

Bachelorarbeit
im Studiengang Audiovisuelle Medien

Immersives Audio über Soundbars

Eigenschaften und Vergleich
zur diskreten Lautsprecherwiedergabe
am Beispiel von Dolby Atmos

vorgelegt von Jascha Guckel
an der Hochschule der Medien Stuttgart am 22.09.2025
zur Erlangung des akademischen Grades eines Bachelor of Engineering

Erstprüfer: Prof. Dr. Frank Melchior
Zweitprüfer: Prof. Oliver Curdt

Ehrenwörtliche Erklärung

Hiermit versichere ich, Jascha Guckel, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Immersives Audio über Soundbars – Eigenschaften und Vergleich zur diskreten Lautsprecherwiedergabe am Beispiel von Dolby Atmos“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Ebenso sind alle Stellen, die mit Hilfe eines KI-basierten Schreibwerkzeugs erstellt oder überarbeitet wurden, kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§ 24 Abs. 2 Bachelor-SPO (Vollzeit)) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

A handwritten signature in black ink, appearing to read 'Jascha Guckel', is written over a horizontal line.

Stuttgart, 22.09.2025

Kurzfassung

Soundbars gelten als beliebtes und benutzerfreundliches Endgerät zur räumlichen Wiedergabe immersiver Audioinhalte. Die vorliegende Arbeit untersucht die spezifischen Eigenschaften dieses Wiedergabesystems im Vergleich zur diskreten Lautsprecherwiedergabe am Beispiel von Dolby Atmos. Der Fokus liegt dabei auf der Identifizierung von Variablen innerhalb des Produktions- und Distributionsprozesses, welche eine Einflussnahme auf die Soundbar-Wiedergabe ermöglichen, um die künstlerische Intention Tonschaffender auch über diesen Ausspielweg transportieren zu können.

Ein vorgestellter praktischer Versuchsaufbau ermöglicht den direkten Vergleich zwischen diskreter Wiedergabe und Soundbar-Wiedergabe. Anhand illustrativer Hörbeispiele werden so Gemeinsamkeiten und Unterschiede zwischen den verglichenen Wiedergabesystemen aufgezeigt. Dabei werden primär die subjektiven Wahrnehmungsattribute Lokalisierbarkeit, Apparent Source Width und Listener Envelopment betrachtet. Diese werden um objektive Messwerte in Form von Raumimpulsantworten und binauralen Raumimpulsantworten beider Systeme ergänzt.

Die Ergebnisse legen nahe, dass insbesondere Abweichungen hinsichtlich der Lokalisation von Hörereignissen zwischen den Wiedergabesystemen zu erwarten sind und die Soundbar-Wiedergabe eine höher empfundene Apparent Source Width begünstigen kann.

Abstract

Soundbars are considered a popular and user-friendly device for spatial reproduction of immersive audio content. This thesis examines the specific characteristics of this playback system in comparison to discrete loudspeaker reproduction, using Dolby Atmos as an example. Focus lies on identifying variables within the production and distribution process, that allow for affecting soundbar playback in order to convey the sound creator's artistic intention through this medium.

A practical experimental setup is presented, that enables direct comparison between discrete playback and soundbar playback. Using illustrative audio examples, similarities and differences between the compared playback systems are demonstrated. In context, the subjective perceptual attributes of localizability, apparent source width and listener envelopment are of interest. These are complemented by objective measurements of both playback systems, namely room impulse responses and binaural room impulse responses.

The results suggest that deviations in localization of auditory events between both playback systems are to be expected, and that soundbar playback could enhance a higher perceived apparent source width.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung	2
Kurzfassung	3
Abstract	3
Abbildungsverzeichnis	7
Tabellenverzeichnis	8
Abkürzungsverzeichnis	9
Gender-Hinweis	11
1. Einleitung	12
2. Grundlagen	13
2.1 Psychoakustik und räumliches Hören	13
2.1.1 Hörereignis	13
2.1.2 Kopfbezogenes Koordinatensystem	13
2.1.3 Räumliches Hören mit einer Schallquelle	14
2.1.3.1 Schalleinfall aus der Medianebene	14
2.1.3.2 Schalleinfall aus der Horizontalebene	15
2.1.3.3 Interaurale Laufzeitdifferenzen	16
2.1.3.4 Interaurale Pegeldifferenzen	16
2.1.4 Räumliches Hören mit mehreren Schallquellen	17
2.1.4.1 Summenlokalisation und Phantomschallquellen	17
2.1.4.2 Präzedenzeffekt und Haas-Effekt	19
2.1.4.3 Primärschallunterdrückung	20
2.1.4.4 Interaurale Kohärenz	22
2.2 Immersives Audio	23
2.2.1 Wahrnehmungsattribute immersiver Höreindrücke	23
2.2.1.1 Lokalisierbarkeit	23
2.2.1.2 Apparent Source Width	23
2.2.1.3 Listener Envelopment	24
2.3 Räumliche Tonwiedergabe	24
2.3.1 Lautsprecher-Stereofonie	24
2.3.1.1 Mehrkanal-Stereofonie	24
2.3.1.2 Diskrete Lautsprecher-Anordnungen	28
2.3.1.3 Downmix	31
2.4. Objektbasiertes Audio	31
2.4.1 Dolby Atmos	32

2.5 Soundbar.....	33
2.5.1 Funktionsprinzipien	34
2.5.1.1 Direktabstrahlung	34
2.5.1.2 Beamforming	36
2.6 Schall im Raum	36
2.6.1 Zeitlicher Aufbau des Schallfeldes	37
2.6.1.1 Direktschall.....	38
2.6.1.2 Frühe Reflexionen	39
2.6.1.3 Diffusschall.....	41
2.7 Wiedergabekompatibilität	41
3. Signalfluss – Von der Postproduktion bis zur Soundbar- Wiedergabe	42
3.1 Postproduktion.....	44
3.1.1 Content-Erstellung in der DAW	44
3.1.2 Dolby Atmos Renderer	45
3.1.2.1 Eingangskonfiguration.....	45
3.1.2.2 Ausgangskonfiguration.....	46
3.1.2.3 Trim- und Downmix-Editor.....	46
3.1.2.4 Monitoring-Optionen	48
3.1.3 Monitoring in der Tonregie.....	48
3.1.3.1 Anforderungen an Tonregien.....	49
3.2 Export-Formate und Codecs.....	52
3.2.1 Master File-Formate.....	52
3.2.1.1 Dolby Atmos Master Fileset.....	52
3.2.1.2 ADM BWF	52
3.2.1.3 IMF IAB	53
3.2.2 Consumer-orientierte Formate	53
3.2.2.1 Spatial Coding	53
3.2.2.2 Dolby Digital Plus Joint Object Coding.....	56
3.2.2.3 Dolby TrueHD	56
3.2.2.4 Dolby AC-4	57
3.3 Decodierung und Prozessierung durch die Soundbar.....	57
3.3.1 Decodierung und Rendering	58
3.3.2 Zusätzliches Processing	58
3.3.2.1 Automatische Raumkorrektur.....	59
3.4 Akustische Wiedergabe über die Soundbar.....	60
4. Analyse der akustischen Soundbar-Wiedergabe und Vergleich mit diskreter Lautsprecherwiedergabe	60
4.1 Methodik.....	60
4.1.1 Subjektive Kriterien – Hörbeispiele.....	60

4.1.2 Objektive Kriterien – Messungen	61
4.2 Versuchsaufbau	61
4.3 Überblick	64
4.4 Wiedergabe von Einzelkanälen und Kanalgruppen	64
4.4.1 Front-Kanäle	65
4.4.2 Surround-Kanäle	68
4.4.3 Top Surround-Kanäle.....	79
4.5 Mehrkanal-Wiedergabe – kohärente Signale	87
4.5.1 Statisches Panning.....	87
4.5.2 Dynamisches Panning	88
4.5.2.1 Dynamisches Panning im Middle Layer.....	88
4.5.2.2 Dynamisches Panning zwischen Middle Layer und Upper Layer.....	89
4.5.2.3 Dynamisches Panning im Upper Layer	92
4.6 Mehrkanal-Wiedergabe – teilkohärente und inkohärente Signale	93
4.6.1 Object Size	93
4.6.3 Teilkohärente Signale	93
4.6.2 Inkohärente Signale.....	95
5. Diskussion	98
5.1 Zusammenfassung	98
5.2 Limitationen und Empfehlungen für zukünftige Arbeiten	101
6. Fazit	102
7. Hörbeispielverzeichnis.....	105
Literaturverzeichnis.....	109
Digitaler Anhang	118

Abbildungsverzeichnis

Abbildung 1: Kopfbezogenes Koordinatensystem	14
Abbildung 2: Richtungsbestimmende Bänder	15
Abbildung 3: Stereodreieck	17
Abbildung 4: Summenlokalisationskurven.....	18
Abbildung 5: Hörereignisauslenkung durch ICTD	20
Abbildung 6: ITD/ILD-Trading	21
Abbildung 7: Lokalisation bei ICTD und ICLD	22
Abbildung 8: 3/2-Lautsprecher-Anordnung nach ITU-R BS.775-4.....	25
Abbildung 9: Vector Base Amplitude Panning	27
Abbildung 10: Lautsprecher im Middle Layer für diskrete Dolby Atmos 7.1.4-Anordnung	29
Abbildung 11: Lautsprecher im Upper Layer für diskrete Dolby Atmos 7.1.4-Anordnung	30
Abbildung 12: Schematische Darstellung einer direkt abstrahlenden Soundbar.....	35
Abbildung 13: Schematisches Reflektogramm mit zeitlichem Aufbau des Schallfeldes	37
Abbildung 14: Prinzip der Spiegelschallquelle	40
Abbildung 15: Signalkette zwischen Postproduktion und Endnutzer	43
Abbildung 16: Toleranzbereiche für die Betriebsschallpegelkurve in Referenz-Abhörräumen.....	50
Abbildung 17: Toleranzbereiche für Nachhallzeiten in Referenz-Abhörräumen	51
Abbildung 18: Audio-Objekt Clustering	54
Abbildung 19: JOC-Encoding	55
Abbildung 20: JOC-Decoding	58
Abbildung 21: Explosionsdarstellung der Soundbar	62
Abbildung 22: Verwendeter Versuchsaufbau für Hörvergleiche und Messungen	63
Abbildung 23: ETCs des Center-Kanals von diskretem System und Soundbar.....	66
Abbildung 24: Klarheitsmaß C_{50} für den Center-Kanal von diskretem System und Soundbar	67
Abbildung 25: ETCs des Left Surround-Kanals von diskretem System und Soundbar	70
Abbildung 26: Klarheitsmaß C_{80} der Surround-Kanäle von diskretem System und Soundbar	72
Abbildung 27: Annotierte ETC des Kanals L_s der Soundbar	74
Abbildung 28: Gefilterte ETC des Left Surround-Kanals der Soundbar.....	76
Abbildung 29: ETCs des L_s -Kanals der Soundbar mit reflektivem vs. absorbierendem Reflexionspunkt	78
Abbildung 30: ETC des Left Top Rear-Kanals der Soundbar	80
Abbildung 31: ETCs des Left Top Front-Kanals von diskretem System und Soundbar	82
Abbildung 32: Klarheitsmaß C_{80} für die Top Front-Kanäle von diskretem System und Soundbar	83
Abbildung 33: Virtual Height Filter	85
Abbildung 34: Ausschnitt der Amplitudenfrequenzgänge der Top Surround-Kanäle der Soundbar ...	86

Tabellenverzeichnis

Tabelle 1: Überblick über Lautsprecherpositionen für Dolby Atmos 7.1.4.....	28
Tabelle 2: $IACC_{Early}$ von Soundbar und diskretem System für die Wiedergabe der Surround-Kanäle L_s und R_s	79
Tabelle 3: $IACC_{Early}$ von Soundbar und diskretem System für die Wiedergabe der Kanäle L_{tf} und R_{tf} 81	
Tabelle 4: $IACC_{Early}$ von Soundbar und diskretem System für simultane Wiedergabe über die Kanäle L_s und R_s	91
Tabelle 5: $IACC_{Late}$ von Soundbar und diskretem System für simultane Wiedergabe über die Kanäle L_s , R_s , L_{tf} und R_{tf}	97

Abkürzungsverzeichnis

ADM	Audio Definition Model
A-JOC	Advanced Joint Object Coding
ASW	Apparent Source Width
AVR	Audio-/Video-Receiver
BWF	Broadcast Wave Format
C	Center
CTC	crosstalk cancellation
DAMF	Dolby Atmos Master Fileset
DAR	Dolby Atmos Renderer
DAW	Digital Audio Workstation
DCP	Digital Cinema Package
DD+	Dolby Digital Plus
DD+JOC	Dolby Digital Plus Joint Object Coding
DFT	Diffus Field Transfer Function
DSP	Digital Signal Processor
eARC	enhanced Audio Return Channel
EBU	European Broadcasting Union
ETC	Energy Time Curve
GUI	graphical user interface
HDMI	High-Definition Multimedia Interface
HRTF	head-related transfer function
IAB	Immersive Audio Bitstream
IACC	interaural cross-correlation
IC	interaural coherence
ICC	inter-channel coherence
ICLD	inter-channel level differences
ICTD	inter-channel time differences
ILD	interaural level differences
IMF	Interoperable Master Format
ITD	interaural time differences
ITU	International Telecommunication Union
JOC	Joint Object Coding
L	Left
LEV	Listener envelopment
LFE	Low Frequency Effects

Lrs	Left rear surround
Ls	Left surround
Ltf	Left top front
Ltr	Left top rear
LTC	Linear Time Code
Lw	Left wide
MADI	Multichannel Audio Digital Interface
NGA	Next Generation Audio
OAR	Object Audio Renderer
PCM	Pulse Code Modulation
R	Right
REW	Room EQ Wizard
Rrs	Right rear surround
Rs	Right surround
Rtf	Right top front
Rtr	Right top rear
Rw	Right wide
SMPTE	Society of Motion Picture and Television Engineers
VBAP	Vector Base Amplitude Panning

Gender-Hinweis

Die in dieser Arbeit verwendeten Personenbezeichnungen in Form des männlichen Maskulinums adressieren gleichermaßen weibliche, männliche und diverse Personen. Auf eine Mehrfachnennung wird zugunsten einer besseren Lesbarkeit verzichtet.

1. Einleitung

Immersives Audio und 3D-Audio haben in den letzten Jahren insbesondere durch die Vorstellung und Verbreitung von Next Generation Audio-Formaten an Bedeutung gewonnen. Diese versprechen nicht nur eine intuitive Integration in bestehende Produktionsprozesse bei geringem Mehraufwand, sondern gewährleisten durch ihre Agnostik gegenüber der verwendeten Wiedergabe-Konfiguration auch den Transfer auf eine Vielzahl verschiedener Wiedergabesysteme, was ein Ende für die Notwendigkeit deziderter Mischungen für verschiedene Wiedergabesysteme impliziert.

Die Soundbar stellt ein weit verbreitetes Endgerät für den Konsumenten dar, welches durch den kompakten Formfaktor, eine einfache Installation und Bedienung, sowie seine Ästhetik zu überzeugen vermag und darüber hinaus von vielen Herstellern als vollwertige Lösung zur Wiedergabe von immersivem Audio vermarktet wird. Die besondere Funktionsweise und die undurchsichtige Signalverarbeitung innerhalb des Geräts macht sie jedoch für viele Tonschaffende auf den ersten Blick zu einer „Blackbox“, was die Frage aufwirft, ob der Transfer auf dieses spezielle Wiedergabesystem „beiläufig“ gelingen kann.

Die vorliegende Arbeit versucht dieser Frage nachzugehen, indem die Signalkette zwischen Postproduktion und Wiedergabe beim Endnutzer untersucht wird, um so Variablen zu ermitteln, die einen Einfluss auf die Soundbar-Wiedergabe zulassen. Zudem sollen wesentliche Unterschiede zur Wiedergabe über diskrete Mehrkanalsysteme ermittelt werden, um so antizipieren zu können, wie sich eine in der Tonregie über ein diskretes Mehrkanalsystem abgehörte Mischung auf die Soundbar-Wiedergabe beim Endnutzer übersetzt. Untersuchungen dieser Arbeit beziehen sich dabei beispielhaft auf den Kontext von Dolby Atmos-Produktionen, welche nach Meinung des Autors eine besondere Relevanz im Hinblick auf die Endnutzer-bezogene Wiedergabe haben. Dolby Atmos ist insbesondere aufgrund des offensiven Marketings, eindrucklichen Nutzererfahrungen im Kino und der weiten Verbreitung über Streamingdienste wie z.B. Netflix, Amazon Prime und Apple Music von hoher Bedeutung für den Markt. Ein qualitativ überzeugender Transfer von Dolby Atmos-Inhalten in die Wohnzimmer des Endnutzers hat daher womöglich eine nicht zu unterschätzende Auswirkung auf die Nachfrage und den langfristigen Erfolg von immersiven Audioproduktionen im Allgemeinen.

Die Arbeit richtet sich an Tonschaffende und zielt darauf ab ein tieferes Verständnis für die Soundbar-spezifische Wiedergabe zu vermitteln. Sie soll ihnen ein Rahmenwerk bieten, welches das Treffen von kritischen, produktionsbezogenen Entscheidungen erleichtert, um die kreative Intention auch auf diesem Wiedergabesystem transportieren zu können.

Die zentrale Forschungsfrage lautet hierbei: „Wie lassen sich die besonderen Eigenschaften des Wiedergabesystems Soundbar bereits bei der Produktion von Dolby Atmos-Inhalten berücksichtigen, um die Kompatibilität bei der Wiedergabe über Soundbars im Vergleich zur diskreten Lautsprecherwiedergabe zu optimieren?“.

2. Grundlagen

2.1 Psychoakustik und räumliches Hören

Die Fähigkeit des räumlichen Hörens ist im Bewusstsein des Menschen fest verankert. Blauert (1996) und Blauert & Braasch (2008) beschreiben sie als die Fähigkeit primär anhand der beiden Ohrsignale Rückschlüsse über die Eigenschaften des uns umgebenden dreidimensionalen Raumes und darin stattfindender Schallereignisse zu ziehen. Die physikalische Grundlage dieser Fähigkeit liegt in den Eigenschaften des Hörapparates, welcher als bewegliche Antenne mit richtungs-, entfernungs- und frequenzabhängiger Richtcharakteristik beschrieben werden kann. In Konsequenz sind die durch diese Richtcharakteristik den beiden Ohrsignalen aufgeprägten linearen Verzerrungen dabei spezifisch für den Ort der Schallquellen relativ zu den Ohren, wodurch wahrnehmungsseitig Rückschlüsse über deren räumliche Position und Ausdehnung im Raum ermöglicht werden. Diese Eigenschaften lassen sich in Form von Außenohr-Übertragungsfunktionen (eng. head-related transfer function, HRTF) mathematisch beschreiben und messtechnisch erfassen (Blauert, 1996; Blauert & Braasch, 2008).

