

**Gestaltungsaspekte multimodaler
Interaktion im Fahrzeug
Ein Beitrag aus
ingenieurpsychologischer Perspektive**

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Philosophischen Fakultät II
(Psychologie, Pädagogik und Sportwissenschaft)
der Universität Regensburg

vorgelegt von

Roman Vilimek
aus München

2007

Erstgutachter: Prof. Dr. A. Zimmer
Zweitgutachter: Prof. Dr. K.-H. Bäuml
Obergutachterin: Prof. Dr. M. Hammerl



Vorwort

Die vorliegende Arbeit entstand in den Jahren 2003 bis 2006 im Rahmen einer Kooperation zwischen dem Fachzentrum für User Interface Design der Siemens AG, Corporate Technology (CT IC 7), und dem Institut für Experimentelle Psychologie der Universität Regensburg.

Mein Dank gilt in erster Linie Prof. Dr. Alf Zimmer für die Betreuung dieser Arbeit, insbesondere für seine fachlich kompetente Unterstützung und für viele anregende Diskussionen. Prof. Dr. Karl-Heinz Bäuml danke ich für seine Bereitschaft, die Arbeit zu begutachten.

Besonders bedanken möchte ich mich bei Dr. Thomas Hempel, der das Forschungsprojekt als Betreuer bei der Siemens AG in allen Phasen begleitet hat. Nicht nur sein fachlicher Beistand war exzellent, auch seine persönliche Unterstützung hat maßgeblich zum Gelingen der Arbeit beigetragen. Des Weiteren bedanke ich mich bei Dr. Robert Rauschenberger von Siemens Corporate Research in Princeton für viele inhaltlich äußerst wertvolle Diskussionen. Ferner gilt mein Dank dem ehemaligen Fachzentrumsleiter Dr. Stefan Schoen sowie dem aktuellen Fachzentrumsleiter Dr. Martin Scheurer, die meine Arbeit stets unterstützt haben, sowie den Kolleginnen und Kollegen von CT IC 7.

Ein ganz besonderer Dank geht an meine Familie.

*„Car on peut bien concevoir
qu'une machine soit tellement faite
qu'elle profère des paroles [...];
mais non pas qu'elle les arrange diversement,
pour répondre au sens de tout ce qui se dira en sa présence“*

„Denn man kann sich wohl denken,
dass eine Maschine so eingerichtet ist,
dass sie Worte von sich gibt [...],
keineswegs aber, dass sie die Worte so mannigfaltig anordnet,
um auf den Sinn alles dessen zu antworten,
was man in ihrer Gegenwart sagen wird“

Descartes, Discours de la méthode

Inhalt

1	Einleitung	1
2	Multimodale Mensch-Maschine-Interaktion	4
2.1	Begriffliche Klärungen	5
2.1.1	Grundlegende Begriffe und Konzepte	5
2.1.2	Modalität	8
2.1.2.1	Wahrnehmungsmodalitäten	8
2.1.2.2	Aktionsmodalitäten	10
2.1.3	Multimodale Schnittstellen	11
2.1.3.1	Definition Multimodalität	12
2.1.3.2	Kennzeichen multimodaler Systeme	15
2.2	Ingenieurpsychologische Konzepte	18
2.2.1	Belastung und Beanspruchung	18
2.2.2	Informationsverarbeitung und Aufmerksamkeit	20
2.2.3	Allgemeine Fehlerklassifikation	25
2.3	Grundlagen multimodaler Mensch-Maschine-Interaktion	27
2.3.1	Einsatzgebiete und Entwicklung multimodaler Systeme	28
2.3.2	Visuelles Interface	30
2.3.2.1	Visuelle Wahrnehmung und visuelle Systemausgaben	30
2.3.2.2	Aspekte der Gestaltung von Gestik- und Blickeingabe	33
2.3.3	Auditives Interface	35
2.3.3.1	Auditive Wahrnehmung und akustische Systemausgaben	36
2.3.3.2	Aspekte der Gestaltung von Spracheingabe	39
2.3.4	Haptisches Interface	42
2.3.4.1	Haptische Wahrnehmung und haptische Systemausgaben	43
2.3.4.2	Aspekte der Gestaltung von manueller Eingabe	48
2.3.5	Vorteile und Einschränkungen multimodaler MMI	50

2.3.6	Experimentelle Befunde zur Verarbeitung multimodaler Signale	52
2.3.7	Experimentelle Befunde zu multimodaler Eingabe.....	55
2.3.8	Gestaltungsrichtlinien für multimodale Systeme	59
3	Anwendungsfall Fahrzeugführung	61
3.1	Theoretische Modelle der Fahrzeugführung	61
3.1.1	Die Drei-Ebenen-Modelle der Fahrzeugführung.....	62
3.1.2	Das Handlungsregulationsmodell.....	64
3.2	Handlungssituation.....	65
3.2.1	Primär-, Sekundär- und Tertiäraufgaben	65
3.2.2	Ablenkung und Abwendung.....	66
3.2.3	Fehlhandlungen im Fahrzeug	68
3.3	Bedienkonzepte für Bordsysteme.....	71
4	Untersuchungsabschnitt 1: Aspekte der Ausgabe im multimodalen Umfeld ...	74
4.1	Multimodales Feedback für Touchpads zur zentralen Bedienung von Bordsystemen	75
4.1.1	Problemstellung und Ausgangspunkt der Experimente.....	75
4.1.1.1	Touchpads als Ersatz für Dreh-Drücksteller	76
4.1.1.2	Multimodale Ausgabe für Touchpads und Überblick über die Experimente	77
4.1.2	Experiment 1: Evaluation mit und ohne Fahraufgabe	79
4.1.2.1	Methode.....	79
4.1.2.2	Hypothesen.....	94
4.1.2.3	Ergebnisse	96
4.1.2.4	Diskussion	108
4.1.2.5	Schlussfolgerungen	117
4.1.3	Experiment 2: Evaluation der Blindbedienbarkeit	118
4.1.3.1	Methode.....	118
4.1.3.2	Hypothesen.....	120
4.1.3.3	Ergebnisse	121
4.1.3.4	Diskussion	122
4.1.3.5	Schlussfolgerungen	124
4.1.4	Zusammenfassende Diskussion und Gestaltungsempfehlungen	124
4.2	Sprachliche vs. nicht-sprachliche akustische Ausgaben	130
4.2.1	Problemstellung und Ausgangspunkt des Experiments.....	130
4.2.1.1	Arbeitsgedächtnis und Fahrzeugführung	131
4.2.1.2	Akustische Ausgaben und das Arbeitsgedächtnis.....	133

4.2.2	Experiment 3: Wirkung sprachlicher und nicht-sprachlicher auditiv dargebotener Information auf das Arbeitsgedächtnis	136
4.2.2.1	Methode	136
4.2.2.2	Hypothesen	143
4.2.2.3	Ergebnisse	144
4.2.2.4	Diskussion	150
4.2.3	Zusammenfassende Diskussion und Gestaltungsempfehlungen	155
5	Untersuchungsabschnitt 2: Aspekte der Eingabe im multimodalen Umfeld ..	158
5.1	Multimodales Talk-Ahead bei Spracheingabe	160
5.1.1	Problemstellung und Ausgangspunkte des Experiments	160
5.1.2	Experiment 4: Multimodales Talk-Ahead	165
5.1.2.1	Methode	165
5.1.2.2	Hypothesen	173
5.1.2.3	Ergebnisse	175
5.1.2.4	Diskussion	181
5.1.3	Zusammenfassende Diskussion und Gestaltungsempfehlungen	185
5.2	Verbale versus ikonografische Funktionsdarstellung	186
5.2.1	Problemstellung	186
5.2.2	Grundlagen der Verwendung von Icons und verbalen Kennzeichnungen	189
5.2.2.1	Experimente zu visueller Suche	190
5.2.2.2	Untersuchungen mit simulierten Systemen	192
5.2.2.3	Untersuchungen auf der Ebene einzelner Eingaben	195
5.2.3	Ausgangspunkte der Experimente	198
5.2.4	Experiment 5: Zufällige Itemanordnung	202
5.2.4.1	Methode	202
5.2.4.2	Hypothesen	208
5.2.4.3	Ergebnisse	208
5.2.4.4	Diskussion	212
5.2.4.5	Schlussfolgerungen	216
5.2.5	Experiment 6: Ortsfeste Itemanordnung	217
5.2.5.1	Methode	218
5.2.5.2	Hypothesen	219
5.2.5.3	Ergebnisse	219
5.2.5.4	Diskussion	224
5.2.5.5	Schlussfolgerungen	226
5.2.6	Experiment 7: Ortsfeste und inhaltlich gruppierte Items	228
5.2.6.1	Methode	229

VIII Inhalt

5.2.6.2	Hypothesen.....	230
5.2.6.3	Ergebnisse	230
5.2.6.4	Diskussion	236
5.2.6.5	Schlussfolgerungen	239
5.2.7	Experiment 8: Visuelle Suche bei Kurzdarbietung	239
5.2.7.1	Methode.....	241
5.2.7.2	Hypothesen.....	243
5.2.7.3	Ergebnisse	244
5.2.7.4	Diskussion	246
5.2.8	Zusammenfassende Diskussion und Gestaltungsempfehlungen	247
6	Prozessmodell zur Entwicklung und Evaluation multimodaler Schnittstellen	252
Anhang	259
Literatur	264

Zusammenfassung

Die vorliegende Arbeit befasst sich aus ingenieurpsychologischer Perspektive mit dem Themenbereich multimodale Mensch-Maschine-Interaktion. Die Ziele der Arbeit bestehen darin, sowohl durch einen umfassenden theoretischen Rahmen als auch mit Hilfe experimentalpsychologischer Methoden dem Forschungsbedarf bei der Gestaltung multimodaler Schnittstellen für tertiäre Aufgaben im Kraftfahrzeug nachzukommen sowie in Form von Gestaltungsempfehlungen das Fundament zur Konstruktion eines multimodalen Bordsystems zu legen.

In der Erörterung der theoretischen Grundlagen werden über den Anwendungsfall der Fahrzeugführung hinaus die relevanten technologischen und humanwissenschaftlichen Einflussfaktoren dargelegt. Der experimentelle Teil der Arbeit besteht aus einer Serie von acht Experimenten. In den Untersuchungen zeigte sich, dass die Integration von haptischem und auditivem Feedback für eine Touchpad-basierte Steuerung des Bordsystems zu einer Verbesserung der Interaktionsqualität und der Fahrsicherheit durch Entlastung des visuellen Kanals beitragen kann (Experimente 1-2). Darüber hinaus erwies sich dieses multimodale Touchpad als sinnvolle Alternative gegenüber einem herkömmlichen Dreh-Drücksteller. Für die Gestaltung akustischer Systemausgaben wurde ermittelt, dass lange sprachliche Ausgaben mit kognitiven Vorgängen interferieren können. Deshalb müssen Sprachausgaben für ein solches System kurz gehalten bzw. durch nicht-sprachliche akustische Ausgaben ersetzt werden (Experiment 3). Die gestalterischen Randbedingungen dafür werden diskutiert. Die Sprach-eingabe des geplanten menübasierten multimodalen Systems soll auf dem *Say What You See* Eingabeprinzip beruhen, bei dem sich das Vokabular für die Sprachbedienung aus der Beschriftung von Menüpunkten oder Steuerelementen auf dem Bildschirm ergibt. Kognitionspsychologische Überlegungen legen nahe, dass die Eingabe-geschwindigkeit durch multimodales *Talk-Ahead* verbessert werden kann. In einem weiteren Experiment konnte gezeigt werden, dass multimodales *Talk-Ahead* die benötigte Aufgabenbearbeitungszeit und die visuelle Ablenkung bei *Say What You See* reduziert (Experiment 4). Durch die Festlegung auf *Say What You See* als Eingabestrategie ergibt sich die grundlegende Frage, inwiefern durch einen Verzicht auf bildhafte

Darstellungen und die hohe Textdichte auf der Anzeigeeinheit die visuelle Orientierung auf dem Bildschirm erschwert würde. Eine Reihe von vier Experimenten lieferte dazu aufschlussreiche Ergebnisse. Der bekannte Befund, dass Icons im Gegensatz zu Text bei zufälliger Anordnung zu deutlich geringeren Suchzeiten führen, konnte repliziert werden (Experiment 5). Diese Situation ist zu erwarten, wenn Nutzer auf dem Display eines ihnen nicht bekannten Systems nach bestimmten Funktionen suchen, deren Darstellung sie kennen (z.B. Standardsymbole). Der Vorteil für Icons blieb auch bestehen, wenn der Anzeigort eines Elements konstant und damit erlernbar war (Experiment 6). Sobald jedoch die Nutzer eines Systems zusätzliches Wissen über die inhaltliche Organisation der Elemente auf dem Display besitzen, verschwindet der Vorteil. Icons und textuell beschriftete Elemente führen dann zu nahezu identischen Suchzeiten (Experiment 7). In einem weiteren Experiment sollte als Worst Case Schätzung ermittelt werden, inwiefern der Vorteil von Icons bei zufälliger Anordnung für typische Blickdauern im Fahrzeug eine Bedeutung hat. Es zeigte sich, dass die Fehlerrate im kritischen Bereich einer Blickdauer von 200 bis 800 ms auf das Display bei Icons nur etwa halb so groß ist (Experiment 8). Abschließend wird ein Prozessmodell zur Entwicklung und Evaluation multimodaler Schnittstellen vorgestellt und exemplarisch aufgezeigt, wie durch das darin vorgeschlagene systematische Vorgehen die beschriebenen experimentellen Ergebnisse zur Realisierung eines multimodalen Prototyps genutzt werden können.

Durch die Verbindung von ingenieurpsychologischen Grundlagen mit der Diskussion technologischer Randbedingungen und experimentell gewonnener Verhaltensdaten schlägt die vorliegende Arbeit eine Brücke zwischen Technologie und Psychologie. Diese übergreifende Sichtweise trägt dazu bei, ein besseres Verständnis des komplexen Themenfelds multimodale Mensch-Maschine-Interaktion zu entwickeln.

1 Einleitung

Mit zunehmender Komplexität und Häufigkeit der Interaktion des Menschen mit technischen Systemen vergrößert sich die Notwendigkeit optimaler Schnittstellengestaltung. Die spezifischen Fähigkeiten der Nutzer werden dabei oft nicht ausreichend berücksichtigt. Buxton (1986) illustriert diesen Sachverhalt, indem er spekuliert, welche Annahmen Anthropologen in einer sehr fernen Zukunft über den heutigen Menschen treffen würden, wenn sie einen vollständig erhaltenen Computerarbeitsplatz aus unserer Zeit fänden: „My best guess is that we would be pictured as having a well-developed eye, a long right arm, a small left arm, uniform-length fingers and a "low-fi" ear“ (Buxton, 1986, S.319).

Dieses ironisch überzeichnete zwanzig Jahre alte Bild trifft in vielen Fällen noch heute zu. Allerdings zeichnet sich gegenwärtig getrieben durch technologische Innovationen sowohl in der wissenschaftlichen Forschung als auch zunehmend im industriellen Umfeld ein Trend ab, der das Potential zu bedeutenden Veränderungen aufweist. Durch multimodale Interaktion sollen viele Beschränkungen konventioneller Mensch-Maschine-Schnittstellen überwunden werden. Darunter ist die Aufbereitung von Systemausgaben für verschiedene Sinne der Nutzer sowie die Bereitstellung zusätzlicher alternativer Eingabewege für die Steuerung zu verstehen. Durch haptische, akustische und visuelle Ausgaben sowie beispielsweise Sprachsteuerung oder die Auswertung von Gesten erhofft man sich eine höhere Bedieneffizienz und mehr Intuitivität sowie eine bessere Aufgabenunterstützung bei der Bedienung komplexer Systeme.

Insgesamt hat die bisherige Forschung gezeigt, dass multimodale Interaktion durchaus die Gebrauchstauglichkeit steigern kann. Nicht selten stellen sich die technologischen Lösungen jedoch als schwierig zu bedienen heraus, bzw. erhoffte Förderungseffekte stellen sich nicht ein. Drei Faktoren sind dafür maßgeblich verantwortlich: die Vernachlässigung psychologischer Grundlagenerkenntnisse in der Anlage der Systeme, eine unklare Konzeptualisierung von Multimodalität und eine zu geringe Orientierung an der konkreten Aufgabe des Nutzers. Deutlicher Forschungsbedarf resultiert nicht zuletzt daraus, dass sich ein bedeutender Teil der Arbeiten zu multimodaler Interaktion auf den Einsatz und die Erprobung technologischer Innovati-

onen konzentriert, während systematischen experimentellen Tests der Lösungen zu wenig Raum zugestanden wird. Erschwerend wirkt sich für die experimentelle Methode aus, dass eine kontrollierte Manipulation aller Variablen eines multimodalen Gesamtsystems kaum möglich ist. Werden die vielfältigen Varianten zur Ein- und Ausgabe berücksichtigt, existieren schlichtweg zu viele Kombinationsmöglichkeiten. Dieser Umstand erfordert, experimentelle Fragen in handhabbaren Teiluntersuchungen anzugehen. Die in dieser Arbeit beschriebenen Experimente finden vor diesem Hintergrund statt. Das primäre Ziel ist es, in einem gegebenen Anwendungskontext mit angewandten psychologischen Methoden und Grundlagenerkenntnissen zu einer Lösung von offenen Gestaltungsfragen zu gelangen und zu einer Erweiterung der Wissensbasis über menschliches Verhalten in komplexen technischen Umgebungen beizutragen.

Der Anwendungsfall Fahrzeugführung weist wegen der hohen Anforderungen der visuell-manuell geprägten primären Fahraufgabe ein besonderes Potential für multimodale Schnittstellen zur Bedienung von nicht fahrrelevanten Komfortfunktionen auf. Da eine massiv ansteigende Anzahl derartiger Funktionen im Fahrzeug zu verzeichnen ist, müssen neue Bedienkonzepte entwickelt werden, die möglichst wenige Ressourcen von der Fahraufgabe abziehen.

Die vorliegende Arbeit stellt sich diesem Anspruch. Sie soll dazu beitragen, psychologisch fundierte und experimentell evaluierte multimodale Interaktionsprinzipien zu entwickeln, die in Gestaltungsempfehlungen für die praktische Umsetzung münden. Dazu werden die relevanten theoretischen Konzepte kontextübergreifend erörtert, um ein hinreichendes Verständnis der Möglichkeiten multimodaler Interaktion von psychologischer und technologischer Seite entwickeln zu können (Kapitel 2). Ferner wird der Anwendungsfall Fahrzeugführung in seinen kennzeichnenden Charakteristika dargelegt (Kapitel 3). Darauf aufbauend werden gezielte Fragestellungen zur Gestaltung multimodaler Systeme im Fahrzeug aufgeworfen, die sich auf der Grundlage gegenwärtig verfügbarer Literatur nicht beantworten lassen. Durch eine Serie von insgesamt acht Experimenten soll diesem Forschungsbedarf nachgegangen werden. Dabei werden sowohl Aspekte der Eingabe (Kapitel 4) als auch der Ausgabe (Kapitel 5) angesprochen. Aufgrund der herausragenden Stellung der manuellen und sprachlichen Schnittstelle zur multimodalen Bedienung im Automobil wird dabei auf diese Schnittstellenkomponenten fokussiert. Schließlich wird ein Prozessmodell für ein systematisches und generalisierbares Forschungsvorgehen zur Entwicklung und Evaluation multimodaler Systeme vorgeschlagen (Kapitel 6).

Die experimentellen Untersuchungen gliedern sich dem Gegenstand nach in zwei Untersuchungsabschnitte zu multimodaler Ein- bzw. Ausgabe, die sich jeweils gemäß ihrer Herangehensweise in grundlagenorientierte und anwendungsorientierte Experimente unterteilen lassen. Die anwendungsorientierten Untersuchungen finden in einem sehr praxisnahen Testumfeld statt. Mit Hilfe vereinfachter experimenteller Prototypen

und ingenieurpsychologischer Methoden soll Wissen über die Wirksamkeit konkreter Gestaltungsvarianten multimodaler Bordsysteme für Fahrzeuge gewonnen werden. Für die Eingabeseite wird evaluiert, inwiefern sich eine häufig verwendete und anfängerfreundliche, jedoch für Experten relativ langsame Variante der Sprachsteuerung durch die Einbeziehung kognitionstheoretischen Wissens verbessern lässt. Die angewandte Untersuchung zu multimodaler Ausgabe hat zum Ziel, die Rolle von akustischem, haptischem und visuellem Feedback für die manuelle Interaktion mit einem speziellen Eingabegerät zu erkunden.

Die grundlagenorientierten Untersuchungen zeichnen sich durch eine allgemeine Herangehensweise aus und setzen experimentelle Paradigmen ein, die aus der Grundlagenforschung stammen. Durch ihren höheren Abstraktionsgrad sind diese Experimente stärker verallgemeinerbar und somit für einen größeren Bereich der Mensch-Maschine-Interaktion relevant. Auf der Seite der Eingabe wird die Frage aufgeworfen, wie die grafische Benutzeroberfläche gestaltet sein muss, um sowohl manuelle als auch sprachbasierte Eingabe optimal zu unterstützen. Dazu wird untersucht, ob zwischen ikonografischen oder verbalen Funktionsdarstellungen in einer grafischen Bedienoberfläche stabile Unterschiede existieren. Die Fragestellung der Ausgabeseite fokussiert auf die Interferenz sprachlicher und nicht-sprachlicher akustischer Ausgaben mit Prozessen im Arbeitsgedächtnis.

Ausgehend von den Resultaten dieser Experimente werden Gestaltungsempfehlungen abgeleitet, um für zukünftig zu entwickelnde Bordsysteme in Fahrzeugen multimodale Lösungen realisieren zu können, die auf fundierten psychologischen Konzepten aufbauen, eine gute Bedienbarkeit gewährleisten und klar an der Aufgabe des Nutzers orientiert sind.

2 Multimodale Mensch-Maschine-Interaktion

Multimodalität wird allgemein für viele Anwendungsdomänen als Möglichkeit zur Erweiterung der begrenzten Bandbreite zum Austausch von Information zwischen Mensch und Maschine und zur Verbesserung der Interaktionsqualität anerkannt. In vielen Forschungsbeiträgen über multimodale Mensch-Maschine-Schnittstellen werden die Begriffe Modalität bzw. Multimodalität allerdings nicht explizit definiert oder grundlegende Begriffe werden teilweise in unterschiedlicher Bedeutung verwendet. Neuss (2000, S.392) weist zurecht darauf hin, dass dadurch bei der Interpretation oder der Realisierung multimodaler Systeme zu viel Freiraum existiert. Eine sorgfältige Definition der Grundbegriffe ist damit mehr als nur eine wissenschaftlich wünschenswerte Eingrenzung des Vokabulars, sie ist auch unter dem Gesichtspunkt der nutzerzentrierten Anwendung entsprechender Technologien unerlässlich, um den Einfluss der Systemvariablen auf die Mensch-Maschine-Interaktion abschätzen zu können. Die zugeschriebenen vorteilhaften Eigenschaften gelten nicht für alle Ausprägungen technischer Realisierungen gleichermaßen. Die notwendige begriffliche Klärung zur kategorialen Einordnung von Lösungsvarianten wird deshalb den Ausgangspunkt zu einer detaillierten Besprechung des Themengebiets multimodale Mensch-Maschine-Interaktion bilden.

Bei der Konstruktion von multimodalen Systemen stehen häufig technologische Fragen im Vordergrund, in vielen Fällen stellen die Prototypen die erstmalige Verknüpfung technischer Lösungen dar. Im praktischen Umfeld kommt es dabei häufig aus psychologischer Sicht zu suboptimalen Lösungen, die zumeist die kognitiven Grundlagen menschlicher Informationsverarbeitung nicht ausreichend berücksichtigen oder von falschen Grundannahmen über das typische Verhalten bei multimodaler Interaktion ausgehen (Oviatt, 1999). Die Integration psychologischer Expertise bereits in der Entwicklung technologischer Lösungen trägt zu einer höheren Gebrauchstauglichkeit und Nutzerakzeptanz bei (vgl. Zimmer, 2001b) und kann eine sinnvollere Anlage fortschrittlicher Ein- und Ausgabemechanismen fördern. Neben der Darlegung

technologischer Aspekte von Multimodalität ist aus diesem Grund eine Erörterung grundlegender Konzepte der Ingenieurpsychologie sowie eine Diskussion der Erkenntnisse aus Psychologie und humanwissenschaftlichen Nachbardisziplinen zur Nutzung spezieller Ein- und Ausgabemodalitäten und multimodaler Kombinationen unverzichtbar.

Bei dieser Diskussion darf nicht vergessen werden, dass Multimodalität kein Gütekriterium an sich ist, sondern lediglich eine technologische Lösung darstellt (Bengler, 2001). Gegenwärtig existiert nach wie vor keine umfassende Theorie, aus der eindeutig Konsequenzen multimodaler Gestaltung abgeleitet werden können (Geiser, 1990). Deshalb muss die Wirksamkeit einer derartigen Maßnahme stets in Einzelfalltests nachgewiesen werden. Wenn jedoch der Gegenstand definitorisch eingegrenzt wird und geplante Gestaltungsmaßnahmen im Licht relevanter Informationsverarbeitungsprozesse in Verbindung mit einer klaren Darstellung der Aufgabensituation (vgl. Kapitel 3) vorab geprüft werden, lässt sich mit höherer Wahrscheinlichkeit eine Lösung realisieren, die auf den menschlichen Nutzer gut abgestimmt ist.

2.1 Begriffliche Klärungen

2.1.1 Grundlegende Begriffe und Konzepte

Die Grundannahme jeder Mensch-Maschine-Interaktion¹ (MMI) ist nach Schomaker et al. (1995), dass mindestens zwei so genannte Agenten involviert sind. Unter Agent versteht man in diesem Kontext den menschliche Nutzer oder aber (teil-)autonom agierende Komponenten des maschinellen Systems, bzw. das maschinelle Gesamtsystem. Sie sind physisch getrennt, aber durch eine Schnittstelle unter Verwendung technischer Mittel in der Lage, über verschiedene Kanäle Information auszutauschen. Ein Kanal ist dabei der Übertragungsweg zwischen Sender und Empfänger der Information. Charakterisierend für die Interaktionssituation ist nach Charwat (1994), dass sich Mensch und Maschine wechselseitig als weitgehend unabhängige Funktionseinheiten beeinflussen. Stehen Mensch und Maschine in ihrer Wirkung aufeinander im Sinne einer Zielerreichung in Beziehung, spricht man nach Charwat von einem Mensch-Maschine-System (MMS). Das MMS ist durch die Form des Informationsaustauschs

¹ In der Literatur wird neben Mensch-Maschine-Interaktion auch häufig von Mensch-Maschine-Kommunikation gesprochen. Charwat (1994) verweist darauf, dass der Begriff der Interaktion zutreffender sei, da Kommunikation ein Verstehen des übertragenen Inhalts voraussetzt. Der Begriff der Kommunikation kann dann angemessen erscheinen, wenn der Mensch den Computer beispielsweise im Dialogkontext einer komplexen Datenbankabfrage tatsächlich als Gesprächspartner behandelt (Bunt, 1998). Nachdem der Begriff der Interaktion allgemeiner ist, soll ihm im weiteren Verlauf der Vorzug gegeben werden.

zwischen Mensch und Maschine gekennzeichnet, während sich die Wechselwirkung mit zunehmender Automatisierung verändert (Hoyos, 1990). Je nach Automatisierungsgrad, Systemziel und technischer Ausrüstung eines MMS muss die Schnittstelle zwischen Mensch und Maschine gestaltet werden. Während in frühen technischen Systemen dieser Gestaltung noch wenig Aufmerksamkeit gewidmet wurde, ist mittlerweile unumstritten, dass die Schnittstellengestaltung maßgeblich die Verwendbarkeit eines MMS bestimmt (Streitz, 1990).

Für die ergonomische Analyse kann man vom Grundmodell eines Arbeitsprozesses ausgehen (Reichart & Haller, 1995). Das Mensch-Maschine-System ist danach eingebettet in eine spezifische Arbeitsumgebung als Teil der Umwelt, deren Eigenschaften auf den Prozess von Aufgabenstellung zu Aufgabenerfüllung Einfluss haben (vgl. Abbildung 1). Dabei liegt die Aufgabenstellung entweder als selbst gewählte Aufgabe vor, oder sie besteht aus einer Aufgabe, die sich aus der Situation ergibt. Zur vollständigen Erfassung der relevanten Determinanten des Prozesses müssen zusätzlich sowohl die internen Eigenschaften des menschlichen Operators (z.B. Leistungsfähigkeit und -einschränkungen, Kenntnisse), als auch der Maschine (z.B. technische Rahmenbedingungen) erkannt werden.

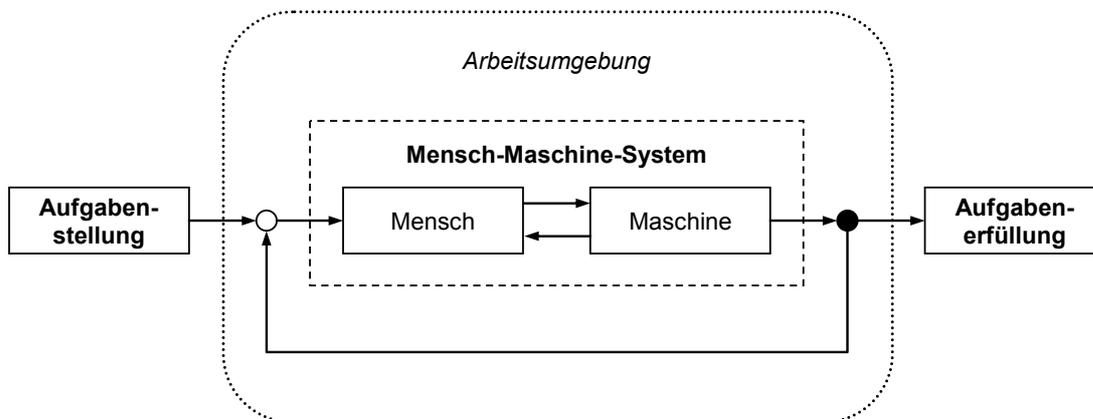


Abbildung 1: Grundmodell eines Arbeitsprozesses (nach Reichart & Haller, 1995, S.200).

Die Güte der Aufgabenerfüllung hängt von der Gebrauchstauglichkeit² der Schnittstelle ab. Nach der ISO-Norm 9241-11 ist zentral, in welchem Ausmaß ein technisches Produkt in der gegebenen Arbeitsumwelt für bestimmte Nutzer effektiv, effizient und zufriedenstellend zum Erfüllen der gesetzten Ziele geeignet ist (ISO 9241-11, 1998). Dabei bezeichnet Effektivität die Genauigkeit und Vollständigkeit, mit der Benutzer ein bestimmtes Arbeitsergebnis erzielen. Effizienz ist der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem Benutzer ein bestimmtes Ziel

² Die Begriffe Benutzungsfreundlichkeit, Benutzerfreundlichkeit, Usability, Benutzbarkeit, Gebrauchstauglichkeit und Gebrauchsfähigkeit werden in der Literatur häufig synonym verwendet.

erreichen. Unter Zufriedenheit versteht man die Akzeptanz, die der Nutzer dem Produkt entgegenbringt, und die subjektive Bewertung des Produkts durch die Nutzer in Bezug auf Beeinträchtigungsfreiheit.

Zur Erhöhung der Usability müssen neben grundlegenden Anforderungen der Informationsdarstellung wie der Erkennbarkeit von Darstellungen, einer Lesbarkeit und Unterscheidbarkeit von Ausgaben (vgl. ISO 9241-12, 1998) insbesondere die Dialoge, die einzelnen Interaktionsschritte, zwischen Mensch und Maschine sinnvoll gestaltet sein. Die ISO-Norm 9241-110³ (2006) listet sieben Grundsätze zur Gestaltung und Bewertung von Dialogen auf. Essentiell für die Gebrauchstauglichkeit interaktiver Systeme sind demnach Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Erwartungskonformität, Lernförderlichkeit, Steuerbarkeit, Fehlertoleranz und Individualisierbarkeit. Diese Empfehlungen müssen mit Blick auf ihre Anwendbarkeit im jeweiligen Nutzungskontext umgesetzt werden, da nicht jeder Grundsatz für jeden Nutzungskontext gleichermaßen sinnvoll ist.

Weitere Leitsätze zu Usability lassen sich in der psychologischen Literatur zur Mensch-Maschine-Interaktion finden (z.B. Norman, 1988; Norman & Draper, 1986; Shneiderman, 1997). Norman (1988) formuliert zwei Imperative benutzerfreundlicher Gestaltung: 1) *Make things visible* und 2) *Provide a good conceptual model*. Die erste Forderung – Dinge sichtbar zu machen – bezieht sich darauf, dass alle verfügbaren Funktionen derart gestaltet sein müssen, dass die Nutzer erkennen können, a) dass sie vorhanden sind und b) welche Aktionen auf den entsprechenden Bedienelementen ausführbar sind. Die zweite Forderung – durch das Design den Nutzern die Ausbildung einer richtigen konzeptuellen Vorstellung vom Funktionieren einer Anwendung zu ermöglichen – zielt darauf ab, dass ein korrektes konzeptuelles Modell insbesondere dann notwendig ist, wenn die Nutzer vor neuen Aufgaben stehen. Nur mit einem guten Modell können korrekte Vorhersagen über die Auswirkung von Aktionen auch in unbekanntem oder wenig genutzten Applikationskontexten getroffen werden. Eine gute Sichtbarkeit von Funktionen vorausgesetzt, sieht Norman einen hohen Aufforderungscharakter (*affordances*) von Objekten, natürliche Abbildungen (*mappings*) und klare Beschränkungen (*constraints*) als wichtigste Quellen für die Vermittlung eines guten konzeptuellen Modells an. Zusätzlich muss eine gute Rückmeldung (*feedback*) über ausgeführte Aktionen oder erzielte Ergebnisse gegeben werden. *Affordances*, die Norman aufbauend auf Gibson (1977) formulierte, sind die Menge der möglichen Aktionen, die das Design eines Objekts dem Nutzer nahe legt. Sie deuten an, was damit getan werden kann. *Mappings* schaffen Verbindungen zwischen dem Bedien-

³ Die Norm ISO 9241-110 von 2006 löst die bekannte und häufig zitierte ISO 9241-10 (1996) ab. Die Normenreihe ISO 9241 wird seitdem unter dem Titel „Ergonomie der Mensch-System-Interaktion“ mit erweitertem Geltungsbereich geführt, d.h. die Einschränkung auf Bildschirmarbeitsplätze aufgehoben.

element und seiner realen Auswirkung. Gute Abbildungen sind *natürlich* durch das Nutzen von physikalischen Analogien und kulturellen Standards, erlauben intuitives Verständnis und sind leicht aus dem Gedächtnis abrufbar. Beispielsweise besteht eine natürliche Abbildung zwischen der Bewegung des Lenkrads und der Fahrtrichtung des Automobils, ein Lenkrad vermittelt darüber hinaus auch die Aufforderung, es zu drehen (und nicht etwa zu ziehen oder zu drücken). Constraints schränken die Benutzung von Objekten ein. Sie ergänzen somit die Affordances aus negativer Sicht und machen deutlich, wofür sich ein Objekt nicht eignet. Feedback ist schließlich erforderlich, damit die Nutzer jederzeit über den Zustand des Systems und den Effekt ihrer Handlungen informiert sind.

2.1.2 Modalität

Um das Konzept der multimodalen Interaktion in das Grundgerüst eines Mensch-Maschine-Systems einzubetten, ist zunächst eine Klärung des Begriffs „Modalität“ notwendig, der in der Mensch-Maschine Forschung uneinheitlich gebraucht wird. In der psychologischen und ergonomischen Forschung wird damit meist auf die menschlichen Sinnesmodalitäten Bezug genommen, während in der Computerwissenschaft Modalität näher am Begriff Modus orientiert ist und allgemein die Art und Weise der Ein- und Ausgabe bezeichnet (Oviatt, 2003; Raisamo, 1999). Zur Abgrenzung hat sich die Unterscheidung nach Wahrnehmungs- und Aktionsmodalitäten als hilfreich erwiesen (Hedicke, 2000).

2.1.2.1 Wahrnehmungsmodalitäten

Der Begriff Modalität stammt ursprünglich aus der Physiologie und bedeutet Wahrnehmung über einen bestimmten Sinneskanal. In der klassischen Medizin des Altertums und der frühen Neuzeit wurden als menschliche Sinne dabei nur die Empfindungskomplexe Sehen, Hören, Riechen, Schmecken und Fühlen unterschieden. Heute sind eine Reihe weiterer Sinne etabliert, z.B. der Gleichgewichtssinn, der Temperatursinn, der Schmerzsin (vgl. Schmidt & Thews, 1997). Die beiden letzteren spiegeln eine umfassendere Aufteilung der Somatosensorik wider. Während in der klassischen Terminologie der Begriff Fühlen den Eindruck einer einheitlichen Funktion nahe legte, ist heute bekannt, dass die Somatosensorik ein multiples System ist. Es setzt sich aus mehreren Submodalitäten zusammen und verarbeitet sowohl exterozeptive (Reize aus der Umwelt) als auch propriozeptive (Informationen über die relative Lage der Körperteile zueinander und der Stellung und Bewegung des Körpers) und interozeptive Information (Registrierung von Vorgängen im Körper, wie z.B. die Höhe des Blutdrucks). Der Vestibularsinn (Gleichgewichtssinn) wird üblicherweise nicht zur Somatosensorik gezählt, sondern getrennt aufgeführt. Trotz des gewachsenen Wissens

über die Art und Arbeitsweise von Sinnesrezeptoren ist jedoch die genaue Anzahl der Sinne, über die der Mensch verfügt, nicht endgültig geklärt und wird nach Schmidt und Thews (1997) wohl stets eine Interpretationsfrage bleiben.

Die Begriffe Wahrnehmungsmodalität, Sinnesmodalität und Ausgabemodalität werden im Kontext der Mensch-Maschine-Interaktion gleichermaßen zur Angabe der involvierten menschlichen Sinnesleistungen bzw. zur Charakterisierung der maschinellen Ausgabe verwendet. Welche Rolle spielen nun die menschlichen Sinnesleistungen bei der Interaktion mit technischen Systemen? Offensichtlich ist der Geschmackssinn hier wenig geeignet, relevante Information zu übermitteln. Auch der Geruchssinn spielt lediglich eine untergeordnete Rolle. Entsprechend definiert Geiser (1990) im Kontext der Darstellung von Information nur optische, akustische und haptische Anzeigen, die die visuelle, auditive und die taktile bzw. kinästhetische Sinnesmodalität involvieren (vgl. Tabelle 1).

Tabelle 1: Zusammenhang zwischen Sinnesorganen (bzw. Sinne des Menschen), Sinnesmodalitäten und Anzeigearten (vgl. Geiser, 1990, S.61)

Sinnesorgane, bzw. Sinne	Sinnesmodalität	Anzeigeart
Sehsinn	visuell	optisch
Hörsinn	auditiv	akustisch
Geruchssinn	olfaktorisch	—
Geschmackssinn	gustatorisch	—
Gleichgewichtssinn	vestibulär	—
Drucksinn Berührungssinn Vibrationssinn Kältesinn Wärmesinn Schmerzsinn	Hautsinn, Tastsinn taktil	haptisch
Stellungssinn Kraftsinn	Proprio- zeptoren kinästhetisch	

Schomaker et al. (1995) heben hervor, dass in Simulationsumgebungen wie Flugsimulatoren oder Anwendungen der virtuellen Realität die Einbindung des Gleichgewichtssinns eine zunehmend bedeutende Rolle spielt. Haptische Displays nehmen eine Sonderstellung ein, da sie nicht eindeutig auf eine Sinnesmodalität abzielen, sondern sowohl für die taktile als auch die kinästhetische Wahrnehmung Information ausgeben können (vgl. Burdea, 1996).

2.1.2.2 Aktionsmodalitäten

In Orientierung an den aufgezeigten Wahrnehmungsmodalitäten lassen sich auch Aktionsmodalitäten (synonym: Eingabemodalitäten) beschreiben, d.h. die Formen der Informationsübertragung von Mensch zu Maschine. Hedicke (2000) schlägt dazu eine Unterscheidung mehrerer technischer Schnittstellenkomponenten einer Mensch-Maschine-Schnittstelle vor, deren Benennung sich bewusst an den menschlichen Sinnen und nicht an der technischen Sensorik orientiert (siehe Abbildung 2). So unterscheidet er ein auditives, ein visuelles und ein haptisches Interface, wobei nicht nur die Art der Wahrnehmung des Menschen, sondern auch die Art der Informationsregistrierung durch die Maschine gemeint ist. Diese Einteilung ist folglich zur Beschreibung von Wahrnehmungs- und Aktionsmodalitäten gleichermaßen geeignet.

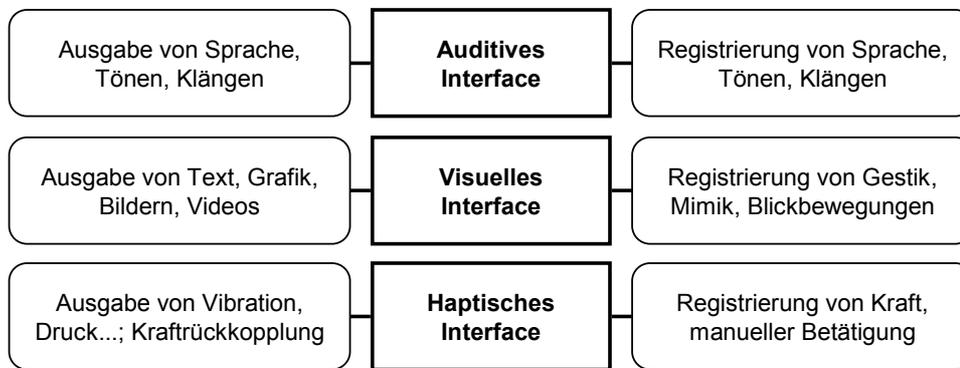


Abbildung 2: Technische Komponenten einer multimodalen Schnittstelle (vgl. Hedicke, 2000, S.210).

Zu den Schnittstellen im Einzelnen:

- Über das auditive Interface werden akustische Informationen durch Luftschwingungen (Schallwellen) übertragen. Eingaben durch den Menschen erfolgen vorwiegend durch Sprache. Zwar sind auch nonverbale Kommandos denkbar, wie z.B. in die Hände klatschen, diese werden aber nur in Ausnahmefällen sinnvoll sein.
- Über das visuelle Interface werden Informationen in Analogie zur menschlichen Wahrnehmung vom „visuellen System“ des Computers, d. h. von den entsprechenden Sensoren, registriert. Darunter fallen Körper- und Körperteilbewegungen, insbesondere Augenbewegungen, Gestik und Mimik.
- Der Begriff des haptischen Interface ist mit dem gesamten Bereich der Somatosensorik und der Gleichgewichtsempfindung weit gefasst. Die relevanten Eingabesignale für die haptische Schnittstelle sind die Veränderung der Lage von Teilen eines Gerätes, das Drücken von Tasten, die Intensität der Betätigung eines Stellteils, die Verformung von Körpern (beispielsweise bei einem CAD-Programm) und ähnliches.

Bei der Benennung der Schnittstellenkomponenten fällt auf, dass die Orientierung an Sinnesmodalitäten bezüglich des Systemoutputs eindeutig und sinnvoll sind, denn der Mensch kann nur diejenigen Signale einer Maschine verarbeiten, die er über seine Sinne wahrnehmen kann. Will man jedoch die Eingabe von Kommandos in die Maschine beschreiben, muss man anstelle von Sinnesleistungen von Sensoren sprechen. Bernsen (1994a, 1994b, 2002) weist darauf hin, dass deshalb zwischen Input- und Outputmodalitäten bedeutsame Unterschiede existieren, die es bei der Mensch-Maschine-Interaktion zu beachten gilt, so dass man nicht einfach die Taxonomie der Output-Modalitäten auf die Input-Modalitäten übertragen kann: Erstens gehe die Leistung vieler Sensoren über den für Menschen wahrnehmbaren Teil der Welt hinaus, beispielsweise können Infrarot-, Radar- oder Ultraschallsignale verarbeitet werden. Zweitens kämen ständig neue technische Möglichkeiten der Verarbeitung von Eingaben hinzu. Vor den 1980er Jahren stand beispielsweise gesprochene Sprache nicht als Möglichkeit zur Kommandoingabe zur Verfügung, mittlerweile wird diese Technologie häufig verwendet. Während also eine Taxonomie der Output-Modalitäten abschließend durch Bezugnahme auf alle menschlichen Sinnesleistungen möglich ist, spiegelt eine Taxonomie der Input-Modalitäten in gewisser Weise immer nur den gegenwärtigen Stand der Technik wider.

2.1.3 Multimodale Schnittstellen

In der multimodalen Mensch-Maschine-Interaktion müssen sowohl Aktions- als auch Wahrnehmungsmodalitäten sinnvoll aufeinander abgestimmt sein. Ihr Zusammenspiel muss so ausgelegt sein, dass Mensch und Maschine ihre vor- bzw. selbst bestimmten Ziele erreichen können (Nigay, Jambon & Coutaz, 1995). Zur Analyse eines gegebenen Mensch-Maschine-Systems, verweisen Schomaker et al. (1995) darauf, dass die beiden grundlegenden psychologischen Prozesse, Wahrnehmung und Steuerung (Eingabe), erfasst sein müssen. Folglich sollten stets sowohl die Aktuatoren als auch die Sensoren beschrieben werden, die an der Interaktion beteiligt sind. Schomaker und Kollegen schlagen ein Grundmodell zur Beschreibung der Mensch-Computer-Interaktion vor, das in Abbildung 3 wiedergegeben ist und die Grundlage für eine Erfassung der Auswirkungen von Multimodalität darstellt. Der Verweis auf die Kognition auf Seiten des Menschen deutet dessen fortlaufende Ziel-Folge-Prüfung in der Interaktion an. Beim Computer findet diese Verarbeitung nur in Form programmierter Routinen statt. Um diesen Unterschied sichtbar zu machen, wurde „cognition“ von den Autoren in Anführungszeichen gesetzt.

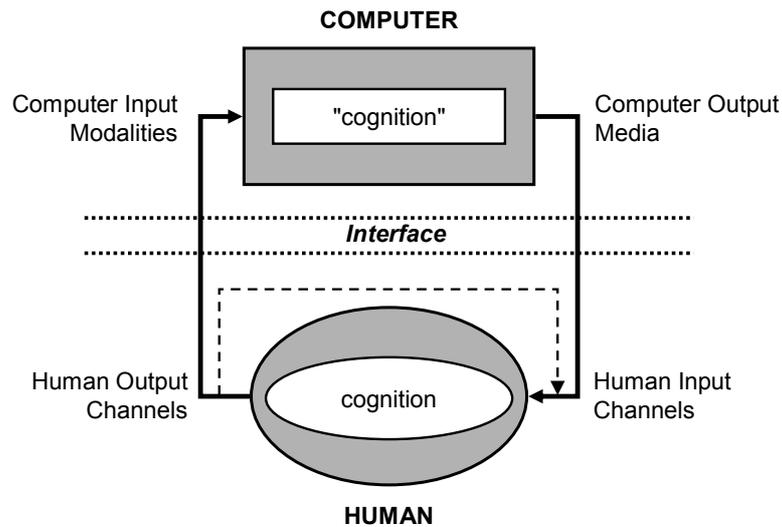


Abbildung 3: Modell zur Prozessidentifikation in der Mensch-Computer-Interaktion (Schomaker et al., 1995, S.2). Die durchgezogenen Pfeile veranschaulichen den Interaktions-/Informationsfluss, die gestrichelten Pfeile veranschaulichen intrinsische Handlungsüberwachungsschleifen.

Schomaker et al. (1995) streben mit diesem Beschreibungsmodell eine möglichst eindeutige Terminologie an. Auf der einen Seite, aus menschlicher Sicht den Steuerungsprozessen zugehörig, listen sie die menschlichen Ausgabekanäle (Human Output Channels) und die Computer-Eingabemodalitäten (Computer Input Modalities) auf. Dazu unterscheiden sie auf der anderen Seite, den Wahrnehmungsprozessen zugehörig, die Ausgabemedien des Computers (Computer Output Media) und die menschlichen Eingangskanäle (Human Input Channels). Aus Gründen der Eindeutigkeit verwenden sie nur im Zusammenhang mit Computer-Eingabemodalitäten den Begriff Modalität. Diese Wahl erscheint etwas willkürlich, da bezüglich der sensorischen Aufnahme des Menschen ebenso gut von Wahrnehmungsmodalitäten gesprochen werden kann. Der Vorteil dieses Modells liegt jedoch weniger in der Terminologie als vielmehr in der vorgenommenen Aufspaltung der Informationsübertragung: So kann ein Kommando vom Menschen über mehrere mögliche Ausgabekanäle erfolgen, die dafür unterschiedlich gut geeignet sind. Dieses Kommando kann wiederum auf Seiten des Computers nur empfangen werden, wenn die vorhandenen Eingabemodalitäten mit den Ausgabekanälen des Menschen zusammen passen. Gleiches gilt für den Informationsfluss in die andere Richtung. Die vollständige Beschreibung einer Interaktionssituation kann nur durch die Angabe aller beteiligten Modalitäten auf jeder Seite – Steuerung oder Wahrnehmung – erreicht werden.

2.1.3.1 Definition Multimodalität

Gegenwärtig existieren sehr heterogene Definitionen des Terminus Multimodalität. Die fehlende Übereinstimmung wird mitunter in der Interdisziplinarität des Gegens-

tands vermutet (vgl. Benoît, Martin, Pelachaud, Schomaker & Suhm, 2000). Eine exklusive Fokussierung der Disziplinen ist zwar zunehmend weniger zu beobachten, jedoch legen disziplinübergreifend viele wissenschaftliche Beiträge ihren Schwerpunkt entweder auf multimodale Ein- oder Ausgabe vom bzw. an den menschlichen Nutzer. Diese Schwerpunkte spiegeln sich in den verwendeten Definitionen wider. So legt Brewster (1994, S.9) fest: „A multimodal interface is defined as one that presents information in different sensory modalities, specifically visual and auditory“. Die Definition von Oviatt (2003, S.286) hingegen betont den Eingabeprozess: „Multimodal systems process two or more combined user input modes – such as speech, pen, touch, manual gestures, gaze, and head and body movements – in a coordinated manner with multimedia system output.“

Zunächst spricht sicherlich nichts gegen eine Eingrenzung des Problembereichs, um den technologischen und versuchsplanerischen Herausforderungen bei der Forschung an multimodalen Schnittstellen Herr zu werden. Jedoch mündet sie in der unbefriedigenden Situation, dass die genannten Definitionen durch ihre Fokussierung auf Teilaspekte multimodaler Interaktion die jeweils andere Form der Multimodalität nicht erfassen können und somit als allgemeine Begriffsbestimmung wenig brauchbar sind. Das European Telecommunications Standards Institute (ETSI) definiert „multimodal“ zwar ebenfalls in Anlehnung an sensorische Modalitäten, versteht darunter aber sowohl die Sensorik des Menschen als auch die der Maschine. „Multimodal“ ist demnach ein „adjective that indicates that at least one of the directions of a two-way communication uses two sensory modalities (vision, touch, hearing, olfaction, speech, gestures, etc.)“ (ETSI EG 202 191, 2003, S.7). An gleicher Stelle werden die Randbedingungen gekennzeichnet: „a) more than one sensory modality is available for the channel (e.g. output can be visual or auditory); or b) within a channel, a particular piece of information is represented in more than one sensory modality (e.g. the command to open a file can be spoken or typed).“ Dieses erweiterte Verständnis erscheint als verbindliche Definition geeignet, da es gleichermaßen den Bereich der Ein- und Ausgabe erfasst.

Schomaker et al. (1995) schlagen zunächst eine nahezu identische Begriffsbestimmung für multimodale Schnittstellen vor, die als Minimalkriterium fordert, dass mehr als eine Interaktionsmodalität auf der Ein- oder der Ausgabeseite bei der Interaktion involviert ist. Allerdings weisen sie auf eine Unschärfe hin. Denn nach diesem Kriterium müsste jede Form von Mensch-Computer-Interaktion als multimodal angesehen werden (S.6): „In this sense every human-computer-interaction has to be considered as multimodal, because the user *looks* at the monitor, *types* in some commands or *moves* the mouse (or some other device) and *clicks* at certain positions, *hears* the reaction (beeps, key clicks, etc.) and so on.“ Schomaker und Kollegen nehmen deshalb eine interessante Einschränkung vor: „our understanding of multimodality is restricted to

those interactions which comprise more than one modality on either the input [...] or the output [...] side of the loop *and* the use of more than one device on either side.” Im Gegensatz zu der Kombination von visuellem, auditivem und taktilem Feedback beim Tippen auf einer Tastatur sei damit die Kombination von auditivem und visuellem Feedback durch einen Monitor und einen Lautsprecher ein echtes multimodales Ereignis. Zwar verhindert diese Einschränkung triviale Multimodalität, weist jedoch an zwei Punkten Unklarheiten auf. Erstens wird keine befriedigende Aussage dazu getroffen, wie mit Bedienelementen umzugehen ist, die durch die Gestaltung ihrer Form oder Mechanik mehrere Sinne des Nutzers gezielt ansprechen. In vielen Fällen wird versucht, die wahrgenommene bzw. die aufgabenbezogene Qualität manueller Stellteile durch spezielle akustische und haptische Eigenschaften zu verbessern. So gibt es beispielsweise gerade im Automobil-Bereich durch psychophysische Experimente gewonnenes ausgereiftes Wissen über optimale Druckpunkte von Tastern, Oberflächentexturen von Bedienelementen oder Widerstände von Drehwählern (vgl. Bubb, 2001; Grunwald & Krause, 2001; Tietz, 2001). Ebenso sind neben der Analyse der wahrgenommenen Qualität von Fahrzeuginnenraumgeräuschen manuelle Stellteile in den Fokus der Aufmerksamkeit auditiver Gestaltung (Sound-Design) gerückt (Hempel, 2003). Dieses Design zielt ebenso wie Multimodalität durch technisch vermittelte Ausgaben auf die komplexen Wahrnehmungseigenschaften des Menschen mit dem Ziel einer Nutzung für die Mensch-Maschine-Interaktion ab. Zweitens ist die Vorgabe, mehr als ein Gerät (*device*) zur Ein- oder Ausgabe zu verwenden, auch aus technologischer Sicht nicht immer nachvollziehbar. So können beispielsweise mit einem geeigneten Lautsprecher durch entsprechende tief- bzw. höherfrequente Signalanteile sowohl vibrotaktile haptische Ausgaben als auch auditiv wahrnehmbare Schallmuster erzeugt werden.

Zur Lösung der Definitionsfrage erscheint es sinnvoll, zunächst der Definition von ETSI zu folgen, jedoch die Einschränkung von Schomaker und Kollegen (1995) bezüglich der Ausgabeseite abzuwandeln und aufzunehmen. Nicht die *Anzahl* der Geräte ist entscheidend. Grundlage der Unterscheidung ist vielmehr, ob eine durch den Systemdesigner manipulierbare *aktive, technisch vermittelte Rückmeldung* einer Nutzeraktion über mehrere Sinne vorliegt oder eine *passive, rein mechanische Auswirkung* dieser Nutzeraktion. Während beide Ereignisse zu einer multimodalen Wahrnehmung des Nutzers führen, sollten diese Aspekte einer gesonderten Betrachtung zugeführt werden. Die vollen Möglichkeiten eines multimodalen Systems können nur erreicht werden, wenn durch eine technisch vermittelte Ausgabe aktiv ein Signal als Reaktion auf einen Systemzustand ausgegeben wird und dieses Signal in seiner Ausprägung beeinflussbar, gegebenenfalls deaktivierbar ist. Unter diesen Vorzeichen kann der Computer als Akteur, wie im oben dargelegten Beschreibungsmodell von Schomaker et al. (1995) angegeben, zielführende Handlungen im Rahmen des programmierten

Rückmeldekonzepts für verschiedene Systemzustände und Umwelten initiieren. Werden durch mechanische Eigenschaften stets dieselben Formen beispielsweise auditiver und haptischer Wahrnehmung hervorgerufen, diese aber nicht aktiv vom System initiiert, kann nicht von einer „Ausgabe“ gesprochen werden. Zur Unterscheidung wird empfohlen, nur im ersten Fall von einem *multimodalen System* zu sprechen und die Berücksichtigung akustischer, visueller oder haptischer mechanischer Eigenschaften unter dem erweiterten Verständnis einer Gestaltung von Schnittstellenkomponenten für *multimodale Wahrnehmung* zu fassen.

Für eine klare Konzeptualisierung multimodaler Schnittstellen empfiehlt sich ferner eine eindeutige Abgrenzung von dem Begriff Multimedia. Sowohl bei Systemen, die bezüglich der Ausgabeseite multimodal sind, als auch bei multimedialer Darbietung erfolgt die Informationsausgabe über mehrere Kanäle. Turk und Robertson (2000) heben jedoch hervor, dass Multimedia auf die reine Präsentation von Information für bestimmte Nutzergruppen entsprechend deren kognitiven und perzeptiven Fähigkeiten abzielt. Buxton (1994) weist darauf hin, dass Multimedia auf das Medium oder die Technologie fokussiert, nicht auf die Anwendung durch den Nutzer. Bei multimodalen Systemen hingegen steht der Aspekt der Interaktivität und damit die Ausrichtung auf den Nutzer im Vordergrund. Deshalb sind bei Multimodalität die verschiedenen Ein- und Ausgabekanäle auf die Optimierung des Interaktionsprozesses, den gegenseitigen Informationsaustausch, ausgerichtet.

2.1.3.2 Kennzeichen multimodaler Systeme

Da die Definition eines multimodalen Systems relativ weit gefasst ist, ist es notwendig, dessen konkrete Eigenschaften präzise zu beschreiben. Durch eine Kennzeichnung der Art multimodaler Ein- und Ausgabe kann ein hilfreiches Kategoriensystem für den praktischen Einsatz und die Abschätzung der Auswirkungen multimodaler Systeme gegeben werden. Zunächst muss bei der Kennzeichnung unterschieden werden, ob das System multimodal bezüglich Ausgabe (Wahrnehmungsmodalitäten) oder Eingabe (Aktionsmodalitäten) ist, bzw. ob Multimodalität auf beiden Seiten vorliegt. In der Literatur existieren ausführliche Vorschläge dazu, wie die Aktionsmodalitäten und deren Zusammenwirken detailliert charakterisiert werden können (Martin, 1997, 1998; Neuss, 2000; Nigay & Coutaz, 1993; Nigay et al., 1995; Oviatt, 2003). Meistens wird die Verwendung der Wahrnehmungsmodalitäten weniger detailliert dargelegt. Mindestens zwei Bestimmungsstücke sollten bei der Beschreibung multimodaler Ausgabe jedoch angegeben werden. Erstens: Welche Wahrnehmungsmodalitäten sind involviert? Zweitens: Sind die Modalitäten äquivalent oder sind bestimmte Informationen für eine Modalität spezialisiert? Bei Äquivalenz kann jedes Ereignis immer durch alle über verfügbaren Modalitäten repräsentiert werden. Falls Äquivalenz vorliegt, muss weiter geklärt werden, ob für Ausgaben die Modalitäten redundant genutzt werden

oder aufgrund welcher Prinzipien die Wahl auf eine der Modalitäten fällt. Weist das System bei der Ausgabe Spezialisierung auf, können bestimmte Ereignisse nur exklusiv in einer zugewiesenen Modalität gemeldet werden.

Die Kennzeichnung von Systemen mit multimodaler Eingabe richtet sich nach der Art der Verwendung der Aktionsmodalitäten. Martin (1997, 1998) nennt sechs Formen, die er als *types of cooperation* zwischen den Modalitäten bezeichnet und die sich mit Einschränkungen auch auf Systemausgaben anwenden lassen.

- Äquivalenz (*equivalence*): Mehrere Modalitäten sind zueinander gleichwertig. Das bedeutet, sie können alternativ für eine Information verwendet werden. Dadurch kann der Nutzer bei der Eingabe die Modalität wählen, die ihm am meisten zusagt bzw. die für eine gegebene Aufgabe am besten geeignet ist.
- Spezialisierung (*specialization*): Wenn die Aufgabenteilung zwischen Modalitäten durch Spezialisierung gekennzeichnet ist, kann eine bestimmte Eingabe nur exklusiv über eine bestimmte dafür vorgesehene Modalität vorgenommen werden. Martin weist darauf hin, dass die Spezialisierung nur in wenigen Systemen absolut gehandhabt werden kann. Bei Spezialisierung muss besonderes Augenmerk darauf gelegt werden, wie gut die jeweilige Modalität die zugewiesene Information repräsentieren kann.
- Redundanz (*redundancy*): Dieselbe Information wird durch mehrere Modalitäten übertragen. Redundante Eingabe kann bei erkenntnisbasierten Technologien für eine Disambiguierung der Eingabe verwendet werden, z.B. durch die Redundanz in der Information eines Sprachkommandos und der zugehörigen Lippenbewegungen.
- Komplementarität (*complementarity*): Die Gesamtinformation ist verteilt über mehrere Modalitäten. Ein prototypisches Beispiel ist die Verbindung von Gestik und Sprache durch einen Befehl, bei dem auf ein Objekt gedeutet und per Sprache die Operation auf diesem Objekt festgelegt wird (Bolt, 1980). Für sich allein trägt weder die Geste noch das Sprachkommando die intendierte Bedeutung. Diese erschließt sich erst aus der zusammengesetzten Interpretation der beiden Ereignisse, die vom System geleistet werden muss.
- Übertragung (*transfer*): Informationen, die in einer Modalität erzeugt wurden, werden durch eine andere Modalität genutzt. Ähnlich wie bei Redundanz kann dadurch bei multimodaler Eingabe die Erkennungsgenauigkeit erhöht werden. Der Unterschied liegt darin, dass bei Übertragung kein inhärenter Zusammenhang der Modalitäten vorliegt und durch Modellannahmen eine semantische Verbindung erst etabliert werden muss. So kann beispielsweise die Registrierung eines Klicks mit der Computermaus auf ein bestimmtes Element dabei

helfen, ein gegebenes Sprachkommando, welches sich auf dieses Element bezieht, einzuordnen.

- Simultanität (*concurrency*): Unterschiedliche Arten von Informationen können über mehrere Modalitäten gleichzeitig verarbeitet werden und sind unabhängig. Eine gemeinsame Interpretation findet nicht statt. Diese Eigenschaft beschleunigt den Interaktionsprozess, wenn mehrere Modalitäten nahtlos parallel genutzt werden können.

Diese Kennzeichen können sowohl auf das gesamte System als auch auf einzelne Subsysteme angewandt werden. So ist denkbar, dass nur in einer Teilapplikation eines Systems manuelle Steuerung und Sprache äquivalent sind und im Gesamtsystem manuelle Eingabe überwiegend exklusiv verwendet wird (Spezialisierung). Eine solche Kennzeichnung der Eigenschaften multimodaler Systeme trägt dazu bei, ein besseres Verständnis über deren Funktionsweise und Eignung für bestimmte Aufgaben zu entwickeln. Aus Sicht der nutzerzentrierten Gestaltung muss jedoch angemerkt werden, dass damit zunächst nur eine formale Kategorisierung vorliegt. So bezeichnet Äquivalenz beispielsweise ausschließlich die Mächtigkeit mindestens zweier Modalitäten, das System von einem Zustand s in den Zustand s' zu überführen. Inwiefern die Modalitäten sich in ihrer Eignung unterscheiden wird dadurch nicht ausgesagt. Hier gilt nach wie vor die eingangs in diesem Kapitel vorgebrachte Forderung, die Wirksamkeit multimodaler Gestaltungsmaßnahmen in Einzelfalltests kritisch empirisch zu prüfen.

Eine der anspruchsvollsten Aufgaben bei der Implementierung multimodaler Systeme ist die so genannte Fusion, die Vereinigung unterschiedlicher Modalitäten auf einem gemeinsamen semantischen Niveau. Oviatt et al. (2000) unterscheiden zwei Subtypen multimodaler Architektur zur gleichzeitigen Verarbeitung mehrerer Eingabesignale. Der eine integriert die Signale auf Merkmalsebene (*feature level, early fusion*), der andere auf semantischer Ebene (*semantic level, late fusion*). Für die Integration auf Merkmalsebene werden zunächst unimodale physikalische Merkmale der Signale aus den Datenströmen extrahiert, die im Anschluss in einer gemeinsamen Repräsentation kombiniert werden und erst dann durch Interpretation ihre Bedeutung erhalten. Dabei beeinflusst typischerweise das Ergebnis der Erkennung in einer Eingabemodalität den Verlauf der Erkennung in der anderen Modalität. Diese Form der Fusion wird hauptsächlich bei Modalitäten verwendet, die zeitlich nah synchronisiert sind (z.B. Lippenbewegungen und Sprache). Sie stößt schnell an ihre Grenzen, wenn der transportierte Informationsgehalt der Modalitäten stark unterschiedlich ist oder wenn sie im Zeitverlauf massiv differieren. Auch bei der Fusion auf semantischer Ebene werden zuerst unimodale Merkmale extrahiert. Allerdings werden diese nicht sofort kombiniert, sondern es werden in jeder Modalität sofort semantische Konzepte extrahiert. Die Fusion auf dieser Ebene entwickelt aus den einzelnen semantischen Konzepten eine übergeordnete Bedeutung. Ein Vorteil dieser Methode ist, dass die

einzelnen Erkennen für Sprache, Gestik, etc. getrennt voneinander mit unimodalen Daten trainiert werden können.

Bezüglich der Nutzung kognitiver Ressourcen ist eine Unterteilung in aktive und passive Modalitäten interessant (Oviatt, 2003). Danach ist unter einer aktiven Modalität zu verstehen, dass der Nutzer intentional und explizit Kommandos zur Steuerung verwendet. Passive Eingabemodalitäten werden hingegen von einem technischen System ohne Zutun des Menschen registriert, beispielsweise unwillkürliche Gesichtsausdrücke, Lippenbewegungen beim Sprechen oder manche Gesten. Ebenso kann ein Gerät registrieren, ob sich der Nutzer innerhalb einer festgelegten Maximaldistanz befindet und nur dann z.B. die Spracherkennung aktivieren. Diese passiven Eingabeformen können bei multimodalen Schnittstellen genutzt werden, um die Erkennungsgüte zu verbessern, ohne bewusste Anstrengungen vom Nutzer zu fordern. Das indirekte Erfassen von Nutzerintentionen zur Ausführung von Systemaktionen schlägt auch Nielsen (1993) unter dem Stichwort *noncommand user interfaces* vor. Bezüglich der Interpretation weist Oviatt (2003) jedoch darauf hin, dass aktive Modalitäten verlässlicher sind. Sie tragen auch weniger das Risiko einer Über-Adaptierung, die aus Usability-Sicht äußerst problematisch ist.

2.2 Ingenieurpsychologische Konzepte

Das Thema und die Aufgabe der Ingenieurpsychologie ist die Untersuchung des menschlichen Verhaltens mit dem Ziel einer Verbesserung der Interaktion zwischen Mensch und Maschine. Sie kann nach Hoyos (1990) in der Anwendung der Psychologie auf die Nutzung, Steuerung und Wartung ingenieurwissenschaftlicher Produkte gesehen werden, damit als Teildisziplin der Angewandten Psychologie. Diese Anwendung geht dabei über ein einfaches empirisches Vergleichen technischer Lösungen hinaus. Die Ingenieurpsychologie arbeitet aktiv mit dem Wissen um die Fähigkeiten und Beschränkungen des Menschen, um gute Gestaltungslösungen bereits im Vorfeld eingrenzen oder sogar a priori aus theoretischen Grundlagen ableiten zu können (Poulton, 1966, zitiert nach Wickens & Hollands, 2000). Eine ingenieurpsychologische Herangehensweise an Multimodalität erfordert daher die Klärung einiger grundlegender Konzepte und Modelle.

2.2.1 Belastung und Beanspruchung

Durch ingenieurpsychologische Maßnahmen bei der Systemgestaltung soll zuvorderst erreicht werden, dass sich der menschliche Nutzer weder geistig noch körperlich bei der Bearbeitung einer Aufgabe überanstrengt. Das ist entscheidend, denn eine zu star-

ke Belastung kann zu Einbußen bei der Qualität der Bearbeitung führen. So vereinfacht sich unter Belastung die Denk- und Handlungsstruktur des Menschen und es kann zu Informationsverzerrungen kommen (Reason, 1994). Darüber hinaus steigt der Anteil suboptimaler Entscheidungen, da zu Gunsten einer schnelleren Entscheidungs- oder Handlungsausführung die Genauigkeit reduziert wird (Wickens & Hollands, 2000). Auf die Anforderungen der gleichen Aufgabe reagieren die Menschen individuell unterschiedlich. Das Belastungs-Beanspruchungskonzept erfasst diese differentielle Betrachtung theoretisch (Rohmert, 1984, 1986). *Belastung* ist danach die objektiv messbare, von außen einwirkende Größe, zum Beispiel die Häufigkeit durch den Operator zu prüfender Systemausgaben. *Beanspruchungen* sind die aus der auftretenden Belastung resultierenden Reaktionen auf subjektiver Ebene. Inwiefern eine Belastung als schwerwiegend empfunden wird, hängt von den psychischen und physiologischen Voraussetzungen der Person ab. Somit kann das objektiv gleiche Maß an Belastung bei verschiedenen Personen eine unterschiedlich hohe Beanspruchung auslösen. Das allgemeine Ziel der ergonomischen Forschung ist nicht, jegliche Beanspruchung zu minimieren, sondern vielmehr ein sinnvolles Niveau einzurichten. Bereits Anfang des 20. Jahrhunderts konnten Yerkes und Dodson einen umgekehrt U-förmigen Zusammenhang zwischen Aktivierung und Leistungsfähigkeit (Yerkes & Dodson, 1908) belegen. Bei Überforderung, aber eben auch bei Unterforderungen, sinkt die Leistung. Eine Minimierung der Beanspruchung kann jedoch zum Primärziel werden, wenn es um eine Nebenaufgabe in einer Mehraufgabensituation geht. Zum Schutz der Hauptaufgabe müssen Nebenaufgaben so gestaltet sein, dass sie möglichst wenige Ressourcen abziehen.

Auch in der deutschsprachigen Literatur wird Begriff *Workload* in zunehmendem Maße verwendet. Einige Autoren argumentieren, dass unter *Workload* sowohl die von außen einwirkenden Belastungsgrößen und -faktoren als auch die subjektiven Reaktionen ohne scharfe Trennung gefasst werden (z.B. Kannheiser, 1990). Die meisten Definitionen ordnen *Workload* jedoch deutlich der Beanspruchung zu (vgl. Gopher & Donchin, 1986; Moray, 1982; O'Donnell & Eggemeier, 1986; Wickens, 2002). So erklären Gopher und Donchin (Gopher & Donchin, 1986, S.41-3) den für die Ingenieurpsychologie besonders wichtigen Unterbegriff der mentalen *Workload* folgendermaßen: „[...] mental workload may be viewed as the difference between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time.“

Drei Klassen von Messverfahren für Beanspruchung bzw. *Workload* können unterschieden werden: subjektive Maße, physiologische Maße und Leistungsmaße (Eggemeier, 1988; O'Donnell & Eggemeier, 1986). Trotz der weit entwickelten verschiedenartigen Techniken innerhalb dieser Klassen lassen sich einige allgemeine Charakteristika der Maße festhalten (vgl. Eggemeier, 1988). Alle subjektiven Prozedu-

ren beinhalten einen Bericht der Nutzer über die erlebte Anstrengung bzw. den gefühlten Aufwand, meist in Form von Rating-Skalen. Physiologische Maße leiten das Workload-Niveau von einer physiologischen Reaktion des Nutzers während der Aufgabenbearbeitung ab (z.B. Hautleitwiderstand oder Puls). Leistungsmaße gehen vom Ergebnis der Aufgabenbearbeitung aus und können als Maße der Primär- oder der Sekundäraufgabe registriert werden. Maße der Primäraufgabe nutzen die Güte der Leistung für die Aufgabe oder Systemfunktion, die im Fokus der Untersuchung steht. Wird gleichzeitig zur Hauptaufgabe eine Nebenaufgabe bearbeitet, kann ein Leistungsabfall bei der Nebenaufgabe zur Abschätzung einer erhöhten Beanspruchung in der zu schützenden Hauptaufgabe herangezogen werden (*subsidiary task paradigm*). Ebenso kann die Anweisung an den Nutzer ergehen, die Leistung in der Nebenaufgabe konstant zu halten (*loading task paradigm*). In diesem Fall werden Änderungen der Leistung in der Primäraufgabe beobachtet und zur Abschätzung der Workload verwendet. Techniken aller Kategorien werden im praktischen Umfeld mit unterschiedlichem Erfolg eingesetzt (O'Donnell & Eggemeier, 1986). Zu den Nachteilen der subjektiven Maße zählt, wie bei allen subjektiven Berichten, dass sie gegenüber Einstellungen und Erwartungen anfällig sind. Die Notwendigkeit, Elektroden oder andere Einrichtungen zur Registrierung relevanter Parameter am Körper der Versuchsperson anzubringen, macht physiologische Maße für manche Aufgaben ungeeignet und verringert die Akzeptanz der Teilnehmer.

2.2.2 Informationsverarbeitung und Aufmerksamkeit

Die Fähigkeit eines menschlichen Nutzers zur Interaktion mit Mensch-Maschine-Schnittstellen und die dabei erlebte Beanspruchung stehen in engem Zusammenhang mit Prozessen der Informationsverarbeitung bei dieser Interaktion. Der Mensch ist kein passiver Empfänger von Reizen oder Information, sondern ein aktives dynamisches System, das Information mit Flexibilität aufsucht, aufnimmt, verarbeitet und umsetzt (Muthig, 1990). Das Vier-Stufen-Modell der Informationsverarbeitung von Wickens (Wickens, 1992; Wickens & Hollands, 2000) repräsentiert eine allgemeine Architektur kognitiver Funktionen, auf der jede Form von Handlung basiert (vgl. Abbildung 4).

Die Grundannahmen des Modells lassen sich direkt aus der Grafik erschließen. Information in Form verschiedener Reize aus der Umwelt wird in einer Reihe von Stufen verarbeitet. Die Reize werden zunächst durch Sinnesorgane aufgenommen und kurz zwischengespeichert (Stufe 1: *sensory processing*). Wickens geht davon aus, dass die reine sensorische Verarbeitung in der Stufe der Signalerkennung weitgehend ohne Ressourcen auskommt. Die Reize werden weiterverarbeitet durch einen Prozess der Mustererkennung, d.h. der Wahrnehmung (Stufe 2: *perception*), in dem sie organisiert

und identifiziert werden. Die Verarbeitung auf der Stufe der Wahrnehmung hat zwei wichtige Eigenschaften. Erstens erfolgt die Verarbeitung schnell und automatisch mit geringer Ressourcenforderung. Zweitens unterliegt sie sowohl Bottom-Up Prozessen durch die Reize selbst, als auch Top-Down Prozessen durch Beiträge des Gedächtnisses im Sinne von Erwartungen darüber, welche Ereignisse wahrscheinlich sind. Basierend auf dem Resultat wird ein Prozess der Entscheidung und Reaktionsauswahl (Stufe 3: *response selection*) initiiert, der letztlich in der Handlungsausführung (Stufe 4: *response execution*) mündet. In Anlehnung an das Arbeitsgedächtnismodell von Baddeley (1986) fasst Wickens als weiteren Einflussfaktor der Entscheidungsfindung Arbeitsgedächtnis und Kognition (zentrale Exekutive) zusammen, die Planungs- und Diagnosevorgänge unterstützen. Die zentrale Exekutive organisiert auch den Informationsaustausch mit dem Langzeitgedächtnis. Entscheidend ist für diese höheren Vorgänge, dass sie als bewusste Aktivitäten ablaufen, ressourcenlimitiert und hochanfällig für Interferenzen sind, wenn Aufmerksamkeitsressourcen für andere kognitive Prozesse abgezogen werden. Daraus lässt sich ableiten, dass eine substantielle Interferenz zwischen ressourcenfordernden Wahrnehmungsaufgaben und kognitiven Aufgaben auftreten wird, wenn beide das Arbeitsgedächtnis zur Speicherung und Manipulation von Information benötigen. Dieser Zusammenhang konnte experimentell belegt werden (Wickens, 2002). Durch die Rückkopplungsschleife können ausgeführte Handlungen überwacht und angepasst werden. Die Feedbackschleife weist auch darauf hin, dass es keinen festen Start- und Endpunkt der Informationsverarbeitung gibt: Der Prozess kann durch Input aus der Umwelt oder durch eine Handlungsintention des Menschen gestartet werden.

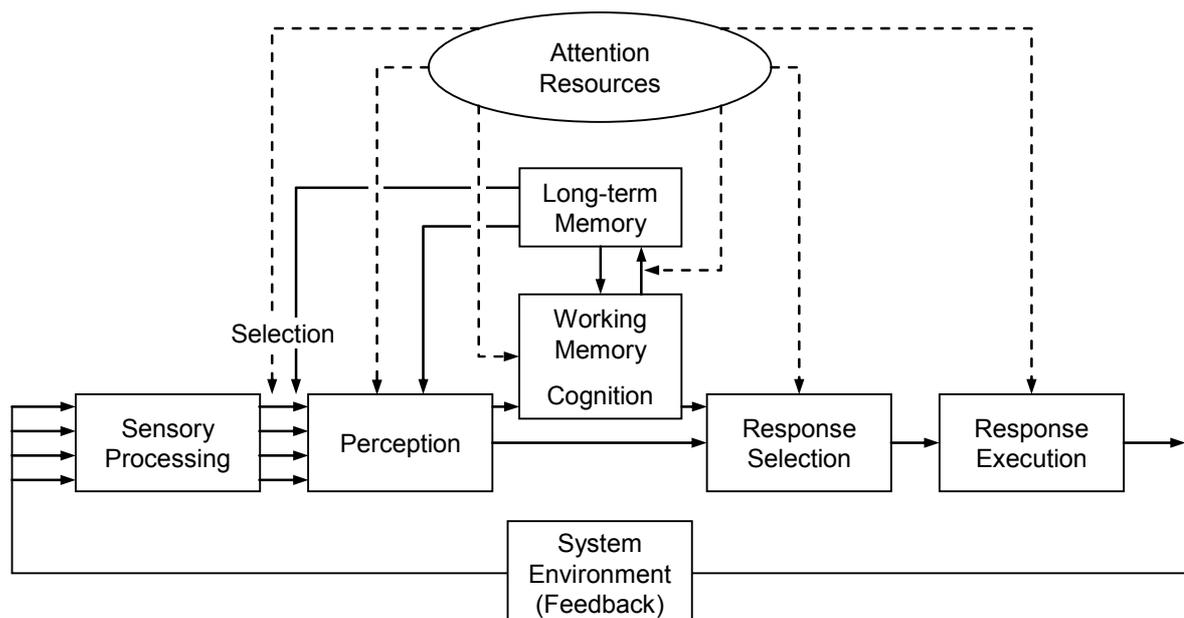


Abbildung 4: Das Vier-Stufen-Modell der Informationsverarbeitung nach Wickens (Wickens & Hollands, 2000, S.11).

Zentral für den Prozess der Informationsverarbeitung ist der Faktor Aufmerksamkeit, dem die bedeutsame Rolle der Selektion von Informationsquellen aus der Umwelt zufällt. Aufmerksamkeit wird oft mit der Suchscheinwerfer-Metapher veranschaulicht, die auf Wachtel (1967) zurückgeht. Neben der Ausrichtung des Scheinwerfers kann auch dessen Beleuchtungsbreite variiert werden. Ähnlich wird zwischen selektiver Aufmerksamkeit mit gebündelter Fokussierung auf einen Ausschnitt der Umwelt und geteilter Aufmerksamkeit bei gleichzeitiger Verarbeitung mehrerer Informationsquellen unterschieden. Für die Arbeit mit maschinellen Systemen ist besonders relevant, dass bei einer zu stark fokussierten selektiven Aufmerksamkeit die Gefahr des *cognitive tunneling* entsteht (Wickens & Hollands, 2000): Wenn das Ablesen einer Anzeige die Aufmerksamkeit zu stark bindet, können wichtige Informationen aus der Umwelt übersehen werden. Bei geteilter Aufmerksamkeit hängt die Qualität der Aufgabenbearbeitung maßgeblich von der Fähigkeit zu *time-sharing*, der Allokation von Bearbeitungsressourcen zwischen zwei oder mehr Aufgaben, ab.

Über die Funktionsweise von Aufmerksamkeit und Informationsverarbeitung, die sich nicht unabhängig voneinander betrachten lassen (Muthig, 1990), existiert eine große Bandbreite von Modellen (für eine umfassende Diskussion siehe Färber, 1987). Experimente mit dichotischer Stimulation führten in den 1950er und 1960er Jahren zu der Erkenntnis, dass die Aufmerksamkeit beträchtlich eingeschränkt ist, wenn sie zwischen mehreren unabhängigen Quellen auditiven verbalen Inputs alternieren muss. Die darauf aufbauenden Theorien nahmen alle einen Engpass (*bottleneck*) in der Informationsverarbeitung an, bei der ein Filter mit begrenzter Kapazität für eine bewusste Verarbeitung verantwortlich ist. Diese Theorien lassen sich danach unterscheiden, wo sich dieser Engpass befindet, d.h. in welcher Verarbeitungsstufe die bewusst zu verarbeitende Information selektiert wird. Nur die selektierte Information kann auf höherer Ebene weiter verarbeitet werden. Broadbent (1958) und Treisman (1960, 1964) gingen von einer frühen Selektion aus, wobei Broadbent einen blockierenden Filter auf der Ebene physikalischer Reizmerkmale als ausschlaggebend ansah. Treisman (1969) hingegen schrieb dem Filter keine blockierende, sondern lediglich eine abschwächende Wirkung zu, da sie in ihren Experimenten nachweisen konnte, dass eine Informationsfilterung neben physikalischen auch auf semantischen Merkmalen erfolgt. Einen anderen Ansatz verfolgen Modelle der späten Selektion (z.B. Deutsch & Deutsch, 1963; Norman, 1968). Sie nehmen an, dass alle Reize verarbeitet, semantisch analysiert und später auf Grundlage ihrer Relevanz selektiert werden. Der Engpass entsteht damit nicht auf sensorischer Ebene, sondern erst an dem Punkt der Verarbeitungssequenz, an dem eine Entscheidung für eine Reaktion initiiert wird.

Eine Wende in der theoretischen Betrachtung wurde in den 60er und 70er Jahren mit den Kapazitäts- oder Ressourcentheorien eingeleitet. Als einen der Ausgangspunkte kann die Forschung zu Leistung unter Single und Dual Task Bedingungen gesehen

werden. Die vorhandenen Theorien konnten Phänomene menschlichen Verhaltens in Mehraufgabensituationen nicht gut erklären. Knowles (1963, zitiert nach Wickens, 1984) stellte ein konzeptuelles Modell vor, das Aufmerksamkeit in den Zusammenhang mit einer begrenzten Verarbeitungskapazität bringt. Ein wichtiges Charakteristikum ist die Zuweisung der Höhe dieser Kapazität zwischen Aufgaben. Damit rückt die Verteilung der Verarbeitungskapazität in den Vordergrund, während die früheren Strukturtheorien davon ausgingen, dass die Aufmerksamkeit global zwischen den Aufgaben alterniert. Das Konzept der Kapazitätstheorien wurde mitunter durch den Beitrag von Kahneman (1973) verfeinert und auch quantitativ formuliert.

Die meisten dieser früheren Kapazitätsmodelle gingen jedoch davon aus, dass zwischen den Aufgaben nur eine einzige, unspezifisch-globale Kapazität graduell verteilt werden kann. Basierend auf der Erkenntnis, dass es zwischen einigen Aufgabenkombinationen keine, zwischen manchen Formen von Aufgaben jedoch starke Interferenz bei gleichzeitiger Bearbeitung auftritt, wurde die Idee mehrerer Ressourcen eingeführt (Kantowitz & Knight, 1976; Navon & Gopher, 1979). Wenn unterschiedliche, unabhängige Ressourcen durch zwei Aufgaben gefordert werden, lässt sich die ausbleibende Interferenz erklären. Darauf aufbauend formulierte Wickens (1984, 1991) ein elaboriertes Konzept der Informationsverarbeitung auf Grundlage mehrerer Ressourcen zur Erklärung der veränderlichen Leistung beim Time-Sharing zwischen zwei Aufgaben, das in Abbildung 5 dargestellt ist.

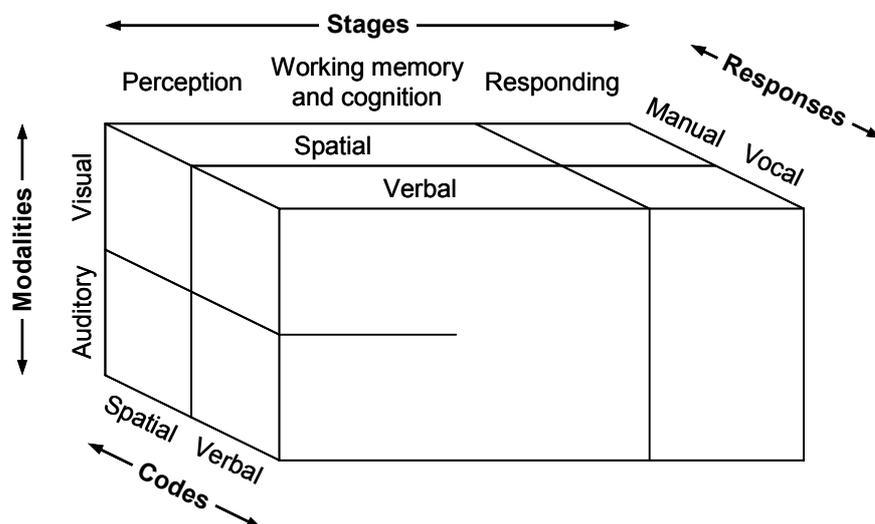


Abbildung 5: Dimensionale Repräsentation der Struktur multipler Ressourcen nach Wickens (vgl. Wickens & Hollands, 2000, S.449).

Das Modell der multiplen Ressourcen (*multiple resource model*, MRM) von Wickens geht von folgendem Verständnis des Begriffs „Ressource“ aus (Wickens, 1991, S.22): „a resource may be defined as one of a small set of scarce commodities within the human information processing system, which is associated with a distinct physio-

logical structure, and with physiological arousal changes as increased demands are placed on it.” Der Grundgedanke des Modells ist, dass bei einer Doppelaufgabe umso besseres Time-Sharing möglich ist, je stärker zwei Aufgaben unterschiedliche Ressourcen nutzen. Werden bei Nutzung unterschiedlicher Ressourcen die Anforderungen einer Aufgabe erhöht, sollte es mit geringerer Wahrscheinlichkeit zu einer Störung der anderen Aufgabe kommen. Wickens geht im MRM von drei strukturell dichotomen Dimensionen aus: den Verarbeitungsstufen (*stages*), den Verarbeitungstypen (*codes*) und den Wahrnehmungsmodalitäten (*modalities*). Die Dimension der Antwortmodalitäten bildet dabei keine eigene strukturelle Dimension, sondern wird als Konsequenz für eine verarbeitungskompatible Antwort verstanden. Die Grafik (Abbildung 5) soll andeuten, dass die Dimensionen in einem gewissen Maß unabhängig voneinander sind. Die unterbrochene Linie bei der Modalitätsdichotomie zwischen auditiven und visuellen Ressourcen drückt aus, dass der Vorteil durch eine Verteilung der Information auf mehrere Modalitäten nur auf perzeptiver Ebene sicher besteht. In welchem Ausmaß er bei der zentralen Verarbeitung (Arbeitsgedächtnis und Kognition) erhalten bleibt, ist nicht vollständig geklärt (vgl. Wickens, 1984). Weiterhin ist die Dimension der Verarbeitungsstufen nur durch zwei Ressourcen repräsentiert, da die Ressourcen für perzeptive und kognitive bzw. Arbeitsgedächtnisprozesse wahrscheinlich die gleichen sind. Sie unterscheiden sich jedoch von den Aktivitäten, die der Auswahl und Ausführung von Handlungen zu Grunde liegen. Wie bereits erwähnt wird angenommen, dass die rein sensorische Verarbeitung von Reizen weitgehend ressourcenfrei abläuft. Deshalb ist diese Stufe nicht repräsentiert.

Eine ausführliche Diskussion experimenteller Belege für die Modellannahmen bieten Wickens und Hollands (2000). Allerdings geben sie einschränkend zu bedenken, dass in vielen Fällen der relative Vorteil von modalitätsübergreifendem gegenüber intramodalem Time-Sharing nicht grundsätzlich das Ergebnis der Nutzung verschiedener Verarbeitungsressourcen sein muss. Darüber hinaus ist nicht immer eine bessere Leistung durch modalitätsübergreifende Informationsdarbietung nachweisbar (Gladstones, Regan & Lee, 1989; Wickens & Liu, 1988). Auch wenn das MRM eines der umfassendsten Modelle darstellt, wurden Aspekte identifiziert, die nicht hinreichend erfasst werden. So weisen Spence und Driver (Driver & Spence, 2004; Spence & Driver, 1999) darauf hin, dass die Rolle der räumlichen Verteilung von Reizen bei der Aufmerksamkeit viele experimentelle Resultate auch ohne die Annahme multipler Ressourcen erklären. Dieser Aspekt sollte entsprechend berücksichtigt werden. Ein weiterer viel versprechender Vorschlag zur Erweiterung des MRM kommt von Dahmen-Zimmer und Kollegen (1999) mit der zusätzlichen Annahme eines semantischen Bottlenecks.

Die Anwendung des MRM zum Verständnis der Interferenz zwischen Aufgaben in Mehraufgabensituationen und damit für die Übertragung auf multimodale Mensch-

Maschine-Interaktion muss einige Voraussetzungen berücksichtigen. Das Modell sagt vorher, dass das Ausmaß der Interferenz von dem Ausmaß gemeinsam genutzter Ressourcen auf allen drei Dimensionen abhängt. Damit kann nicht einfach gefolgert werden, dass zwei Aufgaben die auf einer dieser Dimensionen unterschiedliche Ressourcen nutzen, zu perfektem Time-Sharing führen werden. Des Weiteren können auch aus Sicht des MRM Verarbeitungsengpässe entstehen. Durch die Verwendung unterschiedlicher Modalitäten kann zwar ein Ressourcenkonflikt auf perzeptueller Ebene aufgelöst werden, jedoch ein Bottleneck auf kognitiver Ebene entstehen. Wickens und Hollands (2000) illustrieren diesen Zusammenhang folgendermaßen: Wenn dem Fahrer eines Kraftfahrzeugs Ausgaben in Form von Tönen auditiv dargeboten werden, führt das auditive Display zu einer geringen perzeptuellen Interferenz mit der Fahraufgabe. Wenn jedoch für die Rekonstruktion der Bedeutung größere kognitive Anstrengungen erforderlich sind, kann das die Vorteile der Verwendung unterschiedlicher Wahrnehmungsmodalitäten aufheben. Weiterhin wird häufig die Rolle der Kodierung bei der Gestaltung multimodaler Schnittstellen vernachlässigt. Wird die Information für eine Aufgabe visuell in Textform und für eine andere Aufgabe auditiv, jedoch ebenfalls sprachlich präsentiert, ist eine deutliche Interferenz die Folge. Relevant ist für die Anwendung darüber hinaus, dass automatisierte Prozesse weniger Ressourcen benötigen (Kahneman, 1973) und dass diese selbst in Kombination mit komplexen zusätzlichen Aufgaben zu gutem Time-Sharing führen (Schneider & Fisk, 1982).

Zusammengefasst ist zu erkennen, dass die Voraussagen des MRM nicht derart generell formulierbar sind, dass es als grundlegende Theorie für die multimodale Mensch-Maschine-Interaktion einsetzbar ist. Ferner sind die Aussagen beschränkt auf auditive und visuelle Wahrnehmung bzw. manuelle und sprachliche Reaktionen. Andere relevante Wahrnehmungsmodalitäten wie Haptik oder beispielsweise die Möglichkeit zu einer Steuerung durch Blickbewegungen (*eye gaze input*) sind nicht berücksichtigt. Deshalb hat die einleitende Aussage dieses Kapitels nach wie vor Gültigkeit, nach der technologische Lösungen zu multimodaler Mensch-Maschine-Interaktion stets durch eine experimentelle Prüfung abgesichert werden müssen. Dennoch leisten die Modelle der Informationsverarbeitung einen entscheidenden Beitrag, indem sie mit der Struktur kognitiver Prozesse die relevanten Aspekte zur Prüfung technischer Realisierungen vorgeben.

2.2.3 Allgemeine Fehlerklassifikation

Als typisch menschliches Merkmal muss stets bedacht werden, dass bei der Handlungsausführung oder der Handlungsplanung Fehler entstehen können (Stanton, 2003). Ein grundlegender Bestandteil von gebrauchstauglichen Systemen ist deshalb, dass sie

durch umfassende Rückmeldung die Fehlerwahrscheinlichkeit verringern und die Möglichkeit zu Überprüfung und Korrekturen bieten. Multimodale Systeme können an dieser Stelle einen entscheidenden Beitrag leisten, indem sie eine verbesserte Fehlererkennung durch Feedback in einer für die Aufgabe angemessenen Modalität oder sogar redundant über mehrere Modalitäten ermöglichen. Zur Fehlervermeidung können sie besser als herkömmliche Systeme beitragen, wenn sie für verschiedene Aspekte einer Aufgabe die jeweils optimal geeignete und angepasste Eingabemodalität zur Verfügung stellen. Norman (1998) betont, dass Fehler in der Interaktionen mit Maschinen nur allzu häufig dadurch entstehen, dass Menschen unnatürliche und nicht nachvollziehbare Handlungen zur Bedienung von Systemen ausführen müssen.

Eine klare Fehler-Taxonomie hilft nicht nur bei einer späteren Analyse von Handlungsfehlern nach einem Vorfall, sondern kann bereits im Vorfeld hilfreich sein, um Fehlerpotential im Systemdesign aufzudecken. Reason (1994, S.27ff.) geht bei seiner Klassifikation von Fehlertypen in Anlehnung an die Modellannahmen von Rasmussen (1983, 1986) vor, die er als gefährdende Verhaltensweisen konzeptualisiert. Dabei unterscheidet er zunächst zwischen beabsichtigten und unbeabsichtigten Fehlhandlungen. Unter die nicht intendierten Verhaltensweisen fallen die Aufmerksamkeitsfehler und Gedächtnisfehler. Aufmerksamkeitsfehler (*slips*, Patzer) entstehen auf der Ebene der fähigkeitsbasierten Ausführung einer Handlung, wenn eine notwendige Überwachung der Handlung durch Zuwendung von Aufmerksamkeit ausgelassen wird oder die Aufmerksamkeitskontrolle zu einem nachteiligen Zeitpunkt erfolgt. Ebenfalls auf dieser Ebene kommen Gedächtnisfehler (*lapse*, Schnitzer) zum Tragen. Sie entstehen, wenn Zwischenschritte bei längeren Handlungssequenzen ausgelassen werden, der Abruf temporär relevanter Information scheitert oder Rahmenbedingungen bei der Handlungsausführung vergessen werden. Die Zielsetzungsfehler (*mistake*) sind den Ebenen des regelbasierten und des wissensbasierten Verhaltens zuzuordnen und zählen zu den intendierten Handlungen. Als absichtlich gelten sie deshalb, weil korrekt dem auszuführenden Handlungsplan gefolgt wird. Der Fehler liegt in der Auswahl der Handlung. Ein regelbasierter Fehler besteht in der Anwendung einer schlechten Regel, die inhaltlich falsch ist, oder auf der fehlerhaften Anwendung einer an sich guten Regel. In letzterem Fall wird nach einer Regel vorgegangen, die sich in anderen Situationen als nützlich erwiesen hat, für die aktuelle Aufgabe aber nicht angemessen ist. Wissensbasierte Fehler entstehen durch ein Wissensdefizit aufgrund der Komplexität der Aufgabe, einem fehlenden Verständnis kausaler Zusammenhänge oder einem unvollständigen bzw. unangemessenen mentalen Modell des Anwenders für den Problemraum. Die drei Fehlertypen Aufmerksamkeitsfehler, Gedächtnisfehler und Zielsetzungsfehler bilden die so genannten grundlegenden Fehlertypen im Modell der gefährdenden Verhaltensweisen. Eine besondere Form beabsichtigter Fehlhandlungen sind Verstöße (*violations*), da das Ergebnis der Handlung mit dem angestrebten Ziel-

zustand übereinstimmt. Diese intendierten Handlungen liegen zum Beispiel vor, wenn absichtlich Vorschriften oder Regeln übertreten wurden und haben sich bei der Analyse von Unfällen als unverzichtbar erwiesen.

Im Themengebiet der Mensch-Maschine-Interaktion ist ein Fehler, der bei der Bildung der Intention auftritt, besonders virulent: der Modusfehler (vgl. Norman, 1988, 1990). Zimmer (2001b, S.43) illustriert den Modusfehler anhand einer Fehleinschätzung des Zustands eines Flugmanagement-Systems. Wenn der Pilot annimmt, das Flugzeug befände sich im Modus *descent*, es befindet sich aber tatsächlich im Modus *level flight*, würde eine Eingabe des Werts -3 nicht wie intendiert umgesetzt. Statt einem Sinkwinkel von 3° würde eine Reduktion der Flughöhe um 3000 ft ausgeführt. Gerade zur Vermeidung des Modusfehlers hat sich multimodales Feedback als vorteilhaft erwiesen. So konnte in einem Experiment durch kinästhetische und visuelle Rückmeldung die Anzahl der Modusfehler deutlich reduziert werden (Sellen, Kurtenbach & Buxton, 1992). Ferner lassen sich bei der Eingabe durch die Einbeziehung alternativer Eingabekanäle ungünstige Doppelbelegungen vermeiden. Erst durch Doppelbelegungen der Eingabeinstrumente entsteht die Notwendigkeit eines Moduswechsels.

2.3 Grundlagen multimodaler Mensch-Maschine-Interaktion

Das vorrangige Ziel dieses Abschnitts ist die Besprechung technologischer Rahmenbedingungen sowie psychologischer bzw. physiologischer Grundlagen multimodaler Mensch-Maschine-Interaktion. Dazu werden neben einem kurzen Abriss möglicher Einsatzbereiche die modalitätsspezifischen Charakteristika ausgewählter Interaktionsmodalitäten dargelegt. Die Einteilung und Begrifflichkeit der Gliederung (visuelles, auditives und haptisches Interface) orientiert sich an der Schnittstellendefinition nach Hedicke (2000), wie sie in Abschnitt 2.1.2.2 beschrieben ist. Der Schwerpunkt liegt auf den Ein- und Ausgabemodalitäten, die gegenwärtig bereits ein gewisses Stabilitätsniveau erreicht haben und so Alternativen zu visuell-manueller Interaktion ermöglichen. Nachdem visuelle Ausgaben und manuelle Eingaben Bestandteil nahezu aller multimodaler Lösungen ist, werden auch diese Themen berührt.

Die umfangreichen Ergebnisse sinnesphysiologischer und psychophysischer Untersuchungen der menschlichen Sinnesleistungen und ihre Bedeutung für die Mensch-Maschine-Interaktion haben bereits seit langem Eingang in die entsprechenden Handbücher gefunden (vgl. Muthig, 1990). Die Zusammenstellung im Rahmen dieser Arbeit erhebt keinen Anspruch auf Vollständigkeit, sondern soll nur einige relevante

Funktionen und Beschränkungen der Sinnessysteme aufzeigen. Im Anschluss an die modalitätsspezifischen Grundlagen wird auf deren konzertiertes Zusammenwirken bei multimodaler Eingabe und Wahrnehmung mit einer Diskussion technologischer Vor- und Nachteile und experimenteller Befunde eingegangen. Der Abschnitt schließt mit einem Verweis auf verfügbare Gestaltungsrichtlinien.

2.3.1 Einsatzgebiete und Entwicklung multimodaler Systeme

Multimodale Systeme haben sich in den letzten Jahren insbesondere aufgrund massiver technologischer Fortschritte stark weiterentwickelt. Dabei ist eine generelle Tendenz zu stärkerer Robustheit und Natürlichkeit insbesondere bei der Kombination mit Spracheingabe zu beobachten (Benoît et al., 2000; Kaber, Wright & Sheik-Nainar, 2006; Oviatt, 2003; Reeves et al., 2004). Als einer der Startpunkte moderner multimodaler Schnittstellen wird meist der konzeptuelle „Put That There“-Prototyp von Bolt (1980) genannt. Bolt verknüpfte Spracheingabe mit einfachen deiktischen Gesten, Systemausgaben wurden auf einer großformatigen Leinwand realisiert. Die Befehle konnten rein sprachlich geäußert werden (z.B. „make the blue triangle smaller“), sofern sie eindeutig waren. Durch multimodale Fusion konnten die Nutzer jedoch auch Befehle wie „make that smaller“ äußern, wobei das durch „that“ referenzierte Zielobjekt durch eine Zeigegeste auf das entsprechende Element auf der Leinwand spezifiziert wurde. Durch „Put That There“ wurde eine neue Ära multimodaler Interaktion eingeläutet, da es zum ersten Mal den Aspekt der Fusion mehrerer Eingabemodalitäten demonstrierte.

Dagegen befinden sich multimodale Ausgaben für komplexe Systeme seit langem in der Anwendung. Bereits sehr früh haben Wahrnehmungs- und Vigilanz-Experimente gezeigt, dass eine redundante Ausgabe visueller und akustischer Warnsignale zu höherer Detektionswahrscheinlichkeit führen kann (Buckner & McGrath, 1963; Howarth & Treisman, 1958). In Flugzeugen und Anlagen wurden akustische Warnungen bereits frühzeitig zur Unterstützung des visuellen Kanals eingesetzt, seit Beginn der 1960er Jahre wurden auch taktile Ausgaben vorgeschlagen (Van Laer, Galanter & Klein, 1960). Ein weiterer wichtiger Anwendungsbereich multimodaler Ausgaben kristallisierte sich seit Beginn der 1970er Jahre systematisch heraus: Das Verfügbarmachen von Information für Menschen mit wahrnehmungsseitiger Behinderung. Beispielsweise wurden visuell-taktile Trainingshilfen zum Erlernen des Sprechens für Gehörlose diskutiert (Nickerson & Stevens, 1972). Bei allen frühen Anwendungen multimodaler Ausgabe handelte es sich üblicherweise um einfache zusätzliche Signale. Komplexere Formen der synergistischen Verwendung von Schallen, auch Sprachausgaben, entstanden erst im Bereich der Mensch-Computer-Interaktion in den 1980er Jahren (vgl. Brewster, 2003). Trotz früher interessanter Pro-

totypen (z.B. Optacon, Craig, 1977) war eine wesentliche Entwicklung der technologisch deutlich aufwändigeren haptischen und taktilen Ausgaben erst ab den 1990er Jahren zu beobachten (z.B. Akamatsu & Sato, 1994). Mit der technologischen Weiterentwicklung konnten akustische und haptische Anzeigen neben Warnsignalen auch zunehmend komplexe Information transportieren. Ein bedeutsames Einsatzgebiet ist seitdem die Optimierung von Interaktionssituationen, die eine starke Last auf eine bestimmte Wahrnehmungsmodalität ausüben. Prototypisch kann hier die Fahrzeugführung als maßgeblich visuell geprägte Aufgabe genannt werden. Durch multimodale Ausgabe können Nachrichten übermittelt werden, ohne eine weitere Belastung des visuellen Kanals zu induzieren. Neben den genannten Einsatzgebieten bieten sich multimodale Ausgaben allgemein dann an, wenn besondere oder schwer erfüllbare Anforderungen an die visuellen Anzeigen bestehen. So gestatten mobile Endgeräte und sog. *wearable devices* kaum eine Verwendung größerer, gut ablesbarer Displays.

Mit der Entwicklung der Spracherkennungstechnologie ab den 1980er Jahren und den massiven Fortschritten in den 1990er Jahren bot sich Spracheingabe bei der Mensch-Maschine-Interaktion als Alternative zur Texteingabe über die Tastatur an. Nach Oviatt (2003) bestehen die frühesten und einfachsten multimodalen Eingabesysteme deshalb aus einer Kombination der Standardkomponenten eines Computers (Keyboard, Maus) und Spracheingabe. Dabei wurde in den neueren Systemen neben einfacher kommandosprachlicher Eingabe auch die Verwendung natürlichsprachlich formulierter Eingaben ermöglicht. Zu den komplexeren Systemen gehörten Ende der 1980er Jahre viele Anwendungen zur Interaktion mit Landkartenmaterial, bei denen ähnlich zu „Put That There“ Sprache und Zeigegesten verknüpft werden konnten (z.B. Neal & Shapiro, 1991; Wahlster, 1991). Durch weitere Fortschritte bei der Vorgehensweise zur Fusion paralleler Eingabewege konnten neue Modalitätskombinationen gefunden werden, bei der auch semantisch reichere Information übertragen wird. Dabei werden beispielsweise Sprache und Gesten mit Hilfe eines Stifts (*pen, stylus*) auf einem geeigneten Touchscreen kombiniert (eine ausführliche Übersicht bieten Oviatt et al., 2000) oder Sprachkommandos und Lippenbewegungen zur Erhöhung der Erkennungsqualität erfasst (Benoît et al., 2000). Multimodale Systeme mit Sprachverarbeitung und einfachen zweidimensionalen Gesten zur Steuerung von Komfortfunktionen für Fahrzeuge wurden prototypisch umgesetzt (vgl. Akyol, Libuda & Kraiss, 2001). Die Integration dreidimensionaler Gesten ist hingegen noch nicht weit fortgeschritten (Cohen et al., 1999). Neben Gestenerkennung sind insbesondere die Registrierung von Blick- oder Kopf-/Körperbewegungen viel versprechende videobasierte Verfahren. Gerade in Umgebungen mit hoher Lärmbelastung kann dieser Kanal eine deutliche Erweiterung der Interaktionsmöglichkeiten bieten. Drei Haupteinsatzgebiete lassen sich für Systeme mit multimodaler Eingabe allgemein nennen: Erstens das Ermöglichen der Nutzung technischer Systeme für Menschen mit körperli-

chen Behinderungen durch die Verfügbarkeit alternativer Eingabewege; zweitens eine Steigerung der Effizienz bei der Bearbeitung mehrerer Aufgaben durch Nutzung unabhängiger Interaktionsmodalitäten; drittens die Simulation virtueller Umgebungen, in denen möglichst natürliche und realitätsähnliche Eingabeformen angestrebt werden.

2.3.2 Visuelles Interface

Für den Informationsfluss von der Maschine zum Menschen weist die Gestaltung der visuellen Schnittstelle die längste Tradition auf. Das visuelle System des Menschen nimmt die dargestellte Information (z.B. über Displays oder Kontrollleuchten) auf und verarbeitet sie. Die Eigenschaften dieses Systems müssen bei dem Design der Bedienoberfläche berücksichtigt werden. Neben der Ausgabe von Information an den Menschen kann auch eine Eingabe durch den menschlichen Nutzer an die Maschine über videobasierte Systeme erfolgen. So können beispielsweise Handbewegungen als Gesten durch Mustererkennung identifiziert oder aus Augenbewegungen Steuerimpulse abgeleitet werden.

2.3.2.1 Visuelle Wahrnehmung und visuelle Systemausgaben

Visuelle Wahrnehmung ist das Resultat einer komplexen Wechselwirkung sensorischer und motorischer Leistungen des Auges und des zentralen Nervensystems (Schmidt & Thews, 1997). Das visuelle System dient der Farb-, Objekt- und Bewegungswahrnehmung sowie der Wahrnehmung von räumlicher Tiefe und Größe. Nach Muthig (1990) bietet es unter funktionaler Perspektive damit Orientierung über Helligkeit, Farbigkeit, Größe, Form, Raumlage, Bewegung und die Entfernung und Richtung von Objekten und Flächen. Das visuelle System füllt sowohl bei der Steuerung der Fortbewegung im Raum als auch bei nicht an der Lokomotion beteiligten Körperbewegungen eine wichtige Rolle aus. Ferner ist eine unterstützende Funktion in der Kommunikation, beispielsweise bei der Identifikation von Gestik und Mimik eines Gesprächspartners, festzuhalten. Besonders relevant für die Informationsaufnahme bei der Mensch-Maschine-Interaktion ist die Verarbeitung linguistischer Information, z.B. beim Lesen.

Die folgende Darstellung der visuellen Wahrnehmung und ihre Bedeutung für die MMI orientiert sich an den zusammenfassenden Überblicksdarstellungen von Geiser (1990), Goldstein (1997), Helander (1987), Megaw (1992) und Muthig (1990). Die Wahrnehmung erfolgt durch Photorezeptoren (Stäbchen und Zapfen) im Auge, deren Aufgabe die Umsetzung von Licht (elektromagnetische Schwingungen von ca. 380-780 nm) in elektrische Signale ist. Zum Ablesen von Anzeigen ist eine Abbildung auf der Zone des schärfsten Sehens, der Fovea, auf der Netzhaut notwendig. Die Fovea liegt exakt auf der Blicklinie, sodass das Bild eines direkt betrachteten Objekts genau

in diesen Bereich fällt. Die Ausdehnung beziffert Geiser (1990) mit $1-2^\circ$ und gibt zur Veranschaulichung folgende Faustregel: Bei ausgestrecktem Arm wird in etwa die Größe des Daumnagels scharf gesehen. Dieser relativ kleine Bereich ist ausschließlich durch Zapfen besetzt. Die übrige periphere Netzhaut weist sowohl Stäbchen als auch Zapfen auf. Zusammengefasst sind zapfenförmige Photorezeptoren für das Sehen bei heller Beleuchtung, das Farbsehen und das Detailsehen verantwortlich, während stäbchenförmige Photorezeptoren im Dunkeln äußerst empfindlich reagieren, jedoch keine feinen Details auflösen können. Wenn der Blick stabil auf einem Objekt liegt, spricht man von einer Fixation, die abhängig von der Komplexität des Objekts mindestens 200-600 ms dauert. Die ruckartigen Augenbewegungen zwischen Fixationen werden Sakkaden genannt und erstrecken sich über einen Bereich von bis zu 40° , im Schnitt jedoch $15-20^\circ$, mit einer Dauer von 40-100 ms. Zwischen zwei Sakkaden liegt eine Latenzphase von 100-200 ms. Bei Fixationen wird das Auge durch Akkommodation an die Entfernung eines Objekts angepasst. Die Brechkraft der Linse wird durch muskelgesteuerte Verformung reguliert und durch eine Veränderung der Pupillenweite und des Konvergenzwinkels der beiden Sehachsen begleitet. Der (altersabhängige) Zeitbedarf zur Akkommodation ist für Naheinstellung geringer als für Ferneinstellung und liegt zwischen 0.2 und 1.5 s. Für das sichere Erkennen von Zeichen ist ein Mindestsehwinkel von 12-15 Winkelminuten einzuhalten. Ein weiterer wichtiger sinnesphysiologischer Vorgang ist die Adaptation, die Anpassung an die Helligkeit und den Hintergrund eines Objekts, d.h. die Leuchtdichteverhältnisse im Gesichtsfeld. In einem Leuchtdichtebereich von 10^{-6} bis 10^5 cd/m^2 kann der Mensch Stimuli erkennen und unterscheiden, dabei bleiben ab einer Leuchtdichte von ca. 100 cd/m^2 Sehschärfe und Unterschiedempfinden weitgehend konstant. Beteiligt sind daran die motorische Veränderung der Pupillenweite durch die Iris, der Übergang von der Aktivierung der Zapfen bei Helligkeit zu Stäbchen bei Dunkelheit (bzw. umgekehrt) und die Empfindlichkeitsänderung von Zapfen und Stäbchen. Die benötigte Zeit für diesen Vorgang hängt vom Leuchtdichteunterschied und dem der Richtung (hell nach dunkel bzw. dunkel nach hell) ab. Allgemein dauert die Anpassung umso länger, je größer der Unterschied in der Leuchtdichte ist. Darüber hinaus benötigt die Dunkeladaptation deutlich länger als die Helladaptation. Die Stäbchenanpassung tritt erst nach einigen Minuten auf, wenn die Stäbchen empfindlicher als die Zapfen geworden sind. Das Empfindlichkeitsmaximum der Stäbchen ist erst nach 20 bis 30 Minuten erreicht. Geiser (1990) folgert aus diesem Sachverhalt für Mensch-Maschine-Schnittstellen, dass häufige Fixationswechsel zwischen mehreren Anzeigen mit hohen Leuchtdichteunterschieden vermieden werden müssen. Eine gute tabellarische Zusammenfassung der Eckdaten des visuellen Systems zur Gestaltung visueller Displays bietet Charwat (1994, S.392).

Jede Form der Wahrnehmung ist mehr als eine rein physiologische Verarbeitung von Reizinformation. Durch kognitive Einflüsse entsteht ein aktiver Prozess mit konstruktiven Komponenten, der die Information in der physikalisch gegebenen Umwelt anreichert. Das wird besonders gut bei Phänomenen der Objektwahrnehmung ersichtlich, auf die wegen ihrer hervorstechenden Bedeutung für die Anordnung von Elementen auf Anzeigen kurz eingegangen werden soll. Um die Organisation dieser Elemente für die Wahrnehmung möglichst natürlich zu realisieren, wird häufig auf Gestaltgesetze zurückgegriffen. So betont beispielsweise Helander (1987) die Möglichkeiten einer Nutzung der Gesetze der Nähe, der Ähnlichkeit, der Geschlossenheit, der gestaltgerechten Linienfortsetzung und des Gesetzes des gemeinsamen Schicksals, um Gruppenzugehörigkeit zu betonen oder Beziehungen zwischen Anzeigeelementen hervorzuheben. Für die Detailgestaltung ist ebenso wichtig, Wissen über geometrisch-optische Täuschungen bei der Wahrnehmung von Größe und räumlicher Tiefe zu berücksichtigen (z.B. Müller-Lyersche Täuschung, Ponzo-Täuschung, Oppel-Kundt-Täuschung, Poggendorf-Täuschung, Scheinbewegungen, Amesscher Raum), um eine Fehlinterpretation dargestellter Größen und Verhältnisse von Elementen auf einem Computerbildschirm zu vermeiden.

Die zur Verfügung stehenden Technologien zur visuellen Ausgabe reichen von einzelnen Signallampen über LEDs bis hin zu komplexen Anzeigeeinheiten wie alphanumerischen 7-Segment-Displays, Punkt-Matrix-Displays und Bildschirmen. In den letzten Jahren ist ein zunehmender Einsatz von Flüssigkristallbildschirmen (LCD, *liquid crystal display*) zu beobachten. Dabei wurden nicht nur Röhrenbildschirme (CRT, *cathode ray tube*) abgelöst, sondern auch neue Gebiete der mobilen Mensch-Computer-Interaktion durch LCDs erschlossen (Luczak, Roetting & Oehme, 2003). Für spezielle Anwendungsfelder wie Augmented Reality wird die Form der visuellen Ausgabe besonders angepasst. So können über auf dem Kopf angebrachte Anzeigen (*head mounted displays*, HMD) zusätzliche Informationen über Objekte im Blickfeld überlagernd direkt in das Blickfeld eines Operators eingespiegelt werden. Das HMD kann dabei wie eine normale Brille konstruiert sein, bei der zwei LCDs integriert sind (Siewiorek & Smailagic, 2003). Meist auf experimentellem Niveau werden auch retinale Displays verwendet, die das Bild direkt auf die Netzhaut des Nutzers zeichnen. Unter den Spezialanwendungen sind noch autostereoskopische Displays zu nennen, die dem Nutzer einen dreidimensionalen Eindruck der Bedienoberfläche vermitteln. Diese sind auf keinen Fall mit dreidimensionalen Darstellungen auf normalen zweidimensionalen Medien zu verwechseln. Autostereoskopie bezeichnet eine Methode, mit der ohne Einsatz einer stereoskopischen Sehhilfe Stereobilder zu einem einzigen Raumbild verschmelzen. Das Stereobild besteht dabei aus einem Bildpaar, das aus zwei raumparallaktisch verschiedenen aber weitgehend inhaltsgleichen Halbbildern besteht. Eine weitere besondere Form der Projektionstechnologie wird bei den so

genannten Head-Up Displays (HUD) in Fahr- und Flugzeugen verwendet. Relevante Informationen wie die Geschwindigkeit werden beispielsweise auf die Frontscheibe eines Autos gespiegelt, so dass sie über der Fahrbahn zu schweben scheinen und mit deutlich geringerer Akkommodation abgelesen werden können.

2.3.2.2 Aspekte der Gestaltung von Gestik- und Blickeingabe

Die einfachste Form der Geste ist die Zeigegeste. Streng genommen ist bereits eine herkömmliche Computermaus ein Gerät zur Gesteneingabe, weil auf direkt-manipulativen Benutzeroberflächen durch das „Zeigen“ auf ein Icon, d.h. der Cursor wird darauf bewegt und das Element wird angeklickt, eine Auswahl für weitere Operationen erfolgen kann. Diese Form der Gesteneingabe, ebenso wie Gesten die mit einem Stift auf einem Tablet-PC oder auf einem Touchscreen erzeugt werden können, wird nicht der visuellen Schnittstelle zugeordnet, da sie nicht durch optisch-visuelle Erkennungsverfahren verarbeitet wird. Dynamische Zeigegesten oder auch statische Gesten, die von einem Computer durch visuelle Mustererkennung registriert werden, zeichnen sich dadurch aus, dass sie dem Nutzer eine berührungslose Interaktion mit dem System gestatten. Der Begriff der Geste ist allerdings bei der Mensch-Computer-Interaktion sehr weit gefasst. Nachdem in virtuellen Umgebungen Objekte nicht gegenständlich existieren und damit nicht real in die Hand genommen werden können, wird das Greifen nach einem Gegenstand in der Virtual Reality ebenfalls als Geste verstanden. In der realen Welt hingegen wäre es eine manipulative Handlung.

