Facebook-App angemeldet sind. Sie sollten sich **nicht** über Ihren privaten Account anmelden.
3. Wählen Sie **„Foto“**, um ein neues **Bild hochzuladen**.
 - a. **Bei den Geräten 1,2,3,5,6**: Wählen Sie eines aus den vorhandenen Bildern aus, indem Sie es markieren. Bestätigen Sie Ihre Auswahl, indem Sie auf das viereckige Symbol rechts unten Wählen Sie eines aus den vorhandenen Bildern aus, indem Sie es markieren. Bestätigen Sie Ihre Auswahl, indem Sie auf das viereckige Symbol rechts unten (Android: ; iOS: ) klicken.
 - b. **Bei Gerät 4**: Klicken Sie auf „aus Bibliothek auswählen“, nun auf „Aufnahmen“ und dann darin auf das entsprechende hochzuladende Bild. Wählen Sie danach „Verwenden“.
4. Drücken Sie **„Posten“**, um das Bild hochzuladen.
5. Sie können nun an dem Ladebalken verfolgen, wie weit der Ladevorgang ist.
 - a. **Bei Android-Geräten (Plätze Nr. 1, 2, 3, 6)**: Der Ladebalken läuft im System-Statusbereich. Dieser öffnet sich mit einer Wischbewegung vom oberen (Smartphone) bzw. unteren rechten Rand (Tablet) her.
 - b. **Bei iOS-Geräten (Plätze Nr. 4, 5)**: Der Ladebalken ist im Vordergrund sichtbar.
6. Bitte **warten** Sie, **bis der Ladevorgang abgeschlossen** ist. Wiederholen Sie das Laden bitte nicht!
 - a. **Bei Android-Geräten (Plätze Nr. 1, 2, 3, 6)**: Der Ladevorgang ist erst abgeschlossen, wenn der **Ladebalken** im System-Statusbereich **verschwunden** ist und die Meldung „Facebook-Upload: 1 von 1 wurden hochgeladen“ mit einem **Häkchen** erscheint.
 - b. **Bei iOS-Geräten (Plätze Nr. 4, 5)**: Der Ladevorgang ist dann abgeschlossen, wenn der Ladebalken im Vordergrund nicht mehr sichtbar ist.
7. Nach Beendigung des Ladevorgangs sollte das Bild auf der **Pinnwand** sichtbar sein. Das Bild erscheint ggf. erst nach manueller Aktualisierung (in der Seite Neuigkeiten den Anfang der Liste nach unten ziehen), ggf. auch leicht zeitverzögert.
8. **Schließen** Sie bitte nach Beendigung des Ladevorgangs die **Facebook-Applikation vollständig** und beantworten Sie die unten stehenden **Fragen zur Nutzung von Facebook**.

Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer bis Pinnwand vollständig geladen ist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer bis das Foto vollständig hochgeladen ist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 122: Spezifischer Fragenbogenteil zum Szenario Facebook

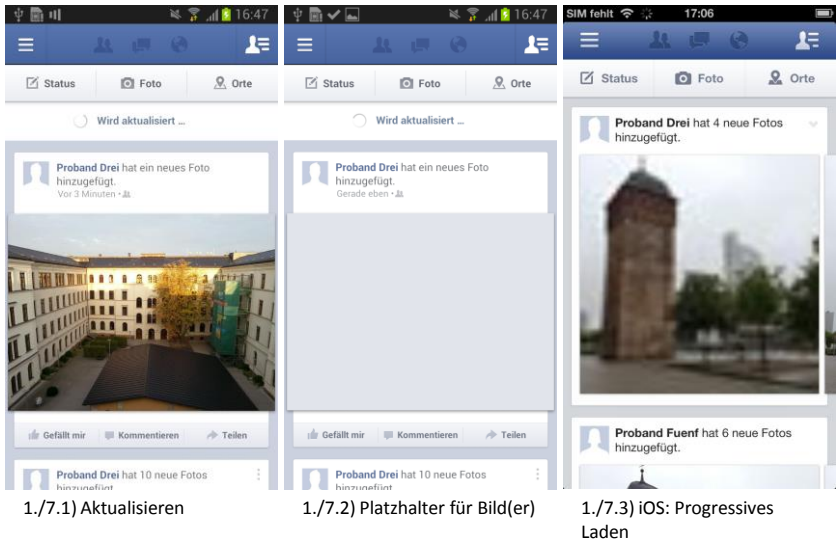


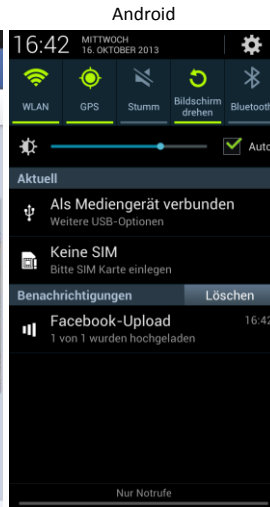
Abbildung 123: Zusammenstellung zur Nutzersicht im Szenario Facebook (Fortsetzung →)
 Schritte 1./7.1, 1./7.2, 1./7.4 und 3 für Android gezeigt, in iOS jeweils vergleichbar



4) Android: Kommentieren des Bildes (optional)
Upload noch nicht gestartet



5.1) Android: Posten ausgelöst
Upload startet, Ablauf im Hintergrund



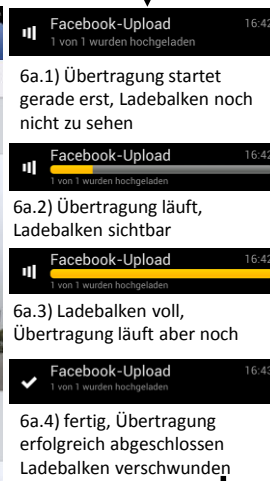
5.2) Android: Upload-Fortschritt im Benachrichtigungsbereich



4) iOS: Kommentieren des Bildes (optional)
Upload bereits gestartet



5 /6.) iOS: Posten ausgelöst
Ladebalken des bereits laufenden Upload



6a.1) Übertragung startet gerade erst, Ladebalken noch nicht zu sehen

6a.2) Übertragung läuft, Ladebalken sichtbar

6a.3) Ladebalken voll, Übertragung läuft aber noch

6a.4) fertig, Übertragung erfolgreich abgeschlossen
Ladebalken verschwunden

7.) nach Upload-Ende: Aktualisierung der Pinwand

Fortsetzung der Abbildung 123

Achtung: unterschiedliche Start-Zeitpunkte für Upload zwischen Android (Schritt 5.1) und iOS (Schritt 4), Screenshots aus der zugehörigen App, Bildausschnitte zum Schritt 6) zur Verdeutlichung

A.4.8 Google Maps

Hinweise

Zur Vergleichbarkeit Ihrer Einschätzung bleiben Sie bitte in der Kartenansicht, wechseln Sie nicht in die Hybridansicht und ändern Sie die Darstellung der Ebenen nicht. Falls die Frage erscheint, ob Ihr aktueller Standort genutzt werden darf, können Sie dem zustimmen.






Aufgabenstellung

1. Öffnen der Anwendung:

- Bei Android-Geräten (Plätze Nr. 1, 2, 3, 6): Bitte öffnen Sie nun die auf der Oberfläche des mobilen Endgerätes installierte Applikation „Maps“.
- Bei iOS-Geräten (Plätze Nr. 4, 5): Bitte öffnen Sie nun die auf der Oberfläche des mobilen Endgerätes installierte Applikation „Google Maps“.

2. Zur Vergleichbarkeit Ihrer Einschätzung bleiben Sie bitte in der Kartenansicht und wechseln Sie **nicht** in die Hybridansicht und ändern Sie die Darstellung der Ebenen nicht. Stimmen Sie ggf. mit „OK“ zu, sollte die Meldung erscheinen, ob GoogleMaps Ihren aktuellen Ort verwenden darf.

