Technische Universität Berlin Fakultät I

Institut für Sprache und Kommunikation Fachgebiet Audiokommunikation

Kompensation von Frequenzgängen im Kontext der Binauraltechnik

Magisterarbeit

vorgelegt von

Zora Schärer

Abgabe am

14.10.2008

Erstgutachter Zweitgutachter Betreuer Prof. Dr. Stefan Weinzierl Dr. Hans-Joachim Maempel Alexander Lindau, M.A.

Eidesstattliche Erklärung

Die selbstständige Anfertigung dieser Arbeit versichere ich an Eides statt.

Berlin, den 13. Oktober 2008 _____

Zora Schärer

Zusammenfassung

Die Arbeit ist eingebettet in ein Forschungsprojekt, in dem mittels dynamischer Binauralsynthese die Rezeption von Medienmusik empirisch untersucht wird. Ein Maßstab für die Qualität von binauralen Simulationen akustischer Umgebungen ist ihre perzeptive Ähnlichkeit zum realen Schallfeld. Es ist daher unerlässlich, spektrale Verzerrungen der Auralisation, hervorgerufen durch nicht-ideale Übertrager, zu kompensieren. Vor diesem Hintergrund war das Ziel der Arbeit die Entzerrung der aufnahme- und wiedergabeseitigen Wandler, namentlich der Kunstkopfmikrofone und des Wiedergabekopfhörers.

Die Messung von Übertragungsfunktionen mehrerer typischer Kopfhörermodelle ermöglichte einen Überblick über das Ausmaß vorhandener Nicht-Linearitäten sowie der interindividuellen Unterschiede zwischen den Modellen. Darüber hinaus zeigte sich die Auswirkung variabler Aufsetzpositionen auf den resultierenden Frequenzgang.

Die Invertierung von Übertragungsfunktionen akustischer bzw. elektroakustischer Systeme stellt insbesondere in Bezug auf die Filterkausalität, Robustheit und perzeptive Eignung keine problemlose Aufgabe dar. Zu deren Lösung existieren zahlreiche Ansätze, die sich hinsichtlich Herangehensweise und Aufwand unterscheiden. Beruhend auf dem aktuellen Stand der Forschung wurden einige Verfahren implementiert, um einen systematischen Vergleich ihrer Effizienz durchzuführen. Anhand berechneter Fehlermaße im Zeit- und Frequenzbereich konnte die Güte der Kompensation objektiv bestimmt werden. Eigentlicher Kernpunkt der Evaluation war jedoch ein Hörversuch, der - unter Einbeziehung von zwei für die Binauraltechnik in Frage kommenden Kopfhörern – zur perzeptiven Bewertung der ausgewählten Methoden durchgeführt wurde. Dabei fand ein direkter Vergleich zwischen einer mit variierenden Filtern entzerrten binauralen Simulation und einer realen Schallquelle statt. Es konnte dadurch einerseits eine Aussage über die erzielbare Ähnlichkeit beider Situationen getroffen sowie andererseits eine Rangfolge der untersuchten Verfahren aufgestellt werden. Dies führte zu Erkenntnissen über die Eignung bestimmter Methoden zur perzeptiv günstigen Entzerrung binaural synthetisierter Signale und über die Grenzen, die einer idealen Kompensation bei Kopfhörerwiedergabe noch gesetzt sind.

Danksagung

Prof. Dr. Stefan Weinzierl bin ich sehr dankbar für das mir entgegengebrachte Vertrauen und Interesse sowie für die inhaltliche Unterstützung während der Arbeit.

Dr. Hans-Joachim Maempel möchte ich insbesondere für die hilfreichen Gespräche über methodische Fragen danken.

Mein spezieller Dank gilt Alexander Lindau für die tatkräfige Unterstützung bei praktischen Belangen und die fachliche Anleitung bei allen anderen Fragen.

Dr. Alexander Raake von den Telekom Laboratories und den dortigen Mitarbeitern danke ich für das unkomplizierte Ausleihen mehrerer Kopfhörer, ohne die wichtige Erkenntnisse ausgelieben wären.

Für das sorgfältige Korrekturlesen meiner Arbeit und den fachlichen Input bin ich Holger Kirchhoff und Stefanie Otto äußerst dankbar. Auch Frank Schlutz möchte ich an dieser Stelle für ergiebige Gespräche über inhaltliche Fragen danken.

Selbstverständlich wäre ein wichtiger Teil dieser Arbeit nicht ohne die Teilnehmer des Hörversuchs möglich gewesen. Ihnen gilt mein großer Dank für die aufgewendete Zeit und das genaue Hinhören.

Meinen Freunden und meiner Familie bin ich für die andauernde Unterstützung, die Ablenkung und aufmunternde Gespräche dankbar. Besonders möchte ich mich bei meinen Eltern für die Anteilnahme und Unterstützung während meines Studiums im Allgemeinen und dieser letzten Phase im Speziellen bedanken.

INHALTSVERZEICHNIS

1	Einl	leitung		1		
	1.1	.1 Kontext der Arbeit				
	1.2	2 Auralisation mittels dynamischer Binauralsynthese				
	1.3	Die Ü	bertragungsstrecke im Simulationssystem	6		
2	Mikrofon- und Kopfhörereigenschaften					
	2.1	Mikrofone				
		2.1.1	Technische Spezifikationen	13		
		2.1.2	Messung des Nenn-Übertragungskoeffizienten	14		
	2.2	Kopfh	örer	16		
		2.2.1	Diffus- und Freifeldentzerrung von Kopfhörern	16		
		2.2.2	Messung von Kopfhörerfrequenzgängen	17		
3	Ent	Entzerrung von Frequenzgängen				
	3.1	Grund	llagen	33		
		3.1.1	Direkte Invertierung	33		
		3.1.2	Zielfunktion	34		
		3.1.3	Minimal- und gemischtphasige Invertierung	36		
	3.2	Entzer	rrungsmethoden	41		
		3.2.1	Stand der Forschung	41		
		3.2.2	Implementierung ausgewählter Verfahren	42		
		3.2.3	Zusammenfassung	58		
4	Eva	luation		60		
	4.1	Objektive Evaluation				
		4.1.1	Fehlermaße	61		
		4.1.2	Parameterauswahl	64		
		4.1.3	Kompensationsgüte	78		
		4.1.4	Rechenaufwand	89		

	4.2	Perzep	ptive Evaluation	89		
		4.2.1	Versuchsbeschreibung	90		
		4.2.2	Versuchsvorbereitungen	99		
		4.2.3	Ergebnisse	104		
	4.3 Diskussion					
		4.3.1	Interpretation der Ergebnisse	112		
		4.3.2	Fehlerquellen	114		
		4.3.3	Schlussfolgerungen	115		
5	Zusa	ammen	fassung und Ausblick	117		
Ał	bild	ungsve	rzeichnis	120		
Ta	belle	nverzei	ichnis	122		
Lit	eratu	irverze	ichnis	123		
A	A Quellcodes					
	A.1	Erzeuş	gung der Entzerrungsfilter	Ι		
	A.2	Evalua	ation	XII		
B	Statistische Daten XIX					
	B.1	Allger	nein	XIX		
	B.2	Varian	nzanalyse	XXII		
C	Vers	uchsan	leitung und Personenfragebogen	XIX		
	C.1	Instru	ktion	XXIX		
	C.2	Frage	oogen	XXX		

Abkürzungen und häufig verwendete Variablen

α	Signifikanzniveau
	oder: Cronbach's Reliabilitätskoeffizient
ABC/HR	Triple Stimulus Hidden Reference
ANOVA	Analysis of Variance
b	Parameter zur Bestimmung eines Glättungsfensters
$b(n), B(\omega)$	Regularisierungsfilter-Impulsantwort bzwÜbertragungsfunktion
β	Regularisierungsgewichtung
BRIR	binaurale Raumimpulsantwort
CTC	Cross Talk Cancellation
$d(n)$, $D(\omega)$	Ziel-Impulsantwort bzwÜbertragungsfunktion
e, E _{tot}	Fehlermaß im Zeit- und Frequenzbereich
$E(f_{\rm c})$	frequenzabhängiges Fehlermaß
ERB	Equivalent Rectangular Bandwidth
ETC	Energy Time Curve
FEC	Free Equivalent Coupling
FFD	Fast Frequency Deconvolution
GUI	Graphical User Interface
HATS	Head and Torso Simulator
$h(n)$, $H(\omega)$	Impulsantwort bzw. Übertragungsfunktion
$h_{\rm c}(n), H_{\rm c}(\omega)$	Kompensationsfilter-Impulsantwort bzwÜbertragungsfunktion
$h_{\rm eq}(n)$, $H_{\rm eq}(\omega)$	entzerrte Impulsantwort bzw. Übertragungsfunktion
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
LS	Least Squares
LTI	Linear Time Invariant
N _c	Länge des Kompensationsfilters $h_{\rm c}(n)$
N _h	Länge der Impulsantwort $h(n)$
RAR	reflexionsarmer Raum
S	Stauchungsfaktor bei Compare and Squeeze
p	Ordnung des All-Pol-Modells
	oder: Irrtumswahrscheinlichkeit bei statistischer Auswertung

1 EINLEITUNG

1.1 Kontext der Arbeit

Die vorliegende Arbeit steht im Kontext eines am Fachgebiet Audiokommunikation der TU Berlin durchgeführten Forschungsprojekts im Bereich der Medienrezeption. Darin geht es um den Vergleich von musikalischen Aufführungen in Konzertsälen und medial vermittelter Musik. Der Prozess der Aufnahme, Bearbeitung und Wiedergabe von Musik steht immer dann zwischen einer musikalischen Darbietung und dem Hörer, wenn sie diesen über Medien erreicht, wobei der größte Teil der gehörten Musik heute über Tonträger rezipiert wird. Im Bereich der Musikproduktion hat sich eine Vielzahl von Aufnahme- und Wiedergabeverfahren etabliert.¹ Bezüglich der Klangbildgestaltung lassen sich in der klassischen Musik seit der Mitte des vergangenen Jahrhunderts drei ästhetische Maximen erkennen (Stolla 2004). Welchen Einfluss einzelne Aspekte der medialen Transformation auf die Beurteilung von Musik haben, ist von wesentlichem Interesse für die Rezeptionsforschung. Die Gegenüberstellung unterschiedlich aufgenommener bzw. wiedergegebener Musik und ihrer Darbietung in natürlicher akustischer Umgebung verspricht hierüber grundlegende Einsichten. Im erwähnten Forschungsprojekt soll die dynamische Binauralsynthese zur Simulation eines Konzertsaals sowie mehrerer medialer Wiedergabesituationen genutzt werden, um durch direktes Umschalten einen unmittelbaren Vergleich zwischen den einzelnen Bedingungen zu ermöglichen.

Ein weiteres, in Planung befindliches Projekt befasst sich mit dem Vergleich von Konzertsälen untereinander, um Erkenntnisse über den Einfluss raumakustischer Parameter auf die Wahrnehmung und Beurteilung von Musikdarbietungen zu gewinnen. Auch hier soll die Binauraltechnik als Instrument zur Simulation der untersuchten Räume dienen. Sie stellt also einen grundlegenden Baustein der geschilderten Forschung dar und ist im Zuge ihrer Optimierung selbst Gegenstand laufender Untersuchungen am Fachgebiet. Auf das Prinzip der dynamischen Binauralsynthese und mögliche Herausforderungen bei ihrer Umsetzung wird im folgenden Abschnitt eingegangen.

¹Eine Übersicht findet man z.B. (Slavik u. Weinzierl 2008) und (Weinzierl 2008).

1.2 Auralisation mittels dynamischer Binauralsynthese

Die dynamische Binauralsynthese ist eine kopfbezogene Methode zur Auralisation – dem Hörbarmachen auditiver Umgebungen durch Simulation –, welche den Ansatz verfolgt, die Signale am Ohr des Hörers zu rekonstruieren. Dabei spielen die Prinzipien von linearen und zeitinvarianten (*linear time invariant*, LTI) Systemen sowie die Eigenschaften des räumlichen Hörens eine zentrale Rolle.

Ein LTI-System ist vollständig charakterisiert durch seine so genannte Impulsantwort h(t), also die Reaktion des Systems auf einen unendlich kurzen Diracstoß. Somit lässt sich der Ausgang eines LTI-Systems darstellen als die Faltung eines Eingangssignals x(t) mit seiner Impulsantwort h(t)

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau = x(t) * h(t)$$
(1.1)

Im Frequenzbereich wird dies durch die Multiplikation der Übertragungsfunktion $H(\omega)$ mit dem Spektrum des Eingangssignals $X(\omega)$ ausgedrückt.

$$Y(\omega) = X(\omega) \cdot H(\omega) \tag{1.2}$$

Kennt man die Impulsantwort bzw. Übertragungsfunktion eines Systems, so können seine Eigenschaften simuliert werden, indem ein beliebiges Eingangssignal mit h(t) gefaltet wird. Da Linearität und Zeitinvarianz grundsätzlich auch für Räume angenommen werden können, ist es möglich, mit Hilfe von (künstlichen oder gemessenen) Raumimpulsantworten akustische Umgebungen zu synthetisieren. Dies ist beispielsweise zur Erzeugung von künstlichem Nachhall in der Tonstudiotechnik gängige Praxis.

Bei der Lokalisation eines Schallereignisses spielen bei seitlichem Schalleinfall vorwiegend interaurale Pegel- und Laufzeitdifferenzen, die durch die Abschattungswirkung des Kopfes zustande kommen, eine Rolle. Auch Kopfbewegungen relativ zur Quelle liefern wichtige Informationen über deren Lage. Zur Orientierung dient hierbei das kopfbezogene Polarkoordinatensystem (Abb. 1.1), dessen Ursprungspunkt sich in der Mitte zwischen beiden Gehörkanaleingängen auf der interauralen Achse befindet. Die Lage einer Quelle sowie die Auslenkungen des Kopfes sind durch das Azimut φ , die Elevation δ und die Entfernung *r* bezüglich des Ursprungs gekennzeichnet (vgl. Blauert u. Braasch 2008).

Betrachtet man den Vorgang des räumlichen Hörens als LTI-System, kann die Schallübertragung von einer Quelle zum Trommelfell eines Hörers durch die so genannte Außenohr-



Abbildung 1.1 – Kopfbezogenes Polarkoordinatensystem, aufgespannt von Horizontal-, Frontal- und Medianebene (aus Blauert u. Braasch 2008)

Übertragungsfunktion oder *head related transfer function* (HRTF)² beschrieben werden. Sie enthält neben spektralen Eigenschaften auch alle Informationen über Entfernung und relative Lage der Quelle zum Kopf des Hörers, ist also abhängig von ω , φ , δ und r. Ihre charakteristische Klangfärbung erhält sie durch Reflexionen, Abschattung und Beugung an Torso, Schultern, Kopf, Ohrmuschel und Ohrkanaleingang sowie durch Resonanzen im Ohrkanal (vgl. Weinzierl 2008, S. 586).

Außenohr-Übertragungsfunktionen werden in reflexionsarmer Umgebung bestimmt. Sind hingegen auch Raumeinflüsse vorhanden, spricht man von binauralen Raumimpulsantworten (BRIRs). Die Messung von HRTFs und BRIRs erfolgt entweder mit Hilfe von Sondenmikrofonen am oder im Ohrkanal eines Hörers³ oder durch den Nachbau eines menschlichen Torsos und Kopfes — *head and torso simulator* (HATS) – mit eingebauten Mikrofonen. In Abb. 1.2 ist der am Fachgebiet entwickelte Messroboter FABIAN dargestellt, der aus dem individuellen Abguss eines Kopfes und einem auf Grundlage antropometrischer Daten konstruierten Torso besteht (Lindau 2006).

Durch ein zusammengehörendes Paar von HRTFs bzw. BRIRs können die oben erwähnten interauralen Differenzen beschrieben werden, indem die Signale für linkes und rechtes Ohr in Beziehung zueinander gesetzt werden. Faltet man ein Paar von binauralen Raumimpulsantworten mit einem nachhallfreien Eingangssignal, simulieren die binauralen Ausgangssignale die akustische Umgebung des ursprünglichen Aufnahmeortes. Die Wiedergabe derartiger Signale erfolgt üblicherweise über Kopfhörer, da sie dadurch direkt an ihren "Bestimmungsort", das Ohr, gelangen. Die sonst bei Kopfhörerwiedergabe übliche Im-Kopf-Lokalisation entfällt hier, genau wie es auch bei herkömmlichen Kunstkopfaufnahmen der Fall ist. Stattdessen ist

²Für den Zeitbereich wird der Begriff *head related impulse response* (HRIR) verwendet, allerdings ist die Bezeichnung im Frequenzbereich geläufiger.

³s. z.B. (Møller et al. 1995c)



Abbildung 1.2 – Messroboter FABIAN, bestehend aus einem Torso, um Schallreflexionen an Schultern und Oberkörper in HRTF-/BRIR-Messungen miteinzubeziehen und einem rotier- und neigbaren Kopf mit in den Ohren eingesetzten Mikrofonen.

infolge der Einflüsse der HRTFs und ihrer interauralen Differenzen eine Ortung der Quelle außerhalb des Kopfes möglich (Externalisierung).

Im bisher beschriebenen Fall ist lediglich die Kopfposition, nicht aber die Kopfbewegung berücksichtigt. Tatsächlich dominiert jedoch die "dynamische propriozeptive und auditive Information, die das Gehör bei bewusst durchgeführten Peilbewegungen erhält, [...] bei der Bildung der Hörereignisorte in der Regel über statisch empfangene [Information]." (Blauert u. Braasch 2008, S. 88) Bei Schallereignissen auf der Medianebene entstehen keine interauralen Laufzeit- und Pegeldifferenzen, sodass es hier im Falle statischer binauraler Simulationen häufig zu Vorne-Hinten-Vertauschungen kommt. Um solche Artefakte zu vermeiden, sollte bei der Binauralsynthese eine dynamische Anpassung der in die Faltung eingehenden Impulsantworten an die Kopfbewegungen des Hörers gewährleistet werden. Karamustafaouglu et al. (1999) verglichen die Lokalisierbarkeit von Schallquellen in statischen und dynamischen Simulationen und stellten fest, dass im Falle nachgeführter BRIRs die Doppeldeutigkeit der Schallereignisse von vorne bzw. hinten praktisch aufgehoben wird. Die laufende Anpassung der binauralen Impulsantworten setzt einerseits einen kompletten Datensatz von BRIRs für verschiedene Kopfpositionen und andererseits das Erfassen der jeweils aktuellen Kopfposition mittels eines so genannten head trackers voraus. Je nach Auflösungsgenauigkeit und Größe des Kopfbewegungsbereichs, die bei der Messung eines BRIR-Datensatzes gewählt werden, können sich über 10.000 Impulsantwortpaare ergeben.⁴ Da die Länge von

⁴Wird beispielsweise ein Bereich von $\pm 75^{\circ}$ Azimut und $\pm 45^{\circ}$ Elevation in jeweils 1° Auflösung vermessen, resultieren 13.741 BRIR-Paare. Dies ist jedoch ein Extremfall, da in der Regel entweder kleinere Bereiche abgedeckt oder geringere Auflösungen gewählt werden. Lindau et al. (2008) ermittelten in Hörversuchen eine Auflösung von 2°x 1° (horizontal x vertikal) für Rauschen und 3°x 2° für ein Gitarrenstimulus als ausreichend genau, um keine hörbaren Artefakte hervorzurufen.



Abbildung 1.3 – System zur dynamischen Binauralsynthese. Links im Bild ist die Aufnahmesituation skizziert, der mittlere Teil kennzeichnet die Faltung eines nachhallfreien Signals mit einem BRIR-Paar, welches durch Befehle des head trackers (rechts im Bild) aus dem gemessenen Datensatz ausgewählt wird. Der rechte Teil der Abbildung zeigt die Kopfhörerwiedergabe.

BRIRs außerdem von der Nachhallzeit des Aufnahmeraumes bestimmt ist, führt dies insgesamt zu einer sehr großen Datenmenge und einem entsprechend hohen Rechenaufwand bei der Faltung. In Abb. 1.3 sind die einzelnen Komponenten der dynamischen Binauralsynthese schematisch dargestellt.

Die Anforderungen, welche ein solches System zu erfüllen hat, beinhalten folgende Punkte (vgl. Slavik u. Weinzierl 2008, S. 672):

- automatische Messung von BRIRs f
 ür verschiedene Kopfpositionen in mindestens zwei Freiheitsgraden (Elevation und Azimut)
- Ausgleich der vorhandenen spektralen Verzerrungen durch elektroakustische Wandler
- möglichst latenzfreie Echtzeitfaltung des durch den head tracker bestimmten BRIR-Paares mit einem nachhallfreien Signal
- möglichst latenzfreier Wechsel zwischen Impulsantworten bei Kopfbewegungen

Im Folgenden soll das am Fachgebiet eingesetzte System vorgestellt werden. Zur rechnergesteuerten Akquise eines kompletten BRIR-Datensatzes kommt der bereits erwähnte Messroboter FABIAN zum Einsatz. Anschließend werden diese in einem *post-processing* von unerwünschten spektralen Einflüssen befreit sowie auf eine Länge von ca. 370 ms gekürzt. Der zweite Schritt beruht auf Ergebnissen von Hörversuchen, laut denen nur dieser anfängliche Ausschnitt der Impulsantworten dynamisch nachgeführt werden muss, während der diffuse Anteil keine richtungsrelevanten Informationen mehr enthält und somit unverändert aus einer einzigen Impulsantwort gewonnen werden kann. Dies ermöglicht eine deutliche Reduktion der Datenmenge. Die Echtzeitfaltung erfolgt durch Multiplikation im Frequenzbereich (schnelle Faltung), wobei jeweils mehrere Impulsantworten um die aktuelle Kopfposition herum in einen dynamischen Cache geladen werden. Ermittelt der Positionssensor eine zu große Entfernung von diesem Punkt, wird ein Teil des Speichers entleert und mit aktuelleren Impulsantworten versehen. Der Wechsel zwischen zwei BRIRs erfolgt durch Überblenden, um Klick-Artefakte zu vermeiden.⁵

Lindau und Weinzierl (2007) evaluierten in einem Hörversuch die Plausibilität des beschriebenen Systems. Eine reale Schallquelle und deren Simulation wurden direkt miteinander verglichen und sollten von den Versuchspersonen richtig benannt werden. Insgesamt lag die Erkennungsrate nur geringfügig über der Ratewahrscheinlichkeit von 50%, diese Abweichung war jedoch statistisch signifikant. Aus einer Befragung der Versuchsteilnehmer ging hervor, dass sich die wahrgenommenen Unterschiede zwischen Realität und Simulation insbesondere in der Klangfarbe manifestieren. Dies deckt sich auch mit den Erkenntnissen einer ähnlichen Untersuchung von Moldrzyk et al. (2005) und führt zur Notwendigkeit, die Entzerrung der involvierten elektroakustischen Wandler zu optimieren.

1.3 Die Übertragungsstrecke im Simulationssystem

In diesem Abschnitt wird näher auf die in Abb. 1.3 dargestellte Übertragungsstrecke eingegangen (nach Møller 1992; Lindau 2006), wobei von der BRIR-Akquise mittels eines Kunstkopfes und der Wiedergabe über Kopfhörer ausgegangen wird. Zunächst wird die Schallübertragung im Außenohr beschrieben, wie sie beim Hören unter Freifeldbedingungen sowie bei Kopfhörerübertragung vorkommt. Mit Hilfe der daraus resultierenden Ausdrücke werden die spektralen Einflüsse auf das Simulationssystem analysiert und ein Filter zu ihrer Kompensation beschrieben.

Abb. 1.4 skizziert die Schallübertragung im Außenohr bei Anregung im Freifeld. Im linken Bild ist das Außenohr dargestellt, darin eingezeichnet sind die bei auftreffender Schallwelle herrschenden Schalldrücke und komplexen Impedanzen.⁶ Am Eingang zum offenen Ohrkanal treten der Schalldruck p_3 und die Impedanz $Z_{ear canal}$ auf. Dieser Punkt ist durch den Ohrkanal (*transmission line*) verbunden mit dem Trommelfell, wo der Schalldruck p_4 und die Impedanz $Z_{eardrum}$ vorzufinden sind. $Z_{radiation}$ kennzeichnet die Abstrahlungsimpedanz, auf welche am Trommelfell reflektierte, rücklaufende Schallwellen treffen. Sie entspricht der Schallkennimpedanz der Luft.

⁵Für eine ausführlichere Beschreibung des Systems siehe (Lindau et al. 2007)

⁶Die Großbuchstaben bezeichnen komplexe Spektren, Kleinbuchstaben werden nachfolgend für die entsprechenden Zeitfunktionen verwendet.

Im Ersatzschaltbild in Abb. 1.4(b) wird ein zusätzlicher Schalldruck, p_2 , eingeführt, den Møller in Analogie zur Elektrotechnik als "Leerlauf"-Schalldruck bezeichnet. Dieser tritt an Stelle von p_3 auf, wenn kein Schall auf $Z_{radiation}$ trifft, wie es im Fall eines verschlossenen Ohrkanals gegeben ist. p_2 bildet nach dem Thévenin-Theorem zusammen mit $Z_{radiation}$ eine "Ersatzspannungsquelle" für die Anregung der am Ohr auftreffenden Schallwelle. Daraus lässt sich die folgende Gleichung ableiten, mit der die Übertragung vom geblockten zum offenen Ohrkanal ausgedrückt wird:

$$\frac{P_3}{P_2} = \frac{Z_{\text{ear canal}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}$$
(1.3)



Abbildung 1.4 – Schallübertragung im Außenohr im freien Schallfeld: (a) anatomische Skizze, (b) elektroakustisches Modell (aus Møller 1992)

Um Übertragungsfunktionen für die beschriebenen Punkte im Außenohr aufstellen zu können, wird der Referenzschalldruck p_1 eingeführt, den Møller als "Schalldruck in der Mitte des Kopfes bei Abwesenheit des Hörers" bezeichnet. Der Ursprung des kopfbezogenen Polarkoordinatensystems (s. Abb. 1.1) definiert diesen Referenzpunkt.

$$\frac{P_4}{P_1} = \frac{\text{Schalldruck am Trommelfell}}{\text{Schalldruck am Referenzpunkt bei abwesendem Hörer}}$$
(1.4)

$$P_3$$
 _ _ _ Schalldruck am offenen Ohrkanal (1.5)

$$P_1$$
 Schalldruck am Referenzpunkt bei abwesendem Hörer
 P_2 Schalldruck am geblockten Ohrkanal

$$\frac{P_1}{P_1} = \frac{P_1}{\text{Schalldruck am Referenzpunkt bei abwesendem Hörer}}.$$
 (1.6)

Die als HRTF definierte Gl. 1.4 kann in einzelne Teile zerlegt werden:

$$\frac{P_4}{P_1} = \frac{P_4}{P_3} \cdot \frac{P_3}{P_2} \cdot \frac{P_2}{P_1}$$
(1.7)

Um bei der BRIR-Akquise den konstruktiven und messtechnischen Aufwand einer Schalldruckbestimmung am Trommelfell zu umgehen, bevorzugt man eine Messung am geblockten Ohrkanal.⁷ Hier sind bereits alle wichtigen Richtungsinformationen der Schallübertragung vorhanden.⁸ Die Übertragungsfunktion am verschlossenen Ohrkanal wird durch den letzten Term in Gl. 1.7 beschrieben. Inwiefern die zwei ersten Terme zur Schallfeldreproduktion am Trommelfell berücksichtigt werden müssen, wird sich weiter unten zeigen.



Abbildung 1.5 – Schallübertragung im Außenohr bei Kopfhörerwiedergabe: (a) anatomische Skizze, (b) elektroakustisches Modell (aus Møller 1992)

Vorerst wird analog zur oben geschilderten Situation im freien Schallfeld die Schallausbreitung im Außenohr bei Kopfhörerwiedergabe erläutert. Die Schalldrücke an Ohrkanal und Trommelfell werden in Abb. 1.5 nun mit p_6 bzw. p_7 bezeichnet, die Impedanzen an diesen Punkten sind dieselben wie im Falle des Freifelds. Statt der Schallkennimpedanz der Luft herrscht jedoch vom Trommelfell her betrachtet am Ohrkanaleingang die Kopfhörerimpedanz $Z_{\text{headphone}}$. In Abb. 1.5(b) ist mit P_5 und $Z_{\text{headphone}}$ eine "Ersatzspannungsquelle" für die Anregung über Kopfhörer dargestellt. p_5 repräsentiert dabei ebenso wie p_2 den Schalldruck am geblockten Ohrkanal.

Aus Abb. 1.5 können den Gln. 1.3-1.7 entsprechende Ausdrücke abgeleitet werden. So gilt für die Übertragung vom geblockten zum offenen Ohrkanal:

$$\frac{P_6}{P_5} = \frac{Z_{\text{ear canal}}}{Z_{\text{ear canal}} + Z_{\text{headphone}}}$$
(1.8)

Die Verhältnisse der Schalldrücke zur Kopfhörereingangsspannung E_{headphone} ergeben für un-

⁷Die weiteren Vorteile fasst (Lindau 2006, S. 51) zusammen.

⁸s. hierzu (Hammershøi u. Møller 1996)

terschiedliche Messpunkte definierte Kopfhörerübertragungsfunktionen.

$$\frac{P_7}{E_{\text{headphone}}} = \frac{\text{Schalldruck am Trommelfell}}{\text{Eingangsspannung des Kopfhörers}}$$
(1.9)

$$\frac{P_6}{E_{\text{headphone}}} = \frac{\text{Schalldruck am offenen Ohrkanal}}{\text{Eingangsspannung des Kopfhörers}}$$
(1.10)

$$\frac{P_5}{E_{\text{headphone}}} = \frac{\text{Schalldruck am geblockten Ohrkanal}}{\text{Eingangsspannung des Kopfhörers}}$$
(1.11)

Die Übertragungsfunktion zum Trommelfell wird aufgespalten in

$$\frac{P_7}{E_{\text{headphone}}} = \frac{P_7}{P_6} \cdot \frac{P_6}{P_5} \cdot \frac{P_5}{E_{\text{headphone}}}$$
(1.12)

Beim Vergleich von Abb. 1.4 und Abb. 1.5 zeigt sich, dass

$$\frac{P_7}{P_6} = \frac{P_4}{P_3}.$$
(1.13)

Zudem ergibt sich ein wichtiger Ausdruck, wenn Gln. 1.3 und 1.8 verbunden werden:

$$\frac{P_3/P_2}{P_6/P_5} = \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}$$
(1.14)

Diese Gleichung resultiert dann zu 1, wenn die Schallkennimpedanz der Luft und die Kopfhörerimpedanz gleich, oder wenn beide klein im Vergleich zur Impedanz des Ohrkanals sind. Møller zeigte, dass die zweite Voraussetzung bei Frequenzen unterhalb von 1 kHz gegeben ist, oberhalb dieser Grenze muss die erste Bedingung erfüllt sein. Kopfhörer mit entsprechenden Eigenschaften werden *free equivalent coupling*⁹ (FEC) Kopfhörer genannt. Møller et al. (1995a) untersuchten diesbezüglich 13 Kopfhörer, wobei das Kriterium $\frac{P_3/P_2}{P_6/P_5} = 1$ lediglich von einem selbst gebauten Modell erfüllt wurde, bei dem zwei kleine Lautsprecher in runden Gehäusen in ca. 25 cm Abstand von der Kopfmitte vor den Ohren platziert wurden. Bei tolerierten Abweichungen von ± 2 dB, konnten nur wenige der kommerziellen Prüflinge als FEC-Kopfhörer bezeichnet werden, dazu gehörte ein Stax Lambda-Kopfhörer¹⁰ sowie das Modell AKG K-1000 (s. auch Abschnitt 2.2).

In Abb. 1.6 ist eine Vereinfachung des in Abschnitt 1.2 beschriebenen Systems zur Binauralsynthese gezeigt. $H_{c_{tot}}$ steht für ein Filter, das die spektrale Beeinflussung der elektroakusti-

⁹Gelegentlich wird auch die Bezeichnung "offen" benutzt. Bei Møller et al. (1995a) wird jedoch vorgeschlagen, diesen Begriff im vorliegenden Zusammenhang zu vermeiden, um eine klare Abgrenzung vom kommerziellen Bereich zu schaffen. Dort bedeute der Begriff lediglich, dass Schall von Außen nicht abgeschirmt wird.

¹⁰Zwei Kopfhörer dieses Typs wurden für den in Abschnitt 4.2 beschriebenen Hörversuch eingesetzt.

schen Wandler ausgleicht und im Folgenden bestimmt werden soll. Zur Vereinfachung wird darauf verzichtet, die elektrischen Übertragungswege, bestehend aus den Verstärkern und A/D-D/A-Wandlern, welche bei Messung und Wiedergabe zum Einsatz kommen, in die Betrachtung miteinzubeziehen, da hier von einer ausreichenden Linearität ausgegangen werden kann.



Abbildung 1.6 – Elektroakustische Übertragung im Simulationssystem. Links ist die BRIR-Akquise mittels Kunstkopf dargestellt, rechts die Wiedergabe über Kopfhörer. Das Filter $H_{c_{tot}}$ dient zum Ausgleich spektraler Verzerrungen.

Vorab sei angemerkt, dass auch das hier nicht skizzierte Abhören über Lautsprecher, die so genannte transaurale Wiedergabe, denkbar wäre. Dabei muss zur korrekten Schallfeldreproduktion am Trommelfell der hinzukommende Signalweg durch den Abhörraum und die Ohren des Hörers berücksichtigt werden. Ferner ist das Übersprechen von den Lautsprechern auf das jeweils gegenüber liegende Ohr zu unterdrücken. Diese so genannte *cross talk cancellation* (CTC) funktioniert nur in einem sehr eingeschränkten *sweet spot* (Møller 1992), was jedoch durch dynamische Nachführung der CTC-Filter überwunden werden kann (Lentz 2006). Eine weitere Methode der transauralen Wiedergabe ist der Binaural Sky. Hier wird mittels Wellenfeldsynthese ein Kopfhörer durch zwei vor den Ohren des Hörers erzeugte Punktquellen simuliert. Diese passen sich der Kopfposition an, die CTC hingegen bleibt konstant (Menzel et al. 2006). Im weiteren Verlauf dieser Arbeit wird nicht mehr auf die transaurale Wiedergabe eingegangen.

Die in Abb. 1.6 dargestellte Signalübertragung von einer Quelle bis zum Trommelfell eines Hörers mit dazwischen liegender BRIR-Messung am geblockten Ohrkanal und Wiedergabe über Kopfhörer kann durch folgende Gleichung ausgedrückt werden:

$$\frac{P_4}{P_1} = H_{\text{LS}}(\omega, \varphi, \delta) \cdot \frac{P_2}{P_1} \cdot M_{\text{KK}}(\omega) \cdot H_{\text{c_tot}}(\omega) \cdot \frac{P_7}{E_{\text{headphone}}}$$
(1.15)

 $M_{\rm KK}(\omega)$ bezeichnet den frequenzabhängigen Übertragungsfaktor¹¹ der Mikrophone, die bei der BRIR-Akquise eingesetzt werden, $H_{\rm LS}(\omega, \varphi, \delta)$ steht für die frequenz- und richtungsabhängige Übertragungsfunktion des Quelllautsprechers. Gl. 1.15 kann nach $H_{\rm c_tot}(\omega)$ umgestellt und mit Hilfe von Gln. 1.7 und 1.12 erweitert werden:

$$H_{c_tot}(\omega) = \frac{P_4/P_1}{H_{LS}(\omega, \varphi, \delta) \cdot [P_2/P_1] \cdot M_{KK}(\omega) \cdot [P_7/E_{headphone}]}$$

= $\frac{[P_4/P_3] \cdot [P_3/P_2] \cdot [P_2/P_1]}{H_{LS}(\omega, \varphi, \delta) \cdot [P_2/P_1] \cdot M_{KK}(\omega) \cdot [P_7/P_6] \cdot [P_6/P_5] \cdot [P_5/E_{headphone}]}$
= $\frac{[P_4/P_3]}{[P_7/P_6]} \cdot \frac{[P_3/P_2]}{[P_6/P_5]} \cdot \frac{1}{H_{LS}(\omega, \varphi, \delta) \cdot M_{KK}(\omega) \cdot [P_5/E_{headphone}]}$ (1.16)

Der erste Term des obigen Ausdrucks ergibt nach Gl. 1.13 den Wert 1 und kann deshalb vernachlässigt werden. Dies ist unter der Voraussetzung, dass bei der Wiedergabe binaural simulierter Quellen FEC-Kopfhörer eingesetzt werden, mit dem zweiten Term ebenfalls möglich. Gl. 1.16 kann dadurch reduziert werden zu:

$$H_{c_tot}(\omega) = \underbrace{\frac{1}{\underbrace{H_{LS}(\omega, \varphi, \delta) \cdot M_{KK}(\omega)}_{\text{Aufnahmeseite}} \cdot \underbrace{[P_5 / E_{\text{headphone}}]}_{\text{Wiedergabeseite}}} (1.17)$$

Es zeigt sich also, dass die nötige Entzerrung aufnahmeseitig den Quelllautsprecher und die bei der BRIR-Messung benutzten Mikrofone sowie wiedergabeseitig die Kopfhörerübertragungsfunktion betrifft.

Den Einfluss des Quelllautsprechers zu eliminieren, ist eine Thematik, die den Rahmen der vorliegenden Arbeit sprengen würde. An dieser Stelle wird kurz auf die verschiedenen Probleme und möglichen Lösungsansätze eingegangen. Einerseits stellt sich die Frage, ob eine Freifeld- oder eine Diffusfeldentzerrung stattfinden soll. Erstere kann bei HRTFs angewendet werden. Lindau (2006, S. 62) weist allerdings darauf hin, dass in diesem Fall die Ausrichtung zur Empfängerposition exakt reproduzierbar sein muss, da $H_{LS}(\omega, \varphi, \delta)$ wie bereits erwähnt richtungsabhängig ist (Lindau 2006, S. 62). Bei BRIR-Messungen, wo raumakustische Informationen ja explizit mit erfasst werden sollen, befindet sich der Kunstkopf in der Regel nicht im Hallradius der Quelle, sodass die Diffusfeldkompensation sinnvoller erscheint. Hier wird jedoch nur eine angenäherte Entzerrung erreicht, da zur Bestimmung des Diffusfeldfrequenzgangs über Leistungsspektren aus mehreren Einfallsrichtungen gemittelt wird. Es entfällt dadurch auch die Phaseninformation, deshalb werden meist minimalphasige Entzerrungsfilter

¹¹s. Abschnitt 2.1 für eine Begriffsklärung

eingesetzt.¹² Eine andere Möglichkeit zum Ausgleich von $H_{LS}(\omega, \varphi, \delta)$ ist die Nachbildung der frequenzabhängigen Richtcharakteristik der zu simulierenden Quelle. Dadurch wird nicht nur eine spektrale Verzerrung vermieden, sondern auch ein natürliches Abstrahlverhalten erreicht. Behler u. Pollow (2008) schlagen hierfür einen Dodekaeder-Lautsprecher mit unabhängig ansteuerbaren Einzelsystemen vor, womit auf Grundlage von sphärischen Harmonischen variable Richtcharakteristiken erzeugt werden können. Die Nachbildung des Abstrahlverhaltens beliebiger Quellen wird in diesem Zusammenhang auch am hiesigen Fachgebiet angestrebt.

Die Kompensation der Mikrofon- und Kopfhörerfrequenzgänge ist in einem Zuge möglich, wie durch die folgende Ausführung erläutert werden soll. Bei der Messung der Kopfhörerübertragungsfunktion muss der Schalldruck p_5 mit Hilfe von Mikrofonen bestimmt werden. Wird hierfür der bei der BRIR-Akquise eingesetzte Kunstkopf (mit gleichen Mikrofonen) verwendet, kann mit $M_{\text{KK}}(\omega) = \frac{E_{\text{microphone}}}{p_5}$ und Gl. 1.17 das Filter $H_c(\omega)$ definiert werden, welches sowohl die Kopfhörer- als auch und die Mikrofonübertragungsfunktion kompensiert.

$$H_{c}(\omega) = \frac{1}{M_{KK}(\omega) \cdot [(E_{microphone}/M_{KK}(\omega))/E_{headphone}]}$$
$$= \frac{E_{headphone} \cdot M_{KK}(\omega)}{E_{microphone} \cdot M_{KK}(\omega)}$$
$$= \frac{E_{headphone}}{E_{microphone}}$$
(1.18)

 $E_{\text{microphone}}$ bezeichnet hier die Ausgangsspannung des Mikrofons. $H_{\text{c}}(\omega)$ kann nicht direkt gemessen werden, stattdessen wird die Funktion

$$H(\omega) = \frac{E_{\text{microphone}}}{E_{\text{headphone}}}$$
(1.19)

bestimmt¹³ und anschließend invertiert. Genau wie die Lautsprecherentzerrung stellt auch die Kompensation von $H(\omega)$ kein triviales Problem dar, wie in Kapitel 3 ausführlich dargelegt wird. Es existieren zahlreiche Verfahren zum Entwurf von geeigneten Entzerrungsfiltern, die sich in Herangehensweise und Aufwand unterscheiden. Die Implementierung verschiedener Methoden und deren empirische Evaluation mit Hilfe eines Hörversuchs ist Gegenstand dieser Arbeit.

¹²s. (Møller 1992) und (Müller u. Massarani 2001)

¹³Ein entsprechender Messaufbau ist in Abschnitt 2.2 beschrieben.

2 MIKROFON- UND KOPFHÖREREIGENSCHAFTEN

2.1 Mikrofone

2.1.1 Technische Spezifikationen

Im Messroboter FABIAN sind zwei Elektretkondensatormikrofone (DPA 4060) mit omnidirektionaler Richtcharakteristik eingebaut. Die sehr kleinen Abmessungen der Kapsel (ca. 5,4 x 12,7 cm) erlauben eine Platzierung in der Mitte der Ohrmuschel an der Stelle des Ohrkanaleingangs, wobei die Kapsel bündig in die umgebende Fläche eingelassen ist. Wie in Abschnitt 1.3 ausgeführt, finden Messungen mit FABIAN also am geblockten Ohrkanal statt. Die technischen Spezifikationen der Mikrofone sind in Tabelle 2.1 aufgelistet.

Frequenzbereich	20 Hz - 20 kHz (±2 dB), mit 3 dB Überhöhung bei 8 - 20 kHz			
Sensitivity	20 mV/Pa; -34 dB (±3 dB) bzgl. 1 V/Pa			
Eigenrauschen	23 dB(A) bzgl. 20 µPa			
THD	<1% THD bis 123 dB SPL peak			
SNR	71 dB(A) bei 1 kHz bzgl. 94 dB SPL			
Dynamic Range	100 dB			

Tabelle 2.1 – Technische Spezifikationen der Messmikrofone im Messroboter FABIAN (Angaben laut Hersteller: DPA Microphones)

In Abb. 2.1 ist das frequenzabhängige Übertragungsmaß¹ der Mikrofone zwischen 10 Hz und 40 kHz dargestellt. Deutlich sichtbar ist die auch vom Hersteller angegebene Überhöhung bei ca. 8 kHz. Der sonst sehr lineare Frequenzgang fällt als spektrale Verzerrung im Vergleich zu den Kopfhörern wenig ins Gewicht (s. Abschnitt 2.2). Wegen des Anstiegs im hohen Frequenzbereich muss er jedoch trotzdem ausgeglichen werden, um eine korrekte Reproduktion der Ohrsignale zu gewährleisten. Der steile Abfall oberhalb von 20 kHz zeigt zudem, dass sich die Mikrofone wenig zur Aufnahme von Signalen mit Anteilen über dieser Grenze eignen, da hier der nötige Aufwand zur Kompensation sehr groß wäre.

¹s. den folgenden Abschnitt



Abbildung 2.1 – Übertragungsmaß der DPA 4060 Messmikrofone von 10 bis 40 kHz (bezogen auf 1V/Pa)

2.1.2 Messung des Nenn-Übertragungskoeffizienten

Wie in Abschnitt 1.2 ausgeführt, tragen interaurale Differenzen zur Schallquellenlokalisation bei. Im Zusammenhang der Binauraltechnik muss also ausgeschlossen sein, dass Ungleichheiten der zusammengehörenden elektromagnetischen Wandler die Wahrnehmung eines Hörereignisortes beeinflusst. Diesbezüglich sei die Empfindlichkeit der Messmikrofone genannt, die im Falle eines zu großen Unterschieds zu unnatürlichen interauralen Pegeldifferenzen führen könnte. Um deren Paarigkeit bei den Mikrofonen im Messroboter FABIAN zu überprüfen, wurde für beide der Nenn-Übertragungskoeffizient gemessen und verglichen. Auch Abweichungen der Phasengänge sind für die Paarigkeit ein wichtiger Aspekt, sie wurden in diesem Rahmen jedoch nicht überprüft (s. hierzu Lindau 2006, S. 100).