Für eine zuverlässige Gestenerkennung muss es gelingen, Gesten trotz individueller Unterschiede zu identifizieren und insbesondere Anfang und Ende einer Eingabeoperation abzustecken. Typischerweise werden deshalb nicht beliebige natürliche Gesten für die Interaktion mit einem System zugelassen, sondern ein Zeichenvorrat vorgegeben. Dadurch wird die Mustererkennung wesentlich vereinfacht. Die Siemens AG hat 1998 einen SIVIT genannten Prototyp vorgestellt (Siemens Virtual Touchscreen), bei dem für ein Informationsterminal die berührungslose Erfassung von Gesten videobasiert realisiert wurde (vgl. Maggioni & Kämmerer, 1998). Im Fahrzeug wurde das Potential einer Gestensteuerung mit Hilfe der Auswertung von Infrarotkamerabildern mit statischen Handgesten unterhalb der Mittelkonsole experimentell demonstriert (Akyol et al., 2001). Aktuelle Entwicklungen weisen viel versprechende Fortschritte auch bei dynamischen Gesten auf, z.B. können zur Unterstützung kleine unauffällige Sensoren in die Kleidung integriert werden und dabei Ankerpunkte für die videobasierte Erkennung liefern (Geer, 2004). Wie bei allen erkenntnisbasierten Technologien ist die Anwendung für sicherheitskritische Funktion dennoch nicht möglich, da aus einer fehlerhaften Erkennung Gefahrenpotential erwächst. Erste kommerzielle Produkte mit Gestenerkennung sind seit einigen Jahren auf dem Markt, zum Beispiel zur Präsentation von Wettervorhersagen (GestureStorm™, Cybernet Systems Corp.)

oder für Eingabefunktionen bei Videospiel-Konsolen (z.B. EyeToy®, Sony Computer Entertainment Inc.).

Aus psychologischer Sicht ist der Unterschied zwischen manipulativen und kommunikativen Gesten bedeutsam (vgl. Pavlovic, Sharma & Huang, 1997). Manipulative Gesten werden wie bereits angesprochen z.B. in der Virtual Reality für Handlungen an Objekten verwendet. Kommunikative Gesten haben vielfältige Funktionen und werden in natürlichen Situationen meist von Sprache begleitet. Pavlovic et al. (1997) differenzieren bei den kommunikativen Gesten zwischen Handlungen und Symbolen. Symbole sind Gesten mit linguistischer Bedeutung, die als Referenz auf ein Objekt oder einen Vorgang verweisen bzw. ihn modellieren. Handlungen sind Gesten, deren Interpretation sich direkt auf ihre Bewegung bezieht. Dabei imitieren mimetische Gesten Handlungen. Die Handlung bei deiktischen Gesten besteht in dem Vorgang des Zeigens auf ein Objekt. In der Literatur wird kontrovers diskutiert, inwiefern Gesten grundsätzlich intendiert sind. Während Pavlovic et al. (1997) absichtlich initiierte Gesten von unabsichtlichen Bewegungen ohne Bedeutung unterscheiden, deuten experimentelle Befunde darauf hin, dass Hand/Arm-Gesten als Begleitung der Sprachproduktion unwillkürlich erfolgen (Krauss & Hadar, 1999).

Systeme zur Registrierung der Blickbewegung (*eye tracking*) als Teil der visuellen Schnittstelle lassen sich in diagnostische und interaktive Systeme unterteilen (Duchowski, 2002). Die interaktive Steuerung von Funktionen im Rahmen multimodaler Eingabekonzepte kann gegenwärtig nicht zum typischen Anwendungsfall für eine Registrierung von Blickbewegungen gezählt werden. Die Verwendung für Leseexperimente in der psychologischen Grundlagenforschung stellt wahrscheinlich den ersten systematischen Einsatz von Eye Trackern in diagnostischer Funktion dar (vgl. Rayner, 1998). Blickbewegungsmessungen werden häufig in Usability-Tests eingesetzt, um den Verlauf der visuellen Informationsaufnahme durch den Nutzer auf grafischen Bedienoberflächen verfolgen zu können. Als Testinstrument kann die Blickbewegungsmessung auch Aufschluss über die kognitive Belastung geben, z.B. bei der Bewertung von Fahrerassistenz- und Fahrerinformationssystemen (Seifert, Rötting & Jung, 2001). Als Eingabemethode wurde Blickbewegungsmessung bereits vor langer Zeit vorgeschlagen (Bolt, 1982) und erste Evaluationen waren erfolgreich (Ware & Mikaelian, 1987). Die direkte Aktivierung von Funktionen durch Blickeingabe weist ein besonderes Potential für Menschen mit motorischen Beeinträchtigungen als alternativer Eingabeweg für die Mensch-Computer-Interaktion auf. Der Aufwand bei der Kalibrierung und der anfänglich notwendigen kopfstationären Montage sprach jedoch gegen einen Ausbau der Idee für einen allgemeinen Nutzerkreis. Moderne Eye Tracker sind in den Bildschirm integriert, wesentlich zuverlässiger und benötigen nur wenige Sekunden zur Kalibrierung. Neben den berührungsfreien videooptischen Verfahren, die z.B. den Cornea-Reflex mit Hilfe einer punktförmigen Infrarotlichtquelle erfassen,

existieren auch elektrophysiologische (EOG) und invasive Verfahren. Eine Übersicht über Methoden zur Registrierung von Augenbewegungen findet sich bei Rötting (1999). Ähnlich wie bei der Gestenerkennung haben in Bezug auf die Nutzerakzeptanz allerdings nur die kontaktlosen Verfahren (visuelles Interface) eine Chance auf allgemeinen Einsatz für die Mensch-Computer-Interaktion. Eine Diskussion technischer Rahmenbedingungen von Eye Trackern und deren Implikationen für die visuelle Gestaltung der Nutzerschnittstelle geben Beinhauer, Vilimek und Richter (2005).

Aus psychologischer Sicht ist die Blickrichtung in erster Linie mit der visuellen Aufmerksamkeit verknüpft (vgl. Kahneman, 1973), allerdings stehen Augenbewegungsmuster auch in engem Zusammenhang mit mentaler Aktivität und Beanspruchung (Yarbus, 1967). Eine Interpretation der Blickrichtung bei Fixationen ist nicht immer eindeutig. Es kann durchaus vorkommen, dass Menschen die Aufmerksamkeit auf ein Objekt im peripheren Sichtfeld gerichtet haben. Ein weiteres Problem für die Gestaltung von blickbasierten Schnittstellen sind unabsichtliche Fixationen, ein sporadisches Verweilen auf einem Objekt. Dies kommt besonders in Phasen hoher mentaler Aktivität vor. Dieser Faktor ist als „Midas Touch“-Problem bekannt (Jacob, Leggett, Myers & Pausch, 1993, S.74): „At first it is helpful to be able simply to look at what you want and have it occur without further action; soon, though, it becomes like the Midas Touch. Everywhere you look, another command is activated; you cannot look anywhere without issuing a command.“ Als häufigste Lösung wird eine Mindestverweildauer von einigen hundert Millisekunden über einem Steuerelement eingerichtet, bevor eine Aktion ausgelöst wird (Salvucci & Anderson, 2000). Wenn mit diesen *dwell thresholds* gearbeitet wird, ist eine ausgefeilte zeitabhängige Rückmeldung an den Nutzer notwendig, die effektiv über die bevorstehende Aktivierung informiert. Ferner muss abhängig von der Komplexität und Dichte von Elementen auf der Bedienoberfläche und von Charakteristika der Nutzergruppe eine optimale Verweildauer bestimmt werden: Bei zu kurzen Schwellenwerten steigt die Zahl unbeabsichtigter Aktivierungen, zu lange Mindestverweildauern belasten die Eingabeeffizienz beträchtlich.

2.3.3 Auditives Interface

Sicherheitskritische Bereiche der Mensch-Maschine-Interaktion sind ohne akustische Warnsignale nicht vorstellbar. Doch neben dieser klassischen Funktion finden Töne, Geräusche und sprachliche Ausgaben zunehmend den Weg in die alltägliche Bediensituation. Für die multimodale Interaktion ist das auditive Interface von besonderer Bedeutung, weil sowohl die Ausgabeseite als auch durch Sprachsteuerung die Eingabeseite sehr reichhaltige Möglichkeiten zur Anpassung an spezielle Bediensituationen bieten.

2.3.3.1 Auditive Wahrnehmung und akustische Systemausgaben

Das auditive System dient der Aufnahme von Schallereignissen, d.h. von Informationen über die Entfernung, Richtung, Bewegung und Art dieses Ereignisses (Muthig, 1990). Die Art des Ereignisses umfasst psychophysische Parameter wie Lautheit, Schärfe, Klangfarbe und Tonhöhe. Physikalisch betrachtet ist Schall eine mechanische Vibration, eine Welle, die sich in einem elastischen Medium fortbewegt. Hörschall wird normalerweise durch die Luft übertragen. Das auditive System ist auf zeitlich strukturierte Ereignisse spezialisiert und spielt eine entscheidende Rolle bei der Steuerung von Erkundungsaktivitäten und der sprachlichen Kommunikation.

Zusammenfassende Darstellungen physiologischer und psychophysischer Befunde zu auditiver Wahrnehmung finden sich bei Brewster (2003), Goldstein (1997), Hempel und Altinsoy (2005), Kohlrausch und van de Par (2005), Muthig (1990), Sorkin (1987) sowie Zwicker und Fastl (1999). Sie bilden die Grundlage für die folgende Erörterung. Die Druckschwankungen einer Schallwelle erreichen das menschliche Ohr und durchlaufen dann mehrere Stufen der Verarbeitung. Zunächst passieren sie das äußere Ohr (Ohrmuschel und Gehörgang), treffen dann auf das Trommelfell und werden über die Gehörknöchelchen im Mittelohr verstärkt. Im Innenohr befindet sich die flüssigkeitsgefüllte Cochlea. Diese Flüssigkeit wird in Schwingung versetzt, was letztlich zu einer Deformation der Sinneshärchen im Cortischen Organ führt. Durch die Auslenkung der Sinneshärchen wird ein Transformationsvorgang angestoßen, durch den ein chemischer Transmitter an eine Nervenfasern ausgeschüttet wird. Das elektrophysiologische Signal der Nervenfasern wird an das Gehirn weitergeleitet. Der Bereich hörbarer Schwingungen beträgt beim Menschen etwa 20 Hz bis 20 kHz. Die Empfindlichkeit hat ein Maximum bei etwa 4 kHz, wobei innerhalb einer Bandbreite von 700 bis 7000 Hz die Empfindlichkeit des Gehörs annähernd konstant ist. Der Schwellenwert der Amplitude eines hörbaren akustischen Signals liegt bei 0 dB SPL, bei 130-140 dB SPL wird die Schmerzgrenze erreicht. Wie laut ein Schall subjektiv empfunden wird, hängt nicht nur vom Schalldruckpegel, sondern auch von der Frequenz ab. So wird beispielsweise ein 1000 Hz Sinuston bei deutlich niedrigerem Schalldruck wahrgenommen als ein 60 Hz Ton. Der 60 Hz Ton muss mit mehr Energie übertragen werden, wenn er als gleich laut beurteilt werden soll. Dieser Zusammenhang kann in so genannten „Kurven gleicher Lautheit“, Hörschwellenkurven und der Hörflächendarstellung in einem Diagramm festgehalten werden, bei dem der Schalldruckpegel gegen die Frequenz aufgetragen ist (siehe z.B. Sorkin, 1987, S.552ff.). Ein wesentliches Merkmal des menschlichen auditiven Systems muss bei der Gestaltung akustischer Ausgaben besonders berücksichtigt werden: Spektrale Maskierung, auch Verdeckung genannt. Maskierung bedeutet, dass die Anwesenheit eines Schallsignals die Fähigkeit zur Wahrnehmung eines anderen Schallsignals einschränkt. Wenn beispielsweise ein

Ton mit einer Frequenz von 1 kHz mit einem Schalldruckpegel von 90 dB anwesend ist, müsste ein 2 kHz Ton eine Amplitude von mehr als 50 dB über der normalen Hörschwelle aufweisen, um gehört werden zu können. Hätte der 1 kHz Maskierungston nur einen Schalldruckpegel von 70 dB, müsste der 2 kHz Ton nur noch 30 dB über seiner eigentlichen Hörschwelle liegen. Diese Zusammenhänge werden in psychophysischen Frequenz-Tuningkurven beschrieben (siehe z.B. Hempel & Altinsoy, 2005, S.138).

Ein Sonderfall der auditiven Wahrnehmung ist aufgrund der Komplexität und Spezialisierung des verarbeiteten Materials die Sprachwahrnehmung. Die Frage, ob Sprachwahrnehmung auf einem eigenen spezifischen Sprachverarbeitungsmechanismus basiert oder auf einem allgemeinen, wenngleich hoch ausgebildeten, auditiven Mechanismus beruht, wird in der Sprachforschung heftig diskutiert. Einige neuropsychologische Befunde wie die überwiegende Verarbeitung der Sprache in spezifischen Kortexregionen der linken Gehirnhälfte (Geschwind, 1979) sprechen dafür. Die Vermutung wird auch von Experimenten zur kategorialen Wahrnehmung und Befunden wie dem McGurk-Effekt bei der audiovisuellen Sprachverarbeitung (McGurk & MacDonald, 1976) gestützt. Allerdings ließen sich stets auch mit nicht-sprachlichen Schallmustern ähnliche Resultate erzeugen (Pisoni, 1977; Saldaña & Rosenblum, 1993) und die Debatte wird weiterhin offen geführt. Für die Gestaltung akustischer Ausgaben ist die Unterscheidung zwischen sprachlichen und nicht-sprachlichen Ausgaben eindeutig relevant. Vor- und Nachteile dieser Gestaltungsvarianten werden auch in Handbüchern diskutiert (z.B. Brewster, 2003; Dix, Finlay, Abowd & Beale, 2004), jedoch existieren zu wenige experimentelle Vergleiche als eindeutige Basis für Hinweise zu differentiellen Einsatzfeldern.

Bei der konkreten Design-Entscheidung für jede Form eines auditiven Stimulus müssen nach Guski (1997) drei Aspekte bedacht werden: 1) die Eignung eines Tons, Geräusches oder auch einer sprachlichen Ausgabe (d.h. die Stimulus-Response Kompatibilität), 2) der Grad der Angenehmheit (Wohlklang), 3) die Identifizierbarkeit. Aures (1985) schlägt ein formales Verfahren zur Bewertung des Wohlklangs vor. Der Aspekt der Eignung und der Identifizierbarkeit sind in der Praxis häufig verbunden und geben oftmals den Ausschlag für die Verwendung von sprachlichen Schallen. Jedoch kann durch den sorgfältigen Aufbau eines semantischen Systems auch mit nicht-sprachlichen Schallen durchaus ein konsistentes, leicht erkennbares und eindeutiges Rückmeldemuster erzeugt werden. Unter dem Stichwort einer semiotischen Matrix diskutiert Hempel (Hempel, 2001, 2002) die Semantik von Hörereignissen und zeigt auf, wie zwischen einem auditiven Zeichen und dem von ihm bezeichneten Objekt indexikalische, ikonische und symbolische Beziehungen beschaffen sind. Brewster (1994, 2003) stellt Gestaltungsrichtlinien vor, wie durch eine Variation von Tonsequenzen, Rhythmen und Klangfarben (sog. *Earcons*) ein effektives Repräsentation

tionssystem geschaffen werden kann bzw. wie durch aufgezeichnete natürliche Geräusche (sog. *Auditory Icons*) Systemereignisse verständlich repräsentiert werden können. Neben einer Anpassung der Lautstärke einer akustischen Ausgabe an situative Gegebenheiten des Aufgabenumfelds ist dieser Parameter besonders geeignet, für die Priorität eines Ereignisses zu kodieren. Obwohl das auditive System absolute Lautstärkepegel nicht gut erinnern kann, ist das Gehör doch sehr empfindlich für relative Veränderungen in der Amplitude eines Schallsignals. Von dieser Möglichkeit wird erstaunlicherweise äußerst selten Gebrauch gemacht.

Sowohl sprachliche als auch nicht-sprachliche Signale werden meist über Lautsprecher oder Kopfhörer ausgegeben. Für besondere Anwendungsfälle können binaurale Ausgaben vorteilhaft sein. Durch eine differentielle Aufbereitung des Signals für beide Ohren kann über die Lokalisierbarkeit des Schalls gegebenenfalls direkte Zusatzinformationen über den Ort des Auftretens z.B. einer Gefahrenmeldung erzeugt werden (vgl. Hempel & Altinsoy, 2005). Akustische Ausgaben über den Computer können durch Wiedergabe einer gespeicherten digitalisierten Aufzeichnung erzeugt werden. Bei der Speicherung kommen meist verlustbehaftete Komprimierungsverfahren zum Einsatz. Einige Datenformate, z.B. MP3 (MPEG-1 Audio Layer 3) machen dabei Gebrauch von Effekten wie der spektralen Maskierung, um die Hörbarkeit des Verlusts durch Kompression gering zu halten. Ebenfalls möglich ist die synthetische Generierung von Klängen, Geräuschen oder Sprachsignalen. Während nicht-sprachliche Schalle seit längerem auf hohem Niveau dynamisch erzeugt werden können (vgl. Brewster, 2003) stellen die komplexen Charakteristika in der Prosodie menschlicher Sprache die Entwicklung vor besondere Schwierigkeiten. Durch voraufgezeichnete Sprachausgaben (*pre-recorded speech*) lassen sich im Allgemeinen hochwertigere und natürlicher klingende Ausgaben erzeugen. Die verwendeten Textbausteine sind jedoch nicht flexibel handhabbar und können keine zur Entwurfszeit unbekannt Texte (z.B. SMS oder E-Mails) vorlesen. Sprachsyntheseverfahren erhalten als Input geschriebenen Text und erzeugen daraus dynamisch das Sprachsignal (TTS, *text-to-speech*). Die Programme müssen dafür den Text mit lexikalischen Auswahlroutinen analysieren und die grammatischen Einheiten erkennen, um ihn in eine Sequenz von Phonemen zu konvertieren (Graphem-Phonem-Übersetzung). Zwei Verfahren können unterschieden werden, die Diphon-Synthese und die Formant-Synthese (Weinschenk & Barker, 2000). Bei der Diphon-Synthese werden gespeicherte Sprachaufnahmen von Diphonen (Nachbar-Laut Kombinationen) verkettet und die Prosodie durch Signalmodulation erzeugt. Bei der Formant-Synthese erzeugt der Computer das Sprachsignal vollständig künstlich durch die Anwendung phonologischer Regeln und physikalischer Modelle. Die Formant-Synthese ist flexibler und benötigt kaum Systemressourcen, da der Aufwand zur Speicherung entfällt. Die erzeugte Sprache klingt jedoch vergleichsweise künstlich. Von der häufig vorgetragenen Idee, voraufgezeichnete Sprache und TTS zu

verknüpfen, damit die Ausgaben zumindest über weite Teile natürlich klingen und nur dynamische Informationen mit TTS erzeugt werden müssen, ist eher abzuraten. Evaluationsstudien haben ergeben, dass die Akzeptanz der Nutzer selbst bei einer künstlich klingenden, aber konsistenten TTS Ausgabe höher ist als bei einer gemischten Ausgabe, sofern die TTS gut verständlich ist (Gong, Nass, Simard & Takhteyev, 2001). Ausgenommen hiervon sind Situationen, in denen durch eine maschinell klingende TTS die Ausgabe wichtiger Parameter aus dem normalen Sprachfluss explizit hervorgehoben werden soll. Ferner bieten einige Hersteller an, denselben menschlichen Sprecher zur Aufzeichnung der Spracheinheiten für die Synthese und zum Einsprechen der vorab aufgezeichneten Ansagen heranzuziehen. Dadurch kann der Grad der Natürlichkeit bei gleichzeitig hoher Flexibilität in beeindruckender Weise hergestellt werden.

2.3.3.2 Aspekte der Gestaltung von Spracheingabe

Spracherkennung weist viele Vorteile für den Einsatz in multimodalen Umgebungen auf, insbesondere weil bei entsprechender Rückmeldung die Hände für andere Aktionen frei bleiben, allgemein Augen- und Handfreiheit ermöglicht werden kann sowie verknüpfte und komplexe Abrufaktionen gestattet werden (Grasso, Ebert & Finin, 1998). Im industriellen Umfeld kann von Vorteil sein, dass durch Berührungsfreiheit weniger Desinfektionsvorgänge notwendig werden und längere Wege zur Konsole für die Eingabe von Parametern durch ein Headset mit Funkverbindung eingespart werden können (Schorn, 2004).

Automatische Spracherkennung (*automatic speech recognition*, ASR) lässt sich als dreistufiger Prozess beschreiben (Ainsworth, 1997). Grundlage ist zunächst die korrekte Transformation gesprochener Sprache in eine maschinenverständliche Form in der Vorverarbeitung (1) mit einer Digitalisierung des Signals und der Merkmalsextraktion (2) mit der Bereinigung des Signals von nicht relevanten Bestandteilen und Störeinflüssen. Die anschließende Mustererkennung (3) vergleicht die extrahierten Merkmale mit Referenzmustern oder nutzt Hidden-Markov-Modelle. Bei diesem Verfahren wird nicht einfach das Eingangssignal regelbasiert mit einem Referenzmuster verglichen, sondern es wird mit Hilfe eines Markov-Modells die Wahrscheinlichkeit des Übergangs von einem gegebenen Phonem auf ein darauf folgendes Phonem berücksichtigt. Auf diese Weise kann die Wahrscheinlichkeit aller möglichen Phonemfolgen berechnet werden. Die Modelle nutzen so genannte „Loop-Next-Skip-Schleifen“ (Spies, 1994). Mit der Loop-Funktion wird das aktuelle Phonem untersucht und seine Eigenschaften gefiltert. Die Next-Funktion gibt die Information an das nächste Phonem weiter, die Skip-Funktion an das übernächste Phonem. Damit bewertet die Schleife das zu betrachtende Phonem zusätzlich auf der Basis der Information über die beiden vorhergehenden Phoneme. Anschließend wird durch ein Sprachmodell

die linguistische Wahrscheinlichkeit einer „Satzhypothese“ ermittelt (Haberland, Kanthak, Overmann, Schütz & Welling, 1998). Dabei wird die Wahrscheinlichkeit für einen Satz als Produkt der bedingten Wahrscheinlichkeiten für die jeweils enthaltenen Wörter berechnet. Jedes Wort im Satz hängt nicht nur vom direkten Vorgängerwort, sondern von allen Vorgängerworten ab. Bei der Analyse werden typischerweise jedoch nur die letzten zwei Vorgänger berücksichtigt, um Rechenleistung zu sparen. Zu den Herausforderungen bei dem Prozess der Spracherkennung zählt die Segmentierung des kontinuierlichen menschlichen Sprachflusses, die Variabilität der Aussprache zwischen einzelnen Sprechern (neben Alter und Geschlecht z.B. Regiolekt, Dialekt, Soziolekt). Auch die Variabilität, die ein einzelner Sprecher bei der Aussprache aufweist, kann beträchtlich sein. Die Situation verschärft sich mit steigender Vokabulargröße (Wortschatz erkennbarer Begriffe) des Erkenners. Eine bisher kaum gelöste Schwierigkeit ist die Berücksichtigung von Information aus der Prosodie der Äußerung zu entnehmen.

Im Kontext der Steuerung von Systemfunktionen kommen meist sprecherunabhängige Systeme zum Einsatz. Im Gegensatz zu sprecherabhängigen Systemen müssen sie nicht auf einen Nutzer trainiert werden, können aber nur mit einem kleineren Vokabular arbeiten und umfassen weniger flexible Regeln zur Kombination der Wörter (Grammatik). Eine Ausnahme bilden Telefondialogsysteme (*interactive voice response*, IVR), die trotz Sprecherunabhängigkeit häufig über ein großes Vokabular verfügen. Allerdings sind meist nicht alle Teile des Vokabulars aktiv. Eine gute Übersicht über technische Rahmenbedingungen und Gestaltungsrichtlinien für Spracheingabe bei IVR-Systemen geben Weinschenk und Barker (2000).

Fehler sind bei der Spracherkennung unvermeidbar, nicht zuletzt aufgrund der probabilistischen Natur des Erkennungsprozesses. Die Forderung nach einem „designing for error“ (Lewis & Norman, 1986) gilt hier wie in kaum einem anderen Bereich der Mensch-Maschine-Interaktion. Schließlich hängt die Akzeptanz von Sprachtechnologie neben der reinen Fehlerkennungsrate immens davon ab, wie leicht Erkennungsfehler zu korrigieren sind, welches Ausmaß die Konsequenzen einer Fehlerkennung annehmen können und wie sich das auf die Möglichkeiten der Nutzer auswirkt, eine Aufgabe zu Ende zu führen (Boyce, 2000; Kamm, 1994; Rhyne & Wolf, 1993). Besonders kritisch sind Fehlererkennungen, wenn sie mehrfach hintereinander vorkommen. Die Nutzer neigen dann zu einer unnatürlich deutlichen, überbetonten und unnötig langsamen Sprechweise (Hyperartikulation), die von den akustischen Modellen des Spracherkenners noch schlechter verarbeitet werden kann und zu weiteren Fehlern führt (Oviatt, Levow, Moreton & MacEachern, 1998). Bei den Erkennungsfehlern werden vier Kategorien unterschieden: Zurückweisungsfehler (*rejection error*, zulässige Äußerung wird nicht verstanden; Nutzer wird darauf aufmerksam gemacht), Löschungs- oder Auslassungsfehler (*deletion error*, System ignoriert eine Äußerung

vollständig ohne Benachrichtigung), Einfügingsfehler (*insertion error*, System wertet ein Hintergrundgeräusch als Spracheingabe) und Verwechslungsfehler (*substitution error*, System interpretiert eine zulässige Äußerung falsch). Diese Fehlerquellen können durch verbesserte technologische Möglichkeiten verringert werden. So können beispielsweise Filtermechanismen und Algorithmen zur Rauschunterdrückung die Anzahl von Einfügingsfehlern vermindern oder die Verwendung von mehreren Mikrofonen (*arrays*) zu weniger Auslassungsfehlern beitragen. Eine weitere Möglichkeit ist die Einführung einer Schlüsselwort-Aktivierung oder die Verwendung so genannter Push-to-Talk Tasten, die man zur Aktivierung der Spracherkennung drücken muss. Ist die Spracherkennung nicht die ganze Zeit über eingeschaltet, können Einfügingsfehler wirksam unterdrückt werden. Versuchen die Nutzer, Wörter zu verwenden, die das System nicht kennt, spricht man von einem *out-of-vocabulary* (OOV) Fehler. Steht eine visuelle Anzeige zur Verfügung, hat es sich bewährt, gültige relevante Sprachbefehle dort explizit aufzuführen. Dieses Prinzip nennt sich *Say What You See* und trägt zu einer massiven Reduktion von OOV-Fehlern bei (vgl. Weinschenk & Barker, 2000; Yankelovich, 1996).

Oft wird mit Spracheingabe die Erwartung einer sehr intuitiven, weil für den Menschen natürlichen Interaktionsform verbunden. Dabei wird die Natürlichkeit der Interaktion häufig mit natürlichsprachlicher Eingabe im Sinne einer Mensch-Mensch-Kommunikation gleichgesetzt. Aus psychologischer Sicht kann diese Gleichsetzung nicht uneingeschränkt unterstützt werden und technisch ist sie auf absehbare Zeit nicht zu verwirklichen. So genannte *fully conversational systems* benötigen nach Heisterkamp (2003) die volle menschliche Verständniskapazität, eine ausgereifte Expertise über alle verfügbaren Funktionen und müssen in der Lage sein, aus den Äußerungen der Nutzer deren wahre Intention zu erschließen. Heisterkamp betont, dass zwar bedeutsame Fortschritte bei der Erfüllung dieser Voraussetzungen gelungen sind, einige Aspekte jedoch nicht durch Technologie allein lösbar sind und so immer mehr der Mensch in den Fokus rückt. Eine nicht zu unterschätzende Anzahl von Misserfolgen bei der sprachbasierten Mensch-Maschine-Interaktion ist danach auf Probleme durch die Nutzer zurückzuführen, die nicht direkt zu dem System sprechen, auf Rückfragen des Systems unscharfe Antworten geben, ihre Anfragen nicht in verständlichen Sätzen artikulieren, viele Selbstkorrekturen durchführen oder schlichtweg nicht das ausdrücken können, was sie eigentlich anstreben. Zu dieser Einschätzung passen auch experimentelle Befunde von Oviatt (1994), wonach die Aussagen von Nutzern mit zunehmender Länge unstrukturierter werden. Gerade bei natürlicher Sprache lässt sich ein steigender Anteil von Selbstkorrekturen, Häitationen und Füll-Lauten bei längeren Sätzen nachweisen. Darüber hinaus ist die Güte der Sprachproduktion von der kognitiven Belastung abhängig. Baber und Noyes (1996) konnten nachweisen, dass in Dual Task Situationen wie der Fahrzeugführung mit gleichzeitiger Systembedienung viele

Unterbrechungen und Wiederholungen zu verzeichnen sind. Die Konstruktion vollständiger Sätze trifft hier auch nicht unbedingt auf die Bedürfnisse der Nutzer. Für den Anwendungsfall einer Steuerung von Infotainment- und Navigationsfunktionen konnte eine deutliche Präferenz der Nutzer von kurzen, kommandosprachlichen Eingabestrategien über konversationale Eingabe nachgewiesen werden (Graham, Aldridge, Carter & Lansdown, 1999). Darüber hinaus ließ sich auch in anderen Fällen nutzerseitig initiiertes Sprachinteraktion nachweisen, dass zur Steuerung von Maschinen oder anderen technischen Einrichtungen ein eher kurzer und kommandobasierter Sprachstil den Erwartungen entspricht (Akyol et al., 2001; Baber & Stammers, 1989; Richards & Underwood, 1984; Zoltan-Ford, 1991). Unter diesen Umständen kann die Möglichkeit zu unstrukturierter natürlichsprachlicher Eingabe mehr Usability-Probleme schaffen, als sie löst. Darüber hinaus wirft Heisterkamp (2003) eine weitere Frage auf: Was ist überhaupt eine „natürliche“ Eingabeform? Als Beispiel führt er an, dass es für uns durchaus natürlich ist, in einem dunklen Raum einen Lichtschalter zu suchen, um Licht zu machen. Er führt aus, dass diese Aktion jedoch keineswegs natürlich im Wortsinn ist, sondern lediglich eine gelernte Konvention darstellt. Viele Probleme mit Sprachbedienungen existieren deshalb, weil noch keine Konventionen existieren. Statt einer einseitigen Ausrichtung auf perfekt diskursfähige Systeme sollte deshalb zunächst versucht werden, gut lernbare sprachliche Standards zu etablieren. Diese sollten auch deutlich robuster gegenüber kognitiven Parallelanforderungen sein. Für multimodale Systeme bietet sich das *Say What You See* Eingabeprinzip als Grundlage einer solchen Konvention an. Auf diesem Weg wird durch die Integration in die visuell-manuelle Interaktion das Sprachvokabular gelernt und kann vom Nutzer leicht memoriert werden. Vilimek und Hempel (2005) führen aus, wie *Say What You See* zu einer Erfüllung von Usability-Anforderungen beiträgt.

2.3.4 Haptisches Interface

Die manuelle Steuerung von Systemfunktionen ist die Urform der Eingabe bei der Mensch-Maschine-Interaktion. Eine klare Besonderheit der manuellen Eingabe ist, dass die Hand die Rolle des Effektors und Rezeptors in sich vereinigt. Aktionsfeedback über den haptischen Kanal wird in vielen Fällen durch mechanische Eigenschaften der Stellteile vermittelt. Bereits sehr früh wurde vorgeschlagen, die Oberflächeneigenschaften wie die Textur und Form eines Bedienelements aktiv in die Gestaltung zu integrieren (formkodierte Oberflächengestaltung, Jenkins, 1947). Ob man einen Schalter wirklich gedrückt hat, kann man dem Kraft-Weg-Verlauf und dem Überwinden des Druckpunkts entnehmen. Ist jedoch das Stellteil nicht mehr direkt mechanisch mit dem System verbunden, wie es zum Beispiel durch Servomotoren der

Fall ist, benötigt die Gestaltung der Stellkräfte eine besondere Zuwendung (Bubb, 2001).

2.3.4.1 Haptische Wahrnehmung und haptische Systemausgaben

In der deutschsprachigen und auch der internationalen Literatur wird mit haptischer Wahrnehmung eine unübersichtliche Menge von Begriffen und Begriffskombinationen in Zusammenhang gebracht (Grunwald, 2001). Dies ist wenig überraschend, wenn man sich vor Augen hält, dass die Frage nach der Einheitlichkeit des Tastsinns und die Diskussion um einen möglichen Primat des Tastsinns bei der Wahrnehmung ein altes philosophisches und erkenntnistheoretisches Problem ist (John, 2001; Zimmer, 2001a, 2002).

Die Arbeiten von Beyer und Weiss (2001), Geiser (1990), Kolb und Wishaw (1996), Schmidh und Thews (1997), Sorkin (1987) sowie Zimmer (1998, 2001a) ermöglichen eine detaillierte Einsicht über den Tastsinn und fassen empirische Befunde zusammen. Auf dieser Basis soll versucht werden, die haptische Wahrnehmung über eine Analyse der Sinnesleistungen des somatosensorischen Systems zu umreißen und in den Zusammenhang zur Mensch-Maschine-Interaktion zu rücken. Wie in Abschnitt 2.1.2.1 bereits kurz angerissen, nimmt die haptische Wahrnehmung im Vergleich zu anderen Sinnesmodalitäten eine Sonderrolle ein, da es kein einzelnes definiertes Wahrnehmungsorgan gibt und sie mit verschiedenen Wahrnehmungsqualitäten betraut ist (Burdea, 1996). Der Terminologie von Geiser (1990) folgend umfasst Haptik die Komponenten kinästhetische und taktile Wahrnehmung.

Kinästhetische Sensibilität wird gleichbedeutend mit den Begriffen Tiefensensibilität oder Propriozeption verwendet (Schmidt & Thews, 1997). Darunter ist die Vermittlung der Positions- und Kraftempfindung durch mechanische Einflüsse auf Muskeln und Gelenke zu verstehen, die eine Wahrnehmung der Stellung und Bewegung des Körpers bzw. von Körperteilen ermöglichen. Gelenkrezeptoren (Ruffini-Körperchen, Pacini-Körperchen, Golgi-Rezeptoren) reagieren auf Bewegung und Lage der Körperextremitäten, während Mechanosensoren in den Gelenkkapseln sowie Muskel- und Sehnenrezeptoren (Rezeptoren der Muskelspindeln und Golgi-Sehnenorgane) Streckungen und Kontraktionen melden. Auch die Mechanosensoren der Haut tragen zur propriozeptiven Wahrnehmung bei, z.B. durch die Rückmeldung der Deformation im Bereich eines Gelenks. Drei kinästhetische Qualitäten werden unterschieden: Stellungs-, Bewegungs- und Kraftsinn. Der Stellungssinn informiert über die Stellung der Gelenke zueinander. Gelenkbewegungen werden durch den Bewegungssinn wahrgenommen, dessen Wahrnehmungsschwelle von der Winkelgeschwindigkeit und von der Lokalisierung abhängt. Die Schwelle für die Bewegung eines Fingergelenks beträgt beispielsweise 1.0 bis 1.3 Grad bei einer Geschwindigkeit von 12.5 Grad/s. Körpernahe Gelenke können kleinere Änderungen des Winkels registrieren, wie z.B. das Schul-

tergelenk mit einer Schwelle von 0.2 bis 0.4 Grad bei 0.3 Grad/s. Der Kraftsinn vermittelt schließlich das Ausmaß an aufgewendeter Muskelkraft für eine Bewegung oder zum Einhalten einer statischen Gelenkstellung. Die Wahrnehmungsleistung bezüglich der Unterschiedsschwellen ist beim freien Halten von Objekten deutlich höher, als wenn nur deren Druck passiv auf die Hautoberfläche ausgeübt wird. Bei der komplexen Integration der verschiedenen sensorischen Eindrücke im Zentralnervensystem wird zusätzlich auch das Vestibularorgan involviert, sodass bei der Propriozeption tatsächlich alle verfügbaren neuronalen Informationen genutzt werden.

Die taktile Wahrnehmung bezieht sich auf den Haut- oder Tastsinn und umfasst unterschiedliche und in der Haut vermischt auftretende Sensoren für Druck, Berührung, Schmerz, Wärme, Kälte und Vibration, die miteinander interagieren. Für Steuerungs- und Regeltätigkeiten in multimodalen Schnittstellen haben die Ausgabe und das Empfinden von Temperatur und Schmerz eine untergeordnete Bedeutung, weshalb hier nicht näher darauf eingegangen wird. Von besonderem Interesse für die Mensch-Maschine-Interaktion sind die Wahrnehmung von Vibration, Druck und als komplexe Empfindung das Ertasten von Oberflächeneigenschaften mit der Hand. Die sensorischen Fähigkeiten sind bemerkenswert: An der empfindlichsten Stelle liegt die Zweipunktschwelle, bei der zwei Reizpunkte gerade noch einzeln identifiziert werden können, bei nur 1 mm. Eine punktförmige Erhebung von einer glatten Oberfläche kann ab einer Höhe von 2 μm erspürt werden. Die Innenhand des Menschen wird von etwa 17000 Nervenfasern versorgt, die die Informationen der empfindlichen Mechanorezeptoren in der haarlosen Haut weiterleiten. Diese Rezeptoren können nach ihrem charakteristischen Reiz-Antwort-Verhalten und der Eigenschaften ihrer rezeptiven Felder in zwei Gruppen eingeteilt werden (vgl. Beyer & Weiss, 2001): Einem andauernden Reiz ausgesetzt reagieren langsam adaptierende Sensoren (*slowly adapting*, SA I und SA II) mit einer konstanten Antwort, während schnell adaptierende Rezeptoren (*fast adapting*, FA I und FA II) nur zu Beginn und häufig auch am Ende eines Reizes antworten, nicht aber während dessen Fortdauern. In der unbehaarten Haut sind SA I Sensoren Merkel-Zellen, SA II Sensoren Ruffini-Körperchen (Nervenendigungen), FA I Sensoren Meissner-Körperchen und FA II Sensoren Vater-Pacini-Körperchen. Teilweise liest man als alternative Bezeichnungsform für die FA I Rezeptoren auch RA (*rapidly adapting*), während die FA II Rezeptoren mit PC (für Pacini-Körperchen) abgekürzt werden. Die Meissner-Körperchen fehlen in der behaarten Haut. Stattdessen befinden sich dort Haarfollikelsensoren. Jedem Rezeptortyp kann eine separate Bedeutung für die Wahrnehmung zugeschrieben werden (vgl. Schmidt & Thews, 1997). Während SA I Rezeptoren mit einem kleinen rezeptiven Feld die Intensität einer zur Hautoberfläche senkrechten Deformation durch Druck melden, geben SA II Rezeptoren mit großem rezeptivem Feld Auskunft über die Intensität von Hautdeformationen durch laterale Dehnung. FA I Rezeptoren reagieren sensitiv auf Druckanstieg wie beim An-

tippen der Haut und können als Geschwindigkeitssensoren bezeichnet werden, da ihre Entladungsfrequenz mit der Geschwindigkeit der Reizbewegung steigt. FA II Rezeptoren antworten nur, wenn bei einem Hautreiz Beschleunigungen auftreten, weshalb sie auch als Beschleunigungssensoren bezeichnet werden. Für die Ausgabe taktiler Information ist besonders relevant, dass FA II Rezeptoren als Vibrationssensoren agieren (Frequenzbereich 50-1000 Hz) und Vibrationen auch dann vermittelt werden, wenn bereits Druck auf der Haut lastet. FA I nehmen niederfrequente Schwingungen im Bereich von 10-60 Hz als Zittern wahr. Die Wahrnehmung von Vibration ist neben der Frequenz abhängig von der Stärke der Vibration, dem Ausmaß der horizontalen oder vertikalen Verschiebung (Hub) und auch von der Temperatur. Darstellungen von Iso-Sensitivitätskurven und der Veränderung der Vibrationsempfindlichkeit in Abhängigkeit von der Temperatur finden sich bei Zimmer (1998). Ein Frequenzbereich von etwa 20-1000 Hz ist wahrnehmbar. Für die absolute Schwelle kann ein eindeutiges Minimum bei 250 Hz angegeben werden, wobei ab 30 dB zwischen 200 und 750 Hz die Iso-Sensitivitätskurven annähernd flach verlaufen. Bereits ab 5 dB kann bei entsprechendem Hub eine stabile überschwellige Wahrnehmung von taktilen Reizen im Frequenzbereich zwischen 25 und 750 Hz ermöglicht werden.

Strukturierte vibrotaktile Reizmuster (sog. *Tactons*) eignen sich zur nicht-visuellen Ausgabe von Information im Rahmen der Mensch-Maschine-Interaktion (Brewster & Brown, 2004; Brown, Brewster & Purchase, 2005). Brewster und Brown (2004) schlagen als relevante Parameter zur Gestaltung Frequenz, Amplitude, Dauer, Rhythmus und den Ausgabeort am Körper vor. Die Wellenform des Signals eignet sich nur bedingt, taktil können höchstens Rechtecks- von Sinuswellen unterschieden werden. Zimmer (1998) gibt zu Bedenken, dass nur bis ca. 50 Hz die taktilen Unterschiedsschwellen in der gleichen Größenordnung liegen und in höheren Frequenzbereichen rapide zunehmen. Gerade ab dem Bereich, bei dem die Schwellensensitivität am höchsten ist (250 Hz), ist praktisch keine sichere Diskrimination mehr möglich. Frequenzkodierung für vibratorische Information wird für Frequenzen bis zu 50 Hz empfohlen. Als besonders problematisch sieht Zimmer (1998) Einbußen durch Maskierungseffekte an. Bei der Diskrimination ist das kritische Zeitfenster zwischen ca. -30 ms (Beginn der Wirkung der Vorwärtsmaskierung) und +30 ms (Ende der Wirkung der Rückwärtsmaskierung). Für die Wiedererkennung bzw. Identifikation liegt das zeitliche Fenster zwischen -50 ms bzw. +150 ms. Aufgrund dieser Trägheit lassen sich auch Scheinbewegungen induzieren, die für Trackingaufgaben entscheidende Informationen übertragen können.

In den Untersuchungen von Gibson (1962) hat sich gezeigt, dass das aktive Berühren von Gegenständen (*active touch*) in vieler Hinsicht vom passiven Berühren (*passive touch*), d.h. dem berührt werden, zu unterscheiden ist. Das aktive Berühren wird meist mit dem Gegenstand assoziiert, während passives Berühren typischerweise eher

mit einer Empfindung auf der Haut verbunden wird. Darüber hinaus ist die Erkennungsleistung beim *active touch* dem *passive touch* massiv überlegen. Entscheidend ist nach Gibson, dass das aktive Berühren mit bewusster Intention erfolgt. Zusätzlich ist bei passivem Berühren nur taktile Wahrnehmung involviert, die Rückmeldung aus den Gelenken und Bändern kommt nicht zum Tragen. Ein weiterer entscheidender Faktor ist multimodaler Natur: Wird ein Gegenstand aktiv abgetastet, empfindet man nicht nur dessen Oberflächentextur, sondern erhält zusätzliche Informationen über die Geräusche, die beim Streichen über diese Oberfläche entstehen (Taylor, Lederman & Gibson, 1973). In einer Erweiterung der Unterscheidung von aktivem und passivem Berühren klassifizierten Klatzky und Lederman (1987, 1992, 2003) aufgrund einer großen Reihe von Detailuntersuchungen die Mechanismen, mit denen taktile und kinästhetische Empfindungen zu der Erkundung von Objekten führen. In einer aktuellen Darstellung von Klatzky und Lederman (2003, S.156f.) lassen sich folgende Explorationsprozeduren unterscheiden:

1. Laterale Bewegung: Exploration von Texturen. Gekennzeichnet durch Verschiebungen zwischen Haut und einem Objekt (Scherkräfte).
2. Druck: Exploration der Härte. Charakterisiert durch Ausüben von Kraft auf ein stabilisiertes Objekt.
3. Statischer Kontakt: Exploration der Temperatur. Charakterisiert durch Berührung eines Objekts mit größtmöglicher Hautoberfläche, jedoch ohne Bewegung.
4. Ungestütztes Halten: Exploration des Gewichts. Charakterisiert durch Anheben eines Objekts von einer Stützfläche, oft verbunden mit einer Armbewegung.
5. Umschließung: Exploration der Größe und grobe Formabschätzung. Charakterisiert durch Umfassen der Hüllfläche eines Objekts oder von Teilen eines Objekts ohne Krafteinwirkung.
6. Kontur nachfahren: Exploration der exakten Form und Identifikation von Einzelteilen. Charakterisiert durch Abtasten der Kanten eines Objekts mit den Fingerspitzen.

Die aus früheren Veröffentlichungen (Klatzky & Lederman, 1987, 1992) bekannte Prozedur der partiellen Bewegung wird in den neueren Abhandlungen nicht mehr angeführt. Die Prozeduren 1, 3 und 5 entsprechen dem *passive touch* nach Gibson (1962), während 2, 4 und 6 dem *active touch* zugerechnet werden (vgl. Zimmer, 1998). Wird Druck (2) hingegen nicht aktiv auf das Objekt ausgeübt, sondern kommt er durch die Kraft oder das Drehmoment eines auf der Haut ruhenden Objekts zu Stande, wird Druck ebenfalls als *passive touch* betrachtet.

Verschiedene Technologien stehen zur Verfügung, um taktile oder kinästhetische Ausgabe zu ermöglichen. Taktile Displays stimulieren die Hautoberfläche teils mit sehr geringen Kräften. Der häufigste Applikationsort ist die Handfläche oder die Fingerspitze. Allerdings existieren auch prototypische taktile Displays, die zum Beispiel mit vibrierenden Brustgurten oder Vibrationen in Sitzflächen arbeiten (Altinsoy, 2006; van Erp, van Veen, Jansen & Dobbins, 2005). Vibrotaktiler Feedback ist relativ einfach zu erzeugen, z.B. durch Elektromotoren mit Unwucht. Insbesondere für blinde Menschen wurden Ausgabegeräte erzeugt, die mit Hilfe einer dichten Anordnung einzeln beweglicher, sehr kleiner Metallstifte (*micropin arrays*) arbeiten. Eine solche Anordnung kann man sich als Rechtecksmatrix von stumpfen Nadeln vorstellen. Die Abstände der Nadeln sind möglichst klein und jede Nadel kann typischerweise nur binär vollständig ein- oder ausgefahren sein. Daneben können mit Hilfe derartiger Nadelanordnungen Texturen und zweidimensionale Geometrien erzeugt werden (Burdea, 1996). Ein sehr bekanntes taktiles Display, das mit Hilfe von Pins vibrotaktile Muster erzeugt, ist das Optacon (Optical-to-Tactile Connector von Telesensory Systems, Inc.). Grafische Muster wie Buchstaben können damit via Videoerfassung in taktile Vibrationsmuster auf einem *micropin array* konvertiert werden, die sowohl von Blinden wie auch Sehenden Versuchspersonen gut erkannt werden können (Craig, 1977). Eher selten verwendete Ausprägungen taktiler Ausgabe sind elektrotaktile Stimulation, bei der durch geringe Stromstärken ein Kribbelgefühl auf der Haut entsteht, und pneumatische Displays, die fokussierte Luftströme direkt auf die Hautoberfläche ausgeben oder die Befüllung von Luftkissen mit Hautkontakt regulieren (Sato, Igarashi & Kimura, 1991; Zhu, 1988). Taktiles Feedback wird, gemessen an den Fähigkeiten der menschlichen Sensorik, insgesamt eher rudimentär eingesetzt. Meist wird eine einfache Vibration durch einen oder mehrere Aktuatoren erzeugt, realistische taktile Hinweisreize wie eine feinauflösende veränderbare Oberflächenstruktur können noch nicht realisiert werden.

Kinästhetische Ausgaben erfolgen durch Geräte mit Force Feedback (Kraftrückkopplung). Um die Kraftrückkopplung zu spüren, interagiert der Nutzer mit beweglichen Teilen eines fest verankerten Eingabegeräts, indem er diese z.B. mit der Hand ergreift und verschiebt oder drückt. Abhängig von der Art der zu simulierenden Kraft kann sich Force Feedback in einem dynamischen Widerstand bei der Bewegung oder in der aktiven Ausgabe makroskopischer Bewegungskräfte äußern. Technisch werden Einflüsse auf die Position häufig mit Elektromagneten realisiert, Widerstand kann ebenfalls über Elektromagneten oder Aktivbremsen eingeleitet werden. Aktive Kräfte werden mit Elektromotoren erzeugt. Für eine realitätsnahe Ausgabe müssen Kräfte sehr gleichmäßig mit einer hohen Signal-Aktualisierungsrate (500-1000 Hz) betrieben werden (Oakley, 2003). Bei der Mensch-Computer-Interaktion und auch für die Steuerung von Maschinen kommen meist Joystick- oder stiftartige Eingabevarianten zur

Steuerung eines Cursors mit einer unterschiedlichen Anzahl von Freiheitsgraden (*degrees of freedom*, DOF) zum Einsatz. Der Freiheitsgrad beschreibt in diesem Zusammenhang die Anzahl und Art der möglichen Bewegungen eines Gelenks. Rotatorische Freiheitsgrade geben die Anzahl möglicher Drehbewegungen an, translatorische Freiheitsgrade beziehen sich auf lineare Bewegungen. Der Bereich verfügbarer Force Feedback Geräte deckt das gesamte Spektrum ab, von nicht-richtungsgebundener Kraftausgabe bis hin zu sehr komplexen Geräten mit maximal 6 DOF (drei translatorisch, drei rotatorisch), die einen Großteil der menschlichen Wahrnehmungskapazität ansprechen und Bewegungen auf allen drei räumlichen Achsen durchführen können. Als Beispiel kann das PHANTOM Device 3.0/6DOF von Sensable Technologies, Inc. genannt werden. Die Eingabe erfolgt in der typischen Ausführungsvariante über einen einzelnen Finger, der in einer Art Fingerhut positioniert wird und der seinerseits an die Gelenkkonstruktion angebunden ist. Mit einer maximalen Ausgabekraft von 22 N (kontinuierlich: 3 N) auf jeder räumlichen Achse, 515 mNm (Gierachse) und 170 mNm (Rollachse) bei einer Aktualisierungsrate des Signals von 1000 Hz und einer sehr geringen Eigenträgheit von weniger als 220 g der Fingerablage kann das Gerät sehr wirklichkeitsgetreue Kräfte simulieren. Bei Zielbewegungen in der Mensch-Computer-Interaktion werden meist kontinuierliche Widerstandskräfte von 0.25 N bis 0.8 N benötigt (Dennerlein & Yang, 2001; Oakley, Adams, Brewster & Gray, 2002). Eine weiterführende Übersicht über Eingabegeräte bietet Iwata (2003). Grundsätzlich verschieden von der Interaktion mit fest verankerten Eingabegeräten ist die Ausgabe von kinästhetischem Feedback über Exoskelett-Konstruktionen. Systeme, die in Handschuhe integriert sind, bieten 1 DOF translatorische Kraftausgabe und können so das Gefühl des Greifens eines Objekts durch Simulation der Gegenkräfte bei Druck vermitteln (z.B. CyberGrasp von Immersion Corp., 12 N pro Finger). Iwata (2003) weist jedoch darauf hin, dass die Akzeptanz der Nutzer zum Tragen von Exoskeletten gering ist und bei großen Kräften die Verletzungsgefahr nicht unterschätzt werden darf. Als viel versprechende Einsatzgebiete zeichnen sich Trainingsmaßnahmen in virtuellen Umgebungen und die Unterstützung von Rehabilitationsmaßnahmen ab.

2.3.4.2 Aspekte der Gestaltung von manueller Eingabe

Die Eignung eines manuellen Eingabeelements hängt maßgeblich von der inhaltlichen Funktion des Stellteils im Zusammenspiel mit der gestellten Aufgabe ab. Folgende Anforderungen haben sich als sinnvoll für eine Unterscheidung erwiesen (Bullinger, Kern & Muntzinger, 1987): Eignung für die Eingabe von zwei Stellungen, für mehr als zwei Stellungen, für kontinuierliche Eingabe, für präzise Feineinstellungen, für schnelle Einstellungen, für die Aufwendung großer Kräfte, für das Spüren von mechanischer taktiler Rückmeldung, für die Sichtbarkeit der vorgenommenen Einstellung und für die Vermeidung unabsichtlicher Betätigung. Als grundlegende Stellteile identifizieren

Bullinger et al. (1987) für lineare Bewegungen Hand- und Zylindergriffe bei senkrechten oder vertikalen Schaltvorgängen, sowie Druckschalter, Drucktaster, Schieberegler, Schiebeschalter und Sensortaster. Stellteile, die Schwenk- und Gelenkbewegungen ausführen, sind Kipphebel, Gelenkhebel, Joysticks (Steuerknüppel), Kippschalter, Wippenschalter, und Wählscheiben. Drehbewegungen können mit Drehknöpfen, Handrädern, Kurbeln, Rändelrädern, Nockenschaltern und Dreh-Drückstellern ausgeführt werden. Diese Elemente können zu komplexen Eingabegeräten kombiniert werden, wie z.B. zu einer Computer-Tastatur zur Texteingabe. Bullinger et al. (1987) geben eine ausführliche Übersicht über die jeweilige Eignung dieser Stellteile für die oben genannten Anforderungen. Eine besondere Bedeutung für die Mensch-Computer-Interaktion besitzen Zeigegeräte, so genannte *pointing devices*, mit denen der Cursor auf einer grafischen Oberfläche bewegt werden kann (Hinckley, 2003). Die bekanntesten Zeigegeräte sind Computer-Mäuse, Touchpads und Trackballs. Darüber hinaus kommen Joysticks, Tablets, die die absolute oder relative Position eines darauf positionierten Stifts registrieren können, und Touchscreens zum Einsatz. Zusätzlich existieren Geräte mit mehreren translatorischen und rotatorischen Freiheitsgraden wie das bereits erwähnte PHANTOM, die sich besonders für dreidimensionale Aufgaben eignen. Durch geeignete technische Vorrichtungen ist eine Implementierung dynamischer haptischer Ausgaben für fast alle dieser Stellteile möglich.

Entscheidend für eine gute Gestaltung ist die Anpassung des Eingabeelements an die Aufgabe und an die menschliche Motorik. Dabei muss insbesondere der Greifraum berücksichtigt werden. Darunter versteht man den Bereich, in dem Greifaktionen bei stabilisierter Rumpfposition durch Arm- und/oder Handbewegungen im Stehen oder Sitzen durchgeführt werden können (Geiser, 1990). Dieser Raum wird begrenzt durch die Körperhaltung und Körperstellung, den Bewegungsumfang der Gelenke, die Richtung von Bewegungen und Kräften, die verwendeten Arbeitsmittel sowie die notwendige Stabilisierung des Gleichgewichts (Bullinger, 1994). Er reduziert sich bei großer Muskelanspannung. Alle für einen Eingabevorgang wesentlichen Orte sollten im Greifraum zu erreichen sein. Die relevanten Maße des menschlichen Körpers sind der ISO-Norm 7250 zu entnehmen (ISO 7250, 1997).

Viele Eingabeverfahren erfordern vom Nutzer manuelle Zielbewegungen. Die notwendige Zeit, um mit einer ballistischen Bewegung von der Startposition zur Zielposition zu gelangen, lässt sich mit dem Gesetz von Fitts modellieren (Fitts, 1954). Fitts erkannte, dass die mittlere Bewegungszeit durch eine logarithmische Funktion des Verhältnisses von Abstand zu Größe des Ziels vorausgesagt werden kann. Je kleiner oder weiter entfernt ein Ziel ist, desto länger dauert die Bewegung. Die mittlere Zeit, um ein Ziel zu treffen, ergibt sich durch $MT = a + b \log_2(2A/W)$. Dabei ist MT (*movement time*) die Zeit, das Ziel zu erreichen, W (*width*) ist die Größe des Ziels und A (*amplitude*) ist die Entfernung. Die beiden Konstanten a und b sind empirisch ange-

passte Regressionsparameter. Der Logarithmus der Formel wird Schwierigkeitsindex *ID* genannt und in Bits gemessen. MacKenzie (1989, 1992) hat eine Modifikation des Schwierigkeitsindex eingebracht, die die Anpassung und die Korrelation mit realen Daten deutlich verbessert: $ID = \log_2(A/W + 1)$. Diese Variante kann als de facto Standard für die Mensch-Computer-Interaktion betrachtet werden (Cockburn & Brewster, 2005). Bei der Anwendung von Fitts' Gesetz müssen einige Vorbedingungen berücksichtigt werden, da sich nicht alle Zielaufgaben damit modellieren lassen. Es ist nur anwendbar, wenn der Arm uneingeschränkt bewegt werden kann. In der oben angegebenen Formulierung können darüber hinaus nur Aussagen über eindimensionale Bewegungen getroffen werden, allerdings wurde von Accot und Zhai (1997) eine Erweiterung für zweidimensionale Bewegungen vorgeschlagen. Ferner können damit keine hochtrainierten Bewegungsabläufe beschrieben werden. Die Bedeutung des Modells erwächst dadurch, dass es sich sowohl auf direkte manuelle Eingaben als auch auf Zielbewegungen mit Pointing Devices am Computer anwenden lässt. Wenngleich Fitts' Gesetz manuelle Zielbewegungen nicht perfekt vorhersagt, ist es dennoch eine sehr gute und in vielen Anwendungsfeldern bewährte Approximation (Meyer, Abrams, Kornblum, Wright & Smith, 1988).

2.3.5 Vorteile und Einschränkungen multimodaler MMI

In vielen Untersuchungen konnte bisher das Potential multimodaler Schnittstellen auf unterschiedlichsten Ebenen nachgewiesen werden (z.B. Bos, Huls & Claassen, 1994; Brewster, 1998c; Brewster & Crease, 1999; Bunt, 1998; Grasso & Finin, 1997; Liu, 2001; Maybury, 1998; Neuss, 2000; Oviatt, 2002; Raisamo, 1999; Stevens, Edwards & Harling, 1997; Vitense, Jacko & Emery, 2003; Wickens, Sandry & Vidulich, 1983). Dabei zeichnen sich einige typische Vorteile ab. Oviatt et al. (2000) listen in einem Überblicksartikel die folgenden Aspekte auf:

- Flexibilität: Eine einzelne Modalität kann nicht den gleichen breiten Bereich von Aufgaben und Umweltbedingungen abdecken wie mehrere alternative Eingabemodalitäten. Multimodale Schnittstellen erlauben die Verwendung situativ angemessener Eingabeformen.
- Barrierefreiheit (Accessibility): Multimodale Schnittstellen können eine breitere Nutzerbasis ermöglichen, insbesondere durch das Verfügbarmachen von alternativen Ein- und Ausgabeformen für Menschen mit körperlichen Beeinträchtigungen.
- Erlernbarkeit / Natürlichkeit: Multimodale Schnittstellen gestatten natürlichere Interaktionsformen und sind somit leichter erlernbar.
- Steigerung der Produktivität: In vielen Studien konnten kürzere Aufgabenbearbeitungszeiten und eine Reduktion von Eingabefehlern festgestellt werden.

Teils wird erst durch multimodale Schnittstellen eine parallele Bearbeitung mehrerer Aufgaben ermöglicht, teils wird die Bearbeitung durch Multimodalität effizienter.

- Verbesserte Fehlerprozeduren: Multimodale Schnittstellen können durch die Interpretation mehrerer Modalitäten eine Reduktion von systemseitigen Erkennungsfehlern erreichen (gegenseitige Disambiguierung).
- Robustheit: Bei multimodaler Interaktion lässt sich die Tendenz feststellen, dass sich die Struktur der sprachlichen Eingaben vereinfacht. Dadurch reduziert sich die Komplexität für natürlichsprachliche Systeme und die Robustheit steigt.
- Benutzerfreundlichkeit: Wenn mehrere Modalitäten verfügbar sind, lässt sich bei über 90% der Nutzer zumindest eine rudimentäre Nutzung der erweiterten Möglichkeiten feststellen. Das deutet auf eine hohe Akzeptanz der Nutzer hin.

Für den Anwendungsfall Fahrzeugführung stellt Bengler (2001) heraus, dass insbesondere die Auswahlmöglichkeit einer geeigneten Modalität das Zusammenspiel zwischen Fahraufgabe und Zusatzaufgaben erfolgreich steigern kann. Dafür sei es jedoch notwendig, dass der Nutzer für dieselbe Aufgabenstellung verschiedene Lösungsstrategien verfolgen kann. Hierfür müssen verschiedene Eingabemodalitäten äquivalente Alternativen (vgl. Martin, 1997) sein und der Nutzer muss über ein adäquates mentales Modell bezüglich der Passung der jeweiligen Modalität zur aktuellen Situation verfügen. Nach Hedicke (2000) sowie Schattenberg und Debus (2001) können multimodale Schnittstellen dazu dienen, den direkten Kontakt zu maschinellen Prozessen zu erhöhen bzw. wiederherzustellen, indem aktionsbezogene sensorische Information präsentiert wird. Durch redundante Informationen und prozesskompatible Information kann die Systemsicherheit erhöht und die Beanspruchung des Operators verringert werden, indem durch Multimodalität einseitig belastende Tätigkeiten zu vielfältig wechselnden Vorgängen transformiert werden.

Diesen Vorteilen stehen einige Einschränkungen gegenüber, die bei der Entscheidung für oder gegen ein multimodales System sorgsam erwogen werden sollten. Niedermaier (2003) verweist auf den deutlich höheren technischen Aufwand, der sich letztlich auf den Verkaufspreis der Systeme auswirkt. Probleme entstehen mitunter durch die Integration zusätzlich benötigter Hardware-Komponenten (Mikrofon-Arrays, Kameras, etc.). Eine weitere Herausforderung stellt die gleichzeitige Ausgabe von multimodalen Signalen dar. Wenn Kombinationen von auditiven, taktilen, haptischen und visuellen Stimuli als simultan wahrgenommen werden sollen, werden hohe Anforderungen an den minimalen Zeitversatz gestellt (vgl. Hempel & Altinsoy, 2005). Zusätzlich werden erhöhte Ansprüche bezüglich der Infrastruktur erhoben, indem mehr Rechenleistung, Speicher und höhere Datentransferraten erforderlich werden.

Schließlich bedingt die Steigerung der Komplexität der Software, insbesondere der erkenntnisbasierten Technologien, einen deutlich höheren Entwicklungsaufwand. Darunter fallen nicht nur zeitliche und finanzielle Ressourcen, sondern auch die zusätzlichen Tests, die für multimodale Systeme in der Entwicklungsphase notwendig sind.

Weitere Schwierigkeiten entstehen, wenn falsche Erwartungen über das Verhalten der Nutzer bei der Konzeption eines solchen Systems zu Grunde gelegt werden. Aus diesen Annahmen erwachsende Verzerrungen führen zu einer deutlich verminderten Bedienbarkeit des Systems. Oviatt (1999) fasst einige der häufigsten falschen Auffassungen zusammen. Erschwerend kommt hinzu, dass starke interindividuelle Unterschiede bei multimodalem Verhalten existieren, z.B. ob Modalitäten bevorzugt synergistisch, komplementär oder sequentiell genutzt werden. Diese Unterschiede werden durch die Art der Aufgabe, die mit dem System bearbeitet werden soll, unter Umständen verstärkt. Gegenwärtig existieren zu wenige experimentelle Studien, die das Verhalten der Nutzer in multimodalen Umgebungen auf überdauernde Interaktionsmechanismen untersuchen (Oviatt, Lunsford & Coulston, 2005) und so zu einer Modellbildung beitragen könnten. Auch Schomaker et al. (1995) weisen darauf hin, dass groß angelegte Experimentalsreihen erforderlich sind, um auch nur basale Prinzipien entdecken zu können. Im Zusammenhang mit Systemtests verweisen Schomaker und Kollegen auf einen inhärenten Nachteil multimodaler Systeme: Mit jeder zusätzlichen freien Variable (d.h. Modalität) steigt die Anzahl der experimentellen Bedingungen extrem an (kombinatorische Explosion), wodurch eine systematische experimentelle Exploration erschwert wird.

2.3.6 Experimentelle Befunde zur Verarbeitung multimodaler Signale

Die Untersuchung multimodaler Stimuli in Laborexperimenten kann auf eine lange Geschichte zurückblicken. Vor fast 100 Jahren führte Todd (1912) ein Experiment durch, bei dem er Versuchspersonen visuelle Reize präsentierte, die gleichzeitig mit oder ohne einem zusätzlichen auditiven Stimulus ausgegeben wurden. Er konnte nachweisen, dass bei der Präsentation von bimodalen Stimuli die Reaktionszeiten kürzer waren als bei unimodalen Reizen. Dieser Effekt ist als *redundant signals effect* (RSE) bekannt und wurde in vielen Studien nicht nur mit visuellen und auditiven Stimuli, sondern auch mit visuell-taktilen Reizen repliziert (zusammenfassend Diederich & Colonius, 2004). Für den Förderungseffekt durch bimodale Redundanz gibt es mehrere Erklärungen. Raab (1962) nimmt an, dass der RSE auf einem statistischen Effekt beruht (*statistical facilitation*). Er geht von einer separaten Aktivierung der einzelnen Kanäle aus, bei der die Verarbeitungsgeschwindigkeit in jedem Kanal normalverteilt

variiert. Der jeweils schnellere Kanal führt zu einer sichtbaren Antwort. Aufgrund der Schwankungen in der Verarbeitungsgeschwindigkeit sagt dieses Modell im Mittel eine kürzere Reaktionszeit auf multimodale Reize vorher. Eine andere Erklärung nimmt an, dass die visuelle und die auditive Reizverarbeitung interagieren und sieht den RSE als Konsequenz einer intersensorischen Erleichterung (*intersensory facilitation*). Dabei wird als Alternative zu einer separaten Aktivierung der Kanäle von einer Koaktivierung ausgegangen, bei der die unimodalen Signale kombiniert werden und gemeinsam zu einer Antwort führen (Miller, 1982). Somit kommt eine höhere Übertragungsgeschwindigkeit als bei jedem einzelnen Kanal zu Stande. In Modelltests konnten Verstöße gegen einen rein statistischen Effekt festgestellt werden (Miller, 1982; Mordkoff & Miller, 1993), während Koaktivierung die experimentellen Daten besser erklären kann (Diederich, 1995; Schwarz, 1989). Für die Anwendung dieser Erkenntnisse als Grundlage der Gestaltung multimodaler Ausgabe ist jedoch weniger der Mechanismus als vielmehr das absolute Ausmaß des Effekts relevant. Denn wenngleich der RSE äußerst stabil ist, findet man typischerweise Effekte zwischen 20 und 90 ms Verkürzung der Reaktionszeit (Gielen, Schmidt & van den Heuvel, 1983). Damit ist der RSE allein nicht ausreichend, multimodale Ausgabe in der Praxis zu rechtfertigen – dafür werden Effekte von mehreren hundert Millisekunden benötigt. Dennoch ist er als Grundlagenbefund relevant, da er Rückschlüsse auf die Art der multimodalen Informationsverarbeitung zulässt. Praxisrelevant ist in dieser Hinsicht auch der Befund, dass der Redundanzeffekt im Alter trotz nachlassender Leistungsfähigkeiten bei der geteilten Aufmerksamkeit voll erhalten bleibt (Bucur, Allen, Sanders, Ruthruff & Murphy, 2005).

Neben einer schnelleren Wahrnehmung von Signalen erhofft man sich auch eine höhere Entdeckungswahrscheinlichkeit bei redundanten multimodalen Ausgaben. Hier scheint die Befundlage nicht so einheitlich zu sein. Grundlagenbefunde deuten in Richtung einer besseren Entdeckungsrate (Buckner & McGrath, 1963; Howarth & Treisman, 1958). Auch in einigen angewandten Studien ließ sich die Entdeckungswahrscheinlichkeit für ein Ereignis bei simultaner visueller und auditiver Ausgabe gegenüber unimodaler Darbietung erhöhen (Liu, 2001; Wickens & Hollands, 2000). Mehrfach gelang dies jedoch nicht (Kobus et al., 1986; Seagull, Wickens & Loeb, 2001; Tzelgov, Srebro, Henik & Kushelevsky, 1987). Allerdings ist die Entdeckungswahrscheinlichkeit bei multimodaler Ausgabe nie signifikant geringer als bei unimodaler Ausgabe. Kobus et al. (1986) schließen deshalb, dass multimodale Ausgabe immer vorzuziehen ist, wenn es die Aufgabenstellung zulässt. So könnte zumindest die Chance einer verbesserten Wahrnehmung von Systemausgaben gewährleistet werden.

Wenn nicht redundante multimodale Ausgabe vorliegt, sondern verschiedene unimodale Informationsquellen unter geteilter Aufmerksamkeit beachtet werden müssen, kann es sogar zu Nachteilen kommen. Zwar konnten mehrere Studien nachweisen,

dass auditive Nachrichten bei starker visueller Belastung als Warnsignale eher wahrgenommen werden als visuelle Hinweisreize (Edworthy, 1994; Simpson & Williams, 1980). Spence und Driver (1997) berichten jedoch in einem Überblicksartikel, dass mit dem Wechsel der Aufmerksamkeit von einer Modalität zu einer anderen Kosten verbunden sind. Dies gilt insbesondere, wenn auf ein Ereignis in einer Modalität reagiert werden muss, während Information aus einer anderen Modalität erwartet wird. Die Aufmerksamkeit kann sehr gut zwischen zwei Modalitäten geteilt werden, wenn die Wahrscheinlichkeit für Ereignisse gleich hoch ist. Wenn jedoch eine bestimmte Modalität öfter relevant wird, fokussiert sich die Aufmerksamkeit darauf und seltene Ereignisse (z.B. Warnungen) in einer anderen Modalität führen zu langsameren Antworten. Die Befunde von Spence und Driver (1997) schränken den Gültigkeitsbereich des Modells der multiplen Ressourcen von Wickens (1984) ein, da anscheinend keine vollständige Unabhängigkeit der Modalitäten als Verarbeitungsressourcen bei geteilter Aufmerksamkeit gegeben ist.

Schwenkt man den Fokus von direkt stimulusabhängigen Maßen wie Reaktionszeit und Entdeckungswahrscheinlichkeit hin zu den Effekten multimodaler Ausgaben bei komplexen Aufgaben in angewandten Untersuchungen, findet man weitgehend positive Resultate. So konnten geringere Fehlerraten, kürzere Aufgabenbearbeitungszeiten, besseres Time-Sharing und geringere Workload durch zusätzliches Feedback zur visuellen Ausgabe demonstriert werden (z.B. Akamatsu, 1992; Arsenault & Ware, 2000; Brewster, 1998a, 1998c; Cockburn & Brewster, 2005; Göbel, Luczak, Springer, Hedicke & Rötting, 1995; Liu, 2001; Oakley, McGee, Brewster & Gray, 2000; Vitense et al., 2003). Einige Befunde deuten darauf hin, dass ältere Menschen besonders von multimodalem Feedback profitieren und ihre Leistungsfähigkeit so steigern können (Jacko et al., 2004; Jacko et al., 2005; Liu, 2001). Hier stellt sich die Frage, weshalb die Befunde in Untersuchungen mit Anwendungshintergrund im Gegensatz zu den Laborstudien relativ eindeutig Vorteile für multimodale Ausgaben berichten. Zunächst stammen viele Effizienznachweise aus der Forschung an Mehraufgaben-Situationen wie der Fahrzeugführung, in der die visuelle Modalität durch die Primäraufgabe fast vollständig absorbiert wird. Der positive Beitrag nicht-visueller Zusatzinformation ist hier klar ersichtlich. Einen weiteren Hinweis kann die mehrfach nachgewiesene geringere Workload geben (Brewster, 1998b; Liu, 2001; Oakley et al., 2002; Oakley et al., 2000; Vitense et al., 2003). Je weniger die experimentellen Aufgaben reinen Reiz-Reaktions-Tests entsprechen und stattdessen komplexe Anwendungssituationen widerspiegeln, desto mehr zusätzliche kognitive Aktivitäten (z.B. Planen, Entscheiden) werden erforderlich. Reduziert multimodale Ausgabe die Workload in einer solchen Situation, verkleinert sie gleichzeitig den kognitiven Bottleneck. Gedächtnispsychologische Grundlagenbefunde deuten ebenfalls auf eine unterstützende Wirkung kognitiver Vorgänge durch multimodale Ausgaben hin. Sowohl episodische

Gedächtnisleistungen (Stefanucci & Proffitt, 2005) als auch die Gedächtnisspanne bei kurzzeitiger Speicherung (Karpicke & Pisoni, 2000) erwiesen sich gegenüber unimodaler Informationsausgabe als überlegen. Damit ist eine klare Implikation für die Gestaltung von Mensch-Maschine-Schnittstellen verbunden, da im praktischen Einsatz Anwender oftmals in kurzer Zeit viele Informationen aufnehmen, verarbeiten und mit bereits gespeicherter Information abgleichen müssen.

Um von multimodalen Schnittstellen profitieren zu können, müssen die Nutzer in der Lage sein, eine kohärente Wahrnehmung der Ausgaben zu generieren. Sobald angestrebt wird, dass Signale von einem technischen System über verschiedene Sinneskanäle als simultan registriert werden, entstehen hohe Anforderungen an die Synchronisierung der Ausgaben. Allgemeine Empfehlungen über maximale Verzögerungen bewegen sich zwischen 25 ms (Altinsoy, 2006) und 50 ms (Card, Moran & Newell, 1986). Der geringe Spielraum ist hauptsächlich durch die Empfindlichkeit der Wahrnehmung bei auditiv-haptischen Reizen bedingt. Verschiedene Methoden werden zur Ermittlung der Kennzahlen genutzt. Häufig wird auf die Beurteilung der zeitlichen Reihenfolge oder Synchron-Asynchron-Urteile zurückgegriffen. Typischerweise zeigt sich bei diesen Studien eine Abhängigkeit von der Reihenfolge der Reizpräsentation. Nach Kohlrausch und van de Par (2005) existiert eine deutliche Asymmetrie in Abhängigkeit davon, ob ein auditiver Stimulus *vor* oder *nach* einem visuellem Stimulus präsentiert wird. Geht das Audiosignal dem visuellen Signal voraus, wird ein Versatz von bis zu 40 ms von geübten Hörern noch als gleichzeitig beurteilt. Ein nachfolgend präsentierter auditiver Reiz hingegen wird bis zu 100 ms nach dem visuellen Stimulus mit diesem zu einem simultanen Ereignis integriert (Hempel & Altinsoy, 2005; Kohlrausch & van de Par, 2005). Bei haptischen Stimuli ist zu beachten, wo die Empfindung erzeugt wird. Von besonderem Interesse ist der sensitive Bereich der Hand. Zwischen taktilen Reizen auf der Fingerspitze und auditiven Ausgaben besteht eine zeitliche Wahrnehmungsschwelle von 50 ms, wenn der taktile Reiz dem auditiven vorausgeht und 25 ms, wenn er danach präsentiert wird (Altinsoy, 2006; Hempel & Altinsoy, 2005). Überraschenderweise lässt sich zwischen visuellen und haptischen Reizen keine Reihenfolgeasymmetrie feststellen. Unabhängig davon, welcher Stimulus vorausgeht, sollten ca. 45 ms (Vogels, 2004) nicht überschritten werden. Zusammenfassend sind zeitliche Verzögerungen stets kritischer, wenn haptische Ausgaben involviert sind.

2.3.7 Experimentelle Befunde zu multimodaler Eingabe

Multimodale Eingabe umfasst bei einem Großteil der Systeme sprachliche und manuelle bzw. manuell-gestische Eingabe. Zu dieser Kombination gibt es entsprechend die meisten experimentellen Befunde, während die Wissensbasis für andere Eingabefor-

men teilweise sehr dürftig ist. Das Verhalten der Nutzer bei der Integration von Aktionsmodalitäten, dem Wechsel zwischen diesen Modalitäten sowie die Identifikation von überdauernden Verhaltensmustern und Veränderungen des Eingabeverhaltens in multimodalen Umgebungen stellen Kernfragen für eine erfolgreiche Konstruktion der multimodalen Systemarchitektur dar.

Wenn Sprache und andere aktive Aktionsmodalitäten synergistisch verwendet werden können, es also auf technischer Seite zu einer Fusion der Eingabeströme kommt, verändern sich linguistische Eigenschaften der Sprache des Nutzers. Oviatt (1997) hat gezeigt, dass weniger Wörter, kürzere Sätze und weniger komplexe Anweisungen gegeben werden, wenn Sprache mit einer manuellen oder gestenbasierten Eingabe kombiniert wird. Dieser Effekt tritt auf, wenn eine Information eingegeben werden muss, für die sich die sprachliche Modalität nicht optimal eignet. So muss beispielsweise für die Angabe einer räumlichen Information per Sprache eine komplexe Beschreibung des Zielorts gegeben werden. Dabei treten sehr lange, natürlichsprachliche Konstruktionen mit vielen Selbstkorrekturen auf (Oviatt, 1997; Oviatt & Kuhn, 1998). Nachdem eine sprachliche Eingabe räumlicher Parameter vergleichsweise umständlich ist, ist dieses Resultat an sich nicht überraschend. Interessant hingegen ist das Ausmaß der Veränderung, da sich bei multimodalen Kombinationen gleichzeitig ein Wandel von natürlichsprachlichen Konstruktionen hin zu kommandosprachlicher Eingabe vollzieht.

Bei der Fusion manueller und sprachlicher Eingabe muss das System eine Annahme über die Intention des Nutzers treffen: Handelt es sich um *ein* Kommando, das integriert werden muss, oder sind es *zwei* voneinander unabhängige Eingaben? Typischerweise wird dazu mit Zeitstempeln und zeitlichen Schwellenwerten gearbeitet (Allen, 1983; Nigay & Coutaz, 1993; Oviatt, 2003). Eine gute Wahl des Schwellenwerts ist essentiell, da sowohl intra- als auch interindividuell unimodale und multimodale Kommandos durchmischt verwendet werden. Der Anteil multimodal integrierter Kommandos schwankt zwischen ca. 20% in verbal-zeitlichen Anwendungen und kann bis 70% oder mehr steigen, wenn die Applikation räumliche Komponenten umfasst (Oviatt et al., 2005). Wenn die Konstruktionen sequentiell sind, das Sprachkommando also vor oder nach der manuellen Eingabe gegeben wird, können Verzögerungen von bis zu vier Sekunden auftreten (Oviatt, DeAngeli & Kuhn, 1997). Bei Kommandos, die Verknüpfungen zwischen den Modalitäten zulassen, muss das System folglich annähernd vier Sekunden warten, ob ein weiterer Input folgt. Daraus resultiert eine Verzögerung des Interaktionsprozesses, die die Effizienz des Systems reduziert. Zusätzlich entstehen häufig Erkennungsfehler, da die Nutzer bei zu langer Wartezeit ihre Eingaben wiederholen. Die Wiederholungen kann das System nicht sinnvoll interpretieren. Oviatt et al. (2005) schlagen vor, sich individuelle Unterschiede im Integrationsmuster zu Nutze zu machen, um den Fusionsprozess zu optimieren. In einer

Längsschnittstudie haben sie untersucht, ob es überdauernde Verhaltensweisen im Umgang mit multiplen Modalitäten zwischen den Versuchsteilnehmern gibt. Grundsätzlich sind zwei Versionen möglich: simultane Integration, bei der bereits während der Eingabe in einer Modalität Aktionen mit Hilfe einer zweiten Modalität ausgeführt werden, und sequentielle Integration, bei der beide Eingaben ohne Überlappung aufeinander folgen. Bereits aus einer Vorstudie war bekannt, dass etwa 70% der Nutzer eine Tendenz zu simultaner Integration aufweisen (Oviatt et al., 2003). Allerdings war unklar, inwiefern sich das Muster über die Zeit erhält. Für die Studie wurden durch ein Prescreening Verfahren zwei annähernd gleichgroße Gruppen gewählt. In einer Analyse der multimodalen Konstruktionen über einen Zeitraum von sechs Wochen haben Oviatt und Kollegen festgestellt, dass jeder Teilnehmer ein dominantes Integrationsmuster aufweist, das sich sehr schnell herausbildet und in 95% der Fälle über die Zeit erhalten blieb. Dieser Befund bietet sich als Grundlage an, um in Zukunft adaptive Zeitschwellen einzurichten. Dazu muss das System in einem Klassifikationsverfahren zunächst den Integrationstyp ermitteln. Darauf basierend kann bei simultaner Integration die Zeitschwelle massiv verringert werden.

Die Studie von Oviatt et al. (2005) erbrachte darüber hinaus weitere interessante Resultate, die aufgrund der Anlage als längsschnittliche Untersuchung besonders wertvoll sind. Dabei handelt es sich um Unterschiede im linguistischen Stil und um Verhaltensunterschiede in Abhängigkeit vom Integrationstyp. Während beide Gruppen vergleichbare Aufgabenbearbeitungszeiten und eine ähnliche Sprachgeschwindigkeit aufwiesen, wurden bei sequentieller Integration signifikant weniger Eingabefehler und Probleme während der Formulierung der Spracheingabe registriert. Dies äußerte sich in einer geringeren Anzahl von Pausen, Interjektionen, Häsitationen und Korrekturen. Auffällig ist auch der Unterschied im Sprachstil zwischen den Gruppen. Etwa nur die Hälfte der Teilnehmer mit dominant simultaner Integration zeigte eine sprachökonomische Reduktion auf Kommandosprache bei multimodalen Eingaben. Dagegen verwendeten alle Nutzer mit sequentiellem Integrationsmuster kommandosprachliche Konstruktionen. Des Weiteren führte sequentielle Integration zu einem wesentlich kleineren und weniger variablen Wortschatz der Nutzer. Aus kognitionsergonomischer Sicht sollten deshalb Anstrengungen unternommen werden, durch die Gestaltung der Schnittstelle sequentielle Integration zu fördern, da besonders in Mehraufgabensituationen und unter hoher Workload die Qualität der Sprachproduktion deutlich nachlässt (Baber & Noyes, 1996). Die Menge daraus resultierender Erkennungsfehler könnte durch die vereinfachte Sprache bei sequentieller Integration reduziert werden.

Hochgradig interessant sind die Beobachtungen von Oviatt und Kollegen (2005) zur Reihenfolge der Aktionsmodalitäten bei der Integration. Die Analyse aller multimodalen Konstruktionen ergab, dass bei simultaner Integration zu 72% Sprache der manuellen Eingabe voranging. In einer früheren Untersuchung der Arbeitsgruppe von Oviatt

wurden bei simultaner Integration hingegen nur 14% der multimodalen Ausdrücke von Sprache eingeleitet (Oviatt et al., 1997). Dieser Unterschied lässt sich möglicherweise dadurch begründen, dass in der Studie von 1997 manuell komplexere Informationen (z.B. Symbole oder geschriebener Text) übermittelt werden mussten und diese zusätzliche Anforderung die sprachliche Formulierung verzögerte. Bei sequentieller Integration zeigen die beiden Studien ein übereinstimmendes Ergebnismuster. In der früheren Studie (Oviatt et al., 1997) ging die manuelle Eingabe zu 99% und in der späteren Studie (Oviatt et al., 2005) zu 82% der Spracheingabe voraus.

Wie bereits eingangs kurz angesprochen sind verschiedene Modalitäten unterschiedlich gut geeignet, bestimmte Inhalte zu transportieren. Wickens und Hollands (2000) weisen darauf hin, dass sich die menschliche Sprache hauptsächlich hinsichtlich einer Kompatibilität für linguistische, symbolische und kategoriale Information (z.B. „erfasse das rote Dreieck“) entwickelt hat. Analog-räumliche Information (z.B. „etwas mehr nach oben“) kann wesentlich besser manuell bestimmt werden. In mehreren Experimenten konnten diese Prinzipien unter Dual Task Bedingungen bestätigt werden (für eine Übersicht siehe Wickens & Hollands, 2000, Kapitel 7 und 10). Ähnliches berichtet Oviatt (2003) in einer Übersicht zu multimodalen Single Task Studien. Danach bevorzugten die Teilnehmer Sprache zur Beschreibung von Objekten und Ereignissen, von Mengen und Teilmengen von Objekten, von Objekten außerhalb des aktuell sichtbaren Bereichs, von zusammengesetzten Informationsstücken, von vergangenen sowie zukünftigen Zustandsformen und zur Spezifikation von Aktionen. Manuelle Eingabe überwog für die Eingabe von Zahlen, Symbolen, grafischen Zeichen und die Angabe von Form und Ort räumlicher Inhalte auf einer grafischen Bedienoberfläche.

Einige der Vorteile, die multimodalen Systemen zugeschrieben werden, haben sich bisher experimentell nicht wie angenommen gezeigt. So berichtet Bengler (2001), dass das Spektrum zur Verfügung stehender Aktionsmodalitäten kaum ausgenutzt wird, wie in mehreren Untersuchungen zu Systembedienung während der Fahrzeugführung gezeigt wurde. Ebenso ist das Argument, die Nutzer würden im Fall einer mehrfachen Fehlererkennung flexibel dieselbe Eingabe in einer anderen Modalität vornehmen, in der Praxis noch nicht belegt. Im Gegenteil, die Nutzer waren nur in seltenen Fällen bereit, die einmal gewählte Modalität für einen bestimmten Eingabeschritt zu wechseln (Althoff & McGlaun, 2001, zitiert nach Bengler, 2001). Anders als bei multimodaler Ausgabe ist bei multimodaler Eingabe auch nicht immer von einer Reduktion der Workload auszugehen. Wenn beispielsweise für Gestensteuerung neue Interaktionsformen gelernt werden müssen, kann selbst über längere Zeit hinweg eine erhöhte geistige und körperliche Beanspruchung nachgewiesen werden (Keates & Robinson, 1999, Testzeitraum drei Monate).

2.3.8 Gestaltungsrichtlinien für multimodale Systeme

Gestaltungsrichtlinien, die sich auf multimodale Interaktion bzw. Systeme als Ganzes beziehen, besitzen allein schon aufgrund der vielfältigen Kombinationsmöglichkeiten verfügbarer Ein- und Ausgabetechnologien zwangsläufig einen sehr abstrakten, allgemeinen Charakter. Um die relevanten technischen Kenngrößen und Interaktionsparadigmen berücksichtigen zu können, müssen dementsprechend zusätzlich die verfügbaren unimodalen Gestaltungsrichtlinien herangezogen werden. Multimodale Richtlinien können nur einen Eindruck davon vermitteln, wie sich die Kombination einzelner Modalitäten idealerweise auswirken sollte und welche Wechselwirkungen zu erwarten sind. Doch auch wenn ein konkretes multimodales System streng nach derartigen Richtlinien implementiert wird, entbindet das nicht von empirischen Wirksamkeitsnachweisen und Tests der Gebrauchstauglichkeit.

Drei Quellen für Gestaltungsrichtlinien verdienen eine besondere Erwähnung. Das European Telecommunications Standards Institute (ETSI) stellt zwei Guideline-Dokumente bereit, die „Guidelines on the multimodality of icons, symbols and pictograms“ von 2002 (ETSI EG 202 048) und „Multimodal interaction, communication and navigation guidelines“ von 2003 (ETSI EG 202 191). Ebenfalls 2003 ist eine Sammlung von Gestaltungsrichtlinien auf einem interdisziplinären Experten-Workshop während der Conference on Human Factors in Computing Systems (CHI'03) erarbeitet worden. Diese Richtlinien wurden später von den beteiligten Forschern veröffentlicht (Reeves et al., 2004). Das neueste Dokument stammt von der Multimodal Interaction Working Group des World Wide Web Consortium (W3C) und wurde von Jim Larson (2006) herausgegeben. Diese Dokumente werden im Folgenden kurz besprochen. Auf eine Auflistung der Richtlinien soll hier jedoch verzichtet werden, da die Prinzipien ohne die teils recht umfangreichen Einführungen und Beispiele nicht verständlich sind.

Beide ETSI-Richtlinien wurden vor dem Hintergrund barrierefreier Gestaltung und einem Design-for-All Ansatz erstellt. Die Richtlinienammlung ETSI EG 202 048 beschäftigt sich mit multimodaler Ausgabe. Sie gibt zunächst Empfehlungen für unimodale Symbole und diskutiert neben visueller, auditiver und haptischer Ausgabe auch die Möglichkeiten einer olfaktorischen und gustatorischen Kodierung. Für die Empfehlungen zu multimodalen Symbolen werden zunächst die Übermittlung von Information über einzelne Sinnesmodalitäten und modalitätsübergreifende Wahrnehmungsphänomene besprochen, bevor schließlich explizite Anweisungen in Form von 19 Regeln formuliert werden. Der Empfehlungskatalog ETSI EG 202 191 geht auf multimodale Ein- und Ausgabe ein und stellt den Ausführungen zwei Zielaussagen voraus: Erstens wie Multimodalität genutzt werden kann, um die Gefahr des sozialen Ausschlusses körperbehinderter Menschen zu reduzieren, und zweitens wie eine Stei-

gerung der Accessibility durch Multimodalität zu erreichen ist. Ausgehend von zwei so genannten Metaprinzipien werden insgesamt 14 Gestaltungsrichtlinien bezüglich Gestaltung und Implementierung multimodaler Systeme abgeleitet. Das erste Metaprinzip besagt, dass die multimodale Ausgabe eine Aufnahme der Information auf eine Weise durch die Nutzer gestatten muss, die deren Fähigkeiten und Präferenzen entspricht. Das zweite Metaprinzip verlangt die Gestaltung der Interaktion so zu realisieren, dass die Nutzer Eingaben in einer Form tätigen können, die auf ihre Wünsche und speziellen Bedürfnisse ausgerichtet ist. Diese Ausgangspunkte erscheinen auf den ersten Blick etwas problematisch, weil sie sich nur auf eine äquivalente Auslegung der Modalitäten beziehen. Das lässt sich in der Praxis kaum durchhalten, weil viele Inhalte nicht ohne weiteres über jede Modalität übermittelbar sind. Vor dem Hintergrund der angestrebten Barrierefreiheit macht jedoch die Forderung Sinn, dies zumindest soweit als möglich anzustreben.

Die Sammlung von Gestaltungsrichtlinien von Reeves und Kollegen (2004) ist unterteilt in die Bereiche Anforderungsspezifikation, Gestaltung von multimodaler Ein- und Ausgabe, Adaptivität, Konsistenz, Feedback und Fehlervermeidung bzw. Fehlerbehebung. Wie auch die ETSI-Richtlinien betonen sie, dass gleichermaßen Ein- und Ausgabe für einen erfolgreichen Interaktionsprozess optimiert sein müssen. In ihren Empfehlungen kulminiert das in der Forderung „maximize human cognitive and physical abilities“ (Reeves et al., 2004, S.58).

Bereits im Titel des W3C-Dokuments (Larson, 2006) wird der Anspruch des Dokuments geklärt: „Common sense suggestions for developing multimodal user interfaces.“ Wenngleich durch den Verweis auf den Common Sense angedeutet wird, dass es sich um völlig offensichtliche Notwendigkeiten handelt, wird gleichzeitig darauf hingewiesen, dass eben diese Notwendigkeiten – so einsichtig sie sein mögen – in der Praxis mehr als nur häufig missachtet werden. Die Empfehlungen werden Designern als Leitfaden für die Entwicklung vorgeschlagen, damit diese stets prüfen können, ob ihre Entscheidungen noch auf den richtigen Fragen beruhen. Die Gestaltungsempfehlungen bauen auf vier Hauptprinzipien auf. Es wird gefordert die Beschränkungen im praktischen Einsatz hinreichend zu berücksichtigen (1), einen eindeutigen, präzisen und konsistenten Kommunikationsstil gegenüber den Nutzern zu etablieren (2), den Nutzern eine schnelle und effiziente Fehlerbehebung zu ermöglichen (3) und eine für die Nutzer angenehme Interaktionssituation durch einfache Erlernbarkeit und Bedienbarkeit zu schaffen (4). In einer weiteren Untergliederung dieser Leitprinzipien listet Larson (2006) insgesamt 36 Gestaltungsempfehlungen auf, die in den Anwendungsbeispielen teils sehr direkte Lösungsansätze beinhalten.

3 Anwendungsfall Fahrzeugführung

Den Mittelpunkt der Forschungstätigkeit an multimodalen Schnittstellen bildet die Frage, wie neue Technologien zu einer Verbesserung der Interaktionsqualität zwischen Mensch und Maschine eingesetzt werden können. Die ISO-Norm 9241-11 gibt einen klaren Anhaltspunkt, was bei der Gestaltung derartiger Lösungen unter dem Stichwort der Gebrauchstauglichkeit zu beachten ist: Das „Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“ muss berücksichtigt werden (ISO 9241-11, 1998, S.4). Der Beitrag von Multimodalität kann folglich nur in engem Bezug auf eine Nutzergruppe und deren Aufgabenumfeld (Nutzungskontext und -ziele) sinnvoll eingeordnet werden.

Der Anwendungsfall Fahrzeugführung bietet aufgrund seiner Charakteristika ein sehr interessantes Forschungsumfeld. Neben den zu erfüllenden Voraussetzungen durch die Straßenverkehrsordnung können nur wenige einschränkende Faktoren für die sehr heterogene Nutzergruppe gefunden werden. Um eine Bedienbarkeit der Komponenten im Fahrzeuginnenraum für alle Nutzer sicherstellen zu können, muss berücksichtigt werden, dass viele Selten-Nutzer, Wechsel-Nutzer und nicht-professionelle Nutzer unter den Fahrern sein werden. Umso entscheidender ist eine detaillierte und eindeutige Beschreibung der Fahraufgaben, des Aufgabenumfelds und der resultierenden Handlungssituation als Grundlage jeder gestalterischen Maßnahme im Fahrzeug. Eine Analyse theoretischer Modelle der Fahrzeugführung sowie der besonderen Mehraufgaben-Situation im Fahrzeug soll diese Grundlage bereiten.

3.1 Theoretische Modelle der Fahrzeugführung

Für eine Aufgliederung der komplexen Tätigkeit des Fahrverhaltens hat sich eine Unterscheidung von drei Ebenen als zweckmäßig erwiesen. Sie werden nachfolgend zusammenfassend und autorenübergreifend dargestellt. Anschließend werden Parallelen zu einem hierarchischen Verhaltensmodell nach Rasmussen aufgezeigt.

3.1.1 Die Drei-Ebenen-Modelle der Fahrzeugführung

Die verschiedenen Varianten der mittlerweile als klassisch zu bezeichnenden Drei-Ebenen-Modelle differieren in den konkreten Bezeichnungen der Ebenen, weisen aber nahezu identische Inhalte in der Beschreibung einer hierarchischen Verhaltenskontrolle auf (siehe z.B. Bernotat, 1970; Johannsen & Rouse, 1979; Kraiss, 1985; Michon, 1985). Die Ebenen (1) Planen bzw. Navigieren, (2) Manövrieren/Führen und (3) Stabilisieren unterscheiden sich in Bezug auf zeitliche Aspekte und die notwendige bewusste, kognitive Zuwendung des Fahrers zu ihrer Durchführung (Reichart & Haller, 1995). Diese Ebenen sind nicht unabhängig: Die übergeordnete Ebene liefert die Vorgaben für die tiefer liegende Ebene. Auf der obersten Ebene der Planung/Navigation (1), die auch als strategische Ebene der Fahrzeugführung bezeichnet wird, werden unter Beteiligung wissensbasierter Prozesse die Fahrtroute und allgemeine Ausführungskriterien wie Risikoverhalten oder Fahrökonomie festgelegt. Planungs- und Entscheidungsprozesse können auf dieser Ebene während einer Fahrsituation flexibel angepasst werden. Auf den untergeordneten Ebenen werden diese Vorgaben umgesetzt. Auf der Ebene der Führung (2) werden Fahrmanöver ausgeführt, die sich neben dem Umsetzen der Route auf situative Faktoren beziehen. Darunter fallen Teilaufgaben wie Abbiegen, Überholen, die Wahl der Geschwindigkeit oder die Reaktion auf Verkehrszeichen. Die Stabilisierungsebene (3) ist dadurch charakterisiert, dass fertigkeitbasierte und bei erfahrenen Fahrern weitgehend automatisierte Handlungen wie Gas geben, Bremsen, Lenken sowie Kuppeln und Schalten vorgenommen werden. Durch diese Aktivitäten werden die Spurhaltung (Querführung) und der Abstand zu vorausfahrenden Verkehrsteilnehmern (Längsführung) durch eine fortwährende Kontrolle der Geschwindigkeit in einem stabilen, mit den übergeordneten Zielen korrespondierenden Bereich gehalten. Die erfolgreiche Ausführung der Führungs- und insbesondere der Stabilisierungsebene steht in besonderer Abhängigkeit von der visuellen Aufmerksamkeit für die erforderliche Hand-Auge-Koordination zum Beispiel bei der Abstimmung der Querführung auf den Straßenverlauf. Visuell-manuelle Zusatzaufgaben weisen entsprechend ein besonderes Interferenzpotential auf dieser Ebene auf (Verwey, 1996).

Die drei Ebenen der Fahrzeugführung stellen unterschiedliche Anforderungen an den Fahrer. Während die Aktivitäten auf der Planungsebene ein hohes Maß bewusster kognitiver Anstrengung zum Festlegen einer bestimmten Route verlangt und auch auf der Ebene der Führung die Wahrnehmung und Interpretation von Verkehrssituationen unter kognitiver Kontrolle erfolgt, überwiegen auf der Stabilisierungsebene Prozesse mit geringer mentaler Kapazitätsforderung. Andererseits sind die Tätigkeiten auf der Stabilisierungsebene von zeitlich sehr hoher Frequenz, die Aktivitäten auf den höheren Ebenen müssen hingegen selten durchgeführt werden. Die relativen Unterschiede auf

den Dimensionen der kognitiven und zeitlichen Anforderungen sind in Abbildung 6 dargestellt. Generell gilt, dass durch Übung und eine höhere Ausbildung von Routine Prozesse von einer höheren auf eine niedrigere Ebene übergehen können.

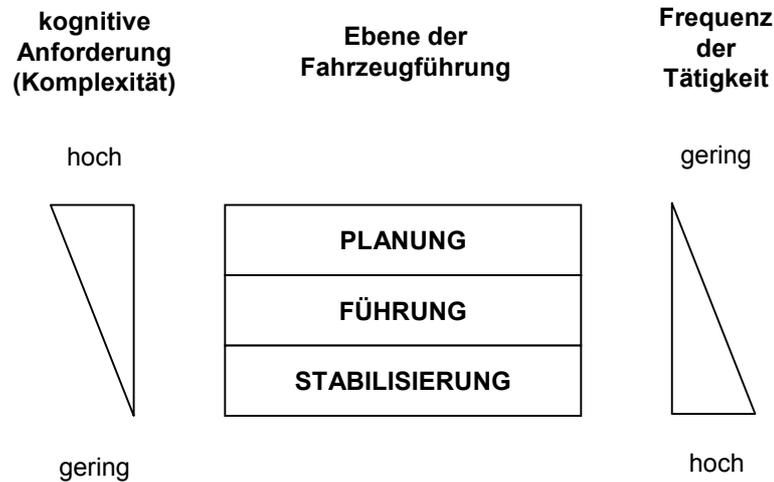


Abbildung 6: Zusammenhang zwischen den Ebenen der Fahrzeugführung und den kognitiven Anforderungen sowie den zeitlichen Randbedingungen (Reichart & Haller, 1995, S.200).

Diese Zusammenhänge sind direkt relevant für die Gestaltung von Mensch-Maschine-Systemen im Kontext der Fahrzeugführung. Die bereits erwähnte Gefahr einer Interferenz durch visuell-manuelle Zusatzaufgaben mit Aktivitäten auf der Stabilisierungsebene spielt aufgrund der Häufigkeit auftretender Fahraufgaben auf dieser Ebene eine besondere Rolle. Entsprechend lässt sich aus dem Modell direkt ableiten, dass derartige nicht direkte fahrrelevante Aufgaben in angemessener Weise gestaltet werden müssen, um die Ablenkung zu reduzieren. Ferner muss beim Design berücksichtigt werden, dass kurze Phasen hoher kognitiver Aktivität auftreten können. Systemausgaben müssen so beschaffen sein, dass sie die Leistungsfähigkeit auf den höheren Ebenen nicht beeinträchtigen.

Entscheidend für das Verständnis der Fahraufgaben ist, dass die Fahrsituation aus einer Wechselwirkung zwischen den Handlungen des Fahrers und den Handlungen weiterer Verkehrsteilnehmer in der Interaktionssituation Fahrer – Fahrzeug – Umwelt resultiert. Dabei können jederzeit Änderungen in den Aufgabenebenen eintreten. So können nach einer Phase des Überwiegens von Aufgaben auf der Stabilisierungsebene Planungs- und Entscheidungsaufgaben hinzukommen, die zum Beispiel durch das Auftreten eines Hindernisses ausgelöst wurden und eine neue Situationsbewertung inklusive der Entscheidung über durchzuführende Manöver unter Beachtung relevanter Randbedingungen notwendig machen.

3.1.2 Das Handlungsregulationsmodell

Nach Sheridan und Ferrell (1974) ist eine Mehrebenen-Hierarchie von Aufgaben typisch für die meisten Formen der Bedienung von Maschinen. Das bedeutet zum einen, dass viele Ergebnisse aus der Forschung in der speziellen Aufgabensituation Fahrzeugführung unter Beachtung situativer Randbedingungen generalisierbar sind. Zum anderen deutet es darauf hin, dass sich die zu Grunde liegenden Verhaltensmuster auch in allgemeinerer Form darstellen lassen. Eine Verhaltensklassifikation, die nicht auf die Fahrzeugführung beschränkt ist, dennoch gleichermaßen gut darauf anwendbar ist, erstellte Rasmussen (1983, 1986) im Modell des wissens-, regel- und fertigkeitstypischen Verhaltens. Dieses Handlungsregulationsmodell wurde im Kontext der Gestaltung von Leitwarten entwickelt und beschreibt Verhalten allgemein in drei kaskadierten Regulationsebenen, die den Mehrebenen-Modellen der Fahrzeugführung ähneln.

Die unterste Ebene bildet bei Rasmussen (1983) das fertigkeitstypische Verhalten (*skill-based behavior*), auch gewohnheitsbasiertes Verhalten genannt, mit automatisierten Reiz-Reaktions-Verknüpfungen und stereotypen Reaktionsmustern. Handlungen basieren auf sensumotorischen Leistungen, die ohne bewusste kognitive Anstrengung reguliert werden. Diese Ebene ist mit der Stabilisierungsebene der Fahrzeugführung eng verwandt. Jedoch können auch Aktivitäten der Planungs- und Führungsebene fertigkeitstypisch erfolgen, wenn es sich um häufig wiederholte, sehr vertraute und so (teil)automatisierte Navigations- oder Entscheidungsaufgaben handelt. Auf der mittleren Ebene des regelbasierten Verhaltens (*rule-based behavior*) werden gespeicherte Regeln zur Aufgabenausführung in Form von Situations-Handlungs-Verknüpfungen abgerufen. In bekannten Arbeitssituationen können so kognitive Routinen regulatorisch angewandt werden. Regelbasiertes Verhalten entspricht der Führungsebene der Fahrzeugführung. Auch hier können Abweichungen existieren. Soll beispielsweise ein unbekanntes Fahrzeug bewegt werden, müssen eigentlich bereits automatisierte Vorgänge erneut regelbasiert ausgeführt werden. Die höchste Ebene des wissens- bzw. modellbasierten Verhaltens (*knowledge-based behavior*) kommt in unbekannten Situationen zum Tragen, für die noch keine bewährten Handlungsmuster vorliegen. Unter Aufwendung bewusster kognitiver Steuerung laufen Planungs- und Entscheidungsprozesse sowie Problemlöseaktivitäten ab, die auf der Grundlage vorhandenen Wissens und einer Situationsanalyse in Handlungsplänen und gegebenenfalls Alternativplänen resultieren. Ein Großteil dieser Prozesse kann analog zur Planungs- und Navigationsebene der Drei-Ebenen-Modelle verstanden werden. Jedoch in unbekannten Situationen wie beispielsweise dem Erlernen des Autofahrens oder bei schlechten Sichtverhältnissen tritt auch bei niedrigeren Ebenen der Fahrzeugführung wissensbasiertes Verhalten auf.

Zusammenfassend erklärt dieses übergreifende Modell, weshalb selbst bei eigentlich einfachen Fahraufgaben durch kognitive Zusatzbelastungen („innere Ablenkung“) Interferenzen entstehen können (z.B. Rumar, 1990). Dieser Aspekt kann nicht ohne weiteres von den Drei-Ebenen-Modellen erfasst werden. Das Modell zeigt darüber hinaus auch den Rahmen auf, innerhalb dessen Befunde aus der Forschung an Mensch-Maschine-Schnittstellen für den Anwendungsfall Fahrzeugführung auf andere Domänen verallgemeinert werden können.

3.2 Handlungssituation

Die theoretischen Modelle der Fahrzeugführung geben ausführliche Auskunft über die Fahraufgabe an sich und beschreiben somit zunächst eine Single Task Situation. Bedenkt man die massive Zunahme von Informationssystemen im Fahrzeug sowie von nicht fahrrelevanten Funktionen (z.B. Komfortfunktionen) und die Nutzung mobiler Geräte wird die resultierende komplexe Mehraufgabensituation offensichtlich. Bei der Analyse des Systems Fahrer-Fahrzeug ist deshalb eine Unterteilung der resultierenden Handlungssituation sinnvoll, die diesen Zusatzbelastungen Rechnung trägt. Ferner muss analysiert werden, in welcher Form durch Zusatzaufgaben die Hauptaufgabe Fahrzeugführung beeinflusst wird. Dies kann sich durch Ablenkung bzw. Abwendung des Fahrers und durch Fehlhandlungen zeigen.

3.2.1 Primär-, Sekundär- und Tertiäraufgaben

Bubb schlägt eine dreigliedrige Unterscheidung der Aufgaben der Fahrzeugführung vor (z.B. Bubb, 2003; Rühmann & Bubb, 2006). Danach kann man unterscheiden zwischen der primären Fahraufgabe, sekundären Aufgaben zur Unterstützung dieser Primäraufgabe und tertiären Aufgaben ohne inhaltlichen Bezug zum Fahren.

Die Fahraufgabe an sich ist dabei bestimmt durch die Notwendigkeit, das Fahrzeug dem Straßenverlauf, den Umgebungsbedingungen und anderen Verkehrsteilnehmern angepasst unter Beachtung der Verkehrsregeln dem Fahrziel zu nähern. Diese Aufgabe ist wie oben ausgeführt durch die Mehrebenen-Modelle der Fahrzeugführung darstellbar.

Sekundäre Aufgaben unterstützen die primäre Fahraufgabe, indem zum einen auf Umgebungsbedingungen reagiert wird und zum anderen aktive Handlungen eingeleitet werden. Während Kuppeln, Schalten, das Einschalten des Tempomat oder das Betätigen der Wischanlage zu den reaktiven Aufgaben zählen, ist z.B. das Setzen des Blinkers als kommunikativer Akt mit anderen Verkehrsteilnehmern eine aktive Aufgabe.

Tertiäre Aufgaben haben keinen inhaltlichen Zusammenhang zur Fahraufgabe. Die hierfür erforderlichen Bedienvorgänge werden zwar während der Fahrt durchgeführt, dienen jedoch ausschließlich Komfort-, Informations- oder Unterhaltungsbedürfnissen. Dazu zählen z.B. die Interaktion mit dem Radio oder Funktionen zur Musikwiedergabe, die Bedienung der Klimaanlage oder Telefonieren. Diesen Aufgaben kommt bei der Gestaltung der Mensch-Maschine-Schnittstelle im Fahrzeug deshalb eine besondere Bedeutung zu, weil sie aus Sicht der Verkehrssicherheit unnötig von der Fahraufgabe ablenken und diese Ablenkungswirkung durch den Einzug moderner Informations- und Kommunikationstechnologie stetig zunimmt.

Die vorliegende Arbeit fokussiert bei der Betrachtung multimodaler Ein- und Ausgabe auf tertiäre Aufgaben. Der Schwerpunkt liegt dabei auf der Interaktion mit dem Bordsystem (z.B. Information, Kommunikation, Komfortfunktionen). Nachdem multimodale Gestaltung das Ablenkungspotential sowohl von Sekundär- als auch von Tertiäraufgaben minimieren soll, ist für die weitere Diskussion die klassische Unterscheidung zwischen der Hauptaufgabe Fahrzeugführung und Nebenaufgaben hinreichend genau. Die Nebenaufgaben werden im weiteren Verlauf nicht explizit bezüglich sekundärer und tertiärer Aufgaben differenziert.

3.2.2 Ablenkung und Abwendung

Bei der Gestaltung von Mensch-Maschine-Schnittstellen für Nebenaufgaben im Fahrzeug ist eine Reduktion der Ablenkungswirkung allein schon deshalb von zentraler Bedeutung, weil visuelle Ablenkung durch Auslöser innerhalb des Fahrzeugs eine der häufigsten Ursachen für Verkehrsunfälle darstellt (Wierwille & Tijerina, 1995, 1997) bzw. zumindest als Teilfaktor zu einem Unfall beiträgt. Selten wird bei dieser Diskussion jedoch der zentrale Begriff der Fahrerablenkung explizit definiert, weshalb keine allgemein akzeptierte Definition dieses Terminus existiert. Gelau (2004) schlägt vor, Fahrerablenkung durch einen Rückgriff auf das Belastungs-Beanspruchungs-Konzept zu definieren: Mit Fahrerablenkung sind danach „durch Eigenschaften der Person vermittelte Wirkungen von Belastungen – als Beanspruchungen – gemeint, die aus den Anforderungen der Aufgaben entstehen, die dem Fahrer neben der reinen Fahrtätigkeit im Fahrzeug gestellt werden oder die er sich selbst stellt“ (Gelau, 2004, S.299). Dieser Ansatz impliziert eine potentielle Ablenkung durch alle Formen der sekundären und tertiären Aufgaben. Gleichzeitig ermöglicht er durch den Verweis auf die subjektiv empfundene Beanspruchung verschiedener Aufgabenformen zu erklären, weshalb teils komplexe Zusatzaufgaben nicht ablenkend wirken. Ist eine derartige Nebenaufgabe in ausreichendem Maß durch Übung automatisiert, wird ihre Interferenzwirkung minimiert.

Ein etwas anderer Blickwinkel ergibt sich, wenn man das Konzept der Aufmerksamkeit aktiv in die definitorischen Bemühungen einbindet. Aufmerksamkeit kann als Vorgang einer selektiven Wahrnehmung verstanden werden, bei dem die Ressourcen zur Informationsverarbeitung auf bestimmte Aspekte der sensorischen Umgebung gerichtet sind, bzw. eine selektive Reaktion auf eine bestimmte Klasse von Reizen bewirkt wird (vgl. Kolb & Wishaw, 1996, S.540). Für eine erfolgreiche Ausführung der Primäraufgabe muss die Aufmerksamkeit auf das Verkehrsgeschehen und die in diesem Kontext notwendigen Reaktionen gerichtet sein. Jede Form der Ressourcenzuwendung zu einer Aufgabe, die damit nicht in direkter Beziehung steht, kann somit als Ablenkung begriffen werden. Bedenkt man, dass unter geteilter Aufmerksamkeit mehrere Aufgaben gleichzeitig bearbeitet werden können, muss Ablenkung als stimulusinduzierte Veränderung der Ressourcenallokation beim Time-Sharing zum Nachteil der Primäraufgabe konzipiert werden. Der Vorteil dieser Konzeptualisierung gegenüber der zuvor genannten Definition ist, dass die direkte Verbindung zu einer Allokation von Ressourcen und damit zu theoretischen Modellen der Informationsverarbeitung wie dem multiplen Ressourcenmodell von Wickens (1984, 2002) ersichtlich wird. Somit können Vorhersagen abgeleitet werden, aufgrund welcher Charakteristika Zusatzaufgaben ablenkend wirken können: Kommen unabhängige perzeptive Ressourcen zum Einsatz und wird kein kognitiver Bottleneck durch die zusätzlich notwendigen mentalen Prozesse erzeugt, sollte keine wesentliche Ablenkung stattfinden. Nachdem bei der Fahraufgabe die visuelle Informationsaufnahme dominiert (Evans, 1991; Hills, 1980), ist durch die zusätzliche Aufnahme nicht-visueller Information wenig Interferenz zu erwarten, sofern durch die akustische oder haptische Ausgabe nicht die visuelle Aufmerksamkeit auf den Ort der Ausgabe, z.B. einer Anzeige, gezogen wird.

Mit dem Begriff Ablenkung allein lassen sich mögliche negative Auswirkungen zusätzlicher Systeme im Fahrzeug nicht ausreichend diskutieren und beschreiben. Zimmer (2001b) weist auf die Notwendigkeit einer Unterscheidung zwischen *Ablenkung* und *Abwendung* hin. Bei Ablenkung löst ein externes Signal, z.B. ein lautes Störgeräusch oder bewegte Inhalte auf der Anzeigeeinheit des Bordsystems, beim Fahrer eine Reaktion aus. Aus der Perspektive der Reizverarbeitung handelt es sich um ein Bottom-Up Phänomen. Bei einer Abwendung hingegen entscheidet sich der Fahrer bewusst aufgrund der augenblicklichen Situation zur Bedienung eines Geräts oder für das Ablesen einer Anzeige, was als Top-Down Prozess eine willentliche Aufmerksamkeitslenkung infolge eines Informationsbedarfs darstellt. Es konnte nachgewiesen werden, dass die unterschiedliche Initiierung dieser beiden Prozesse der Aufmerksamkeitsverschiebung von der Fahraufgabe unterschiedlichen Einfluss auf das Fahrverhalten haben kann (Dahmen-Zimmer et al., 1999). Die Ergebnisse von Dahmen-Zimmer und Kollegen weisen vor allem darauf hin, dass speziell bei Fahranfängern eine überwiegend externale Steuerung der Aufmerksamkeit zu verzeichnen ist und es bei diesen

somit besonders leicht zu Ablenkung kommen kann. Zur Minimierung negativer Auswirkungen durch Abwendung existieren zwei Lösungsansätze. Erstens sollten Bedienvorgänge möglichst gut unterbrechbar sein. Experimentell kann beispielsweise mit Hilfe der Okklusionsmethode erhoben werden, ob Interaktionsvorgänge segmentiert und unterbrechbar sind (vgl. Krems, Keinath, Baumann & Jahn, 2004). Des Weiteren darf der Fahrer nicht zu Eingabevorgängen gezwungen sein, wenn die Verkehrssituation dies nicht gestattet. Zweitens kann im Rahmen multimodaler Lösungsansätze versucht werden, durch eine geeignete Wahl von Modalitäten die visuelle Abwendung zu verkürzen oder unnötig zu machen. Dennoch muss beachtet werden, dass durch eine zu starke kognitive Beanspruchung bei der Bedienung trotz einer weit reichenden Implementierung unabhängiger Modalitäten die perzeptuelle Verarbeitungsfähigkeit des Fahrers auf kognitiver Ebene eingeschränkt werden kann.

Zusammenfassend wirken Bordsysteme dann ablenkend, wenn sie die Aufmerksamkeit des Fahrers entgegen seiner eigentlichen Intention auf sich ziehen. Dieser Aspekt muss bei multimodaler Ausgabe stets bedacht werden. Eine bewusst initiierte Abwendung wird dann problematisch, wenn die Beschaffenheit der Zusatzaufgabe negativen Einfluss auf das Time-Sharing zwischen der Haupt- und der Nebenaufgabe nach sich zieht, indem sie zu viele Ressourcen von der Fahraufgabe abzieht. Durch das Ermöglichen flexibler Eingabewege und durch nicht-visuelle Ausgabe kann multimodale Interaktion eine bessere zeitliche Strategie der Abwendung unterstützen.

3.2.3 Fehlhandlungen im Fahrzeug

Für ein fundiertes Verständnis der Auswirkungen von nicht-automatisierten und damit aufmerksamkeitspflichtigen Nebenaufgaben im Fahrzeug ist eine umfassende Analyse menschlicher Fehlhandlungen erforderlich. Eine systematische Betrachtung von Handlungsabläufen mit einer Aufgliederung der sukzessiven Stadien der Ausführung ermöglicht die Erkenntnis spezifischer Einstiegspunkte für die Optimierung des Mensch-Maschine-Systems. Die allgemeine Unterscheidung zwischen Fehlertypen (vgl. Abschnitt 2.2.3) soll deshalb an dieser Stelle um ein speziell für die Fahrzeugführung adaptiertes Modell zur Fehlerbeschreibung mit einer sequentiellen und kausal verknüpften Verlaufsbeschreibung der Informationsaufnahme und Handlungsausführung ergänzt werden. Aufbauend auf einem Ansatz von Rasmussen (1982) haben Wiegmann und Shappell (1997) den Handlungsablauf bei Fehlern für die Flugzeugführung modelliert. Von Zimmer (2001b) wurde das Modell für die Fahrzeugführung weiterentwickelt (siehe Abbildung 7).

