

Technische Universität Berlin  
Fakultät 1 – Institut für Sprache und Kommunikation  
Fachgebiet Audiokommunikation

# Perzeptiv motivierte Reduktion von Schallfeldbeschreibungen für eine Resynthese in Echtzeit

Masterarbeit

Erstgutachter: Prof. Dr. Stefan Weinzierl  
Zweitgutachter: Alexander Lindau, M.A.

Max Röhrbein



Studiengang: Audiokommunikation und -technologie  
September 14, 2011

# Kurzfassung

Diese Arbeit beschäftigt sich mit der Reduktion modellierter Raumschallfelder auf perzeptiv relevante Anteile durch ein auditorisches Modell. Diese Reduktion ist notwendig bei der Wiedergabe/Reproduktion virtueller Akustiken in realen Räumen durch arraybasierte Lautsprecherverfahren.

In dem auditorischen Modell werden die Verarbeitung und letztlich die Wahrnehmung von räumlichen Schallfeldern durch das menschliche auditorische System nachgebildet, um Entscheidungen zur Hörbarkeit einzelner Schallfeldkomponenten (einzelner Reflexionen) treffen zu können. Der Prototyp des auditorischen Reduktionsalgorithmus wurde darauf ausgerichtet, dass Ergebnisse einschlägiger empirischer Studien reproduziert werden können. Ein Schallfeld wurde in einem detaillierten CAD-Raummodell berechnet und durch das auditorische Modell reduziert. Schließlich wurden unterschiedlich stark auditorisch reduzierte Schallfelder durch dynamische Binauralsynthese hörbar gemacht und in einem Hörversuch mit einem hochaufgelösten Referenz-Schallfeld verglichen. Die Ergebnisse dieser ersten Studie stellen eine potentiell deutliche Reduzierung von Reflexionen für die Darstellung räumlicher Schallfelder ohne perzeptive Einschränkungen in Aussicht.

## Danksagung

Herzlichen Dank an die Mitarbeiter des Fachgebiets Audiokommunikation, insbesondere Alexander Lindau, Frank Schultz, Michael Horn, Fabian Brinkmann und Stefan Weinzierl. Des weiteren bedanke ich mich bei den Mitarbeitern des Projektes “Klangakademie” Antonio Cardenas und Johnny Nahas für die konstruktiven Diskussionen im Verlaufe der Arbeit.

Mein besonderer Dank gilt an dieser Stelle meinen Eltern für die Unterstützung während des gesamten Studiums und der Erstellung meiner Masterarbeit. Für die aufbauenden Worte danke ich meiner Freundin Monique Schulz und all meinen Freunden.

# Erklärung

Die selbstständige und eigenhändige Anfertigung versichere ich an Eides statt.

Max Röhrbein  
Berlin 14. September 2011

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b>	<b>4</b>
<b>1 Einleitung</b>	<b>6</b>
<b>2 Stand der Forschung</b>	<b>8</b>
2.1 Methoden zur Bestimmung von Raumschallfeldern . . . . .	8
2.2 Empirische Untersuchung zur Hörbarkeit von Raumreflexionen . . . . .	12
2.2.1 Schwellwertverfahren und verwendete Apparate . . . . .	12
2.2.2 Ergebnisse . . . . .	14
2.3 Auditorisch motivierte Algorithmen zur Reduktion von Raumschallfeldern	19
2.3.1 Auditorische Parametrisierung realer Schallfelder nach Hulsebos et al. [1] . . . . .	19
2.3.2 Schwellenwertansatz nach Begault et al. [2] . . . . .	20
2.3.3 Geometrischer Ansatz nach Hacıhabiboğlu und Murtagh [3] . . . . .	20
2.3.4 Auditorisches Verdeckungsmodell nach Buchholz et al. [4] . . . . .	22
<b>3 Methodik</b>	<b>26</b>
3.1 Realer Raum . . . . .	27
3.1.1 Frequenzgangsentzerrung des Dodekaeder-Lautsprechers . . . . .	28
3.1.2 Messung der binauralen Raumimpulsantwort . . . . .	31
3.2 Modellierter Raum . . . . .	36
3.2.1 Modellanpassung - 'Fitting' - des Raumakustikmodells an die Realität	37
3.2.2 Berechnung eines Reflektogramms mit EASE - Aura Response . . .	42
3.2.3 Extraktion wichtiger Parameter aus dem Reflektogramm . . . . .	44
3.2.4 Erzeugung von Impulsantworten der extrahierten Amplitudenspektren aus dem Reflektogramm . . . . .	46
3.3 Auditorisch reduzierter modellierter Raum . . . . .	48
3.3.1 Das einfache auditorische Prozessmodell . . . . .	48
3.3.2 Vergleich von Gammaton-Filterbänken . . . . .	53
3.3.3 Kalibrierung des einfachen auditorischen Prozessmodells . . . . .	60
3.3.4 Erweiterung des einfachen auditorischen Modells auf Raumschallfelder	62
3.3.5 Wirkungsweise des erweiterten auditorischen Modells . . . . .	65

<b>4</b>	<b>Perzeptive Evaluation</b>	<b>69</b>
4.1	Auralisation - Binauralsynthese . . . . .	69
4.2	Nachhallsynthese . . . . .	71
4.3	Vorversuch . . . . .	75
4.4	Versuchsbeschreibung . . . . .	80
4.4.1	Versuchsmethode und -design . . . . .	80
4.4.2	Stimuli . . . . .	81
4.4.3	Versuchsaufbau und -durchführung . . . . .	82
4.5	Ergebnisse . . . . .	84
4.5.1	Unterscheidungsmerkmale bezogen auf die Stimuli . . . . .	84
4.5.2	Statistische Auswertung . . . . .	85
4.6	Diskussion der Hörversuchsergebnisse . . . . .	87
<b>5</b>	<b>Zusammenfassung und Diskussion</b>	<b>90</b>
<b>A</b>	<b>Inhalt DVD</b>	<b>93</b>
<b>B</b>	<b>Anhang - Versuchsanleitung und Fragebogen</b>	<b>94</b>
B.1	Instruktionen . . . . .	94
B.2	Fragebogen . . . . .	97
	<b>Abbildungsverzeichnis</b>	<b>98</b>
	<b>Tabellenverzeichnis</b>	<b>101</b>
	<b>Literaturverzeichnis</b>	<b>102</b>

# Kapitel 1

## Einleitung

Diese Arbeit bewegt sich in dem Forschungsbereich der virtuellen Raumakustik. Die virtuelle Akustik bezeichnet die akustische Simulation von Räumen. Ein virtueller akustischer Eindruck eines Raumes kann durch die Reproduktion eines Schallfeldes für eine Quell-, Empfängerkombination mit geeigneten Wiedergabeverfahren erzeugt werden.

Das Schallfeld einer Quelle in geschlossenen Räumen setzt sich aus dem Direktsignal und einer Reihe sich überlagernder Reflexionen in einem Reflexionsmuster zusammen. Die einzelnen Reflexionen, Schallfeldkomponenten, können, nach einer gewissen Ausbreitungsdauer, als ebene Wellen angenähert werden, die jeweils am Empfängerort zu einem bestimmten verzögerten Zeitpunkt, mit unterschiedlichem Einfallswinkel, spezifischen Pegelinformationen und spektralem Inhalt eintreffen. Aus physikalischer Sicht ist das Schallfeld exakt in der Raumimpulsantwort (RIR) in seiner zeitlichen Entwicklung beschrieben. Aus perzeptiver Sicht kann man davon ausgehen, dass nur eine begrenzte Anzahl von sich ausbreitenden Reflexionen für die Wahrnehmung des Schallfeldes durch einen Menschen ausschlaggebend sind. Die Hörbarkeit der Reflexionen ist durch psychoakustische Effekte, wie Verdeckungsmechanismen im menschlichen Hörsystem beeinflusst.

Ziel dieser Arbeit ist die Entwicklung eines auditorischen Algorithmus', welcher eine perzeptiv motivierte Reduktion von Schallfeldbeschreibungen ermöglicht. Als Ausgangsbasis für die Reduktion dient dabei eine hochgenaue numerische Beschreibung des Schallfeldes durch Einzelreflexionen an einem Hörerort. Eine solche Schallfeldbeschreibung muss zunächst durch numerische Verfahren generiert werden. Existierende Methoden zur Bestimmung von Raumschallfeldern werden in dem folgenden Kapitel genannt. Die in dieser Arbeit verwendete Methode wird ansatzweise beschrieben.

Bei der Entwicklung eines auditorischen Algorithmus' zur Vorhersage der Hörbarkeit von Schallfeldkomponenten (Einzelreflexionen), bzw. der Verdeckung derselben durch vorhergehende Reflexionsmuster wird vom Stand der Forschung ausgegangen. Ein geeignetes auditorisches Modell sollte demzufolge die für zeitliche, räumliche und spektrale Verdeckung notwendigen akustischen, mechanischen, neuronalen und kognitiven Prozesse des gesamten menschlichen Hörsystems nachbilden.

Im Hintergrund dieser Arbeit steht die Zusammenarbeit des Fachbereichs Audiokommunikation mit der Firma ADAM Audio GmbH in dem Projekt “Klangakademie”. Das Projekt hat zum Ziel, Voraussetzungen für den Bau eines Konzertsaals zu untersuchen, in dem sowohl beliebige reale als auch künstlerisch-kreativ ‘kreierte’ Raumakustiken resynthetisiert werden können. Damit kann der auditive Eindruck eines anderen Ortes, z.B. einer anderen Spielstätte durch ein zu installierendes Wiedergabesystem in dem Konzertsaal der “Klangakademie” simuliert werden. Der Fachbereich Audiokommunikation unterstützt die technische Planung des in dem geplanten Saal zu installierenden lautsprecherbasierten Wiedergabesystems, mit dem die variablen Akustiken simuliert werden können. Demnach werden Vorschläge für ein räumliches Beschallungskonzept, basierend auf bekannten Wiedergabeverfahren entworfen. Die zu untersuchenden Wiedergabeverfahren sollen bezüglich ihrer klanglichen Eignung überprüft werden, um die gewünschten, in Form parametrischer Beschreibungen vorliegenden Schallfelder, zu resynthetisieren. Die dabei zu untersuchenden Verfahren umfassen die Wellenfeldsynthese [5, 6], aber auch andere Verfahren, wie z.B. Ambisonics-Systeme höherer Ordnung [7]. Neben der Eignung der räumlichen Wiedergabealgorithmen sollen weitere Aspekte, wie z.B. die jeweils geeignete Lautsprecherkonfiguration oder die Abhängigkeit des Höreindrucks von einem Sitzplatz demonstriert und untersucht werden. Dazu sollen die Wiedergabesysteme durch akustische Simulation mit der dynamischen Binauralsynthese hörbar gemacht werden. Die solcherart erzeugten binauralen Simulationen sollen die Entscheidungsfindung bzgl. Qualität und Umsetzbarkeit der in Betracht kommenden Verfahren bereits im Vorfeld erleichtern. Für dieses Projekt gehen arraybasierte Wiedergabeverfahren zunächst von einer aus raumakustischen Raummodellen hervorgegangene Schallfeldbeschreibung aus. Zur Reproduktion des Schallfeldes sind, je nach Verfahren spezifische Ansteuerfilter (sog. Treiberfunktionen) zu berechnen. Dieser Vorgang kann aufgrund der hohen Anzahl von Raumschallfeldkomponenten sehr rechenintensiv werden. Je nach Syntheseverfahren und den zu Verfügung stehenden Berechnungskapazitäten wird es nötig sein, die Schallfeldbeschreibung unter dem Aspekt perceptiver Plausibilität zu vereinfachen. In dieser Arbeit wird davon ausgegangen, dass der Renderingprozess zur Schallfeldreproduktion durch eine vereinfachte Darstellung des Schallfeldes in erlebbarer Zeit ermöglicht werden kann.



# Kapitel 2

## Stand der Forschung

### 2.1 Methoden zur Bestimmung von Raumschallfeldern

Exakte, datenbasierte Schallfeldbeschreibungen kann man durch Schallfeldmessung mit einem Lautsprecher und Mikrofonarrays erhalten [8, 9, 10]. Nach geeigneter Schallfelddekomposition kann die Schallfeldrepräsentation in parametrische Modelle überführt werden, die für eine Datenreduktion nach auditorischen Gesichtspunkten verwendet werden können.

Alternativ kann man Schallfeldbeschreibungen modellbasiert durch Simulationsverfahren mit geometrischen oder wellenbasierten (z.B. Finite-Elemente-Methoden - FEM oder Boundary-Element-Methoden - BEM) Ansätzen berechnen.

Bei den geometrischen Ansätzen werden 3D-Raummodelle durch eine Kombination von graphischen und numerischen Eingaben der Raumgeometrien erstellt [11]. Neben der Geometrie des Raumes (Primärstruktur) werden auch Materialbeschaffenheit und damit die akustischen Eigenschaften der Begrenzungsflächen in dem Modell nachgebildet (Sekundärstruktur). Hauptsächlich werden den Wänden frequenzabhängige Werte für den Absorptionsgrad und Streugrad zugewiesen. Diese Werte erhält man aus Datenbanken, die von Materialherstellern zu Verfügung stehen.

Wenn Geometrie und akustische Eigenschaften der Begrenzungsflächen eines Raummodells definiert sind, können zunächst statistische Parameter, wie z.B. die Nachhallzeit (nach Sabine oder Eyring) direkt berechnet werden. Für die Berechnung einer Schallfeldbeschreibung müssen Quellen und Empfänger in das 3D-Modell des Raumes eingefügt werden. Die Quelle kann dabei natürliche Quellen (z.B. Sänger oder Musikinstrumente), breitbandig abstrahlende Kugelquellen (wie sie i.d. Realität als Dodekaeder-Lautsprecher verwendet werden), oder auch komplette Beschallungsanlagen nachbilden. Der Empfänger wird durch einen Standort bzw. einen Bereich im Raummodell definiert. In dem definierten Bereich des virtuellen Empfängers werden die Reflexionswege einzelner ausgesandter Strahlen registriert, die jeweils zeitliche Entwicklung und räumliche Informationen enthalten.

Geometrische Raumakustikmodelle behandeln Schallwellen als Strahlen, die - analog dem

optischen Fall - an den Grenzflächen reflektiert oder gestreut werden, so dass die Ausbreitung von Schallteilchen auf geraden Linien durch den Raum verfolgt werden kann. Unter Voraussetzung, dass die Dimensionen von Begrenzungsflächen oder Hindernissen groß im Verhältnis zu den Wellenlängen sind, können hinreichend gute Ergebnisse für die Berechnung des Schallfeldes erzielt werden. Algorithmen der geometrischen Raumakustik basieren prinzipiell auf zwei physikalischen Denkmodellen, dem Spiegelschallquellenmodell und dem Strahlenmodell<sup>1</sup>. Aus der Kombination der beiden Verfahren kann auch eine zeitlich ausgedehnte Raumimpulsantwort (RIR) mit vertretbarem Rechenaufwand erzeugt werden. Die RIR wird durch die Berechnung mit dem sog. hybriden Verfahren in einen frühen deterministischen Teil und mit dem Strahlenmodell in einen späten stochastischen Teil getrennt. Für eine genaue Erklärung des Spiegelschallquellen-, Strahlenmodells und dem hybriden Ansatz sei hier u.a. auf [12] (S.181 ff) verwiesen. Aus der Schallfeldberechnung mit hybriden Verfahren und stochastischem Ray-Tracing entsteht ein sogenanntes Reflektogramm. In Abb. 2.1 wird in einer schematischen Darstellung die zeitliche Entwicklung der Schallfeldbeschreibung eines Reflektogramms gezeigt.

In dem Reflektogramm setzt sich das Schallfeld aus diskret beschriebenen Reflexionen zu-

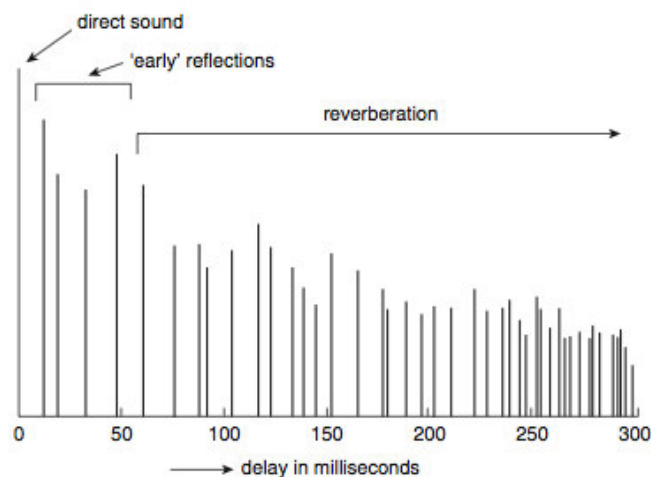
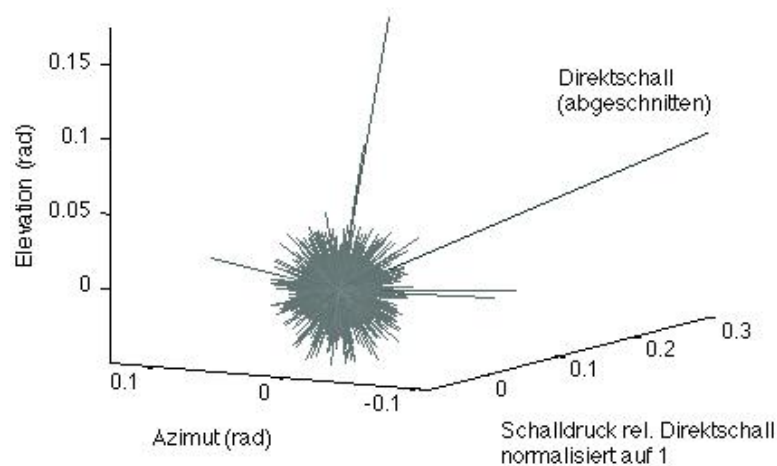


Abbildung 2.1: Schematische Darstellung eines Reflektogramms aus hybriden und stochastischen Verfahren aus [13], S.109, Fig.4.8

sammen. In Abb. 2.2 werden in einem Richtungsigelplot die räumlichen und energetischen Informationen, der an dem Empfängerort registrierten Strahlen schematisch dargestellt. Wobei die energetischen Informationen als Schalldruckanteile relativ zum Direktschall linear auf der X-Achse aufgetragen sind. Der Direktschall wurde auf 1 normalisiert und in der Abb. aus Übersichtlichkeitsgründen abgeschnitten dargestellt. Die Richtungsinformationen des Direktschalls und aller folgenden Reflexionen sind über der Y-, und Z-Achse für die winkelförmige Auslenkung in Radianten für jeweils Azimut und Elevation dargestellt.

<sup>1</sup>respektive engl. 'image source model' und 'ray tracing model'



Abbildungung 2.2: Richtungsigelplot einer Schallfeldbeschreibung mit räumlichen Informationen. Die Richtung des Direktschalls (Einfall von vorne) ist gekennzeichnet.

Eine Raumakustiksimulationssoftware, die auf dem hybriden Ray-Tracing basiert, ist z.B. EASE, die von der Firma Software Design Ahnert GmbH entwickelt wird. In EASE wird das Frequenzspektrum des Direktschalls und aller folgenden Reflexionen bandbeschränkt (125 Hz - 10 kHz) und in einer groben Auflösung (Terzen) berechnet. Weiterhin sind aus dem Reflektogramm für das Direktsignal und aller folgenden Reflexionen die akustische Laufzeit und die entsprechenden Einfallswinkel in Azimut und Elevation bekannt.

Aus pragmatischen Gründen, d.h. aus Ermangelung realer, datenbasierter Schallfeldbeschreibungen (z.B. durch Mikrofonarraymessungen), wurde in dieser Arbeit die Schallfeldbeschreibung anhand geometrischer Schallfeldsimulationen in CAD-Raummodellen benutzt. Es müssen keine zusätzlichen Arraymessungen in dem zu simulierenden Raum durchgeführt werden. Ausserdem standen Mikrofonarrays zu dem Zeitpunkt der Arbeit dem Fachbereich noch nicht zur Verfügung. Das Computerprogramm EASE kann direkt eingesetzt werden, um die Reflektogramme für verschiedene Quell- und Empfängerkombinationen zu berechnen.

Der auditorische Reduktionsalgorithmus soll anhand bestehender Forschungslage zu auditorischen Modellen entwickelt werden. Es wurden solche Modelle betrachtet, die eine perzeptiv motivierte Schallfeldreduktion zum Ziel haben. Im Folgenden werden zunächst Ergebnisse aus empirischen Untersuchungen zusammengefasst, die den Einfluss von Verdeckungseffekten auf Raumreflexionen beschreiben. Diese Ergebnisse dienen zur späteren Validierung von Hörbarkeitsentscheidungen, wie sie im Reduktionsalgorithmus getroffen wurden. Weiterhin wird ein Überblick zu auditorischen Modellen aus relevanter For-

schungsliteratur gegeben. Es werden die Ansätze gekennzeichnet, die zur Entwicklung eines eigenen Algorithmus ausschlaggebend waren.

## 2.2 Empirische Untersuchung zur Hörbarkeit von Raumreflexionen

Klangfelder in Räumen sind sehr kompliziert und von diversen Schallqualitäten beeinflusst. Entsprechend dem Konzept der numerischen Akustik entsteht das Schallfeld in einem Raum aus der mehrmaligen Reflexion eines Direktsignals an den Raumwänden. Ein Hörer hört demnach nicht nur das Direktsignal, sondern auch eine Reihe überlagerter, verzögerter Reflexionen [12]. Die Hörbarkeit einzelner Raumreflexionen kann durch Wahrnehmungsänderungen in räumlichen Klangbildern beim Hörer durch eine kriterienfreie Ermittlung von Unterschiedlichkeitsschwellen oder Wahrnehmungsschwellen bestimmt werden.

Die Wahrnehmungsschwelle kennzeichnet dabei den niedrigsten Schwellwert, bei dem der Pegel einer einzelnen Reflexion gerade zu einer Änderung der Wahrnehmung des Schallfelds führt. Wahrnehmungsschwellen werden in sogenannten einfachen Schallfeldern empirisch ermittelt.

### 2.2.1 Schwellwertverfahren und verwendete Apparate

Bei der Ermittlung von Wahrnehmungsschwellen wird gemessen, unter welchen Bedingungen ein bestimmtes verzögertes Schallsignal eben gerade hörbar wird [14]. Häufige verwendete Begriffe für die Wahrnehmungsschwelle sind “(Reflection) Masking Threshold”[4] oder “absolute Wahrnehmungsschwelle (aWs)”[14, 15].

Wahrnehmungsschwellen für Testreflexionen wurden z.B. als Funktion von

- Verzögerungszeit,
- Einfallswinkel zwischen Maskierer und Testreflexion

in Abhängigkeit

- von dem Pegel des Direktsignals,
- von dem spektralen Unterschied zwischen Direktsignal und Testreflexion,
- für verschiedene Audiostimuli
- und unter dem Einfluss zusätzlichen Nachhalls

erhoben.

In Abschnitt 2.2.2 folgt eine Zusammenfassung der Ergebnisse, bezogen auf die genannten Parameter, aus Messungen, die mit den gleichen Verfahren<sup>2</sup> durchgeführt wurden.

Die Wahrnehmung von Reflexionen wurde in einschlägigen Studien (z.B. [14, 15, 16]) anhand elektroakustisch reproduzierter einfacher Schallfelder untersucht. Dafür wurden einfache Schallfelder in entsprechenden Hörversuchen durch die Anordnung einzelner Laut-

---

<sup>2</sup>Beurteilungsverfahren

sprecher, für hinreichend verschiedene Einfallrichtungen in einem reflexionsarmen Raum simuliert. Die Lautsprecher haben einen definierten Abstand zum Kopf des Hörers. In einigen Studien [14, 15] können die Hörer den Kopf im Raum frei bewegen. Die Schallfelder enthalten ein Direktsignal als Maskierer und eine oder mehrere Testreflexionen. Ein System, wie es typischerweise für die Untersuchungen zur subjektiven Wirkung von Schallfeldern benutzt wird, ist in Abb. 2.3 dargestellt. Eine solche Apparatur ist flexibel ge-

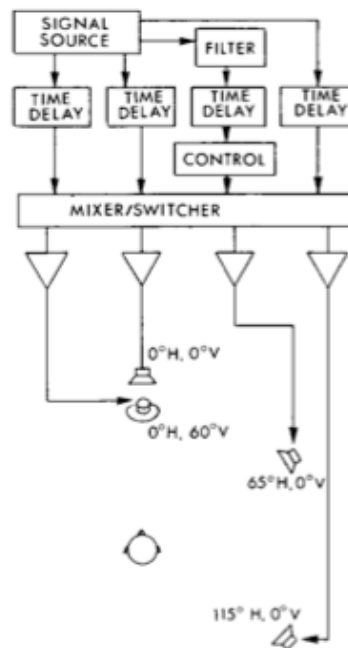


Abbildung 2.3: Messsystem für die Untersuchung zur subjektiven Wirkung von Schallfeldern aus [16], S.540, Fig.1

nug, um Bedingungen, wie Verzögerung, Filterung und Pegel beliebig anpassen zu können. Weiterhin können alle Eingangssignale beliebig auf alle Ausgänge (Lautsprecher) gemischt werden. Das “Control“-Modul in Abb.2.3 ist die Kontrollbox für die Hörer. Durch einen Potentiometer hat ein Hörer die Möglichkeit, den Pegel einer Testreflexion bis zur Wahrnehmungsschwelle, zwischen ‘gerade hörbar’ und ‘gerade nicht mehr hörbar’ zu variieren [16]. Wobei sich die Kriterien ‘gerade hörbar’ und ‘gerade nicht mehr hörbar’ auf jegliche Wahrnehmungsänderung beim Hörer im Schallfeld beziehen. Dieses sogenannte Herstellungsverfahren wird um die Möglichkeit erweitert, die Testreflexion ganz abzuschalten bzw. auf den im Versuch vorgesehenen maximalen Pegel zu erhöhen. Damit sollen Selbsttäuschungen beim Hörer vermieden werden. Das erweiterte Herstellungsverfahren wird “Beurteilungsverfahren” genannt [15]. Es ist kritisch anzumerken, dass dieses veraltete Verfahren methodisch mangelhaft erscheint, da das Kriterium des Hörers (Versuchsperson) nicht überprüfbar ist. Man geht allerdings davon aus, dass durch das Verfahren resultierende Schwellwerte nahe der minimal erreichbaren Werte liegen [16] und somit Rückschlüsse auf die Wahrnehmung einzelner Reflexionen unter definierten Bedingungen zulassen.

## 2.2.2 Ergebnisse

### Signalabhängigkeit

Bei den Untersuchungen in [16] wurde herausgefunden, dass kurze Impulse ( $10 - 40 \mu s$  Rechteckpulse) und 'rosa' Rauschen sehr ähnlich bzw. geeigneter seien als Sprache und Musik, um verzögerte Signale in Wahrnehmungstests abzubilden. Aus entsprechenden Messungen geht hervor, dass Impulse und gefärbtes Rauschen gute Beispiele für jeweils diskontinuierliche und kontinuierliche Signale sind. Schallarm und nah aufgenommene Sprachsignale ergeben Schwellen, die zwischen Schwellen von gefärbtem Rauschen und orchestrale oder quasi allen Signalen mit Nachhall liegen.

Die Abb. 2.4 zeigt die absolute Wahrnehmungsschwelle einer Versuchsperson (VP) für künstliche Seitenwand- und Deckenreflexionen für Impulse und Rauschen. Die Variation

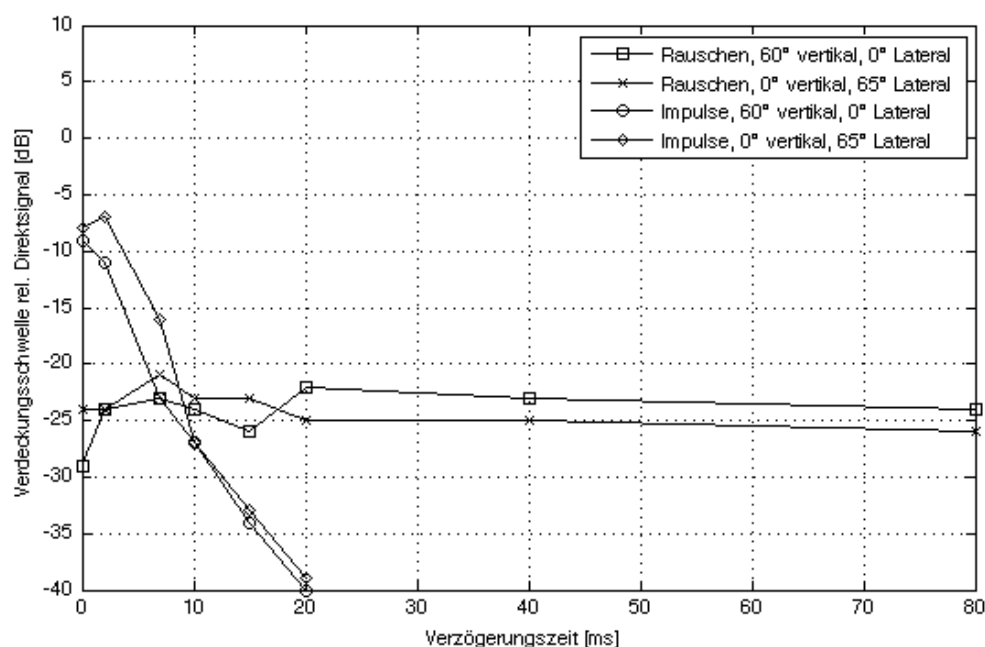


Abbildung 2.4: Signalabhängige Verdeckungsschwelle zwischen Rauschen und Impulsen aus [16], Fig.2

zwischen den sehr unterschiedlichen Einfallswinkeln ist gering. Im Gegensatz dazu sind die Unterschiede zwischen Impuls- und Rauschsignalen sehr groß. Das frühe Absinken der Wahrnehmungsschwelle für Impulse deutet darauf hin, dass diese Signale geeignet für die Echodetektion sind und zur Beschreibung von Nachverdeckungseffekten dienen können. In der Abb. 2.4 liegt die Schwelle für kontinuierliche Signale konstant bei ca. -20 bis -25 dB. Diese Ergebnisse sind vergleichbar mit den Ergebnissen zur Hörbarkeit von Kammfiltereffekten in [17]. Das deutet darauf hin, dass mit kontinuierlichen (Rausch-)Signalen der Einfluss von Kammfiltereffekten beschrieben werden kann. Ausserdem erscheinen solche

Signale geeignet für der Detektion von Klangfarbenveränderung.

In [4] wird eine vereinfachende Darstellung von Wahrnehmungsschwellen unterschiedlicher Signaltypen gezeigt, wie in Abb. 2.5 zu sehen. Daraus geht hervor, dass die Wahrnehmung von Reflexionen besonders kritisch für impulsartige Signale erscheint. Das Verhalten wird als kritisch bezeichnet, da schon nach kurzer Verzögerungszeit zwischen Direktsignal und einer Reflexion die Verdeckungsschwelle drastisch gesunken ist.

Schon nach 10 ms Verzögerung zwischen Direktsignal und Testreflexion wird die

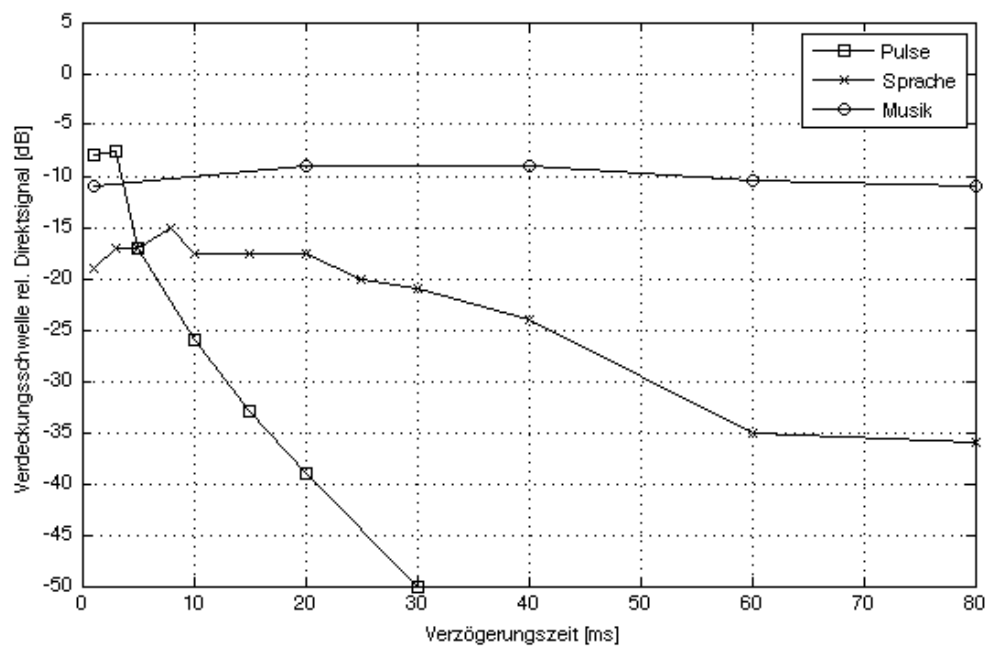


Abbildung 2.5: Signalabhängige Verdeckungsschwelle aus [4], Fig.1

Testreflexion bei -30 dB relativ zum Direktschallpegel wahrgenommen. Aus Abb. 2.4 geht jedoch hervor, dass dies erst für verzögerte Reflexionen nach ca. 10 ms zum Direktsignal gilt. Davor sind Reflexionen mit Rauschsignalen besser detektierbar. Eine Verzögerung ab 10 ms ist jedoch ein relevanter Fall in der Konzertsaalakustik.

### Der Einfluss von Einfallsrichtung

Die geringen Unterschiede der Schwellwerte von Einfallswinkeln stehen im direkten Kontrast zu dem Fall des sehr großen Unterschieds, wenn die Testreflexion aus der gleichen Richtung wie das Direktsignal kommt, wie in Abb. 2.6 zu sehen. In der Abb. 2.6 soll verdeutlicht werden, dass die Verdeckungswirkung von dem Direktsignal wesentlich höher zu sein scheint, wenn die Testreflexion aus der gleichen Richtung kommt. Als Testsignal diene hier gefärbtes 'rosa' Rauschen. Der Grund für die Wahrnehmungsänderung werden wieder Kammfiltereffekte bzw. Klangfarbenunterschiede sein.