Laut DIN EN 60268-4 ist der Übertragungskoeffizient (oder -faktor) definiert als Verhältnis von Ausgangsspannung U_{eff} eines Mikrofons und eintreffendem Schalldruck p_{eff} .

$$M(\omega) = \frac{U_{\text{eff}}}{p_{\text{eff}}} \qquad \left[\frac{V}{Pa}\right]$$
(2.1)

 $M(\omega)$ gibt also an, welche Spannung ein Mikrofon bei 1 Pa Schalldruck liefert. In Abb. 2.1 ist das so genannte Übertragungsmaß $L_{\rm M}$ dargestellt, welches wie folgt definiert ist:

$$L_{\rm M} = 20 \log \frac{M(\omega)}{M_{\rm r}}, \qquad {\rm mit} \quad M_{\rm r} = 1 \, {\rm V/Pa} \tag{2.2}$$

 $M(\omega)$ wird als Freifeld-Übertragungskoeffizient bezeichnet, wenn er "auf den Schalldruck im ungestörten Schallfeld (in Abwesenheit des Mikrofons)" bezogen ist (DIN EN 60268-4, S. 16). Demgegenüber steht der Diffusfeld-Übertragungskoeffizient, der im diffusen Schallfeld ermittelt wird. Vom Nenn-Übertragungsfaktor spricht man im Falle einer Messung bei 1 kHz, diese Größe ist also nicht frequenzabhängig. Hier werden auch die Begriffe Sensitivity oder Empfindlichkeit gebraucht. Zu deren Bestimmung ist es also notwendig, im ungestörten Schallfeld einen Pegel von 94 dB SPL (gemäß 1 Pa) zu erzeugen, um anschließend an derselben Stelle den Prüfling zu fixieren und seine Ausgangsspannung zu messen. Der Aufbau und die Ergebnisse der durchgeführten Messung² sind in den folgenden Abschnitten beschrieben.

2.1.2.1 Messaufbau und Ergebnisse

Die Bestimmung der Mikrofon-Nennübertragungsfaktoren fand im reflexionsarmen Raum (RAR) des Instituts für Technische Akustik, TU Berlin, statt. Um die Handhabung zu vereinfachen und die geforderten Messbedingungen einhalten zu können, wurden die Mikrofone aus dem Kunstkopf ausgebaut.



Abbildung 2.2 – Messaufbau zur Bestimmung des Nenn-Übertragungsfaktors der Kunstkopfmikrofone

Im ersten Teil der Messung wurde das Messsignal, ein 1-kHz-Sinuston, mit Hilfe der Software *Monkey Forest* generiert. Wie in Abb. 2.2 dargestellt, war der signalausspielende Laptop zur D/A-Wandlung mit einem RME Hammerfall DSP Multiface verbunden, welches den Sinuston an einen Lautsprecher (Fostex 6301B) weiterleitete. In 1 m Abstand und auf gleicher Höhe des Lautsprechers wurde mit einem Schallpegelmessgerät (B&K Sound Level Meter 2205), das zur Gewährleistung eines ungestörten Schallfeldes auf einem Stativ befestigt war, der Referenzpegel gemessen. Im zweiten Teil wurden die Mikrofone abwechselnd an derselben Messposition wie bei der Schalldruckbestimmung fixiert. Das Signal des jeweiligen Prüflings wurde nach Versorgung mit Phantomspeisung an ein

²Sie erfolgte, bevor die Messung in Abb. 2.1 zur Verfügung stand. Dort können die gesuchten Werte direkt bei 1 kHz abgelesen werden.

Spannungsmessgerät (Sennheiser UPM 550-1) geleitet, wo die Ausgangsspannungen abgelesen wurden.

Bei einem Schalldruckpegel von 94 dB SPL lieferte die linke Mikrofon eine Spannung von 15,12 mV und das rechte Mikrofon 16,37 mV. Dadurch ergeben sich folgende Übertragungsmaße:

- ♦ links: -36,41 dB bzgl. 1 V/Pa
- ◊ rechts: -35,72 dB bzgl. 1 V/Pa

Es liegt also ein Pegelunterschied von weniger als 1 dB zwischen den beiden Mikrofonen vor, sodass von der Paarigkeit ihrer Sensitivity ausgegangen werden kann. Es ist außerdem anzumerken, dass die gemessenen Werte gut mit den Angaben des Herstellers übereinstimmen (s. Tabelle 2.1).

2.2 Kopfhörer

2.2.1 Diffus- und Freifeldentzerrung von Kopfhörern

Bei der Kopfhörerwiedergabe raumbezogener Signale³ ist eine Kopfhörerentzerrung nötig, die dafür sorgt, dass eine der Lautsprecherwiedergabe ähnliche Übertragung stattfindet (Møller et al. 1995b). Eine wichtige Frage ist hierbei, welche Schallfeldsituation als Referenz dient, das heißt ob die als Ziel dienende Lautsprecherübertragung im Frei- oder im Diffusfeld stattfindet. Theile (1986) schlägt als Standard eine Diffusfeldentzerrung vor und stützt sich dabei auf das so genannte Assoziationsmodell, wonach die Verarbeitung akustischer Reize in zwei Stufen geschieht. Die Ortsassoziationsstufe bestimmt auf Grundlage von Reflexionen an Ohr, Kopf und Oberkörper die Lokalisation des Hörereignisses. Die Gesaltassoziationsstufe bestimmt in einem zweiten Schritt andere Merkmale, wie z.B. die Klangfarbe des Hörereignisses.

"Die empfangenen Ohrsignale sind zurückzuführen auf die beiden voneinander unabhängigen, stets paarweise auftretenden Schallquelleneigenschaften ,Ort' und ,Signal'. Demzufolge sind im Modell die auftretenden Hörereignisse zurückzuführen auf die Wirkung einer ortsbestimmenden und einer gestaltbestimmenden Verarbeitungsstufe." (Theile 1981, S. 159)

Eine zentrale Aussage dieses Modells ist, dass die Ortsassoziationsstufe als eine Art Entzerrungsfilter auf die richtungsbestimmenden spektralen Verfärbungen im Hörereignis wirkt.

³Signale, die für Lautsprecherwiedergabe intendiert sind

Diese haben also keinen Einfluss auf die Klangfarbenwahrnehmung, die in der Gestaltverarbeitung geschieht.

Bei unkompensierter Kopfhörerwiedergabe entfallen die zur Lokalisation dienenden charakteristischen Verfärbungen, da keine Reflexionen an Kopf und Oberkörper stattfinden. Deshalb erfolgt keine Ortsassoziation sondern die kopfhörertypische Im-Kopf-Lokalisation. Durch die fehlende "Entzerrung" auf der Ortsassoziationsstufe wird das Signal aber mit den Verzerrungen des Kopfhörerfrequenzgangs der Gestaltassoziationsstufe zugeführt, die eine Klangfarbenveränderung identifiziert.

Wird ein Kopfhörer freifeldentzerrt, entspricht seine Übertragungsfunktion einer HRTF bezüglich einer Schallquelle im Direktfeld. Auf der Ortsassoziationsstufe wird also eine Hörereignisrichtung passend zu dieser HRTF bestimmt. Theile argumentiert jedoch, dass dies zu einem Konflikt auf der Gestaltassoziationsstufe führe, da die interauralen Differenzen herkömmlicher Stereoaufnahmen der Ortsassoziation einer einzigen frontalen Schallquelle entgegenstünden. Daher müssten die Kopfhörer nicht bezüglich einer spezifischen Richtung sondern bezüglich einer Mittelung aus allen möglichen Richtungen, dem Diffusfeld, entzerrt werden (vgl. Theile 1986, S. 959). Diesbezüglich von ihm durchgeführte Hörversuche ergaben, dass eine Wiedergabe über diffusfeldentzerrte Kopfhörer als "natürlicher" und "angenehmer" bewertet wurde als bei Freifeldentzerrung und zwar für mehrere Musikinhalte und mehrere Kopfhörer (Theile 1986, S. 966).

Diese Methoden der Entzerrung ersetzen jedoch keinesfalls die für eine korrekte binaurale Simulation nötige Kompensation des Kopfhörerfrequenzgangs. Es wäre lediglich denkbar, dass ein einziges Entzerrungsfilter für verschiedene Kopfhörer eingesetzt werden könnte. Dies würde eine, wie von Theile geforderte, Standardisierung von Kopfhörerfrequenzgängen voraussetzen. Obwohl bereits vorab vermutet wurde, dass keine einheitliche Kompensation möglich sein würde, gibt die im folgenden Abschnitt beschriebene Messung verschiedener Kopfhörerfrequenzgänge Aufschluss über deren Varianz.

2.2.2 Messung von Kopfhörerfrequenzgängen

In Abschnitt 2.1 wurde der spektrale Einfluss der Messmikrofone auf ein binaurales Simulationssystem bereits gesondert betrachtet, nun soll auch die Verzerrung durch die Kopfhörer miteinbezogen werden. Hierzu wurden die Frequenzgänge von sieben Kopfhörertypen gemessen (Tabelle 2.2 und Abb. 2.3), wobei selbstverständlich auch die am Fachgebiet verwendeten Modelle Stax SRS 2020 Lambda Basic und Stax SRS 2050 II zu den Prüflingen gehörten. Die ausgewählten Kopfhörer unterscheiden sich insbesondere hinsichtlich dreier Punkte:

- Wandlerprinzip dynamisch oder elektrostatisch
- akustische Kopplung an den Gehörgang akustisch offen, halboffen oder geschlossen⁴
- Bauart circumaural (ohrumschließend), supraaural (ohraufliegend) oder frei vor dem Ohr

Diese Eigenschaften sind in Tabelle 2.2 zusammengefasst.

Kopfhörer (Abkürzung)	Wandlerprinzip	Kopplung	Bauart	Verstärker
Audio-Technica ATH-M40fs (Audiotec)	dynamisch	geschlossen	circumarual	-
AKG K-401 (K401)	dynamisch	halboffen	circumarual	-
Sennheiser Headset H410 (Headset)	dynamisch	halboffen	supraaural	-
Stax Lambda Pro New (Stax Pro)	elektrostatisch	offen	circumarual	SRM 313
Stax SRS 2020 Lambda Basic (Stax I)	elektrostatisch	offen	circumarual	SRM 212
Stax SRS 2050 II (Stax II)	elektrostatisch	offen	circumarual	SRM 252 II
AKG K-1000 (K1000)	dynamisch	offen	frei vor dem Ohr	Carver PM-120

Tabelle 2.2 – Angaben zu den gemessenen Kopfhörern: Wandlerprinzip, Art der akustischen Kopplung an den Gehörgang, Bauart sowie bei der Messung eingesetzter Verstärker. In Klammern neben den Modellen steht jeweils die im Text verwendete Abkürzung.

Es sei an dieser Stelle angemerkt, dass keines der in DIN EN 60268-7 beschriebenen Übertragungsmaße⁵ bestimmt wurde, sondern der in Gl. 1.19 definierte Quotient aus Mikrofonausgangsspannung und Kopfhörereingangsspannung, $H(\omega)$. Diese relative Übertragungsfunktion wird in der vorliegenden Arbeit als Ausgangslage für die Implementierung verschiedener Entzerrungsalgorithmen dienen. Es wurde erläutert, dass bei der Messung von $H(\omega)$ derselbe Kunstkopf wie bei der BRIR-Akquise eingesetzt werden muss, um die gleichzeitige Kompensation von Kopfhörer und Mikrofonen zu ermöglichen. Obwohl also der Einfluss der Messmikrofone in $H(\omega)$ ebenfalls enthalten ist, wird der Einfachheit halber im Folgenden nur noch von Kopfhörerfrequenzgängen (bzw. -übertragungsfunktionen) gesprochen.

Um eine verzerrungsfreie binaurale Wiedergabe auch oberhalb von 22,05 kHz zu ermöglichen, müsste das Kompensationsfilter über diese Grenze hinaus korrekt funktionieren. Da zum Zeitpunkt der Messung eine konkrete Anwendung in Aussicht stand, bei der Kopfhörer bis 48 kHz betrieben werden sollten, wurden alle Frequenzgänge nicht nur mit einer Abtastfrequenz von 44,1 kHz sondern auch mit 96 kHz gemessen.

Bei der Bestimmung von Übertragungsfunktionen führt jede Veränderung in der Ausrichtung zwischen Quelle und Empfänger – in diesem Fall Kopfhörer und Messmikrofon – zu einem anderen Frequenzgang. Um die Auswirkung dieses Phänomens einschätzen und bei

⁴ "Akustisch offen" ist laut Norm definiert als "mit beabsichtigtem akustischem Nebenschluss zwischen der äußeren Umgebung und dem Gehörgang" (DIN EN 60268-7, S. 4). Der Begriff deckt sich also nicht zwingend mit dem in Abschnitt 1.3 definierten Begriff FEC.

⁵Es werden vier (!) unterschiedliche Messverfahren erläutert. Die im Folgenden beschriebene Messung entspricht am ehesten dem Ohrsimulator-Übertragungsmaß, allerdings wurde nicht der als Norm-Messbedingung geforderte Schalldruck von 94 dB (re 20 Pa) bei 500 Hz im Ohrsimulator erzeugt.



(a) Audiotec

(b) K401

(c) Headset



(d) K1000

(e) Stax Pro, Stax I, Stax II

Abbildung 2.3 – Gemessene Kopfhörer auf dem Messroboter FABIAN. In Abb. (d) sind die in kurzer Entfernung vom Ohr freistehenden Ohrhörer des K1000 sichtbar. Ihr Ausrichtungswinkel ist einstellbar, im Bild sind sie maximal geöffnet. Die drei Modelle der Fa. Stax sind in Abb. (e) zusammengefasst.

der Implementierung der Entzerrungsverfahren berücksichtigen zu können, wurden die Messungen bei jedem der sieben Prüflinge zehn Mal durchgeführt, indem der Kopfhörer zwischen zwei Messungen jeweils ab- und ohne allzu sorgfältige Ausrichtung wieder aufgesetzt wurde.

Wie im Falle der transauralen Wiedergabe erläutert wurde, muss bei der Entzerrung binauraler Signale eventuell auch das Übersprechen bedacht werden. Daher fand für die beiden halboffenen Kopfhörer K401 und Headset sowie für die offenen Kopfhörer K1000 und Stax Pro zusätzlich eine Messung der Übersprechdämpfung statt. Die anderen zwei Prüflinge der Fa. Stax wurden hierbei nicht gesondert berücksichtigt, da von einem sehr ähnlichen Ergebnis wie beim Stax Pro ausgegangen wird. Die Übersprechdämpfung ist die frequenzabhängige Differenz zwischen dem Schalldruck, den eine Kopfhörereingangsspannung im zu prüfenden Kanal erzeugt und dem Schalldruck, den dieselbe Spannung in diesem Kanal erzeugt, wenn sie jedoch am anderen Kanal angelegt wird (DIN EN 60268-7, S. 14).

2.2.2.1 Messaufbau



Abbildung 2.4 – Messaufbau zur Bestimmung der Kopfhörerfrequenzgänge

Die Messung fand wie die in Abschnitt 2.1 beschriebene im RAR mit Hilfe der Software Monkey Forest und dem Messroboter FABIAN statt. Ihr Aufbau ist in Abb. 2.4 schematisch dargestellt. Der Messlaptop war wieder mit einem RME-Interface verbunden, welches als D/A- bzw. A/D-Wandler eingesetzt wurde. Über zwei Line-Ausgänge war dieses Interface mit dem zu messenden Kopfhörer verbunden, in vier Fällen war zusätzlich ein Verstärker dazwischen geschaltet (siehe Tabelle 2.2), an dem der Lautstärkeregler jeweils auf ein mittleres Maß eingestellt wurde. Das über diesen Signalweg gesendete Messsignal wurde mit dem Kunstkopf des Messroboters aufgenommen. Die Mikrofone waren über symmetrische Leitungen mit der zum Monkey Forest-System gehörenden ITA ADDA16 Messhardware verbunden, welche als Vorverstärker diente und die benötigte Phantomspeisung lieferte. Das ITA-Frontend war über analoge Line-Ausgänge mit dem A/D-Wandler verbunden, von dort wurde das Signal wieder digital an die Messsoftware zurückgeschickt.

Als Anregungssignale wurden zwei lineare Sweeps⁶ im Bereich von 20 Hz bis 22,05 kHz bzw. bis 48 kHz erzeugt. Die Abtastraten betrugen 44,1 kHz und 96 kHz, die Anzahl der

⁶Zur Theorie des Frequenzgangsmessung mit Sweeps s. (Müller u. Massarani 2001)

samples 2¹⁶ und 2¹⁷. Somit war eine Frequenzauflösung von < 1 Hz gewährleistet:

$$\Delta f_{44} = \frac{44100 \,\mathrm{Hz}}{2^{16}} = 0,67 \,\mathrm{Hz} \qquad \Delta f_{96} = \frac{96000 \,\mathrm{Hz}}{2^{17}} = 0,73 \,\mathrm{Hz}$$

Beiden Sweeps wurde eine Ruhepause der Länge 100 ms hinzugefügt (*stopmargin*), um das Ausschwingen der Kopfhörer bei der Messung zu berücksichtigen. Zudem erfolgte eine Anhebung des Bassbereichs um 20 dB unterhalb von 100 Hz, um auch im tieffrequenten Bereich einen ausreichend großen Störabstand zu ermöglichen.

Die Übertragungsfunktion eines Systems wird nach Umstellen von Gl. 1.2 berechnet, indem das Spektrum eines Ausgangssignals durch das Spektrum eines Eingangssignals komplex dividiert wird:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)}$$
(2.3)

Bei einer FFT-basierten Messung, wie sie hier beschrieben wird, muss also das Spektrum des Eingangssignals bestimmt und anschließend invertiert werden. Die resultierende Funktion wird mit dem Ausgangssignal einer Messung multipliziert, um die gesuchte Übertragungsfunktion zu erhalten. Dieser Vorgang wird Entfaltung oder Dekonvolution genannt (vgl. Müller 2008, S. 1095 f.). Dementsprechend wurde vorab eine Referenzmessung durchgeführt, indem das RME-Interface mit dem ITA-Frontend verbunden wurde (siehe Abb. 2.4: Referenzmessung). Über diesen Signalweg fand eine Aufzeichnung der Frequenzgänge beider Sweep-Eingangssignale zwischen 0 Hz und 22,05 kHz bzw. 48 kHz statt. Die Messsoftware führte bei jeder weiteren Messung eine Multiplikation des Ausgangssignals mit der Inversen des Referenzspektrums durch. Hierbei wurde der Einfluss des Messsystems selbst ebenfalls kompensiert, da dieser ja in den invertierten Eingangssignalen enthalten ist. Zur Kontrolle wurde das kurzgeschlossene System erneut gemessen und entfaltet, wobei sich erwartungsgemäß ein konstanter Amplitudengang ergab.

Die Messungen wurden für den rechten und den linken Kanal gleichzeitig durchgeführt, das heißt das Anregungssignal wurde über beide Ausgangskanäle gesendet und von beiden Eingangskanälen aufgenommen. Bei der Übersprechmessung wurde das Anregungssignal abwechselnd nur über einen Ausgang geschickt, jedoch von beiden Eingangskanälen aufgenommen. Um auch den Einfluss des Raumes einschätzen zu können, wurde vor jeder Übersprechmessung das Grundrauschen aufgezeichnet. Dies geschah, indem die Signalkette am D/A-Wandler-Ausgang unterbrochen wurde, durch die Messmikrofone aber trotzdem für die Dauer des Messsignals eine Aufzeichnung stattfand.

2.2.2.2 Ergebnisse

Frequenzgänge

Zunächst werden in Abb. 2.5 und Abb. 2.6 die zweikanaligen Betragsfrequenzgänge der Prüflinge zwischen 20 Hz und 48 kHz gezeigt. Die Notwendigkeit einer Kopfhörerentzerrung wird auf den ersten Blick ersichtlich, da vor allem im hohen Frequenzbereich starke Schwankungen auftreten, hervorgerufen durch Partialschwingungen der Kopfhörermembranen aber auch durch das Schwingungsverhalten des Luftpolsters zwischen Ohr und Ohrhörer. Es lässt sich eine Ähnlichkeit der Frequenzgänge mit Außenohr-Übertragungsfunktionen erkennen. So weisen sie z.B. alle bei 8-10 kHz einen mehr oder weniger starken Einbruch auf. Dieser gilt bei HRTFs als charakteristisch für den Einfluss der Ohrmuschel. Es wurde also offenbar im Zuge einer Diffusfeldentzerrung eine durch die Partialschwingungen der Kopfhörermembran bedingte Auslöschung auf diesen Bereich abgestimmt. Trotzdem kann nicht von einer standardisierten Kalibrierung gesprochen werden, wie sie von Theile gefordert wurde. Wie schon im Vorfeld vermutet, müssen Kopfhörerfrequenzgänge in der Binauraltechnik also individuell entzerrt werden.

Bei allen Übertragungsfunktionen, insbesondere bei denjenigen des K1000 und des Stax Pro, ist unter 50 Hz ein Störbereich zu sehen. Dieser ist einerseits auf Raumreflexionen zurückzuführen, da die vollständige Absorption der Wände des RAR unter 63 Hz nicht mehr gewährleistet ist. Der K1000 ist durch seine offene Ohrhörerstellung (die Messungen erfolgten mit maximaler Öffnung, s. Abb. 2.3) für solche Störungen besonders anfällig. Weiterhin macht sich bei tiefen Frequenzen die abnehmende Effizienz der Kopfhörer bemerkbar, wodurch der Störabstand trotz erhöhter Energie des Messsignals in diesem Bereich sinkt.

Es wurde in Abschnitt 2.1 bereits angemerkt, dass die Messmikrofone den Übertragungsbereich über 20 kHz deutlich beeinträchtigen. Um von den Kopfhörerübertragungsfunktionen diesbezüglich ein unverfälschtes Bild zu bekommen, wurden sie mit den inversen der Mikrofonfrequenzgänge multipliziert. Das Ergebnis dieser Kompensation ist jeweils im oberen Teil der Grafiken von Abb. 2.5 und Abb. 2.6 dargestellt. Besonders die Stax-Kopfhörer leisten eine relativ gute Übertragung in diesem hochfrequenten Bereich. Allerdings nimmt hier, wie bei allen Prüflingen, die Anzahl der schmalbandigen Einbrüche zu. Diese können bei neuerlichem Aufsetzen des Kopfhörers stark variieren (s. Abb. 2.10-Abb. 2.11) und stellen bei der Kompensation unter Umständen ein Problem dar (s. Abschnitt 3.1).

Die Phasengänge der Prüflinge sind in Abb. 2.7 und Abb. 2.8 gezeigt. Ihre Paarigkeit ist genau wie bei den Messmikrofonen für die Binauraltechnik von Bedeutung. Lindau (2006, S. 60) fasst Forschungsergebnisse zusammen und kommt zum Schluss, dass vor allem im empfindlichen Bereich zwischen 400 Hz und 1 kHz der interaurale Phasenunterschied nicht mehr als 4° betragen sollte. Die Phasendifferenzen der Kanäle sind jeweils in den unteren Grafiken von Abb. 2.7 und Abb. 2.8 zwischen 100 Hz und 4 kHz zu sehen. Die meisten Prüflinge weisen zumindest im kritischen Bereich akzeptable Werte auf, gute Ergebnisse liefert hier vor allem das Headset aber auch der K1000 und die Stax-Kopfhörer. Inwiefern bei der Kompensation von Frequenzgängen auch die Phase entzerrt werden kann und muss, wird in Kapitel 3 näher erläutert.

Die ersten 1,4 ms der Impulsantworten der sieben Kopfhörer sind in Abb. 2.9 dargestellt. Interessanterweise ist der Ausschlag bei einigen Modellen phasengedreht, wobei dies jeweils beide Kanäle betrifft und somit als unkritisch betrachtet werden kann. Zu beachten ist der große Pegelunterschied der Kanäle der Audiotec-Stoßantwort, der bereits im Betragsfrequenzgang zu sehen ist und bezüglich interauraler Differenzen problematisch sein könnte.

In Abb. 2.10 und Abb. 2.11 sind für alle Kopfhörer die Betragsfrequenzgänge von jeweils zehn Messungen, zwischen denen der Prüfling jedes Mal ab- und wieder aufgesetzt wurde, aufgetragen. Hierbei zeigen sich vor allem ab 3-4 kHz relativ große Unterschiede bezüglich Tiefe und exakter Frequenz der schmalbandigen Einbrüche. Sie könnten darauf zurückgeführt werden, dass sich die Größe und damit auch das Schwingungsverhalten des Luftpolsters zwischen Ohr und Ohrhörer bei neuem Aufsetzen des Kopfhörers verändert. Beim Audiotec und den Stax-Kopfhörern sind jedoch auch bei tiefen Frequenzen deutliche Verschiebungen zu erkennen. Sie könnten mit einem undichten Abschluss des Ohrhörers um das Ohr herum – hervorgerufen z.B. durch schlechtes Sitzen des Kopfhörers auf dem Kunstkopf – zusammenhängen, was jedoch nicht bei jedem Aufsetzen gleich ausgeprägt war. Weiterhin sei auf Abb. 2.11(d) hingewiesen, die am Beispiel des Stax II eine weitere Zunahme der Verschiebungen gen oberhalb von 20 kHz zeigt.

Abbildung 2.5 – Betragsfrequenzgänge zwischen 20 Hz und 48 kHz für linken und rechten Kanal. Gezeigt sind eine vom Mikrofonfrequenzgang befreite (oben) und eine unkompensierte (unten) Version.






Abbildung 2.7 – Zweikanalige Phasengänge zwischen 20 Hz und 22 kHz. Die untere Grafik zeigt jeweils die Phasendifferenz der Kanäle zwischen 100 Hz und 4 kHz.



Abbildung 2.8 – Zweikanalige Phasengänge (Fortsetzung)



Abbildung 2.9 – Zweikanalige Impulsantworten aller Prüflinge von 0 bis 1,4 ms





Abbildung 2.11 – (a)-(c): Varianz durch zehnmaliges Ab- und Aufsetzen der Kopfhörer zwischen 20 Hz und 22 kHz (Fortsetzung) (d): Varianz des Stax II zwischen 20 und 48 kHz



Übersprechdämfung

In Abb. 2.12 ist das Übersprechen für die vier halboffenen bzw. offenen Kopfhörer gezeigt. Die durchgezogene blaue Linie ist der am rechten Kanal gemessene Frequenzgang, die gestrichelte blaue Linie das dort gemessene, oktavgemittelte Übersprechen. Die Übersprechdämpfung, also die Differenz der beiden Signale, ist als grüne Linie darüber dargestellt. Das Grundrauschen des Raumes ist jeweils im unteren Bild gezeigt. Dabei ist sichtbar, wie zu tiefen Frequenzen der Rauschpegel der Umgebung die Messung bestimmt hat, am deutlichsten ist dies in Abb. 2.12(c). Wie bereits oben beschrieben, hängt dies mit den Raumreflexionen und der abnehmenden Kopfhörerleistung im Bassbereich zusammen.

Nicht überraschend ist die Tatsache, dass der transaurale Kopfhörer mit 30 dB bis maximal 40 dB die geringste Übersprechdämpfung aufweist. Durch die offene Ohrhörerstellung war hier die Empfindlichkeit gegenüber von Außen eintreffenden Signalen am höchsten. Dies müsste bei einer Entzerrung eventuell berücksichtigt werden. Nicht wesentlich unterschiedlich sind die Ergebnisse der beiden halboffenen Kopfhörer gegenüber dem offenen Stax Pro, hier bewegen sich die Werte um ca. 50 dB, wobei Headset und Stax Pro maximal 60 dB erreichen. Interessant ist zu sehen, dass sich das Maximum aller Prüflinge bei ca. 200 Hz befindet – wobei das Headset hier den weitesten Bereich abdeckt – und die minimale Dämpfung um 2 kHz liegt. Da hier kein Zusammenhang mit dem Grundrauschen des Raums zu erkennen ist, könnte letzteres mit einem charakteristischen Einfluss des Messroboterkörpers zusammenhängen, da ja das Übersprechsignal eine Art Außenohrübertragungsfunktion darstellt.

Die Analyse der Messergebnisse soll an dieser Stelle beendet werden. Zusammenfassend kann festgestellt werden, dass die Amplituden- und Phasengänge unterschiedlicher Kopfhörertypen eine bedeutende Varianz aufweisen. Dies gilt sowohl für die Übertragungsfunktionen untereinander als auch in sich, wie mehrere Messungen am selben Prüfling zeigen. Welche Konsequenzen letzteres für die Entzerrung hat, wird in Kapitel 3 näher beschrieben.



Abbildung 2.12 – Übersprechen der halboffenen und offenen Kopfhörer (rechter Kanal): direktes Signal (durchgezogen, blau), Übersprechen (gestrichelt, blau), Übersprechdämpfung (grün) und Grundrauschen des Raumes (unten, schwarz)

3 ENTZERRUNG VON FREQUENZGÄNGEN

3.1 Grundlagen

3.1.1 Direkte Invertierung

Im vorangegangenen Kapitel wurden die spektralen Einflüsse auf das binaurale Simulationssystem durch die Kopfhörer- und Mikrofonübertragungsfunktion $H(\omega)$ charakterisiert. Zu deren Kompensation ist ein FIR-Filter $H_c(\omega)$ (Gl.1.18) gesucht, sodass gilt

$$H_{\rm eq}(\omega) = H(\omega) \cdot H_{\rm c}(\omega) = 1 \tag{3.1}$$

bzw.

$$h_{\rm eq}(n) = \sum_{k=0}^{N-1} h(k) h_{\rm c}(n-k) = h(n) * h_{\rm c}(n) = \delta(n).$$
(3.2)

Es soll also eine vollständige Entzerrung im Frequenzbereich erfolgen, die einem Delta-Impuls im Zeitbereich entspricht. Die intuitiv einfachste Herangehensweise zur Berechnung von $H_c(\omega)$ aus $H(\omega)$ scheint die direkte Invertierung zu sein:

$$H_{\rm c}(\omega) = \frac{1}{H(\omega)} \tag{3.3}$$

Dabei ergeben sich im Kontext akustischer und elektroakustischer Systeme jedoch die folgenden Probleme:

- Die Linearisierung der kompletten Übertragungsfunktion führt in ihren Randbereichen zu übermäßiger Anhebung des Amplitudengangs.
- Es handelt sich in den meisten Fällen um gemischtphasige Systeme, sodass eine direkte Invertierung nicht zu einem gleichzeitig kausalen und stabilen Filter führt.
- Die exakte Kompensation ist an eine ganz bestimmte Ausrichtung von Sender (Kopfhörer) und Empfänger (Kunstkopfmikrofon bzw. Hörerohr) gebunden, da jede Konfiguration durch einen eigenen Frequenzgang beschrieben wird.

Auf die Lösung der beiden erstgenannten Schwierigkeiten wird in den folgenden Abschnitten genauer eingegangen. Die letztgenannte Problematik wird durch die zehn Einzelmessungen der Kopfhörerübertragungsfunktionen in Abb. 2.10 und Abb. 2.11 veranschaulicht. Die Verschiebung der Sendeposition gegenüber der Messposition ist besonders bei der dort sichtbaren Varianz im hohen Frequenzbereich kritisch. Ein perfektes Entzerrungsfilter gleicht schmalbandige Einbrüche (*dips* oder *notches*) durch entsprechende Überhöhungen (*peaks*) aus. Mit der Positionsänderung können sich die dips leicht verschieben, die peaks im Amplitudengang von $H_{\rm c}(\omega)$ bleiben jedoch an ihrer ursprünglichen Lage und werden als Klingelartefakte deutlich hörbar.¹ Dies kann durch eine Reduktion der Kompensationsgenauigkeit im Bereich der notches umgangen werden. Dadurch sind sie zwar in der entzerrten Funktion $H_{eq}(\omega)$ nach wie vor enthalten, gleichzeitig sind jedoch die durch Verschiebungen entstehenden Artefakte geringer. Ein solcher Ansatz stützt sich auf die in zahlreichen Untersuchungen gezeigte Tatsache, dass schmalbandige Einbrüche weniger störend wahrgenommen werden als entsprechende Überhöhungen (s. z.B. Bücklein 1981; Toole u. Olive 1988; Moore et al. 1989). Zur Umsetzung dieses Lösungsprinzips existieren verschiedene Verfahren, die in Abschnitt 3.2 näher beschrieben werden.

3.1.2 Zielfunktion

Um die exzessive Anhebung des zu kompensierenden Frequenzgangs außerhalb seines normalen Übertragungsbereichs zu vermeiden, muss eine geeignete Zielfunktion definiert werden (Karjalainen et al. 2005). Gln. 3.1 und 3.2 verändern sich zu:

$$H_{\rm eq}(\omega) = H(\omega) \cdot H_{\rm c}(\omega) = D(\omega) \tag{3.4}$$

und

$$h_{\rm eq}(n) = h(n) * h_{\rm c}(n) = d(n).$$
 (3.5)

 $D(\omega)$ und d(n) bezeichnen den Frequenzgang und die Impulsantwort eines Bandpassfilters, das die Entzerrung auf eine obere und untere Grenzfrequenz beschränkt. Abgestimmt auf den von den Kopfhörern abgedeckten Übertragungsbereich wurde der Durchlassbereich der hier eingesetzten Zielfunktion auf 50 Hz bis 21 kHz festgelegt. Eine Sperrdämpfung von 60 dB wurde für die Anwendung als ausreichend erachtet.

Eine grundsätzliche Frage betrifft das Phasenverhalten des Bandpasses, da dieser sowohl

¹Solch starke Überhöhungen wären auch bei gleichbleibender Abhörposition unerwünscht, da sie eine sehr hohe Ordnung des Kompensationsfilters erfordern und im Extremfall auch den Wiedergabewandler beschädigen könnten (Kirkeby u. Nelson 1999).

minimal- als auch linearphasig entworfen werden kann.² Die erste Option zeigt den Vorteil, dass zur Einhaltung eines gegebenen Toleranzschemas eine geringere Filterordnung erforderlich ist als bei linearphasigen Filtern (s. The Mathworks Inc. 2006, S. 1-4). Bei extremen Kriterien, wie beispielsweise hoher Flankensteilheit und Sperrdämpfung, ergeben sich nach Angaben von Lindau (2006, S. 123) jedoch instabile Filter. Ein bedeutender Nachteil betrifft zudem die Abweichung von einer konstanten Laufzeit. Laut Blauert und Laws (1978) beträgt die gerade hörbare monaurale Laufzeitverzerrung elektroakustischer Systeme bei empfindlichen Signalen 0,5 ms. Diese Schwelle wird von einem minimalphasigen Tshebysheff Typ II-Filter mit den oben genannten Eigenschaften unterhalb von 220 Hz überschritten. Gleichwohl handelt es sich hierbei um den Filtertyp mit den geringsten Laufzeitverzerrungen, wie Lindau (2006, S. 124) konstatiert. Perzeptive Untersuchungen von Müller (1999) im Zusammenhang der Lautsprecherentzerrung ergaben, dass gerade im Bassbereich Verzerrungen der Laufzeit hörbar sind, während dies bei Mittel- und Hochtonsystemen weniger ausgeprägt ist. Um bei der Kompenstation der Kopfhörer solche Abweichungen in der resultierenden Übertragungsfunktion zu vermeiden, wurde der Zielbandpass als ein linearphasiges Filter entworfen. Hierbei wird eine gewisse Grundlaufzeit in Kauf genommen, die jedoch unter Umständen von Vorteil ist (s. folgenden Abschnitt).

Die Implementierung erfolgte mit der fir1-Funktion der Signal Processing ToolboxTM von Matlab®, bei der ein FIR-Filter mit idealer Flankensteilheit mit einer anzugebenden Fensterfunktion geglättet wird (The Mathworks Inc. 2006). Es wurde dazu ein Kaiser-Fenster gewählt, dessen Nebenkeulenunterdrückung und Breite der Hauptkeule im Amplitudengang mit Hilfe des Parameters b_{FIR} kontrolliert werden können. Beide vergrößern sich mit steigendem Parameterwert. Wenn *a* die gewünschte Sperrddämfung in dB ist, so lässt sich das dazu optimale b_{FIR} wie folgt berechnen (The Mathworks Inc. 2006, S. 3-9):

$$b_{\rm FIR} = 0,1102 \cdot (a-8,7)$$
 bei $a > 50$ (3.6)

Wichtig zu erwähnen ist die Tatsache, dass die oben angegebenen Eckfrequenzen nicht die 3 dB- sondern die 6 dB-Punkte des Bandpasses bezeichnen, da dies vom fir1-Algorithmus so vorgegeben wird. Der Amplituden- und Phasengang sowie die Impulsantwort des resultierenden Filters sind in Abb. 3.1 gezeigt. Die zwei Betragsfrequenzgänge in Abb. 3.1(a) ergeben sich aus den Filterordnungen 2¹⁰ und 2¹¹. Es wird ersichtlich, dass die gewünschte Sperrdämpfung von 60 dB zwar in beiden Fällen beiden erreicht wird, allerdings mit sehr unterschiedlicher

²ausgeführt in (Lindau 2006, S. 199 ff.); vgl. auch (Müller u. Massarani 2001)



Abbildung 3.1 – Zielbandpass: (a) Betragsfrequenzgang für eine Filterordnung von 2^{10} und 2^{11} ; (b) Impulsantwort und Phasengang bei Ordnung 2^{11}

Flankensteilheit. Die geringere Ordnung wurde als nicht ausreichend eingestuft, sodass der Zielbandpass mit 2049 samples³ implementiert wurde. Die Länge von d(n) sollte diejenige des Entzerrungsfilters nicht überschreiten, da die bei der Berechnung von $h_c(n)$ erzwungene Kürzung (siehe Abschnitt 3.2) ansonsten wichtige Filterkoeffizienten eliminieren würde. Die hier festgelegte Bandpass-Ordnung markiert somit die untere Grenze für die Länge von $h_c(n)$. Bei kürzeren Entzerrungsfiltern müsste unter Umständen doch eine minimalphasige Zielfunktion entworfen werden. Da aber im vorliegenden Fall vom Standpunkt der Kompensationsgüte nicht mit geringeren Ordnungen zu rechnen war,⁴ wurde der linearphasige Bandpass als Zielfunktion beibehalten.

3.1.3 Minimal- und gemischtphasige Invertierung

Um die eingangs gemachte Aussage bezüglich der Stabilität und Kausalität von invertierten, gemischtphasigen Systemen zu erläutern, soll zunächst auf diese Eigenschaften näher eingegangen werden. Die folgenden Ausführungen stützen sich dabei auf Oppenheim und Schafer (1992).

Die Systemfunktion H(z) eines LTI-Systems kann allgemein ausgedrückt werden durch

$$H(z) = \left(\frac{b_0}{a_0}\right) \frac{\prod_{k=1}^{M} (1 - c_k z^{-1})}{\prod_{k=1}^{N} (1 - d_k z^{-1})}$$
(3.7)

³Die ungerade Länge bzw. gerade Ordnung ist wichtig, um einen Abtastwert beim Maximum der Impulsantwort zu erhalten, welches sich per Definition genau in deren Mitte befindet.

⁴siehe Ergebnisse in Abschnitt 4.1

und weist Nullstellen bei $z = c_k$ und z = 0 sowie Pole bei $z = d_k$ und $z = \infty$ auf. Ein solches System kann mehrere gültige Konvergenzgebiete besitzen, also Bereiche der z-Ebene, welche von Polen begrenzt werden, diese jedoch nicht enthalten. Durch die Bedingungen der Kausalität und Stabilität werden die möglichen Konvergenzgebiete jedoch eingeschränkt. Ein System ist dann kausal, wenn $h(n_0)$ nur von Werten $n \le n_0$ abhängt. Die Zeitfunktion muss also eine rechtsseitige Folge sein. Für den Konvergenzbereich von H(z) bedeutet dies, dass er außerhalb des äußersten Pols liegen muss. Die Bedingung der Stabilität wird erfüllt, wenn h(n) absolut summierbar ist:

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty \tag{3.8}$$

Diese Voraussetzung ist äquivalent zu

$$\sum_{n=-\infty}^{\infty} |h(n)z^{-n}| < \infty \qquad \text{mit} \quad |z| = 1,$$
(3.9)

sodass das Konvergenzgebiet stabiler Systeme also den Einheitskreis mit einschließen muss. Zusammengefasst ergibt sich daraus für gleichzeitig stabile und kausale Systeme, dass sich alle ihre Pole innerhalb des Einheitskreises befinden.

Soll nun H(z) wie in Gl. 3.3 direkt invertiert werden, wird aus Gl. 3.7:

$$H_{\rm inv}(z) = \left(\frac{a_0}{b_0}\right) \frac{\prod_{k=1}^{N} (1 - d_k z^{-1})}{\prod_{k=1}^{M} (1 - c_k z^{-1})}$$
(3.10)

Die Nullstellen von H(z) werden hier zu Polen von $H_{inv}(z)$ und umgekehrt. Wenn also das inverse System ebenfalls kausal und stabil sein soll, so müssen auch die Nullstellen von H(z)alle innerhalb des Einheitskreises liegen.

"Deshalb ist ein lineares zeitinvariantes System dann und nur dann stabil und kausal und besitzt ein stabiles und kausales inverses System, wenn die Pole und Nullstellen von H(z) innerhalb des Einheitskreises liegen. Solche Systeme werden *minimalphasige* Systeme genannt." (Oppenheim u. Schafer 1992, S. 239)

Die Pole und Nullstellen maximalphasiger Folgen liegen hingegen alle außerhalb des Einheitskreises. Da diese Systeme aber die Bedingung der Stabilität erfüllen, folgt daraus, dass sie linksseitig – also akausal – sein müssen. Gemischtphasige Systeme sind zwar stabil und kausal, ihre Pole liegen folglich alle innerhalb des Einheitskreises. Da sie aber auch Nullstellen außerhalb von |z| = 1 besitzen, erfüllen sie nicht die obige Voraussetzung der direkten Invertierbarkeit.

Die meisten akustischen und elektroakustischen Übertragungsfunktionen sind gemischtphasig. Neely und Allen (1979) schlagen zur Kontrolle dieser Eigenschaft die Analyse des Nyquist-Diagramms vor. Demnach bezeichnet die Anzahl der positiven Umrundungen des z-Ebenen-Ursprungs die Differenz zwischen Nullstellen und Polen außerhalb des Einheitskreises. Da sich bei stabilen und kausalen Systemen aber keine Pole jenseits dieser Grenze befinden, deuten Umrundungen des Ursprungs auf Nullstellen außerhalb von |z| = 1 und damit auf ein gemischtphasiges System hin. Die derartige Überprüfung der gemessenen Kopfhörerübertragungsfunktionen auf Minimalphasigkeit fiel in allen Fällen negativ aus.

Trotzdem kann aus einem solchen System ein geeignetes Filter zur Entzerrung gewonnen werden. Hierfür gibt es zwei Lösungsmöglichkeiten. Die erste stützt sich auf die Tatsache, dass jede stabile, kausale Funktion H(z) durch die Multiplikation eines minimalphasigen und eines Allpass-Systems ausgedrückt werden kann:⁵

$$H(z) = H_{\min}(z) \cdot H_{ap}(z) \tag{3.11}$$

Daraus folgt mit $z = e^{j\omega}$, dass $|H_{\min}(\omega)| = |H(\omega)|$ gelten muss, da $|H_{ap}(\omega)| = 1$. Der zu $H_{\min}(\omega)$ gehörende Phasengang kann mit Hilfe der Hilbert-Transfomation direkt aus dem Logarithmus des gegebenen Amplitudengangs berechnet werden (Preis 1982):

$$\varphi_{\min}(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\ln |H(\omega')|}{\omega' - \omega} d\omega'$$
(3.12)

Zur Entzerrung des Betragsfrequenzgangs von $H(\omega)$ kann folglich die stabile und kausale Inverse des minimalphasigen Terms aus Gl. 3.11 gebildet werden.

$$H_{\rm c}(\omega) = \frac{1}{H_{\rm min}(\omega)} \tag{3.13}$$

Für das kompensierte System ergibt sich demnach

$$H_{\rm eq}(\omega) = H(\omega) \cdot H_{\rm c}(\omega) = H_{\rm min}(\omega) \cdot H_{\rm ap}(\omega) \cdot \frac{1}{H_{\rm min}(\omega)} = H_{\rm ap}(\omega)$$
(3.14)

Während der Betragsfrequenzgang also exakt kompensiert wird, besitzt $H_{eq}(\omega)$ den (nichtlinearen) Phasengang von $H_{ap}(\omega)$.