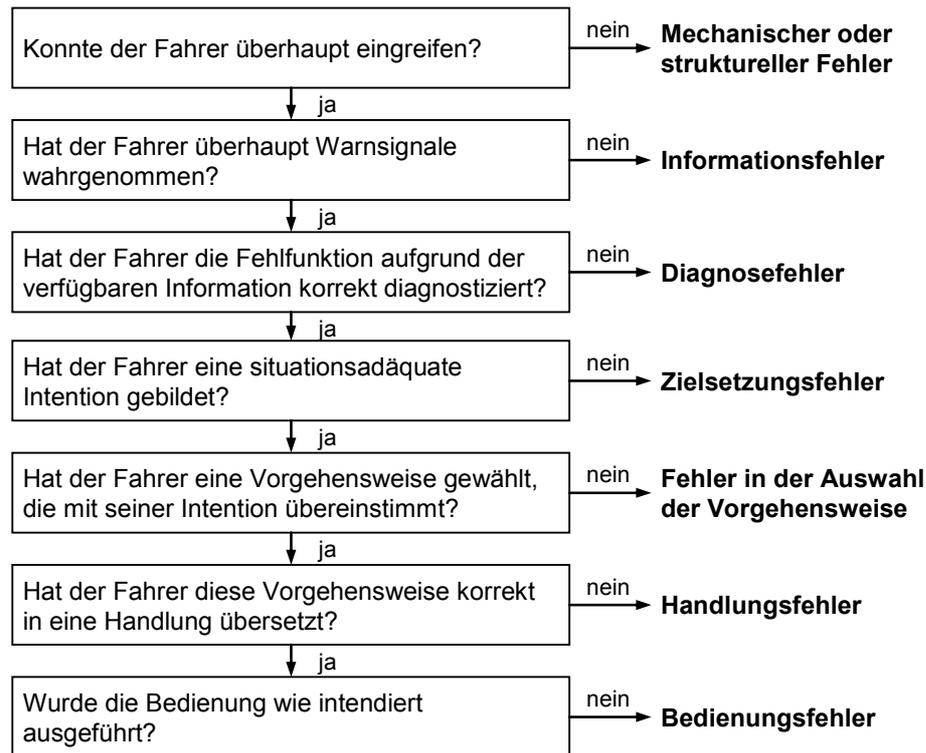


Abbildung 7: Fehlerarten und -ursachen nach dem für die Fahrzeugführung angepassten Modell der internalen Fehlfunktionen (Zimmer, 2001b, S.45).

Der Vorteil dieses Modells ist eine umfassende Darstellung der Vorgänge zur Informationsaufnahme und der Ausführung von Handlungen, die für einen Unfall ursächlich sind. Es ist damit ein besonders wirksames Werkzeug zur Analyse von Unfällen. In einem ersten Schritt wird die Frage aufgeworfen, inwiefern der Fahrer überhaupt eine Chance zum Eingriff hatte. Wenn zum Beispiel unerwartet ein Tier auf der Landstraße vor das Fahrzeug läuft und der Fahrer keine Gelegenheit mehr zum Einleiten von Gegenmaßnahmen hat, kann ihm kein Fehlverhalten vorgeworfen werden. In diesem Fall liegt ein *struktureller Fehler* vor. Ebenso ist bei einem Unfall aufgrund eines technischen Defekts am Auto kein Fehler des Fahrers zu verzeichnen, Unfallverursacher war dann ein *mechanischer Fehler*. Bei einem *Informationsfehler* hat der Fahrer handlungsrelevante Information nicht aufnehmen können, beispielsweise weil er mit einer nicht fahrrelevanten Nebenaufgabe beschäftigt war. Als *Diagnosefehler* kann ein Fehler eingestuft werden, wenn situative Gegebenheiten bei der Entscheidung über Handlungen falsch eingeschätzt werden, so dass es entgegen der Vermutung des Fahrers zu einem Unfall kommt (z.B. Fehleinschätzungen von Geschwindigkeiten und Entfernungen anderer Verkehrsteilnehmer). Bei einem *Zielsetzungsfehler* erkennt der Fahrer die Situation zwar richtig, bildet jedoch eine Handlungsabsicht, die für diese Situation nicht angemessen ist. Diese Absicht ist bei einem *Fehler in der Auswahl der Vorgehensweise* oder auch *Methodenfehler* zwar richtig gebildet, wird aber nicht in der beabsichtigten Form ausgeführt. So mag sich

der Fahrer korrekt für das Ziel „ausweichen“ bei einem Hindernis entschieden haben, bei der Wahl zwischen „links ausweichen“ und „rechts ausweichen“ jedoch die unfallträchtigere Variante gewählt haben. Die Gefahr eines *Handlungsfehlers* besteht dann, wenn alle Entscheidungen zwar richtig getroffen wurden, die Ausführung des Vorgangs „ausweichen“ aber misslingt, beispielsweise durch eine gleichzeitig aus Versehen zu stark ausgeführte Notbremsung. Rutscht der Fahrer bei einer geplanten Notbremsung vom Pedal ab, liegt ein *Bedienungsfehler* vor. Der Unterschied zum Handlungsfehler liegt darin, dass der Bedienungsfehler in einer Störung der Ausführung an der Mensch-Maschine-Schnittstelle liegt. Bei einem Handlungsfehler hingegen ist der Fehler beim Fahrer selbst zu verorten, der die Handlung motorisch unzulänglich initiiert hat.

Neben der Unfallanalyse zeigt dieses Modell klar auf, welche Arten von Fehlern durch eine gute Gestaltung der Schnittstelle für Nebenaufgaben reduziert werden können. Es handelt sich dabei hauptsächlich um diejenigen Fehler, die durch fehlende oder unvollständige Informationen während einer Ablenkung oder Abwendung vom Verkehrsgeschehen entstehen. Sowohl auf der Ebene der Informationsfehler als auch begrenzt auf der Ebene der Diagnosefehler können durch multimodale Interaktion Probleme vermieden werden. So kann durch Anbieten eines Interaktionsprozesses über den akustischen Kanal mit Sprachein- und Ausgabe oder durch zusätzliche nicht-visuelle Systemrückmeldungen bei manueller Eingabe die Wahrscheinlichkeit für eine bessere Konzentration der visuellen Ressourcen auf den Verkehr erhöht und Informationsfehler vermieden werden. Wird durch multimodales Design die Mensch-Maschine-Interaktion effizienter und intuitiver, sind darüber hinaus kognitive Ressourcen in geringerem Ausmaß und zeitlich kürzer gebunden. Dadurch lässt sich die Wahrscheinlichkeit für Diagnosefehler herabsetzen. Durch Nebenaufgaben steigt ferner die Gefahr von Handlungs- und Bedienfehlern bei der Fahrzeugführung, wenn wegen einer ungünstigen Lokalisierung von Schaltern und Anzeigen eine motorisch nachteilige Körperhaltung eingenommen werden muss (z.B. Vorbeugen zur Mittelkonsole). Im Notfall kann dann eine Korrektur des Fahrverhaltens nicht mit der gleichen Güte erfolgen wie bei der Steuerung des Fahrzeugs aus einer ergonomisch entspannten Körperhaltung. Eine gute Sitzposition kann durch die Eingabe per Sprache unterstützt werden. Aktuelle Bedienkonzepte versuchen zusätzlich, möglichst viele Komfortfunktionen durch Einbettung in ein schalterreduziertes Bedienkonzept mit ergonomisch sinnvoller räumlicher Trennung von Anzeige und Eingabe erreichbar zu machen. Dies wird im nächsten Abschnitt thematisiert.

3.3 Bedienkonzepte für Bordsysteme

Frühere Automobil-Generationen wurden gerne in der Ergonomie-Literatur als gutes Beispiel dafür genannt, dass auch komplexe Steuerungsfunktionen relativ leicht erlernbar sind. Wenngleich die Anforderungen an die Geschicklichkeit angehender Fahrer nie unerheblich waren, gehorchte die Gestaltung jeglicher Stellteile immer einem einfachen Prinzip: Ein Schalter, eine Funktion (Norman, 1988). Dieses Prinzip konnte durch die Integration neuer Funktionen nicht mehr aufrechterhalten werden, da einer stetig steigenden Anzahl von Bedienkomponenten und Anzeigen der konstant enge Bauraum auf der Armaturentafel und der Mittelkonsole und der beschränkte Greifraum des Fahrers entgegen standen. Darüber hinaus sind die visuelle Orientierung und die manuelle Zielbewegung in eng gepackten Schalteranordnungen deutlich erschwert. So berichten Schattenberg und Debus (2001) von Fällen mit bis zu 98 Bedienelementen im Cockpit von Oberklassefahrzeugen. Dadurch ist die Mittelkonsole überfrachtet, im Greifraum des Fahrers kein Platz mehr für Erweiterungen, eine Doppelbelegung von Stellteilen erforderlich und eine Bedienung der Systeme im Fahrzeuginnenraum nicht mehr für alle Kundengruppen möglich. Als besonders problematisch erweist sich, dass auch vergleichsweise selten genutzte Komfortfunktionen über eigene Schalter und Funktionsanzeigen verfügen und so den Zugriff auf häufig genutzte, teils fahrerunterstützende Sekundärfunktionen erschweren. Schattenberg und Debus (2001) weisen darauf hin, dass für eine Erhöhung der Sicherheit während des Ablesens der Anzeigen der Straßenverlauf zumindest in Form der peripher wahrgenommenen Spurinformatio n weiterhin verfolgt werden können muss. Als Voraussetzung für die notwendige freie Platzierung des Informationsdisplays geben sie die räumliche Trennung von Anzeige und Bedienkomponenten in einer so genannten „schalterreduzierten integralen Mensch-Maschine-Schnittstelle“ an (Schattenberg & Debus, 2001, S.182).

Im Kern hat sich diese Überlegung in aktuellen Bordsystemen zumindest von Oberklassefahrzeugen, teilweise auch in der oberen Mittelklasse, weitgehend durchgesetzt. Während häufig genutzte Komfortfunktionen (z.B. die Klimaanlage) nach wie vor in der Mittelkonsole oder über das Multifunktionslenkrad per Direktzugriff erreichbar sind, lassen sich Feineinstellungen und alle weiteren Komfortfunktionen mit Hilfe eines zentralen Bedienelements über eine zentrale Anzeigeeinheit in einem menügesteuerten System erreichen (vgl. Ashley, 2001). Dieses Bedienelement ist typischerweise eine Form von Dreh-Drücksteller, teilweise mit zusätzlichen translatorischen Freiheitsgraden. Den ersten Schritt machte BMW mit der Einführung des iDrive-Bedienkonzepts in der 7er Modellreihe des Jahres 2001. Am unteren Ende der Mittelkonsole ist der iDrive Controller über dem Mitteltunnel zwischen Fahrer und Beifahrer montiert und ermöglicht unter anderem den Zugriff auf Navigations-, Telekommunikations-, Audio- und Fahrwerkseinstellungen. Die Anzeigeeinheit für das grafische Be-

diensystem ist in zentraler Position im Armaturenbrett rechts neben dem Lenkrad eingelassen. Durch diese räumliche Trennung rückt die manuelle Eingabe in den primären Greifraum und erhöht so die Bedienergonomie. Die Anzeige befindet sich im primären Sichtfeld und gestattet die periphere Aufnahme verkehrsrelevanter Information auch während der Bedienvorgänge bzw. einen schnellen Wechsel des Fokus der visuellen Aufmerksamkeit zwischen Straße und Display. Dieses Konzept hat den Vorteil, dass nur der zentrale Bildschirm bei der Interaktion mit dem System überprüft werden muss. Die manuelle Interaktion bedarf keiner visuellen Kontrolle, da die wenigen zusätzlich zum Dreh-Drücksteller relevanten Tasten leicht erfüllt werden können. Durch die Auslegung als menübasiertes System lassen sich theoretisch beliebige Erweiterungen integrieren. Eine ähnliche Lösung realisierte Audi mit dem MMI-Bediensystem im A8 des Jahres 2002. Auf Detailunterschiede zwischen den Varianten bezüglich der Funktionsweise des zentralen Eingabeelements und der Anzahl zusätzlicher Bedienelemente soll an dieser Stelle nicht eingegangen werden. Hervorzuheben ist vielmehr, dass auch bei diesem System eine Anzeigeeinheit im primären Sichtfeld mit einem zentralen Dreh-Drücksteller im primären Greifraum existiert. Nachdem seit 2006 auch in der S-Klasse von Mercedes-Benz ein zentraler Controller für das COMAND-System zur Verfügung steht, kann sicherlich von einem allgemeinen Trend zu schalterreduzierten Anzeige- und Bedienkonzepten mit abgesetzter Anzeige für kommende Fahrzeuggenerationen gesprochen werden.

Die Bedeutung dieses Trends konstituiert sich dadurch, dass durch die neuen Bedienkonzepte der Fahrer Arbeitsplatz der Zukunft charakterisiert wird und dass die zu gestaltenden multimodalen Interaktionskonzepte auf dieser Grundlage konzipiert werden müssen. Das gilt insbesondere vor dem Hintergrund experimenteller Evaluationen, die bei Konzepten mit abgesetzter Anzeige und zentralem Bedienelement nachteilige Auswirkungen durch längere Aufgabenbearbeitungszeiten und erhöhte Beanspruchung nachwies (Schattenberg & Debus, 2001). Gerade hier lässt sich durch den Einsatz von Multimodalität eine deutliche Verbesserung der Situation erzielen. Sprachsteuerung kann dazu beitragen, Funktionen über diesen Eingabeweg direkt verfügbar zu machen, die im System erst nach dem Überwinden mehrerer Menüschritte auffindbar sind. Durch die Verwendung einer zentralen Anzeigeeinheit kann gleichzeitig die Spracheingabe vereinfacht werden, indem die Begriffe aus den Menüs gleichzeitig als Sprachbefehle dienen und bei der visuell-manuellen Interaktion nebenbei gelernt werden (*Say What You See*).

Neben der Spracheingabe, die besonders das Potential aufweist, Ressourcenkonflikte auf der Eingabeebene aufzulösen, wird zunehmend Gestensteuerung diskutiert (vgl. u.a. Akyol et al., 2001; Bengler, 2001; Schattenberg & Debus, 2001). Zwar involviert gestenbasierte Eingabe auch die Nutzung der Hände, durch die Berührungslosigkeit entsteht jedoch ein größerer Freiraum bei der Ausführung und dadurch eine geringere

Gefahr der Interferenz mit der Lenkung des Fahrzeugs. So muss zum Beispiel ein Schalter nicht exakt lokalisiert, sondern nur die Hand im Gestenraum bewegt werden. Die Technologien für Gestenerkennung und die entsprechenden Bedienparadigmen sind jedoch gegenwärtig noch nicht produktreif.

Multimodale Rückmeldung muss primär den Zweck erfüllen, dem Nutzer schnell, unmittelbar, verständlich und ohne ablenkend zu wirken eine Rückmeldung über den veranlassten Systemzustand zu geben. So kann die Ausgabe von akustischem und haptischem Feedback für Eingabeoperationen über den zentralen Controller die visuelle Last reduzieren und Blindbedienung ermöglichen. Nachdem gegenwärtig jedoch keine hinreichend gute Möglichkeit zur theoretischen Modellierung der Effekte multimodaler Ein- und Ausgabe auf das Fahrerverhalten existiert, muss jede Maßnahme einer sorgfältigen empirisch-experimentellen Prüfung unterzogen werden. Denn letztlich stellen die multimodalen Bedienvarianten nur Angebote an den Fahrer dar, deren Annahme maßgeblich von der Nutzerakzeptanz abhängt. Akzeptanz wiederum ist nur bei solchen Systemen zu erwarten, bei denen die Verbesserung der Fahrqualität größer ist als die Ablenkung durch diese Systeme (vgl. Zimmer, 2001b) und die nachweislich und dauerhaft zu einer Optimierung der Mensch-Maschine-Interaktion beitragen.

Abschließend soll nicht unerwähnt bleiben, dass es aus verkehrspsychologischer Sicht durchaus bedenklich ist, wenn zu viele Nebenaufgaben neben der Fahrzeugführung entstehen. Sinnvoll wäre in jedem Fall eine Reduktion der Anzahl dieser Funktionen. Allerdings ist davon auszugehen, dass sich aufgrund der Wettbewerbssituation die Hersteller den Nachfragen oder den vermuteten Wünschen der Kunden nicht entziehen können. Solange keine externen Einschränkungen existieren, beispielsweise durch Vorgaben des Gesetzgebers, muss aufgrund des technischen Fortschritts bei mobilen Geräten eher von einer noch zunehmenden Anzahl zusätzlicher Funktionen im Fahrzeug ausgegangen werden. Insofern ist es die Pflicht der angewandten Forschung, neben einer Abschätzung des Gefahrenpotentials durch diese Technologien den Fahrerarbeitsplatz durch neue Interaktionsparadigmen sicherer zu machen und eine gute Bedienbarkeit zu gewährleisten.

4 Untersuchungsabschnitt 1: Aspekte der Ausgabe im multimodalen Umfeld

Der Fokus des ersten Untersuchungsabschnitts liegt auf dem Themenbereich der Gestaltung der Ausgabe von Information eines technischen Systems an den menschlichen Nutzer im multimodalen Kontext. Die Experimente sind darauf ausgerichtet, offene Gestaltungsfragen bei der Implementierung multimodaler Interaktionsprinzipien für Bordsysteme zur Steuerung von Komfortfunktionen im Kraftfahrzeug zu beantworten. Im Detail fokussieren sie auf den Themenbereich der tertiären Aufgaben im Fahrzeug. Eine zentrale Richtvorgabe ist damit, dass die Bedienung von Komfort-, Informations- und Unterhaltungsfunktionen in einer Weise gestaltet wird, die ein Optimum an Ressourcenzuteilung für die Hauptaufgabe Fahrzeugführung gewährleistet. Bedienvorgänge, die damit nicht in direktem Zusammenhang stehen, dürfen nur ein Mindestmaß an Aufmerksamkeit fordern. Aufgrund der Natur der Primäraufgabe gilt es insbesondere, die visuelle Abwendung von der Fahrbahn zu begrenzen. Die zunehmende Integration zusätzlicher Funktionen stellt hier eine Herausforderung an die Gestaltung von Bordsystemen dar. Der Auftrag an die ingenieurpsychologische Forschung lautet deshalb, neue Wege der Interaktion mit den Systemen aufzuzeigen, und diese empirisch zu prüfen. Multimodalität kann ein solcher Weg sein. Das Potential erwächst aus einer Verbesserung des Time-Sharing in der Doppelaufgabensituation durch Verlagerung der Informationsausgabe und Aktionsrückmeldung auf nichtvisuelle Sinneskanäle (vgl. Abschnitt 2.2.2). Berücksichtigt man die mittlerweile verfügbaren Technologien, eignet sich zur Entlastung des visuellen Kanals neben der auditiven Schnittstelle auch eine haptische Ausgabe.

Die anwendungsorientierte Untersuchung dieses Abschnitts beschäftigt sich mit einer konkreten Ausprägung einer multimodalen Schnittstelle. Angetrieben durch die Überlegung, dass sowohl aus technologischen als auch aus ergonomischen Gründen die manuelle Eingabe mit Hilfe eines Touchpads dem typischen zentralen Dreh-Drücksteller in schalterreduzierten Anzeige- und Bedienkonzepten überlegen sein könnte, wurde ein multimodales Rückmeldekonzept für eine Touchpad-Steuerung

entwickelt. Neben visueller Ausgabe verfügt die Schnittstelle über haptisches und auditives Feedback. In zwei Experimenten wurde untersucht, wie sich die Einführung zusätzlicher Ausgabemodalitäten bewährt und inwiefern sich manuelle Eingabe über ein Touchpad bei verschiedenen Ausgabemodalitäten gegenüber einem herkömmlichen Dreh-Drücksteller in Single und Dual Task Situationen eignet.

Die grundlagenorientierte Untersuchung beleuchtet im Detail die Gestaltung von Ausgaben in der auditiven Modalität. Die experimentelle Fragestellung dreht sich dabei nicht um das Zusammenwirken der Modalitäten wie in der anwendungsorientierten Untersuchung, sondern fokussiert auf Forschungsbedarf bei der Rolle der Kodierung akustischer Systemausgaben. Während weitgehend anerkannt ist, dass akustische Ausgaben eine sinnvolle Gestaltungsvariante im multimodalen Umfeld sind, existiert überraschend wenig Wissen darüber, wann sprachlicher oder nicht-sprachlicher Systemoutput zu bevorzugen ist. Beiträge zu diesem Thema beschäftigen sich meist mit elementaren Fragen des Einsatzes von Sprache und nicht-sprachlichen Schallen (z.B. Brewster, 2003), direkte Vergleiche von Gestaltungsvarianten auf experimenteller Basis sind hingegen selten. Für die Implementierung multimodaler Systeme ist dieses Wissen jedoch eine notwendige Voraussetzung. Bei der Ausgabe von Signalen für Nebenaufgaben muss berücksichtigt werden, dass sie nicht mit kognitiven Anforderungen der Hauptaufgabe interferieren dürfen. Dabei muss der aufmerksamkeitslenkende Charakter auditiver Ausgaben besondere Beachtung finden. Durch den Vergleich typischer auditiver Designvarianten in ihren Auswirkung auf das Arbeitsgedächtnis in einem Paradigma der psychoakustischen Forschung soll an dieser Stelle ein Beitrag geleistet werden, der eine Grundlage für fundierte Gestaltungsentscheidungen bilden kann.

4.1 Multimodales Feedback für Touchpads zur zentralen Bedienung von Bordsystemen

4.1.1 Problemstellung und Ausgangspunkt der Experimente

Für viele moderne Bordsysteme existiert als primäres manuelles Eingabeelement zur Steuerung von Komfortfunktionen in schalterreduzierten Anzeige- und Bedienkonzepten neben wenigen Funktionstasten und Softkeys für den Direktzugriff ein zentral über dem Mittelunnel angebrachter Dreh-Drücksteller. Die visuelle Rückmeldung der Interaktion erfolgt über ein Display. Dieses ist typischerweise rechts neben dem Lenkrad in zentraler Position in oder oberhalb der Mittelkonsole montiert, wodurch relevante Verkehrsinformationen im unmittelbaren peripheren Blickfeld sind. Für eine verkehrssichere Bedienung sollte die Interaktion so ausgelegt sein, dass die Anzeige-

einheit trotz der speziellen Positionierung so wenig visuelle Zuwendung wie möglich erfordert.

Das Dreh-Drücksteller-Konzept hat aus psychologischer Sicht den Nachteil, dass bei der Navigation durch Menüs oder dem Einstellen von Funktionsparametern bei den meisten Realisierungen dieser Lösung lineare Bewegungen auf dem Bildschirm durch den Nutzer auf die auszuführende Drehbewegung umgesetzt werden müssen. Darüber hinaus können keine zweidimensionalen Bewegungen realisiert werden, wie sie beispielsweise bei Buchstabenfeldern für die Eingabe eines Navigationsziels sinnvoll wären. Hier muss die Bewegung durch die zweidimensionale Buchstabenmatrix in zeitaufwändiges, zeilenweises Navigieren umgesetzt werden. Im Sinne einer intuitiven Gestaltung wird durch diese Aspekte ferner der Grundsatz des natürlichen Mapping⁴ (Norman, 1988) missachtet. Eine intuitive Zuordnung im gegebenen Kontext wäre zum Beispiel dadurch gegeben, dass die Drehung des Drehwählers im Uhrzeigersinn zu einer Cursorbewegung nach rechts führt, denn diese Bewegung ist analog zur Lenkradbewegung. Gleichzeitig existieren jedoch Steuerelemente wie Listen oder Menüs, die vertikale Cursorbewegungen erfordern. Aus kognitionsergonomischer Sicht ist die resultierende Doppelbelegung der Drehbewegung wenig vorteilhaft. Darüber hinaus können nicht beide Mappings, horizontal und vertikal, natürlich sein. In der Praxis wird ferner die Zuordnung von Drehrichtung und vertikaler Cursorbewegung von verschiedenen Herstellern unterschiedlich gelöst. Erschwerend für die Bedienbarkeit kommt hinzu, dass die Abbildung bei Bewegungen in zweidimensionalen Anordnungen wie dem erwähnten Buchstabenfeld, in denen sich der Cursor in der Matrix spalten- oder zeilenweise bewegt, arbiträr ist.

4.1.1.1 Touchpads als Ersatz für Dreh-Drücksteller

Diese Probleme sind durch ein manuelles Eingabegerät lösbar, das zweidimensionale Bewegungen zulässt und dadurch die Bewegungen auf der Fläche des Displays natürlich abbildet. Als mögliche technische Realisierung wird vorgeschlagen, dass der Dreh-Drücksteller zur zentralen Steuerung von Zusatzfunktionen durch ein Touchpad ersetzt wird. Im Unterschied zum Einsatz beim PC soll das Touchpad keinen Mauscursor bewegen, mit dem Point-and-Click Operationen ausgeübt werden. Vielmehr soll ein Selektionscursor wie bei dem Dreh-Drücksteller-Konzept zwischen anwählbaren Elementen alternieren. Die Eingabe kann im Touchpad-Konzept dadurch sehr einfach erfolgen, dass sich der Selektionscursor auf der Bedienoberfläche kongruent zur Fin-

⁴ Norman (1988, S.23) definiert Mapping im Mensch-Maschine Kontext folgendermaßen: „*Mapping* is a technical term meaning the relationship between two things, in this case between the controls and their movements and the results in the world.[...] Natural mapping, by which I mean taking advantage of physical analogies and cultural standards, leads to immediate understanding. For example, a designer can use spatial analogy: To move an object up, move the control up.”

gerbewegung verhält. Zur Bestätigung der Selektion, d.h. der Auslösung der mit dem markierten Element assoziierten Aktion, wird der Finger kurz auf die berührungssensitive Fläche getippt. Auf diesem Weg entfällt auch das Umgreifen, wie es bei Dreh-Drückregler-Konzepten meist erforderlich ist (siehe Abbildung 8).



Abbildung 8: Eingabe mit Hilfe eines Touchpads (links) und eines Dreh-Drückstellers (rechts). Für Selektions- und Bestätigungsbewegungen kann bei Touchpads dieselbe Handstellung beibehalten werden.

Nachdem das Touchpad wie die Anzeigeeinheit eine rechteckige Form aufweist, wird eine natürliche Abbildung sowohl für vertikale als auch für horizontale Bewegungen geschaffen. Die Eingabe in zweidimensionalen Feldern wird durch das Ermöglichen horizontaler, vertikaler und diagonaler Bewegungen deutlich vereinfacht. Die Art der Eingabe sollte leicht erlernbar sein, da die Verwendung von Touchpads bei Laptops sehr verbreitet ist. Wie beim Dreh-Drücksteller-Konzept wird die ergonomisch sinnvolle Trennung von Anzeige- und Bedienort erhalten. Daneben existieren einige relevante technische Vorteile dieser Lösung. Mit einer entsprechenden Erkennungstechnologie ausgestattet, ist auf Touchpads Handschrifterkennung möglich. Die Buchstaben können direkt mit dem Finger auf die Fläche gezeichnet werden und vereinfachen die Eingabe größerer Textmengen. Zusätzlich können auf dem Touchpad verschiedene Funktionsbereiche implementiert und zweidimensionale Gesten definiert werden, wie sie aus dem Desktop-Bereich als so genannte Mausgesten bekannt sind. Das stellt eine zusätzliche Möglichkeit dar, zukünftige Schnittstellen mit einer leicht realisierbaren Form der Gestenerkennung und damit einer weiteren Eingabemodalität auszustatten.

4.1.1.2 Multimodale Ausgabe für Touchpads und Überblick über die Experimente

Ein massiver Nachteil des Touchpads ist das Fehlen einer unmittelbar wahrnehmbaren Rückmeldung. Im Gegensatz zu einem herkömmlichen Dreh-Drücksteller, der beim Drehen Klick-Geräusche macht und durch seine Rasterung spürbare Aktionsrückkopplung bietet, sind die Aktionen auf dem Touchpad nur auf der grafischen Bedienoberfläche nachvollziehbar. Zur Lösung des Feedback-Problems wurde ein multimodales Touchpad entwickelt. Zusätzlich zur visuellen Rückmeldung von Selektions- und

Bestätigungsvorgängen können akustische Signale an den Nutzer ausgegeben und über entsprechende frei programmierbare Aktuatoren vibrotaktile Rückmeldungen gegeben werden. Die taktile Rückmeldung besteht unter anderem in einem taktilen Impuls, der unterhalb des Touchpads erzeugt wird, wenn der Selektionscursor von einem Steuerelement der Bedienoberfläche auf eine anderes gesetzt wird. Zusätzlich kann ein über Lautsprecher ausgegebenes Klick-Geräusch für die auditive Wahrnehmung das Ereignis signalisieren. Eine genaue Beschreibung des Prototyps und der Aufgaben findet sich in 4.1.2.1 (Methode). Die Idee einer erweiterten Rückmeldung für Touchpads geht auf ein Konzept von MacKenzie und Oniszczak (1997, 1998) für Laptop und Desktop Computer zurück. Sie haben vorgeschlagen, dass die Interaktion mit Touchpads intuitiver wäre, wenn auch bei Touchpads Eingabeoperationen ähnlich wie bei einer Computer-Maus zu einem hör- und spürbaren Klick führen würden. Die auditive Rückmeldung war bei MacKenzie und Oniszczak allerdings nicht frei variierbar, sondern mechanisch an das Auslösen eines taktilen Impulses gekoppelt. Darüber hinaus wurden ausschließlich Klick-Ereignisse rückgemeldet.

Für die hier berichteten Experimente lassen sich zwei Ziele identifizieren. Von primärem Interesse ist eine Analyse des Beitrags verschiedener Ausgabemodalitäten bei der Interaktion mit dem Touchpad. Die zentrale Frage ist, inwiefern zusätzlich zur visuellen Anzeige haptische und/oder akustische Ausgaben von den Versuchspersonen interpretiert und genutzt werden können. Kritische Komponenten im stark visuell geprägten Kontext der Fahrzeugführung sind kürzere Blickabwendungen, eine bessere Spurhaltung bei der Fahraufgabe und kürzere Aufgabenbearbeitungszeiten. So ist bereits bekannt, dass in einer Aufgabenumgebung mit einer dominanten visuellen Komponente akustische Ausgaben die Leistung beim Time-Sharing, der zeitlichen Allokation bei der Bearbeitung paralleler Aufgaben, verbessern können (Liu, 2001; Wickens et al., 1983). Hierfür wird variiert, ob zusätzlich zur visuellen Rückmeldung bei dem Touchpad auditives, taktilen, kombiniert auditives und taktilen oder kein weiteres Feedback gegeben wird. Dadurch soll insbesondere Wissen über die Wirksamkeit taktilen Feedbacks bei der Mensch-Maschine-Interaktion generiert werden. Erst in jüngerer Vergangenheit hat diese Rückmeldevariante unter dem Schlagwort *Tactons*⁵ vermehrt Beachtung gefunden (Brewster & Brown, 2004; Brown et al., 2005). Obwohl schon seit langer Zeit taktile Informationsausgabe zur Entlastung des visuellen Kanals vorgeschlagen wurde (z.B. Zimmer, 1998), werden taktile Displays im Kontext multimodaler Interaktion erstaunlich wenig eingesetzt (van Erp, 2002; Williams, Miche-

⁵ Tactons können nach Brewster und Brown (2004, S.15) als strukturierte, abstrakte Ausgaben definiert werden, die durch Berührungsreize vermittelt werden. Typischerweise handelt es sich dabei um vibrotaktile Muster, die bestimmten Systemereignissen zugeordnet werden.

litsch, van Veen & Goronzy, 2004). Die Erfahrungswerte aus dem Experiment sollen dem Bedarf an systematischer Evaluation Rechnung tragen.

Das zweite Ziel ist eine experimentelle Evaluation des Touchpads als Eingabevariante zur Steuerung von Komfortfunktionen im Fahrzeug. Im Vergleich zu einem herkömmlichen Dreh-Drücksteller als Baseline soll das Auftreten der oben genannten potentiellen Vorteile überprüft werden, die sich durch eine natürlichere Abbildung und schnellere Eingabeform ergeben können. Für das Experiment existieren insgesamt fünf verschiedene Ausprägungen der Schnittstelle zwischen Nutzer und simuliertem System: eine Version des Dreh-Drückstellers und vier Versionen des Touchpads mit verschiedenen Feedbackvarianten.

Im ersten Experiment erfolgt die Untersuchung im Fahrsimulator mit unterschiedlichen Belastungsbedingungen (Bedienung während der Fahrt bzw. Bedienung ohne Zusatzbelastung), um die Effekte auftretender Ressourcenkonflikte bei der Fahrzeugführung zu analysieren. Im zweiten Experiment wird überprüft, inwiefern die verschiedenen Feedbackvarianten der Touchpadsteuerung dazu geeignet sind, Eingaben ohne visuelle Kontrolle zu unterstützen. Dieser Aspekt wird durch einen Versuch zur Einstellung von Parametern ohne Blickkontrolle getestet. Auch im zweiten Experiment findet ein Vergleich des Touchpads zum Dreh-Drücksteller statt. Die Teilnehmer dieses Experiments waren dieselben wie in Experiment 1. Der Versuchsablauf von Experiment 2 war mit der Bedingungsabfolge von Experiment 1 verschränkt. Deutliche Differenzen in der inhaltlichen Herangehensweise lassen es jedoch als sinnvoll erscheinen, die Experimente getrennt zu berichten. Während in Experiment 1 Zeiten und Blickdaten in Fahrt und Stand im Vordergrund stehen, liegt der Fokus bei der Blindbedienung in Experiment 2 ausschließlich auf den Eingabebefehlern.

4.1.2 Experiment 1: Evaluation mit und ohne Fahraufgabe

In einem Experiment mit Fahrsimulation wurde das multimodale Touchpad in verschiedenen Ausprägungen der Rückmeldung während der Fahrt und im Stillstand evaluiert. Relevante Leistungsparameter und eine subjektive Einschätzung der Belastung wurden erhoben. Die Ergebnisse wurden mit dem Datenmuster der Interaktion mit einem herkömmlichen Dreh-Drücksteller verglichen.

4.1.2.1 Methode

Versuchsumgebung und -material

Die Untersuchung wurde im Usability Labor des Fachzentrums für User Interface Design der Siemens AG durchgeführt. Das Labor ist mit einem Cockpit Mock-Up ausgestattet, das den Test unterschiedlicher Bediensysteme und technischer Komponenten für Fahrzeuge gestattet. Das Mock-up ist eine variable Sitzkiste, die unter Be-

rücksichtigung relevanter ergonomischer Grundsätze der Innenraumgestaltung konstruiert wurde. Pedalerie und Lenkrad sind an verschiedene virtuelle Fahrsimulationen anpassbar, die über einen Videobeamer an eine Leinwand ausgegeben werden. Die Vereinfachung des Fahrzeuginnenraums erleichterte die Durchführung des Experiments entscheidend, da die zu vergleichenden zentralen Eingabeelemente Dreh-Drücksteller und Touchpad schnell gewechselt werden konnten. Der Aufbau ist in Abbildung 9 dargestellt. Nachdem die räumlichen Verhältnisse einem realen Fahrzeug nachgebildet waren, kann ein Mindestmaß an Natürlichkeit und Übertragbarkeit der Fahrsituation gewährleistet werden.

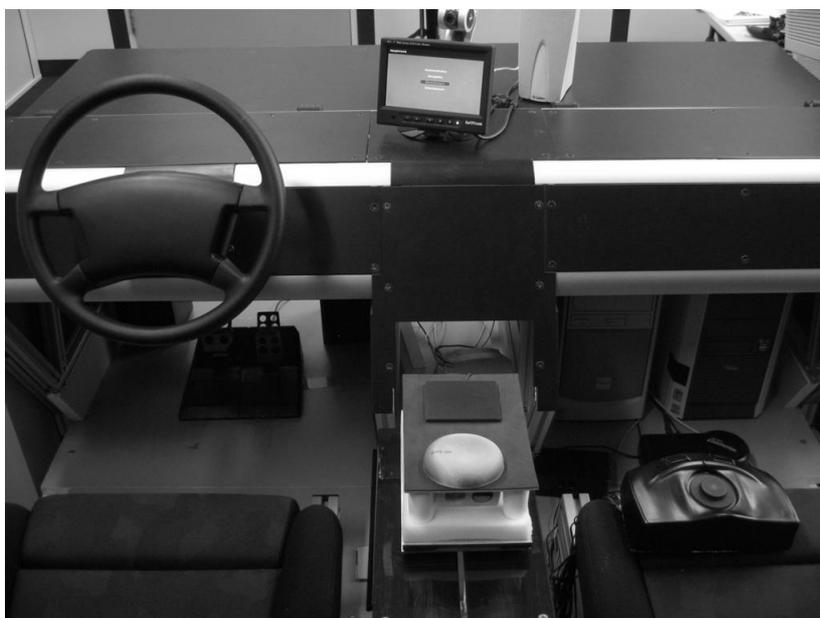


Abbildung 9: Cockpit Mock-up. Rechts neben dem Fahrerplatz ist das Touchpad zu erkennen, auf dem Beifahrersitz der Dreh-Drücksteller.

Für den Versuch wurde kein vollständiges Bordsystem simuliert. Vielmehr repräsentierten die zu erfüllenden Aufgaben typische Eingabemuster wie die Navigation durch ein Menü, die Veränderung von Parametern mit Hilfe von Reglern oder die Eingabe von Buchstaben in einem zweidimensionalen Buchstabenfeld. Das Programm zur Steuerung der Aufgaben lief auf einem Fujitsu Siemens PC mit 3.4 GHz CPU, der mit einer Matrox® G450 Grafikkarte und zwei identischen ESS Technologies ESS1969 Soundkarten ausgestattet war, und wurde vom Autor in Microsoft® Visual Basic® und mit Funktionen des Windows® API programmiert. Dieses Programm verarbeitete auch die Eingaben von Dreh-Drücksteller oder Touchpad und ermöglichte die multimodalen Rückmeldungen. Die Anzeigeeinheit bestand aus einem 7 Zoll TFT-LCD Breitbild-Farbmonitor mit 800×480 Bildpunkten der Firma CarTFT, wie er in ähnlicher Form häufig in Fahrzeugen verbaut ist. Das Touchpad war

über PS/2 an den Rechner angebunden, der Dreh-Drücksteller gab die Signale über ein spezielles Ein-/Ausgabegerät per USB ab.

Zur Aufnahme der Blickbewegungen wurde hinter dem Display eine digitale Videokamera (Philips SPC600NC) positioniert, die das Gesicht der Versuchsperson während der simulierten Fahrt filmte. Die Kamera war an einen weiteren Fujitsu Siemens PC mit 2.8 GHz CPU angeschlossen. Die Blickbewegungen wurden mit einer Auflösung von 640×480 Bildpunkten und 30 Bildern pro Sekunde aufgezeichnet. Als Fahraufgabe kam die Lane Change Task (Mattes, 2003) zum Einsatz. Die Software lief auf einem Fujitsu Siemens PC mit 1.53 GHz CPU, NVIDIA® GeForce™ 6600 GT Grafikkarte und ESS Technologies ESS1969 Soundkarte. Die Simulation wurde mit einem Epson EMP-760 Videobeamer im Sichtabstand von ca. 2.5 m mit einer Auflösung von 1024×768 Bildpunkten auf eine Leinwand vor der Sitzkiste projiziert. Das erzeugte Bild maß etwa 150×133 cm.

Aufgaben

Für den experimentellen Vergleich des Dreh-Drückstellers und des Touchpads mussten die Versuchsteilnehmer mehrere Aufgaben bearbeiten, die Schlüsselcharakteristika typischer Eingabeformen in ein Bordsystem repräsentieren. Um zumindest einen rudimentären Bezug zur Anwendungssituation herzustellen, wurde im oberen Bereich der Anzeigeeinheit eine Art Menüleiste angezeigt, die den Kontext der jeweiligen Aufgabe widerspiegelt (z.B. „Navigation – Ortseingabe“ bei der Buchstabeneingabe, siehe Abbildung 10).

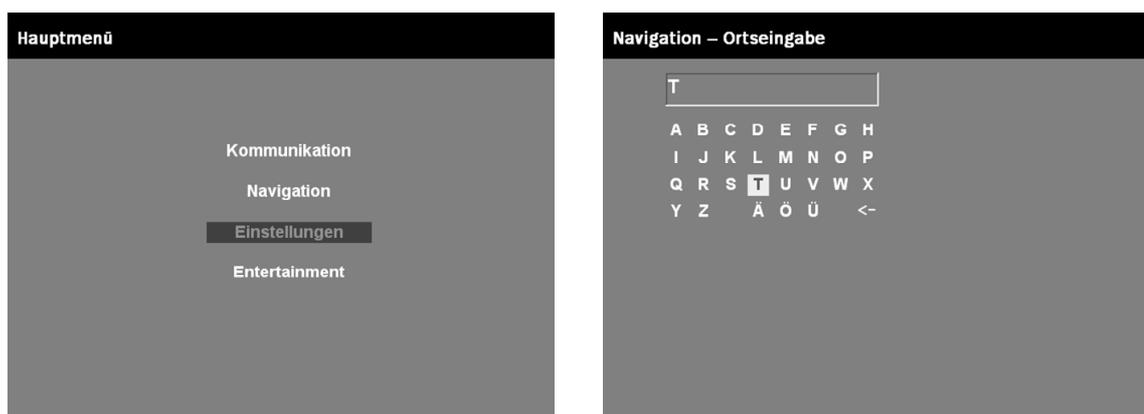


Abbildung 10: Exemplarische Darstellung zweier Aufgabentypen. Links: Listenaufgabe mit Menünavigation; rechts: Eingabe in 2D-Matrix.

Damit die Anforderungen der Aufgabenstellung während der Lane Change Task erfüllt werden können, wurden die Aufgabenanweisungen sehr kurz formuliert und konnten vom Versuchsleiter auf Knopfdruck vom Steuerungsprogramm des Experiments über Lautsprecher ausgegeben werden. Eine Übersicht über alle Anweisungen bietet Anhang A 1. Den Abschluss jeder Ansage bildete ein kurzer Signalton. Ab Ende

des Signaltons wurden die Bearbeitungszeit und andere relevante Parameter gemessen. Der Selektionscursor wurde erst danach freigegeben. Die Versuchsteilnehmer waren instruiert den Signalton abzuwarten. Die Auswahl des Zielitems bildete das Ende der Messzeit.

Bei der Interpretation der Leistung bei diesen Aufgaben muss berücksichtigt werden, dass die Komplexität in der Testsituation stark reduziert ist. Deshalb ist nur eine vergleichende Interpretation der Bedienvarianten sinnvoll, während beispielsweise eine Übertragung der absoluten Aufgabenbearbeitungszeiten auf die reale Anwendungssituation nicht ratsam ist. Vier Aufgabentypen wurden definiert:

- A1: Auswahl eines Items aus einer Liste. Diese Aufgabe war als Navigation durch ein Menü mit drei Ebenen angelegt. Die Menüleiste zeigte jeweils den Pfad an. Die Aufgabe bestand in der Auswahl eines CD-Titels. Dafür musste zunächst in einem Hauptmenü mit vier Items „Entertainment“ ausgewählt werden. Anschließend war der Wechsel zu „CD“ in einer Liste mit fünf Items vorzunehmen. Schließlich musste aus einer Liste mit zehn Items eine bestimmte Titelnummer eingestellt werden.
- A2: Einstellen eines Wertes über einen horizontalen Regler mit neutraler Mitte. Die Aufgabe bildet die Veränderung von Equalizer-Parametern nach (Beschriftung der Menüleiste: „Entertainment – Ton-Einstellungen – Höhen“). Die Skala erstreckte sich über den Bereich -3, -2, -1, neutral, +1, +2, +3. Die Probanden mussten den Selektionscursor auf den Zielwert bewegen und die Auswahl bestätigen.
- A3: Einstellen eines Wertes über einen vertikalen Regler ohne neutrale Mitte. Vorbild dieser Aufgabe war die Manipulation der Ventilationsstärke (Menüleiste: „Einstellungen – Lüftung“). Das untere Ende der Skala bildete der Wert „Aus“, das obere Ende „Maximum“. Dazwischen lagen die numerischen Werte 1 bis 7. Das Vorgehen war analog zu Aufgabe 2.
- A4: Auswahl eines Items aus einem zweidimensionalen Buchstabenfeld. Diese Aufgabe war der Buchstabeneingabe aus Navigationssystemen nachempfunden (Menüleiste: „Navigation – Ortseingabe“). Die Versuchsteilnehmer sollten einen vorgegebenen Buchstaben aus einer Eingabematrix bestehend aus den Buchstaben A-Z und einigen Sonderzeichen eingeben.

Teil der Aufgaben war stets, zunächst den Selektionscursor auf das Zielelement zu bewegen und die Auswahl zu bestätigen. Bei der Touchpad-basierten Eingabevariante erfolgte die Auswahl durch Verschieben des Fingers auf der berührungssensitiven Fläche in die intendierte Richtung der Cursorbewegung und das Bestätigen durch Antippen dieser Fläche an einer beliebigen Position. Drehen als Grundoperationen zur Auswahl beim Dreh-Drücksteller war folgendermaßen definiert: Eine Drehung im

Uhrzeigersinn führte bei horizontaler Anordnung zu einer Cursorbewegung nach rechts bzw. bei vertikal angeordneten Elementen nach unten. Im zweidimensionalen Buchstabenfeld bewegte sich der Cursor entlang einer Zeile und wechselte an deren Ende an den Anfang der nächsten Zeile.

Mit einer Ausnahme war bei den Aufgaben kein „Überdrehen“ des Cursors möglich. Wenn beispielsweise der Selektionscursor auf dem letzten Element der Menüliste war und der Nutzer mit Drehwähler oder Touchpad versuchte, den Cursor weiter in diese Richtung zu drehen, wurde der Cursor nicht auf das erste Element der Liste gesetzt. Dieses Verhalten ist auch bei vielen Systemen in der Praxis zu beobachten. Man erhofft sich dadurch, dass die Bewegungen des Cursors besser vorhersehbar werden und so die Blindbedienbarkeit unterstützt wird. Die Ausnahme bildete der Dreh-Drücksteller bei Aufgabe 4 mit dem zweidimensionalen Feld. Bei einem Drehen über das letzte Element der untersten Zeile hinaus wurde hier der Cursor auf das erste Element der ersten Zeile gesetzt (und umgekehrt). Diese Variante erschien notwendig, weil sie in realen Systemen typischerweise auch umgesetzt wird und der Vergleich zum Touchpad deshalb ansonsten aufgrund der großen Anzahl notwendiger Selektionsschritte willkürlich zum Nachteil des Drehwählers erschwert worden wäre.

Die Aufgabentypen 2-4 wurden je dreimal in jeder Schnittstellenvariante mit bzw. ohne simulierter Fahrt absolviert. Durch die Anlage als Menüaufgabe mit drei Ebenen musste Aufgabe 1 nur einmal bearbeitet werden. Somit ergaben sich für jede Bedienvariante zwölf Datenpunkte pro Versuchsperson pro Bediensituation (Stand, bzw. simulierte Fahrt). Die Zusammenstellung der Aufgaben sollte sicherstellen, dass der Vergleich der Schnittstellenvarianten auf den wesentlichen Eingabeformen für diese Art von Systemen basiert.

In jeder Aufgabe eines Typs musste ein anderer Zielwert eingestellt werden. Die Anzahl der notwendigen Selektionen von Start- zum Zielelement war je nach Aufgabe unterschiedlich. Durch die Experimentalssoftware wurde gesteuert, dass die Anzahl notwendiger Selektionsschritte über die drei Aufgaben hinweg in jeder Eingabevariante gleich war. Zusätzlich wurde darauf geachtet, dass bei jedem Aufgabentyp einmal das zweite oder das vorletzte Element Zielelement war. Diese Manipulation gestattet die Beobachtung, inwiefern die Teilnehmer bei der Zielbewegung die Ränder einer Reihe von Steuerelementen aktiv in die Eingabestrategie einbeziehen.

Touchpad mit multimodaler Rückmeldung

Für den Versuch wurde ein Touchpad konstruiert, das neben visueller Rückmeldung auf der Anzeigeeinheit eines Bordsystems verschiedene Arten von taktilem und auditivem Feedback ermöglicht. Dieser Prototyp wurde in das eingangs beschriebene Cockpit Mock-up integriert. Der Aufbau ist in Abbildung 11 dargestellt und umfasste folgende Komponenten: Die visuelle Ausgabe erfolgte über einen TFT-LCD Farbmo-

nitor, der in zentraler Position oberhalb der Mittelkonsole positioniert wurde. Ein Lautsprecher zur Ausgabe der auditiven Rückmeldungssignale war hinter dem Display aufgestellt. Der Prototyp des Touchpads war über dem Mitteltunnel der Sitzkiste so angebracht, dass die Testpersonen entspannt ihre Hand auf der vorgesehenen Handballenauflage ruhen lassen konnten. In baulicher Einheit mit dem Touchpad waren die Aktuatoren für vibrotaktilen Feedback angebracht: ein Körperschallwandler direkt unterhalb der berührungssensitiven Fläche und zwei Elektromotoren mit verschiedener Unwucht unter der Handballenauflage. Das Touchpad war das Modell TSM9778 der Firma Cirque Corporation. Dieses Modell erschien wegen der großen Abmessungen von 97×78 mm besonders geeignet, da es vergleichsweise weite Bewegungen zulässt.

Cockpit Mock-up (Eingabe: Touchpad)



Handballenauflage

Touchpad



Abbildung 11: links: Cockpit Mock-up; rechts oben: Touchpad mit Handballenauflage (Oberseite); rechts unten: Körperschallwandler und Elektromotoren mit Unwucht für taktilen Feedback (Unterseite).

Die Bewegung des Selektionscursors wurde visuell dadurch angezeigt, dass sich eine dunkle rechteckige Fläche auf den Steuerelementen, z.B. den Menüeinträgen, der Bedienoberfläche bewegte. Zusätzlich änderte sich die Schriftfarbe des aktiven Elements von weiß auf orange (vgl. Abbildung 10, links). Eine Bestätigung der Selektion durch Antippen der Fläche des Touchpads wurde durch einen Wechsel des Cursors von dunkelgrau auf gelb bzw. je nach Aufgabentyp durch einen Wechsel der Menüebene signalisiert (vgl. Abbildung 10, rechts).

Akustische Signale wurden über einen Labtec® LCS 1030 Aktivlautsprecher ausgegeben, der an eine der Soundkarten angeschlossen war. Die Lautstärke war so einge-

stellt, dass die Ausgaben gut hörbar waren, aber nicht als laut empfunden wurden. Eine Bewegung des Selektionscursors wurde durch ein metallisches Klick-Geräusch (Dauer 15 ms) quittiert. Bei Aufgabe 2 war die neutrale Mitte speziell auditiv markiert: Das Klick-Geräusch war von gleicher Länge, mit Hilfe des Programms Cool Edit 2000 (Syntrillium Software Corporation) wurde jedoch eine Tonhöhentransformation auf eine 30% tiefere Frequenz gegenüber dem Ursprungssignal durchgeführt. Dadurch entstand auditiv der Eindruck, dass der Cursor „einrastet“. Die Bestätigung eines Elements hatte ein dumpfes Tock-Geräusch (Dauer 60 ms) zur Folge. Eine besondere Stellung hatten das erste und letzte Element eines Menüs bzw. die Randelemente eines Reglers: Versuchten die Testpersonen, den Selektionscursor über das erste/letzte Element hinaus zu bewegen, informierte sie für die Dauer der Fingerbewegung ein Zisch-Geräusch, dass eine weitere Bewegung in die intendierte Richtung nicht möglich war. Dieses Geräusch hielt mindestens 100 ms an und stoppte danach sofort, falls die Bewegung über den Rand hinaus eingestellt worden war.

Taktile Rückmeldungen der Eingabebewegungen wurden technisch dadurch ermöglicht, dass auf der Unterseite der Trägerplatte unter dem Touchpad ein so genannter Bass Shaker als Körperschallwandler eingesetzt wurde und unter der Handballenauflage zwei Elektromotoren mit Unwucht Vibrationen erzeugen konnten. Im Gegensatz zu normalen Lautsprechern ist in einem Bass Shaker keine Membran vorhanden, die Übertragung der Schwingungen erfolgt durch eine feste Verbindung direkt auf die Trägerplatte. Der verwendete Körperschallwandler „Bass Pump III“ von Fischer Amps (Frequenzbereich 20-80 Hz, Dauereingangsleistung 50 W RMS) war zur Verbesserung der Übertragungsleistung an einen AURA Interactor Bass-Verstärker angeschlossen, der für einen Frequenzbereich von 18-1400 Hz spezialisiert ist. Der Verstärker wurde von der zweiten Soundkarte angesteuert. Die Vibrationsmotoren wurden einem Logitech® Rumblepad™ 2 entnommen, dessen Platine per USB mit dem PC verbunden war. Die Vibrationseffekte wurden mit Hilfe der Software Immersion Studio® (Version 4, Immersion Corp.) generiert und über ein Immersion ActiveX Control in das Experimentalprogramm eingebunden. Für die Gestaltung von Tactons konnte nur auf wenige konkrete Gestaltungsrichtlinien aus der Literatur zurückgegriffen werden. Zwar liegen allgemeine Hinweise von Brown et al. (2005) vor, die Auswahl der letztlich verwendeten vibrotaktilen Muster für die hier benötigten Systemereignisse musste dennoch in einem iterativen Vorgang erfolgen. Dafür wurden in mehreren informellen Vortests mit insgesamt acht Mitgliedern des Fachzentrums für User Interface Design verschiedene Tacton-Varianten erprobt. Die nachfolgend beschriebenen Ausgabevarianten sind das direkte Resultat der Vortests.

Selektionsbewegungen wurden stets durch den Körperschallwandler rückgemeldet. Sobald ein neues Element ausgewählt war, wurde ein 50 Hz Sinus Impuls von 20 ms Länge ausgegeben. Durch die leitende Montage unterhalb des Touchpads spürten die

Testpersonen jede Anwahl eines Elements direkt an ihrer Fingerspitze. Dadurch konnten sie auf taktile Ebene eine Repräsentation der Abstände zwischen einzelnen Elementen entwickeln. Zwischen allen Elementen (Listenitems der Menüs, Regler, Buchstaben in der Matrix) musste eine konstante Strecke mit dem Finger auf dem Touchpad zurückgelegt werden. Analog zur akustischen Ausgabe war auch bei der Gestaltung der taktilen Ausgabe das neutrale Element von Aufgabe 2 besonders hervorgehoben (50 Hz Sinus Grundfrequenz mit überlagerter dritter harmonischer Frequenz, verlängerte Vibration: 200 ms Dauer) und bei allen Aufgaben der Rand der Skalen speziell markiert (80 Hz Sinus, schwächerer Impuls: um 25% verringerte Amplitude). Das taktile Rand-Feedback für mindestens 100 ms ausgegeben und anschließend so lange aufrechterhalten, wie der Nutzer die Bewegung über den Rand hinaus fortsetzte. Die Gestaltung der Rückmeldung für eine Bestätigung der Selektion wurde dadurch erschwert, dass sich die Fingerspitze unmittelbar nach dem Antippen nicht auf dem Touchpad befindet. Nachdem sich jedoch in Vorversuchen gezeigt hat, dass durch die Platzierung des Eingabelements über dem Mitteltunnel die Handballenaufgabe für eine ergonomische Entlastung des Arms während einer Eingabe durchgängig genutzt wurde, konnte dieser Kontaktpunkt zur Ausgabe taktilen Feedbacks genutzt werden. Um sicherzustellen, dass die Vibration am Handballen oder Handgelenk tatsächlich wahrgenommen wurde, musste ein etwas kräftigerer und längerer Impuls gewählt werden. Hierfür wurden die Vibrationsmotoren eingesetzt, die sich für 150 ms bei maximaler Ausgabekraft drehten. Aufgrund einer gewissen unvermeidbaren Trägheit der Elektromotoren beim Anlaufen ging zusätzlich während der ersten 50 ms ein 30 Hz Sinussignal von dem Körperschallwandler aus, um keine spürbare Latenz entstehen zu lassen und um die Intensität in der Anfangsphase zu erhöhen. Nachdem der Vibrationsmotor nicht aktiv gebremst wurde, ebte die Ausgabe nicht abrupt ab. Am Handballen entstand dadurch die Empfindung eines kurzen kräftigen Vibrationsimpulses, der langsam nachließ. Zur Vermeidung von akustischen Artefakten und damit die Vibrationen sich nicht auf den Unterbau übertragen konnten, wurde der Prototyp auf eine 0.5 mm starke Stahlplatte platziert, die auf der Oberseite mit einem festen 2 cm dicken Schaumstoff bezogen war und durch vier Körperschallabsorber aus Gummi von der Standfläche entkoppelt wurde. Bei der Frequenzwahl aller vibrotaktilen Ausgaben wurde darauf geachtet, dass sie auditiv kaum wahrgenommen werden konnten. Die Ausgangsleistung des Bassverstärkers war so gewählt, dass die Rückmeldungen an der Fingerspitze gut zu fühlen, aber möglichst nicht zu hören waren.

Kombiniertes multimodales Feedback wurde durch eine redundante Darbietung der jeweiligen auditiven und haptischen Systemrückmeldungen realisiert. Zusätzlich zur visuellen Ausgabe wurden gleichzeitig die entsprechenden Töne und Vibrationen erzeugt. Durch die Verwendung von identischen Soundkarten konnte sichergestellt werden, dass keine bedeutsame Verzögerung zwischen taktilen und akustischen Aus-

gaben bestand. Auditives und taktiler Feedback waren vollständig äquivalent. Es gab kein Ereignis, das nur in der einen, nicht aber in der anderen Modalität rückgemeldet wurde. Die Programmierung der synchron-parallelen Ausgabe über zwei Soundkarten erfolgte nach einer Vorlage von Chan und Spence (2003).

Dreh-Drücksteller

Für den Vergleich zwischen dem Touchpad mit verschiedenen Feedbackvarianten und dem Dreh-Drücksteller wurde ein Drehimpulsgeber mit Taster der Firma ddm hopt+schuler GmbH zum Einstellen digitaler Werte verwendet (Typ 427). Dieser Dreh-Drücksteller verfügt über einen Gesamtdrehwinkel von 360° (endlos), 30 Einstellpositionen, einem Rastspiel von 12.5° und einer Rastung mit 1.5 Ncm (6 N für Druck). Abbildung 12 zeigt die Integration in den Testaufbau.

Cockpit Mock-up (Eingabe: Dreh-Drücksteller)



Abbildung 12: Cockpit Mock-up mit Dreh-Drücksteller zur Eingabe.

Visuell wurden Cursorbewegungen und Eingabebestätigungen analog zum Touchpad rückgemeldet. Bei diesem konventionellen Dreh-Drücksteller ist im Gegensatz zu sehr fortschrittlichen Varianten wie dem iDrive von BMW keine Ausgabe von frei programmierbaren haptischen Rückmeldungen möglich. Ebenso wurde auf zusätzliches auditives Feedback verzichtet. Der Drehwähler wurde jedoch so gewählt, dass mechanische Effekte (hörbares Klicken bei Drehen oder Drücken und deutlich spürbarer Widerstand) gut wahrnehmbar sind. Diese herkömmliche Variante des Dreh-Drückstellers wird in dieser Form in Automobilen weitaus häufiger verbaut als in der technisch komplexeren Variante. Dies lässt ihn als Baseline für einen Vergleich mit Touchpad als sinnvollsten Kandidaten erscheinen.

Lane Change Task

Ein wesentlicher Bestandteil des Experiments war die Durchführung der Aufgaben unter verschiedenen Belastungsbedingungen im Dual Task Paradigma mit den Komponenten Fahren und Bedienen. Um eine realistische und für die Fahrzeugführung repräsentative Form der visuell-manuellen Belastung zu induzieren, mussten alle Bedieneraufgaben zusätzlich während der gleichzeitigen Fahrt im Simulator absolviert

werden. Da die einzelnen Aufgaben sehr kurz waren, durch die Wiederholungen aber eine relativ hohe Anzahl in jeder Bedienvariante bearbeitet werden sollte, war eine Fahraufgabe mit weitgehend konstantem Schwierigkeitsniveau gefordert. Die Lane Change Task (Mattes, 2003) erfüllt diese Anforderung und ermöglicht so einen guten Vergleich zwischen den verschiedenen Schnittstellenvarianten.

Die Lane Change Task (LCT) ist ein vergleichsweise neuer Ansatz zur Messung der Fahrerablenkung. Sie ist in dem Projekt ADAM (Advanced Driver Attention Metrics) unter der Beteiligung der DaimlerChrysler AG und der BMW Group entstanden. Die Aufgabe besteht darin, auf einer geraden dreispurigen Strecke bei dem Auftauchen eines Hinweisschilds am Straßenrand auf eine vorgegebene Spur zu wechseln. Der Wechsel muss erfolgen, sobald das Schild ablesbar ist. Gleichzeitig müssen Eingaben in das zu testende System getätigt werden. Die Geschwindigkeit beträgt konstant 60 km/h. Die Fahrspur wird aufgezeichnet und mit einem normativen Modell verglichen. Die Abweichung von diesem Modell, ein einzelnes Leistungsmaß, gibt Aufschluss über die beobachtete Fahrgüte. Abbildung 13 veranschaulicht den Zusammenhang zwischen der Aufgabe (links) und der Auswertung durch das Analyseprogramm (rechts).

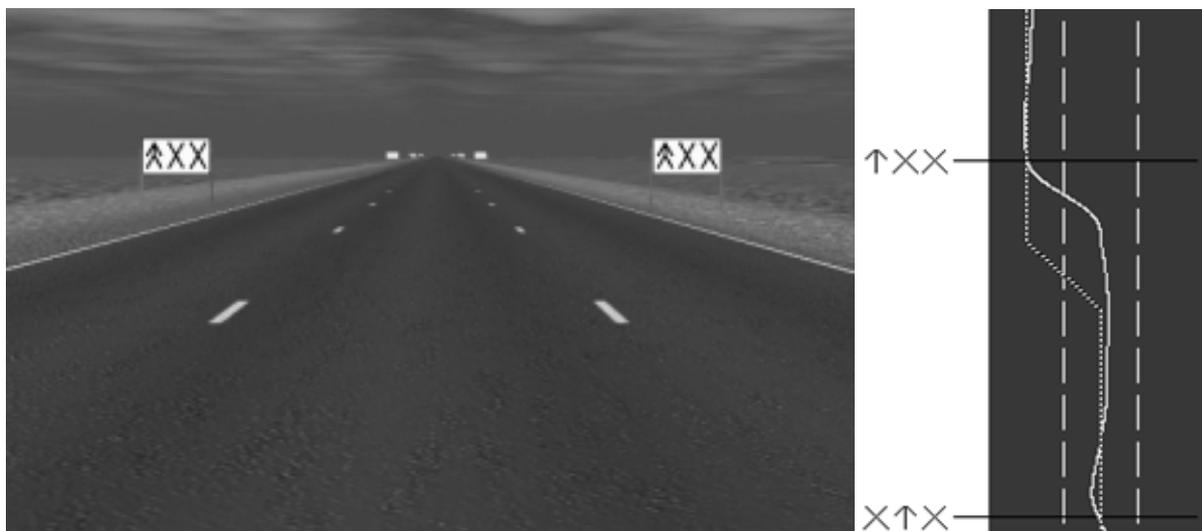


Abbildung 13: Linker Teil: Bildausschnitt der LCT-Simulation. Die Schilder neben der Fahrbahn weisen einen Wechsel auf die linke Spur an. Rechter Teil: Illustration der Bewertung des LCT Analyseprogramms (gepunktete Linie: normatives Modell; durchgezogene Linie: tatsächliche Fahrspur).

Um Lerneffekten vorzubeugen existieren zehn Strecken mit unterschiedlichen Reihenfolgen der Spurwechsel. Jede einzelne Strecke hat 18 Spurwechsel, wobei jede Form des Wechsels gleich oft vorkommt. Das Befahren einer Strecke dauert ca. drei Minuten. Die Komponenten der Aufgabe umfassen die Wahrnehmung des Hinweises, das Einleiten einer Reaktion, das Spurwechsel-Manöver und die Spurhaltung bis zum

nächsten Schild. Jede dieser Komponenten hat Einfluss auf die Güte der Fahrqualität und wird durch das Leistungsmaß erfasst. Es handelt sich damit um ein kombiniertes Maß aus der Güte der Querführung und der Reaktionsfähigkeit. Mattes (2003) hat in mehreren komparativen Experimenten die Validität der LCT evaluiert. Es hat sich gezeigt, dass sie sehr gut die Ablenkung durch verschiedene Zweitaufgaben widerspiegelt. Darüber hinaus korrelierte sie hoch mit Ergebnissen einer Untersuchung mit einem dynamischen Fahrsimulator, bei der die gleichen Zweitaufgaben auf Strecken mit unterschiedlicher Schwierigkeit getestet wurden ($r = .84$). Die Validierungsexperimente wurden mit einem Logitech® Formula GP® Lenkrad durchgeführt, das sich in Handhabung und Verhalten deutlich von dem Fahrzeuglenkrad der Sitzkiste unterscheidet. Deshalb wurde für dieses Experiment das Lenkrad der Sitzkiste temporär durch das Logitech-Modell ersetzt. Abbildung 14 zeigt den resultierenden Aufbau.



Abbildung 14: Versuchsteilnehmerin bei der gleichzeitigen Ausführung einer Bedienungsaufgabe und der LCT Fahrsimulation.

Typischerweise werden mit der LCT verschiedene Formen von Zweitaufgaben geprüft, z.B. unterschiedliche Systemvarianten. In diesem Experiment bleiben hingegen die Aufgaben bis auf ihre Reihenfolge gleich, variiert wird die Art der Ein- und Ausgabe. Für einen reibungslosen Ablauf müssen die verbalen Anweisungen an die Versuchsperson möglichst kurz sein, da die mittlere Entfernung zwischen zwei Hinweisschildern nur 150 m beträgt und so ca. alle 9 s ein Wechsel der Spur notwendig ist. Weiterhin erschien sinnvoll, die Aufgabenansagen möglichst konstant zu halten. Deshalb wurden die Aufgabenanweisungen vorab digital aufgezeichnet und angepasst (mittlere Dauer: 2.1 s). Eine neue Aufgabe wurde erst gestellt, wenn nach dem Abschluss einer Aufgabe der Spurwechsel soweit erfolgt war, dass wieder eine stabile Querführungsphase erreicht war. Bezüglich einer Priorisierung der Aufgaben in

der Mehraufgabensituation waren die Versuchsteilnehmer angewiesen, eine sichere Spurhaltung und rechtzeitige Spurwechsel als primäre Aufgabenanforderung zu sehen.

Versuchspersonen

Zwanzig Versuchspersonen (zwölf Frauen, acht Männer) im Alter zwischen 22 und 36 Jahren (Mittelwert: 28 Jahre) nahmen an dem Experiment teil. Sie erhielten eine angemessene Aufwandsentschädigung und hatten vorab keine Kenntnis über die Hypothesen des Experiments.

Alle besaßen einen gültigen Führerschein und gaben an, regelmäßig Auto zu fahren. Dies war eine notwendige Teilnahmevoraussetzung für die Studie, weil die simulierte Fahraufgabe im Experiment wie eine reale Fahraufgabe zu priorisieren war. Auf diese Weise wurden unübliche Allokationsstrategien beim Time-Sharing verhindert. Elf Teilnehmer legten im Jahr eine Strecke von weniger als 5000 km zurück. Fünf Probanden fuhren bis zu 10000 km. Die verbleibenden vier Teilnehmer gaben an, zwischen 10000 und 50000 km pro Jahr zu fahren. Vier Teilnehmer hatten bereits zuvor an einer Studie mit Fahrsimulation teilgenommen. Da die experimentellen Bedingungen innerhalb der Versuchspersonen variiert werden, sollte sich diese Vorerfahrung nicht verzerrend auswirken. Zu ihrer Computererfahrung befragt, berichteten alle Probanden regelmäßige Nutzung. Sechzehn Teilnehmer gaben ferner häufige Laptop-Nutzung an. Damit ist einem Großteil die Nutzung eines Touchpads gut bekannt.

Versuchsdesign

Ein 2×5 faktorielles Within-Subjects Design kam zum Einsatz. Alle Versuchspersonen absolvierten die Aufgaben mit unterschiedlicher Belastung in zwei Ausprägungen, im Stillstand ohne Fahraufgabe und mit simulierter Fahrt. Neben der unabhängigen Variable Bediensituation wurde auch die Schnittstelle zwischen Mensch und System variiert. Die Probanden nutzten fünf verschiedene Ausprägungen der Mensch-Maschine-Schnittstelle, die durch das Eingabegerät und die Art der Rückmeldung definiert waren. Visuelles Feedback als Ausgabe der Cursoraktionen auf der Anzeigeeinheit wurde immer gegeben. Der konventionelle Dreh-Drücksteller wurde unverändert eingesetzt. Zusammen mit den Touchpad-Varianten enthielt der Faktor Schnittstelle die folgenden Stufen:

- TPsolo: Touchpad, visuelle Ausgabe ohne zusätzliche Rückmeldung
- TPaud: Touchpad, visuelle Ausgabe + auditiv wahrnehmbare Rückmeldung
- TPtakt : Touchpad, visuelle Ausgabe + taktil wahrnehmbare Rückmeldung
- TPaud+takt: Touchpad, visuelle Ausgabe + auditive + taktile Rückmeldung
- DD: Dreh-Drücksteller

Damit lassen sich auf der Dimension der Reichhaltigkeit der Ausgabemodalitäten bei dem Touchpad unterscheiden: Unimodales Feedback bei TPsolo (nur visuelle

Ausgabe); bimodales Feedback als einfachste Form multimodaler Ausgabe bei TPaud und TPtakt (visuell und auditiv bzw. taktil); trimodales Feedback als reichhaltige Form multimodaler Ausgabe (TPaud+takt mit einer vollständigen Kombination der verfügbaren Modalitäten). Die Reihenfolge aller Schnittstellenvarianten war durch ein lateinisches Quadrat ausbalanciert. Jede Bedienvariante kam gleich oft an jeder Position vor, dabei wurde jeweils nach fünf Versuchspersonen eine neue Zufallsfolge gebildet. In jeder Bedienvariante wurden zunächst alle vier Aufgabentypen ohne Fahraufgabe bearbeitet, im Anschluss daran mit zusätzlicher simulierter Fahrt in der LCT. Die Abfolge der Aufgabentypen wurde für jede Testperson bei jeder Schnittstellenvariante neu randomisiert. Zusätzlich wurde eine Baselinefahrt ohne Bedienaufgabe eingestreut, um das individuelle Spurhaltevermögen jedes Teilnehmers zu messen: Je fünf Versuchspersonen absolvierten die Kontrollfahrt vor der ersten, nach der zweiten, nach der dritten oder nach der fünften Bedienvariante. Die LCT-Strecke wurde zufällig unter den zehn vorhandenen Strecken ausgewählt, Wiederholungen waren ausgeschlossen.

Die abhängigen Variablen basierten sowohl auf subjektiven als auch auf objektiven Maßen und werden nachfolgend dargelegt.

- Aufgabenbearbeitungszeit: Die Zeit wurde gemessen zwischen dem akustischen Startsignal und der erfolgreichen Beendigung der Aufgabe (d.h. der Auswahl und Bestätigung des korrekten Elements).
- Fehlerrate: Befand sich der Selektionscursor zum Zeitpunkt der Bestätigung über dem falschen Element, wurde die Eingabe als Fehler gezählt. Fehlerhafte Trials gingen nicht in die Auswertung der anderen Maße ein.
- Anzahl Randberührungen: Wurde bei der Eingabe versucht, den Cursor über das erste bzw. letzte Element hinaus zu bewegen, wurde diese Aktion zur Auswertung der Eingabestrategie aufgezeichnet.
- Blickverhalten: Primär relevant im Rahmen dieser Studie ist, inwiefern die Nutzer ihre Eingaben in das System in Abhängigkeit von der Ausprägung der Schnittstellenvariante visuell überprüfen. Deshalb wurden nur Blicke auf das Display über der Mittelkonsole während der simulierten Fahrt analysiert. Die Dauer pro Blick auf die Anzeige, Blickhäufigkeit und die Gesamtblickdauer (Produkt aus Blickdauer und Häufigkeit) gingen in die Auswertung ein. Daraus resultiert ein valides und akzeptiertes Maß der visuellen Anforderung, die mit einer Sekundäraufgabe verbunden ist, sowie eine Abschätzung des Interferenzpotentials (Haigney & Westerman, 2001).
- Fahrqualität: Die Qualität der Ausführung der LCT soll darüber Auskunft geben, wie gut die Probanden die Primäraufgabe schützen konnten oder ob die

verschiedenen Schnittstellenvarianten Einfluss auf die Fahrqualität hatten. Als Maß diente die mittlere Abweichung von der Normspur.

- Subjektive Workload: Zur Messung der empfundenen Workload wurde vor dem Wechsel der Schnittstellenvariante nach der Fahraufgabe von den Versuchspersonen der RTLX ausgefüllt, eine Variante des NASA Task Load Index (Byers, Bittner & Hill, 1989; Hart & Staveland, 1988).

Die Leistungsmaße im Zusammenhang mit der Sekundäraufgabe (Aufgabenbearbeitungszeit, Fehlerrate und Anzahl der Randberührungen) wurden durch das Experimentalprogramm festgehalten. Das Blickverhalten wurde digital aufgezeichnet und von Hand mit einem speziellen Auswertungsprogramm unter verlangsamter Darstellung ausgewertet. Die Auswertung der Lane Change Task erfolgte durch das dazugehörige Analyseprogramm LCTA, Version 1.99.

Prozedur

Nach der Begrüßung gaben die Teilnehmer in einem Fragebogen zunächst einige Angaben zu ihrer Person ab (Alter und Geschlecht, Fahrpraxis, Vorerfahrung mit Simulatorstudien). Wie in Untersuchungen dieser Art üblich, wurden die Versuchsteilnehmer darauf aufmerksam gemacht, dass der Inhalt des Tests die Evaluation einer Benutzerschnittstelle ist und nicht auf die Erhebung persönlicher Fähigkeiten und Fertigkeiten abzielt. Die Versuchsperson nahm für die Erläuterung der Aufgaben auf dem Fahrersitz der Sitzkiste Platz, wo sie eine funktionsunabhängige Erklärung der Eingabeprinzipien von Touchpad und Dreh-Drücksteller erhielt. Sinn dieser Ausführung war, das Interaktionsmodell der jeweiligen Schnittstelle zu vermitteln. Im Anschluss daran wurden die vier Aufgabentypen erklärt. Zur Eingewöhnung wurden je zwei Aufgaben eines Typs in einem Probedurchgang absolviert. Durch diesen Probedurchgang lernten die Versuchsteilnehmer sowohl die Aufgabentypen, als auch die Art der Rückmeldung und den Ablauf der Aufgabenstellung mit der voraufgezeichneten Audio-Ansage und dem anschließenden Signalton kennen. Dabei kam diejenige Schnittstellenvariante zum Einsatz, mit der laut Versuchsplan das erste Set von Aufgaben zu bearbeiten war. Aufgrund der Balancierung der Abfolge der Schnittstellenvarianten kam jede dieser Varianten gleich oft bei der Einführung und dem Probedurchgang zu Beginn des Experiments zum Einsatz.

Im Anschluss daran machten sich die Versuchspersonen in einer Eingewöhnungsfahrt mit dem Simulator vertraut. Den ersten Teil der Strecke sollten sie nutzen, um sich in freier Fahrt an die Lenkung zu gewöhnen. Das Gaspedal musste während aller Fahrten immer bis zum Maximum gedrückt sein. Die Geschwindigkeit war auf 60 km/h begrenzt und wurde somit konstant gehalten. Hatten sich die Versuchsteilnehmer auf das Rückstellmoment des Lenkrads und den Zusammenhang zwischen Lenkwinkel und Verhalten der Simulation eingestellt, wurde ihnen die Bedeutung der

Spurwechselschilder erläutert. Nach vier bis fünf erfolgreichen Spurwechseln wurden sie ermutigt, gleichzeitig die Sekundäraufgabe zu bedienen. Die Steuerelemente der letzten Beispielaufgabe waren auf dem Bildschirm sichtbar und die Probanden sollten einige beliebige Werte einstellen, um sich an die parallele Anforderung zu gewöhnen. Der Versuchsleiter wies an dieser Stelle darauf hin, dass die Aufgabe der simulierten Fahrt mit einem zügigen Wechsel der Fahrspur und guter Stabilisierung der Querführung wie in der realen Fahrsituation gegenüber der Bedienaufgabe Vorrang hat.

Nachdem alle Komponenten des Experiments besprochen und offene Fragen seitens der Teilnehmer geklärt waren, absolvierten sie für jede Schnittstelle zunächst alle Bedienaufgaben im Stand ohne Fahraufgabe. Daran schloss sich direkt die Simulationsfahrt mit einer Bearbeitung der Aufgaben während der LCT an. Als letzter Schritt des Tests jeder Schnittstellenvariante wurde der RTLX ausgefüllt. Bevor das Experiment mit der nächsten Schnittstelle fortgesetzt wurde, mussten erneut ohne Fahraufgabe zwei weitere Aufgabentypen bearbeitet werden. Diese Aufgaben zählten inhaltlich zu Experiment 2, das in Experiment 1 eingewoben war. Die Gründe für dieses Ineingreifen der Experimente sowie die genaue Natur dieser Zwischenaufgaben werden in der Besprechung von Experiment 2 ausführlich dargelegt.

Als nächstes wurde die Schnittstellenvariante gewechselt, d.h. es fand entweder eine Änderung der Rückmeldungsform beim Touchpad oder ein Wechsel von Touchpad und Dreh-Drücksteller statt. Damit auch hier die Versuchspersonen sich an die neue Ein- und Ausgabe gewöhnen konnten, bearbeiteten sie exemplarisch zwei Aufgabentypen. Eine der Aufgaben bestand dabei stets aus der Eingabe über den horizontalen Regler mit neutraler Mitte, da dabei alle taktilen und auditiven Feedbackformen zum Einsatz kamen. Der andere Aufgabentyp war das zweidimensionale Buchstabenfeld, weil hier die Versuchsteilnehmer beim Dreh-Drücksteller auf die Möglichkeit des Überdrehens aufmerksam gemacht werden mussten. Der weitere Ablauf nach dem verkürzten Probedurchgang war analog zum vorher beschriebenen Vorgehen: Zunächst wurden alle Aufgaben im Stillstand, dann parallel zur LCT bearbeitet, woran sich der RTLX und die eingewobene Aufgabe aus dem zweiten Experiment anschloss.

Nach Beendigung der experimentellen Aufgaben mit allen Schnittstellen erhielten die Versuchsteilnehmer zum Abschluss Gelegenheit, ihre Ansichten zu den Systemvarianten zu äußern und ihre Präferenzen diesbezüglich darzulegen. Das Experiment dauerte insgesamt etwa 70 Minuten. Eine zusammenfassende Übersicht über den Versuchsplan bietet Abbildung 15.

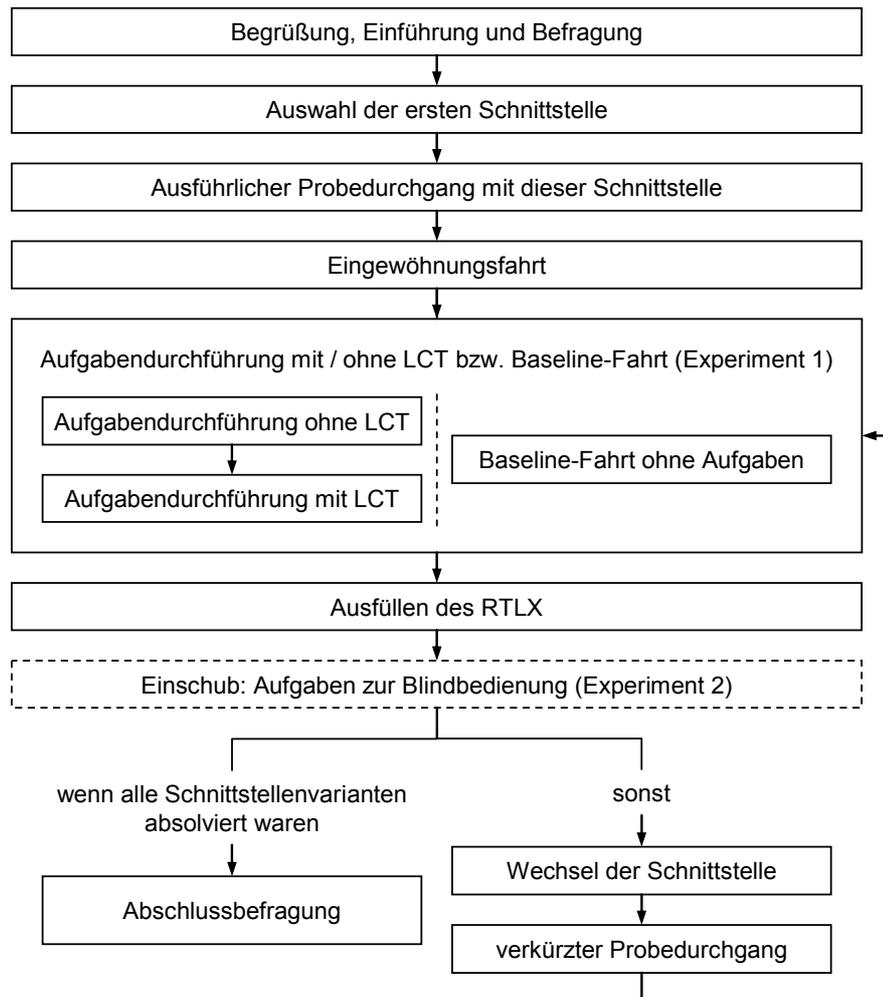


Abbildung 15: Grafische Darstellung des Versuchsplans.

4.1.2.2 Hypothesen

Ein Touchpad sollte aufgrund der natürlicheren Abbildung der Aktionen und des zusätzlichen Freiheitsgrades zur Steuerung des Selektionscursors für die Nutzer intuitiver und schneller zu bedienen sein als ein Dreh-Drücksteller. Zu einer Steigerung der Effizienz trägt weiter bei, dass bei dem Touchpad kein Umgreifen zum Bestätigen oder Auslösen eines ausgewählten Steuerelements notwendig ist. Im Gegensatz zu einem Touchpad gibt ein Dreh-Drücksteller jedoch unmittelbare Rückmeldungen: Das spür- und hörbare Überwinden einer Rasterstufe kann der Bewegung des Selektionscursors um eine Position klar zugeordnet werden. Ebenso weist das Überwinden des Druckpunkts beim Drücken deutlich auf eine erfolgreiche Auslösung des angewählten Elements hin. Bei einem herkömmlichen Touchpad dagegen muss visuell überprüft werden, inwiefern Auswahl und Bestätigung erfolgreich waren. Zum einen zeigt das Touchpad normalerweise weder akustisch noch haptisch eine Veränderung durch Fingerbewegungen an, zum anderen ist technisch bedingt insbesondere bei kapazitiven Touchpads nicht jedes kurze Antippen zur Bestätigung erfolgreich. Die resultierende

visuelle Last reduziert die möglichen Vorteile der Touchpad-basierten Eingabe. Vor dem Hintergrund dieser Konstellation findet die Hypothesenbildung statt.

Die geschilderten Unterschiede sind auf der Ebene der Systemoptimierung angesiedelt. Touchpad und Dreh-Drücksteller sind grundsätzlich geeignete und bewährte Bedienformen zur Benutzung menügesteuerter Systeme. Daraus lässt sich die Hypothese ableiten, dass die Vor- und Nachteile der jeweiligen Varianten kaum zu einem Unterschied führen werden, wenn ihnen die volle Aufmerksamkeit bei der Bedienung gewidmet werden kann. Wenn jedoch, wie bei der (simulierten) Fahrzeugführung eine Hauptaufgabe mit dominierender visueller Komponente existiert, lässt sich basierend auf dem multiplen Ressourcenmodell (Wickens, 1984, 2002) die Vorhersage ableiten, dass vornehmlich eine Nebenaufgabe mit hoher visueller Anforderung durch das Time-Sharing nachteilig betroffen sein wird. In der Fahrsituation sollte sich ein Vorteil für den herkömmlichen Dreh-Drücksteller gegenüber dem einfachen Touchpad (d.h. ohne akustischem oder taktilem Feedback) nachweisen lassen.

Von besonderem Interesse sind die Auswirkungen multimodaler Ausgabe. Es stellt sich hier die Frage, wie sich die geschilderte Situation durch die Einführung von auditivem und/oder taktilem Feedback bei dem Touchpad verändert. Es ist bekannt, dass ballistische Zielaufgaben durch zusätzliches Feedback schneller ausführbar sind (Cockburn & Brewster, 2005; Göbel et al., 1995; Vitense et al., 2003). Darüber hinaus sollte das zusätzliche Feedback die visuelle Last bei der Interaktion mit dem Touchpad verringern, so dass dessen Vorteile bei der Bedienbarkeit besser zum Tragen kommen. Entsprechend wird erwartet, dass die Erweiterung des Touchpads um zusätzliche Ausgabeformen zu einer besseren Leistung führt, operationalisiert durch die Aufgabenbearbeitungszeit und die Fehlerrate. Wenn tatsächlich die Vorteile der natürlichen Abbildung funktional herausgearbeitet werden, sollte sich auch eine bessere Aufgabenleistung als bei dem Dreh-Drücksteller ergeben. Durch die Entlastung des visuellen Kanals sollte beim Touchpad mit zusätzlichen Rückmeldevarianten ferner eine verringerte Blickzuwendung zum Display während der Fahrt nachweisbar sein.

Ergonomische Untersuchungen legen nahe, dass die Workload bei redundanter Rückmeldung über mehrere Sinneskanäle geringer ist (Brewster, 1997; Liu, 2001; Vitense et al., 2003). Damit ergibt sich die Hypothese, dass im Vergleich zu dem Touchpad mit ausschließlich visueller Rückmeldung die erweiterten Varianten (bi- oder multimodal) im TLX Rating zu geringerer subjektiver Workload führen sollten.

Wenn all diese Faktoren zutreffen, ist zu erwarten, dass auch die Fahraufgabe an sich von der multimodalen Ausgabe des Touchpad profitiert. Wenn die Aufgabenbearbeitungszeit kürzer ist, ist auch die Dauer der Abwendung kürzer. Der resultierende geringere Druck beim Time-Sharing und die kürzere visuelle Abwendung von der Fahrbahn sollten zu einer Verbesserung der Querführung und Reaktionsfähigkeit in der LCT führen.

4.1.2.3 *Ergebnisse*

Alle abhängigen Variablen wurden für jede Bedienvariante einzeln erfasst. Die Maße werden bis auf die Bearbeitungszeiten aufgabenübergreifend berichtet. Die Daten für die Bearbeitungszeiten pro Aufgabe werden zusätzlich auf der Ebene einzelner Aufgabentypen analysiert, um insbesondere die Auswirkungen der Art der Eingabe bei der zweidimensionalen Buchstabenmatrix beurteilen zu können.

Die gewonnenen Daten wurden mit Hilfe von Varianzanalysen auf statistisch signifikante Unterschiede untersucht.⁶ Ferner wurden stets drei geplante Vergleiche zwischen einer Teilmenge der Bedienvarianten mit t-Tests für abhängige Stichproben berechnet. Bei signifikanten Unterschieden wird bei Varianzanalysen als Maß der Effektgröße das partielle Eta-Quadrat η_p^2 berichtet, bei t-Tests werden die standardisierten Mittelwertsdifferenzen mit der Effektgröße d angegeben. In jeder Analyse wird die Statistik p_{rep} ausgewiesen, eine relativ neue Methode zur Schätzung der Wahrscheinlichkeit einer Replikation des Effekts (Killeen, 2005a, 2005b). Höhere p_{rep} -Werte ($>.8$) weisen auf eine stärkere Robustheit des gezeigten Effekts hin, mithin auf eine hohe Replikationswahrscheinlichkeit. Dabei hängt die Güte der Vorhersage über eine konkret zu erwartende Replikation eines Effekts selbstverständlich nach wie vor von Faktoren der Versuchsplanung wie Verzerrungen durch die Stichprobe oder Konfundierungen ab. Nachdem p_{rep} noch nicht sehr weit verbreitet ist, werden zusätzlich die traditionellen p -Werte angegeben. Deskriptiv werden im Text Mittelwerte (M) und Standardfehler (SE) berichtet. Um die Ergebnismuster besser erfassbar zu machen, werden für wichtige Befunde stets grafische Darstellungen (Mittelwerte mit Fehlerbalken ± 1 Standardfehler) beigelegt.

In die Gegenüberstellung durch geplante Vergleiche gehen das Touchpad ohne zusätzliches Feedback, das Touchpad mit zusätzlicher auditiver und taktile Rückmeldung sowie der Dreh-Drücksteller ein. Nachdem davon ausgegangen wird, dass sich nur in der Fahrsituation deutliche Unterschiede zeigen werden, werden lediglich diese Daten der Detailanalyse zugeführt. Die Vergleiche zwischen Dreh-Drücksteller und den Touchpad-Varianten sollen Aufschluss darüber geben, inwiefern sich die neu vorgeschlagene Bedienform gegenüber der etablierten Methode bewährt. Das Touchpad mit kombiniert-multimodalem Feedback (visuell, auditiv, taktil) wurde unter den Varianten mit zusätzlichem Feedback gewählt, weil diese Ausprägung maximal von redundanter Signalausgabe Gebrauch macht. Forschung zu uni-, bi- und trimodaler Ausgabe lässt erwarten, dass sich hier die stärkste Förderung zeigen wird (z.B. Diedrich & Colonius, 2004). Wie in den Hypothesen ausgeführt wird erwartet, dass das Touchpad ohne zusätzliches Feedback nicht zu besseren Resultaten führen wird als der

⁶ Bei der Berechnung statistischer Analysen wurde auf die Programme SPSS® für Windows® (Version 11, SPSS Inc.) und |STAT (Version 5, Perlman & Horan, 1986) zurückgegriffen.

Dreh-Drücksteller, das Touchpad mit zusätzlicher nichtvisueller Ausgabe jedoch eine Verbesserung darstellen sollte.

Für dieses und alle folgenden Experimente wurde ein Alpha-Niveau für Signifikanz von .05 gewählt. In der statistischen Literatur wird kontrovers diskutiert, inwiefern das Signifikanzniveau bei multiplen Tests selbst im Fall von hypothesengeleiteten a priori geplanten Vergleichen angepasst werden sollte. An dieser Stelle wird der Argumentation von Keppel (1991, S.187) gefolgt, der empfiehlt, theoretisch begründeten geplanten Vergleichen einen besonderen Status einzuräumen und keine Anpassung vorzunehmen. Mit einem Verweis auf Arbeiten von Davis und Gaito (1984, zitiert nach Keppel, 1991) argumentiert er, dass eine grundsätzlich zu große Besorgnis um Typ I Fehler in Untersuchungen den Fortschritt in diesem Bereich aufhalten könne. Wenn eine begrenzte Anzahl relevanter Vergleiche durchgeführt würde, für die das Experiment eigens konstruiert wurde, bestehe durch die Reduktion der Power bei der Anpassung des Signifikanzniveaus das Risiko, viel versprechende Entdeckungen zu unterbinden. Die geplanten Vergleiche in diesem Experiment werden zur weiteren Absicherung nur gerechnet, wenn die Varianzanalyse zuvor einen statistisch bedeutsamen Einfluss der unabhängigen Variablen nachweist. Sie sollen hauptsächlich dazu dienen, die Effekte der Touchpad-Varianten und des Dreh-Drückstellers im Vergleich zu illustrieren. Die Analyse der Effekte von Multimodalität und deren inhaltliche Beurteilung müssen hingegen durch eine übergeordnete Betrachtung von konstanten Ergebnismustern in den verschiedenen abhängigen Variablen erfolgen. In diesem Sinne sollte der Frage, inwiefern einzelne Vergleiche auf der Ebene der t-Tests auch unter einer konservativen Anpassung des Alpha-Niveaus erreicht würden, keine zu große Bedeutung bei der Interpretation beigemessen werden.

Aufgabenbearbeitungszeit

In jeder Bedienvariante wurden dieselben Aufgaben bearbeitet. Die benötigte Zeit zur Fertigstellung einer Aufgabe ist damit ausschließlich davon abhängig, wie gut der Dreh-Drücksteller bzw. das Touchpad mit den unterschiedlichen Rückmeldungsformen eine zügige Bedienung unterstützt. Für einen Gesamteindruck wird zunächst die Bearbeitungszeit sowohl in der simulierten Fahrt als auch im Stillstand über alle Aufgaben und Teilnehmer gemittelt berichtet. Nachdem die Aufgaben nicht besonders schwierig waren, wird – verglichen mit den anderen erhobenen Maßen – am ehesten auf der Ebene der Bearbeitungszeiten erwartet, dass sich zwischen den verschiedenen Aufgaben in Abhängigkeit von der Bedienvariante Unterschiede zeigen. Für eine detaillierte Analyse wird zusätzlich die durchschnittliche Dauer der Bearbeitung jedes der vier Aufgabentypen separat betrachtet.

Damit die Vergleiche auf der Ebene der einzelnen Aufgabentypen besser mit dem Gesamtmaß vergleichbar sind, wird bei der Bearbeitungszeit über die Aufgabentypen

gemittelt, nicht summiert. Für jeden Aufgabentyp wurde der Mittelwert über die Wiederholungen der jeweiligen Aufgabe berechnet. Über alle Aufgabentypen hinweg zeigt sich, dass die Art der Signalausgabe bei dem Touchpad bzw. der Einsatz eines Dreh-Drückstellers in der Single Task Situation nicht zu merklichen Unterschieden bei der Bearbeitungszeit führt. Wurden Nutzeraktionen beim Touchpad nur visuell rückgemeldet (TPsolo), benötigten die Teilnehmer 2.07 s ($SE = 0.10$) für die Bearbeitung ohne gleichzeitige Fahraufgabe. Bei zusätzlicher auditiv wahrnehmbarer Ausgabe (TPaud) betrug die Zeit 1.97 s ($SE = 0.10$), bei zusätzlicher taktiler Ausgabe (TPtakt) 2.14 s ($SE = 0.10$). Wenn sowohl auditive als auch taktile Rückmeldung zusätzlich zur visuellen ausgegeben wurde (TPaud+takt), ergab sich eine Bearbeitungszeit von 2.03 s ($SE = 0.09$). Die Eingabe über den Dreh-Drücksteller im Stand benötigte im Schnitt 2.05 s ($SE = 0.07$). Ein ganz anderes Bild ergibt sich, wenn gleichzeitig die Hauptaufgabe der Fahrzeugführung zu erfüllen ist. Wurde durch die LCT eine Dual Task Situation herbeigeführt und damit die Notwendigkeit zum Time-Sharing induziert, entstehen klar sichtbare Differenzen zwischen den Schnittstellenausprägungen. Mit 3.41 s ($SE = 0.30$) ist die Bearbeitungszeit bei TPsolo deutlich länger als bei den Touchpad-Varianten mit zusätzlichem Feedback (TPaud: $M = 2.82$ s, $SE = 0.15$; TPtakt: $M = 2.86$ s, $SE = 0.26$; TPaud+takt: $M = 2.96$ s, $SE = 0.23$). Der Dreh-Drücksteller führt während der Fahrsituation zu ähnlich langen Bearbeitungszeiten wie das Touchpad ohne Feedback (DD: $M = 3.39$ s, $SE = 0.27$). Abbildung 16 illustriert diese Zusammenhänge zwischen den Zeiten im Stillstand und bei simulierter Fahrt.

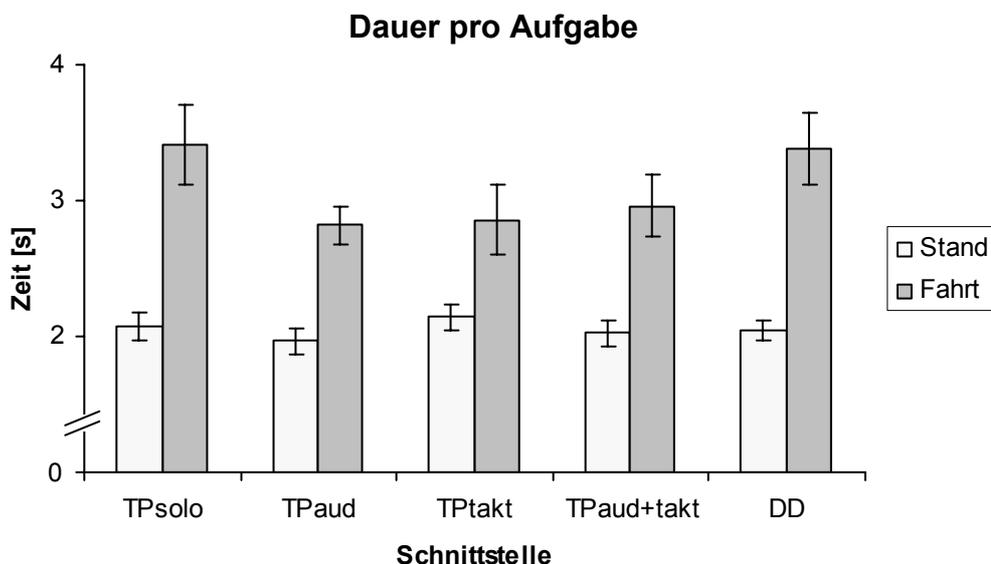


Abbildung 16: Mittlere Bearbeitungszeiten pro Aufgabe über alle Aufgaben für die Schnittstellenvarianten in den Bediensituationen Stand bzw. Fahrt.

Eine zweifaktorielle Varianzanalyse mit Messwiederholung bestätigt den unterschiedlichen Einfluss der Schnittstellenvarianten in Abhängigkeit von der Bediensituation in Fahrt oder Stand durch eine signifikante Interaktion dieser Faktoren

(Schnittstelle \times Bediensituation), $F(4,76) = 2.84$, $p = .03$, $p_{\text{rep}} = .91$, $\eta_p^2 = .13$. Generell sind darüber hinaus die Bearbeitungszeiten während der Fahrt signifikant langsamer als im Stand ($M = 3.09$ s, $SE = 0.11$ vs. $M = 2.05$ s, $SE = 0.04$), $F(1,19) = 34.86$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .65$. Der nachweisbare Unterschied zwischen den Schnittstellenvarianten ($F(4,76) = 2.45$, $p = .05$, $p_{\text{rep}} = .87$, $\eta_p^2 = .11$) resultiert ausschließlich aus den Veränderungen während der simulierten Fahrt.

Für die weitere Untersuchung der Auswirkungen des Unterschieds zwischen Dreh-Drücksteller und Touchpad mit nur visuellem bzw. multimodalem Feedback wurden geplante Vergleiche zwischen DD, TPsolo und TPaud+takt gerechnet (Tabelle 2).

Tabelle 2: Geplante Vergleiche zur Bearbeitungszeit pro Aufgabe, gemittelt über alle Aufgaben.

Vergleich	$t(19)$	p	p_{rep}	d
TPsolo – TPaud+takt	2.23	.038*	.89	0.50
TPsolo – DD	0.11	.917	.16	0.02
DD – TPaud+takt	2.17	.043*	.89	0.49

Nur die Daten während der Fahrsimulation gingen in die Berechnung ein. Signifikante Unterschiede ergaben sich zwischen TPsolo und TPaud+takt sowie zwischen DD und TPaud+takt. Bei TPsolo und DD konnte hingegen kein Unterschied nachgewiesen werden.

Die Aufgabentypen waren derart konstruiert, dass sie relevante Charakteristika typischer Bedienvorgänge repräsentieren. Mit der bisherigen Analyse konnte gezeigt werden, dass die Schnittstellenvarianten in Abhängigkeit von der Bediensituation (Single vs. Dual Task) generell unterschiedlichen Einfluss auf die Bearbeitungszeiten haben. Ein zusätzlicher Analyseschritt soll Auskunft darüber geben, ob dieses Ergebnismuster bei allen Aufgabentypen gleichermaßen zu finden ist. Anhang A 2 gibt eine tabellarische Übersicht der Bearbeitungszeiten aller Aufgabentypen für jede Schnittstelle in Abhängigkeit von der Bediensituation.

Die varianzanalytische Betrachtung zeigt, dass die in den experimentellen Hypothesen erwartete Interaktion Bediensituation \times Schnittstelle nicht bei allen Aufgabentypen statistische Bedeutsamkeit erlangt. In der Menüaufgabe A1 liegen auch bei simulierter Fahrt die mittleren Bearbeitungszeiten der einzelnen Schnittstellenvarianten fast genauso eng zusammen wie im Stillstand (Spannweite Stand: 0.36 s; Fahrt 0.35 s), die Interaktion ist nicht signifikant $F(4,76) = 0.88$, $p = .48$, $p_{\text{rep}} = .51$. Auch der Haupteffekt für Schnittstelle ist nicht signifikant ($F(4,76) = 0.35$, $p = .84$, $p_{\text{rep}} = .24$). In der Regleraufgabe mit neutraler Mitte (A2) lässt sich ebenfalls weder eine Interaktion in den Daten nachweisen ($F(4,76) = 0.75$, $p = .56$, $p_{\text{rep}} = .46$), noch ein Haupteffekt für

Schnittstelle ($F(4,76) = 0.69$, $p = .61$, $p_{\text{rep}} = .43$). Bei beiden Aufgabentypen ändert sich durch eine Veränderung der Ausgabe die Bearbeitungszeit nicht verlässlich. Es lässt sich jedoch nachweisen, dass die Bearbeitungszeiten während der Fahrt deutlich länger werden (A1: $F(1,19) = 28.32$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .60$; A2: $F(1,19) = 13.07$, $p < .01$, $p_{\text{rep}} = .98$, $\eta_p^2 = .41$).

Bei der vertikalen Regleraufgabe ohne Mittenmarkierung (A3) lässt sich ein Effekt der Schnittstelle in Abhängigkeit von der Bediensituation nachweisen ($F(4,76) = 3.23$, $p = .02$, $p_{\text{rep}} = .93$, $\eta_p^2 = .15$), was hauptsächlich an der drastischen Verlängerung der Zeiten bei TPsolo und insbesondere DD von Stand auf Fahrt liegen dürfte. Weiterhin ist erwähnenswert, dass der Haupteffekt für Bediensituation signifikant ist ($F(1,19) = 26.28$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .58$).

Besonders eindrucksvoll sind die Ergebnisse der 2D-Aufgabe (A4). Eine Betrachtung der Mittelwerte im Stand lässt nur marginale Unterschiede erkennen, während bei TPsolo und insbesondere bei DD massive Leistungseinbußen während der Fahrt zu verzeichnen sind (vgl. Anhang A 2). Die Bedienung über die Touchpad-Varianten mit zusätzlichem Feedback hingegen ist deutlich schneller. Zwischen der schnellsten Touchpad-Variante TPaud und DD ergab sich ein Unterschied von 2.07 Sekunden. Die Interaktion Bediensituation \times Schnittstelle ist entsprechend signifikant ($F(4,76) = 3.69$, $p < .01$, $p_{\text{rep}} = .95$, $\eta_p^2 = .16$). Auch ein genereller Unterschied zwischen Stand und Fahrt lässt sich nachweisen ($F(1,19) = 27.59$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .59$). Ebenfalls nachweisbar ist bei dieser Aufgabe ein allgemeiner Einfluss der Schnittstelle auf die Bearbeitungszeit ($F(4,76) = 8.99$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .32$).

Fehlerrate

Als Fehler war in diesem Experiment jede Bestätigungsoperation definiert, die nicht auf dem vorgegebenen Steuerelement erfolgte. Dargestellt wird im Folgenden der Anteil fehlerhafter Eingaben an allen vorgenommenen Eingaben. Dieser Anteil betrug maximal rund 5% und war damit eher niedrig. Ohne Fahraufgabe wurden bei der Schnittstellenvariante TPsolo 1.5% ($SE = 0.69$) derartiger Aktionen falsch ausgeführt. Bei TPaud waren dies 1.9% ($SE = 0.92$), bei TPtakt 3.7% ($SE = 1.68$) und bei TPaud+takt 2.2% ($SE = 1.17$). Die Nutzung des Dreh-Drückstellers zog in 5.4% ($SE = 1.73$) der Fälle eine falsche Bestätigung nach sich.

Während der Fahrsimulation ist keine massive und einheitliche Steigerung der Fehlerraten zu verzeichnen. Bei dem Touchpad ohne zusätzliches Feedback wurden unter Zusatzbelastung 3.2% ($SE = 1.55$) der Eingaben falsch getätigt. In einer vergleichbaren Größenordnung lagen die Fehlerraten beim Touchpad mit zusätzlichem Feedback und

dem Dreh-Drücksteller: TPaud 4.2% ($SE = 1.38$), TPtakt 3.5% ($SE = 1.57$), TPaud+takt 4.6% ($SE = 1.28$) und DD 3.91% ($SE = 1.06$).

Wie man den Standardfehlern entnehmen kann, streuen die Daten der ohnehin eher niedrigen Fehlerraten zum Teil beträchtlich. Es ist deshalb wenig überraschend, dass weder die zweifaktorielle Interaktion Schnittstelle \times Bediensituation ($F(4,76) = 1.30$, $p = .28$, $p_{\text{rep}} = .66$) noch einer der Haupteffekte signifikant ist (Bediensituation: $F(1,19) = 1.42$, $p = .25$, $p_{\text{rep}} = .68$; Schnittstelle: $F(4,76) = 0.68$, $p = .61$, $p_{\text{rep}} = .42$). Aufgrund dieser Ergebnislage werden bezüglich der Fehlerrate keine weiterführenden Paarvergleiche vorgenommen.

Randberührungen

Die Registrierung der Anzahl von Randberührungen soll Aufschluss darüber geben, inwiefern die auditiv und/oder taktil markierten Endpunkte einer Skala oder Liste die Orientierung erleichtern. Befand sich z.B. der Selektionscursor über dem letzten Element eines Reglers und der Nutzer bewegte seinen Finger weiter über das Ende des Reglers hinaus, bekam er in den Bedingungen TPaud, TPtakt und TPaud+takt eine direkte Rückmeldung, dass das Ende bereits erreicht ist. Zusätzlich wurde diese Aktion als Randberührung registriert. Dabei wurden nur die Aufgabentypen A1 bis A3 berücksichtigt, da bei der Eingabe mit dem zweidimensionalen Buchstabenfeld (A4) der Selektionscursor beim Dreh-Drücksteller sich zwangsläufig über die Zeilenenden hinaus bewegen musste, eine weitere Drehbewegung auf dem letzten Element hatte eine Rücksetzung auf das erste Element zur Folge.

Insgesamt konnten Randberührungen sehr selten beobachtet werden. Für die Analyse wurden die Randberührungen über die Aufgabentypen hinweg summiert und als mittlere Anzahl für jede Schnittstelle berechnet. Im Stillstand ergaben sich für die Schnittstellenvarianten (SE in Klammern): TPsolo 1.95 (0.24), TPaud 2.15 (0.40), TPtakt 2.15 (0.31), TPaud+takt 1.40 (0.22) und DD 2.90 (0.68). Nur geringe Unterschiede mit teils recht breiter Streuung sind zu verzeichnen. Das Bild blieb ähnlich, wenn gleichzeitig die LCT bearbeitet wurde: TPsolo 2.50 (0.67), TPaud 2.15 (0.36), TPtakt 3.25 (0.62), TPaud+takt 2.70 (0.36) und DD 2.30 (0.53). Es zeigten sich weder signifikante Haupteffekte (Bediensituation: $F(1,19) = 1.73$, $p = .20$, $p_{\text{rep}} = .72$; Schnittstelle: $F(4,76) = 0.84$, $p = .51$, $p_{\text{rep}} = .50$) noch eine signifikante Interaktion ($F(4,76) = 1.84$, $p = .13$, $p_{\text{rep}} = .79$). Substantiell unterschiedliches Vorgehen der Nutzer bezüglich dieser Strategiekomponente kann folglich nicht konstatiert werden.

Blickverhalten

In diesem Analyseschritt rücken die Dauer einzelner Blicke auf das Display während der simulierten Fahrt, die Dauer der einzelnen Blickzuwendungen und die Gesamtblickdauer bei der Interaktion mit den Schnittstellenvarianten während der simulierten

Fahrzeugführung in den Fokus. Die Grundlage für die Auswertung des Blickverhaltens bildet ein digitales Video des Gesichts der Versuchspersonen. Die Kamera war derart hinter dem Display montiert, dass ein Blick auf die Anzeigeeinheit klar durch die frontale Aufnahme der Pupillen erkennbar war. Das Video wurde manuell mit einem speziellen Programm ausgewertet, welches zur genauen Erfassung der Blicke ein verlangsamtes Abspielen der Bildsequenzen und eine Markierung von Anfang und Ende der Blickzuwendung gestattete. Die Gesamtbetrachtungszeit versteht sich als Summe aller einzelnen Blickdauern über alle Aufgabentypen und Wiederholungen hinweg. Gleiches gilt für die Blickhäufigkeit.

Die mittlere Dauer einzelner Zuwendungen für Kontrollblicke bei der Eingabe während der Fahrt ist bei TPsolo mit 0.79 s ($SE = 0.06$) am längsten. Die kürzeste Zeit wurde mit 0.66 s ($SE = 0.05$) bei TPaud erzielt. Bei den verbleibenden Touchpad-Varianten wurden für TPtakt 0.73 s ($SE = 0.04$) und für TPaud+takt 0.70 s ($SE = 0.04$) gemessen. Mit 0.67 s ($SE = 0.04$) war auch die Zeit bei DD sehr kurz. Abbildung 17 veranschaulicht die Datenlage grafisch. Eine einfaktorielle Varianzanalyse zeigte eine klare Tendenz für einen signifikanten Effekt des 5-stufigen Messwiederholungsfaktors Schnittstelle ($F(4,76) = 2.46, p = .05, p_{rep} = .87, \eta_p^2 = .12$). Da jedoch das Signifikanzniveau nur sehr knapp erreicht wurde, wird auf weitere Analysen (geplante Vergleiche) verzichtet.

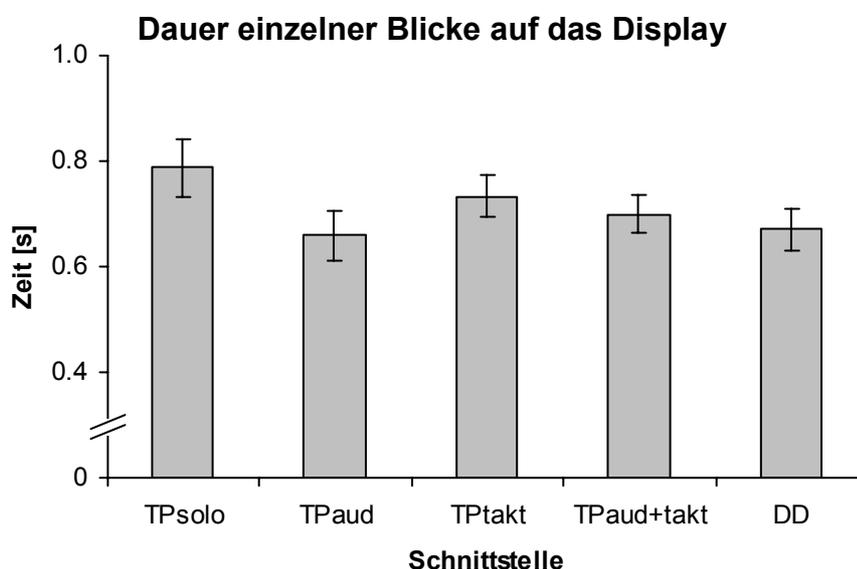


Abbildung 17: Durchschnittliche Dauer einzelner Blickzuwendungen zur Anzeigeeinheit in Abhängigkeit von der Schnittstellenvariante.

Ein etwas anderes Ergebnismuster zeigt sich bei einer Betrachtung der Häufigkeit von Kontrollblicken auf das Display. Unterschiedliche Auswirkungen durch die Schnittstellenvarianten sind hier zum Teil stärker ausgeprägt (siehe Abbildung 18).

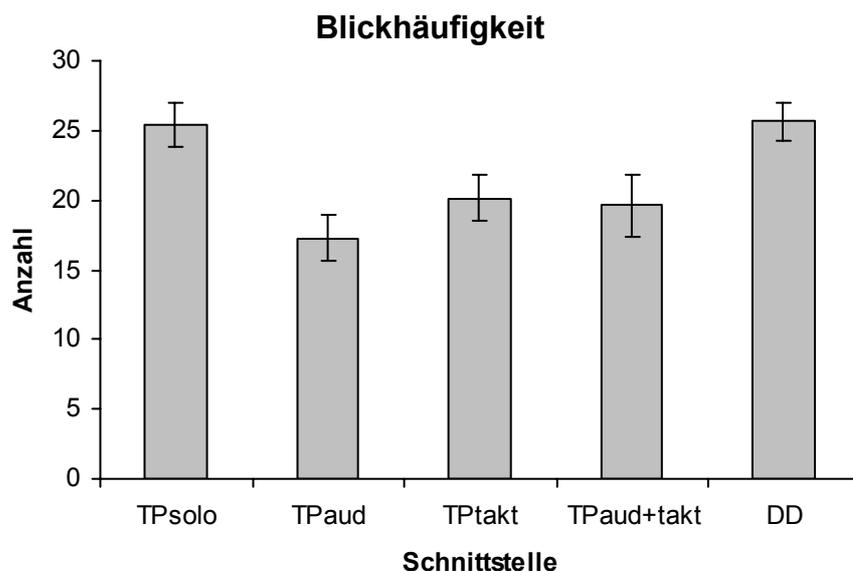


Abbildung 18: Mittlere Anzahl von Blicken auf die Anzeigeeinheit in Abhängigkeit von der Schnittstellenvariante (über Aufgaben summiert).

Das Touchpad mit ausschließlich visueller Rückmeldung und der Dreh-Drücksteller weisen annähernd gleiche Werte auf (TPsolo: $M = 25.40$, $SE = 1.57$; DD: $M = 25.65$, $SE = 1.38$). Die Touchpad-Varianten mit zusätzlicher Rückmeldung benötigen deutlich weniger Blickzuwendungen (TPaud: $M = 17.25$, $SE = 1.64$; TPtakt: $M = 20.15$, $SE = 1.60$; TPaud+takt: $M = 19.65$, $SE = 2.24$). Die Varianzanalyse bekräftigt diesen Eindruck als signifikanten Effekt der Schnittstelle, $F(4,76) = 9.48$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .33$. Auch die geplanten Vergleiche spiegeln den Eindruck der obigen Abbildung klar wider (siehe Tabelle 3). Zwischen TPsolo und DD können keine Unterschiede belegt werden. TPsolo benötigt hingegen signifikant mehr Blickzuwendungen als TPaud+takt. Ebenso wird die Anzeigeeinheit bei DD signifikant häufiger angeblickt als bei TPaud+takt.

Tabelle 3: Geplante Vergleiche zur Blickhäufigkeit.

Vergleich	$t(19)$	p	p_{rep}	d
TPsolo – TPaud+takt	2.70	.014*	.94	0.61
TPsolo – DD	-0.20	.844	0.24	-0.04
DD – TPaud+takt	3.21	.005*	.97	0.72

Die Ergebnisse der Gesamtdauer der Blickzuwendung entsprechen in etwa denen der Blickhäufigkeit, wie Abbildung 19 zu entnehmen ist. Entsprechend lässt sich auch hier ein durch die Schnittstelle verursachter hoch signifikanter Effekt konstatieren ($F(4,76) = 11.80$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .38$). TPsolo benötigt mit 19.15 s ($SE = 1.15$) am meisten visuelle Aufmerksamkeit. Wie bei der Blickhäufigkeit kom-

men die Touchpad-Varianten mit zusätzlichem Feedback mit weniger Zuwendung aus: TPaud 11.84 s ($SE = 1.33$), TPtakt 14.20 s ($SE = 1.07$) und TPaud+takt 13.35 s ($SE = 1.36$). Abweichend vom Datenmuster der Blickhäufigkeiten ist die gemessene Zeit bei DD mit 16.75 s ($SE = 0.97$) deutlich kürzer als der Wert bei TPsolo.

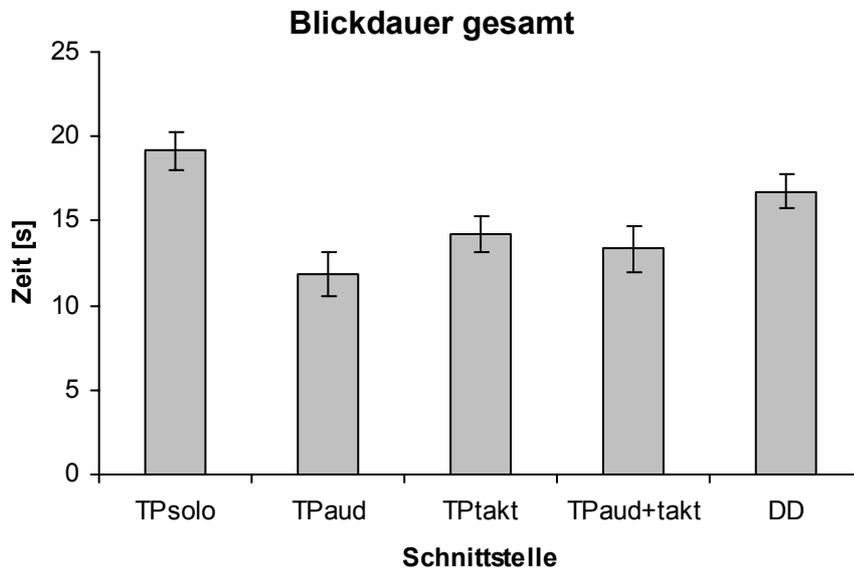


Abbildung 19: Mittlere Gesamtblickdauer auf die Anzeigeeinheit für jede Schnittstellenvariante (summiert über alle Aufgabentypen).

Inwiefern dieser Unterschied belastbar ist, soll durch geplante Vergleiche erörtert werden. Tabelle 4 fasst die Resultate zusammen. Die Vergleiche bestätigen die Vermutung, dass die Gesamtdauer der Blickzuwendung zum Display bei DD kürzer ist als bei TPsolo. Ferner ließ sich nachweisen, dass TPaud+takt die kürzeste Gesamtblickdauer im Vergleich mit DD und TPsolo aufweist.