2.1.1 Hörereignis

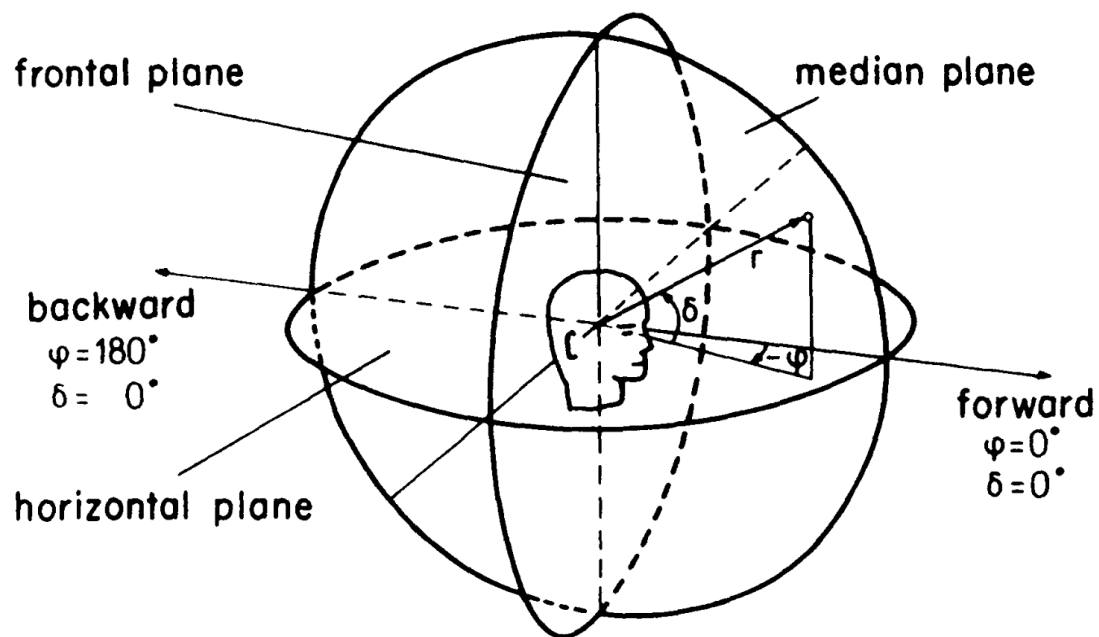
Da die Rückschlüsse wahrnehmungsbedingt nicht zwingend mit den messbaren Eigenschaften der physischen Schallereignisse übereinstimmen, verwendet diese Arbeit in Anlehnung an Blauert (1996) zur Differenzierung den Begriff des Hörereignisses. Ein Hörereignis ist das Konstrukt auditiver Wahrnehmung. Es kann durch physische Schallereignisse initiiert werden und lässt sich in diesem Fall als deren wahrnehmungsabhängige Interpretation verstehen (Blauert, 1996).

2.1.2 Kopfbezogenes Koordinatensystem

Bei der Beschreibung von Hörereignisrichtungen bezieht sich diese Arbeit auf das u.a. von Blauert (1996) verwendete, kopfbezogene Koordinatensystem (siehe Abb. 1), dessen Ursprung mittig zwischen den oberen Begrenzungen der Eingänge beider Ohrkanäle liegt. Gemeinsam mit dem unteren Rand der beiden Augenhöhlen spannen diese die Horizontalebene auf. Rechtwinklig zu dieser liegt die Frontalebene, welche ebenfalls die oberen Begrenzungen der Ohrkanäle und den Ursprung schneidet. Die Frontalebene liegt orthogonal zu den beiden anderen Ebenen und schneidet die Horizontalebene in den Azimuthwinkeln 0° und 180° , wobei 0° in Bezug auf die Blickrichtung des Kopfes „vorne“ angibt und 180° für „hinter dem Kopf“ steht. Alle drei Ebenen schneiden sich dabei im Koordinatenursprung (Blauert, 1996).

Abbildung 1:

Kopfbezogenes Koordinatensystem



Anmerkung. Visualisierung des kopfbezogenen Koordinatensystems mit Horizontalebene, Frontalebene und Medianebene, sowie den richtungsbezogenen Winkeln φ (Azimuth) und δ (Elevation). Adaptiert aus *Spatial Hearing*, von J. Blauert, 1996, S. 14. MIT Press.

2.1.3 Räumliches Hören mit einer Schallquelle

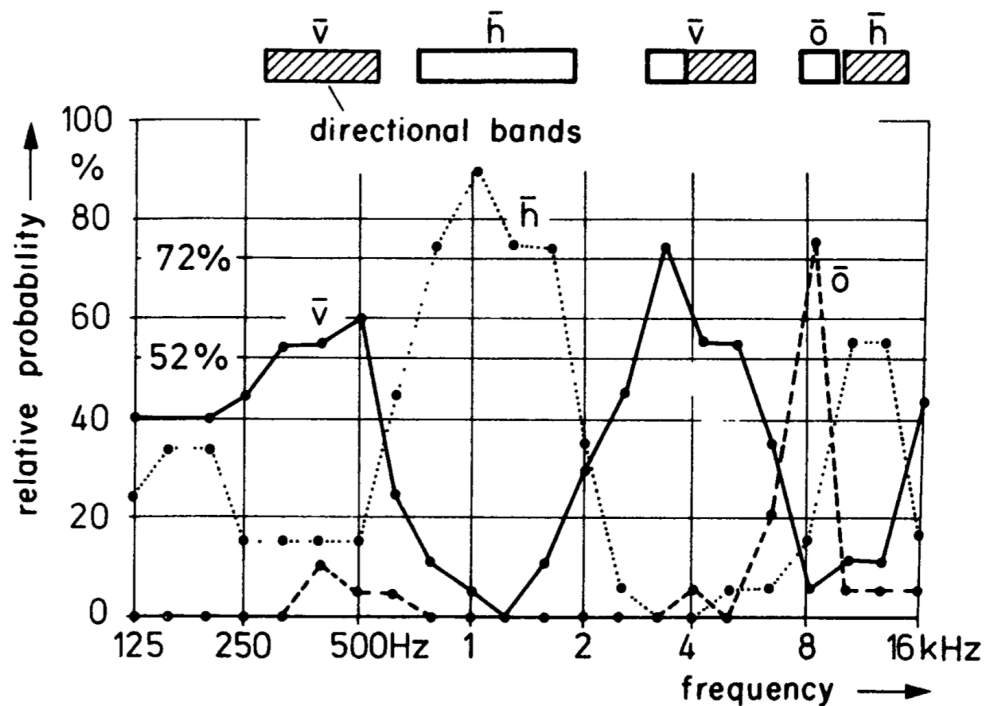
2.1.3.1 Schalleinfall aus der Medianebene

Bei Schalleinfall aus der Medianebene, d.h. aus Azimutwinkeln von 0° oder 180° , sind die beiden Ohrsignale aufgrund der hinreichenden Symmetrie des Kopfes annähernd identisch (Blauert & Braasch, 2008). Die Lokalisation von Hörereignissen beruht deshalb nach Blauert (1996) primär auf monauralen Ohrsignalmerkmalen und einer spektralen Auswertung der Ohrsignale. Maßgeblich entscheidend für die spektralen Differenzen bei unterschiedlichen Schalleinfallsrichtungen ist die Pinna, welche die Ohrsignale aufgrund ihrer komplexen Geometrie ortsspezifisch für verschiedene Schalleinfallsrichtungen linear verzerrt. Abhängig von der individuellen Außenohrgeometrie ergeben sich so verschiedene Minima und Maxima für die HRTFs, welche jeweils charakteristisch für verschiedene Schalleinfallsrichtungen aus der Medianebene sind (Blauert, 1996). Die Lokalisation von Schallereignissen in der Medianebene ist häufig von großen Lokalisationsfehlern begleitet, da die Art des Signals einen großen Einfluss auf dessen Lokalisierbarkeit hat. Blauert (1996) ermittelte über Hörversuche mit Schmalbandrauschen sog. richtungsbestimmende Bänder (siehe Abb. 2), welche dieses Phänomen veranschaulichen. So werden Signale mit einer starken Hervorhebung des Frequenzanteils um 1 kHz häufig „hinten“

lokalisiert, Signale mit einem starken Anteil von 8 kHz häufig „oben“ – unabhängig von der tatsächlichen Schalleinfallrichtung (Blauert, 1996). Je breitbandiger und je bekannter die Art des Signals, desto genauer kann es in der Regel innerhalb der Medianebene lokalisiert werden (Blauert & Braasch, 2008).

Abbildung 2:

Richtungsbestimmende Bänder



Anmerkung. Richtungsbestimmender Bänder in der Frequenzdomäne. Dargestellt sind die relativen Häufigkeiten für die Zuordnung von Terzbandrauschen verschiedener Mittenfrequenzen zu den Hörereignisrichtungen „vorne“ ($\delta = 0-60^\circ$), „oben“ ($\delta = 60-120^\circ$) und „hinten“ ($\delta = 120-180^\circ$) in der Medianebene. Adaptiert aus *Spatial Hearing* von J. Blauert, 1996, S. 110. MIT Press.

2.1.3.2 Schalleinfall aus der Horizontalebene

Weicht die Schalleinfallrichtung von den Azimuthwinkeln 0° oder 180° ab, so ergeben sich Differenzen der beiden Ohrsignale hinsichtlich Zeit bzw. Phase und Pegel bzw. Frequenz, welche maßgeblich für die Lokalisation von Hörereignissen in der Horizontalebene sind (Blauert, 1996; Blauert & Braasch; Görne, 2015; Rumsey, 2001). Die monauralen Ohrsignalmerkmale werden um binaurale Ohrsignalmerkmale ergänzt, was die Präzision der Lokalisation begünstigt (Blauert & Braasch, 2008). Man unterscheidet dabei zwischen interauralen Laufzeitdifferenzen (engl. interaural time difference, ITD) und interauralen Pegeldifferenzen (engl. interaural level difference, ILD).

2.1.3.3 Interaurale Laufzeitdifferenzen

ITD sind nach Görne (2015) in der räumlichen Distanz der beiden Ohren zueinander begründet. Dieser beträgt im Mittel ca. 17 cm mit einem maximalen Umweg um den Kopf von insgesamt ca. 21 cm, woraus unter Annahme von Normalbedingungen mit 20 °C und dem Übertragungsmedium Luft eine maximale ITD von ca. 0,6 ms bei seitlichem Schalleinfall aus einem Azimuthwinkel von $\pm 90^\circ$ resultiert (Görne, 2015). Der physiologisch relevante Wertebereich von ITD ist in der Literatur häufig mit ± 1 ms angegeben (Blauert & Braasch, 2008; Breebart & Faller, 2007). Bis zu einer maximalen Grenzfrequenz von ca. 1,6 kHz mit $\lambda \approx 0,21$ m können interaurale Phasendifferenzen der Ohrsignale zur Bestimmung der Einfallsrichtung herangezogen werden (Blauert, 1996; Görne, 2015). Oberhalb dieser Frequenz ist eine korrekte Lokalisierung von sinusartigen Signalen nicht mehr möglich, da die Phasendifferenz der Ohrsignale nun 360° überschreiten kann, was eine eindeutige Zuordnung verhindert (Blauert, 1996; Görne, 2015). Dennoch ist oberhalb von 1,6 kHz für breitbandige Signale eine Lokalisierung aufgrund von ITD möglich, entscheidend hierfür sind jedoch die Differenzen hinsichtlich der Gruppenlaufzeiten bzw. der Hüllkurven der beiden Ohrsignale (Blauert, 1996; Blauert & Braasch, 2008).

2.1.3.4 Interaurale Pegeldifferenzen

Die zweite Komponente der interauralen Ohrsignalmerkmale bilden ILD. Görne (2015) beschreibt als primäre Ursache von ILD die Dimensionierung des menschlichen Kopfes. Seitlich eintreffender Schall wird am Kopf reflektiert, sofern die Wellenlänge geringer als der im Mittel 17 cm betragende Kopfdurchmesser ist. Am dem Schallereignis zugewandten Ohr entsteht dabei ein Druckstau, am abgewandten Ohr ein Schallschatten, was in ILD resultiert. Als Grenzfrequenz für das Auftreten von ILD kann dabei aufgrund des zuvor erwähnten Kopfumfanges eine Frequenz von ca. 2 kHz mit $\lambda \approx 0,17$ m angenommen werden (Görne, 2015). Nach Dickreiter & Goeres-Petri (2014) kann der Mensch jedoch praktisch im gesamten hörbaren Spektrum ILD auswerten, unterhalb von 300 Hz treten diese jedoch unter natürlichen Bedingungen aufgrund von Beugungseffekten um den Kopf nicht mehr auf. ILD treten jedoch nicht nur in Form absoluter Pegelunterschiede zwischen den Ohrsignalen auf, sondern äußern sich bei breitbandigen Signalen auch in Form stark ausgeprägter spektraler Differenzen. Diese sind wiederum spezifisch für bestimmte Schalleinfallrichtungen aus der Horizontalebene und können über verinnerlichte Lernprozesse zur Lokalisation von Hörereignissen genutzt werden (Dickreiter & Goeres-Petri, 2014).

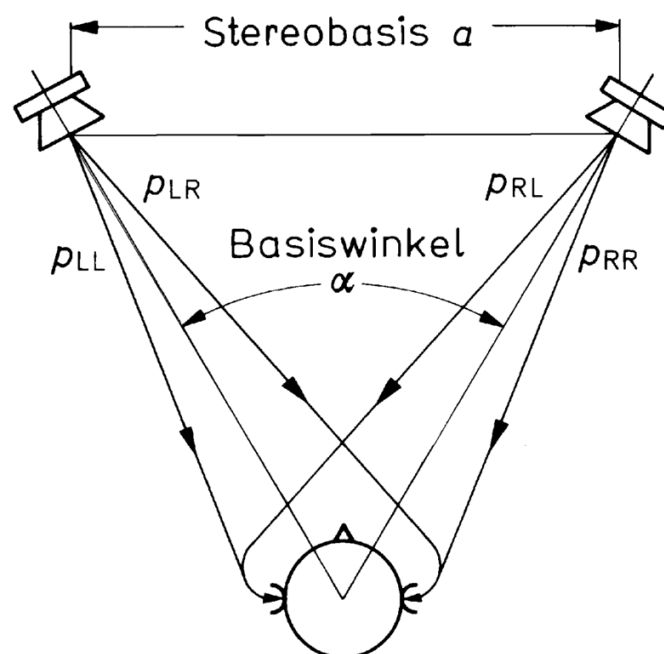
2.1.4 Räumliches Hören mit mehreren Schallquellen

2.1.4.1 Summenlokalisierung und Phantomschallquellen

Der Begriff der Summenlokalisierung beschreibt ein psychoakustisches Phänomen, welches sich anschaulich anhand der stereofonen Tonwiedergabe erläutern lässt (Blauert, 1996; Blauert & Braasch, 2008). Gegeben sind dabei zwei äquidistant vom Hörer platzierte Lautsprecher mit einem Öffnungswinkel von 60° (siehe Abb. 3).

Abbildung 3:

Stereodreieck



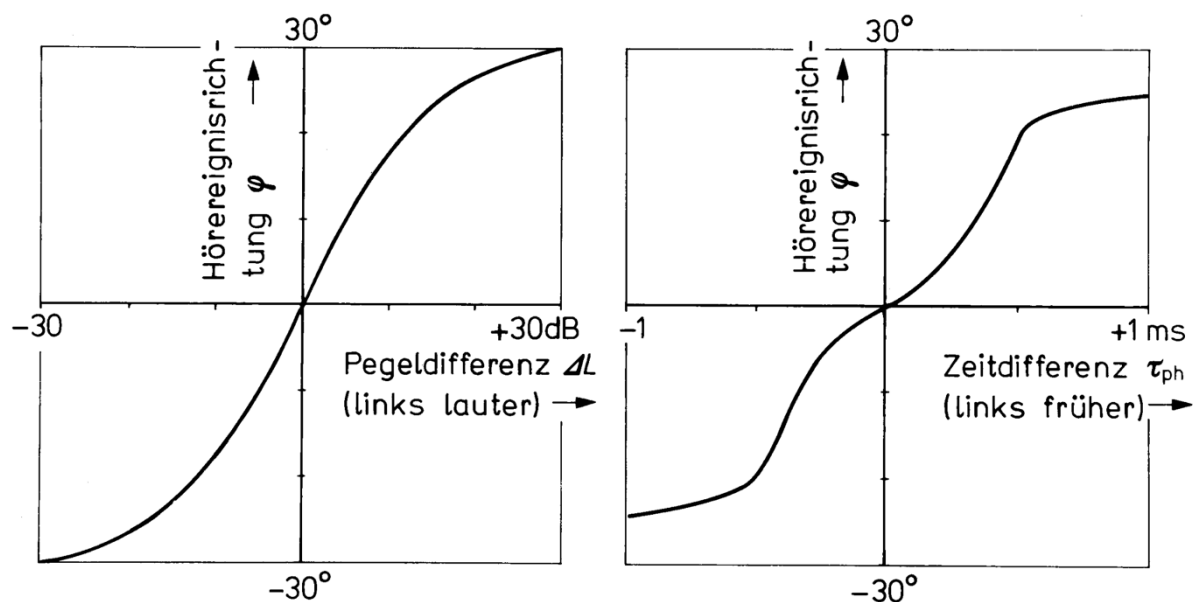
Anmerkung. Standard-Anordnung von zwei Lautsprechern im gleichseitigen Stereodreieck mit Basiswinkel $\alpha = 60^\circ$ und Stereobasis a . Eingezeichnet sind zudem die akustischen Pfade zum linken Gehörgang p_{LL} und p_{RL} , sowie zum rechten Gehörgang p_{LR} und p_{RR} . Adaptiert aus Räumliches Hören, von J. Blauert & J. Braasch, 2008, S. 101, in S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.). Springer.

Spielt man über beide Lautsprecher identische Signale ab, so werden vom Hörer nicht zwei voneinander unabhängige Schallereignisse an den Positionen der Lautsprecher lokalisiert. Sind die Differenzen der Lautsprechersignale hinsichtlich Laufzeit und Pegel vernachlässigbar gering, kommt es zu einer Hörereignisbildung mittig auf der Verbindungsline zwischen den Lautsprechern (Blauert, 1996; Breebart & Faller, 2007). An der Position des Hörereignisses ist keine reale Schallquelle vorhanden, weshalb man zur Beschreibung derartiger Hörereignisse den Begriff der Phantomschallquelle verwendet (Görne,

2015; Theile et al., 2014). Die Einführung von Zeit- und/oder Pegeldifferenzen zwischen beiden Lautsprechersignalen (engl. inter-channel time differences, ICTD und inter-channel level differences, ICLD) wirkt sich in Form von ITD und ILD auf die Ohrsignale des Hörers aus (Breebart & Faller, 2007; Lee & Rumsey, 2013), bewirkt so eine Verschiebung dieser Phantomschallquelle auf der Stereobasis zwischen den beiden Lautsprechern und resultiert in einer Lokalisation des Hörereignisses in Richtung des früher eintreffenden bzw. lauterer Signals (Blauert, 1996; Breebart & Faller, 2007; Theile et al., 2014). Dieser Zusammenhang zwischen Hörereignisauslenkung und ICTD bzw. ICLD lässt sich in Form von Summenlokalisationskurven darstellen (siehe Abb. 4).

Abbildung 4:

Summenlokalisationskurven



Anmerkung. Summenlokalisationskurven für die Wiedergabe von Breitbandsignalen über zwei Lautsprecher in stereofoner Anordnung vor dem Hörer. Dargestellt ist die Hörereignisauslenkung aus $\varphi=0^\circ$ für Pegeldifferenzen ΔL im Bereich $\pm 30\text{ dB}$ und Zeitdifferenzen τ_{ph} im Bereich $\pm 1\text{ ms}$ zwischen den beiden Lautsprechersignalen. Adaptiert aus Räumliches Hören, von J. Blauert & J. Braasch, 2008, S. 101, in *Handbuch der Audiotechnik* (1. Aufl.), von S. Weinzierl (Hrsg.). Springer.