Die absolute Wahrnehmungsschwelle als Funktion vom Einfallswinkel bei fester Verzöge-



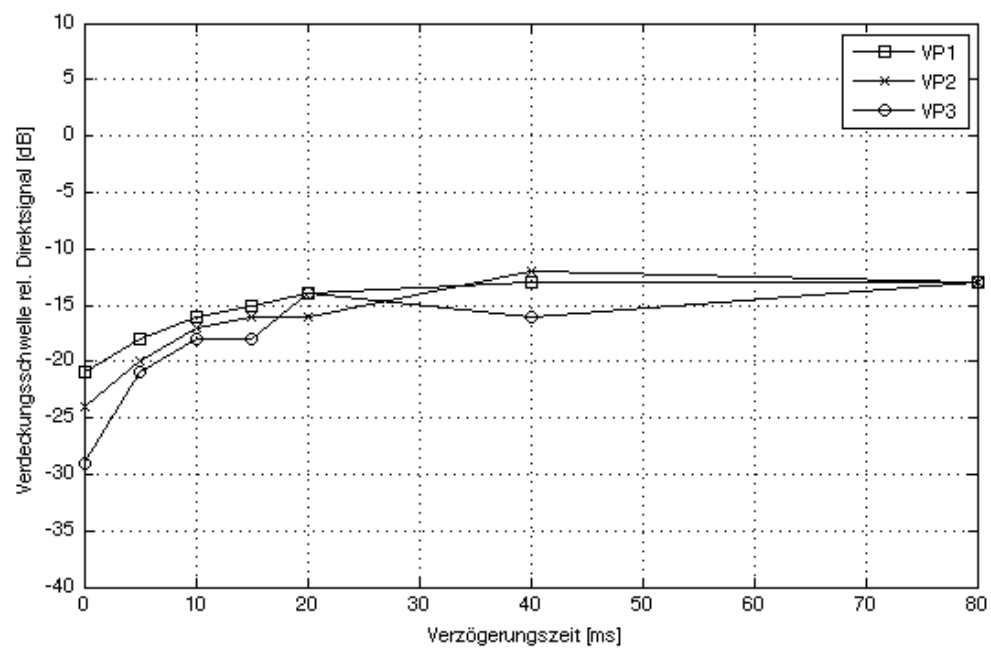


Abbildung 2.6: Verdeckungsschwellen für 3 VPs als Funktion von Verzögerungszeit bei Schalleinfall aus gleicher Richtung von vorne mit dem Stimulus ‘rosa’ Rauschen aus [16], Fig.5

rung wurde u.a. in [14] untersucht. Als Signal wurde das gesprochene Wort ‘Zack’ benutzt. Einmal fiel das Direktsignal aus der  $0^\circ$ -Richtung ein. Eine um 50 ms verzögerte Testreflexion kam aus unterschiedlichen Einfallsrichtungen von  $0^\circ$  -  $180^\circ$ . In dem zweiten Fall war der Einfallswinkel des Direktsignals variabel und die Testreflexion kam aus der  $0^\circ$ -Richtung. Die Wahrnehmungsschwellen sind in der Abb. 2.7 dargestellt. Die Wahrnehmungsschwelle

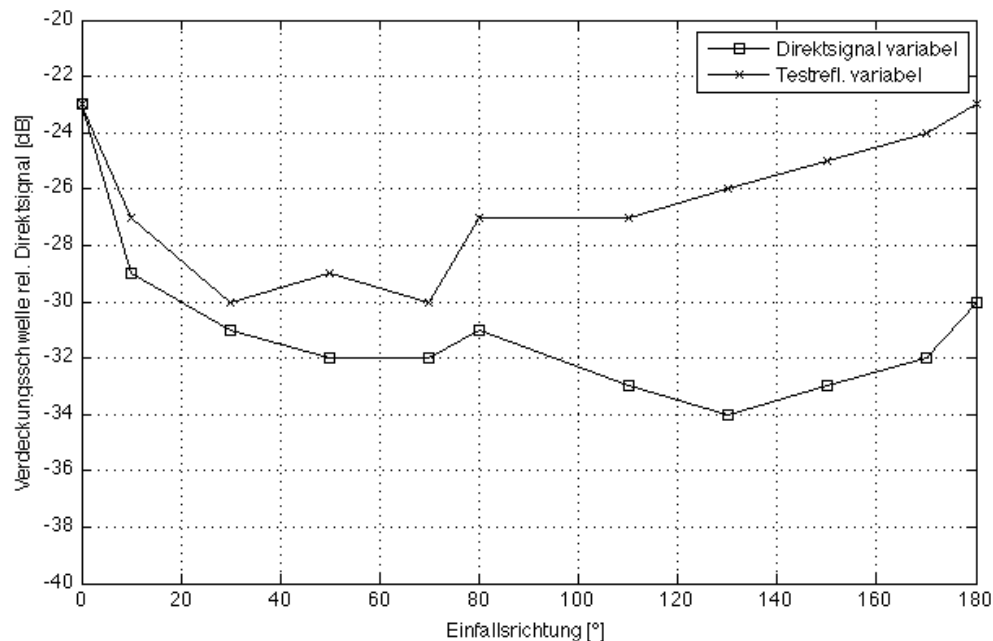


Abbildung 2.7: Verdeckungsschwelle als Funktion von Einfallswinkel bei fester Verzögerung von 50 ms mit dem Stimulus ‘Zack’ aus [14], Fig.22

nimmt bei zunehmendem Auseinanderlaufen der beiden Schallquellen ab. Zwischen  $30^\circ$  und  $70^\circ$  hat sie ein Minimum und steigt dann langsam wieder an bis zu den Winkeln, die einem rückwärtigen Schalleinfall entsprechen. Aber auch beim rückwärtigen Schalleinfall liegt die Verdeckungsschwelle unterhalb des  $0^\circ$ -Einfalls [14]. Man könnte hier vermuten, dass binaurale Mechanismen die Empfindlichkeit der monauralen Echoschwelle verstärken, da bei seitlichem Einfall die Testreflexion lauter wird und dadurch der Schwellwert sinkt. Für den Fall des seitlichen Einfalls des Verdeckers, also dem Direktsignal ist der Effekt noch größer, da der Schwellwert noch weiter sinkt.

Zusammenfassend kann man aus den vorgestellten Ergebnissen festhalten, dass kurze Impulse kritische Signale für die Detektion von Reflexion nach relativ kurzer Verzögerungszeit darstellen und binaurale Mechanismen relevant für die Detektion von Reflexionen aus verschiedenen Richtungen sind. Die Ergebnisse der hier vorgestellten Untersuchungen können benutzt werden, um perzeptiv motivierte Entscheidungen durch ein auditorisches Modell zu überprüfen. Werden sich Ergebnisse aus einem Modell ähnlich der Verdeckungskurven für die verschiedenen Fälle verhalten, kann man zunächst davon ausgehen, dass das Modell

plausible Hörbarkeitsentscheidungen über einzelne Raumschallfeldkomponenten trifft. Die Wirkungsweise des im Verlauf dieser Arbeit beschriebenen Modells wird u.a. anhand der Ergebnisse aus Abb. 2.5 und 2.7 untersucht.

## 2.3 Auditorisch motivierte Algorithmen zur Reduktion von Raumschallfeldern

In diesem Kapitel werden Modelle aus der Forschungsliteratur zusammengefasst, die eine perzeptive Reduktion der physikalischen Komplexität von Raumschallfeldern zum Ziel haben. Alle Modelle gehen stets von der systemtheoretischen Überlegung aus, dass Raumschallfelder als Impulsantworten in numerischer Form beschrieben werden können. Die RIRs hat man in den Modellen auf unterschiedliche Art und Weise erhalten. Wie in Kapitel 2.1 angedeutet, wurden RIRs in real existierenden Räumen mit Lautsprechern und Mikrofonarrays gemessen oder in Computermodellen berechnet. Weiterhin basieren die Modelle auf dem perzeptiven Ansatz, dass ein menschlicher Hörer als Empfänger nicht alle detaillierten Informationen, die in der RIR enthalten sind, wahrnimmt.

### 2.3.1 Auditorische Parametrisierung realer Schallfelder nach Hulsebos et al. [1]

In [1] werden Mikrofonarraymessungen von realen Raumschallfeldern in ebene Wellen zerlegt. Mittels einer vereinfachenden auditorischen Parameterdarstellung können die ebenen Wellen rekonstruiert werden. Dabei wird zunächst eine gemessene Raumimpulsantwort in eine auditorische Filterbank aufgeteilt, wie es bei der Frequenzanalyse auf der Basilar-membran im auditorischen System geschieht. Die auditorischen Filter können durch sog. Pattersonfilter, siehe z.B. [20] (S. 72 ff) angenähert werden. Ähnlich der Frequenzauflösung ist auch die zeitliche Auflösung des auditorischen Systems beschränkt. Dies wurde in Untersuchungen zu Vor- und Nachverdeckung in [18] gezeigt. Die zeitliche Auflösung kann durch eine laufende Mittelung innerhalb eines frequenzabhängig gewählten Zeitfensters beschrieben werden. Aus den Experimenten in [18] ist deutlich geworden, dass sich die Fensterbreiten bei tiefen Frequenzen wesentlich vergrößern, während sie oberhalb von 900 Hz konstant bleiben. In [1] wurden auditorische Filterbänder für Mittenfrequenzen aus der Tabelle 2.3.1 berechnet. Ausserdem sind in der Tabelle interpolierte Längen verwendeter Hanningfenster, die für die Mittelung entsprechender Bänder benutzt wurden, gezeigt. In [1] wurden 38 Filterbänder verwendet, die den hörbaren Frequenzbereich von 20 Hz bis 20 kHz gleichmäßig abdecken.

Anschliessend wurden die Raumimpulsantworten mit weissem Rauschen rekonstruiert. Das Rauschen wurde in entsprechende Filterbänder aufgeteilt. Hörbare Variationen der bandbegrenzten Rauschsignale in den Filterbändern wurden kompensiert. In vorläufigen Hörtests konnte festgestellt werden, dass die Methode der Rekonstruktion von Raumimpulsantworten für frühe Reflexionen gut funktioniert. Das Direktsignal wurde dabei nicht in der genannten Form parametrisiert, da sehr schnelle Einschwingphasen, wie sie bei manchen Signalen auftreten, durch den Rekonstruktionsvorgang verschmiert werden und diese Verschmierung hörbar ist. In Abb. 2.8 werden die originale Raumimpulsantwort und die parametrisierte Version gezeigt.

f [Hz]	t [ms]	f [Hz]	t [ms]	f [Hz]	t [ms]
20	35	807	8.5	4013	8.0
52	35	921	8.0	4555	8.0
86	35	1048	8.0	5186	8.0
125	35	1188	8.0	5924	8.0
168	35	1345	8.0	6797	8.0
214	33.2	1520	8.0	7839	8.0
265	21.7	1716	8.0	9103	8.0
322	15	1935	8.0	10660	8.0
384	13.8	2182	8.0	12610	8.0
453	12.7	2461	8.0	15120	8.0
528	11.7	2776	8.0	18450	8.0
612	10.6	3134	8.0	23063	8.0
705	9.6	3543	8.0		

Tabelle 2.1: Frequenzbänder und zeitliche Integrationslängen aus [1], S.8, Table 2

Das Direktsignal in Abb. 2.8b kommt aus der originalen RIR. Nur die frühen Reflexionen und der späte Nachhall wurden parametrisiert und rekonstruiert.

### 2.3.2 Schwellenwertansatz nach Begault et al. [2]

In [2] wurden durch Hörversuche absolute Wahrnehmungsschwellen<sup>3</sup> für virtuelle akustische Reflexionen als Funktion von räumlicher Position, Zeitverzögerung und Stimulustyp ermittelt. Reflexionen 1.Ordnung wurden über ein Ray-Tracing Modell des verwendeten Hörerraums berechnet. Aus den Ergebnissen wurde in [2] eine einfache Regel zur Hörbarkeit individueller Reflexionen niedriger Ordnungen aus berechneten Raumschallfeldern formuliert. Die Regel sagt, dass frühe Reflexionen nicht mehr hörbar sind, wenn sie 22 dB unterhalb des Direktsignalpegels mit 3 ms Verzögerung bzw. 31 dB unterhalb des Direktsignalpegels bei 15-30 ms Verzögerung auftreten.

### 2.3.3 Geometrischer Ansatz nach Hacıhabiboğlu und Murtagh [3]

In [3] wurde ein psychoakustisches Modell entwickelt, welches eine perzeptiv basierte Auswahl früher Reflexionen trifft, die mit dem Spiegelschallquellenmodell berechnet wurden. Die Auswahl basiert im wesentlichen auf drei psychoakustischen Effekten. Aus dem Gesetz der ersten Wellenfront, der Summenlokalisierung und der Echschwelle<sup>4</sup> wurde eine Funktion

<sup>3</sup>engl. ‘auditory threshold’

<sup>4</sup>respektive engl. ‘law of first wave front’ oder auch ‘precedence effect’, ‘summing localization’, ‘echo threshold’

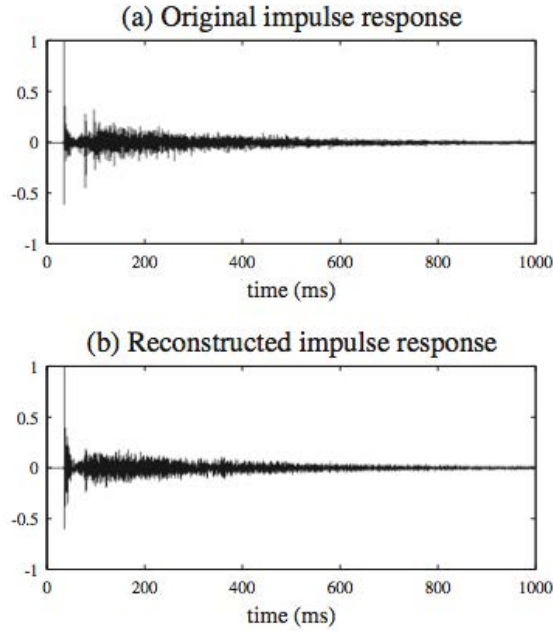


Abbildung 2.8: (a) Originale RIR, (b)rekonstruierte RIR aus [1], S.8, Fig.14

hergeleitet, die einzelne Reflexionen als individuelle Quellen aus dem Spiegelschallquellenmodell untersucht.

Das Gesetz der ersten Wellenfront beschreibt den Effekt, dass wenn zwei Signale mit einer Zeitverzögerung oberhalb eines unteren Schwellenwerts  $\tau_{low}$  am Ohr des Hörers eintreffen, die Position der Quelle aus der Richtung des zuerst eintreffenden Signals wahrgenommen wird. Spätere Signale werden im Interpretationsprozess zur Lokalisation unterdrückt. Bei Impulsen oder breitbandigen Noise-bursts ist  $\tau_{low} \approx 1ms$ .

Ist das Interval der Ankunftszeit von zwei Signalen  $\leq \tau_{low}$ , wird die Quelle aus dem Mittelwert der beiden Signale lokalisiert. Dieser Effekt wird Summenlokalisierung genannt. Siehe hierzu auch [19], S.204 ff. Ist die Zeitverzögerung von zwei Signalen oberhalb eines oberen Schwellenwerts  $\tau_{high}$ , werden sie als separate Klangereignisse wahrgenommen. Bei Impulsen oder breitbandigen Noise-bursts ist  $\tau_{high} \approx 5ms$ . Der Schwellenwert  $\tau_{high}$  wird als Echoschwelle bezeichnet.

Die Funktion ist bei Signalpaaren auf der horizontalen Ebene für eine Unterscheidungs- und Unterdrückungs-Bedingung definiert. Liegt eine verzögerte Reflexion oberhalb  $\tau_{low}$  und unterhalb  $\tau_{high}$ , wird sie den räumlichen Eindruck nicht mit beeinflussen, andernfalls wird sie als Spiegelschallquelle für eine Auralisation mitberechnet. Daraus entsteht ein zeitliches Cluster von hörbaren Reflexionen. Das Modell wurde um ein räumliches Cluster erweitert. Die Spiegelquellen aus dem zeitlichen Cluster werden dabei durch ihre Position, definiert aus Entfernung und Winkeln in Azimut und Elevation zum Hörer, räumlich zusammengefasst. Die Anzahl der Spiegelschallquellen lässt sich durch die vorgeschlagene Methode quantitativ erheblich reduzieren.

Ein durch das Reduktionsmodell berechneter binauraler RIR (BRIR)-Datensatz wurde qualitativ durch eine Auralisation mit originalen und mit der Regel aus [2] reduzierten BRIRs verglichen. In einem ersten Hörversuch wurde die Lokalisationsfähigkeit von virtuellen Quellen für die drei Methoden überprüft. Als Stimuli dienten gefensterte breitbandige Rauschpulse. In einem weiteren Hörversuch wurde die Wahrnehmung von räumlicher Qualität auralisierter Daten untersucht. Dabei wurden die Versuchsperson aufgefordert, drei räumliche Merkmale zu beurteilen. Die Merkmale waren: Präsenz (d.h. wie real scheint die Quelle), Räumlichkeit (d.h. die Breite der Schallquelle) und Umhüllung (d.h. wie sehr umgibt einen Hörer die Quelle)<sup>5</sup>. Zur Analyse der Ergebnisse wird auf die Arbeit [3] verwiesen. Zusammenfassend konnte festgestellt werden, dass in dem durch das vorgeschlagene Reduktionsmodell berechnete Schallfeld sowohl die Lokalisation von den virtuellen Quellen gut funktioniert, als auch kein signifikanter Verlust der wahrgenommenen räumlichen Qualität, beschrieben durch die beurteilten Merkmale, hervorgerufen wird. Durch das Modell wird die Anzahl der frühen Reflexionen allerdings um ca. 30% reduziert. Bei der Evaluation des Reduktionsmodells sind allerdings methodische Mängel festzustellen. Zunächst wurden die mit den unterschiedlichen Methoden berechneten Schallfelder nicht in einem A/B-Vergleich bewertet. Somit wurde die Detektierbarkeit von Wahrnehmungsunterschieden nicht direkt untersucht. Ausserdem scheint die Aussage über Qualität beschrieben durch die beurteilten Merkmale etwas willkürlich, so dass die Unhörbarkeit der manipulierten Schallfelder gegenüber vollaufgelösten Schallfeldern in Frage zu stellen ist.

### 2.3.4 Auditorisches Verdeckungsmodell nach Buchholz et al. [4]

In dem Modell von [4] wird die Hörbarkeit von Reflexionen durch den Vergleich mit dem sog. “Reflection Masked Threshold” (RMT) ermittelt. Auf Grundlage u.a. in Kapitel 2.2 genannter empirischer Daten wurde in [4] eine Funktion für den RMT formuliert, in der relevante Beziehungen für die Beschreibung der Wahrnehmung von Reflexionen in Raumschallfeldern enthalten sind. In der Funktion werden sowohl lineare Abhängigkeiten aus der Kombination von räumlichen, zeitlichen, pegelabhängigen und spektralen Parametern, als auch nicht lineare (signalabhängige) Parameter beschrieben.

Für die Entwicklung eines sog. psychoakustischen Modells werden sowohl diese Parameter, als auch die bekannte Physiologie des auditorischen Systems berücksichtigt. Durch das Modell wird die effektive auditorische Signalverarbeitung des menschlichen Hörsystems nachempfunden. Das vorgeschlagene Modell (“Room Reflection Model” (RMM)) besteht aus den in Abb. 2.9 gezeigten Modulen.

Das “directivity module“ beschreibt die frequenzabhängige Richtwirkung der Aussenohren, welche durch lineare Filter bekannt als Kopfübertragungsfunktionen (HRTF<sup>6</sup>) modelliert werden. Durch dieses Modul wird der Großteil der Richtungsabhängigkeiten des RMT vorausgesagt. Effekte, wie sie durch den Ohrkanal oder das Mittelohr hervorgerufen

---

<sup>5</sup>respektive engl. ‘Presence’, ‘Spaciousness’, ‘Envelopment’

<sup>6</sup>engl. head related transfer function

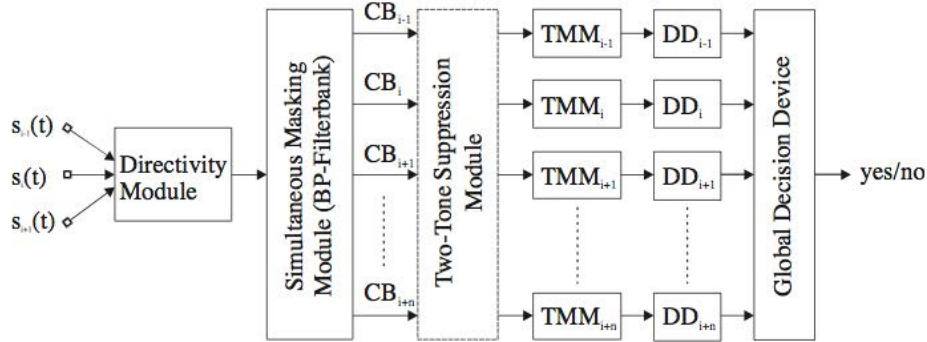


Abbildung 2.9: Globale Struktur des Modells von [4], S.3, Fig.2

werden, wurden vernachlässigt, da sie nicht richtungsabhängig sind.

Die Frequenzanalyse im auditorischen System führt zu einer begrenzten Frequenzauflösung menschlichen Hörens, siehe hierzu auch [20], S.65 ff. In dem “Simultaneous Masking Module“ wird die Frequenzanalyse des auditorischen Systems durch eine auditorische Filterbank simuliert. Es wird eine Gammatone-Filterbank mit 65 Frequenzbändern, verteilt über einen Frequenzbereich von 100 Hz bis 8 kHz u.a. aus [21] verwendet.

Das “Two-Tone Suppression-Module“ bewirkt vor allem eine Reduzierung der Nachverdeckung durch einen schmalbandigen Maskierer, wenn ein Signal mit anderer Frequenz vorliegt. Dieser Effekt wird u.a. in [20] genauer beschrieben.

In dem “Temporal Masking Module“ (TMM) wird die nicht lineare Funktionsweise der Basilarmembran durch eine statische nicht lineare Funktion, wie es in [18] beschrieben wurde, simuliert. Ausserdem werden hier Halbweggleichrichtung und Tiefpassfilterung integriert, wie sie in [22] zur Modellierung der Haarzellen auf dem Cortschen Organ im Innenohr des Menschen vorgeschlagen werden. Weiterhin wird hier die Notwendigkeit einer Funktion zur signalabhängigen Kompression genannt. Die Funktion wird in [23, 24] weiter ausgeführt. Die Funktion beeinflusst kaum schnelle Änderungen von Eingangssignalen, stationäre Signale werden jedoch stark unterdrückt. Das TMM simuliert die in dem RMT genannten Abhängigkeiten und generiert eine angemessene Signaldarstellung für die folgenden Entscheidungsapparate.

In den Entscheidungsapparaten “DD (Decision Device)“ und “Global Decision Device“ werden zunächst Hörbarkeitsentscheidungen pro Filterband getroffen und später eine globale Entscheidung für die gesamte Reflexion. Im “DD“-Modul wird zusätzlich zwischen simultaner Verdeckung und nicht-simultaner Verdeckung unterschieden. Bei simultaner Verdeckung ist das Signal-Rausch-Verhältnis<sup>7</sup>, als Verhältnis von Testsignal (Reflexion) zu Verdeckter (Direktsignal plus zusätzliche Reflexionen) maßgeblich für die Entscheidung. Bei nicht-simultaner Verdeckung ist die Amplitude des Testsignals (Reflexion) der entscheidende Faktor. Nur wenn beide Entscheidungen positiv (d.h. Reflexionen sind hörbar, wenn sie oberhalb des RMT liegen) ausfallen, wird die Reflexion pro Filterband als hörbar markiert.

<sup>7</sup>engl. signal-to-noise ratio



Das vorgeschlagene Modell wurde auf einzelne Raumreflexionen angewandt und mit empirischen Daten verglichen. Als Direktsignal wurde immer ein “Noise-burst” mit einer Länge von 50 ms und einem Schalldruckpegel von 75 dB(SPL) verwendet. Sieht man sich z.B. die Richtungsabhängigkeit der Wahrnehmungsschwelle an, siehe Abb. 2.7, werden empirische Daten sehr gut durch das Modell abgebildet. Die Übereinstimmung ist in Abb. 2.10 zu sehen. Für die Messung wurde einmal die Einfallsrichtung des Direktsignals konstant von

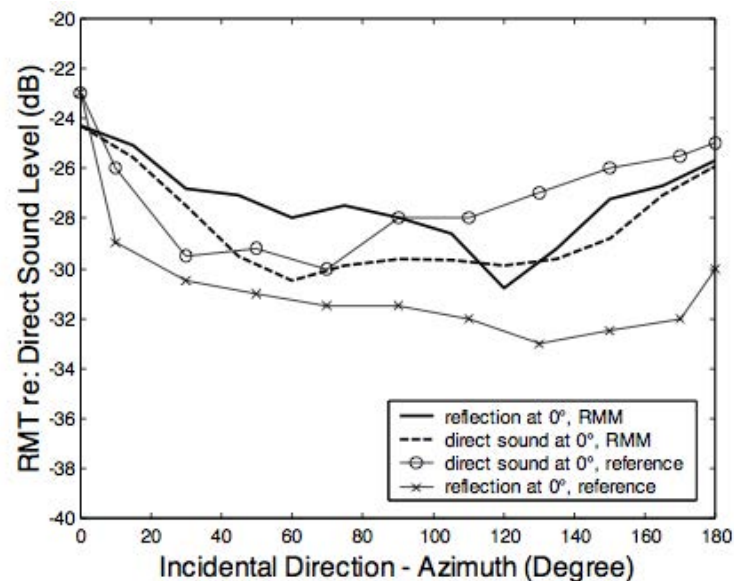


Abbildung 2.10: Verdeckungsschwellen in Abhängigkeit von Einfallswinkel empirischer Daten aus [14], Fig.22 und dem Modell von [4], S.6, Fig.8

vorne und die um 50 ms verzögerte Testreflexion zwischen  $0^\circ$  und  $180^\circ$  variiert und in dem zweiten Fall das Direktsignal variiert, während die Testreflexion aus frontaler Richtung kam. Sowohl in den empirischen Kurven, als auch durch das Modell kann der größte Verdeckungseffekt für gleiche Einfallsrichtung von Direktsignal und Testreflexion beobachtet werden. Auch andere Abhängigkeiten, wie z.B. die Verzögerungszeiten von Testreflexionen für verschiedene Direktsignalpegel oder Länge des Direktsignals konnten gut durch das Modell vorausgesagt werden.

Wie im Anwendungsfall dieser Arbeit wurde das Modell auf eine simulierte RIR (siehe Abb.2.11 (links)) angewandt. In Abb. 2.11 (rechts) wird der RMT in dB SPL der Testreflexion, die in der RIR enthalten ist, dargestellt. Die schwarze Kurve zeigt den RMT für die Reflexion durch die nicht-simultane Verdeckung. Die gestrichelte Kurve zeigt den RMT für simultane Verdeckung. Der RMT wird für Verzögerungen  $\leq 60$  ms hauptsächlich durch die Hörbarkeitsentscheidung der simultanen Verdeckung beeinflusst. Zwischen ca. 60 - 100 ms ist der Einfluss von simultaner und nicht-simultaner Verdeckung auf den RMT in etwa gleich. Danach dominiert die nicht-simultane Verdeckung den RMT. Wird der RMT hauptsächlich durch die Entscheidung des nicht-simultanen DD bestimmt, scheint

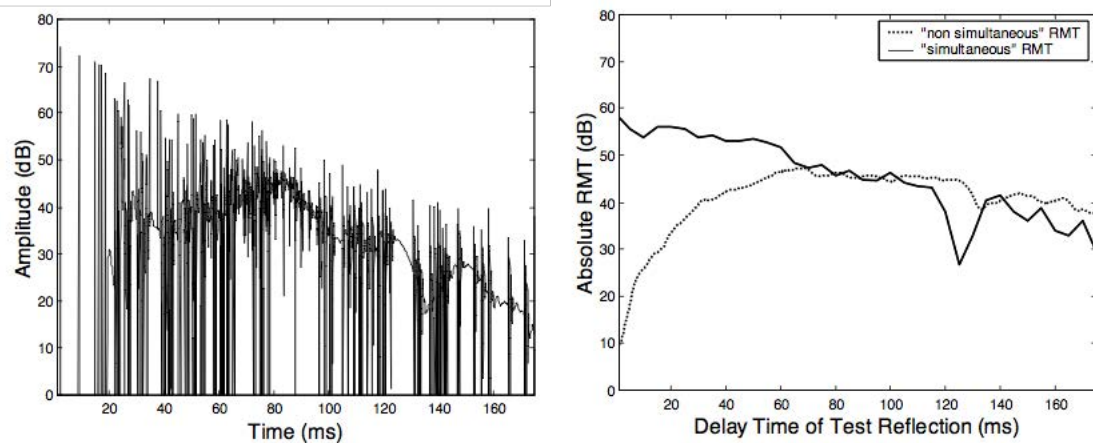


Abbildung 2.11: links: simulierte RIR eines Schuhschachtelraums, rechts: RMT Voraussage aus simultanem und nicht-simultanem DD für die simulierte RIR aus [4], S.6, Fig.10 und 11

die Anwendung eines monauralen Modells ausreichend und interaurale Prozesse können ignoriert werden. Bei simultaner Verdeckung müssen interaurale Prozesse im Modell berücksichtigt werden. Nachverdeckungseffekte spielen meist bei nicht-simultanen Vorgängen eine Rolle, daher ist die Entwicklung eines monauralen Modells ausreichend, um den RMT in einer RIR zu bestimmen.

Auf Basis der gezeigten Evaluationsergebnisse erscheint der Ansatz von [4] am vielversprechendsten, um zu einem eigenen Modell zu gelangen.

# Kapitel 3

## Methodik

Methodisch gliedert sich dieser Teil der Arbeit so, dass zunächst die Messung binauraler Raumimpulsantworten eines Raumes zu Auralisationszwecken beschrieben wird. In einem zweiten Schritt wird dokumentiert, wie anhand eines Raumakustikmodells, entsprechend dem vermessenen Raum, eine diskrete Schallfelddarstellung berechnet und daraus Datensätze zur Auralisation generiert werden. Im dritten Teil wird die Entwicklung des Modells anhand existierender Modelle aus Forschungsliteratur erläutert und auf die berechneten Schallfelddarstellungen angewandt. Aus den daraus resultierenden reduzierten Schallfelddarstellungen wurden auralisierbare Datensätze erstellt.

## 3.1 Realer Raum

Die dynamische, binaurale Simulation der sog. Realität geschieht anhand gemessener binauraler RIRs (BRIRs). Für die (in-situ) Vermessung von BRIRs in realen Räumen wird der akustische Transmissionspfad zwischen einer Klangquelle und einem Empfänger in einem real existierenden Raum gemessen. In der BRIR sind sowohl die akustischen Eigenschaften des Raumes in der RIR, als auch die richtungsabhängigen Eigenschaften der HRTF eines Hörers enthalten.

Man benötigt dafür also einen Messapparat, der sowohl die Klangeigenschaften des Raumes ausreichend anregt, als auch die wahrnehmungsspezifischen Eigenschaften eines Hörers messen kann. Als Quelle wird dafür ein Lautsprecher mit omnidirektionaler Abstrahlcharakteristik verwendet. Als Empfänger wird ein Kopf- und Torsosimulator mit multiplen Freiheitsgraden eingesetzt.

### Dodekaeder-Lautsprecher

Messungen von BRIRs, die zum Zwecke der Auralisation genutzt werden sollen, können mit allseitig abstrahlenden Lautsprechern durchgeführt werden. Der Lautsprecher muss dabei genügend Signalenergie im Frequenzbereich von mindestens 50 Hz bis 12.5 kHz mit hinreichender Abstrahlung erzeugen. Ein solcher Lautsprecher wurde z.B. am Institut für Technische Akustik in Aachen entwickelt, siehe dazu [25]. Es handelt sich dabei um einen Dodekaeder-Lautsprecher, der die genannten Anforderungen erfüllt und darauf ausgelegt ist, variable Richtcharakteristiken für quellenspezifische Impulsantworten zu ermöglichen. Der Dodekaeder-Lautsprecher ist ein 3-Wege Messlautsprecher, bestehend aus Subwoofer, Mitteltonlautsprecher und Hochtöner, die übereinander angeordnet sind. Die omnidirektionale Abstrahlcharakteristik des Lautsprechers verhindert einen richtungsabhängigen Einfluss auf die Messung und ermöglicht eine BRIR-Messung für den allgemeinen Fall der Quellenunabhängigkeit. Vor allem dient die omnidirektionale Richtcharakteristik aber der besseren Modellierbarkeit von Quellen in Computermodellen, wie in EASE zur Berechnung von RIRs.

Ein spezifischer Frequenzgang des Lautsprechers würde die BRIR und damit die Auralisation beeinflussen. Daher wird es nötig sein, den Lautsprecher vorher zu entzerren, mit dem Ziel, ein weisses Spektrum im genannten Frequenzbereich zu erreichen. Der lineare und damit weitestgehend neutrale Frequenzgang kann über die Entzerrung des akustischen Schallleistungsspektrums nach [26] erreicht werden. Die Konstruktion des Leistungsspektrums und die Entzerrung wird im folgenden Kapitel genauer erläutert.

### Kopf- und Torsosimulator

Als Empfänger zur Messung von dynamischen BRIRs dient der am Fachbereich Audio-kommunikation entwickelte Kopf- und Torsosimulator (HATS<sup>1</sup>) FABIAN. In [27] werden die Entwicklung und Konstruktion des Simulators ausführlich beschrieben.

---

<sup>1</sup>engl. ‘head and torso simulator’

Das Messsystem kann idealerweise zur Akquise von kopfbezogenen BRIRs verwendet werden. Es besteht aus einem Kunstkopf und Außenohren als Gipsabguss des Kopfes von Dipl. Ing. C. Moldrzyk. Mit zunehmender Entfernung vom Außenohr nimmt der Einfluss von Körperteilen auf die frequenzspezifische Färbung der HRTFs ab. Die Nachbildung von Kopf & Torso ist demnach mit steigender Entfernung von der Pinna in zunehmenden Maße vereinfacht. Die Messung der BRIRs werden getrennt vom linken und rechten Ohr durchgeführt. Die Messmikrofone befinden sich jeweils am Eingang zum Ohrkanal<sup>2</sup>. Bei den Mikrofonen handelt es sich um vorpolarisierte Elektretkondensatormikrofone in Miniaturbausweise. Sie haben eine omnidirektionale Richtcharakteristik, entsprechend dem Trommelfell, welches als Druckempfänger ohne Vorzugsrichtung arbeitet, einen hinreichend glatten Frequenzgang und eine identische Empfindlichkeit an beiden Kanälen. Durch die Messung mit FABIAN erhält man eine BRIR als Paar von Impulsantworten, in der alle geforderten Beziehungen enthalten sind.

Das Kunstkopffinmesssystem hat ausserdem ein Halsgelenk mit zwei zeitgleich nutzbaren rotatorischen Freiheitsgraden (Drehung um die Hochachse und in der Medianebene). Dadurch können binaurale Impulsantworten physiologiegetreu in allen natürlichen Kopfhaltungen in einem sphärischen Raster von diskreten Messpunkten vermessen werden.

Für eine qualitativ hochwertige Simulation wurden die BRIR-Messungen mit dem automatisierbaren Kopf- und Torsosimulator FABIAN aus [27] im Audimax der TU Berlin durchgeführt. Als Schallquelle diente der 3-Wege Dodekaeder-Lautsprecher aus [25]. Damit kann ein hoher Signalrauschabstand und eine optimale omnidirektionale Richtcharakteristik erreicht werden.

### 3.1.1 Frequenzgangsentzerrung des Dodekaeder-Lautsprechers

Für ein breites Frequenzspektrum der BRIRs und einen minimalen spektralen Einfluss der Messquelle muss der beschriebene Dodekaeder-Lautsprecher entzerrt werden. Die Schallleistungsbestimmung und Frequenzgangentzerrung des Dodekaeder-Lautsprechers geschieht prinzipiell nach dem sogenannten Hallraumverfahren.

#### Aufbau:

Der Lautsprecher und ein Messmikrofon befinden sich für die Diffusfeldmessung im Hallraum des ISTA der TU Berlin. Für den Lautsprecher wird ein Endstufenrack mit Lautsprechercontroller, eine 4-Kanal-Endstufe und ein Messsystem mit XLR-Audioausgang und -eingang benötigt.

Der Signalfluss und die Verkabelung für die Messung sind in Abb. 3.1 zu sehen.

Das Messsignal wird im Messsystem Monkey Forest (MF) als gewichteter Sinussweep erzeugt und auf den Input 1 des Behringer DCX-2496 Controller geführt. Der Controller teilt das Mono-Messsignal in drei Einzelsignale und führt es anschließend bandbegrenzt

---

<sup>2</sup>am geblockten Ohrkanal, engl. ‘blocked ear canal’

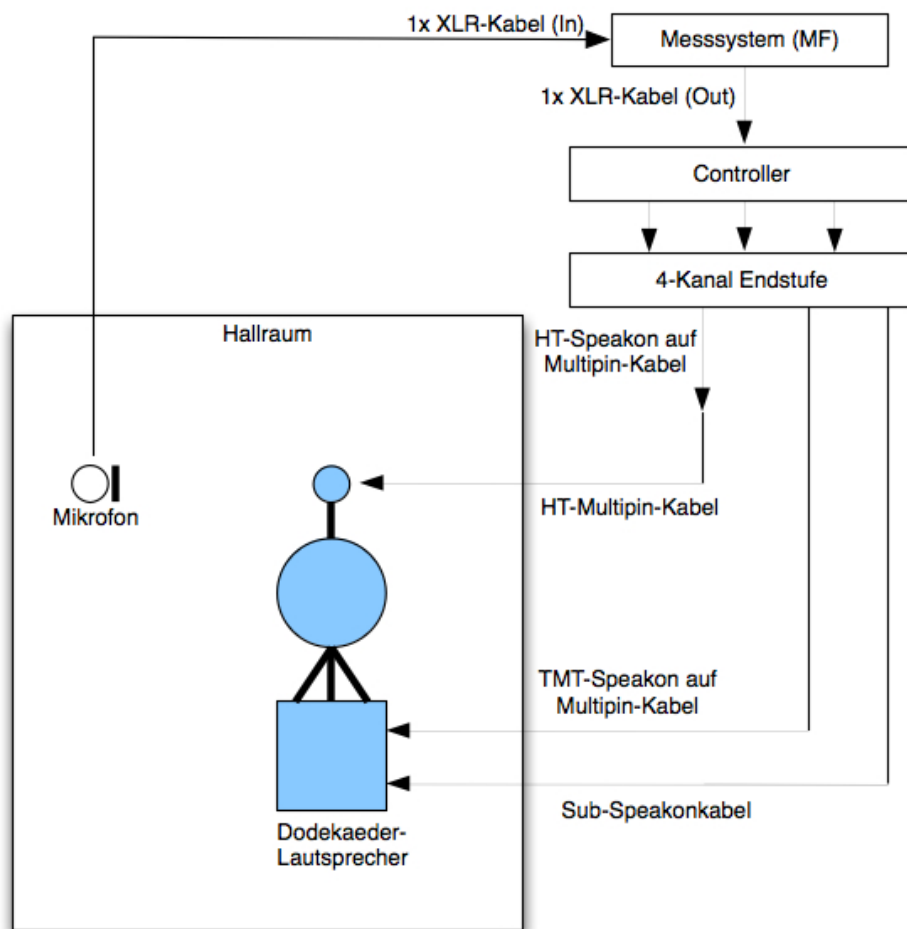


Abbildung 3.1: Signalfluss und Verkabelung für die Vermessung des Dodekaeder-Lautsprechers im Hallraum

auf die Ausgänge 1-3. Diese werden mit den Eingängen 1-3 der T.Amp D4-500 4-Kanal Endstufe verbunden, so dass man an Ausgang A den Subwoofer, an Ausgang B den Mitteltöner und an Ausgang C den Hochtöner betreiben kann.

