

# Über den Einfluss kontextabhängiger Qualitätsparameter auf die Wahrnehmung von Externalität und Hörereignisort

Dissertation zur Erlangung des  
akademischen Grades Doktor-Ingenieur (Dr.-Ing.)

vorgelegt der Fakultät Elektrotechnik und Informationstechnik  
der Technischen Universität Ilmenau  
von Dipl.-Ing. Stephan Werner

1. Gutachter: Prof. Dr.-Ing. Dr. rer. nat. h.c. mult. Karlheinz Brandenburg
2. Gutachter: Prof. Dr. rer. nat. Michael Vorländer
3. Gutachter: Prof. Dr.-Ing. Thomas Sporer

Tag der Einreichung: 10.07.2018

Tag der wissenschaftlichen Aussprache: 13.12.2018

urn:nbn:de:gbv:ilm1-2018000672



---

# Über den Einfluss kontextabhängiger Qualitätsparameter auf die Wahrnehmung von Externalität und Hörereignisort

---

eine Abhandlung von Stephan Werner

**sisko:** ... Jedes Mal, wenn ich diesen Ball werfe, können hunderte von verschiedenen Dingen in diesem Spiel passieren. Vielleicht trifft er den Ball, vielleicht trifft er daneben. Der springende Punkt ist, dass man es niemals weiß. Man versucht zu antizipieren, sich eine Strategie für alle Möglichkeiten, die eintreten können, zurechtzulegen. Aber im Endeffekt kommt es darauf an, einen Ball nach dem anderen zu werfen und zu warten was passiert. Egal was aus dieser Situation wird, das Spiel kommt wieder einen Schritt voran.

**wesen:** Und man hat keine Ahnung, wie das Spiel verläuft, bis es zu Ende gespielt ist?

**sisko:** Ja, stimmt genau! Darum geht es ja, das Spiel würde gar keinen Spaß machen, wenn wir genau wüssten, was passieren wird.

**wesen:** Es bedeutet Dir also etwas nicht zu wissen, was geschehen wird?

**sisko:** Das ist möglicherweise die wichtigste Erkenntnis, die uns Menschen betrifft. Es ist das Unbekannte, durch das unser Leben bestimmt wird. Wir sind dauernd auf der Suche, nicht nur nach Antworten auf unsere Fragen, sondern auch nach neuen Fragen. Wir sind Forscher! Wir erforschen unser Leben, Tag für Tag. Und wir erforschen die Galaxis und versuchen, die Grenzen unseres Wissens zu erweitern.

Dialog aus Star Trek: Deep Space Nine



## Kurzfassung

Das vorliegende Dokument umfasst meine Forschungstätigkeiten auf dem Gebiet des räumlichen Hörens. Es wird der Frage nachgegangen, welche Effekte verschiedene Kontexteinflüsse auf die Herausbildung einer auditorischen Illusion haben. Diese Effekte werden quantitativ erfasst.

Es werden die Wahrnehmung von Externalität, Hörereignisrichtung u. a. bei Verwendung eines Binauralsynthesystems untersucht. Als Ziel steht die Beschreibung und Messbarmachung der Kontexteinflüsse Raumdivergenz, Personalisierung und visuelle Merkmale. Die eigenen Untersuchungen zeigen deutlich, dass für das Zustandekommen einer sogenannten perfekten auditorischen Illusion eine ausreichende technische Realisierung der korrekten Ohrsignale nicht ausreichend ist. Vielmehr ist eine Erfüllung der intrinsischen Erwartungshaltung des Hörers notwendig.

Der Prozess der Herausbildung von Qualität lässt sich durch einen Vergleich und Beurteilungsprozess beschreiben. Hierbei werden die aus den durch die Sinnesorgane aufgenommenen Reizungen abgeleiteten Qualitätsmerkmale mit den abgeleiteten Merkmalen einer inneren, individuellen Referenz verglichen. Dieser Prozess ist ausschlaggebend für die Herausbildung eines Hörereignisses. Mithilfe von qualitätsbewertenden Befragungen (Hörtests) lässt sich dieses Hörereignis als Abbildung messbar und beschreibbar machen.

Die Ergebnisse meiner Arbeit zeigen, dass eine Divergenz akustischer Raumparameter zwischen einer binaural synthetisierten Szene und dem Abhörraum zu einer signifikanten Verringerung der Externalisierung von Hörereignissen führt. Eine Konvergenz raumakustischer Parameter erhöht die wahrgenommene Externalisierung. Die Sichtbarkeit von bspw. Lautsprecherattrappen im Abhörraum lässt die Externalität insgesamt ansteigen. Der sogenannte Raumdivergenzeffekt kann dadurch aber nicht aufgelöst werden. Die Personalisierung eines Binauralsynthesystems zeigt eine Verringerung von Fehllokalisationen und einen Anstieg der Externalisierung. Es wird die Vermutung aufgestellt und untersucht, inwiefern Lokalisierungsfehler und Externalisierung korrelieren. Es wird gezeigt werden, dass eine personalisierte Binauralsynthese den Raumdivergenzeffekt nicht auflösen kann.

Die Untersuchungen zum Raumdivergenzeffekt dienen als Grundlage zur Entwicklung von Methoden, welche verschiedene zeit- und energiebasierte akustische Merkmale der verwendeten Raumimpulsantworten verändern. Das Ziel ist die Anpassung einzelner raumakustischer Parameter der Synthese an die Raumakustik des Abhörraumes zur Erzeugung einer plausiblen auditorischen Illusion.

## **Abstract**

This document covers my research activities in the field of spatial hearing. The research question is examined, which different contextual influences have an effect on the build-up of an auditory illusion. These effects are recorded quantitatively.

The perception of externality, direction of auditory events, etc., when using a binaural synthesis system is investigated. The aim is to describe and measure the contextual influences of spatial divergence, personalisation and visual characteristics. The own investigations clearly show that a sufficient technical realization of the correct ear signals is not sufficient for the formation of a so-called perfect auditory illusion. Rather, the intrinsic expectations of the listener must be fulfilled.

The process of the development of quality can be described by a comparison and assessment process. The quality characteristics derived from the signals of the sensory organs are compared with the derived characteristics of an internal, individual reference. This process is decisive for the development of an auditory event. With the help of quality-assessing surveys (hearing tests), this auditory event can be made measurable and describable as an image.

The results of my work show that a divergence of acoustic room parameters between a binaurally synthesized scene and the listening room leads to a significant reduction of externalization of auditory events. A convergence of room acoustic parameters increases the perceived externalization. The visibility of loudspeaker dummies in the listening room, for example, increases the overall externality. However, the so-called room divergence effect cannot be resolved by this. The personalization of a binaural synthesis system shows a reduction of mislocalizations and an increase in externalization. The assumption is made and investigated to what extent localization errors and externalization correlate. It will be shown that a personalized binaural synthesis cannot resolve the spatial divergence effect.

The investigations of the room divergence effect serve as a basis for the development of methods that change various time- and energy-based acoustic characteristics of the room impulse responses used. The aim is to adapt individual room acoustic parameters of the synthesis to the room acoustics of the listening room to create a plausible auditory illusion.

## Danksagung

Für die inhaltliche und fachliche Unterstützung möchte ich im Besonderen allen Kolleginnen und Kollegen sowie Studierenden danken, welche sich für die in dieser Arbeit angesprochenen Themen und Forschungsbereiche interessiert haben und interessieren. Dies gilt umso mehr für Personen, welche sich aktiv in Form von studentischen Arbeiten, gemeinsamen Veröffentlichungen und/oder durch das Interesse an meiner Forschung beteiligt haben. Diese sind vor allem aber keineswegs ausschließlich: Martina Böhme, Karlheinz Brandenburg, Simone Füg, Georg Götz, Tamás Harczos, Frank Jürgens, Florian Klein, Judith Liebetrau, Thomas Mayenfels, Christina Mittag, Annika Neidhardt, Nicolas Pachatz, Martin Rekitt, Rebecca Sass, André Siegel, Thomas Sporer, Anett Zabel. Die Personen, welche sich ebenfalls durch Coautorenschaften in gemeinsamen Veröffentlichungen beteiligt haben, sind der Literaturliste zu entnehmen.

Im Weiteren soll allen Studierenden gedankt werden, die sich in den durchgeführten Lehrveranstaltungen in Form von Projektentwicklung und -durchführung beteiligt haben. Es sollen hier vor allem die Lehrveranstaltungen Advanced Psychoacoustics, Applied and Virtual Acoustics und die Praxiswerkstatt Hörtests des Studienganges Medientechnologie genannt werden. Ein großer Dank gilt den Testpersonen für ihre Teilnahme an Untersuchungen sowie ihrem Interesse an meiner Forschung.

Meine Forschung konnte nur durchgeführt werden, da ich von der Technischen Universität Ilmenau und dem Fachgebiet Elektronische Medientechnologie die Möglichkeit einer Promotion bekommen habe. Hierfür geht vor allem der Dank an Prof. Karlheinz Brandenburg für sein Vertrauen in mich und für die Freiräume, die er mir in der Auswahl und Bearbeitung des Themas gegeben hat.

Des Weiteren geht der Dank an die Fördermittelgeber und dahinterstehenden Finanzgeber (vorwiegend wohl die Steuerzahler in Deutschland und der EU) der Forschungsprojekte, in denen ich tätig war und bin. Diese sind im Besonderen die Deutsche Forschungsgemeinschaft, das Thüringer Ministerium für Wirtschaft, Wissenschaft und Digitale Gesellschaft, der Europäische Sozialfonds für Thüringen und die TU Ilmenau.

Ein großer Dank geht an meine Eltern sowie an Sara Kepplinger für ihre stetige Unterstützung und den Rückhalt.

## **Eigenständigkeitserklärung**

Die vorliegende Abhandlung wurde von mir selbstständig verfasst. Die für die inhaltliche Erstellung verwendeten Quellen sind den Referenzierungen zu entnehmen. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen oder auch eigenen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Die Arbeit basiert unter anderem auch auf gemeinsamen Arbeiten und Veröffentlichungen mit Kolleginnen und Kollegen sowie Studierenden der Technischen Universität Ilmenau. Die sich daraus ergebenden Arbeiten und eigene Veröffentlichungen sind dem Literaturverzeichnis und der Liste „Eigene Veröffentlichungen“ zu entnehmen. Die von mir thematisch gestellten und betreuten Abschlussarbeiten und deren Bearbeiterinnen und Bearbeiter sind der Liste „Betreute Abschlussarbeiten“ zu entnehmen. Alle weiteren Hilfestellungen und Hilfsmittel dritter Personen wurden ebenfalls kenntlich gemacht.

Die Abhandlung wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt [205].

Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß § 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat [205].

---

Ilmenau, 28. Januar 2019

## **Kurzübersicht**

<b>Kurzfassung</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Danksagung</b>	<b>iii</b>
<b>Eigenständigkeitserklärung</b>	<b>iv</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Physik des räumlichen Hörens</b>	<b>21</b>
<b>3 Hörereignisbildung</b>	<b>55</b>
<b>4 Perzeptive Messung kontextabhängiger Qualitätsparameter</b>	<b>73</b>
<b>5 Methoden einer adaptiven Binauralsynthese</b>	<b>133</b>
<b>6 Zusammenfassung und Ausblick</b>	<b>177</b>
<b>Literaturverzeichnis</b>	<b>192</b>



# Inhaltsverzeichnis

	<b>4</b>
<b>Kurzfassung</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Danksagung</b>	<b>iii</b>
<b>Eigenständigkeitserklärung</b>	<b>iv</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation und Zielstellung . . . . .	2
1.2 Herangehensweise . . . . .	4
1.3 Begrifflichkeiten . . . . .	7
1.4 Denkarten und Sichtweisen . . . . .	11
1.5 Struktur der Arbeit . . . . .	20
<b>2 Physik des räumlichen Hörens</b>	<b>21</b>
2.1 Physiologie . . . . .	22
2.1.1 Das Ohr als sensorisches System . . . . .	22
2.1.2 Auditorischer Pfad . . . . .	28
2.2 Akustische Signalmerkmale . . . . .	31
2.2.1 Monaurale spektrale Merkmale . . . . .	31
2.2.2 Binaurale Merkmale . . . . .	32
2.2.3 Akustische Distanzmerkmale . . . . .	36
2.3 Signalgenerierung . . . . .	40
2.3.1 Mikrofonierung . . . . .	40
2.3.2 Abbildung der Übertragungscharakteristika . . . . .	43
2.4 Binaurale Wiedergabe . . . . .	46
2.4.1 Binaurale Lautsprecherwiedergabe . . . . .	46
2.4.2 Binaurale Kopfhörerwiedergabe . . . . .	47
2.5 Aufnahmeraum und Abhörraum . . . . .	52

<b>3</b>	<b>Hörereignisbildung</b>	<b>55</b>
3.1	Merkmale räumlicher Audioqualität . . . . .	56
3.1.1	Hörereignisrichtung . . . . .	57
3.1.2	Hörereignisdistanz . . . . .	58
3.1.3	Externalität . . . . .	59
3.2	Einflüsse des Verwendungskontextes . . . . .	64
3.2.1	Präzedenzeffekt . . . . .	64
3.2.2	Raumdivergenzeffekt . . . . .	66
3.2.3	Auditorisch-visuelle Interaktion . . . . .	68
<b>4</b>	<b>Perzeptive Messung kontextabhängiger Qualitätsparameter</b>	<b>73</b>
4.1	Ventriloquismus-Effekt . . . . .	74
4.1.1	Exp. I: psychometrische Funktionen . . . . .	74
4.1.1.1	Untersuchungsgegenstand . . . . .	74
4.1.1.2	Versuchsaufbau . . . . .	74
4.1.1.3	Versuchsdurchführung . . . . .	76
4.1.1.4	Ergebnisse . . . . .	77
4.1.2	Exp. II: Zeigermethode . . . . .	80
4.1.2.1	Untersuchungsgegenstand . . . . .	80
4.1.2.2	Versuchsaufbau . . . . .	81
4.1.2.3	Versuchsdurchführung . . . . .	82
4.1.2.4	Ergebnisse . . . . .	82
4.1.3	Zusammenfassung . . . . .	84
4.2	Externalisierung und Hörereignisrichtung . . . . .	86
4.2.1	Exp. III: Raumdivergenzeffekt . . . . .	86
4.2.1.1	Untersuchungsgegenstand . . . . .	87
4.2.1.2	Versuchsaufbau . . . . .	87
4.2.1.3	Versuchsdurchführung . . . . .	89
4.2.1.4	Ergebnisse . . . . .	92
4.2.2	Zusammenfassung . . . . .	105
4.3	Räumliche Szenenkomplexität und raumakustische Disparität . . . . .	110
4.3.1	Exp. IV: Audioszene . . . . .	110
4.3.1.1	Untersuchungsgegenstand . . . . .	110
4.3.1.2	Versuchsaufbau . . . . .	111
4.3.1.3	Versuchsdurchführung . . . . .	113
4.3.1.4	Ergebnisse . . . . .	115
4.3.2	Zusammenfassung . . . . .	118



4.4	Einfluss von Kopfbewegungen auf die Externalisierung . . . . .	121
4.4.1	Exp. V: Head-Tracking . . . . .	121
4.4.1.1	Untersuchungsgegenstand . . . . .	122
4.4.1.2	Versuchsaufbau . . . . .	122
4.4.1.3	Versuchsdurchführung . . . . .	124
4.4.1.4	Ergebnisse . . . . .	126
4.4.2	Zusammenfassung . . . . .	128
4.5	Zusammenfassung kontextabhängige Qualitätsparameter . . . . .	130
<b>5</b>	<b>Methoden einer adaptiven Binauralsynthese</b>	<b>133</b>
5.1	Akustische Adaptationsverfahren . . . . .	134
5.2	Anpassung von BRIRs . . . . .	135
5.2.1	Exp. VI: Perzeptive DRR-Anpassung . . . . .	135
5.2.1.1	Untersuchungsgegenstand . . . . .	135
5.2.1.2	Versuchsaufbau . . . . .	135
5.2.1.3	Versuchsdurchführung . . . . .	137
5.2.1.4	Ergebnisse . . . . .	137
5.2.2	Exp. VII: DRR-Einfluss auf den Raumdivergenzeffekt . . . . .	139
5.2.2.1	Untersuchungsgegenstand . . . . .	139
5.2.2.2	Versuchsaufbau . . . . .	140
5.2.2.3	Versuchsdurchführung . . . . .	141
5.2.2.4	Ergebnisse . . . . .	142
5.2.3	Zusammenfassung . . . . .	149
5.3	Synthese von BRIRs . . . . .	151
5.3.1	Anpassung der Initialzeitlücke . . . . .	151
5.3.2	Interpolationsmethoden . . . . .	152
5.3.3	Exp. VIII: ITDG-basierte Distanzanpassung . . . . .	155
5.3.3.1	Untersuchungsgegenstand . . . . .	155
5.3.3.2	Versuchsaufbau . . . . .	156
5.3.3.3	Versuchsdurchführung . . . . .	156
5.3.3.4	Ergebnisse . . . . .	157
5.3.4	Exp. IX: Interpolationsbasierte Distanzanpassung . . . . .	158
5.3.4.1	Untersuchungsgegenstand . . . . .	159
5.3.4.2	Versuchsaufbau . . . . .	159
5.3.4.3	Versuchsdurchführung . . . . .	160
5.3.4.4	Ergebnisse . . . . .	161
5.3.5	Exp. X: Synthese aus räumlich dünn-besetzten Messungen . . . . .	163
5.3.5.1	Untersuchungsgegenstand . . . . .	164
5.3.5.2	Versuchsaufbau . . . . .	165

5.3.5.3	Versuchsdurchführung . . . . .	167
5.3.5.4	Ergebnisse . . . . .	169
5.3.6	Zusammenfassung . . . . .	174
5.4	Zusammenfassung adaptive Binauralsynthese . . . . .	175
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>177</b>
6.1	Rückschau . . . . .	178
6.2	Neue Fragen . . . . .	181
6.3	Externalisierung als Funktion einzelner kontextabhängiger Qualitätsparameter	183
6.3.1	Motivation . . . . .	183
6.3.2	Kontextparameter . . . . .	184
6.4	Ausblicke . . . . .	188
	<b>Literaturverzeichnis</b>	<b>192</b>

## Kapitel 1

### Einleitung

Zu jener Zeit schlug ihm Orfanik vor, mittels phonografischer Apparate die Hauptnummern ihres Repertoires [...] getreu aufzunehmen. Die betreffenden Apparate waren damals schon wunderbar vervollkommenet, dass die menschliche Stimme weder in Klangfarbe noch in ihrer Reinheit dadurch im Geringsten verändert wurde.

---

*Jules Verne*  
*Das Karpatenschloss, 1892 [218]*

## 1.1 Motivation und Zielstellung

Spätestens seit der Erfindung von technischen Geräten zur Aufnahme und Wiedergabe von Schall lässt sich der Wunsch der Erzeugung einer perfekten auditorischen Illusion für den Hörer formulieren [150, 202]. Das Audiosystem selber soll dabei den Anforderungen und Erwartungen des Nutzers hinsichtlich Immersion [108] und Plausibilität [131, 142] gerecht werden. Durch das technische System soll eine für bestimmte Anforderungen und Erwartungen geeignete auditorische Illusion erzeugt werden. Erfüllt ein solches System diese Erwartungen, dann zeichnet sich das System durch Passung bzw. Viabilität [99] bezogen auf den Anwendungsfall aus.

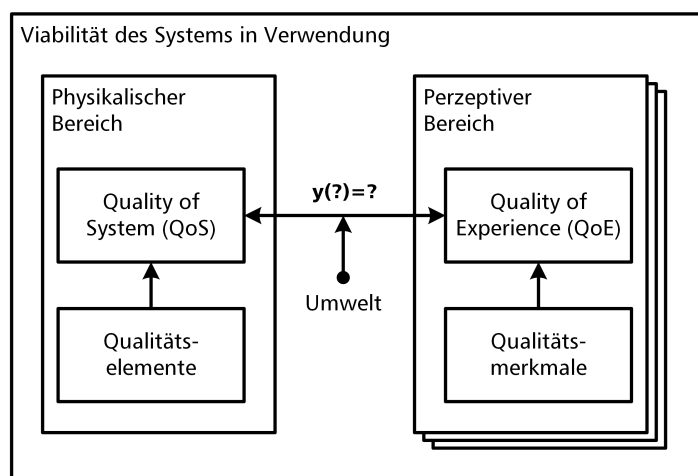


Abbildung 1.1: Systembetrachtung des Zustandekommens einer viablen technisch vermittelten Szene; Abbildung in Anlehnung an Silzle [196] und Jekosch [122].

Ein vielversprechender Ansatz zum grundlegenden Verstehen von Passung im multimedialen Kontext ist die verbundene Untersuchung der an der Entstehung des Wahrnehmungsereignisses beteiligten Systeme und deren Abhängigkeiten untereinander. Es wird davon ausgegangen, dass sich sowohl technische Geräte und Algorithmen als auch die Wahrnehmung des Menschen aus System Sicht betrachten und modellieren lassen.

Diese Sichtweise ist in Abbildung 1.1 verdeutlicht und beinhaltet im Wesentlichen:

- Eigenschaften des technischen Gesamtsystems (QoS) inkl. dessen Einzelkomponenten (physikalischer Bereich)
- erlebte Qualität des perzeptiven Systems (QoE) mit ihren qualitätsbeschreibenden Merkmale der einzelnen Individuen (perzeptiver Bereich)
- Interaktionen zwischen technischem System und perzeptivem System
- Interaktion des technischen und perzeptiven Systems mit seiner Umwelt.

Es ist bei dieser Betrachtung davon auszugehen, dass die beschreibbaren perzeptiven Qualitätsgrößen und Zusammenhänge stets individuell und zeitvariabel sind. Diese zusätzliche

Variabilität scheint vor allem auf die sich ändernde intrinsische Erwartungshaltung des einzelnen Nutzers zurückführbar zu sein. Die individuelle Wahrnehmung und die interne Referenz des Nutzers wird durch kognitive Prozesse der Antizipation und Adaptation stets angepasst.

Das in dieser Arbeit angestrebte Forschungsziel ergibt sich hinsichtlich der systematischen Untersuchung zum Zustandekommen einer auditorischen Illusion. Hierfür werden einzelne Qualitätsmerkmale im Verwendungskontext eines Binauralsynthesystems über Kopfhörer betrachtet. Durch die Forschung soll ein Beitrag zum Verstehen der Herausbildung einer Gesamtqualität in der räumlichen auditorischen Wahrnehmung geschaffen werden. Relevante Systemkomponenten und Kontextparameter des technischen Systems sollen identifiziert werden. Es ist hierbei die Überzeugung, dass die Betrachtung der Zusammenhänge einzelner Systemkomponenten und einzelner perzeptiver Qualitätsmerkmale notwendig ist, um qualifizierte Aussagen über das Zustandekommen einer Gesamtqualität und der Passung eines multimedialen Systems machen zu können.

**Forschungsfragen** - Welchen Einfluss hat der Verwendungskontext eines Binauralsynthesystems auf die Herausbildung einzelner Qualitätsmerkmale? Ergänzende Fragen sind: Wie stark ist dieser Einfluss quantitativ? Wirken die untersuchten kontextabhängigen Qualitätsparameter unabhängig voneinander? Welche akustischen Anpassungen in der Binauralsynthese führen zu einer quantitativen und qualitativen Steigerung der Qualitätsmerkmale?

## 1.2 Herangehensweise

In der Vergangenheit wurden zahlreiche relevante Merkmale von Qualität in Bezug auf das räumliche Hören und die Hörbarkeit von Räumen definiert (bspw. [58]). Diese umfassen die Lokalisierung des Hörereignisses, die Distanzwahrnehmung, die räumliche Umhüllung, die Externalisierung der Hörereignisse, die Wahrnehmung des Raumes, die Abbildbarkeit von Klangfarbe bzw. deren Unterscheidbarkeit und die Wahrnehmung von sonstigen Artefakten (bspw. [174]). Es wird versucht, diese Vielzahl an Qualitätsmerkmalen zu klassifizieren, hierarchisch abzubilden und aufzulisten. Beispielfhaft sind hierzu Arbeiten von Le Bagousse et al. [53], Lindau et al. [140] und Zacharov et al. [235] zu nennen.

Große Fortschritte wurden in der technischen Entwicklung von Audiosystemen zur räumlichen Wiedergabe, wie z.B. Ambisonics, Wellenfeldsynthese und binauraler Kopfhörerwiedergabe gemacht [178, 199]. Die technische und perzeptive Qualität der Synthese wird kontinuierlich verbessert. Dennoch sind zahlreiche Effekte und begrenzende Faktoren unverstanden. Selbst technisch wohlverstandene und beherrschbare Systeme führen je nach ihrer Verwendung zu einer unterschiedlichen Herausbildung von Qualität beim Hörer. Ältere und neuere Erkenntnisse sowie eigene Erfahrungen machen immer wieder deutlich, dass die Darbietung von physikalisch korrekten Signalen alleine nicht ausreichend ist, um eine plausible Simulation oder Synthese einer auditorischen Szene zu erreichen.

Bei der Bildung eines Hörereignisses spielen vor allem kognitive Effekte eine bedeutende Rolle. Diese lassen sich als hypothesengetriebene Prozesse des Hörens zusammenfassen [65, 174]. Die durch die Sinnesorgane aufgenommenen Signale werden durch den Wahrnehmungsapparat analysiert und unterliegen einer von höheren Bereichen des Gehirns gesteuerten Interpretation und Abstraktion der Szene. Das kognitive System erschafft somit eine bewertete Abbildung seiner Umgebung.

Ein technisches System steht immer im Kontext der Verwendung des Systems, bezogen auf die Wiedergabebedingungen und den Zweck der Anwendung. Die in dieser Arbeit untersuchten kontextabhängigen Effekte und Nutzungsbedingungen für ein Audiosystem umfassen vor allem die Wiedergabe. Hierbei werden multimodale Rückkopplungen (bspw. McGurk-Effekt [149] und Ventriloquismus-Effekt [190, 27], u. a.), Vorwissen über z.B. die Raumakustik [77, 22, 31] und damit verbundene Anpassungseffekte [7, 8] betrachtet.

Diese Arbeit soll einen Beitrag zum tiefergehenden Verständnis von kontextabhängigen Qualitätsparametern und die dadurch beeinflussten Qualitätsmerkmale liefern. Die durch den Kontext veränderte Wahrnehmung soll in Hörversuchen quantitativ beschreibbar werden. Es wird der Einfluss auf einzelne als relevant eingestufte Qualitätsmerkmale untersucht. Eine sich anschließende Auswertung zeigt neben dem Einfluss auf die einzelnen Qualitätsmerkmale auch die Beziehungen dieser untereinander auf.

Es werden die Qualitätsmerkmale Wahrnehmung der Schallereignisse außerhalb des Kopfes (Externalisierung), Lokalisationsfähigkeit, Entfernungswahrnehmung, Räumlichkeitswahrnehmung

und ggf. andere untersucht. Abbildung 1.2 verdeutlicht die zu ermittelnden Zusammenhänge grafisch. Die angedeuteten Anpassungsleistungen des kognitiven Systems („Adaptation“ in Abbildung 1.2) werden in dieser Arbeit am Rande betrachtet. Es ist hierzu auf die Arbeiten meines geschätzten Kollegen Florian Klein verwiesen.

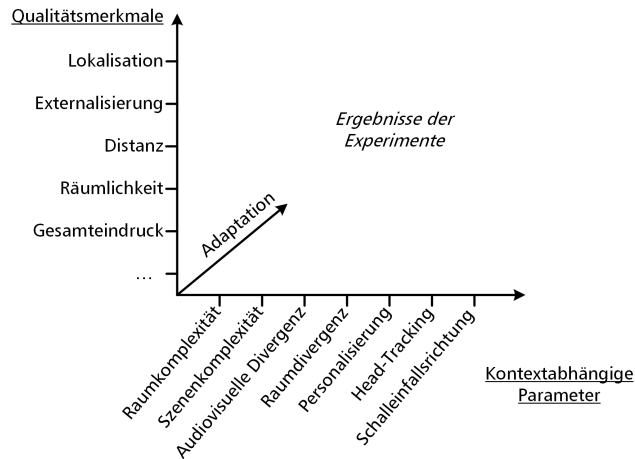


Abbildung 1.2: Zu ermittelnde Zusammenhänge zwischen Qualitätsmerkmalen und kontextabhängigen Qualitätsmerkmalen.

Es wird eine systematische Analyse von Wahrnehmungseffekten bei der Wiedergabe von binauralen Signalen über Kopfhörer angestrebt. Im Rahmen von Hörexperimenten werden die Qualitätsmerkmale gemessen. Die Plausibilität einer binauralen Auralisation in Abhängigkeit von den untersuchten kontextabhängigen Einflussfaktoren soll einschätzbar werden. Zu den Einflussfaktoren zählen vor allem die Variation von Aufnahme- und Wiedergaberaum (Raumdivergenz), die Personalisierung des Aufnahme- und Wiedergabeverfahrens und die visuelle Präsentation der Szenen. Die aus diesen Untersuchungen gewonnenen Erkenntnisse werden genutzt, um die binaurale Simulation einer akustischen Szene an den Kontext der Nutzung anzupassen. Es werden Methoden vorgestellt, die die in der Binauralsynthese verwendeten binauralen Raumimpulsantworten (BRIRs) nachträglich so anpassen, dass die Distanzwahrnehmung sowie die Wahrnehmung von Räumlichkeit gezielt gesteuert werden kann. Hiermit ist eine Anpassung der Simulation in Richtung der Abhörbedingungen (bspw. Abhörraum) möglich.

Die gewonnenen Erkenntnisse können einen Beitrag zur Entwicklung von neuen Methoden zur Evaluierung und Validierung von räumlichen Audiosystemen leisten. Hier sei auf den Standard MPEG-H verwiesen [1, 14, 201], welcher explizit zur Codierung, Speicherung, Übertragung und Wiedergabe von mehrkanaligen und objektbasierten Audiosignalen entwickelt wurde. Weiterhin scheint die Bedeutung einer immersiven und plausiblen räumlichen Audiowiedergabe auch hinsichtlich der audiovisuellen Integration in virtuelle (VR) und augmentierte (AR) Realitäten in Zukunft zuzunehmen. Die Entwicklung von Systemen, welche sich an

den Verwendungskontext anpassen, kann ein Lösungsansatz zur Schaffung einer perfekten auditorischen Illusion für alle Hörer und unter allen Abhörbedingungen sein.

Eine vor fast 35 Jahren getätigte Aussage von Floyd E. Toole [213] legt eine damals schon seit mehr als 10 Jahren und heute immer noch nicht zufriedenstellend gelöste Lücke im grundlegenden Verständnis zur Erzeugung einer plausiblen auditorischen Illusionen bei Verwendung von binauraler Audiotechnik dar. Diese soll mir als Motivation aber auch als Mahnung dienen.

„Over the years there have been several claims that the problems of in-head localization have been solved. Each time, the systems of recordings and reproducing the sounds have moved closer to simulating the performance of real ears. People familiar with the recordings seemed to be most satisfied by the reproductions and, although others could share the experience, some listeners, it seemed, remained unconvinced. Physical and acoustical differences in ears may explain some of the problems but it seems likely that there is also a higher-order psychoacoustical factor: the listeners' experience and expectations.“ Floyd E. Toole [213]



### 1.3 Begrifflichkeiten

Aufgrund der verwirrenden Verwendung und Vielzahl von Definitionen von Begriffen im Umfeld der Qualitätsbewertung und -wahrnehmung ist das Ziel die Schaffung einer klaren Abgrenzung der einzelnen Teilbegriffe. Weiterhin wird eine Trennung zwischen Begriffen zur Beschreibung eines physikalisch, technischen Systems und zur Beschreibung des durch Perzeption und Kognition geprägten Systems (Mensch) angestrebt. Die Trennung soll Unterschiede aber vor allem auch Gemeinsamkeiten erkennbar und formulierbar machen. Im Weiteren wird auf existierende Wahrnehmungsmodelle eingegangen, welche als Grundlage der eigenen Arbeit dienen.

**Auditorische Illusion** - Die Illusion ist eine „Täuschung durch die Wirkung eines Kunstwerkes, das die Darstellung als [plausibel oder gar als] Wirklichkeit erleben lässt. [Sie ist eine] falsche Deutung von tatsächlichen Sinnes[-empfindungen].“ (weitgehend nach [83]). Der Begriff der auditorischen Illusion soll andeuten, dass der Hörer durch ein Audiosystem ertüchtigt werden soll eine individuelle spezifische Hörwahrnehmung für sich zu erschaffen. Hierfür muss das Audiosystem nicht zwingend eine physikalisch korrekte Abbildung eines Schallfeldes, sondern vielmehr ein für die Anforderungen des Hörers geeignetes Schallfeld erzeugen. Technische Audiosysteme erzeugen praktisch immer eine zu einem (angenommenen) originalen Schallfeld abweichende physikalische Repräsentation. Sind diese Abweichungen entweder gering genug oder sind diese, wie in der Stereophonie, aufeinander abgestimmt, werden interferierende sensorische Informationen vom Hörer wissentlich und/oder unwissentlich erduldet, ohne dass die auditorische Illusion zusammenbricht [112]. Die Illusion ist somit nicht in erster Linie von der technischen Realisierung, sondern von der intrinsischen Interpretation und Organisation der Erfahrungswelt (Kognition) des Menschen geprägt.

**Plausibilität** - Die Begrifflichkeit Plausibilität umschreibt, dass die (binaurale) Wiedergabe oder Synthese einer akustischen Szene zu definierende Qualitätskriterien oder -merkmale erfüllen muss. Die Qualitätsmerkmale sind von der Anwendung, z. B. von Anforderungen an die Genauigkeit der Abbildung, abhängig (siehe dazu Pellegrini [165]). Das Ziel der Erzeugung einer plausiblen auditorischen Illusion erfordert eine vorherige Festlegung solcher Qualitätsmerkmale. Bezogen werden diese Kriterien auf die erwartete Wahrnehmung einer (realen) akustischen Umgebung und Szene. Ist eine reale Szene vorhanden, entsteht somit eine authentische Synthese der Szene (Authentizität). Eine authentische Wiedergabe liegt dann vor, wenn bei einem Hörer auf der Wiedergabeseite die gleiche Wahrnehmung hervorgerufen wird, wie wenn er sich auf der Aufnahmeseite befunden hätte [63]. Dem Hörer muss die Möglichkeit gegeben werden, die Synthese mit der entsprechenden realen Szene (quasi externe Referenz) zu vergleichen. Fehlt diese Referenz, da bspw. keine entsprechende reale Szene existiert oder ein Vergleich unmöglich ist, lässt sich die Synthese durch ihre Plausibilität, als Vergleich mit der Erwartungshaltung (quasi interne Referenz) des Hörers, bewerten.

**Plausibilität:** „Plausibilität ist das Ergebnis eines perzeptiven Prozesses, der ermittelt, in welchem Maße ein Wahrnehmungsobjekt mit einer aus individuellen Vorerfahrungen resultierenden, inneren Referenz übereinstimmt. Die Referenz des Plausibilitätsurteils ist das Ereignis aller denkbaren Wahrnehmungsereignisse innerhalb eines jeweiligen Systems von relationaler und klassifizierender Inferenz.“ Kuhn-Rahloff [131].

Wird das Gehörte als plausibel durch den Hörer bewertet, erfüllt das Synthesystem die gestellten Anforderungen und Erwartungen des Hörers hinsichtlich der Qualitätsmerkmale. Das Synthetisierte und damit das System lässt sich als passend oder viabel für die Verwendung beschreiben.

**Qualität** - Zur Definition der Begrifflichkeiten im Umfeld der Qualitätsbeurteilung wird sich im Besonderen auf die Arbeiten und Beschreibungen des „European Network on Quality of Experience in Multimedia Systems and Services“ (Qualinet [86]) gestützt.

**Event / Ereignis:** „An observable occurrence. An event is determined in space, time, and character.“ [86]

Daraus ergeben sich die Definitionen der Begriffe **Empfindung** (engl.: sensation): „refers to the response of sensory receptors and sense organs to environmental stimuli.“ [86], sowie des Begriffs **Wahrnehmung** (engl.: perception): „... is a process which involves the recognition and interpretation of stimuli which register our senses.“ [86].

Ein Ereignis beschreibt jegliches messbare und beobachtbare physikalisches und/oder perzeptives Geschehnis.

**Experience / Erfahrung:** „An experience is an individual’s stream of perception and interpretation of one or multiple events.“ [86]

Die Erfahrung ist damit die nicht messbare Konstruktion von Realität aus den aufgenommenen Sinnesempfindungen und Wahrnehmungen einer Vielzahl von Ereignissen.

**Quality / Qualität:** „Is the outcome of an individual’s comparison and judgment process. It includes perception, reflection about the perception, and the description of the outcome. [...] we consider quality terms of the evaluated excellence or goodness, if the degree of need fulfillment, and in terms of a „quality event“ [...]“ [86]

Die Qualität ist damit eine auf Selbstbeobachtung basierende reflektierte und beschreibbare Erfahrung des Individuums. Es bildet sich ein introspektivistisches Qualitätsereignis aus,

welches sich gemäß Kriterien (excellence, goodness etc.) einordnen lässt. Die Qualität ist der Wunsch der Erreichung einer für individuelle Anforderungen und Erfahrungen passenden (viablen) Konstruktion von Realität durch kognitive Vorgänge im Individuum.

**Quality of Experience (QoE):** „QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.“ [86] und nach ITU-T Rec.P.10: „QoE [describes] the overall acceptability of an application or service, as perceived subjectively by the end user“ mit Ergänzungen [86]: „Note 1: Includes the complete end-to-end system effects; Note 2: May be influenced by user expectations and context.“

Die QoE ist die durch einzelne Merkmale und in Selbstbeobachtung beschreibbare Qualität. Eine geeignete Encodierung, bspw. in einem Wahrnehmungsexperiment, führt zu einer messbaren Abbildung der introspektivistischen Qualitätsereignisse in die Außenwelt.

**Influence Factor on QoE/ Einflussfaktor auf QoE:** „Any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the QoE for the user.“ [86]

Neben Einflussfaktoren, die auf der Charakteristik einer Person (bspw. Demographie oder emotionaler Zustand) oder auf Eigenschaften des technischen Systems (bspw. Codierung und Übertragung) beruhen, lassen sich Einflussfaktoren definieren, die sich aus dem Kontext in dem sich der Nutzer befindet, ergeben. Ulrich Reiter et al. [86] unterteilt hierbei die kontextbedingten Einflüsse in einen physikalischen Kontext, zeitlichen Kontext, ökonomischen Kontext, Kontext der Aufgabenstellung, sowie technischen und Informationskontext.

**QoE Features / Qualitätsmerkmale:** „A perceivable, recognized, and namable characteristic of the individual’s experience of a service which contributes to its quality.“ [86]

Einzelne in Selbstbeobachtung erkennbare Merkmale zur Beschreibung des Qualitätsereignisses. Durch Encodierung lassen sich diese auf Skalen abbilden.

**Quality of Service (QoS):** „Totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service.“ [86, 118].

In dieser Arbeit wird der Begriff System als Synonym zu Service verwendet. Daraus folgend wird in der Arbeit „Quality of System“ verwendet, wenn ein multimediales System betrachtet wird. Dies umfasst die technischen Systemparameter und den Verwendungskontext des Systems sowie die Interaktion zwischen diesen. Die QoS fasst die Eigenschaften der einzelnen Systemkomponenten und die Kontextparameter in denen das System verwendet wird zusammen und beschreibt diese. Das Ziel der Qualität aus technischer Systemsicht ist der Wunsch der Erreichung einer multimedialen (Re-)Synthese mit einer hohen QoE für jeden Nutzer.

**Audio** - Im Bereich von Audiosystemen und der Wahrnehmung von Audiosignalen soll eine Trennung zwischen den Begriffen auf der physikalisch, technischen Systemseite und der perceptiven Nutzerseite erreicht werden.

**Schall:** jegliche sich ausbreitende hörbare Schwingung von Druck- und Dichteschwankungen in einem Medium.

**Hören/Akustik:** aus dem griechischem: akoyein; umfassender Begriff für sämtliche physikalische, biologische und perzeptive Vorgänge bei der Entstehung, Aufnahme und Verarbeitung von Schall.

**akustisches Ereignis:** erfasst die physikalischen Charakteristika von Schall und dem Schallereignis.

**auditiv/auditorisch:** den Gehörsinn und/oder das Gehörorgan betreffende; umfasst die Weiterleitung und Verarbeitung von Schallschwingungen ab den Hörnerven.

**auditorisches Ereignis:** erfasst die psychophysikalischen und perceptiven Charakteristika von Schall im Gehör.

## 1.4 Denkart und Sichtweisen

Im Verlauf der eigenen Forschungstätigkeit und des damit verbundenen Literaturstudiums sind Modelle und Theorien zur Abbildung der kognitiven Prozesse der Entstehung von Qualität, Wissen und Erkenntnis sowie die Verbindung zwischen technischen Systemkomponenten und der Wahrnehmung auffällig geworden. Es zeigt sich, dass die aus unterschiedlichen Forschungsbereichen stammenden Ansätze immer wieder Ähnlichkeiten hinsichtlich Struktur und Grundannahmen aufweisen. Hierzu gehören vor allem die Empfindung der Außenwelt durch die Sinnesorgane, die Ableitung von Perzepten, die Konstruktion einer wie auch immer gearteten internen Repräsentation der Außenwelt und ein Vergleich zwischen Wahrgenommenem und interner Erwartung. Die nachstehenden Abschnitte fassen als relevant eingeschätzte Ansätze und Modelle zum Verstehen der Vorgänge zur Herausbildung von auditiver Qualität zusammen. Abschließend werden diese in einem erweiterten Modell vereinheitlicht. Dieser Modellansatz dient als wesentliche Denkbasis für die nachfolgenden Inhalte der Arbeit.

Der Begriff Modell wird anhand der nachstehenden Definition eines allgemeinen Modellbegriffs von Stachowiak [200] verwendet. Ein Modell besitzt weiterhin verschiedene Merkmale<sup>1</sup>.

**Modell:** Ein Modell stellt eine Vereinfachung und Abbildung eines realen oder fiktiven Originals dar. Die Vereinfachung kann gegenständlicher oder theoretischer Natur sein [200].

**Dreitor-Modell** - Ein von Blauert vorgeschlagenes Modell einer schallwahrnehmenden und berichtenden Person verdeutlicht die durch Wahrnehmung entstehende Herausbildung eines Hörereignisses [64]. Abbildung 1.3 zeigt eine schematische Darstellung der Vorgänge der auditiven Schallaufnahme und Wahrnehmung.

Das akustische Ereignis (S<sub>0</sub>) stellt das physikalische Schallereignis dar. Dieses Ereignis kann durch geeignete physikalische Messgeräte erfasst und aufgenommen werden. Die Abbildung durch ein Messgerät stellt lediglich eine Beschreibung (S) des realen Ereignisses dar und ist durch die Übertragungseigenschaften des Messsystems geprägt. Das akustische Ereignis (S<sub>0</sub>) kann auch durch ein psychophysikalisches System erfasst und aufgenommen werden. Hierbei wird das physikalische Ereignis durch die Sinnesorgane in einen Sinneseindruck gewandelt. Ein sich anschließender wahrnehmender Teil bildet daraus ein auditorisches Ereignis. Dieses Ereignis ist eine durch die Systemeigenschaften der Sinnesorgane und wahrnehmenden

<sup>1</sup>**Abbildungsmerkmal:** Modelle sind Abbildungen, Repräsentationen natürlicher oder künstlicher Originale, die selbst wieder Modelle sein können.

**Verkürzungsmerkmal:** Modelle erfassen im allgemeinen nicht alle Attribute des durch sie repräsentierten Originals, sondern nur solche, die den jeweiligen Modellerschaffern und/oder Modellbenutzern relevant erscheinen.

**Pragmatisches Merkmal** Modelle sind ihren originalen nicht per se eindeutig zugeordnet. Sie erfüllen ihre Ersetzungsfunktion a) für bestimmte - erkennende und/oder handelnde, modellbenutzende - Subjekte, b) innerhalb bestimmter Zeitintervalle und c) unter Einschränkung auf bestimmte gedankliche oder tatsächliche Operationen. [200]

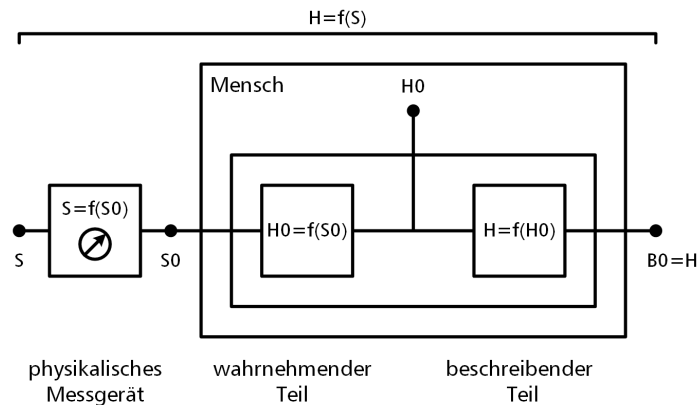


Abbildung 1.3: Dreitor-Modell einer schallwahrnehmenden und berichtenden Person; nachgezeichnet aus Blauert [64].

Verarbeitungsstufen des Gehörs konstruiertes Abbild ( $H_0$ ). Die Eigenschaften dieses Abbildes lassen sich nur durch Introspektion des Individuums erfassen. Ein Zugriff jenseits der eigenen Person erfolgt nur durch eine Beschreibung ( $H$ ) des auditorischen Ereignisses durch die Person selber. Dieser Vorgang beinhaltet ebenfalls eine Abbildung des Hörereignisses auf dessen Beschreibung. Eine sich daran anschließende Befragung der Person führt zu einer messbaren Erfassung des wahrgenommenen akustischen Ereignisses. Anschließend lässt sich ein Vergleich und eine Bewertung mit dem physikalisch gemessenem akustischen Ereignis ( $S$ ) herstellen.

Das Modell zeigt deutlich auf, dass das Hörereignis ein durch kognitive Prozesse entstehendes Konstrukt von Realität ist, welches sich nur durch Introspektion des Individuums erfassen lässt. Die in Hörexperimenten ermittelten Messungen sind immer eine Abbildung dieses Hörereignisses.

**Assoziationsmodell** - Das von Theile vorgeschlagene Modell beschreibt die Vorgänge der Lokalisation von Schallquellen im Raum als einen Wahrnehmungsprozess [206]. Hierbei wird von informationsreduzierenden Vorgängen der assoziativen Mustererkennung ausgegangen. Die Muster selber liegen als Konstrukte im kognitiven System vor und dienen einer Verknüpfung von aktuellen (wahrgenommenen) Mustern mit gespeicherten Mustern [206]. Theile führt hierzu ebenfalls eine Definition der Begriffe Empfindung und Wahrnehmung ein. Empfindungen sind demnach Sinnesereignisse, die sich invariant verhalten und nicht durch Assoziationen zustande kommen. Wahrnehmungen sind Sinnesereignisse, die durch bewusste und unbewusste Interpretationen beeinflusst werden und durch Assoziation zustande kommen [86].

Das vorgeschlagene Assoziationsmodell (siehe Abbildung 1.4) beinhaltet eine periphere Stufe der Aufnahme der Ohrsignale durch das Gehör als Empfindungsreizung. Eine sich anschließende hierarchische Struktur der Wahrnehmung aus einer ortsbestimmenden Musterselektion und einer übergeordneten gestaltsbestimmenden Musterselektion führt zu dem kognitiven Hörereignis. In den Stufen lösen bestimmte Ohrsignale mit darin enthaltenen bestimmten

Mustern eine Ortsassoziation und anschließend eine Gestaltsassoziation aus. Diese Zweiteilung soll die „zwei elementaren Bereiche der Hörerfahrung“, der „zwei voneinander unabhängigen, stets paarweise auftretenden Schallquelleneigenschaften Ort und Signal“ ([206]) modellieren. Theile schlägt weiterhin ein in Wechselwirkung stehendes unspezifisches System vor, welches eine Lenkung der Assoziation bewerkstelligt und in Beziehung zum Kurzzeit- und Langzeitgedächtnis steht. Dieser Ansatz der assoziativen Lenkung wird in nachfolgenden Abschnitten aufgegriffen und genauer betrachtet.

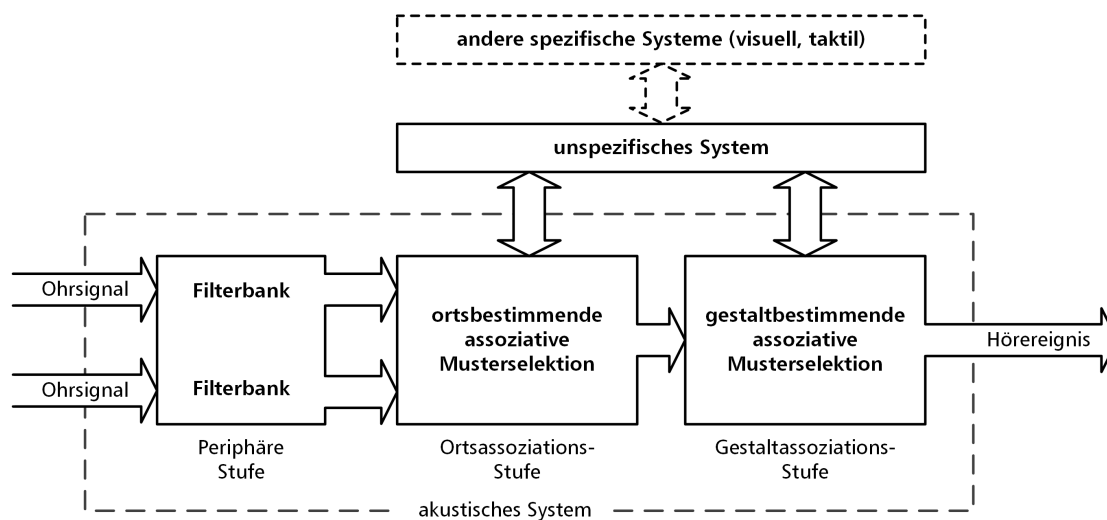


Abbildung 1.4: Assoziationsmodell zur Erklärung der Vorgänge der Lokalisation von Schallquellen im Raum; nachgezeichnet aus Theile [206].

**Taxonomiemodell** - In der Arbeit von Silzle [196] wird ein Klassifizierungsansatz der Qualität auditorischer virtueller Umgebungen vorgestellt. Eine schematische Darstellung der Beziehungen zwischen Elementen und Merkmalen ist in Abbildung 1.5 veranschaulicht. Der Kern ist hierbei eine Verknüpfung der Wahrnehmung durch ihre beschreibenden Qualitätsmerkmale mit dem technischen System und seine beschreibenden Qualitätselemente [121]. Das Modell von Silzle zeichnet sich durch seine Klarheit im Sinne des Versuchs einer direkten Zuordnung zwischen Systemparameter und Wahrnehmungsmerkmal aus. Es stellt gerade dadurch eine Basis für die eigene Arbeit dar.

Die Qualitätsmerkmale umfassen die Wahrnehmung der akustischen virtuellen Umgebung durch den Hörer und bilden zusammen mit einer internen Referenz die QoE. Diese Wahrnehmung wird durch Begriffe wie Lokalisationsgenauigkeit, Hörbarkeit von Artefakten, Dynamikgenauigkeit, Klangfarbe, Halligkeit und Räumlichkeit beschrieben. Qualitätselemente sind die technischen Parameter des Wiedergabesystems und bilden die QoS. Die Qualitätselemente selber wirken sich auf die Qualitätsmerkmale aus. Die bei Silzle verwendeten Qualitätselemente sind z. B. verwendete Außenohrübertragungsfunktionen (eng.: Head-Related Transfer Function, HRTF), Auflösung der HRTFs, Systemlatenz, Raumsimulationsparameter und andere. In dem

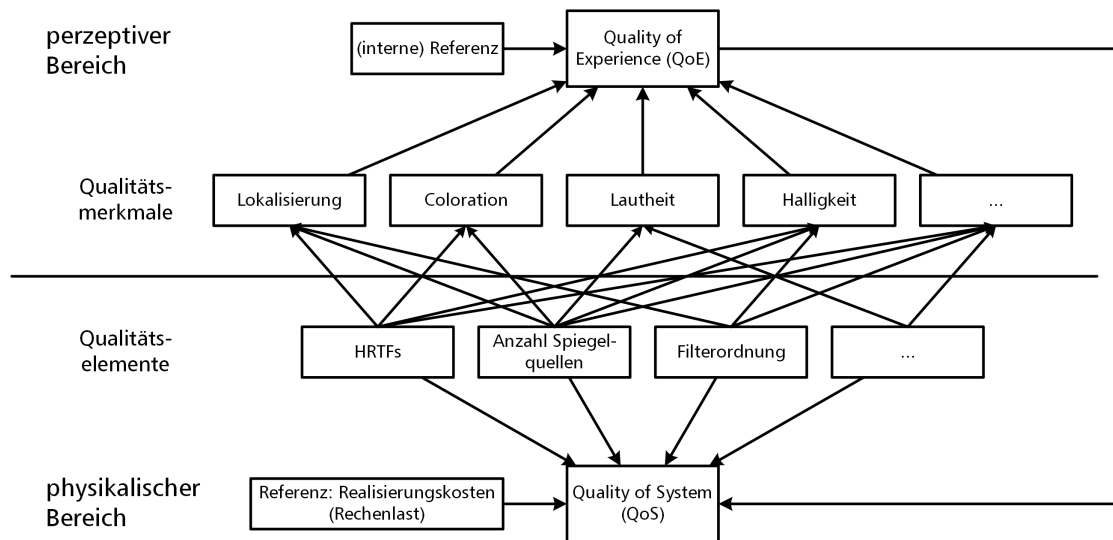


Abbildung 1.5: Schematische Darstellung der Beziehungen von Qualitätselementen und Qualitätsmerkmalen für eine virtuelle auditive Umgebung; nachgezeichnet aus Silzle [196]

vorgeschlagenen Ansatz besteht zusätzlich die Möglichkeit einer Rückkopplung und somit eine durch die QoE gesteuerte Anpassung der QoS. Im Rahmen von Experimenten und Befragungen wurden die Beziehungen zwischen Qualitätsmerkmalen und -elementen quantifiziert [196]. Diese Bestimmung der Beziehungen sind hierbei für verschiedene Anwendungsszenarios vorgenommen worden. Es zeigt sich, dass sich je nach Kontext der Anwendung unterschiedliche Verbindungen zwischen Elementen und Merkmalen herausbilden. Dieser Effekt liegt darin begründet, dass die Testperson die auralisierte virtuelle auditorische Szene mit ihrer persönlichen internen Referenz vergleichen, die vor allem auch die Erwartung der Testperson selber beinhaltet [123].

**Qualitätsbildungsprozess** - In dem von Jekosch [122] und Raake [173] vorgeschlagenen Qualitätsbildungsprozess sind an der Bildung einer Qualitätsbewertung zwei wesentlich Prozesse beteiligt (siehe dazu Abbildung 1.6). Der Wahrnehmungspfad wird durch das physikalisch beschreibbare Ereignis, welches die Sinnesorgane erreicht, getrieben. Unter Beachtung des Referenzpfades bildet sich daraus das wahrgenommene Ereignis ab. Dieses Sinnesereignis unterliegt einem Vergleich- und Beurteilungsprozess mit der internen Referenz der Person. Der Referenzpfad beschreibt hierbei die zeitlichen und kontextuellen Einflussfaktoren auf den Qualitätsbildungsprozess. Er beinhaltet ebenfalls ein Gedächtnis von vorangegangenen wahrgenommenen Qualitäten und beschreibt somit auch Prozesse von Erwartungen. Für die eigene Arbeit ist der beschriebene Qualitätsbildungsprozess vor allem hinsichtlich des Vergleichs von Wahrgenommenem und intern gespeichertem Erwartetem von Relevanz.

**Hörermodell** - Raake, Blauert et al. [63, 174] präsentieren ein Modell zur Einschätzung/Messung der QoE. Es wird eine hierarchische Struktur von Verarbeitungsstufen vor-



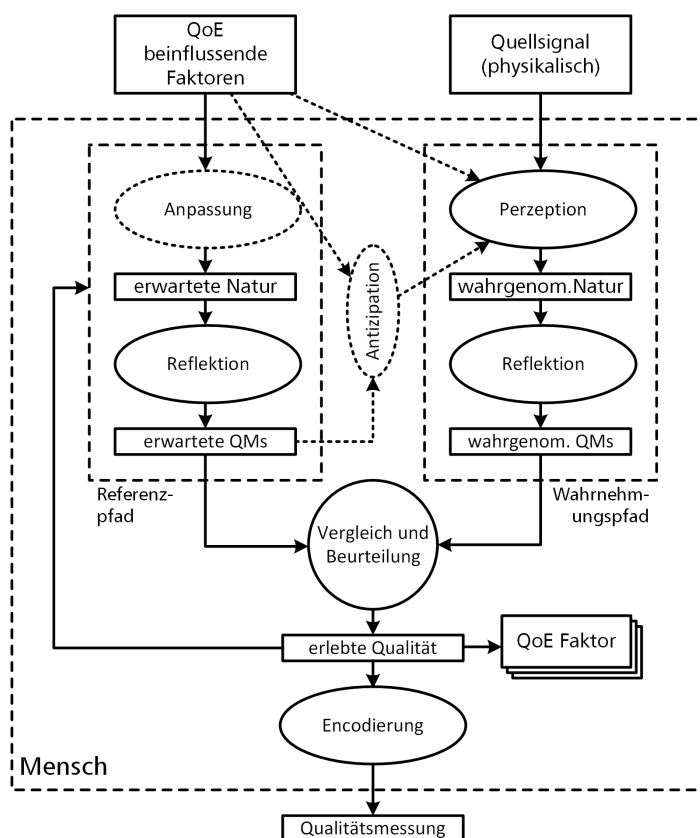


Abbildung 1.6: Qualitätsbildungsprozess; angelehnt an Jekosch [122] und Raake [173].

gestellt, die sich durch eine ansteigende Abstraktion der Merkmalsextraktion sowie einer Zuweisung von Bedeutung auszeichnet. Binaurale und monaurale Merkmale des Schallereignisses werden durch Prozesse der Kategorisierung und Segmentierung in eine symbolische Repräsentation überführt. Nachstehend wird diese Darstellung mithilfe einer Vielzahl von Hypothesen innerhalb der einzelnen Experten-Module auf Plausibilität überprüft. Die Hypothesen selber werden schrittweise evaluiert, angepasst und schließlich verworfen oder angenommen [63]. Die Hypothesen können bspw. Wissen über räumliches Hören, multimodale Integration, Psychoakustik, Kommunikation und Qualität beinhalten [174]. Das Expertensystem ist ein aufgabenspezifisches und kontextspezifisches System, welches dem Schallereignis eine Bedeutung zuordnet und somit das eigentliche Höreignis bildet. Raake und Blauert [174] ermöglichen mit ihrem Modell die Messung von Plausibilität auf einem indirektem Weg, indem der Hörer hinsichtlich seines Involvierens bzw. Immersion in die auditorische Szene, also nach seinem Höreignis, beobachtet bzw. befragt wird.

Das Modell besitzt für die eigene Arbeit vor allem hinsichtlich der beschriebenen symbolischen Repräsentation des Schallereignisses und der internen Plausibilitätsprüfung Relevanz. Weiterhin werden räumliche Qualitätsmerkmale vorgeschlagen, mit denen sich die Plausibilität einer

auditorischen Szene beschreiben lässt. Zu den genannten Merkmalen zählen: Tiefe, Distanz, Lokalisierung, Eingehülltsein, Externalisierung, Räumlichkeit und Timbre-Merkmale [174].

**Schematismus** - Ein weiterer Erklärungsansatz zum Verstehen der kognitiven Vorgänge im Bereich des menschlichen Erkenntnisgewinnes ist in Abbildung 1.7 als Schematismus verdeutlicht. Die Abbildung zeigt ein Modell der Vorgänge der Assimilation und Akkommodation zur Erreichung eines Gleichgewichtszustandes (Äquilibrium) zwischen Perzepten der wahrgenommenen Realität und Perzepten der erwarteten bzw. bis dahin erlernten Realität. Der beschriebene Erklärungsansatz basiert auf den erkenntnistheoretischen Betrachtungen der geistigen Entwicklung beim Menschen von Jean Piaget [166].

Der in dem Modell vorgeschlagene Regelkreis von Assimilation und Akkommodation zwischen wahrgenommenen Objekt und kognitiven Prozessen/Wissensmustern stellt einen Kern des adaptiven Charakters der Wahrnehmung dar. Dieses wird für die eigene Arbeit vor allem hinsichtlich des Abgleichs zwischen wahrgenommenem Hörereignis und intern repräsentiertem erwarteten Hörereignis als relevant eingeschätzt.

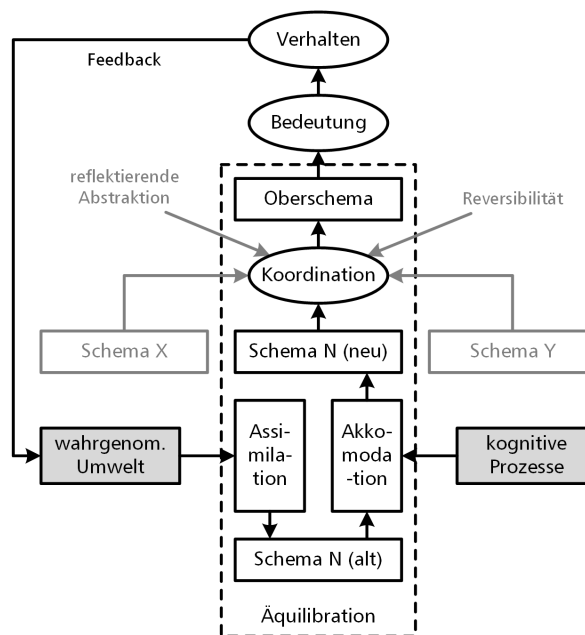


Abbildung 1.7: Schematismus des menschlichen Erkenntnisgewinnes; angelehnt an Piaget [166].

Piaget sieht als grundlegendes Erkenntniselement über das Wissen seiner Umwelt ein im Gehirn gespeichertes Wissens- und Verhaltensmuster an. Die von ihm besprochenen Schemata sind für jedes Individuum verschiedene Kategorien und Verbindungen aus Kategorien. In diese werden Wahrnehmungsereignisse eingeordnet. Die Schemata sind dabei hochgradig zeitvariant. Sie verändern sich aufgrund eines permanenten Abgleichs zwischen wahrgenommenen und

erwarteten Ereignismustern. Diese Anpassung erfolgt durch Prozesse der Assimilation und Akkommodation [107, 166].

Die Assimilation beschreibt die Einpassung neu erlangter Erfahrungen in ein bereits vorliegendes Schema. Eine Besonderheit liegt darin, dass das Wahrgenommene so umgedeutet werden kann, bis es für vorhandene Muster passend ist. Der Prozess ist dabei nicht fest auf eine Wahrnehmungsereignis-Schemata-Relation beschränkt. Die Assimilation wiederholt sich für andere aber ähnliche Anwendungsbereiche und wird dabei für die neuen Wahrnehmungsereignisse generalisiert aber auch differenziert [107, 166].

Die Akkommodation beschreibt die Vorgänge, wenn aufgrund von zu großer Verschiedenheit ein Wahrnehmungsereignis keinem Schema zugeordnet werden kann oder keine Anpassung stattfindet. Das kognitive System versucht ausgehend von einem vorhandenen Schema aus den neu erfahrenen Eigenschaften/Erkenntnissen ein neues Schema zu entwickeln [107, 166].

Die verbundenen Prozesse der Assimilation und Akkommodation streben in Richtung eines Gleichgewichtszustandes (Äquilibrium). Das Streben nach Gleichgewicht wird durch den Aufbau und die Anpassung von neuen und vorhandenen Schemata und durch Verknüpfung der Schemata umgesetzt. Getrieben wird dieser Vorgang durch fehlgeschlagene Assimilations- und Akkommodationvorgänge sowie durch Widersprüche und kognitive Konflikte zwischen wahrgenommener Realität und erwarteter Realität [166].

**Systemansatz der Arbeit** - Der Systemansatz dieser Arbeit ist durch eine Anpassung und Zusammenführung der verschiedenen vorher beschriebenen Modelle geprägt. Hierfür sind im Besonderen die von Silzle [196] vorgeschlagene „general quality taxonomy for an interactive auditory virtual environment“ (Abbildung 1.5) und der von Jekosch [123] und Raake [173] vorgeschlagene Qualitätsentstehungsprozesses einbezogen. Es wird versucht, aus den unterschiedlichen Erklärungsmodellen der Wahrnehmung und unter Beachtung der in der Erkenntnistheorie erlangten Modelle [166] eine eigene Denkweise zu entwickeln. Diese solle einen geeigneten Rahmen für die nachfolgenden Hörversuche und Systemanpassungsmethoden bilden. Abbildung 1.8 verdeutlicht schematisch den Qualitätsbildungsprozess.

Die Anpassung umfasst hierbei in erster Linie nicht den Qualitätsbildungsprozess selber, sondern die Miteinbeziehung des multimedialen technischen Systems bei der Herausbildung der QoE sowie Rückkopplungsmöglichkeiten zum multimedialen System. Das technische System steht dabei in seinem Verwendungskontext. Die QoS des multimedialen Systems wird durch die technischen Qualitätselemente des Systems und durch kontextabhängige Parameter beschrieben. Die Qualitätselemente umfassen die Systemparameter zur Beschreibung der realisierbaren technischen Qualität. Die Kontextparameter beschreiben hingegen die benötigte technische Qualität und damit verbundene benötigte Qualitätselemente. Der Maßstab für den Bedarf der benötigten QEs wird durch die Anforderungen der QoE des einzelnen Nutzers festgelegt.

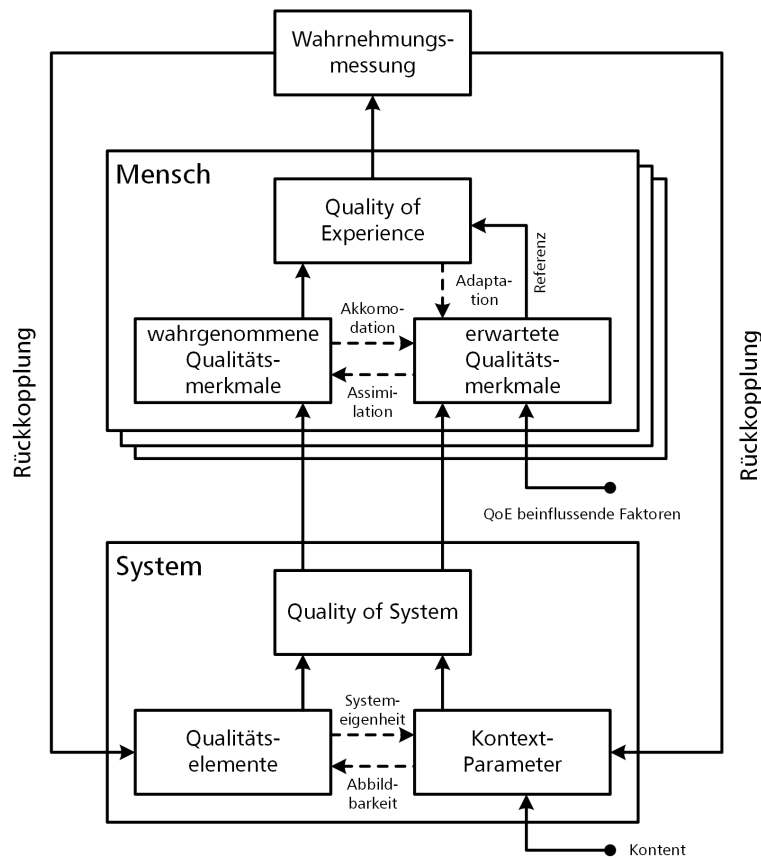


Abbildung 1.8: Angepasster Qualitätsbildungsprozess.

Die durch das multimediale System wiedergegebenen Inhalte werden durch die Sinnesorgane der Rezipienten aufgenommen. In weiteren Verarbeitungsstufen des Gehirns bildet sich das Wahrnehmungsereignis heraus. Das Wahrnehmungsereignis lässt sich durch seine QoE beschreiben, welche in einem Vergleichs- und Beurteilungsprozess gebildet wird und die erlebte Qualität beschreibt. Die wahrgenommenen Qualitätsmerkmale werden mit erwarteten Qualitätsmerkmalen abgeglichen. Hierbei findet ein permanenter Prozess des Angleichens und Anpassens statt.

Die kontextabhängigen Parameter des multimedialen Systems und andere QoE beeinflussende Faktoren wirken sich sowohl auf die wahrgenommenen als auch auf die erwarteten Qualitätsmerkmale aus. Zu den anderen QoE beeinflussende Faktoren können bspw. physikalische, zeitliche, ökonomische und soziale Eigenschaften des Nutzers und/oder Umfeldes zählen [126]. In der vorliegenden Arbeit werden diese Faktoren nicht betrachtet und für die durchgeführten Untersuchungen als wenig kontrollierte aber konstante Einflussparameter angesehen.

Interne Rückkopplungen im Qualitätsentstehungsprozess, wie bspw. Adaptations- und Trainingseffekte, können die interne Referenz verändern und damit den Vergleichs- und Beurteilungsprozess beeinflussen. Weiterhin können Erwartungen des Rezipienten die wahrgenommenen Qualitätsmerkmale beeinflussen. Hierfür sei als Beispiel auf den McGurk-Effekt

[149] verwiesen, in dem die durch visuelle Merkmale gesteuerte auditive Erwartung zu einer geänderten auditiven Wahrnehmung führt.

Mithilfe von perzeptiven Untersuchungen lässt sich die QoE messbar machen. Voraussetzung hierfür ist, dass es den Rezipienten ermöglicht wird, die wahrgenommene Qualität anhand von Qualitätsmerkmalen beschreibbar und auf Skalen bewertbar zu machen. Der Einfluss des Kontextes auf die Ausprägung der Qualitätsmerkmale und der Herausbildung der QoE soll somit beschreibbar werden. Der Einfluss der kontextabhängigen Qualitätsparameter auf den Vergleichs- und den Beurteilungsprozess zwischen den Qualitätsmerkmalen und den technischen Qualitätselementen soll dadurch indirekt darstellbar werden.

Die in den Wahrnehmungsexperimenten gewonnenen Erkenntnisse der Bewertung der QoE sollen eine Rückkopplung in Richtung des technischen Systems ermöglichen. Die Rückkopplung kann dabei zum einen direkt technische Qualitätselemente des Systems und zum anderen den Anwendungskontext selber ändern. Das Optimierungsziel ist hierbei das Erreichen eines hohen Grades an Plausibilität bzw. erlebter Qualität (QoE).

Das technische System interagiert weiterhin hinsichtlich seiner Qualitätselemente und der Kontextparameter. Einzelne oder Gruppen von Qualitätselementen stellen Systemeigenheiten dar, die sich auf den Kontext der Verwendung auswirken. Eine solche Abhängigkeit liegt bspw. dann vor, wenn ein räumliches audiovisuelles System eine Abbildung synthetisiert, in der die räumliche Audioszene nicht mit der visuellen Szene gemäß Objektpositionen übereinstimmt (siehe stereoskopische Videoprojektion und Binauralsynthese oder räumliche Lautsprecherwiedergabe). In diesem Fall würde dies zu einer audiovisuellen Divergenz führen. Im Gegensatz wirken sich Kontextparameter auf die technischen Qualitätselemente hinsichtlich Abbildbarkeit aus. Ein Mangel in der Abbildbarkeit würde bspw. dann vorliegen, wenn eine Synthese einer räumlichen Audioszene mithilfe von BRIRs angestrebt ist und das Audiowiedergabesystem nicht in der Lage ist, einen ausreichend breiten Frequenzbereich abzubilden. In diesem Falle würden bspw. hochfrequente spektrale Merkmale der BRIRs nicht abgebildet werden und es kann dadurch zu Problemen in der räumlichen Audiowahrnehmung kommen. Eine adaptive Steuerung des Systems gemäß festgelegter erreichbarer wahrgenommener Qualität setzt als Regelkreis an dieser Schnittstelle an.

## 1.5 Struktur der Arbeit

Im Kapitel 2 werden die physiologischen Gegebenheiten, die signalverarbeitenden Strukturen im Gehirn sowie die Hörbarmachung über ein Binauralsynthesystem vorgestellt. Kapitel 3 benennt die untersuchten Merkmale räumlicher Audioqualität und Einflüsse des Verwendungskontextes. Kapitel 4 stellt die Hörexperimente vor, welche sich mit der Untersuchung des Einflusses kontextabhängiger Qualitätsparameter auf die Wahrnehmung einer räumlichen Audioszene befassen. Ausgehend davon werden in Kapitel 5 Methoden vorgestellt, welche gemessene BRIRs so anpassen, dass eine veränderte, aber möglichst plausible räumliche Wahrnehmung entsteht. Kapitel 6 schließt die Ausarbeitung mit einer Zusammenfassung ab. Weiterhin wird ein Ausblick auf ein Modell der Externalisierung basierend auf den Erkenntnissen der Hörexperimente gegeben. Die einzelnen Experimente sind in der Arbeit mit „Exp.“ und einer römischen Ziffer gekennzeichnet.

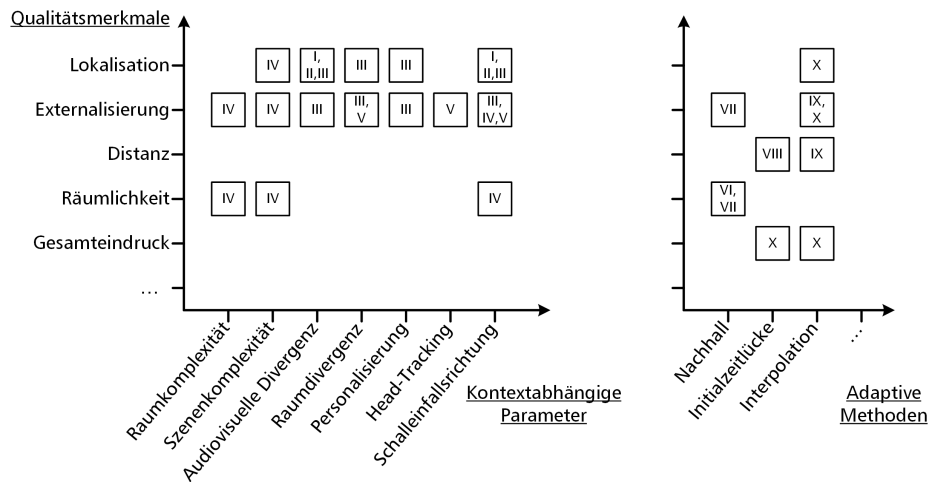


Abbildung 1.9: Zuordnung der Experimente zu den Qualitätsmerkmalen und kontextabhängigen Parametern; Zuordnung der adaptiven Methoden zu den Qualitätsmerkmalen.

Abbildung 1.9 verdeutlicht eine Zuordnung der untersuchten Qualitätsmerkmale und kontextabhängigen Parameter mit den dazugehörigen Experimenten aus Kapitel 4. Weiterhin werden die in Kapitel 5 vorgestellten Methoden einer adaptiven Binauralsynthese den untersuchten Qualitätsmerkmalen zugeordnet.

## Kapitel 2

### Physik des räumlichen Hörens

Das binaurale System [...] ist ein zweikanaliges Aufnahmesystem, bei dem Mikrofone an beiden Ohrkanaleingängen bei einem Menschen oder einem Kunstkopf platziert werden. [...] Es besteht der Wunsch, den Schall von den Ohren am Ort des Originalgeschehens an die Ohren des Hörers zu bringen. Das binaurale System kann die originalen Schallfelder mit den komplexen Spektraldifferenzen und den Laufzeitdifferenzen allein über Kopfhörer eindrucksvoll wiedergeben.

---

*Eberhard Sengpiel*  
*[www.sengpielaudio.com](http://www.sengpielaudio.com) [192]*

## 2.1 Physiologie

Das Hören von Luftschall basiert auf der Wandlung der Bewegung der Luftmoleküle in elektrochemische Signale in den auditiven Nervenfasern. Die sich im cortischen Organ befindlichen Haarzellen transformieren hierbei mechanische Schwingungen in elektrochemische Signale. Stammesgeschichtlich treten die ersten Haarzellen auf der Oberfläche von kieferlosen Fischen vor etwa 450 Mio. Jahren auf [101]. Diese Haarzellen wurden vom Meerwasser umspült und dienen der Lokalisierung von lokalen Wasserströmungen und somit der Orientierung der frühen Fische. Abbildung 2.1 (links) zeigt ein Bündel von Wasser umspülten Haarzellen auf der Oberfläche einer Zebrafischlarve. Die willkürlich über den Körper verteilten Haarzellen ordneten sich allmählich in Linien längs der Fische an. Zuerst entstanden einfache Gruben und daraus dann später, durch weiteren Rückzug ins Gewebe, das heutige Seitenlinienorgan der Fische (siehe Abbildung 2.1).

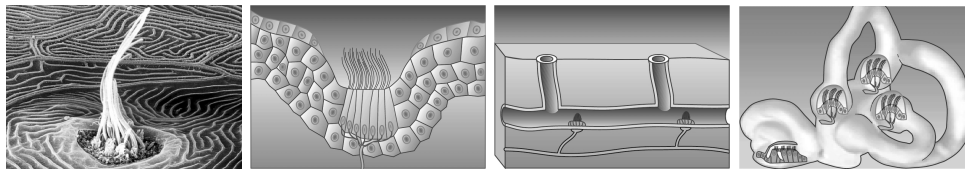


Abbildung 2.1: v.l.n.r.: Haarzelle auf der Oberfläche einer Zebrafischlarve, Grube mit Haarzellen, Seitenlinienorgan, Cochlea und Labyrinth, Abbildungen aus Grothe [101]

Die an Land lebenden Tiere entwickelten aus dem Seitenlinieorgan das neue Lagesinnesorgan (Labyrinth) und die Cochlea. Zur Aufnahme von Luftschall entstand das heutige Mittelohr mit seinen übertragenden Eigenschaften. Auch heute finden sich noch Ähnlichkeiten in Struktur und Wirkung zwischen dem Seitenlinienorgan der Fische und dem menschlichen Innenohr mit seinen Haarzellen (cortische Organ). So sind in beiden Organen die reizwandelnden Sinneszellen linear angeordnet und werden von Flüssigkeit mit darin aufgelösten Salzen umspült. Die Signalgewinnung, Signalableitung und Signalverarbeitung weisen ebenfalls einen sehr hohen Grad an Übereinstimmung auf.

### 2.1.1 Das Ohr als sensorisches System

Nachstehende Abschnitte geben einen Überblick zum Aufbau des Hörorgans sowie weitere Verarbeitungsstufen im Gehirn. Hierbei soll im Besonderen auf die Reizaufnahme, die physiologischen Eigenschaften, der Reizweiterleitung und Reizverarbeitung eingegangen werden, die an einem Zustandekommen eines räumlichen Hörereignisses beteiligt sind.

**Außen- und Mittelohr** - Die Aufgabe des Außenohres ist die gerichtete Weiterleitung des Luftschalles an das Trommelfell. Die Ohrmuschel, der Kopf, der Torso und der Gehörgang verändern den Schall und geben ihn an das Trommelfell weiter. Die Ohrmuschel wirkt vergleichbar einem richtungsabhängigen Filter und stellt neben den interauralen Unterschieden wichtige Merkmale für das räumliche Hören zur Verfügung. Systemtechnisch wird dieses



Verhalten in der kopfbezogenen oder Außenohr-Übertragungsfunktion (engl.: head-related transfer function; HRTF) abgebildet. Eine weiterführende Ausführung hierzu ist Kapitel 2.2.1 zu entnehmen.

Der Gehörgang bildet aufgrund seiner Form einen Resonanzraum, welcher zu einem Anstieg des Schalldruckpegels führt. Seine Wirkung ist unabhängig von der Schalleinfallrichtung. Die Resonanzfrequenz liegt im Durchschnitt bei ca. 2,5 kHz. Der Gehörgangseingang und die Ohrmuschel verursachen Resonanzen bei ca. 2 kHz und ca. 5 kHz [238]. Die Zusammenfassung der beiden Hauptresonanzbereiche resultiert in einem Schalldruckgewinn in einem Bereich von 2 bis 7 kHz [181]. Abbildung 2.2 verdeutlicht die durchschnittlichen Schalldruckgewinne verschiedener Teile des Außenohres.

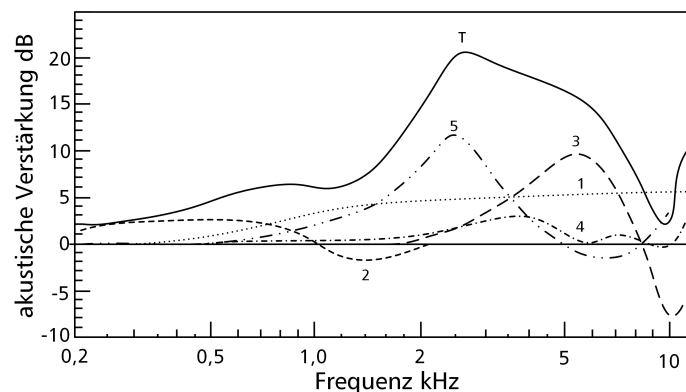


Abbildung 2.2: Schalldruckgewinne verursacht durch verschiedene Teile des Außenohres und Körpers; Schallstimulus bei  $45^\circ$  lateraler Auslenkung; T=gesamt, 1=sphärischer Kopf, 2=Rumpf und Hals, 3=Cavum Conchae, 4=Helix, 5=äußerer Gehörgang und Trommelfell; Abbildung nachgezeichnet aus Shaw [193]

Das Mittelohr dient der Weiterleitung der Schallsignale vom Trommelfell in die Cochlea. Dabei wird die Schwingung der Luftmoleküle in eine mechanische Schwingung des Trommelfells und der Gehörknöchelchen überführt. Die wichtigste Aufgabe ist dabei die Schallimpedanzanpassung zwischen luftgefülltem Außenohr und flüssigkeitsgefüllter Cochlea. Wären die Knöchelchen nicht vorhanden und würde der Luftschall direkt auf die Cochlea auftreffen, würden lediglich 2% der Schallenergie übertragen werden. Aufgrund ihrer Bauform werden ca. 60% der Schallsignalenergie in die Cochlea überführt [238]. Bewirkt wird diese Anpassung durch die Hebelwirkung der Knöchelchen, mit der eine größere Kraft auf das ovale Fenster der Cochlea wirken kann. Weiterhin ist die Fläche der Steigbügelplatte deutlich kleiner als die des Trommelfells, wodurch eine Druckerhöhung erzielt wird. Eine weitere Funktion der Knöchelchen ist die Einstellung der Trommelspannung. Dafür befinden sich Muskeln im Mittelohr. Der Musculus Tensor Tympani setzt am Hammer an und verursacht dadurch eine Änderung der Trommelspannung. Auf Grund seiner im Vergleich zur Schalleinwirkung eher langen Reaktionszeit ist er wahrscheinlich nicht an einer Schutzwirkung für das Innenohr beteiligt. Er vermindert aber vermutlich das Klirren bei der Beschallung mit hohen Frequenzen

[87]. Der Musculus stapedius führt bei Kontraktion zu einem Heraushebeln des Steigbügels aus dem ovalen Fenster. Damit werden zu große Bewegungen des Steigbügels verhindert. Er schützt damit das Innenohr vor zu hohen Schallintensitäten. Für weiterführende Informationen ist hier auf die Literatur verwiesen [87, 238].

**Die Cochlea** - Das Mittelohr wandelt den aufgenommenen Luftschall in eine hydrodynamische Druckänderung in der Cochlea um. Die Cochlea selber besitzt 2,5 Windungen und besteht aus den Bereichen Scala Vestibuli, Scala Tympani und Scala Media. Getrennt sind diese Bereiche durch die Reissnersche Membran und die Basilarmembran (Abbildung 2.3).

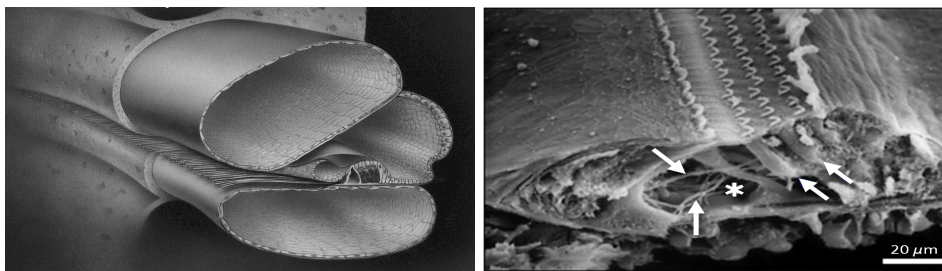


Abbildung 2.3: links: Schema Cochlea-Schnitt, Abbildung aus [69]; rechts: Cortisches Organ einer Ratte, linke Pfeile: Hörnervenableitungen, rechte Pfeile: äußere Haarzellen, Sternchen: Cortischer Tunnel, Abbildung nach [158].

Zur Weiterleitung der mechanischen Schwingung der Gehörknöchelchen drückt der Steigbügel auf das ovale Fenster der Cochlea und verursacht somit eine Druckwelle. Diese Druckwelle pflanzt sich entlang der Basilarmembran fort und verursacht eine Wanderwelle auf der Basilarmembran [62]. Diese Welle erreicht je nach ihrer Frequenz an bestimmten Stellen der Membran ihr Maximum (Ortsprinzip). Hohe Frequenzen haben ihr Maximum am ovalen Fenster, dort ist die Breite der Basilarmembran klein und tiefe Frequenzen an der Spitze der Cochlea, dort ist die Breite der Membran groß. Durch das Schwingen der Basilarmembran werden die auf ihr befindlichen Sinneszellen (Haarzellen) relativ gegenüber der darüber liegenden Tektorialmembran ausgelenkt. Der Ort wo dies passiert wird als Cortische Organ bezeichnet.

**Das Cortische Organ** - Das Cortische Organ stellt das Transformationsorgan dar (Abbildung 2.3). Hier wird mithilfe der Haarzellen eine Druckänderung in elektrochemische Signale umgewandelt. Es besteht beim Menschen aus ca. 3500 inneren Haarzellen in einer Reihe und parallel dazu in drei Reihen ca. 12000 äußeren Haarzellen [238]. Nur die inneren Haarzellen dienen dem direkten Hören. Die äußeren Haarzellen dienen der Erhöhung der Frequenzselektivität und dem Verstärken von kleinen Amplituden [51, 225]. Dabei können sie sich ähnlich einem Muskel strecken und wieder zusammenziehen und somit die Auslenkung sowie Steifigkeit der Basilarmembran gegenüber der Tektorialmembran beeinflussen.