Es ist wichtig zu betonen, dass ICTD und ICLD als weitgehend voneinander unabhängige Komponenten betrachtet werden können, die entweder komplementär oder kompensatorisch auf die Hörereignisauslenkung wirken. Eine komplementäre Wirkung liegt z.B. bei der Äquivalenzstereofonie vor, bei der die Hörereignisauslenkung in etwa zu gleichen Teilen auf gleichgerichteten ICTD und ICLD zwischen

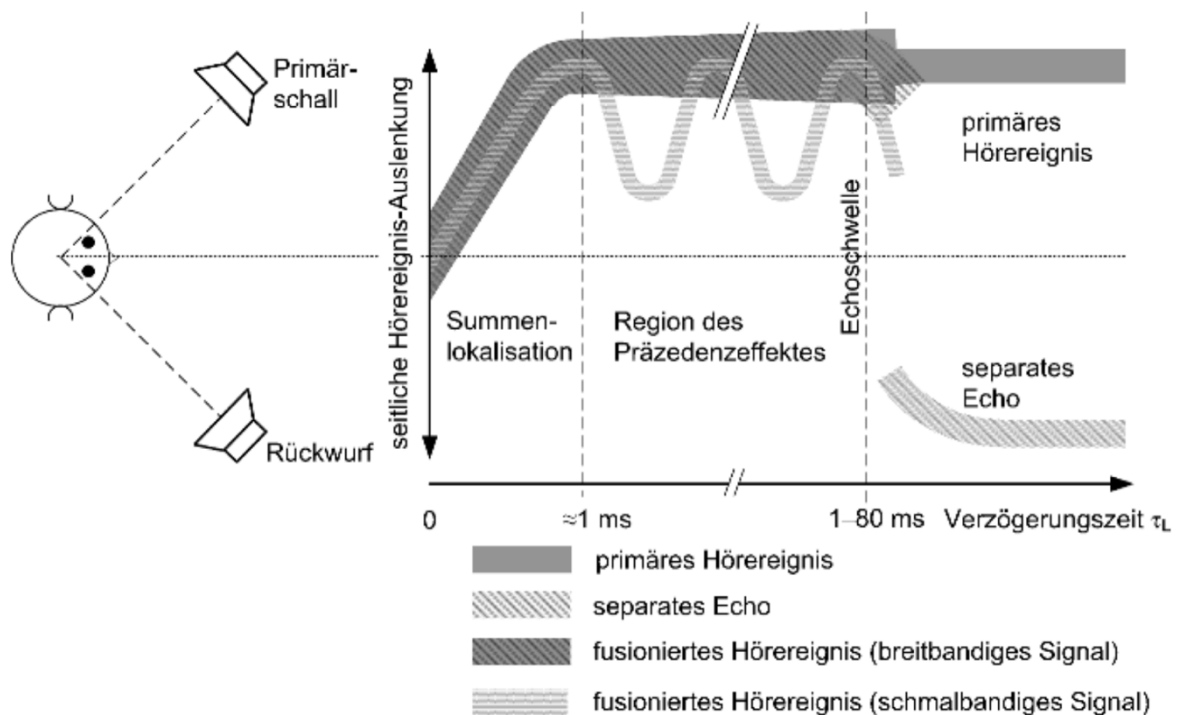
den beiden Wiedergabekanälen beruht (Görne, 2015; Theile et al., 2014). Die kompensatorische Wirkung kann hingegen genutzt werden, um eine wahlweise aufgrund von ICTD oder ICLD außerhalb der Stereomitte platzierte Phantomschallquelle durch die jeweils andere Komponente mit entgegengesetzter Wirkungsweise wieder in die entgegengesetzte Richtung auf der Stereobasis zu bewegen (sog. Trading; Blauert, 1996; Blauert & Braasch, 2008). Letzteres ist jedoch nur begrenzt möglich, da solche widersprüchlichen Merkmale in einer Reduktion der Lokalisationsschärfe resultieren können oder gar einen Zerfall des Hörereignisses bewirken, welcher sich beispielsweise in der Wahrnehmung tiefer Frequenzen in einer Richtung und hoher Frequenzen in der anderen Richtung äußern kann (Blauert & Braasch, 2008; Theile et al., 2014). Prinzipiell sind Phantomschallquellen aus anderen Einfallsrichtungen als von vorne zwar möglich (Breebart & Faller, 2007), bei seitlicher Aufstellung der Lautsprecher in den Azimuthwinkeln $\pm 60^\circ$ und $\pm 120^\circ$ - d.h. symmetrisch zur Frontalebene - verlaufen die Summenlokalisationskurven jedoch deutlich steiler und unterliegen hoher interindividueller Schwankung, sie sind daher wesentlich instabiler als frontale Phantomschallquellen (Blauert & Braasch, 2008; Theile et al., 2014).

2.1.4.2 Präzedenzeffekt und Haas-Effekt

Überschreitet die ICTD bei identischem Pegel der stereofonen Signale einen Betrag von ca. 1 ms, wird das Hörereignis an der Position des Lautsprechers lokalisiert, welcher das früher beim Hörer eintreffende Signal emittiert. Dies wird als Präzedenzeffekt bezeichnet und bewirkt, dass der Mensch innerhalb geschlossener Räume das Hörereignis verlässlich in Richtung der eigentlichen Schallquelle lokalisiert und nicht in Richtung der Wandreflexionen (Blauert & Braasch, 2008; Theile et al., 2014). Bei weiterer Zunahme der ICTD verbleibt das Hörereignis statisch an dieser Position, bis die sog. Echoschwelle erreicht ist (Pulkki & Karjalainen, 2015). Ab dieser können zwei räumlich voneinander getrennte Hörereignisse – Primärschall und Echo – wahrgenommen werden (Blauert & Braasch, 2008). Der Wert der Echoschwelle ist dabei variabel und stark von Signalform und Pegelverhältnissen abhängig (Blauert & Braasch, 2008; Görne, 2015). Eine Besonderheit des Präzedenzeffekts, der sog. Haas-Effekt, gilt im Bereich von ICTD-Beträgen um ca. 10–20 ms: In diesem Wertebereich darf der Pegel des verzögerten Signals um bis zu 10 dB höher als der des primären Schallereignisses sein, ohne dass ein Auswandern des Hörereignisses in Richtung des stärkeren Signals festgestellt werden kann (Blauert & Braasch, 2008; Görne, 2015; Rumsey, 2001). Die Übergänge zwischen den Gültigkeitsbereichen von Summenlokalisation, Präzedenzeffekt und Echo sind dabei gleitend und signalspezifisch (Blauert & Braasch, 2008; Theile et al., 2014) (siehe Abb. 5).

Abbildung 5:

Hörereignisauslenkung durch ICTD



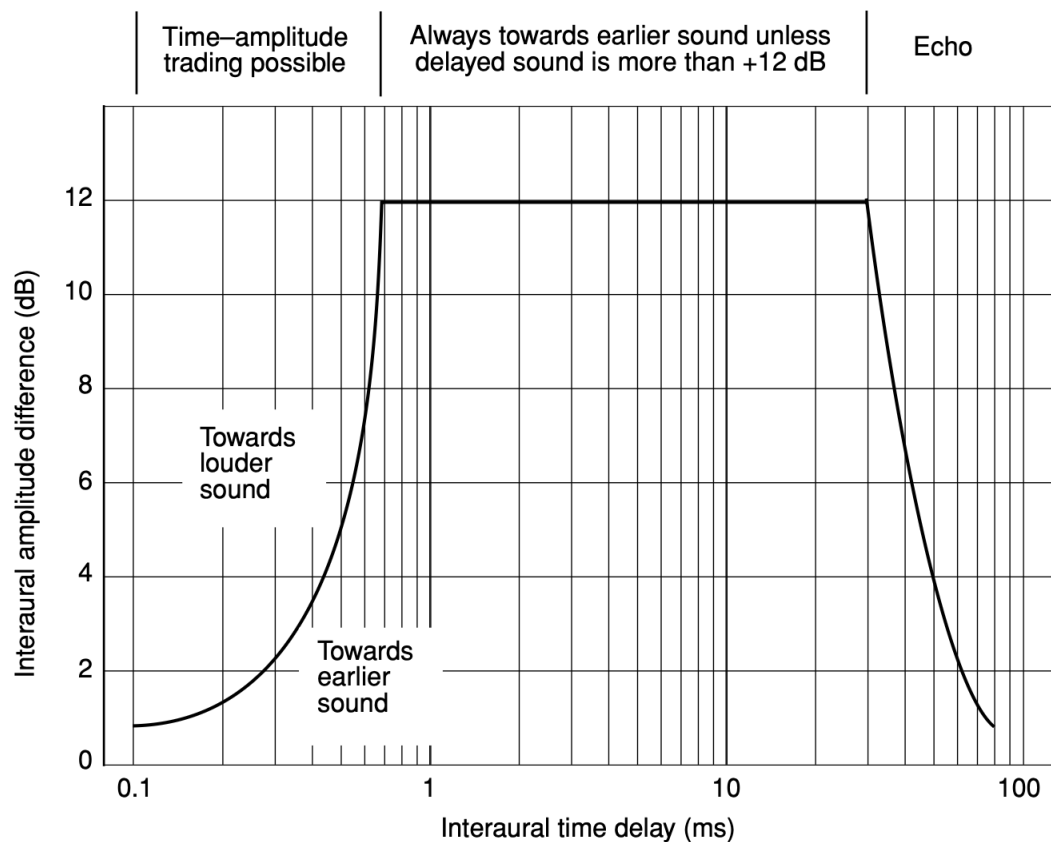
Anmerkung. Zusammenhang zwischen der Verzögerungszeit des rechten Signals einer stereofonen Lautsprecheranordnung und der seitlichen Hörereignisauslenkung zwischen $\pm 30^\circ$ Azimuth. Eingetragen sind die Gültigkeitsbereiche für die psychoakustischen Phänomene der Summenlokalisation und des Präzedenzeffektes, sowie die Echoschwelle. Adaptiert aus Räumliches Hören, von J. Blauert & J. Braasch, 2008, S. 104, in S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.). Springer.

2.1.4.3 Primärschallunterdrückung

Die obige Darstellung beschreibt die Hörereignisauslenkung via ICTD für identische Schalldruckpegel. Wie bereits erwähnt ist jedoch auch ein Trading zwischen ITD und ILD möglich. Ist der Pegel der verzögerten Schallquelle bedeutend höher als der des Primärschallereignisses, so kann dieses wahrnehmungsseitig vollständig verdeckt werden (sog. Primärschallunterdrückung bzw. Rückwärtsmaskierung; Blauert, 1996; Blauert & Braasch, 2008). Im Wirkungsbereich des Präzedenzeffektes ab einer ITD von ca. 1 ms ist dabei eine ILD von mindestens 12 dB zugunsten des verzögerten Signals notwendig, damit die Lokalisation an Position der sekundären Schallquelle erfolgen kann (siehe Abb. 6).

Abbildung 6:

ITD/ILD-Trading

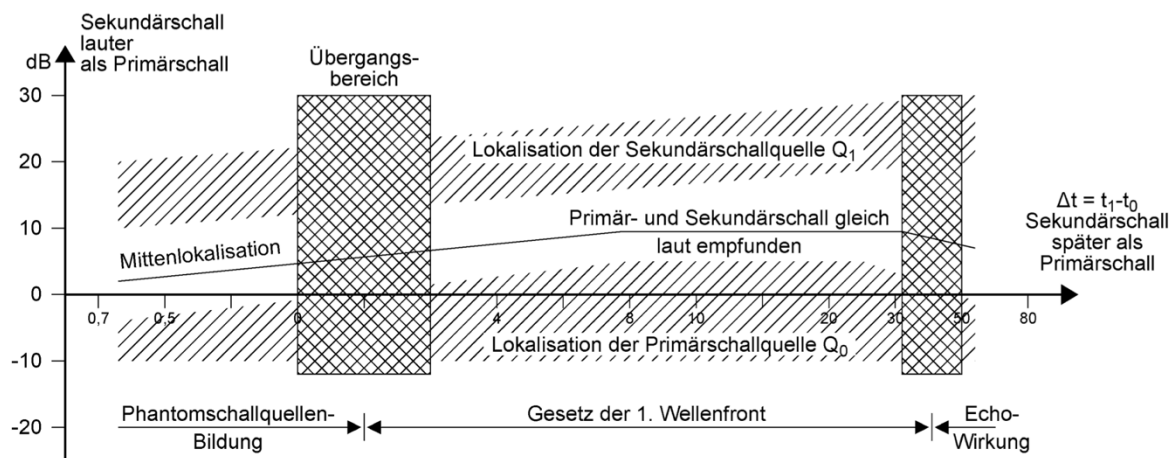


Anmerkung. Visualisierung der Beziehung zwischen kompensatorisch aufeinander wirkenden ITD und ILD. Im Wirkungsbereich des Präzedenzeffekts muss das verzögerte Schallereignis einen um mindestens 12 dB höheren Pegel aufweisen als das primäre Schallereignis, damit das wahrgenommene Hörereignis in dessen Richtung lokalisiert wird. Adaptiert aus *Spatial Hearing* (1. Aufl.), von F. Rumsey, 2001, S. 29. Routledge.

Nach Elliott (1962; 1971) wirkt die sog. Rückwärtsmaskierung bis zu einer ITD von ca. 25 ms zwischen Zielsignal (dem zu maskierenden Signal) und Maskierer, wobei eine höhere ITD eine höhere ILD für die gleiche Dämpfung des Zielsignals erfordert (Elliott, 1962; Elliott, 1971). Abbildung 7 zeigt ergänzend den Zusammenhang der Wirkungsweisen von ICLD und ICTD auf die Lokalisation des damit assoziierten Hörereignisses.

Abbildung 7:

Lokalisation bei ICTD und ICLD



Anmerkung. Darstellung der Beziehung zwischen ICTD und ICLD und deren Einfluss auf die Lokalisation von Hörereignissen. Für eine zwischen 0–30 ms verzögerte Sekundärschallquelle ist demnach ein um ca. 10–28 dB höher Schalldruckpegel erforderlich, damit das Hörereignis an Position der Sekundärschallquelle lokalisiert wird. Adaptiert aus Beschallung, von Schullan et al., 2014, S. 581, in M. Dickreiter V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudioteknik* (Bd. 1) (8. Aufl.). De Gruyter.

2.1.4.4 Interaurale Kohärenz

Für das räumliche Hören ist nach Blauert & Braasch (2008) die Ähnlichkeit der beiden Ohrsignale von großer Bedeutung, wobei als Ähnlichkeitsmaß häufig die normierte Kreuzkorrelationsfunktion der Ohrsignale (engl. interaural cross-correlation, IACC) in der Form für Leistungssignale Verwendung findet, nämlich

$$R_{norm,x,y}(\tau) = \frac{\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) \cdot y(t + \tau) dt}{x(t)_{eff} \cdot y(t)_{eff}}, \quad (1)$$

wobei $x(t)$ das rechte Ohrsignal beschreibt und $y(t)$ das linke Ohrsignal. Liegt τ im physiologisch relevanten Wertebereich für ITD zwischen ± 1 ms, so wird der Betrag der IACC als interauraler Kohärenzgrad (engl. interaural coherence, IC) bezeichnet. Dieser kann Werte zwischen 0 und 1 annehmen, wobei 1 repräsentativ für identische Ohrsignale steht und 0 eine völlige Unähnlichkeit der Ohrsignale beschreibt. Dazwischen liegende IC-Werte, eine Eigenschaft sog. teilkohärenter Ohrsignale, wirken sich in Form von diffus lokalisierbaren Hörereignissen aus, welche räumlich ausgedehnt erscheinen. Eine

solche Dekorrelation der Ohrsignale entsteht unter anderem auf natürliche Weise durch Reflexionen im Raum, welche zu unterschiedlichen Zeiten mit unterschiedlicher Intensität die Ohren erreichen (Blauert & Braasch, 2008).

2.2 Immersives Audio

Abgeleitet vom lateinischen *immersio*, für „Eintauchen“, beschreibt der Begriff der Immersion im auditiven Kontext einen Höreindruck, der beim Hörer das Gefühl auslöst, allseitig von Klang umgeben zu sein – sowohl von separat wahrnehmbaren Schallquellen als auch von ambienten, nicht scharf lokalisierbaren Hintergrundgeräuschen (Wenzel et al., 2017). Eine solche Immersion ist überzeugend, wenn man sich als Hörer in eine akustische Szene eingebettet fühlt und sich selbst als Teil dieser wahrnimmt. Audioinhalte, welche sich in einer Art und Weise repräsentieren lassen, die einen solchen Höreindruck hervorruft, können als *immersives Audio* bezeichnet werden. Schallreproduktionstechniken, welche in der Lage sind, derartige Höreindrücke zu generieren, werden nach Wenzel et al. (2017) unter dem Begriff *Spatial Audio* zusammengefasst. In diesem Kontext ist auch häufig von *3D-Audio* die Rede, da der Eindruck von auditiver Immersion sich besonders dann einstellt, wenn glaubhaft gemacht werden kann, dass die wahrgenommenen Schallquellen und deren Reflexionen ihren Ursprung in allen drei Dimensionen des den Hörer umschließenden Raumes haben (Wenzel et al., 2017).

2.2.1 Wahrnehmungsattribute immersiver Höreindrücke

Im Folgenden sollen einige Attribute definiert werden, die hinsichtlich der Beschreibung immersiver Höreindrücke verwendet werden können.

2.2.1.1 Lokalisierbarkeit

Die Lokalisierbarkeit beschreibt, wie präzise einem Schallereignis innerhalb des dreidimensionalen Raums eine Position in Bezug auf den Hörer, sowie eine Distanz zum Hörer zugeschrieben werden kann (International Telecommunication Union, ITU, 2017). Im Hinblick auf die in 2.1.4.1 erläuterte Summenlokalisation ist aber nicht nur die Lokalisation von realen Schallereignissen relevant, sondern auch die Lokalisierbarkeit von Hörereignissen – im Rahmen der Arbeit insbesondere beim Vergleich der diskreten Lautsprecherwiedergabe mit der Soundbar-Wiedergabe.

2.2.1.2 Apparent Source Width

Das Attribut der Apparent Source Width (ASW) beschreibt die wahrgenommene räumliche Ausdehnung einer lokalisierbaren Schallquelle oder eines Hörereignisses (Toole, 2008; Wenzel et al., 2017). Dieses Attribut wird häufig als positive Eigenschaft wahrgenommen (Beranek, 2004; Toole, 2008).

2.2.1.3 Listener Envelopment

Das Attribut Listener Envelopment (LEV) beschreibt die wahrgenommene auditive Umhüllung. Dieser Höreindruck stellt sich ein, wenn der Hörer das Gefühl hat, allseitig von einer auditiven Szene umgeben zu sein (ITU, 2017). Der Schall scheint dann diffus aus allen Richtungen des dreidimensionalen Raums einzufallen und wird nicht mit einzelnen, diskret lokalisierbaren Schallquellen assoziiert (Toole, 2008). Auch dieses Attribut wird von Hörern häufig als positiv bewertet (Beranek, 2004; Toole, 2008).