### Durchführung und Ergebnisse:

Diffusfeldmessung:

Zunächst wird die RIR des Hallraums mit dem Lautsprecher gemessen. In dem Messsystem Monkey Forest werden die Nachhallzeiten aus der Impulsantwort in Oktavbändern berechnet. Die Nachhallzeiten sind der Tabelle 3.1 zu entnehmen. Die Nachhallzeiten ent-

Freq [Hz]	$T_{30}$ [sek]
31	6.46
63	8.15
125	9.41
250	8.33
500	7.27
1k	5.8
2k	3.98
4k	2.73
8k	1.16

Tabelle 3.1: Berechnete Nachhallzeiten in Oktavbändern aus RIR-Messung im Hallraum

sprechen in etwa den Raumangaben des Hallraumes am Institut der Technischen Akustik. Aus den Nachhallzeiten wird das Kompensationsspektrum manuell in MATLAB mit

$$T_{inv.rev} = 10 * \log_{10}\left(\frac{1}{T_{rev}}\right) \quad (3.1)$$

berechnet.

Dann werden die unentzerrten Übertragungsfunktionen getrennt für Hochtöner (HT) und Mitteltöner (MT) an 10 Messpositionen im Diffusfeld (Messmikrofonabstand ca. 3m vom Lautsprecher) gemessen. In dem Messsystem Monkey Forest wird die Betragsmittelung der gemessenen Übertragungsfunktionen berechnet.

Nahfeldmessung:

Zusätzlich kommt es zu der einmaligen Messung der Übertragungsfunktion vom Tieftöner (TT) des Dodekaeders im Nahfeld.

Der Amplituden-Frequenzgang der zwei Wege von den gemittelten Übertragungsfunktionen aus der Diffusfeldmessung (Abb. links) und des TTs aus der Nahfeldmessung (Abb. rechts) sind in Abb. 3.2 zu sehen.

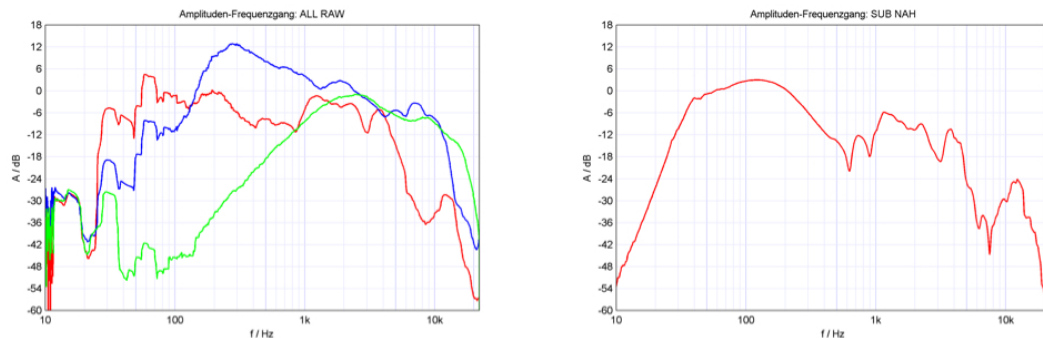


Abbildung 3.2: Rohmessung aller Wege im Diffusfeld (links,), Rohmessung Tieftone im Nahfeld (rechts)

Die Korrekturwerte aus dem Kompensationsspektrum werden in Monkey Forest übertragen. In Monkey Forest wird das Diffusfeld-Schalldruckspektrum mit dem Kompensationsspektrum multipliziert. Man konvertiert damit das Diffusfeld-Schalldruckspektrum in ein Spektrum, welches proportional zur akustischen Leistung ist.

Nun werden die Filter am Controller für die 3 Lautsprecherwege so entworfen, dass ein linearer Frequenzgang mit, soweit möglich, weissem Leistungsspektrum entsteht. Dies geschieht durch einen iterativen Prozess. In der Tabelle 3.2 sind die Einstellungen am Controller für die drei Wege zu sehen, die schließlich verwendet wurden, um den Frequenzgang linear zu entzerren. Das entzerrte Leistungsspektrum zeigt einen nicht linearen Verlauf im unteren Frequenzbereich mit Resonanzspitzen, wahrscheinlich durch Raummoden im Hallraum in dem Bereich. Die Probleme sind in Abb. 3.3 zu sehen.

Ersetzt man das tieffrequente Spektrum des TT mit dem Spektrum aus der Nahfeldmessung, erhält man einen gleichmäßigen Amplituden-Frequenzgang im gewünschten Arbeitsbereich. Der Frequenzgang wird nun in Monkey Forest über 1/3 Oktaven geglättet. Dies führt zu dem Amplituden-Frequenzgang, wie in Abb. 3.4 links zu sehen. Der lineare Frequenzgang zeigt, dass der Lautsprecher durch die Entzerrung einen Arbeitsbereich von 43 Hz bis 15.4 kHz mit  $\pm 2.5\text{dB}$  hat. Das wird aus der vergrößerten Darstellung in Abb. 3.4 rechts deutlich. Der Dodekaeder-Lautsprecher kann zusammen mit dem Controller und den genannten Filtereinstellungen als Quelle verwendet werden, um geeignete dynamische BRIRs in einem Raum zu vermessen.

### 3.1.2 Messung der binauralen Raumimpulsantwort

Die BRIR-Messung fand am 08.05.2010 im Audimax der TU-Berlin statt. Als erstes wurde der Ort für das Messsystem FABIAN bestimmt. Dazu wurden die Quellpositionen auf der Bühne festgelegt. Sie sollten mittig auf der Bühne als klassische Aufstellung eines Streichquartetts mit einer zusätzlichen Quelle in der Mitte der Anordnung (LS Vokal) positioniert werden. Bezüglich der Position der Centerquelle wurde der Sitzplatz von FABIAN fest-



	<b>Sub</b>	<b>MT</b>	<b>HT</b>
<b>Gain</b> in MF	-11	-14	-14
<b>Gain</b> am Controller	0	-2.9	8.5
<b>HP</b> (Butt) -3dB@	2.Ordnung 36kHz	4.Ordnung 220 Hz	4.Ordnung 1.1kHz
<b>LP</b> (Butt) -3dB@	4.Ordnung 200Hz	4.Ordnung 1.2kHz	-
<b>PEQ1</b> Gain [dB] Freq. [Hz] Q	-4.8 120 1	-11 280 1	-9.5 2.3k 1.2
<b>PEQ2</b> Gain [dB] Freq. [Hz] Q	-	1 440 1.6	-1.5 8.5k 3
<b>PEQ3</b> Gain [dB] Freq. [Hz] Q	-	-	6.5 17k 4

Tabelle 3.2: Filtereinstellungen am Controller

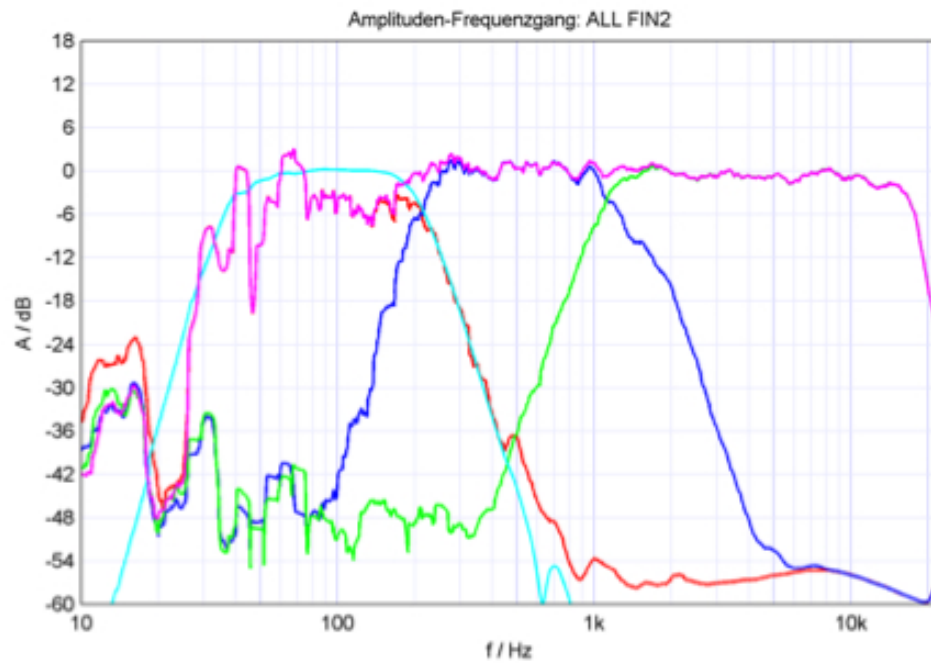


Abbildung 3.3: Finale Frequenzweichen-Filter mit nicht benutzter, aber angepasster Nahfeld-TT-Flanke

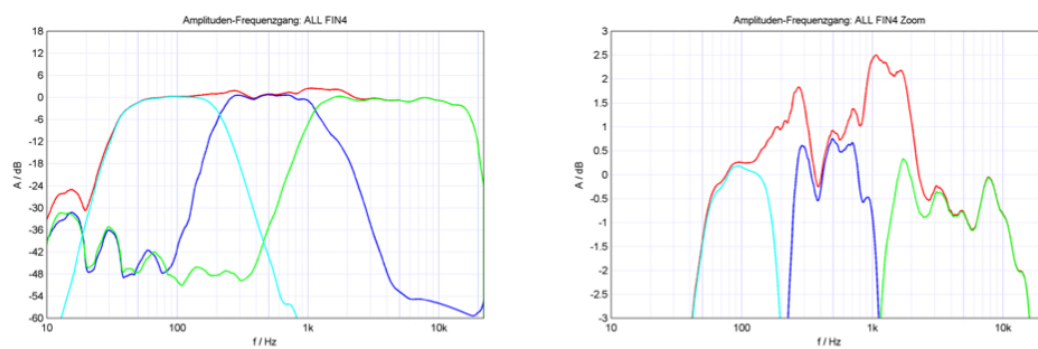


Abbildung 3.4: links: Finale Frequenzweichen-Filter mit benutzter und angepasster Nahfeld-TT-Flanke, 1/3-Oktav geglättet; rechts: Vergrößerte Darstellung der linken Abbildung

gelegt. Dabei sollte ein Abstandskriterium bezüglich des Direkt-/Diffusfeldverhältnisses eingehalten werden. Nach [13] kann man von einem diffusen Schallfeld ausgehen, wenn die Entfernung zwischen Quelle und Empfänger mindestens die doppelte Länge des sog. Hallradius beträgt. Der Hallradius kann mit der Nachhallzeit und dem Raumvolumen in der Form

$$r_c = 0.057 * \sqrt{\frac{V}{T}} \quad (3.2)$$

berechnet werden. Die statistische Nachhallzeit  $T$  wird mit der Sabine'schen Formel berechnet und das Raumvolumen wird in  $m^3$  gemessen [11]. Die statistische Nachhallzeit, siehe Gl. 3.4 und das Volumen des Raumes sind aus vorherigen Messungen in [28] bekannt. Sie betragen  $T_{gem} = 2.08s$  und  $V \approx 8500m^3$ , so dass sich der Hallradius zu  $r_c = 3.64m$  für das Audimax ergibt. Die FABIAN-Position muss also mindestens in  $2 \cdot r_c = 7.28m$  Entfernung längs auf der Symmetrieachse zur Centerquelle aufgestellt werden. Aufgrund der Sitzplatzanordnung im Hörsaal mussten die Quellen etwas von der Bühnenmitte verschoben werden, so dass sich ein Messaufbau, wie in Abb. 3.5 ergibt.

Für die Berechnung der Nachhallzeit des Raumes wurden zunächst drei standardmäßige

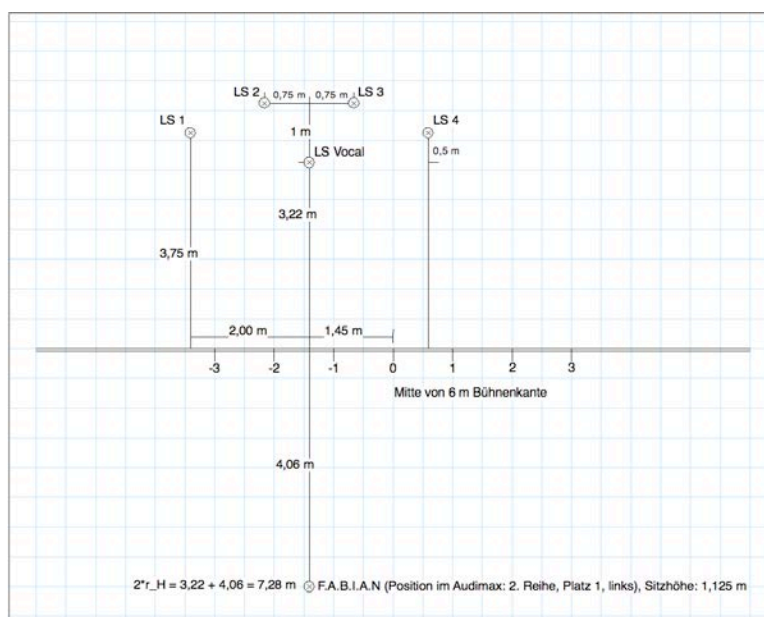


Abbildung 3.5: Bühnensituation für BRIR-Messung

RIR-Messungen mit einem omnidirektionalen Mikrofon für eine Quelle im diffusen Schallfeld durchgeführt. Die Messpositionen waren am Ort des Messsystems FABIAN und zwei Außenpositionen in ausreichender Entfernung zur Quelle im Saal. An der Messposition am Ort des Messsystems wurden die RIR-Messungen zusätzlich für alle 5 Quellen durchgeführt. Das Messsignal wurde mit dem Messsystem als gewichteter Sinussweep von 20 Hz bis 20 kHz erzeugt. Die Nachhallzeit des Raumes ergibt sich aus den über Oktavbändern (von 125 Hz bis 4kHz) und drei Messpositionen für die Centerquelle gemittelten

RIR-Messungen. In Abb. 3.6 sind die Nachhallzeiten in Terzbändern aus [28] zu sehen. Die in Terzbändern aufgelöste Nachhallzeit wird im nächsten Kapitel als Zielkurve zur

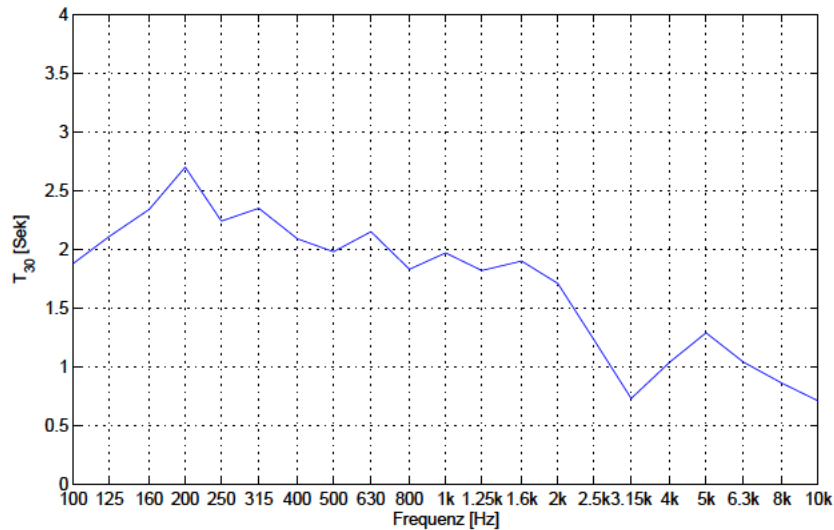


Abbildung 3.6: Gemessene Nachhallzeit vom Audimax aus gemittelten RIR Messungen der Centerquelle

Anpassung des Raummodells an die Realität verwendet.

In einem zweiten Schritt wurde der HATS an der Messposition wie in Abb. 3.5 aufgebaut. Dynamische BRIRs wurden  $\pm 80^\circ$  mit einer Schrittweite von  $1^\circ$  horizontal gemessen. In [29] wurde eine Auflösung von  $5^\circ$ -Schritten horizontal für 95% der Hörer als ausreichend plausible binaurale Simulationen ermittelt. Die verwendete Auflösung von  $1^\circ$ -Schritten ist demnach mehr als ausreichend. Die dynamische binaurale Simulation soll hier nur in der Horizontalebene durchgeführt werden. Somit war eine Messung mit vertikaler Orientierung des Messsystems nicht nötig. Durch die Auflösung (5x161 BRIR-Messungen) ergab sich eine Messdauer von ca. 75 Minuten für alle Quellen. Nachdem die BRIRs in dem Auflösungsraaster vermessen wurden, musste die Quelle an die nächste Position umgesetzt werden, und die Messung wurde wiederholt. Unter Berücksichtigung der Dateinamenskonvention werden die einer Quelle zugeordneten BRIRs jeweils in einem Ordner gespeichert. Sie können anschließend für das Postprocessing, wie z.B. die Kompensation der Übertragungsfunktion des Kopfhörers [30] oder die individuelle ITD-Korrektur [31] verwendet werden. Nicht zuletzt kann der in-situ gemessene BRIR-Datensatz zur dynamischen binauralen Reproduktion [27] über Kopfhörer als Simulation der Realität benutzt werden.

## 3.2 Modellierter Raum

Die Bezeichnung ‘Modellierter Raum’ weist auf die akustische Simulation eines Raumes, in diesem Fall dem Audimax der TU-Berlin, durch die Reproduktion eines Schallfeldes aus modellgenerierten Daten hin. Das aus der Raumakustiksimulation hervorgegangene Schallfeld wird aber nicht nur zur Reproduktion des vollaufgelösten Schallfeldes verwendet, sondern dient in seiner Form, als diskrete Beschreibung von Reflexionen, der Reduktion des Schallfeldes mit einem auditorischen Modell auf perzeptiv relevante Bestandteile. Mit der Raumakustiksimulations-Software EASE können geeignete BRIRs zur Auralisation erzeugt werden. Dabei werden die Raumübertragungsfunktionen (RTFs - RIRs im Zeitbereich) in Form von Reflektogrammen berechnet und mit HRTFs multipliziert, sodass richtungsabhängigen Informationen beim menschlichen Hörer Rechnung getragen werden können. Man erhält daraus die BRIR, welche sich schliesslich aus Direktsignal, frühen Reflexionen und spätem, diffusem Nachhall zusammensetzt. Im Frequenzbereich kann der Prozess durch die Multiplikation der Übertragungs- und Filterfunktion als Repräsentation von Klangübertragung von Quelle zum Empfänger in der Form:

$$H_j|_{left,right} = \frac{e^{-j\omega t_j}}{ct_j} \cdot H_{source}(\theta, \phi) \cdot H_{air} \cdot HRTF(\vartheta, \varphi)|_{left,right} \cdot \prod_{i=1}^{n_j} R_i \quad (3.3)$$

beschrieben werden [12] (Gl.11.50). Wobei  $H_j$  das Spektrum der  $j$ -Reflexion beschreibt.  $t_j$  ist die dazugehörige Zeitverzögerung,  $j\omega t_j$  die Phase,  $1/(ct_j)$  das Abstandsgesetz von sphärischen Wellen.  $H_{source}$  beschreibt die Richtcharakteristik der Quelle innerhalb der Quellkoordinaten,  $H_{air}$  beschreibt die Luftabsorption in Form eines Tiefpasses,  $R_i$  beschreibt den Reflexionsgrad der Wände und  $HRTF$  ist die Übertragungsfunktion des eintreffenden Schalls, zugehörig zu den Empfängerkoordinaten mit einer definierten Orientierung.

Ein CAD-Raummodell des Audimax der TU-Berlin wurde freundlicherweise zur Verfügung gestellt. Wie im Kapitel 2.1 beschrieben, wird die Software EASE in der Version 4.3.8 benutzt, um mit dem Modell eine RIR für entsprechende Quell- und Empfängerkombinationen zu berechnen. Das originale Raummodell des Audimax ist in Abb. 3.7 dargestellt<sup>3</sup>. Durch die blickrichtungsabhängige Gewichtung berechneter Reflektogramme mit HRTF Daten<sup>4</sup> erhält man einen BRIR-Datensatz, der zur dynamischen Auralisation verwendet werden kann. Damit ein perzeptiver Vergleich zur simulierten Realität möglich wird, muss das Modell an die Realität angepasst werden. Dies geschieht durch das Angleichen der Nachhallzeit im Modell an die Nachhallzeit, wie sie aus der Messung hervorgegangen ist. Die Nachhallzeit ist eine der wichtigsten Größen in der Raumakustik und soll hier als Parameter für das Fitting ausreichen, wie es auch in der Praxis üblich ist.

<sup>3</sup>Das Modell ist auf der beigegeführten DVD im soundfield-Repository - 2 EASE soundfield creation - 2 EASEmodel - scattering 060percent correctedVolume - tu audimax ohne-stühle correctedVolume.zip zu finden

<sup>4</sup>Daten können aus Messungen mit einem Kopf- und Torsosimulator im schalltoten Raum (im Freifeld) generiert werden

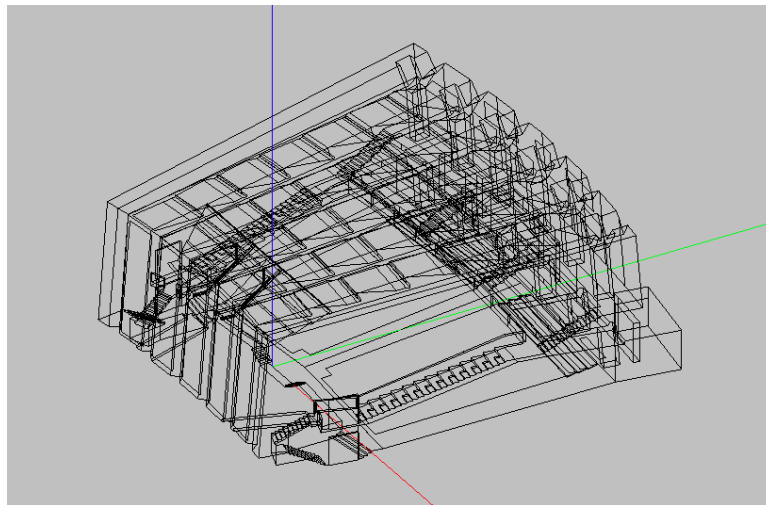


Abbildung 3.7: Das originale Raummodell des Audimax der TU-Berlin als Drahtgitterdarstellung aus EASE

### 3.2.1 Modellanpassung - 'Fitting' - des Raumakustikmodells an die Realität

Das vorliegende Modell des Audimax der TU-Berlin wird zunächst an die Messsituation, wie in Abb. 3.5 angepasst. Um den Anpassungsprozess nicht unnötig zu verkomplizieren, wird das sog. Fitting nur für die Situation einer Quell-Empfänger-Kombination (LS Vocal) durchgeführt. Das angepasste Modell soll später dazu verwendet werden, um für die Auralisation geeignete RIRs zu berechnen. Daher wird ein Empfänger<sup>5</sup> an die FABIAN-Position aus der Messung gesetzt, und alle fünf Quellpositionen werden als Kugelschallquellen<sup>6</sup> entsprechend ihrer Messpositionen eingefügt. Die Elemente, wie sie in das EASE Projekt eingefügt wurden, sind in Abb. 3.8 dargestellt

Beim Einsetzen der Quellen und des Empfängers muss auf die Orientierung des Empfängers zu den Quellen im Koordinatensystem von EASE geachtet werden. Es wurde der Nullpunkt des Raumkoordinatensystems in dem Raummodell auf die Mitte der vorderen Bühnenkante gelegt. Die Empfänger-Position befindet sich im positiven Bereich der Y-Achse mit Blick auf den Ursprung. Die 5 Quellen werden mit Ausrichtung zum Empfänger eingefügt, so dass sie sich auf der Y-Achse im negativen Bereich befinden und die horizontale Ausrichtung um  $180^\circ$  gedreht ist. Wie im weiteren Verlauf der Arbeit ausgeführt, wird die berechnete RIR mit einem HRTF-Datensatz gefaltet, so dass der Richtungsselektivität der Ohren Rechnung getragen werden kann. Dazu müssen die Koordinatenangaben für einzelne Reflexionen in Übereinstimmung mit entsprechenden HRTF-Paaren vorliegen. In EASE wird der horizontale Winkelbereich von  $-180^\circ$  (rechte Hemisphere) bis  $180^\circ$  (linke Hemisphere) und der vertikale Winkelbereich von  $-90^\circ$  (Südpol) bis  $90^\circ$  (Nordpol) in dem

<sup>5</sup>in EASE als 'Listener Seat'

<sup>6</sup>in EASE als Loudspeaker mit Speaker Model: SPHERE

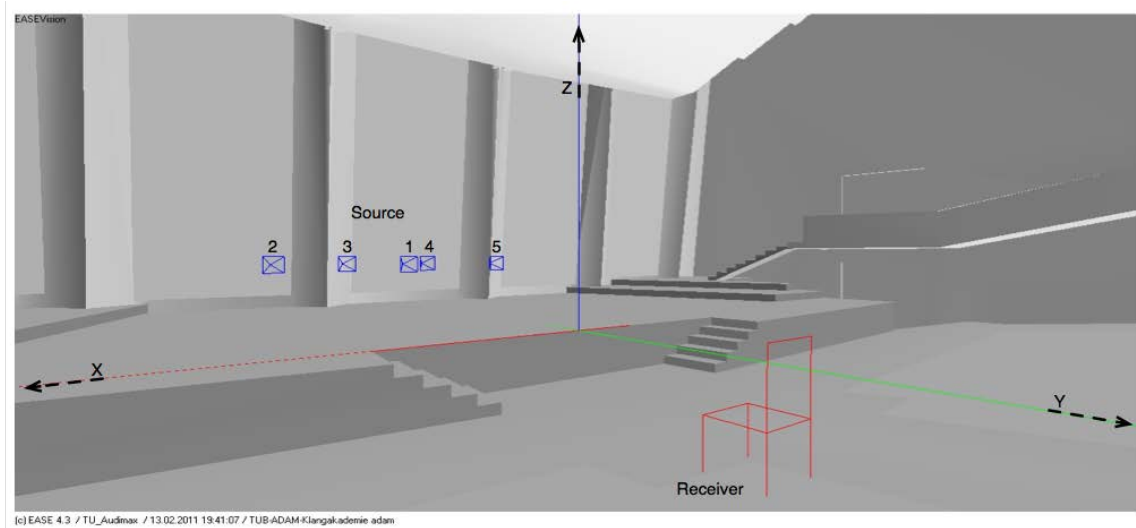


Abbildung 3.8: An die reale Messsituation angepasstes EASE-Raummodell

sog. ‘linkshändigen’ System definiert. Der später verwendete HRTF-Datensatz entstammt aus Freifeld-Messungen aus [32]. Hier wird der Winkelbereich in horizontaler Ausrichtung von  $0^\circ$  bis  $359.5^\circ$  entgegen dem Uhrzeigersinn in dem sog. ‘rechtshändigen’ System definiert. In vertikaler Ausrichtung wird der Winkelbereich von  $-60^\circ$  (unten) bis  $90^\circ$  (oben) definiert. Die unterschiedlichen Winkelangaben in der horizontalen Ausrichtung müssen bei der Verwendung des HRTF-Datensatzes im weiteren Verlauf angepasst werden.

Das originale Modell bestand aus 19 verschiedenen Oberflächenmaterialien. Die Anpassung der Nachhallzeit erfolgt in folgenden Schritten durch (noch realistisch) mögliche Änderungen der Absorptionskoeffizienten ( $\alpha$ ) dieser Oberflächenmaterialien. Ein Anheben des Absorptionskoeffizienten bewirkt dabei eine Verringerung der Nachhallzeit. In Abb. 3.9 ist die Funktion in EASE dargestellt, mit der man durch Wahl von Oberflächenmaterialien und Anpassung ihrer Absorptionskoeffizienten eine gewünschte Nachhallzeit erreichen kann.

Aus der Abb. 3.9 wird deutlich, dass sich die Nachhallzeit in dem entsprechenden Frequenzband verringert, wenn  $\alpha$  angehoben wird. Eine hohe Anzahl der Materialien kann den Anpassungsprozess an die gemessene Nachhallzeit sehr aufwendig und unübersichtlich machen. Daher muss die Anzahl der Absorbertypen reduziert werden. Ausserdem hat ein Großteil der Wandoberflächen in dem originalen Raummodell einen geringen prozentualen Anteil an der Gesamtoberfläche ( $< 5\%$ ), was den Einfluss auf die Nachhallzeit des Raumes durch das Berechnungsverfahren unbedeutend macht. In einem ersten Schritt werden Oberflächenmaterialien, die ähnliche Eigenschaften aufweisen, nach bestem Wissen zusammengefasst. In einem nächsten Schritt wird die Nachhallzeit aus der realen Messung in Terzbändern als Zielkurve (Target Value) definiert. Dann wird über die Seek Funktion in Optimize RT eine weitere Zusammenfassung der Oberflächenmaterialien erreicht. Seek sucht nach Materialien in dem Pool aller für den Raum vorhandenen Materialien, die die



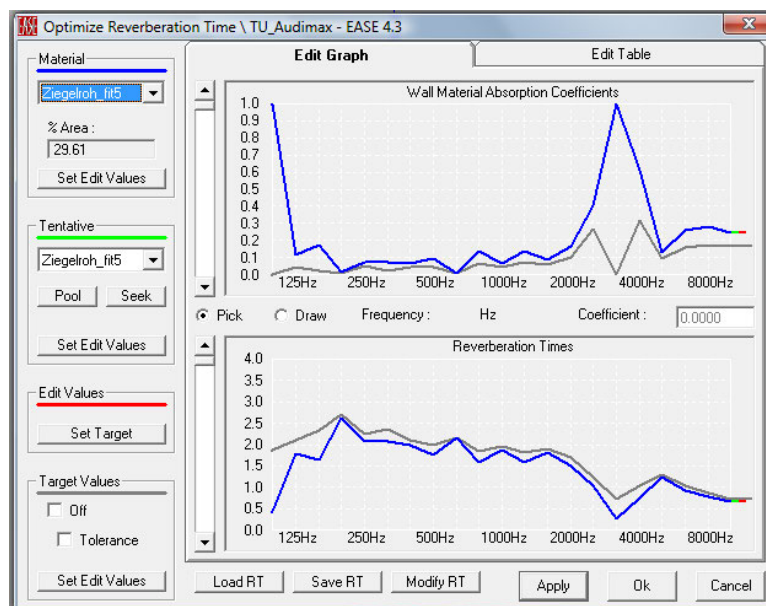


Abbildung 3.9: Optimize RT-Funktion zum Anpassen der Nachhallzeit des Modells an die Realität

angestrebte Nachhallzeit erreicht. Damit sind die Möglichkeiten, Oberflächenmaterialien des detaillierten Modells durch EASE automatisch zu reduzieren, ausgeschöpft.

In einem nächsten Schritt werden fünf der übrig gebliebenen Materialien bestimmt, die die Gesamtoberfläche des Raumes repräsentieren.

In der Tabelle 3.3 sind die Oberflächenmaterialien und ihr prozentualer Anteil an der Gesamtoberfläche von dem ursprünglichen Modell und der aus subjektiven Überlegungen ausgewählten Materialien für das weitergehende Fitting aufgelistet. Die Bezeichnung ‘fit’ weist darauf hin, dass die Materialien, wie im Folgenden beschrieben, schon angepasste Versionen darstellen. Es wird versucht durch die Anpassung des Absorptionskoeffizienten die Daten der Materialien so zu verändern, dass die angestrebte Nachhallzeit bestmöglich erreicht wird.

Mit der Eingabe der Absorptionswerte der Wände lassen sich leicht statistische Ergebnisse der Nachhallzeiten u.a. nach Sabine mit

$$T = 0.163 \frac{V}{A_{ges} + 4mV} \quad (3.4)$$

berechnen. Wobei  $V$  das Raumvolumen,  $A_{ges}$  die äquivalente Schallabsorptionsfläche und  $m$  die unvermeidbare Luftdämpfung beschreibt. Man muss davon ausgehen, dass eine statistische Gleichverteilung der akustischen Gegebenheiten, wie z.B. ein ortsunabhängiger Frequenzgang der Nachhallzeit, überall im Raum nicht gegeben ist. Somit kann dieser schnell zu errechnende Wert nur zur Orientierung dienen[11].

Für eine Detailluntersuchung der Nachhallzeiten wird zunächst die RIR für eine Positi-



<b>Original</b>		<b>Zusammenfassung</b>	
Material	Flächenanteil in %	Material	Flächenanteil in %
$a = 10\%$	15.38	Ziegelroh fit5	29.61
GLATTBETON	15.35	GLATTBETON fit3	21.14
ZIEGELROH	14.23	Klappstuhl fit5	18.28
PARKETT-3	9.56	Holzdielen fit4	17.93
ZIEGELFEIN	8.62	Holz-AB6	13.04
PARKETT	5.51	-	-
HOLZ-AB6	4.95	-	-
RAUPUTZ	4.26	-	-
SPANPLATTE	4.17	-	-
MARMOR	3.75	-	-
PARKETT-2	3.21	-	-
$a = 20\%$	2.64	-	-
HOLZDIELEN	2.56	-	-
RAUBETON	2.49	-	-
STUGIPS 2	1.92	-	-
DOOR SOLID	0.81	-	-
STEEL	0.31	-	-
FENSTER 2	0.27	-	-
KLAPPSTUHL	0.00	-	-

Tabelle 3.3: Originale und angepasst zusammengefasste Materialien und Flächenanteile in [%] des Raummodells vom Audimax der TU-Berlin

on in voller Auflösung berechnet und anschließend nach AFMG EASERA<sup>7</sup> exportiert. In EASERA wird eine breitbandige Nachhallzeit, in Terzen aufgelöst, aus dem sog. Schröder-Diagramm ermittelt. Aus dem ersten Export<sup>8</sup> ergibt sich, dass die Anpassung zu einer Annäherung geführt hat. In dieser Arbeit wird zumindest eine Annäherung der Nachhallzeit mit einem Fehler  $\leq 10\%$  in allen Terzbändern gefordert. Dies konnte in dem ersten Anpassungsschritt noch nicht erreicht werden. Daher werden in der EASE-Funktion, wieder in den einzelnen Terzbändern, Korrekturen an den Absorptionskoeffizienten für die einzelnen Wandoberflächen durchgeführt. Die Änderungen werden als ein neues Wandmaterial gespeichert und gegen das alte ausgetauscht. Ein systematisches Vorgehen beschränkt sich auf die Anpassung der Absorptionskoeffizienten an die grafische Darstellung terzbandaufgelöster Nachhallzeiten aus EASERA. Insgesamt wurde der Prozess sechs mal wiederholt, wobei jedesmal eine neue Impulsantwort errechnet und nach EASERA exportiert werden musste. Ziel war es, durch das Fitting einen gesamten Fehler  $\leq 10\%$  zu erreichen. In Abb. 3.10 ist die gemessene Zielkurve (i.d. Abb. oben ‘measured’), die gefittete Nachhallzeit (i.d. Abb. oben ‘fitted’), die direkte Differenz (i.d. Abb. oben ‘difference’) und der Fehler in Prozent (untere Grafik) dargestellt. Wie in Abb. 3.10 zu sehen konnte ein Fehler  $\leq 10\%$

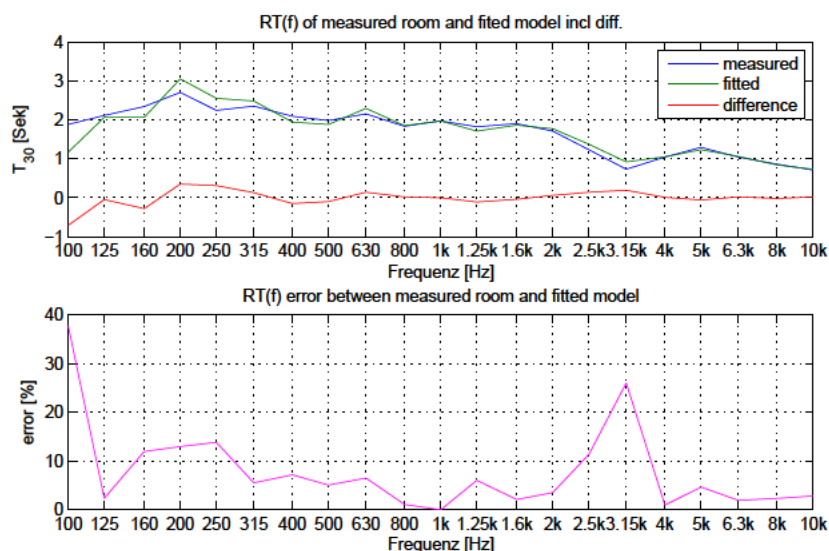


Abbildung 3.10: Terzbandaufgelöste Nachhallzeit des gemessenen Raumes (oben, blau), des gefitteten Modells (oben, grün) und der Fehler in % (unten).

in den meisten Terz-Frequenzbändern erreicht werden. Lediglich bei tiefen Frequenzen und im Terzband von 3.15kHz liegt der Fehler über der Zielvorgabe. Der Fehler im tiefen Frequenzbereich liegt an der limitierten Modellierung von tieffrequenten Absorptionskoeffizienten durch geometrische Raumakustik [12]. Der hohe Fehler im genannten Terzband entsteht durch die starke Welligkeit der Zielkurve aus der in-situ gemessenen Nachhallzeit.

