

Perzeptiv motivierte, parametrische Synthese binauraler Raumimpulsantworten

vorgelegt von
M.Eng.
Philipp Stade
geb. in Haan

von der Fakultät I – Geistes- und Bildungswissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
– Dr.rer.nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende: Prof. Dr. Susann Fegter
Gutachter: Prof. Dr. Stefan Weinzierl
Gutachter: Prof. Dr. Christoph Pörschmann

Tag der wissenschaftlichen Aussprache: 14. Juni 2018

Berlin 2018

Perzeptiv motivierte, parametrische Synthese binauraler Raumimpulsantworten

Philipp Stade

Dissertation, 2014 - 2018



Technology
Arts Sciences
TH Köln

Technische Universität Berlin – Fachgebiet Audiokommunikation

Technische Hochschule Köln – Institut für Nachrichtentechnik

Diese Dissertation ist an der Technischen Universität Berlin – Fakultät I im Fachgebiet Audiokommunikation unter der Betreuung von Prof. Dr. Stefan Weinzierl entstanden.

Die Forschungsaktivitäten an der Technischen Hochschule Köln, Fakultät 7 im Institut für Nachrichtentechnik (INT) wurden vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 03FH005I3 „MoNRa“ gefördert und von Prof. Dr. Christoph Pörschmann geleitet.

Die Thesis ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz (CC-BY 4.0).



Textsatz mit L^AT_EX und Texmaker 4.5.

Referenzen verwaltet mit Papers 3.4 und BibTeX.

Vektor Grafiken erstellt unter Adobe Illustrator CS5.

Pixel Grafiken und Fotos editiert mit Adobe Lightroom und Adobe Photoshop CS5.

Diagramme erstellt mit MATLAB R2015b.

Fotos von Philipp Stade.

für
Caro
Malte
&
Jette

Danksagungen

Ich danke herzlich Stefan Weinzierl für die Betreuung und die Unterstützung dieser Arbeit. Die vertrauensvolle Zusammenarbeit sowie die zahlreichen ergebnisreichen Diskussionen und die Integration in seine Arbeitsgruppe ermöglichten so erst die Forschungsaktivitäten in diesem Umfang.

Vielen lieben Dank an Christoph Pörschmann, ohne dessen Rolle als Visionär, Initiator und Mentor diese Untersuchungen gar nicht hätten durchgeführt werden können. Vielen Dank für das Vertrauen, für die uneingeschränkten Hilfestellungen zu jeder Zeit, für stets neue Ideen, aber auch kritische Fragestellungen sowie für ausgezeichnete Arbeitsbedingungen. Weiterhin danke ich aufrichtig für sein immerwährendes Verständnis und die so mögliche Vereinbarkeit von Familie und Beruf.

Ein besonderer Dank geht an Johannes M. Arend und Benjamin Bernschütz, die beide durch ihren außerordentlichen Einsatz einen maßgeblichen Beitrag zu dieser Arbeit geleistet haben – Danke für alles!

Vielen, vielen Dank an alle aktuellen oder auch ehemaligen Kollegen der TH Köln (in alphabetischer Reihenfolge) für die fantastische Atmosphäre und die gute Mischung aus professioneller Arbeit aber auch freundschaftlichem Miteinander: Johannes M. Arend, Benjamin Bernschütz, Damian Dziwis, Aaron Finkenthe, Antje Goldenberg, Gary Grutzek, Dominic Hemmer, Peter Lieblang, Tim Lübeck, Narea Jantzen, Max Rühl und Arnau Vázquez-Giner.

Weiterhin danke ich allen Studenten, die dieses Projekt durch die Teilnahme an den Hörversuchen, durch Abschlussarbeiten oder aber durch unterstützende Arbeit (Jonas Christofzik, Patrick Pereira, Pia Krauss und Klaus Wunder) vorantreiben konnten.

Ich danke den beteiligten Projektpartnern TU Berlin, Universität Rostock, Westdeutscher Rundfunk Köln, IOSONO/Barco Audio Technologies und Graner+Partner Ingenieure des Forschungsprojektes MoNRa für die fortwährende Unterstützung und den intensiven Austausch.

Vielen Dank an alle Kollegen der virtuellen Akustik (in alphabetischer Reihenfolge) für die Inspiration und den Austausch, die Hilfestellungen und ebenso die zahlreichen fröhlichen Abende: David Ackermann, Jens Ahrens, Fabian Brinkmann, Matthias Geier, Michael Horn, Alexander Lindau, Annika Neidhardt und Frank Schulz.

Ein Hoch auf Annika, Frank und Johannes: Vielen lieben Dank für die Korrekturen und das Aufdecken von kleinen und größeren Fehlern.

Herzliche Grüße und Dank gehen ins Kölner Funkhaus zu den Kollegen des WDR, speziell Markus Haßler und Benedikt Bitzenhofer haben diese Arbeit maßgeblich und in vielen Bereichen unterstützt.

Vielen Dank an Markus Mehnert und Kollegen für die intensive und anspruchsvolle, aber auch sehr ergiebige Zeit in Erfurt. Viele Ideen für die Arbeit sind aus dieser Zusammenarbeit entstanden.

Ich danke Norbert für die Stunden des Zuhörens und die fortwährende Begleitung dieser Arbeit ohne Wiederworte.

Ohne familiäre Unterstützung hätte ich diese Arbeit nie beenden können. Daher danke ich meiner gesamten Familie für den bedingungslosen Beistand in allen Lebenslagen.

Diese Arbeit widme ich Caro, Malte und Jette, den wichtigsten Personen meines Lebens – Danke, dass es euch gibt!

Philipp Stade

Juli 2018

Zusammenfassung

Diese Dissertation beschäftigt sich mit der Entwicklung und Untersuchung von Methoden zur Synthese binauraler Signale für eine dynamische Hörbarmachung von akustischen Gegebenheiten über Kopfhörer. Es werden verschiedene Modelle gebildet, die diffuse und gerichtete Schallfeldkomponenten separat analysieren und mittels Parametern charakterisieren. Anschließend werden diese raumabhängigen Merkmale zur Generierung synthetischer binauraler Raumimpulsantworten (SBRIRs) verwendet.

Im Mittelpunkt der Untersuchungen steht die auditive Wahrnehmung des Menschen. Es wird angestrebt, Schallfeldkomponenten die für die Perzeption irrelevant sind, nicht zu betrachten und so eine wirksame Reduktion der Informationen zu ermöglichen, mit möglichst geringen hörbaren Einflüssen. Die generierten SBRIRs können in dynamischer Binauralsynthese durch Faltung mit nachhallfreien Audiosignalen zur Hörbarmachung des Raumes eingesetzt werden und erlauben bei Änderung der Kopfausrichtung des Rezipienten, eine ortsfeste Wiedergabe der virtuellen Schallquellen. Durch die Gegenüberstellung der SBRIRs mit einer binauralen Referenz auf Basis von Kunstkopfmessungen in einer ABC/HR Testumgebung werden die Modelle perzeptiv evaluiert. Ergebnisse dieser Untersuchungen fließen stets in nachfolgende Modelle ein, sodass ein wachsender Prozess abhängig der auditiven Wahrnehmung entsteht.

Zur Charakterisierung der akustischen Situation werden verschiedene Ansätze kombiniert: Gerichtete Schallfeldkomponenten können mit Methoden der Wellenfeldanalyse auf Basis von Messungen mit sphärischen Mikrofonarrays räumlich und zeitlich detektiert werden. Außenohrübertragungsfunktionen werden anschließend entsprechend der detektierten richtungsabhängigen Komponenten räumlich ausgewählt, verzögert und energetisch sowie spektral angepasst, sodass synthetische Reflexionen erzeugt werden. Die diffusen Komponenten können durch frequenzabhängige Energieabklingkurven und die interaurale Kohärenz repräsentiert werden. Diese Parameter werden verwendet, um weißes Rauschen gemäß der binauralen Referenz zu adaptieren und so synthetisch ein binaurales diffuses Signal zu generieren.

Die Arbeit zeigt, dass eine starke Reduktion des Detailgrades der messtechnisch erfassten Situation erfolgen kann. Die Vereinfachung des Schallfeldes im Raum auf einige wenige Parameter erlaubt weiterhin die adäquate Hörbarmachung der akustischen Situation durch die erzeugten SBRIRs. Dabei kann eine geringere Relevanz der diskreten Reflexionen im Gegensatz zum diffusen Nachhall beobachtet werden. Dieser sollte mit hoher Frequenzauflösung und entsprechend des Verlaufs der binauralen Kohärenz diffuser Schallfelder adaptiert werden. Die Modelle können in Systemen der virtuellen oder erweiterten Realität Anwendung finden und schaffen Möglichkeiten zur flexiblen Adaption des Raumeindrucks bei effektiver Reduktion der Informationen.

Abstract

This thesis discusses the development and the investigation of different approaches for the synthesis of binaural signals to achieve a dynamic auralisation of arbitrary acoustical conditions via headphones. Various models are defined, which analyze diffuse and directional components of the sound field separately and characterize them with individual parameters. After that, these room-dependent features are used to generate synthetic binaural room impulse responses (SBRIRs).

The presented investigations strongly focus on human auditory perception. The parametric models aim for an efficient reduction of the information of the sound field with preferably minor audible influences. Components which are irrelevant for the perception should not be regarded in the synthesis. In order to auralize rooms, the SBRIRs can be applied in dynamic binaural synthesis for the convolution with anechoic audio signals. This method allows for localizable and static virtual sound sources considering the orientation of the listener. An ABC/HR test paradigm is used for the perceptual evaluation of the parametric models. The auralisation based on the SBRIRs is directly compared with the auralisation based on a measured binaural reference (recorded with an artificial head). Results are incorporated in subsequent models with the result of an increasing process dependent on the auditory perception.

Different approaches are combined to characterize the acoustical environment: Directional components of the sound field can be spatial-temporal detected with spherical microphone arrays using wave field analysis. Head-related transfer functions are selected dependent on the detected direction of arrival of the directional components. By delaying, attenuating and spectral filtering of these impulse responses, synthetic reflections which are spatial distributed can be generated. Diffuse components are represented with frequency-dependent energy decay curves and the interaural coherence. These parameters are used to adapt white noise signals according to the binaural reference and to create synthetic decaying signals with the binaural cues of the diffuse sound field of the reference room.

This thesis reveals that the details of the measured information can be considerably reduced. Furthermore, the simplification of the sound field with sparse parameters allows for an adequate auralisation of the acoustical situation using the generated SBRIRs. In contrast to the diffuse components, a minor impact of the directional components can be observed. For the adaptation of the diffuse components, though, a detailed frequency resolution is required and the binaural coherence of the diffuse sound field has to be taken into account. The parametric models can be used in applications for virtual or augmented reality and open up new possibilities for the flexible adaptation of the spatial impression in combination with an effective reduction of the information.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Stand der Forschung	5
1.2	Motivation und Ziele	8
1.3	Aufbau der Thesis	10
1.4	Nomenklatur	11
2	Binauraltechnik für die virtuelle Akustik	15
2.1	Binaurales Hören	17
2.1.1	Wahrnehmung von Reflexionen	20
2.1.2	Wahrnehmung des diffusen Schallfeldes	22
2.1.2.1	Beginn des diffusen Schallfeldes	24
2.1.2.2	Interaurale Kohärenz	25
2.2	Binaurale Messtechnik	27
2.3	Raumakustische Kriterien	31
2.4	Dynamische Binauralsynthese	33
2.5	Wellenfeldanalyse zur Schallfelderlegung	35
2.5.1	Fourier-Transformationen	36
2.5.2	Zerlegung in ebene Wellen	38
3	Methoden, Technologien und Ressourcen	43
3.1	Messsysteme	44
3.1.1	Sequenzielles Mikrofonarray	45
3.1.2	Sequenzielles Binauralsystem	46
3.1.3	Schallquellen	47
3.2	Räumlichkeiten	48
3.2.1	Klaus-von-Bismarck-Saal	48
3.2.2	Kleiner Sendesaal	49
3.3	HRIR-Datensatz	49
3.4	Software	50
3.4.1	SOFiA-Toolbox	51
3.4.2	ITA-Toolbox	51

3.4.3	SoundScape Renderer	53
3.4.4	Scale	54
3.5	Technische Rahmenbedingungen der Versuchsumgebung	54
3.6	Versuchsanordnungen	56
3.6.1	ABC/HR	57
3.6.2	MUSHRA	61
3.7	Statistische Verfahren	62
3.7.1	t-Test	63
3.7.2	Varianzanalyse	65
3.7.3	Post-hoc Tests	65
3.8	Zusammenfassung Methoden, Technologien und Ressourcen	66
4	Signalverarbeitung für die BRIR-Synthese	67
4.1	Gerichtete Komponenten des Schallfeldes	69
4.1.1	Analyse der gerichteten Komponenten	70
4.1.1.1	Schallfeldzerlegung zur Detektion von Maxima	71
4.1.1.2	Generierung gerichteter Impulsantworten	76
4.1.1.3	FIR-Filter für die gerichteten Komponenten	78
4.1.2	Synthese der gerichteten Komponenten	78
4.2	Diffuse Komponenten des Schallfeldes	80
4.2.1	Analyse der diffusen Komponenten	81
4.2.1.1	Energieabklingkurven	81
4.2.1.2	Interaurale Kohärenz	86
4.2.2	Synthese der diffusen Komponenten	87
4.2.2.1	Adaption der Energieabklingkurven	89
4.2.2.2	Adaption der interauralen Kohärenz	89
4.3	Zusammenfassung Signalverarbeitung für die BRIR-Synthese	91
5	Parametrische Modelle für die BRIR-Synthese	93
5.1	Idee und Systemüberblick	96
5.2	Übergangszeit zwischen frühen und späten Anteilen	97
5.3	Späte Anteile (LEA1/2)	102
5.3.1	Modell und Parameter	102
5.3.2	Datensätze	107
5.3.3	Technische Untersuchung	109
5.3.4	Perzeptive Evaluation	116
5.3.5	Diskussion und Zusammenfassung	121
5.4	Frühe Anteile (LEB)	123
5.4.1	Modell und Parameter	124

5.4.2	Datensätze	127
5.4.3	Technische Untersuchung	129
5.4.4	Perzeptive Evaluation	133
5.4.5	Diskussion und Zusammenfassung	137
5.5	Gesamtmodell (LEC)	139
5.5.1	Modell und Parameter	140
5.5.2	Datensätze	142
5.5.3	Technische Untersuchung	142
5.5.4	Perzeptive Evaluation	146
5.5.5	Diskussion und Zusammenfassung	150
5.6	Gesamtmodell mit reduzierter Auflösung (LED)	152
5.6.1	Modell und Parameter	153
5.6.2	Datensätze	156
5.6.3	Technische Untersuchung	156
5.6.4	Perzeptive Evaluation	160
5.6.5	Diskussion und Zusammenfassung	165
6 Ergebnisse und Ausblick		167
A Abbildungen		173
B Abkürzungen, Bezeichnungen und Variablen		229
Literaturverzeichnis		235

KAPITEL **1**

Einleitung



Die zunehmende Digitalisierung nimmt einen großen Einfluss auf viele Bereiche der Gesellschaft und deren Entwicklung. Insbesondere die soziale Interaktion der Menschen mit ihrer Umwelt sowie die Arbeitswelt erfahren einen Wandel durch den Einsatz neuartiger Technologien und Verfahren. Dies eröffnet mitunter, durch den Gebrauch digitaler Medien und Systeme, neue Möglichkeiten für vernetzte Arbeitsprozesse, mit erhöhter Produktivität und Flexibilität. In der Hightech-Strategie der Bundesrepublik Deutschland (BMBF, 2017) nimmt die Digitalisierung der Gesellschaft einen zentralen Stellenwert ein und soll im Rahmen der vierten industriellen Revolution (*Industrie 4.0*¹) beispielsweise Interaktionen zwischen Mensch und Maschine stärken.

Weiterhin ist ein Anstieg der Relevanz von Systemen der sogenannten *virtuellen Realität* (VR) zu verzeichnen. Virtuelle Umgebungen werden in Forschung und Entwicklung bereits seit zahlreichen Jahren angewendet und untersucht. Derweil findet aber zusätzlich eine Kommerzialisierung dieser Thematik durch neue Produkte und Anwendungen statt, sodass VR den Massenmarkt erreicht hat. Im digitalen Zeitalter ist der Einsatz von Technologien der virtuellen Realität zur Darstellung verschiedenster Ebenen der Wahrnehmung ein wichtiger und innovativer Aspekt und wird beispielsweise für veränderte Prozesse der Arbeitswelt (*Arbeit 4.0*²) gesellschaftlich relevant.

Durch die kontinuierliche Erhöhung der verfügbaren Leistungsfähigkeit der Datenverarbeitung und verbesserter Sensorik, kann in diesen Systemen die Wirklichkeit immer präziser und mit steigender Modalität anhand einer virtuellen Umgebung imitiert werden. Neben den visuellen Reizen ist die akkurate Nachbildung der auditiven Komponente essentiell für das Eintauchen in eine virtuelle Umgebung. Dabei gilt es, den Raumeindruck der realen Umgebung gemäß der auditiven Wahrnehmung virtuell abzubilden und die entsprechenden Freiheitsgrade bei Bewegung des Nutzers³ durch Anpassung des Schallfeldes zu berücksichtigen. Zur Integration dieser Systeme in bestehende Funktionalitäten und einer Erhöhung der Praktikabilität können zudem eine Adaption der Komponenten und/oder Möglichkeiten zur Datenreduzierung erforderlich werden.

Um die Entwicklung dieser Technologien zu fördern und eine Verwertbarkeit entsprechender Anwendungen zu steigern, besteht in diesem Kontext weiterhin Forschungsbedarf. So setzt sich diese Arbeit sowohl mit der Untersuchung von Grundlagen der auditiven Perzeption von Räumen, als auch mit praktischen Anwendungen und Ver-

¹„Industrie 4.0 steht für die informations- und kommunikationstechnische Verzahnung der industriellen Produktion. Die Kennzeichen [...] sind die starke Individualisierung der Produkte unter den Bedingungen einer hoch flexibilisierten (Großserien-)Produktion, die weitgehende Integration von Kunden und Geschäftspartnern in Geschäfts- und Wertschöpfungsprozesse sowie die Verkopplung von Produktion und hochwertigen Dienstleistungen.“ (BMBF (2017), S.18)

²„Arbeiten 4.0 ist [...] ein Kürzel für die Veränderungen in der gesamten Arbeitswelt und ihre Folgen für die Gesellschaft [und] beschreibt dabei [...] Perspektiven, Szenarien und Gestaltungschancen für die Zukunft.“ (BMA (2017), S.5)

³Aus Gründen der besseren Lesbarkeit wird in der vorliegenden Arbeit auf die Verwendung geschlechtsspezifischer Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für alle Geschlechter.

fahren zur dynamischen Hörbarmachung in *virtuellen auditiven Umgebungen* (VAEs) auseinander.

In der Audiosignalverarbeitung reichen die Anfänge der sogenannten *Binauraltechnik*, eine Aufnahmetechnologie gemäß der menschlichen Wahrnehmung mit zwei Ohren, mittlerweile über 90 Jahre zurück. So findet sich in Fletcher und Sivian (1927) eine erste Definition und Patentierung eines Kunstkopf-Systems. Dabei handelt es sich um eine Nachbildung eines menschlichen Kopfes mit zwei Mikrofonkapseln, die entsprechend der Trommelfelle positioniert werden. Nachdem sich diese Technologie kommerziell nicht am Markt durchsetzen konnte, gewann sie zum Ende des 20. Jahrhunderts zunehmend an Bedeutung für Forschung und Wissenschaft. Insbesondere in Anwendungen der sogenannten virtuellen Akustik zur räumlichen Wiedergabe von Schallfeldern, erlebte die Binauraltechnik in Kombination mit fortschreitender Sensortechnik eine regelrechte Renaissance (z.B. Begault (1991)), deren Relevanz bis heute anhält. Weiterhin kann zunehmend die Akzeptanz und der Gebrauch von Kopfhörern zur Wiedergabe von Audiosignalen in der Gesellschaft beobachtet werden, der Absatz von Geräten dieser Art nimmt weltweit zu (GfK, 2017).

Durch die Verwendung von mit Kunstköpfen messtechnisch erfassten, binauralen Raumimpulsantworten (BRIRs) zur Faltung mit nachhallfreien Audiosignalen kann eine reale, akustische Gegebenheit per Kopfhörer auditiv wahrnehmbar gemacht werden (*Auralisation* (Kleiner et al., 1993)). In der dynamischen Binauralsynthese (siehe Kapitel 2.4) ist zudem durch die Erfassung der Orientierung des Rezipienten zur Laufzeit und Auswahl der jeweiligen Impulsantworten, eine ortsfeste Wiedergabe der virtuellen Schallquellen entsprechend der realen Situation⁴ möglich. Daher bilden BRIRs heutzutage oftmals die Grundlage für die Berechnung der auditiven Komponente in Systemen der virtuellen Realität. Eine eindeutige akustische Analyse des Schallfeldes ist auf Basis dieser Methoden jedoch nur begrenzt möglich. Richtungsabhängige Komponenten können zwar vom auditiven System ausgewertet, aber nicht technisch in Form einer räumlichen Beschreibung des Schallfeldes extrahiert werden.

Der Einsatz von *sphärischen Mikrofonarrays* und Methoden aus der Fourier-Akustik hingegen, ermöglicht eine richtungssensitive Zerlegung des Schallfeldes im Raum. Dazu muss das Schallfeld akkurat auf einer Kugeloberfläche abgetastet werden. Es können dann unter anderem Aussagen über die räumliche und zeitliche Verteilung von diskreten Reflexionen getroffen werden, sodass eine präzise Analyse der akustischen Gegebenheit möglich wird. Zwar lassen sich auch binaurale Signale aus diesen Datensätzen nach Bernschütz (2016) generieren, eine flexible Adaption der akustischen Situation in

⁴ gemäß der Definition für Realität aus Brockhaus Enzyklopädie Online (2018): „Wirklichkeit im Sinne der Summe alles Vorhandenen, tatsächlich Gegebenen, Gegenständlichen im Unterschied zum lediglich Gedachten oder Vorgestellten“ wird mit realer Situation das auftretende Schallfeld im Raum mit all seinen Eigenschaften verstanden.

Kombination mit Datenreduktion (die z.B. für den Einsatz in VR-Systemen nötig sein könnte), ist nicht ohne Weiteres möglich.

Unter der virtuellen Akustik werden nach Huopaniemi et al. (2000) Methoden zur Erzeugung und Bearbeitung von akustischen Phänomenen durch Computer verstanden: „Virtual acoustics is a general term for acoustical phenomena and systems with the aid of a computer.“ Dabei können zur Hörbarmachung räumliche Wiedergabesysteme auf Basis von Kopfhörern oder Lautsprechern eingesetzt werden, um so z.B. die auditive Komponente in VR-Systemen darzubieten. Durch die Verwendung von Algorithmen zur Berechnung der Schallfelder abhängig des intendierten Wiedergabesystems im sogenannten *Renderer*⁵, findet in der virtuellen Akustik ein Wandel der Definition und des Aufbaus von Audiosignalen statt. So basieren Audio-Produktionen zur konventionellen Wiedergabe auf diskreten Kanälen entsprechend der genutzten Lautsprecher, wohingegen *Rendering-Systeme* für räumliches Audio meist eine objektbasierte Beschreibung (Coleman et al., 2018) verwenden. Diese kann beispielsweise aus einem Audiosignal ohne Rauminformationen (nachhallfrei) in Kombination mit Metadaten zur Definition der akustischen Gegebenheit sowie der Position der virtuellen Schallquelle im Raum bestehen. In aktuellen Standardisierungen finden sich mittlerweile zwar objektbasierte Formate mit *parametrischer Kodierung*, für eine entsprechende Erfassung und Wiedergabe des Raumeindrucks besteht jedoch weiterhin Forschungsbedarf.

In der vorliegenden Arbeit werden diese drei wesentlichen Komponenten - *Binauraltechnik*, *Mikrofonarrays* und *parametrische Kodierung* - zur Hörbarmachung von realen Räumen kombiniert. Eine akustische Umgebung wird dazu unter Einsatz eines sphärischen Mikrofonarrays mit einer parametrischen Beschreibung möglichst präzise, im Bezug auf die menschliche Wahrnehmung, charakterisiert. Diese Beschreibung wird zur Synthese von binauralen Raumimpulsantworten verwendet, die zur Hörbarmachung mittels dynamischer Binauralsynthese eingesetzt werden. Dabei gilt es, die Unterschiede der Auralisation auf Basis der gemessenen BRIRs im Vergleich zu den synthetischen BRIRs technisch und perceptiv zu evaluieren. Eine Synthese auf Basis von Parametern ermöglicht weiterhin die flexible Adaption der Eigenschaften der akustischen Gegebenheiten, sodass eine Vielzahl von neuen Anwendungen möglich werden.

Die entwickelten Methoden zur Hörbarmachung von Räumen grenzen sich maßgeblich von Verfahren der sogenannten geometrischen Akustik ab. Anstatt die gewünschte akustische Gegebenheit auf Basis eines Raummodells zu simulieren, ist deren messtechnische Erfassung stets ein zentraler Bestandteil der hier vorgestellten Untersuchungen. Unter Einbeziehung der auditiven Perzeption werden in dieser Arbeit verschiedene pa-

⁵Ein ursprünglicher Begriff aus der Bildsynthese, bezeichnet dort ein System zur Erzeugung von Bildern mittels Computergrafik.

rametrische Modelle zur Synthese von BRIRs entsprechend der akustischen Merkmale auf Basis von Messdaten gebildet, evaluiert und weiterentwickelt.

1.1 Stand der Forschung

Das beidohrige Hören des Menschen und dessen auditive Wahrnehmung von Räumen sind Forschungsdisziplinen, deren Anfänge nach Wade und Deutsch (2008) bis in das 18. Jahrhundert zurückreichen. Relevante Grundlagenwerke in der Binauraltechnik finden sich von Blauert (1974, 1996) im Bezug auf das räumliche Hören sowie von Toole (2008) im Hinblick auf die menschliche Wahrnehmung von Reflexionen. Weiterhin beschäftigten sich beispielsweise Barron (1971), Bech (1996a,b, 1998), Buchholz et al. (2001) sowie Shinn-Cunningham (2003) oder Shinn-Cunningham und Ram (2003) intensiv mit der auditiven Wahrnehmung in geschlossenen Räumen und damit verbundenen Wahrnehmbarkeitsschwellen. So ist dieser Bereich zwar seit vielen Jahren im Fokus der Wissenschaft, weiterer Forschungsbedarf ist aber aufgrund der Komplexität der Wahrnehmungszusammenhänge noch immer vorhanden. Die Neurowissenschaft setzt sich ebenso mit kognitiven Zusammenhängen der Wahrnehmung von Räumen auseinander, sodass sich hier z.B. von Traer und McDermott (2016) relevante Publikationen finden lassen, die auch Methoden zur Synthese von Raumimpulsantworten einsetzen. Es handelt sich somit um ein interdisziplinäres Forschungsthema, welches eine Vielzahl von Wissensgebieten kombiniert.

Mittels Methoden der digitalen Signalverarbeitung wurden bereits in den 1990er Jahren zahlreiche Systeme zur Erzeugung virtueller akustischer Umgebungen zur Hörbarmachung von Räumen beispielsweise von Wenzel et al. (1990); Foster et al. (1991); Lehnert und Blauert (1992); Kleiner et al. (1993); Begault (1994); Reilly und McGrath (1995); Horbach et al. (1999) und Jot (1999) vorgestellt. Aufgrund dieser neuen Möglichkeiten konnte zudem eine Extensivierung der Untersuchungen zur Binauraltechnik und der Wahrnehmung von Schallfeldern in geschlossenen Räumen mittels der virtuellen Akustik stattfinden, die bis heute anhält, sodass diese Thematik weiterhin forschungsrelevant ist (Møller, 1992; Blauert, 1996; Begault et al., 2001; Zahorik, 2009; Blauert, 2013; Lindau, 2014; Maempel und Horn, 2017). Die datenbasierte, dynamische Binauralsynthese definiert den Stand der Technik dieser Systeme zur Hörbarmachung von akustischen Umgebungen mittels Kopfhörer. Sie ermöglicht nach Lindau und Weinzierl (2012) eine plausible oder mittels Individualisierung der Messdaten nach Fels et al. (2013) und Brinkmann et al. (2017b) gar eine authentische Wiedergabe.

Falls keine binaurale Messung zur Erfassung des Schallfeldes im Raum zum Einsatz kommt, so muss für eine Hörbarmachung die akustische Situation algorithmisch erzeugt

werden. Dies kann durch die Generierung entsprechender Impulsantworten erfolgen. In der Regel betrachten dabei Systeme zur Synthese der Impulsantworten die zeitlich frühen und späten Schallfeldanteile separat (Menzer, 2012). Es existieren Methoden auf Basis einer Raumsimulation oder einer messtechnischen Erfassung vor Ort. Insbesondere der diffuse Nachhall beschäftigt viele Arbeitsgruppen in der Akustik bereits seit zahlreichen Jahren, sodass sich erste Studien zur Generierung von synthetischem Nachhall bereits von Schroeder und Logan (1961) sowie Schroeder (1962) finden lassen. Die Thematik ist aber weiterhin aktuell und relevant für Forschung und Wissenschaft.

Der späte Nachhall wird nach Välimäki et al. (2012) in der Regel mit drei verschiedenen Methoden (oder derer Kombinationen) generiert. Zum einen verzögern und filtern sogenannte *feedback delay networks* (FDN) nach Jot und Chaigne (1991) das Eingangssignal in zahlreichen Rückkopplungs-Schleifen und erzeugen so einen diffusen Nachhall. Beim Faltung-Hall wird das Eingangssignal wiederum mit einer gemessenen oder generierten Impulsantwort der gewünschten akustischen Umgebung gefaltet. Zur Erzeugung können dazu nach Moorer (1979) und Jot et al. (1997) Rauschsignale eingesetzt werden, denen ein frequenzabhängiger Energieabfall aufgeprägt wird. Borß und Martin (2009) sowie Menzer und Faller (2010) kombinieren diesen Ansatz zudem mit einer Anpassung der interauralen Kohärenz entsprechend des idealen Diffusfeldes. Weiterhin existieren geometrische Methoden zur Erzeugung des diffusen Nachhalls auf Basis einer Raumsimulation. Eine umfangreiche Übersicht über die Methoden, Entwicklungen und die Geschichte der artifiziellen Generierung des späten Nachhalls geben Välimäki et al. (2012, 2016).

Der Übergangzeitpunkt in den diffusen Nachhall, die sogenannte *mixing time*, ist maßgeblich für einen adäquaten Einsatz dieser Methoden zur Hörbarmachung. Die *mixing time* wurde in der Vergangenheit beispielsweise von Cremer (1948); Rubak und Johansen (1998); Abel und Huang (2006); Hidaka et al. (2007) als auch Stewart und Sandler (2007a) intensiv untersucht, eine Zusammenfassung sowie perzeptive Evaluation der relevanten Modelle und Methoden zur Bestimmung dieser Übergangszeit findet sich in Lindau et al. (2012).

Die Synthese der frühen Schallfeldanteile dieser Systeme zur Hörbarmachung von Räumen basiert in der Regel auf diskreten Reflexionen, die geometrisch mittels Raytracing (Krokstad et al., 1968) oder Spiegelschallquellen-Verfahren (Allen und Berkley, 1979) mit Hilfe einer Raumsimulation gebildet werden. Weiterhin existieren zahlreiche Ansätze zur Detektion von Reflexionen im Raum auf Basis von Messungen. Daraus resultierende Informationen können ebenso zur Synthese von gerichteten Schallfeldanteilen zwecks Hörbarmachung von Räumen verwendet werden. Es finden sich unterschiedliche Vorgehensweisen, die auf der Auswertung der Kreuzkorrelation (Tervo et al., 2010, 2013), der Schallintensität (Merimaa et al., 2001) oder auf der Erfassung des Schall-

feldes mit Hilfe eines B-Format-Mikrofons (Ambisonic nach Gerzon (1985)) basieren (z.B. SIRR nach Merimaa und Pulkki (2005) sowie Pulkki und Merimaa (2006)). Einen ähnlichen Ansatz auf Basis von B-Format-Signalen präsentieren Menzer et al. (2010).

Auch Mikrofonarrays können zur Detektion von Reflexionen im Raum eingesetzt werden. Der SDM Ansatz nach Tervo et al. (2013) zerlegt das Schallfeld in eine Reihe von Spiegelschallquellen und arbeitet auf beliebigen Array-Anordnungen mit geringer Mikrofon-Anzahl. Hörbarmachungen auf Basis von SIRR und SDM finden sich beispielsweise von Pätynen und Lokki (2011), Laird et al. (2014) als auch Kuusinen und Lokki (2015). Eine Übersicht über Methoden der Detektion von Reflexionen mittels einer Auswertung der Laufzeitunterschiede (TDOA) zeigen Tervo et al. (2012). Remaggi et al. (2015) kombinieren zur Detektion von frühen Reflexionen mehrere Verfahren und bestimmen über den Phasenverlauf nach Naylor et al. (2006) den Zeitpunkt und mittels Beamforming nach Van Veen und Buckley (1988) die Richtung der Reflexionen.

Ansätze nach Rafaely (2004); Duraiswami et al. (2005a) oder Bernschütz (2016), die auf Methoden der Fourier-Akustik gemäß Williams (1999) zur Schallfelderlegung basieren, bedürfen hingegen einer akkuraten Anordnung der Mikrofone mit hoher räumlicher Abtastrate durch eine Vielzahl von Messpositionen. Diese Methoden definieren den Stand der Technik und ermöglichen, bei entsprechender Erfassung, eine hohe räumliche Auflösung und erreichbare Richtwirkung und damit die Zerlegung des Schallfeldes für den gesamten, hörbaren Bereich. Im Allgemeinen werden diese Ansätze aber bisher meist für die raumakustische Analyse und die Visualisierung (siehe z.B. O'Donovan et al. (2008) und Bernschütz et al. (2012)) von Schallfeldern eingesetzt. Verschiedene Möglichkeiten zur Erzeugung von binauralen Signalen auf Basis von Mikrofonarray-Messungen wurden u.a. von Duraiswami et al. (2005b), Bernschütz et al. (2014), Sheaffer et al. (2014, 2015), Ben-Hur et al. (2017) oder Ahrens et al. (2017) vorgestellt und evaluiert.

Weiterhin existieren nach Lokki und Karjalainen (2000) sowie Vesa und Lokki (2006) auditorisch motivierte Verfahren zur Detektion von Reflexionen anhand binauraler Raumimpulsantworten. Neben diesen Methoden, die allesamt auf mehrkanaligen Mikrofonanordnungen basieren, präsentieren Pörschmann und Wiefing (2015) sowie Pörschmann et al. (2017) Ansätze zur dynamischen, binauralen Auralisation von Schallfeldern auf Basis einer einzigen omnidirektionalen Messung. Neben Systemen zur binauralen Wiedergabe lassen sich ebenfalls Untersuchungen zur Synthese von Nachhall zur Auralisation mittels Wellenfeldsynthese von Melchior et al. (2010); Melchior (2011) und Ahrens (2014, 2015, 2016) finden.

Im Zuge der neuen Möglichkeiten von räumlichem Audio (Spors et al., 2013; Savioja et al., 2015), wurden u.a. von Scheirer et al. (1999); Trivi und Jot (2002); Vaananen und Huopaniemi (2004), Potard und Burnett (2004), Peters et al. (2013), Shirley et al. (2013) und Füg et al. (2014) verschiedene Modelle für objektbasierte Beschreibungen

erarbeitet. Neben den Audiosignalen beinhalten diese Modelle zusätzliche Metadaten beispielsweise zur Definition der Schallquellenposition im Raum. Teilweise wurden diese Möglichkeiten bereits in aktuelle Standardisierungen und Codecs wie z.B. MPEG-4 (ISO/IEC 11172-3, 1999), MPEG-H 3D (Herre et al., 2015) oder ADM (EBU, 2014) integriert. Laut Coleman et al. (2017b) ist jedoch insbesondere eine parametrische Beschreibung des Raumeindrucks für die Wiedergabe von objektbasierten Audioproduktionen von Bedeutung und somit weiterhin aktuell von Interesse.

So finden sich parametrische Modelle zur Synthese von akustischen Umgebungen z.B. von Jot (1997) (SPAT), Savioja et al. (2002) (DIVA), Faller (2004), Pulkki (2007); Vilkamo et al. (2009); Pulkki et al. (2017) (DIRAC), Breebaart et al. (2009), Wendt et al. (2014) (RAZR) und Coleman et al. (2016, 2017a) (RSAO). Die Modelle sind zwar mit unterschiedlicher Komplexität aufgebaut, grundsätzlich bedienen sich aber alle Ansätze den zuvor vorgestellten Möglichkeiten zur Analyse und Synthese der gerichteten und diffusen Schallfeldkomponenten.

1.2 Motivation und Ziele

Die zur Verfügung stehenden Kapazitäten in der digitalen Audio-Signalverarbeitung nehmen durch verbesserte Rechenleistung kontinuierlich zu. Dadurch können in der geometrischen Akustik zur Auralisation von Räumen mittels vorverarbeiteter Datensätze im Hinblick auf infrage kommende Algorithmen zur Simulation, keine Grenzen aufgrund der Verarbeitungszeit mehr aufgezeigt werden. Es kommen Verfahren zum Einsatz, die das Schallfeld im Raum mit Methoden der Strahlenverfolgung (Krokstad et al., 1968) physikalisch korrekt nachbilden können und auch komplexe Zusammenhänge wie Diffraktion (Beugung) und Streuung sowie angekoppelte Volumina exakt berechnen lassen, u.a. im RAVEN-System nach Schröder und Vorländer (2011). Trotzdem können noch immer perzeptive Unterschiede zwischen einer Auralisation auf Basis simulierter oder messtechnisch erfasster, binauraler Datensätze beobachtet werden (Brinkmann et al., 2017a). Dabei definiert die individuelle, datenbasierte, dynamische Binauralsynthese nach Brinkmann et al. (2017b) den Stand der Technik, sodass mit dieser Methode weiterhin die geringsten Unterschiede im Vergleich zur realen Situation erreicht werden. Einschränkungen einer geometrischen Raumsimulation ergeben sich z.B. durch eine nicht ideale Abbildung der Geometrie des Raumes anhand des CAD-Modells sowie durch die Definition der Absorptionskoeffizienten der verwendeten Materialien in der Simulation, die Differenzen zur Ist-Situation im Raum aufweisen.

Daher wird in dieser Arbeit ein alternativer Ansatz gewählt und Methoden der geometrischen Raumakustik zur Simulation von Schallfeldern nicht betrachtet. Stattdessen

wird die menschliche Perzeption von Räumen anhand einer messtechnisch erfassten, binauralen Referenz-Raumimpulsantwort in den Fokus gestellt. Einschränkungen für die Hörbarmachung, die sich aus Ungenauigkeiten einer geometrischen Simulation ergeben, werden durch diesen Ansatz vermieden, da die akustischen Eigenschaften des Raumes inhärent in den Messdaten abgebildet werden.

Grundsätzlich soll die Arbeit dem Erkenntnisgewinn über die menschliche Wahrnehmung von Schallfeldern in Räumen dienen. Für die auditive Perzeption essentielle Merkmale sollen festgestellt und anschließend näher untersucht werden. Um die Evaluation systematisch durchführbar zu machen, wird ein System zur Synthese binauraler Raumimpulsantworten benötigt. So wird nicht angestrebt, das Schallfeld physikalisch korrekt abzubilden; vielmehr ist das Ziel, die für die menschliche Wahrnehmung von Räumen relevanten Aspekte aus einer Messung vor Ort zu extrahieren und mittels Parametern zu charakterisieren.

Unter dem Einsatz von sphärischen Mikrofonarrays kann das Schallfeld räumlich zerlegt und detailliert beschrieben werden. Diese Verfahren haben sich bisher zur raumakustischen Analyse bewährt, nun gilt es, deren Praktikabilität zur Hörbarmachung zu evaluieren. Auf Basis von Parametern, die mittels Messung gewonnen werden, soll daraufhin eine neue Raumimpulsantwort algorithmisch und möglichst entsprechend der akustischen Merkmale der Referenz-BRIR generiert werden.

Wie zuvor gesehen, existieren zahlreiche Systeme zur Generierung von synthetischen Raumimpulsantworten auf Basis von Messungen im Raum. Eine direkte Gegenüberstellung der Methoden mit einer messtechnisch erfassten, binauralen Referenz fehlt jedoch in der Regel. Daher wird diese systematische und vergleichende Untersuchung aller Verarbeitungsschritte in dieser Arbeit angestrebt. Die zusätzliche Erfassung einer binauralen Referenz ermöglicht eine präzise Bewertung der Methoden, da die zu entwickelnden Algorithmen perceptiv mit der messtechnisch erfassten BRIR direkt verglichen werden können. Im Falle einer Praktikabilität der angestrebten Verfahrensweisen ließe sich eine akustische Gegebenheit so mit vergleichsweise wenig Aufwand charakterisieren und hörbar machen, bei gleichzeitigem Erhalt der Flexibilität und Möglichkeiten der Adaption der Daten, entsprechend einer geometrischen Simulation.

Aus dieser Motivation heraus, lassen sich eine Reihe von wissenschaftlichen Fragestellungen und Zielen formulieren, die es in dieser Arbeit nach Möglichkeit zu beantworten und erreichen gilt. Es soll geklärt werden, ob die angestrebten Reduzierungen des Schallfeldes entsprechend der Perzeption realisierbar als auch praktikabel sind. Ferner muss beantwortet werden, welche Merkmale des Schallfeldes für die menschliche Wahrnehmung von Räumen essentiell sind und welche hingegen nicht betrachtet werden müssen sowie ob eine hinreichende Approximation dieser Merkmale in Form von Parametern möglich ist. Weiterhin gilt es zu beantworten, ob diese Herangehenswei-

sen zu einer Datenreduzierung führen und welche Messsysteme zur Charakterisierung des Schallfeldes notwendig sind. Die Verwertbarkeit von Mikrofonarrays zur parametrischen Schallfeldbeschreibung muss ebenfalls überprüft werden.

Um die Untersuchungen systematisch durchführen zu können, bedarf es der praktischen Umsetzung des vorgestellten Ansatzes und somit einer nutzbaren Applikation zur Generierung von synthetischen binauralen Raumimpulsantworten zur dynamischen Auralisation auf Basis von akustischen Merkmalen einer realen Situation. Grundsätzlich muss evaluiert werden, welche perceptiven Unterschiede zwischen einer Auralisation mittels messtechnisch erfasster BRIRs im Vergleich zu einer Verwendung von synthetisch generierten BRIRs bestehen. Dabei sollen nach Möglichkeit die verschiedenen Merkmale des Schallfeldes sowie verwendete Algorithmen sowohl unabhängig voneinander, als auch in Kombination evaluiert werden können. Die Verfahrensweisen sollen zudem unter dem Einsatz verschiedener Signale und in unterschiedlichen akustischen Umgebungen untersucht werden, um eine möglichst allgemeingültige Bewertung zu erzielen.

Die Synthese soll die für die menschliche Wahrnehmung von Räumen relevanten Reize betrachten und bei Verwendung in der Auralisation zu möglichst geringen perceptiven Unterschieden im Vergleich zur binauralen Referenz führen. Im Vordergrund steht die menschliche Wahrnehmung, sodass eine Reduktion entsprechend des auditiven Systems angestrebt wird - nicht wahrnehmbare Reize sollen verworfen werden. Weiterhin sollen alle Verarbeitungsschritte und Methoden offen für Erweiterungen und/oder Änderungen sein und der wissenschaftlichen Gemeinschaft Open Source zur Verfügung gestellt werden. Diese Vorgehensweise ist zum einen für die Grundlagenforschung in der Binauraltechnik von Interesse und ist nützlich für das Verständnis der Zusammenhänge der auditiven Perception. Zum anderen kann ein Forschungswerkzeug geschaffen werden, auf dessen Basis weitere Studien durchgeführt werden können. Durch eine Integration der Verfahren in vorhandene Systeme zur virtuellen Realität kann weiterhin dieser aufstrebende Markt unterstützt werden und zu neuen Produkten und Anwendungen führen. Dazu sind jedoch zuvor weitreichende Untersuchungen der Möglichkeiten der BRIR-Synthese vonnöten, die im Folgenden in dieser Arbeit betrachtet werden.

1.3 Aufbau der Thesis

Die Arbeit ist wie folgt strukturiert: In Kapitel 1 findet sich eine kurze Einführung in die Thematik im Hinblick auf die drei relevanten Bereiche Binauraltechnik, Mikrofonarrays und parametrische Kodierung für Audiosignale. Der Stand von Wissenschaft und Technik im Bezug auf diese Forschungsfelder sowie die Synthese von binauralen

Raumimpulsantworten mittels parametrischer Modelle werden vorgestellt. Die Motivation für die durchgeführten Forschungen und die beabsichtigten Ziele werden erläutert. Weiterhin werden für das Verständnis der Arbeit notwendige Bezeichnungen und Konventionen aufgezeigt.

In Kapitel 2 werden grundlegende Aspekte der Binauraltechnik im Bezug auf die virtuelle Akustik und die angestrebten Untersuchungen dargestellt. Die binaurale Wahrnehmung des Menschen sowie die binaurale Messtechnik werden theoretisch betrachtet. In der dynamischen Binauralsynthese werden diese beiden Aspekte kombiniert, sodass diese Vorgehensweise ebenfalls erläutert wird. Das Prinzip der Wellenfeldanalyse mittels sphärischen Mikrofonarrays wird ebenfalls im Hinblick auf dessen Einsatz in der Binauraltechnik beleuchtet und Grenzen der Systeme definiert.

In Kapitel 3 findet sich ein Überblick über die genutzten Methoden, Technologien und Ressourcen. Dies beinhaltet die verwendeten Messsysteme und Räumlichkeiten, externe Datensätze und genutzte Software. Weiterhin werden die technische Umgebung und die Versuchsanordnung der perceptiven Untersuchungen definiert und statistische Analyseverfahren zur Auswertung aufgezeigt.

In Kapitel 4 wird die zur BRIR-Synthese in dieser Arbeit verwendete Signalverarbeitung erläutert. Alle Verarbeitungsschritte werden dabei unabhängig voneinander betrachtet. Die Prozesse sind anhand des Schallfeldes in gerichtete und diffuse Komponenten separiert, entsprechend ist auch das Kapitel aufgebaut.

In Kapitel 5 werden die in dieser Arbeit entwickelten parametrischen Modelle zur Synthese binauraler Raumimpulsantworten vorgestellt und untersucht. So werden die zuvor erläuterten Methoden und Verarbeitungsschritte kombiniert und führen zur Generierung von synthetischen BRIRs mittels vier Modellen, anhand dieser das Kapitel strukturiert ist. Dabei werden die Unterschiede der Synthese der Modelle im Vergleich zu einer messtechnisch erfassten Referenz *technisch* mittels raumakustischer Kriterien sowie *perzeptiv* im Rahmen mehrerer Hörversuche evaluiert.

Kapitel 6 fasst die Arbeit zusammen, nennt konkret erreichte Ziele und legt bisher unbeantwortete Fragen offen. Weiterhin findet sich ein Ausblick und es werden mögliche praktische Anwendungen der entwickelten Verarbeitungsschritte vorgestellt.

1.4 Nomenklatur

In diesem Abschnitt werden wesentliche Begrifflichkeiten, Konventionen und Bezeichnungen erläutert, auf die in dieser Arbeit zurückgegriffen wird. Eine Übersicht der verwendeten Abkürzungen, Bezeichnungen, Variablen und mathematische Symbole findet sich in Anhang B.

Die folgende Arbeit setzt sich mit Wahrnehmungsprozessen des *auditiven* Systems auseinander und behandelt damit Vorgänge „das Gehör betreffend“ (DIN 1320:2009-12, 2009). Der gesunde Mensch nimmt seine Umwelt in der Regel mit zwei Ohren wahr, sodass entsprechend dieser Perzeption meist *binaurale* – „das beidohrige Hören betreffend[e]“ (DIN 1320:2009-12, 2009) – Signale betrachtet werden.

Unter *Synthese* versteht man im Allgemeinen die Zusammenführung von zwei oder mehr Elementen zu einem neuen Gefüge. Der Ausdruck ist zurückzuführen auf die griechische Bezeichnung *σύνθεσις* / *synthesis* und bedeutet „Zusammensetzung, -stellung, -fügung“ (von *syn* „zusammen“ und *thesis* „das Setzen, Stellen, Legen“). In dieser Arbeit werden mit *Synthese* die Prozesse zur Erzeugung von Datensätzen verschiedener Art bezeichnet, in der Regel handelt es sich hierbei um Impulsantworten. Konkret umfassen diese Prozesse alle Methoden zur algorithmischen Generierung von Schallfeldanteilen auf Basis einer parametrischen Beschreibung, mit dem Ziel eine künstliche (= *synthetische*), *binaurale Raumimpulsantwort* zu bilden, die in dieser Arbeit mit *SBRIR* bezeichnet wird. Als *Parameter* werden mathematische Funktionen oder Variablen bezeichnet, die zur Charakterisierung und Approximation der akustischen Merkmale des Schallfeldes eingesetzt werden. Daher findet sich oft die Bezeichnung der *BRIR-Synthese*, um diesen Sachverhalt zu konkretisieren: Verschiedene Parameter (Elemente) werden zur Entwicklung einer neuen, artifiziellen Impulsantwort verwendet (zusammengesetzt), dessen akustische Merkmale perzeptiv möglichst denen einer binauralen Referenz-Messung entsprechen.

Die künstliche Hörbarmachung von akustischen Gegebenheiten mittels virtueller Akustik nennt man auch *Auralisation* (Vorländer, 2007). Bewegungen des Rezipienten werden dabei teilweise (abhängig der gewählten Algorithmen) mit betrachtet und erhöhen so die *Immersion* der VAE durch Abdeckung zusätzlicher Reize. In der Literatur finden sich eine Vielzahl von unterschiedlichen Definitionen für *Immersion* (Nilsson et al., 2016). Die vorliegende Arbeit verwendet hierzu die technische Herangehensweise gemäß Slater et al. (2009):

We reserve the term immersion to stand simply for what the technology delivers from an objective point of view. The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it may be described as being 'immersive'.

So wird *Immersion* als eine objektive Eigenschaft des Rendering-Systems betrachtet, eine Erhöhung der *Immersion* erfolgt durch die Abbildung möglichst vieler Modalitäten beispielsweise durch weitere Sensorik. In Abgrenzung zur *Immersion* berücksichtigt Slater et al. (2009) bei der Definition der *Präsenz* hingegen zusätzlich die subjektive, menschliche Wahrnehmung:

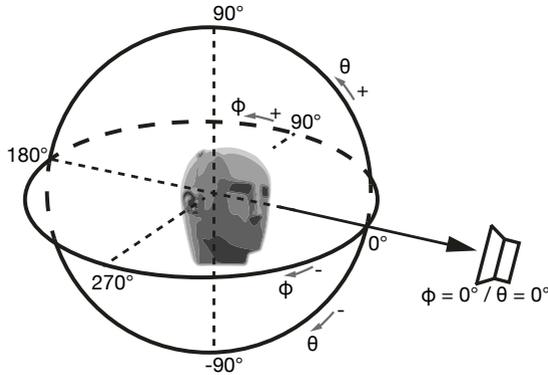


Abbildung 1.1: Darstellung des verwendeten Koordinatensystems nach Blauert (1974), S.11 mit Angabe der Drehrichtung und Position der Schallquelle.

Presence is a human reaction to immersion.

Im Kontext der virtuellen Akustik wird mit Synthese mitunter auch der Prozess zur Auralisation oder gar das Ergebnis der Faltungsoperation im Renderer verstanden. Daher sollten diese Termini hier besser voneinander abgegrenzt werden. Da nur binaurale Systeme zur Auralisation in dieser Arbeit Anwendung finden, kann eine mehrdeutige Bezeichnung vermieden werden, indem diese Systeme stets konkret mit *Binauralsynthese* bezeichnet werden. Zudem arbeiten alle vorgestellten Verfahren auf Basis von Impulsantworten, sodass eine Hörbarmachung erst durch Faltung der Impulsantwort mit einem nachhallfreien Audiosignal, dem sogenannten *Testsignal* möglich wird. Das Ergebnis dieser Operation und damit das Signal welches den Probanden in der dynamischen Binauralsynthese per Kopfhörer präsentiert wird, wird als *Stimulus* bezeichnet. Eine Vielzahl von Wahrnehmungsprozessen sind bei der Perzeption dieser Stimuli durch den Menschen von Bedeutung, sodass im Folgenden gemäß Blauert (1974) alle vom Menschen akustisch wahrgenommenen Reize als *Hörereignisse* bezeichnet werden.

Zur Bewertung einer VAE finden sich häufig die Begrifflichkeiten der *Plausibilität* und der *Authentizität*. Da in dieser Arbeit keine der in Lindau und Weinzierl (2012) und Brinkmann et al. (2017b) vorgeschlagenen Versuchsanordnungen zur Untersuchung dieser Kriterien durchgeführt werden, wird gleichsam auf die Verwendung dieser Begriffe verzichtet. Für das weitere Verständnis ist deren Definition jedoch sinnvoll. Kuhn-Rahloff (2012) definiert das Kriterium der Plausibilität wie folgt:

Plausibilität ist das Ergebnis eines perceptiven Prozesses, der ermittelt, in welchem Maße ein Wahrnehmungsobjekt mit einer aus individuellen Vorerfahrungen resultierenden, inneren Referenz übereinstimmt. Die Referenz des Plausibilitätsurteils ist das Ergebnis aller denkbaren Wahrnehmungsergebnisse innerhalb eines jeweiligen Systems von relationaler und klassifizierender Inferenz. (Kuhn-Rahloff, 2012, S. 99)

Auch Brinkmann et al. (2017b) sowie Lindau und Weinzierl (2012) verwenden ähnliche Begriffe, sodass eine Simulation als plausibel erachtet werden kann, wenn diese mit der Erwartung des Rezipienten im Vergleich zu einer *inneren Referenz* übereinstimmt.

Das Kriterium der Authentizität wird von Blauert (1996) folgendermaßen definiert:

One of the generic tasks of binaural technology [...] is to authentically transmit and reproduce auditory percepts over space and time. Authenticity [...] means that the subjects at the receiving end do not sense a difference between the actual auditory events and those which they would have had at the recording position when the recording was made. (Blauert, 1996, S. 373)

Brinkmann et al. (2017b) verwenden dazu abermals den Begriff der Referenz, sodass eine Simulation als authentisch erachtet wird, wenn diese im direkten Vergleich zur realen Situation (*äußere Referenz*) identisch wahrgenommen wird.

Für die Dokumentation aller räumlichen Komponenten und Orientierungen wird in dieser Arbeit das kopfbezogene Koordinatensystem nach Blauert (1974), S.11 verwendet. Der Azimutwinkel verläuft auf der horizontalen Ebene gegen den Uhrzeigersinn von $0 \leq \phi < 360^\circ$. Änderungen des Elevationswinkel θ auf der Medianebene werden im Bereich von $\pm 90^\circ$ dargestellt, wobei $\theta = 90^\circ$ einer Kopfausrichtung bzw. Position der Schallquelle nach bzw. von oben entspricht. In allen Konfigurationen der Messungen sowie den virtuellen Umgebungen befindet sich die Schallquelle stets frontal zum Messsystem/Nutzer und auf Höhe des Drehmittelpunktes/Ohrs. So kennzeichnet eine Blickrichtung von $\phi = 0^\circ$ und $\theta = 0^\circ$ eine Orientierung des Messsystems/Rezipienten in Richtung Schallquelle. Die räumliche Beschreibung von Schallfeldkomponenten, wie z.B. der Einfallrichtungen von Reflexionen erfolgt gleichsam mit Hilfe dieser Konvention. Eine schematische Darstellung des Koordinatensystems mit Drehrichtung und Position der Schallquelle findet sich in Abbildung 1.1.

Binauraltechnik für die virtuelle Akustik

April 12, 1927.

H. FLETCHER ET AL
BINAURAL TELEPHONE SYSTEM
Filed June 15, 1925

1,624,486

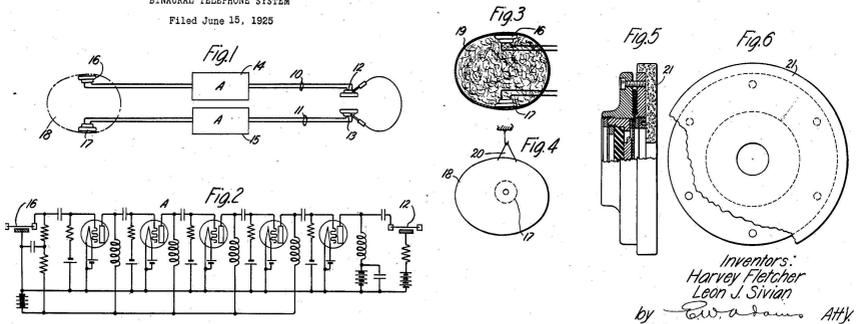


Abbildung 2.1: Ausschnitt aus dem ersten Patent (Anmeldung 1925) zur binauralen Tonaufnahme aus Fletcher und Sivian (1927).

Dieses Kapitel befasst sich mit der Theorie der *Binauraltechnik* im Hinblick auf Anwendungen in der virtuellen Akustik. Es werden grundlegende Aspekte erläutert, die zum Verständnis der in Kapitel 4 vorgestellten Methoden und Signalverarbeitungsalgorithmen erforderlich sind. Weiterhin werden Zusammenhänge und Grenzen der auditiven Wahrnehmung vermittelt, die für die Interpretation der parametrischen Modelle sowie der perzeptiven Untersuchungen in Kapitel 5 von Bedeutung sind.

Die Binauraltechnik bezeichnet im Wesentlichen den Ansatz der Messung oder Synthese von zweikanaligen Ohrsignalen und deren Wiedergabe. Das Ziel ist eine möglichst genaue Reproduktion dieser Signale am Hörort (Møller, 1992). So sollen alle für die auditive Wahrnehmung erforderlichen Informationen und Reize erhalten bleiben. Die binaurale Betrachtung zur Erfassung und Reproduktion von Schallfeldern ermöglicht im Gegensatz zu konventionellen Wiedergabeverfahren eine Reihe von Vorteilen. Besonderheiten und Möglichkeiten, die sich durch das räumliche Hören mit zwei Ohren ergeben und dessen Einfluss auf die Perzeption insbesondere von diffusen Schallfeldern sowie Reflexionen werden in Abschnitt 2.1 erläutert.

Die Binauraltechnik findet heutzutage in ebenfalls Anwendung in der virtuellen Akustik. Dabei werden binaurale Signale oder synthetische Signale, die entsprechende binaurale Eigenschaften aufweisen, verwendet, um mittels dynamischer Rendering-Algorithmen eine VAE zu erzeugen. So kann man sich virtuell in eine zuvor messtechnisch erfasste oder auch gänzlich artifiziell erzeugte akustische Situation auditiv hineinversetzen. Das visuelle Pendant, welches derzeit recht populär ist und durch die Konsumelektronik bereits Einzug in viele Haushalte gefunden hat, ist das *Head-Mounted Display* (HMD) oder die sogenannte *Virtual-Reality*-Brille (Slater, 2009).

Binaurale Rendering-Systeme zur Erzeugung dynamischer VAEs basieren meist auf gemessenen oder synthetischen Impulsantworten. Eine Erläuterung der dazu erforderlichen Messtechnik findet sich in Abschnitt 2.2. Auf Basis von Impulsantworten können Räume anhand raumakustischer Kriterien (siehe 2.3) bewertet werden. In der Binauraltechnik können diese Methoden zur Charakterisierung als auch zur BRIR-Synthese verwendet werden. Die grundsätzliche Funktionsweise der *dynamischen Binauralsynthese* (DBS) und alle erforderlichen Komponenten werden in Abschnitt 2.4 beschrieben.

Die virtuelle Akustik befasst sich neben der Wiedergabe auch mit der Zerlegung von Schallfeldern. Die Wellenfeldanalyse auf Basis von Mikrofonarray-Messungen schafft dabei erweiterte Möglichkeiten zur räumlichen Auswertung und kann für die Binauraltechnik und zur Synthese binauraler Raumimpulsantworten förderlich sein. Daher werden abschließend in Abschnitt 2.5 die Grundlagen dieser Methode zur Schallfeldzerlegung aufgezeigt, die in dieser Arbeit zur Detektion von Reflexionen im Raum und zur Bildung gerichteter Impulsantworten zur Berechnung von Reflexions-Filtern eingesetzt werden.

2.1 Binaurales Hören

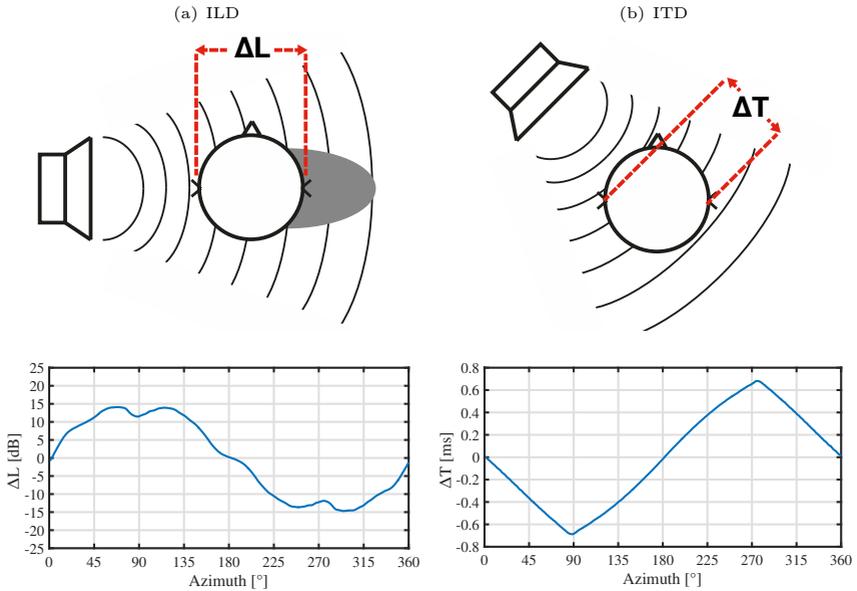


Abbildung 2.2: Schematische Darstellung (oben) von ILD (a) und ITD (b) im Fernfeld in Abhängigkeit vom Einfallswinkel (unten) auf der horizontalen Ebene; berechnet aus HRTF-Daten nach Bernschütz (2013).

Die Aufnahme von Schallereignissen jeglicher Art erfolgt beim Menschen in der Regel binaural, d.h. mit zwei Ohren. Aufgrund des Abstandes der Ohren zueinander, der unterschiedlichen Ausrichtung und Anatomie sowie der Abschattung des Kopfes und des Rumpfes entstehen individuelle Unterschiede zwischen den beiden Ohrsignalen. Dabei können frequenz- und richtungsabhängige *Pegel-* (ILD) und *Laufzeitdifferenzen* (ITD) von eintreffendem Schall auftreten (Blauert, 1996, S.155/S.140), die sogenannten *binauralen Cues*. Insbesondere in Abhängigkeit vom Einfallswinkel der Schallquelle, ergeben sich teilweise große Unterschiede zwischen den beiden Ohrsignalen (siehe Abbildung 2.2). Die Lokalisation in der horizontalen Ebene basiert auf der Auswertung von interauralen Pegel- und Laufzeitdifferenzen.

Neben diesen *interauralen* Differenzen gibt es zudem auch *monaurale* Auslösereize, wie die spektrale und richtungsabhängige Filterung durch Ohrmuschel, Kopf und Rumpf.

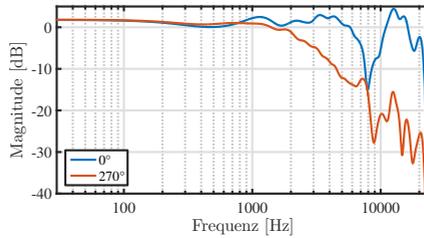


Abbildung 2.3: HRTF-Spektrum für 0° (blau) und 270° (rot) Schalleinfall in der horizontalen Ebene.

Das auditive System kennt die Veränderungen im Amplituden- und Phasenverlauf abhängig von unterschiedlichen Ausrichtungen von Schallquelle und Kopf zueinander. Diese Informationen werden für die Lokalisation in der vertikalen Ebene ausgewertet, da hier aufgrund der Symmetrie des Kopfes nahezu keine interauralen Differenzen auftreten. Blauert (1974), S.30 versteht unter der *Lokalisationsunschärfe* die „kleinste Änderung eines bestimmten Merkmals [...] des Schallereignisses [...] die gerade zu einer Ortänderung des Hörereignisses [...] führt“. Dabei ist Lokalisationsgenauigkeit des auditiven Systems für das *Richtungshören* in der horizontalen Ebene weitaus höher als für vertikalen Schalleinfall. In der Literatur finden sich verschiedene Werte für die Lokalisationsunschärfe, die zudem noch stark signal- und richtungsabhängig sind. In Blauert und Braasch (2008) werden zur Orientierung $\pm 4^\circ$ für die horizontale und $\pm 10^\circ$ für die vertikale Ebene bei frontalem Schalleinfall genannt.

Die Filterwirkung der Außenohren wird durch die *kopfbezogene Außenohrübertragungsfunktion* (HRTF) beschrieben. Messtechnisch wird jedoch meist für verschiedene Schalleinfallrichtungen die *kopfbezogene Impulsantwort* (HRIR) erfasst (siehe Abschnitt 2.2). Durch Überführung der HRIR in den Frequenzbereich wird die HRTF gebildet. Bei Verwendung des kopfbezogenen Koordinatensystems nach Blauert (1974) werden Positionen mittels Polarkoordinaten in Azimuth ϕ , Elevation θ^1 und der Entfernung r zum Ursprung (Kopf-Mittelpunkt) definiert (siehe Abbildung 1.1). Der Gebrauch dieses Koordinatensystems erlaubt eine definierte Beschreibung der Positionen von Schallquellen und Hörereignissen. Eine HRTF ist abhängig von der Schalleinfallrichtung ϕ und θ , der Entfernung r zur Schallquelle sowie der Frequenz f und wird für das linke und rechte Ohr mit $H_l(f, r, \phi, \theta)$ und $H_r(f, r, \phi, \theta)$ bezeichnet. Abbildung 2.3 zeigt den spektralen Verlauf einer HRTF für zwei unterschiedliche Schalleinfallrichtungen (0° und 270°).

¹Hinweis: Blauert (1974) verwendet hierbei δ

Eine weitere Eigenschaft des auditiven Systems ist dessen Fähigkeit, nicht nur die Orientierung von Schallquellen zu bestimmen, sondern in begrenztem Maße auch Aussage über deren Entfernung zu treffen (*Entfernungshören*, siehe Zahorik (2009); Kolarik et al. (2015)). Im Gegensatz zum *Entfernungsssehen* des *binokularen* (beidäugigen) Systems, ist die Entfernungswahrnehmung über binaurale Reize relativ ungenau. Dabei spielen die interauralen Effekte zudem eine untergeordnete Rolle, vielmehr werden die wahrgenommenen Signale in ihrer spektralen Verteilung oder der Lautheit mit bekannten Reizen verglichen. In geschlossenen Räumen unterstützen weiterhin Reflexionen und insbesondere das Verhältnis zwischen Direkt- und Diffusschall (DRR) sowie die *Anfangszeitlücke* (ITDG) – die Zeit zwischen Direktschall und ersten starken Reflexionen – die Entfernungswahrnehmung.

Das menschliche auditive System wertet sowohl monaurale als auch interaurale Merkmale sowie weitere erlernte Muster für die Lokalisation von Schallereignissen und für die Erfassung und Wahrnehmung von Räumen aus. Weiterhin spielen dabei Reflexionen eine große Rolle, daher werden diese Zusammenhänge im kommenden Abschnitt separat erläutert. Durch gezielte Kopfbewegungen (sogenannte Peilbewegungen) können kleine Änderungen der Ohrsignale erzeugt werden, um weitere für das auditive System auswertbare Informationen zu liefern und so die Auflösung und Genauigkeit zu erhöhen.

Zur Darbietung binauraler Signale werden in der Regel kopfhörer-basierte Wiedergabeverfahren verwendet. Man unterscheidet dabei zwischen drei verschiedenen Schalldarbietungsarten, die z.B. in Hörversuchen verwendet und ausgewertet werden können: *monotisch*, *diotisch* und *dichotisch* (DIN 1320:2009-12, 2009). Bei der monotischen Darbietungsart wird nur eine Seite wohingegen bei der diotischen Wiedergabe beide Seiten identisch und bei der dichotischen Reproduktion beide Kanäle des Kopfhörers unterschiedlich beschallt werden (Blauert, 1974, S.77). In der vorliegenden Arbeit werden den Probanden ausschließlich dichotische Signale präsentiert, wie sie in der Realität auch auftreten.

Das binaurale Hören eröffnet neben der Lokalisationsfähigkeit weitere Möglichkeiten. So kann z.B. eine effektive Quelltrennung auch in akustisch anspruchsvollen Umgebungen erzeugt werden. Bei der Anwesenheit mehrerer, gleichzeitig auftretender Signale (wie z.B. Gesprächen), ist es dem Menschen durch das binaurale Hören trotzdem möglich, sich auf ein einzelnes Sprachsignal zu konzentrieren und alle weiteren Störsignale zu unterdrücken (sogenannter Cocktail-Party-Effekt). Des Weiteren kann der Mensch mittels binauralem Hören den wahrgenommen Nachhall unterdrücken und so wiederum die Sprachverständlichkeit in halliger Umgebung erhöhen (Danilenko, 1969). Diese Fähigkeiten des auditiven Systems liegen jedoch nicht im Fokus dieser Arbeit und werden daher nicht weiter betrachtet.

2.1.1 Wahrnehmung von Reflexionen

In geschlossenen Räumen erreicht das von einer Schallquelle abgestrahlte Signal den Hörer nicht nur auf direktem Weg (*Direktschall*), sondern auch nach Rückwurf an den Begrenzungsflächen des Raumes. Dabei entsteht in der Regel eine Pegel- und Zeitdifferenz im Bezug auf den Direktschall. Bei den ersten und stärksten auftretenden Rückwürfen spricht man von den *frühen Reflexionen*. Reflexionen und deren Relation zum Direktschall beeinflussen das Richtungshören, den Raumeindruck, die Klangfarbe und die Distanzwahrnehmung. Damit haben sie eine große Wirkung auf die binaurale Wahrnehmung von Räumen. In Blauert (1974) werden die Zusammenhänge des binauralen Hörens in geschlossenen Räumen dokumentiert. Dabei können dessen Betrachtungen von mehreren Schallquellen auf die Beziehung zwischen Direktschall und im Raum auftretenden Reflexionen – den gerichteten Schallfeldkomponenten – übertragen werden. In Toole (2006) werden zahlreiche Studien zu diesem Thema zusammengefasst.

Für die Wahrnehmung der gerichteten Schallfeldkomponenten und deren Einfluss auf die Richtung des Hörereignisses spielen verschiedene Faktoren eine wichtige Rolle. Dabei werden zunächst ausschließlich die Horizontalebene und zwei nahezu identische Signale, die sich nur durch Pegel- und/oder Laufzeiten unterscheiden, untersucht. Grundsätzlich lässt sich die Betrachtung anhand dieser Differenzen zwischen den Schallereignissen in drei grundlegende Abschnitte einteilen (siehe Abbildung 2.4).

Bei sehr geringen Unterschieden (bis etwa < 1 ms), wird eine *Phantomschallquelle* zwischen den beiden Schallquell-Positionen wahrgenommen, es „entsteht *ein* Hörereignis, dessen Ort von den Orten und den abgestrahlten Signalen beider Schallquellen abhängig ist“ (Blauert (1974), S.163). Dieser Fall wird auch als *Summenlokalisierung* (Warncke, 1941) bezeichnet und das Stereophonieverfahren für die Lautsprecherwiedergabe baut auf diesen Zusammenhängen auf. Durch Variation der Differenzen lässt sich die wahrgenommene Schallquelle zwischen den Lautsprechern verschieben.

Bei größeren Differenzen (> 1 ms) bestimmt die Position der Schallquelle, dessen Signale zuerst am Ohr eintreffen, den Hörereignisort. Nach Blauert (1974), S.163 „entsteht *ein* Hörereignis, dessen Ort durch Ort und abgestrahlte Signale nur einer der beiden Schallquellen bestimmt ist“. Dieser Zusammenhang wird in der Literatur auch als *Gesetz der ersten Wellenfront* (Cremer, 1948) oder als *Präzedenz-Effekt* (Wallach, 1949) beschrieben. So wird ein Schallereignis aus der Richtung der zuerst eintreffenden Wellenfront wahrgenommen. In geschlossenen Räumen führt diese Kausalität trotz auftretender Reflexionen zu einer Wahrnehmung des Hörereignis aus Richtung der Schallquelle. Nach Haas (1951) kann der Pegel eines seitlichen Rückwurfs (bei einer Laufzeitdifferenz < 50 ms) sogar bis zu 6-10 dB höher sein als der Direktschall ohne das ein zweites Hörereignis wahrgenommen wird.

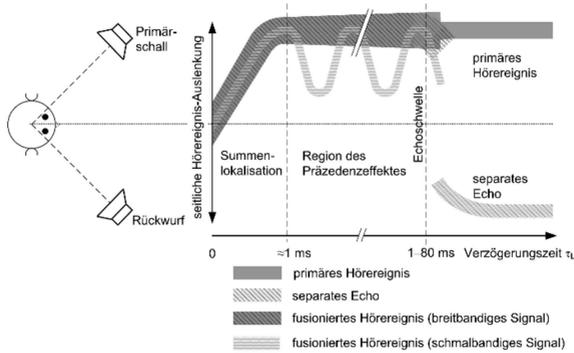


Abbildung 2.4: Hörereignisrichtung in Abhängigkeit von der Laufzeitdifferenz aus Blauert und Braasch (2008).

Überschreitet die Differenz der zwei Schallereignisse die sogenannte *Echo-Schwelle*, so „entstehen *zwei* Hörereignisse. Der Ort des einen hängt mehr oder weniger ausschließlich vom Ort der einen Schallquelle und deren Signalen ab, der des anderen vom Ort der anderen Schallquelle und deren Signalen“ (Blauert, 1974, S.163). Die Echo-Schwelle ist stark abhängig von verschiedenen Variablen wie Signaltyp, Einfallsrichtung und Pegel. In der Literatur finden sich üblicherweise Zeitdifferenzen zwischen Primär-Schall und zweiter Wellenfront von 5 ms für impulshafte Signale, 50 ms für Sprache und bis zu 80 ms für Musik-Signale. Da zwischen Direktschall und Reflexionen meist größere Zeitdifferenzen auftreten, ist diese Betrachtung für die binaurale Wahrnehmung von Reflexionen in geschlossenen Räumen essentiell. In Olive und Toole (1989) werden diese Zusammenhänge für unterschiedliche Konfigurationen intensiv untersucht, erste Abhandlungen dieses Themas geben bereits Seraphim (1961), Burgtorf (1961) und Schubert (1969). Weiterhin finden sich Untersuchungen zu dieser Thematik von Hartmann (1983), Bech (1996a,b, 1998), Gourévitch und Brette (2012) sowie Brown et al. (2014).

Olive und Toole (1989) formulieren die wesentlichen Faktoren zur Wahrnehmung von Reflexionen. Allgemein ist eine gemeinsame Betrachtung von Pegel, Verzögerungszeit, Richtung, Signaltyp und Wiedergaberaum notwendig, um Aussagen über die Wahrnehmbarkeit und die Wirkung von Reflexionen treffen zu können. Die dominante Bedeutung des Direktschalls für die Lokalisation der Schallquelle wird bestätigt. Weiterhin

werden Reflexionen, die aus der selben Richtung wie der Direktschall eintreffen, meist schlechter wahrgenommen als Reflexionen mit anderer Orientierung. Daher können diese 5-10 dB lauter sein, als z.B. seitliche Reflexionen, bis sie wahrgenommen werden. Reflexionen beeinflussen nicht nur das Richtungshören, auch der Raumeindruck infolge der auditiv wahrgenommenen *Räumlichkeit* (spaciousness) wird tangiert. Bradley und Souloudre (1995b) unterteilen den Faktor der Räumlichkeit weiter in die wahrgenommene *Quellbreite* (ASW) und die *Einhüllung* (LEV) des Hörers. Nach DIN EN ISO 3382-1:2009-10 (2009) korreliert der Räumlichkeitseindruck gut mit der Messung der *interauralen Kreuzkorrelation* (IACC). Zur Bestimmung von ASW und LEV werden der frühe bzw. späte Seitenschallgrad herangezogen. Man spricht zudem von einem *Image-Shift*, wenn die wahrgenommene Position der Schallquelle infolge von Reflexionen gemäß deren Einfallsrichtung verschoben oder ausgedehnt wird.

Weiterhin haben Reflexionen Einfluss auf die *Klangfarbe* (timbre) eines Signals. So erhöhen seitliche Reflexionen nach Olive und Toole (1989) die wahrgenommene Räumlichkeit, während vertikale Reflexionen die Klangfarbe beeinflussen. Insbesondere sehr ähnliche und nur gering zeitversetzte Reflexionen, führen zu einer starken Klangverfärbung durch sogenannte Kammfilter. In Rubak und Johansen (2003) findet sich eine umfangreiche Betrachtung dieses Themengebiets. Auch der Signaltyp hat einen signifikanten Einfluss auf die Wahrnehmung von Reflexionen. Impulsive und kontinuierliche Signale führen zu stark unterschiedlichen Wahrnehmbarkeitsschwellen. Das auditive System ist bei pulshaften Signalen weitaus empfindlicher für die Wahrnehmung von Reflexionen als bei Rauschsignalen. Zudem werden Reflexionen in nachhallfreier Umgebung mit geringeren Pegeln wahrgenommen als in mit Nachhall behafteten Situationen. Dabei ist laut Olive und Toole (1989) die Wahrnehmbarkeitsschwelle für kontinuierliche Signale nahezu unabhängig von der Verzögerungszeit.

2.1.2 Wahrnehmung des diffusen Schallfeldes

Man spricht von einem vollkommen diffusen Schallfeld, wenn die folgenden zwei Bedingungen erfüllt sind: Zum einen muss eine gleichmäßige Verteilung der Schalldruckpegel im Raum gegeben sein (Ortsunabhängigkeit). Diese Homogenität hat zur Folge, dass die Schallenergie dichte an jedem Punkt im Raum identisch ist (*räumliche Diffusion*). Zum anderen muss der Energiefluss gleichmäßig über alle möglichen Schalleinfallrichtungen verteilt sein (*gerichtete Diffusion*). Es herrscht somit ein isotropes Schallfeld, d.h. die eintreffende Schallenergie hängt nicht von der Richtung ab und ist für alle Einfallswinkel identisch (Schultz, 1971; Jacobsen, 1979; Jeong, 2016). Oft wird das diffuse Schallfeld daher auch als stochastischer Nachhall bezeichnet. Die stochastische

Betrachtungsweise wurde erstmals von Schroeder (1954) eingeführt, der zudem auch die sogenannte *Schröder-Frequenz*

$$f_{\text{Schroeder}} = 2000 \sqrt{\frac{T}{V}} \quad (2.1)$$

in Abhängigkeit des Volumens V und der Nachhallzeit T definierte (Schroeder, 1996). Diese Anschauungsweise des Schallfeldes ist demnach nur für Frequenzen oberhalb der Schröder-Frequenz gültig.

Im diffusen Schallfeld treten somit keine gerichteten Auslösereize auf, die die Position von Schallquelle und Reflexionen betreffen. Daher dominiert in der Wahrnehmung des späten Nachhalls die auftretende Abklingzeit (Nachhallzeit) und die spektrale Zusammensetzung des Energieabfalls. Die Nachhallzeit T als frequenzabhängige Größe ist definiert als die Zeitspanne, in der der Schalldruck nach Abschalten einer Schallfeldanregung um 60 dB abfällt (Sabine, 1922). Die Bestimmung der Nachhallzeit erfolgt zumeist über die frequenzabhängigen *Energieabklingkurven* (EDC). Nach Schroeder (1965) können diese Kurven durch die Rückwärtsintegration der quadrierten Impulsantwort h folgendermaßen bestimmt werden:

$$\text{EDC}(t) = \int_0^\infty h^2(\tau) d\tau - \int_0^t h^2(\tau) d\tau = \int_t^\infty h^2(\tau) d\tau. \quad (2.2)$$

So kennzeichnet die Funktion $\text{EDC}(t)$ die Summe der in der Impulsantwort auftretenden Energie zum Zeitpunkt t . Durch Bestimmung der Steigung dieser Funktion und Ablesen des Zeitintervalls für den Abfall von 60 dB, kann dann die Nachhallzeit T_{60} ermittelt werden. Laut DIN EN ISO 3382-1:2009-10 (2009) sind durch entsprechende Extrapolation auch geringere Dynamikbereiche möglich. Mit T_{30} wird die Nachhallzeit bezeichnet, die mittels des Intervalls zwischen -5 dB und -35 dB unter dem Maximalpegel und entsprechender Multiplikation bestimmt wird. Analog dazu kennzeichnet T_{20} die Nachhallzeit basierend auf dem Bereich zwischen -5 dB und -25 dB. Des Weiteren ist die *frühe Abklingzeit* (EDT) als die Zeitspanne für einen 60 dB Abfall definiert, die über die Steigung der Abklingkurve zwischen 0 dB und -10 dB bestimmt wird. Nach DIN EN ISO 3382-1:2009-10 (2009), S.19 ist die EDT relevanter für die subjektive Wahrnehmung, während T hingegen den Nachhall eher physikalisch kennzeichnet. In Seraphim (1958) wurde für üblich auftretende Nachhallzeiten (in Räumen oft zwischen 0,5 s und 2 s) eine *relative Wahrnehmbarkeitsschwelle* (JND) von 4 % detektiert.

2.1.2.1 Beginn des diffusen Schallfeldes

Abhängig von den akustischen Gegebenheiten (Raumgröße und Absorptionsverhalten) stellt sich in Räumen bei einer Schallfeldanregung das diffuse Schallfeld nach unterschiedlichen Zeit-Intervallen ein. Der Zeitpunkt, an dem in einer gemessenen Raumimpulsantwort (siehe Abschnitt 2.2) physikalisch das diffuse Schallfeld beginnt, wird in der Literatur meist als *mixing time* (t_m) bezeichnet (Polack, 1992). Es finden sich zahlreiche Publikationen, welche diesen Zeitpunkt raumabhängig anhand verschiedener Eigenschaften schätzen. In der Regel wird hierbei zwischen modell- und datenbasierten Methoden unterschieden. Modellbasierte Prädiktoren verwenden ausschließlich geometrische Eigenschaften des entsprechenden Raumes, wohingegen die datenbasierten Prädiktoren die Raumimpulsantwort analysieren. Die Komplexität ist dabei unterschiedlich hoch und reicht für die modellbasierten Prädiktoren von der Charakterisierung über das Volumen V in m^3 (Cremer, 1948):

$$t_m = \sqrt{V} \text{ [ms]}, \quad (2.3)$$

die freie Weglänge l_m nach Rubak und Johansen (1999):

$$t_m = 4 l_m \frac{10^3}{c} \text{ [ms]} \quad \text{mit } l_m = 4 \frac{V}{S} \text{ [m]} \quad (2.4)$$

in Abhängigkeit von der Oberfläche S des Raumes in m^2 , bis hin zur Berechnung über die Nachhallzeit $T_{500\text{Hz}}$ im 500 Hz Oktavband nach Hidaka et al. (2007):

$$t_m = 80 \cdot T_{500\text{Hz}} \text{ [ms]}. \quad (2.5)$$

Datenbasierte Verfahren analysieren gemessene Impulsantworten und ermitteln auf deren Grundlage die Übergangszeit in den diffusen Nachhall. Abel und Huang (2006) berechnen das Echodichteprofil um den Zeitpunkt zu detektieren, ab dem die Amplituden des Schalldrucks normalverteilt sind. Auch die Methode nach Stewart und Sandler (2007b) basiert auf der Annahme, das die Amplituden eines diffusen Signals im Zeitbereich normalverteilt sind. Anstelle des Echodichteprofiles wird hier jedoch die Kurtosis, einem Maß für die Steilheit einer Häufigkeitsverteilung, gebildet. Weiterhin berechnen Hidaka et al. (2007) die Korrelation zwischen Direktschall und den späten Schallanteilen mittels einer zeitabhängigen Korrelationsfunktion. Der Zeitpunkt, an dem die Korrelation einen definierten Wert unterschreitet, wird dann als *mixing time* angenommen.

Mitunter nimmt der Mensch das Schallfeld bereits zu früheren Zeitpunkten und ohne physikalisch ideale Bedingungen (siehe 2.1.2) diffus wahr und kann nicht mehr

zwischen verschiedenen Orientierungswinkeln oder gar Raumpositionen unterscheiden. Daher wird dieser Zeitpunkt als *perceptual mixing time* (t_{mp}) bezeichnet. So ist aufgrund der menschlichen Wahrnehmungsprozesse t_{mp} immer gleich oder kleiner als t_p . In Lindau et al. (2012) findet sich eine umfangreiche Übersicht und perzeptive Untersuchung aller relevanten Ansätze unter Einsatz der dynamischen Binauralsynthese sowie die Bildung einer Formel mittels linearer Regressionsanalyse zur daten- bzw. modellbasierten Bestimmung der perzeptiven mixing time t_{mp} . Die präziseste Vorhersage der modellbasierten Verfahren erreichte hierbei die Methode in Abhängigkeit des Volumens. Eine Prognose der perzeptiven mixing time kann laut Lindau et al. (2012) mit

$$t_{\text{mp}} = 0,0117 \cdot V + 50,1 \quad (2.6)$$

erfolgen. Der Ansatz von Abel und Huang (2006) brachte für die datenbasierten Methoden die präziseste Vorhersage der perzeptiven Übergangszeit und kann nach Lindau et al. (2012) folgendermaßen berechnet werden:

$$t_{\text{mp}} = 1,8 \cdot t_{m_{\text{Abel}}} - 38. \quad (2.7)$$

Die Kenntnis der wahrgenommenen Übergangszeit in den diffusen Nachhall schafft für den Einsatz in der virtuellen Akustik zahlreiche Möglichkeiten der Datenreduktion und des Rechenaufwandes. Grundsätzlich bauen die Simplifizierungen auf der Annahme auf, dass das Schallfeld ab der perzeptiven mixing time keine für den Hörer wahrnehmbaren Richtungsinformationen beinhaltet und daher auf jegliche Nachführung der Orientierung verzichtet werden kann (siehe Abschnitt 2.4). In Stade (2015) konnten jedoch auch bei sehr späten Übergangszeiten in den diffusen Nachhall weiterhin perzeptive Unterschiede zwischen Referenz-Messungen und den vereinfachten Datensätzen gezeigt werden. Es wurde zwar eine hohe Ähnlichkeit zwischen Referenz und Bearbeitung beobachtet, eine Ununterscheidbarkeit konnte aber selbst für sehr späte Übergangszeiten nicht erreicht werden.

2.1.2.2 Interaurale Kohärenz

Wie zuvor beschrieben ergeben sich durch die Positionierung der Außenohren und der Anatomie des Kopfes bzw. Rumpfs interaurale Unterschiede zwischen den Ohrsignalen. Diese beeinflussen die Wahrnehmung von Schallfeldern und insbesondere perzeptive Bewertungen subjektiver Maße wie *Quellbreite*, *Einhüllung* sowie der *Räumlichkeit* hängen von der Ähnlichkeit der Ohrsignale ab (Blauert, 1996). Bereits Chernyak und Dubrovsky (1968) zeigten den Einfluss auf die Wahrnehmung der Schallquellenposition und deren Ausdehnung abhängig vom Ähnlichkeitsgrad der Ohrsignale bei der

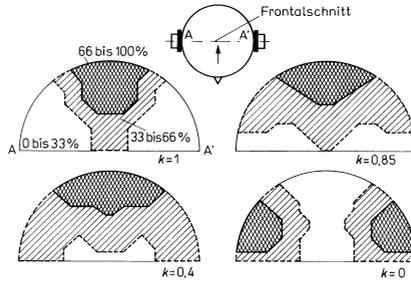


Abbildung 2.5: Wahrnehmung der Schallquellenposition und Ausdehnung in Abhängigkeit der interauralen Kohärenz k aus Chernyak und Dubrovsky (1968).