2.3 Räumliche Tonwiedergabe

2.3.1 Lautsprecher-Stereofonie

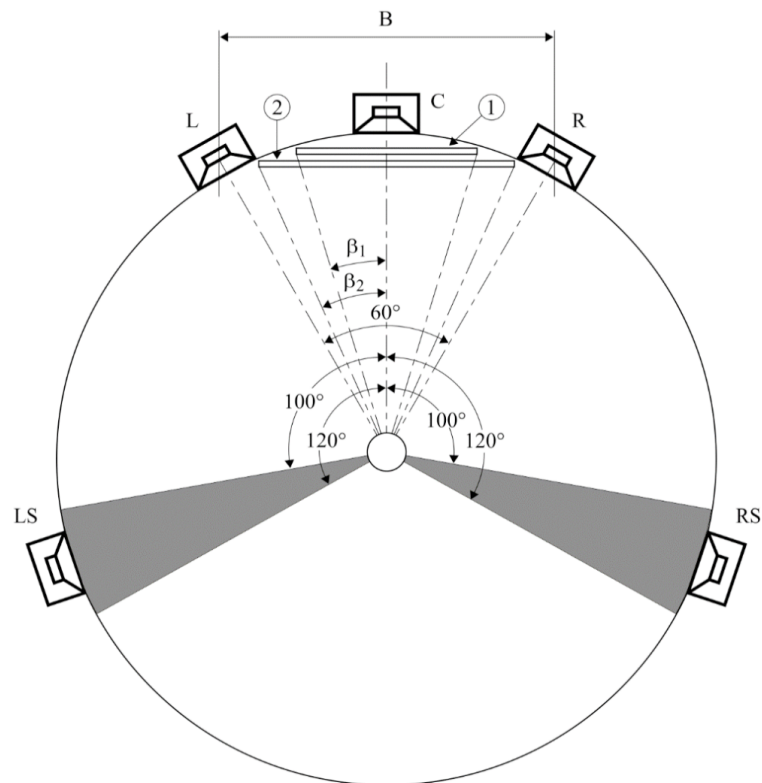
Die Lautsprecher-Stereofonie stellt ein Grundprinzip der räumlichen Tonwiedergabe dar (Theile et al., 2014), dessen Wirkungsweise primär auf der gezielten Ausnutzung psychoakustischer Phänomene, allen voran der eingangs beschriebenen Summenlokalisation, beruht (Blauert & Braasch, 2008; Rumsey, 2001; Theile et al., 2014). Durch die Wiedergabe (teil-)kohärenter Signale über mehrere Lautsprecher und die Überlagerung derer Schallfelder können Hörereignisse auch an Positionen realisiert werden, an welchen keine reale Schallquelle vorzufinden ist (Blauert, 1996; Theile et al., 2014). Rumsey (2001) beschreibt dies Vorgehen als Vermittlung von „glaubhafter Illusion“ (Rumsey, 2001, S. 152). Die einem solchen mittels Stereofonie erzeugten Hörereignis zugrunde liegenden Ohrsignale können sich dabei drastisch von den Ohrsignalen im nachempfundenen, realen Schallfeld unterscheiden (Theile et al., 2014). Eine nachteilige Eigenschaft der stereofonen Wiedergabe ist nach Theile et al. (2014) der begrenzte Abhörpunkt (Sweetspot). Ein Verlassen des Abhörpunktes resultiert in einer Veränderung der ICTD und ICLD der Wiedergabekanäle, sowie einer Veränderung der hörerbezogenen Winkel der Lautsprecher, was sich in Form von Balanceverschiebungen und Abbildungsverzerrungen bemerkbar machen kann. Zudem kann der wahrgenommene Abstand mittels Lautsprecher-Stereofonie erzeugter Hörereignisse nicht geringer sein als der Abstand der Lautsprecherbasis zum Hörer, was die Möglichkeiten zur Darstellung akustischer Nähe einschränkt (Theile et al., 2014).

2.3.1.1 Mehrkanal-Stereofonie

Eine Erweiterung der bekannten und in Kapitel 2.1.4.1 beschriebenen Zweikanal-Stereofonie stellt die Mehrkanal-Stereofonie dar. Hierbei wird die Anzahl diskreter Kanäle gegenüber der Zweikanal-Stereofonie erhöht. Die Referenz-Anordnung der Lautsprecher (sog. 3/2 Layout, auch Surround-Standard) ist in der Empfehlung ITU-R BS-775-4 der ITU (2022a) beschrieben und sieht eine Erweiterung des Stereodreiecks um drei, ebenfalls äquidistant zum Hörer platzierte Lautsprecher auf der Horizontalebene vor: Einen Center-Lautsprecher frontal vor dem Hörer bei 0° Azimuth und zwei sog. Surround-Lautsprecher bei $\pm 110^\circ$ Azimuth $\pm 10^\circ$ (siehe Abb. 8).

Abbildung 8:

3/2-Lautsprecher-Anordnung nach ITU-R BS.775-4



Anmerkung. 3/2 Lautsprecher-Anordnung nach ITU-R BS.775-4 mit den Front-Lautsprechern L, C und R, den Surround-Lautsprechern LS und RS, sowie den zugehörigen Azimuthwinkeln. Adaptiert aus *Recommendation ITU-R BS.775-4: Multichannel stereophonic sound system with and without accompanying picture*, von ITU, 2022a, S. 3.

Ist eine Äquidistanz der Lautsprecher zum Hörer nicht realisierbar, können Abweichungen über individuelle Verzögerung aller Ausgangskanäle auf denjenigen Ausgangskanal kompensiert werden, welcher dem am weitesten vom Hörer entfernten Lautsprecher zugewiesen ist. Die 3/2-Anordnung lässt sich optional um einen Subwoofer zur Wiedergabe eines Kanals für Low Frequency Effects (LFE) erweitern, welcher auf Frequenzen unterhalb von 120 Hz bandbegrenzt ist. Ist ein Subwoofer im System vorhanden, kann aufgrund der geringen Lokalisierbarkeit tiefer Frequenzen wiedergabeseitig Bass-Management angewendet werden, d.h. die tieffrequenten Signalanteile der fünf Hauptkanäle werden mithilfe von Frequenzweichen getrennt, summiert und gemeinsam über den Subwoofer wiedergegeben (ITU, 2022a). Hierdurch können die wiedergabeseitigen Anforderungen an die Hauptlautsprecher hinsichtlich der Wiedergabe tiefer Frequenzen reduziert werden, was in Konsequenz den Einsatz von kompakteren Lautsprechern mit geringerem Membranverschiebevolumen ermöglicht (Maier, 2008; Rumsey, 2001; Slavik & Weinzierl, 2008).

Die Winkelabschnitte zwischen benachbarten Lautsprechern auf Horizontalebene können jeweils als zueinander abgegrenzte Stereobasen betrachtet werden (Pulkki, 1997). Mittels paarweisem Amplituden-Panning können Monosignale so – theoretisch – als Hörereignis an beliebigen Positionen um den Hörer herum abgebildet werden. Die zuvor in 2.1.4.1 erläuterten Einschränkungen für seitliche Phantomschallquellen sollten hierbei jedoch bedacht werden (Rumsey, 2001). Zudem sind in der 3/2-Anordnung die hörerbezo-genen Winkel zwischen den Lautsprechern L und Ls, R und Rs (jeweils $80^\circ \pm 10^\circ$), sowie zwischen Ls und Rs ($140^\circ \pm 20^\circ$) relativ groß, eine exakte Lokalisierbarkeit von Hörereignissen zwischen diesen Lautsprechern sieht der Standard daher explizit nicht vor (ITU, 2022a).

Um dreidimensionale Hörereignisse realisieren zu können, muss eine Einbeziehung der Median- und Frontalebene erfolgen (Slavik & Weinzierl, 2008). Dies kann über zusätzliche Lautsprecher außerhalb der Horizontalebene – d.h. über und/oder unter Ohrhöhe – erreicht werden. Die Empfehlung ITU-R BS.2051-3 der ITU (2022b) stellt eine Erweiterung der ITU-R BS.775-4 dar und liefert Vorschläge für mögliche Lautsprecher-Anordnungen derartiger diskreter Wiedergabesysteme. Sie umfasst die Definition von 54 möglichen, zum Hörer äquidistanten Lautsprecher-Positionen, welche anhand ihrer Elevation in die drei Layer Upper Layer, Middle Layer und Lower Layer unterteilt werden. Lautsprecher auf Ohrhöhe sind dabei Teil des Middle Layers, Lautsprecher mit Elevation sind Teil des Upper Layers und das Lower Layer beschreibt Lautsprecher unterhalb der Ohrhöhe (ITU, 2022b).

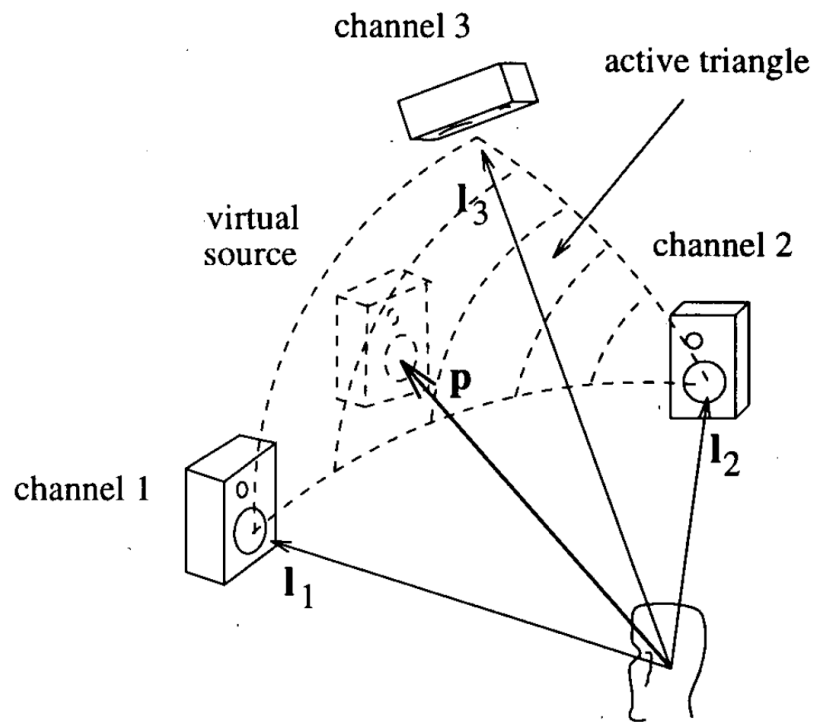
Durch Kombination der im Standard definierten Lautsprecher-Positionen lassen sich verschiedene Referenz-Lautsprecher-Anordnungen zusammenstellen (ITU, 2022b), welche unter Berücksichtigung der Toleranzbereiche eine hohe Kompatibilität mit den separaten Richtlinien proprietärer Systeme wie z.B. Dolby Atmos (Dolby Laboratories, 2021b) aufweisen. Hierdurch wird eine Interoperabilität zwischen verschiedenen kommerziellen Systemen ermöglicht. Dolby (2021b) verwendet dabei zur Bezeichnung von Lautsprecher-Layouts eine Nomenklatur in der Form x.y.z. Hierbei steht x für die Anzahl der Lautsprecher auf Horizontalebene im Middle Layer, y für die Anzahl der Subwoofer und z für Anzahl elevierter Lautsprecher im Upper Layer (Dolby Laboratories, 2021b).

Die zusätzlichen Lautsprecher mit Elevationswinkel liefern die Grundlage für die Erzeugung von Hörereignissen außerhalb der Horizontalebene. Für die freie Platzierung monofoner Signale in Form von Phantomschallquellen ist eine Erweiterung des Amplituden-Pannings notwendig, da paarweises Amplituden-Panning nur eine Hörereignisauslenkung auf der Verbindungslinie zwischen zwei Lautsprechern ermöglicht. Das von Pulkki (1997) vorgestellte Vector base amplitude panning (VBAP) löst diese Beschränkung durch die Bestimmung von Verstärkungsfaktoren für Lautsprecher-Tripel. Beim VBAP wird das verwendete Lautsprecher-Layout trianguliert und in sog. aktive Dreiecke unterteilt, deren Eckpunkte durch die Lautsprecher-Positionen gegeben sind. Das gewünschte Hörereignis und die

Positionen der verfügbaren Lautsprecher lassen sich in Form normierter Einheitsvektoren darstellen (siehe Abb. 9).

Abbildung 9:

Vector Base Amplitude Panning



Anmerkung. Visualisierung des dreidimensionalen Vector Base Amplitude Pannings (VBAP) mit dem Hörereignisvektor p und den Lautsprechervektoren l_1 , l_2 und l_3 des aktiven Dreiecks. Adaptiert aus Virtual Source Positioning Using Vector Base Amplitude Panning, von V. Pulkki, 1997, S. 460. *Journal of the Audio Engineering Society* 45(6).

Der Hörereignisvektor p lässt sich dann auch in Form einer Linearkombination der drei gewichteten Lautsprechervektoren l_1 , l_2 und l_3 beschreiben, welche das ortsspezifische aktive Dreieck bilden:

$$p = g_1 l_1 + g_2 l_2 + g_3 l_3 \quad (2)$$

Die Gewichtungsfaktoren g_1 , g_2 und g_3 definieren dabei die notwendigen Amplitudenverhältnisse zwischen den Lautsprechern, um das Hörereignis an der gewünschten Position zu realisieren.

Eine Normalisierung der Verstärkungsfaktoren in der Form

$$g_1^2 + g_2^2 + g_3^2 = C \quad (3)$$

sorgt für einen konstanten Leistungsausgleich zwischen den Kanälen und ermöglicht ein kontinuierliches Panning ohne Lautheits-Fluktuationen (Pulkki, 1997).

2.3.1.2 Diskrete Lautsprecher-Anordnungen

Im Folgenden sollen einige hinsichtlich der Fragestellung relevanten, diskreten Lautsprecher-Anordnungen vorgestellt werden. Die Richtlinien von Dolby (2021b) empfehlen für das Monitoring von primär an das Home Theater gerichteten Dolby Atmos-Inhalten eine Referenz-Anordnung in der Konfiguration 7.1.4. Diese umfasst gemäß der Dolby-Nomenklatur sieben Lautsprecher auf Horizontalebene (siehe Abb. 10), einen Subwoofer für die Wiedergabe des LFE-Kanals und vier Lautsprecher im Upper Layer (Abb. 11). Tabelle 1 zeigt einen Überblick über die enthaltenen Lautsprecher und deren hörerbezo- gene Winkel, wobei hinsichtlich der Toleranzen Abhängigkeiten zwischen den einzelnen Lautspre- chern existieren, die den Richtlinien zu entnehmen sind (Dolby Laboratories, 2021b).

Tabelle 1:

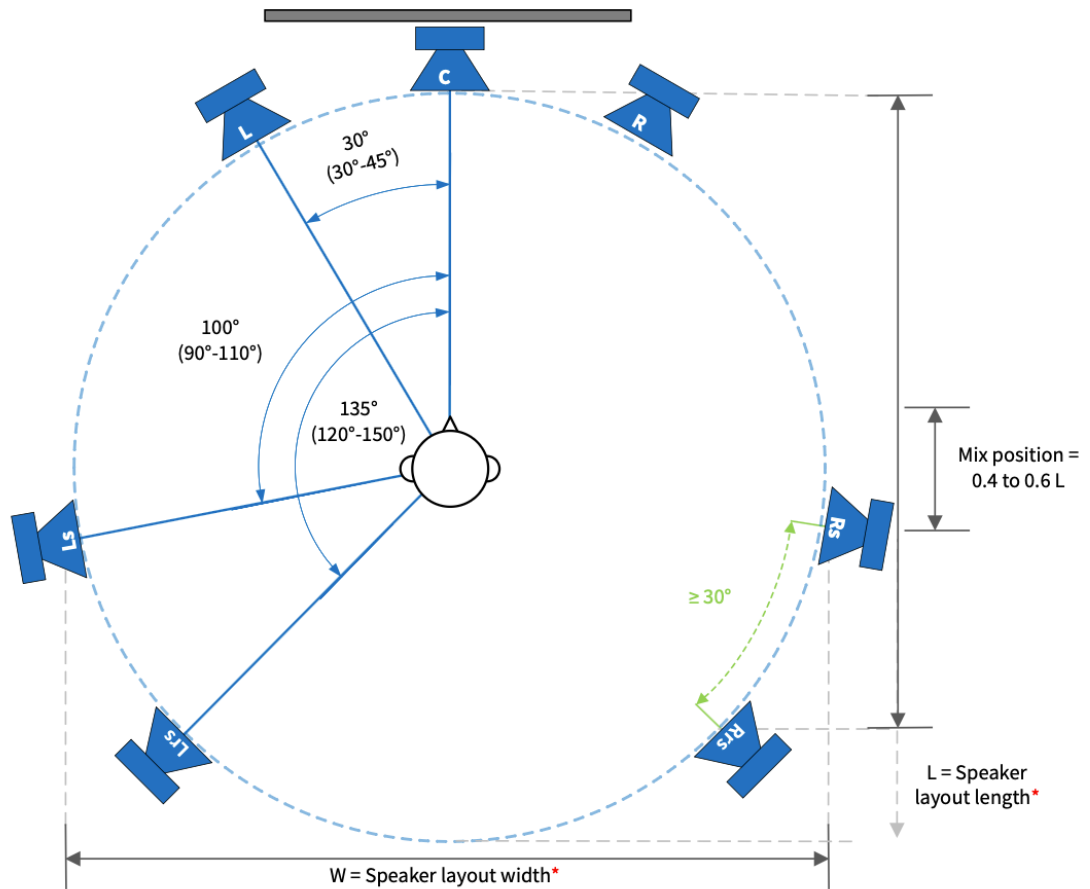
Überblick über Lautsprecherpositionen für Dolby Atmos 7.1.4

#	Name	Kürzel	° Azimuth	° Elevation
1	Left	L	30 (+15)	0 (+20)
2	Right	R	-30 (-15)	0 (+20)
3	Center	C	0	0 (+20)
4	Low frequency effects	LFE		
5	Left surround	Ls	100 (±10)	0 (+25)
6	Right surround	Rs	-100 (±10)	0 (+25)
7	Left rear surround	Lrs	135 (±15)	0 (+25)
8	Right rear surround	Rrs	-135 (±15)	0 (+25)
9	Left top front	Ltf	45 (±15)	45 (±15)
10	Right top front	Rtf	-45 (±15)	45 (±15)
11	Left top rear	Ltr	135 (±15)	45 (±15)
12	Right top rear	Rtr	135 (±15)	45 (±15)

Anmerkung. Übersicht zu Namen und Kürzeln der in einer 7.1.4-Anordnung enthaltenen Lautsprecher, sowie deren zugehörige Azimuth- und Elevationswinkel. In Anlehnung an *Dolby Atmos Home Entertainment Studio Technical Guidelines*, von Dolby Laboratories, 2021b.