3. Planen Sie nun eine **Route mit dem Auto** für den Weg von Ihrem aktuellen Standort zum „Phaeno Wolfsburg“.

- **Plätze Nr. 1, 3, 5, 6:** Zur Routen-Suchfunktion gelangen Sie über das Symbol  bzw. , welches wie ein nach rechts abbiegender Pfeil in einer Raute aussieht. Wählen Sie nach Zieleingabe ggf. die Schaltfläche „Route berechnen“ bzw. „Los“.
- **Platz Nr. 2:** Zur Routen-Suchfunktion gelangen Sie über die „Menü-Taste“. Wählen Sie den Eintrag „Route“. Nach Zieleingabe wählen Sie Schaltfläche „Route berechnen“.
- **Platz Nr. 4:** Zur Routen-Suchfunktion gelangen Sie über das Symbol , welches wie ein nach rechts abbiegender Pfeil aussieht. Wählen Sie ggf. nach Zieleingabe das Symbol . Scrollen Sie danach auf der Routenliste zum oberen Bildende hoch. Wählen Sie dort die Schaltfläche „Kartenansicht“  aus.

4. Schauen Sie sich bitte **die Route als Übersicht** auf der Karte an. Zoomen Sie dafür ggf. heraus.

5. **Zoomen** Sie das **Ziel heran**. Verschieben Sie den Kartenausschnitt, um sich eine Vorstellung der Verkehrsführung für die abschließende Anfahrt in unmittelbarer Nähe des Ziels zu machen und schauen Sie sich nach einem Restaurant in der Nähe um. Zoomen Sie ggf. näher heran.

6. **Schließen** Sie bitte nach Beendigung der Routenplanung die **Google Maps-Applikation (Plätze Nr. 1, 2, 3, 5, 6)** bzw. **Safari (Platz Nr. 4)** und beantworten Sie die unten stehenden **Fragen zur Nutzung von Google Maps**.

Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

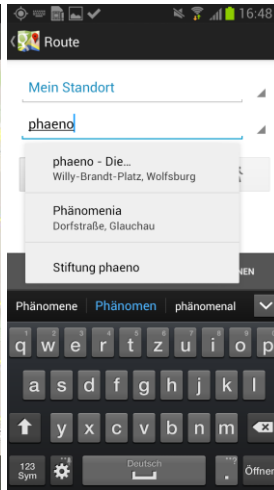
	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer beim Abschicken der Routen-Suchanfrage bis zum Erscheinen der Route	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer beim vollständigen Kartenaufbau (unter Nutzung der Zoom-, Scroll- und Verschiebefunktion)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 124: Spezifischer Fragebogenteil zum Szenario Google Maps

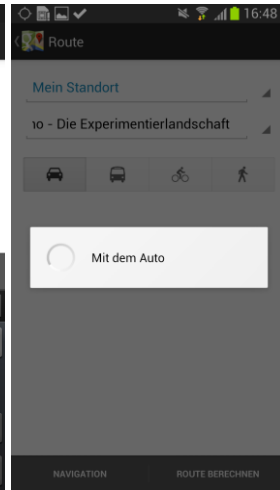
Android-App (iOS mit iPhone-App vergleichbar)



3.1) Startansicht, von aktuellem Standort



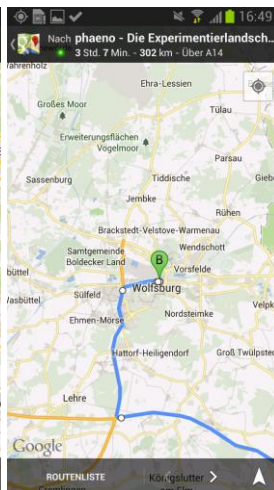
3.2) Zieleingabe mit online Auto-Vervollständigung



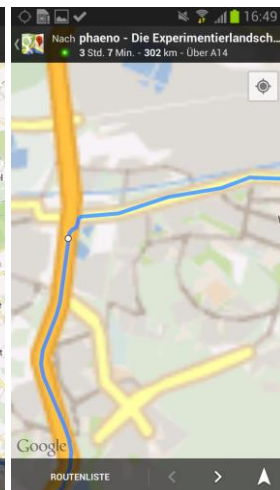
3.3) Laden der Route



4) Kartendarstellung Route Übersicht



5.1) schrittweiser Zoom am Ziel zur Orientierung

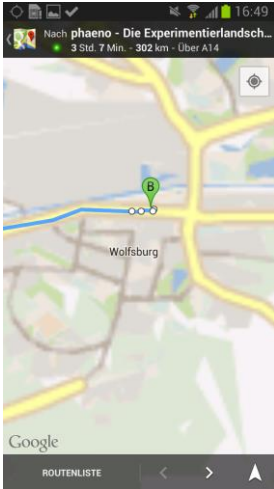


5.2) weiteres Zoomen zur Anfahrt im Zielgebiet

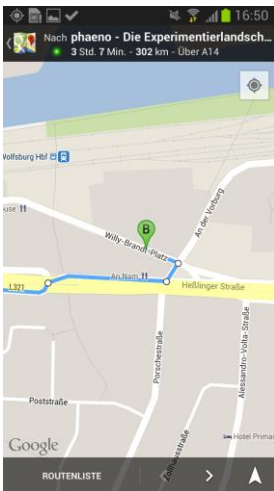
Abbildung 125: Zusammenstellung zur Nutzersicht im Szenario Google Maps (Fortsetzung →) Screenshots aus der zugehörigen App bzw. Browser

Android-App (iOS vergleichbar)

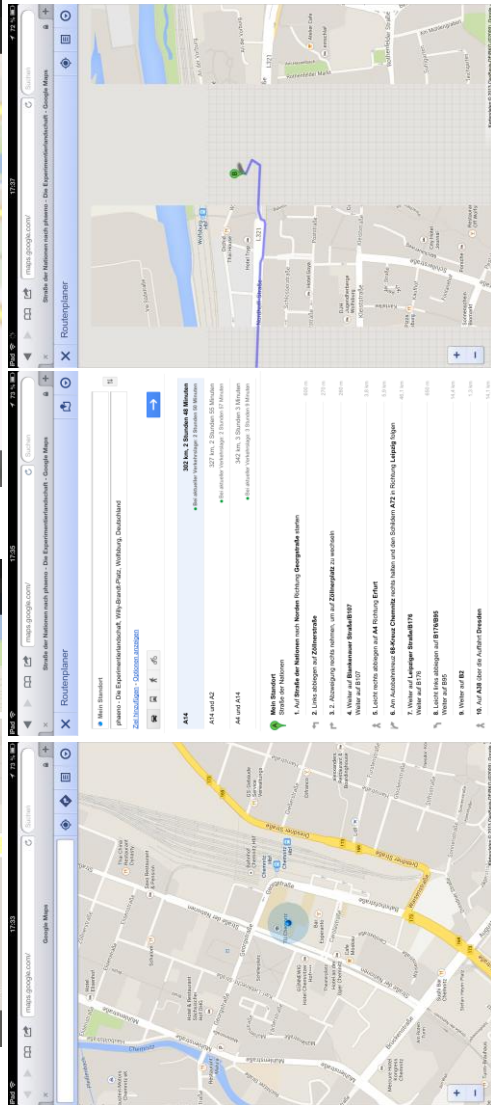
Darstellung im Browser (iPad 4)



5.3) Verschieben des Karten-ausschnitts



5.4) Darstellung des Zieles, Restaurant gefunden, fertig



5) Verschieben/Zoom der Karte zum Ziel

3) Routenplanung als Routenliste

3) Startansicht, von aktuellem Standort



Fortsetzung der Abbildung 125

A.4.9 MTV-Music

Hinweise

Bitte stecken Sie den Kopfhörer in das mobile Endgerät und setzen Sie diesen für die nächste Aufgabe auf. Ändern Sie während der Aufgabe nicht die automatisch gewählte Audio-Qualitätsstufe. Bitte ignorieren Sie mögliche Hilfetexte in der App.