<sup>7</sup>Version 1.1, siehe <http://easera.afmg.eu/> Zugriff 04.08.2011

<sup>8</sup>durch Calculate Aura Response - Save as \*.wav - Import in EASERA und vgl. mit in-situ T30

Das Ergebnis wird aber als ausreichend angesehen, um mit der Berechnung der RIR aus dem Raummodell fortzufahren.

### 3.2.2 Berechnung eines Reflektogramms mit EASE - Aura Response

Eine zu Auralisationszwecken geeignete RIR wurde in EASE mit dem Modul Aura Response berechnet. Das Modul basiert auf dem CAESAR-Algorithmus, entwickelt an der RWTH Aachen. Hier wird das hybride Verfahren zur Berechnung von Impulsantworten, wie in Kap.2.1 angedeutet, angewandt. Dabei entsteht ein AURA-“response file” (\*.rsp Files)<sup>9</sup>, welches die Informationen für die fünf Quell- und Empfängerkombinationen beinhaltet<sup>10</sup>.

Das Modul Aura Response bietet mehrere Optionen für die Berechnung der RIR. Die hier aufgeführten Einstellungen orientieren sich an Ergebnissen aus der Masterarbeit von [33], in der ein systematischer Vergleich von gemessenen BRIRs und durch Raumsimulation erhaltenen BRIRs durchgeführt wurde. Folgende Parameter wurden im Aura Modul benutzt:

#### Tab 1: Response

- Loudspeaker in Use: Lspk S1, S2, S3, S4, S5
- Send Response File to: Only Save To File
- create BIR: uncheck

Mit der letzten Option können BRIRs direkt mit EASE erstellt werden. Dabei werden HRTF-Daten aus der EASE-Datenbank benutzt. In dieser Arbeit soll aber der HRTF-Datensatz aus [32] verwendet werden, der mit dem zur Verfügung gestellten Datensatz nicht übereinstimmt. Somit wird diese Option nicht benutzt.

#### Tab 2: Location

- Walker: uncheck
- List. Seat: 1
- Prefix: reducedModel LSall (Dateinamen sollten nachvollziehbar sein)

#### Tab 3: Calculation

- Particles: Very High Resolution, Slow
- Length [ms]: Long + 30%
- Default Value [%] for Surfaces without Scattering Data: 60%

---

<sup>9</sup>durch Aura berechnete richtungsabhängige monaurale RIRs

<sup>10</sup>Das Aura Response Modul wird in der EASE Hauptoberfläche des geladenen Projekts über das Menü “Calculations”- “Room Mapping”- in “Eyes”- “Tools”- “Calculate Aura Response” geladen.

- Number of Calculation Threads: 32 Threads (8-Kern Rechner)

Die Parameter “Particles” und “Length” sind adaptiv. D.h. sie ergeben sich in Abhängigkeit von Oberfläche und Volumen des simulierten Raumes. “Particles” gibt die Anzahl der ausgesandten Strahlen für die stochastische Berechnung des Schallfeldes an. “Length” gibt die Länge des Zeitfensters für die Berechnung, also das Abbruchkriterium für die Registrierung von (Schall-) Strahlen auf der Zählkugel an. Die Anzahl der Partikel durch die Auswahl von einer sehr hohen Auflösung ergibt sich zu 1.996.000 Partikel. Für die Berechnung wurden schliesslich aber nur 499.000 Partikel verwendet. Nach einer Analyse der berechneten Datensätze stellte sich heraus, dass unabhängig von dem Wert für Partikel die Anzahl der individuell berechneten Reflexionen identisch bleibt. Der Grund dafür blieb trotz Nachfragen beim Hersteller ungeklärt.

Die adaptive Länge für das Abbruchkriterium der Stufe “Long” wird um 30% verlängert und liegt damit bei 3796 ms. Sie sollte aber mindestens so lang wie die zu erwartende Nachhallzeit sein. Aus den vorangegangenen Messungen weiß man, dass die Nachhallzeit um ca. 2,08 Sekunden liegt, so dass der Wert mehr als ausreichend ist. Für die angepassten Oberflächenmaterialien werden die Diffusionsdaten aus der Datenbank nicht verwendet, d.h. sie werden als rein spiegelnde Flächen angenommen. Durch den vorletzten Eintrag im ‘Tab3: Calculation’ kann ein globaler Diffusionsfaktor für alle Materialien festgelegt werden. Hier wird der Diffusionsfaktor von 60% gewählt. In den “Advanced Options” gibt es noch folgende Optionen:

- Max. Diameter after 1s [m]: default
- Diffuse Rain: uncheck
- Advanced Options - Cut Off Order: 16, Density Factor: 10, Tail [ms]: 5

Mit der “Cut-Off Order” wird die Reflexionsordnungsanzahl, mit der der Übergang von Spielschallquellenmodell zu stochastischem Raytracing berechnet wird.

Die Berechnung der RIR als Response-File kann nun durch “Start Simulation” begonnen werden. Die Berechnungszeit wurde nicht dokumentiert. Die Berechnung für alle 5 Quell- und Empfängerpositionen wurde über Nacht durchgeführt.

Die Response-Files können im “Probe”-Modul<sup>11</sup> von EASE geöffnet werden und pro Quelle als Reflektogramme in Form von Textfiles bzw. als RIRs im Wave-Format abgespeichert werden. In den Textfiles wird die RIR, wie in Kap. 2.1 angedeutet, in numerischer Form als Reflektogramm dargestellt.

---

<sup>11</sup>Das Probe Modul wird in der EASE Hauptoberfläche des geladenen Projekts über das Menü “Calculations”- “Probe Processing” geladen.

### 3.2.3 Extraktion wichtiger Parameter aus dem Reflektogramm

Das als Textdatei gespeicherte Reflektogramm wird zur Erzeugung minimalphasiger Impulsantworten benötigt, welche wiederum nach Gewichtung mit HRTFs zu BRIRs zusammengesetzt werden. Zunächst müssen aber relevante Informationen in geeigneter Form aus den Dateien extrahiert werden. Dafür wurde ein sog. ‘Parser’ geschrieben, der \*.mat Files erzeugt, aus denen Impulsantworten für jede Reflexion berechnet werden können. Der ‘Parser’<sup>12</sup> wurde als ausführbares Skript in MATLAB geschrieben. Es werden zunächst die 5 Textfiles geladen und zum Lesen geöffnet. Nicht relevante Informationen, wie sie im ‘Header’ der Textfiles vorhanden sind, werden ignoriert. Relevante Informationen werden durch die ‘textscan’-Funktion in MATLAB extrahiert und entsprechend als aussagekräftige Variablen in geeigneter Form abgespeichert.

So werden Informationen u.a. für

- Ankunftszeit des Direktsignals in [ms]
- Länge der Impulsantwort in [ms]
- gesamte Anzahl der Reflexionen in der RIR
- Mittenfrequenzen der 21 Terzbänder
- Ankunftszeit in [ms], Amplitudenwerte in [dB] pro Frequenzband und Einfallswinkel vertikal und horizontal in [°] für alle individuell berechneten Reflexionen nach dem Direktsignal

pro Quelle vorliegender Textfiles extrahiert. Die Textfiles können nach der Extraktion wieder geschlossen werden.

Als nächstes wird die Zeit bestimmt, nach der man unabhängig von der Berechnungsmethode in Aura Response davon ausgehen kann, dass man den individuellen Nachhall der RIR durch einen für den gesamten Raum repräsentativen Nachhall ersetzen kann. Dieser Zeitpunkt wird in [28] systematisch untersucht und als ‘perceptive Mixingtime’ bezeichnet. Sie ist definiert als der Zeitpunkt, ab dem der Nachhall einer RIR ohne hörbare Konsequenzen durch einen Nachhall einer anderen beliebigen Sender/Empfängerkonstellation ausgetauscht werden kann. In [28] kann die perceptive Mixingtime am ehesten durch das Volumen vorausgesagt werden, mit:

$$t_{mp} = 0.0117 \cdot V + 50.1 \quad (3.5)$$

Für die Anzahl der zu extrahierenden Reflexionen aus dem Reflektogramm wird zunächst der doppelte Wert der  $t_{mp}$  mit 300 ms als ‘t stop’ definiert. Aus dem Reflektogramm ergibt sich, dass (unverständlicherweise) Schallfeldkomponenten vor dem Direktsignal als ‘Vorechos’ berechnet wurden. Es sollen aber nur Komponenten nach dem Direktsignal

---

<sup>12</sup>Der Parser ist auf der beigefügten DVD im soundfield-Repository - 2 EASE soundfield creation - run1 extract soundfield from EASE.m zu finden

betrachtet werden. Daher wird eine Matrix konstruiert, in der nur die Spektren des Direktsignals und der darauffolgenden Reflexionen bis  $t_{mp}$  enthalten sind. Ermittelt wird die Anzahl der zu betrachtenden Reflexionen bis  $2 \cdot t_{mp}$ .

Damit können jeweils Matrizen konstruiert werden, in denen zunächst das Terzband aufgelöste Frequenzspektrum ('spectra') für jede Reflexion enthalten ist.

Aus der Aura-Berechnung ergab sich für die Spektren ein Problem mit Nullstellen. Es kommt vor, dass einige Amplitudenwerte den Wert Null annehmen, bis hin zu dem Fall, dass für komplette Reflexionen keine Amplitudenwerte mit Aura berechnet wurden. Dies führt im Anschluss zu Fehlern bei der Erzeugung von Impulsantworten für die einzelnen Reflexionen als FIR-Filter. In Abb. 3.11 sind die Auswirkungen einzelner Nullstellen im Spektrum beispielhaft für eine Reflexion mit Nullstellen im 3.15kHz Terzband zu sehen.

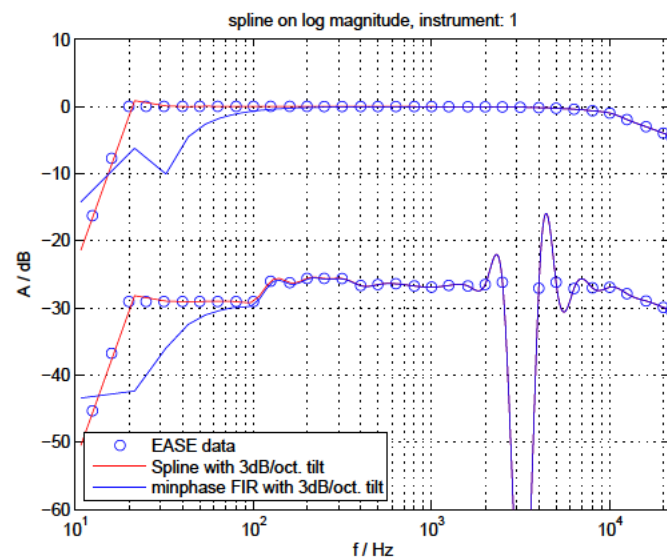


Abbildung 3.11: Amplitudenspektren mit Nullstellen im 3.15 kHz Terzband einer Reflexion für eine Quelle (Src1)

Um die Erzeugung fehlerhafter Impulsantworten aus dem Reflektogramm zu verhindern, werden

- Nullstellen im ersten und letzten Frequenzband durch die Werte aus dem jeweils darüber bzw. darunter liegenden Terzband
- Reflexionen, für die weniger als zwei reelle Werte berechnet wurden, durch die zeitlich davor berechnete Reflexion
- mehrere aufeinanderfolgende Nullstellen durch eine lineare Interpolation mit benachbarten reellen Werten

ersetzt.

Anschließend konvertiert man die Pegel in den Terzbändern in lineare Werte und

normalisiert sie mit maximalen Amplituden zwischen  $\pm 1$ .

Man speichert die normalisierten, linearen, Terzband aufgelösten Frequenzspektren jeweils in einer Matrix als \*.mat Files. Weiterhin speichert man die sog. Metadaten in einem separaten File. In den Metadaten sind für das Direktsignal und jede Reflexion die Dämpfung (zunächst noch mit dem Wert 1 für alle Reflexionen, Erläuterung folgt in Abschnitt 3.2.4), Ankunftszeit, Einfallswinkel in Azimut und Elevation und ein Wert für die Hörbarkeit (zunächst noch mit dem Wert 1 für alle Reflexionen) enthalten.

### 3.2.4 Erzeugung von Impulsantworten der extrahierten Amplitudenspektren aus dem Reflektogramm

In einem MATLAB-Skript<sup>13</sup> generiert man minimalphasige FIR-Filter der extrahierten Amplitudenspektren für jede Reflexion pro Quelle. Die FIR-Filter werden in folgender Abfolge für jede Quelle und jede einzelne Reflexion in dem Skript erzeugt.

1. Erzeugung des Amplitudenspektrums von EASE-Daten und LF/HF-Neigung
2. Interpolation (Spline) von Amplitudenspektrum (logarithmisch oder lineare Amplitude)
3. Erzeugung des FFT-Spektrums mit symmetrischer Amplitude

wähle Modus für die Phasenrekonstruktion: Phasenmodus = 'alex' / 'frank'

4. 'alex' ('unsauber', aber schneller, Amplitudenverlust bis zu 10 dB unterhalb von 80 Hz): Berechnung eines realen FIR-Filters ohne Phase (zerophase)  
'frank': Berechnung eines linearphasigen FIR-Filters (notwendige Funktion: makeLinPhase.m)
5. 'alex': Berechnung eines minimalphasigen FIR-Filters von nullphasigem FIR  
'frank': Berechnung eines minimalphasigen FIR-Filters von linearphasigem FIR (notwendige Funktion: makeMinPhase.m)
6. Ermittlung der Maximalamplitude im minimalphasigen FIR-Filter
7. schreibe das Amplitudenverhältnis relativ zum Direktsignal (=1) in out2 für die Metadaten
8. Normalisierung der Impulsantwort
9. Die normalisierte IR als Wave-File speichern

---

<sup>13</sup>Das Skript ist auf der beigefügten DVD im soundfield-Repository - 2 EASE soundfield creation - run2 make minphase IRs from soundfield.m zu finden

10. Die relativen linearen Amplituden von out2 in einem neuen Metadatensatz “metadata src XX incl damping.mat” speichern

Prinzipiell erhält man durch spektrale Interpolation, unter Annahme minimaler Phasenlage und nach inverser Fouriertransformation die Impulsantwort jeder Reflexion aus dem Reflektogramm. Die Impulsantworten werden normalisiert und in den Metadaten werden zusätzlich die Dämpfungsfaktoren relativ zum Direktsignal in einem linearen Verhältnis von Maximalwert des Direktschalls zum Maximalwert jeder Reflexion aktualisiert.

Die Impulsantworten der Einzelreflexionen werden für den auditorischen Algorithmus benutzt, um die Reflexionen ähnlich der Signalverarbeitung im auditorischen System zu verarbeiten.



### 3.3 Auditorisch reduzierter modellierter Raum

Für die binaurale Reproduktion des auditorisch reduzierten modellierten Raumes wird das individuell beschriebene Schallfeld in der Form von minimalphasigen Impulsantworten für Direktsignal und Reflexionen verwendet. Dazu wird das Modell vorgestellt, welches eine perzeptiv motivierte Reduktion der Schallfeldbeschreibung ermöglicht. Die Reduzierung des Schallfeldes auf die tatsächlich hörbaren Reflexionen wird durch die Modellierung der Hörschwelle und spektraler, zeitlicher und richtungsabhängiger Verdeckungseffekte im auditorischen System erreicht.

#### 3.3.1 Das einfache auditorische Prozessmodell

In Anlehnung an das von [4] vorgeschlagene Modell wird ein einfaches Prozessmodell entwickelt, welches zunächst das Verdeckungsverhalten von Impulspaaren modellieren soll, da, wie in Kapitel 2.2.1 gezeigt, das Verdeckungsverhalten von Impulspaaren besonders kritisch ist.

Es wird der gesamte Signalweg vom Außen- und Innenohr, mechano-neuronaler Schnittstelle der Haarzellen, neuronaler Verarbeitung und Entscheidungsebenen zu Hörbarkeit wie im auditorischen System modelliert.

Das in dieser Arbeit entwickelte Modell ist als Blockdiagramm in Abb. 3.12 schematisch dargestellt. Wie in [4] besteht das Modell aus einzelnen Modulen, die im Folgenden

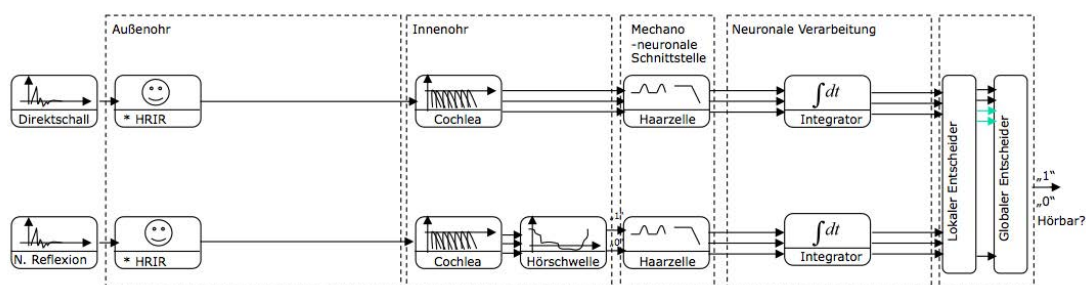


Abbildung 3.12: Einfaches Prozessmodell zur Modellierung des Verdeckungsverhaltens von Impulspaaren

entsprechend ihrer Funktion im auditorischen System und der Modellierung durch digitale Audiosignalverarbeitung genauer beschrieben werden.

#### Normierter Schalldruckpegel

Zunächst muss das im Zeitbereich als Vektoren dargestellte Impulspaar auf einen absoluten Schalldruckpegel ( $ref_{dB}$ ) normiert werden. Diesen Wert kann man selbst wählen. Dann

wird der Effektivwert (rms) der IRs mit

$$p_{rms} = \sqrt{\frac{1}{n} \sum_n IR^2(t)} \quad (3.6)$$

berechnet und der unkorrigierte Schalldruckpegel relativ zum Schalldruck der Bezugshörschwelle  $p_0 = 20\mu PA$  bei 1 kHz mit

$$IR_{SPL} = 20 * \log_{10}\left(\frac{p_{rms}}{p_0}\right) \quad (3.7)$$

berechnet. Daraus wird ein Korrekturwert in dB als Differenz in der Form

$$dL_{corr} = ref_{dB} - IR_{SPL} \quad (3.8)$$

der IRs errechnet. Der Korrekturwert wird wiederum in linearisierter Form mit den im Zeitbereich dargestellten Impulsantwortvektoren multipliziert

$$IR_{norm} = 10^{dL_{corr}/20} * IR(t) \quad (3.9)$$

, so dass das die Impulspaare normiert auf den absoluten Schalldruckpegel von  $ref_{dB}$  vorliegen.

### Außenohr

Die frequenzabhängige Richtwirkung einer Testreflexion ergibt sich aus Beugungseffekten und Reflexionen an den Schultern, dem Kopf und den Außenohren des Menschen. Der Richtungsselektivität des menschlichen Gehörs wird durch die Gewichtung der Testreflexion mit HRTF-Daten Rechnung getragen. Dazu wird die Testreflexion mit der dem Einfallswinkel entsprechenden kopfbezogenen IR<sup>14</sup> gefaltet und die Laufzeitverzögerung zwischen linkem und rechtem Ohr eingeführt.

In dieser Arbeit wird der schon erwähnte HRTF-Datensatz aus [32] verwendet. In einem Teilprojekt wurde der HRTF-Datensatz von Antonio Cardenas in [34] untersucht und für den korrekten Einsatz zur Beschreibung der Richtungsselektivität vorbearbeitet. Die Bearbeitung beinhaltet die Kompensation möglicher Messfehler zum Erhalten des HRTF-Datensatzes. Dabei muss z.B. die Richtwirkung von Messlautsprechern, die Einwirkung von ungewollten Reflexionen aus der Messumgebung und der Einfluss der Messmikrofone kompensiert werden.

Aus der Arbeit von [34] geht der HRTF-Datensatz als minimalphasiger Filter mit einem getrennten Verzögerungsterm hervor, ohne den räumlichen Einfluss der HRTF zu beeinflussen. In den minimalphasigen HRIRs sind die Informationen über den interauralen Pegelunterschied (ILD) von linkem und rechtem Ohr enthalten. Die interaurale Zeitverzögerung (ITD) wird als frequenzunabhängige reine Verzögerung bei der Wiedergabe mit Kopfhörern dem linken und rechten Kanal zugeführt. Die Daten

---

<sup>14</sup>HRIR entspricht der HRTF im Zeitbereich

werden in einem OpenDAFF-Format gespeichert und zugänglich gemacht. OpenDAFF ist ein Open Source Format, welches den Austausch von richtungsspezifischen Audiodaten ermöglicht. Das Format ermöglicht einerseits den Zugriff auf korrekte HRIRs bezüglich der Kopfausrichtung in Echtzeit und stellt andererseits eine Kompatibilität zu Richtungsangaben aus verschiedenen Koordinatensystemen her, wie es z.B. EASE benutzt. Der HRIR-Datensatz aus [32] hat eine begrenzte Auflösung<sup>15</sup>. So kann es dazu kommen, dass für eine Testreflexion kein entsprechender HRIR-Datensatz für die Faltung zur Verfügung steht. Eine eingebaute Funktion in OpenDAFF sorgt dann dafür, dass nach dem ‘nearest-neighbour’-Interpolationsverfahren valide Datensätze erzeugt werden, auch wenn die Winkelangaben ausserhalb des vorliegenden Datensatzes liegen. In Zukunft sollen hier aufwendigere Interpolationsverfahren eingesetzt werden, um die Daten einer HRIR nicht zu verfälschen.

Der Einfluss durch den Ohrkanal und das Mittelohr wird nicht simuliert, wie in Kapitel 2.3.4 angesprochen.

### Basilarmembran

Dem Prozessmodell aus Abb. 3.12 folgend, kommt es nun zu der Modellierung des menschlichen Innenohrs, wie aus der Forschungsliteratur bekannt. Dabei wird die Frequenzauflösung des auditorischen Systems durch die Funktion der Basilarmembran im menschlichen Innenohr bestimmt.

Eine auditorische Filterbank simuliert in dem Modell die Frequenzauflösung des auditorischen Systems. Man geht davon aus, dass die Funktion der Basilarmembran durch eine Reihe von überlappenden Bandpassfiltern ( $\rightarrow$  Filterbank) darstellbar ist. Jeder Bandpassfilter in der Filterbank hat eine eigene Mittenfrequenz, so dass aus der Anzahl berechneter Filter der gesamte Bereich von dem für den Menschen hörbaren Frequenzen abgedeckt wird.

Die außenohr-gewichteten Eingangssignale werden durch die Anwendung einer Gammaton-Filterbank in einzelne Frequenzbänder aufgeteilt. Nach der Auflösung der Signale durch die Filterbank wird das Signal bandweise verarbeitet.

Für eine Gammaton-Filterbank wird die sog. ‘equivalent rectangular bandwidth’ (ERB) zur Definition der Bandbreite eines Bandpassfilters verwendet. Die ERB ist eine psychoakustische Größe, die die Breite eines auditorischen Filters entlang der Cochlea im Innenohr beschreibt. Für die Berechnung eines Bandes muss u.a. die Mittenfrequenz, bzw. eine Anzahl von Mittenfrequenzen des gewünschten Frequenzspektrums festgelegt werden. Für die Herleitung der Filter mit ERBs sei hier u.a. auf [35] oder [36] verwiesen. In dem Abschnitt 3.3.2 werden verschiedene Implementationen von Gammaton-Filterbänken untersucht. Aufgrund ihres Rekonstruktionsverhaltens wird sich für eine Implementation entschieden, um sie in das Modell zu integrieren.

### Hörschwelle

---

<sup>15</sup>in 0.5°-Schritten auf der horizontalen und 5°-Schritten zwischen -60° bis 90° in der vertikalen Ebene

Die Haarzellenbewegung im Cortischen Organ auf der Basilarmembran, in Kombination mit der Ohrflüssigkeitsbewegung im Innenohr, bewirkt einen internen Rauschpegel, der die absolute Hörschwelle des Menschen beeinflusst [37]. Um die Hörschwelle des auditorischen Systems nachzubilden, wird für die bandgefilterte Testreflexion der Effektivwert des Zeitverlaufes<sup>16</sup> berechnet und ins Verhältnis zu dem Effektivwert der Impulsantwort des Direktsignals gesetzt.

Für die Hörbarkeit einer Testreflexion reicht es aus, wenn der Schalldruckpegel innerhalb eines Bandes über dem Schalldruckpegel der ISO-Norm für die Bezugshörschwelle im Diffusfeld [38] liegt. Die Werte werden für gerade eben wahrnehmbare Schalldrücke durch einfache Interpolation auf die Mittenfrequenzen der Filterbänder aus der Gammaton-Filterbank angepasst.

### Haarzellenmodell

Die inneren Haarzellen befinden sich auf dem Cortischen Organ in der Cochlea. Die mechanischen Bewegungen der Haarzellen auf der Cochlea bewirken eine Umsetzung zu neuronaler Aktivität in den auditiven Nervenbahnen [39].

Die mechano-neuronale Schnittstelle zwischen Innenohr und neuronaler Verarbeitung kann durch das sog. Haarzellenmodell simuliert werden.

Darin wird die effektive Signalverarbeitung der inneren Haarzellen durch einen Einweggleichrichter<sup>17</sup> und einen Tiefpassfilter typischerweise bei 1 kHz simuliert. Der Einweggleichrichter bewirkt in etwa die Transformation von Basilarmembran-Schwingungen in Rezeptor-Potentiale der Haarzellen. Der Tiefpassfilter erhält die Hüllkurve für Signale mit hohen Trägerfrequenzen [22].

Die Gleichrichtung und Tiefpassfilterung wird durch die Berechnung, wie in der Toolbox von [39], realisiert. Hier kann man für die Berechnung mit dem auditorischen Modell zwischen Halbweg-Gleichrichtung ('half') und Gleichrichtung ('full') für den 'r type' wählen.

### Neuronale Übertragung

Die zeitliche Auflösung der neuronalen Übertragung wird durch eine Integration über bandweise unterschiedliche Fensterlängen realisiert. Dabei kommt es zu einer Mittelung über gleitende Rechteckfenster mit individuell unterschiedlichen Fensterlängen, wie in [1] vorgeschlagen. Die Fensterlängen, wie sie hier zur Anwendung kommen, sind in der Tabelle 2.3.1 aufgeführt. Die grobe Frequenzauflösung aus der Tabelle wird auf eine feine Auflösung interpoliert. Die Fensterlängen werden für jedes Frequenzband aus der Gammaton-Filterbank berechnet. Die Mittelung geschieht über eine FFT-basierte FIR-Filterung der Testreflexion im Zeitbereich. Der Filter wird durch den Koeffizienten-Vektor aus den interpolierten Fensterlängen beschrieben.

---

<sup>16</sup>engl. RMS: root mean square

<sup>17</sup>engl. 'half-wave rectifier'

### Lokaler Entscheidungsapparat

Kognitive Entscheidungen über die Hörbarkeit einer Testreflexion werden durch den lokalen Entscheidungsapparat simuliert. Hörbarkeitsentscheidungen werden auf der ‘lokalen’ Ebene noch pro Band durchgeführt.

In dem einfachen Prozessmodell wird zunächst die Amplitude des auditorischen Signals vom Verdeckter am Punkt der maximalen Amplitude des auditorischen Signals der Testreflexion verglichen. Als Ausgangspunkt wurde die Empfindlichkeit (Detektionsregel) der Einfachheit halber so gesetzt, dass wenn die maximale Amplitude des auditorischen Signals der Testreflexion größer oder gleich der maximalen Amplitude des auditorischen Signals des Verdeckers ist, wird die Testreflexion in diesem Band als hörbar eingestuft. Die einfache Detektionsregel kann in der Form

$$\frac{\hat{S}_0(t)}{\hat{S}_N(t)} \geq 1 \rightarrow S_N(t) = 1 \quad (3.10)$$

ausgedrückt werden. Wobei  $\hat{S}_0(t)$  die maximale Amplitude des auditorischen Signals des Verdeckers und  $\hat{S}_N(t)$  die maximale Amplitude des auditorischen Signals der Testreflexion beschreibt.  $S_N(t) = 1$  bedeutet, dass die Testreflexion hörbar ist. Die anschließende Evaluation an empirische Daten, siehe Kapitel 3.3.3 machte jedoch erwartungsgemäß deutlich, dass eine Skalierung dieses internen Hörbarkeitskriteriums notwendig war. Die Skalierung wird so ermöglicht, dass die Amplitude des Verdeckers zusätzlich mit einem empirisch gefitteten Korrekturwert ‘at’<sup>18</sup> in der Form

$$\frac{\hat{S}_0(t)}{at \cdot \hat{S}_N(t)} \geq 1 \rightarrow S_N(t) = 1 \quad (3.11)$$

multipliziert wird. Dies stellt eine einfache Implementierung von simultanen und Nachverdeckungseffekten dar, ohne adaptive Signalabhängigkeiten, wie z.B. in [23, 24] zu betrachten. Mit dem Schwellwert ‘at’ existiert ein Faktor, mit dem man Hörbarkeitsentscheidungen entsprechend den Verdeckungskurven aus empirischen Daten beeinflussen kann. In dem folgenden Kalibrierungsprozess, siehe Kapitel 3.3.3 stellte sich heraus, dass ‘at’ maßgeblich entscheidend für die Abbildung empirischer Verdeckungskurven ist. Neben der absoluten Hörschwelle stellt der lokale Entscheidungsapparat die zweite Reduktionsmöglichkeit des auditorischen Modells dar. Das Modell behandelt bis zum jetzigen Zeitpunkt Signale für das linke und rechte Ohr getrennt.

### Globaler Entscheidungsapparat

Ein globaler Entscheidungsapparat sammelt die einzelnen Hörbarkeitsentscheidungen beider Ohren aus dem lokalen Entscheidungsapparat. Die Validierung in Kapitel 3.3.3 zeigte weiterhin, dass es zzgl. zur Schwellwertskalierung mit ‘at’ notwendig war, den Prozentsatz positiver Erkennung als zweiten einstellbaren Empfindlichkeitsparameter einzuführen.

---

<sup>18</sup>at steht hier für audibility threshold, also dem Hörbarkeitsschwellwert

Wenn ein bestimmter Prozentsatz positiver Hörbarkeitsentscheidungen aus den einzelnen Bändern vorliegt, wird die Testreflexion als hörbar eingestuft. Damit kann die Empfindlichkeit des einfachen Prozessmodells für Hörbarkeitsentscheidungen beeinflusst werden. Man erhält also einen weiteren Wert, der die Abbildung von Verdeckungskurven aus empirischen Daten beeinflusst und damit für den Kalibrierungsprozess verwendet werden kann. Im Zuge der Validierung wurden drei Empfindlichkeitsstufen überprüft.

1. Das Modell reagiert besonders empfindlich, wenn die Testreflexion in wenigstens einem Band im lokalen Entscheider als hörbar eingestuft wird.
2. Die mittlere Sensibilitätsstufe wird erreicht, wenn die Testreflexion in 50% aller Filterbänder im lokalen Entscheider als hörbar eingestuft wird.
3. Die Empfindlichkeit des Modells ist gering, wenn die Testreflexion in 100% aller Filterbänder im lokalen Entscheider als hörbar eingestuft wird.

Der globale Entscheidungsapparat stellt die dritte Reduktionsmöglichkeit im auditorischen Modell dar.

### 3.3.2 Vergleich von Gammaton-Filterbänken

In der Literatur existieren Implementationen, die eine Gammaton-Filterbank realisieren und dokumentieren. Eine durchgeführte vergleichende Analyse dient der Entscheidung, welche der hier vorliegenden Implementationen für das auditorische Modell am besten geeignet ist. Folgende Filterbank-Implementationen <sup>19</sup> wurden untersucht:

- Auditory Toolbox Version 2 von M. Slaney, beschrieben in [40]
- Octave Toolbox von C. Couvreur <sup>20</sup>
- Gammaton filterbank von V. Hohmann, beschrieben in [36]
- GammaTone Tool Kit v2.0 Auditory ERB Filterbank von N. Clark in [41]<sup>21</sup>

Das Ziel war es hier, in einem großen Frequenzbereich mit einer variablen Anzahl von Filterbändern eine perfekte Rekonstruktion des Eingangssignals nach Filterung zu erhalten. Ein nahezu perfektes Rekonstruktionsverhalten soll verhindern, dass spektrale Signalanteile durch die Filterbank unterschiedlich bewertet werden. Auch in diesem Fall ist die Filterbank als lineares zeitinvariantes System durch seine Impulsantwort als Reaktion des

<sup>19</sup>Komplette Toolboxes sind auf der beigefügten DVD im Repository ‘third party tools’ zu finden. Für den Vergleich notwendige MATLAB-Funktionen sind jeweils auf der beigefügten DVD im soundfield-Repository - 4 auditory soundfield processing - fb tools zu finden

<sup>20</sup>die Komplette Toolbox mit Kontaktdaten zum Autor und Erstellungsdatum ist auf der beigefügten DVD im Repository ‘third party tools’ zu finden

<sup>21</sup>alternativ ist die Komplette Toolbox auf der beigefügten DVD im Repository ‘third party tools’ zu finden

Systems auf einen unendlich kurzen Diracstoß beschrieben. Die Impulsantwort der Filterbank erhält man durch die Anwendung eines in MATLAB erstellten endlichen Impulses als Annäherung eines Diracstoßes. Das Rekonstruktionsverhalten der Filterbänder kann direkt beobachtet werden. Der Vergleich findet im Zeit- und Frequenzbereich und anhand der Gruppenlaufzeiten der aufsummierten Bänder statt. Dazu dienen folgende Darstellungen:

1. Amplitudenfrequenzgänge aller Einzelfilter
2. Summenamplitudenfrequenzgang
3. die kumulierte Impulsantwort
4. die aufsummierte Gruppenlaufzeit der Einzelfilter

Die Darstellungen sind in den folgenden Abschnitten von allen genannten Filterbank-Implementationen für verschiedene Anzahlen von Bändern zu sehen.

Für einen Vergleich soll das Verhalten aller Filterbänke mit, so weit möglich, identischen (globalen) Parameter untersucht werden.

Globale Parameter zur Berechnung der Filterbänke sind:

- Samplingfrequenz,  $fs = 44,1kHz$
- Länge des Impulses als Eingangssignal  $2^{12}$
- untere Grenzfrequenz  $f_u = 50Hz$ , obere Grenzfrequenz  $f_o = 20kHz$
- jeweils gleiche Anzahl von Bändern

In Abb. 3.13 sind die genannten Darstellungen aus der Implementation von [40] für 30 Bänder zu sehen. Es werden stabile Filter für die gewünschte Anzahl aller Bänder im Frequenzbereich berechnet. Das Amplitudenspektrum ist weitestgehend linear. Die Rekonstruktion der Impulsantwort scheint gut und das Gruppenlaufzeitmaximum von ca. 20 ms bei 55 Hz ist akzeptabel

In Abb. 3.14 sind die genannten Darstellungen aus der Implementation von [36] für 30 Bänder zu sehen. Das Verhalten der Filterbank erscheint im wesentlichen identisch zur Abb. 3.13 von [40]. Gammaton-Filter mit gleichmäßiger Verteilung auf der ERB-Skala werden auf Basis einer Frequenz ("base frequency") berechnet. Die Dichte bzw. Anzahl der Filter ergibt sich auf Grundlage einer definierten Schrittweite [36]. Somit ist die Filterband-Anzahl nicht wie gewünscht frei wählbar. Wie aus Abb. 3.14 oben links deutlich wird, ergeben sich 29 Filterbänder, obwohl 30 gefordert waren.