Abbildung 10:

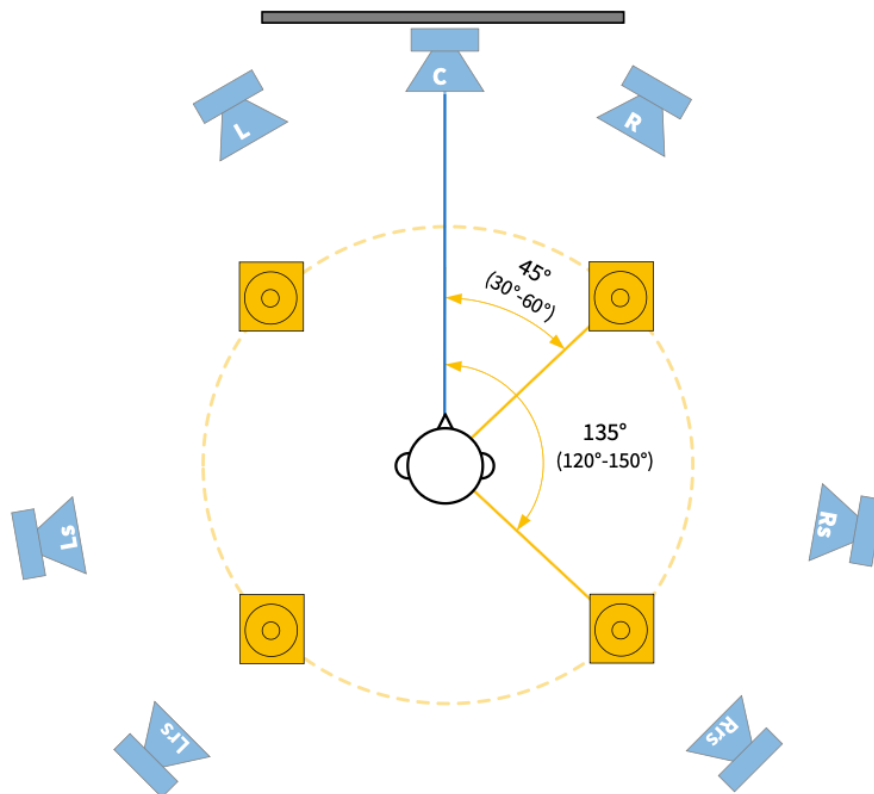
Lautsprecher im Middle Layer für diskrete Dolby Atmos 7.1.4-Anordnung



Anmerkung. Visualisierung der Lautsprecherpositionen auf Horizontalebene für die hinsichtlich des Monitorings von Dolby Atmos Home Entertainment-Inhalten empfohlene äquidistante 7.1.4-Konstellation mit den Lautsprechern L, R, C, Ls, Rs, Lrs und Rrs. Adaptiert aus *Dolby Atmos Home Entertainment Studio Technical Guidelines*, von Dolby Laboratories, 2021b, S. 12

Abbildung 11:

Lautsprecher im Upper Layer für diskrete Dolby Atmos 7.1.4-Anordnung



Anmerkung. Visualisierung der Positionen der elevierten Lautsprecher für die hinsichtlich des Monitorings von Dolby Atmos Home Entertainment-Inhalten empfohlene äquidistante 7.1.4-Konstellation mit den Lautsprechern Ltf, Rtf, Ltr und Rtr. Adaptiert aus *Dolby Atmos Home Entertainment Studio Technical Guidelines*, von Dolby Laboratories, 2021b, S. 18.

Das 7.1.4-Layout stellt eine Erweiterung des 5.1-Layouts dar. Die Lautsprecher Left (L), Right (R), Center (C), Left surround (Ls) und Right surround (Rs) werden dabei um die zwei Rear Surround-Lautsprecher Left rear surround (Lrs) und Right rear surround (Rrs), sowie um die vier Top Surround-Lautsprecher Left top front (Ltf), Right top front (Rtf), Left top rear (Ltr) und Right top rear (Rtr) ergänzt. Bei Wegfall der beiden Rear Surround-Lautsprecher wird daraus ein 5.1.4-Layout. Ein 7.1.4-Layout kann um zwei zusätzliche Lautsprecher im Middle Layer ergänzt werden, welche laut Richtlinien im äquidistanten Layout mit $+15^\circ \pm 5^\circ$ Azimuth-Abweichung von L bzw. R platziert werden sollen. Diese Lautsprecher werden als Wide Surrounds bezeichnet, spezifisch Left wide (Lw) und Right wide (Rw), und sind Bestandteil einer 9.1.4-Konfiguration (Dolby Laboratories, 2021b).

2.3.1.3 Downmix

Die ITU-R BS.775-4 (ITU, 2022a) sieht eine Abwärtskompatibilität zu kompakteren Wiedergabesystemen vor. Diese wird in Form von Downmix-Algorithmen ermöglicht. Sie beschreiben, wie die in diesem Fall fünf diskreten, für die Wiedergabe über eine 3/2-Konfiguration bestimmten Kanäle auf verschiedene Konfigurationen mit geringerer Anzahl an Ausgangskanälen überführt werden können. Hierbei werden die Signale der wiedergabeseitig nicht vorhandenen Kanäle nicht verworfen, sondern mit festen Koeffizienten beaufschlagt und auf wiedergabeseitig vorhandene Kanäle aufsummiert (ITU, 2022a). Diese Form des passiven Downmixes verhindert, dass für jedes Ausgangsformat eine zeit- und kostenintensive individuelle Mischung des Audiomaterials erstellt werden muss (Rumsey, 2017) und soll die Signale der nicht vorhandenen Kanäle so auf die verbleibenden Lautsprecher verteilen, dass ein möglichst äquivalenter Höreindruck entsteht (Lauterbach et al., 2014). Für kohärente, phasenverschobene Signalanteile zwischen den Kanälen resultiert dieses Vorgehen jedoch nach Lauterbach et al. (2014) in mitunter klangfärbenden Kammfiltereffekten. Darüber hinaus kommt die Wiedergabe der vom Downmix betroffenen Signale über räumlich abweichend positionierte Lautsprecher Abbildungsverzerrungen gleich und auch die wahrgenommene Hallbalance kann sich merklich verändern (Lauterbach et al., 2014). Die Eignung der in der Empfehlung dargelegten Koeffizienten ist nach Rumsey (2017) zudem programmabhängig und Gegenstand individueller Präferenzen, weshalb verschiedene Downmix-Algorithmen existieren, die bei der Content-Erstellung mitunter passend zum Material ausgewählt, individuell angepasst und in Form von Metadaten mit dem Audiomaterial gesendet werden können (Rumsey, 2017).

2.4. Objektbasiertes Audio

Die große Auswahl an möglichen Lautsprecher-Layouts erhöht die Möglichkeiten im Hinblick auf die Produktion, sorgt aber auch für zusätzliche Herausforderungen hinsichtlich Distribution und Wiedergabe. Die European Broadcasting Union (EBU) resümiert, dass es sich als zunehmend schwierig darstelle, die steigende Anzahl an Ausgabeformaten und die Varianz in den Lautsprecher-Layouts über eine kanalbasierte Wiedergabe zu bedienen (EBU, 2021). Bei kanalbasierten Wiedergabeverfahren ist jedes Audiosignal einem festen Wiedergabekanal zugeordnet ist, welcher räumlich einer festen Position innerhalb einer fest definierten Lautsprecher-Anordnung entspricht (ITU, 2022b). Es erfolgt somit die direkte Übertragung der Lautsprechersignale für ein spezifisches, vorab definiertes Lautsprecher-Layout, was grundsätzlich eine hohe Übereinstimmung der Lautsprecher-Anordnungen zwischen Produktion und Wiedergabe voraussetzt (Slavik & Weinzierl, 2008; Tsingos, 2017). Kompaktere Lautsprecher-Anordnungen können zwar auch in Form von Downmixes bedient werden, diese unterliegen jedoch den in 2.3.1.3 genannten Nachteilen.

Einen Lösungsansatz für diese Problematik bietet objektbasiertes Audio, welches unter Anderem in der ITU-R BS.2051-3 (ITU, 2022b) definiert ist. Die Bestandteile einer Audiomischung, d.h. die einzelnen Spuren, werden hierbei separat und nicht vorgemischt übertragen und innerhalb des Postproduktionsprozesses um Metadaten ergänzt, welche in Form von Positionsdaten Aufschluss über die intentionierte räumliche Repräsentation des jeweiligen Signals geben (EBU, 2021; ITU, 2022b; Tsingos, 2017). Diese Audioobjekte, d.h. die um Metadaten ergänzten Audiospuren, können nach der ITU-R BS.2051-3 (ITU, 2022b) auch dynamisch sein, z.B. in Form zeitlich variabler Positionsdaten zur Darstellung von Bewegung. Wiedergabeseitig ist nun eine Rendering-Einheit erforderlich, welche die den Audioobjekten zugehörigen Metadaten auswertet und anhand dieser aus den Audiosignalen unter Verwendung von Panning-Algorithmen wie z.B. VBAP die entsprechenden, für die verwendete Lautsprecher-Anordnung spezifischen Lautsprechersignale generiert (Tsingos, 2017). Die Mischung bzw. das Rendering der Audiospuren auf die Ausgangskanäle erfolgt also im Unterschied zu kanalbasierten Verfahren erst im Wiedergabesystem. Dabei wird auf die bestmögliche Repräsentation der Audiosignale abgezielt, die mit dem jeweiligen Wiedergabesystem zu erreichen ist (ITU, 2022b).

Objektbasiertes Audio gilt als Schlüsseltechnologie für den von der EBU geprägten Begriff von Next Generation Audio (NGA), da es die nutzerseitige Personalisierung von Audioinhalten ermöglicht und agnostisch gegenüber der bei der Wiedergabe verwendeten Lautsprecher-Anordnung ist, wodurch die Notwendigkeit zur Erstellung dedizierter Mischungen für verschiedene Wiedergabesysteme entfällt (EBU, 2021).

2.4.1 Dolby Atmos

Dolby Atmos ist eine eingetragene Marke der Dolby Laboratories und fungiert als Oberbegriff für verschiedene proprietäre Technologien zur Erstellung, Encodierung, Bereitstellung und Wiedergabe von immersivem Audio. Es handelt sich somit nicht um ein Audioformat im Sinne eines Codecs, denn Dolby Atmos kann über verschiedene Codecs transportiert werden (Dolby Laboratories, 2025c). Dolby Atmos wurde 2012 von den Dolby Laboratories (2012) ursprünglich als neues Format für Kineton vorgestellt. Neben der Einführung von Deckenlautsprechern umfasst Dolby Atmos dabei im Wesentlichen die Möglichkeit, die bis dahin etablierten kanalbasierten Workflows für Kineton um die Vorzüge objektbasierten Audios zu erweitern (Dolby Laboratories, 2012). Dolby Atmos wird daher häufig als hybrides System bezeichnet, da es sowohl die Nutzung von kanal- als auch von objektbasiertem Audio zulässt (Idrovo, 2025; ITU, 2014; Kim, 2017). Der Einsatz von Audioobjekten und deren Rendering auf die spezifische Lautsprecher-Anordnung vor Ort im Kino ermöglichen laut Dolby (2012) so unter anderem die gezielte Ansteuerung von bis zu 64 einzelnen Lautsprechern, welche in den kanalbasierten Verfahren zuvor nur im Array-Verbund adressierbar waren. Hierdurch können Hörereignisse wie Soundeffekte als Punktschallquelle dargestellt und für ein weiträumig im Kinosaal verteiltes Publikum konsistent und mit hoher Lokalisationsschärfe abgebildet werden. Die simultane Wiedergabe über die

Lautsprecher eines Arrays ist dabei aber weiterhin möglich, indem die entsprechenden Audioinhalte produktionsseitig in ein sogenanntes Bed geroutet werden. Ein Bed ist kanalbasiert und entspricht einem Mehrkanal-Bus mit bis zu zehn Kanälen im Format 7.1.2 (Dolby Laboratories, 2012).

Nach der Einführung im Kino folgte ab 2016 die Erschließung des Marktes für Privatanwender in Form von Dolby Atmos Home Theater (Dolby Laboratories, 2016b). Die Grundlage für diesen Transfer lieferten bereits bestehende Codecs wie Dolby TrueHD und Dolby Digital Plus (DD+), welche um Funktionalitäten zur Unterstützung objektbasierten Audios erweitert wurden – insbesondere die Integration der notwendigen Metadaten in den Bitstream ist hierbei von Bedeutung. Die Art und Weise der Codierung garantiert dabei auch eine Rückwärtskompatibilität zur Ermöglichung der Wiedergabe über rein kanalbasierte Systeme (Dolby Laboratories, 2016b). Distribution und Empfang dieser encodierten Datenströme erfolgen z.B. über Video-on-Demand-Streamingdienste wie Netflix, Disney+ oder Amazon Prime oder in Form von Bluray-Disks (Ziegler & Lerch, 2023). Wiedergabeseitig ist auch hier ein Object Audio Renderer (OAR) erforderlich, welcher die decodierten Metadaten auswertet und die zugehörigen Audiosignale anhand dieser spezifisch für die gegebene Lautsprecher-Anordnung aufbereitet (Dolby Laboratories, 2016b). Der OAR kann dabei z.B. in einen externen Audio-/Video-Receiver (AVR) integriert sein oder ist integraler Bestandteil von All-in-one-Lösungen wie Smartspeakern oder Soundbars.

Mit der Bereitstellung und Integration von Dolby Atmos in Musikstreamingdienste wie Apple Music, Amazon Music oder Tidal (Dolby Laboratories, 2019) erfolgte ab 2019 der nächste Schritt mit der Erschließung des Musikmarktes.

2.5 Soundbar

Für viele private Nutzer scheint die Idee eines umfangreichen diskreten Mehrkanalsystems zur Tonwiedergabe in den eigenen vier Wänden nicht praktikabel zu sein. Die Ursachen hierfür liegen vermutlich in einer Kombination aus dem als unkomfortabel wahrgenommenem Installations- und Einrichtungsaufwand, etwaigen ästhetisch-praktikablen Einschränkungen für Wohnräume (viele Lautsprecher und verlegte Kabel im Wohnzimmer sind häufig nicht erwünscht) und mitunter hohen Einstiegskosten (Baxter, 2022; Hooley, 2006; Walton et al., 2016; Rumsey, 2017). Kommen dennoch solche Systeme zum Einsatz, werden die Lautsprecher häufig nicht konform mit den Vorgaben platziert (Slavik & Weinzierl, 2008; Walton et al., 2016), was die damit theoretisch zu erzielende Wiedergabequalität durch Abbildungsverzerrungen beeinträchtigt (ITU, 2014; Theile et al., 2014).

Die Soundbar ist ein Consumer-orientiertes Wiedergabegerät, welches diese Akzeptanzhürde zu umgehen versucht, indem es die weiträumig verteilten Komponenten eines diskreten Wiedergabesystems in einem einzigen kompakten, vor dem Hörer angeordneten Gehäuse vereint. Dennoch werden sie häufig als Produkte vermarktet, welche imstande sind, ein vollwertiges immersives 3D-Audio-Hörerlebnis zu transportieren (Dolby, 2018b; Sennheiser, 2025a; Sonos, 2024; Walton et al., 2016). Offenbar mit

Erfolg, denn eine von der deutschen TV-Plattform in Auftrag gegebenen Studie – durchgeführt vom Marktforschungsinstitut GfK (2021) – ergab, dass Soundbars mit einem Nutzeranteil von über 58 % das beliebteste zusätzliche Gerät darstellen, wenn eine Optimierung der Tonausgabe gegenüber den integrierten Lautsprechern von Fernsehgeräten erwünscht ist. Die Verwendung mit begleitenden Bildinhalten ist für die Nutzung von Soundbars ein typisches Anwendungsfeld. Als Formfaktor für derartige Wiedergabesysteme hat sich wohl auch deshalb die Anordnung der Lautsprecher in einem flachen, horizontalen Gehäuse etabliert, welches direkt über oder unter einem Fernseher platziert werden kann (Rumsey, 2017). Dieser Formfaktor bedingt die Verwendung kompakter akustischer Treiber mit geringem Membrandurchmesser. Nach dem akustischen Modell berechnet sich die abgestrahlte Schallleistung P eines Lautsprechers zu

$$P = \frac{Z_0 r^4 \omega^4 \xi^2}{2c_0^2} = \underbrace{\frac{8\pi^2 \rho_0}{c_0}}_{\text{const.}} \cdot A^2 \xi^2 f^4, \quad (4)$$

sie steigt demnach quadratisch mit der Membranfläche A und dem Membranhub ξ , jedoch mit der vierten Potenz der Frequenz (Görne, 2015). Das für eine gegebene Schalleistung P notwendige Verschiebevolumen der Membran ist für tiefe Frequenzen also höher, weshalb Soundbars zur Optimierung der Tieftonwiedergabe häufig um einen externen Subwoofer ergänzt werden. Dieser ist nicht nur für die Wiedergabe des LFE-Kanals bestimmt, sondern übernimmt nach angewendetem Bass-Management auch die Wiedergabe der tiefen Frequenzen der übrigen Ausgangskanäle, wodurch die bauformbedingten Limitierungen der Basswiedergabe weitgehend kompensiert werden können.

2.5.1 Funktionsprinzipien

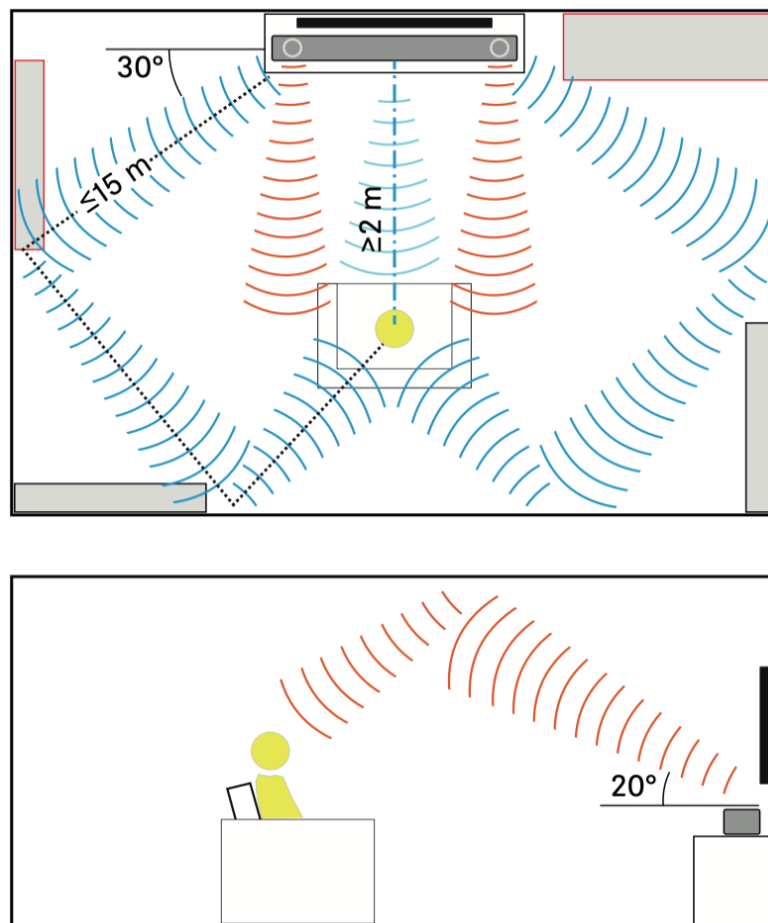
Die am Markt erhältlichen Soundbar-Modelle basieren auf zum Teil unterschiedlichen Funktionsprinzipien zur Erzeugung immersiver Hörerlebnisse, wobei die Hersteller nur selten transparent hinsichtlich der genauen Funktionsweise ihrer Produkte sind. Dieser Abschnitt soll zwei grundsätzliche Funktionsprinzipien vorstellen. Diese werden zum Teil auch kombiniert oder durch zusätzliche, herstelllerspezifische Prozessierung ergänzt.

2.5.1.1 Direktabstrahlung

Die Direktabstrahlung ist ein gängiges Funktionsprinzip von immersiven Soundbars, welches von einer Vielzahl an Modellen am Markt verwendet wird (z.B. Samsung, 2025; Sennheiser, 2025a; Sonos, 2024). Hierbei werden gezielt Raumreflexionen genutzt, um virtuelle Schallquellen an den Begrenzungsflächen des Raumes zu generieren (siehe Abb. 12).

Abbildung 12:

Schematische Darstellung einer direkt abstrahlenden Soundbar



Anmerkung. Schematische Darstellung einer Soundbar nach dem Funktionsprinzip der Direktabstrahlung. Adaptiert aus *Ambeo Soundbar SB01: Bedienungsanleitung*, von Sennheiser, 2023, S.18.

Insbesondere die Surround- und Top-Surround-Kanäle werden dabei mithilfe dedizierter, seitlich und nach oben abstrahlender Lautsprecher (sog. Side-firing und Up-firing Speaker) wiedergegeben, welche in der Soundbar verbaut sind. Sie weisen eine feste Anwinkelung im Gehäuse und eine feste Abstrahlcharakteristik auf (Dolby Laboratories, 2018b). Es handelt sich hierbei folglich ebenfalls um eine Tonwiedergabe nach dem Prinzip der Mehrkanal-Stereofonie, wobei die Wiedergabe der Ausgangskanäle zum Teil indirekt über Wand- und Deckenreflexionen erfolgt.