Aufgabenstellung

- Öffnen** Sie nun bitte die auf der Oberfläche des mobilen Endgerätes installierte **Applikation „MTV Music“**.
- Im Folgenden sollen Sie nun ein bestimmtes Musikstück suchen. **Die Suche per Suchfeld ist entweder nach Künstler oder Titel möglich. Sollte im Suchfeld die Suche nach „Titel“ aktiviert sein, ändern Sie dies durch Antippen „Künstler“**. Suchen Sie nun so auf der Startseite über die **Suchfunktion** nach dem **Künstler „David Bowie“**.
 - Bei **Android-Geräten (Plätze Nr. 1, 2, 3, 6)**: Verwenden Sie zur Suche das „Lupe“-Icon  am oberen rechten Rand.
 - Bei **iOS-Geräten (Plätze Nr. 4, 5)**: Verwenden Sie zur Suche das „Lupe“-Icon  am unteren linken Rand.
- Wählen Sie den entsprechenden Künstler aus. Unter „Top-Titel“ finden Sie eine Reihe von Musikstücken. Wählen Sie das **Musikstück** mit dem **Titel „Heroes“**.
- Beginnen Sie bitte das Musikstück **anzuhören**, indem Sie es kurz antippen und wechseln Sie dabei durch Drücken des Cover-Icons zur **Detaillseite des Musikstücks**:
 - Bei **Android-Geräten (Plätze Nr. 1, 2, 3, 6)**: Wechsel zur Detail-/Player-Seite durch Drücken des Cover-Icons (unten links neben den Steuerelementen)
 - Bei **iOS-Geräten (Plätze Nr. 4, 5)**: Sie sind automatisch auf der Detail-/Player-Seite.
- Pausieren** Sie die Wiedergabe **nach etwa der Hälfte** des Musikstücks.
- Stellen Sie nun sicher, dass die **Musikwiedergabe nicht mehr läuft** (also pausiert ist), **schließen** Sie bitte danach **„MTV Music“** auf dem mobilen Endgerät **vollständig** und beantworten Sie die unten stehenden **Fragen zur Nutzung von MTV Music**.

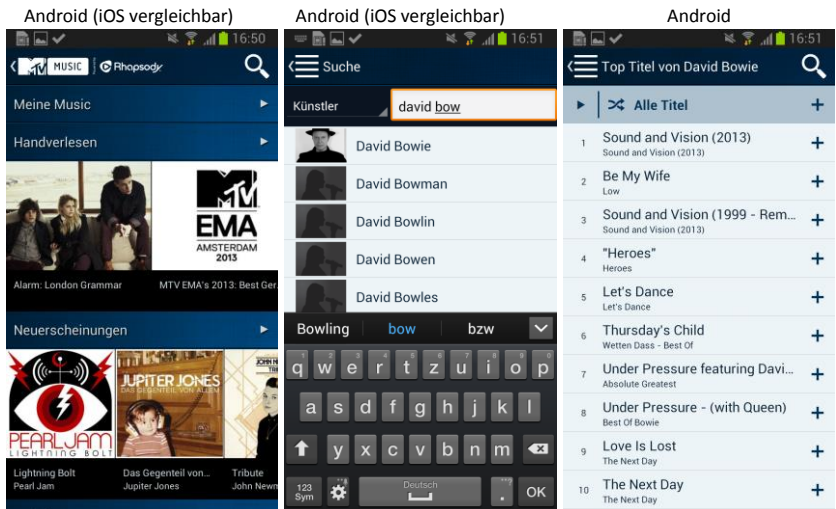
Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer vom Start der Suchanfrage bis zum Erscheinen der Suchergebnisse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer von der Musikauswahl bis zum Start des Musikstücks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Abspielverhalten des Musikstücks	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

* Warum waren Sie mit dem **Abspielverhalten des Musikstücks nicht vollkommen zufrieden**?

Abbildung 126: Spezifischer Fragenbogenteil zum Szenario MTV-Music

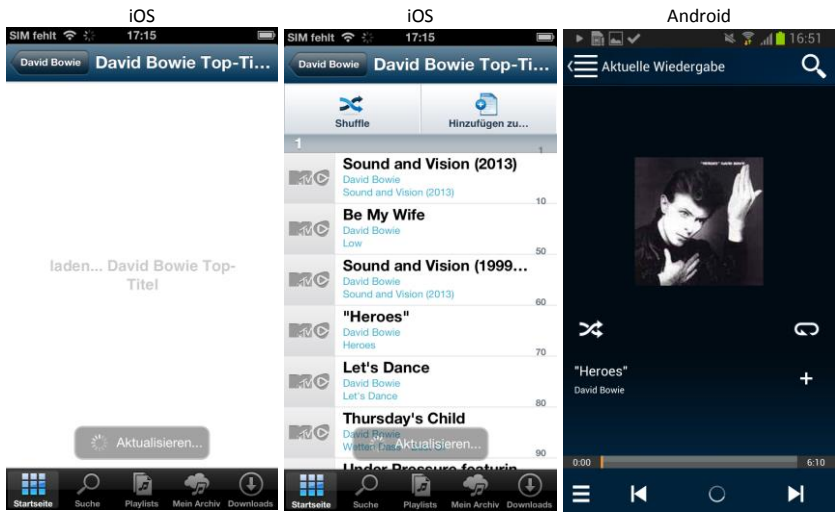
Die Einblendung des Kommentarfeldes zum Abspielverhalten erfolgte, wenn der Proband zum „Abspielverhalten des Musikstücks“ eine andere Bewertung als „vollkommen zufrieden“ wählte.



1.) Laden der Startansicht

2.) Suche nach Künstler (mit online Auto-Vervollständigung)

3.) Laden der Liste der Top-Titel
4.) Start des Titels aus Liste

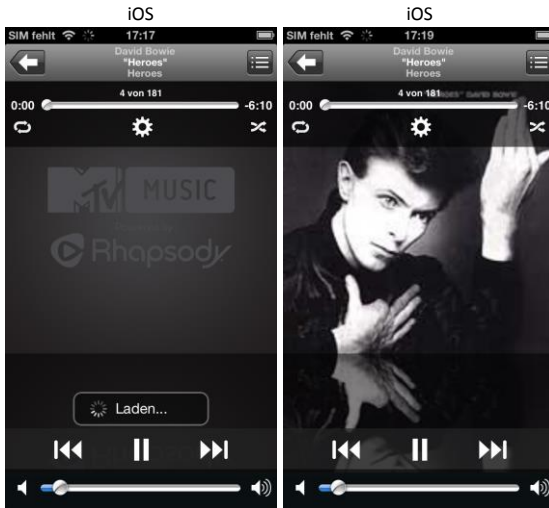


3.1) Laden der Liste der Top-Titel

3.2) Laden der Liste der Top-Titel mit Cover-Icons
Start des Titels aus Liste

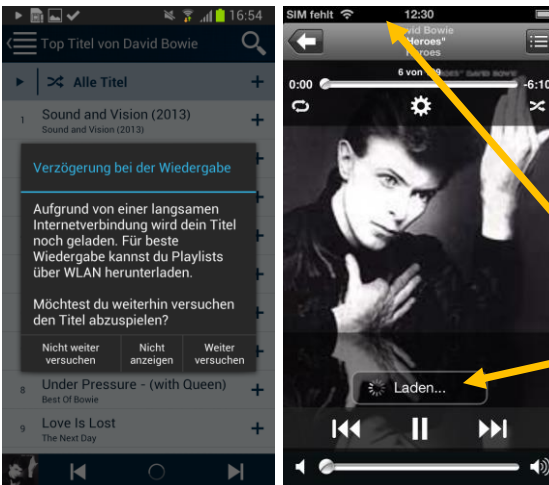
4.a) Player-Seite mit Cover-Icon (Laden optional)

Abbildung 127: Zusammenstellung zur Nutzersicht im Szenario MTV-Music (Fortsetzung →)
Screenshots aus der zugehörigen App



4.1) Laden der Player-Seite, Audio-Puffern, Laden des Covers

4.2) Player-Seite, Cover geladen, Audio gepuffert, Wiedergabe startet, fertig



4.) Android bei M10, Variante a verzögerter Wiedergabestart mit Fehlermeldung, nach weiterer Wartezeit erfolgt Wiedergabestart

4.) iOS bei M10, Variante a

ewiges Puffern ohne automatischen Start der Wiedergabe trotz langer Wartezeit

Übertragungsrate bereits verschwunden = Übertragung des Titels bereits vollständig abgeschlossen

„Laden...“ bleibt ewig, automatischer Wiedergabestart erfolgt nicht, manueller Start wäre sofort möglich

Fortsetzung der Abbildung 127
eingezeichnete Pfeile zur Verdeutlichung

A.4.10 Spiegel.de Hinweise

Bitte beachten Sie, dass Sie nur die mobile Version der Seite nutzen, die automatisch aufgerufen wird. Wechseln Sie bitte nicht die vorausgewählte Version!