Ähnlich wie in Gl. 3.11 kann eine Systemfunktion auch durch die Multiplikation eines minimal- und eines maximalphasigen Terms ausgedrückt werden, sofern sich keine Nullstel-

⁵zum Beweis s. (Oppenheim u. Schafer 1992, S. 271 f.)

len auf dem Einheitskreis befinden:

$$H(z) = H_{\min}(z) \cdot H_{\max}(z) \tag{3.15}$$

Die getrennte Analyse beider Teile soll den Lösungsansatz zur Kompensation der gesamten gemischtphasigen Übertragungsfunktion $H(\omega)$ verdeutlichen (nach Mourjopoulos 1994). Dazu habe die Systemfunktion der Einfachheit halber lediglich zwei Nullstellen, die bei $|c_1| < 1$ und $|c_2| > 1$ liegen:

$$H(z) = H_{\min}(z) \cdot H_{\max}(z) = (1 - c_1 z^{-1})(1 - \frac{1}{c_2} z)$$
(3.16)

Laut der obigen Definition von maximalphasigen Folgen umfasst deren Konvergenzbereich den Einheitskreis, das heißt sie sind stabil aber akausal. Dies wird durch den positiven Exponenten von z im zweiten Term von Gl. 3.16 ausgedrückt. Die Inversen der Teilfunktionen lauten wie folgt:

$$H_{\min_{i}}(z) = \frac{1}{1 - c_{1}z^{-1}} = 1 + c_{1}z^{-1} - c_{1}^{2}z^{-2} + \dots$$
(3.17)

$$H_{\max_i}(z) = \frac{1}{1 - \frac{1}{c_2}z} = \left(\frac{1}{c_2}\right)z + \left(\frac{1}{c_2^2}\right)z^2 - \left(\frac{1}{c_2^3}\right)z^3 + \dots$$
(3.18)

Es resultieren also in beiden Fällen unendlich lange Folgen, die jedoch wegen ihrer Konvergenz gegen Null durch endliche Sequenzen angenähert werden können. Die akausale Folge $h_{\max_i}(n)$ kann außerdem durch ein *modelling delay* um eine Anzahl von *m* samples verzögert werden, indem Gl. 3.18 mit dem Verzögerungsterm z^{-m} multipliziert wird:

$$\tilde{H}_{\max_{i}}(z) = \left(\frac{1}{c_{2}}\right) z^{-m+1} + \left(\frac{1}{c_{2}^{2}}\right) z^{-m+2} - \dots + (-1)^{m} \left(\frac{1}{c_{2}^{m}}\right) + (-1)^{m-1} \left(\frac{1}{c_{2}^{m+1}}\right) z + \dots$$
(3.19)

 $h_{\max_i}(n)$ ist nun also eine zweiseitige Folge, die durch das Eliminieren des akausalen Teils zu einer stabilen und kausalen Impulsantwort wird. Die Länge des verbleibenden Abschnitts hängt von der Größe der Verzögerung *m* ab.

Das modelling delay wird bei der Berechnung von Kompensationsfiltern im Zeitbereich realisiert, indem der Diracstoß in Gl. 3.2 verzögert wird (Clarkson et al. 1985; Kirkeby u. Nelson 1999).

$$h_{\rm eq}(n) = h(n) * h_{\rm c}(n) = \delta(n-m)$$
 (3.20)

Bei der Bestimmung des Entzerrungsfilters im Frequenzbereich nach Gl. 3.3 kann die Kausalität erreicht werden, indem $H_c(\omega)$ in den Zeitbereich transformiert und zirkulär verschoben wird (Kirkeby et al. 1998). Alternativ ist das Einsetzen des Frequenzgangs einer verzögerten Zeitfunktion in den Zähler von Gl. 3.3 möglich (Fielder 2001).

$$H_{\rm c}(\omega) = \frac{D(\omega)}{H(\omega)} \tag{3.21}$$

Bezüglich der Größe von *m* hat sich ein delay um die halbe Filterlänge als Standardwert durchgesetzt (vgl. z.B. ?). Im Falle des oben definierten Zielbandpasses ist eine Verzögerung um die halbe sample-Anzahl bereits in der Impulsantwort enthalten. Bei Übereinstimmung von Bandpassordnung und Länge des Kompensationsfilters ist also keine explizite Einführung eines delays nötig. Bei größeren Filterlängen und damit verbundenen längeren Verzögerungen wurden diese als eine zirkuläre Verschiebung der Bandpassimpulsantwort implementiert (s. Quellcodes in Anhang A).

Die Ergebnisse verschiedener Untersuchungen zeigen, dass die Inverse des minimalphasigen Teils einer Übertragungsfunktion in den meisten Fällen nicht zu deren Kompensation ausreicht (Neely u. Allen 1979; Johansen u. Rubak 1996; Norcross et al. 2002). Der nicht-lineare Phasengang der verbleibenden Allpass-Funktion führt offenbar zu störendem "Verschmieren" des Zeitsignals (Mourjopoulos 1994). Eine Verzögerung der Filterimpulsantwort war in der vorliegenden Arbeit durch die Wahl der Zielfunktion ohnehin festgelegt,⁶ sodass sich vorerst kein Vorteil der minimalphasigen Entzerrung erkennen ließ. Obwohl ihr Vergleich mit der gemischtphasigen Methode eventuell trotzdem lohnenswert wäre, wurde sie hier nicht weiter verfolgt. Dies geschah vor allem auch aus Kapazitätsgründen, da die im nächsten Abschnitt beschriebenen Filterentwurfsmethoden wie erwähnt in einem Hörversuch evaluiert werden sollten. Eine Gegenüberstellung gemischt- und minimalphasiger Entzerrungen hätte eine Verdopplung der Bedingungen bedeutet, was bei der Anzahl der implementierten Kompensationsverfahren nicht realistisch war.⁷

⁶Andernfalls hätte ein minimalphasiger Bandpass mit seinen oben beschriebenen Nachteilen definiert werden müssen.

⁷Auf die Einschränkung der Bedingungsvariationen wird in Abschnitt 4.2 noch näher eingegangen.

3.2 Entzerrungsmethoden

3.2.1 Stand der Forschung

Die Literatur zum Entwurf digitaler Entzerrungsfilter bezieht sich in den meisten Fällen auf die Kompensation von Lautsprecher- und/oder Raumeinflüssen. Nichtsdestotrotz ist es möglich, einige der Methoden auch auf die Fragestellung dieser Arbeit - die Kompensation von Kopfhörer- und Mikrofonfrequenzgängen in der Binauraltechnik - anzuwenden. Die existierenden Verfahren können prinzipiell danach eingeteilt werden, ob die Filterberechnung im Zeit- oder im Frequenzbereich durchgeführt wird. Letztere ist leicht zu verstehen und besteht in der Invertierung der zu entzerrenden Übertragungsfunktion nach Gl. 3.1 unter Einhaltung der oben genannten Randbedingungen bezüglich Zielfunktion und Verzögerung (Leckschat 1992; Kirkeby et al. 1998; Müller 1999). Im Frequenzbereich berechnete Entzerrungsfilter erzielen bei ausreichender Länge laut Norcross et al. (2002) gute Ergebnisse. Eine Methode des Filterentwurfs im Zeitbereich ist die so genannte homomorphe Zerlegung. Dabei werden der minimal- und der maximalphasige Teil der zu entzerrenden Impulsantwort durch Cepstral-Analyse bestimmt, separat invertiert und die Ergebnisse anschließend miteinander gefaltet (Mourjopoulos et al. 1982). Da sich diese Art der Zerlegung jedoch als fehleranfällig herausstellte, wird der Filterentwurf nach dem Kriterium des kleinsten quadratischen Fehlers normalerweise bevorzugt (Mourjopoulos 1994). Dabei soll die Differenz zwischen entzerrter Impulsantwort und Zielfunktion minimiert werden (Mourjopoulos et al. 1982; Kirkeby et al. 1998). Dieses Verfahren erwies sich sowohl bei theoretisch berechneter Güte als auch bei Hörversuchen als sehr effektiv (Mourjopoulos et al. 1982; Clarkson et al. 1985; Norcross et al. 2002).

Die Empfindlichkeit einer exakten Kompensation gegenüber Änderungen der Sender-Empfänger-Anordnung stellt laut Rubak und Johansen (2000) ein gravierendes Problem bei der Kompensation von Lautsprechern und Räumen dar. Dies wurde auch im Falle der hier zu entzerrenden Kopfhörer erkannt. Wie eingangs erwähnt, existieren in dieser Hinsicht mehrere Verfahren, die auf eine höhere Robustheit der Entzerrung abzielen. Eine Möglichkeit besteht beispielsweise in der Invertierung einer geglätteten Übertragungsfunktion, wobei die Glättung entweder nur den Betragsfrequenzgang oder aber die komplexe Übertragungsfunktion betreffen kann (Hatziantoniou u. Mourjopoulos 2000). Diesbezüglich durchgeführte perzeptive Untersuchungen ergaben viel versprechende Resultate (Norcross et al. 2003b; Hatziantoniou u. Mourjopoulos 2004). Die Arbeit von Mourjopoulos und Paraskevas (1991) zeigt, dass auch aus dem All-Pol-Modell des zu entzerrenden Frequenzgangs ein robustes Kompensationsfilter berechnet werden kann. Müller (1999) schlägt die gewichtete Angleichung des geglätteten Amplitudengangs an seinen ursprünglichen Verlauf vor. Dabei werden Überhöhungen perfekt angepasst, während dies bei Einbrüchen nur eingeschränkt geschieht. Die Entzerrung mit der Inversen einer solchen Funktion wurde in Hörversuchen ebenfalls positiv beurteilt. Eine Herangehensweise, die sowohl im Frequenz- als auch im Zeitbereich implementiert werden kann, sieht die Einführung eines Regularisierungsfilters vor, welches den Aufwand der Kompensation in bestimmten Frequenzbereichen beschränkt und somit ihre Positionsempfindlichkeit ebenfalls reduziert (Kirkeby et al. 1998; Kirkeby u. Nelson 1999). Diese Methode lieferte in perzeptiven Untersuchungen von Norcross et al. (2003a) gute Resultate. Außerdem wurde eine Reduktion hörbarer Artefakte im Zeitbereich der kompensierten Funktion festgestellt (Norcross et al. 2004a).

Die hier kurz umrissenen Verfahren wurden zwar einzeln evaluiert, umfangreiche Vergleiche mehrerer Methoden fanden bisher jedoch nur vereinzelt statt (Norcross et al. 2004b). Außerdem ist zu beobachten, dass die Effizienz der Entzerrung in manchen Fällen stark vom Objekt abhängt, auf das sie angewendet wird (Mourjopoulos et al. 1982; Norcross et al. 2002). Es erscheint daher durchaus relevant, eine systematische Untersuchung verschiedener Methoden bezüglich der Kompensation binaural simulierter Signale durchzuführen. Ihr Ziel ist es, eine perzeptiv optimale Filterentwurfsmethode mit möglichst geringem Rechenaufwand und hoher Robustheit gegenüber Positionsänderungen zu finden. Basierend auf den Ergebnissen der genannten Arbeiten wurden mehrere Verfahren ausgewählt und mit Matlab® implementiert. Sie werden nachfolgend ausführlich beschrieben, wobei die zugehörigen Quellcodes in Anhang A zu finden sind.

3.2.2 Implementierung ausgewählter Verfahren

3.2.2.1 Least Squares

Beim Filterentwurf nach dem Kriterium des kleinsten quadratischen Fehlers (*least squares*, LS) wird die Minimierung des Unterschieds zwischen einer erreichbaren und einer idealen Funktion angestrebt. Im hier beschriebenen Kontext der Entzerrung bedeutet dies, dass ein Filter $h_c(n)$ im Zeitbereich gesucht ist, welches den folgenden Ausdruck minimiert (Clarkson et al. 1985):

$$\sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} \left(d(n-m) - h_{\rm eq}(n) \right)^2$$
(3.22)

mit

$$h_{\rm eq}(n) = h(n) * h_{\rm c}(n)$$
 (3.23)

d(n - m) ist hier die Zielfunktion, welche aus den oben genannten Gründen um *m* samples verzögert wird.

Zur Repräsentation der obigen Impulsantworten benutzen Kirkeby und Nelson (1999) die Vektorenschreibweise. Demnach wird beispielsweise h(n) durch den Spaltenvektor **h** mit N_h Elementen dargestellt. Weiterhin kann die Faltung zweier Zeitsignale durch Einführung einer Faltungsmatrix als Matrizenmultiplikation geschrieben werden. Aus Gl. 3.23 wird dann

$$\mathbf{h}_{eq} = \mathbf{h} * \mathbf{h}_{c} = \mathbf{H} \cdot \mathbf{h}_{c} \tag{3.24}$$

mit

$$\mathbf{H} = \begin{bmatrix} h(0) & 0 & \cdots & 0 \\ h(1) & h(0) & \vdots \\ \vdots & h(1) & \ddots & 0 \\ h(N_h - 1) & \vdots & \ddots & h(0) \\ 0 & & \ddots & h(1) \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & h(N_h - 1) \end{bmatrix}$$
(3.25)

Die Faltungsmatrix besitzt die so genannte Toeplitz-Form, das heißt sie hat entlang aller Diagonalen identische Elemente. Bei einer Anzahl von N_c Elementen in \mathbf{h}_c , ist die Anzahl der Zeilen in der Matrix **H** durch $N_h + N_c - 1$ gegeben, während sie N_c Spalten enthält.

Mit dieser Schreibweise wird nun Gl. 3.22 zu

г

$$\mathbf{e}^{\mathrm{T}}\mathbf{e} = (\mathbf{d}_{\mathrm{m}} - \mathbf{H} \cdot \mathbf{h}_{\mathrm{c}})^{\mathrm{T}} (\mathbf{d}_{\mathrm{m}} - \mathbf{H} \cdot \mathbf{h}_{\mathrm{c}})$$
(3.26)

Die Minimierung dieses Ausdrucks führt zur bestmöglichen Kompensation der gesamten Funktion h(n). Es wurde bereits erläutert, dass dies aus Gründen der Filterrobustheit gegenüber Positionsänderungen nicht erwünscht ist. Kirkeby und Nelson (1999) schlagen daher die Erweiterung der obigen Gleichung um einen "effort"-Term vor:

$$I = E + \beta V \tag{3.27}$$

E ist hierbei der quadratische Fehler, während *V* die Energie des Kompensationsfilters kennzeichnet. Letztere stellt ein Maß für den Aufwand dar, der bei der Entzerrung erbracht wird. Bei kleinstmöglichem *I* bestimmt der Faktor β das Gleichgewicht zwischen ausschließlicher Minimierung des Fehlers ($\beta \rightarrow 0$) und alleiniger Minimierung des effort-Terms

 $(\beta \rightarrow \infty)$. Diese so genannte Regularisierung soll nun frequenzabhängig gestaltet werden (Kirkeby u. Nelson 1999). Daher wird **h**_c vor Berechnung seiner Energie mit einem Filter b(n) gefaltet, was zu den folgenden Ausdrücken führt:

$$\mathbf{v}_{\mathbf{b}} = \mathbf{b} * \mathbf{h}_{\mathbf{c}} = \mathbf{B} \cdot \mathbf{h}_{\mathbf{c}} \tag{3.28}$$

$$V = \mathbf{v}_{\mathbf{b}}^{\mathrm{T}} \mathbf{v}_{\mathbf{b}} \tag{3.29}$$

Im Durchlassbereich von b(n) wird der Entzerrungsaufwand sehr stark minimiert, das heißt die Genauigkeit der Linearisierung nimmt dort ab. Die Gln. 3.26 bis 3.29 können zusammengefasst werden zu

$$I = \mathbf{e}^{\mathrm{T}}\mathbf{e} + \beta \mathbf{v}_{\mathbf{b}}^{\mathrm{T}}\mathbf{v}_{\mathbf{b}}$$

= $(\mathbf{d}_{\mathbf{m}}^{\mathrm{T}} - \mathbf{h}_{\mathbf{c}}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}})(\mathbf{d}_{\mathbf{m}} - \mathbf{H} \cdot \mathbf{h}_{\mathbf{c}}) + \beta \cdot \mathbf{h}_{\mathbf{c}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} \cdot \mathbf{B} \cdot \mathbf{h}_{\mathbf{c}}$
= $\mathbf{d}_{\mathbf{m}}^{\mathrm{T}}\mathbf{d}_{\mathbf{m}} - 2 \cdot \mathbf{d}_{\mathbf{m}}^{\mathrm{T}}\mathbf{H} \cdot \mathbf{h}_{\mathbf{c}} + \mathbf{h}_{\mathbf{c}}^{\mathrm{T}} \cdot \mathbf{H}^{\mathrm{T}}\mathbf{H} \cdot \mathbf{h}_{\mathbf{c}} + \mathbf{h}_{\mathbf{c}}^{\mathrm{T}} \cdot \beta \mathbf{B}^{\mathrm{T}}\mathbf{B} \cdot \mathbf{h}_{\mathbf{c}}$
= $\mathbf{h}_{\mathbf{c}}^{\mathrm{T}}(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \beta \mathbf{B}^{\mathrm{T}}\mathbf{B}) \cdot \mathbf{h}_{\mathbf{c}} - 2 \cdot \mathbf{d}_{\mathbf{m}}^{\mathrm{T}}\mathbf{H} \cdot \mathbf{h}_{\mathbf{c}} + \mathbf{d}_{\mathbf{m}}^{\mathrm{T}}\mathbf{d}_{\mathbf{m}}$ (3.30)

Um die Filterkoeffizienten von \mathbf{h}_{c} zu finden, welche den obigen Ausdruck minimieren, muss das folgende lineare Gleichungssystem der Ordung N_{c} gelöst werden:⁸

$$(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \beta \mathbf{B}^{\mathrm{T}}\mathbf{B}) \cdot \mathbf{h}_{\mathbf{c}} = \mathbf{H}^{\mathrm{T}}\mathbf{d}_{\mathbf{m}}$$
(3.31)

Da es sich bei $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ und $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ um quadratische, symmetrische Toeplitz-Matrizen handelt, sind sie invertierbar und das Gleichungssystem hat eine eindeutige Lösung (vgl. Werner 2008, S. 91).

$$\mathbf{h}_{\mathbf{c}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \beta \mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1} \cdot \mathbf{H}^{\mathrm{T}}\mathbf{d}_{\mathbf{m}}$$
(3.32)

Obwohl der Vektor **b** theoretisch beliebig viele Koeffizienten enthalten kann, wird die in ihm enthaltene Information durch die Operation $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ in einer $N_{\mathrm{c}} \times N_{\mathrm{c}}$ -Matrix komprimiert. Dies rührt daher, dass nur die Energie des effort-Terms von Bedeutung ist (s. Gl. 3.29) und führt außerdem zur Feststellung, dass die Phase des Regularisierungsfilters keine Relevanz für die obige Berechnung hat (vgl. Kirkeby u. Nelson 1999, S. 585).

Der Frequenzverlauf von b(n) sollte sich der zu entzerrenden Übertragungsfunktion anpassen. Eine für die hier beschriebene Anwendung sinnvolle Möglichkeit ist daher ein Hoch-

⁸Für den Beweis siehe z.B. (Werner 2008, S. 89 ff.)

passfilter, welches die Kompensationsgenauigkeit der hochfrequenten Einbrüche eingrenzt. Ein solches FIR-Filter wurde, in Anlehnung an das von Lindau et al. (2007) beschriebene, mit einer Sperrdämpfung von 20 dB bis zu 2 kHz entworfen. Die Eckfrequenz wurde als Variable implementiert, welche von der Übertragungsfunktion abhängt, auf die sich die Regularisierung bezieht. Ihr Wert wird von demjenigen dip⁹ mit der tiefsten Frequenz bestimmt, wobei dieser bei praktisch allen gemessenen Kopfhörerübertragungsfunktionen bei ungefähr 8 kHz liegt (s. Abb. 2.5 und 2.5). In Abb. 3.2 sind die so entworfenen Regularisierungsfilter beispielhaft für den K1000- und den Stax I-Kopfhörer dargestellt. Das Filter hat zwei Kanäle, die sich jedoch auf Grund der ähnlichen hochfrequenten Verläufe der Kopfhörerkanäle kaum unterscheiden.



Abbildung 3.2 – Hochpass-Regularisierungsfilter (schwarz) und Betragsfrequenzgänge (linker Kanal: blau, rechter Kanal: grün) zweier Kopfhörer

Norcross et al. (2003a) untersuchten den Effekt unterschiedlicher Regularisierungsfilter, zu denen auch eine terzbreit geglättete Version des zu entzerrenden Frequenzgangs zählte. Die Evaluation dieser Regularisierungsart ergab – zumindest bei geringer Gewichtung – gute Ergebnisse. Die Methode wurde im vorliegenden Fall dahingehend adaptiert, dass die Glättung erstens oktavbreit durchgeführt und zweitens die resultierende Funktion invertiert wurde. Zur Begründung dieser Veränderung wird abermals darauf hingewiesen, dass der geringste Aufwand der Kompensation in dem Bereich erbracht wird, der von der Regularisierung am stärksten betroffen ist. Da es die schmalbandigen Einbrüche sind, die eher ungenau entzerrt werden sollen, muss sich dort der Durchlassbereich von b(n) befinden. Diese Anforderung wird von einer Inverse der Übertragungsfunktion $H(\omega)$ erfüllt und ist sehr individuell an

 $^{^{9}}$ Die Definition eines dips wurde so festgelegt, dass er einen wählbaren Gütefaktor (hier 3) nicht unterschreiten und um mindestens -3 dB vom Mittelwert der Übertragungsfunktion abweichen soll.

jene angepasst (vgl. auch Norcross et al. 2005). Eine eher breitbandige Glättung ist vonnöten, da die Regularisierung auch bei leicht verschobenen Kopfhörerpositionen noch ausreichende Gültigkeit haben soll.

Die entsprechenden Filter sind am Beispiel des Stax II und des Stax Pro in Abb. 3.3 zu sehen. Sie wurden mit dem weiter unten beschriebenen Amplitudenglättungsverfahren erzeugt. Aus Gründen der größeren Übersichtlichkeit sind sie um -15 dB verschoben dargestellt, die Gewichtung von b(n) hängt aber hier und auch beim oben vorgestellten Hochpass vom Faktor β ab. Dieser ist ein frei wählbarer Parameter der LS-Methode und wird in Kapitel 4 näher bestimmt.



Abbildung 3.3 – Inverse, geglättete Regularisierungsfilter (schwarz, zur besseren Darstellung -15 dB), und Betragsfrequenzgänge (linker Kanal: blau, rechter Kanal: grün) zweier Kopfhörer

3.2.2.2 Fast Frequency Deconvolution

Der Entwurf eines geeigneten Entzerrungsfilters kann auch im Frequenzbereich erfolgen. Dies hat den Vorteil eines deutlich geringeren Rechenaufwands, da die im Zeitbereich stattfindenden Faltungsoperationen zum Multiplikationen werden. Kirkeby et al. (1998) beschreiben ein zur oben erläuterten LS-Methode äquivalentes Verfahren, das im Folgenden als *Fast Frequency Deconvolution* (FFD) bezeichnet wird. Die Erweiterung von Gl. 3.21 um den Regularisierungs-Term ergibt analog zu Gl. 3.32 den folgenden Ausdruck:

$$H_{\rm c}(k) = \frac{D(k)H^*(k)}{H(k)H^*(k) + \beta B(k)B^*(k)} = \frac{D(k)H^*(k)}{|H(k)|^2 + \beta |B(k)|^2}$$
(3.33)

D(k) und B(k) sind hier die Frequenzgänge von Zielfunktion respektive Regularisierungsfilter, $H^*(k)$ ist die konjugiert komplexe Funktion von H(k). Der Index k wurde hier verwendet,

da es sich wegen des digitalen Filterentwurfs um diskrete Spektren handelt.

H(k) wird durch Fourier-Transformation der gemessenen Kopfhörerimpulsantwort bestimmt. Kirkeby et al. (1998) schlagen hierzu eine FFT-Ordnung vor, die der gewünschten Filterlänge N_c entspricht. Je nach Größe von N_c , kann dies aber vor allem im Bass zu einer ungenauen Repräsentation des zu kompensierenden Frequenzgangs führen, die sich zwangsläufig auf die invertierte Funktion auswirkt. Es wurde daher die von Leckschat geschilderte Vorgehensweise bevorzugt. Demnach findet die Invertierung selbst mit einer Blockgröße statt, welche "die zu erwartende Länge der Filterimpulsantwort deutlich übersteigt." (Leckschat 1992, S. 35) Anschließend wird $H_c(k)$ zurück in den Zeitbereich transformiert und auf die gewünschte Länge N_c gekürzt. Die zusätzlich nötige Fensterung, welche die Welligkeit im Frequenzgang reduzieren soll, wird weiter unten im Zusammenhang der Nachbearbeitung erörtert.

Ein weiterer Punkt, den es zu beachten gilt, ist eine Dynamikbeschränkung bei der Invertierung (Leckschat 1992; Müller 1999). Zwar werden übermäßige Anhebungen bei hochfrequenten peaks bereits mit Hilfe der Regularisierung vermieden. Allerdings führt die Division durch die Betragsquadrate der sehr kleinen Werte, die am Rande des Übertragungsbereichs von H(k) auftreten, zu extremen und daher untauglichen Ergebnissen. Die Einschränkung der Dynamik kann realisiert werden, indem zunächst der höchste in H(k) enthaltende Pegel identifiziert wird. Werte, die dieses Maximum um einen festgelegten Betrag (hier: 60 dB) unterschreiten, werden nicht invertiert sondern auf den Pegel 60 dB über dem Maximalwert festgesetzt. Dieses *clipping* betrifft nur den Amplitudengang, allerdings folgt die Phase in den betroffenen Randbereichen ohnehin einem zufälligen Verlauf (vgl. Müller 1999, S. 183 f.).

Zur Regularisierung wurden für die FFD-Methode dieselben Filter wie für das LS-Verfahren verwendet. Auch hier stellt β einen noch festzulegenden Parameter dar.

3.2.2.3 Glättung

Die nachfolgend beschriebene Methode zur Kopfhörerentzerrung sieht die Glättung einer Übertragungsfunktion und ihre anschließende Invertierung nach dem FFD-Verfahren (ohne Regularisierung) vor.

$$H_{\rm c}(k) = \frac{D(k)}{H_{\rm smooth}(k)}$$
(3.34)

Bereits die einfache Multiplikation einer Impulsantwort mit einer Fensterfunktion führt zu einer Glättung des Frequenzgangs. Dies ist jedoch ein frequenzlinearer Vorgang, sodass primär der Bassbereich davon betroffen ist. Stattdessen wird hier die frequenzlogarithmische Glättung bevorzugt, welche im einfachsten Fall als gleitende Mittelwertbildung realisiert werden kann. Dabei wird mit wachsender Frequenz über einen immer größeren Bereich des Amplitudengangs gemittelt. "Bessere Resultate liefert eine Fensterfunktion, die mit dem Spektrum gefaltet wird und deren Breite ebenfalls proportional zur Frequenz wächst." (Müller 2008, S. 1127)

Hatziantoniou und Mourjopoulos (2000) bezeichnen die Faltung des Leistungsspektrums $|H(k)|^2$ mit einem Fenster mit zunehmender Breite als "traditionelle" Glättung. Im diskreten Fall handelt es sich hierbei um eine zirkuläre Operation, die von den Autoren folgendermaßen ausgedrückt wird:

$$|H_{\rm sm_amp}(k)|^2 = \sum_{i=0}^{N-1} |H[(k-i) \mod N]|^2 \cdot W_{\rm sm}(m(k), i)$$
(3.35)

m(k) bezeichnet die frequenzabhängige Bandbreite des Glättungsfensters W_{sm} . Eine nach Gl. 3.35 durchgeführte Glättung betrifft lediglich den Amplitudengang von H(k), sodass Phase und Zeitsignal unverändert bleiben. Durch die Faltung mit der komplexen Übertragungsfunktion wirkt die Glättung jedoch auch zeitselektiv.

$$H_{\rm sm_cs}(k) = \sum_{i=0}^{N-1} H[(k-i) \mod N] \cdot W_{\rm sm}(m(k), i)$$
(3.36)

Diese Operation entspricht im Zeitbereich der Multiplikation mit einem Fenster, dessen Breite mit steigender Frequenz abnimmt. Die Komplexität des Zeitsignals wird dadurch reduziert, ohne dass die wichtigen Komponenten der Einschwingphase davon zu sehr betroffen sind (vgl. Hatziantoniou u. Mourjopoulos 2000). Abb. 3.4 verdeutlicht diesen Sachverhalt am Bei-



Abbildung 3.4 – Vergleich der oktavgeglätteten mit der originalen Impulsantwort des Stax Pro (rechter Kanal). Es ist deutlich sichtbar, wie der Einschwingvorgang nahezu unberührt bleibt, da die Glättung erst sukzessive einsetzt.

spiel der oktavgeglätteten Impulsantwort des Stax Pro. Für die angestrebte Robustheit eines Entzerrungsfilters könnte eine derartige Glättung eventuell von Vorteil sein, da sich die Verschiebung der Kopfhörerausrichtung schließlich auch auf das Zeitsignal auswirkt.

Die hier vorgestellte Implementierung des Smoothing-Verfahrens ist auf die Arbeit von Hatziantoniou und Mourjopoulos (2000) gestützt. Dort wird der allgemeine Ausdruck für eine nullphasige Fensterfunktion der Länge *N* wie folgt angegeben:

$$W(k) = b - (1 - b) \cos\left(\frac{2\pi k}{N}\right) \qquad \text{mit} \quad 0 \le k \le N - 1 \tag{3.37}$$

b ist dabei ein Parameter, der die Form des Fensters bestimmt. Bei b = 1 ergibt sich ein Rechteck-Fenster, dessen Flanke mit sinkendem Parameter immer weniger steil wird, wobei b = 0,5 zu einem von-Hann-Fenster führt. Das Glättungsfenster $W_{sm}(m,k)$ in Gl. 3.35 bzw. 3.36 wirkt auf beide Hälften des symmetrischen Spektrums von H(k) und besteht deshalb aus zwei Halbfenstern. Mit der obigen Gleichung wird $W_{sm}(m,k)$ beschrieben durch

$$W_{\rm sm}(m,k) = \begin{cases} \frac{b - (b-1)\cos\left(\frac{\pi k}{m}\right)}{2b(m+1) - 1}, & k = 0, 1, ..., m \\ 0, & k = m+1, ..., N - (m+1) \\ \frac{b - (b-1)\cos\left(\frac{\pi k}{k-N}\right)}{2b(m+1) - 1}, & k = N - m, N - (m-1), ..., N - 1 \end{cases}$$
(3.38)

Die rechte Fensterhälfte glättet das Spektrum zwischen dem tiefsten Frequenzindex und einem festgelegten Punkt *m*, die linke Hälfte besitzt dieselbe Breite und liegt am Ende des symmetrischen Spektrums. *m* kann also maximal den Wert $\frac{N}{2} - 1$ annehmen. In der Mitte zwischen den Halbfenstern ist $W_{sm}(m,k) = 0$. Da die Glättung wie gesagt frequenzabhängig erfolgt, nimmt *m* je nach Frequenzindex *k* verschiedene Werte an. Daher kann $W_{sm}(m(k),k)$ durch die $M \times N$ -Matrix \mathbf{W}_{sm} ausgedrückt werden, die in ihren Zeilen *M* verschiedene Glättungsfenster $W_{sm}(k)$ der Länge *N* enthält. Wie die folgende Ausführung zeigen soll, hängt die Anzahl der Zeilen davon ab, für wieviele unterschiedliche Frequenzpunkte eine Bandbreite berechnet wird.

Die variable Breite des Glättungsfensters kann nach unterschiedlichen Skalen ausgerichtet sein. Oftmals wird die Glättung in Bruchteilen einer Oktave durchgeführt, außerdem sind die Bark- und die ERB-Skala zu nennen, welche an die auditorische Verarbeitung angelehnt sind.¹⁰ Für diese drei genannten Skalierungsmöglichkeiten wurde nach den folgenden For-

¹⁰s. auch Abschnitt 4.1

meln jeweils eine Funktion $P(f_m)$ berechnet, welche die Bandbreite in Abhängigkeit der Mittenfrequenz f_m angibt:

◇ Oktavbruchteil (nach DIN EN 61260):

$$P(f_{\rm m}) = f_{\rm m} \cdot (f_{\rm o} - f_{\rm u})$$

mit
$$f_{\rm o} = \left(10^{3/10}\right)^{0.5 \cdot \text{Oktavbruchteil}}$$
 und $f_{\rm u} = \left(10^{3/10}\right)^{-0.5 \cdot \text{Oktavbruchteil}}$ (3.39)

♦ Bark-Skala (Pohlmann 2005, S. 321):

$$P(f_{\rm m}) = 25 + 75 \cdot \left(1 + 1.4 \cdot (f_{\rm m}/1000)^2\right)^{0.69}$$
(3.40)

◇ ERB-Skala (Moore 1995, S. 175):

$$P(f_{\rm m}) = 24,7 \cdot (4,37 \cdot (f_{\rm m}/1000) + 1) \tag{3.41}$$

Wird im diskreten Fall für jeden Frequenzindex $0 \le k \le \frac{N}{2}$ eine Bandbreite bestimmt, so ist $P(\Delta f \cdot k)$ ein Vektor mit *k* Elementen, welche die Bandbreiten in Hz angeben, wobei $\Delta f = \frac{f_s}{N}$. Daraus kann der Vektor m(k) berechnet werden, dessen einzelne Einträge die Länge eines Halbfensters in Abhängigkeit von *k* bezeichnen.

$$m(k) = 0, 5 \cdot \frac{P(\Delta f \cdot k)}{\Delta f}, \qquad 0 \le k \le \frac{N}{2}$$
(3.42)

Die Matrix **W**_{sm} enthält also $M = \frac{N}{2} + 1$ Zeilen und kann nun vollständig bestimmt werden.

Zur Implementierung der zirkulären Faltung definieren Hatziantoniou und Mourjopoulos zunächst eine quadratische $N \times N$ -Matrix **H**, deren Zeilen die um jeweils ein sample zirkulär verschobene Funktion H(k) enthalten. Die Multiplikation von **H** und **W**_{sm} führt auf die $N \times N$ -Matrix **H**_{sm}, aus welcher die geglättete Funktion $H_{smooth}(k)$ extrahiert werden kann:

$$H_{\text{smooth}}(k) = \mathbf{H}_{\text{sm}}[m(k), k] \qquad \text{mit} \quad \mathbf{H}_{\text{sm}} = \mathbf{W}_{\text{sm}} \cdot \mathbf{H}$$
(3.43)

Da diese Operation relativ rechenaufwändig ist, schlagen die Autoren die getrennte Verarbeitung von Real- und Imaginärteil der zu glättenden Funktion vor. Dadurch muss die Faltung lediglich für $0 \le k \le \frac{N}{2}$ durchgeführt werden, da für $\frac{N}{2} + 1 \le k \le N - 1$ wegen der Symmetrie des Spektrums gilt:

$$\mathcal{R}e\left\{H_{\text{smooth}}(k)\right\} = \mathcal{R}e\left\{H_{\text{smooth}}(N-k)\right\}$$
(3.44)

$$\mathcal{I}m\left\{H_{\text{smooth}}(k)\right\} = -\mathcal{I}m\left\{H_{\text{smooth}}(N-k)\right\}$$
(3.45)

Es wird außerdem darauf hingewiesen, dass zur weiteren Verringerung des Rechenaufwands im Falle der Smoothing-Methode die Fourier-Transformation der Kopfhörerimpulsantwort mit der FFT-Ordnung $N = N_c$ durchgeführt wurde. Die Invertierung von $H_{\text{smooth}}(k)$ erfolgt hier also nicht mit der im obigen Abschnitt beschriebenen hohen Auflösung, da die Funktion bereits auf die Länge des Entzerrungsfilters beschränkt ist. Wegen der ohnehin eher geringen Komplexität des geglätteten Frequenzgangs, schien dies jedoch vertretbar zu sein.

Wenngleich die komplexe Glättung durch ihre Zeitselektivität möglicherweise Vorteile aufweist, ergeben sich bezüglich des geglätteten Amplitudengangs Probleme. Dies sei am Beispiel der gleitenden Mittelwertbildung, als welche der Glättungsvorgang aufgefasst werden kann, veranschaulicht.¹¹ Im Falle der Amplitudenglättung wird das Betragsquadrat einer Übertragungsfunktion über einen bestimmten Bereich gemittelt. Weist das Spektrum dieser Funktion um einen imaginären Nullpunkt herum große Schwankungen auf, so befinden sich "negative" Werte nach der Betragsbildung und Quadrierung über diesem Nullpunkt und der Mittelwert in diesem Abschnitt wird "positiv" sein. Im Gegensatz dazu gehen bei der komplexen Glättung sowohl Werte über als auch unter dem vorgestellten Nullpunkt in die Mittelwertbildung ein, da kein Betragsquadrat gebildet wird. Das Ergebnis wird also niedriger ausfallen, als im vorigen Fall. Bei hoher lokaler Varianz des Spektrums führt die komplexe Glättung daher zu einem abgesenktem Verlauf des Amplitudengangs. Die gemessenen Kopfhörerübertragungsfunktionen weisen insbesondere im hochfrequenten Bereich solche Schwankungen auf. Dort wird der Effekt zusätzlich durch die Tatsache verstärkt, dass die Glättungsbandbreite bei hohen Frequenzen immer größer wird. Das beschriebene Phänomen ist in Abb. 3.5 zu sehen, wo



Abbildung 3.5 – Vergleich von Amplituden- und komplexer Glättung des Stax Pro (rechter Kanal). Letztere zeigt zu hohen Frequenzen hin einen abgesenkten Verlauf.

¹¹Der mathematische Beweis ist bei Hatziantoniou und Mourjopoulos (2000) zu finden.

der Betragsfrequenzgang des Stax Pro im Original sowie nach Amplituden- und komplexer Glättung dargestellt ist.

Im Kontext der Entzerrung ist ein abgesenkter Verlauf der komplex geglätteten Funktion unerwünscht, da ihre Inverse und somit auch der kompensierte Amplitudengang eine Überbetonung hoher Frequenzen aufweisen würden. Hatziantoniou und Mourjopoulos schlagen deshalb die "äquivalente komplexe Glättung" (*equivalent komplex smoothing*) vor. Dabei wird der Amplitudengang von $H_{sm_cs}(k)$ durch denjenigen von $H_{sm_amp}(k)$ ersetzt.

$$H_{\rm sm_eqcmp}(k) = |H_{\rm sm_amp}(k)| \cdot e^{\arg[H_{\rm sm_cs}(k)]}$$
(3.46)

mit

$$|H_{\rm sm_amp}(k)| = \sqrt{\sum_{i=0}^{N-1} |H[(k-i) \mod N]|^2 \cdot W_{\rm sm}(m(k), i)}$$
(3.47)

Bei der Implementierung der Amplitudenglättung wurde die komplexe Funktion $H_{sm_amp}(k)$, ähnlich wie oben, aus $|H_{sm_amp}(k)|$ und der ungeglätteten Phase von H(k) berechnet.

$$H_{\rm sm_amp}(k) = |H_{\rm sm_amp}(k)| \cdot e^{\arg[H(k)]}$$
(3.48)

Es existieren also zwei Varianten der Glättung, aus denen nach Gl. 3.34 ein Entzerrungsfilter berechnet werden kann. Nebst dem Fenster-Parameter *b* kann auch die Smoothing-Bandbreite bei diesem Verfahren frei gewählt werden.

3.2.2.4 Compare and Squeeze

Die von Müller (1999) beschriebene Compare and Squeeze-Methode sieht zur Berechnung eines Kompensationsfilters ebenfalls die Invertierung einer vereinfachten Übertragungsfunktion vor. Die Schritte dieser Simplifizierung sind in Abb. 3.6 zu sehen und werden nachfolgend beschrieben.

Der Frequenzgang von H(k) wird zunächst mit wählbarer Bandbreite geglättet und anschließend der originalen Funktion überlagert (Abb. 3.6, links). Dies bildet die Ausgangslage einer gewichteten Angleichung des geglätteten Amplitudengangs an seinen ursprünglichen Verlauf, wobei die Phase von diesem Vorgang unberührt bleibt. Je nachdem, ob die originale Kurve über oder unter ihrer geglätteten Version liegt, können Überhöhungen und Einbrüche definiert werden. Da erstere durch das Entzerrungsfilter möglichst genau kompensiert werden sollen, wird der geglättete Amplitudengang dort der ursprünglichen Kurve wieder vollstän-



Funktionen fand jedoch mit dem oben beschriebenen FFD-Verfahren (ohne Regularisierung) statt.

3.2.2.5 All-Pol-Modell

In Gl. 3.7 wurde eine Systemfunktion allgemein als Quotient ausgedrückt, dessen Zähler und Nenner die Pole und Nullstellen von H(z) enthalten. Um die Vereinfachung einer Übertragungsfunktion zu erreichen, kann sie durch eine Systemfunktion angenähert werden, welche nur aus Polen oder nur aus Nullstellen besteht. Dabei werden die peaks eines Amplitudengangs grundsätzlich durch die Pole und die notches durch die Nullstellen beschrieben (vgl. Mourjopoulos u. Paraskesvas 1991, S. 283). Um also die Problematik der Klingelartefakte, die bei zu genauer Kompensation von schmalbandigen Einbrüchen auftreten können, zu umgehen, kann das gesuchte Entzerrungsfilter aus dem All-Pol-Modell einer Übertragungsfunktion berechnet werden. Dieses Modell wird durch die folgende Gleichung beschrieben:

$$H_{\text{pole}}(z) = \frac{b(0)}{a(0) \cdot \prod_{k=1}^{p} (1 - d(k) \cdot z^{-1})} = \frac{b(0)}{a(0) + \sum_{k=1}^{p} (a(k) \cdot z^{-k})} = \frac{b(0)}{A(z)}$$
(3.49)

A(z) wird dabei normalisiert, sodass a(0) = 1. Es gilt nun, die Koeffizienten a(k) zu finden, welche das Modell am besten der originalen Funktion H(z) annähern. Dies wird, ähnlich wie im Zusammenhang der LS-Methode beschrieben, durch die Minimierung des quadratischen Fehlers erzielt. Die Vorgehensweise hierzu ist beispielsweise bei Hayes (1996) ausführlich erläutert und wird nachfolgend zusammengefasst.

Für den Fehler E'(z) gilt

$$E'(z) = H(z) - H_{\text{pole}}(z) = H(z) - \frac{b(0)}{A(z)}$$
(3.50)

1 (0)

Durch Multiplikation beider Seiten mit A(z) wird daraus

$$E(z) = A(z) \cdot E'(z) = A(z) \cdot H(z) - b(0)$$

= $\left(1 + \sum_{k=1}^{p} (a(k) \cdot z^{-k})\right) \cdot H(z) - b(0)$ (3.51)

Im Zeitbereich entspricht dies für n > 0 einer Faltung der Impulsantwort h(n) mit dem rekursiven Teil des Modells.

$$e(n) = h(n) + \sum_{l=1}^{p} a(l) \cdot h(n-l)$$
(3.52)

Der quadratische Fehler ist also:

$$\sum_{n=1}^{N-1} e^2(n) = \sum_{n=1}^{N-1} \left[h(n) + \sum_{l=1}^p a(l) \cdot h(n-l) \right]^2$$
(3.53)

Da h(n) = 0 für n < 0, ist der Fehler bei n = 0 gegeben durch h(0) und hängt folglich nicht von a(k) ab. Die gesuchten Koeffizienten minimieren Gl. 3.53 daher nicht nur für $n \ge 1$ sondern auch für $n \ge 0$. Sie können berechnet werden, indem die folgende Gleichung gelöst wird (vgl. Hayes 1996, S. 145 f.)

$$\sum_{l=1}^{p} a(l) \left[\sum_{n=0}^{N-1} h(n-l)h^*(n-k) \right] = -\sum_{n=0}^{N-1} h(n)h^*(n-k)$$
(3.54)

 $h^*(n)$ ist dabei die konjugiert komplexe Entsrpechung von h(n). Durch Definition der Autokorrelationsfolge

$$r_h(k-l) = \sum_{n=0}^{N-1} h(n-l)h^*(n-k)$$
(3.55)

kann Gl. 3.54 auch ausgedrückt werden als

$$\sum_{l=1}^{p} a(l) \cdot r_h(k-l) = -r_h(k), \qquad k = 1, ..., p$$
(3.56)

Mit der Matrix-Schreibweise wird aus $r_h(k-l)$ die Autokorrelationsmatrix $\mathbf{R}_{\mathbf{h}}$

$$\mathbf{R_h} = \begin{bmatrix} r_h(0) & r_h^*(1) & \cdots & r_h^*(p-1) \\ r_h(1) & r_h(0) & \cdots & r_h^*(p-2) \\ \vdots & \vdots & & \vdots \\ r_h(p-1) & r_h(p-2) & \cdots & r_h(0) \end{bmatrix}$$
(3.57)

Da es sich bei h(n) um eine reelle Folge handelt, gilt $r_h = r_h^*$. Somit ist \mathbf{R}_h eine quadratische, symmetrische Toeplitz-Matrix,¹² und zur Bestimmung der Koeffizienten a(k) kann ein lineares Gleichungssystem der Ordnung p gelöst werden:

$$\mathbf{R}_{\mathbf{h}} \cdot \mathbf{a} = -\mathbf{r}_{\mathbf{h}} \tag{3.58}$$

Die Spaltenvektoren **a** und **r**_h enthalten *p* Elemente, wobei h(n) mit steigender Ordnung immer genauer approximiert wird.

Der Koeffizient b(0), der nichts anderes als ein Skalierungsfaktor ist, wird laut Hayes (1996) festgelegt auf $b(0) = \sqrt{\epsilon}$, wobei ϵ den minimalen quadratischen Fehler kennzeichnet:

$$\epsilon = r_h(0) + \sum_{k=1}^p a(k) r_h^*(k)$$
 (3.59)

Das hier beschriebene Verfahren zur Bestimmung der Koeffizienten eines All-Pol-Modells der Ordnung p ist in der Funktion aryule1 der Signal Processing ToolboxTM von Matlab® implementiert. Um ein Entzerrungsfilter zu berechnen, wurde mit Hilfe der resultierenden Koeffizienten a(k) und b(0) die Funktion $H_{\text{pole}}(z)$ bestimmt, die anschließend, wie in den zwei zuletzt beschriebenen Verfahren, nach der FFD-Methode (ohne Regularisierung) invertiert wurde. Beachtenswert ist, dass es sich bei $H_{\text{pole}}(z)$ um eine minimalphasige IIR-Funktion handelt, da sie lediglich bei z = 0, also innerhalb des Einheitskreises, Nullstellen aufweist. Ihre Inverse ist deshalb stets eine stabile und kausale FIR-Funktion.