Die folgenden Betrachtungen innerhalb der Arbeit werden sich auf dieses Prinzip der Direktabstrahlung konzentrieren. Derartige Soundbars werden häufig in Kombination mit zusätzlichen, meist kabellosen, Rücklautsprechern vertrieben. Hierdurch ergibt sich eine Kombination aus diskreter Wiedergabe und Soundbar-Wiedergabe. Um vor dem Hintergrund der Forschungsfrage eine klare Trennung zwischen diesen Arten der Wiedergabe beizubehalten, werden solche Mischformen im Rahmen der Arbeit nicht

weiter behandelt. Eine Ausnahme sollen externe Subwoofer darstellen, da diese aufgrund der Wiedergabe ausschließlich tiefer Frequenzen nur schwer lokalisierbar sind und daher flexibler hinsichtlich ihres Standortes im Wiedergaberaum sind (Baxter, 2022; Görne, 2015; Maier, 2008).

2.5.1.2 Beamforming

Als Beamforming oder Beam-steering bezeichnet man nach Hooley (2006) und Ziomek (1995) ein auf der Verwendung von Lautsprecher-Arrays beruhendes Verfahren nach dem Funktionsprinzip phasengesteuerter Antennen. Hierbei wird ein Eingangssignal über eine Vielzahl akustischer Treiber eines Arrays wiedergegeben, wobei digitale Signalprozessoren (DSP) das Eingangssignal spezifisch für jeden Lautsprecher individuell hinsichtlich Phase und Amplitude so verändern, dass die Überlagerung derer Schallfelder und die daraus resultierenden konstruktiven und destruktiven Interferenzen zu einer Schallbündelung führen. Die individuelle Einstellung der Delays und Amplituden für jeden Lautsprecher ermöglicht es, die Richtung und das Bündelungsmaß des Schallstrahls aktiv zu beeinflussen (Hooley, 2006; Ziomek, 1995).

Dies kann einerseits genutzt werden um – ähnlich dem Prinzip der Direktabstrahlung – gezielt Raumreflexionen zu erzeugen, wobei Abstrahlcharakteristik und Richtung in diesem Verfahren aktiv gesteuert werden können (Hooley, 2006; Yamaha Corporation, 2015).

Ein anderer Anwendungsbereich ist die räumliche Übertragung binauraler Signale, die nach Slavik & Weinzierl (2008) auch als transaurale Wiedergabe bezeichnet wird. Bei räumlicher Übertragung der zwei Binaural-Signale kommt es zur Überlagerung der Schallfelder des linken und rechten Kanals. Das linke Ohr erhält also auch Anteile des für das rechte Ohr bestimmten binauralen Signals und umgekehrt. Dies führt zu einem Verlust der intentionierten räumlichen Perspektive und wird daher in der Regel durch eine Form der Übersprechdämpfung (engl. crosstalk cancellation, CTC) zu kompensieren versucht. Dies erfolgt prinzipiell durch gegenphasiges Zumischen des linken Kanals auf den rechten und umgekehrt, wobei jedoch zusätzliche Prozessierung notwendig ist, da das zu kompensierende Übersprechsignal und das kompensierende Signal zusätzlich den Auswirkungen der Übertragungsstrecke und den ortsspezifischen HRTFs des Hörers unterliegen. Aus diesem Zusammenwirken ergibt sich für die transaurale Wiedergabe auch ein sehr kleiner Sweetspot (Slavik & Weinzierl, 2008), weshalb solche Systeme simultan nur für einen Nutzer funktionieren und in diesem Zusammenhang um Headtracking ergänzt werden können (Razer Inc., 2025).

2.6 Schall im Raum

Das Funktionsprinzip direkt abstrahlender immersiver Soundbars macht deutlich, dass die Wiedergabe durch die aktive Einbeziehung des Raums zwangsläufig einem hohen Einfluss des Wiedergaberaums unterliegt. Im Folgenden sollen daher relevante Aspekte der Schallausbreitung in Räumen erläutert

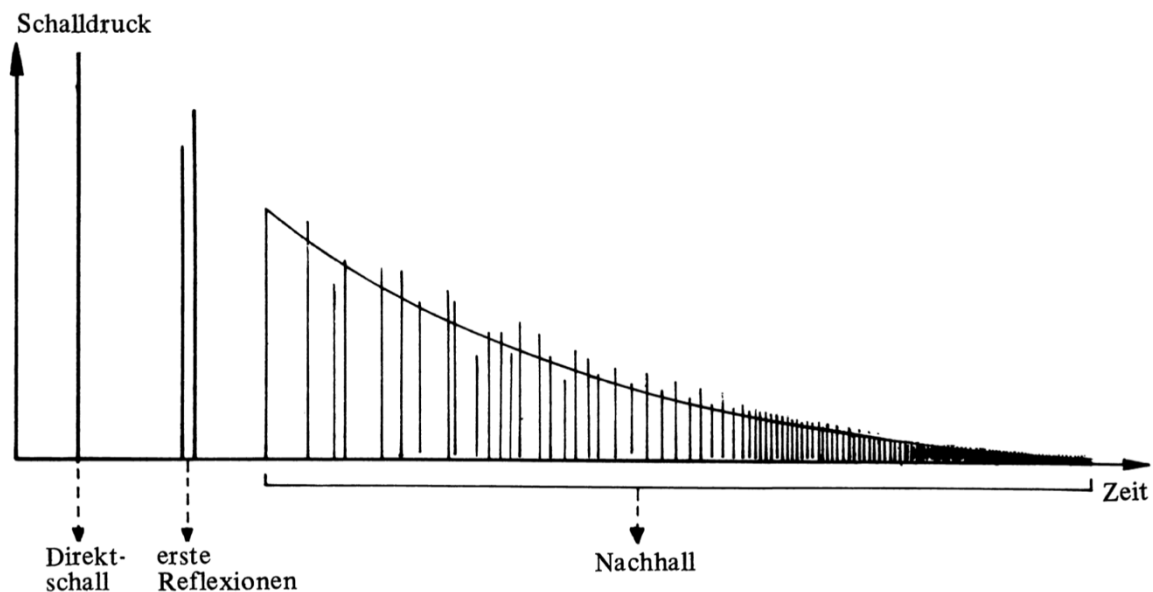
werden, um eine Grundlage für das Verständnis der Interaktion zwischen Soundbar und Wiedergaberaum zu vermitteln.

2.6.1 Zeitlicher Aufbau des Schallfeldes

Ein kugelförmig von einer Schallquelle abgestrahlter Schallimpuls wird nach Dickreiter & Hoeg (2014) an den Begrenzungsflächen des Raumes und an darin befindlichen Gegenständen reflektiert, absorbiert, gebeugt, sowie gebündelt oder zerstreut. Dies geschieht in Abhängigkeit von der Frequenz, weshalb sich die eintretenden Effekte für verschiedene Frequenzen unterscheiden. Der ursprüngliche Impuls wird durch die Einwirkung des Raumes verändert und es ergibt sich ein, für den jeweiligen Raum und die jeweilige Position innerhalb des Raumes, charakteristischer Aufbau des Schallfeldes, welcher grob in drei zeitliche Abschnitte unterteilt werden kann: Direktschall, frühe Reflexionen und diffuser Nachhall (siehe Abb. 13). Diese wirken sich unterschiedlich auf den Höreindruck aus.

Abbildung 13:

Schematisches Reflektogramm mit zeitlichem Aufbau des Schallfeldes



Anmerkung. Visualisierung einer schematischen Reflektogramms mit zeitlicher Unterteilung in die Abschnitte Direktschall, erste Reflexionen und Nachhall. Adaptiert aus Grundlagen der Akustik, von M. Dickreiter & W. Hoeg, 2014, S. 30, in M. Dickreiter V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudioteknik* (Bd. 1) (8. Aufl.). De Gruyter.

Nach Weinzierl (2008) lässt sich die akustische Übertragungsstrecke eines Raumes als System erfassen, welches einen Verlauf des Schalldrucks an Punkt A (z.B. direkt vor der Schallquelle) in einen

abweichenden Schalldruckverlauf an Punkt B (z.B. am Sweetspot eines stereofonen Wiedergabesystems) wandelt. Die Differenz der beiden Punkte ergibt sich dabei im Wesentlichen aus der laufzeitbedingten Verzögerung des Direktschalls zu Punkt B, der frequenzabhängigen Dissipation durch das Übertragungsmedium, sowie der Addition von Rückwürfen durch die Begrenzungsflächen des Raumes und dem daraus resultierenden Nachhall. Besteht Kenntnis darüber, wie ein solches System auf einen infinitesimal kurzen Impuls reagiert, kann die Raumimpulsantwort (RIR) $h(t)$ verwendet werden um die Auswirkungen des Systems auf ein beliebiges Eingangssignal $x(t)$ zu beschreiben. Das Ausgangssignal $y(t)$ berechnet sich so für den Zeitbereich über Faltung des Eingangssignals $x(t)$ mit der Raumimpulsantwort $h(t)$ zu

$$y(t) = L\{x(t)\} = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau . \quad (5)$$

Durch Multiplikation der Fouriertransformierten von Eingangssignal und Raumimpulsantwort lässt sich das Ausgangssignal auch in der Frequenzdomäne darstellen, wobei die Fouriertransformierte der Raumimpulsantwort $h(t)$ in der Form

$$H(\omega) = F\{h(t)\} \quad (6)$$

als Übertragungsfunktion des Raums bezeichnet wird (Weinzierl, 2008).

2.6.1.1 Direktschall

Der Direktschall ist derjenige Teil des Schallfeldes, welcher ohne Umwege in Form von Reflexionen, Beugung oder Streuung beim Hörer eintrifft. Da die Schallwellen den Hörer auf kürzestem Wege erreichen (Dickreiter & Hoeg, 2014), ist der Direktschall aufgrund der für alle Frequenzen konstanten Schallgeschwindigkeit meist als initialer Ausschlag in einer Raumimpulsantwort sichtbar. Die Schallenergiedichte w_d des Direktschalls nimmt unter Annahme einer idealen Kugelwelle nach

$$w_d = \frac{P}{c} \cdot \frac{1}{4\pi r^2} \quad (7)$$

mit dem Quadrat der Entfernung r von der Schallquelle ab, der Schalldruck entsprechend mit $1/r$ (Ahnert & Tennhardt, 2008). Der Primärschall hat aufgrund des in 2.1.4.2 erläuterten Präzedenzeffekts eine hohe Relevanz für die Lokalisation von Schallquellen innerhalb von Räumen.

2.6.1.2 Frühe Reflexionen

Als frühe Reflexionen werden diejenigen Reflexionen bezeichnet, die bis zu einem bestimmten Δt nach dem Eintreffen des Direktschalls beim Hörer ankommen. Dieses Δt ist nicht mit Konsens definiert und variiert, auch in Abhängigkeit vom jeweiligen betrachteten Signal, zwischen 50 ms und 100 ms, wobei für Sprache niedrigere Werte angenommen werden als für Musik (Dickreiter & Hoeg, 2014; Görne, 2015; Pulkki & Karjalainen, 2015; Rumsey, 2001). Es handelt sich dabei um Reflexionen niedrigerer Ordnung, welche im Vergleich zu späteren Reflexionen geringere Energieverluste durch Absorption, Transmission, Diffusion und den entfernungsbedingten Pegelabfall erfahren haben. Innerhalb einer RIR sind sie daher häufig in Form von zeitlich zum Direktschall verzögerten und im Schalldruckpegel reduzierten “Kopien” des initialen Impulses erkenntlich.

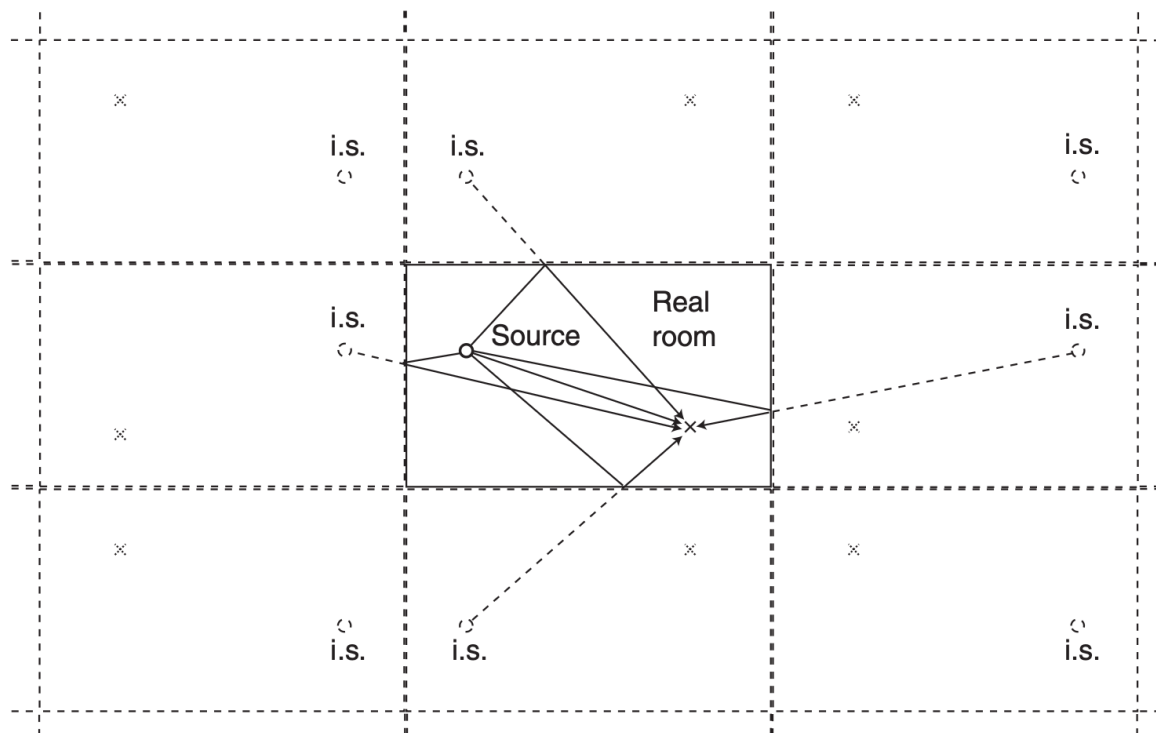
Die Reflexion von Schall ist von der Wellenlänge abhängig, eine Totalreflexion an einer Fläche mit dem Durchmesser d kommt daher nach Dickreiter & Hoeg (2014) nur bis zu einer unteren Grenzfrequenz f_u zustande, ausgedrückt durch

$$f_u = \frac{2c}{(d - \cos\alpha)^2} \cdot \frac{a_1 \cdot a_2}{a_1 + a_2}, \quad (8)$$

mit dem Schalleinfallswinkel α , der Entfernung zwischen Reflektor und Schallquelle a_1 und der Entfernung zwischen Reflektor und Hörer a_2 . Diese Grenzfrequenz liegt demnach umso tiefer, je steiler die Schallwelle auf den Reflektor trifft, je größer die reflektierende Fläche ist, je kleiner ihr Abstand zur Schallquelle ist und je kleiner ihr Abstand zum Hörer ist (Dickreiter & Hoeg, 2014). Für Frequenzen, die diese Bedingung erfüllen, kann die Reflexion analog zu optischen Reflexionen geometrisch beschrieben werden, wobei der Eintrittswinkel α identisch mit dem Austrittswinkel β ist (Ahnert & Tennhardt, 2008). Unter Berücksichtigung des Absorptionsgrades lassen sich Reflexionen so in Stärke, Richtung und Ankunftszeitpunkt akkurat vorhersagen und in Form von Spiegelschallquellen beschreiben (siehe Abb. 14) (Görne, 2015; Pulkki & Karjalainen, 2015).

Abbildung 14:

Prinzip der Spiegelschallquelle



Anmerkung. Visualisierung des Prinzips der Spiegelschallquelle. Der Eintrittswinkel einer Reflexion erster Ordnung am Abhörpunkt lässt sich geometrisch anhand dieses Prinzips vorhersagen, indem die Schallquelle an den Begrenzungsflächen des Raums gespiegelt wird und eine Verbindungslinie zum Abhörpunkt gezogen wird. Der Punkt, an welchem diese Linie die Begrenzungsfläche des Raums schneidet, ist der Reflexionspunkt. Adaptiert aus *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*, von V. Pulkki & M. Karjalainen, 2015, S. 40. Wiley.

Frühe Reflexionen erhöhen einerseits die Lautheit einer Schallquelle (Pulkki & Karjalainen, 2015), können innerhalb der ersten 20 ms einen starken Einfluss auf die wahrgenommene Klangfarbe nehmen (Rumsey, 2001), welcher im Wesentlichen auf stark ausgeprägte Kammfilterstrukturen zurückzuführen ist (Theile et al., 2014; Toole, 2008), und sie beeinflussen die räumliche Wahrnehmung der Schallquelle (Dickreiter & Hoeg, 2014; Pulkki & Karjalainen, 2015; Rumsey, 2001; Toole, 2008). So werden die Reflexionen bis $\Delta t = 80$ ms häufig mit der wahrgenommenen räumlichen Ausdehnung einer Schallquelle – der ASW – in Verbindung gebracht (Rumsey, 2001; Toole, 2008). Insbesondere die lateral auf den Hörer eintreffenden Reflexionen bestimmen diesen Eindruck (Ahnert & Tennhardt, 2008; Dickreiter & Hoeg, 2014) und die über ein Integrationsfenster zwischen $t_1 = 0$ ms und $t_2 = 80$ ms ermittelte $IACC_{\text{Early}}$ korreliert mit dem subjektiven Attribut der ASW (Ahnert & Tennhardt, 2008; Beranek, 2004).

2.6.1.3 Diffusschall

Das diffuse Schallfeld eines Raumes setzt sich aus den späten Reflexionen zusammen, welche gegenüber den frühen Reflexionen primär durch eine höhere zeitliche Dichte und niedrigere Schalldruckpegel charakterisiert sind. Die Schallenergiedichte und der Schalldruckpegel eines idealen diffusen Schallfeldes sind dabei im gesamten Raum identisch und die einzelnen Reflexionen erreichen den Hörer zu gleichen Anteilen aus allen möglichen Schalleinfallrichtungen (Dickreiter & Hoeg, 2014). Nach Abschalten einer den Raum anregenden Schallquelle fällt die Schallenergiedichte exponentiell ab, der Schalldruckpegel linear (Dickreiter & Hoeg, 2014). Zur Berechnung der Nachhallzeit T_{60} , welche ausdrückt wie viel Zeit vergeht, bis die mittlere Schallenergiedichte $w(t)$ innerhalb eines Raumes auf ein Millionstel (-60 dB) des Ausgangswertes w_0 abgeklungen ist (Ahnert & Tennhardt, 2008; Görne, 2015), wird häufig die Formel nach Eyring verwendet, in der Form

$$T_{60} = 0,163 \cdot \frac{V}{- \ln(1 - \bar{\alpha}) S_{ges} + 4mV} , \quad (9)$$

wobei V für das Raumvolumen steht, $\bar{\alpha}$ für den gemittelten Absorptionskoeffizienten der gesamten Raumbooberfläche S_{ges} und m für die Dämpfungskonstante des Übertragungsmediums Luft (Ahnert & Tennhardt, 2008; Görne, 2015).