Wiedergabe über Kopfhörer (siehe Abbildung 2.5). Dabei vergrößerte sich die wahrgenommene Ausdehnung der Schallquelle mit abnehmender Ähnlichkeit bis schließlich zwei separierte Hörereignisse wahrgenommen wurden. Um die Unterschiede zwischen den beiden Signalen zu dokumentieren, wird die Gleichartigkeit des linken und rechten Ohrsignals mit Hilfe eines Einzahlwertes zwischen 0 und 1 bewertet. Dabei kennzeichnet der Wert 0 vollkommen unähnliche Signale, während der Wert 1 bei identischen Signalen auftritt.

In der Binauraltechnik und der Synthese von diffusen Schallfeldern wird dazu meist die *interaurale Kohärenz* (IC) zur Analyse der Gleichheit der Ohrsignale verwendet. Dabei werden mit S_l und S_r die Fouriertransformierten der linken und rechten Ohrsignale $p_l(t)$ und $p_r(t)$ gebildet und so mit

$$IC_{S_l, S_r}(f) = \frac{|\langle S_l(f) S_r^*(f) \rangle|}{\sqrt{\langle S_l(f) S_l^*(f) \rangle \langle S_r(f) S_r^*(f) \rangle}} \quad (2.8)$$

die frequenzabhängige Kohärenz (Bendat und Piersol, 1971) des binauralen Signalpaares gebildet. Dabei bezeichnet $(\cdot)^*$ die konjugiert komplexe Zahl und $\langle x \rangle$ den Erwartungswert von x . Menzer und Faller (2010) erweitern diese Vorgehensweise durch die Verwendung des Realteils anstelle des Betrages im Zähler in Formel (2.8), da laut Borß und Martin (2009), der Imaginärteil klein gegenüber dem Realteil ist. So werden zwar zu Formel 2.8 vergleichbare Werte ermittelt, jedoch wird die interaurale Kohärenz nun im Wertebereich zwischen -1 und 1 berechnet, sodass auch eine negative Korrelation zwischen den Ohrsignalen darstellbar ist. In Cook (1955) wird ebenfalls dieser Be-

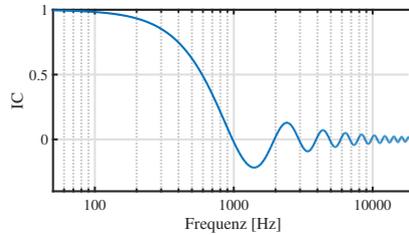


Abbildung 2.6: Theoretische interaurale Kohärenz in einem idealen diffusen Schallfeld nach Cook (1955) für zwei omnidirektionale Mikrofone im Abstand von $d = 17,5$ cm.

reich verwendet um die interaurale Kohärenz über der Frequenz f eines ideal diffusen Schallfeldes mit

$$IC(f) = \frac{\sin\left(\frac{2\pi}{\lambda}d\right)}{\frac{2\pi}{\lambda}d} \quad (2.9)$$

zu definieren. Dabei ist d der Ohrabstand und $\lambda = c/f$, c bezeichnet die Schallgeschwindigkeit. Es zeigt sich der typische Verlauf der interauralen Kohärenz über der Frequenz im diffusen Schallfeld: Zu tieferen Frequenzen unterhalb von 1000 Hz werden die Ohrsignale immer ähnlicher, bis sie bei 100 Hz nahezu identisch sind (siehe Abbildung 2.6).

Die interaurale Kohärenz ist zudem gemäß Abbildung 2.7 a stark zeitabhängig. Während sich im späten Nachhall in der Regel der in Abbildung 2.6 gezeigte Verlauf einstellt, sind die Ohrsignale bei der Anwesenheit von Direktschall und ersten Reflexionen meist deutlich ähnlicher. Des Weiteren hängt die Kohärenz aber für frühe Schallfeldanteile in zunehmenden Maße von der Quellorientierung ab (siehe Abbildung 2.7 b), im diffusen Schallfeld ist diese nach Menzer (2010) vernachlässigbar (siehe Abbildung 2.7 c). In der Binauraltechnik ist die interaurale Kohärenz eine wichtige Eigenschaft, welche die Wahrnehmung von diffusen Schallfeldern mit zwei Ohren entscheidend charakterisiert.

2.2 Binaurale Messtechnik

Lineare, zeitinvariante (LTI)-Systeme lassen sich durch Impulsantworten beschreiben. Diese können z.B. zur Bildung der Übertragungsfunktion eines Systems eingesetzt werden. So können Zeit- und Frequenzverhalten eines Systems dargestellt und analysiert werden (Schüßler, 1994, S.221). Bei konstanten klimatischen Bedingungen und ohne die Anwesenheit von Störquellen im Raum können die Übertragungspfade der Quelle

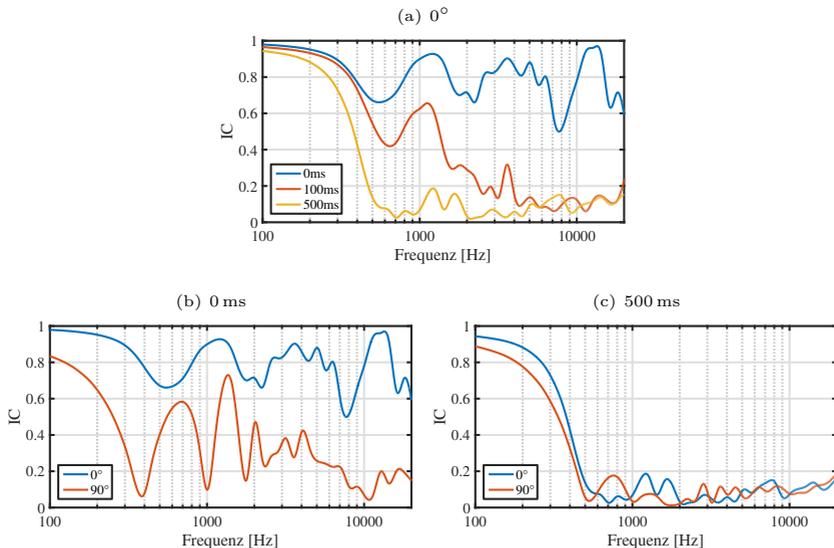


Abbildung 2.7: IC für verschiedene Integrationszeiten (Startzeitpunkt) und horizontale Orientierungen: (a) 0° mit 0 ms (blau), 100 ms (rot) bzw. 500 ms (gelb); (b) 0 ms mit 0° (blau) bzw. 90° (rot); (c) 500 ms mit 0° (blau) bzw. 90° (rot); BRIRs aus KVB-Saal.

zum Empfänger als LTI-Systeme aufgefasst werden, die sich mittels einer Raumimpulsantwort (RIR) charakterisieren lassen (Schüfler, 1994, S.223). Eine RIR bildet den zeitlichen Verlauf des Schalldruckes in einem Raum nach einer Schallanregung ab und liefert damit Informationen über seine akustischen Eigenschaften, wie z.B. auftretende Reflexionen, das diffuse Abklingverhalten oder die Nachhallzeit T .

Nach Anregung des Raumes mittels eines Dirac-Impulses ergibt sich typischerweise der in Abbildung 2.8 aufgezeigte Verlauf mit Direktschall (grün), frühen Reflexionen (rot) und diffusem Nachhall (blau). Da sich ein idealer Dirac-Impuls nur unzureichend erzeugen lässt, erfolgt die Messung einer RIR jedoch zumeist mit anderen Methoden. Zum einen können stochastische Rauschsignale verwendet werden und das Abklingen des Raumes nach Abschalten der breitbandigen Anregung wird aufgezeichnet. Durch Berechnung der Kreuzkorrelation des Eingangsrauschens und des aufgezeichneten Signals kann die Impulsantwort gebildet werden. Wie beim Verfahren mittels Dirac-Impuls, ist die Energieeffizienz bei der Methode des abgeschalteten Rauschen unzureichend, da das gesamte Spektrum gleichzeitig abgespielt werden muss und so die verwendeten

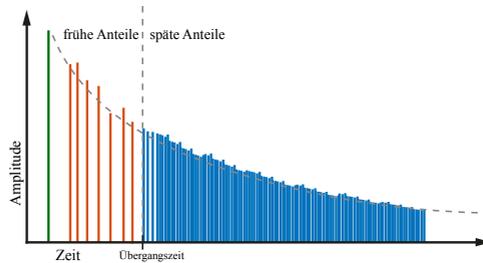


Abbildung 2.8: Typischer Verlauf einer Raumimpulsantwort als Funktion des Schalldrucks über der Zeit mit Direktschall (grün), frühen Reflexionen (rot) und diffusem Nachhall (blau); Trennung in frühe und späte Anteile anhand der Übergangszeit.

Gerätschaften unnötig belastet werden. Daher haben sich mittlerweile in der raumakustischen Messtechnik Verfahren mit deterministischen Signalen z.B. mittels Sinussweep-Anregung und Entfaltung gemäß Müller (1999) etabliert. Insbesondere bei der Anwendung von langen Sweep-Signalen ist dieses Verfahren zudem recht robust gegenüber bei der Messung auftretende Störsignale und ermöglicht höhere Signal-Rauschabstände. In dieser Arbeit wird zur Bestimmung der RIRs ausschließlich die Anregung mittels Sinussweep verwendet (siehe Abbildung 2.10). Das Frequenzverhalten des zur Anregung verwendeten Lautsprechers sollte möglichst linear sein und gegebenenfalls durch eine Entzerrung optimiert werden.

Klassischerweise werden RIRs mit Mikrofonen, die eine omnidirektionale Richtcharakteristik aufweisen, aufgezeichnet. In der Binauraltechnik werden jedoch auch sogenannte Kunstköpfe zur Bestimmung der Impulsantworten eingesetzt. Dabei handelt es sich um Nachbildungen des menschlichen Kopfes in Originalgröße mit Ohrmuscheln und in den Gehörgängen positionierten Mikrofonen. Die Mikrofone besitzen ebenfalls eine Kugel-Richtcharakteristik, durch Abschattung und Positionierung entstehen jedoch entsprechend des menschlichen Hörorgans interaurale Unterschiede zwischen den Mikrofonensignalen und eine richtungsabhängige (monaurale) Filterwirkung aufgrund der Außenohren. Herstellerspezifisch erfolgt die Ausführung des Kunstkopfes unterschiedlich detailgetreu, sodass sich Modelle mit stark vereinfachten bis sehr akkuraten Ohrmuscheln finden lassen. Zudem werden von einigen Herstellern komplette Torso-Nachbildungen geliefert, wohingegen der in dieser Arbeit verwendete *Neumann KU100* Kunstkopf ausschließlich die Ausmaße des menschlichen Kopfes stilisiert. Die Außenohren sind bei diesem Modell sehr präzise nachgebildet. In reflexionsfreier Umgebung können Kunstköpfe zur HRIR-Messung verwendet werden. Dabei werden zwar keine individuellen, anatomischen Unterschiede wie Ohrabstand oder Aussehen der Gehör-



Abbildung 2.9: BRIR-Messung für verschiedenen Orientierungen im KVB-Saal mit *Neumann KU100* Kunstkopf aus Stade et al. (2012).

muscheln erfasst, jedoch bietet dieses Verfahren die Möglichkeit zur Aufnahme von Impulsantworten mit hohem Signal-Rauschabstand und nahezu beliebiger räumlicher Abtastung.

Bei der Messung von individuellen HRIRs werden wiederum kleine Mikrofone in den Gehörgängen der Probanden platziert, die dann möglichst ohne Bewegung die Messung absolvieren. Möchte man nun Impulsantworten für viele Orientierungen erhalten, so sind komplexe Multi-Lautsprecher-Systeme nötig, welche die Messdauer möglichst kurz (und für den Probanden durchführbar) halten (siehe Fuß et al. (2015) oder Richter et al. (2016)). In der vorliegenden Arbeit werden ausschließlich nicht-individuelle, von Bernschütz (2013) mit einem *Neumann KU100* Kunstkopf gemessene HRIRs verwendet. Dies hat zudem den Vorteil, dass die perzeptive Evaluation für jeden Probanden auf den selben Daten basiert. So werden weiterhin nicht die individuellen HRIR-Unterschiede untersucht, da dies auch nicht im Fokus dieser Arbeit liegt. Der Gebrauch individueller Daten ist jedoch denkbar und zudem (wie später gezeigt werden wird) leicht in die untersuchten Modelle integrierbar. Weitere, mit Kunstköpfen gemessene HRIR-Datenbanken finden sich z.B. von Gardner und Martin (1994) (KEMAR Kunstkopf) und Brinkmann et al. (2017c) (FABIAN Kunstkopf), individuelle Datensätze z.B. bei Algazi et al. (2001) (45 Probanden) und Bomhardt et al. (2017) (48 Probanden) oder in der LISTEN² und der ARI³ HRTF-Datenbank.

Wird ein Kunstkopf zur Messung von Impulsantworten nicht in reflexionsfreier Umgebung, sondern in gewöhnlichen Räumlichkeiten eingesetzt, so werden zusätzlich zu

²<http://recherche.ircam.fr/equipes/salles/listen/>

³<https://www.kfs.oeaw.ac.at/index.php?lang=de>

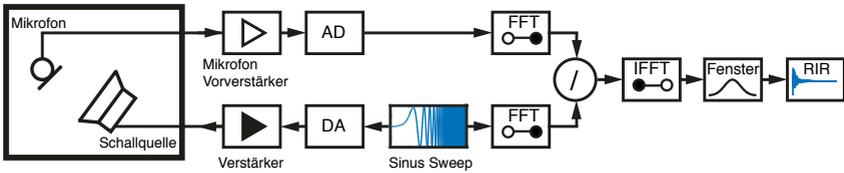


Abbildung 2.10: Schematische Darstellung der RIR-Messung mittels Sinus-Sweep Anregung und Entfaltung im Frequenzbereich nach Müller (1999).

den binauralen Merkmalen auch die akustischen Eigenschaften des Raumes erfasst. Man spricht dann von binauralen Raumimpulsantworten (BRIRs), eine exemplarische Mess-Situation findet sich in Abbildung 2.9 mit *Neumann KU100* Kunstkopf als Empfänger und *SonicBall* Dodekaeder als Schallquelle (aus Stade et al. (2012)). Eine BRIR kennzeichnet ausschließlich die Raumeigenschaften am Messort und für die in der Messung verwendete Kombination aus Sender und Empfänger. Somit handelt es sich bei einer BRIR immer zwingend um zwei Impulsantworten (rechtes und linkes Ohrsignal), die fest miteinander verknüpft sind. Weiterhin können BRIRs sequenziell für verschiedene Orientierungswinkel in Relation zur Schallquelle durch Rotation des Kunstkopfes erfasst werden (siehe 3.1.2). Auch BRIRs können durch die Verwendung von Miniaturmikrofonen in den Gehörgängen der Probanden individuell erfasst werden. Die akkurate Messung von zahlreichen Orientierungswinkeln in hoher räumlicher Auflösung ist jedoch wenig praktikabel, da Bewegungen der Probanden nicht ausgeschlossen werden können. Zudem ist die Messdauer aufgrund der auftretenden Nachhallzeiten erheblich länger als bei HRIR-Messungen in nachhallfreier Umgebung, wodurch die Messung zusätzlich erschwert wird. Wie bereits vorgestellt, können BRIRs aber auch synthetisch erzeugt werden, die vorliegende Arbeit intendiert die perzeptive Untersuchung und Praktikabilität dieser Methoden.

2.3 Raumakustische Kriterien

Durch die Auswertung von Raumimpulsantworten können eine Vielzahl sogenannter raumakustischer Kriterien bestimmt werden. Diese Kennzahlen haben sich zur Charakterisierung von akustischen Gegebenheiten bewährt und werden insbesondere bei der Konstruktion und Bewertung von Konzertsälen und Auditorien eingesetzt (Beranek, 2004; Kuttruff, 2009). Weiterhin können sie auch zu Vergleichszwecken herangezogen werden, um die akustischen Eigenschaften von Räumen gegenüberzustellen. Typische raumakustische Parameter sind laut Rossing (2015) S.313 die Nachhallzeit T , die frü-

Tabelle 2.1: Unterscheidungsschwellen raumakustischer Kriterien nach DIN EN ISO 3382-1:2009-10 (2009).

EDT	<i>G</i>	<i>C80</i>	<i>D50</i>	T_S	<i>IACC</i>
5 %	1 dB	1 dB	5 %	10 ms	0,075

he Abklingzeit EDT, das Stärkemaß G , das Klarheitsmaß $C80$, der Seitenschallgrad LEF , die interaurale Kreuzkorrelation $IACC$ und die frühe Support-Zeit ST_{early} . Die entsprechenden Definitionen dieser Kriterien finden sich ebenfalls in Rossing (2015). Weiterhin gibt Vorländer (2011) eine Übersicht des aktuellen Wissensstandes in der Raumakustik und nennt als bedeutsamste Faktoren zur Charakterisierung von akustischen Umgebungen deren Lautheit, Halligkeit und Raumeindruck. Viele akustische Kriterien sind zudem in DIN EN ISO 3382-1:2009-10 (2009) standardisiert und von Kuusinen und Lokki (2017) im sogenannten „wheel of concert hall acoustics“ anhand von Attributen miteinander verknüpft.

Es finden sich zahlreiche Publikation zur Untersuchung der JNDs akustischer Kriterien (Seraphim, 1961; Cox et al., 1993; Bradley und Soulodre, 1995a; Bradley et al., 1999; Okano, 2002; Zahorik, 2009). Aufgrund differierender Signale und Frequenzbereiche, sind diese Schwellen jedoch nur bedingt auf andere Untersuchungen übertragbar. Nach Bradley (2011) ist das Wissen über JNDs raumakustischer Kriterien bisher limitiert und daher ausbaufähig. Weiterhin finden sich laut Bradley verschiedene Methoden zur Berechnung eines über der Frequenz gemittelten Einzahlwertes, was eine Vergleichbarkeit erschwert. Daher müssen die Wahrnehmbarkeitsschwellen bedachtsam interpretiert werden. So bedarf die Verknüpfung dieser technischen Kriterien mit der auditiven Wahrnehmung weiterer Forschung. Tabelle 2.1 zeigt JNDs für ausgesuchte raumakustische Kriterien nach DIN EN ISO 3382-1:2009-10 (2009), die im Folgenden als Kennzahlen zur Bewertung herangezogen werden können.

Es gilt zudem zu beachten, das für die verschiedenen Kriterien mitunter unterschiedliche Integrationszeiten verwendet werden. Die Kriterien basieren außerdem zum Teil auf unterschiedlichen Eingangssignalen, insbesondere die geforderte Richtcharakteristik für die Erfassung der RIRs unterscheidet sich im Detail. So werden omnidirektionale (Kugel-Charakteristik), binaurale oder auch bi-direktionale (Acht-Charakteristik) Impulsantworten benötigt. Üblicherweise werden die Maße breitbandig bestimmt. So wird die Nachhallzeit zwar oft in Terzbändern kalkuliert, eine Betrachtung der weiteren Kriterien mit einer höheren Auflösung als Oktav-Bandbreite macht laut Rossing (2015), S.315 nur in begrenztem Maße Sinn, da die Kombination schmaler Frequenzbänder und Zeitintervalle zu großen Schwankungen führt. Diese stimmen jedoch nicht mit der subjektiven Wahrnehmung überein. Üblicherweise liegt der betrachtete Frequenzbe-

reich für raumakustische Messungen zwischen 125 - 4000 Hz, eine Erweiterung auf 63 - 8000 Hz ist meist sinnvoll und wird von Rossing (2015) empfohlen.

2.4 Dynamische Binauralsynthese

In der virtuellen Akustik existieren verschiedene Methoden, die mittels digitaler Signalverarbeitung die Hörbarmachung von akustischen Gegebenheiten ermöglichen. Eine Unterscheidung kann anhand der eingesetzten Wiedergabesysteme (Lautsprecher oder Kopfhörer) und den verwendeten Algorithmen erfolgen. In dieser Arbeit wird für die Auralisation in den perzeptiven Experimenten ausschließlich auf die *dynamische Binauralsynthese* (DBS) mittels Kopfhörer zurückgegriffen. Grundlage dieser Rendering-Methode sind gemessene oder simulierte binaurale Impulsantworten für verschiedene Kopf-Orientierungen. Dabei kann es sich um Außenohrimpulsantworten (HRIRs) aus nachhallfreier Umgebung oder aber um Impulsantworten mit Rauminformationen (BRIRs) handeln. Auch die Verwendung individueller, am Probanden gemessener Datensätze ist möglich. In dieser Arbeit finden ausschließlich mittels Kunstkopf, in Nachhall-behafteten Umgebungen, messtechnisch erfasste BRIRs und deren synthetische Pendant (SBRIRs) Anwendung. Abhängig von den gewünschten Freiheitsgraden der möglichen Bewegung und der Auflösung, die bei der Auralisation berücksichtigt werden sollen, sind die Impulsantworten im Vorfeld entsprechend zu messen oder synthetisch zu generieren. Systeme zur dynamischen Binauralsynthese finden sich in der Literatur bereits vor nahezu 30 Jahren von Wenzel et al. (1990), Reilly und McGrath (1995) oder Horbach et al. (1999). Heutzutage definieren diese Verfahren den Stand der Technik und werden auf dem Gebiet der virtuellen Akustik von praktisch allen relevanten Forschungsinstituten eingesetzt sowie erhalten langsam Einzug in die Konsumelektronik (z.B. *Facebook Spatial Workstation*, *Waves Nx*).

Durch Faltung der HRIRs/BRIRs mit einem nachhallfreien Testsignal, kann virtuell nahezu der Raumeindruck erzeugt werden, den die Wiedergabe des Signals im realen Raum (über den bei der Messung verwendeten Lautsprecher) hervorrufen würde. Ein Positionssensor (Headtracker) erfasst dabei in Echtzeit die Orientierung des Kopfes und gibt diese Informationen an den Binauralrenderer weiter. Dadurch kann entsprechend der Positionierung des Probanden, das jeweilig zugehörige binaurale Signalpaar für die Faltung ausgewählt werden (siehe Abbildung 2.11). Eine ausreichende Auflösung muss gewährleistet werden, sodass für den Nutzer keine Abtastung wahrnehmbar ist. Sind nur wenige Messdaten verfügbar, so muss gegebenenfalls zwischen diesen interpoliert werden, um die Auflösung zu erhöhen. Die Faltung muss weiterhin in Echtzeit und mit geringer Latenz durchgeführt werden, damit das wiedergegebene Signal immer der Orientierung des Nutzers entspricht. Durch die Verwendung der segmentierten Faltung

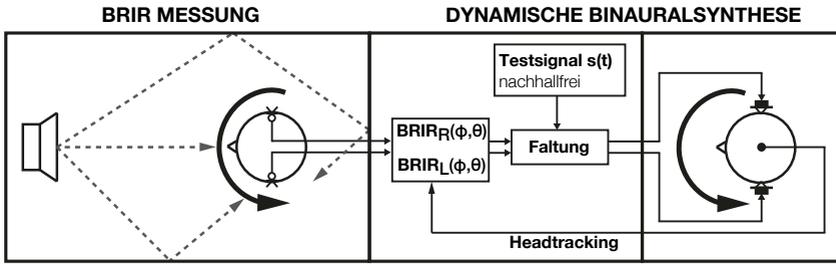


Abbildung 2.11: Schematische Darstellung der datenbasierten, dynamischen Binauralsynthese: sequenzielle BRIR-Messung (links) und Faltung des gemessenen BRIR-Paares mit nachhallfreiem Testsignal mit dynamischer Nachführung per Headtracking entsprechend der Kopforientierung (rechts).

wird dies ermöglicht. Werden diese Faktoren in der Implementierung des Renderers und den Messdaten berücksichtigt, so kann dann ein Höreindruck erzeugt werden, der weitgehend der Wirklichkeit entspricht: Durch Nachführen der binauralen Impulsantworten⁴ bleiben die virtuellen Schallquellen bei Drehungen des Probanden ortsfest stehen und sind für den Nutzer lokalisierbar (entsprechend einer realen Wiedergabe im Raum). Kommen BRIRs zum Einsatz, so wird zusätzlich auch der Raumeindruck (Nachhall, Reflexionen) der akustischen Umgebung auralisiert. Bei konventioneller Kopfhörerwiedergabe hingegen hat die Orientierung keinen Einfluss auf das Hörereignis, das bedeutet Schallquellen drehen sich hier entsprechend der Bewegungen des Hörers mit.

Die Verwendung binauraler Reize und deren dynamische Anpassung zur Auralisation ermöglichen eine *Externalisierung* der dargebotenen Schallsignale. So können in der VAE Hörereignisse erzeugt werden, die außerhalb des Kopfes liegen und klar einer virtuellen Schallquelle zugeordnet werden können. Die sogenannte Im-Kopf-Lokalisation kann dabei wirksam unterdrückt werden. Weiterhin ermöglicht die Verwendung von Bewegungssensoren zur Anpassung des Schallfeldes nach Mackensen et al. (1998) die Verhinderung der sogenannten Vorne-Hinten-Vertauschung, einem bekannten Problem bei der klassischen, statischen Wiedergabe von Kunstkopf-Aufnahmen.

Lindau und Weinzierl (2012) zeigen, dass mittels datenbasierter, dynamischer Binauralsynthese plausible virtuelle Umgebungen geschaffen werden können. Werden zudem individuell gemessene BRIRs verwendet, so ist laut Brinkmann et al. (2017b) unter verschiedenen Randbedingungen sogar eine authentische Wiedergabe möglich. In dieser Studie konnte dies durch eine Individualisierung der dynamischen Binauralsynthese

⁴Dynamische Auswahl oder Interpolation der binauralen Impulsantworten, die der aktuellen Kopfposition am besten entspricht.

gezeigt werden. Weiterhin wurde jedoch eine Abhängigkeit vom verwendeten Testsignal und dessen spektraler Zusammensetzung beobachtet. Die individuelle DBS basiert, wie bereits erörtert, auf einer diffizilen und selten praktikablen Datenerhebung mittels spezifischer Messung in den Gehörgängen aller Probanden. Daher könnten für eine breitere Anwendbarkeit, alternative Ansätze zur Individualisierung vonnöten sein.

Systeme zur dynamischen Binauralsynthese erfordern wenige Komponenten, die mittlerweile alle am Markt erhältlich sind. Die Faltungsoperation ist auf konventionellen Computern und heute auch Smartphones möglich, Bewegungssensoren finden sich ebenfalls in einer Vielzahl von Geräten oder lassen sich mit entsprechenden Inertialsensoren leicht realisieren.

2.5 Wellenfeldanalyse zur Schallfeldzerlegung

Konventionelle, akustische Messverfahren, die meist auf omnidirektionalen Mikrofonen oder Mikrofon-Anordnungen niedriger Ordnung basieren, haben eine eingeschränkte räumliche Auflösung und können nur den Verlauf des Schalldruckes über der Zeit oder der Frequenz darstellen. So können zwar Raumimpulsantworten gebildet werden, sodass sich deren Spektrum als auch eine Vielzahl raumakustischer Kriterien berechnen lassen (Beranek, 2004). Eine richtungssensitive Auswertung des Schallfeldes ist jedoch nicht möglich und es kann weiterhin keine genaue Aussage über dessen räumliche Verteilung getroffen werden. Die *Wellenfeldanalyse* (WFA) bezeichnet Verfahren zur Zerlegung von Schallfeldern im Raum und setzt sich zum Ziel, das Schallfeld räumlich beschreibbar zu machen, um beispielsweise Reflexionen im Raum richtungssensitiv detektieren zu können. So kann ein Echogramm der akustischen Gegebenheit mit räumlichen Informationen gebildet und zur Synthese von diskreten Schallfeldanteilen verwendet werden. Bereits in Berkhout et al. (1997) findet sich auf Basis eines planaren Mikrofonarrays die Formulierung der Wellenfeldanalyse. In der vorliegenden Arbeit hingegen werden ausschließlich sphärische (kugelförmige) Mikrofonarrays betrachtet. Dazu kann beispielsweise die Fourier-Akustik (Williams, 1999) eingesetzt werden, die eine Vielzahl von Signaltransformationen bereitstellt.

Die WFA kann dabei mathematische als das Gegenstück zur mittlerweile recht populären *Wellenfeldsynthese* (WFS), einem lautsprecherbasiertem Wiedergabeverfahren für räumliche Audioinhalte (Spors und Rabenstein, 2008) aufgefasst werden. Die WFS macht sich das Huygens-fresnelsche Prinzip (Huygens, 1690) zu Nutze, um mittels einer Vielzahl von Lautsprechern durch Überlagerung einzelner Elementarwellen, beliebige Wellenfronten zu erzeugen. So kann ein physikalisch korrektes Schallfeld innerhalb einer großen Hörfläche gebildet werden. Dies schafft (bei exakter Konfiguration des Systems)

im Gegensatz zu konventionellen, kanalbasierten Wiedergabeverfahren eine Vergrößerung des *sweet spots*, dem optimalen Hörort.

WFA als auch WFS basieren auf dem *Kirchhoff-Helmholz-Integral* (KHI), welches den Schalldruck innerhalb eines quellfreien Körpers (Volumen) mit dem Schalldruck und der Schallschnelle auf dessen Oberfläche verknüpft. Es besagt, dass bei Kenntnis des Schalldrucks und der Schallschnelle auf allen Punkten der Oberfläche, das Schallfeld ebenso an jedem Punkt innerhalb des Volumens bestimmt werden kann. Dabei kann die Definition von innerhalb und außerhalb nach Williams (1999), S.123 getauscht werden. Williams spricht von einem „interior problem“, wenn sich alle Quellen außerhalb des Volumens befinden, sodass das Schallfeld innerhalb des Volumens bestimmt werden kann. Liegen die Quellen jedoch alle innerhalb des Volumens, so lässt sich durch Lösen des „exterior problem“ das Schallfeld außerhalb des Volumens bestimmen. Dies bedeutet, dass bei Erfassung des Schallfeldes auf einer definierten Oberfläche, ebenso das Schallfeld außerhalb dieser Oberfläche mathematisch berechnet werden kann. Die Erfassung des Schallfeldes in Räumen kann dabei auf einer imaginären oder auch existenten Oberfläche erfolgen. Hierzu werden meist sphärische Mikrofonarrays verwendet, die eine Kugeloberfläche gleichzeitig oder sequentiell abtasten und für jeden Abtastpunkt eine Raumimpulsantwort bestimmen (Meyer und Elko, 2002). Eine ausführliche Betrachtung dieser Thematik und genaue Erläuterung aller notwendigen Verarbeitungsschritte für die Schallfeldzerlegung mittels Wellenfeldanalyse findet sich in Bernschütz (2016). Die nachfolgende für diese Arbeit wichtige Zusammenfassung der relevanten Transformationen basiert auf dieser Ausführung.

2.5.1 Fourier-Transformationen

Die Zeit-Frequenz Fourier-Transformation \mathcal{F}_t eines Signals $p(t)$ in Abhängigkeit der Zeit t ist definiert mit

$$P(\omega) = \mathcal{F}_t\{p(t)\} = \int_{-\infty}^{\infty} p(t)e^{-i\omega t} dt. \quad (2.10)$$

Dabei bezeichnet $P(\omega)$ das Signal in Abhängigkeit der Kreisfrequenz ω , mit $\omega = 2\pi f$. Eine Rücktransformation vom Frequenz- in den Zeitbereich ist durch die inverse Fourier-Transformation \mathcal{F}_t^{-1} mit

$$p(t) = \mathcal{F}_t^{-1}\{P(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\omega)e^{i\omega t} d\omega \quad (2.11)$$

definiert.

Des Weiteren kann die Fourier-Transformation zur Überführung in die räumliche Domäne verwendet werden. Diese als *räumliche* (oder *sphärische*) Fourier-Transformation bezeichnete Operation ist essentiell für die Wellenfeldanalyse. Bei der Definition wird zwischen den verwendeten Koordinatensystemen unterschieden sowie es lassen sich in der Literatur verschiedene Konventionen für Raum \mathbf{x} und Zeit t finden, siehe Tourbabin und Rafaely (2015). Die räumliche Fourier-Transformation $\tilde{\mathcal{F}}_s$ in einem kartesischen Koordinatensystem ist laut Williams (1999), S.2 bei eindimensionaler Betrachtung der x -Komponenten des Ortsvektors x definiert mit

$$\tilde{P}(k_x) = \tilde{\mathcal{F}}_x\{P(x)\} = \int_{-\infty}^{\infty} P(x)e^{ik_x x} dx. \quad (2.12)$$

Die entsprechende Rücktransformation lautet:

$$P(x) = \tilde{\mathcal{F}}_x^{-1}\{P(k_x)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{P}(k_x)e^{-ik_x x} dk_x. \quad (2.13)$$

Dabei bezeichnet $\tilde{P}(k_x)$ das Signal in der räumlichen Domäne des Wellenspektrums und k_x die Wellenzahl entlang der x -Achse des kartesischen Koordinatensystems. Die Wellenzahl ist definiert mit

$$k = \frac{\omega}{c} \quad (2.14)$$

und

$$k = \sqrt{k_x^2 + k_y^2 + k_z^2}. \quad (2.15)$$

Erweitert man die Betrachtung um die Komponenten in y - und z -Richtung so erhält man

$$\tilde{P}(\mathbf{k}_c) = \tilde{\mathcal{F}}_x\{P(\mathbf{x}_c)\} = \iiint_{-\infty}^{\infty} P(\mathbf{x}_c)e^{i\mathbf{k}_c^T \mathbf{x}_c} dx dy dz, \quad (2.16)$$

mit der inversen Transformation

$$P(\mathbf{x}_c) = \tilde{\mathcal{F}}_x^{-1}\{P(\mathbf{k}_c)\} = \frac{1}{(2\pi)^3} \iiint_{-\infty}^{\infty} \tilde{P}(\mathbf{k}_c)e^{-i\mathbf{k}_c^T \mathbf{x}_c} dx dy dz. \quad (2.17)$$

Die Zeit-Frequenz Fourier-Transformation kann zudem mit der räumlichen Fourier-Transformation zusammengeführt werden, sodass man

$$\tilde{P}(\mathbf{k}_c, \omega) = \tilde{\mathcal{F}}_{xt}\{p(\mathbf{x}_c, t)\} = \iiint_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\mathbf{x}_c, t)e^{i(\mathbf{k}_c^T \mathbf{x}_c - \omega t)} dx dy dz dt, \quad (2.18)$$

und die entsprechende inverse Transformation

$$p(\mathbf{x}_c, t) = \tilde{\mathcal{F}}_{\mathbf{x}t}^{-1}\{P(\mathbf{k}_c, \omega)\} = \frac{1}{(2\pi)^4} \iiint \int_{-\infty}^{\infty} \tilde{P}(\mathbf{k}_c, \omega) e^{-i(\mathbf{k}_c^T \mathbf{x}_c - \omega t)} \, dx \, dy \, dz \, dt \quad (2.19)$$

erhält (Bernschütz, 2016, S. 33-34).

Für die Anwendung der Wellenfeldanalyse mittels sphärischen Mikrofonarrays wird jedoch auf die räumliche Fourier-Transformation $\mathring{\mathcal{F}}_s$ für ein sphärisches Koordinatensystem zurückgegriffen sowie auf die entsprechende Rücktransformation $\mathring{\mathcal{F}}_s^{-1}$. Dabei wird das Signal auf der Kugeloberfläche des Arrays in das sphärische Wellenspektrum transformiert. Daher werden im Gegensatz zum kartesischen Koordinatensystem anstelle der Exponentialfunktionen, die sogenannten *Kugelflächenfunktionen* Y_n^m (sphärisch Harmonische - SH) als Basisfunktionen der räumlichen Fourier-Transformation verwendet. Die Kugelflächenfunktionen sind für den Polar- und den Azimuth-Winkel θ und ϕ laut Williams (1999), S. 186 definiert mit

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m!)}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi}. \quad (2.20)$$

So werden die Kugelflächenfunktionen durch komplexe Exponentialfunktionen und die sogenannten *assozierten Legendre-Funktionen* P_n^m mit der Ordnung n und dem Modus m gebildet.

Unter Verwendung der Kugelflächenfunktionen ist die räumliche Fourier-Transformation in sphärischen Koordinaten für einen Radius r_0 definiert mit

$$\mathring{P}_{nm}(r_0, \omega) = \int_0^{2\pi} \int_0^\pi P(\theta, \phi, r_0, \omega) Y_n^m(\theta, \phi)^* \sin\theta \, d\theta \, d\phi. \quad (2.21)$$

Mit $(\cdot)^*$ wird die komplex Konjugierte bezeichnet und $\mathring{P}_{nm}(r_0, \omega)$ ist das Signal in der räumlichen Domäne des sphärischen Wellenspektrums. So kann bei Abtastung mit einem sphärischen Mikrofonarray durch eine Zeit-Frequenz Fourier-Transformation und anschließender räumlicher Fourier-Transformation der entsprechenden Signale das Schallfeld im Raum mittels räumlicher Fourier-Koeffizienten P_{nm} dargestellt werden.

2.5.2 Zerlegung in ebene Wellen

Auf Basis der Koeffizienten P_{nm} ist es möglich, das Schallfeld in *ebene Wellen* zu zerlegen (PWD), um so die räumliche Struktur des Schallfeldes auswerten zu können (Rafaely, 2004; Duraiswami et al., 2005a). Die PWD ermöglicht in der Theorie die

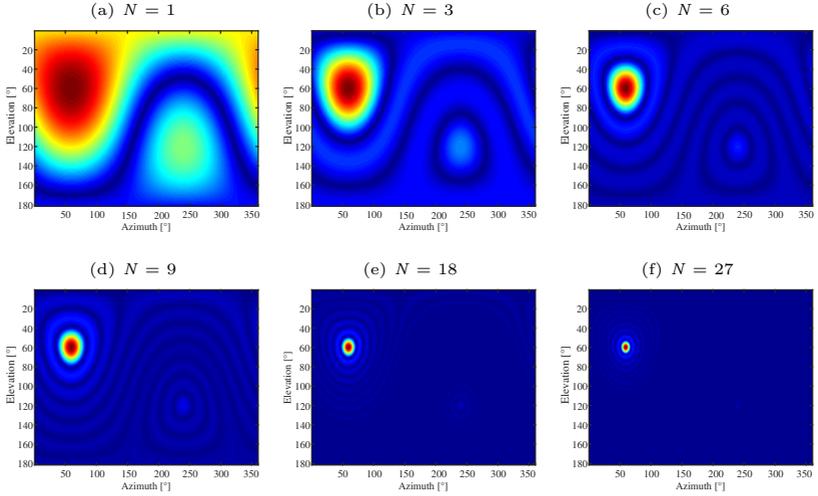


Abbildung 2.12: Simulation der räumlichen Antwort einer ebenen Welle aus 60° Azimuth / Elevation für verschiedene Ordnungen N .

ideale räumliche Abtastung des Schallfeldes, indem für alle Orientierungen θ_d, ϕ_d die Schallfeld-Komponenten D berechnet werden. Die Summierung aller Kugelflächenfunktionen führt dabei zu einem räumlichen Dirac-Impuls mit unbegrenzter Richtwirkung. Die PWD ist definiert als

$$D(\theta_d, \phi_d, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{1}{i^n j_n(\frac{\omega}{c} r_0)} \mathring{P}_{nm}(r_0, \omega) Y_n^m(\theta_d, \phi_d). \quad (2.22)$$

Mit j_n wird dabei eine sphärische Bessel-Funktion erster Art bezeichnet. Bei kleiner Wellenzahl müssen die Modenbeiträge jedoch teilweise erheblich verstärkt werden, um diese im Ausgangssignal sinnvoll nutzen zu können. Es werden modale Radialfilter d_n zur Kompensation eingesetzt, die von der Konfiguration des Mikrofonarrays (open/rigid) und des verwendeten Radius abhängen (Rafaely, 2005). Dazu wird für Formel 2.22 eine leicht abweichende Schreibweise verwendet:

$$D(\theta_d, \phi_d, \omega) = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^n d_n(\frac{\omega}{c} r_0) \mathring{P}_{nm}(r_0, \omega) Y_n^m(\theta_d, \phi_d). \quad (2.23)$$

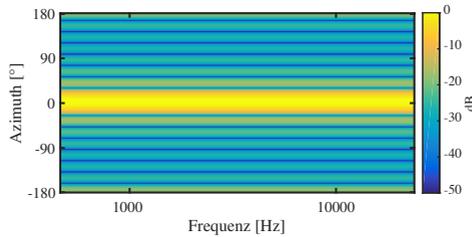


Abbildung 2.13: Antwort eines sphärischen Mikrofonarrays in der Azimuth-Ebene für eine aus $\phi = 0^\circ$ und $\theta = 0^\circ$ eintreffende breitbandige ebene Welle bei $N = 7$.

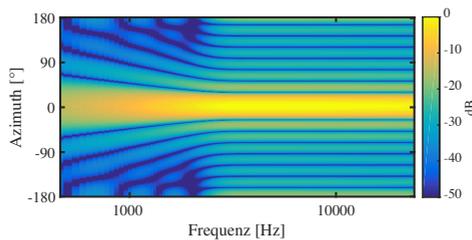


Abbildung 2.14: Antwort eines Mikrofonarrays mit 1202 Abtastpunkten und einem Radius von 8,75 cm in der Azimuth-Ebene für eine aus $\phi = 0^\circ$ und $\theta = 0^\circ$ eintreffende breitbandige ebene Welle bei $N = 7$ und modaler Verstärkung von +18 dB.

In Bernschütz et al. (2011b) werden die Radialfilter zudem als FIR-Filter implementiert und deren Verstärkung mit einem Soft-Knee-Verhalten limitiert. Entsprechende Radialfilter finden in dieser Arbeit Anwendung.

Die Summierung aller Ordnungen der sphärisch Harmonischen ist jedoch nicht möglich, da dazu das Schallfeld ideal auf einer Kugeloberfläche abgetastet werden müsste. Eine ideale Abtastung ist praktisch aber nicht realisierbar und kann nur an diskreten Abtastpunkten erfolgen. Die hat eine Reduzierung der räumlichen Auflösung und Begrenzung der Bandbreite zu hohen Frequenzen zur Folge, da die Ordnung der SH begrenzt wird. Dadurch ergibt sich inhärent eine Begrenzung der möglichen Richtwir-

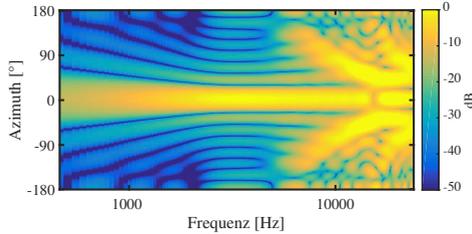


Abbildung 2.15: Antwort eines Mikrofonarrays mit 86 Abtastpunkten und einem Radius von 8,75 cm in der Azimuth-Ebene für eine aus $\phi = 0^\circ$ und $\theta = 0^\circ$ eintreffende breitbandige ebene Welle bei $N = 7$ und modaler Verstärkung von +18 dB.

kung der PWD. Die maximale, mögliche Ordnung N des Mikrofonarrays hängt dabei mit

$$N = \sqrt{\frac{n_{\text{sp}}}{\eta_g}} - 1 \quad (2.24)$$

von der Anzahl der Abtastpunkte n_{sp} und dem Faktor η_g , der die Effizienz der Abtastung der verwendeten Quadratur beziffert. Dabei bezeichnet $\eta_g = 1$ die höchste Effizienz und $\eta_g > 1$ eine ansteigende Ineffizienz. Eine *Gauß-Quadratur* (Stroud und Secrest, 1966) tastet die Kugeloberfläche gleichwinklig ab und erreicht $\eta_g = 2$. Effizienter ist die Verwendung einer *Lebedev-Quadratur* (Lebedev, 1977), da diese die Kugeloberfläche äquidistant abtastet und entgegen der Gauß-Abtastung keine Häufung der Punkte an den Polen hervorruft. In dieser Arbeit finden daher ausschließlich Lebedev-Quadraturen mit $\eta_g = 1, 3$ Anwendung. Bei Kenntnis von N wird die Frequenz f_A , ab der *räumliches Aliasing* (Rafaely et al., 2007) auftritt mit

$$f_A = \frac{Nc}{2\pi r_0} \quad (2.25)$$

in Abhängig von der Schallgeschwindigkeit c und dem Radius r_0 des Arrays berechnet. Es existieren zwar Möglichkeiten der Kompensation des räumlichen Aliasing (z.B. von Bernschütz (2012)), diese werden in dieser Arbeit aber nicht betrachtet.

Die PWD kann anschließend zur Visualisierung des Schallfeldes im Raum verwendet werden. Dazu wird ein sphärisches Raster gewählt, welches die Orientierungen definiert (nach Azimuth ϕ_d und Elevation θ_d), in deren Richtung die ebenen Wellen gebildet werden. Für das nachfolgende Beispiel wurde ein Gauss-Raster mit 360×181 (1° -Auflösung) Abtastpunkten und daraus resultierenden 65.160 ebenen Wellen verwen-

det. Abbildung 2.12 zeigt eine Simulation der räumlichen Antwort aus Richtung 60° Azimuth / Elevation für verschiedene Ordnungen N . Dabei wird deutlich, wie die Richtwirkung der PWD mit der Erhöhung der Ordnung ansteigt. Durch die PWD ist eine richtungs-sensitive Auswertung des Schallfeldes möglich. Wird die Schallfelderlegung mit einer zeitlichen Fensterung kombiniert, so kann die Auswertung auch zeit-sensitiv erfolgen, sodass Reflexionen im Raum mit Pegel, Orientierung und Zeitpunkt detektiert werden können (siehe Kapitel 4.1.1).

Die Bandbreite der Wellenfeldanalyse mit Mikrofonarrays ist durch die notwendige Kompensation der Radial-Filter auch zu den tiefen Frequenzen begrenzt. Hierbei handelt es sich um eine Verstärkung der Modenbeiträge. Bei der idealen Betrachtung ist eine unbegrenzte Verstärkung möglich, sodass stabile Impulsantworten erzeugt werden können. In gemessenen Impulsantworten ist die maximale Verstärkung jedoch durch den Signal-Rauschabstand begrenzt, sodass eine ideale Kompensation nicht durchgeführt werden kann (Bernschütz et al., 2011b).

Abbildung 2.13 zeigt die Antwort eines idealen Mikrofonarrays in der Azimuth-Ebene für eine aus $\phi = 0^\circ$ und $\theta = 0^\circ$ eintreffende breitbandige ebene Welle bei $N = 7$. Es zeigt sich eine stabile Auflösung über das gesamte Spektrum. Bei Verwendung eines realen Arrays mit 1202 Punkten auf einer Lebedev-Quadratur und einem Radius $r_0 = 8,75$ cm zeigt sich im Vergleich zum idealen Array bei einer modalen Verstärkung von +18 dB eine Reduzierung der tieffrequenten Auflösung. Aufgrund des hoch-aufgelösten Abtastgitters tritt in dieser Konfiguration im hörbaren Bereich kein räumliches Aliasing auf. Auswirkungen von räumlichem Aliasing auf die Richtwirkung, verursacht durch eine niedrigere Abtastung der Kugeloberfläche zeigt hingegen Abbildung 2.15 für ein Mikrofonarray mit 86 Abtastpunkten (Lebedev) auf einem Radius $r_0 = 8,75$ cm. In dieser Konfiguration arbeitet das Array nur noch bis zu einer Frequenz $f_\Lambda = 4450$ Hz stabil, oberhalb von f_Λ zeigen sich deutliche Artefakte durch räumliches Aliasing.

KAPITEL 3

Methoden, Technologien und Ressourcen

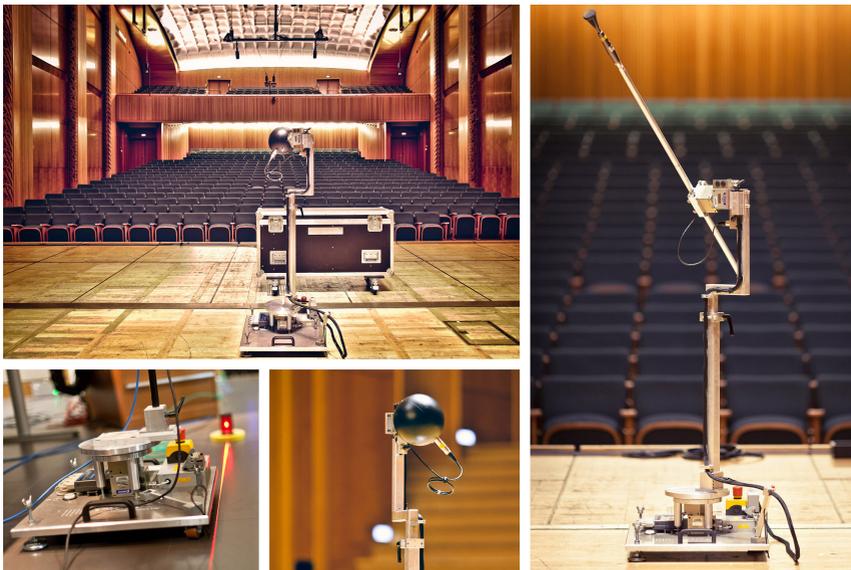


Abbildung 3.1: *VariSpear* Mikrofonarray in verschiedenen Konfigurationen: geschlossene (*rigid sphere* - links) und offene (*open sphere* - rechts) Anordnung.

Dieses Kapitel befasst sich mit allen relevanten Methoden, technischen Verfahren und Ressourcen, die in dieser Arbeit Anwendung finden und die die Grundlage für die weiteren Untersuchungen bilden. Dies betrifft im Einzelnen alle technischen Messsysteme zur akustischen Erfassung und Analyse von Räumen (siehe Kapitel 3.1) sowie die Räumlichkeiten (siehe 3.2), die synthetisiert, auralisiert und dann evaluiert werden. Die zur BRIR-Synthese verwendeten HRIR-Daten und deren Messung werden ebenfalls vorgestellt (siehe 3.3). Es wird auf maßgeblich zurückgegriffene Software für die später zu erläuternde Signalverarbeitung (siehe 3.4) eingegangen. Weiterhin werden die technische Umgebung der perzeptiven Untersuchungen (siehe 3.5) mittels dynamischer Binauralynthese als auch die verwendete Versuchsanordnung und der Ablauf und die Fragestellung der Experimente genauer erläutert (siehe 3.6). Abschließend werden Verfahren aus der Statistik beschrieben, die in der Auswertung und Interpretation der durchgeführten Hörversuche Anwendung finden (siehe 3.7). Damit erörtert dieses Kapitel alle für die Entwicklung als auch die Analyse der erarbeiteten und untersuchten parametrischen Modelle notwendigen Werkzeuge, die zur synthetischen BRIR-Erzeugung und zur Erfassung entsprechender Referenz-Datensätze eingesetzt werden.

3.1 Messsysteme

Verschiedene akustische Messsysteme werden für die Entwicklung und Validierung der vorgestellten parametrischen Modelle verwendet. Darüber hinaus werden zum Vergleich Referenz-Datensätze zur perzeptiven Beurteilung der entwickelten Verfahren mit diesen Messsystemen gewonnen. Es handelt es sich um Systeme zur Aufzeichnung (*Empfänger*) als auch zur Anregung (*Quellen*), die gemeinsam in unterschiedlicher Kombination verwendet wurden. In allen Räumlichkeiten werden mit diesen Messsystemen zahlreiche Raumimpulsantworten und binaurale Raumimpulsantworten (siehe Kapitel 2.2) mit Sinussweeps mit einem +20 dB Kuhschwanzfilter bei 100 Hz (*colored sine sweep*) und einer Länge von 2^{19} Samples bei einer Abtastrate von $f_s = 48$ kHz gewonnen. Weiterhin wird ein *RME Fireface UCX*-Audiointerface, inklusive der integrierten Mikrofonvorverstärker und AD/DA Wandler, eingesetzt. Alle Messsysteme verwenden in den jeweiligen Räumen die identischen Positionen und Drehpunkte (*pivot point*) und können daher direkt aufeinander bezogen und gemeinsam in den Modellen verwendet werden. Die Messdaten wurden von Stade et al. (2012) vorgestellt und sind unter einer CC BY-SA 3.0 Lizenz frei verfügbar¹.

¹http://audiogroup.web.th-koeln.de/wdr_irc.html

3.1.1 Sequenzielles Mikrofonarray

Ein sequenzielles Mikrofonarray ermöglicht die Messung von Raumimpulsantworten auf der Oberfläche einer gewünschten geometrischen Form in nahezu beliebiger Auflösung. Dabei sind Konfigurationen mit einer sehr hohen Anzahl von Abtastpunkten möglich, wodurch sich technische Limitationen wie z.B. räumliches Aliasing weitestgehend vermeiden lassen. Da die Erfassung der Impulsantworten zwar sequenziell erfolgt, die Auswertung den Datensatz aber geschlossen betrachtet, werden eine hohe Präzision sowie nahezu gleichbleibende klimatische Bedingungen über die Messdauer gefordert. Unter Anwendung der räumlichen Fourier-Zerlegung (siehe Abschnitt 2.5) können Mikrofonarray-Daten zur Schallfeldzerlegung (Williams, 1999), zur Analyse (Farina et al., 2011; Bernschütz et al., 2012) als auch zur Auralisation (Duraiswami et al., 2005b; Bernschütz et al., 2014) eingesetzt werden. Sphärische Mikrofonarrays haben sich hierzu in der Literatur etabliert (Meyer und Elko, 2002; Rafaely, 2005; Duraiswami et al., 2005a).

Das an der Technischen Hochschule Köln entwickelte und gefertigte sphärische Mikrofonarray *VariSphear*² (Bernschütz et al., 2010) wird zur sequenziellen Messung von Impulsantworten auf einer Kugeloberfläche verwendet (Aufbau mit unterschiedlichen Konfigurationen siehe Abbildung 3.1). Zwei *Schunk/Amtec Robotics PR70* Motoren mit einer sehr hohen Wiederholgenauigkeit (Winkelfehler $< 0,01^\circ$) erlauben die Positionierung eines Mikrofons in zwei Freiheitsgraden (Azimuth ϕ und Elevation θ) auf unterschiedlichen Radien und Höhen. Ein Mikrofon mit Kugel-Richtcharakteristik wird dazu in einer Holzkugel (rigid sphere) montiert (Mikrofonkapsel plan auf Kugeloberfläche, siehe Abbildung 3.1 links), sodass die Kapsel mit Hilfe der zwei Motoren an beliebige Positionen auf dieser Oberfläche bewegt werden kann. Zudem ist auch eine Konfiguration ohne physischen Körper (open sphere) und Verwendung eines Mikrofons mit Nieren-Richtcharakteristik möglich (siehe Abbildung 3.1 rechts). Das Mikrofonarray wird über Netzwerk angesteuert, die integrierte Audiosignalverarbeitung zur Messung der Impulsantworten basiert auf einer Anregung mit Sinussweep und Entfaltung (Müller, 1999).

In der vorliegenden Arbeit finden Abtastgitter mit 1202 und 86 Samplingpunkten und einer Lebedev Quadratur (Lebedev, 1977) in der geschlossenen Konfiguration mit physischem Körper Anwendung. Ein *Earthworks M30* Messmikrofon mit Kugel-Richtcharakteristik wird dazu auf einem Radius von 8,75 cm an die verschiedenen Abtastpunkte der Kugeloberfläche bewegt. Der Theorie nach arbeitet das Mikrofonarray mit diesen Konfigurationen stabil bis $f_\Lambda = 18$ kHz (1202 Punkte) bzw. $f_\Lambda = 4450$ Hz (86 Punkte) und ermöglicht maximale modale Ordnungen von $N = 29$ bzw. $N = 7$.

²VariSphear - variable spherial ear: <http://audiogroup.web.th-koeln.de/varisphear.html>

3.1.2 Sequenzielles Binauralsystem

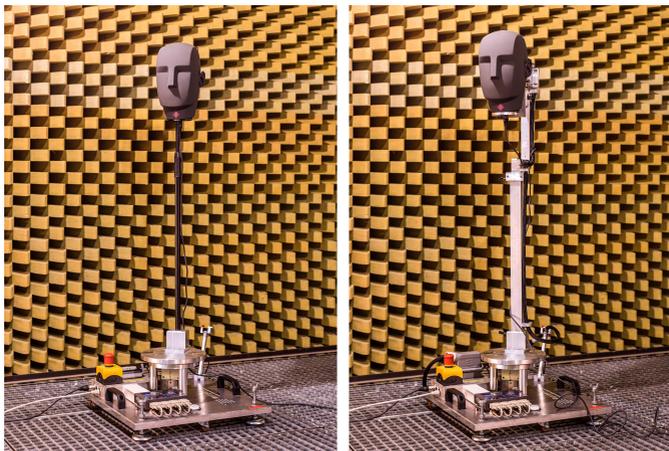


Abbildung 3.2: *Neumann KU100* Kunstkopf auf *VariSpear* Mikrofonarray für zylindrische (links) und sphärische (rechts) Abtastgitter.

Für die datenbasierte dynamische Binauralsynthese mit Nachführung der Kopfbewegungen des Rezipienten werden eine Vielzahl von BRIRs für verschiedene Orientierungswinkel benötigt (siehe Kapitel 2.4). Zur Messung dieser BRIRs kann auf Funktionalitäten des *VariSpear* Mikrofonarrays zurückgegriffen werden. Mit einer speziellen Halterung wird ein Kunstkopf anstelle des Mikrofons im Drehmittelpunkt des Arrays befestigt, sodass sphärische BRIR- (oder auch HRIR-) Messungen möglich werden (siehe Abbildung 3.2 rechts). Da in dieser Arbeit nur die Anpassung an Änderungen der Orientierung auf der horizontalen Ebene erfolgt, werden die BRIRs gleichsam nur auf der horizontalen Ebene erfasst. Dazu wird eine alternative Halterung verwendet, die weniger Reflexionsflächen besitzt. Ausschließlich der Motor für die Azimuthal-Ebene wird verwendet, auf der Grundplatte des Systems wird ein Kunstkopf auf einem konventionellen Mikrofonstativ montiert. So kann der Kunstkopf an beliebige Winkelpositionen auf der horizontalen Ebene bewegt werden, bei gleichbleibendem Drehmittelpunkt im Vergleich zu den Mikrofonarray-Messungen (siehe Abbildung 3.2 links). Es gilt zu beachten, dass die identische Höhe der Mikrofonkapseln verwendet werden muss.

Für die Messungen der Referenz-BRIRs wird ein *Neumann KU100* Kunstkopf eingesetzt, der in 1° -Schritten auf der horizontalen Achse rotiert, sodass 360 BRIRs pro Raum und Schallquelle gewonnen werden (Stade et al., 2012).

3.1.3 Schallquellen



Abbildung 3.3: Verwendete Schallquellen mit unterschiedlicher Richtcharakteristik zur Messung von Raumimpulsantworten, gerichtet (links) und omnidirektional (rechts).

Bei der Messung von RIRs wird eine Schallquelle zur Wiedergabe des Messsignals (z.B. Sinussweep) benötigt. Die verwendete Schallquelle hat insbesondere durch ihre frequenzabhängige Abstrahlcharakteristik einen erheblichen Einfluss auf die Anregung des Raumes und somit auf die gemessene und später auralisierte Raumimpulsantwort. Insbesondere das Verhältnis zwischen Direktschall und Nachhall (DRR) wird durch die Anregungsquelle beeinflusst. Zudem wird für eine adäquate Auralisation entsprechend der originalen Richtcharakteristik des Testsignals, eine möglichst ähnliche Quellcharakteristik bei der Messung gefordert. Um Unterschiede aufgrund von Richtcharakteristiken untersuchen zu können, werden für alle Messungen mit den Empfänger-Systemen sequenziell zwei verschiedene Lautsprecher als Quelle verwendet (Stade et al., 2012). Als gerichtete Schallquelle kommt ein *AD Systems Stium* Mittelhochtonsystem in Verbindung mit 3 *AD Systems Flex 15B* Subwoofern zum Einsatz. Das System wird mittels einer *Camco Vortex* Endstufe in Verbindung mit einem *XTA* Lautsprechercontroller verstärkt und entzerrt. Als omnidirektionale Schallquelle wird der an der Technischen Hochschule Köln entwickelte und gefertigte Mehrwege-Dodekaeder *Sonicball* (Meulemann et al., 2011) mit speziell entwickelter und DSP-gesteuerter Endstufe verwendet. Entgegen üblicher (z.B. in der Bauakustik) verwendeter Dodekaeder, handelt es sich hierbei um ein leistungsfähiges Breitband-System mit nahezu linearem Frequenzgang von 50 - 18000 Hz. Aufgrund der Spezifikationen der verwendeten Schallquellen, geringen Ruhegeräuschpegeln in den Konzertsälen und langer Anregungssignale, können in beiden Räumlichkeiten hochqualitative Raumimpulsantworten mit einem sehr hohen Signal-Rauschabstand (> 90 dB) gewonnen werden.

3.2 Räumlichkeiten



Abbildung 3.4: KVB-Saal (links) und kleiner Sendesaal (rechts), Blick in Richtung Bühne.

Die zuvor beschriebenen Messsysteme werden in den folgenden zwei Räumlichkeiten kombiniert eingesetzt. Bei beiden Räumen handelt es sich um Konzert-Säle, die sich im Funkhaus des Westdeutschen Rundfunks (WDR) Köln befinden. Das Funkhaus existiert seit Anfang der 1950er Jahre am Wallrafplatz in der Kölner Innenstadt und wurde mehrfach umgebaut und modernisiert (NWDR, 1953; WDR, 1993). Die Säle werden von Ensembles verschiedener Größen und Genres für Musikveranstaltungen und deren Aufnahme sowie für Wort-Produktionen genutzt (Stade et al., 2012). Jährlich finden hier etwa 250 Veranstaltungen statt, sodass die Räumlichkeiten zu den am häufigsten frequentierten Veranstaltungssälen Nordrhein-Westfalens gehören. Grundlegende Eigenschaften der Räume finden sich in Tabelle 3.1.

3.2.1 Klaus-von-Bismarck-Saal

Bei dem *Klaus-Von-Bismarck-Saal* (KVB)³, der auch unter dem Namen *Großer Sendesaal* bekannt ist, handelt es sich um einen klassischen Konzertsaal für Orchestermusik. Der Saal bietet bestuhlt Platz für insgesamt 637 Besucher, verteilt auf zwei Ebenen. Die Bühne ist mit ca. 210 m² ausreichend dimensioniert und kann auch große Orchester oder Chöre aufnehmen. Im Raum finden sich überwiegend Materialien aus Holz. Im Grundriss (siehe Abbildung 3.6) zeigt sich ein nahezu symmetrischer Aufbau des Saals. Die Nachhallzeit des Raumes über der Frequenz findet sich in Abbildung 3.5 links, im Mittel beträgt diese 1,46 s, der mittlere Absorptionsgrad α liegt bei 0,24. An der Vorderwand des Saals befindet sich eine *Klais-Konzertorgel*.

³Klaus von Bismarck (*1912 - †1997) war von 1961-1976 Intendant des WDR.

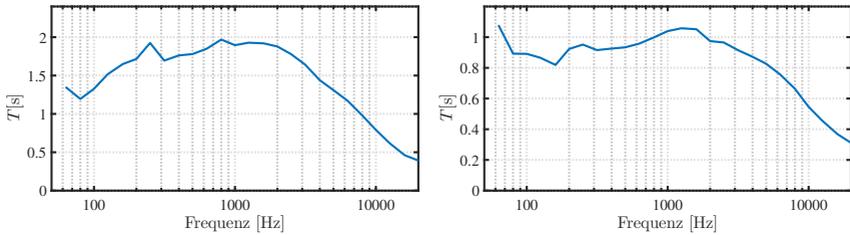


Abbildung 3.5: Frequenzabhängige Nachhallzeit T in s für KVB-Saal (links) und kleiner Sendesaal (rechts).

Tabelle 3.1: Eigenschaften der untersuchten Räume.

Raum	Volumen	Fläche	Höhe	T_{mean}	Sitzplätze
KVB-Saal	6098 m ³	480 m ²	11 m	1,46 s	637
kleiner Sendesaal	1247 m ³	204 m ²	6 m	0,83 s	160

3.2.2 Kleiner Sendesaal

Beim *kleinen Sendesaal* (KSS) handelt es sich verglichen mit dem KVB-Saal um einen erheblich kleineren Raum, in dem meist Instrumental-Ensembles (Kammermusik) auftreten und produziert werden. Neben Wortbeiträgen oder Kabarett-Aufzeichnungen finden hier auch des Öfteren Rock-Pop Konzerte und Live-Übertragungen statt. Während der Messungen ist der Saal mit ca. 160 Stühlen ausgestattet. Die Bestuhlung ist jedoch variabel und lässt sich den jeweiligen Veranstaltungen anpassen oder gar ganz entfernen. Bühne und Bestuhlung sind schräg im Raum liegend angeordnet. Wie auch im KVB-Saal, bestehen die überwiegenden Materialien und Oberflächen aus Holz. Die Nachhallzeit des Raumes über der Frequenz findet sich in Abbildung 3.5 rechts, im Mittel beträgt diese 0,83 s, der mittlere Absorptionsgrad α liegt bei 0,32.

3.3 HRIR-Datensatz

Für die BRIR-Synthese wird auf einen von Bernschütz (2013) an der Technischen Hochschule Köln gemessenen sphärischen HRIR-Datensatz zurückgegriffen. Eingesetzt wurde ein *Neumann KU100* Kunstkopf. Die Messung erfolgt auf einer Lebedev Quadratur mit 2702 Abtastpunkten im reflexionsarmen Raum der Hochschule unter Anwendung des *VariSphear* Systems und der entsprechenden Halterung gemäß Abbildung 3.2

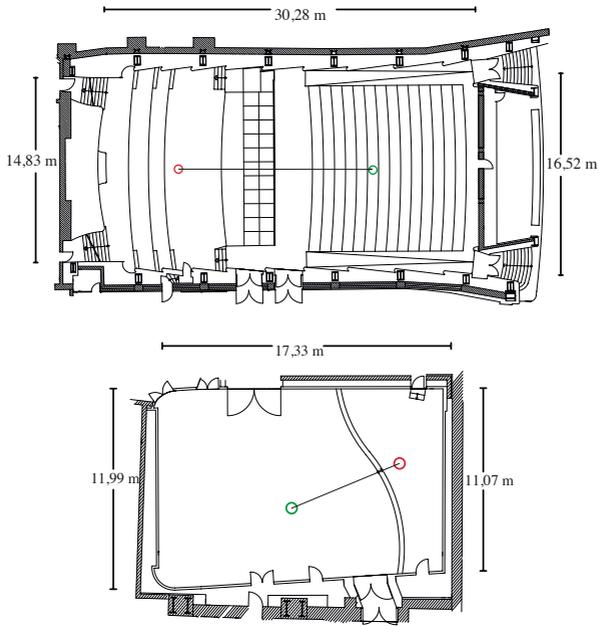


Abbildung 3.6: Grundriss KVB-Saal (oben) und kleiner Sendesaal (unten) mit Quell- (rot) und Empfänger-Position (grün) aus Stade (2013).

rechts. Als Anregungsquelle wird ein aktiver *Genelec 8260A* DSP-gesteuerter 3-Wege Studiolausprecher eingesetzt. Die Impulsantworten werden unterhalb von 200 Hz mit dem ALFE⁴ Verfahren korrigiert (Bernschütz, 2013). Durch eine räumliche Fourier-Transformation wird der Datensatz als SH-Koeffizienten abgespeichert, sodass sich über eine Interpolation HRIRs mit beliebigen Orientierungen erzeugen lassen. Der HRIR-Datensatz ist in verschiedenen Formaten (miro, SOFA, DAFF) unter einer CC BY-SA 3.0 Lizenz frei verfügbar⁵.

3.4 Software

Die Entwicklung und Evaluierung der vorgestellten parametrischen Modelle zur Synthese binauraler Raumimpulsantworten erfolgt im Wesentlichen mit Hilfe der kom-

⁴ALFE - *a*daptive *l*ow *f*requency *e*xtension

⁵<http://audiogroup.web.th-koeln.de/ku100hrir.html>

merziellen Software *MATLAB* (R2015b) des Unternehmens *MathsWorks* und eigens erstellten Funktionen, die in Kapitel 4 erläutert werden. Des Weiteren wird aber auch auf externe Software zurückgegriffen, die im Folgenden vorgestellt wird.

3.4.1 SOFiA-Toolbox

SOFiA⁶ ist eine Toolbox für *MATLAB*, die zur Erweiterung der beinhaltenden Funktionalitäten verwendet wird. Sie bietet Funktionen zur Wellenfeldanalyse und zur Simulation von sphärischen Mikrofonarrays. Einzelne Verarbeitungsschritte sind zur Steigerung der Performance in C++ programmiert und werden über sogenannte *MEX* Dateien in *MATLAB* eingebunden. SOFiA wird in dieser Arbeit für die Weiterverarbeitung der gemessenen Mikrofonarray-Daten verwendet, dies betrifft im Detail die Detektion von Direktschall und Reflexionen (siehe 4.1.1.1) und die Erzeugung von gerichteten Impulsantworten (siehe 4.1.1.2) zur Generierung von Reflexions-Filtern (siehe 4.1.1.3). Auch für Transformationen zwischen Zeit- und Frequenzbereich sowie der räumlichen Domäne wird SOFiA eingesetzt. Die Toolbox ist unter einer MIT-Lizenz frei nutzbar und wurde von Bernschütz et al. (2011a) entwickelt. Mittlerweile ist SOFiA auch unter der frei-verfügbaren Programmiersprache *Python* einsetzbar (Hohnerlein und Ahrens, 2017)⁷. Einen Überblick über alle enthaltenden sowie die in dieser Arbeit maßgeblich verwendeten Funktionalitäten (rot) gibt Abbildung 3.7.

3.4.2 ITA-Toolbox

Die ITA-Toolbox⁸ ist eine ebenfalls frei-verfügbare *MATLAB* Toolbox für eine Vielzahl von Anwendungen in der Akustik. Sie wird entwickelt und vertrieben vom Institut für Technische Akustik (ITA)⁹ der RWTH Aachen University (Dietrich et al., 2013; Berzborn et al., 2017). Die ITA-Toolbox bietet u.a. Routinen zur akustischen Messtechnik, Psychoakustik, DSP-Signalverarbeitung, Fourier-Akustik, Virtuelle Akustik und der Raumakustik.

In dieser Arbeit findet die ITA-Toolbox¹⁰ Anwendung bei der Bestimmung und Auswertung raumakustischer Kriterien nach DIN EN ISO 3382-1:2009-10 (2009) der gemessenen BRIRs als auch der generierten SBRIRs (siehe 2.3). Es wird in der Regel der Frequenzbereich von 63 Hz bis 20 kHz in Oktavbändern ausgewertet, da diese Betrachtung

⁶SOFiA - sound field analysis: http://audiogroup.web.th-koeln.de/SOFiA_wiki/WELCOME.html

⁷https://github.com/QUlab/sound_field_analysis-py

⁸<http://www.ita-toolbox.org>

⁹<http://www.akustik.rwth-aachen.de>

¹⁰Download am 03.02.2018, Datei: toolbox-release-9653e3164adf0e359b1252b649c95d8c057234a8.zip

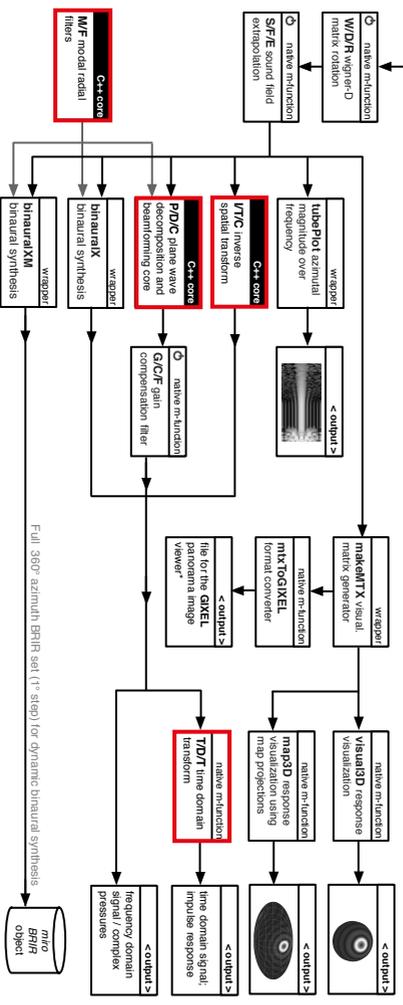
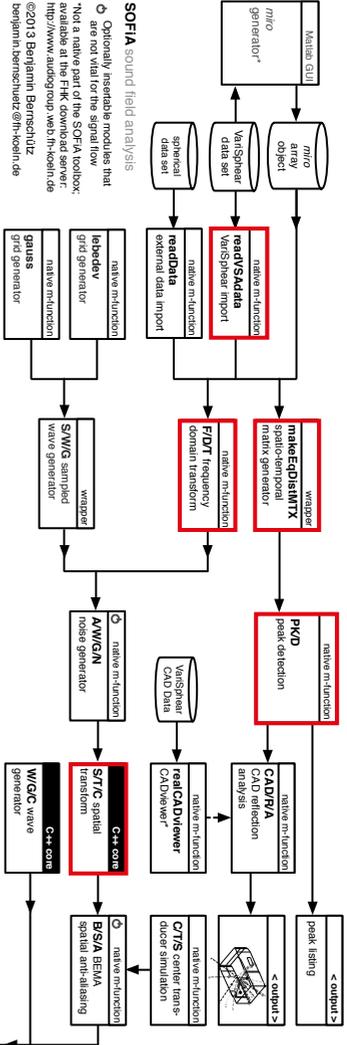


Abbildung 3-7: SOFIA Blockschaltbild von Bernschütz et al. (2011a) mit in dieser Arbeit maßgeblich verwendeten Funktionalitäten (rot).

laut Rossing (2015), S.315 mit der subjektiven Wahrnehmung besser übereinstimmt als eine feinere Auflösung. Teilweise ist jedoch auch eine höhere Frequenzauflösung nötig, um Unterschiede zwischen den Konfigurationen besser herausarbeiten zu können. Betrachtet wird die Nachhallzeit $T30$, die frühe Abklingzeit EDT, der frühe und späte interaurale Kreuzkorrelationskoeffizient $IACC_{\text{early}}$, $IACC_{\text{late}}$, die Schwerpunktzeit T_S , das Klarheitsmaß $C80$, die Deutlichkeit $D50$ und das Bass- und Höhenverhältnis BR und TR .

Die Auswertung des Seitenschallgrades oder ähnlicher Maße benötigt die Verwendung eines Mikrofons mit Achter-Richtcharakteristik. Die Mikrofonarray-Messungen erlauben zwar die Generierung von unterschiedlichen Richtcharakteristika, da aber entsprechende Daten auf Seiten der synthetischen Impulsantworten fehlen, werden diese Kriterien nicht betrachtet. Weiterhin gilt zu beachten, dass viele Maße eine omnidirektionale Richtcharakteristik fordern. Da in dieser Arbeit die raumakustischen Kriterien jedoch ausschließlich zum direkten Vergleich zwischen Referenz-Messung und SBRIR und nicht zur absoluten Bewertung angewendet werden, kann dieser Sachverhalt vernachlässigt werden. Ein direkter Vergleich mit Kriterien anderer Räumlichkeiten ist daher jedoch nur bedingt möglich. Betrachtet wird ein Kanal der Datensätze (linkes Ohr, Blickrichtung in Richtung Schallquelle mit $\phi = 0^\circ$ und $\theta = 0^\circ$), lediglich die binauralen Maße verwenden selbstredend beide Ohrsignale. Diese technische Untersuchung analysiert die gesamte Impulsantwort, da diese auch in den Hörversuchen zur Auralisation benutzt wird. Auf in der Toolbox enthaltene Methoden zur grafischen Darstellung der Ergebnisse wird nicht zurückgegriffen. Maßgeblich werden die Funktionen `ita_roomacoustics.m`, `ita_roomacoustics_tonal_color.m` und `ita_roomacoustics_IACC.m` verwendet.

3.4.3 SoundScape Renderer

Der *SoundScape Renderer* (SSR) ist ein vom *Quality and Usability Lab* der Technischen Universität Berlin¹¹ und dem *Institut für Nachrichtentechnik* der Universität Rostock¹² entwickeltes Werkzeug zur Echtzeit-Wiedergabe von räumlichen Audioverfahren (Geier et al., 2008; Geier und Spors, 2012). Dazu werden verschiedene Algorithmen zur Verfügung gestellt die z.B. Wellenfeldsynthese, Higher Order Ambisonics (HOA) oder binaurales Rendering ermöglichen. In der vorliegenden Arbeit wird ausschließlich der *Binaural Room Synthesis* (BRS) - Renderer auf Basis von BRIRs eingesetzt. In diesem Modul ermöglicht der SoundScape Renderer bei Verwendung eines Kopfhörers mit Orientierungssensor und spezifischen BRIR-Datensätzen die dynamische Binauralsyn-

¹¹<http://qu.tu-berlin.de>

¹²<http://www.int.uni-rostock.de>

these mit Anpassung an die Kopfdrehung des Rezipienten auf der horizontalen Ebene. Unterstützt werden verschiedene Headtracker-Protokolle wie z.B. *Polhemus Fastrak*, *Razor AHRs* oder auch das *Virtual Reality Peripheral Network* (VRPN). Mit Hilfe einer Netzwerkschnittstelle (TCP/IP) ist es weiterhin möglich den SoundScape Renderer anzusteuern. Unter einer GNU General Public License (GPL) ist der SoundScape Renderer für *Linux* und *Mac OS X* frei verfügbar¹³. Im wissenschaftlichen Kontext wird der SoundScape Renderer von einer Vielzahl weiterer Forschungsinstitute eingesetzt (z.B. SEACEN Group¹⁴, BBC R&D¹⁵, Chalmers University of Technology¹⁶). Für diese Arbeit wird die Versionsnummer 0.4.2 unter Mac OS X 10.10.5 verwendet.

3.4.4 Scale

Bei der Software *Scale*¹⁷ handelt es sich um eine an der Technischen Hochschule Köln von Vazquez (2013) unter MATLAB entwickelte Umgebung zur Implementierung, Durchführung und Auswertung von perzeptiven Untersuchungen. Scale bietet eine Vielzahl implementierter Versuchsanordnungen wie z.B. ABC/HR (ITU-R BS.1116-3, 2015), MUSHRA (ITU-R BS 1534-3, 2015) oder auch SAQI (Lindau et al., 2014). Weiterhin ist eine direkte Anbindung an den SoundScape Renderer vorgesehen. Dies ermöglicht die Durchführung von kopfhörer-basierten Hörversuchen unter Anwendung der dynamischen Binauralsynthese (Vazquez, 2015). Scale wird in dieser Arbeit für die Implementierung, Durchführung und Datenerhebung aller Experimente verwendet.

3.5 Technische Rahmenbedingungen der Versuchsumgebung

Ein schematischer Überblick über die technische Umgebung der Hörversuche findet sich in Abbildung 3.8. Alle in dieser Arbeit vorgestellten perzeptiven Untersuchungen werden in einer kopfhörer-basierten virtuellen akustischen Umgebung unter Einsatz des SoundScape Renderers, *AKG K601* Kopfhörern und einem *Polhemus Fastrak* Headtracking-System durchgeführt. Die dynamische Binauralsynthese passt die Auralisation in Echtzeit an Drehungen und Kopfbewegungen des Hörers auf der gesamten horizontalen Ebene in 1° Auflösung an. Dies ermöglicht eine ortsfeste Abbildung von virtuellen Schallquellen ohne eingeschränkten (horizontalen) Winkelbereich. Eine Adaption an vertikale Kopfbewegungen oder auch translatorische Bewegungen der Nutzer

¹³<http://spatialaudio.net/ssr/>

¹⁴<http://www.seacen.tu-berlin.de/seacen/>

¹⁵http://www.bbc.co.uk/rd/work_streams/audio-research

¹⁶<http://www.chalmers.se/en/departments/ace/research/appliedacoustics/Pages/default.aspx>

¹⁷Scale - setup, conduction, and analysis of listening experiments: <http://audiogroup.web.th-koeln.de/scale.html>

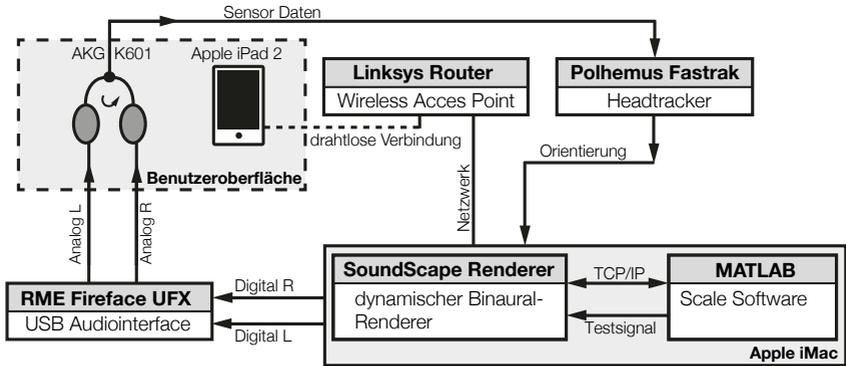


Abbildung 3.8: Schematischer Aufbau des technischen Systems für die perzeptive Untersuchung mittels dynamischer Binauralsynthese.

ist im Aufbau nicht vorgesehen. Auf eine Entzerrung der Kopfhörer wird verzichtet, da alle Stimuli über den selben Kopfhörer wiedergegeben werden, sodass dieser immer Teil der Signalverarbeitungskette ist.

Die Experimente werden im reflexionsarmen Raum der Technischen Hochschule Köln (11,7m × 4,5m × 2,3m) durchgeführt, sodass eine ruhige Umgebung mit wenigen visuellen Reizen gewährleistet werden kann. Es wird ein Ruhegeräuschpegel von unter 20dB(A) erreicht. Ein Drehstuhl ermöglicht es den Probanden während der Versuche beliebige Dreh-Positionen einzunehmen und diese auch zu variieren.

Ein *Apple iMac* Computer mit *RME Fireface UFX* USB-Audiointerface wird als zentraler Rechner sowohl für das binaurale Rendering als auch für die Steuerung der Experimente verwendet. Der Hörversuch und dessen grafische Benutzeroberfläche (GUI) wird unter MATLAB mit Hilfe der Software Scale kontrolliert. Die nachhallfreien Testsignale werden intern von MATLAB über das *JACK Audio Connection Kit* an den SoundScape Renderer weitergeleitet und dort abhängig von der Dreh-Position mit den entsprechenden BRIRs gefaltet. Dann wird das zwei-kanalige, binaurale Ausgangssignal am USB-Interface analog ausgegeben und über den Kopfhörer wiedergegeben. Zudem ermöglicht das System ein direktes Umschalten zwischen verschiedenen BRIR-Datensätzen ohne Verzögerung. Dazu werden die BRIRs in den Arbeitsspeicher (RAM) geladen und können so unverzüglich und ohne Ladezeiten abgerufen werden. Dabei hängt deren maximale Anzahl nur von dem zur Verfügung stehenden RAM ab, das hier eingesetzte System ist mit 16 GB ausreichend dimensioniert.

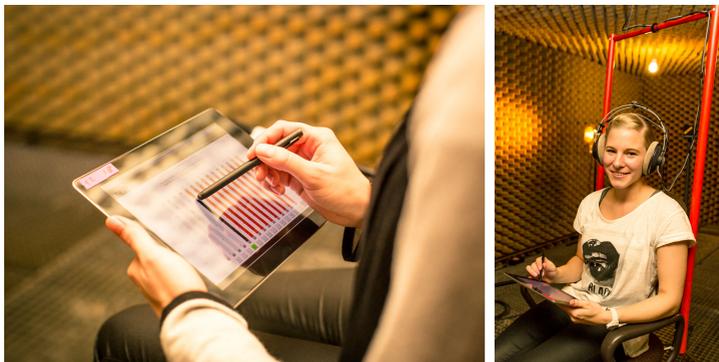


Abbildung 3.9: Umgebung der Hörversuche: Teilnehmer auf Drehstuhl im reflexionsarmen Raum, Wiedergabe über Kopfhörer mit Headtracker, Eingabe über Tablet.

Es werden zwei verschiedene Testsignale verwendet, die sich spektral als auch im Zeitverlauf unterscheiden: Eine Schlagzeug-Sequenz bestehend aus Kick-Drum, Snare-Drum und Hi-Hats mit einer Länge von 3 s sowie ein Folge von zwei weißen Rauschimpulsen mit 500 ms Dauer, 1 ms Fade-In/Out und 1000 ms Pause zwischen den Impulsen (siehe Abbildungen 3.10 und 3.11). Die Stimuli sind lautheits-normalisiert gemäß ITU-R BS 1770-4 (2015), die entsprechende Kalibrierung berücksichtigt jeweils das Testsignal als auch den zur Faltung verwendeten BRIR-Datensatz. Die Wiedergabepegel liegen bei 64,5 dB(A) bzw. 69,5 dB(A) (nur Experiment LEA1/2). Die GUI des Hörversuchs wird mit Hilfe einer VNC (Virtual Network Computing) Software drahtlos auf ein *Apple iPad2* gespiegelt und ermöglicht so den Probanden eine komfortable Bedienung und Eingabe der Bewertungen direkt während der dynamischen Auralisation im Hörversuch. Bei jedem Versuch ist ein Versuchsleiter anwesend, um ein einwandfreies technisches System als auch die korrekte Durchführung der Versuche zu gewährleisten.

3.6 Versuchsanordnungen

Im Folgenden Abschnitt werden die zwei Versuchsanordnungen, die in dieser Arbeit zur perceptiven Evaluation der entwickelten Methoden und Modelle Anwendung finden, erläutert.

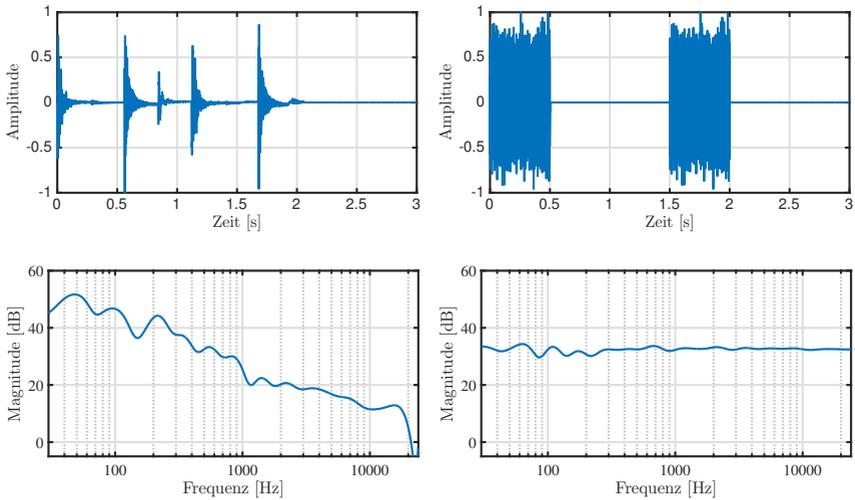


Abbildung 3.10: Zeitverlauf (oben) und Frequenzgang (unten) der Testsignale Schlagzeug (links) und Rauschimpulse (rechts).

3.6.1 ABC/HR

Der *double-blind triple-stimulus with hidden reference* (oder kurz ABC/HR) Test ist in ITU-R BS.1116-3 (2015) standardisiert und findet meist Anwendung in der Bewertung von verlustbehafteten Audiokodern wie z.B. dem mp3-Standard. Hörtests dieser Art sind im Speziellen dazu gedacht, perzeptiv geringe Beeinträchtigungen zu untersuchen. Diese Versuchsanordnung findet im überwiegenden Teil, der im Rahmen dieser Arbeit durchgeführten Hörversuche (siehe Kapitel 5.3.4, 5.4.4, 5.5.4 und 5.6.4) Anwendung und wird daher ausführlich erläutert.