Aufgabenstellung

1. Starten Sie bitte über das **Spiegel-Shortcut** die Nachrichtenseite und laden Sie diese vollständig. Bitte beachten Sie, dass Sie nur die mobile Version der Seite nutzen, die automatisch aufgerufen wird. Wechseln Sie bitte **nicht** die vorausgewählte Version!
2. Klicken Sie bitte auf die Schaltfläche „**Rubriken**“ (am oberen Rand der Webseite) oder scrollen Sie bis zum unteren Rand der Webseite und öffnen Sie bitte die **Rubrik „einestages“**. Öffnen Sie dort einen **beliebigen Artikel** Ihrer Wahl **mit einer Fotoserie – außer Artikel** mit dem Titel „Auf den ersten Blick“.
3. Wenn Sie einen Artikel gefunden haben, der eine Fotoserie enthält, **öffnen** Sie bitte die **Fotoserie**, in dem Sie das Bild anklicken.
4. **Sehen** Sie sich die **ersten 4 Bilder der Fotoserie an** und **lesen** Sie die unter den Bildern stehenden **Texte kurz an**.
5. Danach **schließen** Sie bitte **den Browser vollständig** und beantworten Sie die unten stehenden **Fragen zur Nutzung von Spiegel**.

Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer bis die Startseite vollständig geladen ist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer bis der ausgewählte Artikel vollständig geladen ist	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dauer bis die entsprechenden Fotos und Textelemente der Fotostricke vollständig geladen sind	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 128: Spezifischer Fragebogenteil zum Szenario Spiegel.de

iOS mit iPhone (Android vergleichbar)



1.1) Laden der Startseite

1.2) sichtbarer Bereich geladen
Android (zum Vergleich)



1.3) Laden abgeschlossen

2.1) Laden der Liste der
Artikel der Rubrik „eines
Tages“, noch unvollständig

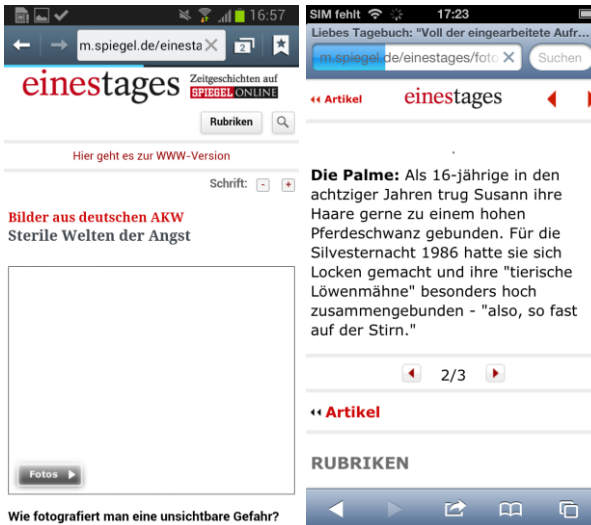
Abbildung 129: Zusammenstellung zur Nutzersicht im Szenario Spiegel.de (Fortsetzung →)
Screenshots der mobilen Webseite aus dem Browser

iOS mit iPhone (Android vergleichbar)



2.2) nach Laden weiterer Elemente abgeschlossen

3.1) Laden eines Artikels, noch unvollständig, iOS



Wie fotografiert man eine unsichtbare Gefahr?

3.2) Laden eines Artikels, noch unvollständig, Android

4) Fotostrecke des Artikels, Laden mehrerer Bilderseiten, nach Bildladen fertig, iOS



Fortsetzung der Abbildung 129

A.4.11 Youtube

Hinweise

Bitte stecken Sie den Kopfhörer in das mobile Endgerät und setzen Sie diesen für die nächste Aufgabe auf. Wechseln Sie während der Aufgabe nicht die automatisch gewählte Video-Qualitätsstufe. Wir haben für Sie Inhalte vorausgewählt; nutzen Sie diese bitte über den beschriebenen Weg.

Aufgabenstellung

- Öffnen Sie nun bitte die auf der Oberfläche des mobilen Endgerätes installierte **Applikation „Youtube“**.
- Wir haben für Sie Inhalte über eine **Videoliste** vorausgewählt. Nutzen Sie diese daher bitte über den beschriebenen Weg. So öffnen Sie die Videoliste:
 - Android-Geräte (Plätze Nr. 1, 2, 3, 6):** Wählen Sie am oberen linken Bildrand das YouTube-Icon  und anschließend bei „Konto“ die angezeigte E-Mail-Adresse aus (der erste Eintrag unter Konto). In der Folge erscheinen nun 5 Kategorien. Wählen Sie da die **Kategorie „Später ansehen“** aus.
 - iOS-Geräte (Plätze Nr. 4, 5):** Wählen Sie am oberen Bildrand das Menü-Icon  (3 horizontale Striche). Wählen Sie am oberen linken Bildrand bei „Konto“ die angezeigte E-Mail-Adresse aus (der erste Eintrag unter Konto). In der Folge erscheinen nun verschiedene Kategorien. Wählen Sie da die **Kategorie „Später ansehen“** aus.
- Bitte wählen Sie folgendes **Video** aus der Liste aus: „**STAR TREK INTO DARKNESS**“ (Länge ca. 1:19 min).
- Schauen** Sie sich dieses **Video** bitte **vollständig** an. Bitte wechseln Sie dabei **nicht** die automatisch gewählte Video-Qualitätsstufe.
- Danach **schließen** Sie bitte **Youtube** auf dem mobilen Endgerät und beantworten Sie die unten stehenden **Fragen zur Nutzung von YouTube**.

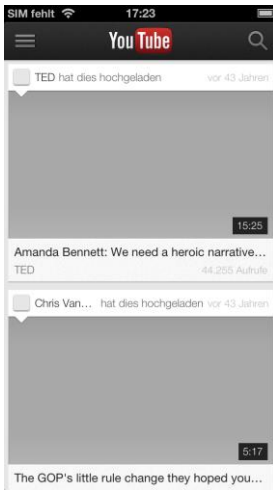
Bitte geben Sie an, wie **zufrieden** Sie mit folgenden **Aspekten während der Aufgabenbearbeitung** waren.

	vollkommen zufrieden	eher zufrieden	weder zufrieden, noch unzufrieden	eher unzufrieden	vollkommen unzufrieden
Dauer von der Videoauswahl bis zum Start des Videos	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Abspielverhalten des Videos	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

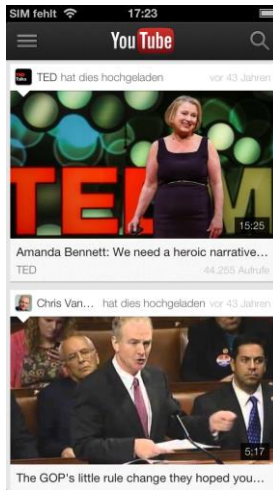
* Warum waren Sie mit dem **Abspielverhalten des Videos nicht vollkommen zufrieden**?

Abbildung 130: Spezifischer Fragenbogenteil zum Szenario Youtube

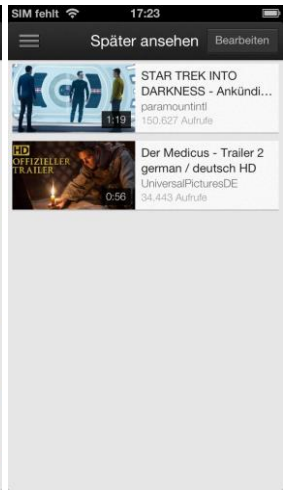
Die Einblendung des Kommentarfeldes zum Abspielverhalten erfolgte, wenn der Proband zum „Abspielverhalten des Videos“ eine andere Bewertung als „vollkommen zufrieden“ wählte.