**Die inneren Haarzellen** - Die inneren Haarzellen sind in ihrer Funktion ein mechanisch-elektrischer Wandler. Mit ihren Härchen (Stereocilien) folgen sie der Schwingung der Basilarmembran. Gedämpft werden sie dabei durch die Zähigkeit der Lympheflüssigkeit, in der sie sich bewegen und der Eigensteifigkeit der Härchen selbst. Die Haarzellen selbst erregen die an sie über einen synaptischen Spalt angeschlossenen auditorischen Nervenfasern (ANFs). Die von Békésy [62] vorgeschlagene Wanderwellentheorie mit ihrem Ortsprinzip scheint allerdings unzureichend zur Erklärung der Sinneswahrnehmung von komplexen Signalen und bei Vorhandensein von Hintergrundgeräuschen zu sein. In diesen Situationen sind nahezu alle Haarzellen und ANFs in Sättigung [238]. In Untersuchungen von Sachs und Young [185] wird gezeigt, dass die Aktivierung der ANFs dem Frequenzverlauf des Audiosignales folgen. Erklärt wird dies durch die Depolarisierung der inneren Haarzellen lediglich bei Bewegung der Stereocilien in eine Richtung und Hemmung bei Bewegung in die andere Richtung. In den Zellen findet dadurch ein Enkodieren der Phase und somit eine Periodizitätsanalyse des Signals statt (Abbildung 2.4). Diese Phasenkopplung lässt sich bis zu einer Audiosignalfrequenz von ca. 5 kHz nachweisen [238]. Die Fähigkeit des auditorischen Systems zeitliche Informationen zu extrahieren und zu verarbeiten, wird hauptsächlich auf die Phasenkopplung der ANFs zurückgeführt [221]. Sie stellt somit eine wesentliche Voraussetzung zur binauralen Signalverarbeitung dar.

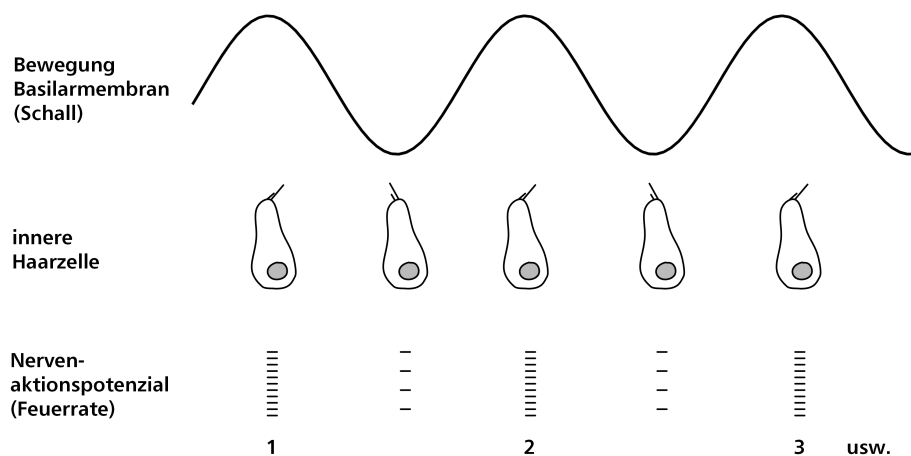


Abbildung 2.4: Phasenkopplung einer Haarzelle; Abbildung nachgezeichnet aus Zenner [238]

Die Stereocilien bestehen aus drei Reihen mit langen, mittleren und kurzen Härchen. Diese Härchen sind über Tips miteinander verbunden. Die Bewegung der Stereocilien lässt sich über ein System gekoppelter Oszillatoren modellieren [203]. Befindet sich die Haarzelle in Ruhe, so herrscht im Inneren der Zelle eine geringere Konzentration an  $K^+$ -Ionen als außerhalb der Zelle (Abbildung 2.5). Es liegt das Ruhepotenzial an. Werden die Stereocilien medial in Bewegung versetzt, so öffnen sich am oberen Ende der Zelle Ionenkanäle und es strömen  $K^+$ -Ionen in die Zelle hinein. Durch diesen Ionenfluss steigt die Spannung innerhalb der Zelle.

Es findet eine Depolarisierung statt. Werden die Stereocilien in die andere Richtung bewegt sinkt die Spannung [238].

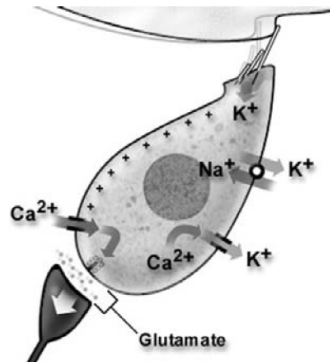


Abbildung 2.5: Entstehung eines Aktionspotenzials, Abbildung aus [158]

Diese Depolarisierung öffnet in der Zellmembran weitere Ionenkanäle durch die  $Ca^{2+}$ -Ionen einströmen können. Dies hat zur Folge, dass Neurotransmitter Richtung präsynaptische Membran wandern und dort in den Synaptischenspalt entlassen werden. Diese setzen sich dann auf Rezeptorproteinen auf der postsynaptischen Membran der auditiven Nervenfasern ab und veranlassen das Öffnen von  $Na^+$ -Ionenkanälen in der postsynaptischen Membran. Steigt der Einstrom von  $Na^+$ -Ionen auf ein bestimmtes Niveau und erreicht die damit entstehende Spannung einen Schwellwert, entsteht ein postsynaptisches Aktionspotenzial (Spike) in der auditiven Nervenfasern [203].

**Aktionspotenziale in den auditiven Nervenfasern** - Eine auditive Nervenfasern ist nicht in der Lage kontinuierlich ein Signal zu übertragen. Nach der Entstehung eines Aktionspotenzials (Spike) braucht die Nervenfasern eine gewisse Zeit, bis sie erneut einen Spike feuern kann. Diese Zeit der Inaktivität wird als Refraktärphase bezeichnet. Eine Nervenfasern überträgt somit Zeitpunkte von gefeuerten Spikes. Ein akustisches Signal wird durch eine Sequenz von Feuerzeitpunkten für jede auditive Nervenfasern  $i$  beschrieben [203].

$$T_i = \{t_1^1, \dots, t_i^{n_i}\}$$

Weiterhin weisen die auditiven Nervenfasern eine hohe Frequenzselektivität auf. Jede Nervenfasern besitzt eine für sie charakteristische Frequenz, bei der sie den höchsten Grad an Aktivität erreicht. Eine innere Haarzelle leitet ihre Erregung in mehrere Nervenfasern des Typs I ab. Dagegen steht eine Anzahl von äußeren Haarzellen, die ihre Erregungen gebündelt in je eine Nervenfasern des Typs II ableiten. Typ I und Typ II Nervenfasern besitzen einen unterschiedlichen Aufbau und werden in der späteren Signalverarbeitung auch getrennt bewertet [158].

**Audiorepräsentation** - Ein Schallereignis wird durch eine Sequenz von Spikefeuerzeitpunkten für jede auditive Nervenfaser (ANF) repräsentiert. Für jede Nervenfaser zeigt sich eine spezifische Antwortzeit. Werden die Aktionen aller Nervenfasern betrachtet, ergibt sich das in Abbildung 2.6 unten dargestellte Diagramm.

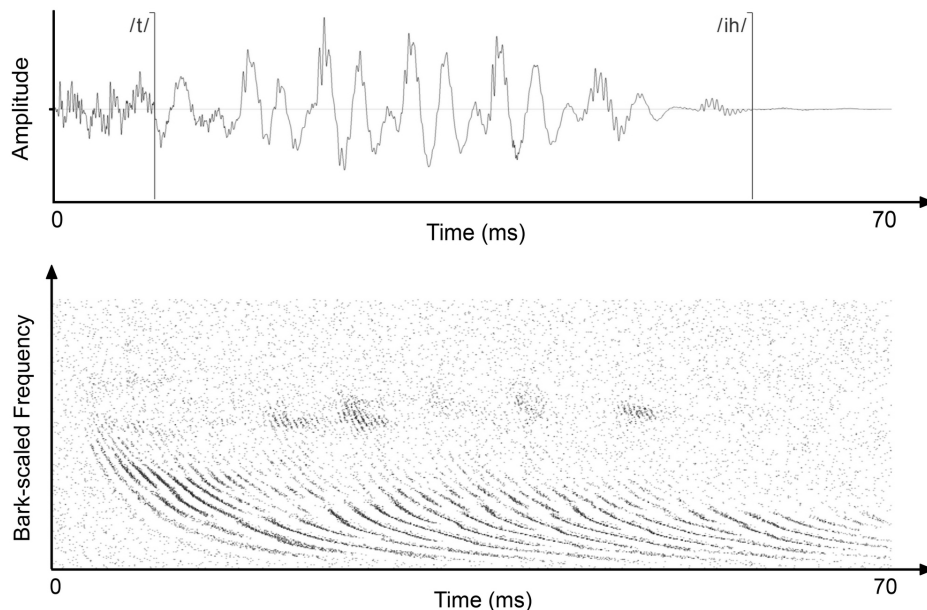


Abbildung 2.6: Computersimulation der Antwort einer Population auditorischer Nervenfasern auf die Silbe 'ih'; Abbildung aus Harcos [104]

Auf der x-Achse ist die Antwortzeit der Nervenfasern aufgetragen. Als Datenpunkte sind die Feuerzeitpunkte der ANFs abgebildet. Auf der y-Achse der unteren Grafik befinden sich die einzelnen Nervenfasern mit ihren charakteristischen Frequenzen. Hierdurch ist auch der Ort auf der Basilarmembran bestimmbar. Als Skala der Frequenz ist eine Einteilung gemäß der Analysefähigkeit des Gehörs von Frequenzgruppen, der Bark-Skala, gewählt [240, 241]. Ein Schallereignis repräsentiert sich als ein spezifisches Muster in den Nervenfasern. Es findet eine Spikecodierung in jeder auditiven Nervenfaser statt. Die Spikecodierung stellt die Zeit als zusätzliche Kodierungsdimension für den neuronalen Wahrnehmungsprozess des Hörens zur Verfügung. Die Betrachtung der mittleren Feuerraten der Nervenfasern würde für unterschiedliche Reizungen nahezu identisch sein. Auch würde die Betrachtung der Gesamtaktivität aller Nerven über die Zeit allein ebenfalls keine wesentlichen Unterschiede für verschiedene Reizungen aufzeigen können. Werden beide Dimensionen betrachtet, ergibt sich ein raumzeitliches Muster oder Perzept, welches bei unterschiedlichen Reizungen auch unterschiedlich strukturiert ist.

### 2.1.2 Auditorischer Pfad

Abbildung 2.7 zeigt schematisch den aufsteigenden auditorischen Pfad für die linke Seite beginnend mit den auditorischen Nerven. In den nachstehenden Abschnitten wird im besonderen auf den Cochlear Nucleus und weitere neuronale Bereiche als relevante Verarbeitungsstufen zum räumlichen Hören eingegangen.

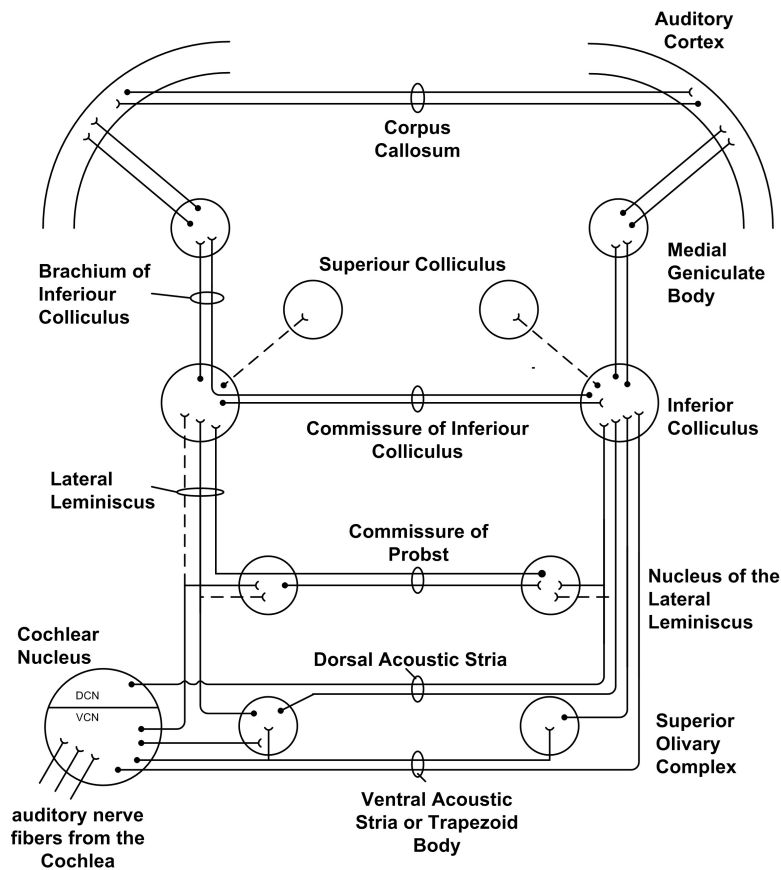


Abbildung 2.7: Aufsteigender auditorischer Pfad (linke Seite); Abbildung nachgezeichnet aus Ehret und Romand [84]

Die aufsteigenden auditorischen Nervenfasern (ANFs) leiten die Information aus den Haarzellen der Cochlea in den Cochlea Nucleus weiter. Neben diesen aufsteigenden Nerven existieren absteigende Nervenfasern aus dem Tractus olivocochlearis, die an den Haarzellen im cortischen Organ enden. Ihre Aufgabe besteht in einer Regulierung der Empfindungsfähigkeit der Haarzellen und dient somit wahrscheinlich der Adaption der Sinneszellen. Ein Beispiel hierfür ist das aktive Heraushören einer Stimme aus Hintergrundgeräuschen [87].

**Cochlear Nucleus** - Der CN ist ein tonotop und somit anatomisch gemäß der Frequenz des Signals organisiertes peripheres Zentrum im zentralen auditorischen System. Er umfasst die erste Merkmalskodierung der Audiosignale nach dem Cortischen Organ. Detailliertere Informationen zum Aufbau und Wirkung des CN ist der Veröffentlichung [18] zu entnehmen.

Die auditorischen Nervenfasern aus der Cochlea stellen die Haupteingangssignale dar. Die ANFs sind dabei gemäß ihrer spontane Feueraktivität in „high“, „middle“, und „low spontaneous rate“ (H-, M-, LSR) eingeteilt. Die verschiedenen Regionen des CN werden durch unterschiedliche ANF-Typen in Abhängigkeit von der Intensität des Schallsignals angeregt. Weiterhin zeigen sich zahlreiche Eingänge von nicht auditorischen Nervenfasern aus höheren Regionen des auditorischen Pfades und von außerhalb des auditorischen Systems [84].

Hauptmerkmale, die innerhalb des CN enkodiert werden sind bspw. die Amplitudenmodulation (AM) und die Periodizität des Audiosignals. Frisina et al. [95] fanden Neuronen im ventralen CN (VCN) zum Enkodieren der AM. Diese Zellen sind sensibel auf eine charakteristische Frequenz der AM. Im posterior-ventralen Teil des CN (PVCN) sind Zellen identifiziert, die den Beginn („onset“) eines Sinnesereignisses detektieren und ebenfalls eine wichtige Rolle in der Erkennung von Periodizität spielen [84]. Im anterior-ventralen CN (AVCN) liegen Zellen vor, welche Eingänge von zahlreichen ANFs mit unterschiedlicher charakteristischer Frequenz erhalten [74]. Diese Neuronen verhalten sich ähnlich wie eine Kreuzkorrelation zur Erkennung von Periodizitätsmustern [224].

Die Ausgänge des CN enden u. a. in der medialen superioren Olive (MSO) und der lateralen superioren Olive (LSO) des superioren Olivenkomplexes (SOC), sowie im Inferior Corriculus (IC) der ipsi- und kontralateralen Seite. Diese Regionen stellen die Hauptareale zur Enkodierung von binauralen Merkmalen, wie Laufzeit- und Intensitätsunterschieden, dar.

Die Ausgänge der Zellen im AVCN weisen eine präzisere Enkodierung der Phase der akustischen Schallwellen als die ANFs selber auf [125, 145]. Die zwei CNs bilden eine Vorverarbeitungsstufe zur Extraktion von binauralen Merkmalen (siehe [84, 221]). Oertel und Young [163] wiesen nach, dass die hemmenden Schaltungen zwischen den Neuronen im dorsalen CN (DCN) eine Enkodierung von lokalen Maxima und Minima im Amplitudenspektrum eines Schallereignisses ermöglichen. Lokale Maxima und Minima entstehen beispielsweise durch richtungsabhängige Filterungen mit den HRTFs. Der DCN stellt somit wahrscheinlich eine entscheidende Vorverarbeitung zur Erkennung von spektralen Merkmalen im IC dar. Weiterhin wird der DCN von zahlreichen multimodalen Eingängen aus höheren Bereichen des Gehirns angeregt. Dies lässt den Schluss zu, dass diese Eingänge Informationen von „oben nach unten“ abbilden. .

**Binaurale Merkmalsextraktion im auditorischen Pfad** - In den nächsten Stufen des aufsteigenden auditorischen Pfades werden die binauralen Merkmale Laufzeitdifferenz und Intensitätsdifferenz zwischen contra- und ipsilateraler Seite ausgewertet. Hierfür sind die Bereiche der MSO und LSO des SOC innerhalb des zentralen Nervensystems lokalisiert [84, 221] u. a.. Sie erhalten ihre Eingänge aus den CNs beider Seiten und sind tonotop strukturiert. Der Inferior Colliculus stellt (IC) eine weitere signalverarbeitende Struktur innerhalb des aufsteigenden auditorischen Pfades dar [84]. Der IC erhält seine Eingänge vom CN, SOC und dem lateralen Lemniscus. Ein Viertel der im IC befindlichen Neuronen sind empfindlich für Signale, die von der kontralateralen Seite stammen. Diese Neuronen reagieren auf monaurale

Signale eines Ohres unter Beachtung der Signale des gegenüberliegenden Ohres. Die restlichen Neuronen reagieren auf binaurale Signale, die von beiden Seiten des auditorischen Pfades stammen. Die Population an Neuronen unterteilt sich in Neuronen, die sensitiv für interaurale Intensitätsunterschiede, sowie Neuronengruppen, die sensitiv für interaurale Phasenunterschiede sind. Neben zeitlichen und Intensitätsunterschieden ist der IC auch sensitiv gegenüber spektralen Unterschieden zwischen den Audioinformationen des linken und rechten Ohres und enkodiert somit wahrscheinlich auch spektrale Unterschiede, die durch die Außenohrübertragungsfunktionen verursacht werden. Der IC enkodiert ebenfalls auch die Modulationsfrequenz von amplitudenmodulierten Schallereignissen. Neben rein auditorischen Information ist der IC auch sensibel für taktile Informationen. Der IC formiert somit verschiedene neuronale Merkmale, von denen die Qualität von Schallereignissen abgeleitet werden kann [84].



## 2.2 Akustische Signalmerkmale

Die auditive Wahrnehmung von Richtung und Distanz eines Hörereignisses basiert auf der Auswertung von verschiedenen Merkmalen durch das auditive System im Gehirn. Das Hörereignis setzt sich aus der Auswertung von zeitlichen, intensitätsabhängigen, spektralen und räumlichen Merkmalen zusammen. Neben Merkmalen, die sich aus der Analyse der Signalstruktur des Gehörten zusammensetzen, sind für die Wahrnehmung von Richtung und Distanz primär monaurale und binaurale Merkmale relevant. Monaurale Merkmale basieren auf Signalen von einem Ohr, während binaurale oder interaurale Merkmale die Signale von beiden Ohren benötigen. Die Merkmale, die sich aus der Analyse der Signalstruktur ergeben, umfassen u. a. Periodizität, Amplitudenmodulation und „Onset“-Erkennung. Neben rein akustisch getriebenen Merkmalen bildet sich in der Regel ein Hörereignis vor allem auch durch multimodale Interaktion mit anderen Sinneswahrnehmungen aus.

### 2.2.1 Monaurale spektrale Merkmale

Beim Auftreffen einer sich im Raum ausbreitenden Schallwelle auf das Trommelfell eines Hörers wird der Schall durch die Form von Kopf, Torso, Ohrmuschel und Ohrkanal in seiner spektralen und zeitlichen Zusammensetzung verändert, bevor er das Trommelfell erreicht. Die Gesamtheit aller dieser, durch Beugungen, Abschattungen und Reflexionen verursachten Veränderungen lässt sich durch die Außenohrübertragungsfunktion (engl.: Head-Related Transfer Function, HRTF) beschreiben. Der so veränderte Schall ist der wesentliche Merkmalsträger für die räumliche Wahrnehmung. Je nach Schalleinfallrichtung, Distanz zur Schallquelle und Anthropometrie des Individuums ergeben sich verschiedene Übertragungscharakteristiken (siehe als Übersicht bspw. [64, 220]).

Obwohl die Übertragungseigenschaften sich zum Teil deutlich zwischen den Individuen unterscheiden, zeigen sich auch Gemeinsamkeiten, die durch anthropometrische Maße bestimmt sind. Kopf, Schulter und Torso bestimmen den Frequenzverlauf in einem Bereich von ca. 700 Hz bis ca. 3,5 kHz [49]. In einem Bereich von ca. 3 kHz bis 4 kHz verursacht der Ohrkanal eine Resonanz. Zu höheren Frequenzbereichen hin beeinflusst vor allem die individuelle Form der Ohrmuschel die HRTF. Je nach Hörer und Schalleinfallrichtung ergeben sich charakteristische Überhöhungen und Einbrüche innerhalb der HRTF, die Informationen über die Position des Schallereignisses beinhalten. Abbildung 2.2 gibt eine schematische Übersicht über die Schalldruckgewinne am Trommelfell, welche durch verschiedene Teile des Außenohres und des Körpers verursacht werden.

**Individuelle Merkmale** - Die Geometrie und der Aufbau des Kopfes und der Ohrmuscheln unterscheiden sich von Mensch zu Mensch. Dies hat zwangsweise Auswirkungen auf die Ausprägung von monauralen und binauralen Merkmalen. Die Außenohrübertragungsfunktionen des einen Menschen können sich erheblich von denen eines anderen Menschen unterscheiden. Die Unterschiede erreichen vor allem im hochfrequenten Bereich bis zu 30 dB, welche sich auf

die Ausprägung der interauralen Intensitätsunterschiede auswirkt [90, 137]. In Untersuchungen von Seeber [189] wurden bei Messungen an neun Testpersonen maximale interauralen Laufzeitunterschiede in einem Bereich von ca. 0,68 bis 0,8 Millisekunden ermittelt. Bei der binauralen Synthese spielen die interauralen Differenzen vor allem für die Lokalisation von Schallereignissen in der Horizontalebene eine entscheidende Rolle. Durch die Verwendung von nicht individuellen Übertragungsfunktionen treten Lokalisationsfehler auf, die sich durch die Unterschiede in der Physiologie des Körpers der Personen begründen lassen. In Untersuchungen zur Lokalisation von Schallquellen in der Horizontal- und Medianebene konnte gezeigt werden, dass bei der Verwendung von fremden Übertragungsfunktionen die Lokalisationsschärfe deutlich abnimmt und es vermehrt zu Vorne-Hinten-Vertauschungen und vertikalen Lokalisationsfehlern kommt [103, 154, 191, 228]. Die Lokalisationsschärfe bei individuell angepasster Binauralsynthese ist vergleichbar mit denen in realen Hörumgebungen [189, 191].

In engem Zusammenhang mit der Lokalisationsgenauigkeit steht die Externalität von Hörereignissen. Reale Schallquellen werden in der Regel als extern wahrgenommen. Bei der Kopfhörerwiedergabe tritt der entgegengesetzte Fall der Im-Kopf-Lokalisation auf. Das auditive System hat Probleme, die Ohrsignale bei der Kopfhörerwiedergabe zu interpretieren. Alle Schallquellen scheinen sich im Kopf zu befinden [168]. Die externe Wahrnehmung von Hörereignissen (Externalisierung) wird in der Literatur weitestgehend als unabhängig von der Verwendung individueller HRTFs betrachtet [103, 136, 154]. In einem Gegensatz dazu konnte in Experimenten aber auch gezeigt werden, dass bei der Verwendung von nicht individuellen binauralen Raumimpulsantworten bzw. HRTFs aber individueller Entzerrung des Kopfhörersystems schlechtere Externalisierungs- und Lokalisationsergebnisse erzielt werden als bei Verwendung von gleichsam nicht individueller Kopfhörerentzerrung [139].

Aufgrund des engen Zusammenhangs zwischen Lokalisation und Externalität ist es schwierig, diese Parameter getrennt voneinander zu bewerten. Auf dieses Problem gehen Hartman und Wittenberg [105] näher ein. In ihrem Experiment konnte gezeigt werden, dass bei erlaubter Rotation des Kopfes beim Hörtest die Genauigkeit der Lokalisation in der Elevation positiv beeinflusst wird. Besonders deutlich ist der Effekt bei der Betrachtung von verzerrten Ohrsignalen (Veränderung der individuellen Merkmale der HRTF). Trotz der Verzerrung kann eine deutlich verbesserte Lokalisation erreicht werden. Durch Unterschiede im Amplitudenspektrum zwischen einer generalisierten und einer individuellen HRTF sind deutliche Unterschiede in der Klangfarbe wahrnehmbar. Die Klangfarbe kann als ein Parameter zur Beurteilung der Ähnlichkeit von Amplitudenspektren dienen [196]. Deutliche Klangfarbenveränderungen können störend sein, vor allem dann, wenn eine authentische Wiedergabe angestrebt wird.

### 2.2.2 Binaurale Merkmale

Binaurale Merkmale ergeben sich bei Betrachtung der Ohrsignale von linkem und rechtem Ohr. Abbildung 2.8 verdeutlicht schematisch die Entstehung von interauralen Phasen- (IPD) bzw.

Laufzeit- (ITD) und Pegeldifferenzen (ILD) an einem idealen sphärischen Kopf bei seitlichem Eintreffen einer ebenen Schallwelle.

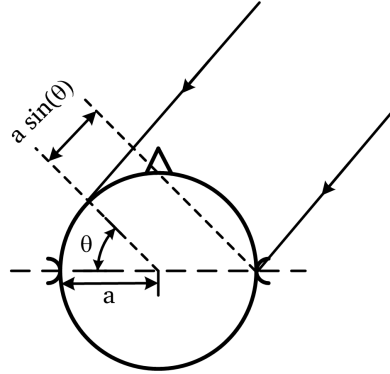


Abbildung 2.8: Interaurale Laufzeit- (ITD) und Pegeldifferenzen (ILD) an einem idealen sphärischen Kopf bei seitlichem Eintreffen einer ebenen Schallwelle; Abbildung neu gezeichnet nach Wang und Braun [223]

Die binauralen Merkmale können im Frequenzbereich aus den gemessenen oder simulierten HRTFs errechnet werden. Die interaurale Übertragungsfunktion (ITF) errechnet sich als Verhältnis der HRTFs des linken und rechten Ohres zu:

$$ITF(\alpha, \varphi, f) = \frac{HRTF_L(\alpha, \varphi, f)}{HRTF_R(\alpha, \varphi, f)} \quad (2.1)$$

mit  $\alpha$  und  $\varphi$  als Azimut- und Elevationswinkel in Grad und  $f$  als Frequenz in Hz.

**Interaurale Pegeldifferenz** - Interaurale Pegeldifferenzen (engl.: Interaural Level Difference, ILD) entstehen durch frequenz- und richtungsabhängige Beugung und Abschattung der Schallwellen am Kopf des Hörers. Ist die Wellenlänge des Schallsignals deutlich kleiner als die Ausdehnung des Kopfes, kommt es zur Abschattung der Schallwelle. Ist die Ausdehnung des Kopfes gering gegenüber der Wellenlänge, kommt es zur Beugung des Schallfeldes um das Hindernis. Hieraus ergibt sich, dass für lateral ausgelenkte Schallquellen die ILD zu niedrigeren Frequenzen geringer als zu höheren Frequenzen ist. Die durch die Wegdifferenz zwischen den Ohren hervorgerufene zusätzliche Pegeldifferenz (quadratisches Abstandsgesetz) ist im Gegensatz dazu gering und kann unter Annahme eines Fernfeldes vernachlässigt werden [234]. Bei Beschallung in der Medianebene und unter Annahme einer symmetrischen Anordnung (Kopf und Ohrmuscheln) ergeben sich keine ILDs.

Die ILD lässt sich als Differenz der Amplitudenspektren der linken und rechten HRTF definieren. Die ILD weist somit frequenzabhängige und richtungsabhängige Unterschiede auf und ergibt zu:

$$ILD(\alpha, \varphi, f) = 20 \cdot \log_{10}(|ITF(\alpha, \varphi, f)|) \quad (2.2)$$

**Interaurale Laufzeitdifferenz** - Interaurale Laufzeitdifferenzen (engl.: Interaural Time Difference, ITD) werden durch unterschiedliche Laufzeiten der Schallsignale für rechtes und linkes Ohr gebildet. Die Laufzeit zu den Ohren hängt dabei zum einen von der Schallgeschwindigkeit und zum anderen von der Entfernung der Schallquelle zu den Ohren ab. Befindet sich die Schallquelle auf der Medianebene ist die Laufzeit zu den beiden Ohren gleich. Ist die Schallquelle seitlich ausgelenkt, ergeben sich in Abhängigkeit des Schalleinfallswinkels, der Distanz und von frequenzabhängiger Beugung der Schallwellen am Kopf Laufzeitunterschiede zwischen den Ohren. ITDs entstehen über den gesamten Frequenzbereich eines Schallsignals. Ist allerdings die ITD geringer als die halbe Periode eines periodischen Schallsignals entstehen Uneindeutigkeiten. Es wird die Phasendifferenz detektiert (IPD). Die maximal mögliche ITD liegt bei Betrachtung eines gemittelten Kopfes bei ca.  $660 \mu\text{s}$  [223]. Für schmalbandige Schallsignalen werden für die Lokalisation ITDs bis zu einer Frequenz von ca. 1,5 kHz ausgewertet [223]. Diese Grenze ergibt sich aus der effektiven Distanz zwischen den beiden Ohren und der Wellenlänge des Signals. Bei höheren Frequenzen treten Uneindeutigkeiten in der Erkennung des Phasenunterschiedes beider Ohrsignale auf (engl. Interaural Phase Differenz, IPD). Für breitbandige Signale dienen die IPDs auch bei höheren Frequenzen für die Lokalisation, in dem die IPDs über mehrere Frequenzbänder aufsummiert werden [223]. Die ITD ist auf Grund von Beugungseffekten am Kopf nicht frequenzunabhängig [146]. Abbildung 2.9 verdeutlicht die Frequenzabhängigkeit für drei Frequenzbereiche unter Annahme eines idealen sphärischen Kopfes.

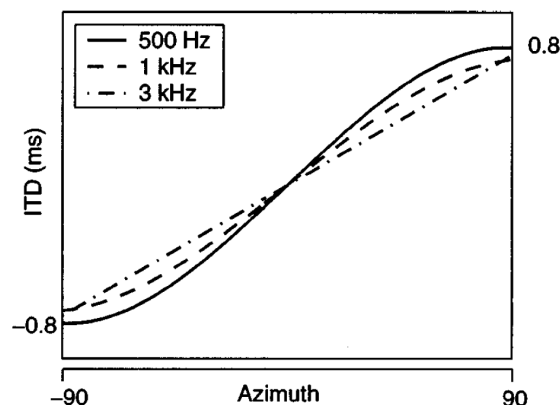


Abbildung 2.9: Funktion von Azimut-Richtung und ITD für drei Frequenzbereiche (eng.: auditory channels) mit den Mittenfrequenzen 500 Hz, 1 kHz und 3 kHz; Abbildung aus Roman et al. [179]

Die ITD lässt sich aus der HRTF des linken und rechten Ohres als Differenz der Phasenverzögerung  $\tau_{Phase}$  berechnen. Die Phasenverzögerung berechnet sich mit nachstehender Vorschrift zu:

$$\tau_{Phase} = \frac{\phi(f)}{2\pi f} \quad (2.3)$$

mit  $f$  Frequenz in Hz und  $\phi(f)$  Phasenspektrum.

Zur Berechnung der IPD bzw. ITD aus der interauralen Übertragungsfunktion lassen sich nachstehende Formulierungen nach [128] angeben:

$$IPD(\alpha, \varphi, f) = \arg(ITF(\alpha, \varphi, f)) \quad (2.4)$$

$$ITD(\alpha, \varphi, f) = -\frac{1}{2\pi} \frac{d}{df} IPD(\alpha, \varphi, f) \quad (2.5)$$

Unter der Annahme eines kugelförmigen Kopfmodells und bei Betrachtung von tiefen Frequenzen ( $ka \ll 1$ ) vereinfacht sich die ITD für tiefe Frequenzen (tF) zu [128]:

$$ITD_{tF}(\alpha) \approx \frac{6\pi a}{c} \sin(\alpha) \quad (2.6)$$

mit  $c$  Schallgeschwindigkeit in  $m/s$  und  $a$  Kopfdurchmesser in  $m$ .

Für hohe Frequenzen (hF) breiten sich die Schallwellen um das kugelförmige Kopfmodell mit Schallgeschwindigkeit aus [128]. Die ITD lässt sich somit aus der Geometrie des Kugelkopfes und dem Pfad entlang der Oberfläche des Kopfes berechnen:

$$ITD_{hF}(\alpha) \approx \frac{4\pi a}{c} \sin(\alpha) \quad (2.7)$$

**Flächen gleicher interauraler Merkmale** - Die interauralen Laufzeit- und Pegeldifferenzen sind gemäß Definition an einem kugelsymmetrischen Kopf ohne Beachtung der richtungsabhängigen HRTFs auf der Oberfläche eines Hyperboloiden konstant. Die Lage des Hyperboloiden mit dessen Rotationsachse ist an der interauralen Achse orientiert. Eine eindeutige Richtungs-zuweisung ist auf Grund der Mehrdeutigkeit nicht möglich. Diese Mehrdeutigkeiten führen zu dem Begriff „cone-of-confusion“. Eine Unterscheidung der Richtung durch Änderungen der ILDs und ITDs treten bei Berücksichtigung der Richtwirkung des Außenohres auf. Diese Änderungen fallen allerdings gering aus [232] und führen daher nicht zu einer vollständigen Auflösung der Mehrdeutigkeiten. Die Beachtung von monauralen spektralen Merkmalen im Audiosignalen kann ebenfalls zu einer Auflösung des „cone-of-confusion“ beitragen. Hierfür wird eine Trennung der Richtungswahrnehmung in Elevation und Azimut vorgenommen. Mit Hilfe der interauralen Merkmale wird ein horizontaler Schalleinfallswinkel durch Projektion in die Ebene bestimmt. Die Mehrdeutigkeiten reduzieren sich somit auf Vorne-Hinten und Oben-Unten und können durch Bekanntheit der Richtungsabhängigkeit des monauralen Spektrums, eben durch die HRTFs, aufgelöst werden. Das monaurale Spektrum des Signals an den Ohren hängt allerdings auch vom Spektrum des Audiosignals der Quelle ab [114]. Ist dieses dem Gehör nicht bekannt, kommt es zu vermehrten Vorne-Hinten-Vertauschungen.

Kopfbewegungen führen zu einer eindeutigen Änderung der ITD und ILD und können somit sicher zur Auflösung der Vorne-Hinten-Vertauschung beitragen [57].

### 2.2.3 Akustische Distanzmerkmale

Neben der Zuordnung einer Schalleinfallrichtung ist es dem Menschen möglich, die Distanz eines Schallereignisses einzuschätzen. Die an den Trommelfellen anliegenden Schallsignale beinhalten entfernungsabhängige Merkmale, die in gewisser Näherung auf eine bestimmte Distanz eines Hörereignisses schließen lassen. Diese auditiven Distanzmerkmale sind dabei von der Schallquelle selber und von der Umgebung abhängig. Die Distanz zu einem Hörereignis wird als Entfernung zwischen wahrgenommenen Hörereignisort und dem Mittelpunkt der interauralen Achse definiert [64]. Die metrische Distanz ist als die Entfernung vom Mittelpunkt der interauralen Achse bis zur Schallquelle definiert [64]. Eine Entfernung von Null ist hierbei als Im-Kopf-Lokalisation definiert. Die metrische Entfernung einer Schallquelle zum Empfänger muss dabei nicht der wahrgenommenen Distanz des Hörereignisses entsprechen. In der Regel besteht allerdings ein Zusammenhang zwischen Distanz des Schallereignisses und Distanz des Hörereignisses.

In nachhallfreien Umgebungen kann die Distanz zu einem Schallobjekt lediglich über die Schallintensität am Ohr geschätzt werden. Eine genaue Zuordnung einer Distanz ist nur dann möglich, wenn dem Hörer ein Vergleich oder Vorwissen über die Schallintensität vorliegt. Bei unbekanntem Audiosignal konvergiert die wahrgenommene Distanz bei nachhallfreien Bedingungen zu einem Grenzwert [79]. In nachhallbehafteten Umgebungen kann die Distanz auch über nicht schallquellbezogene aber raumbezogene Merkmale eingeschätzt werden. Nachstehende Abschnitte geben einen Überblick über die in der Arbeit relevanten entfernungsabhängigen Merkmale, mithilfe derer das Gehör die Distanz eines Hörereignis einschätzen kann. Es werden Effekte aus physikalischer und perzeptiver Sicht betrachtet. In nachstehenden Abschnitten finden ausschließlich akustische Distanzmerkmale Beachtung.

**Schallintensität und Schalldruck** - Bei zunehmender Distanz zwischen Schallquelle und Empfänger nimmt die Schallintensität ab. Bei Annahme eines Kugelstrahlers und bei einer Schallausbreitung unter Freifeldbedingungen ist diese Abnahme über das Entfernungsgesetz definiert. Die Intensität verteilt sich hierbei gleichmäßig auf einer größer werdenden Kugeloberfläche, welche proportional mit dem Quadrat des Abstandes  $r$  größer wird. Hieraus resultiert eine Intensitätsabnahme mit dem Proportionalitätsfaktor  $1/r^2$ . Die Betrachtung des Schalldrucks als Schallfeldgröße führt zu einer Abnahme des Effektivwert des Schalldrucks proportional zu  $1/r$ . Die wahrgenommene Distanz des Hörereignisses muss allerdings nicht mit der geometrischen Entfernung übereinstimmen. Eine Abnahme des Schalldruckpegels um 6 dB führt nicht zwangsläufig zu einer Verdoppelung der wahrgenommenen Entfernung. Es muss eine deutlich größerer Abnahme vorliegen, um den Eindruck einer verdoppelten Entfernung hervorzurufen [64].

In reflektionsbehafteten Umgebungen muss neben dem Direktschallpegel auch die Abhängigkeit des Nachhallpegels von der Distanz betrachtet werden. Bei Annäherung des Nachhalls durch ein Diffusschallfeld ist definitionsgemäß die Energie unabhängig von der Quellposition im Raum. Die Intensität des Nachhalls ändert sich somit nicht mit einer Distanzveränderung. Bei realen Messungen zeigen sich allerdings Abhängigkeiten. Eine Verdoppelung der Distanz führt hier zu einem Abfall der Nachhallintensität um ca. 1 dB [236].

**Lautheit** - Die Größen Schallintensität und Schalldruck scheinen aufgrund ihres physikalischen Charakters ungenügend zur Beschreibung der wahrgenommenen Distanz zu sein. Es scheint daher sinnvoller, die perzeptiven Größen Lautstärke und Lautheit in Bezug zur Distanzwahrnehmung zu setzen. Die Lautstärke stellt ein psychologisches Konzept zur Beschreibung des Ausmaßes der Empfindung des Schalldruckpegels eines Schallereignisses dar [92]. Der Lautstärkepegel eines Schallereignisses wird in *phon* angegeben. Hierfür sei ebenfalls auf die Kurven gleicher Lautheit oder Isophone verwiesen [233]. Mit Hilfe der Isophone, des Lautstärkepegels lässt sich ein Pegel ermitteln, der bei einer Verdopplung der Lautstärke empfunden wird. Für das Beispiel der Distanzverdoppelung entspricht dies nach den Isophonen für Einzeltöne im Bereich von ca. 400 Hz bis 5000 Hz bei einem Schalldruckpegel von 40 dB bis 100 dB etwa einer Pegelminderung von 10 dB [64].

**Direktschall- zu Nachhallenergie-Verhältnis** - In natürlichen Umgebungen breiten sich Schallwellen nach allen Richtungen aus. Dabei treffen diese auf Objekte im Raum, durch welche sie gebeugt, absorbiert und/oder reflektiert werden. Es entsteht eine komplexe Überlagerung von Schallwellen, welche vom Hörer wahrgenommen werden. Der Schallanteil, welcher direkt von der Schallquelle zum Empfänger gelangt, wird als Direktschall bezeichnet. Der Nachhall ergibt sich aus den an Objekten und Wänden reflektierten Schallwellen und wird zeitverzögert und aus verschiedenen Richtungen vom Empfänger wahrgenommen [176]. Der Nachhall beinhaltet alle Anteile einer Raumimpulsantwort bis auf den Direktschall. Der indirekte Schall enthält Informationen über die Umgebung und die relative Position des Schallereignisses. Das Verhältnis der Intensitäten zwischen direktem und reflektiertem Schall (englisch: Direct-to-Reverberant-Energy-Ratio, D/R oder DRR) spielt eine bedeutende Rolle bei der Distanzwahrnehmung (siehe [79, 133, 236] u. a.).

Mit zunehmender Distanz zwischen Schallquelle und Empfänger nimmt das DRR systematisch ab. Die Intensität des Direktschalles nimmt mit steigender Entfernung gemäß  $1/r^2$  ebenfalls ab. Bei Annahme eines Diffusschallfeldes ist die Intensität des Nachhalls im Diffusfeld unabhängig von der Distanz. Das Verhältnis aus der Intensität des Direktschalls und der Intensität des Nachhalls nimmt unabhängig von der Amplitude des ausgesendeten Schallsignals mit ca.  $1/r^2$  ab. Bei Auswertungen von Messungen mit individuellen BRIRs und verschiedenen Schalleinfallrichtungen konnte eine Abnahme des DRR um ca. 3-4 dB bei Entfernungsverdoppelung festgestellt werden [236]. Bei vergleichbaren Messungen aus einer Richtung und Verwendung

eines Kunstkopfes wurde eine Absenkung des DRR für den untersuchten Raum um 4,7 dB bei Entfernungsverdoppelung festgestellt [133]. Die Unterschiede in den DRR liegen innerhalb der JND für die Wahrnehmung von Halligkeitsunterschieden bei Lautsprecherbeschallung [176, 237] und bei Anwendung der Binauralsynthese [133]. Ein Einfluss der Personalisierung der Binauralaufnahme wurde in verschiedenen Untersuchungen ausgeschlossen [133, 237].

In halligen Umgebungen hängt die Distanzwahrnehmung nicht zwangsläufig von den absoluten Intensitäten des Direktschalls und des Nachhalls ab. Vielmehr ist das Verhältnis zwischen den beiden, also das DRR, entscheidend. Befinden sich zwei Schallquellen in einem reflektionsbehafteten Raum, die an den Ohren eines Hörers den gleichen Schalldruckpegel hervorrufen, ist der Schallintensitätspegel des Nachhalls bei der näher gelegenen Schallquelle relativ geringer gegenüber dem Schallintensitätspegel des Direktschalls bei der weiter entfernt liegenden Quelle. Dem Hörer ist es so möglich zu erkennen, dass es sich bei der entfernteren Quelle um eine Quelle handelt, die lauter spielt und weiter entfernt ist, obwohl der Pegel des Direktschalls am Ohr mit der näher gelegenen Quelle identisch sind.

**Anfangszeitlücke** - Die Laufzeitdifferenz zwischen dem Eintreffen des Direktschalls und der ersten Reflexion am Empfänger wird als Anfangszeitlücke (englisch: Initial Time Delay Gap, ITDG) bezeichnet [58]. Sie entspricht unter Beachtung der Schallgeschwindigkeit der Differenz aus dem Laufweg der Reflexion und des Laufweges des Direktschalls. Die Größenordnung der ITDG ist abhängig von der Distanz zwischen Empfänger und Schallquelle im Raum. Bei zunehmender Distanz wird die ITDG kleiner, da die relative Zeitdifferenz zwischen dem Durchlaufen des Direktschallweges und des Weges der ersten Reflexion abnimmt. Im Gegensatz dazu ist die ITDG groß, wenn sich die Schallquelle nahe am Empfänger befindet. Der Weg des Direktschalls ist dann kurz gegenüber dem Weg der ersten Reflexion. "Eine Ausnahme, bei der die Anfangszeitlücke entfernungsunabhängig ist, ist ggf. bei einer speziellen Anordnung von Hörer und Quelle im Raum gegeben. Liegt die interaurale Achse parallel zur Wand im Rücken des Hörers und befindet sich die Quelle frontal zum Hörer, so ist es möglich, dass die erste Reflexion von der Rückwand kommt und entfernungsunabhängig ist." [33] Die ITDG steht in einem Zusammenhang mit der wahrgenommenen akustischen Intimität [58]. Dieses Qualitätsmerkmal beschreibt die Raumwahrnehmung und ist eng mit dem Gefühl der Nähe zur Quelle verbunden [164].

**Spektraler Inhalt** - Änderungen in der spektralen Zusammensetzung des Schallsignals können zu einem Unterschied in der Distanzwahrnehmung führen. Zum einen werden mit größer werdenden Distanz zwischen einer Schallquelle und dem Hörer hochfrequente Signalanteile durch die frequenz- und entfernungsabhängige Luftabsorption gedämpft [64]. Bei Entfernung unter 10 m können allerdings keine perzeptiv relevanten Einflüsse festgestellt werden [72]. In Messungen von BRIRs und einer Entfernungsänderung von 30 cm auf ca. 14 m zeigt sich eine Abnahme von lediglich 1,6 dB in den hohen Frequenzen [236].



Zum anderen wird bei einem nahen Hörereignis die Klangfarbe als dunkler gegenüber einem weiter entfernten Hörereignis wahrgenommen. Dieses wird als "darkening in timbre of very near sound sources" bezeichnet [73]. Dieser Lautheitsanstiegs lässt sich durch die Kurven gleicher Lautheit begründen. Bei einer näher kommenden Schallquelle steigt die Lautheit der tiefen Frequenzen stärker an als bei den hohen Frequenzen [72, 73]. Eine sich ergänzende Erklärung ist die relative stärkere Abnahme der Anteile hoher Frequenzen bei sehr nahen Quellen ( $\ll 1$  m) gegenüber weit entfernten Quellen (10 m) bei Vorhandensein eines Kugelkopf-Modells [73].

**Monaurale und binaurale Distanzmerkmale** - Die interaurale Zeitdifferenz (ITD) wird als vorherrschendes Merkmal für die horizontale Richtungslokalisierung von Schallquellen angesehen ([64, 111] u. a.). Darüber hinaus ist diese auch abhängig von der Distanz. Bei geringerer Distanz zum Kopf nimmt die absolute Zeitdifferenz leicht zu. Durch die Geometrie eines realen Kopfes tritt dieser Effekt bei gemessenen HRTFs viel stärker auf als bei HRTFs, die auf einem symmetrischen Kugelkopf-Modell basieren [73]. Die ITD ändert sich über die Entfernung nur in einem geringen Maß, da sie von der relativ kleinen arithmetischen Entfernungsdifferenz von der Quelle zum ipsilateralen und zum kontralateralen Ohr abhängt.

Die interaurale Pegeldifferenz lässt sich gemäß Abschnitt 2.2.2 über das Verhältnis der HRTFs von kontra- und ipsilateraler Seite berechnen. Das Betragsspektrum der HRTF selber weist je nach Einfallrichtung eine Abhängigkeit von der Distanz auf. Brungart und Rabinowitz [73] untersuchten diesen Effekt sowohl an einem Kunstkopf als auch an einer Simulation eines Kugelkopfmodells. Die HRTFs des kontralateralen Ohres weisen bei sich verringernder Distanz aufgrund der größer werdenden Abschattung ein sich verringerndes Betragsspektrum mit verstärktem Tiefpassverhalten auf. Die HRTF des der Quelle zugewandten Ohres zeigt dahingegen eine Anhebung des Betragsspektrums mit einer zusätzlichen Anhebung von hohen Frequenzen bei sich verringernden Distanzen. Die sich daraus ergebenden ILDs sind bei näher gelegenen Quellen größer als bei weiter entfernt liegenden Quellen.

Generell nehmen die Zeit- und Pegeldifferenzen mit steigender seitlichen Auslenkung der Schallquelle zu und mit zunehmender Distanz ab. Die binauralen Merkmale stellen robuste Distanzmerkmale im nahen Bereich dar. Sie sind nicht von der Intensität des Schalls, sondern nur von der Quellposition abhängig [194]. Das Schallsignal der Schallquelle hat auf die binauralen Merkmale lediglich einen Einfluss aufgrund seiner Frequenzzusammensetzung.

In der Literatur werden verschiedene maximale Distanzen angegeben, in denen ILDs und ITDs als Distanzmerkmale vom Gehör ausgewertet werden können. Diese umfassen einen Bereich von ca. 1 m bis 2 m vom Kopfmittelpunkt in denen systematische Entfernungsabhängigkeiten aufgezeigt werden (siehe dazu [72, 73, 96, 127, 194]). Bei größeren Distanzen liefern die binauralen Merkmale fast ausschließlich räumliche Richtungsinformationen.

## 2.3 Signalgenerierung

Die (Re-)Synthese einer akustischen Szene basiert auf der Erzeugung geeigneter Ohrsignale für den Hörer. Hierfür werden akustische Übertragungsfunktionen zwischen den einzelnen Quellen in einem Raum und den Empfängerpositionen (linkes und rechtes Ohr) benötigt. Diese Filterfunktionen werden anschließend für das linke und rechte Ohr mit dem gewünschten Audiomaterial gefaltet und über Kopfhörer wiedergegeben. Der aufgesetzte Kopfhörer stellt ein zusätzliches Übertragungssystem dar. Seine akustischen Eigenschaften müssen kompensiert werden, wenn keine Veränderung der wiederzugebenden Ohrsignale gewünscht wird. Abbildung 2.10 verdeutlicht schematisch diese Zusammenhänge. Für eine dynamische Binauralsynthese ist die Generierung der Übertragungsfunktionen für jede sich ergebende Quelle-Empfänger-Positionsrelation notwendig.

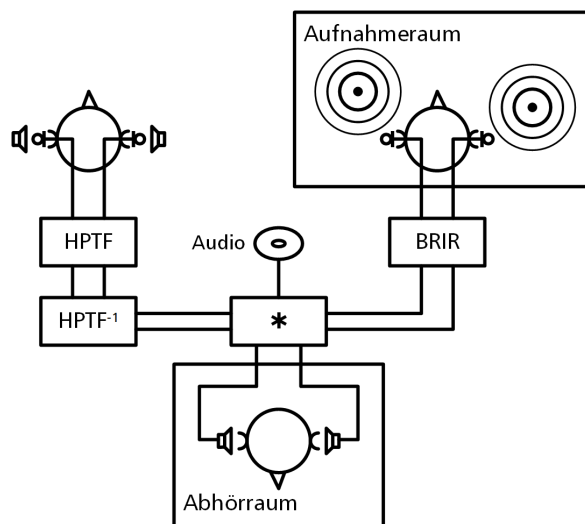


Abbildung 2.10: Schematische Darstellung eines Binauralsynthesystems für Kopfhörer.

### 2.3.1 Mikrofonierung

Durch die Anwendung verschiedener Methoden der räumlichen Audioaufnahme lassen sich für eine Binauralsynthese geeignete Signale erzeugen. Eine wesentliche Voraussetzung ist die Abbildung von richtungs- und frequenzabhängigen Intensitäts- und Laufzeitunterschieden. Diese Gruppe der Mikrofonierung lässt sich als Trennkörperstereofonie bezeichnen. Hierbei werden die Intensitätsunterschiede zwischen den Stereokanälen nicht durch die Richtcharakteristik der Mikrofone, sondern vor allem durch Einbringen eines akustischen Trennkörpers generiert. Das physikalische Verhalten von Kopf, Torso und Ohrmuschel mit den darin enthaltenen binauralen und monauralen Merkmalen soll so angenähert werden.

Die Verfahren lassen sich gemäß der mit ihnen möglichen Übereinstimmung der darstellbaren auditorischen Merkmale gegenüber den Ohrsignalen einer realen Person einteilen. Stereo-Aufnahmeverfahren ohne Trennkörper, wie z.B. das Blumlein-Verfahren oder ORTF-Verfahren

bilden einfache Laufzeit- oder/und Intensitätsunterschiede aufgrund der Mikrofonrichtcharakteristika und Mikrofonanordnung ab. Die Übereinstimmung vor allem der spektralen Merkmale mit denen eines realen Hörers ist als gering einzuschätzen. Trennkörperverfahren, wie z.B. dem OSS-Verfahren (Jecklin-Scheibe) [120] oder Kugelflächenmikrofonen [208] erreichen einen höheren Grad an Übereinstimmung. Deutlich größere Übereinstimmungen können durch die Verwendung von Kunstköpfen oder durch individuelle Messungen mithilfe von Mikrofonen in Trommelfellnähe oder am Eingang zum Gehörgang erreicht werden [153, 154].

Die Mikrofonierung lässt sich zum einen dazu verwenden, die bspw. in einem Raum befindlichen Schallquellen direkt aufzunehmen. Sollen zum anderen die Audiosignale der Schallquellen austauschbar werden und eine Simulation bzw. (Re-)Synthese der akustischen Umgebung realisiert werden, müssen die Übertragungsfunktionen der Schallquellen im Raum durch das Mikrofonsystem abgebildet bzw. gemessen werden.

Eine Übersicht über gängige binaurale Verfahren liefert Hammershøi [103] und Nicol [159]. Mithilfe von Mikrofonarraymessungen [160] oder Raumsimulationstechniken können ebenfalls richtungsabhängige Raumimpulsantworten erzeugt werden. Eine anschließende Faltung mit dazugehörigen HRTFs liefert ebenfalls Ohrsignale für eine Wiedergabe.

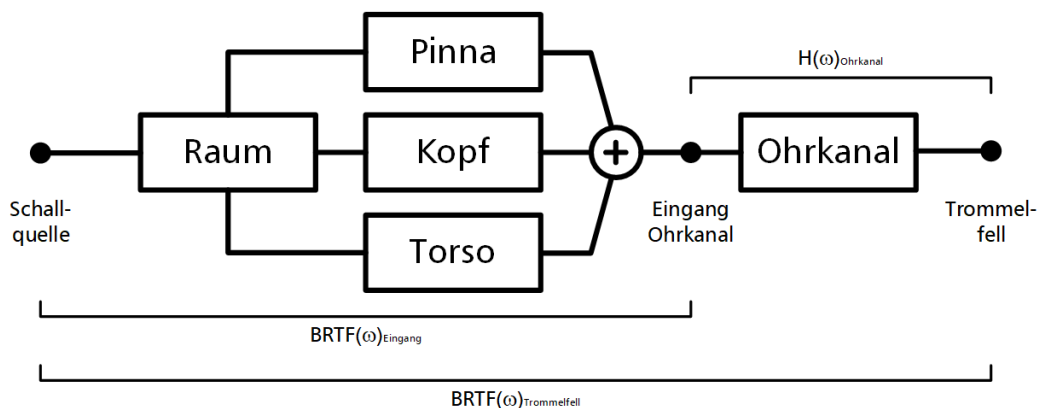


Abbildung 2.11: Schematische Darstellung der Messpunkte und Einflüsse auf eine binaurale Raumübertragungsfunktion; Abbildung angelehnt an Vorländer [220].

In dem theoretischen Fall, dass eine Messung der Übertragungsfunktionen im Freifeld und bei Schalleinfall einer ebenen Welle stattfindet, wird diese Übertragungsfunktion als Außenohrübertragungsfunktion bzw. head-related transfer function (HRTF) bezeichnet. Diese strikte Bedingung an eine HRTF kann in der Praxis nicht gehalten werden und wird im Allgemeinen auch nicht gefordert. Der Begriff HRTF wird verwendet, wenn die akustischen Übertragungseigenschaften des Ohres, Kopf und Torso in einer nachhallarmen Umgebung ermittelt werden. Hierbei ist oft auch die akustische Übertragungscharakteristik des verwendeten Schallwandlers beinhaltet, sofern diese nicht explizit entzerrt wird.

Die HRTFs werden idealerweise durch das Verhältnis von Schalldruck an einem Messpunkt bspw. am Trommelfell oder am geblockten Gehörgang, und eines Messpunktes am Mittelpunkt

des Kopfes in Abwesenheit des Hörers definiert. Der Schalldruckpegelverlauf am Messpunkt lässt sich im Zeitbereich durch die kopfbezogene Impulsantwort (engl.: Head-Related Impulse Response, HRIR) beschreiben. Der Gehörgang wird als weitestgehend richtungsunabhängiger Filter angenommen [220, 229]. Er kann daher aus der Messung des richtungsabhängigen Einflusses der Pinna, Kopf und Torso herausgenommen werden. Abbildung 2.11 verdeutlicht schematisch die Komponenten dieser Übertragungsfunktionen.

Werden die Übertragungseigenschaften in einem akustisch aktiven Raum ermittelt, wird der Begriff der binauralen Raumübertragungsfunktion bzw. binaural room transfer function (BRTF) verwendet. Im Zeitbereich ergibt sich die binaural room impulse response (BRIR). Der Begriff der BRTF ist in seiner Benennung ungünstig, da er eine Verbindung zwischen linken und rechten Ohr impliziert, welche in erster Linie so nicht nötig ist. So wie die HRTF im Freifeld für linkes und rechtes Ohr definiert ist, sollte eine Übertragungsfunktion in einem Raum ebenso für beide Ohren separat definiert sein. Es wäre damit eine Benennung in Form einer room- and head-related transfer function (RHRTF) sinnhaft. In der Praxis haben sich allerdings die Begriffe HRTF/HRIR und BRTF/BRIR durchgesetzt und werden auch in dieser Arbeit verwendet.

Eine gemessene BRIR enthält neben den richtungsabhängigen Eigenschaften des Außenohres, Kopf und Torsos auch die richtungs- und distanzabhängigen Eigenschaften des Raumes. Diese sind vor allem durch Reflexions- und Absorptionseigenschaften der Raumgeometrie und Raummaterialien bestimmt. Abbildung 2.12 zeigt beispielhaft eine mittels Kunstkopf (ohne Gehörgang) gemessene BRIR aus 2 m Abstand in einem realen Raum.

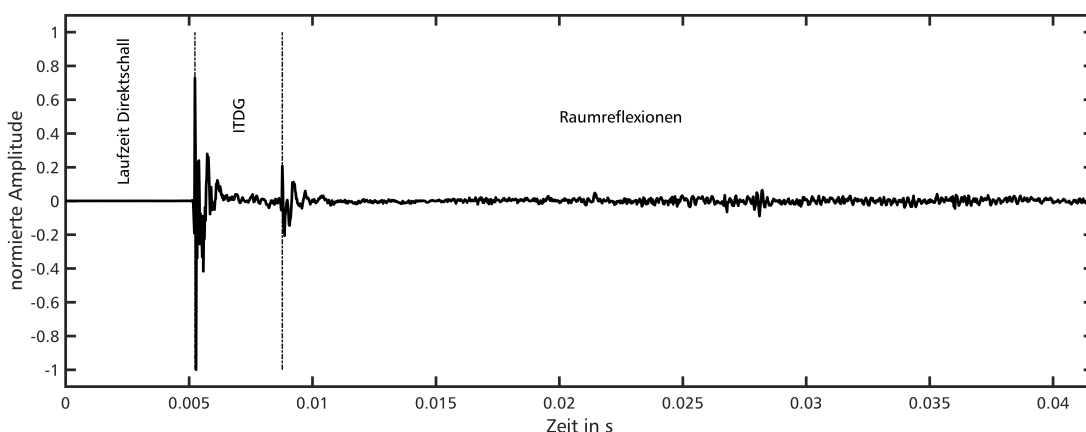


Abbildung 2.12: Gemessene binaurale Raumimpulsantwort für das linke Ohr eines Kunstkopfes im Abstand von 2 m zur Quelle.

Der Abstand zwischen Messposition und Schallquellposition führt zu einer Signallaufzeit des direkten und reflektierten Schalls. Die erste Reflexion trifft nach dem Direktschall ein und bildet die Initialzeitlücke (engl. initial time delay gap; ITDG). Die ITDG stellt einen wichtigen Parameter zur Wahrnehmung von Distanz in reflexionsbehafteten Umgebungen dar (siehe

dazu Kapitel 2.2.3). Das Verhältnis der Energien des Direktschalls und aller Reflexionen wird durch das DRR abgebildet. Das DRR ist ebenfalls ein wichtiger akustischer Parameter der Raum- und Distanzwahrnehmung. Mit zunehmendem zeitlichen Verlauf der BRIR sinkt die Nachhallenergie allmählich bis auf ein Rauschniveau ab. Dieses Abklingen wird durch die Energieabklingkurve (engl. energy decay curve; EDC) beschrieben. Aus ihr lassen sich weitere akustische Parameter, wie bspw. Nachhallzeit und Deutlichkeitsmaße berechnen.

Diese raumakustischen Parameter einer aufgenommenen BRIR lassen sich nachträglich verändern und anpassen. Es lassen sich dadurch bspw. die wahrgenommene Distanz durch Veränderung von akustischen Distanzmerkmalen oder die Anpassung der wahrgenommenen Halligkeit der Synthese an den Abhörraum realisieren. Weiterhin ist eine gezielte Veränderung der zeitlichen Reflexionsmuster in der BRIR möglich. Derartige Ansätze zielen darauf ab, eine perzeptive Passung zwischen der Synthese und dem Abhörraum zu erzeugen. Hierzu werden Methoden vorgestellt, die einen Einfluss auf Qualitätsmerkmale, wie wahrgenommenen Raumeindruck, Distanz und Externalität haben. Die Methoden umfassen verschiedene Interpolationsverfahren, die Anpassungen der Initialzeitlücke, die Anpassung der EDC und des DRR. Die Beschreibung und Evaluierung ausgewählter Methoden erfolgt in Kapitel 5.

### 2.3.2 Abbildung der Übertragungscharakteristika

**Impulsantwort** - Ein zeit-invariantes und lineares System lässt sich durch eine Übertragungsfunktion  $h(t)$  beschreiben. Das System kann zusätzlich einen Rauschanteil  $n(t)$  beinhalten. Das Rauschsignal wird als normal verteilt und als unabhängig vom Eingangssignal betrachtet. Das System lässt sich somit über nachstehende Formel beschreiben:

$$y(t) = n(t) + x(t) * h(t) \quad (2.8)$$

Zur Bestimmung der Übertragungsfunktion wird ein bekanntes Signal  $x(t)$  als Eingang in das System gegeben. Am Ausgang des Systems wird das veränderte Signal  $y(t)$  gemessen. Als Messsignal kann bspw. ein Sinus-Sweep verwendet werden. Dieses Signal zeichnet sich durch seine Breitbandigkeit und Determiniertheit aus. Zur Verbesserung des Signal-Rausch-Abstandes (eng.: signal-to-noise-ratio, SNR) wird die Messung mehrfach durchgeführt und das Ausgangssignal des Systems im Zeitbereich gemittelt. Das gemittelte Signal lässt sich als  $\hat{y}(t)$  beschreiben. Unter der Annahme, dass sich das normal-verteilte systeminherente Rauschen durch die Mehrfachmessung minimiert, lässt sich die Impulsantwort des Systems durch seine Fourier-Transformierte und durch nachstehende Entfaltung ermitteln:

$$h(t) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\hat{y}(t))}{\mathcal{F}(x(t))}\right) \quad (2.9)$$

Reale Systeme beinhalten neben einem linearen Teil auch einen nicht linearen Anteil, welcher ebenfalls die Übertragungsfunktion beeinflusst. Dieser Anteil lässt sich in der Regel schwer

erfassen. Im Falle einer Messung einer Impulsantwort eines akustischen Raumes, welcher Reflexionen der Schallwellen an den Begrenzungsflächen beinhaltet und welcher durch einen Lautsprecher angeregt wird, lassen sich vereinfachende Annahmen formulieren. Die Nichtlinearitäten treten zu Beginn des Messsignals auf und werden vor allem durch den Lautsprecher verursacht. Die Ausbreitung der Schallwellen im Raum folgt dahingegen wieder einem linearen Prozess, welcher die Reflexionen, Echos und Nachhall des Raumes beinhaltet [88]. Besonders zu beachten ist, dass die Dauer des sich wiederholenden Eingangssignals nicht kürzer als die Impulsantwort des Systems ist. Ist dies der Fall treten Aliasingfehler in der berechneten Impulsantwort auf. Verhindern lässt sich dies, indem bei der Erstellung bspw. eines Sinus-Sweeps als Anregungssignal entsprechend viele Nullen nach jedem Sweep-Signal angehängt werden.

In der vorliegenden Arbeit werden raumakustische Messungen mithilfe eines Kunstkopfes oder durch individuelle Messungen in den Ohren des Hörers durchgeführt. Für die Bestimmung der Übertragungscharakteristik des Systems Raum-Kopf wird ein logarithmischer Sinus-Sweep mit einer gegebenen unteren und oberen Grenzfrequenz genutzt.

**Invertierung einer Impulsantwort** - Für eine Binauralsynthese über Kopfhörer muss das System des aufgesetzten Kopfhörers entzerrt bzw. kompensiert werden (siehe Abschnitt 2.4.2). Dazu ist eine Filterfunktion  $h_{komp}$  gesucht, die nachstehende Bedingung theoretisch erfüllt:

$$h_{entzerrt} = h(n) * h_{komp} = \delta(n) \quad (2.10)$$

mit  $\delta(n)$  = Impulse im abgetasteten Zeitbereich,  $h(n)$  = Impulsantwort des Kopfhörersystems.

Eine direkte Invertierung wird durch nachstehende Formel im Frequenzbereich beschrieben:

$$H_{komp}(\omega) = \frac{1}{H(\omega)} \quad (2.11)$$

mit  $H(\omega)$  als Übertragungsfunktion des Kopfhörersystems. Diese direkte Invertierung führt vor allem an den Grenzen des Frequenzspektrums zu ungewollten übermäßigen Anhebungen des Amplitudenfrequenzganges. Ein weiteres Problem besteht darin, dass der so invertierte Filter immer nur für eine Aufsetzposition des Kopfhörers gültig ist. Eine Repositionierung des Kopfhörers führt in der Regel immer zu einer leicht veränderten Aufsetzposition und somit auch zu einem Verschieben von Frequenzeinbrüchen und Frequenzüberhöhungen innerhalb der Kopfhörerübertragungsfunktion. Bei der Anwendung des oben beschriebenen Kompensationsfilters würden somit unter Umständen nicht zu einer Kompensation, sondern zu einer Verstärkung schmalbandiger Frequenzbereiche führen. Dieser Effekt wirkt sich vor allem im hochfrequenten Bereich deutlich aus.

Zur Minderung der beiden genannten Effekte wird eine Kompensationsmethode verwendet, bei der die Kompensationsgenauigkeit im hochfrequenten Bereich verringert wird. Die Fre-

quenzeinbrüche werden damit zwar weniger gut kompensiert, allerdings kommt es auch zu deutlich weniger Artefakten. Weiterhin wird nicht der gesamte Frequenzbereich kompensiert. Es wird ein Bandpass als Zielfunktion definiert, der den zu kompensierenden Frequenzbereich an den Rändern einschränkt.

Die in dieser Arbeit verwendete Methode zur Kopfhörerentzerrung beruht auf Arbeiten von Schärer [187, 188], in denen verschiedene Ansätze zur Invertierung von Kopfhörerübertragungsfunktionen ausführlich vorgestellt und verglichen werden. Darauf aufbauend wird die Methode nach dem Kriterium des kleinsten quadratischen Fehlers zur Invertierung ausgewählt. Nachstehend wird die Funktionsweise dieser Methode genauer ausgeführt.

Bei der Invertierung nach dem kleinsten quadratischen Fehler wird die zu erstellende Funktion an die vorgegebene Zielfunktion mit einer Minimierung des Unterschiedes angenähert. Der Entzerrfilter  $h_{entzerrt}(n)$  soll somit nachstehende Formulierung erfüllen:

$$\sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} (d(n-m) - h_{entzerrt}(n))^2 \quad (2.12)$$

mit  $h_{entzerrt} = h(n) * h_{komp}(n)$  und Zielfunktion  $d(n-m)$ .

Durch die Anwendung der Matrixmultiplikation und der Überführung in Faltungsmatrizen lässt sich Gleichung 2.12 darstellen durch:

$$e^T e = (d_m - H \cdot h_{komp})^T (d_m - H \cdot h_c) \quad (2.13)$$

Der Filter  $h_{komp}$  lässt sich nach Umformung auf folgende Weise berechnen:

$$h_{komp} = (H^T \cdot H) / (H^T \cdot d_m) \quad (2.14)$$

Zur Realisierung der frequenzabhängigen Kompensation (Regularisierung) wird der Kompensationsfilter  $h_{komp}$  mit einem Filter  $b(n)$  gefaltet. Dieser Filter selber besitzt eine Hochpasscharakteristik. Im Durchlassbereich des Filters wird die Entzerrung verringert. Die Berechnung der Entzerrung mit Regularisierung wird analog zur Berechnung ohne Regularisierung mittels Matrixmultiplikation durchgeführt:

$$h_{komp} = (H^T \cdot H + G \cdot B^T \cdot B) / (H^T \cdot d_m) \quad (2.15)$$

hierbei entspricht  $B$  dem Regularisierungshochpass als Faltungsmatrix (Toeplitz-Form) und  $G$  einem frei wählbaren Dämpfungsfaktor.

## 2.4 Binaurale Wiedergabe

Die gehörrichtige Wiedergabe einer drei-dimensionalen akustischen Szene lässt sich durch Kopfhörersysteme ([82, 230], u. a.) oder durch Lautsprecher Systeme ([52, 137], u. a.) realisieren. In beiden Systemen kommen Verfahren zur Anwendung, die es ermöglichen, die Audiosignale so aufzubereiten, dass am linken bzw. rechten Ohr nur die zu diesem Ohr gehörenden Signale ankommen [220]. Es muss dabei großer Wert auf die Kompensation der Übertragungseigenschaften der Systeme gelegt werden. Die binauralen Signale sollen nicht durch die Eigenschaften der Wiedergabesysteme verfälscht werden. Weiterhin stellt die Personalisierung der Systeme eine große Herausforderung für eine plausible Synthese der akustischen Szene dar [155, 191]. Neben diesen und weiteren technischen Einflussfaktoren stellen vor allem wahrnehmungsbezogene Einflussfaktoren eine Herausforderung zur Erreichung eines plausiblen räumlichen Höreindrucks dar. Eine Benennung derartiger Herausforderungen und Einflüsse ist der Literatur zu entnehmen [2, 3, 19].

Die Position einer Schallquelle im Raum lässt sich durch ihren Abstand zum Mittelpunkt des Kopfes des Zuhörers (Koordinatenursprung), des Azimutwinkel und des Elevationswinkels beschreiben. An den Trommelfellen des Zuhörers liegt das zeitveränderliche und richtungsabhängige Schalldrucksignal an, welches sich durch die HRTFs beschreiben lässt. Können keine Freifeldbedingungen angenommen werden, kommen zusätzlich Schallanteile an den Ohren an, die durch Reflexionen an den Begrenzungsflächen des Raumes verursacht werden. Die Gesamtheit der HRTFs und der Raumakustik kann durch binaurale Raumimpulsantworten beschrieben werden.

Nachstehende Ausführungen zeigen Wiedergabesysteme auf, die in unterschiedlichem Maße gehörrichtige Signale an die Trommelfelle eines Hörers liefern. Es wird dabei näher auf die binaurale Lautsprecherwiedergabe und binaurale Kopfhörerwiedergabe eingegangen. Weiterführende Übersichten über räumliche Audioaufnahme- und Audiowiedergabesysteme finden sich u. a. in Zhang et al. [239] und Spors et al. [199].

### 2.4.1 Binaurale Lautsprecherwiedergabe

Bei der binauralen Auralisation über z. B. zwei Lautsprecher ist das Ziel, am linken bzw. rechten Ohr des Hörers nur die Signale, die für das linke bzw. das rechte Ohr bestimmt sind, ankommen zu lassen. Die von den beiden Lautsprechern abgestrahlten Signale erreichen aber jeweils beide Ohren des Hörers. Es findet somit ein Übersprechen der Signale des linken bzw. rechten Ohres auf das jeweils gegenüberliegende Ohr statt. Weiterhin erreichen beide Ohren zusätzliche Schallsignale, die durch Reflexionen an den Wänden des Wiedergaberaumes entstehen. Es kommt zu Verfärbungen des Klanges, sowie zu einem Zusammenbruch der im binauralen Signal enthaltenen räumlichen Signalmerkmale [197]. Das Übersprechen des linken Lautsprechers auf das rechte Ohr und das Übersprechen des rechten Lautsprechers auf das linke Ohr muss kompensiert werden [52, 137, 220].



Es müssen Filterfunktionen erstellt werden, die die Signalpfade des Übersprechens kompensieren. Der rechte Lautsprecher muss ein zeitlich versetztes, phaseninvertiertes und durch die Abschattung des Kopfes eingefärbtes Signal abstrahlen, welches dem Signal entspricht, das vom linken Lautsprecher am rechten Ohr des Hörers eintrifft. Für den rechten Lautsprecher gilt dies in umgekehrter Weise. Dieses Kompensationssignal verursacht nun selber wieder ein Übersprechen auf das gegenüberliegende Ohr und muss durch den gegenüberliegenden Lautsprecher auf gleiche Weise kompensiert werden. Der Prozess der Übersprechkompensation ist somit ein iterativer Prozess, der aber aufgrund der Pegelabnahme der Übersprechsignale, welcher durch die Abschattung des Kopfes verursacht wird, abgebrochen werden kann. Eine mathematische Beschreibung zur Erzeugung der Kompensationssignale findet sich in der Literatur [153, 155, 220]. Weitere Ansätze zur binauralen Wiedergabe in 2-kanal Stereoanordnung [38] sowie zur Kompensation nicht optimal angeordneter Lautsprecher [41] sind der Literatur zu entnehmen.

#### 2.4.2 Binaurale Kopfhörerwiedergabe

Bei der binauralen Wiedergabe über Kopfhörer wird das vom Kopfhörer abzugebende Signal  $s(t)$  durch den Frequenzgang des Kopfhörers, durch die Filtereigenschaften der Pinna und durch Resonanzen im Gehörgang des linken und rechten Ohres verändert. Je nach Bauform und Typ des Kopfhörers ergeben sich zum Teil sehr unterschiedliche Frequenzgänge [110]. An den Trommelfellen liegt somit ein verändertes Signal  $s_{KT|l,r}(t)$  an. Die Kopfhörerübertragungsfunktion (HPTF)  $H_{K|l,r}(w)$  lässt sich durch den Quotienten der Fouriertransformierten von Kopfhörersignal  $s(t)$  und Signal am Trommelfell  $s_{KT|l,r}(t)$  beschreiben. Die richtungsabhängigen HRTFs bzw. BRIRs müssen für eine gehörrichtige Wiedergabe des Audiosignals mit der Inversen der HPTF bewertet werden. Die Einflüsse des Kopfhörers werden somit weitestgehend kompensiert und es ergeben sich die kompensierten Übertragungsfunktionen. Die kompensierte Übertragungsfunktion  $H_{C|l,r}(w, r, \theta, \varphi)$  ergibt sich aus dem Quotienten der HRTF bzw. BRIR und der Kopfhörerübertragungsfunktion. Die Inverse der HPTF kann bspw. durch eine Methode der kleinsten Fehlerquadrate mit einer Minimalphaseninvertierung realisiert werden [188] (siehe 2.3.2). Die entzerrten Kopfhörersignale  $s_{K|l,r}(t)$  ergeben sich nun aus der Filterung des Quellsignals  $s(t)$  mit den kompensierten Übertragungsfunktionen  $H_{C|l,r}$ .

Für eine korrekte Entzerrung müssen zusätzlich die Aufsetzposition und die Art des verwendeten Kopfhörers berücksichtigt werden. Die HPTFs können entweder an einem Kunstkopf oder am Hörer individuell gemessen werden. Bei individueller Messung können In-Ohr-Mikrofone verwendet werden. Hierbei kann sowohl am geblockten Ohrkanal oder aber auch mittels Sondenmikrofonen in der Nähe des Trommelfells gemessen werden. Werden die Messungen der BRIRs und der HPTFs in einem Durchgang durchgeführt, empfiehlt es sich die Mikrofone zwischen den Messungen nicht neu zu platzieren. Eine Repositionierung führt in der Regel zu unerwünschten Veränderungen des Frequenzganges der Kopfhörerentzerrung. Bei Verwendung

der selben Mikrofone zur Messung der HRTFs bzw. BRIRs und der HPTFs wird der Frequenzgang des verwendeten Mikrofones durch die HPTF-Invertierung ebenfalls mit entzerrt. Eine separate Kompensation des Mikrofonfrequenzganges ist somit nicht nötig. Eine weiterführende Übersicht über binaurale Techniken liefert z.B. Roginska [177] und Nicol [159].

**Schalldruckpegel-Lautheits-Divergenz** - Bei der Audiowiedergabe über Kopfhörer (auch ohne Anwendung der Binauralsynthese) ist bekannt, dass im Vergleich zum Hören über Lautsprecher ein höherer Schalldruckpegel am Ohr des Hörers benötigt wird, um die gleiche Lautheitswahrnehmung hervorzurufen. Dieser Effekt wird als Schalldruckpegel-Lautheitsdivergenz (SLD) bezeichnet [209]. In Untersuchungen von Fastl et al. [210] ergeben sich frequenzabhängig erhöhte Schalldruckpegel von im Mittel 5 dB. In einer Auflistung von Untersuchungen in Theile [207] werden SLD-Pegel von 3,4 dB bis 12,8 dB berichtet.

Die Ursache für diesen perceptiven Effekt ist bisher nicht eindeutig geklärt. Es wird vermutet, dass dieser Effekt auf der beim Lautstärkevergleich wahrgenommenen Entfernung der Hörereignisse zurückführbar ist [209, 210]. Eine hohe Entfernungsdivergenz der Hörereignisse liegt beim Hören über Kopfhörer (Im-Kopf-Lokalisation bei nicht binauraler Auralisation) und über Lautsprecher im Raum vor. Es wird weiterhin angenommen, dass Schallquellen größerer Entfernung bei Lautsprecherwiedergabe einen geringeren Schalldruck aufweisen können, um eine gleiche Lautstärkewahrnehmung hervorzurufen, als bei der Kopfhörerwiedergabe. Nach Rudmose [182] besitzt eine weiter entfernte Schallquelle eine ausgedehntere auditorische Größe, als eine nahe Schallquelle, welche als wesentlich schmaler wahrgenommen wird. Somit ist der SLD-Effekt nicht zwingend an einen Kopfhörer gebunden, sondern vielmehr an die wahrgenommene Entfernung des Hörereignisses.

Der SLD-Effekt sollte demnach aufgelöst werden können, wenn eine Binauralsynthese der realen Lautsprecher im Raum stattfindet. Die Anforderung an die Synthese ist hierbei eine authentische Resynthese der Lautsprecher, welches eine gleiche Entfernungswahrnehmung beim Hörer hervorruft. Eine Verringerung des SLD-Effektes wurde auch für binaurale Kopfhörerwiedergabe im Diffusfeld nachgewiesen [207].

Die Relevanz des SLD-Effektes für diese Arbeit liegt darin, dass eine authentische Binauralsynthese im Prinzip nur dann möglich ist, wenn der SLD-Effekt aufgelöst werden kann. Ist dies nicht der Fall, scheint damit im auditorische System, selbst bei einer plausiblen Wahrnehmung der Objektpositionen im Raum (und anderer Qualitätsmerkmale), ein Missverhältnis zwischen binaural synthetisierter Szene und erwarteter zu konstruierender Szene vorzuliegen. Eine Beschreibung der kognitiven Vorgänge und eine Abschätzung der Relevanz in Bezug auf Plausibilität/Authentizität bleibt in dieser Arbeit offen, sollte allerdings weiter untersucht werden. Eine Auffälligkeit zur Untersuchung des SLD-Effektes soll aber in nachstehender Zusammenfassung eigener Arbeiten dazu dargestellt werden.

In Untersuchungen von Jürgens und Werner [37, 6] wurden die Kurven gleicher Lautheit (KgL) bei Darbietung von frequenzgruppenbreiten Rauschen über Lautsprecher in einem

Raum (Hörlabor, siehe Abschnitt 2.5; N=18 Testpersonen) und bei Binauralsynthese dieser Lautsprecher über Kopfhörer (N=10 Testpersonen) gemessen. Weiterhin wurde in einem zweiten Experiment die SLD-Pegel direkt gemessen (N=13 Testpersonen).

Abbildung 2.13 verdeutlicht die KgL für frontale Beschallung und für frequenzgruppenbreites Rauschen bei Wiedergabe über Lautsprecher und über Kopfhörer. Die Kopfhörerwiedergabe erfolgt hierbei unter Verwendung von gemessenen Kunstkopf-BRIRs (KEMAR) aus und in dem zu auralisierenden Raum. Das für die KgL-Messung benötigte Referenzsignal ist ein frequenzgruppenbreites Rauschen mit einer Mittenfrequenz von 1 kHz und 60 dB Schalldruckpegel an den Ohren des Kunstkopfes. Die Wiedergabe des Referenzsignals und der Testsignale erfolgt innerhalb des jeweiligen Wiedergabesystems. Referenz- und Testsignale sollten dadurch in vergleichbarer Entfernung wahrgenommen werden. Im oberen Teil der Abbildung sind die Mediane und 25/75 Quartile der KgL bei binauraler Kopfhörerwiedergabe, im mittigen Teil die KgL bei Lautsprecherwiedergabe und im unteren Teil die Differenz der Mediane (SLD) dargestellt.

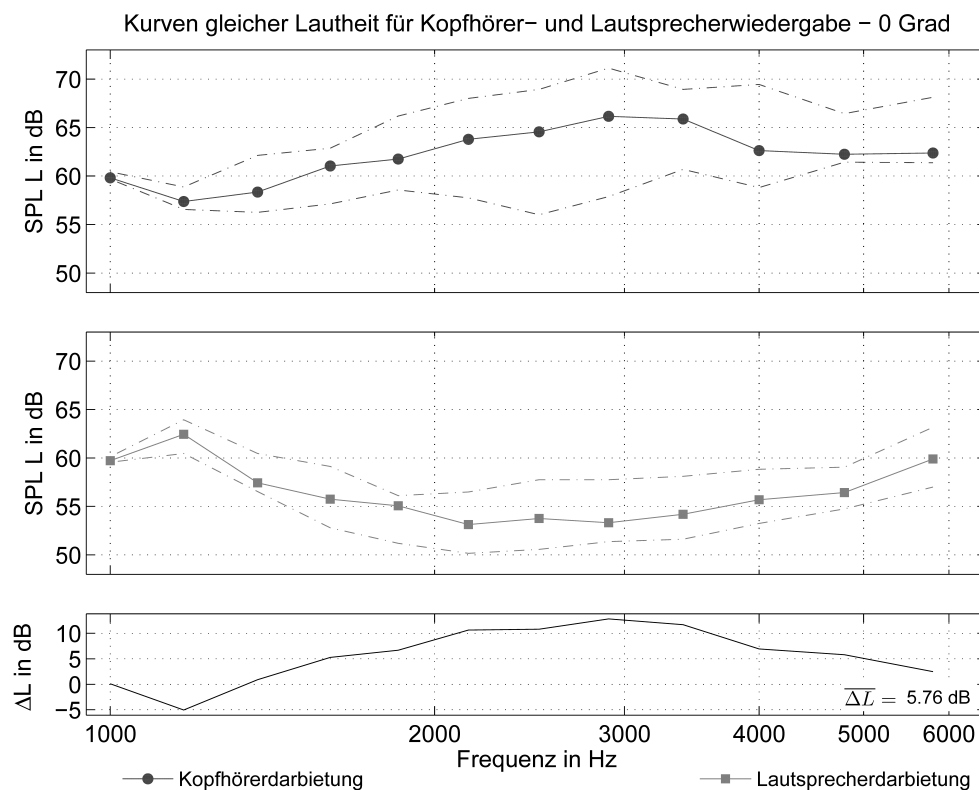


Abbildung 2.13: Kurven gleicher Lautheit für binaurale Kopfhörerdarbietung (oben; 10 Testpersonen) und Lautsprecherwiedergabe (Mitte; 18 Testpersonen) sowie die Differenz der Lautheitskurven (SLD); durchgezogene Linien = Mediane, strichpunktiierte Linien = 25/75 Quartile; Abbildung aus Jürgens [37].

Es zeigen sich signifikante Unterschiede (WilcoxonTest,  $p < 0,05$ ) in den KgLs für die Frequenzgruppen 1175 Hz und von 1600 Hz bis 4800 Hz. Die Abweichungen erreichen zum Teil Werte

von 12 dB. Im Mittel wird eine Differenz von 5,8 dB ermittelt. Die Werte liegen somit in einer vergleichbaren Größenordnung der SLD-Pegel, welche in der Auflistung von Theile angegeben wurden [207].

Abbildung 2.14 zeigt die als gleich-laut wahrgenommenen Schalldruckpegel für frontalen Schalleinfall für frequenzgruppenbreites Rauschen. Die Referenzsignale sind in diesem Experiment dieselben Rauschbänder, wie die Testsignale bei variabler Mittenfrequenz, jedoch mit einem fixen Schalldruckpegel von 60 dB. Der obere Teil zeigt die Schalldruckpegeldifferenzen (SLD) bei einem Lautheitsvergleich mit Wiedergabe der Referenzsignale über den realen Lautsprecher im Raum und Wiedergabe der Testsignale über den Kopfhörer (Binauralsynthese der Lautsprecher). Der mittlere Teil zeigt die SLD-Pegel bei Darbietung der Referenzsignale über den Kopfhörer und Wiedergabe der Testsignale über den Lautsprecher. Der untere Teil zeigt die Differenz der Mediane der beiden oberen SLD-Kurven. Der während des Hörversuchs von der Testperson aufbehaltene offene Kopfhörer (Stax Lambda Pro) wurde durch Messung der Lautsprecher-Kopfhörer-Übertragungsfunktion auf einem Kunstkopf entzerrt.