¹²s. LS-Methode

3.2.2.6 Vor- und Nachbearbeitung der Impulsantworten

In den oben dargestellten Verfahren zum Entwurf digitaler Kompensationsfilter dient stets die zu entzerrende Impulsantwort h(n) als Grundlage der Koeffizientenberechnung. Es stellt sich also im Zusammenhang der zehn gemessenen Kopfhörerübertragungsfunktionen (Abschnitt 2.2) die Frage, wonach ihre Kompensation ausgerichtet sein soll. Die Auswahl eines spezifischen Frequenzgangs erscheint unangemessen, zumal die Einzelmessungen nicht reproduzierbar sind, sondern lediglich einen Überblick über mögliche Frequenzverläufe geben. Es wurde daher eine Mittelung der zehn Messungen durchgeführt und das Ergebnis als Basis der Filterberechnung verwendet. Dies führt zwar bei keiner der zehn gemessenen Frequenzgänge zu einer genauen Entzerrung, vor dem Hintergrund der geschilderten Verschiebungsproblematik ist dies aber auch nicht wünschenswert. Die folgende Formel beschreibt die arithmetische Mittelung, in welche n Funktionen eingehen (Müller 2008, S. 1129):

$$A_n = A_{n-1} \left(1 - \frac{1}{n} \right) + A_{\text{neu}} \cdot \frac{1}{n}$$
(3.60)

 A_{n-1} ist das Ergebnis einer mit n-1 Funktionen durchgeführten Mittelung, A_{neu} ist eine neu hinzukommende Funktion. Bei räumlichen Mittelungen über mehrere Messpunkte mit unterschiedlichen Laufzeiten muss zur Vermeidung von Interferenzeffekten ein energetische Mittelung erfolgen (vgl. Müller 2008, S. 1130). Im Falle der Kopfhörer gingen jedoch die komplexen Übertragungsfunktionen in Gl. 3.60 ein, sodass auch die Zeitsignale gemittelt wurden. Das Ergebnis dieser Bearbeitung ist in Abb. 3.7 anhand des Stax Pro zu sehen.



Abbildung 3.7 – Komplexe, arithmetische Mittelung am Beispiel des Stax Pro (linker Kanal). Die grauen Linien zeigen die zehn Einzelmessungen, die gemittelte Kurve ist schwarz dargestellt.

Bei der Messung von Übertragungsfunktionen mit Sweep-Signalen, wie sie auch im vorliegenden Fall zum Einsatz kamen, ist zur Entfernung von Störkomponenten die nachträgliche Fensterung des Zeitsignals nötig (vgl. Müller 2008, S. 1115). Um den Hauptbestandteil des Zeitsignals dabei möglichst unverändert zu belassen, können laut Müller rechte Fensterhälften eingesetzt werden. Abb. 3.8(a) zeigt das hier verwendete Blackmanharris-Fenster, dessen Amplitude für die halbe Fensterlänge zu 1 gesetzt wurde. Die gemittelten Impulsantworten wurden damit auf eine Länge von 4096 samples gekürzt. In Abb. 3.8(b) ist am Beispiel des Stax II sichtbar, dass zu späteren Zeitpunkten nur noch Rauschen auftritt.



Abbildung 3.8 – Vorfensterung: (a) rechter Kanal des gemittelten Stax II zusammen mit der verwendeten Fensterfunktion, (b) ETC des Stax II bis 2¹⁴ samples. Nach der eingezeichneten Grenze von 4096 samples enthält die Impulsantwort keine relevanten Informationen mehr.

Im Zusammenhang der FFD-Methode wurde bereits erwähnt, dass nach Berechnung der N_c Filterkoeffizienten eine Fensterung der Impulsantwort nötig ist. In Abb. 3.9(a) sind verschiedene hierfür denkbare Fensterfunktionen sichtbar. Zwei Ausschnitte des Bassbereichs der dazugehörigen Entzerrungsergebnisse¹³ sind in Abb. 3.9(b) zu sehen. Es wird deutlich, dass zwischen Hauptkeulenbreite und Nebenkeulenunterdrückung abgewägt werden muss, da bei hoher Dämpfung der Roll-off des Zielbandpasses unzureichend angenähert wird. Anhand des dargestellten Vergleichs wurde für die nachträgliche Fensterung der Filterimpulsantworten ein Tukey-Fenster – die Kombination einer Rechteck- mit einer Fensterfunktion – gewählt. Dessen Parameter r gibt an, zu wieviel Prozent sich das Rechteck einem von-Hann-Fenster angleicht und wurde auf 0,75 festgesetzt.

Die Fensterung einer Impulsantwort führt oft zu einem unerwünschten Gleichanteil, da sich ihre positiven und negativen Halbwellen nicht mehr ausgleichen. Dies kann durch die

¹³Faltung der gefensterten Filterimpulsantworten mit h(n)

von Müller beschriebene "Colva-Fensterung" vermieden werden. Dabei wird von der gefensterten Impulsantwort ein weiteres Fenster subtrahiert, dessen Abtastwert-Summe dem unerwünschten Gleichanteil entspricht (vgl. Müller 1999, S. 189 ff.). Eine solche DC-Korrektur wurde bei der hier beschriebenen Implementierung der Filterentwurfsmethoden nach jeder durchgeführten Fensterung angewendet.



Abbildung 3.9 – Nachträgliche Fensterung: (a) Verschiedene Fensterfunktionen über dem rechten Kanal des Stax II, (b) Entzerrungsergebnis im Bassbereich bei Anwendung der Fensterfunktionen aus (a)

3.2.3 Zusammenfassung

Abb. 3.10 zeigt einen Überblick über die im vorangegangenen Abschnitt beschriebenen Entzerrungsverfahren und ihre Parameter. Ein systematischer Vergleich der Methoden soll Aufschluss über ihre Eignung zur Kompensation von Frequenzgängen in der Binauraltechnik geben. Dabei sind mehrere Ebenen der Gegenüberstellung denkbar. Einerseits können Zeit- und Frequenzbereichsverfahren, repräsentiert durch LS- und FFD-Methode, verglichen werden. Weiterhin lässt sich eine Gruppierung vornehmen, wie sie für die Abbildung gewählt wurde. Die Regularisierungsverfahren (LS und FFD) erzielen während der Filterberechnung selbst eine Reduktion der Kompensationsgenauigkeit. Die restlichen Verfahren beruhen hingegen auf einem *pre-processing* der zu kompensierenden Funktion, dessen vereinfachtes Ausgangssignal anschließend im Frequenzbereich invertiert wird.¹⁴ Beides soll zu einer erhöhten Robustheit gegenüber Verschiebungen der Abhörposition führen. Auf einer dritten Ebene könnten die minimal- und die gemischtphasige Kompensation gegenübergestellt werden, was im Rahmen dieser Arbeit jedoch aus oben genannten Gründen nicht erfolgte.

¹⁴Selbstverständlich wäre auch die Inversion nach der LS-Methode denkbar. Wegen ihres höheren Rechenaufwandes wurde sie jedoch hier nicht eingesetzt.



Abbildung 3.10 - Zusammenfassung der Filtermethoden und ihrer Parameter

Die in der Abbildung sichtbare zusätzliche Gliederung soll den Vergleich weiterer Aspekte veranschaulichen. Erstens ist von Interesse, welches der vorgestellten Regularisierungsfilter besser zur Kompensation von Kopfhörerfrequenzgängen geeignet ist. Zudem soll geklärt werden, ob sich möglicherweise eine der beschriebenen Vereinfachungen von H(k) besonders von den anderen oder der Regularisierung abhebt und welche der beiden Smoothing-Methoden zu bevorzugen ist. Die acht dargestellten Varianten werden nachfolgend der besseren methodischen Vergleichbarkeit wegen als einzelne Filteralgorithmen behandelt, obwohl sie streng genommen lediglich auf fünf Verfahren beruhen.

4 EVALUATION

4.1 **Objektive Evaluation**

Um die Effizienz der in Abschnitt 3.2 beschriebenen Entzerrungsmethoden bereits vor ihrer empirischen Evaluation einschätzen zu können, wurde der Rechenaufwand der einzelnen Verfahren untersucht sowie die Güte der Kompensation mit Hilfe von Fehlermaßen ermittelt. Die Faltung der Kopfhörer- mit den Filterimpulsantworten lieferte für letztere die Ausgangslage. Ein berechneter Wert ist zwar ein wichtiger Aspekt der Bewertung, er lässt aber keinen simplen Schluss auf die perzeptive Beurteilung zu. Der hier verwendete Begriff der Objektivität impliziert daher nicht, dass eine perzeptive Evaluation weniger aussagekräftig wäre. Er soll lediglich den Vorgang der Berechnung abgrenzen von der subjektiven Bewertung, die in dem in Abschnitt 4.2 beschriebenen Hörversuch stattfand.

Aus zeitlichen Gründen konnten nicht bei allen Methoden perzeptive Vorversuche zur Ermittlung geeigneter Filterparameterwerte durchgeführt werden. Auch musste die Anzahl der Verfahren für den Hörversuch eingeschränkt werden. Auf Grundlage der berechneten Faltungsergebnisse und Fehlermaße wurde deshalb eine Vorauswahl der Verfahren und ihrer Parametergrößen für die perzeptive Evaluation getroffen.

Es wurde bereits angeführt, dass die Güte einer Kompensation unter Umständen von der Übertragungsfunktion abhängt, die es zu invertieren gilt. Daher wurden sowohl bei der objektiven Evaluation als auch im Hörversuch die Entzerrungsergebnisse zweier Kopfhörer miteinander verglichen. Wie in Abschnitt 1.3 erwähnt, sind nur FEC-Kopfhörer für die Wiedergabe binaural synthetisierter Signale geeignet. Unter diesem Gesichtspunkt kamen die Stax-Kopfhörer und der AKG K-1000 für eine Gegenüberstellung in Frage. Bei letzterem spielen jedoch zusätzlich Aspekte des Raumeinflusses und des Übersprechens eine Rolle (s. Abschnitt 2.2), sodass die Kompensation mit einem Entzerrungsfilter allein eventuell nicht ausreichend wäre für die korrekte Wiedergabe der Ohrsignale. Da der Stax I und der Stax II sehr ähnliche Frequenzgänge aufweisen, und ihr Vergleich somit keine wertvollen Erkenntnisse geliefert hätte, fiel die Wahl auf den Stax II und den Stax Pro. Letzterer hebt sich insbesondere durch die kaum vorhandene Bassanhebung von den anderen Stax-Modellen ab, seine
Empfindlichkeit gegenüber veränderten Aufsetzpositionen scheint im hochfrequenten Bereich jedoch etwas stärker ausgeprägt zu sein (s. Abb. 2.11(c)).

4.1.1 Fehlermaße

Das Ergebnis einer Entzerrung entspricht im Idealfall der vorher festgelegten Zielfunktion. Um die Abweichungen der kompensierten Impulsantworten bzw. Frequenzgänge von dieser Zielfunktion zu quantifizieren, wurde sowohl für den Frequenz- als auch den Zeitbereich ein Fehlermaß definiert. Obwohl ersteres gewisse psychoakustische Kriterien mit einschließt, sollen sie jedoch wie gesagt keinen Ersatz für Hörversuche darstellen. Die Entwicklung auditorischer Modelle ist ein eigener Aspekt der Forschung, als Beispiel sei hier die Arbeit von Dau et al. (1996) genannt, welche auf die Vorhersage von auditiven Schwellwerten abzielt. Mehrere Verarbeitungsstufen simulieren darin unter anderem die Frequenzselektivität der Basilarmembran sowie die Transformation ihrer mechanischen Schwingungen in die Reizimpulse der Haarzellen. Ein ähnliches Modell wurde von Huopaniemi und Smith (1999) zur gehörgerechten Evaluation von unterschiedlich angenäherten HRTFs eingesetzt, wobei hier auch die Lautheitswahrnehmung einbezogen wurde. Eine derartig genaue Nachbildung der auditorischen Verarbeitung hätte den Rahmen dieser Arbeit gesprengt und war wegen des durchgeführten Hörversuchs auch nicht vonnöten.

4.1.1.1 Zeitbereich

Im Zeitbereich wurde die Abweichung von der Zielfunktion, wie von Mourjoupoulos et al. (1982) vorgeschlagen, als Leistung des Fehlers berechnet:

$$e = \frac{1}{N} \sum_{n=0}^{N-1} \left(h_{\text{eq}}(n) - d(n) \right)^2$$
(4.1)

Für eine gehörgerechte Beurteilung müssten zusätzlich Aspekte der zeitlichen Maskierung berücksichtigt werden. Fielder (2001) schlägt hierzu beispielsweise als Kompromiss aus mehreren Forschungsergebnissen ein rückwärts gerichtetes Zeitintervall von 15 ms vor, innerhalb dessen von einer Vorverdeckung durch den Hauptimpuls von $h_{eq}(n)$ ausgegangen werden kann. Weiterhin definiert er eine Dauer von 4 ms, während der simultane Verdeckung auftritt, die anschließend mit 35 dB/Dekade abklingt. Wie der Autor selbst feststellt, hängen solche Werte jedoch stark von den involvierten Signalen ab. Kriterien, welche Maskierungseffekte betreffen, gingen im vorliegenden Fall nicht in die Berechnung eines Zeitfehlermaßes ein. Sie können aber bei der allgemeinen Beurteilung der berechneten Kompensationsergebnisse herangezogen werden.

4.1.1.2 Frequenzbereich

Bei der Bestimmung eines Fehlermaßes für den Frequenzbereich wurde die Art und Weise der Frequenzanalyse durch das Gehör mit berücksichtigt. Hier spielen die so genannten auditorischen Filter eine wichtige Rolle, wie bereits 1940 von Harvey Fletcher festgestellt wurde (vgl. Moore u. Glasberg 1983). Er beschrieb die menschliche Frequenzselektivität als eine Filterbank aus überlappenden Bandpassfiltern. Deren konstante relative Breite wird als *critical band* oder Frequenzgruppe bezeichnet und ist die Basis der von Zwicker definierten Bark-Skala.¹ Glasberg, Moore und Patterson untersuchten in zahlreichen Experimenten die Form der auditorischen Filter und charakterisierten sie durch den folgenden Ausdruck (vgl. Moore 1995; Salomons 1995):

$$C(f, f_{\rm c}) = \left(1 + \frac{4|f - f_{\rm c}|}{W(f_{\rm c})}\right) \cdot \exp\left(-\frac{4|f - f_{\rm c}|}{W(f_{\rm c})}\right)$$
(4.2)

Dabei ist f_c die Mittenfrequenz eines Filters und $W(f_c)$ die *equivalent rectangular bandwidth* (ERB), welche die Bandbreite eines rechteckförmigen Filters angibt, in das dieselbe Energie fällt wie in ein auditorisches Filter bei identischer Mittenfrequenz. Es ist also

$$\int_{-\infty}^{\infty} C(f, f_{\rm c})df = W(f_{\rm c}).$$
(4.3)

Als Berechnungsformel für die ERB gibt Moore (1995, S. 175) für mittlere Pegel

$$ERB = 24,7(4,37F_c + 1)$$
(4.4)

an, wobei F_c die Mittenfrequenz eines auditorischen Filters in kHz ist. Die ERB unterscheidet sich besonders im tiefen Frequenzbereich von den Bandbreiten der Bark-Skala, da letztere ab ca. 500 Hz konstant bleiben, während die ERB mit sinkender Frequenz weiter abnimmt (Moore u. Glasberg 1983).

Um die Auswirkung der auditorischen Filterbank auf ein Signal S(f) zu beschreiben, lässt sich laut Salomons (1995) das so genannte interne Leistungsspektrum dieses Signals berechnen:

$$|S_{\rm c}(f_{\rm c})|^2 = \frac{\int\limits_{-\infty}^{\infty} C(f, f_{\rm c})|S(f)|^2 df}{\int\limits_{-\infty}^{\infty} C(f, f_{\rm c}) df}$$
(4.5)

¹Für eine genaue Beschreibung siehe (Zwicker u. Fastl 1999, S. 149-174)

Aus dem Leistungsspektrum $|S(f)|^2$ wird also das von den Mittenfrequenzen der auditorischen Filter abhängige interne Leistungsspektrum $|S_c(f_c)|^2$, das auf die jeweilige ERB normiert ist (vgl. Gl. 4.3). Zur Bestimmung von Klangverfärbungen in einem Signal schlägt Salomons die Berechnung der maximalen Modulationstiefe – der Unterschied zwischen dem Pegel des globalen Maximums und Minimums – des internen Leistungsspektrums vor. Sie definiert daraufhin auf Grundlage empirischer Daten eine Schwelle, ab der eine Veränderung der Klangfarbe wahrgenommen wird (A_0 -Kriterium; Salomons 1995, S. 73 ff.). Dieser Ansatz wurde zur Berechnung des Fehlermaßes im Frequenzbereich adaptiert, indem statt der maximalen Modulationstiefe innerhalb eines Signals der Unterschied zwischen den internen Leistungsspektren der entzerrten Übertragungsfunktion und des Zielbandpasses bestimmt wurde.

$$E(f_{c}) = 10 \log \left(\frac{\int_{-\infty}^{\infty} C(f, f_{c}) |H_{eq}(f)|^{2} df}{\int_{-\infty}^{\infty} C(f, f_{c}) df} \right) - 10 \log \left(\frac{\int_{-\infty}^{\infty} C(f, f_{c}) |D(f)|^{2} df}{\int_{-\infty}^{\infty} C(f, f_{c}) df} \right)$$

= $10 \log \left(|H_{eq_{c}c}(f_{c})|^{2} \right) - 10 \log \left(|D_{c}(f_{c})|^{2} \right)$ (4.6)

Die obigen Integrale werden im diskreten Fall zu Summen, deren Grenzen hier auf die Eckfrequenzen des Zielbandpasses festgesetzt wurden, da außerhalb dieser eine Fehlerberechnung wenig sinnvoll erschien. Bei einer Anzahl von *n* Mittenfrequenzen in der auditorischen Filterbank besteht $E(f_c)$ also aus *n* Werten, die für jedes Band die Abweichung des kompensierten Spektrums von der Zielfunktion angeben. Daraus kann in Anlehnung an Gl. 4.1 ein Gesamtfehler berechnet werden:

$$E_{\rm tot} = \frac{1}{n} \sum_{c=1}^{n} |E(f_c)|$$
(4.7)

Es ist wichtig zu bemerken, dass hierbei der Unterschied in der Wahrnehmung von peaks und dips nicht berücksichtigt wird. Dies könnte beispielsweise durch die Gewichtung der Überhöhungen mit einem Faktor > 1 geschehen. Allerdings hängt die Detektierbarkeit von peaks und notches außer von ihrem Pegel auch von ihrer Breite ab und ist zudem über der Frequenz variabel, was die Wahl eines festen Faktors erschwert. In den erwähnten Untersuchungen zu diesem Thema fand lediglich bei Bücklein – sehr eingeschränkt – ein direkter Vergleich der Wahrnehmungsschwellen von Anhebungen und Einbrüchen statt. Dessen Ergebnisse implizieren nur bis ca. 1 kHz einen linearen Zusammenhang. Auditorische Modelle, wie sie oben beschrieben wurden, scheinen diesen Aspekt ebenfalls nicht einzubeziehen. Es fand bei der Berechnung von E_{tot} daher keine Gewichtung statt, diese Thematik könnte jedoch bei Weiterentwicklung des Frequenzfehlermaßes näher bearbeitet werden. Die Berechnung der ERB-Filter $C(f, f_c)$ in Gl. 4.6 erfolgte mit Hilfe der Auditory Toolbox für Matlab® (Slaney 1993; Slaney 1998). Die Anzahl der Bänder ist dort ein frei wählbarer Parameter und wurde auf Basis der folgenden Gleichung berechnet (Moore 1995, S. 176):

$$n(F) = 21, 4 \cdot \log(4, 37F_{\rm c} + 1) \tag{4.8}$$

Der Ausdruck gibt an, dass bei der Mittenfrequenz *F* (in kHz) das *n*-te Band der auditorischen Filterbank wirksam ist. Die Anzahl der zwischen 50 Hz und 21 kHz – den Eckfrequenzen des Zielbandpasses – aktiven Bänder wurde daher folgendermaßen berechnet:

$$n(21) - n(0,05) = 21, 4 \cdot \log(4, 37 \cdot 21 + 1) - 21, 4 \cdot \log(4, 37 \cdot 0, 05 + 1)$$

= 42, 1 - 1, 8 = 40, 3 \approx 40 (4.9)

Dies ist mit der von Huopaniemi und Smith (1999) verwendeten Anzahl von Bändern vergleichbar, die n = 44 bei einer oberen Grenzfrequenz von 25 kHz wählten. Die in dieser Arbeit eingesetzte auditorische Filterbank ist in Abb. 4.1 dargestellt.



Abbildung 4.1 – ERB-Filterbank mit 40 Bändern zwischen 50 Hz und 21 kHz

Der Quellcode zur Berechnung der beschriebenen Fehlermaße kann in Anhang A eingesehen werden.

4.1.2 Parameterauswahl

In Abschnitt 3.2 wurde deutlich, dass die Variation der Filterparameter bei allen beschriebenen Kompensationsmethoden auf deren Robustheit gegenüber Positionsverschiebungen des Kopfhörers abzielt. Um diesen Gesichtspunkt bei der Auswahl von passenden Parameterwerten einzubeziehen, wurden die Filterimpulsantworten, die ja auf einer Mittelung aus zehn gemessenen Kopfhörerimpulsantworten beruhen, jeweils mit den zehn Messungen² einzeln gefaltet. Dies sollte die reale Situation, bei der Versuchspersonen die Kopfhörer auf unterschiedliche Weise aufsetzen, simulieren. Die Faltungsergebnisse sind also einerseits ein Indikator für die erzielbare Robustheit einer Kompensationsmethode und dienten gleichzeitig zur Bestimmung diesbezüglich geeigneter Parameterwerte. Falls nicht anders angegeben, beträgt die Filterlänge bei den anschließend vorgestellten Berechnungen jeweils 2¹¹ samples, da auch im Hörversuch nur Filter dieser Ordnung eingesetzt wurden. Auf die Gründe hierfür wird in Abschnitt 4.2 eingegangen.

Im Falle der LS- und FFD-Methode sollte der Parameter β in einem perzeptiven Vorversuch bestimmt werden, da hier laut Norcross et al. (2003a) ein optimaler Wert für die Entzerrungsgüte von großer Bedeutung ist. Dieser hänge von der zu invertierenden Übertragungsfunktion ab, sodass bei jeder neuen Kompensation die vorherige empirische Ermittlung von β nötig sei. Die beiden Entzerrungsverfahren involvieren also einen relativ großen Aufwand, was die Festlegung der Regularisierungsstärke angeht. Im Falle der vorliegenden Untersuchung wurde davon abgesehen, für beide Kopfhörer ein eigenes β zu bestimmen. Die Unterschiedlichkeit beider Übertragungsfunktionen hält sich in Grenzen, sodass von einer zumindest ähnlichen optimalen Gewichtung ausgegangen werden kann. Zudem hätten individuelle Werte einen direkten Vergleich der Kopfhörer erheblich verkompliziert, da die Güte ihrer Kompensation dadurch nicht nur von der Filtermethode selbst sondern auch von variablen Parametergrößen innerhalb eines Verfahrens abhängig gewesen wäre. Dieselben Uberlegungen treffen übrigens auf die einzelnen Kanäle zu. Mit Blick auf die Messergebnisse in Abschnitt 2.2, könnten für rechten und linken Kanal jeweils eigene Parameterwerte eingesetzt werden. Die sich dabei ergebenden diversen Kombinationsmöglichkeiten würden das erwähnte methodische Problem jedoch weiter vergrößern. In zukünftigen Untersuchungen wären diese Aspekte mit mehr zeitlichem und methodischem Aufwand möglicherweise einzubeziehen.

Für den geplanten Vorversuch zur Bestimmung von β wurden auf Grundlage der nachfolgend beschriebenen Ergebnisse vier Werte ausgewählt. Da für die LS- und die FFD-Methode identische Parameter zur Verfügung stehen, wurden die beiden Verfahren dabei nicht getrennt behandelt.

In Abb. 4.2 sind die Resultate einer FFD-Entzerrung beider Kopfhörer mit dem in Abschnitt 3.2 beschriebenen Hochpass-Regularisierungsfilter und verschiedenen Gewichtungsfaktoren als graue Kurven dargestellt. In der oberen Grafik sind jeweils die linken Kanäle,

²Diese wurden vorab wie in Abschnitt 3.2 beschrieben auf 2¹² samples gekürzt und gefenstert.



Abbildung 4.2 – FFD-Kompensation mit Hochpass-Regularisierung und verschiedenen β für linken (oben) und rechten (unten) Kanal. Die grauen Linien zeigen jeweils zehn entzerrte Übertragungsfunktionen, die schwarze Linie ist der Zielbandpass.



Abbildung 4.3 – FFD-Kompensation des Stax Pro mit einer Filterlänge von 2^{15} samples ($\beta = 0, 4$)

in der unteren Grafik die rechten Kanäle zu sehen. Die schwarze Linie zeigt den Zielbandpass. Auffällig sind die "Ausreißer" im tieffrequenten Bereich des linken Stax Pro-Kanals und, etwas schwächer, des rechten Stax II-Kanals. Diese können auch in den Messergebnissen in Abschnitt 2.2 beobachtet werden und hängen mit den Positionsänderungen der Kopfhörer zusammen. Offenbar ist die starke Anhebung, welche hier zur Angleichung an den Zielbandpass nötig wäre, vom Kompensationsfilter nicht zu leisten. Abb. 4.3 zeigt, dass sich diese Problematik auch mit einer Filterlänge von 2¹⁵ samples nicht ausreichend eliminieren lässt.

Die zu den Kompensationsergebnissen gehörenden frequenzabhängigen Fehlermaße $E(f_c)$ sind in Abb. 4.4 aufgetragen. Für jede der zehn Entzerrungen ist der Fehler in dB in den 40 Bändern der auditorischen Filterbank zu sehen. Wie bereits festgestellt, verändert sich der



Abbildung 4.4 – Fehler in dB im Frequenzbereich bei FFD-Kompensation mit Hochpass-Regularisierung und verschiedenen β für linken (oben) und rechten (unten) Kanal. Für alle zehn Entzerrungen ist die Abweichung vom Zielbandpass in den 40 Frequenzbändern der auditorischen Filterbank zu sehen.

Frequenzgang des Stax Pro bei Neuausrichtung des Kopfhörers grundsätzlich stärker als derjenige des Stax II, was in einer entsprechenden Varianz der Entzerrungsgüte resultiert.

Wenig überraschend ist die Tatsache, dass bei stärkerer Regularisierung die peaks im hohen Frequenzbereich abnehmen, dafür die dips aber tiefer und breiter werden. Dies spiegelt sich auch in Tabelle 4.1 wider, wo für beide Kopfhörer und Kanäle das Maximum und das Minimum aus allen zehn Fehlervektoren $E(f_c)$ für mehrere β -Werte eingetragen sind. Dies soll Aufschluss darüber geben, welche positiven und negativen Abweichungen vom Zielbandpass bei einem bestimmten β maximal auftreten können. Der praktisch gleichbleibende Wert für das Minimum im linken Kanal des Stax Pro bezieht sich offensichtlich auf die große Abweichung im Bass und ist deshalb nicht besonders aussagekräftig bezüglich des Einflusses von β . Genauso verhält es sich mit dem Maximum des rechten Kanals des Stax II, welches auf eine unveränderte Überhöhung bei ca. 90 Hz (s. Abb. 4.2 und Abb. 4.4) zurückzuführen ist. Wie bereits in Abschnitt 2.2 bemerkt wurde, verschiebt sich die dem Stax II eigene Bassanhebung bei wechselnder Aufsetzposition des Kopfhörers ziemlich stark. Weicht sie, wie es hier der Fall ist, zu sehr von dem Mittel ab, welches als Ausgangslage für das Entzerrungsfilter diente, lässt sie sich nicht ausreichend kompensieren.

		Stax Pro		Stax II	
		links	rechts	links	rechts
$\beta = 0, 1$	$\max[E(f_c)]$	3,42	1,81	1,97	1,86
	$\min[E(f_c)]$	-7,50	-4,42	-3,61	-4,16
$\beta = 0, 2$	$\max[E(f_c)]$	2,70	1,27	1,48	1,86
	$\min[E(f_c)]$	-7,52	-5,27	-4,51	-5,56
$\beta = 0, 3$	$\max[E(f_c)]$	2,04	1,07	1,11	1,86
	$\min[E(f_c)]$	-7,54	-5,92	-5,30	-6,52
$\beta = 0, 4$	$\max[E(f_c)]$	1,64	1.00	1,03	1,85
	$\min[E(f_c)]$	-7,56	-6,55	-6,00	-7,28

Tabelle 4.1 – Fehlerwerte in dB bei FFD-Entzerrung mit Hochpass-Regularisierung und verschiedenen β . Es sind die größten positiven sowie negativen Abweichungen vom Zielbandpass aus zehn Kompensationen aufgelistet.

Es ist offenbar notwendig, bei der Wahl des β -Wertes einen Kompromiss zwischen zu großen Überhöhungen und zu tiefen Einbrüchen zu finden. Unklar ist dabei, wie stark sich die jeweiligen Abweichungen auf die perzeptive Beurteilung auswirken. Die hier dargestellten Werte sind nicht direkt mit den Ergebnissen von Bücklein (1981), Toole und Olive (1988) oder Moore et al. (1989) zu vergleichen, da dort die Größe der hörbaren peaks und notches ohne Gewichtung mit einer perzeptiven Filterbank angegeben wurden. Ohne Zweifel bleibt indes, dass Überhöhungen deutlicher wahrnehmbar sind als Einbrüche, negative Abweichungen vom Zielbandpass also tolerierbarer sind als positive. Unter diesem Gesichtspunkt wurden für den Vorversuch zur Bestimmung von β die vier in Tabelle 4.1 aufgelisteten Werte ausgewählt. Auf dieselbe Weise wie oben sind nachfolgend (Abb. 4.5, Abb. 4.6, Tabelle 4.2) die Ergebnisse einer FFD-Entzerrung mit einer oktavgeglätteten Inversen als Regularisierungsfilter (vgl. Abschnitt 3.2) dargestellt. Es erscheint hier als Herausforderung, eine Gewichtung zu finden, die zu ausreichender Regularisierung bei hohen Frequenzen führt, eine exakte Entzerrung im mittleren Frequenzbereich jedoch nicht zu sehr beeinträchtigt. Da die Kopfhörerübertragungsfunktionen zwischen 100 Hz und 1 kHz eine leichte Einbuchtung aufweisen, erfahren sie dort wegen des invertierten Regularisierungsfilters eine weniger genaue Kompensation, was sich bei hohen β -Werten negativ auswirkt. Vor allem beim linken Kanal des Stax Pro ist der starken Varianz der hochfrequenten Einbrüche mit dem hier angewendeten $B(\omega)$ offenbar schwer beizukommen. Dies wird in Abb. 4.6 durch die hohen Fehlermaxima oberhalb von 10 kHz deutlich, die trotz zunehmender Größe von β beinahe unverändert bestehen bleiben.



Abbildung 4.5 – FFD-Kompensation mit oktavgeglättetem, invertiertem Regularisierungsfilter und verschiedenen β für linken (oben) und rechten (unten) Kanal. Die grauen Linien zeigen jeweils zehn entzerrte Übertragungsfunktionen, die schwarze Linie ist der Zielbandpass.

Die geschilderten Beobachtungen weisen darauf hin, dass eine eher "globale" Regularisierung, wie z.B. durch ein Hochpassfilter, besser geeignet ist, als eine individuell an die Übertragungsfunktion angepasste. Die für den Vorversuch mit den gleichen Überlegungen wie oben gewählten Gewichtungsfaktoren und ihre Fehlerwerte sind in Tabelle 4.2 aufgelistet.



Abbildung 4.6 – Fehler in dB im Frequenzbereich bei FFD-Kompensation mit oktavgeglättetem, invertiertem Regularisierungsfilter und verschiedenen β für linken (oben) und rechten (unten) Kanal. Für alle zehn Entzerrungen ist die Abweichung vom Zielbandpass in den 40 Frequenzbändern der auditorischen Filterbank zu sehen.

		Stax Pro		Sta	ax II
		links	rechts	links	rechts
$\beta = 0,02$	$\max[E(f_c)]$	4,12	2,39	2,99	2,00
	$\min[E(f_c)]$	-7,67	-3,76	-3,09	-3,40
$\beta = 0,05$	$\max[E(f_c)]$	3,87	2,18	2,28	1,79
	$\min[E(f_c)]$	-7,94	-4,33	-3,27	-4,61
$\beta = 0,07$	$\max[E(f_c)]$	3,73	2,09	2,13	1,77
	$\min[E(f_c)]$	-8,10	-4,81	-3,48	-5,09
$\beta = 0, 1$	$\max[E(f_c)]$	3,52	1,96	1,96	1,74
	$\min[E(f_c)]$	-8,33	-5,27	-3,85	-5,61

Tabelle 4.2 – Fehlerwerte in dB bei FFD-Entzerrung mit oktavgeglättetem, invertiertem Regularisierungsfilter und verschiedenen β . Es sind die größten positiven sowie negativen Abweichungen vom Zielbandpass aus zehn Kompensationen aufgelistet.

Bei den Glättungsmethoden nach Hatziantoniou und Mourjopoulos (2000) wurde als Fensterfunktion ein von-Hann-Fenster (b = 0, 5) verwendet, was eine mittlere Glättung bewirkt (vgl. Hatziantoniou u. Mourjopoulos 2000, S. 274 f.). Die Autoren konstatieren, dass die Veränderung der Glättungsbandbreite eine erheblich größere Auswirkung auf das Kompensationsergebnis hat, als verschiedene Fensterfunktionen. Da eine Einschränkung der experimentellen Parameterbestimmung im Rahmen dieser Arbeit notwendig war, wurden also lediglich unterschiedliche Bandbreiten verglichen. In Abb. 4.7 und Abb. 4.8 sind die Ergebnisse für oktav- und ERB-Glättungen zu sehen. Es ist offensichtlich, dass eine relativ große Bandbreite nötig ist, um starke peaks bei hohen Frequenzen ausreichend zu vermeiden. Dies trifft wiederum vor allem auf den Stax Pro zu, wodurch sich die bereits oben gemachte Feststellung bestätigt, dass die ausgeprägte Verschiebung der Frequenzgänge dieses Kopfhörers im hochfrequenten Bereich für die Kompensation problematisch ist. Es können nur mit einer oktavbreiten Glättung Ergebnisse erzielt werden, die mit denjenigen der bisher beschriebenen Methoden vergleichbar sind. Aus diesem Grund wurde diese Bandbreite für die Smoothing-Verfahren festgelegt. Dies gilt sowohl für die Amplituden- als auch für die äquivalente komplexe Glättung, die sich im Betragsfrequenzgang per Definition nicht unterscheiden.

Für die **Compare and Squeeze-Methode** wurde auf die Angaben von Müller (1999, S. 194) zurückgegriffen, der einen Stauchungsfaktor von s = 0,3 zur Angleichung der geglätteten notches an die originale Übertragungsfunktion empfiehlt. Die Glättung wurde allerdings nicht wie von ihm vorgeschlagen terz- sondern oktavbreit durchgeführt, um dies konsistent mit den anderen Glättungsmethoden zu halten. Die Ergebnisse hierfür sind in Abb. 4.9 und Abb. 4.10 dargestellt. Zwar weisen die kompensierten Frequenzgänge relativ starke Überhöhungen im hohen Frequenzbereich auf. Ein Vergleich mit einem Faktor von 0,1 zeigte dahingehend jedoch kaum Veränderungen (s. Tabelle 4.3), sodass der Stauchungsparameter bei dem von Müller

angegebenen Wert belassen wurde. Je kleiner *s* ist, desto mehr gleichen die dips einer Übertragungsfunktion ihrer geglätteten Version. Bei kleinerem Faktor müssten also die peaks in der hier simulierten Entzerrung abnehmen, da die Einbrüche der Ausgangsfunktion ungenauer kompensiert werden und sich somit Verschiebungen der Kopfhörerposition weniger negativ auswirken. Dass die maximalen Abweichungen vom Zielbandpass in Tabelle 4.3 selbst bei geringem Stauchungsfaktor bei beiden Kopfhörern um 2-3 dB liegen, weist auf eine geringe Robustheit dieser Kompensationsmethode gegenüber Verschiebungen hin, wie dies auch bei der Regularisierung mit einem geglätteten, invertierten Filter der Fall war.



Abbildung 4.7 – Smoothing-Kompensation mit verschiedenen Glättungsbandbreiten für linken (oben) und rechten (unten) Kanal. Die grauen Linien zeigen jeweils zehn entzerrte Übertragungsfunktionen, die schwarze Linie ist der Zielbandpass. Zu beachten ist die abweichende Ordinatenskalierung in (a)/(b) gegenüber (c)/(d)



Abbildung 4.8 – Fehler in dB im Frequenzbereich bei Smoothing-Kompensation mit verschiedenen Glättungsbandbreiten für linken (oben) und rechten (unten) Kanal. Für alle zehn Entzerrungen ist die Abweichung vom Zielbandpass in den 40 Frequenzbändern der auditorischen Filterbank zu sehen. Zu beachten ist die abweichende Ordinatenskalierung in (a)/(b) gegenüber (c)/(d)



Abbildung 4.9 – Compare and Squeeze-Kompensation mit einem Stauchungsfaktor von 0,3 für linken (oben) und rechten (unten) Kanal. Die grauen Linien zeigen jeweils zehn entzerrte Übertragungsfunktionen, die schwarze Linie ist der Zielbandpass.



Abbildung 4.10 – Fehler im Frequenzbereich bei Compare and Squeeze-Kompensation mit einem Stauchungsfaktor von 0,3 für linken (oben) und rechten (unten) Kanal. Für alle zehn Entzerrungen ist die Abweichung vom Zielbandpass in den 40 Frequenzbändern der auditorischen Filterbank zu sehen.

		Stax Pro		Stax II	
		links	rechts	links	rechts
<i>s</i> = 0, 1	$\max[E(f_c)]$	3,67	2,16	1,98	2,15
	$\min[E(f_c)]$	-7,35	-4,29	-3,79	-2,63
<i>s</i> = 0, 3	$\max[E(f_c)]$	3,79	2,85	2,10	2,23
	$\min[E(f_{c})]$	-7,39	-4,10	-3,63	-2,31

Tabelle 4.3 – Fehlerwerte in dB bei Compare and Squeeze-Kompensation mit zwei verschiedenen Stauchungsfaktoren. Es sind die größten positiven sowie negativen Abweichungen vom Zielbandpass aus zehn Kompensationen aufgelistet.

Die All-Pol-Methode erscheint von den beschriebenen Verfahren am problematischsten. Je höher die Ordnung des All-Pol-Modells gewählt wird, desto genauer wird die Übertragungsfunktion des Kopfhörers angenähert. Dies hat eine exaktere Kompensation zur Folge, was im hohen Frequenzbereich wegen der dortigen notches eher unerwünscht ist. Eine geringe Modellordnung wirkt sich aber auf den gesamten Frequenzbereich aus und hat daher im vorliegenden Fall ein ungenaue Repräsentation der Stax-typischen Anhebung im Bass zur Folge. Wie in Abb. 4.11(a)-(b) zu sehen ist, führt dies zu einer unzureichenden Kompensation der tiefen Frequenzen, was beim Stax II durch seine stärkere Bassüberhöhung ausgeprägter ist. Wird die Modellordnung erhöht (Abb. 4.11(c)-(d)), verschiebt sich die Problematik in den hohen Frequenzbereich, da die Kompensation dort wie erläutert dann zu exakt erfolgt. Bei Kopfhörern mit glattem Frequenzgang im Bass wäre eine All-Pol-Kompensation sicherlich einfacher. Es ist allerdings aus Abb. 4.11(a) ersichtlich, dass sogar bei kleinem p die Anhebungen nicht nur bei tiefen sondern auch bei hohen Frequenzen unzureichend vermieden werden. Um ein Bild von weniger extremen Werten als den in Abb. 4.11 und Abb. 4.12 dargestellten zu bekommen, sind in Tabelle 4.4 die Fehlerwerte für drei mittlere Ordnungen zu sehen. Da nicht alle in Abschnitt 3.2 beschriebenen Verfahren im geplanten Hörversuch evaluiert werden konnten, wurde die All-Pol-Methode auf Basis der hier dargestellten Ergebnisse als ungeeignet für die Kompensation der verwendeten Kopfhörer eingestuft. Die Betrachtungen zu diesem Verfahren werden an dieser Stelle beendet.

		Stax Pro		Stax II	
		links	rechts	links	rechts
p=128	$\max[E(f_c)]$	5,71	7,05	4,69	5,66
	$\min[E(f_c)]$	-5,65	-2,92	-2,81	-1,35
p=256	$\max[E(f_c)]$	5,67	10,04	3,77	4,14
	$\min[E(f_c)]$	-6,44	-2,92	-2,78	-1,45
p=512	$\max[E(f_c)]$	5,32	12,15	3,47	2,93
	$\min[E(f_c)]$	-6,87	-3,20	-3,08	-1,35

Tabelle 4.4 – Fehlerwerte in dB bei All-Pol-Kompensation für verschiedene Ordnungen p. Es sind die größten positiven sowie negativen Abweichungen vom Zielbandpass aus zehn Kompensationen aufgelistet.



Abbildung 4.11 – All-Pol-Kompensation mit verschiedenen Ordnungen *p* für linken (oben) und rechten (unten) Kanal. Die grauen Linien zeigen jeweils zehn entzerrte Übertragungsfunktionen, die schwarze Linie ist der Zielbandpass. Zu beachten ist die abweichende Ordinatenskalierung in (a)/(b) gegenüber (c)/(d)



Abbildung 4.12 – Fehler in dB im Frequenzbereich bei All-Pol-Kompensation mit verschiedenen Ordnungen *p* für linken (oben) und rechten (unten) Kanal. Für alle zehn Entzerrungen ist die Abweichung vom Zielbandpass in den 40 Frequenzbändern der auditorischen Filterbank zu sehen. Zu beachten ist die abweichende Ordinatenskalierung in (a)/(b) gegenüber (c)/(d)

4.1.3 Kompensationsgüte

Zur Bestimmung der Kompensationsgüte der Verfahren wurden die aus zehn Messungen gemittelten Kopfhörerimpulsantworten mit den jeweiligen Filterimpulsantworten gefaltet. Die Anwendung der Entzerrungsfilter auf eben jene Übertragungsfunktion, die als Basis der Koeffizientenberechnung diente, ermöglicht eine theoretischere Aussage über die Güte der einzelnen Methoden als die oben durchgeführte Faltung mit zehn Einzelmessungen. Für die entsprechenden Faltungsergebnisse wurden die beschriebenen Fehlermaße im Zeit- und Frequenzbereich bestimmt. Tabelle 4.5 zeigt eine Liste der Verfahren, ihrer im Folgenden verwendeten Abkürzungen sowie der Parameterwerte, die bei Faltung und Fehlerberechnung eingesetzt wurden.³ Die Filterlänge betrug jeweils 2048 samples.

Verfahren	Abkürzung	Parameter
FFD, Regularisierung: Hochpass	ffd_hp	$\beta = 0, 4$
FFD, Regularisierung: Geglättete Inverse	ffd_oct	$\beta = 0, 1$
LS, Regularisierung: Hochpass	ls_hp	$\beta = 0, 4$
LS, Regularisierung: Geglättete Inverse	ls_oct	$\beta = 0, 1$
Amplitude Smoothing	sm_amp	Oktavglättung, $b = 0, 5$
Equivalent Complex Smoothing	sm_eqcmp	Oktavglättung, $b = 0, 5$
Compare and Squeeze	c&s	Oktavglättung, $s = 0, 3$

Tabelle 4.5 – Abkürzungen der Verfahren und in diesem Abschnitt verwendete Parameterwerte

In Tabelle 4.6 ist der Fehler im Zeitbereich, in Tabelle 4.7 der Gesamtfehler im Frequenzbereich dargestellt, wobei die Verfahren nach ihrer daraus resultierenden Güte geordnet wurden. In Abb. 4.13 ist der frequenzabhängige Fehler für einige Verfahren sichtbar. Die Ergebnisse der FFD-Methode und der Amplitudenglättung erscheinen darin nicht explizit, da sie praktisch identisch mit der LS-Methode bzw. der äquivalenten komplexen Glättung sind, wie aus Tabelle 4.7 ersichtlich wird.