Tabelle 4: Geplante Vergleiche zur Gesamtblickdauer.

Vergleich	$t(19)$	p	p_{rep}	d
TPsolo – TPaud+takt	4.16	.001*	.99	.93
TPsolo – DD	2.63	.016*	.93	.59
DD – TPaud+takt	3.46	.003*	.98	.77

Subjektive Workload

Mit Hilfe des NASA Task Load Index (TLX, Hart & Staveland, 1988) wurde die subjektiv wahrgenommene Workload bei der Interaktion mit jeder Schnittstellenvariante in einem mehrdimensionalen Beanspruchungskonzept erfasst. Dabei wird der relative Beitrag von sechs zu Grunde liegenden psychologischen Faktoren in je einer entsprechenden Subskala zu einer Gesamtbewertung zusammengefasst. Die ersten drei Subskalen haben eine aufgabenbezogene Ausrichtung: geistige Anforderungen,

körperliche Anforderungen und zeitliche Anforderungen. Die verbleibenden drei Subskalen sind verhaltensbezogen und reflektieren die Interaktion zwischen der Aufgabe und dem Operator: Ausführung der Aufgaben, Anstrengung und Frustration. Durch diese Konzeptualisierung der Komponenten von Workload ergibt sich die Möglichkeit einer vertiefenden Analyse auf Subskalen-Ebene, wenn im Gesamtergebnis keine erhellenden Befunde möglich sind. Zum Einsatz kam in diesem Experiment die Variante RTLX, die auf eine Gewichtung des Beitrags der Subskalen verzichtet. Stattdessen wird einfach der Durchschnitt der sechs Skalen berechnet. Es hat sich gezeigt, dass auf diesem Weg ebenso valide Resultate erzielt werden können (Byers et al., 1989). Die hier verwendete Übersetzung ist der Arbeit von Seifert (2002) entnommen und geht auf Unema, Rötting, Sepher-Willeberg, Strümpfel und Kopp (1988, zitiert nach Seifert, 2002) zurück. Die einzelnen Subskalen des eingesetzten RTLX weisen einen Bereich von 0 bis 20 auf, wobei 20 die Maximalangabe der erlebten Belastung darstellt. Beim NASA-TLX wird der Gesamtscore zwischen 0 und 100 ermittelt, indem die Bewertung auf jeder Dimension zunächst mit 5 multipliziert und anschließend das Produkt dieser skalierten Bewertungen und der Dimensionsgewichte durch die Summe der Gewichte (15) geteilt wird (vgl. Johannson et al., 2004, S.62). Beim RTLX werden Einheitsgewichte vergeben, sodass statt durch 15 durch 6 geteilt wird. Dies entspricht dem Mittelwert der auf den Bereich 0 bis 100 skalierten Bewertungen.

Nach Beendigung der jeweiligen Bedingung mit einer Schnittstellenvariante in Stand und Fahrt füllten die Teilnehmer den Fragebogen aus. Die Ergebnisse sind in Abbildung 20 dargestellt. Bei TPsolo berichteten die Teilnehmer mit 35.6 ($SE = 4.8$) die höchste Workload. Die anderen Touchpad-Varianten blieben deutlich unter diesem Wert: TPaud 23.9 ($SE = 3.4$), TPtakt 28.0 ($SE = 3.4$), TPaud+takt 24.9 ($SE = 3.3$).

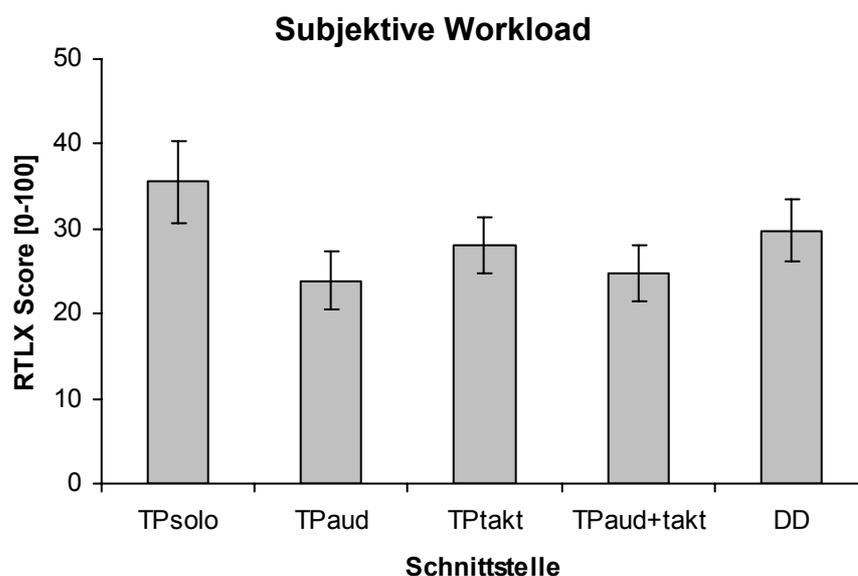


Abbildung 20: Mittlere RTLX Gesamtbewertung für jede Schnittstellenvariante (höhere Werte bedeuten höhere Workload).

Beim Dreh-Drücksteller wurde im Schnitt ein Wert von 29.8 ($SE = 3.7$) ermittelt. Die Unterschiede in den mittleren Gesamtwerten wurden durch eine einfaktorielle Varianzanalyse als signifikant bestätigt ($F(4,76) = 4.67$, $p < .01$, $p_{rep} = .98$, $\eta_p^2 = .18$). Eine Darstellung der Ergebnisse auf Subskalen-Ebene ist in Anhang A 3 zu finden. Wie in den vorigen Analyseschritten mit signifikanten Unterschieden wurden auch hier geplante Vergleiche zwischen TPsolo, TPaud+takt und DD gerechnet. Die Ergebnisse der t-Tests sind in Tabelle 5 zusammengefasst.

Tabelle 5: Geplante Vergleiche zu Workload.

Vergleich	$t(19)$	p	p_{rep}	d
TPsolo – TPaud+takt	2.54	.020*	.93	0.57
TPsolo – DD	1.76	.095	.82	0.39
DD – TPaud+takt	2.12	.048*	.88	0.47

Das Touchpad mit ausschließlich visuellem Feedback führt demnach zu höherer Workload als das Touchpad mit multimodaler Rückmeldung. Ebenso ist das Touchpad mit multimodaler Rückmeldung gegenüber dem Dreh-Drücksteller in Vorteil. Zwischen DD und TPsolo ließen sich keine stabilen Unterschiede nachweisen.

Fahrqualität

Die Abweichung der tatsächlichen Fahrspur von einem normativen Modell gibt in der Lane Change Task Auskunft über das Ausmaß der Ablenkung des Fahrers durch eine Zweitaufgabe und deren Einfluss auf die Fahrqualität. Für jeden Versuchsteilnehmer wurde ein Durchschnittswert für die Baselinefahrt ohne Zweitaufgabe und für jede Fahrt mit gleichzeitiger Bedienung über eine der Schnittstellenvarianten berechnet. Anschließend wurde der Abweichungswert für jede Schnittstelle durch den Abweichungswert der Baselinefahrt geteilt. Dies liefert ein Maß der relativen Leistungsabnahme durch die Zweitaufgaben (vgl. Mattes, 2003). Abbildung 21 zeigt die relative Abweichung vom normativen Modell für alle Schnittstellenvarianten.

Bei TPsolo ergibt sich ein Wert von 1.00 ($SE = 0.03$), ebenso bei TPaud ($M = 1.00$, $SE = 0.03$). Der Quotient erreicht bei TPtakt 1.03 ($SE = 0.03$), bei TPaud+takt 1.05 ($SE = 0.03$) und bei DD 1.04 ($SE = 0.04$). Niedrigere Werte deuten auf eine bessere Fahrqualität hin. Es fällt ins Auge, dass der Quotient bei allen Bedingungen annähernd 1 ergibt. Das bedeutet, dass es allgemein kaum Einflüsse durch die Zweitaufgabe gab. Darüber hinaus legt die Grafik nahe, dass nur sehr schwache Unterschiede in den Auswirkungen der Schnittstellen existieren. In diese Richtung weist auch das Ergebnis der Varianzanalyse: Die leichten Schwankungen erreichen trotz geringer Streuungen in den Mittelwerten nicht statistische Signifikanz ($F(4,76) = 1.7$, $p = .15$, $p_{rep} = .77$).

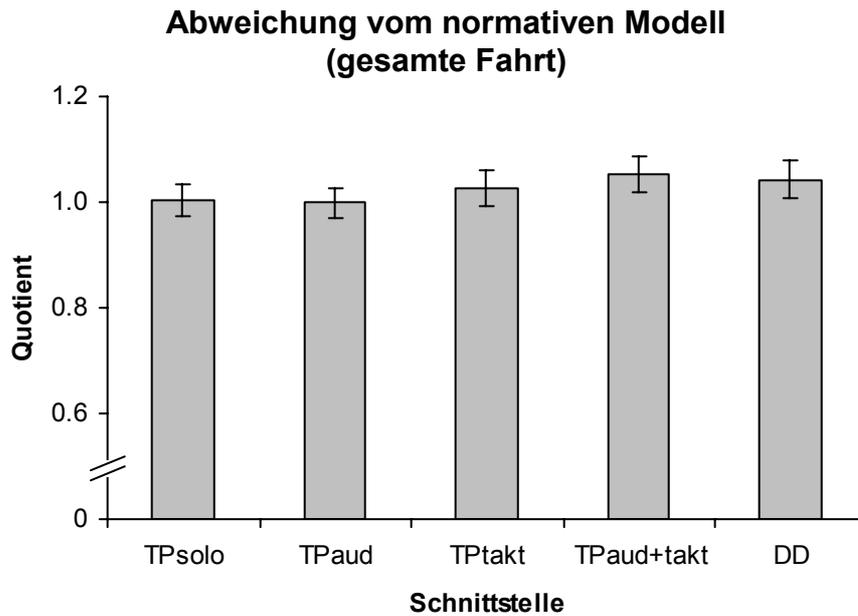


Abbildung 21: Ergebnisse der LCT für jede Schnittstellenvariante während der Fahrt (Quotient aus der Abweichung mit/ohne Zweitaufgabe).

Die LCT ermöglicht durch das Setzen von Markern, nur die Messpunkte während der Sekundäraufgabe zu betrachten. Setzt man diese Daten in Bezug zur Baseline ergibt sich folgende Sachlage (Standardfehler in Klammern): TPsolo 0.85 (0.04), TPaud 0.69 (0.05), TPtakt 0.78 (0.06), TPaud+takt 0.78 (0.06) und DD 0.80 (0.04). Abbildung 22 gibt eine grafische Veranschaulichung der Daten.

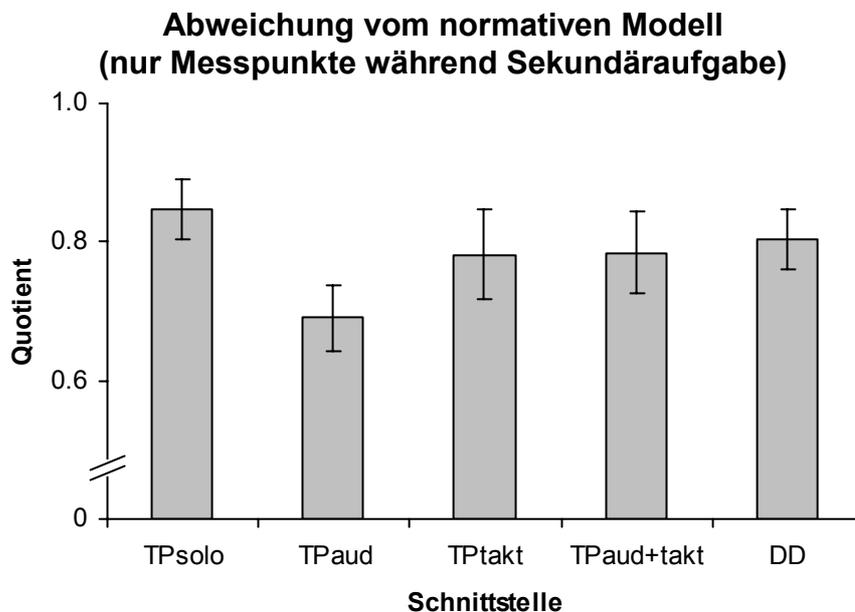


Abbildung 22: Ergebnisse der LCT während der Bearbeitung der Sekundäraufgaben für alle Schnittstellen (Quotient aus Abweichung der Baseline/Messpunkte der Sekundäraufgaben in der Experimentalfahrt).

Typischerweise werden bei dieser Art der Analyse Werte größer als 1 gefunden, da die Fahrqualität während der Bearbeitung der Sekundäraufgabe schlechter ist als die Baseline (gesamt). Wegen der Kürze der Bearbeitungszeit der einzelnen Aufgaben konnten diese in sehr vielen Fällen abgeschlossen werden, bevor ein Spurwechsel stattfand. Damit gingen in die Baseline-Daten wesentlich mehr Spurwechsel ein und die Daten während der Sekundäraufgabe zeigen hauptsächlich die Leistung bei der Stabilisierung der Querführung an. Bei Spurwechseln ist die Abweichung vom normativen Modell größer als auf der Geraden. Das erklärt, weshalb die hier gefundenen Werte unter 1 liegen.

Die Varianzanalyse zeigt, dass sich Unterschiede in der Fahrqualität nachweisen lassen: $F(4,76) = 2.47$, $p = .05$, $p_{\text{rep}} = .87$, $\eta_p^2 = .12$. Überraschend ist jedoch, dass nur bei TPaud eine deutliche Veränderung der Fahrqualität erkennbar ist. Aufgrund des Ergebnismusters, das bei den kritischen geplanten Vergleichen kaum Unterschiede andeutet, und des Resultats der Varianzanalyse wurden keine Vergleiche zwischen den einzelnen Schnittstellen berechnet.

Präferenzurteile

Nach Abschluss des Experiments wurden die Teilnehmer gebeten, die Schnittstellenvariante zu benennen, die ihnen am meisten zugesagt hat. Diese Ergebnisse werden nur deskriptiv berichtet. Am häufigsten wurden TPaud und TPaud+takt mit jeweils sechs Nennungen gezählt. Ebenfalls gleichauf nehmen TPtakt und DD den zweiten Rang mit jeweils vier Nennungen ein. TPSolo wurde nicht genannt. Ferner wurde von fünf Teilnehmern in der abschließenden Besprechung zusätzlich geäußert, dass der Dreh-Drücksteller mit auditivem Feedback eine sehr reizvolle Alternative wäre. Da diese Aussage jedoch nur auf Vorstellung, nicht auf Erfahrung beruht, wird sie nur als Randnotiz vermerkt.

4.1.2.4 Diskussion

Im vorliegenden Experiment wurde untersucht, ob multimodale Ausgabe die Interaktion mit einem Touchpad zur Bedienung von Komfortfunktionen im Fahrzeug in einem schalterreduzierten Anzeige- und Bedienkonzept verbessern kann. Ferner sollte erschlossen werden, inwiefern sich ein Touchpad generell zur Bedienung dieser Funktionen eignet. Anhand einer Reihe verschiedener Maße wurden hierzu erste Ergebnisse gesammelt, die im Folgenden diskutiert werden. Die Aufgabe der Testpersonen im Experiment bestand darin, einmal ohne und einmal mit Fahraufgabe diverse Eingaben in ein vereinfachtes Bordsystem zu tätigen. Dazu nutzten sie entweder einen herkömmlichen Dreh-Drücksteller oder ein Touchpad, mit dessen Hilfe sie den Selektionscursor auf einer grafischen Bedienoberfläche auf das jeweils gewünschte Element bewegen konnten. Zur Untersuchung der Effekte multimodaler Rückmeldung wurden

mehrere Varianten des Touchpads getestet: ein Touchpad mit ausschließlich visueller Rückmeldung, eine Touchpad-Variante mit zusätzlich auditiv wahrnehmbarer Ausgabe, eine mit zusätzlicher taktiler Ausgabe und eine mit zusätzlicher auditiver und taktiler Rückmeldung.

Zusammengefasst lässt sich festhalten, dass Rückmeldung in einer weiteren Modalität neben der visuellen Ausgabe generell die Qualität der Mensch-Maschine-Interaktion deutlich verbessert hat. Bimodale Rückmeldung ist damit der unimodalen Rückmeldung vorzuziehen. Eine weitere Verbesserung durch trimodale Rückmeldung gegenüber der bimodalen Form war nicht konsistent erkennbar. Wenn mindestens bimodale Rückmeldung gegeben wurde, führte darüber hinaus die Touchpad-basierte Eingabe auf mehreren Dimensionen zu besseren Resultaten als der Dreh-Drücksteller. Das Touchpad mit ausschließlich visueller Rückmeldung hingegen war dem Dreh-Drücksteller fast durchgehend unterlegen. Diese Ergebnisse spiegeln sich auch in den Präferenzurteilen der Versuchsteilnehmer wider.

Aufgabenbearbeitungszeit

Einer der wichtigsten Aspekte jeder Schnittstellengestaltung zur Bedienung von Sekundär- und Tertiäraufgaben ist einen schnellen Arbeitsfluss zu unterstützen. Je kürzer die Interaktion zur Ausführung dieser Aufgaben dauert, umso geringer die Zeitspanne der Abwendung von der Primäraufgabe. Deshalb wird der Diskussion der Bearbeitungszeiten besonderer Raum eingeräumt. In unterschiedlichen Kontexten konnte bereits nachgewiesen werden, dass eine zusätzliche nicht-visuelle Form der Informationsausgabe die Aufgabenzeit reduzieren kann. So konnten Akamatsu und MacKenzie (1996) mit einer speziellen Computermaus demonstrieren, dass die Zeit für Zielbewegungen zuverlässig durch taktilen Feedback verringert wurde. Sklar und Sarter (1999) fanden in einer Überwachungsaufgabe mit visueller Primäraufgabe eine höhere Entdeckungsrate und schnellere Aufgabenbearbeitung durch taktile Ausgaben. Göbel et al. (1995) registrierten in einer Reihe von Experimenten zwar Geschwindigkeitseinbußen in Tracking-Aufgaben durch taktilen Feedback, wiesen aber dafür in ballistischen Aufgaben bis zu 29% schnellere Aufgabenbearbeitung nach. Auch mit auditiver Rückmeldung lassen sich diese Vorteilseffekte erzielen. In den Untersuchungen von Brewster und seiner Arbeitsgruppe konnte mehrfach nachgewiesen werden, dass auch nicht-sprachliche Audiosignale als Unterstützung der visuellen Informationsaufnahme einen vorteilhaften Effekt auf die Interaktionsgeschwindigkeit aufweisen. Sie zeigten unter anderem, dass Zielbewegungen oder Drag-and-Drop Operationen mit auditiver Rückmeldung effizienter ausgeführt werden können (z.B. Brewster, 1998a, 1998b). Auch in Studien zu Automatisierung liefern redundante auditive Ergänzungen konsistent bessere Entdeckungsraten und schnellere Aufgabenzeiten (vgl. Kaber et al., 2006). Allerdings fokussiert die meiste Literatur zu auditiver Optimierung der

Mensch-Maschine-Interaktion mehr auf Fehlerreduktion als auf Selektionszeiten, so dass diesbezüglich keine klare Aussage möglich ist (vgl. Brewster & Crease, 1997; Cockburn & Brewster, 2005).

In einem direkten Vergleich der Ausgabeformen in multimodalen Experimenten mit mehr als zwei Modalitäten hat sich in einer Studie von Akamatsu und Kollegen gezeigt, dass taktiler Feedback als Hinweis zur Positionierung über einem Target gegenüber akustischer und rein visueller Ausgabe bezüglich der zeitlichen Dimension im Vorteil ist (Akamatsu, MacKenzie & Hasbrouq, 1995). Wenngleich ein signifikanter Effekt nachgewiesen werden konnte, waren die Unterschiede zwischen den multimodalen Varianten jedoch eher gering, allein die rein visuelle Ausgabeform führte zu deutlich langsameren Resultaten. Cockburn und Brewster (2005) demonstrierten, dass durch nicht-sprachliche akustische oder taktile Ausgabe die Erfassung sehr kleiner Zielelemente auf grafischen Benutzeroberflächen beschleunigt werden kann. Auch hier gestattete die taktile Modalität die schnellste Positionierung.

Daneben existieren jedoch Befunde, die zu einer weniger eindeutigen Situation beitragen. In einer Studie von Vitense und Kollegen hat haptisches Feedback sogar zu einer vergleichsweise langen Bearbeitungszeit geführt (Vitense et al., 2003). Sie konnten nachweisen, dass die Gesamtzeit für die Aufgabenbearbeitung mit haptischem Feedback schlechter ist als mit auditiver oder sogar ohne zusätzliche Rückmeldung. Nur eine Komponente der Interaktion profitierte von haptischem Feedback: die Zeit über dem Zielitem vor der Drop-Operation in Drag-and-Drop. In weiteren Studien ergaben sich zwar Vorteile bei der Fehlervermeidung, aber nicht in der Bearbeitungszeit (z.B. McGee, 1999; Oakley et al., 2000).

Ein maßgeblicher Unterschied zwischen den geschilderten Untersuchungen und dem vorliegenden Experiment besteht in der hier vollzogenen Anlage in einer Single- und Dual Task Situation. Die Hypothese wurde aufgestellt, dass unter Aufwendung der vollen Aufmerksamkeit mit jeder Schnittstellenvariante ähnlich schnelle Bearbeitungszeiten erzielt werden können. Wenn gleichzeitig die simulierte Fahraufgabe als Hauptaufgabe bewältigt werden muss, entstehen durch die Verknappung informativischer Ressourcen so genannte *dual task costs*, zu deren Ausgleich die multimodale Ausgabe dienen kann. Über die Auswirkungen einer Unterstützung der Eingabe durch akustische und haptische Rückmeldungen auf die Bearbeitungszeit existieren in diesem Kontext bisher kaum gezielte Studien. Untersuchungen zu Multimodalität im Fahrzeug beziehen sich zumeist entweder auf eine Verbesserung der Leistungen bei der Fahrzeugführung durch multimodale Eingabe (z.B. Sprache oder Gestik, vgl. Akyol et al., 2001; Graham et al., 1999; Herfet, Kirste & Schnaider, 2001) oder im Fall von multimodaler Ausgabe direkt auf fahrrelevante Aspekte wie Warnhinweise oder Informationsausgaben (Liu, 2001; Vitense et al., 2003).

Betrachtet man zunächst die Bearbeitungszeiten im Stillstand und bei gleichzeitiger Fahrzeugführung in der Simulation, lassen die generell höheren Zeiten während der Fahrsituation auf den vermuteten Ressourcenkonflikt schließen. Da die Teilnehmer die Primäraufgabe zu schützen hatten, passten sie ihre Strategie beim Time-Sharing zu Ungunsten der Sekundäraufgabe an und arbeiteten länger an einer Aufgabe. Dabei gilt es zu beachten, dass die simulierte Fahrt grundsätzlich erst nach der Aufgabenbearbeitung mit der jeweiligen Schnittstelle im Stand erfolgte und die Versuchspersonen also schon entsprechend Übung mit der Eingabe hatten. Dieser Übungseffekt konnte die Belastung durch die multiple Aufgabensituation jedoch nicht ausgleichen. Die erhöhten Bearbeitungszeiten während der Fahrt sind nicht an sich von Bedeutung, da sie im Grunde nur eine Replikation des bekannten Befunds darstellen, dass die Leistung bei Ressourcenkonflikten und dem daraus erwachsenden Time-Sharing für die niedriger priorisierte Aufgabe Einbußen bringt. Sie sind hingegen durchaus relevant als Nachweis, dass in dem Experiment durch die Lane Change Task erfolgreich eine derartige Konfliktsituation etabliert wurde.

Inhaltlich deutlich relevanter ist die vorgefundene Interaktion zwischen Schnittstelle und Bediensituation. Während ohne Fahraufgabe kaum Unterschiede zwischen den Schnittstellen existieren, ergeben sich durch die Interferenz der visuellen Last der Lane Change Task mit der gleichzeitigen Kontrolle der Bedienvorgänge über die Anzeigeeinheit deutliche Differenzen in den Bearbeitungszeiten der Schnittstellenvarianten. Dabei ließ sich nachweisen, dass die Ergänzung des Touchpads um zusätzliche Ausgaben einen deutlich positiven Effekt hatte und eine schnellere Bearbeitung als mit dem Touchpad ohne zusätzlichem Feedback oder dem Dreh-Drücksteller während der Fahrt ermöglichte. Ein besonderer Vorteil der kombinierten Ausgabe von zusätzlichem auditivem und taktilem Feedback war nicht erkennbar.

Auffällig war weiterhin, dass nicht in allen Aufgabentypen der Vorteil multimodaler Ausgabe gleichermaßen vorlag. Nur in den Aufgabentypen 3 (vertikaler Regler) und 4 (2D-Eingabe) wurde der Effekt deutlich. Interessant ist, dass Aufgabe 3 deutliche Parallelen zu der Menüaufgabe (Aufgabe 1) in der Eingabeoperation auf einem vertikalen eindimensionalen Feld aufweist. In Aufgabe 1 wirkte sich die multimodale Ausgabe jedoch auch während der Fahrt nicht erwartungsgemäß aus. Eine mögliche Erklärung liegt darin, dass die Menüaufgabe aus mehreren Eingabeschritten besteht, die leicht segmentierbar waren. Die Nutzer konnten möglicherweise darauf die Strategie ihrer Eingabeschritte während der Fahrt besser wählen, weil sie vorausschauend arbeiten und so ihr Vorgehen besser mit den Erfordernissen der Fahrsituation in Einklang bringen konnten. Des Weiteren wies die Regler-Steuerung deutlich mehr Elemente auf als die einzelnen Menülisten. Die geringere Komplexität der in dieser Hinsicht einfacheren Menüaufgabe kann ebenfalls dazu beigetragen haben, Unterschiede zu verwischen. Diese These wird dadurch gestützt, dass die schwierigste Auf-

gabe, die Eingabe im zweidimensionalen Buchstabenfeld, zu den deutlichsten Unterschieden geführt hat. Darüber hinaus zeigte sich bei dieser Aufgabe erwartungsgemäß, dass das Touchpad selbst ohne zusätzliches Feedback aufgrund der besseren Abbildung der Erfordernisse einer zweidimensionalen Eingabe zu kürzeren Bearbeitungszeiten als der Dreh-Drücksteller führte. Besonders interessant ist an diesem Aufgabentyp, dass während der Fahrt eklatante Nachteile des Dreh-Drückstellers nachweisbar waren, diese jedoch im Stillstand ohne Fahraufgabe durch fokussierten Ressourceneinsatz kompensiert werden konnten.

Erklärenswert ist weiterhin, weshalb bei der horizontalen Regleraufgabe mit auditiver bzw. taktiler Markierung der neutralen Mitte keine Effekte nachgewiesen werden konnten, während sich bei dem vertikalen Regler ohne Markierung Unterschiede einstellten. Die Beobachtung der Probanden bei der Interaktion mit dem Touchpad während des Experiments und die Aussagen einiger Teilnehmer nach dem Experiment legen nahe, dass die Markierung der neutralen Mitte nicht optimal gelöst war und einen unbeabsichtigten Effekt bewirkte. Aufgrund der Intensität und der gegenüber anderen Menüelementen verlängerten Dauer des haptischen Signals bei der neutralen Mitte (200 ms) tendierten die Versuchspersonen dazu, das Ende dieser Ausgabe abzuwarten, bevor sie ihre Fingerbewegung fortführten. Ferner war aufgrund der Gestaltung der Ausgabe die Semantik nicht eindeutig. Mehrfach wurde von den Probanden angegeben, dass die verlängerte und intensivere Vibration einen warnenden Charakter aufweise, ähnlich einer Fehlermeldung. Es ist somit wahrscheinlich, dass in mehreren Fällen die Eingabeoperation über der markierten Mittelposition für eine kognitive Prüfung der Bedeutung des Signals unterbrochen wurde. Die entstandenen kurzen Pausen durch die genannten Faktoren können dazu geführt haben, dass die Bedienung über das Touchpad mit taktiler Ausgabe bzw. mit akustischer und taktiler Ausgabe mehr Zeit benötigte und so Unterschiede zu der Touchpad-Variante ohne zusätzliche Rückmeldung bzw. zum Dreh-Drücksteller vermindert wurden.

Fehlerrate

Die Analyse der Fehler wies keine signifikanten Unterschiede aus. Das Ergebnismuster ist insofern überraschend, weil auch kein einheitlicher Anstieg der Fehlerrate während der Fahrsimulation festgestellt werden konnte. Die Versuchsteilnehmer waren offensichtlich in der Lage, die Qualität der Bearbeitung trotz der Zusatzbelastung aufrecht zu erhalten. Ein Teil der beobachteten längeren Aufgabenbearbeitungszeiten kann den dafür notwendigen Bemühungen zugeschrieben werden. Die relativ geringe Fehlerrate mit maximal rund 5% Falscheingaben deutet darauf hin, dass alle Eingabeformen generell eine fehlerarme Bedienung zulassen. Daraus lässt sich eine alternative Erklärung für fehlende Unterschiede ableiten: Ein Ceiling-Effekt bezüglich der Leistung könnte vorliegen. Falls die vorgefundenen Fehlerraten während der Fahrt eine Art

untere Grenze für diese Art von Aufgaben darstellen, ist nicht zu erwarten, dass es im Stand weitere Reduktionen gibt. Ein solcher Ceiling-Effekt würde auch erklären, warum multimodale Rückmeldung keine weitere Fehlerreduktion bewirkt hat, wie man sie typischerweise bei Experimenten mit zusätzlicher redundanter Ausgabe akustischer oder haptischer Ausgaben findet (Brewster & Crease, 1997; Oakley et al., 2000).

Als Fehler wurden ausschließlich Bestätigungsoperationen registriert, die nicht auf dem vorbestimmten Zielelement stattfanden⁷. Ignoriert wurden hingegen zusätzliche Selektionsbewegungen, die nicht notwendig gewesen wären, z.B. so genannte *overshoots*. Da diese zusätzlichen Selektionsbewegungen ebenfalls korrigiert werden müssen, hätte man sie als Ausführungsfehler einordnen können. Aus zwei Gründen wurde darauf verzichtet. Erstens ist diese Form des Fehlers in der Praxis für den gegebenen Nutzungskontext vergleichsweise unproblematisch, da sie sehr leicht vom Nutzer zu korrigieren ist. Bei einer falschen Bestätigung hingegen kann es passieren, dass der Nutzer z.B. in das falsche Menü gelangt oder einen falschen Buchstaben in das Eingabefeld eingibt. Das hat im Vergleich zu einer unabsichtlich ausgeführten Selektion eine aufwändigere Fehlerbehebungsprozedur zur Folge. Zweitens ist bei zusätzlichen Selektionen nicht unbedingt klar, ob tatsächlich eine Fehlhandlung vorliegt oder inwiefern sie Teil einer bewussten Strategie sind. Dieser Aspekt wird im folgenden Abschnitt bei der Diskussion der Randberührungen vertieft.

Randberührungen

Zur Illustration des Terminus Randberührung soll folgendes Beispiel dienen: Ein Nutzer befindet sich bei der Auswahl eines Menü-Items auf dem ersten Element und versucht, den Cursor weiter nach oben zu verschieben oder er versucht auf dem letzten Element eine Bewegung nach unten auszuführen. In beiden Fällen ist keine weitere Cursorbewegung möglich, der Cursor stößt gewissermaßen am „Rand“ des Menüs an.

Diese Randberührungen können unter Umständen auch als Eingabefehler, als nicht optimales Erreichen des Zielelements definiert werden. Gegen die Interpretation als Fehler spricht die Möglichkeit, dass der Rand als Orientierungshilfe strategisch bei der Blindbedienung genutzt werden kann, um ohne visuelle Kontrolle einen definierten Ausgangspunkt für eine einfachere Zielwahl festzulegen. So kann ein Nutzer, der z.B. das vorletzte Menüelement anwählen möchte, bewusst zunächst mit einer langen Bewegung den Selektionscursor komplett an das Ende des Menüs manövrieren, um dann mit einem gezielten Rückschritt das gewünschte Element auszuwählen. Durch die Ausgabe eines haptischen bzw. akustischen Signals bei einer Randberührung in den multimodalen Varianten des Touchpads bestand die Möglichkeit, aktiv mit dieser

⁷ In der Terminologie von Reason (1994, S.27ff.) handelt es sich bei diesem Fehlertyp genau genommen um „Patzer“ (*slips*), die auf der Ebene der fähigkeitsbasierten Ausführung auftreten. Vereinfachend wird im Text weiterhin der Begriff „Fehler“ verwendet.

Strategie zu arbeiten. Bei den getesteten Bedienungsaufgaben konnte eine derartige Strategie jedoch nicht eindeutig nachgewiesen werden. Wenn tatsächlich die Endpunkte eines Menüs oder Reglers für die Blindbedienung genutzt worden wären, hätte dieses Verhalten während der Fahraufgabe deutlich häufiger zu beobachten sein müssen. Das war nicht der Fall, zwischen Fahrt und Stillstand ließ sich kein Unterschied nachweisen.

Des Weiteren konnten auch keine eindeutig unterschiedlichen Auswirkungen der Schnittstellenvarianten identifiziert werden. Das wirft die Frage auf, ob die Aufgabentypen in diesem Experiment dazu geeignet waren, eine bewusste Strategie der Versuchsteilnehmer zu fördern. Die Menüs und Regler hatten alle gleichermaßen feste Endpunkte, ein Rücksprung vom letzten zum ersten Element oder umgekehrt durch Überdrehen war nicht möglich. Damit konnten die Teilnehmer den Cursor an einem der Endpunkte auch mit dem Dreh-Drücksteller oder dem Touchpad mit unimodalem visuellem Feedback ohne Blickkontakt platzieren. Sie konnten dies zwar nicht so effizient wie bei den multimodalen Touchpad-Varianten ausführen, weil sie wegen der fehlenden Rückmeldung über das Erreichen der Endpunkte möglicherweise eine unnötig intensive Bewegung initiierten. Dieser Nachteil war jedoch nur mit zeitlichen Kosten verbunden. Hochgradig interessant wäre an dieser Stelle ein Vergleich von Menüs mit festen Endpunkten und von Menüs mit der Möglichkeit zu Überdrehen. Erst in letzterem Fall wäre der Informationsvorteil der multimodalen Varianten für die Blindbedienbarkeit von einflussreichem Vorteil.

Für dieses Experiment muss auf der Basis der beschriebenen Daten gefolgert werden, dass sich keine unterschiedliche Strategie der Teilnehmer durch die Rückmeldung der Randpositionen in den multimodalen Touchpad-Varianten beobachten ließ. Solange konsistent feste Endpunkte in Menüs oder Reglern verwendet werden, die ein Überdrehen nicht zulassen, profitieren die Nutzer nicht von dieser zusätzlichen Information.

Blickverhalten

Die weitgehende Fokussierung der visuellen Aufmerksamkeit auf die Fahraufgabe ist für eine sichere Fahrzeugführung unabdingbar. Einige Forscher gehen soweit, von einem visuell aufgenommenen Informationsanteil von 90% beim Fahren zu sprechen (Hills, 1980). Diese Einschätzung wird zwar bezüglich der Höhe der Quantifizierung nicht von allen Befunden gestützt (z.B. Fairclough & Maternaghan, 1993), aber die Dominanz visueller Informationsaufnahme für die unmittelbare Gefahrenerkennung, vorausschauendes Fahren und Spurhaltung ist unbestritten (Evans, 1991). Die Analyse von Blickdauer und -häufigkeit auf Orte der Informationsausgabe im Fahrzeug ist deshalb für die Beurteilung der Schnittstellengestaltung unerlässlich. Analysen zu

typischen Blickzeiten finden sich z.B. bei Wierwille, Antin, Dingus und Hulse (1988), Rockwell (1988) sowie Zwahlen, Adams und DeBald (1988).

Im Mittelpunkt der Analyse dieses Experiments steht die Rolle der multimodalen Ausgabe des Touchpads bei der Allokation der visuellen Ressourcen. Bereits bei der Analyse der Dauer einzelner Blicke hat sich als Tendenz abgezeichnet, dass die multimodale Ausgabe gegenüber der rein visuellen Ausgabe zu kürzeren Blickzuwendungen führt. Die anschließende Betrachtung der Blickhäufigkeit und der Gesamtblickdauer zeigte dann eindeutig die hypothesenkonformen Unterschiede zwischen dem Touchpad mit unimodaler visueller Ausgabe und den multimodalen Varianten. Zusätzliches Feedback führte bei der Blickhäufigkeit zu einer Verringerung der Kontrollblicke zwischen 21% und 32% sowie bei der Gesamtblickdauer zu einer Verkürzung zwischen 26% und 38%. Dieses Resultat demonstriert eindrucksvoll das Potential multimodaler Ausgabe zur Entlastung des visuellen Kanals durch Redundanz. Es trägt auch zur Erklärung der geringen Fehlerrate bei: Fehler konnten selbst beim unimodalen Touchpad erfolgreich während der Fahrt vermieden werden, jedoch nur auf Kosten einer höheren visuellen Kontrolle und damit unter Vernachlässigung eines wichtigen Aspekts der Primäraufgabe.

Eine weitere Grundfrage des Experiments betrifft den Vergleich zwischen den Eingabeformen Touchpad und Dreh-Drücksteller. Die Hypothese wurde aufgestellt, dass ein Touchpad für die typischen Eingabeoperationen besser geeignet sei als ein Dreh-Drücksteller, dass jedoch diese Vorteile durch die Kosten einer vermehrten visuellen Zuwendung kompensiert würden, sofern keine zusätzliche Ausgabemodalität involviert ist. Auch diese Hypothese wird von den Testresultaten untermauert. Zwar benötigten die Teilnehmer nahezu gleich viele Kontrollblicke bei dem Dreh-Drücksteller und bei dem unimodalen Touchpad, aufgrund der kürzeren Kontrollblicke zeigt sich jedoch bei der Gesamtbetrachtungszeit eine signifikant kürzere Zuwendung als beim Dreh-Drücksteller. In beiden Fällen erwies sich das multimodale Touchpad als überlegen.

Subjektive Workload

Die Messung von Workload ist bereits seit langem ein fester Bestandteil der Evaluation von komplexen interaktiven Systemen, da eine Optimierung auf dieser Dimension die Wahrscheinlichkeit für Fehlhandlungen senken und gleichzeitig die Sicherheit des Systems sowie die menschliche Leistungsfähigkeit erhöhen kann (Gopher & Donchin, 1986; Moray, 1982). Gerade bei Experimenten zu multimodaler Interaktion wird sehr häufig die Höhe der subjektiv erlebten Workload festgestellt, um vor dem Hintergrund des multiplen Ressourcenmodells mögliche Kosten oder Vorteile für die Informationsverarbeitung zu erkennen. Es hat sich gezeigt, dass der NASA-TLX und einige seiner Varianten sehr sensitiv auf experimentelle Manipulationen in Untersuchungen zu

Multimodalität reagieren (z.B. Brewster, 1997; Oakley et al., 2002; Vitense et al., 2003).

Auch im vorliegenden Experiment konnten mit Hilfe einer TLX-Variante, dem RTLX, Unterschiede in der subjektiven Workload bei der Interaktion mit uni- vs. multimodalen Schnittstellen nachgewiesen werden. Das Ergebnismuster spiegelt hier die Resultate der objektiven Maße der Blick- und Aufgabenbearbeitungszeiten während der Fahrt wider. Das Touchpad mit unimodaler visueller Ausgabe verursacht die höchste empfundene Workload. Bei den Varianten mit zusätzlichem Feedback gaben die Teilnehmer geringere empfundene Beanspruchung an. Dabei fallen die Bewertungen des Touchpads ohne zusätzliche Rückmeldung gleichzeitig signifikant schlechter aus als die des Dreh-Drückstellers. Typischerweise erhofft man sich von den Workload-Daten zusätzliche Information, die noch nicht durch die anderen Maße gegeben wurde. Das kann im vorliegenden Fall nicht berichtet werden. Es ist jedoch durchaus beachtenswert, wie deutlich sich das Einflussmuster der Modalitätsmanipulation beim RTLX erhält. Nicht nur, dass das Touchpad ohne zusätzliches Feedback am schlechtesten abschneidet, sondern es bleibt auch das Touchpad mit akustischer Ausgabe die beste Ergänzungslösung. Das kombinierte-multimodale Touchpad hat sich nicht am besten bewährt, steht aber immer noch über der taktilen Variante.

Fahrqualität

Der Einfluss der Schnittstellenvarianten auf die Fahrqualität wurde mittels der Lane Change Task evaluiert. Die Hypothese, dass die Leistung bei der LCT dann besser ist, wenn multi- statt unimodaler Rückmeldung bei der Touchpad-Eingabe realisiert wird, hat sich nicht bestätigt. Die Hypothese basierte auf der Vermutung, durch eine unterstützende Wirkung der zusätzlichen Rückmeldung auf den bisher genannten Maßen könnten besseres Time-Sharing ermöglicht und so mehr Ressourcen auf die Fahraufgabe allokiert werden. Obwohl sich bei der Aufgabenbearbeitungszeit, den Blickdaten und der Workload klare Vorteile für die auditive, taktile oder kombinierte Ergänzung der Ausgabe ergeben haben, zeigten sich nicht die erwarteten Unterschiede. Betrachtet man die dreiminütige Fahrt als Ganzes, ist im Abweichungswert vom normativen Modell weder durch die uni- und multimodalen Varianten des Touchpads noch durch den Dreh-Drücksteller ein klarer Einfluss erkennbar. Selbst wenn nur die Streckenabschnitte betrachtet werden, in denen eine Bearbeitung der Zweitaufgabe stattgefunden hat, erreicht der Unterschied zwischen den Schnittstellen nur knapp statistische Bedeutsamkeit. Erstaunlicherweise wird dieser tendenzielle Effekt durch das Touchpad mit zusätzlich auditivem Feedback erzeugt, nicht von der kombinierte-multimodalen Variante.

Mangelnde Sensitivität des Tests ist als Erklärung für das Ergebnismuster unwahrscheinlich, da diese hinreichend belegt ist (Mattes, 2003). In jedem Fall wurden erfolg-

reich fahrtypische Ressourcen gebunden, wie sich an den Resultaten der Aufgabenbearbeitungszeiten und der Blickdaten gut erkennen lässt. Das spricht dagegen, dass die LCT zu einfach gewesen sein könnte. Die Erklärung kann eher in der Aufgabenpriorisierung gesehen werden und damit in einem klassischen Trade-off zwischen den Aufgaben. Der zügige Wechsel auf die vorgegebene Fahrspur mit stabiler Querführung auf dieser Spur war als Primäraufgabe vorgegeben. Offensichtlich haben sich die Versuchsteilnehmer bestmöglich an diese Anweisung gehalten und die Primäraufgabe geschützt. Vor diesem Hintergrund ist keine starke Variation der Fahrqualität zu erwarten, wohl aber ist mit den beobachteten deutlichen Auswirkungen auf die Sekundäraufgabe zu rechnen.

Präferenzurteile

Die Präferenzurteile spiegeln das Ergebnismuster der anderen abhängigen Variablen weitgehend wider. Die Touchpad-Varianten mit zusätzlichen Ausgaben wird gegenüber der unimodalen Variante der Vorzug gegeben. Dabei verteilen sich die Präferenzen nahezu gleichmäßig auf die verbleibenden Alternativen des auditiven, taktilen oder kombiniert-multimodalen Feedbacks sowie auf den Dreh-Drücksteller.

4.1.2.5 Schlussfolgerungen

Insgesamt kann das Experiment die Hypothese bestätigen, dass ein Touchpad zur zentralen Steuerung eines Bordsystems im Fahrzeug aufgrund der natürlicheren Abbildung der Eingabebewegung auf die Cursorbewegung besser geeignet ist als ein herkömmlicher Dreh-Drücksteller. Diese Aussage gilt jedoch nur, wenn neben einer visuellen Ausgabe die Nutzeraktionen durch mindestens eine zusätzliche Modalität rückgemeldet werden. Ohne zusätzliches Feedback geht der Vorteil des Touchpads während der Fahrt durch die notwendige visuelle Kontrolle der Eingabe verloren. Mit dem herkömmlichen Dreh-Drücksteller lassen sich dann genauso gute oder sogar bessere Resultate erzielen, wie mit den Direktvergleichen bezüglich Workload und Gesamtblickdauer klar belegt werden konnte. Nachteile dieser Touchpad-Variante und des Dreh-Drückstellers können durch fokussierte Anstrengungen weitgehend kompensiert werden, wenn nicht gleichzeitig eine Fahraufgabe zu erfüllen ist.

Bei dem Vergleich der Modalitäten zur Unterstützung der grafischen Benutzeroberfläche war erwartet worden, dass die kombiniert-multimodale Variante mit synchronisierter visuell, auditiv und taktil wahrnehmbarer Ausgabe stets zu den besten Resultaten führt. Das konnte nicht in dieser Form beobachtet werden. Im Hinblick auf die Bearbeitungszeit während der Fahrt gab es kaum Unterschiede zwischen den Varianten auditiv, taktil und auditiv plus taktil. Die Blickdaten weisen das zusätzliche auditive Feedback als beste Lösung aus, sowohl bei der Dauer pro Blick, der Blickhäufigkeit als auch entsprechend bei der Gesamtblickdauer. Die kombiniert-multimodale

Rückmeldung war hier entweder gleichauf mit der taktilen Variante oder etwas besser. Ein ähnliches Bild ergab sich bei der subjektiven Workload. Das auditive Feedback führte zu der geringsten berichteten Belastung, die taktile und die kombinierte Ausgabe waren dahinter etwa gleichauf. Schließlich wird das gute Abschneiden des auditiven Feedbacks durch die Fahrqualität während der Bearbeitung der Sekundäraufgabe unterstützt. Wenngleich es insgesamt kaum Unterschiede in der Fahrqualität gab, konnte sich die zusätzliche auditive Rückmeldung während dieser kritischen Phase der Doppelbelastung auszeichnen.

4.1.3 Experiment 2: Evaluation der Blindbedienbarkeit

Die Analyse der Blickdaten während der Fahrt in Experiment 1 gibt bereits einige Anhaltspunkte über die Blindbedienbarkeit des Systems mit den Schnittstellenvarianten. Allerdings lässt die Auswertung des Blickverhaltens primär Rückschlüsse über die Strategie des Time-Sharing zur Ressourcenverteilung zu. Über das Ausmaß der notwendigen Blickzuwendung zur Bedienung eines Systems gibt sie indirekt dadurch Auskunft, dass sie erkennen lässt, wie viel visuelle Kontrolle der Nutzer für notwendig erachtet. Es wird eine umfassende Analyse der Blindbedienbarkeit der Schnittstelle und damit der Beiträge der Ausgabemodalitäten angestrebt, die einen weiteren Untersuchungsschritt sinnvoll macht.

In Experiment 2 mussten die Versuchsteilnehmer den Selektionscursor von einer ihnen bekannten Startposition auf einem Regler ohne Blickkontakt auf eine definierte Zielposition manövrieren. Dazu wurde ihnen zunächst der Regler mit dem Cursor in der Startposition auf dem Display dargeboten, damit sie sich die visuelle Szene merken konnten. Anschließend wurde der Bildschirm verdunkelt, und die Teilnehmer mussten versuchen, das Zielelement auf dem Regler zu markieren. Dieses Experiment soll zeigen, mit welcher Güte die Operationen ohne jegliche visuelle Zuwendung ausgeführt werden können und mit welcher Schnittstellenvariante dies optimal gelöst wird.

4.1.3.1 Methode

Die Versuchsumgebung und die Schnittstellenvarianten sind identisch zu Experiment 1. Alle 20 Versuchspersonen aus Experiment 1 haben auch an Experiment 2 teilgenommen. Die Durchführung dieses zweiten Experiments erfolgte nicht separat, sondern war in das erste Experiment eingeflochten. Sobald die Teilnehmer in Experiment 1 die Aufgabenbearbeitung im Stand und in simulierter Fahrt beendet hatten, absolvierten sie mit derselben Schnittstellenvariante die Aufgaben zur Evaluation der Blindbedienbarkeit (siehe auch Abbildung 15, S.94). Dieses Design erschien notwendig, um den Probanden hinreichend gute Vorkenntnisse zur Bedienung der jeweiligen

Schnittstelle einzuräumen. Nur durch die ausgiebige Interaktionserfahrung durch wiederholte Eingaben konnte einem Floor-Effekt in der deutlich schwierigeren Aufgabe in Experiment 2 vorgebeugt werden.

Als Aufgaben kamen die horizontale Regleraufgabe mit neutraler Mittenmarkierung und die vertikale Regleraufgabe (A2 und A3) zum Einsatz. Von besonderem Interesse ist die Auswirkung der Mittenmarkierung für die non-visuelle Bedienung. Auf die zweidimensionale Aufgabe (A4) wurde verzichtet, weil diese aufgrund ihrer Komplexität kaum Aussicht auf Erfolg bei der Blindbedienung hatte. Die Navigation über mehrere Menüebenen (A1) war ebenfalls ohne visuelle Rückmeldung nicht möglich. Die Aufgabenstellung und -formulierung war mit Ausnahme der vollständigen Verdunklung (Maskierung) der Anzeige identisch zu Experiment 1. Zur Eingabe dienten dieselben vier Varianten des Touchpads (TPsolo, TPaud, TPtakt, TPaud+takt) und wie zuvor der Dreh-Drücksteller. Details zu den Ausgabevarianten sind in Abschnitt 4.1.2.1 (Methode) dargestellt. Die Reihenfolge der Aufgaben war ausbalanciert, die Reihenfolge der Schnittstellenvarianten wurde durch Experiment 1 vorgegeben.

Die Versuchsteilnehmer hatten 2.5 Sekunden Zeit, sich die aktuelle Cursorposition auf dem Display einzuprägen. Da die Elemente des Reglers visuell deutlich separiert waren, konnten sie leicht erkennen, wie viele Selektionsschritte sie bis zum Zielelement zurücklegen mussten. Während der Sichtbarkeit des Displays erfolgte auch die akustische Ansage des Zielelements. Danach wurde das Display durch eine einheitliche schwarze Fläche maskiert und ein Signalton markierte den Beginn der Aufgabe. Erst nach diesem Signalton konnte der Cursor bewegt werden. Die Teilnehmer versuchten dann, das gewünschte Element zu selektieren. Sobald sie sich der korrekten Position sicher waren, bestätigten sie die Eingabe durch Antippen des Touchpads bzw. Drücken des Dreh-Drückstellers. Vor Beginn der nächsten Aufgabe wurde der Cursor stets automatisch auf das korrekte Zielelement gesetzt. Dieser Vorgang blieb den Probanden verborgen. Sie erhielten keine Rückmeldung, ob ihre Eingabe richtig oder falsch war.

Als abhängige Variable werden in diesem Experiment ausschließlich die Fehlerraten erhoben. Sie geben Aufschluss darüber, ob es möglich ist, auch ohne Blickkontakt die gewünschte Einstellung vorzunehmen. Die Versuchsteilnehmer wurden vor den Aufgaben zu Experiment 2 darauf hingewiesen, dass in diesem Fall die Genauigkeit Vorrang hat. Sie wurden ferner aufgefordert, die Eingaben zwar zügig, aber ohne Zeitdruck abzuschließen.

Die Bedingung TPsolo weiterhin mitzuführen mag auf den ersten Blick wenig relevant erscheinen. Wenn das Touchpad an sich nur durch eine visuelle Ausgabe die Nutzeraktionen rückmeldet und just diese Ausgabe nicht zur Verfügung steht, wirkt eine schlechte Leistung über diese Schnittstellenvariante trivial. Weit weniger klar ist hingegen das Ausmaß der Leistungseinbuße. Das Experiment wurde so angelegt, dass

mit den jeweiligen Schnittstellenvarianten möglichst viel durch die Aufgaben in Experiment 1 geübt werden konnte. Um den Selektionscursor um eine Position zu verschieben, musste stets eine konstante Wegstrecke mit dem Finger zurückgelegt werden. Durch eine einzige Bewegung, die horizontal oder vertikal vom einen Rand des Touchpads zum anderen verläuft, konnte bei beiden Reglervarianten der Cursor vom ersten bis zum letzten Element bewegt werden. Es war also möglich, eine innere Repräsentation der notwendigen Bewegungslänge herauszubilden. So ist durchaus von Interesse, bis zu welchem Grad durch den Beitrag des motorischen Gedächtnisses die Blindbedienbarkeit selbst bei einem Touchpad ohne zusätzliche Rückmeldung erreicht werden kann.

4.1.3.2 Hypothesen

Durch die Maskierung des Displays in Experiment 2 werden die eigentlich bimodalen Varianten TPtakt und TPaud auf eine unimodale Rückmeldung reduziert, lediglich TPaud+takt bietet reduziert auf eine bimodale Schnittstelle mehrere Modalitäten. Ein Vergleich von Schnittstellen, die ohne visuelle Ausgabe auskommen müssen, ist in der Literatur zu Mensch-Maschine-Interaktion recht selten. Das erschwert die Hypothesenbildung. Die meisten Untersuchungen, die durch multimodale Ausgaben bessere Interaktionsergebnisse nachweisen konnten, beziehen sich auf einen Vergleich zwischen unimodaler visueller Ausgabe und bi- oder trimodalen Schnittstellen. Eine Ausnahme bildet zum Beispiel die Untersuchung von Vitense et al. (2003), die visuelles, auditives und haptisches Feedback einzeln und in allen Kombinationen experimentell analysierte. Die besten Ergebnisse erzielten sie, sofern nur eine Modalität zum Einsatz kam, mit visuellem oder mit haptischem Feedback, deutlich bessere Resultate jedoch mit bimodalen Ausgaben. Ein solcher Befund ist allerdings aufgabenabhängig und damit nicht generell übertragbar. Ausgehend von der Annahme, dass in sehr vielen Fällen Förderung durch reichhaltigere Ausgabe gezeigt werden konnte, lässt sich jedoch die Vermutung ableiten, dass das kombiniert auditiv-taktile Feedback zu den besten Resultaten führen dürfte. Ebenso ist offensichtlich, dass das Touchpad ohne zusätzliche Rückmeldung die größten Fehlerraten produzieren wird. Das Experiment soll auch zeigen, inwiefern die Probanden durch die vorangegangene Bedienung in Experiment 1 ein Gefühl für die Schnittstelle entwickeln konnten.

Weitgehend unklar ist, wie sich die Varianten auditive und taktile Rückmeldung zueinander verhalten. Das hängt hauptsächlich davon ab, ob die Kodierung der Ereignisse durch beide Modalitäten gleichermaßen erfassbar ist. In dieser Hinsicht ist das Experiment explorativer Natur und soll Hinweise auf die Eignung der Modalitäten liefern. Auch in Bezug auf den Vergleich zum Dreh-Drücksteller herrscht der explorative Charakter vor. Der Vorteil der besseren Abbildung könnte durch die Vorerfahrung der Nutzer in Experiment 1 reduziert sein. Inwiefern die Touchpad-Varianten zur

Blindbedienung besser einsetzbar sind als der Dreh-Drücksteller, hängt davon ab, wie gut die verschiedenen Rückmeldeformen ein Abzählen einzelner Auswahlsschritte unterstützen. Falls die Probanden die Klicks des Dreh-Drückstellers leichter mitzählen können, könnte dieser eine bessere Orientierung bieten.

Eine gerichtete a priori Hypothese betrifft den Unterschied zwischen den Aufgabentypen. Durch die Mittenmarkierung in der horizontalen Regleraufgabe wird in dieser Aufgabe mehr Information zur Orientierung über die zusätzlichen Ausgaben des Touchpads transportiert. Das sollte bei dieser Aufgabe zu einer verringerten Fehlerrate bei den Touchpad-Varianten mit auditiver, taktiler und kombinierter Rückmeldung führen.

4.1.3.3 Ergebnisse

Die Auswertung der Ergebnisse erfolgt analog zu Experiment 1. Im Text und in den Grafiken werden Mittelwerte und Standardfehler berichtet. Aufgrund des explorativen Charakters dieses Experiments wird im Gegensatz zum ersten Experiment auf geplante Vergleiche zwischen einzelnen Schnittstellenvarianten verzichtet.

Als Fehler war jede Bestätigungsoperation definiert, die nicht auf dem vorgegebenen Zielelement stattfand. Es wird der Anteil fehlerhafter Eingaben an allen vorgenommenen Eingaben angegeben. Zunächst soll die Aufgabe mit neutraler Mittenmarkierung betrachtet werden. Diese Ergebnisse sind im linken Teil der Abbildung 23 veranschaulicht.

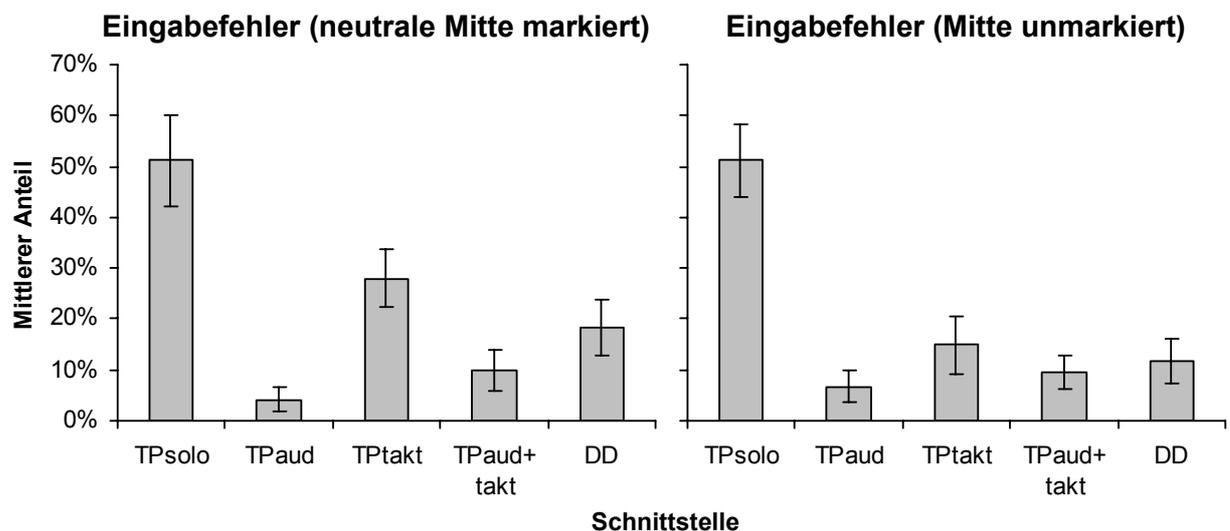


Abbildung 23: Prozentueller mittlerer Anteil fehlerhafter Eingaben an allen Eingaben. Links: Regleraufgabe mit neutraler Mittenmarkierung. Rechts: Regleraufgabe ohne spezielle Hervorhebung der Mitte.

Wie erwartet ergab sich bei TPsolo mit 51.3% ($SE = 8.93$) eine sehr hohe Fehlerrate. Der geringste Anteil an fehlerhaften Eingaben liegt bei TPaud mit 4.2%

($SE = 2.30$). TPtakt hingegen weist einen hohen Fehleranteil von 28.0% ($SE = 5.59$) auf. TPaud+takt ordnet sich mit 10.0% ($SE = 4.08$) dazwischen ein. Die Eingabe über den Dreh-Drücksteller zog eine Fehlerrate von 18.3% ($SE = 5.66$) nach sich. Eine einfaktorielle Varianzanalyse mit dem Messwiederholungsfaktor Schnittstelle ergab einen hoch signifikanten Effekt ($F(4,76) = 13.37, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .41$). Dass bei TPsolo eine sehr hohe Fehlerrate resultieren wird, war bereits im Vorfeld vermutet worden. Der signifikante Effekt ist jedoch auch dann nachweisbar, wenn man TPsolo von der Analyse ausschließt ($F(3,57) = 5.74, p < .01, p_{\text{rep}} = .98, \eta_p^2 = .23$).

Die Analyse der Aufgabe ohne Mittenmarkierung lieferte ebenfalls eine recht hohe Fehlerrate für TPsolo von 51.2% ($SE = 7.11$). Die Touchpad-Varianten mit zusätzlicher Rückmeldung waren sich bei dieser Aufgabe durchwegs ähnlicher; TPaud 6.7% ($SE = 3.06$), TPtakt 15.0% ($SE = 5.66$) und TPaud+takt 9.6% ($SE = 3.38$). Beim Dreh-Drücksteller wurde eine Fehlerrate von 11.7% ($SE = 4.38$) erzielt (vgl. Abbildung 23). Unter Einbeziehung aller Schnittstellen zeigt die Varianzanalyse einen hoch signifikanten Effekt ($F(4,76) = 14.71, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .44$). Dieser ist jedoch im Gegensatz zu dem Resultat bei der Regleraufgabe mit markierter neutraler Mitte nur von TPsolo getragen. Ohne TPsolo lassen sich keine statistisch bedeutsamen Unterschiede mehr nachweisen ($F(3,57) = 0.73, p = .54, p_{\text{rep}} = .47$).

Ein direkter Vergleich der beiden Aufgaben zeigt, dass die Leistung bei TPsolo und TPaud+takt sehr ähnlich ist. Eine vergleichsweise kleine Änderung ergab sich bei TPaud und DD, während TPtakt in der Aufgabe ohne neutrale Mittenmarkierung zu deutlich weniger Fehlern geführt hat. Um zu überprüfen, ob es eine statistisch belastbare Veränderung zwischen den Aufgabentypen gab, wurde eine zweifaktorielle Varianzanalyse mit den Messwiederholungsfaktoren Aufgabentyp und Schnittstelle berechnet. Weder der Haupteffekt für Aufgabentyp ($F(1,19) = 1.08, p = .31, p_{\text{rep}} = .64$) noch die Interaktion Aufgabentyp \times Schnittstelle ($F(4,76) = 0.89, p = .48, p_{\text{rep}} = .52$) konnten jedoch statistisch belegt werden. Der signifikante Einfluss der Schnittstelle war wiederum aufgrund der oben berichteten Resultate ohnehin zu erwarten ($F(4,76) = 25.13, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .57$).

4.1.3.4 Diskussion

Experiment 2 verfolgte einen vergleichsweise strengen Ansatz, um die Bedienbarkeit durch die Schnittstellenvarianten ohne visuelle Kontrolle zu prüfen. Nachdem den Probanden Zeit gegeben wurde, sich auf dem Display zu orientieren sowie die aktuelle und die Zielposition des Selektionscursors zu erfassen, wurde die Anzeige verdunkelt. Die Aufgabe bestand darin, vollständig ohne visuelle Unterstützung das Zielelement einzustellen. Bei Verfehlen dieses Zielelements spielte die Distanz keine Rolle, ein direkt an das Zielelement grenzendes benachbartes Element wurde gleichermaßen als

Fehler gewertet. Diese Auslegung ist der Bedeutung von Fehleingaben in der Praxis geschuldet: Sobald das falsche Element zum Beispiel in einem Menü angewählt wurde, wird eine aufmerksamkeitsfordernde Fehlerbehandlung notwendig. Dabei macht es keinen Unterschied, ob das korrekte Element knapp oder deutlich verpasst wird.

Die Versuchspersonen gaben übereinstimmend an, zur Bearbeitung die Anzahl der notwendigen Schritte während der visuellen Darbietungsphase mit Hilfe der numerischen Beschriftung der Elemente abgezählt zu haben, sodass sich die Aufgabe in der Art „drei Schritte nach oben“ oder „drei Schritte nach unten“ vereinfachte. Bei dem Touchpad ohne zusätzliches Feedback erhielten die Teilnehmer keine Ausgabe, die diese Zähloperation während der anschließenden Dunkelphase eindeutig unterstützte. Damit gibt diese Bedingung darüber Aufschluss, ob nach einiger Bedienerfahrung – hier die Aufgabendurchführung in Experiment 1 – durch den Beitrag des motorischen Gedächtnisses bereits ausreichend Wissen über die Abstandsverhältnisse vorliegt, um eine weitgehende Blindbedienbarkeit zu ermöglichen. Auf diesem Wege kann die Relevanz des Beitrags der multimodalen Ausgabe eingeordnet werden; handelt es sich lediglich um eine marginale Verbesserung der Bedienung ohne visuelle Kontrolle oder ist sie essentiell, um die Arbeit mit einer Touchpad-basierter Steuerung während der Fahrt zuzulassen?

Zusammengefasst sprechen die Befunde dafür, dass es zumindest nicht unmöglich ist, mit der Touchpad-Variante, die kein zusätzliches Feedback neben der visuellen Ausgabe bietet, auch ohne visuelle Kontrolle einfache Bedienvorgänge zu absolvieren. Die Fehlerrate von rund 51% in beiden Aufgabentypen ist zwar hoch, sie zeigt aber auch, dass ein großer Teil der Bedienschritte durchführbar war. Im gegenteiligen Fall wäre ein deutlich höherer Anteil an Fehlern zu beobachten gewesen. Aufgrund des angelegten strengen Bewertungskriteriums ist dies ein durchaus beachtliches Ergebnis. Darüber hinaus wurden in diesem Experiment teils relativ schwierige Bedienschritte forciert ohne Blickkontakt durchgeführt. Einfache Eingabeoperationen, die nur wenige Elemente überspringen, oder das Einstellen von Maximal- und Minimalwerten sollten folglich möglich sein. Die Eingabequalität bleibt insgesamt jedoch deutlich hinter dem zurück, was mit zusätzlicher akustischer und/oder haptischer Ausgabe möglich ist. Selbst die rein taktile Ausgabe, die unter den Varianten mit zusätzlicher Rückmeldung im ersten Experiment am schlechtesten abgeschnitten hat, brachte eine dramatische Reduktion der Fehlerrate mit sich. Unter diesen Varianten trägt die auditive Rückmeldung am besten zur Blindbedienbarkeit bei. Etwas überraschend ist nicht die kombiniert-multimodale Ausgabe wie vermutet klar im Vorteil, sondern ordnet sich bei beiden Aufgabentypen zwischen der auditiven und der taktilen Variante ein. Der herkömmliche Dreh-Drücksteller verfügt neben der hier nicht verfügbaren visuellen Rückmeldung über keine technisch vermittelte Ausgabe. Dennoch lässt sich damit ein vergleichbar gutes Resultat wie mit den Touchpads mit zusätzlicher Ausgabe erzielen.

Ein unerwarteter Befund soll abschließend noch diskutiert werden. Wenn die neutrale Mitte speziell markiert ist, schneidet das taktile Feedback auffällig schlecht ab. Das deutet darauf hin, dass die Nutzer bei taktilem Darbietung von dieser Zusatzmarkierung eher irritiert waren. Die Teilnehmer taten sich bei dem Touchpad mit taktilem Feedback ohnehin schwerer, die einzelnen Elemente der Regler gut anzusteuern, wie die allgemein höheren Fehlerraten bei taktilem Ausgabe in beiden Aufgabentypen zeigen. Gegeben, dass die Strategie bei der Blindbedienung in einem Abzählen der wahrgenommenen Ausgabeereignisse (Klick-Töne oder taktile Impulse) bestand, liegt die Annahme nahe, dass dieses Vorgehen durch die Gestaltung der taktilem Impulse nicht optimal unterstützt wurde. Ein Grund könnte sein, dass die für einzelne Impulse verwendete vollständige Sinus-Welle von 20 ms Dauer (50 Hz) nicht pointiert genug war. Gegenwärtig existieren keine Gestaltungsrichtlinien in der Literatur, die hier eine klare Antwort geben könnten. Einen Überblick über manipulierbare Parameter für Tactons geben beispielsweise Brown et al. (2005). Allerdings sind die Ergebnisse aufgrund des besonderen hier verwendeten Eingabegeräts nicht einfach übertragbar. Mehr Klarheit können nur weitere Experimente in dem noch jungen Feld des Tacton-Designs geben.

4.1.3.5 *Schlussfolgerungen*

Eingabeelemente für Bordsysteme sollten für eine Minimierung der visuellen Abwendung von der Fahraufgabe die Blindbedienbarkeit bestmöglich unterstützen. Aus dem vorliegenden Experiment kann gefolgert werden, dass sich außer der Touchpad-Variante ohne zusätzliche Ausgabemodalität alle getesteten Eingabeelemente ähnlich eigneten. In der Aufgabe ohne neutrale Mittenmarkierung waren keine auffälligen Unterschiede zu erkennen. Der Aufgabentyp mit neutraler Mittenmarkierung führte zwar zu Leistungsunterschieden, nicht jedoch zu dem erwarteten Ergebnismuster. Während das Touchpad mit auditivem Feedback zumindest etwas von dieser zusätzlichen Information profitieren konnte, verschlechterte sich die Fehlerrate bei der taktilem Variante. Gegenwärtig existiert noch zu wenig Wissen darüber, welche Gestaltungsparameter in welcher Weise verändert werden müssen, um sowohl von der Semantik als auch von der Wahrnehmbarkeit her derartige Ereignisse optimal zu kodieren. Der überraschend nachteilige Effekt bei der Mittenmarkierung zeigt, dass beim Einsatz von Tactons als neuer Ausgabeform in Mensch-Maschine-Schnittstellen eine systematische empirische Evaluation unerlässlich ist und mehr Experimente zu Gestaltungsparametern notwendig sind.

4.1.4 *Zusammenfassende Diskussion und Gestaltungsempfehlungen*

In zwei Experimenten wurde untersucht, ob durch redundante multimodale Ausgabe eine Touchpad-basierte manuelle Eingabeschnittstelle zur Steuerung von Komfort-

funktionen im Fahrzeug einen höheren Grad an Bedienbarkeit und Aufgabenangemessenheit erreichen kann. Zusätzlich zur visuellen Ausgabe wurde für relevante Interaktionsereignisse auditives, taktiles oder kombiniert auditives und taktiles Feedback implementiert. Diese multimodalen Ausgabevarianten wurden zum einen gegen eine unimodale Touchpad-Variante verglichen, die nur über visuelle Rückmeldung verfügte. Zum anderen fand der Vergleich gegen einen herkömmlichen Dreh-Drücksteller statt, dessen Aktionen ebenfalls auf dem Display ausgegeben wurden, der aber nicht über mechanisch vermitteltes haptisches oder auditives Feedback verfügte. Es zeigte sich, dass durch die Ausgabe über zusätzliche Modalitäten wesentliche Parameter der Bedienbarkeit eines solchen Systems für die Mehraufgaben-Situation im Fahrzeug positiv beeinflusst wurden. Dabei war zu beobachten, dass Multimodalität dann Vorteile entfaltet, wenn durch die gleichzeitige Belastung mit der Primäraufgabe Fahrzeugführung und der Sekundäraufgabe Bedienung Ressourcenkonflikte bei der Informationsverarbeitung entstehen. Diese Aussage bezieht sich sowohl auf die bimodalen Ausgabeformen (visuell und taktil bzw. visuell und auditiv) als auch auf die trimodale Rückmeldung (visuell, auditiv und taktil). Somit wurden die Voraussagen des multiplen Ressourcenmodells erfüllt (Wickens, 2002). Besonders aufschlussreich ist eine übergeordnete Betrachtung der sich ergebenden Datenmuster über die abhängigen Maße in beiden Experimenten hinweg. Nahezu durchgehend zeigt sich das gleiche Bild: Bei der Bedienung mit dem unimodalen Touchpad ohne zusätzliches Feedback ergibt sich die schlechteste Leistung bzw. die größte Belastung. Der Dreh-Drücksteller ist typischerweise etwas besser, aber auf einigen Maßen nur in unwesentlichem Ausmaß. Die Touchpad-Varianten mit zusätzlichem Feedback schneiden am besten ab, dabei ebenfalls fast durchgängig in folgender Rangfolge: Zusätzliches auditives Feedback hat die besten fördernden Auswirkungen, zusätzlich taktiles Feedback führt zur geringsten Förderung, das kombiniert-multimodale Touchpad ordnet sich stets dazwischen ein. Diese Befunde bedürfen einer weitergehenden Klärung.

Taktiler Feedback war der auditiven Ausgabe unterlegen. Wie bereits dargelegt ist schwer zu entscheiden, ob auditives Feedback generell für diese Art von Ausgabe besser geeignet war, oder ob das taktile Feedback nicht optimal gestaltet war. Die Art der Schwierigkeiten mit der Verwendbarkeit des taktilen Feedbacks durch die Probanden, wie sie insbesondere in Experiment 2 sichtbar geworden sind, spricht für ein Gestaltungsproblem. Obwohl durch separate Vortests versucht wurde, das Design der taktilen Ausgabe gut zu lösen, ist die Realisierung der Mittenmarkierung offensichtlich nicht gelungen. Ob auch für die Markierung der einzelnen Elemente eine bessere Gestaltungsform möglich gewesen wäre oder ob bereits die maximale Leistungsfähigkeit des taktilen Feedbacks erreicht war, kann nicht beantwortet werden. Dennoch sollen zumindest einige Anhaltspunkte zur Diskussion gestellt werden. Zum schlechteren Abschneiden der taktilen Erweiterung gegenüber der auditiven Erweiterung kann

beigetragen haben, dass die auditiven Stimuli grundsätzlich besser gestaltet waren, da für deren Design auf deutlich mehr Erfahrungswissen aus der Literatur zurückgegriffen werden konnte als bei taktilen Mustern. Des Weiteren könnten die Versuchsteilnehmer mehr Übung darin gehabt haben, auditive Systemausgaben als Hinweisreize zu verarbeiten. Auditive Systemausgaben sind weit verbreitet und bekannt und lassen sich deshalb unter Umständen besser in Handlungsschemata integrieren. Durchaus möglich ist ferner, dass die taktilen Impulse an den Fingerspitzen nicht optimal wahrnehmbar waren. Die Fingerspitze wurde während der Ausgabe des taktilen Reizes gleichzeitig bewegt. Der taktile Impuls sollte in Analogie zu der Empfindung stehen, die auftritt, wenn der Finger über eine Schwelle oder eine ähnliche Oberflächenmarkierung gleitet. Diese Analogie ist auf semantischer Ebene sicherlich tragfähig, auf physiologischer Ebene möglicherweise nicht. Im Gegensatz zum Spüren einer Erhebung beim untersuchenden Tasten heben sich die taktilen Markierungen nicht tatsächlich durch eine Änderung der Oberflächeneigenschaft hervor. Bei der Bewegung des Fingers über die Fläche signalisieren die schnell adaptierenden Mechanorezeptoren vom Typ II (FA II) und die langsam adaptierenden Mechanorezeptoren SA II eine laterale Verformung der Haut und transportieren auch die Richtungsinformation (für die physiologischen Grundlagen vgl. Beyer & Weiss, 2001 und Kapitel 2). Ferner ist bekannt, dass FA I und SA I Rezeptoren eine besondere Empfindlichkeit für scharfe Kanten aufweisen. Reize bei diesen Rezeptoren lösen eine stärkere Reaktion aus, wenn eine scharfe Kontur statt einer flächigen Belastung auf das rezeptive Feld wirkt. Während also bei realen Markierungen über die FA I und SA I Rezeptoren Konturen gefühlt werden, werden bei den hier verwendeten virtuellen Markierungen durch vibrotaktile Ausgaben primär die FA II Rezeptoren stimuliert. Die Art der Empfindung ist somit auf physiologischer Ebene unterschiedlich und deshalb möglicherweise nicht so prägnant.

Auch der Befund zu den unterschiedlichen Förderungseffekten multimodaler Ausgabe bei den Touchpad-Varianten soll abschließend diskutiert werden. Zunächst kann festgehalten werden, dass die bi- und trimodalen Schnittstellen zu der typischen und zu erwartenden Förderung der Interaktionsqualität gegenüber unimodalem Feedback geführt haben. Entgegen der Hypothesen des Experiments konnten bei trimodaler Ausgabe keine besseren Resultate erzielt werden als bei den bimodalen Varianten. Die Vermutung liegt zwar nahe, dass keine zusätzliche Förderung sichtbar war, weil es sich schließlich bei allen Varianten bereits um multimodales – mindestens bimodales – Feedback handelt. Einige Befunde deuten aber darauf hin, dass beispielsweise die Reaktionszeiten bei der Informationsverarbeitung durchaus von trimodalem Feedback zusätzlich profitieren (Diederich & Colonius, 2004) und so schnellere Eingaben zu erreichen sein sollten. Stattdessen konnte durch eine Kombination aller Feedback-Modalitäten keine zusätzliche Verbesserung erreicht werden. Cockburn und Brewster (2005) fanden bei einer Untersuchung, bei der gleichfalls ballistisches Zielverhalten

im Vordergrund stand, ein ähnliches Resultat. Ebenfalls entgegen ihrer Vermutungen brachte die Kombination von auditivem und taktilem Feedback keinen Vorteil, während auditiv bzw. taktil allein als Ergänzung der visuellen Ausgabe durchaus von Vorteil waren. Wenngleich durch auditives und taktiles Feedback die Zeiten um 4.2% und 3.5% gegenüber rein visueller Rückmeldung verbessert wurden, konnte die Kombination der beiden zusätzlichen Ausgabemodalitäten nur eine deutlich geringere Verbesserung von 1.7% erbringen. Cockburn und Brewster vermuteten, dass durch zu viel redundante und damit überflüssige Information die Nutzer überlastet worden sind. Diese Erklärung ist für das vorliegende Experiment ebenfalls denkbar. Da jedoch keine Verschlechterung gegenüber beiden bimodalen Ausgaben eintrat, sondern sich die trimodale Variante meist zwischen der schlechteren taktilen und der besseren auditiven Eingabe eingeordnet hat, kann eine einfachere Erklärung herangezogen werden. Nachdem die zusätzliche taktile Ausgabe an sich bei weitem nicht so leistungsfähig in der Förderung war wie die auditive Variante, verringerte sie auch den Förderungseffekt der kombiniert-multimodalen Ausgabe. Anders ausgedrückt: Superadditive Effekte sind von trimodaler Ausgabe eher zu erwarten, wenn die bimodalen Ausgaben vergleichsweise ähnlich in ihrem Beitrag sind. Unterscheiden sie sich so deutlich wie hier, verringert sich die Leistungsfähigkeit der kombinierten Form durch das schwächere Glied in der Kette. Weitere Klärung kann ein Aspekt aus der Grundlagenforschung einbringen, der als *inverse effectiveness rule* bekannt ist (Stein & Meredith, 1993): Eine Verbesserung durch multimodale Ausgabe (*cross-modal enhancement*) erreicht ein umso höheres Ausmaß, je schwächer bei einem bimodalen Stimulus die Wahrnehmbarkeit der einzelnen Modalitäten ist. Die Förderung ist relativ gesehen maximal, wenn beide Modalitäten allein kaum die Wahrnehmungsschwelle erreichen. Diese Regel basiert auf einer eindimensionalen Verbesserung der Reaktionszeiten und ist nicht für komplexe Bediensituationen erstellt worden. Dennoch lässt sich daraus ableiten, dass durch die ungehinderte Wahrnehmbarkeit beider Modalitäten in den vorliegenden Experimenten kaum eine massive Förderung resultieren kann.

Was kann aus den Ergebnissen der Experimente im Hinblick auf den Einsatz multimodalen Feedbacks im Allgemeinen und einer Touchpad-basierten Steuerung von Komfortfunktionen im Fahrzeug im Besonderen gefolgert werden? Bei der Beurteilung der Eignung muss zuvorderst ein Aspekt beachtet werden, auf den Vitense und Kollegen (2003) bei der Interpretation ihrer Untersuchungen zu multimodalem Feedback nachdrücklich hinweisen: Die Wirksamkeit einer Kombination von Modalitäten ist inhärent aufgaben- und nutzerabhängig. Die hier verwendeten zusätzlichen nichtvisuellen Rückmeldemodalitäten zur Unterstützung der Interaktion mit einem Touchpad für eine Nebenaufgabe der Fahrzeugführung können grundsätzlich als aufgabentauglich angesehen werden. Bei der Beurteilung des Beitrags der einzelnen Modalitäten und der Kombinationen daraus muss die vereinfachte Laborsituation im Auge behalten

werden. In diesem Setting konnte die kombiniert-multimodale Variante ihre Vorteile nicht voll ausspielen und erwies sich nicht als beste Lösung. Das lag mitunter daran, dass die Systemausgaben über alle Sinneskanäle in vollem Umfang wahrgenommen werden konnten. In der realen Anwendungssituation im Fahrzeug hingegen existieren Störquellen, die diese Wahrnehmung gegebenenfalls beeinträchtigen. Nutzer hören oftmals Musik während der Fahrt oder es findet eine Unterhaltung zwischen den Fahrzeuginsassen statt. Das kann die Wahrnehmbarkeit einer auditiven Ausgabe herabsetzen. Eine unterschiedliche Beschaffenheit der Fahrbahnoberfläche hingegen kann zu Vibrationen im Fahrzeuginnenraum führen und damit die Wahrnehmbarkeit einer taktilen Ausgabe vermindern. Nachdem sich in diesem Experiment alle Touchpad-Varianten mit zusätzlichen Modalitäten gegenüber der unimodalen Variante und dem herkömmlichen Dreh-Drücksteller als Orientierungspunkt bewährt haben, wird unter dem erweiterten Gesichtspunkt der Anwendungssituation die Verwendung des kombiniert-multimodalen Touchpads empfohlen. Berücksichtigt man die angesprochene Nutzerabhängigkeit einer Wirksamkeit der Rückmeldungen, müssen die Präferenzurteile aus dem Experiment beachtet werden. Eine Mehrzahl der Nutzer hat sich verglichen mit dem Dreh-Drücksteller für eine Touchpad-Variante ausgesprochen. Dabei verteilt sich die Bevorzugung auf die drei Varianten mit zusätzlichem Feedback etwa gleich, das Touchpad ohne zusätzliches Feedback hingegen wurde abgelehnt.

Basierend auf den Vortests und den Experimenten lassen sich folgende Empfehlungen für die Gestaltung von Feedback zur Entlastung des visuellen Kanals bei Touchpad-basierter Eingabe geben:

- Die erwarteten Vorteile einer Touchpad-basierten Steuerung gegenüber einem herkömmlichen Dreh-Drücksteller (z.B. natürliche Abbildung von Eingabebewegungen und zweidimensionale Bewegungen) kommen in Dual-Task Situation nur zum Tragen, wenn die damit durchgeführten Aktionen durch eine zusätzliche Ausgabemodalität rückgemeldet werden. Unter diesen Umständen wirkt sich das Touchpad besonders vorteilhaft bei Eingabevorgängen aus, die eine Navigation auf zweidimensionalen Steuerelementen erfordern.
- Wenn mit dem Ziel einer personalisierbaren Schnittstelle dem Nutzer die Möglichkeit eingeräumt wird, Ausgabeformen zu deaktivieren, muss eine sicherheitsbewusste Bedienbarkeit gesichert bleiben. Entsprechend sollte die Konfiguration so eingerichtet werden, dass obligatorisch eine nicht-visuelle (auditive oder taktile) Rückmeldung über Eingabeaktionen aktiviert ist. Diese Maßnahme wird voraussichtlich auf Akzeptanz stoßen, da die Nutzer in ihren Präferenzurteilen ein Touchpad mit nur visuellem Feedback ablehnten.
- Kombiniert-multimodales Feedback hat das Potential, selbst bei einer Vielzahl von Störeinflüssen durch die Umwelt (Lärm, Vibrationen) wirksam zu sein.

- Während taktiles Feedback durch die Applikation auf die Fingerspitze nur direkt vom Nutzer wahrgenommen werden kann, werden Audiosignale von allen Fahrzeuginsassen vernommen und können unter Umständen als störend empfunden werden.
- Bei redundanter bi- oder trimodaler Ausgabe muss auf eine gute Synchronisierung der Signale geachtet werden, die intermodale Verzögerung sollte so kurz wie möglich gehalten werden. Als Richtwert sollte der zeitliche Abstand 50 Millisekunden nicht übersteigen (vgl. Card et al., 1986), um die Wahrnehmung als ein einzelnes Ereignis zu gewährleisten. Für eine perfekte Synchronisierung, die unabhängig davon ist, welches Signal zuerst ausgegeben wird, empfehlen Hempel und Altinsoy (2005) eine maximale Verzögerung von 25 Millisekunden.
- Neben den Tests zur Gebrauchstauglichkeit in frühen Entwicklungsphasen muss die Semantik haptischer und akustischer Ausgaben vor einem Einsatz gesondert auf Verständlichkeit geprüft werden. Die Ergebnisse zur Markierung der neutralen Mitte insbesondere in Experiment 2 haben gezeigt, dass durch eine nicht optimale Repräsentation eines Systemereignisses negativere Konsequenzen für die Qualität der Interaktion resultieren können als durch eine fehlende Repräsentation dieses Ereignisses.

Bei den Vergleichen zwischen Touchpad und Dreh-Drücksteller kam in den Experimenten als Dreh-Drücksteller stets ein herkömmliches Bauteil zum Gebrauch, das neben der visuellen Ausgabe nur über ein nicht beeinflussbares mechanisches Rückmeldung verfügte. Als Anstoß für weitere Untersuchungen auf dem Feld der Multimodal-Forschung wäre ein Vergleich verschiedener Rückmeldevarianten des Touchpads mit einem Drehwähler hochgradig interessant, der wie beispielsweise der iDrive Controller von BMW über haptische Ausgabe (Force Feedback) verfügt (Ashley, 2001).

Das Ziel dieser Studie mit einem Touchpads zur Steuerung von Komfortfunktionen im Fahrzeug war nicht die Realisierung eines produktreifen Prototyps. Der vorrangige Inhalt war ein erweitertes Verständnis des Beitrags der auditiven und taktilen Modalität als Ergänzung der visuellen Modalität im direkten Vergleich zu erreichen. Bevor ein vorgeschlagenes multimodales Touchpad tatsächlich zum Einsatz kommen kann, sind noch weitere Untersuchungen über die Gestaltung der Ausgabe notwendig. Neben der bereits erwähnten Problematik einer vibrotaktilen Markierung von speziellen Schnittstellenkomponenten wie einer neutralen Mitte wird insbesondere Wissen darüber notwendig, wie viele und welche Ereignisse sinnvoll kodiert werden können. Bei den gegenüber Tactons bekannteren Earcons ist diese Diskussion bereits lange im Gange, für Tactons hat sie gerade erst begonnen. In diesem Sinne sieht sich diese Arbeit als einer der ersten Beiträge, die nachgewiesen haben, dass für Selektionsereig-

nisse im beschriebenen Anwendungskontext Tactons zu einer Verbesserung der Interaktionsqualität beitragen können.

4.2 Sprachliche vs. nicht-sprachliche akustische Ausgaben

4.2.1 Problemstellung und Ausgangspunkt des Experiments

Die Erweiterung einer einfachen grafischen Anzeige um akustische Ausgaben ist unter Berücksichtigung gegenwärtig verfügbarer Hardware der einfachste Weg, eine multimodale Schnittstelle zu realisieren. Auditiv wahrnehmbare Systemausgaben können maßgeblich dazu beitragen, die visuelle Last des Nutzers zu reduzieren. Neben den im vorigen Abschnitt (4.1) besprochenen direkten Rückmeldungen einer Eingabeoperation kann das System über akustische Ereignisse auch von sich aus auf Systemzustände aufmerksam machen. Der aufmerksamkeitslenkende und omnidirektionale Charakter akustischer Ausgaben gestattet dabei die Übermittlung von Information an den Nutzer auch dann, wenn dessen visuelle Aufmerksamkeit anderweitig konzentriert ist. Auf diese Eigenschaft wird besonders bei der Gestaltung von Warnsignalen zurückgegriffen, wie sie auch zum Einsatz in Bordsystemen von Fahrzeugen diskutiert werden (Baber, 1994; Dingus, McGehee, Manakkal & Jahns, 1997; Edworthy & Adams, 1996; Hirst & Graham, 1997; Roßmeier, Grabsch & Rimini-Döring, 2005). Aufgrund der zunehmenden Anzahl sekundärer und tertiärer Aufgaben im Fahrzeug werden neben Warnungen vermehrt auch nicht-kritische Systeminformationen auditiv vermittelt. Durch die Einführung von Fahrerassistenz- und Fahrerinformationssystemen sowie der Integration moderner Kommunikationslösungen und mobiler Geräte in das Bordsystem ist der Bedarf an zusätzlichen Ausgabeformen gestiegen. Würden diese Komponenten regelmäßig visuell überprüft, wäre die Blickabwendungszeit von der Fahrbahn schnell auf einem kritischen Niveau.

Im Gegensatz zu Warnungen ist bei Informationsausgaben die Dringlichkeit geringer. Wenngleich es für den Nutzer in seiner Handlungssituation hilfreich sein kann, über das entsprechende Ereignis informiert zu werden, kann diese Ausgabe zum gegenwärtigen Zeitpunkt ebenso irrelevant oder sogar störend sein. Zimmer (2001b) betont den Unterschied zwischen *Ablenkung* und *Abwendung*. Während bei einer Ablenkung ein externes Signal die Reaktion bei dem Fahrer auslöst, versteht man unter Abwendung die bewusste Entscheidung die Aufmerksamkeit für einen kurzen Moment von der Verkehrssituation zu nehmen, um z.B. ein Gerät zu bedienen. Gerade akustische Reize können nach Zimmer aufgrund einer starren Reiz-Reaktions-Kopplung (insbesondere bei Fahranfängern) ablenkend wirken. Durch das Aktivieren fester

Handlungsmuster haben ablenkende externe Reize das Potential, kognitive Aktivitäten zu unterbrechen und so bei Entscheidungs- oder Problemlöseprozessen zu stören. Nicht-kritische Informationen sollten deshalb derart ausgegeben werden, dass der Bedarf nach Eingriff flexibel beurteilt werden kann. Die Gestaltung auditiver Informationsmeldungen als Komponente eines multimodalen Ausgabekonzepts in Mehraufgabensituationen steht damit unter einer besonderen Fragestellung: Wie können akustische Ausgaben so gestaltet werden, dass sie möglichst wenig stören? Sind sprachliche oder nicht-sprachliche Ausgaben diesbezüglich zu bevorzugen? Diesem Problem widmet sich das vorliegende Experiment.

Der Zweck der Untersuchung ist ein verbessertes Verständnis über die Wirkung typischer Gestaltungsvarianten auditiv dargebotener Informationsausgaben zu erlangen. Vier Gruppen von Schallen (zwei sprachliche, zwei nicht-sprachliche) wurden hinsichtlich ihres Interferenzpotentials mit Arbeitsgedächtnisprozessen und Unterschieden in der Erkennbarkeit evaluiert. Mit einem angepassten Paradigma der psychoakustischen Grundlagenforschung soll verallgemeinerbares Gestaltungswissen gewonnen werden, das auch für den Anwendungsfall Fahrzeugführung relevant ist.

4.2.1.1 Arbeitsgedächtnis und Fahrzeugführung

Die Fokussierung auf das Arbeitsgedächtnis erfolgt im Hinblick auf dessen prominente Rolle bei der Informationsverarbeitung und dem Aufrechterhalten eines aktuellen Situationsmodells (vgl. Wickens & Hollands, 2000, Kapitel 7) sowie der Modulation der Aufmerksamkeit (Baddeley, 2003). Die Bedeutung des Arbeitsgedächtnisses für die Mensch-Maschine-Interaktion wird auch dadurch klar, dass Kurzzeitgedächtnisleistungen, Aufmerksamkeit und Bewusstsein miteinander verwobene Konstrukte sind und dass *working memory* unter einigen Gesichtspunkten besser als *working attention* konzeptualisierbar wäre (für eine tiefere Darstellung dieser Zusammenhänge vgl. Shah & Miyake, 1999).

Nach Baddeley und Hitch (1974) besteht das Arbeitsgedächtnis aus drei Komponenten (siehe auch Baddeley, 1986). Den Kern des Systems bildet die zentrale Exekutive (*central executive*), die die Arbeitsgedächtnisaktivität kontrolliert und den Speichermodulen limitierte Aufmerksamkeitsressourcen und Verarbeitungsprioritäten zuordnet. Der zentralen Exekutive unterstehen zwei Subsysteme zur Speicherung von Information. Eine, die phonologische Schleife (*phonological loop*), ist mit der Verarbeitung von akustischer und verbaler Information betraut. Die andere, der visuell-räumliche Notizblock (*visuo-spatial sketchpad*), erfüllt seinem Namen gemäß dieselbe Funktion für visuelle und räumliche Information. Die Subsysteme speichern die Informationen in unterschiedlichen Codes und arbeiten unabhängig, d.h. die Arbeit mit Information in einem Subsystem wird nicht durch die Aktivität des anderen Subsystems beeinflusst. In der Erweiterung des Mehrkomponentenmodells wird ferner davon

ausgegangen, dass durch einen episodischen Puffer (*episodic buffer*) Wissen aus dem Langzeitgedächtnis integriert wird (Baddeley, 2000a, 2002).

Im Kontext akustischer Ausgaben liegt das Interesse primär auf der phonologischen Schleife. Darin gespeicherte Informationen verfallen innerhalb einer Periode von ungefähr zwei Sekunden, wenn sie nicht durch einen aktiven Prozess aufrechterhalten werden. Dieser ressourcenfordernde Prozess hat eine begrenzte Kapazität, nur eine gewisse Anzahl von Items kann gleichzeitig in der Schleife gehalten werden. Kommt neue auditive Information hinzu, kann das dazu führen, dass alte Items nicht mehr behalten werden. Da akustische Ausgaben aufmerksamkeitsfordernd sind, können sie unter Umständen auch ohne Intention des Nutzers Zugang zur phonologischen Schleife erhalten und bei ausgelastetem Arbeitsgedächtnis alte Items verdrängen. Es kann damit zu Informationsverlust kommen.

Wie beeinflusst ein solcher Informationsverlust die Fahraufgabe? Eine Antwort geben die Modelle der Fahrzeugführung, die in Abschnitt 3.1 im Detail dargelegt sind. Die Aktivitäten im Fahrzeug lassen sich kurz gesagt in Kontrolltätigkeiten auf der einen Seite und in Entscheidungs- und Problemlösetätigkeiten auf der anderen Seite einordnen (Johannsen, 1990). Kontrolltätigkeiten wie Greifen, Schalten, manuelle und überwachend-eingreifende Regelung – bei geübten Fahrern weitgehend automatisierte motorische Tätigkeiten – sind von einer Belastung des Arbeitsgedächtnisses nicht betroffen. Ganz anders hingegen Prozesse des Entscheidens und Problemlösens wie Fehlermanagement und Planung. Auf die Ebenen der Fahrzeugführung bezogen ist die kognitive Beanspruchung auf der Navigationsebene maximal und nimmt von der Ebene der Führung auf die Stabilisierung stetig ab (Reichart & Haller, 1995). Diejenigen Prozesse, die unter bewusster kognitiver Kontrolle ablaufen, erfordern Aufmerksamkeit und sind durch Ressourcen eingeschränkt (Dahmen-Zimmer & Zimmer, 1997). Genau hier führen Interferenzen im Arbeitsgedächtnis zu suboptimaler Aufgabebearbeitung. Navigationsinformationen zu unbekanntem Routen können zum Beispiel im Arbeitsgedächtnis in dem Format „in ein paar hundert Metern rechts in die Ludwigstraße einbiegen, dann an der Ampel links und anschließend lange geradeaus und an der Brücke wieder rechts“ aufrechterhalten werden. In diesem Beispiel wird deutlich, dass neben den Items auch die Reihenfolge korrekt gespeichert werden muss. Bei unbekanntem Strecken ist eine bildliche Speicherung dieser Inhalte nicht möglich, die Abfolge der Items muss verbal memoriert werden. Zusätzlich werden kontinuierlich verkehrsrelevante Informationen aufgenommen, bei der Regulation des Verhaltens verwendet und anschließend wieder fallengelassen. In die Entscheidung über das Einleiten eines Überholvorgangs auf der Landstraße geht so Wissen darüber ein, ob ein Hintermann im Rückspiegel bereits zum Überholen angesetzt hat, ob der Abstand zu entgegenkommenden Fahrzeugen ausreicht oder genügend Sicht besteht, ob die Stre-

cke eine zügige Beschleunigung zulässt, wie groß die gegenwärtige eigene Geschwindigkeit ist und wie groß die des Vordermannes.

Es ist möglich, dass in Abhängigkeit von der Strategie bei der Enkodierung einige dieser Informationen bildlich memoriert werden und nicht in der phonologischen Schleife aufrechterhalten werden. In diesen Fällen ist keine oder deutlich weniger Interferenz mit auditiven Ausgaben zu erwarten. Die genannten Beispiele illustrieren dennoch die Problemsituation. Ziel des vorliegenden Experiments ist eine grundsätzliche Untersuchung der Störwirkung akustischer Ausgaben auf das Arbeitsgedächtnis als Beitrag zum Design der auditiven Komponente für multimodale Mensch-Maschine-Interaktion.

4.2.1.2 Akustische Ausgaben und das Arbeitsgedächtnis

Bei der Gestaltung für den auditiven Sinneskanal steht man zunächst vor einer grundlegenden Frage: Sollen sprachliche oder nicht-sprachliche Schalle eingesetzt werden? Grundlegende Fragen zur Nützlichkeit und Einsetzbarkeit dieser Gestaltungsvarianten wurden zwar bereits mehrfach in entsprechenden Handbüchern zu Mensch-Maschine-Interaktion thematisiert (z.B. Brewster, 2003; Dix et al., 2004). Wie Fröhlich und Pucher (2005) jedoch anlässlich einer Begutachtung der empirischen Literatur zur Einsetzung einer Arbeitsgruppe auf der internationalen Fachtagung für auditive Displays (ICAD) feststellten, existieren auffällig wenige formale und vergleichende experimentelle Studien.

Zur Präsentation von Information für den auditiven Sinneskanal stehen sowohl auf sprachlicher als auch auf nicht-sprachlicher Seite verschiedene Gestaltungsvarianten zur Verfügung. Als wesentliche Alternativen für Sprachausgabe können Sätze oder Schlüsselwörter genannt werden (National Research Council, 1997), z.B. „Sie haben eine neue Nachricht auf Ihrem Anrufbeantworter“ vs. „Neue Nachricht“. Dabei fallen elliptische Sätze meist in die erste Kategorie. Für die nichtverbale akustische Informationsdarbietung existieren seit Ende der 1980er Jahre zwei Alternativen: *Earcons* und *Auditory Icons*. Earcons sind abstrakte, nicht-sprachliche auditive Botschaften, die aus musikalischen Tonfolgen in strukturierter Kombination bestehen (Blattner, Sumikawa & Greenberg, 1989). Sie können kurz als melodieartige Sequenzen beschrieben werden. Die hierzu verwendeten Töne haben zumeist künstlichen oder synthetischen Charakter oder es kommen tonale Klänge von Musikinstrumenten zum Einsatz (vgl. Hempel & Altinsoy, 2005). Typischerweise besteht zwischen einem Earcon und dem bezeichneten Objekt oder Ereignis keine natürliche, intuitive Relation. Die Zuordnung ist arbiträr und muss deshalb gelernt werden. Dennoch konnte mehrfach gezeigt werden, dass sie ein wirksamer Weg sind, auditive Rückmeldungen zu geben (z.B. Brewster, 1998c; Brewster, Wright & Edwards, 1993, 1994). Im Gegensatz zu Earcons sind Auditory Icons natürliche Alltagsgeräusche, die in der Umwelt aufgenommen

oder synthetisiert wurden und durch Analogie auf die Modellwelt des Computers abgebildet werden (Gaver, 1989). So kann beispielsweise das Geräusch, das beim Zerknüllen von Papier entsteht, das Löschen einer Datei kodieren. Im direkten Vergleich weisen Auditory Icons den Vorteil auf, dass sie aufgrund der sinntragenden semantischen Relation sehr leicht erlernbar sind. Gleichzeitig tragen sie den Nachteil, dass es praktisch unmöglich ist, für abstrakte Interaktionsvorgänge oder -objekte ein repräsentatives natürliches Geräusch zu finden. Earcons hingegen gestatten problemlos eine beliebige Erweiterung der Schemata.

Diese Gestaltungsvarianten wurden bisher nicht direkt bezüglich ihrer Ablenkungswirkung miteinander verglichen. Ein Befund der psychoakustischen Grundlagenforschung kann jedoch möglicherweise dazu beitragen, das Verhältnis zwischen sprachlichen und nicht-sprachlichen Schallen im Allgemeinen zu erhellen. In der Diskussion von Vor- und Nachteilen sprachlicher und nicht-sprachlicher Ausgaben spricht Brewster (2003) den so genannten *Unattended Speech Effect* – möglicherweise besser bekannt unter dem Namen *Irrelevant Speech Effect* (ISE) – an und führt ihn als Argument für die Nutzung nicht-sprachlicher Schalle an. Der Effekt verweist auf das Phänomen, dass die Kurzzeitgedächtnisleistung durch Sprache im Hintergrund verschlechtert wird – obwohl den Versuchspersonen die Irrelevanz der Hintergrundsprache für die Gedächtnisaufgabe mitgeteilt wird und sie die Anweisung haben, diese zu ignorieren (Colle & Welsh, 1976; Salamé & Baddeley, 1982). Diese experimentelle Situation ist zwar nicht deckungsgleich mit der geschilderten Problematik der auditiven Darbietung nicht-kritischer Information, trägt aber gemeinsame Züge. Der Nutzer sollte beim Hören der Information in der Lage sein, schnell zu erkennen, ob sie für seine Aufgabe gegenwärtig relevant ist oder nicht. Ist sie es nicht oder befindet sich der Nutzer gerade in einer Situation hoher Belastung, sollte er fähig sein, die Ausgabe ohne maßgeblichen Verlust der Aufmerksamkeit auf die Erstaufgabe zu ignorieren. Der ISE deutet darauf hin, dass dies bei sprachlichen Systemausgaben nicht ohne weiteres gelingen könnte.

Wenn, wie Brewster (2003) argumentiert, tatsächlich nur Sprache diesen störenden Effekt auslösen würde und nicht-sprachliche Signale nicht, wäre das ein entscheidendes Argument zur Verwendung nicht-sprachlicher Ausgaben. Experimentelle Befunde zeigen, dass dies nicht der Fall ist. Zunächst steht im Raum, dass die Größe des Effekts vom semantischen Inhalt der Sprache kaum beeinflusst wird. Einsprachig aufgewachsene englischsprachige Versuchspersonen wurden in gleichem Ausmaß von Englisch, Walisisch oder sogar rückwärts gesprochener Sprache gestört (Jones, Miles & Page, 1990). Offensichtlich ist ein aktives Verstehen und Verarbeiten der Hintergrundsprache nicht für den Effekt notwendig. Ferner konnten Salamé und Baddeley (1982) nachweisen, dass nicht beachtete vokale Musik in gleichem Ausmaß, instrumentale Musik in einem etwas schwächeren Ausmaß wie Hintergrundsprache stört. Das weist

bereits darauf hin, dass nicht-sprachliche Schalle ebenfalls einen nachteiligen Effekt auf das Arbeitsgedächtnis haben können. Insbesondere durch Experimente von Jones und Kollegen verdichten sich die Hinweise darauf, dass nicht-sprachliche Schalle das gleiche Störpotential aufweisen (Jones & Macken, 1993; Jones, Madden & Miles, 1992; Jones & Morris, 1992). Sie werfen die Überlegung auf, dass die Beschaffenheit der akustischen Stimuli ausschlaggebend sei, da ein gleichförmiger Hintergrundton oder weißes Rauschen ebenso wie eine monotone Wiederholung von Silben keine Interferenz verursacht. Mit Hilfe frequenzmodulierter Sinustöne und ähnlichem Reizmaterial hingegen konnte mehrfach ein Effekt erzielt werden und es wurde die Hypothese aufgestellt, dass die Störwirkung von Tönen und Sprache bei entsprechender Beschaffenheit der nicht-sprachlichen Schalle gleich ist (Jones & Macken, 1993). Diese hat sich bisher gut bewährt und die notwendigen Bedingungen zum Erzeugen eines gleichstarken *Irrelevant Sound Effect* konnten zusammengefasst werden (Banbury, Macken, Tremblay & Jones, 2001): Der hauptsächliche Faktor ist nicht Sprache an sich, sondern eine Veränderung des akustischen Signals über die Zeit. Diese Veränderung kann sich durch Wechsel in Tempo, Tonhöhe oder Klangfarbe manifestieren. Wenn nicht-sprachliche Schalle in ihrer akustischen Variation entsprechend angepasst werden, sind die Effekte identisch zu Sprache. Nur eine Veränderung im Schalldruckpegel oder ein einfache Wiederholung von Schallmustern führt nicht zu einer Interferenz. Die Lautstärke der Ausgabe spielt dabei keine Rolle, auch bei relativ geringem Schalldruckpegel ließ sich der ISE erzeugen (Ellermeier & Hellbrück, 1998). Sehr laute gleichförmige Störgeräusche wie weißes Rauschen bei höheren Schalldruckpegeln (>85 dB, vgl. Broadbent, 1981) und Umweltlärm können zwar auch zu einer Leistungsminderung führen, sind aber für die gegenwärtige Untersuchung nicht von Interesse. Zur Einordnung der Befunde ist zu erwähnen, dass der ISE bisher nur bei solchen Aufgaben zuverlässig nachweisbar war, die Reihenfolgeinformation erfordern (vgl. Jones, Beaman & Macken, 1996). Doch gerade diese Komponente ist für mentale Aufgaben im Alltag sehr wichtig (Gathercole & Baddeley, 1993) und kann wie oben ausgeführt auch auf der Navigationsebene der Fahrzeugführung durchaus relevant sein.

Zusammengefasst gibt es überzeugende Hinweise darauf, dass der *Irrelevant Speech Effect* nicht auf Sprache begrenzt ist (mit der Abkürzung ISE wird deshalb sowohl auf *Speech* als auch auf *Sound* Bezug genommen). Unklar ist hingegen, inwiefern die genannten typischen Gestaltungsvarianten für die auditive Schnittstelle einen störenden Einfluss auf das Arbeitsgedächtnis aufbauen können. Diesen Zusammenhang soll das vorliegende Experiment klären.

4.2.2 Experiment 3: Wirkung sprachlicher und nicht-sprachlicher auditiv dargebotener Information auf das Arbeitsgedächtnis

Das Standardparadigma des ISE ist ein guter Ausgangspunkt für einen experimentellen Aufbau zur Klärung der Störwirkung sprachlicher und nicht-sprachlicher akustischer Systemmeldungen auf Arbeitsgedächtnisprozesse. Nicht-kritische Informationsausgaben sind keine „irrelevanten“ Schalle, sollten aber weitgehend unauffällig im Hintergrund erfolgen und nicht mit Planungsprozessen der Hauptaufgabe interferieren. Wie bereits ausgeführt, können sie abhängig von der konkreten Handlungs- und Belastungssituation in einer Mehraufgaben-Umgebung temporär irrelevant werden. Das erfordert, dass ihr Inhalt klar und schnell erfassbar ist und ohne größere Ablenkung eine Entscheidung über ihre Bedeutung getroffen werden kann. Um diese Aspekte in einem einzigen Test erfassen zu können, wurde das ISE Paradigma für das vorliegende Experiment modifiziert. Das Hauptinteresse liegt auf dem Vergleich der Wirkung der Schalle auf das Arbeitsgedächtnis. Die Schnelligkeit der Informationsaufnahme ist dem untergeordnet.

4.2.2.1 Methode

Überblick

Ein klassisches ISE-Experiment (vgl. Jones & Macken, 1993; Salamé & Baddeley, 1982) beinhaltet eine Gedächtnisaufgabe, bei der den Versuchspersonen visuell zufällig permutierte Items einzeln nacheinander präsentiert werden. Dabei kommt meist eine Reihe von Buchstaben oder einstelligen Ziffern zum Einsatz, die an der Kapazitätsgrenze oder leicht darüber liegt. Nach einem Behaltensintervall von mehreren Sekunden muss das dargebotene Material in der Präsentationsreihenfolge erinnert werden. Die Hintergrundschalle können während der Itempräsentation, während des Behaltensintervalls oder in beiden Intervallen ausgegeben werden. Die Teilnehmer werden angehalten, jedes Geräusch zu ignorieren und sich nur auf die Arbeitsgedächtnisaufgabe zu konzentrieren. Im Detail müssen sie die Items in der korrekten Reihenfolge memorieren und sie während des Behaltensintervalls aufrechterhalten. In einer Kontrollbedingung wird die Aufgabe ohne Hintergrundschall in Stille ausgeführt. Die abhängige Variable ist die Zahl der Positionsfehler für jede Schallbedingung.

Dieses Paradigma wurde für die hier berichtete Untersuchung abgewandelt, um den Einfluss verschiedener auditiv präsentierter Systemmeldungen auf die Arbeitsgedächtnisleistung auf der einen Seite und auf die notwendige Zeit zum Verstehen (Dekodieren) der Nachricht auf der anderen Seite zu erfassen. Die serielle Gedächtnisaufgabe aus dem Standardexperiment wurde dazu um eine Wahlreaktionsaufgabe während des Behaltensintervalls erweitert. In diesem Intervall erfolgte die Darbietung der unterschiedlichen Audionachrichten. Statt jedoch die Versuchspersonen anzuweisen, die

Töne bzw. Sprachnachrichten zu ignorieren, sollten sie schnellstmöglich durch Drücken der zugeordneten Taste auf einem Tastenbrett das Verstehen der Nachricht bestätigen. Dabei mussten sie darauf achten, die Gedächtnisaufgabe nicht zu vernachlässigen. Die Teilnehmer erhielten die Instruktion, die Zahlenfolge mit höchster Priorität aufrechtzuerhalten. Dadurch wurde eine Doppelaufgabe etabliert, die wesentliche Elemente mit der anvisierten Anwendungssituation unter geteilter Aufmerksamkeit aufweist: Situationsrelevante Information muss im Arbeitsgedächtnis aufrechterhalten und gleichzeitig eine auditive Nachricht beurteilt werden. Im direkten Vergleich konnten so die Gestaltungsvarianten Earcons, Auditory Icons, Schlüsselwörter und vollständige Sätze untersucht werden.

Neben den Reproduktionsdaten wurden Reaktionszeiten und Fehler bei der Wahlreaktionsaufgabe ermittelt. Diese sollen darüber Auskunft geben, wie lange es dauert, ein auditives Ereignis wahrzunehmen und dessen semantischen Gehalt zu verarbeiten, während gleichzeitig kognitive Ressourcen anderweitig eingesetzt werden. Das Drücken der Taste signalisiert den Abschluss dieses Prozesses. Die Fehlerrate gibt einen Hinweis auf die Güte der Assoziation zwischen dem auditiven Reiz und der zugehörigen Antwort. Damit auch in der Baseline-Bedingung ohne akustische Ausgabe dieser kognitive Prozess stattfindet und eine motorische Antwort erfolgt, mussten die Versuchspersonen auf einen visuellen Stimulus (Piktogramm) reagieren. Das Ausmaß der Ablenkung durch Auswahl und Drücken der Taste war somit über alle Schallbedingungen konstant. Da die Versuchsteilnehmer bei der Gewichtung der Aufgaben die Gedächtnisaufgabe als Primäraufgabe behandeln sollten, können Veränderungen der Gedächtnisleistung bei der Reproduktion direkt dem unterschiedlichen Einfluss von Earcons, Auditory Icons, kurzen und langen Sprachausgaben zugeschrieben werden. Auf diese Weise kann die Forschungsfrage beantwortet werden, ob diese Arten von Schallen zu einer ISE-artigen Störung des Arbeitsgedächtnisses führen und inwieweit sie sich dabei unterscheiden.

Versuchspersonen

An der Untersuchung nahmen 25 Versuchspersonen (11 Frauen, 14 Männer) im Alter zwischen 20 und 50 Jahren (Mittelwert: 26 Jahre) teil. Sie erhielten eine entgeltliche Aufwandsentschädigung. Alle Probanden hatten zum Zeitpunkt des Tests keine Kenntnis über die Hypothesen des Experiments und berichteten normale Hörfähigkeit.

Versuchsdesign und Aufgaben

Die Systemausgaben wurden als unabhängige Variable (Schallbedingung) in einem Within-Subjects Design miteinander verglichen. Als nicht-sprachliche Stimuli kamen Earcons und Auditory Icons zum Einsatz. Sprachliche Ausgaben wurden mittels vollständiger Sätze oder der Ausgabe von einfachen Schlüsselwörtern realisiert. Wie beim

ISE Standardparadigma fungierte Stille als Kontrollbedingung zur Überprüfung der Reproduktionsleistung ohne interferierendes Schallmaterial.

Inhaltlich war die Gedächtnisaufgabe ebenfalls wie in einem typischen ISE-Experiment angelegt. In jedem Durchgang mussten sich die Versuchspersonen eine seriell dargebotene, zufällige Zahlenfolge bestehend aus den Ziffern 1 bis 9 in der richtigen Reihenfolge merken und nach einem Behaltensintervall wiedergeben (*Serial Recall*). Von Interesse war als abhängige Variable die Anzahl der Fehler bei der Reproduktion. Diese Aufgabe wurde in gleicher Form für alle Schallbedingungen durchgeführt. Um die Übertragbarkeit der experimentellen Aufgabe zur Fahrsituation zu erhöhen, wäre ebenso denkbar gewesen, Navigationsinformation zum seriellen Behalten vorzugeben. Das Material wäre jedoch während des Experiments schwierig zu handhaben gewesen, da verbale Anweisungen wie „die nächste Straße rechts und dann an der Ampel geradeaus weiter“ nicht ohne weiteres in einer standardisierten und damit vergleichbaren Form und Länge über die Durchgänge hinweg austauschbar sind. Es erschien deshalb vorteilhaft, bewährtes abstraktes Versuchsmaterial zu verwenden, dessen Effekt in der Grundlagenforschung bereits bekannt ist. Ohnehin existieren einige Untersuchungen, die statt Zahlen oder Buchstaben Items aus der direkten Umwelt der Aufgabe verwendet haben. So mussten zum Beispiel in einem Experiment Koordinaten und in einem anderen Navigationsinformationen von Flugzeugen in einer Fluglotsenaufgabe aufrechterhalten werden (Banbury, Jones & Emery, 1999). Auch unter diesen Bedingungen gab es ISE Effekte, obwohl die Aufgaben in begrenztem Maß räumliche Vorstellungen beim Aufrechterhalten der Information ermöglichten. Es deutet vieles darauf hin, dass entsprechend auch das Verwenden von seriellen Navigationsanweisungen zu denselben Resultaten führen würde.

Die Wahlreaktionsaufgabe unterschied sich zwischen den Bedingungen mit Audionachricht und der Stille-Bedingung. In den vier experimentellen Bedingungen wurde durch die sprachliche bzw. nicht-sprachliche akustische Ausgabe signalisiert, welche Taste auf einem Tastenbrett zu drücken war. Insgesamt gab es vier verschiedene akustische Ereignisse, die durch auditive Stimuli in jeder Gruppe von Reizen repräsentiert wurden. Entsprechend waren vier Tasten auf dem Tastenbrett angebracht, deren Bedeutung durch ein Piktogramm markiert wurde. Die Assoziation zwischen Piktogramm und akustischer Ausgabe wurde im Vorfeld jeder Bedingung gelernt. In der Stille-Bedingung wurde statt dem akustischen Signal direkt das Piktogramm als visueller Hinweisreiz angezeigt. Sobald die Probanden die Bedeutung des akustischen Signals verstanden bzw. das Bild identifiziert hatten, sollten sie die zugehörige Taste drücken. In beiden Fällen waren sie angewiesen, bei der Ressourcenallokation die Gedächtnisaufgabe absolut zu bevorzugen. Während die Stille-Bedingung in der Serial Recall Aufgabe die Kontrollbedingung darstellt, sind deren Reaktionszeit- und Fehlerdaten bei der Wahlreaktionsaufgabe von mindermem Interesse. Relevant sind hier ausschließ-

lich die unterschiedlichen Zeiten zum Dekodieren der Earcons, Auditory Icons und der Schlüsselwörter.

Jede Versuchsperson absolvierte fünf experimentelle Blöcke bestehend aus je zwei Durchgängen pro Schallbedingung. Innerhalb eines Blocks blieb die Schallbedingung gleich. Die Reihenfolge der Blöcke (Schallbedingungen) wurde mit einem balancierten lateinischen Quadrat bestimmt. Ein Durchgang bestand aus der Präsentation der Zahlenfolge, der Durchführung von zwei Wahlreaktionsaufgaben in der zugeordneten Schallbedingung während des Behaltensintervalls und der anschließenden Reproduktion der Zahlenfolge.

Stimuli

Vier verschiedene Ereignisse wurden in jeder experimentellen Bedingung durch einen auditiven Stimulus repräsentiert und waren einem Piktogramm auf dem Tastenbrett zugeordnet. Nachdem das Experiment im Kontext der ingenieurpsychologischen Forschung an Kfz-Bordsystemen stattfand, wurden die Symbole und deren Bedeutung aus diesem Bereich entliehen. Die Piktogramme zeigten vereinfachte Darstellungen einer Batterie, eines Sitzgurtes, eines Tachometers und einer Zapfsäule (einer Tankstelle). Die Meldungen waren jedoch für die Erfordernisse des Experiments konstruiert und haben artifiziellen Charakter. Sie sollten auf die folgenden Ereignisse aufmerksam machen: 1) die Tankfüllung hat die Reserve erreicht; 2) der Gurt ist nicht korrekt geschlossen; 3) die Batterie hat einen niedrigen Ladestand; 4) eine Geschwindigkeitsbeschränkung ist zu beachten. Alle akustischen Ausgaben wurden auf dieselbe Lautstärke normalisiert. Die Sprachausgaben wurden vorab aufgezeichnet, auf ausreichend deutliche Aussprache wurde dabei geachtet.