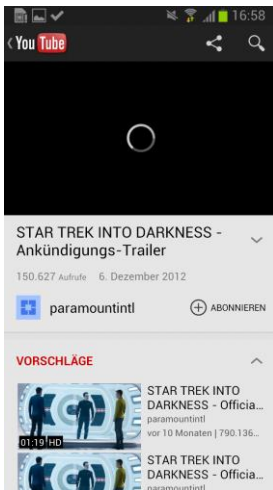
1.1) Laden der Startansicht mit Vorschau-Bildchen



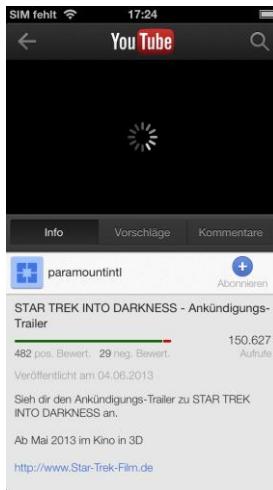
1.2) Startansicht geladen



2) Wahl eines der Videos aus vorausgewählter Liste im eigenen Youtube-Profil



3.a) Android
 Videoseite, Vorpuffern des Videos, Wiedergabe startet nachfolgend automatisch



3.b) iOS
 Videoseite, Vorpuffern des Videos, Wiedergabe startet nachfolgend automatisch

iOS (Android vergleichbar)

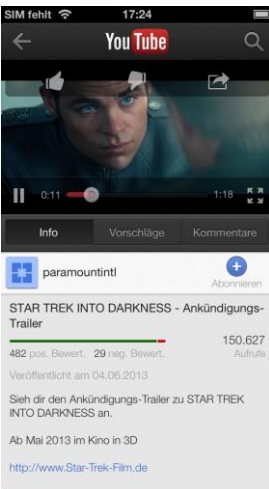
Darstellung Smartphone, Tablet sehr ähnlich

bei 3) und 4) Wechsel zum Querformat und damit Vollbildansicht möglich (siehe nachfolgende Teile)

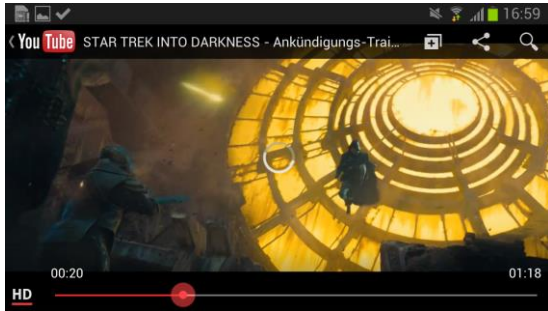
Abbildung 131: Zusammenstellung zur Nutzersicht im Szenario Youtube (Fortsetzung →)
 Screenshots aus der zugehörigen App



4.a) Android Hochformat



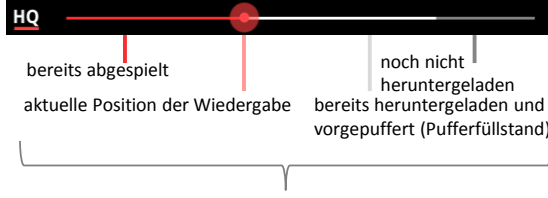
4.b) iOS Hochformat



4.c) Android quer: Qualitätsstufe und Button zum manuellen Wechsel, hier: Video in HD verfügbar und HD aktiv (Galaxy SIII)



4.d) hier: Video in HQ verfügbar und HQ aktiv (Galaxy Ace 2)



einzelne Komponenten der Zeitleiste zur Abschätzung des aktuellen Füllstands des Puffers



4.e) iOS quer (iPhone): keine Anzeige der Qualitätsstufe

Fortsetzung der Abbildung 131

4.a und 4b: Videoseite, Sicht während der Wiedergabe des bereits gestarteten Videos mit eingebledeten Controls (nach Berührung des Bildschirms, sonst ohne Einblendung)

4.c und 4.e: Wiedergabe gerade unterbrochen (Stall), da Puffer entleert, Apps beim Nachpuffern, erkennbar am Ladekreis in der Bildmitte, Ausschnitt Ladebalken HQ zur Verdeutlichung

A.4.12 Nachbefragung

Nachbefragung

*

Wie **alt** sind Sie?

In dieses Feld dürfen nur Zahlen eingegeben werden.

* Bitte geben Sie Ihr **Geschlecht** an.

weiblich männlich

* Was ist Ihr **höchster Bildungsabschluss**?
Bitte wählen Sie eine der folgenden Antworten:

Schüler / ohne Schulabschluss
 Hauptschulabschluss
 Mittlere Reife
 Fachhochschulreife / Abitur
 Hochschulabschluss (erworben an einer Universität, (Fach-)Hochschule, etc.)

*

Welche der folgenden Aussagen trifft auf Ihren **derzeitigen beruflichen Status bzw. Ihre Ausbildung** zu?

Bitte wählen Sie eine der folgenden Antworten:

nicht berufstätig
 teilweise berufstätig
 voll berufstätig
 in Berufsausbildung/Wehrpflicht/Zivildienst
 in Hochschulausbildung (Student)
 Sonstiges: _____

* Welche der folgenden **Geräte** nutzen Sie im **privaten Gebrauch**?
Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

Smartphone
 Tablet-PC

Abbildung 132: Nachbefragung im Fragebogen Teil 1

allgemeine Angaben zur Person und Abfrage zur Nutzung mobiler Geräte

Die Auswahl zur Nutzung eines entsprechenden mobilen Gerätes im privaten Gebrauch war Voraussetzung für die Einblendung des Teils 2 der Nachbefragung für die jeweilige Geräteklasse.

* Welche der folgenden **Geräte** nutzen Sie im **privaten Gebrauch**?

Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

- Smartphone
- Tablet-PC

* Über welches Betriebssystem verfügt Ihr **Smartphone**?

Bitte wählen Sie eine der folgenden Antworten:

- iOS
- Android
- sonstiges (z.B. Symbian, Windows)
- weiß nicht

* Welches **Smartphone-Modell** nutzen Sie?

* Haben Sie bei Ihrem Mobilfunkanbieter auch einen **Datenvertrag** oder eine **Datenflatrate** abgeschlossen (Prepaid-Karte oder Mobilfunkvertrag), mit der Sie unterwegs das **mobile Internet** (kein W-LAN) auf dem **Smartphone** nutzen?

- Ja
- Nein

* Über welches Betriebssystem verfügt Ihr **Tablet**?

Bitte wählen Sie eine der folgenden Antworten:

- iOS
- Android
- sonstiges (z.B. Symbian, Windows)
- weiß nicht

* Welches **Tablet-Modell** nutzen Sie?

* Haben Sie bei Ihrem Mobilfunkanbieter auch einen **Datenvertrag** oder eine **Datenflatrate** abgeschlossen (Prepaid-Karte oder Mobilfunkvertrag), mit der Sie unterwegs das **mobile Internet** (kein W-LAN) auf dem **Tablet** nutzen?

- Ja
- Nein

Abbildung 133: Nachbefragung im Fragebogen Teil 2 Angaben zum genutzten mobilen Gerät

Die Angabe „Ja“ zur Nutzung des mobilen Internets auf dem jeweiligen Gerät war Voraussetzung für die Einblendung des Teils 3 der Nachbefragung für die jeweilige Geräteklasse.

* Seit wann nutzen Sie das mobile Internet auf dem **Smartphone**?
 Bitte wählen Sie eine der folgenden Antworten:

- seit bis zu 3 Monaten
- bis zu 6 Monaten
- bis zu 12 Monaten
- mehr als 12 Monate

* Welches Netz nutzen Sie für das mobile Internet auf dem **Smartphone**?
 Bitte wählen Sie eine der folgenden Antworten:

- E-Plus
- O2
- T-Mobile
- Vodafone
- weiß nicht

* Welchen Mobilfunkanbieter nutzen Sie für das mobile Internet auf ihrem **Smartphone**?