Dazu werden den Probanden drei Stimuli vorgespielt, wobei Stimulus C der Referenz (unbearbeitetes Audiosignal) entspricht. Als Stimuli A und B werden randomisiert erneut die Referenz C (*hidden reference*) sowie das kodierte Audiosignal präsentiert. Der Proband hat die Aufgabe, den veränderten Stimulus zu erkennen und dessen Beeinträchtigung im Vergleich zur Referenz auf einer standardisierten Skala (siehe Tabelle 3.6.1) zu bewerten. Dabei steht 5.0 für eine unhörbare Beeinträchtigung, 1.0 bezeichnet die größtmögliche Störung. Die Eingabe erfolgt in der Regel über einen entsprechenden Schieberegler. Klassischerweise werden die Ergebnisse eines ABC/HR Experiments als sogenannte *difference grades* berechnet, indem die Bewertung der versteckten Referenz

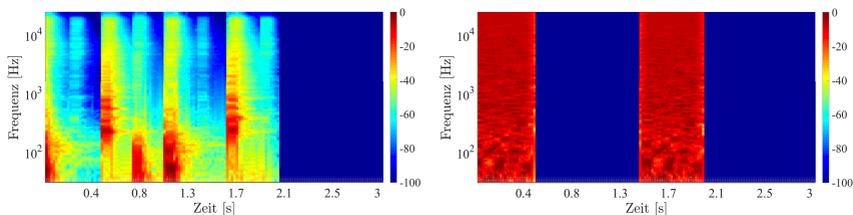


Abbildung 3.11: Spektrogramme der verwendeten Testsignale Schlagzeug (links) und Rauschimpulse (rechts).

renz von der Bewertung des kodierten Reizes subtrahiert wird. Dadurch weisen negative Werte auf eine korrekte Identifikation der kodierten Stimuli hin, während positive Werte auftreten, wenn der Proband die Referenz als beeinträchtigt bewertet. Je näher die Werte bei 0 liegen, desto geringer wurde die Beeinträchtigung des Stimulus bewertet. Zudem ist gemäß ITU-R BS.1116-3 (2015) eine Trainings-Phase vorgeschrieben, die der eigentlichen Bewertungs-Phase vorangestellt wird, sodass sich die Probanden mit der Benutzeroberfläche des Versuchs, der Skala und der Bandbreite der verwendeten Stimuli vertraut machen können.

Für die Hörversuche dieser Arbeit wird das standardisierte ABC/HR Versuchsdesign in leicht angepasster Form verwendet. Die Untersuchungen zielen somit nicht auf die Bewertung der Beeinträchtigung¹⁸ der Stimuli aufgrund eines Algorithmus ab. Stattdessen soll unter Einsatz der dynamischen Binauralsynthese die Ähnlichkeit von Stimuli basierend auf einer gemessenen Referenz-BRIR und Stimuli basierend auf SBRIRs evaluiert werden. So soll allgemein untersucht werden, wie unterschiedlich gemessene und synthetische BRIRs im direkten Vergleich zueinander bewertet werden. Dazu wird die Referenz immer mit Reiz A abgespielt, der einer Faltung der gemessenen BRIRs mit dem Testsignal zur jeweiligen Orientierung des Rezipienten entspricht. Stimuli B und C entsprechen wiederum randomisiert zum einen erneut der Referenz und zum anderen einer Faltung der synthetischen BRIRs auf Basis der parametrischen Modelle und gemäß der Orientierung mit dem Testsignal.

Es gilt sicherzustellen, dass jeder Proband die identische Versuchseinführung und Vorbereitung erhält. Daher wird die folgende, ausformulierte Aufgabenstellung jedem Probanden zu Beginn des Experimentes visuell vorgelegt und wörtlich vorgelesen:

¹⁸laut ITU-R BS.1116-3 (2015): „impairment“

Tabelle 3.2: Skala der Beeinträchtigung nach ITU-R BS.1116-3 (2015).

1.0	very annoying
2.0	annoying
3.0	slightly annoying
4.0	perceptible, but not annoying
5.0	imperceptible

Ihre Aufgabe besteht darin, jedes Paar (B und C) mit der Referenz A zu vergleichen und zu entscheiden, welcher Stimulus NICHT der Referenz entspricht. Ihre Entscheidung äußern Sie, indem Sie am entsprechenden Schieberegler einstellen, wie deutlich Sie den wahrgenommenen Unterschied (von identisch bis sehr unterschiedlich) empfinden.

Analog zu Brinkmann (2011) und Lindau und Brinkmann (2012) wird zur Bewertung die ITU-Skala nach Tabelle 3.6.1 aufgrund unterschiedlicher sprachlicher Bedeutungen der Attribute, unzureichender Äquidistanz und fehlendem Bezug zur Ähnlichkeit, leicht modifiziert. Die Endpunkte der Skala werden in dieser Studie nur mit den Attributen *identisch* und *sehr unterschiedlich* gekennzeichnet, dazwischen liegenden Bezeichner werden entgegen ITU-R BS.1116-3 (2015) nicht verwendet. Lediglich äquidistante Markierungen zur verbesserten Orientierung finden Anwendung. Die für die Probanden während des Versuchs dauerhaft sichtbar ausformulierte Fragestellung lautet:

*Welcher Stimulus (B oder C) ist unterschiedlich zur Referenz A?
Bitte bewerten Sie den wahrgenommenen Unterschied (von identisch bis sehr unterschiedlich).*

Abbildung 3.12 zeigt die Benutzerschnittstelle des Versuchs, die den Probanden für jeden Vergleich auf dem Tablet-PC dargeboten wird und über die die Bewertungen abgegeben werden. Es wird zunächst eine definierte Sequenz abgespielt (A, B, A, C), die zudem noch durch eine farbliche Hervorhebung (grün) des entsprechenden Tasters visuell unterstützt wird. Nach dieser Sequenz ist eine beliebige Wiederholung der Stimuli möglich, wobei jedes Signal immer komplett abgespielt wird und zur Laufzeit nicht unterbrochen werden kann. Zudem wird durch Pausen am Ende der Testsignale sichergestellt, dass jeder auralisierte Raum komplett ausklingen kann. Insbesondere bei langen Nachhallzeiten ist dies essenziell. Die Abgabe der Bewertungen erfolgt stufenlos mit Hilfe von zwei Schieberegler (oben B, unten C), eine ausreichend hohe Abstufung wird verwendet. Weiterhin können die Probanden durch einen Taster mit

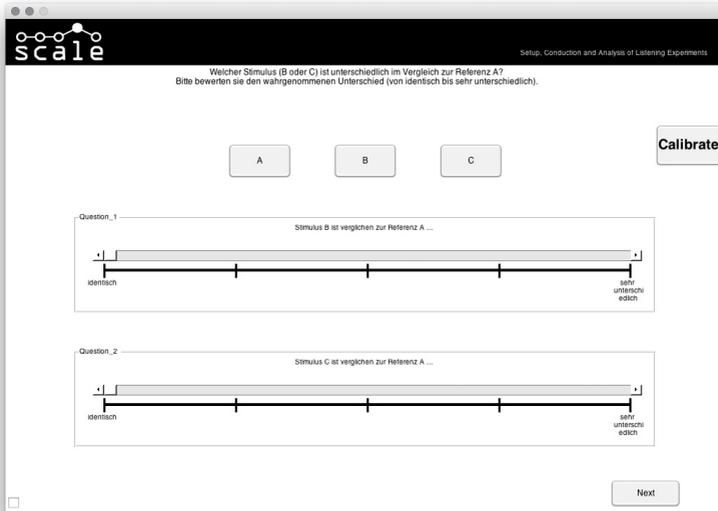


Abbildung 3.12: Bildschirmfoto Benutzerschnittstelle der ABC/HR Versuchsumgebung. Eingabe der Bewertungen der Probanden über Tablet-Computer.

der Bezeichnung „Calibrate“ die Orientierung des Binauralsystems zurücksetzen: Dies hat zur Folge, dass die auralisierte Schallquelle bei Aktivierung des Tasters auf 0° gesetzt wird (relativ gesehen zur Orientierung des Kopfhörers). So kann beispielsweise sichergestellt werden, dass die Probanden die Schallquelle immer aus Richtung der Vorderseite des reflexionsarmen Raumes wahrnehmen. Im Rahmen des Trainings werden die Probanden entsprechend instruiert und auf die Besonderheiten der dynamischen Binauralsynthese hingewiesen.

Der Versuchsplan ist balanciert, d.h. es wird für jede Bearbeitung die gleiche Anzahl von Bewertungen (mit n Wiederholungen, siehe Tabellen 5.8, 5.10, 5.15, 5.19 und 5.23) erhoben. Die Wiedergabe erfolgt randomisiert, sodass jeder Proband die Stichproben in einer zufälligen Reihenfolge bewerten muss.

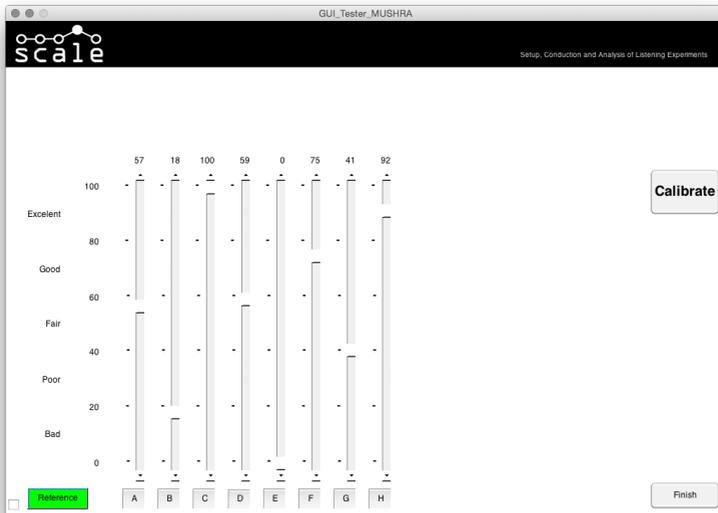


Abbildung 3.13: Bildschirmfoto Benutzerschnittstelle der MUSHRA Versuchsumgebung. Eingabe der Bewertungen der Probanden über Tablet-Computer.

3.6.2 MUSHRA

Beim *Multi Stimulus test with Hidden Reference and Anchor* (MUSHRA), der in ITU-R BS 1534-3 (2015) standardisiert ist, handelt es sich um einen Hörtest, mit dem zahlreiche Stimuli hinsichtlich ihrer Audioqualität vergleichend beurteilt werden können. Wie auch der ABC/HR Test wird eine MUSHRA Versuchsanordnung meist für die Bewertung von verlustbehafteten Audiokodecs eingesetzt, entgegen ITU-R BS.1116-3 (2015) wird jedoch der Einsatz von MUSHRA für eine mittlere Audioqualität empfohlen. In dieser Arbeit wird diese Versuchsanordnung zur Untersuchung verschiedener Übergangszeiten in Experiment MT1/2 verwendet (siehe Abschnitt 5.2).

Der Proband hat die Aufgabe, den Unterschied von mehreren Stimuli im Vergleich zu einer Referenz auf einer kontinuierlichen Qualitäts-Skala (CQS) von 0 - 100 zu bewerten. Die Skala ist weiterhin mit den Attributen *bad* (CQS = 0 - 20), *poor* (CQS = 20 - 40), *fair* (CQS = 40 - 60), *good* (CQS = 60 - 80) und *excellent* (CQS = 80 - 100) beschriftet. Eine Bewertung von 100 entspricht einer Ununterscheidbarkeit der Modifikation zum

Referenzstimulus. Weiterhin sind eine oder mehrere Anker-Signale vorgesehen, die in der Regel durch eine Bandbegrenzung des Stimulus erzeugt werden. Zusätzlich ist zur Verifikation der Ergebnisse noch eine versteckte Referenz integriert, die die Probanden mit mehr als 90 Punkten bewerten sollten. Entgegen der ABC/HR Versuchsanordnung, erlaubt MUSHRA ein beliebiges Umschalten zwischen den dargebotenen Stimuli zur Laufzeit. Das bedeutet, dass auch während der Wiedergabe eines Stimulus, ein Wechsel auf einen anderen Stimulus erfolgen kann. Weiterhin können in diesem Versuch mehrere Bearbeitungen gleichzeitig miteinander verglichen werden und es gibt keine Limitierung der Wiederholungen. Der Versuch ist somit erst abgeschlossen, wenn der Proband die Bewertungen seiner Meinung nach sinnvoll zueinander angeordnet hat und so das Ergebnis bestätigt. Laut ITU-R BS 1534-3 (2015) wird jedoch empfohlen nicht mehr als 12 Signale pro Probe miteinander zu vergleichen (9 Bearbeitungen, ein versteckter schlechter Anker, ein versteckter mittlerer Anker und eine versteckte Referenz).

In der vorliegenden Arbeit wird dieser Versuchsplan leicht modifiziert angewendet. Die Probanden müssen die globale Ähnlichkeit der Auralisation auf Basis der dargebotenen modifizierten BRIRs (basierend auf Datensätzen mit unterschiedlichen Übergangszeiten) zur Auralisation mittels messtechnisch erfasster Referenz-BRIR bewerten. Pro Vergleich werden ausschließlich BRIRs aus dem selben Raum untersucht. Abweichend von der ITU-Empfehlung wird als Anker eine Mono-Impulsantwort aus den jeweiligen Räumen verwendet. Entsprechend Abschnitt 3.6.1 ist der Versuchsplan ebenso balanciert sowie es erfolgt eine randomisierte Wiedergabe der zu bewertenden Stimuli. In Abbildung 3.13 findet sich die auf dem Tablet-PC dargebotenen GUI zur Eingabe der Bewertungen durch die Probanden.

3.7 Statistische Verfahren

Da die perceptiven Experimente dieser Arbeit nur mit zwei unterschiedlichen Versuchsanordnungen durchgeführt werden, kann die statische Auswertung auf die identischen Werkzeuge zurückgreifen. Lediglich die Faktoren und deren Faktorstufen und somit die Anzahl der untersuchten Bedingungen unterscheidet sich zwischen den Experimenten. Es werden ausschließlich sogenannte *within-(subject)*-Testanordnungen betrachtet, da in jedem Versuch nur eine Probandengruppe existiert, d.h. alle Probanden bewerten stets die identischen Bedingungen. Ein Signifikanzniveau von $\alpha = 0,05$ wird für alle statistischen Auswertungen verwendet. Die Überprüfung der Gesamtverteilung auf Normalität wird nach Shapiro und Wilk (1965) durchgeführt. Die Auswertung erfolgt unter MATLAB und der *Statistics and Machine Learning Toolbox*. Zudem werden alle Ergebnisse mit der Statistik-Software *SPSS* vergleichend verifiziert.

3.7.1 t-Test

Der Begriff t -Test wurde von Student (1908)¹⁹ eingeführt und benennt mehrere Arten von Signifikanztests zur Überprüfung von Unterschiedshypothesen. Zwei Varianten des t -Tests werden im Folgenden vorgestellt:

1-Stichproben t-Test: Mit Hilfe des 1-Stichproben t -Test (engl. one sample t -test) wird überprüft, ob „eine Stichprobe von einer Populationsverteilung mit einem bestimmten Populationsmittelwert abweicht bzw. ob die Stichprobe einer speziellen Referenzpopulation entstammt“ (Rasch et al., 2014, S.67). Somit wird die Nullhypothese H_0 (es besteht kein Unterschied zwischen dem Mittelwert μ der Grundgesamtheit und dem vorgegebenen Wert μ_0) gegen die Alternativhypothese H_1 (es besteht ein Unterschied, $\mu \neq \mu_0$) überprüft. Dabei gilt zwischen drei verschiedenen Hypothesenpaaren zu unterscheiden:

$$\begin{aligned} H_0 : \mu = \mu_0 \text{ und } H_1 : \mu \neq \mu_0 & \quad (\text{zweiseitiger Test, ungerichtet}) \\ H_0 : \mu \leq \mu_0 \text{ und } H_1 : \mu > \mu_0 & \quad (\text{rechtsseitiger Test, gerichtet}) \\ H_0 : \mu \geq \mu_0 \text{ und } H_1 : \mu < \mu_0 & \quad (\text{linksseitiger Test, gerichtet}) \end{aligned}$$

Der t -Wert wird durch den Mittelwert \bar{x} einer Zufallsstichprobe des Umfangs n und dessen Stichprobenstandardabweichung s mit

$$t = \sqrt{n} \left(\frac{\bar{x} - \mu_0}{s} \right) \quad (3.1)$$

und

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (3.2)$$

gebildet.

Laut Student (1908) können abhängig der verwendeten Freiheitsgrade $df = n - 1$ und des gewählten Signifikanzniveaus α verschiedene Funktionen der t -Verteilung definiert werden. Anhand dieser Funktionen können sogenannte t -Tabellen erstellt werden (Glass und Stanley, 1970, S. 521), mit Hilfe dieser der Ablehnungsbereich (bei Beachtung des verwendeten Hypothesenpaares) der Nullhypothese (hier liegen signifikante Ergebnisse) abgelesen werden kann. Dies gibt jedoch lediglich Auskunft, ob die Nullhypothese

¹⁹Pseudonym von William Sealy Gosset, englischer Statistiker

bei gewähltem Signifikanzniveau abgelehnt oder beibehalten wurde. Ist ein feineres Maß gewünscht, z.B. um zu erfahren, ob Nullhypothese auch für ein kleineres Signifikanzniveau abgelehnt worden wäre, so wird der p -Wert, „der auch als 'beobachtetes Signifikanzniveau' bezeichnet wird, verwendet. Der p -Wert entspricht dem kleinsten Wert von α , für den das Testergebnis gerade noch Signifikanz erreicht“ (Bortz und Schuster, 2010, S.106). Er wird anhand des Integrals des Ablehnungsbereiches berechnet.

2-Stichproben t-Test für Beobachtungspaare: Der 2-Stichproben t-Test für Beobachtungspaare (engl. two sample t -test) untersucht den Unterschied zwischen zwei abhängigen Gruppen. Mithilfe der Mittelwerte der abhängigen Stichproben \bar{x}_1 und \bar{x}_2 wird untersucht, ob ein Unterschied zwischen den Mittelwerten μ_1 und μ_2 der zugehörigen Grundgesamtheiten besteht. Dabei wird die Nullhypothese (es besteht kein Unterschied zwischen den Mittelwerten der Grundgesamtheiten, $H_0 : \mu_1 = \mu_2$) der Alternativhypothese (es besteht ein Unterschied, $H_1 : \mu_1 \neq \mu_2$) gegenübergestellt. Wie beim 1-Stichproben t -Test können die entsprechenden Alternativhypothesen für das einseitige Testen aufgestellt werden. Auch die Berechnung folgt ähnlichen Regeln wie zuvor gesehen. Die Differenz d_i kann für jedes Messwertpaar x_{i1} und x_{i2} mit

$$d_i = x_{i1} - x_{i2} \tag{3.3}$$

gebildet werden. Mit i wird dabei der Proband und mit 1/2 die Bedingung gekennzeichnet. Die Mittelwertsdifferenz \bar{d} ist somit:

$$\bar{d} = \frac{1}{n} \sum_i x_{i1} - x_{i2}. \tag{3.4}$$

Der t -Wert berechnet sich aus:

$$t = \sqrt{n} \left(\frac{\bar{d}}{s_d} \right) \tag{3.5}$$

mit

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}. \tag{3.6}$$

In einer t -Tabelle kann wiederum abgelesen werden, ob das Ergebnis signifikant ist, falls der errechnete t -Wert größer ist, als der entsprechende Eintrag (bei Beachtung des Signifikanzniveaus α und des Freiheitsgrades $df = n - 1$).

3.7.2 Varianzanalyse

Eine erhobene Stichprobe kann mit Hilfe einer mehrfaktoriellen Varianzanalyse mit Messwiederholung (rmANOVA) auf signifikante Unterschiede zwischen den Bedingungen untersucht werden (Bortz und Schuster, 2010, S.285). Entgegen dem t -Test, können mit einer Varianzanalyse auch die Mittelwerte von mehr als zwei Bedingungen sinnvoll verglichen werden. Für Faktorstufen größer 3 sind die Freiheitsgrade der Varianzanalyse nach Geisser und Greenhouse (1958) korrigiert. Diese Korrektur ist bei Verletzung der Annahme der Sphärizität nötig.

Mit Hilfe der Ergebnisse der Varianzanalyse können gemäß Bakeman (2005) unter Verwendung des partiellen Eta-Quadrats η_p^2 die Effektgrößen berechnet und damit die Relevanz der signifikanten Ergebnisse beurteilt werden. Zur Interpretation dieser Werte hat Cohen (1988) die folgende Klassifikation entwickelt (S.414-415): Für $\eta_p^2 = 0,02$ spricht Cohen von einer kleinen Effektstärke, für $\eta_p^2 = 0,13$ von einer mittleren Effektstärke und für $\eta_p^2 = 0,26$ von einer großen Effektstärke. Alle durchgeführten ABC/HR Experimente werden in der vorliegenden Arbeit mit einer rmANOVA ausgewertet und die Effektgrößen η_p^2 bestimmt.

3.7.3 Post-hoc Tests

Entgegen den Ergebnissen der Varianzanalyse, die ausschließlich untersucht, ob es signifikante Unterschiede in einer Gruppe von Mittelwerten gibt, kann mitunter auch der direkte Vergleich zweier Bedingungen von Interesse sein. Mit Hilfe dieser sogenannten Post-hoc Tests kann somit untersucht werden, welche Mittelwerte sich im speziellen voneinander unterscheiden. Dabei werden alle Mittelwerte paarweise miteinander verglichen und so signifikante Unterschiede zwischen einzelnen Bedingungen gefunden. Dabei kann z.B. ein 2-Stichproben t -Test zum Einsatz kommen. Werden dabei m Bedingungen mit m einzelnen t -Tests (mit gleichen Signifikanzniveau α) überprüft, so führt dies zu einem Anstieg des Gesamt-Signifikanzniveaus α_{gesamt} , der sogenannten α -Fehler-Kumulierung:

$$\alpha_{\text{gesamt}} = 1 - (1 - \alpha)^m. \quad (3.7)$$

Daher wird das Signifikanzniveau nach Hochberg (1988) korrigiert, um die α -Fehler-Kumulierung der multiplen Vergleiche zu unterbinden.

3.8 Zusammenfassung Methoden, Technologien und Ressourcen

Eine Vielzahl von akkuraten, akustischen Messsystemen, wie z.B. ein sequenzielles, sphärisches Mikrofonarray, ein sequentiell bewegbarer Kunstkopf und leistungsfähige Schallquellen mit unterschiedlicher Richtcharakteristik (omnidirektional und gerichtet) wurden kombiniert verwendet, um in zwei Konzertsälen des Westdeutschen Rundfunks Köln Referenz-Messdaten mit hoher Qualität aufzuzeichnen. Es gilt später, diese Daten mithilfe eigens entwickelter Verfahren möglichst adäquat synthetisch nachzubilden, um diese zur Auralisation verwenden zu können. Dazu müssen die Messdaten mit der entsprechenden vorgestellten Software ausgewertet und mit Methoden der räumlichen Fourier-Akustik analysiert werden. So können die gewonnenen Messdaten, als auch die synthetisch erzeugten Datensätze in der vorgestellten und eigens aufgebauten Hörversuchsumgebung mittels dynamischer Binauralsynthese auralisiert, im Rahmen eines Hörversuchs miteinander verglichen und mit der vorgestellten Methodik statistisch ausgewertet werden. Mit dem ABC/HR Test wird eine standardisierte Versuchsanordnung verwendet, die oftmals für die Untersuchung der Beeinträchtigung einer akustischen Wiedergabe aufgrund unterschiedlicher Audiokodecs eingesetzt wird. In den vorliegenden Experimenten wird das Verfahren leicht modifiziert verwendet, um die Ähnlichkeit zwischen den gemessenen Referenz-Datensätzen und deren synthetisch generierten Pendanten zu evaluieren.

KAPITEL 4

Signalverarbeitung für die BRIR-Synthese

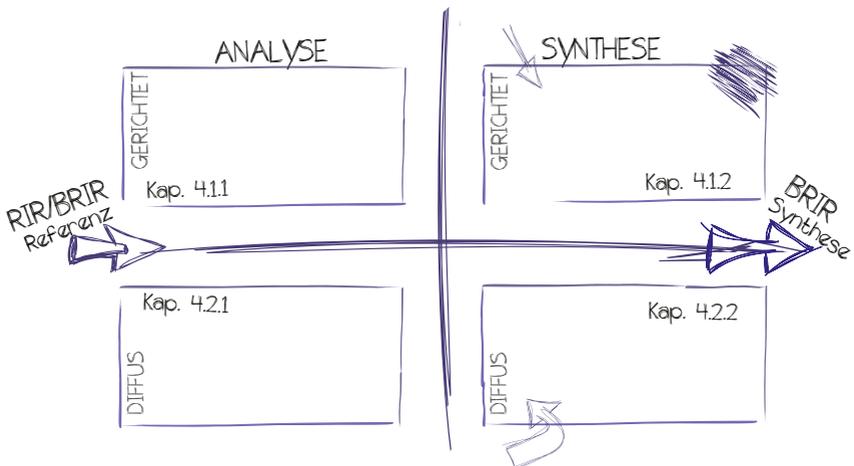


Abbildung 4.1: Konzeptioneller Aufbau der Signalverarbeitung gerichteter und diffuser Komponenten mit Analyse- und Synthese-Modul inklusive Kapitelstruktur.

Dieses Kapitel befasst sich mit allen relevanten Aspekten der digitalen Signalverarbeitung, die in dieser Arbeit zur Synthese von binauralen Datensätzen (SBRIRs) zur dynamischen Auralisation implementiert werden. Alle Verarbeitungsschritte finden kombiniert Anwendung in den parametrischen Modellen (siehe Kapitel 5), werden aber im Vorfeld zum besseren Verständnis jeweils separat für sich erläutert. Grundsätzlich lässt sich anhand des Schallfeldes die Signalverarbeitung zur BRIR-Synthese in zwei Komponenten – *diffus* und *gerichtet* – aufteilen, entsprechend dieser Bezeichnungen ist auch das folgende Kapitel strukturiert. Eine Visualisierung des konzeptionellen Aufbaus und der Kapitelstruktur findet sich Abbildung 4.1 anhand einer Skizze.

Das Kapitel setzt sich im ersten Schritt zum Ziel, das Konzept der BRIR-Synthese zu erörtern. Im zweiten Schritt wird die konkrete Realisierung erläutert, die in dieser Arbeit Anwendung findet. Möglich ist jedoch auch eine andere Umsetzung, durch Austausch einzelner Verarbeitungsschritte.

Entgegen der in der Literatur üblicherweise verwendeten Anschauungsweise (siehe z.B. Välimäki et al. (2012, 2016)) erfolgt keine zeitliche Trennung der Impulsantworten in Direktschall, erste Reflexionen und später Nachhall wie in Abbildung 2.8 gezeigt. Vielmehr wird das Schallfeld so betrachtet, das zu jedem Zeitpunkt diffuse als auch gerichtete Schallfeldanteile auftreten können. Daher erfolgt die Verarbeitung auch jeweils für die gesamte Impulsantwort, eine zeitliche Fensterung ist aber anschließend möglich und wird teilweise in den parametrischen Modellen angewendet. Die gesamte Signalverarbeitung basiert auf Impulsantworten, eine Betrachtung und Verarbeitung von zur Laufzeit im Raum aufgezeichneten Signalen ist bisher nicht vorgesehen.

Weiterhin lässt sich die Signalverarbeitung der diffusen und gerichteten Schallfeldanteile jeweils in zwei Blöcke segmentieren: Für beide Komponenten existiert ein *Analyse*- und ein *Synthesemodul*. Prinzipiell ist diese zweistufige Vorgehensweise für die einfache BRIR-Generierung nicht erforderlich. Um ein geeignetes Parametermodell zur Synthese gemessener Schallfelder zu entwickeln, welches sich zudem komprimiert speichern lässt und eine beliebige Modifikation der Parameter ermöglicht, ist dieses Verfahren aber sinnvoll und nötig. Der Fokus dieses Kapitels liegt jedoch auf einer allgemeinen Beschreibung der nötigen Signalverarbeitung. Die Beleuchtung der verwendeten Parameter und Konfigurationen sowie die Erläuterung der Verknüpfung von Analyse und Synthese findet sich erst in den entsprechenden Modellen in Kapitel 5.

Die Analyse bestimmt die gerichteten (siehe 4.1.1) und diffusen Komponenten (siehe 4.2.1) eines messtechnisch erfassten Schallfeldes und speichert diese in geeigneten Variablen. Dabei basiert das System sowohl auf binauralen Messdaten, als auch auf Impulsantworten die mit einem sphärischen Mikrofonarray erfasst wurden. Hierbei werden Methoden der Wellenfeldanalyse zur Schallfelderlegung (siehe Abschnitt 2.5) angewendet. Das Synthese-Modul der gerichteten (siehe 4.1.2) und diffusen Kompo-

nenten (siehe 4.2.2) befasst sich mit der Generierung synthetischer BRIR-Datensätze zur dynamischen Auralisation. Zunächst wird die Erzeugung der SBRIRs unabhängig vom Analyse-Modul betrachtet, obwohl die Synthese grundsätzlich auf der Struktur von Schallfeld-Komponenten basiert, wie sie in der Analyse bestimmt werden. Dabei müssen die Komponenten aber nicht zwingend dem Analyse-Modul entspringen, eine beliebige Wahl ist ebenfalls möglich. Weiterhin werden diffuse und gerichtete Komponenten separat verarbeitet und es besteht keine Abhängigkeit, sodass sie auch gesondert synthetisiert werden können.

In Abschnitt 4.1 wird die Signalverarbeitung der gerichteten Komponenten erläutert. Mittels Schallfelderlegung werden Direktschall und Reflexionen örtlich detektiert und deren Pegel, Einfallsrichtung und Zeitpunkt in einer Tabelle abgelegt. Die spektrale Zusammensetzung der gerichteten Anteile kann weiterhin zur Bildung von Reflexions-Filtern bestimmt werden. BRIRs mit gerichteten Komponenten können durch die Verwendung eines sphärischen HRIR-Datensatzes und eines Echogramms mit räumlicher Beschreibung in Kombination mit entsprechenden Filtern generiert werden.

Die Erläuterung der Signalverarbeitung der diffusen Schallfeldanteile findet sich in Abschnitt 4.2. Es werden die frequenzabhängigen Abklingkurven einer Impulsantwort analysiert und anschließend mit möglichst wenigen Werten approximiert. Weiterhin wird die interaurale Kohärenz eines gemessenen binauralen Signalpaares bestimmt und deren Verlauf über der Frequenz ebenfalls nachgebildet. Binaurale, diffuse Schallfeldanteile können durch die Anpassung von Rauschsignalen mittels frequenzabhängigem Energieabfall und einer Adaption der Kohärenz generiert werden.

4.1 Gerichtete Komponenten des Schallfeldes

Als gerichtete Komponenten werden alle Anteile des Schallfeldes verstanden, die für den Rezipienten Hörereignisse mit wahrnehmbaren Richtungsinformationen erzeugen. Das bedeutet, die gerichteten Komponenten bilden Schallereignisse, denen der Hörer eine diskrete Richtung, aus der er den Schall wahrnimmt, zuordnen kann oder aber die gemäß der Wahrnehmung von Reflexionen die Hörereignisrichtung beeinflussen. Dabei handelt es sich um den am Hörort eintreffenden Direktschall, der nach einer Schallfeldanregung des Raumes ohne Umwege den Hörer erreicht, sowie um Rückwürfe des Schalls, die erst nach einfacher oder mehrmaliger Reflexion an Objekten und/oder Begrenzungsflächen des Raumes hörbar werden. Entsprechend der Betrachtungsweise der diffusen Komponenten, werden auch für die gerichteten Komponenten nur die für das auditive System wahrnehmbaren und nicht alle physikalisch korrekt auftretenden Anteile beleuchtet. So ergeben sich wiederum Möglichkeiten, die Informationen des Schall-

feldes im Raum entsprechend stark zu vereinfachen und auf die wahrnehmbaren Anteile zu reduzieren. Für die Wahrnehmung der gerichteten Komponenten spielen maßgeblich deren Pegel, deren Einfallrichtung und deren Verzögerungszeit im Vergleich zu zuvor auftretenden Schallereignissen eine Rolle (siehe Abschnitt 2.1.1). Daher gilt es, die gerichteten Schallfeldanteile immer geschlossen zu analysieren, um auftretende Maskierungseffekte in die Betrachtung und spätere Auralisation mit einzubeziehen.

So ist das Ziel dieses Verarbeitungsschrittes, Reflexionen im Schallfeld zu detektieren und Aussagen über deren Pegel, Einfallrichtung, Auftrittszeitpunkt und spektraler Färbung zu erhalten. Die Bestimmung der gerichteten Komponenten erfolgt mit Methoden aus der Wellenfeldanalyse (siehe Kapitel 2.5) auf Basis eines Datensatzes, der mit einem sphärischen Mikrofonarray gewonnen wurde. So wird ein Echogramm des Raumes erzeugt, welches weiterhin noch Informationen über die räumliche Anordnung der gerichteten Schallfeldanteile liefert. Durch Bildung entsprechender *gerichteter Impulsantworten* (DIR) kann zudem selektiv die spektrale Verteilung der gerichteten Komponenten bestimmt werden. Durch die Verwendung eines räumlich aufgelösten Echogramms in Kombination mit sphärisch, gemessenen HRIR-Datensätzen und den entsprechenden Reflexions-Filtern, können synthetische BRIRs mit gerichteten Komponenten zur dynamischen Auralisation generiert werden.

4.1.1 Analyse der gerichteten Komponenten

Das Blockschaltbild in Abbildung 4.2 zeigt schematisch den Aufbau der Analyse der gerichteten Komponenten. Das Modul benötigt im Eingang einen geschlossenen Datensatz von Raumimpulsantworten, der mit einem sphärischen Mikrofonarray im zu synthetisierenden Raum gewonnen wurde. Abhängig von der Konfiguration des Arrays (Radius, Anzahl/Anordnung der Messpunkte) kann mit unterschiedlicher Genauigkeit und Bandbreite das Schallfeld mittels Wellenfeldanalyse zerlegt werden.

Dazu kann eine richtungssensitive Intensitätsmatrix des Schallfeldes im Raum erzeugt werden. Ein spezieller Klassifizierungs-Algorithmus durchsucht diese Matrix nach lokalen Maxima und detektiert so den Pegel, den Zeitpunkt und die Einfallrichtung von gerichteten Komponenten im Bezug auf die gewählte Empfängerposition (das Zentrum des Mikrofonarrays, siehe Abschnitt 4.1.1.1). Anschließend werden gerichtete Impulsantworten für jede Reflexion gebildet (siehe 4.1.1.2), um deren spektrale Färbung mittels eines Filters zu charakterisieren (siehe 4.1.1.3).

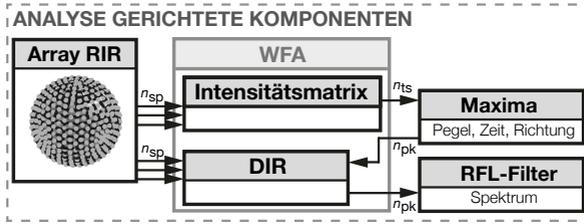


Abbildung 4.2: Schematische Darstellung der Analyse der gerichteten Komponenten (Ausschnitt aus Parameter-Modell) auf Basis einer sphärischen Mikrofonarray-Messung: Methoden der Wellenfeldanalyse (WFA) werden angewendet um eine zeit- und richtungssensitive Intensitätsmatrix des Schallfeldes zu erzeugen. Ein Algorithmus klassifiziert Maxima und detektiert so Pegel, Zeitpunkt und Richtung von Reflexionen. Gerichtete Impulsantworten (DIR) werden mittels WFA gebildet und zur Bestimmung der spektralen Färbung der Reflexionen verwendet.

4.1.1.1 Schallfelderlegung zur Detektion von Maxima

Mit Methoden der Wellenfeldanalyse kann das Schallfeld auf Basis eines entsprechend gemessenen Datensatzes in ebene Wellen zerlegt werden (siehe Kapitel 2.5.2). Dies erlaubt eine Auswertung des Schallfeldes hinsichtlich des Zeit- und Pegelverlaufs sowie der räumlichen Anordnung eintreffender Schallfeldanteile. Das Verfahren hat sich in der Vergangenheit für die raumakustische Analyse bewährt (Duraiswami et al., 2005a; Farina et al., 2011; Stade et al., 2014; Spors und Rettberg, 2018) und bietet die Möglichkeit der Detektion von Reflexionen im Raum, sodass ein räumlich aufgelöstes Echogramm erzeugt werden kann.

Intensitätsmatrix

Zunächst muss eine räumlich und zeitlich aufgelöste Intensitätsmatrix des Schallfeldes erzeugt werden. Dazu werden die Impulsantworten des Mikrofonarray-Datensatzes mit Funktionen aus der SOFiA-Toolbox (siehe Abschnitt 3.4.1, Bernschütz et al. (2011a) prozessiert (`sofia_`). Im ersten Schritt erfolgt eine block-basierte, zeitliche Segmentierung der Impulsantworten. Je kleiner die Blöcke, desto genauer ist das zeitliche Auflösungsvermögen der Reflexionsdetektion. Jeder Block wird mittels einer zeitlichen Fourier-Transformationen vom Zeit- in den Frequenzbereich überführt (`sofia_fdt`). Anschließend erfolgt eine Transformation in die räumliche Domäne mittels räumlicher Fourier-Transformation (`sofia_sdt`), sodass das Schallfeld mittels räumlicher Fourier-

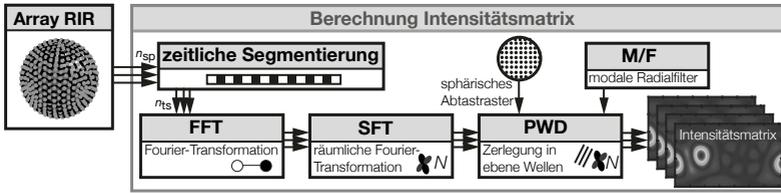


Abbildung 4.3: Aufbau Schallfeldzerlegung zur Bildung der Intensitätsmatrix aus Stade et al. (2017b): Die Impulsantworten werden zeitlich segmentiert und mittels zeitlicher und räumlicher Fourier-Transformation in die räumliche Domäne überführt. Anschließend erfolgt eine Zerlegung in ebene Wellen anhand eines gewählten Gitters. Dies erzeugt eine zeitlich und räumlich aufgelöste Intensitätsmatrix des Schallfeldes.

Koeffizienten beschrieben wird. Diese Koeffizienten im sphärischen Wellenspektrum werden verwendet, um entsprechend des gewählten Gitters, das Schallfeld jedes Blockes in ebene Wellen zu zerlegen. Das Raster und die Ordnung bestimmen das räumliche Auflösungsvermögen dieses Verarbeitungsschrittes.

In Abbildung 2.12 wurde diese Methodik bereits auf Basis einer Gauss-Quadratur visualisiert. Entsprechend der verwendeten FFT Bins der räumlichen Fourier-Koeffizienten wird die Frequenz der ebenen Welle definiert. So erhält man für jeden zeitlichen Block des Array-Datensatzes und den entsprechenden Abtastpunkt des Rasters, die Energie des Schallfeldes in diesem Frequenzbereich. Dies führt zur Bildung einer dreidimensionalen Matrix, die den Intensitätsverlauf des Schallfeldes zeit- und räumlich beschreibt (`sofia_makeEqDistMTX`).

Abbildung 4.4 zeigt exemplarisch den Aufbau der Raum-Zeit-Intensitätsmatrix des Schallfeldes: Pro Zeitblock werden die Intensitäten des Schallfeldes auf einer Kugeloberfläche dargestellt. Die Auflösung ist dabei abhängig von dem gewählten sphärischen Gitter für die Erzeugung der ebenen Wellen. Hohe Intensitäten werden mit roten Bereichen visualisiert, niedrige mit blauen Bereichen.

Detektion von Maxima

In der zuvor berechneten Intensitätsmatrix des Schallfeldes zeigen sich Reflexionen als lokale Maxima. Daher gilt es die multidimensionale Matrix nach diesen Maxima zu durchsuchen, um so gerichtete Komponenten des Schallfeldes zu detektieren. Es muss beachtet werden, dass neben den Reflexionen auch der Direktschall in der Matrix als Maximum repräsentiert wird. Dabei handelt es sich um das absolute Maximum

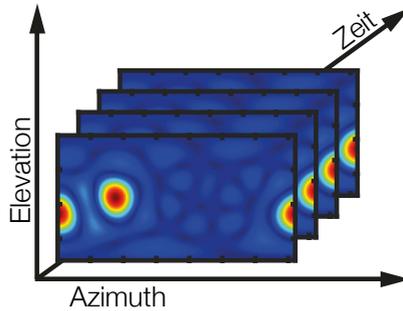


Abbildung 4.4: Prinzipieller Aufbau der Raum-Zeit-Intensitätsmatrix: Pro Zeitblock werden die Intensitäten des Schallfeldes in Abhängigkeit des Gitters der PWD entsprechend ihrer Orientierung in Azimuth- und Elevationswinkel dargestellt.

der Gesamt-Matrix. Die Syntax des Algorithmus zur Maxima-Suche ist im Folgenden kurz zusammengefasst und besteht aus verschiedenen Stadien, die iterativ aufeinander aufbauen. Der Algorithmus wurde erstmalig von Rühl (2012) vorgestellt und von Bernschütz et al. (2012); Stade et al. (2014); Stade und Arend (2016a) weiter beschrieben und angewendet sowie ist auch in der SOFiA-Toolbox enthalten (`sofia_pkd`). Abbildung 4.5 zeigt exemplarisch das Konzept des Algorithmus zur Klassifizierung von Reflexionen in der Intensitätsmatrix des Schallfeldes: Aus ursprünglich sechs lokalen Maxima klassifiziert der Algorithmus in diesem Beispiel eine valide Reflexion.

Im ersten Schritt muss ein potentielles Maximum einen gewählten Schwellwert überschreiten. Dieser Schwellwert bezieht sich jeweils auf die mittlere Energie innerhalb eines Zeitblockes (*sensitivity*). Danach wird die Intensität dieses potenziellen Maximums innerhalb eines gewählten Bereichs mit den Intensitäten von benachbarten Abtastpunkten des Gitters verglichen (*surface range*). So schätzt der Algorithmus, ob die Intensitäten einer oder mehreren eng beieinander liegenden Reflexionen zugeordnet werden können. Dabei gilt zu beachten, dass abhängig der gewählten Ordnung N die räumliche Auflösung und Richtwirkung der Analyse begrenzt ist. Daher führen Reflexionen in der Praxis immer zu einer unscharfen, räumlichen Antwort (siehe Abbildung 2.12). Neben der räumlichen Unschärfe, existiert auch eine zeitliche Ungenauigkeit, sodass eine diskrete Reflexion auch über verschiedene Blöcke zeitlich verschmieren kann. Daher wird im dritten Schritt das potenzielle Maximum mit den Intensitäten der umliegenden Zeitscheiben innerhalb eines gewählten Bereichs (*time range*) verglichen. So schätzt der Algorithmus, ob es sich um eine Reflexion handelt oder ob mehrere Reflexionen zeitlich dicht hintereinander auftreten. Treffen all diese Bedingungen zu und

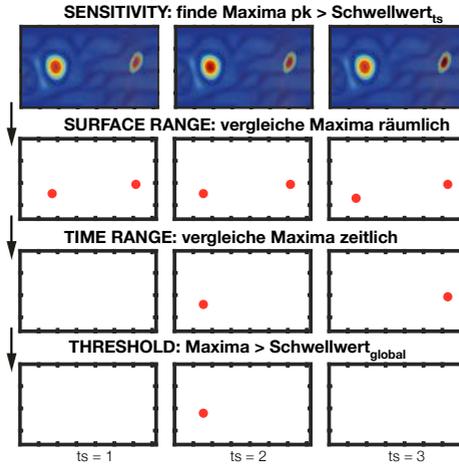


Abbildung 4.5: Schematische Darstellung des Algorithmus zur Detektion von Maxima aus Stade und Arend (2017). Bei Überschreiten eines Schwellwertes werden potenzielle Maxima pro Block ($ts = 1, 2, 3$) detektiert und mit benachbarten Intensitäten räumlich und zeitlich verglichen. Überschreitet der Kandidat auch einen globalen Schwellwert, so wird er als valides Maximum in einer Liste abgelegt.

überschreitet zudem die Intensität des Abtastpunktes auch einen globalen Schwellwert (*threshold*), der sich auf die normalisierte Energieabklingkurve bezieht, so wird der potenzielle Kandidat als valides Maximum und somit als detektierte Reflexion $pk(n)$ in einer Liste der gerichteten Komponenten abgelegt. Dabei gibt die Orientierung des Abtastpunktes auf dem Gitter Auskunft über die Einfallsrichtung $Az_{pk(n)}$ und $El_{pk(n)}$ der Reflexion. Der Zeitblock in Abhängigkeit von Blocklänge und Abtastfrequenz kennzeichnet wiederum den Zeitpunkt $\tau_{pk(n)}$ der gerichteten Komponente, die Intensität $dB_{pk(n)}$ deren Amplitude.

In Abbildung 4.6 findet sich eine exemplarische Darstellung von in der Intensitätsmatrix detektierten Maxima und deren Energie-Verlauf über die Zeit im Vergleich zu einer messtechnisch erfassten Referenz-RIR.

Verifikation

Anhand eines systematischen Versuchsaufbaus wurde die Funktionsweise der Detektion der gerichteten Schallfeldkomponenten verifiziert. Dazu wurde ein Mikrofonarray-

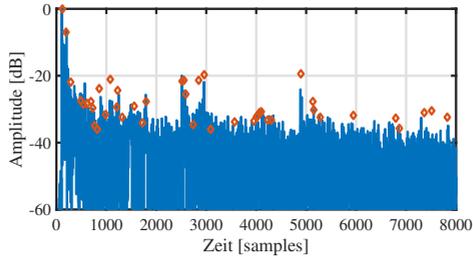


Abbildung 4.6: Exemplarische Darstellung detektierter Maxima im Zeit- und Energieverlauf (rot) im Vergleich zur Referenz-RIR.

Datensatz mit 110 Abtastpunkten auf einer Lebedev-Quadratur (schallharte Kugel mit Radius $r_0 = 8,75$ cm) verwendet, der vier räumlich verteilte und zeitlich verzögerte Schallquellen in reflexionsfreier Umgebung aufzeichnete¹. Diese Schallquellen simulieren vier diskrete Reflexionen, die es im Schallfeld zu detektieren galt. Alle Quellen befanden sich auf der selben vertikalen Ebene mit $\theta = 0^\circ$ und hatten die identische Entfernung (2,5 m) zum Mittelpunkt des Mikrofonarrays. Die Orientierungen auf der horizontalen Ebene betragen $\phi = 45^\circ, 15^\circ, 345^\circ$ und 315° mit den Verzögerungszeiten 0 ms, 16 ms, 32 ms und 48 ms.

Die zuvor erläuterte Signalverarbeitungs-Kette wurde auf diesen Datensatz angewendet, sodass die räumlich und zeitlich aufgelöste Intensitätsmatrix des Schallfeldes erzeugt und anschließend nach Maxima durchsucht wurde. Die Blocklänge für die Intensitätsmatrix betrug 24 Samples, daraus resultierte bei der verwendeten Abtastfrequenz $f_s = 44,1$ kHz eine zeitliche Auflösung von 0,54 ms. Die Zerlegung des Schallfeldes in ebene Wellen erfolgte mit einem Lebedev-Gitter mit 3074 Punkten, einer Frequenz $f_{\text{pwd}} = 4823$ Hz, der Ordnung $N = 5$ und einer modalen Verstärkung von +10 dB für die Radialfilter. Der Algorithmus zur Klassifizierung der gerichteten Anteile detektierte gemäß Tabelle 4.1 vier Maxima im Schallfeld. Es zeigte sich eine genaue räumliche und zeitliche Ortung der vier Reflexionen im Schallfeld. Die geringen Ungenauigkeiten sind auf die limitierte Zeit- und Ortsauflösung des Systems zurückzuführen. Für in der Praxis nutzbare Konfigurationen des Messaufbaus als auch des Analyse-Systems lassen sich diese Abweichungen jedoch nicht sinnvoll und effektiv eliminieren, sodass im Folgenden weiterhin mit diesen Konfigurationen gearbeitet wird.

¹siehe http://audiogroup.web.th-koeln.de/SOFiA_wiki/AE5.html

Tabelle 4.1: Verifikation: Detektierte Maxima $p_k(n)$ des Aufbaus mit Zeitpunkt t , Orientierung Az und $E1$ und Amplitude dB .

n	t_{pk} [ms]	Az_{pk} [°]	$E1_{pk}$ [°]	dB_{pk} [dB]
1	0,5	48,2	2,3	-2,1
2	16,9	13,7	0	-3
3	32,7	346,3	0	0
4	49,0	315,0	0	0,7

4.1.1.2 Generierung gerichteter Impulsantworten

Für eine synthetische Generierung der gerichteten Schallfeldanteile zur adäquaten Auralisation fehlt neben den zuvor bestimmten Merkmalen – Zeitpunkt des Eintreffens am Hörort, Einfallswinkel und Amplitude – noch die spektrale Komponente. Abhängig der verwendeten Materialien im Raum und deren frequenzabhängigem Absorptionsvermögen, besitzt jede gerichtete Komponente eine eigene spektrale Färbung. So weist z.B. eine Reflexion an einer schallharten Betonfläche einen anderen spektralen Verlauf als ein Rückwurf an einer Holzwand auf. Das Spektrum des Direktschalls als gerichtete Komponente mit dem höchsten Pegel ist in der Regel ebenso individuell und unterschiedlich im Vergleich zu späteren Reflexionen. Diese spektralen Eigenschaften sollen in der angestrebten Synthese berücksichtigt werden und daher muss der Frequenzverlauf für jede gerichtete Komponente separat bestimmt werden.

Dazu können mittels einer Zerlegung in ebene Welle, fokussierte Impulsantworten in Richtung der detektierten Einfallrichtungen gebildet werden. In der Literatur findet sich hierfür die Formulierung des „akustischen Teleskops“ (Möser, 2009, S.365), die diese Vorgehensweise treffend beschreibt. Weiterhin wird dieser Verarbeitungsschritt mit einer zeitlichen Fensterfunktion kombiniert, sodass *räumlich gerichtete Impulsantworten* (DIR) für einzelne zeitliche Blöcke berechnet werden. Abbildung 4.7 zeigt schematisch den Aufbau der Berechnung der gerichteten Impulsantworten für jedes im vorangegangenen Verarbeitungsschritt detektierte Maximum des Schallfeldes.

Das System benötigt zur Erzeugung der DIRs zum einen den identischen Mikrofonarray-Datensatz, der zuvor zur Bildung der dreidimensionalen Intensitätsmatrix verwendet wurde und zum anderen die Liste der detektierten Reflexionen $p_k(n)$, die der Algorithmus zur Klassifizierung der Maxima erstellt. Die nachfolgend erläuterte Signalverarbeitung erfolgt in einer Schleife und für jeden Eintrag der Liste identisch. Im ersten Schritt wird eine auf die Ankunftszeit t_{pk} zentrierte Fensterfunktion mit der Länge $\pm\Delta t_w/2$ auf die Impulsantworten angewendet, sodass nur ein definierter Zeitbereich der Länge t_w weiterverarbeitet wird. Dieser Block wird mittels zeitlicher und anschließender räumlicher Fourier-Transformation in das sphärische Wellenspektrum überführt.

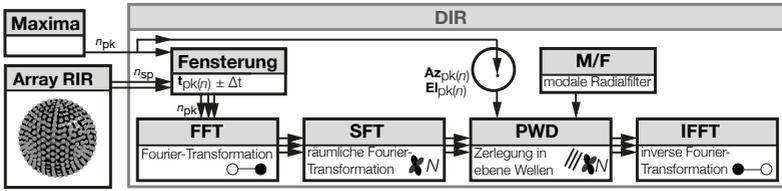


Abbildung 4.7: Schematische Darstellung der Bildung gerichteter Impulsantworten (DIR) aus Stade et al. (2017b): Die Impulsantworten eines Mikrofonarray-Datensatzes werden entsprechend der detektierten Ankunftszeiten $t_{pk(n)}$ der gerichteten Anteile zeitlich gefenstert und in die räumliche Domäne transformiert. Auf Basis der räumlichen Fourier-Koeffizienten wird für jedes Maximum eine ebene Welle mit der jeweiligen Orientierung $Az_{pk(n)}$ und $El_{pk(n)}$ gebildet und mittels inverser Fourier-Transformation in den Zeitbereich überführt.

Im nächsten Schritt erfolgt auf Basis der räumlichen Fourier-Koeffizienten abermals eine Zerlegung in ebene Wellen. Entgegen der Signalverarbeitung zur Erzeugung der Intensitätsmatrix (siehe Abschnitt 4.1.1.1) werden hierbei jedoch nicht zahlreiche ebene Wellen entsprechend eines sphärischen Rasters, sondern pro Maximum $pk(n)$ jeweils ausschließlich eine ebene Welle mit dessen Orientierung $Az_{pk(n)}$ und $El_{pk(n)}$ gebildet. Die PWD führt bei Verwendung entsprechender Ordnungen zu einer hohen räumlichen Richtwirkung. Anschließend werden die Komponenten D der ebenen Welle im Frequenzbereich mittels inverser Fourier-Transformation zurück in den Zeitbereich überführt. Der Ausgang dieser Verarbeitungsschritte ist für jedes Maxima $pk(n)$ eine zeitlich segmentierte, gerichtete Impulsantwort $DIR(n)$. Die Kombination von Fensterfunktion und hoher Richtwirkung führt zu einer auf die jeweilige Reflexion räumlich und zeitlich fokussierte Impulsantwort und effektiver Dämpfung der übrigen Schallfeldanteile. Diese Vorgehensweise kann mit der *In-Situ*-Methode (Brandão et al., 2011) verglichen werden, die in der akustischen Messtechnik zur Bestimmung von Absorptions-Koeffizienten von Materialien Anwendung findet.

Aufgrund der großen energetischen Differenz zwischen Direktschall und den restlichen Schallfeldanteilen, kann die Fensterung des absoluten Maximums vernachlässigt werden, da das Energieverhältnis bereits eine effektive Fokussierung auf den Direktschall bewirkt. So ist es auch weiterhin möglich, tieffrequente Anteile des Direktschalls genau abbilden zu können.

4.1.1.3 FIR-Filter für die gerichteten Komponenten

Auf Basis der zuvor berechneten gerichteten Impulsantworten wird für jedes Maximum separat ein digitales Filter mit endlicher Impulsantwort (FIR) gebildet. Dieses Filter beschreibt den jeweiligen Frequenzgang der DIRs und charakterisiert somit deren frequenzabhängige Filterwirkung. Somit können diese FIR-Filter in der Synthese verwendet werden, um den künstlichen, gerichteten Komponenten den entsprechenden Frequenzgang der realen Reflexionen und des Direktschalls aufzuprägen. In der weiteren Betrachtung werden die Filter mit $\text{FIR}_{\text{pk}(n)}$ bezeichnet.

Entsprechend der beabsichtigten Verwendung und Wirkung können unterschiedliche Filterarten von Nutzen sein. In der vorliegenden Arbeit werden daher zwei verschiedene Filtertypen für diesen Verarbeitungsschritt benutzt. Zum einen können minimalphasige FIR-Filter mit Von-Hann-Fenster gebildet werden. Diese Filter vermeiden wirksam sogenannte *Pre-Ringing* Artefakte bei tiefen Frequenzen sowie weisen keine Gruppenlaufzeit auf, sodass keine zusätzliche Latenz erzeugt wird. Weiterhin können auch linearphasige FIR-Filter mit konstanter, frequenzunabhängiger Gruppenlaufzeit (ebenfalls mit Von-Hann-Fenster) auf Basis der DIRs gebildet werden. Da die Filter ausschließlich zur Adaption des Frequenzgangs verwendet werden, erfolgt eine Normalisierung, sodass die Filter keine Änderungen im Pegel hervorrufen.

4.1.2 Synthese der gerichteten Komponenten

Die Synthese der gerichteten Komponenten erfolgt auf Basis eines sphärischen HRIR-Datensatzes und der Struktur von gerichteten Komponenten, wie sie bereits in der Analyse Anwendung fand. So müssen die gerichteten Komponenten nur diesem Aufbau folgen, aber nicht einer Analyse von Messdaten entspringen, sondern können ebenfalls beliebig erzeugt werden. Die Struktur muss für jede gerichtete Komponente $\text{pk}(n)$ deren Orientierung $A_{z_{\text{pk}(n)}}$ und $E_{l_{\text{pk}(n)}}$, Ankunftszeit $\tau_{\text{pk}(n)}$, Pegel $dB_{\text{pk}(n)}$ und das gewünschte Spektrum $\text{FIR}_{\text{pk}(n)}$ enthalten. Das System erzeugt daraus eine synthetische, binaurale Raumimpulsantwort, die ausschließlich aus einzelnen, diskreten Impulsen besteht. Diese kann mittels Faltung mit einem nachhallfreien Audiosignal zu Auralisationszwecken verwendet werden und führt bei der Kopfhörer-Wiedergabe zu räumlich verteilten Reflexionen entsprechend des beabsichtigten räumlichen Echogramms. Das Blockschaltbild in Abbildung 4.8 zeigt schematisch die Verarbeitungsschritte der Synthese der gerichteten Komponenten.

Im ersten Schritt wird gemäß der Einfallswinkel $A_{z_{\text{pk}(n)}}$ und $E_{l_{\text{pk}(n)}}$ der gerichteten Komponente die zugehörige HRIR mit identischer Orientierung aus einem HRIR-Datensatz ausgewählt, der auf einem sphärischen Abtastgitter erfasst wurde. Anschlie-

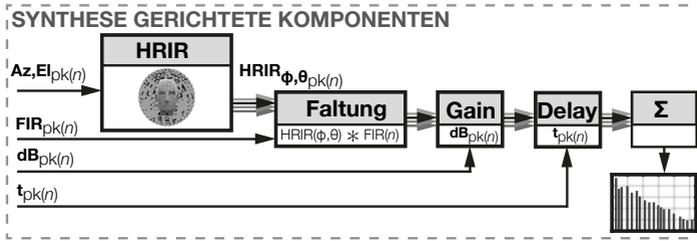


Abbildung 4.8: Schematische Darstellung der Synthese der gerichteten Komponenten (Ausschnitt aus Parameter-Modell): Gemäß der Orientierung der gerichteten Komponente wird die entsprechende HRIR aus einem sphärischen Datensatz ausgewählt und mit dem FIR-Filter gefaltet, im Pegel angepasst und verzögert. Anschließend werden alle Impulse aufsummiert.

end werden die Signale mit dem entsprechenden Filter $FIR_{pk(n)}$ gefaltet, sodass die HRIR den spektralen Verlauf der gerichteten Komponente erhält. Werden linear- und minimalphasige Filter kombiniert zur Anpassung der Signale verwendet, so muss deren unterschiedliche Latenz beachtet und entsprechend korrigiert werden. Gemäß des Faktors $dB_{pk(n)}$ wird das Signal abgeschwächt und so der Pegel angepasst. Eine Verzögerung um $\tau_{pk(n)}$ verschiebt die Signale anschließend zeitlich an den entsprechenden Auftretzeitpunkt der gerichteten Komponente. Diese Signalverarbeitung wird für alle Maxima wiederholt und abschließend werden alle Signale summiert, sodass eine synthetische BRIR gemäß des Echogramms erzeugt wird. In Abbildung 4.9 ist das Ergebnis beispielhaft dargestellt.

Um eine dynamische Auralisation zu ermöglichen, muss eine neue SBRIR entsprechend der gewünschten Auflösung und des Freiheitsgrades für jeden Orientierungswinkel erzeugt werden. Die Signalverarbeitung ist identisch wie zuvor erläutert, lediglich in die Auswahl der HRIR gemäß des Einfallswinkels muss der Versatz auf der entsprechenden Ebene mit einfließen. Dies ergibt einen Datensatz, der für die dynamische Binauralsynthese verwendet werden kann und eine ortsfeste Auralisation der gerichteten Schallfeldanteile ermöglicht. Dabei behalten die gerichteten Komponenten ihre räumliche Anordnung auch bei Änderungen des Orientierungswinkels des Rezipienten. In dieser Arbeit werden nur Änderungen auf der horizontalen Ebene betrachtet. Die Generierung von sphärischen Datensätzen, die weitere Freiheitsgerade (*tilt* und *roll*) bereitstellen, sind aber ebenfalls möglich.

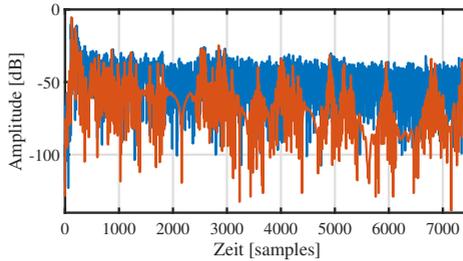


Abbildung 4.9: Beispielhafte Darstellung einer RIR aus synthetischen, gerichteten Komponenten (rot) im Vergleich zur messtechnisch erfassten Referenz-RIR (blau).

4.2 Diffuse Komponenten des Schallfeldes

Da in der vorliegenden Arbeit angestrebt wird, die menschliche Wahrnehmung ins Zentrum aller entwickelter Algorithmen zu rücken, wird in diesem Verarbeitungsschritt nicht das physikalisch diffuse Schallfeld (wie in Abschnitt 2.1.2 definiert) betrachtet. Vielmehr werden mit den *diffusen Komponenten* alle Anteile im Schallfeld, die für den Hörer keine wahrnehmbaren Richtungsinformationen beinhalten, verstanden. Somit ist keine Lokalisation von Schallquellen und/oder Reflexionen im diffusen Anteil möglich. Dadurch ergeben sich Möglichkeiten zur substanziellen Reduktion des Aufwandes und dem benötigten Speicherbedarf zur Synthese mittels Parametern.

Das Ziel dieses Verarbeitungsschrittes ist die Generierung eines binauralen Signalpaares, welches der menschlichen Wahrnehmung des diffusen Schallfeldes im jeweiligen Raum entspricht. Des Weiteren soll das Signal zeitlich mit der Referenz-BRIR übereinstimmen und ein möglichst identisches Frequenzspektrum sowie vergleichbare interaurale Eigenschaften aufweisen. Es handelt sich um einen hörerzentrierten Ansatz; es wird versucht nur die Komponenten zu detektieren, die für das Hörereignis relevant sind und anschließend ausschließlich diese zur Synthese zu verwenden. Daher lässt sich das Vorgehen gut mit dem Grundgedanken der *Irrelevanzreduktion* (wie z.B. in mp3 angewendet, siehe ISO/IEC 14496-3 (1993)) vergleichen: Für das auditive System nicht wahrnehmbare Informationen können verworfen werden und müssen nicht für die Synthese verwendet werden. Diese Vorgehensweise soll SBRIRs erzeugen, die im Idealfall in der Auralisation bei Faltung mit einem Testsignal zu identischen Hörereignissen führen, wie die Faltung des Testsignals mit entsprechenden gemessenen Referenz-BRIRs.

In der Literatur finden sich zahlreiche Publikationen zur synthetischen Erzeugung von

diffusum Nachhall (siehe Abschnitt 1.1). Diese Arbeit verfolgt einen Ansatz, den Moorer bereits 1979 vorstellte, indem er folgendermaßen anmerkte:

While digitizing the impulse responses from concert halls around the world, we kept noticing that the responses in the finest concert halls sounded remarkably similar to white noise with an exponential amplitude envelope.²

Dabei ist die grundsätzliche Idee, den Energieabfall frequenzabhängig zu bestimmen und diesen Verlauf einem weißen Rauschsignal aufzuprägen (Jot et al., 1997). Dies kann in der einfachsten Form mittels einer Multiplikation mit der Nachhallzeit (bei logarithmischer Darstellung beschrieben durch eine abfallende Gerade) erfolgen.

Kombiniert wird der Ansatz von Moorer und Jot in dieser Arbeit zusätzlich mit der Anpassung der *interauralen Kohärenz* (IC) nach Borß und Martin (2009). Da die Kohärenz von zwei Rauschsignalen frequenzunabhängig bei ≈ 0 liegt, erzeugt die Synthese mittels weißem Rauschen jedoch bei Faltung dieser Impulsantwort mit einem Testsignal zur Hörbarmachung mittels Kopfhörer zwei getrennte, laterale Hörereignisse (siehe Abbildung 2.5, unten rechts) anstelle eines umhüllenden diffusen Nachhalls. Wie in Cook (1955) gezeigt, entspricht die Kohärenz von binauralen Signalen im idealen Diffusfeld einer *si-Funktion*, sodass die Ohrsignale zu tiefen Frequenzen stärker korrelieren. Daher ist zusätzlich die frequenzabhängige Adaption der Kohärenz des Rauschsignals entsprechend des idealen diffusen Schallfeld wie von Menzer und Faller (2009) gezeigt, zur Erzeugung adäquater synthetischer BRIRs notwendig.

4.2.1 Analyse der diffusen Komponenten

Das Blockdiagramm in Abbildung 4.10 zeigt den grundlegenden Aufbau der Analyse der diffusen Komponenten. Die Signalverarbeitung unterteilt sich in die frequenzabhängige Bestimmung und Approximation der *Energieabklingkurven* (EDC) und der mittleren Energie sowie in die Kalkulation und Approximation der interauralen Kohärenz. Alle Bestandteile werden im Folgenden getrennt voneinander erläutert. Das System wurde in dieser Form von Stade und Arend (2016a,b) vorgestellt sowie perzeptiv evaluiert.

4.2.1.1 Energieabklingkurven

Die Energieabklingkurven einer gemessenen Referenz-BRIR müssen bestimmt und mit Hilfe einer Ausgleichsrechnung geeignet approximiert werden. So ergeben sich geglä-

² *eigene Übersetzung:* Während der Erfassung von Impulsantworten von Konzertsälen aus aller Welt beobachteten wir, dass die Impulsantworten klanglich sehr weißem Rauschen mit einer exponentiellen Abklingkurve ähnelten.

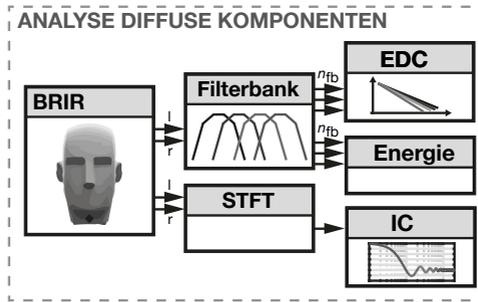


Abbildung 4.10: Schematische Darstellung der Analyse der diffusen Komponenten (Ausschnitt aus Parameter-Modell). Die Referenz-BRIR wird in einzelne Frequenzbänder zerlegt und pro Band werden die Energieabklingkurven approximiert und deren mittlere Energie bestimmt. Weiterhin erfolgt eine Bestimmung und Approximation der interauralen Kohärenz mit Hilfe einer STFT.

tete Kurvenverläufe der Impulsantworten, die inhärent die Nachhallzeit repräsentieren. Zudem wird die mittlere Energie ermittelt, um eine Pegel-Anpassung der SBRIR an die Referenz-BRIR sowie zwischen gerichteten und diffusen Schallfeldanteilen zu ermöglichen. Als Eingangsgröße erhält dieses Modul idealerweise eine binaurale Impulsantwort, die im zu synthetisierenden Raum mit einem Kunstkopf erfasst wurde. Alternativ können auch zwei, entsprechend des Ohrabstandes, räumlich verteilte Mikrofonsignale oder auch omnidirektionale RIRs verarbeitet werden. Unter der Annahme, dass im Diffusfeld keine Lokalisation von Schallquellen und/oder Reflexionen möglich ist, können hierfür Impulsantworten mit beliebiger Blickrichtung ausgewählt werden.

Die Impulsantworten werden zunächst für die Weiterverarbeitung angepasst. Es erfolgt eine Normalisierung und zeitliche Kürzung der Signale. Dabei werden alle Samples bis zum Überschreiten eines gewählten Schwellwertes (*threshold onset-detection*) entfernt. Diese zeitliche und energetische Anpassung der Impulsantworten ist nötig, damit die Analyse von der Konfiguration der Messung unabhängige Ergebnisse liefert. Zur besseren Übersicht wird exemplarisch die Approximation der Energieabklingkurven und deren Energiebestimmung ausschließlich für eine Impulsantwort erläutert und in Abbildung 4.11 dargestellt. Für die frequenzabhängige Analyse werden beide Kanäle der BRIR mittels einer Filterbank zerlegt, das weitere Vorgehen pro Frequenzband ist identisch wie im Folgenden vorgestellt.

Die Ausgleichsrechnung (oder Kurvenanpassung) verfolgt das Ziel, die Impulsantworten mithilfe einer einfachen mathematischen Funktion möglichst genau nachzubilden und

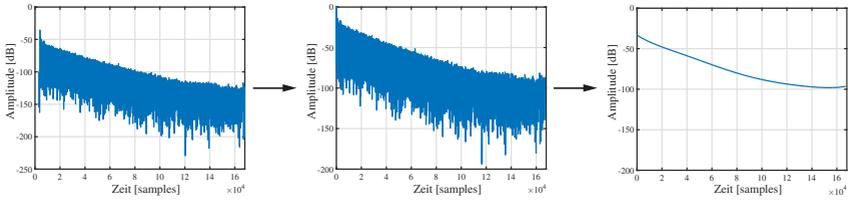


Abbildung 4.11: Prinzip der Bestimmung der Energieabklingkurven: Eine BRIR (links) wird zeitlich und energetisch angepasst (Mitte) und anschließend dessen Energieabfall mittels Polynomfunktion approximiert (rechts).

mittels einzelner Variablen zu schätzen (Puate León und Kiencke, 2011, S.25-52). Dieses Verfahren wird oft als *Fitting* bezeichnet. Für die Bestimmung der Abklingkurven wird der Weg der Ausgleichsrechnung anstelle der Rückwärtsintegration nach Schroeder (1965) gewählt, da die Schätzung inhärent eine Datenreduzierung ermöglicht und zudem die Präzision der Kurvenanpassung variabel ist. Es existieren verschiedene Ansätze und Möglichkeiten zur Ausgleichsrechnung, in dieser Arbeit wird die Interpolation durch Polynome verwendet. Dabei wird der Kurvenverlauf mittels einer Polynomfunktion des Grades p_n folgendermaßen beschrieben:

$$f(x) = a_{p_n} \cdot x^{p_n} + a_{p_n-1} \cdot x^{p_n-1} + \dots + a_2 \cdot x^2 + a_1 \cdot x + a_0 \quad (4.1)$$

Der Polynomgrad p_n bestimmt dabei durch $p_n - 2$ die maximale Anzahl der möglichen Wendepunkte der Polynomfunktion. Das bedeutet, die Wahl des Polynomgrades hat direkten Einfluss auf die mögliche Genauigkeit der Approximation. So ergibt eine Polynomfunktion ersten Grades eine einfache Gerade als Tangente an die Abklingkurve, während mit höheren Polynomgraden auch komplexere Verläufe möglich sind. Insbesondere bei Änderungen der Neigung der Abklingkurve über der Zeit, die bei Unterschieden zwischen EDT und T z.B. durch gekoppelte Räume entstehen können, kann ein höherer Polynomgrad sinnvoll sein, um den Kurvenverlauf hinreichend genau nachbilden zu können. Es muss beachtet werden, dass hohe Polynomgrade vor allem an den Endpunkten der Abklingkurven zu Überschwingern führen können und auch der allgemeine Verlauf schneller oszilliert. Den Einfluss verschiedener Polynomgrade auf die Kurvenanpassung einer RIR zeigt Abbildung 4.12. Aufgrund der Glättung durch das Fitting, entstehen insbesondere für den frühen Abschnitt der Impulsantwort Ungenauigkeiten, sodass der Amplitudenverlauf früher Reflexionen nicht nachgebildet werden kann. Da dieses Verfahren jedoch ausschließlich zur Approximation des diffusen Schallfeldverlaufs eingesetzt wird, ist der Aspekt der Glättung nicht nur vernachlässigbar,

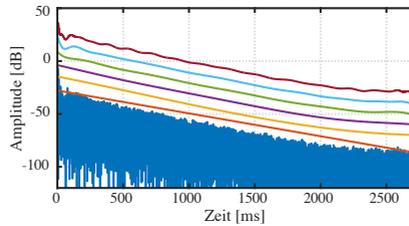


Abbildung 4.12: Einfluss des Polynomgrads $p_n = 1$ (rot), 3 (gelb), 6 (lila), 12 (grün), 24 (hellblau) und 48 (dunkelrot) auf die Kurvenanpassung einer RIR (blau) mittels Polynomfunktion; zur besseren Lesbarkeit wurden die Kurven vertikal verschoben.

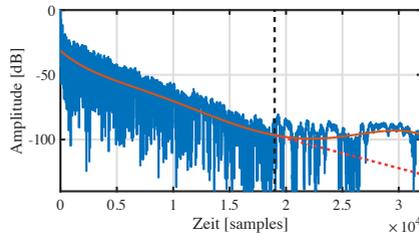


Abbildung 4.13: SNR-Erhöhung durch Extrapolation (rot gestrichelt) der Polynomkurve (rot) ab Beginn des Grundrauschens (schwarz gestrichelt).

sondern sogar förderlich. So werden starke Reflexionen in der Approximation ausgespart und ein gleichmäßiger Energieabfall ohne Sprungstellen wird erzeugt.

Weiterhin wird der Zeitpunkt bestimmt, an dem die Impulsantwort in das Grundrauschen übergeht. Dazu wird ein iterativer Algorithmus nach Lundeby et al. (1995) verwendet, der den Pegel des Grundrauschens detektiert. So kann die mittlere Energie des diffusen Nachhalls zu einem definiertem Zeitpunkt (z.B. 10 dB über dem Grundrauschen) berechnet werden. Es wird dadurch gewährleistet, dass die Energie-Berechnung abhängig vom vorhandenen *Signal-Rauschabstand* (SNR), möglichst spät (und damit in der Regel auch im Diffusfeld) erfolgt. Die Bestimmung der mittleren Energie ermöglicht eine genaue Pegelanpassung zwischen gerichteten und diffusen Komponenten und wird daher in einer eigenen Variable abgelegt.

Des Weiteren erlaubt dieses Vorgehen durch Bestimmung des Schnittpunktes von Grundrauschen und der Abklingkurve eine Expansion des SNRs. Die Polynomfunktionen können ab dem Schnittpunkt dahingehend korrigiert werden, dass die Kurven nicht dem

gemessenen Verlauf folgen, sondern entsprechend des ursprünglichen Abfalls verlängert werden (siehe Abbildung 4.13). So wird gewährleistet, dass die approximierten Energieabklingkurven auch jenseits des gemessenen Grundrauschens gleichmäßig abnehmen, wodurch später synthetische BRIRs mit hohem SNR möglich werden (entsprechend der Auflösung des Systems). Weiterhin können damit auch Referenz-BRIRs mit ursprünglich für eine adäquate Auralisation unzureichendem SNR analysiert werden und anschließend deren Eigenschaften zur Synthese und Hörbarmachung verwendet werden. Ein ähnlicher Ansatz findet sich auch in Dietrich et al. (2013).

Filterbank

Aufgrund des frequenzabhängigen Absorptionsverhaltens der im Raum verwendeten Materialien und der Dämpfung zu hohen Frequenzen durch die Luftabsorption, stellen sich in Räumlichkeiten frequenzabhängige Unterschiede der Energieabklingkurven ein. Insbesondere in großen Räumlichkeiten ist typischerweise ein starker Abfall der Nachhallzeiten zu hohen Frequenzen und damit auch steilere Energieabklingkurven zu verzeichnen.

Daher ist es notwendig, die Energieabklingkurven als auch die mittlere Energie und den Beginn des Grundrauschens frequenzabhängig zu bestimmen. Die BRIR muss zunächst in einzelne Frequenzbänder aufgespalten werden, um dann die zuvor erläuterte Vorgehensweise auf jedes einzelne Band anwenden zu können. Klassische Bandpassfilter in Oktav- oder Terzband Auflösung sind ebenso denkbar, wie auch eine auditorisch motivierte Gammatone-Filterbank. Um die frequenzabhängigen Unterschiede der Abklingkurven genau bestimmen zu können, ist eine ausreichende Auflösung mittels schmaler Frequenzbänder vonnöten. Da die Unterschiede zu hohen Frequenzen zudem oftmals größer werden (siehe z.B. T in Abbildung 3.5), kann eine alternative Frequenzauflösung sinnvoll sein.

In dieser Arbeit wird daher eine mehrphasige Filterbank nach Lubberhuizen (2010) verwendet, die eine nahezu ideale Rekonstruktion bei flexibler Frequenztrennung ermöglicht. Die Filterbank bietet eine hohe Kanaltrennung und geringe Rekonstruktionsfehler (mittlerer Fehler = -300,14 dB, siehe Abbildung 4.14 a). Dadurch ergibt sich für ein weißes Rauschsignal bei Aufspaltung und anschließender Zusammensetzung mittels dieser Filterbank ein nahezu identisches Spektrum (siehe Abbildung 4.14 b). Zudem können Breite und Mittenfrequenz der Frequenzbänder variabel gewählt werden. Dank der hohen Kanaltrennung der Filterbank kann die Frequenz-Aufspaltung in beliebigen Frequenz-Bereichen erfolgen, sodass auch flexible Intervalle mit unterschiedlicher Bandbreite möglich sind. Nach der Zerlegung in die einzelnen Frequenzbänder können dann

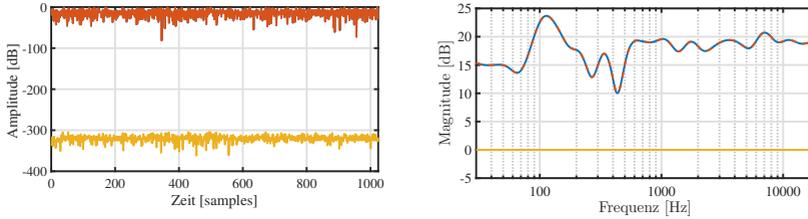


Abbildung 4.14: Analyse Filterbank mit Rauschsignal: Eingang (blau), Ausgang (rot) und Differenz (gelb) im Zeit- (links) und Frequenzbereich (rechts).

in den gefilterten Impulsantworten die Abklingkurven frequenzabhängig approximiert, sowie deren Grundrauschpegel und mittlere Energie bestimmt werden.

4.2.1.2 Interaurale Kohärenz

Die interaurale Kohärenz (IC) ist eine wichtige Eigenschaft für die binaurale Wahrnehmung diffuser Schallfelder und soll daher bei der Synthese der diffusen Komponenten beachtet werden. Im ersten Schritt muss die interaurale Kohärenz in einer gemessenen BRIR frequenzabhängig bestimmt werden. Ebenfalls kann die Kohärenz von zwei räumlich (idealerweise im Ohrabstand d) verteilten Mikrofonsignalen berechnet werden. Nach Formel 2.8 ist die Kalkulation des Verlaufs der interauralen Kohärenz über der Frequenz durch Bildung der Fouriertransformierten S_l und S_r des linken und rechten Ohrsignals möglich. Gemäß Menzer und Faller (2010) wird in der Praxis zur Bestimmung der interauralen Kohärenz $\Phi(i)$ eine *Kurzzeit-Fourier-Transformation* (STFT) mit K Blöcken auf die BRIR angewendet:

$$\Phi(i) = \frac{|\sum_{k=1}^K S_l(i, k) S_r(i, k)^*|}{\sqrt{\sum_{k=1}^K |S_l(i, k)|^2 \sum_{k=1}^K |S_r(i, k)|^2}} \quad (4.2)$$

Dabei bezeichnen i und k die Indizes von Frequenz und Zeit. Erneut kann die Kurvenanpassung mittels Polynomfunktion verwendet werden, um den Verlauf der interauralen Kohärenz über der Frequenz zu approximieren. Zwar ist auch eine sample-genaue Speicherung der IC-Kurve möglich, im Hinblick auf ein datenreduzierendes Parametermodell soll aber versucht werden, den Kurvenverlauf mit wenigen Stützstellen nachzu-

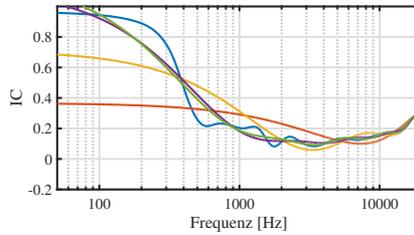


Abbildung 4.15: Einfluss des Polynomgrads $p_n = 3$ (rot), 6 (gelb), 12 (lila), 16 (grün) auf die Kurvenanpassung des IC-Verlaufs (blau) mittels Polynomfunktion.

bilden. Im Gegensatz zu den Energieabklingkurven zeigt sich hierbei eine komplexere Funktion mit mehr Wendepunkten, daher ist ein höherer Polynomgrad nötig, um eine hinreichend genaue Approximation des IC-Verlaufs zu erreichen. Abbildung 4.15 zeigt den Einfluss verschiedener Polynomgrade auf die Genauigkeit der Reproduktion der Kurve mittels einer Polynomfunktion. Dabei führen in der Praxis geringe Polynomgrade $p_n < 9$ zu unzureichenden Ergebnissen.

Wie in Abbildung 2.7 aufgezeigt, ist die Bestimmung der interauralen Kohärenz stark abhängig von der Integrationszeit. Da in diesem Verarbeitungsschritt die diffusen Komponenten analysiert werden, wird für die IC-Berechnung nur der Teil der Impulsantwort nach der theoretischen mixing time des Raumes verwendet. So wird sichergestellt, dass der IC-Verlauf ausschließlich im diffusen Nachhall bestimmt wird und nicht in Zeitbereichen mit starken Reflexionen oder gar unter Anwesenheit von Schallfeldanteilen des Direktschalls. Es kann zudem für die Bestimmung der interauralen Kohärenz ausschließlich eine Blickrichtung verwendet werden, da im diffusen Nachhall die Kohärenz nicht mehr von der Orientierung abhängt (siehe Abbildung 2.7 c).

4.2.2 Synthese der diffusen Komponenten

Weißes Rauschen kann wirksam als artifizieller Nachhall zur Auralisation von Räumen verwendet werden, indem ein exponentieller Energieabfall entsprechend der frequenzabhängigen Nachhallzeit dem Rauschsignal aufgeprägt wird. Auch der in dieser Arbeit verwendete Ansatz basiert auf Rauschsignalen und deren Adaption zur Synthese der diffusen Schallfeld-Komponenten. Der grundlegende Aufbau des Systems findet sich in Abbildung 4.16. Der Gebrauch von weißem Rauschen zur Synthese bietet verschiedene Vorteile: Zum einen kann der Nachhall breitbandig und ohne Begrenzung im Frequenzbereich synthetisiert werden, sodass beliebige spektrale Verteilungen möglich werden.

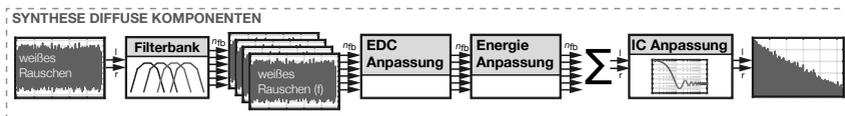


Abbildung 4.16: Schematische Darstellung der Synthese der diffusen Komponenten (Ausschnitt aus Parameter-Modell). Weißes Rauschen wird in Frequenzbänder zerlegt und pro Band wird der Energieabfall aufgeprägt sowie die Energie angepasst. Nach Summierung aller Frequenzbänder erfolgt die Anpassung der interauralen Kohärenz.

Zum anderen kann weißes Rauschen mittels einer Zufallsreihe leicht und mit beliebiger Länge erzeugt werden, So kann das benötigte Rauschsignal erst bei Bedarf entsprechend der benötigten Länge (abhängig der zu synthetisierenden Nachhallzeit) für den jeweiligen Raum prozessiert werden.

Gemäß der zuvor definierten Struktur der diffusen Komponenten im Analyse-Modul wird ein zwei-kanaliges weißes Rauschsignal entsprechend der frequenzabhängigen Energieabklingkurven, dem Faktor für die mittlere Energie und dem Verlauf der interauralen Kohärenz über der Frequenz angepasst und kann dann als diffuse Komponenten einer synthetischen BRIR verwendet werden. Die Komponenten müssen jedoch nicht der Analyse einer gemessenen BRIR entspringen, sondern können auch frei gewählt werden.

Die Betrachtungsweise der diffusen Komponenten ermöglicht für die in dieser Arbeit angestrebte Synthese mittels Parametern ein erhebliches Einsparungspotential. Lindau et al. (2012) zeigen, dass eine orts- und richtungsselektive BRIR-Synthese jenseits von $t > t_{mp}$ nicht notwendig ist, da Unterschiede zwischen verschiedenen Orientierungswinkeln nicht mehr wahrnehmbar sind. In Stade (2015) konnte zwar keine Ununterscheidbarkeit gezeigt werden, jedoch führten entsprechende Datensätze, die ab t_{mp} auf eine dynamische Nachführung der Kopf-Orientierung verzichteten, zu hohen Bewertungen auf der kontinuierlichen Qualitäts-Skala im Rahmen eines Hörversuchs mit MUSHRA (*Multi-Stimulus Test with Hidden Reference and Anchor*, ITU-R BS 1534-3 (2015)) Anordnung (siehe Abschnitt 5.2). Dabei zeigten spätere Übergangszeiten die größten Übereinstimmungen zwischen Messung und Bearbeitung. Dieser Sachverhalt kann ebenso für die Synthese der diffusen Komponenten konstruktiv genutzt werden, da auch hier keine Richtungsinformationen wahrnehmbar sein sollen. So kann die Synthese der diffusen Komponenten ausschließlich für eine Orientierung erfolgen. Das bedeutet, bei richtungs-sensitiver Auralisation mittels dynamischer Binauralsynthese auf Basis synthetischer BRIRs mit diffusen Komponenten wird für alle Blickrichtungen das identische binaurale Signalpaar verwendet. Dies ermöglicht im Gegensatz zum gemes-

senen Referenz-Datensatz, der abhängig von der verwendeten Auflösung, verschiedene BRIRs für die einzelnen Orientierungen bereitstellt, eine erhebliche Reduzierung der Rechenleistung sowie des Speicherbedarfs.

4.2.2.1 Adaption der Energieabklingkurven

Durch Bildung des *Hadamard*- (oder *Schur*-) Produkts (Horn, 1990) der Rauschsignale A mit den Kurvenverläufen B der Energieabklingkurven gemäß

$$A \circ B = a_{ij} \cdot b_{ij} = \begin{pmatrix} a_{11} \cdot b_{11} & \cdots & a_{1n} \cdot b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} \cdot b_{m1} & \cdots & a_{mn} \cdot b_{mn} \end{pmatrix} \quad (4.3)$$

können abklingende Rauschsignale erzeugt werden (siehe Abbildung 4.17). Um die Signale frequenzabhängig formen zu können, muss dieser Verarbeitungsschritt für jedes Frequenzband separat erfolgen. Dazu wird das Rauschsignal zuerst mit der identischen Filterbank prozessiert, die bereits für die Analyse der gemessenen Impulsantworten und Bestimmung der Energieabklingkurven Anwendung fand. Die gefilterten Rauschsignale $A(f)$ werden dann gemäß Formel 4.3 elementweise mit den dazugehörigen Polynom-Funktionen $B(f)$ der Abklingkurven multipliziert, um den Rauschsignalen den entsprechenden frequenzabhängigen Abfall aufzuprägen. Anschließend erfolgt eine frequenzabhängige Pegel-Anpassung der Signale mittels Multiplikation mit dem Faktor der mittleren Energie. Im letzten Verarbeitungsschritt werden alle Frequenzbänder aufsummiert, sodass sich ein breitbandiges Rauschsignal, welches entsprechend der gewünschten frequenzabhängigen Nachhallzeit abklingt, ergibt.

4.2.2.2 Adaption der interauralen Kohärenz

Aufgrund der stochastischen Verteilung der Werte eines zwei-kanaligen Rauschsignals, besteht keine Beziehung zwischen den Kanälen, sodass diese zueinander inkohärent sind. Dadurch ergibt sich bei der Bestimmung der Kohärenz solcher Signale ein frequenzunabhängiger Verlauf gemäß Abbildung 4.18 b, blaue Kurve, der gleichmäßig bei ~ 0 liegt. Wie zuvor gezeigt, ist dies jedoch für binaurale Signale kein üblicher Verlauf eines diffusen Schallfeldes (siehe Abbildung 2.6) und führt bei Verwendung entsprechender Impulsantworten zu klanglichen Beeinträchtigungen der Auralisation. Daher sollte die Kohärenz der Rauschsignale entsprechend der interauralen Kohärenz eines binauralen Signalpaares im Diffusfeld angepasst werden.