Die Ereignisse wurden durch folgende Schlüsselwörter signalisiert (Dauer in Klammern):

- Tank: „*Tanken*“ (440 ms)
- Gurt: „*Gurt*“ (280 ms)
- Batterie: „*Batterie*“ (550 ms)
- Geschwindigkeitsbeschränkung: „*Geschwindigkeit*“ (920 ms)

Bei der Konstruktion von Sprachausgaben mit vollständigen Sätzen stand das Interesse an der Gedächtnisaufgabe im Vordergrund. A priori ist klar, dass Reaktionszeiten auf solche Sätze von ihrer Konstruktion abhängen und damit davon, an welcher Stelle im Satzbau ihr Inhalt eindeutig verständlich wird. Weiterhin ist klar, dass Ausgaben mit vollständigen Sätzen stets so konstruiert werden sollten, dass möglichst früh die Bedeutung vermittelt wird und der verbleibende Satzteil für eine Klärung des vermittelten Inhalts genutzt werden kann. Unter den bereits erwähnten Vorbedingungen wurden die Sätze einheitlich so gestaltet, dass als erstes Wort das oben angegebene Schlüsselwort genannt wurde, während der verbleibende Satzteil Zusatzinformationen

dazu gab. Dadurch sollte erreicht werden, dass die Reaktionszeiten auf Sätze mit den anderen Bedingungen vergleichbar werden. Die Versuchspersonen erhielten zu diesem Reizmaterial die Erklärung, dass sie nur auf das erste Wort reagieren und den Rest des Satzes soweit es geht ignorieren sollten. Auf diese Weise wurde eine Situation simuliert, in der der Nutzer bereits die Grundbotschaft der Nachricht erkannt hat, ihr aber wegen vordringlicher Anforderungen der Hauptaufgabe keine weitere Aufmerksamkeit zukommen lassen möchte. Nachdem die Aufgabe eher abstrakt war und nicht an Bord eines Fahrzeugs durchgeführt wurde, wurden die Satzaussagen für die Wahlreaktionsaufgabe und nicht auf tatsächliche Aufgaben während der Fahrzeugführung angepasst. Über den funktionsunabhängigen Charakter ihrer Handlungen waren die Versuchspersonen informiert.

Die einzelnen Meldungen waren wie folgt aufgebaut:

- Tank: „*Tanken* drücken, um die Tankfüllung des Fahrzeugs zu überprüfen.“ (3.6 s)
- Gurt: „*Gurt* drücken, um den Verschluss des Gurtes zu überprüfen.“ (3.7 s)
- Batterie: „*Batterie* drücken, um den Ladezustand der Batterie zu überprüfen.“ (3.7 s)
- Geschwindigkeitsbeschränkung: „*Geschwindigkeit* drücken, um die Geschwindigkeit des Fahrzeugs zu überprüfen.“ (4.2 s)

Bei der Auswahl der nicht-sprachlichen Stimuli wurde darauf geachtet, dass sie einen sehr eigenständigen Klang hatten und nicht leicht zu verwechseln waren. Diese Anforderung besteht grundsätzlich bei der auditiven Gestaltung. Die verwendeten Earcons kamen bereits bei einer Untersuchung von Leplâtre und Brewster (2000) zum Einsatz und haben sich dort bewährt. Sie wurden mit Erlaubnis von Brewster genutzt und sind auf dessen Homepage verfügbar (Brewster, 2000). Obwohl die Dauer der Earcons unterschiedlich lang war, war ihr individueller Charakter durchwegs nach etwa 200 ms etabliert. Die Tonfolgen der Earcons können folgendermaßen beschrieben werden:

- Tank: hart einsetzender Mollakkord, E-Piano (1.4 s)
- Gurt: gebrochener Akkord mit Nonencharakter, E-Piano/Orgel (780 ms)
- Batterie: pizzicatoartiges Geigenmotiv aus drei rasch aufeinander folgenden Noten (effektiv ca. 400 ms; zusätzlich Ausklingen und Nachhall, summiert ca. 1 s)
- Geschwindigkeitsbeschränkung: auf Klavierklang beruhende offene Quinte, Klangeinsatz unterstützt durch einen Tamburinschlag (900 ms)

Die Auditory Icons stammen aus der Datenbank für Sound-Design des Fachzentrums für User Interface Design der Siemens AG. Nachfolgend eine Beschreibung des Klangbilds:

- Tank: Geräusch von schnell strömender Flüssigkeit, die ein Gefäß füllt (350 ms)
- Gurt: Metallisches Klickgeräusch, ähnlich dem Schließen eines Schlosses (420 ms)
- Batterie: Geräusch eines von Hochspannung induzierten Funkenflugs bei Kontakt zu elektrischem Leiter (370 ms)
- Geschwindigkeitsbeschränkung: Geräusch eines sehr schnell vorbeifahrenden Fahrzeugs, abnehmende Tonhöhe (220 ms)

Versuchsaufbau

Der Versuch fand in ruhigen Laborräumen statt. Jeder Versuchsdurchgang lief vollständig computergesteuert ab, lediglich die Versuchsbedingung (Schallbedingung) musste eingestellt werden. Das Experimentalprogramm wurde vom Autor in Microsoft® Visual Basic® implementiert. Der verwendete Computer war ein Dell Laptop mit 1.2 GHz Intel Pentium M® CPU (Chipsatz Intel 855GM®), 12 Zoll TFT Monitor und einer SigmaTel C-Major on-board Audio Soundkarte. Der Laptop wurde so auf einem Tisch vor der Versuchsperson platziert, dass sie die Bildschirminhalte gut lesen konnte. Vor dem Laptop war das Tastenbrett für die Wahlreaktionsaufgabe platziert. Die Zahlenfolge wurde mit Hilfe einer herkömmlichen drahtlosen Computermaus eingegeben, indem Zahlenfelder auf dem Bildschirm angeklickt wurden. Die Tastatur des Laptops wurde nicht benötigt. Der Aufbau ist in Abbildung 24 dargestellt.



Abbildung 24: Versuchsperson während der Darbietung der Ziffernfolge. Die rechte Hand ruht auf dem Tastenbrett für die Wahlreaktionsaufgabe. Rechts befindet sich die Maus zur anschließenden Zifferneingabe.

Die auditiven Stimuli wurden diotisch über lärmämpfende dynamische Sennheiser HD 25-1 Studiokopfhörer ausgegeben (geschlossene Bauform, ohraufliegende An-

kopplung mit Ringpolstern, Andruckkraft ca. 2.5 N). Mit Hilfe dieser Kopfhörer konnten eventuelle Umgebungsgeräusche während des Experiments nahezu vollständig eliminiert werden.

Prozedur

Nachdem die Versuchsperson begrüßt wurde und einige Angaben zu ihrer Person gemacht hat (Alter, Geschlecht, Hörfähigkeit), wurden ihr zunächst die experimentellen Aufgaben erklärt. Zu Beginn des Experiments absolvierten die Versuchspersonen zwei Probedurchgänge in einer jeweils zufällig ausgewählten Bedingung. In den Probedurchgängen mussten sie noch nicht die richtige Antwort auf den auditiven bzw. visuellen Cue der Wahlreaktionsaufgabe geben, sie sollten eine beliebige Taste drücken. Diese Durchgänge dienten nur zur Eingewöhnung an das Vorgehen.

Jeder Durchgang war gleich aufgebaut. Die Teilnehmer saßen vor dem Computerbildschirm und trugen Kopfhörer. In der Mitte des Bildschirms erschien für 2 Sekunden ein Fixationskreuz, das die Probanden anblicken sollten. Danach wurden seriell die Ziffern 1 bis 9 in zufälliger Reihenfolge präsentiert. Jede Zahl erschien nur einmal und wurde für 800 ms mit einem Interstimulus-Intervall von 200 ms angezeigt. Die Versuchspersonen waren im Vorfeld angewiesen worden, sich die Reihenfolge zu merken und sie aufmerksam im Geist zu wiederholen. Dabei sollten sie es vermeiden, die Zahlenfolge laut auszusprechen.

Nach der letzten Zahl erschien ein Pfeil in der Mitte des Bildschirms. Er zeigte nach unten und deutete so an, dass die Versuchspersonen nun das Tastenbrett nutzen mussten, um auf den Stimulus in der Wahlreaktionsaufgabe zu reagieren. Je nach experimenteller Bedingung wurde den Versuchsteilnehmern dann auditiv ein Earcon, ein Auditory Icon, ein Schlüsselwort oder ein vollständiger Satz dargeboten. Ein visueller Cue wurde nicht gegeben. In der Kontrollbedingung ohne auditiven Stimulus zeigte hingegen ein visueller Hinweis ohne auditive Ausgabe die zu drückende Taste an. Diese Reaktionsaufgabe wurde im selben Behaltensintervall einmal wiederholt, wobei stets eine andere Systemnachricht ausgegeben wurde. Das Behaltensintervall dauerte 15 Sekunden.

Im Anschluss an das Behaltensintervall mit den Wahlreaktionsaufgaben erschien ein 3×3 Ziffernblock auf dem Bildschirm. Das Zahlenlayout entsprach der Tastenkonfiguration des Nummernblocks auf einer PC-Tastatur ohne die Ziffer Null. Mit Hilfe der Maus musste die zuvor gelernte Zahlenfolge in der erinnerten Reihenfolge eingegeben werden. Jede Zahl konnte nur einmal angeklickt werden.

Der Durchgang endete mit dieser Reproduktion der Zahlenfolge. Zwischen zwei Durchgängen pausierten die Probanden für ein paar Minuten. Vor jeder neuen Schallbedingung, d.h. vor jedem experimentellen Block und damit auch zu Beginn des Experiments, fand ein Stimulus-Response Training statt. Die auditive Ausgabe wurde dazu

vorgespielt und das zugehörige Piktogramm auf dem Bildschirm angezeigt. Nach einigen dieser Trainingsdurchgänge, wenn die Versuchsperson den Eindruck hatte, die Verbindung gelernt zu haben, wurde nur noch der auditive Stimulus gegeben. Dieses Training wurde fortgeführt, bis jeder Stimulus dreimal nacheinander korrekt ohne Hilfe identifiziert wurde. Aufgrund der engen Verbindung zwischen den Piktogrammen und den sprachlichen Ausgaben wurde das Training in diesen Bedingungen sowie in der visuellen Cueing-Bedingung auf eine korrekte Antwort begrenzt.

4.2.2.2 Hypothesen

Vorrangiges Ziel des Experiments war zu klären, inwiefern sich typische Gestaltungsvarianten akustischer Systemausgaben negativ auf die Arbeitsgedächtnisleistung auswirken können. Die Kontrollbedingung Stille mit visuellem Cue für die Wahlreaktionsaufgabe wird wegen der Darbietungsmodalität das Aufrechterhalten der Information in der phonologischen Schleife am wenigsten beeinträchtigen. Ausgehend von der Grundlagenforschung zum ISE wird erwartet, dass als vollständige Sätze formulierte Informationsausgaben einen massiv nachteiligen Einfluss auf die Behaltensleistung ausüben werden. Dies sollte ersichtlich sein an einem deutlichen Unterschied zwischen der Kontrollbedingung und der Schallbedingung mit vollständigen Sätzen. Nachdem bekannt ist, dass das Ausmaß des ISE abhängig ist von der Anzahl eingestreuter Wörter (Bridges & Jones, 1996), sollte in der Schlüsselwort-Bedingung eine deutlich schwächere Störwirkung der Sprachausgaben zu verzeichnen sein. Die nicht-sprachlichen Schallbedingungen sollten ebenfalls deutlich schwächer stören als vollständige Sätze. Das lässt sich daraus ableiten, dass die Forschungsgruppe um Jones als verbales Material sinnlose Silben oder Ein-Silben-Wörter zum Nachweis der gleichen Wirksamkeit von sprachlichen und nicht-sprachlichen Schallen nutzte. LeCompte, Neely und Wilson (1997) konnten hingegen zeigen, dass „richtige“ Wörter meist zu etwas stärkeren Effekten führen. Ferner ist die Unterbrechung des Rehearsal durch auditive Ereignisse wie Auditory Icons und Earcons deutlich kürzer als durch vollständige Sätze. Somit sollte sich nur ein marginaler Effekt gegenüber der Baseline ergeben.

Für die Wahlreaktionsaufgabe wird erwartet, dass aufgrund der willkürlichen Bedeutungszuweisung bei Earcons – trotz des ausführlichen Trainings – die Reaktionszeiten eindeutig am Längsten sind. Wenn die Zeiten länger sind spricht das für erhöhte kognitive Aktivität während der Dekodierphase. Das könnte theoretisch einen nachteiligen Einfluss auf die Behaltensleistung bei Earcons haben. Sollte die Behaltensleistung bei den Earcons dennoch nicht verschlechtert sein, wäre das ein Hinweis auf die Unabhängigkeit der bei der Dekodierung der Audionachricht notwendigen kognitiven Leistungen und der Aktivität in der phonologischen Schleife. Zwischen den Varianten der Sprachausgabe wird wegen der spezifischen Konstruktion der Sätze mit einem

Schlüsselwort zu Beginn und einer nachfolgenden „Erklärung“ kein Unterschied erwartet. Die Reaktionszeiten von Sprache und Auditory Icons sollten sich nicht deutlich unterscheiden, weil beide intuitiv zugängliche Formen der Abbildungsrelation zwischen Bezeichnendem und Bezeichnetem sind. In den direkten Vergleichen werden die nicht-sprachlichen Stimuli nur gegen die Schlüsselwort-Bedingung getestet. Zum einen weisen beide sprachliche Varianten als ersten Begriff das Schlüsselwort auf, weshalb ein Vertreter für den Vergleich ausreicht. Zum anderen ist nur für diesen Vergleich eine allgemeine Aussage möglich, da bei Sätzen stets die Art der grammatischen Konstruktion berücksichtigt werden müsste. Bezüglich der Fehlerraten in diesen Bedingungen wird ebenfalls unter Verwendung derselben Vergleiche wie bei den Reaktionszeiten angenommen, dass es zwischen den Varianten mit intuitiver Abbildung nur wenig Unterschiede gibt. Denkbar ist jedoch, dass Auditory Icons nicht mit der gleichen Güte auf das zu repräsentierende Ereignis verweisen wie die sprachlichen Stimuli und so Unterschiede entstehen. Grundsätzlich sollte die Fehlerrate bei der nicht-sprachlichen abstrakten Variante, den Earcons, am höchsten sein. Auch hier ist wieder das Ausmaß von Interesse. Insgesamt hat die Untersuchung der Leistung bei der Wahlreaktionsaufgabe deutlich explorativeren Charakter als die Arbeitsgedächtnisaufgabe.

4.2.2.3 *Ergebnisse*

Für alle Schallbedingungen (Earcons, Auditory Icons, vollständige Sätze, Schlüsselwörter und Stille) wurden die mittlere Anzahl an Positionsfehlern in der Arbeitsgedächtnisaufgabe und die mittleren Fehlerraten sowie die Reaktionszeiten in der Wahlreaktionsaufgabe ermittelt. Die Daten wurden mit Varianzanalysen auf statistisch signifikante Unterschiede geprüft. Wies die Varianzanalyse einen belastbaren Effekt aus, wurden mit paarweisen t-Tests für abhängige Stichproben ausgewählte Vergleiche betrachtet, um detaillierte Aussagen über die aufgestellten Hypothesen zu ermöglichen. Wenngleich a priori Tests ein besonderer Status bei der Notwendigkeit einer Anpassung des Signifikanzniveaus in der Diskussion multipler Tests eingeräumt wird (vgl. die Ausführungen in Abschnitt 4.1.2.3), erscheint aufgrund der etwas höheren Anzahl geplanter Vergleiche dennoch die Berücksichtigung der *family wise error rate* (FWE) angebracht. Die Vergleiche werden mit einer Einschränkung der FWE auf $p_{\text{FWE}} = .10$ mit der sequentiellen Bonferroni-Holm-Prozedur (vgl. Shaffer, 1995) abgesichert. Die statistischen Kennwerte werden dazu geordnet und ein Vergleich der p -Werte gegen einen kritischen Grenzwert p_{crit} durchgeführt. Im Gegensatz zur einfachen Bonferroni-Korrektur handelt es sich bei der Holm-Variante um kein konstantes, sondern um ein abgestuftes Kriterium. Auf der ersten Stufe ergibt sich p_{crit} durch Teilen des globalen Signifikanzniveaus p_{FWE} durch die Anzahl aller Vergleiche (m), d.h. $p_{\text{crit}_1} = p_{\text{FWE}} / m$. Die Nullhypothese wird abgelehnt, wenn der p -Wert kleiner ist als

p_{crit} . Nur wenn ein signifikanter Unterschied nachweisbar war, wird für den zweiten Vergleich p_{crit} berechnet als $p_{\text{crit}_2} = p_{\text{FWE}} / (m - 1)$. Allgemein ausgedrückt kann für jede Stufe j der Grenzwert berechnet werden als $p_{\text{crit}_j} = p_{\text{FWE}} / (m - j + 1)$. Nach einem nichtsignifikanten Ergebnis werden ohne weitere Prüfung alle verbleibenden Unterschiede ebenfalls als nichtsignifikant betrachtet und müssten nicht gesondert bearbeitet werden. Für eine bessere Übersicht über die Resultate werden in den entsprechenden Tabellen dennoch die Ergebnisse aller Paarvergleiche notiert.

In jeder Analyse werden zusätzlich die Statistik p_{rep} (Killeen, 2005a, 2005b) und die Effektstärken berichtet (bei Varianzanalysen das partielle Eta-Quadrat η_p^2 , bei t-Tests die Effektgröße d). Zur Beschreibung der Resultate werden stets Mittelwerte (M) und Standardfehler (SE) angegeben. Bei grafischen Darstellungen weisen die Fehlerbalken Standardfehler aus.

Interferenz der Ausgaben mit dem Arbeitsgedächtnis

In der *Serial Recall*-Gedächtnisaufgabe wurden angelehnt an das übliche Analyse-schema von ISE Experimenten alle Bedingungen mit akustischer Ausgabe gegen die Kontrollbedingung Stille verglichen. Die Versuchspersonen mussten eine Sequenz von neun Ziffern in der richtigen Reihenfolge memorieren und wiedergeben. Während des Behaltensintervalls erfolgte die Ausgabe der auditiven Signale bzw. in der Kontrollbedingung Stille des visuellen Signals. Als Fehler wurde bei der reproduzierten Zahlenkette jede Zahl an einer falschen Position gewertet. Wenn nicht alle Zahlen an der korrekten Position wiedergegeben wurden, resultieren durch die Vertauschung mindestens zwei Fehler. Das Ergebnismuster ist in Abbildung 25 dargestellt.

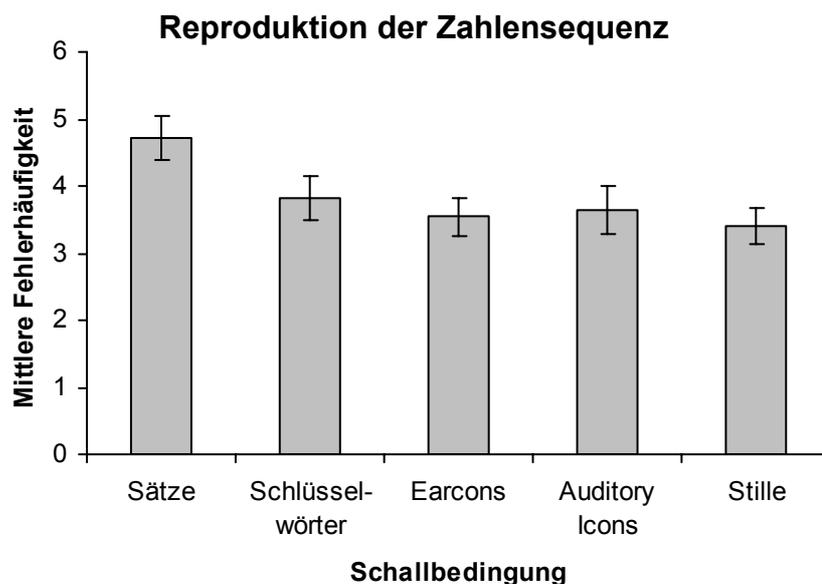


Abbildung 25: Mittlere Anzahl an Fehlern bei der Gedächtnisaufgabe in Abhängigkeit von der Schallbedingung.

Die visuelle Inspektion lässt vermuten, dass sich außer der Schallbedingung mit vollständigen Sätzen keine der anderen Bedingungen deutlich von der Baseline unterscheidet. Auf der Seite der sprachlichen Ausgaben machten die Probanden bei vollständigen Sätzen im Mittel 4.72 ($SE = 0.33$) Positionsfehler. Wurden Schlüsselwörter ausgegeben, reduzierte sich die Fehlerzahl auf 3.82 ($SE = 0.34$). Bei nicht-sprachlichen Ausgaben ergaben sich für Earcons 3.54 ($SE = 0.28$) und für Auditory Icons 3.64 ($SE = 0.35$) Fehler. Stille führte durchschnittlich zu 3.40 ($SE = 0.27$) Positionsfehlern. Zunächst wurde eine einfaktorielle Varianzanalyse mit dem Messwiederholungsfaktor Schallbedingung gerechnet. Diese zeigte einen signifikanten Unterschied zwischen den Schallbedingungen an ($F(4,96) = 3.25, p = .02, p_{rep} = .94, \eta_p^2 = .12$).

Die Einzelvergleiche aller Bedingungen mit auditiver Stimuluspräsentation während des Behaltensintervalls gegen die Kontrollbedingung bestätigen den Eindruck der alleinigen Interferenzwirkung vollständiger Sätze. Tabelle 6 zeigt die Analyseergebnisse der sequentiellen Bonferroni-Holm-Prozedur.

Tabelle 6: Geplante Vergleiche der Reproduktionsleistungen gegen die Baseline Stille (geordnet nach statistischen Kenngrößen).

Vergleich	$t(24)$	p	p_{crit}	p_{rep}	d
Sätze – Stille	3.16	.004*	.025	.97	0.63
Schlüsselwörter – Stille	1.10	.281	.033	.66	0.22
Auditory Icons – Stille	0.56	.579	-	.44	0.11
Earcons – Stille	0.38	.707	-	.35	0.08

Anmerkung: * signifikant bei $p < p_{crit}$ (Bonferroni-Holm, $p_{FWE} = .10$).

Zwischen der Bedingung mit langer Sprachausgabe und Stille zeigt sich ein deutlicher Effekt, die Ausgabe ganzer Sätze während des Behaltensintervalls verschlechterte die Reproduktionsleistung beträchtlich. Alle weiteren Vergleiche zeigten keine statistisch belastbare Differenz zur Kontrollbedingung. Selbst wenn die nicht korrigierten p -Werte als Grundlage für die Entscheidung herangezogen würden, ließe sich kein Unterschied nachweisen. Ferner sind die Effekte in diesen Bedingungen eher schwach (d maximal .22).

Leistung bei der Wahlreaktionsaufgabe

In die inferenzstatistische Analyse bei der Wahlreaktionsaufgabe gingen nur die Bedingungen mit akustischer Ausgabe ein. Die versuchsplanerische Einbindung einer Reaktion auf visuelle Stimuli bei Stille hatte ausschließlich den Zweck, auch in dieser Bedingung während des Behaltensintervalls eine Auswahl Aufgabe zu realisieren und so eine ähnliche Belastungssituation für den Vergleich der Reproduktionsleistung zu

schaffen. Die Leistung bei visuellen Cues in der Wahlreaktionsaufgabe wird deshalb nur deskriptiv dargelegt.

Bei der Auswertung der Reaktionszeiten wurden nur die Durchgänge berücksichtigt, in denen die Teilnehmer richtig geantwortet hatten. Abbildung 26 fasst die Resultate zusammen.

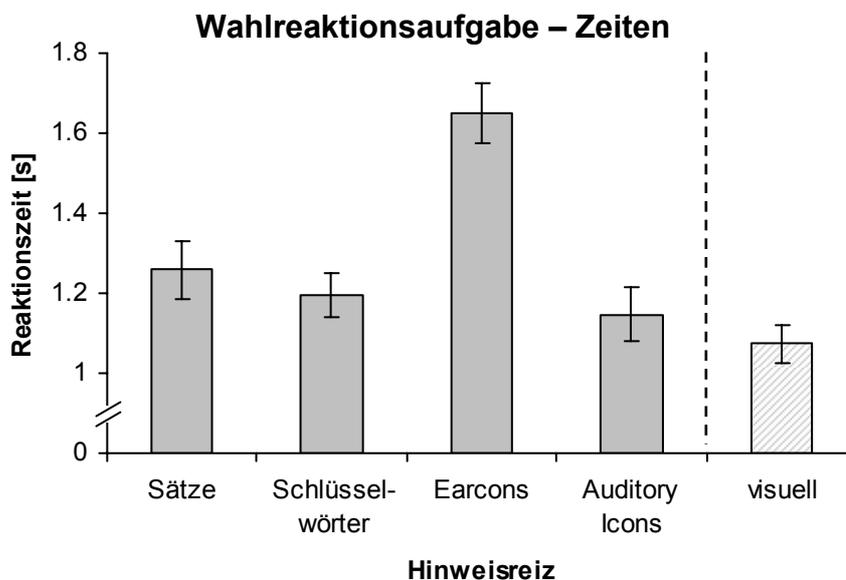


Abbildung 26: Mittlere Reaktionszeit auf auditive und visuelle Stimuli in der Wahlreaktionsaufgabe.

Die Ausgabe vollständiger Sätze führte zu einer mittleren Reaktionszeit von 1.26 s ($SE = 0.07$). Ähnlich, aber etwas kürzer ist die Reaktionszeit auf Schlüsselwörter mit 1.20 s ($SE = 0.05$). Deutliche Unterschiede gibt es in der Gruppe der nicht-sprachlichen Schalle. Die Probanden benötigten im Schnitt 1.65 s ($SE = 0.08$), um auf Earcons zu reagieren. Auditory Icons hingegen bedingen eine Reaktionszeit von nur 1.15 s ($SE = 0.07$). Ohne Einfluss auf die nachfolgende Analyse ist die Zeit von 1.07 s ($SE = 0.05$) auf visuelle Stimuli. Eine einfaktorielle Varianzanalyse mit dem vierstufigen Messwiederholungsfaktor akustischer Hinweisreiz belegt einen unterschiedlichen Einfluss der Ausgabeform ($F(3,72) = 13.66$, $p < .001$, $p_{rep} > .99$, $\eta_p^2 = .36$). Dieser Effekt dürfte durch die langsamen Reaktionszeiten bei Earcons ausgelöst worden sein. Nachdem es bei der Wahlreaktionsaufgabe keine eigentliche Kontrollbedingung gab, sind die geplanten Vergleiche nach den einzelnen Hypothesen in Abschnitt 4.2.2.2 ausgerichtet. Relevant sind damit zunächst mögliche Unterschiede innerhalb der sprachlichen und innerhalb der nicht-sprachlichen Stimuli (Sätze – Schlüsselwörter; Earcons – Auditory Icons). Ferner werden Auditory Icons und Earcons noch gegen die Schlüsselwörter-Bedingung als Vertreter der sprachlichen Schalle getestet. Die Ergebnisse sind in Tabelle 7 zusammengefasst.

Tabelle 7: Geplante Vergleiche der Reaktionszeiten (geordnet nach statistischen Kenngrößen).

Vergleich	$t(24)$	p	p_{crit}	p_{rep}	d
Earcons – Schlüsselwörter	6.12	<.001*	.025	>.99	1.22
Earcons – Auditory Icons	4.85	<.001*	.033	>.99	0.97
Sätze – Schlüsselwörter	0.74	.468	.050	.52	0.15
Schlüsselwörter – Auditory Icons	0.62	.543	-	.47	0.12

Anmerkung: * signifikant bei $p < p_{crit}$ (Bonferroni-Holm, $p_{FWE} = .10$).

Danach wird durch die Paarvergleiche bestätigt, dass Earcons zu signifikant langsameren Reaktionszeiten führen als Auditory Icons oder Schlüsselwörter. Zwischen Auditory Icons und den Schlüsselwörtern lässt sich kein Unterschied feststellen. Auch innerhalb der Gruppe der sprachlichen Ausgaben ist die Differenz in den Reaktionszeiten nicht signifikant.

Neben den Reaktionszeiten ist bei der Auswertung der Wahlreaktionsaufgabe die Fehlerrate von Interesse. Für jede Bedingung wurde der mittlere Fehleranteil berechnet. Die Analyse erfolgt analog zu den Reaktionszeitdaten. Abbildung 27 veranschaulicht die Ergebnisse.

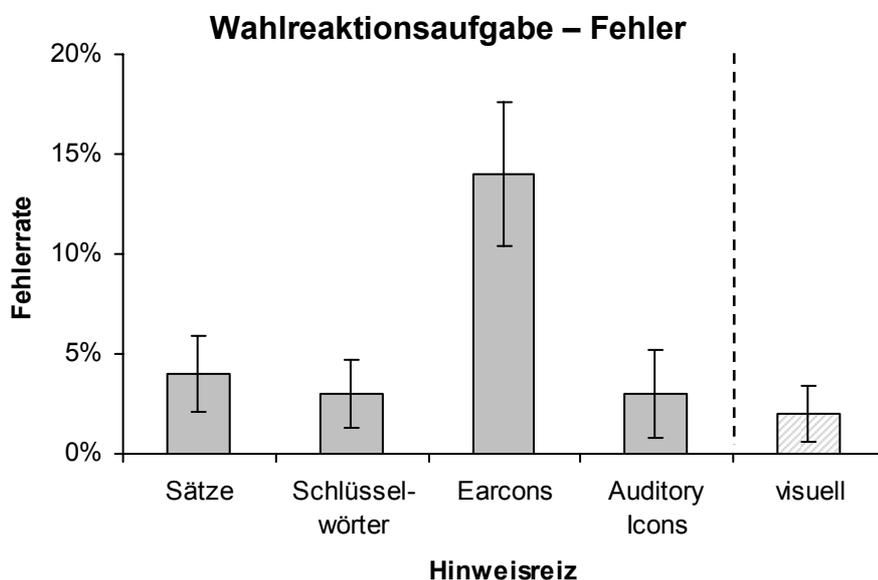


Abbildung 27: Mittlere Fehlerrate bei auditiven und visuellen Stimuli in der Wahlreaktionsaufgabe.

Wurden als auditiver Hinweisreiz vollständige Sätze ausgegeben, machten die Probanden im Durchschnitt 4% Fehler ($SE = 1.9$). Bei Schlüsselwörtern waren es 3% ($SE = 1.7$). Earcons führten zu einer deutlich höheren Fehlerrate von 14% ($SE = 3.6$), während Auditory Icons nur 3% ($SE = 2.2$) fehlerhafte Antworten zur Folge hatten. In den Daten lässt sich folglich kein Speed-Accuracy Trade-off finden. Wiederum ohne

weitere Bedeutung für die Vergleiche zwischen den akustischen Ausgaben ist die Fehlerrate bei visuellen Cues von 2% ($SE = 1.4$). Auch bei den Fehlerraten zeigt die Varianzanalyse einen signifikanten Einfluss der Art des auditiven Hinweisreizes ($F(3,72) = 5.87, p = .001, p_{\text{rep}} = .99, \eta_p^2 = .20$). Wie bei den Reaktionszeiten fallen die Earcons durch deutlich schlechtere Werte auf. Die anderen Varianten der akustischen Ausgabe sind dagegen recht ähnlich.

Mit Hilfe derselben Paarvergleiche wie zuvor kann dieser Eindruck untermauert werden (siehe Tabelle 8. Hinweis: Bei Schlüsselwörtern und Auditory Icons ergab sich aufgrund derselben Fehlerrate keine Mittelwertsdifferenz, weshalb für dieses Paar kein Vergleich gerechnet wurde.). Die relevante Erkenntnis aus den Paarvergleichen ist das nachweisbar schlechte Abschneiden der Earcons.

Tabelle 8: Geplante Vergleiche der Fehlerraten (geordnet nach statistischen Kenngrößen).

Vergleich	$t(24)$	p	p_{crit}	p_{rep}	d
Earcons – Schlüsselwörter	2.86	.009*	.025	.95	0.57
Earcons – Auditory Icons	2.40	.024*	.033	.92	0.48
Sätze – Schlüsselwörter	1.00	.327	.050	.62	0.20

Anmerkung: * signifikant bei $p < p_{\text{crit}}$ (Bonferroni-Holm, $p_{\text{FWE}} = .10$).

Detailanalyse Earcons

Ein überraschendes Ergebnismuster ist das gleichzeitige Vorliegen sehr langer Reaktionszeiten und hoher Fehlerraten bei Earcons und gleichzeitig eine gute Reproduktionsleistung in dieser Schallbedingung bei der *Serial Recall* Aufgabe. Für die Antwortgüte in der Gedächtnisaufgabe scheint es also keinen Unterschied zu machen, ob man sich bei der vorangegangenen Wahlreaktionsaufgabe besonders häufig in der Antwort geirrt hat oder nicht. Ebenso stören die Prozesse, die eine lange Antwortlatenz bedingen, nicht bei der Reproduktion. Um diese Vermutung zu überprüfen, wurde ein Versuchspersonen-Split durchgeführt. Die Teilnehmer wurden einmal danach aufgeteilt, ob sie einen Fehler bei der Antwort auf die Earcons machten oder nicht. Für einen zweiten Test wurden sie danach aufgeteilt, ob sie besonders schnell oder besonders langsam auf Earcons reagierten. Sollten tatsächliche Fehlerrate und Reaktionszeit keinen Einfluss auf die Leistung bei der Reproduktion der Zahlenfolge haben, dann dürfte sich das auch nicht in diesen Subgruppen mit besonders hervorgehobener Merkmalsausprägung zeigen. Vorweggenommen ließ sich bei den Versuchspersonen, die Fehler in der Wahlreaktionsaufgabe bei der Reaktion auf Earcons machten, tatsächlich kein Unterschied in der Reproduktionsleistung bei *Serial Recall* nachweisen.

Hohe Reaktionszeiten gingen jedoch mit einer erhöhten Fehlerrate bei der Gedächtnisaufgabe einher.

Von den 25 Versuchspersonen blieben 14 ohne Fehler bei der Tastenwahl nach Earconpräsentation. Die verbleibenden 11 Probanden machten mindestens einen Fehler bei der Wahlreaktionsaufgabe. In der fehlerfreien Gruppe konnten im Schnitt 3.32 ($SE = 0.31$) Positionsfehler bei der Reproduktion der Zahlenfolge festgestellt werden. Die Gruppe mit mindestens einem Fehler in der Wahlreaktionsaufgabe machten 3.81 ($SE = 0.51$) Positionsfehler. In diesen beiden Gruppen wurden die Fehler in der Gedächtnisaufgabe mit Hilfe eines t-Tests für unabhängige Stichproben analysiert. Es ergab sich kein signifikanter Unterschied, $t(23) = 0.87$, $p = .40$, $p_{\text{rep}} = .57$.

Bei dem Split basierend auf Reaktionszeiten wurden die Versuchspersonen in zwei gleichgroße Gruppen geteilt. Die Daten der zwölf langsamsten und der zwölf schnellsten Teilnehmer bildeten die Grundlage für die Bewertung der Gedächtnisleistung bei der Arbeitsgedächtnisaufgabe mit Earcons im Behaltensintervall. In der schnellen Gruppe ergab sich eine mittlere Reaktionszeit von 1.35 s ($SE = 0.05$) und in der langsamen Gruppe von 1.95 s ($SE = 0.08$). In der Gedächtnisaufgabe wies die schnelle Gruppe mit 2.83 ($SE = 0.33$) Positionsfehlern eine deutlich bessere Reproduktionsleistung auf als die langsame Gruppe, die 4.17 ($SE = 0.40$) Fehler machte. Der t-Test für unabhängige Stichproben weist einen signifikanten Effekt aus ($t(22) = 2.70$, $p = .02$, $p_{\text{rep}} = .94$, $d = 1.04$).

4.2.2.4 Diskussion

Ziel des Experiments war eine Erweiterung des Gestaltungswissens über das ergonomische Design akustischer Ausgaben. In einer multimodalen Arbeitsumgebung sind akustische Systemmeldungen insbesondere für Mehraufgabensituationen mit starker visueller Komponente der Primäraufgabe sinnvoll und werden zunehmend eingesetzt. Aus der Grundlagenliteratur ist bekannt, dass auditive Stimuli mit Arbeitsgedächtnisprozessen und damit mit einer wesentlichen Komponente der Informationsverarbeitung interferieren können. Die bisherige Forschung zur Gestaltung akustischer Ausgaben beschäftigt sich fast ausschließlich mit dem Themenbereich auditiver Warnungen (Baber, 1994; Edworthy, 1994, 1997; Edworthy & Adams, 1996; Edworthy, Loxley & Dennis, 1991; Hirst & Graham, 1997). Die Frage nach einer Störung durch die Ausgabe betrifft dort primär die Verständlichkeit weiterer, nachfolgender Informationen. Über die Ablenkungswirkung zum Beispiel bei diagnostischen Vorgängen oder anderen komplexen kognitiven Tätigkeiten mit Arbeitsgedächtnisforderung ist wenig bekannt. Sofern die akustischen Benachrichtigungen nicht Warnungen höchster Priorität sind, sondern ausschließlich der Übertragung nicht-kritischer Information dienen, muss eine derartige Interferenz weitestgehend minimiert werden.

Deshalb wurden im Experiment sprachliche und nicht-sprachliche Gestaltungslösungen für akustische Ausgaben in ihrer Ablenkungswirkung auf das Arbeitsgedächtnis evaluiert. Für den Anwendungsfall Fahrzeugführung wurde als Beispiel die Situation eingeführt, dass akustische Ausgaben von Komfortkomponenten des Bordsystems das Aufrechterhalten von Information stören können, die die strategische Ebene der Fahraufgabe betrifft. Ein weiteres Ziel des Experiments war zu testen, ob bei Gestaltungsvarianten Unterschiede in der benötigten Zeit und Fehlerrate bei der semantischen Dekodierung bestehen, operationalisiert durch die Leistung in einer Wahlreaktionsaufgabe. Die auditiven sprachlichen Stimuli waren entweder einzelne Schlüsselwörter oder vollständige Sätze. Als nicht-sprachliche Stimuli kamen Earcons und Auditory Icons zum Einsatz. Zusammengefasst hat sich gezeigt, dass sprachliche Ausgaben (Sätze) zu einer schlechteren Leistung bei der Arbeitsgedächtnisaufgabe führen können, während bei der Wahlreaktionsaufgabe nicht-sprachliche Ausgaben (Earcons) von Nachteil sein können.

Interferenz der Ausgaben mit dem Arbeitsgedächtnis

Das Experiment wurde in Anlehnung an Untersuchungen zum *Irrelevant Speech/Sound Effect* (ISE) konstruiert. Die dabei verwendete Arbeitsgedächtnisaufgabe reagiert sensitiv auf sprachliche und nicht-sprachliche Hintergrundschalle, auch wenn sie nur während des kurzen Behaltensintervalls eingestreut werden. Als Grund für die Störung kann die Interferenz der auditiven Stimuli mit dem Aufrechterhalten von Information der Gedächtnisaufgabe in der phonologischen Schleife gesehen werden. Für den genauen Mechanismus gibt es jedoch gegenwärtig mitunter wegen unterschiedlichen Vorstellungen über die Struktur des Arbeitsgedächtnisses keine einheitliche Sichtweise in der Grundlagenforschung (Baddeley, 2000b; Jones & Tremblay, 2000; Neath, 2000; Page & Norris, 2003). Dennoch wird übergreifend anerkannt, dass nicht-sprachliche Schalle in gleichem Ausmaß wie sprachliche Schalle den Rehearsal Prozess im Arbeitsgedächtnis unterbrechen können.

Anders als bei einem ISE Experiment sollten in der vorliegenden experimentellen Aufgabe die akustischen Ausgaben nicht vollständig ignoriert werden. Die Single Task Gedächtnisaufgabe des ISE, in der die Schalle nur einen irrelevanten Hintergrund bilden, wurde in ein Dual Task Setting umgewandelt: Die Primäraufgabe des Experiments bestand im Merken einer Zahlenfolge, die Reaktion auf auditive Stimuli während des Behaltensintervalls war die Sekundäraufgabe. Auf diese Weise entstand eine Situation, die auf viele Anwendungsszenarien übertragbar ist. Während mit oberster Priorität eine kognitive Aufgabe unter Beteiligung des Arbeitsgedächtnisses bewältigt wird, müssen nebenbei informative akustische Ausgaben wahrgenommen werden. Ferner muss eine Entscheidung getroffen werden, wie mit dieser Information zu ver-

fahren ist, ohne die Primäraufgabe zu vernachlässigen. Die akustischen Systemmeldungen müssen so gestaltet werden, dass die Interferenzwirkung minimiert wird.

Hypothesenkonform zeigte sich im Experiment, dass lange sprachliche Ausgaben in Form von Sätzen vor diesem Hintergrund weniger geeignet sind als die anderen getesteten Lösungen. Das Resultat ist insbesondere deshalb interessant, weil nicht die gesamte auditive Nachricht relevant war. Der bedeutungstragende Teil befand sich in Form eines Schlüsselworts zu Beginn der Ausgabe. Dieses Schlüsselwort war mit dem Begriff aus der Schlüsselwort-Bedingung identisch, die verglichen mit der Kontrollbedingung Stille nicht gestört hat. Der Rest der langen sprachlichen Ausgabe durfte von den Versuchsteilnehmern ignoriert werden. Das bedeutet, selbst wenn nur ein Teil einer längeren Sprachnachricht relevant ist, werden die Nutzer dennoch von dem irrelevanten Teil abgelenkt. Mit höchster Wahrscheinlichkeit wäre der nachteilige Einfluss der langen Sprachausgabe noch deutlicher hervorgetreten, wenn die Bedeutung des Satzes nicht schon nach dem ersten Wort, sondern erst mitten im Satz klar geworden wäre. Denn in diesem Fall wären die Anforderungen an die Aufmerksamkeitsteilung deutlich höher gewesen. Das wirft ein besonderes Licht auf die grammatikalische Gestaltung sprachlicher Ausgaben. Die Empfehlung liegt nahe, möglichst kurze Ausgaben zu konstruieren und notfalls sprachästhetische Gesichtspunkte zu vernachlässigen. Diese Empfehlung wird von Grundlagenbefunden gestützt, die das Ausmaß des ISE in direkten Zusammenhang mit der Anzahl der Wörter im Hintergrundschaall stellen konnten (Bridges & Jones, 1996).

Bei keiner der nicht-sprachlichen Gestaltungsvarianten konnte ein störender Einfluss auf die Arbeitsgedächtnisaufgabe nachgewiesen werden, wie der Vergleich mit der Baseline ergab. Wie Banbury und Kollegen (2001) ausführen, produzieren nicht-sprachliche Schalle dann ähnliche Effekte wie Sprache, wenn sie eine merkbliche Variation im Schallmuster aufweisen. Zumindest für Earcons ist das eine typische Eigenschaft. Insofern ist der Befund bemerkenswert, dass sich auch mit Earcons kein Störeffekt produzieren ließ. Allerdings ist bei der geringen Dauer keine besonders anhaltende akustische Variation zu etablieren, wie sie vermutlich notwendig gewesen wäre. Höchstwahrscheinlich ist aufgrund der Kürze typischer Earcons oder Auditory Icons keine Interferenz aufgetreten. Entsprechend kann argumentiert werden, dass bei längeren nicht-sprachlichen Ausgaben ein analoges Ergebnismuster wie bei den sprachlichen Stimuli denkbar wäre, bei denen ein kurzes Schlüsselwort nicht gestört hat während ein langer Satze sehr wohl nachteilige Effekte hatte. Dabei muss jedoch beachtet werden, dass es ein grundsätzliches Gestaltungsprinzip ist, Earcons und Auditory Icons zeitlich kurz zu halten (vgl. Hempel & Altinsoy, 2005). Um diese Schlüsselanforderung auch für komplexe Audiobotschaften zu realisieren, wurde mit so genannten parallelen Earcons eigens eine Lösung entwickelt (Brewster, Wright & Edwards, 1995). Ein Effekt besonders langer nicht-sprachlicher Ausgaben wäre aus

Sicht der angewandten Forschung deshalb wenig relevant. Längere Earcons kommen in der Praxis nur dadurch zu Stande, dass das Signal durch Halleffekte oder einen Nachklang verlängert ist. Eine Dauer von über ein bis zwei Sekunden wird jedoch auch dadurch nicht erzielt.

Zusammengenommen konnte gezeigt werden, dass lange Sprachausgaben mit den hier untersuchten Arbeitsgedächtnisprozessen interferieren, während Schlüsselwörter, Earcons und Auditory Icons das nicht tun. Die fehlende Störwirkung wurde mit einem Nulleffekt gegenüber der Baseline belegt. Wie bei allen Nulleffekten kann das natürlich nicht als zweifelsfreier Beleg gewertet werden. Dennoch gibt es guten Grund zu der Annahme, dass der gefundene Nulleffekt die wahre Sachlage repräsentiert. Erstens konnte selbst ohne Korrektur für multiple Tests bei diesen Ausgabeformen kein Effekt festgestellt werden. Zweitens ist die fehlende Störwirkung eines einzelnen Wortes und der kurz andauernden Earcons und Auditory Icons durchaus im Einklang mit der existierenden Literatur zur Störwirkung von Schallen auf das Arbeitsgedächtnis (insbes. Bridges & Jones, 1996). Darüber hinaus sind die gefundenen Effekte durchwegs schwach und die Streuung der Daten ist nicht groß. In der Praxis wären die Unterschiede also von relativ geringer Bedeutung. Ferner gestattet die Statistik p_{rep} durch das Einbeziehen zusätzlicher Information in den Test nach Killeen (2005b) eine weitergehende Interpretation als einfache p -Werte. Ein p_{rep} -Wert zwischen .2 und .8 kann so interpretiert werden, dass kein guter Hinweis für irgendeinen Unterschied in den Daten existiert. Das war durchgehend der Fall.

Leistung bei der Wahlreaktionsaufgabe

Die Daten aus der Wahlreaktionsaufgabe geben Auskunft über den Prozess des semantischen Dekodierens auditiv präsentierter Information. Die benötigte Zeit zum Verstehen der Bedeutung wird durch die Reaktionszeiten erfasst. Fehler bei der Tastenauswahl werden als Maß der Güte der Assoziation zwischen Stimulus und Antwort analysiert. Nicht überraschend war der Befund eines fehlenden Unterschieds in Reaktionszeiten und Fehlern zwischen den sprachlichen Varianten der Ausgabe. Schließlich war die Systemnachricht in Satzform so konstruiert worden, dass bereits zu Beginn der Ausgabe die Bedeutung klar wurde. In gewisser Hinsicht waren die Antwortzeiten bei den Sätzen somit trivial kurz. Ist eine Nachricht nicht in einem Schlüsselwort subsumierbar und muss der Botschaft von Beginn bis zum Ende vollständig zugehört werden, wären entsprechend sehr lange Reaktionszeiten zu erwarten, bzw. forcierbar. In diesen Fällen ist ein klarer Vorteil nicht-sprachlicher Stimuli aufgrund ihrer hohen und komprimierten Ausdruckskraft zu erwarten (vgl. Brewster, 2003). Der Aspekt der Konstruktion sprachlicher Ausgaben stand jedoch nicht im Fokus der Untersuchung. Für die Vergleiche der Reaktionszeiten waren primär die auditiven Nachrichten vergleichbarer Länge relevant, d.h. Schlüsselwörter gegen nicht-sprachliche

Stimuli. Erwähnenswert bei den sprachlichen Ausgaben ist das gleichzeitige Vorliegen eines Nulleffekts in Reaktionszeiten und Fehlern, während ein deutlicher Unterschied zwischen diesen Varianten bei dem Einfluss auf das Arbeitsgedächtnis vorlag.

Der direkte Vergleich von Auditory Icons, Earcons und Schlüsselwörtern unterstützte die Hypothese, dass ein inhärenter Zusammenhang zwischen bezeichnendem auditiven Stimulus und bezeichnetem Ereignis oder Objekt für einen reibungslosen Ablauf des Interaktionsprozesses äußerst sinnvoll ist. Während sich weder auf der Ebene der Reaktionszeiten noch bei den Fehlerraten nachweisbare Unterschiede zwischen Schlüsselwörtern und Auditory Icons ergaben, ließ sich zwischen diesen Bedingungen auf beiden Dimensionen ein signifikanter Vorteil gegenüber den Earcons zeigen. An sich ist das Ergebnis zwar zu erwarten, da das Mapping bei Earcons arbiträr ist. Aber nachdem die Versuchsteilnehmer ein ausführliches Stimulus-Response-Training erhielten, überrascht die Fehlerrate durchaus. Aufgrund des Kriteriums ein Element dreimal in Folge korrekt identifizieren zu müssen, ist unwahrscheinlich, dass das Training nicht ausreichend war. Darüber hinaus kann für gewöhnliche Anwendungssituationen und Nutzer nicht beliebig viel Trainingsaufwand spendiert werden. Im Bereich der Mensch-Computer-Interaktion werden Earcons mit gleichzeitiger visueller Ausgabe genutzt. Zur Optimierung der Interaktion mit koordinierter grafischer Darstellung des Ereignisses werden Earcons basierend auf den vorliegenden Ergebnissen eher empfohlen als für eine unimodale Kodierung von Ereignissen. In einer solchen Anwendung wird sich die Bindungslosigkeit bei der Repräsentation von Earcons als Vorteil z.B. zur Repräsentation abstrakter Vorgänge als Vorteil nutzen lassen.

Die Datenlage deutet darauf hin, dass im Vergleich zu den anderen auditiven Stimuli bei Earcons eine deutlich höhere kognitive Aktivität bei der semantischen Einordnung des Reizes mit gleichzeitig geringerer Erfolgswahrscheinlichkeit stattgefunden hat. Dennoch war überraschenderweise die Fehlerrate bei der Arbeitsgedächtnisaufgabe im Vergleich zur Baseline nicht erhöht. Im Sinne einer Interpretation nach dem multiplen Ressourcenmodell (Wickens, 1984, 2002) liegt die Vermutung nahe, dass unabhängige kognitive Ressourcen bei der Verarbeitung auditiver Stimuli und dem gleichzeitigen Aufrechterhalten von Information im Arbeitsgedächtnis – in der phonologischen Schleife – beteiligt sind. Die detaillierte Analyse des Zusammenhangs zwischen Reaktionszeiten und Fehler bei der seriellen Reproduktion mit Hilfe eines Versuchspersonen-Splits ergab jedoch, dass eine derartige Verallgemeinerung nicht zulässig ist, da deutlich erhöhte Reaktionszeiten sehr wohl mit einer schlechteren Erinnerungsleistung bei der Arbeitsgedächtnisaufgabe einhergehen. Bei dieser Analyse konnte auch sichergestellt werden, dass die Fehlerrate bei der Wahlreaktionsaufgabe nicht in einem derartigen Zusammenhang mit der Leistung bei *Serial Recall* steht. Überraschend an dem Resultat des Splits ist insbesondere, wie gering die Fehlerrate bei der seriellen Reproduktion bei kurzen Reaktionszeiten ist. Wenn sich der kognitive

Aufwand bei der Verarbeitung der Earcons reduzieren ließe, könnten sie sogar zu einer besseren Behaltensleistung als die anderen auditiven Varianten führen. In der Praxis ließe sich das beispielsweise durch eine geringe Anzahl repräsentierter Ereignisse und auditiv sehr verschiedenartige Earcons oder durch eine redundante Ausgabe in einer anderen Modalität realisieren.

4.2.3 Zusammenfassende Diskussion und Gestaltungsempfehlungen

Das Experiment untersuchte die Eignung typischer Gestaltungsvarianten sprachlicher und nicht-sprachlicher akustischer Ausgaben zur Übermittlung nicht-kritischer Systemnachrichten. Dabei standen Überlegungen für den Einsatz in einer Mehraufgabensituation mit stark ausgeprägter visueller Komponente, wie sie beispielsweise bei der Fahrzeugführung vorliegt, im Mittelpunkt der Versuchsplanung. Wenn nicht-kritische Ereignisse auditiv gemeldet werden, sollte die Ausgabe so beschaffen sein, dass sie trotz des aufmerksamkeitsfordernden Charakters akustischer Benachrichtigungen möglichst wenig stören. Nichtsdestotrotz sollten sie schnell wahrnehmbar und verständlich sein. Zur Evaluation dieser Komponenten wurden Auditory Icons, Earcons, kurze Sprachnachrichten in Form von Schlüsselwörtern und lange Sprachausgaben als Sätze in einer Aufgabe zum seriellen Arbeitsgedächtnis mit einer eingebetteten Wahlreaktionsaufgabe getestet. Anzustreben ist bei der Gestaltung eine möglichst geringe Störwirkung von Arbeitsgedächtnisprozessen.

Hintergrund des Experiments ist nicht zuletzt die Fragestellung, in welcher Form auditive Information in einer multimodalen Arbeitsumgebung präsentiert werden soll, wenn eine Ausgabe in dieser Modalität vom Nutzer ohne Zuhilfenahme weiterer Modalitäten interpretierbar sein soll. Die Reaktionszeiten der Wahlreaktionsaufgabe geben darüber Aufschluss, wie lange es dauert, die Nachricht zu verstehen, während die Fehlerrate anzeigt, wie gut die Nachricht generell verstanden wurde. Die Ergebnisse lassen sich folgendermaßen zusammenfassen und können in dieser Form einen Beitrag zur Gestaltung künftiger Mensch-Maschine-Schnittstellen leisten:

- Sprachausgabe – vollständige Sätze: Nachrichten in Satzform wurden generell gut verstanden und führten kaum zu Eingabefehlern. Die notwendige Dauer, um die Bedeutung eines Satzes zu verstehen, hängt maßgeblich von dessen Konstruktion ab. Sinnvoll ist sicherlich, zu Beginn in möglichst knapper Form die grundlegende Botschaft zu kommunizieren und weitere Erklärungen in einen Nachsatz zu geben. In jedem Fall führt diese Gestaltungsvariante zur größten Interferenz mit Arbeitsgedächtnisprozessen und sollte deshalb mit großer Umsicht eingesetzt werden. Wenn Sätze für nicht-kritische Informationsausgaben Anwendung finden, sollten diese so kurz wie möglich formuliert

- werden. Das mag trivial klingen, aber bei vielen gegenwärtig verfügbaren Systemen mit Sprachausgabe wird diesem Aspekt zu wenig Beachtung geschenkt.
- Sprachausgabe – Schlüsselwörter: Akustische Ausgaben in Form von Schlüsselwörtern haben sich auf allen erhobenen Maßen bewährt. Sie interferieren im Vergleich mit einer Baseline-Bedingung nicht mit dem Aufrechterhalten von Information im Arbeitsgedächtnis, führen zu einer sehr geringen Reaktionsfehlerrate und können schnell verarbeitet werden. Als Einschränkung muss genannt werden, dass sich nicht alle Systemereignisse in ein oder zwei Schlüsselwörtern ausdrücken lassen.
 - Nicht-sprachliche Ausgabe – Auditory Icons: Wie die Schlüsselwörter erwiesen sich auch Auditory Icons im Test als vorteilhaft. Sie wurden schnell erkannt, führten zu wenigen Verständnisfehlern und störten den Rehearsal Prozess im Arbeitsgedächtnis ebenfalls kaum. In der Praxis kann es sich jedoch als schwierig erweisen, stets intuitive natürliche Geräusche zur Abbildung von Systemereignissen zu finden. Insbesondere abstrakte Ereignisse lassen sich nicht ohne weiteres durch Auditory Icons ausdrücken.
 - Nicht-sprachliche Ausgabe – Earcons: Earcons beeinträchtigten zwar das Arbeitsgedächtnis nicht mehr als Auditory Icons oder Schlüsselwörter, führten aber zu einer deutlich schlechteren Leistung bei der Wahlreaktionsaufgabe. Sie wurden sowohl deutlich langsamer als auch mit geringerer Wahrscheinlichkeit richtig erkannt.

Diese Kriterien für den Einsatz akustischer Ausgaben basieren auf den Erfahrungen des vorliegenden Experiments. Die Entscheidung für oder gegen eine dieser Gestaltungslösungen ist darüber hinaus von konkreten Randbedingungen des anvisierten Einsatzbereichs abhängig. So konnte im Experiment zwar gezeigt werden, dass die Zeit zur Verarbeitung und Reaktion auf Earcons signifikant länger ist, ob aber das Ausmaß des Unterschieds von ca. 500 ms relevant ist, hängt von der Art der übermittelten Information ab. Wenn tatsächlich die Übertragung nicht-kritischer Information das Ziel ist, können die zeitlichen Nachteile und sogar die höhere Fehlerrate in den Hintergrund treten. Von zentralem Interesse sind häufig Designkriterien wie die Fähigkeit von Earcons, hierarchische Strukturen zu repräsentieren (Lepître & Brewster, 2000) sowie die Möglichkeit, bei späteren Erweiterungen eines Systems eine konsistente Klangzuordnung zu erhalten. Dabei sollte jedoch nicht vernachlässigt werden, dass auch ein längeres Nachdenken über die Bedeutung von Earcons mit Kosten verbunden ist, wie der Versuchspersonen-Split in der Detailanalyse gezeigt hat. Ein weiteres Hindernis kann der Lernaufwand zum Verstehen der Signalkodierung sein.

Bei der Verwendung verbaler Ausgaben beeinflusst die Designüberlegungen, dass längere Sprachausgaben nicht nur auf das Arbeitsgedächtnis nachteilig wirken. Gerade

bei Bordsystemen in Kraftfahrzeugen spielen Komfortaspekte eine große Rolle. Eine Konversation der Fahrzeuginsassen wird durch die Ausgabe vollständiger Sätze deutlich stärker beeinträchtigt als durch kurze Schlüsselwörter. So konnte gezeigt werden, dass Ausgaben von einer Länge zwischen 4 und 8 Silben schnellen Informationstransport gewährleisten und deutlich weniger mit gleichzeitiger sprachlicher Kommunikation der Nutzer interferieren als längere Ausgaben (Simpson, McCauley, Roland, Ruth & Williges, 1987). Gleichzeitig sollte beachtet werden, dass mehrsilbige Wörter besser verstanden werden als einsilbige Wörter (National Research Council, 1997). Ein altbekannter Befund ist ferner, dass Wörter in Sätzen verständlicher sind als isolierte Wörter (Miller, Heise & Lichten, 1951). Diese Eigenschaften treten besonders bei ungünstigen Umweltbedingungen wie Hintergrundlärm zu Tage. Wenn keine maskierenden Schalle zu erwarten sind oder durch einen klaren Interpretationskontext die Auswahlmenge der möglichen Schlüsselwörter begrenzt wird, sollten sich mehrsilbige Schlüsselwörter bestens eignen. Viele der früheren technischen Überlegungen zur Verständlichkeit sprachlicher Ausgaben für die Mensch-Maschine-Interaktion, z.B. dass synthetische Sprache mehr Verarbeitungsressourcen benötigt als natürliche Sprache (Pisoni, 1982), dürften durch den Fortschritt bei der Ausgabequalität überholt sein.

Für den Anwendungsfall der akustischen Ausgabe nicht-kritischer Information durch ein Bordsystem im Fahrzeug empfehlen sich zusammenfassend kurze Schlüsselwörter und/oder griffige Auditory Icons. Bei Auditory Icons überwiegen die gefundenen Vorteile den Nachteil, dass nicht alle Ereignisse durch ein natürliches Geräusch abgebildet werden können. Bei typischen Bordsystemen gibt es nur sehr wenige derart abstrakte Ereignisse und Auditory Icons können gegebenenfalls mit Schlüsselwörtern kombiniert werden. Darüber hinaus erfordern beide Lösungen keinen oder kaum Lernaufwand, um die Bedeutung einer Nachricht in einem dem Nutzer bekannten Kontext zu verstehen. Abschließend soll auf einige grundlegende Randbedingungen für die zielgerichtete Gestaltung nicht-kritischer akustischer Ausgaben hingewiesen werden (vgl. Hempel, 2002). Die Rückmeldung muss den Erwartungen und Anforderungen des Nutzers entsprechen sowie zugleich sinnvoll, unmissverständlich und möglichst intuitiv sein. Ferner muss eine auditive Rückmeldung zu produktinhärenten Geräuschen passen und darin wahrnehmbar sein. Nutzer des Systems sollten stets in der Lage sein, diese Ausgaben schnell und unaufwändig abzustellen (Stummschaltung).

5 Untersuchungsabschnitt 2: Aspekte der Eingabe im multimodalen Umfeld

Der zweite Untersuchungsabschnitt befasst sich mit dem Themenbereich der Gestaltung für Eingabe im multimodalen Umfeld zum Einsatz im Fahrzeug. In der Diskussion um Gebrauchstauglichkeit und Zugänglichkeit technologischer Innovationen wie Fahrerinformationssystemen und Komfortfunktionen taucht Sprachbedienung immer häufiger als viel versprechende Maßnahme auf, Technik bedienerfreundlich zu gestalten. Ausschlaggebend ist in der Argumentation häufig, dass durch sprachbasierte Interaktion weniger Ablenkung des Fahrers von seiner Primäraufgabe zu erwarten ist und dass sie diesem gestattet, auf besonders einfache und direkte Weise seine Intention auf die Funktionen einer sprachgesteuerten Anwendung abzubilden (Akyol et al., 2001; Bengler, 2001; Pieraccini et al., 2004; Schattenberg & Debus, 2001). Aufgrund dieser Schlüsselrolle der Modalität Sprache steht diese im Fokus des Untersuchungsabschnitts zur Eingabe. Der Erfolg von Sprachbedienung im Automobil und das Ausschöpfen des genannten Potentials hängen jedoch kritisch von einer umsichtigen gestalterischen Umsetzung ab, die die Fähigkeiten, Erwartungen und Präferenzen der Nutzer berücksichtigt.

Die Experimente des vorliegenden Untersuchungsabschnitts widmen sich der Optimierung einer konkreten Ausprägung einer Gestaltungsvariante für sprachgesteuerte Anwendungen. Neben den in Abschnitt 2.3.3.2 erwähnten Gestaltungsaspekten ist es eine der großen Herausforderungen für das Design sprachgesteuerter Systeme, den Nutzer wissen zu lassen, welche Sprachbefehle zu einem gegebenen Zeitpunkt möglich sind. Das so genannte *Say What You See* (SWYS) Prinzip ist eine bekannte und häufig angewandte Lösungsstrategie für multimodale Interaktion mit visueller Ausgabe (vgl. Yankelovich, 1996): Der Nutzer kann jedes Kommando sprechen, das z.B. in Menüs oder auf Schaltflächen lesbar ist. Eine gut strukturierte Benutzeroberfläche mit hinreichend großen und gut ablesbaren Beschriftungen vorausgesetzt, ist diese Lösung sehr gut geeignet für Komfort- und Infotainmentsysteme im Fahrzeug, die über eine abgesetzte grafische Ausgabe in einem integrativen Anzeige- und Bedienkonzept

verfügen. Dadurch wird auch unerfahrenen Nutzern ohne zusätzliche Hilfsfunktion Sprachbedienung ermöglicht und das System muss nur bei unvollständiger oder nicht eindeutiger Eingabe durch gerichteten Dialog oder visuelle Ausgabe eine Klärung einleiten (Pieraccini et al., 2004). Gleichzeitig wird eine Einschränkung des nutzerseitig verwendeten Vokabulars erreicht, die für gute Erkennungsergebnisse notwendig ist. Ohne entsprechende Gestaltungsmaßnahmen ist bei der Interaktion die für menschliches Kommunikationsverhalten typische Variabilität des Wortschatzes zu erwarten (Furnas, Landauer, Gomez & Dumais, 1987). Zusätzlich entsteht die Möglichkeit, die relevanten Begriffe bei manuell-visueller Erkundung des Systems nebenbei zu erlernen. Diese Darstellungsform kann dazu beitragen, dem Nutzer auch verständlich zu machen, warum bei fehlerhafter Spracherkennung aufgrund unbekanntes Vokabulars ein Befehl nicht erkannt wurde (Schattenberg & Debus, 2001). Allerdings weist das Konzept einige Probleme auf, die durch mehrere Experimente adressiert werden sollen.

Die anwendungsorientierte Untersuchung dieses Abschnitts evaluiert eine Option, die Defizite der Eingabeeffizienz unter SWYS ausgleichen soll. Viele Bordsysteme besitzen eine tiefe, hierarchisch gegliederte Menüstruktur zum Zugriff auf Komfortfunktionen, da dies sowohl Vorteile für die Spracherkennungsqualität durch Einschränkung des pro Schritt aktiven Vokabulars als auch für die Erlernbarkeit des Systems durch unerfahrene Nutzer bringt. Existieren weniger Auswahlmöglichkeiten an einem Knotenpunkt, ist der Planungs- und Entscheidungsaufwand geringer (Norman, 1988). SWYS setzt bei komplexen Systemen notwendigerweise auf menübasierte Navigation, da sonst alle Sprachbefehle als textuelle Elemente gleichzeitig auf dem Bildschirm angezeigt werden müssten. Allerdings unterstützt der menübasierte Aufbau nicht die Bedürfnisse erfahrener Nutzer nach zeiteffizienter Navigation. Wenngleich es möglich ist, manche Funktionen durch sprachlichen Direktzugriff schnell verfügbar zu machen, verbleiben viele Funktionen bei SWYS nur durch schrittweises Navigieren in der Menühierarchie erreichbar. Auf jeder Menüebene tritt jedoch eine technisch bedingte Latenzzeit durch den Spracherkennungsvorgang auf, die sich kumuliert deutlich negativ auf die Bedieneffizienz der Spracheingabe auswirkt. In einem Simulationsexperiment wurde eine Expertenoption getestet, die als multimodales *Talk-Ahead* beschrieben werden kann und die aufgrund lernpsychologischer Annahmen zum Ausgleich dieses Defizits viel versprechend erscheint.

Die grundlagenorientierte Untersuchung setzt sich mit der visuellen Ausgabe und Strukturierung von Information auseinander. Zur Steigerung der Memorierbarkeit und für eine möglichst schnelle Orientierung werden häufig piktografische Funktionsdarstellungen empfohlen (z.B. Stauer, 1987). Um die Dauer der visuellen Abwendung des Fahrers von der Straße zu minimieren, ist diese Designoption für manuelle Eingabe möglicherweise einer textuellen Beschriftung von Steuerelementen vorzuziehen.

Das hätte jedoch auch zur Folge, dass sprachbasierte Interaktion ausschließlich über die auditive Schnittstelle erfolgen und die *Say What You See* Eingabemethode aufgegeben werden müsste. Diese hat sich aber aus Sicht der Gebrauchstauglichkeit sehr gut bewährt. Eine tiefere Betrachtung der relevanten Forschungsarbeiten zeigt jedoch, dass der geschilderte Vorteil kürzerer Suchzeiten für bildhafte Funktionsangabe in der Literatur nicht einheitlich nachvollziehbar ist. Es stellt sich die Frage, inwiefern die Angabe verbaler Beschriftungen tatsächlich zu längeren visuellen Suchzeiten führt. In mehreren Experimenten wurde dieser Sachverhalt systematisch untersucht und dabei auf die relevanten Experimente und Paradigmen zu Grundlagenbefunden bei visueller Suche und auf Studien zur Wirkung piktografischer und verbaler Darstellung Bezug genommen.

5.1 Multimodales Talk-Ahead bei Spracheingabe

5.1.1 Problemstellung und Ausgangspunkte des Experiments

Dillon und Norcio (1997) führten Untersuchungen zum Verhalten verschiedener Nutzergruppen bei Spracheingabe im Kontext medizinischer Diagnostik und Dokumentation im Krankenhaus durch. Unerfahrene Nutzer ohne Vorerfahrung mit Spracheingabe profitierten dabei von einer hierarchischen Menüstrukturierung, die zu jedem Zeitpunkt nur eine begrenzte Auswahl von Kommandos anbietet. Erfahrene Bediener hingegen nahmen dieses schrittweise Navigieren als träge und nicht effektiv wahr. In vielen Fällen kann durch kontextunabhängig gültige Kommandos ein Direktzugriff auf tiefer im Menü befindliche Funktionen ermöglicht werden. Ist dem Nutzer das entsprechende Kommandowort bekannt, kann er beispielsweise in einem Fahrzeug-Infotainmentsystem durch Angabe des Schlüsselwortes „CD“ bereits im Hauptmenü auf den CD-Player zugreifen, ohne erst in den Entertainment-Kontext wechseln zu müssen. Diese Möglichkeit ergibt sich jedoch nur, wenn das Kommando ohne nähere Angabe des Kontextes bereits eindeutig ist oder wenn nur ein Parameter zur vollständigen Funktionsausführung benötigt wird. Zwei Beispiele sollen zur Beleuchtung dieser Sonderfälle dienen. Erstens: Der Befehl „Titel fünf“ im Kontext des Hauptmenüs ist zur Einstellung des gewünschten Musiktitels nicht unbedingt eindeutig, sofern noch andere titelbasierte Auswahloptionen existieren. So kann sich „Titel fünf“ auch auf einen im Bordsystem integrierten MP3-Player beziehen. Es ist damit notwendig, zunächst in den CD-Kontext zu wechseln und erst dann die Titelangabe vorzunehmen sowie gegebenenfalls noch einen Startbefehl abzusetzen. Zweitens: Soll in einem mehrstufigen Dialog, wie er typischerweise zur Einstellung des Navigationsziels über die Adresseingabe vorliegt, das Fahrziel festgelegt werden, muss dem Sys-

tem der Zielort, die Zielstraße und möglicherweise die Hausnummer übergeben werden. Je nach System muss vor diesen Schritten erst explizit in den Navigationskontext gewechselt werden. Natürlich wäre eine grundsätzlich kontextfreie Verwendbarkeit aller Kommandos aus Nutzersicht optimal, allerdings lässt sich diese in zunehmend komplexen Systemen schwer verwirklichen. Darüber hinaus ergab sich in Studien zu entsprechenden Systemen, dass die Nutzer ohnehin dazu tendieren, den Kontext explizit anzugeben (vgl. Neuss, 2000).

Durch umsichtiges Sprachdialog-Design kann man die Anzahl der notwendigen Einzelkommandos teilweise minimieren. Trotzdem bleibt ein technisch bedingter Faktor vorhanden, der maßgeblich zur Wahrnehmung der Trägheit eines solchen System beiträgt: Bei jeder Spracheingabe existiert eine Latenzphase von einigen hundert Millisekunden, bis das Spracherkennungssystem reagiert. Da diese Latenz auf jeder Menüebene bzw. bei jedem einzelnen Bedienschritt eintritt, wirkt sie sich kumuliert stark negativ auf die Bedienerffizienz der Sprachinteraktion aus. Je nach Umfang des aktiven Vokabulars kann die Latenz nach einem einzelnen Dialogakt zwischen 300 und 1000 ms, teilweise sogar darüber liegen. Im Detail bedeutet das in einem Menüsystem mit SWYS Konzept, dass zunächst ein Kommando gesprochen wird und der Nutzer dann auf das Spracherkennungssystem für Erkennung und Ausführung wartet und ihm erst im Anschluss von der grafischen Benutzeroberfläche die nächste zur Verfügung stehende Menge von Kommandos angeboten wird⁸. Ein derartiger Prozess kann schnell eine kritische Grenze des Zeitkonsums übersteigen und insbesondere von erfahrenen Nutzern durch die Anzahl separater Schritte als unbefriedigend angesehen werden.

Dieser Umstand lässt sich in seiner Auswirkung durch die Integration psychologischen Wissens bei der Systemgestaltung deutlich abmindern. Ausgangspunkt der Überlegung muss dabei der Erhalt der bewährten Gliederung des Systems in hierarchische Menüabschnitte zur Unterstützung der Erlernbarkeit sein. Auch Norman (1988) betont, dass eine gut strukturierte Menüführung besonders bei vergleichsweise selten genutzten Funktionen sinnvoll ist: Je weniger Optionen pro Knotenpunkt existieren, desto weniger Planungsaufwand ist erforderlich. Der weiterführende Lösungsansatz besteht darin, eine Verknüpfung von Sprachkommandos zu ermöglichen, um den Interaktionsprozess zu beschleunigen. Basierend auf der Annahme, dass mit zunehmender Erfahrung die Sequenz aufeinander folgender Kommandos immer besser memoriert wird und diese Sequenzen einzelner Kommandos mit der Zeit in einer Art

⁸ Derartige Wartezeiten bei Sprachsteuerung treten nicht nur in hierarchischen Menüsystemen auf. Auch innerhalb einer Menüebene kann durch Angabe eines Exekutivkommandos mit daran anschließender Parameterspezifikation diese Latenz entstehen. Doch gerade bei mehrfachen Menüwechseln kommt es durch kontextabhängige Umstellungen in Vokabular und Grammatik zu relativ langen Latenzen, weshalb dieser Fall hier besonders betont wird.

Chunk-Bildung (Servan-Schreiber & Anderson, 1990) zu Entitäten integriert werden, sollte die Zeit zwischen einzelnen Sprachkommandos mit zunehmender Erfahrung der Nutzer kürzer werden. Das bedeutet, dass aus Nutzersicht mehrere oder sogar alle Kommandos, die in der sequentiellen Menühierarchie aufeinander folgen, als *ein* Kommando gegeben werden können, ohne den entsprechenden Bildschirmaufbau abzuwarten. Abbildung 28 veranschaulicht das Prinzip für obiges Navigationsbeispiel.

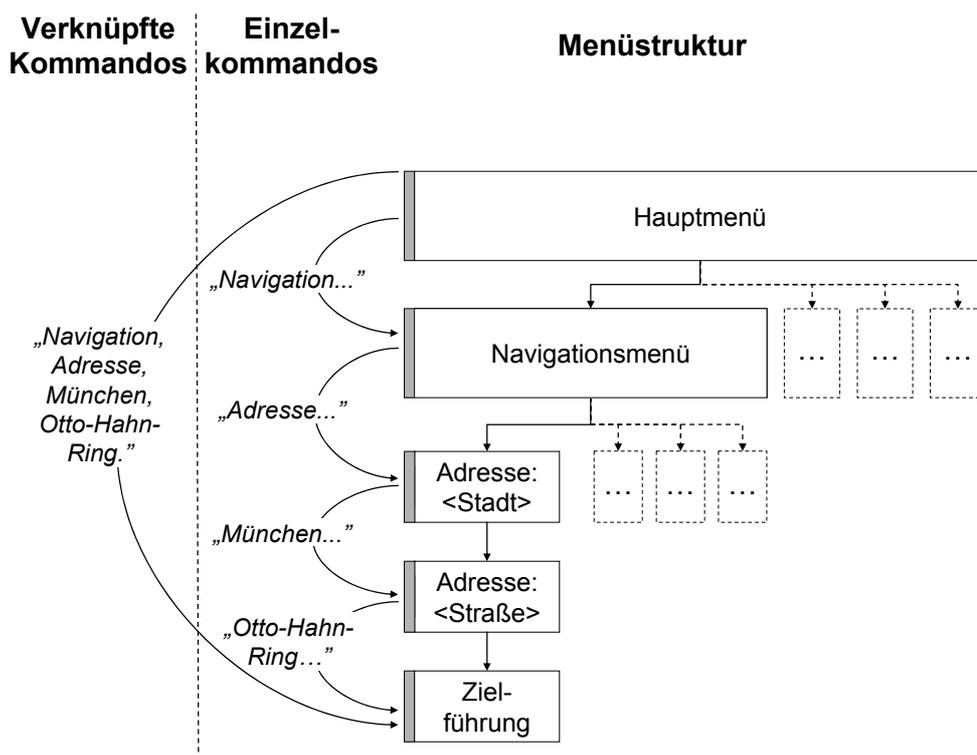


Abbildung 28: Das *Talk-Ahead* Prinzip gestattet die Verknüpfung von Spracheingaben für fortgeschrittene Nutzer (Darstellung: Maximalbeispiel).

Statt die Folge „Navigation“ [Pause] „Adresse“ [Pause] „München“ [Pause] „Otto-Hahn-Ring“ zu sagen, kann der Nutzer eine Kette bilden, die im Extremfall aus allen vier Elementen besteht „Navigation, Adresse, München, Otto-Hahn-Ring“. Selbstverständlich muss es auch möglich sein, nur Teilschritte in Ketten zusammenzuführen und die Reihenfolge der Kommandos zu variieren. Für eine benutzerfreundliche Ausführung ist darüber hinaus entscheidend, dass der Spracherkenner Keyword Spotting beherrscht. So sollte z.B. nach einem Wechsel zur Adresseingabe der Nutzer sagen können: „Ich möchte nach München in den Otto-Hahn Ring“. In dieser Aussage stellen nur der Ort und die Straße relevante Schlüsselwörter da, die anderen Teile des Satzes sollten vom System ignoriert werden. Dabei ist entscheidend, dass die Zahl von Ersetzungsfehlern durch Keyword-Spotting nicht ansteigt.

Bereits mit existierenden Spracherkennungstechnologien kann bei entsprechender Gestaltung diese Eingabevariante des Zulassens von verknüpften Sprachkommandos aus mehreren Menüebenen unterstützt werden. Da es für die Güte der Spracherken-

nungsqualität von Vorteil ist, auf einer Menüebene immer nur die dort benötigte Teilmenge des verfügbaren Gesamtvokabulars aktiv zu halten, sollte dieser Umstand bei der Implementierung der Verknüpfungsoption berücksichtigt werden. Dies kann dadurch geschehen, dass die Spracheingabe zwischengespeichert und anschließend seriell abgearbeitet wird: Mit dem internen Erreichen jeder neuen Menüebene erfolgt ein Zurücksetzen des Zeigers im noch abzuarbeitenden Sprachsignal, d.h. die bereits abgearbeiteten Begriffe werden ignoriert, und die verbleibenden Begriffe werden mit Hilfe des durch die neue Menüebene aktualisierten Vokabulars und der entsprechenden Grammatik neu analysiert. Die Reihenfolge der Kommandos kann in diesem Fall natürlich nicht tolerant gehandhabt werden. Zur Abgrenzung von Spracherkennungsvarianten, die den erhaltenen Input sofort verarbeiten oder ein zwischengespeichertes Signal analysieren, unterscheidet Neuss (2000) zwischen einstufiger Online- und potentiell mehrstufiger Offline-Erkennung, die quasidynamische Vokabularanpassung gestattet.

Im Bereich sprachgesteuerter Telefonservices, die ohne grafische Rückmeldung rein auf die auditive Schnittstelle angewiesen sind, ist diese Form der beschleunigten Eingabe bereits fest im Katalog anerkannter Gestaltungsrichtlinien als *Talk-Ahead* Prinzip verankert (Gardner-Bonneau, 1999). Der Integration des *Talk-Ahead* Prinzips in die *Say What You See* Strategie und damit die Einsetzbarkeit für multimodale Interaktion mit grafischer und/oder akustischer Rückmeldung kann zur Unterscheidung von Voice-only Systemen durch den Terminus „multimodales *Talk-Ahead*“ Rechnung getragen werden.

Zur Gestaltung multimodaler Eingabe im Fahrzeug wurde *Talk-Ahead* bereits in der Arbeit von Neuss (2000) und in ähnlicher Form von Schattenberg und Debus (2001) vorgeschlagen. Neuss spricht dabei von sprachlichen „Shortcuts“. Dieser Begriff ist jedoch etwas unscharf, da er nicht eindeutig macht, ob ein einfacher kontextfreier Sprachbefehl einen abgekürzten Bedienweg erwirkt, oder ob die Abkürzung durch eine Verknüpfung mehrerer Exekutivkommandos und Parameter in einer Äußerung erfolgt. Über die Auswirkungen von multimodalem *Talk-Ahead* für diesen Anwendungsfall existiert bisher wenig systematisch experimentell gewonnenes Wissen. Wenngleich Neuss (2000) *Talk-Ahead* implementiert und in einer Untersuchung neben anderen Systemvariablen evaluiert hat, war es doch nicht alleiniger Schwerpunkt des Tests. Da das Verhalten von Nutzern bei multimodaler Interaktion mit einem Bordcomputer und dabei entstehende Nutzungsmuster im Vordergrund standen, konnte die Anwendung der *Talk-Ahead* Option nur in diesem Kontext diskutiert werden. Wie erwartet wurde *Talk-Ahead* von den Versuchspersonen gut angenommen und zeigte das Potential, den Dialogablauf zu beschleunigen. Problematisch hingegen war, dass die Versuchspersonen häufig das Vokabular vergaßen und deshalb Hilfsanfragen an das System stellen mussten. Dadurch lässt sich nicht klar erkennen, ob *Talk-Ahead*

tatsächlich nutzbar ist und ob sich die zeitlichen Vorteile wirklich einstellen können. Es erscheint auch unwahrscheinlich, dass die Versuchspersonen im Rahmen eines relativ kurzen Experiments mit vielen Eingabemöglichkeiten bezüglich einer besonderen Variante einer Schnittstelle eine gute Nutzungsstrategie entwickeln können. Deshalb war das vorliegende Experiment so konzipiert, dass *Talk-Ahead* im alleinigen Fokus der Untersuchung lag und keine zusätzlichen Eingabekanäle ermöglicht wurden.

Neuss (2000) schlägt vor, durch das Zulassen von Synonymen die Nutzer auch bei *Talk-Ahead* zu unterstützen. Dies ist eine grundsätzlich sinnvolle Forderung, die auch den Bedienkomfort allgemein steigern kann. Allerdings sinkt die Erkennungsgüte eines Spracherkenners mit jedem weiteren Begriff im Vokabular und gerade bei Exekutivkommandos muss darauf geachtet werden, dass sie phonologisch gut differenzierbar sind. Darüber hinaus besteht das Risiko, dass die Schnittstelle für den Nutzer inkonsistent wirkt: Es ist fast nicht möglich, alle Synonyme zu berücksichtigen, die den Nutzern sinnvoll erscheinen. Deshalb können diese bei der Eingabe leicht die Erfahrung machen, dass sie an einigen Stellen mehrere Begriffe für ein und denselben Befehl verwenden können, während sie an anderen Stellen zwingend genau einen Begriff verwenden müssen. In dieser Studie wurde deshalb durch die Einbettung des *Talk-Ahead* Prinzips in den Kontext einer *Say What You See* Steuerung versucht, den Versuchspersonen die relevanten Begriffe eindeutig zu präsentieren. Darüber hinaus wurden mehrere Blöcke mit ähnlichen Aufgaben durchgeführt. So konnten die Teilnehmer in relativ kurzer Zeit eine gewisse Nutzungserfahrung entwickeln. Basierend auf oben ausgeführten Überlegungen zu der Bildung von Chunks sollte es so möglich sein, eine größere Chance für das Behalten des relevanten Vokabulars zu erzielen, da die Kommandos nicht separat, sondern zu Einheiten integriert gespeichert werden.

An dieser Stelle erscheint der erneute Hinweis sinnvoll, dass *Talk-Ahead* nicht als Ersatz für kontextfreie Direktzugriffe eingesetzt, sondern dem Nutzer in einer Schnittstelle zusätzlich angeboten werden sollte. *Talk-Ahead* soll die schnelle Nutzung der sprachlichen Schnittstelle gerade in den Fällen ermöglichen, in denen kein Direktzugriff möglich ist und deshalb zeitaufwändig schrittweise navigiert werden muss. Reeves et al. (2004) betonen in einem Artikel zu Gestaltungsrichtlinien, dass sich die Vorteile von Multimodalität nur dann voll ausschöpfen lassen, wenn die Nutzer volle Kontrolle bei der Wahl der Modalitäten haben und die für sie bestmögliche Variante in einer gegebenen Aufgabensituation wählen können. Damit sie jedoch vollwertige Alternativen darstellen, müssen die einzelnen Schnittstellenkomponenten soweit wie möglich optimiert werden. Das an diese Stelle besprochene Experiment ist in diesem Sinn als Beitrag zur Optimierung der sprachlichen Schnittstelle zu sehen.

5.1.2 Experiment 4: Multimodales Talk-Ahead

Um multimodales *Talk-Ahead* empirisch testen zu können, wurden in einem Simulationsexperiment folgende Forschungsfragen adressiert: Inwiefern machen Testteilnehmer mit zunehmender Erfahrung tatsächlich von der *Talk-Ahead* Option Gebrauch? Wie wirkt sich das auf das Blickverhalten sowie auf die Workload aus? Inwiefern kann die Aufgabenbearbeitungszeit positiv beeinflusst werden?

5.1.2.1 Methode

Überblick

Die Versuchspersonen in dem hier beschriebenen Experiment mussten mit Hilfe von Spracheingabe nach dem *Say What You See* Prinzip eine Reihe von Aufgaben durchführen, wie sie für Infotainment- und Navigationssysteme oder zur Steuerung von Komfortfunktionen im Fahrzeug typisch sind. Die Hälfte der Teilnehmer konnte dabei multimodales *Talk-Ahead* einsetzen, die andere Hälfte erhielt diese Möglichkeit nicht.

Der Autor entwickelte für diesen Versuch einen vereinfachten Prototyp eines entsprechenden Systems, der zu einer Versuchsplattform mit Analysefunktionen für Leistungsmaße und mit experimentellen Manipulationsoptionen ausgebaut wurde. Über eine multimodale Mensch-Maschine-Schnittstelle kann in einem hierarchischen menü-basierten System die simulierte Einstellung von Klima-, Entertainment- und Navigationsparametern sowohl per Sprache als auch manuell über ein spezielles Eingabeelement getätigt werden. Die visuelle Gestaltung und das Interaktionskonzept wurden in einem Reviewprozess von mehreren Designern und Ergonomen des Fachzentrums für User Interface Design der Siemens AG überprüft und in einem iterativen Prozess angepasst. Dadurch sollte sichergestellt werden, dass die gestellten Aufgaben in ihrer Ausprägung und Funktionsweise typischen Fahrzeugsystemen ähneln und keine Probleme durch missverständliche Systemgestaltung während des Tests auftreten. Obwohl die Qualität des verwendeten Spracherkenners sich in informellen Vortests als sehr gut herausstellte, wurde im Experiment die Methode der Wizard-of-Oz Simulation (siehe unten) gewählt, damit die Erfolgsrate von Sprachbefehlen an die Anforderungen des experimentellen Aufbaus angepasst und ein möglicher unterschiedlicher Einfluss der Erkennungsgüte zwischen den experimentellen Gruppen ausgeschlossen werden kann.

Versuchspersonen

An dem Experiment partizipierten dreißig Versuchspersonen (19 Frauen, 11 Männer) im Alter zwischen 19 und 32 Jahren (Mittelwert: 24 Jahre), die für ihre Teilnahme eine entgeltliche Aufwandsentschädigung erhielten. Teilnahmevoraussetzung war Deutsch als Muttersprache. Dadurch sollten verzerrende Effekte bei der Spracheingabestrategie

vermieden werden. Keiner der Teilnehmer hatte Gebrauchserfahrung mit sprachgesteuerten Bordsystemen. Eine diesbezüglich unerfahrene Stichprobe war für das Experiment sinnvoll, um Lernfortschritte unverzerrt beobachten zu können. Fünfzehn Teilnehmer wurden zufällig der Kontrollgruppe zugeordnet, die die herkömmliche Einzelwort-Eingabe vornehmen musste. Die andere Hälfte war Teil der Experimentalgruppe, die nach der oben beschriebenen Methode die Möglichkeit hatte, durch *Talk-Ahead* ihre Sprachkommandos zu verknüpfen. Die Versuchspersonen waren nicht über die Hypothesen des Experiments informiert.

Wizard-of-Oz Simulation

Bei einem Wizard-of-Oz Experiment wird den Versuchspersonen gegenüber behauptet, ihre Sprachbefehle würden von einem Computerprogramm aufgenommen und interpretiert, obwohl tatsächlich ein Versuchsleiter – meist in einem anderen Raum – das Verhalten des Spracherkennungssystems simuliert. Dahlbäck, Jönsson und Ahrenberg (1993) betonen in einem Methodenartikel, dass es entscheidend sei, Versuchspersonen das Gefühl zu geben, mit einem Computer statt mit einem Menschen zu interagieren. Das gilt insbesondere, wenn die sprachliche Struktur nutzerseitiger Eingabe Grundlage der Untersuchung ist. Bei der Kommunikation von Mensch zu Mensch ist ein grundsätzlich anderes Verhalten zu beobachten als bei sprachbasierter Mensch-Maschine-Interaktion: Intuitiv wählen Menschen eher kurze, kommandobasierte Sprache, wenn sie mit einem Computer kommunizieren (Baber & Stammers, 1989; Richards & Underwood, 1984).

Durch die zunehmende Verbreitung von Untersuchung zu Spracheingabetechnologien und damit einer steigenden Gefahr der Bekanntheit der Wizard-of-Oz Methode bei den Versuchsteilnehmern erscheint es sinnvoll, zusätzliche Maßnahmen zu ergreifen, um das Vertrauen der Versuchspersonen in die Interaktionssituation mit einem technischen System zu stärken. Nachdem der experimentelle Prototyp über einen funktionierenden Spracherkenner verfügt, wurde den Versuchspersonen vor Beginn des Experiments ein Blick „hinter die Kulissen“ gestattet, d.h. der Arbeitsplatz des Versuchsleiters gezeigt. Dazu lief der Prototyp nicht im Wizard-Modus, sondern mit aktiviertem Spracherkenner. Für die Versuchspersonen war damit das Programmfenster auf der grafischen Benutzeroberfläche sichtbar, das eine 3-elementige n-best Liste der erkannten Sprachbefehle nebst dem dazugehörigen Konfidenzniveau (einer Zahl zwischen 1 und 10000) angab. Die Versuchspersonen wurden eingeladen, mit Hilfe einiger Kommandos den MP3-Player des Systems zu bedienen. Sie konnten dann sehen, dass ihre Kommandos tatsächlich das Verhalten des MP3-Players beeinflussten und wie z.B. undeutlichere Aussprache die Zusammensetzung der n-best Liste veränderte. Den Teilnehmern der Experimentalgruppe wurde zusätzlich das *Talk-Ahead* Prinzip erklärt und unter ihrer aktiven Beteiligung vorgeführt. Sobald die Versuchsperson

ihren Platz eingenommen hatte, deaktivierte der Versuchsleiter die Spracherkennung und schaltete in den Wizard-Modus um.

Alle Funktionen des Systems waren als einzelne Schaltflächen auf einem 19 Zoll Touchscreen repräsentiert und die Bedienoberfläche für beidhändige Touchscreen-Eingabe optimiert, so dass der Versuchsleiter sehr schnell, lautlos und ohne zusätzliche manuelle Eingabelemente das Verhalten des Systems beeinflussen konnte. Vor den Versuchen wurde der Wizard in mehreren Sitzungen darauf trainiert, mit möglichst konstanter Reaktionszeit auf Sprachbefehle zu reagieren. Dies konnte dadurch sichergestellt werden, dass ihm die Abfolge der Aufgaben vorher bekannt gemacht wurde und so die Menge der möglichen Kommandos zu einem Zeitpunkt beträchtlich eingegrenzt wurde. Für das Experiment von Bedeutung war dabei, dass der Wizard nicht möglichst schnell, sondern möglichst gleichmäßig mit einer gewissen typischen Verzögerung reagierte. Neuss (2000) berichtet, dass Latenzen von 300 bis 1000 ms, meist jedoch etwas über 500 ms üblich sind. Der Wizard wurde im Vorfeld darauf trainiert, diese Zeit vor der Auslösung eines Kommandos möglichst konstant einzuhalten. Erleichtert wurde das Vorgehen dadurch, dass die Abfolge der Aufgaben und damit der Sprachbefehle bekannt war.

Okklusionsmethode

Um Sicherheitsrisiken bei der Bedienung von Informationsanzeigen im Fahrzeug zu vermeiden, ist eine der wichtigsten Anforderungen an die Systemgestaltung, dass visuelle Zuwendungen zu diesen Anzeigen möglichst kurz sind. Die Okklusionsmethode bietet die Möglichkeit, diese Zeit zu erfassen (Krems, Keinath, Baumann, Bengler & Gelau, 2000; Krems et al., 2004). Charakteristisch für diese Methode ist die systematische Kontrolle der Zeiteinheiten, die einer Person für die Erfassung einer Vorlage zur Verfügung gestellt werden. In verschiedenen Varianten wird also die Sichtbarkeit z.B. der Anzeigeeinheit für die Versuchsteilnehmer eingeschränkt. Zum einen besteht die Möglichkeit, die für eine fehlerfreie Informationsaufnahme benötigte Betrachtungszeit von den Versuchsteilnehmern selbst wählen zu lassen und diese Zeit als abhängiges Maß zu verwenden (*subject-paced* Verfahren). Zum anderen kann die Information systemgesteuert zeitlich begrenzt in mehreren Präsentations- und Okklusionszyklen dargeboten und dabei die Genauigkeit und die Güte der Informationsaufnahme erhoben werden (*system-paced* Verfahren). Dieser Ansatz ist mit den seit langem eingesetzten tachistoskopischen Techniken zur Untersuchung der Lesbarkeit von Informationsdarstellungen verwandt. Gelau, Keinath, Baumann, Bengler und Krems (1999) konnten in einem Experiment unter Variation der Aufgabenkomplexität zeigen, dass die beiden Verfahren analoge Ergebnisse liefern. Die Okklusionsmethode wird bei der Evaluation von Schnittstellenvarianten typischerweise eingesetzt, um entweder auf der Ebene einer Komplexitätsabschätzung zwischen unterschiedlichen

visuellen Gestaltungsvarianten zu trennen oder auf der Ebene der Unterbrechbarkeit von Aufgaben eine Schätzung über das Ausmaß der Abwendung während der Fahrt zu gewinnen.

Die wichtigsten Parameter in einem Versuch mit der Okklusionsmethode sind die Gesamtzeit, die für die Bearbeitung einer Aufgabe benötigt wird (total task time, TTT) und die Gesamtzeit, in der die jeweilige Vorlage während einer Aufgabe sichtbar ist (total shutter open time, TSOT). Darüber hinaus sind die einzelnen Betrachtungs- und Okklusionszeiten relevant, die je nach Durchführungsvariante konstante Werte annehmen, einer Verteilung folgen oder von den Versuchspersonen selbst gesteuert werden können. So können beispielsweise Betrachtungszyklen von 200 bis 1200 ms untersucht werden, die für einzelne Blickabwendungen von der Fahrbahn als akzeptabel angesehen werden (Zwahlen et al., 1988).

Krems et al. (2004) merken kritisch an, dass durch die Simulation der Abfolgen von Blickzuwendung und Blickabwendung bei der Okklusionsmethode die reale Fahrsituation nur partiell abgebildet wird. Die Okklusionsphase unterbricht wie bei der Blickabwendung vom Display zur Fahrbahn den Wahrnehmungsvorgang für die Verarbeitung der Anzeige. Die Zuwendung zu einer anderen Szene wie dem Verkehrsgeschehen wird jedoch nicht simuliert und sogar der Fixationspunkt kann während der Okklusion häufig beibehalten werden. Durch die Installation eines Doppelaufgabenparadigmas kann die externe Validität erhöht werden, indem eine zweite Aufgabe mit visueller Komponente die Blickzuwendung erfordert.

In dem vorliegenden Experiment kam die *subject-paced* Variante zum Einsatz. Anders als bei typischen Anwendungen der Okklusionsmethode war allerdings nicht die Trennung unterschiedlicher visueller Designausprägungen von Interesse, sondern das Informationsaufnahmeverhalten unter SWYS-basierter Sprachsteuerung. Der Fokus dieser Untersuchung bestand darin, inwiefern multimodales *Talk-Ahead* die Interaktion mit dem visuellen Display beeinflusst, genauer: Inwiefern sich die Zuwendung zum Display zur Aufnahme von Information über die Verfügbarkeit von Sprachbefehlen und deren Ausführungsstatus nach Kommandoabgabe in Abhängigkeit von der Nutzungserfahrung verändert. Da zwei Gestaltungsalternativen direkt miteinander verglichen wurden, konnte auch das sonst übliche Vorgehen, die Bedienung einmal mit und einmal ohne Okklusion durchzuführen, verzichtet werden. Nicht der Quotient aus TTT und TSOT wird in diesem Experiment gegenübergestellt, sondern direkt die TSOT bzw. die einzelnen Blickdauern.

Versuchsprototyp und -material

Die Experimentalplattform des Systems zur simulierten Steuerung von Navigations-, Unterhaltungs- und Komfortfunktionen wurde vom Autor in Microsoft® Visual Basic® realisiert. Einen Überblick über die verfügbaren Funktionen gibt Anhang B 1.

Zusätzlich konnte der Versuchsleiter über das Wizard-of-Oz Modul jeden zur Versuchsdurchführung relevanten Systemzustand per Fernzugriff herstellen. Das Verhalten des Prototyps war dabei identisch zu einer Nutzereingabe per Sprachbefehl. Sprachliche Kommandos wurden über ein angebundenes C-Programm mit dem Spracherkennung VoCon® 3200 der Firma ScanSoft, Inc. (jetzt: Nuance Communications, Inc.) ermöglicht. Dieser Erkennung muss vor der Verwendung nicht trainiert werden. Da das aktive Vokabular für die Vorführung klein war, wurde multimodales *Talk-Ahead* zur Demonstration für die Experimentalgruppe durch eine optionale Verknüpfung der Kommandos in der Grammatik realisiert (Online-Erkennung).

Das Programm lief auf einem Fujitsu Siemens PC mit 2.8 GHz CPU, ausgestattet mit einer Matrox® G450 Grafikkarte und einer ESS Technologies ESS1969 Soundkarte. Die Anzeigeeinheit bestand aus einem 7 Zoll TFT-LCD Breitbild-Farbmonitor mit 800 × 480 Bildpunkten der Firma CarTFT. Zur Spracheingabe diente das rauscharme Mikrofon eines Sennheiser PC130 Headsets. Das Display wurde auf einem Tisch zur rechten Seite der Versuchspersonen angebracht. Dabei wurde in etwa der Abstand und Winkel eingehalten, wie er bei einer Montage in der Mittelkonsole auftreten würde. Direkt vor der Versuchsperson befand sich ein 17 Zoll CRT Farbmonitor, der für die Bearbeitung einer visuell-manuellen Nebenaufgabe benötigt wurde. Diese lief separat, aber zeitlich synchronisiert auf einem Dell Laptop mit 1.2 GHz CPU (Chipsatz Intel® 855GM). Ebenfalls der Bearbeitung dieser Nebenaufgabe diente ein Tastenelement, dessen Taster von der Versuchsperson mit der linken Hand bedient wurden. Zur Steuerung der Sichtbarkeit der grafischen Benutzeroberfläche im Rahmen der Okklusionsmethode verfügte die Versuchsperson über einen einzelnen Taster, den sie mit der rechten Hand nutzen konnte. Eine bildliche Darstellung dieses Aufbaus ist Abbildung 29 A) zu entnehmen. Alle verfügbaren Sprachbefehle wurden auf der Anzeigeeinheit in weißer Schrift dargestellt, bzw. orange hinterlegt, wenn das entsprechende Element ausgewählt wurde. Diese Markierung fand sowohl bei der sprachlichen als auch bei der hier nicht getesteten manuellen Interaktion statt, um die Auswirkungen der verschiedenen Eingabemodalitäten auf die visuelle Ausgabe parallel zu halten. Abbildung 29 B) zeigt exemplarisch zwei der für die Aufgabendurchführung relevanten Menüebenen.

Die erfolgreiche Auslösung eines Sprachbefehls wurde nicht durch Sprachausgabe quittiert, sondern nur durch ein einfaches auditives Ereignis (Klickgeräusch) und auf der Anzeigeeinheit durch die Darstellung des neuen Inhalts signalisiert. Multimodale Bordsysteme verfügen zwar typischerweise über komplexe auditive Ausgaben z.B. in Form von Earcons oder über Sprachausgaben zur Rückmeldung. Im vorliegenden Experiment wurde jedoch aufgrund versuchsplanerischer Veranlassung die Möglichkeit zur vollständig non-visuellen Interaktion eingeschränkt: Das Ziel der Studie war nicht, eine absolute Abschätzung dafür zu erhalten, ob und wie lange der Nutzer bei

Systemen mit Sprachsteuerungen den Blick von der Fahrbahn abwendet. Ebenso wenig war es von Interesse, einen Usability-Test eines produktreifen Systems durchzuführen. Es sollte stattdessen ermittelt werden, wie sich die zwei Varianten der SWYS Steuerung mit und ohne *Talk-Ahead* Option bezüglich des Blickverhaltens unterscheiden. Die in dieser Studie gewonnenen Blickdaten müssen vor dem Hintergrund des eingeschränkten auditiven Feedbacks interpretiert und für eine Extrapolation auf den realen Anwendungsfall als Maximaldauer gesehen werden. Allerdings muss angemerkt werden, dass eine auditive Ausgabe nicht immer ihren Zweck erfüllen kann: Es lässt sich häufig beobachten, dass bei der Interaktion mit multimodalen Systemen im Fahrzeug, insbesondere wenn Mitfahrende anwesend sind, die auditive Ausgabe sehr leise oder falls möglich sogar völlig ausgeschaltet wird, weil sie von vielen als störend empfunden wird.

(A) Versuchsaufbau



(B) Anzeigeeinheit

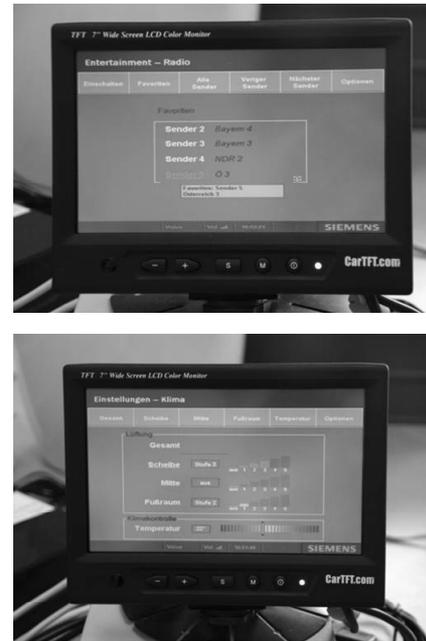


Abbildung 29: A) Versuchsteilnehmer in der Experimentalsituation (in einem Vortest); B) oben: Senderauswahl des Radios, B) unten: Klimafunktionen.

Versuchsdesign

Die Studie war als 2×3 gemischt faktorielles Design konzipiert. Die Art der Spracheingabe wurde als unabhängige Between-Subjects Variable zwischen den Probanden variiert und bestand aus zwei Stufen: Einzelworteingabe ohne Verknüpfungsoption (Kontrollgruppe) und *Talk-Ahead* mit Verknüpfungsoption der Sprachkommandos (Experimentalgruppe). Es erschien ratsam, die Art der Spracheingabe als Between-Faktor anzulegen. Bei einer Within-Subjects Variation wären Carry-over-Effekte insbesondere dann sehr wahrscheinlich, wenn eine Versuchsperson nach der Eingabevariante mit *Talk-Ahead* Option auf eine funktional eingeschränktere Schnittstelle

umsteigen muss. Die Durchgangsnummer des experimentellen Blocks stellte mit drei Stufen die unabhängige Within-Subjects Variable dar: erster, zweiter und dritter Block. Basierend auf der Annahme, dass die Nutzung von *Talk-Ahead* einige Erfahrung mit dem System, der Menüstruktur und dem Vokabular zur Steuerung erfordert, bestand das Experiment aus mehreren experimentellen Blöcken. Dadurch sollten zum einen der allgemeine Lernfortschritt bei der Steuerung des Systems und zum anderen mögliche Unterschiede zwischen den Eingabevarianten über die Zeit sichtbar gemacht werden.

Als abhängige Variablen wurden die Aufgabenbearbeitungszeit, visuelle Abwendung (mittlere Betrachtungszeiten und TSOT) und subjektive Workload erhoben. Zur Workloadmessung kam der RTLX (Byers et al., 1989) eine Abwandlung des NASA Task Load Index (TLX, Hart & Staveland, 1988) zum Einsatz, der eine Abschätzung der gesamten mit der Aufgabenbearbeitung assoziierten Workload gestattet. Zusätzlich wurde in der Experimentalgruppe der Anteil eingesparter einzelner Sprachkommandos durch die Verwendung von Kommandoverknüpfung bei *Talk-Ahead* analysiert.

Prozedur und Aufgaben

Nach der Begrüßung durch den Versuchsleiter füllten die Teilnehmer zunächst einen demografischen Fragebogen aus. Sie wurden darüber informiert, dass dieser Test ausschließlich der Evaluation einer Benutzerschnittstelle dient und keine Untersuchung über die Fähigkeiten der beteiligten Personen zur Nutzung technischer Produkte darstellt. Auf eine kurze Schilderung des Kontextes der Untersuchung – Spracheingabe zum Einsatz im Automobil – folgte die oben erwähnte Demonstration des Spracherkenners, während der die Teilnehmer der Experimentalgruppe auf die Möglichkeit des *Talk-Ahead* aufmerksam gemacht wurden. Ferner wurde ihnen mitgeteilt, dass sie diese Option nutzen können, falls sie ihnen sinnvoll und hilfreich erscheint, aber nicht nutzen müssen. Lediglich bei der Vorführung des Spracherkenners wurde ihnen nahe gelegt, *Talk-Ahead* auszuprobieren. Zusätzlich wurde allen Teilnehmern das *Say What You See* Prinzip erklärt. Exemplarisch anhand einer Beispielaufgabe, die im späteren Test nicht vorkam, wurde den Versuchspersonen der Ablauf einer Aufgabe im Experiment vorgeführt. Sie bekamen die Anweisung, die Aufgaben möglichst zügig, aber auch ohne zeitlichen Druck und möglichst fehlerfrei zu erledigen.

Alle Versuchsteilnehmer mussten drei Blöcke zu je sechs Aufgaben absolvieren. Die Aufgaben bestanden darin, ausschließlich per Spracheingabe nach dem SWYS Prinzip typische Vorgänge bei der Bedienung von Komfortsystemen durchzuführen. In jedem Block galt es, je zwei Aufgaben aus dem Kontext Navigation, Entertainment und Klimaeinstellungen auszuführen. Eine Auflistung der Aufgaben findet sich in Anhang B 2. Die Aufgaben unterschieden sich in Details zwischen den Blöcken, waren aber so konzipiert, dass die Aufgabenschwierigkeit nahezu äquivalent war. Es

handelte sich um dieselben Typen von Operationen, die mit unterschiedlichen Argumenten ausgeführt werden mussten (z.B. Titelnnummer bei CD ändern in Block 1 und anderen Radiosender einstellen in Block 2). Die Reihenfolge der Aufgaben wurde für jede Versuchsperson durch einen Algorithmus randomisiert, der sicherstellte, dass nicht zwei Aufgaben aus demselben Kontext (d.h. Navigation, Klima oder Entertainment) nacheinander vorkamen. Dadurch konnte erreicht werden, dass die Versuchspersonen bei der Bearbeitung immer einen Weg über mehrere Menüebenen nehmen mussten und immer mit einem Kontextwechsel starteten. Um zu vermeiden, dass die erhobenen Maße durch interindividuell verschiedenartige Verwendung von Abkürzungen durch Direktzugriffe beeinflusst wurden, waren diese Abkürzungen nicht zugelassen und der erste Sprachbefehl musste explizit einen Kontextwechsel herbeiführen. Alle Aufgaben erforderten zwischen vier und sechs Kommandos. Die Anzahl notwendiger Kommandos, insgesamt 27, war zwischen den experimentellen Blöcken gleich. Jede Aufgabe wurde eingeleitet durch einen Signalton, der unmittelbar auf die Anweisung des Versuchsleiters folgte, und wurde beendet, wenn der Zielzustand im System hergestellt war. Jedes Sprachkommando wurde wie bereits dargelegt vom Wizard ausgeführt und durch einen einfachen Signalton bestätigt.

Erst nachdem von den Probanden eindeutig Verständnis über den Ablauf der Aufgaben und die Art der Eingabe signalisiert wurde, erfolgte die Einweisung in das Vorgehen bei der Okklusionsmethode. Die Anzeigeeinheit blieb während des Experiments standardmäßig okkludiert, d.h. eine schwarze Maske lag über der grafischen Benutzeroberfläche, außer wenn die Versuchspersonen eine Taste zur Freigabe drückten. Während die Taste gedrückt war, blieben die Bildschirminhalte erkennbar und die Betrachtungszeit konnte gemessen werden. Es wurde betont, dass die Teilnehmer die Dauer der Freigabe so wählen sollten, dass sie die gewünschte Information tatsächlich einholen können. Gleichzeitig sei es jedoch entscheidend, dass sie die Taste unmittelbar nach dem Einholen dieser Information wieder losließen.

Parallel zu der Systembedienung musste eine visuell-manuelle Zweitaufgabe bearbeitet werden. Dieses Vorgehen empfiehlt sich bei der Okklusionsmethode, um die externe Validität zu steigern. Des Weiteren existiert der Vorteil in einem Doppelaufgabenparadigma, Rückschlüsse auf die Workload durch die Leistung bei der Zweitaufgabe ziehen zu können. Diese Variante ist besonders sensitiv, auch sehr kurze Intervalle erhöhter Workload zu detektieren (Verwey & Veltman, 1996). Die Zweitaufgabe bestand in einer einfachen Wahlreaktionsaufgabe. Auf einem dafür zusätzlich bereitgestellten Monitor wurde mittig ein Fixationskreuz angezeigt. Zu einem zufälligen Zeitpunkt in jedem Fünf-Sekunden-Intervall ersetzte ein Pfeilsymbol dieses Kreuz. Die Probanden mussten beim Erscheinen eines nach Links weisenden Pfeils die linke Taste auf einem Tastenbrett drücken, bei einem Pfeil nach rechts die rechte Taste. Zwischen den Intervallen bestand eine Pause von einer Sekunde, damit

zwei Stimuli nicht unmittelbar aufeinander folgen konnten. Bezüglich der Aufgabenhierarchie wurde den Teilnehmern mitgeteilt, dass die Systembedienung zu priorisieren sei. Die Versuchspersonen waren angewiesen, die Freigabe-Taste zum Sichtbarmachen des Bildschirms bei der Okklusionsmethode loszulassen, wenn sie auf das Display für die Zweitaufgabe blickten. Dieses Vorgehen wurde im Vorfeld geübt und während des Versuchs vom Versuchsleiter besonders kontrolliert. Im Gegensatz zum üblichen Vorgehen bei der Okklusionsmethode (Krems et al., 2004) wurde jedoch die Zweitaufgabe während der Shutter-Open Phase nicht angehalten, um eine realistischere Belastungssituation zu schaffen.

Vor der Durchführung der drei experimentellen Aufgabenblöcke absolvierten die Versuchspersonen zwei Blöcke zu ebenfalls je sechs Aufgaben als Probedurchgänge. Während des ersten Blocks wurde die Okklusionsmethode nicht eingesetzt, um den Probanden die Möglichkeit zu geben, das System kennen zu lernen. Auch die Zweitaufgabe musste nicht bearbeitet werden. Der zweite Block beinhaltete sowohl die Okklusionsmethode als auch die Zweitaufgabe, um die Teilnehmer an die Betätigung der Freigabetaste zu gewöhnen. In beiden Probedurchgängen simulierte der Wizard jeweils einen leicht zu korrigierenden Ersetzungsfehler des Sprachsystems. Dadurch sollte der Eindruck verstärkt werden, dass mit einem realen System gearbeitet wurde. In den darauf folgenden experimentellen Durchgängen lag die simulierte Erkennungsrate bei hundert Prozent.

Zwischen allen experimentellen Blöcken wurde eine kurze Pause von ungefähr drei Minuten Dauer eingefügt. Am Ende des Experiments füllten die Teilnehmer einen Fragebogen zu subjektiver Workload (NASA Task Load Index) aus. Sie wurden darauf hingewiesen, dass nur die Spracheingabe in das System zu bewerten sei und nicht die Zweitaufgabe. Anschließend erhielten sie Gelegenheit, ihre Gedanken zu der Ausführung des Systems und der Steuerungsoption per Sprache zu äußern. Das Experiment dauerte etwas über eine Stunde.

5.1.2.2 Hypothesen

Aufbauend auf den Befunden von Neuss (2000) wird erwartet, dass die Verknüpfung von Sprachkommandos (*Talk-Ahead*) zu einer substantiellen Reduktion der Aufgabebearbeitungszeit führt. Wenn weniger Einzelkommandos geäußert werden, sollte allein schon durch die Einsparung der jeweiligen Erkennungslatenzen eine zeitliche Reduktion zu erreichen sein. Darüber hinaus ist allerdings zu erwarten, dass sich das Nutzerverhalten grundsätzlich ändert: Die typische Trägheit, die sich bei der Mensch-Maschine-Interaktion einstellt, wenn Nutzer erst darauf warten müssen, dass die eben eingeleitete Aktion abgeschlossen wird, bevor eine neue initiiert werden kann, sollte durch *Talk-Ahead* minimiert werden. Wenn diese Option tatsächlich dem Bedürfnis nach einer effektiveren Eingabeform bei dem Übergang vom unerfahrenen zum erfah-

renen Nutzer entspricht, sollte das einen übergreifenden Einfluss auf das Verhalten haben. Die Nutzer sollten im Rahmen eines Strategiewechsels ihre Interaktionsgeschwindigkeit allgemein steigern, weil sie die Repräsentation eines Systems aufbauen, das sich ihrer persönlichen Geschwindigkeit besser anpasst und bei dem sie nicht warten müssen, bis es bereit ist. Trifft diese Annahme zu, sollten größere Einspareffekte als nur der Wegfall der einzelnen Latenzen des Erkenners (bzw. hier des Wizards) bei der Aufgabenbearbeitungszeit zu erzielen sein. Nachdem es sich bei *Talk-Ahead* um eine Expertenoption handelt, sollten sich die Effekte allerdings erst in den späteren experimentellen Blöcken einstellen.

Bezüglich der Blickabwendungen, wie sie durch die Gesamtbetrachtungszeit (TSOT) und die einzelnen mittleren Betrachtungsdauern bei der Okklusionsmethode gemessen werden, sollten sich ebenfalls differentielle Effekte durch die Art der Eingabe ergeben. Zwar steht zu erwarten, dass die reine Informationsaufnahme sowie das Ablesen und Lernen der Sprachbefehle durch SWYS ähnlich viel Betrachtungszeit in Anspruch nehmen. Allerdings lässt sich bei der Sprachinteraktion mit visueller Ausgabe immer wieder beobachten, dass die Nutzer Kontrollblicke über den Ausführungsstatus ihrer Eingabe tätigen. Durch das Nutzen der *Talk-Ahead* Option fallen einige Menüebenen und Einzelschritte weg, wodurch sich schlichtweg weniger Gelegenheiten ergeben, auf das Display zu sehen. Das führt zu der Hypothese, dass die Teilnehmer nur kurz am Ende einer Befehlskette den Status quo überprüfen, was die TSOT gering halten sollte. Sollte darüber hinaus tatsächlich die Interaktionsgeschwindigkeit durch *Talk-Ahead* wie dargelegt durch bessere Nutzerunterstützung zu einer allgemeinen Erhöhung der Dynamik führen, sollte sich das auch an den einzelnen Blickdauern ablesen lassen. Die Dauer einzelner Kontrollblicke müsste dann kürzer werden.

Für die Steuerung für Komfortfunktionen im Fahrzeug sollte jede Eingabeform unter der Prämisse gestaltet sein, möglichst wenige Ressourcen des Fahrers von der Hauptaufgabe abzuziehen. Norman (1988, S.125) fordert für alltägliche Bedienungsaufgaben, dass sie mit wenig bewusster mentaler Aktivität zu bewältigen sein müssen. Das bedeutet konkret: „[...] they must minimize planning (and especially any planning with extensive looking ahead and backing up) and mental computation“. Damit die *Talk-Ahead* Option tatsächlich als vorteilhaft angesehen werden kann, darf die Verknüpfung von Sprachkommandos keinen zusätzlichen kognitiven Planungsaufwand erfordern. Da bei *Talk-Ahead* die Nutzer eher die Chance haben, die Interaktionsgeschwindigkeit auf ihre eigenen Bedürfnisse hin anzupassen und weniger auf den Bereitschaftszustand des Systems achten müssen, wird die Hypothese aufgestellt, dass *Talk-Ahead* mit verminderter Workload einhergeht. Zur Kontrolle dieser Variable dient auf subjektiver Ebene der RTLX, auf objektiver Ebene die Leistung in der Wahlreaktionsaufgabe.

5.1.2.3 Ergebnisse

Aufgabenbearbeitungszeit, Gesamtbetrachtungszeit (TSOT), mittlere Blickdauer sowie die Leistungen in der Wahlreaktionsaufgabe wurden für die Experimentalgruppe und die Kontrollgruppe über alle Aufgaben in jedem Block gemittelt. Diese Ergebnisse und der Anteil verknüpfter Sprachkommandos durch *Talk-Ahead* in der Experimentalgruppe wurden mittels Varianzanalysen auf statistische Unterschiede geprüft. Für die inferenzstatistische Auswertung der subjektiven Workload und der mittleren Fehlerrate bei der Durchführung der Aufgaben über alle experimentellen Blöcke kamen t-Tests für unabhängige Stichproben zum Einsatz. Bei signifikanten Unterschieden wird als Maß der Effektgröße das partielle Eta-Quadrat η_p^2 bei Varianzanalysen berichtet, bei t-Tests Cohens d . Jede Analyse beinhaltet zusätzlich eine Schätzung der Wahrscheinlichkeit einer Replikation des Effekts p_{rep} (Killeen, 2005a, 2005b). Da in dem Experiment Vergleiche zwischen Gruppen vorkommen, wurde in diesen Fällen zusätzlich der Levene-Test zur Überprüfung auf Varianzhomogenität gerechnet. Die Ergebnisse des Tests werden nur im Fall heterogener Varianzen berichtet. Wenn nicht anders angegeben, werden im Text Mittelwerte (M) und Standardfehler (SE) berichtet. Ebenso weisen bei grafischen Darstellungen die Fehlerbalken den Standardfehler aus.