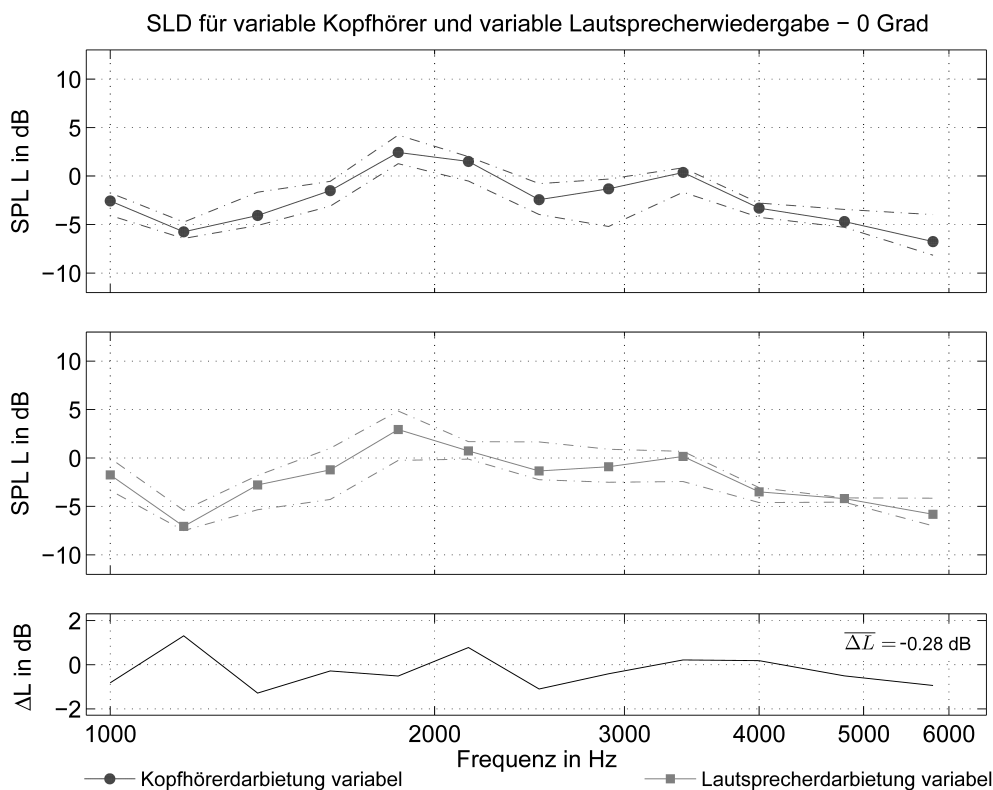


Abbildung 2.14: Schalldruckpegel-Lautheits-Divergenz ( $0^\circ$ ) für schalldruckpegelvariable Kopfhörer (oben)- und Lautsprecherwiedergabe (Mitte) sowie die Differenz der Kurven (13 Testpersonen), durchgezogene Linien = Mediane, strichpunktierte Linien = 25/75 Quartile; Abbildung aus Jürgens [37].

Eine Betrachtung der SLD-Kurven zeigt lediglich für die Bänder von 1000 Hz bis 1370 Hz und für 4000 Hz bis 5800 Hz größere Unterschiede zwischen Lautsprecher- und Kopfhörer-

darbietung. Die Magnitude der SLD beträgt in diesen Bereichen 2 dB bis 5 dB. In einem Frequenzbereich von 1850 Hz bis 3400 Hz treten lediglich geringe Schalldruckpegeldifferenzen auf. Der SLD-Effekt fällt somit deutlich niedriger als in dem ersten Hörversuch bei Messung der KgLs innerhalb eines Wiedergabesystems und als in der Literatur angegeben aus [207]. Ein Vergleich der beiden SLD-Kurven ergibt eine mittlere Abweichung über alle Frequenzbänder von 0,3 dB. Es scheint somit irrelevant in welchem Wiedergabesystem das Referenz- und Testsignal wiedergegeben wird. Eine weitere Auffälligkeit sind die teilweise sehr geringen Interquartilabstände. Dies deutet darauf hin, dass sich die Testpersonen bei den Lautheitsvergleichen relativ einig waren, unabhängig davon auf welchem Wiedergabesystem sich die Referenz bzw. das Testsignal befindet. Ein Test auf Signifikanz der Unterschiede zwischen den SLD-Kurven in Abbildung 2.14 ergibt keine signifikanten Unterschiede (WilcoxonTest,  $p < 0,05$ ) zwischen den beiden Messungen. Eine nach dem Hörversuch durchgeführte Befragung der Testpersonen ergab, dass mehr als die Hälfte der Probanden ( $N=13$ ) während der SLD-Hörversuche das Wiedergabesystem nicht identifizieren konnten. Sie konnten folglich nicht erkennen, ob das Referenz- oder Testsignal aus dem Lautsprecher oder aus dem Kopfhörer kommt.

Die bisher vorgestellten Hörversuche verwenden für die Binauralsynthese Messungen der BRIRs an einem Kunstkopf (KEMAR). Folglich ist die Binauralsynthese nicht auf die einzelne Testperson personalisiert. Dieses kann zu einer verfälschten räumlichen Wahrnehmung bzgl. Lokalisation und Distanzwahrnehmung der Hörereignisse führen, welches den SLD-Effekt ggf. unterstützen kann. Daher wurde ein, den vorangestellten Ergebnissen, unterstützender Hörversuch durchgeführt. Das Versuchsdesign ist hierbei identisch. Es wurde lediglich eine personalisierte Binauralsynthese auf Basis von Messungen individueller BRIRs angewendet. Die Ergebnisse bzgl. ermittelter KgL und SLD-Pegel sind nahezu identisch mit dem vorangegangenen Hörversuch. Lediglich die Interquartilabstände der Bewertungen konnten teils deutlich verringert werden. Weiterführende Informationen dazu sind der Veröffentlichung Jürgens und Werner [6] zu entnehmen.

Die beiden vorgestellten Hörversuche zur Messung des SLD-Effektes werfen Fragen auf, welche bisher nicht zufriedenstellend beantwortet sind. Die Bestimmung der SLD-Pegel bei Abfrage der KgLs innerhalb eines Lautsprecherwiedergabesystems und innerhalb einer Binauralsynthese ergeben SLD-Pegel, welche vergleichbar mit denen in der Literatur sind. Dahingegen wird bei direkter Messung der SLD-Pegel bei gleichzeitiger Darbietung über Lautsprecher und Kopfhörer und an sonst identischer Binauralsynthese kein SLD-Effekt festgestellt. Eine Verwendung von Kunstkopf- oder individuell gemessenen BRIRs scheint keinen Einfluss auf die Ausprägung des SLD-Effektes zu haben. Diese Ergebnisse widersprechen in gewisser Weise den Ergebnissen aus anderen Arbeiten in denen allerdings entweder Diffusfeldbedingungen [207] und/oder keine das Wiedergabesystem übergreifende Testdesigns [219] verwendet wurden.

## 2.5 Aufnahmeraum und Abhörraum

Für die in dieser Arbeit vorgestellten Untersuchungen und damit verbundenen Hörversuche, werden unterschiedliche Räume für die Aufnahme und Wiedergabe verwendet. Die Räume unterscheiden sich vor allem hinsichtlich ihrer geometrischen Dimensionen und raumakustischen Eigenschaften. Nachstehend soll eine Übersicht der verwendeten Räume gegeben werden. In den sich in späteren Kapiteln anschließenden Beschreibung der Hörversuche wird bei der Beschreibung der Räume auf diesen Abschnitt verwiesen.

**Hörlabor (HL)** - Das Hörlabor stellt einen Aufnahme- und Testraum für überwiegend akustische und auditive Experimente an der Technischen Universität Ilmenau dar. Er ist weiterhin eines der Labore des Instituts für Medientechnik zur Untersuchung von multimodalen Wahrnehmungsvorgängen beim Menschen. Seine Fertigstellung erfolgte im Jahr 2003. Die akustischen Eigenschaften des Raumes entsprechen den Forderungen des Standards ITU-R BS.1116-1 [116] gemäß Nachhallverhalten und Aufstellmöglichkeiten eine sechs-kanaligen Lautsprecheraufbaus (5.1 Surround Setup). Weiterhin erfüllt es die Anforderungen an die Raumakustik gemäß Standard EBU Tech. 3276 [216].

Seine Raummaße betragen an den Innenwänden 8,4 m x 7,6 m x 2,8 m mit einem Raumvolumen gemäß Raummaß von 179 m<sup>3</sup> bzw. gemäß der Maße der Innenseite der im Gebäude angebrachten Vorsatzschale von 233 m<sup>3</sup>. Der Unterschied ergibt sich aus Zwischenräumen mit u. a. akustisch wirkenden Geometrien und Auskleidungen mit verschiedenen akustisch wirkenden Materialien. Das Labor besitzt weiterhin eine Trennung zwischen Hörraum und Technik-/Regieraum, welches für die Durchführung verschiedener Audioaufnahmen und Hörversuche von Vorteil ist. Die Belüftung des Raumes erfolgt über eine akustisch optimierte Belüftungsanlage, welche bzgl. Temperaturregelung und Luftdurchsatz einstellbar ist.

Raumakustische Messungen gemäß Standard EN ISO 3382:2000/3.5 [115] sind im fertigen Zustand inkl. diverser Einbauten (bspw. Beamerkasten) und ohne Personen im Raum durchgeführt worden. Es ergibt sich eine über 16 Terzbänder gemittelte Nachhallzeit (T60) von 0,34 s. Das Deutlichkeitsmaß C50 ergibt sich zu 24 dB, das Klarheitsmaß C80 zu 33 dB.

**Seminarraum (SR)** - Der Seminarraum stellt einen, zu einem universell einsetzbaren Lehrraum umfunktionierten Raum der TU Ilmenau dar. Die hauptsächliche Verwendung liegt in der Durchführung von diversen Praktika im Bereich Videoprojektion. Er wird weiterhin als Testraum zur Wahrnehmungsuntersuchung von Videobildschirmen und andere Dinge verwendet. Seine Raumakustik und Raummaße erfüllen keinen gesetzten Standard.

Seine Raummaße betragen an den Innenwänden 10,1 m x 5,0 m x 3,4 m mit einem Raumvolumen gemäß Raummaß von ca. 171,7 m<sup>3</sup>. An einer Seite befindet sich eine weitestgehend durchgängige Fensterfront mit angebrachter textiler Verdunkelungsvorrichtung (lichtdicht schließende Rollos). Alle anderen Wände und des Raumes sind verputztes und tapeziertes Mauerwerk. Die Decke ist eine verputzte Betondecke. Der Boden ist mit einem PVC-Bodenbelag

ausgestattet. Neben eingebauten Rasteranbauleuchten an der Decke und einer Waschbeckenecke befinden sich keine akustisch übermäßig relevanten Ausstattungen in dem Raum.

Raumakustische Messungen gemäß Standard EN ISO 3382:2000/3.5 [115] sind für den Raum vor einem durchgeführten Umbau des Raumes durchgeführt worden. Diese ergeben für die Nachhallzeit (T60) einen gemittelten Wert von 2,0 s.

**Medienlabor (ML)** - Das Medienlabor stellt einen TV-Studiosaal an der TU Ilmenau dar. Seine Raummaße an den Innenwänden betragen 19,5 m x 11,5 m x 5,5 m mit einem Raumvolumen gemäß Raummaß von 1233,4 m<sup>3</sup>. Die Wände und Decke sind mit Lochblechen und akustisch wirkenden Materialien gleichmäßig ausgestattet. Der Boden ist mit einem PVC-Bodenbelag ausgeführt. In dem Raum befindet sich ein aufgeständertes und den Raum umlaufendes Traversensystem. An diesem sind zahlreiche Studioscheinwerfer angebracht. Weiterhin befindet sich in einer Hälfte des Raumes eine der Innenwand vorgesetzte Blauwand in L-Ausführung (Länge ca. 14 m, Höhe 4 m). In der anderen Hälfte des Raumes befindet sich einer der Innenwand vorgesetzte von hinten beleuchtbare Operafolie in L-Ausführung (Länge ca. 14 m, Höhe 4 m).

Raumakustische Messungen sind für den Raum im Zusammenhang mit den dort durchgeführten Hörversuchen durchgeführt worden. Die Messung entspricht nicht in vollem Umfang den Forderungen in Standard EN ISO 3382:2000/3.5 [115]. Die verwendete Schallquelle muss als gerichtete Schallquelle angenommen werden (Lautsprecher Genelec 1030A). Als Messposition des Messmikrofons wird ausschließlich die spätere Abhörposition gewählt. Als Senderposition wird ausschließlich die direkte frontale Richtung und ein Abstand von 2 m zum Empfänger gewählt. Der Lautsprecher ist in seiner Hauptabstrahlrichtung auf das Messmikrofon ausgerichtet. Es ergibt sich für die Nachhallzeit (T60) ein gemittelter Wert von 0,7 s. Das Deutlichkeitsmaß C50 ergibt sich zu 15 dB, das Klarheitsmaß C80 zu 17 dB.

**LeonardoDaVinci-Raum (LdV)** - Der LeonardoDaVinci-Raum stellt einen Vorlesungssaal der TU Ilmenau dar. Baulich ist er mit einer Schrägstellung des Zuhörerbereichs ausgeführt. Seine Raummaße an den Innenwänden betragen (ohne Beachtung der Schräge) 10,4 m x 11,9 m x 8 m mit einem Raumvolumen gemäß Raummaß von ca. 980 m<sup>3</sup> (mit Abzug der Schräge).

Raumakustische Messungen sind für den Raum im Zusammenhang mit den dort durchgeführten Hörversuchen durchgeführt worden. Die Messung entspricht nicht in vollem Umfang den Forderungen in Standard EN ISO 3382:2000/3.5 [115]. Die verwendete Schallquelle muss als gerichtete Schallquelle angenommen werden (Lautsprecher Geithain Mo-2). Als Messposition des Messmikrofons wird ausschließlich die spätere Abhörposition gewählt. Als Senderposition wird ausschließlich die direkte frontale Richtung und ein Abstand von 2 m zum Empfänger gewählt. Der Lautsprecher ist in seiner Hauptabstrahlrichtung auf das Messmikrofon ausgerichtet. Es ergibt sich für die Nachhallzeit ein gemittelter Wert von 1,5 s. Das Deutlichkeitsmaß C50 ergibt sich zu 8 dB, das Klarheitsmaß C80 zu 9,5 dB.

**Humboldt-Raum (Hu201)** - Der Humboldt-Raum stellt einen regulären Seminarraum der TU Ilmenau dar. Seine Raummaße an den Innenwänden betragen 7,6 m x 10,1 m x 3,7 mit einem Raumvolumen gemäß Raummaß von 284 m<sup>3</sup>. Er ist mit mehreren Tischreihen und Stühlen in Parallelanordnung ausgestattet.

Raumakustische Messungen sind für den Raum im Zusammenhang mit den dort durchgeführten Hörversuchen durchgeführt worden. Die Messung entspricht nicht in vollem Umfang den Forderungen in Standard EN ISO 3382:2000/3.5 [115]. Die verwendete Schallquelle muss als gerichtete Schallquelle angenommen werden (Lautsprecher Geithain Mo-2). Als Messposition des Messmikrofons wird ausschließlich die spätere Abhörposition gewählt. Als Senderposition wird ausschließlich die direkte frontale Richtung und ein Abstand von 2 m zum Empfänger gewählt. Der Lautsprecher ist in seiner Hauptabstrahlrichtung auf das Messmikrofon ausgerichtet. Es ergibt sich für die Nachhallzeit ein gemittelter Wert von 1,1 s. Das Deutlichkeitsmaß C50 ergibt sich zu 7 dB, das Klarheitsmaß C80 zu 9,2 dB.



## Kapitel 3

### Hörereignisbildung

Perhaps more than any other discipline, audio engineering involves not only purely objective characterization but also subjective interpretations. It is the listening experience, that personal and most private sensation, which is the intended result of our labors in audio engineering. No technical measurement, however glorified with mathematics, can escape that fact.

---

*Richard C. Heyser*  
*aus einem Anschreiben zur Heyser Memorial Lecture Series*  
*142. AES Convention [113]*

### 3.1 Merkmale räumlicher Audioqualität

In der Literatur liegen zahlreiche Untersuchungen vor, welche sich mit der Ermittlung von akustischen und auditiven Qualitätsmerkmalen des räumlichen Hörens beschäftigen. Im Speziellen werden die Wahrnehmung von Konzertsälen nach Beranek [58], die „multi-level auditory assessment language“ (MURAL) von Letowsky [138], die Attributerhebung von Berg und Rumsey [59, 60], das „spatial auditory quality inventory“ (SAQI) von Lindau et al. [140] sowie das „lexicon for spatial sound quality„ von Zacharov et al. [235] in einer Übersicht betrachtet. Nachfolgend werden die für die eigene Arbeit als relevant eingeschätzten Qualitätsmerkmale des räumlichen Hörens tiefergehend erfasst.

Beranek betrachtet vor allem auch die auditive Wahrnehmung in Konzertsälen [58]. Er beschreibt hierbei u. a. die musikalische Akustik in der, ausgehend von akustischen Eigenschaften der Schalle, die Wirkung der Audiosignale beschrieben wird. Dies umfasst bspw. Merkmale wie Räumlichkeit, Wärme, Intimität/Präsenz oder auch Brillanz. Er ordnet weiterhin den Qualitätsmerkmalen akustische Parameter zu. Er beschreibt diese Parameter und zeigt Methoden zur Messung dieser auf. Zu diesen gehören bspw. die Initialzeitlücke, Nachhallumhüllung und Energieabklingkurven.

Letowski betrachtet in seinen Untersuchungen die Herausbildung von Klangqualität als perceptiven Vorgang und definiert dafür Begriffe, mit denen sich das im Gehirn entstehende Hörereignis nach außen hin beschreiben lässt [138]. Hierbei unterteilt er die Gesamtaudioqualität in eine Qualität basierend auf Timbre-Merkmalen und in eine räumliche Audioqualität. Gemäß dieser Annahme formulierte er eine hierarchische Struktur von Attributen, welche die Audioqualität als multidimensionales Konstrukt abbildet. Hieraus entwickelt er die „multi-level auditory assessment language“ (MURAL), welche sich als kreisförmige Anordnung von Qualitätsmerkmalen mit verschiedenen Hierarchieebenen abbildet [138]. Unabhängige und sich ergänzende Qualitätsmerkmale liegen auf gleichen Radien, während voneinander abhängige Merkmale in einem radialen Bezug abgebildet sind. Eine Gewichtung bzw. Relevanz der einzelnen Merkmale durch das Individuum werden in diesem Modellansatz nicht abgebildet.

Berg und Rumsey untersuchen in ihrer Forschung Erhebungsmethoden und Strukturierungen von Qualitätsmerkmalen bei Wiedergabe von Surround Sound [59, 60]. Sie unterteilen die wahrgenommene Audioqualität in die Kategorien „Timbral Quality“, „Spatial Quality“ und „Technical Quality“. In ihren Experimenten zur räumlichen Audioqualität definieren sie verschiedene Merkmale, welche den Raum, Quelle, Präsenz und Natürlichkeit abdecken [60].

Lindau et al. erstellen in ihrem „spatial auditory quality inventory“ (SAQI) eine umfassende Zusammenfassung von Merkmalen zur Beschreibung von räumlicher Audioqualität. Weiterhin formulieren sie Methoden zur Erhebung von Bewertungen dieser Qualitätsmerkmale [140]. Es wird der Versuch unternommen, die übergeordneten Merkmale Plausibilität und Authentizität in eine Vielzahl von räumlichen Qualitätsmerkmalen zu unterteilen. Die Zielstellung ist hierbei Attribute und Methoden zu bieten, die eine technische Verbesserung des Systems, eine Ab-

schätzung der wahrnehmungsbezogenen Notwendigkeit von technischen Systemkomplexitäten oder einen Vergleich von Systemen zu ermöglichen [140]. Die Qualitätsmerkmale sind durch eine Experten-Fokusgruppe entwickelt worden. Zielstellung ist die „Creation of a consensus vocabulary for evaluating apparatus-related perceptual differences between technically generated acoustic environments or with respect to a presented or imagined acoustic reality.“ [140]. Es werden Merkmale und deren Ausprägungen aus verschiedenen Kategorien vorgeschlagen. Dies umfassen Timbre- und Tonalitätsmerkmale, geometriebasierte Merkmale, wie bspw. Richtung, Distanz und Externalität, raumbezogenen Merkmale, zeitbasierte Merkmale, Artefakte und andere. Es wird weiterhin eine Messmethode der Merkmale in Hörversuchen vorgestellt.

Zacharov et al. formulieren einen hierarchischen Ansatz zur Beschreibung von Basic Audio Quality durch einzelnen Qualitätsmerkmale [235]. Ähnlich zu den vorangestellten Arbeiten werden Timbre-Merkmale, räumliche Merkmale, Transparenz, Dynamik-Merkmale, Lautheit und Artefakte definiert.

Eine grundlegende Annahme in den vorangestellten hierarchischen Ansätzen ist, dass die einzelnen Merkmale auf gleicher Hierarchieebene unabhängig voneinander sind. Nur durch diese Annahme lassen sich derartige Merkmalsmodelle aufbauen. Es ist allerdings kritisch anzumerken, dass es zu prüfen gilt, unter welchen Kontextbedingungen diese Grundannahme Gültigkeit besitzt und wann nicht. Hierzu sind offensichtlich weitere Forschungen nötig.

### 3.1.1 Hörereignisrichtung

In der vorliegenden Arbeit wird die Fähigkeit zur Lokalisierung von Hörereignissen definiert als die Wahrnehmung der Richtung des Direktschalls eines Schallereignisses in Azimut und Elevation. Eine für die Anwendung passende Genauigkeit in der Lokalisierung eines Hörereignisses wird als ein zwingend notwendiges Qualitätsmerkmal für ein binaurales Kopfhörersystem angesehen. Besonders in der binauralen Simulation von Multikanallautsprecheranordnungen [201] müssen Fehllokalisationen minimiert werden. Das Richtungshören der Synthese soll vergleichbar mit dem Hören über reale Lautsprecheraufbauten sein [1]. Es soll eine korrekte Resynthese der Position von Audioobjekten und sogenannten Phantomschallquellen gewährleistet werden. Dies umfasst sowohl die Lokalisationsgenauigkeit in Azimut und Elevation, als auch die Vorne-Hinten-Lokalisation.

Die Verwendung eines Binauralsynthesystems und nicht individueller Übertragungsfunktionen lässt vor allem eine Verschlechterung der Lokalisationsleistung in der Elevation und einen Anstieg der Anzahl der Vorne-Hinten-Vertauschungen erwarten [103, 154, 228]. Die Verwendung von individuellen Übertragungsfunktionen führt zu einer verringerten Lokalisationsabweichung und Anzahl an Vorne-hinten-Vertauschungen gegenüber der Lokalisationsleistung in realen akustischen Schallfeldern [189]. Die Streuung der Lokalisationsleistung scheint gemäß Wenzel et al. [228] keine oder lediglich geringe Abhängigkeit von den verwendeten Übertragungsfunktionen und realen Hörumgebungen aufzuweisen.

Weiterhin findet eine Überschätzung seitlicher Richtungen bei Verwendung einer nicht individualisierten Binauralsynthese gegenüber einer realen Hörumgebung statt. In Versuchen zur Auswahl von Übertragungsfunktionen durch Testpersonen wurde festgestellt, dass Testpersonen vermehrt Übertragungsfunktionen auswählen, welche zu einer Überschätzung führen [189].

### 3.1.2 Hörereignisdistanz

Die Wahrnehmung der Hörereignisdistanz lässt sich in eine absolute und eine relative Wahrnehmung unterteilen [55]. Die absolute Distanzwahrnehmung soll hierbei die Fähigkeit beschreiben, dass Distanzen eines Schallereignisses direkt beim erstmaligen Hören angegeben werden können. Das relative Distanzhören beinhaltet dahingegen Merkmale, die aus einem mehrmaligen Hören und aus einem Vergleich der Hörereignisse resultieren.

Eine Übersicht über die Distanzwahrnehmung für kleine Distanzen ( $<2\text{ m}$ ) wird in der Arbeit von Brungart gegeben [72]. In den dort genannten Studien zu gerade wahrnehmbaren Änderungen zeigen sich für nachhallarme Umgebungen Schwellwerte von 6% zur festgelegten Referenzdistanz von bis zu 2 m. Werden die Intensitätsmerkmale der unterschiedlich weit entfernten Quellen ausgeglichen steigen die Schwellen auf bis zu 16% an.

In nachhallarmen Umgebungen und eher geringen Entfernungen ( $<10\text{ m}$ ) basiert die Distanzwahrnehmung auf der Auswertung von relativen Intensitätsunterschieden und auf der Auswertung von binauralen Unterschieden, sollten diese durch seitlich ausgelenkte Quellenpositionen gegeben sein. Direkt frontal angeordnete Schallquellen sind nicht ohne weitere Vorkenntnisse in ihrer Distanz in nachhallfreien Umgebungen lokalisierbar. Allerdings können binaurale Merkmale durch Kopfbewegung und Peilbewegung zur Distanzschätzung herangezogen werden. In Experimenten von Mershon und King [152] konnte gezeigt werden, dass große Ungenauigkeiten in der Distanzbewertung von Schallereignissen in nachhallarmer Umgebung und bei erstmaliger Präsentation aus frontaler Richtung bestehen. Wird das gleiche Schallsignal aus anderen Richtungen und Distanzen weitere Male präsentiert, steigt die Anzahl korrekter Distanzbewertungen an. Es wird geschlussfolgert, dass dieser Anstieg vor allem durch einen relativen Vergleich zwischen den verschiedenen Hörereignissen ermöglicht wird. Weiterhin muss ein erstmaliger Zusammenhang zwischen Schallintensität und Distanz erkannt werden. Dies kann durch visuelle Merkmale erreicht werden oder auch dadurch, dass die Schallsignal-Entfernungs-Relation, wie bspw. bei Sprache, bekannt ist.

In reflexionsbehafteten Umgebungen sind neben den Intensitätsunterschieden vor allem raumakustische Parameter wie das DRR und die ITDG relevant [57, 152]. Die Korrektheit der Distanzwahrnehmung erhöht sich dadurch gegenüber nachhallfreien Situationen erheblich [56]. Bei größeren Entfernungen ( $>10\text{ m}$ ) kommen weitere Merkmale durch bspw. frequenzabhängige Luftdämpfung hinzu.

### 3.1.3 Externalität

Der Begriff Externalität oder Externalisierung beschreibt die Wahrnehmung des Ortes eines Hörereignisses außerhalb des Kopfes. Das Gegenstück dazu ist die Im-Kopf-Lokalisation (IKL). Per Definition findet eine IKL dann statt, wenn das Hörereignis innerhalb des Kopfes lokalisiert wird. Die Geometrie des Kopfes ist damit eindeutig als Grenze definiert und dem Qualitätsmerkmal Externalität wird eine bimodale Ausprägung zugewiesen. Die Wahrnehmung von Hörereignissen außerhalb des Kopfes wird als ein zwingend notwendiges Qualitätsmerkmal eines binauralen Kopfhörersystems zur Erzeugung einer plausiblen auditorischen Illusion angesehen.

In Untersuchungen von Toole [212] und Plenge [168] wird deutlich, dass der Effekt der IKL nicht zwingend von der Verwendung eines Kopfhörersystems abhängig ist. Toole konnte in seinen Experimenten zeigen, dass IKL auch bei Verwendung von Lautsprechern in nachhallarmer Umgebung stattfindet [212]. Den Testpersonen wurden sowohl Audiosignale aus einzelnen Lautsprechern vor und hinter der Person, als auch aus bis zu vier Lautsprechern um die Testperson herum dargeboten. Es erfolgte eine gleichzeitige Beschallung aus einem, zwei oder vier Lautsprechern mit Rauschsignalen unterschiedlicher Bandbreite. Es zeigt sich, dass bei Präsentation der Audiosignale über zwei und vier Lautsprecher eine Häufigkeit der Externalisierung der Hörereignisse von kleiner 20% erreicht wird. Bei Präsentation von Audiosignalen aus einem Lautsprecher hinter der Testperson werden für breitbandiges weißes Rauschen Externalisierungswerte von ca. 80% erreicht. Die Darbietung der Audiosignale aus einem frontalen Lautsprecher führt zu Werten von ca. 65% für breitbandiges weißes Rauschen.

Toole untersuchte weiterhin den Einfluss geringfügiger Kopfbewegungen zur Identifikation der Hörereignisorte. Hierzu minimierte er die möglichen Kopfbewegungen der Testpersonen [212]. Bei Ausschluss von relativen Änderungen in der Schalleinfallrichtung zeigen sich vergleichbare Bewertungen bzgl. der Externalität wie in dem vorangegangenen Experiment. Toole schlussfolgert, dass in ihrer Auslenkung große suchende Kopfbewegungen zu einer korrekten Lokalisierung der Schallereignisse führen. Dahingegen scheinen kleine Auslenkungen nicht ausreichend für einen Anstieg an Externalisierung von Hörereignissen zu sein [212]. Die Einflüsse von Kopfbewegungen auf die Externalisierung von Hörereignissen werden in der vorliegenden Arbeit im Kapitel 4.4 näher untersucht.

In einer weiterführenden Untersuchung zur Entstehung der IKL durch Plenge [168] wird die Hypothese aufgestellt, dass die IKL durch eine fehlende Gewöhnung oder einen unzureichenden Lernprozess entsteht. Der Lernprozess umfasst hierbei das kurzzeitige Lernen von Eigenschaften der Schallquelle und des Abhörtraumes. Es werden Experimente vorgestellt, um a) „... einen gleitenden Übergang zwischen Außer-Kopf-Lokalisation (AKL) und IKL zu schaffen ...“ [168], b) einen „Vergleich kopfbezogener elektroakustischer Speicherung mit dem Originalsignal [durchzuführen]“ [168] und c) „Schließlich [werden] die oben erwähnten

Störungen bei der Lokalisation künstlich herbeigeführt, indem ein Erlernen von Schallquellen- und Raumeigenschaften unterbunden wurde.“ [168].

Im Experiment a) zeigt sich, dass ein gleitender Übergang zwischen einer Außer-Kopf-Lokalisation und einer Im-Kopf-Lokalisation bei Lautsprecherwiedergabe in einem reflexionsarmen Raum nicht eindeutig herstellbar ist [168]. Bereits kleine Veränderungen in den Testsignalen führen entweder zur IKL oder zur Wahrnehmung von Externalität. Die Wahrnehmung von Externalität lässt sich anhand der Ergebnisse dieses Experimentes als ein bimodales Qualitätsmerkmal auffassen. Experiment b) befasst sich mit dem Vergleich einer Lautsprecherwiedergabe mit der binauralen Synthese der gleichen Lautsprecherwiedergabe über Kopfhörer. Die Ergebnisse zeigen, dass das Auftreten der Im-Kopf-Lokalisation „unabhängig von der Kopfhörerwiedergabe“ [168] ist. Als eine Schlussfolgerung wird von Plenge festgehalten, dass die IKL dann entsteht, wenn die Ohrsignale „keiner realen Quelle außerhalb des Kopfes zugeordnet werden können und dürfen - sie dürfen mit keiner realen Quelle verwechselbar sein -, müssen sozusagen am einzig verbleibenden Ort, dort wo keine Schallquelle auftreten kann, nämlich im Kopf „untergebracht“ werden“ [168]. Im ersten Versuchsteil des Experimentes c) wird in einem reflexionsarmen Raum ein synthetisches Schallfeld eines Konzertsaaes erzeugt. Direktschall, frühe Reflexionen und diffuser Nachhall werden durch Lautsprecherwiedergabe angenähert. Als Testsignal wird die nachhallfreie Wiedergabe eines Sprechers über einen Lautsprecher in Blickrichtung im Raum verwendet. Die Testpersonen sollen bewerten, ob sich der Sprecher Im-Kopf, Am-Kopf oder Außer-Kopf befindet. In lediglich 1 von 68 Bewertungen wird der Sprecher als extern bewertet [168]. Im zweiten Versuchsteil wird den Testpersonen ein über Kunstkopf aufgezeichnetes Schallfeld eines Konzertsaaes oder einer Außenaufnahme (Straßenlärm, usw.) über Kopfhörer binaural präsentiert. Das Testsignal wird unverändert über einen Lautsprecher wiedergegeben, beinhaltet aber nun entweder nachhallfreie oder nachhallbehaftete Sprach- oder Musiksignale. Plenge gibt an, dass eine Im- bzw. Am-Kopf-Lokalisation dann auftritt, wenn es eine „fehlende, mangelhafte oder falsche Schallquellen und -feldkenntnis und/oder die Signale und damit die Reize so geartet sind, dass sie keinem im Langzeitspeicher enthaltenem Reizmuster zugeordnet werden können“ [168]. Die Ergebnisse lassen vermuten, dass das Qualitätsmerkmal Externalität vom Kontext der Wiedergabe und der Abhörsituation beeinflusst wird.

Abbildung 3.1 zeigt die in den eigenen Untersuchungen angewendete mehrstufige Bewertungsskala für Externalität in prototypischer Darstellung. Auf Variationen der Bewertungsskala in den einzelnen Untersuchungen wird in den einzelnen Kapiteln gesondert eingegangen. Eine weiterführende Darstellung zu alternative Bewertungsskalen findet sich am Ende dieses Abschnittes.

Die Bewertung der Externalität eines Hörereignisses wird durch Auswahl der inneren, mittleren oder äußeren Region auf der Oberfläche durchgeführt. Hierbei werden für die einzelnen Bereiche folgende Definitionen in dem Test verwendet:

- a) Mittelpunkt: „Das Hörereignis ist vollständig in meinem Kopf und sehr diffus.“
- b) innerer Kreis: „Das Hörereignis ist vollständig in meinem Kopf und einfach zu lokalisieren.“
- c) mittlerer Kreis: „Das Hörereignis ist extern aber sehr nahe an meinen Ohren oder Kopf.“
- d) äußerer Kreis: „Das Hörereignis ist extern und einfach zu lokalisieren.“
- e) äußerer Punkt: „Das Hörereignis ist extern und sehr diffus.“

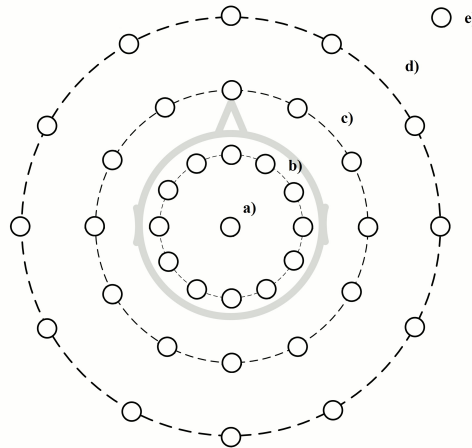


Abbildung 3.1: Bewertungsskala für Externalität in den Hörtests.

Aufgrund des perceptiven Zusammenhangs von Lokalisierung und Externalisierung eines Hörereignisses vereint die Skala diese beiden Qualitätsmerkmale. Diese Vereinigung bezieht sich aber darauf, dass es der Testperson ermöglicht werden soll, verschiedene Kombinationen von Externalisierung und Lokalisierung getrennt zu bewerten und auswertbar zu machen. Diese Trennung soll durch die Skalenpunkte a-b und d-e deutlich werden.

Gemäß der Definition liegt Externalisierung dann vor, wenn das Hörereignis außerhalb der geometrischen Ausdehnung des Kopfes wahrgenommen wird. Die Erfahrung zeigt allerdings, dass die geometrische Dimension des Kopfes ungeeignet als Entscheidungspunkt zur Bewertung von Externalität ist. Die Lage des perceptiven Entscheidungspunktes ist von Person zu Person verschieden. So kann für die eine Person bereits eindeutig Externalität vorliegen, wenn das Hörereignis sehr nahe am Kopf lokalisiert wird, während für eine andere Person noch eindeutig Im-Kopf-Lokalisation vorliegen kann. Aufgrund der individuellen Entscheidung wird das Merkmal Externalität auf einer Mehrpunkte-Skala bewertet. Hartmann und Wittenberg [105] verwenden in ihren Untersuchungen bspw. 4-Punkt-Skalen zur Bewertung. In den eigenen Untersuchungen wird ebenfalls eine derartige Skalierung verwendet (siehe hierzu Skalenpunkt c). Durch die Verwendung dieser Skalen und der individuellen Lage des Entscheidungspunktes ist es möglich, eine eher strenge oder weniger strenge Auswertungen der Bewertungen vorzunehmen. Eine weniger strenge Auswertung bzgl. Externalität würde dann vorliegen, wenn Bewertungen „Am-Kopf“ als extern bewertet würden. Eine strenge Auswertung liegt dann vor wenn „Am-Kopf“-Bewertungen als „Im-Kopf“-Lokalisation bewertet werden.

Es wird weiterhin versucht auf der Skala die Anzahl der Auswahlmöglichkeiten für Externalität und IKL gleich gewichtet zu gestalten. Es soll keine Verzerrung der Wahrnehmungsbewertung erzeugt werden, die allein durch die Anzahl der Bewertungsmöglichkeiten hervorgerufen wird. In eigenen Untersuchungen hat sich allerdings gezeigt, dass es in seltenen Fällen zu der Wahrnehmung von Externalität und diffuser Richtungslokalisierung des Hörereignisses kommen kann. Dies betrifft vor allem Hörsituationen, in denen sehr hallige und/oder sehr weit entfernte Audioobjekte dargeboten werden. Zur Erfassung dieser Wahrnehmung wird die Auswahlmöglichkeit e) (äußerer Punkt) mit aufgenommen. Aufgrund der geringen Anzahl an derartigen Bewertungen in den Hörversuchen wird davon ausgegangen, dass diese Asymmetrie, wenn überhaupt, lediglich einen geringen Effekt in Richtung einer Mehrbewertung von Externalität hat.

Die vorgestellte Bewertungsmethode ist eine Möglichkeit zur Abfrage von Externalität als bimodales Merkmal. Variationen dieser Skala finden in Untersuchungen von bspw. Hendrickx et al. [109], Gil-Carvajal et al. [98], Brimijoin et al. [71], Begault et al. [57], Hartmann und Wittenberg [105] und anderen, Verwendung. Die Unterschiede ergeben sich hinsichtlich Anzahl an Zwischenstufen und der Art der Auswertung in Form eines Indexwertes oder durch die Skalenstufen selber. Der Vorteil in Bezug auf die eigene Arbeit ergibt sich hinsichtlich Vergleichbarkeit der Untersuchungen.

Neben der Bewertungsmethode unter Beachtung des bimodalen Charakters des Externalisierungsmerkmals werden in der Literatur alternative Befragungsmethoden verwendet. Gängige Ansätze sind die Angabe der Ausprägung der Externalisierung von Hörereignissen auf mehrstufigen Skalen. In diesen Ansätzen entsprechen ein kleiner Zahlenwert einer geringen und ein hoher Zahlenwert einer hohen wahrgenommenen Externalität. Ein Beispiel hierfür ist die 6-Punkte Skala von Udesen et al. [215]. Es ist bei derartigen eher ordinal skalierten Skalen durchaus zu hinterfragen, inwieweit das Konstrukt Externalität überhaupt bewertet werden soll. Ist gemäß der obenstehenden Definition die Annahme, dass Externalität entweder vorliegt oder nicht vorliegt, scheint die Erstellung einer Rangordnung ungeeignet. Was soll bspw. eine mittlere Externalisierung eines Hörereignisses von 2,5 von 5,0 Skalenpunkten bedeuten?

Abbildung 3.2 verdeutlicht den Zusammenhang zwischen wahrgenommene Distanz und Externalität von Hörereignissen. Es wurde ein Hörversuch durchgeführt, in dem mittels Binaural-synthese Audioobjekte aus zwei akustisch verschiedenen Räumen und aus unterschiedlichen Entfernungen zur Aufnahme-position resynthetisiert wurden. Die beiden Räume sind ein halliger Seminarraum (SR) und ein wenig halliges Hörlabor (HL). Weitere Informationen zu den Räumen sind Kapitel 2.5 zu entnehmen. Der Hörtest wurde im halligen Seminarraum durchgeführt. Hierdurch ergeben sich Abhörsituationen mit Konvergenz und Divergenz zwischen synthetisiertem Raum und Abhör-raum. Die abzubildenden Distanzen sind 0,7 m, 0,9 m, 1,3 m, 2,1 m und 3,7 m in beiden Aufnahme-räumen. Die abzubildenden Richtungen sind  $210^\circ$  und  $330^\circ$ . Es wurden individuelle BRIRs (ind.) und Kunstkopf-BRIRs (Kemar) verwendet.



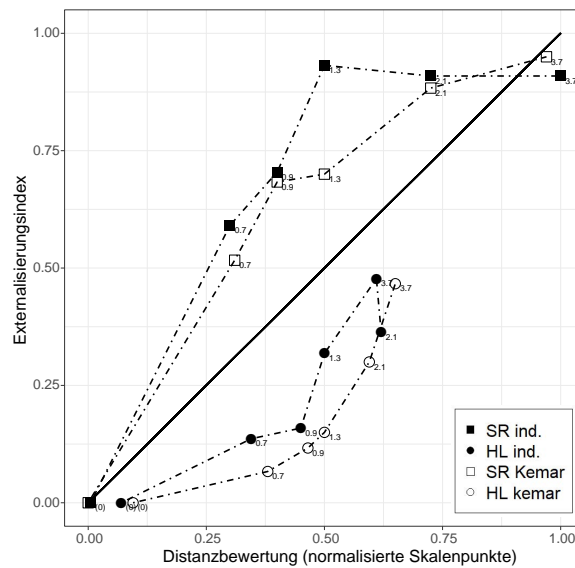


Abbildung 3.2: Zusammenhang von Externalisierung und Distanzwahrnehmung bei Synthese eines halligen Raumes (SR) und eines weniger halligen Raumes (HL) in dem halligen Raum (SR); Zahlen an den Symbolen entsprechen den geometrischen Distanzen der abzubildenden Hörereignisse in Meter; Mediane der Bewertungen für die Distanz und Mittelwerte die Externalisierung (auf Konfidenzintervalle wurde der Übersichtlichkeit wegen verzichtet).

Im ersten Versuchsteil bewerteten die Testpersonen die relative Distanz der Stimuli gegenüber einer gegebenen Referenz bei 1,3m für jeden Raum. Diese Bewertungen wurden über alle Bewertungen auf eine Skala von 0 (kleine Entfernung) bis 1 (große Entfernung) normalisiert. In einem zweiten Versuchsteil bewerteten die Testpersonen die Externalität der Hörereignisse. Aus diesen Bewertungen wird ein Externalitätsindex von 0 (keine Externalisierung) bis 1 (alle Stimuli der Kondition sind extern) erstellt.

Es zeigen sich je nach getesteter Kondition unterschiedliche Korrelationen zwischen Distanzbewertung und Externalisierung. Während die raumdivergenten Stimuli (Synthese HL in Abhörraum SR) eine geringe Externalisierung und eine geringere Distanzbewertung erreichen, werden die raumkonvergenten Stimuli (Synthese SR in Abhörraum SR) häufiger extern und als weiter weg bewertet. Es zeigen sich weiterhin bei den raumkonvergenten Stimuli Sättigungseffekte bzgl. der Externalisierung.

Die Grafik zeigt deutlich, dass es je nach Art der Stimuli und Testkondition zu unterschiedlichen Korrelationen zwischen Distanzwahrnehmung und Externalisierung kommt. Dieses Verhalten lässt den Schluss zu, dass eine Ableitung der Externalisierung von Hörereignissen über die Bewertung der Distanz nicht hinreichend ist. Es ist durchaus denkbar und die Erfahrung bestätigt dies, dass eine Distanzwahrnehmung von Hörereignissen gegeben sein kann, wengleich die Ereignisse im Kopf lokalisiert werden. Ein Beispiel hierfür ist die stereofone Audiowiedergabe über Kopfhörer bei Darbietung räumlicher Audioaufnahmen.

## 3.2 Einflüsse des Verwendungskontextes

Die räumliche auditive Wahrnehmung von Schallereignissen ist ein kognitiver Prozess, welcher zum einen von der verwendeten technischen Realisierung der Signalgenerierung und Audiowiedergabe und zum anderen von verschiedenen kontextabhängigen qualitätsbeeinflussenden Parametern der Audio(re-)synthese abhängt. Nachstehend werden die für die Arbeit grundlegenden Effekte vorgestellt und bisherige Untersuchungen dazu angeführt.

### 3.2.1 Präzedenzeffekt

Der auditive Präzedenzeffekt beschreibt das Priorisieren der zuerst bei einem Hörer eintreffenden Schallwellen eines Schallereignisses in der Wahrnehmung des Schallereignisses [222]. Danach eintreffende Schallwellen werden bis zum Erreichen einer Echoschwelle vom Gehör dem Präzedenzschall zugeordnet. Dieser Effekt bezieht sich im Besonderen auf die Lokalisierung und Richtungszuordnung von Hörereignissen in einer durch Schallreflexionen behafteten Umgebung. Es ist hierbei zwischen der Fusion und der Lokalisationsdominanz zu unterscheiden. Die Fusion beschreibt die Verschmelzung von Direktschall und Rückwürfen zu einem Hörereignis. Die Reflexionen verlieren hierbei aber nicht ihre Eigenschaften als Merkmalsträger für bspw. den Raumeindruck. Die Lokalisationsdominanz beschreibt die Unterdrückung der Richtungsinformation der Rückwürfe zur Herausbildung der Hörereignisrichtung des Direktschalls. Der zeitliche Bereich der Präzedenz bzw. der Fusion zwischen ersten Wellenfronten und danach eintreffenden Wellenfronten hängt von der Art des Stimulus, der spektralen Zusammensetzung der später eintreffenden Schallwellen in Bezug zum Direktschall und der Adaptation an die raumzeitlichen Muster von Direktschall und später eintreffendem Schall ab. In Untersuchungen bei Verwendung von Click-Folgen ergeben sich Echoschwellen zwischen 5-10 ms [143, 144]. Für Sprachsignale ergeben sich Echoschwellen von 30-40 ms und mehr [102]. Die Rückwurfunterdrückung verhält sich hierbei proportional zur spektralen Überlappung des Rückwurfs zum Direktschall [67].

Die zeitliche Größenordnung der genannten Echoschwellen unterliegt einem Aufbauprozess. Dieser Aufbau des Präzedenzeffektes ist in Experimenten von Clifton, Freyman, Litovsky und McGall [77, 78, 94] und durch Thurlow und Parks [211] intensiv untersucht worden. Eine mehrmalige Wiederholung von gleichen Mustern von Direktschall und Rückwürfen führte zu einem Anstieg der Echoschwelle. Dies deutet auf einen Adaptations- und Lernprozess hin, der weniger von der Zeitdauer, sondern vielmehr von der Anzahl an vergleichbaren Reflexionsmustern abhängig ist [81]. In Schlussfolgerung bedeutet dies, dass es bei einer sich ändernden akustischen Umgebung nicht schlagartig zu einem Zusammenbruch (oder Neuaufbau) des Präzedenzeffektes kommt.

Eine wesentliche Erweiterung des Präzedenzeffektes sind die von Clifton et al. [76] und Litovsky et al. [143] durchgeführten Experimente zur räumlichen Variation des Musters aus Direktschall und Reflexion. Eine Änderung des Musters führt zu einer Verringerung der Echoschwelle und

damit zu einem Zusammenbruch des Präzedenzeffektes. Nach der Änderung baut sich ein neuer Präzedenzeffekt auf. Dieser Effekt wird gemeinhin als Clifton-Effekt bezeichnet.

Studien von McGall et al. [148] untersuchten den Zusammenbruch einer aufgebauten Echoschwelle, wenn sich das erlernte Muster von Direktschall und Reflexion nicht nur zeitlich, sondern auch spektral ändert. Den Testpersonen wurde bspw. Stimulipaare mit einem breitbandigen Direktschall und einem hoch- bzw. tiefpassgefilterten Reflexionsschall präsentiert. Wurde den Testpersonen keine Möglichkeit zum Erlernen dieser Muster gegeben, sinkt die Echoschwelle und die Unterscheidbarkeit der Richtung von Direktschall und Reflexion steigt an. Dieses ist ebenfalls der Fall, wenn die Testpersonen ein Reflexionsmuster erlernt hatten aber ihnen ein abweichendes Muster, bspw. durch spektrale Änderung des reflektierten Schalls, präsentiert wird. Ähnliche Untersuchungen zur spektralen Abhängigkeit mit vergleichbaren Erkenntnissen sind von Tsakostas und Blauert [214] und Blauert und Braasch [66] durchgeführt worden.

In weiteren Untersuchungen von Keen (früher Clifton) und Feyman [129] wird der Aufbau und Abbau der Echoschwelle bei Variation des Musters aus Direktschall und Reflexionen untersucht. Die Variation besteht dabei in der Darbietung von Klickpaaren aus akustisch unterschiedlichen simulierten Räumen. Die gemessenen Echoschwellen sinken ab, wenn Divergenz zwischen den erlernten Klickpaaren aus einem Raum zu den ansonsten identischen Klickpaaren aus einem anderen Raum besteht. Es wird geschlussfolgert, dass die Hörer ein Modell ihrer akustischen Umgebung auf Basis von Lern- und Adaptationsvorgängen erstellen. Eine Änderung der akustischen Umgebung führt zu einer Divergenz zwischen dem aktuell vorliegende Modell und der sich geänderten Umgebung. Keen und Freyman beschreiben dies als „... a competition between two room models ...“ [129] bei dem der zuletzt erlernte Raum gewinnt. Dies führt schließlich zu einem Zusammenbruch des Präzedenzeffektes.

Eine geänderte aber dann stabile akustische Situation führt zu einem erneuten Aufbau der Echoschwelle und zu einem neuen Modell der akustischen Umgebung. Der Hörer erlernt immer wieder neu die raumzeitlichen Reflexionsmuster des Raumes in dem er sich befindet. Dies lässt die Vermutung zu, dass diese Muster auch einen Beitrag zur internen Referenz des räumlichen Hörens beitragen. Stimmen die Muster bspw. einer Synthese nicht mit den gespeicherten oder erwarteten Mustern überein, kann es zu einem Missverhältnis und somit zu fehlender Plausibilität kommen (siehe hierzu auch Plenge [168, 169]). Die dem Präzedenzeffekt zugrundeliegenden Wahrnehmungsprinzipien werden somit auch als Grundlage für den beschriebenen Raumdivergenzeffekt angesehen. Der Raumdivergenzeffekt ist eine Erweiterung des Präzedenzeffektes auf komplexe räumliche reale und resynthetisierte virtuelle Szenen, wenn die raumakustischen Parameter der zu synthetisierende Szene nicht mit den raumakustischen Parametern des Abhörtraumes übereinstimmen.

### 3.2.2 Raumdivergenzeffekt

Der Raumdivergenzeffekt (RDE) beschreibt den Einfluss einer abweichenden Raumakustik des zu synthetisierenden Raumes mit dem Abhörraum. Liegt eine solche Divergenz vor, kommt es zu einer Verringerung der wahrgenommenen Externalität des Hörereignisses. Die Begründung dieses Effektes liegt in einem kognitiven Missverhältnis zwischen erwartetem Hörereignis und aktuellem wahrgenommenen Hörereignis. Ein grundlegender Ansatz zur Erklärung findet sich im Präzedenzeffekt und dessen Erweiterung, dem Clifton-Effekt. Die im auditorischen System gespeicherten Muster zur Erkennung des Raumes und der Audioszene stimmen nicht mit denen aus der Synthese abgeleiteten Mustern überein. Sind diese Abweichungen ausreichend groß, ist das Gehirn nicht mehr in der Lage, perzeptive Fusion zwischen Synthese und Abhörraum herzustellen. Die Assimilation des aktuell Wahrgenommenen auf ein gespeichertes Schema/Mustern versagt.

Wie in vorangegangenen Ausführungen bereits dargelegt, scheint dieser Effekt durch raumakustische Divergenzen zwischen Synthese und Abhörraum ausgelöst zu werden. Eine auditorisch-visuelle Divergenz scheint den Effekt zu verstärken, ist aber selber nicht hinreichend für eine schlüssige Erklärung.

Nachstehend werden aktuelle Forschungsarbeiten zu diesem Thema vorgestellt und näher betrachtet. Hierfür werden die Arbeiten von Udesen et al. [215] und Gil-Carvajal et al. [98] herangezogen. Relevante Vorarbeiten von Plenge [168, 169] (siehe Abschnitt 3.1.3) und grundlegende Untersuchungen von Clifton/Keen [76] (siehe Abschnitt 3.2.1) sind den jeweiligen Veröffentlichungen und Abschnitten in dieser Arbeit zu entnehmen. Die eigenen Untersuchungen zum Raumdivergenzeffekt werden in Abschnitt 4.2 im Detail diskutiert. Erste eigene Untersuchungen zum RDE wurden bereits im Jahr 2011 durchgeführt und in sich anschließenden Experimenten bis 2016 verfeinert. Hierzu sei auch auf die eigene Literatur als Referenz verwiesen [19, 20, 22, 23, 31].

**RDE als visueller Effekt** - In der Arbeit von Udesen et al. [215] aus dem Jahr 2015 wird postuliert, dass die Verringerung der Externalisierung von Hörereignissen unter anderem auf einem Missverhältnis zwischen visuellem Eindruck des Abhörraumes und dem dargebotenen binaural synthetisiertem Raum zurückzuführen ist. Als weitere Einflüsse wurde der Hörer selbst und die dargebotene Schalleinfallrichtung identifiziert.

Es wird ein wohnzimmerähnlicher Raum (LR; Raum nach IEC60268-13 mit  $T60 = 0,4s$  [215]) mit darin befindlichen Lautsprechern binaural auralisiert. Die untersuchten Testkonditionen umfassen eine visuelle Variation des Abhörraumes. Das (virtuelle) Audiosignal des Wohnzimmers wird konstant gehalten. Es ergeben sich damit die Konditionen reales Wohnzimmer (RLR, real living room) mit Audiodarbietung über reale Lautsprecher im Raum und virtuelles Wohnzimmer (VLR, virtual living room) mit binauralsynthetisierten Lautsprechern. Neben dem Abhörraum Wohnzimmer wird eine Halle (Raumdimension größer als LR;  $T60 = 0,9s$

[215]) als Abhörraum verwendet. In dieser Halle sind die Positionen der zu synthetisierenden Lautsprecher als Attrappen sichtbar (VHWS, virtual hall with speakers) oder nicht sichtbar (VHWOS, virtual hall without speakers) [215]. Das Audiosignal ist jeweils identisch mit der Kondition VLR. Gemäß der Definition zum Raumdivergenzeffekt stellen die Konditionen RLR und VLR raumkonvergente Situationen dar. Die Konditionen VHWS und VHWOS sind raumdivergente Situationen.

Die Bewertungen zur Externalisierung der Hörereignisse zeigen, dass die raumkonvergenten Konditionen RLR und VLR hohe Externalisierungswerte erreichen. Die raumdivergenten Konditionen VHWS und VHWOS erreichen jeweils signifikant geringere Werte gegenüber der Kondition VLR. Udesen et al. schlussfolgern aus dem Unterschied zwischen VLR und VHWS: „Der einzige Parameter, der zwischen den virtuellen Testumgebungen geändert wurde, waren die visuellen Reize während der Tests. In der VLR-Testumgebung entsprach der Nachhall dem visuellen Eindruck des Raumes, während in der VHWS-Umgebung eine Diskrepanz zwischen dem virtuellen Nachhall und dem visuellen Eindruck des Raumes bestand.“ [215] (Übersetzung aus dem Englischen). Zwischen den Konditionen VHWS und VHWOS wird kein signifikanter Unterschied ermittelt. Udesen et al. schlussfolgern hieraus: „Dies deutet darauf hin, dass der visuelle Eindruck des Raumes die Externalisierung des Klangs mehr beeinflusst als die visuelle Präsenz der Lautsprecher.“ [215] (Übersetzung aus dem Englischen).

Gemäß der eigenen Hypothese zum Raumdivergenzeffekt müssen diese Aussagen relativiert werden. Es wird vielmehr behauptet, dass der Unterschied in der Externalisierung durch ein Missverhältnis der erlernten und durch die Audiosynthese aktuell erfahrenen und auf der Raumakustik basierenden Muster im Gehirn hervorgerufen wird. Das Vorhandensein von passenden visuellen Reizungen erhöht zwar den Grad der Externalisierung, erklärt aber nicht den Raumdivergenzeffekt. Zur Klärung dieser Behauptung sollen die eigenen Untersuchungen dienen (siehe Abschnitt 4.2).

Udesen et al. stellen weiterhin einen Einfluss des Hörers und der dargebotenen Schalleinfallrichtung fest. Schallquellen aus der Richtung  $0^\circ$  externalisieren signifikant schlechter als Quellen aus  $180^\circ$  und  $90^\circ$ . Dieser Effekt wird in den eigenen Untersuchungen ebenfalls adressiert und bestätigt (siehe Abschnitt 4.2).

**RDE als akustischer und visueller Effekt** - Eine weitere relevante Arbeit zum Raumdivergenzeffekt wurde von Gil-Carvajal et al. im Jahr 2016 veröffentlicht [98]. Hierbei werden Lautsprecherpositionen aus verschiedenen Räumen und Richtungen binaural resynthetisiert. Die Testkonditionen umfassen die getrennte und gemeinsame visuelle und auditorische Darbietung verschiedener Räume. Als Räume werden ein Referenzraum ( $T60 = 0,4s$ ,  $V = 99m^3$ ), ein kleiner und halliger Raum ( $T60 = 2,8s$ ,  $V = 43m^3$ ) und ein großer und nachhallarmer Raum ( $T60 < 0,01s$ ,  $V = 330m^3$ ) verwendet. Die zu synthetisierenden Schallquellen befinden sich in  $1,5m$  Entfernung in den Richtungen  $0^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $210^\circ$  und  $270^\circ$ . Für die Binauralsynthese werden individuelle BRIRs aus dem Referenzraum verwendet. Der Abhörraum ist

stets der Referenzraum. Hierdurch lassen sich visuelle und/oder akustische raumdivergente und raumkonvergente Testkonditionen bilden. Die Bewertung der Externalität wird durch eine 6-Punkte-Distanzskala von 0 „Im-Kopf“ bis 4 „Hörereignis von der Lautsprecherdistanz“ und 5 „Hörereignis größer als Lautsprecherdistanz“ verwendet.

Die Ergebnisse zeigen, dass für die Bewertungen der Konditionen mit visuellen und auditorischen Raummerkmalen ein Effekt gemäß Raumkonvergenz bzw. -divergenz auftritt. Die Hörereignisse werden bei raumdivergenten Konditionen als näher im Vergleich zur raumkonvergenten Referenzkondition bewertet [98]. Die größten Effekte werden zwischen der Referenzkondition und dem kleinen-halligen Raum ermittelt. Für die Bewertungen der Distanz bei den Testkonditionen mit rein auditorischen Merkmalen ergeben sich ebenfalls signifikante Unterschiede zwischen raumkonvergenten und raumdivergenten Konditionen. Die Betrachtung rein visueller Merkmale und dadurch verursachter Raumkonvergenz bzw. Raumdivergenz ergibt keine signifikanten Unterschiede [98].

Gil-Carvajal et al. schlussfolgern, dass auditorische Merkmale einen größeren Einfluss auf die Externalisierung der Hörereignisse haben als visuelle Merkmale [98]. Sie merken weiterhin an, dass bei den visuellen Merkmalen zwischen raumbezogenen Merkmalen und schallquellenbezogenen Merkmalen unterschieden werden sollte. Gil-Carvajal et al. stellen weiterhin fest, dass der Raumdivergenzeffekt für vordere und hintere Schallquellenpositionen stärker ausfällt als für seitliche Positionen.

In den eigenen Untersuchungen werden im Speziellen schallquellenbezogene visuelle Merkmale (Sichtbarkeit der Lautsprecher) und der Einfluss der Individualisierung der Binauralsynthese in Bezug zum Raumdivergenzeffekt untersucht (siehe Abschnitt 4.2). Es wird weiterhin eine direkte Messung der Externalisierung von Hörereignissen ohne Verwendung einer Distanzbewertung verwendet. Der Raumdivergenzeffekt wird nicht nur für einen Abhörraum, sondern für zwei zu synthetisierende Räume und Abhörräume untersucht. Hierdurch ergibt sich ein vollständiges Testdesign mit allen vier Möglichkeiten von Raumdivergenz und Raumkonvergenz für die Untersuchung.

### 3.2.3 Auditorisch-visuelle Interaktion

**Multimodale Wahrnehmung** - Die multimodale Wahrnehmung beschreibt die zeitgleiche Aufnahme von Informationen über ein Ereignis mit zwei oder mehreren Sinnesorganen. Bei der menschlichen Wahrnehmung ist in der Regel immer mit multimodalen Effekten zu rechnen. Nur in Ausnahmefällen steht ausschließlich das akustische Signal für die Informationsgewinnung zur Verfügung. Visuelle Hinweise sind im Normalfall vorhanden und werden somit zwangsläufig zur Herausbildung einer internen Realität hinzugezogen.

Eine prominente Verknüpfung von Empfindungen unterschiedlicher Sinne beschreibt die Synästhesie, in der sich bspw. Farb-, Audio- und Räumlichkeitswahrnehmung gegenseitig aktivieren. Wenngleich die Synästhesie eine direkte Verknüpfung von an sonst getrennten Bereichen der

Wahrnehmung beschreibt, legt doch die im Gehirn angenommene Strukturähnlichkeit und starke Verknüpfung der verschiedenen sinnesreizverarbeitenden Bereiche allgemeine synästhetische Effekte nahe. Beispielhaft sollen hier die Wirkung unterschiedlicher Farben auf die Lautheitswahrnehmung, Raumwirkung und Verknüpfungen von Ereignissen, Erfahrungen und Gefühlen genannt sein [186]. In Untersuchungen von Abou-Elleal [46] konnte gezeigt werden, dass die visuelle Darbietung von Räumen mit bspw. geöffneter oder geschlossener Decke und/oder Wänden einen Effekt auf die „akustische Qualität“ (Benennung nach [46]) in Abhängigkeit der dargebotenen Raumakustik (hallig vs. weniger hallig) hat.

**Ventriloquismus-Effekt** - Ein weiterer prominenter Effekt der Kopplung zwischen visueller und auditiver Wahrnehmung ist der Ventriloquismus- oder Bauchredner-Effekt [61, 190]. Dieser beschreibt die audiovisuelle Beeinflussung der wahrgenommenen Position eines Sinnesereignisses. Die auditive Positionswahrnehmung wird hierbei durch eine gleichzeitige Präsentation einer visuellen Reizung beeinflusst. Ist die örtliche Diskrepanz zwischen akustischem und optischem Stimulus groß genug, werden beide Ereignisse als separate Ereignisse wahrgenommen. Wird die Diskrepanz zwischen den beiden Stimuli geringer, wird der akustische Reiz verstärkt von der visuellen Reizung dominiert, bis es schließlich zu einer perceptiven Fusion der wahrgenommenen Orte der beiden Stimuli kommt. Eine detaillierte Ausführung zur perceptiven Wirkung des Ventriloquismus-Effektes findet sich in Abschnitt 4.1. In dieser Arbeit wird dieser Effekt für die vertikale auditive Richtungswahrnehmung bei Verwendung eines Binauralsynthesystems detaillierter untersucht.

Eine Vielzahl von Untersuchungen liegen für diesen Wahrnehmungseffekt in der Horizontalebene vor. Bertelson und Radeau [61] fanden Abweichungen in der Schalllokalisierung von ca.  $4^\circ$  für eine Diskrepanz von  $7^\circ$  zwischen akustischem Stimulus und optischem Stimulus,  $6,3^\circ$  für  $15^\circ$  und  $8,2^\circ$  für  $25^\circ$  Diskrepanz. Die Magnitude der Abweichung der wahrgenommenen Position des Hörereignisses nimmt mit steigender audiovisuelle Diskrepanz zu. In den Experimenten wurden Lautsprecher und Blitzlichter als Wiedergabeobjekte verwendet. Die Quellen waren in der Horizontalebene angeordnet. Die Bewertung der Richtung wurde durch Handzeigen abgegeben. Seeber und Fastl [190] verwendeten eine Zeigemethode, um die audiovisuelle Diskrepanz in realen und virtuellen Umgebungen zu untersuchen. In realen Umgebungen wurde eine mittlere Abweichung in der Lokalisierung von  $4,3^\circ$ ,  $1,9^\circ$  und  $4,2^\circ$  für die horizontalen Blickrichtungen  $-40^\circ$ ,  $0^\circ$  und  $40^\circ$  gemessen. Die Medianebene wurde nicht untersucht. Ähnliche Ergebnisse wurden in Experimenten unter Verwendung einer personalisierten Binauralsynthese über Kopfhörer gefunden. Eine geringere Abweichung wurde bei Verwendung nicht-individueller HRTFs festgestellt. Bohlander [68] ermittelte in seinen Versuchen Abweichungen von  $1,5^\circ$  bis  $5,9^\circ$  für  $45^\circ$  intersensorischer Diskrepanz für die Medianebene und reale Umgebung. Alais und Burr [47] führten Experimente zur Bestimmung psychometrischer Funktionen und Punkten subjektiver Gleichheit in Bezug auf den Ventriloquismus-Effekt in der Horizontalebene und in Abhängigkeit des Durchmessers der visuellen Quelle (Lichtpunkt) durch. Sie fanden einen

starken Einfluss des Durchmessers der Lichtquelle. Für kleine Lichtquellen ( $4^\circ$ ) verändert sich die wahrgenommen auditive Richtung wie erwartet direkt mit dem visuellen Stimulus. Für große Lichtflecke ( $64^\circ$ ) wird ein inverses Verhalten des Ventriloquismus-Effektes („... indicating that the click dominated perceived position ...“ [47]) gemessen.

In Untersuchungen von Kunze et al. [132] zur Bestimmung der wahrgenommenen Qualität einer audiovisuellen Szene bei Wiedergabe über Wellenfeldsynthese und einem 47“ Stereoskopischen TV Gerät ergibt sich ein Versatz von  $5^\circ$  bis  $10^\circ$ , ab wann eine Abweichung zwischen visueller und auditiver Szene als „störend“ beschrieben werden. Rummukainen et al. [183] untersuchen in ihren Experimenten die Lokalisation von audiovisuellen Ereignissen in der Horizontalebene bei Verwendung von Directional Audiocoding (DirAc, [172]) und einer 2D Videoprojektion bei unterschiedlichen optimalen und nicht optimalen Abhörpositionen. Es werden Abweichungen in der Lokalisation von  $\pm 8^\circ$  festgehalten. In ihren Untersuchungen konnte kein Einfluss des Schalldruckpegels und der Diffusität nachgewiesen werden. Der Ventriloquismus-Effekt scheint somit nicht maßgeblich von der Wiedergabelautstärke und der in den Signalen enthaltenen Halligkeit abzuhängen, so lange den Hörereignissen eine Richtung zugeordnet werden kann.

**Aufmerksamkeit** - Die räumliche Aufmerksamkeitssteuerung bei bimodaler Anregung wurde in Arbeiten von MacDonald und Ward [147] untersucht. Sie fanden heraus, dass ein auditiver Stimulus die Reaktionszeit auf einen kurz darauffolgenden visuellen Stimulus an gleicher Position deutlich verkürzt. Rummukainen und Mendonca untersuchten in [184] Methoden zur Verstärkung der Erkennung eines visuellen Zielobjektes in einer immersiven audiovisuellen Umgebung. Hierbei präsentierten sie zu dem visuellen Objekt dazugehörige und aufgaben-relevante räumliche auditorische Informationen. Die zu dem Zielobjekt gehörenden auditorischen Informationen erhöhten die Erkennung des Zielobjektes. Dieser Effekt wurde für seitliche aber nicht für zentral ausgerichtete visuelle Objekte beobachtet. Eine zusätzlich Verstärkung wurde bei dem visuellen Objekt vorausgehenden auditiven Informationen beobachtet.

Es wird geschlussfolgert, dass die auditive und visuelle Informationsverarbeitung im Gehirn entweder eng miteinander verknüpft sind oder gar auf gemeinsame Gehirnmechanismen zurückzuführen sind. Bei der auditiven Raumwahrnehmung scheinen daher nicht nur die akustischen Signale einen Einfluss zu haben. Vielmehr tragen auch die visuellen Merkmale des Raumes zur Gesamtheit der Raumwahrnehmung bei.

**Bimodale Übereinstimmung** - Eine Steigerung der Erinnerungs- und Orientierungsleistung in virtuellen auditorischen und visuellen Umgebungen durch zusätzliche auditive Hinweise werden von Larsson et al. [135] berichtet. Dieser Effekt ist umso stärker, je mehr Übereinstimmung zwischen akustischen und optischen Signalen vorhanden ist. Es wird bestätigt, dass auditive Ereignisse einen messbaren Einfluss auf die visuelle Raumwahrnehmung haben. Die Autoren stellen die Hypothese auf, dass anders herum auch visuelle Hinweise die auditive Raumwahrnehmung maßgeblich beeinflussen und sich somit der Ventriloquismus-Effekt auf



die Schallquellenlokalisierung in Räumen erklären lässt [135]. Diese Hypothese wird durch eine zweite Studie unterstützt, in der die Einschätzung der auditiven Merkmale Schallquellenbreite, Entfernung zur Schallquelle und Raumgröße bei unterschiedlich detaillierter visueller Präsentation untersucht werden [134]. Die Schallquellenbreite wird bspw. im realen Raum und in einer dreidimensionalen virtuellen Umgebung signifikant höher eingeschätzt als wenn lediglich eine zweidimensionale Abbildung des Raumes vorliegt. Die Entfernung der Schallquelle und die Raumgröße werden im Originalraum deutlich geringer eingeschätzt als bei allen anderen visuellen Stimuli. Es wird geschlossen, dass die Raumwahrnehmung vom zunehmenden Realismus der visuellen Präsentation abhängt. In Bezug auf den Ventriloquismus-Effekt wird geschlossen, dass bei nicht übereinstimmenden auditiven und visuellen Merkmalen die visuellen Merkmale dominieren. Ein ähnlicher Zusammenhang wird in Untersuchungen von Valente et al. gefunden [217]. Bei der Wiedergabe einer auditiven Szene mittels Binauralsynthese und unterschiedlich breiten optischen Abbildungen von Schallquellen steigt die wahrgenommene Schallquellenbreite proportional mit der zunehmenden optischen Abbildungsbreite.



## Kapitel 4

### Perzeptive Messung kontextabhängiger Qualitätsparameter

Das aller sicherste was wir wissen, das ist zweifellos das, was wir selber an unserem eigenen Leibe erfahren. [...] Daher bildet die erlebte Sinnwelt die einzige unangreifbare Grundlage für die Arbeit der exakten Wissenschaft.

---

*Max Planck*

*Audioaufnahme aus Wissenschaft und Leben [167]*

## 4.1 Ventriloquismus-Effekt

Nachstehende Experimente überprüfen die Abweichung in der Richtungswahrnehmung von Hörereignissen bei in ihrer Richtung abweichenden visuellen Stimuli. Es wird überprüft, ob diese Abweichungen bei binauraler Kopfhörerwiedergabe vergleichbar mit denen unter natürlichen Abhörbedingungen sind. Es werden psychometrische Funktionen der Richtungswahrnehmung für frontale, seitliche und in der Vertikalen ausgelenkte Schallereignisse bestimmt. Diese Abweichungen werden in einem weiteren Experiment bei Abzug der auditiven Richtungswahrnehmung ohne visuelle Stimuli gemessen und ein intersensorischer Bias/Versatz nach [189] bestimmt.

### 4.1.1 Exp. I: psychometrische Funktionen

Es wird die perzeptive örtliche Fusion zwischen konkurrierenden visuellen Stimuli und mittels Binauralsynthese simulierten Lautsprecherpositionen untersucht. Dies beinhaltet die Entwicklung einer Testmethode, die Durchführung des Experimentes sowie dessen Auswertung.

#### 4.1.1.1 Untersuchungsgegenstand

Es wird die wahrgenommene örtliche Diskrepanz in der Richtungswahrnehmung zwischen optischen und akustischen Reizen bei binauraler Kopfhörerwiedergabe gemessen. Das nachstehende Experiment zeigt den Einfluss visueller Wahrnehmungsereignisse auf die Richtungswahrnehmung von Hörereignissen am Beispiel des Ventriloquismus-Effektes. Weiterführende Informationen sind der dazugehörigen Veröffentlichung zu entnehmen [27].

Nachstehende Fragen sollen beantwortet werden: Welche Abweichung in Grad in der Wahrnehmung der vertikalen Richtung eines Hörereignisses liegt vor, wenn in der Richtung konkurrierende optische und akustische Stimuli präsentiert werden? In welcher Form bildet sich ein psychometrische Funktion heraus?

#### 4.1.1.2 Versuchsaufbau

**Binauralsynthesystem** - Es werden mehrere Schallquellenpositionen mittels Binauralsynthese über Kopfhörer wiedergegeben. Das statische Binauralsynthesystem beinhaltet keine Abbildung einer Positions- oder Kopfposenänderung der Testperson. Zur Generierung der Testsignale werden Messungen von individuellen binauralen Raumimpulsantworten (BRIRs) für die verwendeten Raum- und Schallquellenpositionendurchgeführt. Das binaurale System wird für jeden Teilnehmer individuell angepasst, um Quadrantenfehler in der Richtungslokalisierung zu minimieren sowie die perzeptive Ähnlichkeit der Simulation mit den realen Lautsprechern zu erhöhen. Es wird ein Hörlabor mit definierter Raumakustik und ausreichendem Quellempfängerabstand gewählt (siehe hierzu Kapitel 2.5). Der Abstand zwischen Empfänger und Quelle wird so gewählt, dass sich Empfänger bzw. Quelle jeweils im Fernfeld des Lautsprechers sowie des Empfängers (Kopfes) befindet. Es sollen möglichst keine Veränderung der binauralen

Signale in Abhängigkeit von der Entfernung vorliegen. Die Kopfhörer werden mit individuellen Kopfhörer-Übertragungsfunktionen (HPTFs) entzerrt. In-Ohr-Mikrofone (Eigenbau [13]) werden verwendet, um einzelne BRIRs und einzelne HPTFs in der Nähe des Trommelfells jeden Probanden zu messen. Die Mikrofone werden zwischen den BRIR- und HPTF-Messungen nicht entfernt. Die Messungen der HPTFs werden über fünf Aufnahmen gemittelt, wobei die Kopfhörer für jede Aufnahme neu positioniert werden. Die Invertierung einer HPTF wird nach der Methode der kleinsten Quadrate mit minimaler Phaseninversion berechnet. Ein Bandpassfilter wird zwischen 80 Hz und 18 kHz eingesetzt. Die Messungen der BRIRs werden über drei Aufzeichnungen gemittelt. Für die Wiedergabe wird ein Stax Lambda Pro Kopfhörer verwendet.

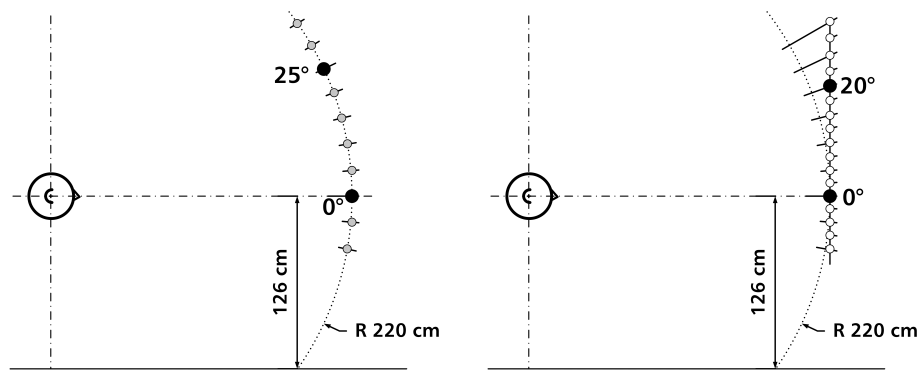


Abbildung 4.1: Positionen der akustischen und optischen Quellen für die Experimente I und II; Schallquellenpositionen für die Wiedergabe über Kopfhörer als schwarze Punkte bei  $0^\circ$  und  $+25^\circ$  (Experiment I links), sowie bei  $0^\circ$  und  $+20^\circ$  (Experiment II rechts); optische Quellpositionen (LEDs) als helle Punkte in einem Bereich von  $-10^\circ$  bis  $+35^\circ$  mit  $5^\circ$  Intervallen (Experiment I links) und in einem Bereich von  $-10^\circ$  bis  $+30^\circ$  mit  $2.5^\circ$  Intervallen (Experiment II rechts); In Experiment I sind die Quellen auf Kreissegmenten, in Experiment II auf einer Ebene angeordnet.

Ein Geithain Mo-2 Lautsprecher wird für die Messungen der BRIRs für jede der vier Positionen im Hörlabor (HL, siehe Kapitel 2.5) verwendet. Der Abstand von den Lautsprecherpositionen zum Messpunkt beträgt 2,2 m. Die Höhe der Quellpositionen beträgt für  $0^\circ$  Elevation 1,26 m (durchschnittliche Ohrposition einer sitzenden Person). Die Aufnahmepositionen der BRIRs sind identisch zur späteren Abhörposition im Test.

**Quellpositionen** - Die zu untersuchenden Quellpositionen für auditive und visuelle Quellen sind auf vertikalen Kreissegmenten angeordnet. Die Testperson befindet sich im Mittelpunkt der Anordnung. Die Quellen sind für den Hörer bei einer gleichbleibenden Distanz positioniert. Der linke Teil der Abbildung 4.1 gibt einen Überblick über den Aufbau. Neben den binaural zu synthetisierenden Lautsprecherpositionen befinden sich weiß leuchtende LEDs auf den Kreissegmenten. Bei den horizontalen Richtungen  $0^\circ$  und  $30^\circ$  sind jeweils zehn LEDs

angeordnet. Sie bedecken einen Bereich von  $-10^\circ$  bis  $35^\circ$  mit einem gleichförmigen Abstand von  $5^\circ$  in der Elevation. Der durch eine LED entstehende Leuchtfleck umfasst ca. 5 mm im Durchmesser und besitzt eine Lichtstärke von ca. 15 cd. Die einzelnen LEDs werden durch eine MATLAB gesteuerte Arduino-Mega Plattform [50] angesteuert. Die gesamte LED-Anordnung ist während des Tests sichtbar. Das Umgebungslicht im Raum ist auf ein Minimum reduziert, um anderweitige visuelle Einflüsse weitestgehend auszuschließen.

Tabelle 4.1 gibt eine Übersicht über die Bezeichnung der verwendeten Schallquellpositionen. Es werden Kombinationen aus vier Schallquellpositionen und 20 visuellen Quellpositionen im Test untersucht.

Tabelle 4.1: Azimut und Elevation der virt. Schallquellpositionen aus Experiment I.

Name	H0V0	H30V0	H0V25	H30V25
Azimuth	$0^\circ$	$+30^\circ$	$0^\circ$	$+30^\circ$
Elevation	$0^\circ$	$0^\circ$	$+25^\circ$	$+25^\circ$

**Audiosignale** - Für jede der beiden horizontalen Richtungen werden alle Kombinationen von akustischen und optischen Quellpositionen im Test untersucht. Als Testsignale kommen eine nachhallfreie Aufnahme eines Saxofones (6 s) und eine Rauschpulsfolge (5 Rauschpulse mit je 30 ms Dauer, 3 ms Ein- und Ausblendung mit Cosinus-Fenster und 70 ms Stille zwischen den Rauschpulsen) zur Verwendung. Das Saxofonsignal wird wegen seiner zur Sprache ähnlichen spektralen und tonalen Charakteristik gewählt [162, 204] ohne dass die Distanzwahrnehmung durch Artikulation oder Erfahrung beeinflusst wird [64]. Die Rauschpulsfolge wird aufgrund ihrer gut zu lokalisierenden Merkmale gewählt [189]. Die Reihenfolge der Stimuli im Test ist für jede Testperson zufällig angeordnet.

#### 4.1.1.3 Versuchsdurchführung

**Testpersonen** - Zwei weibliche und drei männliche Testpersonen mit keinen bekannten Hörschädigungen (Audiometrie bis 8 kHz) und mit einem Alter von 29 bis 33 Jahren nehmen am Test teil. Die Testpersonen sind erfahren in der Teilnahme an Wahrnehmungstests. Vor der Durchführung des Tests wird ein Training durchgeführt, um die Teilnehmer mit dem Testaufbau, Testablauf und Teststimuli vertraut zu machen. Zusätzlich erhalten die Testpersonen eine verbale und schriftliche Testeinführung sowie eine Definition der Begriffe Lokalisation und Externalität (siehe [105, 151]). Jeder Teilnehmer soll eine Auswahl von Teststimuli bei Übereinstimmung und Divergenz zwischen auditiver und visueller Position, gemäß dem im Haupttest verwendeten Bewertungsschema, bewerten.

**Durchführung** - Das Experiment umfasst eine Sitzung zur Untersuchung des angenommenen Einflusses visueller Merkmale auf die auditive Lokalisation. Die Sitzung ist in drei Teile

unterteilt. Teil eins beinhaltet das Training der Testpersonen zur Etablierung der Perzepte Lokalisierung und Externalität. Die Training-Stimuli umfassen die vier akustischen Richtungen, die zwei Audiosignale und Übereinstimmung bzw. Divergenz zwischen auditiven und visuellem Stimulus. Der zweite und dritte Teil besteht aus drei Wiederholungen der Teststimuli mit einer ca. fünf minütigen Pause. Die Gesamtanzahl der Stimuli beträgt 256 ( $3_{\text{Wiederholungen}} \times 2_{\text{Audiosignale}} \times 4_{\text{akustischePositionen}} \times 10_{\text{visuellePositionen}} = 240$  plus 16 Trainingsstimuli) pro Testperson. Das Durchlaufen der gesamten Sitzung dauert ca. 60 Minuten.

Die Testpersonen sollen die nachstehende Frage beantworten: Nehmen Sie den akustischen Stimulus unterhalb, auf gleicher Höhe oder oberhalb des optischen Stimulus wahr? Die Testpersonen werden instruiert, ihren Kopf während der Präsentation und Bewertung stillzuhalten. Weiterhin sollen Sie erst, nachdem Sie das komplette Signal gehört haben ihre Bewertung abgeben. Um mögliche Bewegungen oder sonstige Störungen der Testpersonen durch ein Eingabegerät zu minimieren, werden die Bewertungen verbal von den Testpersonen geäußert. Der Versuchsleitende überträgt die Äußerungen in einen Erfassungsbogen. Augenbewegungen sind während des Tests explizit erlaubt, um die Fixierung auf das Sinnesereignis zu erhöhen und um eine bessere Lokalisierung der beiden Stimuli zu ermöglichen. Ein wiederholtes Anhören der Stimulus-Paare ist auf Anfrage der Testperson ebenfalls möglich.

#### 4.1.1.4 Ergebnisse

Die Bewertungen zur Richtungswahrnehmung sind als normalisierte Häufigkeiten angegeben. Es werden keine signifikanten Unterschiede (McNemar Test, Signifikanzniveau 5%) zwischen den Bewertungen der Signale Saxofon und Rauschpulsfolge für die Lokalisationsbewertung „gleich“ und „nicht gleich“ (Summe von „unterhalb“ oder „oberhalb“) festgestellt. Die Bewertungen der beiden Signale werden zur weiteren Analyse zusammengefasst. Die Ergebnisse aus dem ersten Teil (Training) des Experimentes zeigen, dass alle Testpersonen die Stimuli mit 0° Abweichung zwischen akustischer und visueller Position auch als solche bewerten konnten. Die maximalen Abweichungen werden ebenfalls von allen Testpersonen erkannt.

Abbildung 4.2 zeigt die normalisierten Häufigkeiten der Bewertungen für die akustischen Positionen H0V0, H30V0, H0V25 und H30V25 als Funktion der Divergenz zwischen akustischer und visueller Quellposition in Grad. Es sind jeweils die Häufigkeiten für die Bewertungen „unterhalb“, „gleich“ und „oberhalb“ abgebildet. Die horizontale Linie markiert 50% der Bewertungen.

Für die vertikal ausgelenkten akustischen Positionen H0V25 und H30V25 zeigen die Kurven für die Bewertung „gleich“ einen breiteren Verlauf auf als die Bewertungen „gleich“ für die 0° vertikalen akustischen Positionen. Die Testpersonen tolerieren eine größere Abweichung zwischen den optischen und akustischen Quellpositionen, wenn die Schallquellen in der Elevation ausgelenkt sind.

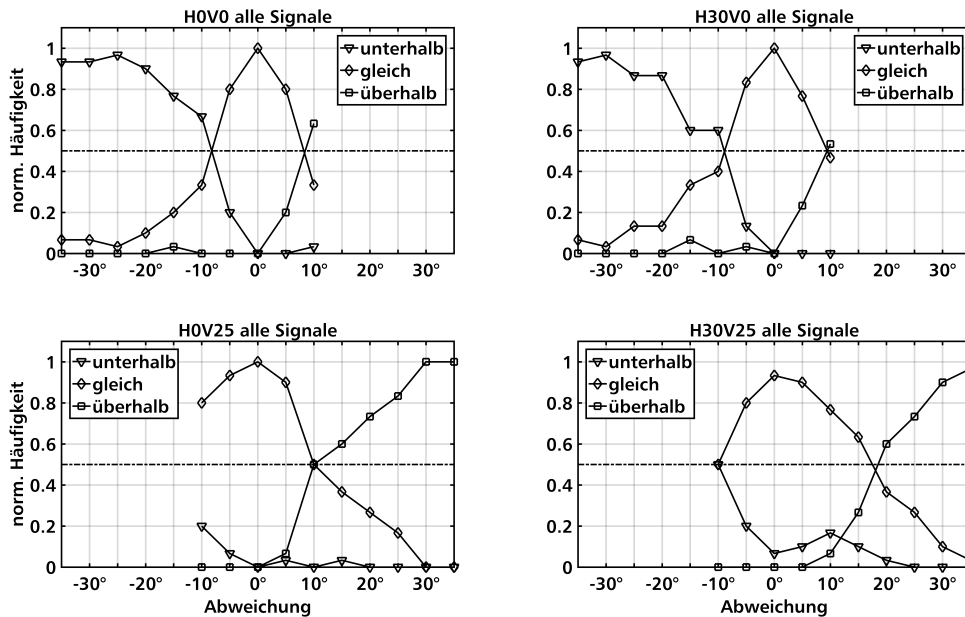


Abbildung 4.2: Lokalisierungsbewertungen als normalisierte Häufigkeiten für die akustischen Positionen H0V0, H30V0, H0V25, H30V25 und für beide Signaltypen zusammen; x-Achse: Abweichungen zwischen Audiostimulus und visuellem Stimulus; negative Werte für Audiostimulus unterhalb des visuellen Stimulus; horizontale Linie markiert 50% der Bewertungen.