Der diffuse Nachhall eines Raumes kann die Sprachverständlichkeit negativ beeinflussen (Pulkki & Karjalainen, 2015), trägt aber wesentlich zur Umhüllung des Hörers – der LEV – bei. Maßgeblich sind dabei die lateralen Reflexionen, welche ab $\Delta t = 80$ ms nach dem initialen Schallereignis beim Hörer eintreffen (Rumsey, 2001; Toole, 2008). Die $IACC_{Late}$ mit den Integrationszeitgrenzen $t_1 = 80$ ms und $t_2 = 500$ ms korreliert nach Beranek (2004) mit der LEV.

2.7 Wiedergabekompatibilität

Im Kontext dieser Arbeit und im Hinblick auf die Fragestellung ist der Begriff der Wiedergabekompatibilität so zu verstehen, dass die vom Tonschaffenden intentionierten auditiven Eigenschaften und Attribute immersiver Audioinhalte auch bei einem Wechsel zwischen verschiedenen Wiedergabesystemen in möglichst ähnlicher Ausprägung erhalten bleiben sollen. Bei Stereo-Produktionen wird häufig ein Erhalt der Monokompatibilität angestrebt, welche erfüllt ist, wenn das aus einer Summierung der beiden Kanäle resultierende Monosignal keine wesentliche Entwertung, z.B. in Form von Phasenauslöschungen oder Balanceverschiebungen, erfährt und die grundsätzliche Intention erhalten bleibt (Theile et al., 2014). NGA-Formate zielen auf die Bedienung eines breiten Angebots verschiedener Wiedergabesysteme mit nur einem Delivery-File ab, die Intention sollte daher unabhängig von der Form der Wiedergabe – sei es diskret über verschiedene Mehrkanalsysteme, binaural über Kopfhörer, über

Soundbars oder Smartspeaker – transportiert werden können (EBU, 2019). Im Kontext der Arbeit ist dabei der Transfer von diskreten Mehrkanalsystemen auf die Soundbar von Relevanz.

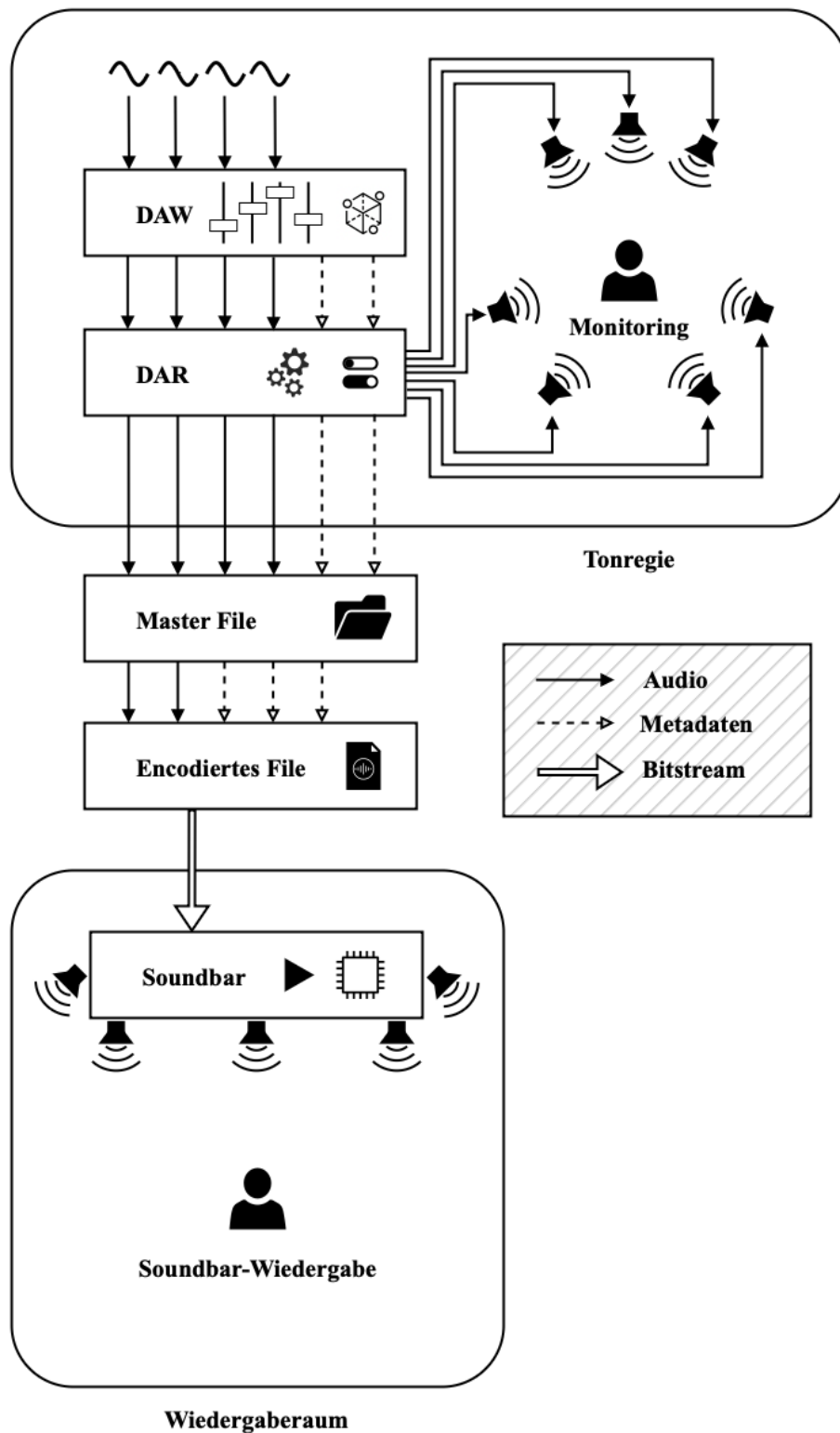
3. Signalfluss – Von der Postproduktion bis zur Soundbar-Wiedergabe

Die Signalkette zwischen Postproduktion in der Tonregie und der Wiedergabe beim Endnutzer beinhaltet eine Vielzahl von Komponenten und Signalverarbeitungsschritten, welche sich potenziell auf den Höreindruck auswirken können (Rumsey, 2001). Ein Verständnis über diese ist notwendig, um im Postproduktionsprozess Entscheidungen treffen zu können, die auf eine Optimierung der Soundbar-Wiedergabe abzielen.

In diesem Kapitel soll daher anhand einer Signalflussanalyse untersucht werden, welche Faktoren den immersiven Höreindruck von Dolby Atmos-Inhalten zwischen diskretem Monitoring im Postproduktions-Umfeld und dem nutzerseitigen Hörerlebnis bei Wiedergabe über eine Soundbar beeinflussen können. Hierbei wird einerseits aufgezeigt, welche Signalverarbeitungsschritte erforderlich sind, um Dolby Atmos-Inhalte auf einer Soundbar wiedergeben zu können. Andererseits sollen Variablen ermittelt werden, welche es Tonschaffenden ermöglichen aktiv Einfluss auf die Soundbar-Wiedergabe nehmen zu können. Die Signalflussanalyse folgt dabei einer Strukturierung, welche schematisch in Abbildung 15 visualisiert ist.

Abbildung 15:

Signalkette zwischen Postproduktion und Endnutzer



Anmerkung. Schematische Visualisierung der Signalkette zwischen Postproduktion und Endnutzer im Kontext der Soundbar-Wiedergabe von Dolby Atmos-Inhalten. Eigene Darstellung in Anlehnung an *Spatial Audio* (1. Aufl.), von F. Rumsey, S. 47. Routledge.

3.1 Postproduktion

Im ersten Abschnitt wird die Postproduktionsumgebung näher betrachtet. Dies umfasst die Art und Weise der Erstellung von Dolby Atmos-Inhalten in der Digital Audio Workstation (DAW), die verfügbaren Optionen im Dolby Atmos Renderer und das Monitoring in der Tonregie.

3.1.1 Content-Erstellung in der DAW

Die Erstellung von Dolby Atmos-Inhalten ist hinsichtlich des Workflows ähnlich zu kanalbasierten digitalen Postproduktionsverfahren. Sämtliche Editierung und Bearbeitung von Audiosignalen kann hierbei innerhalb einer DAW erfolgen. Die Möglichkeiten der individuellen Signalbearbeitung werden dabei vom Funktionsumfang der DAW und den zur Verfügung stehenden Werkzeugen, z.B. Plugins von Drittanbietern, vorgegeben. Der wesentliche Unterschied zur Produktion kanalbasierter Formate liegt in der Form des Zielroutings (Rothermich, 2021). Während in kanalbasierten Verfahren in einen vorab fest konfigurierten, kanalbasierten Summen-Bus gemischt wird (z.B. 5.1 oder Stereo), erfolgt im objektbasierten Dolby Atmos-Workflow das Routing von bis zu 128 Ausgangskanälen in die Eingänge des Dolby Atmos Renderers (DAR) (Dolby Laboratories, 2018c). Bei DAWs mit nativer Dolby Atmos-Unterstützung (z.B. Protools, Cubase/Nuendo, Logic Pro, Studio One) geschieht dies meist über ein integriertes DAR-Plugin, welches in den Main-Output der Session inseriert wird (Apple, 2025a; Steinberg, 2025). Nutzt man die dedizierte externe DAR-Anwendung, sind laut Dolby (2025a) zwei grundsätzliche Anwendungsfälle für das Routing zwischen DAW und DAR zu unterscheiden. Die erste Variante erfolgt über hardwarebasierte Verbindungen wie Dante oder MADI, für den Fall, dass der DAR auf einem dedizierten separaten Rechner läuft. Im zweiten Fall kann der externe DAR unter macOS auch auf dem gleichen Rechner wie die DAW laufen. Hierbei erfolgt das Routing zwischen DAW und DAR über einen virtuellen Audiotreiber wie z.B. die im Software-Paket enthaltene Dolby Audio Bridge. Der DAR muss in beiden Fällen mit der DAW synchronisiert werden. Dies geschieht über die Ausgabe eines Linear Time Codes (LTC) auf einer dedizierten Audiospur in der DAW, welche ebenfalls in den DAR geroutet wird (Dolby Laboratories, 2025a). Umfangreichere Ausführungen zur Einrichtung von Dolby Atmos-Sessions sind frei verfügbar und sollen deshalb nicht Teil dieser Arbeit sein. Im weiteren Verlauf der Arbeit sind sämtliche Betrachtungen auf den externen DAR in der zum Zeitpunkt der Verfassung dieser Arbeit aktuellen Version v5.3 bezogen, da dieser universell mit verschiedenen DAWs verwendbar ist und über zum Teil exklusive Funktionen verfügt (Dolby Laboratories, 2021a, Dolby Laboratories, 2025b).

Die ersten zehn Eingangskanäle des DAR sind grundsätzlich für ein kanalbasiertes Bed reserviert, standardmäßig in einer 7.1.2-Konfiguration. Es verbleiben somit 118 Audioeingänge, die sich im DAR wahlweise kanalbasiert in zusätzliche Beds routen oder Audioobjekten zuweisen lassen. Bei der Erstellung von Dolby Atmos-Inhalten stellt sich somit hinsichtlich jeder Spur des Projekts die Frage, ob diese einem Objekt zugewiesen wird, oder ob das Routing kanalbasiert in ein Bed erfolgen soll. Das Panning

von in Beds gerouteten Signalen erfolgt kanalbasiert in einem Bus identischer Breite innerhalb der DAW. Wird ein Signal in der grafischen Benutzeroberfläche (engl. graphical user interface, GUI) eines kanalbasierten Panners platziert, werden die mit den Parametern des Panners korrespondierenden Verstärkungsfaktoren umgehend angewendet, bevor die Ausgangskanäle des Busses in die dem jeweiligen Bed zugewiesenen Eingangskanäle des DAR geroutet werden.

Wird ein Track hingegen einem Objekt zugewiesen, so erfolgt das Panning nicht über eine unmittelbare Anwendung von Verstärkungsfaktoren durch den Panner. Die GUI objektbasierter Panner wie z.B. dem Dolby Atmos Music Panner gleicht zwar prinzipiell denen kanalbasierter Panner, es werden hierbei jedoch lediglich Timecode- und positionsbezogene Metadaten in Form normierter, dreidimensionaler, kartesischer Koordinaten für das jeweilige Audioobjekt generiert, welche separat an den DAR übermittelt werden. Der Panner kann daher auch auf einer anderen Spur inseriert werden, solange die Zuweisung der generierten Metadaten zur gleichen Objekt-ID erfolgt, wie das zu pannende Signal. Die vom Objekt-Panner generierten Koordinaten für X und Y können dabei Werte zwischen -1 und 1 annehmen, der Wert für Z ist auf einen Bereich zwischen 0 und 1 begrenzt, da Dolby Atmos keine Hauptlautsprecher im Lower Layer vorsieht und der Koordinatenursprung (0, 0, 0) auf den Abhörpunkt verweist. Jedes Audioobjekt verfügt darüber hinaus über einen *Object Size*-Parameter, welcher ebenfalls über den Objekt-Panner variiert werden kann und Bestandteil der generierten Metadaten ist. Wiedergabeseitig bewirkt eine Erhöhung des Object Size-Parameters laut Dolby (2018d) die zunehmende Verteilung des Signals auf angrenzende Lautsprecher des Wiedergabesystems mit vorheriger, individueller Dekorrelation der Lautsprechersignale. Für den Hörer resultiert dies in einer zunehmenden wahrgenommenen Ausdehnung des Hörereignisses (Dolby Laboratories, 2018d).

3.1.2 Dolby Atmos Renderer

Der DAR fungiert als Schnittstelle zwischen den in der DAW erstellten Inhalten und deren Monitoring in der Tonregie, indem er das Rendering der Eingangskanäle des DAR auf die gewählte Konfiguration von Ausgangskanälen ermöglicht (Dolby Laboratories, 2025b). Dabei bieten sich dem Tonschaffenden eine Reihe von Optionen.

3.1.2.1 Eingangskonfiguration

Im Dialog *Input Configuration* des DAR erfolgt die Audio-Objekt-Zuweisung der Eingangskanäle. Die ersten zehn Kanäle sind dabei wie zuvor geschildert nur einem Bed zuweisbar, die Kanäle 11 bis 128 können wahlweise weiteren Beds oder Audio-Objekten zugewiesen werden. Ein jedes Bed umfasst dabei maximal zehn Kanäle mit den möglichen Konfigurationen 7.1.2, 7.0.2, 7.1, 7.0, 5.1, 5.0, 3.0 und 2.0.

3.1.2.2 Ausgangskonfiguration

Im Dialog *Room Setup* erfolgen einige der für das Monitoring relevanten Einstellungen. Im Tab *Speaker* werden dabei aus der dort abgebildeten Übersicht von 22 Lautsprechern in der maximalen Konfiguration 11.1.10 alle diejenigen Lautsprecher ausgewählt, die in der Lautsprecher-Anordnung der Tonregie vorhanden sind. Im Tab *Routing* wird anschließend das Ausgangsrouting für diese gewählte Lautsprecher-Anordnung vorgenommen. Hierbei wird entsprechend für jeden Lautsprecher derjenige Eingang des in den Präferenzen des DAR festgelegten Ausgabegeräts ausgewählt, welcher mit den physischen Eingängen des jeweiligen Lautsprechers verbunden ist. Über den Tab *Speaker Delay* lassen sich die Ausgänge des DAR zudem individuell verzögern, wodurch etwaige Distanzunterschiede der Lautsprecher zum Abhörpunkt kompensiert werden können, um so eine virtuelle Äquidistanz zu erzielen. Der Tab *Monitoring* ermöglicht es dem Anwender individuelle Sub-Layouts des verfügbaren physischen Layouts zu erstellen. Dies geschieht durch die Abwahl einzelner Lautsprecher oder Lautsprecher-Paare. Die gewählten Sub-Layouts stehen dem Anwender dann im Hauptfenster des DAR im Pulldown-Menü *Monitoring* zur Verfügung.

3.1.2.3 Trim- und Downmix-Editor

Wichtige Optionen stehen dem Anwender im Dialog *Trim and downmix controls* zur Auswahl. Im Abschnitt *Downmix controls* wird dabei festgelegt, welche der verfügbaren Algorithmen bei einem Downmix der Dolby Atmos-Inhalte verwendet werden sollen. Die hier gewählten Einstellungen können dabei umgehend geprüft werden, indem für das Monitoring ein Sub-Layout gewählt wird, welches einen vorherigen Downmix erfordert.

Ein Downmix ist grundsätzlich für Bed-Kanäle notwendig, sofern die bei der Wiedergabe verwendete Lautsprecher-Anordnung weniger Kanäle umfasst als das jeweilige Bed. Die gewählten Einstellungen haben aber auch Auswirkungen auf Audioobjekte, insbesondere wenn wiedergabeseitig kein OAR vorhanden ist und nur ein kanalbasierter Downmix wiedergegeben werden kann. Zudem liefern sie auch bei Audioobjekten unterschiedliche Verhaltensweisen hinsichtlich des Pannings bei der Wiedergabe über verschiedene Lautsprecher-Konfigurationen.

Die Erstellung des Downmix erfolgt dabei nach Dolby (2024a) in mehreren, aufeinander folgenden Schritten. Standardmäßig erfolgt der Downmix nach dem Algorithmus *Lo/Ro*. Dabei erfolgt initial das Rendering der bis zu 128 Kanäle auf eine 7.1-Konfiguration bestehend aus den Kanälen L, R, C, LFE, Ls, Rs, Lrs und Rrs, gefolgt von einem Downmix auf 5.1, welchem sich ein Downmix auf Stereo bzw. 2.0 anschließt. Dem Stereo-Downmix geht somit grundsätzlich ein Downmix auf 5.1 voraus. Hinsichtlich der Soundbar-Wiedergabe sind insbesondere die Optionen für den 5.1-Downmix relevant. Der Anwender hat dabei die Wahl zwischen den Algorithmen *Lo/Ro*, *Dolby Pro Logic IIx*, *Direct render* und *Direct render with room balance* (Dolby Laboratories, 2024a). Bei *Lo/Ro* und *Dolby Pro Logic IIx* handelt es sich um passive Downmix-Algorithmen, welchen jeweils ein Rendering auf 7.1 voraus geht.

Für *Lo/Ro* werden dabei die folgenden Koeffizienten für den 7.1 zu 5.1-Downmix angewendet:

$$L_s = 0 \text{ dB} \times L_{ss} + 0 \text{ dB} \times L_{rs}$$

$$R_s = 0 \text{ dB} \times R_{ss} + 0 \text{ dB} \times R_{rs}$$

Für *Dolby Pro Logic IIx* gilt:

$$L_s = L_{ss} + (-1.2 \text{ dB} \times L_{rs}) + (-6.2 \text{ dB} \times R_{rs})$$

$$R_s = R_{ss} + (-6.2 \text{ dB} \times L_{rs}) + (-1.2 \text{ dB} \times R_{rs})$$

Der Unterschied der beiden Algorithmen liegt im Umgang mit Signalen auf den Rear-Surrounds einer 7.1-Anordnung. Während bei *Lo/Ro* beispielsweise die Signalanteile des L_{rs} nach dem Downmix nur auf L_s präsent sind, werden sie bei *Dolby Pro Logic IIx* durch Amplituden-Panning als Phantomschallquelle auf der Lautsprecherbasis zwischen L_s und R_s umgesetzt, wobei das Signal auf R_s einen um 5 dB niedrigeren Pegel aufweist (Dolby Laboratories, 2024a). In beiden Fällen führt die Anwendung der statischen Downmix-Koeffizienten zu Pegelsprüngen bei Pan-Automationen im hinteren Teil des Raums, da sie keinen konstanten Leistungsausgleich zwischen den vom Downmix betroffenen Kanälen gewährleisten, sowie zu Abbildungsverzerrungen in Bezug auf den Abhörpunkt.