Ein Vergleich des Zeit- und Frequenzfehlers macht deutlich, dass diese beiden Maße zusammenhängen. Verfahren, die im Frequenzbereich eine gute Annäherung an die Zielfunktion leisten, zeigen entsprechende Ergebnisse im Zeitbereich, da die entzerrte Funktion dort ebenfalls am ehesten der Impulsantwort eines Bandpasses gleicht. Dieser Zusammenhang ist zwar nicht allzu erstaunlich, müsste jedoch nicht zwingend bestehen. Hierzu sei die äquivalente komplexe Glättung als Gegenbeispiel angeführt. Sie liefert im Zeitbereich deutlich schlechtere Ergebnisse als im Frequenzbereich, was an der zeitselektiven Glättung liegt, welche dieses Verfahren kennzeichnet. Die nicht ganz exakte Kompensation findet hier nicht nur bezüglich des Amplitudengangs sondern auch bezüglich des Zeitsignals statt.

³Im Falle der LS- und der FFD-Methode wurde der jeweils höchste der für den Vorversuch ausgewählten Werte verwendet.



Abbildung 4.13 – Frequenzabhängiger Fehler ausgewählter Verfahren für beide Kanäle (links: schwarz, rechts: grau) des Stax Pro (oben) und Stax II (unten)

Daher erhöht sich der Fehler in der entzerrten Impulsantwort, im Frequenzbereich entspricht er hingegen demjenigen der Amplitudenglättung.

Beim Vergleich der Methoden untereinander ist vor allem der große Sprung zwischen den Regularisierungs- und den restlichen Verfahren in Tabelle 4.7 auffällig. Weiterhin überrascht die große Übereinstimmung der LS- und FFD-Entzerrung bei beiden Fehlermaßen, da sie sich nicht mit den Ergebnissen von Norcross et al. (2002) deckt. Dort lieferte die FFD-Methode sowohl in perzeptiven Tests als auch objektiv berechnet schlechtere Ergebnisse, die sich vor allem bei tiefen Frequenzen manifestierten. Allerdings wurden in der genannten Untersuchung Lautsprecherübertragungsfunktionen entzerrt, die deutlich größere Schwankungen im Bass aufwiesen als die Kopfhörerfrequenzgänge in der vorliegenden Arbeit. Es könnte also an der linearen Auflösung der Fourier Transformation liegen, dass diese Methode bei tiefen Frequenzen eine unzureichend genaue Entzerrung erreichte. Es wird angenommen, dass dieser Aspekt im vorliegenden Fall durch die Charakteristik der zu entzerrenden Frequenzgänge nicht ins Gewicht fiel.

	Stax Pro		Stax II		
Verfahren	links	rechts	links	rechts	
c&s	$0.0556 \cdot 10^{-5}$	$0.1884 \cdot 10^{-5}$	$0.0996 \cdot 10^{-5}$	$0.1562 \cdot 10^{-5}$	
sm_amp	$0.1789 \cdot 10^{-5}$	$0.4199 \cdot 10^{-5}$	$0.3258 \cdot 10^{-5}$	$0.4226 \cdot 10^{-5}$	
ffd_oct	$0.1746 \cdot 10^{-5}$	$0.6117 \cdot 10^{-5}$	$0.0154 \cdot 10^{-4}$	$0.1050 \cdot 10^{-4}$	
ls_oct	$0.1746 \cdot 10^{-5}$	$0.6117 \cdot 10^{-5}$	$0.0154 \cdot 10^{-4}$	$0.1050 \cdot 10^{-4}$	
ffd_hp	$0.0686 \cdot 10^{-4}$	$0.1211 \cdot 10^{-4}$	$0.0784 \cdot 10^{-4}$	$0.1596 \cdot 10^{-4}$	
ls_hp	$0.0686 \cdot 10^{-4}$	$0.1211 \cdot 10^{-4}$	$0.0784 \cdot 10^{-4}$	$0.1596 \cdot 10^{-4}$	
sm_eqcmp	$0.2145 \cdot 10^{-4}$	$0.2475 \cdot 10^{-4}$	$0.2630 \cdot 10^{-4}$	$0.2726 \cdot 10^{-4}$	

Tabelle 4.6 – Fehler e im Zeitbereich für alle Kompensationsverfahren

	Stax	Pro	Stax II		
Verfahren	links	rechts	links	rechts	
sm_amp	0.1130	0.1066	0.1102	0.1422	
sm_eqcmp	0.1135	0.1067	0.1090	0.1413	
c&s	0.1347	0.1984	0.1661	0.1783	
ffd_hp	0.6422	0.7593	0.6087	0.8713	
ls_hp	0.6421	0.7595	0.6089	0.8718	
ffd_oct	0.7565	0.8403	0.4010	0.6889	
ls_oct	0.7575	0.8422	0.4014	0.6895	

Tabelle 4.7 – Fehler *E*tot in dB im Frequenzbereich für alle Kompensationsverfahren

Eine Gegenüberstellung der beiden Kopfhörer lässt keine eindeutigen Schlüsse zu. Allerdings scheint der Fehler im Zeitbereich beim Stax Pro tendenziell kleiner auszufallen als beim Stax II. Besonders deutlich ist dies bei Regularisierung mit oktavgeglätteter Inverse. Im Frequenzbereich ist der Unterschied der Fehlerwerte beider Kopfhörer nicht systematisch ausgeprägt.

Nach den dargestellten Fehlermaßen bringen eben jene Methoden und jener Kopfhörer die geringste Güte hervor, welche im Zusammenhang der Filterparameterwahl trotz veränderter Kopfhörerausrichtungen eine relativ gute Kompensation zeigten. Eine exakte theoretische Kompensation bedeutet also keinesfalls eine ausreichende Unempfindlichkeit gegenüber Positionsverschiebungen. Es kann vermutet werden, dass niedrige Werte von E_{tot} mit mehr unerwünschten peaks bei "realer" Entzerrung, also bei beliebig aufgesetztem Kopfhörer, in Zusammenhang gebracht werden müssen.

In Abb. 4.14-4.20 sind die Betragsfrequenzgänge der Entzerrungsfilter selbst sowie die Faltungsergebnisse im Zeit- und Frequenzbereich grafisch dargestellt. Anstatt des in Ab-

schnitt 2.2 gezeigten Phasengangs ist jeweils die Gruppenlaufzeit aufgetragen, die wegen des modelling delays eine Grundverzögerung der halben Filterlänge, also

$$\Delta t = \frac{1024 \,\text{samples}}{44100 \,\text{samples/s}} = 23,2 \,\text{ms}$$
(4.10)

erkennen lässt. Es ist zu sehen, dass die Entzerrungsfilter unter dem Gesichtspunkt der von Blauert und Laws (1979) angegebenen Schwelle von 0,5 ms eine weitgehend konstante Gruppenlaufzeit – und somit eine Linearisierung der Phase – bewirken. Lediglich unterhalb von ca. 100 Hz ist dies nicht mehr gewährleistet, wobei hier die untere Eckfrequenz des Zielbandpasses bereits beinahe erreicht ist, und die beschränkte Filterlänge zusätzlich eine Rolle spielt. Das in Abschnitt 2.2 angesprochene Problem der interauralen Phasendifferenz zeigt nach der Kopfhörerentzerrung ebenfalls eine deutliche Verbesserung. Die hier angesetzte Schwelle von 4° wird bis zur unteren Eckfrequenz des Bandpasses in den dargestellten Ergebnissen praktisch nicht überschritten. Lediglich der "Ausreißer" des Stax Pro bei ca. 8700 Hz, der auch im Betragsfrequenzgang und der Gruppenlaufzeit des linken Kanals zu sehen ist, lässt sich bei dieser Filterlänge offenbar nicht von allen Methoden ausreichend kompensieren. Ob dies einen hörbaren Einfluss hat, ist jedoch zu bezweifeln.

Die Kompensation mit Hilfe äquivalenter komplexer Glättung führt im Vergleich zu den restlichen Verfahren zu einem geringeren Ausgleich der interauralen Phasendifferenz und der Gruppenlaufzeitverzerrungen. Wie im Zusammenhang der Fehlerwerte in Tabelle 4.6 erläutert wurde, findet bei dieser Methode wegen der komplexen Glättung eine unpräzise Linearisierung der gesamten Übertragungsfunktion, also auch der Phase und der Impulsantwort, statt. Die Höhe der in Tabelle 4.6 dargestellten Fehlerleistung dieses Verfahrens hängt offenbar vor allem mit dem Ausschwingverhalten der entzerrten Impulsantwort zusammen. Das Einschwingen ist hingegen von geringerem Rauschpegel bestimmt als bei den anderen Methoden. Da die zeitliche Nachverdeckung stärker ausgeprägt ist als die Vorverdeckung, könnte die perzeptive Beurteilung dieses Verfahrens eventuell besser ausfallen, als die Berechnung auf den ersten Blick vermuten lässt. Diese Frage wird sich im folgenden Abschnitt klären.



Abbildung 4.14 – Kompensationsergebnisse FFD mit Hochpass-Regularisierung: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.15 – Kompensationsergebnisse LS mit Hochpass-Regularisierung: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.16 – Kompensationsergebnisse FFD mit geglättetem, invertiertem Regularisierungsfilter: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.17 – Kompensationsergebnisse LS mit geglättetem, invertiertem Regularisierungsfilter: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.18 – Kompensationsergebnisse Amplitudenglättung: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.19 – Kompensationsergebnisse äquivalente komplexe Glättung: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II



Abbildung 4.20 – Kompensationsergebnisse Compare and Squeeze: (a) Betragsfrequenzgänge der Kopfhörer und der Entzerrungsfilter, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (b) ETC der kompensierten Kopfhörer, jeweils linker Kanal (Stax Pro: oben, Stax II: unten); (c) Gruppenlaufzeit der kompensierten Kopfhörer (Stax Pro: oben, Stax II: unten); (d) Phasendifferenz der kompensierten Kopfhörerr (Stax Pro: oben, Stax II: unten); (e) Betragsfrequenzgang des kompensierten Stax Pro; (f) Betragsfrequenzgang des kompensierten Stax II

4.1.4 Rechenaufwand

Zwar geschieht die Kompensation der elektroakustischen Übertrager nicht in Echtzeit sondern im post-processing der BRIR-Messung (s. Abschnitt 1.2), sodass es dabei nicht auf eine unhörbare Latenz ankommt. Trotzdem spielt auch der Rechenaufwand eine gewisse Rolle bei der Beurteilung der Verfahren, da exzessive Berechnungszeiten grundsätzlich unpraktikabel sind. Die nachfolgende Tabelle gibt Aufschluss über diese Frage, wobei die dargestellten Werte mit einem Intel Core Duo 2,16 GHz MacBook mit 2 GB RAM ermittelt wurden.

	Filterlänge			
Verfahren	1024	2048	4096	8192
FFD	50,69 ms	50,74 ms	56,76 ms	71,61 ms
LS	9,797 s	40,40 s	3,530 min	(17,65 min)
Amplitude Smoothing	16,51 s	2,092 min	16,57 min	(2,18 h)
Equivalent Complex Smoothing	49,78 s	6,354 min	49,34 min	(6,50 h)

Tabelle 4.8 – Rechenaufwand der Verfahren bei verschiedenen Filterlängen

Die Compare and Squeeze-Methode konnte nicht auf diese Weise evaluiert werden, da der Ablauf zur Filterberechnung von demjenigen der restlichen Verfahren abweicht, wie in Abschnitt 3.2 erläutert wurde. Die Angaben für eine Filterlänge von 8192 samples beim LSund den Smoothing-Verfahren sind lediglich geschätzt, was daher rührt, dass hier der Speicher des verwendeten Computers für die Erzeugung der Faltungsmatrizen (s. Abschnitt 3.2) nicht ausreichte. Ein solches Problem lässt sich natürlich durch einen größeren Arbeitsspeicher oder auch durch eine blockweise Berechnung der Filter beheben, ist aber trotzdem als negativer Aspekt dieser Verfahren zu erwähnen. Beachtenswert ist, dass sich die Rechenzeiten bei den Smoothing-Verfahren mit verdoppelter Filterlänge jeweils beinahe verachtfachen, während dieser Faktor bei der LS-Methode bei ungefähr 5 liegt. Insgesamt hebt sich die bereits im Namen des Verfahrens enthaltene Geschwindigkeit der FFD-Methode deutlich von den anderen ab.

4.2 **Perzeptive Evaluation**

Die perzeptive Evaluation der implementierten Entzerrungsalgorithmen stellt einen zentralen Punkt dieser Arbeit dar. Zwar ließen sich bereits im vorangegangenen Abschnitt auf Basis einer objektiven Bewertung der Verfahren einige Vermutungen über ihre Effizienz formulieren. Im eingangs beschriebenen Kontext dieser Arbeit geht es jedoch um die Frage, ob die binaurale Simulation einer akustischen Umgebung perzeptiv vom realen Schallfeld unterscheidbar ist, wenn die elektroakustische Übertragungsstrecke des Simulationssystems ausreichend kompensiert wird. Um die Auswirkung der Entzerrungsverfahren auf die Wahrnehmung zu untersuchen und zu ermitteln, welcher Algorithmus sich perzeptiv am besten zur originalgetreuen Wiedergabe der Ohrsignale eignet, wurde ein Hörversuch zum direkten Vergleich von Simulation und Realität durchgeführt. Nachfolgend wird diese Evaluation zusammen mit ihren Ergebnissen vorgestellt.

4.2.1 Versuchsbeschreibung

4.2.1.1 Versuchsmethode und -design

Als Methode zur Beantwortung der oben genannten Fragestellung wurde ein ABC/HR-Test (*triple stimulus hidden reference*) gewählt. Dabei handelt es sich um ein Forced-Choice-Verfahren, bei dem einer Versuchsperson mindestens zwei Reize dargeboten werden, wobei einer entsprechend der Fragestellung manipuliert wurde, während der oder die anderen eine unveränderte Referenz darstellen. Von der Versuchsperson wird die Kennzeichnung eines der Stimuli als veränderte Version verlangt, wofür es eine objektiv richtige bzw. falsche Antwort gibt.



Abbildung 4.21 – Grafische Benutzeroberfläche für den ABC/HR-Hörversuch

Beim ABC/HR-Verfahren werden dem Probanden drei Reize dargeboten, von denen einer eindeutig als Referenz gekennzeichnet ist (vgl. "Ref"-Button in Abb. 4.21) und mit den zwei unmarkierten Stimuli verglichen wird. Von diesen entspricht einer der Referenz, der andere hingegen stellt einen manipulierten Vergleichsreiz dar. Die Versuchsperson soll entscheiden, welcher der nicht gekennzeichneten Stimuli mit dem Original identisch ist. Die Zuordnung wird vorgenommen, indem der als verändert empfundene Reiz auf einer Intervallskala quantitativ bewertet wird (*rating*), das Forced-Choice-Verfahren wird also mit einem absoluten Urteil verbunden. Für die Versuchsperson sind mehrere Bedingungsvariationen gleichzeitig verfügbar (acht Gruppen in Abb. 4.21), sodass sie beliebig oft zwischen allen Reizen vergleichen kann. Dies ergibt den Vorteil, dass die Werte des absoluten Urteils relativ gesehen stimmig sind. Da für beide ungekennzeichneten Stimuli eine Bewertungsskala zur Verfügung steht, kann statt des veränderten Stimulus' auch die Referenz von der Versuchsperson als verändert eingestuft werden. Dies entspricht einer objektiv falschen Antwort und lässt den Schluss zu, dass Original und Vergleichsreiz nicht unterschieden werden konnten. Zur Auswertung von ABC/HR-Tests werden Differenzgrade (*diffgrades*) gebildet, indem der Unterschied zwischen der Bewertung des manipulierten Stimulus' und der Referenz berechnet wird. Positive Differenzgrade bedeuten demnach, dass der veränderte Reiz nicht als solcher erkannt wurde.

Die beschriebene Versuchsmethode wurde in dieser Arbeit ausgewählt, um einerseits durch eine direkte Gegenüberstellung zu ermitteln, ob einige der Kompensationsalgorithmen die perzeptive Gleichheit von Simulation und Realität tatsächlich gewährleisten können. Auf Grund des Ratings sollte andererseits eine perzeptive Klassifizierung dieser Methoden vorgenommen werden. Der als Referenz definierte Stimulus wurde über einen Lautsprecher – die reale Schallquelle – dargeboten. Der manipulierte Vergleichsreiz war eine über Kopfhörer wiedergegebene Simulation dieses Lautsprechers und seiner akustischen Umgebung,⁴ wobei die Übertragungsstrecke des Simulationssystems abwechselnd mit den in Tabelle 4.5 aufgelisteten Entzerrungsverfahren kompensiert wurde.

Empfohlene Richtlinien zum ABC/HR-Verfahren sind in Rec. ITU-R BS.1116-1 "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems" beschrieben. Darin wird die Bezeichnung der Ratingskala von "imperceptible", entsprechend eines Skalenwerts von 5, bis "very annoying" mit Skalenwert 1, empfohlen. Diese Benennungen waren im Zusammenhang des durchgeführten Versuchs jedoch in zweierlei Hinsicht problematisch. Erstens wird die Empfehlung grundsätzlich kritisch gesehen, da die zwei genannten Begriffe aus unterschiedlichen semantischen Kategorien stammen und deshalb nicht ohne weiteres als Gegensatzpaar benutzt werden können. Zweitens sollte nicht die Qualitätsverminderung der Entzerrungsmethoden eingeschätzt werden, sondern explizit die Ähnlichkeit der kompensierten Simulation mit einer realen Schallquelle. Die genannten Bezeichnungen erschienen hierfür unpassend. In Rec. ITU-R BS.1248-1 werden weitere Skalenbeschriftungen vorgeschlagen. So wird "excellent" bis "bad" für die Bewer-

⁴Dazu kam das am Fachgebiet entwickelte System zur dynamischen Binauralsynthese zum Einsatz (Lindau et al. 2007).

tung von Qualität und "much better" bis "much worse" für Vergleichstests genannt. Da diese für die geplante Anwendung ebenfalls nicht geeignet waren, wurde die in Abb. 4.21 sichtbare Bezeichnung von "identisch" bis "sehr unterschiedlich" - mit numerischen Werten zwischen 5 und 1 - festgelegt. Zu beachten ist, dass die Zwischenstufen der Skala entgegen Rec. ITU-R BS.1116-1 verbal unbezeichnet belassen wurden, da ansonsten ihre Äquidistanz hätte in Frage gestellt werden müssen. Es wurden lediglich in der Mitte und an den Endpunkten der Skala Markierungen angebracht, um eine Orientierung zu vereinfachen. Da die Versuchspersonen verschiedene subjektive Auffassungen des Begriffs "sehr unterschiedlich" haben mögen, ist das untere Extrem der Skala nicht eindeutig definiert. Dieses Problem ist jedoch kaum zu beheben und tritt auch bei den in Rec. ITU-R BS.1248-1 genannten Bezeichnungen auf. Um der Schwierigkeit entgegenzuwirken, wurde ein versteckter Ankerreiz in den Hörversuch integriert, der einer unentzerrten Version der simulierten Schallquelle entsprach. Wie oben erwähnt, erfolgen die Ratings mehrerer Bedingungsvariationen im ABC/HR-Test relativ zueinander, sodass davon ausgegangen wurde, dass die Bewertung des Ankers jeweils den tiefsten Punkt einer subjektiven Skala markieren würde. Damit war zumindest für die einzelnen Versuchspersonen der untere Extrempunkt festgelegt.

Um die perzeptive Eignung der Entzerrungsmethoden in Abhängigkeit des dargebotenen Audioinhalts zu untersuchen, war die Verwendung zweier Testreize mit unterschiedlichen Eigenschaften beabsichtigt. Als besonders empfindlicher Stimulus wurde rosa Rauschen ausgewählt, ferner war ein natürlicher Content mit transienten Anteilen wünschenswert. Letzteres war der Absicht geschuldet, dass auch die Wahrnehmung der Entzerrung im Zeitbereich erfasst werden sollte. Durch die geforderte Natürlichkeit als Gegensatz zum künstlichen Rauschen wurde eine zusätzliche Verallgemeinerung der Ergebnisse ermöglicht. In die engere Auswahl für den zweiten Stimulus kamen eine akustische Gitarre sowie ein Schlagzeug. Zum Vergleich sind in Abb. 4.22 die Betragsfrequenzgänge⁵ der drei erwähnten Contents nach Glättung mit einer Bandbreite von ¹/12 Oktave dargestellt. Um die Bandbeschränkung der Kopfhörer als offensichtliches Erkennungsmerkmal für den Unterschied zur Wiedergabe über Lautsprecher auszuschließen, fand eine Beschränkung der Reize auf die Bandbreite des Zielbandpasses statt. Zudem wurden sie zur Darbietung im Hörversuch auf eine Länge von 5 s gekürzt.

Der Gitarrenstimulus⁶ hatte sich in der erwähnten Arbeit von Lindau und Weinzierl (2007),

⁵Fourier-Transformation der gesamten Zeitsignale

⁶Er wurde dem Sound Quality Assessment Material der European Broadcasting Union (EBU SQAM) entnommen.



Abbildung 4.22 – Betragsfrequenzgänge der drei bandbeschränkten Stimuli: Schlagzeug (blau), Gitarre (grün), rosa Rauschen (schwarz)

in der die Plausibilität des binauralen Simulationssystems am Fachgebiet untersucht wurde, unter den natürlichen Testreizen als besonders geeignet zur Detektion von Artefakten erwiesen. Daher wurde er auch für die hier beschriebene perzeptive Evaluation der Entzerrungsalgorithmen verwendet. Er vereint in sich einerseits harmonische Komponenten sowie andererseits transiente Anteile, die durch Zupfgeräusche zustande kommen.

Der Schlagzeugausschnitt wurde als weniger günstig betrachtet, weil darin die spektralen Anteile der vorhandenen Bassdrum eine prominente Rolle spielen (s. Abb. 4.22). Durch tiefe Frequenzen hervorgerufene Vibrationen, die bei Lautsprecherwiedergabe zustande kommen, können jedoch über Kopfhörer nicht angemessen simuliert werden. Diese Tatsache machte sich beim Probehören der drei Stimuli trotz Bandbeschränkung beim Schlagzeugsample am deutlichsten bemerkbar.⁷

Wie in Abschnitt 4.1 erläutert wurde, kamen im Hörversuch zwei Kopfhörermodelle, Stax SRS 2050 II und Stax Lambda Pro New, zum Einsatz. Dadurch ergaben sich also zusammen mit den zwei Testreizen, den sieben Kompensationsverfahren und dem Anker $2 \cdot 2 \cdot 8 = 32$ Bedingungsvariationen. Die abhängigen Variablen (Faktoren) waren dabei die Entzerrungsmethode, der Content und der verwendete Kopfhörer. Die unabhängige Variable war die Bewertung, die auf der Ratingskala abgegeben wurde, wobei aus ihr auch entnommen werden konnte, ob eine korrekte Identifikation der simulierten Quelle erfolgt war. Da

⁷Da dieses Phänomen aber auch bei den anderen Stimuli spürbar war, könnte in künftigen Untersuchungen ein bei Kopfhörerwiedergabe zugeschalteter Subwoofer miteinbezogen werden. Dessen Wirksamkeit bzw. Plausibilität müsste jedoch vorab untersucht werden.

alle 32 Bedingungsvariationen von allen Versuchspersonen beurteilt wurden, handelt es sich um einen dreifaktoriellen Versuch mit Messwiederholungen und abhängigen Stichproben.

Ein weiterer Faktor, der mindestens zweistufig hätte variiert werden können, ist die Filterlänge. Eine Verdopplung der Bedingungsvariationen erschien aber aus Gründen der Konzentrationsanforderung und des Zeitaufwands unzumutbar für die Versuchspersonen. Um diese Einschränkung zu umgehen, hätte eine Variable – vermutlich der Content – als Gruppierungsfaktor eingesetzt werden können und wäre von zwei separaten Gruppen beurteilt worden. Wie weiter unten ausgeführt wird, wurde diese Option jedoch wegen der zu großen Anzahl an erforderlichen Probanden verworfen. Stattdessen wurde eine einzige Filterlänge von 2¹¹ samples festgelegt. Diese wurde als praktikabel beurteilt und ermöglicht laut den Ergebnissen aus Abschnitt 4.1 eine gute Annäherung an die Zielfunktion.

4.2.1.2 Stichprobenumfang und -beschreibung

Durch eine varianzanalytische Auswertung der Versuchsergebnisse sollten die Haupteffekte und Interaktionen der oben beschriebenen Faktoren überprüft werden. Um dabei bestimmte Effektgrößen auf einem Signifikanzniveau von $\alpha = 0,05$ abzusichern, wurde der optimale Stichprobenumfang für diese Untersuchung berechnet. Zu seiner Bestimmung bei einem mehrfaktoriellen Design mit Messwiederholungen geben Bortz und Döring (2002, S. 616) die folgende Formel an:

$$n_{\text{opt}} = \frac{(n-1) \cdot (df+1)}{p \cdot q \cdot r} + 1 \tag{4.11}$$

df bezeichnet dabei die Freiheitsgrade der fraglichen Interaktion, während p, q und r die Stufenanzahl der einzelnen Faktoren definieren. n ist hier die optimale Stichprobengröße für eine einfaktorielle Varianzanalyse mit Messwiederholungen und wurde zur Berechnung von n_{opt} in Abhängigkeit der Freiheitsgrade einer Tabelle aus Bortz und Döring (2002, S. 615) entnommen. Die Teststärke, also die Wahrscheinlichkeit, dass eine richtige Alternativhypothese H₁ angenommen wurde, ist dort stets auf den allgemein akzeptierten Wert von 0,8 festgelegt (vgl. Bortz und Döring 2002, S. 612).

Im vorliegenden Fall nahmen die Variablen *p*, *q* und *r* in Gl. 4.11 die folgenden Werte an:

- p = 7 Anzahl der Stufen des Faktors Entzerrungsmethode (ohne Anker)
- q = 2 Anzahl der Stufen des Faktors Kopfhörer
- r = 2 Anzahl der Stufen des Faktors Content

Da der Ankerreiz – die unkompensierte Simulation – keine eigentliche Bedingungsvariation darstellte, sondern auf jeder Stufe konstant blieb, wurde er bei der Varianzanalyse und damit auch bei der Berechnung des optimalen Stichprobenumfangs nicht miteinbezogen. Für die möglichen Interaktionen galten die unten aufgeführten Freiheitsgrade.

$$\begin{aligned} \text{df}(\text{Entzerrungsmethode x Kopfhörer x Content}) &= (p-1) \cdot (q-1) \cdot (r-1) \\ &= (7-1) \cdot (2-1) \cdot (2-1) = 6 \\ \\ \text{df}(\text{Entzerrungsmethode x Kopfhörer}) &= (p-1) \cdot (q-1) \\ &= (7-1) \cdot (2-1) = 6 \\ \\ \\ \text{df}(\text{Entzerrungsmethode x Content}) &= (p-1) \cdot (r-1) \\ &= (7-1) \cdot (2-1) = 6 \\ \\ \\ \text{df}(\text{Content x Kopfhörer}) &= (r-1) \cdot (q-1) \\ &= (2-1) \cdot (2-1) = 1 \end{aligned}$$

Aus Gl. 4.11 ist ersichtlich, dass der gesuchte Stichprobenumfang mit steigendem df wächst. Um auch bei den Interaktionen mit hohen Freiheitsgraden kleine Effekte mit $\alpha = 0,05$ absichern zu können, wurde n_{opt} daher mit df = 6 und – gemäß der oben erwähnten Tabelle für einfaktorielle Versuchspläne – mit n = 99 berechnet:

$$n_{\rm opt} = \frac{(99-1)\cdot(6+1)}{7\cdot 2\cdot 2} + 1 = 25, 5 \approx 26 \tag{4.12}$$

Die Haupteffekte beruhen damit pro untersuchter Stufe auf 26 Versuchspersonen, sodass sich durch die Messwiederholungen für sie die folgenden Stichprobenumfänge ergeben:

> Entzerrungsmethoden: $2 \cdot 2 \cdot 26 = 104$ Kopfhörer: $2 \cdot 7 \cdot 26 = 364$ Content: $2 \cdot 7 \cdot 26 = 364$

Diese reichen laut der Tabelle für einfaktorielle Pläne mit Messwiederholungen, abhängig von den Freiheitsgraden, in allen drei Fällen zur Absicherung von kleinen Effekten auf dem festgelegten Signifikanzniveau von 0,05 (Entzerrungsmethoden) bzw. sogar 0,01 (Kopfhörer und Content) aus.

Wäre die Filterlänge als zusätzlicher Faktor eingeführt und die Contents an zwei getrennten Personengruppen untersucht worden, wäre der Gruppierungsfaktor nicht in die Messwiederholungen eingeschlossen gewesen. Dadurch hätte sich der nötige Stichprobenumfang verdoppelt, was im Rahmen dieser Arbeit leider nicht zu erreichen war. Dies lag nicht zuletzt daran, dass für den durchgeführten Versuch nur geschulte Hörer in Frage kamen (siehe unten), und die Auswahl damit eingeschränkt war.

An der perzeptiven Evaluation der Kompensationsmethoden nahmen 28 Personen im Alter zwischen 24 und 43 Jahren (Durchschnitt: 30,21) teil. Davon waren 25 männlich und 3 weiblich. Zur Frage der Selektion geeigneter Probanden steht in Rec. ITU-R BS.1116-1:

"The outcome of subjective tests of sound systems with small impairments utilizing a selected group of listeners is not primarily intended for extrapolation to the general public. Normally the aim is to investigate whether a small group of expert listeners, under certain conditions, are able to perceive relatively subtle degradations..."(Rec. ITU-R BS.1116-1, S. 3)

Unter diesem Gesichtspunkt wurden als Versuchspersonen Mitarbeiter und Studenten des Fachgebiets ausgewählt, da sie als Expertenhörer bezeichnet werden können. Es wurde außerdem in einem Fragebogen (s. Anhang C) die musikalische Ausbildung und die Erfahrung mit Hörversuchen abgefragt. 21 der Teilnehmer stuften sich sowohl als musikalisch gebildet wie auch als vertraut mit Hörversuchen ein. Vier der Probanden hatten bereits früher an perzeptiven Versuchen teilgenommen ohne allerdings musikalische Kenntnisse zu besitzen. Schließlich waren drei Versuchspersonen musikalisch geschult, hatten aber noch keine Erfahrung mit Hörversuchen. Insgesamt kann also von qualifizierten Teilnehmern mit angemessener Hörerfahrung ausgegangen werden.

4.2.1.3 Versuchsaufbau und -ablauf

Der Hörversuch fand im kleinen elektronischen Studio des Fachgebiets Audiokommunikation der TU Berlin statt. Dieser Raum ist in Abb. 4.23(a) zu sehen und hat bei einem Volumen von 160 m³ eine Nachhallzeit von 0,7 s. Der als Referenz dienende Quelllautsprecher (Meyer Sound UPL-1) war in 2,9 m Entfernung und 2 m Höhe frontal vor der Versuchsperson platziert. Als Benutzerschnittstelle diente ein Computer mit der in Abb. 4.21 gezeigten grafischen Benutzeroberfläche (*graphical user interface*, GUI).

Der technische Versuchsaufbau ist in Abb. 4.23(b) schematisch dargestellt. Darin sind Signalwege, die nur bei Lautsprecherwiedergabe aktiv waren mit blauen Pfeilen, solche die nur für die Kopfhörerwiedergabe gelten mit grünen Pfeilen gekennzeichnet. Gemeinsame Signalwege sind schwarz dargestellt. Vom Versuchsrechner wurden je nach aktiviertem Button in der GUI bestimmte OSC (*open sound control*) Steuerbefehle über Netzwerk an




Abbildung 4.23 – (a) Versuchsraum: kleines Studio des Fachgebiets Audiokommunikation (b) Versuchsaufbau: blaue Pfeile kennzeichnen die Lautsprecher-, grüne Pfeile die Kopfhörerwiedergabe

den Rechner gesendet, der die Faltung ausführte. Sie sorgten zum einen für die korrekte Auswahl eines der beiden dort gespeicherten Stimuli. Im Falle der Lautsprecherwiedergabe wurde der entsprechende Reiz direkt über eine optische Verbindung zu einem D/A-Wandler (marian ADCON) und von dort über eine symmetrische Leitung zum Lautsprecher gesendet. Startete die Versuchsperson durch anwählen des Vergleichsreizes die binaurale Simulation, wurden mehrere Prozesse gestartet. Der ständig aktive head tracker, der auf dem Kopfhörer fixiert war und mit magnetischen Signalen die aktuelle Kopfposition an einen Empfänger übermittelte, lieferte Informationen zur korrekten Auswahl eines zur Kopfausrichtung passenden BRIR-Paares. Dieses wurde mit dem Testreiz gefaltet und als binaurales Signal einer zweiten Faltungsoperation zugeführt, welche die Entzerrung mit einem der Kompensationsfilter erzielte. An dieser Stelle bewirkten wiederum OSC-Befehle die Auswahl des von der Versuchsperson aktivierten und zum Kopfhörer passenden Filters aus einem Datensatz.⁸ Die so erzeugten Ohrsignale wurden ebenfalls über die optische Leitung an den D/A-Wandler gesendet und von dort über den Kopfhörerverstärker dem Kopfhörer zugeführt.

Die grafische Benutzeroberfläche sowie die Ablaufsteuerung des Versuchs wurden in Matlab® programmiert. Die Bedingungsvariationen wurden den Probanden dabei in Achtergruppen präsentiert (s. Abb. 4.21), wobei eine solche Gruppe aus einer zufälligen Reihenfolge der sieben Filter und des Ankers bestand. Der Vergleich der Entzerrungsmethoden untereinander war also sehr empfindlich, da sie relativ zueinander beurteilt wurden. Durch die – ebenfalls randomisierte – Darbietung der beiden Stimuli ergaben sich zwei Achtergruppen. Deren Rating wurde bei neu gemischter Ordnung von Filtern und Contents für den zweiten Kopfhörer wiederholt, auch die Reihenfolge dieses Faktors war zufällig. In der Mitte des Versuchs entstand durch den Kopfhörerwechsel eine Pause. Es ist wichtig festzustellen, dass die Teilnehmer den Kopfhörer jedoch während der Lautsprecherwiedergabe nicht absetzten, da ansonsten die Unterscheidung von simulierter und realer Quelle trivial gewesen wäre. Dieser Umstand wurde bei der Messung des BRIR-Datensatzes für die Simulation berücksichtigt, wie weiter unten näher ausgeführt wird.

Zu Beginn des Hörversuchs erhielten die Teilnehmer eine schriftliche Instruktion, die sie über den Versuchsablauf und ihre Aufgabe informierte (s. Anhang C). Darauf folgte ein Training, bei der die Filter anhand eines Stimulus' probeweise beurteilt wurden. Dies diente einerseits dem Vertrautwerden mit dem Prozess des relativen Ratings, andererseits konnten sich die Probanden dadurch an die Konzentration auf Merkmale zur Detektion von Unterschieden gewöhnen. Als Trainingsreiz wurde die Gitarre ausgewählt, weil dieser weniger empfindlich zur Detektion von Artefakten war und somit mehr Übung bedurfte. Nach der Durchführung des Hauptversuchs wurden die Teilnehmer gebeten, den oben erwähnten Fragebogen auszufüllen. Darin wurde nebst persönlicher Angaben erfasst, anhand welcher Merkmale die Unterschiede zwischen Simulation und Referenz erkannt worden waren. Bei Mehrfachnennungen sollten sie in eine Rangfolge der Offensichtlichkeit gebracht werden.

⁸Die unkompensierte Variante, die als Ankerreiz diente, wurde als verzögerter Diracstoß realisiert, der als ein "Filter" in der Datenbank gespeichert war.

4.2.2 Versuchsvorbereitungen

4.2.2.1 BRIR-Messung

Grundlage der binauralen Stimulation im Hörversuch war ein BRIR-Datensatz, der im Versuchsraum mit dem Messroboter FABIAN aufgenommen wurde. Da die Probanden direkt zwischen realem und simuliertem Schallfeld umschalten sollten, war es von großer Wichtigkeit, dass Mess- und Hörversuchsumgebung möglichst identisch waren, um Störvariablen konstant zu halten. Dies beinhaltete die Verwendung des Wiedergabelautsprechers als Messquelle, wodurch die in Abschnitt 1.3 angesprochene Quellkompensation hinfällig wurde. Der Messlautsprecher entsprach der zu simulierenden Quelle und übte somit keinen verfälschenden Einfluss auf die Übertragungsstrecke des Simulationssystems aus. Weiterhin war die Einhaltung derselben Sitzposition des Messroboters und der Versuchspersonen sowie eine konstante Lage der beweglichen Gegenstände im Raum zu beachten. Ersteres wurde durch eine entsprechende Markierung und die Verwendung ein und desselben Stuhls gewährleistet. Die Einrichtung des Studios wurde eingehend dokumentiert und wiederhergestellt. Auch der Versuchslaptop wurde bereits bei der Messung an seine spätere Position gesetzt, wie in Abb. 4.23(a) zu sehen ist. Da die Probanden im Hörversuch die Kopfhörer während der Lautsprecherdarbietung nicht abnahmen, wurde der Stax II⁹ für die BRIR-Messung auf den Kunstkopf gesetzt. Somit konnte ein identisches Beugungs- und Reflexionsverhalten am Hörerkopf bei realer und simulierter Wiedergabe eingehalten werden. Mit den beschriebenen Maßnahmen wurde angestrebt, dass nur das jeweilige Entzerrungsverfahren die Unterscheidbarkeit von Vergleichs- und Referenzreiz beeinflussen sollte.

Die BRIR-Messung erfolgte automatisch und für zwei Freiheitsgrade. Es wurde ein Bewegungsbereich von $\pm 65^{\circ}$ in horizontaler Ebene und -40° bis $+30^{\circ}$ in vertikaler Richtung abgedeckt. Sowohl Azimut als auch Elevation wurden mit einer Auflösungsgenauigkeit von 1° vermessen, sodass ein Datensatz von 9301 binarualen Raumimpulsantworten entstand. Laut den Versuchsergebnissen von Lindau et al. (2008) produziert bereits ein Raster von $2^{\circ}x 1^{\circ}$ (horizontal x vertikal) keine hörbaren Artefakte, sodass diese folglich bei der hier gewählten Auflösung ausgeschlossen werden konnten. Als Messsignal kam ein Sweep mit Bassanhebung, der FFT-Ordnung 17 und einer Abtastrate von 44,1 kHz zum Einsatz. Die Verteilung der erzielten Signal-Rausch-Abstände ist in Abb. 4.24(a) zu sehen und zeigt gute Resultate.

⁹Da sich der Stax Pro in seiner äußeren Form nicht vom Stax II unterscheidet, war keine getrennte Messung für die beiden Modelle nötig.



Abbildung 4.24 – (a) Verteilung der Signal-Rauschabstände der BRIR-Messung im Versuchsraum für linkes (schwarz) und rechtes (grau) Ohr; (b) ETC der linken BRIR für frontalen Schalleinfall

4.2.2.2 Lautheitskalibrierung

Im beschriebenen Versuchsdesign war es notwendig, dass die mit den Entzerrungsfiltern kompensierten Signale sowohl untereinander als auch im Vergleich zum Lautsprechersignal die gleiche Lautheit aufwiesen. Um diese Anforderung zu erfüllen, wurden in einem ersten Schritt die Filter auf Basis eines in Rec. ITU-R BS.1770 beschriebenden Algorithmus' kalibriert. Demnach ergibt sich die Lautheit eines Signals, indem es durch Hochpassfilterung spektral gewichtet und der zeitliche Mittelwert seiner Energie gebildet wird. Daraus kann ein gewichtetes äquivalentes Lautstärkemaß (*Equivalent Sound Level*) Leq(W) berechnet werden (Rec. ITU-R BS.1770, S. 8):

$$Leq(W) = 10 \log \left[\frac{1}{T} \int_0^T \frac{x_w^2}{x_{ref}^2} dt \right]$$
 (4.13)

 x_{ref} ist eine beliebige Referenz und x_w das hochpassgefilterte Eingangssignal. Die erforderlichen Filterkoeffizienten sind in Rec. ITU-R BS.1770 zwar für eine Abtastrate von 48 kHz angegeben sind, es stellte sich jedoch heraus, dass die Angaben auch bei $f_s = 44,1$ kHz gut funktionieren. Zur Kalibrierung eines diskreten Signals x(n) auf eine Referenzlautheit l_{ref} wurde Gl. 4.13 leicht abgewandelt und die folgende Berechnung ausgeführt:

$$l_x^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_w^2(n)$$
(4.14)

$$x_{\text{kal}}(n) = x(n) \cdot \frac{l_{\text{ref}}}{l_x}$$
(4.15)

Zunächst wurden so die beiden Testreize – rosa Rauschen und Gitarre – bezüglich der Lautheit des Gitarrensamples einander angeglichen. Diese Signale wurden anschließend mit den Impulsantworten der Entzerrungsfilter gefaltet, wobei hierzu auch der Diracstoß zählte, der zur Realisierung der ungefilterten Simulation verwendet wurde. Die Ausgangssignale der Faltungsoperationen wurden spektral gewichtet und dienten zur Lautheitsberechnung nach Gl. 4.14. Der geringste¹⁰ sich ergebende Wert wurde in Gl. 4.15 für l_{ref} eingesetzt, sodass die Filterimpulsantworten als Eingangssignale x(n) auf gleiche Lautheit kalibriert wurden. Die nochmalige Berechnung aller l_x zur Kontrolle ergab durchweg identische Werte. Auch der Höreindruck lieferte sehr zufriedenstellende Resultate.

Zur Angleichung der Lautheit des Kopfhörer- an diejenige des Lautsprechersignals, dessen Pegel 70 dB(A) betrug, wurde eine dem Herstellungsverfahren ähnliche Methode gewählt. Zwei Expertenhörer verglichen dabei im direkten Umschalten beide Reize und näherten sie mit einem fein aufgelösten Regler sukzessive einander an. Dies war ein genauer und ohne großen Aufwand realisierbarer Vorgang. Durch den numerisch ausgegebenen Wert des Reglers war es unproblematisch, die ermittelten Pegel im Hörversuch wiederherzustellen und für alle Versuchspersonen konstant zu halten. Insgesamt hatten also alle dargebotenen Signale die gleiche subjektive Lautstärke, sodass auch dies als offensichtliches Unterscheidungsmerkmal ausgeschlossen war.

4.2.2.3 Elektrische und elektroakustische Übertragung

Im Zuge der Versuchsvorbereitungen wurde ein Sweep-Signal in die Faltung mit den Entzerrungsfiltern (Faltung II, Abb. 4.23) eingespeist und direkt am Ausgang des Faltungsrechners wieder aufgenommen. Anhand der Übertragungsfunktion dieses Vorgangs – also <u>Sweep-Output</u> – konnte die korrekte Auswahl der Filter aus der Datenbank überprüft sowie eventuelle Übersteuerungen erkannt und behoben werden. Des Weiteren wurden die Kopfhörerübertragungsfunktionen wie in Abschnitt 2.2 mit dem Kunstkopf gemessen. Das Messsignal wurde diesmal jedoch vorher mit der Impulsantwort eines Entzerrungsfilters gefaltet, sodass die effektiv wirksame Kompensation erkennbar wurde. In Abb. 4.25 sind auf der linken Seite die Betragsfrequenzgänge des FFD-Filters mit Hochpass-Regularisierung¹¹ vor und nach Durchlaufen der Faltung abgebildet. Auf der rechten Seite sind fünf gemessene Ausgangsspektren nach elektroakustischer Übertragung durch die Kopfhörer und Kunstkopfmikrofone zu sehen, wobei jede Kurve eine unterschiedliche Kopfhörerposition repräsentiert.

Während die Filterfrequenzgänge vor und nach der elektrischen Übertragung sehr gut übereinstimmen, zeigen die gemessenen Entzerrungen deutliche Abweichungen vom defi-

¹⁰Damit sollten Übersteuerungen bei der späteren Faltung mit den Raumimpulsantworten vermieden werden.

¹¹Dieses hatte sich in einem ersten auditiven Vergleich der Verfahren als das optimalste herausgestellt.



Abbildung 4.25 – Links: Die schwarze Linie zeigt den Betragsfrequenzgang des FFD-Filters, die graue Linie die gemessene Übertragungsfunktion nach Durchlaufen der Faltung. Rechts: Fünf mit dem Kunstkopf gemessene Entzerrungsergebnisse: Die oberen Grafiken zeigen den linken Kanal des Stax Pro, unten ist der rechte Kanals des Stax II zu sehen.