Tabelle 4.2 listet die geschätzten Winkel der Abweichungen für die Bewertungen „gleich“ aller Testpersonen am 50% Punkt der normalisierten Häufigkeiten auf. Ein Anstieg der wahrgenommenen Abweichung ist für die vertikal ausgelenkten Positionen der akustischen Quellen bei H0V25 und H30V25 sichtbar.

Tabelle 4.2: Geschätzte Abweichungen in Grad für den 50% Punkt der Häufigkeiten der Bewertungen „gleich“; \* keine Schätzung möglich.

Name	H0V0	H30V0	H0V25	H30V25
50% Punkt	+8°/-8°	+9°/-9°	+10°/*	+17°/-10°

Zur Schätzung der Signifikanz der Unterschiede zwischen den Häufigkeiten für die Bewertung „gleich“ und „nicht gleich“ für die einzelnen Konditionen wird ein McNemar Test durchgeführt. Die Bewertung „nicht gleich“ wird definiert als die Summe der Bewertungen „oberhalb“ und „unterhalb“. Signifikante Unterschiede ( $p < 0,05$ ,  $N = 500$ ,  $DF = 1$ ) zeigen sich zwischen den Konditionen H0V0 und H0V25, sowie zwischen H0V0 und H30V25 und zwischen H30V0 und H30V25 (siehe Tabelle 4.3).

Die Zuverlässigkeit der Bewertungen über alle Testpersonen sind in Abbildung 4.3 für die 0° Elevation (V0) sowie für die 25° Elevation (V25) dargestellt. Sie wird aus der Anzahl von



Tabelle 4.3: Chi-Werte und Phi-Werte (in Klammern) für die Unterschiedsanalyse (McNemar Test) zwischen den Bewertungen für „gleich“ und „nicht gleich“ der Konditionen; signifikante Werte sind *kursiv* gedruckt ( $p < 0,05$ ,  $N=500$ ,  $DF=1$ ).

Name	H0V0	H30V0	H0V25	H30V25
H0V0	-	1,13 (0,04)	<i>8,64 (0,12)</i>	<i>14,73 (0,16)</i>
H30V0		-	3,53 (0,08)	<i>7,78 (0,11)</i>
H0V25				0,81 (0,04)

gleichbleibenden Antworten der Testpersonen der drei Wiederholungen der Stimuli berechnet. Die Zuverlässigkeit ist 100% für 0° vertikale Abweichung zwischen Audio- und Videostimulus und für alle Testsignale. Bei der Kondition H30V25 sind geringere Werte von 87% für das Testsignal Saxofon festzuhalten. Ein Abfall der Zuverlässigkeit der Bewertungen ist für steigende Abweichung zwischen visuellem und akustischem Objekt zu erkennen. Die visuellen und auditiven Richtungen können nicht deutlich von den Testpersonen getrennt oder als fusionierte Richtung wahrgenommen werden. Bei einem weiteren Anstieg der Abweichung steigt die Zuverlässigkeit der Bewertungen wieder an und erreicht Werte von nahe 100%. Den Testpersonen gelingt es einfacher, visuelle und auditive Richtung getrennt wahrzunehmen. Die hier verwendete Zuverlässigkeit der Bewertungen kann als Indikator für einen visuellen Einfluss auf die Lokalisierung eines auditiven Ereignisses gesehen werden. Die Minima der Kurven beschreiben ähnliche Werte für die Richtungswahrnehmung des auditiven Ereignisses bei gleichzeitiger Präsentation eines visuellen Ereignisses mit abweichender Richtung.

Als Ergebnis lässt sich zusammenfassen, dass eine audiovisuelle Diskrepanz in der Positionierung für obere seitliche (H30V25) und für obere frontale (H0V25) Positionen für die Testpersonen mehr tolerierbar sind als für die Konditionen mit 0° Elevation (H0V0 und H30V0). Die geschätzten Abweichungen überdecken einen Bereich von 8° für Positionen mit 0° Elevation und bis zu 17° für seitliche und obere Positionen. Es ist allerdings deutlich anzumerken, dass die in Experiment I gemessenen Abweichungen zusätzlich noch von der Lokalisierungsgenauigkeit ohne Einfluss von visuellen Merkmalen beeinflusst sind. Der Ventriloquismus-Effekt ist somit als geringer einzuschätzen. In dem durchgeführten Experiment können aufgrund der groben Winkelauflösung von 5° keine Aussagen über die exklusive auditive Lokalisierungsgenauigkeit getroffen werden. Weiterhin sind die vertikalen Positionen  $> 30^\circ$  für einige Testpersonen mit Brille schwer ohne Kopfbewegungen zu erfassen, da die Objekte sich hinter dem Rand der Brille befinden. Aufgrund der Einschränkungen in Experiment I wird Experiment II zur weiterführenden Untersuchung des Ventriloquismus-Effektes in der Elevation bei binauraler Kopfhörerwiedergabe durchgeführt.

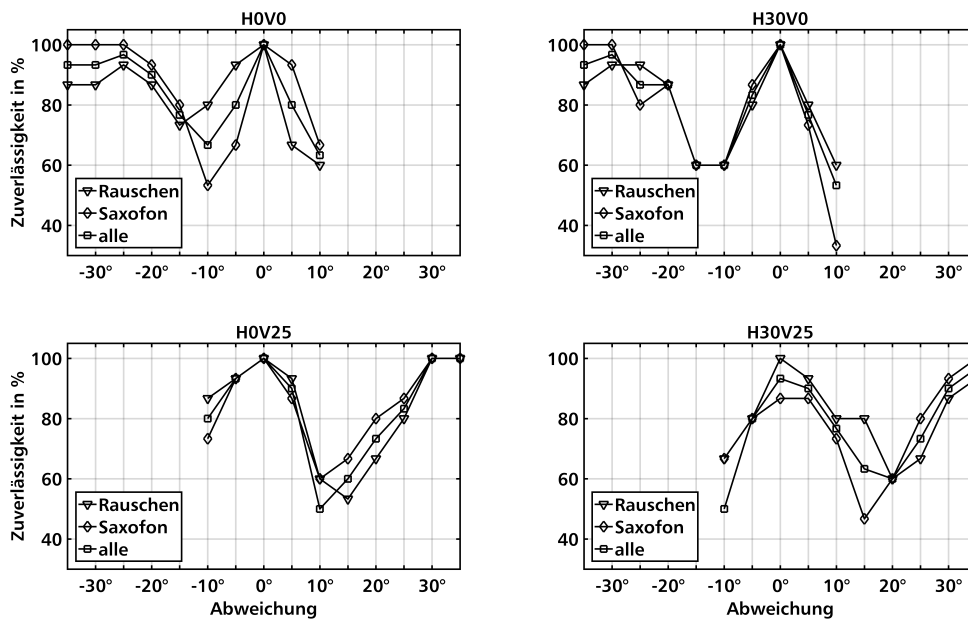


Abbildung 4.3: Zuverlässigkeit der Bewertungen aller Testpersonen für die akustischen Positionen H0V0, H30V0, H0V25, H30V25 und für die Testsignale; die Abweichung in Grad zwischen akustischen und optischen Stimulus ist auf der x-Achse abgetragen; eine positive Abweichung zeigt an, dass der akustische Stimulus über dem optischen Stimulus angeordnet ist.

#### 4.1.2 Exp. II: Zeigermethode

Zur Untersuchung des Ventriloquismus-Effekt wird eine neue Methode zur Bewertung der Richtungswahrnehmung verwendet. Die wahrgenommene Richtung des Hörereignisses soll durch eine Lichtzeigermethode angezeigt werden. Hierfür erhalten die Testpersonen einen Laserpointer, mit denen sie die Richtung auf einem Vorhang markieren können.

##### 4.1.2.1 Untersuchungsgegenstand

In Analogie zu Experiment I wird die wahrgenommene örtliche Diskrepanz in der Richtungswahrnehmung zwischen optischen und akustischen Reizen bei binauraler Kopfhörerwiedergabe gemessen. Die Messung erfolgt hier unter Beachtung der durch die Hörwahrnehmung ohne visuelle Einflüsse auftretenden Ungenauigkeiten. Die prinzipielle Gestaltung des Experimentes ist vergleichbar zu Experiment I. Die Hauptunterschiede liegen in einer neuen Anordnung der akustischen und optischen Quellen auf einer Ebene (siehe Abbildung 4.1 rechts), einer erhöhten Anzahl der optischen Quellen und in der Verwendung eine Zeigermethode.

Nachstehende Fragen sollen beantwortet werden: Welche Abweichung in Grad in der Wahrnehmung der vertikalen Richtung eines Hörereignisses liegt vor, wenn in der Richtung konkurrierende optische und akustische Stimuli präsentiert werden? Wie groß ist die Abweichung

abzüglich der Ungenauigkeiten in auditiven Richtungswahrnehmung ohne konkurrierende visuelle Stimuli?

#### 4.1.2.2 Versuchsaufbau

**Binauralsynthesystem** - Das Binauralsynthesystem entspricht in weiten Teilen dem in Experiment I beschriebenen System. Die Änderungen sind die Verwendung von Lautsprechern des Typs Genelec 8030BPM, die Verwendung von In-Ohr-Mikrofonen des Typs Svantek SV-25S zur Aufnahme der individuellen BRIRs und eine Variation in der Anordnung der Lautsprecher und LEDs. Die optischen und akustischen Quellen, die Zeigevorrichtung und die Aufzeichnung der Bewertungen werden durch MATLAB und durch eine MATLAB gesteuerte Arduino-Mega Plattform [50] kontrolliert.

**Quellpositionen** - Es werden vier Schallquellpositionen und 34 visuelle Quellpositionen in diesem Experiment untersucht. Die Positionen der Schallquellen sind in Tabelle 4.4 zusammengefasst.

Tabelle 4.4: Azimut und Elevation der virt. Schallquellpositionen aus Experiment II.

Name	H0V0	H20V0	H0V20	H20V20
Azimuth	0°	+20°	0°	+20°
Elevation	0°	0°	+20°	+20°

Die Entfernung zum Lautsprecher bei der Richtung H0V0 beträgt ca. 2,2 m. Die Höhe der Quellen bei 0° Elevation beträgt ca. 1,26 m. Siebzehn vertikale Positionen bei den horizontalen Richtungen 0° und 20° werden als optische Quellen (LEDs) verwendet. Sie bedecken einen Bereich von  $-10^\circ$  bis  $+30^\circ$  mit Schritten von  $2,5^\circ$ . Ein schwarzer schalltransparenter Vorhang ist direkt vor den LEDs platziert. Der Durchmesser der Lichtpunkte beträgt ca. 10 mm (ca.  $0,26^\circ$ ) auf der Vorderseite des Vorhangs. Alle Kombinationen zwischen akustischen und optischen vertikalen Richtungen für jede der beiden horizontalen Richtungen werden im Test verwendet.

**Audiosignale** - Zwei Audiosignale werden verwendet: Nachhallfreie Aufnahme eines männlichen Sprechers (4 s Dauer) und Rauschpulsfolge aus Experiment I. Die optischen und akustischen Stimuli werden mit einem zeitlichen Versatz präsentiert. Der Audiostimulus ist gegenüber dem optischen Stimulus um 150 ms verzögert. Diese Verzögerung rührt von technischen Begrenzungen bei der Präsentation der optischen und akustischen Stimuli sowie der Aufnahme der Bewertungen mit einer IP Kamera her. Aufgrund dieser Verzögerung wird eine verringerte wahrgenommene Fusion von auditivem und visuellem Stimulus erwartet, welches zu einem geringer ausgeprägten Ventriloquismus-Effekt führt [61].

### 4.1.2.3 Versuchsdurchführung

**Testpersonen** - Zwei weibliche und vier männliche normal hörende Testpersonen mit einem Alter von 21 bis 30 Jahren nahmen an dem Hörversuch teil. Die Teilnehmer sind erfahren in der Teilnahme an Wahrnehmungsexperimenten. In Übereinstimmung mit Experiment I absolvierten alle Testpersonen ein Training, um sich mit den Teststimuli, der Testprozedur und dem Eingabegerät vertraut zu machen, sowie um eine individuelle internen Referenz zu entwickeln. Die Training-Stimuli bestehen aus einer Auswahl von Test-Stimuli mit Übereinstimmung und Divergenz zwischen akustischer und optischer Quellposition und mit Test-Stimuli, die aus der akustischen Quelle alleine bestehen.

**Durchführung** Das Experiment besteht aus einem Hörtestdurchgang bei gleichzeitiger Präsentation der optischen und akustischen Quellen zur Untersuchung der Änderung in der Richtungswahrnehmung des Hörereignisses. Weiterhin wird die auditive Lokalisationsgenauigkeit für Hörereignisse in der Elevation ohne visuellen Einfluss gemessen. Das Experiment ist in drei Teile unterteilt. Teil eins beinhaltet das Training. Der zweite und dritte Teil umfasst zwei Wiederholungen der Test-Stimuli aller Kombinationen optischer und akustischer Quellpositionen in zufälliger Reihenfolge. Weiterhin werden die akustischen Positionen ohne visuelle Merkmale ebenfalls zweimal präsentiert. Zwischen den einzelnen Teilen wird eine Pause von ca. fünf Minuten eingefügt, um einer Erschöpfung der Testpersonen entgegenzuwirken.

Die Anzahl der Stimuli beträgt insgesamt 320 pro Testperson ( $2_{\text{Wiederholungen}} \times 2_{\text{Audiosignale}} \times 4_{\text{akustischePositionen}} \times 17_{\text{visuellePositionen}} = 272$  plus  $2_{\text{Wiederholungen}} \times 2_{\text{Audiosignale}} \times 4_{\text{akustischePositionen}} = 16$  plus 32 Training-Stimuli). Die Dauer eines Durchgangs beträgt ca. 60 Minuten.

Die Testteilnehmer bewerten das Hörereignis durch Anzeigen der wahrgenommenen Richtung mittels eines Laserpointers in ihrer rechten bzw. linken Hand auf einem schwarzen Vorhang. Der Vorhang ist direkt vor den LEDs angeordnet. Eine IP Kamera zeichnet die Bewertungen der Testpersonen auf. Ausgelöst wird die Kamera durch Betätigung eines Knopfes durch die Testperson. Die Testpersonen wurden darauf hingewiesen, dass sie den Kopf während der Bewertung still halten und nach vorne schauen sollten. Weiterhin sollten sie das Teststimulus-Paar komplett anhören und dann bewerten. Augenbewegungen sind erlaubt. Ein wiederholtes Abspielen des Stimuluspaares ist auf Anfrage möglich.

### 4.1.2.4 Ergebnisse

Zur Auswertung der Bewertungen wird ein Gitternetz mittels Videoprojektor auf den Vorhang projiziert und durch die im Test verwendete IP-Kamera aufgezeichnet. Das projizierte Gitter wurde so verzerrt, dass es die korrekten horizontalen und vertikalen Winkelangaben eines Kugelsegmentes mit seinem Zentrum in der Hörposition widerspiegelt. Die Winkelauflösung des Gitters beträgt  $1^\circ$ . Das Gitter wird einmalig projiziert und aufgezeichnet. Während der Tests

ist es für die Testpersonen nicht sichtbar. Der Lichtpunkt des Laserpointers wird durch ein Bildschirmfoto der IP-Kamera aufgezeichnet. Durch eine Bildanalyse werden die Koordinaten des Lichtpunktes auf das Gitternetz übertragen.

Die Quantile der Verteilung der Bewertungen aus dem Lokalisationstest mit Präsentation von visuellen Merkmalen (Testdurchgang oder „test trial“) werden mit den Bewertungen aus dem Lokalisationstest ohne Präsentation von visuellen Merkmalen (Kontrolldurchgang oder „control trial“) normalisiert. Der Einfluss der visuellen Merkmale auf die Lokalisierung des Hörereignisses (Abweichung) wird als Differenz der Mediane zwischen normalisierten Testergebnissen (normalisierter Testdurchgang oder „normalized test trial“) und dem Kontrolldurchgang („control trial“) für jede Audioposition berechnet. Weiterhin wird eine mittlere absolute Abweichung (mad) der Mediane über alle visuellen Richtungen und über einen Bereich von  $+10^\circ$  bis  $+30^\circ$  für die V0 Kondition und über einen Bereich von  $-10^\circ$  bis  $+10^\circ$  für die V20 Kondition berechnet. Die Festlegung der Grenzen ist durch die Ergebnisse aus Experiment I für den 50% Punkt der Bewertungen motiviert. Signifikante Ergebnisse (einseitiger Vorzeichenrangtest,  $p < 0,05$ ) für die Hypothese einer  $0^\circ$  Abweichung sind durch Kreuz-Symbole in Abbildungen 4.4 gekennzeichnet.

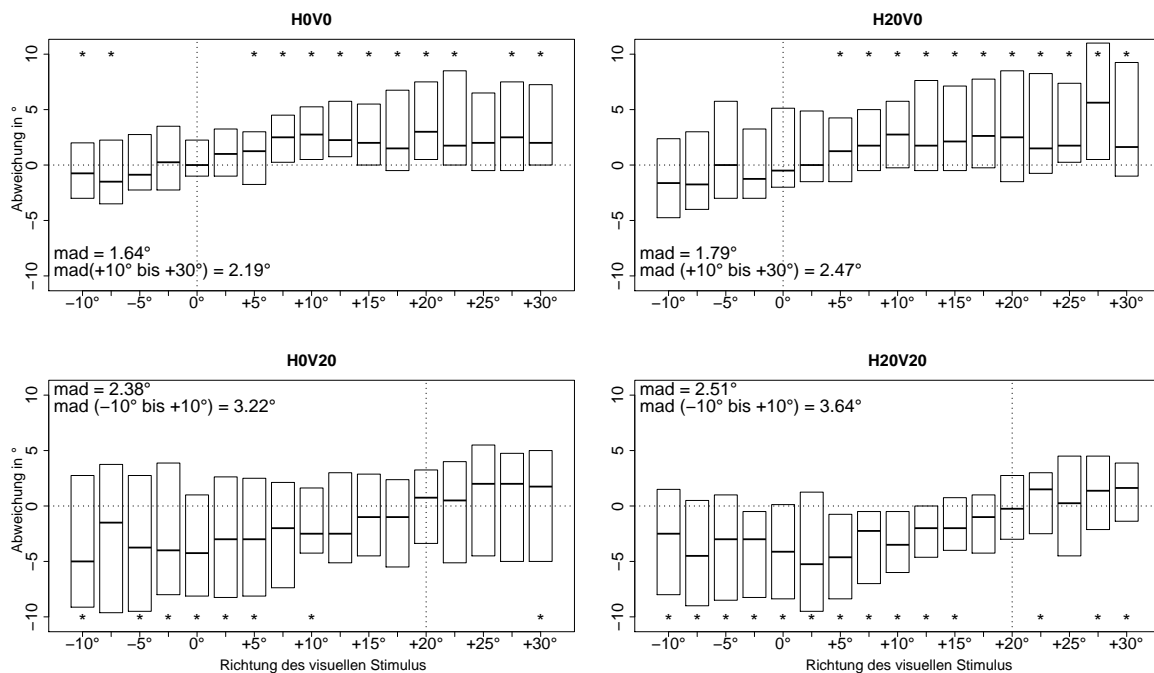


Abbildung 4.4: Vertikale Abweichung in Grad für die Konditionen H0V0, H20V0, H0V20, H20V20 bezogen auf die Richtung des visuellen Stimulus; mad = mittlere absolute Abweichung; \*  $p < 0,05$  bei einseitigem Vorzeichenrangtest.

Abbildung 4.4 zeigt die vertikalen Abweichungen für die Konditionen H0V0, H20V0 sowie H0V20 und H20V200 unter dem Einfluss der Präsentation von visuellen Stimuli. Signifikante vertikale Abweichungen bei virtuellen Schallquellen in der Horizontalebene sind für Richtungen

des visuellen Stimulus größer gleich  $+5^\circ$  (ausgenommen  $+25^\circ$  bei H0V0) und kleiner gleich  $-7,5^\circ$  für H0V0 zu erkennen. Die mittlere absolute Abweichung (mad) steigt leicht für die seitliche Kondition (H20V0) an.

Für die nach oben ausgelenkten virtuelle Schallquellen ergeben sich signifikante vertikale Abweichungen für alle visuellen Richtungen kleiner gleich  $+15^\circ$  und größer gleich  $+22,5^\circ$  (ausgenommen  $+25^\circ$  für H20V20). Die Kondition H0V20 zeigt denselben Trend aber mit nicht signifikanten Ergebnissen für einige Richtungen. Die mittlere absolute Abweichung steigt für die obere seitliche Kondition (H20V20) an. Ein starker Anstieg ist zwischen der frontalen und oberen Bedingung beobachtbar.

Weiterhin wird ein intersensorischer Bias aus der Division der Mediane der Abweichungen und der intersensorischen Richtungsdiskrepanz zwischen akustischem und optischem Stimulus berechnet. Der Bias stellt einen direkten Bias mit einem minimalen Einfluss von Adaptions-effekten dar [61]. Abbildung 4.5 zeigt den intersensorischen Bias für die vier untersuchten Konditionen.

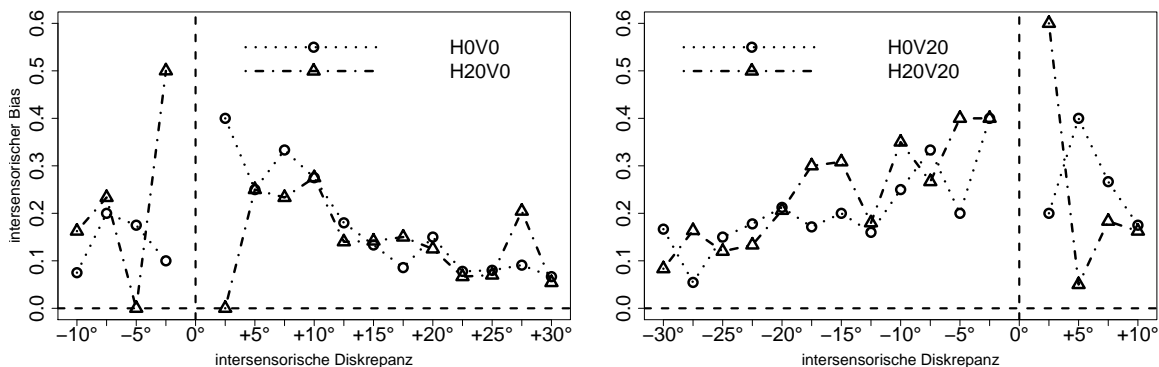


Abbildung 4.5: Intersensorischer Bias für die Konditionen H0V0 und H20V0 (links), sowie H0V20 und H20V20 (rechts).

Der beobachtete Bias ist konsistent mit Ergebnissen aus der Literatur [61, 190] für intersensorische Diskrepanzen in der Horizontalebene für reale Schallquellen und binaural synthetisierte Quellen. Eine Schiefelage kann zwischen positiven und negativen Diskrepanzen beobachtet werden. Dies wird nicht in Experimenten für die Horizontalebene berichtet. Weiterhin wird ein geringfügig höherer Bias für die V20 Kondition festgestellt.

#### 4.1.3 Zusammenfassung

Es wurden zwei Experimente zur Bestimmung von psychometrischen Funktionen und des intersensorischen Bias bei konkurrierender Darbietung von auditiven und visuellen Stimuli vorgestellt. Es wurde der Ventriloquismus-Effekt für vertikal ausgelenkte Positionen in der Medianebene und für seitliche vertikale Richtungen untersucht. Für eine plausible Simulation eines realen Lautsprechers im Raum wurde eine individualisierte Binauralsynthese über

Kopfhörer verwendet. Die Ergebnisse aus Experiment I deuten darauf hin, dass eine größere Richtungsdiskrepanz zwischen akustischem und optischem Stimuli bei vertikal und seitlich vertikal ausgelenkten Quellpositionen für die Testpersonen tolerabel sind, ohne dass die wahrgenommene Fusion zwischen den beiden Stimuli gestört wird. Die Abweichungen betragen ca.  $8^\circ$  für nicht vertikal ausgelenkte Positionen und ca.  $17^\circ$  für seitlich vertikal ausgelenkte Quellpositionen. Diese Abweichungen sind allerdings zusätzlich durch die Lokalisationsgenauigkeit ohne visuelle Merkmale beeinflusst. Es wurde daher ein zweites Experiment durchgeführt, indem dieser Tatsache durch eine veränderte Versuchsgestaltung Rechnung getragen wurde. Für die einzelnen untersuchten Quellpositionen ergeben sich mittlere Abweichungen von  $2,2^\circ$  und  $2,5^\circ$  für frontal und seitlich aber nicht vertikal ausgelenkte Schallquellpositionen (H0V0 und H20V0), sowie  $3,2^\circ$  und  $3,6^\circ$  für zusätzlich vertikal ausgelenkte Schallquellpositionen (H0V20 bzw. H20V20). Die Ergebnisse von Experiment II zeigen, dass die beobachteten mittleren Abweichungen in der Lokalisation der auditiven Quellen geringfügig kleiner, als die in früheren Experimenten berichteten Abweichungen für die Horizontalebene sind [61, 190]. Diese Beobachtung lässt sich durch eine verringerte Fusion zwischen auditivem und visuellem Stimulus, verursacht durch einen zeitlichen Versatz von 150 ms, erklären. Dennoch kann deutlich gezeigt werden, dass der für eine individualisierbare Binauralsynthese über Kopfhörer gemessene Ventriloquismus-Effekt für vertikale Quellpositionen eine vergleichbare Größenordnung, wie in der Horizontalebene bei virtuellen und realen Umgebungen aufweist.

## 4.2 Externalisierung und Hörereignisrichtung

Der in der Literatur beschriebene Raumdivergenzeffekt [168, 212] soll bei Verwendung eines Binauralsynthesystems und bei Beachtung kontextabhängiger Qualitätsparameter untersucht werden. Die untersuchten Qualitätsmerkmale sind beeinflusst durch eine Divergenz zwischen synthetisierter Szene und Abhörraum, der Methode der Personalisierung der Binauralsynthese und durch die Sichtbarkeit des Abhörtraumes. Die Ergebnisse zeigen, dass eine Raumdivergenz zu einem Abfall der wahrgenommenen Externalisierung der Hörereignisse führt. Raumkonvergenz führt dahingegen zu einem Anstieg der Externalität. Weiterführende Informationen sind den dazugehörigen Veröffentlichungen zu entnehmen [19, 20, 22, 23, 31].

### 4.2.1 Exp. III: Raumdivergenzeffekt

Das nachstehende Experiment untersucht kontextabhängige Qualitätsparameter innerhalb einer Binauralsynthese. Es wird eine binaurale Simulation einer akustischen Szene über ein personalisierbares Kopfhörersystem verwendet. Die untersuchten Qualitätsparameter sind Divergenz und Konvergenz zwischen synthetisierter Szene und Abhörraum, Sichtbarkeit der Szene und Personalisierung der Binauralsynthese.

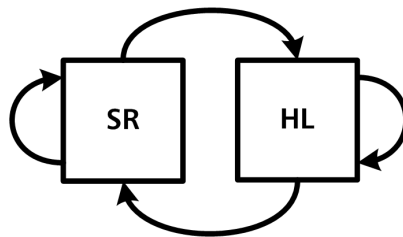


Abbildung 4.6: Kombinationen von Übereinstimmung und Divergenz zwischen synthetisiertem Raum und Abhörraum; SR=Seminarraum, HL=Hörlabor.

Zwei Räume mit unterschiedlichen akustischen Parametern werden sowohl als Aufnahme- als auch als Abhörraum verwendet. Die Testpersonen hören somit entweder denselben Raum oder den jeweiligen anderen Raum. Die wahrgenommene Plausibilität der auditorischen Szene wird mithilfe der Qualitätsmerkmale Externalität und wahrgenommene Schalleinfallrichtung beschrieben. In eigenen vorangegangenen Untersuchungen [31] konnte dieser Einfluss bereits nachgewiesen werden. Eine Auswertung hinsichtlich des Einflusses der Sichtbarkeit des Abhörtraumes und der Schallquellpositionen auf die Qualitätsmerkmale ist allerdings aufgrund der damaligen Testgestaltung nicht möglich. In dem hier vorgestellten Experiment werden zwei Gruppen von Testpersonen gebildet. Die erste Gruppe hat bei der Wiedergabe keine visuellen Merkmale (dunkler Abhörraum) zur Verfügung, während die zweite Gruppe die synthetisierten Schallquellpositionen und den Abhörraum sieht. Die Auswertung der Ergebnisse zeigt, dass signifikante Unterschiede ( $p < 0,05$ ) in der wahrgenommenen Externalität und Schalleinfallrichtung in Abhängigkeit des synthetisierten Raumes und Abhörtraumes, zwischen den beiden Probandengruppen und der Personalisierung der Binauralsynthese vorliegen.



#### 4.2.1.1 Untersuchungsgegenstand

Die Hörversuche evaluieren kontextabhängige Qualitätsparameter und ihren Einfluss auf die wahrgenommene Externalität und Schalleinfallrichtung von Hörereignissen. Das Experiment untersucht Kombinationen von Abhörraum und synthetisiertem Raum, Sichtbarkeit des Abhörraumes und der zu synthetisierenden Schallquellpositionen sowie unterschiedliche Personalisierung der Binauralsynthese.

Es wird untersucht, inwieweit die Verwendung individueller BRIRs gegenüber Kunstkopf-BRIRs zu einem Anstieg der Externalisierung der Hörereignisse führt. Weiterhin soll nachgewiesen werden, inwieweit eine Übereinstimmung des Aufnahme- und Wiedergaberaumes zu einem Anstieg der Externalität gegenüber keiner Übereinstimmung führt. Neben der Externalisierung werden Lokalisationsfehler in der Richtungswahrnehmung bei den verschiedenen Testkonditionen untersucht. Es werden alle Kombinationen von Abhörraum, synthetisierten Raum und Personalisierung für zwei Räume betrachtet.

Nachstehende Fragen sollen beantwortet werden: In welchem Umfang ändert sich die Externalisierung der Hörereignisse zwischen Testkonditionen mit Divergenz und Konvergenz zwischen Abhörraum und synthetisiertem Raum? In welchem Umfang ändert sich die Externalisierung der Hörereignisse zwischen individualisierter Binauralsynthese und Verwendung von Kunstkopf BRIRs in Situation der Raumdivergenz und Raumkonvergenz? In welchem Umfang ändert sich die Externalisierung der Hörereignisse zwischen Testkonditionen mit Sichtbarkeit und Nicht-Sichtbarkeit des Abhörraumes und der Schallereignispositionen in Situation der Raumdivergenz und Raumkonvergenz? Wirken die kontextabhängigen Qualitätsparameter Raumdivergenz, Personalisierung und visuelle Merkmale unabhängig voneinander?

#### 4.2.1.2 Versuchsaufbau

**Binauralsynthesystem** - Für die Kopfhörerwiedergabe werden individuelle und mithilfe eines KEMAR Kopf- und Torsosimulators aufgezeichneten BRIRs für die ausgewählten Räume, Schallquellen und Positionen verwendet. Das statische binaurale System ist personalisierbar, um Verwechslungsfehler in der Richtungswahrnehmung zu reduzieren. Weiterhin soll so die Wiedergabe der Simulationen im Vergleich zu echten Lautsprechern erhöht werden. Die Räume sind so gewählt, dass sie unterschiedliche raumakustische Eigenschaften wie Nachhall und Quellen-Empfänger-Abstände aufweisen. Der Nachhall soll die Externalisierung der Hörereignisse und die Wahrnehmung von Distanz unterstützen. Die Kopfhörer werden über individuelle Kopfhörer-Übertragungsfunktionen (HPTFs) entzerrt, wenn individuelle BRIRs verwendet werden. HPTFs des Kunstkopfes werden verwendet, wenn künstliche BRIRs verwendet werden. In-Ear-Mikrofone werden zur Messung individueller BRIRs und HPTFs am Eingang des blockierten Gehörgangs jeder Testperson verwendet. Die Mikrofone werden zwischen den BRIR- und HPTF-Messungen nicht entfernt. Die Invertierung einer HPTF wird nach der Methode der kleinsten Quadrate berechnet. Für die Wiedergabe werden je

nach Experiment ein Stax Lambda Pro Kopfhörer oder ein BK211 extra-auraler Kopfhörer verwendet, die die Anforderungen an offene Kopfhörer erfüllen. Die Verzerrung des Schalleinfalls von den realen Quellen im Raum zu den Ohren des Hörers wird speziell bei Verwendung extra-auraler Kopfhörer minimiert. Dies ermöglicht ein Testdesign mit Präsentation realer und virtueller Schallquellen im Raum.

**Quellpositionen** - Der Einfluss auf die Qualitätsmerkmale wird für fünf Schallquellpositionen in der Horizontalebene untersucht. Abbildung 4.7 gibt eine Übersicht über diese Positionen. Es sollen vor allem mögliche Fehler in der Richtungswahrnehmung und Wahrnehmung von Externalität bei binauraler Synthese untersucht werden. Diese umfassen Vorne-Hinten-Vertauschungen ( $0^\circ$  und  $180^\circ$  Richtung), Cone-of-Confusion-Fehler ( $330^\circ$  und  $120^\circ$ ), Lokalisationsgenauigkeiten (alle Positionen) und Externalisierung der Hörereignisse (alle Positionen).

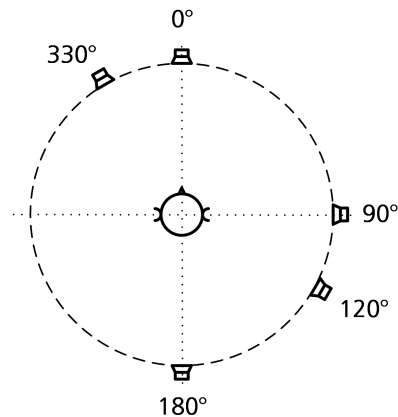


Abbildung 4.7: Schallquellpositionen für die Wiedergabe über Lautsprecher und Binaural-synthese über Kopfhörer; der Abstand der Schallquellen zum Kopfmittelpunkt beträgt 2,2 m.

Genelec 1030A Lautsprecher werden zur Messung der BRIRs für alle Positionen verwendet. Der Abstand der Lautsprecher zum Abhörpunkt beträgt ca. 2,2 m. Die Höhe der Quellpositionen beträgt ca. 1,3 m (in etwa Ohrposition einer sitzenden Person). Die BRIRs werden für jede Position und jede Testperson in den beiden Räumen gemessen. Die Aufnahmeposition im Raum ist dabei nahezu identisch mit der Abhörposition im Test.

**Kombinationen von Aufnahmeräumen und Abhörräumen** - Ein Hörlabor (HL; siehe Kapitel 2.5) und ein leerer Seminarraum (SR; siehe Kapitel 2.5) mit unterschiedlichen raumakustischen Eigenschaften werden für den Hörversuch und zur Messung der BRIRs verwendet. Die Hörversuche werden im selben Hörlabor (HL) und im selben Seminarraum (SR) durchgeführt. Abbildung 4.6 verdeutlicht die Kombinationen von Abhörraum und synthetisiertem Raum im Experiment.

**Sichtbarkeit des Abhörtraumes und der Schallquellpositionen** - Die Testpersonen werden gemäß zufälliger Auswahl in zwei Gruppen unterteilt. Für die erste Gruppe wird die Beleuchtung des Abhörtraumes während des Hörversuchs minimiert (nahezu komplette Dunkelheit). Ein schalltransparenter schwarzer Vorhang wird in einem Abstand von ca. 2,2 m um die Testpersonen positioniert. Die Testpersonen sollen keinen visuellen Eindruck und visuelle Merkmale vom Abhörtraum im Hörversuch haben. Die Testpersonen in der zweiten Gruppe werden im beleuchteten Abhörtraum platziert. Zusätzlich werden Lautsprecher als Attrappen an jeder Stundenposition eines Uhrkreises als zusätzliche visuelle Merkmale für die zu synthetisierenden Schallquellpositionen platziert.

**Personalisierung** - Es werden BRIRs für jede zu synthetisierende Schallquellenposition im Raum mittels Kunstkopf (KEMAR Kopf und Torsosimulator 45BA) und individuelle BRIRs für jeden Probanden gemessen. Die individuellen BRIRs werden mittels Sondenmikrofone am geblockten Ohrkanal in beiden Räumen gemessen. Zusätzlich werden für die individuelle Binauralsynthese individuell gemessene und invertierte HPTFs zur Kompensation des Kopfhörersystems verwendet. Bei nicht individueller Synthese kommen invertierte HPTFs des Kunstkopfes zur Anwendung. Weiterhin werden Freifeld-Kunstkopf-HRIRs [128] aus den fünf Richtungen im Test mit abgefragt. Diese dienen als Anker, da bei den so synthetisierten Stimuli keinerlei Raumanteil enthalten ist. Es ist zu erwarten, dass die wahrgenommene Externalität bei diesen Stimuli geringer als bei den Stimuli mit Raumanteil ausfällt [57, 136].

**Audiosignale** - Es kommen nachhallfreie breitbandige Rauschsignalpulse und ein Sprachsignal (männlicher Sprecher) zur Anwendung. Die Rauschsignalpulsfolgen bestehen aus fünf aufeinander folgenden weißen Rauschsignalen mit je 30 ms Dauer und 3 ms Ein- und Ausblendung mit einem Cosinus-Fenster. Die Pulse folgen in einem Abstand von 70 ms.

**Testitems** - Die zu untersuchenden Teststimuli im Experiment ergeben sich aus:

- 5 Richtungen: 0°, 90°, 120°, 180°, 330° (Orientierung im Uhrzeigersinn)
- 2 Signaltypen: Sprache, Rauschpulsfolge
- 4 Raumkombinationen: HL in HL, HL in SR, SR in HL, SR in SR
- 3 BRIRs: individuelle BRIR, Kunstkopf-BRIR, Kunstkopf-Freifeld-HRIR

Für die Abfrage aller Kombinationen der genannten Variablen ergeben sich 5 Richtungen x 2 Signale x 4 Raumkombination x 3 BRIRs = 120 Faktorstufen für die Bewertung je Proband. Das Experiment stellt ein vierfaktorielles Versuchsdesign mit Messwiederholung dar.

#### 4.2.1.3 Versuchsdurchführung

**Stichprobenumfang** - Für die Abschätzung der nötigen Stichprobengröße werden das Signifikanzniveau mit 0,05 und die Teststärke mit 0,95 festgelegt. Es sollen kleine Effekte gemessen werden. Durch die Messwiederholung ergibt sich eine höhere Empfindlichkeit des Tests ge-

genüber den vermuteten Effekten. Dies spiegelt sich in einer korrigierten Effektgröße und somit einem geringeren Stichprobenumfang wider. Bei der Annahme einer Effektgröße von  $\epsilon = 0,05$  und einer Korrelation zwischen den Wiederholungen von  $\rho = 0,25$  ergibt sich nach der Beziehung  $\epsilon = \epsilon/\sqrt{1-\rho}$  eine korrigierte Effektgröße von  $\epsilon = 0,06$ . Die Freiheitsgrade der Untersuchung ergeben sich aus den theoretisch maximal vorkommenden Kombinationen von Richtung (5 Stufen), Signal (2 Stufen), Raumkombinationen (4 Stufen) und BRIR-Art (3 Stufen) zu  $df = (5 - 1) \cdot (2 - 1) \cdot (4 - 1) \cdot (3 - 1) = 24$ . Der mithilfe des Statistik-Programms G-Power [89, 171] geschätzte optimale Stichprobenumfang für einen exakten Binomialtest für das dichotome Merkmal Externalität wird anschließend mit der Anzahl an Faktorstufen dividiert. Es ergibt sich eine Stichprobengröße von acht Testpersonen ( $902(\text{Stichprobenumfang nach G-Power})/120(\text{Faktorstufen})$ ). Für die Bewertung der Varianz der wahrgenommenen Schalleinfallrichtung als Merkmale ergibt sich ein optimaler Stichprobenumfang von 23 Testpersonen ( $2727(\text{Stichprobenumfang nach G-Power})/120(\text{Faktorstufen})$ ).

**Bewertungsskala** - In beiden Tests werden die Hörereignisrichtung und Externalität abgefragt. Die Hörereignisrichtung wird über eine nicht-diskrete Plantechnik abgefragt. Die Externalität wird mithilfe einer dreistufigen Skala durch die Probanden bewertet. Die Eingabe erfolgt über einen berührungssensitiven Bildschirm mit grafischer Benutzeroberfläche (Abbildung 4.8). Die in der Abbildung dargestellten Schieberegler zur Bewertung der Breite und Elevation des Hörereignisses werden im Hörversuch nicht benutzt.

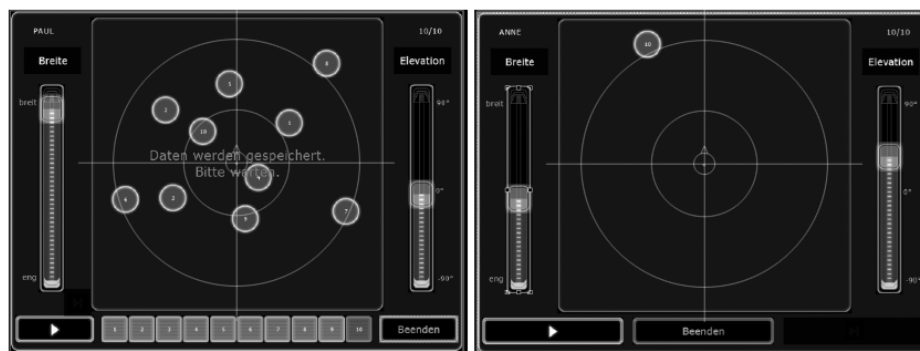


Abbildung 4.8: Eingabeoberflächen des berührungssensitiven Bildschirms mit beispielhaften Bewertungen zweier Probanden; links=Mehrfachbewertung während der Phase I des Haupttestes, rechts=Einzelbewertung während des Vortests und Phase II des Haupttestes; die Regler an den Seiten (Breite, Elevation) werden in dem Hörversuch nicht verwendet.

**Durchführung Vortest** - Der Vortest umfasst die Bewertung der wahrgenommenen Schalleinfallrichtung und Externalität der Hörereignisse. Hierbei kommen die Wiedergabe von Schallereignissen mittels Intensitätsstereopanning und die binaurale Synthese der Lautsprecherpositionen mittels Kunstkopf-BRIRs zur Anwendung. Der Abhörraum und der wiederzugebende Raum stimmen hierbei überein. Lautsprecherpositionen und der Abhörraum sind den

Testpersonen sichtbar. Der Vortest wird unmittelbar nach den Messungen der individuellen BRIRs und HPIRs durchgeführt.

Die Testpersonen können ihre Bewertung gemäß der in Abbildung 4.8 links dargestellten Benutzeroberfläche in einem Multi-Stimulus-Test abgeben. Die Anzahl der Stimuli ergeben sich aus den fünf wiederzugebenden Positionen, der zwei Audiosignaltypen und der Wiedergabe über Kopfhörer für Stereo-Signale und Verwendung von Kunstkopf-BRIRs. Jedes Testitem wird zweimal wiederholt. Es ergeben sich somit 40 Teststimuli pro Testperson.

Die Bewertungen dienen der Einschätzung der einzelnen Testpersonen und der gesamten Probandengruppe hinsichtlich ihrer Lokalisationsfähigkeit und der Wahrnehmung von Externalität bei nicht-individuell synthetisierten Schallquellen. Weiterhin wird den Testpersonen die Hörtestmethodik des Haupttestes bekannt gemacht.

**Durchführung Haupttest** - Es werden die wahrgenommene Schalleinfallrichtung und Externalität der dargebotenen Hörereignisse bewertet. Die Wiedergabe erfolgt über Binauralsynthese unter Verwendung von Kunstkopf-BRIRs und individuellen BRIRs. Den Testpersonen werden Hörereignisse dargeboten, bei denen Raumdivergenz oder Raumkonvergenz vorliegt. Der Haupttest wird sowohl im Raum HL als auch im Raum SR durchgeführt.

Die Ergebnisse umfassen Aussagen über den Einfluss der Personalisierung der Binauralsynthese, der Divergenz zwischen Abhörraum und wiederzugebendem Raum, den Einfluss einer visuellen Referenz (Sichtbarkeit der Lautsprecher) und den Einfluss der Sichtbarkeit des Abhörraumes auf die Lokalisationsfähigkeit und Externalität der Hörereignisse.

Der Haupttest selber umfasst zwei Phasen. In der Phase I werden alle Schallquellpositionen und Audiosignaltypen bei individualisierter und nicht-individualisierter Synthese und für alle Kombinationen der Divergenzen zwischen Abhörraum und wiederzugebendem Raum für die fünf Richtungen präsentiert. Die Stimuli-Anzahl für Phase I ergibt sich somit aus den fünf Richtungen, zwei Signaltypen, zwei BRIR-Typen und zwei Räumen zu 40 ( $5_{\text{Richtungen}} \times 2_{\text{Signale}} \times 2_{\text{Räume}} \times 2_{\text{BRIR}} = 40$ ). Die Testpersonen können ihre Bewertungen über die in Abbildung 4.8 links dargestellte Oberfläche abgeben. Den Probanden wird es ermöglicht, die Teststimuli wiederholt anzuhören und relativ miteinander zu vergleichen. Pro Durchgang (Oberfläche) werden zehn Stimuli präsentiert. Insgesamt ergeben sich für Phase I vier Durchgänge, die die Stimuli der beiden Räume für jeweils einen BRIR-Typ und jeweils eine Audiosignaltyp enthalten ( $5_{\text{Richtungen}} \times 1_{\text{Signal}} \times 2_{\text{Räume}} \times 1_{\text{BRIR}} = 10$ ). Es werden somit pro Durchgang die Räume miteinander verglichen.

In der sich nach einer kurzen Pause anschließenden Phase II werden die Teststimuli einzeln hintereinander in zufälliger Reihenfolge präsentiert (Abb. 4.8 rechts). Ein erneutes Anhören des aktuellen Stimulus ist möglich. Die Testperson kann hier absolute Bewertungen abgeben. Für Phase II ergibt sich eine Stimuli-Anzahl aus den fünf Positionen, zwei Audiosignalarten, zwei

wiedergehenden Räumen, drei BRIR-Typen und zwei Wiederholungen zu 120 ( $5_{\text{Richtungen}} \times 2_{\text{Signale}} \times 2_{\text{Räume}} \times 3_{\text{BRIR}} \times 2_{\text{Wiederholung}} = 120$ ).

Das Ziel der Unterteilung des Haupttestes in zwei Phasen ist es, die Testpersonen auf die Teststimuli in Phase I zu trainieren. Weiterhin soll ein Vergleich der Bewertungsmethodiken zwischen Phase I und Phase II ermöglicht werden. In Phase I werden relative Bewertungen zu den Stimuli abgegeben, während in Phase II absolute Bewertungen abgegeben werden.

Der Hörversuch wird im HL und im SR in getrennten Durchgängen an verschiedenen Tagen durchgeführt. In jedem Durchgang werden den Testpersonen individuell synthetisierte und mittels Kunstkopf synthetisierte Schallquellpositionen aus beiden Aufnahmeräumen vorgespielt. Die Stimuli werden zweimal in zufälliger Reihenfolge präsentiert. Der wahrgenommene Schalleinfallswinkel wird durch Wahl der entsprechenden Richtung gemäß Abbildung 4.8 in einer Draufsicht bewertet. Die wahrgenommene Externalität des Hörereignisses wird durch Auswahl des Mittelpunktes, inneren Kreise bzw. des äußeren Kreises gemäß Abbildung 4.8 bewertet. Die Skala des Qualitätsmerkmals Externalität orientiert sich an Definitionen von Hartmann und Wittenberg [105]. Folgende Definitionen werden im Hörversuch verwendet: a) Mittelpunkt: „Das Hörereignis ist vollständig in meinem Kopf oder es ist sehr diffus.“, b) innerer Kreis: „das Hörereignis ist extern aber sehr nahe an meinen Ohren oder Kopf.“, c) äußerer Kreis: „Das Hörereignis ist extern und gut lokalisierbar.“ Weitere Ausführungen hierzu sind Kapitel 3.1.3 zu entnehmen.

Am Hörversuch nehmen 21 Testpersonen teil, welches zu einer Teststärke von 0,93 führt. Die Testpersonen sind erfahren in der Teilnahme an Hörversuchen und werden trainiert. Das Training besteht aus einer mündlichen und schriftlichen Einführung und einer Definition der Qualitätsmerkmale Lokalisation und Externalität. Weiterhin hört sich jede Testperson die Testitems in Phase I des Haupttestes an. Die Testperson kann jedes Item mit den anderen Items gemäß oben beschriebener Multistimulus-Methode vergleichen und bewerten. Die Testperson soll somit auch eine interne Referenz und eine eigene Definition zwischen den Items für die Merkmale Lokalisation und Externalität erstellen.

#### 4.2.1.4 Ergebnisse

Die Bewertungen für das Merkmal Externalität werden als Häufigkeiten gezählt. Die Verteilung der Häufigkeiten zeigt keine signifikanten ( $p < 0,05$ ) Unterschiede zwischen den beiden verwendeten Audiosignalen. Die Bewertungen beider Signale werden für die weitere Analyse zusammengefasst. Der Externalitätsindex berechnet sich aus dem Verhältnis der Bewertungen für „extern“ (äußerer Kreis auf der Bewertungsoberfläche) und aller abgegebenen Bewertungen. Ein Index von „0“ bedeutet vollständige In-Kopf-Lokalisation. Ein Index von „1“ deutet auf Lokalisation des Hörereignisses außerhalb des Kopfes hin. Die Bewertungen der Richtungswahrnehmung werden über die Auswertung der Häufigkeit der Quadrantenfehler dargestellt.

**Lokalisation** - Zur Auswertung der wahrgenommenen Hörereignisrichtungen werden lediglich die Bewertungen derjenigen Stimuli verwendet, die von den Testpersonen als „extern“ bewertet wurden. Zum einen ist aufgrund der Eigenschaften der Bewertungsoberfläche für die Testpersonen eine präzise Richtungseingabe für einen als „intern“ wahrgenommenen Stimuli nur eingeschränkt möglich. Zum anderen sollen lediglich die Bewertungen der Stimuli ausgewertet werden, die auch als „extern“ und somit für dieses Qualitätsmerkmal als plausibel beschrieben werden können. Diese Auswahl führt dazu, dass für die kritischen Richtungen  $0^\circ$  und  $180^\circ$  zum Teil deutlich weniger Bewertungen als für die andere Richtungen vorliegen.

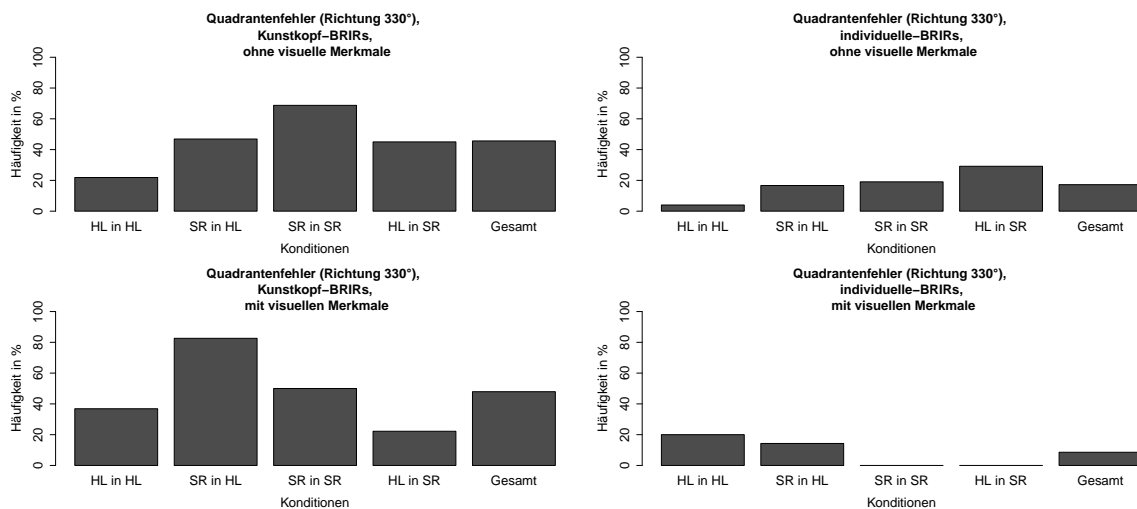


Abbildung 4.9: Quadrantenfehler für die Bewertungen der Stimuli aus der  $330^\circ$  Richtung als Häufigkeiten für die untersuchten Raumkombinationen, Personalisierungsmethoden und Vorhandensein von zusätzlichen visuellen Merkmalen.

Abbildung 4.9 verdeutlicht die in den Bewertungen auftretenden Quadrantenfehler. Ein Quadrantenfehler wird hierbei für die  $330^\circ$  Richtung angenommen, wenn die Winkelangabe der Bewertungen kleiner als  $270^\circ$  und größer  $180^\circ$  beträgt. Es werden somit Fehler in der Lokalisation erfasst, die auf Vorne-Hinten-Vertauschungen zurückzuführen sind. Es zeigen sich hohe Fehler bei Verwendung von Kunstkopf-BRIRs gegenüber individuellen BRIRs. Bei der Testbedingung ohne zusätzliche visuelle Merkmale ergibt sich für die Verwendung von Kunstkopf-BRIRs ein über alle Raumkombinationen gemittelter Quadrantenfehler von ca. 46%. Bei individuellen BRIRs zeigt sich ein Fehler von ca. 17%. Bei den Bewertungen der Probandengruppe mit zusätzlichen visuellen Merkmalen zeigt sich bei Verwendung von Kunstkopf-BRIRs keine Verringerung des gemittelten Fehlers (ca. 47%). Bei der Verwendung individueller BRIRs verringert sich der gemittelte Quadrantenfehler in den Bewertungen auf ca. 8%. Für den Seminarraum als Abhörraum ergeben sich Quadrantenfehler von 0%.

Abbildung 4.10 verdeutlicht die in den Bewertungen auftretenden Vorne-Hinten-Vertauschung. Eine Vorne-Hinten-Vertauschung wird hierbei für die  $0^\circ$  Richtung angenommen, wenn die Winkelangabe der Bewertungen größer  $90^\circ$  und kleiner  $270^\circ$  beträgt. Es werden somit Fehler in

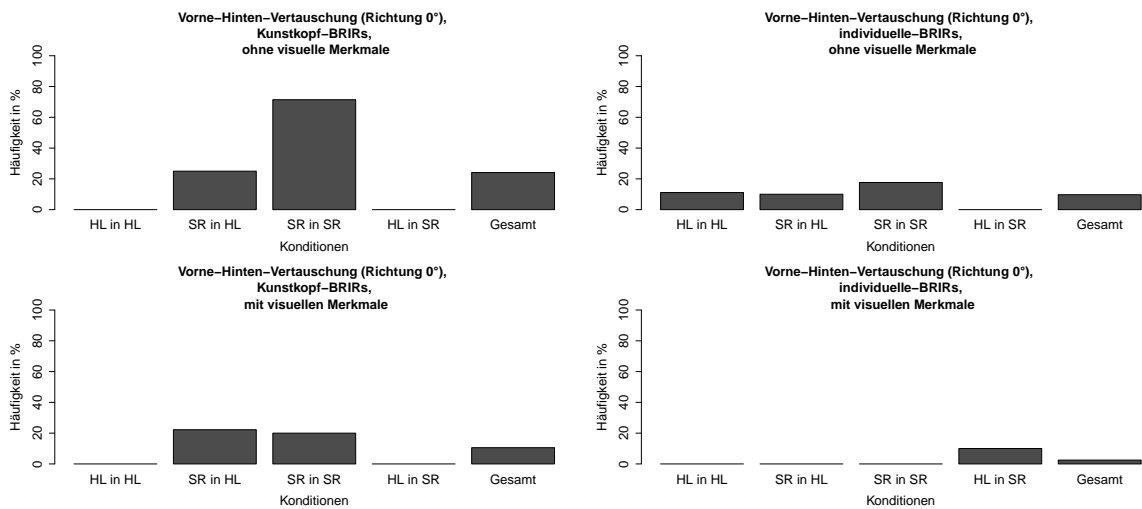


Abbildung 4.10: Vorne-Hinten-Vertauschung für die Bewertungen der Stimuli aus der 0° Richtung als Häufigkeiten für die untersuchten Raumkombinationen, Personalisierungsmethoden und Vorhandensein von zusätzlichen visuellen Merkmalen.

der Lokalisation erfasst, die auf einen Sonderfall der cone-of-confusion Fehler zurückzuführen sind, wenn sich das Schallobjekt direkt frontal befindet. Es zeigen sich hohe Vorne-Hinten-Vertauschungen bei Verwendung von Kunstkopf-BRIRs gegenüber individuellen BRIRs bei den Raumkombinationen „SR in HL“ und „SR in SR“. Zusätzlich lässt sich eine Verringerung der Anzahl an Vorne-Hinten-Vertauschungen bei Vorhandensein von zusätzlichen visuellen Merkmalen feststellen.

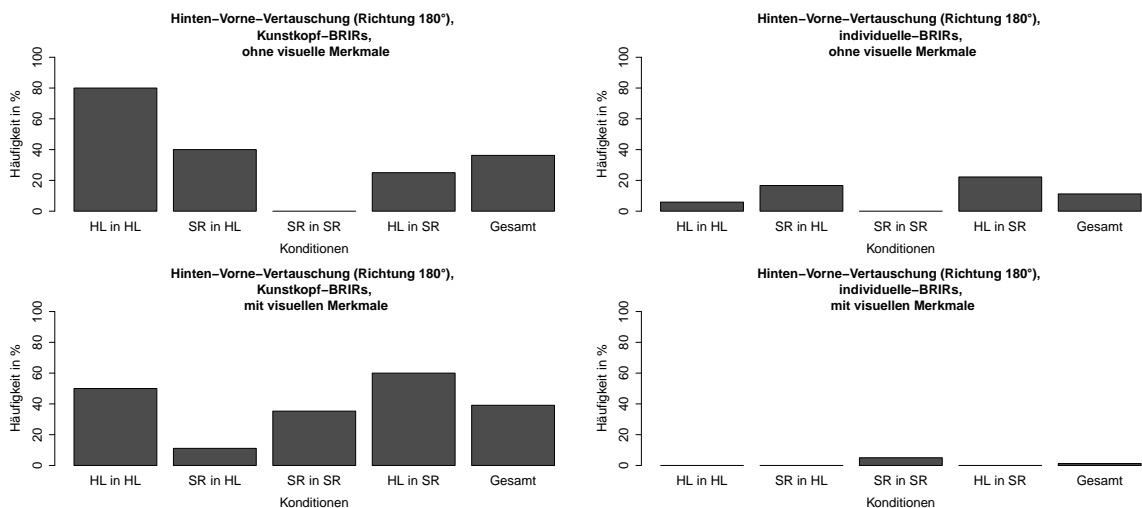


Abbildung 4.11: Hinten-Vorne-Vertauschung für die Bewertungen der Stimuli aus der 180° Richtung als Häufigkeiten für die untersuchten Raumkombinationen, Personalisierungsmethoden und Vorhandensein von zusätzlichen visuellen Merkmalen.



Abbildung 4.11 verdeutlicht die in den Bewertungen auftretenden Hinten-Vorne-Vertauschung. Eine Hinten-Vorne-Vertauschung wird hierbei für die  $180^\circ$  Richtung angenommen, wenn die Winkelangabe der Bewertungen kleiner  $90^\circ$  oder größer  $270^\circ$  beträgt. Es werden somit Fehler in der Lokalisation erfasst, die auf einen Sonderfall der cone-of-confusion Fehler zurückzuführen sind, wenn sich das Schallobjekt direkt hinter einem befindet. Es zeigen sich hohe Hinten-Vorne-Vertauschungen bei Verwendung von Kunstkopf-BRIRs gegenüber individuellen BRIRs bei allen Raumkombinationen. Eine Verringerung der Anzahl an Hinten-Vorne-Vertauschungen wird bei Vorhandensein von zusätzlichen visuellen Merkmalen und bei Verwendung von individuellen BRIRs festgestellt.

Nachstehende Tabelle 4.5 gibt die Signifikanz der Unterschiede zwischen den untersuchten Bedingungen für den Quadrantenfehler, Vorne-Hinten-Vertauschungen und Hinten-Vorne-Vertauschungen an. Als Signifikanztest kommt ein einseitiger exakter Test nach Barnhard [54] zur Anwendung. Der Test selber wird in der Statistiksoftware R durchgeführt. Die Implementierung kann unter nachstehendem Verweis nachgeschlagen [97] werden.

Tabelle 4.5: Signifikanz der Unterschiede in den Bewertungen zwischen der untersuchten Bedingung Kunstkopf-BRIR gegenüber individueller BRIRs für den Quadrantenfehler (QF), Vorne-Hinten-Vertauschungen (VH) und Hinten-Vorne-Vertauschungen (HV); Signifikanztest: einseitiger exakter Test nach Barnhard [54]; \*signifikanter Unterschied bei  $\alpha = 0,05$  wenn nicht anders in Klammern angegeben; HL=Hörlabor, SR=Seminarraum, ov=ohne zusätzliche visuelle Merkmale, mv=mit zusätzlichen visuellen Merkmalen.

Kondition	QF	QF	VH	VH	HV	HV
	ov	mv	ov	mv	ov	mv
HL in HL	*(0,1)	*	-	-	*	*
SR in HL	*(0,1)	*	-	*(0,1)	-	-
SR in SR	*	*	*	*	-	*
HL in SR	-	*(0,1)	-	-	-	*

Die festgestellte Häufigkeit eines Quadrantenfehler bei Verwendung einer Binauralsynthese mit individuellen BRIRs liegt in einer vergleichbaren Größenordnung wie in vorangegangenen Untersuchungen. Wenzel et al. [227, 228], sowie Wightman und Kistler [231], ermittelten in ihren Experimenten für erfahrene Hörer durchschnittlich 11% Vorne-Hinten-Vertauschungen bei Verwendung einer individualisierten Binauralsynthese (Freifeld-HRTFs) ohne Einbeziehung von Kopfbewegungen. Bei Betrachtung der Bewertungen von nicht erfahrenen Hörern ermittelten sie Vertauschungen von bis zu 31%. Begault [57] ermittelte Vorne-Hinten- und Hinten-Vorne-Vertauschungen von 59% für eine Binauralsynthese ohne Head-Tracking. Møller et al. ermittelten einen „Within Cone“ Fehler (vgl. Quadrantenfehler) von 1,5% bei individueller Binauralsynthese, Sichtbarkeit des Raumes, Übereinstimmung von Abhörraum und zu

synthetisierendem Raum und ohne Einbeziehung von Kopfdrehungen. Im hier vorliegenden Test treten keine Quadrantenfehler für die genannten Bedingungen auf. Seeber [189] ermittelte bei seiner Binauralsynthese Vorne-Hinten-Vertauschungen von 0,25% bei individueller Synthese. Seeber unterscheidet in seinen Experimenten nicht zwischen Vorne-Hinten-Vertauschungen, die sich bei Präsentation von virtueller Schallquellposition bei direkt 0° bzw. 180° oder bei seitlich ausgelenkten Schallquellen (siehe „within-cone“ Fehler bzw. Quadrantenfehler) ergeben. Die dort gemachten Angaben sind somit eine Kombination von Vorne-Hinten-, Hinten-Vorne-Vertauschungen und Quadrantenfehler.

**Externalisierung** - Für den Vortest ergeben sich die in Abbildung 4.12 gezeigten Bewertungen für die wahrgenommene Externalisierung der Hörereignisse als Indices. Im Vortest wurden die Bewertungen bei Übereinstimmung des Abhörtraumes mit dem zu synthetisierendem Raum, bei Verwendung von Kunstkopf-BRIRs und bei Sichtbarkeit des Raumes sowie der Schallquellpositionen (Lautsprecheratmosphäre) abgegeben. Zusätzlich wurden im Vortest nicht binaurale Audiosignale mit geringer räumlicher Qualität als Anker verwendet. Diese Signale sind mittels Intensitätsstereophonie für die jeweiligen Richtungen gepannte Stimuli. Zusätzlich wurden diese Signale mit einer monauralen Raumimpulsantwort des jeweiligen Abhörtraumes gefaltet. Diese Signale enthalten Richtungs- und Rauminformationen. Diese entsprechen aber nicht den spektralen und zeitlichen Merkmalen einer binauralen Signalgenerierung.

Für die als „stereo“ gekennzeichneten Stimuli ergeben sich Externalitätsindizes  $< 0,2$ . Dies entspricht den Erwartungen, da die Stimuli lediglich schwache binaurale und monaurale Merkmale enthalten. Für die Binauralsynthese mittels Kunstkopf-BRIRs ergeben sich für die 0° und 180° Richtungen Indexwerte von ca. 0,4 für den Seminarraum (SR) als Abhörtraum und Werte von ca. 0,2 für das Hörlabor (HL) als Abhörtraum. Indexwerte von  $> 0,8$  ergeben sich für die 90° und 120° Richtung und Werte von ca. 0,7 für die 330° Richtung für beide Abhörträume. Die Bewertungen des Vortests zeigen, dass den Testpersonen eine Unterscheidung zwischen den beiden Signaltypen hinsichtlich Externalisierung möglich ist.

Die Bewertungen des Haupttests werden zunächst bzgl. der Raumdivergenz zwischen Abhörtraum und synthetisiertem Raum ausgewertet. Hierfür wird getrennt auf die Bewertungen der Versuchsgruppen mit bzw. ohne Sichtbarkeit des Abhörtraumes eingegangen. Abbildung 4.13 verdeutlicht die Externalitätsindizes für die Gruppe ohne zusätzliche visuelle Merkmale für die Verwendung von individuellen und Kunstkopf-BRIRs.

Es zeigt sich ein signifikanter Anstieg (Test nach Fisher) bei Übereinstimmung zwischen Abhörtraum und synthetisiertem Raum im Vergleich mit der divergenten Raumsituation. Die Effektstärke lässt sich durch das Quotenverhältnis der Bewertungen ausdrücken. Eine Effektstärke von größer als Eins deutet auf einen positiven Einfluss der konvergenten Raumkondition in Bezug zur divergenten Kondition auf das Qualitätsmerkmal Externalisierung hin.

Dieser Effekt ist bspw. für die 0° Richtung und bei Anwendung einer individualisierten Binauralsynthese und Synthese des Seminartraumes erkennbar. Für diese Konditionen „SR(I)

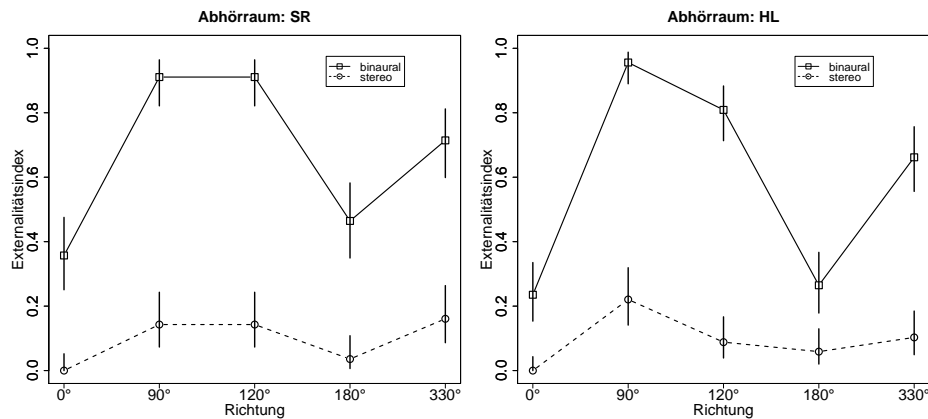


Abbildung 4.12: Bewertungen der wahrgenommenen Externalität aus dem Vortest als Externalitätsindex in Abhängigkeit vom Abhörraum und bei raumkonvergenten Abhörbedingungen; Verwendung von Kunstkopf(KEMAR)-BRIRs und Vorhandensein von visuellen Merkmalen mit 95% Konfidenzintervall; SR=Seminarraum, HL=Hörlabor.

in SR“ und „SR(I) in HL“ ergibt sich bspw. ein bei einem Signifikanzwert  $p < 0,1$  eine Effektgröße als Quotenverhältnis von 3,3. Vergleichbare Anstiege der Externalisierung sind weiterhin zwischen der individualisierten Synthese „HL(I) in HL“ und „HL(I) in SR“ erkennbar. Hierfür ergeben sich ein Signifikanzwert von  $p < 0,01$  und eine Effektstärke von 5,0. Für die  $180^\circ$  Richtung zeigen ebenfalls deutlich signifikante Unterschiede in den Bewertungen. Für die anderen Richtungen und bei der Verwendung von Kunstkopf-BRIRs zeigen sich teils geringere Signifikanzen. Die Werte dazu sind der Abbildung zu entnehmen.

Abbildung 4.14 verdeutlicht die Externalitätsindizes für die Gruppe mit Sichtbarkeit des Abhörortes für die Verwendung von individuellen und Kunstkopf-BRIRs. Das Vorhandensein von zusätzlichen visuellen Merkmalen erhöht die externalisierte Wahrnehmung der Hörereignisse für alle Raumkonditionen und Richtungsdarbietungen. Für die  $0^\circ$  und  $180^\circ$  Richtungen zeigen sich aber auch hier teils signifikante Unterschiede zwischen Raumkonvergenz und Raumdivergenz. Für die seitlichen Richtungen fallen diese Unterschiede hingegen geringer aus. Der Einfluss der Sichtbarkeit des Abhörortes wird in nachstehendem Abschnitt genauer diskutiert.

Der Einfluss der Sichtbarkeit des Abhörortes auf die Externalisierung der Hörereignisse ist in Abbildung 4.15 dargestellt. Die Abbildung zeigt den Anstieg zwischen den Bewertungen der Gruppen mit und ohne zusätzliche visuellen Merkmalen für die verschiedenen Richtungen, Personalisierungen und Raumkonditionen (oberer Bereich: Hörlabor als Abhörort; unterer Bereich: Seminarraum als Abhörort). Der Anstieg ist in Prozent sowie mit einer Bewertung der Signifikanz in der Abbildung angegeben.

Es zeigen sich signifikante Anstiege (Test nach Fisher) in der Externalisierung im Besonderen für die  $0^\circ$  und  $180^\circ$  Richtungen sowie für seitliche Richtungen bei Wiedergabe im nachhallarmen

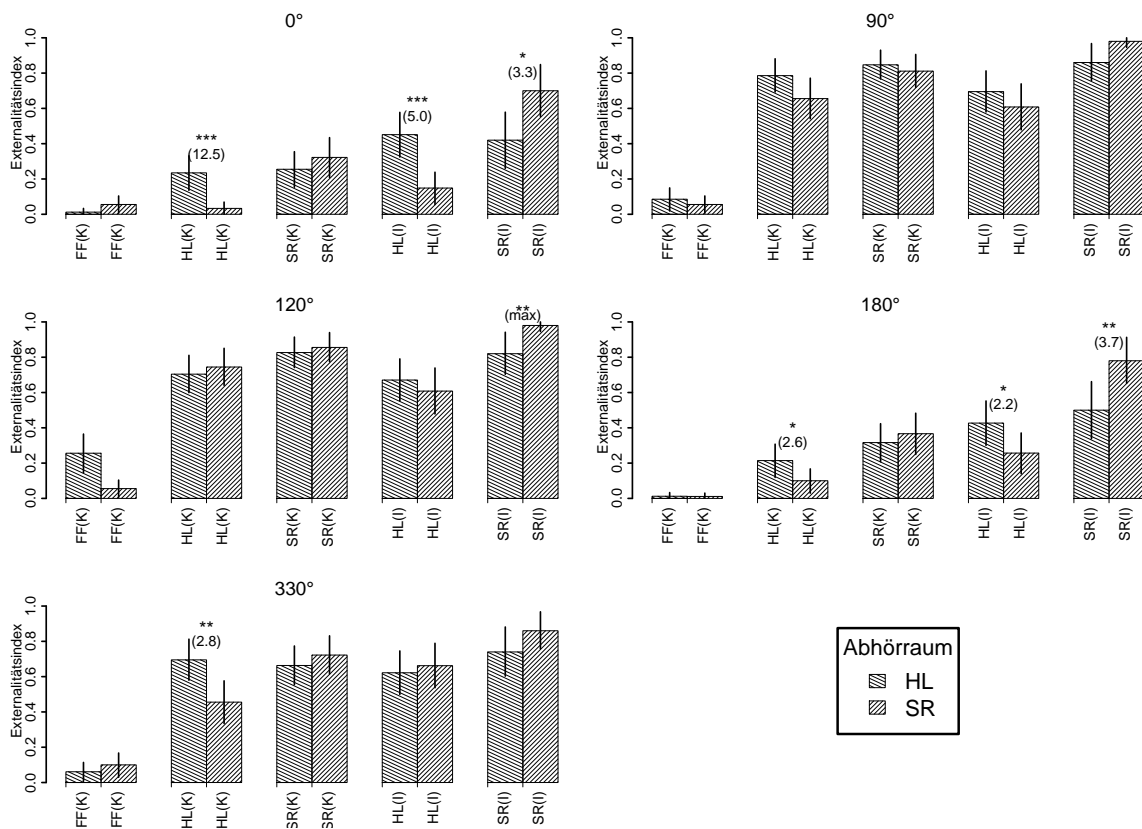


Abbildung 4.13: Externalitätsindizes für Kombination von Abhörraum und synthetisiertem Raum, Personalisierung der Binauralsynthese und bei nicht Vorhandensein von visuellen Merkmalen mit 95% Konfidenzintervall; SR=Seminarraum, HL=Hörlabor, FF=Freifeld, K=Kunstkopf BRIRs, I=individuelle BRIRs; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ , \* signifikanter Unterschied bei  $p < 0,1$ ; Zahlen in Klammern ist die Effektgrößen als Quotenverhältnis.

Hörlabor. Eine Betrachtung aller Konditionen liefert allerdings keine klaren Tendenzen bzgl. Personalisierungsmethode und Raumkonditionen. Über alle Testkonditionen hinweg ist ein mittlerer Anstieg der Externalität von 16% erkennbar. Ein höherer Anstieg von 26% ist für die 180° Richtung unabhängig von der Raumkondition und Personalisierung erkennbar. Der Effekt ist in sechs von acht Testkonditionen bei einem Signifikanzwert von mindestens  $p < 0,1$  signifikant. Der überdurchschnittlich hohe Anstieg der Externalisierung in rückseitiger Richtung ist unerwartet. Die Testpersonen wurden angewiesen, während des Tests den Kopf stillzuhalten und sich vor allem nicht umzudrehen.

In den bisherigen Betrachtungen ergeben sich zum Teil größerer Abweichungen in den Bewertungen zwischen einer individualisierter Binauralsynthese und bei Verwendung von Kunstkopf-BRIRs. Dieser Einfluss soll hinsichtlich der unterschiedlichen Raumkonditionen genauer betrachtet werden. Abbildung 4.16 verdeutlicht den Einfluss der Personalisierung auf die Bewertung von Externalität.

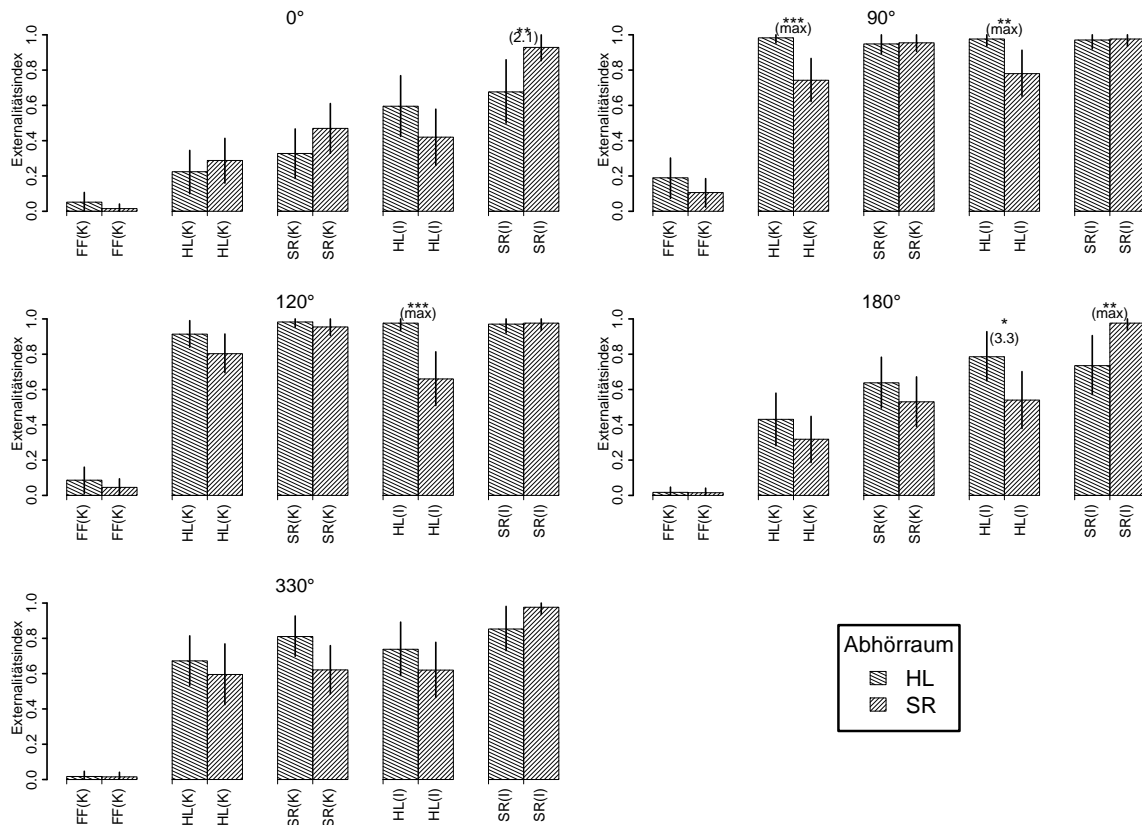


Abbildung 4.14: Externalitätsindizes in Abhängigkeit der Kombination von Abhörraum und synthetisiertem Raum, Personalisierung der Binauralsynthese und bei Vorhandensein von visuellen Merkmalen mit 95% Konfidenzintervall; SR=Seminarraum, HL=Hörlabor, FF=Freifeld, K=Kunstkopf BRIRs, I=individuelle BRIRs; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ , \* signifikanter Unterschied bei  $p < 0,1$ ; Zahlen in Klammern sind die Effektgrößen als Quotenverhältnis.

Die Externalisierung steigt bei Verwendung einer individualisierten Binauralsynthese im Vergleich zur Verwendung von Kunstkopf-BRIRs an. Signifikante Unterschiede (Test nach Fisher) ergeben sich für die  $0^\circ$  und  $180^\circ$  Richtungen mit einem Signifikanzwert von  $p < 0,01$  und Effektgrößen als Quotenverhältnisse von 3,4 bis zu 10. Die Effektgrößen auf die Externalisierung scheinen für die Personalisierungsmethoden größer auszufallen, als die für den Raumdivergenzeffekt. Beispielsweise wird eine Effektgröße von 5 für die Synthese des Seminarraumes bei Verwendung individueller BRIRs für die  $180^\circ$  Richtung bei Raumdivergenz beobachtet. Eine Effektgröße von 10 wird dahingegen für die Testkondition „SR in SR“ für die gleiche Richtung zwischen den Personalisierungsmethoden festgestellt. Vergleichbare Tendenzen sind für die  $0^\circ$  Richtung und andere Raumkonditionen erkennbar. Weiterhin werden größere Effekte für konvergierende Raumkonditionen im Vergleich zu divergenten Raumkonditionen für die  $0^\circ$  und  $180^\circ$  Richtungen beobachtet.

Testkondition		0°	90°	120°	180°	330°	alle Richtungen
individuelle BRIRs	SR in HL	●* 27	● 13	● 17	● 25	● 13	● 19
	HL in HL	● 15	●*** 30	●*** 33	●*** 38	● 13	● 26
Kunstkopf-BRIRs	SR in HL	● 7	● 11	●** 17	●*** 33	● 15	● 17
	HL in HL	○ -2	●* 21	●** 22	●* 22	○ -2	● 12
Σ in HL		● 12	● 19	● 22	● 30	● 10	● 19
individuelle BRIRs	SR in SR	●** 24			●** 21	● 13	● 12
	HL in SR	●** 28	● 18	● 6	●** 30	○ -4	● 16
Kunstkopf-BRIRs	SR in SR	● 15	●* 15	● 11	● 17	○ -10	● 10
	HL in SR	●*** 26	● 9	● 6	●** 22	● 15	● 16
Σ in SR		● 23	● 11	● 6	● 23	● 4	● 13
Σ alle		● 18	● 15	● 14	● 24	● 7	● 16

Abbildung 4.15: Anstieg des Externalitätsindex in Prozent bei Vorhandensein von visuellen Merkmalen in Bezug auf keine visuellen Merkmale im Test; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ , \* signifikanter Unterschied bei  $p < 0,1$ .

Ein Vergleich der Externalitätsindizes bei Verwendung von Kunstkopf-BRIRs und bei den raumkonvergenten Konditionen „HL in HL“ und „SR in SR“ weisen mit den Bewertungen aus dem Vortest eine sehr hohe Übereinstimmung der Mediane auf. Die Konfidenzintervalle sind bei den Bewertungen aus dem Vortest teils kleiner als im Haupttest. Die im Vortest verwendete Multistimulus-Methode zur Bewertung der Testitems scheint somit ebenfalls für die Bewertung der wahrgenommenen Externalität geeignet. Ein Vorteil dieser Methode liegt in der geringeren benötigten Zeitdauer zur Bewertungsabgabe. Die hohe Übereinstimmung der Ergebnisse aus den Vortests mit denen aus den Haupttests bestärkt die Annahme einer hohen Zuverlässigkeit beider Testmethoden.

Abbildung 4.17 zeigt die Korrelation zwischen den Bewertungen der im Haupttest vorkommenden Wiederholungen der Teststimuli. Zur Berechnung der Korrelationen wird ein Korrelationstest nach Spearman für verbundene Stichproben durchgeführt. Die Ergebnisse können als Anhaltspunkt zur Einschätzung der Intrarater-Reliabilität der Bewertungen für Externalität über alle Testpersonen angesehen werden. Ein  $\rho$ -Wert von eins deutet auf vollständige Übereinstimmung zwischen den Wiederholungen hin. Zusätzlich wird der  $p$ -Wert als statistisches Maß der Signifikanz der Korrelationen mit angegeben.

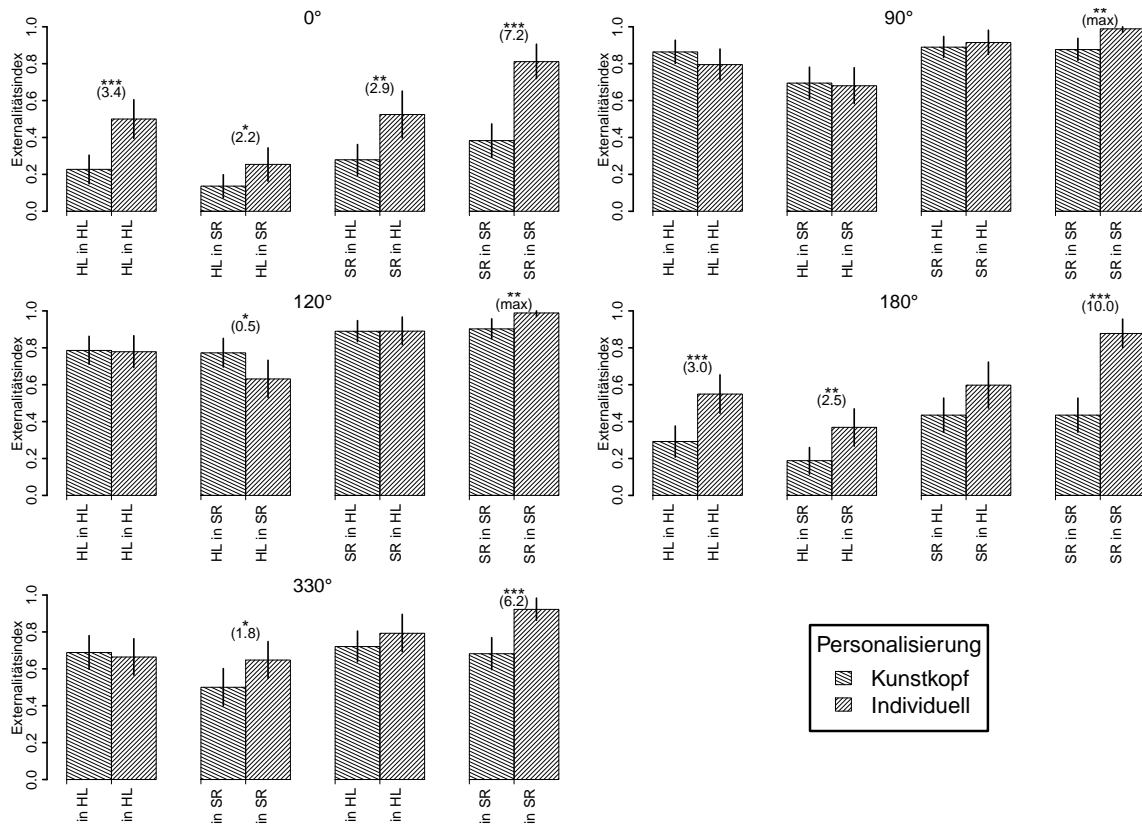


Abbildung 4.16: Externalitätsindizes in Abhängigkeit der Personalisierung der Binauralsynthese durch Verwendung von Kunstkopf- und individuellen BRIRs mit 95% Konfidenzintervall; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ , \* signifikanter Unterschied bei  $p < 0,1$ ; Zahlen in Klammern sind die Effektgrößen als Quotenverhältnis

Bei der Probandengruppe mit zusätzlichen visuellen Merkmalen im Test (Abb. 4.17 oben) zeigen sich überwiegend hohe und signifikante ( $p < 0,05$ ) Korrelationswerte zwischen den Bewertungen der Wiederholungen. Die größten Unterschiede in den Bewertungen der Wiederholungen sind bei der untersuchten Raumkombination „HL in SR“ bei Verwendung von Kunstkopf-BRIRs zu erkennen. Geringere Korrelationen sind ebenfalls tendenziell für die Wiedergabe im „SR“ zu erkennen. Weitere Unterschiede in den Bewertungen ergeben sich für die Bedingungen „SR in HL“ mit individuellen BRIRs und „SR in SR“ bzw. „HL in HL“ bei Kunstkopf-BRIRs. Die Abweichungen in den Bewertungen sind verstärkt für die  $0^\circ$  und  $180^\circ$  Richtung festzustellen.