Nutzung der Talk-Ahead Option

Von Interesse ist zunächst, in welchem Ausmaß die *Talk-Ahead* Option von den Teilnehmern der Experimentalgruppe genutzt wurde. Da bei Eingabefehlern, z.B. durch Versprechen oder falsches Vokabular, keine eindeutige Einordnung der Befehlskette mehr hergestellt werden konnte, wurden nur fehlerfreie Durchgänge berücksichtigt. Für die Analyse wurde die Differenz aus der ohne *Talk-Ahead* notwendigen Anzahl von Einzelkommandos und aus der tatsächlich abgegebenen Anzahl von Kommandos für jeden Block errechnet. Wenn eine Versuchsperson beispielsweise Titel fünf der CD mit dem *Talk-Ahead* Befehl „CD, Titel fünf“ statt durch „CD“ [Pause] „Titel 5“ gestartet hat, wurden zwei Einzelkommandos zu einem verknüpften Kommando zusammengeschlossen und ein Einzelkommando eingespart. Erfolgt die Eingabe des Navigationsziels durch „Navigation“ [Pause] „Adresse, München, Otto-Hahn-Ring“, wurden gegenüber der Version ohne *Talk-Ahead* zwei Befehle gespart („Navigation“ [Pause] „Adresse“ [Pause] „München“ [Pause] „Otto-Hahn-Ring“). Auf diese Weise erhält man eine Abschätzung, inwiefern *Talk-Ahead* effektiv genutzt wurde.

Die Ergebnisse sind in Abbildung 30 dargestellt. Insgesamt zeigte sich, dass bereits im ersten experimentellen Block *Talk-Ahead* eingesetzt wurde. Im Schnitt wurden im ersten Block 4.6 ($SE = 0.96$) Kommandos eingespart. Die Nutzung stieg jedoch über den Verlauf des Experiments weiter an: In Block 2 wurden im Mittel 5.5 ($SE = 1.09$) im 3. Block 7.7 ($SE = 1.15$) Kommandos eingespart.

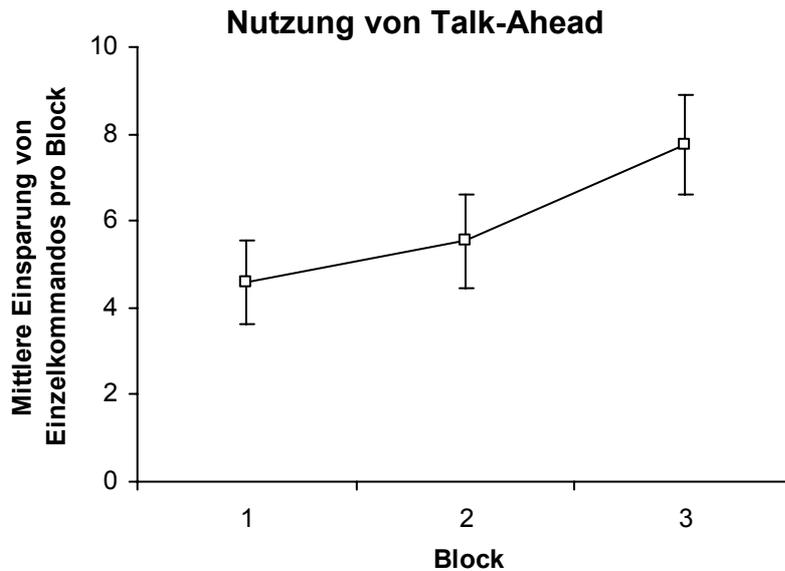


Abbildung 30: Mittlere Anzahl durch *Talk-Ahead* eingesparter Einzelkommandos summiert über alle fehlerfreien Durchgängen eines Blocks.

Es lässt sich ein signifikanter Unterschied zwischen der Anzahl gesparter Kommandos in den einzelnen Blöcken nachweisen, $F(2, 28) = 6.21$, $p < .01$, $p_{\text{rep}} = .96$, $\eta_p^2 = .31$.

Leistung bei der Aufgabenbearbeitung

Multimodales *Talk-Ahead* führte zu einer deutlich geringeren Aufgabenbearbeitungszeit als eine Spracheingabestrategie, die schrittweise ein Kommando nach dem anderen abarbeitet: Die Experimentalgruppe mit *Talk-Ahead* konnte die Aufgaben schneller bearbeiten als die Kontrollgruppe ohne diese zusätzliche Option. Abbildung 31 zeigt diesen Zusammenhang für den Verlauf des Experiments.

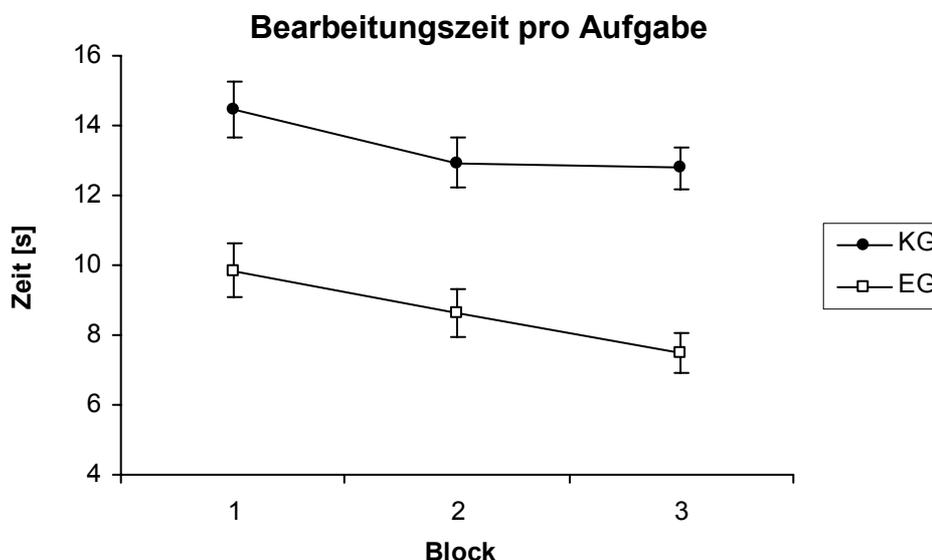


Abbildung 31: Mittlere Aufgabenbearbeitungszeiten der Experimentalgruppe (EG) mit *Talk-Ahead* Option und der Kontrollgruppe (KG).

Die Daten aus Durchgängen mit Fehlern wurden nicht aus der Analyse ausgenommen, da eine Betrachtung der tatsächlich möglichen Bearbeitungszeiten im Vergleich angestrebt wurde und die Fehlerraten sehr ähnlich waren (siehe unten). Die Teilnehmer der Kontrollgruppe benötigten im ersten experimentellen Block 14.45 ($SE = 0.79$), im zweiten 12.93 ($SE = 0.70$) und im dritten 12.78 ($SE = 0.58$) Sekunden. Die korrespondierenden Zeiten der Experimentalgruppe sind 9.86 ($SE = 5.18$), 8.62 ($SE = 0.45$) und 7.49 ($SE = 0.42$) Sekunden. Die mittlere Bearbeitungszeit in der Kontrollgruppe betrug 13.39 Sekunden ($SE = 0.41$), die Experimentalgruppe war mit 8.65 Sekunden ($SE = 0.30$) ungefähr 1/3 schneller. Eine zweifaktorielle Varianzanalyse mit Messwiederholung bestätigt den Haupteffekt für Gruppe, $F(1,28) = 45.96$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .62$. Wie ebenfalls aus Abbildung 31 ersichtlich ist, gibt es einen deutlichen Übungseffekt in beiden Gruppen. Die durchschnittliche Aufgabenbearbeitungszeit fiel von 12.16 Sekunden ($SE = 0.63$) auf 10.77 Sekunden ($SE = 0.57$) in Block 2 und 10.14 Sekunden ($SE = 0.61$) in Block 3. Dieser Haupteffekt ist ebenfalls signifikant: $F(2,56) = 13.29$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .32$. Eine Interaktion Gruppe \times Block hingegen konnte nicht festgestellt werden, $F(2,56) = 0.80$, $p = .46$, $p_{\text{rep}} = .53$.

Trotz der Übungsdurchgänge machten die Teilnehmer in beiden Gruppen vereinzelt Fehler bei der Aufgabenbearbeitung. Es kam nicht vor, dass eine Aufgabe nicht gelöst werden konnte. Vielmehr versprachen sich die Teilnehmer bei der Formulierung der Sprachbefehle oder benutzten nicht genau die Kommandovokabel, die durch SWYS vorgegeben wurde. Synonyme waren im Test nicht zugelassen. Es handelt sich somit um vergleichsweise unkritische Fehler. Betrachtet man die Durchgänge, in denen derartige Fehler aufgetreten sind, findet sich zwischen den Gruppen kein deutlicher Unterschied. In der Experimentalgruppe kam dies in 7.7% ($SE = 1.94$) der Fälle vor, in der Kontrollgruppe wiesen 9.3% ($SE = 1.85$) der Durchgänge einen Fehler auf. Entsprechend kann kein statistisch verlässlicher Unterschied nachgewiesen werden, $t(28) = .55$, $p = .59$, $p_{\text{rep}} = .44$.

Blickdaten

Die Zeit, die Nutzer eines Bordsystems im Fahrzeug ihren Blick von der Straße abwenden, ist ein wesentlicher Faktor für die Fahrgüte und damit für die Eignung dieses Systems. Mit Hilfe des *subject-paced* Verfahrens der Okklusionsmethode wurden für beide Gruppen die Dauer einzelner Blicke auf das Display und die TSOT, die Gesamtzeit der Blickzuwendung, ermittelt.

Die Gesamtblickdauer auf das Display während der Sprachbedienung unterschied sich deutlich zwischen den beiden Gruppen, über die Blöcke hinweg ist ein Lerneffekt ersichtlich (siehe Abbildung 32). In der Kontrollgruppe wurden von Block 1 nach 3 Zeiten von 4.35 ($SE = 0.67$), 3.46 ($SE = 0.70$) und 3.23 ($SE = 0.57$) Sekunden gemes-

sen, in der Experimentalgruppe 2.50 ($SE = 0.47$), 2.23 ($SE = 0.38$) und 1.20 ($SE = 0.20$) Sekunden.

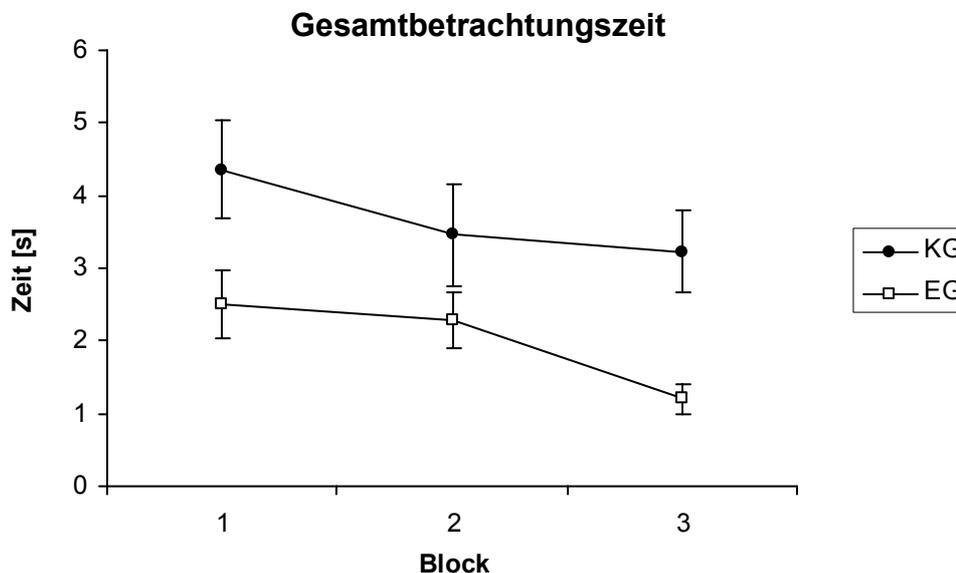


Abbildung 32: Durchschnittliche Gesamtbetrachtungszeit der Anzeigeeinheit pro Aufgabe für Kontroll- (KG) und Experimentalgruppe (EG).

Die Haupteffekte erwiesen sich als signifikant, eine Interaktion Gruppe \times Block war nicht nachweisbar ($F(2,56) = 1.21$, $p = .31$, $p_{\text{rep}} = .64$). Von Block 1 zu Block 3 sank die TSOT in beiden Gruppen stetig ab: 3.4, 2.9 und 2.2 Sekunden, $F(2,56) = 8.68$, $p = .001$, $p_{\text{rep}} = .99$, $\eta_p^2 = .24$. In der Kontrollgruppe ergab sich eine mittlere Gesamtbetrachtungsdauer von 3.68 Sekunden ($SE = 0.37$), unter Verfügbarkeit der *Talk-Ahead* Option reduzierte sich die TSOT auf 2.00 Sekunden ($SE = 0.22$), $F(1,28) = 6.43$, $p = .02$, $p_{\text{rep}} = .93$, $\eta_p^2 = .19$. Der Levene-Test auf Gleichheit der Fehlervarianzen zeigt bei den Blöcken 2 und 3 unterschiedliche Varianzen an (Block 2: $F(1,28) = 8.44$, $p < .01$; Block 3: $F(1,28) = 21.87$, $p < .001$). Wengleich die Varianzanalyse ein sehr robustes Verfahren ist und ungleiche Varianzen den F -Test nicht erheblich beeinflussen, wenn die Stichproben gleichen Umfang haben (Glass, Peckham & Sanders, 1972), besteht die Gefahr eines positiven Bias (Wilcox, 1987). Um auszugleichen, dass die Varianzanalyse bei Varianzheterogenität zu liberal ist, wird häufig die Verwendung des Welch-Test empfohlen (Keppel, 1991, Kapitel 6). Das Ergebnis wurde entsprechend mit Hilfe des Welch-Tests überprüft und konnte auch hier bestätigt werden, $df_{\text{corr}} = 21.40$, $p = .019$.

Betrachtet man die Dauer einzelner Blicke auf das Display, ergibt sich ein etwas anderes Bild. Abbildung 33 gibt eine grafische Darstellung des Ergebnismusters. Während in den ersten beiden Blöcken die Einzelblickdauern in beiden Gruppen sehr ähnlich waren (Kontrollgruppe: $M = 1.54$ s, $SE = 0.36$; $M = 1.31$ s, $SE = 0.28$; Experimentalgruppe: $M = 1.53$ s, $SE = 0.45$; $M = 1.58$ s, $SE = 0.35$), konnte in Block 3

eine substantielle Reduktion der Dauer eines Kontrollblicks bei den Teilnehmern der Experimentalgruppe festgestellt werden ($M = 0.79$ s, $SE = 0.18$), während die Versuchspersonen aus der Kontrollgruppe diesen Effekt nicht zeigten ($M = 1.22$ s, $SE = 0.22$).

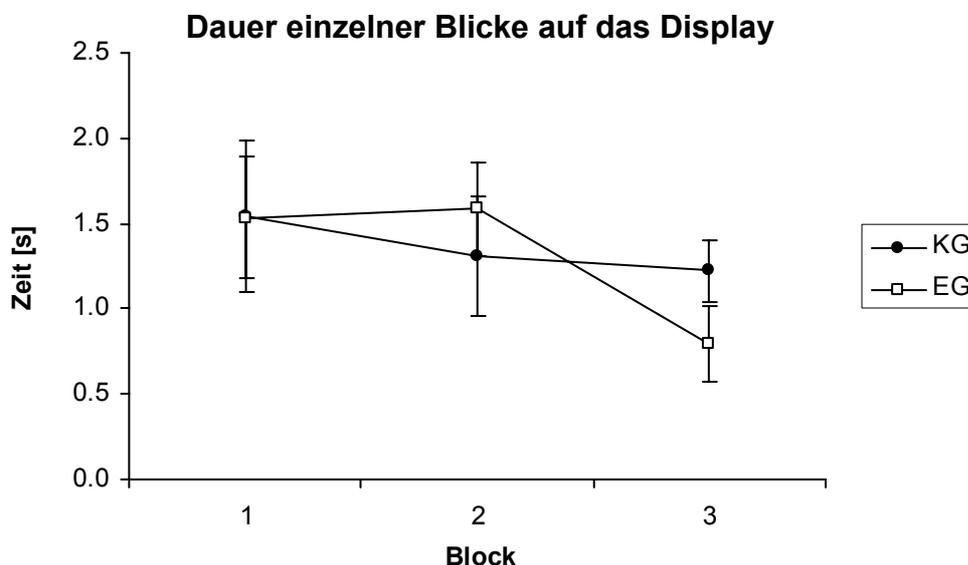


Abbildung 33: Durchschnittliche Dauer einzelner Blicke auf das Display für Kontrollgruppe (KG) und Experimentalgruppe (EG).

Die statistische Analyse zeigt eine Tendenz zu einer signifikanten Gruppe \times Block Interaktion: $F(2,56) = 2.71$, $p = .075$, $p_{\text{rep}} = .85$, $\eta_p^2 = .09$. Der signifikante Haupteffekt für Block wird aufgrund dieser tendenziellen Wechselwirkung (disordinale Interaktion) nicht weiter interpretiert ($F(2,56) = 7.01$, $p < .01$, $p_{\text{rep}} = .98$). Ein allgemeiner Unterschied zwischen den Gruppen kann nicht statistisch belegt werden ($F(1,28) = 0.02$, $p = .90$, $p_{\text{rep}} = .19$).

Workload

Die bei der Aufgabenbearbeitung erlebte Workload wurde über zwei Methoden ermittelt: Zum einen berichteten die Versuchsteilnehmer mit Hilfe der RTLX Variante des NASA Task Load Index die subjektive Beanspruchung, zum anderen sollte die Leistung bei der Sekundäraufgabe im Subsidiary Task Paradigma Auskunft geben. Allgemeine Anmerkungen zur Auswertung des TLX sind in Abschnitt 4.1.2.3 (Ergebnisse – Subjektive Workload) zu finden.

Der RTLX zeigt keine signifikanten Unterschiede, die Differenz betrug lediglich einen Skalenpunkt ($t(28) = 0.34$, $p > .7$, $p_{\text{rep}} = .33$). In der Kontrollgruppe ergab sich ein mittlerer Wert von 36.3 ($SE = 3.6$), für die Experimentalgruppe wurde ein Score von 34.6 ($SE = 3.5$) errechnet. Die Mittelwerte der Subskalen sind in Tabelle 9 zusammenfassend dargestellt.

Tabelle 9: Übersicht über die Bewertungen auf den RTLX Subskalen (Wertebereich 0-20) für Kontroll- und Experimentalgruppe (KG, EG).

Dimension	KG		EG		t(28)	p
	M	(SE)	M	(SE)		
Geistige Anforderungen	11.5	(1.2)	11.5	(1.1)	0.04	.97
Körperliche Anforderungen	3.1	(0.8)	4.3	(0.9)	-0.95	.35
Zeitliche Anforderungen	7.7	(1.1)	7.4	(1.4)	0.19	.85
Anstrengung	7.5	(1.0)	8.5	(1.1)	-0.63	.53
Frustration ^{*)}	7.4	(1.6)	3.7	(1.1)	1.84	.08
Ausführung der Aufgaben	6.2	(1.2)	6.1	(1.2)	0.04	.97

^{*)}: Welch-Test, $df_{\text{corr}} = 24.71$

Auf der von 0 bis 100 reichenden Skala des Gesamtscore, bzw. 0 bis 20 bei den Subskalen, zeigt ein höherer Wert höhere Workload an. Von den sechs Subskalen zeigten fünf keine signifikanten Unterschiede (min. $p > .35$). Lediglich auf der Dimension „Frustration“ deutete sich ein Effekt an. Ist *Talk-Ahead* nicht verfügbar, ist die Frustration bei der Bearbeitung von Aufgaben deutlich höher, 7.4 (Kontrollgruppe) vs. 3.7 (Experimentalgruppe). Da bei diesem Vergleich der Levene-Test Varianzheterogenität anzeigt ($F(1,28) = 6.94$, $p = .01$), wurde die Welch-Formel zur Korrektur der Freiheitsgrade angewendet: $t(24.71) = 1.84$, $p = .08$, $p_{\text{rep}} = .84$, $d = 0.67$. Der Unterschied zeigt eine Tendenz zu statistischer Signifikanz.

In der visuell-manuellen Wahlreaktionsaufgabe ergaben sich keine signifikanten Unterschiede zwischen den Gruppen. Allgemein war der Anteil richtiger Antworten sehr hoch und in beiden Gruppen äußerst ähnlich (Kontrollgruppe: $M = 94.4\%$, $SE = 1.3$; Experimentalgruppe: $M = 92.3\%$, $SE = 2.0$; $F(1,28) = 0.53$, $p = .47$, $p_{\text{rep}} = .52$). Eine leichte Verbesserung über die Zeit kann konstatiert werden, die jedoch nur als Tendenz statistisch bestätigt wird (Block 1: $M = 91.0\%$, $SE = 2.7$; Block 2: $M = 92.8\%$, $SE = 1.9$; Block 3: $M = 96.3\%$, $SE = 1.4$; $F(2,56) = 2.39$, $p = .10$, $p_{\text{rep}} = .82$, $\eta_p^2 = .08$). Eine Interaktion liegt nicht vor ($F(2,56) = 0.37$, $p = .69$, $p_{\text{rep}} = .37$). In beiden Gruppen nahmen die Reaktionszeiten signifikant ab (Block 1: $M = 1081$ ms, $SE = 73.4$; Block 2: $M = 929$ ms, $SE = 57.3$; Block 3: $M = 875$ ms, $SE = 53.7$; $F(2,56) = 6.59$, $p < .01$, $p_{\text{rep}} = .97$, $\eta_p^2 = .19$). Auch bei den Zeiten sind keine Unterschiede zwischen den Gruppen (Kontrollgruppe: $M = 995$ ms, $SE = 53.4$; Experimentalgruppe: $M = 929$ ms, $SE = 50.3$; $F(1,28) = 0.38$, $p = .54$, $p_{\text{rep}} = .47$) nachweisbar. Eine Interaktion kann ebenfalls nicht gefunden werden ($F(2,56) = .44$, $p = .64$, $p_{\text{rep}} = .40$).

5.1.2.4 Diskussion

Um ungünstige Eigenschaften von menübasierten sprachgesteuerten Bordsystemen auszugleichen, wurde zur Verbesserung der Bedienbarkeit *Talk-Ahead* als Expertenoption vorgeschlagen und in einem Wizard-of-Oz Simulationsexperiment getestet. Im Folgenden werden die Ergebnisse im Detail diskutiert.

Nutzung der Talk-Ahead Option

Die Versuchsteilnehmer der Experimentalgruppe wurden zu Beginn des Experiments über die *Talk-Ahead* Option informiert, die tatsächliche Verwendung wurde ihnen freigestellt. Eine eingehende Betrachtung der angewandten Eingabestrategie ergab einen klaren Anstieg der Nutzungshäufigkeit von *Talk-Ahead*. Folglich kann von einer guten Akzeptanz der Maßnahme und spontaner Verwendung ausgegangen werden.

Einschränkend muss angemerkt werden, dass durch die Ausführung als Wizard-of-Oz Studie mit einer simulierten Spracherkennerrate von 100% während der experimentellen Durchgänge das Vertrauen der Teilnehmer in die Fähigkeiten der sprachlichen Schnittstelle stark angehoben wurde. Die Vermutung liegt nahe, dass bei schlechten Erkennungsergebnissen die Nutzer zu einem vorsichtigen Verhalten tendieren werden. Typischerweise sprechen sie nach einem Erkennungsfehler sehr langsam (Stifelman, 1993) und/oder neigen zu Hyperartikulation (Oviatt, MacEachern & Levow, 1998). Inwiefern tatsächlich Erkennungsfehler zu einer verminderten Nutzung von *Talk-Ahead* führen, hängt sowohl von der Häufigkeit als auch von der Art der Fehler ab. Bei einer insgesamt niedrigen Erkennungsrate ist die Nutzung der Sprachschnittstelle im Allgemeinen verringert (Kamm, 1994; Kamm & Helander, 1997) und entsprechend auch die Ausbildung von Expertenverhalten. Als besonders problematisch dürften sich Verwechslungsfehler des Spracherkenners (*substitution errors*) erweisen, wenn sie den Nutzer unter Veränderung des Vokabulars in einen anderen Applikationskontext führen. Für *Talk-Ahead* ist dabei von besonderem Interesse, dass die häufig eingesetzten Hidden-Markov-Modelle dazu führen können, dass sich Spracherkennungsfehler begrenzt ausbreiten. Das bedeutet ein nicht erkanntes Wort kann die Erkennung für Wörter in seiner Nachbarschaft verschlechtern (Rhyne & Wolf, 1993). Auf der anderen Seite zählt die exakte Wiederholung einer Eingabe zu den am häufigsten angewandten Fehlerbehebungsstrategien (Shin, Narayanan, Gerber, Kazemzadeh & Byrd, 2002). Somit könnte *Talk-Ahead* auch dazu beitragen, Fehler leichter zu beheben, wenn die Befehlskette den Kontext des richtigen Vokabulars wiederherstellt. Diese stark von der eingesetzten Technologie abhängigen Fragen könnte nur ein Vergleichstest mit aktuellen Spracherkennern beantworten. Ein wichtiger Beitrag zur Erweiterung des Wissens über die Einsetzbarkeit von *Talk-Ahead* wurde mit der vorliegenden Studie geleistet: Es konnte festgestellt werden, dass *Talk-Ahead* grundsätzlich bei den Nutzern auf Akzeptanz stößt und somit eine tatsächliche

Verwendung unter entsprechenden technologischen Rahmenbedingungen wahrscheinlich ist. Erst mit diesem Wissen über das Nutzerverhalten macht es Sinn, aufwendige Tests zur technologischen Realisierung anzugehen.

Aufgabenbearbeitung und Blickdaten

Stand *Talk-Ahead* bei der sprachbasierten Eingabe in einem Menüsystem zur Verfügung, konnte eine Verringerung der Aufgabenbearbeitungszeit um etwa ein Drittel nachgewiesen werden. Ein ähnlicher Effekt stellte sich bei der Gesamtblickdauer auf die Anzeigeeinheit ein. Die TSOT wurde fast um 50% gesenkt. Auch die Dauer der Kontrollblicke nahm deutlich ab. Damit trägt die Implementierung einer *Talk-Ahead* Option zu einem deutlich schnelleren und sichereren Betrieb des Bordsystems bei. Dabei ist zu beachten, dass der Effekt auf die Betrachtungszeit transient sein könnte. Außer der Tatsache, dass sich bei der Verwendung von *Talk-Ahead* schlicht weniger Möglichkeiten für Kontrollblicke ergeben, gibt es kaum zwingende Gründe, warum die Bedienung eines Systems ohne *Talk-Ahead* dauerhaft unter verstärkter visueller Kontrolle erfolgen sollte. Wenn die Sprachkommandos bei SWYS einmal gelernt sind, müssten die Nutzer für eine erfolgreiche Sprachbedienung auch bei mehreren Schritten einer menübasierten Navigation keinen Blickkontakt zum Bildschirm aufnehmen. Allerdings lässt sich in der Praxis häufig beobachten, dass Nutzer von sprachgesteuerten Systemen mit *Say What You See* basierter Eingabe selbst bei eindeutigem akustischem Feedback nach einer Kommandoausführung einen kurzen Orientierungsblick auf die Anzeigeeinheit werfen. Möglicherweise entsteht dieses Verhalten zumindest in der ersten Zeit der Benutzung dadurch, dass die Nutzer einen kurzen Blick auf das Display zum Ablesen der korrekten Kommandovokabel fest in ihre Interaktionsstrategie und den daraus resultierenden Bewegungsablauf integriert haben. In der vorliegenden Studie scheint dies sehr prägnant ausgeprägt gewesen zu sein. Bei einer stärker fordernden Zusatzaufgabe, wie sie die Fahrzeugführung darstellt, könnte ein solches Verhalten jedoch deutlich abgeschwächt werden.

Bei der Analyse der Gesamtblickdauer auf das Display wurde Varianzheterogenität festgestellt und die statistische Analyse daraufhin angepasst. Bryk und Raudenbush (1988) weisen darauf hin, dass Varianzheterogenität kein Zufallsprodukt sein muss und grundsätzlich die Möglichkeit bedacht werden sollte, dass unterschiedliche Varianzen eine systematische Reaktion der Versuchspersonen auf experimentelle Manipulationen reflektieren können. Bei der Entwicklung der Blickdauern über die Zeit in der Experimentalgruppe fällt auf, dass nicht nur die mittlere Dauer kontinuierlich abnimmt, sondern dass auch die Streuung um diese Mittelwerte fortschreitend kleiner wird. Das deutet darauf hin, dass das Nutzen der *Talk-Ahead* Option die Variabilität des Blickverhaltens in der Experimentalgruppe reduziert.

Eine Interpretation der Aufgabenbearbeitungszeit legt die Vermutung nahe, dass sich die kürzere Zeit bei *Talk-Ahead* aus zwei Komponenten zusammensetzt. Ein Teil ist im engeren Sinn technisch bedingt: Wenn zwei Befehle zu einem zusammengefasst werden, verkürzt sich wie eingangs beschrieben die Bearbeitungszeit zwangsläufig um die Dauer einer Verarbeitungslatenz. Es sei hier noch einmal darauf hingewiesen, dass im Vorfeld des Experiments der Wizard auf die Einhaltung möglichst konstanter Auslöselatenzen trainiert wurde. Der vorgefundene Effekt – eine mittlere Reduktion um 4.7 Sekunden – ist jedoch zu groß, als dass er nur durch diese Komponente erklärt werden könnte. Wie die Nutzungsdaten zeigen, wurde im Mittel pro Aufgabe nur etwa ein Kommando eingespart. Nahe liegender ist deshalb, dass ein zweiter Faktor die Bearbeitungszeit zusätzlich reduziert hat: Aus Studien zu Verhaltensänderungen beim Umgang mit sprachgesteuerten Schnittstellen ist bekannt, dass sich Menschen nicht nur den Eigenschaften eines menschlichen Gegenübers anpassen, sondern auch dazu tendieren, eine systematische Adaptation ihrer Sprache an Eigenschaften eines Computerpartners vorzunehmen (Darves & Oviatt, 2002; Zoltan-Ford, 1991). Dazu zählt u.a. die Akkomodation zeitlicher Parameter von sprachlichen Äußerungen, insbesondere die Antwortlatenzen vor der (erneuten) Initiierung eines Dialogschritts. Erlebt folglich der Nutzer das System als einen Dialogpartner, auf dessen Bestätigung (im Fall des Experiments der Signalton) er jedes Mal vor einer Aktion explizit warten muss, sollte sich dies auf die Dynamik des Interaktionsverhaltens in der hier vorgefundenen Weise auswirken. Weil das System ohne *Talk-Ahead* keine schnelle Eingabefolge zulässt und regelmäßig kurze Warteintervalle verlangt, ist es sehr wahrscheinlich, dass der Nutzer, möglicherweise unwillkürlich, sein Eingabeverhalten mehr als notwendig verlangsamt hat. Die Schlussfolgerung liegt nahe, dass die Nutzer bei Verfügbarkeit der *Talk-Ahead* Option auf Dauer eine andere Strategie entwickeln, an die Schnittstelle heranzugehen. Sie müssen weniger Rücksicht auf Rückmeldungen des Systems nehmen und können besser ihr eigenes Interaktionstempo etablieren. Inwiefern durch die Unterstützung eines höheren Interaktionstempos auf Dauer mehr nutzerseitige Fehler, z.B. bei der Zielsetzung oder der sprachlichen Formulierung, entstehen, kann nur durch Feldstudien und systematische längsschnittliche Beobachtung im praktischen Einsatz überprüft werden.

Als eine der Hypothesen des Experiments wurde die Vermutung aufgestellt, dass die Versuchsteilnehmer etwas Übung mit Spracheingabe und Erfahrung mit der Menüstruktur sammeln müssen, bevor sie von *Talk-Ahead* Gebrauch machen können. Vor diesem Hintergrund wäre anzunehmen, dass sich bei Blickdauer und Aufgabenbearbeitungszeit anfangs noch kein Unterschied zwischen den Gruppen ergeben sollte und erst in Block 2 oder 3 Differenzen entstünden. Weder bei der Gesamtbetrachtungszeit noch bei den Bearbeitungszeiten war eine entsprechende Interaktion nachweisbar, die Dauer einzelner Blicke wies sie nur tendenziell auf. Die Begründung hierfür liegt mit

hoher Wahrscheinlichkeit in den sehr ausführlichen Probedurchgängen und in dem relativ geringen Schwierigkeitsgrad der Aufgaben. Faktisch haben die Testpersonen fünf Blocks mit sehr ähnlichen, zumindest vom Prinzip her identischen Aufgaben absolviert. Da die ersten beiden Blöcke explizit als Probedurchgänge angekündigt wurden und während der Bearbeitung der Aufgaben in diesen Durchgängen zwischen Versuchsleiter und Teilnehmer Klärungsgespräche stattgefunden haben, ist eine Auswertung dieser Daten weder möglich noch sinnvoll. Streng genommen konnten die Teilnehmer sogar bereits bei der Demonstration des Spracherkenners beginnen, eine interne Repräsentation von der Funktionsweise der Schnittstelle aufzubauen. Somit ist durchaus möglich, dass die gesammelte Erfahrung während der Probedurchgänge ausreichend war, um bereits im ersten experimentellen Block von der Expertenoption Gebrauch machen zu können. Dies wird durch die nachgewiesene Nutzung von *Talk-Ahead* bereits im ersten Block unterstützt. Auf Basis der vorhandenen Daten ist dieser Punkt nicht abschließend zu klären.

Workload

Weder die Ergebnisse in der Sekundäraufgabe noch die Selbstauskünfte der Versuchsteilnehmer ließen auf erhöhte Workload bei *Talk-Ahead* schließen. In der Experimentalgruppe wurden auch nicht mehr Eingabefehler als in der Kontrollgruppe festgestellt. Die Befunde deuten an, dass sich diese Expertenoption tatsächlich durch Automatisierung mit fortschreitender Kenntnis der Menüstruktur entwickelt und dass das Bilden der Befehlsketten nicht in einem Ressourcen fordernden Prozess entsteht. Diese Interpretation stützt sich bezüglich der Sekundäraufgabe auf die sehr ähnlichen Reaktionszeiten. Der Anteil korrekter Antworten war in beiden Bedingungen sehr hoch, so dass ein Ceiling Effekt wahrscheinlich ist. Die mittleren Reaktionszeiten hingegen sind mit etwas über 900 Millisekunden als Antwort auf einen visuellen Stimulus in dieser einfachen Wahlreaktionsaufgabe mit nur zwei Alternativen vergleichsweise lang und können somit eher als sinnvolles Maß herangezogen werden.

Die Raw-Variante des NASA-TLX zeigte auf der Ebene der Gesamtworkload keine Unterschiede an. Dies ist insofern beachtenswert, als seit seinem Erscheinen viele Studien den Nachweis geführt haben, dass der TLX eines der validesten Workloadmaße ist, sehr sensitiv auf experimentell manipulierte Workloadunterschiede anspricht und gleichzeitig eine sehr geringe Between-Subject Variabilität aufweist (Aretz, Johannsen & Ober, 1996; Byers, Bittner, Hill, Zaklad & Christ, 1988; Hart & Wickens, 1990; Vidulich & Tsang, 1986). Unter Einbeziehung der Daten aus der Sekundäraufgabe erscheint es durchaus angemessen, diesen Nulleffekt des Gesamtscore wie geschehen inhaltlich zu interpretieren. Auf Basis der Einzeldimensionen hätte sich bei erhöhtem Planungsaufwand unter *Talk-Ahead* zumindest bei „geistige Anforderungen“ ein Unterschied ergeben müssen. Vor dem Hintergrund einer gebrauchsfreundlichen

Gestaltung ist hingegen durchaus relevant, dass allein die Dimension „Frustration“ nachweisbare Unterschiede aufwies. Als Kerneigenschaft von *Talk-Ahead* wurde postuliert, dass es dem Bedarf der Nutzer nach effizienterer Menünavigation nachkommt. Bei der Sprachschnittstelle mit eingeschränktem Funktionsspektrum in der Kontrollgruppe hat dieses Defizit zu einer negativeren Bewertung des erlebten Interaktionsverlaufs geführt.

5.1.3 Zusammenfassende Diskussion und Gestaltungsempfehlungen

Multimodales *Talk-Ahead* bezeichnet die Verknüpfung von Sprachbefehlen in einer Bedienungsumgebung, die neben der auditiven Bedienführung auch visuelle Informationen zur Bedienung der Sprachschnittstelle durch *Say What You See* ermöglicht. Da in komplexen Systemen *Say What You See* mit hierarchischen Menüs einhergeht, ist eine Expertenoption notwendig, die eine schnelle Navigation auch per Sprache zulässt. In einem Simulationsexperiment wurde untersucht, inwiefern *Talk-Ahead* diesem Bedarf nachkommt. Die Ergebnisse deuten darauf hin, dass die Möglichkeit einer Verknüpfung von Sprachkommandos bei *Talk-Ahead* gut angenommen wird, sowie positive Effekte auf den Interaktionsprozess und die visuelle Abwendung hat. Entscheidend ist, dass diese Effekte nicht mit erhöhtem Planungsaufwand einhergehen, wie die erhobenen Workloadmaße zeigen.

Zusammenfassend hat sich *Talk-Ahead* als sinnvolle Einrichtung bewährt, die im Systemdesign bedacht werden sollte. Damit multimodales *Talk-Ahead* in der Praxis erfolgreich zum Einsatz kommen kann, sollten die folgenden Aspekte bedacht werden:

- Damit die Nutzer mit zunehmender Erfahrung aus den Begriffen der Menüebenen Chunks bilden können, die die kognitive Voraussetzung für *Talk-Ahead* sind, muss ein klar gegliedertes, hierarchisches Menüsystem geschaffen werden.
- Um eine maximale Übertragbarkeit des Wissens aus der Bedienung der manuellen Schnittstelle für die Sprachschnittstelle zu gewährleisten, müssen diese möglichst parallel angelegt werden. Eine Gestaltung der grafischen Bedienoberfläche zur Eingabe über *Say What You See* kann maßgeblich dazu beitragen.
- Die Kommandos im Rahmen von *Say What You See* müssen kurz und eindeutig sein, um auf der grafischen Oberfläche mit hinreichend großer Schrift dargestellt werden zu können. Darüber hinaus müssen die Formulierungen der einzelnen Menüebenen oder Befehlsflächen derart gestaltet sein, dass sie sich einfach in einer Kette hintereinander sprechen lassen.

- Wie bereits dargelegt zählt die Wiederholung einer Eingabe zu den häufigsten Strategien zur Fehlerbehebung (Shin et al., 2002). Die Verkettung von Befehlen unter *Talk-Ahead* führt häufig zu einem Kontextwechsel in der Applikation und damit auch zu einem Wechsel von Grammatik und Vokabular. Um im Fall von Fehlerkennungen die Korrektur durch Wiederholung zuzulassen, sollte die letzte Befehlskette trotz eines Kontextwechsels aktiv gehalten werden.
- Die tatsächliche Anwendung von *Talk-Ahead* ist von der Güte der Spracherkennungsqualität und -geschwindigkeit abhängig. Ist das System wenig performant, werden die Nutzer sehr zurückhaltend explorieren und kaum Expertenverhalten entwickeln. Deshalb sollte grundsätzlich bereits in einer frühen Entwicklungsphase die Eignung der eingesetzten Engine für *Talk-Ahead* unter realistischen Bedingungen evaluiert werden.
- Wenngleich durch *Talk-Ahead* die Nachteile von multiplen Erkennerlatenzen bei Spracheingabe in menübasierten Systemen vermindert werden kann, wird zusätzlich eine wohl durchdachte Shortcut-Strategie empfohlen. In separaten Tests muss vorab ermittelt werden, welche Elemente des Systems in der Repräsentation des Nutzers auf „oberster Ebene“ anzusiedeln sind. Sprachliche Direktzugriffe auf diese Funktionen bleiben trotz *Talk-Ahead* obligatorisch.

5.2 Verbale versus ikonografische Funktionsdarstellung

5.2.1 Problemstellung

Variabilität im verbalen Ausdruck ist ein typisches Charakteristikum menschlichen Kommunikationsverhaltens und gleichzeitig eine der großen Herausforderungen für sprachgesteuerte Systeme. Mit Hilfe des *Say What You See* (SWYS) Eingabepinzips kann im multimodalen Umfeld eine Einschränkung des Wortschatzes erreicht werden, die zu einer deutlichen Reduktion von Out-of-Vocabulary Fehlern führen kann. Problematisch hingegen ist bei der Anwendung im Fahrzeug, dass die explizite verbale Angabe verfügbarer Funktionen auf dem Bildschirm des Bordsystems zu einer sehr großen Anzahl textueller Elemente auf der Benutzeroberfläche führt. Einige Anhaltspunkte deuten darauf hin, dass möglicherweise die Verwendung von bildhaften Funktionsdarstellungen unter dem Aspekt der visuellen Suche vorteilhaft wäre und dadurch insbesondere die visuell-manuelle Interaktion besser unterstützen würde. Unter diesen Umständen wäre die sinnvolle Anwendung von SWYS in Frage gestellt.

Die ergonomische Literatur empfiehlt insbesondere bei begrenzter räumlicher Darstellungsfläche die Verwendung bildlichen Funktionsdarstellungen (gute Übersichten bieten u.a. Cher, Campbell & Mitchell, 1998; Kebeck, Cieler & Pohlmann, 1997;

Shneiderman, 1987; Stauer, 1987). Das Aufzeigen verfügbarer Funktionen durch Icons birgt zusätzlich das Potential, das visuelle Auffinden einer Funktion für die manuelle Auswahl zu beschleunigen (z.B. Arend, Muthig & Wandmacher, 1987)⁹. Darüber hinaus kann die bildliche Darstellung unter Nutzung des Bildüberlegenheitseffekts die Memorierbarkeit von Bildschirmhalten erhöhen (vgl. Clark & Craig, 1992; für eine Diskussion des Bildüberlegenheitseffekts siehe Engelkamp, 1991; Paviio, 1971). Die Kapazität des Gedächtnisses für visuell-bildhaftes Material hat sich in Grundlagenuntersuchungen sowohl bei der Kurzzeitspeicherung als auch bei der Langzeitspeicherung von Information in eindrucksvoller Weise demonstrieren lassen (Nickerson, 1965; Standing, Conezio & Haber, 1970). In angewandten Studien konnte zwar kein derart drastischer Effekt von Icons auf die Behaltensleistung gezeigt werden, eine Verbesserung und Unterstützung des Suchprozesses ist jedoch trotzdem zu erwarten (Lansdale, 1988). Wenn eine bedeutungshaltige Verbindung zwischen dem Icon und seinem Zweck geknüpft werden kann, wird die Behaltensleistung deutlich verbessert (Lansdale, Simpson & Stroud, 1990). Die Verwendung von Metaphern bei der Konstruktion von Icons kann darüber hinaus die Erlernbarkeit des Systems steigern (Carroll & Mack, 1999). Eine Übersicht über weitere Vorteile von Icons gibt Horton (1994). Damit diese zum Tragen kommen, ist zum einen notwendig, dass bei den Rezipienten *Visual Literacy* angenommen werden kann, also gewisse Fähigkeiten zum Entschlüsseln bildlicher Codes ausgeprägt sind (Weidenmann, 1988). Diese müssen erlernt werden, sind kulturell beeinflusst und können nicht immer vorausgesetzt werden. Zum anderen muss über die allgemeine Befähigung zur Dekodierung hinaus die spezielle Bedeutung eines Icons verstanden oder erschlossen werden können. Sind diese Bedingungen gegeben, wäre die Verwendung von Icons durchaus vorteilhaft: Durch den Gedächtnisvorteil wäre der Aufbau des Systems schneller zu erlernen und die kürzere visuelle Suche würde die manuelle Eingabe vereinfachen sowie für kürzere

⁹ Icons sind Bildzeichen, die in der Mensch-Maschine-Interaktion Anwendung finden. Sie werden meistens durch den Abstraktionsgrad der Abbildung zwischen Bild und Referent klassifiziert. Dabei ist zu beachten, dass im engeren Sinn alle Icons dahingehend Abstraktionen sind, dass sie nur eine Teilmenge der Eigenschaften des Referenten ausdrücken (Familiant & Detweiler, 1993). Charwat (1994) unterscheidet unter dem Ordnungsbegriff Bildzeichen Piktogramme und Ideogramme (Sinnbilder). Piktogramme geben reale Gegenstände oder beobachtbare Tätigkeiten bildhaft stilisiert und vereinfacht wieder und sind meist selbsterklärend. Ideogramme sind abstrakte geometrische Formen oder grafische Symbole, die für eine Operation stehen und häufig einer Erklärung bedürfen. Grafisches Symbol und Ideogramm werden nach Charwat synonym verwendet. Blattner et al. (1989) fügen der Terminologie noch eine Mischklasse hinzu. Neben konkret-bildhaften, repräsentationalen Icons (Piktogrammen) und abstrakten Icons (Ideogrammen) unterscheiden sie semiastrakte Icons: Hierbei handelt es sich um die Kombination eines bildhaften Elements mit einem abstrakten Symbol. Häufig drückt dabei das Symbol, z.B. ein Pfeil, die Art der Aktion mit oder auf dem Objekt aus. Diese klare Terminologie ist in der Literatur üblicherweise nicht in dieser Schärfe anzutreffen. In vielen Fällen wird „Piktogramm“ als Oberbegriff synonym mit „Icon“ verwendet.

Blickdauern auf die Anzeigeeinheit sorgen. Wiedenbeck (1999) weist darauf hin, dass all diese Ansprüche über die Potentiale von Icons implizit die Verwendung von Text im Interface als Baseline einbezögen und die meisten von ihnen auf psychologischer oder perzeptueller Ebene auf eine Grundformel reduzierbar seien: Icons wären einfacher zu verarbeiten als Text. Gegenstand der bisherigen Untersuchungen war dabei meist der Kontext einer Bildschirmarbeitsplatzumgebung und die Optimierung der Darstellung für manuelle Eingabe. Im Kontext multimodaler Systeme, die auf *Say What You See* zurückgreifen wollen und bei denen die visuelle Zuwendung zur Anzeigeeinheit kürzestmöglich sein sollte, stellt sich die Frage unter einem anderen Blickwinkel mit neuer Tragweite.

Einige Anzeichen sprechen dafür, dass Icons in einem System, das auch Spracheingabe zulässt, weniger von Vorteil sind. Grundsätzlich kann SWYS bei bildhaftem Material nicht greifen, da Icons offensichtlich keinen eindeutigen Sprachbefehl nahe legen. Bereits die Natur von Icons als symbolische Abbildungen lässt je nach Ausprägung einen geringen bis teilweise deutlichen Spielraum bei der Interpretation des repräsentierten Konzepts zu. Selbst wenn das repräsentierte Konzept eines Icons bekannt ist, wie z.B. das Dreieck als Wiedergabe-Symbol bei Stereoanlagen, sind weiterhin verschiedenste Verbalisierungen denkbar („Wiedergabe“, „Start“, „Play“, etc.). Ives (1982) merkt dazu an, dass ein Icon zwar mehr wert sei als tausend Worte, es seien aber nicht immer genau die tausend Worte, die der Designer im Kopf habe. Eine feste Verbindung von Bild und Wort kann auch aus gedächtnispsychologischer Sicht nicht erwartet werden, wenn man bedenkt, dass die Verarbeitung von Information für verbale und nonverbale Information relativ unabhängig ist und auch die Speicherung in unterschiedlichen Repräsentationsformaten erfolgt. Nach der multimodalen Gedächtnistheorie (Engelkamp, 1991) führt die Darbietung von Bildern ausschließlich zur Aktivierung der Bildmarken, d.h. der bildhaft gespeicherten Einheiten sensorischen Wissens, sowie deren Konzepte. Auf experimentelle Befunde gestützt argumentiert Engelkamp, dass bei der Wahrnehmung Bilder auch nicht implizit benannt werden (weder Wortmarken noch Sprechprogramme aktiviert werden). Generell wird angenommen, dass man beim Abruf nur eine Abstraktion des Bildes erinnert, die die allgemeine Bedeutung erfasst (Anderson, 2000). Damit ist ebenfalls aus dieser Perspektive unwahrscheinlich, dass mehrere Nutzer dasselbe Wort mit einem Icon verbinden würden bzw. dass ein Nutzer immer dasselbe Wort verwendet, da keine feste Bindung zwischen Bild und Wort existiert. Ohne klare Einschränkungen oder Vorgaben ist die für menschliches Kommunikationsverhalten typische Variabilität im Wortschatz zu erwarten (Furnas et al., 1987).

Basierend auf empirischen Befunden wird in der Usability Literatur empfohlen, Icons und verbale Beschriftungen gemeinsam anzubieten (Bewley, Roberts, Schroit & Verplank, 1983; Egido & Patterson, 1988; Haramundanis, 1996). Der geringe Platz auf

der Anzeigeeinheit eines Bordsystems im Kfz mit meist nur sieben oder acht Zoll Bilddiagonale erschwert allerdings eine durchgängig redundante Darbietung von Icons und Beschriftung, bei denen dann kaum die entsprechenden Gestaltungsrichtlinien bezüglich Schriftgrößen und Abständen eingehalten werden können (Campbell, Carney & Kantowitz, 1997; Campbell, Richman, Carney & Lee, 2004).

Wie bereits dargelegt, ist einer der entscheidenden Schlüsselfaktoren für die Gebrauchstauglichkeit einer Mensch-Maschine-Schnittstelle im Fahrzeug die Unterstützung einer möglichst kurzen visuellen Abwendungszeit von der Primäraufgabe. Ein multimodales System muss ferner sicherstellen, dass alle Eingabemodalitäten sinnvoll nutzbar sind. Wenn folglich textbasierte Funktionsdarstellung auf dem Bildschirm mit gravierenden Nachteilen für die visuelle Orientierung einhergeht, sollte bei der Gestaltung für die visuell-manuelle Interaktion auf ikonografische Funktionsdarstellung zurückgegriffen werden. Für die Strategie zur sprachlichen Interaktion würde das bedeuten, dass vollständig auf die auditive Schnittstelle ausgewichen werden müsste und SWYS nicht als grundlegendes Interaktionsprinzip empfohlen werden könnte. Damit klare und praxisnahe Gestaltungshinweise für diese Konfliktsituation gegeben werden können, soll zunächst eine Zusammenfassung bisheriger empirischer Studien zum Vergleich piktografischer und textueller Schnittstellen den Stand der Forschung darlegen. Darauf aufbauend wird der Ansatz für die Experimente dieses Abschnitts abgeleitet, deren Ziele eine systematische Analyse der visuellen Suche bei bildlicher vs. textueller Funktionsdarstellung und die Ableitung von Empfehlungen für das Design der visuellen Schnittstelle sind.

5.2.2 Grundlagen der Verwendung von Icons und verbalen Kennzeichnungen

Piktografische Kennzeichnungen für Instrumente und Maschinen wurden vor allem seit den siebziger Jahren verstärkt eingeführt und untersucht (z.B. Cahill, 1975; Cahill, 1976; Easterby, 1970). Mit Hilfe neuer Evaluations- und Prüfverfahren hat u.a. Green erste systematische Tests für Piktogramme im Kraftfahrzeug durchgeführt (Green, 1981; Green & Pew, 1978).

Der Großteil der Untersuchungen zu Icons für den Computer wurde in den achtziger, teilweise in den frühen neunziger Jahren mit der beginnenden Verfügbarkeit höher auflösender Anzeigen durchgeführt. Als einer der Startpunkte der umfassenden Einführung von Icons im Bereich Mensch-Computer-Interaktion gilt die Arbeit von Smith und Kollegen zur Benutzeroberfläche des Xerox 8010 Star Computers (Smith, Irby, Kimball, Verplank & Harslem, 1982) mit den anschließenden systematischen Evaluationen der Icons (Bewley et al., 1983). Viele Studien fokussieren nicht primär auf den Vergleich der Eignung von Icons und verbalen Kennzeichnungen, sondern auf grund-

sätzliche Fragen der Gestaltung wie Layout, Verständlichkeit, Identifizierbarkeit, semantische Konstruktion und allgemeine Eigenschaften piktografischer Darstellungen (vgl. Gittins, 1986; Horton, 1994; Kebeck et al., 1997; Rogers, 1989; Shneiderman, 1987; Stauer, 1987). Dabei wurde beispielsweise in der Studie von Bewley et al. (1983) mit Hilfe verschiedener Icon Sets die Unterscheidbarkeit, Bekanntheit und Wiedererkennbarkeit von Icons evaluiert. Eines der Schlüsselergebnisse war die Erkenntnis, dass Icons mit gleichzeitiger verbaler Beschriftung zwar am besten verstanden werden, aber auch die längste Verarbeitungszeit erfordern. In einer groß angelegten Befragungsstudie unter Anlagenfahrern von Gebäudeautomationssystemen fanden Guastello und Kollegen bei dem direkten Vergleich von Icons, Icons mit Beschriftung und Beschriftung ohne Icon ebenfalls einen Vorteil bei der Verständlichkeit für die redundante Icon-plus-Text Darstellung (Guastello, Traut & Korienek, 1989). Bei einfacher Darstellung ließ sich jedoch keine klare Vorrangstellung für entweder Icons oder Beschriftung nachweisen, tendenziell ergab sich eine Präferenz der Teilnehmer für textuelle Angaben. Beschriftungen, die mit starken Abkürzungen arbeiten, haben sich in dieser Studie als problematisch erwiesen. Kacmar und Carey (1991) weisen auf die besondere Sorgfaltspflicht bei der Konstruktion der Icons hin: Ist die Abbildung des Referenten nicht gelungen oder der Kontext der Funktion oder des abgebildeten Elements nicht eindeutig, können sogar sehr einfache Alltagssymbole falsch interpretiert werden.

Trotz des hohen Verbreitungsgrads von Icons für diverse Anwendungen existiert nur eine vergleichsweise geringe Zahl an kontrollierten experimentellen Studien, die die Erlernbarkeit und Leistungsfähigkeit dieses Ansatzes untersucht haben. Vor allem für den direkten Vergleich der Effekte von Icons und verbalen Beschriftungen auf Such- oder Aufgabenbearbeitungszeiten besteht Forschungsbedarf.

5.2.2.1 Experimente zu visueller Suche

Experimentelle Studien zu visueller Suche haben entscheidend dazu beigetragen, relevante Faktoren für das Design von Elementen auf grafischen Benutzeroberflächen zu identifizieren und ihre Auswirkungen auf das Suchverhalten zu evaluieren. Dazu zählen grundlegende Eigenschaften wie Farbe, Form, Orientierung, Linienverläufe und -unterbrechungen, Größe und Tiefenhinweise, die Treisman als Elementarmerkmale bezeichnet (Treisman & Gelade, 1980). Viele Forschungsarbeiten haben sich damit beschäftigt, wie diese Merkmale dazu führen, dass ein Reiz aus der Umgebung hervorsticht. Es konnte z.B. gezeigt werden, dass Farbe einer der effektivsten Wege ist, einen Stimulus aus seiner Umgebung hervorzuheben (Wolfe, 1994): Egal, ob man einen roten Kreis unter fünf oder unter fünfzig blauen Kreisen finden muss, die Suche wird immer sehr effizient sein. Sie erfordert kein schrittweises Absuchen des Bildschirms und findet präattentiv statt. Kombiniert man mehrere Elementarmerkmale in einem

Stimulus, wird die Suche weniger effizient. Der Bildschirminhalt kann nicht mehr als Ganzes erfasst werden, die Elemente müssen unter Aufwendung selektiver Aufmerksamkeit einzeln sukzessive untersucht werden.

Die Befunde aus der Grundlagenforschung zu visueller Suche sind wegen des speziellen Reizmaterials nicht einfach auf den Vergleich Icons vs. verbale Beschriftung von Funktionen verallgemeinerbar. Denkbar ist natürlich die Annahme, dass Icons, die aus wenigen Elementarmerkmalen bestehen, schneller zu finden sein sollten als Wörter. Trotzdem erscheinen eigens konstruierte Untersuchungen zu Icons und Text aus mehreren Gründen notwendig. Im Standardparadigma müssen die Versuchspersonen ein Target-Item unter mehreren Distraktoren finden, Reaktionszeit und Genauigkeit dienen als abhängige Variable (Wolfe, 1998). Wenngleich viele Variationen dieses Paradigmas existieren, werden aus wenigen Elementarmerkmalen zusammengesetzte Elemente oder nur einzelne Buchstaben verwendet, die keine semantische Bedeutung wie Icons oder Wörter haben. Sobald der Faktor Semantik selbst auf der Ebene sehr einfacher visueller Stimuli hinzukommt, können die Ergebnisse von üblichen Befunden der Forschung zu visueller Suche deutlich und überraschend abweichen. Dies zeigte sich in einer Studie von Rauschenberger und Yantis (2006): Wird ein Kreis mit einer überschneidenden Linie so rotiert, dass er wie ein „Q“ aussieht, war die Sucheffizienz deutlich höher als bei anderen Rotationen. Dieser semantische Effekt konnte in einer Studie mit chinesischen Versuchspersonen, die keinen Kontakt mit dem lateinischen Alphabet hatten, nicht hergestellt werden. Sie profitierten nicht von der Ähnlichkeit des Stimulus zu dem Buchstaben „Q“ (Rauschenberger & Chu, 2006).

Neben der Komplexität und Semantik der Stimuli kommt ein wesentlicher Unterschied im Aufbau von Grundlagenexperimenten zu visueller Suche und angewandten Studien hinzu: Im angewandten Kontext kann jedes Element auf dem Bildschirm in einem Durchgang Target und im nächsten Durchgang bereits wieder Distraktor sein. Bei Standardexperimenten existiert diese Doppelbelegung des Reizmaterials typischerweise nicht. In der Literatur wird das Problem diskutiert, dass ein Target unter Non-Targets extrem hervorgehoben und die Suche damit willkürlich effizient gemacht werden könnte (Duncan & Humphreys, 1989). Wechseln die Items hingegen häufig den Status zwischen Ziel und Distraktor, wird ein einzelnes, besonders hervorgehobenes Element die Suche nicht unbedingt insgesamt verbessern. Im Gegenteil, die ablenkende Wirkung dieses Items kann sogar zu Interferenzen führen. Durch diesen Umstand muss entsprechend neben der Qualität des Targets zu einem gegebenen Zeitpunkt auch die der Distraktoren betrachtet werden. Aus der Grundlagenliteratur können kaum Anregungen gefunden werden, da die Non-Targets zumeist als reine Distraktoren gesehen worden, die keinem besonderen Enkodierungsprozess unterzogen werden. Erst in jüngerer Vergangenheit wird argumentiert, dass die perzeptuelle

Qualität (Prägnanz) der Non-Targets die visuelle Suche entscheidend beeinflussen kann (vgl. Rauschenberger & Yantis, 2006).

5.2.2.2 Untersuchungen mit simulierten Systemen

Wie bereits beschrieben können aus der Forschung mit Mensch-Maschine-Hintergrund ebenfalls nicht ohne weiteres eindeutige Rückschlüsse auf die optimale Funktionsdarstellung gezogen werden: Während auf der einen Seite berichtet wird, dass gut gestaltete Symbole schneller und genauer als Zeichen mit verbaler Beschriftung erkannt werden, berichten andere Studien, dass kein Unterschied festgestellt werden kann. Schwierig einzuordnen sind hier zum Teil die Befunde, die auf einem Test vollständiger Systeme basieren (z.B. Whiteside, Jones, Levy & Wixon, 1985). Diese haben den Vorteil, die Komplexität vieler Programme zumindest ansatzweise nachzubilden. Allerdings sind teilweise die Ursachen bei fehlenden Unterschieden nicht klar und die Analyseebene der Aufgabenbearbeitungszeiten erscheint sehr grobkörnig.

Whiteside et al. (1985) testeten sieben verschiedene Systeme, die speziell für eine Eingabe über die Kommandozeile, Icons oder Menüs ausgelegt waren. Weder für Experten noch für Novizen konnten Whiteside und Kollegen Vorteile bei der Bedienung (Genauigkeit und Zeit) für ikonografische Darstellung finden. Der Aufbau der Studie lässt allerdings keinen genauen Rückschluss darauf zu, welche Faktoren dieses Ergebnis verursacht haben. Verbale Beschriftungen kamen nur in Menüs vor, während Icons auf der Benutzeroberfläche direkter Manipulation zugänglich waren. Benbasat und Todd (1993) argumentieren, dass der eigentliche Unterschied zwischen menübasierter Eingabe und direkter Manipulation und nicht zwischen Icons und Text zu finden ist. In ihren Experimenten, die diese Faktoren getrennt untersuchten, ergab sich ein deutlicher Unterschied für die Art der Schnittstelle (direkt vs. Menüs), aber kein Unterschied für die Art der Darstellung (Icons vs. Text). Es fällt jedoch auf, dass die Aufgabenbearbeitungszeiten insgesamt vergleichsweise lang waren. Betrachtet man das insgesamt etwas schnellere Direct Manipulation Interface, ergaben sich bei Benbasat und Todd für den ersten Durchgang 495 (Icons) und 565 Sekunden (Text). Diese Zeit nahm monoton bis zum dritten letzten Durchgang ab und fiel auf immerhin noch 246 (Icons) und 257 (Text) Sekunden. In dieser Studie konnte klar gezeigt werden, dass sich auf der Ebene längerer Aufgabenbearbeitung die Art der Darstellung nicht bedeutsam auswirkt und die Erwartungen, die an Icons gestellt werden, in mancher Hinsicht nicht erfüllt werden. Inwiefern für einzelne Verarbeitungsschritte nicht doch zeitliche Unterschiede entstehen, bleibt hingegen unklar. Leider wird die Analyse der Aufgabenbearbeitungszeiten in der Studie bezüglich einzelner Nutzeraktionen nicht weiter vertieft. Beachtlich ist immerhin, dass im ersten Durchgang ein Unterschied von fast einer Minute zugunsten der Icons existiert. Weiterhin wäre durchaus relevant, inwiefern einzelne Auswahlsschritte – die Zeit von der Sichtbarkeit des Bildschirm-

halts bis zur Anwahl des Elements – Unterschiede erbracht hätten. Auf der relativ grobkörnigen Analyseebene der Gesamtbearbeitungszeit von Aufgaben kann diese Information untergehen. Hier sind Unterschiede im Bereich um oder unter einer Sekunde zu erwarten. Die Streuung der Aufgabenbearbeitungszeiten kann einen kumulativen Effekt in der Größenordnung einiger Sekunden verdecken. Für den Anwendungsfall Bildschirmarbeitsplatz ist diese Differenz natürlich nicht besonders relevant, bei einer stabilen Verkürzung der Blickabwendung von der Straße auf einen Bordcomputer um mehrere hundert Millisekunden pro Interaktionsschritt wäre jedoch ein deutlicher Gewinn erzielt.

Ein ähnliches Problem lässt sich in der Studie von Wiedenbeck (1999) erkennen. Ihr Experiment untersuchte die Effekte von piktografischer, verbaler und kombinierter (Icons plus Beschriftung) Repräsentation von Funktionen sowohl auf der Ebene von Leistungsparametern bei der Bearbeitung als auch unter subjektiven Gesichtspunkten. Der Test war über zwei Testsitzungen mit einer Woche Abstand verteilt. Auch hier ergaben sich relativ lange Aufgabenbearbeitungszeiten von mindestens 165 bis maximal 719 Sekunden. Während die Leistungsparameter in der ersten Sitzung darauf hindeuteten, dass die Darstellung mit Text allein und Icons plus Text zu besseren Ergebnissen führte, zeigte sich in der zweiten Sitzung kein Unterschied mehr. Der Unterschied in der Aufgabenbearbeitungszeit konnte zwischen den Gestaltungsvarianten allerdings bereits in den späteren Blöcken der ersten Sitzung nicht mehr nachgewiesen werden. Insgesamt wurde in dieser Studie bezüglich der Bearbeitungszeit tendenziell ein Nachteil für die Darstellungsvariante „Icons allein“ gefunden, während „Icons plus Text“ und „Text allein“ besser abschnitten. Bezieht man die anderen Leistungsparameter mit ein, wird die Interpretierbarkeit dieses Befunds allerdings eingeschränkt: Auf der Ebene der Fehler und auch bei der Häufigkeit des Aufrufs der Hilfefunktion ergaben sich signifikante Unterschiede, Icons allein führten zu deutlich mehr Fehlern und einem erhöhten Bedarf an externer Hilfe. Es stellt sich die Frage, inwiefern das Gestaltungsmerkmal ikonografische Darstellung zu den langsameren Aufgabenbearbeitungszeiten geführt hat oder ob die Ausprägung der verwendeten Icons dies verursacht hat. Die Tatsache, dass sowohl Hilfeaufrufe als auch Fehler deutlich höher waren, deutet zumindest darauf hin, dass es mit der Verständlichkeit der Icons Probleme gab. Nur von gut gestalteten Icons ist zu erwarten, dass sie schneller und genauer erkannt werden als verbale Marken. Auf der anderen Seite ist lange bekannt, dass bei der Suche nach grafischen Symbolen hohe Fehlerraten mit hohen Reaktionszeiten einhergehen (Williams & Falzon, 1963). Um also den Effekt von Icons und Text auf die Bearbeitungszeit isoliert betrachten zu können, sollten Icons verwendet werden, die von den Nutzern hinreichend gut verstanden werden. Nur dann kann klar ausgesagt werden, ob selbst gut verständliche Icons zu längerer oder gleicher

Bearbeitungszeit führen, und ein relativ trivialer Effekts zu Ungunsten der Icons könnte ausgeschlossen werden.

In einer Studie von Egido und Patterson (1988) wurden ebenfalls Bilder, verbale Marken und die Kombination Bilder plus verbale Marken in einer Art Navigationsaufgabe verglichen. Ziel der Aufgaben war, ein Bild in einer Datenbank zu finden. Auch in dieser Studie fand sich tendenziell ein Nachteil für die Darstellung von bildhaftem Material ohne zusätzliche Beschriftung, während Text oder die kombinierte Darstellung von Text und Bild zu kürzeren Suchzeiten führte. Wenngleich diese Studie in der HCI Literatur häufig zitiert wird, muss doch kritisch angemerkt werden, dass es sich nicht um typische Computer-Icons handelte und die Ausführung der Aufgaben kritisch am semantischen Wissen der Versuchspersonen hing. Als Icons dienten Bilder von z.B. Vögeln und Meerestieren, diese wurden mit und ohne verbale Beschriftung angezeigt. Dieses Reizmaterial ist zwar eine bildlich-symbolische Repräsentation, berücksichtigt aber in keiner Weise die notwendigen Gestaltungsrichtlinien, die für den vorteilhaften Einsatz von Icons notwendig sind (Shneiderman, 1987). Mit Hilfe der Bilder bzw. der verbalen Marken musste der Zusammenhang der angezeigten Elemente untereinander erkannt werden, da es die Aufgabe der Teilnehmer war, durch das System zu navigieren und ein spezielles Bild am Ende der Menühierarchie zu finden. Bei der Bedingung „Icons allein“ musste also bei der Menüauswahl zwischen einer Qualle und einem Wal erkannt werden, dass es sich um ein wirbelloses bzw. Wirbeltier handelt (diese Information war in den anderen Bedingungen textuell gegeben). In den eigentlichen Vergleich der Darstellungsformen spielt hier als Störvariable das Hintergrundwissen der Versuchspersonen hinein, korrekt die Eigenschaften des Zielelements auf die übergeordneten Prototypen abzubilden bzw. sie in diesen zu erkennen. Diese Aufgabe teilt nur wenige Merkmale mit den typischen Problemen bei der Mensch-Maschine-Interaktion.

Insgesamt kann man jedoch aus den Studien folgern, dass sich die Erwartungen in die Effekte piktografischer Darstellung wohl nicht uneingeschränkt erfüllen lassen. Wenn Vorteile vorhanden sind, zeigen sie sich zumindest nicht ohne weiteres auf der Analyseebene eines Gesamtsystems. Bei dem hier zugrunde liegenden Anwendungsfall der visuell-manuellen Bedienung von Komfortfunktionen im Fahrzeug ist jedoch die Analyseebene einzelner Interaktionsschritte entscheidend. Gut gestaltete Systeme erlauben Unterbrechungen der Eingabe mit anschließend nahtloser Fortsetzung der Bedienung. Dadurch wird die Aufgabe segmentiert und die Analyse muss die Ebene separater Bedienschritte in den Vordergrund rücken, da die Reduktion der Dauer eines Arbeitsschritts um einige hundert Millisekunden bei der Zweitaufgabe das Time-Sharing und die Ressourcenallokation deutlich zu Gunsten der Hauptaufgabe Fahren beeinflussen kann.

5.2.2.3 Untersuchungen auf der Ebene einzelner Eingaben

Nur wenige Forschungsbeiträge untersuchten die Auswirkungen von Icons und/oder verbalen Markierungen auf einzelne Selektionszeiten. Drei dieser Studien sollen hier Erwähnung finden. Während in dem Experiment von Kacmar und Carey (1991) keine Unterschiede zwischen Icons- und Textvarianten gefunden wurden, konnten in der Untersuchung von Arend und Kollegen (Arend et al., 1987) stabile Effekte zwischen den Gestaltungsvarianten nachgewiesen werden. Aufbauend auf den Resultaten von Arend et al. untersuchten Green und Barnard (Green & Barnard, 1990) die Rolle der Anordnung und visuellen Unterscheidbarkeit von Elementen auf dem Bildschirm und lieferten dadurch wertvolle Anhaltspunkte zur Interpretation der Arend-Studie. Kacmar und Carey (1991) variierten in einem Experiment zur Menüdarstellung ebenfalls die Darstellungsformen Icons allein, Icons plus verbale Beschriftung sowie verbale Beschriftung allein. Die Versuchsteilnehmer mussten zunächst an einer Voruntersuchung teilnehmen, in der ihnen alle Funktionen dargelegt wurden, die von den Icons oder den verbalen Beschriftungen repräsentiert wurden. Nur wenn über 90% der Funktionen einer Versuchsperson geläufig waren, wurde sie für die Hauptuntersuchung zugelassen. Je nach Darstellungsbedingung wurden die Platzhalter einer 3×5 Suchmatrix auf dem Bildschirm mit piktografischen, textuellen oder kombinierten Elementen gefüllt und für jede Aufgabe neu zufällig angeordnet. Die Teilnehmer erhielten eine Beschreibung der gesuchten Funktion und sollten das Element anwählen, das diese Funktion für sie am besten repräsentiert. Im Gegensatz zu den anderen bisher dargelegten Studien führten Kacmar und Carey die Untersuchung auf der Ebene einzelner Selektionszeiten für die Menüelemente aus. Allerdings war auch bei dieser Studie der semantische Aspekt und der Aspekt der visuellen Suche bei der Verwendung von Icons vermengt: Es zeigt sich ein sehr hoher Fehleranteil bei Icons (nur 58% korrekt), während bei Text und Icons plus Text über 80% korrekt ausgewählt wurden. Die Versuchspersonen absolvierten auch nur einen experimentellen Block ohne Probedurchgang, ferner erhielten sie kein Feedback über das Ergebnis ihrer Auswahl. Sie konnten also keine Form von Erfahrung mit dem System aufbauen. Unter diesen Umständen und bei Berücksichtigung der Fehlerzahlen erstaunt es wenig, dass keine Unterschiede in den (stark streuenden) Selektionszeiten zwischen den Gestaltungsvarianten nachgewiesen werden konnte. Des Weiteren wurde den Versuchspersonen in der Instruktion nicht mitgeteilt, dass sie die Suche nach dem richtigen Element und die manuelle Auswahl möglichst schnell durchführen sollen. Kacmar und Carey argumentieren, dass auf diesem Weg die Interaktionssituation für die Teilnehmer natürlicher sei und die Ergebnisse deshalb besser verallgemeinerbar seien. Andererseits wirft dieses Vorgehen die Frage auf, inwiefern der experimentelle Auf-

bau dazu geeignet ist, eventuell vorhandene Unterschiede in den Selektionszeiten aufzudecken.

Ausgangspunkt des Experiments von Arend et al. (1987) war primär die Frage nach dem Einfluss globaler und lokaler Merkmale von Icons auf die visuelle Suche. Zwei verschiedene Icon Sets, abstrakte und konkret-bildhafte Icons, wurden dazu gegeneinander und gegen ein Set bestehend aus verbalen Kennzeichnungen verglichen. Die abstrakten Icons unterschieden sich in globalen Merkmalen, d.h. es wurden für die äußere Form sehr unterschiedliche Gestaltungselemente (Elementarmerkmale wie Form, Größe, Abschluss oder freie Linienenden, etc.) verwendet, um die Unterscheidbarkeit zu erhöhen. Um ein Icon hingegen konkret-bildhaft zu gestalten, müssen lokale Merkmale verwendet werden. Lokale Merkmale (wie Linien und Strukturen innerhalb eines Objekts) erhöhen zwar die Wiedererkennbarkeit, sollten aber die Suche verlangsamen, weil mehr Information verarbeitet werden muss. Im so genannten Search-and-Select Paradigma wurde den Versuchspersonen ähnlich dem Vorgehen in der Studie von Kacmar und Carey zunächst eine Suchaufgabe gestellt. Alle Aufgaben stammten aus dem Umfeld Textverarbeitung, z.B. ein Wort im Text löschen oder eine Zeile verschieben. Auf dem Bildschirm musste dann nach dem Element gesucht werden. Um die Effizienz der Suche abschätzen zu können, gab es zwei verschieden große Suchmengen (6 oder 12 Elemente), die in einer entsprechenden Matrixanordnung dargestellt wurden (2×3 , bzw. 4×3). Je flacher die Steigung, desto effizienter ist die Suche (Wolfe, 1998), desto weniger hängt das Auffinden eines Elements von der Anwesenheit zusätzlicher Elemente ab. Auch in diesem Experiment wurden die Elemente nach jeder Aufgabe neu zufällig angeordnet, um die visuelle Suche zu forcieren. Die Autoren argumentieren, damit den Fall eines ungeübten, neuen Nutzers abzubilden. Die Versuchspersonen waren angewiesen, schnell zu arbeiten, aber die Fehlerrate gering zu halten. Im Gegensatz zu den bisher besprochenen Studien wurde den Versuchspersonen das Itemmaterial im Vorfeld erklärt. Je nach Bedingung bekamen die Teilnehmer das vollständige Set von abstrakten Icons, konkreten Icons oder verbalen Funktionsdarstellungen vorgelegt und erläutert und mussten diese Zuordnungen im Anschluss selbst wiedergeben. Dementsprechend war die Fehlerrate sehr niedrig (zwischen 1.1% und 2.3%) und es ergaben sich diesbezüglich keine Unterschiede zwischen den Funktionsdarstellungen. Hochinteressant sind die gewonnenen Antwortzeiten: Abstrakte Icons konnten deutlich schneller gefunden werden als konkrete Icons und auch als verbale Kennzeichnungen. Der zeitliche Vorteil von abstrakten Icons gegen die anderen Varianten lag bei etwa 800 ms, während sich konkrete Icons und die textuelle Beschriftung nicht unterschieden. Arend et al. schließen, dass konkret-bildhafte Icons durch die Verwendung von lokalen Merkmalen gegenüber verbalen Markierungen nicht im Vorteil sind, außer man würde durch gestalterische Maßnahmen eine verbesserte Unterscheidbarkeit auf globaler Ebene einrichten. Zusätzlich konnte ge-

zeigt werden, dass sich die Erhöhung der Suchmenge auf zwölf Elemente nur bei konkreten Icons und bei Text auswirkt, bei abstrakten Icons konnte kein Effekt nachgewiesen werden. Die Autoren argumentieren, dass abstrakte Icons im Gegensatz zu den anderen Varianten relativ unabhängig von der Suchmenge parallel abgesucht werden können.

Unter den bisher besprochenen Untersuchungen war die Arbeit von Arend und Kollegen die einzige, die einen stabilen Unterschied in der erwarteten Richtung nachweisen konnte. Als einer der Gründe wurde bereits die Verlagerung der Analyse auf einzelne Interaktionsschritte statt auf den Verlauf einer langwierigen Interaktionskette mit den unterschiedlichsten kognitiven und motorischen Prozessen angesprochen. Darüber hinaus ermöglichten sie den Versuchspersonen, das Iconmaterial vorab zu lernen. Einen weiteren Hinweis liefert ein Experiment von Green und Barnard (1990). Auf dem Search-and-Select Paradigma und dem Iconmaterial von Arend und Kollegen aufbauend, interessierten sie sich für die Rolle der Anordnung von Icons: Die Leistung bei visueller Suche hängt nicht nur von den eigenen Merkmalen des Elements ab, sondern auch von Eigenschaften des weiter gefassten Applikationskontexts, d.h. von dem Aufbau der Suchmatrix. Anstatt wie Arend et al. (1987) alle Elemente nach jedem Durchgang zufällig anzuordnen, führten sie für eine Teilmenge Ortskonstanz ein. Einige Icons blieben damit immer an einer festen Position, während die anderen nach jeder Aufgabe an zufälligen Orten innerhalb der Matrix auftauchten. Auch in diesem Experiment erhielten die Versuchspersonen die Gelegenheit, das Iconmaterial zuvor zu sichten. Erst nachdem jedes Item dreimal korrekt erkannt wurde, begann der eigentliche Versuch. Je mehr Icons fest angeordnet waren, desto schneller waren erwartungsgemäß die Suchzeiten. Green und Barnard konnten darüber hinaus aber zeigen, dass der relative Unterschied in den Suchzeiten zwischen den Icontypen mit zunehmender Anzahl fester Icons reliabel abnahm. Interessant ist zusätzlich, dass bereits bei der örtlichen Fixierung der Hälfte der Icons der Unterschied zwischen konkreten und abstrakten Icons fast so klein war wie bei der Fixierung aller Icons. Einen deutlichen Effekt auf die Suchzeiten erhielten sie wie Arend et al. nur bei vollständig zufälliger Anordnung. In den hier dargelegten Studien zu Gesamtsystemen wurden stets vergleichsweise kleine Mengen an Icons verwendet und die Systeme erreichten nicht die Komplexität realer Anwendungen. Das macht die Interpretation wahrscheinlich, dass in diesen Experimenten kaum Aspekte der visuellen Suche, sondern primär Gedächtniseffekte zum Tragen kamen. Bedauerlicherweise untersuchten Green und Barnard nur abstrakte und konkrete Icons, keine verbalen Beschriftungen. Folglich ist nicht vollständig geklärt, ob die gefundene partielle Kompensation von Unterschieden in der Erkennbarkeit von Icons auch auf dieses Material übertragbar ist.

5.2.3 *Ausgangspunkte der Experimente*

Auf der Ebene einzelner Selektionsschritte lässt sich wie im vorigen Abschnitt dargestellt auch in angewandten Untersuchungen durchaus ein Vorteil in den Antwortzeiten für Iconmaterial finden. Zusammengefasst lassen sich drei Gründe ausmachen, warum in vielen Studien keine Unterschiede gefunden wurden. Erstens: Bei dem Test von Gesamtsystemen fand wahrscheinlich aufgrund der festen Anordnung und der geringen Anzahl an Steuerelementen keine visuelle Suche statt. Ein Vorteil von Icons sollte sich aber, von der Wiedererkennbarkeit abgesehen, hauptsächlich bei den Suchzeiten in komplexen Umgebungen zeigen. Zweitens waren die Icons häufig vorher unbekannt, was zu entsprechend hohen Fehlerraten geführt hat. Für das Design von Mensch-Maschine-Schnittstellen muss natürlich obligatorisch berücksichtigt werden, dass Icons unter Umständen vorher gelernt werden müssen – je nach Anwendungsdomäne kann sich das als gravierender Nachteil erweisen und deren Einsatz von vornherein verbieten. Für einen sinnvollen Vergleich der Auffindbarkeit von Text oder Symbol auf einer Benutzerfläche muss jedoch sichergestellt werden, dass das Bildmaterial verständlich ist. Dieser Aspekt kann nicht sinnvoll gleichzeitig mit visueller Suche getestet werden. Drittens wurde die gesamte Aufgabenbearbeitungszeit betrachtet, bei der eventuell vorhandene Unterschiede zwischen bildhafter und verbaler Darstellung durch andere motorische und kognitive Aktivitäten überlagert werden können.

Insbesondere die Experimente von Arend und Kollegen (1987) sowie von Green und Barnard (1990) haben diese Fallstricke vermieden und hilfreiche Ansätze für Gestaltungshinweise geliefert. Dennoch kann die Frage nach der Auswirkung von Icons und verbalen Beschriftungen nicht als geklärt angesehen werden. Forschungsbedarf besteht schon allein aus Gründen der Aktualität. Sehr viele Studien wurden deutlich in den achtziger und frühen neunziger Jahren durchgeführt. Zu diesem Zeitpunkt waren Piktogramme bei weitem nicht so verbreitet wie heute insbesondere durch die Verbreitung von Computerprogrammen, und die angesprochene *Visual Literacy* sollte sich diesbezüglich verändert haben. Piktografische Funktionsdarstellungen und die Arbeit mit grafischen Benutzeroberflächen sind Teil des Alltags geworden. Der Zeichenvorrat an symbolischen Darstellungen zur Dekodierung von Systemausgaben dürfte sich bei den Nutzern entsprechend erweitert haben. Experiment 5 strebt deshalb eine Replikation der Befunde von Arend und Kollegen an.

Darüber hinaus kam sehr spezielles Iconmaterial in diesen Studien zum Einsatz. Zum einen existiert eine derart scharfe Trennung von abstrakten und konkreten Icons in praxisnahen Anwendungsfällen kaum, wie auch schon allein die Existenz der Kategorie der semiabstrakten Icons nahe legt. Zum anderen werden innerhalb einer grafischen Benutzerschnittstelle häufig abstrakte und konkrete Icons gemischt. Es stellt sich also die Frage, inwiefern sich der von Arend et al. gefundene Unterschied für

abstrakte Icons auch mit normalen, für den alltäglichen Einsatz typischen Iconsätzen finden lässt. In vielen dieser Iconsätze findet man – entgegen der Argumentation von Arend und Kollegen – durchaus konkrete Icons, die sich in globalen Merkmalen unterscheiden und so leicht erkennbar sein sollten. Für die hier durchgeführten Experimente wurde deshalb entschieden, dass nur Icons verwendet werden sollen, die sich in vielen Anwendungen finden lassen. Die Icons stammen aus der Norm ISO 7000 (2004) zu grafischen Symbolen für technische Geräte oder sind allgemein bekannte Piktogramme aus Kommunikations- und Unterhaltungssoftware. Diese zum Teil sehr heterogen wirkenden Icons wurden zusätzlich von Designern des Fachzentrums für User Interface Design der Siemens AG überprüft und z.B. bezüglich Linienstärke, Größe und Prägnanz derart für das Experiment angeglichen, dass sie ein Mindestmaß an visueller Stimmigkeit aufweisen, wie es für reale Softwareprodukte vorausgesetzt wird.

Auch die verbalen Markierungen in der Arend-Studie bedürfen einer Überarbeitung. Sie bestanden stets aus zwei Begriffen und waren entsprechend zweizeilig auf der Schaltfläche angelegt. Zusätzlich war bei einigen Elementen bereits beim ersten Begriff die assoziierte Funktion klar (z.B. „Text speichern“, „Rand ausgleichen“). Bei sehr vielen Items erfolgte die Disambiguierung aber erst in der zweiten Zeile (z.B. „Wort ersetzen“, „Wort suchen“ oder „Wort löschen“). Der Vergleich zwischen Icons und verbalen Markierungen scheint so etwas unfair, da bei Textdarbietung eine zusätzliche Schwierigkeit entstand. Auf diese Weise ist die Suche für Text willkürlich erschwert. Wenngleich als inhärenter Vorteil von Icons gilt, Sachverhalte in einem Bild auszudrücken, würde man doch unter Usability-Gesichtspunkten beim so genannten Wording der Funktionen immer versuchen, einen klaren und trotzdem kurzen Ausdruck zu finden. Das gilt in besonderem Maße, wenn diese verbalen Kennzeichnungen für Sprachbefehle im SWYS Kontext stehen sollen. Selbstverständlich lässt sich nicht jede Funktion immer nur mit einem Wort angeben, dennoch wird bei der Systemgestaltung angestrebt, wenige Mehrwort-Konstruktionen einzubinden. Deshalb sollen in den Experimenten nur einzeilige verbale Markierungen bestehend aus jeweils nur einem Wort zum Einsatz kommen.

Ebenfalls unklar ist die Rolle strukturierender Faktoren wie der ortskonstanten Anordnung von Bedienelementen. In der Studie von Green und Barnard (1990) wurden nur unterschiedliche Arten von Icons analysiert und eine kompensierende Wirkung räumlich konstanter Anordnung bei visuellen Nachteilen für die Suche festgestellt. Unterschiede in der Art der Informationsverarbeitung zwischen bildhaftem und verbalem Material lassen es nicht ohne weiteres zu, dieses Ergebnis auf Beschriftungen auszuweiten. Durchaus wahrscheinlich ist, dass grundsätzlich unterschiedliche Suchstrategien bei ortsfesten Icons oder verbalen Elementen zu einem anderen Resultat führen können. Darüber hinaus kann auf Basis der Studie von Green und Barnard kein guter Rückschluss auf die Auswirkung unterschiedlich großer Suchmengen gezogen

werden, wie sie bei einer Beurteilung von Such- und Gedächtnisparametern sinnvoll ist. Dieser Fragestellung soll in Experiment 6 nachgegangen werden.

Im Zusammenhang mit strukturierenden Faktoren stellt sich die grundsätzliche Frage, inwiefern sich ein möglicher Unterschied zwischen Icons und verbalen Marken erhält, wenn über Ortskonstanz hinaus zusätzliche Designprinzipien zum Tragen kommen. Bei Green und Barnard (1990) waren die Elemente auf dem Bildschirm vergleichsweise sinnlos angeordnet. Trotz Ortskonstanz ergaben sich immer noch Anordnungen, die übliche Gestaltungsprinzipien missachten. So waren viele Funktionen die semantisch zueinander gehört hätten (z.B. „Bild vor“, „Bild zurück“), durch die Zufallsanordnung beliebig in der Matrix verteilt. Das schränkt die ökologische Validität des Vergleichs drastisch ein. Im Rahmen eines Usability Engineering Prozesses würde immer ein Mindestmaß an innerer Logik des Aufbaus verfolgt. Es soll folglich ein Experiment durchgeführt werden, bei dem neben der Ortskonstanz inhaltlich zusammengehörige Elemente gruppiert werden (Experiment 7).

Bei näherer Betrachtung der hier dargelegten Literatur ist zu erkennen, dass zur Diskussion der Rolle von Icons und verbalen Beschriftungen im Hinblick auf kognitive Prozesse zwei Faktoren diskutiert werden müssen: Gedächtnispsychologische und wahrnehmungspsychologische Aspekte. Diese lassen sich zum Teil getrennt analysieren. Zu den gedächtnispsychologischen Fragestellungen lassen sich die Wiedererkennbarkeit von visuellem Material, das semantische Verständnis von Symbolen oder das Erlernen des Layouts einer dargestellten Matrix aus Funktionsanzeigen zählen. Eine wahrnehmungspsychologische Ausrichtung findet sich bei allen Untersuchungen, die ausschließlich auf die Auffindbarkeit eines visuellen Stimulus fokussieren. Im Anwendungsfall Bildschirmgestaltung sind diese zwei Faktoren verwoben und beeinflussen die Mensch-Maschine-Interaktion in unterschiedlicher Weise. Um ein tieferes Verständnis dieses Zusammenspiels zu entwickeln, soll eine Serie von insgesamt drei Experimenten mit demselben Stimulusmaterial und gleicher Durchführungsprozedur den Übergang von rein visueller Suche bei Icons und textuellen Elementen hin zu einer stark gedächtnisgestützten Suche vollziehen.

Basierend auf den dargelegten Anforderungen wurden diese Experimente folgendermaßen konstruiert: Im ersten Experiment werden alle Elemente zufällig angeordnet, was maximale visuelle Suche erforderlich macht (Experiment 5). Im zweiten Experiment wird die visuelle Suche eingeschränkt, indem die Darbietung der Elemente ortskonstant erfolgt und somit Gedächtnisprozesse zum Ergebnis beitragen (Experiment 6). Im dritten Experiment wird die visuelle Suche weiter zurückgenommen, indem alle Items ortskonstant und zusätzlich inhaltlich gruppiert dargeboten werden (Experiment 7). Dieses Vorgehen ermöglicht, direkte Vergleiche zwischen den Experimenten und damit den Einflussfaktoren zu ziehen, was bei den existierenden Studien durch Unterschiede in Vorgehen und Material nur sehr schwierig möglich ist. Alle

Experimente nutzten das Search-and-Select Paradigma, das feinkörnige Analysen der Reaktionszeiten auf der Ebene einzelner Eingaben zulässt. Der Aufbau orientiert sich dabei grundsätzlich an den Studien von Arend et al. (1987) und Green und Barnard (1990), das Iconmaterial wurde allerdings wie beschrieben angepasst und die verbalen Kennzeichnungen auf ein Wort reduziert. Das Lernen der Semantik von Icons wird ausgeklammert, da dies die Interpretation der Suchzeiten einseitig beeinflussen würde. Im Vorfeld eines Experiments wird die Bedeutung der Icons entsprechend kurz dargelegt.

Arend et al. (1987) haben das Argument vorgebracht, dass die visuelle Suche in einer zufälligen Anordnung das Verhalten eines völlig unerfahrenen Nutzers simuliert. In diesem Sinne simuliert die hier geplante Serie verschiedene Stadien der Nutzererfahrung mit einem System: Die Teilnehmer in Experiment 5 haben ähnlich zu Arend et al. nur das Wissen über die Existenz von Funktionen, wie es neue Nutzer nach einer kurzen Einführung besitzen. Sie müssen jedoch auf dem ihnen noch unbekanntem Bildschirm danach suchen. Dieser Fall tritt zum Beispiel bei der Nutzung von Mietwagen ein: Die Icons sind weitgehend bekannt, der spezifische Aufbau der Bedienoberfläche jedoch nicht. Experiment 6 kann man mit Einschränkungen als intermediären Zustand betrachten, bei dem sich zwar schon einige Aspekte der grafischen Oberfläche eingeprägt haben, aber noch kein funktionaler Überblick besteht. In Experiment 7 hingegen hat sich Wissen über inhaltliche Zusammenhänge herausgebildet, die die Suche nach verwandten Funktionen erleichtern. Der untersuchte Gedächtnisfaktor ist dabei die Wiedererkennbarkeit visueller Muster, der am Lernfortschritt über experimentelle Durchgänge hinweg ab Experiment 6 thematisiert wird.

Experiment 8 fokussiert auf einen bisher vernachlässigten Aspekt im Vergleich von Icons und verbalen Kennzeichnungen, der im Hinblick auf eine möglichst kurze visuelle Interaktionsdauer hochgradig relevant ist. In den Experimenten 5 bis 7 hatten die Versuchspersonen unbegrenzt Zeit, das Display zu betrachten und ein Element auszuwählen. Bei der Fahrzeugführung gelten Blickabwendungen von der Fahrbahn bis maximal 1200 Millisekunden als akzeptabel (Zwahlen et al., 1988). Berücksichtigt man die benötigte Zeit zur motorischen Veränderung des Blickortes, Orientierung in der neuen visuellen Szene mit Fokussierung und Akkommodation sowohl zum Display als auch zurück zur Straße, kann deutlich weniger als eine Sekunde zur Verfügung stehen. In Experiment 8 wird deshalb systematisch die Zeit variiert, die zur Erkennung der Elemente auf dem Display zur Verfügung steht. In einer Art Worst-Case-Szenario wird dabei von visueller Suche ausgegangen (simuliert durch zufällige Anordnung der Elemente) und die Frage beantwortet, mit welcher Güte bei sehr kurzen Darbietungsdauern die Position eines Elements identifiziert werden kann. Die Identifikation der Position ist in dem gegebenen Anwendungskontext entscheidend, damit z.B. eine

zugeordnete frei belegbare Taste (Softkey) ohne Blickzuwendung erfüllt und betätigt werden kann.

5.2.4 Experiment 5: Zufällige Itemanordnung

Experiment 5 zielt auf eine Replikation der Ergebnisse von Arend et al. (1987) mit abgewandeltem Stimulusmaterial ab. In dem Experiment fand ein Vergleich von Icons und verbalen Kennzeichnungen im Search-and-Select Paradigma mit zufälliger Anordnung der Elemente nach jedem Durchgang statt. Zwei verschieden große Suchmengen mit sechs bzw. zwölf Elementen kamen zum Einsatz. Die experimentelle Fragestellung lässt sich folgendermaßen umreißen: Wie unterscheiden sich Icons und verbale Beschriftungen in verschieden großen Suchmengen, wenn die Suche forciert ist?

Die Ähnlichkeit im Aufbau dient dazu, einen direkten Vergleich zu der Studie von Arend und Kollegen herstellen zu können. Anpassungen der Versuchsprozedur und des Search-and-Select Paradigmas werden im folgenden Abschnitt (5.2.4.1) ausführlich dargelegt. Die Icons auf den Schaltflächen waren nicht artifiziell, sondern bekannte Standardsymbole zur Gestaltung der visuellen Schnittstelle. Die alternativ zu den Icons verwendeten verbalen Funktionskennzeichnungen bestanden aus Gründen der ökologischen Validität immer nur aus einem Wort. Die Zweiwort-Konstruktionen von Arend und Kollegen wurden vor dem bereits genannten Hintergrund abgelehnt, dass bei der Gestaltung von textbasierten Oberflächen stets möglichst einfache Kommandowörter verwendet werden. Das gilt insbesondere, wenn die Kommandowörter auch zur Spracheingabe benutzt werden sollen.

5.2.4.1 Methode

Search-and-Select Paradigma

Als Grundlage für dieses sowie für die folgenden Experimente diente das Search-and-Select Paradigma (Arend et al., 1987). Es gestattet die realitätsnahe Durchführung von Aufgaben, wie sie in Menüauswahl-Situationen vorkommen, bietet aber größere Kontrolle über relevante Umgebungsvariablen. Dadurch können die Reaktions- und visuellen Suchzeiten für Nutzereingaben auf der Ebene einzelner Aktionen erhoben werden. Das ermöglicht eine vergleichende Einschätzung des Ablenkungspotentials verschiedener Darstellungsvarianten. Ferner ist eine Überprüfung des Itemmaterials möglich: Durch die Analyse der mittleren Antwortzeiten für jedes einzelne Item können besonders gute oder schlechte Gestaltungsvarianten erkannt werden.

Der Standardablauf sieht vor, dass in jedem Durchgang der Versuchsperson zuerst die Aufgabe präsentiert und danach der Menübildschirm, eine Matrixanordnung von Schaltflächen, angezeigt wird. Die Versuchsteilnehmer müssen dann diejenige Schalt-

fläche auswählen, die eine vorgegebene Funktion repräsentiert. Die Funktion auf dem Steuerelement kann entweder durch Icons oder verbale Kennzeichnungen signalisiert werden. Das Target ist dabei in allen Durchgängen präsent, Catchtrials sind nicht vorgesehen. Bei Arend und Kollegen erhielten die Versuchspersonen die Anweisungen zur aktuellen Aufgabe in Schriftform auf dem Bildschirm. Um jede Form von visuellem Priming zu vermeiden, wurde im vorliegenden Experiment auditiven Anweisungen über Kopfhörer der Vorzug gegeben.

Wie in typischen Grundlagenexperimenten zu visueller Suche variiert die Größe der Suchmenge, damit eine Aussage über die entstehende Steigung und damit über die Natur des Suchprozesses möglich wird. Sternberg (1969) beschreibt im Detail die Überlegungen hinter einer Interpretation der Geradengleichung in Scanning-Experimenten. Der y-Achsenabschnitt gibt die Gesamtdauer aller Prozesse an, die nur einmal unabhängig von der Suchmenge stattfinden. Für die Charakterisierung der Suche ist die Steigung primär relevant, da sie die Dauer der Verarbeitungsprozesse misst, die pro Item stattfinden. Darunter fallen z.B. Vergleichsoperationen oder ggf. die Zeit, um von einem auf ein anderes Item zu schwenken. Flache Steigungen zeigen, dass der Prozess relativ unabhängig von der Anzahl der Elemente in der Suchmenge ist. Dies wird üblicherweise als Hinweis auf einen parallelen Suchprozess gesehen. Steile Steigungen hingegen deuten danach auf serielle Suche hin. Wolfe (1998) führt aus, dass die theoretischen Annahmen hinter einer strikten Trennung in serielle und parallele Prozesse problematisch sind und schlägt vor, die Suche auf einem Kontinuum der Effizienz zu bewerten. Sticht ein Target aus den anderen Elementen heraus und kann so unabhängig von deren Anzahl schnell gefunden werden, ist die Suche effizient. Nimmt mit zunehmender Anzahl der Non-Targets die Suchzeit pro Item zu, muss sie als entsprechend weniger effizient eingestuft werden. Im Bereich der Grundlagenforschung existieren Schwellenwerte für die entsprechenden Bezeichnungen. Bei einer Steigung von wenigen Millisekunden pro Item ist die Suche *effizient*, ab 10 ms/Item wird sie noch als *ziemlich effizient* beschrieben, während sie bei mehr als 25 ms/Item *ineffizient* ist (Wolfe, 1998). Wegen des andersartigen Stimulusmaterials in angewandten Untersuchungen ist es nicht sinnvoll, diese Grenzwerte direkt zu übertragen. Für den Vergleich in der vorliegenden Studie ist darüber hinaus primär die relative Effizienz von Interesse.

Versuchspersonen

An dem Experiment nahmen zwölf Versuchspersonen (sechs Frauen, sechs Männer) im Alter zwischen 22 und 36 Jahren (Mittelwert: 26 Jahre) teil, die für ihre Teilnahme eine angemessene Aufwandsentschädigung erhielten. Alle verfügten über normale oder voll korrigierte Sehschärfe und sprachen Deutsch als Muttersprache, um möglichen Variationen in der Antwort auf verbales Material entgegenzuwirken. Die Ver-

suchspersonen nutzten aus beruflichen Gründen oder für das Studium regelmäßig bis häufig Computer. Sie waren nicht über die Hypothesen des Experiments informiert.

Material und Stimuli

Zwei Itemsätze mit je zwölf korrespondierenden Elementen kamen in dem Experiment zum Einsatz. Ein Satz bestand aus Icons, der andere aus verbalen Beschriftungen. Aufgrund des anvisierten Anwendungsfalls Fahrzeugführung repräsentierten die zwölf Items typische Komfortfunktionen, wie sie z.B. im Bordsystem eines PKW vorkommen können. Jeweils drei Items waren einem von vier Funktionsbereichen zugeordnet. Sie sind nachfolgend mit der als Beschriftung verwendeten Verbalisierung aufgezählt:

- Musikfunktionen: Start, Stop, Lautstärke
- Kommunikationsfunktionen: Adressbuch, Telefon, SMS
- Klimafunktionen: Klimaanlage, Sitzheizung, Lüftung
- Systemfunktionen: Einstellungen, Zurück, Hilfe

Die Icons waren Standardsymbole (ISO 7000, 2004) oder Symbole, deren Bekanntheit aus dem Kommunikations- und Unterhaltungskontext weitgehend vorausgesetzt werden kann. Sowohl abstrakte, semiabstrakte, als auch konkrete Icons wurden verwendet, um die häufig anzutreffende Durchmischung bei grafischen Benutzeroberflächen nachzubilden.

Icons und Textelemente sind in der Reihenfolge der obigen Aufzählung in Abbildung 34 dargestellt. Im Experiment sind die Elemente zufällig angeordnet. Die Abbildung zeigt die Benutzeroberfläche, mit der die Versuchsteilnehmer arbeiteten, für zwölf Elemente. Wenn nur sechs Elemente angezeigt wurden, waren auch die entsprechenden Schaltflächen (die weißen Platzhalter) ausgeblendet. Die Ausgabe dieser grafischen Oberfläche erfolgte auf einem 8 Zoll TFT-LCD Touchscreen der Firma CarTFT mit einer Auflösung von 640×480 Bildpunkten (Breite \times Höhe). Jedes Steuerelement hatte eine Größe von 130×80 Bildpunkten. Bei der gewählten Monitorgröße und Auflösung ergab sich eine Abmessung von 3.5×2.2 cm pro Element. Das Format der Beschriftung war Arial, 12 Punkt, Fettdruck. Das Experiment wurde in einem ruhigen Raum durchgeführt, die Beleuchtungsverhältnisse unterstützten das blendfreie Ablesen der Bildschirminhalte. Die jeweilige Search-and-Select Aufgaben wurde den Versuchspersonen auditiv über dynamische Sennheiser HD 25-1 Studio-Kopfhörer angekündigt, die Umgebungsgeräusche sehr gut abdämpfen (geschlossene Bauform, ohraufliegende Ankopplung mit Ringpolstern, Andruckkraft ca. 2.5 N).

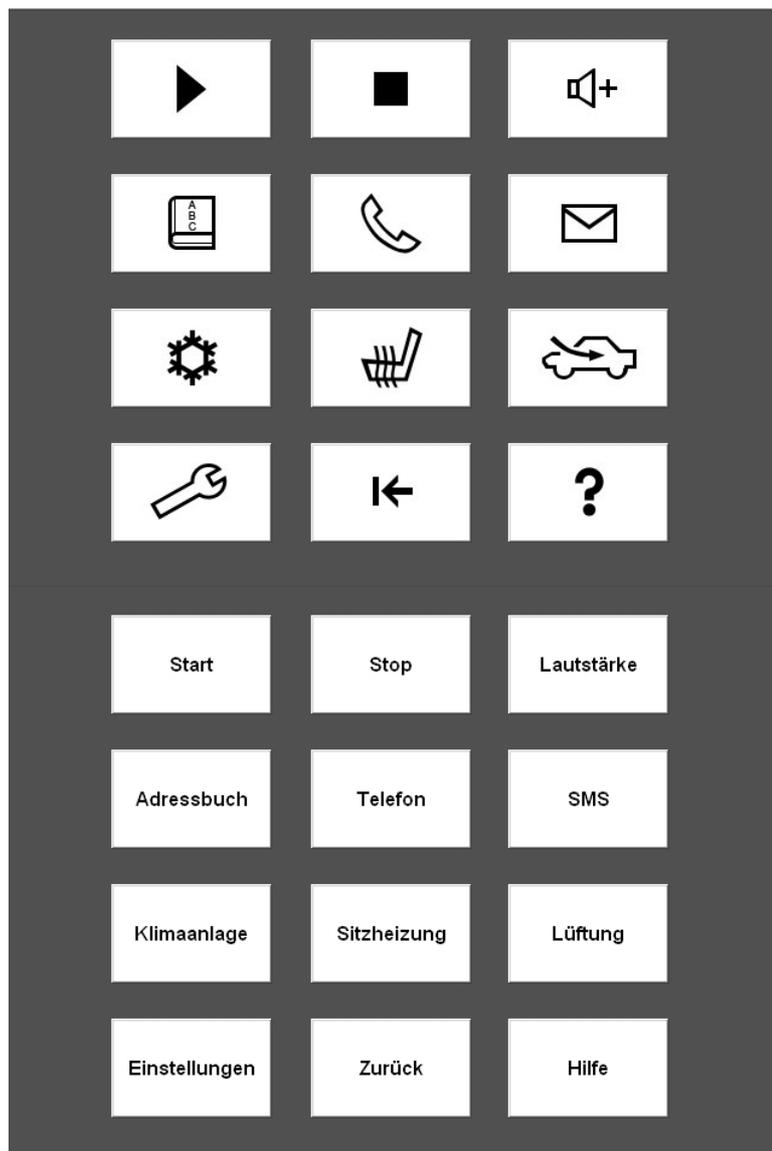


Abbildung 34: Ikonografische und verbale Funktionsdarstellungen, die Stimuli des Experiments, als 12-elementige Suchmenge.

Das Programm, das die Ausgabe der Anweisungen und der Stimuli kontrollierte, lief auf einem Fujitsu Siemens Laptop mit 1.73 GHz CPU, Intel® 915GM Chipsatz und Realtek ALC 260 Soundkarte. Die Testsoftware wurde vom Autor mit Microsoft® Visual Basic® und Windows® API Funktionen programmiert. Der Versuchsleiter konnte den teilautomatisierten Ablauf der Untersuchung über eine hierfür erstellte Benutzeroberfläche steuern, die auf dem Bildschirm des Laptops angezeigt wurde. Die Anzeige auf dem 8 Zoll Touchscreen der Versuchsperson stellte ausschließlich Inhalte des Experiments dar und zeigte in Zwischenpausen einen dunklen Hintergrund. Die Blickdistanz wurde anders als in den meisten Experimenten zu visueller Suche nicht mit einer Kopfstütze konstant gehalten (siehe Abbildung 35).



Abbildung 35: Versuchsteilnehmer beim Ausführen einer Aufgabe.

Die Versuchspersonen waren angehalten, sich so zu positionieren, dass sie bequem die 2×3 , bzw. 4×3 Suchmatrix überblicken und die Beschriftungen ohne Schwierigkeiten lesen konnten. Durch den Verzicht auf eine Fixierung von Blickwinkel und -abstand wird auf einen Aspekt der experimentellen Kontrolle verzichtet. Das kann zu einer erhöhten Streuung in den Daten und so zu einer Überlagerung der Effekte der Darstellungsform verbal versus bildlich führen. Im realen Anwendungsfall ist jedoch davon auszugehen, dass auch hier die Blickdistanz zwischen und innerhalb von Personen merklich variiert. Könnte ein Effekt nur bei Konstanthalten dieses Parameters gezeigt werden, hätte er für praktische Belange der Systemgestaltung keine Bedeutung.

Manuelle Antworten, d.h. die Selektion des Items auf dem Bildschirm, erfolgten durch direkte Anwahl auf dem Touchscreen. Die Versuchspersonen wurden instruiert, mit dem Zeigefinger ihrer dominanten Hand das Element anzutippen. Als Suchzeit wurde der Zeitraum zwischen dem Erscheinen der Matrix und dem Zeitpunkt des Auftreffens des Fingers auf der Schaltfläche festgehalten.

Versuchsdesign

Ein $2 \times 2 \times 3$ faktorielles Within-Subjects Design wurde genutzt. Jede Versuchsperson absolvierte mehrere experimentelle Blöcke (Stufen: 1., 2., 3. Block) mit unterschiedlich großer Suchmenge (Stufen: 6 oder 12 Elemente) und unterschiedlichem Präsentationsformat (Stufen: Icons oder Text). Betrug die Größe der Suchmenge sechs Elemente, wurde das Gesamtset der zwölf Elemente in zwei 6-elementige Subsets A und B aufgeteilt, bestehend aus je zwei vollständigen Funktionsbereichen (Musik, Kommunikation, Klima und System). Die Anordnung der Items in den Subsets war wie bei der 12-elementigen Variante permutiert.

In jedem experimentellen Block wurde jede Aufgabe einmal gestellt, d.h. jedes Item musste einmal gefunden werden. Damit bestand bei der 12-elementigen Menge ein Block aus 12 Aufgaben, bei der 6-elementigen Menge entsprechend aus 6 Durchgängen. Zur Vermeidung von Gruppierungseffekten war die Reihenfolge der Aufgaben zufällig. Um jedes Item gleich häufig im Experiment testen zu können, ergab sich folgende Aufteilung: Bei 12-elementiger Menge wurden 3 Blöcke mit je 12 Aufgaben absolviert, bei 6-elementigen Mengen insgesamt 6 Blöcke mit je 6 Aufgaben. Die Blöcke mit 6-elementigen Sets folgten stets aufeinander. Während der ersten drei Blöcke kam ausschließlich eines der beiden Subsets zum Einsatz. Erst wenn alle Blöcke in einer Darstellungsvariante absolviert waren, fand diesbezüglich ein Wechsel statt. Dabei waren die Abfolge der vollständigen Sets und der aufgeteilten Sets sowie die Funktionsdarstellung zwischen den Versuchspersonen ausbalanciert.

Als abhängige Variable ist in dem gewählten Paradigma primär die Reaktionszeit von Interesse. Zusätzlich wurde zur Kontrolle die Fehlerrate erhoben.

Prozedur

Vor dem Experiment wurde den Versuchsteilnehmern zunächst die Aufgabe erläutert. Sie wurden darüber informiert, dass sie nach auditiver Anweisung ein bestimmtes Element auf dem Touchscreen auswählen müssen. Wie oben dargelegt, musste die Vertrautheit der Probanden mit dem Iconmaterial sichergestellt werden. Dazu wurde ihnen ein Ausdruck vorgelegt, auf dem die Elemente aus Abbildung 34 zu sehen waren. Der Versuchsleiter las jede der zwölf Aufgaben vor und verwies auf die zugehörige textuelle oder bildliche Funktionsdarstellung. Eine Auflistung aller zwölf Aufgaben findet sich in Anhang C 1. Anschließend wiederholten die Probanden die Funktion der Icons mit eigenen Worten und mögliche Unklarheiten konnten geklärt werden.

Jede Aufgabe begann mit der auditiven Anweisung. Die Versuchspersonen sollten versuchen sich während dieser Ansage vorzustellen, wie das zugehörige textuelle oder bildliche Element auf dem Bildschirm aussieht. Erst wenn sie die Aufgabe klar verstanden hatten, sollten sie die mittig auf dem Touchscreen sichtbare „OK“ Schaltfläche betätigen. Haben sie die Anweisung akustisch nicht verstanden, konnten sie die darunter befindliche Schaltfläche „Anweisung wiederholen“ drücken. Bei inhaltlichen Problemen sollten sie den Versuchsleiter fragen, was jedoch während des Experiments nicht vorkam. Durch dieses Vorgehen und die zufällige Anordnung war sichergestellt, dass ausschließlich Aspekte der visuellen Suche zum Tragen kamen und weder semantische noch episodische Gedächtnisleistungen interferierten.

Nach dem Drücken der Bestätigungstaste durch die Versuchsperson erschien für 1.5 Sekunden an zentraler Position ein Fixationskreuz. Die Probanden waren angehalten, dieses Kreuz anzublicken und von diesem Mittelpunkt die visuelle Suche zu starten. Unmittelbar im Anschluss erschien die Suchmatrix. Die Teilnehmer sollten schnellst-

möglich die korrekte virtuelle Schaltfläche betätigen, dabei aber die Fehlerrate gering halten. Bei Berührung des Steuerelements wurde unmittelbar durch Veränderung der Farbe Rückmeldung über die Eingabe gegeben. Im Fall einer falschen Eingabe erschien anschließend für zwei Sekunden das Zielitem allein zentral auf dem Bildschirm. Bei einer richtigen Eingabe erschien sofort erneut das Fixationskreuz. Die nächste Aufgabe folgte nach einer Pause von 1.5 Sekunden.

Vor dem ersten experimentellen Block wurden zwei Aufgaben zufällig zur Übung herausgegriffen. Letzte Unklarheiten wurden beseitigt und die Teilnehmer ermutigt, sich mit dem Touchscreen vertraut zu machen. Nach jedem Block wurde eine Pause von 30 Sekunden systemgesteuert eingehalten. Beim Wechsel von Icons zu textueller Darstellung, also nach der Hälfte des Experiments, pausierten die Probanden fünf Minuten. Danach wurde der Wechsel der Darstellungsform angekündigt und zur Orientierung noch einmal der Papierausdruck mit den abgebildeten Varianten vorgelegt. Insgesamt dauerte dieses Experiment ca. 35 Minuten.

5.2.4.2 Hypothesen

Basierend auf der Studie von Arend et al. (1987) und den dargelegten Befunden zur visuellen Suche wird erwartet, dass sich ein Unterschied zwischen Icons und verbalen Funktionsbeschriftungen zeigt. Dieser sollte darüber hinaus abhängig von der Größe der Suchmenge ausfallen: Je schwieriger die Suchaufgabe wird, desto stärker sollte sich der Vorteil für bildhaftes Material abzeichnen. Selbst bei ikonografischem Material ist jedoch eine effiziente Suche im Sinne Wolfes (1998) unwahrscheinlich, da jedes Element Target oder Non-Target in kurzer Folge sein kann. Entsprechend sollte sich für beide Darstellungsvarianten ein Effekt für den Faktor Größe der Suchmenge nachweisen lassen. Aufgrund der zufälligen Anordnung der Elemente sind zudem keine bedeutsamen Lerneffekte über die Zeit hinweg zu erwarten.

5.2.4.3 Ergebnisse

In einem ersten Analyseschritt werden Reaktionszeiten und Fehlerraten über alle Aufgaben und Blöcke hinweg für beide Darstellungsformen gemittelt betrachtet. Ein zweiter Schritt soll Aufschluss über mögliche Lerneffekte geben. Schließlich werden die einzelnen Items einer differenziellen Analyse unterzogen. Bei diesem und allen folgenden Experimenten werden die Ergebnisse inferenzstatistisch mit Varianzanalysen auf Unterschiede geprüft. Wenn nicht anders angegeben, werden im Text stets Mittelwerte (M) und Standardfehler (SE) berichtet. In Grafiken weisen Fehlerbalken den Standardfehler aus. Zusätzlich werden die Replikationswahrscheinlichkeit p_{rep} (Killeen, 2005a, 2005b) und bei signifikanten Unterschieden als Maß der Effektgröße das partielle Eta-Quadrat η_p^2 berichtet. Bei Gruppenvergleichen wurde der Levene-Test

zur Überprüfung auf Varianzhomogenität in die Analyse einbezogen. Die Ergebnisse des Tests werden nur im Fall heterogener Varianzen berichtet.

Reaktionszeiten und Fehler

Das Zusammenwirken der Suchmengengröße und der Darstellungsform steht bei dieser Analyse im Vordergrund. Gemessen wurde die Reaktionszeit ab Sichtbarkeit der Suchmatrix bis zum Auftreffen des Fingers auf der Schaltfläche des Touchscreen. Jeder Mittelwert basiert auf 432 Beobachtungen. Für die Kombinationen von Suchmenge und Darstellung sind die mittleren Antwortzeiten in Abbildung 36 aufgetragen. Wurden die Funktionen durch Text dargestellt, benötigten die Teilnehmer für 6 Items 991 ms ($SE = 37.10$) bzw. 1503 ms ($SE = 56.18$) für 12 Items. Bei einer Darstellung durch Icons wurden bei 6 Items 873 ms ($SE = 55.64$) und bei 12 Items 1205 ms ($SE = 61.51$) gemessen. Durchgänge, in denen ein Fehler gemacht wurde, gingen nicht in die Analyse mit ein. Die Auswirkungen der Within-Faktoren Darstellungsformat (Text, Icons) und Suchmenge (6 oder 12 Elemente) auf die Reaktionszeit wurden mit einer zweifaktoriellen Varianzanalyse analysiert. Sowohl Haupteffekte als auch die Interaktion Darstellungsformat \times Suchmenge erwiesen sich als signifikant.

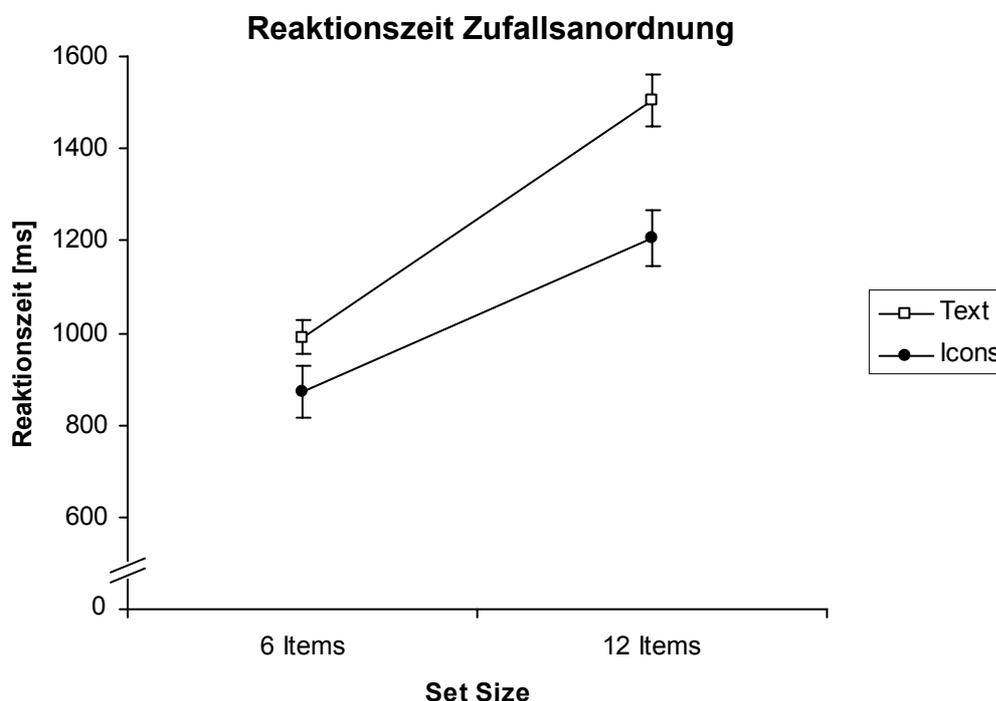


Abbildung 36: Mittlere Reaktionszeiten von Text und Icons als Funktion der Größe der Suchmenge bei zufälliger Anordnung der Items.

Es zeigt sich, dass verbale Beschriftungen signifikant langsamer gefunden werden als Icons. Während bei Icons im Schnitt 1039 ms ($SE = 53.32$) benötigt werden, muss bei Text 1247 ms ($SE = 62.67$) gesucht werden ($F(1,11) = 22.84$, $p = .001$, $p_{rep} = .97$, $\eta_p^2 = .68$). Eine Erhöhung der Itemanzahl führt zu deutlich gesteigerten Suchzeiten.