Abbildung 4.26 – Fünf mit dem Kunstkopf gemessene FFD-Entzerrungsergebnisse des K1000 (rechter Kanal)

nierten Zielbandpass. Dies ist im Bassbereich besonders gravierend. Hier wird die untere Eckfrequenz von 50 Hz nur in Einzelfällen erreicht, und es ergeben sich zudem Überhöhungen, die beim Stax II beinahe 10 dB erreichen. Dies hängt wie bereits mehrfach erwähnt mit der starken Varianz der Frequenzgänge in Abhängigkeit der Kopfhörerposition zusammen. Dass die vorhandene Bassanhebung der Kopfhörer davon mitbetroffen ist, erhöht die Problematik zusätzlich, da diese bei zu großer Verschiebung nicht ausgeglichen werden kann. Die Abbildung zeigt für beide Kopfhörer denjenigen Kanal mit den schwerwiegenderen Abweichungen. Der hier nicht dargestellte rechte Kanal des Stax Pro weist auch ohne Entzerrung praktisch keine Überhöhung tiefer Frequenzen auf und war deshalb einfacher zu kompensieren. Es ist also davon auszugehen, dass ein Kopfhörer ohne Anhebung und mit geringen Verschiebungen im Bass eine deutlich bessere Entzerrung ermöglichen würde. Als Beispiel sind hierfür in Abb. 4.26 fünf gemessene Entzerrungen des K1000 gezeigt, die diese Vermutung bestätigen. Dort wird auch die untere Eckfrequenz des Zielbandpasses bei gleicher Filterlänge wie oben gut angenähert. Allerdings weist der Frequenzgang dieses Kopfhörers bereits ab ca. 15 kHz einen steilen Abfall auf, der vom Kompensationsfilter unzureichend angehoben wird. Dieser Umstand machte sich beim auditiven Vergleich mit dem Lautsprechersignal deutlich bemerkbar. Grundsätzlich waren die beschriebenen Messungen und das erstmals direkt vor dem Hörversuch erfolgte Probehören ernüchternd bezüglich der erwarteten Ergebnisse der perzeptiven Evaluation.

4.2.2.4 Vorversuch

Die optimalen β -Werte der Regularisierungsverfahren wurden in einem Vorversuch mit drei Expertenhörern ermittelt. Der Ablauf war prinzipiell gleich wie im Hauptversuch (siehe oben), allerdings wurde nur das rosa Rauschen als Testreiz verwendet, da es in diesem Zusammenhang um besonders empfindliche Ergebnisse ging. Für beide Regularisierungsfilter wurden die Gewichtungsfaktoren aus den Tabellen 4.1 und 4.2 eingesetzt und anhand des FFD-Verfahrens für beide Kopfhörer miteinander verglichen. Wie in Abschnitt 4.1 wurde die LS-Methode nicht getrennt untersucht, sondern die ermittelten Werte für sie übernommen. Der versteckte Anker wurde auch hier in die Bedingungsvariationen integriert.

Abb. 4.27 zeigt die Differenzgrad-Mittelwerte der Beurteilungen, die zur Ähnlichkeit von realem Schallfeld und dessen Simulation von den drei Versuchspersonen abgegeben wurden. Im linken Teil der Grafik sind die Ergebnisse der Kopfhörer einzeln zu sehen. Auffällig ist die sehr gute Bewertung des Stax Pro bei $\beta = 0, 4$ mit Hochpass-Regularisierung,¹² während die Resultate beim Stax II nur geringfügige Unterschiede aufweisen. Eine ähnliche Tendenz zeigt sich beim zweiten Regularisierungsfilter, allerdings ist hier die Bewertung des Stax II eindeutiger als die des Stax Pro. Da die Ergebnisse nur auf einer geringen Anzahl von Versuchspersonen basieren, kann hier schwerlich eine Aussage über deren Signifikanz getroffen

 $^{^{12}}$ Eine entsprechende Überprüfung zeigte jedoch keine weitere Steigerung bei größeren β -Werten.



(b) Regularisierung mit geglätteter Inverse

Abbildung 4.27 – Ergebnisse des Vorversuchs: Links sind die Differenzgrade von Stax Pro (grau) und Stax II (schwarz) für die Bewertungen der β -Werte getrennt aufgetragen, rechts sind die Mittelwerte über die Kopfhörer zu sehen.

werden. Wie oben erläutert, sollten bei beiden Kopfhörern identische Parameterwerte verwendet werden, sodass jeweils das β mit dem höchsten mittleren Differenzgrad ausgewählt wurde (Abb. 4.27, rechts). Für die Hochpass-Regularisierung wurde der Hauptversuch demnach mit $\beta = 0, 4$ durchgeführt, die Regularisierung mit geglätteter Inverse erfuhr eine Gewichtung mit $\beta = 0,07$.

4.2.3 Ergebnisse

4.2.3.1 Unterscheidungsmerkmale

Zunächst wird nachfolgend eine Übersicht über die Merkmale gegeben, welche die Versuchspersonen als relevant zur Detektion von Unterschieden zum realen Schallfeld nannten. Insgesamt war die korrekte Identifikation der Simulation ohne Probleme möglich, wenngleich dies einigen Teilnehmern beim Gitarrenstimulus erst nach kurzer Eingewöhnung gelang.¹³ Mehrere Personen bemerkten außerdem, dass die Vergleichsreize untereinander teilweise sehr ähnlich waren und deshalb das relative Rating keine leichte Aufgabe darstellte.

¹³siehe Trainingsphase (oben)



Abbildung 4.28 – Zusammenfassung der von den Versuchspersonen angegebenen Unterscheidungsmerkmale

In allen Fällen spielten klangliche Merkmale die größte Rolle, da sie ausschließlich oder in der geforderten Rangfolge an oberster Stelle genannt wurden. Einige Teilnehmer spezifizierten diesen Begriff, wobei anscheinend ausgeprägte Höhenanteile oder sogar Klingelartefakte dominanter waren als der fehlende Bass. Das Merkmal der Präsenz könnte sich ebenfalls auf hochfrequente Anteile beziehen. Die Wahrnehmung der entzerrten Simulation im Zeitbereich spielte offenbar eine Rolle bei der Bewertung, da dies bei der Gitarre mehrmals als Unterscheidungsmerkmal genannt wurde. Die Unterschiede in Lokalisation/Externalisierung, Quellenentfernung und Räumlichkeit hängen höchstwahrscheinlich nicht mit den Filtermethoden zusammen. Diese Aspekte werden in der Fehlerdiskussion weiter ausgeführt.

4.2.3.2 Post-screening

Zur nachträglichen Identifikation von Versuchspersonen, die aufgrund ihres Antwortverhaltens als ungeeignet eingestuft werden müssen, schlägt Rec. ITU-R BS.1116-1 ein *post-screening* vor. Darin soll mit Hilfe eines *t*-Tests die Nullhypothese geprüft werden, dass ein Proband primär geraten hat, da die Differenzgrade seiner Bewertungen in zufälliger Weise um den Wert 0 herum schwanken. Diese Art der Kontrolle ist jedoch nur dann möglich, wenn die Unterschiede zum Referenzstimulus kaum hörbar sind, denn andernfalls können auch gering geschulte Hörer korrekte Zuordnungen vornehmen (vgl. Rec. ITU-R BS.1116-1, S. 22). Da im vorliegenden Fall das Erkennen der Simulation wie gesagt leicht möglich war, konnte die beschriebene Methode also nicht angewandt werden. Nicht nur die Identifikation, sondern auch das zuverlässige Rating der Filtermethoden war jedoch ein wichtiger Bestandteil der Untersuchung, sodass das differenzierte Hören durch Experten trotzdem sichergestellt werden musste. Daher wurden die Antworten jeder Versuchsperson gesichtet, um offenkundige Fehler nach Möglichkeit auszuschließen.

Es fielen die Daten von zwei Probanden auf, die in Abb. 4.29 nach Contents getrennt dargestellt sind. Links sind jeweils die Differenzgrade der Bewertungen für Rauschen zu sehen, rechts diejenigen für Gitarre. Auf der *x*-Achse sind die Abkürzungen der Filter gemäß Tabelle 4.5 angezeigt. Bei beiden Personen waren die Beurteilungen des Rauschens ausschlag-



Abbildung 4.29 – Ergebnisse zweier Versuchspersonen, die aus dem Datensatz ausgeschlossen wurden: Links sind die Differenzgrade der Bewertungen für Rauschen zu sehen, rechts sind die Differenzgrade für Gitarre dargestellt.

gebend für den Zweifel an der Richtigkeit der Daten. Im einen Fall (Abb. 4.29(a), links) erfolgte anscheinend beim Stax II durchweg eine Verwechslung der Schieberegler für Referenz und Vergleichsreiz. Die Bewertungskurve erscheint daher an einer Achse, die ungefähr durch -3 verläuft, gespiegelt. Der positive Differenzgrad der zweiten Versuchsperson (Abb. 4.29(b), links) für die amplitudengeglättete Entzerrung des Stax II bedeutet eine falsche Zuordnung des Vergleichsreizes. Da alle anderen Teilnehmer die Simulation beim Rauschen stets erkannten und das betroffene Filter auch nicht als eines der besten eingestuft wurde, lässt diese Antwort auf mangelnde Konzentration oder Differenzierbarkeit schließen. Das sehr ähnliche Rating der restlichen Entzerrungen dieses Kopfhörers unterstützt diese Annahme. Entsprechend verhält es sich mit der Bewertung des äquivalent komplex geglätteten Filters im rechten Teil von Abb. 4.29(a), obwohl hier daran erinnert werden muss, dass die korrekte Identifikation der Referenz beim Gitarrenstimulus schwerer fiel. Der positive Differenzgrad in Abb. 4.29(b), rechts, ist daher vermutlich kein Fehler, zumal dieses Filter grundsätzlich hohe Bewertungen erhielt. Man hätte also bei der zweiten Versuchsperson lediglich die Daten bezüglich des Rauschens ausschließen können. Um jedoch Probleme mit *missing values* zu umgehen, und da auch nach Weglassen zweier kompletter Datensätze der Stichprobenumfang immer noch ausreichend groß war (s. Gl. 4.12), wurden alle Werte dieser beiden Versuchspersonen unberücksichtigt gelassen.

4.2.3.3 Statistische Auswertung

Die durchschnittlichen Differenzgrade der verbleibenden 26 Probanden sind in Abb. 4.30(a) für die einzelnen Faktorstufen mit 95%-igen Konfidenzintervallen als Maß für die Streuung zu sehen. Es wurde oben angesprochen, dass die Definition des unteren Skalenendes nicht eindeutig war und daher ein versteckter Ankerreiz eingeführt wurde, bei dem man von der jeweils schlechtesten Beurteilung ausgehen konnte. Die Abbildung bestätigt diese Annahme.

Die diffgrades jeder Versuchsperson wurden nachträglich bezüglich des geringsten Wertes normiert und in Prozent umgerechnet, sodass sich eine Skala von 0%-iger bis 100%-ige Ähnlichkeit zwischen simuliertem und realem Schallfeld ergab. Diese Werte sind in Abb. 4.30(b) dargestellt. Demnach betrug die Ähnlichkeit zwischen beiden Bedingungen beim Content Rauschen im besten Fall 54,85% (Stax Pro) und beim Content Gitarre 65,60% (Stax II). Dies entspricht leider nicht der erhofften kompletten Übereinstimmung, wobei sich hierfür verschiedene Gründe finden lassen, die weiter unten näher besprochen werden. Augenfällig ist zunächst die Tatsache, dass die Filtermethoden mit Hochpass-Regularisierung bessere Beurteilungen erhielten als die restlichen Verfahren. Die folgende Analyse wird Aufschluss über die Signifikanz der einzelnen Unterschiede geben.

Die Auswertung der Daten erfolgte mit Hilfe von SPSS®, wobei die ausführlichen Ergebnisse in Anhang B einzusehen sind. Es wurde bereits im Zusammenhang der Berechung der optimalen Stichprobengröße erklärt, dass die Bewertungen des Ankerreizes hierbei nicht mit eingeschlossen waren. Zunächst wurde die Reliabilität der Versuchspersonen untersucht, indem die Stichprobe geteilt und für beide Gruppen der α -Koeffizient nach Cronbach berechnet wurde. Damit kann die Verallgemeinerbarkeit eines aus mehreren Variablen gebildeten Er-



Abbildung 4.30 – Ergebnisse des Hauptversuchs für Rauschen (links) und Gitarre (rechts): (a) Mittelwerte der Differenzgrade; (b) Bewertungen bezogen auf das Rating des Ankers in %

gebnisses überprüft werden (vgl. Bortz 2005, S. 559).

$$\alpha = \frac{p}{p-1} \cdot \left(1 - \frac{\sum_{i} s_i^2}{s_{\text{tot}}^2} \right)$$
(4.16)

p ist dabei die Anzahl der Variablen (hier der Versuchspersonen), s_i^2 ihre Varianz und s_{tot}^2 die Varianz ihrer Summe. Die Berechnung ergab für die beiden Gruppen Werte von $\alpha = 0,86$ bzw. $\alpha = 0,94$, das heißt die Reliabilität der Daten ist hoch und hängt nicht von einem bestimmten Teil der Probanden ab. Dies ist sicherlich auch auf den Umstand der Messwiederholung zurückzuführen.

Weiterhin wurden die Bewertungen jeder Faktorstufe auf Normalverteilung geprüft, da dies eine Voraussetzung für ihre Analyse auf Grundlage von Mittelwerten war. Ein Kolmogoroff-Smirnov-Test bestätigte für alle Bedingungsvariationen die Nullhypothese (H₀) einer Normalverteilung.¹⁴

Mit Hilfe einer dreifaktoriellen Varianzanalyse (*analysis of variance*, ANOVA) wurde die Nullhypothese geprüft, dass sich die Mittelwerte der Differenzgrade nicht voneinander unterscheiden. Die nachfolgend angeführten Ergebnisse basieren auf den unveränderten Daten, allerdings ergaben sich bei einer Prüfung der skalierten Bewertungen kaum Abweichungen, sodass die Aussagen für beide gelten können. Als erstes wurde die H₀ "global" für die Haupteffekte und Interaktionen getestet. Erstere waren bei allen Faktoren hoch signifikant (p < 0,01), während dies bei letzteren nicht für alle Kombinationen zutraf:

- ♦ Methode x Kopfhörer: signifikant (p = 0,024)
- ♦ Methode x Content: hoch signifikant (p < 0,01)
- \diamond Kopfhörer x Content: nicht signifikant (p = 0, 222)
- \diamond Methode x Kopfhörer x Content: nicht signifikant (p = 0, 189)

Es kann also konstatiert werden, dass die wahrgenommene Güte der Kompensationsmethoden bezüglich der zu entzerrenden Übertragungsfunktionen differierte, und auch je nach dargebotenem Inhalt unterschiedlich beurteilt wurde. Dass die Bewertung der Kopfhörer nicht vom Teststimulus abhing, ist ein Beweis für die Konsistenz der Antworten, da ein gegenteiliges Resultat eher fragwürdig gewesen wäre. Die zur obigen Auflistung gehörenden Interaktionsdiagramme sind in Abb. 4.31 zu sehen. Daraus wird deutlich, dass die Entzerrungen des Stax II im Mittel über beide Stimuli meist bessere Bewertungen erhielten, und die Unterschiede zur Referenzsituation beim Rauschen erwartungsgemäß deutlicher wahrnehmbar waren.

Um zu ergründen, welche Mittelwertsunterschiede im Falle der Filtermethoden und der Interaktion Methode x Kopfhörer im Wesentlichen zum Verwerfen der H₀ beigetragen hatten, wurden a-posteriori-Einzelvergleiche durchgeführt. Dabei fand eine α -Fehler-Korrektur statt, die auch als Sidak-Korrektur bezeichnet wird und durch folgende Gleichung gegeben ist (Bortz 2005, S. 271):

$$\alpha' = 1 - (1 - \alpha)^{\frac{1}{m}} \tag{4.17}$$

 α ist das Signifikanzniveau (hier 0,05) und *m* die Anzahl der Einzelvergleiche. Es ergab sich, was auch aus Abb. 4.31 abzuleiten ist. Erstens lieferten die Methoden mit Hochpass-Regularisierung gemittelt über beide Kopfhörer und Stimuli hoch signifikant bessere Ergeb-

¹⁴Die Verteilungen sind ebenfalls in Anhang B zu finden.



(c) Kopfhörer x Content

Abbildung 4.31 – Interaktionsdiagramme

nisse als die anderen Verfahren, allerdings unterschied sich die Variante im Zeitbereich (LS) nicht signifikant von der Variante im Frequenzbereich (FFD). Die Beurteilungen der restlichen Filter wichen ebenfalls nicht signifikant voneinander ab. Zweitens führte eine Entzerrung der beiden Kopfhörer mit den besten Verfahren offenbar zu ihrer gleichwertigen Beurteilung, da die Nullhypothese beim Vergleich von Stax II und Stax Pro in diesen Fällen beibehalten wurde.

Da Abb. 4.30 vermuten lässt, dass die obigen Aussagen nicht unter beiden Stimulus-Bedingungen gleichermaßen zutreffen, wurden zusätzlich zwei nach Contents getrennte, zweifaktorielle Varianzanalysen sowie Einzelvergleiche¹⁵ auf allen Faktorstufen durchgeführt. Interessanterweise war der Haupteffekt des Faktors Kopfhörer im Falle des Rauschens nicht mehr signifikant, wies jedoch mit p = 0,063 immer noch eine klare Tendenz auf. Die ANOVA bezüglich des Gitarrenstimulus' ergab zwar nach wie vor hoch signifikante Unterschiede zwischen Stax II und Stax Pro (p < 0,01), allerdings traf dies nicht mehr auf die Interaktion Methode x Kopfhörer zu (p = 0,511).

Zunächst werden nun die einzelnen Bewertungen unter der Bedingung Rauschen näher

¹⁵mit Korrektur nach Gl. 4.17

betrachtet. Obwohl sich die Kopfhörer hier global offenbar nur tendenziell unterschieden, legt der linke Teil von Abb. 4.30 nahe, dass dies nicht bei jeder Entzerrungsmethode der Fall war. Es wurde oben festgestellt, dass die Entzerrung des Stax II derjenigen des Stax Pro im Mittel mindestens gleichgestellt, meistens jedoch überlegen war. Gerade bei einem der besten Verfahren, der FFD-Entzerrung mit Hochpass-Regularisierung, wurde der Stax Pro bei Darbietung des Rauschens jedoch hoch signifikant besser beurteilt. Seine Mittelwertskurve folgt der beschriebenen allgemeinen Tendenz, nach der die Kompensation mit Hochpass-Regularisierung am besten funktionierte, sich ihre Zeit- und Frequenzbereichsverfahren aber nicht unterschieden. Dies wurde auch durch Einzelvergleiche überprüft und bestätigt. Die Kurve des Stax II fällt jedoch bei der fraglichen FFD-Kompensation im Vergleich zur allgemeinen Tendenz leicht ab, was zur genannten Abweichung vom Stax Pro führt. Beim Stax II hebt sich in Abb. 4.30, links, dadurch nur noch die LS-Hochpass-Methode klar von den anderen ab. Einzelvergleiche der Filterbewertungen für diesen Kopfhörer legten ihre Einteilung in drei Gruppen nahe. Zur ersten gehört demzufolge nur die LS-Hochpass-Regularisierung, die sich von allen Methoden außer ihrer Variante im Frequenzbereich und der Amplitudenglättung signifikant unterschied.¹⁶ Diese beiden Verfahren bilden die zweite Gruppe, da ihre Bewertungen weder von der höchsten noch von den anderen überzufällig abwichen. Die letzte Gruppe wiederum beinhaltet die restlichen Filter, die alle signifikant schlechter als die LS-Methode waren.

Die Einzelvergleiche der Methoden unter der Bedingung Gitarre bestätigten weiter die geringere Empfindlichkeit dieses Stimulus'. Demnach waren die Differenzen unter den Stax II-Entzerrungen alle statistisch nicht signifikant, obwohl im rechten Teil von Abb. 4.30 in der Grundtendenz ähnliche Ergebnisse wie beim Stax Pro zu sehen sind. Hier schnitt wieder die FFD-Hochpass-Entzerrung am besten ab, wobei für den Vergleich zur Version im Zeitbereich die Nullhypothese wie bisher angenommen werden musste. Das LS-Hochpass-Verfahren hob sich lediglich von den Regularsierungsmethoden mit oktavgeglätteter Inverse und von der äquivalent komplexen Glättung signifikant bzw. in Tendenz ab. Die Amplitudenglättung und das Compare and Squeeze-Filter können ihr also gleichgestellt werden, allerdings wiesen diese überzufällige Unterschiede zur FFD-Hochpass-Methode auf. Die Vergleiche der Methoden beim Gitarrenstimulus sind insgesamt mit Vorsicht zu betrachten, da hier wie erwähnt laut Aussage der Versuchspersonen das relative Rating schwierig war. Überdies könnte die eher strenge Sidak-Korrektur ebenfalls einen Einfluss auf die Ergebnisse gehabt haben.

¹⁶Es könnte allerdings sein, dass die fehlende Signifikanz auf die eher konservative α -Fehler-Korrektur zurückzuführen ist.

4.3 Diskussion

4.3.1 Interpretation der Ergebnisse

Bezüglich der dargebotenen Testreize wurde festgestellt, dass bei der Gitarre weniger deutliche Unterschiede zur Referenzsituation wahrgenommen wurden als beim rosa Rauschen. Dies ist kaum verwunderlich und wurde bereits bei der Auswahl der Stimuli erwartet. In Abb. 4.22 ist zu sehen, dass das natürliche Instrument gerade in den für die Kopfhörerentzerrung als problematisch identifizierten Bereichen Bass/sehr hohe Frequenzen wenig Energie enthält. Dies führte einerseits dazu, dass die Filtermethoden untereinander schwer zu unterscheiden waren, wie die Ergebnisse der Einzelvergleiche unter dieser Bedingung nahe legen. Die Prüfung der Interaktionen ergab darüber hinaus, dass die insgesamt festgestellte Abhängigkeit der Kompensationsgüte vom entzerrten Frequenzgang bei diesem Stimulus nicht wahrnehmbar war. Das Rauschen ermöglicht also eher Aussagen über die relative Bewertung der Entzerrungsverfahren, während die Ergebnisse der Gitarre anzeigen, wie gut das binaural simulierte Schallfeld von der realen Situation bei Darbietung eines natürlichen Stimulus' unterscheidbar ist. Die erzielbare Ähnlichkeit liegt hier relativ zum Ankerreiz bei maximal 65,60%.

Bei der Festlegung der Filterparameter in Abschnitt 4.1 wurde anhand des berechneten frequenzabhängigen Fehlermaßes festgestellt, dass die Frequenzgangsvarianz in Abhängigkeit der Kopfhörerposition beim Stax Pro schwer zu kompensieren ist. Bei Verfahren mit geringer Robustheit gegenüber verschobenen dips wies dieser Kopfhörer höhere Fehlerwerte auf als der Stax II. Die objektive Berechnung wurde durch die perzeptive Evaluation bestätigt, da die Entzerrungen des Stax II von den Versuchspersonen eher besser beurteilt wurden. Dies traf allerdings nicht bei den zwei optimalsten Verfahren zu, was dafür spricht, dass Methoden mit ausreichender Robustheit auch stark variierende Frequenzgänge gut kompensieren können. Die schlechte perzeptive Beurteilung der restlichen Entzerrungen des Stax Pro lässt vermuten, dass bei wenig robusten Filtern auftretende Klirrartefakte dominante Unterscheidungsmerkmale beim relativen Rating waren. Dies wird durch Angaben der Versuchspersonen weiter bestätigt (s. Abb. 4.28). Es ist möglicherweise denkbar, dass die in manchen Fällen vorhandene Bassanhebung im entzerrten Frequenzgang des Stax II¹⁷ dessen unzureichende Annäherung an die untere Grenzfrequenz des Zielbandpasses perzeptiv "vertuschte", und er auch deshalb als ähnlicher beurteilt wurde als der Stax Pro.

¹⁷vgl. Abb. 4.25

Bezüglich der Verfahren stellte sich die Regularisierung mit Hochpassfilter klar als die perzeptiv am besten geeignete Methode zur Kopfhörerkompensation heraus. Diese hatte sich auch in Abschnitt 4.1 als ziemlich robust erwiesen. Zwischen der LS- und FFD-Variante war kein Unterschied festzustellen. Etwas relativiert wurde diese Aussage durch die schlechtere Bewertung der Stax II-FFD-Entzerrung unter der Bedingung Rauschen, die derjenigen des Zeitbereichsverfahrens zwar immer noch gleichzustellen war, sich allerdings auch nicht von den restlichen Filterbeurteilungen abhob. Es könnte daher sein, dass sich – ähnlich wie in der erwähnten Untersuchung von Norcross et al. (2002) - Schwankungen im Bassbereich, wie sie beim Stax II in Form einer Anhebung auftreten, für den Filterentwurf im Frequenzbereich als problematisch erweisen, da die lineare Auflösung der Fourier Transformation für eine exakte Repräsentation tiefer Frequenzen ungeeignet ist. Dieses Resultat der perzeptiven Evaluation erstaunt insofern, als dass die FFD- und LS-Methoden in Abschnitt 4.1 durchweg identische Fehlermaße aufwiesen. Aus der Perspektive des Rechenaufwands ist die Gleichwertigkeit der beiden Varianten erfreulich, da der Algorithmus im Frequenzbereich erheblich schneller ist. Die Amplitudenglättung reichte zwar im globalen Ergebnis nicht an die Regularisierungsmethoden mit Hochpassfilter heran, erhielt jedoch im Einzelfall der Stax II-Entzerrung unter der Bedingung Rauschen eine mittlere Bewertung. Für Frequenzgänge mit wenig positionsbedingter Varianz im hochfrequenten Bereich könnte die Oktavglättung also ebenfalls zur Kompensation geeignet sein. Hier ist noch zu erwähnen, dass zwar transiente Anteile zur Detektion von Unterschieden zum realen Schallfeld dienten,¹⁸ zwischen der Bewertung der Amplituden- und der äquivalent komplexen Glättung aber keinerlei Differenz beobachtet werden konnte. Dieses Ergebnis deckt sich nicht mit dem berechneten Zeitbereichsfehlermaß, wo erhebliche Abweichungen zwischen beiden Verfahren zu beobachten waren. Die Amplitudenglättung wäre wegen ihrer immerhin dreimal kürzeren Rechenzeit zu bevorzugen.

Wie bereits vermutet wurde, stellt die in den Tabellen 4.6 und 4.7 aufgeführte Rangfolge der Entzerrungsverfahren keine zuverlässige Vorhersage der perzeptiven Evaluation dar, da die Robustheit der Filter offenbar ein sehr dominantes Kriterium ist. Die Aussagen, die diesbezüglich anhand des frequenzabhängigen Fehlermaßes getroffen wurden, stimmen jedoch gut mit den Resultaten des Hörversuchs überein.

Die hier vorgestellten Ergebnisse decken sich zwar insoweit mit den diversen Untersuchungen von Norcross et al.¹⁹ zur Filterinversion, als dass die Regularisierung mit Hochpassfilter dort ebenfalls sehr gute Ergebnisse hervorbrachte. Es ergaben sich aber grundsätzlich

¹⁸s. Abb. 4.28

¹⁹zusammengefasst in (Norcross et al. 2004b)

bessere Bewertungen der entzerrten Übertragungsfunktionen als im vorliegenden Versuch. Allerdings wurde als Teststimulus meist nur ein Castagnettensample verwendet, das im Zeitbereich sicher empfindlich war, dem aber wichtige Frequenzkomponenten zur zuverlässigen Bewertung spektraler Filtereigenschaften fehlten. Außerdem wurde die Entzerrung von Lautsprechern untersucht, wo das hier dominante Problem der Frequenzgangsverschiebungen in anderem Maße ausgeprägt ist.

4.3.2 Fehlerquellen

Nebst der trotz Kopfhörerentzerrung weiterhin vorhandenen spektralen Einflüsse auf die binaurale Simulation, sind ein paar andere Fehlerquellen zu nennen, die möglicherweise ebenfalls zur Unterscheidung vom realen Schallfeld beitrugen.

Dies betrifft als erstes die relative Lage der Versuchsteilnehmer im Schallfeld. Wie oben erwähnt, wurde zwar für die Verwendung desselben Stuhls und die Einhaltung seiner einheitlichen Position während BRIR-Akquise und Hörversuch gesorgt. Die Größe der Probanden und dadurch ihre Sitzhöhe sowie die Neigung des Oberkörpers beim Hören waren jedoch schwer kontrollierbare Störvariablen, die zu einer Inkongruenz der verglichenen Schallfelder führten.

Genauso verhält es sich mit der Tatsache, dass die BRIRs aus Messungen mit einem HATS stammten und daher nicht mit den individuellen HRTFs der Versuchspersonen übereinstimmten. Dass die "aufgesetzte" Außenohr-Übertragungsfunktion nicht mit der gewohnten zusammenpasst, führt bei binauralen Simulationen oftmals zu erschwerter Lokalisation (vgl. Weinzierl 2008, S. 673). Dieser Aspekt wurde hier möglicherweise noch durch den Umstand verstärkt, dass der verwendete Kunstkopf breiter ist als der Bevölkerungsdurchschnitt, wodurch mit ihm gemessene BRIRs eine tendenziell zu große interaurale Laufzeitdifferenz aufweisen.²⁰ Es wurde in Abschnitt 1.2 erklärt, dass diese bei der Ortung von Schallereignissen eine wichtige Rolle spielt. Unter den genannten Gesichtspunkten können die von den Probanden angegebenen Unterscheidungsmerkmale, welche die Lokalisation bzw. Entfernung der Quelle betreffen, also leicht erklärt werden.

Zuletzt sei noch auf die Latenz des Simulationssystems hingewiesen. Aktuelle Untersuchungen am Fachgebiet ergaben, dass das System selbst eine durchschnittliche Latenz von 50,6 ms aufweist, die zwischen 47,8 ms und 66,4 ms schwanken kann.²¹ Hierzu muss die vom modelling delay der Entzerrungsfilter verursachte Verzögerung von 23,2 ms (vgl. Gl. 4.10) ad-

²⁰Derzeit finden am Fachgebiet Bemühungen zur Anpassung an durchschnittliche Werte statt.

²¹Diese Angaben gelten für eine wie im Hörversuch verwendete Blockgröße von 256 samples bei der schnellen Faltung.

diert werden, sodass sich eine Gesamtlatenz von 73,8 ms ergibt. Dieser Wert liegt über der in Hörversuchen ermittelten Schwelle von 60 ms (vgl. Weinzierl 2008, S. 677). Verzögerungen des Simulationssystems machen sich dann bemerkbar, wenn durch den head tracker die Faltung mit einem neuen BRIR-Paar ausgelöst wird, also bei jeder Kopfbewegung, deren Größe die gemessene Winkelauflösung übersteigt. Bei der Evaluation der Filtermethoden führten die Hörer zwar keine exzessiven Kopfbewegungen aus, trotzdem kann auch die Systemlatenz zu hörbaren Abweichungen von der Realität geführt haben. Zudem stellt dies ein grundsätzliches Problem dar, da es eine Einschränkung der zur Entzerrung verwendbaren Filterlängen bedeutet. Unter diesem Gesichtspunkt wäre eventuell eine minimalphasige Kompensation eher geeignet. Wie in Abschnitt 3.1 erwähnt wurde, bringt sie jedoch andere Probleme mit sich, außerdem müsste dann auch der Zielbandpass minimalphasig sein. Die Eignung einer solchen Entzerrung könnte auf Grundlage der als optimal ermittelten Filtermethoden (FFD bzw. LS) Gegenstand weiterer Untersuchungen sein. In diesem Zusammenhang sei die aktuelle Arbeit von Norcross et al. (2006) erwähnt, welche darauf abzielt, die Regularisierung direkt in eine minimalphasige Zielfunktion einzuschließen. Dies soll ein perzeptiv störendes Einschwingverhalten der entzerrten Impulsantwort vermeiden, das bei der herkömmlichen Regularisierungsmethode auftritt. Eine eingehende Evaluation hierzu steht allerdings noch aus.

4.3.3 Schlussfolgerungen

Insgesamt stellte die Veränderung der Kopfhörerfrequenzgänge durch variable Aufsetzpositionen ein gravierendes Problem dar, das sich nicht nur im hoch- sondern auch im tieffrequenten Bereich manifestierte. Der Verschiebung im Bass ist mit einer Regularisierung jedoch schwer beizukommen, da dort keine schmalbandigen Einbrüche auftreten, deren Kompensation aus perzeptiven Gründen eher vernachlässigt werden kann. Vielmehr variiert bei den untersuchten Kopfhörern die untere Grenze ihres Übertragungsbereichs und ist teilweise gepaart mit einer sich ebenfalls stark verschiebenden Anhebung der tiefen Frequenzen. Der fehlende Bassanteil trug im Hörversuch maßgeblich dazu bei, dass die Simulation im Vergleich zum realen Schallfeld sofort als solche erkannt wurde. Das allgemein existierende Problem, dass bei Lautsprecherwiedergabe Vibrationen auftreten, die bei Kopfhörerwiedergabe nicht vorhandenen sind, verstärkte den Eindruck fehlender Tiefen sicherlich.

Die geschilderten Schwierigkeiten führen zur Schlussfolgerung, dass sich die verwendeten Kopfhörer auf Grund ihrer Eigenschaften nicht befriedigend entzerren lassen. Für die Anwendung in der Binauraltechnik, wo eine Linearisierung der spektralen Einflüsse nötig ist, wären Modelle besser geeignet, die zumindest im Bass keine starke Verschiebungen und einen glatten Frequenzgang aufweisen. Denkbar wäre als Alternative der AKG K-1000, dessen Bandbeschränkung oberhalb von 15 kHz jedoch ungenügend kompensiert werden konnte.²² Eine weitere Variante wäre ein selbst gebautes Modell, ähnlich wie das von Møller et al. (1995a) beschriebene,²³ bei dem zwei kleine Lautsprecher gänzlich frei vor den Ohren des Hörers platziert wurden. Dadurch wären die Frequenzgänge weniger stark den Schwankungen durch veränderte Ohrhörerposition unterworfen. Allerdings müsste hier das Übersprechen bedacht werden, was auch schon im Zusammenhang des K1000 weiter oben erwähnt wurde.

An dieser Stelle sei nochmal an das hochempfindliche Versuchsdesign der vorgestellten perzeptiven Evaluation erinnert. Durch den unmittelbaren Vergleich mit einer gekennzeichneten Referenz waren Abweichungen der Simulation vom realen Schallfeld leicht erkennbar. Auch wurde nicht wie bei Moldrzyk et al. (2005) oder Lindau und Weinzierl (2007) die Plausibilität des Systems untersucht, sondern nach seiner Ähnlichkeit mit der Referenzsituation gefragt. Der ambitionierte Anspruch, durch Entzerrung ein binaural simuliertes Signal seiner realen Entsprechung bis zur Unhörbarkeit anzunähern, war wegen der genannten Probleme jedoch nicht möglich.

²²s. Abb. 4.26

²³vgl. Abschnitt 1.3

5 ZUSAMMENFASSUNG UND AUSBLICK

Bei der Auralisation mittels dynamischer Binauralsynthese tragen besonders spektrale Verzerrungen zur Unterscheidbarkeit vom realen Schallfeld bei. Diese Einflüsse sind auf die aufnahme- und wiedergabeseitigen elektroakustischen Wandler zurückzuführen und müssen kompensiert werden. Die Analyse der Übertragungsstrecke eines Simulationssystems zeigte, dass sich die bei der BRIR-Akquise eingesetzten Kunstkopfmikrofone und die Wiedergabekopfhörer gleichzeitig entzerren lassen, sofern ihre Messung simultan erfolgt.

Es wurden die Frequenzgänge mehrerer Kopfhörer mit verschiedenen Eigenschaften bestimmt, wobei sich deutliche interindividuelle Unterschiede erkennen ließen. Dies führte zur Feststellung, dass für einen Kopfhörer, der zur Wiedergabe binaural simulierter Signale verwendet werden soll, auf Grundlage seiner gemessenen Übertragungsfunktion ein spezifisches Entzerrungsfilter entworfen werden muss. Des Weiteren zeigten sich bei unterschiedlichen Aufsetzpositionen der Kopfhörer große intraindividuelle Differenzen der Frequenzgänge. Sie manifestierten sich vor allem in einer variablen Lage und Tiefe schmalbandiger Einbrüche bei hohen Frequenzen. In manchen Fällen war auch der Bassbereich von einer nicht unerheblichen Verschiebung betroffen.

Die Schwierigkeiten, welche sich im Zusammenhang der Filterinversion ergeben, wurden thematisiert und Ansätze zu ihrer Lösung aufgezeigt. So kann die Kausalität eines Kompensationsfilters entweder durch minimalphasige Invertierung oder durch zeitliche Verzögerung der akausalen Impulsantwort gewährleistet werden. Um bei der Entzerrung exzessive Anhebungen außerhalb eines bestimmten Bereichs zu vermeiden, muss eine geeignete Zielfunktion definiert werden, die hier als ein linearphasiges Bandpassfilter entworfen wurde. Im Bezug auf eine erhöhte Robustheit gegenüber Verschiebungen des zu kompensierenden Frequenzgangs, wurden mehrere Herangehensweisen vorgestellt. Sie alle beruhen auf einer Verringerung der Entzerrungsgenauigkeit schmalbandiger Einbrüche, um im Gegenzug perzeptiv störende Überhöhungen im kompensierten Amplitudengang zu reduzieren. Die Implementierung und Gegenüberstellung einiger Verfahren sollte Aufschluss über deren perzeptive Eignung zum Ausgleich von spektralen Verzerrungen im Kontext der Binauraltechnik geben. Erste Hinweise über die Effizienz der ausgewählten Entzerrungsalgorithmen lieferte die Bestimmung von Fehlermaßen im Zeit- und Frequenzbereich sowie ein Vergleich der Berechnungszeiten. Auf Basis simulierter Kompensationsergebnisse konnten die Parameter der meisten Filtermethoden festgelegt und Aussagen über die erzielbare Robustheit getroffen werden. Dabei zeigte sich, dass die geschilderte Verschiebung schmalbandiger Einbrüche bei veränderten Aufsetzpositionen eine Herausforderung darstellt, der nur mit erheblicher Reduktion der Kompensationsgenauigkeit begegnet werden kann. Die hierbei erreichbaren Resultate hängen überdies von der zu entzerrenden Übertragungsfunktion ab, wie die Gegenüberstellung zweier Kopfhörer deutlich machte.

Mit Hilfe eines Hörversuchs im ABC/HR-Design erfolgte die perzeptive Evaluation der implementierten Kompensationsmethoden. Die Filter wurden abwechselnd zur Entzerrung einer binauralen Simulation verwendet, welche direkt mit dem realen Schallfeld verglichen wurde. Auch hier kamen zwei Kopfhörer zum Einsatz, ferner wurden zwei Testreize mit differierenden Eigenschaften verwendet. Ein Gitarrensample repräsentierte einen natürlichen Stimulus mit transienten Anteilen, rosa Rauschen sollte die mühelose Detektierbarkeit vorhandener Artefakte ermöglichen. Die Abhängigkeit der Entzerrungsgüte vom eingesetzten Wiedergabekopfhörer konnte zumindest bei letzterem Stimulus empirisch bestätigt werden. Der Hörversuch zeigte außerdem, dass von den verglichenen Verfahren der Filterentwurf mit frequenzabhängiger Regularisierung, die als Hochpassfilter ausgelegt ist, zur perzeptiv besten Kompensation binauraler Simulationen führt. Bemerkenswert ist, dass diese Methode in einem weiten Bereich der hohen Frequenzen eine ungenaue Kompensation bewirkt. Die restlichen Verfahren scheinen dagegen zu individuell an eine spezifische Übertragungsfunktion angepasst zu sein, obwohl auch sie deren Verzerrungen nicht perfekt ausgleichen. Die genannte Kompensationsmethode führte auch bei demjenigen Kopfhörer zu guten Ergebnissen, der eine ausgeprägte Verschiebung schmalbandiger Einbrüche bei variablen Aufsetzpositionen aufwies. Ein Nachteil der Regularisierung besteht allerdings in der Notwendigkeit, ihre optimale Gewichtung perzeptiv zu ermitteln. Dieser Umstand macht sie zu einem relativ aufwändigen Verfahren, wenngleich die Berechnungszeit des Filters bei einer Implementierung im Frequenzbereich sehr kurz ist.

Die Bewertungen der Ähnlichkeit von realem und simuliertem Schallfeld lagen insgesamt auf einem recht tiefen Niveau. Die angestrebte perzeptive Gleichheit ließ sich mit keinem der implementierten Methoden erzielen. Dies ist auf darauf zurückzuführen, dass die Verschiebung des tieffrequenten Bereichs bei veränderter Kopfhörerposition ein schwer lösbares aber sehr dominantes Problem darstellte. Offenbar setzte hier die Charakteristik der verwendeten Kopfhörer ihrer zufrieden stellenden Entzerrung gewisse Grenzen. Die variable untere Grenze des Übertragungsbereichs in Zusammenhang mit einer Bassanhebung erwies sich hierbei als äußerst kritisch.

Die Entzerrungsmethode, die sich im Rahmen dieser Arbeit als besonders geeignet herausstellte, kann als Grundlage für künftige Untersuchungen zur weiteren Annäherung binauraler Simulationen an ihre realen Entsprechungen dienen. Diesbezüglich sind mehrere Aspekte der Optimierung denkbar. Zum einen ist die Wahl oder Entwicklung eines Kopfhörers zu nennen, der sich für die Wiedergabe verzerrungsfreier binauraler Signale besser eignet. Dies setzt in erster Linie einen Frequenzgang voraus, dessen Bassbereich keine Anhebung und einen von der Aufsetzposition unabhängigen Verlauf aufweist. Ein Kopfhörer mit frei vor den Ohren fixierten Ohrhörern scheint hierbei eine viel versprechende Option zu sein. Darüber hinaus wäre der Einsatz eines Subwoofers möglich, der die bei Kopfhörerwiedergabe fehlenden tieffrequenten Vibrationen ausgleicht. Bezüglich des Kompensationsverfahrens selbst, könnte die Regularisierungsgewichtung eingehender untersucht und individuell für beide Kopfhörerkanäle bestimmt werden. Vor dem Hintergrund der unerwünschten Latenz müsste die minimalphasige mit der hier verwendeten gemischtphasigen Entzerrung verglichen und evaluiert werden. Dies würde auch die Definition eines minimalphasigen Zielbandpasses beinhalten, der eventuell die Regularisierung direkt enthalten könnte und dessen perzeptive Wirkung ebenfalls untersucht werden müsste.