Für die Probandengruppe ohne zusätzliche visuelle Merkmale (Abb. 4.17 unten) zeigen sich keine signifikanten Korrelationen ( $p > 0,05$ ) zwischen den Bewertungen der Stimuli-Wiederholungen bei Wiedergabe der Raumkombination „SR in SR“ bei Verwendung von individuellen und Kunstkopf-BRIRs für die Richtungen  $0^\circ$ ,  $120^\circ$ ,  $180^\circ$  und  $330^\circ$ . Weitere Unterschiede in den Bewertungen ergeben sich für die Bedingungen der Wiedergabe mittels

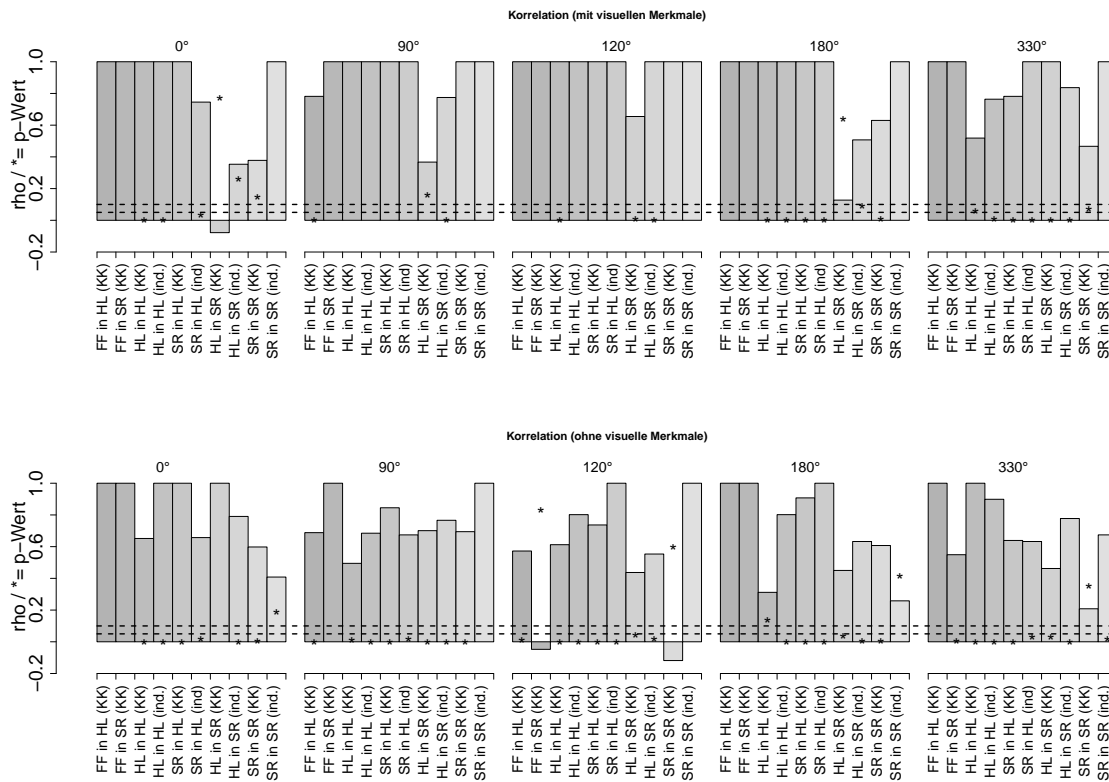


Abbildung 4.17: Korrelationen der Bewertungen der im Test vorkommenden zwei Wiederholungen für untersuchte Richtungen, Raumkombinationen und Personalisierungsmethoden; oben: mit zusätzlichen visuellen Merkmalen, unten: ohne zusätzliche visuelle Merkmale; rho-Wert nach Spearman und p-Wert (\*) mit Kennzeichnung der Signifikanzniveaus 5% und 10% als gestrichelte Linien; HL=Hörlabor, SR=Seminarraum, FF=Freifeld, KK=Kunstkopf-(BR)IRs, ind.=individuelle BRIRs.

Freifeld-Kunstkopf-HRTFs im Seminarraum (SR\_FF\_KK), sowie bei „HL in HL“ ebenfalls bei Kunstkopf-BRIRs.

Bei den untersuchten Divergenzen kommt es offensichtlich zur Konfusionen in der auditiven Wahrnehmung von Externalität, welches sich in ungleichen Bewertungen für die beiden Wiederholungen widerspiegelt. Die Probandengruppe mit zusätzlichen visuellen Merkmalen profitiert anscheinend von den zusätzlichen Merkmalen, welches sich in einer höheren Korrelation zwischen den Bewertungen der beiden Wiederholungen widerspiegelt.

**Zusammenhang Lokalisation und Externalisierung** - Nachstehende Abbildungen verdeutlichen den Zusammenhang zwischen wahrgenommener Externalität und Lokalisation für verschiedene Richtungen, Testbedingungen und die beiden Probandengruppen. Die Lokalisationsfähigkeit wird als Häufigkeit des Quadrantenfehlers angegeben. Für eine dargebotene Schalleinfallrichtung von 0° und 330° entspricht der Fehler einer Vorne-Hinten-Vertauschung,



für eine Schalleinfallrichtung von  $180^\circ$  einer Hinten-Vorne-Vertauschung in Prozent angegeben.

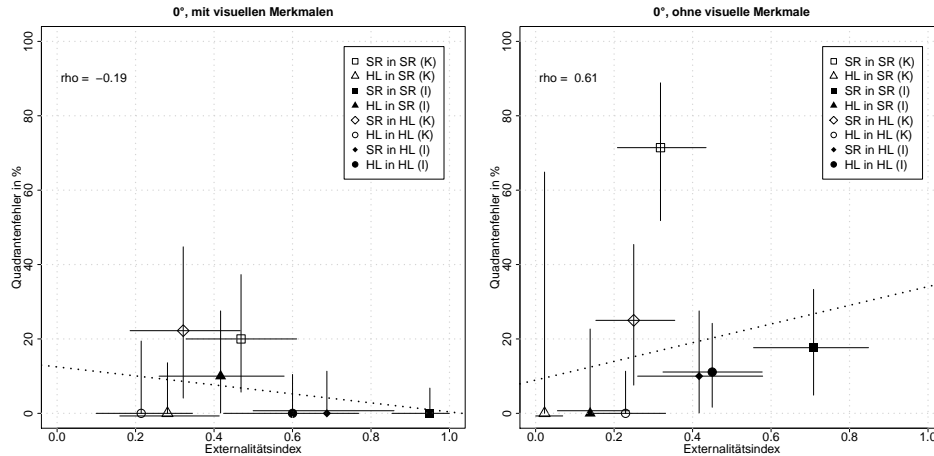


Abbildung 4.18: Externalitätsindizes und Quadrantenfehler (Vorne-Hinten-Vertauschung) für die Darbietung der  $0^\circ$  Richtung; links=mit zusätzlichen visuellen Merkmalen; rechts=ohne zusätzliche visuelle Merkmale; 95% Binominalintervall; rho-Wert nach Spearman; lineare Regressionsgerade (gepunktete Linie); SR=Seminarraum, HL=Hörlabor, K=Kunstkopf-BRIRs, I=individuelle BRIRs.

Die Bewertungen für die wahrgenommene Externalität und Anzahl an Richtungsvertauschnungen für die  $0^\circ$  Richtung sind in Abbildung 4.18 dargestellt. Geringere Lokalisationsfehler werden bei Verwendung individueller BRIRs gegenüber der Verwendung von Kunstkopf-BRIRs in drei von vier Raumkombinationen beobachtet. Die Anzahl an Richtungsvertauschnungen sinkt für die Gruppe mit zusätzlichen visuellen Merkmalen für alle Raumkombinationen und Methoden der Personalisierung. Der Externalitätsindex steigt in sieben von acht Testbedingungen bei der Gruppe mit zusätzlichen visuellen Merkmalen an. Dieses Verhalten ist in allen Fällen bei Anwendung einer personalisierten Synthese zu beobachten.

Abbildung 4.19 veranschaulicht die Bewertungen der wahrgenommenen Externalisierung und Häufigkeit der Hinten-Vorne-Vertauschungen für eine Schalleinfallrichtung von  $180^\circ$ , verschiedene Raumkombinationen, Personalisierungen und Probandengruppen. Die Externalitätsindizes steigen in sieben von acht Testbedingungen für die Gruppe mit zusätzlichen visuellen Merkmalen gegenüber der anderen Gruppe an. Bei allen Testbedingungen unter Verwendung individueller BRIRs steigt die Externalisierung an. Die Anzahl an Hinten-Vorne-Vertauschungen sinkt in drei von vier Testbedingungen bei Verwendung individueller BRIRs und für die Gruppe mit zusätzlichen visuellen Merkmalen. Ein starker negativer Zusammenhang zwischen Lokalisationsfehler und Externalitätsindex ist für beide Gruppen erkennbar. Der Rang-Korrelationskoeffizient nach Spearman beträgt  $-0,76$  und  $-0,75$  für die beiden Gruppen.

Abbildung 4.20 zeigt die Bewertungen der wahrgenommenen Externalisierung und Häufigkeit der Vorne-Hinten-Vertauschungen für die  $330^\circ$  Richtung, verschiedene Testbedingungen und

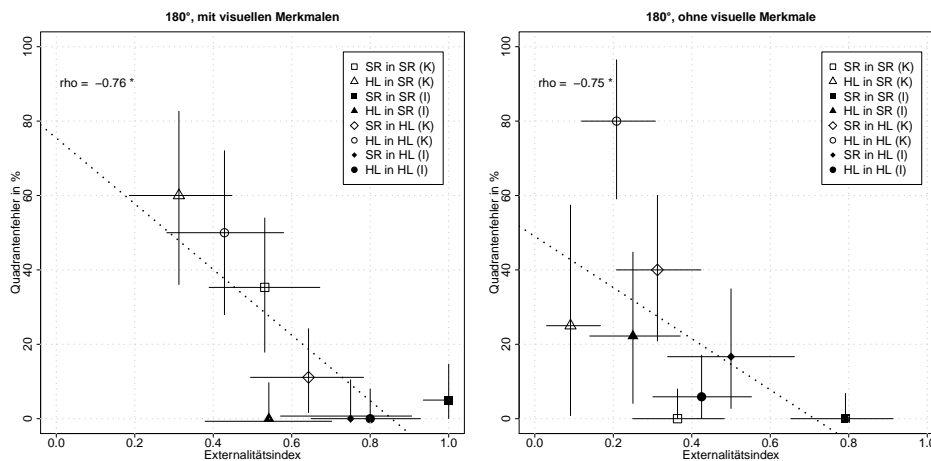


Abbildung 4.19: Externalitätsindizes und Quadrantenfehler (Hinten-Vorne-Vertauschung) für die Darbietung der 180° Richtung; links=mit zusätzlichen visuellen Merkmalen; rechts=ohne zusätzliche visuelle Merkmale; 95% Binominalkonfidenzintervall; rho-Wert nach Spearman, \* signifikantes rho mit  $p < 0,05$ ; lineare Regressionsgerade (gepunktete Linie); SR=Seminarraum, HL=Hörlabor, K=Kunstkopf-BRIRs, I=individuelle BRIRs.

für die beiden Probandengruppen. Ein schwacher negativer Zusammenhang zwischen Lokalisationsfehler und Externalitätsindex ist zu beobachten. Die Externalitätsindizes steigen in drei von vier Testbedingungen bei Anwendung einer personalisierten Binauralsynthese und in null von vier Testbedingungen bei Verwendung von Kunstkopf-BRIRs an.

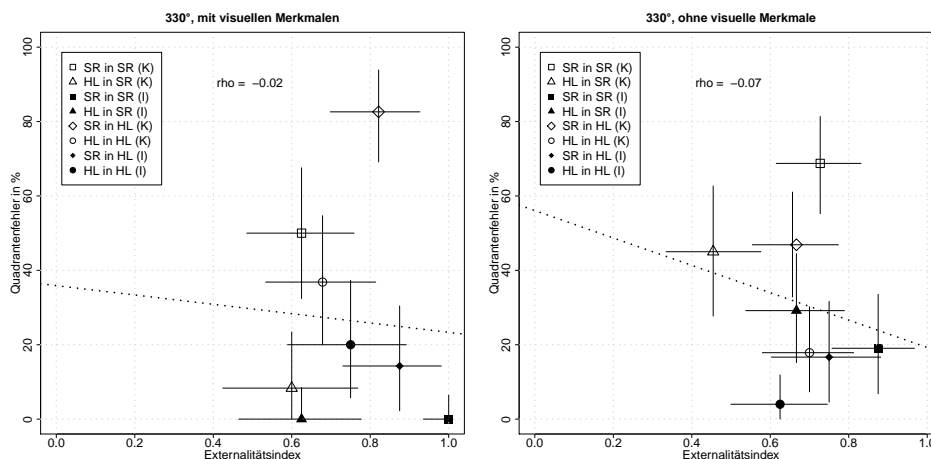


Abbildung 4.20: Externalitätsindizes und Quadrantenfehler (Vorne-Hinten-Vertauschung) für die Darbietung der 330° Richtung; links=mit zusätzlichen visuellen Merkmalen; rechts=ohne zusätzliche visuelle Merkmale; 95% Binominalkonfidenzintervall; rho-Wert nach Spearman; lineare Regressionsgerade (gepunktete Linie); SR=Seminarraum, HL=Hörlabor, K=Kunstkopf-BRIRs, I=individuelle BRIRs.

Geringere Ungenauigkeiten in der Lokalisierung werden bei einer personalisierten Binauralsynthese gegenüber der Verwendung von Kunstkopf-BRIRs beobachtet (Vergleich zu [154]). Das

Vorhandensein von zusätzlichen visuellen Merkmalen führt zu einer deutlichen Verringerung der Vorne-Hinten-Vertauschungen bei einer individuellen Synthese. Dieses Verhalten ist nicht bei der Verwendung generischer BRIRs festzustellen. Tendenziell weisen die Bewertungen einen Zusammenhang zwischen der Anzahl der Lokalisierungsfehler und Externalität auf. Eine starke negative Korrelation ist für die 180° Richtung erkennbar. Eine höhere Anzahl an Lokalisierungsfehler und geringer Externalitätsindizes lassen sich bei Verwendung von Kunstkopf-BRIRs beobachten. Eine personalisierte Synthese führt zu einem Anstieg der Externalisierung und zu geringeren Lokalisierungsungenauigkeiten. Es werden keine Zusammenhänge zwischen der Divergenz des Abhörtraumes und synthetisiertem Raum in Bezug auf die Häufigkeit des Auftretens von Lokalisierungsfehlern für beide Personalisierungsmethoden beobachtet.

#### 4.2.2 Zusammenfassung

**Raumdivergenzeffekt** - Ein Abfall der Externalität wird für Divergenz zwischen Abhörtraum und synthetisiertem Raum gegenüber Konvergenz der Räume beobachtet. Dieser Effekt ist dominant für vordere und hintere Richtungen des Schallereignisses. Im Gegensatz zur Divergenz zwischen den Räumen werden hohe Externalitätsindizes ( $>0.5$ ) für Kongruenz zwischen den Räumen erreicht. Im Allgemeinen führt die Synthese des halligeren Seminarraums zu höheren Externalitätsindizes im Vergleich zur Synthese des weniger halligen Hörlabors. Der Raumdivergenzeffekt ist aber dennoch in beiden Abhörträumen sichtbar.

Es wird weiterhin vermutet, dass ein Training auf die Synthese des Hörlabors bzw. des Seminarraumes eine unterschiedliche Adaptation der Hörer an die Synthese hervorruft. Diese Änderung führt zu einem Verschieben der Erwartung und internen Referenz in Richtung konvergenter oder divergenter Raumbedingung in der Bewertung der wahrgenommenen Externalität.

In der Arbeit von Klein et al. [8] wird eine vermutete Verschiebung der Erwartung und internen Referenz in Richtung konvergenter oder divergenter Raumbedingung in der Bewertung der wahrgenommenen Externalität untersucht. Es werden zwei Gruppen von Hörern untersucht, welche zum einen ein Hörtraining auf Raumkonvergenz und zum anderen auf Raumdivergenz erhalten. Das Ziel ist die Adaptation der Hörer auf die jeweilige Raumkondition. Ein sich anschließender Hörtests misst die Externalisierung der Hörereignisse bei Raumkonvergenz und bei Raumdivergenz.

Abbildung 4.21 verdeutlicht den Trainingseffekt auf die Bewertung der Externalisierung der Hörereignisse für die zwei Gruppen von Testpersonen. Den Testpersonen beider Gruppen werden Hörstimuli vorgespielt, welche eine Synthese eines Lautsprechers aus unterschiedlichen Richtungen und Räumen umfasst. Der Abhörtraum ist für beide Gruppen ein halliger Seminarraum (SR, siehe Kapitel 2.5).

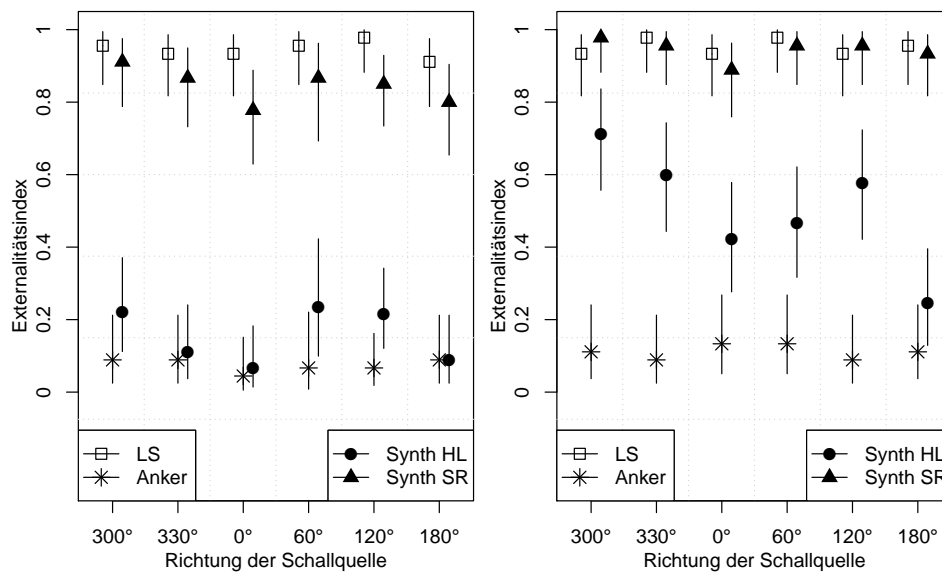


Abbildung 4.21: Externalitätsindizes der Bewertungen von zwei unterschiedlich trainierten Gruppen von Testpersonen, links=konvergente Gruppe, rechts=divergente Gruppe; Abbildungen aus Klein et al.[8].

In dem vorangestellten Trainingsdurchgang werden der einen Gruppe raumkonvergente Audiosignale (konvergente Gruppe) und der zweiten Gruppe raumdivergente Signale (divergente Gruppe) vorgespielt. Die Testpersonen bewerten die wahrgenommene Schalleinfallrichtung und Externalität der Signale. Für jede Bewertung im Training wird eine visuelle Bestätigung der zu synthetisierenden Richtung auf der Bewertungsfläche (vgl. siehe Kapitel 3.1.3) gegeben. Für die konvergente Gruppe werden die Lautsprecherpositionen des Seminarraumes im Seminarraum synthetisiert. Für die divergente Gruppe werden die Lautsprecherpositionen des Hörlabors (HL, siehe Kapitel 2.5) im Seminarraum synthetisiert.

Im anschließenden Hörtest werden beiden Gruppen konvergente und divergente Stimuli sowie die realen Lautsprecher im Seminarraum und Ankersignale vorgespielt. Die Ankersignale sind mittels Intensitätsstereophonie in ihrer Richtungswahrnehmung eingestellte Signale, welche mit einer Raumimpulsantwort (RIR) der jeweiligen Lautsprecherposition gefaltet wurden. Für die konvergente Gruppe wird die RIR aus dem SR, für die divergente Gruppe die RIR aus dem HL genommen. Der Anker enthält somit zum jeweiligen Training passende Merkmale der Halligkeit, aber schwache binaurale und monaurale Merkmale für Direktschall und Reflexionen. Der Anker sollte ein gering externalisiertes Hörereignis auslösen.

Das linke Diagramm in Abbildung 4.21 zeigt die Bewertungen für die konvergente Gruppe. Das Training führt zu einer Ausdifferenzierung. Die raumdivergenten Stimuli (Synth HL) werden in ihrer Externalisierung deutlich abgewertet und zeigen keine signifikanten Unterschiede ( $p > 0,05$ ) mit den Ankersignalen. Die konvergenten Stimuli (Synth SR) werden deutlich

höher bewertet und zeigen keine signifikanten Unterschiede ( $p > 0,05$ ) mit den Bewertungen der realen Lautsprecher.

Das rechte Diagramm in Abbildung 4.21 zeigt die Bewertungen für die divergente Gruppe. Das Training führt zu einer deutlicher externeren Bewertung der divergenten Stimuli (Synth HL). Es zeigen sich hier signifikante Unterschiede ( $p < 0,05$ ) zu den Ankersignalen. Die konvergenten Stimuli (Synth SR) werden analog zur ersten Gruppe höher bewertet und zeigen keine signifikanten Unterschiede ( $p > 0,05$ ) mit den Bewertungen der realen Lautsprecher.

Das Training führt dazu, dass die konvergente Gruppe divergente Stimuli deutlich abwertet. Die divergente Gruppe wertet divergente Stimuli dahingegen nicht in dem Ausmaß ab. Die raumkonvergenten Stimuli werden allerdings von beiden Gruppen ähnlich hoch bewertet. Es scheint so zu sein, dass die divergente Gruppe tolleranter gegenüber divergenten Raumkonditionen ist ohne aber konvergente Stimuli abzuwerten.

**Visuelle Merkmale** - Ein mittlerer Anstieg von 16% in der Wahrnehmung von Externalität wird bei Sichtbarkeit des Abhörtraumes und der Positionen der virtuellen Schallquellen als Lautsprecherattrappen im Vergleich zu keinen zusätzlichen visuellen Merkmalen beobachtet. Dieser Effekt ist unabhängig bezogen auf die Personalisierung der Binauralsynthese und die Kombination von Abhörtraum und synthetisierten Raum.

**Personalisierung** - Im Allgemeinen sind die Externalitätsindizes kleiner bei Verwendung von Kunstkopf-BRIRs im Vergleich zu einer individualisierten Binauralsynthese. Der Raumdivergenzeffekt ist weniger ausgeprägt bei Verwendung von Kunstkopf-BRIRs. Dennoch ist der Effekt in beiden Bedingungen sichtbar. Es wird geschlossen, dass die Personalisierung des Binauralsynthesystems den Raumdivergenzeffekt überlagert.

**Externalisierung vs. Lokalisation** - Abbildung 4.22 verdeutlicht den Zusammenhang zwischen Externalisierung und Quadrantenfehler für 25 Richtungen um den Hörer [43, 29]. In der Untersuchung wurden ein Saxofon-Signal, ein tiefpassgefiltertes (80 Hz bis 600 Hz) und ein hochpassgefiltertes (6 kHz bis 10 kHz) weißes Rauschen als Audiosignale in der Binauralsynthese verwendet. Die BRIRs der verschiedenen Richtungen wurden an einem Kunstkopf (KEMAR) und an einem Kugelflächenmikrofon (vgl. MTB-Mikrofonarray aus Hesse [36]) gemessen. In dem Hörversuch lag Raumkonvergenz vor. Weiterhin waren der Abhörtraum und Lautsprecherattrappen den Testpersonen sichtbar.

Die Analyse der Bewertungen aller Testsignale bei Verwendung der Kunstkopf-BRIRs zeigt eine signifikante negative Korrelation ( $p < 0,05$ , Spearmans rho) zwischen der Häufigkeit der Quadrantenfehler und der Externalisierung. Diese bestätigt die vorstehende Annahme eines derartigen Zusammenhangs. Im deutlichen Gegensatz dazu stehen die Bewertungen bei Verwendung der BRIRs des Kugelflächenmikrofons. Hierfür zeigt sich kein Zusammenhang zwischen Quadrantenfehler und Externalisierung. Die Verteilung der Quadrantenfehler folgt

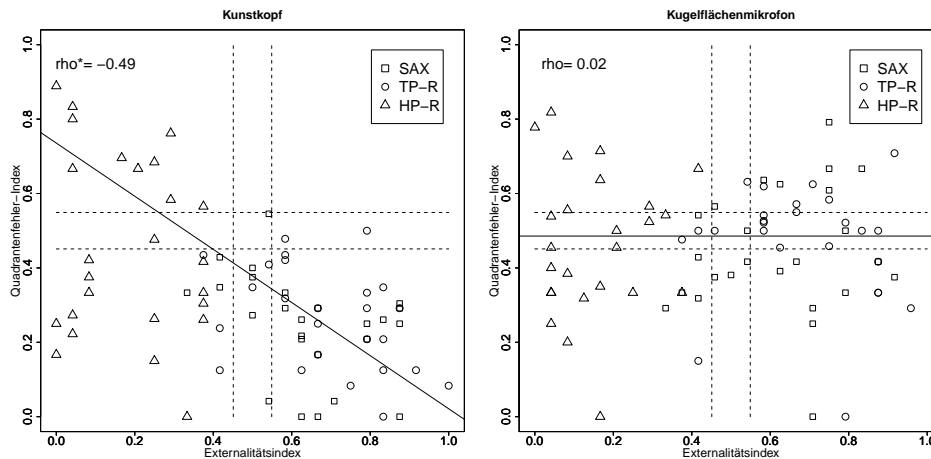


Abbildung 4.22: Externalitätsindizes und Quadrantenfehler für die Darbietung verschiedener Richtungen; links=Kunstkopf; rechts=Kugelflächenmikrofon; 95% Binominalkonfidenzintervall; rho-Wert nach Spearman, \* signifikantes rho bei  $p < 0,05$ ; lineare Regressionsgerade (durchgezogene Linie); Ratebereiche (gestrichelte Linien); SAX=Saxophon, TP-R=tiefpassgefiltertes Rauschen, HP-R=hochpassgefiltertes Rauschen; Abbildungen nach [29].

einer Normal-Verteilung mit einem Mittelwert von 0,5. Dies deutet auf eine zufällige Verteilung der Richtungswahrnehmung hin. Die wahrgenommene Externalisierung der Stimuli ist hingegen mit den Kunstkopf-BRIRs vergleichbar.

Georg Plenge formulierte in seiner Veröffentlichung [168]: „[Im-Kopf-Lokalisation tritt dann auf] ... wenn fehlende, mangelhafte oder falsche Schallquellen- und feldkenntnis ... und/oder die Signale und damit Reize so geartet sind, dass sie keinem im Langzeitspeicher enthaltenem Reizmuster zugeordnet werden können.“ Die auditorischen Merkmale der Kunstkopfmessungen passen nur in geringerem Maße mit den individuellen Hörmustern überein. Hierdurch liegt u. U. ein falsches Wissen über das Schallfeld vor, welches in der auditorischen Wahrnehmung zu Lokalisationsfehlern führt. Dem folgend sollte ein Anstieg der Lokalisationsfehler auch zu einem Anstieg der Im-Kopf-Lokalisation führen. Diese Annahme scheint für den Kunstkopf zuzutreffen, ist aber für das Kugelflächenmikrofon nicht ohne weiteres haltbar. Die Testpersonen raten in der Richtungswahrnehmung, bewerten aber die Externalisierung vergleichbar zu den Kunstkopfsignalen.

Welch und Warren formulieren in ihrer Arbeit [226] eine Hypothese der Angemessenheit der Sinnesmodalitäten. Das Wahrnehmungssystem wählt demnach diejenige Modalität aus, welche die höchste Zuverlässigkeit in Bezug auf eine plausible Wahrnehmung der Umwelt erwarten lässt. Dieser Idee folgend, kann das Fehlen von auditorischen Merkmalen, bspw. durch die schwachen spektralen Merkmale des Kugelflächenmikrofons, die Sichtbarkeit des Abhörortes, der Lautsprechertrappen sowie weitestgehender akustischer Raumkonvergenz dennoch zu

einer plausiblen räumlichen Audiowahrnehmung führen. Hierzu bieten sich weiterführende Untersuchungen an.

Abschließend wird festgestellt, dass der beobachtete Raumdivergenzeffekt, verursacht durch verschiedene Kombinationen von Abhörraum und synthetisierten Raum als ein akustisch basierter kontextabhängiger Qualitätsparameter einer Binauralsynthese angenommen wird. Der Parameter hat einen Einfluss auf das Qualitätsmerkmal Externalisierung. Der Einfluss basiert auf den Erwartungen des Hörers an die Raumakustik des Abhörzimmers.

**Unabhängigkeit der Kontextparameter** - Die Überprüfung auf Unabhängigkeit erfolgt durch den Vergleich der Auftretenshäufigkeit von extern bewerteten Stimuli für die verschiedenen kontextabhängigen Qualitätsparameter. Die Überprüfung erfolgt durch einen McNemar-Test für verbundene Stichproben für das dichotome Merkmal Externalität [91]. In den aufgestellten Vierfeldertafeln stehen die Anzahl an extern und im-Kopf bewerteten Stimuli für den einen Kontextparameter und die Bewertungen für einen anderen Parameter. Die Vergleiche werden für alle Richtungen durchgeführt. Die untersuchten Parameterkombinationen sind:

- Raumdivergenz (Abhörzimmer) vs. Personalisierung
- Raumdivergenz (synthetisierter Raum) vs. Personalisierung
- Raumdivergenz (Abhörzimmer) vs. visuelle Merkmale
- Raumdivergenz (synthetisierter Raum) vs. visuelle Merkmale
- Personalisierung vs. visuelle Merkmale

Für die Vergleiche ergeben sich keine signifikanten Unterschiede bei einem Signifikanzniveau von 5% (die statistischen Prüfwerte sind kleiner als der kritische Wert). Dies deutet auf Unabhängigkeit der einzelnen kontextabhängigen Qualitätsparameter hin.

### 4.3 Räumliche Szenenkomplexität und raumakustische Disparität

Es wird der Einfluss der Komplexität der akustischen Szene und der Raumakustik der wiedergabenden Räume auf die Wahrnehmung verschiedener Qualitätsmerkmale bei Verwendung eines Binauralsynthesystems untersucht. Die Qualitätsmerkmale räumliche Präsenz und Nachhalleinhüllung werden neben dem Merkmal Externalität und Richtungswahrnehmung untersucht. In den Untersuchungen werden drei verschiedene Räume mit unterschiedlicher raumakustischen Charakteristik und verschiedenen Szenen mit unterschiedlicher räumlicher Komplexität betrachtet. Es wird die Frage adressiert, inwiefern die verschiedenen Qualitätsmerkmale von den unterschiedlichen Audioszenen beeinflusst werden. Die Ergebnisse zeigen, dass räumliche Präsenz durch die räumliche Komplexität, und dass die Nachhalleinhüllung durch die Unterschiede der Akustik der Räume beeinflusst werden. Externalität und Lokalisierung werden nicht durch die räumliche Komplexität, ungeachtet der Personalisierungsmethode der Binauralsynthese, beeinflusst. Weiterführende Informationen sind der dazugehörigen Veröffentlichung zu entnehmen [21].

#### 4.3.1 Exp. IV: Audioszene

Der Hörversuch evaluiert den Einfluss der räumlichen Komplexität und der raumakustischen Disparität der synthetisierten Szene auf die Wahrnehmung. Der Einfluss auf die Qualitätsmerkmale „räumliche Präsenz“ und „Nachhallumhüllung“ sowie „Externalisierung“ und „Lokalisation“ wird untersucht. Die räumliche Komplexität beschreibt die unterschiedliche räumliche Anordnung von mehreren Schallquellen um den Hörer herum. Die raumakustische Disparität beschreibt die Wiedergabe von Räumen mit unterschiedlichen raumakustischen Eigenschaften.

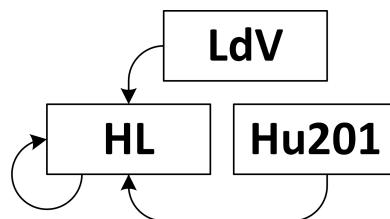


Abbildung 4.23: Kombinationen von Abhörraum und synthetisiertem Raum innerhalb des Hörtests; HL = Hörlabor, LdV = Leonardo DaVinci Hörsaal, Hu201 = Seminarraum.

#### 4.3.1.1 Untersuchungsgegenstand

Es wird der Frage nachgegangen, ob die untersuchten Qualitätsmerkmale von der Wiedergabe der verschieden komplexen Audioszenen beeinflusst werden oder nicht. Das Experiment ermittelt den Unterschied in der Wahrnehmung verschiedener Qualitätsmerkmale zwischen der binauralen Resynthese von drei verschiedenen Räumen sowie von verschiedenen Szenen mit unterschiedlicher räumlicher Komplexität. Im Detail wird der Effekt einer räumlichen Verteilung



der Audioobjekte in unterschiedlichen Räumen auf die Qualitätsmerkmale Externalisierung, räumliche Präsenz und Nachhallumhüllung untersucht.

Nachstehende Fragen sollen beantwortet werden: Führt eine verschiedene Anzahl an synthetisierten Audioobjekten zu einer Änderung in der Externalisierung des Höreignisses eines Zielobjektes? Führt eine verschiedene Anzahl an synthetisierten Audioobjekten zu einer Änderung der Auftretenshäufigkeit eines Quadrantenfehlers in der Richtungswahrnehmung eines Zielobjektes? In welchem Umfang ändert sich die wahrgenommene räumliche Präsenz einer Audioszene in Abhängigkeit der unterschiedlichen räumlichen Anordnung und einer unterschiedlichen Anzahl an Audioobjekten um den Hörer? Inwieweit hängt die wahrgenommene Nachhallumhüllung von der Anzahl und räumlichen Anordnung der Audioobjekte um den Hörer ab?

#### 4.3.1.2 Versuchsaufbau

**Binauralsynthesystem** - Es werden individuelle und mittels KEMAR aufgezeichnete BRIRs für die ausgewählten Räume, Schallquellen und Positionen verwendet. Das binaurale System ist personalisierbar, um Verwechslungsfehler in der Richtungswahrnehmung zu reduzieren sowie die Wiedergabetreue der Simulationen im Vergleich zu echten Lautsprechern zu erhöhen. Die Räume sind so gewählt, dass sie unterschiedliche raumakustische Eigenschaften wie Nachhall und Quellen-Empfänger-Abstände aufweisen. Die Kopfhörer werden über individuelle Kopfhörer-Übertragungsfunktionen (HPTFs) entzerrt, wenn individuelle BRIRs verwendet werden. HPTFs aus dem Head-and-Torso-Simulator werden verwendet, wenn künstliche BRIRs verwendet werden. In-Ear-Mikrofone werden zur Messung individueller BRIRs und HPTFs am Eingang des blockierten Gehörgangs verwendet. Die Mikrofone werden zwischen den BRIR und HPTF Messungen nicht entfernt. Die Inversion einer HPTF wird nach der Methode der kleinsten Quadrate berechnet. Für die Wiedergabe wird ein Stax Lambda Pro Kopfhörer verwendet, der die Anforderungen an offene Kopfhörer erfüllt.

**Räumliche Szenenkomplexität** - Die Testkonditionen beinhalten unterschiedliche Audioszenen mit verschiedener Anordnung einer Zielschallquelle und einer, zwei oder drei Störschallquellen. Abbildung 4.24 gibt einen Überblick über die verwendeten Szenen. Szene A besteht aus einer Zielschallquelle in verschiedenen Richtungen. Szene B beinhaltet eine seitliche, frontale oder rückseitige Zielschallquelle und einer richtungsvariablen Störschallquelle. Szene C beinhaltet eine Zielschallquelle und zwei seitliche Störschallquellen. Die Szenen werden als dünn besetzte Szenen angesehen. Die einzelnen Schallquellen können vom Menschen gut lokalisiert und voneinander getrennt wahrgenommen werden. Das Ziel der Untersuchung ist herauszufinden, ob verschiedene Schallquellrichtungen und Anzahl an Schallquellen im jeweiligen resynthetisierten Raum einen unterstützenden oder begrenzenden Einfluss auf die Wahrnehmung der Qualitätsmerkmale besitzen.

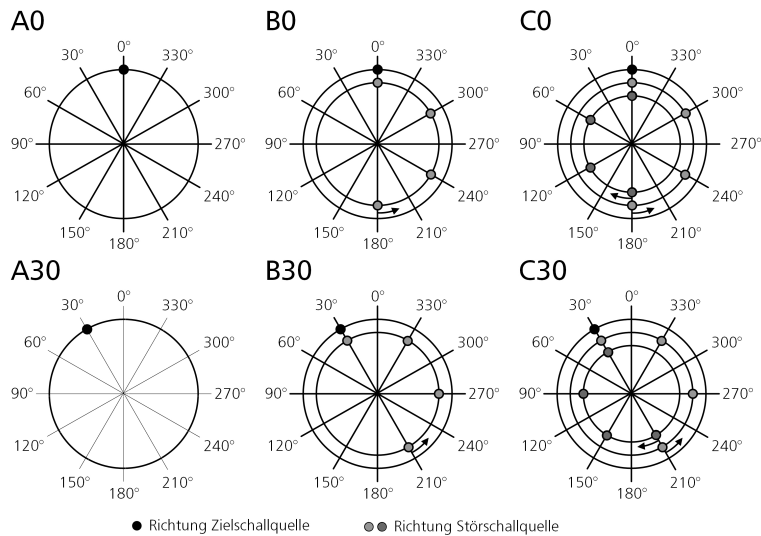


Abbildung 4.24: Audioszenen mit verschiedener räumlicher Komplexität; zusätzliche Szenen durch Spiegelung an der Interauralenachse gespiegelt (weiterführende Details in Tabelle 4.6).

Für Untersuchungen von Quadrantenfehlern in der Lokalisation werden an der interauralen Achse gespiegelte Szene mit frontaler und rückseitiger Ausrichtung verwendet. Die Positionen der einzelnen Schallquellen sind durch ihren Azimuthwinkel und den Radius bestimmt. Ein Winkel von  $0^\circ$  wird für die Elevation gewählt. Die möglichen horizontalen Richtungen liegen in  $60^\circ$  Schritten auf einen Kreis mit dem Radius 2,5 m um die Hörposition. Eine detaillierte Auflistung der Winkel von Zielschallquelle und Störschallquellen für die einzelnen Szenen findet sich in Tabelle 4.6

Als Audiosignale kommen eine weibliche Sprechstimme und zwei männliche Sprechstimmen zur Anwendung. Die Dauer der Signale beträgt 6 s. Die weibliche Sprechstimme wird im Versuch als die Zielschallquelle definiert. Die beiden männlichen Sprechstimmen stellen die Störsignale dar. Es werden dänische Sprecher ausgewählt, um eine eventuelle Beeinflussung durch den Inhalt des Gesagten zu minimieren.

**Raumakustische Disparität** - Es werden drei unterschiedliche Räume zur Resynthese verwendet. Die Räume unterscheiden sich vor allem hinsichtlich ihrer raumakustischen Eigenschaften. Die Räume sind das Hörlabor (HL; siehe dazu Kapitel 2.5), ein leerer Seminarraum (Hu201; siehe dazu Kapitel 2.5) und ein leerer Vorlesungssaal (LdV; siehe dazu Kapitel 2.5) an der TU Ilmenau. Sie unterscheiden sich hinsichtlich verschiedener raumakustischer Parameter wie Nachhallzeit, Direktschall- zu Nachhallenergieverhältnis, interauraler Korrelation und anderen. Eine Zusammenfassung des zeitlichen Verhaltens des Nachhalls ist in Abbildung 4.25 als Energieabklingkurven dargestellt. Der LdV besitzt einen längeren Nachhall als das HL mit dem geringsten Nachhall. Der Raum HL wird als Abhörraum verwendet.

Tabelle 4.6: Audioszenen und Schallquellrichtungen; ZQ=Zielschallquelle, SQ=Störschallquelle; linker Teil = frontale Richtung, rechter Teil = rückseitige Richtung.

Szene	ZQ	SQ1	SQ2	Szene	ZQ	SQ1	SQ2
A0	0°	-	-	A180	180°	-	-
A30	30°	-	-	A150	150°	-	-
B0_1	0°	180°	-	B180_1	180°	0°	-
B0_2	0°	240°	-	B180_2	180°	300°	-
B0_3	0°	300°	-	B180_3	180°	240°	-
B0_4	0°	0°	-	B180_4	180°	180°	-
B30_1	30°	210°	-	B150_1	150°	330°	-
B30_2	30°	270°	-	B150_2	150°	270°	-
B30_3	30°	330°	-	B150_3	150°	210°	-
B30_4	30°	30°	-	B150_4	150°	150°	-
C0_1	0°	180°	180°	C180_1	180°	0°	0°
C0_2	0°	240°	120°	C180_2	180°	300°	60°
C0_3	0°	300°	60°	C180_3	180°	240°	120°
C0_4	0°	0°	0°	C180_4	180°	180°	180°
C30_1	30°	210°	210°	C150_1	150°	330°	330°
C30_2	30°	270°	150°	C150_2	150°	270°	30°
C30_3	30°	330°	90°	C150_3	150°	210°	90°
C30_4	30°	30°	30°	C150_4	150°	150°	150°

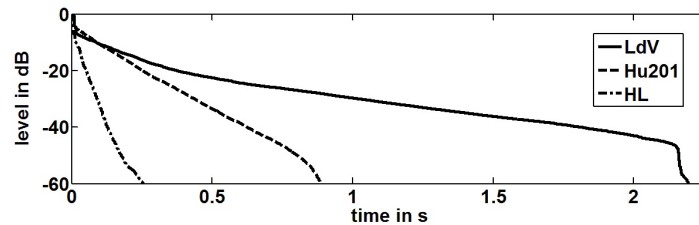


Abbildung 4.25: Energieabklingkurven für die verwendeten Räume bei 0° Beschallungsrichtung unter Verwendung eines omnidirektionalen Messmikrofons.

Zur Messung der BRIRs werden Lautsprecher des Typs Genelec 1030A verwendet. Die Lautsprecher sind an den Stundenpositionen eines Uhrenkreises mit einem Radius von 2,5 m positioniert und auf den Mittelpunkt ausgerichtet. Individuelle BRIRs der Testpersonen werden lediglich im Hörlabor gemessen. BRIRs unter Verwendung eines Kopf- und Torso-Simulator (Typ KEMAR) werden in allen drei Räumen gemessen.

#### 4.3.1.3 Versuchsdurchführung

**Testpersonen** - Drei Frauen und zwanzig Männer, mittleres Alter von 27 Jahren (Standardabweichung 3 Jahre), nehmen als Testpersonen an dem ersten und zweiten Untertest teil. 70%

der Teilnehmenden berichten von Erfahrung mit perzeptiven Untersuchungen. Zwei Frauen und dreizehn Männer, mittleres Alter von 28 Jahren (Standardabweichung 3 Jahre), nehmen als Testpersonen am dritten Untertest teil. Alle Teilnehmenden berichten von Erfahrung mit perzeptiven Untersuchungen.

**Durchführung** - Abbildung 4.23 gibt einen Überblick über die im Test verwendeten Kombinationen von synthetisiertem Raum und Abhörraum. Nur das Hörlabor („HL“) wird als Abhörraum verwendet. Die im Test abgefragten Raumkombinationen sind somit: „HL in HL“, „LdV in HL“, „Hu201 in HL“. Die Teststimuli ergeben sich aus den Kombinationen der drei Aufnahmeräume, drei Audioszenen mit 36 Unterszenen, individuellen BRIRs und Kunstkopf-BRIRs.

Der Hörtest ist in Untertests unterteilt. Der erste Test ist ein Single-Stimulus-Hörtestdesign zur Bewertung der Qualitätsmerkmale Lokalisation und Externalität. Als Stimuli werden die Resynthese der verschiedenen Audioszenen aus den drei Räumen bei Verwendung individueller und Kunstkopf BRIRs präsentiert. Die Bewertung der Schalleinfallrichtung wird durch Anzeigen der Richtung auf der Bewertungsoberfläche realisiert (siehe Kapitel 3.1.3 Abbildung 3.1). Die Bewertungsoberfläche ist eine Draufsicht mit einem symbolisierten Kopf als Mittelpunkt. Verschiedene Regionen sind durch den Probanden zur Bewertung der wahrgenommenen Richtung und Externalität auswählbar. Die wahrgenommene Lokalisierung und Externalität wird in dem Hörversuch für die Zielschallquelle durchgeführt. Die Gestaltung der Oberfläche folgt früheren Untersuchungen, wie sie bspw. in [20] beschrieben wird.

Die Bewertung der Externalität eines auditorischen Ereignisses wird durch Auswahl der inneren, mittleren oder äußeren Region auf der Oberfläche durchgeführt. Nachstehende englische Definitionen werden in dem Test verwendet: a) midpoint: „The auditory event is entirely in my head and very diffuse.“ b) inner circle: „The auditory event is entirely in my head and easy to locate.“; c) middle circle: „The auditory event is external but it is next to my ears or head.“; d) outer circle: „The auditory event is external and easy to locate.“; e) outer cloud: „The auditory event is external and very diffuse“. Die Definitionen werden den Testpersonen in der Versuchseinführung schriftlich vorgelegt. Zusätzlich werden diese vor dem Hörversuch nochmals mündlich an der Bewertungsoberfläche genannt. In dem Versuch werden die oben genannten Definitionen in englischer Sprache verwendet. Die Gruppe der Testpersonen setzt sich aus Personen unterschiedlicher sprachlicher Herkunft zusammen. Alle Testpersonen verstehen die englische Sprache. Keiner der Testpersonen ist englischer Muttersprachler.

Der zweite und dritte Untertest ist ein Hörversuch zur Bewertung der räumlichen Präsenz und der Nachhallumhüllung innerhalb eines referenzfreien Multi-Stimulus-Testdesigns. Die Nachhallumhüllung wird für die verschiedenen raumakustisch unterschiedlichen Räume untersucht. Die Bewertung der räumlichen Präsenz wird für die verschiedenen Szenen und somit räumlichen Komplexitäten durchgeführt. In beiden Hörtests werden ausschließlich Stimuli unter Verwendung der Kunstkopf-BRIRs verwendet. Eine individuelle Messung der BRIRs für

die drei Räume und alle Testpersonen ist zeitlich und organisatorisch schwer durchführbar und es wurde daher darauf verzichtet. Im Gegensatz zum ersten Untertest sollen im zweiten und dritten Test die gesamte Audioszene durch die Testpersonen bezüglich der Qualitätsmerkmale bewertet werden.

#### 4.3.1.4 Ergebnisse

**Externalisierung** - Die Bewertungen des Qualitätsmerkmals Externalität werden als Häufigkeiten gezählt und als Index dargestellt. Der Externalitätsindex berechnet sich aus dem Verhältnis zwischen der Anzahl der Bewertungen für „extern“ und der Gesamtanzahl an Bewertungen. Ein Index von „0“ deutet auf eine komplette Im-Kopf-Lokalisation hin. Ein Index von „1“ deutet auf vollständig externe Wahrnehmung des Hörereignisses hin.

Der obere Teil der Abbildung 4.26 zeigt die Externalitätsindizes für die Szenen A, B und C bei Synthese des Hörlabors (HL) und bei Verwendung individueller und Kunstkopf-BRIRs. Die Bewertungen zeigen einen Anstieg des Externalitätsindex bei Verwendung einer personalisierten Binauralsynthese gegenüber der Verwendung von Kunstkopf-BRIRs. Geringere Indizes sind für die Richtungen  $0^\circ$  und  $180^\circ$  gegenüber den seitlichen Richtungen erkennbar. Die ermittelten Indizes sind vergleichbar mit Indizes aus vorangegangenen eigenen Untersuchungen (siehe [20] und Abschnitt 4.2). Die verschiedenen untersuchten Szenen zeigen keinen signifikanten Einfluss auf die wahrgenommene Externalität ( $\alpha = 5\%$ , Wilcoxon Vorzeichenrangtest). Es werden weiterhin keine signifikanten Unterschiede bezüglich der Anzahl an Schallquellen und ihrer räumlichen Verteilung gefunden. Im oberen Teil der Abbildung 4.26 sind daher die Bewertungen je einzelner Szene zusammengefasst.

**Quadrantenfehler** - Die Bewertungen des Qualitätsmerkmals Lokalisierung werden als Richtungsangaben gezählt. Die Winkelauflösung der Bewertungsfläche liegt bei  $30^\circ$  an den Stundenpositionen eines Uhrenkreises. Zur weiteren Analyse wird ein Quadrantenfehlerindex berechnet. Der Index ergibt sich aus dem Verhältnis zwischen der Anzahl an Quadrantenvertauschungen an der interauralen Achse und der Gesamtanzahl an Richtungsbewertungen. Ein Index von „0“ deutet auf keine Richtungsvertauschungen hin. Ein Index von „1“ deutet auf vollständige Vertauschung der Richtungen bei allen Teststimuli hin. Der untere Teil der Abbildung 4.26 zeigt die Indizes des Quadrantenfehlers für die Szenen A, B und C bei Synthese des Hörlabors (HL) und bei Verwendung individueller und Kunstkopf-BRIRs.

Die Bewertungen zeigen einen Anstieg der Anzahl an Quadrantenfehler für die Richtungen  $0^\circ$  und  $180^\circ$  im Vergleich zu den anderen Richtungen bei Verwendung einer nicht personalisierten Binauralsynthese. Geringere Quadrantenfehler sind bei einer personalisierten Binauralsynthese im Vergleich zur Verwendung von Kunstkopf BRIRs zu erkennen. Diese Ergebnisse sind konform zu anderen Lokalisationsuntersuchungen [57, 154, 20]. Ähnlich zu den Bewertungen für das Qualitätsmerkmal Externalität lassen sich keine signifikanten Unterschiede für das Auftreten

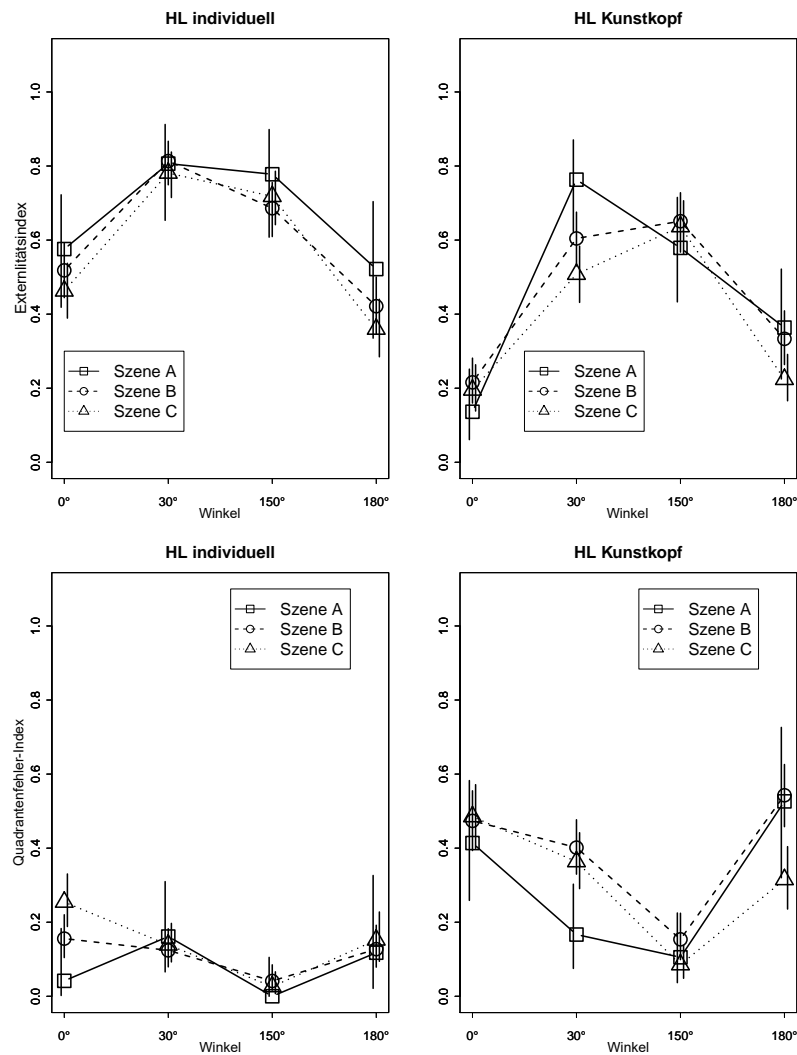


Abbildung 4.26: Externalitätsindizes (oben) und Quadrantenfehler-Index (unten) mit 95% binominal Konfidenzintervallen für Synthese des Hörlabors (HL) und der Szenen A, B, C, für verschiedene Quellrichtungen, bei Verwendung eines Kunstkopfes (KEMAR) und individuellen BRIRs im Abhörraum Hörlabor (HL).

von Quadrantenfehlern aufgrund der Anzahl und räumlichen Anordnung der Schallquellen feststellen.

**Räumliche Präsenz** - Zum Vergleich der verschiedenen Audioszenen untereinander wird ein referenzfreier Multi-Stimulus Test durchgeführt. Jeder Vergleich besteht aus den drei Audioszenen A, B und C für die verschiedenen Richtungen. Neben den genannten Stimuli werden zusätzlich Audiosignale mit geringer räumlicher Qualität als Ankersignale mit abgefragt. Das Ankersignal ergibt sich aus den Anordnungen der Schallquellen gemäß der Szene A (eine Schallquelle) aber unter Verwendung von Freifeld-HRTFs des Kopf- und Torsosimulators. Die Bewertungen können auf einer Skala von „0“ bis „100“ vorgenommen werden. Die Skalen-

endpunkte begrenzen das Qualitätsmerkmal räumliche Präsenz von „less present“ bis „more present“ ab. Aufgrund der unterschiedlichen sprachlichen Herkunft der Probanden werden die Benennungen in englischer Sprache realisiert. Alle Testpersonen beherrschen die englische Sprache ausreichend. In der Auswertung der Bewertungen werden keine signifikanten Unterschiede zwischen den verschiedenen Richtungen ermittelt ( $\alpha = 5\%$ , Wilcoxon Vorzeichenrangtest). Für die weitere Analyse werden die Bewertungen für die verschiedenen Richtungen zusammengefasst. Abbildung 4.27 verdeutlicht die Bewertungen des Qualitätsmerkmals räumliche Präsenz als Median und 25/75 Quartile für die verschiedenen Audioszenen, Anzahl und räumliche Anordnung der Schallquellen.

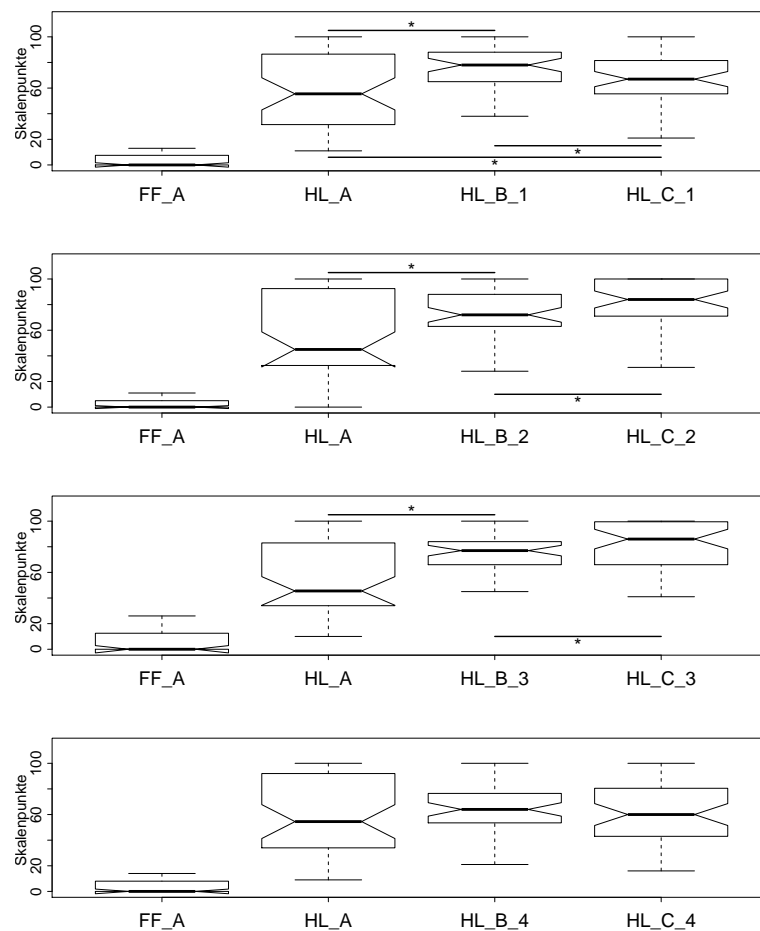


Abbildung 4.27: Räumliche Präsenz als Median und 25/75 Quartile für die verschiedenen Audioszenen und Anzahl an Schallquellen; \* = signifikanter Unterschied bei  $\alpha=5\%$  (Wilcoxon Vorzeichenrangtest); HL=Hörlabor, FF=Freifeld.

Kleine Interquartilabstände (IQAs) sind für die Bewertungen der Ankersignale („FF\_A“) ersichtlich. Die Bewertungen zeigen eine hohe Reliabilität zwischen den Einzelbewertungen der Testpersonen. Die größten IQAs werden für die Audioszene A („HL\_A“) ermittelt. Kein signifikanter Unterschied liegt zwischen den verschiedenen Szenen vor, wenn die Audioobjekte

innerhalb der Szenen sich an der gleichen Position befinden (Konditionen mit Bezeichnung „\_4“). Im Gegensatz dazu werden signifikante Unterschiede zwischen den Szenen mit seitlich angeordneten Audioobjekten ermittelt (Konditionen mit Bezeichnung „\_2“ und „\_3“). Weiterhin existieren signifikante Unterschiede bei Betrachtung der Anzahl der Audioobjekte (bspw. Konditionen mit Bezeichnung „B\_2“ und „C\_2“). Eine geringere räumliche Präsenz wird für die Szene C mit zwei Störschallquellen an der gleichen Position und der Zielschallquelle an der gegenüberliegenden Position im Vergleich zur Szene B mit lediglich zwei Schallquellen an gegenüberliegenden Positionen berichtet (Konditionen mit Bezeichnung „B\_1“ und „C\_1“). Die räumliche Präsenz steigt, wenn die Anzahl an Schallobjekten ansteigt und wenn die Schallobjekte räumlich verteilt um den Hörer angeordnet sind.

**Nachhallumhüllung** - Zur Bewertung der Nachhallumhüllung wird ein referenzfreies Multi-stimulus Testdesign ähnlich wie zur Bewertung der räumlichen Präsenz verwendet. In dem Test vergleichen die Testpersonen jeweils eine Szene bei Resynthese der verschiedenen Räume. Zusätzliche Audioszenen bei Resynthese einer nachhallfreien Situation werden als Ankersignal mit geringer räumlicher Qualität verwendet. Die wahrgenommene Nachhallumhüllung wird für die Ankersignale als gering angenommen, da akustische Informationen des Raumes bei Verwendung von Freifeldbedingungen ausgeschlossen werden können [70].

Die Bewertungen der Nachhallumhüllung erfolgen auf einer kontinuierlichen Skala mit den Eckpunkten „0“ (less enveloped) und „100“ (more enveloped). Bei Auswertung der Bewertungen für die verschiedenen Quellrichtungen zeigen sich keine signifikanten Unterschiede ( $\alpha = 5\%$ , Wilcoxon Vorzeichenrangtest). Zur weiteren Analyse werden die Bewertungen der einzelnen Richtungen zusammengefasst und gemeinsam ausgewertet. Abbildung 4.28 zeigt die Bewertungen als Median und 25/75 Quantile der verschiedenen Szenen, Räume und Anzahl an Schallquellen. Die Bewertungen zeigen, dass das Qualitätsmerkmal Nachhallumhüllung mit dem präsentierten Raum, aber nicht mit der präsentierten Szene ändert. In Übereinstimmung mit Untersuchungen zur Konzerthallenakustik und virtuellen Akustik bei Verwendung von Wellenfeldsynthesensystemen steigt die wahrgenommene Nachhallumhüllung an, wenn der Anteil der Energie in den späten Reflektionen und die Nachhallzeit ansteigt [70, 161]. Nowak et al. konnten zeigen, dass eine Abnahme der Ensemblebreite (im Sinne von Quellenbreite) von  $12^\circ$  auf  $4^\circ$  bei frontaler Ausrichtung keinen Einfluss auf wahrgenommene Nachhallumhüllung hat [161]. In der hier vorgestellten Untersuchung kann gezeigt werden, dass die Nachhallumhüllung sich auch dann nicht signifikant ändert, wenn sich die Schallquellen an verschiedenen Positionen und Anordnungen um den Hörer herum befinden.

### 4.3.2 Zusammenfassung

Es wurde der Einfluss von kontextabhängigen Qualitätsparametern auf die Wahrnehmung von Hörereignissen bei Wiedergabe von Audioszenen über ein Binauralsynthesystem untersucht. Die in dem Test untersuchten Räume zeigen keinen Einfluss auf die Qualitätsmerkmale



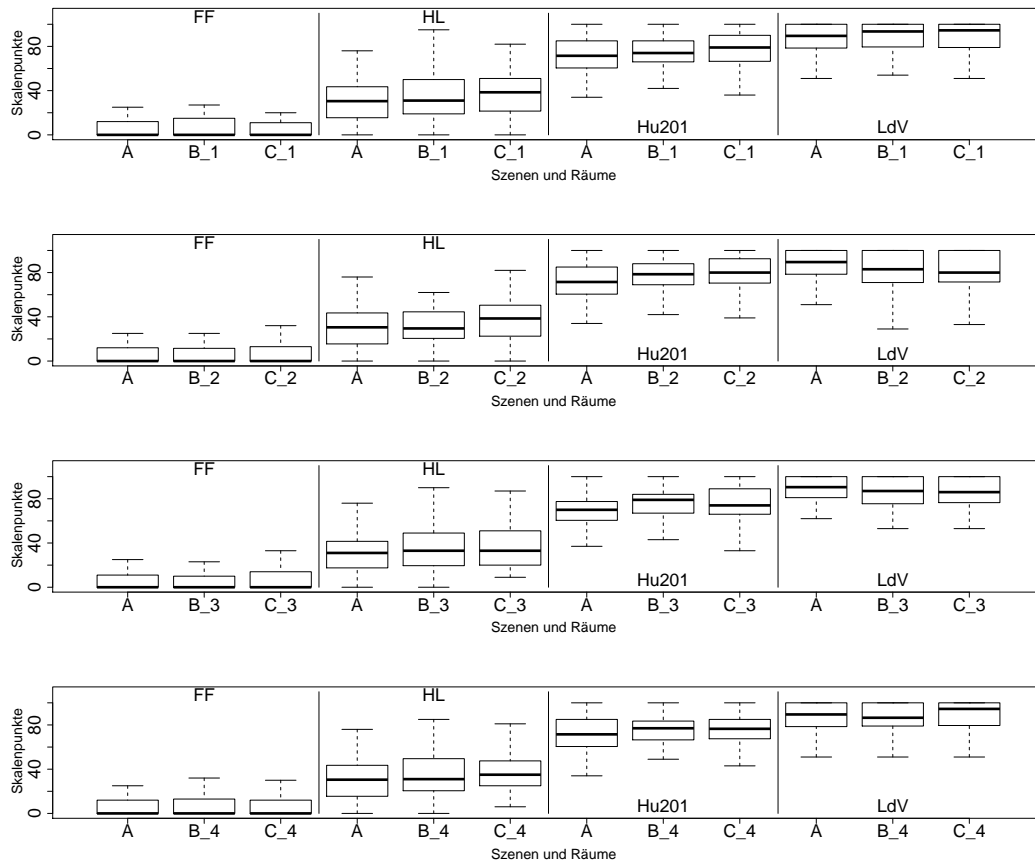


Abbildung 4.28: Nachhallumhüllung als Median und 25/75 Quantile für die verschiedenen Audioszenen und Räume; FF=Freifeld, HL=Hörlabor, Hu201=Seminarraum, LdV=Hörsaal.

Externalität, Lokalisation oder Nachhallumhüllung. Dahingegen zeigt sich ein Effekt auf das Qualitätsmerkmal räumliche Präsenz. Räumlich stärker verteilte Szenen lassen die Wahrnehmung von räumlicher Präsenz ansteigen. Die Ausrichtung der Szenen in frontaler, rückwärtiger oder seitlicher Richtung zeigt dahingegen keinen Einfluss.

Die untersuchte raumakustische Disparität zeigt einen Einfluss auf die Wahrnehmung der Nachhallumhüllung. Halligere Räume werden mit höherer Nachhallumhüllung bewertet als weniger hallige Räume. Die Anordnung der Audioobjekte in den verschiedenen Audioszenen in frontalen, rückwärtigen oder seitlichen Richtungen zeigen keinen signifikanten Einfluss auf die Wahrnehmung der Nachhallumhüllung.

Die Ergebnisse des Experiments für die Qualitätsmerkmale Lokalisation und Externalisierung liegen in der Messgenauigkeit mit Experimenten von früheren Untersuchungen trotz verschiedener Testdesigns und Untersuchungskonditionen (siehe [57, 154, 20] und andere). Es wurden keine signifikanten Einflüsse der untersuchten Szenen und Anzahl an Audioobjekten auf das Auftreten von Quadrantenfehlern und wahrgenommener Externalität festgestellt. Es muss allerdings beachtet werden, dass die untersuchten Szenen, als dünn besetzte Szenen angesehen

werden müssen in denen eine Lokalisierung und Separierung der einzelnen Objekte problemlos möglich ist. Weiterführende Untersuchungen mit dichter besetzten Szenen, in welchen die Lokalisierung und Separierung der einzelnen Schallobjekte nicht mehr einfach möglich ist, sind notwendig. Es bleibt unklar wie räumliche Präsenz wahrgenommen wird, wenn keine klar lokalisierbaren einzelnen Audioobjekte vorliegen.

Die Ergebnisse der vorgestellten Untersuchung und die Ergebnisse von früheren Untersuchungen zeigen eine sehr gute Reliabilität der untersuchten Qualitätsmerkmale. Die Qualitätsmerkmale scheinen geeignet zur Beschreibung der räumlichen Wahrnehmung im Kontext eines binauralen Resynthesystems zu sein. Allerdings muss die wahrgenommene räumliche Präsenz als ein mehrdimensionales Qualitätsmerkmal angesehen werden. Ist die Audioszene im Test dünn besetzt (Szene A) lassen sich deutlich weniger konsistente Bewertungen über die Testpersonen feststellen. Die Testpersonen berichten in einem sich dem Test anschließenden informellen Interview, dass sie für diese Szene andere Qualitätsmerkmale wie Lokalisation, Externalität, Räumlichkeit und Eingehülltsein bewertet hätten. Räumliche Präsenz beschreibt somit wahrscheinlich eine Ansammlung von Untermerkmalen. Welche diese sind gilt es in weiterer Forschung herauszufinden.

## 4.4 Einfluss von Kopfbewegungen auf die Externalisierung

Das Ziel von binauralen Kopfhörersystemen ist die Reproduktion von Hörereignissen einer echten akustischen Umgebung oder die Synthese virtuell erzeugter Audioszenen. In dem hier vorgestellten Experiment wird sich auf den Einfluss der Kopfbewegung und der Kopfpennachverfolgung konzentriert. Es wird untersucht, inwieweit Kopfbewegungen bei an- und abgeschalteter Nachverfolgung die Wahrnehmung von Externalität bei Raumdivergenz bzw. Raumkonvergenz beeinflusst.

Head-Tracking ist ein Merkmal eines dynamischen binauralen Synthesesystem, das ein zur Kopfbewegung des Benutzers korrekte Signalgenerierung ermöglicht. Die in der virtuellen Szene abzubildenden Audioobjekte werden aus der perspektivisch korrekten Richtung synthetisiert. Es ist damit ein relevanter Baustein für die Schaffung einer plausiblen Resynthese eines räumlichen Hörereignisses. Die Implementierung von Head-Tracking führt zu einer Verringerung der Vertauschung zwischen Vorder- und Rückseite bei Lokalisierungsaufgaben [57]. Diese Verringerung lässt gemäß der Hypothese, dass die wahrgenommenen Wahrnehmungsmuster mit den gespeicherten Mustern im Gehirn übereinstimmen müssen [168], eine Verringerung der Im-Kopf-Lokalisation erwarten [29].

In den Arbeiten von Begault und Wenzel [57] wird jedoch kein signifikanter Effekt auf die Externalisierung berichtet. Es bleibt in dieser Arbeit allerdings unklar, welche Art und welches Ausmaß die Kopfbewegung hatten. Es ist weiterhin nicht eindeutig, ob eine Kopfbewegung des Hörers in dem Fall, dass kein Head-Tracking verwendet wird auch stattfindet. In einem Experiment von Brimijoin et al. [71] wird die Wirkung von Head-Tracking unter raumkonvergenten, aber nicht unter raumdivergenten Hörbedingungen untersucht. Die dort angewendete dynamische Binauralsynthese führt zu einer signifikant höheren Externalisierung im Vergleich zu einer statischen Synthese, wenn eine Kopfbewegung vorliegt. Sie kommen zu dem Schluss, dass die Position eines Audioobjekts während der Kopfbewegung an einem festen Ort im Raum liegen muss.

### 4.4.1 Exp. V: Head-Tracking

Das folgende Experiment untersucht den Einfluss von erbetenen Kopfbewegungen auf die Wahrnehmung von Externalität von Hörereignissen. Die weiteren Testvariablen sind zum einen das Vorhandensein eines Head-Tracking innerhalb des Binauralsynthesystems und zum anderen Divergenz bzw. Konvergenz zwischen synthetisierter Audioszene und Abhörraum. Die Untersuchung zeigt, dass die Externalisierung der Hörereignisse aus frontalen Richtungen bei Kopfbewegung und Head-Tracking ansteigt. Der in Kapitel 4.2 dargelegte Raumdivergenzeffekt konnte durch die Kopfbewegungen und Head-Tracking nicht aufgelöst werden. Weiterführende Informationen sind der dazugehörigen Veröffentlichung [17] sowie der durch Georg Götz durchgeführten Bachelorarbeit [35] zu entnehmen .

#### 4.4.1.1 Untersuchungsgegenstand

Das Experiment ermittelt den Unterschied in der Wahrnehmung von Externalität zwischen Hörereignissen mit und ohne Head-Tracking und mit und ohne erzwungener Kopfbewegung. Die synthetisierten Audioobjekte sind Lautsprecherpositionen sowie 5-kanal stereogepannte Audioobjekte in einer 5.0 Lautsprecheraufstellung.

Nachstehende Fragen sollen beantwortet werden: Wird durch die Anwendung eines Head-Tracking die Externalisierung der Hörereignisse bei Raumkonvergenz und bei Raumdivergenz erhöht? Führt eine Hinzunahme von auditiven Merkmalen basierend auf Kopfdrehungen zu einer Auflösung des Raumdivergenzeffektes?

#### 4.4.1.2 Versuchsaufbau

**Binauralsynthesesystem** - Als binaurales Synthesesystem kommt der Smyth Realiser A8 zum Einsatz [198]. Es handelt sich hierbei um ein kommerzielles System zur Auralisierung von Mehrkanallautsprecheraufstellungen mittels dynamischer Binauralsynthese. Die Kopfdrehungen werden via Infrarottracking erfasst. Die dafür benötigten Geräte sind am Kopfhörer und am Center-Lautsprecher angebracht. Die für jeden Lautsprecher benötigten BRIRs werden für die Blickrichtungen  $30^\circ$ ,  $0^\circ$  und  $30^\circ$  individuell gemessen. Zwischen diesen Richtungen werden BRIRs mittels Interpolation berechnet. Dies ermöglicht eine Kompensation der Kopfdrehung im Bereich des Stereodreiecks. Die Anpassung an die Ausrichtung des Kopfes erfolgt in Echtzeit und hat eine Genauigkeit von  $0,25^\circ$  [198].

Neben der Individualisierung der BRIRs bietet der Smyth Realiser A8 eine Kompensation der genutzten Kopfhörer durch Messung und Entzerrung der individuellen HPTF. Für die Untersuchung wird ein der extra-auraler Kopfhörer BK211 verwendet [85]. Durch die akustische Offenheit des Systems wird ein direkter Vergleich von realen Lautsprechern und binaural synthetisierten Audioobjekten ermöglicht.

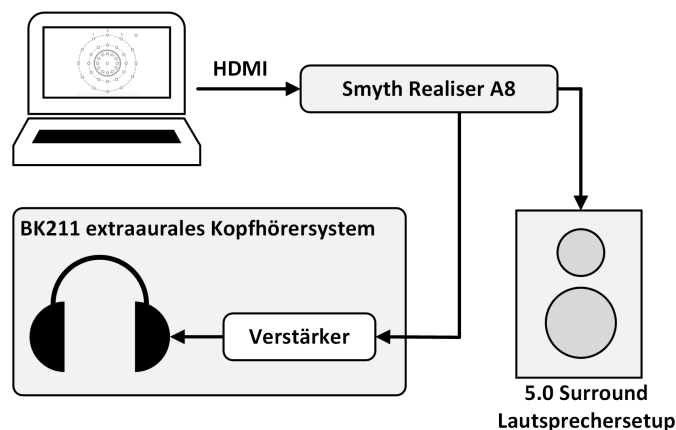


Abbildung 4.29: Schematische Darstellung des im Hörversuch verwendeten Systems.

Abbildung 4.29 verdeutlicht die einzelnen Systemkomponenten schematisch. Der Smyth Realizer dient zur Aufnahme der individuellen BRIRs und zur Resynthese des verwendeten 5.0 Lautsprecher setups.

**Raumdivergenz** - Zur Realisierung der in der Untersuchung gewünschten Raumdivergenz zwischen resynthetisiertem Raum und dem Abhörraum wird ein Hörlabor (HL; siehe dazu Kapitel 2.5) und ein leerer Seminarraum (SR; siehe dazu Kapitel 2.5) für die Messungen der BRIRs und als Hörtestraum verwendet. Die beiden Räume besitzen sehr verschiedene raumakustische Charakteristika, wie bspw. Nachhallzeiten und DRR (siehe Kapitel 2.5).

In dem verwendeten Audiosystem Smyth Realizer A8 wurde keine Möglichkeit gefunden, um Voreinstellungen zum verwendeten Raum zu programmieren und somit im Hörversuch zwischen Räumen hin und her zu schalten. Dementsprechend werden die Teilnehmer gemäß dem wiederzugebendem Raum in zwei Gruppen aufgeteilt. Die BRIRs für die erste Gruppe (konvergente Gruppe) werden im Seminarraum gemessen. Der Raum und die BRIRs der zweiten Gruppe (divergente Gruppe) werden im Hörlabor gemessen. Beide Gruppen führen den Hörtest im Seminarraum durch.

**Objektpositionen** - Der Smyth Realizer A8 wird für die Synthese eines 5.0 Surround-Sounds verwendet. Die Lautsprecheranordnung entspricht der in Rec. ITU-R BS.775-2 [117] mit Surround-Lautsprechern bei  $110^\circ$  und  $250^\circ$ . Weiterhin werden zwei Lautsprecheratrappen bei  $15^\circ$  und  $345^\circ$  platziert. Diese dienen der visuellen Unterstützung der von den Probanden verlangten Kopfdrehungen.

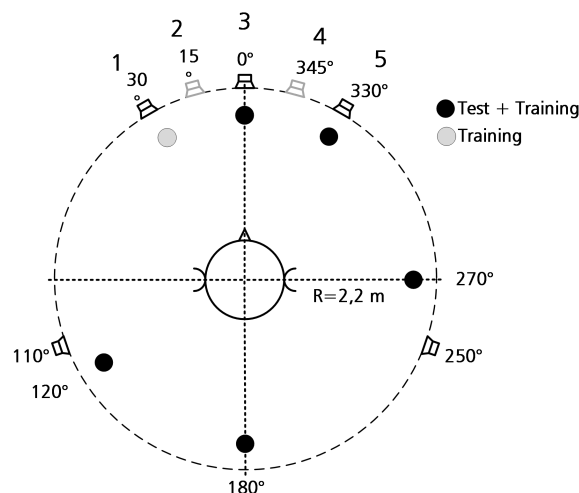


Abbildung 4.30: Lautsprechersetup und Audioobjekte; 5.0 Lautsprecher-Setup durch schwarze Lautsprechersymbole angezeigt; Lautsprecheratrappen mit grauen Markierungssymbole; synthetisierte Audioobjekte für Training und Test als graue und schwarze Kreise (Abstand bei  $R = 2,2 \text{ m}$ ); die Lautsprecher sind durch Nummern von 1 bis 5 benannt.

Abbildung 4.30 zeigt neben den Lautsprecherpositionen auch die verwendeten Positionen der Audioobjekte innerhalb des Lautsprecher-Setups. Der Abstand der Lautsprecher zur Mess- und Abhörposition des Hörers beträgt ca. 2,2 m. Die Höhe der Lautsprecher beträgt ca. 1,3 m (Ohrposition einer sitzenden Person). Die Positionen der Audioobjekte wurden gewählt, um frontale, laterale und dorsale Richtungen und Schallquellen für den Hörversuch bereitzustellen.

**Kopfbewegung** - Die Hauptherausforderung im Testdesign besteht darin, die Testpersonen zu einer kontrollierten und reproduzierbaren Kopfdrehung zu animieren. Vor jeder Präsentation des Teststimulus wird eine Zahl als Sprachsignal dargeboten. Die Zahlen entsprechen den Markierungen der Lautsprecher vor dem Probanden (siehe Abbildung 4.30). Die Frontallautsprecher sind von  $-30^\circ$  bis  $30^\circ$  in  $15^\circ$  Schritten angeordnet. Die Probanden sind aufgefordert, ihren Kopf zwischen den genannten Lautsprecherpositionen hin und her zu drehen. Der Winkel der Kopffrotation variierte zwischen  $15^\circ$  und maximal  $60^\circ$ . Als zusätzliche Hilfestellung zur Beurteilung der Blickrichtung ist auf dem Kopfhörer ein Laserpointer aufgesetzt. Dies ermöglicht den Probanden ein genaues Zielen auf den jeweiligen Lautsprecher bzw. Attrappe. Die Geschwindigkeit der Kopfbewegung ist durch den Probanden frei wählbar.

**Audiosignale** - Die dargebotenen Audiosignale sind entweder ein Sprachsignal (ca. fünf Sekunden Länge, männlicher niederländischer Sprecher) oder eine Saxophon-Aufnahme (ca. neun Sekunden Länge). Beide Audiosignale sind hallfreie Aufnahmen. Für die Schallquellenrichtungen  $0^\circ$  und  $30^\circ$  stimmen die Lautsprecher des 5.0 Setups mit der Richtung der Synthese überein. Für die anderen synthetisierten Richtungen wird ein intensitätsbasiertes Surround-Panning angewendet, um mehrkanalige Audiodateien zu erzeugen. Diese werden als Eingangssignale für das Smyth-System verwendet.

Der Hörtest selber ist ein Einzel-Stimulus-Test mit zufälliger Reihenfolge der Teststimuli für jede Testperson. Die Bewertung der Stimuli erfolgt gemäß der in Kapitel 3.1.3 beschriebenen und in Abbildung 3.1 dargestellten Benutzeroberfläche. Im Hörtest sind die Teilnehmer gebeten, die Externalisierung und die wahrgenommene Einfallrichtung der jeweiligen Teststimuli zu bewerten. Der wahrgenommene Einfallswinkel wird durch die Wahl der jeweiligen diskreten Richtung in einer Draufsicht ermöglicht. Die Externalisierung wird durch die Wahl des Mittelpunktes des inneren oder äußeren Kreises ermöglicht. Die Skalierung und die Definition des Qualitätsmerkmals Externalisierung ist im Detail Kapitel 3.1.3 zu entnehmen.

#### 4.4.1.3 Versuchsdurchführung

Das gesamte Hörexperiment teilt sich in zwei Testtage auf. Die individuellen BRIRs der Probanden werden für beide Räume am Testtag eins gemessen. Der zweite Testtag umfasst den eigentlichen Hörversuch mit den Bewertungen durch den Probanden. Während des Hörversuchs (zweiter Termin) werden verschiedene Wiedergabelisten in zufälliger Anordnung zwischen

den Probanden verwendet. Abbildung 4.31 verdeutlicht die verwendeten Listen schematisch. Insgesamt ergeben sich sechs Wiedergabelisten. Die gesamte Hörtestdauer beträgt inkl. einer kurzen Pause zwischen Teil eins und zwei ca. 60 Minuten.

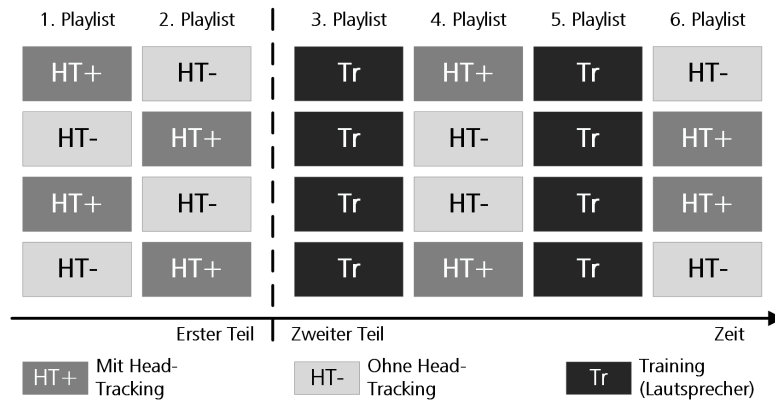


Abbildung 4.31: Struktur der im Test verwendeten Wiedergabelisten.

Der Bewertungsteil besteht aus zwei direkt aufeinander folgenden Sitzungen. In beiden Sitzungen werden die beiden Räume und die fünf Richtungen mit und ohne Head-Tracking den Probanden dargeboten. Die erste Sitzung ist durch eine Gewöhnung der Probanden an die Audiosignale, Richtungen, erzwungen Kopfdrehung und Bewertung motiviert. Zusätzlich dienen die Ergebnisse der ersten Sitzung als Vergleichsgrundlage mit früheren Studien zum Raumdivergenzeffekt. Die Probanden sind in der ersten Sitzung noch nicht oder zumindest unbestimmt trainiert und haben ihre anfängliche individuellen Erwartungshaltung an die dargebotenen Stimuli. Die Testpersonen sind nicht gemäß konvergenter oder divergenter Raumkondition vorgeprägt. Die Unterschiede zwischen den Sitzungen ermöglichen dahingegen Einblicke in mögliche Anpassungseffekte.

Die zweite Sitzung beinhaltet zwei Trainings-Playlists mit Darbietung über reale Lautsprecher im Abhörraum. Die Erwartungshaltung der Testpersonen soll im Sinne einer Anpassung ihrer internen Referenz bezüglich der Wahrnehmung der realen 5.0 Lautsprecherkonfiguration im Hörraum harmonisiert werden. Dieser Ansatz ist durch die Beobachtung motiviert, dass der Raumdivergenzeffekt in Richtung der kongruenten oder divergenten Raumbedingung verschoben werden kann [8]. Es wird hypothetisiert, dass die im Gehirn gespeicherten Muster aus der Binauralsynthese mit den Mustern aus dem realen Setup überschrieben werden. Die wahrgenommene Externalisierung der binaural resynthetisierten Signale sollte daher für die folgenden Stimuli vermindert werden. Die Fähigkeit zur Differenzierung zwischen Lautsprechern und der Resynthese wird für beide Gruppen höchstwahrscheinlich zunehmen. Es wird erwartet, dass nach jeder Trainings-Playlist ähnliche Ausgangsbedingungen für den anschließenden Bewertungsteil herrschen. Eine exklusive Differenzierung in Bezug auf die Head-Tracking-Bedingung ist beabsichtigt. Weiterhin soll eine zusätzliche Abhängigkeit der

Bewertungen durch ein direktes Aufeinanderfolgen der Stimuli mit und ohne Head-Tracking minimiert werden.

Die Trainings-Playlisten bestehen aus 36 Stimuli in zufälliger Anordnung mit sechs Richtungen, zwei Audiosignalen und drei Iterationen. Die Wiedergabelisten der Binauralsynthese bestehen aus 20 Stimuli in zufälliger Reihenfolge mit fünf Richtungen, zwei Audiosignale und zwei Iterationen.

#### 4.4.1.4 Ergebnisse

Abbildung 4.32 zeigt die Bewertungen der Testpersonen aus der zweiten Sitzung als Externalisierungsindizes für die fünf Richtungen. Externalisierungsindizes von mehr als 0,9 werden für die binaurale Resynthese mittels Head-Tracking und bei Konvergenz zwischen Synthese- und Abhörraum für alle Richtungen beobachtet. Für die konvergente Raumkondition, aber ohne Head-Tracking, ist eine niedrigere Externalisierung sichtbar, insbesondere für die Richtung 0° und 30°. Indizes nahe eins sind für die seitlichen Richtungen von 90° und 240° sichtbar. Die divergente Raumkondition mit Head-Tracking erreicht tendenziell eine höhere Externalisierung als der divergierende Raumzustand ohne Head-Tracking.