* Welche Art von Mobilfunkkarte nutzen Sie für Ihr **Smartphone**?
 Bitte wählen Sie eine der folgenden Antworten:

- Wiederaufladbare Karte (sogenannte Prepaid Karte ohne Vertragsbindung und ohne Grundgebühr)
- Mobilfunkvertrag mit Mindestlaufzeit und monatlicher Rechnung

Wie häufig nutzen Sie das mobile Internet über Ihren Mobilfunkanbieter für die folgenden Anwendungen auf ihrem Smartphone?

	mehrmals täglich	mehrmals pro Woche	mehrmals pro Monat	seltener als einmal im Monat	nutze ich gar nicht
E-Mails abrufen und beantworten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Messaging über Messaging App (z.B. WhatsApp)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ortsbezogene Dienste (Location Based Services, z.B. lokale Information, Routenplanung über GoogleMaps)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nutzung sozialer Netzwerke (z.B. Facebook)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fotos teilen (z.B. Picasa)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online-Spiele für Smartphones	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Internettelefonie (Voice over IP, z.B. Skype)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nachrichten lesen (z.B. Spiegel Online)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recherche, Wissen, Informationen nachschlagen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online-Shopping (z.B. über Amazon)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Online-Banking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Videostreaming, also das Ansehen von Filmen direkt online (z.B. YouTube)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Musikstreaming, also das Hören von Musik direkt online (z.B. MTV-Music)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Dokumente aus der Cloud ansehen und bearbeiten (z.B. iCloud, Google Drive)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Applikationen (Apps) Download und Update	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 134: Nachbefragung im Fragebogen Teil 3

Erfassung von Nutzungserfahrung, Netzanbieter und Nutzungshäufigkeit ausgewählter mobiler Dienste mit der jeweiligen Geräteklasse. Bei der Angabe von „weiß nicht“ zum genutzten Netz wurde der Mobilfunkanbieter per Freitextfeld erfasst.

A.5 Rohergebnisse

A.5.1 Übersichtstafeln

*„Stellen Sie multivariate Daten dar – zeigen Sie also mehr als eine oder zwei Variablen“
nach Edward Tufte²²⁹*

Die von den Probanden generierten Rohergebnisse werden in Form von Übersichtstafeln angegeben, die den Datensatz nach der Vorvalidierung darstellen. Aus ihnen ist die Generierung einer Urliste möglich.

Nicht erschrecken, die nachfolgenden Abbildungen sind keine DNA-Bandenmuster, sondern ein eigenes Konzept, in dem alle maßgeblichen Variablen mit hoher Codierungsdichte dargestellt werden:

- Testslot: Nummer (woraus Datum, Startzeit und Panel der Probanden abzuleiten sind)
- Testplatz: Belegung und berücksichtigte Testläufe
- mit dem Testplatz festgelegt: Endgerät, Endgerätyp (Format), OS (mobiles Betriebssystem)
- Testlauf durch Testslot (Spalte) und Testplatz (Zeilenblock) als Spaltenabschnitt innerhalb der Matrix bestimmt
- jedem Testlauf ist ein Proband zugeordnet (Spaltenabschnitt innerhalb des Zeilenblocks)
- Proband
 - Alter und Geschlecht (Farbkodierung)
 - privat genutztes Netz
- jedem Testlauf sind 12 Aufgabenausführungen (Zellen) zugeordnet

²²⁹ mit Verweis auf die Grundprinzipien für die analytische Informationsgestaltung, entnommen aus [357] S. 420

- jede Aufgabenausführung beinhaltet
 - Aufgabennummer und Reihenfolge der Aufgaben implizit durch die Position innerhalb der Spalte
 - Aufgabe mit Szenario und Netzparameterstufe durch Angabe des Pre-sets
 - Aufgabenvariante (A, B) als Unterstreichung
 - Bearbeitungserfolg als Schriftgrad
 - Bewertung Gesamtzufriedenheit als Farbkodierung gemäß der zugeordneten Skalenwerte
 - optional: Validität der Aufgabenausführung bzw. Bewertung: durchgestrichen, wenn invalid (hier nicht dargestellt, da zum Zweck der Rohergebnisdarstellung vor der weiteren Validierung)

Für unvollständige Aufgabenausführungen liegt jeweils nur die Bewertung der Gesamtzufriedenheit vor. Eine dritte Teilbewertung wurde nur für die Szenarien MTV-Music und Spiegel.de erfasst, entsprechen bleiben die Felder für die anderen Szenarien ohne entsprechende Teilbewertung leer.

4	iPad 4 iOS Tablet															5	iPhone 5 iOS Smartphone															6	Sill Android smartphone																																							
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1		15	14	13	12	11	10	9	8	7	6	5	4	3	2	1		15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1										
1	Y40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	1	Y40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	1	Y40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	1	Y40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40						
2	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	2	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	2	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	2	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	
3	D10	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	3	D10	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	3	D10	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	3	D10	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
4	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	4	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	4	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	4	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
5	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	5	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	5	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	5	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
6	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	6	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	6	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	6	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
7	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	7	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	7	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	7	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
8	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	8	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	8	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	8	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
9	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	9	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	9	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	9	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
10	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	10	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	10	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	10	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
11	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	11	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	11	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	11	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40
12	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	12	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	12	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	12	G50	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40	G40

Fortsetzung der Abbildung 138

Text: Preset mit Kürzel zu Aufgabe und Parameterstufe
 D Drive 50 (hoch)
 F Facebook 40
 G Google Maps 30
 M MTV-Music 20
 S Spiegeldat 10 (niedrig)
 Y Youtube

Farbkodierung: Bewertung Gesamtzufriedenheit
 vollkommen zufrieden
 eher zufrieden
 weder zufrieden, noch unzufrieden
 vollkommen unzufrieden
 (abwertung 3 je Aufgabe →)

Schriftschnitt / -effekt: Erfolg der Aufgabenbearbeitung
 Aufgabe im ersten Anlauf gelöst
 Aufgabe nach mehreren Anläufen gelöst
 Aufgabe nicht begonnen oder nicht gelöst

Unterschiede: Aufgabenvariable
 ohne Variante A der Aufgabe
 Variable B der Aufgabe
 M Abspielverhalten des Musikstücks
 S Dauer zum vollständigen Laden der Fotos
 und Textelemente der Robocheck

A.5.2 Anzahl Bearbeitungen, Bearbeiter und Bearbeitungserfolg

Es erfolgte eine Überprüfung und Rekodierung der Angaben „Sonstiges“ mit Interpretation des zugehörigen Freitextes um die Angaben einer der vorgegebenen Kategorien zuzuordnen, siehe Tabelle 73. Dabei interessierten vor allem solche Angaben, in denen die Probanden Probleme mit der Internetverbindung beschreiben, allerdings ohne dies als eigentlichen Grund angegeben zu haben.

Drive: Von 11 Angaben „Sonstiges“ konnten mit der Internetverbindung in Zusammenhang stehend identifiziert werden:

- „ja, Aufgabe nach mehreren Versuchen gelöst“
 - mit Internetverbindung in Zusammenhang, aber vom Probanden so nicht angegeben: 2
 - eventuell mit der Internetverbindung in Zusammenhang, aber nicht so angegeben: 1
- „nein, Aufgabe nicht beendet/abgebrochen“
 - eventuell mit der Internetverbindung in Zusammenhang, aber nicht so angegeben: 2

Facebook: Von den 16 Angaben „Sonstiges“ ohne weiteren Grund konnten 5 im Zusammenhang mit der Internetverbindung identifiziert werden, die die Probanden so nicht direkt angegeben haben.

Google Maps: Von den drei Angaben „Sonstiges“ ohne weiteren Grund konnten keine mit der Internetverbindung als Problemursache in Zusammenhang gebracht werden. Ein Kommentar unter „Sonstiges“ mit Problemangabe „Internetverbindung“ deutet eher auf ein App-Problem (Absturz) hin.

MTV-Music: Von den sechs Angaben „Sonstiges“ ohne weiteren Grund konnte eine mit der Internetverbindung als Problemursache im Sinne der beabsichtigten Parametrisierung in Zusammenhang gebracht werden. Zwei weitere Angaben „Sonstiges“ beziehen sich indirekt auf die Internetverbindung bzw. Probleme bei der Suche, wahrscheinlich durch Server-Probleme beim Anbieter.