In Abb. 3.15 sind die genannten Darstellungen aus der Implementation von [41] zu sehen. Im unteren Frequenzbereich kommt es zu Überhöhungen einzelner Frequenzbänder. Das lässt auf nicht stabile Filter schließen. In einem Pol-/Nullstellen-Diagramm für die einzelnen Filter lässt sich, anhand der Lage der Pole bezüglich des Einheitskreises, die Stabilität der Filter zeigen. In Abb. 3.16 sind die Polstellen markiert, die sehr nahe des Einheitskreises liegen. Die detaillierten Filter Informationen des "fvtool"<sup>22</sup> in MATLAB ge-

---

<sup>22</sup>engl Filter Visualization Tool

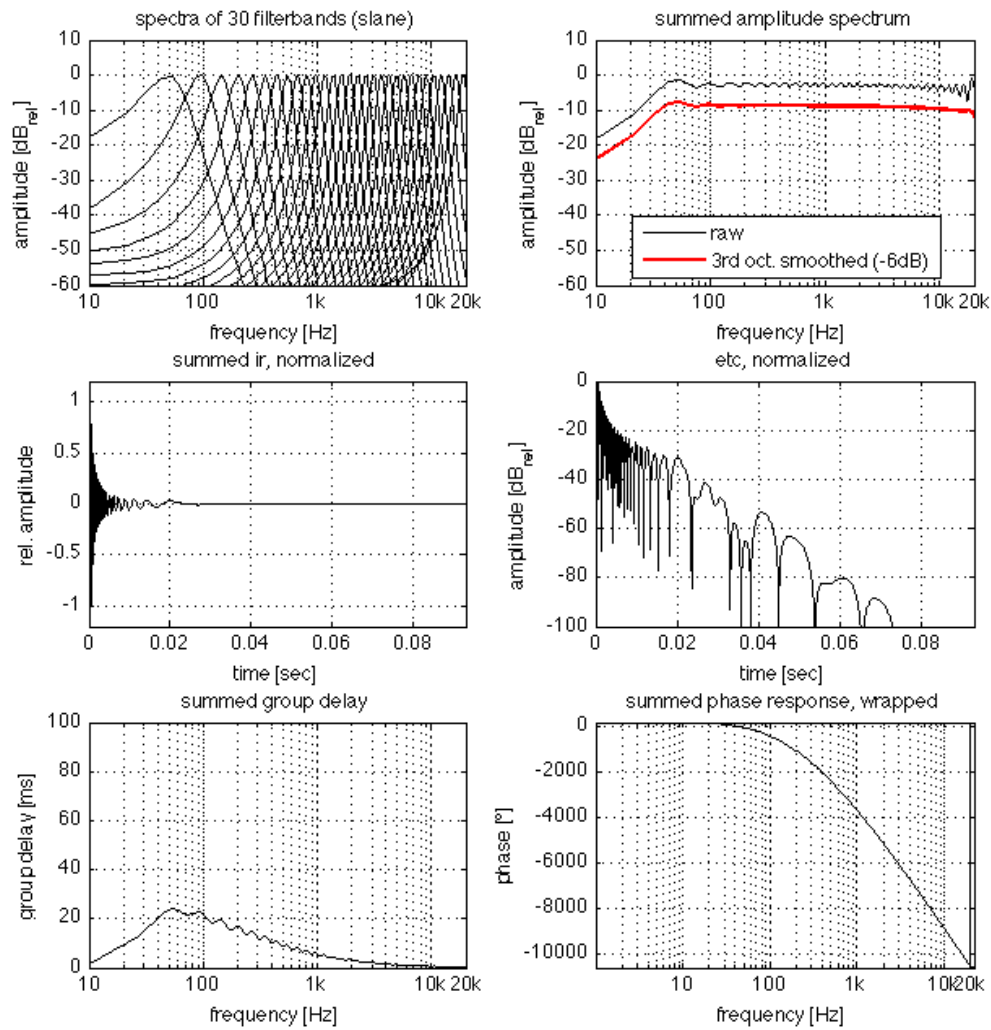


Abbildung 3.13: Rekonstruktionsverhalten der Gammaton-Filterbank aus [35]



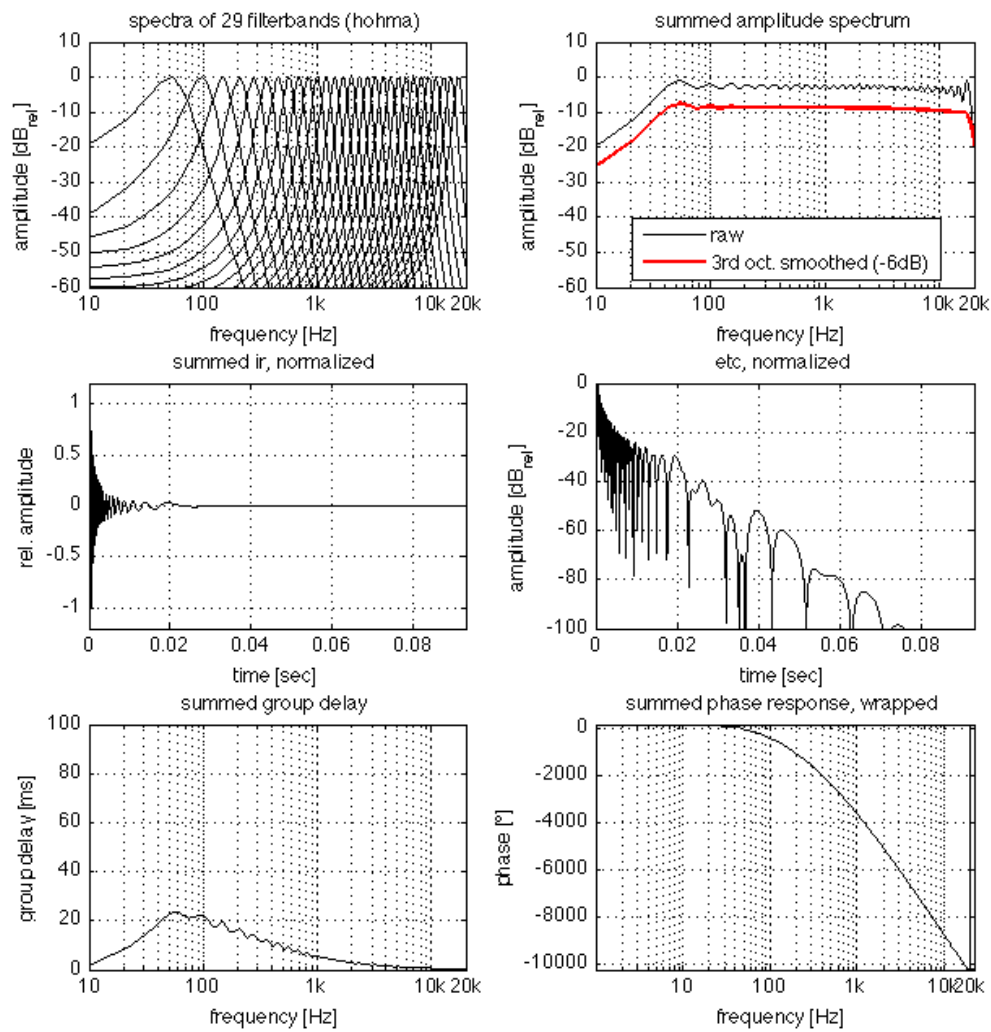


Abbildung 3.14: Rekonstruktionsverhalten der Gammaton-Filterbank aus [36]

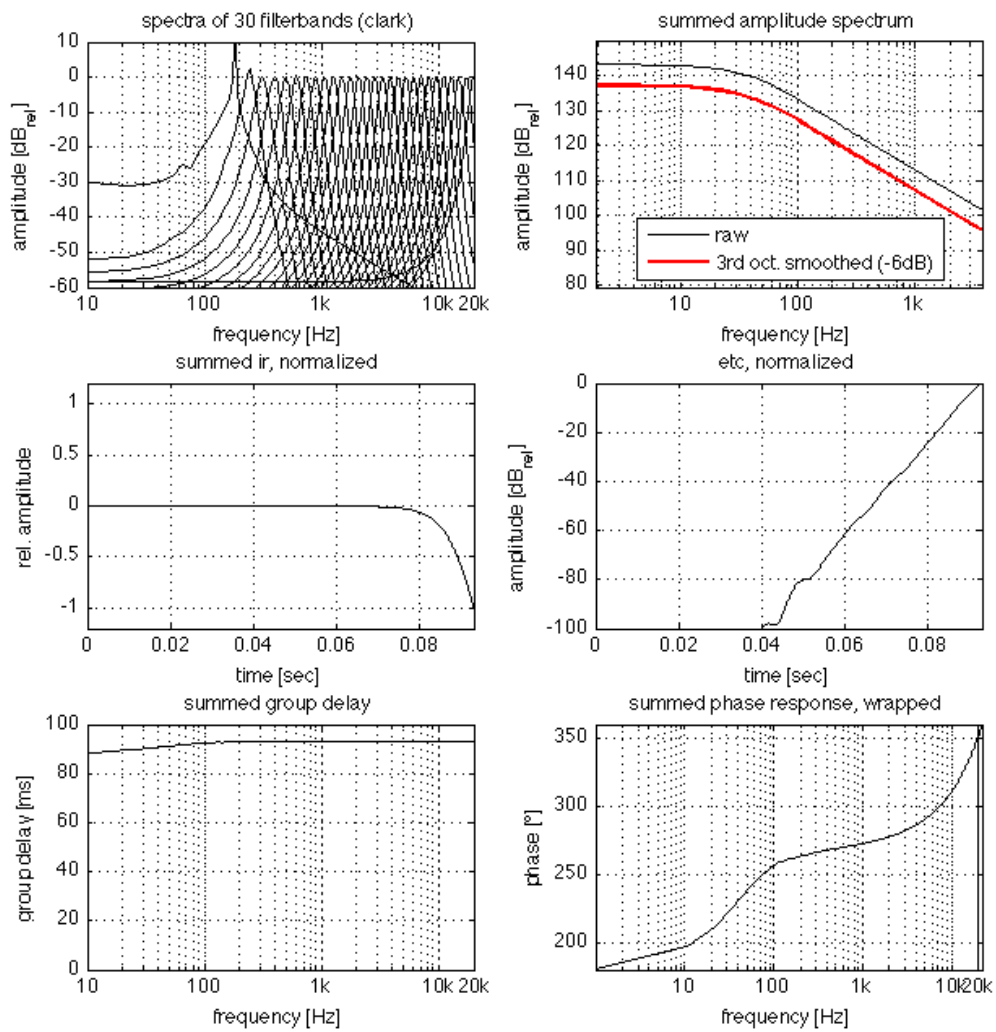


Abbildung 3.15: Rekonstruktionsverhalten der Gammaton-Filterbank [41]

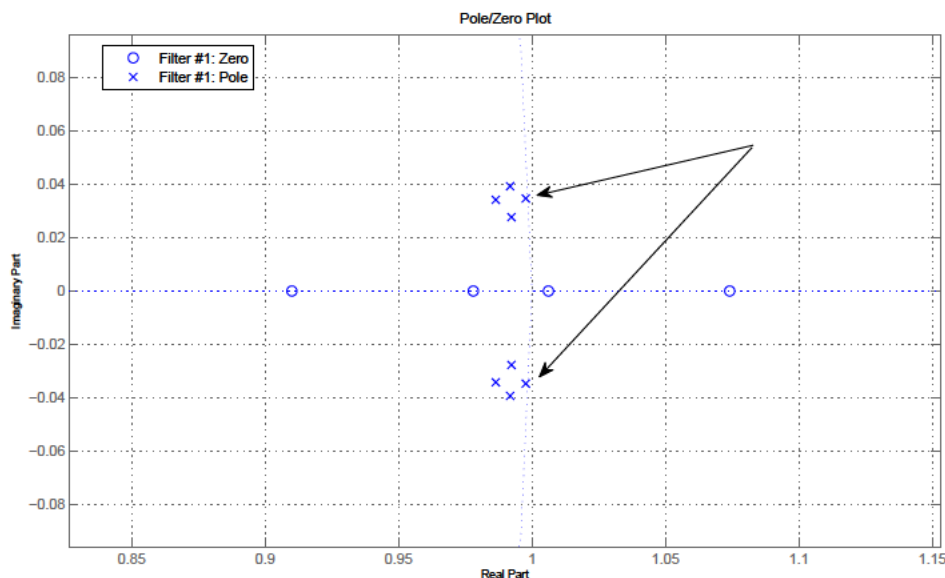


Abbildung 3.16: Pol-/Nullstellendiagramm des vierten Filterbands (von 30), Mittenfrequenz bei 240 Hz, berechnet mit der Gammaton Filterbank aus [41]

ben an, dass der Filter für das entsprechende Frequenzband (mit der Mittenfrequenz von 240 Hz) nicht stabil ist. Für die Berechnung einer Filterbank mit 30 Bändern gilt dies für die unteren vier Bänder. Berechnet man eine Gammaton-Filterbank mit 60 ERB-Bändern werden instabile Filter bis zum siebten Filterband berechnet. Durch die instabilen Filter kommt es zu unvorhersehbarem Verhalten in den anderen Darstellungen der Abb. 3.15. In [40], S.24 wird auf eine Korrektur zur Berechnung von Filterkoeffizienten hingewiesen. Durch die Korrektur werden Rundungsfehler bei der Koeffizientenberechnung für niedrige Frequenzen vermieden. Clark benutzt sowohl die Herleitung der Filterkoeffizienten, als auch die Anwendung der Filter auf Eingangssignale aus [35]. Scheinbar wurde hier aber nicht die korrigierte Version 2, aus [40] in die Implementation integriert.

In Abb. 3.17 sind die genannten Darstellungen aus der Implementation von C. Couvreur (Octave Toolbox) für 26 Bänder (als Terzband-Auflösung) zu sehen.

Diese Filterbank berechnet keine ERB-Bänder, sondern teilt das Signal in Terz- bzw. Oktavbänder. In einer vereinfachten Darstellung könnte man auch die Frequenzauflösung des auditorischen Systems darauf beschränken. Hier scheint die Filter-Berechnung unterhalb von 100 Hz instabil. Dies wird sowohl in den Darstellungen für das Frequenzverhalten (Abb. 3.17 oben links), als auch im aufsummierten Amplitudenspektrum (Abb. 3.17 oben rechts) im unteren Frequenzbereich deutlich. Oberhalb von 100 Hz hat die Filterbank in Bezug auf Frequenz-, Amplitudenspektrum und IR-Rekonstruktion keine Einschränkungen. Dies muss für spätere Anwendungen, wie z.B. in Kapitel 4.2 beachtet werden. Die Couvreur Filterbank wäre hier nur eine Ausweichmöglichkeit gewesen, falls keine der anderen Implementationen zu brauchbaren Ergebnissen geführt hätten.

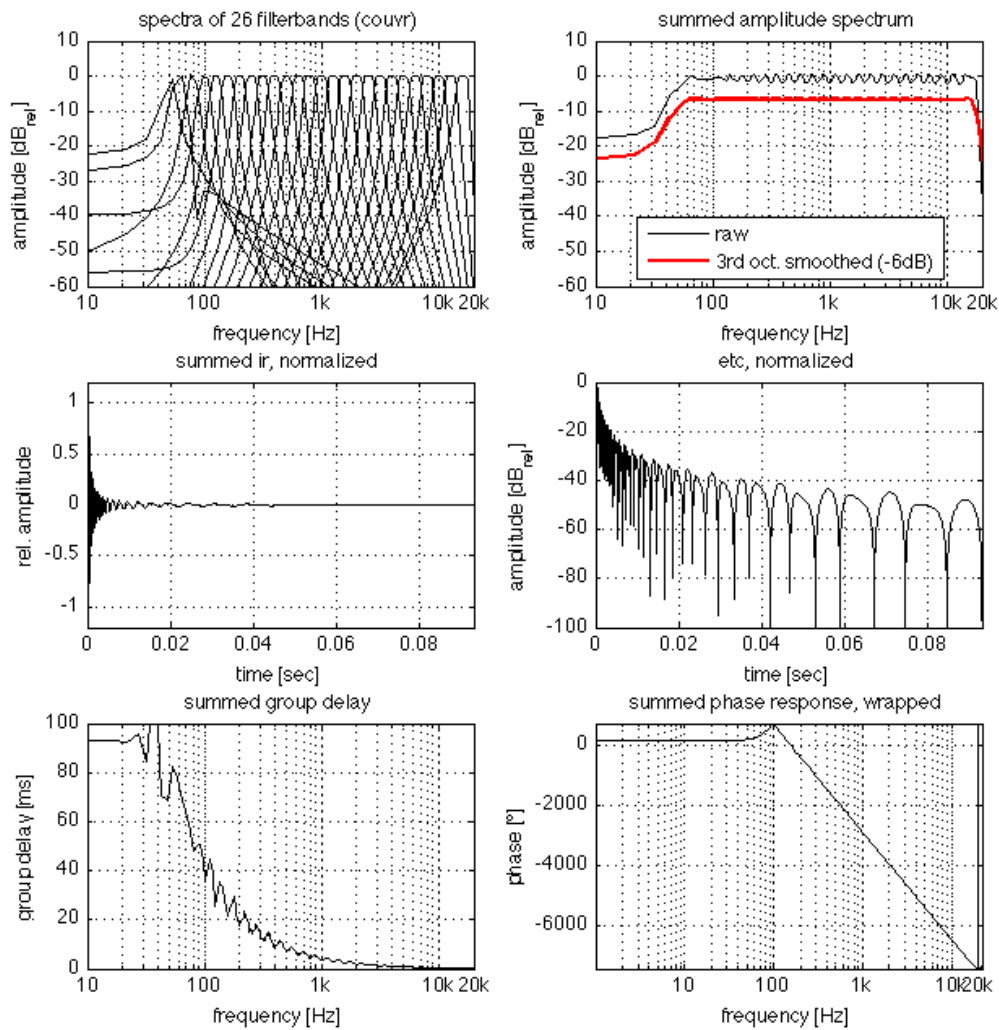


Abbildung 3.17: Rekonstruktionsverhalten der Filterbank von C. Couvreur

Die Filterbank von [40] erfüllt alle Kriterien, zeigt keine Einschränkungen bezüglich der Parameterwahl und plausibles Filterverhalten. Sie wird also im weiteren Verlauf für die Signalverarbeitung im auditorischen Modell bevorzugt und verwendet.

### 3.3.3 Kalibrierung des einfachen auditorischen Prozessmodells

Wie aus der Modulbeschreibung des einfachen Prozessmodells deutlich wird, kommt es an drei Punkten im Modell zu einer möglichen Reduktion. Die Reduktion durch die Hörschwelle ist festgelegt durch die Schalldruckpegel in der ISO-Norm. Die anderen zwei Reduktionsmöglichkeiten bieten einstellbare Parameter, die sich auf Hörbarkeitsentscheidungen auswirken. Zum einen gibt es den Schwellwert ‘at’ aus dem lokalen Entscheidungsapparat. Zum anderen gibt es den wählbaren Prozentsatz positiver Hörbarkeitsentscheidungen aus dem globalen Entscheidungsapparat.

Für die Kalibrierung des Modells werden Verdeckungskurven für mehrere ‘at’-Schwellwerte und Sensibilitätsstufen aus dem globalen Entscheider als Funktion von Verzögerungszeiten bzw. Einfallswinkeln für Impulspaare durch das auditorische Modell berechnet. Das Ziel ist es, durch entsprechende Einstellparameter das Modell so zu trainieren, dass die Verdeckungskurven aus empirischen Daten abgebildet werden.

Dafür werden Schallfeldparameter und Testparameter entsprechend den empirischen Untersuchungen definiert<sup>23</sup>. Weiterhin müssen hier die Parameter für das auditorische Modell eingestellt werden und das Modell auf die Impulspaare angewandt werden. Dazu wird eine MATLAB-Funktion benötigt, in der das implementierte auditorische Modell aufgerufen werden kann<sup>24</sup>.

Die Definition der Parameter wird im Folgenden für einen Fall beschrieben. Dabei soll die Messituation zur Ermittlung der Verdeckungskurve, wie in Abb. 2.4 aus den Untersuchungen von [16] nachgebildet werden.

- Schallfeldparameter:
  - Direktsignal und Testreflexion: Impulse mit gleicher Länge
  - Referenz-Schalldruckpegel des Direktsignals:  $ref_{dB} = 70dB_{SPL}$
  - Einfallswinkel der Testreflexion: ist fest aus  $-65^\circ$ -Azimut und  $0^\circ$ -Elevation
- Testparameter:
  - at-Schwellwert: von 0.6 bis 1.4, mit einer Schrittweite von 0.2
  - Verzögerungszeiten an diskreten Punkten: 2, 4, 6, 12, 15 und 20 ms
  - Pegel der Testreflexion: von 0 bis -100 dB in 1 dB Schritten

<sup>23</sup>Dies geschieht in einem MATLAB Skript, welches auf der beigelegten DVD im soundfield-Repository - 4 auditory soundfield processing - calibration - calibrate aud mod.m zu finden ist

<sup>24</sup>Die Funktion ist auf der beigelegten DVD im soundfield-Repository - 4 auditory soundfield processing - calibration - tools - reflections audible.m zu finden

- Einstellungen für das auditorische Modell
  - Filterbank: ‘Slaney93’
  - Anzahl der Filterbänder: 26
  - Empfindlichkeit des globalen Entscheiders: mittlere Empfindlichkeitsstufe<sup>25</sup>
  - Gleichrichter-Modus des inneren Haarzellenmodells: ‘full’

Die Verdeckungsschwellen, wie sie durch das auditorische Modell für mehrere ‘at’-Schwellwerte berechnet wurden, ist zusammen mit der Verdeckungskurve aus den empirischen Untersuchungen aus [16] in der linken Grafik von Abb.3.18 dargestellt. In der rechten Grafik wird zusätzlich der Fall einer Testreflexion aus vertikaler Richtung gezeigt. In der

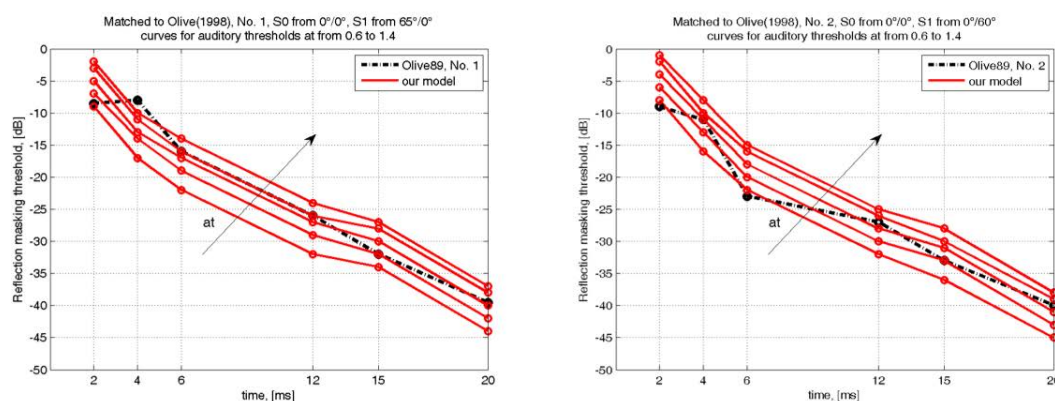


Abbildung 3.18: Verdeckungsschwellen als Funktion von Verzögerungszeit aus Modellberechnung (rot) und empirischen Untersuchungen (schwarz) aus [16], S.541, Fig.2 für Impulspaare. Links: Verdeckter von vorne, Testreflexion von 65° rechts. Rechts: Verdeckter von vorne, Testreflexion von 60° oben

linken Darstellung wird das Direktsignal aus frontaler Richtung und die Reflexion 65° rechts vom Empfänger präsentiert. Rechts ist die Verdeckungsschwelle für ein Direktsignal wieder aus frontaler Richtung und einer Testreflexion von 60° oben vom Empfänger zu sehen. Die schwarzen Kurven zeigen die Ergebnisse aus empirischen Untersuchungen, die roten Kurven zeigen das Modell. Die Abb. 3.18 zeigt, dass die Schwellen durch das Modell zunächst dem Verlauf aus empirischen Untersuchungen folgen. Der variable ‘at’-Schwellwert bewirkt eine mehr oder weniger parallele Verschiebung der Verdeckungsschwelle auf der Y-Achse. Somit hat man mit dem Schwellwert einen Modellparameter, mit dem man die Abweichung des Modells zu empirischen Verdeckungsdaten bestimmen kann. Dies gilt zunächst für den Fall von Verdeckung einzelner Impulspaare als Funktion von Verzögerungszeit. Als Ansatz für das kalibrierte Modell wird nun der Schwellwert so gewählt, dass das Modell

<sup>25</sup>‘mid sens’, d.h. positive Hörbarkeitsentscheidung wird getroffen, wenn mindestens ein Band im lokalen Entscheider als hörbar markiert wurde.

im überwiegenden Fall empfindlicher als laut empirischen Daten bekannt, reagiert. Daraus ergibt sich, dass das Modell für den ‘at’ – *Schwellwert* = 0.6 mit einer globalen mittleren Empfindlichkeit die kalibrierte Version darstellt.

Das Verhalten des Modells wurde ausserdem auf die Nachbildung von Verdeckungsschwellen als Funktion von Einfallswinkeln überprüft, wie in Abb. 2.7 aus Untersuchungen von [14] dargestellt. Dafür wurde, ähnlich der Untersuchung, das Wort ‘Zack’ aufgezeichnet und in das entsprechende Skript<sup>26</sup> eingelesen. Als Schallfeldparameter wurde daraus ein Impulspaar entsprechend der Länge des aufgezeichneten Files als Eingangssignale erzeugt. Der Referenz-Schalldruckpegel für das Direktsignal wurde auf  $ref_{dB} = 55dB_{SPL}$  und der Einfallswinkel in Elevation auf  $0^\circ$  (frontaler Einfall) gesetzt. In den Testparametern lag der Pegelbereich der Testreflexion wieder bei 0 bis -100 dB in 1 dB Schritten. Die Verzögerungszeit war entsprechend der Untersuchung festgelegt auf 50 ms. Die Verdeckungsschwelle wurde als Funktion des Einfallswinkels in horizontaler Ebene untersucht, so dass der Azimut-Winkel entsprechend einen Wertebereich von  $0^\circ$  -  $180^\circ$  in  $10^\circ$ - $20^\circ$ -Schritten erhält. Bei den Einstellparametern für das auditorische Modell wird die Empfindlichkeit des globalen Entscheidungsapparates auf hoch gesetzt und das innere Haarzellenmodell mit einer Gleichrichtung (‘full’) berechnet. Das Ergebnis ist in Abb. 3.19 zu sehen. In Abb. 3.19 sind die empirischen Kurven wieder in schwarz dargestellt. Die roten Kurven zeigen das Modell. In der oberen Darstellung wird der horizontale Einfallswinkel für das Direktsignal variiert. In der unteren Darstellung ändert sich der Einfallswinkel der Testreflexion. Auch wenn man wohlwollend behaupten könnte, dass die Kurven, wie sie durch das Modell berechnet wurden, in etwa dem Verlauf aus empirischen Daten folgen, konnte selbst für hohe ‘at’-Schwellwerte keine Übereinstimmung oder Annäherung der Verdeckungsschwellwerte erreicht werden. D.h. die Abbildung von Verdeckungskurven als Funktion von Einfallswinkel durch das auditorische Modell funktioniert hier nur bedingt. Es können für diesen Fall keine Modellparameter ermittelt werden, die zu einer kalibrierten Version führen. Daher gelten weiterhin die Parameter zur Abbildung von Verdeckungsschwellen als Funktion von Verzögerungszeit für ein kalibriertes Modell.

### 3.3.4 Erweiterung des einfachen auditorischen Modells auf Raumschallfelder

Das Ziel war es, mit dem Modell zu Entscheidungen über die Hörbarkeit einer Einzelreflexion in einem Raumschallfeld zu kommen. Das Direktsignal und die Einzelreflexionen liegen als minimalphasige IRs im Wave Format vor. Zusätzlich gehört zu jedem dieser Signale ein Metadatensatz, in dem Dämpfungsfaktoren relativ zum Direktsignal enthalten sind. Weiterhin enthält der Metadatensatz Ankunftszeit, Einfallswinkel in Azimut und Elevation und einen Wert zur Hörbarkeit für alle Signale. Sowohl der Ort der Schallfelddaten, als

---

<sup>26</sup>calibrate aud mod.m



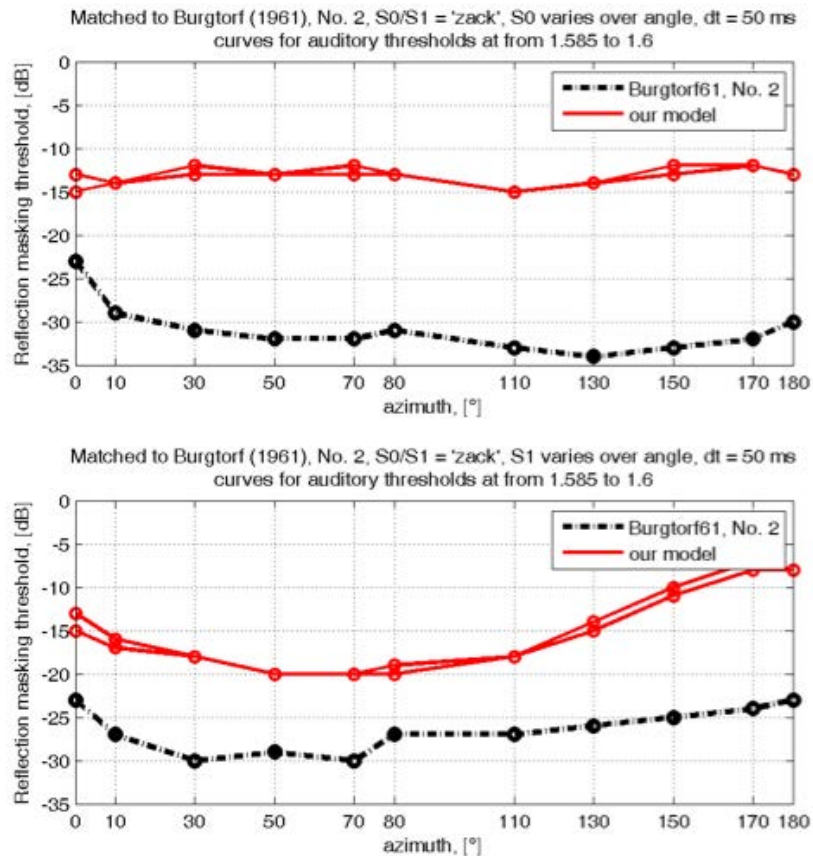


Abbildung 3.19: Verdeckungsschwellen als Funktion von Einfallswinkel aus Modellberechnung (rot) und empirischen Untersuchungen (schwarz) aus [14], S.107, Bild 22 für Impulspaare. Oben: Die Richtung des Verdeckers wird variiert, Unten: Die Richtung der Testreflexion wird variiert



auch der Metadatenatz muss dem Algorithmus<sup>27</sup> zur Verfügung stehen.

In dem Skript müssen Parameter für das erweiterte Modell festgelegt werden, mit denen die auditorische Verarbeitung realisiert werden soll.

Das erweiterte Modell bereitet zunächst das Direksignal auditorisch auf. Dazu wird das Schallfeld auf einen absoluten Schalldruckpegel ( $ref_{dB}$ )<sup>28</sup> normiert. Durch die Gewichtung mit dem HRIR-Datenatz und der korrekten Zuordnung der Verzögerung für linkes und rechtes Ohr wird ein BRIR-Datenatz angelegt, in dem beide Ohrkanäle nacheinander bearbeitet werden. Somit wird zunächst das Direksignal in einer Matrix für den BRIR-Datenatz gespeichert und in den Metadaten als hörbar markiert. Für die Betrachtung von darauffolgenden Reflexionen aus dem numerischem Schallfeld kommt es zu einer iterativen Erweiterung des Modells. Im Vergleich zu Abb. 3.12 ist die Erweiterung schematisch in Abb. 3.20 zu sehen. Wie in Abb. 3.20 dargestellt, wird zunächst die erste Reflexion au-

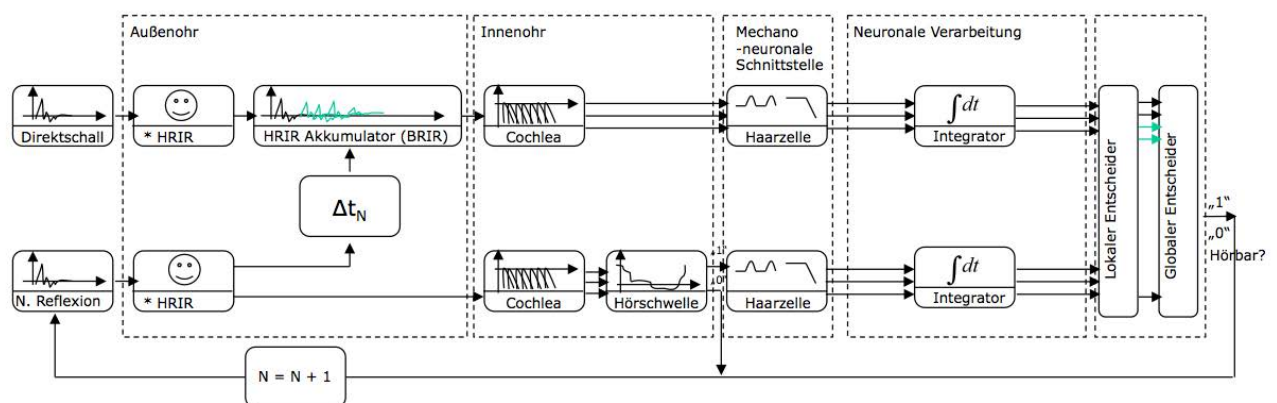


Abbildung 3.20: Erweitertes Prozessmodell zur Modellierung des Verdeckungsverhaltens in Raumschallfeldern

ditorisch aufgearbeitet. D.h. sie wird außenohrgewichtet in eine Filterbank aufgeteilt, auf Hörbarkeit in Bezug auf die Hörbarkeitsschwelle nach ISO-Norm überprüft, bandweise über gleitende Rechteckfenster gefiltert und durch den lokalen und globalen Entscheidungsapparat gegenüber dem Direksignal jeweils für linkes und rechtes Ohr auf Hörbarkeit überprüft. Wird die erste Reflexion als hörbar eingestuft, muss sie im auditorischen Gesamtschallfeld für die Betrachtung der nächsten Reflexion enthalten sein. Somit wird iterativ das Gesamtschallfeld jeweils um die hörbaren Reflexionen erweitert (im HRIR Akkumulator) und mit der nächstfolgenden Reflexion ( $N = N + 1$ ) bandweise verglichen. Im lokalen Entscheidungsapparat gilt die gleiche Regel, wie im einfachen Modell, nur dass der Verdeckter durch das kumulierte vorhergehende Gesamtschallfeld beschrieben und mit der nächstfolgenden Reflexion verglichen wird, ohne dass diese im Gesamtschallfeld enthalten ist. Das Gesamtschallfeld wird aus jeder Entscheidung für die individuelle Reflexion, je nach Hör-

<sup>27</sup>Das MATLAB-Skript für das erweiterte auditorische Modell liegt auf der beigefügten DVD im soundfield-Repository - 4 auditory soundfield processing - start AudMod.m

<sup>28</sup>zunächst normiert auf  $ref_{dB} = 85db_{SPL}$

barkeit, um die Reflexion erweitert oder eben nicht. Der Wert zur Hörbarkeit (0/1) im Metadatensatz wird je nach Entscheidung für jede Reflexion aktualisiert.

In Kapitel 3.2.3 wird beschrieben, das mit Aura Response berechnete Reflexionen nur bis zu dem Zeitpunkt ( $t_{mp}$ , siehe Gl. 3.5), der perzeptiv als früh bewertet wird aus dem Reflektogramm extrahiert werden. Der Prozess liefert demnach ‘nur’ ein auditorisch reduziertes Schallfeld von frühen Reflexionen.

### 3.3.5 Wirkungsweise des erweiterten auditorischen Modells

Mit der iterativen Erweiterung des auditorischen Prozessmodells können diskret beschriebene Raumschallfelder auf den perzeptiv relevanten Inhalt überprüft werden. Aus dem Kalibrierungsprozess für Impulspaare liegen Parameter für das Modell vor, mit denen empirische Daten zur Beschreibung von Verdeckung von Reflexionen als Funktion von Verzögerungszeit abgebildet werden können. Wie aus Kapitel 3.3.3 hervorging, dient vor allem der Schwellwert ‘at’ aus dem lokalen Entscheidungsapparat einer sinnvollen Änderung einzelner Hörbarkeitsentscheidungen. Somit wurden mit dem Skript ‘Start AudMod.m’ für mehrere solcher Schwellwerte Schallfeldreduktionsstufen berechnet. Anhand grafischer Darstellungen kann überprüft werden, welche Reflexionen in Abhängigkeit von ‘at’ als hörbar markiert werden.

Dazu werden sog. ETC-Plots und Richtungsigel erstellt<sup>29</sup>. In dem ETC-Plot wird die Energie des Schallfelds in dB relativ zum Direktschall über der Zeit dargestellt. In den Richtungsigelplots werden Direktsignal und Reflexionen in einem kartesischen Koordinatensystem entsprechend der sphärischen Koordinaten für die Einfallsrichtung in Azimut und Elevation und Pegel dargestellt. Die Reduktion früher Reflexionen durch das auditorische Modell wird in den folgenden Darstellungen deutlich.

In Abb. 3.21 sieht man links in rot das Originalschallfeld als Energie über der Zeit für die ersten 150 ms. In schwarz wird nur der Direktschall hervorgehoben. Rechts ist in einem Richtungsigelplot das Originalschallfeld mit räumlichen und energetischen Informationen mit in schwarz hervorgehobenem Direktschall dargestellt. Nun werden hörbare Reflexionen durch das Modell mit variablem Schwellwert berechnet. In der Abb. 3.22 ist das Schallfeld nach auditorischer Reduktion durch das kalibrierte Modell, also  $at = 0.6$  und mittlerer globaler Empfindlichkeitsstufe, in schwarz dargestellt. Durch das Modell werden nur sehr wenige Reflexionen als hörbar eingestuft. Aus dem Richtungsigelplot in Abb. 3.22 (rechts) wird deutlich, dass zwei Decken- und eine Seitenwandreflexion hörbar sind.