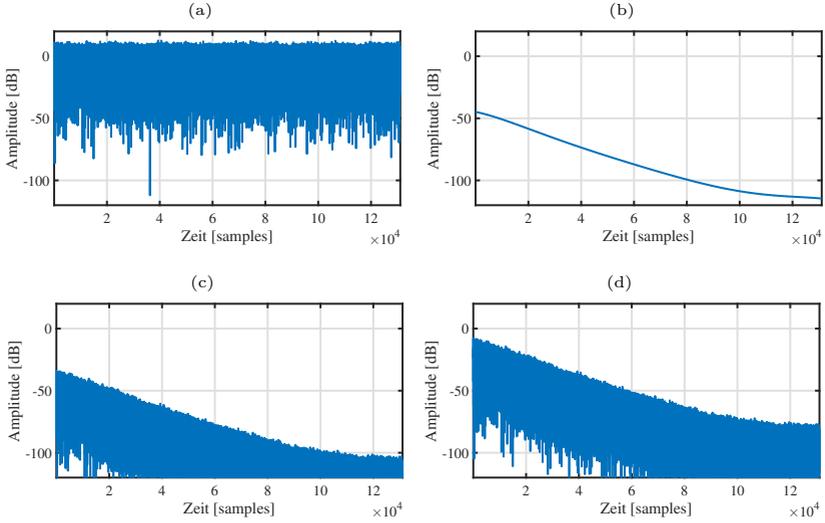


Abbildung 4.17: Hadamard-Produkt (c) aus Rauschsignal (a) und Polynomfunktion des Energie-Abfalls (b) mit Energieanpassung (d).

In dieser Arbeit wird ein an Menzer und Faller (2009) angelehntes Verfahren zur Anpassung der Kohärenz angewendet. Der Verlauf der interauralen Kohärenz Φ (siehe Formel 4.2) als Funktion über der Frequenz wird dazu verwendet. Mit Hilfe dieser Funktion können zur Kohärenz-Anpassung die Filter $u(n)$ und $v(n)$ berechnet werden. Diese sind im Frequenzbereich definiert mit:

$$\begin{aligned}
 U(w) &= \sqrt{\frac{1 + \Phi(w)}{2}} \\
 V(w) &= \sqrt{\frac{1 - \Phi(w)}{2}}
 \end{aligned}
 \tag{4.4}$$

Anschließend werden die unkorrelierten, abfallenden Rauschsignale $d_l(n)$ und $d_r(n)$ des linken und rechten Kanals linear im Zeitbereich mit den Filtern $u(n)$ und $v(n)$ gefaltet:

$$\begin{aligned}
 b_l(n) &= (u * d_l + v * d_r)(n) \\
 b_r(n) &= (u * d_l - v * d_r)(n)
 \end{aligned}
 \tag{4.5}$$

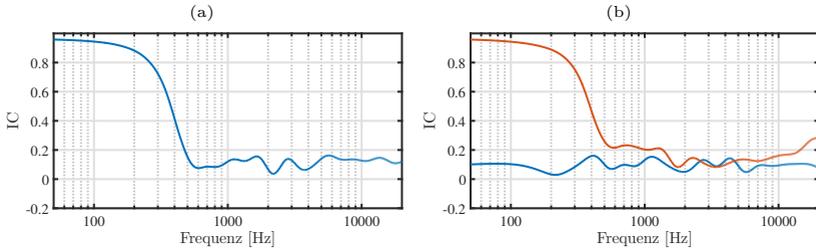


Abbildung 4.18: Verlauf der Kohärenz einer Referenz-BRIR (Diffusfeld, ab 250 ms) (a) und eines zwei-kanaligen, abklingenden Rauschsignals (b) vor (blau) und nach (rot) der Kohärenz-Anpassung gemäß Menzer und Faller (2010).

So ergeben sich zwei abfallende Rauschsignale $b(n)$, deren Kohärenz nun dem gewünschten Verlauf folgt. Abbildung 4.18 a zeigt beispielhaft die interaurale Kohärenz einer gemessenen Referenz-BRIR, 4.18 b den Verlauf der Kohärenz eines abfallenden Rauschsignals $d(n)$ vor (blau) und nach (rot) der Faltung mit den Filtern $u(n)$ und $v(n)$.

4.3 Zusammenfassung Signalverarbeitung für die BRIR-Synthese

In diesem Kapitel wurden Methoden der Signalverarbeitung zur Synthese binauraler Raumimpulsantworten vorgestellt und erläutert. Dabei betrachtet die Signalverarbeitung die Impulsantworten isoliert in zwei eigenständigen und unabhängigen Verarbeitungsketten und trennt das Schallfeld in gerichtete und diffuse Komponenten. Eine zeitliche Unterteilung erfolgt dabei nicht, beide Komponenten werden für die gesamte Impulsantwort analysiert und gebildet. Zudem definiert die Signalverarbeitung die Struktur der Schallfeldkomponenten und reduziert diese auf einige wenige Variablen. Grundsätzlich wird angestrebt, den Aufwand der Synthese möglichst gering zu halten und das Schallfeld im Hinblick auf die Grenzen der menschlichen, auditiven Wahrnehmung zu reduzieren.

Die gerichteten Schallfeldanteile werden durch ein räumlich-aufgelöstes Echogramm bestimmt, welches für jede gerichtete Komponente $p_k(n)$ deren Orientierung Az_{pk} und El_{pk} , Zeitpunkt t_{pk} Pegel dB_{pk} und spektralen Verlauf FIR_{pk} definiert. Synthetische BRIRs mit gerichteten Anteilen werden auf Basis von HRIR-Datensätzen erzeugt, deren Impulsantworten entsprechend dieser Variablen ausgewählt, abgeschwächt, verzögert und spektral verändert werden.

Die diffusen Komponenten werden durch die frequenzabhängigen Energieabklingkur-

ven $EDC(f)$, den Faktor $mE(f)$ für die mittlere Energie (ebenfalls frequenzabhängig) und dem Verlauf der interauralen Kohärenz $IC(f)$ über der Frequenz charakterisiert. Zur Synthese der BRIRs mit diffusen Schallfeldanteilen werden ausschließlich die entsprechenden Variablen in Kombination mit weißen Rauschsignalen verwendet. Durch Aufprägen des frequenzabhängigen Energieabfalls und Anpassung der interauralen Kohärenz, werden abklingende Rauschsignale mit binauralen Eigenschaften gebildet. Die Synthese der diffusen Anteile folgt dabei aus der Literatur bekannten Modellen und kombiniert mehrere Ansätze miteinander. Insbesondere die Vereinfachung der Schallfeldanteile auf einige wenige Variablen ermöglicht eine effektive Reduktion der Informationen. Zudem schafft die Betrachtung der perzeptiven mixing time und eine daraus mögliche Begrenzung der räumlichen Auflösung weitere Möglichkeiten zur Simplifizierung.

Weiterhin wurde die Signalverarbeitung für die Analyse einer Referenzmessung vorgestellt, die das gegebene Schallfeld entsprechend der gerichteten und diffusen Komponenten untersucht und mittels der definierten Struktur charakterisiert. Diese Komponenten können dann zur Synthese eines binauralen Datensatzes eingesetzt werden, welcher zur dynamischen Auralisation verwendet werden kann und dessen zeitliche, spektrale und räumliche Struktur weitestgehend der Referenz entspricht. Insbesondere die Analyse der gerichteten Komponenten folgt innovativen Ansätzen durch Verwendung der Schallfelderlegung mittels Wellenfeldanalyseverfahren.

Alle vorgestellten Verarbeitungsschritte können isoliert oder auch kombiniert angewendet werden und bieten so viele Möglichkeiten, um BRIRs zu charakterisieren und/oder synthetisch zu generieren. Dabei können die Variablen der Synthese leicht verändert werden, um so auch beliebige Räumlichkeiten nachzubilden. Weiterhin können alle Module mit unterschiedlichen Eingangsdaten wie omnidirektionalen und binauralen Raumimpulsantworten als auch mit sphärischen Mikrofonarray-Messungen arbeiten. Dabei skaliert sich dann die mögliche Auflösung der Analyse oder einzelne Verarbeitungsschritte wie z.B. die Kohärenz-Anpassung müssen idealisiert erfolgen. Die Leistungsfähigkeit dieser Signalverarbeitung wird nun im weiteren Verlauf dieser Arbeit technisch untersucht sowie perzeptiv evaluiert und ausgewertet.

KAPITEL 5

Parametrische Modelle für die BRIR-Synthese

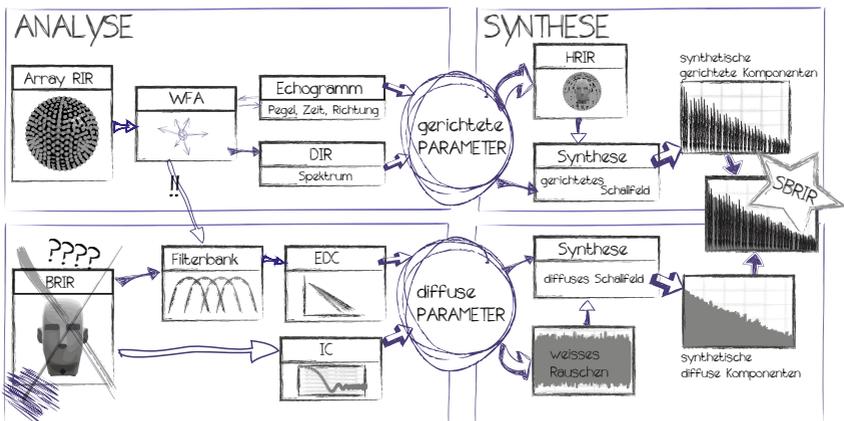


Abbildung 5.1: Skizze eines parametrischen Modells zur BRIR-Synthese.

Dieses Kapitel befasst sich mit der differenzierten Untersuchung verschiedener *parametrischer Modelle* zur Synthese von binauralen Raumimpulsantworten (SBRIRs) auf Basis messtechnisch erfasster Schallfelder. Als parametrisches Modell wird dabei das Gesamtsystem verstanden, welches zur Analyse eines Referenz-Schallfeldes im Raum und dessen Charakterisierung mittels einer parametrischen Beschreibung sowie zur Generierung von SBRIRs mit den maßgeblichen Merkmalen der Referenz auf Basis ebendieser Beschreibung, eingesetzt wird. Dabei verbindet die parametrische Kennzeichnung der wesentlichen Schallfeldkomponenten die Analyse mit der Synthese. Die Generierung der SBRIRs erfolgt ausschließlich mittels dieser Parameter, ohne Kenntnis der originalen Referenz-Signale. Das Ziel sind SBRIRs, die zur Auralisation von akustischen Umgebungen mittels dynamischer Binauralsynthese eingesetzt werden können und eine adäquate Hörbarmachung des entsprechenden Raumes ermöglichen. Die Kopfausrichtung der Rezipienten soll dabei berücksichtigt werden, um eine ortsfeste Abbildung der virtuellen Schallquellen in der Auralisation zu erreichen. Daher werden SBRIR-Datensätze benötigt, die richtungsabhängige Informationen und Reize für verschiedene Orientierungen des Hörers beinhalten. Alle nötigen Verarbeitungsschritte der Signalverarbeitung zur Analyse und Synthese wurden bereits erläutert (siehe Kapitel 4) und werden nun in diesem Kapitel in unterschiedlicher Kombination und Auflösung angewendet. Daraus werden vier parametrische Modelle gebildet, die im Folgenden untersucht werden.

Das Kapitel ist wie folgt strukturiert: In Abschnitt 5.1 wird ein allgemeiner Überblick über das System gegeben und dessen grundlegende Bestandteile erläutert. Es findet sich zunächst kein konkretes Modell, vielmehr gilt es den Grundgedanken der parametrischen Betrachtung zu vermitteln. Im Anschluss erfolgt in Abschnitt 5.2 die Darstellung einer Studie zur Wahrnehmung der Übergangszeit in einen diffusen Nachhall (Experimente MT1/2). Auch hier wird noch kein parametrisches Modell gebildet, die Untersuchung ist aber essentiell für die weitere parametrische Betrachtung und die folgende Unterteilung. Entsprechend der Signalverarbeitung und der Segmentierung des Schallfeldes in diffuse und gerichtete Komponenten sowie der zeitlichen Betrachtung einer Raumimpulsantwort gemäß Abbildung 2.8 werden anschließend verschiedene parametrische Modelle gebildet und perzeptiv mit der selben Versuchsanordnung evaluiert. Die Unterteilung der BRIRs erfolgt zunächst auf Basis der Ergebnisse aus Abschnitt 5.2 zeitlich anhand einer im Modell definierten Übergangszeit (t_{mm}) in das diffuse Schallfeld (siehe 2.1.2.1). So werden frühe (Beginn der Impulsantwort bis t_{mm}) und späte (ab t_{mm} bis Ende der Impulsantwort) Anteile getrennt voneinander betrachtet und synthetisiert. Da in den Modellen teilweise verschiedene Übergangszeiten pro Raum Anwendung finden, wird bewusst die Bezeichnung t_{mm} verwendet, anstelle der zuvor definierten perzeptiven (t_{mp}) oder physikalischen Übergangszeit (t_m).

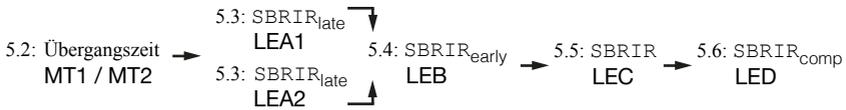


Abbildung 5.2: Abfolge der Experimente mit Kapitelstruktur und verwendeter parametrischer Modelle.

In Abschnitt 5.3 findet sich die Untersuchung des parametrischen Modells $\text{SBRIR}_{\text{late}}$ zur Synthese der späten Anteile einer BRIR anhand der zwei Experimente **LEA1** und **LEA2**. Frühe Anteile entsprechen dabei der Referenz-Messung und werden nicht synthetisiert. Diese Betrachtungsweise wird in Abschnitt 5.4 gespiegelt, sodass im Modell $\text{SBRIR}_{\text{early}}$ nur die frühen Schallfeldanteile bis zu einer definierten Übergangszeit synthetisiert werden und die späten Anteile der Referenz entnommen werden. Das entsprechende Experiment wird mit **LEB** bezeichnet. So können in **LEB** und **LEA1/2** die Verarbeitungsschritte der frühen und späten Anteile unabhängig voneinander untersucht und evaluiert werden, ohne eine wechselseitige Einflussnahme. Das parametrische Modell SBRIR in Abschnitt 5.5 kombiniert beide Ansätze und untersucht daher in Experiment **LEC** vollständig synthetische BRIRs. Da dieses System jedoch auf Referenz-Datensätzen mit hohem messtechnischem Aufwand basiert, wird in Abschnitt 5.6 mit dem Modell $\text{SBRIR}_{\text{comp}}$ ein vereinfachter Ansatz vorgestellt und dessen Leistungsfähigkeit bei unterschiedlichen Konfigurationen in Experiment **LED** evaluiert.

Alle Modelle werden jeweils mit identischer Struktur erläutert und zunächst technisch untersucht. Diese Untersuchung setzt sich zum Ziel, die generierten Datensätze anhand allgemein anerkannter raumakustischer Maße (siehe Abschnitt 2.3) zu untersuchen. So wird eine Vielzahl raumakustischer Kriterien (deren Definitionen finden sich in Rossing (2015)) auf Basis der Referenz-BRIRs als auch der synthetischen Pendanten mithilfe der ITA-Toolbox (Dietrich et al., 2013; Berzborn et al., 2017) berechnet und verglichen (siehe Abschnitt 3.4.2). Durch Bildung der Differenzbeträge der Kriterien zwischen Referenz-BRIR und SBRIR können Abweichungen herausgearbeitet werden.

Die perzeptive Evaluation der entsprechenden Modelle in den Experimenten **LEA1/2**, **LEB**, **LEC** und **LED** basiert auf einer ABC/HR Versuchsanordnung (siehe Abschnitt 3.6.1) gemäß ITU-R BS.1116-3 (2015) mit dynamischer Binauralsynthese unter Einsatz des SoundScape Renderers (Geier et al., 2008). Dabei wird die Auralisation auf Basis der SBRIR einer Hörbarmachung mittels einer messtechnisch im Raum erfassten BRIR gegenübergestellt und deren Unterschiede zueinander bewertet. Zudem bauen die Untersuchungen aufeinander auf und die zeitliche Bearbeitung entspricht der Kapitelstruktur, sodass resultierende Erkenntnisse (insbesondere der Hörversuche) direkt in den Aufbau nachfolgender Modelle einfließen können. Es entsteht ein wachsender

Prozess, abhängig von der menschlichen Wahrnehmung. Weiterhin nutzen die Modelle Wahrnehmungsgrenzen des auditiven Systems zur Reduktion der Informationen, sodass die Formulierung der *perzeptiv motivierten* Synthese verwendet wird. Abbildung 5.2 skizziert die Abfolge und Entwicklung der parametrischen Modelle sowie der entsprechenden Experimente und gibt eine Übersicht über die Kapitelstruktur.

5.1 Idee und Systemüberblick

Die folgende Betrachtungsweise wird für die Entwicklung der parametrischen Modelle zu Grunde gelegt: Das Schallfeld einer realen, akustischen Situation kann auf einige wenige, primäre und die für die auditive Wahrnehmung relevante Merkmale, reduziert werden. Diese können wiederum (entsprechend ihrer Komplexität) mit mathematischen Formeln oder Einzahlwerten beschrieben werden. Die Variablen, die das Schallfeld im Bezug auf das auditive System möglichst ausreichend charakterisieren, werden im Folgenden *Parameter* genannt. So handelt es sich um eine definierte Struktur von Werten, welche die für die menschliche Perzeption maßgeblichen Eigenschaften, eines messtechnisch erfassten Schallfeldes, kennzeichnen. Welche Eigenschaften dabei konkret gewonnen und mittels der Parameter beschrieben werden müssen, gilt es, in den folgenden Experimenten zu untersuchen und gegebenenfalls zusätzlich deren Auflösung zu definieren. Die Relevanz der Eigenschaften für die auditive Wahrnehmung wird anhand von Hörversuchen mittels dynamischer Binauralsynthese bestimmt.

Um eine akustische Situation mit Parameter zu kennzeichnen, müssen diese zunächst durch die Analyse einer Referenz-Messung bestimmt werden. Zur sinnvollen Speicherung der Parameter, können anschließend Methoden zur Datenreduzierung notwendig werden. Es wird zwar vorerst eine definierte Struktur von Parametern verwendet, doch da die Signalverarbeitung in mehrere unabhängige Verarbeitungsschritte unterteilt ist, sind Erweiterungen oder Beschränkungen der Prozesse leicht möglich. Daher ist die folgende segmentierte Betrachtungsweise der Modelle auch möglich, sodass einzelne Elemente der Datenverarbeitung separat untersucht und evaluiert werden können.

Wird ein Schallfeld im Raum messtechnisch erfasst und mittels Parametern charakterisiert, so kann eine Synthese auf Basis dieser Parameter erfolgen und möglicherweise SBRIRs erzeugen, die eine adäquate Auralisation für beliebige Orientierungswinkel erlauben. Die perzeptiven Unterschiede zwischen der Auralisation auf Basis der Messung oder dem synthetischen Äquivalent sollen möglichst gering sein. Eine gar authentische Wiedergabe nach Brinkmann et al. (2017b) ist weiterhin wünschenswert, der Fokus der Untersuchungen liegt jedoch auf einer wirksamen Reduktion der Informationen des Schallfeldes entsprechend der Wahrnehmung. So werden die Verarbeitungsschritte hin-

sichtlich ihres Aufwandes mit Blick auf das Ergebnis abgewägt. Die Entwicklung der Modelle und die entsprechende Analyse sowie Synthese erfolgt generisch, ohne spezifische Anpassung der Datenverarbeitung an die jeweilige akustische Situation.

Analog zu der in Abschnitt 4 bereits vorgestellten Differenzierung der Signalverarbeitung, betrachten die parametrischen Modelle das Schallfeld isoliert anhand von zwei grundlegenden Eigenschaften – den diffusen und gerichteten Anteilen. Diese Komponenten des messtechnisch erfassten Schallfeldes werden separat bestimmt, gespeichert als auch synthetisiert. Alle in dieser Arbeit enthaltenden Blockschaltbilder der parametrischen Modelle sind identisch aufgebaut, damit die einzelnen Experimente anschaulich untereinander verglichen werden können. Dabei findet sich dieselbe Struktur entsprechend der vorherigen Skizze in Abbildung 5.1. Die parametrische Beschreibung liegt im Zentrum und verbindet die Analyse (links) mit der Synthese (rechts). Der obere Teil der Blockschaltbilder befasst sich mit den gerichteten, der untere mit den diffusen Komponenten des Schallfeldes. Schwarze Pfeile symbolisieren den Audio-Signalweg der Impulsantworten, graue Pfeile den Weg der parametrischen Beschreibungen. Eine Unterscheidung zwischen breitbandigen und frequenzabhängigen Signalen und Parametern erfolgt durch die Anzahl der Pfeile. Einzelne Pfeile symbolisieren dabei frequenzunabhängige Prozesse.

Die Synthese auf Basis von Parametern ermöglicht des Weiteren eine Adaption oder Modifikation der Daten, sodass die Eigenschaften der akustischen Situation systematisch verändert werden können. Diese Anwendung wird in dieser Arbeit jedoch nicht betrachtet, da es zunächst gilt, die Leistungsfähigkeit der Modelle zu evaluieren und die für eine angemessene Hörbarmachung relevanten Merkmale zu extrahieren. So erfolgt die Untersuchung der SBRIRs kritisch durch direkten Vergleich mit der entsprechenden erfassten Referenz-Messung.

5.2 Übergangszeit zwischen frühen und späten Anteilen

Da zunächst die Betrachtung der frühen und späten Schallfeldanteile und deren Synthese anhand der parametrischen Modelle separat erfolgen soll, ist die Kenntnis der Übergangszeit in ein diffuses Schallfeld essentiell. Daher wird in dieser Arbeit auf eine Studie von Stade (2015) zurückgegriffen, in der verschiedene Übergangszeiten in den selben Räumlichkeiten (die in den parametrischen Modellen betrachtet werden) bereits untersucht wurden. Der folgende Exkurs fasst die wesentlichen Ergebnisse dieser Untersuchung zusammen. Da sich in der Literatur stark unterschiedliche Prädiktoren der Übergangszeit in ein ausschließlich diffus wahrnehmbares Schallfeld finden lassen und diese mitunter für eigene Anwendungen nicht zu zufriedenstellenden Lösungen führ-

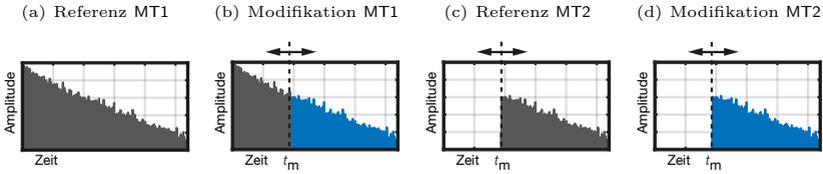


Abbildung 5.3: Schematische Darstellung der Datensätze aus Stade (2015): Referenz in Experiment MT1 (a) und in Experiment MT2 (c); Modifikation in Experiment MT1 (b) und in Experiment MT2 (d); grau = dynamisch, blau = statisch; ab t_m keine Anpassung an die Orientierung, t_m variabel.

ten, wurde ein alternativer Ansatz verfolgt. In einer perceptiven Untersuchung wurden mittels MUSHRA-Versuchsordnung und unter Anwendung der dynamischen Binauralsynthese mittels des SoundScape Renderers (Geier et al., 2008), sechs verschiedene Übergangszeiten im KVB-Saal als auch im kleinen Sendesaal mit unterschiedlichen Schallquellen (omnidirektional und gerichtet) verglichen (Eigenschaften siehe Tabelle 5.1). Gemäß der Annahme, dass im diffusen Schallfeld die eintreffende Schallenergie nicht von der Richtung abhängt und für alle Einfallswinkel identisch ist, wurde ab der jeweiligen Übergangszeit analog zu Lindau et al. (2012) auf eine dynamische Anpassung der Binauralsynthese verzichtet und ausschließlich die BRIR in Blickrichtung zur Anregungsquelle ($\phi = 0^\circ$, $\theta = 0^\circ$) zur Faltung verwendet (Experiment MT1). Tabelle 5.2 gibt einen Überblick über Übergangszeiten ins diffuse Schallfeld aus relevanten Quellen für die untersuchten Räumlichkeiten.

In einem zweiten Experiment MT2 wurden zudem als Erweiterung, alle Schallfeldanteile vor den Übergangszeiten aus den Impulsantworten entfernt, sodass nur die zeitlich späten Anteile zur Auralisation verwendet wurden. Die Referenz-Datensätze wurden dazu gleichermaßen gekürzt, analog zu Experiment MT1 wurden für die Bearbeitungen ausschließlich statische späte Anteile der frontal orientierten BRIR eingesetzt. Eine schematische Darstellung der untersuchten Datensätze findet sich in Abbildung 5.3, eine detaillierte Erläuterung der Versuchsanordnung und der technischen Rahmenbedingungen in Kapitel 3.5 und 3.6.2. 16 Probanden (\varnothing 30 Jahre, 1 w) nahmen an zwei Terminen an den Experimenten teil, 9 konnten bereits Erfahrung in Hörversuchen unter dynamischer Binauralsynthese vorweisen.

In Experiment MT1 zeigte sich bei beiden Räumlichkeiten und Schallquellen, entsprechend der Erwartung, mit ansteigender Übergangszeit tendenziell eine zunehmende Bewertung der Modifikationen (siehe Abbildung 5.4). Die Mittelwerte der Bewertungen im kleinen Sendesaal lagen für beide Anregungsquellen bereits ab einem Zeitpunkt

Tabelle 5.1: Eigenschaften Experiment MT1/2.

Experiment	MT1/2
Kurzbeschreibung	Vergleich Übergangszeiten in diffusen Nachhall
Versuchsdesign	MUSHRA
Bedingungen	24
Faktoren	3
Faktorstufen	
Übergangszeit	6 $t_m = 5/10/20/40/80/160$ ms (KSS) $t_m = 10/20/40/80/160/320$ ms (KVB)
Raum	2 KSS / KVB
Schallquellen	2 gerichtet / omnidirektional
Testsignal	1 Schlagzeug
Probanden	16 (15 m / 1 w) / \varnothing 30 Jahre

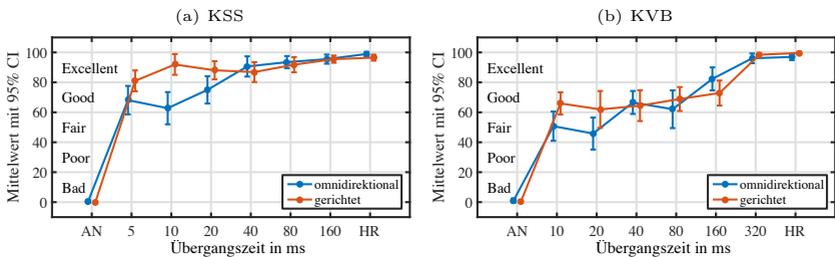


Abbildung 5.4: Experiment MT1: Bewertete Ähnlichkeit der unterschiedlichen Übergangszeiten für omnidirektionale (blau) und gerichtete (rot) Quelle; AN = Anker, HR = versteckte Referenz aus Stade (2015).

von 5 ms bei CQS > 60 und entsprachen somit dem Attribut *good*. Im KVB-Saal wurde diese Schwelle bei 40 ms (für die omnidirektionale Quelle) bzw. bei 10 ms (für die gerichtete Quelle) überschritten. Im KVB-Saal wurden alle Übergangszeiten bis 160 ms signifikant schlechter im Vergleich zur Referenz bewertet, eine Ununterscheidbarkeit zeigte sich somit in diesem Raum nur bei 320 ms. Im kleinen Sendesaal fanden sich größere Unterschiede zwischen den Schallquellen. Bei Verwendung der omnidirektionalen Schallquelle, konnte anhand der untersuchten Übergangszeiten keine Ununterscheidbarkeit zwischen Referenz und Modifikation gezeigt werden. Selbst die Bearbeitung mit 160 ms Übergangszeit wurde signifikant unterschiedlich im Vergleich zur Referenz bewertet. Bei Verwendung der gerichteten Quelle hingegen, wurde ausschließlich die Übergangszeit von 5 ms als signifikant unterschiedlich bewertet, sodass in allen weiteren Modifikationen den Probanden keine Unterscheidung möglich war.

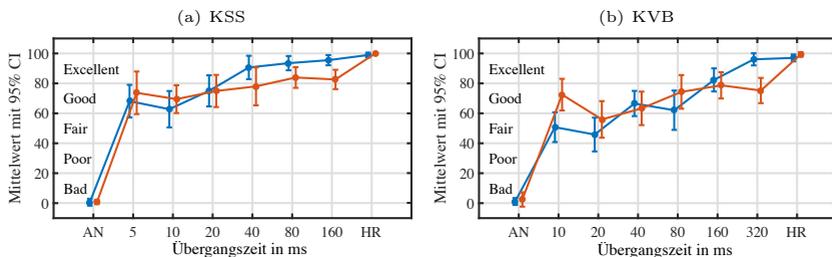


Abbildung 5.5: Vergleich Experiment MT1/MT2: Bewertete Ähnlichkeit der unterschiedlichen Übergangszeiten für MT1 (blau) und MT2 (rot) bei omnidirektionaler Quelle; AN = Anker, HR = versteckte Referenz aus Stade (2015).

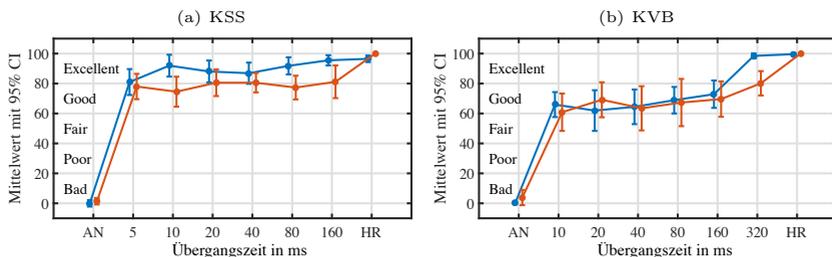


Abbildung 5.6: Vergleich Experiment MT1/MT2: Bewertete Ähnlichkeit der unterschiedlichen Übergangszeiten für MT1 (blau) und MT2 (rot) bei gerichteter Quelle; AN = Anker, HR = versteckte Referenz aus Stade (2015).

Experiment MT2 zeigte bei beiden Räumlichkeiten und Schallquellen einen ähnlichen Verlauf wie Experiment MT1 (siehe Abbildung 5.5 für die omnidirektionale und 5.6 für die gerichtete Schallquelle). Die Bewertungen stiegen tendenziell mit der Übergangszeit in den statischen Nachhall an und befanden sich auch hier bereits ab einem Zeitpunkt von 5 ms im Bereich des Attributs *good*. Im kleinen Sendesaal lagen die Mittelwerte der Bewertungen von Experiment MT2 jedoch nahezu alle unterhalb der von Experiment MT1. Zudem war den Probanden hier bei Verwendung der gerichteten Schallquelle nun ebenfalls eine signifikante Unterscheidung für alle Übergangszeiten möglich. Die Bewertungen im KVB-Saal differierten zwischen den Versuchen MT1 und MT2 marginal, allerdings konnten nun auch bei Verwendung der spätesten Übergangszeit von 320 ms signifikante Unterschiede detektiert werden.

Tabelle 5.2: Übergangszeiten ins diffuse Schallfeld in ms, basierend auf ausgewählter Literatur aus Cremer (1948), Rubak und Johansen (1999), Hidaka et al. (2007) und Lindau et al. (2012).

Raum	Cremer	Rubak und Johansen	Hidaka et al.	Lindau et al. _{model}	Lindau et al. _{data}
KSS	35	77	74	65	62
KVB	78	102	140	121	74

Einerseits wurden somit alle untersuchten Modifikationen mit dem Attribut *fair* oder besser bewertet ($CQS > 40$). Dies war insbesondere bei frühen Übergangszeiten überraschend, da hier mitunter durch die Modifizierung der BRIR-Datensätze direkt in erste Reflexionen eingegriffen wurde. Andererseits zeigte sich, dass teilweise selbst späte Übergangszeiten (z.B. kleiner Sendesaal bei 160 ms) signifikant unterschiedlich zur Referenz bewertet wurden, d.h. den Probanden war eine Unterscheidung der Stimuli weiterhin möglich und eine Ununterscheidbarkeit wurde nicht erreicht, sodass kein diffuses Schallfeld angenommen werden konnte.

Zudem konnte eine Maskierungswirkung des Direktschalls gezeigt werden: Zum einen wurden bei der Auralisation ohne Direktschall in allen Konstellationen signifikante Unterschiede zur Referenz festgestellt, während mit Direktschall mitunter keine Unterscheidung möglich war. Zum anderen wurden die Stimuli ohne Direktschall in der Regel signifikant unterschiedlicher bewertet als die entsprechenden Stimuli mit Direktschall. Aufgrund der Maskierung durch den Direktschall wurden Veränderungen im diffusen Nachhall teilweise unhörbar. Somit könnte das Verhältnis zwischen Direkt- und Diffusschall ein relevanter Faktor für die Bestimmung der perzeptiven Übergangszeit sein. Insbesondere bei frühen Übergangszeiten in einen statischen Nachhall konnten signifikante Unterschiede zwischen den Richtcharakteristika nachgewiesen werden. Dabei wurde im Vergleich zur Referenz die gerichtete Schallquelle tendenziell ähnlicher als die omnidirektionale Schallquelle bewertet. Auch dies spricht für eine hohe Relevanz des Verhältnisses von Direkt- zu Diffusanteil. Bei der gerichteten Quelle dominierte der Direktschall, was möglicherweise zu einer stärkeren Maskierungswirkung führte.

Diese Studie legt somit dar, dass die Auswahl geeigneter Übergangszeiten nicht trivial ist und von verschiedenen Faktoren, wie z.B. auch der Richtcharakteristik der Schallquelle, abhängt. Im Vergleich zu den Übergangszeiten aus relevanter Literatur (siehe Tabelle 5.2) zeigen sich weiterhin teilweise deutliche Unterschiede durch andere Übergangszeitpunkte. Daher sollte abgewägt werden, ob für die Anwendung der Übergangszeiten in das diffuse Schallfeld in den parametrischen Modellen diese konservativeren Zeitpunkte gemäß der Ergebnisse aus Stade (2015) gewählt werden sollten.

5.3 Späte Anteile (LEA1/2)

Das parametrische Modell $SBRIR_{late}$ befasst sich mit der Generierung der zeitlich späten Anteile einer binauralen Raumimpulsantwort. Bis zu einer im Modell gewählten Übergangszeit t_{mm} entsprechen die Impulsantworten zunächst der messtechnisch erfassten Referenz-BRIR. Ab der Übergangszeit wird die jeweilige SBRIR, die durch das parametrische Modell erzeugt wird, anstelle der Messung verwendet. So werden synthetische und messtechnisch erfasste BRIRs gemeinsam in einer Impulsantwort verknüpft. Das Modell wurde von Stade und Arend (2016a,b) vorgestellt und perceptiv evaluiert. Aufgrund zuvor getätigter Untersuchungen (siehe Stade (2015)) und dem Bedarf an synthetischem Nachhall in weiteren Forschungsprojekten der TH Köln wird mit der Untersuchung der späten Anteile begonnen.

Die folgende Untersuchung ist in zwei separate Experimente LEA1 und LEA2 unterteilt, die zwar auf dem selben parametrischen Modell $SBRIR_{late}$ zur Synthese der späten Anteile basieren, jedoch dazu unterschiedliche Konfigurationen nutzen und sich weiterhin in den auralisierten Bedingungen unterscheiden. LEA1 synthetisiert nur eine akustische Situation, es werden die Signalverarbeitungsschritte im Detail untersucht, durch Variation der Auflösung einzelner Prozesse und der Konfiguration des Modells. In Experiment LEA2 hingegen wird das Modell anhand verschiedener Räumlichkeiten und Schallquellen evaluiert.

Im Allgemeinen soll untersucht werden, inwieweit sich die Auralisation unter Anwendung von BRIRs mit synthetischen späten Anteilen von einer Hörbarmachung auf Basis messtechnisch erfasster BRIRs unterscheidet und ob das gewählte Modell die reale, akustische Situation hinreichend charakterisiert. Dazu werden zunächst das Modell an sich, dessen Konfiguration und die verwendeten Parameter erläutert und die für die perceptiv Untersuchung generierten Datensätze vorgestellt. Anschließend werden die BRIRs mit synthetischen späten Anteilen technisch anhand raumakustischer Kriterien untersucht und den messtechnisch erfassten Pendanten gegenübergestellt, bevor sie dann in zwei getrennten Experimenten mit ABC/HR Versuchsanordnung perceptiv untersucht und ausgewertet werden.

5.3.1 Modell und Parameter

In Abbildung 5.7 findet sich das Blockschaltbild des parametrischen Modells $SBRIR_{late}$ zur Generierung binauraler Raumimpulsantworten mit synthetischen späten Anteilen, welches den folgenden Experimenten LEA1 und LEA2 zu Grunde liegt. Da dieses Modell die Synthese erst ab der Übergangszeit t_{mm} verwendet, werden ausschließlich diffuse Komponenten des Schallfeldes betrachtet und synthetisiert. Der Theorie nach treten

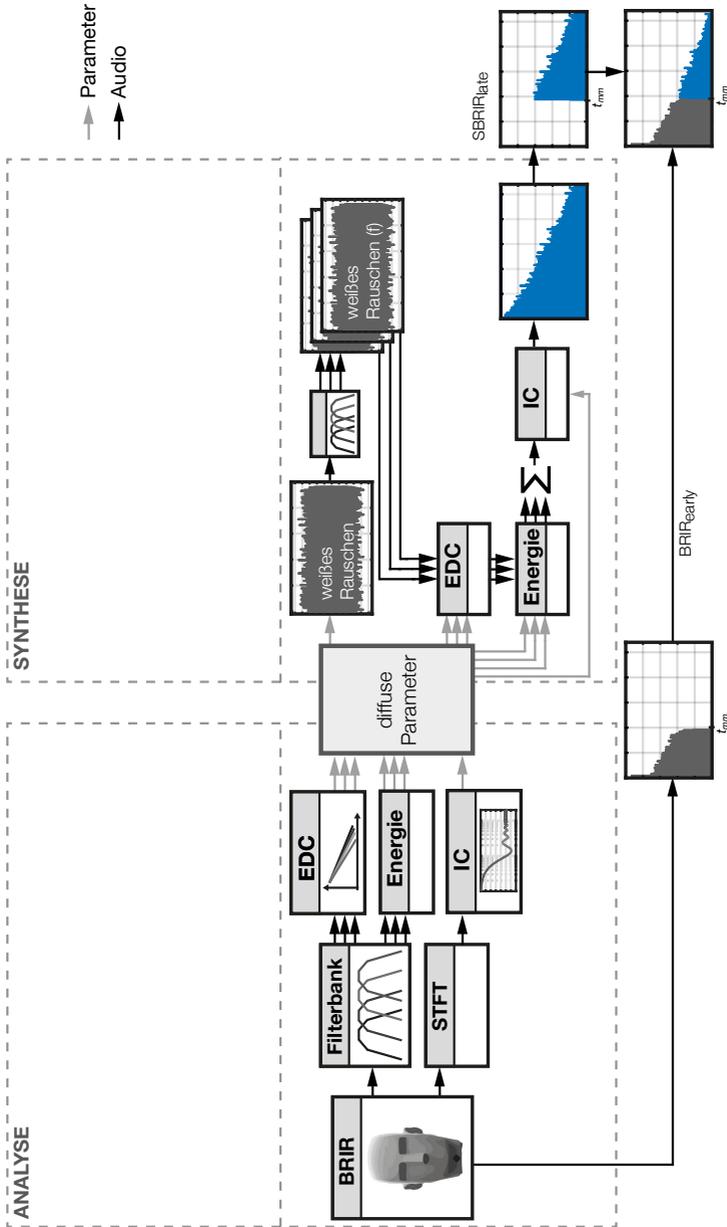


Abbildung 5.7: Blockschaltbild $\text{SBRIR}_{\text{late}}$: Diffuse Komponenten einer Referenz-BRIR werden analysiert und mittels Parameter beschrieben. Weißes Rauschen wird entsprechend der Parameter angepasst und als binärer, diffuser Nachhall verwendet (blau). Frühe Schallfeldanteile entsprechen der Referenz-BRIR. (Stade und Arend, 2016b)

Tabelle 5.3: Diffusen Parameter des parametrischen Modells und benötigte Werte in Abhängigkeit vom Polynomgrad der Abklingkurven $p_{n_{\text{EDC}}}$ und des IC-Verlaufs $p_{n_{\text{IC}}}$ sowie von der Anzahl der Frequenzbänder n_{fb} der verwendeten Filterbank.

Parameter	Beschreibung	Werte
$\text{EDC}(f)$	Polynome Abklingkurve pro Frequenzband	$(p_{n_{\text{EDC}}} + 1) \cdot n_{\text{fb}}$
	BRIR Länge in Samples	1
$\text{mE}(f)$	mittlere Energie pro Frequenzband	$2 \cdot n_{\text{fb}}$
	Beginn Grundrauschen in Samples	n_{fb}
$\text{IC}(f)$	Polynom interaurale Kohärenz	$p_{n_{\text{IC}}} + 1$
	IC-Verlauf Länge in Samples	1

bei Wahl der entsprechenden Übergangszeiten in das diffuse Schallfeld, im Hinblick auf die Geometrie des Raumes und dessen Reflexions-Struktur ($t_{\text{mm}} \equiv t_{\text{mp}}$), keine wahrnehmbaren, richtungssensitiven Reize im Schallfeld auf (siehe 2.1.2.1). Daher sollte t_{mm} nach Möglichkeit t_{mp} entsprechen, sodass die Betrachtung aller gerichteten Komponenten in diesem Modell verworfen werden kann. Es soll zusätzlich untersucht werden, welche Auswirkungen die Wahl abweichender Übergangszeiten hat ($t_{\text{mm}} \neq t_{\text{mp}}$). Die diffusen Komponenten werden mit den in Abschnitt 4.2.1 erläuterten Methoden detektiert und approximiert. Als Eingangsvariable benötigt das System für verschiedene Orientierungswinkel messtechnisch erfasste, binaurale Raumimpulsantworten des zu synthetisierenden Raumes. In dieser Arbeit werden dazu Datensätze mit 1° horizontaler Auflösung (360 BRIRs) verwendet. Aufgrund der ausschließlichen Betrachtung diffuser Komponenten, bedarf es für die Analyse nur einer Orientierung im Raum. Somit werden zur Analyse der diffusen Komponenten ausschließlich BRIRs in Blickrichtung zur Anregungsquelle ($\phi = 0^\circ$, $\theta = 0^\circ$) aus Stade et al. (2012) verwendet.

Die parametrische Beschreibung charakterisiert das diffuse Schallfeld im Raum gemäß Tabelle 5.3 mit den drei Parametern $\text{EDC}(f)$, $\text{mE}(f)$ und $\text{IC}(f)$. Die diffusen Schallfeldanteile werden so mit einigen wenigen Werten beschrieben, deren Umfang hängt dabei von der Frequenzauflösung entsprechend der Anzahl der verwendeten Frequenzbänder (n_{fb}) sowie dem Grad des Polynomfits der Abklingkurven ($p_{n_{\text{EDC}}}$) und des IC-Verlaufs ($p_{n_{\text{IC}}}$) ab.

Für die folgenden Experimente werden Polynomgrade von $p_{n_{\text{EDC}}} = 6$ und $p_{n_{\text{IC}}} = 18$ verwendet, sodass die entsprechenden Kurvenverläufe hinreichend genau approximiert werden können. Der Verlauf der interauralen Kohärenz über der Frequenz wird ausschließlich für den späten Nachhall der Referenz-BRIR bestimmt¹. Die angewandte STFT zur Kalkulation der Kohärenz basiert auf einem Hamming-Fenster mit 256 Samples und 50 % Überschneidung der Segmente.

¹Beginn Zeitfenster: ab 320 ms nach Direktschall; Ende Zeitfenster: Länge der BRIR

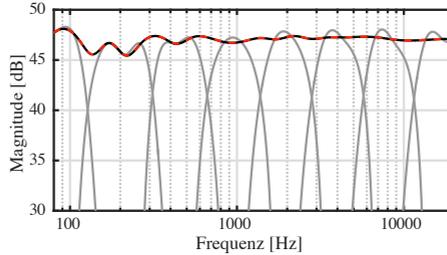


Abbildung 5.8: Darstellung der Filterbank im Frequenzbereich für $fb = 8$: Frequenzbänder (grau), original Signal (schwarz), mit Filterbank zerlegt und anschließend zusammengesetzt (rot).

Um eine akkurate Frequenz-Auflösung zu ermöglichen, wird zunächst eine Filterbank mit ansteigender Güte (und damit schmaler werdenden Frequenzbändern) zu hohen Frequenzen verwendet. Die Bandbreite der Frequenzbänder beträgt bis 6300 Hz $1/3$ -, bis 16000 Hz $1/6$ -, bis 19000 Hz $1/12$ - und oberhalb von 19000 Hz $1/24$ -Oktav. Dabei wird die BRIR in insgesamt $n_{fb} = 32$ Frequenzbänder zwischen 100 Hz und 20 kHz aufgespalten. Diese feine Auflösung ermöglicht eine sehr präzise Abbildung der frequenzabhängigen Nachhallzeiten, insbesondere bei hohen Frequenzen. In Experiment LEA1 wird das Modell zusätzlich mit einer Oktav-Filterbank untersucht und die Analyse in 8 Frequenzbändern zwischen 125 Hz und 16 kHz durchgeführt. So kann die Präzision der Abbildung der Nachhallzeiten für unterschiedliche Frequenzauflösungen der Filterbank verglichen und deren Einfluss auf die Auralisation bewertet werden. Die hohe Frequenzauflösung zur Adaption der Abklingkurven unterscheidet diese Studie von bisherigen Arbeiten, in der Regel lassen sich in den entsprechenden Systemen wie z.B. RSAO (Remaggi et al., 2015) Oktavbandbreiten finden.

Das Modell reduziert bei Verwendung der Filterbank mit 32 Bändern das Schallfeld auf insgesamt 341 Werte, bei Gebrauch einer Oktav-Filterbank auf lediglich 101 Werte. Zum Vergleich: Eine gemessene BRIR mit der Länge von 2 Sekunden umfasst bei einer Abtastrate von 48 kHz 192.000 Samples, ein BRIR-Datensatz zur dynamischen Auralisation mit 1° horizontaler Auflösung 69.120.000 Samples.

Die Synthese verwendet ausschließlich die genannten Parameter zur Generierung von SBRIRs mit diffusen Komponenten, ohne Kenntnis der originalen Referenz-Messung. Die Signalverarbeitung basiert ebenfalls auf den in Abschnitt 4.2.2 vorgestellten Methoden. Die Verarbeitung erfolgt grundsätzlich für die gesamte Impulsantwort, eine zeitliche Kürzung kann später durchgeführt werden. Entsprechend der Länge der Referenz wird ein zwei-kanaliges weißes Rauschsignal erzeugt. Für die Evaluation der

entwickelten Algorithmen wird jedoch auf eine individuelle Erzeugung der Rauschsignale pro Raum und Experiment verzichtet. Um Unterschiede aufgrund verschiedener Rauschsignale zu vermeiden (die aufgrund der stochastischen Verteilung unvermeidlich wären), wird für diese Arbeit ausschließlich ein einziges zwei-kanaliges Rauschsignal generiert, das in allen Experimenten Anwendung findet. Es erfolgt lediglich eine Kürzung des Signals entsprechend der Länge der Impulsantworten der Räumlichkeiten. Um eine systematische Analyse aller Verarbeitungsschritte zu gewährleisten, sollten Unterschiede verschiedener Zufallssignale nicht mit einfließen, obwohl diese äußerst gering sind. Die Rauschsignale werden mit derselben Filterbank und Konfiguration prozessiert, die für die Analyse Anwendung findet. Deren Frequenzauflösung und Mittenfrequenzen müssen identisch sein, damit die Parameter die jeweils korrespondierenden Frequenzbänder der Rauschsignale bearbeiten.

Mittels Polynomauswertung der Parameter $EDC(f)$ und $IC(f)$ können die Verläufe der Abklingkurven und der interauralen Kohärenz der Referenz wiederhergestellt werden. Den Rauschsignalen (Anzahl $n_{fb} \times 2$) wird der frequenzabhängige Energieabfall der Referenz aufgeprägt und die Amplitude mittels der mittleren Energie $m_E(f)$ an die Messung angepasst. Anschließend werden alle Frequenzbänder aufsummiert, sodass wieder ein zwei-kanaliges Rauschsignal entsteht. Die Kohärenz der beiden Kanäle kann gemäß der binauralen Messung mit $IC(f)$ adaptiert werden, sodass sich abklingende Rauschsignale mit den binauralen Merkmalen des diffusen Schallfeldes der Referenz ergeben. Die Relevanz dieses Verarbeitungsschrittes wird in LEA1 gesondert untersucht, indem zusätzlich Datensätze ohne Kohärenzanpassung erzeugt werden.

Im letzten Verarbeitungsschritt werden Referenz-BRIRs und SBRIRs miteinander verbunden. Ab der gewählten Übergangszeit t_{mm} werden dabei für alle Orientierungen anstelle der Referenz-BRIR die synthetischen Impulsantworten verwendet. Da die Synthese entsprechend der Länge der Referenz erfolgt, kann die Übergangszeit beliebig gewählt und verändert werden. So werden SBRIRs gebildet, die bis zu der Übergangszeit alle diffusen und richtungs-sensitiven Schallfeldkomponenten (entsprechend der Messung) enthalten und eine dynamische Auralisation mit einer ortsfesten Abbildung der Schallquellen ermöglichen, da jeder Orientierung eine eigene BRIR zu Grunde liegt. Ab der Übergangszeit wird für jede dieser Orientierungen die identische SBRIR mit ausschließlich diffusen Komponenten verwendet.

5.3.2 Datensätze

Tabelle 5.4: Generierte Datensätze für LEA1, abhängig von Übergangszeit t_{mm} , Anzahl Frequenzbänder n_{fb} und Kohärenzanpassung (IC).

Raum	Schallquelle	t_{mm} [ms]	n_{fb}	IC
KSS	direktional	20	8	aktiv
KSS	direktional	160	8	aktiv
KSS	direktional	320	8	aktiv
KSS	direktional	20	32	aktiv
KSS	direktional	160	32	aktiv
KSS	direktional	320	32	aktiv
KSS	direktional	20	8	inaktiv
KSS	direktional	160	8	inaktiv
KSS	direktional	320	8	inaktiv
KSS	direktional	20	32	inaktiv
KSS	direktional	160	32	inaktiv
KSS	direktional	320	32	inaktiv

Im Folgenden werden die für die technische und perzeptive Untersuchung des parametrischen Modells $\text{SBRIR}_{\text{late}}$ generierten Datensätze erläutert. Um die Leistungsfähigkeit des Modells zu evaluieren, werden SBRIRs mit unterschiedlichen Konfigurationen der Algorithmen und auf Basis verschiedener akustischer Situationen gebildet. Eine genaue Erläuterung der Methoden, der verwendeten Messsysteme und deren Konfiguration sowie eine Beschreibung der Räumlichkeiten findet sich in Kapitel 3. Es gilt, Bedingungen und Datensätze zu kreieren, mit denen Aussagen über die Wirkung der Konfiguration der Algorithmen sowie den Einfluss verschiedener Räume, Schallquellen und Übergangszeiten auf das perzeptive Modell und damit auf das Ergebnis der Synthese getroffen werden können.

LEA1: In Tabelle 5.4 sind alle in Experiment LEA1 untersuchten Konditionen und damit die erzeugten synthetischen Datensätze aufgelistet. Der Fokus dieser Studie liegt auf der Beurteilung der Signalverarbeitung des Modells und dessen Auflösung. Es soll untersucht werden, welchen Einfluss unterschiedliche Konfigurationen auf die Präzision der Synthese nehmen und inwieweit diese die Auralisation beeinflussen. Daher wird zunächst nur ein Raum (kleiner Sendesaal - KSS, mit gerichteter Schallquelle) mit unterschiedlichen Konfigurationen des parametrischen Modells synthetisiert.

Die drei Faktoren *Modell*, *Filterbank* und *IC* werden mit drei, bzw. zwei Faktorstufen betrachtet, sodass 12 ($3 \times 2 \times 2$) Datensätze erzeugt werden. Es gilt in LEA1 zum einen, den Einfluss der Frequenz-Auflösung der Filterbank zu evaluieren (Faktor *Filterbank*),

Tabelle 5.5: Generierte Datensätze für LEA2, abhängig von Schallquelle, Übergangszeit t_{mm} , Anzahl Frequenzbänder n_{fb} und Kohärenzanpassung (IC).

Raum	Schallquelle	t_{mm} [ms]
KVB	direktional	40
KVB	direktional	320
KVB	omnidirektional	40
KVB	omnidirektional	320
KSS	direktional	20
KSS	direktional	160
KSS	omnidirektional	20
KSS	omnidirektional	160

sodass zwei verschiedene Konfigurationen (Anzahl Frequenzbänder $n_{\text{fb}} = 8$ oder 32) zum Einsatz kommen, deren Aufbau bereits im Vorfeld erläutert wurde. Weiterhin soll untersucht werden, welchen Einfluss die Anpassung der interauralen Kohärenz der Impulsantworten auf das Synthese-Ergebnis nimmt (Faktor *IC*). Daher wird dieser Sachverhalt ebenfalls in zwei Faktorstufen untersucht, indem Datensätze sowohl mit als auch ohne aktiver Anpassung der Kohärenz erzeugt und bewertet werden. Der Faktor *Modell* ist in drei Faktorstufen unterteilt und variiert die Übergangszeit t_{mm} zwischen Referenz-BRIR und SBRIR mittels drei verschiedener Zeitpunkte.

Aufgrund der Ergebnisse aus Stade (2015) (siehe Abschnitt 5.2), die einerseits teilweise Unterschiede zur Literatur in Richtung spätere Übergangszeiten aber auch andererseits teilweise bei sehr frühen Übergangszeiten keine signifikanten Unterschiede zur Referenz aufzeigte, werden in LEA1 drei unterschiedliche Übergangszeiten gegenübergestellt. Für die Trennung von Referenz-BRIR und SBRIR werden die Zeitpunkte $t_{\text{mm}} = 20/160/320$ ms benutzt. Dabei repräsentiert der Zeitpunkt von $t_{\text{mm}} = 160$ ms eine verhaltene und im Hinblick auf die Ergebnisse aus Tabelle 5.2 relativ späte Wahl der Übergangszeit. Doch da in Stade (2015) selbst bei 160 ms keine Ununterscheidbarkeit (bei gerichteter Schallquelle) gezeigt werden konnte, wird als Erweiterung mit $t_{\text{mm}} = 320$ ms ein noch späterer Zeitpunkt gewählt. Andererseits erreichten jedoch in Stade (2015) selbst sehr frühe Übergangszeiten qualitativ hohe Bewertungen, daher wird in LEA1 zusätzlich der Zeitpunkt $t_{\text{mm}} = 20$ ms evaluiert.

LEA2: In Tabelle 5.5 sind alle in Experiment LEA2 untersuchten Konditionen und die erzeugten synthetische Datensätze aufgelistet. In Abgrenzung zu LEA1 wird der Algorithmus des Modells $\text{SBRIR}_{\text{late}}$ nun in verschiedenen akustischen Umgebungen evaluiert und nicht primär Konfigurationen der Signalverarbeitung. Es wird untersucht, wie

die Synthese unter veränderten akustischen Bedingungen arbeitet. Dabei wird auf die komplexeste Konfiguration des Modells aus Experiment LEA1 zurückgegriffen, sodass die Synthese auf einer Filterbank mit 32 Frequenzbändern und aktivierter Kohärenz-Anpassung der Rauschsignale basiert.

Die drei Faktoren *Modell*, *Raum* und *Schallquelle* werden mit je zwei Faktorstufen betrachtet, sodass acht ($2 \times 2 \times 2$) Datensätze erzeugt werden. Die späten Anteile von BRIRs aus zwei Räumlichkeiten (Faktor *Raum* mit den Faktorstufen *KVB* und *KSS*) werden synthetisiert. Beide Säle werden mit je zwei Anregungsquellen (Faktor *Schallquelle*) unterschiedlicher Richtcharakteristik (Faktorstufen *gerichtet* und *omnidirektional*) untersucht. Entgegen LEA1 wird die Bezeichnung der Faktorstufe *Übergangszeit* nicht verwendet und durch *Modell* ersetzt, da LEA2 nur eine Variable des parametrischen Modells variiert. Für die Faktorstufe *Modell* werden zwei verschiedene Übergangszeiten eingesetzt, die sich aufgrund unterschiedlicher Dimensionen und Nachhallzeiten zwischen den Räumlichkeiten unterscheiden. Dabei repräsentiert $t_{\text{mm late}}$ erneut eine verhaltene und späte Wahl der Übergangszeiten ($KSS = 160 \text{ ms}$ und $KVB = 320 \text{ ms}$), wohingegen mit $t_{\text{mm early}}$ das Modell auch bei sehr frühen Übergangszeiten evaluiert wird ($KSS = 20 \text{ ms}$ und $KVB = 40 \text{ ms}$). Auf die Untersuchung von weiteren Übergangszeiten muss aufgrund des Umfangs verzichtet werden.

5.3.3 Technische Untersuchung

Im Folgenden werden die Ergebnisse der technischen Untersuchung des parametrischen Modells $\text{SBRIR}_{\text{late}}$ vorgestellt. Es wird eine Reduktion auf die wesentlichen Erkenntnisse angestrebt, um so einige wenige Hauptaussagen zu treffen. Aufgrund der Vielzahl an Bedingungen des parametrischen Modells und der Fülle von raumakustischen Kriterien (die zudem frequenzabhängig bestimmt werden), ist diese Auswertung komplex und vielschichtig, sodass eine Reduktion auf die essentiellen Gesichtspunkte nötig ist. Daher werden anhand absoluter Differenzbeträge die Unterschiede zwischen Messung und Modell im Mittel über alle Frequenzen als Einzahlwerte berechnet. Die Differenzen geben den Betrag der mittleren Abweichung der SBRIR (Ist-Wert) von der messtechnisch erfassten Referenz-BRIR (Soll-Wert) an.

LEA1: In Tabelle 5.6 finden sich die absoluten Differenzbeträge zwischen Referenz-BRIR und SBRIR im Mittel über alle Frequenzen für die untersuchten Bedingungen des parametrischen Modells. Eine frequenzabhängige Betrachtung kann anhand der Plots gemäß Abbildungen A.1-A.8 (siehe Anhang) erfolgen.

Tabelle 5.6: LEA1 - Absolute Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ der raumakustischen Kriterien im Mittel über alle Frequenzen, Vergleich zwischen allen Bedingungen.

Modell			Spektrum	$T30$	EDT	$IACC_{\text{early}}$	$IACC_{\text{late}}$
t_{mm}	n_{fb}	IC	[dB]	[%]	[%]		
20	8		0,18	18,0	28,2	0,11	0,21
160	8		0,02	11,3	11,3	0	0,06
320	8		0,003	11,5	0,8	0	0,009
20	32		0,1	13,0	24,0	0,14	0,25
160	32		0,02	7,4	7,5	0	0,07
320	32		0,003	6,9	1,0	0	0,01
20	8	×	0,18	18,1	31,5	0,08	0,1
160	8	×	0,02	11,3	10,0	0,0001	0,03
320	8	×	0,004	11,1	0,6	0	0,004
20	32	×	0,12	14,4	26,8	0,09	0,12
160	32	×	0,02	7,0	8,0	0	0,03
320	32	×	0,003	6,8	0,8	0	0,004

Modell			T_s	$C80$	$D50$	BR	TR
t_{mm}	n_{fb}	IC	[ms]	[dB]	[%]		
20	8		10,1	4,1	10,6	0,01	0,002
160	8		3,6	0,6	1,7	0,03	0,003
320	8		0,6	0,4	0,1	0,09	0,03
20	32		6,3	3,6	8,1	0,006	0,001
160	32		2,8	0,5	1,3	0,06	0,001
320	32		0,6	0,04	0,1	0,06	0,004
20	8	×	9,8	4,0	11,2	0,04	0,06
160	8	×	3,5	0,5	1,5	0,01	0,03
320	8	×	0,6	0,03	0,1	0,05	0,05
20	32	×	7,1	3,8	9,6	0,03	0,05
160	32	×	2,7	0,5	1,3	0,05	0,02
320	32	×	0,5	0,03	0,09	0,06	0,01

Im Allgemeinen werden die Unterschiede zwischen Messung und Modell mit steigender Übergangszeit erwartungsgemäß geringer. Weiterhin zeigt sich mit steigender Komplexität des parametrischen Modells, eine präzisere Übereinstimmung der Kriterien. So weist generell die Verwendung von $t_{\text{mm}} = 320$ ms und $n_{\text{fb}} = 32$ mit aktivierter Anpassung der interauralen Kohärenz die geringsten Differenzbeträge zwischen Ist- und Soll-Wert auf. Nahezu alle Differenzen liegen unterhalb bekannter Wahrnehmbarkeitsschwellen der entsprechenden raumakustischen Kriterien. Weiterhin erweckt die Betrachtung der raumakustischen Maße den Eindruck, dass die Übergangszeit t_{mm} der bestimmende Faktor für die Differenzen zwischen BRIR und SBRIR ist. Die größten Differenzbeträge zeigen sich für alle Kriterien stets bei Verwendung von $t_{\text{mm}} = 20$ ms, bekannte JNDs nach DIN EN ISO 3382-1:2009-10 (2009) werden dabei in der Regel überschritten.

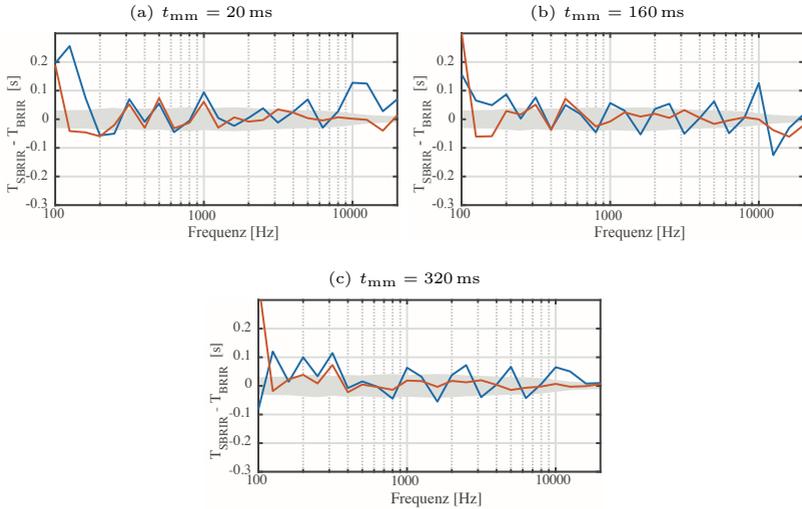


Abbildung 5.9: Differenz der Nachhallzeiten zwischen synthetischem und originalem Nachhall: $n_{fb} = 8$ (blau) und $n_{fb} = 32$ (rot), Wahrnehmbarkeitsgrenze Nachhall $\pm 4\%$ (grauer Bereich) nach Seraphim (1958).

Es treten geringe spektrale Unterschiede auf, lediglich bei Verwendung früher Übergangszeiten ($t_{mm} = 20$ ms) zeigen sich größere Differenzen (siehe Abbildung A.1). Für die Betrachtung des Spektrums gilt jedoch zu beachten, dass der identische Direktschall maßgeblich den Frequenzgang aufgrund des hohen Energiegehalts prägt.

Die Verläufe der Nachhallzeiten über der Frequenz stimmen im Allgemeinen für alle Bedingungen überein (siehe Abbildung A.2). Größere Differenzen finden sich in allen Bedingungen tieffrequent unterhalb von 125 Hz sowie zusätzlich bei Verwendung früher Übergangszeiten oberhalb von 12 kHz. Augenscheinlich nimmt die verwendete Filterbank einen hohen Einfluss auf die resultierende Nachhallzeit der SBRIR. Daher wird dieser Sachverhalt in Abbildung 5.9 genauer analysiert und die zwei verschiedenen Auflösungen der Filterbank gegenübergestellt. Die Verwendung der Oktav-Filterbank (blau) zeigt stärkere Schwankungen um den Soll-Wert, wohingegen die Filterbank mit 32 Frequenzbändern (rot) oberhalb von 125 Hz für nahezu alle Frequenzen Ist-Werte innerhalb der Wahrnehmbarkeitsschwelle von $\pm 4\%$ für Nachhall gemäß Seraphim (1958) generiert (graue Hüllkurve). Die Abweichungen bei Oktav-Auflösung sind partiell größer als 4 % und können somit vermutlich wahrgenommen werden. Es ergibt sich eine

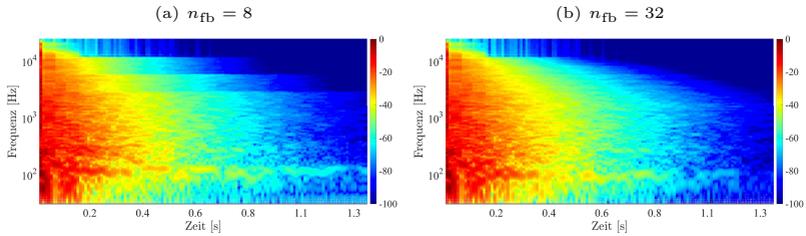


Abbildung 5.10: Vergleich der Auflösung der Filterbank mit 8 (a) oder 32 (b) Frequenzbändern mittels Spektrogramm.

„Treppenstruktur“, die auch im Spektrogramm in Abbildung 5.10 a deutlich sichtbar wird. Die reale Nachhallzeit wird in der Synthese mit 32 Frequenzbändern und hoher Güte insbesondere für hohe Frequenzen präziser adaptiert (siehe 5.10 b). Für die Übereinstimmung der frühen Abklingzeit zeigt sich generell derselbe Sachverhalt, jedoch ist folgerichtig der Faktor der Übergangszeit weitaus relevanter als die Frequenzauflösung der Filterbank (siehe Abbildung A.3).

Schwerpunktzeit (siehe A.4), Deutlichkeit (siehe A.6) und das Klarheitsmaß (siehe A.5) der SBRIR zeigen keine besonderen Auffälligkeiten und nähern sich mit steigender Übergangszeit und Komplexität des Algorithmus weiter an Werte der Referenz an. Die Übergangszeit nimmt dabei mehr Einfluss als die Frequenz-Auflösung oder die Kohärenzanpassung. Insbesondere für $t_{\text{mm}} = 160$ und 320 ms sind die Unterschiede als sehr gering zu bewerten, entsprechende Wahrnehmbarkeitsschwellen werden unterschritten. Die Auswertung der Klangfarbe des Nachhalls anhand der Kriterien des Bass- und Höhenverhältnisses (BR/TR) zeigt keine nennenswerten Unterschiede zwischen Messung und Modell sowie unter den Bedingungen.

Aufgrund der Integrationszeiten von $\text{IACC}_{\text{early}}$, weist ausschließlich die Verwendung von $t_{\text{mm}} = 20$ ms unterhalb von 400 Hz große Unterschiede auf, für alle weiteren Bedingungen sind keine nennenswerten Differenzbeträge zu verzeichnen (siehe A.7). Die Analyse mittels $\text{IACC}_{\text{late}}$ offenbart Unterschiede aufgrund der Anpassung der interauralen Kohärenz (siehe A.8). Eine aktivierte Adaption entsprechend der Ohrsignale führt für alle Bedingungen zu einer präziseren Übereinstimmung der Verläufe unterhalb von 400 Hz. Vor allem frühe Übergangszeiten profitieren von diesem Algorithmus. Die Betrachtung der Differenzen des Verlaufs der interauralen Kohärenz über der Frequenz gemäß Abbildung 5.11 verdeutlicht diesen Sachverhalt weiter. Mit IC-Anpassung (rot) stimmen Referenz-BRIR und SBRIR tieffrequent deutlich besser überein.

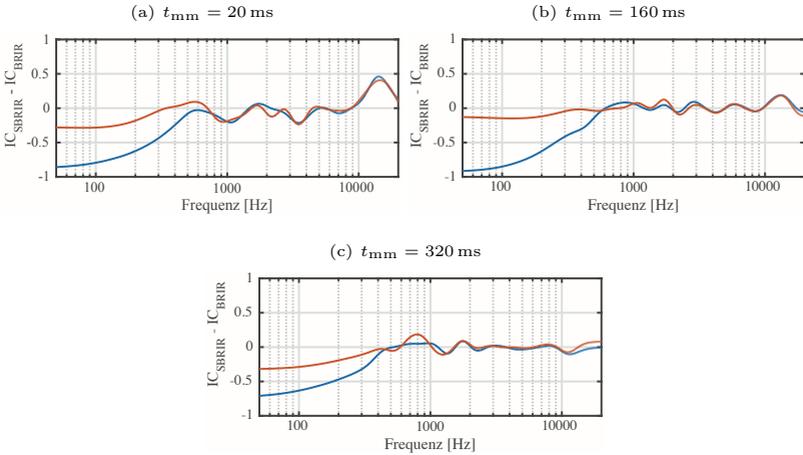


Abbildung 5.11: Differenz der interauralen Kohärenz zwischen synthetischem und originalem Nachhall: IC-Anpassung inaktiv (blau) und IC-Anpassung aktiv (rot).

So dokumentiert die technische Analyse mittels raumakustischer Kriterien einen Gewinn für das Modell sowohl durch Nutzung der detaillierten Frequenz-Auflösung als auch durch Aktivierung der Anpassung der interauralen Kohärenz der Rauschsignale an die binaurale Referenz. In der Regel werden bei Verwendung der Übergangszeiten von 160 und 320 ms die Wahrnehmbarkeitsschwellen der Kriterien unterschritten. Zu beachten gilt jedoch, dass es sich hierbei um einen Mittelwert über der Frequenz handelt. So treten für nahezu alle Kriterien weiterhin Frequenzbereiche auf, in denen bekannte JNDs überschritten werden.

LEA2: In Tabelle 5.7 finden sich die absoluten Differenzbeträge zwischen Referenz-BRIR und SBRIR im Mittel über alle Räumlichkeiten, Schallquellen und Frequenzen für die zwei untersuchten Übergangszeiten $t_{\text{mm early}}$ und $t_{\text{mm late}}$ des parametrischen Modells. Eine frequenzabhängige Betrachtung aller untersuchten Kriterien (im Mittel über Räumlichkeiten und Schallquellen) kann anhand der Plots gemäß Abbildung 5.12 erfolgen. Weiterhin finden sich in A.9 - A.16 eigenständige Abbildungen für alle Kriterien und akustische Situationen.

Im Allgemeinen zeigt die Verwendung der späten Übergangszeit $t_{\text{mm late}}$ eine zufriedenstellende Übereinstimmung aller untersuchten raumakustischen Kriterien über der

Tabelle 5.7: LEA2 - Absolute Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ der raumakustischen Kriterien im Mittel über alle Frequenzen, Vergleich zwischen $t_{\text{mm early}}$ und $t_{\text{mm late}}$.

Modell	Spektrum [dB]	<i>T30</i> [%]	EDT [%]	<i>IACC</i> _{early}	<i>IACC</i> _{late}
$t_{\text{mm early}}$	0,6	5,1	16,2	0,06	0,12
$t_{\text{mm late}}$	0,02	4,5	5,3	0	0,02
Modell	<i>T</i> _s [ms]	<i>C80</i> [dB]	<i>D50</i> [%]	<i>BR</i>	<i>TR</i>
$t_{\text{mm early}}$	4,8	1,5	4,6	0,025	0,027
$t_{\text{mm late}}$	2,5	0,2	0,9	0,038	0,025

Frequenz mit geringen Abweichungen vom Soll-Wert. Bei Betrachtung der Einzahlwerte werden für alle nahezu alle Kriterien die Wahrnehmbarkeitsschwellen laut DIN EN ISO 3382-1:2009-10 (2009) unterschritten. Lediglich die Nachhallzeit sowie die frühe Abklingzeit weisen Differenzbeträge auf, die leicht über bekannten JNDs liegen. Bei einer frequenz-sensitiven Betrachtung finden sich insgesamt die größten Differenzbeträge für die Nachhallzeit unterhalb von 125 Hz ($> 5\%$), für die frühe Abklingzeit unterhalb von 400 Hz ($> 5\%$) sowie für $IACC_{\text{late}}$ im Bereich zwischen 300 und 4000 Hz. Alle weiteren Frequenzbereiche und Kriterien weichen nur marginal vom Soll-Wert der Referenz ab.

Bei Verwendung der frühen Übergangszeit $t_{\text{mm late}}$ zeigen sich im Vergleich dazu für nahezu alle Kriterien deutlich größere Unterschiede zwischen Messung und Modell, entsprechende Wahrnehmbarkeitsschwellen (vor allem EDT und $IACC_{\text{late}}$) werden nicht erreicht und überboten. Lediglich die Nachhallzeit wird bei beiden Übergangszeiten nahezu identisch nachgebildet.

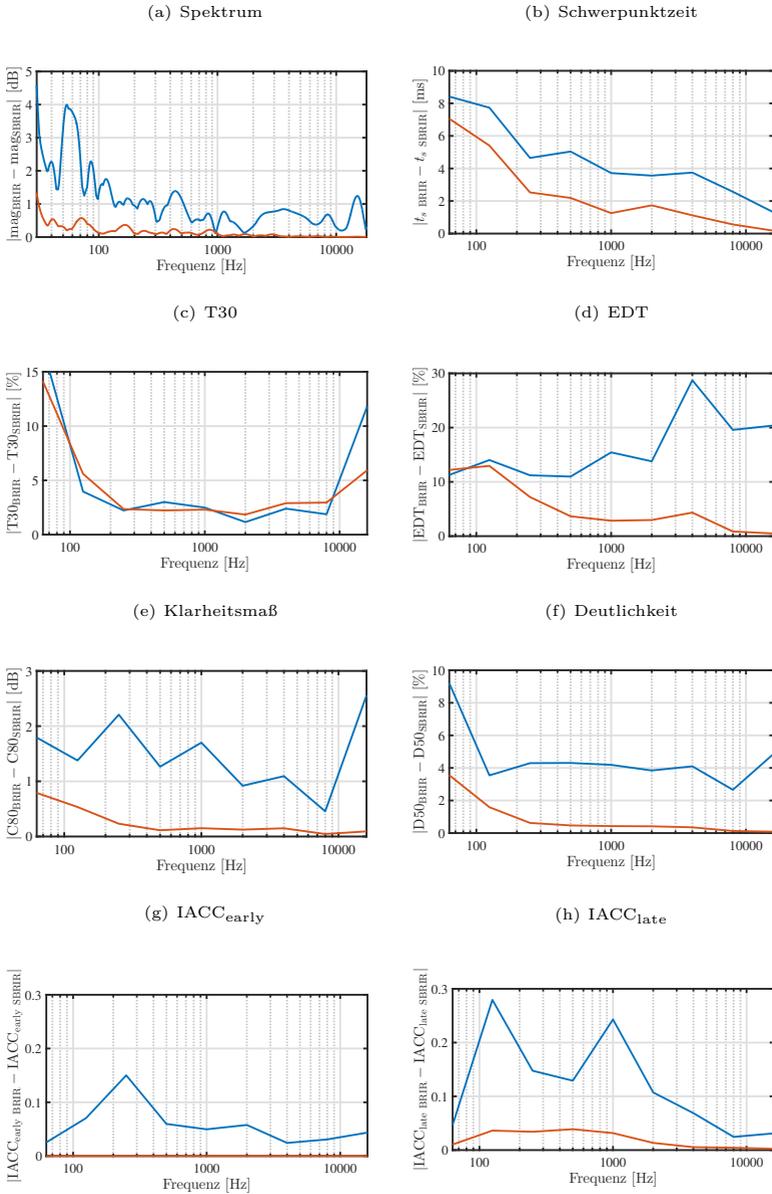


Abbildung 5.12: Experiment LEA2: Mittlere Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ über der Frequenz für $t_{\text{mm}}^{\text{early}}$ (blau) und $t_{\text{mm}}^{\text{late}}$ (rot).

5.3.4 Perzeptive Evaluation

Tabelle 5.8: Eigenschaften LEA1.

Experiment	LEA1	
Modell	SBRIR _{late}	
Kurzbeschreibung	Synthese der späten Anteile; Vergleich Algorithmen	
Versuchsdesign	ABC/HR	
Bedingungen	12	
Faktoren	3	
Faktorstufen		
Übergangszeit	3	$t_{mm} = 20/160/320$ ms
Filterbank	2	$n_{fb} = 8/32$ Frequenzbänder
IC	2	aktiv / inaktiv
Raum	1	KSS
Schallquelle	1	gerichtet
Testsignal	1	Schlagzeug
Wiederholungen	10	
Bewertungen / Proband	120	
Probanden	28 (25 m / 3 w)	19 - 36 Jahre / \varnothing 24,1 Jahre / σ 4 Jahre

LEA1: In Tabelle 5.8 finden sich die relevanten Eigenschaften des Experiments LEA1 mit Fokussierung auf Algorithmen der Signalverarbeitung zur Synthese binauraler, diffuser Schallfeldkomponenten. 28 (25 männlich, 3 weiblich) normal-hörende Probanden im Alter von 19 bis 36 Jahren (\varnothing 24,1 Jahre) nehmen an der Untersuchung teil, 24 können bereits Erfahrung in Hörversuchen mittels des verwendeten dynamischen Binauralsystems vorweisen. Eine detaillierte Erläuterung der Versuchsanordnung und der technischen Rahmenbedingungen findet sich in Kapitel 3.5 und 3.6.1. Gemäß Abschnitt 5.3.2 werden insgesamt 12 verschiedene synthetische Datensätze zur Auralisation verwendet und dem messtechnisch erfassten Datensatz mittels Faltung mit einem Schlagzeug Signal gegenübergestellt. Alle Bedingungen werden 10 mal wiederholt (randomisiert), sodass pro Proband in ca. 60 Minuten Hörversuch (mit Pause) 120 Bewertungen abgegeben werden.

Untersucht werden die drei Faktoren *Übergangszeit*, *Filterbank* und *IC*. Der Faktor *Übergangszeit* unterteilt sich in die drei Faktorstufen $t_{mm} = 20/160/320$ ms, *Filterbank* in die zwei Faktorstufen $n_{fb} = 8/32$ sowie *IC* ebenfalls in die zwei Stufen aktiv bzw. inaktiv. Für jeden Probanden werden gemäß ITU-R BS.1116-3 (2015) die *difference grades* der Bewertungen im Mittel über alle Wiederholungen durch Subtraktion der Bewertung der versteckten Referenz von der Bewertung der Bearbeitung (SBRIR) gebildet und im Weiteren statistisch analysiert.

Eine drei-faktorielle Varianzanalyse mit Messwiederholung (rmANOVA) wird zur Auswertung der drei Faktoren ($3 \times 2 \times 2$) verwendet (siehe Abschnitt 3.7.2). Gemäß eines

Tabelle 5.9: Zusammenfassung rmANOVA LEA1; bei $\alpha = 5\%$ signifikante Werte symbolisiert mit *; Freiheitsgrade nicht korrigiert, p -Werte korrigiert nach Greenhouse und Geisser (1959).

Faktor(en)	F-Wert	p-Wert	Effektstärke
Übergangszeit	$F(2,54) = 95,0$	$p < 0,001^*$	$\eta_p^2 = 0,78$
Filterbank	$F(1,27) = 66,7$	$p < 0,001^*$	$\eta_p^2 = 0,71$
IC	$F(1,27) = 45,9$	$p < 0,001^*$	$\eta_p^2 = 0,63$
Übergangszeit \times Filterbank	$F(2,54) = 27,4$	$p < 0,001^*$	$\eta_p^2 = 0,50$
Übergangszeit \times IC	$F(2,54) = 18,6$	$p < 0,001^*$	$\eta_p^2 = 0,41$
Filterbank \times IC	$F(1,27) = 30,9$	$p < 0,001^*$	$\eta_p^2 = 0,53$
Übergangszeit \times Filterbank \times IC	$F(2,54) = 1,08$	$p = 0,35$	$\eta_p^2 = 0,04$

Shapiro-Wilk-Test (Shapiro und Wilk, 1965) sind die Bewertungen pro Bedingung normalverteilt. Die Ergebnisse dieser statistischen Analyse der Bewertungen des Experiments LEA1 finden sich in Tabelle 5.9. Es zeigen sich signifikante Haupteffekte für die Faktoren *Übergangszeit*, *Filterbank* und *IC*. Weiterhin können signifikante Effekte für die Interaktion zwischen *Übergangszeit* und *Filterbank*, *Übergangszeit* und *IC* sowie für *Filterbank* und *IC* beobachtet werden. Einzig die Interaktion von allen drei Faktoren ist nicht signifikant. Die Haupteffekte können interpretiert werden, da alle Kombinationen der Faktoren ordinale Interaktionen (siehe Bortz und Schuster (2010), S.245) aufweisen. Angesichts der η_p^2 -Werte hat der Faktor der *Übergangszeit* zwar die höchste Effektstärke, aber trotzdem sollten die starken Interaktionseffekte weiter betrachtet werden. Daher werden im Folgenden weitere Post-Hoc-Tests aufgeführt, um die Ergebnisse besser interpretieren zu können. Insgesamt zeigen sich für die signifikanten Effekte entsprechend der Klassifizierung nach Cohen (1988) ausschließlich hohe Effektstärken.

Die Mittelwerte der Ergebnisse der perceptiven Untersuchung LEA1 über alle Probanden und Wiederholungen als *difference grades* finden sich in Abbildung 5.13 mit entsprechenden 95% Konfidenzintervallen (CI). Eine Übersicht über die Verteilung der Bewertungen gibt Abbildung 5.14 in Form eines Boxplots. Die Antennen (*whisker*) sind definiert als $Q \pm 1,5 \cdot IQR$. Dabei bezeichnet Q das untere bzw. obere Quartil sowie IQR den Interquartilsabstand, Ausreißer werden mit einem + Zeichen dargestellt. Die x-Achse beziffert die drei untersuchten Übergangszeiten t_{mm} , die y-Achse die jeweiligen *difference grades* der Bewertungen. Die Filterbank mit 8 Frequenzbändern wird mittels eines Kreises dargestellt, die Filterbank mit 32 Frequenzbändern mittels eines Dreiecks. In blau bzw. rot wird die inaktive oder aktive Anpassung der Kohärenz visualisiert.

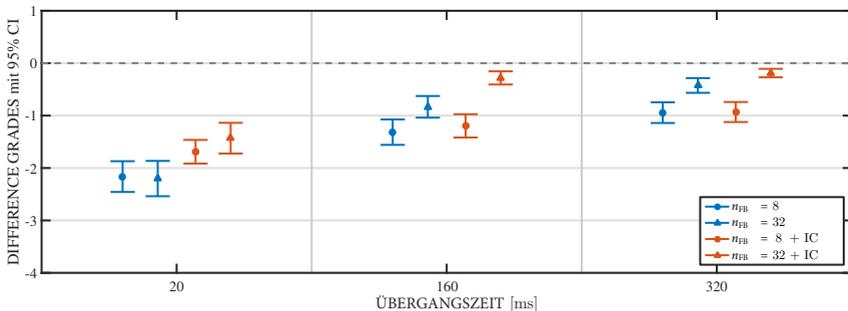


Abbildung 5.13: LEA1 - Mittelwerte der difference grades mit 95% CI (aus 1-Stichproben t -Test); Filterbank mit 8 Bändern (Kreis), Filterbank mit 32 Bändern (Dreieck), ohne (blau) und mit Kohärenzanpassung (rot) für verschiedene Übergangszeiten.

Insgesamt kann bei allen Bedingungen die Auralisation auf Basis der synthetischen BRIRs signifikant detektiert werden, sodass im Mittel ausschließlich negative difference grades auftreten. Alle Bewertungen liegen in der oberen Hälfte der Skala, insbesondere bei Verwendung der späten Übergangszeiten gar im oberen Drittel. Mittels paarweiser t -Tests mit Korrektur nach Hochberg (1988) können die einzelnen Bedingungen direkt miteinander verglichen werden, um signifikante Unterschiede aufgrund einzelner Signalverarbeitungsblöcke zu detektieren.

Die größten Beeinträchtigungen können für alle Bedingungen bei Verwendung der frühen Übergangszeit von $t_{mm} = 20$ ms beobachtet werden. Hier bewerten die Probanden die Auralisation auf Basis der SBRIRs mit der geringsten Ähnlichkeit im Vergleich zur Referenz. Ohne Anpassung der interauralen Kohärenz kann zudem kein signifikanter Unterschied aufgrund der Filterbank für diese Übergangszeit detektiert werden. Zwar führt die IC-Anpassung zu einer Verbesserung bei $t_{mm} = 20$ ms, die Mittelwerte der Bewertungen zeigen jedoch weiterhin größere Unterschiede zur Referenz, als bei $t_{mm} = 160$ oder 320 ms. Bei diesen Übergangszeiten führt die Verwendung der Filterbank mit 32 Frequenzbändern weiterhin zu einer signifikant höheren Ähnlichkeit im Vergleich zur Oktav-Filterbank, sodass diese Filterbank-Auflösung hier sinnvoll erscheint. Signifikante Unterschiede aufgrund der IC-Anpassung können bei Verwendung der Oktav-Filterbank und $t_{mm} = 160$ und 320 ms nicht beobachtet werden, wohingegen die Kombinationen aus $n_{FB} = 32$, $t_{mm} = 160$ oder 320 ms und aktivierter Kohärenzanpassung zu den höchsten perceptiven Bewertungen führen. Nur äußerst geringe Unterschiede zeigen sich für diese Konfigurationen zwischen den beiden Übergangszeiten. Zwar erreicht $t_{mm} = 320$ ms absolut betrachtet die höchste Ähnlichkeit, ein signifi-

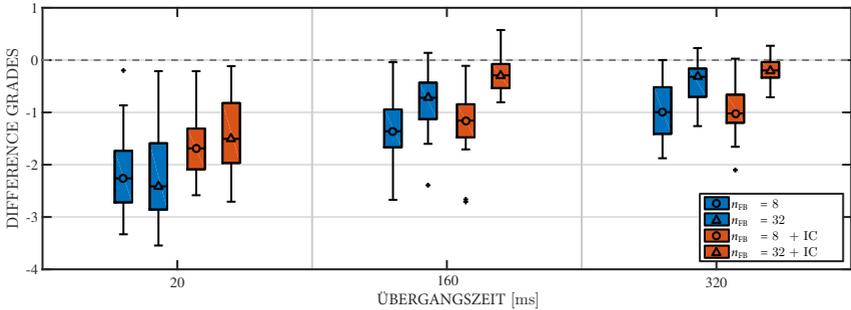


Abbildung 5.14: LEA1 - Verteilung der difference grades in Form eines Boxplots; Filterbank mit 8 Bändern (Kreis), Filterbank mit 32 Bändern (Dreieck), ohne (blau) und mit Kohärenzanpassung (rot) für verschiedene Übergangszeiten.

kanter Unterschied zu $t_{mm} = 160$ ms wird aber nicht erreicht. So erzeugt die höchste Komplexität des parametrischen Modells ($n_{fb} = 32$ und IC aktiv) für alle Übergangszeiten maßgeblich die am geringsten bewerteten Unterschiede zwischen Messung und Modell. Insbesondere bei $t_{mm} = 160$ und 320 ms werden nur äußerst geringe Differenzen festgestellt.

LEA2: In Tabelle 5.10 finden sich die relevanten Eigenschaften dieser perceptiven Evaluation des parametrischen Modells $SBRIR_{late}$ mit Fokussierung der Untersuchung auf die Synthese verschiedener akustischer Umgebungen. 22 (alle männlich) normalhörende Probanden im Alter von 20 bis 45 Jahren ($\bar{\varnothing}$ 25,6 Jahre) nehmen an der Untersuchung teil, 16 können bereits Erfahrung in Hörversuchen auf Basis des verwendeten dynamischen Binauralsystems vorweisen. Gemäß Abschnitt 5.3.2 werden insgesamt 8 verschiedene synthetische Datensätze zur Auralisation mittels einer Schlagzeug-Sequenz verwendet. Alle Bedingungen werden 10 mal wiederholt (randomisiert), sodass jeder Proband in ca. 45 Minuten Hörversuch (mit Pause) 80 Bewertungen abgeben muss.

Untersucht werden die drei Faktoren *Modell*, *Raum* und *Schallquelle*. Der Faktor *Modell* unterteilt sich in zwei Faktorstufen anhand zwei verschiedener Übergangszeiten t_{mm} *early* und t_{mm} *late*; Faktor *Raum* untersucht zwei Räumlichkeiten (KVB-Saal und kleiner Sendesaal) sowie Faktor *Schallquelle* die zwei Richtwirkungen gerichtet und omnidirektional.

Eine drei-faktorielle Varianzanalyse mit Messwiederholung (rmANOVA) wird zur Auswertung der drei Faktoren ($2 \times 2 \times 2$) verwendet (siehe Abschnitt 3.7.2). Gemäß eines

Tabelle 5.10: Eigenschaften LEA2.