Spiegel.de: Von den 8 Angaben „Sonstiges“ ohne weiteren Grund konnte keine weitere mit der Internetverbindung als Problemursache in Zusammenhang gebracht werden (nur Bedienprobleme bzw. im Zusammenhang mit der Aufgabenstellung).

Youtube: Von den drei Angaben „Sonstiges“ ohne weiteren Grund konnte keine weitere mit der Internetverbindung als Problemursache in Zusammenhang gebracht werden.

Tabelle 73: Übersicht zum Bearbeitungserfolg, Problemen, Anzahl Bearbeitungen, Bearbeitern und validen Bewertungen

Bearbeitungserfolg und genannte Probleme	Dienst																													
	Drive				Facebook				Google Maps				MTV Music				Spiegel.de				Youtube									
	D10	D20	D30	D40	D50	F10	F20	F30	F40	F50	G10	G20	G30	G40	G50	M10	M20	M30	M40	M50	S10	S20	S30	S40	S50	Y10	Y20	Y30	Y40	Y50
ja, Aufgabe beim ersten Versuch gelöst																														
Anzahl Bearbeitungen	78	66	77	73	78	65	79	76	80	72	68	83	78	82	80	67	84	66	79	86	79	84	77	83	78	86	68	85	84	80
ja, Aufgabe nach mehreren Versuchen gelöst																														
Anzahl Bearbeitungen	11	8	11	4	6	13	6	5	2	7	6	6	4	1	3	7	2	3	7	6	6	4	4	1	2	5	1	1	4	6
Verbindung	3	2				10	2	2	1		4	5	1	1		5	1	1			3					2				1
App-Bedienung	5	8	4	3	2	3	4	2	1	6	2	5	4	2	2	2	1	3	4		3	2	3	1	2	1	1		2	4
Gerätebedienung	3	3	1								1	1	1				1	2	2	1	1	1				1	2	1		1
Aufgabenstellung	1	1	2	1	3		1	2			1	1	1	1			1	2			2					2				
Sonstiges	2	4	3			4	4	1	1	1	1	1	1	1	1	3	1	2	3	1	1	3	1	1	1	1	1	1	1	1
nein, Aufgabe nicht beendet/abgebrochen																														
Anzahl Bearbeitungen	1	1	2	2	2	11	1	2	1		6	1	1			4	1	4	1	2		1	1							
Verbindung	1	1	1			11	1	1			4	1	1			4	1	1												
App-Bedienung	1	1	1	1	2						1						2	1												
Gerätebedienung	1					1					1								1											
Aufgabenstellung	1						1				1						1	1												
Sonstiges	1	1				4	1	1			1	1	1			1	1	1	1					1	1					
nein, Aufgabe konnte gar nicht erst gestartet werden																														
Anzahl Bearbeitungen											1					1														
Verbindung																1														
App-Bedienung																														
Gerätebedienung											1																			
Aufgabenstellung																														
Sonstiges																1														
Gesamt: Anzahl Bearbeitungen	90	75	90	79	86	89	86	83	83	79	80	91	82	84	83	78	87	74	87	94	85	88	82	85	80	91	69	86	88	86
Gesamt: Anzahl mit Problemen	12	9	13	6	8	24	7	7	3	7	12	8	4	2	3	11	3	8	8	8	6	4	5	2	2	5	1	1	4	6
Gesamt: Verbindung	3	1	2	1		21	3	3	1		8					9	2	3			3					2				1
Gesamt: App-Bedienung	6	8	5	4	4	3	4	2	1	6	3	5	4	2	2	2	1	2	4	3	3	2	3	1	2	1	1		2	4
Gesamt: Gerätebedienung	3	4	1			3	4	1			2	1	1			1	2	3	1		2	1				1	2	1		1
Gesamt: Aufgabenstellung	1	1	3	1	3	1	1	2			2	1	1	1		2	2	1			2					2				
Gesamt: Sonstiges	2	1	4	1	3	8	4	2	1	1	2	1	1	1	1	4	1	3	2	4	1	3	1	2	1	1	1	1		1
Teilbewertung 1 valid von Anzahl Bearbeitern	82	68	85	73	81	89	86	82	82	79	78	89	82	83	83	77	86	71	87	93	85	88	81	85	80	91	69	86	88	86
Teilbewertung 2 valid von Anzahl Bearbeitern	80	67	79	65	79	79	85	79	82	76	77	83	80	77	80	77	81	69	83	83	82	84	75	83	77	81	68	82	87	85
Teilbewertung 3 valid von Anzahl Bearbeitern	84	71	86	76	81	89	86	82	82	79	78	89	82	83	83	77	86	71	87	94	85	88	81	85	80	91	69	85	87	84
Gesamtzufriedenheit valid von Anzahl Bearbeitern	82	69	79	66	80	79	85	79	82	76	77	83	80	77	80	77	81	69	83	83	82	84	75	83	77	81	68	81	86	83
Gesamtzufriedenheit valid von Anzahl Bearbeitern	78	64	81	70	77	89	86	82	82	79	78	89	82	83	83	77	86	71	87	93	85	88	81	85	80	91	69	86	88	86
	76	63	75	63	76	79	85	79	82	76	77	83	80	77	80	77	81	69	83	83	82	84	75	83	77	81	68	82	87	85

Die Angaben erfolgen als absolute Häufigkeit. Bei den Angaben der Probleme sind Mehrfachnennungen möglich, Werte "Sonstiges" rekodiert.

A.5.3 Zusatzanmerkungen zur Validierung

Die Gegenüberstellung der Anzahl Aufgabenausführungen (insgesamt) und valide (Teil-) Bewertungen ist der Tabelle 73 zu entnehmen. An dieser Stelle werden, soweit notwendig, ergänzende Anmerkungen aus der Validierung der Rohergebnisse vermerkt.

MTV-Music: Einige der Angaben „Internetverbindung“, die einen Ausfall der Suchfunktion beschreiben, wurden durch (kurzzeitige) Server-Probleme beim Anbieter verursacht und sind damit als invalid markiert (externer Störeinfluss), da dies nicht Teil der gezielten QoS-Variation ist. In der Validierung wurde für diese Aufgabenausführungen, die durch externe Faktoren verzerrt worden, auch nicht von einer validen Bewertung der Gesamtzufriedenheit ausgegangen.

Youtube: Für HD bei Y30 sind 8 unerwartete Bewertungen zum Abspielverhalten zu finden, die von „vollkommen zufrieden“ abweichen. Eine Bewertung steht nicht im Zusammenhang mit der QoS-Parametrisierung (invalid):

- „angezeigtes Format etwas zu klein; Video etwas zu dunkel“ (normale Darstellung des Videos, kein Funktionseffekt beschrieben)

Die restlichen 7 abweichenden Bewertungen sind durch Unterbrechungen begründet. Dies war ursprünglich nicht geplant, die maximale Download-Datenrate bei Y30 ist aber recht knapp an der notwendigen Inhaltsdatenrate der ausgewählten Videos. Schon geringe kurze äußere Einflüsse (Youtube-Server, Netz, ...) ändern den relevanten Nutzerstimulus der Unterbrechungsfreiheit und damit die Zufriedenheit deutlich.

Für HD bei Y50 sind 2 unerwartete Bewertungen zum Abspielverhalten zu finden, die von „vollkommen zufrieden“ abweichen. Beide Bewertungen stehen aber nicht im Zusammenhang mit der beabsichtigten Variation der Nutzerstimuli zum Abspielverhalten, da von den Probanden wohl anders verstanden und begründet (invalid):

- „am Anfang erschienen Steuerelemente, nachdem Smartphone waagrecht gedreht wurde“ (normales Verhalten der Youtube-App)

- „dunkle Flächen waren schlecht zu sehen, Helligkeit des Displays könnte höher sein, Anzeigequalität sehr gut, Display war schmutzig“ (Geräteeinstellungen und normale Fettfinger durch die Nutzung durch den Probanden selbst, da in der Vorbereitung die Geräte jeweils per Mikrofasertuch gereinigt wurden)

Für HQ sind insgesamt 3 unerwartete Bewertungen zum Abspielverhalten zu finden, die von „vollkommen zufrieden“ abweichen. Zwei der Bewertungen beziehen sich auf unerwartete Unterbrechungen der Wiedergabe, eine ist ohne Zusammenhang zu den beabsichtigten Nutzerstimuli zu sehen (invalid):

- „Der Pause-Button stört“ (eine kurze Einblendung der Kontrollelemente ins Video ist normal)

A.6 Zusatzerkenntnisse

Nachfolgend sind einige Auszüge aus den Nutzerkommentaren zu speziellen Aspekten aufgeführt.