ABBILDUNGSVERZEICHNIS

1.1 1.2	Kopfbezogenes Polarkoordinatensystem	3 4
1.3	System zur dynamischen Binauralsynthese	5
1.4	Schallübertragung im Außenohr im freien Schallfeld	7
1.5	Schallübertragung im Außenohr bei Kopfhörerwiedergabe	8
1.6	Elektroakustische Übertragung im Simulationssystem	10
2.1	Übertragungsmaß DPA 4060 Messmikrofone	14
2.2	Messaufbau Nenn-Übertragungsfaktoren	15
2.3	Gemessene Kopfhörer	19
2.4	Messaufbau zur Bestimmung der Kopfhörerfrequenzgänge	20
2.5	Betragsfrequenzgänge von Audio-Technica, AKG K-401, Sennheiser Headset und AKG K-1000	24
2.6	Betragsfrequenzgänge von Stax SRS 2020 Lambda Basic, Stax SRS 2050 II und Stax Lambda Pro New	25
2.7	Phasengänge von Audio-Technica, AKG K-401, Sennheiser Headset und AKG K-1000	26
2.8	Phasengänge von Stax SRS 2020 Lambda Basic, Stax SRS 2050 II und Stax Lamb-	20
20	da Pro New	27
2.9	Varianz von Audia Tachnica, AVC V 401 Sannhaiser Haadaat und AVC V 1000	20
2.10 2.11	Varianz von Audio-Technica, AKG K-401, Sentineiser Headset und AKG K-1000 Varianz von Stax SRS 2020 Lambda Basic, Stax SRS 2050 II und Stax Lambda	29
0.10	Pro New	30
2.12	Ubersprechen	32
3.1	Zielbandpass	36
3.2	Hochpass-Regularisierungsfilter	45
3.3	Inverses Regularisierungsfilter	46
3.4	Zeitselektive Glättung	48
3.5	Amplituden- und komplexe Glättung	51
3.6	Compare and Squeeze	53
3.7	Mittelung	56
3.8	Vortensterung	57
3.9		58
3.10	Zusammenfassung der Filtermethoden	59
4.1	ERB-Filterbank	64
4.2	FFD-Kompensation mit Hochpass-Regularisierung und verschiedenen β	66
4.3	FFD-Kompensation des Stax Pro mit Filterlänge 2^{15}	66
4.4	Fehler im Frequenzbereich bei FFD-Kompensation mit Hochpass- Regularisierung und verschiedenen β	67

4.5	FFD-Kompensation mit oktavgeglättetem, invertiertem Regularisierungsfilter und verschiedenen β
16	Eabler im Erequenzbereich hei FED Komponsation mit ektavgeglättetem inver
4.0	tiertem Bogularisierungsfilter und verschiedenen β 70
17	Smoothing Komponentian mit verschiedenen Clättungehandbreiten 72
4.7	Shoothing-Kompensation interesting Kompensation wit searching of 2
4.0	Clätten sehen densiten
4.0	Gattungsbandbreiten
4.9	Compare and Squeeze-Kompensation
4.10	Fenier im Frequenzbereich bei Compare and Squeeze-Kompensation
4.11	All-Pol-Kompensation mit verschiedenen Oranungen p
4.12	Fenier im Frequenzbereich bei All-Pol-Kompensation mit verschiedenen Ord-
1 1 0	nungen p
4.13	Frequenzabhangiger Fehler ausgewahlter Verfahren
4.14	Kompensationsergebnisse FFD mit Hochpass-Kegularisierung
4.15	Kompensationsergebnisse LS mit Hochpass-Regularisierung
4.16	Kompensationsergebnisse FFD mit invertiertem Regularisierungsfilter 84
4.17	Kompensationsergebnisse LS mit invertiertem Regularisierungsfilter
4.18	Kompensationsergebnisse Amplitudenglättung
4.19	Kompensationsergebnisse äquivalente komplexe Glättung
4.20	Kompensationsergebnisse Compare and Squeeze
4.21	GUI des ABC/HR-Tests
4.22	Stimuli
4.23	Versuchsraum und -aufbau
4.24	BRIR-Messung
4.25	Elektrische und elektroakustische Übertragung Stax Pro und Stax II 102
4.26	Elektroakustische Übertragung K1000
4.27	Ergebnisse des Vorversuchs
4.28	Unterscheidungsmerkmale
4.29	Ergebnisse der ausgeschlossenen VPs 106
4.30	Ergebnisse des Hauptversuchs 108
4.31	Interaktionsdiagramme
B.1	Verteilung der Differenzgrade
В.2	Verteilung der angepassten Differenzgrade

TABELLENVERZEICHNIS

2.1	Technische Spezifikationen der Messmikrofone	13
2.2	Eigenschaften der gemessenen Kopfhörer	18
4.1	Fehlerwerte bei FFD-Entzerrung mit Hochpass-Regularisierung und verschie-	
	denen β	68
4.2	Fehlerwerte bei FFD-Entzerrung mit oktavgeglättetem, invertiertem Regulari-	
	sierungsfilter und verschiedenen β	71
4.3	Fehlerwerte bei Compare and Squeeze-Kompensation	74
4.4	Fehlerwerte bei All-Pol-Kompensation für verschiedene Ordnungen p	75
4.5	Abkürzungen der Verfahren	78
4.6	Zeitbereichsfehler	80
4.7	Frequenzbereichsfehler	80
4.8	Rechenaufwand	89

LITERATURVERZEICHNIS

- Behler GK, Pollow M (2008) Variable Richtcharakteristik mit Dodekaeder-Lautsprechern. DAGA Dresden, Fortschritte der Akustik, S. 67-68
- Blauert J, Laws P (1978) Group delay distortions in electroacoustical systems. J. Acoust. Soc. Am. 63/5:1478-1483
- Blauert J, Braasch J (2008) Räumliches Hören. In: Weinzierl S (Hrsg.) Handbuch der Audiotechnik. Springer, Berlin/Heidelberg, S. 87-121
- Bortz J, (2005) Statistik für Human- und Sozialwissenschaftler. 6. Auflage, Springer, Berlin/Heidelberg
- Bortz J, Döring N (2002) Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 3. Auflage, Springer, Berlin/Heidelberg
- Bücklein R (1981) The audibility of frequency response irregularities. J. Audio Eng. Soc. 29/3:126-131
- Clarkson PM, Mourjopoulos JN, Hammond JK (1985) Spectral, phase, and transient equalizsation for audio systems. J. Audio Eng. Soc. 33/3:127-132
- Dau T, Püschel D, Kohlrausch A (1996) A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. J. Acoust. Soc. Am. 99/6:3615-3622 dpa
- DPA Microphones, Miniature Microphones, Specifications: http://www.dpamicrophones.com/ (letzter Zugriff: 24.09.2008)
- Fielder LS (2001) Practical limits for room equalization. 111th AES Convention, New York, Preprint 5481
- Hammershøi D, Møller H (1996) Sound transmission to and within the human ear canal. J. Acoust. Soc. Am. 100:408-427
- Hatziantoniou PD, Mourjopoulos JN (2000) Generalized fractional-octave smoothing of audio and acoustic responses. J. Audio Eng. Soc. 48/4:259-280
- Hatziantoniou PD, Mourjopoulos JN (2004) Real-time room equalization based on complex smoothing: Robustness results. 116th AES Convention, Berlin, Preprint 6070
- Hayes MH (1996) Statistical Digital Signal Processing and Modelling. John Wiley & Sons, New York/Chichester
- Huopaniemi J, Smith JO (1999) Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters. Proc. AES 16th Int. Conference on Spatial Sound Reproduction, Rovaniemi, S. 301-312
- Johansen LG, Rubak P (1996) The excess phase in loudspeeker/room transfer functions: Can it be ignored in equalization tasks? 100th AES Convention, Kopenhagen, Preprint 4181
- Karamustafaoglu A, Horbach U, Pellegrini R, Mackensen P, Theile G (1999) Design and applications of a data-based auralisation system for surround sound. 106th AES Convention, München, Preprint 4976
- Karjalainen M, Paatero T, Mourjopoulos JN, Hatziantoniou PD (2005) About room response equalization and dereverberation. IEEE Workshop on ASPAA, S. 183-186

- Kirkeby O, Nelson PA, Hamada H, Orduna-Bustamante F (1998) Fast deconvolution of multichannel systems using regularization. IEEE Trans. on SAP 6/2:189-194
- Kirkeby O, Nelson PA (1999) Digital filter design for inversion problems in sound reproduction. J. Audio Eng. Soc. 47/7:583-595
- Leckschat D (1992) Verbesserung der Wiedergabequalität von Lautsprechern mit Hilfe von Digitalfiltern. Dissertation, RWTH Aachen
- Lentz T (2006) Dynamic cross talk cancellation for binaural synthesis in virtual reality environments. J. Audio Eng. Soc. 54/4:283-294
- Lindau A (2006) Ein Instrument zur softwargestützten Messung binauraler Raumimpulsantworten. Magisterarbeit, Technische Universität Berlin
- Lindau A, Weinzierl S (2007) FABIAN. Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden. DAGA Stuttgart, Fortschritte der Akustik 33/2:633-534
- Lindau A, Hohn T, Weinzierl S (2007) Binaural resympthesis for comparative studies of acoustical environments. 122nd AES Convention, Wien, Preprint 7032
- Lindau A, Maempel HJ, Weinzierl S (2008) Minimum BRIR grid resolution for dynamic binaural synthesis. Proc. of the Acoustics '08 Paris, S. 3851-3856
- Menzel D, Wittek H, Fastl H, Theile G (2006) Binaurale Raumsynthese mittels Wellenfeldsynthese. Realisierung und Evaluierung. DAGA Braunschweig, Fortschritte der Akustik 32/1:255-256
- Moldrzyk C, Lentz T, Weinzierl S (2005) Perzeptive Evaluation binauraler Auralisationen. DAGA München, Fortschritte der Akustik 31/2:545-546
- Møller H (1992) Fundamentals of binaural technology. Applied Acoustics 36:171-218
- Møller H, Hammershøi D, Jensen CB, Sørensen MF (1995a) Transfer characteristics of headphones measured on human ears. J. Audio Eng. Soc. 43/4:203-217
- Møller H, Jensen CB, Hammershøi D, Sørensen MF (1995b) Design criteria for headphones. J. Audio Eng. Soc. 43/4:218-232
- Møller H, Sørensen MF, Hammershøi D, Jensen CB (1995c) Head related transfer functions of human subjects. J. Audio Eng. Soc. 43/5:300-321
- Moore BCJ (1995) Frequency Analysis and Masking. In: Moore BCJ (Hrsg.) Hearing. Academic Press, San Diego/New York, S. 161-205 (=Handbook of Perception and Cognition. Bd. 2, 2. Auflage)
- Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths ans excitation patterns. J. Acoust. Soc. Am. 74/3:750-753
- Moore BCJ, Oldfield SR, Dooley GJ (1989) Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. J. Acoust. Soc. Am. 85/2:820-836
- Mourjopoulos JN (1994) Digital equalization of room acoustics. J. Audio Eng. Soc. 42/11:884-900
- Mourjopoulos JN, Clarkson PM, Hammond JK (1982) A comparative study of leas-squares and homomorphic techniques for the inversion of mixed phase signals. Proc. of the 1982 IEEE Int. Conference on ASSP, Paris
- Mourjopoulos JN, Paraskevas MA (1991) Pole and zeros modelling of room transfer functions. Journal of Sound and Vibration 146/2:281-302
- Müller S (1999) Digitale Signalverarbeitung für Lautsprecher. Dissertation, RWTH Aachen
- Müller S (2008) Messtechnik. In: Weinzierl S (Hrsg.) Handbuch der Audiotechnik. Springer, Berlin/Heidelberg, S. 1087-1169

- Müller S, Massarani P (2001) Transfer-function measurement with sweeps. J. Audio Eng. Soc. 49/6:443-471
- Neely ST, Allen JB (1979) Invertibility of a room impulse response. J. Acoust. Soc. Am. 66/1:165:169
- Norcross SG, Soulodre AG, Lavoie MC (2002) Evaluation of inverse filtering techniques for room/speaker equalization. 113th AES Convention, Los Angeles, Preprint 5662
- Norcross SG, Soulodre AG, Lavoie MC (2003a) Subjective effects of regularization on inverse filtering. 114th AES Convention, Amsterdam, Preprint 5848
- Norcross SG, Soulodre AG, Lavoie MC (2003b) Further investigations of inverse filtering. 115th AES Convention, New York, Preprint 5923
- Norcross SG, Soulodre AG, Lavoie MC (2004a) Distortion audibility in inverse filtering. 117th AES Convention, San Francisco, Preprint 6311
- Norcross SG, Soulodre AG, Lavoie MC (2004b) Subjective investigations of inverse filtering. J. Audio Eng. Soc. 52/10:1003-1028
- Norcross SG, Bouchard M, Soulodre AG (2005) Adaptive strategies for inverse filtering. 119th AES Convention, New York, Preprint 6563
- Norcross SG, Bouchard M, Soulodre AG (2006) Inverse filtering design using a minimal-phase target function from regularization. 121st AES Convention, San Francisco, Preprint 6929
- Opppenheim AV, Schafer RW (1992) Zeitdiskrete Signalverarbeitung. Oldenbourg, München
- Pohlmann KC (2005) Principles of Digital Audio. 5. Auflage, McGraw-Hill, New York/Chicago
- Preis D (1982) Phase distortion and phase equalization in audio signal processing A tutorial review. J. Audio Eng. Soc. 30/11:774-794
- Rubak P, Johansen LG (2000) Design and evaluation of digital filters applied to loudspeaker/room equalization. 108th AES Convention, Paris, Preprint 5172
- Slaney M (1993) An efficient implementation of the Patterson-Holdsworth auditory filter bank. Apple Computer Technical Report #35, Perception Group – Advanced Technology Group, Apple Computer Inc.
- Slaney M (1998) Auditory toolbox. Version 2. Technical Report #1998-010, Interval Research Corporation
- Slavik KM, Weinzierl S (2008) Wiedergabeverfahren. In: Weinzierl S (Hrsg.) Handbuch der Audiotechnik. Springer, Berlin/Heidelberg, S. 609-685
- Salomons AM (1995) Coloration and Binaural Decoloration of Sound Due to Reflections. Dissertation, Delft University of Technology
- Stolla J (2004) Abbild und Autonomie. Zur Klanggestaltung bei Aufnahmen klassischer Musik 1950-1994. Tectum Verlag, Marburg (Zugl.: TU Berlin, Univ. Diss. 2002)
- Theile G (1981) Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer. Rundfunktechnische Mitteilungen 25/4:155-170
- Theile G (1986) On the standardization of the frequency response of high-quality studio headphonse. J. Audio Eng. Soc. 34/12:956-969
- The Mathworks Inc. (2008) Signal Processing Toolbox 6. User's Guide. Revised for Version 6.9 (Release 2008a), nur online verfügbar: http://www.mathworks.com/access/helpdesk/help/pdf_doc/signal/signal_tb.pdf
- Toole FE, Olive SE (1988) The Modification of Timbre by Resonances. Perception and Measurement. J. Audio Eng. Soc. 36/3:122–142

Weinzierl S (2008) Aufnahmeverfahren. In: Weinzierl S (Hrsg.) Handbuch der Audiotechnik. Springer, Berlin/Heidelberg, S. 551-607

Werner M (2008) Digitale Signalverarbeitung mit Matlab-Praktikum. Zustandsraumdarstellung, Lattice-Strukturen, Prädiktion und adaptive Filter. Vieweg, Wiesbaden

Zwicker E, Fastl H (1999) Psychoacoustics. Facts and Models. 2. Auflage, Springer, Berlin/Heidelberg

Normen und Standards

DIN EN 60268-4:2004-07	Elektroakustische Geräte Teil 4: Mikrofone. Deutsches Institut für Nor- mung, Berlin
DIN EN 60268-7:1996-10	Elektroakustische Geräte Teil 7: Kopfhörer und Ohrhörer. Deutsches Institut für Normung, Berlin
DIN EN 61260:2003-03	Bandfilter für Oktaven und Bruchteile von Oktaven. Deutsches Institut für Normung, Berlin
Rec. ITU-R BS.1116-1	Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Broad- casting Union, Genf, 1994-1997
Rec. ITU-R BS.1248-1	General methods for the subjective assessment of sound quality. Inter- national Broadcasting Union, Genf, 1997-2003
Rec. ITU-R BS.1770	Algorithms to measure audio programme loudness and true-peak audio level. International Broadcasting Union, Genf, 2006

A QUELLCODES

A.1 Erzeugung der Entzerrungsfilter

```
%
      Calculate equalization filters with different methods
                                                             %
%-----%
clear all; close all; clc;
global setup
%-----
h = average('II');
                      % average from 10 measurements
\% % minimum phase version of h
% for i = 1:setup.channels
% [r, h(:,i)] = rceps(h(:,i));
% end
% normalizing h
h = h./max(max(abs(h)));
filterlen = 2^11;
setup.fs = 44100;
setup.channels = size(h,2);
setup.b = firfilt(2^11,50,21000,60); % target bandpass
if filterlen <= 4096
  1 = 4096;
elseif filterlen > 4096
  l = filterlen;
end
hw = pre_windowing(h,l); % window h
% % compare and squeeze
% hwcs = wavread('II_CS3.WAV');
% hwcs = hwcs./max(max(abs(hwcs)));
% regularization filter
reg = regfilter(hw,3);
% reg = load('reg_oct4096_II'); reg = reg.h_c;
%calculate h_c
tic
% h_c = allpole(hw,filterlen,filterlen/2,512);
h_c = ffd(hw,filterlen,filterlen/2,reg,.4);
% h_c = ffd(hwcs,filterlen,filterlen/2);
% h_c = ls(hw,filterlen,filterlen/2,reg,.4);
% h_c = smoothing(h,filterlen,filterlen/2,'oct','eqcmp',1);
toc
h_c = post_windowing(h_c,filterlen); % window h_c
```

A.1.1 Pre- und post-processing

A.1.1.1 Mittelung

```
\% Averaging of 10 (complex) transfer functions - inputs are wavfiles.
% A(n) = A(n-1)*(1-1/n) + H_new*(1/n)
% [Mueller (2008), Messtechnik. In: Handbuch der Audiotechnik (Weinzierl (Hrsg.))]
% syntax: h_av = average(hph)
%
% hph
            - string; name of headphone-file
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_av = average(hph)
cd(strcat('wavfiles/',hph))
h(:,:,1) = wavread('44_STE_0.WAV');
h(:,:,2) = wavread('44_STE_1.WAV');
h(:,:,3) = wavread('44_STE_2.WAV');
h(:,:,4) = wavread('44_STE_3.WAV');
h(:,:,5) = wavread('44_STE_4.WAV');
h(:,:,6) = wavread('44_STE_5.WAV');
h(:,:,7) = wavread('44_STE_6.WAV');
h(:,:,8) = wavread('44_STE_7.WAV');
h(:,:,9) = wavread('44_STE_8.WAV');
h(:,:,10) = wavread('44_STE_9.WAV');
cd ..; cd ..
N = length(h);
channels = size(h,2);
H = fft(h);
% complex averaging
A = zeros(N, channels, 11);
H_av = zeros(N, channels);
h_av = H_av;
for i = 1:channels
   for n = 1:10
        A(:,i,n+1) = A(:,i,n)*(1-1/n)+H(:,i,n)/n;
    end
    H_av(:,i) = A(:,i,11);
    h_av(:,i) = ifft(H_av(:,i));
end
```

A.1.1.2 Zielbandbpass

```
\% Creates a linear phase bandpass with the matlab function fir1 and a
% kaiser window
%
% syntax: b = firfilt(n,fL,fH,StopdB)
%
% n
            - filterorder, should be even
            - lower cut-off frequency
% fL
% fH
            - higher cut-off frequency
% StopdB
            - stopband attenuation in dB
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function b = firfilt(n,fL,fH,StopdB)
error(nargchk(3, 4, nargin))
global setup
beta = 0.1102*(StopdB-8.7);
w = kaiser(n+1,beta);
b = fir1(n,[fL/(setup.fs/2) fH/(setup.fs/2)], w);
b = DCext(b',w);
                   % dc-extinction
```

A.1.1.3 Vorfensterung

```
% Windows the input impulse response [h] with a right-sided window,
% based on a blackmanharris window, equal to 1 for half the window length.
% Only for minimum phase impulse responses!
%
% syntax: h_w = pre_windowing(h,len)
% h
            - multichannel impulse response to be windowed
% len
            - length of window
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_w = pre_windowing(h,len)
channels = size(h,2);
% windowing
w = blackmanharris(len);
w(1:len/2) = 1;
h_w = zeros(len,channels);
for i = 1:channels
   h_w(:,i) = h(1:len,i).*w;
end
h_w = DCext(h_w,w);
                        % dc-extinction
```

A.1.1.4 Nachträgliche Fensterung

```
\% Windows the input impulse respnse [h] with a tukey window (r = 0.75), by
% placing the window around the maximum of the impulse response
%
% syntax: h_w = post_windowing(h,len)
% h
            - multichannel impulse response to be windowed
% len
            - length of window, must be same as h
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_w = post_windowing(h,len)
channels = size(h,2);
\% windowing (placing window around maximum of h)
[m,ind] = max(abs(h(:,1)));
w1 = tukeywin(ind*2,.75);
w2 = tukeywin((len-ind)*2,.75);
w = [w1(1:length(w1)/2); w2(length(w2)/2+1:end)];
h_w = zeros(len,channels);
for i = 1:channels
   h_w(:,i) = h(1:len,i).*w;
end
h_w = DCext(h_w,w);
                        % dc-extinction
```

A.1.1.5 Gleichanteil-Korrektur

```
% Applies dc-correcting colva window on input
% [Müller S (1999) Digitale Signalverarbeitung für Lautsprecher]
%
% syntax: h = DCext(h,w)
%
% h
            - multichannel impulse response to be corrected
% w
            - window function to be used
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h = DCext(h, w)
N = length(h);
channels = size(h,2);
w_sum = sum(w);
h_sum = sum(h);
```

```
div = h_sum./w_sum;
mult = zeros(N,channels);
for i = 1:channels
    mult(:,i) = div(i).*w;
    h(:,i) = h(:,i) - mult(:,i);
end
```

A.1.2 Methoden

A.1.2.1 Least Squares

```
\% Returns the LS-optimal compensation filter [h_c] to an input file [h]
\% as solution of H*h_c = d(n-delay) when H is convolution matrix
\% and d is a target function delayed by [delay] number of samples. If the
% paramter for [phase] is chosen to be 'min', the minimal phase transfer
% function of [h] is calculated before the inversion, so that the minimal
% phase inverse is returned.
\% [reg] determines a filter used for frequency dependant regulariation and
% can be scaled by [gain].
%
\% [Kirkeby O, Nelson PA (1999) Digital Filter Design for Inversion Problems
% in Sound Reproduction]
%
% syntax: h_c = ls(h,filterlen,delay,reg,gain)
%
% h
           - multichannel impulse response to be inverted
\% filterlen – length of h_c
% delay
           - modelling delay in h_c
           - impulse response of a filter regularizing the optimzation effort
% reg
%
            spectrally
% gain
           - optional [0;1] scaling the effort while using regularization,
%
             default is [1]
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_c = ls(h,filterlen,delay,reg,gain)
global setup
error(nargchk(3, 5, nargin))
if nargin == 3
    method = 'no_regularization';
end
if nargin == 4
    gain = 1;
    method = 'with_regularization';
end
if nargin == 5
   method = 'with_regularization';
end
%------
if size(h,2) > size(h,1)
   h = h';
end
N = length(h);
if delay > N
    error('Delay must be smaller than lengtht of h.')
end
\% target function must have same length as h and maximum at delay
d = [setup.b; zeros(N+filterlen-1-length(setup.b),1)];
[m,i] = max(d):
d = circshift(d,delay+1-i);
%-----
```

```
%inversion
switch method
    case 'no regularization'
        display('ls: Inverting without regularization')
        h_c = zeros(filterlen,setup.channels);
        for i = 1:setup.channels
              H = convmtx(h(:,i),filterlen);
              h_c(:,i) = (H.'*H) \setminus (H.'*d);
        end
    case 'with_regularization'
        display('ls: Inverting with regularization')
        if size(reg,2) > size(reg,1)
            reg = reg';
        end
        if size(reg,2) < setup.channels</pre>
            reg = repmat(reg,1,setup.channels);
        end
        h_c = zeros(filterlen,setup.channels);
        for i = 1:setup.channels
              H = convmtx(h(:,i),filterlen);
              B = convmtx(reg(:,i),filterlen);
              h_c(:,i) = (H.'*H + gain*B.'*B)\(H.'*d);
        end
```

end

A.1.2.2 Fast Frequency Deconvolution

```
\% Returns the compensation filter [h_c] to an input file [h]
\% as solution of H*H_c = D when H and H_c are the frequency responses of h
\% and h_c. D is the frequency response of a target function delayed by
\% [delay] number of samples. If the paramter for [phase] is chosen to be
\% 'min', the minimal phase transfer function of [h] is calculated before
% the inversion, so that the minimal phase inverse is returned.
\% [reg] determines a filter used for frequency dependant regulariation and
% can be scaled by [gain].
%
% [Kirkeby O et al. (1998) Fast Deconvolution of Multichannel Systems Using
% Regularization; Leckschat D (1992) Verbesserung der Wiedergabequalität von
% Lautsprechern mit Hilfe von Digitalfiltern]
%
% syntax: h_c = ffd(h,filterlen,delay,reg,gain)
%
% h
            - multichannel impulse response to be inverted
% filterlen - length of h_c
% delay
           - modelling delay in h_c
% reg
            - impulse response of a filter regularizing the optimzation effort
              spectrally (must have the same length as h)
%
% gain
            - optional [0;1] scaling the effort while using regularization,
%
              default is [1]
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_c = ffd(h,filterlen,delay,reg,gain)
global setup
error(nargchk(3, 5, nargin))
if nargin == 3
   method = 'no_regularization';
end
if (nargin > 3) && (length(reg) ~= length(h))
    error('Regularization filter must have same length as h')
end
if (nargin == 4) && (length(reg) == length(h))
    gain = 1;
    method = 'with_regularization';
```

```
end
if (nargin == 5) && (length(reg) == length(h))
    method = 'with_regularization';
end
%-----
if size(h,2) > size(h,1)
   h = h';
end
N = length(h);
if delay > N
    error('Delay must be smaller than length of h')
end
H = fft(h);
\% zeropad bandpass coeffitients to the same length as h and circshift their
% maximum to the index specified by delay
setup.b = [setup.b; zeros(N-length(setup.b),1)];
[m,i] = max(setup.b);
setup.b = circshift(setup.b,delay+1-i);
% target bandpass
band = freqz(setup.b,1,N,'whole');
H_c = zeros(N,setup.channels);
%-----
% inversion
switch method
    case 'no_regularization'
       display('ffd: Inverting without regularization')
        for k = 1:setup.channels
           \% restriction of dynamic to 60 dB
           q = abs(20*log10(abs(H(:,k)))) > max(20*log10(abs(H(:,k))))+60;
           ind1 = find(q == 1);
           ind0 = find(q == 0);
           H_cabs = 10<sup>((max(20*log10(abs(H(:,k))))+60)/20);</sup>
           H_c(ind1,k) = (band(ind1).*H_cabs).*exp(sqrt(-1)*angle(band(ind1)./H(ind1,k)));
           H_c(ind0,k) = band(ind0)./H(ind0,k);
        end
    case 'with_regularization'
        display('ffd: Inverting with regularization')
        if size(reg,2) > size(reg,1)
           reg = reg';
        end
       if size(reg,2) < setup.channels</pre>
           reg = repmat(reg,1,setup.channels);
        end
       B = fft(reg);
        for k = 1:setup.channels
           % restriction of dynamic to 60 dB
           q = abs(20*log10(abs(H(:,k)))) > max(20*log10(abs(H(:,k))))+60;
           ind1 = find(q == 1);
           ind0 = find(q == 0);
           H_cabs = 10<sup>((max(20*log10(abs(H(:,k))))+60)/20);</sup>
           H_c(ind1,k) = (band(ind1).*H_cabs).*exp(sqrt(-1)*angle(band(ind1).*conj(H(ind1,k))...
            ./(abs(H(ind1,k)).^2 + gain*(abs(B(ind1,k)).^2))));
           H_c(ind0,k) = band(ind0).*conj(H(ind0,k))./(abs(H(ind0,k)).^2 + gain*(abs(B(ind0,k)).^2));
        end
```

```
end
```

%-----

% calculating h_c
h_c = real(ifft(H_c));
```
% truncation of filter impulse response
h_c = h_c(1:filterlen,:);
```

Regularisierungs-Hochpassfilter

```
% A high-pass filter [reg] is returned, intended to serve as a regularization
% filter for lms- or ffd-inversion of the input impulse response [h].
\% The stopband of the high-pass filter is set to be between 0 and 2000 Hz
\% with a stopband attenuation of -20 dB. The cut-off frequency is
\% calculated depending on the input impulse response and the parameter
\% [qfact]. The high-pass filter is intended to have its passband
\% where the severe dips of the transfer function start.
\% Dips are defined by a minimum Q-factor [qfact] and as being at least
\% 3 dB below the mean value of the transfer function (between 20 and fs/2).
\% The cut-off frequency of the high-pass filter is the frequency of the
% first dip (i.e. with the lowest frequency) meeting these conditions. If
% there is no dip that meets the 3dB-condition, this condition is dropped
% but a warning is displayed.
%
% syntax: reg = regfilter(h,qfact)
%
% h
            - multichannel impulse response for which regularization
             high-pass filter is to be calculated
%
% qfact
           - minimum Q-factor of the dip, whose frequency shall serve as
%
             cut-off frequency of the high-pass filter
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function reg = regfilter(h,qfact)
global setup
%-----
if size(h,2) > size(h,1)
   h = h';
end
N = length(h);
H = fft(h);
%------
d_f = setup.fs/N;
                               % frequency increment
ind_low = round(20/d_f);
                               % sample at 20 Hz
ind_high = round((setup.fs/2)/d_f); % sample at fs/2
m = mean(20*log10(abs(H(ind_low:ind_high,:)))); % mean value of H between [20, fs/2] Hz
reg = zeros(N,setup.channels);
for q = 1:setup.channels
    index = 0;
    ind_m = 20*log10(abs(H(:,q))) < m(q);
                                               % H smaller than its mean value
    zer = ind_low;
    while zer < N/2+1
       one = find(ind_m(zer:end),1)+zer-1;
                                               % sample where H is < mean
       zer = find(ind_m(one:end)==0,1)+one-1; % sample where H is > mean
        cond1 = sqrt(zer*one)/(zer-one) >= qfact; % dip must have higher Q-factor than qfact
       cond2 = zer-one \ge 10;
                                              % dip must be wider 10 samples
        cond3 = min(20*log10(abs(H(one:zer,q)))) <= (m(q)-3); % dip must be at least 3 dB below mean</pre>
        if cond1 == 1 && cond2 ==1 && cond3 == 1
           index = one;
                                               % sample from which Fpass is calculated
           break
        end
    end
    \% if Fpass cannot be calculated with all three conditions being true,
    % cond3 is dropped
    if index == 0
       zer = ind_low;
       while zer < N/2+1
           one = find(ind_m(zer:end),1)+zer-1;
                                                   % sample where H is < mean
           zer = find(ind_m(one:end)==0,1)+one-1; % sample where H is > mean
           cond1 = sqrt(zer*one)/(zer-one) >= qfact; % dip must have higher Q-factor than qfact
```

end

A.1.2.3 Glättung

b = fir2(50, f, m);

reg(:,q) = [b'; zeros(N-length(b),1)];

```
\% Calculates the smoothed transfer function to an input [h] and inverts it
% with the frequency domain inversion method ffd. Returns the compensation
% filter [h_c] and the impulse response of the smoothed transfer function
% [h_smooth].
% The smoothed function is calculated by convolving the transfer function
% of h with a window funktion W. The width of the window increases with
\% increasing frequency, [band] determines the frequency scale used (fraction
\% of octave, bark or ERB). The convolution is either conducted with the
\% magnitude transfer function of h or the complex transfer function. In the
\% latter case, the window is also applied to the time domain, increasing in
\% width with decreasing frequency. In the case of equivalent complex
\% smoothing, h_smooth is a comibination of the amplitude smoothed magnitude
% and the complex smoothed phase.
%
%
 [Hatziantoniou PD, Mourjopoulos JN (2000) Generalized Fractional-Octave
% Smoothing of Audio and Acoustic Responses]
%
% syntax: smoothing(h,filterlen,delay,band,mode,frac,b)
%
% h
            - multichannel impulse response to be inverted
% filterlen - length of compensation filter h_c
            - 'oct' for octave, 'bark' for bark or 'ERB'
% band
% mode
            - 'cpm' for complex smoothing (phase and amplitude) or 'amp'
%
              for smoothing amplitdue only
% frac
            - if scale=octave: fraction of octave (e.g. 1/3)
%Ъ
            - coefficient for smoothing function
              W = b - (1-b)*\cos(2*pi*k/N), for k = [0,filterlen-1]. The
%
%
              default value of 0.5 produces a Hann window function
%
% see also: ffd
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function [h_c,h_smooth] = smoothing(h,filterlen,delay,band,mode,frac,b)
global setup
error(nargchk(5, 7, nargin))
if (strcmp(band,'ERB') == 1) || (strcmp(band,'bark') == 1)
    if nargin == 5
       b = 0.5;
    end
    if nargin == 6
       b = frac;
    end
end
if strcmp(band, 'oct') == 1
    if nargin == 5
    error('Please enter fraction of octave.')
    end
```

```
if nargin == 6
       b = 0.5;
    end
end
%------
if size(h,2) > size(h,1)
   h = h';
end
N = length(h);
if delav > N
    error('Delay must be smaller than length of h')
end
% n-point fft of h, with n=filterlen
h = pre_windowing(h,filterlen);
H = fft(h,filterlen);
if strcmp(mode,'amp') == 1
   H2 = abs(H).^{2};
    display('Amplitude smoothing')
end
if strcmp(mode,'cmp') == 1
   H_r = real(H);
    H_im = imag(H);
    display('Complex smoothing')
end
if strcmp(mode,'eqcmp') == 1
   H_r = real(H);
   H_im = imag(H);
    H2 = abs(H).^2;
   display('Equivalent complex smoothing')
end
%-----
%smoothing
\% generate vector with bandwidths according to the selected scale
d_f = setup.fs/filterlen;
switch band
    case 'oct'
       display(strcat('bandwidth: ',num2str(frac),' octave'))
       bandwidth = zeros(filterlen/2+1,1);
        for k = 1:filterlen/2+1
           f_{up} = 10^{(3/10)}(0.5*frac);
           f_{low} = 10^{(-3/10)^{(0.5*frac)}};
           bandwidth(k)= k*d_f*(f_up - f_low);
        end
    case 'bark'
       display('bandwidth: critical band')
        bandwidth = zeros(filterlen/2+1,1);
       for k = 1:filterlen/2+1
           bandwidth(k) = 25 + 75*(1 + 1.4*(k*d_f/1000)^2)^0.69;
       \operatorname{end}
    case 'ERB'
        display('bandwidth: ERB')
       bandwidth = zeros(filterlen/2+1,1);
        for k = 1:filterlen/2+1
           bandwidth(k) = 24.7*(4.37*(k*d_f/1000) + 1);
        end
end
m = zeros(filterlen/2+1,1);
for k = 1:filterlen/2+1
   m(k) = 0.5*bandwidth(k)/d_f;
end
```

```
m = ceil(m);
clear bandwidth
% generate windowfunction W
W = zeros(length(m),filterlen);
for j = 1:length(m)
    for k = 0:m(j)
       W(j,k+1) = (b - (1-b)*cos((pi/m(j))*(m(j)-k))) ./ (2*b*(m(j)+1) - 1);
    end
    for k = (filterlen-m(j)):(filterlen-1)
        W(j,k+1) = (b - (1-b)*cos((pi/m(j))*((filterlen-m(j))-k))) ./ (2*b*(m(j)+1) - 1);
    end
end
\% generate circular convolution matrix H_circ
if strcmp(mode,'cmp') == 1 || strcmp(mode,'eqcmp') == 1
    H_circ_r = zeros(filterlen,filterlen,setup.channels);
    H_circ_im = H_circ_r;
    for j = 1:setup.channels
        for 1 = 0:filterlen-1
            H_circ_r(l+1,:,j) = circshift(H_r(:,j),l);
            H_circ_im(l+1,:,j) = circshift(H_im(:,j),l);
        end
    end
end
if strcmp(mode,'amp') == 1 || strcmp(mode,'eqcmp') == 1
   H_circ2 = zeros(filterlen,filterlen,setup.channels);
    for j = 1:setup.channels
        for l = 0:filterlen-1
            H_circ2(l+1,:,j) = circshift(H2(:,j),l);
        end
    end
end
clear H_r H_im H_r2 H_im2
\% circular convolution of W and H_circ
if strcmp(mode,'cmp') == 1 || strcmp(mode,'eqcmp') == 1
    H_smooth_circ_r = zeros(filterlen/2+1,filterlen,setup.channels);
    H_smooth_circ_im = H_smooth_circ_r;
    for j =1:setup.channels
        H_smooth_circ_r(:,:,j) = W * H_circ_r(:,:,j);
       H_smooth_circ_im(:,:,j) = W * H_circ_im(:,:,j);
    end
end
if strcmp(mode,'amp') == 1 || strcmp(mode,'eqcmp') == 1
H_smooth_circ2 = zeros(filterlen/2+1,filterlen,setup.channels);
    for j =1:setup.channels
        H_smooth_circ2(:,:,j) = W * H_circ2(:,:,j);
    end
end
clear H_circ_r H_circ_im H_circ_r2 H_circ_im2 W
\% derive the smoothed function from <code>H_smooth_circ</code>
U = zeros(filterlen,filterlen);
v = zeros(1,filterlen/2+1);
if strcmp(mode,'cmp') == 1 || strcmp(mode,'eqcmp') == 1
   H_smooth_r = zeros(filterlen, setup.channels);
    H_smooth_im = H_smooth_r;
    for l = 1:setup.channels
        for k = 1:filterlen/2+1;
            U(k,k) = 1;
            v(m(k)) = 1;
            H_smooth_r(k,l) = sum(v*H_smooth_circ_r(:,:,l)*U);
            H_smooth_im(k,1) = sum(v*H_smooth_circ_im(:,:,1)*U);
            U(k,k) = 0;
            v(m(k)) = 0;
        end
    end
end
if strcmp(mode, 'amp') == 1 || strcmp(mode, 'eqcmp') == 1
   H_smooth2 = zeros(filterlen, setup.channels);
    for l = 1:setup.channels
```

```
for k = 1:filterlen/2+1;
            U(k,k) = 1;
            v(m(k)) = 1;
            H_smooth2(k,1) = sum(v*H_smooth_circ2(:,:,1)*U);
            U(k,k) = 0;
            v(m(k)) = 0;
        end
    end
end
clear U v H_smooth_circ_r H_smooth_circ_im
if strcmp(mode,'cmp') == 1 || strcmp(mode,'eqcmp') == 1
    for l = 1:setup.channels
        for k = filterlen/2+2:filterlen
           H_smooth_r(k,1) = H_smooth_r(filterlen-k+2,1);
            H_smooth_im(k,1) = -H_smooth_im(filterlen-k+2,1);
        end
    end
    H_smoothc = H_smooth_r + sqrt(-1).*H_smooth_im;
end
if strcmp(mode, 'amp') == 1 || strcmp(mode, 'eqcmp') == 1
    for l = 1:setup.channels
        for k = filterlen/2+2:filterlen
            H_smooth2(k,1) = H_smooth2(filterlen-k+2,1);
        end
    end
end
switch mode
    case 'amp'
       H_smooth = sqrt(H_smooth2).*exp(sqrt(-1).*angle(H));
    case 'cmp'
       H_smooth = H_smoothc;
    case 'eqcmp'
       H_smooth = zeros(filterlen,setup.channels);
        for k = 1:setup.channels
            H_smooth(:,k) = sqrt(H_smooth2(:,k)).*exp(sqrt(-1).*angle(H_smoothc(:,k)));
        end
end
%-----
h_smooth = real(ifft(H_smooth));
%inversion with ffd-method
h_c = ffd(h_smooth,filterlen,delay);
```

A.1.2.4 All-Pol-Modell

```
\% Calculates all-pole model of the input impulse response [h] with the
% matlab funcion aryule and inverts this model with the frequency domain
\% inversion method ffd. Returns the compensation filter [h_c].
% [Mourjopoulos J, Paraskevas MA (1991) Pole and Zero Modeling of Room
% Transfer Functions; Hayes MH (1996) Statistical Digital Signal Processing
% and Modelling]
%
% syntax: h_c = allpole(h,filterlen,delay,p)
%
% h
           - multichannel impulse response to be inverted
% filterlen - length of h_c
% delay
           - modelling delay in h_c
%р
            - order of all-pole model
%
\% see also: aryule, ffd
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function h_c = allpole(h,filterlen,delay,p)
global setup
```

%-----if size(h,2) > size(h,1)
 h = h';
end
H = fft(h);
N = length(h);
%-----% all-pole model
H_pole = zeros(N,setup.channels);
for i = 1:setup.channels;
 [a,e] = aryule(h(:,i),p);
 H_pole(:,i) = freqz(e,a,N,'whole');
 G = abs(H(round(1000/(setup.fs/N)),i))/abs(H_pole(round(1000/(setup.fs/N)),i)); % scaling factor
 H_pole(:,i) = H_pole(:,i)*G;
end

```
%-----
```

h_pole = ifft(H_pole);

%inversion with ffd-method
h_c = ffd(h_pole,filterlen,delay);

A.2 Evaluation

A.2.1 Objektive Evaluation

A.2.1.1 Berechnung der Fehlermaße

```
\% Returns the time domain and frequency domain error between an (equalized)
% impulse response [h_eq] and a target function [d]. Must have same length!
\% The time domain error is calculated as the sum of the squared difference
% between impulse response and target function:
% err.time = sum((h_eq - d).^2);
\% For calculating frequency domain error (err.freq), the power spectra of h_eq
% and d are filtered with an ERB filterbank, the number of bands being
\% calculated with the following equation:
% n = 21.4*log10(0.004367*f_up+1)-21.4*log10(0.004367*f_low+1) (f_up =
% upper edge frequency, f_low = lower edge frequency)
\% The algorithm integrates over each band within the frequency range of the
\% target function and then calculates the magnitude difference in dB
\% between H_eq and the target function for each integration result.
% The total frequency error is the mean of the modulus of err.freq:
% err.tot = mean(abs(err.freq))
%
\% [Solomons AM (1995) Coloration and binaural decoloration of sound due
% to reflections; Moore BCJ (1995) Hearing]
%
% syntax: [err,f_c,n] = errcalc(h_eq,d)
%
% h_eq
            - multichannel impulse response to be compared to target
%
              function
% d
            - target function
%
% see also: MakeERBFilters, ERBFilterBank
%
% Zora Schaerer, zora.schaerer@tu-berlin.de, 05.08.2008
function [err,f_c,n] = errcalc(h_eq,d)
global setup
Q = size(h_eq, 1);
s = size(h_eq,3);
% ----- 1. time domain error -----%
err.time = zeros(setup.channels,1);
```

```
for i = 1:setup.channels
    for j = 1:s
        err.time(i,j) = mean((h_eq(:,i,j) - d).^2);
    end
end
% ----- 2. frequency domain error -----%
% ERB Filters
n = round(21.4*log10(0.004367*21000+1)-21.4*log10(0.004367*50+1));
[fcoefs,f_c] = MakeERBFilters(setup.fs,n,50);
x = zeros(Q,1);
x(1) = 1;
x_ERB = ERBFilterBank(x',fcoefs);
C = abs(fft(x_ERB'));
i_50 = round(50/(setup.fs/Q));
i_21k = round(21000/(setup.fs/Q));
\% filter h_eq with filerbank and calculate error in each band
err.freq = zeros(n,setup.channels,s);
H_eq_ERB = zeros(Q,n,s);
D_{ERB} = zeros(Q,n);
for i = 1:setup.channels
    for j = 1:s
        for k = 1:n
            H_eq_ERB(:,k,j) = C(:,k).*(abs(fft(h_eq(:,i,j))).^2);
            D_ERB(:,k) = C(:,k).*(abs(fft(d)).^2);
            H_eq_c = sum(H_eq_ERB(i_50:i_21k,k,j));
            D_c = sum(D_ERB(i_50:i_21k,k));
            err.freq(k,i,j) = 10*log10(abs(H_eq_c)) - 10*log10(abs(D_c));
        end
        err.tot(i,j) = mean(abs(err.freq(:,i,j)));
    end
end
```

A.2.2 Perzeptive Evaluation

A.2.2.1 Lautheitskalibrierung

```
% Purpose: ITU-R BS 1770 sytle loudness calibration of n-channel audiosignals for listening tests
%
\% Calibrates all filters for same loudness. All channels are
% calibrated regarding the lowest loudness as reference. Steady state is assumed,
\% calibration will be erroneous for longer signals with deviation of dynamic over time.
clear all; clc; close all
filterlen = 2^{11};
fs = 44100;
r = wavread('2.wav');
% load delta peaks (for anchor stimulus)
u = load(strcat('d_',num2str(filterlen))); filt(:,:,5,1) = u.d;
filt(:,:,5,2) = filt(:,:,5,1);
% load all filters
% Stax II
s = load('lms_reg_hp_II.mat'); filt(:,:,1,1) = s.h_c;
s = load('lms_reg_oct_II.mat'); filt(:,:,2,1) = s.h_c;
s = load('ffd_reg_hp_II.mat'); filt(:,:,3,1) = s.h_c;
s = load('ffd_reg_oct_II.mat'); filt(:,:,4,1) = s.h_c;
s = load('smooth_amp_II.mat'); filt(:,:,6,1) = s.h_c;
s = load('smooth_eqcmp_II.mat'); filt(:,:,7,1) = s.h_c;
s = load('cs_II.mat'); filt(:,:,8,1) = s.h_c;
% Stax Pro
s = load('lms_reg_hp_pro.mat'); filt(:,:,1,2) = s.h_c;
s = load('lms_reg_oct_pro.mat'); filt(:,:,2,2) = s.h_c;
s = load('ffd_reg_hp_pro.mat'); filt(:,:,3,2) = s.h_c;
s = load('ffd_reg_oct_pro.mat'); filt(:,:,4,2) = s.h_c;
s = load('smooth_amp_pro.mat'); filt(:,:,6,2) = s.h_c;
s = load('smooth_eqcmp_pro.mat'); filt(:,:,7,2) = s.h_c;
s = load('cs_pro.mat'); filt(:,:,8,2) = s.h_c;
```

```
% -----
                     -----
for k = 1:size(filt,4)
   for j = 1:size(filt,3)
      for i = 1:2
         r_(:,i,j,k) = fftfilt(filt(:,i,j,k),r);
      end
   end
end
% --
     _____
                      -----
% RLB/Soulodre-filter
af=[1 -1.99004745483398 0.99007225036621];
bf=[1 -2 1];
% measure each channel's loudness
for k = 1:size(filt,4)
   for j = 1:size(filt,3)
      for i = 1:2
          loud(j,i,k)=mean(filter(bf,af,r_(:,i,j,k)).^2);
      end
   end
end
loud_norm = min(min(min(min(abs(loud)))));
save('loud_norm','loud_norm')
% -----
                                ------
for k = 1:size(filt,4)
   for j = 1:size(filt,3)
      for i = 1:2
          filt(:,i,j,k)=filt(:,i,j,k)*sqrt((loud_norm/loud(j,i,k)));
      end
   end
end
```