Durch eine weitere stufenlose Vergrößerung der Empfindlichkeit des Modells kann man zeigen, dass zunehmend mehr Reflexionen als hörbar eingestuft werden. Die Empfindlichkeit wird durch kleinere ‘at’-Werte erhöht. In der Abb. 3.23 werden hörbare Reflexionen mit  $at = 0.39$  gezeigt. Durch den kleineren ‘at’-Wert werden nun zehn Reflexionen als hörbar eingestuft. Es kommen zusätzlich Reflexionen aus frontaler Richtung und weiterer Seiten-

<sup>29</sup>Das entsprechende Skript und alle nötigen Daten zum Erstellen der Plots sind auf der beigelegten DVD im Ordner visualize - visualize stimuli double subplots.m zu finden

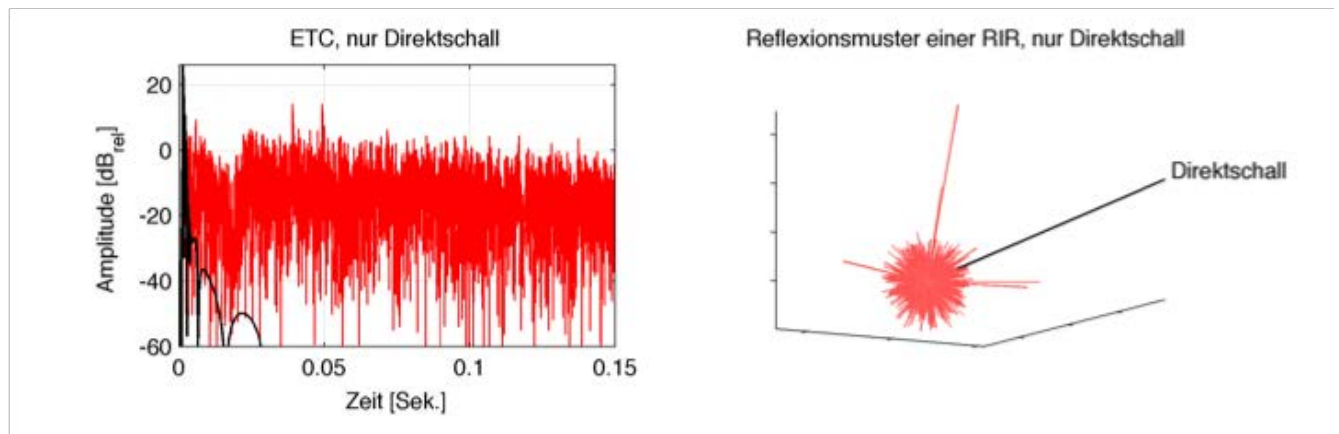


Abbildung 3.21: links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vollaufgelöstes Schallfeld, schwarz: nur Direktschall

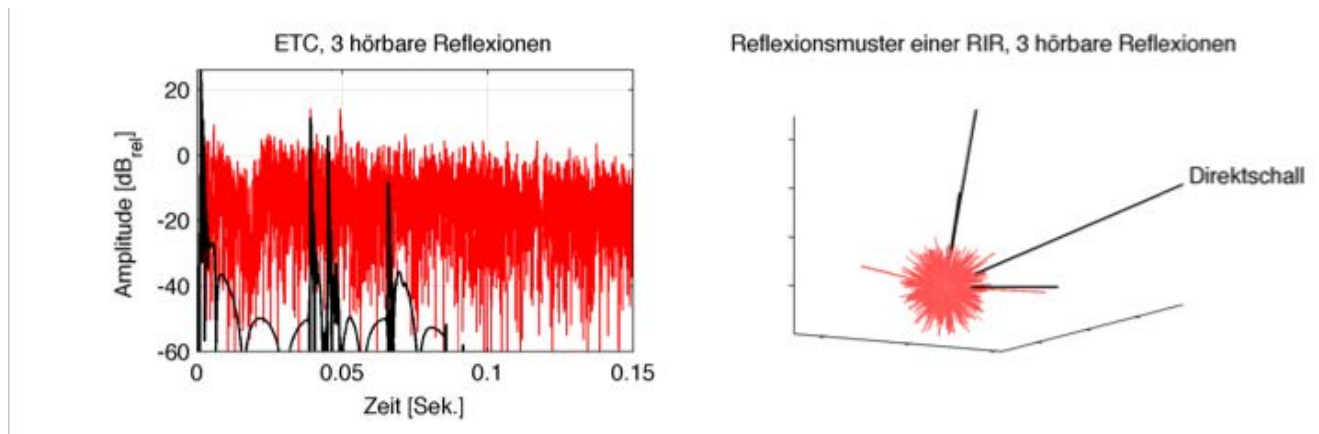
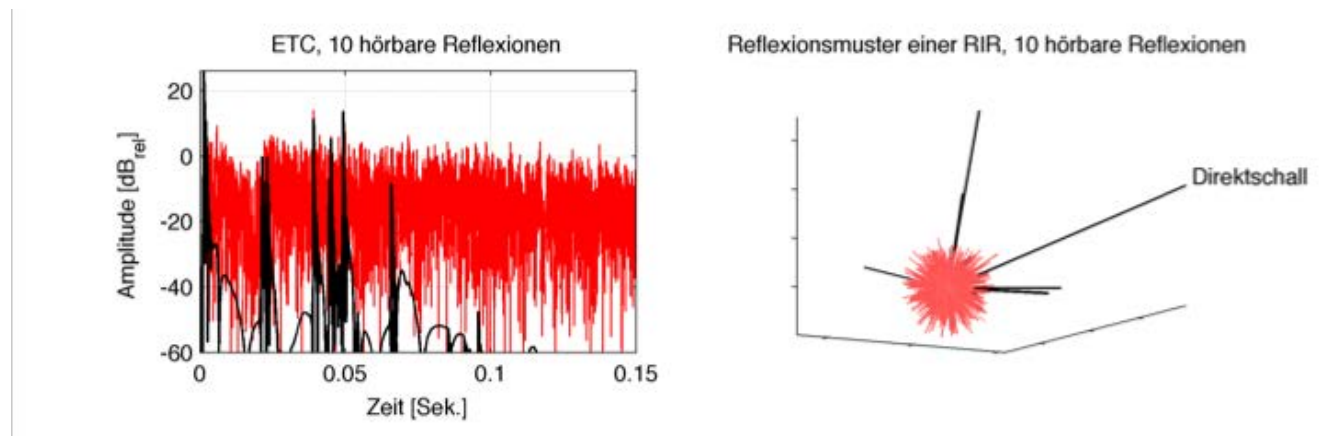


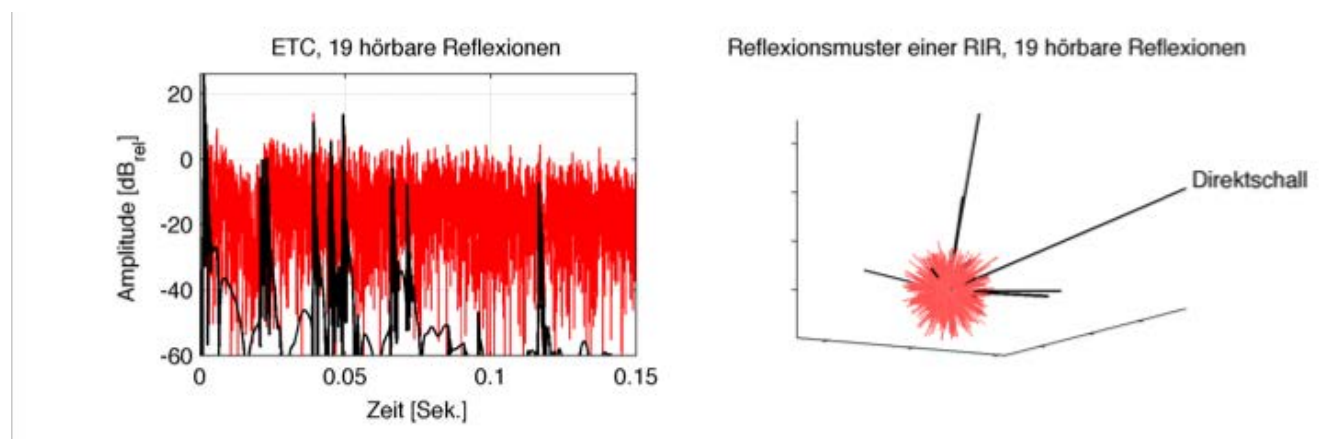
Abbildung 3.22: links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vollaufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit 'at' = 0.6



Abbildungung 3.23: links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vollaufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit  $at = 0.39$

wandreflexionen auch aus der anderen Richtung hinzu.

Eine Berechnung mit  $at = 0.345$  ergibt ein Reflexionsmuster, siehe Abb. 3.24. Wieder wird



Abbildungung 3.24: links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vollaufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit  $at = 0.345$

die Anzahl der Reflexionen durch die höhere Empfindlichkeit des Modells auf 19 Reflexionen erhöht. Hörbare Reflexionen werden aus zunehmend unterschiedlichen Richtungen detektiert. In Abb. 3.24 werden auch Reflexionen von der Rückwand in das Schallfeld mit einbezogen. Ausserdem zeigt die Abb. links in dem ETC-Plots hörbare Schallfeldkomponenten mit geringeren Pegeln, da die Zeitverzögerung zum Direktschall groß genug zu sein scheint.

In Abb. 3.25 wird das Schallfeld mit einem  $at = 0.228$  berechnet. In der Abb. besteht das perzeptiv reduzierte Schallfeld schon aus 263 Reflexionen, die in dem gesamten dargestell-

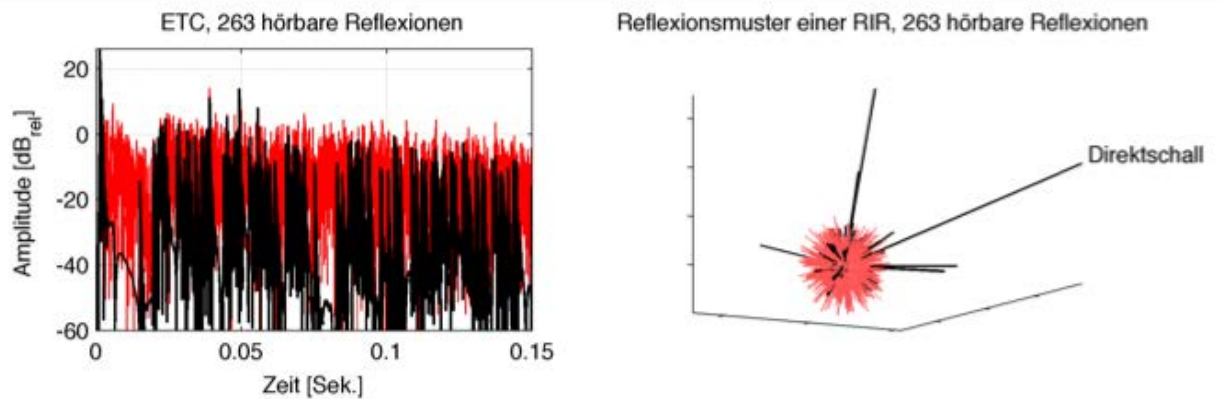


Abbildung 3.25: links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vollaufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit  $\text{'at'} = 0.228$

ten Zeitbereich aus scheinbar allen Richtungen kommen.

Die Auswahl hörbarer Reflexionen durch den Algorithmus ist nach Augenschein zunächst plausibel. Es werden prominente Reflexionen, wie sie aus der Schallfeldberechnung hervorgehen detektiert. Mit zunehmend höherer Empfindlichkeit des Modells steigt die Anzahl der Reflexionen. Zunehmend mehr Einfallsrichtungen nehmen einen hörbaren Einfluss. Nach genügender Verzögerungszeit werden auch Reflexionen mit geringerem Pegel als hörbar eingestuft.

# Kapitel 4

## Perzeptive Evaluation

In diesem Teil der Arbeit wird das auditorische Modell perzeptiv evaluiert. In Kapitel 3.3.5 konnte gezeigt werden, dass die Auswahl früher Reflexionen aus einem Schallfeld durch das Modell zu plausiblen Ergebnissen geführt hat. Ob diese objektive Auswertung auch auf die perzeptive Wahrnehmung zu übertragen ist, soll durch einen Hörversuch (HV) geklärt werden.

In dem HV soll die Frage beantwortet werden, ob ein Wahrnehmungsunterschied zwischen verschiedenen aufgelösten Schallfeldsimulationen, wie sie durch das auditorische Modell berechnet wurden, bestimmt werden kann. Die Ergebnisse aus dem HV sollen Aufschluss darüber liefern, ob und in wie weit man mit dem auditorischen Modell ein Schallfeld reduzieren kann, bis ein perzeptiver Unterschied bezüglich eines Referenz-Schallfeldes vorliegt. Insbesondere interessiert, ob ein perzeptiver Unterschied zwischen dem Referenz-Schallfeld und der Schallfeldauflösung, wie sie durch das an empirische Daten angepasste Modell berechnet wurde, festzustellen ist.

### 4.1 Auralisation - Binauralsynthese

Eine Evaluation des Reduktionsalgorithmus geschieht durch den Vergleich verschiedener Simulationsszenarien. Die Schallfeldbeschreibung in Form einer RIR muss dafür ‘auralisiert’ werden.

In [12] wird der Begriff Auralisation folgendermaßen beschrieben:

“Auralization is the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data.”[12], S.3.

Das Ziel der Auralisation ist also die “Hörbarmachung” von Räumen anhand numerischer Daten. Das Resultat der Auralisation ist ein hörbarer Eindruck einer Quelle innerhalb eines existierenden oder virtuellen Raumes. Der Raum wird durch die RIR beschrieben und der Empfänger kann durch die Anwendung der HRTF modelliert werden. In der HRTF sind spezifische, frequenzabhängige Färbung, wie sie durch Beugung und Reflexion

am Torso, Kopf und an den Aussenohren eines Hörers entstehen, und Laufzeit- und Pegelunterschiede, die sich aus der Trennung der beiden Ohren durch den Kopf ergeben, enthalten [27]. Für die Auralisation wird es daher nötig, sowohl die Eigenschaften des Raumes, als auch des Empfängers zu übertragen. Diese Informationen sind in der binauralen RIR enthalten. Zur Zeit stehen auralisierbare Datensätze in Form von BRIRs aus der Messung im Audimax mit FABIAN, eine vollaufgelöste Schallfeldbeschreibung aus dem raumakustischen Modell und Schallfeldreduktionsstufen aus modellgenerierten Daten durch das auditorische Modell zur Verfügung.

BRIRs gefaltet mit nachhallfreiem Audiosignal ermöglichen die Simulation eines hoffentlich identischen Höreindrucks mit beliebig wählbaren akustischen Inhalten.

### Binauralsynthese

Mit der dynamischen Binauralsynthese existiert ein Verfahren, dass einen direkten Vergleich (in Echtzeit) von perzeptiv plausiblen Simulationen mit den beschriebenen Methoden zum Erhalt der BRIRs erlaubt [42, 12]. Die binaurale Repräsentation eines Schallfeldes wird über Kopfhörer wiedergegeben. Alternativ kommen transaurale (lautsprecherbasierte) Verfahren zum Einsatz, die ein “Crosstalk Cancellation”-Verfahren anwenden. Das Ziel ist es, die gleichen Signale an den Ohren des Hörers zu reproduzieren, wie sie im realen Raum durch einen Hörer wahrgenommen werden. Bei der Binauralsynthese über die Kopfhörerwiedergabe werden die Signale am Trommelfell eines Hörers durch die Verwendung von BRIRs modelliert. Ein Monosignal kann durch die Faltung mit einem entsprechenden BRIR-Paar in jede Richtung verlagert werden [12].

Die perzeptive Qualität kann durch eine dynamische Nachführung der Simulation auf die Kopfbewegung des Hörers verbessert werden. Dies geschieht durch ein sog. Head-Tracking-System. Der Head-Tracker besteht aus einem Empfänger üblicherweise am Kopfhörer montiert und einem Transmitter vor oder über dem Kopfhörersystem angeordnet. Mit dem System wird die momentane Kopfposition ermittelt und die zur Position gehörenden BRIR-Daten mit dem Audiosignal gefaltet und reproduziert. Existierende Head-Tracking-Systeme basieren auf Ultrasound-, elektromagnetischen, optischen oder mechanischen Technologien. In dieser Arbeit wird ein elektromagnetisches System verwendet, wie es u.a. in [32] vorgestellt wird.

Das binaurale Syntheseverfahren über Kopfhörer, wie es in dieser Arbeit verwendet wurde, wird um die getrennte Verarbeitung zeitlich abgestimmter BRIRs und interauraler Zeitverzögerung (ITD<sup>1</sup>) erweitert. Damit können hörbare Kammfiltereffekte während dem Faltungsprozess bei der Überblendung von HRIRs in Abhängigkeit der Kopfposition vermieden werden. Das korrekte ITD-Verhältnis wird nach dem Faltungsprozess in dem daraus entstandenen BRIR-Datensatz separat und samplegenau wiedereingeführt. Dieser Ansatz wird in der Arbeit von [31] erläutert, wobei in der vorliegenden Arbeit die BRIRs nicht individualisiert werden, sondern nur die Trennung der ITD-Informationen während der Faltung angewendet wird, da auch diese schon zu perzeptiven Vorteilen, durch eine Reduktion

---

<sup>1</sup>engl. interaural time difference

von Schallartefakten führt. Die Individualisierung wird durch die Tatsache verhindert, dass die synthetisch erzeugten BRIRs nicht mit den HRTFs des Messroboters FABIAN erzeugt wurden, sondern mit dem Datensatz eines Vorgängermodells, für das die Korrekturformel evtl. nicht zu den selben Ergebnissen führen würde. Die Abb. 4.1 illustriert die dynamische Binauralsynthese mit getrennter ITD-Verarbeitung aus [31].

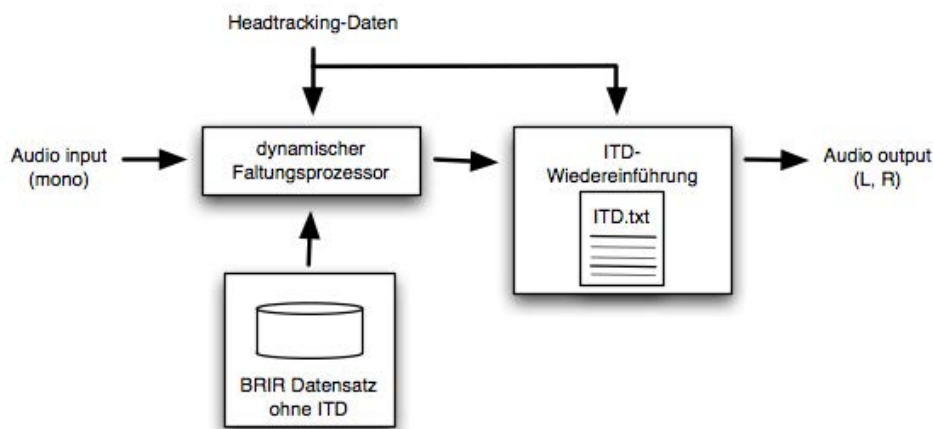


Abbildung 4.1: Schematische Darstellung der dynamischen Binauralsynthese mit getrennter ITD-Verarbeitung aus [31], wie Fig.4, S.7

## 4.2 Nachhallsynthese

Modellgenerierte Daten liegen nur für den frühen Teil der BRIR vor, da sie bis zur doppelten Mixingtime aus dem EASE-Reflektogramm extrahiert wurden. Für eine komplette BRIR muss den Daten noch der Nachhallbereich in geeigneter Form angehängt werden. Aufgrund von steigender Reflexionsdichte und Diffusität des ausklingenden Schallfeldes werden individuelle Reflexionen nicht mehr wahrgenommen, so dass dieser Bereich durch einen stochastischen Prozess beschrieben werden kann und durch einen für den gesamten Raum repräsentativen synthetischen Nachhall ersetzt werden kann.

Der Nachhallbereich (Tail) wird auf Basis eines weißen, gaussverteilten Rauschens modelliert<sup>2</sup>. Mit der `randn`-Funktion in MATLAB wird ein Vektor von weißem Rauschen erzeugt. Das Rauschen wird mit einer Diffusfeld-HRTF beaufschlagt, um einen korrekten Bezug vom diffusen Schallfeld des zu simulierenden Raumes herzustellen. Durch die Bildung der Diffusfeld-HRTF auf in folgender beschriebener Weise ist auch gleichzeitig die reale Höhendämpfung des späten Nachhalls korrekt mitmodelliert. Im Sinne einer vollständigen Simulation sollte man jedoch in Zukunft Diffusfeld-HRTF und Höhendämpfung

<sup>2</sup>siehe dazu die MATLAB-Funktion auf der beigefügten DVD im Ordner Nachhallsynthese - make stochastic tail.m



getrennt modellieren. Die Diffusfeld-HRTF wurde aus dem Nachhallbereich einer im Audimax gemessenen BRIR (linkes Signal) geschätzt<sup>3</sup>. Dazu wurde die gemessene BRIR als Wave-File eingelesen. In Abb.4.2 ist sie oben links als ETC-Plot (normalisiert) und oben rechts als Spektrum von der frontal gemessenen BRIR für Quelle 1 in voller Länge und unten für einen Zeitausschnitt von 100ms bis 1s gezeigt. Die Abb. unten links zeigt deutlich die bereits erwähnte Höhendämpfung. Das Amplitudenspektrum der zu verwendenden

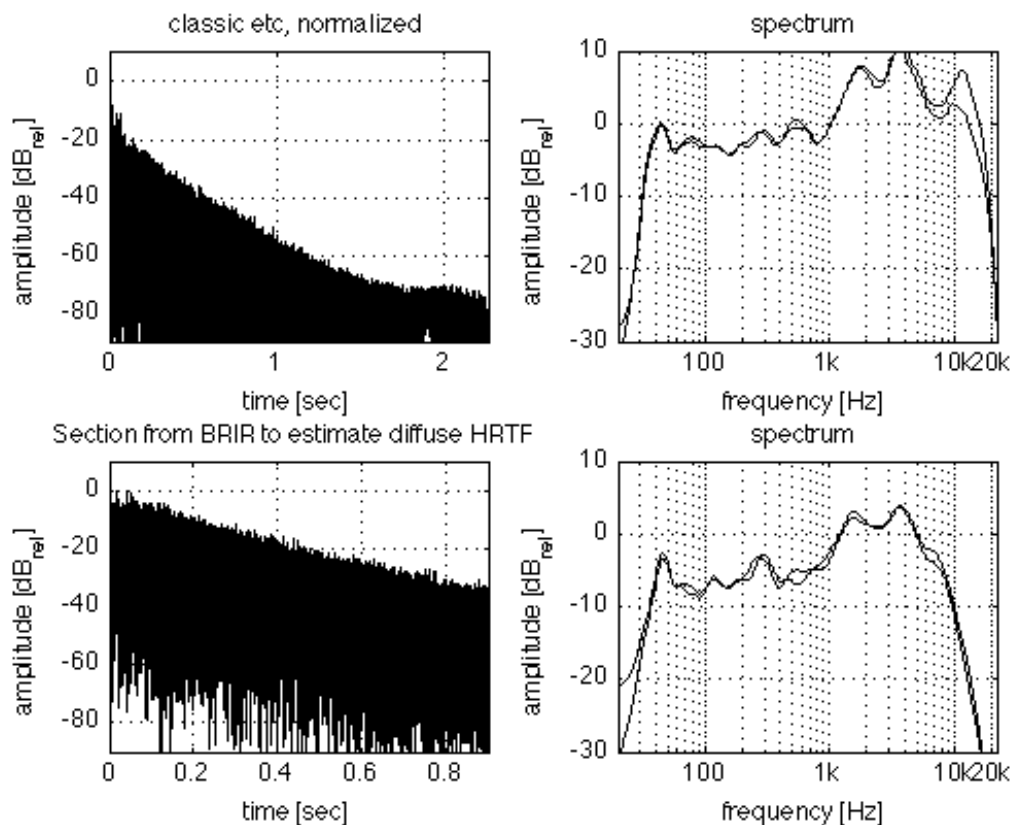


Abbildung 4.2: oben: modellgenerierte BRIR, links: ETC, rechts: Amplitudenspektrum, unten: geschätzte Diffusfeld-HRTF aus Zeitbereich 100ms bis 1s der BRIR, links: ETC, rechts: Amplitudenspektrum

Diffusfeld-HRTF wird durch den Zeitraum von 100ms bis 1s der linken BRIR beschrieben. Der untere Frequenzbereich wird bis ca. 500 Hz ‘flat’ gesetzt. Von dem Amplitudenspektrum wird ein minimalphasiger FIR-Filter erstellt. In der Abb. 4.3 ist oben das Amplitudenspektrum der geschätzten Diffusfeld-HRTF mit glattem unteren Frequenzgang von 100 ms bis 1 s zu sehen. In der mittleren Darstellung ist das Amplitudenspektrum des minimalphasigen Filters gezeigt. In der unteren Darstellung wird die Impulsantwort des minimalphasigen Filters dargestellt. Die Impulsantwort des minimalphasigen FIR-Filters

<sup>3</sup>Das entsprechende MATLAB-Skript ist auf der beigefügten DVD in dem Ordner Nachhallsynthese - estimate diffusefield HRTF old.m zu finden

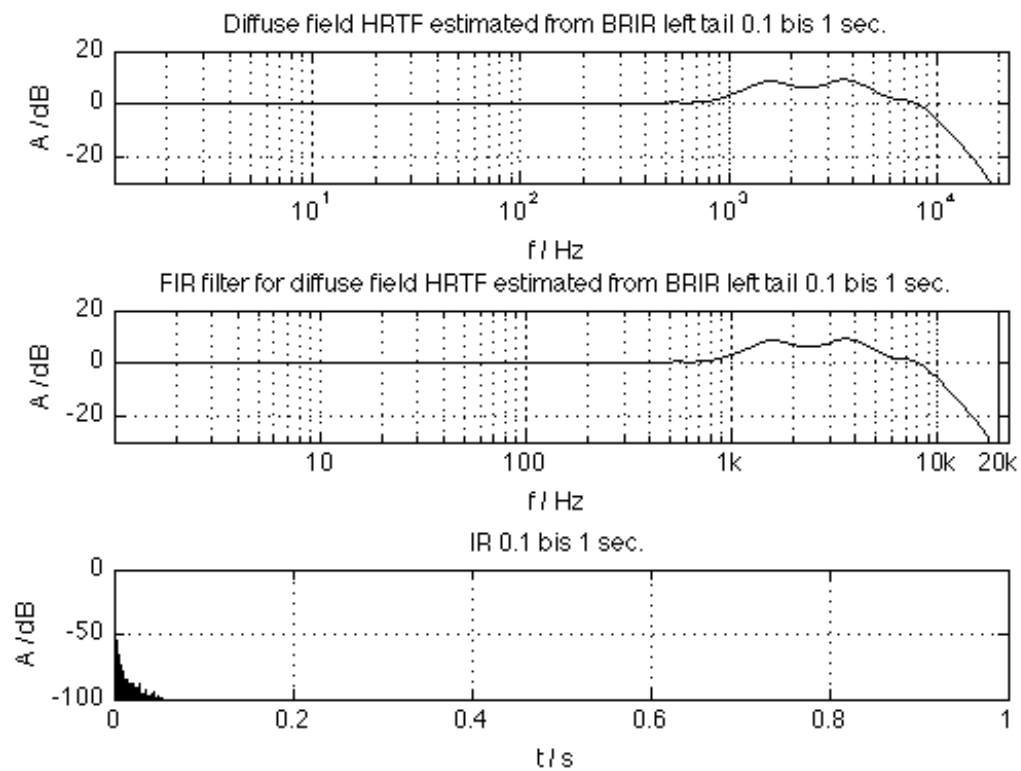


Abbildung 4.3: oben: geschätzte Diffusfeld-HRTF von BRIR-Ausschnitt (100ms bis 1s) mit geglättetem tiefen Frequenzbereich, mitte: FIR-Filter von Diffusfeld-HRTF, unten: Impulsantwort des FIR-Filters von 100 ms bis 1s

(siehe Abb. 4.3 unten) wird gespeichert und mit dem Rauschsignal gefiltert. Das mit der Diffusfeld-HRTF gefilterte Rauschsignal wird durch die Verwendung einer Filterbank in Terzband aufgelöste Bandpassfilter aufgeteilt. Dazu werden zunächst die Mittenfrequenzen und die dazugehörigen Nachhallzeiten aus einer Analyse im Messsystem Monkey Forest von real gemessenen Daten aus der Nachhallzeit-Messung im Audimax entnommen. Wobei die in Terzbändern aufgelösten Nachhallzeiten aus den Oktavband aufgelösten Nachhallzeiten interpoliert wurden. In Abb. 4.4 sind in der oberen Darstellung die Terzband aufgelösten Nachhallzeiten in schwarz und die über die Oktavbänder interpolierten Nachhallzeiten in rot zu sehen. In der unteren Darstellung sieht man die Oktavband-aufgelösten Nachhallzeiten, über die interpoliert wurde.

Die Filterkoeffizienten für die interpolierten Mittenfrequenzen werden mit der Couvreur

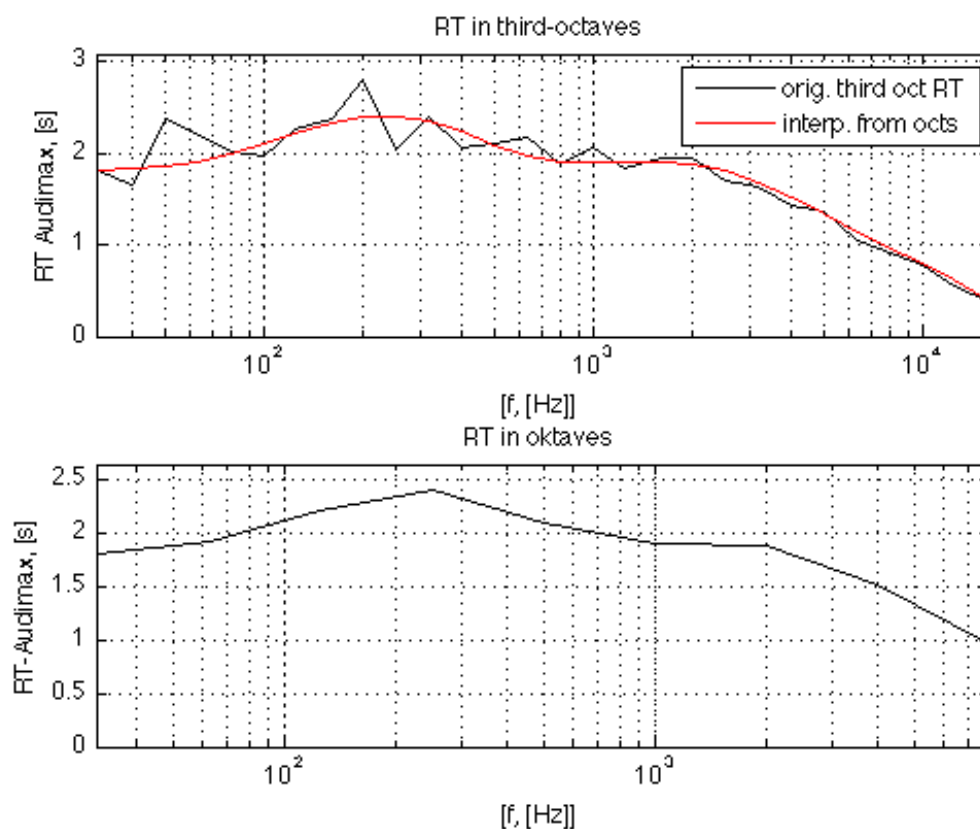


Abbildung 4.4: oben: terzbandaufgelöste Nachhallzeiten aus Messung im Audimax, schwarz: original gemessene Daten, rot: über Oktavbänder interpolierte Daten, unten: gemessene Nachhallzeiten in Oktavbänder aufgelöst, über die die obere rote Kurve interpoliert wurde.

Filterbank, wie aus Kapitel 3.3.2 berechnet und auf das mit der Diffusfeld-HRTF gefilterte Rauschsignal angewendet. Mit den Gleichungen in [13], S. 130-131, Gl. 5.7 - 5.9 werden

die zeitlichen Abklingraten (decay-rates) des Nachhallbereichs im Zeitbereich in der Form

$$d = \frac{0.161}{8} \cdot c_0 \cdot \frac{1}{T_{3rd}} \quad (4.1)$$

theoretisch berechnet und mit dem bandgefilterten Rauschen multipliziert.  $c_0$  ist die Schallgeschwindigkeit mit 343 m/s und  $T_{3rd}$  ist die Nachhallzeit pro Terzband in Sekunden. Aus der Summe der Terzbänder entsteht der ‘fullrange’ synthetisierte Nachhallbereich, der wiederum in doppelter Ausführung individuell (durch die randn-Funktion) für linkes und rechtes Ohr, wie für die BRIR nötig, berechnet wird. Für ein korrektes Direkt-/Diffusschall-Verhältnis werden die Effektivwerte im Zeitbereich (RMS-Level), der aus EASE erhaltenen unreduzierten nachbearbeiteten BRIRs und der beiden Tails berechnet<sup>4</sup>. Aus dem Verhältnis  $\frac{BRIR_{RMS}}{Tail_{RMS}}$  wird ein Gain-Faktor gebildet, der jeweils mit den Tails multipliziert wird, um die Pegel der Tails entsprechend einem korrekten Verhältnis zu kalibrieren. In Abb.4.5 wird ein ETC-Plot des synthetisierten Tails in schwarz und der unreduzierten BRIR aus EASE in rot für den linken und rechten Kanal gezeigt. In der Abb. ist zu sehen, dass die Abklingraten und der Pegel der Tails (schwarz) mit der berechneten BRIR ausreichend übereinstimmen.

Zur Kombination mit dem ‘sparsam besetzten’ auditorisch reduzierten frühen Schallfelds wird der modellierte Tail mit einer linearen Fade-In Funktion (siehe hierzu u.a. [43]) multipliziert, so dass die Reflexionsdichte-Zunahme über der Zeit simuliert wird und ein sprunghafter Übergang von reduziertem frühen BRIR-Teil zum stochastischen Nachhallbereich vermieden werden kann. Dieser Fade-In ist in Abb. 4.6 zu sehen. In dieser Darstellung wird in rot der Nachhallschwanz mit dem Fade-In und in schwarz die unreduzierte BRIR aus EASE dargestellt. Die lineare Fade-In Funktion zeigt sich über der logarithmischen Amplituden-Darstellung als exponentielle Funktion.

Mit der beschriebenen Nachhallsynthese wird ein Tail jeweils für den linken und rechten Kanal erstellt. Er kann für die Auralisation, der mit dem auditorischen Modell reduzierten BRIR-Datensätzen, verwendet werden.