A.6.1 Netzwahrnehmung durch die Probanden

Wie wird die genutzte Zugangstechnologie im Testbed von den Probanden wahrgenommen? Die Angaben zu Endgerät und Dienst erfolgen jeweils in Klammern hinter den Kommentaren.

8 x **KONTRA** (Wahrnehmung als WLAN):

- „In der Anwendung über 3G hätte ich diese Vorgänge abgebrochen, da die lange Ladezeit auf eine sehr große Datengröße hindeutete. Solche großen Dateien lade ich nur über WLAN.“ (iPad, Drive)
- „Der Upload und der Download haben viel zu lange gedauert und das trotz WLAN.“ (iPhone, Drive)
- „zügig und reibungslos, aber bei w-lan auch kein wunder“ (iPhone, Facebook)
- „beim ersten versuch schlug das hochladen fehl, was ich mit w-lan nicht erwartet hatte“ (iPhone, Facebook)
- „dauer des kartenaufbaus beim zoomen zu lang trotz wlan verbindung.“ (iPhone, Google Maps)

- „nach mehreren Minuten immernoch kein Suchergebnis, WLAN ist aktiviert, daran dürfte es also nicht liegen“ (Android Tab 2, MTV-Music)
- „für wlan zu langsam“ (iPhone, Spiegel.de)
- „Trotz Voller Wlan Verbindung ‚gepuffert.‘“ (Android SIII, Youtube)

5 x **PRO** (Wahrnehmung als mobiles Internet):

- „Das mobile Internet funktionierte einwandfrei, was sich jedoch immer ändert, kommt man in Gebiete, in denen die Verbindung nicht so gut ist.“ (Android Ace 2, Google Maps)
- „Es dauerte schon eine ganze Weile bis das Musikstück geladen hat. Ich denke dass die Anwendung nur im WLAN Betrieb Sinn macht und nicht z. B. wenn man mit dem Zug unterwegs ist.“ (Android Tab 2, MTV-Music)
- „[...] Ich habe leider keine Auskunft über die verwendete internetleitung“ (Android Tab 2, Youtube)
- „Geschwindigkeit des mobilen Internet mäßig schnell“ (Android SIII, Google Maps)
- „langsame Geschwindigkeit des mobilen Internet -> längere Ladezeit des Musikstücks“ (Android SIII, MTV-Music)

A.6.2 Wahrnehmung der Qualität des Inhalts

Werden die Aspekte der Inhaltsqualität wahrgenommen bzw. bei der Bewertung honoriert?

MTV-Music:

- „Die Bewertung zum Abspielverhalten ist nur schwer vorzunehmen, da die Qualität meines Erachtens nach nur über entsprechende Lautsprecher oder Kopfhörer möglich ist.“
- „Tonqualität des Musikstücks war nicht wirklich überzeugend.“
- „Nach dem kurzen Fehlverhalten beim Liedstart spielt es ohne Probleme das Lied [...]“
- „[...] und es spielt ohne stocken in guter Qualität.“
- „Klang war nicht besonders beeindruckend“
- „Die Trommeln im Hintergrund haben sehr gedöhnt.“

- „spezielles soundverhalten des Liedes fehlt. (alle Instrumente und Stimmen klingen gleich laut) Es ist eine spezielle Einstellung erst notwendig.“

Youtube:

- „[...] wenn viel Dunkle Bereiche im Video kamen und ich hatte das Gefühl als wäre das Bild ein bisschen pixelig bei schnellen Bildfolgen.“ (Ace 2, HQ)
- „[...] die Qualität war trotzdem bestens!“ (Android Tab 2, HD)
- „Videos sollten bei einer Videoapp richtig laufen, auch wenn es dabei die Verminderung der Qualität zur Folge hat.“ (iPad, HD)
- „Video lief fließend, kein stocken und ist geprägt von guter Qualität“ (Android SIII, HD)
- „Das Video musste während des Abspielens 2 mal nachgeladen werden (Qualität HD) Ich habe leider keine Auskunft über die verwendete internetleitung“ (Android Tab 2, HD)
- „Sehr kurze Ladezeiten, leichte Bedienung, super Bildqualität.“ (iPhone, HD)
- „wirklich sehr schnell und top qualität des videos“ (Android Tab 2, HD)
- „[...] super Bildqualität“ (Android SIII, HD)
- „Video musste 4 mal nachgeladen werden; ca doppelte abspieldauer (Qualität HD)“ (Android Tab 2, HD)
- „Es musste Zwischenladen, damit wurde das ansehen unterbrochen und ich fühlte mich leicht genervt. Die Qualität an sich war aber sehr gut.“ (iPad, HD)
- „[...] Anzeigequalität sehr gut [...]“ (Android Tab 2, HD)
- „Die Bildqualität war zwar top, es gab allerdings 6 Ladepausen von jeweils mehreren Sekunden. Also kein flüssiges Abspielverhalten.“ (Android SIII, HD)

A.7 Credits und Disclaimer

Genutzte Ressourcen

Einige der Inhalte, Programme und verwendeten Datenquellen wünschen bzw. erfordern eine explizite oder gesonderte Nennung.

Zur Erstellung der Illustrationen und in der Web-GUI wurde teilweise das FatCow Icon Pack²³⁰ (angepasst) genutzt (Creative Commons Attribution 3.0 License).

Cisco attribution: Cisco Visual Networking Index (VNI) Global Mobile Data Forecast genutzt.

Geschützte Bezeichnungen, Warenzeichen und Wortmarken

Weitere genannte, hier aber nicht aufgeführte Warenzeichen und Handelsmarken etc. sind Eigentum ihrer jeweiligen Inhaber, auch wenn diese an entsprechender Stelle innerhalb des Textes ggf. nicht als solche gekennzeichnet sind.

Screenshots und Illustrationen

Die Screenshots zu den einzelnen mobilen Diensten, von Anwendungen und darin gezeigten Inhalten dienen lediglich zu Dokumentations- bzw. Belegzwecken. Die Bildrechte und sonstigen Rechte wie Urheberrechte etc. der darauf gezeigten Inhalte liegen bei den jeweiligen Rechteinhabern.

Die symbolische Darstellung einzelner Dienste, Anwendungen (z. B. per App-Icon) oder Endgeräte dient im Allgemeinen jeweils nur zur Illustration, stellvertretend für die Gesamtheit aller Dienste, zugehöriger Apps, mobiler Plattformen und Endgeräte.

²³⁰ <http://www.fatcow.com/free-icons> abgerufen am 19.05.2014

Dissertationen der Medieninformatik

- (1) Kürsten, Jens (2012)
A Generic Approach to Component-Level Evaluation in Information Retrieval
ISBN 978-3-941003-68-2
Volltext: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-96344>

- (2) Berger, Arne (2014)
Prototypen im Interaktionsdesign : Klassifizierung der Dimensionen von Entwurfsartefakten zur Optimierung der Kooperation von Design und Informatik
ISBN 978-3-944640-00-6
Volltext: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-126344>

- (3) Ritter, Marc (2014)
Optimierung von Algorithmen zur Videoanalyse : Ein Analyseframework für die Anforderungen lokaler Fernsehsender
ISBN 978-3-944640-09-9
Volltext: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-133517>

- (4) Kurze, Albrecht (2016)
Modellierung des QoS-QoE-Zusammenhangs für mobile Dienste und empirische Bestimmung in einem Netzemulations-Testbed
ISBN 978-3-944640-60-0
Volltext: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-195066>