Experiment	LEA2	
Modell	SBRR _{late}	
Kurzbeschreibung	Synthese der späten Anteile	
Versuchsdesign	ABC/HR	
Bedingungen	8	
Faktoren	3	
Faktorstufen		
Modell	2	$t_{\text{mm early}} / t_{\text{mm late}}$
Raum	2	KVB / KSS
Schallquelle	2	gerichtet / omnidirektional
Testsignal	1	Schlagzeug
Wiederholungen	10	
Bewertungen / Proband	80	
Probanden	22 (22 m)	20 - 45 Jahre / \varnothing 25,6 Jahre / σ 5,6

Shapiro-Wilk-Test (Shapiro und Wilk, 1965) sind die Bewertungen pro Bedingung normalverteilt. Die Ergebnisse dieser statistischen Analyse der Bewertungen des Experiments LEA2 finden sich in Tabelle 5.11. Es zeigen sich signifikante Haupteffekte für die Faktoren *Modell*, *Raum* und *Schallquelle*. Weiterhin kann ein signifikanter Effekt für die Interaktion zwischen *Schallquelle* und *Modell* beobachtet werden, im Vergleich zu den Haupteffekten jedoch mit erheblich geringerer Effektstärke. Alle weiteren Interaktionen sind nicht signifikant. Angesichts der η_p^2 -Werte hat der Faktor *Modell* die höchste Effektstärke, *Raum* und *Schallquelle* zeigen leicht niedrigere Werte. Gemäß Cohen (1988) sind die Effektstärken der Haupteffekte als hoch zu deuten. Um die Ergebnisse besser interpretieren zu können, werden im Folgenden paarweise t -Tests zum direkten Vergleich aller Bedingungen aufgeführt.

Die Ergebnisse der perzeptiven Untersuchung LEA2 über alle Probanden und Wiederholungen als *difference grades* finden sich in Abbildung 5.15, eine Übersicht über die Verteilung der Bewertungen gibt ein Boxplot in Abbildung 5.16. Die x-Achse beziffert die vier untersuchten akustischen Gegebenheiten (Raum/Schallquelle) die y-Achse die jeweiligen *difference grades* der Bewertungen. Die frühe Übergangszeit $t_{\text{mm early}}$ wird blau, die späte Übergangszeit $t_{\text{mm late}}$ rot visualisiert.

Bei allen untersuchten Bedingungen kann die Auralisation auf Basis synthetischer BRIRs signifikant detektiert werden. Alle Bewertungen liegen im oberen Drittel der Skala. Die frühe Übergangszeit wird in allen akustischen Gegebenheiten mit einem signifikant größeren Unterschied zur Referenz evaluiert im Vergleich zur späten Übergangszeit. Der kleine Sendesaal wird tendenziell etwas ähnlicher bewertet als der KVB-Saal, obwohl die Raumgröße durch spätere Übergangszeiten kompensiert wurde. In beiden Räumlichkeiten wird die omnidirektionale Schallquelle signifikant unterschied-

Tabelle 5.11: Zusammenfassung rmANOVA LEA2; bei $\alpha = 5\%$ signifikante Werte symbolisiert mit *; Freiheitsgrade nicht korrigiert, p -Werte korrigiert nach Greenhouse und Geisser (1959).

Faktor(en)	F-Wert	p-Wert	Effektstärke
Modell	$F(1,21) = 47,2$	$p < 0,001^*$	$\eta_p^2 = 0,69$
Raum	$F(1,21) = 30,0$	$p < 0,001^*$	$\eta_p^2 = 0,59$
Schallquelle	$F(1,21) = 24,0$	$p < 0,001^*$	$\eta_p^2 = 0,53$
Modell \times Raum	$F(1,21) = 0,39$	$p = 0,54$	$\eta_p^2 = 0,02$
Modell \times Schallquelle	$F(1,21) = 4,70$	$p = 0,04^*$	$\eta_p^2 = 0,18$
Raum \times Schallquelle	$F(1,21) = 0$	$p = 0,98$	$\eta_p^2 = 0$
Modell \times Raum \times Schallquelle	$F(1,21) = 0,78$	$p = 0,39$	$\eta_p^2 = 0,04$

licher als die gerichtete Schallquelle bewertet. Es kann angenommen werden, dass das Verhältnis zwischen Direktschall und spätem Nachhall hierbei den kritischen Faktor darstellt – bei geringerer Ratio (die bei Verwendung omnidirektionaler Schallquellen auftritt) wird die Auralisation auf Basis der SBRIRs eindeutiger detektiert.

5.3.5 Diskussion und Zusammenfassung

In diesem Abschnitt wurde das parametrische Modell $SBRIR_{late}$ zur Synthese der späten Anteile einer binauralen Raumimpulsantwort vorgestellt und erläutert sowie technisch anhand raumakustischer Kriterien und perceptiv im Rahmen von zwei Hörversuchen evaluiert. Die SBRIRs können zur dynamischen Auralisation verwendet werden und erlauben die Änderung der Orientierung des Rezipienten auf der horizontalen Ebene, bei ortsfester Wiedergabe der virtuellen Schallquellen. Eine Erweiterung um zusätzliche Freiheitsgrade ist weiterhin möglich. Das Modell generiert ausschließlich diffuse Schallfeldkomponenten und verwendet bis zu einer beliebig wählbaren Übergangszeit die messtechnisch erfasste Referenz. Ab der Übergangszeit wird die Referenz-Impulsantwort durch eine synthetische Impulsantwort ausgetauscht. Die Synthese basiert auf einem Ansatz nach Moorer (1979) und kombiniert diesen mit Untersuchungen nach Borß und Martin (2009) sowie Menzer (2010). Entgegen vieler Systeme in der Literatur finden FDN-Methoden keine Anwendung sowie es wird eine höhere Frequenzauflösung verwendet. Der Energieabfall und die interaurale Kohärenz der binauralen Referenz werden dazu weißen Rauschsignalen in bis zu 32 Bändern frequenzabhängig aufgeprägt. Weiterhin ermöglicht die parametrische Betrachtung eine beliebige Anpas-

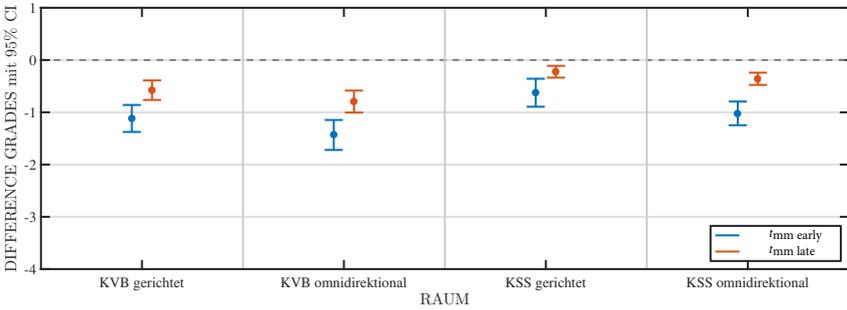


Abbildung 5.15: LEA2 - Mittelwerte der difference grades mit 95% CI (aus 1-Stichproben t -Test); $t_{\text{mm early}}$ (blau) und / $t_{\text{mm late}}$ (rot) für verschiedene Räume und Schallquellen.

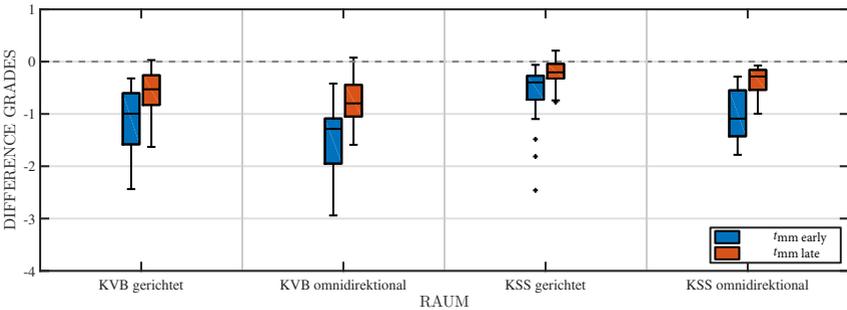


Abbildung 5.16: LEA2 - Verteilung der difference grades in Form eines Boxplots; $t_{\text{mm early}}$ (blau) und / $t_{\text{mm late}}$ (rot) für verschiedene Räume und Schallquellen.

sung des frequenzabhängigen Energieabfalls über der Zeit als auch der interauralen Kohärenz der Signale.

Zur Untersuchung der Algorithmen wurden SBRIR-Datensätze verschiedener Konfigurationen des Modells sowie unterschiedlicher akustischer Gegebenheiten erstellt. Im Allgemeinen zeigten sich mit steigender Übergangszeit und erhöhter Komplexität des Modells, geringere Unterschiede zwischen Messung und Modell. So hängen Beeinträchtigungen der Auralisation mittels SBRIRs stark von der in der Synthese verwendeten Frequenzauflösung als auch der Anpassung der interauralen Kohärenz ab. Insbesondere bei hohen Frequenzen ist eine detaillierte Frequenzauflösung vonnöten, um die Nachhallzeit akkurat adaptieren zu können und spektrale Unterschiede zwischen Referenz-

BRIR und SBRIR zu verringern. Die Kohärenzanpassung der Ohrsignale ist essentiell für eine adäquate Synthese tieffrequenter Anteile und der Vermeidung lateraler Hörereignisse durch inkohärente Rauschsignale. Daher wird in den folgenden Untersuchungen ausschließlich diese Konfiguration zur Synthese diffuser Schallfeldkomponenten verwendet. In Anbetracht der möglichen Verbesserung durch diese Signalverarbeitungsschritte bei nur wenigen zusätzlich benötigten Parametern, erscheint diese Vorgehensweise folgerichtig.

In beiden perceptiven Untersuchungen LEA1 und LEA2 konnte zwar ein signifikanter Unterschied zwischen der Auralisation auf Basis messtechnisch erfasster BRIRs und deren synthetischer Pendanten, die mit dem parametrischen Modell generiert wurden, beobachtet werden, jedoch sind die evaluierten Differenzen mitunter als gering aufzufassen. Im Allgemeinen stimmen die Ergebnisse der technischen und perceptiven Betrachtung weitestgehend überein. Weiterhin erreicht das Modell tendenziell vergleichbare Ergebnisse in den verschiedenen akustischen Situationen, obgleich ein signifikanter Einfluss von Raum und Schallquelle beobachtet werden kann. Auffällig ist der geringe Unterschied der perceptiven Bewertungen zwischen den beiden Übergangszeiten des Modells in Experiment LEA2. Obwohl die technische Analyse insbesondere für die Nachhallzeit, die frühe Abklingzeit und die interaurale Kreuzkorrelation größere Differenzen zwischen den Übergangszeiten aufzeigt, sind die perceptiven Unterschiede im Vergleich dazu als geringer zu interpretieren. Im Gegensatz zur Frequenzauflösung der Filterbank oder der Kohärenzanpassung liegt ein erhebliches Potential der Dateneinsparung bei der Wahl der Übergangszeit. Daher kann im Hinblick auf die spätere Anwendung des Modells, auch die Verwendung sehr früher Übergangszeiten möglich und sinnvoll sein. Im nächsten Schritt gilt es, die Vorgehensweise zu spiegeln und anstelle der späten Anteile einer BRIR, deren frühe Anteile auf Basis eines parametrischen Modells zu generieren und das Verfahren technisch sowie perceptiv zu evaluieren.

5.4 Frühe Anteile (LEB)

Das parametrische Modell $SBRIR_{\text{early}}$ befasst sich mit der Generierung der zeitlich frühen Anteile einer binauralen Raumimpulsantwort. Bis zu einer im Modell gewählten Übergangszeit t_{mm} werden die jeweiligen SBRIRs, die durch das parametrische Modell erzeugt werden, anstelle der Messung verwendet. Ab der Übergangszeit entsprechen die Impulsantworten wieder der originalen, messtechnisch erfassten Referenz-BRIR. So werden synthetische und messtechnisch erfasste BRIRs gemeinsam in einer Impulsantwort verknüpft. Das Modell wurde erstmalig von Stade und Arend (2017) vorgestellt und von Stade et al. (2017b) perceptiv evaluiert.

In Experiment LEB wird das Modell anhand verschiedener Räumlichkeiten, Schallquellen, Testsignale und Konfigurationen evaluiert. Allgemein soll untersucht werden, inwieweit sich die Auralisation unter Anwendung von BRIRs mit synthetischen, frühen Anteilen von einer Hörbarmachung auf Basis messtechnisch erfasster BRIRs unterscheidet und ob das gewählte Modell die reale, akustische Situation hinreichend charakterisiert.

5.4.1 Modell und Parameter

Tabelle 5.12: Gerichtete Parameter des parametrischen Modells und benötigte Werte in Abhängigkeit von der Anzahl n der detektierten Maxima p_k und der verwendeten Filterlänge t_{FIR} .

Parameter	Beschreibung	Werte
$Az_{p_k(n)}$	Einfallswinkel in Azimutwinkel	n
$El_{p_k(n)}$	Einfallswinkel in Elevationswinkel	n
$t_{p_k(n)}$	Ankunftszeit	n
$dB_{p_k(n)}$	Amplitude	n
$t_{p_k(n)}$	spektraler Verlauf der gerichteten Komponente	$n \cdot t_{\text{FIR}}$

In Abbildung 5.17 findet sich das Blockschaltbild des parametrischen Modells $\text{SBRIR}_{\text{early}}$ zur Generierung binauraler Raumimpulsantworten mit synthetischen frühen Anteilen, welches Experiment LEB zu Grunde liegt. Das Modell betrachtet diffuse und gerichtete Komponenten des Schallfeldes separat und analysiert, parametrisiert und synthetisiert diese in zwei isolierten Signalverarbeitungswegen.

Die Verwendung diffuser Schallfeldkomponenten auch für den frühen Anteil einer binauralen Raumimpulsantwort verfolgt das Ziel, die synthetischen Reflexionen besser zu verbinden. Aufgrund der Signalverarbeitung und Reduzierung auf wenige gerichtete Anteile, können mitunter zwischen einzelnen Reflexionen Bereiche mit geringer Energie erzeugt werden. Dies hat zur Folge, dass möglicherweise die Echoschwelle unterschritten wird, wodurch starke klangliche Beeinträchtigungen resultieren. Daher wird versucht werden, die diffusen Schallfeldanteile als Maskierer einzusetzen.

Die Signalverarbeitung der Analyse der diffusen Schallfeldkomponenten entspricht dem zuvor erläuterten und evaluierten Modell $\text{SBRIR}_{\text{late}}$ und verwendet aufgrund der Erkenntnisse aus LEA1 eine Filterbank mit 32 Frequenzbändern sowie adaptiert die Kohärenz der Rauschsignale entsprechend der binauralen Referenz. So wird die identische Verarbeitungskette, die in LEA1 und LEA2 zur Generierung binauraler, diffuser Schallfeldkomponenten Anwendung fand, aufgrund überwiegend zufriedenstellender perzeptiver Ergebnisse nun ebenfalls für den zeitlich frühen Bereich der Impulsantworten ver-

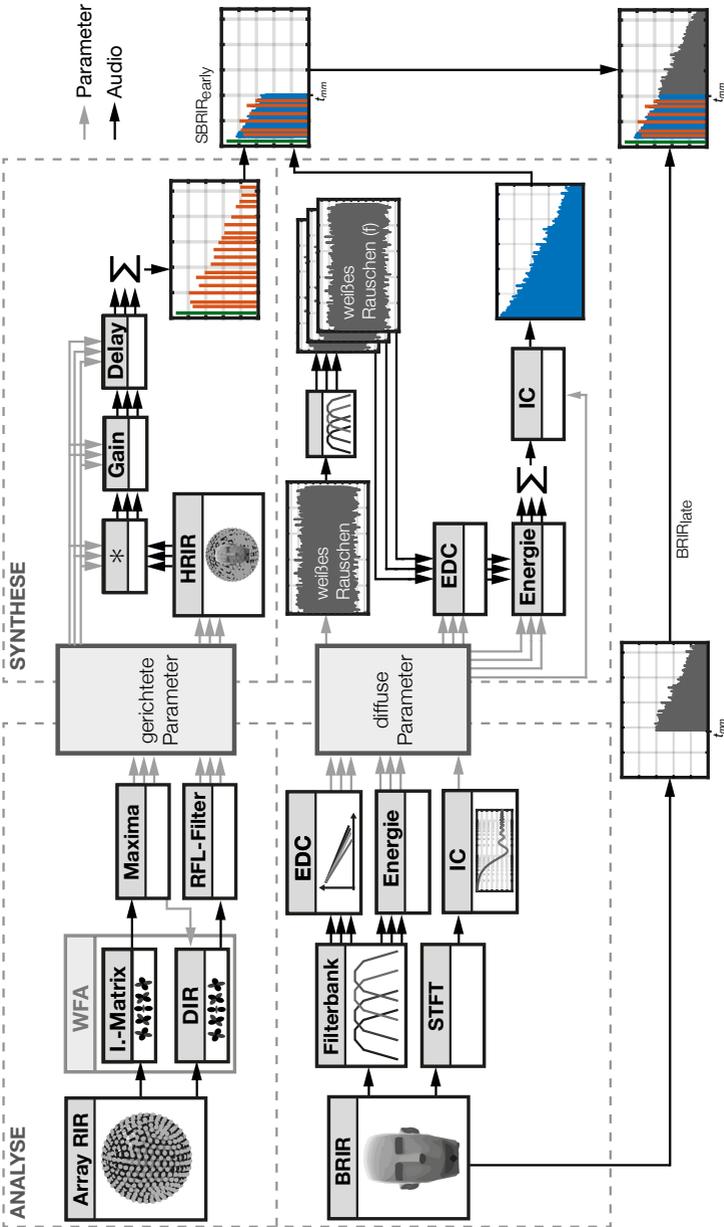


Abbildung 5.17: Blockschaltbild SBRIR_{early}: Gerichtete Komponenten werden auf Basis von Mikrofonarray-Daten mit Wellenfeldanalyse (WFA) detektiert und zur Synthese mittels sphärischer HRIRs verwendet. Diffuse Komponenten werden in einer Referenz-BRIR detektiert und zur Synthese mittels sphärischer HRIRs verwendet. Diffuse Komponenten werden in einer Referenz-BRIR detektiert und zur Synthese mittels sphärischer HRIRs verwendet. Synthetische Reflexionen (rot) und diffuser Nachhall (blau) werden zusammengefügt und bis zur Übergangszeit t_{nm} als frühe Anteile der SBRIR verwendet, ab t_{nm} wird die Referenz-BRIR (grau) eingesetzt (Stade et al., 2017b).

wendet. Das Modell basiert weiterhin auf für verschiedene Orientierungswinkel (in 1° horizontaler Auflösung, 360 BRIRs), messtechnisch erfassten, binauralen Raumimpulsantworten des zu synthetisierenden Raumes. Abermals bedarf es – unter der Annahme eines diffusen Schallfeldes – für die Analyse der diffusen Komponenten lediglich einer Orientierung im Raum, sodass BRIRs in Blickrichtung zur Anregungsquelle ($\phi = 0^\circ$, $\theta = 0^\circ$) verwendet werden. Die parametrische Beschreibung der diffusen Komponenten und deren Anwendung zur Generierung von synthetischen BRIRs mit ausschließlich diffusen Anteilen entspricht ebenfalls dem Modell aus Abschnitt 5.3.

Für die Analyse der gerichteten Komponenten des Schallfeldes werden die in Abschnitt 4.1.1 vorgestellten Methoden mittels WFA verwendet. Daher benötigt dieser Verarbeitungsblock einen Datensatz von Raumimpulsantworten, der im zu synthetisierenden Raum mit einem sphärischen Mikrofonarray gewonnen wurde. Um Einschränkungen in der Analyse hinsichtlich der Präzision durch unzureichend aufgelöste Datensätze zu vermeiden, wird zunächst mit nahezu idealen Messdaten gearbeitet. So basiert die Untersuchung auf Array-Datensätzen mit 1202 Impulsantworten auf einer Lebedev-Quadratur und einem Radius von 8,75 cm einer schallharten Kugel. Daraus resultieren für die vorgestellten Anwendungen keine spektralen Beschränkungen und/oder Ungenauigkeiten in der räumlichen Auflösung, da diese Array-Konfiguration mittels WFA zu stabilen Datensätzen im gesamten hörbaren Bereich mit hoher modaler Ordnung führt (siehe Kapitel 3.1.1).

Die Analyse der gerichteten Komponenten detektiert mit Hilfe der Schallfeldzerlegung in ebene Wellen (PWD), zeit- und ortssensitiv n Intensitäts-Maxima im Schallfeld im Raum. Die Maxima $p_k(n)$ repräsentieren den Direktschall und am Hörort einfallende Reflexionen. Auf Basis dieser Maxima kann die spektrale Filterwirkung der gerichteten Anteile, durch Berechnung gerichteter Impulsantworten und Bildung entsprechender Reflexions-Filter, bestimmt werden. Somit hängt der benötigte Speicherbedarf für die Parameter der gerichteten Komponenten direkt von der Anzahl der detektierten Reflexionen ab. Die parametrische Beschreibung charakterisiert das gerichtete Schallfeld anhand dieser Maxima $p_k(n)$ mit den Parametern gemäß Tabelle 5.12.

In der vorliegenden Arbeit wird zur Erzeugung der zeit- und richtungssensitiven Intensitätsmatrix mit einer Blocklänge von 24 Samples gearbeitet. Bei einer Abtastfrequenz von 48 kHz resultiert so eine zeitliche Auflösung von 0,5 ms. Um eine hohe örtliche Auflösung zu erreichen, wird zur Zerlegung des Schallfeldes in ebene Wellen eine Lebedev-Quadratur mit 3074 Punkten eingesetzt. Weiterhin wird für die Schallfeldzerlegung eine modale Ordnung von $N = 5$, eine Frequenz von 4500 Hz und eine maximale modale Verstärkung von 10 / 18 dB für die Radialfilter verwendet (siehe 4.1.1.1). Diese Konfiguration führt im KVB-Saal / kleinen Sendesaal zur Detektion von ca. 200 bzw. 50 Reflexionen.

Zur Generierung der gerichteten Impulsantworten (siehe 4.1.1.2) wird die folgende Konfiguration verwendet: Die Fensterlänge t_w für die zeitliche Segmentierung beträgt 192 Samples und damit 4 ms bei einer Abtastfrequenz von 48 kHz. Für die ebene Wellenzerlegung wird ebenfalls mit der Ordnung $N = 5$ gearbeitet. Bei der Generierung der FIR-Filter der gerichteten Komponenten (siehe 4.1.1.3) findet eine Unterscheidung in der Signalverarbeitung zwischen Direktschall und Reflexionen statt. In der Entwicklung zeigten sich unterschiedliche Anforderungen, sodass eine verschiedenartige Ausführung der Filter sinnvoll ist. Für den Direktschall wird ein minimalphasiges FIR-Filter mit Von-Hann-Fenster gebildet. Damit können sogenannte *Pre-Ringing* Artefakte bei tiefen Frequenzen wirksam vermieden werden. Für die Reflexionen wird mit linearphasigen FIR-Filtern (ebenfalls mit Von-Hann-Fenster) gearbeitet. Aufgrund der spektralen Zusammensetzung der Reflexionen, sind deren tiefe Frequenzen unproblematischer. Die Phase der Filter ist jedoch weitaus kritischer, da die Reflexionen teilweise zeitlich dicht beieinander liegen, sodass sich auch die Filter überlappen. Daher werden linearphasige FIR verwendet, um spektrale Auslöschungen durch unterschiedliche Phasengänge zu vermeiden. Bei der gemeinsamen Verwendung von linear- und minimalphasigen Filtern muss eine Kompensation der unterschiedlichen Latenz der Filter erfolgen. Die Filterlänge t_{FIR} des Systems kann variabel gestaltet werden, verwendet werden in dieser Arbeit 8192 Samples.

Die Synthese der gerichteten und diffusen Schallfeldanteile auf Basis der Parameter erfolgt gemäß Abschnitt 4.2.2 und 4.1.2, unabhängig voneinander. Um eine ortsfeste Auralisation der virtuellen Schallquellen im Binauralrenderer zu ermöglichen, muss die Synthese der gerichteten Anteile entsprechend der beabsichtigten Auflösung der Orientierung für jede Blickrichtung separat erfolgen. Im letzten Verarbeitungsschritt werden gerichtete und diffuse Komponenten überlagert und als zeitlich frühe Anteile der synthetischen Impulsantwort bis zur Übergangszeit verwendet. Analog zu LEA2 werden im KVB-Saal / kleinen Sendesaal Übergangszeiten von 320 bzw. 160 ms verwendet. Um eine Beeinträchtigung der Anfangszeitlücke (*initial time delay gap* - ITDG) zu vermeiden, werden die synthetischen, diffusen Anteile vor der ersten synthetischen Reflexion entfernt.

5.4.2 Datensätze

Im Folgenden werden die für die technische und perzeptive Untersuchung des parametrischen Modells $\text{SBRIR}_{\text{early}}$ generierten Datensätze erläutert. Um die Leistungsfähigkeit des Modells zu untersuchen, werden SBRIRs mit unterschiedlichen Konfigurationen der Algorithmen und auf Basis verschiedener akustischer Situationen gebildet. Es werden Bedingungen und Datensätze erstellt, mit denen Aussagen über die Wir-

Tabelle 5.13: Generierte Datensätze für LEB, abhängig von Modell, Schallquelle und Raum.

gerichtet	diffus	Raum	Schallquelle
×		KVB	direktional
	×	KVB	direktional
×	×	KVB	direktional
×		KVB	omnidirektional
	×	KVB	omnidirektional
×	×	KVB	omnidirektional
×		KSS	direktional
	×	KSS	direktional
×	×	KSS	direktional
×		KSS	omnidirektional
	×	KSS	omnidirektional
×	×	KSS	omnidirektional

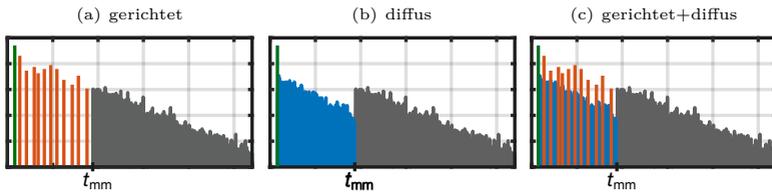


Abbildung 5.18: Schematische Darstellung der resultierenden Impulsantworten durch Synthese von gerichteten (a), diffusen (b) und gerichteten+diffusen (c) Komponenten aus Stade et al. (2017b). Identische Synthese des Direktschalls sowie Übereinstimmung mit der Referenz ab t_{mm} .

kung der Konfiguration der Algorithmen sowie den Einfluss verschiedener Räume und Schallquellen auf das parametrische Modell und damit auf das Ergebnis der Synthese getroffen werden können. Insbesondere die Notwendigkeit diffuser und gerichteter Schallfeldkomponenten sowie deren Kombination gilt es in diesem Experiment zu untersuchen. Weiterhin soll als Erweiterung zu den vorherigen Experimenten das Modell nun auch anhand verschiedenartiger Testsignale evaluiert werden.

Für alle untersuchten, akustischen Gegebenheiten werden drei verschiedene Datensätze erzeugt. In Abbildung 5.18 sind diese Konfigurationen anhand resultierender Impulsantworten schematisch dargestellt. Es werden (a) SBRIRs mit ausschließlich gerichteten Komponenten, (b) SBRIRs mit ausschließlich diffusen Komponenten sowie (c) SBRIRs auf Basis des vollständigen parametrischen Modells $SBRIR_{early}$ mit gerich-

teten und diffusen Schallfeldanteilen generiert, sodass insgesamt 12 unterschiedliche SBRIR-Datensätze erzeugt werden. Der Direktschall wird basierend auf dem absoluten Maximum der Intensitätsmatrix synthetisiert und erlaubt so, mit allen Datensätzen die identische, ortsfeste Abbildung der Schallquelle in der Auralisation. Ein Ausschluss des Direktschalls bei der Synthese mit ausschließlich diffusen Komponenten schien aufgrund zu großer klanglicher Beeinflussung nicht sinnvoll. Zur Untersuchung des Einflusses verschiedener Testsignale, müssen keine zusätzlichen Datensätze erzeugt werden, da die Anwendung des Testsignals erst im Binauralrenderer erfolgt und damit die SBRIR-Datensätze nicht tangiert.

5.4.3 Technische Untersuchung

Tabelle 5.14: LEB - Absolute Differenzbeträge $|BRIR - SBRIR|$ der raumakustischen Kriterien im Mittel über alle Frequenzen, Räume und Schallquellen, Vergleich zwischen Modell Konfigurationen des Modells.

Modell		Spektrum	<i>T30</i>	EDT	<i>IACC</i> _{early}	<i>IACC</i> _{late}
Gerichtet	Diffus	[dB]	[%]	[%]		
×		4,8	21,8	111,2	0,07	0,06
	×	4,1	11,8	60,4	0,07	0,06
×	×	2,8	11,8	37,1	0,1	0,05
Modell		<i>T_S</i>	<i>C80</i>	<i>D50</i>	<i>BR</i>	<i>TR</i>
Gerichtet	Diffus	[ms]	[dB]	[%]		
×		15,1	2,8	13,1	0,05	0,09
	×	18,4	2,1	9,6	0,05	0,06
×	×	13,5	2,0	9,5	0,05	0,04

Im Folgenden werden die Ergebnisse der technischen Untersuchung des parametrischen Modells $SBRIR_{\text{early}}$ vorgestellt (siehe Tabelle 5.14 und Abbildung 5.20).

Im Allgemeinen zeigen sich zwischen BRIRs und SBRIRs große absolute Differenzbeträge der raumakustischen Kriterien. Für nahezu alle Maße und Bedingungen werden die bekannten Wahrnehmbarkeitsschwellen nach DIN EN ISO 3382-1:2009-10 (2009) (zum Teil deutlich) überschritten. Die größten Differenzen treten für die ausschließliche Verwendung gerichteter Anteile auf, die geringsten Unterschiede erzeugen Datensätze unter Anwendung des kompletten parametrischen Modells mit gerichteten und diffusen Komponenten. Die Differenzbeträge der Kriterien von Datensätzen mit ausschließlich diffusen Anteilen liegen im Mittel näher bei den Werten des kompletten Modells, als bei den Werten mit ausschließlich gerichteten Komponenten. Die Analyse der frühen Abklingzeit weist die absolut größten Differenzen aller untersuchten Kriterien auf. Dieser Sachverhalt ist jedoch nicht verwunderlich, da die Synthese speziell den für die Be-

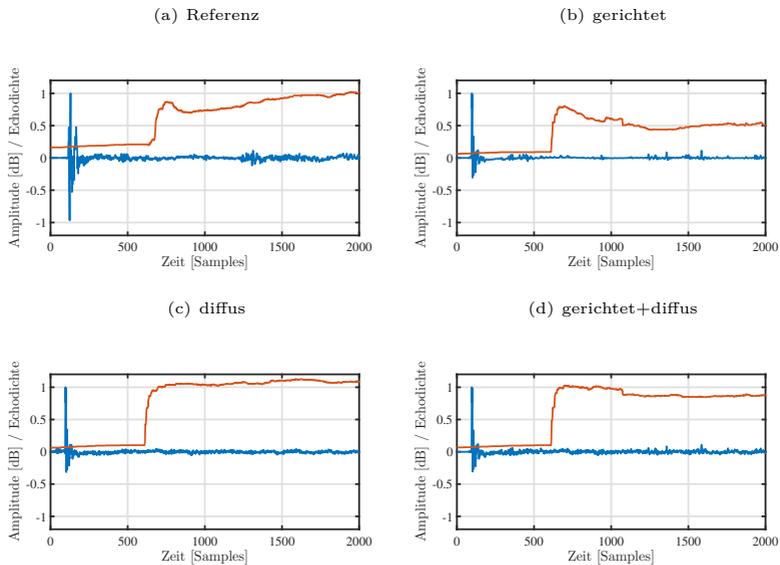


Abbildung 5.19: Echodichteprofil nach Abel und Huang (2006) für KVB-Saal und gerichteter Schallquelle; Vergleich zwischen Referenz (a) und Synthese mit gerichteten (b), diffusen (c) und gerichteten+diffusen (d) Komponenten.

rechnung der EDT relevanten Bereich betrifft. Weiterhin ist vermutlich der Übergang zwischen den synthetischen frühen und den messtechnisch erfassten späten Anteilen verantwortlich für die großen Differenzen zwischen den raumakustischen Maßen. Liegt hier kein idealer Übergang der Energieverhältnisse zwischen den beiden Impulsantworten vor, so führt dies insbesondere zu stark differierenden Werten für die Nachhallzeiten bzw. die frühen Abklingzeiten aufgrund der verwendeten Berechnungsmethoden.

Bei frequenzabhängiger Betrachtung der Räumlichkeiten im Detail gemäß Abbildungen A.17-A.28 (siehe Anhang) weist das Modell im kleinen Sendesaal unter Anwendung der gerichteten Schallquelle die geringsten Differenzen zur Messung auf. Im Gegensatz zu den vorherigen Untersuchungen LEA1 und LEA2 zeigen sich bei den Datensätzen aus LEB deutlich größere Unterschiede anhand der raumakustischen Maße. Zum einen betrifft eine Synthese der frühen Anteile einer BRIR maßgeblich die Kalkulationen der meisten Kriterien und zum anderen ist im Hinblick auf die Signalverarbeitung, ein Eingriff in die frühen Anteile des Schallfeldes als stärker zu beurteilen im Gegensatz zu den späten Anteilen. Weiterhin kann eine nicht ideale, energetische Verknüpfung von BRIR

und SBRIR an der Übergangszeit, die Berechnung der akustischen Maße beeinflussen und zu größeren Unterschieden zwischen Referenz und Modell führen.

Es ist zudem möglich, anhand des Echodichteprofiles nach Abel und Huang (2006) die Impulsantworten untereinander zu vergleichen. Abbildung 5.19 zeigt exemplarisch für den KVB-Saal mit gerichteter Schallquelle die Echodichteprofile der Referenz-BRIR sowie der drei verschiedenen SBRIRs (Betrachtung des linken Kanals). Offensichtlich erreicht bei der Referenz das Echodichteprofil innerhalb der ersten 2000 Samples den Wert 1, was nach Abel und Huang (2006) für ein vollkommen diffuses Schallfeld spricht. Bei Verwendung ausschließlich gerichteter Anteile wird dieser Wert (in diesem Zeitbereich) jedoch nie erreicht, das Schallfeld wird also folglich nie diffus, da dafür zu wenige Reflexionen synthetisiert werden. Bei Verwendung ausschließlich diffuser sowie gerichteter und diffuser Schallfeldkomponenten hingegen wird das Schallfeld im Vergleich zur Referenz deutlich früher diffus. So bestehen technisch betrachtet große Differenzen zwischen den messtechnisch erfassten Referenz-BRIRs und deren synthetischer Gegenstücke. Den Einfluss dieser Differenzen auf die menschliche Wahrnehmung gilt es nun in der folgenden perceptiven Untersuchung zu evaluieren.

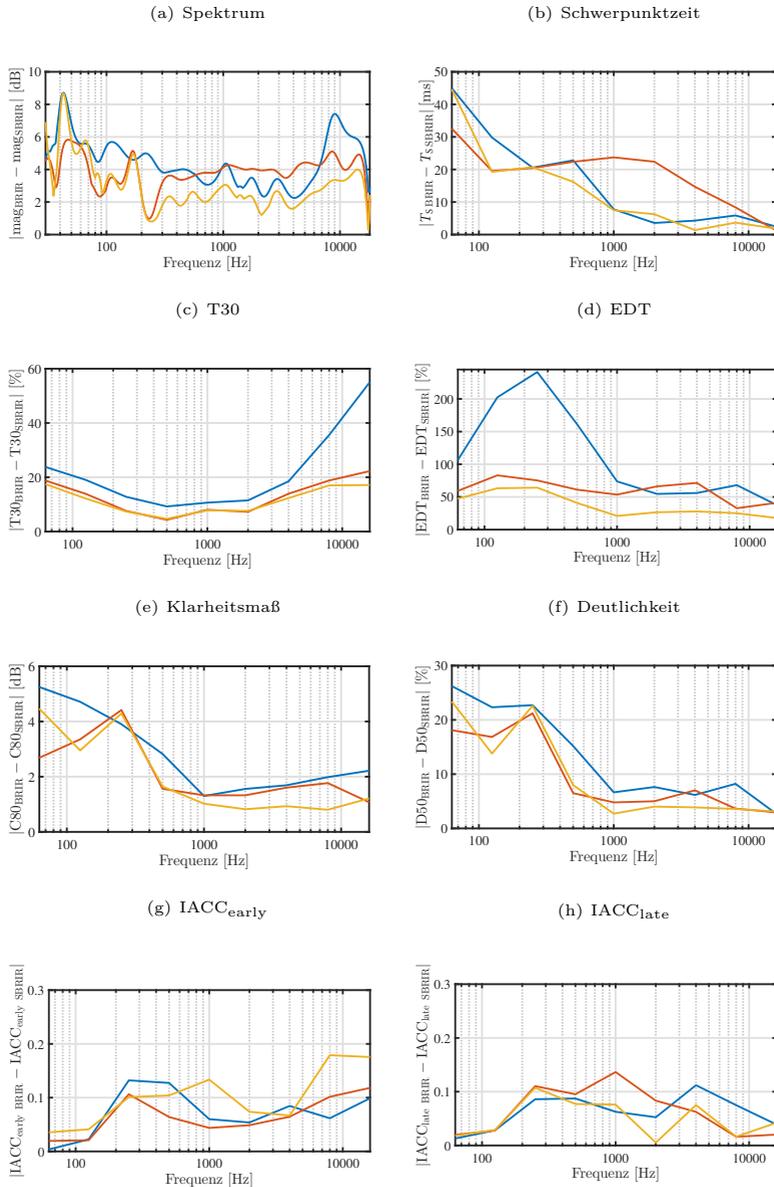


Abbildung 5.20: LEB: Mittlere Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ über der Frequenz für gerichtete (blau), diffuse (rot) und gerichtete+diffuse Anteile (gelb).

5.4.4 Perzeptive Evaluation

Tabelle 5.15: Eigenschaften LEB.

Experiment	LEB
Modell	SBRIR _{early}
Kurzbeschreibung	Synthese der frühen Anteile
Versuchsdesign	ABC/HR
Bedingungen	24
Faktoren	4
Faktorstufen	
Modell	3 gerichtet / diffus / gerichtet+diffus
Raum	2 KVB / KSS
Schallquelle	2 gerichtet / omnidirektional
Testsignal	2 Schlagzeug / Rauschimpulse (weiß)
Wiederholungen	5
Bewertungen / Proband	120
Probanden	32 (27 m / 5 w) 18 - 46 Jahre / $\bar{\sigma}$ 25,1 Jahre / σ 5,4 Jahre

In Tabelle 5.15 finden sich die relevanten Eigenschaften des Experiments LEB zur Untersuchung der perzeptiven Unterschiede zwischen Referenz-BRIRs und SBRIRs mit synthetischen, frühen Anteilen. 32 (27 männlich, 5 weiblich) normal-hörende Probanden im Alter von 18 bis 46 Jahren ($\bar{\sigma}$ 25,1 Jahre) nehmen an der Untersuchung teil, alle Probanden können bereits Erfahrung in Hörversuchen mittels des verwendeten dynamischen Binauralsystems vorweisen, da die Studie im Rahmen einer Reihe von Experimenten durchgeführt wird. Gemäß Abschnitt 5.4.2 werden insgesamt 12 verschiedene synthetische Datensätze für die Auralisation erzeugt und dem messtechnisch erfassten Datensatz mittels Faltung mit zwei verschiedenen Signalen (Schlagzeug/weiße Rauschimpulse) gegenübergestellt. Die resultierenden 24 Bedingungen werden je 5 mal wiederholt (randomisiert), sodass pro Proband in ca. 60 Minuten Hörversuch (mit Pause) 120 Bewertungen abgeben werden.

Untersucht werden die vier Faktoren *Modell*, *Raum*, *Schallquelle* und *Testsignal*. Der Faktor *Modell* unterteilt sich in drei Faktorstufen entsprechend der verschiedenen generierten Abstufungen des Modells nach gerichtet/diffus/gerichtet+diffus. Faktor *Raum* und *Schallquelle* sind abermals in zwei Faktorstufen (KVB-Saal/kleiner Sendesaal bzw. gerichtet/omnidirektional) unterteilt. Die Stufen des Faktors *Testsignal* entsprechen den beiden Audiosignalen (Schlagzeug/Rauschimpulse).

Eine vier-faktorielle Varianzanalyse mit Messwiederholung (rmANOVA) wird zur Auswertung der vier Faktoren ($3 \times 2 \times 2 \times 2$) verwendet (siehe Abschnitt 3.7.2). Gemäß eines Shapiro-Wilk-Test (Shapiro und Wilk, 1965) sind die Bewertungen pro Bedingung normalverteilt. Die Ergebnisse dieser statistischen Analyse der Bewertungen des Experiments LEB finden sich in Tabelle 5.16. Es werden signifikante Haupteffekte für

Tabelle 5.16: Zusammenfassung rmANOVA LEB; bei $\alpha = 5\%$ signifikante Werte symbolisiert mit *; Freiheitsgrade nicht korrigiert, p -Werte korrigiert nach Greenhouse und Geisser (1959).

Faktor(en)	F -Wert	p -Wert	Effektstärke
Modell	$F(2,62) = 518$	$p < 0,001^*$	$\eta_p^2 = 0,94$
Raum	$F(1,31) = 0$	$p = 0,97$	$\eta_p^2 = 0$
Schallquelle	$F(1,31) = 57,6$	$p < 0,001^*$	$\eta_p^2 = 0,65$
Testsignal	$F(1,31) = 4,83$	$p = 0,04^*$	$\eta_p^2 = 0,13$
Modell \times Raum	$F(2,62) = 19,5$	$p < 0,001^*$	$\eta_p^2 = 0,39$
Modell \times Schallquelle	$F(2,62) = 20,5$	$p < 0,001^*$	$\eta_p^2 = 0,40$
Modell \times Testsignal	$F(2,62) = 136$	$p < 0,001^*$	$\eta_p^2 = 0,81$
Raum \times Schallquelle	$F(1,31) = 40,3$	$p < 0,001^*$	$\eta_p^2 = 0,57$
Raum \times Testsignal	$F(1,31) = 130$	$p < 0,001^*$	$\eta_p^2 = 0,81$
Schallquelle \times Testsignal	$F(1,31) = 3,40$	$p = 0,07$	$\eta_p^2 = 0,10$
Modell \times Raum \times Schallquelle	$F(2,62) = 33,3$	$p < 0,001^*$	$\eta_p^2 = 0,52$
Modell \times Raum \times Testsignal	$F(2,62) = 2,51$	$p = 0,11$	$\eta_p^2 = 0,07$
Modell \times Schallquelle \times Testsignal	$F(2,62) = 1,71$	$p = 0,19$	$\eta_p^2 = 0,05$
Raum \times Schallquelle \times Testsignal	$F(1,31) = 7,21$	$p = 0,01^*$	$\eta_p^2 = 0,19$
Modell \times Raum \times Schallquelle \times Testsignal	$F(2,62) = 2,73$	$p = 0,07$	$\eta_p^2 = 0,08$

die Faktoren *Modell*, *Schallquelle* und *Testsignal* beobachtet. Der Faktor *Raum* ist hingegen nicht signifikant. Weiterhin werden verschachtelte rmANOVAs (*nested repeated measures ANOVA*) verwendet, um zu prüfen, ob immer ein Effekt für den Faktor *Modell* gezeigt werden kann. In jeder Betrachtungsweise der verschachtelten Varianzanalyse kann eine hohe Signifikanz für *Modell* gezeigt werden. Weiterhin bringt die Analyse signifikante Interaktionen zwischen einer Reihe von Faktoren zum Vorschein. Daher wird anhand von Interaktionsdiagrammen nach Bortz und Schuster (2010), S.245 überprüft, ob die Haupteffekte eindeutig interpretierbar sind. Da ausschließlich ordinale oder zumindest hybride Interaktion auftreten, kann eine Interpretation der Haupteffekte vorgenommen werden. Anhand der η_p^2 -Werte zeigen sich für die zwei Haupteffekte *Modell* ($\eta_p^2 = 0,94$) und *Schallquelle* ($\eta_p^2 = 0,65$) sehr hohe Effektstärken, der Faktor *Testsignal* erreicht einen geringen Wert ($\eta_p^2 = 0,13$).

Die Mittelwerte der Ergebnisse der perceptiven Untersuchung LEB über alle Probanden und Wiederholungen als *difference grades* finden sich in Abbildung 5.21 mit entspre-

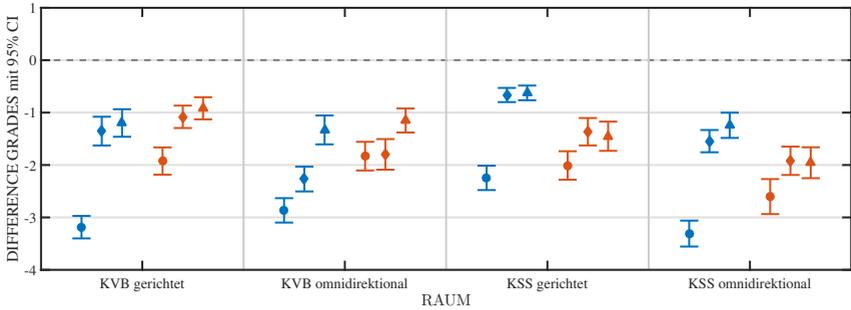


Abbildung 5.21: LEB - Mittelwerte der difference grades mit 95% CI (aus 1-Stichproben t -Test); gerichtete Anteile (Kreis), diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug (blau) und Rauschimpulse (rot) für verschiedene Räume und Schallquellen.

chenden 95% Konfidenzintervallen. Eine Übersicht über die Verteilung der Bewertungen gibt Abbildung 5.22 in Form eines Boxplots. Die x-Achse beziffert die vier untersuchten akustischen Gegebenheiten (Raum/Schallquelle), die y-Achse die jeweiligen difference grades der Bewertungen. Die Ergebnisse der Synthese mit ausschließlich gerichteten Schallfeldkomponenten werden mittels eines Kreises, die der Synthese mit ausschließlich diffusen Komponenten mittels einer Raute und die der Synthese auf Basis des kompletten Modells (gerichtet+diffus) mittels eines Dreiecks dargestellt. In verschiedenen Farben werden die Testsignale Schlagzeug (blau) und weiße Rauschimpulse (rot) visualisiert.

Die Auralisation auf Basis der SBRIRs wird in allen untersuchten Bedingungen detektiert und signifikant unterschiedlich im Vergleich zur Referenz-BRIR bewertet. Insgesamt liegt die Mehrzahl der Mittelwerte der Bewertungen in der oberen Hälfte der Skala, die Hälfte der Bedingungen zeigen gar Bewertungen im oberen Drittel der Skala. In nahezu allen untersuchten akustischen Gegebenheiten sind jedoch die Mittelwerte der Bearbeitung auf Basis der Synthese mit ausschließlich gerichteten Komponenten auffällig und zeigen im Gegensatz zu den zwei weiteren untersuchten Abstufungen des Modells deutlich größere Unterschiede im Vergleich zur Referenz.

Weiterhin werden paarweise t -Test mit Korrektur nach Hochberg (1988) als Post-Hoc-Analyse durchgeführt. Alle Abstufungen des parametrischen Modells werden direkt miteinander verglichen, um die Unterschiede besser interpretieren zu können. Die Synthese auf Basis gerichteter+diffuser Komponenten führt bei allen Räumen, Schallquellen und Testsignalen im Vergleich zur Synthese mit ausschließlich gerichteten Komponenten zu einer Bewertung von signifikant geringeren Unterschieden zwischen Referenz

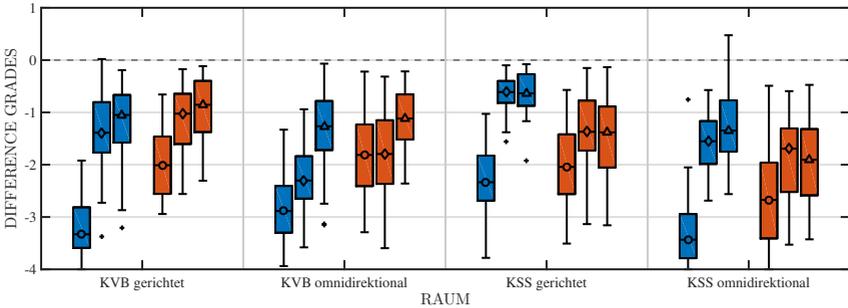


Abbildung 5.22: LEB - Verteilung der difference grades in Form eines Boxplots; gerichtete Anteile (Kreis), diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug (blau) und Rauschimpulse (rot) für verschiedene Räume und Schallquellen.

und Bearbeitung. Weiterhin zeigt die Synthese auf Basis diffuser Komponenten in 7 von 8 Bedingungen signifikant geringere Differenzen im Vergleich zur Referenz als die Synthese mit gerichteten Komponenten. In 3 von 8 Bedingungen werden Stimuli auf Basis der Synthese des kompletten Modells signifikant ähnlicher zur Referenz bewertet als Stimuli mit ausschließlich diffusen Schallfeldkomponenten. Außerdem deutet die Mehrheit der Bewertungen auf denselben Trend hin: In 14 von 16 Vergleichen zeigen die Mittelwerte der Auralisation auf Basis von gerichteten+diffusen Komponenten die größte Ähnlichkeit zur Referenz.

Bei der Verwendung der omnidirektionalen Schallquelle im KVB-Saal können bei beiden Testsignalen für die Synthese mit diffusen Komponenten auffallend große Unterschiede der Bewertungen im Vergleich zur Synthese mit gerichteten+diffusen Komponenten beobachtet werden. In allen weiteren Konditionen liegen die Mittelwerte der Bewertungen für diese zwei Abstufungen des Modells eng beieinander. Daher werden diese Bedingungen im Detail weiter untersucht, indem eine zusätzliche vier-faktorielle Varianzanalyse mit Messwiederholung ohne die Bewertungen der Synthese mit ausschließlich gerichteten Komponenten durchgeführt wird ($2 \times 2 \times 2 \times 2$). Auch diese Auswertung zeigt einen stark signifikanten Haupteffekt für den Faktor *Modell* ($F(1,31) = 99, p < 0,001, \eta_p^2 = 0,76$).

Die omnidirektionale Schallquelle wird im kleinen Sendesaal signifikant unterschiedlicher im Vergleich zur gerichteten Schallquelle bewertet, wohingegen im KVB-Saal kein signifikanter Einfluss der Richtcharakteristik der Schallquelle beobachtet werden kann. Weiterhin werden signifikante Unterschiede zwischen den Testsignalen festgestellt, jedoch ist ausschließlich bei der Synthese mit gerichteten Komponenten ein

Trend ersichtlich. Die Verwendung des Schlagzeug Signals führt hierbei zu einer stärkeren Beeinträchtigung im Vergleich zu den weißen Rauschimpulsen. Für die weiteren Bedingungen kann jedoch kein Trend eindeutig erkannt werden.

5.4.5 Diskussion und Zusammenfassung

Es wurde das parametrische Modell $SBRIR_{\text{early}}$ zur Synthese der frühen Anteile einer binauralen Raumimpulsantwort vorgestellt und erläutert, sowie technisch anhand raumakustischer Kriterien und perzeptiv im Rahmen eines Hörversuchs evaluiert. Entsprechend der Synthese der späten Anteile, kann auch die Synthese früher Anteile zur dynamischen Auralisation verwendet werden und erlaubt die Änderung der Orientierung des Rezipienten auf der horizontalen Ebene, bei ortsfester Wiedergabe der virtuellen Schallquellen. Eine Erweiterung um zusätzliche Freiheitsgrade ist möglich und erfordert ausschließlich wiedergabeseitig die Erfassung der Orientierung des Hörers in den entsprechenden Dimensionen. Die Synthese stellt bereits alle Funktionalitäten bereit, es bedarf lediglich der Bildung weiterer Impulsantworten entsprechend der beabsichtigten Auflösung. Das Modell detektiert, parametrisiert und generiert gerichtete und diffuse Schallfeldkomponenten unabhängig voneinander und verwendet diese bis zu einer beliebig wählbaren Übergangszeit anstelle der messtechnisch erfassten Referenz. Ab der Übergangszeit wird die Referenz-BRIR eingesetzt.

Die Synthese der diffusen Komponenten basiert, wie bereits in Abschnitt 5.3 vorgestellt, auf weißen Rauschsignalen, denen der Energieabfall sowie die interaurale Kohärenz einer binauralen Referenz frequenzabhängig aufgeprägt werden. Die Synthese der gerichteten Komponenten setzt Methoden der Wellenfeldanalyse zur richtungs-sensitiven Detektion von Reflexionen im Schallfeld auf Basis sphärischer Mikrofonarray-Messungen ein. So kann ein Echogramm mit zusätzlicher räumlicher Auflösung erstellt und zur Generierung von synthetischen Schallfeldanteilen auf Basis sphärischer HRIR-Datensätze verwendet werden. Diese Methode ist in der Literatur bisher selten zu finden, üblich werden zur Analyse von Reflexionen niedrigere Mikrofonordnungen verwendet, sodass Verfahren zur Detektion des Einfallzeitpunktes (TOA) mit Beamforming-Methoden zur Detektion der Einfallrichtung (DOA) kombiniert werden (z.B. in RSAO nach Remaggi et al. (2015)). Der Ansatz nach Melchior (2011) kommt den hier vorgestellten Methoden recht nah und basiert ebenfalls auf sphärischen Mikrofonarray-Messungen in Kombination mit einer Zerlegung in ebene Wellen zur Synthese von Raumimpulsantworten für binaurales Rendering. Hier zeigen sich für diese Methoden zwar zufriedenstellende Ergebnisse, es muss jedoch beachtet werden, dass die Bewertungen nicht im Vergleich zu einer binauralen Referenz erfolgten. Weiterhin unterteilen Melchior et al. (2010) die Impulsantworten zeitlich in drei Abschnitte (diskret, früh und spät).

Für eine systematische Untersuchung der Algorithmen wurden in dieser Arbeit SBRIR-Datensätze verschiedener Konfigurationen des Modells, sowie unterschiedlicher akustischer Gegebenheiten erstellt. Dabei galt es, die Verarbeitungsschritte der diffusen und gerichteten Komponenten sowohl separat als auch simultan zu evaluieren. Im Allgemeinen zeigten sich anhand der Auswertung raumakustischer Kriterien und der perceptiven Evaluation technische als auch wahrnehmbare Unterschiede zwischen Messung und Modell. Da ein sehr hoher Einfluss des Faktors *Modell* in der Varianzanalyse beobachtet wurde, muss dieser im Folgenden genauer betrachtet werden. Die größte Ähnlichkeit im Vergleich zur Referenz und damit die beste Performanz der drei untersuchten Abstufungen wurde bei Verwendung der gerichteten+diffusen Komponenten beobachtet. Alle dargebotenen Stimuli auf Basis dieser Synthese wurden im oberen Drittel der Skala bewertet. Überraschenderweise wurden Stimuli auf Basis ausschließlich diffuser Komponenten perceptiv größtenteils vergleichbar bewertet. Das bedeutet, für die Mehrzahl der untersuchten akustischen Bedingungen, konnte die Synthese mittels Direktschall in Kombination mit diffusem, binauralem Nachhall ebenfalls eine adäquate Wiedergabe erzielen. Bei dynamischer Auralisation ist die ortsfeste Wiedergabe des Direktschalls essentiell, da andernfalls das Headtracking keinen Einfluss auf die Auralization nehmen würde und dem Rezipienten keine Lokalisation der Schallquelle möglich wäre. Die technische Analyse zeigt für die Konfiguration mit ausschließlich diffusen Komponenten speziell für die frühe Abklingzeit, die Schwerpunkzeit und das Spektrum größere Abweichungen, die offenbar nur geringfügig in die perceptive Bewertung einfließen.

Der Unterschied zur Referenz wurde hingegen bei der Synthese mittels gerichteter Komponenten für alle Räume, Schallquellen und Testsignale erheblich größer bewertet im Vergleich zu den beiden anderen Konfigurationen. Diese Bewertung deckt sich mit der Betrachtung anhand raumakustischer Maße. Vermutlich war aufgrund der geringen Reflexionsdichte in diesem Zustand das Echogramm hörbar, sodass einzelne Reflexionen als Echos wahrgenommen wurden. Gespräche mit den Probanden im Anschluss an das Experiment, als auch die theoretische Betrachtung der zugrunde liegenden Impulsantworten unterstützen diese These. Entsprechend der höheren Empfindlichkeit des auditiven System für die Wahrnehmung von frühen Reflexionen bei transienten Signalen nach Olive und Toole (1989), wurden die größten Unterschiede für die Verwendung gerichteter Komponenten in Kombination mit dem Schlagzeug Signal beobachtet. Die angestrebte Maskierungs-Wirkung der diffusen Komponenten konnte gezeigt werden, da mit zusätzlicher Anwendung von Diffusschall für den frühen Anteil der Impulsantworten, keine Beeinträchtigung durch die synthetischen gerichteten Anteile beobachtet werden konnten.

Weiterhin konnten Variationen der Bewertungen aufgrund der Richtwirkung der Schallquelle festgestellt werden. Der vorgestellte Syntheseansatz zeigte die größten Unter-

schiede in Kombination mit der omnidirektionalen Schallquelle. Wie bereits zuvor kann angenommen werden, dass das Verhältnis zwischen Direktschall und später eintreffender Energie einen Einfluss hat. Bei geringeren Differenzen zwischen Direktschall und ersten Reflexionen ist die SBRIR leichter zu identifizieren. Der Einsatz der omnidirektionalen Schallquelle führt in beiden Räumlichkeiten zu einem geringen Energie-Verhältnis zwischen Direktanteilen und dem restlichen Schallfeld. Die frequenzabhängige Betrachtung des Klarheitsmaß $C80$ legt dies offen. Außerdem wird auffällig, dass ausschließlich bei Verwendung der omnidirektionalen Quelle, die Synthese mittels diffuser Komponenten signifikant größere perzeptive Unterschiede erzeugt. Abermals ist vermutlich das Energieverhältnis zwischen Direktschall und ersten Reflexionen entscheidend. Speziell in dieser akustischen Situation sind unter Betrachtung des Echogramms der Referenz BRIR zahlreiche starke Reflexionen zu verzeichnen. Daher kann für Räume mit starken, ersten Reflexionen und einem geringen DRR, der Gebrauch des Gesamt-Modells die perzeptiven Ergebnisse vermutlich weiter verbessern.

Für den Faktor *Testsignal* wurde im Vergleich zu den weiteren Haupteffekten *Raum* und *Schallquelle* eine erheblich kleinere Effektstärke beobachtet. So arbeitet das Modell in der Auralisation der untersuchten Räumlichkeiten zwar nicht unabhängig von den verwendeten Testsignalen, die Relevanz ist jedoch geringer im Gegensatz zu den weiteren Faktoren (ausgenommen für die Kombination von gerichteten Komponenten und Schlagzeug-Signal, siehe oben). So kann nun das Modell zur Synthese einer kompletten BRIR ausgeweitet werden. Die Synthese mit ausschließlich gerichteten Komponenten findet dabei aufgrund der großen perzeptiven Beeinträchtigung keine Anwendung.

5.5 Gesamtmodell (LEC)

Das parametrische Modell SBRIR befasst sich mit der Generierung von synthetischen binauralen Raumimpulsantworten auf Basis einer messtechnisch erfassten Referenz. Es kombiniert die zuvor vorgestellten und evaluierten Ansätze für die Synthese der zeitlich frühen (siehe Abschnitt 5.4) und späten (siehe Abschnitt 5.3) Anteile. Das Modell wurde von Stade et al. (2017a) vorgestellt und perzeptiv evaluiert. Entgegen den vorherigen Untersuchungen basiert nun die gesamte generierte Impulsantwort auf der Synthese mittels Parametern und es werden keine Bereiche der messtechnisch erfassten Referenz-BRIR in der SBRIR verwendet, sodass diese nun vollkommen synthetisch ist.

In Experiment LEC wird das Modell anhand verschiedener Räumlichkeiten, Schallquellen, Testsignale und Konfigurationen evaluiert. Im Allgemeinen gilt es, zu untersuchen, inwieweit sich die Auralisation unter Anwendung von SBRIRs von einer Hörbarmachung auf Basis messtechnisch erfasster BRIRs unterscheidet und ob das gewählte Mo-

dell die reale, akustische Situation hinreichend charakterisiert. Dazu werden zunächst das Modell an sich, dessen Konfiguration und die verwendeten Parameter erläutert und die für die systematische Untersuchung erzeugten Datensätze vorgestellt. Anschließend werden die synthetischen BRIRs technisch untersucht und den messtechnisch erfassten Pendanten gegenübergestellt, bevor sie dann in einem Experiment mit ABC/HR Testanordnung perzeptiv evaluiert werden.

5.5.1 Modell und Parameter

In Abbildung 5.23 findet sich das Blockschaltbild des parametrischen Modells SBRIR zur Generierung synthetischer binauraler Raumimpulsantworten, welches Experiment LEC zu Grunde liegt. Das Modell betrachtet diffuse und gerichtete Komponenten des Schallfeldes separiert und analysiert, parametrisiert und synthetisiert diese in zwei isolierten Signalverarbeitungswegen. Es findet keine zeitliche Trennung zwischen den Schallfeldanteilen statt. Die Synthese beider Komponenten erfolgt gleichsam für die gesamte Länge der Impulsantwort. Dies bedeutet, dass im Gegensatz zu den vorherigen Modellen, keine Übergangszeit zum Diffusfeld Anwendung findet. Es können sowohl gerichtete Anteile in zeitlich späten Bereichen der Impulsantwort detektiert, als auch diffuse Komponenten für die gesamte Länge der Impulsantwort erzeugt werden. Ausschließlich die zuvor vorgestellten Parameter der Modelle $\text{SBRIR}_{\text{late}}$ und $\text{SBRIR}_{\text{early}}$ werden zur Charakterisierung und Synthese verwendet.

Die Signalverarbeitung und parametrische Beschreibung der diffusen Schallfeldkomponenten entspricht damit den Methoden und Konfigurationen des Modells $\text{SBRIR}_{\text{late}}$. Weiterhin wird eine Filterbank mit 32 Frequenzbändern eingesetzt sowie die Kohärenz der Rauschsignale entsprechend der binauralen Referenz angepasst. Abermals bedarf es – unter der Annahme eines diffusen Schallfeldes – für die Analyse der diffusen Komponenten lediglich einer Orientierung im Raum, sodass BRIRs in Blickrichtung zur Anregungsquelle ($\phi = 0^\circ$, $\theta = 0^\circ$) als Referenz verwendet werden.

Die Signalverarbeitung und parametrische Beschreibung der gerichteten Schallfeldkomponenten entspricht den im Modell $\text{SBRIR}_{\text{early}}$ angewendeten Algorithmen. So basiert die Analyse des Schallfeldes im Hinblick auf eine Detektion der gerichteten Komponenten weiterhin auf Methoden der Wellenfeldanalyse mittels Raumimpulsantworten, die im zu synthetisierenden Raum mit einem sphärischen Mikrofonarray gewonnen wurden. Abermals wird ausschließlich mit sehr hoch aufgelösten Messdaten (1202 Lebedev) gearbeitet, um Einschränkungen in der Analyse hinsichtlich der Präzision durch unzureichend aufgelöste Datensätze zu vermeiden. Die Konfiguration der PWD entspricht ebenfalls dem Modell $\text{SBRIR}_{\text{early}}$. Es wird jedoch keine zeitliche Limitierung für die

Schallfeldzerlegung anhand einer Übergangszeit angewendet. So können Reflexionen für die gesamte Länge der Impulsantwort detektiert werden. In dieser Untersuchung werden für den KVB-Saal ca. 500 und für den kleinen Sendesaal ca. 100 synthetische Reflexionen generiert und zur Auralisation verwendet.

5.5.2 Datensätze

Im Folgenden werden die für die technische und perzeptive Untersuchung des parametrischen Modells *SBRIR* generierten Datensätze erläutert. Um die Leistungsfähigkeit des Modells zu untersuchen, werden *SBRIRs* mit unterschiedlichen Konfigurationen der Algorithmen und auf Basis verschiedener akustischer Situationen gebildet. Es sollen Bedingungen und Datensätze erzeugt werden, mit denen Aussagen über die Wirkung der Konfiguration der Algorithmen sowie den Einfluss verschiedener Räume, Schallquellen und Testsignale auf das perzeptive Modell und damit auf das Ergebnis der Synthese getroffen werden können. Insbesondere die Notwendigkeit diffuser und gerichteter Schallfeldkomponenten sowie deren Kombination gilt es zu untersuchen. Da in Experiment *LEB* für die ausschließliche Verwendung gerichteter Schallfeldkomponenten zur Synthese große Differenzen im Vergleich zur Referenz beobachtet werden konnten, wird diese Konfiguration im Folgenden nicht weiter evaluiert.

So werden für die untersuchten Räume und Schallquellen je zwei verschiedene Datensätze erzeugt. In Abbildung 5.24 sind die zwei verschiedenen untersuchten Konfigurationen anhand resultierender Impulsantworten schematisch dargestellt. Es werden (a) *SBRIRs* mit ausschließlich diffusen Komponenten und (b) *SBRIRs* auf Basis des vollständigen parametrischen Modells *SBRIR* mit gerichteten+diffusen Schallfeldanteilen generiert, sodass insgesamt 8 unterschiedliche Datensätze synthetischer *BRIRs* erzeugt werden. Der Direktschall basiert wie zuvor in allen Datensätzen auf dem absoluten Maximum der Intensitätsmatrix und erlaubt so eine ortsfeste Abbildung der virtuellen Schallquelle in der dynamischen Auralisation.

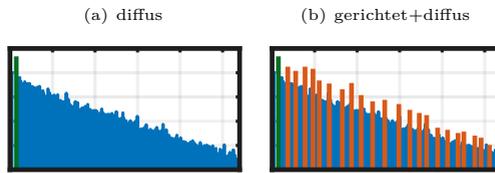
5.5.3 Technische Untersuchung

Im Folgenden werden die Ergebnisse der technischen Untersuchung des parametrischen Modells *SBRIR* vorgestellt (siehe Tabelle 5.18 und Abbildung 5.25).

Im Allgemeinen zeigen sich bei den ausgewerteten raumakustischen Kriterien nur geringe Differenzbeträge zwischen den Signalen. Bekannte Wahrnehmbarkeitsschwellen nach DIN EN ISO 3382-1:2009-10 (2009) werden zwar teilweise überschritten, entgegen den Ergebnissen aus *LEB* sind aber deutlich geringere Abweichungen zu verzeichnen.

Tabelle 5.17: Generierte Datensätze für LEC, abhängig von Modell, Raum und Schallquelle.

gerichtet	diffus	Raum	Schallquelle
	×	KVB	direktional
×	×	KVB	direktional
	×	KVB	omnidirektional
×	×	KVB	omnidirektional
	×	KSS	direktional
×	×	KSS	direktional
	×	KSS	omnidirektional
×	×	KSS	omnidirektional

**Abbildung 5.24:** Schematische Darstellung der resultierenden Impulsantworten durch Synthese diffusen (a) und gerichteten+diffusen (b) Komponenten aus Stade et al. (2017a).

Insbesondere die Nachhallzeit wird präzise in der Synthese adaptiert. Auffällig sind Abweichungen für die frühe Abklingzeit. Bei der frequenzabhängigen Betrachtung der untersuchten Konfigurationen des Modells, Räumlichkeiten und Schallquellen im Detail anhand der Abbildungen A.29-A.36 (siehe Anhang) wird jedoch der Grund dafür deutlich: Bei einigen akustischen Situationen zeigt die Referenz bei einer Frequenz von 16 kHz Werte der EDT, die nahezu bei 0 liegen. Die Synthese hingegen erzeugt Impulsantworten, die in diesem Frequenzband leicht höhere Werte der frühen Abklingzeit aufzeigen. So entstehen prozentual gesehen starke Abweichungen, die gleichsam Einfluss auf den frequenzunabhängigen Einzahlwert in Tabelle 5.18 nehmen. Wird dieser Frequenzbereich für die Kalkulation des Einzahlwertes ausgespart, so werden deutlich geringere Abweichungen dokumentiert. Ob sich Unterschiede in diesem Frequenzbereich auch perceptiv auswirken ist fraglich und muss näher untersucht werden.

Die Gegenüberstellung der Differenzbeträge der Konfigurationen des Modells zeigt keine klare Tendenz und allgemein vergleichbare Werte. Dabei finden sich sowohl Vorteile durch kleinere Differenzen für die Synthese mittels ausschließlich diffuser Komponenten

Tabelle 5.18: LEC - Absolute Differenzbeträge $|BRIR - SBRIR|$ der raumakustischen Kriterien im Mittel über alle Frequenzen, Räume und Schallquellen, Vergleich zwischen Konfigurationen des Modells.

Modell		Spektrum	$T30$	EDT	$IACC_{early}$	$IACC_{late}$
gerichtet	diffus	[dB]	[%]	[%]		
	×	3,6	3,9	165 (19,6)	0,1	0,1
×	×	2,3	4,9	190 (20,7)	0,1	0,08

Modell		T_S	$C80$	$D50$	BR	TR
gerichtet	diffus	[ms]	[dB]	[%]		
	×	13,2	2,2	8,8	0,08	0,03
×	×	14,3	2,1	8,5	0,04	0,06

($T30$, EDT, T_S , TR), als auch für das vollständige Modell (Spektrum, $IACC_{late}$, $C80$, $D50$, BR). Auch der direkte Vergleich der Abstufungen gemäß Abbildung 5.25 lässt für die untersuchten Kriterien keine eindeutige Aussage zu. Offensichtlich bestehen für die Kriterien nur geringe Unterschiede zwischen den zwei Variationen des Modells.

Beim Vergleich der Ergebnisse dieser Untersuchung mit den Differenzbeträgen der Kriterien aus **LEB** nach Tabelle 5.14 werden zum Teil große Unterschiede zwischen den Experimenten deutlich. Allgemein führt das Modell **SBRIR** zu Datensätzen mit geringeren Differenzbeträgen als das Modell **SBRIR_{early}**, obwohl dieses weiterhin für Teilbereiche der Impulsantworten die messtechnisch erfasste Referenz-BRIR verwendet. Hierfür ist aber vermutlich eine nicht ideale energetische Anpassung zwischen den diffusen Anteilen der Synthese und dem späten Nachhall der Messung in **SBRIR_{early}** verantwortlich, was eine starke Beeinflussung der Kalkulation der raumakustischen Maße hervorrufen kann. Dieses Problem tritt bei **SBRIR** nicht auf, da hier nur ein (synthetischer) diffuser Energieabfall für die gesamte Impulsantwort verwendet wird. Dadurch können keine Differenzen aufgrund nicht ideal zusammengefügter Abklingkurven auftreten. Mit einer spezifischen Anpassung der Energieverhältnisse, wäre dieser Sachverhalt in **LEB** vermeidbar gewesen. Es galt jedoch einen generischen Algorithmus zu verwenden, sodass eine gezielte Adaption der Prozesse nicht zugelassen wurde.

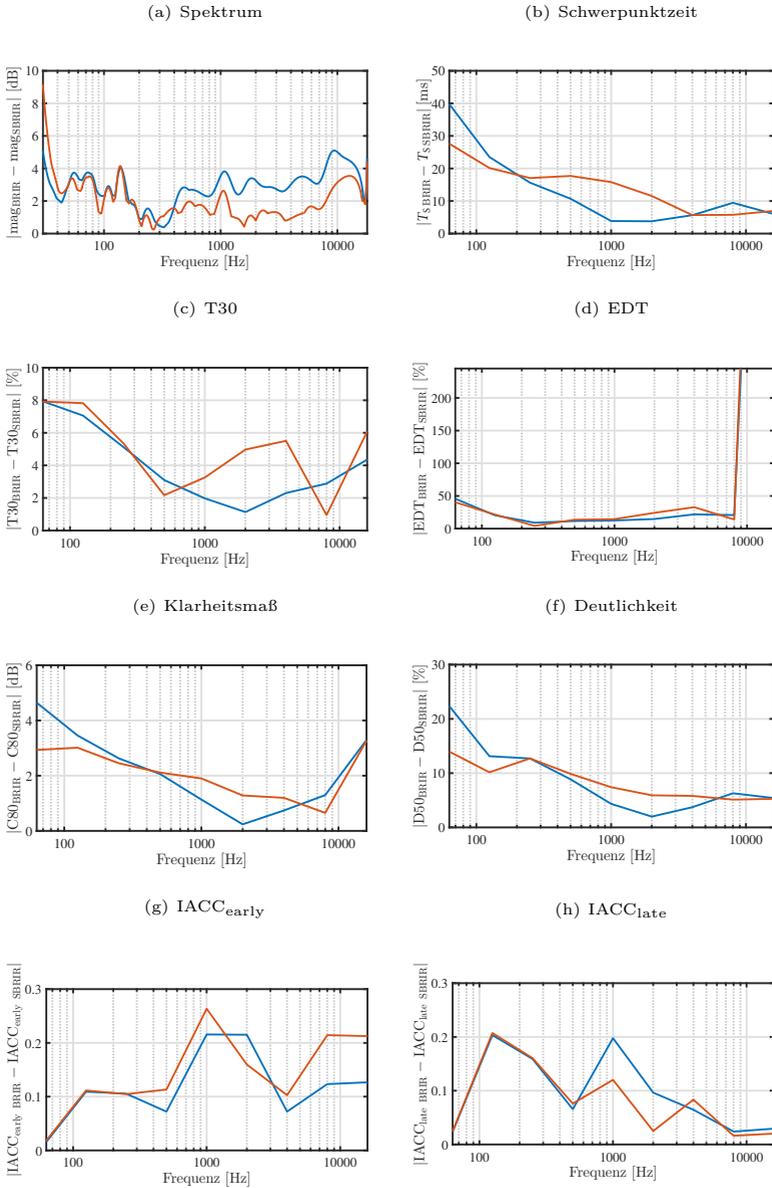


Abbildung 5.25: LEC: Mittlere Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ über der Frequenz für diffuse (blau) und gerichtete+diffuse Anteile (rot).

5.5.4 Perzeptive Evaluation

Tabelle 5.19: Eigenschaften LEC.

Experiment	LEC	
Modell	SBRIR	
Kurzbeschreibung	voll-synthetische BRIRs	
Versuchsdesign	ABC/HR	
Bedingungen	16	
Faktoren	4	
Faktorstufen		
Modell	2	diffus / gerichtet+diffus
Raum	2	KVB-Saal / KSS
Schallquellen	2	gerichtet / omnidirektional
Testsignal	2	Schlagzeug / Rauschimpulse (weiß)
Wiederholungen	8	
Bewertungen / Proband	128	
Probanden	28 (21 m / 7 w)	19 - 47 Jahre / $\bar{\mu}$ 25,8 Jahre / σ 5,3 Jahre

In Tabelle 5.19 finden sich die Eigenschaften des Experiments LEC zur Untersuchung der wahrnehmbaren Unterschiede zwischen einer Auralisation auf Basis der Referenz-BRIRs im Vergleich zu SBRIRs, die mit dem Modell generiert wurden. 28 (21 männlich, 7 weiblich) normal-hörende Probanden im Alter von 19 bis 47 Jahren ($\bar{\mu}$ 25.8 Jahre) nehmen an der Untersuchung teil. 22 Probanden können bereits Erfahrung in Hörversuchen mittels der verwendeten DBS vorweisen. Gemäß Abschnitt 5.5.2 werden insgesamt 8 verschiedene synthetische Datensätze für diese Studie generiert und in der Auralisation dem messtechnisch erfassten Datensatz mittels Faltung mit zwei verschiedenen Testsignalen (Schlagzeug/weiße Rauschimpulse) gegenübergestellt. Die resultierenden 16 Bedingungen werden je 8 mal wiederholt (randomisiert), sodass jeder Proband in ca. 60 Minuten Hörversuch (mit Pause) 128 Bewertungen abgeben muss.

Untersucht werden die vier Faktoren *Modell*, *Raum*, *Schallquelle* und *Testsignal*. Der Faktor *Modell* unterteilt sich in diesem Experiment in zwei Faktorstufen entsprechend der verschiedenen generierten Abstufungen des Modells nach diffus/gerichtet+diffus. Faktor *Raum* und *Schallquelle* sind unterteilt in zwei Faktorstufen (KVB-Saal/kleiner Sendesaal bzw. gerichtet/omnidirektional). Die Stufen des Faktors *Testsignal* entsprechen den beiden im Binauralrenderer zur Hörbarmachung eingesetzten Audiosignalen (Schlagzeug/Rauschimpulse).

Eine vier-faktorielle Varianzanalyse mit Messwiederholung (rmANOVA) wird zur Auswertung der vier Faktoren ($2 \times 2 \times 2 \times 2$) verwendet (siehe 3.7.2). Gemäß Shapiro und Wilk (1965) sind die Bewertungen pro Bedingung normalverteilt. Die Ergebnisse dieser statistischen Analyse der Bewertungen des Experiments LEC finden sich in Tabelle 5.20. Es werden signifikante Haupteffekte für die Faktoren *Modell*, *Schallquelle* und *Raum*

Tabelle 5.20: Zusammenfassung rmANOVA LEC; bei $\alpha = 5\%$ signifikante Werte symbolisiert mit *; Freiheitsgrade nicht korrigiert, p -Werte korrigiert nach Greenhouse und Geisser (1959).

Faktor(en)	F -Wert	p -Wert	Effektstärke
Modell	$F(1,27) = 15,6$	$p < 0,001^*$	$\eta_p^2 = 0,37$
Raum	$F(1,27) = 7,86$	$p = 0,009^*$	$\eta_p^2 = 0,23$
Schallquelle	$F(1,27) = 15,7$	$p < 0,001^*$	$\eta_p^2 = 0,37$
Testsignal	$F(1,27) = 1,38$	$p = 0,25$	$\eta_p^2 = 0,05$
Modell \times Raum	$F(1,27) = 14,8$	$p < 0,001^*$	$\eta_p^2 = 0,35$
Modell \times Schallquelle	$F(1,27) = 0,27$	$p = 0,61$	$\eta_p^2 = 0,01$
Modell \times Testsignal	$F(1,27) = 1,96$	$p = 0,17$	$\eta_p^2 = 0,07$
Raum \times Schallquelle	$F(1,27) = 65,7$	$p < 0,001^*$	$\eta_p^2 = 0,71$
Raum \times Testsignal	$F(1,27) = 43,3$	$p < 0,001^*$	$\eta_p^2 = 0,62$
Schallquelle \times Testsignal	$F(1,27) = 35,7$	$p < 0,001^*$	$\eta_p^2 = 0,57$
Modell \times Raum \times Schallquelle	$F(1,27) = 14,8$	$p < 0,001^*$	$\eta_p^2 = 0,35$
Modell \times Raum \times Testsignal	$F(1,27) = 0,69$	$p = 0,41$	$\eta_p^2 = 0,03$
Modell \times Schallquelle \times Testsignal	$F(1,27) = 0,83$	$p = 0,37$	$\eta_p^2 = 0,03$
Raum \times Schallquelle \times Testsignal	$F(1,27) = 7,34$	$p = 0,01^*$	$\eta_p^2 = 0,21$
Modell \times Raum \times Schallquelle \times Testsignal	$F(1,27) = 8,67$	$p = 0,007^*$	$\eta_p^2 = 0,24$

beobachtet. Der Faktor *Testsignal* ist nicht signifikant. Anhand der η_p^2 -Werte zeigen sich für die drei Haupteffekte *Modell* ($\eta_p^2 = 0,37$), *Schallquelle* ($\eta_p^2 = 0,37$) und *Raum* ($\eta_p^2 = 0,23$) hohe Effektstärken. Weiterhin bringt die Analyse teilweise stark signifikante Interaktionen zwischen einer Reihe von Faktoren zum Vorschein. Besonders die Interaktionen von *Raum* \times *Schallquelle*, *Raum* \times *Testsignal* und *Schallquelle* \times *Testsignal* zeigen hohe Effektstärken. Daher wird anhand von Interaktionsdiagrammen nach Bortz und Schuster (2010), S.245 überprüft, ob die Haupteffekte eindeutig interpretierbar sind. Die Interaktionsdiagramme für den Vergleich der Stufen der Faktoren *Modell* und *Raum* als auch der Vergleich der Stufen von *Modell* und *Schallquelle* ergeben ordinale oder zumindestens hybride Interaktion. Die Linienzüge im Vergleich der Stufen der Faktoren *Raum* und *Schallquelle* verlaufen jedoch teilweise disordinal, sodass sie nach Bortz und Schuster (2010), S.245 nur in Kombination interpretierbar sind.

Die Mittelwerte der Ergebnisse der perzeptiven Untersuchung LEC über alle Probanden und Wiederholungen als *difference grades* finden sich in Abbildung 5.26 mit entspre-

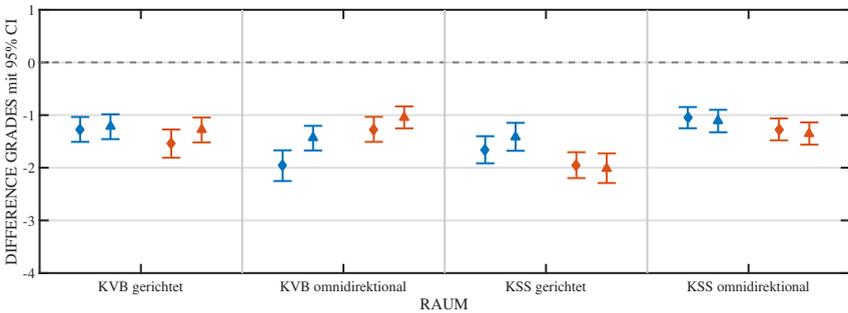


Abbildung 5.26: LEC - Mittelwerte der difference grades mit 95% CI (aus 1-Stichproben t -Test); diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug (blau) und Rauschimpulse (rot) für verschiedene Räume und Schallquellen.

chenden 95% Konfidenzintervallen. Eine Übersicht über die Verteilung der Bewertungen gibt Abbildung 5.27 in Form eines Boxplots. Die Mittelwerte der Ergebnisse der Synthese mit ausschließlich diffusen Komponenten werden abermals mittels Raute und die Ergebnisse des kompletten Modells (gerichtet+diffus) mittels Dreieck dargestellt. Verschiedene Farben visualisieren die zwei unterschiedlichen Testsignale Schlagzeug (blau) und weiße Rauschimpulse (rot).

Die Auralisation auf Basis der SBRIRs wird in allen untersuchten Bedingungen detektiert und signifikant unterschiedlich im Vergleich zur Referenz-BRIR bewertet. Insgesamt liegen nahezu alle Mittelwerte der Bewertungen im oberen Drittel der Skala. Weiterhin können auf den ersten Blick keine bemerkenswerten Unterschiede zwischen den untersuchten Bedingungen beobachtet werden, da die Mittelwerte der Bewertungen hierbei recht eng beieinander liegen.

Des Weiteren werden paarweise t -Tests mit Korrektur nach Hochberg (1988) als Post-Hoc-Analyse durchgeführt. Alle Faktorstufen von *Modell*, *Raum*, *Schallquelle* und *Testsignal* werden untereinander verglichen. Gemäß der Interaktionsdiagramme kann der Faktor *Modell* für sich alleine eindeutig interpretiert werden. In fünf von acht Bedingungen zeigt die Verwendung des Modells mit gerichteten+diffusen Komponenten in der Hörbarmachung geringere Unterschiede zur Referenz im Vergleich zu SBRIRs mit ausschließlich diffusen Komponenten. Es werden jedoch nur in drei Bedingungen (KVB-Saal gerichtet mit Rauschsignal, KVB-Saal omnidirektional mit Schlagzeug und KVB-Saal omnidirektional mit Rauschsignal) signifikante Unterschiede zwischen den beiden Abstufungen des Modells erreicht. So zeigt sich eine signifikante Verbesserung aufgrund der zusätzlichen Synthese von gerichteten Anteilen ausschließlich im größeren Raum

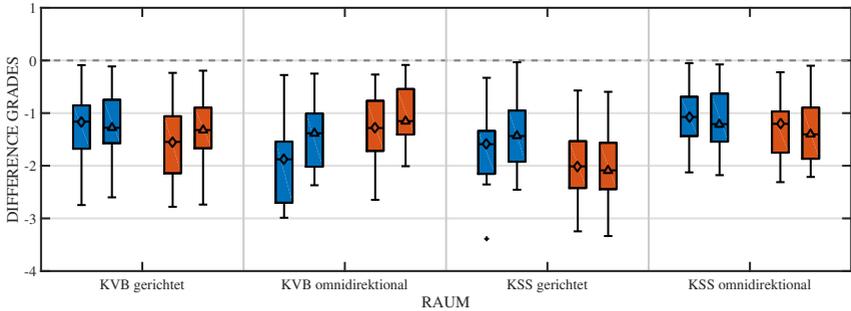


Abbildung 5.27: LEC - Verteilung der difference grades in Form eines Boxplots; diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug (blau) und Rauschimpulse (rot) für verschiedene Räume und Schallquellen.

(KVB-Saal). Die Verwendung von diffusen Komponenten wird im Vergleich zum Gesamtmodell nie signifikant ähnlicher zur Referenz bewertet. Dies deutet entsprechend der Ergebnisse aus Experiment LEB erneut an, dass für eine adäquate Darbietung mittels synthetischer BRIRs das abklingende diffuse Schallfeld essentiell ist, jedoch abhängig von den akustischen Gegebenheiten eine zusätzliche Synthese diskreter Reflexionen die Auralisation aufwerten kann. Die Unterschiede zwischen den Mittelwerten der Bewertungen der beiden Abstufungen sind sehr gering. Auch diese Evaluation deckt sich mit vorherigen Erkenntnissen. Für die Mehrzahl der untersuchten Bedingungen erreicht somit auch die Synthese mit diffusen Komponenten eine gute Abbildung der Räumlichkeiten. Es muss betont werden, dass beide Abstufungen eine ortsfeste Auralisation des Direktschalls ermöglichen, da dieser richtungsabhängig synthetisiert wird. Informelle Hörversuche dokumentierten im Vorfeld eine starke Beeinträchtigung der Auralisation ohne dynamische Nachführung des Direktschalls.

Aufgrund disordinaler Interaktionen kann der Faktor *Schallquelle* und dessen Stufen nicht eindeutig interpretiert werden, der Faktor *Raum* muss in die Betrachtung einbezogen werden. In 6 der 8 Bedingungen wird die omnidirektionale Schallquelle signifikant ähnlicher als die gerichtete Schallquelle bewertet, die verbleibenden zwei Bedingungen zeigen signifikant höhere Bewertungen für die gerichtete Schallquelle (KVB-Saal mit Schlagzeug). Da signifikante Bewertungen wechselweise für beide Schallquellen beobachtet werden, ist kein klarer Trend für den Faktor *Schallquelle* erkennbar im Gegensatz zum Faktor *Modell*. Die Mehrzahl der *t*-Tests zum Vergleich der Schallquellen ergeben weiterhin geringe *p*-Werte.

Auch der Faktor *Raum* kann aufgrund der Interaktionseffekte nicht für sich alleine betrachtet werden. In 6 von 8 Gegenüberstellungen werden signifikant unterschied-

liche Bewertungen für diesen Faktor anhand paarweiser t -Tests beobachtet. Dabei wird der größere Raum (KVB-Saal) in vier Bedingungen (diffus mit gerichteter Quelle und Schlagzeug; diffus+gerichtet mit gerichteter Schallquelle und Rauschsignal; diffus mit gerichteter Schallquelle und Rauschsignal; diffus+gerichtet mit omnidirektionaler Schallquelle und Rauschsignal) signifikant ähnlicher als der kleine Sendesaal bewertet. Unter Anwendung des Schlagzeug-Signals und der omnidirektionalen Schallquelle wird die Auralisation des kleinen Sendesaals mittels SBRIRs bei beiden Abstufungen des Modells signifikant ähnlicher im Vergleich zur Referenz bewertet. Analog zum Faktor *Schallquelle* kann somit auch für den Faktor *Raum* keine eindeutige Aussage aufgrund starker Interaktionseffekte und wechselseitiger Signifikanz getroffen werden. Wiederholt werden zudem ausschließlich geringe p -Werte beobachtet.