### 4.3 Vorversuch

In einem (internen) Vorversuch sollte die Frage geklärt werden, wie die objektiven Unterschiede, d.h. die Schallfelder in unterschiedlichen auditorischen Reduktionsstufen aus Kapitel 3.3.5 hörbar gemacht werden können. Dafür mussten Entscheidungen über die genaue Anzahl und Auflösung von zu vergleichenden Reduktionsstufen aus den auralisierbaren Datensätzen getroffen werden und bedingt durch das Versuchsdesign, welches im in Kapitel 4.4 genauer erläutert wird, das Referenz-Schallfeld bestimmt werden. Dazu wurde ein direkter Vergleich zwischen BRIR-Datensätzen der sog. Realität, des vollaufgelösten

---

<sup>4</sup>siehe dazu das MATLAB-Skript create MODELLEDtail for mult sources HV Roehrbein.m auf der beigefügten DVD im Ordner Nachhallsynthese

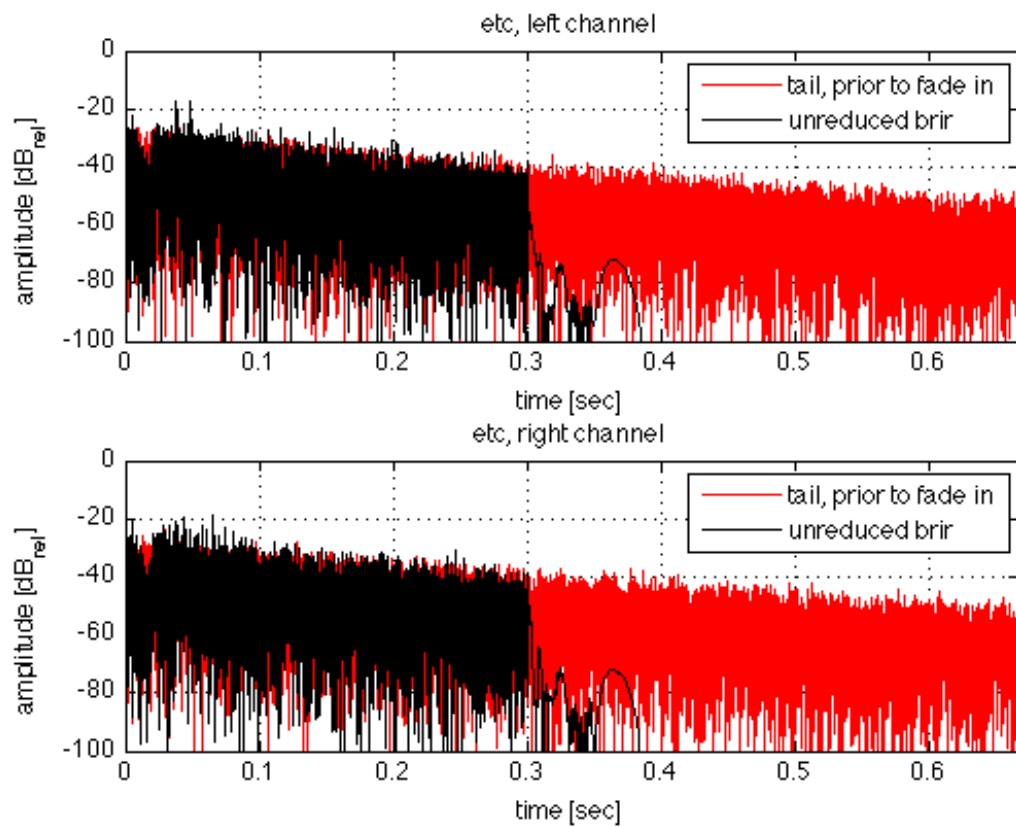


Abbildung 4.5: ETC-Plots von modelliertem Nachhall und unreduziertem modellgenerierten BRIR-Datensatz, oben: linker Kanal (Ohr), unten: rechter Kanal (Ohr)

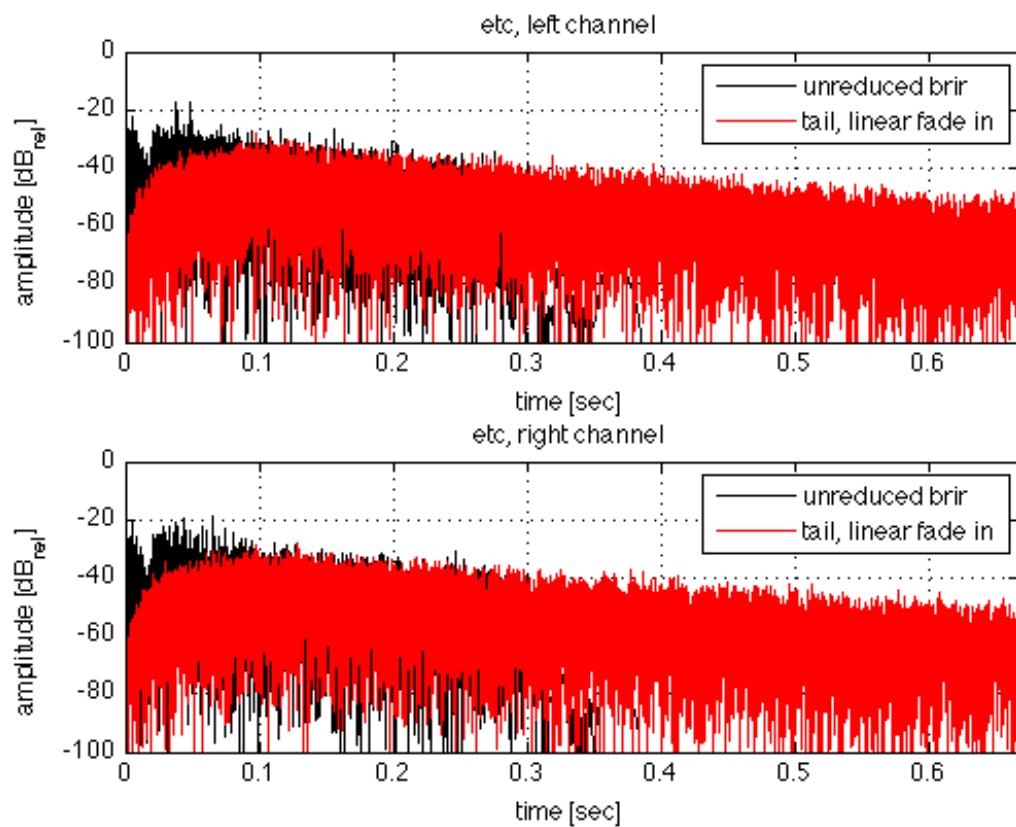


Abbildung 4.6: ETC-Plots von modelliertem Nachhall (rot) und unreduziertem modellgenerierten BRIR-Datensatz (schwarz) mit linearem Fade-In, oben: linker Kanal (Ohr), unten: rechter Kanal (Ohr)

modellgenerierten Datensatzes und einer repräsentativen Auswahl von reduzierten Datensätzen durchgeführt.

Für den Vorversuch wurde ein puredata-Patch implementiert, welches über das OSC-Protokoll mit der Faltungssoftware fWONDER kommuniziert. fWONDER realisiert sowohl die latenzfreie Wahl eines BRIR-Datensatzes und eines Stimulus', als auch die Faltung der Daten inklusive modelliertem Nachhall, um den virtuellen Raumeindruck über Kopfhörer wiedergeben zu können. Wie für den Hörversuch standen auch hier drei Stimuli zur Verfügung, die in Kapitel 4.4.2 noch genauer erläutert werden. Der perzeptive Vergleich wurde durch den Projektleiter Alexander Lindau und den Autor der Arbeit vorgenommen. Die Ergebnisse führten zur Bestimmung des Referenzschallfeldes und der zu vergleichenden Reduktionsstufen. Sie wurden jedoch nicht weiter protokolliert.

Die perzeptiven Eindrücke der Beteiligten ergaben, dass die sog. Realität ganz unähnlich zu allen Simulationen war. Dies war z.B. wegen des modellierten Nachhalls in den Simulationen zu vermuten. Ähnlich deutliche Abweichungen von Realität und numerischer Simulation mit EASE sind z.B. auch schon in der Masterarbeit von J. Blickensdorff [33] beschrieben worden. Es war jedoch nicht Gegenstand dieser Untersuchung, modellgenerierte Daten gegenüber gemessenen Daten qualitativ zu bewerten, so dass die Realität als Referenz-Schallfeld zunächst ausgeschlossen wurde. Weiterhin ergab sich aus den Vorversuchen, dass auch das vollaufgelöste modellgenerierte Schallfeld sehr unähnlich zu allen Simulationen in der Nähe der durch das kalibrierte Modell vorhergesagten Reduzierbarkeit war. Auch wenn der Nachhall für alle Simulationen identisch war. Es zeigt sich hier, dass die Pegelung des diffusen Nachhalls im Verhältnis zum frühen Teil des Schallfeldes individuell an vollaufgelöste Schallfelder und reduzierte Varianten angepasst werden sollte. Die Bewertung der Reduktionsstufen gegenüber einem Referenz-Schallfeld würde stets von dieser Unähnlichkeit beeinflusst werden, so dass ein vollaufgelöstes modellgeneriertes Schallfeld auch nicht als Referenz-Schallfeld in Frage zu kommen scheint.

Zu den modellgenerierten Daten muss erwähnt werden, dass in EASE mit der Berechnung durch Aura Response ein vollaufgelöstes Schallfeld sich aus ca. 3200 Reflexionen bis zum Zeitpunkt  $t = 300ms$  zusammensetzt. Nach [13], S.110, Gl. (4.5) kann die theoretische Reflexionsdichte mit

$$\frac{dN_r}{dt} = 4\pi \frac{c^3 t^2}{V} \quad (4.2)$$

berechnet werden. Integriert man die Gl. 4.2 über  $t$ , erhält man die Anzahl der Reflexionen zum genannten Zeitpunkt in der Form

$$N_r = \int_0^t \frac{dN}{dt} dt = \frac{4\pi}{3} \frac{c^3 t^3}{V} \quad (4.3)$$

Geht man nun, wie im Raumakustikmodell von einem Raumvolumen  $V = 8500m^3$  und einer Schallgeschwindigkeit von  $c = 343m/s$  aus, ergibt sich eine theoretische Anzahl von Reflexionen am Zeitpunkt  $t = 300ms$  von  $N_r \approx 520$ . Diese hohe Diskrepanz zwischen modellgenerierten Daten und theoretisch berechneter Anzahl kann ein Grund für den hohen perzeptiven Unterschied gegenüber reduzierten Datensätzen sein. Die Ursache für die Diskrepanz bleibt jedoch ungeklärt, auch wenn aus Gesprächen mit den Entwicklern von

EASE bestätigt wurde, dass der theoretische Ansatz von [13] bei der Berechnung mit Aura Response berücksichtigt wird. Es gilt weiterhin die Frage zu beantworten, wie sich die perzeptiven Unterschiede zwischen den reduzierten Datensätzen auswirken. Da die Vorhersage des kalibrierten auditorischen Modells auf ein sehr stark reduziertes (nur Direktschall plus drei Reflexionen) vorhergesagtes Schallfeld führte, wurde sich darauf zurückgezogen die Unterscheidbarkeit gegenüber weniger stark reduzierter Schallfeldversionen, die aber dabei nicht mehr als 500 Reflexionen enthielten, zu prüfen.

Weiterhin zeigte sich in den Vorversuchen, dass erst nach deutlicher Reduktion der Anzahl früher Reflexionen ein Bereich gefunden werden konnte, in dem eine zunehmende Reflexionsdichte zu einem konstanten Klangeindruck führte. Die Hörbarkeit der verschiedenen Schallfeldreduktionsstufen waren sehr stark von der Pegelung des Diffushalls abhängig. Die exakte Herstellung des Diffusschallpegels, wie in Kapitel 4.2 dokumentiert, ist wesentlich für eine korrekte Vorhersage der Hörbarkeit der Reflexionen.

Der Bereich eines konstanten Höreindrucks ergab sich bei einem BRIR-Datensatz mit 263 Reflexion, wie er durch das auditorische Modell mit einem *at*-Schwellwert von 0.228 berechnet wurde, plus dem modellierten Nachhall. Dieser BRIR-Datensatz wurde schließlich als Referenz-Schallfeld bestimmt.

In der Tabelle 4.1 sind die weiteren Reduktionsstufen aufgeführt, die im Hörversuch zur Unterschiedlichkeitsbewertung angeboten wurden. Sie entsprechen den Stufen, wie sie in

Raum	'at'-Schwellwert	Anzahl der Reflexionen ( $N_r$ )
1	0.228	263 (Referenz)
2	0.273	76
3	0.345	19
4	0.39	10
5	0.5	4
6	0.6	3
7	$\infty$	nur Direktsignal

Tabelle 4.1: Verwendete Reduktionsstufen im Hörversuch berechnet durch variablen '*at*'-Schwellwert im auditorischen Modell und daraus resultierende Anzahl von Reflexionen  $N_r$

Kapitel 3.3.5 zur objektiven Auswertung anhand grafischer Darstellungen berechnet wurden.

Raum 7 besteht nur aus Direktsignal plus modelliertem Nachhall. Er stellt einen Ankerreiz dar, mit dem versucht wurde, das untere Ende der Bewertungsskala eindeutig zu definieren. Bei der Auswahl der reduzierten Schallfelder wurde versucht, einerseits eine ungefähr logarithmisch gleichmäßig zunehmende Abstufung der Reflexionszahlen zu erhalten und andererseits die Anzahl von Bedingungsvariationen (Reduktionsstufen) gering zu halten, damit die Versuchspersonen (VP) bei der Unterschiedlichkeitsbewertung im Hörversuch nicht unnötig überfordert werden.



## 4.4 Versuchsbeschreibung

### 4.4.1 Versuchsmethode und -design

Als Methode zur Beantwortung der eingangs in diesem Kapitel genannten Fragestellung wurde ein ABC/HR-Test gewählt. Dabei handelt es sich um ein Forced-Choice-Verfahren, bei dem einer Versuchsperson drei Stimuli (A,B,C) zur Auswahl stehen. Die bekannte Referenz ist stets als Stimulus A zur Verfügung. Die versteckte Referenz und eine Schallfeldreduktionsstufe (das Objekt) stehen gleichzeitig der VP zur Verfügung, sind jedoch zufällig B oder C zugeordnet. Von der VP wird eine quantitative Bewertung der Unterschiedlichkeit beim Vergleich von B zu A und C zu A anhand einer kontinuierlichen Intervallsskala erwartet. Die Skala geht in diesem Fall von identisch (5.0) bis sehr unterschiedlich (1.0). Jeglicher wahrgenommener Unterschied zwischen Referenz und den anderen Stimuli kann zur Bewertung benutzt werden. Die Zuordnung der Reize B und C erfolgt dabei zufällig, wie es das Versuchsdesign vorsieht. Es handelt sich in diesem speziellen Fall um ein Mehrfach-Stimulus-Verfahren, da pro Audiostimulus alle sieben Reduktionsstufen als Bedingungsvariationen in einem Durchgang gleichzeitig bewertet werden. Da für alle ungekennzeichneten Stimuluspaare eine Bewertungsskala zur Verfügung steht, kann statt des veränderten Stimulus' auch die Referenz von der VP als verändert eingestuft werden. Dies entspricht einer objektiv falschen Antwort und lässt den Schluss zu, dass Referenz und Vergleichsreiz nicht unterschieden werden konnten.

Zur Auswertung von ABC/HR-Tests werden Differenzgrade 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 Benutzeroberfläche für das beschriebene Versuchsdesign ist in der Abb.4.7 zu sehen.

In der Abb.4.7 wird Stimulus A, also die Referenz, als 'Ref' bezeichnet. Die versteckte Referenz und das Objekt (B und C) verbergen sich hinter den beiden 'Play'-Buttons.

Aus der Darstellung wird deutlich, dass die Bewertung in einem ABC/HR Test mehrfach (7x) durchgeführt wird. Der einfache ABC/HR Test besteht nur aus einem Vergleichspaar mit der Referenz.

Der Mehrfach-ABC/HR-Test wird für drei Audiobeispiele in zufälliger Reihenfolge wiederholt. Die Audiobeispiele werden jedoch innerhalb eines 'Panels' nicht variiert. Der Bewertungsprozess kann solange fortgeführt werden, bis die Versuchsperson meint, alle Reizpaare beurteilt zu haben und die Werte des absoluten Urteils relativ gesehen stimmig sind, so dass eine Rangfolge unter den Bedingungsvariationen hergestellt wird.

Empfohlene Richtlinien zum ABC/HR-Verfahren sind in der Norm [44] beschrieben. Die Benennung der Ratingskala wurde für den Versuch verändert, da die vorgeschlagenen Bezeichnungen nicht der Fragestellung des Versuchs entsprechen. Weiterhin wurden die Zwischenstufen der Skala entgegen der Norm nicht beschriftet. Es wurde lediglich der Mittelpunkt und die Endpunkte der Skala markiert, um eine Orientierung für die VP zu vereinfachen. Ein Problem tritt am unteren Ende der Skala auf, da die subjektive Auffas-

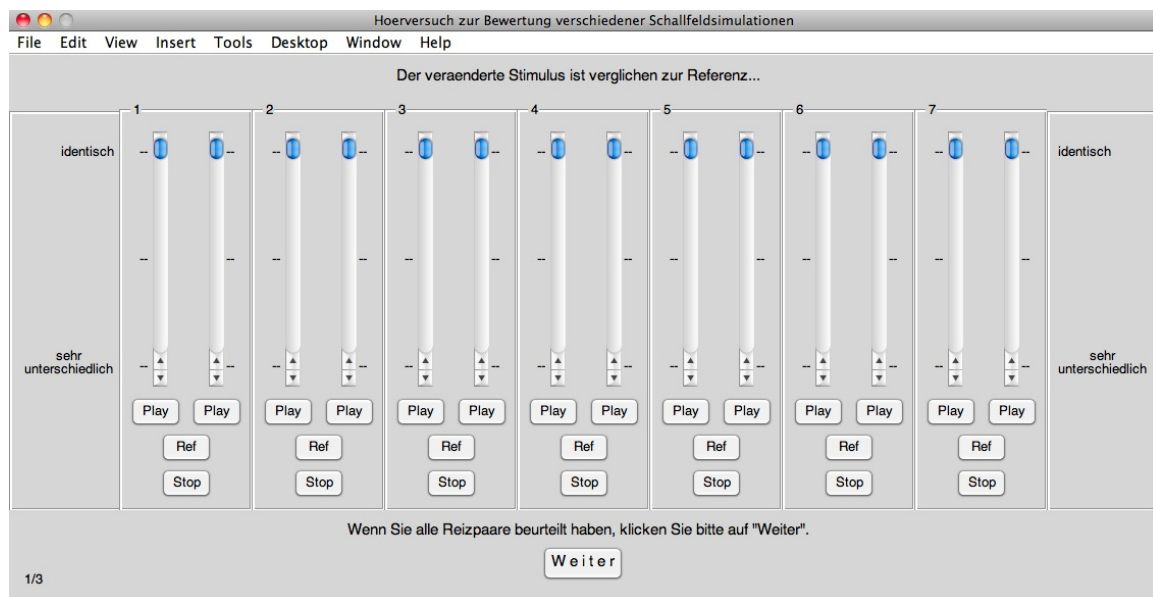


Abbildung 4.7: Hörversuchspanel

sungen der VPs gegenüber dem Begriff “sehr unterschiedlich” variieren kann. Somit ist das untere Extrem nicht eindeutig definiert. Um dem entgegenzuwirken, wurde ein versteckter Ankerreiz in den Hörversuch integriert, der sich nur aus Direktsignal und modelliertem Nachhall zusammensetzt. Man kann also davon ausgehen, dass die Bewertung des Ankers jeweils den tiefsten Punkt einer subjektiven Skala markieren wird.

In dem Hörversuch ergeben sich aus den drei Testreizen und den sieben Reduktionsstufen  $7 \times 3 = 21$  Bedingungsvariationen. Die abhängigen Variablen (Faktoren) sind dabei die verschiedenen aufgelösten Schallfeldvarianten und der Audioinhalt. Die unabhängige Variable ist die Bewertung, die auf der Ratingskala abgegeben wird, wobei aus ihr entnommen werden kann, ob ein reduzierter Raum identifiziert werden kann. Da alle 21 Bedingungsvariationen von allen VPs beurteilt werden, handelt es sich um einen zweifaktoriellen Versuch mit Messwiederholungen und abhängigen Stichproben. Die Teststärke dieses Messwiederholungsdesigns ist als außerordentlich hoch einzustufen, wenn die Anzahl der VPs entsprechend hoch ist. Für eine ausreichende Teststärke muss die Anzahl von mindestens 17 VPs erreicht werden. Dies wurde mit gPower<sup>5</sup> [45] für mittlere Effekte bezüglich des Faktors ‘Schallfeldreduktion’ berechnet.

## 4.4.2 Stimuli

Perzeptive Unterschiede der reduzierten Schallfeldsimulationen werden in Abhängigkeit vom Audioinhalt untersucht. Die Wahl des Audioinhalts bildet daher eine wichtige Grundlage für die Bewertung im Hörversuch. Es sollte ein Spektrum an Audiomaterial angeboten

<sup>5</sup><http://www.psych.uni-duesseldorf.de/aap/projects/gpower/index.html>, Zugriff: 11.09.2011

werden, welches sich untereinander sowohl in der zeitlichen und spektralen Zusammensetzung, als auch in der räumlichen Abbildung unterscheidet.

Wie schon in Kapitel 2.2.1 festgestellt werden konnte, stellt ein impulsartiges Signal einen besonders kritischen Stimulus dar, um die Wahrnehmung von Raumschallfeldkomponenten zu untersuchen. Somit wurde in dem Hörversuch ein Impulstrain dargeboten, bei dem sich synthetisch hergestellte Pulse über einen Zeitraum von 30 Sekunden wiederholten. Weiterhin sollte die Bewertung anhand von natürlichen Stimuli geschehen. Daher wurden nachhallfreie Aufnahmen von einem männlichen deutschen Sprecher und einem klassischen Streichquartett einer Mozartkomposition für die Simulationen verwendet. Musiksignale stellen spektral komplexe und im Zeitbereich kontinuierliche Signale dar, so dass marginale raumspezifische Unterschiede schwer zu ermitteln sein werden. Das Sprachsignal wurde als Zwischenstufe von kritischen Impulssignalen und dem kontinuierlichen Musiksignal ausgewählt.

### 4.4.3 Versuchsaufbau und -durchführung

Der Hörversuch fand im kleinen Studio des Fachgebiets Audiokommunikation der TU-Berlin statt. Die Abb.4.8 zeigt schematisch den dortigen technischen Aufbau. Die Benutzerschnittstelle wurde in Anlehnung an die Arbeit von [30] in MATLAB implementiert und entsprechend diesem Hörversuch angepasst<sup>6</sup>. Das dabei entstandene HV-Panel (siehe Abb.4.7) sendet in Abhängigkeit vom gewählten Reiz A (Ref), B oder C (2xPlay) OSC-Steuerdaten über eine Netzwerkverbindung an den Faltungsrechner. Durch Drücken eines der drei Buttons wird das aktuelle Klangbeispiel mit den BRIR-Datensätzen von allen Räumen und dem modellierten Tail in der Faltungssoftware fWONDER parallel gefaltet. Nach der Wiedereinführung des korrekten ITD-Verhältnisses zwischen binauralen RIRs und einer weiteren Faltung mit den Kompensationsfiltern für den Kopfhörer, ebenfalls in fWONDER, wird das Klangmaterial über einen dynamischen Kopfhörer (AKG K-601) der Versuchsperson dargeboten.

Je nach gewähltem Reiz werden jeweils die nicht aktivierten Räume stumm geschaltet. Die Realisierung der OSC-Schnittstelle im Faltungsrechner, die Wiedergabe und das Senden der Klangbeispiele an die Faltungseingänge, sowie die automatische Stummschaltung erfolgte in der Opensource Software puredata<sup>7</sup>.

Der Headtracker ist zur Feststellung der Kopfposition der VP direkt am Kopfhörerbügel befestigt. Er sendet fortlaufend magnetische Positionsdaten an einen Empfänger, der wiederum Informationen an die Faltungssoftware zur korrekten Wahl eines zur Kopfrichtung der VP passenden BRIR-Paares von dem aktivierten Raum übermittelt, wo diese dann parallel gefaltet werden. Die parallele Faltung aller Räume ermöglicht ein artefaktfreies Umschalten zwischen den Räumen. Die Wiedergabe des aktuellen Klangbeispiels konnte

---

<sup>6</sup>Das entsprechende MATLAB-Skript ist auf der beigelegten DVD unter HV Roehrbein - HV Implementation - abchr.m zu finden

<sup>7</sup>siehe <http://puredata.info/> Zugriff 04.08.2011, der pd-Patch ist auf der beigelegten DVD in HV Roehrbein - HV Implementation - HVRoehrbein.pd zu finden

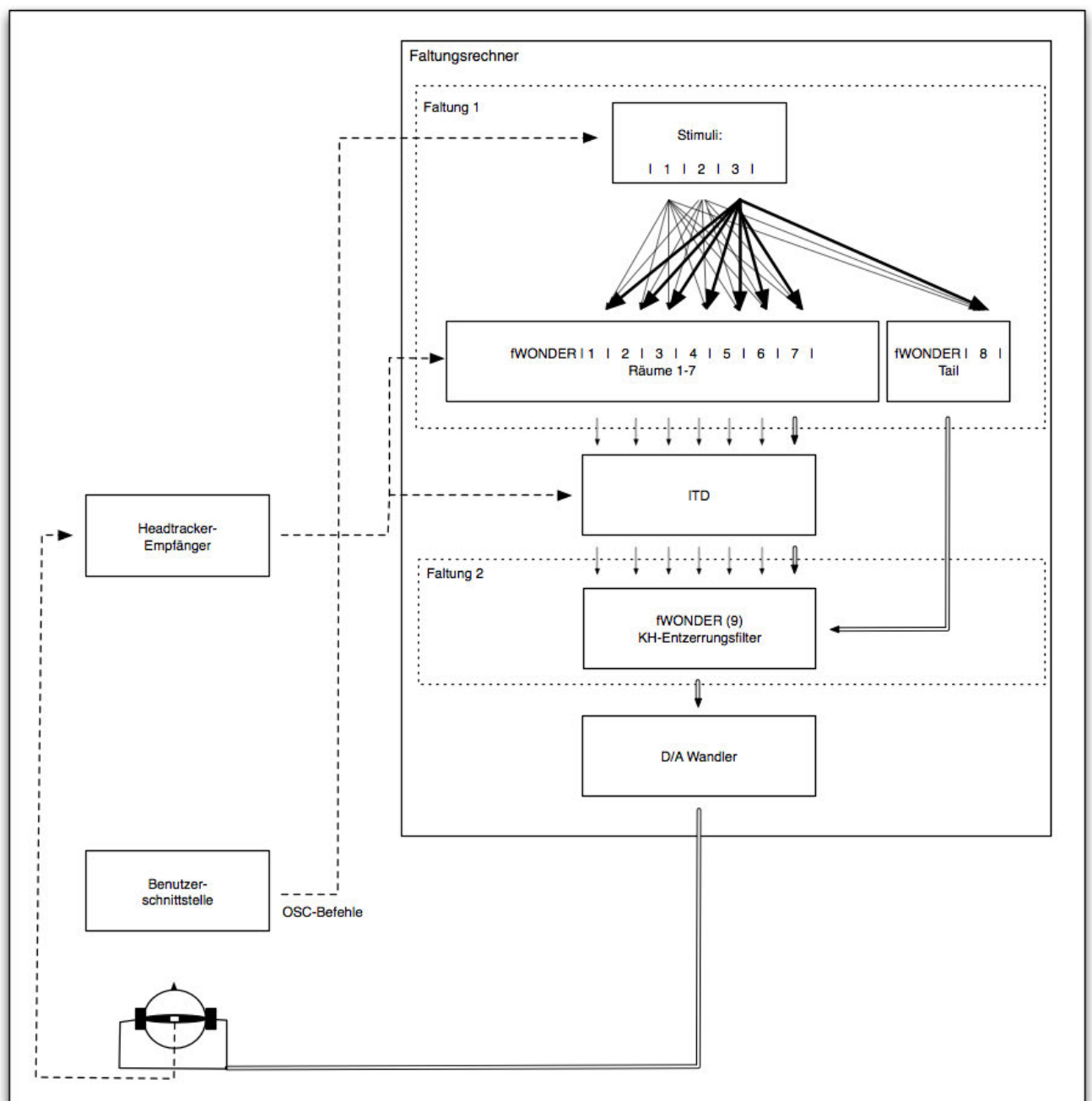


Abbildung 4.8: Hörversuchs-Aufbau im kleinen Studio am Fachgebiet Audiokommunikation. Die gestrichelten Linien zeigen OSC- und Headtracker-Befehle. Die fetten Linien deuten den Fall an, dass Stimulus 3 mit Raum 7 und dem Tail auralisiert wird, während die anderen Räume parallel gefaltet werden, durch das puredata-Routing aber stumm geschaltet werden.

von den VPs jederzeit gestoppt werden und durch erneutes Drücken der Buttons gestartet werden. Die Wiedergabe des Klangbeispiels wurde bei einem Wechsel des Raumes neu gestartet. Wie sich später bei der Befragung der VPs herausgestellt hat, wäre es aus Rücksicht auf das zeitlich begrenzte perzeptive Erinnerungsvermögen und die Marginalität der Unterschiede besser gewesen, beim Umschalten die Wiedergabe des Klangbeispiels fortzusetzen.

Zu Beginn des Hörversuches wurden den VPs die Instruktionen zu dem Versuch vorgelegt, die sie über Ablauf und Aufgabe aufklärte. Die Instruktionen sind im Anhang B.1 zu finden. Darauf folgte eine Trainingsphase, um die VP mit dem Spektrum der auftretenden klanglichen Unterschiede und dem Bewertungsprozess vertraut zu machen. In einer ersten Trainingsphase wurde der Versuch mit der Folge von Impulsen durchgeführt. In einer zweiten Phase hatte die VP die Wahl, mit einem der anderen beiden Stimuli oder beiden nacheinander das Training fortzusetzen.

Nach der Trainingsphase wurde der eigentliche Hörtest durchgeführt, wobei alle Simulationen in Bezug auf Ähnlichkeit gegenüber der Referenz mit allen drei Stimuli in jeweils zufälliger Reihenfolge bewertet wurden. Nach dem Hörversuch wurden die VPs gebeten, einen Fragebogen auszufüllen. Hier wurde anonym nach Alter, musikalischer Ausbildung und Erfahrung mit Hörtests gefragt. Ausserdem wurden die VPs dazu aufgefordert, auditive Merkmale für die Unterschiedlichkeitsbewertung zu nennen. Sie sollten die Merkmale in der Reihenfolge ihrer Relevanz, bezogen auf den Stimulus benennen. Der Fragebogen ist im Anhang B.2 beigefügt.

## 4.5 Ergebnisse

An dem Hörversuch nahmen insgesamt 29 VPs teil.

Die Auswertung der Fragebögen ergab, dass:

- 83% der VPs waren männlich.
- Das Durchschnittsalter lag bei 31 Jahren.
- Die VPs hatten durchschnittlich eine musikalische Ausbildung von 12 Jahren.
- 76% der VPs hatten vorherige Erfahrungen mit Hörversuchen.

### 4.5.1 Unterscheidungsmerkmale bezogen auf die Stimuli

Die auditiven Merkmale für die Unterschiedlichkeitsbewertung konnten durch eine quantitative Analyse der Angaben im Fragebogen in folgender Weise zusammengefasst werden.

	<b>Impulse</b>	<b>Sprache</b>	<b>Musik</b>
Unterschiedlichkeitsbewertung	leicht - nur für diesen Stimulus	schwer - kaum möglich	sehr schwer - nicht möglich
Merkmale:			
Räumlichkeit	14 (45.16%)	16 (59.26%)	10 (33.33%)
Klangfarbe	12 (38.71%)	4 (14.81%)	8 (26.67%)
spektrale Eigenschaften	-	5 (18.52%)	4 (13.33%)
Lautheit	-	-	8 (26.67%)
ohne Zuordnung	5 (16.13%)	2 (7.41%)	5 (16.67%)

Tabelle 4.2: Ergebnisse der quantitativen Analyse von genannten auditiven Merkmalen für die Unterschiedlichkeitsbewertung bezüglich der Stimuli

## 4.5.2 Statistische Auswertung

Als Grundlage für die statistische Auswertung dienen die erfassten Daten, wie sie an der Bewertungsskala von allen VPs vorgenommen wurden. Daraus werden zunächst von allen VPs durchschnittliche Differenzgrade für die einzelnen Faktor- (Reduktions-)stufen mit 95%-igen Konfidenzintervallen als Maß für die Streuung berechnet. Die Ergebnisse sind in Abb. 4.9 dargestellt. Über der Abszisse sind die Bewertungen nach abnehmender Reflexionsanzahl der Schallfelder geordnet. Die Bewertungen der drei Audioinhalte sind jeweils farbkodiert. Die ersten drei Balken zeigen die Bewertung des Referenzschallfeldes, die vorletzten drei die des mit dem kalibrierten auditorischen Modell reduzierten Schallfelds. Negative Differenzgrade bedeuten, dass die manipulierte Variante erkannt und als unterschiedlich zur Referenz bewertet wurde.

Die weitere Auswertung der Daten wurde in MATLAB und mit SPSS® von dem Betreuer der Arbeit (Alexander Lindau) durchgeführt. An dieser Stelle werden nur einige Resultate kurz genannt. Für die genaue Berechnung und Bedeutung der Analysemethoden und der darin verwendeten Begriffe wird auf die Bücher [46] und [47] verwiesen.

Zunächst wurde die Nullhypothese ( $H_0$ ) aus einer mehrfaktoriellen Varianzanalyse (ANOVA) "global" geprüft, dass sich die Mittelwerte der Differenzgrade nicht voneinander unterscheiden. Dabei wird die  $H_0$  über alle Faktoren für die Haupteffekte und Interaktion getestet. Daraus ergab sich, dass sowohl Haupteffekte, als auch Interaktion signifikant sind. Aufgrund der vielen Faktorstufen ist die Interpretation der Interaktion schwierig. Signifikante Haupteffekte deuten darauf hin, dass die wahrgenommene Qualität der verschiedenen Faktorstufen je nach Audioinhalt unterschiedlich bewertet wurde.

Ein 'posthoc'- Paarvergleich hat gezeigt, dass der Stimulus Impulstrain unterschiedlich von Sprache und Musik bewertet wurde. Daraus ergibt sich, dass die Bewertung besser je Stimulus ausgewertet werden sollte.

Die getrennte ANOVA für den Stimulus Impulstrain zeigt einen multivariaten Effekt. Der Effekt beschreibt, dass für die Bewertung der Faktorstufen eine Gruppenbildung festgestellt

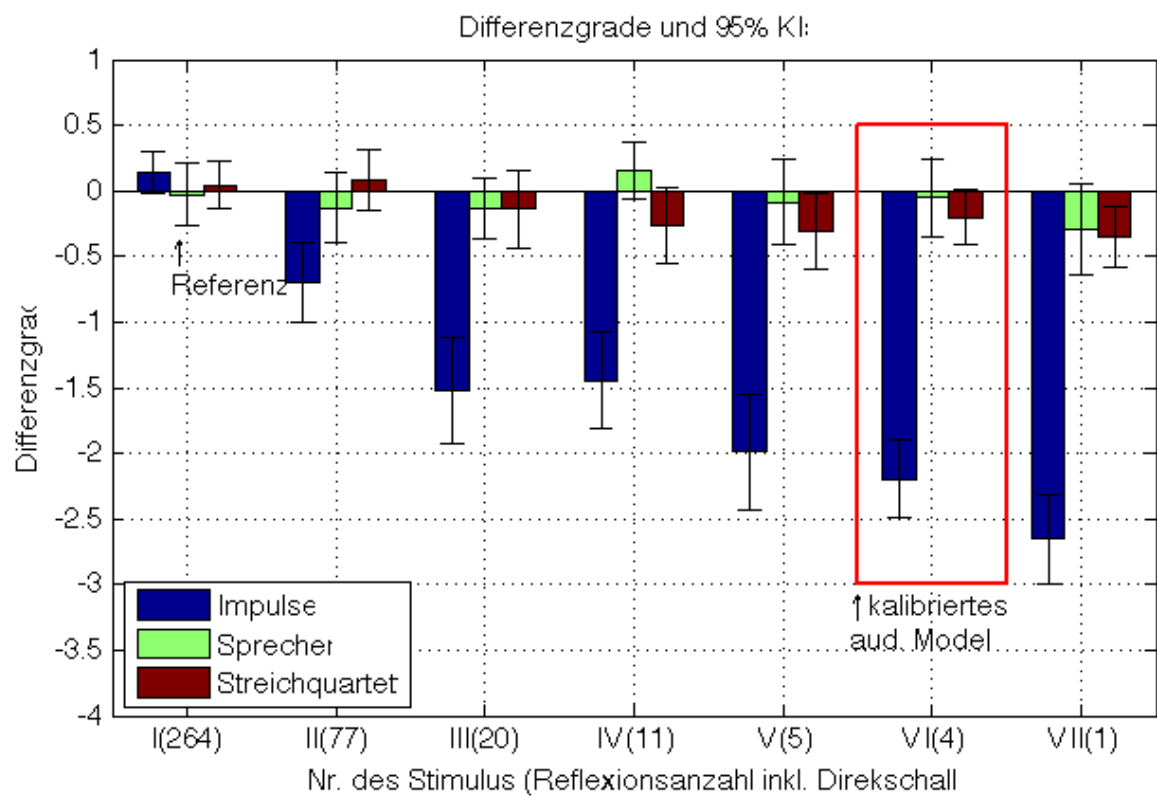


Abbildung 4.9: Ergebnisse des Hörversuchs als durchschnittliche Differenzgrade mit 95%-igen Konfidenzintervallen

werden konnte. Diesen Effekt kann man auf die ungleiche Verteilung der Reduktionsstufen zurückführen. Das Referenzschallfeld mit 263 Reflexionen und die nächste Reduktionsstufe mit 76 Reflexionen bilden jeweils eine Gruppe. Die Unterscheidbarkeit der Gruppe III (19 Reflexionen) - IV (10 Reflexionen) und V (4 Refl.) - VI (3 Refl.) können jeweils zusammengefasst werden. Die letzte Gruppe besteht wiederum isoliert aus dem Ankerreiz.

Aus der ANOVA für jeweils Sprache und Musik ergeben sich keine multivariaten Effekte. D.h. alle Faktorstufen wurden ähnlich bewertet. Für die Sprache konnten auch keine univariaten Effekte oder signifikanten Erkenntnisse aus dem ‘posthoc’ - Paarvergleich ermittelt werden. Nur für den Stimulus Musik war der Paarvergleich einer Faktorstufe (für den Ankerreiz) signifikant.

Individuelle t-Tests wurden jeweils für alle Faktorstufen auf Unterscheidbarkeit zur Referenz durchgeführt. Für das Impulstrain-Signal waren sie unter allen Schallfeldbedingungen signifikant. D.h. jede Reduktionsstufe konnte von der Referenz unterschieden werden. Für den Stimulus Musik waren sie nur unter der Direktschallbedingung (Ankerreiz) signifikant. Diese detaillierten Ergebnisse lassen eine Interpretation der Ergebnisse zu, wie sie sich für die durchschnittlichen Differenzgrade in Abb.4.9 zeigen, da die wichtigsten Effekte auch hierin abgebildet werden. Die Interpretation der Ergebnisse findet nun im folgenden Kapitel statt.