Die Verwendung von Head-Tracking erhöht die Externalisierung nur für die Richtungen 0° und 30°. Der Effekt des Head-Trackings (Verhältnis der Wechsler zwischen Externalisierung und Im-Kopf-Lokalisation) ist bei der konvergenten Gruppe höher als bei der divergenten Gruppe. Die Tabelle 4.7 zeigt die Ergebnisse einer statistischen Analyse unter Verwendung des exakten McNemar Tests zwischen dem Zustand mit und ohne Head-Tracking. Die p-Werte und das Verhältnis der Wechsler zwischen den Zuständen werden angezeigt. Das Verhältnis der Wechsler basiert auf den Bewertungen zwischen den Bedingungen mit und ohne Head-Tracking. Werte über eins bedeuten, dass sich mehr Ratings von in-Kopf auf außerhalb-Kopf geändert haben als umgekehrt. Bei Werten unter 1 verringert sich die Anzahl der In-Kopf-Ratings.

Tabelle 4.7: p-Werte und Verhältnis der Wechsler eines exakten McNemar-Tests auf Unabhängigkeit zwischen den Zuständen mit und ohne Head-Tracking für die konvergente und divergierende Gruppe.

Richtung	konvergent		divergent	
	p-value	Wechsler	p-value	Wechsler
0°	<0.001	13.0	<0.001	5.7
30°	<0.001	18.0	<0.001	4.4
90°	0.18	0.3	0.48	1.6
180°	0.77	0.7	0.55	1.8
240°	0.07	0.1	0.80	1.3

Die Analyse der Bewertung zwischen der divergenten und konvergenten Raumkondition ergibt sehr signifikante (exakter Fisher-Test,  $p < 0,001$ ) Unterschiede zwischen den Externalisierungsindizes. Darüber hinaus ist der Raumdivergenzeffekt für die Bedingungen mit und ohne



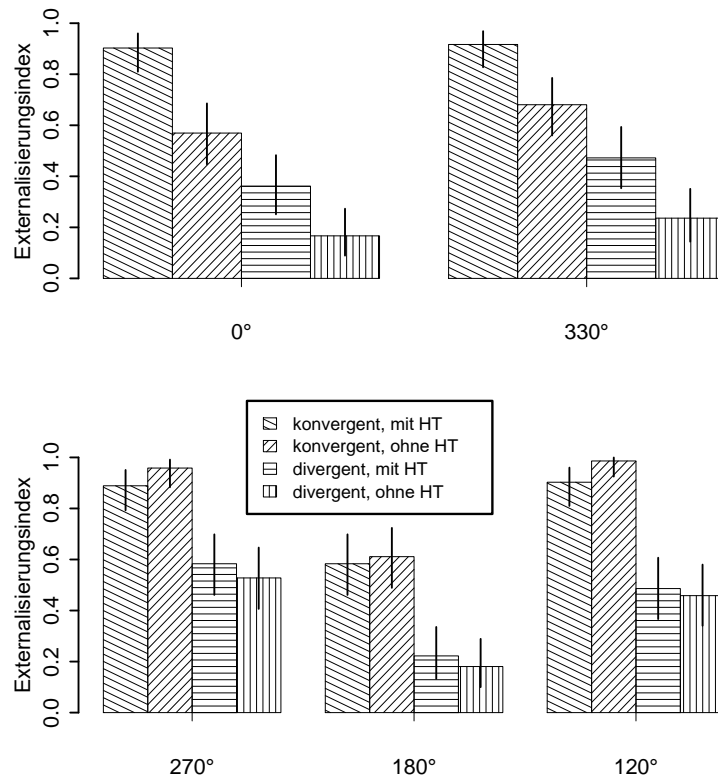


Abbildung 4.32: Externalisierungsindeks mit 95% binomischen Konfidenzintervallen für die verschiedenen Testbedingungen und -richtungen; oben: Ratings für die Resynthese der Quellen auf den Lautsprecherpositionen; unten: Ratings für die Resynthese der Phantomquellen; HT=Head-Tracking.

Kopfnachführung sichtbar. Die Tabelle 4.8 zeigt die p-Werte und die Quotenverhältnisse zwischen den Zuständen. Das Quotenverhältnis ist ein Maß für die Effektgröße. Ein Wert von 1 bedeutet keine Wirkung und Werte größer als 1 entsprechen einer größeren Effektgröße.

Tabelle 4.8: p-Werte und Odds Ratio eines exakten Fisher-Tests zwischen den Ratings der konvergenten und divergenten Gruppe in Abhängigkeit vom Head-Tracking; HT=Head-Tracking.

Richtung	mit HT		ohne HT	
	p-value	Ratio	p-value	Ratio
0°	<0.001	16.1	<0.001	6.5
30°	<0.001	12.1	<0.001	6.8
90°	<0.001	5.6	<0.001	20.2
180°	<0.001	4.8	<0.001	7.0
240°	<0.001	9.7	<0.001	81.6

Der Unterschied zwischen dem Zustand konvergent ohne Kopftracking und divergent mit Kopftracking ist ebenfalls sehr signifikant (exakter Fisher-Test,  $p < 0,001$ ) für die meisten Richtungen. Der abweichende Raumzustand wird mit höheren Externalisierungsindizes bewertet als der abweichende Zustand mit Kopfnachführung. Das Vorhandensein von Head-Tracking für den raumdivergenten Zustand löst den Raumdivergenzeffekt nicht auf. Die Tabelle 4.9 zeigt die p-Werte und die Quotenverhältnisse zwischen den Bedingungen.

Tabelle 4.9: p-Werte und Odds Ratio eines exakten Fisher-Tests zwischen den Ratings der konvergenten Gruppe ohne Head-Tracking und der divergenten Gruppe mit Head-Tracking.

Richtung	p-value	Ratio
0°	0.02	2.3
30°	0.02	2.4
90°	<0.001	16.1
180°	<0.001	5.4
240°	<0.001	73.1

#### 4.4.2 Zusammenfassung

Die Ergebnisse zeigen, dass die Externalisierung bei Kopfbewegungen für frontale Richtungen erhöht werden kann, wenn eine Kompensation mit Head-Tracking vorgesehen ist. Für laterale und dorsale Richtungen konnte kein signifikanter Unterschied zwischen den beiden Head-Tracking-Zuständen festgestellt werden. Es ist derzeit nicht sicher, ob dies an den synthetisierten Schallquellenrichtungen liegt oder daran, dass für diese Richtungen Phantomquellen resynthetisiert werden. Es wird jedoch deutlich, dass die Kompensation von Kopfbewegungen mittels Head-Tracking den Effekt der Raumdivergenz nicht überwinden kann, da es immer noch sehr signifikante Unterschiede gibt.

Für eine Abschätzung des Einflusses der erzwungenen Kopfbewegung auf die Externalisierung, lassen sich die Ergebnisse dieses Experimentes mit früheren Ergebnissen [8] verglichen. Der linke Teil der Abbildung 4.21 zeigt die Ergebnisse eines Hörversuchs, der dem hier vorgestellten ähnlich ist [8, 39]. Es werden dieselben Räume, dieselben Audiosignale, individualisierte binaurale Synthese, ähnliche Richtungen, gleiche Quellen-Empfänger-Abstände und ein ähnliches Testdesign verwendet.

Es zeigen sich ähnliche Größenordnung an Externalität bei raumkonvergenten Bedingungen zwischen der Anwendung von Head-Tracking und erzwungener Kopfbewegung und keinem Head-Tracking und bei nicht erzwungener Kopfbewegung [8]. Dies könnte bedeuten, dass der Mangel an Head-Tracking nur dann kritisch wird, wenn die Kopfbewegung erzwungen wird. Der Einfluss kleiner unbewusster Kopfbewegungen, die nicht verfolgt werden, ist für die Externalisierung möglicherweise unkritisch. Weiterführende Hörtests mit einer Winkelauflösung der Binauralsynthese und eines Head-Tracking kleiner als der gerade noch wahrnehmbare

Winkelunterschied in der Richtungswahrnehmung sowie eine Beachtung kleiner translatorischer Bewegungen und ein Kippen des Kopfes können hier Aufklärung schaffen.

Andererseits kann diese Hypothese bei der Untersuchung der Bewertungen der divergierenden Raumkondition nicht bestätigt werden. In beiden Tests ist die Externalisierung ohne Head-Tracking relativ gering, unabhängig von einer erzwungenen Kopfbewegung. Basierend auf dieser Beobachtung kann davon ausgegangen werden, dass die Externalisierung für divergierende Raumkonditionen ohne Head-Tracking und ohne Kopfbewegung höher ist als für eine divergierende Raumkonditionen ohne Head-Tracking aber mit Kopfbewegung. Dies ist bei diesem Experiment aber nicht der Fall. Weitere Untersuchungen zu diesem Thema scheinen hier notwendig.

#### 4.5 Zusammenfassung kontextabhängige Qualitätsparameter

In den Untersuchungen zum Ventriloquismus-Effekt konnte gezeigt werden, dass der durch die Fusion zwischen visuellen Sehereignis und auditiven Hörereignis, aber aus unterschiedlichen Richtungen kommenden Ereignissen verursachte Winkelabweichung, in der Richtungswahrnehmung des Hörereignisses in vergleichbarer Größenordnung liegt, wie dieser in Abhörbedingungen ohne Binauralsynthese gemessen wurde. Es zeigen sich größere Winkelabweichungen für horizontal und horizontal-vertikal ausgelenkte Darbietungsrichtungen gegenüber frontalen Richtungen. Die Schlussfolgerung ist, dass das verwendete Binauralsynthesystem bzgl. dieses Wahrnehmungsmerkmals valide im Vergleich zur Wahrnehmung ohne ein derartiges technisches System ist.

In den Untersuchungen zum Raumdivergenzeffekt konnte dieser in der Literatur bereits 1972 beschriebene Effekt erneut nachgewiesen und quantitativ erfasst werden. Es wurde zusätzlich untersucht, welche anderen Parameter einen Einfluss auf die Externalisierung von Hörereignissen bei Raumdivergenz haben. Es konnte gezeigt werden, dass die Richtung der zu synthetisierenden Schallquelle einen Einfluss auf die Externalisierung und auf den Raumdivergenzeffekt hat. Der Raumdivergenzeffekt ist besonders stark für frontale und rückseitige Syntheserichtungen ausgeprägt. Er verliert für seitlich ausgelenkte Richtungen an Stärke. Der Raumdivergenzeffekt ist weiterhin bei Darbietung eines weniger akustisch aktiven Raumes (i.d.R. weniger hallig) in einem akustisch aktiven Raum (i.d.R. hallig) stärker als in anderer Richtung. Die Personalisierung der Binauralsynthese hat einen Einfluss auf die Lokalisierungsfähigkeit (siehe Quadrantenfehler). Es wird aber kein Zusammenhang zur Raumdivergenz bzw. Raumkonvergenz erkannt. Im Allgemeinen ist die Externalisierung der Hörereignisse für frontale und rückseitige Syntheserichtungen und bei Verwendung individueller BRIRs größer als bei Verwendung von Kunstkopf-BRIRs. Es hat sich gezeigt, dass in einigen Testkonditionen eine deutliche negative Korrelation zwischen der Auftretenshäufigkeit eines Quadrantenfehlers und der Externalisierung der Hörereignisse besteht. Andererseits konnte dieser Zusammenhang bei einer Binauralsynthese mit schwachen spektralen Merkmalen in den BRIRs nicht bestätigt werden, obwohl die Externalisierung hoch ist. Die Experimente zur Untersuchung des Einflusses der Sichtbarkeit des Abhörortes auf die Externalisierung bei Raumdivergenz ergibt keinen Zusammenhang. Die Sichtbarkeit des Abhörortes und von Lautsprecherattrappen an den zu synthetisierenden Schallobjektpositionen im Raum führt zu einem generellen Anstieg der Externalität der Hörereignisse. Der Raumdivergenzeffekt löst sich dadurch aber nicht auf. Ein interessantes Phänomen ist hierbei, dass die Externalisierung für rückseitige Synthesepositionen bei Sichtbarkeit des Raumes mehr ansteigt als für frontale oder seitliche Positionen.

In den Untersuchungen zur Prüfung des Einflusses der räumlichen Komplexität der zu au-ralisierenden Szene (Anzahl an Schallquellen und räumliche Anordnung) zeigen sich bzgl. der Externalisierung keine Abhängigkeiten von der dargebotenen Szene. Die Anzahl (hier

eins bis drei Quellen) und die räumliche Anordnung dieser Quellen um den Hörer werden bei Verwendung individueller und Kunstkopf-BRIRs nahezu identisch bewertet. Zusätzliche Audioobjekte sind bzgl. der Externalisierung der Hörereignisse anscheinend weder förderlich noch störend. Die Bewertungen zur räumlichen Präsenz zeigen signifikante Unterschiede in Bezug auf die Anzahl der Audioobjekte und deren Anordnung. Audioszenen mit mehreren Objekten und einer gestreuten räumlichen Anordnung um den Hörer, erreichen eine höhere Bewertung der Präsenz als Szenen mit wenigen und/oder gering räumlich gestreuten Audioobjekten. Das auditive Merkmal der Nachhallumhüllung hängt lediglich vom untersuchten Raum (und dessen Halligkeit) aber nicht von der Anzahl oder Anordnung der Audioobjekte ab.

Das Hinzufügen der Möglichkeit zur Verwendung von auditiven Merkmalen durch Kopfdrehung zeigt einen Anstieg der Externalisierung für konvergente und divergente Raumkonditionen bei der Synthese diskreter Lautsprecherpositionen. Es bleibt offen, warum dies nicht für zwischen Lautsprechern stereogepannte Audioobjekte beobachtet wurde. Bezüglich des Raumdivergenzeffektes zeigt sich kein Einfluss in der Hinzunahme derartiger dynamischer Merkmale. In allen untersuchten Fällen wird die raumkonvergente Kondition signifikant höher als die divergente Kondition bewertet. Weiterhin zeigt sich die Hinzunahme von Head-racking nur dann als für eine plausible Synthese notwendig, wenn auch perspektivische Änderungen zwischen Hörerausrichtung und Audioobjekt abzubilden sind. Kleine Kopfdrehungen scheinen keinen Einfluss auf die Externalisierung der Hörereignisse zu haben.

Abschließend zeigt sich, dass verschiedene Parameter der Binauralsynthese die Externalisierung der Hörereignisse beeinflussen. Personalisierung, visuelle Merkmale und eine dynamische Binauralsynthese steigern die Externalisierung. Sie überdecken in manchen Fällen den Raumdivergenzeffekt aber lediglich schlecht. In vielen Fällen lösen sie ihn nicht auf und er bleibt in seinem Prinzip bestehen. Die auditorische Übereinstimmung der synthetisierten Audioszene mit der erwarteten Szene, also Raumkonvergenz, scheint ein notwendiges Kriterium für die Herausbildung einer plausiblen Hörwahrnehmung und somit für ein Binauralsynthesesystem mit hoher Passung/Viabilität in Bezug auf Plausibilität zu sein.



## Kapitel 5

### Methoden einer adaptiven Binauralsynthese

Was macht ein Steuermann, der sein Schiff sicher in den Hafen hineinmanövrieren möchte? Er absolviert kein ein für allemal festgelegtes Programm, sondern er variiert dies permanent. Wenn das Boot vom Kurs und seinem Ziel nach links abweicht [...] schätzt er diese Kursabweichung ein, so dass er weiterhin auf den Hafen zufährt. Er versucht, den Fehler zu korrigieren. Und vielleicht steuert er etwas zu stark gegen. Das Ergebnis ist womöglich eine Kursabweichung nach rechts - und die Notwendigkeit, erneut gegenzusteuern. In jedem Moment wird die Abweichung in Relation zu dem ins Auge gefassten Ziel [...] korrigiert.

---

*Heinz von Foerster*  
*Interview [93]*

## 5.1 Akustische Adaptationsverfahren

Die aus den Hörversuchen in Kapitel 4 erlangten Erkenntnisse bestätigen die Annahme, dass die Konstruktion von räumlicher Audioqualität nicht ausschließlich von der technischen Synthese der Ohrsignale, sondern gerade auch von verschiedenen kontextabhängigen Qualitätsparametern abhängt. Eine daraus gezogene Schlussfolgerung ist, dass die akustischen Eigenschaften der in der Binauralsynthese verwendeten Übertragungsfunktionen so verändert werden können, dass die aufgrund von verschiedenen Kontextabhängigkeiten erzeugten Divergenzen gemindert oder gar aufgelöst werden können.

Mithilfe von BRIRs können Schallquellen im Raum abgebildet werden. Die BRIRs können dabei aus raumakustischen Simulationen oder aus Messungen von realen Schallquellen in realen Räumen stammen. Eine umfassende Synthese einer auditorischen Szene mit einer Vielzahl von Schallquellen, Raumakustik und Bewegungen der Quellen und des Empfängers benötigt eine hohe Anzahl an BRIRs. Eine Minimierung der Anzahl bei gleichzeitiger Beibehaltung einer hohen wahrgenommenen Qualität ist hier wünschenswert. Weiterhin ist es eine Herausforderung, die Binauralsynthese an die herrschenden Abhörbedingungen und somit an die kontextabhängigen Qualitätsparameter anzupassen.

Es werden Herangehensweisen vorgestellt, welche einzelne akustische Parameter der (Re-)Synthese an die akustischen Parameter der Abhörsituation anpassen. Es werden Methoden entwickelt, die energiebasierte Parameter, zeitbasierte Parameter oder Kombinationen aus beiden Parameter anpassen. Gemäß der Annahme einer perzeptiven Passung zwischen wahrgenommener resynthetisierte Audioszene und erwarteter internen Repräsentation wird untersucht, welche akustischen Parameter des Binauralsynthesystems geeignet zur Schaffung einer plausiblen auditorischen Illusion sind.

Die Zielstellung ist die Entwicklung von Methoden zur

**Anpassung an den Abhörraum** Die in einem Raum gemessenen BRIRs sollen in ihren raumakustischen Eigenschaften an den Abhörraum angepasst werden. Die Anpassung erfolgt durch Veränderung einzelner oder mehrerer raumakustischer Parameter der aufgenommenen BRIRs bis diese dem Abhörraum entsprechen.

**Synthese von neuen BRIRs** Es sollen neue BRIRs an verschiedenen Positionen im Raum aus gemessenen BRIRs von anderen Positionen synthetisiert werden. Die Methoden haben zum Ziel, eine Vielzahl an BRIRs aus räumlich dünn-besetzten Messungen und somit wenigen BRIRs zu erzeugen.

Die für die Synthese und Anpassung der BRIRs untersuchten raumakustischen Parameter sind das Direktschall- zu Nachhallenergieverhältnis (DRR) und die Initialzeitlücke (ITDG). Die Parameter decken zum einen intensitätsbasierte Akustikparameter und zum anderen die zeitliche Struktur einer Raumimpulsantwort ab. Beide Gruppen sind gemäß den Untersuchungen von Begault et al. [57] zum Einfluss des Nachhalls und nach Wallach et al.[222] und



Clifton [76] zum Präzedenzeffekt als relevant einzuschätzen. Weiterhin sind diese Parameter direkt für Anwendungen einer adaptiven Binauralsynthese geeignet [4].

## 5.2 Anpassung von BRIRs

Die Zielstellung der nachfolgend beschriebenen Methode ist die Anpassung der in der Binauralsynthese verwendeten BRIRs an die akustischen Eigenschaften des Abhörtraumes. Die anzupassenden BRIRs können beispielsweise aus einem zum Abhörtraum akustisch verschiedenen Raum kommen. Der in Kapitel 4.2 dargelegte Raumdivergenzeffekt soll durch diese Anpassung minimiert werden.

Es werden zwei Experimente vorgestellt, in denen es Hörern ermöglicht wird, das Direktschall- zu Nachhallenergieverhältnis (DRR) der Synthese an das des Abhörtraumes anzupassen. Das vorgestellte Verfahren verändert somit energiebasierte Eigenschaften der BRIRs. Die zeitliche Struktur und somit auch die zeitlichen Reflexionsmuster des zu synthetisierenden Raumes und Szene werden nicht verändert. Der Höreindruck der abzubildende Szene und des Raumes soll dadurch erhalten bleiben.

### 5.2.1 Exp. VI: Perzeptive DRR-Anpassung

Es soll untersucht werden, inwiefern Testpersonen in der Lage sind, das DRR der Synthese an das DRR des Abhörtraumes dynamisch anzupassen. Weiterführende Informationen sind ebenfalls den Veröffentlichungen zu entnehmen [25, 32].

#### 5.2.1.1 Untersuchungsgegenstand

In diesem Experiment wird der Frage nachgegangen, wie zuverlässig die Testpersonen das DRR der Synthese bei Raumdivergenz und Raumkonvergenz anpassen können. Neben den eingestellten Medianen des DRR werden vor allem auch die ermittelten Interquartilabstände und somit die Reliabilität der Einstellung diskutiert.

Die an dem Experiment teilnehmenden Testpersonen haben die Aufgabe, das DRR des zu synthetisierenden Raumes so einzustellen, bis perzeptive Übereinstimmung zwischen der Simulation und der individuellen Erwartung erreicht wird. Die Referenzbildung der Einstellung erfolgt anhand eines vorher stattgefundenen Trainings mit realen Lautsprechern im Abhörtraum. Die Testpersonen werden dadurch auf den Abhörtraum trainiert.

Nachstehenden Fragen sollen beantwortet werden: Auf welchen mittleren DRR-Wert stellen Testpersonen das DRR der Synthese bei Raumdivergenz und Raumkonvergenz ein? Welche Varianz ergibt sich für die Einstellungen?

#### 5.2.1.2 Versuchsaufbau

**Binauralsynthesesystem** - Zur Generierung der Testsignale werden künstliche BRIRs unter Verwendung eines KEMAR Kopf-und-Torso-Simulators für die ausgewählten Räume und

Positionen gemessen. Die Entzerrung der Kopfhörer erfolgt über eine am Kunstkopf gemessenen und invertierten Kopfhörer-Übertragungsfunktionen. Für die Wiedergabe wird ein Stax Lambda Pro Kopfhörer verwendet. Die Testpersonen können mithilfe eines Drehknopfes die aktuell verwendeten BRIRs auswählen. Die hierfür benötigte Faltung und Filterauswahl wird durch das Programm pyBinSim [157] realisiert. Die Bereitstellung der Filter erfolgt durch eine DRR-Skalierung der aufgenommenen BRIRs.

**Objektpositionen** - Es werden die BRIRs für die horizontalen Richtungen  $330^\circ$  und  $90^\circ$  bei einer Entfernung von 2 m zum Aufnahmeort gemessen. Die verwendeten Richtungen beinhalten binaurale Merkmale und die Ergebnisse können mit denen aus anderen Untersuchungen verglichen werden (bspw. [31]).

**Raumdivergenz** - Für die Aufnahme der BRIRs und als Abhörräume kommen ein Hörlabor (HL, siehe auch Kapitel 2.5) und ein leergeräumter Seminarraum (SR, siehe auch Kapitel 2.5) mit unterschiedlichen raumakustischen Eigenschaften zur Anwendung. Es werden verschiedene Räume mit definierter Raumakustik und ausreichenden Quellen-Empfänger-Abständen gewählt, um den Nachhall in diesen Räumen berücksichtigen zu können.

**DRR Einstellung** - Das DRR wird auf einer logarithmischen Skala durch Verstärkung oder Dämpfung des Nachhallanteils relativ zum Direktschallanteil von gemessenen BRIRs eingestellt. Die Änderung in der BRIR setzt 5 ms nach dem Direktschall ein. Eine Beeinflussung des richtungsabhängigen kopfbezogenen Anteils (HRTF) wird dadurch minimiert. Ein ansteigender bzw. abfallender Kurvenverlauf eines Hann-Fensters wird im Übergangsbereich zur Überblendung angewendet. Die unterschiedlichen DRRs sind in 33 Schritte eingeteilt. Es ergeben sich dadurch 25 Dämpfungsschritte bei einer normalisierten Amplitude von Null bis Eins und sieben Verstärkungsstufen bei einer normalisierten Amplitude von 1,0 bis 1,3. Weiterhin liegt die unveränderte gemessene BRIR vor. Abbildung 5.1 verdeutlicht die im Experiment möglichen einstellbaren Stufen mit den dazugehörigen DRR-Werten für die beiden untersuchten Räume und beiden Richtungen. Die Testpersonen können das DRR des zu synthetisierenden Raumes mithilfe eines Drehknopfes gemäß den in Abbildung 5.1 angegebenen Grenzen frei anpassen. Das DRR soll dabei soweit angepasst werden, bis perzeptive Übereinstimmung zwischen der Simulation und der individuellen Erwartung erreicht wird.

**Audiosignale** - Die verwendeten Audiosignale sind ein Saxofonstück und ein nachhallfreies Sprachsignal eines männlichen Sprechers. Die Verwendung der Sprache wird in diesem Experiment als unkritisch im Sinne der Distanzwahrnehmung betrachtet, da lediglich eine fixe Distanz im Test abgefragt wird.

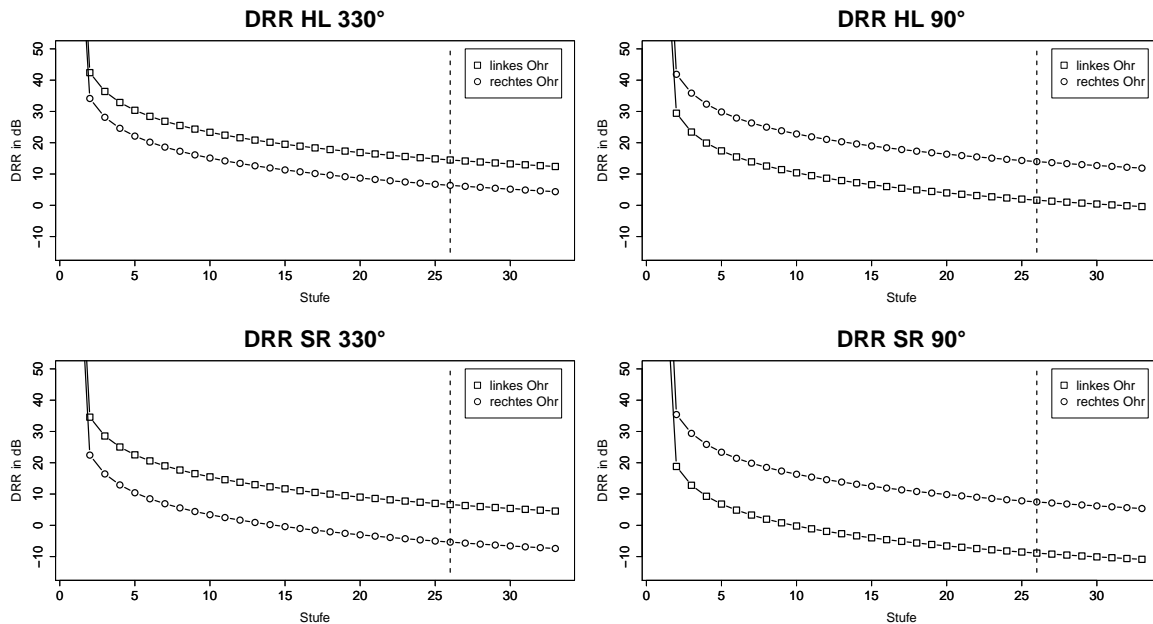


Abbildung 5.1: Im Experiment mögliche DRR-Stufen am ipsilateralen Ohr; Stufe 1 entspricht einer nachhallfreien Situation; gestrichelte Linie entspricht dem DRR der gemessenen BRIR bei 2 m Entfernung.

### 5.2.1.3 Versuchsdurchführung

Die Richtung der DRR Einstellung am Drehknopf wird nach jedem Stimulus zufällig geändert, um Adaptionseffekte aufgrund gleicher Drehrichtung am Knopf zu minimieren. Die Testpersonen müssen ihre Bewertung durch drücken eines Knopfes bestätigen. Der aktuelle Stimulus wird in einer Schleife wiederholt abgespielt, um genügend Zeit zur Einstellung zur Verfügung zu stellen. Der nachfolgende Stimulus wird nach einer Pause von drei Sekunden präsentiert. Ein wiederholtes Abspielen des vorangegangenen Stimulus ist nicht möglich.

Zur Herausbildung der Referenz zum Vergleich des eingestellten DRRs der Synthese werden den Testpersonen zu Beginn des Testes zweimal die Testsignale über reale Lautsprecher im Abhörraum aus den Richtungen  $330^\circ$  und  $90^\circ$  vorgespielt. Die Testpersonen sollen einen Eindruck der dargebotenen Richtung, Distanz und der Raumakustik des Abhörraumes für dem im Test geforderten Vergleich bekommen. Alle Tests werden in allen vier möglichen Kombinationen von zu synthetisierenden Raum und Abhörraum durchgeführt.

### 5.2.1.4 Ergebnisse

Die Ergebnisse beinhalten die Bewertungen von zwei weiblichen und 12 männlichen Testpersonen. Die Testpersonen sind erfahren in der Teilnahme an Wahrnehmungsexperimenten. Abbildung 5.2 zeigt die Bewertungen als Mediane und 25/75 Quartile für die verschiedenen Richtungen und Raumkombinationen. Eine Auswertung der Bewertungen über die Audiosigna-

le ergibt keine signifikanten Unterschiede zwischen den Audiosignalen. Zur weiteren Analyse werden die Bewertungen zusammengefasst.

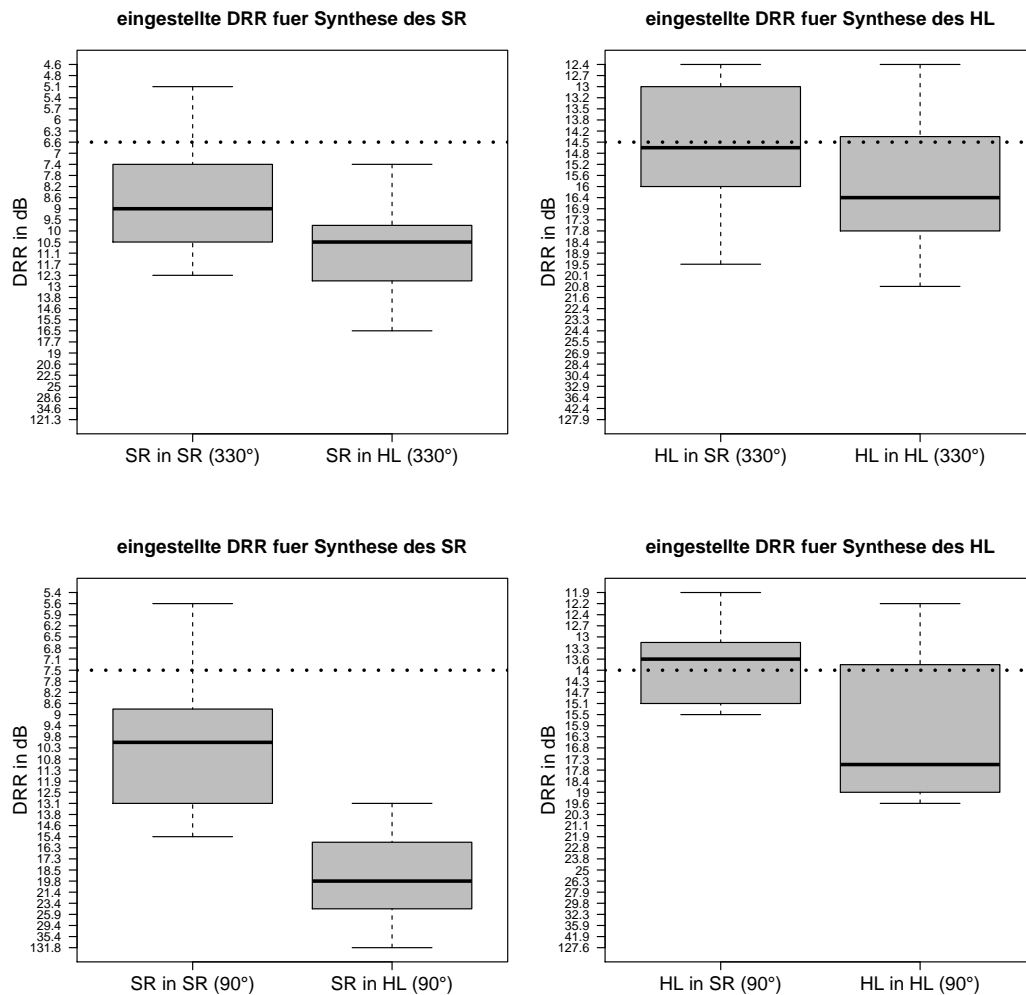


Abbildung 5.2: Im Hörtest von den Testpersonen eingestellte DRR für die untersuchten Testkonditionen als Mediane und Quantile; gepunktete Linie bezeichnet das gemessene (unveränderte) DRR des synthetisierten Raumes; SR=Seminarraum, HL=Hörlabor.

Die eingestellten DRRs sollen erwartungsgemäß vergleichbar zu der gemessenen und unveränderten DRR des synthetisierten Raumes bei Übereinstimmung zwischen synthetisiertem Raum und Abhörraum sein. Der Median des eingestellten DRR für die Synthese des Seminarraumes im Seminarraum („SR in SR“) für die 330° Richtung liegt bei ca. 9 dB und für die 90° Richtung bei ca. 10 dB. Dies entspricht einer Differenz von ca. 2,5 dB gegenüber der Messung im SR für beide Richtungen. Das eingestellte DRR für die Synthese des Hörlabors im Hörlabor („HL in HL“) liegt bei ca. 16,4 dB für die 330° Richtung und bei ca. 17,5 dB für die 90° Richtung. Die DRR Differenz zur Messung im HL beträgt ca. 2 dB für 330° und 3,5 dB für die 90° Richtung.

Die Testpersonen wählen für beide Raumkombinationen ein niedrigeres DRR bei perzeptiver Übereinstimmung von synthetisiertem Raum und Abhörraum.

Demgegenüber steht ein signifikant ( $p < 0,05$ ,  $\alpha = 5\%$ ) höheres DRR von ca. 10,5 dB für die 330° Richtung und 19,8 dB für die 90° Richtung bei Wiedergabe des Seminarraums im Hörlabor („SR in HL“). Die gemessenen DRRs werden für die Wiedergabe des Hörlabors im Seminarraum („HL in SR“) für beide Richtungen ausgewählt. Die Abweichung beträgt maximal 0,5 dB und liegt damit deutlich unter der JND (2,5 dB bis 3,8 dB [133]). Es wird erwartet, dass die Testpersonen für eine perzeptive Übereinstimmung ein geringeres DRR auswählen, da das Hörlabor deutlich weniger Nachhall als der Seminarraum besitzt. Das Fehlen von ausreichender Nachhallenergie in den BRIRs des Hörlabors machte es allerdings unmöglich, das gemessene DRR des Seminarraumes zu erreichen (siehe Abbildung 5.1).

Folgende Interquartilabstände (IQAs) der Bewertungen ergeben sich für die untersuchten Raumkombinationen für die Richtungen 330° und 90°: „HL in SR“ 3 dB und 2 dB, „SR in SR“ 3 dB und 4,3 dB, „HL in HL“ 3,5 dB und 5,2 dB, „SR in HL“ 2,5 dB und 8,2 dB. Die in diesem Test ermittelten IQAs liegen in einer vergleichbaren Größenordnung mit den JNDs für einen Bereich der DRR von 0 dB bis 20 dB bei Lautsprecherwiedergabe (siehe [176], [236]) und JNDs von 2,4 dB bis 8,7 dB bei Verwendung einer Binauralsynthese (siehe [133]).

Das Experiment zeigt, dass die Hörer das DRR der Synthese innerhalb des gerade noch wahrnehmbaren Unterschiedes (JND) an den Abhörraum anpassen können. Diese Anpassung erfolgt im Vergleich zu einer internen Referenz der Hörer gemäß erwarteter Raumübereinstimmung. Die Ergebnisse zeigen weiterhin einen Versatz des eingestellten DRR von ca. 2,5 dB für alle untersuchten Konditionen.

### 5.2.2 Exp. VII: DRR-Einfluss auf den Raumdivergenzeffekt

Es soll untersucht werden, ob eine DRR Anpassung zu einer Minimierung des Raumdivergenzeffektes führt. Weiterführende Informationen sind der dazugehörigen Veröffentlichung zu entnehmen [24]. Ein dazugehöriger Datensatz an gemessenen und mithilfe der Änderung des DRR angepassten BRIRs ist der Veröffentlichung [15] zu entnehmen.

#### 5.2.2.1 Untersuchungsgegenstand

Dieses Experiment setzt auf den Ergebnissen des Experiments VI auf. Der Schwerpunkt der Untersuchungen liegt hier neben der Einstellung des DRR auf der Evaluierung der wahrgenommenen Externalität der eingestellten Szenen bei Konvergenz bzw. Divergenz zwischen synthetisiertem Raum und Abhörraum. Die Qualitätsmerkmale Lokalisation, Externalität und wahrgenommene Raumübereinstimmung zwischen synthetisiertem Raum und Abhörraum werden in einem Hörversuch untersucht.

Nachstehenden Fragen sollen beantwortet werden: Auf welchen mittleren DRR-Wert stellen Testpersonen das DRR der Synthese bei Raumdivergenz und Raumkonvergenz ein? Welche

Varianz ergibt sich für die Einstellungen? In welchem Umfang führt eine DRR-Einstellung zu einer Änderung in der Externalisierung der Hörereignisse bei Raumdivergenz und Raumkonvergenz?

### 5.2.2.2 Versuchsaufbau

**Binauralsynthesystem** - Das System verwendet Messungen von individuellen und Kunstkopf-BRIRs für die ausgewählten Räume, Schallquellen und Positionen. Ein nicht-dynamisches System ohne Head-Tracking wird verwendet, um zu vermeiden, dass dynamische Cues wahrnehmungsbedingte Mehrdeutigkeiten, wie Quadrantenfehler und Im-Kopf-Lokalisierung auflösen und somit die Effekte der DRR Einstellung ggf. überdecken. Ein personalisierbares binaurales System wird verwendet, um die Wiedergabetreue der Simulation im Vergleich zu echten Lautsprechern zu erhöhen. Verschiedene Räume werden ausgewählt, um unterschiedliche raumakustische Eigenschaften, wie Nachhallzeit und Quellen-Empfänger-Abstände zu berücksichtigen.

Bei Verwendung individueller BRIRs werden die Kopfhörer über individuelle HPTFs entzerrt. HPTFs aus dem Head-and-Torso-Simulator werden verwendet, wenn künstliche BRIRs verwendet werden. In-Ohr-Mikrofone werden zur Messung einzelner BRIRs und HPTFs am Eingang des blockierten Gehörgangs jeder Testperson verwendet. Die Mikrofone werden zwischen den BRIR- und HPTF-Messungen nicht entfernt. Die Invertierung einer HPTF wird nach der Methode der kleinsten Quadrate berechnet. Für die Wiedergabe wird ein extra-auraler Kopfhörer BK211 verwendet, der die Anforderungen an einen offenen Kopfhörer erfüllt. Dies ermöglicht ein Testdesign mit dem Hören des realen Raumes und der Resynthese dieser Lautsprecher über Kopfhörer.

**Raumdivergenz** - Es werden zwei Räume zur Synthese und als Abhörraum verwendet. Die Räume unterscheiden sich hinsichtlich ihrer raumakustischen Eigenschaften teils erheblich. Es kommt ein Hörlabor (HL; siehe dazu Kapitel 2.5) sowie ein leergeräumter Seminarraum (SR; siehe dazu Kapitel 2.5) zur Verwendung.

Die verwendeten Kombinationen von Abhörraum und synthetisierendem Raum ergeben sich zu: Synthese des Seminarraums im Seminarraum, Synthese des Seminarraums im Hörlabor, Synthese des Hörlabors im Hörlabor und Synthese des Hörlabors im Seminarraum. Die durch die Testpersonen vorgenommene Einstellung des DRR der Synthese führt zu weiteren Kombinationen von Abhörraum und synthetisiertem Raum. Folgende Nomenklatur wird als Bezeichnung verwendet: „synthetisierter Raum (DRR-Einstellung auf Raum)\_Abhörraum“. Die einzelnen Bezeichnungen der Konditionen sind:

**HL(SR)\_SR** Synthese des Hörlabors im Seminarraum mit DRR Einstellung der Synthese auf den Seminarraum

- SR(HL)\_HL** Synthese des Seminarraums im Hörlabor mit DRR Einstellung der Synthese auf das Hörlabor
- HL(HL)\_HL** Synthese des Hörlabors im Hörlabor mit DRR Einstellung der Synthese auf das Hörlabor (keine Veränderung des DRR)
- SR(SR)\_SR** Synthese des Seminarraums im Seminarraum mit DRR Einstellung der Synthese auf den Seminarraum (keine Veränderung des DRR)
- HL(HL)\_SR** Synthese des Hörlabors im Seminarraum mit DRR Einstellung der Synthese auf das Hörlabor (keine Veränderung des DRR)
- SR(SR)\_HL** Synthese des Seminarraums im Hörlabor mit DRR Einstellung der Synthese auf den Seminarraum (keine Veränderung des DRR)
- HL(SR)\_HL** Synthese des Hörlabors im Hörlabor mit DRR Einstellung der Synthese auf den Seminarraum
- SR(HL)\_SR** Synthese des Seminarraums im Seminarraum mit DRR Einstellung der Synthese auf das Hörlabor

Abbildung 5.3 verdeutlicht schematisch die im Test verwendeten Raumkombinationen mit Kongruenz des DRR zwischen dem eingestellten synthetisierten Raum und dem Abhörraum.

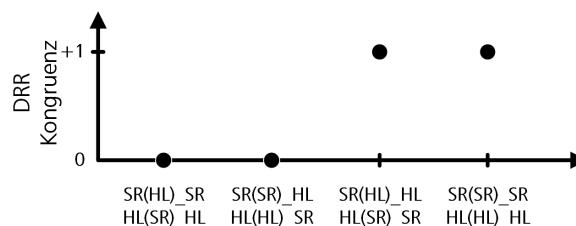


Abbildung 5.3: Kongruenz des Direktschall- zu Nachhallenergieverhältnisses (DRR) zwischen dem DRR des eingestellten, synthetisierten Raumes und dem Abhörraum für verschiedenen Raumkombinationen; Raumname in Klammern bezeichnet die DRR-Einstellung auf diesen Raum; SR = Seminarraum, HL = Hörlabor.

**Objektpositionen** - Die synthetisierten Schallquellen befinden sich an sechs Positionen um den Hörer herum. Zur Erstellung der Binauralsynthese werden die BRIRs aus den Richtungen  $0^\circ$ ,  $30^\circ$ ,  $60^\circ$ ,  $180^\circ$ ,  $240^\circ$  und  $300^\circ$  im Hörlabor und im Seminarraum gemessen. Der Abstand der gemessenen Lautsprecher zum Aufnahmeort beträgt 2,2 m. Als Lautsprecher wird ein Genelec 1029A verwendet, welcher auf den Aufnahmeort ausgerichtet ist und sich in der Medianebene der Testperson befindet. Abbildung 5.4 verdeutlicht schematisch den Messaufbau.

### 5.2.2.3 Versuchsdurchführung

Die Testpersonen können das DRR der Synthese einstellen bis perzeptive Übereinstimmung zwischen der Synthese und dem Abhörraum als interne Referenz wahrgenommen wird. Die Einstellung des DRR soll anhand des Qualitätsmerkmals wahrgenommene Raumkonvergenz

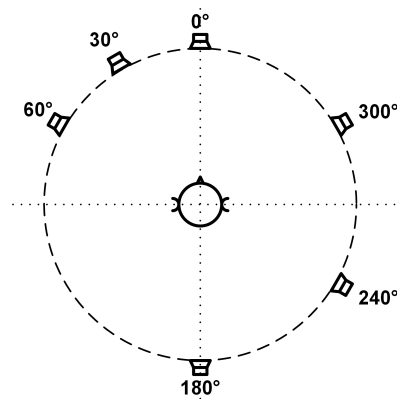


Abbildung 5.4: Schematische Darstellung des Messaufbaus mit Anordnung der Lautsprecher zur Messung der BRIRs; Radius = 2,2 m.

(engl. room convergence) erfolgen. Raumkonvergenz beschreibt die Wahrnehmung von auditiver Übereinstimmung zwischen synthetisierter Szene mit dem Abhörraum. Die Erwartungen basierend auf der Kenntnis des Abhörraumes und wird als interne Referenz für den Vergleich angenommen. Raumkonvergenz lässt sich mit den Attributen „Passgenauigkeit“, „Natürlichkeit“, „Wohlklang“ oder „angenehm“ beschreiben [45].

Die verschiedenen DRR-skalierten BRIRs werden durch eine Verstärkung bzw. Dämpfung des Nachhallanteils relativ zum Direktschallanteil der individuell gemessenen BRIRs verändert. Der Nachhallanteil wird in diesem Fall ab 3 ms nach dem Direktschall definiert. Die BRIRs mit angepasstem DRR werden für 70 Stufen mit 46 Dämpfungsstufen bei einer normierter Amplitude von Null bis Eins, 23 Verstärkungsstufen bei einer normierten Amplitude von Eins bis 1,3 und der gemessenen BRIR berechnet. Abbildung 5.5 verdeutlicht die in dem Hörversuch für die Synthese des Seminarraums und Hörlabors auswählbaren DRR-Stufen mit ihren dazugehörigen Pegelangaben. Die DRRs sind als Mediane und 25/27 Quartile für die Messungen aller Testpersonen für das linke und rechte Ohr dargestellt. DRR-Stufe 1 entspricht einer weitestgehend nachhallfreien Synthese während Stufe 70 der halligsten Kondition entspricht. Die Stufe 47 entspricht dem gemessenen DRR der BRIR der originalen Aufnahme.

#### 5.2.2.4 Ergebnisse

**DRR Einstellung** - Die Abbildungen 5.6 und 5.7 verdeutlichen die von den Testpersonen eingestellten DRR Pegel. Abbildung 5.6 zeigt die Einstellung bei der Synthese des Seminarraums beziehungsweise des Hörlabors bei Wiedergabe über Kopfhörer im Seminarraum als Abhörraum. Der DRR Pegel des Abhörraumes ist in der Abbildung 5.6 als gestrichelte Linie eingezeichnet. Die Testpersonen stellen bei der Resynthese des Seminarraums den entsprechenden DRR Pegel des Seminarraums ein (Abbildung 5.6 links). Der Median der eingestellten Pegel liegt für alle Richtungen im Mittel 1,3 dB über den Pegeln der Messungen in diesem



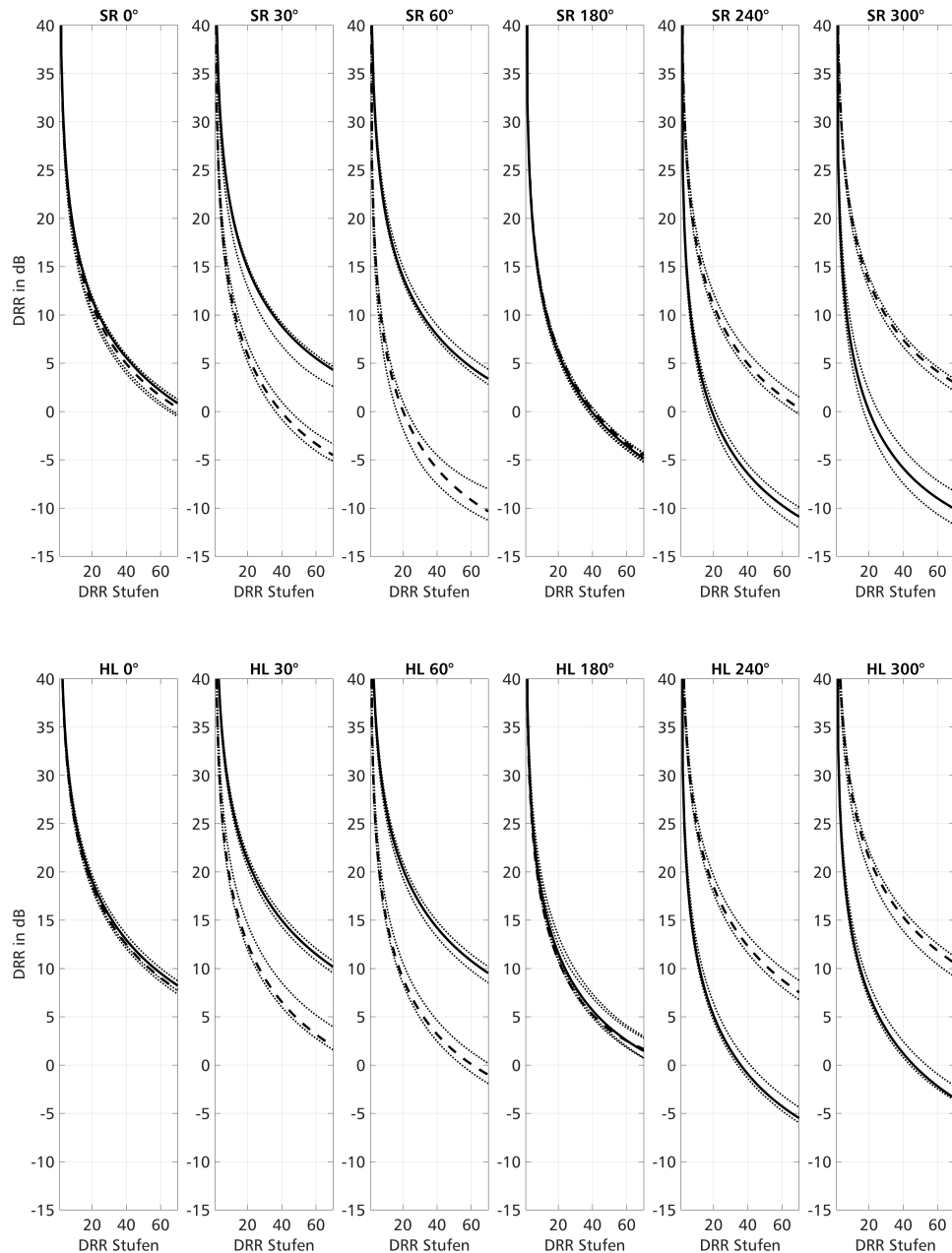


Abbildung 5.5: Im Hörversuch mögliche DRR-Stufen für die Synthese des Seminarraums (SR, oben) und des Hörlabors (HL, unten) zur Einstellung durch die Testpersonen; durchgezogene Linie=linkes Ohr, gestrichelte Linie=rechtes Ohr, gepunktete Linie=25/75 Quartile.

Raum. Die Testpersonen wählen eine Synthese mit geringfügig weniger Nachhall aus. Dies entspricht in seiner Tendenz den Ergebnissen aus dem Experiment I. Allerdings liegen die eingestellten höheren DRR Pegel innerhalb der gerade wahrnehmbaren Unterschiede (JNDs) in der DRR Wahrnehmung [133, 176, 236]. Die Interquartilabstände (IQAs) der Einstellungen

liegen im Mittel über alle Richtungen bei 4,7 dB. Die in diesem Experiment ermittelten IQAs liegen ca. 1 dB über den in Experiment I ermittelten IQAs.

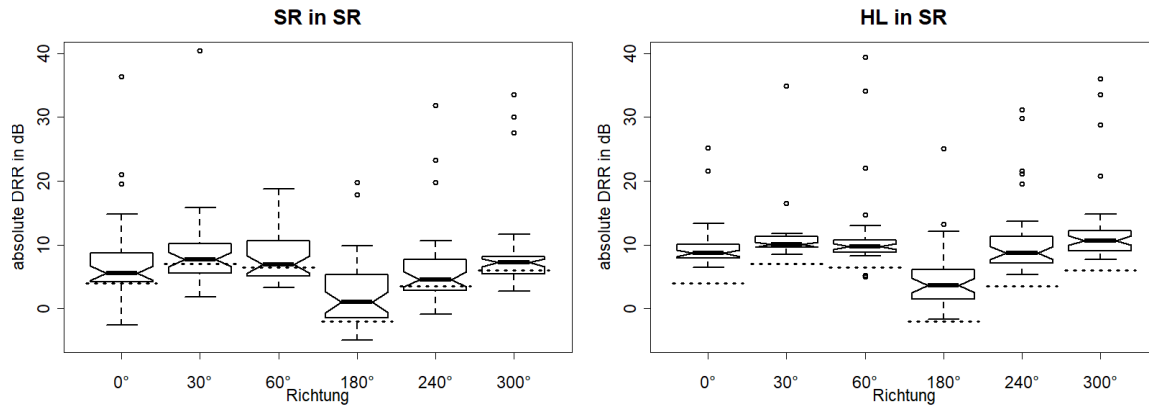


Abbildung 5.6: Eingestellte DRRs für verschiedene Richtungen und Seminarraum als Abhör- raum.

Bei der Resynthese des Hörlabors im Seminarraum (Abbildung 5.6 rechts; „HL in SR“) liegt der Median der eingestellten DRR Pegel im Mittel über alle Richtungen 4,4 dB über den Pegeln der Messungen. Die Testpersonen weichen gegenüber der Resynthese des Seminarraums deutlich in ihrer DRR Einstellung ab. Der Grund hierfür liegt in der begrenzten Auswahlmöglichkeit der DRR Pegel bei der Synthese des Hörlabors. Die im Hörversuch auswählbaren DRR Stufen erreichen nicht die geringen DRR Pegel der Messungen im Seminarraum (siehe hierzu Abbildungen 5.5). Auf eine weitere Verstärkung der Hallanteile der BRIRs des Hörlabors wurde aufgrund der dadurch störenden Anhebung des Rauschens verzichtet.

Die ermittelten IQAs der Einstellungen liegen im Mittel über alle Richtungen bei ca. 2,8 dB und fallen somit geringer als bei der Wiedergabe des Seminarraumes aus. Ein Grund hierfür ist ebenfalls in der begrenzten Auswahlmöglichkeit der DRR Pegel zu sehen. Dennoch sind die in Experiment VII gefundenen IQAs von 2,8 dB nahezu identisch mit den IQDs aus Experiment VI mit 2,5 dB.

Abbildung 5.7 zeigt die Einstellung bei der Synthese des Seminarraums beziehungsweise des Hörlabors bei Wiedergabe über Kopfhörer im Hörlabor als Abhör- raum. Der DRR Pegel des Abhör- raumes ist in der Abbildung 5.7 als gestrichelte Linie eingezeichnet. Im rechten Teil der Abbildung sind die eingestellten DRR Pegel bei Konvergenz zwischen Abhör- raum und synthetisierten Raum dargestellt. Den Testpersonen gelingt es sehr gut, den DRR Pegel der Synthese an den gemessenen Pegel des Abhör- raumes einzustellen. Die gemittelte Abweichung der Mediane über alle Richtungen ergibt eine Abweichung von 0 dB gegenüber dem gemessenen Pegel. Die über alle Richtungen gemittelten IQAs liegen bei 3,8 dB und sind damit nahezu identisch mit den aus Experiment VI ermittelten und über die beiden Richtungen gemittelte IQA für die Raumkondition „HL in HL“ von 4,3 dB.

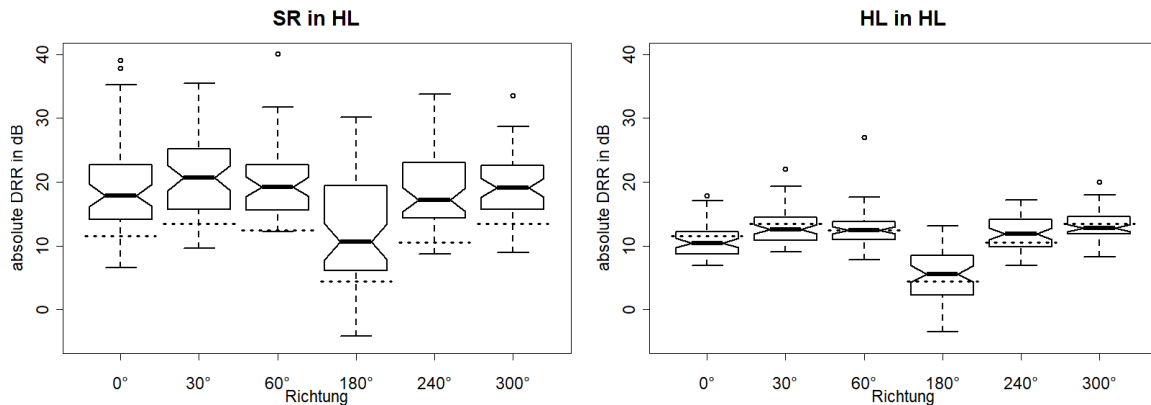


Abbildung 5.7: Eingestellte DRRs für verschiedene Richtungen und Hörlabor als Abhörraum.

Der linke Teil der Abbildung 5.7 zeigt die eingestellten DRR Pegel für die Resynthese des Seminarraums im Hörlabor als Abhörraum („SR in HL“). Die über alle Richtungen gemittelte Abweichung des eingestellten Medians gegenüber der gemessenen DRR Pegel des Hörlabors liegt bei 9 dB. Die Testpersonen wählen eine weniger hallige Resynthese des Seminarraums als es dem Hörlabor als Abhörraum entsprechen würde. Die über alle Richtungen gemittelten IQAs liegen bei 17,5 dB. Der IQA liegt damit deutlich höher als die in der Literatur angegebenen JNDs [133, 176, 236] und dem in Experiment VI ermittelten IQA von ca. 5,4 dB.

**Vortest** - Der Vortest findet nach den Messungen der individuellen BRIRs und HPIRs im Seminarraum statt. Er dient dem Kennenlernen der Test-GUI durch die Testpersonen. Weiterhin soll überprüft werden, ob die Ankersignale, Wiedergabe über die realen Lautsprecher im Raum und die Binauralsynthese bei Verwendung von Kunstkopf-BRIRs unterschieden werden können.

Abbildung 5.8 zeigt die Bewertungen der Testpersonen aus dem Vortest. Die Wiedergabe über die realen Lautsprecher im Raum wird mit den höchsten Externalitätsindizes bewertet. Sie liegen im Mittel über alle Richtungen bei 0,9. Die realen Lautsprecher werden als weitestgehend extern bewertet. Lediglich die Richtung 180° wird mit einem Index von 0,7 und somit als weniger extern bewertet. Die Audiostimuli bei Wiedergabe mittels Binauralsynthese und bei Verwendung von Kunstkopf-BRIRs werden im Mittel über alle Richtungen mit einem Index von 0,7 bewertet. Die Richtungen 0° und 180° werden mit einem geringeren Index von 0,5 und 0,6 bewertet. Die Resynthese der Lautsprecher wird im Vergleich mit den realen Lautsprechern als weniger extern bewertet. Dies liegt in der Verwendung von Kunstkopf-BRIRs begründet. Weiterhin werden im Test auch die realen Lautsprecher dargeboten. Es ist nicht auszuschließen, dass diese Signale als interne Referenz für externe Signale dienen.

Die in dem Test verwendeten Ankersignale mit geringer räumlicher Qualität werden erwartungsgemäß mit geringer Externalität bewertet. Der über alle Richtungen gemittelte Index liegt bei 0,25. Im Vergleich zu anderen Untersuchungen zeigt die Auswertung höhere Externalitätsin-

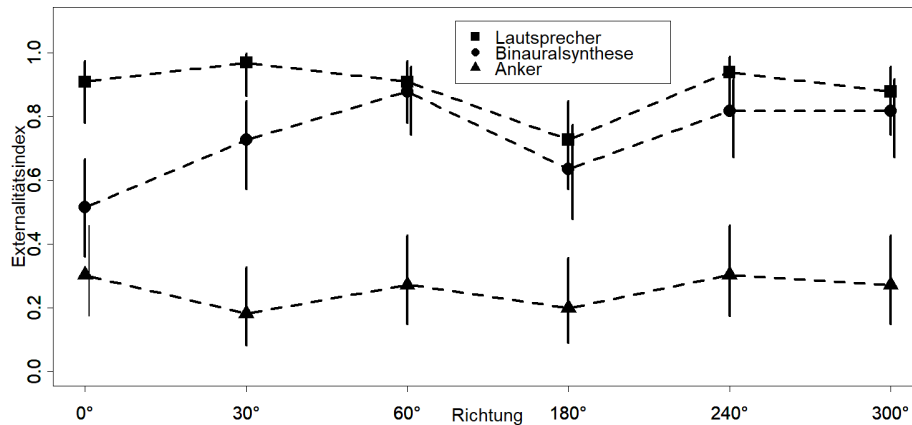


Abbildung 5.8: Externalitätsindizes des Vortests bei Wiedergabe der resynthetisierten Lautsprecher mittels Kunstkopf-BRIRs, der realen Lautsprecher im Raum sowie der Ankersignale mit 95% Binominalkonfidenzintervall.

dizes (siehe hierzu Experiment III, Kapitel 4.2). Die in diesem Test verwendeten Ankersignale sind mittels Intensitätsstereophonie in ihrer Richtungswahrnehmung eingestellte und mit monauralen Raumimpulsantworten des jeweiligen Lautsprechers gefaltete Audiosignale. Der Anker enthält somit Merkmale der Halligkeit, aber schwache binaurale und monaurale Merkmale für Direktschall und Reflexionen. Ihre Externalisierung fällt dennoch geringer gegenüber der Bewertung der realen Lautsprecher und der mittels Kunstkopf-BRIRs resynthetisierten Lautsprecher ab. Die Signale sind somit als Ankersignale für die Untersuchung geeignet.

**Hörversuch Externalität** - Abbildungen 5.9 und 5.10 verdeutlichen die Bewertungen der Testpersonen aus dem Qualitätstests bei Bewertungen des Qualitätsmerkmals Externalität. Die Ergebnisse sind als Externalitätindizes mit dazugehörigen 95% Konfidenzintervall für die verschiedenen Richtungen und Raumkombinationen dargestellt.

Abbildung 5.9 zeigt die Indizes bei Wiedergabe im Seminarraum. Die höchsten Indizes werden bei Resynthese des Seminarraumes im Seminarraum und bei DRR Kongruenz („SR(SR)\_SR“) erreicht. Im Mittel über alle Richtungen liegt der Index bei ca. 0,9. Geringere Indizes von ca. 0,7 sind bei den Richtungen mit erwartbaren Lokalisationsfehlern bei 0° und 180° festzustellen. Ein vergleichbares Verhalten des Abfalls der Indizes für frontale und rückseitige Richtungen zeigt sich für die Resynthese des Seminarraumes im Seminarraum aber bei DRR Divergenz („SR(HL)\_SR“). Für diese Raumkombination liegen die Indizes für die 0° und 180° Richtung bei ca. 0,1. Über alle Richtungen gemittelt ergibt sich ein Index von 0,3. Bei der Resynthese der Lautsprecher aus seitlichen Richtungen ergibt sich ein höherer gemittelter Index von ca. 0,4. Für die Richtungen 30° und 60° liegen die Indizes allerdings im Bereich der Ratewahrscheinlichkeit für das dichotome Merkmal Externalität.

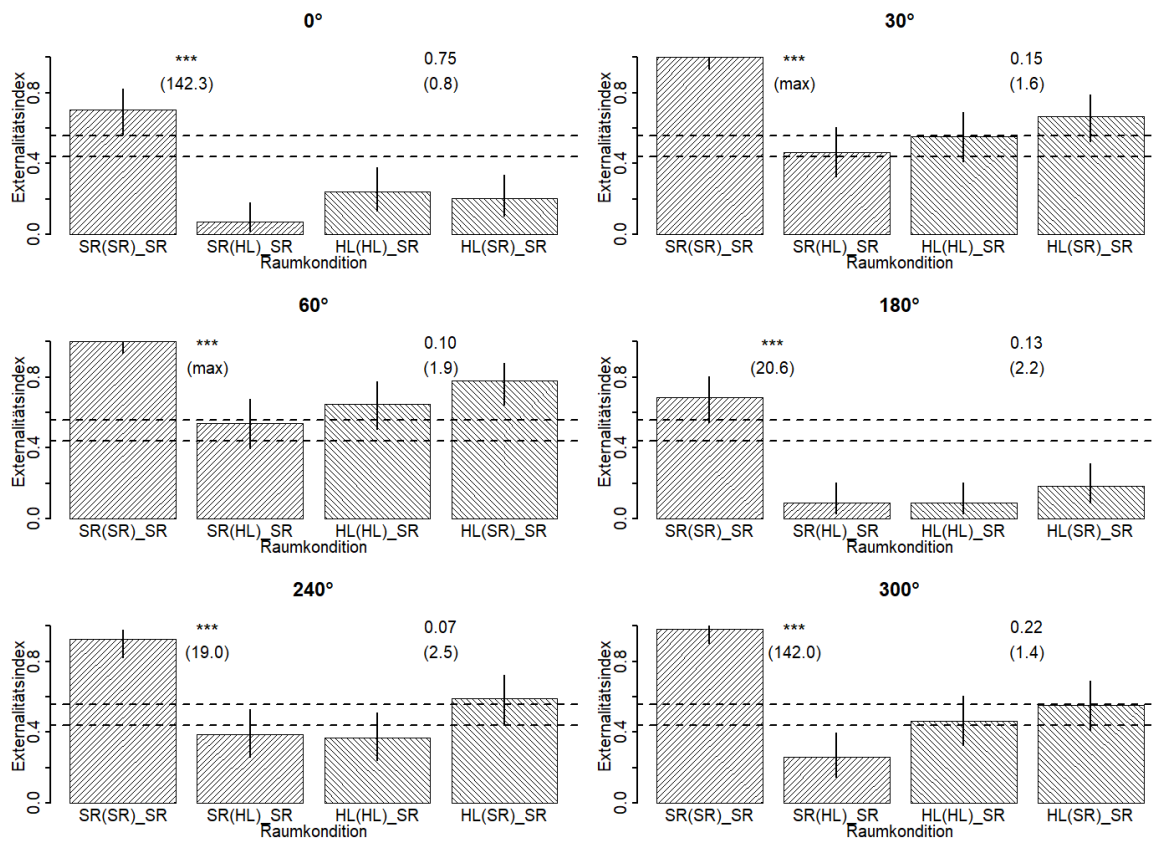


Abbildung 5.9: Externalitätsindizes des Qualitätstestes bei Wiedergabe im Seminarraum mit 95% Konfidenzintervallen; gestrichelte Linie: 95% Konfidenzintervall eines Binomialtests für eine Auswahlwahrscheinlichkeit von 0,5 (Ratewahrscheinlichkeit bei  $N=216$ ); SR=Seminarraum, HL=Hörlabor; \*\*\* signifikanter Unterschied zwischen den Konditionen ( $p < 0,01$ ) bzw. Angabe des Signifikanzniveaus mit Quotenverhältnis.

Die Raumkondition mit Raumdivergenz und DRR Divergenz zwischen resynthetisierten Raum und Abhörraum („HL(HL)\_SR“) erreichen im Mittel über alle Richtungen einen Index von ca. 0,4. Auch hier fallen die Indizes für die  $0^\circ$  und  $180^\circ$  Richtung gegenüber den seitlichen Richtungen stark ab. Die Indizes für die Richtungen  $30^\circ$  und  $300^\circ$  liegen im Bereich der Ratewahrscheinlichkeit für das dichotome Merkmal Externalität. Für die Raumkondition mit Divergenz zwischen resynthetisierten Raum und Abhörraum aber DRR Konvergenz zum Abhörraum („HL(SR)\_SR“) wird im Mittel über alle Richtungen ein Index von ca. 0,5 erreicht. Auch hier fallen die Indizes für die Richtungen  $0^\circ$  und  $180^\circ$  mit ca. 0,2 gegenüber den seitlichen Richtungen geringer aus. Für die Richtungen  $30^\circ$  und  $60^\circ$  ergeben sich die höchsten Indizes von ca. 0,7 und ca. 0,8. Die Indizes für die Raumkondition „HL(SR)\_SR“ liegen somit für alle Richtungen über den Indizes der Raumkondition „HL(HL)\_SR“ bei denen keine DRR Anpassung an den Abhörraum vorgenommen wurde.

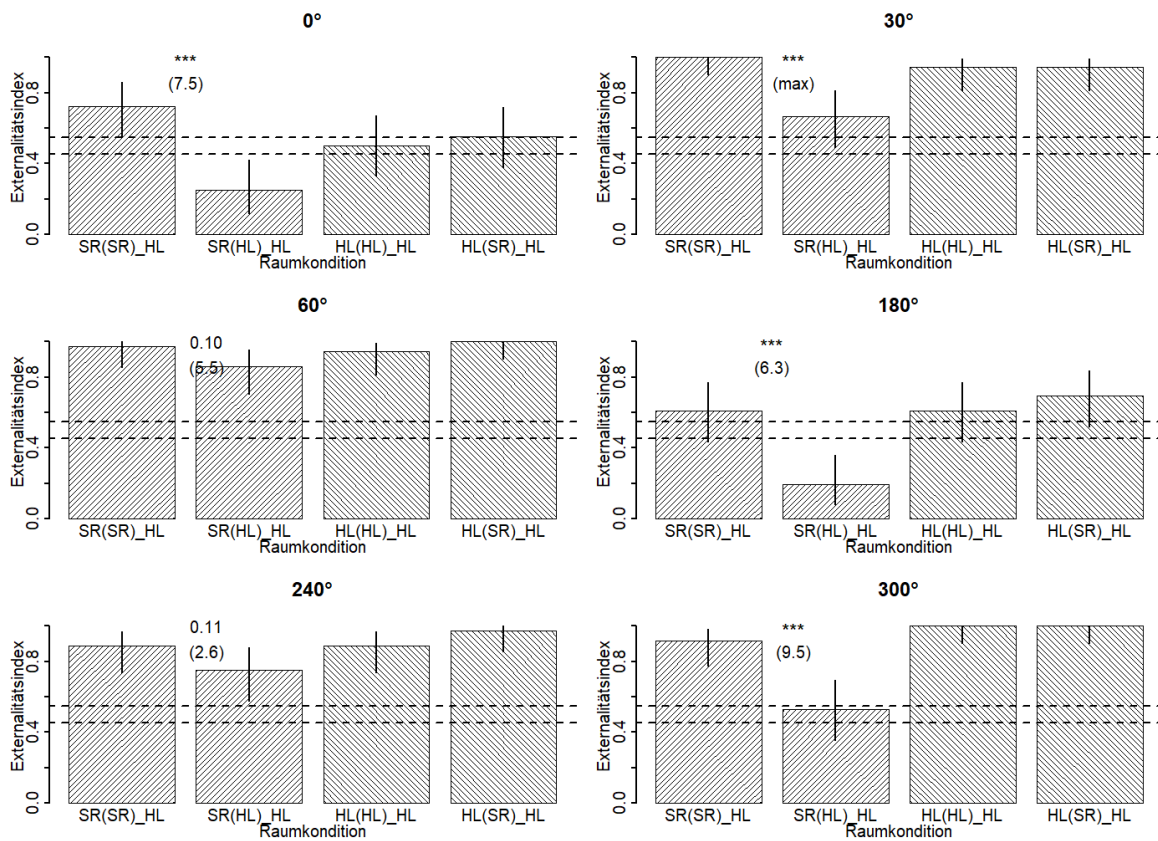


Abbildung 5.10: Externalitätsindizes des Qualitätstestes bei Wiedergabe im Hörlabor mit 95% Konfidenzintervallen; gestrichelte Linie: 95% Konfidenzintervall eines Binominaltests für eine Auswahlwahrscheinlichkeit von 0,5 (Ratewahrscheinlichkeit bei N=324); SR=Seminarraum, HL=Hörlabor; \*\*\* signifikanter Unterschied zwischen den Konditionen ( $p < 0,01$ ) bzw. Angabe des Signifikanzniveaus mit Quotenverhältnis.

Abbildung 5.10 zeigt die Bewertungen der Testpersonen als Externalitätindizes für die Resynthese des Seminarraums bzw. des Hörlabors im Hörlabors bei DRR Konvergenz und DRR Divergenz der Synthese gegenüber dem Abhörraum. Im Vergleich zu den Bewertungen im Seminarraum werden bei allen Raumkonditionen höhere Indizes erreicht. Die geringsten Indizes zeigen sie erwartungsgemäß bei den Richtungen 0° und 180°. Bei der Resynthese des Seminarraums im Hörlabor und bei DRR Divergenz („SR(HL)\_HL“) ergibt sich im Mittel über alle Richtungen ein Index von ca. 0,5. Für die Richtungen 0° und 180° fällt dieser auf 0,25 und 0,2 ab. Für die anderen Raumkonditionen zeigen sich keine signifikanten Unterschiede. Die Resynthese des Hörlabors im Hörlabor bei DRR Konvergenz („HL(HL)\_HL“) erreicht für seitliche Richtungen hohe Indizes von über 0,9.

Für die frontale und rückwärtige Richtung ergeben sich geringere Indizes von 0,5 und ca. 0,6. Ähnliche Werte sind auch für die Resynthese des Hörlabors bei DRR Anpassung an den Seminarraum („HL(SR)\_HL“) und Resynthese des in seinem DRR nicht veränderten

Seminarraums („SR(SR)\_HL“) festzustellen. Im Vergleich zu Abbildung 5.9 zeigt sich, dass die Synthese eines halligen Raumes in einem weniger halligen Raum zu höheren Indizes und somit zu einer erhöhten Externalität führt. Der Raumdivergenzeffekt tritt vor allem bei der Synthese eines weniger halligen Raumes in einem halligeren Abhörraum auf (siehe hierzu auch 4.2).

### 5.2.3 Zusammenfassung

Die Bewertungen zur Einstellung des DRR aus Experiment VI zeigen, dass es den trainierten Testpersonen zuverlässig gelingt, das DRR der Synthese an den Abhörraum anzupassen. Die Anpassung erfolgt anhand der erwarteten Übereinstimmung der Halligkeit. Es liegt kein Vergleich oder relativer Abgleichprozess zwischen Abhörraum und synthetisiertem Raum vor. Die Hörer hören zu Beginn des Tests den realen Raum über Lautsprecher und stellen anschließend sequentiell die Binauralsynthese ein.

Die erreichten IQAs der Bewertungen sind vergleichbar mit den in der Literatur gefundenen JNDs für die DRR Wahrnehmung [133, 176, 236]. Dies deutet auf eine hohe Reliabilität zwischen den Testpersonen und auf ein geeignetes steuerbares akustisches Qualitätselement der Synthese hin. Die Einstellung des DRR scheint weiterhin eine valide Methode zur Anpassung einer Binauralsynthese an den Nachhall des Abhörraumes als kontextabhängiger Qualitätsparameter zu sein. Der IQA bei der Einstellung des DRR kann eine Indikation auf die JND der DRR Wahrnehmung mit der Beziehung  $JND \leq IQA$  sein. Bei Übereinstimmung des synthetisierten Raumes mit dem Abhörraum stellen die Testpersonen für die Synthese ein um 2 dB bis 3,5 dB höheres DRR gegenüber des realen Abhörraumes ein. Die Testpersonen wählen zuverlässig eine weniger hallige Synthese. Dieser Effekt ist als anekdotische Evidenz auch bei Lautsprecher- und Kopfhörerwiedergabe ohne Verwendung einer Binauralsynthese nachweisbar.

Der Hörversuch in Experiment VII untersucht den Effekt eines bei Raumdivergenz auf den Abhörraum eingestelltes DRR der Synthese auf die Wahrnehmung von Externalität der Hörereignisse. Die in diesem Hörversuch eher untrainierten Testpersonen erreichen in der DRR-Einstellung geringfügig größere IQDs als in Experiment VI. Die Mediane liegen gemäß Experiment VI im Bereich des DRR des Abhörraumes. Es werden tendenziell ebenfalls größerer DRR-Werte (weniger hallige Synthese) eingestellt.

Die Abfrage bzgl. der Externalität der Hörereignisse bei Raumdivergenz, aber auf den Abhörraum eingestelltem DRR der Synthese liefert nicht den erwarteten Anstieg an Externalisierung. Tendenziell erreichen zwar die auf DRR-Raumkongruenz eingestellten Teststimuli höhere Externalitätsindizes als die raumdivergenten Signale, allerdings fallen diese Anstiege im Allgemeinen nicht signifikant aus ( $p < 0,05$ ). Das DRR ist ein reliabler, anhand der internen Referenz, einstellbarer raumakustischer Parameter der Synthese. Er wirkt sich allerdings nur geringfügig auf die Externalisierung und den Raumdivergenzeffekt aus.

Es wird vermutet, dass die Anpassung eines reinen energiebasierten raumakustischen Parameters nicht zur gewünschten Übereinstimmung der Reizmuster der Synthese mit den gespeicherten Reizmuster bzw. der Erwartungshaltung führt. Dieses lässt sich in argumentative Übereinstimmung mit dem Clifton-Effekt bringen, in dem die zeitliche Struktur der Rückwürfe der entscheidende Merkmalsträger ist. Lediglich eine Verhallung der Signale führt nicht zwangsläufig zu einem Anstieg an Externalisierung.



### 5.3 Synthese von BRIRs

Nachstehende Abschnitte beschreiben verschiedene Methoden zur Anpassung von zeit- und energiebasierten akustischen Parametern aufgenommener BRIRs. Hierdurch werden neue BRIRs synthetisiert, die beim Hörer einen veränderten Distanzeindruck hervorrufen sollen.

#### 5.3.1 Anpassung der Initialzeitlücke

Nachstehende Methode synthetisiert neue BRIRs für gewünschte Positionen im Raum aus einer einzelnen gemessenen BRIR. Die gemessene BRIR wird durch Veränderung der Initialzeitlücke (ITDG) an die gewünschte neue Position im Raum angepasst. Abbildung 5.12 verdeutlicht den Ansatz als Blockschaltbild. Die beschriebene Methode basiert auf Arbeiten von Simone Füg [33, 5] und wurde von Christina Mittag weiterentwickelt und evaluiert [40, 12].

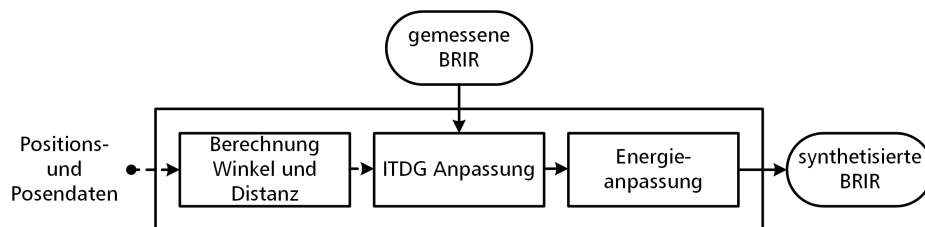


Abbildung 5.11: Methode zur Synthese neuer BRIRs basierend auf der Messung einer BRIR und Anpassung der Initialzeitlücke (ITDG); Abbildung nach Mittag [40].

Für die Erstellung der neuen BRIR werden die Positions- und Posendaten der gemessenen Position und der zu synthetisierenden neuen Position benötigt. An der Messposition liegt ein Satz von BRIRs für die verschiedenen Posenrichtungen vor. Die Winkelauflösung entspricht hierbei der gewünschten Auflösung. Eine Interpolation des Direktschallanteils zur Erhöhung der Auflösung ist hierfür ebenfalls denkbar. Anschließend wird die Distanz zwischen der zu synthetisierenden Position und der Messposition berechnet. Ausgehend von der gewünschten Posenrichtung an der Syntheseposition wird der entsprechende Winkel an der Messposition bestimmt.

Auf Grundlage der berechneten Distanzänderung zwischen Mess- und Syntheseposition wird die ITDG der gemessenen BRIR entsprechend angepasst. Hierfür wird eine Methode verwendet, welche von Simone Füg entwickelt und evaluiert wurde [33, 5, 16]. Gemäß der synthetisierten Distanz wird abschließend eine Energieanpassung der neuen BRIR vorgenommen. Hierbei wird der Direktschall gemäß einem Energieabfall von 6 dB und der Nachhall mit einem Abfall von 1,5 dB je Entfernungsverdoppelung angepasst. Hierdurch wird auch eine implizite distanzabhängige Anpassung des DRR erreicht. Weitere Details zu der verwendeten Methode sind ebenfalls der Arbeit von Simone Füg [33, 5] zu entnehmen.

Die angewendete Energieanpassung führt allerdings dazu, dass der zeitliche Energieverlauf (EDC) weitestgehend dem der Messposition entspricht. Lediglich durch die ITDG-Anpassung und damit verbundene Stauchung bzw. Streckung des Verlaufs der ersten Reflexionen wird

eine Veränderung der EDC vorgenommen. Diese Änderung entspricht aber nicht zwingend dem EDC-Verlauf an der Syntheseposition.

Weiterhin wird durch die Streckung oder Stauchung des zeitlichen Verlaufs der ersten Reflexionen eine Veränderung der spektralen Zusammensetzung und somit eine klangliche Verfärbung verursacht. In Untersuchungen von Hassager et al. [106] konnte belegt werden, dass eine spektrale Glättung und somit Veränderung der spektralen Zusammensetzung des Nachhalls keinen oder lediglich geringen Einfluss auf die Externalisierung von Hörereignissen hat.

Die Relevanz dieser akustisch nicht korrekten Anpassung auf die räumliche auditive Wahrnehmung lässt sich anhand von Untersuchungen zur Identifikation der Abhörposition im Raum abschätzen. Shinn-Cunningham [195] schlussfolgert aus ihrer Untersuchung, dass es trainierten Hörern möglich ist zu unterscheiden, ob sie sich mitten im Raum oder nahe an einer Wand befinden. Allerdings ist eine weiterführende Unterscheidbarkeit oder Identifikation zwischen mehreren Positionen im Raum nicht ohne weiteres möglich.

Aufbauend darauf untersuchten Neidhardt und Klein et al. den Einfluss von Kopfbewegungen [156] und Training [130] auf die Identifikationsleistung von Positionen im Raum. Es zeigt sich, dass die gezielte Zunahme von Kopfbewegungen keinen signifikanten Einfluss hat. Die Untersuchungen zum Einfluss von Training bestätigen, dass es ohne Training den Hörern nicht gelingt, zwischen verschiedenen Positionen im Raum zu unterscheiden. Ein Training führt bei 6 von 21 (29%) Hörern zu einer signifikanten Unterscheidbarkeit von Positionen im Raum. Es zeigte sich allerdings auch eine deutliche Abhängigkeit von den getesteten Positionen im Raum (bspw. Wandposition und mitten im Raum; siehe hierzu auch [195]).

Weitere Studien zeigen ergänzend dazu, dass es Hörern schwer fällt, sich auf Grundlage von akustischen Merkmalen korrekt im Raum einzuordnen [34, 170]. Es bleibt dennoch offen, inwieweit bspw. in augmentierten Audioumgebungen ein Vergleich von akustischen Merkmalen zwischen realen Schallquellen im Raum mit synthetisierten virtuellen Schallquellen zu einer erhöhten Diskriminationsfähigkeit führt. Es ist weiterhin offen, inwieweit eine reale Schallquelle, aber mit unterschiedlicher Position, akustischen Eigenschaften und Audiosignal als Referenz für eine virtuelle Schallquelle dienen kann.

### 5.3.2 Interpolationsmethoden

**Lineare Interpolation** - Die vorgestellte Methode der Interpolation soll einen schrittweisen Übergang zwischen zwei gemessenen BRIRs (Start- und Ziel-BRIR) unterschiedlicher Distanz ermöglichen. In einem ersten Schritt wird eine lineare Interpolation im Zeitbereich zwischen den Amplitudenwerten einer BRIR realisiert. Der Abtastwert  $y_t(a)$  des Interpolationsindex  $a$  zum Zeitpunkt  $t$  wird linear interpoliert. Für zwei beliebige Start- und Endpunkte der Interpolation zwischen  $(x_1, y_1)$  und  $(x_{anz+1}, y_{anz+1})$  mit  $anz$  Interpolationsschritten ergibt sich nachstehende Berechnungsvorschrift, wobei  $x_a$  mit  $a=0, 1, \dots, anz$  der x-Wert der Stützstelle am Interpolationsindex  $a$  ist.

$$y_t(a) = y_t(x1) + \frac{y_t(x_{anz+1}) - y_t(x1)}{x_t(x_{anz+1}) - x_t(x1)}(x_a - x1) \quad (5.1)$$

In den später beschriebenen Synthesemethoden werden ggf. verschiedene Bereiche der BRIRs getrennt betrachtet. Für die eigentliche Interpolation zwischen zwei BRIRs, werden die BRIRs in eine Direktschallanteil und einen Raumanteil aufgetrennt. Lediglich der Raumanteil soll durch die beschriebenen Methoden neu synthetisiert werden. Der Direktschall wird als HRIR ohne Raumanteil betrachtet und als konstant angenommen. Die Auftrennung wird in diesem Fall bei einem Zeitpunkt (bspw.  $t_{trenn}=1,5$  ms) nach dem Maximum im Direktschallteil definiert. Innerhalb dieses Zeitraums sollen alle durch den Körper induzierten Reflexionen das Trommelfell erreicht haben [124]. Die Trennung selber wird durch zwei sich überlappende Fensterfunktionen realisiert. Die Fenster überschneiden sich bei  $t_{trenn}$  und einer normalisierten Amplitude von  $\sqrt{0,5}$ .

Es findet abschließend eine Energieanpassung des Direktschallanteils und der Gesamtverzögerung bezogen auf die abzubildende Distanz statt. Die Direktschallenergie wird gemäß dem quadratischen Entfernungsgesetz und die Verzögerung gemäß der Laufzeit angepasst. Der Direktschallanteil und der neu synthetisierte Nachhallanteil werden anschließend wieder zusammengeführt und für die weitere Binauralsynthese zur Verfügung gestellt. Eine genauere Beschreibung der Methoden, sowie eine Gegenüberstellung mit anderen Verfahren sind der Masterarbeit von Rebecca Sass [44] zu entnehmen.

**MTB-Ansatz** - Die folgende Methode stellt einen Sonderfall der Interpolation zwischen zwei Übertragungsfunktionen dar. Er basiert auf den Ansätzen des Motion-Tracked Binaural Mikrofonarrays von Algazi et al. [48]. Eine Realisierung dieses Arrays sowie eine perzeptive Evaluierung einzelner Systemkomponenten findet sich bspw. in der Magisterarbeiten von Ross [180] sowie in der Diplomarbeit von Hesse [36].

Die Kernidee des Ansatzes ist eine Interpolation benachbarter Mikrofon-signale im MTB-Mikrofon-Array zur Erzeugung neuer virtueller Mikrofonpositionen. Hierfür werden die Signale (Impulsantworten oder am Mikrofon anliegende Audiosignale) einer Tiefpass- und Hochpassfilterung zugeführt. Die Übergabefrequenz sollte im Bereich von etwa 1,5 kHz liegen. Diese Frequenz ist aus der unterschiedlichen frequenzabhängigen Relevanz der binauralen Merkmale ITD und ILD zur Wahrnehmung der Schalleinfallrichtung motiviert.

Abbildung 5.12 verdeutlicht eine Adaptation des MTB-Ansatz zur Synthese neuer BRIRs als Blockschaltbild. Die hier beschriebene Methode wurde von Christina Mittag entwickelt und evaluiert [40, 12] und basiert auf Arbeiten von Simone Füg [33].