Der Anstieg von 932 ms ($SE = 34.93$) auf 1354 ms ($SE = 51.25$) erwies sich als signifikant ($F(1,11) = 95.07, p < .001, p_{rep} > .99, \eta_p^2 = .90$). Eine Betrachtung der Steigung gibt Aufschluss über die Effizienz der Suche für die verschiedenen Präsentationsformate. Bei Icons ergibt sich eine Steigung von 55 ms/Item, bei verbalen Beschriftungen hingegen 85 ms/Item. Die Steigungen unterscheiden sich also deutlich, was durch die signifikante Interaktion Präsentationsformat \times Suchmenge bestätigt wird ($F(1,11) = 5.35, p = .04, p_{rep} = .89, \eta_p^2 = .33$).

Fehler waren insgesamt äußerst selten. Maximal in einem Prozent der Fälle wurde die falsche Schaltfläche gewählt. Bei Icons wurden 1% (6 Items) bzw. 0.5% (12 Items) der Antworten falsch gegeben. Beschriftungen wurden in 1% (6 Items) bzw. 0.2% (12 Items) falsch angeklickt. Aufgrund der geringen Häufigkeit – über alle Probanden summiert ist die Anzahl der Fälle pro Zelle nie größer als vier – werden diese Daten keiner weiteren Analyse zugeführt.

Lerneffekte

Um zu testen, inwiefern die Reaktionszeiten sich mit fortschreitender Übung verändern, wurden die Antwortzeiten in jedem der drei experimentellen Blöcke für Icons und Text gemittelt. Da bei der größeren Suchmenge mit 12 Elementen Lernfortschritte besser zu erkennen sein sollten, beschränkt sich die Analyse auf diese Durchgänge. Durchgänge mit 6 Elementen gingen nicht mit ein.

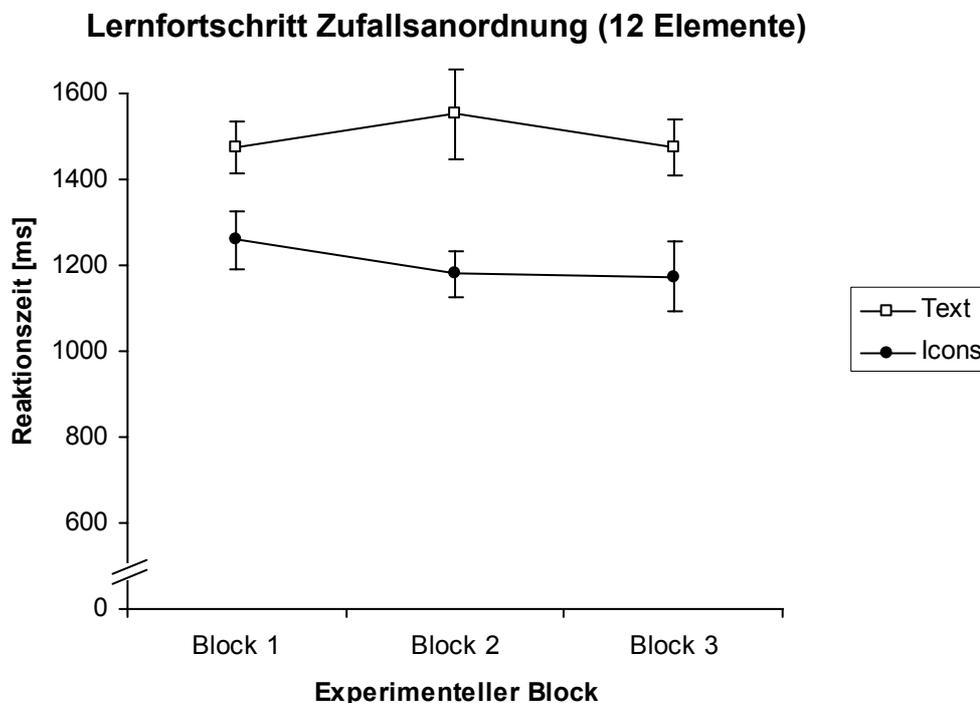


Abbildung 37: Lernfortschritt als Veränderung der mittleren Reaktionszeiten bei 12-elementigen Suchmengen für die beiden Darstellungsformate über die Zeit.

Wie Abbildung 37 zu entnehmen ist, zeigt sich weder bei Text, noch bei Icons eine sichtbare Abnahme der mittleren Reaktionszeit über den Verlauf des Experiments hinweg. Die mittleren Reaktionszeiten (Standardfehler in Klammern) von Block 1 nach Block 3 betragen bei Text 1478 ms (60.48), 1554 ms (106.15) und 1477 ms (65.79), bei Icons 1260 ms (66.93), 1181 ms (53.27) und 1175 ms (81.90).

Diese Daten wurden einer zweifaktoriellen Varianzanalyse zugeführt. Weniger von Interesse ist der signifikante Haupteffekt für das Präsentationsformat (Text: $M = 1504$ ms, $SE = 45.32$, Icons: $M = 1205$ ms, $SE = 38.88$), da dieser aufgrund der obigen Betrachtung der Reaktionszeit bei 6- und 12-elementigen Mengen bereits vorweggenommen ist ($F(1,11) = 25.66$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .70$). Die Reaktionszeiten veränderten sich kaum während des Experiments; von Block 1 nach 3 (Standardfehler in Klammern): 1369 ms (49.66), 1368 ms (69.89), 1326 ms (60.32). Der Haupteffekt für Block erwies sich als nicht signifikant ($F(2,22) = 0.47$, $p = .63$, $p_{\text{rep}} = .41$). Ebenso ist die Interaktion Block \times Präsentationsformat nicht signifikant ($F(2,22) = 0.81$, $p = .46$, $p_{\text{rep}} = .53$).

Itemanalyse

Aufgrund der unterschiedlichen visuellen Beschaffenheit der Stimuli ist zu erwarten, dass auch innerhalb der Gruppe der Icons und textuellen Beschriftungen eine beträchtliche Streuung in den Suchzeiten existiert. Die Itemanalyse verfolgt zwei Ziele. Erstens soll durch eine Rangordnung der Stimuli nach Suchzeiten untersucht werden, welche Elemente besonders gut und welche besonders schlecht gefunden wurden. Zweitens soll der Zusammenhang zwischen den Rangreihen der Icons und der textuellen Elemente in den Fokus rücken. Bei der Planung des Experiments wurde besonderer Wert darauf gelegt, dass die Suchzeiten nur durch visuelle, nicht durch konzeptuelle Eigenschaften der Stimuli bestimmt sind. Sollten konzeptuelle Einflüsse derart im Experiment zum Tragen gekommen sein, dass die Reaktionszeit bei manchen Items unabhängig vom Präsentationsformat besonders gut oder schlecht war, sollte sich ein Zusammenhang zwischen den Rangreihen nachweisen lassen. Eine geordnete Auflistung der Mittelwerte für alle Items lässt sich Tabelle 10 entnehmen.

Zunächst wurden zwei einfaktorielle Varianzanalysen mit den mittleren Reaktionszeiten der Items innerhalb des jeweiligen Sets berechnet. Die Reaktionszeiten wurden dazu über alle Blöcke und über die Größen der Suchmengen hinweg gemittelt. Sowohl bei Icons als auch bei Text lassen sich deutliche Unterschiede erkennen. Bei Icons ergab sich im Schnitt eine Reaktionszeit von 1039 ms mit einer Standardabweichung von 131 ms. Bei Text benötigten die Probanden 1247 ms bei einer Standardabweichung von 115 ms. Der Unterschied zwischen der langsamsten und der schnellsten Antwort beträgt bei Icons 497 ms, bei Text 338 ms. Die Varianzanalysen zeigen so-

wohl bei Icons ($F(11,121) = 5.07, p < .001, p_{rep} > .99, \eta_p^2 = .32$) als auch bei Text ($F(11,121) = 2.68, p < .01, p_{rep} = .97, \eta_p^2 = .20$) einen signifikanten Effekt an.

Tabelle 10: Rangordnung der Items basierend auf dem Mittelwert der Reaktionszeit bei Zufallsanordnung für Icons und textuelle Darstellung.

Rangplatz	Icons	Reaktionszeit [ms] (Min; Max)	Text	Reaktionszeit [ms] (Min; Max)
1	■	796 (572; 1048)	Klimaanlage	1091 (909; 1474)
2	▶	876 (643; 1349)	Adressbuch	1098 (901; 1500)
3	☎	983 (755; 1469)	SMS	1119 (807; 1486)
4	🔧	1015 (817; 1283)	Hilfe	1169 (728; 2048)
5	⚙️	1015 (760; 1784)	Stop	1203 (907; 1563)
6	⬅️	1020 (789; 1263)	Einstellungen	1227 (843; 1651)
7	?	1030 (623; 1648)	Telefon	1274 (954; 1727)
8	✉️	1035 (632; 1437)	Sitzheizung	1285 (846; 1764)
9	🖱️	1081 (857; 1568)	Start	1321 (1075; 1829)
10	📄	1133 (672; 1756)	Lüftung	1361 (1011; 1860)
11	👉	1194 (806; 1740)	Zurück	1390 (1113; 1812)
12	🔊+	1293 (800; 2318)	Lautstärke	1429 (925; 2632)

Im Anschluss wurde für eine Überprüfung des Itemmaterials die Korrelation der Rangpositionen mit Spearmans Rho berechnet. Es ergab sich eine geringe Korrelation von $r_s = .21$, die nicht signifikant ist ($r_{crit} = .58$ für $p < .05$).

5.2.4.4 Diskussion

Experiment 5 stellt einen Replikationsversuch des Experiments von Arend et al. (1987) zu einem Vergleich des visuellen Suchverhaltens bei Icons und verbalen Beschriftungen dar. Eine Wiederholung dieser Untersuchung mit abgewandelten Stimuli erschien notwendig, da sowohl bei der Gestaltung von bildhaften Funktionsdarstellungen als auch bei der Erfahrung der Nutzer mit diesem Material verglichen mit dem Zeitpunkt der Studie wesentliche Veränderungen stattgefunden haben. Darüber hinaus wird mit dieser Untersuchung der Forderung nach mehr Replikationsexperimenten in der psychologischen Forschung Rechnung getragen (Cohen, 1994). Die Studie von

Arend und Kollegen wird seit ihrem Erscheinen häufig zitiert und bildet nicht selten die Grundlage für weiterführende Gestaltungsüberlegungen. Im vorliegenden Experiment hat sich dasselbe Muster ergeben: Die Suche nach Icons verläuft schneller und effizienter als die visuelle Suche nach textuellen Elementen.

Um eine bessere Generalisierbarkeit der Ergebnisse zu ermöglichen, wurden standardisierte Icons aus einem anderen Applikationskontext verwendet. Die strikte Trennung von abstrakten und konkreten Icons wurde dabei aufgehoben, da sie als Gestaltungsprinzip für Computeranwendungen nicht in dieser Form aufrecht zu erhalten ist. Die verbalen Beschriftungen wurden als Einwort-Konstruktionen realisiert, im Gegensatz zu den Zweiwort-Konstruktionen der Originalstudie.

Inhaltlich geht das Experiment der Frage nach, ob Icons besser als verbale Beschriftungen zur Orientierung auf grafischen Bedienoberflächen beiträgt. Für den Anwendungsfall der Steuerung eines multimodalen Bordsystems im Fahrzeug stellt sich diese Frage mit besonderem Hintergrund: Sollten Icons die visuell-manuelle Interaktion deutlich besser unterstützen und zu kürzeren Blickabwendungen von der Fahrbahn beitragen, kann das *Say What You See* Prinzip als Gestaltungsgrundlage der visuellen Schnittstelle zur Unterstützung von Spracheingabe nicht mehr ohne weiteres empfohlen werden. Interaktionsdesigner müssten auf andere Wege setzen, um einfach zu signalisieren, welche Sprachbefehle zu einem gegebenen Zeitpunkt verfügbar sind. Eine weitgehende Parallelität der sprachlichen und der manuell-visuellen Schnittstelle würde dadurch erschwert.

Itemanalyse

Mit unterschiedlich großen Suchmengen, 6 vs. 12 Elemente wurde der Schwierigkeitsgrad der visuellen Suche variiert. Als Stimuli dienten zwei Sets von Darstellungen, entweder verbale Funktionsbeschreibungen oder Icons. Die Reaktionszeiten geben Aufschluss über die Dauer der visuellen Suche. Wenn semantische Effekte im Vorfeld tatsächlich ausgeschlossen werden konnten, kann die Variation in den Reaktionszeiten nur auf die unterschiedliche Komplexität der Vorlage attribuiert werden (Darstellungsform bzw. Größe der Suchmenge). Wäre beispielsweise die inhaltliche Bedeutung eines Items besonders leicht oder schwierig zu erfassen gewesen, müsste sich das auf dessen textuelle und bildliche Repräsentation gleichermaßen auswirken. Wäre dieser semantische Effekt dominant, sollten die Rangreihen entsprechend eine gewisse Korrelation aufweisen. Die Itemanalyse zeigt durch die fehlende Korrelation der Rangreihen, dass die Suche tatsächlich unabhängig von konzeptuellen Eigenschaften der Items war und vielmehr aufgrund rein visueller Charakteristika erfolgte. Für die weiteren Experimente kann das Itemmaterial deshalb unverändert bleiben.

Neben der Unabhängigkeit der Rangreihen lässt die Analyse der einzelnen Items im Vergleich von Sprache und Bild interessante Rückschlüsse auf die Wirksamkeit be-

sonderer visueller Charakteristika zu. Insgesamt ergab sich eine geringere Spannweite zwischen den Items als bei Arend und Kollegen (1987), die in jeder Gruppe ca. 800 ms berichten. Verbale Beschriftungen streuten dabei deutlich weniger (338 ms), auch das Icon-Material zeigt einen geringeren Abstand zwischen Minimal- und Maximalwert (497 ms). Offensichtlich war das hier verwendete Material homogener. Im Hinblick auf die Beschriftungen ist die geringere Spannweite zwar nicht so überraschend, trotzdem ist bemerkenswert, dass ein visuell auffälliger Reiz wie „SMS“ nicht zu deutlich kürzeren Reaktionszeiten und damit einer größeren Streuung beigetragen hat. „SMS“ belegt sogar nur den dritten Rang nach „Klimaanlage“ und „Adressbuch“. Besondere Auffälligkeiten in der visuellen Struktur der verbalen Stimuli haben also im Gegensatz zu der Arend-Studie nicht zu auffällig kurzen Suchzeiten geführt. Bei der Betrachtung der Rangstruktur fällt auf, dass weder eine Rolle gespielt hat, ob die Wörter zusammengesetzt oder aus einem Begriff waren, noch ob sie besonders lang oder kurz waren.

Bei den Icons hingegen lassen sich deutlich eher Begründungen für die Reihenfolge der Rangliste geben. Besonders kurze Suchzeiten ließen sich für die abstrakten Symbole „Stop“ und „Start“ nachweisen. Dabei ist wahrscheinlich weniger die Abstraktheit ein Vorteil – der Pfeil für „Zurück“ und das Fragezeichen für „Hilfe“ finden sich nur im Mittelfeld – sondern vielmehr gestaltpsychologische Aspekte wie visuelle Prägnanz und Geschlossenheit. „Start“ und „Stop“ waren die einzigen Symbole, die nicht aus Linien bestanden, sondern ausgefüllt waren. Sie waren visuell sehr einfach und haben sich deutlich abgehoben. Wesentlich komplexer hingegen war das langsamste Symbol, „Lautstärke“. Es bestand aus zwei Elementen (Lautsprecher und Pluszeichen). Die zeitliche Ordnung spiegelt bei den Icons relativ anschaulich wieder, wie sich die Gestaltqualität von Icons auf das Suchverhalten auswirkt. Inwiefern es sich um abstrakte oder konkrete Icons handelte, scheint erst an zweiter Stelle relevant zu sein.

Lerneffekte

Lerneffekte ließen sich in der vorliegenden Studie nicht nachweisen. In der Arend-Untersuchung hingegen waren insbesondere nach dem ersten Block für Icons deutliche Reduktionen der Reaktionszeiten zu erkennen. Das lässt sich nur dadurch erklären, dass das Mapping zwischen Icons und Referenten in dieser Studie schwieriger zu lernen war als in dem vorliegenden Experiment. Durch die zufällige Anordnung der Items war ein Lernen der Position nicht möglich, somit scheidet dieser Faktor für eine Verkürzung der Reaktionszeiten über die Zeit aus. Entsprechend bleibt nur die Erlernbarkeit der Elemente als Faktor. Aufgrund der sorgfältigen Gestaltung der Icons und weil die einzelnen Funktionen der Items deutlicher voneinander abgegrenzt waren als

bei Arend und Kollegen, gab es kaum Steigerungspotential. Deshalb war keine substantielle Abnahme der Reaktionszeiten zu verzeichnen.

Reaktionszeiten

Bezüglich der Grundaussage, dass unter bestimmten Bedingungen Icons schneller gefunden werden können als textuelle Elemente, kann basierend auf vorliegendem Material die Studie von Arend et al. (1987) als erfolgreich repliziert angesehen werden. Es ergab sich ebenfalls ein deutlicher Unterschied in den Suchzeiten zu Gunsten der Icons, der mit ansteigender Größe der Suchmenge zunahm. Darauf deuten die unterschiedlichen Steigungen hin, gestützt durch die signifikante Interaktion von Präsentationsformat und Suchmenge. Auch die sehr geringen Fehlerraten sind ähnlich. Das Material war folglich in beiden Experimenten hinreichend gut bekannt. In einiger Hinsicht differieren die Studien jedoch: Arend und Kollegen fanden für konkrete Icons keinen Unterschied zu verbalen Elementen. In der vorliegenden Studie wurden sowohl konkrete als auch abstrakte Icons durchmischt verwendet. Der Effekt konnte trotzdem gezeigt werden. Berücksichtigt man, dass die die textuellen Bezeichnungen in diesem Experiment sogar deutlich einfacher waren als in der Arend-Studie, ist dies umso erstaunlicher. Es scheint sich also um einen sehr mächtigen Vorteilseffekt für Icons zu handeln, der bei komplexerem verbalem Material noch deutlicher ausgeprägt sein dürfte. Dass Arend et al. für konkrete Icons keinen Vorteil gefunden haben, ist vermutlich durch den konzeptuellen Aufbau ihres Experiments bedingt. Sie etablierten eine strikte Trennung von konkreten und abstrakten Icons und gestalteten Unterschiede der konkreten Icons nur durch schwer erkennbare lokale Features. Korrekt ist, dass nur bei sehr abstrakten Icons mit ausschließlich globalen Features gearbeitet werden kann und damit pop-out Effekte herstellbar sind. Das muss im Umkehrschluss jedoch nicht bedeuten, dass sich konkrete Icons nur durch schwer differenzierbare lokale Features unterscheiden müssen. Bei entsprechender Gestaltung können lokale Merkmale derart verwendet werden, dass die Gesamterscheinung einzelner Icons hinreichend unterschiedlich ist. Damit sind zwar keine pop-out Effekte zu erwarten, aber deutlich stärkere Effekte von Icons gegenüber verbalen Kennzeichnungen, wie dieses Experiment gezeigt hat.

Eine Analyse der Steigung ermöglicht die Einschätzung der Sucheffizienz. Bedauerlicherweise gaben Arend und Kollegen weder die exakten Mittelwerte für die jeweiligen Suchmengen noch die Steigungen an. Deshalb konnten diese nur auf Basis ihres grafischen Materials durch Abmessung der Mittelwerte geschätzt werden. Ein direkter Vergleich unterstützt die Vermutung, dass das zweizeilige verbale Material in ihrer Studie deutlich schwieriger zu erkennen war: Während in der vorliegenden Untersuchung eine Steigung von 85 ms/Item für Text ermittelt wurde, ergab sich bei Arend et al. (1987) eine Steigung von ca. 215 ms/Item. Bereits bei der Betrachtung des Graphen

(S.419) fällt auf, dass der Anstieg von ca. 1330 ms auf ca. 2610 ms extrem deutlich ist. Des Weiteren wird die oben getroffene Annahme unterstützt, dass sich das verwendete Itemmaterial im vorliegenden Experiment zwischen den Extrempolen der abstrakten und konkreten Icons von Arend et al. einordnet. Der hier gefundenen Steigung von 55 ms/Item steht eine geschätzte Steigung von ca. 22 ms/Item für abstrakte Icons und ca. 140 ms/Item für konkrete Icons bei Arend und Kollegen gegenüber.

Orientiert an den Standards der Grundlagenliteratur, die im Abschnitt 5.2.4.1 (Methode, Search-and-Select Paradigma) dargelegt wurden, tendiert selbst die geringe Steigung von ca. 22 ms/Item bereits zu ineffizienter Suche. Nach dieser Einteilung muss folglich die Interpretation von Arend und Kollegen, abstrakte Icons könnten unabhängig von der Größe der Suchmenge parallel abgesucht werden, zumindest in Frage gestellt werden. Nichtsdestotrotz ist diese Steigung im Vergleich zu den anderen von Arend et al. und den hier ermittelten Geradengleichungen bemerkenswert gering. Ebenfalls bemerkenswert ist, dass in beiden Studien nur piktografisches Material das Potential hatte, relativ geringe Steigungen zu erzeugen. Verbales Material hingegen führte immer zu relativ großen Steigungen. Dies lässt sich – im gegebenen experimentellen Setting – durchaus als klarer Vorteil für bildliche Funktionsdarstellungen interpretieren.

Neben dieser Erkenntnis zur Gestaltung sollte aus der Diskussion der Steigung eine Konsequenz für weitere Studien gezogen werden. Viele Arbeiten beschäftigen sich mit der Güte der Gestaltung von Icons. Einheitliche Vorgehensweisen oder übergreifende Maße existieren jedoch kaum. Eine Ausnahme bilden die Prüfungen zur Verständlichkeit nach ISO 9186 (2001). An dieser Stelle scheint es empfehlenswert, über die Verständlichkeitsprüfung hinaus experimentelle Methoden zu verwenden, die eine Abschätzung der Steigung erlauben. Sobald über mehrere Studien hinweg in einem einheitlichen Vorgehen für verschiedene Arten von Icons oder auch verbalen Funktionsbeschreibungen systematisch die Steigungen analysiert werden, lässt sich ähnlich zu den Kategorien der Sucheffizienz in der Grundlagenforschung (Wolfe, 1998) ein Qualitätskriterium etablieren. Dadurch können mit einem einfachen Testaufbau wie dem Search-and-Select Paradigma ganze Itemsets auf ihre Tauglichkeit für zeitkritische Aufgaben geprüft und übergreifend verglichen werden. Zusätzlich kann durch ein einheitliches Maß besser nach generellen Ergebnismustern und erfolgreichen Gestaltungsvarianten über Studien hinweg gesucht werden. Die vorliegende Arbeit versteht sich als erster Beitrag zu diesem Ansatz und will in diesem Sinne einen Einstieg leisten.

5.2.4.5 Schlussfolgerungen

Die zufällige Anordnung von Icons und textuellen Funktionsbeschriftungen in einer grafischen Benutzeroberfläche führt zu kürzeren Reaktionszeiten für Icons. Als Vor-

aussetzung muss eingeräumt werden, dass die symbolische Darstellung und das Abbildungsverhältnis zwischen Icons und Referent bekannt sein muss. Dieses Ergebnis ist eine Replikation der Studie von Arend et al. (1987), die mit den geänderten Voraussetzungen bei den Nutzern nach fast 20 Jahren notwendig war. Es stellt in der Hinsicht eine Ausweitung dar, als dasselbe Ergebnismuster mit kürzeren verbalen Beschriftungen erreicht wurde und genormte Symbole ohne klare Auftrennung in konkrete und abstrakte Icons zum Einsatz kamen. Bei zufälliger Anordnung werden Icons – sofern ihre Bedeutung bekannt ist – schneller gefunden als Elemente mit textueller Beschriftung.

5.2.5 Experiment 6: Ortsfeste Itemanordnung

Wissen über den Anzeigeort eines Instruments ist eine der wichtigsten Voraussetzungen für zeiteffiziente Ablese- und Bedienprozesse. Ortsinformationen werden sehr gut memoriert, es gibt unter bestimmten Randbedingungen sogar Hinweise auf eine weitgehend automatische Speicherung der Position von Objekten (Ellis, 1991; Hasher & Zacks, 1979; Shadoin & Ellis, 1992). Bei der Usability-Optimierung von Produkten wird diesem Aspekt der visuellen Gestaltung ein hoher Stellenwert eingeräumt, der entsprechend auch in einem experimentellen Vergleich von Icons und textueller Beschriftung berücksichtigt werden muss.

Green und Barnard (1990) sowie Lansdale (1988) geben an, dass Ortskonsistenz anfängliche Unterschiede in der Erkennbarkeit von Icons teilweise kompensieren kann. Bestehen bei zufälliger Anordnung noch massive Vorteile für abstrakte Icons gegenüber konkreten Icons in der visuellen Suchzeit, schwächt sich dieser Effekt durch konstante Anzeigeorte deutlich ab. Die Suche nach einem Item ist in diesem Fall nicht mehr ausschließlich auf visuelle Charakteristika angewiesen. Je nach Komplexität und Größe der Suchmenge kann sie vollständig oder zumindest partiell auf Gedächtnisleistungen basieren.

Bisher existieren nur Untersuchungen, die die Wirkung einer ortsfesten Anordnung bei unterschiedlichen Formen von Icons analysiert haben. Ein direkter Vergleich von Icons und verbalen Beschriftungen steht noch aus. A priori lässt sich sicher aussagen, dass durch viel Erfahrung im langzeitlichen Umgang mit einer konkreten grafischen Benutzeroberfläche die Positionen der Steuerelemente auswendig bekannt sind. Dann macht es sicherlich keinen Unterschied mehr, ob Icons oder textuelle Beschriftungen die Funktion bezeichnen. Nicht nur für den hier dargestellten Anwendungsfall der Steuerung von Komfortfunktionen bei der Fahrzeugführung, auch für andere Kontexte gilt jedoch, dass nicht beliebig viel Zeit für diesen Lernprozess vorausgesetzt werden darf. Damit steht in erster Linie die Frage im Raum, ob überhaupt bei ortskonstanter Darbietung der in Experiment 5 gefundene Unterschied zwischen Icons oder Text

fortbesteht und wenn ja, ob dieser nach wenigen Lerndurchgängen zu eliminieren ist. Dieser Frage widmet sich Experiment 6.

Im Sinne einer Erweiterung der Untersuchung von Green und Barnard (1990) lassen sich folgende weiterführende Merkmale aufzählen. Statt eines Vergleichs von bildhaftem Material untereinander werden wie im vorherigen Experiment Text und Bild einander gegenübergestellt. Es werden nur drei experimentelle Blöcke absolviert, da bei dem Icon-Vergleich von Green und Barnard bereits nach drei Blöcken kein weiterer Lernerfolg zu erkennen war, die Reaktionszeiten also nicht mehr weiter abnahmen. Diese Anzahl an Wiederholungen erscheint auch aus der Sicht einer gebrauchstauglichen Schnittstellengestaltung noch akzeptabel. Darüber hinaus werden ebenfalls abweichend von Green und Barnard wieder unterschiedlich große Suchmengen in den Versuchsplan miteinbezogen, um durch eine vergleichende Analyse der Steigungen den Beitrag der Gedächtnisaspekte auf die Sucheffizienz gegenüber Experiment 5 ermitteln zu können. Verzichtet wird hingegen auf eine Variation der Anzahl der örtlich fixierten Elemente. Insofern wird dieses Experiment mit demselben Aufbau wie Experiment 5 angegangen, nur die zufällige Anordnung der Items nach jeder Aufgabe wird durch einen ortskonstanten Aufbau ersetzt. Dies soll zum einen die Frage beantworten, wie sich Gedächtnisaspekte bei der gegebenen Aufgabensituation auswirken, und zum anderen, wie leicht sich die Position von Icons und verbalen Beschriftungen im Vergleich erlernen lässt.

5.2.5.1 Methode

An Experiment 6 nahmen zwölf neue Versuchspersonen (acht Frauen, vier Männer) teil. Ihr Alter lag zwischen 22 und 35 Jahren (Mittelwert: 28 Jahre). Sie erhielten für ihre Teilnahme eine angemessene Aufwandsentschädigung. Alle Probanden verfügten über normale oder voll korrigierte Sehschärfe, sprachen Deutsch als Muttersprache und gaben regelmäßige bis häufige Computernutzung an. Sie waren nicht über die Hypothesen des Experiments informiert.

Das Vorgehen (Versuchsmaterial und -stimuli sowie das Design und die Ausbalancierung der Bedingungsabfolgen) war bis auf nachfolgend beschriebene Ausnahme identisch mit Experiment 5: In jeder experimentellen Bedingung wurde nur einmal das Layout der Items zufällig angeordnet, nicht nach jeder Aufgabe wie in Experiment 5. Dieser zufällige Aufbau erfolgte jeweils bei der ersten Aufgabe des ersten experimentellen Blocks. Alle Elemente blieben dann sowohl für den Rest dieses Blocks als auch für die zwei folgenden Blöcke derselben Bedingung örtlich fixiert. Das bedeutet, dass nur bei einem Wechsel der Suchmenge (12 Elemente, 6 Elemente Subset A, 6 Elemente Subset B) und bei einem Wechsel des Präsentationsformats (Icons, Text) ein neuer zufälliger Aufbau stattfand. In jeder Variante wurden die Probanden so mit der Not-

wendigkeit konfrontiert, ein neues Layout zu erlernen. Dieser Umstand wurde den Probanden bekannt gegeben.

5.2.5.2 Hypothesen

Aufgrund der bekannten Gedächtnisvorteile für bildhaftes Material (vgl. Engelkamp, 1991) liegt nahe, dass das Layout einer Matrix von Icons schneller gelernt wird als eine Anordnung von textuellen Elementen. Dies sollte sich in schneller abnehmenden Reaktionszeiten der Icons über die experimentellen Blöcke hinweg zeigen.

Nachdem die Probanden beim Suchen nicht mehr nur allein auf visuelle Merkmale angewiesen sind, sollten die Suchzeiten im Vergleich zu Experiment 5 kürzer sein. Darüber hinaus ist zu erwarten, dass der Unterschied in den Reaktionszeiten zwischen Icons und verbalen Funktionskennzeichnungen dadurch generell abnimmt.

5.2.5.3 Ergebnisse

Reaktionszeiten und Fehler

Die Reaktionszeiten wurden wie zuvor zunächst über alle Blöcke hinweg differentiell für Suchmengen und Präsentationsformat gemittelt. Wenn die Funktionen durch textuelle Beschriftungen repräsentiert waren, ergab sich eine mittlere Reaktionszeit von 617 ms ($SE = 60.21$) bei 6 Items und 1062 ms ($SE = 90.20$) bei 12 Items. Icons führten zu einer mittleren Reaktionszeit von 540 ms ($SE = 39.81$) bei 6 Items und 797 ms ($SE = 54.51$) bei 12 Items. Das Ergebnismuster ist in Abbildung 38 dargestellt.

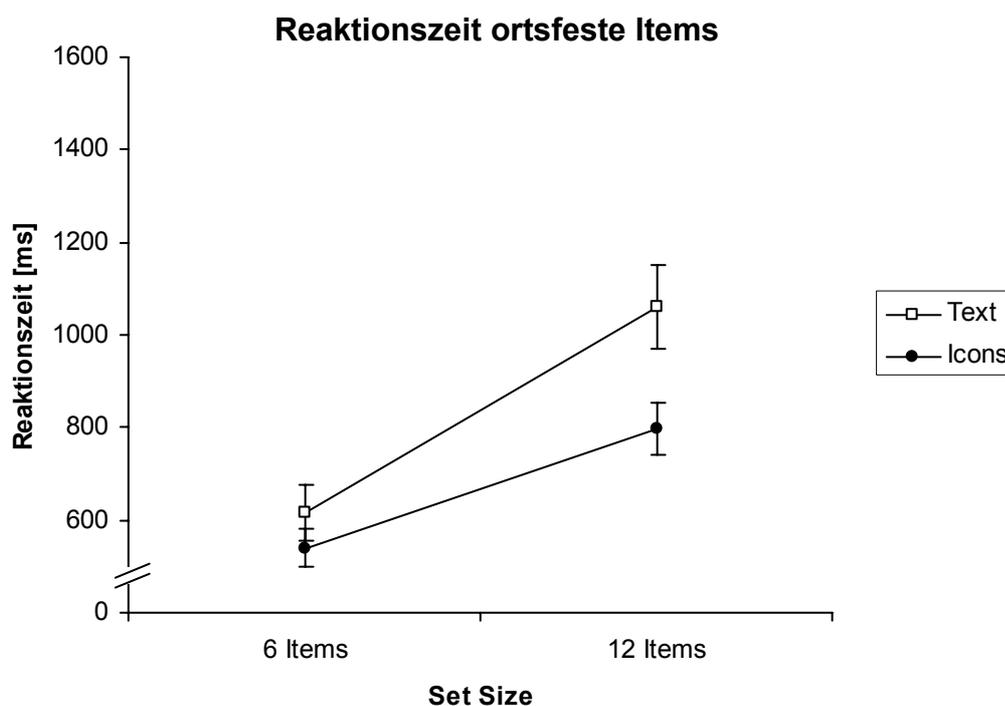


Abbildung 38: Mittlere Reaktionszeiten von Text und Icons als Funktion der Größe der Suchmenge bei ortskonstanter Anordnung der Items.

Bereits die visuelle Inspektion der Grafik zeigt, dass nach wie vor ein von der Größe der Suchmenge abhängiger Unterschied zwischen Icons und Text existiert. Visuelle Suche in einer ortsfesten Anordnung von Icons führte zu einer Steigung von 43 ms/Item, textuelle Elemente dagegen hatten 74 ms/Item zur Folge. Eine zweifaktorielle Varianzanalyse (Suchmenge \times Präsentationsformat) bestätigt dieses Bild, es ergab sich eine signifikante Interaktion der beiden Within-Faktoren ($F(1,11) = 11.59$, $p < .01$, $p_{\text{rep}} = .96$, $\eta_p^2 = .51$). Die Art der Interaktion lässt auch hier eine globale Interpretation der Haupteffekte zu. Auf dieser Ebene kann ein allgemeiner Vorteil für Icons statistisch nachgewiesen werden, die mit 669 ms im Mittel ($SE = 42.51$) schnellere Antworten ermöglichten als Texte mit 839 ms ($SE = 70.48$), $F(1,11) = 14.11$, $p < .01$, $p_{\text{rep}} = .97$, $\eta_p^2 = .56$. Unter Berücksichtigung der Interaktion wird klar, dass dieser Effekt primär durch die Suchmenge mit 12 Items hervorgerufen wird. Auch der Haupteffekt für die Größe der Suchmenge ist signifikant. Bei 6 Items erfolgte die Auswahl bereits nach 578 ms ($SE = 36.19$), bei 12 Elementen erst nach 929 ms ($SE = 58.47$), $F(1,11) = 97.55$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .90$.

Die Fehlerrate war durchweg unter 3% und damit erneut sehr gering in diesem Experiment. Bei textueller Beschriftung wählten die Probanden in 0.7% der Fälle das falsche Item (sowohl bei 6 als auch bei 12 Items). Während bei 6 Icons ebenfalls nur 1.0% Fehler gemacht wurden, liegt die Rate bei 12 Icons mit 2.6% etwas höher. In Häufigkeiten ausgedrückt handelt es sich um 11 Fehler von 432 Eingaben. Aufgrund dieser Seltenheit von Fehlerereignissen wird von einer tiefer gehenden Analyse Abstand genommen.

Besonders interessant ist das Ergebnis der Reaktionszeiten im direkten Vergleich mit Experiment 5. Hierfür wurde eine dreifaktorielle Varianzanalyse mit den Within-Faktoren Präsentationsformat (Icons, Text), Größe der Suchmenge (6, 12) und dem Between-Faktor Experiment (Experiment 5, Experiment 6) gerechnet. Grundsätzlich unterscheiden sich die beiden Experimente in ihrem Wertebereich. In Experiment 5 sind die Zeiten generell langsamer als in Experiment 6, wie der signifikante Haupteffekt für Experiment bestätigt ($F(1,22) = 33.06$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .60$). In beiden Experimenten werden die Icons schneller erinnert, der Haupteffekt für Präsentationsformat ist signifikant ($F(1,22) = 36.22$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .62$). Ebenso lässt sich in beiden Experimenten ein signifikanter Einfluss der Größe der Suchmenge feststellen: Bei zwölf Items wird deutlich länger gesucht als bei sechs Elementen ($F(1,22) = 190.52$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .90$). Wie sich die Größe der Suchmenge auswirkt, hängt jedoch vom Präsentationsformat ab: Mit zunehmender Größe nimmt der Unterschied zwischen Icons und Text zu, die Interaktion Präsentationsformat \times Suchmenge ist wie in den einzelnen Experimenten signifikant:

$F(1,22) = 14.87, p = .001, p_{\text{rep}} = .97, \eta_p^2 = .40$. Die zweifaktorielle Interaktion Präsentationsformat \times Experiment $F(1,22) = 0.35, p = .56, p_{\text{rep}} = .46$ ist nicht signifikant. Ebenso lässt sich keine Signifikanz für die Interaktion Größe der Suchmenge \times Experiment feststellen ($F(1,22) = 1.62, p = .22, p_{\text{rep}} = .71$), d.h. die Suche durch ortskonstante Items ist *nicht* effizienter. Die dreifaktorielle Interaktion Präsentationsformat \times Größe der Suchmenge \times Experiment erreicht ebenfalls nicht statistische Signifikanz: $F(1,22) = 0.07, p > .9, p_{\text{rep}} = .14$.

Lerneffekte

Für jeden der drei experimentellen Blöcke wurden die mittleren Reaktionszeiten wiederum nur für Durchgänge mit 12 Items getrennt für Icons und Text berechnet. Die mittleren Reaktionszeiten (Standardfehler in Klammern) von Block 1 nach Block 3 betragen bei Text 1230 ms (93.31), 1060 ms (117.21) und 895 ms (93.37), bei Icons 1051 ms (63.27), 721 ms (62.39) und 619 ms (56.32). Diese Daten wurden mit einer zweifaktoriellen Varianzanalyse mit den Within-Faktoren Präsentationsformat und Block ausgewertet. Im Gegensatz zum ersten Experiment ließen sich durchaus Übungseffekte nachweisen. Von Block 1 bis 3 nimmt die Reaktionszeit stetig ab (siehe Abbildung 39): 1141 ms ($SE = 58.21$), 891 ms ($SE = 73.92$), 757 ms ($SE = 60.63$). Der Haupteffekt für die Blocknummer ist signifikant, $F(2,22) = 25.99, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .70$.

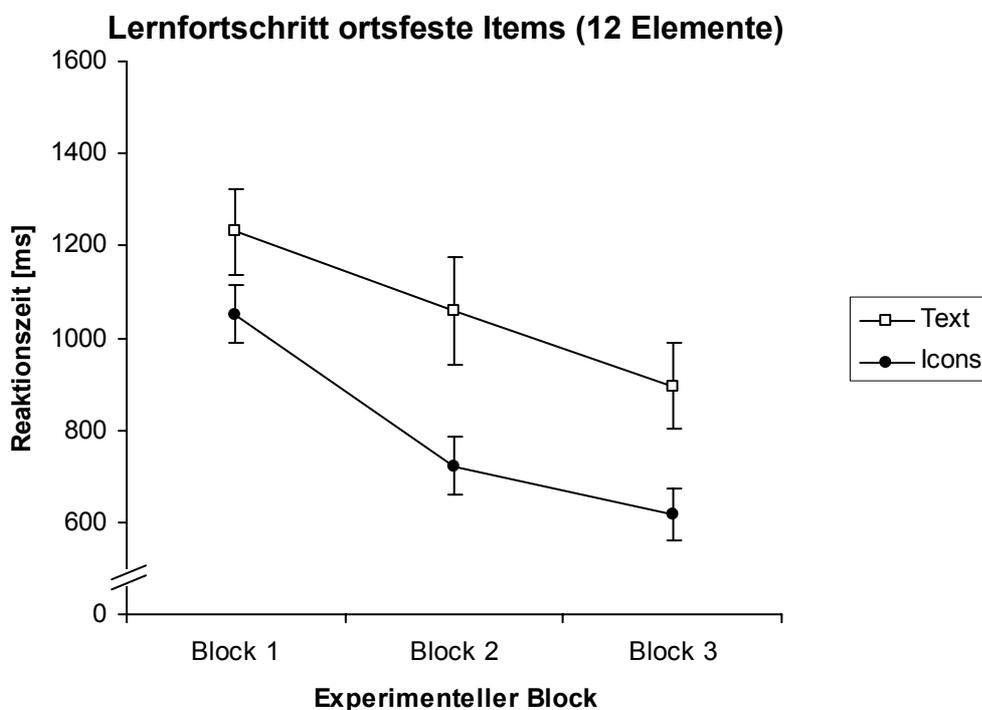


Abbildung 39: Lernfortschritt als Veränderung der mittleren Reaktionszeiten über die Zeit bei ortsfester Anordnung der Items.

Der an dieser Stelle weniger relevante Haupteffekt für Präsentationsformat wird ebenfalls signifikant, die Reaktionszeitdaten im ersten Analyseschritt ließen dies bereits erkennen (Text: $M = 1062$ ms, $SE = 61.65$; Icons: $M = 797$ ms, $SE = 46.17$), $F(1,11) = 15.62$, $p < .01$, $p_{\text{rep}} = .98$, $\eta_p^2 = .59$. Betrachtet man alle drei experimentellen Blöcke, kann keine Interaktion festgestellt werden ($F(2,22) = 2.11$, $p = .15$, $p_{\text{rep}} = .77$, $\eta_p^2 = .16$). Die Suche wird also nicht effizienter, nur kürzer. Nachdem ein möglicher verbesserter Lernvorteil für Icons transient ist, sollte sich dieser am ehesten nach dem ersten Block zeigen. Rechnet man eine Varianzanalyse nur für die ersten zwei Blöcke, d.h. Präsentationsformat (Text, Icons) \times Block (1,2), lässt sich jedoch ebenfalls keine Interaktion nachweisen ($F(1,11) = 2.98$, $p = .11$, $p_{\text{rep}} = .80$).

Zusätzlich wurde eine dreifaktorielle Varianzanalyse mit den Within-Faktoren Präsentationsformat (Icons, Text), Block (1-3) und Experiment (Exp. 5, Exp. 6; Between-Faktor) durchgeführt. Grundsätzlich besteht ein Unterschied zwischen den Experimenten (Haupteffekt Experiment signifikant, $F(1,22) = 25.68$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .54$). Anhand einer signifikanten Block \times Experiment Interaktion ($F(2,44) = 11.39$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .34$) lässt sich erkennen, dass sich in Experiment 6 das Suchverhalten im Vergleich zu Experiment 5 über die experimentellen Blöcke deutlich verändert hat. Während in Experiment 5 ein weitgehendes Gleichbleiben der Reaktionszeiten zu verzeichnen war, sinken diese im zweiten Experiment kontinuierlich. Ein Haupteffekt für Präsentationsformat ($F(1,22) = 39.85$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .64$) ist nachweisbar, ebenso der Haupteffekt für Block ($F(2,44) = 16.87$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .43$). Das Präsentationsformat hat sich in den Experimenten nicht unterschiedlich ausgewirkt (Präsentationsformat \times Experiment, $F(1,22) = 0.14$, $p = .71$, $p_{\text{rep}} = .35$) und auch über die Blöcke keine Wechselwirkung verursacht (Präsentationsformat \times Block, $F(2,44) = 2.37$, $p = .11$, $p_{\text{rep}} = .81$). Ebenso gibt es keine differentiellen Effekte für das Präsentationsformat in den verschiedenen Blöcken über die Experimente hinweg. Die dreifaktorielle Interaktion ist nicht signifikant ($F(2,44) = 0.004$, $p > .99$, $p_{\text{rep}} = .03$).

Itemanalyse

Durch die ortsfeste Anordnung der Elemente bestimmen sich die Suchzeiten der Items nicht mehr rein durch visuelle Charakteristika. Trotzdem zeigen einfaktorielle Varianzanalysen, dass sich die Items in den jeweiligen Formaten signifikant bezüglich der Suchzeiten unterscheiden. Wie in Experiment 5 gehen sowohl Daten beider Suchmenügrößen und aller Blöcke in die Analyse ein. Zwischen der langsamsten und der schnellsten Antwort ergab sich bei Icons eine Differenz von 291 ms und bei Text von 365 ms. Die Standardabweichung bei Icons lag bei 81 ms (Mittelwert: 669 ms) und bei

Text bei 116 ms (Mittelwert: 839 ms). Wenngleich die Unterschiede in den Mittelwerten bei den Icons damit geringer sind als im ersten Experiment, lässt sich immer noch ein signifikanter Einfluss der einzelnen Items mit vergleichbarer Effektstärke nachweisen ($F(11,121) = 3.65, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .25$). Ein ebensolches Ergebnis lässt sich für Text finden ($F(11,121) = 3.60, p < .001, p_{\text{rep}} > .99, \eta_p^2 = .25$). Für eine geordnete Auflistung der Mittelwerte siehe Tabelle 11.

Betrachtet man Experiment 5 und 6 im Vergleich, steht hier die folgende Frage im Vordergrund: Wie verändert sich das Suchverhalten auf der Ebene einzelner Items durch Informationen über den Anzeigeort der Items? Zur Beantwortung wurde die Korrelation zwischen der Rangposition der jeweiligen Itemsätze berechnet (Spearman Rangkorrelationskoeffizient). Für die Rangplätze der Beschriftungen ergab sich eine signifikante mittlere Korrelation von $r_s = .62$ ($r_{\text{crit}} = .58$ für $p < .05$) zwischen den Experimenten. Bei den Icons war sogar eine hohe Korrelation von $r_s = .79$ nachweisbar, die sehr signifikant ist ($r_{\text{crit}} = .71$ für $p < .01$).

Tabelle 11: Rangordnung der Items basierend auf dem Mittelwert der Reaktionszeit bei ortsfestem Layout für Icons und textuelle Darstellung.

Rangplatz	Icons	Reaktionszeit [ms] (Min; Max)	Text	Reaktionszeit [ms] (Min; Max)
1	■	581 (404; 1195)	Stop	683 (443; 1313)
2	▶	602 (393; 1117)	SMS	699 (460; 1295)
3	⚙	606 (392; 887)	Hilfe	723 (374; 1240)
4	?	607 (348; 949)	Start	769 (467; 1443)
5	🔧	620 (391; 1019)	Adressbuch	770 (429; 1464)
6	⬅	656 (380; 1026)	Einstellungen	831 (428; 1690)
7	📞	659 (486; 897)	Telefon	849 (474; 1897)
8	📞	660 (413; 1044)	Klimaanlage	857 (499; 1503)
9	🚗	703 (363; 1370)	Sitzheizung	926 (635; 1332)
10	✉	715 (398; 1171)	Lautstärke	949 (457; 1812)
11	👉	741 (429; 1374)	Zurück	968 (397; 1429)
12	🔊+	872 (504; 1228)	Lüftung	1048 (517; 1404)

5.2.5.4 Diskussion

Experiment 6 ging der Frage nach, ob sich der Unterschied zwischen Icons und textuellen Beschriftungen erhält, wenn nicht mehr ausschließlich aufgrund visueller Charakteristika der Stimuli eine Suche über die Items stattfindet, sondern Wissen über den Anzeigort hinzukommt. Statt einer rein zufälligen Anordnung der Elemente vor jeder Aufgabe blieb die Anordnung der Suchmatrix nach einer initialen Zufallsanordnung konstant. Dieses Experiment simuliert damit die Situation eines Systemnutzers, der einige Aspekte der grafischen Oberfläche gelernt hat, der aber noch nicht über ein ausgereiftes mentales Modell der Funktionszusammenhänge verfügt.

Wie im vorangegangenen Experiment wurden über mehrere experimentelle Blöcke hinweg unter Variation des Präsentationsformats und der Größe der Suchmenge die Reaktionszeiten zum Auswählen des gesuchten Elements gemessen. Ein Effekt der ortskonstanten Anordnung ließ sich sowohl durch einen Lernfortschritt über die Zeit als auch durch generell niedrigere Reaktionszeiten in Experiment 6 nachweisen. Der Vorteil für Icons gegenüber Texten blieb in überraschender Deutlichkeit erhalten.

Lerneffekte

Durch die Einführung der Ortskonstanz verkürzten sich mit fortschreitender Zeit die Reaktionszeiten der Versuchspersonen. Damit ließ sich ein deutlicher Lerneffekt der Position der Items nachweisen, die experimentelle Manipulation war erfolgreich. Aufgrund des Bildüberlegenheitseffekts war erwartet worden, dass die Anordnung von Icons schneller gelernt werden würde als die Anordnung von textuellen Elementen. Dies hätte sich in einer schnelleren Abnahme der Reaktionszeiten für Icons zeigen müssen, insbesondere im Übergang von Block 1 auf Block 2. Wenngleich in der grafischen Darstellung der Daten (Abbildung 39) dieser Eindruck entsteht, ließ sich diese Vermutung nicht varianzanalytisch untermauern.

Überraschend ist die beträchtliche Kluft zwischen Icons und Text, die auch im dritten Block besteht. Zunächst stellt sich hier die Frage nach einer Generalisierbarkeit des Befundes, da nur eine recht begrenzte Lernerfahrung im Experiment etabliert wurde. Zwar kann mit Sicherheit angenommen werden, dass nach einer großen Zahl von Lerndurchgängen die Positionen gut genug gemerkt werden, um eine weitestgehende Angleichung zu erwirken. Allerdings sollte hier der Anwendungskontext der Untersuchung beachtet werden: Für eine nutzerfreundliche und verkehrssichere Bedienbarkeit eines Bordsystems im Fahrzeug sollte nicht von einer großen Anzahl von Lernokkasionen ausgegangen werden. Erstens liegt die Aufmerksamkeit der Nutzer auf der Primäraufgabe Fahrzeugführung, zweitens werden viele Komfortfunktionen eher selten genutzt und drittens wechseln manche Nutzer, z.B. Mietwagenfahrer, häufig das Fahrzeug. Damit ändert sich selbst bei genormten Symbolen und Begrifflichkeiten die konkrete Erscheinungsform der Bedienoberfläche. Insofern kann die vorgefundene

Datenlage durchaus direkt übertragen werden. Unter dieser Prämisse stellt sich weitergehend die Frage nach dem Grund für das Fortbestehen des Unterschieds, da durch die Gedächtniskomponente eine Annäherung der Reaktionszeiten zu erwarten stand. Zur Beantwortung sollen zunächst die Reaktionszeiten in Abhängigkeit der Suchmenge erörtert werden, die weiter gehende Rückschlüsse auf die Charakteristika der Item-Suche ermöglicht.

Reaktionszeiten

Das Muster der Reaktionszeiten bei verschiedenen großen Suchmengen ähnelt stark den Resultaten des ersten Experiments. Die Items werden grundsätzlich bei größeren Suchmengen langsamer gefunden, Icons sind dabei generell schneller als Texte. Dabei wirken diese Faktoren zusätzlich zusammen: Je größer die Suchmenge, desto stärker tritt der Vorteil für Icons in Erscheinung. Durch Ortskonstanz wurden die Steigungen fast nicht beeinflusst. Bei zufälliger Anordnung wurden 55 ms/Item für Icons und 85 ms/Item für Text erzielt, ortsfestes Layout hatte entsprechende Steigungen von 42 ms/Item und 74 ms/Item zur Folge. Der Unterschied bei der großen Suchmenge erhielt sich sogar fast im selben Ausmaß von ca. 300 ms. Diese Ergebnisse stehen im Widerspruch zu der geäußerten Hypothese, dass sich der Unterschied zwischen Icons und Text durch Ortskonstanz reduzieren lassen sollten.

Eigentlich wurde erwartet, dass durch die Mitwirkung von Gedächtnisleistungen bei der Suche der Einfluss visueller Charakteristika in ihrer Bedeutung abnimmt und damit in Bezug auf Reaktionszeiten Icons und verbales Material näher zusammenführt. Eine Analyse über die Experimente hinweg wies als einzigen Unterschied zwischen den Studien durch den signifikanten Haupteffekt für den Faktor für Experiment aus, dass eine generelle Verkürzung der Reaktionszeiten mit der ortsfesten Anordnung einhergeht. Für die Interpretation ist primär relevant, dass keiner der Interaktionseffekte, die den Faktor Experiment berücksichtigen, statistische Signifikanz erreichte. Weder ändern sich die Unterschiede im Präsentationsformat über die Experimente, noch der Einfluss, den die Größe der Suchmenge hat. Die fehlende dreifaktorielle Interaktion zeigt darüber hinaus, dass sich auch der Unterschied durch die Suchmengen in den verschiedenen Präsentationsformaten nicht über die Experimente hinweg verändert. Der letzte Punkt deutet darauf hin, dass Ortskonstanz die Effizienz der Suche nicht entscheidend beeinflusst hat. Diese Schlussfolgerung ist besonders überraschend, wenn Erkenntnisse der Grundlagenliteratur einbezogen werden. In einem Experiment von Chun und Jiang (1998) führte die räumliche Fixierung von Elementen durchaus zu effizienterer Suche. Selbstverständlich existieren weit reichende Unterschiede zwischen dieser Studie und dem vorliegenden Experiment. So wussten beispielsweise die Versuchspersonen von Chun und Jiang nicht von vornherein, dass in bestimmten Trials Ortskonstanz vorliegen wird. Deshalb sprechen Chun und Jiang auch von einem

impliziten Lernprozess. Ferner sind das Stimulusmaterial und das Vorgehen im Versuch aufgrund der Natur der Studie zu dem hier verwendeten sehr unterschiedlich, weshalb ein Vergleich der Ergebnisse nicht ohne weiteres möglich ist. Trotzdem bleibt der fehlende Effekt auf die Sucheffizienz in vorliegendem Experiment ein recht erstaunlicher Befund.

Itemanalyse

Auch bei ortskonstanter Anordnung der Items waren die visuellen Charakteristika der Items relevant für schnellen Sucherfolg. Zwar ging die Spannweite gegenüber dem ersten Experiment ein wenig zurück, in beiden Präsentationsformaten zeigten sich aber noch signifikante Unterschiede in den Antwortzeiten der einzelnen Elemente.

Den nach wie vor vorhandenen Einfluss visueller Charakteristika der Items insbesondere bei bildlicher Darstellung dokumentiert hier die hohe Korrelation der Rangpositionen zwischen den beiden Experimenten. Icons, die im ersten Experiment schnell gefunden wurden, hatten auch im zweiten Experiment einen Zeitvorteil; besonders schwierige Items behielten diesen Status ebenfalls bei. Das deutet auf eine ähnliche Suchstrategie hin, die von Gedächtnisfaktoren wenig beeinflusst ist. Bei verbalen Funktionsbeschreibungen war diese Relation nicht so eindeutig, hier haben mehr Elemente ihren Rangplatz verändert. Allerdings war immer noch eine mittlere Korrelation nachweisbar.

5.2.5.5 Schlussfolgerungen

Eine räumlich konstante Anordnung der Items hat in dem vorliegenden Experiment zwar zu allgemein verkürzten Reaktionszeiten, nicht jedoch zu einer grundlegenden Veränderung im Suchverhalten geführt. Entgegen der zu Grunde liegenden Hypothese dieser Untersuchung, blieb der Vorteil für Icons unverändert erhalten. Nachfolgend sollen mögliche Ursachen diskutiert werden.

Vorweg ist erwähnenswert, dass auch in der Studie von Green und Barnard (1990) die ortsfeste Anordnung von Elementen den Unterschied zwischen abstrakten und konkreten Icons nur teilweise kompensieren konnte. Sie beobachteten eine deutliche Abnahme des Unterschieds mit zunehmender Anzahl ortsfester Items. Aber auch wenn alle zwölf Items fixiert waren, bestand noch ein Unterschied zwischen den Gestaltungsvarianten. Dabei sind die Auswirkungen verschiedener Iconvarianten wahrscheinlich ähnlicher als zwischen Icons und Text. Wenn also schon zwischen verschiedenen Iconvarianten ein Unterschied bei ortsfester Anordnung bestehen bleibt, ist dies für Icons und Text ebenfalls zumindest nicht unwahrscheinlich.

Die Versuchspersonen im vorliegenden Experiment waren über die ortsfeste Anordnung der Items informiert. Ein denkbarer Einwand ist, dass es einer gewisser geistigen Anstrengung bedarf, um den Bildschirmaufbau zu lernen und dass dies zum

Erfüllen der Aufgabe nicht unbedingt notwendig ist. Hinzu kommt, dass zwar Ortsinformationen von realen Objekten im Allgemeinen gut behalten werden, aber in experimentellen Studien zu Gedächtnisleistungen bei Ortsinformationen der Anzeigen für Instrumente gezeigt werden konnte, dass dieses Erinnern relativ schwierig und fehleranfällig ist (vgl. Kebeck et al., 1997). Unter diesen Umständen hätten die Probanden doch primär aufgrund der visuellen Eigenschaften der Stimuli gesucht. Da sie dann von den Ortsinformationen kaum Gebrauch gemacht hätten, konnten die Suchzeiten für verbale Beschriftungen nicht reduziert werden. Diese Erklärung kann jedoch nicht begründen, wie die deutliche allgemeine Verkürzung der Reaktionszeiten gegenüber Experiment 5 zu Stande kam.

Wahrscheinlicher ist hingegen, dass in beiden Darstellungsbedingungen durchaus ein Gedächtniseffekt zum Tragen gekommen ist, die Icons aber gerade bei größeren Suchmengen mehr davon profitieren konnten. Dadurch, dass wie in Experiment 5 gezeigt Icons schneller wahrgenommen werden können, können sie auch eher bei einem vorbeistreichenden Blickkontakt erkannt werden. Verbale Beschriftungen hingegen können nicht so einfach identifiziert werden. Auf der Suche nach dem Zielitem bestand bei Icons also eher die Chance, bei der Prüfung (und dem Verwerfen) eines Non-Targets dieses semantisch zu verarbeiten und sich dessen Position zu merken. Nach mehreren Aufgaben wurde die Position vieler Icons auf diesem Weg wiederholt geübt und mit Icons in der Nachbarschaft in Verbindung gebracht. Bei Text hingegen wird der Prüfungsvorgang bei einem Non-Target schon früh abgebrochen, sobald der Wortanfang nicht mit dem Ziel übereinstimmt. Bei den sehr kurzen Betrachtungszeiten dieser Prüfung wird bei Text vermutlich keine vollständige Repräsentation des Non-Targets aufgebaut. Es kommt zu einer deutlich schlechteren Verknüpfung von Ort, Inhalt und Umgebung der Items.

Wenn diese Vermutung zutrifft, sollten besonders kurze textuelle Elemente zu kürzeren Suchzeiten führen, weil diese ähnlich der Icons bei der visuellen Suche eher gestatten, ihre Position „nebenbei“ zu enkodieren. Ein Blick auf Tabelle 11 (S.223) stützt diese Interpretation. Im Gegensatz zum ersten Experiment befinden sich alle vier der sehr kurzen Wörter (Start, Stop, SMS, Hilfe) auf den ersten vier Rangpositionen. In diesem Zusammenhang erklärt sich auch, warum die Rangpositionen der Icons mit denen aus Experiment 5 sehr hoch korrelieren, während es bei den Texten eine Veränderung in der Reihenfolge gibt, die nur zu mittleren Korrelationen geführt hat.

Fasst man die Erkenntnisse aus Experiment 5 und 6 zusammen, liegt die Schlussfolgerung nahe, dass Icons tatsächlich einen deutlichen Zeitvorteil für die visuell-manuelle Interaktion im Fahrzeug bringen können. Wie bereits wiederholt betont wurde, gilt dies selbstverständlich nur, wenn die Symbole bekannt sind. Durch die Verbreitung von normierten Symbolen kann das gewährleistet werden. Als Konsequenz hätte die extensive Verwendung von textuellen Funktionskennzeichnungen eine

nachteilige Wirkung, wenn es um das schnelle Auffinden von Funktionen geht. Bevor diese Schlussfolgerung gezogen wird, soll in Experiment 7 neben einer einfachen Ortskonstanz ein weiterer Aspekt ergonomischer Gestaltung in den Vergleich von Icons und Text einfließen.

5.2.6 Experiment 7: Ortsfeste und inhaltlich gruppierte Items

In Experiment 5, der erweiterten Replikation von Arend et al. (1987), wurden vor jeder Aufgabe die Items auf dem Bildschirm zufällig angeordnet. Der experimentelle Zweck war eine Forcierung der visuellen Suche. Übertragen auf eine reale Nutzungssituation bildet diese Manipulation den Fall eines Nutzers ab, der zwar die Symbole, aber nicht den Bildschirmaufbau kennt. Experiment 6 hingegen zieht in Betracht, dass die Anordnung der Elemente auf dem Bildschirm von den Nutzern gelernt werden kann. Wie bei Green und Barnard (1990) war jedoch das Layout der Elemente nicht sinnvoll aufgebaut. Vielmehr wurde zu Beginn des Experiments eine einmalige zufällige Anordnung bestimmt, die dann konstant blieb. Funktionale oder semantische Zusammenhänge zwischen den Items wurden dabei nicht berücksichtigt. Dies repräsentiert in etwa den Eindruck, den ein Nutzer von einem System gewinnt, mit dem er selten interagiert und in dem er deshalb funktionale Zusammenhänge nicht als Orientierungshilfe nutzen kann.

In der Literatur zu nutzerzentrierter Gestaltung sind bereits seit langem detaillierte Gestaltungsrichtlinien zum Aufbau grafischer Bedienoberflächen etabliert. Shneiderman (1987, S.111) listet einige dieser Grundsätze auf. Eines der Prinzipien, konsistentes Layout zu verwenden, wurde bereits in Experiment 6 umgesetzt. Ein weiteres wichtiges Prinzip hingegen blieb unbeachtet: Die semantische oder funktionale Gruppierung von Items. Wie bereits in Abschnitt 5.2.3 ausgeführt kamen sowohl in den Studien von Arend und Kollegen (1987) als auch in der von Green und Barnard (1990) Anordnungen auf dem Bildschirm vor, die semantisch zusammengehörige Funktionen weit voneinander separiert hatten. Gleiches trifft auf die Experimente 5 und 6 zu. Durch diese Maßnahme konnte visuelle Suche erzwungen werden bzw. unterschiedliche Nutzererfahrung simuliert werden. Eine Verallgemeinerbarkeit der Ergebnisse ist jedoch schwierig, da jede Form des nutzerzentrierten Designs von Mensch-Maschine-Schnittstellen neben einer ortskonstanten Anordnung auch eine Gruppierung von verwandten Inhalten vornehmen wird. Selbstverständlich kommen in der realen Anwendung noch weitere Gestaltungselemente hinzu, wie farbliche Hervorhebungen, visuelle Strukturierungen durch Linien, etc. Aber die Prinzipien Ortskonstanz und Gruppierung bilden den kleinsten gemeinsamen Nenner, somit einen Minimalanspruch. In Experiment 7 soll deshalb erörtert werden, wie sich eine zusätzliche inhaltliche Gruppierung der Items auf den Unterschied zwischen Icons und Text auswirkt. Erst wenn die Frage

geklärt ist, ob sich der bisher gefundene Vorteil für Icons auch dann noch erhält, kann eine generelle Empfehlung für den vorliegenden Gegenstand abgeleitet werden.

5.2.6.1 Methode

An Experiment 7 nahmen ebenfalls zwölf neue Versuchspersonen (drei Frauen, neun Männer) teil, sie waren zwischen 20 und 29 Jahre alt (Mittelwert: 24 Jahre). Für ihre Teilnahme erhielten sie eine angemessene Aufwandsentschädigung. Alle Versuchsteilnehmer sprachen Deutsch als Muttersprache. Wie in den Experimenten zuvor wurde darauf geachtet, dass normale oder voll korrigierte Sehschärfe vorlag. Die Versuchspersonen gaben regelmäßige Computernutzung an. Sie waren nicht über die Hypothesen des Experiments informiert.

Das Vorgehen (Versuchsmaterial und -stimuli sowie das Design und die Ausbalancierung der Bedingungsabfolgen) war bis auf nachfolgend beschriebene Ausnahme identisch mit den vorangegangenen Experimenten. In Experiment 7 gab es keine völlig zufällige Anordnung der Items auf der Benutzeroberfläche. Alle Funktionen, die in der Suchmatrix dargestellt werden, lassen sich einer von vier Gruppen zuordnen (vgl. Ausführungen zum Stimulus-Material, Abschnitt 5.2.4.1). Je drei Items gehören zu Musikfunktionen, Kommunikationsfunktionen, Klimafunktionen oder Systemfunktionen. Die Zuordnung existierte bereits in den Experimenten 5 und 6, kam dort aber nicht strukturierend zum Einsatz. Die funktionalen Gruppen blieben im dritten Experiment zeilenweise erhalten, jedoch ohne weitere visuelle Hervorhebung. Die zeilenweise Gruppenzugehörigkeit für Icons und Text entspricht der in Abbildung 34 (S.205) mit einem entscheidenden Unterschied: Damit auch in diesem Experiment die Größe der Suchmenge und das Präsentationsformat in gleicher Art wie in den vorangegangenen Experimenten in einem Within-Subjects Design variiert werden können, wird das Layout nicht völlig unverändert über die Bedingungen hinweg erhalten. Analog zu Experiment 6 gibt es einen (eingeschränkt) zufälligen Aufbau jeweils bei jeder ersten Aufgabe des ersten experimentellen Blocks einer neuen Bedingung, d.h. bei einem Wechsel der Suchmenge oder des Präsentationsformats. Dabei wird die Zeilennummer einer funktionellen Gruppe zufällig ausgewählt. Zusätzlich wird die Position der Items innerhalb einer Zeile zufällig festgelegt. Dieser Aufbau bleibt für alle Aufgaben einer Bedingung konstant. Auf diese Weise existiert zwischen den experimentellen Bedingungen noch genug Variation, damit im ersten Block Lernen stattfindet und keine Störeffekte durch spezielle Abfolgen der Elemente das Experiment einschränken. Darüber hinaus minimiert die Ausbalancierung der Abfolge der Versuchsbedingungen einen möglichen Einfluss von Transfereffekten. Die Versuchsteilnehmer wurden über die Existenz der Gruppen nur kurz mündlich bei der Erklärung der Items informiert. Ferner war ihnen bekannt, dass die Anordnung der Elemente konstant blieb, bis sie über eine Änderung unterrichtet werden (analog zu Experiment 6).

5.2.6.2 Hypothesen

Durch die stärkere Einbindung von Gedächtnisaspekten sollte die Suche nach dem Zielitem eine stärkere Unabhängigkeit von visuellen Charakteristika der Stimuli erlangen. Das sollte zu einer starken Reduktion der Unterschiede in den Reaktionszeiten von Icons und textuellen Beschriftungen führen. Insbesondere sollten sich ähnliche Steigungen und Lernverläufe beobachten lassen. Deutliche Unterschiede im Suchmuster zu den vorangegangenen Experimenten sind in der Itemanalyse zu erwarten, wenn tatsächlich eine Verschiebung der Primärinformation über das Ziel auf die Gedächtnisebene stattfindet.

5.2.6.3 Ergebnisse

Reaktionszeiten und Fehler

Die Analyse aller Ergebnisse erfolgt wie in den Experimenten 5 und 6. Abbildung 40 stellt die Reaktionszeiten auf Icons und Text in Abhängigkeit der Größe der Suchmenge dar. Während die Größe der Suchmenge die Suchzeiten nach wie vor beeinflusst, scheint sich das Präsentationsformat kaum noch auszuwirken.

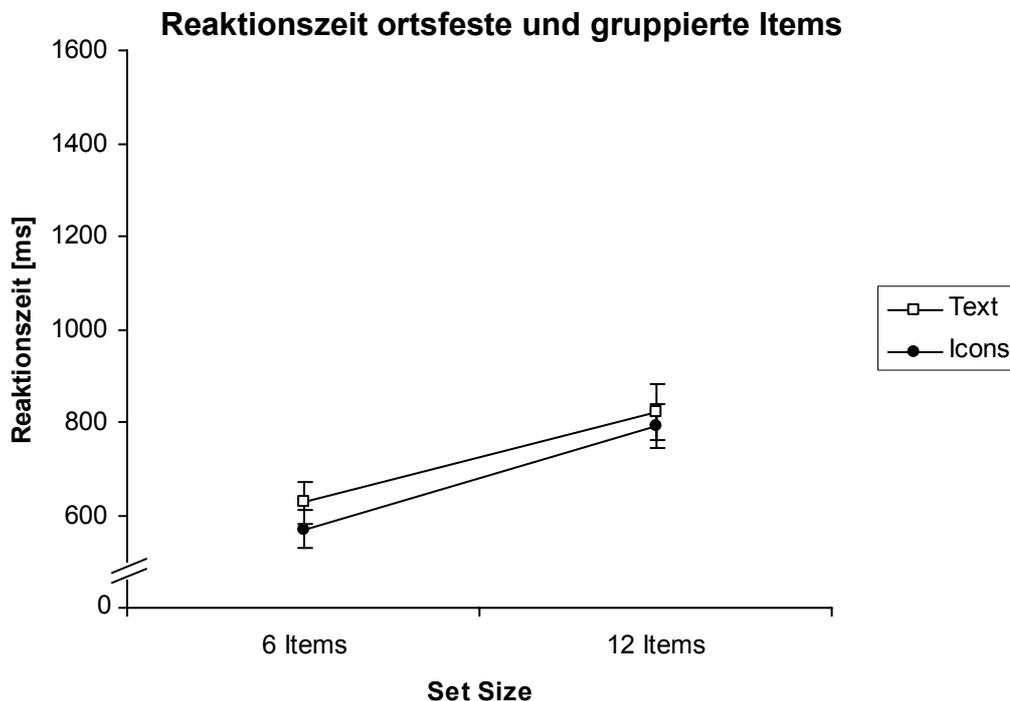


Abbildung 40: Mittlere Reaktionszeiten von Text und Icons als Funktion der Größe der Suchmenge bei ortskonstanter und gruppierter Anordnung der Items.

Bei textueller Funktionsdarstellung wurden Reaktionszeiten von 625 ms ($SE = 45.83$) bei 6 Items und 823 ms ($SE = 60.10$) bei 12 Items registriert, während Icons zu Zeiten von 570 ms ($SE = 40.27$) bei der kleinen Suchmenge und 791 ms

($SE = 46.52$) bei der großen Suchmenge führten. Bereits auf den ersten Blick fällt auf, dass sich das Muster sehr angepasst hat. Die statistische Analyse bestätigt den Eindruck. Wurden die Elemente als Icons präsentiert, führten sie zu einer durchschnittlichen Reaktionszeit von 680 ms ($SE = 37.88$). Nur wenig unterschiedlich ist das Resultat bei textueller Darstellung: 725 ms ($SE = 42.31$). Dieser geringe Unterschied ist nicht statistisch belastbar, $F(1,11) = 1.54$, $p = .24$, $p_{\text{rep}} = .69$. Ebenso verhält es sich bei den Steigungen. Icons bedingen eine Steigung von 33 ms/Item. Die von textuellen Elementen produzierte Steigung ist mit 37 ms/Item sehr ähnlich, die Interaktion Präsentationsformat \times Suchmenge ist nicht signifikant ($F(1,11) = 0.13$, $p = .73$, $p_{\text{rep}} = .33$). Betrachtet man die Größe der Suchmenge, suchten die Probanden bei 6 Items 598 ms ($SE = 30.40$). Bei 12 Items ergab sich eine Reaktionszeit von 807 ms ($SE = 37.32$). Dieser Haupteffekt ist signifikant, $F(1,11) = 47.87$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .81$.

Die Fehlerrate in diesem Experiment war äußerst niedrig und wird deshalb nur deskriptiv berichtet. Nur in maximal einem Prozent der Fälle wurde sowohl bei Icons (6 Items: 0.7%; 12 Items 1%) als auch bei Text (6 Items: 1%; 12 Items: 0.2%) das falsche Element ausgewählt.

Vergleich der Reaktionszeiten über alle Experimente

Die Reaktionszeitdaten der Experimente 5 bis 7 wurden weiter mit einer dreifaktoriellen Varianzanalyse mit den Within-Faktoren Präsentationsformat (Icons, Text) und Größe der Suchmenge (6, 12) und dem Between-Faktor Experiment (Experiment 5, 6, 7) analysiert. Abbildung 41 fasst die Ergebnisse der drei Experimente grafisch zusammen.

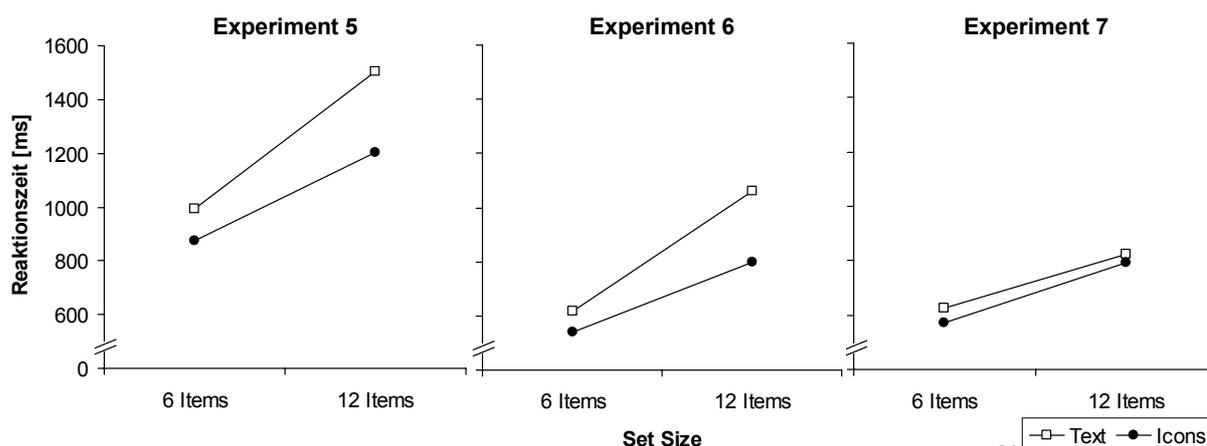


Abbildung 41: Vergleichende Darstellung der mittleren Reaktionszeiten von Text und Icons als Funktion der Größe der Suchmenge (vgl. Abbildung 36, Abbildung 38, Abbildung 40).

Es ist gut zu erkennen, dass in den ersten beiden Experimenten die Steigungen von Icons und Text deutlich unterschiedlich sind. In Experiment 7 hingegen verlaufen die

Geraden parallel. Zusätzlich nimmt über die Experimente die Steigung ab, der Abstand zwischen den Geraden verringert sich dabei. Analysiert man die Daten im Detail, zeigt die Varianzanalyse eine signifikante Interaktion zwischen Präsentationsformat, Größe der Suchmenge und Experiment: Der Unterschied durch das Präsentationsformat in den Suchmengen wird über die drei Experimente hinweg verändert ($F(2,33) = 3.24$, $p = .05$, $p_{\text{rep}} = .87$, $\eta_p^2 = .16$). Dieser Effekt kommt erst durch Experiment 7 zu Stande, da sich diese Interaktion bei der Betrachtung der Experimente 5 und 6 nicht als signifikant herausgestellt hat (vgl. die Reaktionszeit-Analyse in Abschnitt 5.2.5.3). Faszinierend ist, dass sich zwischen Experiment 6 und 7 faktisch keine Veränderung in der Sucheffizienz der Icons ergeben hat (43 ms/Item, bzw. 37 ms/Item). Durch die inhaltliche Gruppierung konnte jedoch die Steigung bei Text mehr als halbiert werden (74 ms/Item, bzw. 33 ms/Item).

Alle Haupteffekte sind signifikant, müssen aber vor dem Hintergrund der gefundenen Interaktionen betrachtet werden. Der Haupteffekt für Experiment reflektiert die allgemeine Abnahme der Reaktionszeiten ($F(2,33) = 28.36$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .63$). Die Zeiten sind gemittelt über Suchmengen und Formate in Experiment 5 am längsten, in Experiment 6 kürzer und in Experiment 7 am kürzesten. Bei 12-elementigen Suchmengen lassen sich allgemein längere Reaktionszeiten nachweisen als bei 6-elementigen Mengen (Größe der Suchmenge, $F(1,33) = 238.18$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .88$). Insgesamt wirkt sich auch das Präsentationsformat aus: Texte sind generell etwas langsamer als Icons ($F(1,33) = 34.10$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .51$).

Die signifikante Interaktion Präsentationsformat \times Suchmenge weist darauf hin, dass die Art der Darstellung sich auf die Reaktionszeiten in verschiedenen Suchmengen unterschiedlich auswirkt. Bei Text wird der Unterschied zu Icons mit zunehmender Suchmenge größer ($F(1,33) = 8.96$, $p < .01$, $p_{\text{rep}} = .97$, $\eta_p^2 = .21$). Dieser Effekt wurde ausschließlich von den Experimenten 5 und 6 erzeugt.

Das Präsentationsformat wirkt sich weiterhin über die Experimente unterschiedlich aus. In Experiment 5 und 6 existieren noch große Unterschiede, in Experiment 7 nicht mehr. Auch diese Interaktion ist signifikant (Präsentationsformat \times Experiment: $F(2,33) = 4.19$, $p = .02$, $p_{\text{rep}} = .92$, $\eta_p^2 = .20$). Wie sich die Größe der Suchmenge auswirkt, hängt ebenfalls vom Experiment ab. Von Experiment 5 bis 7 sind signifikant abnehmende Unterschiede zwischen 6 und 12-elementigen Suchmengen zu verzeichnen (Suchmenge \times Experiment: $F(2,33) = 8.71$, $p = .001$, $p_{\text{rep}} = .99$, $\eta_p^2 = .35$).

Lerneffekte

Wie in den vorangegangenen Experimenten wurden Übungseffekte durch die Analyse der mittleren Reaktionszeiten für Durchgänge mit 12 Items für Icons und Text berechnet. Von Block 1 nach Block 3 ergaben sich im Mittel die folgenden Reaktionszeiten (Standardfehler in Klammern) bei textueller Darstellung: 1053 ms (99.24), 768 ms (59.11) und 650 ms (51.12). Die korrespondierenden Zeiten bei Icons waren 1019 ms (76.03), 723 ms (44.34) und 630 ms (48.90). Auch hier zeigt sich wie im zweiten Experiment eine klare Abnahme der Reaktionszeiten über den Verlauf des Experiments, wie Abbildung 42 zu entnehmen ist. Ein deutlicher Unterschied fällt in der Betrachtung der Effekte für das Präsentationsformat ins Auge: In keinem der Blöcke lässt sich eine bedeutsame Differenz in den Zeiten durch die verschiedenen Darstellungsweisen ausmachen. Die statistische Analyse mit einer zweifaktoriellen Varianzanalyse zeigt entsprechend nur einen signifikanten Haupteffekt für Block: Die Abnahme der Reaktionszeiten sinkt von 1036 ms ($SE = 61.24$) in Block 1 über 745 ms ($SE = 36.44$) in Block 2 auf 640 ms ($SE = 34.66$) in Block 3 ab, $F(2,22) = 43.42$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .80$. Ein allgemeiner Vorteil für Icons ist nicht nachweisbar (Icons: $M = 790$ ms, $SE = 49.91$; Text: $M = 823$ ms, $SE = 42.98$), $F(1,11) = 0.331$, $p = .58$, $p_{\text{rep}} = .44$. Die Interaktion Präsentationsformat \times Block ist nicht signifikant ($F(2,22) = 0.03$, $p = .97$, $p_{\text{rep}} = .01$).

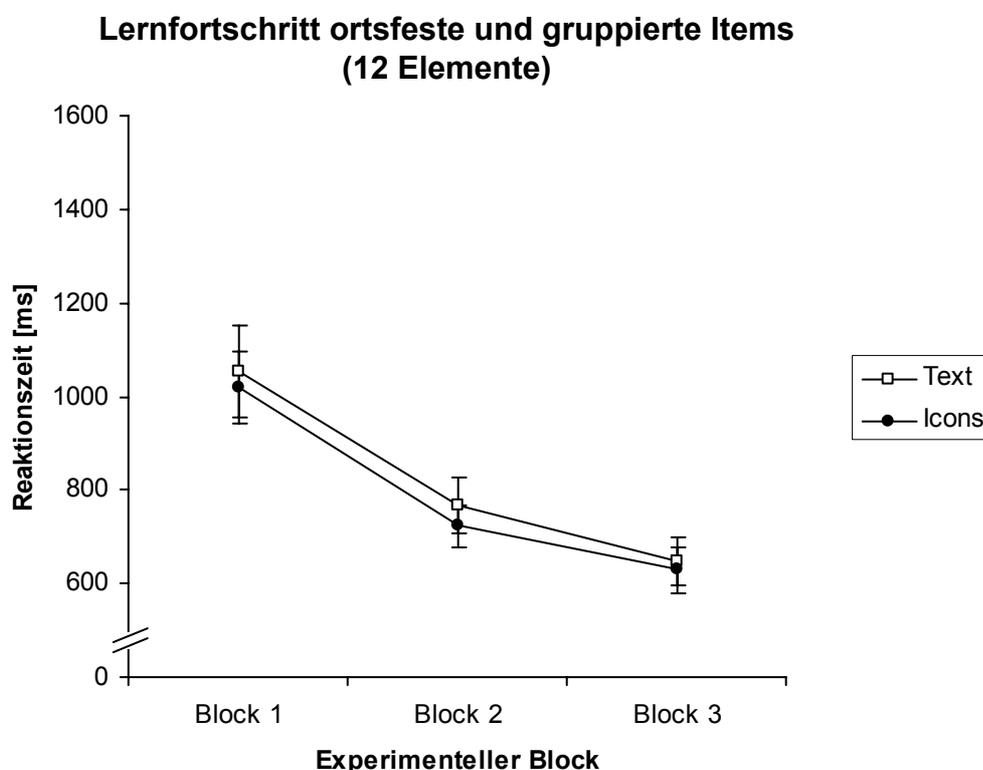


Abbildung 42: Lernfortschritt als Veränderung der mittleren Reaktionszeiten über die Zeit bei gruppierter und ortsfester Anordnung der Items.

Vergleich der Lerneffekte über alle Experimente

Für einen Vergleich der Lerneffekte über die Experimente hinweg wurde erneut eine dreifaktorielle Varianzanalyse mit den Within-Faktoren Präsentationsformat (Icons, Text), Block (1-3) und Experiment (5, 6, 7; Between-Faktor) berechnet. Die Reaktionszeiten in den einzelnen Blöcken haben sich über die Experimente hinweg deutlich verändert (vgl. Abbildung 43). Während in Experiment 5 kein Lernfortschritt vorlag, war dieser in den Experimenten 6 und 7 durchaus vorhanden. Dies wird durch eine signifikante Block \times Experiment Interaktion bestätigt ($F(4,66) = 9.05$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .35$). Weiter lässt sich ein allgemeiner Unterschied zwischen den Experimenten konstatieren (Haupteffekt Experiment, $F(2,33) = 27.21$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .62$). Mit zunehmender Übung werden die Reaktionszeiten kürzer, wobei dieser Effekt nur durch die Experimente 6 und 7 erzeugt wurde (Haupteffekt Block: $F(2,66) = 47.33$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .59$). Der generell nachweisbare Unterschied zwischen Icons und Text hingegen wurde von den Experimenten 5 und 6 getragen (Haupteffekt Präsentationsformat: $F(1,33) = 31.70$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .49$).

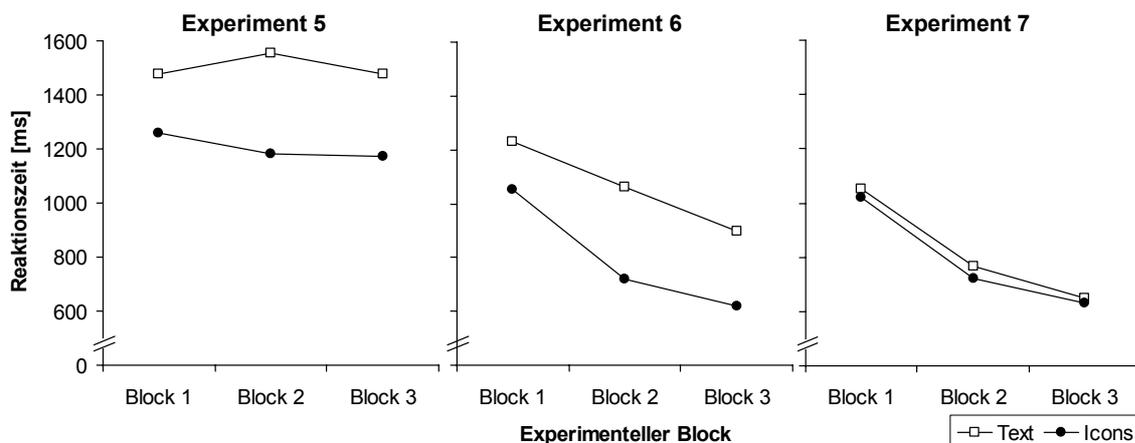


Abbildung 43: Vergleich des Lernfortschritts (mittlere Reaktionszeiten) über die Experimente (vgl. Abbildung 37, Abbildung 39, Abbildung 42).

Im Vergleich von Experiment 5 und 6 beim Lernfortschritt (siehe 5.2.5.3) wurde die Interaktion Präsentationsformat \times Experiment nicht signifikant, da der Unterschied zwischen Icons und Labels erhalten blieb. In Experiment 7 hingegen liegen die Lernkurven fast aufeinander, so dass bei der Betrachtung über alle Experimente eine signifikante Interaktion Präsentationsformat \times Experiment festzustellen ist, $F(2,33) = 5.58$, $p < .01$, $p_{\text{rep}} = .96$, $\eta_p^2 = .25$. Das Ausbleiben einer Veränderung der Reaktionszeiten über die Blöcke in Experiment 5 bei gleichzeitig deutlicher Abnahme der Zeiten in Experiment 6 und 7 spiegelt sich in einer signifikanten Interaktion Block \times Experiment wider ($F(4,66) = 9.05$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .35$). Die Interaktion Präsentations-

format \times Block über alle drei Experimente erweist sich als nicht signifikant ($F(2,66) = 1.72, p = .19, p_{\text{rep}} = .73$). Ebenso kann keine Interaktion höherer Ordnung nachgewiesen werden ($F(4,66) = 0.37, p = .83, p_{\text{rep}} = .25$).

Itemanalyse

Die Itemanalyse soll in Experiment 7 klären, inwiefern eine zur Ortskonstanz hinzukommende inhaltliche Gruppierung der Items den Unterschied in der Reaktionszeit bei einzelnen Elementen innerhalb der Präsentationsformate weiter reduziert. Hierfür wurden wieder einfaktorielle Varianzanalysen über die Mittelwerte der Antwortzeiten gerechnet. Diese Mittelwerte sind aufgetrennt nach Format in Tabelle 12 dargestellt.

Tabelle 12: Rangordnung der Items (Mittelwert der Reaktionszeit) bei ortsfestem und gruppiertem Layout für Icons und textuelle Darstellung.

Rangplatz	Icons	Reaktionszeit [ms] (Min; Max)	Text	Reaktionszeit [ms] (Min; Max)
1		499 (337; 760)	Start	631 (335; 859)
2		576 (396; 897)	SMS	679 (300; 1000)
3		649 (365; 1071)	Sitzheizung	700 (319; 1274)
4		661 (316; 949)	Stop	711 (303; 1047)
5		674 (296; 1336)	Zurück	713 (317; 1201)
6		692 (413; 941)	Einstellungen	719 (387; 1103)
7		707 (413; 926)	Lautstärke	721 (354; 1357)
8		712 (376; 959)	Adressbuch	752 (398; 1072)
9		713 (402; 1133)	Klimaanlage	757 (346; 1330)
10		724 (359; 1341)	Telefon	761 (504; 1066)
11		744 (352; 1211)	Hilfe	773 (288; 1191)
12		810 (446; 1048)	Lüftung	780 (401; 1244)

Sofort fallen die deutlich veränderten Suchzeiten innerhalb der Gruppe der Texte auf. Bei einem Mittelwert von 725 ms ergeben sich für Text eine Standardabweichung von 43 ms und eine Spannweite von 149 ms, eine deutliche Reduktion gegenüber Experiment 6 und eine deutlich geringere Streuung als bei den Icons. Für Icons (Mittelwert: 680 ms) werden eine Standardabweichung von 80 ms und eine Spannweite von 311 ms errechnet. Aufgrund der geringen Mittelwertsunterschiede in der Gruppe

der textuellen Schaltflächen ist nicht überraschend, dass eine einfaktorielle Varianzanalyse über die 12 Items keinen signifikanten Unterschied anzeigt ($F(11,121) = 0.65$, $p = .79$, $p_{\text{rep}} = .29$). Die Reaktionszeiten in der Gruppe der Icons hingegen unterscheiden sich nach wie vor: $F(11,121) = 3.02$, $p = .001$, $p_{\text{rep}} = .99$, $\eta_p^2 = .22$.

Bereits im Anschluss an Experiment 6 wurde ein erster Vergleich der Rangpositionen der Items innerhalb der jeweiligen Sets über die Experimente durchgeführt. Bei Texten wurde eine mittlere, signifikante, bei Icons sogar eine hohe, signifikante Korrelation festgestellt. Bei dem Suchverhalten auf der Ebene einzelner Items gibt es durch Experiment 7 einen Bruch, denn sowohl bei Icons als auch bei Text lassen sich nur noch geringe nichtsignifikante Rangkorrelationen feststellen (siehe Tabelle 13). Offensichtlich war erst die Manipulation in Experiment 7 in der Lage, die Rangordnung der Items neu aufzubauen.

Tabelle 13: Ergebnisse der Korrelationsanalyse (Spearman's Rho) für die Rangordnung der Items innerhalb der Icon- und Text-Sets, basierend auf den Mittelwerten der Reaktionszeiten über die Experimente 5-7.

	Icons		Text	
	E1	E2	E1	E2
Experiment 5 (E1)				
Experiment 6 (E2)	.79**		.62*	
Experiment 7 (E3)	.29	.31	-.10	.36

Anmerkung: * $p < .05$ ($r_{\text{crit}} = .58$), ** $p < .01$ ($r_{\text{crit}} = .71$).

5.2.6.4 Diskussion

Zufällige Anordnungen oder ortsfeste Anordnungen ohne inhaltlich begründetes Layout von Funktionsschaltflächen innerhalb einer Suchmatrix führten in den Experimenten 5 und 6 zu einem klaren Vorteil ikonografischer Funktionsdarstellung gegenüber textueller Beschriftung. Um zu klären, inwieweit der Vorteil für Icons allgemein in üblichen Anwendungsszenarien der Mensch-Maschine-Interaktion Bestand hat, wurde Experiment 7 durchgeführt. Wie in Experiment 6 wurden die Items in der Suchmatrix nach einer initialen Zufallsanordnung konstant gehalten. Im Gegensatz zu Experiment 6 erfolgte die anfängliche Zufallsanordnung in bestimmten Grenzen: Je drei der zwölf Items waren einer inhaltlich funktionalen Gruppe zugeordnet (Musik-, Kommunikations-, Klima- oder Systemfunktionen). Diese Gruppen blieben stets erhalten. Die Versuchsteilnehmer waren nur mündlich über die Existenz der Gruppen informiert und hatten keine weiteren Hilfsmittel zur Strukturierung. Mit Einschränkungen kann dieses Experiment als die Simulation des Nutzungsverhaltens gesehen werden, das sich her-

ausbildet, sobald der Nutzer sich nicht nur vereinzelte Aspekte der grafischen Oberfläche merkt, sondern auch Wissen über Funktionszusammenhänge ausgebildet hat.

Bereits mit der ortskonstanten Anordnung in Experiment 6 ließ sich ein Lerneffekt durch sukzessiv kürzere Reaktionszeiten für Icons und Text beobachten. Dieser Lerneffekt stellte sich auch in Experiment 7 ein. Im Gegensatz zu den Experimenten 5 und 6 ließ sich jedoch kein Vorteil für Icons in den Reaktionszeiten nachweisen. Vielmehr haben sich Icons und Text sowohl bezüglich unterschiedlich großer Suchmengen als auch bezüglich des Lernfortschritts nahezu identisch verhalten.

Reaktionszeiten

Mit der zusätzlichen Einführung der inhaltlichen Gruppierung veränderte sich das Muster der Reaktionszeiten gegenüber den ersten beiden Experimenten grundlegend. In Experiment 7 konnte nur noch ein Einfluss der Größe der Suchmenge festgestellt werden. Die zusätzliche Strukturierung des Materials führte zu einem schwächeren Anteil der visuellen Suche am Ergebnismuster und zu einem stärkeren Anteil von Gedächtnisprozessen. Da der Effekt der Suchmenge über alle experimentellen Blöcke gemittelt betrachtet wird, ist der gefundene Anstieg in der Reaktionszeit bei größeren Suchmengen nicht überraschend. Es steht zu vermuten, dass eine länger andauernde Lernperiode zu einer Eliminierung dieses Unterschieds führen kann. Für eine optimale Gestaltung ist jedoch die Betrachtung der ersten, wenigen Kontakte mit einer grafischen Benutzeroberfläche ebenfalls entscheidend.

Als wichtigste Erkenntnis dieses Experiments ist die Auflösung des Vorteilseffekts für Icons gegenüber verbalen Funktionsbeschreibungen zu sehen. Die Steigungen zwischen Icons und Text waren äußerst ähnlich, ebenso ließ sich kein allgemeiner Vorteil für bildhafte Darstellung mehr nachweisen. Im Vergleich zu den anderen Experimenten zeigt sich klar der Bruch, der sich durch die inhaltliche Gruppierung ergeben hat. Während von Experiment 5 auf 6 zwar die Reaktionszeiten generell durch Gedächtnisleistungen reduziert wurden, blieb ein deutlicher Unterschied zwischen Icons und Text bestehen, der in Experiment 7 auf eindrucksvolle Weise aufgelöst wurde. Interessant ist dabei, dass offensichtlich die Icons weniger von der zusätzlichen Information profitiert haben. Von Experiment 6 auf 7 ergab sich für Icons kaum eine Veränderung. Bildhaftes Material bietet insgesamt durchaus Vorteile für die Gedächtnisleistung, da die Positionsinformation allein schon zu recht guten Leistungen geführt hat. Textuelle Beschriftungen hingegen benötigen weitere Strukturierungen wie die inhaltliche Gruppierung, um Anschluss an diese kurzen Reaktionszeiten halten zu können. Können entsprechende Strukturierungen ermöglicht werden, lässt sich entgegen der in der Literatur weit verbreiteten Meinung kein allgemeiner Vorteil für Icons über Text etablieren.

Lerneffekte

Auch die Analyse der Lerneffekte deutet in diese Richtung. Durch die Gruppierung wurde ein Faktor eingeführt, der den Lernfortschritt von Icons und Text massiv vereinheitlichte. Gegenüber Experiment 5 haben Experiment 6 und 7 zunächst beide zu einer Reduktion der Reaktionszeiten in den einzelnen Blöcken beigetragen. In Experiment 7 ist nicht nur der Haupteffekt für das Präsentationsformat verschwunden. Bereits im ersten experimentellen Block ergaben sich fast identische Reaktionszeiten zwischen Icons und Text. Inhaltliche Gruppierung ist also nicht nur in der Lage, allgemein den Unterschied zwischen Icons und Text aufzulösen, sie ermöglicht auch eine sehr leichte Erlernbarkeit des Layouts der Items. Dieser Aspekt ist hochgradig relevant, denn für die visuelle Gestaltung eines Bordsystems im Fahrzeug müssen kurze Lernzeiten für die grafische Benutzeroberfläche vorausgesetzt werden. Wäre zu Beginn des Experiments noch ein Unterschied zwischen Icons und Text erkennbar gewesen, hätte die Aussage einer Äquivalenz der Gestaltungsformen bei ortsfester Anordnung mit inhaltlicher Gruppierung relativiert werden müssen. So kann sie in vollem Umfang aufrechterhalten werden.

Wie bei den Reaktionszeiten in Bezug auf verschieden große Suchmengen zeigt sich auch bei der Betrachtung über die Zeit, dass in Experiment 7 eine Veränderung gegenüber Experiment 6 nur durch die textuellen Beschriftungen verursacht wurde. Die Lernkurven der Icons sind in Experiment 6 und 7 kaum unterscheidbar, Text hingegen weist eine merkbare Veränderung auf. Auch hier muss einschränkend bemerkt werden, dass die Reaktionszeiten bereits sehr kurz waren und eine weitere Reduktion eventuell nicht möglich war.

Itemanalyse

In der Itemanalyse wird sichtbar, wie sehr sich das Selektionsverhalten durch Gruppierung verändert hat. Die fehlende Korrelation der Rangfolgen der Icons und der verbalen Beschriftungen zusammen mit den bereits dargelegten allgemein gleichartigen Reaktionszeiten der beiden Präsentationsformate deuten darauf hin, dass nicht mehr visuelle Suche sondern Gedächtnisprozesse bestimmend waren.

Unklar hingegen ist, weshalb die Reaktionszeiten der Texte deutlich ähnlicher sind als die der Icons. Die Spannweite der mittleren Reaktionszeiten bei Text ist nur etwa halb so groß. Entsprechend zeigt sich in den varianzanalytischen Befunden, dass sich nur die Icons in ihren Reaktionszeiten signifikant voneinander unterscheiden, bei Text ist das nicht der Fall. Der Rückgriff auf Gedächtnisprozesse könnte dies erklären. Ausgehend von der Gedächtnistheorie von Engelkamp (1991) ist anzunehmen, dass die Verknüpfung der textuellen Items mit dem Kategoriebegriff einfacher möglich sein sollte als eine Verknüpfung der Symbole mit diesem Oberbegriff. Durchaus nahe liegend ist deshalb, dass die Probanden die Suche bei Text dadurch eingeleitet haben,

dass sie zunächst ein Item einer Zeile dahingehend betrachtet haben, ob es zu der gesuchten Funktionskategorie gehört. Hat das Item den Kategorie Cue erfolgreich aktiviert, sind sie in der entsprechenden Zeile geblieben. Ist das Item nicht aus der gesuchten Kategorie, haben sie den Prozess in der nächsten Zeile wiederholt. Basiert die Suche auf einem solchen Prozess, sollte sie relativ unabhängig von den konkreten Beschriftungen funktionieren, was die geringe Streuung erklärt. Bei Icons erscheint diese Strategie aufgrund der schwierigeren kognitiven Verknüpfungsoperation weniger wahrscheinlich. Da jedoch die Korrelation der Rangpositionen mit den Vorgängerexperimenten fehlt, ist eine rein visuelle Suche trotz ähnlicher Reaktionszeiten mit Experiment 6 eher unwahrscheinlich. Vermutlich gab es bei Icons eher eine Mischung aus visuellen Suchanteilen und Gedächtniseinflüssen. Dies ist aber allein auf der Basis der vorliegenden Daten schwierig zu entscheiden. Wenn dem so wäre, hätten eventuell noch kürzere Reaktionszeiten bei den Icons zu beobachten sein müssen. Es ist möglich, dass dieser zusätzliche Effekt aufgrund der ohnehin bereits kurzen Reaktionszeiten für Icons nicht mehr nachweisbar war.

5.2.6.5 Schlussfolgerungen

Experiment 7 zeigt, dass der Vorteil von Icons gegenüber Text bei der visuellen Suche mit einfachen gestalterischen Mitteln durch eine Verlagerung der Anforderungen der Aufgabe an Gedächtnisprozesse aufzulösen ist. Andererseits müssen die Anwender dazu in der Lage sein, funktionale Zusammenhänge und Strukturen in der Bedienoberfläche zu erkennen. Das wiederum kann durchaus Einarbeitungszeit erfordern, während der die visuellen Charakteristika der Items die primäre Quelle der Orientierung darstellen. Wenn angenommen wird, dass eher Gelegenheits- oder Seltennutzer mit dem System arbeiten und bekannte Standard-Icons eingesetzt werden können, ist ein Vorteilseffekt zu erwarten. Gleiches gilt, wenn eine inhaltliche Strukturierung der Oberfläche nicht möglich ist. So ist es bei Softkey-Bedienkonzepten oft schwierig, eine einheitliche Ordnung zu etablieren. Bevor basierend auf den Befunden der Experimente abschließende Schlussfolgerungen gezogen und Gestaltungsrichtlinien abgeleitet werden, soll der mögliche Unterschied von Icons und Text in seiner Bedeutung für die praktische Anwendung durch ein letztes Experiment beleuchtet werden.

5.2.7 Experiment 8: Visuelle Suche bei Kurzdarbietung

In Experiment 5 hat sich gezeigt, dass eine rein visuelle Suche ohne Strukturierungshilfe auf einem Set von Icons oder textuellen Funktionskennzeichnungen deutlich längere Suchzeiten bei den verbalen Kennzeichnungen zur Folge hat. Selbst wenn die Elemente an konstanten Anzeigeorten ausgegeben werden, bleibt dieser Unterschied erhalten (Experiment 6). In beiden Fällen gilt, dass der Vorteil von Icons mit zuneh-

mender Größe der Suchmenge ansteigt. Da in Experiment 7 gezeigt wurde, dass der Vorteil aufgelöst werden kann, scheint das Darstellungsformat ein nicht zu unterschätzender Faktor zu sein. Nachdem selbst bei nur einigermaßen großen Displays mit zwölf Informationsitems, wie sie hier zum Einsatz kamen, bereits ein stabiler Unterschied von 300 ms existiert, soll in einem weiteren Experiment geklärt werden, ob und wie sich das bei der Bedienung eines Bordsystems auswirken könnte.

Als eine der wichtigsten Eigenschaften der grafischen Benutzeroberfläche eines Bordsystems wurde die Unterstützung der Bedienvorgänge mit sehr kurzen Blickzuwendungen identifiziert. In den bisherigen Experimenten wurden die Reaktionszeiten gemessen, um Aufschluss über den Eingabeprozess an sich zu erlangen. Unter der Annahme konstanter Reaktionszeiten lassen diese Untersuchungen den Rückschluss zu, dass textuelle Beschriftungen einen längeren Zeitraum visueller Zuwendung benötigen. Experiment 8 soll durch eine Änderung des Versuchsablaufs diesen Aspekt verifizieren.

Wolfe (1998) unterscheidet zwei Methoden im Standardparadigma zur visuellen Suche. Bei der bekannteren *reaction time method* bleibt die Anzeige sichtbar, bis die Versuchsperson reagiert hat. Die Reaktionszeit wird als Funktion der Größe der Suchmenge analysiert. Darüber hinaus existiert die *accuracy method*, bei der die Richtigkeit der Antworten als abhängiges Maß fungiert. Der Suchstimulus wird nur kurz gezeigt und wird von einer Maske gefolgt, die die visuelle Suche beendet sowie gegebenenfalls die Inspektion von Nachbildern unterbindet. Der zeitliche Abstand zwischen Stimulus und Maske (*stimulus onset asynchrony*, SOA) wird variiert und die Reaktionsgenauigkeit als Funktion dieser Abstände analysiert. In der Grundlagenforschung wird diese Methode häufig verwendet, um durch ultrakurze Darbietung willkürliche Augenbewegungen bei der Suche zu unterbinden. Das zuvor angewandte Search-and-Select Paradigma entspricht im Grundsatz der *reaction time method*. Auch zur *accuracy method* gibt es eine wesensverwandte Form der Untersuchung in der angewandten Forschung, die Okklusionsmethode (siehe Abschnitt 5.1.2 für eine nähere Ausführung zu dieser Methode).

Experiment 8 orientiert sich im Versuchsaufbau an der *accuracy method*. Wie in den vorigen Experimenten wird den Versuchsteilnehmern eine Suchmatrix mit zwölf Items gezeigt, die Größe der Suchmenge bleibt diesmal jedoch unverändert. Um einen optimalen Ergebnistransfer für den zu Grunde liegenden Anwendungsfall Fahrzeugführung zu gewährleisten, werden die SOAs in einem Bereich variiert, der für die Fahrzeugführung als kritisch gilt. Nach Zwahlen und Kollegen (1988) sind Blickabwendungen von der Fahrbahn zwischen 200 und 1200 Millisekunden tolerierbar. Dabei gilt, dass die erste Blickabwendung bis zu 1200 ms betragen darf, die dritte Blickabwendung innerhalb der Bearbeitung einer Aufgabe sollte jedoch unter 900 ms liegen. Nachdem zu berücksichtigen ist, dass zwischen dem Wechsel der Blickorte im

Fahrzeug eine gewisse Totzeit existiert, wird die Grenze für die längste SOA in diesem Experiment etwas niedriger angesetzt. Damit erhalten die Versuchspersonen zwischen 200 und 800 Millisekunden Zeit, ein Zielitem zu identifizieren und dessen Position auf dem Bildschirm zu lokalisieren. Nach dieser Zeit bleibt die visuelle Schablone der Suchmatrix erhalten, die Schaltflächen zeigen aber statt der Icons oder Texte nur noch die Maske. Die Versuchspersonen müssen auf die Schaltfläche drücken, die zuvor das gesuchte Element gezeigt hat. Die korrekte Lokalisierung des Items wird im Standardparadigma selten verlangt, dort muss ausschließlich erkannt werden, ob ein Item unter Distraktoren präsent oder absent war. Für eine Übertragbarkeit der Ergebnisse in Hinblick auf die Blindbedienbarkeit eines Systems ist jedoch entscheidend, dass nicht nur wahrgenommen wird, *dass* ein Element auf dem Bildschirm vorhanden ist, sondern auch *wo* es sich befindet. Das Experiment beschäftigt sich mit dem anwendungsrelevanten Szenario, dass ein Nutzer kurz auf das Display blickt, sich dann visuell wieder der Fahraufgabe zuwendet und ohne weiteren Blickkontakt zu der Anzeigeeinheit z.B. den zugehörigen Softkey betätigt.

Da die Darbietungszeiten mit einer unteren Grenze von 200 ms sehr kurz sind, ist die Bekanntheit des Itemmaterials eine Voraussetzung zur Versuchsdurchführung. Deshalb nahmen an diesem Experiment nur Probanden teil, die bereits Experiment 5 oder 6 absolviert hatten. Sie absolvierten dieses Experiment nach einer längeren Erholungspause im Anschluss an Experiment 5 bzw. 6. Zu diesem Zeitpunkt hatten sie hinreichend oft Gelegenheit, die visuelle Gestalt der Icons und der Beschriftungen zu trainieren. Die zwölf Elemente sind wie in Experiment 5 zufällig angeordnet, um als extreme Abschätzung die Frage beantworten zu können, wie stark der Unterschied zwischen Icons und Text in der Antwortgüte ist, wenn visuelle Suche stattfindet und gleichzeitig kritisch kurze Blickdauern eingehalten werden sollen.

5.2.7.1 Methode

Versuchspersonen

Die Versuchsteilnehmer dieses Experiments setzten sich aus den Probanden der Experimente 5 und 6 zusammen. Experiment 8 wurde immer nach einer Pause von mehreren Minuten im direkten Anschluss durchgeführt. Insgesamt bestand die Stichprobe damit aus 24 Versuchspersonen (14 Frauen, 10 Männer) im Alter zwischen 22 und 36 Jahren (Mittelwert: 27 Jahre).

Stimuli

Versuchsaufbau und Stimuli waren identisch zu den vorangegangenen Experimenten. Der grundlegende Unterschied bestand in der Maskierung der Elemente nach einer Darbietungsdauer (SOA) von 200, 400, 600 oder 800 ms. Als Maske für textuelle Elemente diente eine Reihe mehrerer „X“-Buchstaben. Icons wurden durch eine Flä-

che mit weißem Rauschen (zweidimensionales graustufiges Rauschsignal) maskiert. Abbildung 44 zeigt beide Maskierungsarten. Die Items in der Suchmatrix wurden nach jeder Aufgabe neu zufällig angeordnet.

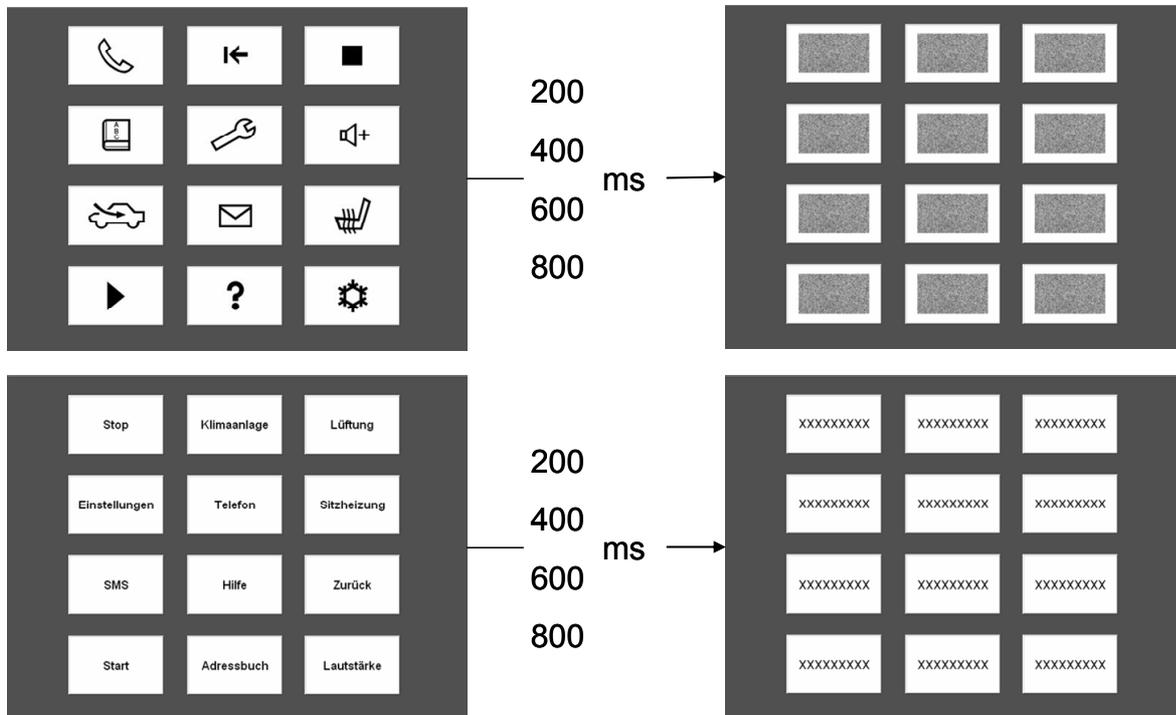


Abbildung 44: Suchmatrix mit Schaltflächen und Maskierung der Stimuli nach unterschiedlicher SOA (obere Hälfte: Icons; untere Hälfte: Text).

Versuchsdesign

Ein 2×4 faktorielles Within-Subjects Design kam in dem Experiment zum Einsatz. Jeder Versuchsteilnehmer bearbeitete Search-and-Select Aufgaben mit unterschiedlichem Präsentationsformat (Stufen: Icons, Text) und einer variierenden Anzeigelänge der Suchmatrix, bevor die Maskierungsmatrix diese ersetzte (SOA, Stufen: 200, 400, 600, 800 ms).

Das Präsentationsformat blieb konstant, bis alle vier SOA Stufen absolviert waren. Für jede Stufe der *stimulus onset asynchrony* wurden einmal alle zwölf Aufgaben bearbeitet, d.h. jedes Item musste pro Anzeigelänge einmal gefunden werden. Die Reihenfolge der Aufgaben war zufällig. Wenn alle Aufgaben bearbeitet waren, änderte sich die SOA für die nächsten zwölf Aufgaben. Die Abfolge der SOA Stufen war ebenfalls zufällig. Die Hälfte der Versuchspersonen begann mit Icons, die andere Hälfte mit Text (vollständig ausbalanciert). Als abhängige Variable diente die Fehler-rate.

Prozedur

Zu Beginn des Experiments wurde den Versuchspersonen der Wechsel in der Aufgabe erläutert. Während sie im Experiment zuvor so schnell wie möglich reagieren sollten,

mussten sie jetzt versuchen, die visuelle Szene so gut es geht zu erfassen und das gesuchte Item in der verfügbaren Anzeigezeit zu lokalisieren. Sobald sie das Item erkannt hatten, sollten sie es durch Antippen der entsprechenden Position in der maskierten Matrix auswählen. Sie wurden darauf aufmerksam gemacht, dass die Anzeigezeiten teilweise sehr kurz sind. Deshalb wurden sie ermutigt, dass sie gegebenenfalls raten sollten, wenn sie ein Item nicht sicher erkennen konnten. Ferner wurde betont, dass die Reaktionszeiten irrelevant waren und die Ressourcen darauf konzentriert werden sollten, das Zielitem visuell zu suchen. Die Teilnehmer wurden informiert, dass die Anordnung der Items nach jeder Aufgabe zufällig erfolgt. Dieser Punkt wurde insbesondere gegenüber den Probanden aus Experiment 6 betont, die zuvor mit ortskonstanten Layouts konfrontiert waren.

In gleicher Weise wie bei den Reaktionszeitexperimenten wurde jede Aufgabe von der auditiven Ansage eingeleitet. Wenn die Versuchsteilnehmer die Aufgabe verstanden und sich das zu suchende Element vergegenwärtigt hatten, sollten sie die „OK“-Schaltfläche betätigen. Anschließend erschien mittig für 1.5 Sekunden ein Fixationskreuz, das die Probanden fokussieren sollten, um von diesem Punkt die Suche zu starten. Nachfolgend erschien die 4×3 Suchmatrix und blieb für das vorbestimmte Anzeigeintervall sichtbar, bevor sie durch die Maskierungsmatrix ersetzt wurde. Diejenige virtuelle Schaltfläche musste betätigt werden, die vor der Maske das gesuchte Element gezeigt hat. Bei Berührung des Steuerelements wurde durch eine Veränderung der Farbe der Schaltfläche Rückmeldung über die erfolgte Eingabe gegeben. Im Gegensatz zu den Reaktionszeitexperimenten gab es jedoch anschließend keine Ausgabe des korrekten Zielitems.

Vor dem ersten Durchgang wurden fünf Aufgaben zufällig als Übungsdurchgänge mit einer SOA von 600 ms herausgegriffen. Während und nach den Übungsdurchgängen hatten die Probanden Gelegenheit, Unklarheiten anzusprechen. Die Aufgaben folgten mit einem Leerintervall von 1.5 Sekunden aufeinander (Ende der Rückmeldung bis Beginn der Ansage). Nach je zwölf Aufgaben und damit verbundener Veränderung der SOA wurde durch das Programm eine Unterbrechung von 30 Sekunden vorgegeben. Die Versuchsteilnehmer durften diese Pause gegebenenfalls verlängern. Nach vier mal zwölf Aufgaben, der Hälfte des Experiments, wurde eine Mindestpause von zwei Minuten eingehalten. Danach wurde der bevorstehende Wechsel des Darstellungsformats (Icons/Text) angekündigt. Die Gesamtdauer von Experiment 5 bzw. 6 und Experiment 8 betrug in etwa eine Stunde.

5.2.7.2 Hypothesen

Ausgehend von der Studie von Arend et al. (1987) wird die Hypothese aufgestellt, dass die Iconmatrix schneller abgesucht werden kann als die Textmatrix. Damit sollte insbesondere bei den sehr kurzen Darbietungszeiten (200 und 400 ms) eine geringere

Fehlerrate bei Icondarbietung zu beobachten sein. Bei längerer Darbietungszeit sollte der Abstand zwischen Icons und Text geringer werden. Allerdings ist a priori keine Aussage darüber möglich, inwiefern diese Annäherung bereits im kritischen Zeitbereich bis 800 ms zu erreichen ist. Durch den Versuchsaufbau mit zufälliger Itemanordnung wird die Situation eines ungeübten Nutzers simuliert, der sich auf der grafischen Benutzeroberfläche primär durch visuelle Suche orientiert und der wenig inhaltlich-funktionale Kenntnis des Systems hat. Sollte sich die Antwortgüte von Icons und Text bei höchstens 800 ms bereits eindeutig angenähert haben, könnte kein substantieller Vorteil für bildhaftes Material konstatiert werden.

5.2.7.3 Ergebnisse

Die Analyse fokussiert auf einen Vergleich der Fehlerraten zwischen Icons und textuellen Beschriftungen bei unterschiedlich langen Präsentationszeiten der Suchmatrix. Bei 12 Elementen existierte eine Zufallswahrscheinlichkeit von .08, das Zielitem zufällig zu treffen. Für die verschiedenen Anzeigedauern (SOA 200, 400, 600, 800 ms) wurden die mittleren Fehlerraten getrennt für die Präsentationsformate berechnet. Wie Abbildung 45 zeigt, ist auf allen SOA Stufen ein deutlicher Unterschied zwischen Icons und Text zu erkennen.