A.2.2.2 Ablaufsteuerung des Hörversuchs

Hauptfunktion

% Hörversuch zur Evaluation digitaler Entzerrungsfilter

```
function abchr
clc; close all; clear all
global setup
setup.server_pd = osc_new_address('130.149.50.180',7777);
setup.server_fwonder = osc_new_address('130.149.50.180',6666);
                   % number of pannels in each window
% number of windows
setup.ngroups = 8;
setup.nwindows = 2;
setup.nfilters = 8; % number of filters
setup.nstim = 2;
                      % number of stimuli
setup.slidemin = 1;
                      % mininum slider value
setup.slidemax = 5;
                      % maximum slider value
setup.nhph = 2;
setup.vallinks = setup.slidemax*ones(setup.ngroups,1);
setup.valrechts = setup.slidemax*ones(setup.ngroups,1);
setup.pressed_button = 0;
setup.col = [.8 .8 .8];
setup.bcol = [.9 .9 .9];
setup.VP = input('==> VP.nr: ');
%-----
setup.hphperm = randperm(setup.nhph); % random distribution of head phones
for q = 1:setup.nhph
   % current head phone
   setup.chph = setup.hphperm(q);
   if setup.chph == 1
       hphname = 'Stax II';
   elseif setup.chph == 2
```

```
hphname = 'Stax Pro';
end
% random distribution of all filters on the pannels
for l = 1:setup.nwindows
    setup.filterperm(1,:,setup.chph) = randperm(setup.nfilters);
    % random distribution of conditions for each pannel
    for p = 1:setup.ngroups
        setup.condperm(p,:,l,setup.chph) = randperm(2)-1; % 0 = unfiltered, 1 = filtered
    end
end
% random distribution of filterlengths on the windows
setup.stimperm(setup.chph,:) = randperm(2); % 1 = Rauschen; 2 = Gitarre
% start screen
f_ = figure('ToolBar', 'None', 'MenuBar', 'None', 'Name',...
    'H^rversuch zur Bewertung verschiedener Entzerrungsfilter', 'NumberTitle',...
    'off', 'Position', [340 200 600 400], 'Color', setup.col);
uicontrol(f_,'Style','text','String','Bitte setzen Sie den folgenden Kopfh^rer auf:','fontsize',14,...
    'BackGroundColor', setup.col, 'Position', [100 265 400 60]);
uicontrol(f_,'Style','text','String',num2str(setup.chph),'fontsize',14,...
    'BackGroundColor', setup.col, 'Position', [100 235 400 60]);
uicontrol(f_, 'Style', 'Push Button', 'Position', [225 175 150 50], 'String',...
    {'START'},'fontsize',16, 'Callback','global setup;setup.pressed_button = 1;');
while setup.pressed_button ~= 1
  pause (0.1)
end
set(f_,'Visible','Off')
setup.pressed_button = 0;
%-----
% variables for while-loop
setup.k = 1;
slinks = zeros(setup.nwindows,1);
srechts = slinks;
while setup.k <= setup.nwindows
    f = figure('Name',...
    'H^rversuch zur Bewertung verschiedener Entzerrungsfilter', 'NumberTitle',...
    'off','Position',[640-(220+setup.ngroups*130)/2 400-(230+setup.slidemax*50)/2 220+setup.ngroups...
    *130 285+setup.slidemax*50],'Color',setup.col);
   uicontrol(f,'Style','text','String','Der ver%.nderte Stimulus ist verglichen zur Referenz...',...
        'fontsize',12,'BackGroundColor', setup.col,'Position', [(210+setup.ngroups*130)/2-300 255+...
        setup.slidemax*50 600 20]);
    uicontrol(f,'Style','text','String',...
        Wenn Sie alle Reizpaare beurteilt haben, klicken Sie bitte auf "Weiter".',...
        'fontsize',12,'BackGroundColor',setup.col,'Position',[(210+setup.ngroups*130)/2-250 60 500 20]);
    % "Weiter" button
   hweiter = uicontrol(f,'Style','pushbutton','String','W e i t e r','fontsize',12,'Position',...
        [(220+setup.ngroups*130)/2-40 20 80 35]);
    % counting the windows
    uicontrol(f,'Style','text','String', strcat(num2str(setup.k),'/',...
        num2str(setup.nwindows)),'BackGroundColor',setup.col,...
        'Position',[10 10 30 20]);
   % labelling scales
    p = uipanel('Parent',f,'BackGroundColor',setup.col,'Units',...
         'pixels', 'Position', [2 90 110 141.5+setup.slidemax*50]);
    uicontrol(p,'Style','text','String','identisch','BackGroundColor',setup.col,...
         'Position', [60 127.7+(setup.slidemax-1)*52.5 45 20]);
    uicontrol(p,'Style','text','String','sehr unterschiedlich','BackGroundColor',setup.col,...
         'Position', [7 127.7+(setup.slidemin-1)*52.5 100 20]);
    p = uipanel('Parent',f,'BackGroundColor',setup.col,'Units',...
         'pixels','Position',[110+setup.ngroups*130 90 110 141.5+setup.slidemax*50]);
     uicontrol(p,'Style','text','String','identisch','BackGroundColor',setup.col,...
         'Position', [5 127.7+(setup.slidemax-1)*52.5 45 20]);
    uicontrol(p,'Style','text','String','sehr unterschiedlich','BackGroundColor',setup.col,...
```

```
XVI
```

end

```
'Position',[3 127.7+(setup.slidemin-1)*52.5 100 20]);
        % panels
        for j = 1:setup.ngroups
            % buttons
            panel = uipanel('Parent',f,'Title',num2str(j),'BackGroundColor',setup.col,'Units',...
                'pixels', 'Position', [110+(j-1)*130 90 130 150+setup.slidemax*50]);
            uicontrol(panel,'Style','pushbutton','BackGroundColor',[1 1 1],'String','Play',...
                'Tag',num2str(j),'Position',[10 80 50 30],...
                'Callback', 'global setup; setup.pressed_button = 10; play;');
            uicontrol(panel,'Style', 'pushbutton', 'BackGroundColor', [1 1 1], 'String', 'Play',...
                'Tag',num2str(j),'Position',[70 80 50 30],...
                'Callback', 'global setup; setup.pressed_button = 20; play;');
            uicontrol(panel,'Style','pushbutton','BackGroundColor', setup.bcol,'String','Ref',...
                'Position',[40 45 50 30],'Callback','global setup; setup.pressed_button = 30; play;');
            uicontrol(panel,'Style','pushbutton',...
                'Position', [40 10 50 30], 'String', 'Stop', 'Callback', ...
                'global setup; setup.pressed_button = 40; play;');
            % sliders
            setup.slinks(j) = uicontrol(panel,'Style','slider','Min',setup.slidemin,...
                'Max', setup.slidemax, 'SliderStep', [0.1 0.1]./(setup.slidemax-setup.slidemin),...
                'Value',setup.slidemax,'Position',[27 120 20 setup.slidemax*50],'Tag',num2str(j),...
                'Callback', 'slide_1;');
            setup.srechts(j) = uicontrol(panel,'Style','slider','Min',setup.slidemin,...
                'Max', setup.slidemax, 'SliderStep', [0.1 0.1]./(setup.slidemax-setup.slidemin),...
                'Value', setup.slidemax, 'Position', [82 120 20 setup.slidemax*50], 'Tag', num2str(j),...
                'Callback', 'slide_r;');
            % label sliders
            for i = setup.slidemax:-2:setup.slidemin
                uicontrol(panel,'Style','text','String','--','BackGroundColor',setup.col,...
                'Position', [17 137.7+(i-1)*52.5 10 10]);
uicontrol(panel,'Style','text','String','--','BackGroundColor', setup.col,...
                    'Position', [102 137.7+(i-1)*52.5 10 10]);
            end
        end
        %-
                      -----
        set(hweiter,'Callback','global setup; setup.pressed_button = 1;');
        while setup.pressed_button ~= 1
            pause (0.1)
        end
        setup.pressed_button = 0;
        set(f,'Visible','off')
        % get values from sliders
        for i = 1:setup.ngroups
            setup.val(setup.filterperm(setup.stimperm(setup.chph,setup.k),i,setup.chph),...
                setup.condperm(i,1,setup.k,setup.chph)+1,setup.stimperm(setup.chph,setup.k),...
                setup.chph) = get(setup.slinks(i),'Value');
            setup.val(setup.filterperm(setup.stimperm(setup.chph,setup.k),i,setup.chph),...
                setup.condperm(i,2,setup.k,setup.chph)+1,setup.stimperm(setup.chph,setup.k),...
                setup.chph) = get(setup.srechts(i),'Value');
        end
        setup.k = setup.k+1;
                                     % next window
    end
% save results
name = strcat('VP',num2str(setup.VP));
save(name, '-struct', 'setup','val','condperm','filterperm','stimperm','hphperm')
% end screen
f_ende = figure('ToolBar', 'None', 'MenuBar', 'None', 'Name',...
    'H^rversuch zur Bewertung verschiedener Entzerrungsfilter','NumberTitle',...
    'off', 'Position', [340 200 600 400], 'Color', setup.col);
uicontrol(f_ende,'Style','text','String','Vielen Dank f,r Ihre Zeit.',...
    'fontsize',14,'BackGroundColor',setup.col,'Position',[20 100 560 150]);
```

```
uicontrol(f_ende, 'Style', 'Push Button', 'Position', [225 30 150 50], 'String', {'ENDE'}, 'fontsize', 16, ...
```

```
'Callback','global setup;setup.pressed_button = 1;');
while setup.pressed_button ~= 1
    pause (0.1)
end
set(f_ende,'Visible','Off')
setup.pressed_button = 0;
osc_free_address(setup.server_pd);
osc_free_address(setup.server_fwonder);
```

end

play-Skript

global setup

```
cstim = setup.stimperm(setup.chph,setup.k);
if setup.pressed_button == 10 || setup.pressed_button == 20
   g = gcbo;
    c = get(g,'Tag');
    c = str2double(c);
    ccond = setup.condperm(c,:,setup.k,setup.chph);
    cfilt = setup.filterperm(setup.stimperm(setup.chph,setup.k),c,setup.chph);
    if cfilt == 1
       cpan = -4;
    elseif cfilt == 2
       cpan = -3;
    elseif cfilt == 3
       cpan = -2;
    elseif cfilt == 4
       cpan = -1;
    elseif cfilt == 5
       cpan = 0;
    elseif cfilt == 6
       cpan = 1;
    elseif cfilt == 7
       cpan = 2;
    elseif cfilt == 8
       cpan = 3;
    end
    if setup.chph == 1 && setup.stimperm(setup.chph,setup.k) == 1
       ctilt = 1;
    elseif setup.chph == 1 && setup.stimperm(setup.chph,setup.k) == 2
       ctilt = -1:
    elseif setup.chph == 2 && setup.stimperm(setup.chph,setup.k) == 1
       ctilt = 0;
    elseif setup.chph == 2 && setup.stimperm(setup.chph,setup.k) == 2
       ctilt = -2;
    end
end
if setup.pressed_button == 10
    if ccond(1) == 1
       % OSC senden an falter
       osc_data = struct('path', '/tilt' , 'data', {{int32(ctilt)}});
       osc_send(setup.server_fwonder, osc_data);
       pause(0.1)
       osc_data = struct('path', '/pan' , 'data', {{int32(cpan)}});
       osc_send(setup.server_fwonder, osc_data);
    end
    % OSC senden an pd
    osc_data = struct('path', '/file', 'data', {{int32(cstim)}});
    osc_send(setup.server_pd, osc_data);
   pause(0.05)
    osc_data = struct('path', '/play' , 'data', {{int32(ccond(1))}});
    osc_send(setup.server_pd, osc_data);
elseif setup.pressed_button == 20
    if ccond(2) == 1
       % OSC senden an falter
       osc_data = struct('path', '/tilt' , 'data', {{int32(ctilt)}});
```

```
osc_send(setup.server_fwonder, osc_data);
      pause(0.1)
       osc_data = struct('path', '/pan' , 'data', {{int32(cpan)}});
       osc_send(setup.server_fwonder, osc_data);
    end
    % OSC senden an pd
    osc_data = struct('path', '/file', 'data', {{int32(cstim)}});
    osc_send(setup.server_pd, osc_data);
    pause(0.05)
    osc_data = struct('path', '/play' , 'data', {{int32(ccond(2))}});
    osc_send(setup.server_pd, osc_data);
elseif setup.pressed_button == 30
                                        % reference
    % OSC senden an pd
    osc_data = struct('path', '/file' , 'data', {{int32(cstim)}});
    osc_send(setup.server_pd, osc_data);
   pause(0.05)
    osc_data = struct('path', '/play' , 'data', {{int32(0)}});
    osc_send(setup.server_pd, osc_data);
elseif setup.pressed_button == 40
                                        % stop
    % OSC senden an pd
    osc_data = struct('path', '/stop' , 'data', {{int32(1)}});
    osc_send(setup.server_pd, osc_data);
end
```

```
setup.pressed_button = 0;
```

B STATISTISCHE DATEN

Allgemein **B.1**

B.1.1 Deskriptiva

m1kh1c1

m2kh1c1

m3kh1c1

m4kh1c1

m5kh1c1

m6kh1c1

m7kh1c1

m1kh2c1

m2kh2c1

m3kh2c1

m4kh2c1

m5kh2c1

m6kh2c1

m7kh2c1

m1kh1c2

m2kh1c2

m3kh1c2

m4kh1c2

m5kh1c2

m6kh1c2

m7kh1c2

m1kh2c2

m2kh2c2

m3kh2c2

m4kh2c2

m5kh2c2

m6kh2c2

m7kh2c2

Valid N (listwise)

Descriptive Statistics Skewness stic Std. Error Kurtosis tic Std. Error Std. Ν Range Minimum Maximum Mean Statistic Statisti Statistic Statistic Statisti Statistic Statistic Statistic 3,37143 -3,50476 -,13333 1,6681315 ,92527006 ,456 -,897 26 -,124 ,90304028 ,456 26 3,25714 -3,40952 -,15238 -2,1316854 ,468 -,643 26 3,12381 -3,48571 -,36190 -1,9904754 88402415 ,362 ,456 -,936 26 26 3 23810 -3 69524 -,45714 -,55238 -2 2688646 94190560 .263 ,456 - 988 2,91429 ,122 -3,46667 -2,1333335 ,90738717 ,456 -1,091 26 3,48095 -4,00000 -,51905 -2,2446888 ,93693180 -,065 ,456 -,879 26 26 26 3,46667 3,44762 -4,00000 -,53333 -2,2525646 ,95357129 -,045 ,456 ,456 -,923 -3.61905-.17143 -1.7166673 83332619 -,634 ,291 2,57143 -3,63810 -1,06667 -2,4842488 ,70227961 297 456 -,639 26 3,14286 -3,37143 -,22857 1,6633700 77438804 -,528 ,456 ,006 26 26 26 2,70477 -3,63810 ,93333 -2,4175827 ,76268410 ,215 -,312 ,456 -1,003 456 -4 00000 -,236 -,847 2 81905 -1.18095-2 4399262 68162854 -2,4930404 ,078 ,456 2,76190 -4,00000 -1,23810 ,73191103 3,20000 -4,00000 -,80000 -2,6183150 , 85811988, ,224 , 456 -,889 26 26 26 26 3,12381 -3,12381 ,00000 -1,0717954 79189594 -,974 ,456 ,635 -,285 -,730 ,456 ,456 1,3245427 3 02857 -3.20000 -,731 -,424 -,17143 82339080

-,9855308

1,2974358

1,2655681

-1,1379119

-1,1996338

1,2300369

-1,6791215

1.1604396

1,5934073

1,5371796

-1.5992677

-1,5010981

64262277

,88154600

.87213195

72023492

,78226251

74843644

81694026

73429844

76485031

,86296118

83264948

87870354

-,591

,408

-.777

,024

-1,544

-,759 -1,577 ,077

-1,118

-,873

-,935

,456

,456 ,456

,456

,456

,456 ,456

,456

,456

,456

456

-,677

-,898

-.659

-1,304

4,775

1,160

4,398 -,010

,768

,040

.947

-,20952

-,09524

-,19048

-.36190

,13333

-,13333

-,22857

-,11429

-,13333

-,43810

-.68571

-,22857

-2,45714

-3,08571

-3,10476

-2.80000

-2,43810

-3,82857

-4,00000

-3.67619

-3,37143

-3,75238

-3,71429

-3,98095

2,24762

2,99047

2.91428

2,43810

2,57143

3,69524

3,77143

3,56190

3,23810

3,31428

3.02858

3,75238

26

26 26

26

26

26 26 26

26

26 26

26

,887

,887

,887

887

.887

,887

,887

.887

.887

,887

,887

887

,887

,887

,887

.887

,887

,887

,887

.887

,887

. ,887

,887

.887

,887

. ,887

.887

.887

B.1.2 Verteilung

		m3kb2c1	m4kb2c1	m5kb2c1	m6kb2c1	m7kb2c1	m1kb1c2	m2kb1c2	m3kb1c2	m4kb1c2
		makiizut	III4KII261	makiizut	mokiizei	III/MIZCI	III MITCZ	IIIZKII IGZ	III JAIL 162	III4KII162
N		26	26	26	26	26	26	26	26	26
Normal Parameters ^{a,b}	Mean	-1,6633700	-2,4175827	-2,4399262	-2,4930404	-2,6183150	-1,0717954	-1,3245427	-,9855308	-1,2974358
	Std. Deviation	,77438804	,76268410	,68162854	,73191103	,85811988	,79189594	,82339080	,64262277	,88154600
Most Extreme	Absolute	,157	,122	,099	,134	,143	,118	,141	,191	,155
Differences	Positive	,069	,120	,068	,134	,143	,089	,098	,114	,086
	Negative	-,157	-,122	-,099	-,112	-,129	-,118	-,141	-,191	-,155
Kolmogorov-Smirnov Z		,801	,624	,507	,681	,727	,602	,717	,971	,790
Asymp. Sig. (2-tailed)		.543	.831	.960	.743	.665	.861	.682	,302	.560

		m6kh2c2	m7kh2c2
N		26	26
Normal Parameters ^{a,b}	Mean	-1,5992677	-1,5010981
	Std. Deviation	,83264948	,87870354
Most Extreme	Absolute	,184	,150
Differences	Positive	,136	,075
	Negative	-,184	-,150
Kolmogorov-Smirnov Z		,937	,766
Asymp. Sig. (2-tailed)		,344	,600

a. Test distribution is Normal.

b. Calculated from data.



Abbildung B.1 – Verteilung der Differenzgrade für alle Faktorstufen



Abbildung B.2 – Verteilung der angpassten Differenzgrade für alle Faktorstufen

B.1.3 Reliabilität

Reliability Statistics

Cronbach's Alpha	Part 1	Value	,862
		N of tems	13 ^a
	Part 2	Value	,937
		N of tems	13 ^b
	Total N of tems	3	26
Correlation Between	n Forms		,943
Spearman-Brown	Equal Length		,970
Coefficient	Unequal Length	1	,970
Guttman Split-Half (Coefficient		,924

Scale	Statistics
Scale	Jungaroa

	Mean	Variance	Std. Deviation	N of Items
Part 1	-24,3449	27,844	5,27678	13 ^a
Part 2	-21,2534	66,660	8,16456	13 ^b
Both Parts	-45,5983	175,726	13,25616	26

Item-Total Statistics

		Scale	Corrected	Squared	Cronbach's
	Scale Mean if	Variance if	Item-Total	Multiple	Alpha if Item
	Item Deleted	Item Deleted	Correlation	Correlation	Deleted
var001	-42,3689	165,758	,539		,952
var002	-43,4296	168,084	,551		,952
var003	-44,6316	163,606	,717		,951
var004	-43,2881	165,375	,444		,954
var005	-43,5145	166,911	,554		,952
var006	-44,0752	171,846	,258		,954
var007	-44,7956	170,312	,438		,953
var008	-43,9046	167,199	,551		,952
var009	-44,0058	161,198	,572		,953
var010	-44,3032	163,215	,649		,951
var011	-43,2990	162,786	,805		,950
var012	-43,9527	166,589	,637		,952
var013	-42,8643	160,335	,759		,950
var014	-43,8221	148,683	,860		,949
var015	-44,5352	164,469	,616		,952
var016	-44,1636	158,193	,847		,949
var017	-43,2942	160,283	,690		,951
var018	-44,3551	160,093	,905		,949
var019	-44 2031	148 155	875		949
var020	-43,3963	165,172	,591		,952
var021	-43,7527	163,000	,742		,951
var022	-44 8378	165 404	787		951
var023	-43,6534	146,603	,855		,950
var024	-43,0371	161,621	,907		,949
var025	-44,3381	170,846	,391		,953
var026	-44,1359	166.355	.614		.952

Varianzanalyse **B.2**

B.2.1 Dreifaktorielle ANOVA

Mauchly's Test of Sphericity^b

Measure:	MEASURE	1
mououro.		

					Epsilon ^a			
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh-Feldt	Lower-bound	
method	,159	41,933	20	,003	,645	,778	,167	
kh	1,000	,000	0		1,000	1,000	1,000	
content	1,000	,000	0	-	1,000	1,000	1,000	
method * kh	,085	56,119	20	,000	,486	,558	,167	
method * content	,279	29,094	20	,089	,760	,951	,167	
kh * content	1,000	,000	0		1,000	1,000	1,000	
method * kh * content	,307	26,877	20	,142	,746	,929	,167	

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table. b.

Design: Intercept Within Subjects Design: method+kh+content+method*kh+method*content+kh*content+method*kh*content

Tests of Within-Subjects Effects

Measure: MEASURE_1

		Type III Sum					Partial Eta	Noncent.	Observed
Source		of Squares	df	Mean Square	F	Sig.	Squared	Parameter	Power
method	Sphericity Assumed	29,735	6	4,956	22,429	,000	,473	134,575	1,000
	Greenhouse-Geisser	29,735	3,872	7,680	22,429	,000	,473	86,844	1,000
	Huynh-Feldt	29,735	4,670	6,367	22,429	,000	,473	104,746	1,000
	Lower-bound	29,735	1,000	29,735	22,429	,000	,473	22,429	,995
Error(method)	Sphericity Assumed	33,143	150	,221					
	Greenhouse-Geisser	33,143	96,798	,342					
	Huynh-Feldt	33,143	116,752	,284					
	Lower-bound	33,143	25,000	1,326					
kh	Sphericity Assumed	9,281	1	9,281	11,077	,003	,307	11,077	,892
	Greenhouse-Geisser	9,281	1,000	9,281	11,077	,003	,307	11,077	,892
	Huynh-Feldt	9,281	1,000	9,281	11,077	,003	,307	11,077	,892
	Lower-bound	9,281	1,000	9,281	11,077	,003	,307	11,077	,892
Error(kh)	Sphericity Assumed	20,947	25	,838					
	Greenhouse-Geisser	20,947	25,000	,838					
	Huynh-Feldt	20,947	25,000	,838					
	Lower-bound	20,947	25,000	,838					
content	Sphericity Assumed	132,379	1	132,379	47,551	,000	,655	47,551	1,000
	Greenhouse-Geisser	132,379	1,000	132,379	47,551	,000	,655	47,551	1,000
	Huynh-Feldt	132,379	1,000	132,379	47,551	,000	,655	47,551	1,000
	Lower-bound	132,379	1,000	132,379	47,551	,000	,655	47,551	1,000
Error(content)	Sphericity Assumed	69,599	25	2,784					
	Greenhouse-Geisser	69,599	25,000	2,784					
	Huynh-Feldt	69,599	25,000	2,784					
	Lower-bound	69,599	25,000	2,784					
method * kh	Sphericity Assumed	4,023	6	,671	3,371	,004	,119	20,226	,932
	Greenhouse-Geisser	4,023	2,919	1,379	3,371	,024	,119	9,839	,733
	Huynh-Feldt	4,023	3,346	1,202	3,371	,018	,119	11,280	,777
	Lower-bound	4,023	1,000	4,023	3,371	,078	,119	3,371	,423
Error(method*kh)	Sphericity Assumed	29,839	150	,199					
	Greenhouse-Geisser	29,839	72,968	,409					
	Huynh-Feldt	29,839	83,653	,357					
	Lower-bound	29,839	25,000	1,194					

Source		Type III Sum of Squares	df	Mean Square	F	Sia.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
method * content	Sphericity Assumed	4,879	6	,813	5,124	,000	,170	30,742	,993
	Greenhouse-Geisser	4,879	4,561	1,070	5,124	,000	,170	23,369	,975
	Huynh-Feldt	4,879	5,704	,855	5,124	,000	,170	29,226	,991
	Lower-bound	4 879	1 000	4 879	5 124	033	170	5 124	586
Error(method*content)	Sphericity Assumed	23,806	150	,159					
	Greenhouse-Geisser	23,806	114,026	,209					
	Huynh-Feldt	23,806	142,603	,167					
	Lower-bound	23,806	25,000	,952					
kh * content	Sphericity Assumed	,710	1	,710	1,569	,222	,059	1,569	,226
	Greenhouse-Geisser	,710	1,000	,710	1,569	,222	,059	1,569	,226
	Huynh-Feldt	,710	1,000	,710	1,569	,222	,059	1,569	,226
	Lower-bound	,710	1,000	,710	1,569	,222	,059	1,569	,226
Error(kh*content)	Sphericity Assumed	11,318	25	,453					
	Greenhouse-Geisser	11,318	25,000	,453					
	Huynh-Feldt	11,318	25,000	,453					
	Lower-bound	11,318	25,000	,453					
method * kh * content	Sphericity Assumed	1,477	6	,246	1,480	,189	,056	8,878	,562
	Greenhouse-Geisser	1,477	4,477	,330	1,480	,208	,056	6,624	,474
	Huynh-Feldt	1,477	5,574	,265	1,480	,194	,056	8,248	,539
	Lower-bound	1,477	1,000	1,477	1,480	,235	,056	1,480	,216
Error(method*kh*content)	Sphericity Assumed	24,947	150	,166					
	Greenhouse-Geisser	24,947	111,919	,223					
	Huynh-Feldt	24,947	139,348	,179					
1	Lower-bound	24,947	25,000	,998					

a. Computed using alpha = ,05

B.2.1.1 Einzelvergleiche

method

Pairwise Comparisons

Measure: MEASURE_1

		Mean			95% Confiden Differ	ice Interval for rence ^a
(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
1	2	,483*	,062	,000	,274	,693
	3	,028	,062	1,000	-,180	,237
	4	,473*	,076	,000	,217	,729
	5	,422*	,082	,001	,147	,698
	6	,447*	,059	,000	,247	,647
	7	,471*	,078	,000	,208	,734
2	1	-,483*	,062	,000	-,693	-,274
	3	-,455*	,075	,000	-,709	-,201
	4	-,011	,040	1,000	-,146	,125
	5	-,061	,074	1,000	-,309	,187
	6	-,036	,050	1,000	-,204	,131
	7	-,012	,068	1,000	-,241	,217
3	1	-,028	,062	1,000	-,237	,180
	2	,455*	,075	,000	,201	,709
	4	,444*	,080	,000	,175	,713
	5	,394*	,079	,001	,128	,661
	6	,419*	,065	,000	,200	,638
	7	,443*	,073	,000	,196	,689
4	1	-,473*	,076	,000	-,729	-,217
	2	,011	,040	1,000	-,125	,146
	3	-,444*	,080,	,000	-,713	-,175
	5	-,050	,065	1,000	-,271	,170
	6	-,026	,048	1,000	-,187	,135
-	/	-,001	,059	1,000	-,199	,196
5	1	-,422*	,082	,001	-,698	-,147
	2	,061	,074	1,000	-,187	,309
	3	-,394*	,079	,001	-,661	-,128
	4	,050	,065	1,000	-,1/0	,2/1
	7	,025	,051	1,000	-,148	,197
6	1	,049	,054	1,000	-,132	,230
•	2	-,447	,059	,000	-,047	-,247
	2	,030	,000	1,000	-,131	,204
	4	-,419	,005	1,000	-,030	-,200
	5	,020	,040	1,000	-,133	,107
	7	-,025	,031	1,000	-,197	,140
7	. 1	,024	,045	1,000	-,121	- 209
l .	2	012	,078	1 000	- 217	-,200
	3	- 4/3*	073	000	- 680	- 106
	4	001	059	1,000	-,009	199
	5	- 049	,054	1,000	- 230	132
	6	- 024	045	1,000	- 175	127

Based on estimated marginal means *. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Sidak.

method * kh * content

Pairwise Comparisons

Measure: MEASURE_1

				Mean			95% Confiden Differ	ce Interval for ence ^a
method	content	(I) kh	(J) kh	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
1	1	1	2	,049	,121	,691	-,200	,297
		2	1	-,049	,121	,691	-,297	,200
	2	1	2	,158	,136	,255	-,122	,438
		2	1	-,158	,136	,255	-,438	,122
2	1	1	2	,353*	,119	,006	,108	,597
		2	1	-,353*	,119	,006	-,597	-,108
	2	1	2	,355*	,137	,016	,072	,637
		2	1	-,355*	,137	.016	637	-,072
3	1	1	2	-,327*	,134	,022	-,604	-,050
		2	1	,327*	,134	,022	,050	,604
	2	1	2	.175	,110	,123	051	.401
		2	1	-,175	,110	,123	401	.051
4	1	1	2	,149	,139	,294	-,137	,434
		2	1	-,149	,139	.294	-,434	.137
	2	1	2	,296*	,134	,037	,019	,573
		2	1	-,296*	,134	,037	-,573	-,019
5	1	1	2	,307	,158	,063	-,018	,631
		2	1	-,307	,158	.063	631	.018
	2	1	2	,272	,139	,062	-,014	,558
		2	1	-,272	,139	,062	-,558	,014
6	1	1	2	.248	,135	.079	031	.527
		2	1	-,248	,135	,079	-,527	,031
	2	1	2	.461*	,118	,001	,219	,703
		2	1	-,461*	,118	.001	-,703	-,219
7	1	1	2	.366	,189	.064	023	.754
		2	1	-,366	,189	,064	-,754	,023
	2	1	2	,301	,151	,056	-,009	,612
		2	1	-,301	,151	,056	-,612	,009

Based on estimated marginal means *. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Sidak.

		_						
				Mean Difference			95% Confiden Differ	rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
1	1	1	2	,464*	,120	,015	,059	,868
			3	,322	,115	,187	-,066	,711
			4	,601*	,127	,002	,172	1,030
			5	,465	,162	,159	-,081	1,011
			6	,577*	,132	,004	,130	1,023
			7	,584*	,159	,023	,049	1,120
		2	1	-,464*	,120	,015	-,868	-,059
			3	-,141	,124	,998	-,560	,277
			4	,137	,070	,729	-,098	,372
			5	,002	,157	1,000	-,528	,531
			6	,113	,117	1,000	-,281	,507
			7	,121	,165	1,000	-,434	,676
		3	1	-,322	,115	,187	-,711	,066
			2	,141	,124	,998	-,277	,560
			4	,278	,144	,758	-,208	,765
			5	,143	,166	1,000	-,418	,703
			6	,254	,115	,545	-,134	,643
			7	,262	,159	,917	-,274	,798
		4	1	-,601*	,127	,002	-1,030	-,172
			2	-,137	,070	,729	-,372	,098
			3	-,278	,144	,758	-,765	,208
			5	-,136	,149	1,000	-,638	,366
			6	-,024	,114	1,000	-,407	,359
			7	-,016	,159	1,000	-,551	,518
		5	1	-,465	,162	,159	-1,011	,081
			2	-,002	,157	1,000	-,531	,528
			3	-,143	,166	1,000	-,703	,418
			4	,136	,149	1,000	-,366	,638
			6	,111	,100	,999	-,227	,450
			7	,119	,077	,951	-,140	,378

Pairwise Comparisons

Measure: MEASURE_1

				Mean Difference			95% Confider Diffe	nce Interval for rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
1	1	6	1	-,577*	,132	,004	-1,023	-,130
			2	-,113	,117	1,000	-,507	,281
			3	-,254	,115	,545	-,643	,134
			4	,024	,114	1,000	-,359	,407
			5	-,111	,100	,999	-,450	,227
			7	,008	,088	1,000	-,288	,304
		7	1	-,584*	,159	,023	-1,120	-,049
			2	-,121	,165	1,000	-,676	,434
			3	-,262	,159	,917	-,798	,274
			4	,016	,159	1,000	-,518	,551
			5	-,119	,077	,951	-,378	,140
			6	-,008	,088	1,000	-,304	,288
	2	1	2	,253	,106	,418	-,106	,611
			3	-,086	,152	1,000	-,597	,425
			4	,226	,112	,694	-,152	,603
			5	,194	,131	,967	-,246	,634
			6	,066	,143	1,000	-,417	,549
			7	,128	,131	1,000	-,313	,569
		2	1	-,253	,106	,418	-,611	,106
			3	-,339	,151	,516	-,848	,170
			4	-,027	,082	1,000	-,305	,251
			5	-,059	,125	1,000	-,481	,363
			6	-,187	,137	,986	-,648	,275
			7	-,125	,113	,999	-,505	,255
		3	1	,086	,152	1,000	-,425	,597
			2	,339	,151	,516	-,170	,848
			4	,312	,148	,624	-,187	,811
			5	,280	,130	,589	-,159	,719
			6	,152	,083	,826	-,129	,434
			7	,214	,122	,866	-,197	,625

				Mean			95% Confiden Differ	ice Interval for rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
1	2	4	1	-,226	,112	,694	-,603	,152
			2	,027	,082	1,000	-,251	,305
			3	-,312	,148	,624	-,811	,187
			5	-,032	,128	1,000	-,463	,400
			6	-,160	,134	,997	-,612	,293
			7	-,098	,100	1,000	-,433	,238
		5	1	-,194	,131	,967	-,634	,246
			2	,059	,125	1,000	-,363	,481
			3	-,280	,130	,589	-,719	,159
			4	,032	,128	1,000	-,400	,463
			6	-,128	,137	1,000	-,590	,335
			7	-,066	,112	1,000	-,445	,313
		6	1	-,066	,143	1,000	-,549	,417
			2	,187	,137	,986	-,275	,648
			3	-,152	,083	,826	-,434	,129
			4	,160	,134	,997	-,293	,612
			5	,128	,137	1,000	-,335	,590
			7	,062	,096	1,000	-,262	,386
		7	1	-,128	,131	1,000	-,569	,313
			2	,125	,113	,999	-,255	,505
			3	-,214	,122	,866	-,625	,197
			4	,098	,100	1,000	-,238	,433
			5	,066	,112	1,000	-,313	,445
			6	-,062	,096	1,000	-,386	,262
2	1	1	2	,768*	,111	,000	,394	1,141
			3	-,053	,066	1,000	-,275	,169
			4	,701*	,126	,000	,276	1,126
			5	,723*	,121	,000	,315	1,131
			6	,776*	,116	,000	,384	1,169
			7	,902*	,147	,000	,407	1,396

				Mean			95% Confider Diffe	ice Interval for rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
2	1	2	1	-,768*	,111	,000	-1,141	-,394
			3	-,821*	,098	,000	-1,151	-,491
			4	-,067	,066	1,000	-,288	,155
			5	-,044	,098	1,000	-,374	,285
			6	,009	,106	1,000	-,348	,365
			7	,134	,143	1,000	-,349	,617
		3	1	,053	,066	1,000	-,169	,275
			2	,821*	,098	,000	,491	1,151
			4	,754*	,104	,000	,403	1,105
			5	,777*	,102	,000	,433	1,120
			6	,830*	,106	,000	,474	1,186
			7	,955*	,140	,000	,483	1,426
		4	1	-,701*	,126	,000	-1,126	-,276
			2	,067	,066	1,000	-,155	,288
			3	-,754*	,104	,000	-1,105	-,403
			5	,022	,119	1,000	-,379	,424
			6	,075	,126	1,000	-,350	,501
			7	,201	,161	,995	-,341	,742
		5	1	-,723*	,121	,000	-1,131	-,315
			2	,044	,098	1,000	-,285	,374
			3	-,777*	,102	,000	-1,120	-,433
			4	-,022	,119	1,000	-,424	,379
			6	,053	,052	1,000	-,123	,229
			7	,178	,088	,680	-,117	,474
		6	1	-,776*	,116	,000	-1,169	-,384
1			2	-,009	,106	1,000	-,365	,348
1			3	-,830*	,106	,000	-1,186	-,474
1			4	-,075	,126	1,000	-,501	,350
1			5	-,053	,052	1,000	-,229	,123
1			7	,125	,057	,540	-,066	,316

				1				
				Mean			95% Confider Differ	ice Interval for rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
2	1	7	1	-,902*	,147	,000	-1,396	-,407
			2	-,134	,143	1,000	-,617	,349
			3	-,955*	,140	,000	-1,426	-,483
			4	-,201	,161	,995	-,742	,341
			5	-,178	,088	,680	-,474	,117
			6	-,125	,057	,540	-,316	,066
	2	1	2	,449*	,109	,008	,081	,817
			3	-,070	,072	1,000	-,312	,173
			4	,363	,110	,058	-,007	,734
			5	,307	,105	,140	-,046	,661
			6	,369	,113	,064	-,012	,750
			7	,271	,100	,220	-,065	,607
		2	1	-,449*	,109	,008	-,817	-,081
			3	-,519*	,123	,006	-,935	-,103
			4	-,086	,113	1,000	-,465	,294
			5	-,142	,135	,999	-,596	,312
			6	-,080	,103	1,000	-,426	,267
			7	-,178	,110	,930	-,550	,194
		3	1	,070	,072	1,000	-,173	,312
			2	,519*	,123	,006	,103	,935
			4	,433*	,104	,007	,083	,783
			5	,377*	,103	,025	,029	,724
			6	,439*	,127	,041	,010	,867
			7	,341*	,098	,038	,011	,671
		4	1	-,363	,110	,058	-,734	,007
			2	,086	,113	1,000	-,294	,465
			3	-,433*	,104	,007	-,783	-,083
			5	-,056	,118	1,000	-,454	,341
			6	,006	,112	1,000	-,370	,382
			7	-,092	,110	1,000	-,464	,279

				Mean Difference			95% Confider Differ	ice Interval for rence ^a
kh	content	(I) method	(J) method	(I-J)	Std. Error	Sig. ^a	Lower Bound	Upper Bound
2	2	5	1	-,307	,105	,140	-,661	,046
			2	,142	,135	,999	-,312	,596
			3	-,377*	,103	,025	-,724	-,029
			4	,056	,118	1,000	-,341	,454
			6	,062	,113	1,000	-,319	,443
			7	-,036	,103	1,000	-,382	,310
		6	1	-,369	,113	,064	-,750	,012
			2	,080	,103	1,000	-,267	,426
			3	-,439*	,127	,041	-,867	-,010
			4	-,006	,112	1,000	-,382	,370
			5	-,062	,113	1,000	-,443	,319
			7	-,098	,100	1,000	-,434	,237
		7	1	-,271	,100	,220	-,607	,065
			2	,178	,110	,930	-,194	,550
			3	-,341*	,098	,038	-,671	-,011
			4	,092	,110	1,000	-,279	,464
			5	,036	,103	1,000	-,310	,382
			6	,098	,100	1,000	-,237	,434

Based on estimated marginal means *. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Sidak.

B.2.2 Zweifaktorielle ANOVA – Rauschen

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

						Epsilon ^a	
		Approx.			Greenhouse		
Within Subjects Effect	Mauchly's W	Chi-Square	df	Sig.	-Geisser	Huynh-Feldt	Lower-bound
method	,092	54,364	20	,000	,575	,678	,167
kh	1,000	,000	0		1,000	1,000	1,000
method * kh	025	84 057	20	000	430	484	167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table. b. Design: Intercept Within Subjects Design: method+kh+method*kh

Tests of Within-Subjects Effects

Measure: MEASURE_1

		Type III Sum					Partial Eta	Noncent.	Observed
Source		of Squares	df	Mean Square	F	Sig.	Squared	Parameter	Power
method	Sphericity Assumed	26,922	6	4,487	24,530	,000	,495	147,182	1,000
	Greenhouse-Geisser	26,922	3,448	7,807	24,530	,000	,495	84,585	1,000
	Huynh-Feldt	26,922	4,067	6,619	24,530	,000	,495	99,767	1,000
	Lower-bound	26,922	1,000	26,922	24,530	,000	,495	24,530	,997
Error(method)	Sphericity Assumed	27,437	150	,183					
	Greenhouse-Geisser	27,437	86,205	,318					
	Huynh-Feldt	27,437	101,677	,270					
	Lower-bound	27,437	25,000	1,097					
kh	Sphericity Assumed	2,428	1	2,428	3,779	,063	,131	3,779	,464
	Greenhouse-Geisser	2,428	1,000	2,428	3,779	,063	,131	3,779	,464
	Huynh-Feldt	2,428	1,000	2,428	3,779	,063	,131	3,779	,464
	Lower-bound	2,428	1,000	2,428	3,779	,063	,131	3,779	,464
Error(kh)	Sphericity Assumed	16,063	25	,643					
	Greenhouse-Geisser	16,063	25,000	,643					
	Huynh-Feldt	16,063	25,000	,643					
	Lower-bound	16,063	25,000	,643					
method * kh	Sphericity Assumed	4,660	6	,777	3,765	,002	,131	22,591	,958
	Greenhouse-Geisser	4,660	2,582	1,805	3,765	,019	,131	9,723	,745
	Huynh-Feldt	4,660	2,906	1,604	3,765	,015	,131	10,940	,783
	Lower-bound	4,660	1,000	4,660	3,765	,064	,131	3,765	,463
Error(method*kh)	Sphericity Assumed	30,941	150	,206					
	Greenhouse-Geisser	30,941	64,558	,479					
	Huynh-Feldt	30,941	72,644	,426					
	Lower-bound	30,941	25,000	1,238					

a. Computed using alpha = ,05

B.2.3 Zweifaktorielle ANOVA – Gitarre

Mauchly's Test of Sphericity^b

l	leasure: MEASURE_1								
							Epsilon ^a		
	Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse -Geisser	Huynh-Feldt	Lower-bound	
	method	,301	27,314	20	,130	,742	,923	,167	
	kh	1,000	,000	0		1,000	1,000	1,000	
	method * kh	,406	20,541	20	,429	,750	,935	,167	

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix. a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept Within Subjects Design: method+kh+method*kh

Measure: MEASURE_1

Tests of Within-Subjects Effects

C		Type III Sum			F	C i-	Partial Eta	Noncent.	Observed
Source		of Squares	ar	Mean Square	F	Sig.	Squared	Parameter	Power
method	Sphericity Assumed	7,692	6	1,282	6,516	,000	,207	39,097	,999
	Greenhouse-Geisser	7,692	4,452	1,728	6,516	,000	,207	29,009	,994
	Huynh-Feldt	7,692	5,536	1,390	6,516	,000	,207	36,071	,998
	Lower-bound	7,692	1,000	7,692	6,516	,017	,207	6,516	,689
Error(method)	Sphericity Assumed	29,512	150	,197					
	Greenhouse-Geisser	29,512	111,295	,265					
	Huynh-Feldt	29,512	138,391	,213					
	Lower-bound	29,512	25,000	1,180					
kh	Sphericity Assumed	7,564	1	7,564	11,671	,002	,318	11,671	,907
	Greenhouse-Geisser	7,564	1,000	7,564	11,671	,002	,318	11,671	,907
	Huynh-Feldt	7,564	1,000	7,564	11,671	,002	,318	11,671	,907
	Lower-bound	7,564	1,000	7,564	11,671	.002	,318	11,671	.907
Error(kh)	Sphericity Assumed	16,202	25	,648					
	Greenhouse-Geisser	16,202	25,000	,648					
	Huynh-Feldt	16,202	25,000	,648					
	Lower-bound	16,202	25,000	,648					
method * kh	Sphericity Assumed	,840	6	,140	,881	,511	,034	5,285	,341
	Greenhouse-Geisser	,840	4,501	,187	,881	,488	,034	3,965	,289
	Huynh-Feldt	,840	5,611	,150	,881	,505	,034	4,943	,328
	Lower-bound	,840	1,000	,840	,881	,357	,034	,881	,147
Error(method*kh)	Sphericity Assumed	23,845	150	,159					
	Greenhouse-Geisser	23,845	112,525	,212					
	Huynh-Feldt	23,845	140,283	,170					
	Lower-bound	23,845	25,000	,954					

a. Computed using alpha = ,05

C VERSUCHSANLEITUNG UND PERSONENFRAGEBOGEN

C.1 Instruktion

Liebe/r Versuchsteilnehmer/in,

Sie nehmen heute an einem Hörversuch zur Evaluation von Kopfhörer-Entzerrungsmethoden teil. Dazu hören Sie eine reale Schallquelle im direkten Vergleich mit simulierten Versionen derselben Schallquelle. Sie werden die Methoden über zwei Kopfhörertypen und anhand von zwei Stimuli (rosa Rauschen, akustische Gitarre) beurteilen.

Die unterschiedlich entzerrten Simulationen werden jeweils paarweise mit der realen Schallquelle präsentiert. Ihre Aufgabe besteht darin, jedes Paar mit dem als Referenz markierten Stimulus (der realen Schallquelle) zu vergleichen und zu entscheiden, welcher Stimulus nicht dieser Referenz entspricht. Ihre Entscheidung äußern Sie, indem Sie am entsprechenden Schieberegler einstellen, als wie deutlich Sie den wahrgenommenen Unterschied empfinden. Die Beurteilung innerhalb eines Fensters sollten relativ zueinander gemacht werden. Bitte vergessen Sie hierbei jedoch nicht, dass nach dem Unterschied zur Referenz und nicht nach dem Gefallen gefragt wird.



In der Mitte des Hörversuchs werden Sie gebeten, den Kopfhörer zu wechseln. Bitte melden Sie sich dann bei der Versuchsleiterin.

C.2 Fragebogen

Wie alt sind Sie?	Jahre
Geschlecht:	weiblich männlich
Sind Sie musikalisch ausgebildet? (z B Instrumentalkenntnisse, Musikstudium, Tonmeister)	☐ nein ☐ ja,
Haben Sie Erfahrung mit Hörversuchen?	☐ nein ☐ ja
Kererenz erkannt haben (diese mussen nicht nu Merkmale waren, bringen Sie diese bitte in ein lichsten Unterschied.	r die Klangfarbe betreffen). Falls es mehrer e Rangfolge, beginnend mit dem offensicht