Die für die Erzeugung neuer BRIRs notwendigen gemessenen BRIRs (mindestens 2) werden gemäß ihrer neuen Position in ihren akustischen Eigenschaften (ITDG und Intensität) angepasst. Anschließend werden die BRIRs in einen Direktschallanteil, frühe Reflexionen und Nachhallanteil aufgetrennt. Die Abtrennung des Nachhalls wird durch die perceptual mixing time

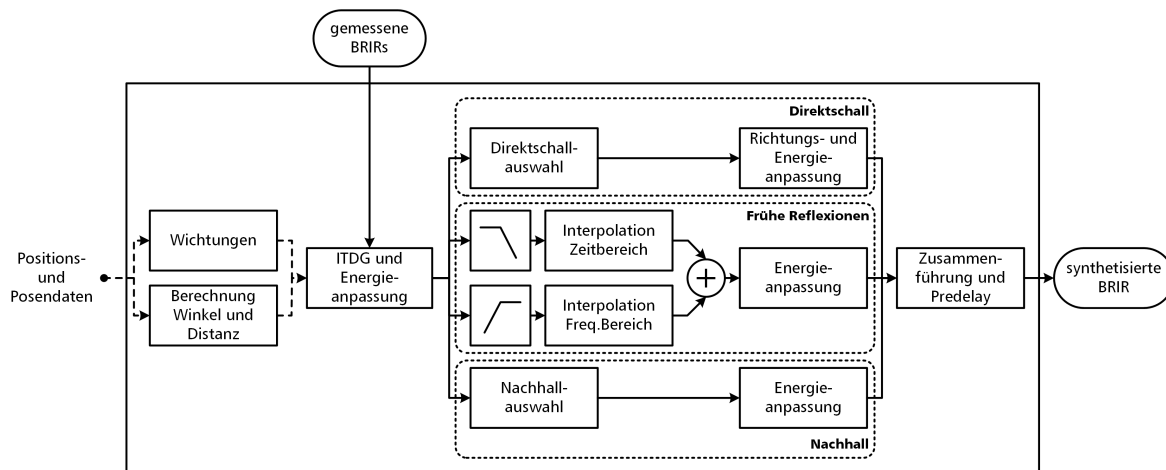


Abbildung 5.12: Methode zur Synthese neuer BRIRs basierend auf der Messung zwei oder mehrerer BRIRs und Interpolation im Zeit- und Frequenzbereich; Abbildung nach Mittag [40].

bestimmt. Diese bestimmt den Zeitpunkt des Übergangs von noch der Richtungswahrnehmung zuträglichen Reflexionen und dem diffusen Nachhall [141].

Der für die neue BRIR benötigte Direktschallanteil wird aus der BRIR der nächstgelegenen gemessenen BRIR entnommen. Bei Betrachtung der BRIRs für linkes und rechtes Ohr sind hier die richtungsrelevanten Merkmale ITD und ILD für den Direktschall enthalten. Die Signalanteile der frühen Reflexionen werden gemäß dem MTB-Ansatzes in einen tieffrequenten und einen hochfrequenten Anteil aufgetrennt und einer entsprechenden Interpolation zugeführt. Die tieffrequenten Signalanteile durchlaufen eine lineare Interpolation im Zeitbereich zur Anpassung der zeitbasierten Merkmale. Die hochfrequenten Anteile dienen der Anpassung frequenzabhängiger intensitätsbasierter Merkmale durch Interpolation im Frequenzbereich. Die hochfrequenten Signalanteile werden in komplexe Spektren überführt. Die Realanteile werden linear interpoliert. Die fehlende Phaseninformation wird ebenfalls der nächstgelegenen gemessenen BRIR entnommen. Anschließend wird das so interpolierte komplexe Spektrum in den Zeitbereich überführt. Die einzelnen Zeitbereiche der neuen BRIR werden in ihrer Energie an die neue Position im Raum angepasst und zusammengeführt [40]. Abschließend erfolgt die Anpassung der Vorverzögerung gemäß der Laufzeit bzw. Distanz der neuen Position zur Schallquelle.

**Dynamic Time Warping** - Die Abbildung 5.13 verdeutlicht den Ansatz zur Interpolation bei vorheriger Neuordnung der zeitlichen Struktur der Reflexionen in einer BRIR. Die Neuordnung erfolgt durch Anwendung der Prinzipien des Dynamic Time Warping (DTW). Der DTW-Ansatz ist eine Erweiterung der zeitlichen Interpolation der frühen Reflexionen zwischen einer Start- und einer Ziel-BRIR. Die vorgestellte Methode wurde von Rebecca Sass [44] entwickelt und von Nicolas Pachatz weiter angepasst [42].

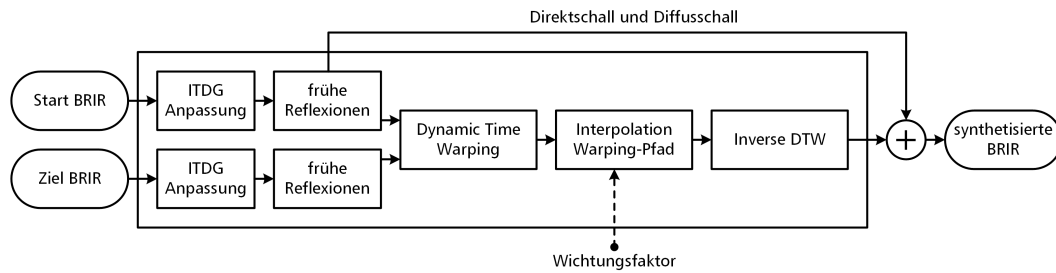


Abbildung 5.13: Methode zur Synthese neuer BRIRs basierend auf der Interpolation im Zeitbereich zwischen einer Start- und Ziel-BRIR bei Anwendung der Dynamic Time Warping (DTW) Methode; Abbildung nach Pachatz [42].

Die DTW-Methode beinhaltet eine Neuordnung der sich in einer BRIR befindlichen Abtastwerte. Dieses erfolgt durch Stauchung und Streckung der zugrundeliegenden Zeitachse. Das Kriterium der Transformation ist die Minimierung der Distanz zwischen einer Start- und Ziel-BRIR. Es werden Korrespondenzen und Ähnlichkeiten in den Signalsequenzen gefunden. Hierfür wird bspw. euklidische Distanz als Distanzmaß zwischen den Samplewerten der BRIRs verwendet. Die ursprüngliche Anwendung der Methode ist die Signalvorverarbeitung von Sprachsignalen zur Verringerung des Einflusses der Sprechgeschwindigkeit in der Spracherkennung [175]. Eine optimale zeitliche Anordnung der zeitlichen Struktur zweier BRIRs ist eine Voraussetzung für die Interpolation. Die Start- und Ziel-BRIRs werden so zueinander angeordnet, dass sich entsprechende Reflexionen auf einer gemeinsamen Zeitachse an gleicher Position befinden. Dieses verringert das zeitliche Verschmieren von einzelnen Reflexionen bei der Interpolation. Für die Verwendung in der hier vorliegenden Binauralsynthese werden die Ziel- und Start-BRIR zu Beginn durch die Anpassung der ITDG auf einen gleichen Startwert gesetzt.

### 5.3.3 Exp. VIII: ITDG-basierte Distanzanpassung

Für eine Verwendung in einem adaptiven Binauralsynthesystem wird eine ITDG-Anpassung gemessener BRIRs zur Erzeugung neuer BRIRs mit verschiedener Distanzwahrnehmung untersucht. Hierfür wird ein von Simone Füg entwickelter Algorithmus verwendet. Ausführlichere Informationen zur Signalverarbeitung sind den Literaturquellen [33, 5] zu entnehmen.

#### 5.3.3.1 Untersuchungsgegenstand

Es wird die Synthese von BRIRs an unterschiedlichen Positionen im Raum evaluiert. Die Distanzwahrnehmung eines Hörereignisses in Abhängigkeit von der Distanz der virtuellen Quelle zum Hörer bei gemessenen und neu geformten BRIRs wird untersucht.

Nachstehende Frage soll beantwortet werden: Welcher mittlere Fehler in der Distanzwahrnehmung ergibt sich für Hörereignisse bei Synthese mittels gemessenen und angepassten BRIRs?

### 5.3.3.2 Versuchsaufbau

**Anpassungsmethode** - Aus einer an einer bestimmten Distanz zur Messposition gemessenen BRIR werden neue BRIRs an verschiedenen Distanzen synthetisiert. Die Veränderung der gemessenen BRIR erfolgt durch Änderung der ITDG. In einem ersten Schritt wird eine allgemeine normalisierte ITDG-Kurve in Abhängigkeit der Distanz aus gemessenen BRIRs unterschiedlicher Distanz berechnet. In der zu bearbeitenden BRIR wird ein Bereich definiert, in dem die Formung angewendet wird (hier: 1,5 ms - 80 ms nach dem Direktschall). Der ITDG-Formungsalgorithmus berechnet mithilfe der allgemeinen ITDG-Kurve eine zeitliche Dehnung bzw. Kompression der BRIR für eine vorgegebene neue Distanz. Die Kompression bzw. Dehnung wird durch eine dynamische Kompressionscharakteristik und eine inverse Zuordnung der Amplitudenwerte innerhalb der BRIR durchgeführt.

**Audiosignale** - Es werden als Audiosignale ein nachhallfreies Saxofonstück und eine weißgefärbte Rauschpulsfolge (fünf Rauschpulse mit je 30 ms Dauer und 3 ms cosinusförmige Ein- und Ausblendung, 70 ms Pause zwischen den Pulsen und einem Frequenzbereich von 125 Hz bis 20 kHz) verwendet. Das Saxofonsignal wird aufgrund seiner zur menschlichen Sprache ähnlichen spektralen und tonalen Zusammensetzung [162, 204], aber ohne die Distanzwahrnehmung beeinflussende Faktoren, wie Artikulation und Bekanntheit, verwendet [64].

**Räume** - Ein Hörlabor (HL; siehe Kapitel 2.5) und ein TV Studio (ML; siehe Kapitel 2.5) mit unterschiedlichen raumakustischen Eigenschaften werden für die Messung der BRIRs aus verschiedenen Richtungen verwendet. Der Hörversuch wird im Hörlabor durchgeführt.

**Objektpositionen** - Die trockenen Audiosignale werden mit BRIRs aus der horizontalen Richtung  $330^\circ$  (Rechtsorientierung) gefaltet. Die gemessenen Distanzen im Hörlabor sind: 0,7 m, 1,0 m, 1,4 m, 2 m, 2,8 m und 4,0 m. Im TV Studio sind die gemessenen Entfernungen 0,7 m, 1,4 m, 2,0 m, 4,0 m, 5,6 m und 7,9 m. Die Richtung wird aufgrund der in den BRIRs enthaltenen binauralen Merkmalen gewählt. Gemessene und neu geformte BRIRs werden für die Erzeugung der Teststimuli verwendet. Die Schallintensität wird als ein weiteres Distanzmerkmal durch Normalisierung der Gesamtenergie ausgeschlossen.

### 5.3.3.3 Versuchsdurchführung

**Testpersonen** - An dem Hörversuch nehmen 22 Testpersonen im Alter von 20 und 31 Jahren (14 männliche und 8 weibliche Teilnehmer) teil. Ihr mittleres Alter beträgt 25,5 Jahre (Varianz=9,4 Jahre). Zwölf von 22 Testpersonen gaben Erfahrungen an der Teilnahme an Hörversuchen an. Eine anhand der Bewertungen durchgeführte Nach-Selektion der Testpersonen, stellt der späteren Auswertung die Bewertungen von 16 Testpersonen zur Verfügung. In der Arbeit von Füg [33] wird diese Auswahlprozedur genauer beschreiben. Sie basiert im Wesentlichen darauf,

dass die Bewertungen derjenigen Testpersonen ausgeschlossen werden, die die Referenz in mindestens drei von zwölf Fällen um mindestens +/-10 Skalenwerte verfehlt haben.

**Bewertungsskala** - Ein Multi-Stimulus-Test mit versteckter Referenz wird für den Hörtest verwendet. Die wahrgenommene Distanz des Hörereignisses wird im Vergleich zu einer Referenzdistanz bewertet. Der minimale und maximale Skalenwert wird mit „näher“ und „weiter entfernt“ beschrieben. Es wird eine 100 Punkteskala auf der „0“ näher, „50“ die Referenzdistanz von 2 m und „100“ weiter entfernt entspricht. Das Referenzsignal wird durch eine gemessene BRIR aus 2 m Entfernung repräsentiert. Ein Lautsprecher wird als Attrappe an der Referenzdistanz als visuelles Merkmal positioniert. Stimuli, die aus gemessenen BRIRs und durch das ITDG-Verfahren neugeformten BRIRs erstellt werden, dienen als Teststimuli.

Der zweite Durchgang ist in zwei Teile unterteilt. Im ersten Teil bewerten die Testpersonen die Stimuli entsprechend der wahrgenommenen Distanz relativ zu einer gegebenen Referenz mit der Multi-Stimulus-Testmethode. Gemessene BRIRs aus dem Hörlabor und zusätzliche nicht binaurale mittels Intensitäts-Stereo-Panning erzeugte Stimuli werden in diesem Test verwendet. Die nicht binauralen Stimuli stellen Stimuli für In-kopf-Lokalisierung dar und bilden in diesem Testdurchgang einen Anker für die Bewertungen. Jede Testperson bewertet zehn Stimuli während des ersten Teiles des zweiten Durchganges ( $(4_{\text{Distanzen}} + 1_{\text{nicht binaural}}) \times 2_{\text{Audiosignale}}$ ).

#### 5.3.3.4 Ergebnisse

Die Analyse der Hörversuche basiert auf den Medianen und Quartilen der abgegebenen Bewertungen für den verwendeten Algorithmus zur Formung neuer BRIRs, den untersuchten Räumen, Distanzen und Audiosignalen auf einer 100 Punkteskala zur Bewertung der wahrgenommenen Distanz. Zur Berechnung eines Fehlerindex wird eine Zielkurve mit linearem Verhalten zwischen den Skalenpunkten  $d_{min} = 0$  Skalenwerte und  $d_{ref} = 50$  Skalenwerte und zwischen  $d_{ref} = 50$  Skalenwerte und  $d_{max} = 100$  Skalenwerte definiert. Der Fehlerindex für jede Testperson errechnet sich aus der mittleren Abweichung der Bewertungen ( $d_n$ ) mit der Zielkurve ( $d_{soll,n}$ ) und wird formal beschrieben durch ( $N = \text{Anzahl an Bewertungen}$ ):

$$\frac{1}{N} \sum_{n=1}^N d_n - d_{soll,n} \quad (5.2)$$

Ein Fehlerindex größer als Null bezeichnet eine Überschätzung der Distanz. Ein Index kleiner als Null deutet auf eine Unterschätzung hin. Die Auswertung ergibt keine signifikanten Unterschiede zwischen den beiden verwendeten Audiosignalen. Für die weitere Darstellung werden die Bewertungen der beiden Audiosignale zusammengefasst.

Abbildung 5.14 zeigt die Fehlerindizes für den verwendeten Algorithmus und die beiden Räume im Test. Der Index für die gemessenen BRIRs zeigt eine leichte Überschätzung der

dargebotenen Distanz. Die Interquartilabstände (IQA) sind vergleichbar für die Auralisierung der Audiosignale in den beiden Räumen. Die IQA steigt für die Auralisierung mit den geformten BRIRs und für die Wiedergabe des TV Studios an. Weiterhin ist eine Unterschätzung der Distanz für die geformten BRIRs sichtbar. Die Ergebnisse zeigen dennoch, dass eine kontrollierte Steuerung der Distanzwahrnehmung durch eine Formung gemessener BRIRs möglich ist.

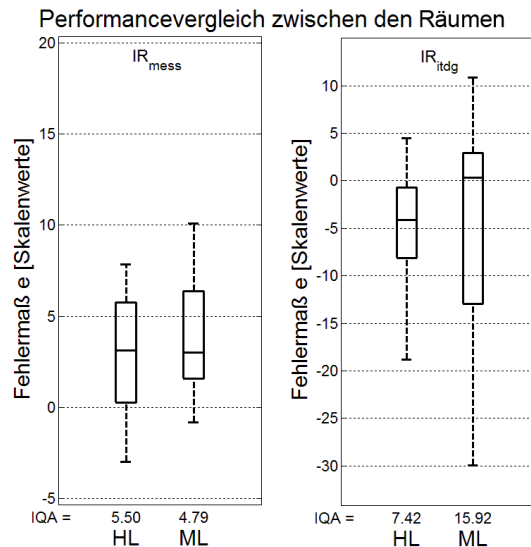


Abbildung 5.14: Fehlerindex als Median und Quartile in Skalenwerten (100 Punkteskala); links: gemessene BRIRs aus den beiden Räumen, rechts: ITDG-geformte Referenz-BRIRs aus den beiden Räume; HL=Hörlabor, ML=TV-Studio; IQA=Interquartilabstand; Abbildung aus Füg [33].

Die Ergebnisse zeigen eine geringe Überschätzung der Distanz bei der Verwendung der gemessenen BRIRs. Die Verwendung der neu geformten BRIRs führt zu einer Unterschätzung der Distanz. Dennoch ist durch Anwendung synthetisierter und geformter BRIRs eine Steuerung der Wahrnehmung von Distanz eines Hörereignisses möglich. Die Verwendung von Formungsalgorithmen zur Manipulation der Distanzwahrnehmung bei gleichzeitiger Reduktion der Anzahl an nötigen real gemessenen BRIRs ist für die Binauralsynthese geeignet.

### 5.3.4 Exp. IX: Interpolationsbasierte Distanzanpassung

Es werden Methoden zur Synthese binauraler Raumimpulsantworten (BRIRs) mittels Interpolation aus gemessenen Datensätzen evaluiert. Es wird auf die Interpolation zwischen BRIRs mit unterschiedlicher Distanz in zwei verschiedenen Räumen eingegangen. Die Ausführungen beziehen sich auf Experimente und Arbeiten, die im Rahmen der vorliegenden Dissertation entstanden sind und unter [22, 30] veröffentlicht wurden. Die Methoden zur Interpolation zwischen BRIRs ist ebenfalls in der Masterarbeit von Rebecca Sass [44] dokumentiert.



#### 5.3.4.1 Untersuchungsgegenstand

Es werden zwei Algorithmen zur Synthese von BRIRs untersucht. Diese sind zum einen eine lineare Interpolation im Zeitbereich sowie eine Interpolation im Zeitbereich, der eine zeitliche Anpassung der Samplewerte basierend auf Dynamic Time Warping vorangestellt ist. Die Anwendbarkeit der BRIR-Synthesemethoden wird anhand der Qualitätsmerkmale wahrgenommene Distanz und Externalität nachgewiesen.

Nachstehende Fragen sollen beantwortet werden: Lässt sich durch Methoden der Interpolation zwischen zwei gemessenen BRIRs eine gezielte Distanzwahrnehmung hervorrufen? Wie ändert sich die Externalisierung der Hörereignisse bei Verwendung neu synthetisierte BRIRs für unterschiedliche Distanzen bei Raumdivergenz und Raumkonvergenz?

#### 5.3.4.2 Versuchsaufbau

**Binauralsynthesystem** - Es werden Messungen von Kunstkopf-BRIRs aus einem Hörlabor (HL, siehe Kapitel 2.5) und einem Seminarraum (SR, siehe Kapitel 2.5) an verschiedene Schallquellpositionen verwendet. Für die Synthese über Kopfhörer werden invertierte HPTFs eines STAX Lambda Pro bei Messung auf dem Kunstkopf verwendet. Die HPTFs sind über drei Aufnahmen, bei Repositionierung des Kopfhörers, gemittelt. Die Inverse der HPTF wird durch einen Least-Square-Ansatz erstellt [188]. Die Synthese der neuen BRIRs wird durch lineare Interpolation mit und ohne Anwendung des DTW-Ansatzes zwischen den Amplitudenwerten einer Start- und Ziel-BRIR durchgeführt. Die angewendeten Methoden sind den vorangestellten Abschnitten 5.3.2 und 5.13 zu entnehmen.

**Messpositionen** - Die Messpositionen im Hörlabor und im Seminarraum befinden sich in einer seitlichen Richtung von 30° links der Blickrichtung des Kunstkopfes. Die Richtung wird ausgewählt, damit auch binaurale Richtungsmerkmale in der Synthese enthalten sind. Es werden in einer Entfernung zum Kunstkopf von 1 m und 5 m die Übertragungsfunktionen gemessen. Als Schallquelle kommt ein Geithain Mo-2 Lautsprecher zur Anwendung. Die Ohrhöhe des Kunstkopfes sowie die akustische Mitte des Lautsprechers sind auf ca. 1,26 m angeordnet.

**Audiosignale** - Ein Saxofonstück ohne Nachhall und eine weißgefärbte Rauschpulsfolge (fünf Pulse mit je 70 ms Dauer, 3 ms cosinusgefenstertes Ein- und Ausblenden und 500 ms Stille zwischen den Pulsen mit einer Bandbreite von 125 Hz bis 20 kHz) werden als Audiosignale verwendet. Das Saxofonsignal wird aufgrund seiner zur menschlichen Sprache ähnlichen spektralen und tonalen Zusammensetzung gewählt [162, 204], ohne dass die Distanzwahrnehmung durch Artikulation oder Bekanntheit beeinflusst wird [64]. Die Rauschpulsfolge wird aufgrund der guten Lokalisierbarkeit gewählt.

### 5.3.4.3 Versuchsdurchführung

**Bewertung Distanz** - Ein Multi-Stimulus Testdesign mit versteckter Referenz wird für die Evaluierung der wahrgenommenen Distanz verwendet. Die wahrgenommene Distanz der Stimuli wird mit einer Referenzdistanz auf einer Skala von „näher“ bis „weiter entfernt“ abgefragt. Auf der verwendeten 100 Punkteskala entspricht die 0 „näher“, 50 Skalenwerte einer mittleren Distanz und 100 Skalenwerte „weiter entfernt“. Die im Test vorkommende nächste Distanz von 1 m entspricht der gemessenen Start-BRIR. Die größte Distanz von 5 m entspricht der gemessenen Ziel-BRIR. Die anderen Distanzen sind durch die neu-synthetisierten BRIRs realisiert. Die am weitesten entfernte BRIRs wird ebenfalls als Referenz und versteckte Referenz verwendet. Ein Lautsprecher wird als Attrappe bei der Referenzdistanz als visuelles Merkmal platziert. Als Training werden den Testpersonen die Teststimuli ohne Abgabe einer Bewertung durch die Testpersonen, aber mit sichtbarer Benennung der Distanz, vorgespielt.

**Bewertung Externalität** - Die Externalität des Hörereignisses wird in einem Single-Stimulus Testdesign und einer Forced-Choice-Methode bewertet. Die Testperson gibt an, wo sich das Hörereignis befindet, in dem sie einen von drei Möglichkeiten auf einem graphischen Eingabegerät auswählt. Nachstehende Definitionen werden verwendet: a) „Die Schallquelle befindet sich vollständig in meinem Kopf oder ist sehr diffus.“, b) „Die Schallquelle ist extern. Sie befindet sich aber in der Nähe meiner Ohren oder meines Kopfes.“, c) „Die Schallquelle ist sehr extern und gut lokalisierbar.“. Der gesamte Stimulus muss angehört bis eine Bewertung abgegeben werden kann. Die synthetisierten BRIRs aus verschiedenen Distanzen aus dem Hörlabor und dem Seminarraum werden als Stimuli im Test verwendet. Eine Präsentation von nicht binauralen mittels Intensitätsstereophonie gepannten Audiosignalen, eine Wiedergabe über den Referenzlautsprecher und eine binaurale Synthese des Referenzlautsprechers werden als Stimuli im Training verwendet. Die Bezeichnungen der Trainingstimuli sind nicht sichtbar. Die Bewertung der Trainingsstimuli geschieht durch die oben beschriebene Methode.

**Testpersonen** - Vier weibliche und 12 männliche Testpersonen nehmen am Test teil. Das durchschnittliche Alter beträgt 24,5 Jahre. Sieben der 16 Testpersonen geben an, dass sie Erfahrungen an der Teilnahme in Wahrnehmungstest haben. Drei der Testpersonen geben an, Erfahrungen im Hören einer Binauralsynthese zu haben.

**Durchführung** - Die Hörtests werden in beiden Räumen in zwei Versuchsteilen zur Bewertung der Distanz und der Externalität durchgeführt. Die Reihenfolge der Teile wird für die Testpersonen zufällig angeordnet. Für den Distanztest ergeben sich acht Testreihen mit jeweils fünf zu bewertenden Stimuli in zufälliger Anordnung für jede Testperson. Die Testreihen sind die Kombinationen aus 2 Audiosignalen, 2 Räumen und 2 Synthesemethoden. Die Anzahl der Stimuli ergibt sich aus den gemessenen Positionen bei 1 m und 5 m sowie aus drei dazwischen synthetisierten Distanzen bei 2 m, 3 m und 4 m. Die identifizierte Referenzdistanz soll gemäß Anweisung mit 100 Skalenpunkten, das am nächsten wahrgenommene Hörereignis

mit 0 Skaleneinheiten bewertet werden. Die anderen Stimuli sind dazwischen anzuordnen. Den Testpersonen ist die Referenzdistanz durch eine aufgestellte Lautsprecheratmosphäre in 5 m Entfernung bekannt. Für die Bewertung der Externalität der Hörereignisse werden in beiden Testräumen 80 Stimuli in zufälliger Reihenfolge präsentiert. Die Anzahl ergibt sich aus 2 Audiosignalen, 2 Räumen, 2 Synthesemethoden und zweimaliger Präsentation.

#### 5.3.4.4 Ergebnisse

Die Auswertung der Distanzbewertungen basiert auf den Quantilen der Bewertungen für die einzelnen Distanzen, Kombination der Räume, Synthesemethoden sowie Audiosignal auf einer 100 Punkte Skala. Abbildung 5.15 zeigt die Bewertungen der wahrgenommenen Distanz aller Probanden für unterschiedliche Kombinationen von Abhörort und auralisiertem Raum. Eine signifikante Überschätzung der Distanz ist für alle Raumkombinationen sichtbar. Dies zeigt sich speziell für die mittleren Distanzen von 3 m und 4 m. Diese Ergebnisse sind vergleichbar mit eigenen anderen Experimenten mit gemessenen BRIRs [16]. Tendenziell sind Unterschiede in den Bewertungen für die verwendeten Audiosignale erkennbar. Zwischen den Bewertungen der beiden Algorithmen werden keine signifikanten Unterschiede gefunden. Beide Synthesemethoden sind für die Synthese von BRIRs und für eine Steuerung der Distanzwahrnehmung geeignet.

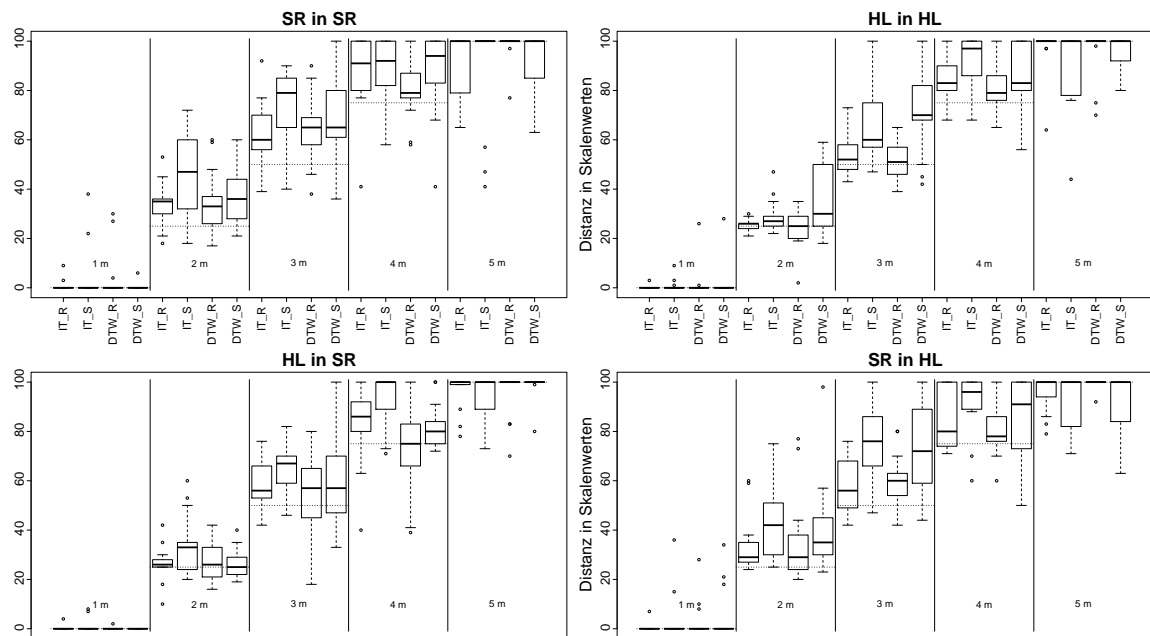


Abbildung 5.15: Distanzbewertungen als Quartile für die Interpolation der Distanz bei verschiedenen Kombinationen von Abhörort und auralisiertem Raum; SR=Seminarraum, HL=Hörlabor; DTW=Dynamic Time Warping, IT=Interpolation; gepunktete Linie bezeichnet die Zielbewertung bei gegebener Distanz.

Abbildung 5.16 zeigt die Bewertungen der Testpersonen für die Externalisierung der Hörereignisse. Die Analyse der Bewertungen basiert auf einem Externalitätsindex, welcher sich aus der Summe der Bewertungen „extern“ geteilt durch alle Bewertungen einer Testkondition ergibt. Der Externalitätsindex ist für die weniger nachhallbehafteten Konditionen (Synthese Hörlabor) konsistent niedriger als für den halligeren Seminarraum. Äußerungen der Testpersonen bestätigen dieses Ergebnis. Es wird konsistent berichtet, dass die Auralisierung des halligeren Raumes im weniger halligen Raum plausibler als die Wiedergabe des weniger halligen Raumes im halligen Raum ist. Es ist eine Sättigung des Indexes für die Wiedergabe des halligen Seminarraums im selben Seminarraum in Abhängigkeit der Distanz erkennbar. Keine signifikanten Unterschiede können zwischen der Verwendung der beiden Synthesemethoden für alle Raumkombinationen bestimmt werden. Unterschiede werden für die verwendeten Audiosignale (Rauschpuls und Saxofon) speziell für die Kondition “HL in HL” sichtbar. Die Größenordnung der abgebildeten Externalitätsindizes für die verschiedenen Raumkombinationen sind vergleichbar mit den Ergebnissen aus anderen Experimenten bei Verwendung von ebenfalls nicht individuellen BRIRs, aber gemessenen BRIRs aus verschiedenen Distanzen (siehe [16, 26]).

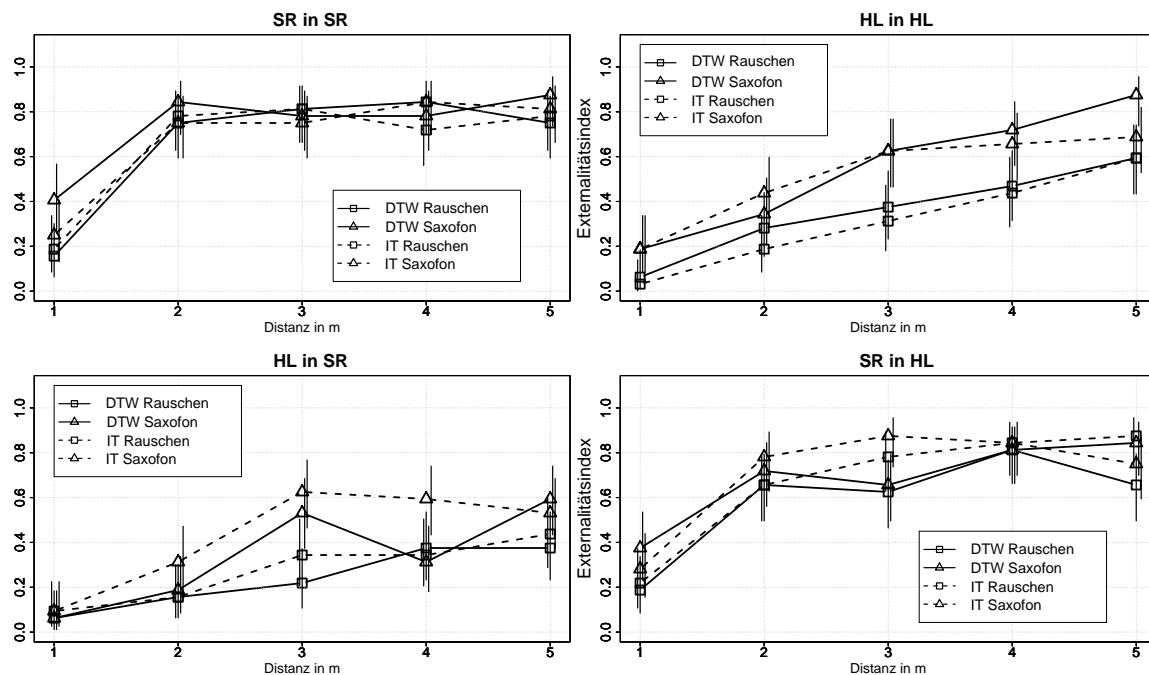


Abbildung 5.16: Externalitätsindizes und 95% Konfidenzintervalle für die Interpolation der Distanz bei verschiedenen Kombinationen von Abhörraum und auralisiertem Raum; SR=Seminarraum, HL=Hörlabor; DTW=Dynamic Time Warping, IT=Interpolation.

Die Ergebnisse des Experimentes zeigen eine geringe Distanzüberschätzung bei Verwendung neu synthetisierter BRIRs (siehe Abbildung 5.15). Keine signifikanten ( $p < 0,05$ ) Unterschiede werden zwischen den Synthesemethoden gefunden. Mit den beiden vorgestellten Algorithmen ist

eine gesteuerte Distanzwahrnehmung möglich. Eine Substitution von realen BRIR-Messungen durch Verwendung von Interpolationsmethoden ist möglich.

Es zeigt sich, dass der Anteil des Nachhalls einen Einfluss auf die wahrgenommene Externalität eines Hörereignisses hat. Der Externalitätsindex ist weitestgehend konstant für größere Distanzen. Bei kleineren Distanzen sinkt die wahrgenommene Externalität deutlich ab. Dieser Abfall tritt für den verwendeten halligeren Raum früher als bei dem weniger halligen Raum auf. Der Externalitätsindex ist am geringsten für die weniger halligen Bedingungen. Die Externalität steigt, wenn der synthetisierte Raum gleich dem Abhörraum ist (siehe Abbildung 5.16). Es scheint weiterhin eher plausibel zu sein, einen halligen Raum in einem weniger halligen Raum zu auralisieren als anders herum.

### 5.3.5 Exp. X: Synthese aus räumlich dünn-besetzten Messungen

Nachstehende Experimente und Untersuchungen basieren auf einer Zusammenführung verschiedener Methoden zur Synthese und Anpassung von gemessenen BRIRs. Die Methoden hierfür sind den vorangestellten Kapiteln und Abschnitten zu entnehmen.

**Motivation** - Die vorgestellte Methode ermöglicht die Auralisation virtueller Audioobjekte im realen Raum. Der Hörer soll sich in einer augmentierten auditorischen Umgebung frei im Raum bewegen können und eine plausible Hörwahrnehmung in Bezug auf die Lokalisation, Externalisierung und Passung der Hörereignisse zum Abhörraum haben. Zur Erreichung einer plausiblen und mit hoher Passung versehenen Audiowiedergabe sollen reale Messungen von BRIRs im Abhörraum verwendet werden. Hierdurch soll der Raumdivergenzeffekt minimiert werden. Es wird weiterhin ein extra-auraler Kopfhörer zur Wiedergabe verwendet, um die Schalle im realen Abhörraum während der Wiedergabe wahrnehmbar zu halten. Der Kopf des Hörers wird während der Wiedergabe in seiner Pose und Position im Raum durch ein Tracking-System erfasst. Die Berechnung der binauralen Audiosignale erfolgt in weitestgehender Echtzeit positions- und posengetreu.

Für eine plausible räumliche auditorische Wahrnehmung werden theoretisch für jede Sender-Empfänger-Positionsrelation ein Satz an BRIRs für das linke und rechte Ohr benötigt. Die benötigte Anzahl lässt sich gemäß der Lokalisationsungenauigkeiten und Toleranzen in der auditorischen Wahrnehmung reduzieren. Beispielsweise werden für die horizontale Richtungsabbildung eine Winkelauflösung der BRIRs von  $5^\circ$  als ausreichend angenommen. Dies trifft dann zu, wenn in der Systemevaluierung keine Paarvergleiche und weniger kritische Audiosignale (keine Rauschpulsfolgen) verwendet werden. Auf eine vertikale Richtungsabbildung soll aufgrund der Komplexität der Messungen der Übertragungsfunktionen zunächst verzichtet werden. Weiterhin sind die hier beschriebenen Methoden prinzipiell in gleicher Art für vertikal ausgerichtete Audioobjekte geeignet. Für die Positionen der Messpunkte im Raum werden Auflösungen von  $0,5\text{ m} \times 0,5\text{ m}$  und  $0,25\text{ m} \times 0,25\text{ m}$  in der nachstehenden Untersuchung evaluiert. Eine schematische Darstellung für das Experiment ist Abbildung 5.17 zu entnehmen.

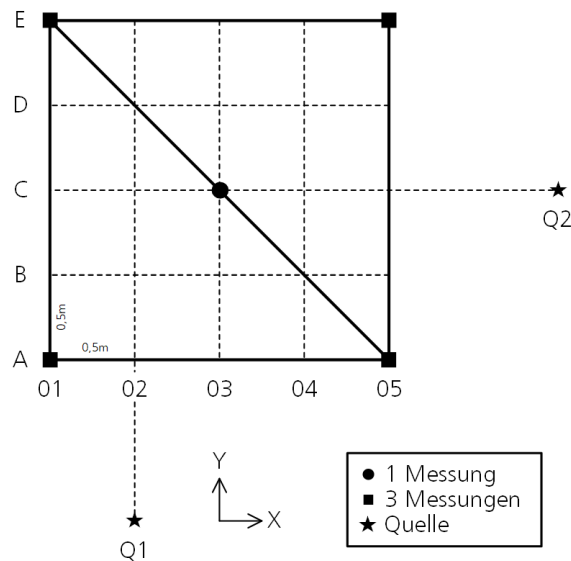


Abbildung 5.17: Schema der in Experiment A verwendeten Mess- und Auralisierungspositionen; Synthese von BRIRs an den Kreuzungspunkten des Gitters; Messpositionen gemäß Legende.

Selbst diese perceptiv motivierte Reduktion der Positionen bedeutet bspw. für eine Fläche von 2 m x 2 m mit einer Positionsauflösung von 0,5 m und bei einer horizontalen Richtungsauflösung von 5° insgesamt 1800 BRIR-Messungen (25 Positionen x 72 Winkel) für linkes und rechtes Ohr für jede abzubildende Schallquelle. Es ist daher das Ziel diese Anzahl deutlich zu reduzieren. Es wird untersucht, wie sich Methoden der Interpolation und ITDG-Anpassung für den Anwendungsfall anwenden lassen. Die benötigten BRIRs werden aus drei bzw. aus lediglich einer Messposition neu synthetisiert. Bei einer horizontalen Winkelauflösung ergibt dies für linkes und rechtes Ohr je Schallquelle 216 bzw. 72 BRIR-Messungen. Die Messung der verschiedenen Richtungen lässt sich durch einen Drehteller und geeigneter Messprozedure bequem automatisieren.

Die vorgestellten Methoden basieren auf der Masterarbeit von Christina Mittag [40] sowie der weiteren Veröffentlichung [12]. Das Experiment A entstand im Rahmen der Lehrveranstaltung “Advanced Psychoacoustics (Sommersemester 2017)“. Das Experiment B wurde in der Lehrveranstaltung “Praxiswerkstatt Hörtests (Wintersemester 2017/2018)“ durchgeführt. Allen teilhabenden Studierenden sowie Thomas Sporer, Annika Neidhardt und Florian Klein sei hier besonders gedankt.

### 5.3.5.1 Untersuchungsgegenstand

Es werden zwei Experimente vorgestellt, die die räumliche auditorische Wahrnehmung einer Synthese von neuen BRIRs aus wenigen Messpostionen untersuchen. Es werden die Effekte auf die Bewertungen der Qualitätsmerkmale Externalisierung, Lokalisation und Gesamteindruck untersucht.

Nachstehende Fragen sollen für das **Experiment A** beantwortet werden: In welchem Umfang unterscheiden sich die Bewertungen einzelner Qualitätsmerkmale zwischen der Verwendung synthetisierter BRIRs und gemessenen BRIRs? Zeigen sich für die beabsichtigte Binauralsynthese Abhängigkeiten von der Entfernung zwischen Audioobjekt und synthetisierter Abhörposition?

Nachstehende Fragen sollen für das **Experiment B** beantwortet werden: Ab welcher Distanz zwischen synthetisierter Position und für die Synthese verwendeter Messposition wird ein Unterschied zwischen der Synthese und Referenzmessungen wahrgenommen? Wie groß ist dieser Unterschied? Unterscheidet sich dieser Unterschied bei Verwendung unterschiedlicher Messpositionen?

**Experiment A** - Es soll die prinzipielle Machbarkeit der Ansätze für eine Fläche von 2 m x 2 m mit einer Positionsauflösung von 0,5 m aufgezeigt werden. Es werden Synthesemethoden der Interpolation der BRIRs aus drei Messpunkten und der Anpassung der ITDG aus einer Messposition evaluiert. Es werden die Qualitätsmerkmale Externalisierung, Lokalisierung sowie der allgemeine Eindruck untersucht. Weitere Details sind der Veröffentlichung [28] zu entnehmen.

**Experiment B** - Das zweite Experiment baut auf den Ergebnissen des ersten Experimentes auf und erweitert die augmentierte Fläche auf 4 m x 3,75 m mit einer Positionsauflösung von 0,25 m und bei Anwendung der ITDG-Anpassung. Die Reduktion liegt in diesem Fall bei einem Wert von 272/1 (16 x 17 abzubildende Positionen zu einer Messposition). Während in Experiment A die Fläche quasi mittig im Raum mit gleichmäßigen Abständen zu den Wänden liegt, ist in Experiment B an einer Seite der Fläche direkt eine Wand platziert. Die für die Synthese der BRIRs verwendete Messposition sowie die Synthesepositionen werden an unterschiedliche Abstände zur augmentierten Wand gesetzt. Es soll der Einfluss der Messposition auf die Wahrnehmung untersucht werden. Die Qualitätsmerkmale sind hierbei die Erkennung der neu synthetisierten Position aus räumlich benachbarten Referenzmessungen. Die Ausprägung der Verschlechterung wird ebenfalls abgefragt.

### 5.3.5.2 Versuchsaufbau

**Experiment A** - Es werden in einer 2 m x 2 m großen Fläche mit einer Positionsauflösung von 0,5 m BRIRs für linkes und rechtes Ohr für zwei Schallquellen im Raum bei einer horizontalen Winkelauflösung von 5° gemessen. Aus den 25 Messpositionen werden für die Synthese neuer BRIRs drei bzw. eine Position ausgewählt.

Abbildung 5.17 verdeutlicht die Positionen schematisch. Für die Messungen wird ein Kunstkopf (KEMAR) verwendet. Mithilfe eines fernsteuerbaren Drehtellers werden die jeweilige horizontale Ausrichtung des Kunstkopfes eingestellt. Als Schallquellen kommen Lautsprecher (ME Geithain MO-2) zur Anwendung. Die Ohrhöhe des Kunstkopfes und des akustischen

Zentrums der Lautsprecher entspricht ca. 1,6 m (in etwa Ohrhöhe einer stehenden Person). Der für alle 25 Positionen gemessene BRIR-Datensatz kann unter [11] heruntergeladen werden. Eine weiterführende Beschreibung findet sich in [10].

Es werden drei binaurale Wiedergabesysteme untersucht. Als Referenzsystem dient die Auralisierung aller 25 Positionen basierend auf den BRIR-Messungen an diesen Positionen (Abkürzung: 25M). Das erste Testsystem ist eine Auralisierung der Positionen basierend auf der Interpolation von BRIRs nach dem MTB-Ansatz (siehe Kapitel 5.3.2) bei Verwendung von drei Messungspositionen (Abkürzung: 3M). Das zweite Testsystem ist eine Auralisierung nach dem ITDG-Ansatz (siehe Kapitel 5.3.1) bei Verwendung einer Messposition (Abkürzung: 1M). Als Audiosignal wird ein Ausschnitt aus einem englischsprachigem Hörspiel (männlicher Sprecher) verwendet. Die Dauer des Audiosignal beträgt 9,8 Minuten. Spätestens nach dieser Zeit soll die Bewertung durch die Testperson abgeschlossen sein.

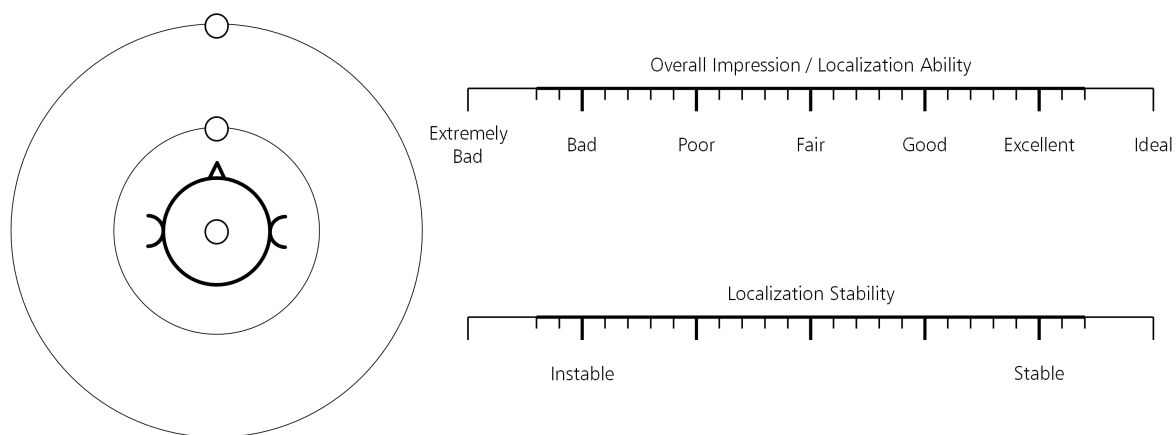


Abbildung 5.18: Skalen für die in Experiment A verwendeten Skalen; linke Skala zur Abfrage der Externalisierung, rechte Skalen angelehnt an ITU-T P.851 [119].

**Experiment B** - Der im Experiment A verwendete Testaufbau wird auf einen größeren Raum übertragen. Für die Auralisierung kommen reale Messungen mit einem Kunstkopf und  $5^\circ$  horizontaler Winkelauflösung für eine Schallquelle im Raum zur Anwendung. Als Testsystem wird der Ansatz der ITDG-Anpassung zur Synthese neuer BRIRs untersucht. Die zu auralisierende Fläche wird an einer Seite direkt von einer Wand begrenzt. Hierdurch wird es möglich, dass die für die BRIR-Synthese verwendete gemessene Position in ihrem Abstand zur Wand variieren kann. In dem Experiment werden verschiedene Kombinationen von Messposition und Syntheseposition untersucht. Es wird sich hierfür auf die Auralisierung eines Pfades beschränkt.

Abbildung 5.19 verdeutlicht die in dem Hörversuch untersuchten Konditionen. Die dargestellten Pfade von „B“ bis „Q“ entsprechen einem Bewegungspfad von der linken Wand in den Raum. Die räumliche Auflösung der Positionen beträgt 0,25 m. Für jede dieser Positionen wird eine Kunstkopfmessung für die Schallquelle „S1“ durchgeführt. Diese sind als Referenz „Ref“





Im Gegensatz dazu steht der Pfad II, welcher in weiterer Entfernung zur Schallquelle verläuft. Der Pfad III beschreibt eine diagonale Bewegung durch die zu auralisierende Fläche. Dieser wird ebenfalls als kritisch für Lokalisationsinstabilitäten angesehen, da dieser zusätzlich die größten Sprünge zwischen den synthetisierten Punkten im Raum erfasst. Der Pfad IV bewegt sich auf die Schallquelle in Ausrichtung der Quelle zu. Hierdurch sollen vor allem frontale und rückseitige Auralisierungen der Schallquelle untersucht werden. Es wird angenommen, dass dieser Pfad kritisch bzgl. der Externalisierung ist. Instabilitäten in der Lokalisation sollten hier weniger deutlich auffallen.

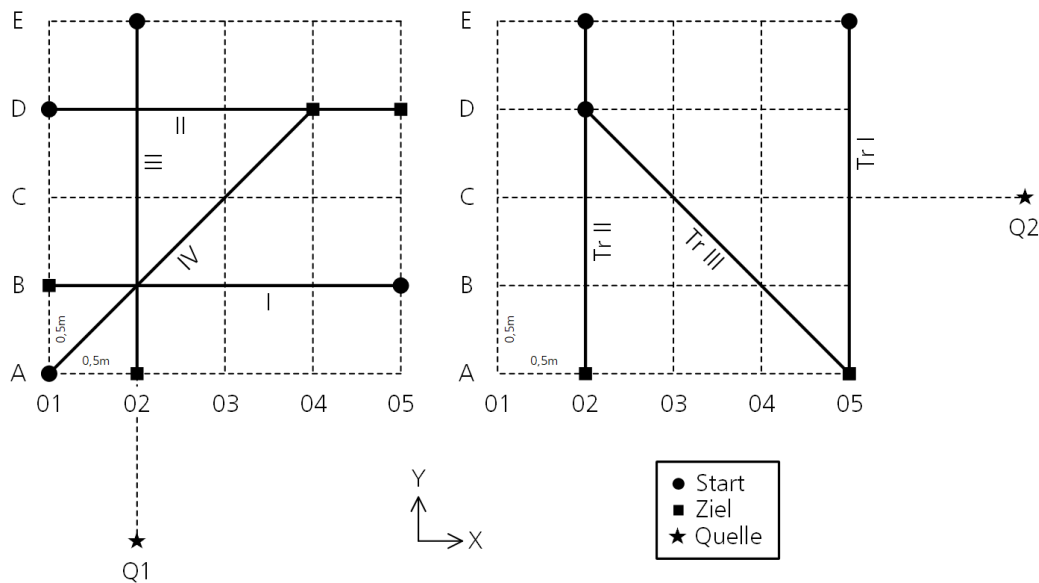


Abbildung 5.20: Pfade zum Ablaufen durch die Testpersonen in Experiment A; links=Pfade im Test, rechts=Pfade im Training.

Vor der Durchführung des eigentlichen Testes wird den Testpersonen die Möglichkeit eines Kennenlernens der Audiowiedergabe, der Qualitätsmerkmale und der Bewertungsmöglichkeit gegeben. Hierfür laufen die Testpersonen die in Abbildung 5.20 (rechts) angegebene Trainingspfade für die Auralisierung der zweiten Schallquelle ab. Die Testpersonen starten ihre Bewegung auf den Pfaden an den angegebenen Start-Positionen und durchlaufen den Pfad bis zur Ziel-Position. Anschließend können die Testpersonen frei entscheiden, wie sie sich auf dem Pfad hin und her bewegen wollen. Komplette Drehung des Körpers oder Rückwärtslaufen ist erlaubt. Die Reihenfolge der Präsentation der Pfade ist für jede Testperson zufällig angeordnet.

**Experiment B** - In dem Hörversuch können sich die Testpersonen auf dem auralisierten Pfad frei bewegen. Kopfdrehung und Körperdrehungen sind möglich. Die Testpersonen sollen die augmentierte Hörumgebung erkunden. In zufälliger Reihenfolge werden einzelne Positionen des Pfades durch synthetisierte BRIRs ersetzt. Die anderen Positionen entsprechen der

Auralisierung basierend auf der Referenzmessung. Die Aufgabenstellung ist es, diese Position bei der Bewegung auf dem Pfad zu erkennen. Ist ein Unterschied in der Wahrnehmung erkannt, soll dieser in seiner Stärke bewertet werden. Dies erfolgt auf einer Beeinträchtigungs-Skala mit den Skalenpunkten: „5.0 - nicht wahrnehmbar“, „4.0 - wahrnehmbar aber nicht störend“, „3.0 - geringfügig störend“, „2.0 - annehmbar aber störend“, „1.0 - nicht annehmbar“ (Skala gemäß ITU-R BS.1116 [116]).

Vor der eigentlichen Testdurchführung werden die Testpersonen bzgl. der auftretenden Unterschiede zwischen Synthese und Referenzmessung trainiert. Dies beinhaltet ein Ablauf des Pfades gemäß der Vorgabe im Test. Die Pfade beinhalten neu synthetisierte BRIRs aus den Messpositionen „B09“ und „P09“ mit Entfernungen zwischen Synthesposition und Messposition von  $>1,5$  m. Die Trainingsstimuli stellen somit große Unterschiede dar und sollten durch Testpersonen wahrnehmbar sein.

#### 5.3.5.4 Ergebnisse

**Experiment A** - Für dieses Experiment werden die Bewertungen von 18 Testpersonen (drei weiblich) mit einem mittleren Alter von 29 Jahren (Standardabweichung von 9 Jahren) ausgewertet. Elf von 18 Testpersonen berichten Erfahrung in der Teilnahme an Wahrnehmungsuntersuchungen. Sieben von 18 Testpersonen berichten Erfahrung in der Verwendung der Binauralsynthese. Keine Testperson berichtet über vorliegende Hörschädigungen.

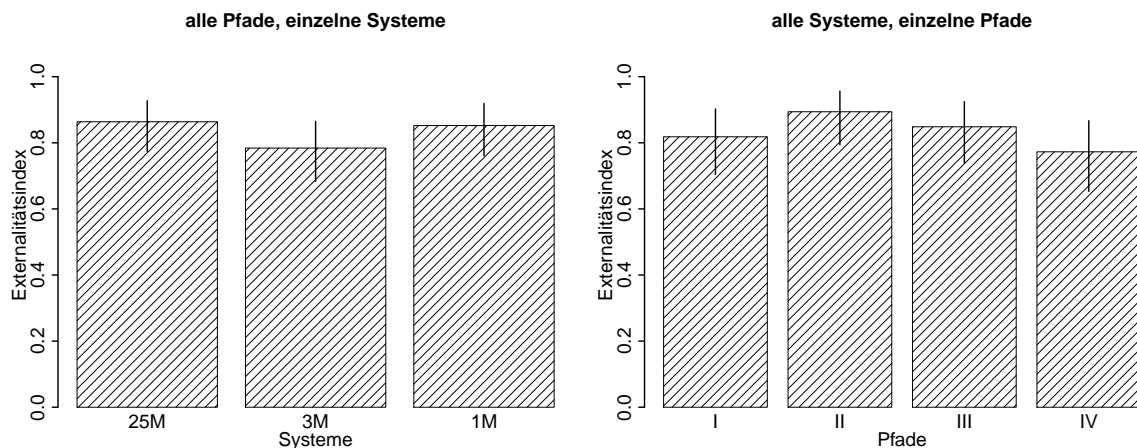


Abbildung 5.21: Externalitätsindizes der Bewertungen der Probanden mit 95% Konfidenzintervall; links: Zusammenfassung aller Pfade, rechts: Zusammenfassung aller Systeme.

Abbildung 5.21 verdeutlicht die Bewertungen der Testpersonen bzgl. der Externalisierung der Hörereignisse für die verschiedenen Systeme und Pfade. Es zeigen sich in keinem Vergleich zwischen den Systemen oder Pfaden signifikante Unterschiede ( $p < 0,05$ ) in den Externalisierungsindizes. Im Allgemeinen liegt die Externalisierung in einem Bereich größer oder knapp unter 0,8. Im Vergleich mit den Bewertungen aus anderen Hörversuchen entspricht dies relativ hohen Werten.

Bei den Pfaden weisen die Pfade I und IV die geringste Externalisierung auf. Pfad I ist der Pfad mit der geringsten Distanz zur Quelle. Hier kann sich die geringe Entfernung auf die Bewertung der Externalität (im Sinne von nah-am-Kopf) auswirken. Pfad IV ist der Pfad mit frontaler Ausrichtung auf die Quelle. Die Kopfausrichtungen der Testpersonen ist auf diesem Pfad überwiegend frontal oder rückseitig zur Schallquelle. Diese Richtungen sind bezogen auf die Externalisierung kritische Richtungen und weisen tendenziell geringere Externalisierung als seitliche Richtungen auf.

Bei der Betrachtung der Systeme liegt das System bei Anwendung des MTB-Ansatzes und bei Verwendung von drei Messpositionen (Abkürzung: 3M) hinter den Referenzmessungen (Abkürzung: 25M) und der Synthese aus einer Messposition unter Anwendung des ITDG-Ansatzes (Abkürzung: 1M).

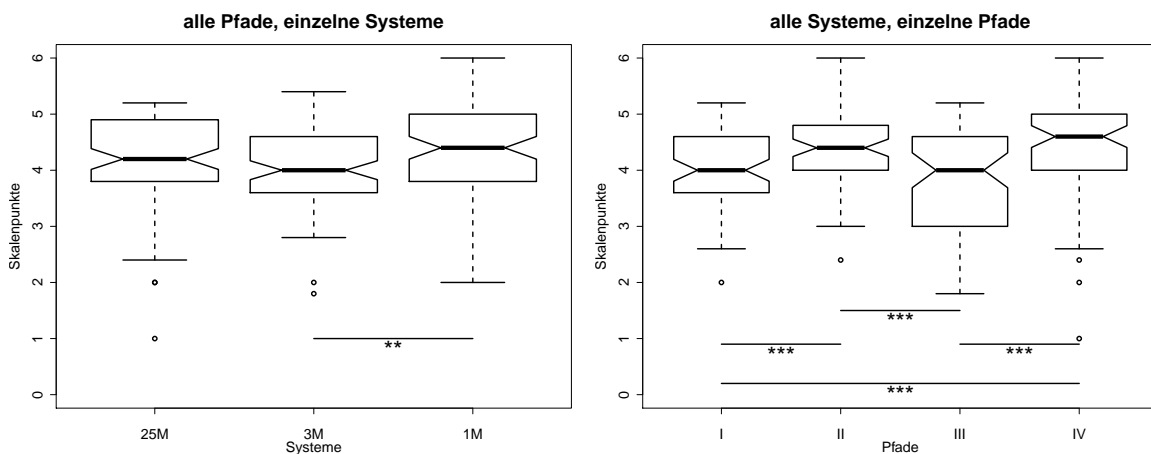


Abbildung 5.22: Gesamteindruck der Auralisierung als Boxplots; links: Zusammenfassung aller Pfade, rechts: Zusammenfassung aller Systeme; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ .

Abbildung 5.22 zeigt den im Hörversuch bewerteten Gesamteindruck der Audiowiedergabe. Die Skalenpunkte entsprechen gemäß der Skala für „Overall Impression“ in Abbildung 5.18 den Hauptintervallmarkierungen der Skala. Die Mediane der Bewertungen für die Systeme und die Pfade liegen gleich oder oberhalb der Bewertung gut („4 - good“). Zwischen den Systemen zeigt sich ein signifikanter ( $p < 0,05$ ) Unterschied zwischen der Methode 3M und 1M. Die Synthese aus der Verwendung von drei Messpositionen wird als leicht schlechter bewertet. Als Ursache wird die Interpolation vor allem der zeitlichen Verläufe der frühen Reflexionen gesehen. Es entstehen hierdurch u.a. Klangverfärbungen, die vor allem in dem durchgeführten Vergleich störend wirken. Für die Pfade I und III zeigen sich signifikant ( $p < 0,01$ ) schlechtere Bewertungen gegenüber den anderen Pfaden. Ein Grund kann in der ebenfalls geringeren Lokalisationstabilität (siehe nachstehende Abschnitte) gesehen werden. Für diese Pfade zeigen sich für dieses Qualitätsmerkmale Auffälligkeiten, welche zu einer Abwertung des Gesamteindrucks führen können.

Der obere Teil der Abbildung 5.23 zeigt die Lokalisationsfähigkeit der Hörereignisse als Boxplots. Die Skalenpunkte entsprechen gemäß der Skala für „Localization Ability“ in Abbildung 5.18 den Hauptintervallmarkierungen der Skala. Die Mediane der Bewertungen für die Systeme und die Pfade liegen oberhalb der Bewertung gut („4 - good“). Es zeigen sich keine signifikanten Unterschiede ( $p < 0,01$ ) zwischen den Systemen oder Pfaden. In allen Systemen wurde der Direktschallanteil der BRIRs mit den darin kodierten monauralen und binauralen Merkmalen auf gleiche Weise erstellt. Die hohe Übereinstimmung ist somit erwartbar. Lediglich das System 3M weist einen geringfügig höheren IQD als die anderen Systeme auf. Diese kann in der Verwendung von drei Messungen zur Interpolation der frühen Reflexionen begründet sein. Die in den frühen Reflexionen kodierten Richtungsinformationen werden durch die Interpolation verwischt.

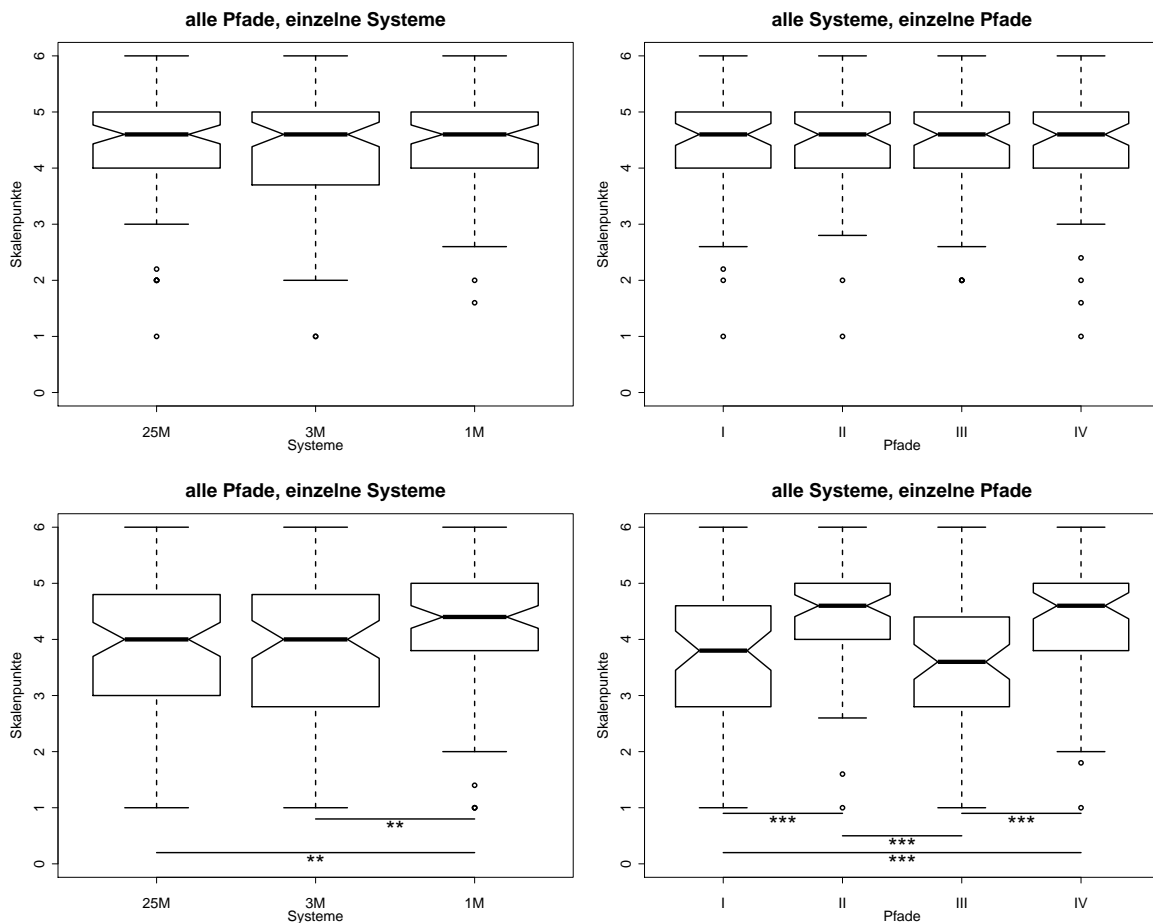


Abbildung 5.23: Lokalisationsfähigkeit (oben) und Lokalisationsstabilität (unten) der Hörereignisse als Boxplots; links: Zusammenfassung aller Pfade, rechts: Zusammenfassung aller Systeme; \*\*\* signifikanter Unterschied bei  $p < 0,01$ , \*\* signifikanter Unterschied bei  $p < 0,05$ .

Der untere Teil der Abbildung 5.23 zeigt die Stabilität in der Richtungsabbildung der Hörereignisse. Das Hörereignis soll an einem festen Punkt im augmentierten Raum wahrgenommen werden. Eine wahrgenommene Bewegung des Hörereignisses führt zu einer Verringerung der Lokalisationsstabilität. Die Skalenpunkte entsprechen den Markierungen der Hauptintervalle der Skala für Localization Stability in Abbildung 5.18. Die Mediane der Bewertungen für die Systeme und die Pfade liegen im Bereich von gut („4 - good“).

Es zeigt sich ein signifikanter ( $p < 0,05$ ) Unterschied in den Bewertungen zwischen dem System 1M und den anderen beiden Synthesemethoden. Interessanterweise wird die Synthese, basierend auf einer Messung bzgl. der Lokalisationsstabilität besser bewertet als die Referenzmessung. Die in der Messung abgebildeten frühen Reflexionen werden durch die ITDG-Methode in ihrem zeitlichen Auftreten gestaucht bzw. gestreckt. Es werden aber keine Reflexionen von anderen Messpositionen verwendet. Die zeitlichen Reflexionsmuster unterscheiden sich zwischen den verschiedenen Messpositionen, bspw. in der Methode 3M, teils deutlich voneinander. Es kommt somit in der Methode 1M zu einer gleichmäßigeren Änderung der Reflexionsmuster über die einzelnen Synthesepositionen.

Zwischen den Bewertungen der einzelnen Pfade zeigen sich signifikante ( $p < 0,01$ ) Unterschiede. Die Pfade direkt auf die Quelle zu (Pfad III) und der nahe der Quelle verlaufende Pfad (Pfad I) zeigen gegenüber den anderen beiden Pfaden deutliche geringere Bewertungen. Für den Pfad III wird basierend auf Aussagen der Probanden geschlussfolgert, dass es vor allem bei nahen Entfernungen zur Quelle, es zu Sprüngen in der Distanzwahrnehmung kommt. Weiterhin sind für diesen Pfad die relativ geringe Winkelauflösungen der Synthese der Kopfpose von  $5^\circ$  aufgefallen. Für den Pfad I führte die geringen räumliche Auflösung der Synthesepositionen in Verbindung zur relativ geringen Entfernung zur Quelle zu Sprüngen in der Richtungswahrnehmung. Dieser Effekt fällt für den Pfad II deutlich geringer aus.

**Experiment B** - Die Auswertung basiert auf den Bewertungen von 14 Testpersonen (9 weiblich, 5 männlich). Das Durchschnittsalter beträgt 26 Jahre (Standardabweichung 6 Jahre). Eine vorherige Hörtesterfahrung wird von 41% der Probanden berichtet. Erfahrung mit Binauralsynthese liegt bei 35% der Testpersonen vor. In Bezug auf VR und AR Systeme berichten 82% bzw. 30% der Teilnehmer eine vorherige Benutzung. Die Durchschnittsgröße der Teilnehmer beträgt 1,74 m mit einer Standardabweichung von 0,1 m. Der Durchschnitt der Testpersonen liegt somit in der des verwendeten Kunstkopfes.

Die Bewertungen der Testpersonen sind in Abbildung 5.24 als Boxplots dargestellt. Es werden die Bewertungen für die Testkondition „B09“ und „P09“ dargestellt. Für jede Kondition sind die unterschiedlichen Distanzen zwischen synthetisierter Position auf dem Pfad und der Position der Synthesemessung aufgetragen. Die ebenfalls in dem Hörtest vorkommende Referenz, basierend auf Messungen an allen zu synthetisierenden Positionen, ist als „REF“ dargestellt.

In einer vergleichenden Betrachtung der Mediane der beiden Testkonditionen zeigt sich eine generell schlechtere Bewertung für die Kondition „B09“ gegenüber „P09“. „B09“ wird als „geringfügig störend“ oder schlechter bewertet, während „P09“ als „wahrnehmbar aber nicht störend“ beschrieben wird. Die Referenz wird im Median als „nicht wahrnehmbar“ bewertet. Speziell für „B09“ ist eine zunehmend schlechter werdende Bewertung bei steigender Distanz zwischen Mess- und Syntheseposition zu erkennen.

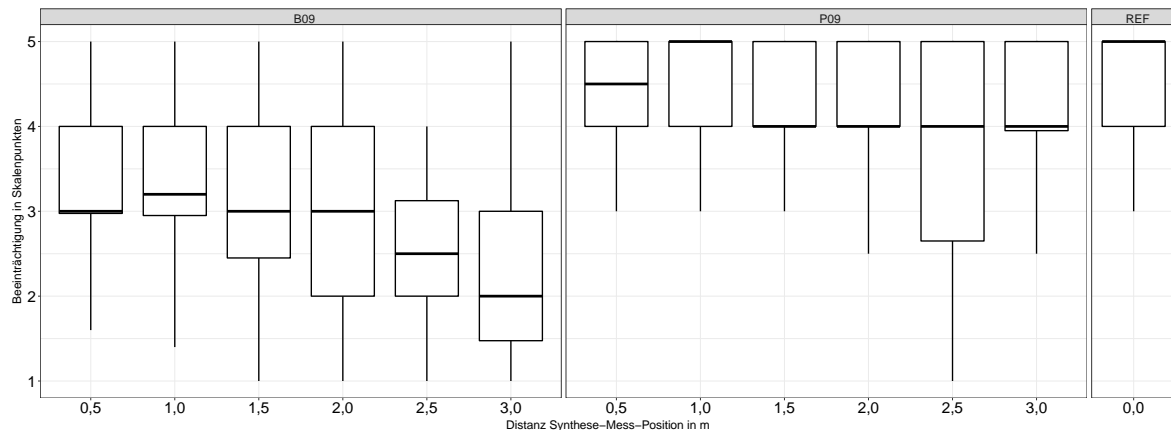


Abbildung 5.24: Bewertete Beeinträchtigung der Synthese als Boxplots; links: B09, mittig: P09, rechts: Referenz.

Die Messposition der BRIRs für die Synthese liegt in Kondition „B09“ direkt (0,25 m) vor der im Messaufbau vorhandenen Wand. Somit sind die ersten durch die Wand induzierten Reflexionen in der Messung und in den synthetisierten BRIRs enthalten. Ihre relative Intensität gegenüber der des Direktschalls und anderer Raumreflexionen wird zwar durch den ITDG-Ansatz verringert aber nicht explizit an die zu synthetisierende Position angepasst. Es tritt somit vor allem auch für die von der Messposition weiter weg gelegenen Synthesepositionen ein deutlicher Unterschied in der Distanzwahrnehmung auf. Die Hörereignisse der benachbarten Referenzmessungen werden als weiter weg wahrgenommen als die synthetisierten Positionen.

Für die Kondition „P09“ zeigt sich eine Verschlechterung der Bewertungen bei steigender Synthese-Mess-Position Distanz nicht. Es wird geschlussfolgert, dass für eine plausible Synthese in dem Anwendungsfall eines AR-Systems Messpositionen zur Synthese von neuen BRIRs sich im Raum und nicht nahe von Wänden befinden sollten. Mit diesen Positionen sind plausible Auralisierungen von Positionen im Raum aber auch nahe der Wände möglich. BRIRs von Positionen nahe der Wände führen zu deutlichen Effekten in der Distanzwahrnehmung.

Die ermittelten IQAs der Bewertungen überdecken Bereiche von eins bis über zwei Skaleneinheiten. Die Whisker der Boxplots überdecken teils den gesamten Skalenbereich. Eine Analyse der Verteilung der Bewertung zeigt eine große Streuung der Bewertungen. Die Ursache hierfür wird in einem unzureichenden Training der Testpersonen gemäß dem zu messenden Artefakt gesehen. Es war den Testpersonen teils nicht klar genug, auf welche Merkmale gehört

werden sollte. Auch waren einige Testpersonen nicht mit der verwendeten Skala vertraut und konnten diese schlecht in dem geforderten Fall anwenden. Es wird davon ausgegangen, dass die Testpersonen-Population nicht als Experten-Hörer im Bereich der vorgestellten Binauralsynthese angesehen werden kann. Deutlich mehr als die Hälfte der Testpersonen weisen keine Erfahrung mit Binauralsynthese und Hörversuchen auf.

Der hier präsentierte Hörtest ist lediglich eine Trendermittlung von Effekten und ein Beginn zu weiterführenden Untersuchungen für eine augmentierte auditorische Umgebung basierend auf räumlich dünn-besetzten BRIR Messungen. In zukünftigen Untersuchungen sollte eine Testprozedur mit einem ausgeprägterem Training verwendet werden. Als Untersuchungsgegenstand in den weiterführenden Arbeiten kann unter anderem der Übergang zwischen Bereichen, in denen neue BRIRs durch unterschiedliche Methoden synthetisiert wurden, dienen. Zusätzlich können weitere Messpositionen für die Synthese an anderen Stellen in einem akustisch höher komplexeren Raum dienen. Es sei somit auf zukünftige Forschung verwiesen.

### **5.3.6 Zusammenfassung**

Die Ergebnisse aus Experiment VIII zur Untersuchung einer ITDG-basierten (siehe Kapitel 5.3.3) und die Ergebnisse aus Experiment IX bzgl. einer interpolationsbasierten Anpassung akustischer Distanzmerkmale einer BRIR (siehe Kapitel 5.3.4) verdeutlichen, dass es mit den dort untersuchten Methoden möglich ist, den Distanzeindruck beim Hören zielgerichtet zu verändern. Hierfür wurden Methoden der Interpolation zwischen zwei (oder mehreren) und der Anpassung einer BRIR untersucht.

Die Ansätze ermöglichen in einer sich anschließenden Anwendung die Anpassung des Qualitätsmerkmals an die Kontextbedingungen der Verwendung des Audiosystems. Beispielsweise kann die komplette zu auralisierende Szene in ihrer räumlichen Dimensionierung skaliert werden. Hierdurch können perzeptive Kontexteinflüsse, wie bspw. Passung der Audioszene mit den visuellen Dimensionen des Abhörtraumes oder die Erwartungshaltung an die Distanzwahrnehmung der abzubildenden Objekte, adressiert und ihre Effekte minimiert werden.

Die Ergebnisse aus Experiment X (siehe Kapitel 5.3.5) verdeutlichen, dass es durch die Anwendung verschiedener Methoden zur Synthese von BRIRs aus räumlich dünn-besetzten Messungen möglich ist, eine augmentierte auditorische Hörumgebung zu erstellen. Die Bewertungen der verschiedenen untersuchten Qualitätsmerkmale zeigen gute bis teils sehr gute Werte bzgl. einer räumlichen und plausiblen Hörwahrnehmung. Die verwendeten Synthesemethoden ermöglichen eine deutliche Reduktion an akustischen Messungen zur Erstellung der für die Binauralsynthese benötigten BRIRs. Es konnte ein geeigneter Weg zur Erstellung einer AR-Anwendung bei Minimierung des Raumdivergenzeffektes aufgezeigt werden. Eine Validierung der Methoden hinsichtlich Verwendbarkeit in einer prototypischen Anwendung soll hier als weiterführende Untersuchung genannt sein.



## 5.4 Zusammenfassung adaptive Binauralsynthese

Die Untersuchungen in diesem Kapitel legen nahe, dass es durch Anpassung von raumakustischen Parametern in gemessenen BRIRs möglich ist, eine plausible räumliche Hörwahrnehmung durch Anwendung der Binauralsynthese zu realisieren. Dieses konnte für die Qualitätsmerkmale Raumübereinstimmung, Distanzwahrnehmung, Externalität, Lokalisation und Gesamtqualität gezeigt werden. Die untersuchten akustischen Parameter sind energie- und zeit-basierte Parameter von BRIRs.

Die Anpassung des DRR als energiebasierter Parameter zeigt, dass es den Hörern in hohem Maße möglich ist, die Halligkeit der Synthese an die des Abhörtraumes anzupassen. Das Kriterium ist hierbei das Erreichen hoher Raumkonvergenz mit der Synthese. Die DRR-Einstellung erfolgt hierbei nicht durch einen direkten Vergleich der Synthese mit bspw. Lautsprechern im realen Raum, sondern durch Hören der Synthese und Vergleich mit der Erfahrung durch vorheriges Hören des realen Raumes. Die gemessenen Mediane der eingestellten DRR Pegel der Synthese weichen lediglich um wenige dB von den gemessenen DRR-Werten des Abhörtraumes ab. Die IQAs liegen im Bereich der JNDs der DRR-Wahrnehmung für reale und virtuelle Hörumgebungen. Dennoch spiegelt sich diese hohe Sensitivität der Hörer nicht in einer signifikanten Verringerung des Raumdivergenzeffektes wider. Es wird vermutet, dass gemäß der Annahmen zum Präzedenzeffekt lediglich energiebasierte Parameter nicht ausreichen. Die Abbildung eines akustischen Raumes in der Hörwahrnehmung basiert anscheinend vor allem auf der zeitlichen Struktur der aus der Raumakustik abgeleiteten Mustern.

In Bezug darauf sind verschiedene weitere Anpassungsmethoden zur Synthese von neuen BRIRs aus gemessenen BRIRs entwickelt und untersucht worden. Das Ziel ist hier die Anpassung der Distanzwahrnehmung zur Abbildung von neuen Sender-Empfänger-Positionen. Die Verwendung liegt in der Skalierung der zu synthetisierenden Szene und in der Realisierung einer Hörertranslation in einer augmentierten Hörumgebung.

Die zeitbasierte Anpassungsmethode auf Basis einer ITDG-Anpassung ermöglicht eine Änderung in der Wahrnehmung von Distanz. Gegenüber einer Auralisierung basierend auf gemessenen BRIRs, zeigt sich eine leichte Unterschätzung der abzubildenden Distanz. Für hallige Räume wird ein größerer IQA gegenüber einem weniger halligen Raum festgestellt. Die untersuchten Methoden, basierend auf Interpolation zwischen zwei gemessenen BRIRs zeigt ebenfalls die Möglichkeit einer gesteuerten Distanzwahrnehmung auf. In den Experimenten zeigt sich hierfür eine Überschätzung der abzubildenden Distanz. Es soll hier erwähnt werden, dass in diesen Experimenten eine explizite Referenzposition (i.d.R. Lautsprecheratrappe und Binauralsynthese dieser Position) zur relativen Bewertung der Distanz vorhanden ist. Ist diese an der maximal möglichen abzubildenden Distanz platziert, kommt es zur Distanzüberschätzung der näher gelegenen Hörereignisse. Ist diese bspw. in der Mitte der abzubildenden Distanzen positioniert, kommt es zur Unterschätzung der Distanz für weiter weg gelegene Hörereignisse. Dies deutet auf eine audio-visuelle Interaktion hin, welche sich auf eine Fusion der

Hörereignisorte mit dem visuellen Ereignisort zurückführen lässt (siehe Ventriloquismus-Effekt für Distanzwahrnehmung).

Die Methoden zur Synthese von neuen BRIRs zeigen bzgl. des Merkmals Externalität keine signifikanten Unterschiede zu gemessenen BRIRs. Die Untersuchung der Externalisierung der Hörereignisse von verschiedenen Bewegungspfaden des Hörers relativ zu der abzubildenden Schallquelle zeigt eine, wenn gleich nicht signifikante, geringe Abhängigkeit vom jeweiligen Pfad. Der Gesamteindruck der Hörer, die Beurteilung der Lokalisationsfähigkeit und Lokalisationsstabilität liegt für alle Synthesemethoden vergleichbar zu den Referenzmessungen. Bzgl. der Stabilität der Lokalisierung der Hörereignisse zeigen sich signifikante Unterschiede zwischen den untersuchten Bewegungspfaden. Die Ursache hierfür wird vor allem in der teils unzureichenden räumlichen Dichte der zu auralisierenden Positionen im Raum relativ zur Schallquellen-Hörer-Entfernung gesehen. Eine Erhöhung der Auflösung verringert diesen Effekt deutlich. Eine Weiterentwicklung des verwendeten Audiosystems beinhaltet eine je nach Distanz zwischen Hörer- und Schallquellposition skalierbare räumliche Auflösung der BRIRs.

Es wird empfohlen, dass die für die Synthese benötigte/n Messposition/en sich nicht nahe der Wände des Raumes befinden. Werden diese Positionen verwendet, lässt sich mit den vorgestellten Methoden keine ausreichend plausibel wirkende Auralisierung ermöglichen. Die kurz nach dem Direktschall von den Wänden kommenden Reflexionen können nicht adäquat angepasst werden, um weiter im Raum liegende Positionen zu synthetisieren. Anders herum scheinen Messungen relativ weit weg ( $>1,5$  m) von den Wänden durchaus geeignet, um Positionen nahe der Wände ( $<1,5$  m) zu auralisieren.

## Kapitel 6

### Zusammenfassung und Ausblick

Wissen wird nicht passiv aufgenommen, weder durch die Sinnesorgane noch durch Kommunikation. Wissen wird vom denkenden Subjekt aktiv aufgebaut. Die Funktion der Kognition ist adaptiver Art, und zwar im biologischen Sinn des Wortes, und zielt auf Passung oder Viabilität. Kognition dient der Organisation der Erfahrungswelt des Subjekts und nicht der Erkenntnis einer objektiven, ontologischen Realität.

---

*Ernst von Glasersfeld*  
[99]

## 6.1 Rückschau

Die vorliegende Arbeit stellt zahlreiche Hörversuche zur Untersuchung der räumlichen auditiven Wahrnehmung bei Verwendung einer Binauralsynthese über Kopfhörer vor. Es wird der Einfluss verschiedener kontextabhängiger Qualitätsparameter auf die Wahrnehmung untersucht. Es wird versucht, diese Parameter getrennt voneinander zu untersuchen. Dieser Ansatz soll ihren Einfluss auf eine plausible räumliche Hörwahrnehmung und die Beziehungen untereinander erkennen lassen. Die untersuchten Qualitätsmerkmale sind:

- **Lokalisation:** Bewertung der wahrgenommenen Richtung des auditorischen Ereignisses. Dies beinhaltet ggf. die Angabe der Richtung oder der Stabilität der wahrgenommenen Richtung.
- **Distanz:** Bewertung des wahrgenommenen Distanzunterschiedes zwischen einem Testsignal und einer Referenz.
- **Externalisierung:** Bewertung einer wahrgenommenen Entfernung eines Hörereignisses im Vergleich zur intrinsischen Grenze zwischen Im-Kopf und Außerhalb-Kopf (external).
- **Räumliche Präsenz:** Relative Bewertung der Wahrnehmung, sich in einer räumlichen Szene im Vergleich zu einer anderen räumlichen Szene zu befinden.
- **Nachhallumhüllung:** Bewertung der Wahrnehmung, sich im Vergleich zwischen verschiedenen Räumen und Szenen vom Nachhall umhüllt zu sein.
- **Gesamteindruck:** Bewertung der Wahrnehmung, einen plausiblen räumlichen Höreindruck in einer audio-augmentierte Wiedergabesituation zu haben.

Die kontextabhängigen Qualitätsparameter sind:

- **Audio-visuelle Divergenz:** zum Schallereignis in ihrer Position variierende visuelle Merkmale
- **Schalleinfallrichtung:** verschiedene Schallereignisrichtungen von Einzelereignissen
- **Raumdivergenz:** konvergierende und divergierende Kombinationen von Abhörraum und synthetisiertem Raum
- Sichtbarkeit des Abhörtraumes
- **Personalisierung:** individuelle Anpassung der binauralen Wiedergabe
- **Szenenkomplexität:** Anzahl und räumliche Anordnung der Audioobjekte in der dargebotenen Szenen
- **Raumkomplexität:** vergleichende Gegenüberstellung mehrerer Räume mit verschiedener Raumakustik
- **Head-Tracking:** Binauralsynthesystem mit Kopfposennachverfolgung bei Kopfbewegung

Die Details zu den Effekten und die Beantwortung der Forschungsfragen sind den Zusammenfassungen und den einzelnen Experimenten in den jeweiligen Kapiteln zu entnehmen. Abbildung 6.1 stellt eine zusammenfassende Einschätzung der Einflüsse der kontextabhän-

gigen Qualitätsparameter auf die untersuchten Qualitätsmerkmale dar. Es wird gemäß den Kategorien „Gesichert und in der Literatur von anderen bestätigt“, „Gesichert aber mit Uneinheitlichkeiten in der Literatur“, „Offen mit weitere Forschung nötig“ eingeteilt. Nicht belegte Kreuzungspunkte wurden in dieser Arbeit nicht erfasst.

Für den Zusammenhang zwischen Szenenkomplexität und den Merkmalen Externalisierung und Lokalisierung bleibt offen, ob räumlich verteilte Audioobjekte einen für die Qualitätsmerkmale förderlichen Einfluss haben, wenn andere Kontextparameter variiert werden. Der Einfluss der Personalisierung auf die Externalisierung bleibt dahingehend offen, dass in Experimenten Korrelationen zwischen Lokalisierungsfehlern (Quadrantenfehler) und Externalisierung der Hörereignisse für unterschiedliche Personalisierungen gezeigt oder eben auch nicht gezeigt werden konnten (siehe hierzu Abschnitte 4.2.1.4 und 4.2.2).