## 4.6 Diskussion der Hörversuchsergebnisse

Zunächst fällt auf, dass die Erkennbarkeit des Referenzschallfelds mit natürlichen Stimuli (Sprache / Musik) kaum gegeben war. Die Konfidenzintervalle umfassen in nahezu allen Fällen die 0-Linie. Einzig bei Musik konnte eine statistisch signifikante Unterscheidbarkeit der einfachsten Schallfeldvariante (Ankerreiz bestehend aus nur Direktschall plus modelliertem Nachhall) vom Referenzschallfeld gezeigt werden. Damit bestätigt sich die Annahme, dass zumindest für den natürlichen Stimulus Musik der Ankerreiz zur schlechtesten Beurteilung führte. Wie vom Modell vorhergesagt, konnten höhere Schallfeldauflösungen nicht vom Referenzschallfeld unterschieden werden. Beim Sprachstimulus konnte noch nicht mal der einfachste Schallfeldtyp, also der Ankerreiz, sicher erkannt werden. Diese Ergebnisse sind vergleichbar mit Ergebnissen der Untersuchung zur perzeptiven Mixingtime in [28]. Dort wurde indirekt gezeigt, dass Schallfelddetails nach der 2.-3. Wandreflexion kaum erkennbar sind.

Ein völlig anderes Bild ergab sich für den Pulsetrain-Stimulus. Sowohl Unterscheidbarkeit als auch Rangfolgenbildung war nahezu eindeutig und mit hoher Reliabilität möglich. Keines der Schallfelder wurde mit dem Referenzschallfeld verwechselt. Eine mit abnehmender Reflexionsanzahl abnehmende Ähnlichkeit zur Referenz wurde unerwartet gut reproduziert. Das auditorische Modell wurde auf die Vorhersage der Vedeckungseigenschaften von Impulspaaren ausgelegt. In dem Modell werden durch die Hörbarkeitsregel im lokalen Entscheidungsapparat simultane und Nachverdeckungseffekte berücksichtigt. Es wurde vereinfachend angenommen, dass eine iterative Erweiterung des Prozesses auf die Vorher-



sage für das Verdeckungsverhalten komplexer Reflexionsmuster möglich sei. Da bei dem Impulstrain-Stimulus die Reduktionsstufen in allen Fällen eindeutig erkannt wurden und mit abnehmender Anzahl von individuellen Reflexionen eine steigende Unähnlichkeit zur Referenz festgestellt wurde, erscheint das Verdeckungsverhalten von Reflexionsmustern hier deutlich komplexer als das von einfachen Impulspaaren.

Wie in Kapitel 2.2 bereits gezeigt, beschränken sich die Untersuchungen hauptsächlich auf die Beschreibung des Verdeckungsverhaltens von einzelnen Reflexionen. Auch bei dem Versuch, die Frage nach dem Einfluss früher Reflexionen auf den gesamten Raumklang zu beantworten, scheint der Einfluss einzelner Reflexionen die einzige Möglichkeit, solche Effekte zu ermitteln. Siehe hierzu auch [48]. D.h. empirische Daten dazu scheinen nicht zu existieren, bzw. konnten vom Autor nicht gefunden werden.

Die Auswahl früher Reflexionen durch das Modell scheint zwar plausibel, die Ergebnisse aus dem Hörversuch lassen allerdings den Schluss zu, dass z.B. für Sprachsignale die Übertragung von Direktsignal und synthetischem Nachhall ausreicht, um einen identischen Höreindruck zu hochaufgelösten Schallfeldbeschreibungen zu generieren. Zumindest für quasistatische natürliche Stimuli, wie Sprache oder Musik zeigen auditorisches Modell und auch die Hörversuchs Ergebnisse eine geringe Detektierbarkeit früher Reflexionen. In der Literatur geht man allerdings davon aus, dass frühe Reflexionen einen starken Einfluss insbesondere auf

- Lokalisation und Charakterisierung von Quellen
- den räumlichen Eindruck
- charakteristische Eigenschaften von Räumen, wie Wandoberflächen und geometrische Anordnung der Raumelemente

haben. Punkt 1 und 3 beziehen sich dabei implizit nur auf sehr frühe Reflexionen, d.h. solche 1. Ordnung.

In [12] wird der Einfluss von frühen Reflexionen folgendermaßen zusammengefasst:

“However, the reflections within an interval of so-called “early reflections” contribute in a specific way to the direct sound impression. They enhance the loudness, support the intelligibility of speech, the clarity of music and the impression of the auditory source width.”[12], S.93.

In [48] wird der Begriff “spatial impression” benutzt, um den subjektiven Effekt in Zusammenhang mit frühen seitlichen Reflexionen zu beschreiben. Zumindest für den Stimulus Impulse scheint also fraglich, den Ergebnissen aus dem HV allgemeine Gültigkeit zuzusprechen.

In [49] wurden synthetische Raumimpulsantworten berechnet, mit denen ein natürlicher Klangeindruck eines Raumes durch Faltung mit einem Musiksignal erreicht werden konnte. In die RIR-Berechnung durch IIR-Filter für den stochastischen Nachhallbereich wurden zwei frühe Reflexionen aus einem geometrischen Raummodell als FIR-Filter integriert. Sie wurden nach dem ‘trial and error’-Prinzip so ausgewählt, dass gute klangliche Ergebnisse

erzielt werden konnten. Der natürliche Klangeindruck konnte jedoch keinem spezifischen Raum zugeordnet werden.

Versucht man diese Erkenntnis auf die Ergebnisse des HVs zu verallgemeinern, kann angenommen werden, dass mit der Auswahl früher Reflexionen durch das Modell zumindest für Musik und Sprache ein natürlicher Raumklang hergestellt werden konnte.

Das Modell ermöglicht somit die Erzeugung einer reduzierten BRIR für eine realistisch klingende Raumsimulation. Die Akustik ist aber nicht unbedingt mit dem Klang eines spezifischen Schallfelds, wie hier von dem Audimax, vergleichbar. Dieser Eindruck betrifft jedoch eher das Problem der Unähnlichkeit von gemessenen Daten (Simulation der Realität) zu modellgenerierten Daten (Simulation der modellierten Realität), wie schon in Kapitel 4.3 angesprochen.

Eine scheinbar zufällige Wahl von frühen Reflexionen könnte daher eventuell mit der perzeptiv motivierten Auswahl durch das Modell ersetzt werden. Ob sich die Auswahl auch für jeden Stimulustyp auf den perzeptiv relevanten Inhalt bezieht, muss jedoch in Frage gestellt werden. Die Ergebnisse für den kritischen pulsartigen Stimulus zeigen, dass dies nicht der Fall sein kann, da sich der qualitative Raumeindruck für solche Signale eindeutig ändert.

Geht es aber darum, für anwendungstypische Stimuli natürlich klingende Raumsimulationen zu erzeugen, scheint eine Anwendbarkeit des erweiterten Algorithmus' zur Schallfeldreduktion denkbar.

# Kapitel 5

## Zusammenfassung und Diskussion

Für die vorliegende Arbeit wurden raumakustische Messungen im Audimax der TU-Berlin durchgeführt. Dabei wurde zunächst die Nachhallzeit mit dem entzerrten Lautsprecher und einem omnidirektionalen Messmikrofon gemessen.

Weiterhin wurden für fünf Quellpositionen binaurale Raumimpulsantworten mit dem Kopf- und Torsosimulator ‘FABIAN’ in einer feinen Auflösung auf horizontaler Ebene gemessen. Die BRIRs wurden für die Auralisation der Realität durch die dynamische Binauralsynthese verwendet.

In einem weiteren Teil der vorliegenden Arbeit wurde die Software EASE benutzt, um diskrete Raumschallfeldbeschreibungen von dem dreidimensionalen, raumakustischen Modell des Audimax für einzelne Quell- und Empfängerkombinationen zu berechnen. Dazu wurde die Sekundärstruktur des Modells, unter Berücksichtigung einflussnehmender Oberflächenstrukturen und nach bestem Wissen vereinfacht. Die akustischen Eigenschaften der Oberflächenmaterialien wurden derart verändert, dass die berechnete Nachhallzeit bestmöglich der in der Realität gemessenen Nachhallzeit entspricht. Das sogenannte Modell-Fitting an die Realität ist im Hinblick auf geringe Abweichungen von gemessener und aus dem Modell berechneter Nachhallzeit gut gelungen. Die entstandenen Fehler sind hauptsächlich auf die Einschränkungen durch EASE im unteren Frequenzbereich zurückzuführen.

Die Berechnung einer Raumimpulsantwort in EASE lieferte Reflektogramme, in der die zeitliche Struktur und Zusammensetzung des Schallfeldes in parametrisierter Form beschrieben werden. Relevante Informationen aus den Reflektogrammen wurden bis zur doppelten perzeptiven Mixingtime extrahiert. Aus den Amplitudenspektren diskreter Reflexionen wurden Impulsantworten als FIR-Filter berechnet. In einem Metadatensatz wurden relative Pegel, zeitliche und räumliche Informationen für jede Reflexion gespeichert. Aus der Faltung aller Reflexionen als FIR-Filter mit HRIR-Daten wurde ein BRIR-Datensatz für horizontale Kopfdrehungen generiert, der für die Auralisation mit dynamischer Binauralsynthese verwendet wurde. Da die BRIR-Daten nur den frühen Teil einer Raumimpulsantwort enthalten, wurde ihnen ein synthetisch erzeugter Nachhallbereich in geeigneter Form angehängt. Aus pragmatischen Gründen wurde aus einem heuristischen Ansatz die

Nachhallsynthese implementiert.

In einem Vorversuch wurden die modellgenerierten Datensätze mit der dynamischen Binauralsynthese auralisiert. Dabei hat sich herausgestellt, dass sich ein kontinuierlicher Klangeindruck zwischen den durch das auditorische Modell reduzierten Schallfeldern mit synthetischem Nachhall erst bei erheblich reduzierten Schallfeldarstellungen ergibt und dass die korrekte Herstellung des Direkt-/Diffusschallverhältnisses wesentlich ist. Die Erzeugung eines geeigneten Nachhallbereichs und die Verbindung dessen mit dem frühen Teil der BRIR muss eingehender untersucht werden. Als Ausgangspunkt für weiterführende Ansätze der Nachhallsynthese sei u.a. auf die Arbeit von [50] verwiesen. Hier wird eine Übersicht über die hauptsächlichen Disziplinen, die sich mit der Nachhallerzeugung beschäftigen, gegeben.

Die als FIR-Filter beschriebenen Reflexionen wurden verwendet, um mit einem auditorischen Modell Entscheidungen über Hörbarkeit einzelner Reflexionen in dem gesamten Schallfeld zu treffen. Dazu wurde zunächst ein einfaches auditorisches Modell anhand vorliegender Forschungsliteratur implementiert. Als Vorbild für die Entwicklung des Modells diene vor allem der Ansatz von [4]. Wie in dieser Arbeit dokumentiert, modelliert der Algorithmus Teile des auditorischen Systems. Auf der Entscheidungsebene zur Hörbarkeit von Schallfeldkomponenten im auditorischen Modell werden nur simultane und Nachverdeckungseffekte berücksichtigt. Die Hörbarkeitsentscheidungen werden aus dem Pegelverhältnis von auditorisch aufgearbeiteten Signalpaaren getroffen. Ein adaptives Verhalten in Abhängigkeit vom Eingangssignal wurde nicht implementiert. Ansätze dafür sind z.B. in den Arbeiten [23, 24] zu finden. Die Wirkungsweise des Modells wurde durch entsprechende Einstellparameter an empirischen Daten zur Beschreibung von Verdeckung einzelner künstlicher Reflexionen angepasst. Zumindest für Verdeckungskurven als Funktion von der Verzögerungszeit können empirische Daten durch das Modell gut abgebildet werden. Daraus haben sich Modellparameter ergeben, die zu einem kalibrierten auditorischen Modell führten. Unter der vereinfachenden Annahme, dass die Abbildung der empirischen Daten verallgemeinerbar auf die Vorhersage für das Verdeckungsverhalten komplexer Reflexionsmuster seien, wurde das Modell um einen iterativen Prozess erweitert. Durch die Erweiterung kommt man zu Entscheidungen über die Hörbarkeit einzelner Raumreflexionen in der modellgenerierten Schallfeldbeschreibung. Die Wirkungsweise des Modells wurde anhand grafischer Darstellungen objektiv überprüft. Insbesondere die Darstellungen in Richtungsigelplots führte dazu, dass die Auswahl von perzeptiv relevanten frühen Reflexionen als plausibel beurteilt wurde.

In einem Hörversuch wurde überprüft, ob die objektive plausible Auswahl früher Reflexionen einen subjektiven Einfluss auf die Unterscheidbarkeit zu hochaufgelösten Schallfeldern nehmen. Die Ergebnisse zeigen, dass eine Unterscheidbarkeit von Reduktionsstufen stark stimulusabhängig ist. Reduktionsstufen können mit dem Impulstrain-Stimulus eindeutig von der Referenz unterschieden werden. Man kann also davon ausgehen, dass die verallgemeinernde Annahme nicht unbedingt für jeden Stimulustyp gültig ist. Somit scheint der auditorische Algorithmus die Schallfeldbeschreibung nicht auf den perzeptiv relevanten Inhalt zu reduzieren.

In Bezug auf einen natürlichen Klangeindruck von Raumsimulationen bedeutet dies aber nicht, dass die reduzierten Schallfelder mit synthetisiertem Nachhall keine guten Datensätze für die Simulationen darstellen. Es zeigt vielmehr, dass für natürlich klingende Raumsimulationen mit kontinuierlichen Signalen die Anordnung früher Reflexionsmuster keine große Rolle spielen. Es muss jedoch in Frage gestellt werden, ob die Akustik spezifischer Räume eindeutig simuliert werden können. Dies wurde in dieser Arbeit nicht weiter untersucht. Die Ergebnisse aus dem Vorversuch dieser Arbeit deuten daraufhin, dass selbst vollaufgelöste modellgenerierte Datensätze qualitativ nicht mit gemessenen Datensätzen vergleichbar sind. Es wundert also nicht, dass die reduzierten Datensätze spezifische Raumsimulation nicht ermöglichen. Auch wenn es fragwürdig ist, ob das Modell die Schallfeldbeschreibung auf den perzeptiv relevanten Inhalt reduziert, kann ein natürlicher Höreindruck mit anwendungstypischen Stimuli, wie Musik oder Sprache für die Auralisation erzeugt werden. Die Ergebnisse des Hörversuchs zeigen, dass die reduzierten Schallfelder mit den entsprechenden Stimuli nicht von der Referenz unterschieden werden konnten.

Das Ziel dieser Arbeit wurde also insofern erreicht, dass durch ein kalibriertes auditorisches Modell reduzierte und evaluierte Schallfeldbeschreibungen in geeigneter Form zur Verfügung gestellt werden konnten, um die Berechnung von Treiberfunktionen für lautsprecherbasierte Wiedergabesysteme zur Reproduktion von Raumschallfeldern in erlebbarer Zeit zu ermöglichen.

# Anhang A

## Inhalt DVD

Auf der beigelegten DVD befinden sich folgende Anhänge:

- PDF-Version der vorliegenden Arbeit
- Verwendetes EASE-Modell in dem Ordner - Quellcodes - 2 EASE soundfield creation - 2 EASEmodel - scattering 060percent correctedVolume.zip
- Audiofiles, pd-Patch, Konfigurationsfiles und MATLAB-Skripte, die für den Hörversuch verwendet wurden in dem Ordner Quellcodes - HV Roehrbein
- Alle MATLAB-Skripte aus dem soundfield-Repository in dem Ordner Quellcodes - soundfield. (Alternativ kann auf das Repository unter <https://svr2.ak.tu-berlin.de/soundfield> zugegriffen werden.)
- zusätzliche MATLAB-Skripte in den Ordnern Quellcodes - Visualize: Für die Erzeugung von ETC-, Richtungsigelplots und Nachhallsynthese: Für die Erzeugung des in dem Hörversuch verwendeten diffusen Nachhalls
- Alle Volltexte bzw. den in der Arbeit referenzierten Abschnitt der im Literaturverzeichnis aufgeführten Publikationen im PDF-Format im Ordner Literatur - Pdfs
- Formatierte BibTex Datei: BibliographyI.bib (für das in der Arbeit abgedruckte Literaturverzeichnis) in dem Ordner Literatur, zusätzlich das Literaturverzeichnis im Citavi-Format (Formatierung nicht geprüft) in dem Ordner Literatur - Citavi - MSc MaxRöhrbein Literaturverzeichnis.ctv3

## Anhang B

### Anhang - Versuchsanleitung und Fragebogen

#### B.1 Instruktionen

Liebe/r Versuchsteilnehmer/In!

In diesem Hörversuch geht es um die Bestimmung von Wahrnehmungsunterschieden zwischen verschiedenen binauralen Schallfeldsimulationen. Ziel ist es, eine Referenzsimulation zu erkennen und das Ausmaß der klanglichen Abweichungen der veränderten Simulationen zu bewerten.

#### Versuchsablauf:

Der Versuch wird mithilfe der unten dargestellten grafischen Benutzeroberfläche durchgeführt. Die verschiedenen Simulationen werden dort jeweils paarweise in zufälliger Kombination mit der Referenzsimulation dargeboten und können über die beiden Play-Tasten abgespielt werden. Zusätzlich kann die Referenzsimulation immer über eine eigene Taste ('Ref') abgespielt werden.

Deine Aufgabe ist es,

1. wenn möglich, die manipulierte, d.h. die sich von der Referenz unterscheidende Simulation zunächst zu erkennen und dann
2. anhand des der Simulation zugeordneten Schiebereglers - im Bereich von "identisch" bis "sehr unterschiedlich" - zu bewerten, wie stark sich diese Deiner Meinung nach vom Schallfeld der Referenzsimulation unterscheiden.

Bei dieser Bewertung ist zu beachten, dass die gegebenen Bewertungen auch eventuelle Ähnlichkeitsunterschiede **zwischen** den einzelnen manipulierten Simulationen widerspiegeln, so dass sich, wenn möglich, eine nach abnehmender Ähnlichkeit gestufte Reihenfolge der Audiobeispiele ergibt.





Zur Erhöhung der Verlässlichkeit können alle Klänge beliebig oft angehört werden. Da die Simulation zudem interaktiv auf Kopfbewegungen reagiert, kannst Du, z. B. um die Unterschiede besser zu erkennen, Deinen Kopf in der Horizontalebene drehen ( $\pm 80^\circ$ ). Der Versuch wird für drei verschiedene Audiobeispiele durchgeführt. Diese werden jedoch innerhalb eines „Panels“ der Benutzeroberfläche nicht variiert.

**Training:**

Der Versuch beginnt mit einer kurzen Trainingsphase, um Dich mit dem Spektrum der auftretenden klanglichen Unterschiede und dem Bewertungsprozess vertraut zu machen. In einer ersten Trainingsphase wird der Versuch mit einer Folge von Impulsen durchgeführt. In einer zweiten Trainingsphase hast Du die Wahl mit einem der anderen Stimuli (Sprache, Musik) fortzusetzen.

## B.2 Fragebogen

### Fragebogen zu statistischen Daten

VP Nr. \_\_\_\_\_

Alter \_\_\_\_\_

Geschlecht \_\_\_\_\_

Musikalische Ausbildung oder Erfahrung :

---



---

Dauer/seit : \_\_\_\_ Jahr(en)

Ich habe bereits an einem Hörversuch teilgenommen: ja / nein

Falls die Audiobeispiele unterschiedlich bewertet wurden : Welche auditiven Merkmale waren für die unterschiedliche Bewertung verantwortlich? Bitte nenne die Merkmale in der Reihenfolge ihrer Relevanz bezogen auf den Stimulus?

---



---



---



---



---



---



---



---

# Abbildungsverzeichnis

2.1	Schematische Darstellung eines Reflektogramms aus hybriden und stochastischen Verfahren aus [13], S.109, Fig.4.8 . . . . .	9
2.2	Richtungsigelplot einer Schallfeldbeschreibung mit räumlichen Informationen. Die Richtung des Direktschalls (Einfall von vorne) ist gekennzeichnet.	10
2.3	Messsystem für die Untersuchung zur subjektiven Wirkung von Schallfeldern aus [16], S.540, Fig.1 . . . . .	13
2.4	Signalabhängige Verdeckungsschwelle zwischen Rauschen und Impulsen aus [16], Fig.2 . . . . .	14
2.5	Signalabhängige Verdeckungsschwelle aus [4], Fig.1 . . . . .	15
2.6	Verdeckungsschwellen für 3 VPs als Funktion von Verzögerungszeit bei Schalleinfall aus gleicher Richtung von vorne mit dem Stimulus ‘rosa’ Rauschen aus [16], Fig.5 . . . . .	16
2.7	Verdeckungsschwelle als Funktion von Einfallswinkel bei fester Verzögerung von 50 ms mit dem Stimulus ‘Zack’ aus [14], Fig.22 . . . . .	17
2.8	(a) Originale RIR, (b)rekonstruierte RIR aus [1], S.8, Fig.14 . . . . .	21
2.9	Globale Struktur des Modells von [4], S.3, Fig.2 . . . . .	23
2.10	Verdeckungsschwellen in Abhängigkeit von Einfallswinkel empirischer Daten aus [14], Fig.22 und dem Modell von [4], S.6, Fig.8 . . . . .	24
2.11	links: simulierte RIR eines Schuhschachtelraums, rechts: RMT Voraussage aus simultanem und nicht-simultanem DD für die simulierte RIR aus [4], S.6, Fig.10 und 11 . . . . .	25
3.1	Signalfluss und Verkabelung für die Vermessung des Dodekaeder-Lautsprechers im Hallraum . . . . .	29
3.2	Rohmessung aller Wege im Diffusfeld (links,), Rohmessung Tieftön im Nahfeld (rechts) . . . . .	31
3.3	Finale Frequenzweichen-Filter mit nicht benutzter, aber angepasster Nahfeld-TT-Flanke . . . . .	33
3.4	links: Finale Frequenzweichen-Filter mit benutzter und angepasster Nahfeld-TT-Flanke, 1/3-Oktav geglättet; rechts: Vergrößerte Darstellung der linken Abbildung . . . . .	33
3.5	Bühnensituation für BRIR-Messung . . . . .	34

3.6	Gemessene Nachhallzeit vom Audimax aus gemittelten RIR Messungen der Centerquelle . . . . .	35
3.7	Das originale Raummodell des Audimax der TU-Berlin als Drahtgitterdarstellung aus EASE . . . . .	37
3.8	An die reale Messsituation angepasstes EASE-Raummodell . . . . .	38
3.9	Optimize RT-Funktion zum Anpassen der Nachhallzeit des Modells an die Realität . . . . .	39
3.10	Terzbandaufgelöste Nachhallzeit des gemessenen Raumes (oben, blau), des gefitteten Modells (oben, grün) und der Fehler in % (unten). . . . .	41
3.11	Amplitudenspektren mit Nullstellen im 3.15 kHz Terzband einer Reflexion für eine Quelle (Src1) . . . . .	45
3.12	Einfaches Prozessmodell zur Modellierung des Verdeckungsverhaltens von Impulspaaren . . . . .	48
3.13	Rekonstruktionsverhalten der Gammaton-Filterbank aus [35] . . . . .	55
3.14	Rekonstruktionsverhalten der Gammaton-Filterbank aus [36] . . . . .	56
3.15	Rekonstruktionsverhalten der Gammaton-Filterbank [41] . . . . .	57
3.16	Pol-/Nullstellendiagramm des vierten Filterbands (von 30), Mittenfrequenz bei 240 Hz, berechnet mit der Gammaton Filterbank aus [41] . . . . .	58
3.17	Rekonstruktionsverhalten der Filterbank von C. Couvreur . . . . .	59
3.18	Verdeckungsschwellen als Funktion von Verzögerungszeit aus Modellberechnung (rot) und empirischen Untersuchungen (schwarz) aus [16], S.541, Fig.2 für Impulspaare. Links: Verdecker von vorne, Testreflexion von 65° rechts. Rechts: Verdecker von vorne, Testreflexion von 60° oben . . . . .	61
3.19	Verdeckungsschwellen als Funktion von Einfallswinkel aus Modellberechnung (rot) und empirischen Untersuchungen (schwarz) aus [14], S.107, Bild 22 für Impulspaare. Oben: Die Richtung des Verdeckers wird variiert, Unten: Die Richtung der Testreflexion wird variiert . . . . .	63
3.20	Erweitertes Prozessmodell zur Modellierung des Verdeckungsverhaltens in Raumschallfeldern . . . . .	64
3.21	links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vol- laufgelöstes Schallfeld, schwarz: nur Direktschall . . . . .	66
3.22	links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vol- laufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit 'at' = 0.6 . . . . .	66
3.23	links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vol- laufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit 'at' = 0.39 . . . . .	67
3.24	links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vol- laufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit 'at' = 0.345 . . . . .	67
3.25	links: ETC-Plot für die ersten 150 ms, rechts: Richtungsigel-Plot, rot: Vol- laufgelöstes Schallfeld, schwarz: Durch das Modell als hörbar detektierte Reflexionen mit 'at' = 0.228 . . . . .	68

4.1	Schematische Darstellung der dynamischen Binauralsynthese mit getrennter ITD-Verarbeitung aus [31], wie Fig.4, S.7 . . . . .	71
4.2	oben: modellgenerierte BRIR, links: ETC, rechts: Amplitudenspektrum, unten: geschätzte Diffusfeld-HRTF aus Zeitbereich 100ms bis 1s der BRIR, links: ETC, rechts: Amplitudenspektrum . . . . .	72
4.3	oben: geschätzte Diffusfeld-HRTF von BRIR-Ausschnitt (100ms bis 1s) mit geglättetem tiefen Frequenzbereich, mitte: FIR-Filter von Diffusfeld-HRTF, unten: Impulsantwort des FIR-Filters von 100 ms bis 1s . . . . .	73
4.4	terzbandaufgelöste Nachhallzeiten . . . . .	74
4.5	ETC-Plots von modelliertem Nachhall und unreduziertem modellgenerierten BRIR-Datensatz, oben: linker Kanal (Ohr), unten: rechter Kanal (Ohr) . .	76
4.6	ETC-Plots von modelliertem Nachhall (rot) und unreduziertem modellgenerierten BRIR-Datensatz (schwarz) mit linearem Fade-In, oben: linker Kanal (Ohr), unten: rechter Kanal (Ohr) . . . . .	77
4.7	Hörversuchspanel . . . . .	81
4.8	Hörversuchs-Aufbau im kleinen Studio am Fachgebiet Audiokommunikation. Die gestrichelten Linien zeigen OSC- und Headtracker-Befehle. Die fetten Linien deuten den Fall an, dass Stimulus 3 mit Raum 7 und dem Tail auralisiert wird, während die anderen Räume parallel gefaltet werden, durch das puredata-Routing aber stumm geschaltet werden. . . . .	83
4.9	Ergebnisse des Hörversuchs als durchschnittliche Differenzgrade mit 95%-igen Konfidenzintervallen . . . . .	86

# Tabellenverzeichnis

2.1	Frequenzbänder und zeitliche Integrationslängen aus [1], S.8, Table 2 . . .	20
3.1	Berechnete Nachhallzeiten in Oktavbändern aus RIR-Messung im Hallraum	30
3.2	Filtereinstellungen am Controller . . . . .	32
3.3	Originale und angepasst zusammengefasste Materialien und Flächenanteile in [%] des Raummodells vom Audimax der TU-Berlin . . . . .	40
4.1	Verwendete Reduktionsstufen im Hörversuch berechnet durch variablen ‘at’- Schwellwert im auditorischen Modell und daraus resultierende Anzahl von Reflexionen $N_r$ . . . . .	79
4.2	Ergebnisse der quantitativen Analyse von genannten auditiven Merkmalen für die Unterschiedlichkeitsbewertung bezüglich der Stimuli . . . . .	85

# Literaturverzeichnis

- [1] Hulsebos, Edo; de Vries, Diemer (2002): “Parameterization and reproduction of concert hall acoustics measured with a circular microphone array”. In *112th AES Convention, Munich* (Preprint 5579).
- [2] Begault, Durand R.; McClain, Bryan U.; Anderson, Mark R. (2004): “Early reflection Thresholds for anechoic and Reverberant Stimuli within a 3D Sound Display”. In *18th International Congress on Acoustics* (Proceedings):1267–1270.
- [3] Hacıhabiboğlu, Hüseyin; Murtagh, Fionn (2008): “Perceptual simplification for model-based binaural room auralization”. In *Applied Acoustics* **69**:715–727.
- [4] Buchholz, Jörg; Mourjopoulos, John; Blauert, Jens (2001): “Room Masking: Understanding and Modelling the Masking of Room Reflections”. In *110th AES Convention, Amsterdam* (Preprint 5312).
- [5] Berkhout, A.J.; de Vries, Diemer; Vogel, P. (1993): “Acoustic control by wave field synthesis”. In *J. Acoust. Soc. Am.* **93**(5):2764–2778.
- [6] Berkhout, A.J.; de Vries, Diemer; Sonke, J.J. (1997): “Array technology for acoustic wave field analysis in enclosures”. In *J. Acoust. Soc. Am.* **102**(5):2757–2770.
- [7] Gerzon, Michael A. (1985): “Ambisonics in Multichannel Broadcasting and Video”. In *J. Audio Eng. Soc.* **33**(11):859–871.
- [8] Hulsebos, Edo (2004): *Auralization using Wave Field Synthesis*. Dissertation, Technische Universiteit Delft.
- [9] Duraiswami, Ramamani; et al. (2005): “High Order Spatial Audio Capture and its Binaural Head-Trackable Playback over Headphones with HRTF Cues”. In *Proceedings of the 119th AES Convention, New York* (6540).
- [10] Zotter, Franz (2009): *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*. Dissertation, University of Music and Performing Arts, Graz, Austria.
- [11] Ahnert, Wolfgang; Tennhardt, Hans Peter (2008): “Raumakustik”. In *Stefan Weinzierl (Hg.): Handbuch der Audiotechnik*, Kapitel 5. Heidelberg: Springer, 181–266.
- [12] Vorländer, Michael (2008): *Auralization*. 1. Auflage. Heidelberg, Berlin: Springer.

- [13] Kuttruff, Heinrich (2009): *Room Acoustics*. 5. Auflage. London and New York: Spon Press.
- [14] Burgdorf, W. (1961): “Untersuchungen zur Wahrnehmbarkeit verzögerter Schallsignale”. In *Acoustica* **11**:97–111.
- [15] Seraphim, H. P. (1961): “Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall”. In *Acoustica* **11**:80–91.
- [16] Olive, Sean E.; Toole, Floyd E. (1989): “The Detection of Reflections in Typical Rooms”. In *J. Audio Eng. Soc.* **37**(7/8).
- [17] Brunner, Stefan; Maempel, Hans-Joachim; Weinzierl, Stefan (2007): “On the Audibility of Comb Filter Distortions”. In *122nd AES Convention, Vienna* (Preprint 7047).
- [18] Plack, Christopher J.; Oxenham, Andrew J. (1998): “Basilar-membrane nonlinearity and the growth of forward masking”. In *J. Acoust. Soc. Am.* **103**(3):1598–1608.
- [19] Blauert, Jens (1997): *Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition)*. Revised Auflage. Cambridge u.a.: The MIT Press.
- [20] Moore, Brian C.J. (2004): *An Introduction to the Psychology of Hearing*. 5. Auflage. London u.a.: Elsevier Academic Press.
- [21] Glasberg, Brian R.; Moore, Brian C.J. (1990): “Derivation of auditory filter shapes from notched-noise data”. In *Hearing Research* **47**(1-2):103–138.
- [22] Dau, Torsten; Püschel, Dirk; Kohlrausch, Armin (1996): “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure”. In *J. Acoust. Soc. Am.* **99**(6):3615–3622.
- [23] Buchholz, Jörg M.; Mourjopoulos, John (2004): “A Computational Auditory Masking Model Based on Signal-Dependent Compression. I. Model Description and Performance Analysis”. In *Acta Acustica united with Acustica* **90**:873–886.
- [24] Buchholz, Jörg M.; Mourjopoulos, John (2004): “A Computational Auditory Masking Model Based on Signal-Dependent Compression. II. Model Simulations and Analytical Approximations”. In *Acta Acustica united with Acustica* **90**:887–900.
- [25] Behler, Gottfried K. (2007): “Dodekaeder-Lautsprecher mit variabler Richtcharakteristik”. In *Fortschritte der Akustik: Tagungsband d. DAGA* **33**:887–888.
- [26] Müller, Swen; Massarani, Paulo (2001): “Transfer-Function Measurement with Sweeps”. In *J. Audio Eng. Soc.* **49**(6):443–471.
- [27] Lindau, Alexander (2006): *Ein Instrument zur softwaregestützten Messung binauraler Raumimpulsantworten in mehreren Freiheitsgraden*. Magisterarbeit, Technische Universität Berlin, Institut für Sprache und Kommunikation, FB Kommunikationswissenschaft.



- [28] Lindau, Alexander; Kosanke, Linda; Weinzierl, Stefan (2010): "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses". In *Presented at the 128th AES Convention, London* .
- [29] Lindau, Alexander; Weinzierl, Stefan (2009): "ON THE SPATIAL RESOLUTION OF VIRTUAL ACOUSTIC ENVIRONMENTS FOR HEAD MOVEMENTS IN HORIZONTAL, VERTICAL AND LATERAL DIRECTION". In *Proc. of the EAA Symposium of Auralization* .
- [30] Schärer, Zora (2008): *Kompensation von Frequenzgängen im Kontext der Binauraltechnik*. Magisterarbeit, Technische Universität Berlin, Institut für Sprache und Kommunikation, Fachgebiet Audiokommunikation.
- [31] Lindau, Alexander; Estrella, Jorgos; Weinzierl, Stefan (2010): "Individualization of dynamic binaural synthesis by real time manipulation of the ITD". In *Presented at the 128th AES Convention, London* .
- [32] Moldrzyk, Christoph; Ahnert, Wolfgang; Feistel, Stefan; et al. (2004): "Head-Trackted Auralization of Acoustical Simulation". In *Presented at the 116th AES Convention, San Francisco* .
- [33] Blickensdorff, Johannes (2010): *Raumakustik im Wiener Burgtheater: Ein Vergleich von Messwerten und modellierten Parametern*. Magisterarbeit, Technische Universität Berlin, Institut für Sprache und Kommunikation, Fachgebiet Audiokommunikation.
- [34] Cardenas, Antonio (2010): *Simulation of a sound field reproduction approach for an innovative musical venue (II): Pre-processing of Head-Related Impulse Responses*. Praktikumsbericht, Technische Universität Berlin, Fachgebiet Audiokommunikation.
- [35] Slaney, Malcolm (1993): "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank". In *Apple Computer Technical Report Nr.35, Perception Group - Advanced Technology Group, Apple Computer Inc.* .
- [36] Hohmann, Volker (2002): "Frequency analysis and synthesis using a Gammatone filterbank". In *Acta Acustica united with Acustica* **88**:433–442.
- [37] Robinson, David J. M.; Hawksford, Malcolm J. (1999): "Time-domain auditory model for the assessment of high-quality coded audio". In *Presented at 107th AES Convention, New York* :5017–5040.
- [38] DIN EN ISO 389-7: 2006-04 (2006): *Akustik – Standard-Bezugspegel für die Kalibrierung von audiometrischen Geräten Teil 7: Bezugshörschwellen unter Freifeld- und Diffusfeldbedingungen*. Deutsches Institut für Normierung, Berlin.
- [39] Härmä, Aki; Palomäki, Kalle (2000): "HUTear – a Free Matlab Toolbox for Modeling of Human Auditory System". In *Proceedings 1999 Matlab DSP Conference* (<http://www.acoustics.hut.fi/software/HUTear/> Zugriff 25.07.2010).

- [40] Slaney, Malcolm (1998): “Auditory Toolbox. Version 2”. In *Technical Report Nr.1998-010, Interval Research Corporation* .
- [41] Clark, Nick (2007): “GammaTone Tool Kit v2.0”. In <http://www.mathworks.com/matlabcentral/fileexchange/15313> **Zugriff: 12.09.2011.**
- [42] Nicol, Rozenn (2010): *Binaural Technology*. 1. Auflage. New York: Audio Engineering Society Inc.
- [43] Borß, Christian (2009): “A NOVEL APPROACH FOR OPTIMALLY MATCHING A LATE REVERBERATION MODEL TO AN IMAGE SOURCE MODEL - OR: WHAT DOES A FOOTBALL HAVE TO DO WITH SHOEBOX SHAPED ROOMS?” In *Proc. of the EAA Symposium of Auralization* .
- [44] Rec. ITU-R BS.1116-1 (1994-1997): *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. International Broadcasting Union, Genf.
- [45] Erdfelder, Edgar; Faul, Franz; Buchner, Axel (1996): “GPOWER: A general power analysis program”. In *Behaviour Research Methods, Instruments & Computers* **28**(1):1–11.
- [46] Bortz, Jürgen (2005): *Statistik für Human- und Sozialwissenschaftler*. 6. Auflage. Heidelberg: Springer Medizin Verlag.
- [47] Bortz, Jürgen; Döring, Nicola (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage. Heidelberg: Springer Medizin Verlag.
- [48] Barron M.; Marshall A.H. (1981): “Spatial Impression due to early lateral reflections in concert halls: The derivation of a physical measure”. In *Journal of Sound and Vibration* **77**(2):211–232.
- [49] Moorer, James A. (1979): “About This Reverberation Business”. In *Computer Music Journal* **3**(2):13–28.
- [50] Blesser, Barry (2001): “An Interdisciplinary Synthesis of Reverberation Viewpoints”. In *J. Audio Eng. Soc.* **49**(10):887–903.