5.5.5 Diskussion und Zusammenfassung

In diesem Abschnitt wurde das parametrische Modell SBRIR zur voll-synthetischen Generierung von binauralen Raumimpulsantworten auf Basis einer messtechnisch erfassten Referenz vorgestellt. Das Modell und die zugrunde liegende Signalverarbeitung wurden erläutert, sowie technisch anhand raumakustischer Kriterien und perzeptiv im Rahmen eines Hörversuchs evaluiert. Die so erzeugten SBRIRs können zur dynamischen Auralisation von Räumen verwendet werden und erlauben bei Änderung der Orientierung des Rezipienten auf der horizontalen Ebene, eine ortsfeste Wiedergabe des Schallfeldes entsprechend der Realität. Eine Erweiterung um zusätzliche Freiheitsgrade ist weiterhin leicht durchführbar. Das Modell detektiert, parametrisiert und generiert gerichtete und diffuse Schallfeldkomponenten unabhängig voneinander und überlagert beide Anteile zur Bildung der synthetischen Impulsantworten.

Die Signalverarbeitung des Modells SBRIR entspricht einer Kombination der vorherigen Modelle SBRIR_{late} und SBRIR_{early}. Die Synthese der diffusen Komponenten basiert weiterhin auf weißen Rauschsignalen, denen der Energieabfall sowie die interaurale Kohärenz der binauralen Referenz frequenzabhängig aufgeprägt werden. Die Synthese der gerichteten Komponenten verwendet Methoden der Wellenfeldanalyse zur richtungs-sensitiven Detektion von Reflexionen im Schallfeld auf Basis sphärischer Mikrofonarray-Messungen und generiert mit Hilfe von HRIR-Daten synthetische Reflexionen entsprechend der detektierten gerichteten Parameter.

Es findet keine zeitliche Trennung zwischen Reflexionen und diffusen Anteilen statt, beide Komponenten werden für die gesamte Impulsantwort synthetisiert, sodass gleichsam der diffuse binaurale Nachhall im frühen Bereich verwendet wird als auch Reflexionen zu späten Zeitpunkten generiert werden können. Dies vermeidet Beeinträchtigungen der

Hörbarmachung durch nicht ideal gewählte Übergangszeiten in ein diffuses Schallfeld. Wie bereits gezeigt, ist diese Wahl nicht trivial und mitunter nicht zufriedenstellend. Daher sieht das grundsätzliche Konzept der vorgestellten Methode keine zeitliche Trennung zwischen frühen Reflexionen und spätem Nachhall anhand einer Übergangszeit vor. Es wird zwar beobachtet, dass zu späten Zeitpunkten deutlich weniger diskrete Reflexionen detektiert werden, jedoch werden durch diese Vorgehensweise keine späten Reflexionen (in Konzertsälen beispielsweise typischerweise Rückwürfe der Rückwand) übergangen. Im Modell nach Coleman et al. (2017a) (RSAO) findet sich ein ähnlicher Ansatz, auch hier werden diffuse Anteile bereits zu frühen Zeitpunkten in den Impulsantworten verwendet. Der diffuse Nachhall beginnt ebenfalls ab der ersten detektierten Reflexion, die Abklingkurven werden jedoch mittels einer Fensterfunktion eingeblendet und erreichen erst zur mixing time den gewünschten Pegel.

Zur Untersuchung des Modells wurden SBRIR-Datensätze verschiedener Variationen der Algorithmen sowie unterschiedlicher akustischer Gegebenheiten erstellt. Dabei galt es, wie in Experiment LEB die Verarbeitungsschritte der diffusen und gerichteten Komponenten sowohl separat als auch simultan zu evaluieren. Ausschließlich gerichtete Komponenten wurden aufgrund vorheriger Evaluation weiterhin nicht zur Synthese verwendet. Im Allgemeinen zeigen sich anhand raumakustischer Kriterien und der perceptiven Evaluation, Unterschiede zwischen Messung und Modell. Ein signifikanter Haupteffekt für den Faktor *Modell* kann beobachtet werden, die Differenzen der Ergebnisse zwischen den beiden untersuchten Abstufungen sind jedoch als gering zu bewerten. So wurde zwar in der Regel tendenziell die größte Ähnlichkeit und damit die beste Performanz im Vergleich zur Referenz bei Verwendung der gerichteten+diffusen Komponenten beobachtet. Signifikante Unterschiede zur Auralisation auf Basis diffuser Komponenten konnten jedoch nur in drei der acht Bedingungen gezeigt werden. Diese Beobachtung deckt sich mit Ergebnissen der vorherigen Experimente. Allgemein betrachtet wurden die Unterschiede der dargebotenen Stimuli auf Basis der Synthese im Vergleich zur Referenz zumeist im oberen Drittel der Skala bewertet. Insbesondere bei Verwendung der Schlagzeug-Signale zur Auralisation konnten jedoch größere Differenzen im Vergleich zu Experiment LEB beobachtet werden. Hierfür ist vermutlich die Kürzung der Untersuchung um die Faktorstufe der ausschließlich gerichteten Komponenten verantwortlich. Wie in LEB gezeigt, führt diese Konfigurationen zu deutlich größeren Unterschieden und starker Beeinträchtigung der Auralisation. In Experiment LEC fehlt diese Verankerung und es werden nur Konfigurationen evaluiert, die schon technisch betrachtet geringe Unterschiede aufweisen. Eine unterschiedliche Verankerung kann zur einer Verschiebung der Ergebnisse (Bias) führen.

Eine große Einschränkung des Modells liegt bisher in der Notwendigkeit sowohl von einer Mikrofonarray-Messung mit großer Bandbreite als auch von einer Kunstkopf-

Messung als Eingangs-Signal für die Charakterisierung des Schallfeldes mit Parametern. Daher wird im folgenden Kapitel eine Methode erarbeitet, die sowohl den messtechnischen Aufwand als auch die Audio-Signalverarbeitung reduzieren soll.

5.6 Gesamtmodell mit reduzierter Auflösung (LED)

Das parametrische Modell $SBRIR_{comp}$ befasst sich ebenfalls wie das zuvor erläuterte Modell $SBRIR$ mit der Generierung von voll-synthetischen binauralen Raumimpulsantworten auf Basis einer messtechnisch erfassten Referenz. Es wird nun angestrebt, den benötigten Aufwand der Analyse der akustischen Situation erheblich zu verringern. Aus dieser Vereinfachung des Modells resultierende technische als auch perzeptive Unterschiede gilt es, zu evaluieren. Daher liegt der Fokus in Experiment LED nicht mehr auf den Algorithmen an sich, vielmehr werden unterschiedliche Konfigurationen zur Charakterisierung der Räumlichkeiten mittels Parametern verglichen.

Die bisher vorgestellten Modelle benötigen für die Entwicklung der parametrischen Beschreibung zum einen eine hochaufgelöste Mikrofonarray-Messung für die Methoden der Wellenfeldanalyse und zum anderen eine binaurale Referenz (Kunstkopfmessung) für die Bestimmung der diffusen Abklingkurven sowie der interauralen Kohärenz. Diese Vorgehensweise scheint für Anwendungen am Markt nicht praktikabel, daher soll das Modell vereinfacht werden. Dazu können zwei Variablen verändert werden, die eine erhebliche Reduzierung des messtechnischen Aufwandes mit sich ziehen.

Die Verwendung geringerer Abtastgitter des Mikrofonarrays kann die Messdauer herabsetzen oder gar den Einsatz eines Echtzeit-Mikrofonarrays ermöglichen. Im Gegensatz zum in dieser Arbeit eingesetzten sequenziellen Mikrofonarray, besteht ein Echtzeit-Mikrofonarray aus einer Vielzahl von Mikrofonen, sodass die verschiedenen Impulsantworten auf der Kugeloberfläche gleichzeitig gewonnen werden können. Gerätschaften dieser Art sind mittlerweile zwar bereits verfügbar², eine breite Kommerzialisierung hat jedoch noch nicht stattgefunden, sodass es sich noch um Nischenprodukte handelt. Um den Einfluss der Auflösung des Mikrofonarrays auf die Performanz des parametrischen Modells zu evaluieren, werden im Folgenden verschiedene Abtast-Muster verglichen. Weiterhin wird angestrebt, auf die Messung der binauralen Referenz zu verzichten, sodass mit dem Mikrofonarray nur noch eine Mess-Apparatur (im Idealfall ein Echtzeit-Mikrofonarray) benötigt wird. Dazu müssen die diffusen Parameter der interauralen Kohärenz sowie die binauralen Abklingkurven auf Basis des Mikrofonarrays berechnet werden. Im Allgemeinen werden so weiterhin die zuvor vorgestellten und evaluierten Ansätze für die Synthese der gerichteten und diffusen Schallfeldkomponen-

²z.B. *mh acoustics Eigenmike* / *VisiSonics Audio/Visual Camera* / *Yamaha ViReal Mic*

ten verwendet. Auf Unterschiede in der Signalverarbeitung bezüglich der angestrebten Vereinfachungen wird im Folgenden Abschnitt näher eingegangen.

5.6.1 Modell und Parameter

In Abbildung 5.29 findet sich das Blockschaltbild des parametrischen Modells $SBRIR_{comp}$ zur Generierung synthetischer binauraler Raumimpulsantworten, welches Experiment LED zu Grunde liegt. Dieses Modell entspricht weitestgehend Modell $SBRIR$, zum Einsatz niedrigerer Auflösungen des Mikrofonarrays sind kleine Abänderungen der Signalverarbeitung notwendig.

Um die Parameter ausschließlich auf Basis des Mikrofonarrays bestimmen zu können und den Einsatz von reduzierten Abtastmustern zu ermöglichen, ergeben sich die folgenden Unterschiede zu den vorherigen Modellen. Eine Verringerung der Abtastung der Kugeloberfläche hat, wie in Kapitel 2.5 gezeigt, eine Einschränkung der effektiv nutzbaren Bandbreite der Wellenfeldanalyse zur Folge. Gemäß Abbildung 2.15 arbeitet beispielsweise die Schallfeldzerlegung auf Basis eines Mikrofonarray mit 86 Abtastpunkten (Lebedev) und einem Radius von 8,75 cm ab einer Frequenz von $f_a = 4450$ Hz nicht mehr ideal, da räumliches Aliasing auftritt. Da nun eine entsprechende Array-Konfiguration zur Bestimmung der Parameter des Schallfeldes verwendet wird, bedarf es einer Anpassung der Signalverarbeitung.

Für die Zerlegung des Schallfeldes in ebene Wellen zur Bildung der Intensitätsmatrix wird nun eine geringere Frequenz von $f_a = 3750$ Hz verwendet, um weiterhin in einem stabilen Frequenzbereich des Arrays zu arbeiten. Die modale Ordnung $N = 5$ muss hingegen nicht reduziert werden. Die Verschiebung des Frequenzbereichs zur Detektion der Schallfeldmaxima zeigt keine auffälligen Unterschiede zu den vorherigen Untersuchungen. Durch Erhöhung des globalen Schwellwertes für die Klassifikation der Maxima (siehe 4.5) wird in dieser Untersuchung zudem die Anzahl der zu synthetisierenden Reflexionen herabgesetzt, um eine weitere Reduktion des Aufwandes der Synthese zu erzielen. In jeder akustischen Situation werden für Experiment LED zwischen 50 und 60 Reflexionen synthetisch generiert.

Darüber hinaus kann die Signalverarbeitung zur Bildung der gerichteten Impulsantworten für die Berechnung der spektralen Färbung der gerichteten Anteile für Alias-behaftete Signale nicht mehr verwendet werden. Da das auftretende räumliche Aliasing den Frequenzgang der Impulsantworten beeinflusst, hätte dies ebenfalls eine Verzerrung der Frequenzfilter zur Folge. Daher werden keine gerichteten Impulsantworten mit Methoden der Wellenfeldanalyse erzeugt, sondern stattdessen entsprechend der detektierten Einfallsrichtung der Reflexion, die Array-Impulsantwort mit der gerings-

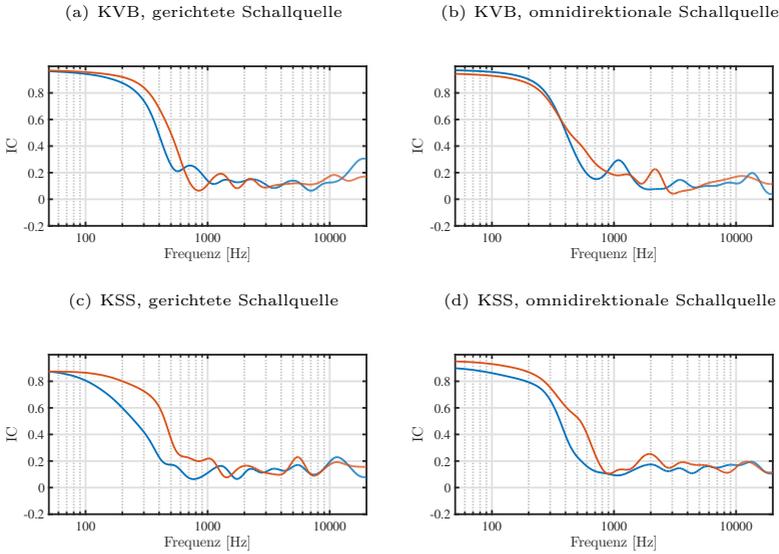


Abbildung 5.28: Vergleich der interauralen Kohärenz (Diffusfeld ab 320 ms) über der Frequenz für Referenz-BRIRs (blau) und auf Basis von Mikrofonarray-Daten (rot).

ten Differenz der Orientierung zur Reflexion bestimmt. Diese Impulsantwort wird ohne weitere Signalverarbeitung zur Bestimmung der spektralen Färbung der Rückwürfe verwendet. Aus dieser Vorgehensweise ergeben sich zum einen abhängig der räumlichen Abtastung des Mikrofonarrays Abweichungen der detektierten Einfallswinkel, zum anderen hängt die Richtwirkung lediglich von der Mikrofonkapsel in Verbindung mit der mechanischen Konstruktion des Arrays ab. Eine Erhöhung der Richtwirkung durch Methoden der Wellenfeldanalyse kann entgegen vorheriger Experimente nicht durchgeführt werden. Daher sinkt die räumliche Fokussierung auf die jeweilige Reflexion, sodass zeitgleich auftretende Schallfeldanteile mehr Einfluss auf die Filter nehmen.

Die Bestimmung der diffusen Parameter erfolgt ebenfalls auf Basis der Impulsantworten der Mikrofonarray-Messung. Hierfür werden die Impulsantworten detektiert, deren Orientierung möglichst genau den beiden Ohrsignalen bei Blickrichtung zur Schallquelle entsprechen (linkes Ohr: $\phi = 90^\circ$, $\theta = 0^\circ$, rechtes Ohr: $\phi = 270^\circ$, $\theta = 0^\circ$). Dieses Signalpaar wird dann als binaurale Referenz aufgefasst und zur Bestimmung der diffusen Parameter (Abklingzeit und Kohärenz) verwendet. Ein Vergleich der interauralen Kohärenz zwischen dieser Methode und einer Kopfkopfmessung findet sich in Abbildung

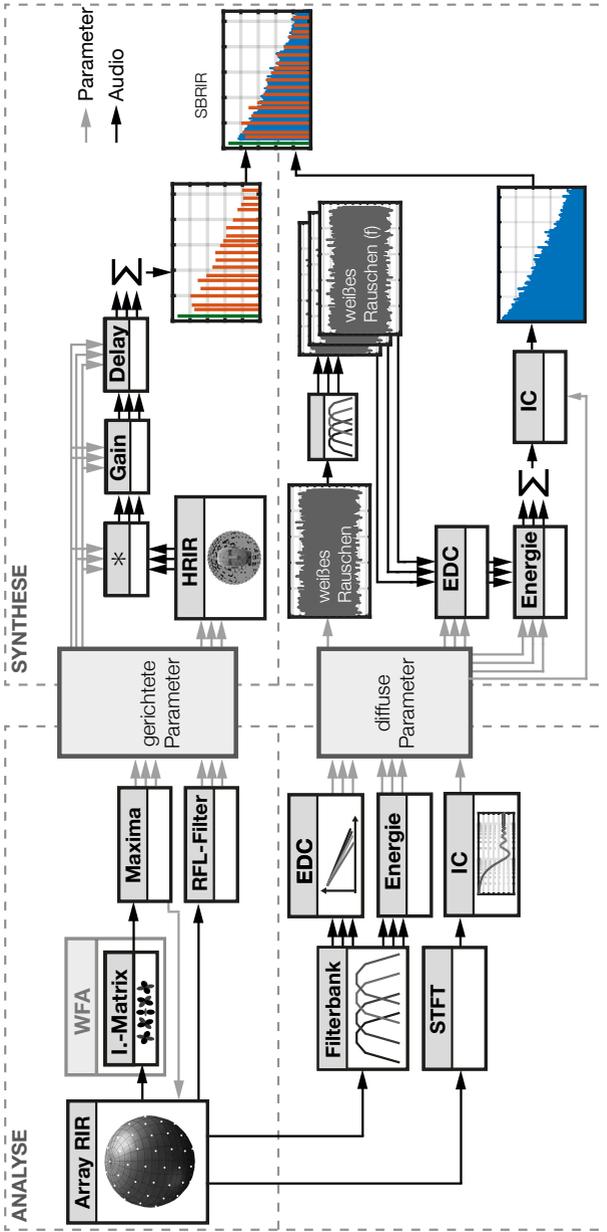


Abbildung 5.29: Blockschaltbild **SBRIR_{comp}**: Gerichtete Komponenten werden auf Basis von Mikrofonarray-Daten mit Wellenfeldanalyse (WFA) detektiert und zur Synthese von Reflexionen mittels sphärischer HRIRs verwendet. Diffuse Komponenten werden ebenfalls in den Array-Daten detektiert und zur Adaption von weißem Rauschen eingesetzt. Synthetische Reflexionen (rot) und diffuser Nachhall (blau) werden zur Bildung einer SBRIR zusammengefügt.

5.28. Allgemein zeigen sich in allen untersuchten Räumlichkeiten übereinstimmende Verläufe über der Frequenz für beide Eingangssignale (Kunstkopf und Mikrofonarray), sodass keine großen Beeinträchtigungen aufgrund dieser alternativen Signalverarbeitung zu erwarten sind. Die Bestimmung der frequenzabhängigen Abklingkurven auf Basis der Impulsantworten des Mikrofonarrays ist ebenfalls unkritisch.

5.6.2 Datensätze

Im Folgenden werden die für die technische und perzeptive Evaluation des parametrischen Modells $SBRIR_{comp}$ generierten Datensätze (siehe Tabelle 5.21) erläutert. Der Fokus liegt dabei auf der Untersuchung der Differenzen, die aus der Verwendung verschiedener Mikrofonarrays resultieren. Daher werden zum einen Datensätze erzeugt, die auf der Analyse bisher genutzter Array-Konfigurationen (1202 Lebedev) basieren. Weiterhin werden Datensätze generiert, die zur Analyse des Schallfeldes ein Mikrofonarray mit lediglich 86 Lebedev Abtastpunkten verwenden. Die Radien der Mikrofonarrays (8,75 cm) und Ausführungen (Mikrofon planar in geschlossener Holzkugel) sind identisch. Um einen systematischen Vergleich zu ermöglichen, wird für beide Konfigurationen abermals auf das sequenzielle Mikrofonarray *VariSphear* zurückgegriffen. Die Impulsantworten der geringeren Auflösung könnten jedoch ebenfalls mit Hilfe eines Echtzeit-Arrays gewonnen werden. Weiterhin werden SBRIRs mit unterschiedlichen Konfigurationen der Algorithmen und auf Basis verschiedener akustischer Situationen gebildet. Als Referenz wird eine binaurale Kunstkopf-Messung verwendet, die nun aber nicht mehr in das parametrische Modell eingespeist wird.

Es werden pro Kombination aus Raum und Schallquelle zwei verschiedene Abstufungen des Modells (diffus/gerichtet+diffus) mit den zwei verschiedenen Array-Konfigurationen (1202/86 Abtastpunkte) erzeugt. So ergeben sich insgesamt 16 unterschiedliche SBRIR-Datensätze, die näher untersucht werden.

5.6.3 Technische Untersuchung

In Tabelle 5.22 und Abbildung 5.30 finden sich die Ergebnisse der technischen Untersuchung des parametrischen Modells $SBRIR_{comp}$.

Im Allgemeinen zeigen sich für die ausgewerteten raumakustischen Kriterien nur geringe Differenzbeträge zwischen den Signalen. Bekannte Wahrnehmbarkeitsschwellen nach DIN EN ISO 3382-1:2009-10 (2009) werden teilweise geringfügig überschritten und die Abweichungen entsprechen weitgehendst den in LEC beobachteten Werten. Es ist ein leichter Vorteil für die Verwendung der Konfiguration mit 1202 Punkten zu

Tabelle 5.21: Generierte Datensätze für LED.

Abtastpunkte Array	gerichtet	diffus	Raum	Schallquelle
86		×	KVB	direktional
86	×	×	KVB	direktional
86		×	KVB	omnidirektional
86	×	×	KVB	omnidirektional
86		×	KSS	direktional
86	×	×	KSS	direktional
86		×	KSS	omnidirektional
86	×	×	KSS	omnidirektional
1202		×	KVB	direktional
1202	×	×	KVB	direktional
1202		×	KVB	omnidirektional
1202	×	×	KVB	omnidirektional
1202		×	KSS	direktional
1202	×	×	KSS	direktional
1202		×	KSS	omnidirektional
1202	×	×	KSS	omnidirektional

verzeichnen. Hierbei erfolgt eine präzisere Nachbildung des spektralen Verlaufs, der Nachhallzeit, der Schwerpunktzeit sowie von Deutlichkeit und Klarheit. Ebenso ist eine leicht genauere Adaption der raumakustischen Kriterien durch Verwendung der Synthese von gerichteten+diffusen Komponenten im Vergleich zur Synthese ausschließlich diffuser Anteile zu beobachten. Die Differenzen zwischen den beiden Abstufungen sind jedoch ebenfalls gering. So zeigt sich auch in der frequenzabhängigen Betrachtung gemittelt über alle Räumlichkeiten und Schallquellen keine klare Tendenz für eine bestimmte Konfiguration des Modells. Eine frequenzabhängige Betrachtung im Detail für jede akustische Situation und Abstufung des Modells kann anhand der Plots gemäß Abbildungen A.37-A.52 (siehe Anhang) erfolgen.

Wie bereits in Experiment LEC gesehen, entstehen große Abweichungen bei der frequenzabhängigen Betrachtung der frühen Abklingzeit. Abermals misslingt eine genaue Adaption dieses Maßes bei einer Frequenz von 16 kHz, sodass sich hohe prozentuale Abweichungen zeigen. Wird dieser Frequenzbereich für die Kalkulation des Einzahlwertes ausgespart, so werden wiederum deutlich geringere Abweichungen dokumentiert. Jedoch können für das Maß der frühen Abklingzeit die mit Abstand größten Differenzbeiträge aller untersuchten Kriterien beobachtet werden, sodass die Wahrnehmbarkeitsschwelle hierbei vermutlich überschritten wird.

Ein Vergleich der detektierten Reflexionen auf Basis von 1202 oder 86 Abtastpunkten ist aufgrund der verschiedenen Dimensionen der Parameter (Orientierung, Zeit und Amplitude) komplex und schwer zu visualisieren. Eine getrennter Vergleich der unter-

Tabelle 5.22: LED - Absolute Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ der raumakustischen Kriterien im Mittel über alle Frequenzen, Räume und Schallquellen, Vergleich zwischen Array Auflösung und Konfiguration des Modells.

Array	Modell		Spektrum	<i>T30</i>	EDT	<i>IACC</i> _{early}	<i>IACC</i> _{late}
	gerichtet	diffus	[dB]	[%]	[%]		
86		×	2,6	4,4	116 (19)	0,11	0,09
86	×	×	2,2	5,0	196 (17)	0,11	0,09
1202		×	2,4	4,2	175 (19)	0,15	0,1
1202	×	×	2,0	4,3	199 (17)	0,15	0,09

Array	Modell		<i>T_s</i>	<i>C80</i>	<i>D50</i>	<i>BR</i>	<i>TR</i>
	gerichtet	diffus	[ms]	[dB]	[%]		
86		×	12,1	2,0	7,6	0,06	0,05
86	×	×	10,6	1,8	7,2	0,05	0,06
1202		×	11,8	2,0	7,8	0,06	0,04
1202	×	×	10,7	1,9	7,0	0,05	0,05

schiedlichen Dimensionen ist sinnvoll. In Abbildung A.53 findet sich für jede Räumlichkeit und Schallquelle eine Gegenüberstellung der detektierten Reflexionen in Abhängigkeit von Zeit und Amplitude. In allen Bedingungen zeigt sich ein sehr ähnlicher Verlauf, sodass bei allen Konfigurationen für diese Betrachtungsweise vergleichbare Reflexionen detektiert werden. Mittels eines Histogramms kann die Häufigkeitsverteilung der Orientierungen in den jeweiligen Maxima-Listen analysiert werden. Abbildungen A.54 und A.55 visualisieren diesen Zusammenhang für Azimuth- und Elevations-Winkel separat. Eine Interpretation dieser Diagramme ist im Vergleich zur energetischen Betrachtung weitaus schwieriger, da pro Konfiguration nie dieselbe Anzahl an Reflexionen detektiert werden. Daher geben die Histogramme nur einen groben Überblick über die unterschiedlichen Verteilungen der Orientierungen für 1202 oder 86 Abtastpunkte. Insgesamt zeigen sich jedoch vergleichbare Verteilungen, insbesondere im KVB-Saal.

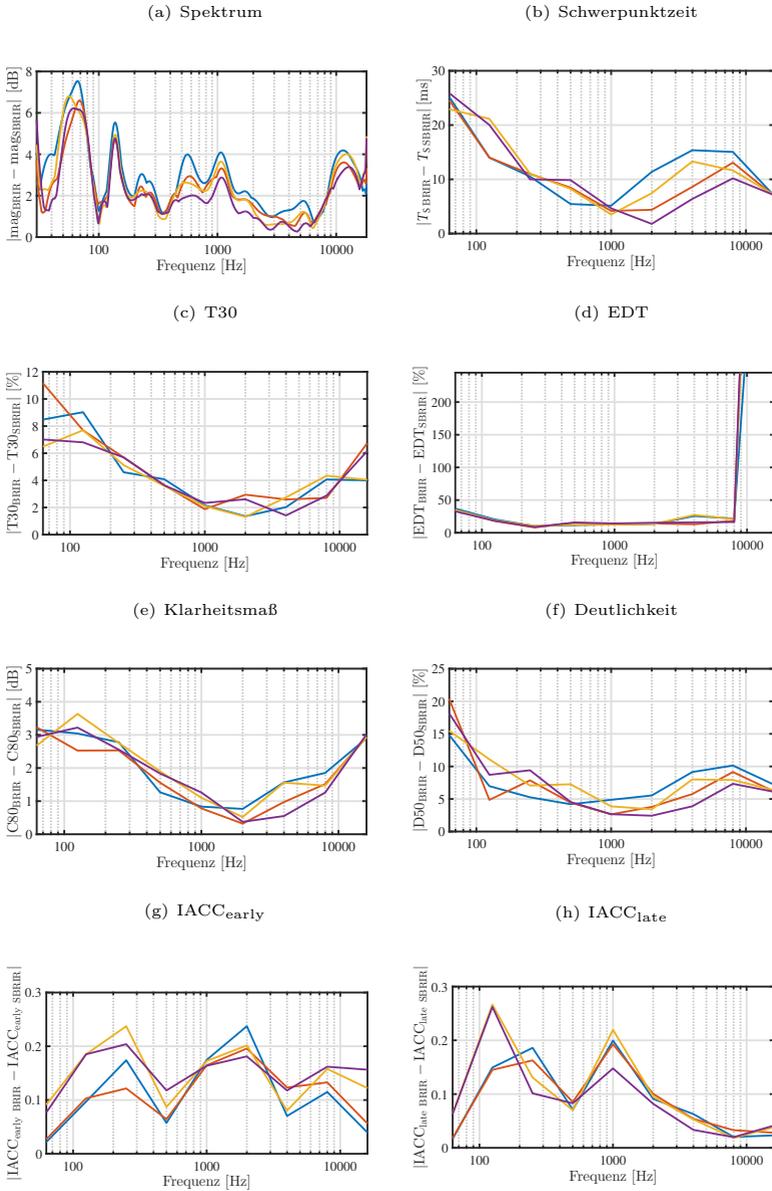


Abbildung 5.30: LED: Mittlere Differenzbeträge $|\text{BRIR} - \text{SBRIR}|$ Vergleich Array, diffuse (blau: 86, gelb: 1202) und gerichtete+diffuse Anteile (rot: 86, lila: 1202).

5.6.4 Perzeptive Evaluation

Tabelle 5.23: Eigenschaften LED.

Experiment	LED	
Modell	SBRIR _{comp}	
Kurzbeschreibung	voll synthetische BRIRs; Vergleich Array-Auflösung	
Versuchsdesign	ABC HR	
Bedingungen	32	
Faktoren	5	
Faktorstufen		
Modell	2	diffus / gerichtet+diffus
Raum	2	KVB-Saal / KSS
Schallquelle	2	gerichtet / omnidirektional
Testsignal	2	Schlagzeug / Rauschimpulse (weiß)
Array	2	1202 / 86 Abtastpunkte
Wiederholungen	5	
Bewertungen / Proband	160	
Probanden	25 (19 m / 6 w)	19 - 32 Jahre / $\bar{\mu}$ 23,7 Jahre / σ 3,2 Jahre

In Tabelle 5.23 finden sich die relevanten Eigenschaften des Experiments LED zur Untersuchung der wahrnehmbaren Unterschiede zwischen einer Auralisation auf Basis der Referenz-BRIRs im Vergleich zu SBRIRs, die mit dem Modell generiert wurden. 25 (19 männlich, 6 weiblich) normal-hörende Probanden im Alter von 19 bis 32 Jahren ($\bar{\mu}$ 23,7 Jahre) nehmen an der Untersuchung teil. 20 Probanden können bereits Erfahrung in Hörversuchen mittels des verwendeten dynamischen Binauralsystems vorweisen. Gemäß Abschnitt 5.6.2 werden insgesamt 16 synthetische Datensätze für diese Studie generiert und in der Auralisation dem messtechnisch erfassten Datensatz mittels Faltung mit zwei verschiedenen Testsignalen (Schlagzeug/weiße Rauschimpulse) gegenübergestellt. Die resultierenden 32 Bedingungen werden je 5 mal wiederholt (randomisiert), sodass jeder Proband in ca. 75 Minuten Hörversuch (mit Pause) 160 Bewertungen abgeben muss.

Untersucht werden die fünf Faktoren *Modell*, *Raum*, *Schallquelle*, *Testsignal* und *Array*. Die Faktoren *Modell*, *Raum* und *Schallquelle* sind unterteilt in je zwei Faktorstufen (diffus/gerichtet+diffus bzw. KVB-Saal/kleiner Sendesaal bzw. gerichtet/omnidirektional). Die Stufen des Faktors *Testsignal* entsprechen den beiden im Binauralrenderer eingesetzten Audiosignalen (Schlagzeug/Rauschimpulse). Der Faktor *Array* wird anhand der beiden untersuchten Lebedev-Gitter mit 1202 bzw. 86 Abtastpunkten unterteilt.

Eine fünf-faktorielle Varianzanalyse mit Messwiederholung (rmANOVA) wird zur Auswertung der fünf Faktoren ($2 \times 2 \times 2 \times 2 \times 2$) angewendet (siehe Abschnitt 3.7.2). Gemäß Shapiro und Wilk (1965) sind die Bewertungen pro Bedingung normalverteilt. Die Ergebnisse dieser statistischen Analyse des Experiments LED finden sich in Tabelle 5.24.

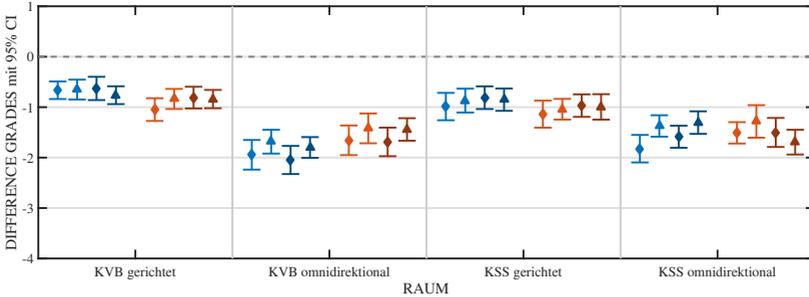


Abbildung 5.31: LED - Mittelwerte der difference grades mit 95% CI (aus 1-Stichproben t -Test); diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug + Array 1202 (hellblau), Schlagzeug + Array 86 (dunkelblau), Rauschimpulse + Array 1202 (orange) und Rauschimpulse + Array 86 (dunkelrot) für verschiedene Räume und Schallquellen.

Es werden signifikante Haupteffekte für die Faktoren *Schallquelle* und *Modell* beobachtet. Die Faktoren *Raum*, *Testsignal* und *Array* sind hingegen nicht signifikant. Die höchste Effektstärke anhand der η_p^2 -Werte wird dabei für den Faktor *Schallquelle* ($\eta_p^2 = 0,83$) beobachtet. Faktor *Raum* ($\eta_p^2 = 0,47$) zeigt ebenfalls eine hohe Effektstärke. Allerdings belegt die statistische Auswertung zusätzlich eine Vielzahl von stark signifikanten Interaktionen zwischen den Faktoren mit hohen Effektstärken. Es wird anhand von Interaktionsdiagrammen nach Bortz und Schuster (2010), S.245 überprüft, ob die Haupteffekte eindeutig interpretierbar sind. In der Mehrheit der Interaktionsdiagramme zeigen sich zwar ordinale Verläufe, da jedoch für jede Kombination auch mitunter disordinale Kurvenzüge beobachtet werden können, ist eine eindeutige Interpretation der Haupteffekte nicht möglich. Weiterhin zeigt sich anhand der Interaktionseffekte, dass offenbar alle Faktoren Einfluss nehmen, obgleich nicht immer signifikante Haupteffekte belegt werden können.

Die Mittelwerte der Ergebnisse der perceptiven Untersuchung des Experiments LED über alle Probanden und Wiederholungen als *difference grades* finden sich in Abbildung 5.31, eine Übersicht über die Verteilung der Bewertungen gibt Abbildung 5.32. Die Mittelwerte der Bewertungen der Synthese mit ausschließlich diffusen Komponenten werden in den Diagrammen abermals mittels Raute, die Bewertungen der Synthese auf Basis des kompletten Modells (gerichtet+diffus) mittels Dreieck dargestellt. Verschiedene Farben visualisieren die Testsignale sowie die Array-Konfigurationen. Dabei steht ein blauer Farbton für die Verwendung der Schlagzeug-Signale und ein roter Farbton für die der Rauschimpulse. Eine Unterteilung in hell- und dunkelblau differenziert

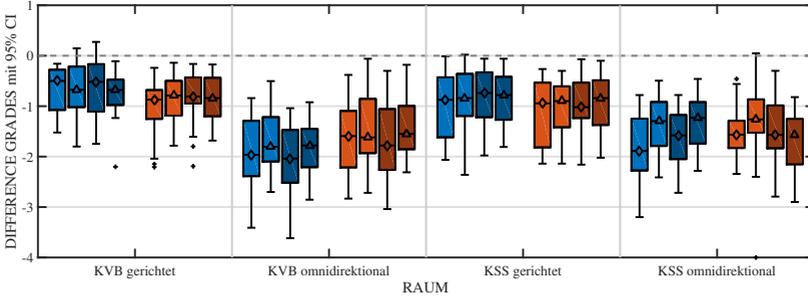


Abbildung 5.32: LED - Verteilung der difference grades in Form eines Boxplots; diffuse Anteile (Raute) und gerichtete+diffuse Anteile (Dreieck) für Schlagzeug + Array 1202 (hellblau), Schlagzeug + Array 86 (dunkelblau), Rauschimpulse + Array 1202 (orange) und Rauschimpulse + Array 86 (dunkelrot) für verschiedene Räume und Schallquellen.

zwischen 1202 bzw. 86 Abtastpunkten des Arrays für die Auralisation mittels Schlagzeug. In orange (1202) oder dunkelrot (86) werden die Ergebnisse der verschiedenen Array-Auflösungen bei Verwendung der Rauschimpulse dargestellt.

Im Allgemeinen zeigen sich zunächst zu Experiment LEC vergleichbare Mittelwerte der Ergebnisse, die in der oberen Hälfte der Skala liegen. Aufgrund starker Interaktionseffekte werden alle Faktoren anhand paarweiser t -Tests mit Korrektur nach Hochberg (1988) näher analysiert. Es werden jeweils pro Faktor alle Stufen gegenübergestellt, bei gleichzeitiger Fixierung der restlichen Bedingungen. So müssen pro Faktor 16 Vergleiche durchgeführt werden, deren Ergebnisse nun vorgestellt werden.

Die Betrachtung von *Schallquelle* dokumentiert für 15 der 16 Kombinationen einen signifikanten Unterschied der Bewertungen aufgrund der Richtcharakteristik. Dabei werden für die Verwendung der omnidirektionalen Schallquelle größere Unterschiede zur Referenz beobachtet. Ausschließlich der Vergleich der Bewertungen der Schallquellen im kleinen Sendesaal bei Verwendung des Modells mit gerichteten+diffusen Anteilen, einer Abtastung von 1202 Punkten und Rauschimpulsen als Testsignal, führt zu keinem signifikanten Unterschied der Schallquellen. Die Mittelwerte zeigen aber auch hier eine größere Ähnlichkeit zur Referenz für Verwendung der gerichteten Schallquelle.

Der paarweise Vergleich der unterschiedlichen Abstufungen des Modells anhand der verwendeten Schallfeldanteile zur Synthese (diffus bzw. gerichtet+diffus) zeigt ausschließlich für eine Kombination (kleiner Sendesaal, omnidirektionale Quelle, 1202 Abtastpunkte, Schlagzeug) einen signifikanten Unterschied mit höherer Ähnlichkeit der

Auralisation auf Basis der Synthese mit gerichteten+diffusen Komponenten. Insgesamt kann für 11 der 16 Vergleiche tendenziell eine höhere Ähnlichkeit im Vergleich zur Referenz bei Verwendung beider Schallfeldkomponenten beobachtet werden.

Zwar können anhand der Varianzanalyse für die weiteren Faktoren *Raum*, *Testsignal* und *Array* keine signifikanten Effekte gezeigt werden. Da aber signifikante Interaktionen dieser Faktoren auftreten, werden die paarweisen Vergleiche ebenfalls aufgeführt. Im Allgemeinen kann bei Betrachtung der Mittelwerte anhand Abbildung 5.31 für keinen dieser Faktoren eine klare Tendenz beobachtet werden. Der paarweise Vergleich des Faktors *Raum* dokumentiert in zwei Kombinationen der Bedingungen einen signifikanten Unterschied aufgrund des Raums. Bei Verwendung der omnidirektionalen Schallquelle, der Schlagzeug-Signale und einer Array-Konfiguration mit 86 Abtastpunkten, werden für beide Abstufungen des Modells die Unterschiede zur Referenz im KVB-Saal größer bewertet als im kleinen Sendesaal. Die nähere Betrachtung des Faktors *Testsignal* zeigt in 3 von 16 Vergleichen einen signifikanten Unterschied durch Variation der Stufe, für 2 Kombinationen erreichen die SBRIRs mit Schlagzeug-Signalen signifikant geringere Unterschiede zur Referenz-BRIR im Gegensatz zu den Rauschsignalen. Anhand der *t*-Tests für den Faktor *Array* wird in einer Kombination ein signifikanter Unterschied durch Variation der Konfiguration des Array beobachtet (KVB-Saal, diffuse Komponenten und Rauschsignale), mit höherer Bewertung für Datensätze auf Basis des Array mit 86 Abtastpunkten. So zeigen sich auch in den Post-Hoc-Tests für die Faktoren Raum, Testsignal und Arrays keine eindeutigen Tendenzen, sodass deren Relevanz trotz zu beobachtender Interaktionseffekte als gering aufgefasst werden kann.

Tabelle 5.24: Zusammenfassung rmANOVA LED; bei $\alpha = 5\%$ signifikante Werte symbolisiert mit *; Freiheitsgrade nicht korrigiert, p -Werte korrigiert nach Greenhouse und Geisser (1959).

Faktor(en)	F-Wert	p-Wert	Effektstärke
Modell	$F(1,24) = 21,0$	$p < 0,001^*$	$\eta_p^2 = 0,47$
Raum	$F(1,24) = 0,09$	$p = 0,77$	$\eta_p^2 = 0$
Schallquelle	$F(1,24) = 119$	$p < 0,001^*$	$\eta_p^2 = 0,83$
Testsignal	$F(1,24) = 0$	$p = 0,94$	$\eta_p^2 = 0$
Array	$F(1,24) = 0,25$	$p = 0,62$	$\eta_p^2 = 0,01$
Modell \times Raum	$F(1,24) = 0,11$	$p = 0,74$	$\eta_p^2 = 0$
Modell \times Schallquelle	$F(1,24) = 11,0$	$p = 0,003^*$	$\eta_p^2 = 0,31$
Modell \times Testsignal	$F(1,24) = 1,36$	$p = 0,26$	$\eta_p^2 = 0,05$
Modell \times Array	$F(1,24) = 12,6$	$p = 0,002^*$	$\eta_p^2 = 0,34$
Raum \times Schallquelle	$F(1,24) = 16,1$	$p < 0,001^*$	$\eta_p^2 = 0,40$
Raum \times Testsignal	$F(1,24) = 3,66$	$p = 0,07$	$\eta_p^2 = 0,13$
Raum \times Array	$F(1,24) = 0,80$	$p = 0,38$	$\eta_p^2 = 0,03$
Schallquelle \times Testsignal	$F(1,24) = 24,9$	$p < 0,001^*$	$\eta_p^2 = 0,51$
Schallquelle \times Array	$F(1,24) = 6,77$	$p = 0,016^*$	$\eta_p^2 = 0,22$
Testsignal \times Array	$F(1,24) = 0,44$	$p = 0,52$	$\eta_p^2 = 0,02$
Modell \times Raum \times Schallquelle	$F(1,24) = 0,85$	$p = 0,37$	$\eta_p^2 = 0,03$
Modell \times Raum \times Testsignal	$F(1,24) = 7,67$	$p = 0,01^*$	$\eta_p^2 = 0,24$
Modell \times Schallquelle \times Testsignal	$F(1,24) = 6,85$	$p = 0,02^*$	$\eta_p^2 = 0,22$
Modell \times Raum \times Array	$F(1,24) = 1,14$	$p = 0,3$	$\eta_p^2 = 0,05$
Modell \times Schallquelle \times Array	$F(1,24) = 0,08$	$p = 0,78$	$\eta_p^2 = 0$
Modell \times Testsignal \times Array	$F(1,24) = 0,41$	$p = 0,53$	$\eta_p^2 = 0,02$
Raum \times Schallquelle \times Testsignal	$F(1,24) = 6,52$	$p = 0,02^*$	$\eta_p^2 = 0,21$
Raum \times Schallquelle \times Array	$F(1,24) = 0,06$	$p = 0,81$	$\eta_p^2 = 0$
Raum \times Testsignal \times Array	$F(1,24) = 7,25$	$p = 0,01^*$	$\eta_p^2 = 0,23$
Schallquelle \times Testsignal \times Array	$F(1,24) = 4,42$	$p = 0,05^*$	$\eta_p^2 = 0,16$
Modell \times Raum \times Schallquelle \times Testsignal	$F(1,24) = 0,82$	$p = 0,37$	$\eta_p^2 = 0,03$
Modell \times Raum \times Schallquelle \times Array	$F(1,24) = 3,66$	$p = 0,07$	$\eta_p^2 = 0,13$
Modell \times Raum \times Testsignal \times Array	$F(1,24) = 0,22$	$p = 0,64$	$\eta_p^2 = 0$
Modell \times Schallquelle \times Testsignal \times Array	$F(1,24) = 0,18$	$p = 0,67$	$\eta_p^2 = 0$
Raum \times Schallquelle \times Testsignal \times Array	$F(1,24) = 2,02$	$p = 0,17$	$\eta_p^2 = 0,08$
Modell \times Raum \times Schallquelle \times Testsignal \times Array	$F(1,24) = 1,16$	$p = 0,29$	$\eta_p^2 = 0,05$

5.6.5 Diskussion und Zusammenfassung

In diesem Abschnitt wurde das parametrische Modell $SBRIR_{comp}$ zur voll-synthetischen Generierung einer binauralen Raumimpulsantwort auf Basis von sphärischen Mikrofonarray-Messungen vorgestellt. Das Modell und die zugrunde liegende Signalverarbeitung wurde erläutert sowie technisch anhand raumakustischer Kriterien und perzeptiv im Rahmen eines Hörversuchs evaluiert. Die so erzeugten $SBRIRs$ können entsprechend der vorherigen Modelle zur dynamischen Auralisation von Räumen verwendet werden und erlauben bei Änderung der Orientierung des Rezipienten auf der horizontalen Ebene, eine ortsfeste Wiedergabe des Schallfeldes entsprechend der Realität. Das Modell detektiert, parametriert und generiert gerichtete und diffuse Schallfeldkomponenten unabhängig voneinander und überlagert beide Anteile zur Bildung der synthetischen Impulsantworten.

Im Gegensatz zu den zuvor vorgestellten Modellen, arbeitet die Parametrierung eines messtechnisch erfassten Schallfeldes mittels $SBRIR_{comp}$ ausschließlich auf Mikrofonarray-Datensätzen, sodass keine binaurale Messung zur Parametrierung benötigt wird. Weiterhin sind nun auch Array-Konfigurationen mit deutlich niedrigerer Auflösung einsetzbar, die Echtzeit-Arrays möglich machen. Dadurch wird der messtechnische Aufwand erheblich reduziert und die Praktikabilität des Modells erhöht.

Die gesamte Signalverarbeitung, die in diesem Modell Anwendung findet, entspricht weitgehend den vorherigen Modellen $SBRIR_{late}$ und $SBRIR_{early}$. So basiert die Synthese der diffusen Komponenten weiterhin auf weißen Rauschsignalen, denen der Energieabfall sowie die interaurale Kohärenz des messtechnisch erfassten Referenz-Schallfeldes frequenzabhängig aufgeprägt werden. Als Referenz findet nun keine binaurale Messung Anwendung, stattdessen werden entsprechende Signale aus der Arraymessung gebildet. Die resultierenden Differenzen sind äußerst gering im Vergleich zu einer Kunstkopfmessung.

Die Synthese der gerichteten Komponenten verwendet Methoden der Wellenfeldanalyse zur richtungs-sensitiven Detektion von Reflexionen im Schallfeld auf Basis sphärischer Mikrofonarray-Messungen und generiert mit Hilfe sphärischer HRIR-Daten synthetische Reflexionen. Da die Array-Datensätze mitunter aufgrund der verwendeten Konfiguration mit räumlichem Aliasing behaftet sind, werden entgegen der Experimente **LEB** und **LEC** keine gerichteten Impulsantworten gebildet. Stattdessen werden die Reflexionsfilter lediglich auf Basis der gemessenen Impulsantworten (mit möglichst geringem Winkelunterschied zur Einfallrichtung) ohne weitere Signalverarbeitung gebildet. Eine geringere räumliche Richtwirkung ist dabei die Folge im Gegensatz zu **LEB** und **LEC**.

Es findet weiterhin wie in Experiment **LEC** keine zeitliche Trennung zwischen Reflexionen und diffusen Anteilen statt, beide Komponenten werden für die gesamte Im-

pulsantwort synthetisiert, sodass gleichsam der diffuse binaurale Nachhall im frühen Bereich verwendet wird und auch Reflexionen zu späten Zeitpunkten generiert werden können. Zusätzlich erfolgt eine Reduzierung auf die stärksten gerichteten Anteile (im Mittel ca. 55 Reflexionen).

Es wurden zu Experiment LEC vergleichbare Ergebnisse der Bewertungen beobachtet, die alle in der oberen Hälfte der Skala liegen. Das Modell arbeitet weitestgehend unabhängig von Raum, Testsignal und Array. Es zeigen sich aber Interaktionseffekte für diese Faktoren. Auch der signifikante Einfluss des Modells bei Verwendung diffuser oder gerichteter+diffuser Schallfeldanteile ist als äußerst gering zu bewerten, obgleich sich eine Tendenz zum Vorteil der Gesamt-Modells beobachten lässt. Der Einfluss der Richtwirkung auf die Ergebnisse ist stark signifikant, sodass die omnidirektionale Schallquelle weitaus größere Differenzen im Vergleich zur Referenz als die gerichtete Schallquelle offenlegt. In Experiment LEC konnte diese klare Tendenz mit Vorteilen für die gerichtete Schallquelle nicht beobachtet werden, in LEB ausschließlich im kleinen Sendesaal. Wie bereits zuvor angemerkt, ist vermutlich abermals die DRR kritisch, sodass Unterschiede zwischen Messung und Modell bei geringer Ratio leichter wahrgenommen werden können. Offenbar sind diese Differenzen weiterhin bei nicht ganz akkurater Synthese (aufgrund geringerer räumlicher Auflösung und Einschränkungen der Reflexionsfilter) leichter detektierbar und beeinflussen stärker die Hörbarmachung.

Im Allgemeinen erlaubt das Modell $SBRIR_{comp}$ den Einsatz von niedrigeren Auflösungen des Mikrofonarrays und ermöglicht eine ausreichende Erfassung und Parametrierung der Räumlichkeiten zur adäquaten binauralen Synthese. Limitierungen der Signalverarbeitung aufgrund der nutzbaren Bandbreite durch räumliches Aliasing konnten durch Abänderung der Prozesse effektiv unterbunden werden, ohne dass ein wahrnehmbarer Einfluss beobachtet werden konnte. Dabei wurde weiterhin der messtechnische Aufwand als auch die notwendige Signalverarbeitung stark reduziert, sodass kein signifikanter Unterschied zwischen den verschiedenen Ansätzen gezeigt werden konnte. Somit führt diese Vereinfachung des Modells zu keiner Verschlechterung der Auralisation und kann gleichsam zur synthetischen Generierung binauraler Raumimpulsantworten verwendet werden.

KAPITEL 6

Ergebnisse und Ausblick



In Forschung und Entwicklung als auch in der Konsumelektronik ist weiterhin eine hohe Relevanz von Anwendungen der virtuellen Realität zu verzeichnen. Die Hörbarmachung von Schallfeldern in Räumen ist dabei ein zentrales Element der virtuellen Umgebung und es existieren verschiedene Ansätze zur Realisation. Die datenbasierte, dynamische Binauralsynthese verwendet dazu in der Regel im Raum erfasste, binaurale Impulsantworten. Methoden zur Synthese dieser Impulsantworten sind von hohem Interesse und fanden in dieser Arbeit eine umfangreiche Betrachtung.

Die den Untersuchungen vorausgehende These konnte anhand perceptiver Experimente bestätigt werden: Es ist möglich, das Schallfeld in Räumen mit einigen wenigen Merkmalen zu charakterisieren und diese zur adäquaten Reproduktion des Raumeindrucks in Form der Synthese einer binauralen Raumimpulsantwort zu verwenden. Möglich wird diese Vereinfachung aufgrund der auditiven Wahrnehmung des Menschen. Dazu wird das Schallfeld messtechnisch erfasst und durch diskrete Werte beschrieben, die weiterhin zur Datenreduktion approximiert werden können. Diese Parameter erlauben die Generierung von synthetischen, binauralen Datensätzen, die zu einer adäquaten Auralisation des untersuchten Raumes mittels Binauralsynthese verwendet werden können. Weiterhin ist auf Basis dieser Daten eine dynamische Hörbarmachung durch die Anpassung der Signale im Binauralrenderer entsprechend der Orientierung des Nutzers realisierbar. Entgegen der physikalisch, idealen Betrachtung des Schallfeldes in einer Raumsimulation, ermöglicht die perceptiv motivierte Synthese die Entfernung von für die Wahrnehmung irrelevanter Bestandteile.

Diese Vorgehensweise wurde in verschiedenen Modellen zur Synthese der Impulsantworten technisch anhand raumakustischer Kriterien als auch perceptiv mittels einer ABC/HR Versuchsumgebung evaluiert. Obgleich die Schritte der Signalverarbeitung zu einer starken Reduktion der Informationen führen und mitunter große technische Differenzen beobachtet werden können, sind die perceptiven Unterschiede bei entsprechender Ausführung der Synthese vergleichsweise gering. Der vorgestellte Ansatz separiert das Schallfeld in Direktschall sowie diffuse und gerichtete Komponenten, die unabhängig erfasst, mittels Parametern beschrieben und synthetisiert werden. Die Methoden wurden dabei kritisch untersucht und jeweils im direkten Vergleich einer messtechnisch erfassten Referenz gegenübergestellt und bewertet.

Eine zeitliche Unterteilung der Komponenten konnte als nicht praktikabel beobachtet werden, sodass gleichsam gerichtete als auch diffuse Anteile für die gesamte Länge der Raumimpulsantwort generiert werden. Die Anwendung der perceptiven mixing time als Übergangzeitpunkt zwischen den beiden Schallfeldkomponenten zeigte sich aufgrund mehrerer Gesichtspunkte nachteilig. Zum einen misslingt teilweise eine zufriedenstellende Bestimmung der perceptiven mixing time mittels bekannter Modelle. Dabei wurden sowohl negative (eine Ununterscheidbarkeit konnte erst zu späteren Zeitpunkten oder

nie gezeigt werden) als auch positive Abweichungen (eine Ununterscheidbarkeit konnte bereits deutlich früher gezeigt werden) von der mixing time beobachtet werden. Interessant ist weiterhin, dass wenn keine Ununterscheidbarkeit erreicht werden konnte, meist trotzdem eine große Akzeptanz der Bearbeitungen durch hohe Bewertungen evaluiert wurde. So führen zwar Vereinfachungen des diffusen Schallfeldes in der Regel zu einer Unterscheidbarkeit zwischen Referenz und Bearbeitung, die Unterschiede sind jedoch gering. Diese Betrachtungsweise wurde auch in dieser Arbeit verfolgt, sodass keine Ununterscheidbarkeit, sondern lediglich eine adäquate Auralisation bei gleichzeitiger starker Vereinfachung angestrebt wurde. Dies ist vermutlich für eine Vielzahl der intendierten Anwendungen zulänglich. Weiterhin zeigte sich bei Abwesenheit von diffusen Schallfeldanteilen im zeitlich frühen Bereich der Impulsantworten eine große Beeinträchtigung der Auralisation. Die Reflexionsdichte ist meist zu gering, sodass synthetische Reflexionen durch Überschreiten der Echoschwelle als einzelne Hörereignisse wahrgenommen werden können. Abhilfe schafft dabei der Einsatz von diffusen Schallfeldanteilen als Maskierer.

Der Direktschall kann entsprechend der Orientierung des Hörers zur Schallquelle synthetisch durch die Nutzung von Außenohrübertragungsfunktionen wiedergegeben werden. Die dynamische Nachführung des Direktschalls ist essentiell und für eine adäquate Auralisation in der virtuellen auditiven Umgebung unverzichtbar. Außerdem muss der Frequenzgang des synthetischen Direktschalls an die Referenz-Situation angepasst werden, um spezifische Klangeigenschaften zu berücksichtigen.

Die Synthese der weiteren Komponenten erlaubt große Vereinfachungen bei weiterhin zufriedenstellender Auralisationsergebnisse. Die präzise Adaption der frequenzabhängigen Abklingkurven sowie des Verlaufs der interauralen Kohärenz der diffusen Schallfeldanteile ist maßgeblich für eine adäquate Wiedergabe der akustischen Situation. Dazu werden lediglich zwei Rauschsignale benötigt, die energetisch über der Frequenz entsprechend der Nachhallzeit der realen Situation angepasst sowie deren Ähnlichkeit zueinander gemäß der binauralen Merkmale verändert werden. In den Experimenten zeigt sich der Bedarf einer schmalbandigen Frequenzauflösung für die Adaption des Energieabfalls insbesondere für hohe Frequenzen. Die Anpassung der Kohärenz der Rauschsignale hingegen ist unkritischer und sollte lediglich den für binaurale Signale üblichen Verlauf über der Frequenz aufweisen.

Zusätzlich können Außenohrübertragungsfunktionen, die entsprechend der Einfallrichtung detektierter Reflexionen verzögert sowie energetisch und spektral adaptiert werden, als gerichtete Komponenten verwendet und dem diffusen Schallfeld überlagert werden. Entgegen der diffusen Komponenten erfordert die Synthese der gerichteten Anteile die Generierung einer Vielzahl von Impulsantworten entsprechend der beabsichtigten Bewegungsauflösung in der Hörbarmachung, um bei Änderungen der Orientierung des

Nutzers, die ortsfeste Wiedergabe der synthetischen Reflexionen zu ermöglichen. Dieser Prozess führt nur teilweise zu einer signifikanten Verbesserung der Auralisation. Die perzeptive Relevanz der gerichteten Schallfeldanteile für eine adäquate Auralisation ist im Allgemeinen überraschend gering. Bei der Anwesenheit von starken ersten Reflexionen und einem geringen Verhältnis zwischen Direktschall und Nachhall erreicht diese Vorgehensweise zwar eine höhere Ähnlichkeit im Vergleich zur Referenz-Messung im Gegensatz zur Verwendung von ausschließlich diffusen Anteilen, für eine Vielzahl von Anwendungen kann aber womöglich auf die Synthese der gerichteten Anteile verzichtet werden. Denkbar ist auch der Einsatz von erheblich weniger gerichteten Anteilen, z.B. durch die ausschließliche Betrachtung von Reflexionen 1. Ordnung. Weiterhin sind alternative Ansätze zur Bestimmung der gerichteten Komponenten sinnvoll, um den Messaufwand gering zu halten.

Da die entwickelten Modelle zur Detektion der gerichteten Komponenten Methoden der Wellenfeldanalyse einsetzen, sind zwingend Messungen mit sphärischen Mikrofonarrays notwendig. Diese Systeme können wirksam zur Detektion von gerichteten Schallfeldanteilen verwendet werden und erlauben eine akkurate, räumliche Zerlegung sowie die Berechnung der spektralen Eigenschaften von Reflexionen. Auf Basis dieser Beschreibungen kann ein gerichtetes Schallfeld entsprechend der realen Situation synthetisiert werden. Ebenso konnte die Verwertbarkeit von Mikrofonarrays niedriger Ordnung für diese Methoden gezeigt werden, signifikante Unterschiede der Synthese basierend auf Systemen mit 1202 bzw. 86 Mikrofonen wurden nicht beobachtet. Bei der Verwendung entsprechender Radien (gemäß des Ohrabstandes) können auch binaurale Merkmale wie die interaurale Kohärenz präzise bestimmt werden. Die Unterschiede des Kohärenzverlaufs von zwei gegenüberliegenden Mikrofonen des Mikrofonarrays im Vergleich zu dem einer Kunstkopf-Messung sind gering. Dies ermöglicht den Einsatz von Echtzeit-Mikrofonarrays zur Synthese von Schallfeldern gemäß der vorgestellten, parametrischen Modelle.

Dieser Ansatz zur Synthese der binauralen Raumimpulsantworten schafft eine Reihe von neuen Möglichkeiten und Anwendungen. So können anhand der Modelle durch die Kombination eines zwei-kanaligen Rauschsignals, einem sphärischen Datensatz von Außenohrübertragungsfunktionen und einer entsprechenden parametrischen Beschreibung beliebige binaurale Raumimpulsantworten generiert werden. Eine flexible Adaption der akustischen Merkmale eines realen Raumes wie z.B. eine Veränderung der Nachhallzeit, des Reflexionsmusters oder der Diffusität ist leicht realisierbar. Dies ermöglicht die Hörbarmachung von raumakustischen Modifikationen (z.B. durch bauliche Veränderungen) oder die gezielte Generierung raumakustischer Kriterien zur perzeptiven Untersuchung (z.B. von Wahrnehmbarkeitsschwellen). Weiterhin erlaubt der Einsatz von individuellen Außenohrübertragungsfunktionen die Personalisierung der Hörbar-

machung. Wurde ein Raum parametrisch erfasst, so kann eine Individualisierung der binauralen Raumimpulsantworten erfolgen, sodass man sich in eine beliebige akustische Umgebung hineinversetzen kann, bei gleichzeitiger Verwendung der eigenen Außenohrübertragungsfunktionen. Kreative Anwendungen der parametrischen Synthese binauraler Raumimpulsantworten sind ebenfalls denkbar, z.B. in Form eines parametrischen Faltungshalls oder eines Werkzeugs zur gezielten Abänderung der Raumeindrucks einer realen Umgebung.

Ein Ausbau des Systems, um weitere Freiheitsgrade der Kopforientierung des Nutzers in der Auralisation zu berücksichtigen, ist leicht möglich. Die Beschränkung der dynamischen Adaption auf die horizontale Ebene begründet sich ausschließlich durch die Verwendung des Binauralrenderers und dessen Funktionalitäten. Da sphärische Außenohrübertragungsfunktionen bereits benutzt werden, ist eine Nachführung von vertikalen Kopfbewegungen leicht integrierbar. Eine Berücksichtigung von translatorischen Änderungen könnte ebenso auf Basis der Parameter erfolgen, hierzu werden aber Anpassung des Systems erforderlich.

So präsentiert diese Arbeit praktische Ansätze zur erheblichen Vereinfachung des Schallfeldes hinsichtlich der auditiven Perzeption und soll für eine weitere Verwendung und Untersuchung dieser Funktionalitäten motivieren. Dabei können VR-Anwendungen oder gar Systeme zur *erweiterten Realität* (AR) eine große Rolle spielen. Die Verknüpfung der Realität mit virtuellen Elementen ist visuell als auch auditiv hoch anspruchsvoll. So sollen beispielsweise virtuelle, visuelle Objekte im Raum erscheinen und entsprechend der realen Lichtsituation erzeugt werden. Weiterhin muss auch die adäquate Integration einer virtuellen Schallquelle in die reale Situation entsprechend des Raumeindrucks erfolgen. Dabei können in dieser Arbeit entwickelte Methoden und deren Ergebnisse zuträglich sein. Es konnte eine starke Relevanz des diffusen Nachhalls gezeigt werden, wohingegen den frühen Reflexionen eine deutlich geringe Bedeutung zugesprochen werden kann. Für AR-Anwendungen bedeutet dies, dass der technische Aufwand zur Detektion von Reflexionen und dessen Genauigkeit womöglich stark vereinfacht werden kann, sodass alternative Methoden denkbar werden, wie z.B. optische Ansätze zur Analyse des Reflexionsverhalten. Daher sollten diese Möglichkeiten der Synthese auf Basis einer parametrischen Beschreibung nun anwendungsorientiert implementiert und weiterentwickelt werden, sodass sie für den innovativen und aufstrebenden Markt der VR/AR-Technologien relevant werden können und so im Allgemeinen die Klangqualität der auditiven Komponente in diesen Systemen fördern.

ANHANG **A**

Abbildungen

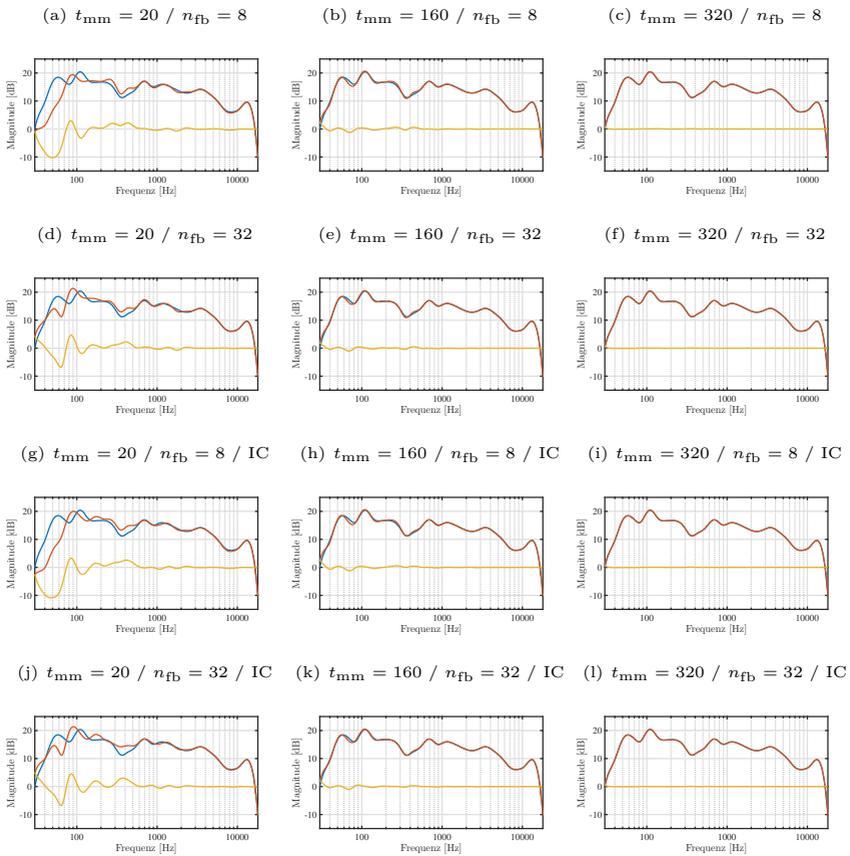


Abbildung A.1: Experiment LEA1: Spektrum Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

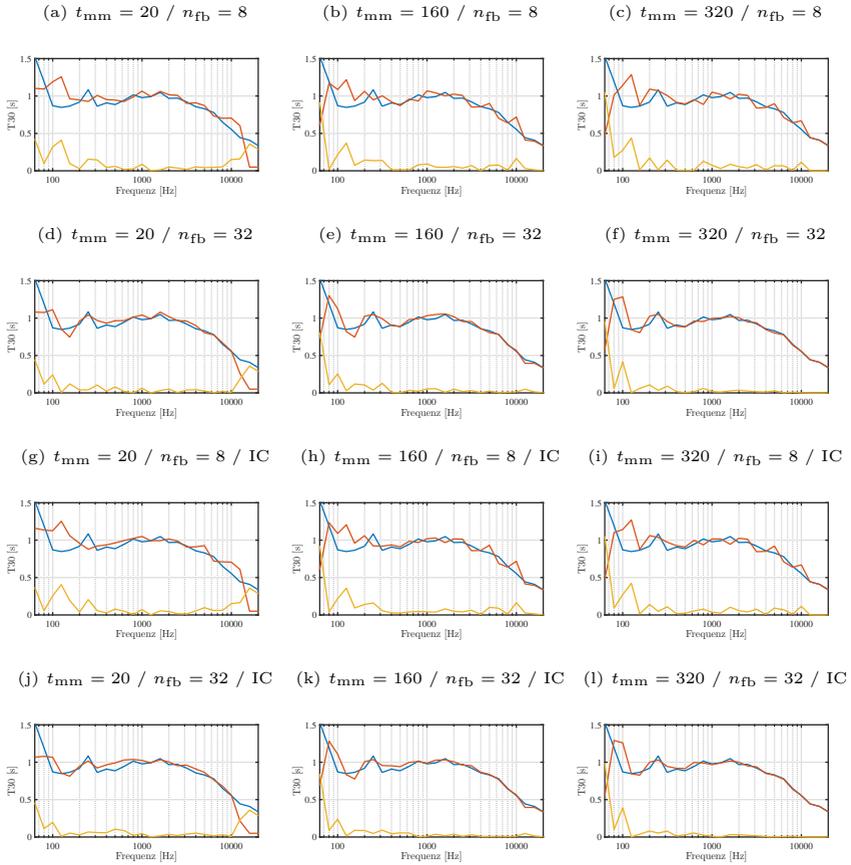


Abbildung A.2: Experiment LEA1: Nachhallzeit T_{30} Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

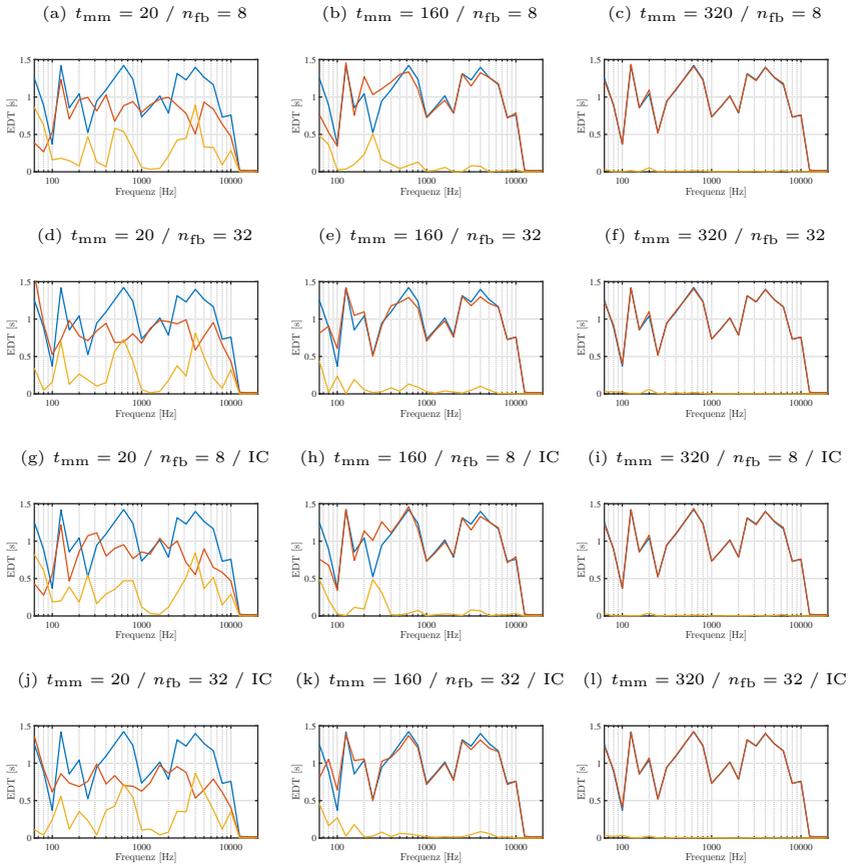


Abbildung A.3: Experiment LEA1: Frühe Abklingzeit EDT Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

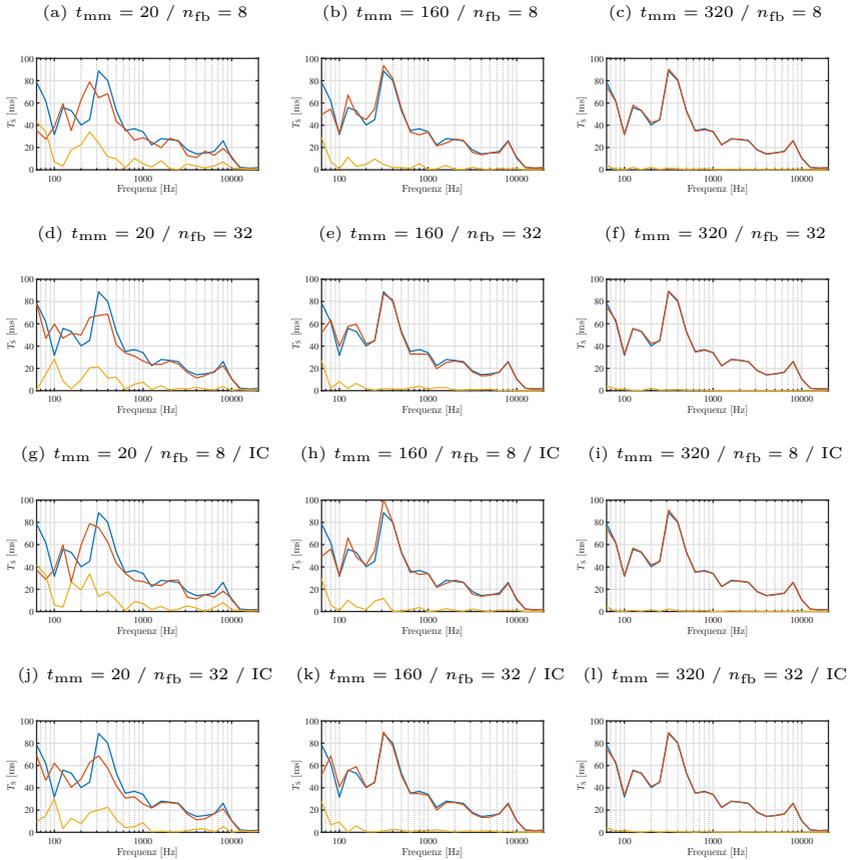


Abbildung A.4: Experiment LEA1: Schwerpunktzeit T_S Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

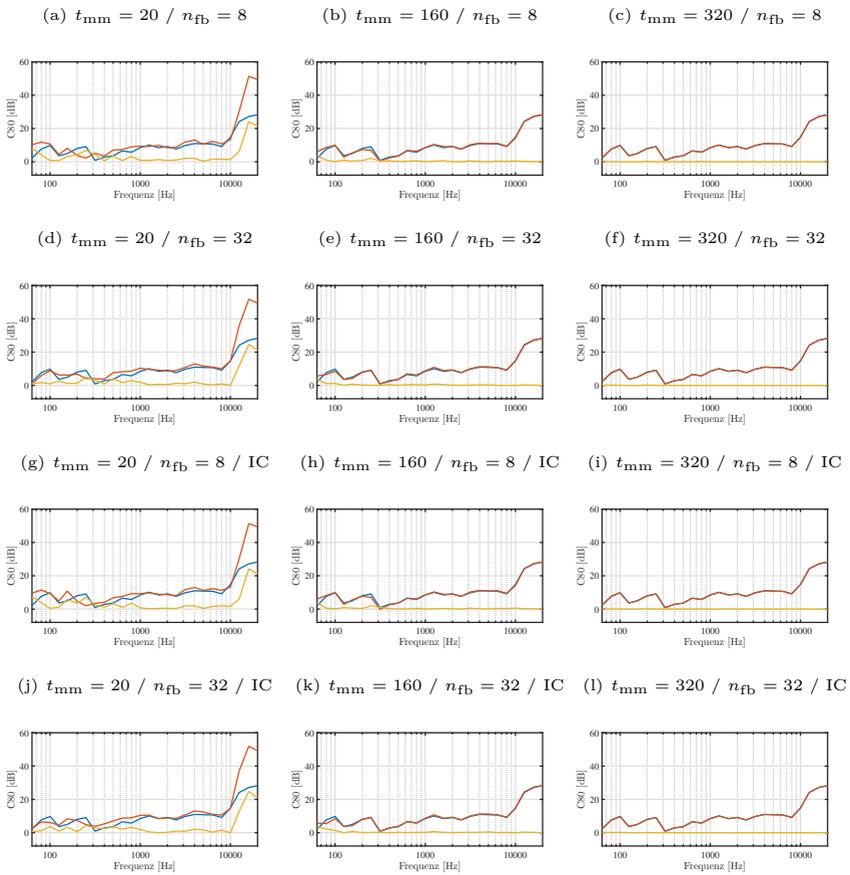


Abbildung A.5: Experiment LEA1: Klarheitsmaß $C80$ Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

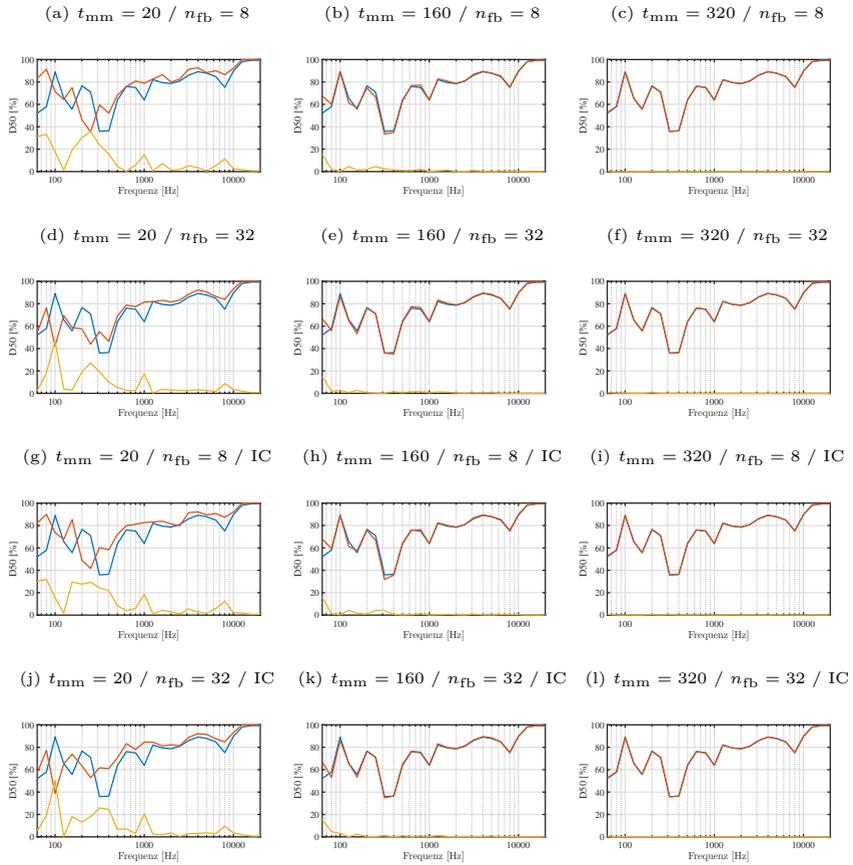


Abbildung A.6: Experiment LEA1: Deutlichkeit $D50$ Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

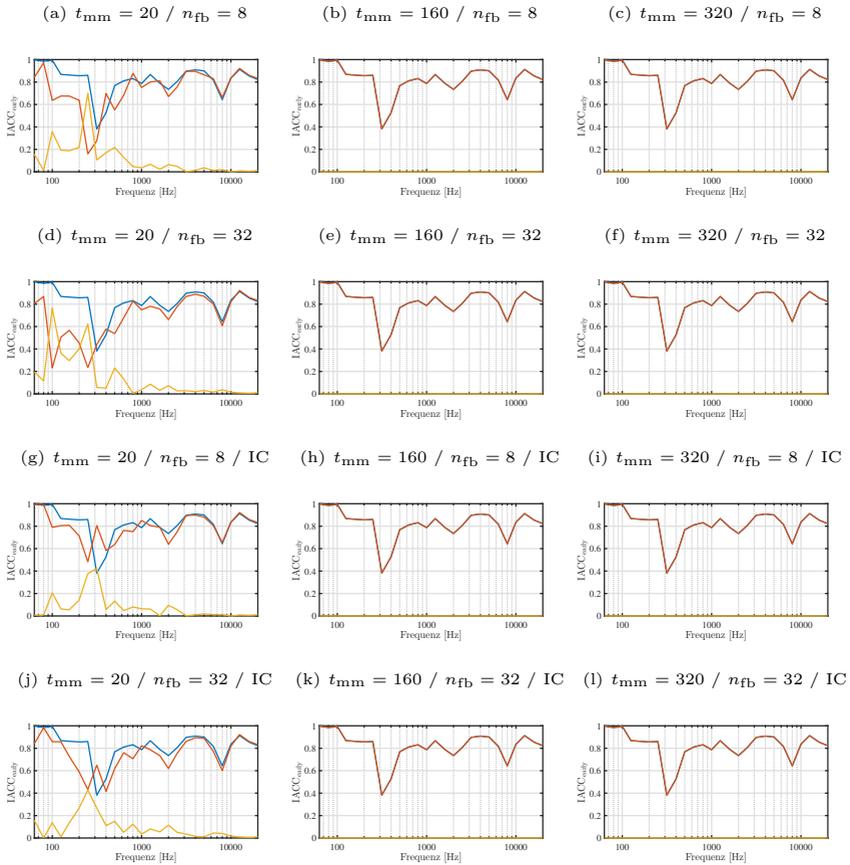


Abbildung A.7: Experiment LEA1: $IACC_{early}$ Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

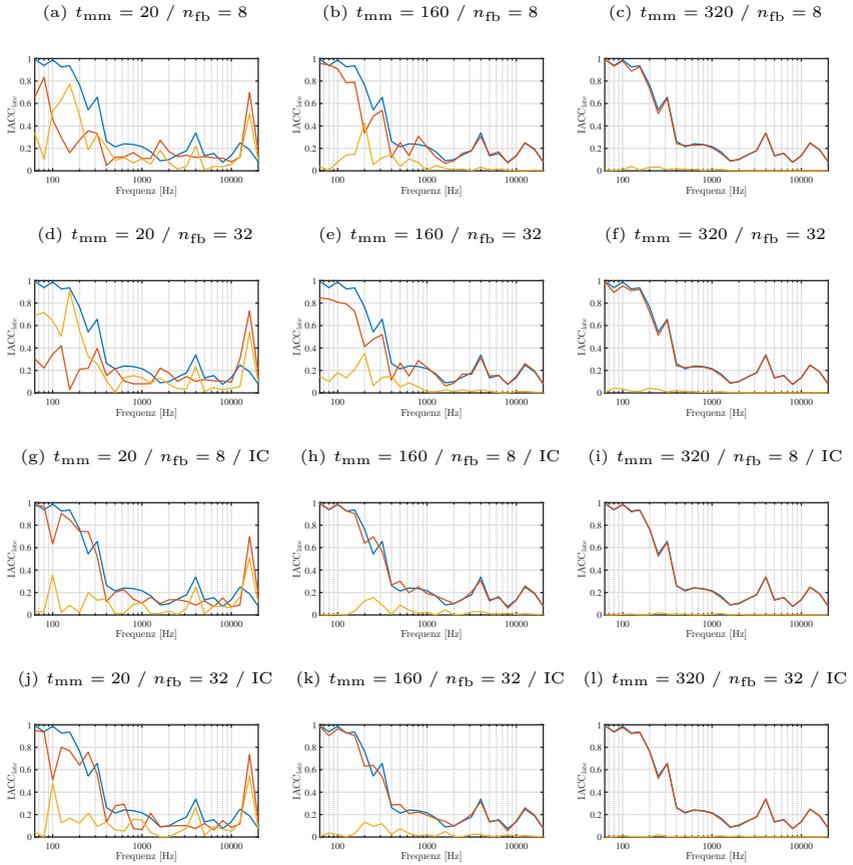


Abbildung A.8: Experiment LEA1: $IACC_{late}$ Synthese (KSS direktional, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

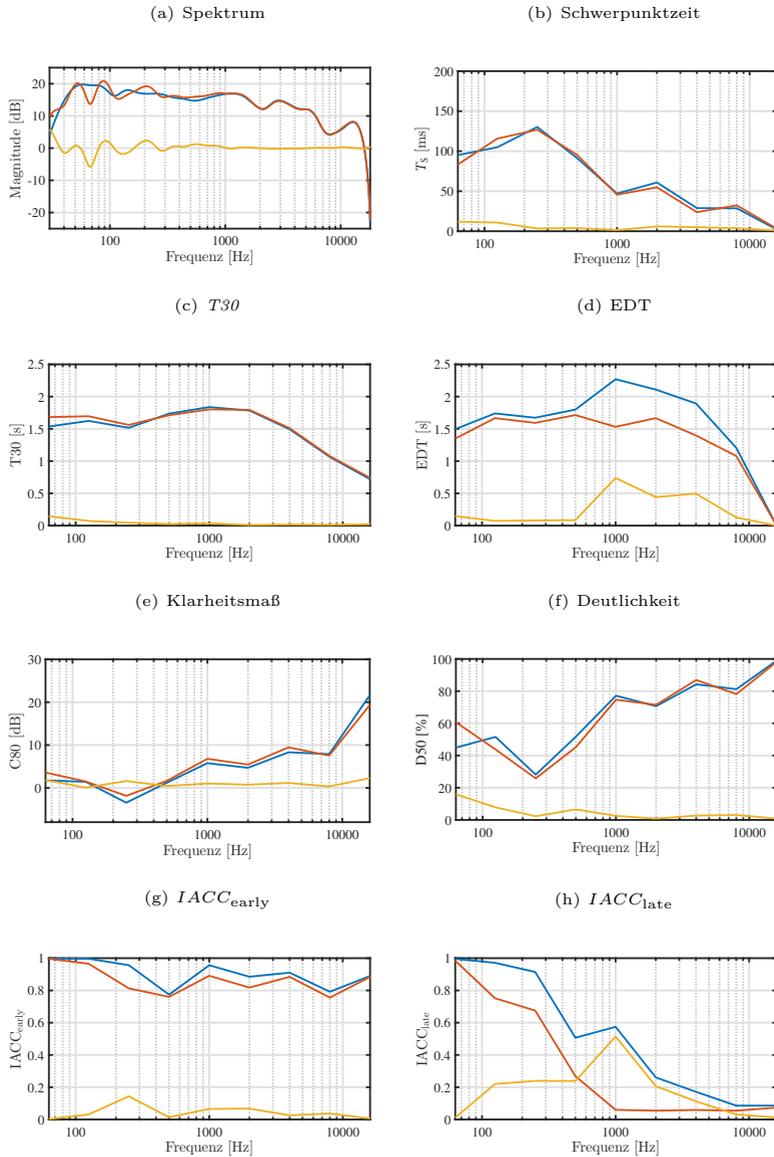


Abbildung A.9: Experiment LEA2: Raumakustische Kriterien Synthese (KVB direk-
tional $t_{mm} = 40$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

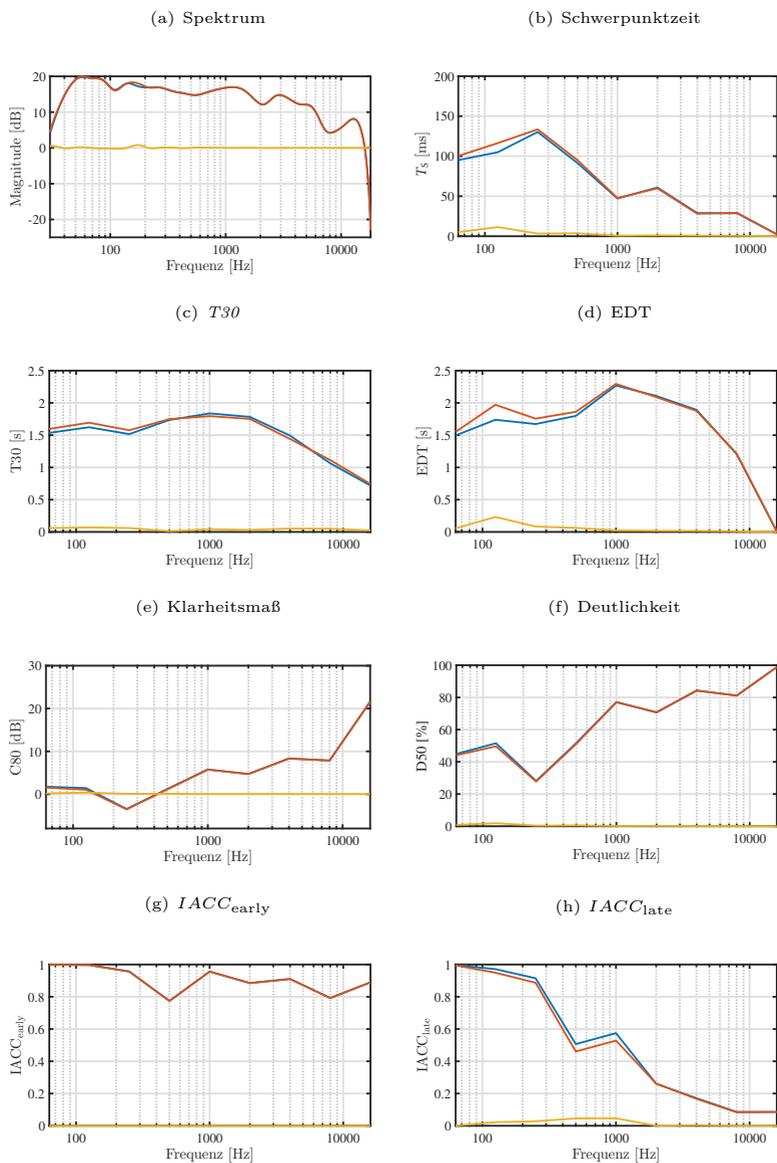


Abbildung A.10: Experiment LEA2: Raumakustische Kriterien Synthese (KVB di-rektional $t_{mm} = 320$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