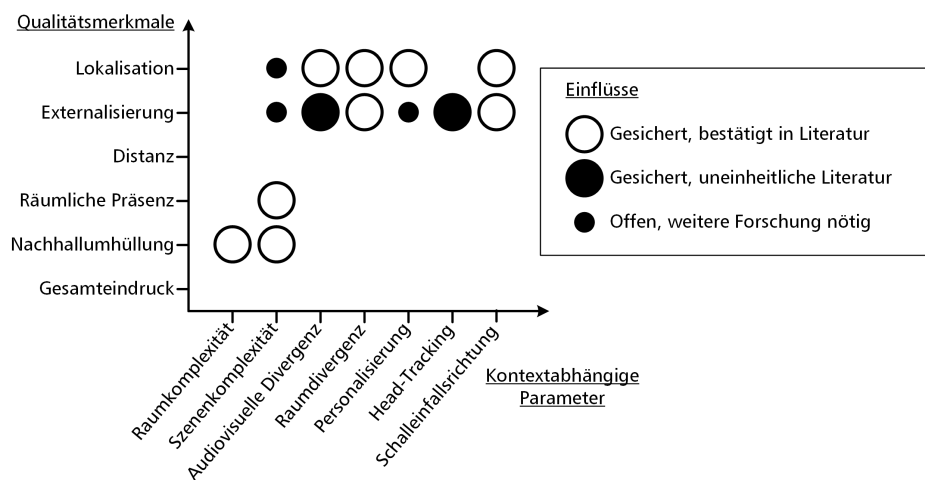


Abbildung 6.1: Einschätzung der Einflüsse der kontextabhängigen Qualitätsparameter auf die Qualitätsmerkmale.

Für den Zusammenhang zwischen der Sichtbarkeit bzw. Nicht-Sichtbarkeit des Abhörtraumes (audiovisuelle Divergenz) und der Externalisierung zeigen sich Uneinheitlichkeiten mit der bisherigen Literatur. Es wird in dieser Arbeit gezeigt, dass zusätzliche visuelle Merkmale den Raumdivergenzeffekt nicht zwingend auflösen. Weiterhin wird in dieser Arbeit gezeigt, dass die Verwendung eines Head-Trackings den Raumdivergenzeffekt ebenfalls nicht zwingend auflösen kann.

Es soll hier festgehalten werden, dass die Einschätzung einer, durch ein Binauralsynthesystem erreichbaren, perzeptiven räumlichen Qualität nicht durch einfache Implikationen basierend auf den technischen Parametern und Komponenten des Systems beruhen kann. Es ergibt sich zum einen von selbst, dass es eine geeignete technische Realisierung als notwendige Bedingung braucht, wengleich zum anderen der Kontext der Verwendung als hinreichende Bedingung im Gesamtsystem Beachtung finden muss. Der Kontext aber auch die technische Realisierung wirken stets gemeinsam. Das kognitive System nutzt alle ihm zur Verfügung stehenden und

geeigneten sensorischen Reize sowie die Erfahrung und damit verbundene Erwartungshaltung zur Konstruktion und fortlaufenden Anpassung eines Hörereignisses.

Es zeigt sich durchaus die unbequeme Situation, dass bei Fragen nach der Eignung eines Systems oder nach notwendigen Systemkomponenten im Prinzip niemals eine einfache Antwort gegeben werden kann. Die Passung ist immer an den Verwendungskontext und das Individuum gebunden, welche mit Sicherheit deutlich komplexer sind, als diese in dieser Arbeit untersuchbar wären.

## 6.2 Neue Fragen

Während der verschiedenen Untersuchungen sind einige Effekte auffällig geworden, die hier als neue Fragen formuliert werden sollen. Es ist in keinster Weise auszuschließen, dass diese Fragen entweder bereits beantwortet sind oder deren Existenz durch anderweitige Erkenntnisse nicht gerechtfertigt werden kann.

**Schalldruckpegel-Lautheitsdivergenz-Effekt** - Es zeigt sich, dass je nach verwendeter Testmethode ein Schalldruckpegel-Lautheitsdivergenz-Effekt (SLD-Effekt) in der Verwendung eines ansonst identischen Binauralsynthesystems festgestellt werden kann oder auch nicht. Es bleibt verwunderlich, warum dieser Effekt bei Darbietung einer Referenz über Binauralsynthese und der Teststimuli über reale Schallquellen (oder anders herum) verschwindet, während er bei Messung im gleichen technischen System auftritt. Eine Sammlung an Fragen soll sein: Auf welche perzeptiven Qualitätsmerkmale lässt sich der SLD-Effekt zurückführen? Was ist die interne Referenz für einen Lautheitsvergleich und wie ändert sich diese bei verändertem technischen System? Wie wirkt sich ein Konflikt in einem Lautheitsvergleich auf die Validität der Binauralsynthese aus?

**Fitness von sensorischen Merkmalen** - In der Gegenüberstellung des Auftretens eines Quadrantenfehlers und der Externalisierung zeigt sich oft eine starke negative Korrelation; je mehr Lokalisierungsfehler, umso geringer die Externalisierung der Hörereignisse. Dieser Zusammenhang konnte aber für binaurale Signale mit schwachen spektralen Merkmalen (Kunstkopfes vs. Kugelflächenmikrofon) nicht nachgewiesen werden. Weiterhin zeigt sich in Untersuchungen bei Verwendung von Kunstkopf-BRIRs und Hinter-dem-Ohr-BRIRs ein Wechsel der kritischen Richtungen für Quadrantenfehler in der Lokalisation von Hörereignissen [9]. Es stellt sich die Frage: Welche sensorischen Merkmale werden unter welchen Umständen in welcher Wichtung für die Herausbildung eines Hörereignisses und in Bezug auf Externalisierung und Lokalisation verwendet?

Eine weitere Auffälligkeit zeigte sich für visuelle Merkmale. Ist während des räumlichen Hörens der Abhörraum sichtbar und ggf. weitere visuelle Merkmale vorhanden, steigt die Externalisierung der Hörereignisse für rückseitige Audioobjektrichtung gegenüber frontalen Richtungen verstärkt an. Eine Betrachtung der rückseitigen Richtungen durch den Hörer sei hier ausgeschlossen. Es stellt sich die Frage: Verändern sich für rückseitige Hörereignisse die Wichtung und Relevanz der aus den sensorischen Informationen abgeleiteten Merkmale, wenn kongruente visuelle Reize vorhanden sind?

**Externalisierung und Kopfbewegung** - In der Untersuchung zum Einfluss von Head-Tracking auf die Externalisierung zeigt sich, dass die Head-Tracking einen signifikanten positiven Einfluss auf die Externalisierung hat, wenn diskrete Schallquellen (hier Lautsprecher) synthetisiert werden. Dieser positive Effekt ist bei zwischen synthetisierten Lautsprechern stereogepannten

Audiosignalen nicht feststellbar. Die Externalisierung ist unabhängig vom Tracking ähnlich hoch wie für diskrete Lautsprecher. Frage: Werden für die Wahrnehmung von stereogepannten Signalen dieselben auditorischen Merkmale in gleicher Wichtung verwendet, wie für diskrete Lautsprecher?

**Strukturmerkmale von BRIRs** - Es wurde gezeigt, dass eine Binauralsynthese einer Vielzahl von Positionen im Raum durch Anpassung einer einzelnen gemessenen BRIR in diesem Raum als plausibel wahrgenommen wird. Es wird gemäß dem Clifton-Effektes und des Raumdivergenzeffektes geschlussfolgert, dass ein Mustervergleich von wahrscheinlich überwiegend der zeitlichen Struktur der Raumreflexionen zwischen Gehörtem (Synthese) und Erwartetem (Kurzzeitspeicher im Gehirn; Formulierung gemäß Plenge [168]) stattfindet. Es bleibt offen: Welche Strukturähnlichkeiten in den Reflexionsmustern werden ausgewertet? Auf Grundlage welcher Signalmerkmale werden diese Muster erkannt?



### 6.3 Externalisierung als Funktion einzelner kontextabhängiger Qualitätsparameter

Die in den Hörexperimenten untersuchten Einflüsse auf die Externalisierung von Hörereignissen sollen in einem Modell zusammengeführt werden. Das Modell soll die Externalisierung als lineare Kombination nicht-linearer Einzelkomponenten erfassen. Die einzelnen Kontextparameter werden durch nicht-lineare psychometrische Funktionen angenähert. Diese Funktionen basieren auf Schätzungen und Interpretation der in den einzelnen Experimenten gemessenen Einflüsse. Diese Ergebnisse dienen als Stützstellen der einzelnen psychometrischen Funktionen. Die gesamte psychometrische Funktion der Externalität ergibt sich als Produkt der einzelnen Funktionen.

#### 6.3.1 Motivation

Die Motivation dieses Ansatzes ergibt sich aus Modellen von sensorischen Systemen aus den Neurowissenschaften. Das Ziel ist hierbei die Charakterisierung der funktionalen Zusammenhänge zwischen einwirkendem Stimuli und neuronalem Antwortverhalten des sensorischen Systems. Die Antwortverhalten der gesamten Neuronenpopulation wird als beliebig komplex und nicht-linear angesehen. Die Annahme ist weiterhin, dass das Antwortverhalten eines Neurons aus einer linearen Kombination (Summe) einer Vielzahl von Signalgebern/Eingangssignalen (bspw. synaptische Reaktionspotentiale) und einer spezifischen Verhaltenscharakteristik des Neurons besteht [75]. Der augenblickliche Zusammenhang zwischen Eingangssignal und Ausgangssignal entspricht der Charakteristik des Neurons und kann linear oder nichtlinear sein. Die Kodierung der Eingangssignale wird als Kombination einer ersten linearen Filterung des Reizes und einer nichtlinearen Funktion zur Erzeugung des Ausgangssignals des Neurons modelliert [80].

Das von Chichilinsky [75] angenommene neuronale Antwortverhaltens ist die Summe des Produktes aus Feuerzeitpunkten (engl.: spike count) und einer Wahrscheinlichkeit, dass ein Neuron auf einen Stimulus aktiv wird. Die Summe wird über die Anzahl der Feuerzeitpunkte gebildet. Nachstehende Formel beschreibt diesen Zusammenhang:

$$R(s) = \sum_f f P(f|s) \quad (6.1)$$

Mit  $f$  als die Spike-Anzahl,  $s$  als Stimulusvektor aus einem  $k$ -dimensionalen Stimulusraum (bspw. Repräsentation der tonotopen Struktur im auditorischen System) und  $R(s)$  als Antwortverhalten (bspw. Spike-Rate des Neurons).

Das Antwortverhalten  $R(s)$  soll nun abgeschätzt werden. Hierbei gilt die Festlegung, dass  $R(s)$  eine statische nichtlineare Funktion einer linearen Funktion des Stimulusvektors ist [75].

$$R(s) = (w \cdot s)N \quad (6.2)$$

$w$  entspricht einer Wichtung der in  $s$  befindlichen Stimuliintensitäten.  $w \cdot s$  ist damit das Eingangssignal. Die Werte für  $N$  und  $w$  können aus Messungen der neuronalen Aktivität abgeleitet werden [75].  $N$  entspricht einer beliebigen nichtlinearen Funktion und repräsentiert bspw. ein Schwellwertverhalten oder Sättigungsverhalten des Neurons.

Dieser Lineare-Nichtlineare Modellansatz wird als Gedankengerüst auf die untersuchten kontextabhängigen Einflüsse auf die Externalisierung von Hörereignissen übertragen.

### 6.3.2 Kontextparameter

Die kontextabhängigen Qualitätsparameter Raumdivergenz, Personalisierung, visueller Einfluss und Schalleinfallrichtung werden als unabhängig voneinander angesehen. Dies ergibt sich direkt aus der „Untersuchung zur Wahrnehmung von Externalität und Hörereignisrichtung“ (Kapitel 4.2.1.4). Der Einfluss der einzelnen kontextabhängigen Qualitätsparameter wird als nicht-linear angenommen. Gemäß der Modellannahme in Formel 6.3.1 wird für jeden Kontextparameter eine gewichtete logistische Verteilungsfunktion in Form einer Sigmoid-Funktion angenommen und definiert als:

$$E_{Parameter} = \Delta \cdot \frac{1}{1 + e^{-\frac{x-\alpha}{\beta}}} \quad (6.3)$$

mit den Funktionsparametern  $\alpha$  für die zentrale Lage und  $\beta$  für die Breite der Funktion.

Die Funktionsparameter ergeben sich aus der in den Experimenten ermittelten Einflüsse auf die Externalisierung als Verteilung über die Magnitude des jeweiligen Kontextparameters. Der Wichtungsfaktor  $\Delta$  ergibt sich aus den in den Experimenten ermittelten Effektstärken bzw. der Änderung in der wahrgenommenen Externalisierung der Hörereignisse für die untersuchten Testkonditionen.

Auf der x-Achse ist die normierte Magnitude des jeweiligen kontextabhängigen Qualitätsparameters aufgetragen. Auf der y-Achse ist der Externalisierungsindex abgetragen. Eine Berücksichtigung eines Rate-Fehlers und eines Lapsus-Fehlers findet in erster Linie nicht statt. Diese Fehler werden nicht explizit in den einzelnen Experimenten untersucht, können aber in weiterführenden Untersuchungen erhoben werden. Weiterhin sind die hier aufgestellten Funktionen eine durch Messwerte getriebene qualitative Beschreibung der zugrundeliegenden psychometrischen Funktionen. Weitere Einflüsse, wie bspw. die genannten Fehler oder eine andere Verteilungsfunktion können jederzeit integriert werden.

Die kumulierte Funktion der Externalisierung ergibt sich aus der Addition der Einflüsse der einzelnen kontextabhängigen Qualitätsparameter zu:

$$E_{gesamt} = E_{Raumdivergenz} + E_{Person} + E_{visuell} + E_{Azimut} \quad (6.4)$$

Es wird eine Addition angenommen, da in den Experimenten gezeigt werden konnte, dass die untersuchten Kontextparameter unabhängig voneinander wirken (siehe Abschnitt 4.2.2). Nachstehend werden die geschätzten Verteilungen der einzelnen kontextabhängigen Qualitätsparameter vorgestellt und beschrieben.

**Raumdivergenz** - Der Einfluss der Raumdivergenz auf die Externalisierung wird gemäß vorherigen Ausführungen mit folgenden Funktionsparametern geschätzt:  $\alpha = 1/3$ ;  $\beta = 0,5$ ;  $\Delta = 0,4$

Die Magnitude lässt sich als Verhältnis zwischen synthetisiertem Raum und Abhörraum auffassen. Bei einer konvergenten Raumkondition ergibt sich die Magnitude zu „1“. Dies ist bspw. bei den Raumkonditionen „SR in SR“ oder „HL in HL“ aus den Hörexperimenten der Fall. Bei totaler Divergenz ergibt sich die Magnitude zu „0“. Dies ist bspw. für die Raumkonditionen „FF in SR“ oder „SR in FF“ (FF=Freifeld) der Fall. Die Raumkonditionen „HL in SR“ und „SR in HL“ ordnen sich zwischen diesen beiden Grenzen ein. Als eine raumakustisch-motivierte Magnitude bietet sich das Verhältnis des DRRs von Synthese und Abhörraum an. Aus den Experimenten ergeben sich somit drei Stützstellen zur Schätzung der Funktion.

In den Experimenten zur Raumdivergenz hat sich gezeigt, dass eine Konvergenz zwischen Abhörraum und synthetisiertem Raum zu einem Anstieg der Externalität von ca. 0,4 Indexpunkten gegenüber einer komplett divergenten Raumkondition führt. Das Maximum der angenäherten psychometrischen Funktion des Raumdivergenzeffektes liegt somit bei  $\Delta = 0,4$ .

**Personalisierung** - Der Parameter Personalisierung erstreckt sich von nicht personalisiert bis vollständig personalisiert. Eine individualisierte Binauralsynthese führt zu einem mittleren Anstieg der Externalität von 0,17 Indexpunkten. Der Übergang von nicht personalisiert zu vollständig personalisiert wird als kontinuierlich angenommen und der 50%-Punkt der Funktion auf die Mitte der Magnitudenskala gesetzt. Die Parameter der psychometrischen Funktion sind damit:  $\alpha = 0$ ;  $\beta = 1$ ;  $\Delta = 0,17$ .

**Visuelle Merkmale** - Ein Vorhandensein von den Audioobjekten dazugehörigen visuellen Objekten führt zu einem mittleren Anstieg des Externalisierungsindex von 0,16 Punkten gegenüber keinen visuellen Merkmalen. Der Übergang wird in diesem Fall als weitestgehend sprunghaft angenommen. Dies entspricht dem Fall, dass entweder eine perzeptive Fusion zwischen visuellen und auditorischen Objekt vorliegt oder nicht. Es soll hier aber darauf hingewiesen werden, dass es ähnlich dem Venriloquismuseffekt für die Lokalisation ein intersensorischen Bias auch für die Externalisierung vorliegen kann. Eine mehr oder weniger starke Fusion zwischen visuellem und auditorischen Objekt kann zu einer Beeinflussung der Externalisierung führen. Weiterführende Untersuchungen hierzu sind sinnhaft. Die Parameter der psychometrischen Funktion sind festgelegt mit:  $\alpha = 0$ ;  $\beta = 0,005$ ;  $\Delta = 0,16$ .

**Schalleinfallrichtung** - Der Einfluss des Azimuts der synthetisierten Schallquelle erstreckt sich auf der normierten Magnitudenachse von einer frontalen ( $0^\circ$ ) bzw. direkt rückwärtigen ( $180^\circ$ ) Schallquellenrichtung bis zu einer vollständig seitlichen Richtung ( $90^\circ$  bzw.  $270^\circ$ ). Dieser Annahme liegen die Ergebnisse der Untersuchungen zugrunde, dass frontale bzw. rückwärtige Schallquellen geringer externalisiert werden als seitliche Schallquellen. Der 50% Punkt der Funktion wird mehr in Richtung frontaler bzw. rückwärtiger Richtungen verschoben, da bereits bei geringen seitlichen Auslenkungen von bspw.  $30^\circ$  maximale Externalisierungsindizes erreicht werden können. Der maximale Einfluss der horizontalen Schalleinfallrichtung wird mit ca. 0,3 Indexpunkten angenommen. Der Einfluss von in der Elevation ausgelenkten Schalleinfallrichtungen ist hier nicht erfasst. Deren Einfluss sollte in weiterführenden Untersuchungen ermittelt werden. Die Parameter der psychometrischen Funktion ergeben sich zu:  $\alpha = 1/4$ ;  $\beta = 0,5$ ;  $\Delta = 0,3$ .

**Funktion der Externalisierung** - Eine Aufsummierung über alle psychometrischen Funktionen ergibt die geschätzte gesamte psychometrische Funktion der Externalisierung. Abbildung 6.2 verdeutlicht schematisch die aus den Hörexperimenten ermittelten Externalisierungsfunktionen der einzelnen kontextabhängigen Qualitätsparameter.

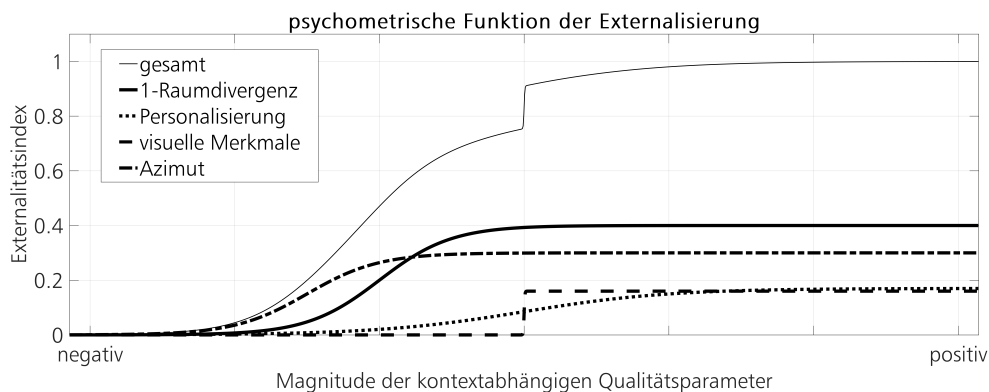


Abbildung 6.2: Schätzung der psychometrischen Funktion der Externalisierung basierend auf einzelnen kontextabhängigen Qualitätsparametern.

Die vorgestellte Idee einer mathematischen Beschreibung der Externalisierung muss hinsichtlich ihrer Validität überprüft werden. Hierzu sind weiterführende Untersuchungen durchzuführen. Der Schwerpunkt sollte hierbei auf der Bestimmung weiterer Stützstellen der dem Modell zugrundeliegenden psychometrischen Funktionen liegen. Es sollte weiterhin überprüft werden, welche Start- und Endpunkte die Ausprägungen der einzelnen kontextabhängigen Qualitätsparameter besitzen können und wie eine Abstufung der Parameter realisiert werden kann. Weiterhin gilt es zu untersuchen, welche weiteren Qualitätsparameter für die Wahrnehmung eines externalisierten Hörereignisses relevant sind. Beispielfhaft sei hier auf den Einfluss der Kopfbewegung und der translatorischen Bewegung des Hörers in der Audioszene verwiesen. Hieraus ergeben sich auch Fragestellungen zum Einfluss der bewussten und unbewussten

Exploration einer Audioszene durch den Hörer und der Vorgänge der Adaptation des Gehörs während des Hörens einer Audioszene.

## 6.4 Ausblicke

Der in Kapitel 1.4 und in Abbildung 1.8 vorgestellte Qualitätsbildungsprozess soll hinsichtlich möglicher Anwendungen abschließend genauer betrachtet werden. Ein Bestandteil des Prozesses ist die systemische Betrachtung von perzeptiven Vorgängen im Menschen und technischen Vorgängen im multimedialen System. Hierbei sind Interaktionen zwischen den Qualitätselementen und den Kontextparametern aufgezeigt (siehe Abbildung 6.3).

Die Systemeigenschaften wirken sich auf den Kontext der Verwendung des technischen Systems aus. In der anderen Richtung wirken sich Kontextparameter auf die technischen Qualitätselemente hinsichtlich Abbildbarkeit aus. An dieser Interaktion besteht die Möglichkeit einer adaptiven Steuerung des Systems gemäß festgelegter erreichbarer wahrgenommener Qualität (QoE). Sollte es dem technischen System möglich sein, die Zusammenhänge und Wirkung seiner multimedialen Synthese auf die Wahrnehmung von Qualität abzuschätzen, kann es eigenständig die Parameter seiner Qualitätselemente aber auch die Kontextparameter anpassen. Hieraus können sich interessante und neuartige technische Systeme ergeben.

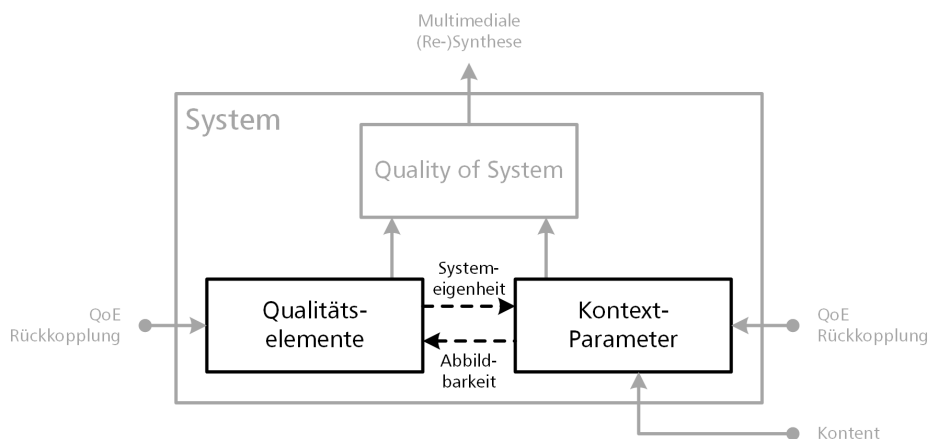


Abbildung 6.3: Interaktion der technischen Qualitätselemente und der Kontextparameter.

**Ein sich selbst adaptierendes technisches System** - So wie in kognitiven Systemen Prozesse der Assimilation und Akkommodation mit dem Ziel der Erreichung eines Gleichgewichtszustandes wirken [166], lassen sich vergleichbare Prozesse in technischen Systemen formulieren. In Analogie zum perzeptiven Qualitätsbildungsprozess [122], entsprechen die technischen Qualitätselemente den Parametern der realisierbaren technischen Qualität. Die Kontextparameter beschreiben die benötigte technische Qualität und damit benötigten Qualitätselementen. Die Systemeigenschaften und Abbildbarkeiten sind die steuernden (Anpassungs-)Prozesse des technischen Systems. Der Wirkungssinn ist die gezielte Erreichung von Wahrnehmungsereignissen mit einer beschreibbaren QoE beim Nutzer bei Minimierung der kontextabhängigen Effekte. Dieses fortlaufende Anpassen müsste von im System vorhandenen (Wahrnehmungs-)Modellen gesteuert werden. Die Modelle sollten selber, gemäß der Wahrnehmung beim Nutzer, durch

Rückkopplungen adaptiv sein. Kurzum, es müsste ein intelligentes (kognitives) technisches System geschaffen werden, welches die Benutzer und sich selber beobachtet, daraus Schlüsse zieht, um die Parameter seiner Systemkomponenten im Kontext in dem das System steht anzupassen. Solche Systeme sind in Ihrer Idee keineswegs neu. Sie sind Bestandteil des Denkens in den Bereichen der Kybernetik, Robotik und künstlichen Intelligenz.

Nachstehend sollen zwei miteinander verwandte Ansätze vorgestellt werden. Es wird lediglich die Grundidee skizziert. Die dafür nötigen Komponenten u.a. zur Messung der Wahrnehmung beim Nutzer und der detaillierten Abbildung im technischen System seien hier ausgeklammert.

**Adaptive Steuerung der technischen Qualitätselemente eines audio-visuellen Systems zur Resynthese räumlicher audio-visueller Szenen** - Bei Binauralsynthesystemen sollen die räumliche Auflösung sowie die Frequenzauflösung der verwendeten HRTFs bzw. BRIRs adaptiv steuerbar sein. Raumbereiche, in denen keine bzw. eine geringe Aufmerksamkeit des Nutzers liegt, können mit einer geringeren Auflösung synthetisiert werden. Bereiche mit hoher Aufmerksamkeit des Nutzers können mit einer höheren Auflösung synthetisiert werden, um einen hohen Grad an Immersion und Plausibilität der Szene zu erreichen. Weiterhin können die Qualitätselemente einer vorhandenen akustischen Raumsimulation steuerbar werden. Raumbereiche mit gering bzw. keiner Aufmerksamkeit des Nutzers können mit geringerer Komplexität simuliert werden, als Bereiche in denen eine hohe Aufmerksamkeit des Nutzers liegt. Als Ziel steht, die Qualitätsmerkmale, wie bspw. Externalität, Lokalisation, Distanz, Klangfarbe und Räumlichkeit in ihrer wahrgenommenen Qualität zu maximieren. Es soll gleichzeitig eine Reduktion der Komplexität der durchzuführenden Berechnungen erzielt werden.

**Steuerung des Ortes der Aufmerksamkeit eines Nutzers innerhalb einer räumlichen audio-visuellen Szene** - Die Blickrichtung des Nutzers soll durch eine gezielte Anpassung von räumlichen, auditiven Merkmalen der Synthese verändert werden. Hierfür sollen monaurale und binaurale Merkmale in ihrer perzeptiven Wirkung verstärkt bzw. verringert werden. Weiterhin soll eines dem visuellen Ereignis vorausseilendes auditiven Hörereignis den Ort der Aufmerksamkeit des Nutzers auf das Objekt lenken. Diese Veränderungen beziehen sich auf einzelne zu synthetisierende Objekte oder Raumbereiche der wiederzugebenden auditiven Szene. Es sollen gezielt Wahrnehmungseffekte, wie bspw. Nähe/Entfernung, Im-Kopf-Lokalisation/Externalität, Lokalisationunschärfe/Lokalisationschärfe erzeugt werden. Weiterhin sollen die zeitlichen Abfolgen und Synchronizitäten des Auftretens von visuellen Objekten und dazugehörigen auditiven Objekten, bspw. Automobil mit dazugehörigen Motor- und Rollgeräuschen, steuerbar werden.





**chekov:** Welchen Kurs, Captain?

**kirk:** Der zweite Stern von rechts. Bis zum Morgengrauen. Direkter Kurs.

Dialog aus Star Trek VI: Das unentdeckte Land

## Literaturverzeichnis

### Eigene Veröffentlichungen

- [1] Carsten Boensel, Martina Böhme, David Goecke, Thomas Mayenfels, M. Schnemilich, Thomas Sporer, Stephan Werner, M. Wolf und A. Würsig. *Study on effects of BRIR set on Quality Assessment*. Techn. Ber. ISO/IEC JTC1/SC29/WG11 MPEG2013/M30222. Wien, Österreich: Fraunhofer IDMT und Technische Universität Ilmenau, 2013 (siehe S. 5, 57).
- [2] Karlheinz Brandenburg, Florian Klein, Annika Neidhardt und Stephan Werner. “Auditory Illusion over Headphones Revisited”. In: *International Congress on Acoustics, Boston, USA*. 2017 (siehe S. 46).
- [3] Karlheinz Brandenburg, Stephan Werner, Florian Klein und Christoph Sladeczek. “Auditory Illusions through Headphones: History, Challenges, and New Solutions”. In: *International Congress on Acoustics, Buenos Aires, Argentina*. 2016 (siehe S. 46).
- [4] Karlheinz Brandenburg, Stephan Werner und Christoph Sladeczek. *Determining and using room-optimized transfer functions*. US Patent 10,003,906, Erteilungstag 19. Juni 2018. 2018 (siehe S. 135).
- [5] Simone Füg, Stephan Werner und Karlheinz Brandenburg. “Controlled Auditory Distance Perception using Binaural Headphone Reproduction - Algorithms and Evaluation”. In: *VDT Int. Conv., 27. Tonmeistertagung*. 2012, S. 614–620 (siehe S. 151, 155).
- [6] Frank Jürgens und Stephan Werner. “Loudness Curves for binaural playback”. In: *27. Tonmeistertagung des VDT, Köln, Deutschland*. 2012 (siehe S. 48, 51).
- [7] Florian Klein und Stephan Werner. “HRTF adaptation and pattern learning”. In: *4th International Symposium on Auditory and Audiological Research (ISAAR), Denmark*. 2013, S. 69–76 (siehe S. 4).
- [8] Florian Klein, Stephan Werner und Thomas Mayenfels. “Influences of training on externalization in binaural synthesis in situations of room divergence”. In: *Journal of the Audio Engineering Society* 65.3 (2017), S. 178–187. DOI: 10.17743/jaes.2016.0072 (siehe S. 4, 105 f., 125, 128).
- [9] Florian Klein, Stephan Werner und Martin Rekitt. “Evaluierung einer virtuellen akustischen Umgebung für Hörgeräte”. In: *44. Jahrestagung für Akustik, DAGA, Garching bei München, Deutschland*. 2018 (siehe S. 181).

- [10] Christina Mittag, Martina Böhme und Stephan Werner. “Dataset of Binaural Room Impulse Responses at Multiple Recording Positions, Source Positions, and Orientations in a Real Room”. In: *43. Jahrestagung für Akustik, DAGA 2017, Kiel, Deutschland*. 2017 (siehe S. 166).
- [11] Christina Mittag, Martina Böhme und Stephan Werner. *Dataset of KEMAR-BRIRs measured at several positions and head orientations in a real room*. Dataset, Technische Universität Ilmenau. 2016. DOI: 10.5281/zenodo.206860 (siehe S. 166).
- [12] Christina Mittag, Stephan Werner und Florian Klein. “Development and Evaluation of Methods for the Synthesis of Binaural Room Impulse Responses based on Spatially Sparse Measurements in Real Rooms”. In: *43. Jahrestagung für Akustik, DAGA, Kiel, Deutschland*. 2017 (siehe S. 151, 153, 164).
- [13] Rebecca Sass, Stephan Werner und André Siegel. “Comparison of recording methods for measurements of individualized head-related transfer functions”. In: *VDT Int. Conv., 26. Tonmeistertagung, Leipzig, Deutschland*. 2010 (siehe S. 75).
- [14] Thomas Sporer u. a. *Study on localization of audio objects*. Techn. Ber. ISO/IEC JTC1/SC29/WG11 MPEG2014/M32328. San Jose, USA: Fraunhofer IDMT und Technische Universität Ilmenau, 2014 (siehe S. 5).
- [15] Stephan Werner. *DRR-scaled Individual Binaural Room Impulse Responses*. Dataset, Technische Universität Ilmenau. 2016. DOI: 10.5281/zenodo.61072 (siehe S. 139).
- [16] Stephan Werner und Simone Fueg. “Controlled Auditory Distance Perception using Binaural Headphone Reproduction - Evaluation via Listening Tests”. In: *VDT Int. Convention, 27. Tonmeistertagung, Cologne, Germany*. 2012, S. 622–629 (siehe S. 151, 161 f.).
- [17] Stephan Werner, Georg Götz und Florian Klein. “Influence of Head Tracking on the Externalization of Auditory Events at Divergence between Synthesized and Listening Room Using a Binaural Headphone System”. In: *142nd Convention of the Audio Engineering Society (AES), Berlin, Deutschland*. 2017 (siehe S. 121).
- [18] Stephan Werner, Tamás Harczos und Karlheinz Brandenburg. “Overview of numerical models of cell types in the cochlear nucleus”. In: *2nd International Symposium on Auditory and Audiological Research ISAAR, Denmark*. 2009, S. 61–70 (siehe S. 28).
- [19] Stephan Werner und Florian Klein. “Context and Plausibility in Binaural Synthesis”. In: *2nd Int. Conf. on Spatial Audio (ICSA), Erlangen, Deutschland*. 2014 (siehe S. 46, 66, 86).
- [20] Stephan Werner und Florian Klein. “Influence of Context Dependent Quality Parameters on the Perception of Externalization and Direction of an Auditory Event”. In: *Audio Engineering Society Conference: 55th International Conference: Spatial Audio, Helsinki, Finnland*. 2014 (siehe S. 66, 86, 114 f., 119).

- [21] Stephan Werner, Florian Klein und Karlheinz Brandenburg. “Influence of Spatial Complexity and room Acoustic Disparity on Perception of Quality Features Using a Binaural Synthesis System”. In: *7th International Workshop on Quality of Multimedia Experience (QoMEX), Greece*. 2015 (siehe S. 110).
- [22] Stephan Werner, Florian Klein und Tamás Harcos. “Context-dependent quality parameters and perception of auditory illusions”. In: *4th International Symposium on Auditory and Audiological Research ISAAR, Denmark*. 2013, S. 445–452 (siehe S. 4, 66, 86, 158).
- [23] Stephan Werner, Florian Klein, Thomas Mayenfels und Karlheinz Brandenburg. “A Summary on Acoustic Room Divergence and its Effect on Externalization of Auditory Events”. In: *8th International Conference on Quality of Multimedia Experience (QoMEX), Portugal*. 2016. DOI: 10.1109/QoMEX.2016.7498973 (siehe S. 66, 86).
- [24] Stephan Werner, Florian Klein und Thomas Sporer. “Adjustment of the Direct-to-Reverberant-Energy-Ratio to Reach Externalization within a Binaural Synthesis System”. In: *AES Conference on Audio for Virtual and Augmented Reality, Los Angeles, CA, USA*. 2016 (siehe S. 139).
- [25] Stephan Werner. und Judith Liebetrau. “Adjustment of Direct-to-Reverberant-Energy-Ratio and the Just-Noticable-Difference”. In: *6th International Workshop on Quality of Multimedia Experience (QoMEX), Singapore*. 2014 (siehe S. 135).
- [26] Stephan Werner und Judith Liebetrau. “Effects of Shaping of Binaural Room Impulse Responses on Localization”. In: *5th International Workshop on Quality of Multimedia Experience (QoMEX), Österreich*. 2013, S. 88–93 (siehe S. 162).
- [27] Stephan Werner, Judith Liebetrau und Thomas Sporer. “Vertical Sound Source Localization Influenced by Visual Stimuli”. In: *Signal Processing Research 2.2 (2013)*, S. 29–38 (siehe S. 4, 74).
- [28] Stephan Werner, Annika Neidhardt, Florian Klein und Karlheinz Brandenburg. “Comparison of Different Methods to Create an Interactive Augmented Auditory Reality Scenario Using Sparse Binaural Room Impulse Response Measurements”. In: *44. Jahrestagung für Akustik, DAGA, Garching bei München, Deutschland*. 2018 (siehe S. 165).
- [29] Stephan Werner, Martin Rekitt und Florian Klein. “Distribution of Quadrant Errors in Auditory Localization using a Binaural Headphone system”. In: *41. Jahrestagung für Akustik, DAGA, Nürnberg, Deutschland*. 2015 (siehe S. 107 f., 121).
- [30] Stephan Werner und Rebecca Sass. “Synthesis of Binaural Room Impulse Responses”. In: *AIA-DAGA, 39th German Annual Conference on Acoustics, Merano, Italien*. 2013, S. 572–575 (siehe S. 158).

- [31] Stephan Werner und André Siegel. “Effects of binaural auralization via headphones on the perception of acoustic scenes”. In: *3rd International Symposium on Auditory and Audiological Research ISAAR, Denmark*. 2011, S. 215–222 (siehe S. 4, 66, 86, 136).
- [32] Stephan Werner und Anett Zabel. “Effects of the Direct-to-Reverberant-Ratio and Divergence Between Synthesized and Listening Room on Perception of Virtual Acoustic Environment”. In: *AIA-DAGA, 39th German Annual Conference on Acoustics, Merano, Italien*. 2013, S. 566–568 (siehe S. 135).

### Betreute Abschlussarbeiten

- [33] Simone Füg. *Untersuchungen zur Distanzwahrnehmung von Hörereignissen bei Kopfhörerwiedergabe*. Masterarbeit, Fachgebiet Elektronische Medientechnik, Technische Universität Ilmenau, Deutschland. 2012 (siehe S. 151, 153, 155 f., 158).
- [34] Bernhard Fiedler und T. Heidl. *Einfluss der Simulationsparameter auf die Wahrnehmung virtueller Räume für bewegliche Hörer*. Medienprojekt, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2015 (siehe S. 152).
- [35] Georg Götz. *Untersuchung zum Einfluss von Head-Tracking auf die Externalisierung von Hörereignissen bei Divergenz zwischen synthetisierter Szene und Abhörraum unter Verwendung eines binauralen Kopfhörersystems*. Bachelorarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2016 (siehe S. 121).
- [36] Markus Hesse. *Mikrofonierung mit dem Motion-Tracked Binaural Verfahren*. Diplomarbeit, Technische Universität Ilmenau, Fakultät für Elektrotechnik und Informationstechnik, Fachgebiet Elektronische Medientechnik. 2012 (siehe S. 107, 153).
- [37] Frank Jürgens. *Kurven gleicher Lautheit bei binauraler Kopfhörerwiedergabe*. Masterarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2012 (siehe S. 48–50).
- [38] Johannes Maggi. *Binaurale Wiedergabe über Lautsprecher in Stereoanordnung*. Bachelorarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2010 (siehe S. 47).
- [39] Thomas Mayenfels. *Untersuchung zum Einfluss von Training auf die Wahrnehmung von Externalität*. Masterarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2015 (siehe S. 128).
- [40] Christina Mittag. *Entwicklung und Evaluierung eines Verfahrens zur Synthese von binauralen Raumimpulsantworten basierend auf räumlich dünnbesetzten Messungen in realen Räumen*. Masterarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2016 (siehe S. 151, 153 f., 164).

- [41] Christoph Neukam. *Kompensation nicht optimaler Stereoanordnungen für die Wiedergabe von stereophonen/binauralen Signalen*. Bachelorarbeit, Technische Universität Ilmenau, Fachgebiet Elektronische Medientechnik. 2010 (siehe S. 47).
- [42] Nicolas Pachatz. *Untersuchungen zur Relevanz raumakustischer Parameter bei Anpassung eines Binauralsynthesystems an die Raumakustik des Abhörtraumes*. Bachelorarbeit, Fachgebiet Elektronische Medientechnik, Technische Universität Ilmenau, Deutschland. 2017 (siehe S. 154 f.).
- [43] Martin Rekitt. *Bestimmung der Häufigkeitsverteilung von Quadrantenfehlern bei der Lokalisation von Hörereignissen unter Verwendung einer binauralen Kopfhörerwiedergabe*. Bachelorarbeit, Fachgebiet Elektronische Medientechnik, Technische Universität Ilmenau, Deutschland. 2014 (siehe S. 107).
- [44] Rebecca Sass. *Synthese binauraler Raumimpulsantworten*. Masterarbeit, Fachgebiet Elektronische Medientechnik, Technische Universität Ilmenau, Deutschland. 2012 (siehe S. 153 f., 158).
- [45] Anett Zabel. *Vergleich von Hörtestmethodiken zur Beurteilung der räumlichen Wahrnehmung bei binauraler Kopfhörerwiedergabe*. Bachelorarbeit, Fachgebiet Elektronische Medientechnik, TU Ilmenau, Deutschland. 2012 (siehe S. 142).

### Andere Veröffentlichungen

- [46] Esam Ragab Esmail Abou-Elleal. “Raumakustik - Interaktion visueller und auditiver Wahrnehmung”. Diss. Aachen, Deutschland: Rheinisch-Westfälische technische Hochschule Aachen, Fakultät für Architektur, 2003 (siehe S. 69).
- [47] David Alais und David Burr. “The ventriloquist effect results from near-optimal bimodal integration”. In: *Current Biology* 14.3 (2004), S. 257–262. DOI: 10.1016/j.cub.2004.01.029 (siehe S. 69 f.).
- [48] Ralph Algazi, Richard O. Duda, und Dennis M. Thompson. “Motion-Tracked Binaural Sound”. In: *Journal of the Audio Engineering Society* 52.11 (2004), S. 1142–1156 (siehe S. 153).
- [49] Ralph Algazi, Richard O. Duda, Dennis M. Thompson und C. Avendano. “The CIPIC HRTF Database”. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York*. 2001, S. 99–102 (siehe S. 31).
- [50] Arduino. <http://www.arduino.cc>. Arduino-Mega. 2013 (siehe S. 76, 81).
- [51] Jonathan Ashmore. “Cochlear Outer Hair Cell Motility”. In: *Physiological Reviews* 88 (2008), S. 173–210. DOI: 10.1152/physrev.00044.2006 (siehe S. 24).
- [52] Bishnu S. Atal und Manfred R. Schröder. *Nachahmung der Raumakustik durch den Elektronenrechner*. Gravesaner Blätter, 27/28, Ars Viva Verlag. 1966 (siehe S. 46).

- [53] Sarah Le Bagousse, Catherine Colomes und Mathieu Paquier. “State of the Art on Subjective Assessment of Spatial Sound Quality”. In: *38th International Conference: Sound Quality Evaluation, Audio Engineering Society*, 2010. URL: <http://www.aes.org/e-lib/browse.cfm?elib=15477> (siehe S. 4).
- [54] George A. Barnhard. “A New Test for 2 x 2 Tables”. In: *Nature* 156 (1945), S. 783–784 (siehe S. 95).
- [55] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia*. National Aeronautics und Space Administration Ames Research Center California USA, 2000 (siehe S. 58).
- [56] Durand R. Begault. “Tutorial on distance perception”. In: *Audio Engineering Society 118th Convention, Barcelona, and 119th Convention, New York*. 2005 (siehe S. 58).
- [57] Durand R. Begault, Elizabeth M. Wenzel und Mark R. Anderson. “Direct Comparison of the Impact of Head Tracking, Reverberation and Individualized Head-Related Transfer Functions on the Spatial Perception of Virtual Speech Source”. In: *J. Audio Eng. Soc.* 10.49 (2001), S. 904–916 (siehe S. 36, 58, 62, 89, 95, 115, 119, 121, 134).
- [58] Leo L. Beranek. *Concert halls and opera houses: music, acoustics, and architecture*. 2. Ausgabe. Springer Verlag, New York, 2004. ISBN: 0387955240 (siehe S. 4, 38, 56).
- [59] Jan Berg. “Evaluation of Perceived Spatial Audio Quality”. In: *Systemics, Cybernetics and Informatics* 2 (2006), S. 10–14 (siehe S. 56).
- [60] Jan Berg und Francis Rumsey. “Systematic Evaluation of Perceived Spatial Quality”. In: *AES 24th International Conference on Multichannel Audio*. 2003 (siehe S. 56).
- [61] Paul Bertelson und Monique Radeau. “Cross-modal bias and perceptual fusion with auditory-visual spatial discordance”. In: *Perception and Psychophysics* 6.29 (1981), S. 578–584 (siehe S. 69, 81, 84 f.).
- [62] Georg von Békésy. *Physiologie der Sinneshemmung*. Das wissenschaftliche Taschenbuch, Goldmann, München, 1970 (siehe S. 24 f.).
- [63] Jens Blauert. “Analysis and Synthesis of Auditory Scenes”. In: *Communication Acoustics*. Hrsg. von J. Blauert. Springer, 2005, S. 1–25. ISBN: 354022162 (siehe S. 7, 14 f.).
- [64] Jens Blauert. *Spatial Hearing - The Psychophysics of Human Sound Localization*. Mit University Press Group Ltd, Cambridge, London, 1996. ISBN: 9780262024136 (siehe S. 11 f., 31, 36–39, 76, 156, 159).
- [65] Jens Blauert. *The Technology of Binaural Listening (Modern Acoustics and Signal Processing)*. Springer Berlin Heidelberg, 2013. ISBN: 9783642377624 (siehe S. 4).
- [66] Jens Blauert und Jonas Braasch. “The Precedence Effect”. In: *Acoustic Communication* (2005) (siehe S. 65).
- [67] Jens Blauert und P. L. Divenyi. “Spectral selectivity in binaural contralateral inhibition”. In: *Acoustics* 66 (1988), S. 267–274 (siehe S. 64).

- [68] Robert W. Bohlander. “Eye position and visual attention influence perceived auditory direction”. In: *Percep. Mot. Skills* 59.2 (1984), S. 483–510. DOI: 10.2466/pms.1984.59.2.483 (siehe S. 69).
- [69] BoysTownHospital. <http://www.audres.org/>. 2018 (siehe S. 24).
- [70] John S. Bradley und Gilbert A. Soulodre. “The influence of late arriving energy on spatial impression”. In: *J. Acoust. Soc. Am.* 97.4 (1995), S. 2263–2271. DOI: 10.1121/1.411951 (siehe S. 118).
- [71] W. Owen Brimijoin, Alan W. Boyd und Michael A. Akeroyd. “The Contribution of Head Movement to the Externalization and Internalization of Sounds”. In: *PLoS ONE* 8.12 (2013). DOI: 10.1371/journal.pone.0083068 (siehe S. 62, 121).
- [72] Douglas S. Brungart. “Near-Field Auditory Localization”. Diss. Massachusetts Institute of Technology, Dept. of Electrical Engineering und Computer Science, USA, 1998 (siehe S. 38 f., 58).
- [73] Douglas S. Brungart und William M. Rabinowitz. “Auditory localization of nearby sources. Head-related transfer functions”. In: *J. Acoust. Soc. Am.* 106.3 (1999), S. 1465–1479 (siehe S. 39).
- [74] Laurel H. Carney. “Sensitivities of cells in anteroventral cochlear nucleus of cat to spatio-temporal discharge patterns across primary afferents”. In: *Journal of Neurophysiology* 64.2 (1990), S. 437–456. DOI: 10.1152/jn.1990.64.2.437 (siehe S. 29).
- [75] E. J. Chichilnisky. “A simple white noise analysis of neuronal light responses”. In: *Network: Computation in Neural Systems* 12 (2001), S. 199–213 (siehe S. 183 f.).
- [76] Rachel K. Clifton. “Breakdown of echo suppression in the precedence effect”. In: *J. Acoust. Soc. Am.* 82.5 (1987), S. 1834–1835 (siehe S. 64, 66, 135).
- [77] Rachel K. Clifton und Richard L. Freyman. “Effect of click rate and delay on breakdown of the precedence effect”. In: *Percept. Psychophys.* 2.49 (1989), S. 139–145 (siehe S. 4, 64).
- [78] Rachel K. Clifton, Richard L. Freyman, Ruth Y. Litovsky und Daniel McCall. “Listeners’ expectations about echoes can raise or lower echo threshold”. In: *J. Acoust. Soc. Am.* 3.95 (1993), S. 1525–1533 (siehe S. 64).
- [79] Paul D. Coleman. “Failure to localize the source distance of an unfamiliar sound”. In: *J. Acoust. Soc. Am.* 34 (1962), S. 345–346 (siehe S. 36 f.).
- [80] Johannes C. Dahmen, Peter Keating, Fernando R. Nodal, Andreas L. Schulz und Andrew J. King. “Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space”. In: *Neuron* 66.6 (2010), S. 937–947. DOI: 10.1016/j.neuron.2010.05.018 (siehe S. 183).



- [81] Thomas Djelani und Jens Blauert. “Investigations Into the Build-Up and Breakdown of the Precedence Effect”. In: *acta acustica - Acustica* 87 (2001), S. 253–261 (siehe S. 64).
- [82] Thomas Djelani, Christoph Pörschmann, Jörg Sahrhage und Jens Blauert. “An interactive virtual environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization”. In: *Acustica* 6.86 (2000), S. 1046–1053 (siehe S. 46).
- [83] Duden. *Illusion, die*. Eintrag im Duden zum Wort Illusion; <https://www.duden.de/rechtschreibung/Illusion>, abger. 26.06.2018. 2018 (siehe S. 7).
- [84] Günter Ehret und Raymond Romand. *The Central Auditory System*. Oxford University Press, 1997. ISBN: 0195096843 (siehe S. 28–30).
- [85] Vera Erbes, Frank Schultz, Alexander Lindau und Stefan Weinzierl. “An extraural headphone system for optimized binaural reproduction”. In: *38. Jahrestagung für Akustik, DAGA, Darmstadt, Deutschland*. 2012 (siehe S. 122).
- [86] European Network on Quality of Experience in Multimedia Systems and Service. *Qualinet White Paper on Definitions of Quality of Experience*. Techn. Ber. Output from the 5th Qualinet Meeting, Novi Sad, 2013 (siehe S. 8 f., 12).
- [87] Jochen Fanghänel, Franz Pera, Friedrich Anderhuber und Robert Nitsch, Hrsg. *Waldeyer Anatomie des Menschen - Lehrbuch und Atlas in einem Band*. De Gruyter, Berlin, 2009. ISBN: 9783110193534 (siehe S. 24, 28).
- [88] Angelo Farina. “Simultaneous measurement of impulse response and distortion with a swept-sine technique”. In: *Audio Engineering Society Convention 108th, Italy*. 2000 (siehe S. 44).
- [89] Franz Faul, Edgar Erdfelder, Albert-Georg Lang und Axel Buchner. “G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences”. In: *Behavior Research Methods* 39.2 (2007), S. 175–191 (siehe S. 90).
- [90] Janina Fels. “From Children to Adults: How Binaural Cues and Ear Canal Impedances Grow”. Diss. Rheinisch-Westfälische Technische Hochschule Aachen, Fakultät für Elektrotechnik und Informationstechnik, 2008 (siehe S. 32).
- [91] Andy Fields, Jeremy Miles und Zoe Field. *Discovering Statistics Using R*. SAGE Publications, 2012. ISBN: 9781446200452 (siehe S. 109).
- [92] Harvey Fletcher. *Speech and Hearing in Communication*. D. van Nostrand Company, Inc. New York, USA, 1953 (siehe S. 37).

- [93] Heinz von Foerster und Bernhard Pörksen. *Wir sehen nicht, dass wir nicht sehen*. Interview veröffentlicht in <https://heise.de/-3446178> als Vorabdruck aus dem Buches: Heinz von Foerster und Bernhard Pörksen: Wahrheit ist die Erfindung eines Lügners. Gespräche für Skeptiker. Carl-Auer-Systeme Verlag, Heidelberg, 2006. 1998 (siehe S. 133).
- [94] Richard L. Freyman, Rachel K. Clifton und Ruth Y. Litovsky. “Dynamic processes in the precedence effect”. In: *J. Acoust. Soc. Am.* 90 (1991), S. 874–884 (siehe S. 64).
- [95] Robert D. Frisina, Robert L. Smith und Steven C. Chamberlain. “Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement”. In: *Hearing Research* 44.2-3 (1990), S. 99–122. DOI: 10.1016/0378-5955(90)90074-Y (siehe S. 29).
- [96] Takuya Fukuda, Toshiharu Horiuchi, Haruhide Hokari und Shoji Shimada. “Relative distance perception by manipulating the ILD of HRTFs”. In: *Journal of the Acoustical Society of Japan (Acoust. Sci. and Tech.)* 24.5 (2003), S. 325–326. DOI: 10.1250/ast.24.325 (siehe S. 39).
- [97] Tal Galili. *Barnards exact test - a powerful alternative for Fishers exact test*. <http://www.r-statistics.com/>. 2010 (siehe S. 95).
- [98] Juan C. Gil-Carvajal, Jens Cubick, Sebastien Santurette und Torsten Dau. “Spatial Hearing with Incongruent Visual or Auditory Room Cues”. In: *Nature Scientific Reports* 6:37342 (2016). DOI: 10.1038/srep37342 (siehe S. 62, 66–68).
- [99] Ernst von Glasersfeld. *Radikaler Konstruktivismus: Ideen, Ergebnisse, Probleme*. Suhrkamp Taschenbuch Wissenschaft, 1997. ISBN: 9783518289266 (siehe S. 2, 177).
- [100] Tillmann Gronert. *Distanzwahrnehmung bei virtueller Tondarstellung mittels Wellenfeldsynthese*. Diplomarbeit, Technische Fachhochschule Berlin, Institut für Rundfunktechnik. 2003.
- [101] Benedikt Grothe. *Wie die Evolution die Ohren aufsperrt*. Im Reich der Sinne. Max Planck Forschung, 2001 (siehe S. 22).
- [102] Helmut Haas. “Über den Einfluss eines Einfach-Echos auf die Hörsamkeit von Sprache”. In: *Acoustics* 1 (1951) (siehe S. 64).
- [103] Dorte Hammershøi. “Binaural Technique - a method of true 3D sound reproduction”. Diss. Department of Electronic Systems, Universität Aalborg, 1995 (siehe S. 32, 41, 57).
- [104] Tamás Harczos, Waldo Nogueira, Gero Szepannek und Frank Klefenz. “Comparative evaluation of successive cochlear modeling stages as possible front-ends for automatic speech recognition”. In: *19th Int. Congress on Acoustics (ICA), Madrid, Spain*. Bd. 38. 3-4. 2007 (siehe S. 27).

- [105] William M. Hartmann und Andrew T. Wittenberg. “On the externalization of sound images”. In: *J. Acoust. Soc. Am.* 6.99 (1996), S. 3678–3688. DOI: 10.1121/1.414965 (siehe S. 32, 61 f., 76, 92).
- [106] Henrik G. Hassager, Frederik Gran und Torsten Dau. “The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment”. In: *The Journal of the Acoustical Society of America* 139.5 (2016), S. 2992–3000. DOI: 10.1121/1.4950847 (siehe S. 152).
- [107] Kerstin Heckler. *Jean Piagets Theorie der geistigen Entwicklung*. Zusammenfassung für das Vordiplom, Universität Koblenz. elektronische Veröffentlichung Universität Koblenz, abgerufen 07.12.2017. 2008 (siehe S. 17).
- [108] Carrie Heeter. “Being There: The subjective experience of presence”. In: *Presence: Teleoperators and Virtual Environments* 1.2 (1992), S. 262–271 (siehe S. 2).
- [109] Etienne Hendrickx, Peter Stitt, Jean-Christophe Messonnier, Jean-Marc Lyzwa, Brian F. G. Katz und Catherine De Boisheraud. “Improvement of Externalization by Listener and Source Movement Using a Binauralized Microphone Array”. In: *Journal of the Audio Engineering Society* 65 (7/8), S. 590–599. DOI: 10.17743/jaes.2017.0018 (siehe S. 62).
- [110] H. Henrik Møller, Dorte Hammershøi, Clemen B. Jensen und Michael F. Sørensen. “Transfer characteristics of headphones measured on human ears”. In: *J. Audio Eng. Soc.* 43 (1995), S. 203–217 (siehe S. 47).
- [111] Wolfgang Hess. “Time-Variant binaural-Activity Characteristics as Indicator of Auditory Spatial Attributes”. Diss. Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 2006 (siehe S. 39).
- [112] Richard C. Heyser. *Hearing vs. Measurement*. Audio Magazine, Diamandis Communications, Inc. 1988 (siehe S. 7).
- [113] Richard C. Heyser. *Zitat*. Anschreiben zur Heyser Memorial Lecture Series auf der 142. AES Convention, Berlin, Deutschland. 2017 (siehe S. 55).
- [114] Paul Hofman und John van Opstal. “Comparative evaluation of successive cochlear modeling stages as possible front-ends for automatic speech recognition”. In: *International Work-Conference on Artificial Neural Networks, IWANN 1997: Biological and Artificial Computation: From Neuroscience to Technology*. 1997, S. 1126–1135 (siehe S. 35).
- [115] Radio Communication Assembly International Telecommunication Union. *Acoustic Measurement of the reverberation time of rooms with reference to other acoustical parameters*. BS EN ISO 3382-2000, ISBN: 0580346986, Publisher: BSI. 2000 (siehe S. 52–54).

- [116] Radio Communication Assembly International Telecommunication Union. *Methods for the subjective assessment of small impairments in audio systems including Multichannel Sound Systems*. Recommendation ITU-R BS.1116-1 (10/1997). 1997 (siehe S. 52, 169).
- [117] Radio Communication Assembly International Telecommunication Union. *Multichannel stereophonic sound system with and without accompanying picture*. Recommendation ITU-R BS.775-2 (08/2012). 2012 (siehe S. 123).
- [118] Telecommunication Standardization Sector International Telecommunication Union. *Definitions of terms related to quality of service*. Recommendation ITU-T E.800 (09/2008). 2008 (siehe S. 9).
- [119] Telecommunication Standardization Sector International Telecommunication Union. *Subjective quality evaluation of telephone services based on spoken dialogue systems*. Recommendation ITU-T P.851 (11/2003). 2003 (siehe S. 166).
- [120] Jürg Jecklin. *3. Mikrofone, Vorlesungsskript Universität für Musik und darstellende Künste Wien*. <http://www.mdw.ac.at/I101/iea/tm/scripts/jecklin/tt03mikrofon.pdf>. 2003 (siehe S. 41).
- [121] Ute Jekosch. “Basic concepts and terms of quality, reconsidered in the context of product-sound quality”. In: *Acta Acustica united with Acustica* 6.90 (2004), S. 999–1006 (siehe S. 13).
- [122] Ute Jekosch. *Voice and Speech Quality perception - Assessment and Evaluation*. Springer Series in Signals and Communications Technology, Springer-Verlag Berlin Heidelberg, 2005. ISBN: 9783540240952 (siehe S. 2, 14 f., 188).
- [123] Ute Jekosch und Jens Blauert. “Einige Grundüberlegungen zur Qualität von Geräuschen”. In: *31. Deutsche Jahrestagung für Akustik, DAGA, München, Deutschland*. 2005 (siehe S. 14, 17).
- [124] H. Jo, Y. Park und W.L. Martens. “Elimination of minor peaks and notches in head-related transfer functions supports improved angular identification and discrimination on the median plane”. In: *J. Audio Eng. Soc.* (2011) (siehe S. 153).
- [125] John H. Wittig Jr. “Encoding and Enhancing Acoustic Information at the First Stages of the Auditory System”. Diss. University of Pennsylvania, USA, 2004 (siehe S. 29).
- [126] Satu Jumisko-Pyykkö. “User-Centered Quality of Experience and Its Evaluation Methods for Mobile Television”. Diss. Tampere University of Technology, 2011 (siehe S. 18).
- [127] Bill Kapralos, Minchael Jenkin und Evangelos Miliotis. *Auditory Perception and Spatial (3D) Auditory Systems*. Techn. Ber. Department of Computer Science, University York, Canada, 2003 (siehe S. 39).

- [128] Hendrik Kayser, Stephan D. Ewert, Jörn Anemüller, Thomas Rohdenburg, Volker Hohmann und Birger Kollmeier. “Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses”. In: *EURASIP Journal on Advances in Signal Processing* 2009. Article ID 298605 (2009) (siehe S. 35, 89).
- [129] Rachel Keen und Richard L. Freyman. “Release and re-build of the listern’s models of auditory space”. In: *J. Acoust. Soc. Am.* 125(5) (2009), S. 3243–3251 (siehe S. 65).
- [130] Florian Klein, Annika Neidhardt, Marius Seipel und Thomas Sporer. “Training on the Acoustical Identification of the Listening Position in a Virtual Environment”. In: *143rd Audio Engineering Society Convention*. 2017 (siehe S. 152).
- [131] Clemens Kuhn-Rahloff. “Prozesse der Plausibilitätsbeurteilung am Beispiel ausgewählter elektroakustischer Wiedergabesituationen”. Diss. Berlin, Deutschland: Technische Universität Berlin, Fakultät I - Geisteswissenschaften, 2011 (siehe S. 2, 8).
- [132] Kristina Kunze, Judith Liebetau und Thomas Korn. “Evaluating Spatial Congruency of 3D Audio and Video Objects”. In: *132nd Audio Engineering Society Convention*. 2012. URL: <http://www.aes.org/e-lib/browse.cfm?elib=16317> (siehe S. 70).
- [133] Erik Larsen, Nandini Iyer, Charissa R. Lansing und Albert S. Feng. “On the minimum audible difference in direct to reverberant energy ratio”. In: *J. Acoust. Soc. Am.* 124.1 (2008), S. 450–461 (siehe S. 37 f., 139, 143, 145, 149).
- [134] Pontus Larsson, Daniel Västfjäll und Mendel Kleiner. “Auditory visual interaction in real and virtual rooms”. In: *Forum Acusticum, Sevilla, Spanien*. 2002 (siehe S. 71).
- [135] Pontus Larsson, Daniel Västfjäll und Mendel Kleiner. “Ecological acoustics and the multimodal perception of rooms: real and unreal experiences of auditory-visual virtual environments”. In: *International Conference on Auditory Display, Espoo, Finland*. 2001, S. 245–249 (siehe S. 70 f.).
- [136] Peter Laws. “Entfernungshören und das Problem der Im-Kopf-Lokalisiertheit von Hörereignissen (Auditory distance perception and the problem of in head localization of sound images)”. In: *Acustica, NASA Technical Translation TT-20833* 29 (1973), S. 243–259 (siehe S. 32, 89).
- [137] Tobias Lentz. “Binaural Technology for Virtual Reality”. Diss. Lehrstuhl und Institut für Technische Akustik, Rheinisch-Westfälische Technische Hochschule Aachen, 2007 (siehe S. 32, 46).
- [138] Tomasz Letowski. “Sound Quality Assessment: Concepts and Criteria”. In: *87th AES Convention, New York, USA*. 1989 (siehe S. 56).
- [139] Alexander Lindau und Fabian Brinkmann. “Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings”. In: *3rd ISCA/DEGA Tutorial and Research Workshop on Perceptual Quality of Systems* (2010), S. 137–142 (siehe S. 32).

- [140] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman und Stefan Weinzierl. “A Spatial Audio Quality Inventory for Virtual Acoustic Environments (SAQI)”. In: *Acta Acustica united with Acustica* 100.5 (2014), S. 984–994 (siehe S. 4, 56 f.).
- [141] Alexander Lindau, Linda Kosanke und Stefan Weinzierl. “Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses”. In: *128th Audio Engineering Society Convention, London, UK*. 2010 (siehe S. 154).
- [142] Alexander Lindau und Stefan Weinzierl. “Assessing the plausibility of virtual acoustic environments”. In: *Forum Acusticum, European Acoustic Association, Aalborg, Denmark*. 2011, S. 1187–1192 (siehe S. 2).
- [143] Ruth Y. Litovsky, H. Steven Colburn, William A. Yost und Sandra J. Guzman. “The precedence effect”. In: *J. Acoust. Soc. Am.* 106.4 (1999), S. 1633–1654. DOI: 10.1121/1.427914 (siehe S. 64).
- [144] Ruth Y. Litovsky und Barbara G. Shinn-Cunningham. “Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression”. In: *J. Acoust. Soc. Am.* 109.1 (2001), S. 346–358 (siehe S. 64).
- [145] Dries H. G. Louage, Marcel van der Heijden und Philip X. Joris. “Enhanced Temporal Response Properties of Anteroventral Cochlear Nucleus Neurons to Broadband Noise”. In: *The Journal of Neuroscience* 6.25 (2005), S. 1560–1570 (siehe S. 29).
- [146] Ewan A. Macpherson. “A Computer Model of Binaural Localization for Stereo Imaging Measurement”. In: *J. Audio Eng. Soc* 39.9 (1991), S. 604–622. URL: <http://www.aes.org/e-lib/browse.cfm?elib=5969> (siehe S. 34).
- [147] John J. McDonald und Lawrence M. Ward. “Involuntary Listening Aids Seeing: Evidence From Human Electrophysiology”. In: *Psychological Science, Vancouver, Kanada* 11.2 (2000), S. 167–171. DOI: 10.1111/1467-9280.00233 (siehe S. 70).
- [148] Daniel D. McGall, Richard L. Freyman und Rachel K. Clifton. “Sudden changes in spectrum of an echo cause a breakdown of the precedence effect”. In: *Perception and Psychophysics* 60(4) (1998), S. 593–601 (siehe S. 65).
- [149] Harry McGurk und John MacDonald. “Hearing Lips and Seeing Voices”. In: *Nature* 264 (1976), S. 746–748. DOI: 10.1038/264746a0 (siehe S. 4, 19).
- [150] Jan McKee. *Is It Live or Is It Edison?* Blog der Library of Congress, USA, <https://blogs.loc.gov/now-see-hear/2015/05/is-it-live-or-is-it-edison/>, abgerufen 07.06.2018. 2015 (siehe S. 2).
- [151] Juhaa Merimaa und Wolfgang Hess. “Training of Listeners for Evaluation of Spatial Attributes of Sound”. In: *117th AES Convention, San Francisco, USA*. 2004 (siehe S. 76).

- [152] Donald H. Mershon und L. Edward King. “Intensity and reverberation as factors in the auditory perception of egocentric distance”. In: *Perception and Psychophysics* 6.18 (1975), S. 409–415 (siehe S. 58).
- [153] Henrik Møller. “Fundamentals of binaural technology”. In: *Applied Acoustics* 36.3-4 (1992), S. 171–218 (siehe S. 41, 47).
- [154] Henrik Møller, Michael F. Sørensen, Clemen B. Jensen und Dorte Hammershøi. “Binaural Technique: Do We Need Individual Recordings?” In: *J. Audio Eng. Soc.* 44.6 (1996), S. 451–469 (siehe S. 32, 41, 57, 104, 115, 119).
- [155] Alastair H. Moore, Anthony I. Tew und Rozenn Nicol. “An initial validation of individualised crosstalk cancellation filters for binaural perceptual experiments”. In: *AES 125th Convention, San Francisco, USA*. 2008 (siehe S. 46 f.).
- [156] Annika Neidhardt. “Perception of the reverberation captured in a real room, depending on position and direction”. In: *22nd Int. Congress on Acoustics (ICA), Buenos Aires, Argentinien*. 2016 (siehe S. 152).
- [157] Annika Neidhardt, Florian Klein und Thomas Köllmer. “Flexible python tool for dynamic binaural synthesis applications”. In: *142nd Convention of the Audio Engineering Society (AES), Berlin, Deutschland*. 2017 (siehe S. 136).
- [158] Association NeurOreille. <http://www.cochlea.eu/>. Neufassung von Promenade around the Cochlea, Webseite abgerufen am 08.02.2018. 2018 (siehe S. 24, 26).
- [159] Rozen Nicol. “Binaural Technology”. In: *AES Monograph, Audio Engineering Society* (2010) (siehe S. 41, 48).
- [160] Johannes Nowak. “P4 - Psychoacoustic analysis and spatial coding of microphone array signals”. In: *SEACEN - Simulation und Evaluation akustischer Umgebungen, DFG-Projekt* (2011) (siehe S. 41).
- [161] Johannes Nowak, Judith Liebetrau und Thomas Sporer. “On the perception of apparent source with and listener envelopment in wave field synthesis”. In: *5th International Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Austria*. 2013, S. 83–87 (siehe S. 118).
- [162] Arne Nykänen und Örjan Johannson. “Development of a language for specifying saxophone timbre”. In: *Stockholm Music Acoustics Conference (SMAC), Stockholm, Schweden: SMAC*. 2003 (siehe S. 76, 156, 159).
- [163] Donata Oertel und Eric D. Young. *What is a cerebellar circuit doing in the auditory system?* TRENDS in Neurosciences Vol.27 No.2 February 2004, 2004 (siehe S. 29).
- [164] Djamel Ouis. “Study on the Relationship between Some Room Acoustical Descriptors”. In: *J. Audio Eng. Soc* 51.6 (2003), S. 518–533 (siehe S. 38).

- [165] Renato S. Pellegrini. “Quality assessment of auditory virtual environments”. In: *2001 International Conference on Auditory Displays*. 2001, S. 161–168 (siehe S. 7).
- [166] Jean Piaget. *Meine Theorie der geistigen Entwicklung*. Hrsg. R. Fatke; Deutsche Übersetzung aus dem Englischen. Beltz Taschenbuch / Psychologie, 2016. ISBN: 9783407221421 (siehe S. 16 f., 188).
- [167] Max Planck. *Max Planck - Wissenschaft und Leben, Originaltonaufnahmen 1936-1945*. Herausgegeben von Klaus Sander, Verlag: c und p 2003, Suppose, Köln. 2003 (siehe S. 73).
- [168] Georg Plenge. “Über das Problem der Im-Kopf-Lokalisation”. In: *Acustica* 26.5 (1972), S. 241–252 (siehe S. 32, 59 f., 65 f., 86, 108, 121, 182).
- [169] Georg Plenge. “On the differences between localization and lateralization”. In: *J. Acoust. Soc. Am.* 56 (1974), S. 944–951 (siehe S. 65 f.).
- [170] Christoph Pörschmann und Stephan Wiefling. “Perceptual aspects of dynamic binaural synthesis based on measured omnidirectional room impulse responses”. In: *Int. Conf. on Spatial Audio (ICSA), Graz, Österreich*. 2015 (siehe S. 152).
- [171] Institut für experimentelle Psychologie-Friedrich Heine Universität Düsseldorf. *G\*Power 3*. <http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>. 2013 (siehe S. 90).
- [172] Ville Pulkki. “Spatial sound reproduction with directional audio coding”. In: *J. Audio Eng. Soc.* 55.6 (2007), S. 503–516 (siehe S. 70).
- [173] Alexander Raake. *Speech Quality of VoIP - Assessment and Prediction*. John Wiley und Sons, Chichester, West Sussex, 2006 (siehe S. 14 f., 17).
- [174] Alexander Raake und Jens Blauert. “Comprehensive modelling of the formation process of sound-quality”. In: *5th International Workshop on Quality of Multimedia Experience (QoMEX), Klagenfurt, Österreich*. 2013, S. 76–81 (siehe S. 4, 14–16).
- [175] Lawrence Rabiner und Biing-Hwang Juang. *Fundamentals of Speech Recognition*. Englewood Cliffs, USA :Prentice Hall PTR, 1993, S. 200–238. ISBN: 0130151572 (siehe S. 155).
- [176] Walter Reichard und W. Schmidt. “Die hörbaren Stufen des Raumeindrucks bei Musik”. In: *Acustica* 17 (1966), S. 175–179 (siehe S. 37 f., 139, 143, 145, 149).
- [177] Agnieszka Roginska. “Binaural Audio Through Headphones”. In: *Immersive Sound - The Art and Science of Binaural and Multi-Channel Audio*. Hrsg. von Agnieszka Roginska und Paul Geluso. Routledge, Audio Engineering Society - A Focal Press book, 2018, S. 88–123. ISBN: 9781138900011 (siehe S. 48).



- [178] Agnieszka Roginska und Paul Geluso, Hrsg. *Immersive Sound - The Art and Science of Binaural and Multi-Channel Audio*. Routledge, Audio Engineering Society - A Focal Press book, 2018. ISBN: 9781138900011 (siehe S. 4).
- [179] Nicoleta Roman, DeLiang Wang und Guy J. Brown. "Speech segregation based on sound localization". In: *J. Acoust. Soc. Am.* 4.114 (2003), S. 2236–2252 (siehe S. 34).
- [180] Sebastian Roos. *Implementierung und Evaluation eines Aufnahme- und Wiedergabesystems nach dem Motion-Tracked Binaural Verfahren*. Magisterarbeit, Technische Universität Berlin, Fakultät I, Fachgebiet Audiokommunikation. 2011 (siehe S. 153).
- [181] John J. Rosowski, Laurel H. Carney, T.J. T. J. Lynch und William T. Peake. "The effectiveness of external and middle ears in coupling acoustic power into the cochlea". In: *Allen, J.B., Hall, A., Hubbard, A., Neely, S.T., Tubis, A.: Peripheral Auditory Mechanism*. Springer, Berlin (1986) (siehe S. 23).
- [182] Wayne Rudmose. "The case of the missing 6 dB". In: *Journal Acoustic Society of America* 71(3) (1982) (siehe S. 48).
- [183] Olli Rummukainen, Javier G. Bolanos und Ville Pulkki. "Horizontal localization of auditory and visual events with directional audio coding and 2D video". In: *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)* (2013), S. 94–99 (siehe S. 70).
- [184] Olli Rummukainen und Caterina Mendonca. "Task-relevant spatialized auditory cues enhance attention orientation and peripheral target detection in natural scenes". In: *Journal of Eye Movement Research* 1.9 (2016), S. 01–10 (siehe S. 70).
- [185] Murray B. Sachs. und Eric D. Young. "Effects of nonlinearities on speech encoding in the auditory nerve". In: *J. Acoust. Soc. Amer.* 68 (1980), S. 858–875 (siehe S. 25).
- [186] Markus Schlegel. *Farbe und Wirkung - Architektur*. Skript der Hochschule für angewandte Wissenschaft und Kunst, Fakultät Gestaltung - Farbdesign, HAWK Hildesheim; <http://www.tsp-design.de/>; abgerufen am 21.12.2017. 2017 (siehe S. 69).
- [187] Zora Schärer. *Kompensations von Frequenzgängen im Kontext der Binauraltechnik*. Magisterarbeit, Fachgebiet Audiokommunikationstechnik, Technische Universität Berlin, Deutschland. 2008 (siehe S. 45).
- [188] Zora Schärer und Alexander Lindau. "Evaluation of Equalisation Methods for Binaural Signals". In: *126th AES Conv., München, Deutschland*. 2009 (siehe S. 45, 47, 159).
- [189] Bernhard Seeber. "Untersuchung der auditiven Lokalisation mit einer Lichtzeiger-methode". Diss. Technische Universität München, 2002 (siehe S. 32, 57 f., 74, 76, 96).

- [190] Bernhard Seeber und Hugo Fastl. “On auditory-visual interaction in real and virtual environments”. In: *18th Int. Congress on Acoustics (ICA), Kyoto, Japan*. 2004, S. 2293–2296 (siehe S. 4, 69, 84 f.).
- [191] Bernhard Seeber und Hugo Fastl. “Subjective selection of non-individual head-related transfer functions”. In: *International Conference on Auditory Display, Boston, USA*. 2003 (siehe S. 32, 46).
- [192] Eberhard Sengpiel. *Stereofonie für Lautsprecher und Stereofonie für Kopfhörer 1*. <http://www.sengpielaudio.com/StereoFuerLautsprUndKopfhoerer1.pdf>. 2013 (siehe S. 21).
- [193] Edgar A. G. Shaw. “The external ear”. In: *Keidel, W.D., Neff, W.D.: Handbook of sensory physiology. Vol. V/2, Springer, Berlin* (1974), S. 455–490 (siehe S. 23).
- [194] Barbara Shinn-Cunningham. “Distance Cues for Virtual Auditory Space”. In: *1st IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM), Special Session on Virtual Auditory Space, Sydney, Australien*. 2000, S. 227–230 (siehe S. 39).
- [195] Barbara Shinn-Cunningham. “Identifying where you are in the room: Sensitivity to room acoustics”. In: *International Conference on Auditory Display, Boston, USA*. 2003 (siehe S. 152).
- [196] Andreas Silzle. “Generation of Quality Taxonomies for Auditory Virtual Environments by Means of Systematic Expert Survey”. Shaker Verlag. Diss. Ruhr-Universität Bochum, 2007 (siehe S. 2, 13 f., 17, 32).
- [197] Karl M. Slavik und Stefan Weinzierl. “Wiedergabeverfahren”. In: *Handbuch der Audio-technik*. Hrsg. von Stefan Weinzierl. 9783510343004. Springer, 2008, S. 609–685 (siehe S. 46).
- [198] Stephen Smyth, Mike Smyth und Steve Cheung. “Smyth SVS - Headphone Surround Monitoring For Studios”. In: *Music Everywhere - AES 23rd UK Conference, UK*. 2008 (siehe S. 122).
- [199] Sascha Spors, Hagen Wierstorf, Alexander Raake, Frank Melchior, Matthias Frank und Franz Zotter. “Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State”. In: *Proceedings of the IEEE* 101.9 (2013), S. 1920–1938 (siehe S. 4, 46).
- [200] Herbert Stachowiak. *Allgemeine Modelltheorie*. Buchkopie auf <https://archive.org/>, Abruf 11.12.2017. Springer-Verlag, Wien, New York, 1973. ISBN: 3211811060 (siehe S. 11).
- [201] International Organisation for Standardisation. *Coding of Moving Pictures and Audio, ISO/IEC JTC1/SC29/WG11/w13194*. Draft Call for Proposals for 3D Audio, Shanghai, China. 2012 (siehe S. 5, 57).

- [202] Stanley S. Stevens. *Schall und Gehör*. Deutsche Ausgabe von Rowohlt Taschenbuch Verlag, TIME-LIFE-BOOKS, 1970. ISBN: 3499180154 (siehe S. 2).
- [203] Gero Szepannek, Frank Klefenz und Claus Weihs. “Schallanalyse, Neuronale Repräsentation des Hörvorgangs als Basis”. In: *Informatik-Spektrum* 28.5 (2005), S. 389–395. DOI: 10.1007/s00287-005-0030-3 (siehe S. 25 f.).
- [204] Larry Teal. *The Art of Saxophone Playing*. Alfred Music, 1963. ISBN: 9781457400261 (siehe S. 76, 156, 159).
- [205] Technische Universität Ilmenau - Der Rektor (Peter Scharff). *Promotionsordnung der Technischen Universität Ilmenau - Allgemeine Bestimmungen*. in der Fassung der zehnten Änderung vom 13. Oktober 2017. 2017 (siehe S. iv).
- [206] Günther Theile. “Über die Lokalisation im überlagertem Schallfeld”. Diss. Fachbereich 21 (Umwelttechnik), Technische Universität Berlin, 1980 (siehe S. 12 f.).
- [207] Günther Theile. “Beurteilungskriterien für Kopfhörer unter Berücksichtigung verschiedener Anwendungsbereiche”. In: *NTG Fachberichte - Hörrundfunk* 7 (1986), S. 290–301 (siehe S. 48, 50 f.).
- [208] Günther Theile. “Das Kugelflächenmikrofon”. In: *Bericht 14. Tonmeistertagung des VDT*. 1987, S. 277–293 (siehe S. 41).
- [209] Günther Theile. “On the Standardization of the Frequency Response of High-Quality Studio Headphones”. In: *Journal of the Audio Engineering Society* 34(12) (1986) (siehe S. 48).
- [210] Günther Theile. “Schallpegel im Gehörgang für gleichlaute Schalle auf Kopfhörern bzw. Lautsprechern”. In: *Fortschritte der Akustik, DAGA, Bad Honnef, Deutschland*. 1985 (siehe S. 48).
- [211] Willard R. Thurlow und Theodore E. Parks. “Precedence-suppression effects for two click sources”. In: *Percept. Mot. Skills* 13 (1961), S. 7–12 (siehe S. 64).
- [212] Floyd E. Toole. “In-Head Localization of Acoustic Images”. In: *Journal of the Acoustical Society of America* 48(4) (1970), S. 943–949 (siehe S. 59, 86).
- [213] Floyd E. Toole. “The Acoustics and Psychoacoustics of Headphones”. In: *2nd AES International Conference, Anaheim, USA*. 1984 (siehe S. 6).
- [214] Christos Tsakostas und Jens Blauert. “Some New Experiments On The Precedence Effect”. In: *27. Deutschen Jahrestagung für Akustik, DAGA, Hamburg, Deutschland*. 2001 (siehe S. 65).
- [215] Jesper Udesen, Tobias Piechowiak und Fredrik Gran. “The Effect of Vision on Psychoacoustic Testing with Headphone-Based Virtual Sound”. In: *Journal of the Audio Engineering Society* 63.7/8 (2015), S. 552–561 (siehe S. 62, 66 f.).

- [216] European Broadcasting Union. *EBU Doc. Tech 3276-1998 (second edition): Listening conditions for the assessment of sound program material: monophonic and two-channel stereophonic: Supplement 1: Multichannel Sound*. EBU Doc. Tech 3276-1999. 1998 (siehe S. 52).
- [217] Daniel L. Valente und Jonas Braasch. “Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues”. In: *J. Acoust. Soc. Am.* 128.4 (2010), S. 1952–1964 (siehe S. 71).
- [218] Jules Verne. *Das Karpatenschlos*. Kindle eBook Version. Boston: Pierre-Jules Hetzel, 1892 (siehe S. 1).
- [219] Florian Völk, Andreas Dunstmair, Tobias Riesenweber und Hugo Fastl. “Bedingungen für gleichlaute Schalle aus Kopfhörern und Lautsprechern”. In: *37. Deutschen Jahrestagung für Akustik, DAGA, Düsseldorf, Deutschland*. 2011 (siehe S. 51).
- [220] Michael Vorländer. *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, 2007. ISBN: 3540488294 (siehe S. 31, 41 f., 46 f.).
- [221] Kyriakos Voutsas und Jürgen Adamy. “A Biologically Inspired Spiking Neural Network for Sound Source Lateralization”. In: *IEEE Transactions on Neural Networks* 6.18 (2007), S. 1785–1799 (siehe S. 25, 29).
- [222] Hans Wallach, Edwin B. Newman und Mark R. Rosenzweig. “The precedence effect in sound localization”. In: *The American Journal of Psychology* 62 (1949), S. 315–336 (siehe S. 64, 134).
- [223] DeLiang Wang und Guy J. Brown. *Computational Auditory Scene Analysis - Principles, Algorithms, and Applications*. IEEE Press, John Wiley und Sons, New Jersey, 2006. ISBN: 9780471741091 (siehe S. 33 f.).
- [224] DeLiang Wang und Guy J. Brown. “Separation of Speech from Interfering Sound Based on Oscillatory Correlation”. In: *IEEE Transactions on Neural Networks* 10.3 (1999) (siehe S. 29).
- [225] Stefan Weinzierl. “Aufnahmeverfahren”. In: *Handbuch der Audiotechnik*. Hrsg. von S. Weinzierl. 9783510343004. Springer, 2008, S. 551–607 (siehe S. 24).
- [226] Robert B. Welch und David. H. Warren. “Immediate perceptual response to intersensory discrepancy”. In: *Psychological Bulletin* 88 (1980), S. 638–667 (siehe S. 108).
- [227] Elizabeth M. Wenzel. “Three-Dimensional Virtual Acoustic Displays”. In: *NASA Technical Memorandum 103835* (1991) (siehe S. 95).
- [228] Elizabeth M. Wenzel, Marianne Arruda, Doris J. Kistler und Frederic L. Wightman. “Localization using nonindividualized head-related transfer functions”. In: *J. Acoust. Soc. Am.* 1.94 (1993), S. 111–123 (siehe S. 32, 57, 95).

- [229] Francis M. Wiener und Douglas A. Ross. “The Pressure Distribution in the Auditory Canal in a Progressive Sound Field”. In: *The Journal of the Acoustical Society of America* 18 (1946). <https://doi.org/10.1121/1.1916378>, S. 401–408 (siehe S. 42).
- [230] Frederic L. Wightman und Doris J. Kistler. “Headphone Simulation of Free-field Listening: I Stimulus Synthesis”. In: *J. Acoust. Soc. Am.* 2.85 (1989), S. 858–867 (siehe S. 46).
- [231] Frederic L. Wightman und Doris J. Kistler. “Headphone Simulation of Free-field Listening: II Psychophysical Validation”. In: *J. Acoust. Soc. Am.* 2.85 (1989), S. 868–878 (siehe S. 95).
- [232] Frederic L. Wightman und Doris J. Kistler. “Sound Localization”. In: *Yost, W., A. Popper und R. Fay (Hrsg): Human Psychophysics, Springer Handbook of Auditory Research, Springer, New York* (1993), S. 155–192 (siehe S. 35).
- [233] Yoiti Suzuki (Research Coordinator) et al. *Precise and Full-range Determination of Two-dimensional Equal Loudness Contours*. Techn. Ber. <https://www.mp3-tech.org/programmer/docs/IS-01Y-E.pdf>, Abruf 12.02.2018. Research Institute of Electrical Communication, Tohoku University, Japan, 2003 (siehe S. 37).
- [234] William A. Yost. *Fundamentals of Hearing - An Introduction*. Academic Press, 3. Auflage, San Diego, 1994 (siehe S. 33).
- [235] Nick Zacharov, Torben Pedersen und Chris Pike. “A common lexicon for spatial sound quality assessment - latest developments”. In: *8th International Conference on Quality of Multimedia Experience (QoMEX), Portugal*. 2016. DOI: 10.1109/QoMEX.2016.7498967 (siehe S. 4, 56 f.).
- [236] Pavel Zahorik. “Assessing auditory distance perception using virtual acoustics”. In: *J. Acoust. Soc. Am.* 4.111 (2002), S. 1832–1846 (siehe S. 37 f., 139, 143, 145, 149).
- [237] Pavel Zahorik. “Auditory display of sound source distance”. In: *International Conference on Auditory Display, Kyoto, Japan*. 2002 (siehe S. 38).
- [238] Hans P. Zenner. *Hören-Physiologie, Biochemie, Zell- und Neurobiologie*. Georg Thieme Verlag Stuttgart - New York, 1994. ISBN: 3131390018 (siehe S. 23–26).
- [239] Wen Zhang, Parasanga N. Samarasinghe, Hanchi Chen und Thushara D. Abhayapala. “Surround by Sound: A Review of Spatial Audio Recording and Reproduction”. In: *MDPI Applied Science* 7.532 (2017). doi:10.3390/app7050532 (siehe S. 46).
- [240] Eberhard Zwicker und Hugo Fastl. *Psychoacoustics - Facts and Models*. 2nd Edition. Springer-Verlag Berlin, Heidelberg, New York, 1990. ISBN: 3540650636 (siehe S. 27).
- [241] Eberhard Zwicker und Richard Feldtkeller. *Das Ohr als Nachrichtenempfänger*. Zweite neubearbeitete Auflage. S. Hirzel Verlag Stuttgart, 1967 (siehe S. 27).