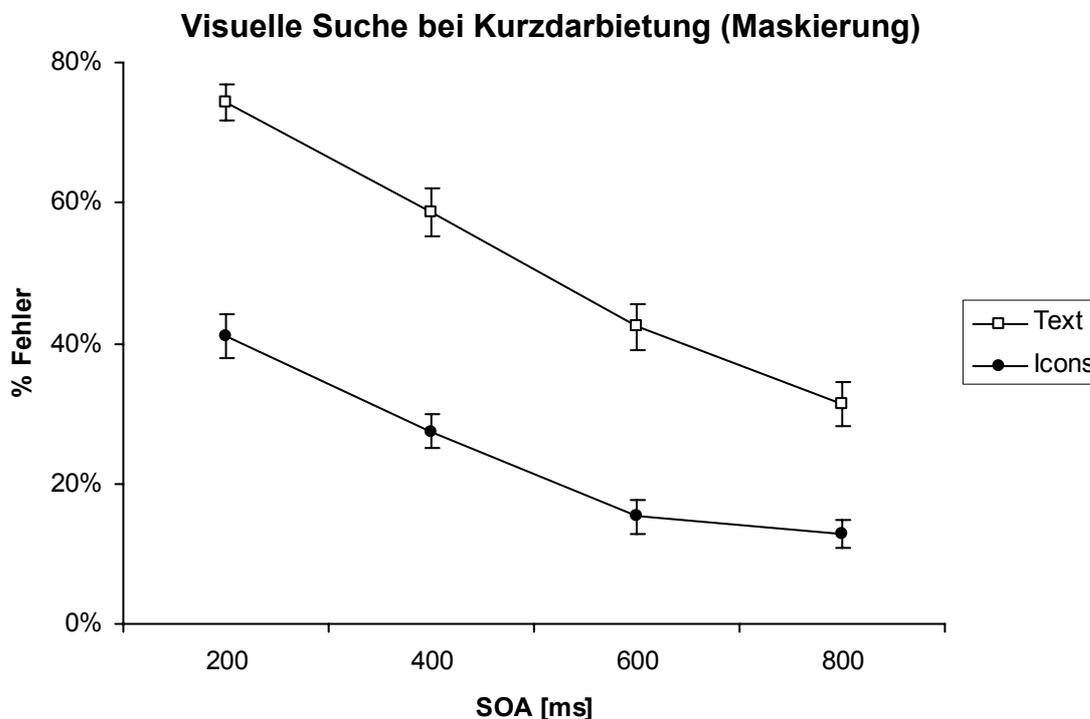


Abbildung 45: Mittlere Fehlerraten (mit Standardfehlern) von Text und Icons als Funktion Zeit zwischen Reizdarbietung und Maskierung.

Waren die Funktionsschaltflächen beschriftet, betrug der Fehleranteil bei den Eingaben bei 200 ms SOA 74.3% ($SE = 2.65$), bei 400 ms SOA 58.7% ($SE = 3.46$), bei

600 ms SOA 42.4% ($SE = 3.25$) und bei der maximalen Zeit zwischen Stimulus und Maskierung von 800 ms SOA 31.3% ($SE = 3.10$). Icons hingegen führten bereits bei der kurzen Anzeigedauer von 200 ms SOA nur zu 41.0% falschen Eingaben ($SE = 3.17$). Diese Fehlerrate reduzierte sich weiter bei 400 ms SOA auf 27.4% ($SE = 2.33$), bei 600 ms SOA auf 15.3% ($SE = 2.49$) und schließlich bei 800 ms SOA auf 12.8% ($SE = 1.94$).

Diese Daten wurden einer zweifaktoriellen Varianzanalyse zugeführt. Bereits durch die Betrachtung der obigen Abbildung wird klar, dass es neben den deutlichen Unterschieden zwischen den Präsentationsformaten auch eine massive Abnahme der Fehleraten mit längerer Anzeigedauer gibt. Wurden Funktionen als Icons dargestellt, machten die Probanden im Mittel 24.1% Fehler ($SE = 1.69$). Bei textueller Darstellung waren es mit 51.6% ($SE = 2.27$) mehr als doppelt so viele. Der Haupteffekt für Präsentationsformat ist entsprechend signifikant: $F(1,23) = 153.35$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .87$. Ebenfalls signifikant ist die Abnahme der Fehlerrate mit zunehmender Anzeigedauer ($F(3,69) = 88.394$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .79$). Bei nur 200 ms Sichtbarkeit der Anzeige wurden im Schnitt 57.6% ($SE = 3.18$) falsche Eingaben getätigt, bei 400 ms 43.1% ($SE = 3.07$), bei 600 ms 28.8% ($SE = 2.83$) und schließlich bei 800 ms 22.1% ($SE = 2.25$). Abbildung 45 lässt weiter die Vermutung zu, dass sich die Reduktion der Fehlerraten durch zunehmende Darbietungslänge bei Text massiver auswirkt als bei Icons. Die Fehlerraten bei Text scheinen steiler abzufallen als bei Icons. Dies wird bestätigt durch die signifikante Präsentationsformat \times SOA Interaktion ($F(3,69) = 2.96$, $p = .038$, $p_{\text{rep}} = .90$, $\eta_p^2 = .11$).

Die Ergebnisgrafik legt ferner nahe, dass die Fehlerrate bei Text linear abnimmt. Bei Icons gibt es eine derart generelle Abnahme jedoch nur bis 600 ms. Zwischen 600 und 800 ms hingegen ist der Rückgang deutlich schwächer ausgeprägt. Es stellt sich die Frage, ob die Abnahme der Fehlerraten bei Icons noch linear ist, oder ob ab 600 ms bereits keine drastische Reduktion mehr zu erwarten ist und eine Asymptote erreicht wurde. Mit Hilfe einer Trendanalyse soll ermittelt werden, ob eine Gerade das Verhältnis zwischen Fehlerrate und SOA bei Icons voll erfassen kann oder eine quadratische Trendkomponente in den Daten nachgewiesen werden kann. Die Trendanalyse wies für Text lineare Trendkomponenten nach: $F(1,23) = 236.41$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .91$. Quadratische Komponenten konnten nicht nachgewiesen werden: $F(1,23) = 0.58$, $p = .46$, $p_{\text{rep}} = .53$. Bei Icons zeigte sich ebenfalls ein linearer Trend: $F(1,23) = 79.87$, $p < .001$, $p_{\text{rep}} > .99$, $\eta_p^2 = .78$. Darüber hinaus konnten jedoch zusätzlich quadratische Trendkomponenten gefunden werden: $F(1,23) = 6.49$, $p = .018$, $p_{\text{rep}} = .93$, $\eta_p^2 = .22$. Kubische Komponenten waren hingegen nicht vorhanden: $F(1,23) = 1.17$, $p = .29$, $p_{\text{rep}} = .65$. Bei Icons wird die Vermutung der visuellen Inspek-

tion belegt: Die Linearität der Abnahme ist unterbrochen. Der signifikante quadratische Trend deutet darauf hin, dass diesbezüglich eine Veränderung in der Abnahmerate stattgefunden hat.

5.2.7.4 Diskussion

Experiment 8 untersuchte die Frage, inwiefern eine visuelle Vorlage bei verschiedenen langen Darbietungszeiten erfolgreich nach einem Zielitem abgesucht werden kann. Die visuelle Vorlage bestand aus einer Suchmatrix mit zwölf Elementen. Bei typischen Kfz-Bordsystemen kann durchaus angenommen werden, dass in etwa diese oder eine höhere Informationsdichte erreicht wird. Basierend auf den Empfehlungen von Zwahlen et al. (1988) zu tolerierbaren Blickabwendungsdauern von der Fahrbahn wurden die Darbietungsdauern festgelegt. Die Probanden hatten 200, 400, 600 oder 800 ms Zeit, das Zielitem zu lokalisieren, bevor die Items der Suchmatrix durch eine Maskierungsmatrix ersetzt wurden. Diese Lokalisierung ist eine der wesentlichen Facetten der Blindbedienbarkeit eines Komfortsystems im Fahrzeug. Wird z.B. ein Softkey-Konzept verwendet, muss mit einem kurzen Blick klar werden, mit welcher Taste das Bildelement verknüpft ist, damit diese anschließend ohne weitere Blickabwendung erfüllt und betätigt werden kann.

Mit zunehmender Darbietungslänge verbesserte sich die Fehlerrate deutlich für beide Präsentationsformate. Darüber hinaus ist erwartungsgemäß die Leistung bei Icons wesentlich besser als bei Text. Es zeigte sich jedoch auch, dass selbst bei einer relativ langen Zeit zwischen Reiz und Maske von 800 ms die Fehlerrate von Text immer noch mehr als doppelt so hoch war wie die von Icons. Diese 800 ms waren als kritische Grenze angesehen worden, bis zu der die Fehlerrate bei Text sich den Icons angeglichen haben sollte, um beide Gestaltungsvarianten als gleichermaßen geeignet ansehen zu können. Das Ergebnis legt nahe, dass bei ikonografischer Funktionsdarstellung deutlich häufiger ein einzelner Blick auf die Anzeige des Bordsystems ausreicht, um das gesuchte Element zu lokalisieren. Bei textueller Beschriftung der Funktionsschaltflächen hingegen ist wahrscheinlich, dass entweder kritisch lange Blickabwendungen von der Fahrbahn resultieren oder mehrere Einzelblicke notwendig werden.

Besonders interessant ist das Ergebnis, dass bei 800 ms die Fehlerleistung bei Icons nicht mehr so deutlich abnahm wie bei den 200 ms-Schritten zuvor. Wie von der Trendanalyse bestätigt wurde, nähert sich die Leistung der Icons hier langsam einer Asymptote. Diese Tendenz kann bei Text nicht ausgemacht werden. Gegeben, dass bei 800 ms die Fehlerrate bei Text immer noch über 30% war, ist ein Abflachen der Kurve auch nicht zu erwarten. Somit deutet einiges darauf hin, dass deutlich längere Betrachtungszeiten notwendig sind, um für textuelle Elemente eine zufrieden stellende Antwortgüte herzustellen. Bei der Umsetzung dieser Erkenntnisse im Sinne von Gestaltungsrichtlinien darf jedoch nicht vergessen werden, dass es sich um eine Worst-

Case-Schätzung handelt: Die Untersuchungssituation mit zufälliger Itemanordnung beruht ausschließlich auf visueller Suche über die Items, eine Zuhilfenahme von Gedächtnisprozessen war nicht möglich. Dieser Schritt wurde jedoch bewusst angestrebt, um das maximale Ausmaß der Auswirkungen der Gestaltungsvarianten in dem gegebenen kritischen Zeitbereich beobachten zu können.

Erklärenswert scheint die doch relativ hohe Fehlerrate selbst bei Icons bei 800 ms SOA. Sie ist deshalb erstaunlich, weil in Experiment 5 die Items unbegrenzt betrachtet werden konnten. Bereits bei einer Reaktionszeit von 1200 ms war die Fehlerrate der Icons auf ca. 1% gefallen. Wie verträgt sich dieser Befund mit der vermuteten asymptotischen Annäherung bei etwas über 10% in Experiment 8? Zwei Gründe können dafür ausgemacht werden. Erstens existierten in Experiment 5 und 8 unterschiedliche Zielsetzungen, die sich auf das Verhalten der Versuchsteilnehmer niederschlagen können. Das Wissen um die sehr kurzen Darbietungszeiten und die damit verbundene Notwendigkeit, besonders schnell zu suchen, kann zu einer höheren Stressbelastung mit suboptimaler Strategiewahl geführt haben. Durch diese spezielle Form der kognitiven Belastung kann es im Vergleich mit dem ersten Experiment zu einer Reduktion der Antwortgüte gekommen sein. Zweitens fand Experiment 8 immer erst im Anschluss an Experiment 5 bzw. 6 statt. Das kann zu Positionseffekten geführt haben, die sich z.B. durch ein Nachlassen der Konzentration äußern und mehr Fehler bedingen. Nimmt man diese beiden Faktoren zusammen, lassen sich die erhöhten Fehlerraten durchaus erklären. Darüber hinaus deutet die signifikante quadratische Trendkomponente nur an, dass die weitere Abnahme der Fehlerraten nicht in dem Ausmaß fortgeführt wird wie zuvor. Von der längsten SOA-Stufe bis zu den beobachteten 1200 ms Reaktionszeit in Experiment 5 fehlen noch 400 ms. Wenn nun die Abnahme der Fehlerraten nur in dem geringen Ausmaß wie von 600 auf 800 ms SOA fortgesetzt wird, d.h. ungefähr 2-3% pro 200 ms, ist eine Stabilisierung bei 5-6% durchaus denkbar. Unter Berücksichtigung der genannten Einflussfaktoren kognitive Belastung und nachteilige Rahmenbedingungen für die Konzentration kann die Interpretation aufrechterhalten werden, dass sich die Fehlerrate der Icons bei 800 ms SOA bereits in der Annäherung an die Asymptote befand.

5.2.8 Zusammenfassende Diskussion und Gestaltungsempfehlungen

In einer Serie von vier Experimenten wurde der Frage nachgegangen, ob zwischen ikonografischen oder verbalen Funktionsdarstellungen in einer grafischen Bedienoberfläche stabile Unterschiede existieren (Experimente 5 - 7) und wie sich diese Unterschiede im Anwendungskontext relevant auswirken können (Experiment 8). Das inhaltliche Ziel der Experimente ist eine empirisch begründete Designentscheidung zur

Gestaltung des grafischen User Interface im Umfeld multimodaler Bedienvorgänge bei einem Bordsystem im Fahrzeug. Im Mittelpunkt steht dabei folgende Problematik: Um dem Nutzer das zur Spracheingabe notwendige Vokabular bereits während manueller Interaktionsvorgänge nahe zu legen, wird häufig die bewährte *Say What You See* Strategie angewandt. Auf diesem Weg wird die Variabilität des Nutzerwortschatzes reduziert und die sprachbasierte Interaktion verbessert. Der Nachteil ist, dass sehr viele Textelemente auf der Bedienoberfläche zu finden sind. Berücksichtigt man, dass Bedienschritte unter visueller Kontrolle so kurz wie möglich sein sollten, kann die verbale Beschriftung der Steuerelemente von Nachteil sein. Als alternative Lösung für die visuell-manuelle Schnittstelle ist der verstärkte Einsatz von Icons denkbar, wodurch jedoch die erstrebenswerte enge Parallelität von sprachlicher und manueller Schnittstelle aufgegeben werden müsste. Wie einleitend gezeigt (vgl. Abschnitte 5.2.2 und 5.2.3), kann auf der Basis der gegenwärtig verfügbaren Literatur keine klare Gestaltungsempfehlung gegeben werden. Mit Hilfe eines einheitlichen experimentellen Ansatzes sollte die Untersuchungsreihe diesem Forschungsbedarf nachkommen. Die ersten drei Experimente variierten systematisch den Anteil von visuellen Such- und Gedächtnisprozessen beim Auffinden einer Schaltfläche mit Icons oder verbalen Beschriftungen. In Experiment 5 wurden die Items nach jeder Aufgabe zufällig angeordnet, wodurch zufällige visuelle Suche forciert wurde. In Experiment 6 blieb das Layout nach einer einmalig zufälligen Anordnung konstant. Dadurch konnte begrenzt auf Gedächtnisleistungen zurückgegriffen werden. Zusätzlich zu dieser ortskonstanten Anordnung kam in Experiment 7 der Aspekt einer inhaltlichen Gruppierung hinzu und ermöglichte einen stärkeren Rückgriff auf Gedächtnisinhalte und eine größere Unabhängigkeit von visuellen Charakteristika der Items.

In Experiment 5 konnte gezeigt werden, dass die Ergebnisse der Untersuchung von Arend und Kollegen (1987) auch mit aktuellen standardisierten Icons und dem seit den achtziger Jahren stark veränderten Erfahrungshintergrund der Nutzer Bestand haben. Obwohl einfacheres verbales Material verwendet wurde und die Icons sich nicht optimal in globalen Merkmalen unterschieden, konnte ein deutlicher Vorteil für Icons in den Reaktionszeiten nachgewiesen werden. Dieser Vorteil war abhängig von der Komplexität der Suchvorlage. Mit zunehmender Größe der Suchmenge stieg die Suchzeit bei Text schneller als bei Icons.

Experiment 6 untersuchte die Fragestellung, ob die bei unterschiedlichem Iconmaterial nachgewiesene kompensierende Wirkung von ortsfester Anordnung bezüglich der Suchzeiten (Green & Barnard, 1990) auch bei Icons und Text greift. Es zeigte sich, dass durch die Einführung einer Gedächtniskomponente in beiden Darstellungsvarianten eine substantielle Reduktion der Reaktionszeiten gegenüber Experiment 5 zu erreichen ist. Der zunehmende Vorteil von Icons bei größer werdenden Suchmengen hingegen bleibt unverändert erhalten. Schnelleres Lernverhalten war bei Icons entge-

gen der geäußerten Hypothese nicht zu beobachten. Stattdessen zeigte sich, dass Icons in allen experimentellen Blöcken schneller gefunden werden. Ortskonstanz hat bei einem direkten Vergleich von Icons und Text also keine kompensierende Wirkung.

Experiment 7 ging noch einen Schritt weiter und führte zusätzlich eine inhaltliche Gruppierung der Items ein. Funktional ähnliche Elemente blieben innerhalb einer Zeile der Suchmatrix gruppiert angeordnet. Der einzige objektive Unterschied zu Experiment 6 war damit eine rudimentäre Logik in der Anordnung, wie sie von jedem Systemdesigner verfolgt werden würde. Allein dadurch ließ sich der Vorteilseffekt für Icons vollständig eliminieren. Die Suchzeiten von Icons und Text unterschieden sich weder bei der Betrachtung von Lerneffekten, noch wenn die Auswirkung verschiedener Suchmengen im Fokus stand. Bei der visuellen Gestaltung von Mensch-Maschine-Schnittstellen gelangen neben einer einfachen Gruppierung noch weitere strukturierende Faktoren basierend auf gestalttheoretischen Überlegungen zur Anwendung. Damit kann entgegen der häufig verbreiteten Meinung geschlossen werden, dass unter diesen Bedingungen kein gravierender Unterschied zwischen Icons und Text in den Suchzeiten eintreten wird.

Zusammengefasst zeigen diese Experimente auch, warum in der Forschung bisher Uneinigkeit darüber bestand, ob Icons tatsächlich das Potential haben, den Interaktionsprozess durch geringere Suchzeiten zu verkürzen: Die Beiträge der visuellen Charakteristika des Reizmaterials wurden bisher nicht systematisch von den Beiträgen des Gedächtnisses getrennt. Diese Lücke kann durch die vorliegenden Erkenntnisse geschlossen werden.

Experiment 8 versuchte einzuordnen, in welchem Ausmaß sich ein möglicher Unterschied zwischen Icons und Text in der Anwendungssituation auswirkt. Die Probanden hatten auch hier die Aufgabe, das Zielitem in der Suchmatrix zu lokalisieren. Diese Suchmatrix war aber nur für den kurzen Zeitraum von 200-800 ms sichtbar. Der Zeitraum wurde gewählt, da er basierend auf Sicherheitserwägungen (Zwahlen et al., 1988) als unkritisch für die Dauer einer Blickabwendung zur Interaktion mit einem Bordsystem im Fahrzeug gelten kann. Aufgabe des Experiments war eine Abschätzung, ob bei dem Suchmuster wie es von unerfahrenen Nutzern zu erwarten ist, überhaupt ein Unterschied zwischen Icons und Text zu allen Messpunkten innerhalb des kritischen Zeitraums besteht. Wie in Experiment 5 wurde hierzu die visuelle Suche durch zufällige Itemanordnung forciert. Für eine Verallgemeinerbarkeit der Resultate auf die Gestaltung der visuellen Schnittstelle im Fahrzeug ist eine Untersuchung dieses Aspekts zwingend notwendig. Das Ergebnis war überraschend eindeutig: Ist visuelle Suche auf der grafischen Oberfläche erforderlich, besteht selbst bei 800 ms eine mehr als doppelt so hohe Fehlerrate für textuelle Beschriftungen der Steuerelemente. Während Icons sich bei dieser Darbietungsdauer mit knapp 13% einer ansatzweise akzeptablen Fehlerrate angenähert haben, sind für Text noch 31% Fehler zu verzeichnen.

Dabei nähert sich im kritischen Betrachtungsbereich die Fehlerrate der Icons langsam der Asymptote an. Das bedeutet, dass unter den gegebenen Bedingungen – hoher Zeitdruck und eine gewisse Ermüdung, da die Versuchspersonen bereits ein vollständiges Experiment zuvor absolviert hatten – bei 800 ms unter Icondarbietung bereits ein gutes Leistungsniveau erreicht ist. Die Daten lassen die Schlussfolgerung zu, dass durch visuelle Suche bei Text ein einzelner kurzer Blick auf das Display in deutlich weniger Fällen ausreichend sein wird als bei Icons.

Alle Ergebnisse der Experimente beruhen darauf, dass das verwendete Iconmaterial den Nutzern bekannt ist. Wie in Abschnitt 5.2.2 dargelegt, ist mit einer sehr hohen Fehlerrate zu rechnen, wenn die Abbildung zwischen Bild und Referent nicht eindeutig oder unklar ist. Durch die Verwendung von Standardsymbolen kann diese Voraussetzung erfüllt werden. Letztlich muss jedoch immer ein Nutzertest einer Gestaltungsentscheidung vorausgehen. Nur so kann sichergestellt werden, dass der Erfahrungshintergrund der Zielgruppe erfasst wurde.

Aus den Experimenten lassen sich einige konkrete Gestaltungsempfehlungen ableiten. Unter der Annahme, dass die Icons der Nutzergruppe bekannt sind, können folgende Empfehlungen geben werden:

- Wenn davon ausgegangen werden kann, dass die Orientierung auf der grafischen Benutzeroberfläche durch unsystematische visuelle Suche erfolgt, sind Icons vorteilhaft. Dies kann bei Nutzern eintreten, denen zwar generell die Funktionen des Systems bekannt sind, die aber aufgrund fehlender Erfahrung nicht wissen, wo sie eine gesuchte Funktion auslösen können. Ferner kann dies der Fall sein, wenn Informationen an wechselnden Orten ausgegeben werden (z.B. Kombiinstrument oder Infotainment-Display).
- Der Vorteil von Icons war von der Größe der Suchmenge abhängig. Bei einer geringen Anzahl von Informationselementen auf dem Bildschirm wird nur ein geringer oder kein Unterschied in der Praxis auftreten. Wenn nur eine überschaubare Anzahl von Steuerelementen dargestellt werden muss, können grundsätzliche textuelle Beschriftungen ohne Bedenken eingesetzt werden. Die genaue Quantifizierung einer „geringen Anzahl“ sollte jedoch wiederum in Nutzertests überprüft werden, da die Zahl von sechs Items als kleine Suchmenge aus diesem Experiment nicht beliebig übertragbar ist.
- Wenn keine inhaltlich unmittelbar verständliche Anordnung der Steuerelemente möglich ist, führen Icons ebenfalls zu geringeren Suchzeiten. Eine örtlich konstante Anordnung von Steuerelementen ruft zwar allgemein kürzere Suchzeiten hervor, der Vorteil für Icons bleibt jedoch bei Ortskonstanz erhalten. Aufgrund der feststellbaren allgemeinen Reduktion für beide Darbietungsvarianten sollte in jedem Fall überprüft werden, ob nicht bereits eine zu

erreichende kritische Grenze unterschritten wurde. Dafür bietet sich ein Vorgehen wie in Experiment 8 an.

- Sofern sichergestellt wird, dass ein hoher Strukturierungsgrad der Benutzeroberfläche vorliegt und eine inhaltliche Erlernbarkeit des Aufbaus gewährleistet ist, lässt sich zumindest ein deutlicher Vorteil für Icons weitgehend ausschließen. Mit textueller Beschriftung kann dann eine ähnliche Leistung erzielt werden.

Auf Grundlage der Experimente sollte entschieden werden, ob die verhältnismäßig große Anzahl verbaler Beschriftungen, die *Say What You See* mit sich bringt, als problematisch angesehen werden muss. Dies kann verneint werden, solange dafür gesorgt wird, dass – wie im letzten Gestaltungshinweis angegeben – die Aufbereitung der visuellen Oberfläche klar inhaltlich strukturiert und in ihrem Aufbau lernförderlich ist.

6 Prozessmodell zur Entwicklung und Evaluation multimodaler Schnittstellen

Die in den vorangegangenen Kapiteln aufgeworfenen Fragestellungen im Zusammenhang mit der Gestaltung multimodaler Schnittstellen machen deutlich, dass systematische Nutzertests zur Verbesserung der Interaktionsqualität zwischen Mensch und Maschine unerlässlich sind. Ein mögliches Vorgehen, das sich in der Praxis häufig beobachten lässt, besteht darin, zunächst einen multimodalen Prototyp technisch vollständig zu realisieren und diesen im Anschluss experimentell zu evaluieren. Dieses Vorgehen birgt jedoch mehrere Nachteile. Als besonders schwierig erweist sich das bereits angesprochene Problem der kombinatorischen Explosion (Schomaker et al., 1995), da bei mehreren gleichzeitig zu untersuchenden Ein- und Ausgabemodalitäten inklusive ihrer Gestaltungsvarianten extrem viele experimentelle Bedingungen notwendig wären. Sowohl die hohe Anzahl erforderlicher Versuchsteilnehmer als auch eine strikte experimentelle Kontrolle sind kaum erreichbar. Darüber hinaus dürften während der Konstruktion eines komplexen Prototyps viele Designentscheidungen implizit getroffen werden, die nicht weiter evaluiert werden. Ferner ist der technologische Aufwand bei diesem Vorgehen bereits in einer frühen Projektphase sehr hoch, was ein effizientes Zusammenspiel von Technikern und Humanwissenschaftlern erschwert. Nutzertests können erst nach weitgehender Beendigung der komplexen Konstruktions- und der technischen Testphase stattfinden. Schließlich muss bedacht werden, dass nachträgliche Änderungen an einem System von technischer Seite umso schwieriger und kostenintensiver sind, je höher der Komplexitätsgrad des Systems ist.

Im Rahmen der vorliegenden Arbeit wurde deshalb ein alternatives Prozessmodell zur Konstruktion multimodaler Schnittstellen entworfen. Das Ziel war dabei, ein systematisches und generalisierbares Vorgehen zu entwickeln, das den versuchsplanerischen und technologischen Herausforderungen bei der Anlage multimodaler Systeme Rechnung trägt. Durch eine Zerlegung der Implementierung und des Tests von Schnittstellenkomponenten in handhabbare Einzelschritte soll sichergestellt werden,

dass die in den Planungsphasen des Systems aufgestellten Kriterien nachprüfbar erfüllt werden. Abbildung 46 veranschaulicht den Prozess.

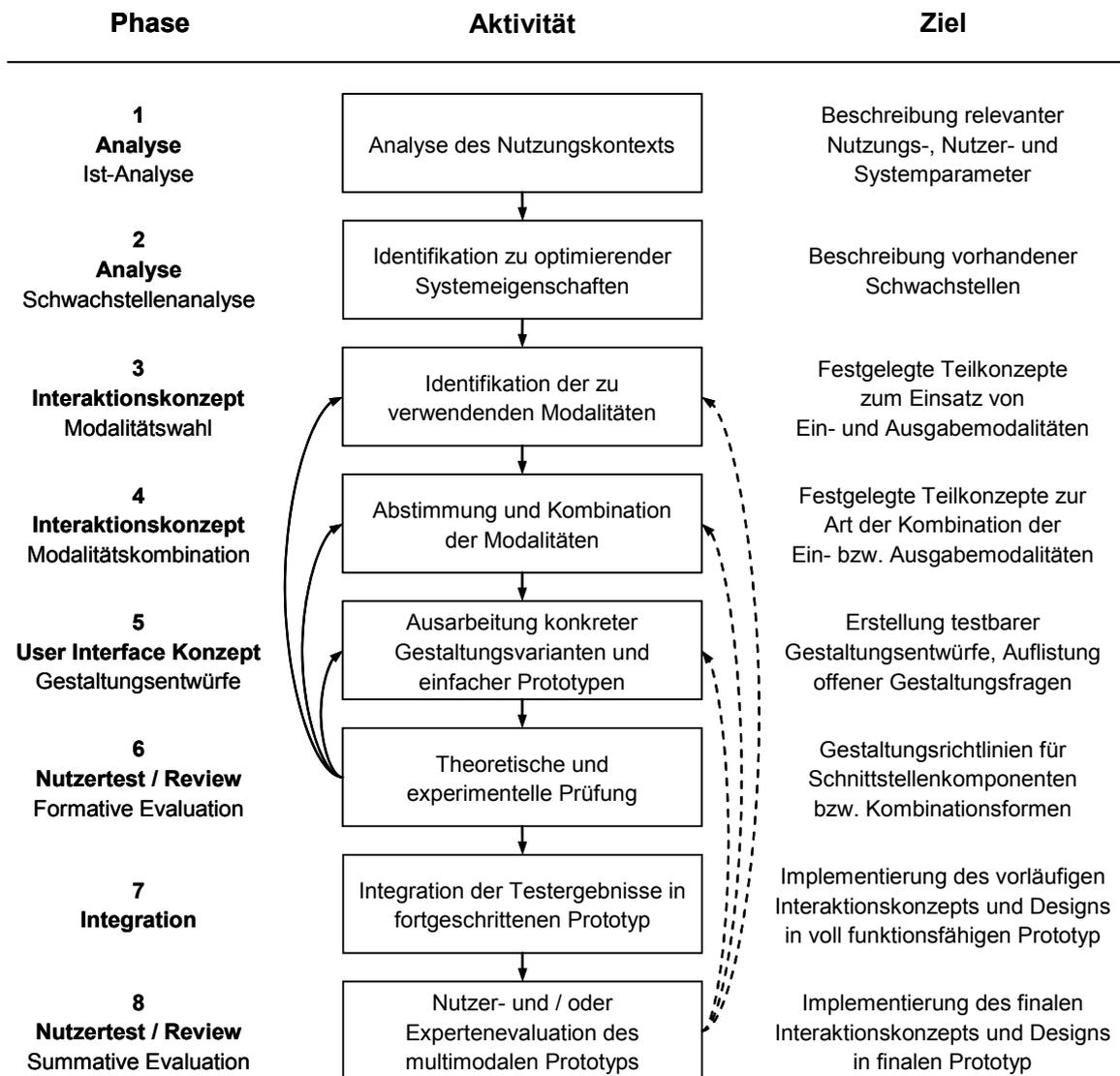


Abbildung 46: Prozessmodell zur Forschung an multimodalen Schnittstellen.

Das Kernstück des Prozessmodells ist die Aufspaltung der empirischen Evaluation. Statt direkt an einem finalen Prototyp zu arbeiten, werden geeignete Fragestellungen in einem möglichst frühen Stadium des Entwicklungsprozesses in separaten Untersuchungen zu Teilaspekten iterativ angegangen. Eine abschließende Gesamtevaluation prüft die Qualität des integrierten Finalsystems. Die explizite Angabe der einzelnen Phasen, der notwendigen Aktivitäten in jeder Phase und deren Ergebnisse soll dazu beitragen, die Projekt- und Forschungsplanung für multimodale Schnittstellen über unterschiedliche Anwendungsdomänen hinweg transparent zu gestalten. Durch Aufzeigen der jeweils im Zentrum stehenden Frage werden die Phasen nachfolgend unter Angabe von Beispielen ausgeführt.

Phase 1 – Analyse: Ist-Analyse. Die erste Phase des Modells beantwortet die Frage *Wodurch ist der gegebene Nutzungskontext charakterisiert?* In Orientierung an ISO 9241-11 (1998) müssen zur Erfassung des Nutzungskontexts die Benutzer des Systems, deren Arbeitsaufgaben, Arbeitsmittel und die Umwelt, in der sie sich dabei bewegen, identifiziert werden. Ein solides Verständnis des Nutzungskontexts muss die Grundlage jeder optimierenden Maßnahme für ein System bilden, da dieser letztlich im Gesamtest als Prüfstein fungiert. Im Kontext der Fahrzeugführung ist beispielsweise ein Kennzeichen, dass eine große Anzahl von Komfortfunktionen über das Bordsystem zur Verfügung steht (für weitere Charakteristika siehe Abschnitt 3.1).

Phase 2 – Analyse: Schwachstellenanalyse. In Phase 2 wird darauf aufbauend erarbeitet, welche Aspekte in der Interaktionssituation zu verbessern sind, d.h. die Frage *Was führt zu Problemen?* wird für das zu optimierende System beantwortet. So entsteht etwa durch die große Anzahl von Komfortfunktionen eine zu lange visuelle Abwendung von der Fahraufgabe bei der visuell-manuellen Bedienung des Bordsystems. Möglicherweise sind die Schwachstellen bereits im Vorfeld bekannt und bilden den Ausgangspunkt für den geplanten Einsatz multimodaler Lösungen. Dennoch sollten die ersten zwei Phasen im Prozess bewusst durchschritten werden, um frühzeitig falsche Vorstellungen über Optimierungspotentiale aufdecken zu können.

Phase 3 – Interaktionskonzept: Modalitätswahl. Die dritte Phase wirft die Frage *Welche Teile des Problemraums können sich durch multimodale Interaktion verbessern lassen?* auf. Theoretische Modelle wie das Modell der multiplen Ressourcen (vgl. Abschnitt 2.2.2) und empirische Befunde aus der Literatur zur Mensch-Maschine-Interaktion geben einen Rahmen bei der Suche vor, die in einem oder mehreren Teilkonzepten zum Einsatz von Ein- und Ausgabemodalitäten mündet. Für das Bordsystem kann zum Beispiel der verstärkte Einsatz von akustischen Ausgaben oder die Implementierung von Spracheingabe vorgeschlagen werden. Empfehlenswert ist bereits hier eine Prüfung auf technische Realisierbarkeit als Abschluss dieser Phase.

Phase 4 – Interaktionskonzept: Modalitätskombination. Phase 4 setzt die Optimierungsüberlegungen fort, indem ebenfalls unter Berücksichtigung der technischen Machbarkeit die Frage gestellt wird *Wie können die Modalitäten zusammenwirken?* In Form noch relativ abstrakter Teilkonzepte wird erarbeitet, wie die verfügbaren Ein- bzw. Ausgabemodalitäten kombiniert werden sollen. Dabei können unterschiedliche Vorschläge für die Kombination von Modalitäten entstehen und für die jeweiligen Modalitäten zur Ein- und Ausgabe eigene Kombinationskonzepte notwendig werden. In dieser Phase wird zum Beispiel festgelegt, ob redundante Ausgabe durch äquivalente Modalitäten oder eine Spezialisierung der Ausgabemodalitäten vorliegt. Auf der Eingabeseite muss beispielsweise entschieden werden, inwiefern eine Fusion der Ein-

gabemodalitäten sinnvoll bzw. nutzerseitig zu erwarten ist und wie die Aufgabenteilung unter den Eingabemodalitäten konzeptualisiert ist. Weitergehende Ausführungen zu Kombinationen von Modalitäten finden sich in Abschnitt 2.1.3.2. Für ein nahtloses, nutzungsorientiertes Ineinandergreifen der Modalitäten muss ferner dafür gesorgt werden, dass von vornherein ein gemeinsames Bedienkonzept entwickelt wird (holistischer Ansatz). Gerade im Produktgeschäft ist hingegen häufig zu beobachten, dass multimodale Lösungen nicht ganzheitlich entwickelt werden. So wird beispielsweise Spracheingabe teils erst nachträglich integriert, wenn bereits die Benutzerführung über die grafische Bedienoberfläche eingerichtet ist. Unter diesen Bedingungen wird die sinnvolle Nutzung paralleler Eingabewege wie bei *Say What You See* dadurch eingeschränkt, dass das für manuelle Handlungsabläufe optimierte grafische Interface nicht darauf abgestimmt ist, effizient sprachliche Abläufe zu unterstützen.

Phase 5 – User Interface Konzept: Gestaltungsentwürfe. Mit Phase 5 wird die Ausformung der gesammelten Konzeptideen konkretisiert. Die Leitfrage ist dabei *Welche konkreten Gestaltungsformen sind für die gegebenen Ein- und Ausgabemodalitäten bzw. für deren Kombinationen denkbar?* Unter den abgesteckten Randbedingungen durch Phase 3 und 4 werden die Schnittstellenkomponenten ausgestaltet. Dabei können sich gegebenenfalls alternative Varianten ergeben, die in technologisch möglichst einfachen Prototypen einem Test zugeführt werden sollten. Beispielsweise kann der Vorschlag entstehen, die akustischen Rückmeldungen im System durch sprachliche Ausgaben oder durch nicht-sprachliche Auditory Icons oder Earcons zu realisieren. Neben unimodalen Gestaltungsfragen können auch genuin multimodale Aspekte zu dem Vorschlag unterschiedlicher Ausgestaltung führen, z.B. bezogen auf den Zeitversatz bei der Fusion sprachlicher und manueller Eingabe. Zusätzlich zu diesen Gestaltungsvarianten können ferner ungeklärte Gestaltungsaspekte in dieser Phase ans Tageslicht treten. So kann beispielsweise die Frage auftauchen, wie viele abstrakte Earcons der Nutzer überhaupt unterscheiden kann.

Phase 6 – Nutzertest/Review: Formative Evaluation. In Phase 6 wird zwischen den Gestaltungsalternativen entschieden, bzw. die offenen Aspekte aus Phase 5 werden einer Klärung zugeführt. In Form von Gestaltungsrichtlinien muss eine Antwort auf die Frage *Welche Form der Gestaltung ist geeignet?* gefunden werden. Sofern auf Grundlage der relevanten Literatur eine Entscheidung möglich ist, können diese Richtlinien theoretisch ermittelt werden. Anderenfalls wird mit Hilfe von mehreren einzelnen Experimenten eine Bewertung der zu prüfenden Designalternativen vorgenommen. Für ein systematisches Vorgehen erscheint es sinnvoll, zunächst innerhalb einer zu verwendenden Modalität die bestmögliche Gestaltungsvariante ausfindig zu machen, bevor modalitätsübergreifende Aspekte der Kombination adressiert wer-

den. So kann durch geeignete objektive und subjektive Maße zunächst ermittelt werden, ob sich zum Beispiel Spracheingabe als sinnvolle Modalität im gegebenen Kontext bewährt oder welche von mehreren Varianten einer sprachbasierten Schnittstelle zu den besten Interaktionsresultaten führt. Im Anschluss an diese unimodalen Fragestellungen kann dann experimentell entschieden werden, welche Form der Kombination von Sprache und manueller Interaktion in einem gegebenen Kontext den Nutzern die beste Unterstützung bietet. Aufgrund der Aufspaltung in mehrere Einzelexperimente können die erforderlichen Prototypen oder Systemsimulationen vergleichsweise einfach, abstrakt und funktional eingeschränkt sein. Teilweise sind sogar unimodale Prototypen für die Tests ausreichend. Das bedeutet, dass in dieser Phase der Konstruktion eines multimodalen Prototyps noch mit relativ geringem Aufwand und geringen Kosten ungeeignete Gestaltungsalternativen systematisch ausgesiebt werden können. Durch fokussierte Fragestellungen in den Einzelexperimenten lässt sich ferner ein hoher Grad versuchsplanerischer Kontrolle aufrechterhalten. Die Ergebnisse der Teilexperimente können eine Neukonzeptualisierung der verwendeten Modalitäten bzw. ihrer Kombinationen oder von Detailspekten der Gestaltung erforderlich machen. Wie bei jeder formativen Evaluationsprozedur sind eventuelle Änderungen anschließend wieder zu prüfen. Die grafische Darstellung des Prozesses (Abbildung 46) weist darauf mit durchgezogenen Rückführungspfeile hin.

Phase 7 – Integration. In Phase 7 steht zunächst die Frage *Welche Gestaltungsformen haben sich in der Prüfung bewährt und stehen für eine Integration zur Verfügung?* im Mittelpunkt. Mit der Integration wird ein abgestimmtes Konzept realisiert, das die Grundlage für die Architektur und die Eigenschaften des angestrebten multimodalen Systems bildet. Erst in dieser Phase wird der vollständige multimodale Prototyp erstellt. Dieser muss soweit als möglich unter typischen Umweltbedingungen im intendierten Nutzungskontext getestet werden.

Phase 8 – Nutzertest/Review: Summative Evaluation. Phase 8 dient der abschließenden Evaluation mit Usability-Tests und Experimenten. Die grundlegende Frage ist hier: *Wie agieren die Nutzer unter Realbedingungen mit dem prototypischen System?* Dabei soll vor allem erkannt werden, wie das neu entstandene System tatsächlich genutzt wird, welche Interaktionspräferenzen die Nutzer aufweisen und welche technologischen bzw. interaktionsgestalterischen Aspekte bei der Bedienung Schwierigkeiten machen. Ein besonderes Augenmerk ist darauf zu richten, ob sich die erfolgreichen Lösungen aus den Einzeltests auch im Gesamttest als sinnvoll erweisen oder ob sich durch die Integration in ein multimodales System ein Zustand ergeben hat, in dem die Nutzer ihr Verhalten maßgeblich ändern. Die Ergebnisse der Einzeltests (Phase 6) können hier bei geeigneter Anlage der Untersuchungen als Baseline dienen. Abhängig

vom Erfolg des Gesamtsystems ist möglicherweise trotz des summativen Charakters dieser Testphase eine iterative Revision der Modalitätsauswahl, ihrer Kombinationen oder von Gestaltungsaspekten notwendig (veranschaulicht durch gestrichelte Rückführungspfeile in Abbildung 46).

Die vorliegende Arbeit hatte mitunter das Ziel, das Fundament zur Konstruktion eines multimodalen Systems zu legen. Die Vorgehensweise dafür war an dem hier vorgeschlagenen Prozessmodell orientiert und vollzieht einen großen Teil der notwendigen Entwicklungsarbeit. Für die Phasen 1-4 wurde eine umfangreiche Analyse der wissenschaftlichen Literatur durchgeführt (vgl. Kapitel 2 und 3 sowie die Einführungen zu den jeweiligen Experimenten). Der experimentelle Teil deckt die Phasen 5 und 6 mit den dort dargelegten Gestaltungsrichtlinien ab. Der Schwerpunkt der Arbeit liegt auf diesen Experimenten und deren Ergebnissen. Einige dieser Erkenntnisse konnten jedoch bereits in einem abgestimmten Interaktionskonzept für einen Prototyp realisiert werden, weshalb diese Implementierung hier zur Veranschaulichung dargelegt wird. Der Prototyp verfügt über manuelle Eingabe und Spracheingabe. Im Kern ist er eine Verschmelzung und gestalterische Ausarbeitung der Versuchsprototypen aus den Experimenten zu Touchpad-Steuerung und Spracheingabe (Funktionsumfang vgl. Anhang B 1). Das Touchpad wurde aufgrund der Ergebnisse der Experimente 1 und 2 zur zentralen manuellen Steuerung einem Dreh-Drücksteller vorgezogen. Zusätzlich sind vier Taster als Softkeys und eine größere Hauptmenü-Taste angebracht (siehe Abbildung 47).



Abbildung 47: Überarbeiteter multimodaler Prototyp im Cockpit Mock-up.

Das Touchpad bietet wie in den Experimenten zur zusätzlichen nicht-visuellen Aktionsrückmeldung redundantes vibrotaktilen und akustisches Feedback. Die vibrotaktilen Rückmeldungen entsprechen weitgehend denen aus den Experimenten 1 und 2. Allerdings wurde auf die zusätzliche Markierung der neutralen Mitte bei virtuellen Schieberegler verzichtet, da sich diese in den Experimenten nicht bewährt hat. Die akustischen Rückmeldungen wurden gegenüber der Experimentalversion erweitert. Experiment 3 zeigte, dass nicht-sprachliche Schalle gegenüber sprachlichen Ausgaben möglicherweise vorzuziehen sind, aber dass einfache Earcons Probleme hinsichtlich ihrer Verständlichkeit verursachen können. Neben einfachen Auditory Icons bei Selektion und Bestätigung wurden deshalb Earcons genutzt, die hierarchische Strukturen repräsentieren können. In menübasierten Systemen können hierarchische Earcons eine bessere Orientierung bieten als einfache Sounds und führen zu einer sehr hohen Wiedererkennungsrates (Brewster, Ratty & Kortekangas, 1996). Der Vorteil dieses Feedback-Konzepts ist, dass diese vergleichsweise kurzen Ausgaben sowohl zur Bestätigung manueller Aktionen als auch zur Rückmeldung der korrekten Erkennung bei Spracheingabe geeignet sind.

Aufbauend auf der Software-Plattform von Experiment 4 wurde eine Sprachsteuerung mit kleinem Vokabular (ca. 200 Wörter) implementiert, die multimodales *Talk-Ahead* mit Online-Erkennung ermöglicht und auf das *Say What You See* Prinzip aufsetzt. Spracheingabe und manuelle Eingabe sind vollständig äquivalent angelegt. Darüber hinaus existieren für die wichtigsten Funktionen sprachliche Direktzugriffe. Die Aktionsmodalitäten können als Alternativen sequentiell verwendet werden, eine Fusion der Eingabeströme findet nicht statt. Diese Entscheidung erfolgte vor dem Hintergrund, dass sequentielle Multimodalität für Mehraufgabensituationen wie zum Beispiel im Fahrzeug besonders gut geeignet ist (vgl. Kapitel 2 sowie Bengler, 2001). Die Experimente 5-7 haben gezeigt, dass durch eine klare, bedeutungstragende Strukturierung der Bedienoberfläche die Verwendung verbaler Beschriftungen nicht zu einer Verschlechterung der visuellen Orientierung führt. Entsprechend wurde versucht, eine übersichtliche Einteilung der grafischen Bedienoberfläche zu gewährleisten. Sprechbare Elemente wurden zusätzlich durch Schriftfarbe und -form hervorgehoben.

Diese Auflistung zeigt beispielhaft, wie die Ergebnisse aus den theoretischen Analysen und den experimentellen Phasen des Prozessmodells für ein abgestimmtes Interaktionskonzept interpretiert werden können, welches dann wiederum einer abschließenden Prüfung unterzogen werden muss. Durch das vorgeschlagene systematische Vorgehen war es möglich, innerhalb relativ kurzer Zeit zu einem multimodalen System zu gelangen, dessen Komponenten auf nachvollziehbare Weise gestaltet wurden. Somit empfiehlt sich das Prozessmodell auch für die zukünftige Forschung an multimodalen Schnittstellen, deren technologische und gestalterische Komplexität unvermindert zunehmen werden.

Anhang

Anhang A

A 1: Aufgaben und Anweisungen an die Versuchsperson (Exp. 1 und 2)

- A1 (Menü): „CD, Titel 4“; „CD, Titel 5“; „CD, Titel 6“
- A2 (horizont. Regler): „Nächster Wert: Neutral“; „Nächster Wert: -1“; „Nächster Wert: +2“
- A3 (vert. Regler): „Nächster Wert: 5“; „Nächster Wert: 7“; „Nächster Wert: Maximum“
- A4 (2D): „Nächster Buchstabe: B, wie Berta“; „Nächster Buchstabe: R, wie Richard“; „Nächster Buchstabe: T, wie Theodor“

A 2: Bearbeitungszeiten [s] pro Aufgabe für die einzelnen Aufgabentypen (Exp. 1)

Aufgabe	Fahr- bedingung	Schnittstelle									
		TPsolo		TPaud		TPtakt		TPaud+takt		DD	
		M	(SE)	M	(SE)	M	(SE)	M	(SE)	M	(SE)
A1: Menü (Liste)	Stand	1.93	(0.13)	2.23	(0.29)	2.07	(0.11)	1.87	(0.11)	1.90	(0.12)
	Fahrt	3.27	(0.36)	2.96	(0.22)	3.14	(0.30)	3.10	(0.28)	2.92	(0.27)
A2: h. Regler (neutr. Mitte)	Stand	1.72	(0.07)	1.68	(0.07)	1.85	(0.14)	1.77	(0.14)	1.91	(0.17)
	Fahrt	2.78	(0.34)	2.28	(0.19)	2.43	(0.27)	2.35	(0.16)	2.48	(0.38)
A3: v. Regler (ohne Mitte)	Stand	1.84	(0.08)	1.77	(0.07)	1.87	(0.11)	1.94	(0.11)	1.56	(0.07)
	Fahrt	2.98	(0.34)	2.50	(0.16)	2.12	(0.15)	2.39	(0.18)	2.55	(0.18)
A4: 2D- Eingabe	Stand	2.77	(0.22)	2.20	(0.10)	2.77	(0.23)	2.52	(0.15)	2.83	(0.13)
	Fahrt	4.62	(0.45)	3.53	(0.30)	3.73	(0.52)	4.00	(0.47)	5.60	(0.48)

A 3: RTLX-Ergebnisse für jede Schnittstellenvariante auf Subskalenebene (Exp. 1 u. 2)

Subskalen-Wertebereich: 0-20.

Dimension	Schnittstelle										<i>F</i> (4,76)	<i>p</i>
	TPsolo		TPaud		TPtakt		TPaud+takt		DD			
	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)	<i>M</i>	(<i>SE</i>)		
Geistige Anforderungen	9.9	(1.3)	6.4	(1.1)	7.1	(0.9)	5.5	(0.9)	8.1	(1.2)	7.99	<.001
Körperliche Anforderungen	5.1	(1.1)	5.2	(1.0)	5.9	(1.1)	5.3	(0.9)	6.0	(1.1)	0.32	.86
Zeitliche Anforderungen	7.2	(1.2)	5.3	(1.0)	5.3	(0.9)	5.4	(1.1)	6.3	(1.2)	2.53	.05
Ausführung der Aufgaben	6.1	(1.0)	4.3	(1.0)	5.6	(1.1)	6.0	(1.0)	4.9	(0.8)	1.79	.14
Anstrengung	8.6	(1.2)	5.5	(0.9)	6.6	(0.9)	6.0	(1.0)	7.5	(1.0)	2.99	.02
Frustration	6.0	(1.2)	2.1	(0.5)	3.3	(0.7)	1.8	(0.4)	3.1	(0.7)	7.35	<.001

Anhang B

B 1: Überblick über die Funktionen des Prototyps

Menü	Submenü	Funktionsbeschreibung
Hauptmenü		- Auswahl der Funktionsbereiche (Entertainment, Navigation, Einstellungen, Kommunikation)
	Radio	- Radiosender aus Liste wählen - Senderauswahl durch Funktion voriger / nächster Sender - Anwahl: Favoriten (6 Sender) / alle Sender (10 Sender) - Radio ein / ausschalten
Entertainment	CD	- Titel aus Liste wählen (12 Titel verfügbar) - Titelauswahl durch Funktion voriger / nächster Titel - Start / Stop der Wiedergabe - Pause-Funktion
	MP3	- <wie CD; 3 Titel verfügbar>
Kommunikation		- Nur Auswahl der Submenüs SMS, Telefon, Adressbuch möglich. Keine simulierten Funktionen hinterlegt.
	Adresseingabe	- Ort, Straße, Zwischenziel festlegen - Zieleingabe (für Ort, Straße bzw. Zwischenziel) durch zweidimensionales Speller-Feld mit Buchstabeneingabe oder Auswahl aus Ortsliste - simulierter Start der Zielführung
Navigation	Zielliste	- Anwahl: Letzte Ziele / gespeicherte Ziele - Auswahl aus Liste mit je 6 Einträgen - Übernehmen des Ziels in Adresseingabe (Einleiten Zielführung / Detailänderung des Ziels)
	Sonderziele	- Auswahl nächstgelegener Sonderziele: Hotel, Restaurant, Tankstelle, Bahnhof, Flughafen, Werkstatt - Übernehmen des Ziels in Adresseingabe (Einleiten Zielführung / Detailänderung des Ziels)
	Klima	- Festlegen der Luftmenge für Scheibe, Mitte, Fußraum und Gesamt (Summe) über Schieberegler (Aus, 5 Werte) - Temperatureinstellung (Schieberegler, 20 Werte)
Einstellungen	Sitzheizung	- Temperatureinstellung für Sitz links und Sitz rechts über Schieberegler (Aus, 5 Werte)
	Standheizung	- Temperatureinstellung der Standheizung (Schieberegler, 20 Werte) - Timer-Einstellung (Schieberegler, 30 Werte)

B 2: Aufgaben und Anweisungen an die Versuchsperson (Exp. 4)

Unter den Anweisungen sind jeweils die Lösungsschritte angegeben.

Version a

- 1: Ent-1: „Sie möchten Radio hören. Stellen Sie im Radio den 2. Sender ein und schalten Sie das Radio ein.“
Entertainment – Radio – Sender 2 – Einschalten.
- 2: Ent-2: „Schalten Sie im Radio einen Sender weiter und wechseln Sie dann einen Bildschirm zurück auf den Auswahlschirm des Entertainment-Bereichs.“
Entertainment – Radio – nächster Sender / Sender 3 - Escape
-
- 3: Ein-1: „Erhöhen sie die Luftzufuhr zur Scheibe auf die 4. Stufe.“
Einstellungen – Klima – Scheibe – Stufe 4.
- 4: Ein-2: „Stellen Sie die Laufzeit der Standheizung auf 14 Minuten.“
Einstellungen – Standheizung – Timer – 14 Minuten.
-
- 5: Nav-1: „Wählen Sie das nächstgelegene Restaurant als Ziel aus, und leiten Sie die Zielführung ein.“
Navigation – Sonderziele – Restaurant – Ziel übernehmen – Zielführung starten.
- 6: Nav-2: „Sie möchten nach Aachen fahren. Geben Sie Aachen in das Navigationssystem ein.“
Navigation – Adresseingabe – Ort – Aachen – Ziel übernehmen

Version b

- 1: Ent-1: „Sie möchten den 2. Titel von der CD hören. Stellen Sie den Titel ein und schalten Sie den CD-Player ein“
Entertainment – CD – Titel 2 – Start.
- 2: Ent-2: „Sie möchten CD hören, aber nicht den aktuell eingestellten Titel, sondern den Titel danach. Stellen Sie diesen Titel ein und wechseln Sie dann ins Hauptmenü.“
Entertainment – CD – nächster Titel / Titel 3 - Hauptmenü
-
- 3: Ein-1: „Stellen Sie die Temperatur im Fahrzeug auf 24 Grad“
Einstellungen – Klima – Temperatur – 24 Grad.
- 4: Ein-2: „Stellen Sie die Laufzeit der Standheizung auf 23 Minuten.“
Einstellungen – Standheizung – Timer 23 Minuten.
-
- 5: Nav-1: „Wählen Sie die nächstgelegene Werkstatt als Ziel aus, und leiten Sie die Zielführung ein.“
Navigation – Sonderziele – Werkstatt – Ziel übernehmen – Zielführung starten.
- 6: Nav-2: „Sie möchten in den Ahornweg fahren. Geben Sie Ahornweg in das Navigationssystem ein.“
Navigation – Adresseingabe – Straße – Ahornweg – Ziel übernehmen

Version c

- 1: Ent-1: „Sie möchten Radio hören. Stellen Sie im Radio den 4. Sender ein und schalten Sie das Radio ein.“
Entertainment – Radio – Sender 4 – Einschalten.
- 2: Ent-2: „Schalten Sie im Radio einen Sender weiter und wechseln Sie dann ins Hauptmenü.“
Entertainment – Radio – nächster Sender / Sender 3 – Hauptmenü
-
- 3: Ein-1: „Vermindern sie die Luftzufuhr für den Fußbereich auf die niedrigste Stufe, die erste.“
Einstellungen – Klima – Scheibe – Stufe 1.
- 4: Ein-2: „Stellen Sie die Temperatur der Standheizung auf 28 Grad.“
Einstellungen – Standheizung – Temperatur – 28 Grad.
-

- 5: Nav-1: „Wählen Sie die nächstgelegene Tankstelle als Ziel aus, und leiten Sie die Zielführung ein.“
Navigation – Sonderziele – Tankstelle – Ziel übernehmen – Zielführung starten.
- 6: Nav-2: „Sie möchten in nach Abensberg fahren. Geben Sie Abensberg in das Navigationssystem ein.“
Navigation – Adresseingabe – Ort – Abensberg – Ziel übernehmen
- Version d**
- 1: Ent-1: „Sie möchten den 3. Titel von der CD hören. Stellen Sie den Titel ein und schalten Sie den CD-Player ein.“
Entertainment – CD – Titel 3 – Start.
- 2: Ent-2: „Sie möchten CD hören, allerdings nicht den aktuell eingestellten Titel, sondern den Titel danach. Stellen Sie diesen Titel ein und wechseln Sie dann einen Bildschirm zurück auf den Auswahlbildschirm des Entertainment-Bereichs.“
Entertainment – CD – nächster Titel / Titel 2 – Escape
-
- 3: Ein-1: „Stellen Sie die Temperatur im Fahrzeug auf 20 Grad.“
Einstellungen – Klima – Temperatur – 20 Grad.
- 4: Ein-2: „Stellen Sie die Temperatur der Standheizung auf 24 Grad.“
Einstellungen – Standheizung – Temperatur 24 Grad.
-
- 5: Nav-1: „Wählen Sie den nächstgelegenen Flughafen als Ziel aus, und leiten Sie die Zielführung ein.“
Navigation – Sonderziele – Flughafen – Ziel übernehmen – Zielführung starten.
- 6: Nav-2: „Sie möchten in die Abteistraße fahren. Geben Sie Abteistraße in das Navigationssystem ein.“
Navigation – Adresseingabe – Straße – Abteistraße – Ziel übernehmen

Anhang C

C 1: Aufgaben und Anweisungen an die Versuchsperson (Exp. 5-8)

- „Starten Sie die CD Wiedergabe.“
- „Stoppen Sie die CD Wiedergabe.“
- „Erhöhen Sie die Lautstärke.“
- „Öffnen Sie Ihr Adressbuch.“
- „Benutzen Sie das Telefon.“
- „Lesen Sie Ihre neuen SMS.“
- „Schalten Sie die Klimaanlage ein.“
- „Schalten Sie die Sitzheizung ein.“
- „Schalten Sie die Lüftung ein.“
- „Ändern Sie Einstellungen des Systems.“
- „Navigieren Sie eine Ebene zurück.“
- „Rufen Sie die Hilfe-Funktion auf.“

Literatur

- Accot, J. & Zhai, S. (1997). Beyond Fitts' law: Models for trajectory-based HCI tasks. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 295-302). New York: ACM Press.
- Ainsworth, W. A. (1997). Some approaches to automatic speech recognition. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 721-743). Oxford, UK: Blackwell Publishing.
- Akamatsu, M. (1992). The influence of combined visual and tactile information on finger and eye movements during shape tracing. *Ergonomics*, 35(5), 646-660.
- Akamatsu, M. & MacKenzie, I. S. (1996). Movement characteristics using a mouse with tactile and force feedback. *International Journal of Human-Computer Studies*, 45, 483-493.
- Akamatsu, M., MacKenzie, I. S. & Hasbrouq, T. (1995). A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38, 816-827.
- Akamatsu, M. & Sato, S. (1994). A multi-modal mouse with tactile and force feedback. *International Journal of Human-Computer Studies*, 40(3), 443-453.
- Akyol, S., Libuda, L. & Kraiss, K. F. (2001). Multimodale Benutzung adaptiver Kfz-Bordsysteme. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 137-154). Berlin: Springer.
- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832-843.
- Altinsoy, E. (2006). *Auditory-tactile interaction in virtual environments*. Aachen: Shaker.
- Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th ed.). New York: Worth Publishers, Inc.
- Arend, U., Muthig, K.-P. & Wandmacher, J. (1987). Evidence for global feature superiority in menu selection by icons. *Behaviour & Information Technology*, 6(4), 411-426.

- Aretz, A. J., Johannsen, C. & Ober, K. (1996). An empirical validation of subjective workload ratings. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 91-95). Santa Monica, CA: HFES.
- Arsenault, R. & Ware, C. (2000). Eye-hand co-ordination with force feedback. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 408-414). New York: ACM Press.
- Ashley, S. (2001). Simplifying controls. *Automotive Engineering International*, March 2001, 123-126.
- Aures, W. (1985). Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale. *Acustica*, 59, 130-141.
- Baber, C. (1994). Psychological aspects of conventional in-car warning devices. In N. A. Stanton (Ed.), *Human factors in alarm design* (pp. 193-205). London: Taylor & Francis.
- Baber, C. & Noyes, J. (1996). Automatic speech recognition in adverse environments. *Human Factors*, 38(1), 142-155.
- Baber, C. & Stammers, R. B. (1989). Is it natural to talk to computers? An experiment using the 'Wizard of Oz' technique. In E. D. Megaw (Ed.), *Contemporary Ergonomics 1989* (pp. 234-239). London: Taylor & Francis.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- Baddeley, A. D. (2000a). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.
- Baddeley, A. D. (2000b). The phonological loop and irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, 7(3), 544-549.
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, 7(2), 85-97.
- Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.
- Banbury, S. P., Jones, D. M. & Emery, L. (1999). Extending the 'irrelevant sound-effect': The effects of extraneous speech on aviation-related tasks. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Job design, product design and human-computer interaction* (Vol. 4, pp. 199-206). Aldershot: Ashgate.
- Banbury, S. P., Macken, W. J., Tremblay, S. & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, 43(1), 12-29.
- Beinhauer, W., Vilimek, R. & Richter, A. (2005). Eye-controlled applications - opportunities and limits [CD-ROM]. In *Proceedings of Human-Computer Interaction International 2005*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Benbasat, I. & Todd, P. (1993). An experimental investigation of interface design alternatives: Icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies*, 38(3), 369-402.
- Bengler, K. (2001). Aspekte der multimodalen Bedienung und Anzeige im Automobil. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 195-205). Berlin: Springer.
- Benoît, J., Martin, C., Pelachaud, C., Schomaker, L. & Suhm, B. (2000). Audio-visual and multimodal speech-based systems. In *Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation* (pp. 102-203). Boston: Kluwer Academic Publishers.
- Bernotat, R. (1970). Anthropotechnik in der Fahrzeugführung. *Ergonomics*, 13(3), 353-377.
- Bernsen, N. O. (1994a). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers*, 6, 347-371.
- Bernsen, N. O. (1994b). *A revised generation of the taxonomy of output modalities* (CCI Working Papers in Cognitive Science and HCI No. WPCS-94-7): Roskilde University, Centre for Cognitive Science.
- Bernsen, N. O. (2002). Multimodality in language and speech systems - from theory to design support tool. In B. Granström, D. House & I. Karlsson (Eds.), *Multimodality in language and speech systems* (pp. 93-148). Dordrecht: Kluwer Academic Publishers.
- Bewley, W. L., Roberts, T. L., Schroit, D. & Verplank, W. L. (1983). Human factors testing in the design of Xerox's 8010 'Star' office workstation. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72-77). Boston, MA: ACM Press.
- Beyer, L. & Weiss, T. (2001). Elementareinheiten des somatosensorischen Systems als physiologische Basis der taktil-haptischen Wahrnehmung. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 25-38). Basel: Birkhäuser Verlag.
- Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.
- Bolt, R. A. (1980). Put that there: Voice and gesture at the graphics interface. *ACM Computer Graphics*, 14, 262-270.
- Bolt, R. A. (1982). Eyes at the interface. In *Proceedings of the 1982 conference on human factors in computing systems* (pp. 360-362). New York: ACM Press.
- Bos, E., Huls, C. & Claassen, W. (1994). EDWARD: Full integration of language and action in a multimodal user interface. *International Journal of Human-Computer Studies*, 40, 473-495.
- Boyce, S. J. (2000). Natural spoken dialogue systems for telephony applications. *Communications of the ACM*, 43(9), 29-34.
- Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces*. Unpublished dissertation, University of York, York, UK.

- Brewster, S. A. (1997). Using non-speech sound to overcome information overload. *Displays*, 17, 179-189.
- Brewster, S. A. (1998a). The design of sonically-enhanced widgets. *Interacting with Computers*, 11(2), 211-235.
- Brewster, S. A. (1998b). Sonically-enhanced drag and drop. In *Proceedings of the International Conference on Auditory Display (ICAD'98)*. London: British Computer Society.
- Brewster, S. A. (1998c). Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction*, 5(2), 224-259.
- Brewster, S. A. (2000). *Principles for improving interaction in telephone-based interfaces*. Retrieved April 12, 2005, from <http://www.dcs.gla.ac.uk/~stephen/research/telephone/simulator.shtml>
- Brewster, S. A. (2003). Nonspeech auditory output. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 220-239). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brewster, S. A. & Brown, L. M. (2004). Tactons: Structured tactile messages for non-visual information display. In *Proceedings of Australasian User Interface Conference 2004* (pp. 15-23). Dunedin, New Zealand: Australian Computer Society.
- Brewster, S. A. & Crease, M. G. (1997). Making menus musical. In *Proceedings of IFIP Interact'97* (pp. 389-396). London: Chapman & Hall.
- Brewster, S. A. & Crease, M. G. (1999). Correcting menu usability problems with sound. *Behaviour & Information Technology*, 18(3), 165-177.
- Brewster, S. A., Raty, V.-P. & Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy. In A. Sasse, R. Cunningham & R. Winder (Eds.), *Proceedings of BCS HCI'96* (pp. 169-183). London: Springer.
- Brewster, S. A., Wright, P. C. & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In S. Ashlung, K. Mullet, A. Henderson, E. Hollnagel & T. White (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 222-227). New York: ACM Press.
- Brewster, S. A., Wright, P. C. & Edwards, A. D. N. (1994). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), *Auditory display. Sonification, audification and auditory interfaces*. (pp. 471-498). Reading, MA: Addison-Wesley.
- Brewster, S. A., Wright, P. C. & Edwards, A. D. N. (1995). Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies*, 43(2), 153-175.
- Bridges, A. M. & Jones, D. M. (1996). Word dose in the disruption of serial recall by irrelevant speech: Phonological confusions or changing state? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 49(4), 919-939.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

- Broadbent, D. E. (1981). The effects of moderate levels of noise on human performance. In J. Tobias & E. Schubert (Eds.), *Hearing: Research and Theory* (Vol. 1, pp. 165-186). New York: Academic Press.
- Brown, L. M., Brewster, S. A. & Purchase, H. C. (2005). A first investigation into the effectiveness of Tactons. In *Proceedings of WorldHaptics* (pp. 167-176). Pisa, Italy: IEEE.
- Bryk, A. S. & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, 104(3), 396-404.
- Bubb, H. (2001). Haptik im Kraftfahrzeug. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 155-175). Berlin: Springer.
- Bubb, H. (2003). Fahrerassistenz - primär ein Beitrag zum Komfort oder für die Sicherheit? In *VDI-Berichte* (Vol. 1768, Der Fahrer im 21. Jahrhundert, pp. 25-44). Düsseldorf: VDI-Verlag.
- Buckner, D. N. & McGrath, J. J. (Eds.). (1963). *Vigilance: a symposium*. New York: McGraw-Hill.
- Bucur, B., Allen, P. A., Sanders, R. E., Ruthruff, E. & Murphy, M. D. (2005). Redundancy gain and coactivation in bimodal detection: Evidence for the preservation of coactive processing in older adults. *Journals of Gerontology: Psychological Sciences and Social Sciences*, 60(5), 279-282.
- Bullinger, H.-J. (1994). *Ergonomie. Produkt- und Arbeitsplatzgestaltung*. Stuttgart: Teubner.
- Bullinger, H.-J., Kern, P. & Muntzinger, W. F. (1987). Design of controls. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 577-600). Oxford, UK: John Wiley & Sons.
- Bunt, H. (1998). Issues in multimodal human-computer communication. In H. Bunt, R. J. Beun & T. Borghuis (Eds.), *Multimodal human-computer communication: Systems, techniques, and experiments* (pp. 1-12). Berlin: Springer.
- Burdea, G. C. (1996). *Force and touch feedback for virtual reality*. New York: John Wiley & Sons.
- Buxton, W. (1986). There's more to interaction than meets the eye: Some issues in manual input. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 319-337). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Buxton, W. A. (1994). Human skills in interface design. In L. W. MacDonald & J. Vince (Eds.), *Interacting with virtual environments* (pp. 1-13). New York: Wiley.
- Byers, J. C., Bittner, A. C. & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety, I* (pp. 481-485). London: Taylor & Francis.
- Byers, J. C., Bittner, A. C., Hill, S. G., Zaklad, A. L. & Christ, R. E. (1988). Workload assessment of a remotely piloted vehicle (RPV) system. In *Proceedings of the Human Factors Society 32nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

- Cahill, M.-C. (1975). Interpretability of graphic symbols as a function of context and experience factors. *Journal of Applied Psychology*, 60(3), 376-380.
- Cahill, M.-C. (1976). Design features of graphic symbols varying in interpretability. *Perceptual and Motor Skills*, 42(2), 647-653.
- Campbell, J. L., Carney, C. & Kantowitz, B. H. (1997). *Human Factors design guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO)* (Technical Report FHWA-RD-98-057). Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration.
- Campbell, J. L., Richman, J. B., Carney, C. & Lee, J. D. (2004). *In-vehicle display icons and other information elements. Volume I: Guidelines* (Technical Report No. FHWA-RD-03-065). McLean, VA: U.S. Department of Transportation, Federal Highway Administration.
- Card, S. K., Moran, T. P. & Newell, A. (1986). The model human processor: An engineering model of human performance. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2: Cognitive processes and performance, chap. 45, pp. 1-35). Oxford, UK: John Wiley & Sons.
- Carroll, J. M. & Mack, R. L. (1999). Metaphor, computing systems, and active learning. *International Journal of Human-Computer Studies*, 51(2), 385-403.
- Chan, J. S. & Spence, C. (2003). Presenting multiple auditory signals using multiple sound cards in Visual Basic 6.0. *Behavior Research Methods, Instruments, & Computers*, 35(1), 125-128.
- Charwat, H. J. (1994). *Lexikon der Mensch-Maschine-Kommunikation* (2nd ed.). München: Oldenbourg.
- Cher, C., Campbell, J. L. & Mitchell, E. A. (1998). *In-vehicle display icons and other information elements: literature review* (Technical Report FHWA-RD-98-164). McLean, VA: U.S. Department of Transportation, Federal Highway Administration.
- Chun, M. M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36(1), 28-71.
- Clark, R. & Craig, T. G. (1992). Research and theory on multimedia learning effects. In M. Giardina (Ed.), *Interactive multimedia learning environments. Human factors and technical considerations on design issues* (pp. 19-30). Berlin: Springer.
- Cockburn, A. & Brewster, S. A. (2005). Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48(9), 1129-1150.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12), 997-1003.
- Cohen, P. R., McGee, D., Oviatt, S. L., Wu, L., Clow, J., King, R., Julier, S. & Rosenblum, L. (1999). Multimodal interaction for 2D and 3D Environments. *IEEE Computer Graphics and Applications*, 19(4), 10-13.
- Colle, H. A. & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15, 17-31.
- Craig, J. C. (1977). Vibrotactile pattern perception: Extraordinary observers. *Science*, 196(4288), 450-452.

- Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993). Wizard of Oz studies - Why and how. In W. D. Gray, W. E. Hefley & D. Murray (Eds.), *Proceedings of the 1st international conference on intelligent user interfaces* (pp. 193-200). New York: ACM Press.
- Dahmen-Zimmer, K., Huber, M., Kaiser, I., Kostka, M., Scheufler, I., Piechulla, W., Praxenthaler, M. & Zimmer, A. (1999). *Definition und Validierung von Kriterien für die Ablenkungswirkung von MMI-Lösungen* (Endbericht MoTiV-Teilprojekt MMI AP/5). Regensburg: Universität Regensburg, Lehrstuhl für Angewandte und Experimentelle Psychologie.
- Dahmen-Zimmer, K. & Zimmer, A. (1997). *Situationsbezogene Sicherheitskenngrößen im Straßenverkehr* (Berichte der Bundesanstalt für Straßenwesen, No. M 78). Bergisch Gladbach: BASt.
- Darves, C. & Oviatt, S. L. (2002). Adaptation of users' spoken dialogue patterns in a conversational interface. In J. Hansen & B. Pellom (Eds.), *Proceedings of the 7th International Conference on Spoken Language Processing ICSLP'2002* (pp. 561-564). Denver, CO: Casual Prod. Ltd.
- Dennerlein, J. T. & Yang, M. C. (2001). Haptic force-feedback devices for the office computer: Performance and musculoskeletal loading issues. *Human Factors*, 43(2), 278-286.
- Deutsch, J. A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 80-90.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, 39(2), 197-215.
- Diederich, A. & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8), 1388-1404.
- Dillon, T. W. & Norcio, A. F. (1997). User performance and acceptance of a speech-input interface in a health assessment task. *International Journal of Human-Computer Studies*, 47, 591-602.
- Dingus, T. A., McGehee, D. V., Manakkal, N. & Jahns, S. K. (1997). Human factors field evaluation of automotive headway maintenance/collision warning devices. *Human Factors*, 39(2), 216-229.
- Dix, A., Finlay, J., Abowd, G. D. & Beale, R. (2004). *Human-Computer Interaction* (3rd ed.). Harlow: Prentice Hall.
- Driver, J. & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179-220). Oxford, UK: Oxford University Press.
- Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4), 455-470.
- Duncan, J. & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458.

- Easterby, R. S. (1970). The perception of symbols for machine displays. *Ergonomics*, 13(1), 149-158.
- Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics*, 25(4), 202-210.
- Edworthy, J. (1997). Cognitive compatibility and warning design. *International Journal of Cognitive Ergonomics*, 1(3), 193-209.
- Edworthy, J. & Adams, A. (1996). *Warning design: A research prospective*. London: Taylor & Francis.
- Edworthy, J., Loxley, S. & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33(2), 205-231.
- Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. M. Meshkati (Eds.), *Human mental workload* (pp. 41-62). Amsterdam: North Holland B.V.
- Egido, C. & Patterson, J. (1988). Pictures and category labels as navigational aids for catalog browsing. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 127-132). Washington, D.C.: ACM Press.
- Ellermeier, W. & Hellbrück, J. (1998). Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1406-1414.
- Ellis, N. R. (1991). Automatic and effortful processes in memory for spatial location. *Bulletin of the Psychonomic Society*, 29(1), 28-30.
- Engelkamp, J. (1991). *Das menschliche Gedächtnis. Das Erinnern von Sprache, Bildern und Handlungen* (2nd ed.). Göttingen: Hogrefe.
- van Erp, J. B. F. (2002). Guidelines for the use of vibro-tactile displays in human-computer interaction. In S. A. Wall, B. Riedel, A. Crossan & M. R. McGee (Eds.), *Proceedings of Eurohaptics 2002* (pp. 18-22). Edinburgh, UK: University of Edinburgh.
- van Erp, J. B. F., van Veen, H. A. H. C., Jansen, C. & Dobbins, T. (2005). Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception*, 2(2), 106-117.
- ETSI EG 202 048. (2002). *Human Factors (HF); Guidelines on the multimodality of icons, symbols and pictograms*. Sophia-Antipolis Cedex, France: ETSI. Retrieved December 10, 2006, from http://docbox.etsi.org/EC_Files/EC_Files/eg_202048v010101p.pdf.
- ETSI EG 202 191. (2003). *Human Factors (HF); Multimodal interaction, communication and navigation guidelines*. Sophia-Antipolis Cedex, France: ETSI. Retrieved December 10, 2006, from http://docbox.etsi.org/EC_Files/EC_Files/eg_202191v010101p.pdf.
- Evans, L. (1991). *Traffic safety and the driver*. New York: Van Nostrand Reinhold.
- Fairclough, S. H. & Maternaghan, M. (1993). Changes in drivers' visual behaviour due to the introduction of complex versus simple route navigation information. In D.

- Brogan, A. G. Gale & K. Carr (Eds.), *Visual Search 2* (pp. 419-432). London: Taylor & Francis.
- Familant, M. E. & Detweiler, M. C. (1993). Iconic reference: Evolving perspectives and an organizing framework. *International Journal of Man-Machine Studies*, 39(5), 705-728.
- Färber, B. (1987). *Geteilte Aufmerksamkeit: Grundlagen und Anwendung im motorisierten Straßenverkehr*. Köln: TÜV Rheinland.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381-391.
- Fröhlich, P. & Pucher, M. (2005). *Combining speech and sound in the user interface*. Retrieved August 15, 2005, from http://userver.ftw.at/~frohlich/icad05_workshop_outline.htm
- Furnas, G. W., Landauer, T. K., Gomez, L. M. & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11), 964-971.
- Gardner-Bonneau, D. (1999). Guidelines for speech-enabled IVR application design. In D. Gardner-Bonneau (Ed.), *Human factors and voice interactive systems* (pp. 147-161). Boston: Kluwer Academic Publishers.
- Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Lawrence Erlbaum Associates.
- Gaver, W. A. (1989). The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4(1), 67-94.
- Geer, D. (2004). Will gesture-recognition technology point the way? *Computer*, 37(10), 20-23.
- Geiser, G. (1990). *Mensch-Maschine-Kommunikation*. München: Oldenbourg.
- Gelau, C. (2004). Fahrerablenkung durch Informations- und Kommunikationssysteme im Fahrzeug: Auswirkungen auf das Fahrerverhalten und die Verkehrssicherheit. In B. Schlag (Ed.), *Verkehrspsychologie. Mobilität - Sicherheit - Fahrerassistenz* (pp. 297-316). Lengerich: Papst Science Publishers.
- Gelau, C., Keinath, A., Baumann, M., Bengler, K. & Krems, J. F. (1999). Die Okklusionsmethode als Verfahren zur Bewertung von visuellen Displaydarstellungen in Kraftfahrzeugen. *Zeitschrift für Arbeitswissenschaft*, 53, 41-57.
- Geschwind, N. (1979). Specializations of the human brain. *Scientific American*, 241(3), 180-199.
- Gibson, J. J. (1962). Observations on active touch. *Psychological Review*, 69(6), 477-491.
- Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gielen, S. C., Schmidt, R. A. & van den Heuvel, P. J. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics*, 34(2), 161-168.

- Gittins, D. (1986). Icon-based human-computer interaction. *International Journal of Man-Machine Studies*, 24(6), 519-543.
- Gladstones, W. H., Regan, M. A. & Lee, R. B. (1989). Division of attention: The single-channel hypothesis revisited. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 41(1), 1-17.
- Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42(3), 237-288.
- Göbel, M., Luczak, H., Springer, J., Hedicke, V. & Rötting, M. (1995). Tactile feedback applied to computer mice. *International Journal of Human-Computer Interaction*, 7(1), 1-24.
- Goldstein, E. B. (1997). *Wahrnehmungspsychologie. Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.
- Gong, L., Nass, C., Simard, C. & Takhteyev, Y. (2001). When non-human is better than semi-human: Consistency in speech interfaces. In M. J. Smith, G. Salvendy, D. Harris & R. J. Koubek (Eds.), *Proceedings of HCI International 2001. Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (Vol. 1, pp. 290-294). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gopher, D. & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2: Cognitive processes and performance, chap. 41, pp. 1-49). Oxford, UK: John Wiley & Sons.
- Graham, R., Aldridge, L., Carter, C. & Lansdown, T. C. (1999). The design of in-car speech recognition interfaces for usability and user acceptance. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Job design, product design and human-computer interaction* (Vol. 4, pp. 313-320). Aldershot: Ashgate.
- Grasso, M. & Finin, T. W. (1997). Task integration in multimodal speech recognition environments. *Crossroads*, 3(3), 19-22.
- Grasso, M. A., Ebert, D. S. & Finin, T. W. (1998). The integrality of speech in multimodal interfaces. *ACM Transactions on Computer-Human Interaction*, 5(4), 303-325.
- Green, A. J. K. & Barnard, P. J. (1990). Iconic interfacing: The role of icon distinctiveness and fixed or variable screen locations. In D. Diaper, D. Gilmore, G. Cockton & B. Shackel (Eds.), *Human-Computer Interaction - INTERACT '90* (pp. 457-462). Amsterdam: Elsevier.
- Green, P. (1981). Displays for automotive instrument panels: Production and rating of symbols. *HSRI Research Review*, 12(1), 12.
- Green, P. & Pew, R. W. (1978). Evaluating pictographic symbols: An automotive application. *Human Factors*, 20(1), 103-114.
- Grunwald, M. (2001). Begriffsbestimmung zwischen Psychologie und Physiologie. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 1-14). Basel: Birkhäuser Verlag.

- Grunwald, M. & Krause, F. (2001). Haptik-Design im Fahrzeugbau. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 171-176). Basel: Birkhäuser Verlag.
- Guastello, S. J., Traut, M. & Korienek, G. (1989). Verbal versus pictorial representations of objects in a human-computer interface. *International Journal of Man-Machine Studies*, 31(1), 99-120.
- Guski, R. (1997). Psychological methods for evaluating sound quality and assessing acoustic information. *Acustica united with Acta Acustica*, 83(5), 765-774.
- Haberland, N., Kanthak, S., Overmann, J., Schütz, K. & Welling, L. (1998). Sprachunterricht. Wie funktioniert computerbasierte Spracherkennung. *c't*, 1998/5, 120-125.
- Haigney, D. & Westerman, S. J. (2001). Mobile (cellular) phone use and driving: A critical review of research methodology. *Ergonomics*, 44(2), 132-143.
- Haramundanis, K. (1996). Why icons cannot stand alone. *ACM SIGDOC Asterisk Journal of Computer Documentation*, 20(2), 1-8.
- Hart, S. G. & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. M. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland B.V.
- Hart, S. G. & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.), *MANPRINT: An approach to systems integration* (pp. 257-296). New York: Van Nostrand Reinhold.
- Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108(3), 356-388.
- Hedicke, V. (2000). Multimodalität in Mensch-Maschine-Schnittstellen. In K. P. Timpe, T. Jürgensohn & H. Kolrep (Eds.), *Mensch-Maschine-Systemtechnik. Konzepte, Modellierung, Gestaltung, Evaluation* (pp. 203-232). Düsseldorf: Symposion Publishing GmbH.
- Heisterkamp, P. (2003). "Do not attempt to light with match!" Some thoughts on progress and research goals in spoken dialog systems. In *Eurospeech 2003* (pp. 2897-2900). Switzerland: ISCA.
- Helander, M. G. (1987). Design of visual displays. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 507-548). Oxford, UK: John Wiley & Sons.
- Hempel, T. (2001). *Untersuchungen zur Korrelation auditiver und instrumenteller Messergebnisse für die Bewertung von Fahrzeuginnenraumgeräuschen als Grundlage eines Beitrags zur Klassifikation von Hörereignissen*. München: Herbert Utz Verlag.
- Hempel, T. (2002). Auditive Rückmeldungen und Produktgeräuschgestaltung: Zur Steigerung der Gebrauchstauglichkeit technischer Produkte. *Forum Ware*, 30, 5-7.
- Hempel, T. (2003). Sound-Design für Benutzungsschnittstellen: zur semantischen Betrachtung von Geräuschen. *Forum Ware*, 31, 15-19.

- Hempel, T. & Altinsoy, E. (2005). Multimodal user interfaces: Designing media for the auditory and the tactile channel. In R. W. Proctor & K.-P. L. Vu (Eds.), *Handbook of human factors in web design* (pp. 134-155). Mahwah, NJ: Lawrence Erlbaum Associates.
- Herfet, T., Kirste, T. & Schnaider, M. (2001). Elektronische Multimodale Bedien- und Service-Assistenz - Das Leitprojekt EMBASSI. *i-com*, 0(1), 12-16.
- Hills, B. L. (1980). Vision, visibility, and perception in driving. *Perception*, 9(2), 183-216.
- Hinckley, K. (2003). Input technologies and techniques. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 151-168). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hirst, S. & Graham, R. A. (1997). The format and presentation of collision warnings. In Y. I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces* (pp. 203-219). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Horton, W. (1994). *The icon book: Visual symbols for computer systems and documentation*. New York: Wiley.
- Howarth, C. I. & Treisman, M. (1958). The effect of warning interval on the electric phosphene and auditory thresholds. *Quarterly Journal of Experimental Psychology*, 10, 130-141.
- Hoyos, C. Graf. (1990). Menschliches Handeln in technischen Systemen. In C. Graf Hoyos (Ed.), *Enzyklopädie der Psychologie. Ingenieurpsychologie* (Vol. D/III/2, pp. 1-30). Göttingen: Verlag für Psychologie, Hogrefe.
- ISO 7000. (2004). *Graphical symbols for use on equipment -- index and synopsis*. Geneva: ISO copyright office.
- ISO 7250. (1997). *Wesentliche Maße des menschlichen Körpers für die technische Gestaltung*. Berlin: Beuth.
- ISO 9186. (2001). *Graphical symbols -- Test methods for judged comprehensibility and for comprehension*. Geneva: ISO copyright office.
- ISO 9241-10. (1996). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 10: Grundsätze der Dialoggestaltung*. Berlin: Beuth.
- ISO 9241-11. (1998). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze*. Berlin: Beuth.
- ISO 9241-12. (1998). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Teil 12: Informationsdarstellung*. Berlin: Beuth.
- ISO 9241-110. (2006). *Ergonomie der Mensch-System-Interaktion. Teil 110: Grundsätze der Dialoggestaltung*. Berlin: Beuth.
- Ives, B. (1982). Graphical user interfaces for business information systems. *Management Information Systems Quarterly*, 6(5), 15-48.

- Iwata, H. (2003). Haptic interfaces. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 206-219). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jacko, J. A., Emery, V. K., Edwards, P. J., Ashok, M., Barnard, L., Kongnakorn, T., Moloney, K. P. & Sainfort, F. (2004). The effects of multimodal feedback on older adults' task performance given varying levels of computer experience. *Behaviour & Information Technology*, 23(4), 247-264.
- Jacko, J. A., Moloney, K. P., Kongnakorn, T., Barnard, L., Edwards, P. J., Leonard, V. K., Sainfort, F. & Scott, I. U. (2005). Multimodal feedback as a solution to ocular disease-based user performance decrements in the absence of functional visual loss. *International Journal of Human-Computer Interaction*, 18(2), 183-218.
- Jacob, R. J., Leggett, J. J., Myers, B. A. & Pausch, R. (1993). Interaction styles and input/output devices. *Behaviour & Information Technology*, 12(2), 69-79.
- Jenkins, W. O. (1947). The tactual discrimination of shapes for coding aircraft-type controls. In P. M. Fitts (Ed.), *Psychological research on equipment design* (pp. 199-205). Washington, D.C.: US Army Air Forces Aviation Psychology Program.
- Johannsen, G. (1990). Fahrzeugführung. In C. Graf Hoyos (Ed.), *Enzyklopädie der Psychologie. Ingenieurpsychologie* (Vol. D/III/2, pp. 426-454). Göttingen: Verlag für Psychologie, Hogrefe.
- Johannsen, G. & Rouse, W. B. (1979). Mathematical concepts for modeling human behavior in complex man-machine systems. *Human Factors*, 21(6), 733-747.
- Johannson, E., Engström, J., Cherri, C., Nodari, E., Toffetti, A., Schindhelm, R. & Gelau, C. (2004). *Review of existing techniques and metrics for IVIS and ADAS assessment* (Report No. AIDE IST-1-507674-IP D2.2.1). AIDE Consortium, www.aide-eu.org.
- John, M. (2001). Historisch-philosophischer Exkurs über den Tastsinn. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 15-24). Basel: Birkhäuser Verlag.
- Jones, D. M., Beaman, P. & Macken, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209-237). Hove, UK: Psychology Press.
- Jones, D. M. & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 369-381.
- Jones, D. M., Madden, C. & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 44(4), 645-669.
- Jones, D. M., Miles, C. & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, 4(2), 89-108.
- Jones, D. M. & Morris, N. (1992). Irrelevant speech and cognition. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 1: The physical environment, pp. 29-53). San Diego: Academic Press, Inc.

- Jones, D. M. & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, 7(3), 550-558.
- Kaber, D. B., Wright, M. C. & Sheik-Nainar, M. A. (2006). Investigation of multi-modal interface features for adaptive automation of a human-robot system. *International Journal of Human-Computer Studies*, 64(6), 527-540.
- Kacmar, C. J. & Carey, J. M. (1991). Assessing the usability of icons in user interfaces. *Behaviour & Information Technology*, 10(6), 443-457.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kamm, C. (1994). User interfaces for voice applications. In D. Roe & J. Wilpon (Eds.), *Voice communication between humans and machines* (pp. 422-442). Washington, D.C.: National Academy of Sciences Press.
- Kamm, C. & Helander, M. G. (1997). Design issues for interfaces using voice input. In M. G. Helander, T. K. Landauer & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., pp. 1043-1059). Amsterdam: Elsevier.
- Kannheiser, W. (1990). Methoden der Ingenieurpsychologie. In C. Graf Hoyos (Ed.), *Enzyklopädie der Psychologie. Ingenieurpsychologie* (Vol. D/III/2, pp. 55-91). Göttingen: Verlag für Psychologie, Hogrefe.
- Kantowitz, B. H. & Knight, J. L. (1976). Testing tapping timesharing: II Auditory secondary task. *Acta Psychologica*, 40(5), 343-362.
- Karpicke, J. & Pisoni, D. B. (2000). *Memory span and sequence learning using multi-modal stimulus patterns: Preliminary findings in normal-hearing adults* (Research on spoken language processing, progress report No. 24). Bloomington, IN: Indiana University.
- Keates, S. & Robinson, P. (1999). Gestures and multimodal input. *Behaviour & Information Technology*, 18(1), 36-44.
- Kebeck, G., Cieler, S. & Pohlmann, S. (1997). *Vergessene Ergonomie. Gedächtnispsychologische Aspekte der Gestaltung von Displays*. Münster: Lit Verlag.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Killeen, P. R. (2005a). An alternative to null-hypothesis significance tests. *Psychological Science*, 16(5), 345-353.
- Killeen, P. R. (2005b). Replicability, confidence, and priors. *Psychological Science*, 16(12), 1009-1012.
- Klatzky, R. L. & Lederman, S. J. (1987). The intelligent hand. In G. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 121-151). New York: Academic Press.
- Klatzky, R. L. & Lederman, S. J. (1992). Stages of manual exploration in haptic object identification. *Perception & Psychophysics*, 52(6), 661-670.
- Klatzky, R. L. & Lederman, S. J. (2003). Touch. In I. B. Weiner, A. F. Healy & R. W. Proctor (Eds.), *Handbook of Psychology, Vol. 4. Experimental Psychology* (pp. 147-176). New York: John Wiley & Sons.

- Kobus, D. A., Russotti, J., Schlichting, C., Haskell, G., Carpenter, S. & Wojtowicz, J. (1986). Multimodal detection and recognition performance of sonar operators. *Human Factors*, 28(1), 23-29.
- Kohlrausch, A. G. & van de Par, S. L. (2005). Audio visual interaction in the context of multi-media applications. In J. Blauert (Ed.), *Communication acoustics*. Berlin: Springer.
- Kolb, B. & Wishaw, I. Q. (1996). *Neuropsychologie* (2nd ed.). Heidelberg: Spektrum Akademischer Verlag.
- Kraiss, K. F. (1985). *Fahrzeug- und Prozeßführung. Kognitives Verhalten des Menschen und Entscheidungshilfen*. Berlin: Springer.
- Krauss, R. M. & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 93-116). New York: Oxford University Press.
- Krems, J. F., Keinath, A., Baumann, M., Bengler, K. & Gelau, C. (2000). Die Bewertung von visuellen Displaydarstellungen in Kraftfahrzeugen: Vor- und Nachteile der Okklusionsmethode. In Bundesanstalt für Straßenwesen (Ed.), *Informations- und Assistenzsysteme im Auto benutzergerecht gestalten* (pp. 25-32). Bergisch Gladbach: Wirtschaftsverlag.
- Krems, J. F., Keinath, A., Baumann, M. & Jahn, G. (2004). Die Okklusionsmethode: Ein einfaches und valides Verfahren zur Bewertung der visuellen Beanspruchung von Zweitaufgaben? In B. Schlag (Ed.), *Verkehrspsychologie. Mobilität - Sicherheit - Fahrerassistenz* (pp. 335-349). Lengerich: Papst Science Publishers.
- Lansdale, M. W. (1988). On the memorability of icons in an information retrieval task. *Behaviour & Information Technology*, 7(2), 131-152.
- Lansdale, M. W., Simpson, M. & Stroud, T. R. (1990). A comparison of words and icons as external memory aids in an information retrieval task. *Behaviour & Information Technology*, 9(2), 111-131.
- Larson, J. (2006). *Common sense suggestions for developing multimodal user interfaces (W3C Working Group Note)*. Retrieved December 10, 2006, from <http://www.w3.org/TR/2006/NOTE-mmi-suggestions-20060911/>
- LeCompte, D. C., Neely, C. B. & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the Irrelevant Speech Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 472-483.
- Leplâtre, G. & Brewster, S. A. (2000). Designing non-speech sounds to support navigation in mobile phone menus. In *Proceedings of the International Conference on Auditory Display ICAD 2000* (pp. 190-199). Atlanta, USA: ICAD.
- Lewis, C. & Norman, D. A. (1986). Designing for error. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 411-432). Mahwah, NJ: Lawrence Erlbaum Associates.
- Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on driver's performance in advanced traveller information systems. *Ergonomics*, 44, 425-442.

- Luczak, H., Roetting, M. & Oehme, O. (2003). Visual displays. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 187-205). Mahwah, NJ: Lawrence Erlbaum Associates.
- MacKenzie, I. S. (1989). A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior*, 21(3), 323-330.
- MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7(1), 91-139.
- MacKenzie, I. S. & Oniszczak, A. (1997). The tactile touchpad. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems - CHI '97* (pp. 309-310). New York: ACM Press.
- MacKenzie, I. S. & Oniszczak, A. (1998). A comparison of three selection techniques for touchpads. In *Proceedings of the ACM Conference on Human Factors in Computing Systems - CHI '98* (pp. 336-343). New York: ACM Press.
- Maggioni, C. & Kämmerer, B. (1998). GestureComputer - history, design and applications. In R. Cipolla & A. Pentland (Eds.), *Computer vision for human-machine interaction* (pp. 23-52). Cambridge, UK: Cambridge University Press.
- Martin, J.-C. (1997, August). *Towards "intelligent" cooperation between modalities: The example of a system enabling multimodal interaction with a map*. Paper presented at the IJCAI'97 workshop on intelligent multimodal systems, Nagoya, Japan.
- Martin, J.-C. (1998, May). *Types of cooperation and referenceable objects: Implications on annotation schemas for multimodal language resources*. Paper presented at the LREC 2000 pre-conference workshop, Athens, Greece.
- Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. In H. Strasser, K. Kluth, H. Rausch & H. Bubb (Eds.), *Quality of work and products in enterprises of the future* (pp. 57-60). Stuttgart: Ergonomia Verlag.
- Maybury, M. T. (1998). Toward cooperative multimedia interaction. In H. Bunt, R. J. Beun & T. Borghuis (Eds.), *Multimodal human-computer communication: Systems, techniques, and experiments* (pp. 13-38). Berlin: Springer.
- McGee, M. R. (1999). A haptically enhanced scrollbar: Force feedback as a means of reducing the problems associated with scrolling. In *Proceedings for the First PHANToM Users Research Symposium (PURS'99)* (pp. 17-20). Heidelberg: DKFZ.
- McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 246, 746-748.
- Megaw, E. (1992). The visual environment. In A. P. Smith & D. M. Jones (Eds.), *Handbook of human performance* (Vol. 1, pp. 261-296). London: Academic Press.
- Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E. & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, 95(3), 340-370.
- Michon, J. A. (1985). A critical view on driver behavior models: What do we know, what should we do? In L. Evans & R. Schwing (Eds.), *Human behavior and traffic safety* (pp. 485-520). New York: Plenum Press.

- Miller, G. A., Heise, G. A. & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 41, 329-335.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14(2), 247-279.
- Moray, N. (1982). Subjective mental workload. *Human Factors*, 24(1), 25-40.
- Mordkoff, J. T. & Miller, J. (1993). Redundancy gains and coactivation with two different targets: The problem of target preferences and the effects of display frequency. *Perception & Psychophysics*, 53(5), 527-535.
- Muthig, K.-P. (1990). Informationsaufnahme und Informationsverarbeitung. In C. Graf Hoyos (Ed.), *Enzyklopädie der Psychologie. Ingenieurpsychologie* (Vol. D/III/2, pp. 92-120). Göttingen: Verlag für Psychologie, Hogrefe.
- National Research Council. (1997). *Tactical display for soldiers. Human factors considerations*. Washington, D.C.: National Academy Press.
- Navon, D. & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214-255.
- Neal, J. G. & Shapiro, S. C. (1991). Intelligent multimedia interface technology. In J. W. Sullivan & S. W. Tyler (Eds.), *Intelligent user interfaces* (pp. 11-43). New York: ACM Press.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7(3), 403-423.
- Neuss, R. (2000). *Usability Engineering als Ansatz zum Multimodalen Mensch-Maschine-Dialog*. Unpublished dissertation, Technische Universität München.
- Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 19(2), 155-160.
- Nickerson, R. S. & Stevens, K. N. (1972). Teaching speech to the deaf: can a computer help. In *Proceedings of the ACM annual conference* (Vol. 1, pp. 240-251). Boston: ACM Press.
- Niedermaier, F. B. (2003). *Entwicklung und Bewertung eines Rapid-Prototyping Ansatzes zur multimodalen Mensch-Maschine-Interaktion im Kraftfahrzeug*. Unpublished dissertation, Technische Universität München.
- Nielsen, J. (1993). Noncommand user interfaces. *Communications of the ACM*, 36(4), 83-99.
- Nigay, L. & Coutaz, J. (1993). A design space for multimodal systems - concurrent processing and data fusion. In *INTERCHI '93: Proceedings of the Conference on Human Factors and Computing Systems* (pp. 172-178). New York: ACM Press.
- Nigay, L., Jambon, F. & Coutaz, N. (1995, April). *Formal specification of multimodality*. Paper presented at the CHI'95 workshop on formal specification of user interfaces, Denver, Colorado.
- Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, 75(6), 522-536.

- Norman, D. A. (1988). *Psychology of everyday things*. New York: Basic Books.
- Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. In D. E. Broadbent, J. Reason & A. D. Baddeley (Eds.), *Human factors in hazardous situations* (pp. 137-145). New York: Clarendon Press.
- Norman, D. A. (1998). *The invisible computer: Why good products can fail, the personal computer is so complex, and information appliances are the solution*. Cambridge: MIT Press.
- Norman, D. A. & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction*. Mahwah, NJ: Lawrence Erlbaum Associates.
- O'Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2: Cognitive processes and performance, chap. 42, pp. 1-49). Oxford, UK: John Wiley & Sons.
- Oakley, I. (2003). *Haptic augmentation of the cursor: Transforming virtual actions into physical actions*. Unpublished dissertation, University of Glasgow, Glasgow, UK.
- Oakley, I., Adams, A., Brewster, S. A. & Gray, P. (2002). Guidelines for the design of haptic widgets. In *Proceedings of BCS HCI 2002* (pp. 195-212). London: Springer.
- Oakley, I., McGee, M. R., Brewster, S. A. & Gray, P. D. (2000). Putting the feel in 'Look and Feel'. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 415-422). New York: ACM Press.
- Oviatt, S. L. (1994). Interface techniques for minimizing disfluent input to spoken language systems. In *Proceedings of the SIGCHI conference on human factors in computing systems: Celebrating interdependence (CHI'94)* (pp. 205-210). New York: ACM Press.
- Oviatt, S. L. (1997). Multimodal interactive maps: Designing for human performance. *Human-Computer Interaction*, 12, 93-129.
- Oviatt, S. L. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42(11), 74-81.
- Oviatt, S. L. (2002). Breaking the robustness barrier: Recent progress on the design of robust multimodal systems. *Advances in Computers*, 56, 305-341.
- Oviatt, S. L. (2003). Multimodal interfaces. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 286-304). Mahwah, NJ: Lawrence Erlbaum Associates.
- Oviatt, S. L., Cohen, P. R., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J. & Ferro, D. (2000). Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Human-Computer Interaction*, 15, 263-322.
- Oviatt, S. L., Coulston, R., Tomko, S., Benfang, X., Lunsford, R., Wesson, M. & Carmichael, L. (2003). Toward a theory of organized multimodal integration patterns

- during human-computer interaction. In *Proceedings of the 5th international conference on multimodal interfaces* (pp. 44-51). New York: ACM Press.
- Oviatt, S. L., DeAngeli, A. & Kuhn, K. (1997). Integration and synchronization of input modes during human-computer interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 415-422). New York: ACM Press.
- Oviatt, S. L. & Kuhn, K. (1998). Referential features and linguistic indirection in multimodal language. In *Proceedings of the international conference on spoken language processing* (pp. 2339-2342). Sydney: ASSTA, Inc.
- Oviatt, S. L., Levow, G.-A., Moreton, E. & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human-computer error resolution. *Journal of the Acoustical Society of America*, 104(5), 3080-3098.
- Oviatt, S. L., Lunsford, R. & Coulston, R. (2005). Individual differences in multimodal integration patterns: What are they and why do they exist? In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 241-249). New York: ACM Press.
- Oviatt, S. L., MacEachern, M. & Levow, G.-A. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication*, 24(2), 87-110.
- Page, M. P. A. & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 56(8), 1289-1300.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Pavlovic, V. I., Sharma, R. & Huang, T. (1997). Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 677-695.
- Perlman, G. & Horan, F. L. (1986). Report on UNIX|STAT release 5.1: Data analysis programs for UNIX and MSDOS. *Behavior Research Methods, Instruments, & Computers*, 18(2), 168-176.
- Pieraccini, R., Dayanidhi, K., Bloom, J., Dahan, J. G., Phillips, M., Goodman, B. R. & Prasad, K. V. (2004). Multimodal conversational systems for automobiles. *Communications of the ACM*, 47(1), 47-49.
- Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61(5), 1352-1361.
- Pisoni, D. B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, 1, 10-23.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, 24(5), 574-590.
- Raisamo, R. (1999). *Multimodal human-computer interaction: A constructive and empirical study*. Unpublished doctoral dissertation, University of Tampere, Finland.

- Rasmussen, J. (1982). Human errors: A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, 4, 311-333.
- Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13(3), 257-267.
- Rasmussen, J. (1986). *Information processing and human-machine interaction. An approach to cognitive engineering*. New York: North Holland.
- Rauschenberger, R. & Chu, H. (2006). The effects of stimulus rotation and familiarity in visual search. *Perception & Psychophysics*, 68(5), 770-775.
- Rauschenberger, R. & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General*, 135(1), 116-131.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Reason, J. (1994). *Menschliches Versagen. Psychologische Risikofaktoren und moderne Technologien*. Heidelberg: Spektrum Akademischer Verlag.
- Reeves, L. M., Lai, J., Larson, J. A., Oviatt, S. L., Balaji, T. S., Buisine, S., Collings, P., Cohen, P. R., Kraal, B., Martin, J. C., McTear, M., Raman, T. V., Stanney, K. M., Su, H. & Wang, Q. Y. (2004). Guidelines for multimodal user interface design. *Communications of the ACM*, 41(1), 57-59. Retrieved December 10, 2006, from <http://www.niceproject.com/publications/CACM2004.pdf>.
- Reichart, G. & Haller, R. (1995). Mehr aktive Sicherheit durch neue Systeme für Fahrzeug und Straßenverkehr. In W. Fastenmeier (Ed.), *Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme* (pp. 199-215). Köln: TÜV Rheinland.
- Rhyne, J. R. & Wolf, C. G. (1993). Recognition-based user interfaces. In H. R. Hartson & D. Dix (Eds.), *Advances in Human-Computer Interaction* (Vol. 4, pp. 191-250). Norwood, NJ: Ablex Publishing.
- Richards, M. & Underwood, K. (1984). Talking to machines: How are people naturally inclined to speak? In E. D. Megaw (Ed.), *Contemporary Ergonomics 1984* (pp. 62-67). London: Taylor & Francis.
- Rockwell, T. H. (1988). Spare visual capacity in driving - revisited: New empirical results for old ideas. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith & S. P. Taylor (Eds.), *Vision in vehicles II* (pp. 317-324). Amsterdam: Elsevier.
- Rogers, Y. (1989). Icons at the interface: Their usefulness. *Interacting with Computers*, 1(1), 105-117.
- Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept. *Zeitschrift für Arbeitswissenschaft*, 38(4), 193-200.
- Rohmert, W. (1986). Ergonomics: Concept of work, stress and strain. *International Review of Applied Psychology*, 35(2), 159-181.
- Roßmeier, M., Grabsch, H.-P. & Rimini-Döring, M. (2005). Blind flight: Do auditory lane departure warnings attract attention or actually guide action? In *Proceedings of*

- the 11th International Conference on Auditory Display ICAD 2005* (pp. 43-48). Limerick, Ireland: ICAD.
- Rötting, M. (1999). Methoden zur Registrierung von Augenbewegungen. In M. Rötting & K. Seifert (Eds.), *ZMMS-Spektrum: Vol. 8. Blickbewegungen in der Mensch-Maschine-Systemtechnik* (pp. 19-34). Sinzheim: Pro Universitate Verlag.
- Rühmann, H. & Bubb, H. (2006). Fahrerassistenzsysteme. Ein Gewinn an Sicherheit und Komfort oder elektronischer Schnickschnack? *Ergonomie aktuell*, 7, 2-13.
- Rumar, K. (1990). The basic driver error: Late detection. *Ergonomics*, 33(10), 1281-1290.
- Salamé, P. & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 150-164.
- Saldaña, H. M. & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, 54(3), 406-416.
- Salvucci, D. D. & Anderson, J. R. (2000). Intelligent gaze-added interfaces. In *CHI '00: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 273-280). New York: ACM Press.
- Sato, K., Igarashi, E. & Kimura, M. (1991). Development of non-constrained master arm with tactile feedback device. In P. Dario (Ed.), *Proceedings of the 5th International Conference on Advanced Robotics (ICAR)* (Vol. 1, pp. 334-338). Pisa: IEEE.
- Schattenberg, K. & Debus, G. (2001). Multimodale Anzeige- und Bedienkonzepte zur Steuerung technischer Systeme während der Fahrt im Kraftfahrzeug: Evaluationsbefunde zur Systemweiterentwicklung mit paralleler Sprachbedienung. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 177-193). Berlin: Springer.
- Schmidt, R. F. & Thews, G. (1997). *Physiologie des Menschen* (27th ed.). Berlin: Springer.
- Schneider, W. & Fisk, A. D. (1982). Concurrent automatic and controlled visual search: Can processing occur without resource cost? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(4), 261-278.
- Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoît, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defée, I., Münch, S., Hartung, K. & Blauert, J. (1995). *A taxonomy of multimodal interaction in the human information processing system. Multimodal integration for advanced multimedia interfaces* (Report of the Esprit Project 8579 MIAMI No. WP 1). Nijmegen, Netherlands: University of Nijmegen.
- Schorn, H. J. (2004). Sprachsteuerung in industrieller Umgebung. *A&D Newsletter*, 6(2), 40-43.
- Schwarz, W. (1989). A new model to explain the redundant-signals effect. *Perception & Psychophysics*, 46(5), 498-500.
- Seagull, F. J., Wickens, C. D. & Loeb, R. G. (2001). When is less more? Attention and workload in auditory, visual and redundant patient-monitoring conditions. In *Proceedings of the Human Factors Society 45th Annual Meeting* (pp. 1395-1399). Santa Monica, CA: Human Factors and Ergonomics Society.

- Seifert, K. (2002). *Evaluation multimodaler Computer-Systeme in frühen Entwicklungsphasen*. Unpublished doctoral dissertation, Technische Universität Berlin, Berlin.
- Seifert, K., Rötting, M. & Jung, R. (2001). Registrierung von Blickbewegungen im Kraftfahrzeug. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 207-228). Berlin: Springer.
- Sellen, A. J., Kurtenbach, G. P. & Buxton, W. A. (1992). The prevention of mode errors through sensory feedback. *Human-Computer Interaction*, 7(2), 141-164.
- Servan-Schreiber, E. & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 592-608.
- Shadoin, A. L. & Ellis, N. R. (1992). Automatic processing of memory for spatial location. *Bulletin of the Psychonomic Society*, 30(1), 55-57.
- Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46, 561-584.
- Shah, P. & Miyake, A. (1999). Models of working memory. An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control*. (pp. 1-27). Cambridge, UK: Cambridge University Press.
- Sheridan, T. B. & Ferrell, W. R. (1974). *Man-machine systems. Information, control, and decision models of human performance*. Cambridge, MA: MIT Press.
- Shin, J., Narayanan, S. S., Gerber, L., Kazemzadeh, A. & Byrd, D. (2002). Analysis of user behavior under error conditions in spoken dialogs. In *Proceedings of the 7th International Conference on Spoken Language Processing ICSLP 2002* (pp. 2069-2072). Denver, CO: ICSA.
- Shneiderman, B. (1987). *Designing the user interface: Strategies for effective human-computer interaction*. Reading, MA: Addison-Wesley.
- Shneiderman, B. (1997). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Reading, MA: Addison-Wesley.
- Siewiorek, D. & Smailagic, A. (2003). User-centered interdisciplinary design of wearable computers. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 635-655). Mahwah, NJ: Lawrence Erlbaum Associates.
- Simpson, C. A., McCauley, M. E., Roland, E. F., Ruth, J. C. & Williges, B. H. (1987). Speech controls and displays. In G. Salvendy (Ed.), *Handbook of human factors*. (pp. 1490-1525). Oxford, UK: John Wiley & Sons.
- Simpson, C. A. & Williams, D. H. (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *Human Factors*, 22(3), 319-330.
- Sklar, A. E. & Sarter, N. B. (1999). Good vibrations: Tactile feedback in support of attention allocation and human-automation coordination in event-driven domains. *Human Factors*, 41(4), 543-552.

- Smith, D. C., Irby, C. H., Kimball, R. B., Verplank, W. L. & Harslem, E. F. (1982). Designing the Star user interface. *Byte*, 7(4), 242-282.
- Sorkin, R. D. (1987). Design of auditory and tactile displays. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 549-576). Oxford, UK: John Wiley & Sons.
- Spence, C. & Driver, J. (1997). Cross-modal links in attention between audition, vision, and touch: Implications for interface design. *International Journal of Cognitive Ergonomics*, 1(4), 351-373.
- Spence, C. & Driver, J. (1999). Multiple resources and multimodal interface design. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Transportation systems, medical ergonomics and training* (Vol. 3, pp. 305-312). Aldershot: Ashgate.
- Spies, M. (1994). Grundzüge der Spracherkennung in einem Diktiersystem. *Spektrum der Wissenschaft*, 1994/3, 90-95.
- Standing, L., Conezio, J. & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73-74.
- Stanton, N. A. (2003). Human error identification in human-computer interaction. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 371-383). Mahwah, NJ: Lawrence Erlbaum Associates.
- Staufer, M. J. (1987). *Piktogramme für Computer. Kognitive Verarbeitung, Methoden zur Produktion und Evaluation*. Berlin: Walter de Gruyter.
- Stefanucci, J. K. & Proffitt, D. R. (2005). Multimodal interfaces improve memory [CD-ROM]. In *Proceedings of Human-Computer Interaction International 2005*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Stein, B. E. & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- Sternberg, P. (1969). Memory-scanning: Mental process revealed by reaction-time experiments. *American Scientist*, 57(4), 421-457.
- Stevens, R. D., Edwards, A. D. N. & Harling, P. A. (1997). Access to mathematics for visually disabled students through multimodal interaction. *Human-Computer Interaction*, 12, 47-92.
- Stifelman, L. J. (1993). *User repairs of speech recognition errors: An intonational analysis* (MIT Media Laboratory Technical Report). Cambridge, MA: Massachusetts Institute of Technology.
- Streitz, N. A. (1990). Psychologische Aspekte der Mensch-Computer-Interaktion. In C. Graf Hoyos (Ed.), *Enzyklopädie der Psychologie. Ingenieurpsychologie* (Vol. D/III/2, pp. 240-284). Göttingen: Verlag für Psychologie, Hogrefe.
- Taylor, M. M., Lederman, S. J. & Gibson, R. H. (1973). Tactual perception of texture. In E. Carterette & M. Friedman (Eds.), *Handbook of perception* (Vol. 3, pp. 251-272). New York: Academic Press.

- Tietz, W. (2001). Haptische Auslegung der Fahrzeuginnenausstattung bei AUDI. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 177-186). Basel: Birkhäuser Verlag.
- Todd, J. W. (1912). *Reactions to multiple stimuli* (Archives of psychology, No. 25). New York: Science Press.
- Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.
- Treisman, A. (1964). Selective attention in man. *British Medical Bulletin*, 20, 12-16.
- Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, 76(3), 282-299.
- Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136.
- Turk, M. & Robertson, G. (2000). Perceptual user interfaces. *Communications of the ACM*, 43(3), 33-34.
- Tzelgov, J., Srebro, R., Henik, A. & Kushlevsky, A. (1987). Radiation search and detection by ear and by eye. *Human Factors*, 29(1), 87-95.
- Van Laer, J., Galanter, E. H. & Klein, S. J. (1960). Factors relevant to the development of aircraft warning and caution signal systems. *Aerospace Medicine*, 31, 31-39.
- Verwey, W. B. (1996). *Evaluating safety effects of in-vehicle information systems. A detailed research proposal* (Technical Report TM-96-C045). Soesterberg, Netherlands: TNO Human Factory Research Institute.
- Verwey, W. B. & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, 2(3), 270-285.
- Vidulich, M. A. & Tsang, P. S. (1986). Techniques of subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods. *Ergonomics*, 29(11), 1385-1398.
- Vilimek, R. & Hempel, T. (2005). Eine nutzerzentrierte Analyse von Erfolgsfaktoren zur Sprachbedienung im Automobil. *Forum Ware*, 32(1-4), 47-51.
- Vitense, H. S., Jacko, J. A. & Emery, V. K. (2003). Multimodal feedback: An assessment of performance and mental workload. *Ergonomics*, 46, 68-87.
- Vogels, I. M. L. C. (2004). Detection of temporal delays in visual-haptic interfaces. *Human Factors*, 46(1), 118-134.
- Wachtel, P. L. (1967). Conceptions of broad and narrow attention. *Psychological Bulletin*, 68(6), 417-429.
- Wahlster, W. (1991). User and discourse models for multimodal communication. In J. W. Sullivan & S. W. Tyler (Eds.), *Intelligent user interfaces* (pp. 45-67). New York: ACM Press.
- Ware, C. & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. *ACM SIGCHI Bulletin*, 18(4), 183-188.
- Weidenmann, B. (1988). *Psychische Prozesse beim Verstehen von Bildern*. Bern: Verlag Hans Huber.

- Weinschenk, S. & Barker, D. T. (2000). *Designing effective speech interfaces*. New York: John Wiley & Sons.
- Whiteside, J., Jones, S., Levy, P. S. & Wixon, D. (1985). User performance with command, menu, and iconic interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 185-191). San Francisco, CA: ACM Press.
- Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63-102). London: Academic Press.
- Wickens, C. D. (1991). Processing resources and attention. In D. Damos (Ed.), *Multiple-task performance* (pp. 3-34). London: Taylor & Francis.
- Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins Publishers.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.
- Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.
- Wickens, C. D. & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors*, 30(5), 599-616.
- Wickens, C. D., Sandry, D. & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25(2), 227-248.
- Wiedenbeck, S. (1999). The use of icons and labels in an end user application program: An empirical study of learning and retention. *Behaviour & Information Technology*, 18(2), 68-82.
- Wiegmann, D. A. & Shappell, S. A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *International Journal of Aviation Psychology*, 7(1), 67-81.
- Wierwille, W. W., Antin, J. F., Dingus, T. A. & Hulse, M. C. (1988). Visual attention demands of an in-car navigation display system. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith & S. P. Taylor (Eds.), *Vision in vehicles II* (pp. 307-316). Amsterdam: Elsevier.
- Wierwille, W. W. & Tijerina, L. (1995). Eine Analyse von Unfallberichten als ein Mittel zur Bestimmung von Problemen, die durch die Verteilung der visuellen Aufmerksamkeit und der visuellen Belastung innerhalb des Fahrzeugs verursacht wird. *Zeitschrift für Verkehrssicherheit*, 41, 164-168.
- Wierwille, W. W. & Tijerina, L. (1997). Darstellung des Zusammenhangs zwischen der visuellen Beanspruchung des Fahrers im Fahrzeug und dem Eintreten eines Unfalls. *Zeitschrift für Verkehrssicherheit*, 43, 67-74.
- Wilcox, R. R. (1987). New designs in analysis of variance. *Annual Review of Psychology*, 38, 29-60.
- Williams, J., Michelitsch, G., van Veen, J.-H. & Goronzy, S. (2004). MIAMM: Lessons learned in haptic & tactile interaction. In M. Buss & M. Fritschi (Eds.), *Pro-*

- ceedings of Eurohaptics 2004* (pp. 30-35). München: Technische Universität München.
- Williams, J. R. & Falzon, R. P. (1963). Relationship of display system variables to symbol recognition and search time. *Journal of Engineering Psychology*, 2(3), 97-111.
- Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1(2), 202-238.
- Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-73). London: University College London Press.
- Yankelovich, N. (1996). How do users know what to say? *ACM Interactions*, 3(6), 32-43.
- Yarbus, A. L. (1967). Eye movements during perception of complex objects. In L. A. Riggs (Ed.), *Eye movements and vision* (pp. 171-196). New York: Plenum Press.
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, 18, 459-482.
- Zhu, H. (1988). Electrotactile stimulation. In J. G. Webster (Ed.), *Tactile sensors for robotics and medicine* (pp. 341-354). New York: John Wiley & Sons.
- Zimmer, A. (1998). *Verhalten, Handlung und Kognition - Grundlagen und Anwendung: Anwendbarkeit taktiler Reize für die Informationsverarbeitung*. (Forschungsberichte vom Lehrstuhl Psychologie II). Regensburg: Universität Regensburg.
- Zimmer, A. (2001a). Gestaltpsychologische Ansätze zur Analyse der haptischen Wahrnehmung. In M. Grunwald & L. Beyer (Eds.), *Der bewegte Sinn. Grundlagen und Anwendungen zur haptischen Wahrnehmung* (pp. 77-88). Basel: Birkhäuser Verlag.
- Zimmer, A. (2001b). Wie intelligent darf/muss ein Auto sein? Anmerkungen aus ingenieurpsychologischer Sicht. In T. Jürgensohn & K. P. Timpe (Eds.), *Kraftfahrzeugführung* (pp. 39-55). Berlin: Springer.
- Zimmer, A. (2002). Berkeley's touch. Is only one sensory modality the basis of the perception of reality? In L. Albertazzi (Ed.), *Unfolding perceptual continua* (pp. 205-221). Amsterdam: John Benjamins.
- Zoltan-Ford, E. (1991). How to get people to say and type what computers can understand. *International Journal of Man-Machine Studies*, 34(4), 527-547.
- Zwahlen, H. T., Adams, C. C., jr. & DeBald, D. P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A. G. Gale, M. H. Freeman, C. M. Haslegrave, P. Smith & S. P. Taylor (Eds.), *Vision in vehicles II* (pp. 335-344). Amsterdam: Elsevier.
- Zwicker, E. & Fastl, H. (1999). *Psychoacoustics. Facts and models*. (2nd ed.). Berlin: Springer.