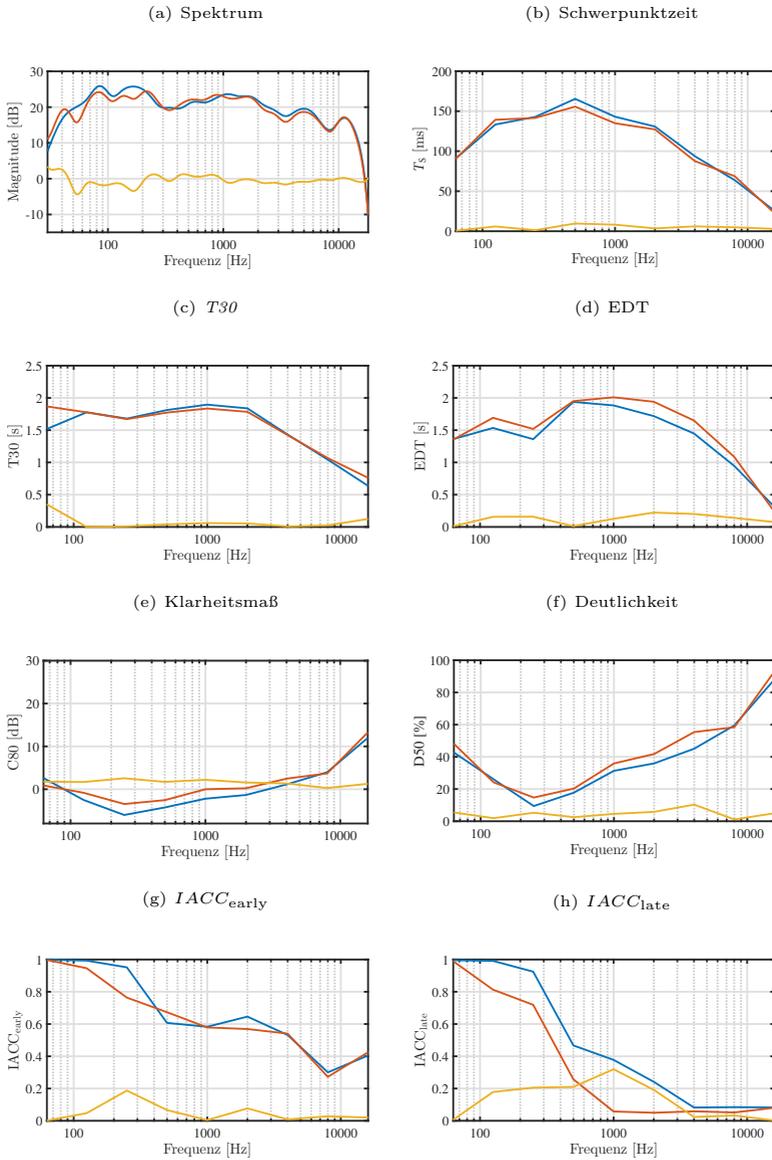


Abbildung A.11: Experiment LEA2: Raumakustische Kriterien Synthese (KVB omnidirektional $t_{mm} = 40$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

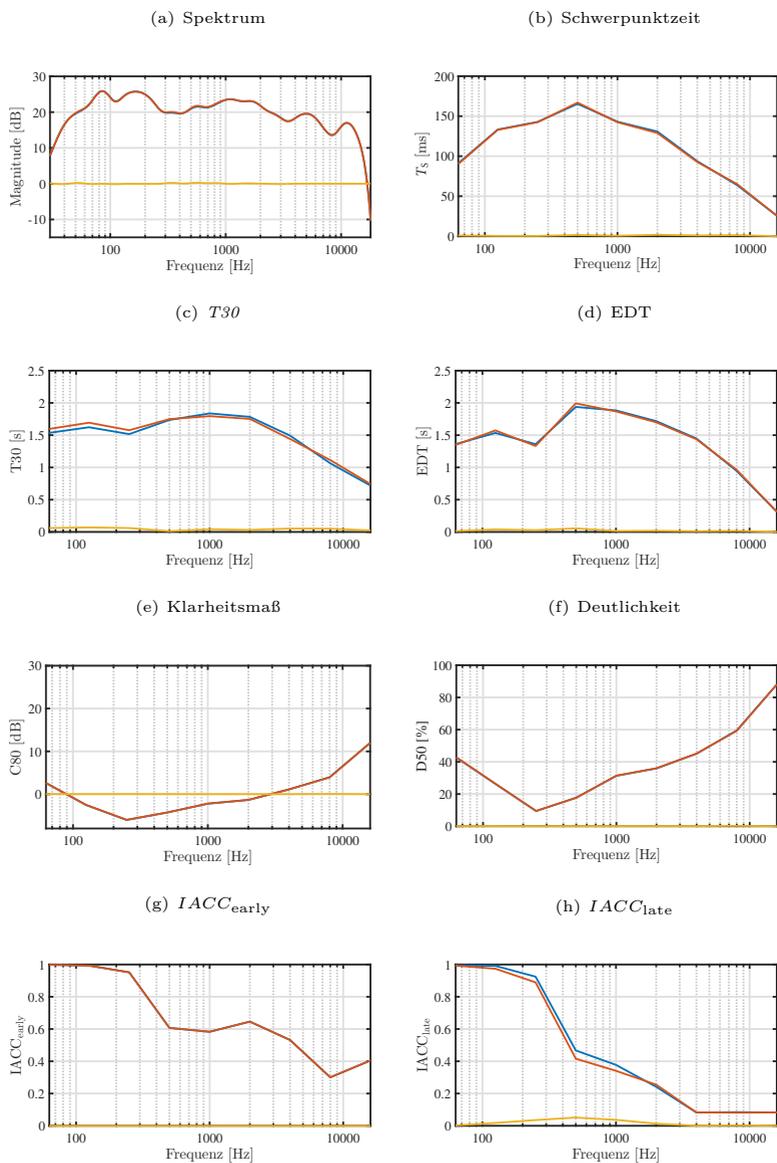


Abbildung A.12: Experiment LEA2: Raumakustische Kriterien Synthese (KVB omnidirektional $t_{mm} = 320$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

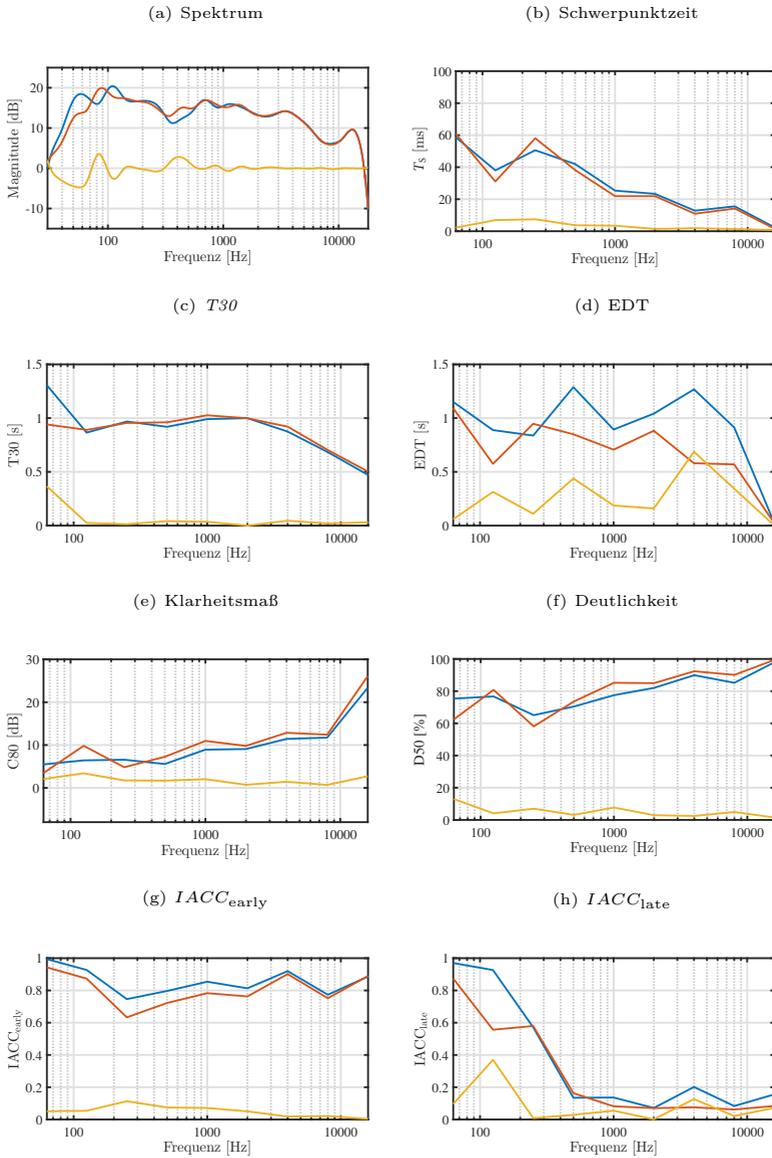


Abbildung A.13: Experiment LEA2: Raumakustische Kriterien Synthese (KSS directional $t_{mm} = 20$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

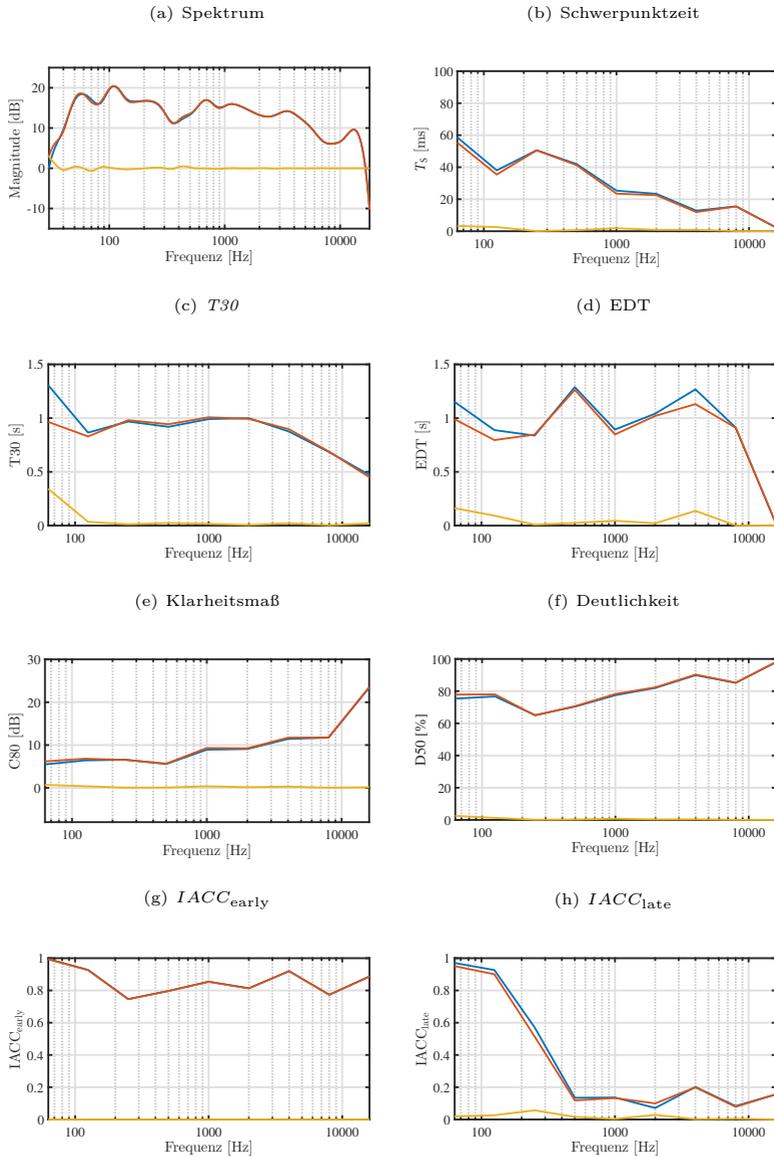


Abbildung A.14: Experiment LEA2: Raumakustische Kriterien Synthese (KSS di-rektional $t_{\text{mm}} = 160$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

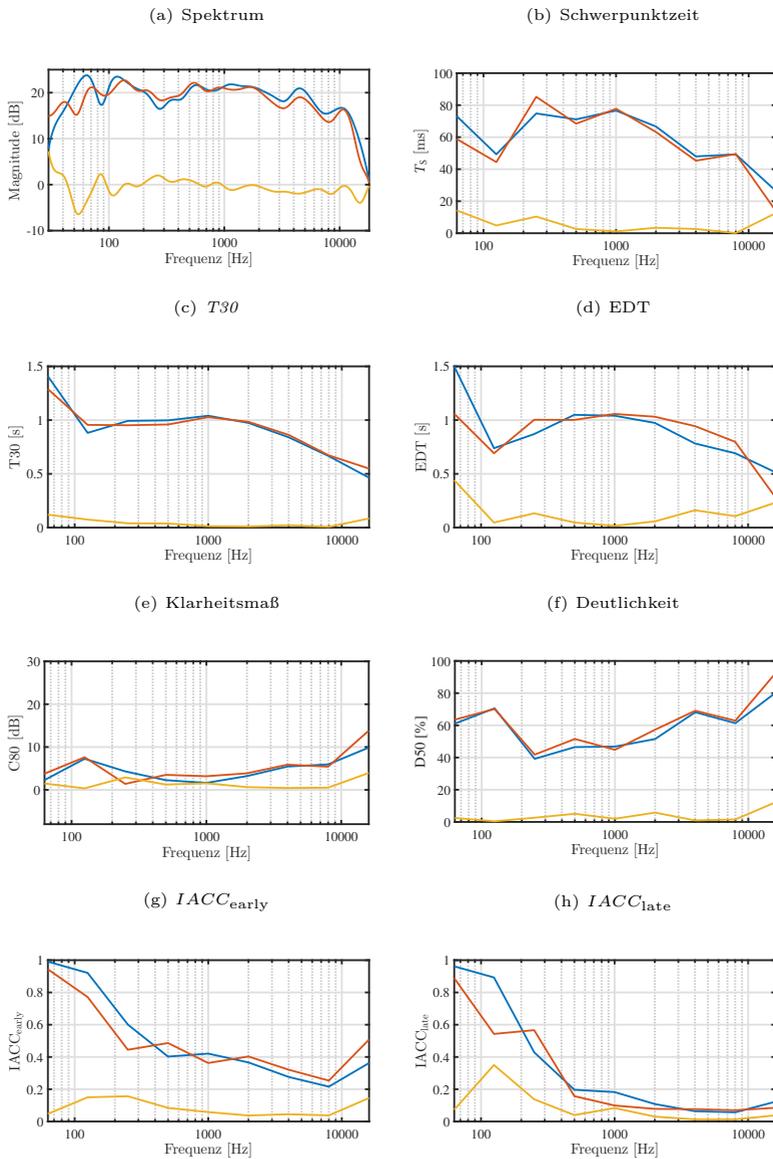


Abbildung A.15: Experiment LEA2: Raumakustische Kriterien Synthese (KSS omnidirektional $t_{mm} = 20$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

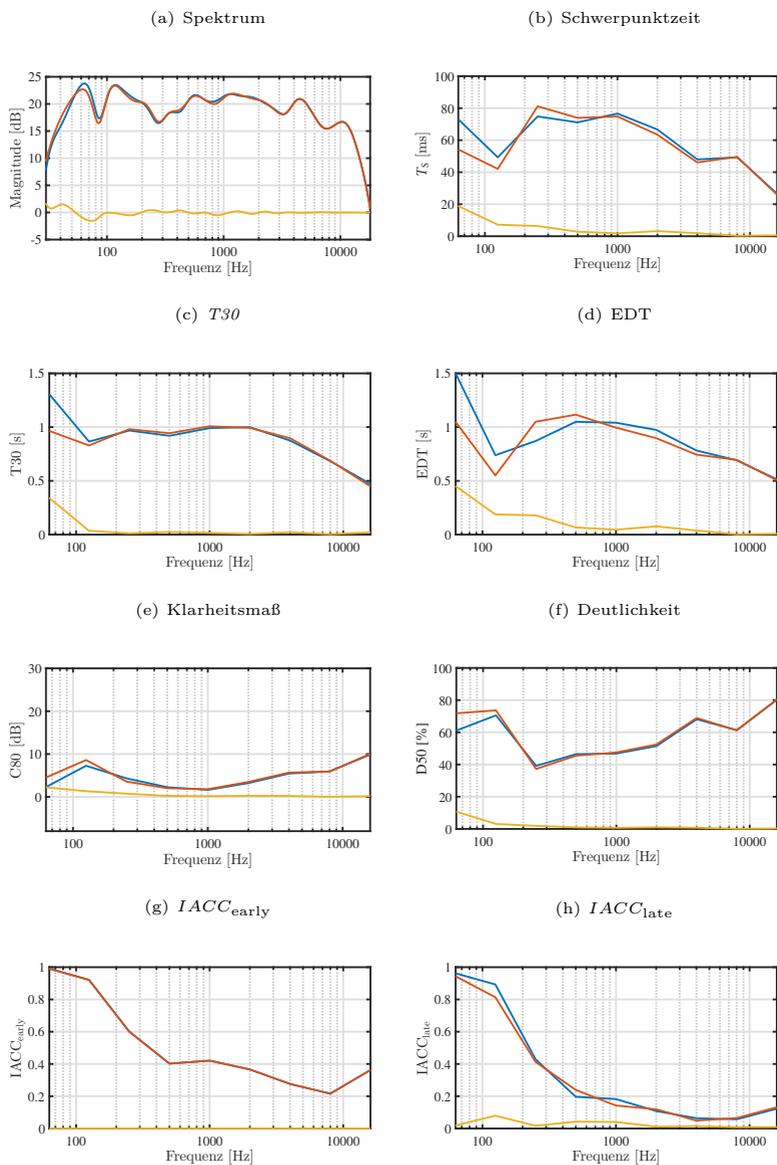


Abbildung A.16: Experiment LEA2: Raumakustische Kriterien Synthese (KSS omnidirektional $t_{imm} = 160$, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

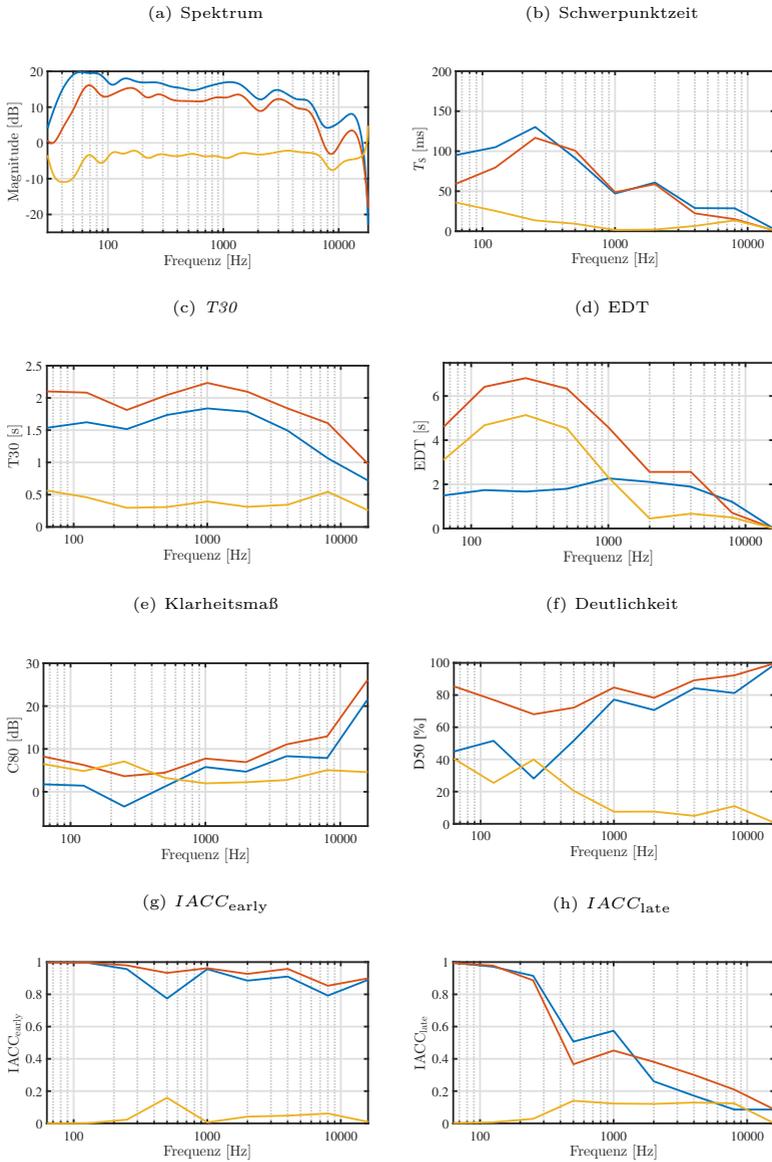


Abbildung A.17: Experiment LEB: Raumakustische Kriterien Synthese gerichtete Anteile (KVB directional RFL, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

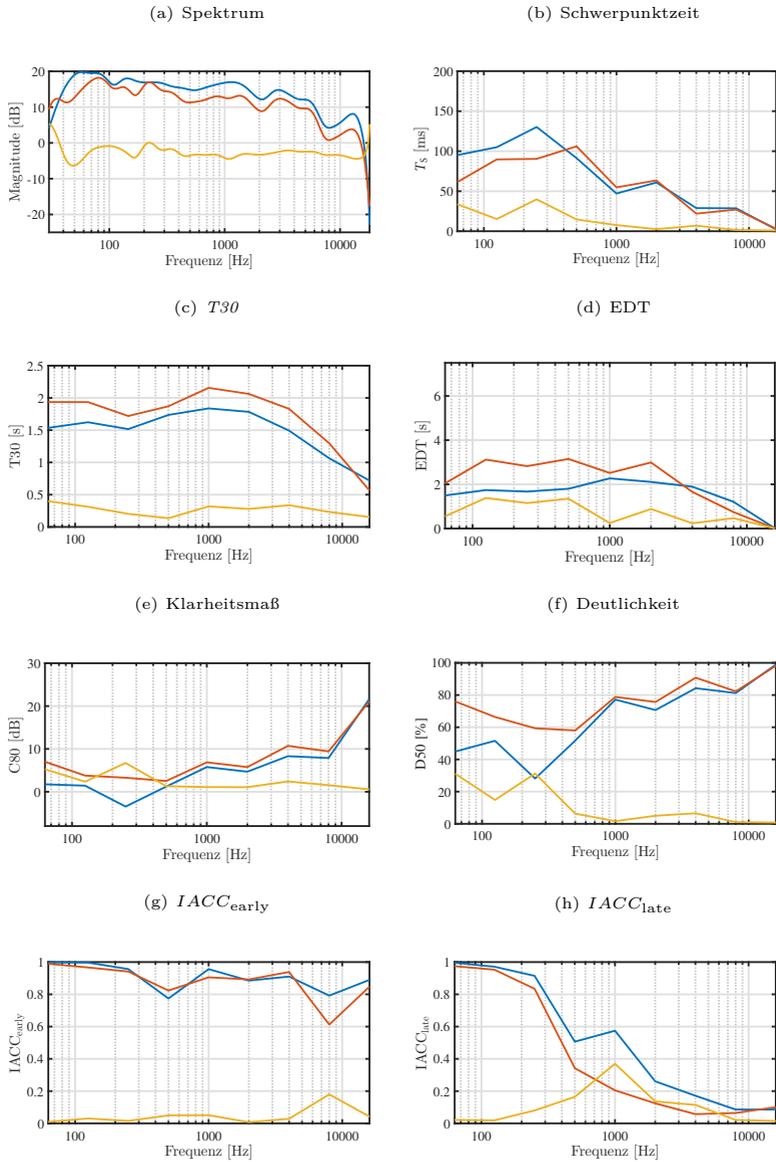


Abbildung A.18: Experiment LEB: Raumakustische Kriterien Synthese diffuse Anteile (KVB direktional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

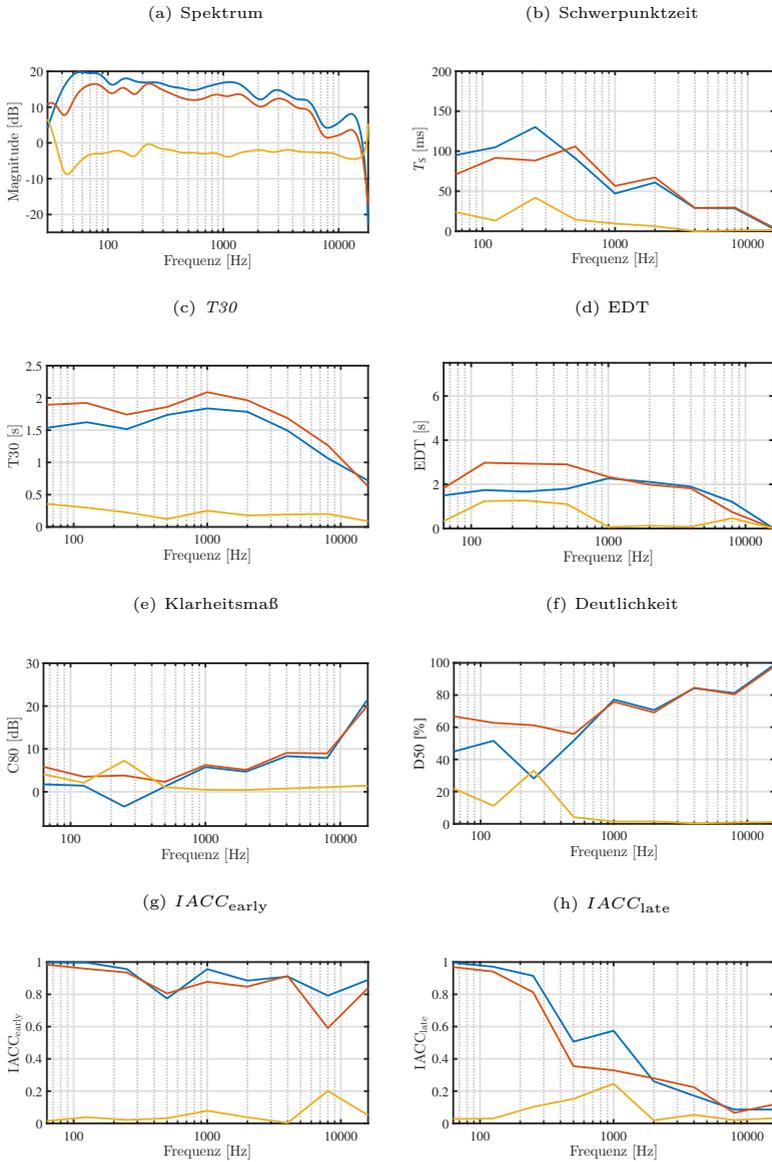


Abbildung A.19: Experiment LEB: Raumakustische Kriterien Synthese gerichte+diffuse Anteile (KVB direktional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

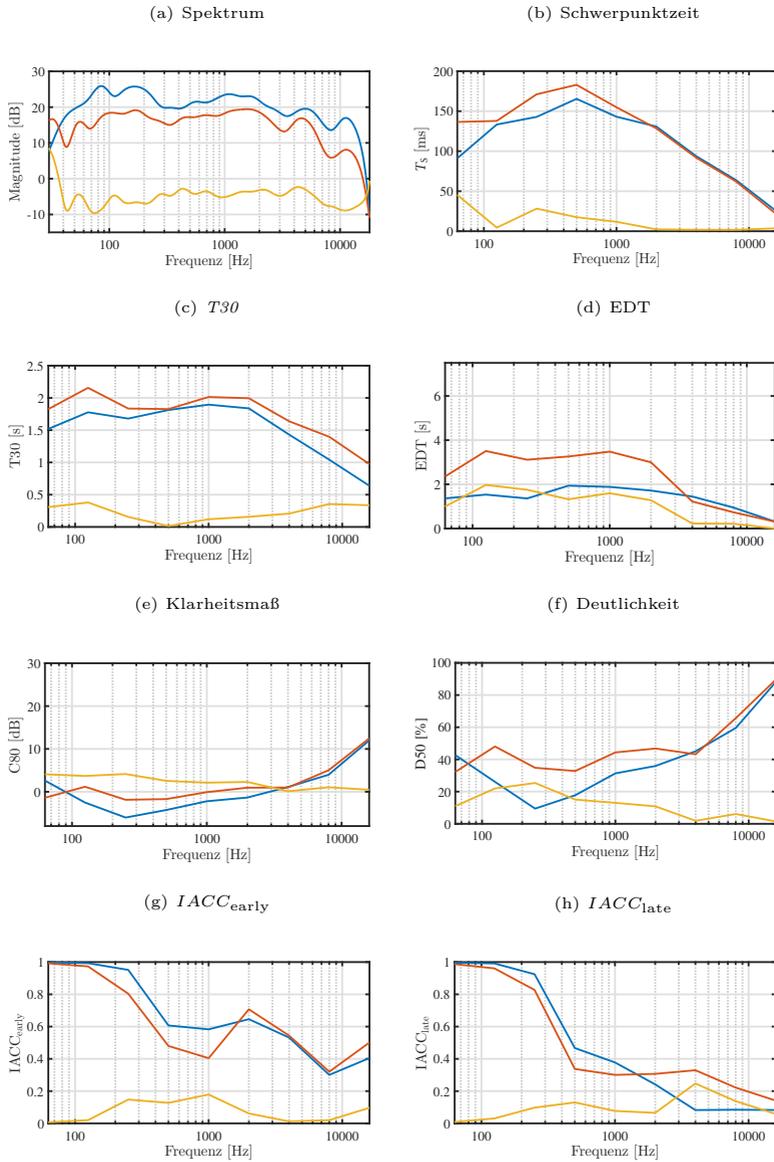


Abbildung A.20: Experiment LEB: Raumakustische Kriterien Synthese gerichtete Anteile (KVB omnidirektional RFL, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

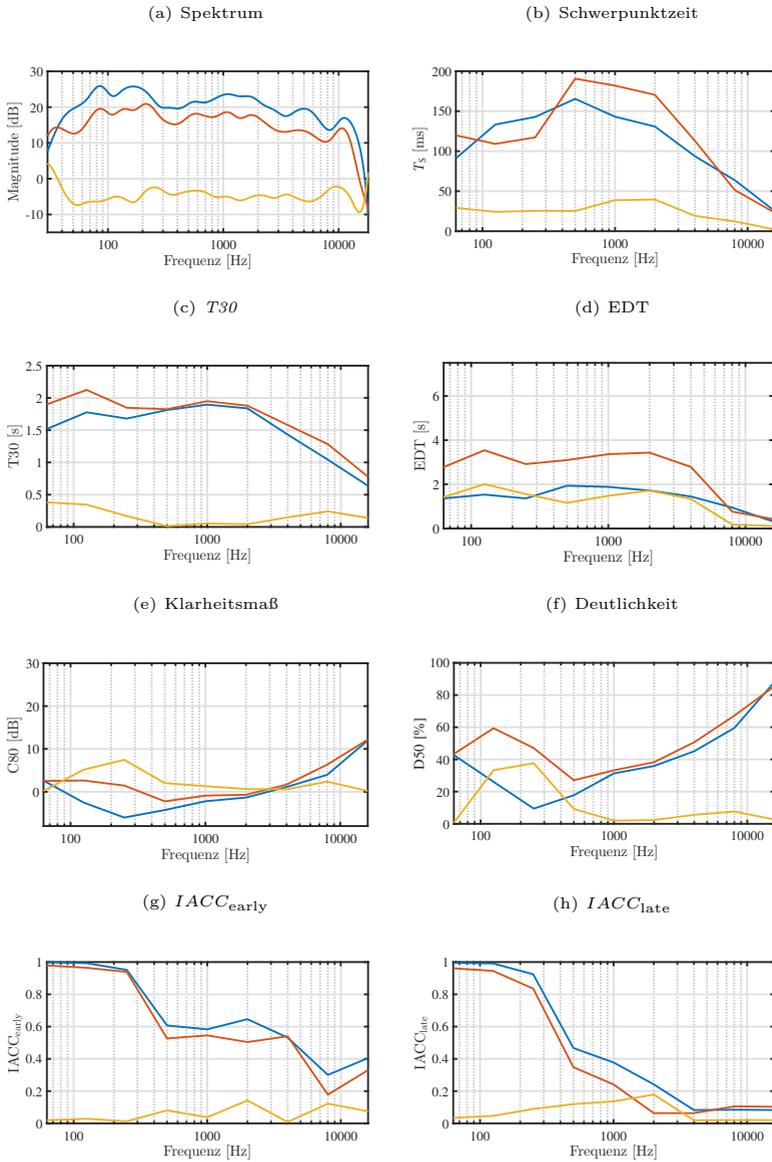


Abbildung A.21: Experiment LEB: Raumakustische Kriterien Synthese diffuse Anteile (KVB omnidirektional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

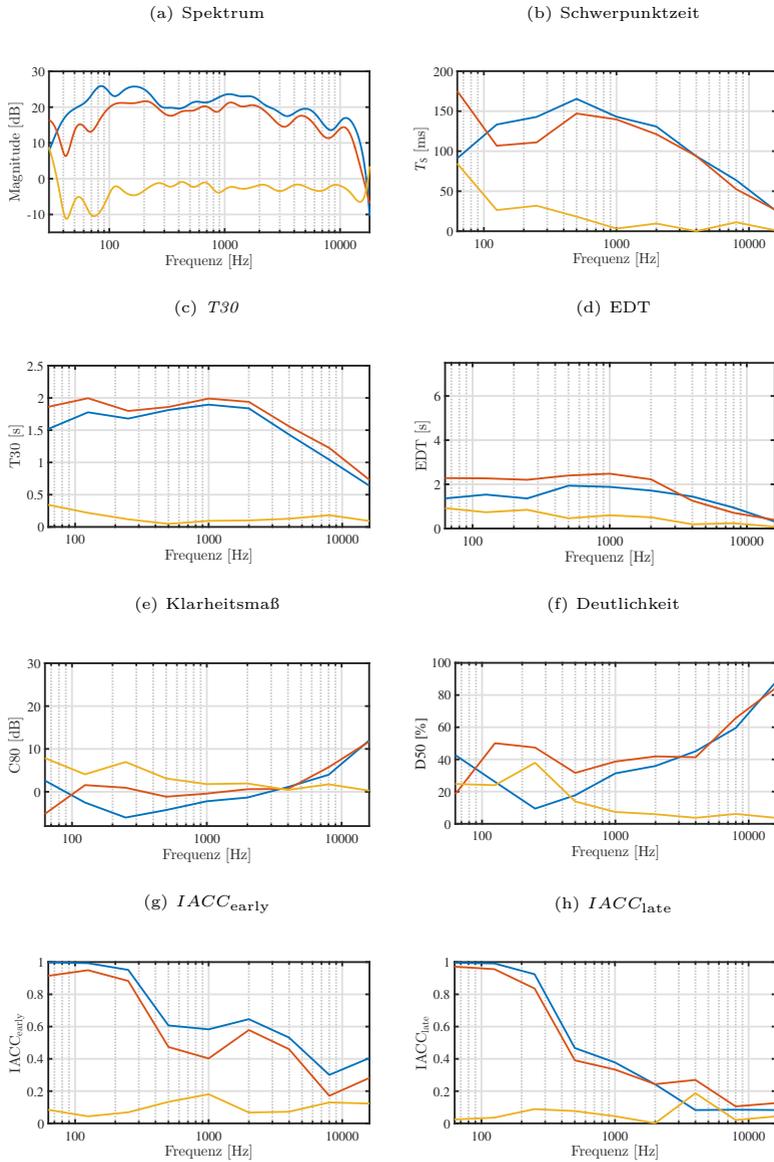


Abbildung A.22: Experiment LEB: Raumakustische Kriterien Synthese gerichte-te+diffuse Anteile (KVB omnidirektional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

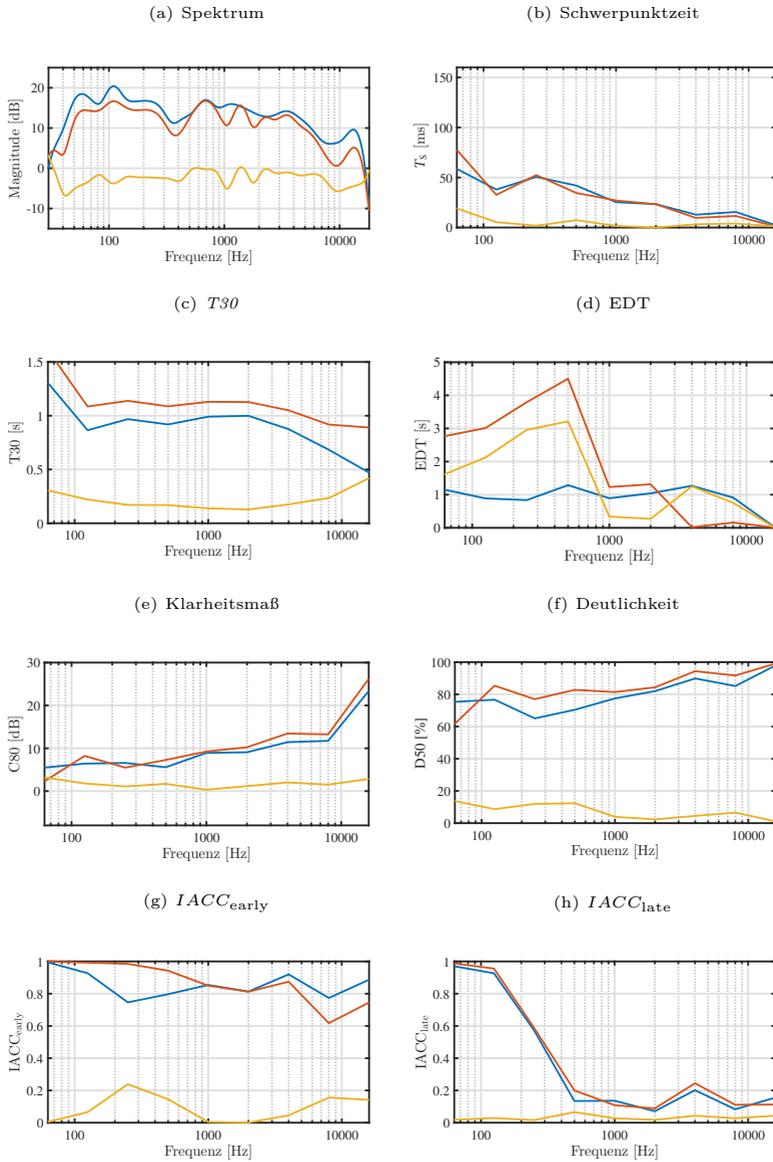


Abbildung A.23: Experiment LEB: Raumakustische Kriterien Synthese gerichtete Anteile (KSS directional RFL, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

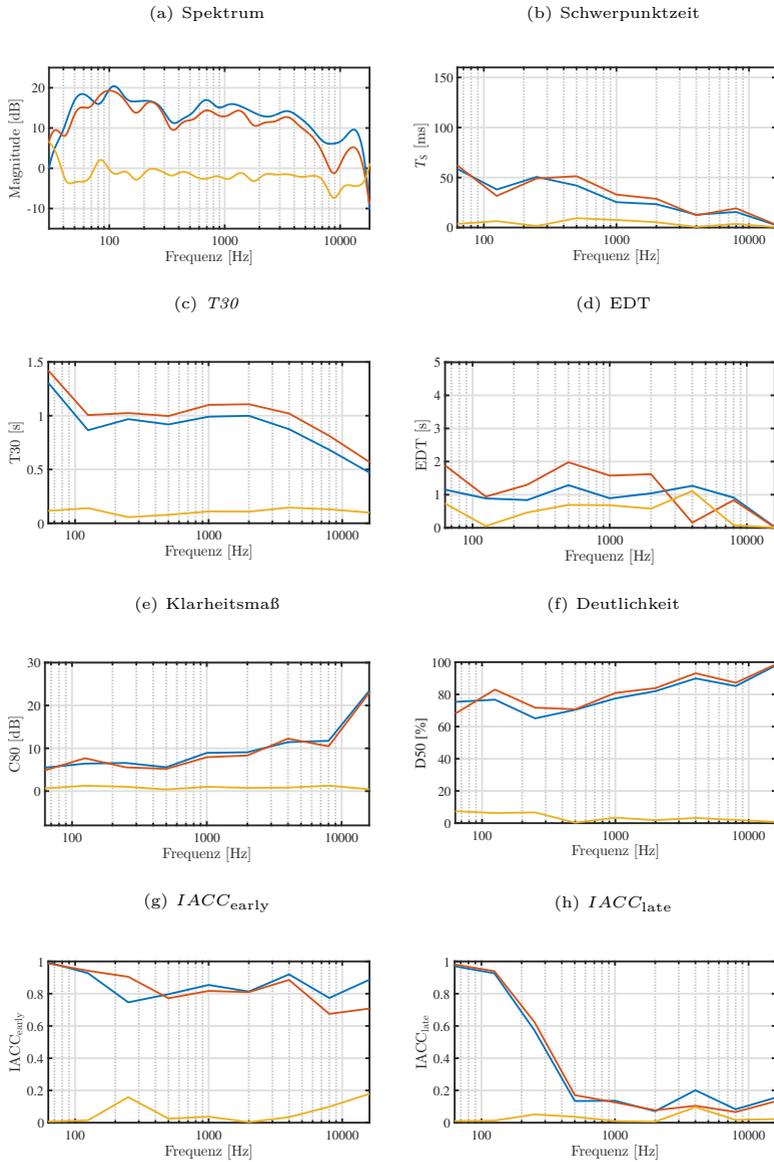


Abbildung A.24: Experiment LEB: Raumakustische Kriterien Synthese diffuse Anteile (KSS direktional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

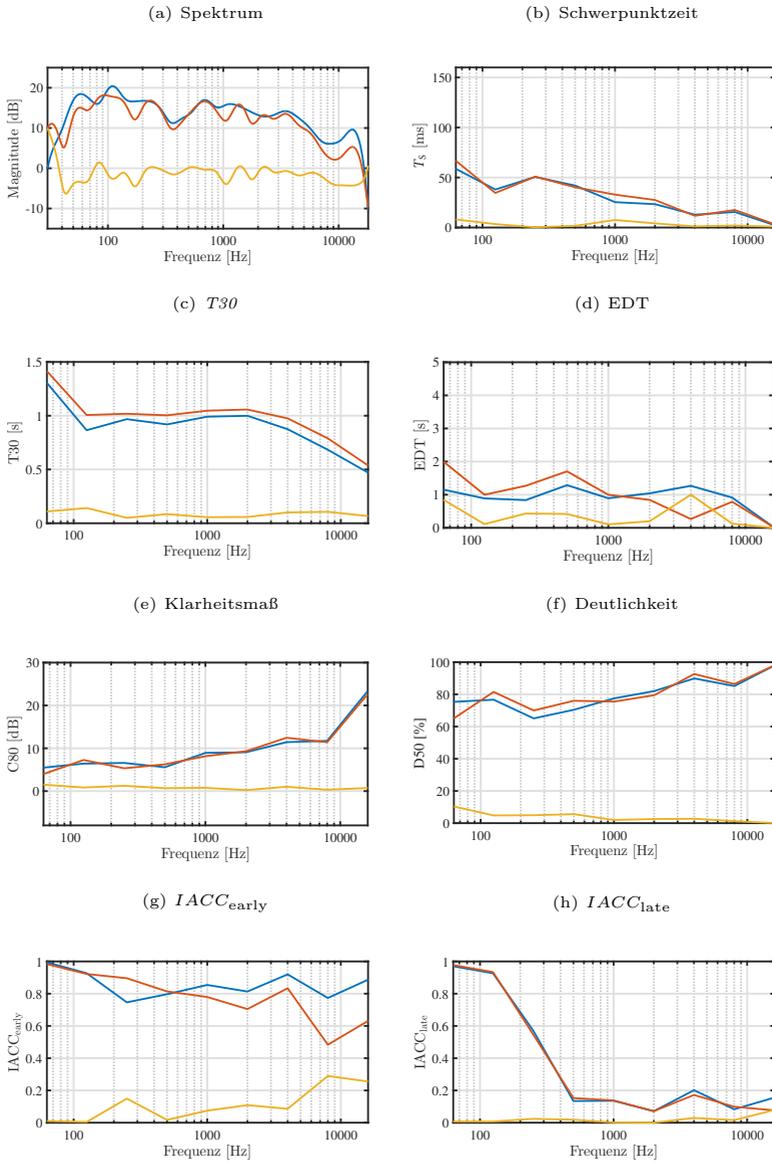


Abbildung A.25: Experiment LEB: Raumakustische Kriterien Synthese gerichte+diffuse Anteile (KSS direktional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

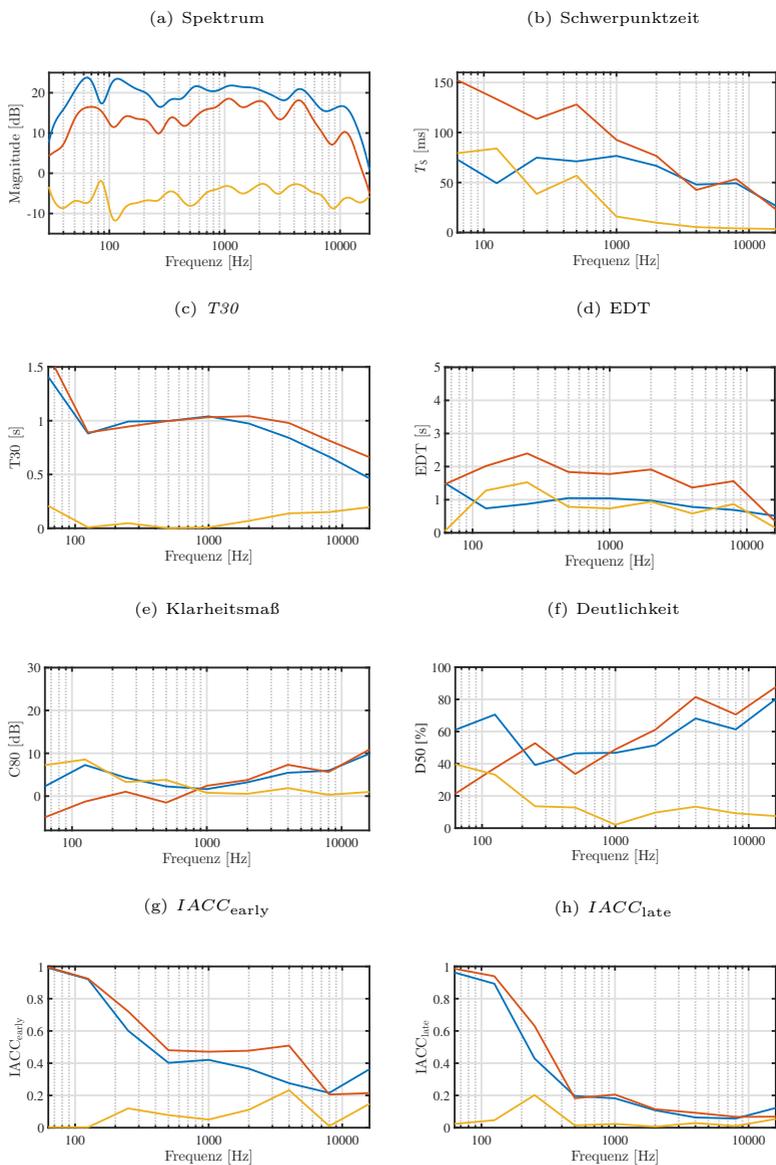


Abbildung A.26: Experiment LEB: Raumakustische Kriterien Synthese gerichtete Anteile (KSS omnidirektional RFL, rot), Vergleich mit Referenz-BRIR (blau) und Differezbetrag (gelb)

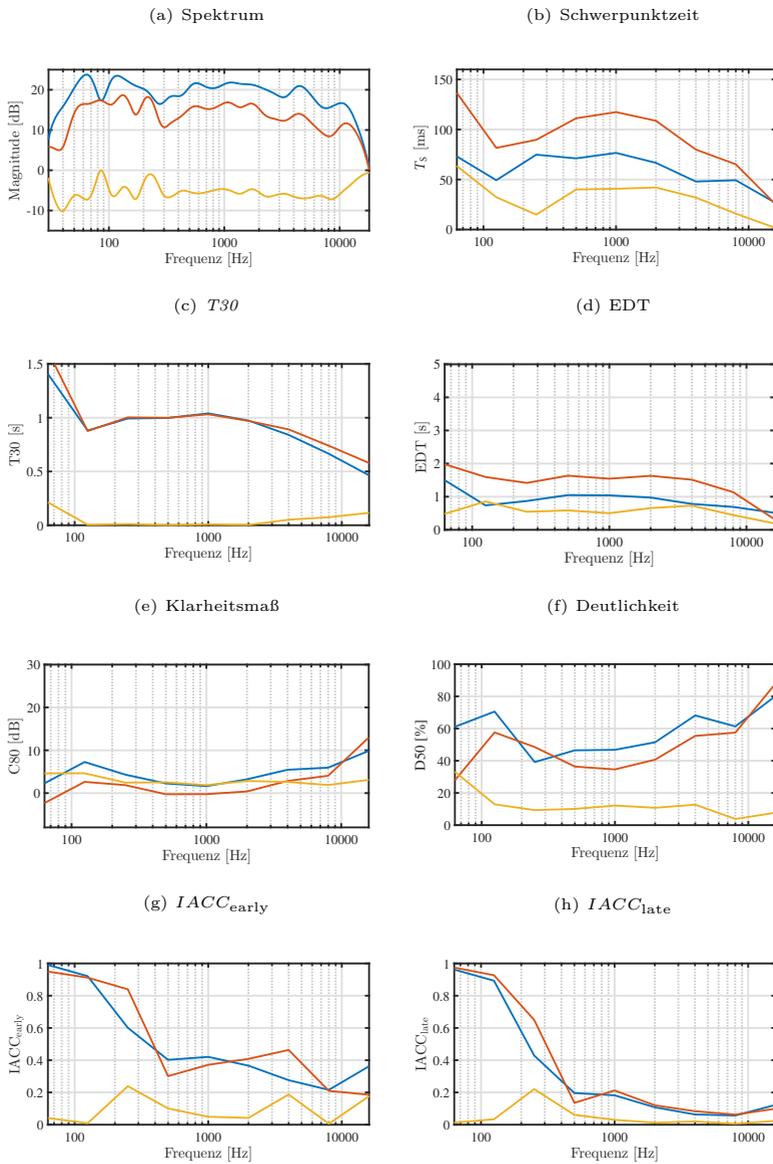


Abbildung A.27: Experiment LEB: Raumakustische Kriterien Synthese diffuse Anteile (KSS omnidirektional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

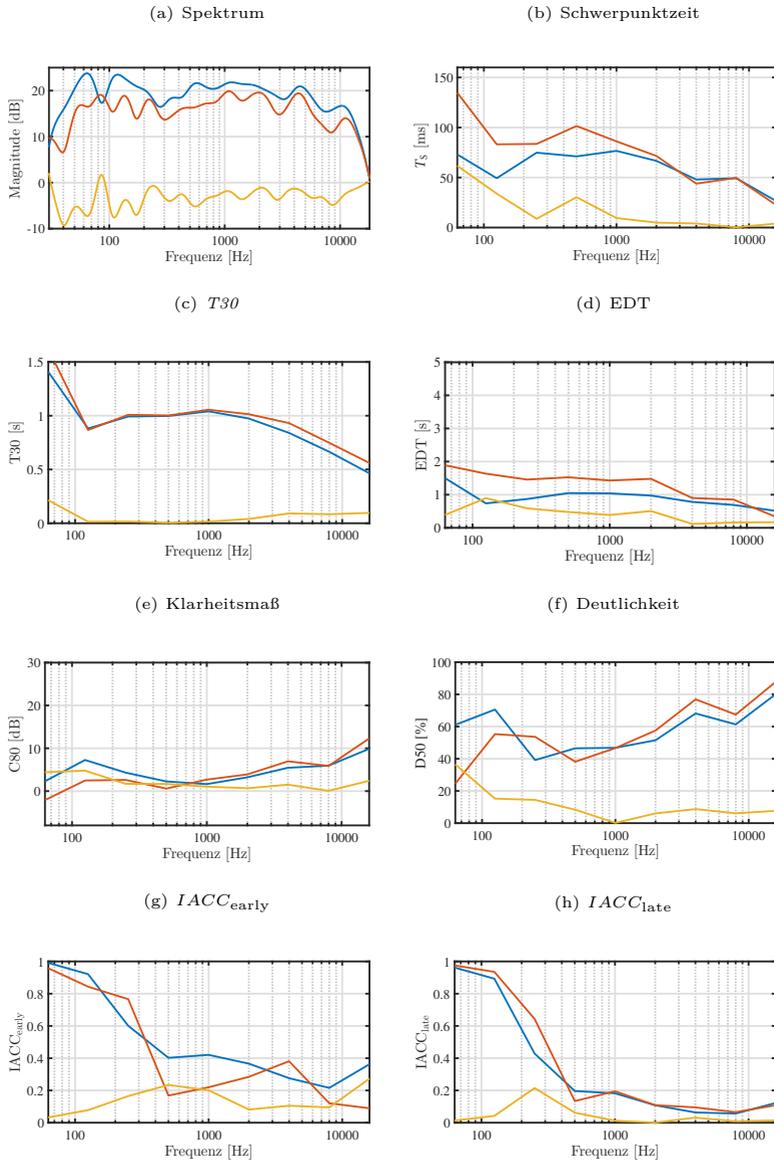


Abbildung A.28: Experiment LEB: Raumakustische Kriterien Synthese gerichte+diffuse Anteile (KSS omnidirektional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

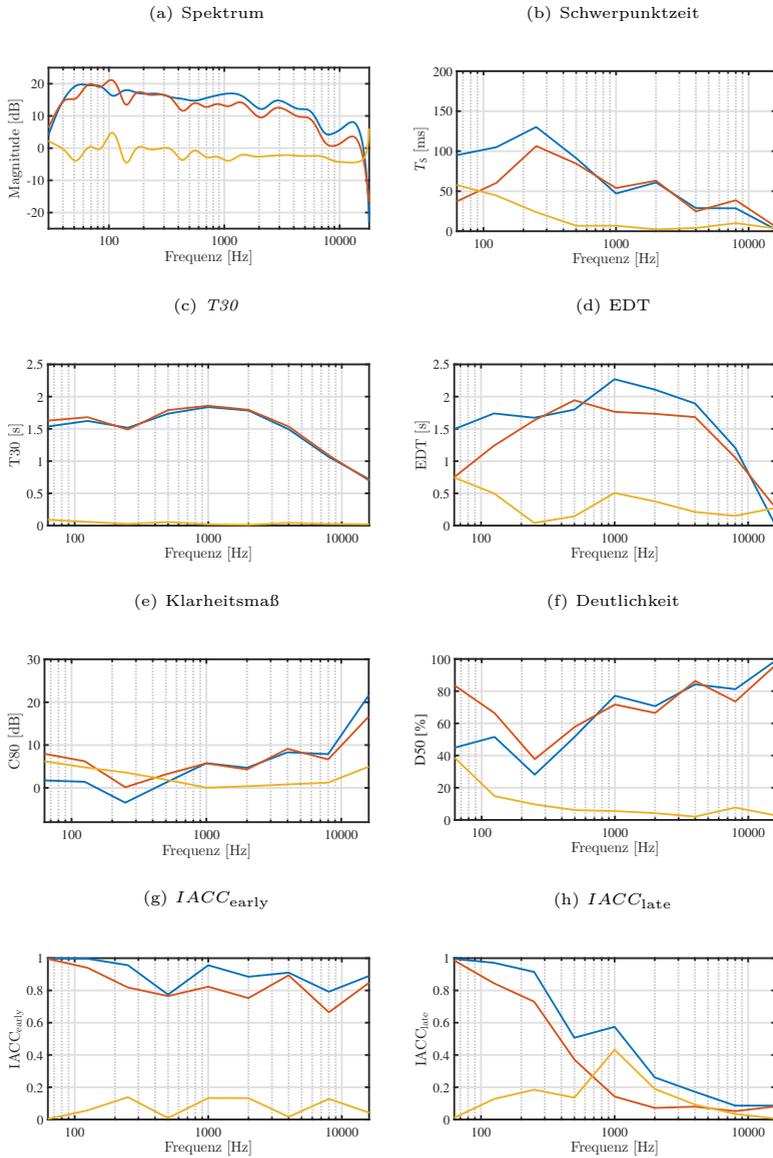


Abbildung A.29: Experiment LEC: Raumakustische Kriterien Synthese diffuse Anteile (KVB *directional* RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

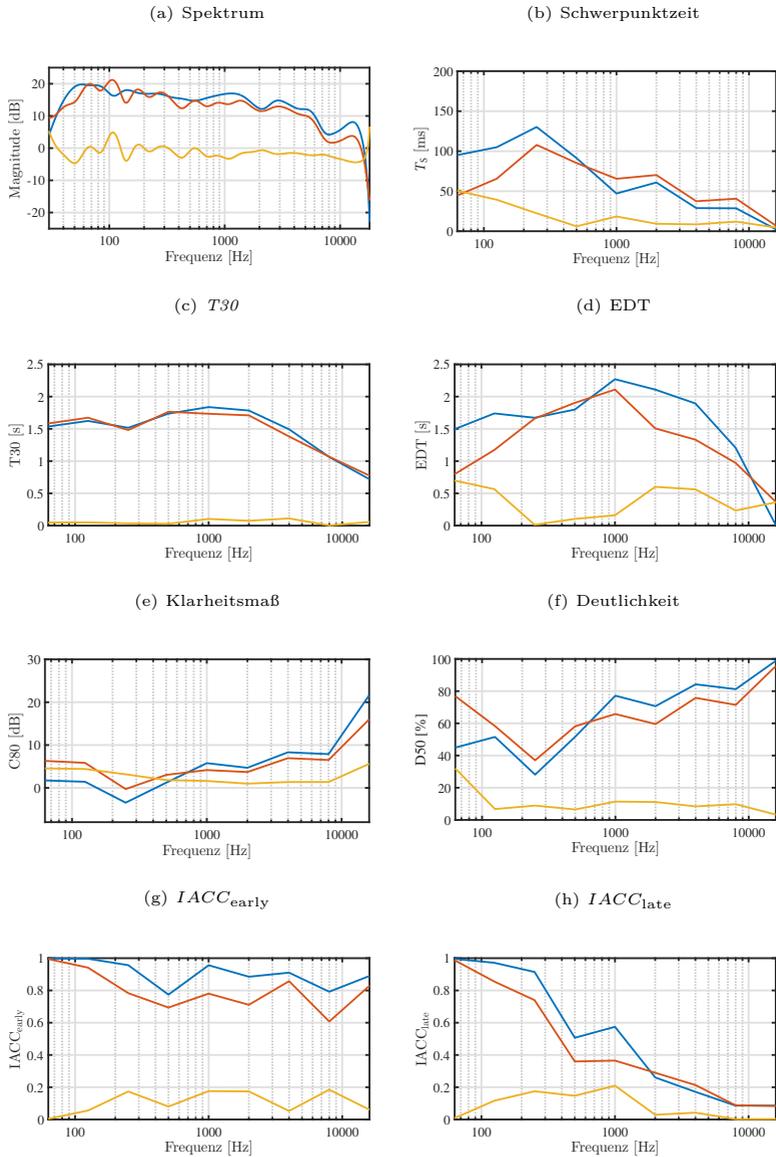


Abbildung A.30: Experiment LEC: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KVB direktional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

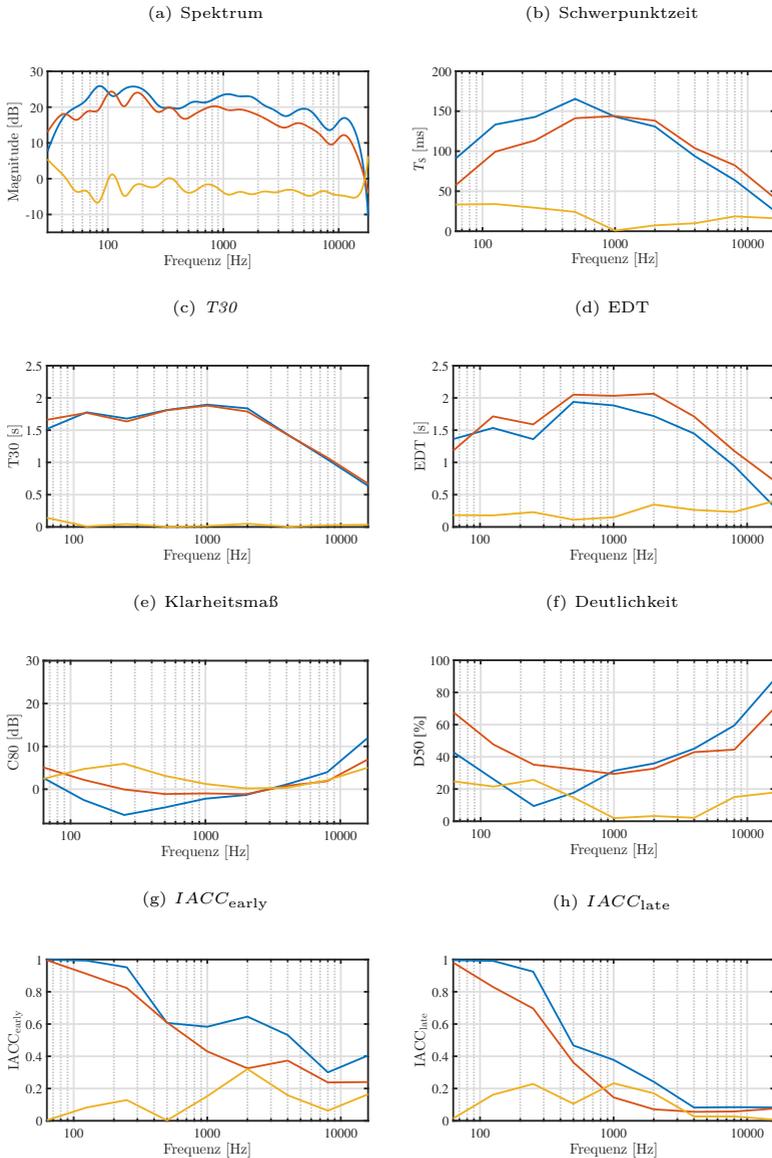


Abbildung A.31: Experiment LEC: Raumakustische Kriterien Synthese diffuse Anteile (KVB omnidirektional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

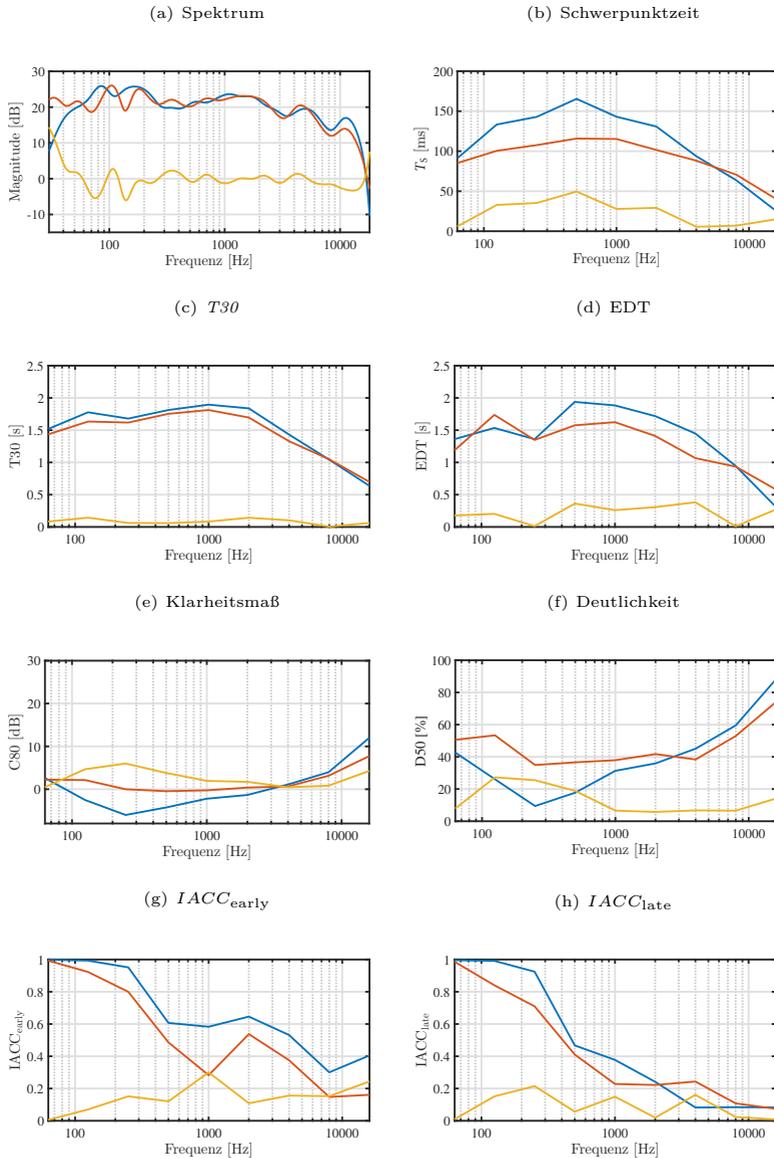


Abbildung A.32: Experiment LEC: Raumakustische Kriterien Synthese diffuse+gerichtete (KVB omnidirectional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

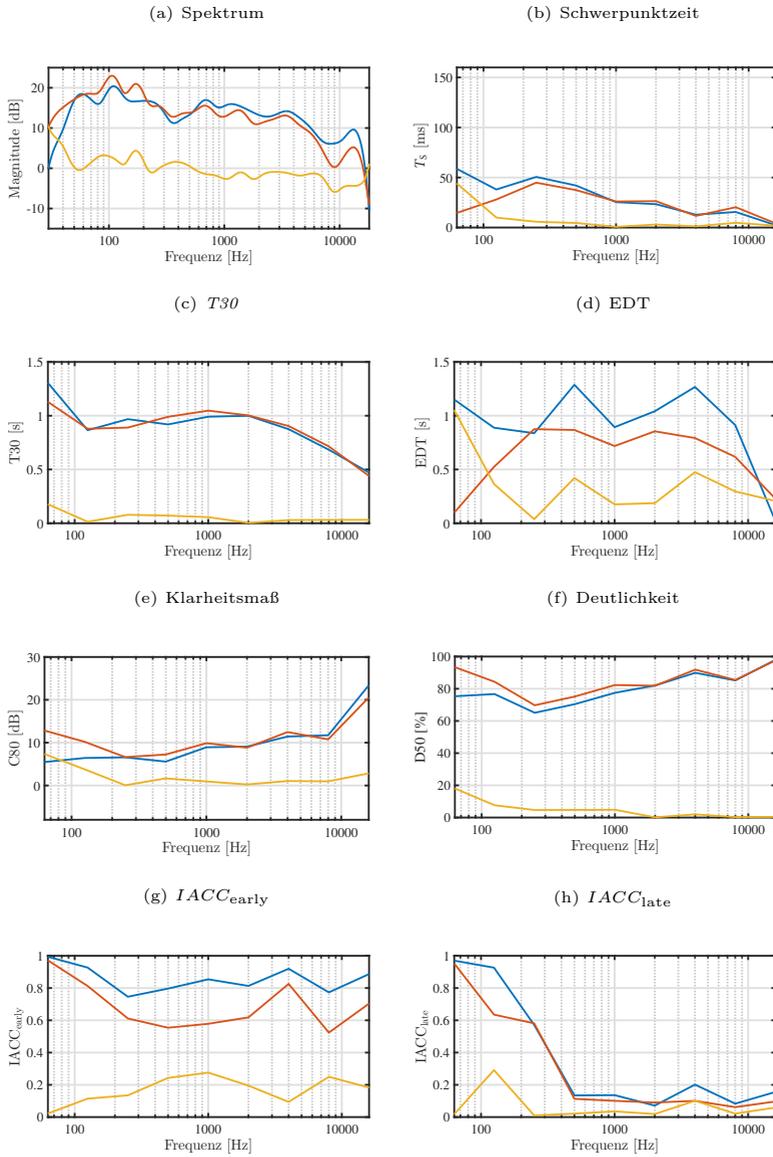


Abbildung A.33: Experiment LEC: Raumakustische Kriterien Synthese diffuse Anteile (KSS direktional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

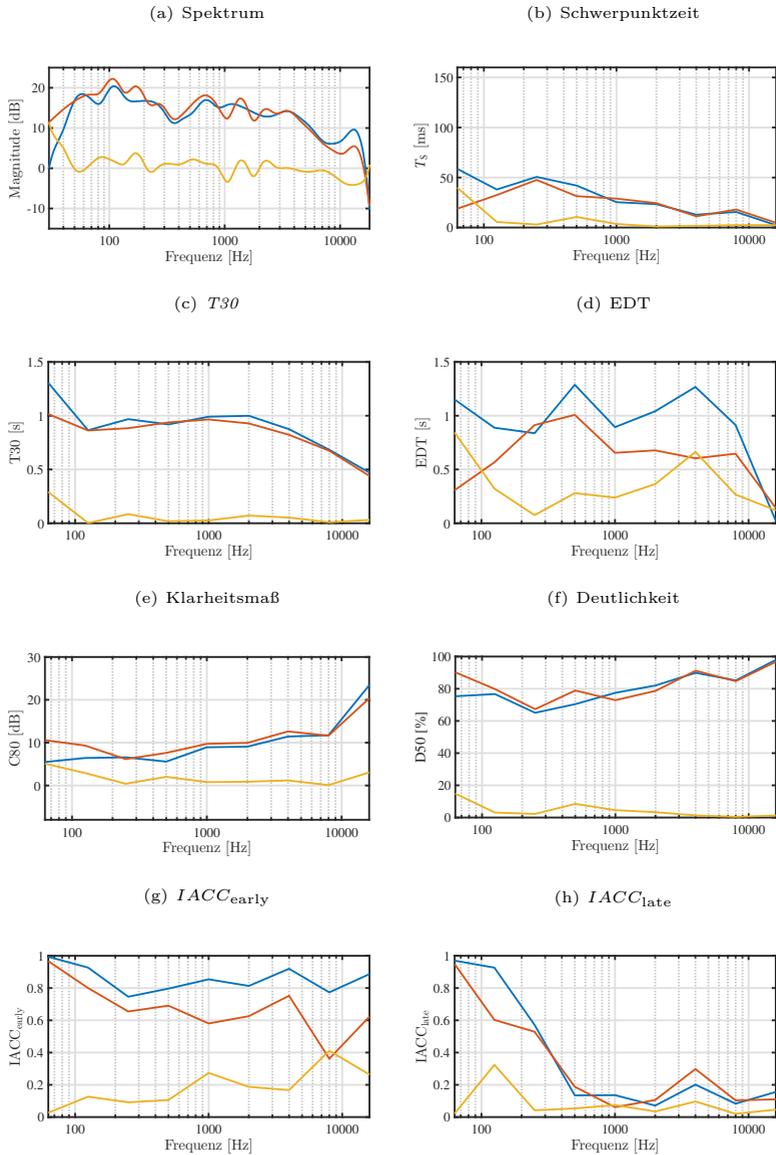


Abbildung A.34: Experiment LEC: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS direktional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

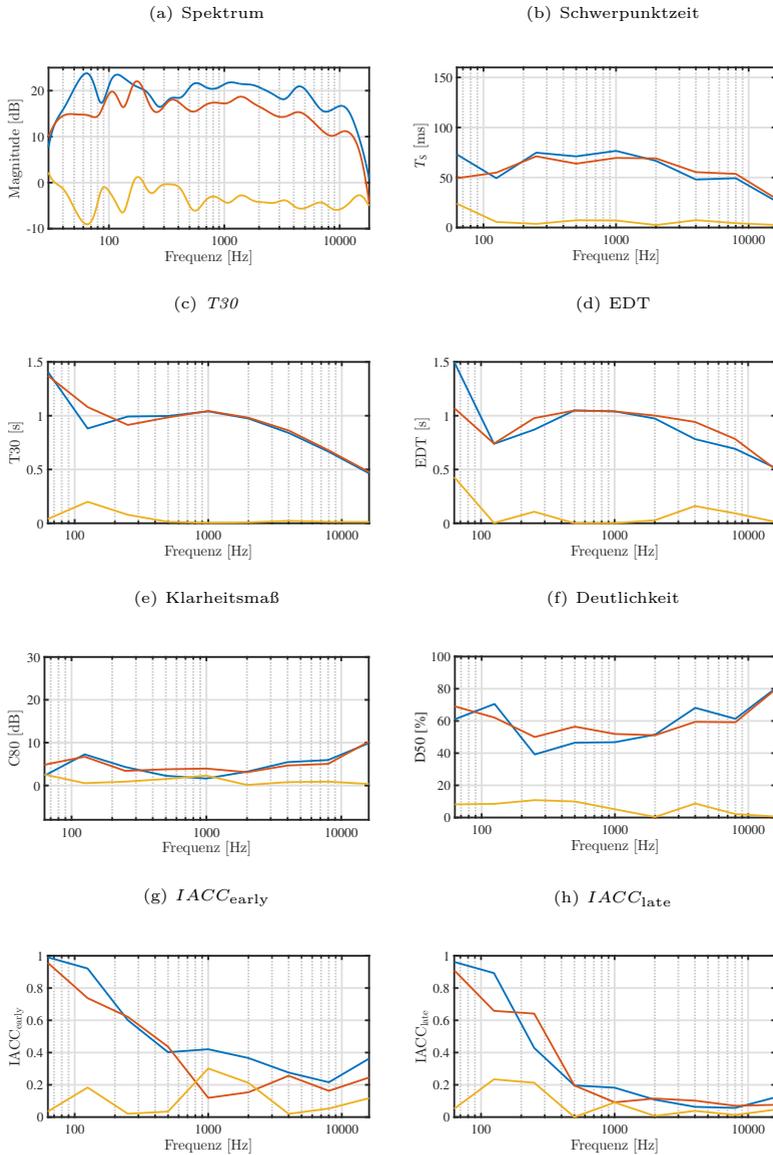


Abbildung A.35: Experiment LEC: Raumakustische Kriterien Synthese diffuse Anteile (KSS omnidirektional RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

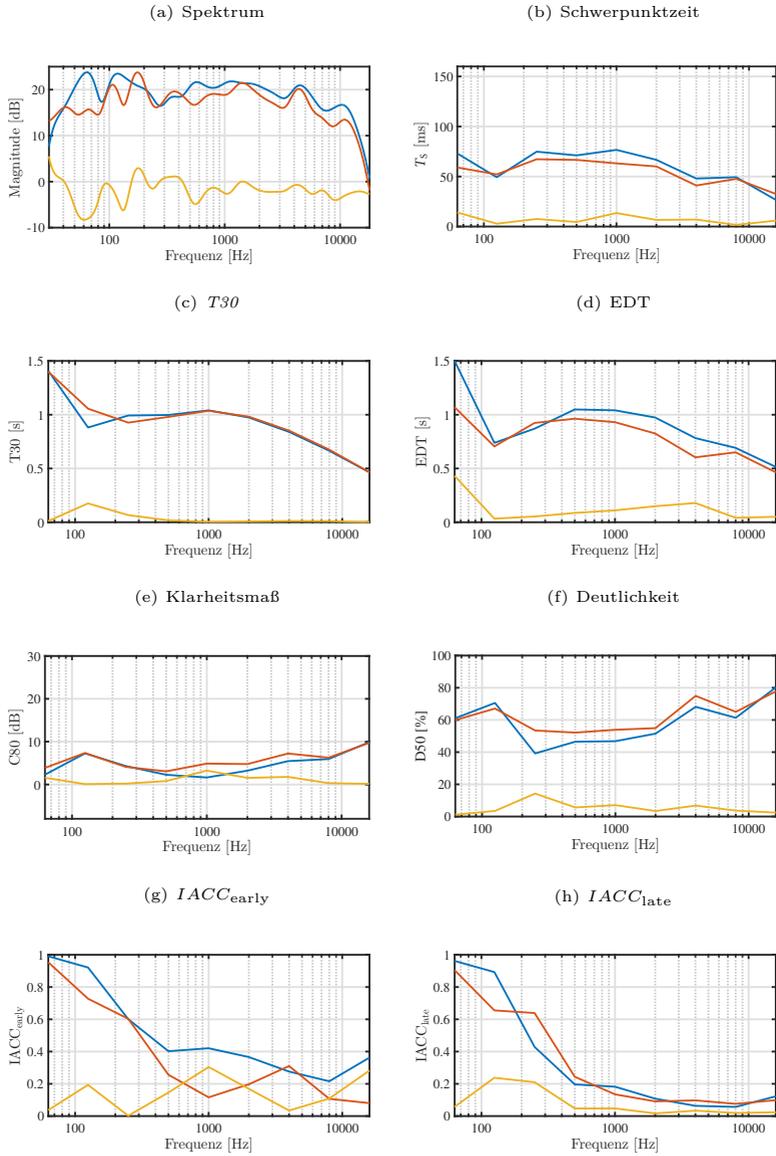


Abbildung A.36: Experiment LEC: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS omnidirektional RFL RVB, rot), Vergleich mit Referenz-BRIR (blau) und Differenzbetrag (gelb)

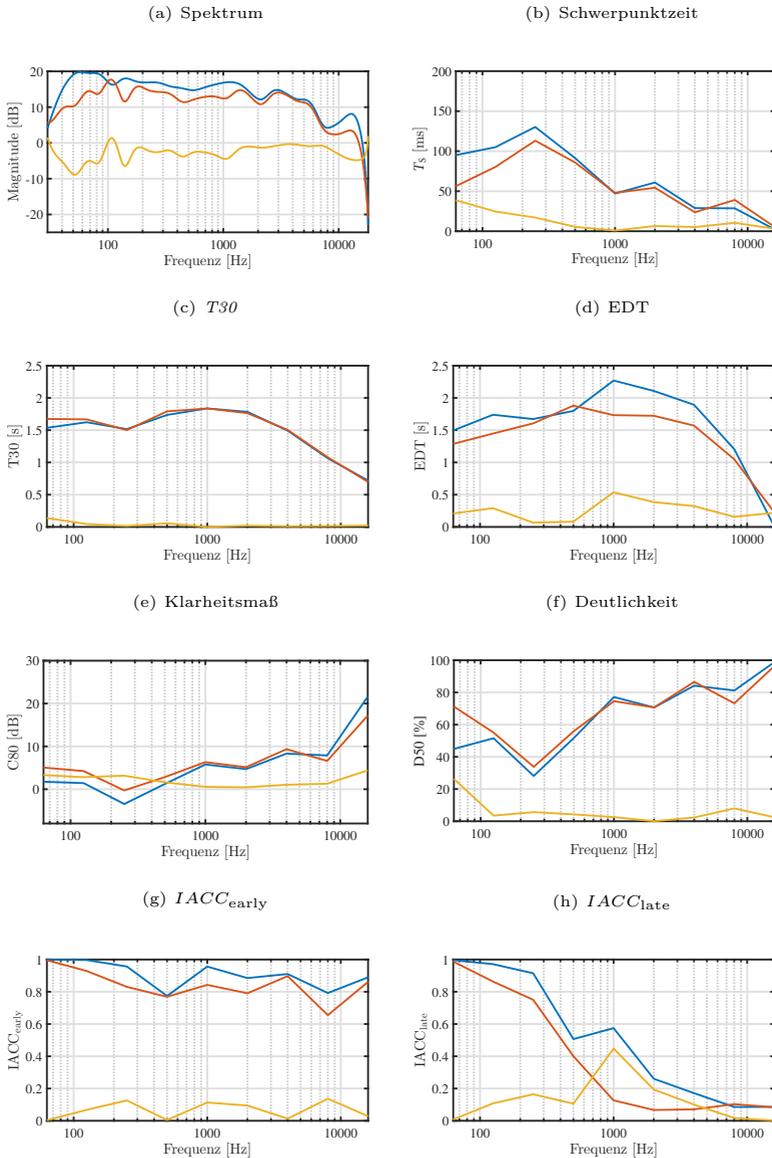


Abbildung A.37: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KVB direktional RVB), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

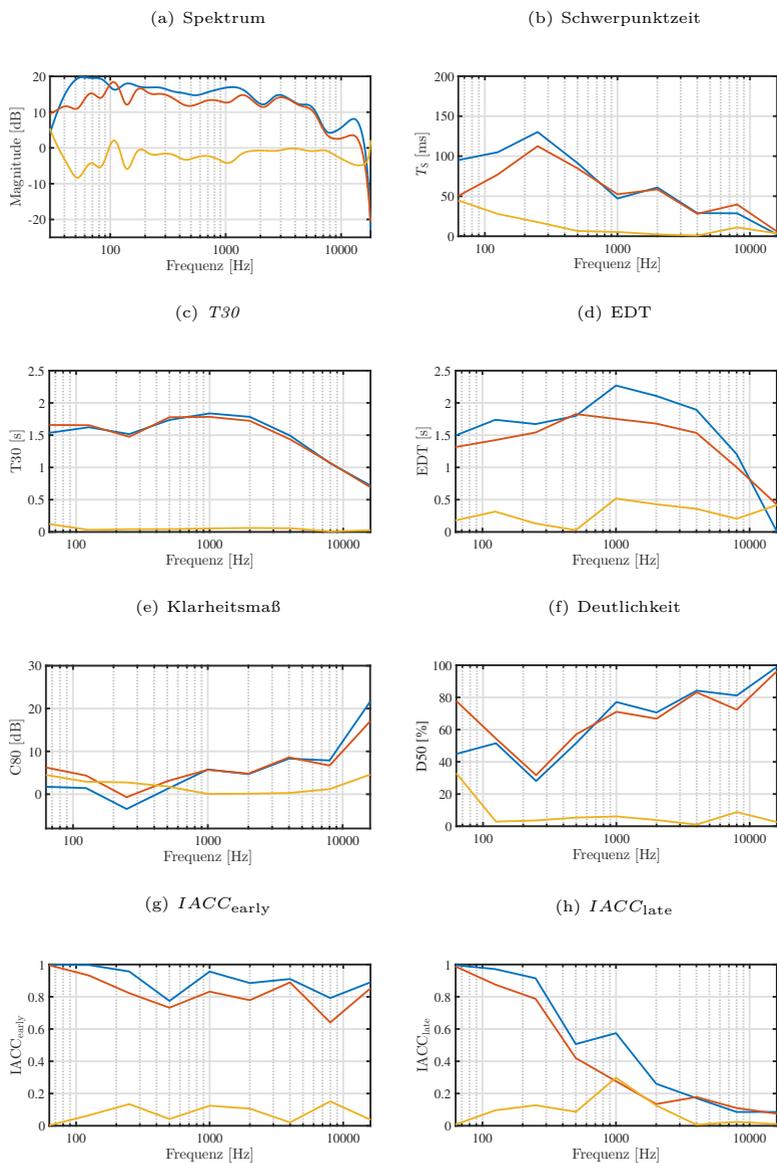


Abbildung A.38: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KVB *directional* RFL RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

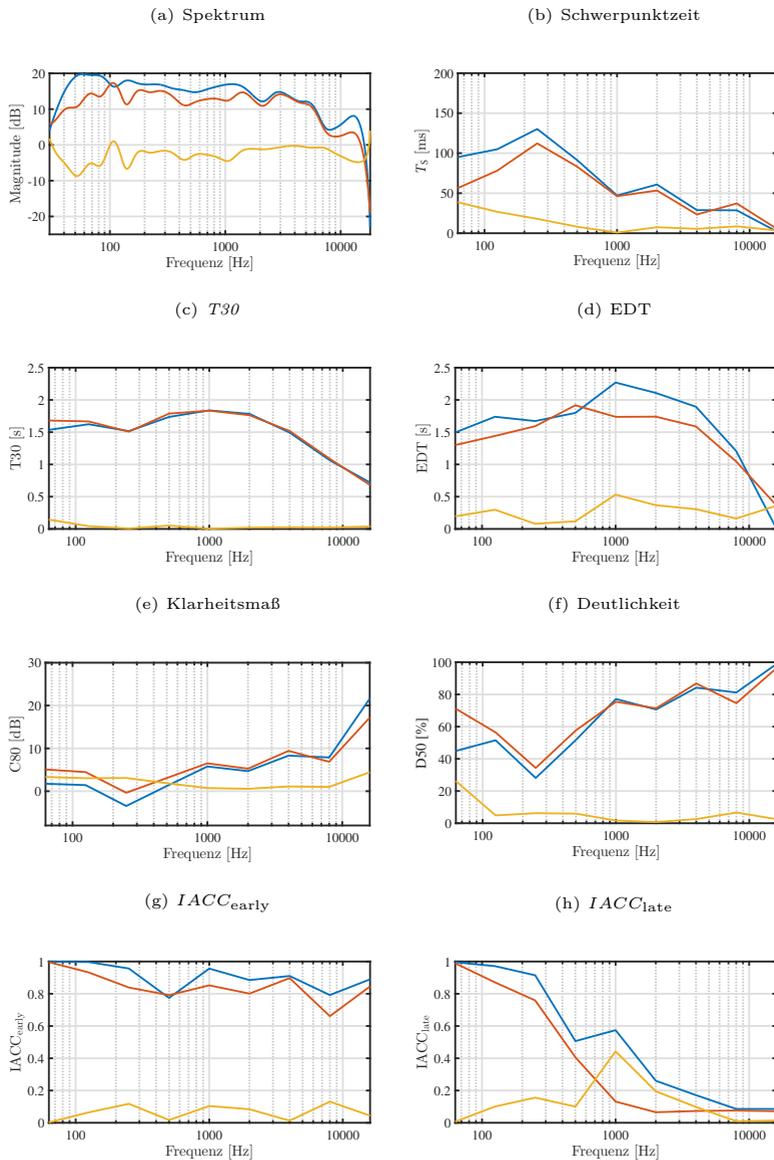


Abbildung A.39: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KVB *direktional* RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

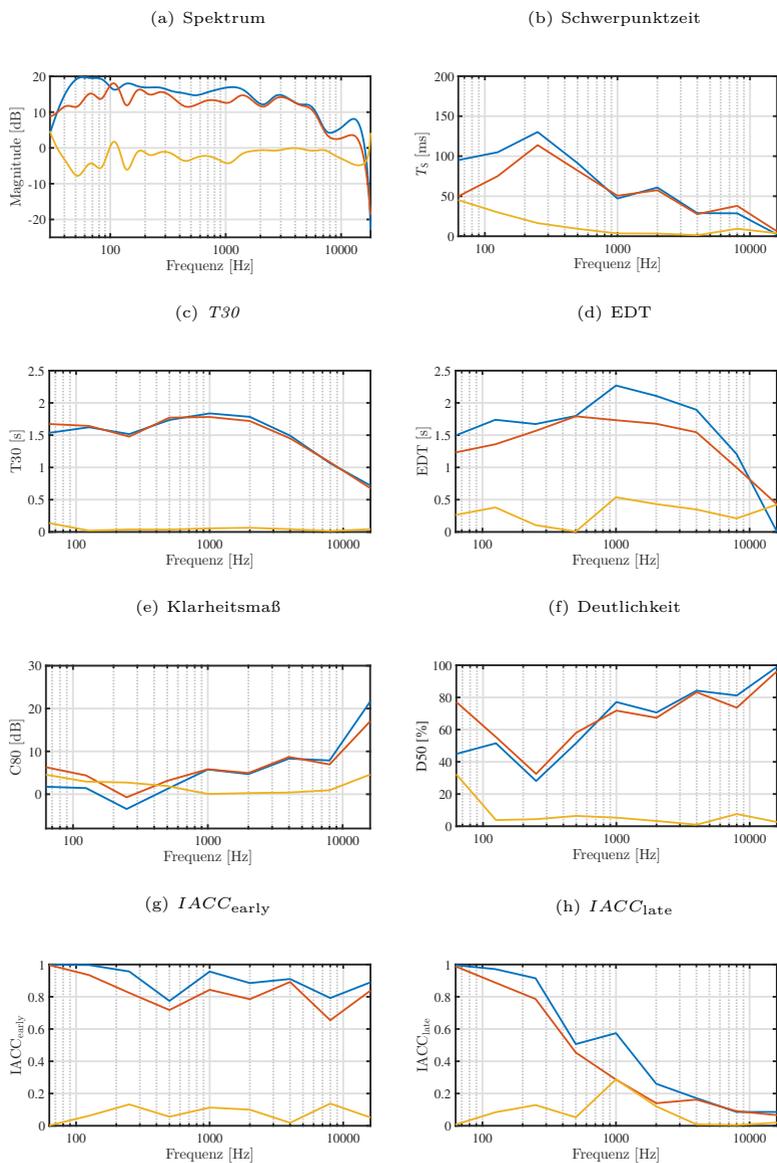


Abbildung A.40: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KVB direktional RFL RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

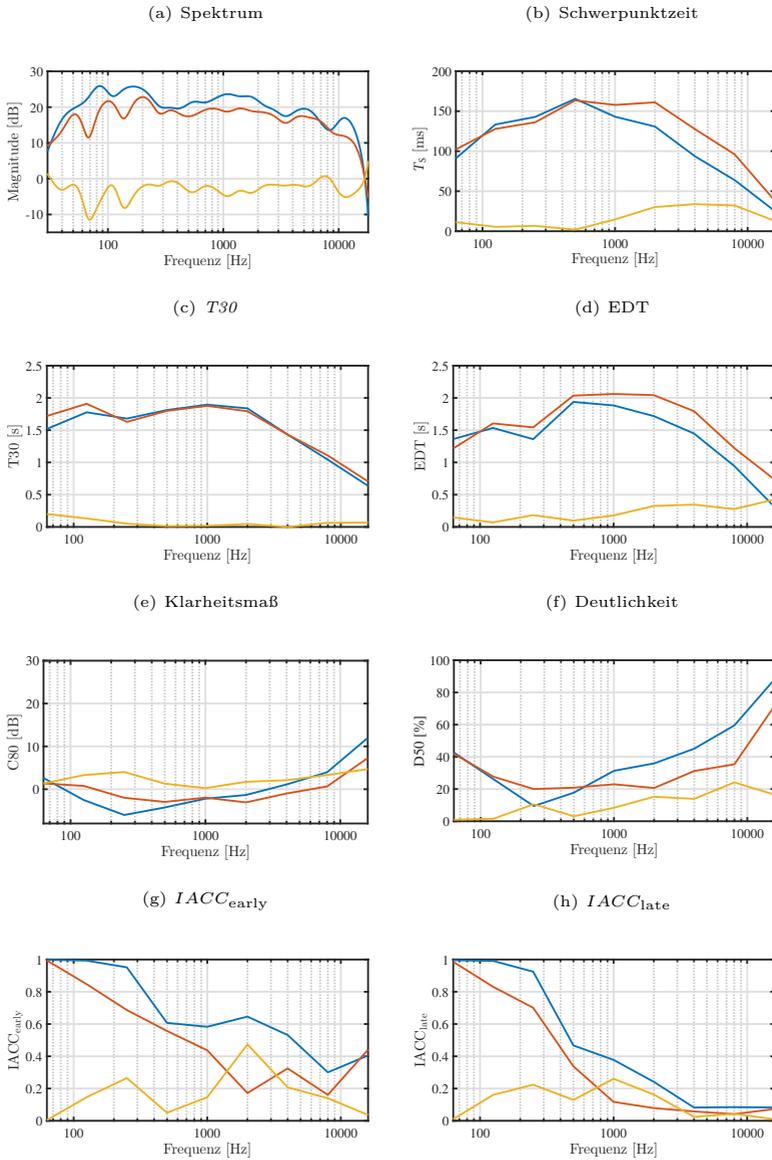


Abbildung A.41: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KVB omnidirektional RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

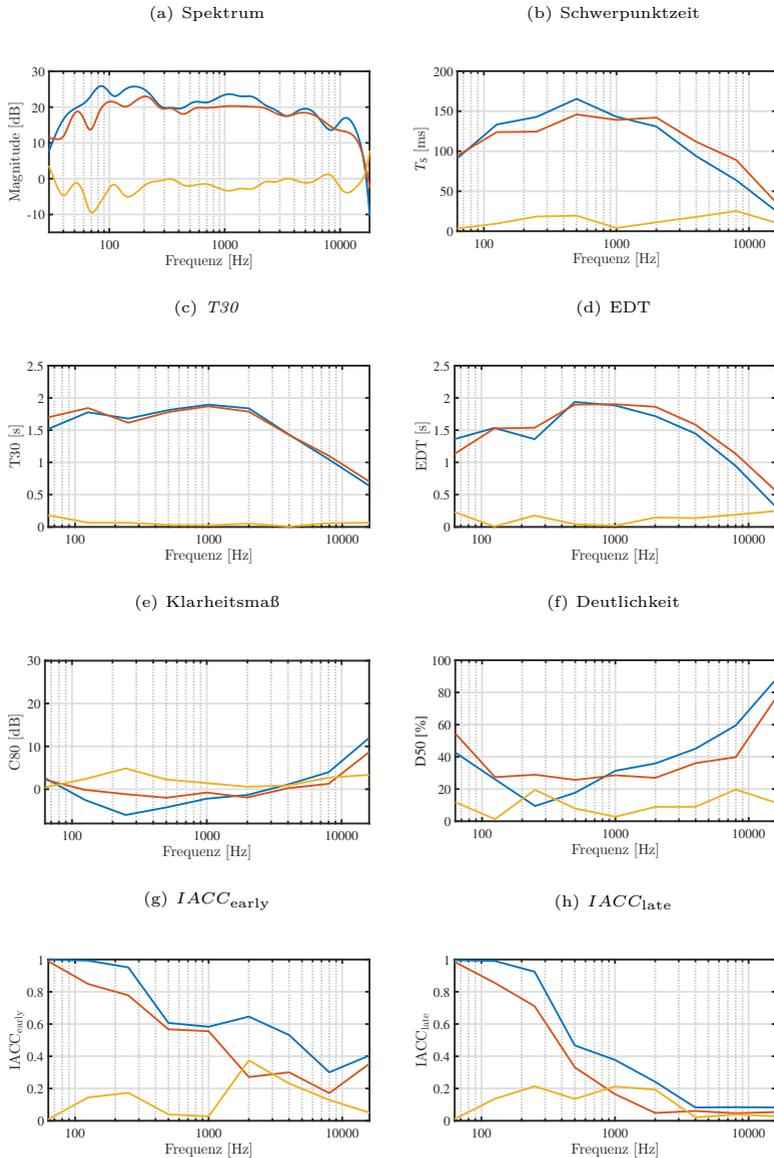


Abbildung A.42: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete (KVB omnidirektional RFL RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

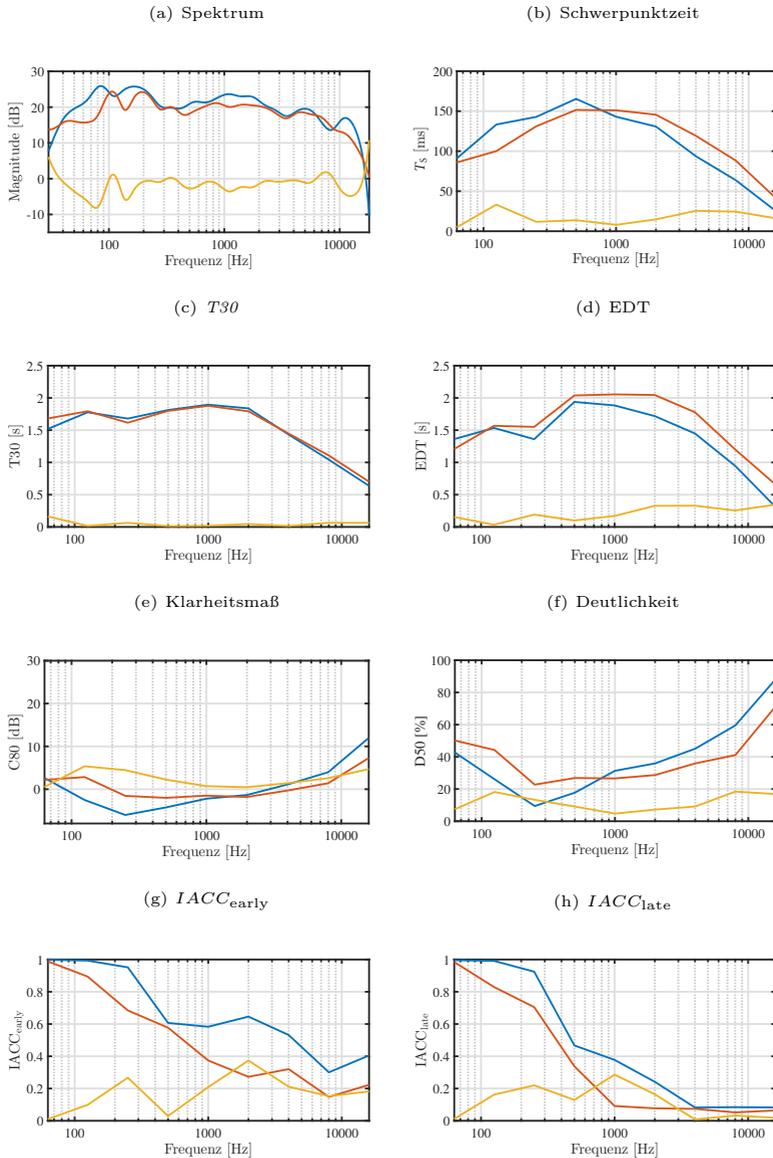


Abbildung A.43: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KVB omnidirektional RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

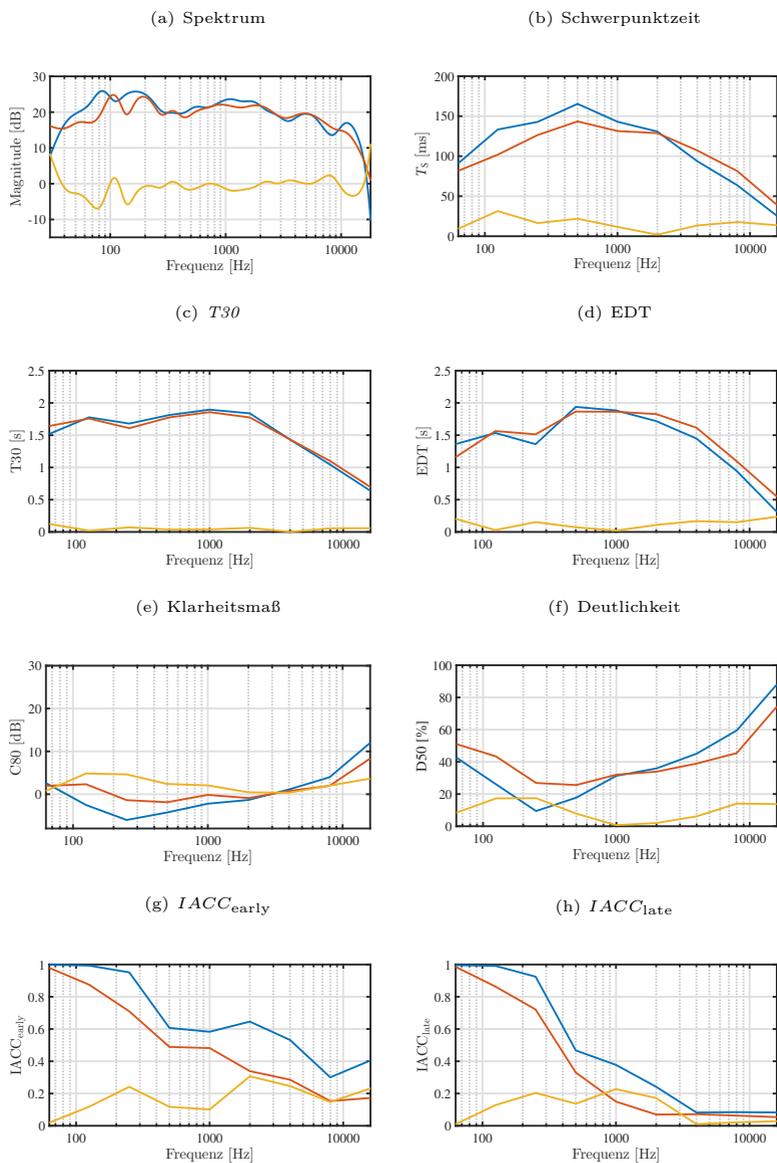


Abbildung A.44: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete (KVB omnidirektional RFL RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

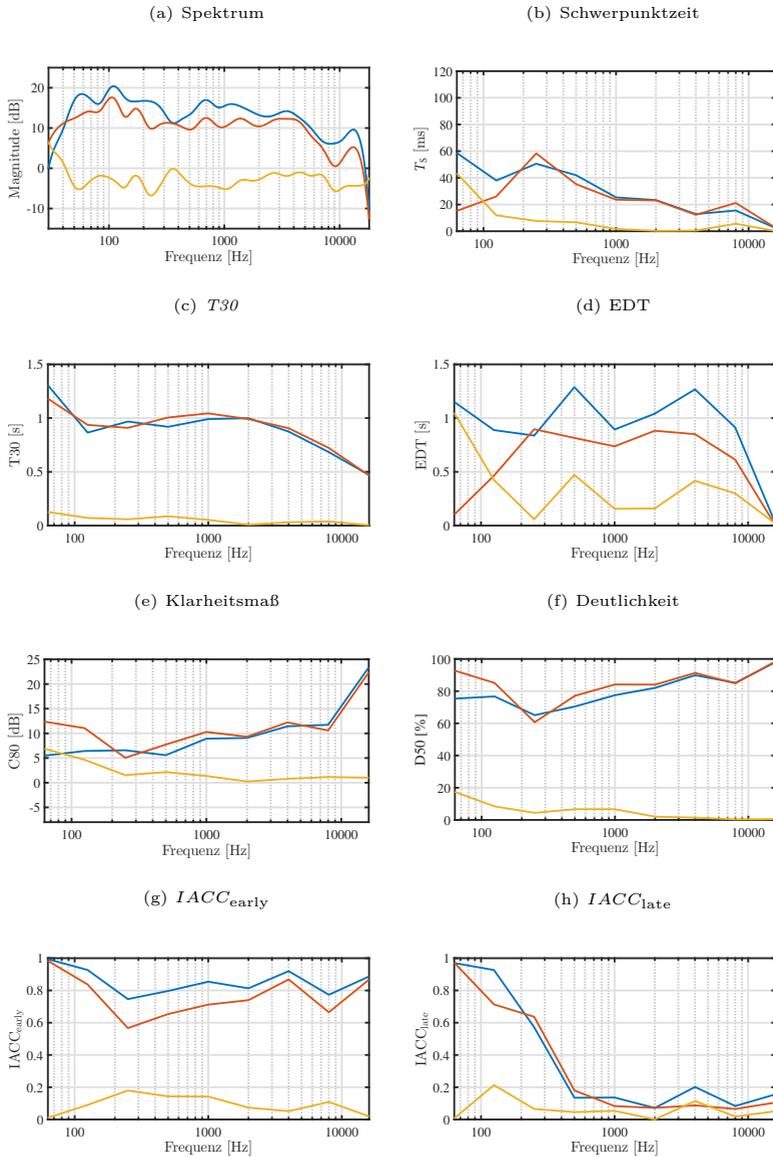


Abbildung A.45: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KSS direktional RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

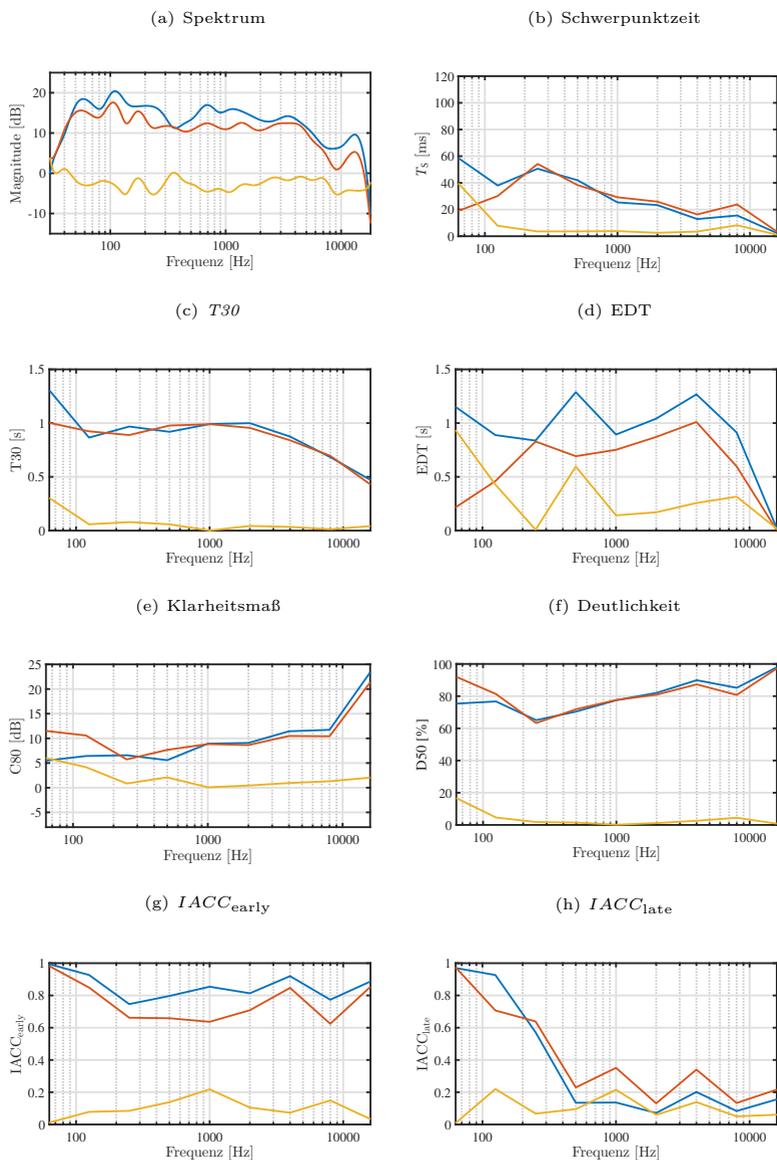


Abbildung A.46: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS direktional RFL RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

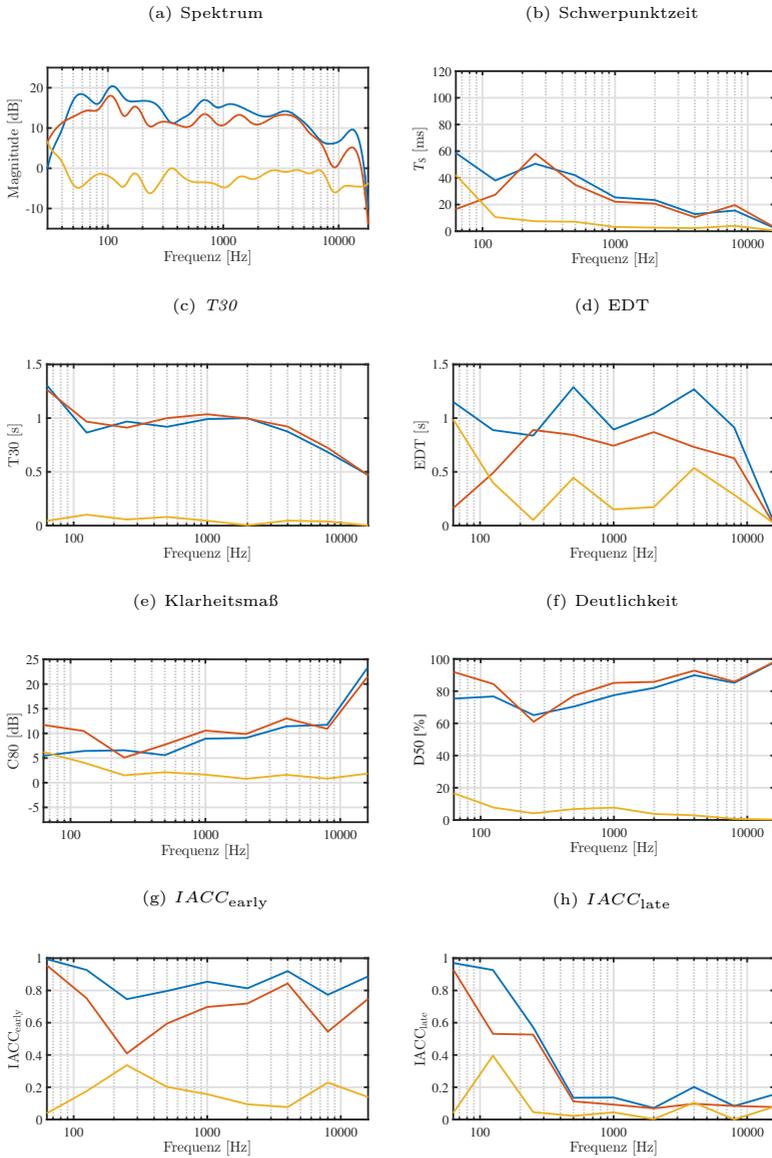


Abbildung A.47: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KSS direktional RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

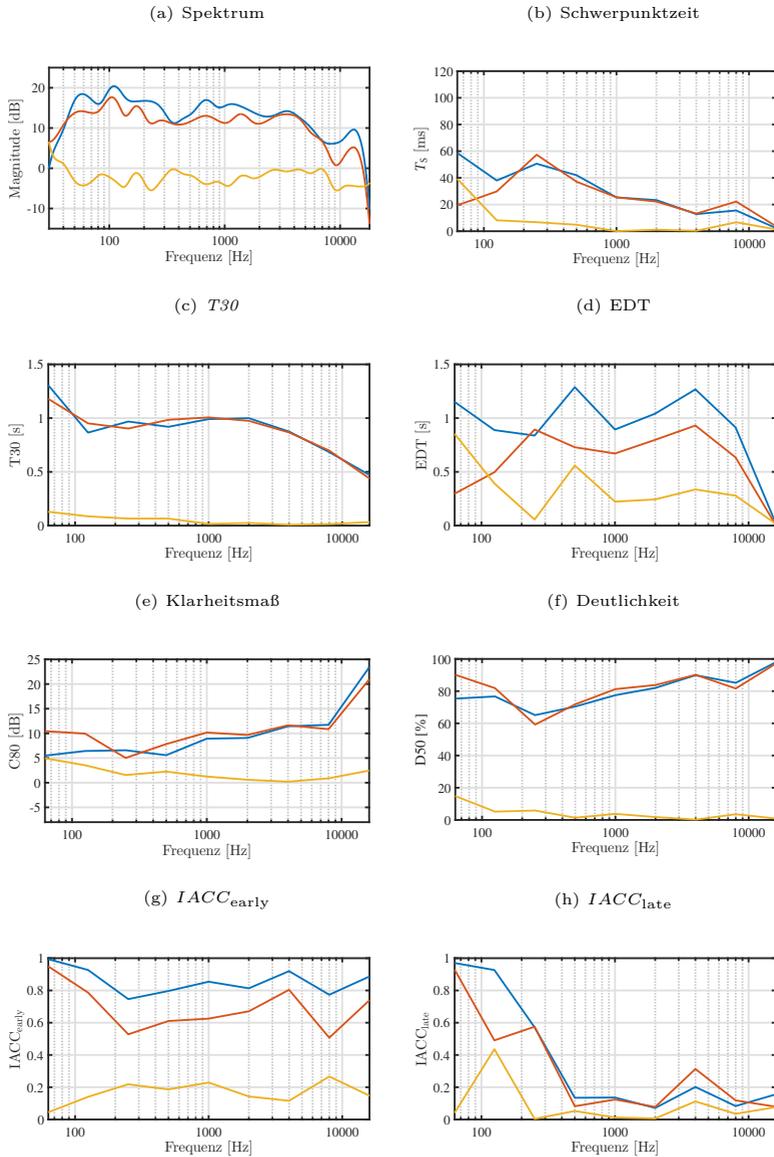


Abbildung A.48: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS direktional RFL RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

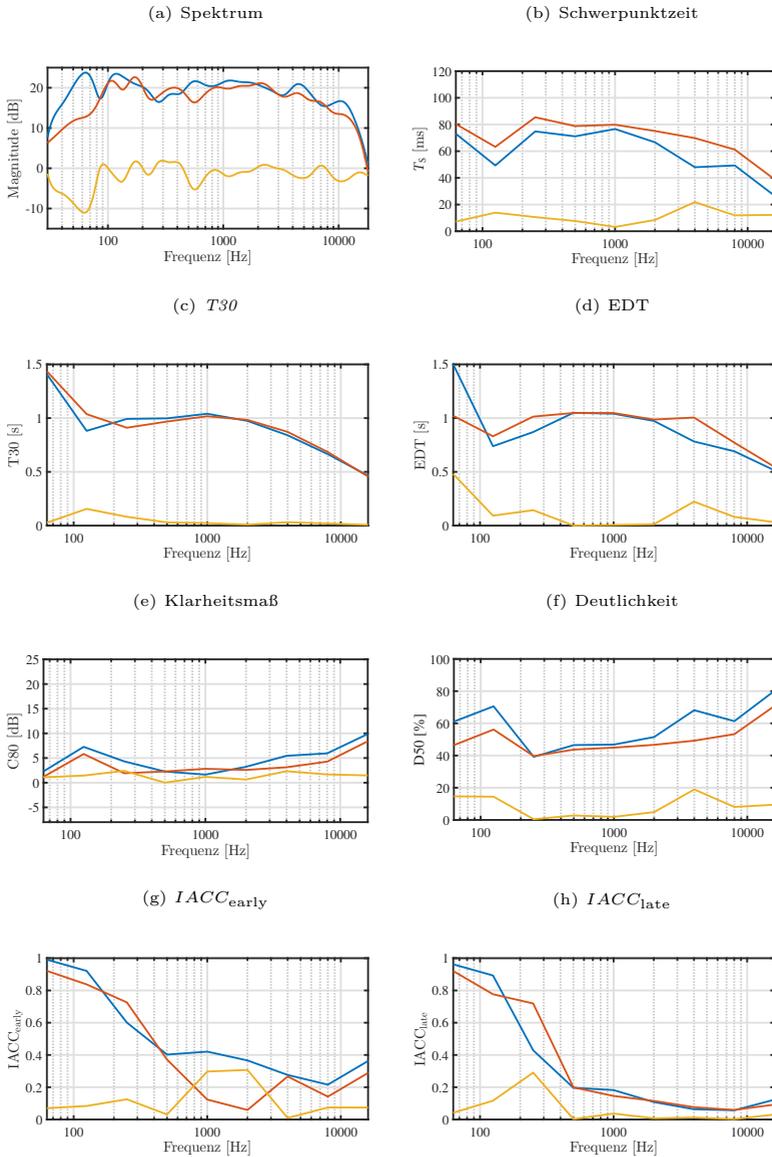


Abbildung A.49: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KSS omnidirektional RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

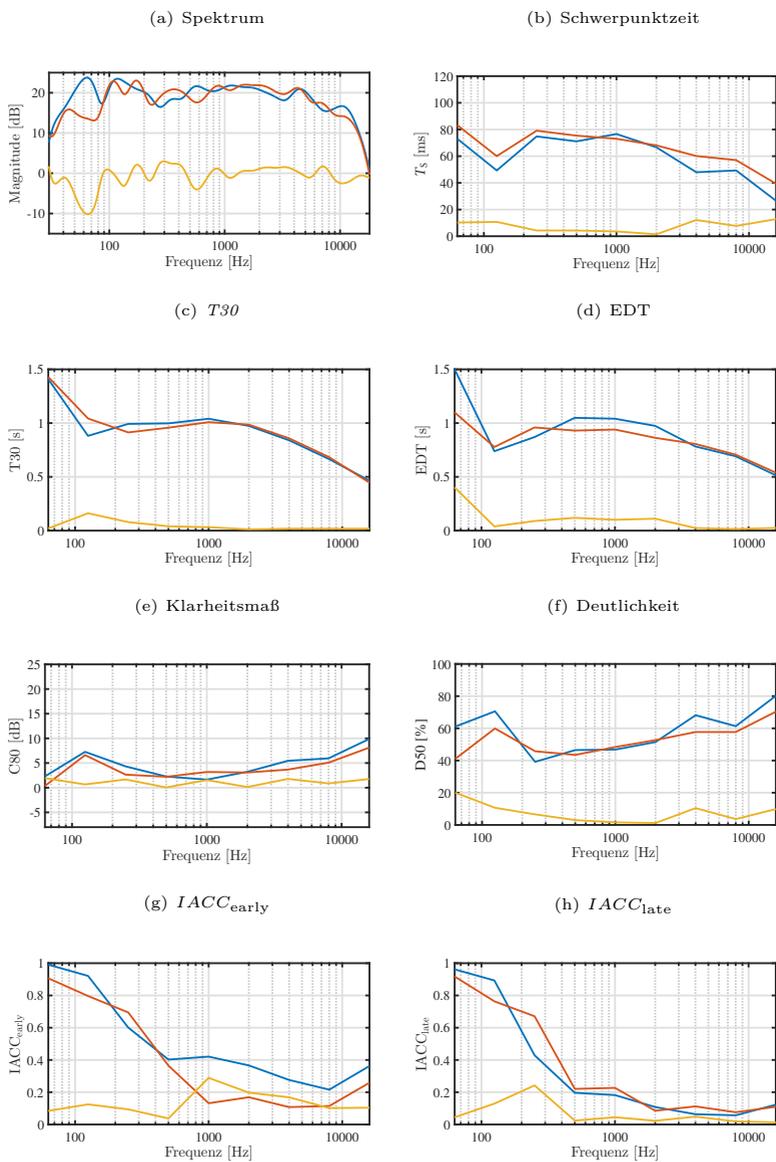


Abbildung A.50: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS omnidirektional RFL RVB, rot), Vergleich Array 86 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

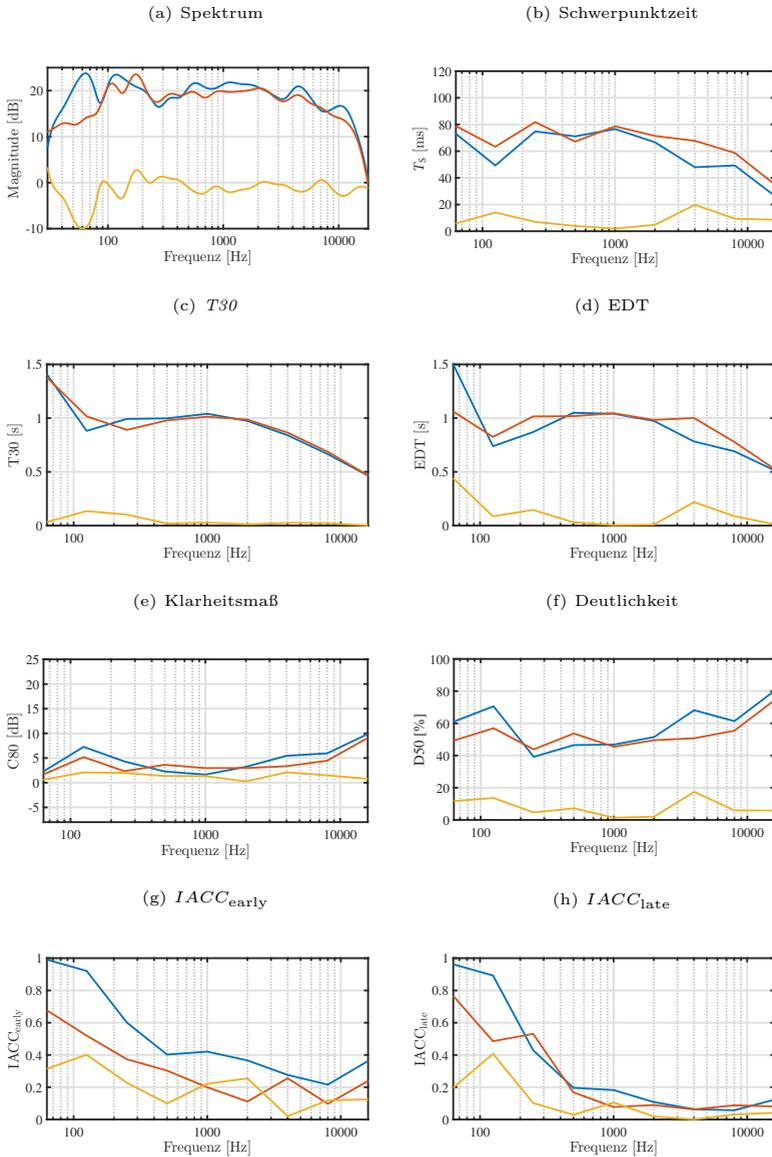


Abbildung A.51: Experiment LED: Raumakustische Kriterien Synthese diffuse Anteile (KSS omnidirektional RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

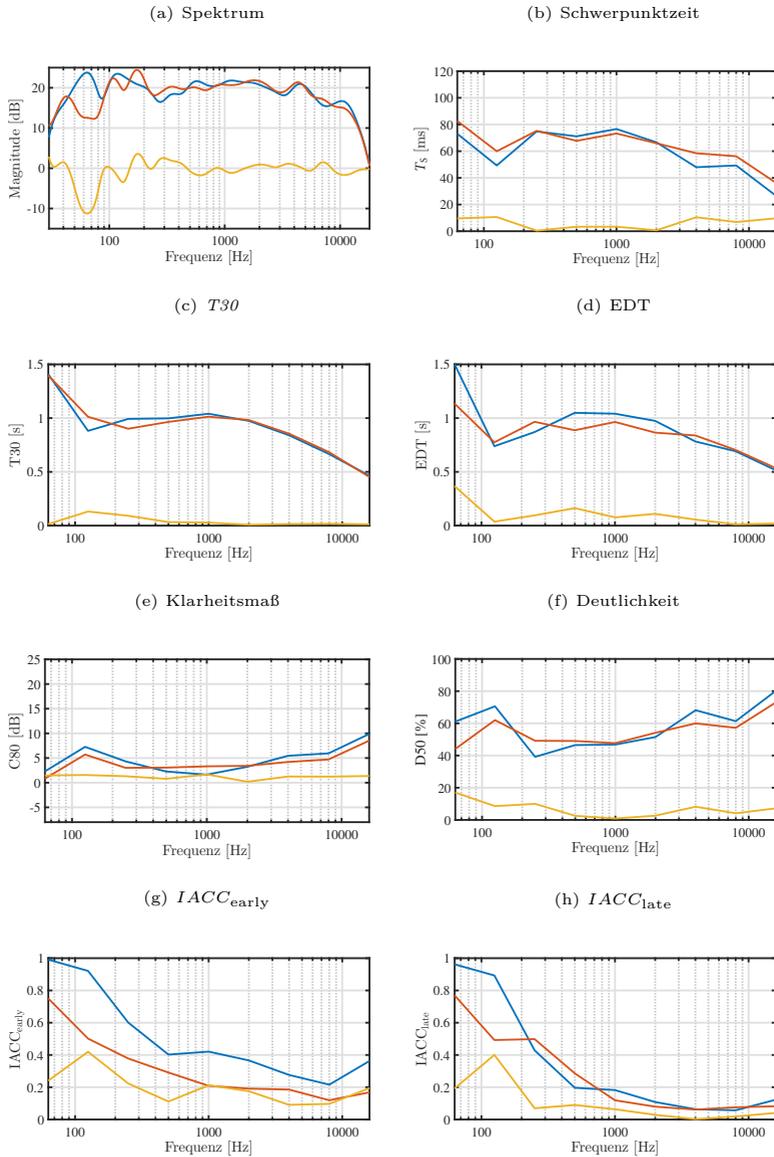


Abbildung A.52: Experiment LED: Raumakustische Kriterien Synthese diffuse+gerichtete Anteile (KSS omnidirektional RFL RVB, rot), Vergleich Array 1202 mit Referenz-BRIR (blau), Differenzbetrag (gelb)

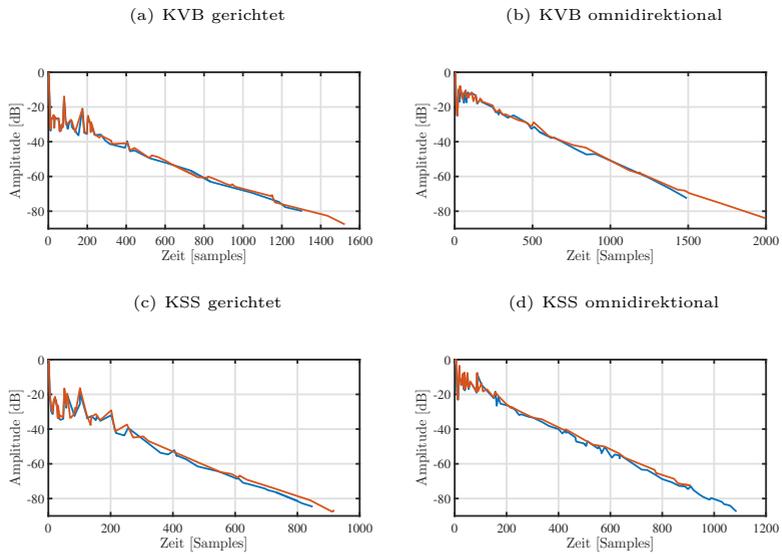


Abbildung A.53: Experiment LED: Vergleich der detektierten Reflexionen über Zeit und Amplitude mit 1202 (rot) und 86 (blau) Abtastpunkten.

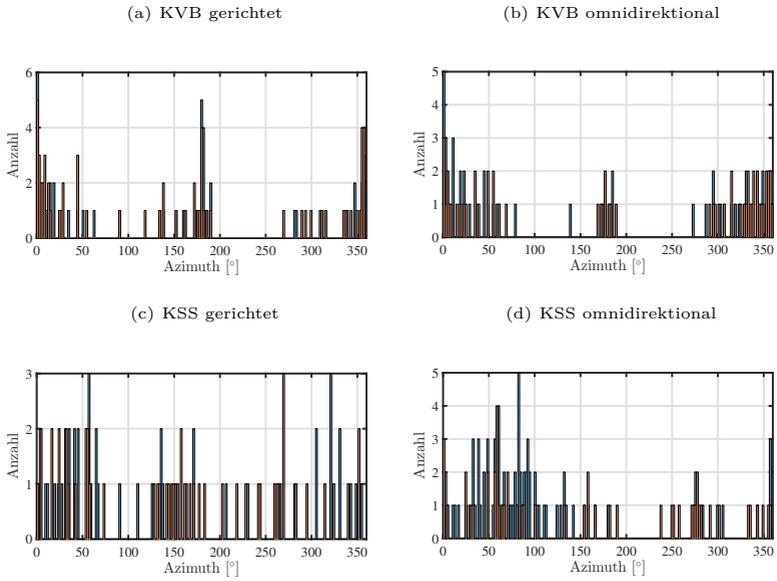


Abbildung A.54: Experiment LED: Histogramm Azimuth-Winkel der detektieren Reflexionen mit 1202 (rot) und 86 (blau) Abtastpunkten.

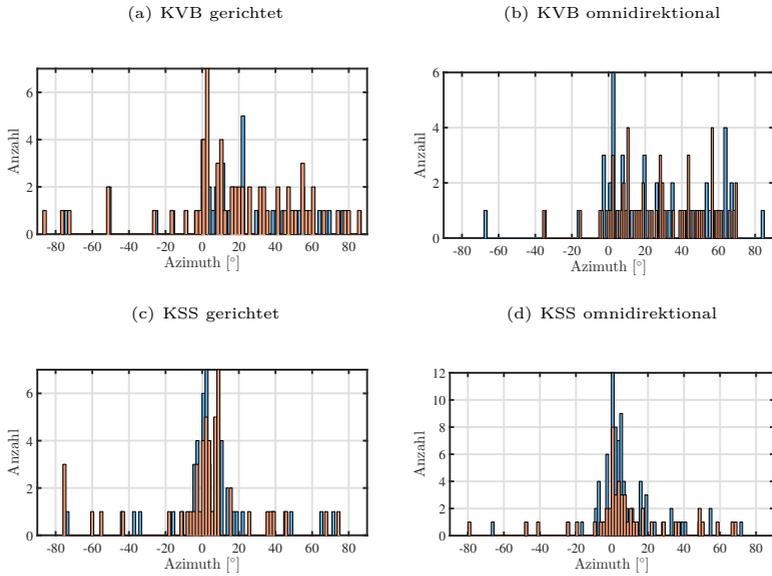


Abbildung A.55: Experiment LED: Histogramm Elevations-Winkel der detektieren Reflexionen mit 1202 (rot) und 86 (blau) Abtastpunkten.

ANHANG **B**

**Abkürzungen,
Bezeichnungen und
Variablen**

Abkürzungen

ABC/HR	double-blind triple-stimulus test with hidden reference / Hörversuch nach ITU 1116-3
AR	augmented reality / erweiterte Realität
ASW	apparent source width / wahrgenommene Quellbreite
BR	bass ratio / Bassverhältnis
BRIR	binaural room impulse response / binaurale Raumimpulsantwort
C80	Klarheitsmaß in dB
CQS	continuous quality scale/ kontinuierliche Qualitäts-Skala
D50	Deutlichkeit in %
DBS	dynamische Binauralsynthese
DIR	directional impulse response / gerichtete Impulsantwort
DOA	direction of arrival / Einfallsrichtung
DRR	direct to reverberant ratio / Verhältnis zwischen Direkt- und Diffusschall
EDC	energy decay curve / Energie-Abklingkurve
EDT	early decay time / frühe Abklingzeit
FFT	fast Fourier transform / schnelle Fourier-Transformation
FIR	finite impulse response / endliche Impulsantwort
G	strength / Stärkemaß in dB
GUI	graphical user interface / grafische Benutzeroberfläche
HRIR	head related impulse response / Außenohrimpulsantwort
HRTF	head related transfer function / Außenohrübertragungsfunktion
IACC	interaural cross-correlation / interaurale Kreuzkorrelation
IACF	interaural cross-correlation function / interaurale Kreuzkorrelationsfunktion
IC	interaural coherence / interaurale Kohärenz
IFFT	inverse fast Fourier transform / inverse schnelle Fourier-Transformation
ILD	interaural level difference / interauraler Pegelunterschied
ITD	interaural time difference / interauraler Laufzeitunterschied
ITDG	initial time delay gap / Anfangszeitlücke
JND	just noticeable difference / differentielle Wahrnehmbarkeitsschwelle
KSS	Kleiner Sendesaal
KVB	Klaus-Von-Bismarck-Saal
LEV	listener envelopment / Umhüllung
MUSHRA	Multi-Stimulus Test with Hidden Reference and Anchor
PWD	plane-wave decomposition / ebene Wellenzerlegung
RIR	room impulse response / Raumimpulsantwort
rmANOVA	repeated measures analysis of variance / Varianzanalyse mit Messwiederholung
SBRIR	synthetic binaural room impulse response / synthetische binaurale Raumimpulsantwort
SDM	spatial decomposition method

SIRR	spatial impulse response rendering
SNR	signal-to-noise-ratio / Signal-Rausch-Abstand
SFT	spatial Fourier transform / räumliche Fourier-Transformation
SSR	SoundScape Renderer
STFT	short-time Fourier transform / Kurzzeit-Fourier-Transformation
T	Nachhallzeit in s
TDOA	time difference of arrival / Differenz der Ankunftszeit
TOA	time of arrival / Ankunftszeit
TR	treble ratio / Höhenverhältnis
T_s	Schwerpunktzeit in ms
VAE	virtual acoustic environment / virtuelle akustische Umgebung
VR	virtual reality / virtuelle Realität
WDR	Westdeutscher Rundfunk
WFA	wave field analysis / Wellenfeldanalyse
WFS	wave field synthesis / Wellenfeldsynthese

Bezeichnungen Parameter, Modelle und Experimente

SBRIR	Modell zur Synthese von BRIRs
SBRIR _{comp}	Modell zur Synthese von BRIRs mit reduzierter Auflösung
SBRIR _{early}	Modell zur Synthese von der frühen Anteile von BRIRs
SBRIR _{late}	Modell zur Synthese von der späten Anteile von BRIRs
$p_{\mathbf{k}}(n)$	gerichtete Komponenten mit Anzahl n
$Az_{p_{\mathbf{k}}(n)}$	gerichteter Parameter Azimuth-Winkel
$El_{p_{\mathbf{k}}(n)}$	gerichteter Parameter Elevations-Winkel
$dB_{p_{\mathbf{k}}(n)}$	gerichteter Parameter Pegel
$t_{p_{\mathbf{k}}(n)}$	gerichteter Parameter Zeitpunkt
$FIR_{p_{\mathbf{k}}(n)}$	gerichteter Parameter spektrale Färbung
EDC(f)	diffuser Parameter der Abklingkurven
IC(f)	diffuser Parameter der interauralen Kohärenz
mE(f)	diffuser Parameter der mittleren Energie
LEA1	perzeptive Untersuchung Synthese später Anteile, Vergleich Algorithmen
LEA2	perzeptive Untersuchung Synthese später Anteile, Vergleich Räume
LEB	perzeptive Untersuchung Synthese früher Anteile
LEC	perzeptive Untersuchung Synthese Gesamtmodell
LED	perzeptive Untersuchung Synthese Gesamtmodell mit reduzierter Auflösung
MT1	perzeptive Untersuchung Übergangszeit
MT2	perzeptive Untersuchung Übergangszeit ohne Direktschall

Variablen und mathematische Symbole

c	Schallgeschwindigkeit in m/s
d	Ohrabstand
d_n	Radialfilter
η_g	Effizienz Abtastquadratur
η_p^2	Effektgröße partial Eta-squared
f	Frequenz
f_A	Frequenz räumliches Aliasing in Hz
f_{pwd}	Frequenz zur Zerlegung in ebene Wellen
f_s	Abtastrate
$\mathcal{F}_t\{\cdot\}$	Zeit-Frequenz Fourier-Transformation
$\mathcal{F}_t^{-1}\{\cdot\}$	inverse Zeit-Frequenz Fourier-Transformation
$\tilde{\mathcal{F}}_x\{\cdot\}$	räumliche Fourier-Transformation in kartesischen Koordinaten
$\tilde{\mathcal{F}}_x^{-1}\{\cdot\}$	inverse räumliche Fourier-Transformation in kartesischen Koordinaten
$\hat{\mathcal{F}}_s\{\cdot\}$	räumliche Fourier-Transformation in sphärischen Koordinaten
$\hat{\mathcal{F}}_s^{-1}\{\cdot\}$	inverse räumliche Fourier-Transformation in sphärischen Koordinaten
$H_l(f, r, \phi, \theta)$	HRTF linkes Ohr
$H_r(f, r, \phi, \theta)$	HRTF rechtes Ohr
λ	Wellenlänge
l_m	freie Weglänge
N	Ordnung
n_{fb}	Auflösung der Filterbank in Frequenzbändern
n_{sp}	Anzahl der Abtastpunkte Mikrofonarray
n_{ts}	Anzahl der Zeitscheiben
Φ	interaurale Kohärenz
ϕ	Azimuth-Winkel in rad (Formeln) oder $^\circ$ (Abbildungen/Erläuterungen)
p_n	Polynomgrad
$P_n^m(\cdot)$	assoziierte Legendre-Funktionen
q_{sp}	Faktor der verwendeten Abtastquadratur des Mikrofonarrays
r_0	Radius Mikrofonarray
S	Oberfläche in m^2
S_l	Fouriertransformierte des linken Ohrsignals
S_r	Fouriertransformierte des rechten Ohrsignals
t	Zeit in s
t_{FIR}	Filterlänge
θ	Elevations-Winkel in rad (Formeln) oder $^\circ$ (Abbildungen/Erläuterungen)
t_{mp}	perzeptive mixing time in ms
t_m	mixing time in ms
t_{mm}	Übergangszeit Modell in ms

$t_{\text{mm early}}$	Übergangszeit Modell früh in ms
$t_{\text{mm late}}$	Übergangszeit Modell spät in ms
t_w	Fensterlänge
V	Volumen in m^3
ω	Kreisfrequenz
Y_n^m	Kugelflächenfunktionen

Literaturverzeichnis

- Abel, Jonathan S und P Huang (2006): „A simple, robust measure of reverberation echo density.” In: *Proc. of the 121st AES Convention*. San Francisco.
- Ahrens, Jens (2014): „Challenges in the Creation of Artificial Reverberation for Sound Field Synthesis: Early Reflections and Room Modes.” In: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*. Berlin.
- Ahrens, Jens (2015): „Perceptual Evaluation of the Diffuseness of Synthetic Late Reverberation Created by Wave Field Synthesis at Different Listening Positions.” In: *Fortschritte der Akustik - DAGA 2015*. Nürnberg, S. 1634–1637.
- Ahrens, Jens (2016): „On the Generation of Virtual Early Reflections in Wave Field Synthesis.” In: *Fortschritte der Akustik - DAGA 2016*. Aachen, S. 1315–1318.
- Ahrens, Jens; Christoph Hohnerlein und Carl Andersson (2017): „Auralization of acoustic spaces based on spherical microphone array recordings.” In: *Journal of the Acoustical Society of America*, 141(5) S. 3512–3512.
- Algazi, V Ralph; Richard O Duda; D M Thompson und Carlos Avendano (2001): „The CIPIC HRTF database.” In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. New York, S. 99–102.
- Allen, Jont B und David A Berkley (1979): „Image method for efficiently simulating small-room acoustics.” In: *Journal of the Acoustical Society of America*, 65(4) S. 943–950.
- Bakeman, Roger (2005): „Recommended effect size statistics for repeated measures designs.” In: *Behavior Research Methods*, 37(3) S. 379–384.
- Barron, Michael F E (1971): „The subjective effects of first reflections in concert halls - The need for lateral reflections.” In: *Journal of Sound and Vibration*, 15(4) S. 475–494.
- Bech, Søren (1996a): „Audibility of individual reflections in a complete sound field, III.” In: *Proc. of the 140th AES Convention*. Copenhagen.
- Bech, Søren (1996b): „Timbral aspects of reproduced sound in small rooms. II.” In: *Journal of the Acoustical Society of America*, 99(6) S. 3539–3549.
- Bech, Søren (1998): „Spatial aspects of reproduced sound in small rooms.” In: *Journal of the Acoustical Society of America*, 103(1) S. 434–445.
- Begault, Durand R (1991): „Challenges to the Successful Implementation of 3-D Sound.” In: *Journal of the Acoustical Society of America*, 39(11) S. 864–870.
- Begault, Durand R (1994): *3-D Sound for Virtual Reality and Multimedia*. Cambridge: AP Professional.

- Begault, Durand R; Brian U McClain und Mark R Anderson (2001): „Early Reflection Thresholds For Virtual Sound Sources.” In: *Proc. of the 2001 International Workshop on Spatial Media*. Aizu-Wakamatsu.
- Ben-Hur, Zamir; Fabian Brinkmann; Jonathan Sheaffer; Stefan Weinzierl und Boaz Rafaely (2017): „Spectral equalization in binaural signals represented by order-truncated spherical harmonics.” In: *Journal of the Acoustical Society of America*, 141(6) S. 4087–4096.
- Bendat, Julius S und Allan G Piersol (1971): *Random Data: Analysis and Measurement Procedures*. New York: John Wiley & Sons.
- Beranek, Leo (2004): *Concert Halls and Opera Houses*. Music, Acoustics, and Architecture. Springer Science & Business Media.
- Berkhout, A J; Diemer de Vries und J J Sonke (1997): „Array technology for acoustic wave field analysis in enclosures.” In: *Journal of the Acoustical Society of America*, 102(5) S. 2757–2770.
- Bernschütz, Benjamin (2012): „Bandwidth Extension for Microphone Arrays.” In: *133rd AES Convention*. San Francisco.
- Bernschütz, Benjamin (2013): „A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100.” In: *Fortschritte der Akustik - AIA DAGA 2013*. Meran, S. 592–595.
- Bernschütz, Benjamin (2016): *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*. Dissertation.
- Bernschütz, Benjamin; Christoph Pörschmann; Sascha Spors und Stefan Weinzierl (2010): „Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio.” In: *Fortschritte der Akustik - DAGA 2010*. Berlin, S. 717–718.
- Bernschütz, Benjamin; Christoph Pörschmann; Sascha Spors und Stefan Weinzierl (2011a): „SOFiA sound field analysis toolbox.” In: *Proc. of the International Conference on Spatial Audio - ICSA 2011*. Detmold.
- Bernschütz, Benjamin; Christoph Pörschmann; Sascha Spors und Stefan Weinzierl (2011b): „Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren.” In: *Fortschritte der Akustik - DAGA 2011*. Düsseldorf, S. 661–662.
- Bernschütz, Benjamin; Philipp Stade und Max Rühl (2012): „Sound Field Analysis in Room Acoustics .” In: *Proc. of the 27th VDT International Convention*. Köln.
- Bernschütz, Benjamin; Arnau Vazquez; Christoph Pörschmann und Johannes M Arend (2014): „Binaural Reproduction of Plane Waves With Reduced Modal Order.” In: *Acta Acustica united with Acustica*, 100(5) S. 972–983.
- Berzborn, Marco; Ramona Bomhardt; Johannes Klein; Jan-Gerrit Richter und Michael Vorländer (2017): „The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing.” In: *Fortschritte der Akustik - DAGA 2017*. Kiel, S. 222–225.
- Blauert, Jens (1974): *Räumliches Hören*. Stuttgart: Hirzel.

- Blauert, Jens (1996): *Spatial Hearing. The Psychophysics of Human Sound Localization*. MIT Press.
- Blauert, Jens (2013): *The Technology of Binaural Listening*. Springer Science & Business Media.
- Blauert, Jens und Jonas Braasch (2008): „Räumliches Hören.“ In: *Handbuch der Audiotechnik*. Springer-Verlag, S. 87–121.
- BMAS (2017): *Weißbuch Arbeiten 4.0*. Berlin: Bundesministerium für Arbeit und Soziales.
- BMBF (2017): *Fortschritt durch Forschung und Innovation - Bericht zur Umsetzung der Hightech-Strategie*. Berlin: Bundesministerium für Bildung und Forschung.
- Bomhardt, Ramona; Matias de la Fuente Klein und Janina Fels (2017): „A high-resolution head-related transfer function and three-dimensional ear model database.“ In: *Proceedings of Meetings on Acoustics*. Honolulu.
- Borß, Christian und Rainer Martin (2009): „An Improved Parametric Model for Perception-Based Design of Virtual Acoustics.“ In: *Proc. of the 35th International AES Conference*. London.
- Bortz, Jürgen und Christof Schuster (2010): *Statistik für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer-Verlag.
- Bradley, John S (2011): „Review of objective room acoustics measures and future needs.“ In: *Applied Acoustics*, 72(10) S. 713–720.
- Bradley, John S; R Reich und S G Norcross (1999): „A just noticeable difference in C50 for speech.“ In: *Applied Acoustics*, 58(2) S. 99–108.
- Bradley, John S und G A Soulodre (1995a): „Objective measures of listener envelopment.“ In: *Journal of the Acoustical Society of America*, 98(5) S. 2590–2597.
- Bradley, John S und G A Soulodre (1995b): „The influence of late arriving energy on spatial impression.“ In: *Journal of the Acoustical Society of America*, 97(4) S. 2263.
- Brandão, Eric; Emiel Tijs; Arcanjo Lenzi und Hans-Elias de Bree (2011): „A Comparison of Three Methods to Calculate the Surface Impedance and Absorption Coefficient from Measurements Under Free Field or in situ Conditions.“ In: *Acta Acustica united with Acustica*, 97(6) S. 1025–1033.
- Breebaart, Jeroen; F Nater und Armin Kohlrausch (2009): „Parametric binaural synthesis: Background, applications and standards.“ In: *NAG/DAGA International Conference on Acoustics*.
- Brinkmann, Fabian (2011): *Individual headphone compensation for binaural synthesis*. Masterarbeit.
- Brinkmann, Fabian; David Ackermann; Lukas Aspöck und Stefan Weinzierl (2017a): „First international round robin on auralization: Results of the perceptual evaluation.“ In: *Journal of the Acoustical Society of America*, 141(5) S. 3996.

- Brinkmann, Fabian; Alexander Lindau und Stefan Weinzierl (2017b): „On the authenticity of individual dynamic binaural synthesis.” In: *Journal of the Acoustical Society of America*, 142(4) S. 1784–1795.
- Brinkmann, Fabian et al. (2017c): *The FABIAN head-related transfer function data base*. Tech. rep. URL <http://dx.doi.org/10.14279/depositonce-5718>.
- Brockhaus Enzyklopädie Online (2018): *Realität (allgemein)*. URL <https://brockhaus.de/ecs/permalink/29EB8654FB0E8B11C374D2C51F3F74A4.pdf>. Zugriff 20.04.2018.
- Brown, Andrew D; G Christopher Stecker und Daniel J Tollin (2014): „The Precedence Effect in Sound Localization.” In: *Journal of the Association for Research in Otolaryngology*, 16(1).
- Buchholz, Jörg M; Jens Blauert und John Mourjopoulos (2001): „Room Masking: Understanding and Modelling the Masking of Reflections in Rooms.” In: *Proc. of the 140th AES Convention*. Amsterdam.
- Burgtorf, Wolfgang (1961): „Untersuchungen zur Wahrnehmbarkeit verzögerter Schallsignale.” In: *Acoustica*, 11 S. 97–111.
- Chernyak, R I und N A Dubrovsky (1968): „Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise .” In: *Proc. of the 6th international Congress on Acoustics*. Tokyo.
- Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Coleman, Philip; Andreas Franck; Philip J B Jackson; Luca Remaggi und Frank Melchior (2016): „On object-based audio with reverberation.” In: *AES 60th Conference on Dereverberation and Reverberation of Audio, Music, and Speech*. Leuven, S. 1–10.
- Coleman, Philip; Andreas Franck; Dylan Menzies und Philip J B Jackson (2017a): „Object-Based Reverberation Encoding from First-Order Ambisonic RIRs.” In: *Proc. of the 142nd AES Convention*. Berlin.
- Coleman, Philip et al. (2017b): „Object-Based Reverberation for Spatial Audio.” In: *Journal of the Audio Engineering Society*, 65(1/2) S. 66–77.
- Coleman, Philip et al. (2018): „An Audio-Visual System for Object-Based Audio: From Recording to Listening.” In: *IEEE Transactions on Multimedia*.
- Cook, Richard K (1955): „Measurement of Correlation Coefficients in Reverberant Sound Fields.” In: *Journal of the Acoustical Society of America*, 27(6) S. 1072–1077.
- Cox, T J; William J Davies und Y W Lam (1993): „The sensitivity of listeners to early sound field changes in auditoria.” In: *Acustica*, 79 S. 27–41.
- Cremer, Lothar (1948): *Die wissenschaftlichen Grundlagen der Raumakustik Bd. 1*. Stuttgart: S. Hirzel Verlag.
- Danilenko, Leo (1969): „Binaurales Hören im nichtstationären diffusen Schallfeld.” In: *Kybernetik*, 6(2) S. 50–57.

- Dietrich, Pascal; Martin Guski; Johannes Klein; M Müller-Trapet und Michael Vorländer (2013): „Measurements and room acoustic analysis with the ita-toolbox for matlab.” In: *Fortschritte der Akustik - AIA DAGA 2013*. Meran, S. 1391–1394.
- DIN 1320:2009-12 (2009): Akustik - Begriffe. DIN Deutsches Institut für Normung e.V.
- DIN EN ISO 3382-1:2009-10 (2009): Akustik – Messung von Parametern der Raumakustik – Teil 1: Aufführungsräume. DIN Deutsches Institut für Normung e.V.
- Duraiswami, Ramani; Zhiyun Li; Elena Grassi; Nail A Gumerov und Dmitry N Zotkin (2005a): „Plane-wave decomposition analysis for spherical microphone arrays.” In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* S. 150–153.
- Duraiswami, Ramani; Dmitry N Zotkin; Elena Grassi; Nail A Gumerov und Larry S Davis (2005b): „High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues.” In: *Proc. of the 140th AES Convention*. New York.
- EBU (2014): Tech 3364 Audio Definition (Metadata) Model.
- Faller, Christof (2004): „Parametric coding of spatial audio.” In: *Proc. of the 7th Int. Conference on Digital Audio Effects*. Naples.
- Farina, Angelo; Alberto Amendola; Andrea Capra und Christian Varani (2011): „Spatial analysis of room impulse responses captured with a 32-capsules microphone array.” In: *Proc. of the 140th AES Convention*. London.
- Fels, Janina; Josefa Oberem und Bruno Masiero (2013): „Experiments on authenticity and naturalness of binaural reproduction via headphones.” In: *ICA 2013 Montreal*. ASA, S. 050044–050044.
- Fletcher, Harvey und Leon J Sivian (1927): „Binaural telephone system.” US Patent 1624486.
- Foster, Scott H; Elizabeth M Wenzel und R Michael Taylor (1991): „Real Time Synthesis of Complex Acoustic Environments.” In: *Final Program and Paper Summaries 1991 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE.
- Füg, Simone et al. (2014): „Design, Coding and Processing of Metadata for Object-Based Interactive Audio.” In: *Proc. of the 140th AES Convention*. Los Angeles.
- Fuß, Alexander; Fabian Brinkmann; T Jürgensohn und Stefan Weinzierl (2015): „Ein voll-sphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen.” In: *Fortschritte der Akustik - DAGA 2015*. Nürnberg, S. 1114–1117.
- Gardner, Bill und Keith Martin (1994): *HRTF Measurements of a KEMAR Dummy-Head Microphone*. Tech. rep.
- Geier, Matthias; Jens Ahrens und Sascha Spors (2008): „The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods.” In: *Proc. of the 124th AES Convention*. Amsterdam.
- Geier, Matthias und Sascha Spors (2012): „Spatial Audio with the SoundScape Renderer.” In: *Proc. of the 27th VDT International Convention*. Köln.

- Geisser, Seymour und Samuel W Greenhouse (1958): „An Extension of Box’s Results on the Use of the F Distribution in Multivariate Analysis.” In: *The Annals of Mathematical Statistics*, 29(3) S. 885–891.
- Gerzon, Michael A (1985): „Ambisonics in Multichannel Broadcasting and Video.” In: *Journal of the Audio Engineering Society*, 33(11) S. 859–871.
- GfK (2017): *Absatz von Kopfhörern/ Headsets weltweit von 2013 bis 2017*.
 URL <https://de.statista.com/statistik/daten/studie/326416/umfrage/entwicklung-der-verkaufszahlen-von-kopfhoeerern-weltweit/>. Zugriff 06.04.2018.
- Glass, Gene und Julian Stanley (1970): *Statistical methods in education and psychology*. Englewood Cliffs, N.J., Prentice-Hall.
- Gourévitch, Boris und Romain Brette (2012): „The impact of early reflections on binaural cues.” In: *Journal of the Acoustical Society of America*, 132(1) S. 9–19.
- Greenhouse, Samuel W und Seymour Geisser (1959): „On methods in the analysis of profile data.” In: *Psychometrika. A Journal Devoted to the Development of Psychology as a Quantitative Rational Science*, 24(2) S. 95–112.
- Haas, Helmut (1951): „Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache.” In: *Acustica*, 1 S. 49–58.
- Hartmann, William Morris (1983): „Localization of sound in rooms.” In: *Journal of the Acoustical Society of America*, 74(5) S. 1380–1391.
- Herre, Jürgen; J Hilpert; A Kuntz und Jan Plogsties (2015): „MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio.” In: *IEEE Journal of Selected Topics in Signal Processing*, 9 S. 770–779.
- Hidaka, T; Y Yamada und T Nakagawa (2007): „A new definition of boundary point between early reflections and late reverberation in room impulse responses.” In: *Journal of the Acoustical Society of America*, 122(1) S. 326–332.
- Hochberg, Yosef (1988): „A Sharper Bonferroni Procedure for Multiple Tests of Significance.” In: *Biometrika*, 75(4) S. 800–802.
- Hohnerlein, Christoph und Jens Ahrens (2017): „Spherical Microphone Array Processing in Python with the sound field analysis-py Toolbox.” In: *Fortschritte der Akustik - DAGA 2017*. Kiel, S. 1033–1036.
- Horbach, Ulrich; A Karamustafaoglu und Renato Pellegrini (1999): „Design and applications of a data-based auralization system for surround sound.” In: *Proc. of the 106th AES Convention*. München.
- Horn, Roger A (1990): „The Hadamard product.” In: *Proceedings of Symposia in Applied Mathematics*. Providence: American Mathematical Society, S. 87–169.
- Huopaniemi, Jyri; Lauri Savioja; Tapio Lokki und Riitta Väänänen (2000): „Virtual acoustics - applications and technology trend.” In: *Proc. the European Signal Processing Conference*. Tampere, S. 2201–2208.
- Huygens, Christiaan (1690): *Traité de la lumière*. Leyde.

- ISO/IEC 11172-3 (1999): Information technology – Coding of audio-visual objects – Part 3: Audio.
- ISO/IEC 14496-3 (1993): Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3: Audio.
- ITU-R BS 1534-3 (2015): Method for the subjective assessment of intermediate quality level of audio systems. International Telecommunication Union.
- ITU-R BS 1770-4 (2015): Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunication Union.
- ITU-R BS.1116-3 (2015): Methods for the subjective assessment of small impairments in audio systems. International Telecommunication Union.
- Jacobsen, Finn (1979): *The diffuse sound field: Statistical considerations concerning the reverberant field in the steady state*. Acoustics Laboratory, Technical University of Denmark.
- Jeong, Cheol-Ho (2016): „Diffuse sound field: Challenges and misconceptions.” In: *Proceedings of the INTER-NOISE 2016 - 45th International Congress and Exposition on Noise Control Engineering: Towards a Quieter Future*. Danmarks Tekniske Universitet, Lyngby, Denmark, S. 1015–1021.
- Jot, Jean Marc (1997): „Efficient models for reverberation and distance rendering in computer music and virtual audio reality.” In: *Proc. of the COST G-6 Conference on Digital Audio Effects*. Thessaloniki.
- Jot, Jean Marc (1999): „Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces.” In: *Multimedia systems*, 7(1) S. 55–69.
- Jot, Jean Marc; Laurent Cerveau und O Warusfel (1997): „Analysis and synthesis of room reverberation based on a statistical time-frequency model.” In: *Proc. of the 103rd AES Convention*. New York.
- Jot, Jean Marc und Antoine Chaigne (1991): „Digital delay networks for designing artificial reverberators.” In: *Proc. of the 90th AES Convention*. Paris.
- Kleiner, Mendel; Bengt-Inge Dalenbäck und U P Svensson (1993): „Auralization An Overview.” In: *Journal of the Audio Engineering Society*, 41(11) S. 861–875.
- Kolarik, Andrew J; Brian C J Moore; Pavel Zahorik; Silvia Cirstea und Shahina Pardhan (2015): „Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss.” In: *Attention, Perception, & Psychophysics*, 78(2) S. 373–395.
- Krokstad, A; S Strom und S Sørsdal (1968): „Calculating the acoustical room response by the use of a ray tracing technique.” In: *Journal of Sound and Vibration*, 8(1) S. 118–125.
- Kuhn-Rahloff, Clemens (2012): *Realitätstreue, Natürlichkeit, Plausibilität*. Perzeptive Beurteilungen in der Elektroakustik. Berlin, Heidelberg: Springer-Verlag.
- Kuttruff, Heinrich (2009): *Room Acoustics, Sixth Edition*. 5. Spon Press.

- Kuusinen, Antti und Tapio Lokki (2015): „Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics.” In: *Journal of the Acoustical Society of America*, 138(5) S. 3148–3159.
- Kuusinen, Antti und Tapio Lokki (2017): „Wheel of Concert Hall Acoustics.” In: *Acta Acustica united with Acustica*, 103(2) S. 185–188.
- Laird, I; D Murphy und Paul Chapman (2014): „Comparison of Spatial Audio Techniques for Use in Stage Acoustic Laboratory Experiments.” In: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*.
- Lebedev, Vyacheslav Ivanovich (1977): „Spherical quadrature formulas exact to orders 25–29.” In: *Siberian Mathematical Journal*, 18(1) S. 99–107.
- Lehnert, Hilmar und Jens Blauert (1992): „Principles of binaural room simulation.” In: *Applied Acoustics*, 36(3-4) S. 259–291.
- Lindau, Alexander (2014): *Binaural Resynthesis of Acoustical Environments. Technology and Perceptual Evaluation*. Dissertation, Berlin.
- Lindau, Alexander und Fabian Brinkmann (2012): „Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings.” In: *Journal of the Audio Engineering Society*, 60(1/2) S. 54–62.
- Lindau, Alexander; Linda Kosanke und Stefan Weinzierl (2012): „Perceptual evaluation of model-and signal-based predictors of the mixing time in binaural room impulse responses.” In: *Journal of the Audio Engineering Society*, 60(11) S. 887–898.
- Lindau, Alexander und Stefan Weinzierl (2012): „Assessing the plausibility of virtual acoustic environments.” In: *Acta Acustica united with Acustica*, 98(5) S. 804–810.
- Lindau, Alexander et al. (2014): „A spatial audio quality inventory (SAQI).” In: *Acta Acustica united with Acustica*, 100(5) S. 984–994.
- Lokki, Tapio und Matti Karjalainen (2000): „An auditorily motivated analysis method for room impulse responses.” In: *Proc. of the COST G-6 Conference on Digital Audio Effects*. Verona.
- Lubberhuizen, Wessel (2010): *Near Perfect Reconstruction Polyphase Filterbank*. URL <http://de.mathworks.com/matlabcentral/fileexchange/15813-near-perfect-reconstruction-polyphase-filterbank>. Zugriff 06.04.2018.
- Lundebj, Anders; Tor Erik Vigran; Heinrich Bietz und Michael Vorländer (1995): „Uncertainties of measurements in room acoustics.” In: *Acta Acustica united with Acustica*, 81(4) S. 344–355.
- Mackensen, Philip; Klaus Reichenauer und Günther Theile (1998): „Einfluss der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören.” In: *Proc. of the 20th VDT International Convention*.
- Maempel, Hans-Joachim und Michael Horn (2017): „The Virtual Concert Hall-A Research Tool for the Experimental Investigation of Audiovisual Room Perception.” In: *International Journal on Stereo & Immersive Media*, 1(1) S. 78–98.

- Melchior, Frank (2011): *Investigations on spatial sound design based on measured room impulse responses*. Dissertation.
- Melchior, Frank; Christoph Sladeczek; Andreas Partzsch und Sandra Brix (2010): „Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis.” In: *Proc. of the 40th AES International Conference*. Tokyo.
- Menzer, Fritz (2010): *Binaural audio signal processing using interaural coherence matching*. Dissertation.
- Menzer, Fritz (2012): „Efficient Binaural Audio Rendering Using Independent Early and Diffuse Paths.” In: *AES 132nd Convention*. Budapest.
- Menzer, Fritz und Christof Faller (2009): „Binaural Reverberation Using a Modified Jot Reverberator with Frequency-Dependent Interaural Coherence Matching.” In: *Proc. of the 140th AES Convention*. München.
- Menzer, Fritz und Christof Faller (2010): „Investigations on an Early-Reflection-Free Model for BRIRs*.” In: *Journal of the Audio Engineering Society*, 58(9) S. 709–723.
- Menzer, Fritz; Christof Faller und Hervé Lissek (2010): „Obtaining Binaural Room Impulse Responses From B-Format Impulse Responses Using Frequency-Dependent Coherence Matching.” In: *IEEE Transactions on Audio, Speech, and Language Processing*, 19(2) S. 396–405.
- Merimaa, Juha; Tapio Lokki; Timo Peltonen und Matti Karjalainen (2001): „Measurement, analysis, and visualization of directional room responses.” In: *Proc. of the 140th AES Convention*. New York.
- Merimaa, Juha und Ville Pulkki (2005): „Spatial impulse response rendering I: Analysis and synthesis.” In: *Journal of the Audio Engineering Society*, 53(12) S. 1115–1127.
- Meulemann, Jan; Benjamin Bernschütz und Christoph Pörschmann (2011): „Entwurf und Aufbau eines konzentrischen Mehrwegedodekaeders.” In: *Fortschritte der Akustik - DAGA 2011*. Düsseldorf, S. 719–720.
- Meyer, Jens und Gary Elko (2002): „A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield.” In: *Proc. of ICASSP '02*, 2 S. II–1781–II–1784.
- Møller, Henrik (1992): „Fundamentals of binaural technology.” In: *Applied Acoustics*, 36(3–4) S. 171–218.
- Moorer, James A (1979): „About this reverberation business.” In: *Computer music journal*, 3(2) S. 13–28.
- Möser, Michael (2009): *Messtechnik der Akustik*. Berlin: Springer-Verlag.
- Müller, Swen (1999): *Digitale Signalverarbeitung für Lautsprecher*. Dissertation.
- Naylor, Patrick A; Anastasis Kounoudes; Jon Gudnason und Mike Brookes (2006): „Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm.” In: *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1) S. 34–43.

- Nilsson, Niels Christian; Rolf Nordahl und Stefania Serafin (2016): „Immersion Revisited: A review of existing definitions of immersion and their relation to different theories of presence.” In: *Human Technology*, 12(2) S. 108–134.
- NWDR (1953): „Das neue Kölner Funkhaus des nordwestdeutschen Rundfunks.” In: *Technische Hausmitteilungen des Nordwestdeutschen Rundfunks*, 5(5/6) S. 78–107.
- O’Donovan, A; Ramani Duraiswami und Dmitry N Zotkin (2008): „Imaging concert hall acoustics using visual and audio cameras.” In: *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, S. 5284–5287.
- Okano, Toshiyuki (2002): „Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections.” In: *Journal of the Acoustical Society of America*, 111(1) S. 217–229.
- Olive, Sean E und Floyd E Toole (1989): „The detection of reflections in typical rooms.” In: *Journal of the Audio Engineering Society*, 37(7/8) S. 539–553.
- Pätynen, Jukka und Tapio Lokki (2011): „Evaluation of Concert Hall Auralization with Virtual Symphony Orchestra.” In: *Building Acoustics*, 18(3-4) S. 349–366.
- Peters, Nils; Trond Lossius und Jan C Schacher (2013): „The Spatial Sound Description Interchange Format - Principles, Specification, and Examples.” In: *Computer music journal*.
- Polack, Jean-Dominique (1992): „Modifying Chambers to play Billiards: the Foundations of Reverberation Theory.” In: *Acustica*, 76 S. 257–272.
- Pörschmann, Christoph; Philipp Stade und Johannes M Arend (2017): „Binauralization of Omnidirectional Room Impulse Responses-Algorithm and Technical Evaluation.” In: *Proc. of the 20th International Conference on Digital Audio Effects*. Edinburgh, S. 345–352.
- Pörschmann, Christoph und Stephan Wiefing (2015): „Perceptual Aspects of Dynamic Binaural Synthesis based on Measured Omnidirectional Room Impulse Responses.” In: *International Conference on Spatial Audio - ICSA*. Graz.
- Potard, Guillaume und Ian Burnett (2004): „An XML-based 3D Audio Scene Metadata Scheme.” In: *Proc. of the 25th AES Conference*. London: Audio Engineering Society.
- Puente León, Fernando und Uwe Kiencke (2011): *Messtechnik*. Systemtheorie für Ingenieure und Informatiker, 8. Springer.
- Pulkki, Ville (2007): „Spatial Sound Reproduction with Directional Audio Coding.” In: *Journal of the Audio Engineering Society*, 55(6) S. 503–516.
- Pulkki, Ville; Symeon Delikaris-Manias und Archontis Politis (2017): *Parametric Time-Frequency Domain Spatial Audio*. Chichester, UK: John Wiley & Sons.
- Pulkki, Ville und Juha Merimaa (2006): „Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests.” In: *Journal of the Audio Engineering Society*, 54(1-2) S. 3–20.
- Rafaely, Boaz (2004): „Plane-wave decomposition of the sound field on a sphere by spherical convolution.” In: *Journal of the Acoustical Society of America*, 116(4) S. 2149–2157.

- Rafaely, Boaz (2005): „Analysis and design of spherical microphone arrays.” In: *IEEE Transactions on Speech and Audio Processing*, 13(1) S. 135–143.
- Rafaely, Boaz; Barak Weiss und Eitan Bachmat (2007): „Spatial Aliasing in Spherical Microphone Arrays.” In: *IEEE Transactions on Signal Processing*, 55(3) S. 1003–1010.
- Rasch, Björn; Malte Friese; Wilhelm Hofmann und Ewald Naumann (2014): *Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Springer-Verlag.
- Reilly, Andrew und David McGrath (1995): „Real-time auralization, with head tracking.” In: *Proc. of the 5th Australian Regional AES Convention*. Sydney.
- Remaggi, Luca; Philip J B Jackson und Philip Coleman (2015): „Estimation of room reflection parameters for a reverberant spatial audio object.” In: *Proc. of the 138th AES Convention*. Warschau.
- Richter, Jan-Gerrit; Gottfried Behler und Janina Fels (2016): „Evaluation of a fast HRTF measurement system.” In: *Proc. of the 140th AES Convention*. Paris.
- Rossing, Thomas (2015): „Acoustics in Halls for Speech and Music.” In: *Springer Handbook of Acoustics*. Springer, S. 301–350.
- Rubak, Per und Lars G Johansen (1998): „Artificial reverberation based on a pseudo-random impulse response Part I.” In: *Proc. of the 140th AES Convention*. Amsterdam.
- Rubak, Per und Lars G Johansen (1999): „Artificial Reverberation Based on a Pseudo-Random Impulse Response II.” In: *Proc. of the 106th AES Convention*. München: Audio Engineering Society.
- Rubak, Per und Lars G Johansen (2003): „Coloration in Natural and Artificial Room Impulse Responses.” In: *23rd International AES Conference*. Denmark: Audio Engineering Society.
- Rühl, Max (2012): *Entwicklung eines Verfahrens zur Detektion von Reflexionen in messtechnisch erfassten Schallfeldern*. Bachelorarbeit, Köln.
- Sabine, Wallace Clement (1922): *Collected papers on acoustics*. Cambridge Harvard University Press.
- Savioja, Lauri; Akio Ando; Ramani Duraiswami; E A P Habets und Sascha Spors (2015): „Introduction to the Issue on Spatial Audio.” In: *IEEE Journal of Selected Topics in Signal Processing*, 9(5) S. 767–769.
- Savioja, Lauri; Tapio Lokki und Jyri Huopaniemi (2002): „Auralization applying the parametric room acoustic modeling technique-the DIVA auralization system.” In: *Proc. of the 8th International Conference on Auditory Display*.
- Scheirer, Eric D; R Vaananen und Jyri Huopaniemi (1999): „AudioBIFS: Describing audio scenes with the MPEG-4 multimedia standard.” In: *IEEE Transactions on Multimedia*, 1(3) S. 237–250.
- Schröder, Dirk und Michael Vorländer (2011): „RAVEN: A real-time framework for the auralization of interactive virtual environments.” In: *Forum Acusticum* S. 1541–1546.

- Schroeder, Manfred R (1954): „Die statistischen Parameter der Frequenzkurven von grossen Räumen.“ In: *Acoustica*, 4 S. 594–600.
- Schroeder, Manfred R (1962): „Natural Sounding Artificial Reverberation.“ In: *Journal of the Acoustical Society of America*, 10 S. 219–223.
- Schroeder, Manfred R (1965): „New Method of Measuring Reverberation Time.“ In: S. 409–412.
- Schroeder, Manfred R (1996): „The “Schroeder frequency” revisited.“ In: *Journal of the Acoustical Society of America*, 99(5) S. 3240–3241.
- Schroeder, Manfred R und B F Logan (1961): „Colorless artificial reverberation.“ In: *IRE Transactions on Audio*, AU-9(6) S. 209–214.
- Schubert, P (1969): „Die Wahrnehmbarkeit von Rückwürfen bei Musik.“ In: *Z. Hochfrequenztechn. u. Elektroakust.*, 78 S. 230–245.
- Schultz, T J (1971): „Diffusion in reverberation rooms.“ In: *Journal of Sound and Vibration*, 16(1) S. 17–28.
- Schüller, Hans Wilhelm (1994): *Digitale Signalverarbeitung 1. Analyse Diskreter Signale Und Systeme*. Springer.
- Seraphim, Hans-Peter (1958): „Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen.“ In: *Acta Acustica united with Acustica*, 8 S. 280–284.
- Seraphim, Hans-Peter (1961): „Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall.“ In: *Acustica*, 11 S. 80–91.
- Shapiro, Samuel S und Martin B Wilk (1965): „An analysis of variance test for normality (complete samples).“ In: *Biometrika*, 52(3/4) S. 591.
- Sheaffer, Jonathan; Maarten van Walstijn; Boaz Rafaely und K Kowalczyk (2015): „Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing.“ In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(12) S. 2125–2135.
- Sheaffer, Jonathan; S Villeval und Boaz Rafaely (2014): „Rendering Binaural Room Impulse Responses from Spherical Microphone Array Recordings Using Timbre Correction.“ In: *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*. Berlin.
- Shinn-Cunningham, Barbara (2003): „Acoustics and perception of sound in everyday environments.“ In: *Proc. of 3rd International Workshop on Spatial Media*. Aisu-Wakamatsu.
- Shinn-Cunningham, Barbara und S Ram (2003): „Identifying where you are in a room: Sensitivity to room acoustics.“ In: *Proc. of the International Conference on Auditory Display*. Boston, S. 21–24.
- Shirley, Ben; Rob Oldfield; Frank Melchior und Johann-Markus Batke (2013): „Platform Independent Audio.“ In: *Media Production, Delivery and Interaction for Platform Independent Systems*. Chichester, UK: John Wiley & Sons, Ltd, S. 130–165.

- Slater, Mel (2009): „Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments.” In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535) S. 3549–3557.
- Slater, Mel; Beau Lotto; Maria Marta Arnold und Maria V Sanchez-Vives (2009): „How we experience immersive virtual environments: The concept of presence and its measurement.” In: *Anuario de Psicología*, 40(2) S. 193–210.
- Spors, Sascha und Rudolf Rabenstein (2008): „The theory of wave field synthesis revisited.” In: *124th AES Convention*. Amsterdam.
- Spors, Sascha und Till Rettberg (2018): „On the Estimation of Acoustic Reflection Coefficients from In-Situ Measurements using a Spherical Microphone Array.” In: *Fortschritte der Akustik - DAGA 2018*. München, S. 1326–1329.
- Spors, Sascha et al. (2013): „Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State.” In: *Proceedings of the IEEE*, 101(9) S. 1920–1938.
- Stade, Philipp (2013): *Zur Untersuchung der Akustik von Studios im WDR-Funkhaus unter Anwendung von Wellenfeldanalyseverfahren*. Masterarbeit, Köln.
- Stade, Philipp (2015): „Perzeptive Untersuchung zur Mixing Time und deren Einfluss auf die Auralisation.” In: *Fortschritte der Akustik - DAGA 2015*. Nürnberg, S. 1103–1106.
- Stade, Philipp und Johannes M Arend (2016a): „A Perception-Based Parametric Model for Synthetic Late Binaural Reverberation.” In: *Fortschritte der Akustik - DAGA 2016*. Aachen, S. 63–66.
- Stade, Philipp und Johannes M Arend (2016b): „Perceptual Evaluation of Synthetic Late Binaural Reverberation Based on a Parametric Model.” In: *Proc. of the AES International Conference on Headphone Technology*. Aalborg, S. 80–87.
- Stade, Philipp und Johannes M Arend (2017): „Synthetic Reflections for Binaural Rendering using Sound Field Analysis.” In: *Fortschritte der Akustik - DAGA 2017*. Kiel, S. 1089–1092.
- Stade, Philipp; Johannes M Arend und Karl Goebels (2014): „Anwendung neuer Methoden zur raumakustischen Analyse in einem Regieraum der WDR-Hörspielstudios.” In: *Fortschritte der Akustik - DAGA 2014*. Oldenburg, S. 602–603.
- Stade, Philipp; Johannes M Arend und Christoph Pörschmann (2017a): „A parametric model for the synthesis of binaural room impulse responses.” In: *Proceedings of Meetings on Acoustics*, 30.
- Stade, Philipp; Johannes M Arend und Christoph Pörschmann (2017b): „Perceptual Evaluation of Synthetic Early Binaural Room Impulse Responses Based on a Parametric Model.” In: *Proc. of the 142nd AES Convention*. Berlin.
- Stade, Philipp; Benjamin Bernschütz und Max Rühl (2012): „A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios.” In: *Proc. of the 27th VDT International Convention*. Köln.
- Stewart, Rebecca und Mark Sandler (2007a): „Statistical measures of early reflections of room impulse responses.” In: *Proc. of the 10th international Conference on Digital Audio Effects*. Bordeaux.

- Stewart, Rebecca und Mark Sandler (2007b): „Statistical measures of early reflections of room impulse responses.” In: *Proceedings of the 10th International Conference on Digital Audio Effects, DAFx 2007*. Queen Mary, University of London, London, United Kingdom, S. 59–62.
- Stroud, A H und Don Secrest (1966): „Gaussian Quadrature Formulae.” Prentice-Hall.
- Student (1908): „The probable error of a mean.” In: *Biometrika*, 6(1).
- Tervo, Sakari; T Korhonen und Tapio Lokki (2010): „Estimation of Reflections from Impulse Responses.” In: *Proc. of the International Symposium on Room Acoustics*. Melbourne.
- Tervo, Sakari; Jukka Pätynen; Antti Kuusinen und Tapio Lokki (2013): „Spatial Decomposition Method for Room Impulse Responses.” In: *Journal of the Audio Engineering Society*, 61(1/2) S. 17–28.
- Tervo, Sakari; Jukka Pätynen und Tapio Lokki (2012): „Acoustic Reflection Localization from Room Impulse Responses.” In: *Acta Acustica united with Acustica*, 98(3) S. 418–440.
- Toole, Floyd E (2006): „Loudspeakers and rooms for sound reproduction - A scientific review.” In: *Journal of the Audio Engineering Society*.
- Toole, Floyd E (2008): *Sound reproduction: loudspeakers and rooms*. Burlington, MA: Elsevier.
- Tourbabin, V und Boaz Rafaely (2015): „On the Consistent Use of Space and Time Conventions in Array Processing.” In: *Acta Acustica united with Acustica*, 101(3) S. 470–473.
- Traer, James und Josh H McDermott (2016): „Statistics of natural reverberation enable perceptual separation of sound and space.” In: *PNAS*, 113(48) S. E7856–E7865.
- Trivi, J M und Jean Marc Jot (2002): „Rendering MPEG-4 AABIFS content through a low-level cross-platform 3D audio api.” In: *Multimedia and Expo* S. 513–516.
- Vaananen, R und Jyri Huopaniemi (2004): „Advanced AudioBIFS: Virtual Acoustics Modeling in MPEG-4 Scene Description.” In: *IEEE Transactions on Multimedia*, 6(5) S. 661–675.
- Välimäki, Vesa; Julian D Parker; Lauri Savioja; Julius O Smith und Jonathan S Abel (2012): „Fifty Years of Artificial Reverberation.” In: *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5) S. 1421–1448.
- Välimäki, Vesa; Julian D Parker; Julius O Smith und Jonathan S Abel (2016): „More Than Fifty Years of Artificial Reverberation.” In: *AES 60th Conference on Dereverberation and Reverberation of Audio, Music, and Speech*. Leuven.
- Van Veen, Barry D und Kevin M Buckley (1988): „Beamforming: a versatile approach to spatial filtering.” In: *IEEE Acoustic, Speech and Signal Processing Magazine*, 5(2) S. 4–24.
- Vazquez, Arnau (2013): „Scale - A Software Tool for Listening Experiments.” In: *Fortschritte der Akustik - AIA DAGA 2013*. Meran, S. 1316–1319.

- Vazquez, Arnau (2015): „Scale - Conducting Psychoacoustic Experiments with Dynamic Binaural Synthesis.” In: *Fortschritte der Akustik - DAGA 2015*. Nürnberg, S. 1128–1130.
- Vesa, Sampo und Tapio Lokki (2006): „Detection of room reflections from a binaural room impulse response.” In: *Proc. of the COST G-6 Conference on Digital Audio Effects*. S. 215–220.
- Vilkamo, Juha; Tapio Lokki und Ville Pulkki (2009): „Directional audio coding: Virtual microphone-based synthesis and subjective evaluation.” In: *Journal of the Audio Engineering Society*, 57(9) S. 709–724.
- Vorländer, Michael (2007): *Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin, Heidelberg: Springer Science & Business Media.
- Vorländer, Michael (2011): „What do we know in room acoustics?” In: *Forum Acusticum 2011*. Aalborg.
- Wade, Nicholas J und Diana Deutsch (2008): „Binaural Hearing—Before and After the Stethophone.” In: *Acoustics Today*, 4(3) S. 16–14.
- Wallach, Hans (1949): „The precedence effect in sound localization.” In: *The American Journal of Psychology*, 62 S. 315–336.
- Warncke, Hans (1941): „Die Grundlagen der raumbezüglichen stereophonischen Übertragung im Tonfilm.” In: *Akust. Z.* S. 174–188.
- WDR (1993): „Funkhaus Wallrafplatz.” In: *WDR Magazin*.
- Wendt, Torben; Steven van de Par und Stephan D Ewert (2014): „A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation.” In: *Journal of the Audio Engineering Society*, 62(11) S. 748–766.
- Wenzel, Elizabeth M; Scott S Fisher; Philip K Stone und Scott H Foster (1990): „A System for Three-Dimensional Acoustic Visualization in a Virtual Environment Workstation.” In: *Proc. of the first IEEE Conference on Visualization*.
- Williams, Earl George (1999): *Fourier Acoustics. Sound Radiation and Nearfield Acoustical Holography*. Academic Press.
- Zahorik, Pavel (2009): „Perceptually relevant parameters for virtual listening simulation of small room acoustics.” In: *Journal of the Acoustical Society of America*, 126(2) S. 776–16.