Die beiden Optionen *Direct render* und *Direct render with room balance* sind Ansätze diese Problematiken des passiven Downmixes – wie die Bezeichnung bereits suggeriert – durch direktes Rendering der bis zu 128 Kanäle auf eine 5.1-Konfiguration zu umgehen. Dem 5.1-Output geht somit kein Rendering auf 7.1 voraus. *Direct render* sorgt dabei dafür, dass Audioobjekte, welche an den Randbereichen des Objekt-Panners platziert sind, als Phantomschallquellen auf den Basen benachbarter Lautsprecher im Middle Layer umgesetzt werden. Laut Dolby (2024a) verfolgt dies den Zweck, die intentionierte räumliche Position in Bezug auf den Abhörpunkt möglichst beizubehalten.

Folgendes Beispiel zeigt, dass das nicht immer funktioniert: Ein auf die Objekt-Koordinaten (-1, 0, 0) gepanntes Audioobjekt wird bei Wiedergabe über eine 7.1.4-Anordnung nur über den Lautsprecher L_s wiedergegeben. Bei Wiedergabe über eine diskrete 5.1-Anordnung unter Anwendung von *Direct render* erfolgt die Wiedergabe des gleichen Audioobjektes mit jeweils -3 dB (bei Trim-Settings von 0 dB) über die Lautsprecher L und L_s . Es kommt folglich auch hier zu Abbildungsverzerrungen gegenüber der diskreten Wiedergabe in 7.1.4. *Direct render* ermöglicht jedoch ein kontinuierliches Panning bei Pan-Automationen im Middle Layer ohne wahrnehmbare Pegelsprünge.

Direct render with room balance unterscheidet sich von *Direct render* hinsichtlich des Verhaltens beim Panning in der hinteren Hälfte des Raumes. Laut Dolby (2024a) zielt der Algorithmus darauf ab Kammfiltereffekte zu verhindern, welche bei der Nutzung seitlicher Phantomschallquellen entstehen können. Das im vorherigen Beispiel genannte Audioobjekt mit den Objekt-Koordinaten (-1, 0, 0) wird im 5.1-Downmix – identisch mit der Wiedergabe über 7.1.4 – nur über den Lautsprecher L_s wiedergegeben. Eine kontinuierliche Verringerung des Wertes für die Y-Koordinate bis zum Wert -1 bewirkt hier keine

Veränderung der Ausgangspegel, das Audio-Objekt wird statisch über Ls wiedergegeben. Erst in der vorderen Hälfte der GUI des Panners, bei Erhöhung der Y-Koordinate auf Werte >0 , erfolgt die Wiedergabe via Amplituden-Panning zwischen Ls und L.

Die *Trim Controls* sind eine zusätzliche Funktion, welche es dem Anwender ermöglicht Einfluss darauf zu nehmen, wie sich die Mischung bei Wiedergabe über kompaktere Lautsprecher-Anordnungen mit geringerer Anzahl an Ausgangskanälen überträgt. Die Anwendung der Trim-Parameter geht dabei dem Downmix voraus (Dolby 2024b). Der Nutzer hat hier über *Trims* die Möglichkeit für verschiedene Lautsprecher-Konfigurationen individuellen Einfluss auf die Pegelreduktion von Signalen zu nehmen, welche anteilmäßig über die Surround- und Höhen-Kanäle wiedergegeben werden. Zusätzlich kann über *Front/back balance* die Lautstärke-Balance für elevierte Signale und für Signale auf dem Middle Layer zwischen Front und Rückseite des Raums individuell angepasst werden.

Die in diesem Dialog verfügbaren Optionen sollten grundsätzlich in Abhängigkeit der jeweiligen Mischung ausgewählt werden, wobei verschiedene Mischungen unter Umständen verschiedene Downmix-Algorithmen erfordern und besonders die Wahl geeigneter Trim-Settings stark von der Art der Signale, deren Priorität in der Mischung und deren Beziehung zueinander abhängig ist (Rothermich, 2021). Gegebenenfalls muss auch beachtet werden, dass diverse Labels Vorschriften hinsichtlich der zu verwendenden Downmix-Algorithmen machen (z.B. Universal Media Group, 2025).

3.1.2.4 Monitoring-Optionen

Hinsichtlich des Monitorings hat der Anwender im Hauptfenster des DAR die Möglichkeit den Transfer der Mischung auf kompaktere Lautsprecher-Anordnungen zu überprüfen, indem im Pulldown-Menü *Monitoring* das entsprechende Sub-Layout ausgewählt wird. Die in 3.1.2.3 erläuterten Einstellungen aus dem Dialog *Trim and downmix controls* werden dabei umgehend angewendet, wodurch sich deren Auswirkungen überprüfen lassen. Zusätzlich können einzelne Ausgangskanäle solo oder stumm geschaltet werden, sowie global alle Beds und Objects. Im Dialog *Preferences* hat der Nutzer darüber hinaus die Möglichkeit Bass-Management für die verwendete Lautsprecher-Anordnung zu aktivieren, indem unter *Speaker bass management frequency* die entsprechende Crossover-Frequenz zwischen Subwoofer und den Hauptlautsprechern des verwendeten diskreten Mehrkanalsystems festgelegt wird. Darüber hinaus kann das *Output limiting* aktiviert werden, welches laut Dolby (2025a) eine Emulation des Limitings darstellt, das beim Encoding in Consumer-Formate Anwendung findet.

3.1.3 Monitoring in der Tonregie

Dem Thema der Arbeit liegt die Frage zugrunde, wie sich die im Postproduktionsprozess erstellten Audioinhalte auf das Consumer-orientierte Wiedergabesystem Soundbar übertragen. Der Vergleich zwischen diskreter Mehrkanal-Wiedergabe und Soundbar-Wiedergabe ist deshalb insbesondere für derartige diskrete Wiedergabesysteme relevant, wie sie in professionellen Tonregien zum Einsatz kommen.

Sämtliche Betrachtungen stehen daher unter der Annahme, dass das betrachtete diskrete Wiedergabesystem die Richtlinien hinsichtlich der Lautsprecher-Anordnungen einhält, wie sie in Kapitel 2.3.1.2 dargestellt wurden. Darüber hinaus wird vorausgesetzt, dass das Wiedergabesystem in einem raumakustisch optimierten Umfeld verwendet wird, das die in den etablierten Richtlinien definierten Kriterien für professionelle Abhörräume erfüllt. Diese sind im Folgenden zusammengefasst.

3.1.3.1 Anforderungen an Tonregien

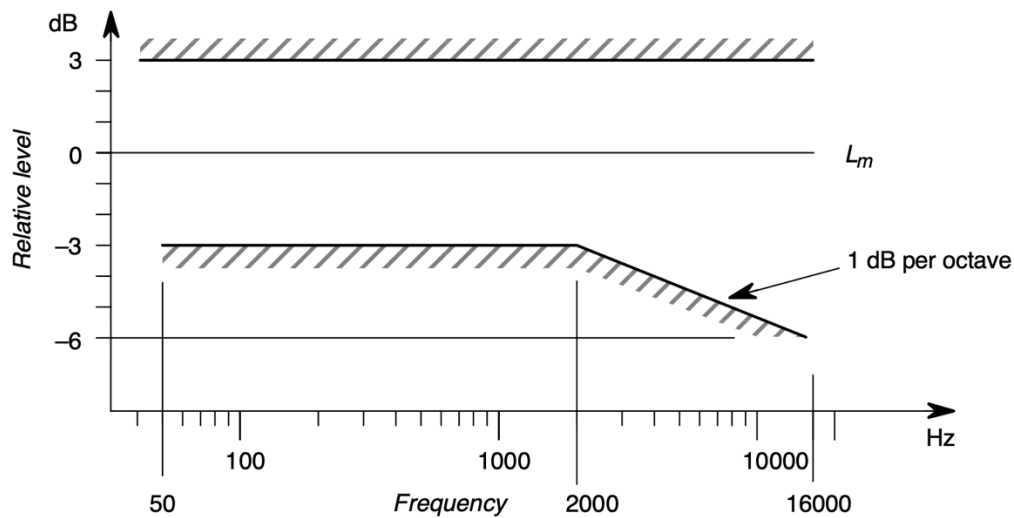
An professionelle Abhörräume werden hinsichtlich der Audiowiedergabe strenge Anforderungen gestellt, da die wechselseitige Beziehung zwischen Lautsprechern und Abhörraum einen hohen Einfluss auf die auditive Wahrnehmung der reproduzierten Inhalte hat (Maier, 2008; Toole, 2008). Da sämtliche klangästhetischen Entscheidungen des Tonschaffenden im Postproduktionsprozess auf dessen Wahrnehmung des Schallfeldes im Abhörraum beruhen, muss gewährleistet werden, dass diese nicht in einer Kompensation lokaler Eigenschaften des Raums begründet liegen (Everest & Pohlmann, 2022; Holman, 2008). Angestrebt wird eine *Neutralität* und eine *Transparenz* der Wiedergabe, d.h. der Einfluss der aus Wiedergabesystem und Raum bestehenden Einheit auf die auditive Wahrnehmung des Tonschaffenden soll so weit reduziert werden, dass eine zuverlässige Beurteilung der Tonwiedergabe ermöglicht wird (Hoeg et al., 2014; Maier, 2008).

Hierfür ist ein kontrolliertes Schallfeld notwendig, sowohl hinsichtlich des Direktschalls, früher Reflexionen, des diffusen Nachhalls, des Ruhegeräuschpegels und der Frequenzabhängigkeit des stationären Schallfeldes (Maier, 2008). Etablierte Standards wie die ITU-R BS.1116-3 und die EBU Tech. 3276 geben diesbezüglich Angaben, welche Messkriterien ein Abhörraum für die Lautsprecher-Wiedergabe zu erfüllen hat (EBU, 1998; ITU, 2015).

Der Direktschall wird weitgehend von den Eigenschaften der verwendeten Lautsprecher bestimmt. (Maier, 2008). Diese sollen nach der ITU-R BS.1116-3 (ITU, 2015) unter Freifeldbedingungen gemessen auf der 0°-Achse einen nahezu linearen Frequenzgang von maximal ± 2 dB Abweichung im Wiedergabebereich zwischen 40 Hz und 16 kHz aufweisen und ein gleichmäßiges Abstrahlverhalten garantieren, das bis $\pm 30^\circ$ horizontaler Auslenkung keine Abweichung vom Frequenzgang der 0°-Achse um mehr als ± 4 dB zulässt. Außerdem sollen sie ein direktionales Abstrahlverhalten mit einem Schallbündelungsmaß C von $4 \text{ dB} \leq C \leq 12 \text{ dB}$ im Frequenzbereich zwischen 250 Hz und 4 kHz aufweisen (ITU, 2015). Für sämtliche verwendeten Lautsprecher soll die durch die Lautsprecher-Raum-Interaktion beeinflusste Betriebsschallpegelkurve am Abhörpunkt für jedes Terzband zwischen 50 Hz und 4 kHz innerhalb eines Toleranzbereiches von ± 3 dB liegen (siehe Abb. 16), wobei zusätzlich eine hohe Übereinstimmung zwischen den einzelnen Lautsprechern des Systems mit max. 2 dB Abweichung je Terzband vorgegeben ist (ITU, 2015; EBU, 1998).

Abbildung 16:

Toleranzbereiche für die Betriebsschallpegelkurve in Referenz-Abhörräumen



Anmerkung. Darstellung der Toleranzbereiche für die Betriebsschallpegelkurve jeden Lautsprechers am Abhörpunkt in Referenz-Abhörräumen nach EBU Tech. 3276. Dargestellt sind die zulässigen Abweichungen vom durchschnittlichen Schalldruckpegel L_m , welcher über die Messung von Terzbandrauschen mit Mittenfrequenzen zwischen 200 Hz und 4 kHz ermittelt wird. Adaptiert aus *EBU Tech. 3276 - Listening conditions for the assessment of sound programme material - monophonic and two-channel stereophonic*, von EBU, 1998, S. 7.

Aufgrund der in 2.6.1.2 genannten Auswirkungen früher Reflexionen auf den Höreindruck, ist es von Bedeutung diese für eine verlässliche Beurteilung des Direktschalls ab Abhörpunkt zu unterdrücken. Die ITU-R BS.1116-3 und die EBU Tech. 3276 fordern daher, dass der gemessene Schalldruckpegel früher Reflexionen, welche in den ersten 15 ms nach dem Direktschall ab Abhörpunkt eintreffen, mindestens 10 dB unterhalb des Direktschalls liegen muss (EBU, 1998; ITU, 2015). In der Praxis wird dieses Kriterium wahlweise raumgeometrisch durch gezielte Lenkung des reflektierten Schalls vorbei am Abhörpunkt erfüllt (reflexionsfreie Zone), oder durch die Platzierung schallabsorbierender Begrenzungsflächen an den Reflexionspunkten, wobei insbesondere bei Mehrkanal-Konfigurationen auf eine homogene Verteilung der Absorptionsflächen zu achten ist (Everest & Pohlmann, 2022; Maier, 2008; Toole, 2008).

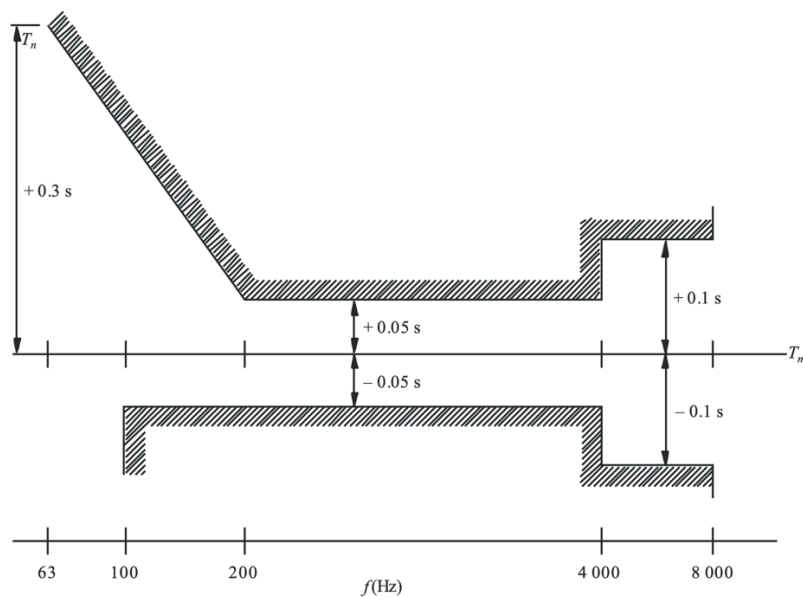
Bezüglich des Nachhalls werden Anforderungen an die frequenzabhängigen Nachhallzeiten gestellt. In Abhängigkeit vom Raumvolumen V ist der Sollwert für das arithmetische Mittel der Nachhallzeiten T_m in den Terzbändern zwischen 200 Hz und 4 kHz durch die Formel

$$T_m = 0,25 \left(\frac{V}{V_0} \right)^{\frac{1}{3}} \quad (10)$$

definiert, wobei V_0 ein Referenzvolumen von 100 m^3 beschreibt (EBU, 1998; ITU, 2015). Für den Frequenzbereich zwischen 63 Hz und 8 kHz ist zudem ein Toleranzbereich für frequenzspezifische Abweichungen der Nachhallzeiten vom Durchschnittswert T_m definiert (siehe Abb. 17). Dieser stellt sicher, dass die frequenzabhängigen Nachhallzeiten eine geringe Varianz aufweisen. Für den Frequenzbereich zwischen 200 Hz und 4 kHz ist dabei eine enge Toleranz von lediglich $\pm 50 \text{ ms}$ vorgesehen, womit die Nachhallzeit in diesem Bereich nahezu frequenzunabhängig zu sein hat (EBU, 1998; ITU, 2015).

Abbildung 17:

Toleranzbereiche für Nachhallzeiten in Referenz-Abhörräumen



Anmerkung. Darstellung der Toleranzbereiche für Abweichungen frequenzspezifischer Nachhallzeiten vom im Frequenzbereich zwischen 200 Hz und 4 kHz ermittelten Durchschnittswert T_m nach ITU-R BS.1116-3. Adaptiert aus *Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems*, von ITU, 2015, S. 14.

Es lässt sich festhalten, dass professionelle Abhörräume, die sich an den geschilderten Standards orientieren, ein hoch kontrolliertes Arbeitsumfeld darstellen, welches eine verlässliche Beurteilung der produzierten Audioinhalte ermöglicht.

3.2 Export-Formate und Codecs

Nach der Postproduktion gilt es zu betrachten, wie die erstellten Dolby Atmos-Inhalte zum Endnutzer gelangen und welche Optionen sich dabei bieten. Ausschlaggebend sind hier der Übertragungsweg, sowie die Wiedergabefähigkeiten von Soundbars.

3.2.1 Master File-Formate

Master File-Formate werden primär für den Austausch und die Archivierung von Dolby-Atmos Inhalten innerhalb der Produktionskette genutzt und sind nicht für die Wiedergabe auf Consumer-Geräten intendiert. Sie garantieren den unkomprimierten Erhalt aller bis zu 128 Kanäle einer Dolby Atmos-Mischung, sowie deren begleitende Metadaten. Eine Konvertierung zwischen verschiedenen Master File-Formaten ist mit dem Dolby Atmos Conversion Tool möglich (Dolby Laboratories, 2025c). Die gängigsten Master File-Formate sollen im Folgenden erläutert werden.

3.2.1.1 Dolby Atmos Master Fileset

Das Dolby Atmos Master Fileset (DAMF) ist das proprietäre und native Master File-Format von Dolby (Dolby Laboratories, 2025c; Dolby Laboratories, 2025d). Es besteht aus drei separaten Dateien, welche in einem gemeinsamen Ordner gesichert werden. Bei der ersten Datei mit der Endung *.atmos* handelt es sich um eine XML-Datei, welche auf die beiden anderen Dateien referenziert und globale Metadaten über den Master enthält. Dazu zählen unter anderem die im DAR festgelegten Einstellungen bezüglich Sample Rate, Frame Rate, Eingangskonfiguration, Downmix und Trims. Die Datei mit der Endung *.atmos.audio* enthält die bis zu 128 Audiokanäle in Form von unkomprimiertem, linearem PCM-Audio innerhalb eines Core Audio File-Containers, spezifiziert durch Apple (2011). Die dritte Datei mit der Endung *.atmos.metadata* ist eine weitere XML-Datei, welche sämtliche dynamischen Metadaten der bis zu 128 Audio-Objekte enthält, so unter anderem zu den kartesischen Objekt-Koordinaten und dem Size-Parameter. Ein DAMF kann über den DAR geöffnet werden und ermöglicht dabei die nachträgliche Editierung von Metadaten, sowie die Ersetzung einzelner Sektionen des Masters durch Punch-In-Recordings aus der DAW (Dolby Laboratories, 2025c; Dolby Laboratories, 2025d).

3.2.1.2 ADM BWF

Das ADM BWF ist ein nicht proprietäres Master File-Format, welches als offener Standard für die Speicherung und den Austausch objektbasierten Audios entwickelt wurde (Dolby Laboratories, 2025c). Es stellt eine Erweiterung des in der EBU Tech. 3285 spezifizierten Broadcast Wave Formats (BWF)

dar (EBU, 2011), welches um das durch die ITU standardisierte Audio Definition Model (ADM) zur Beschreibung von Metadaten für NGA ergänzt wurde (ITU, 2025). Hierfür wird ein dediziertes ADM-Profil für Dolby Atmos Master verwendet, welches öffentlich zugänglich ist (Dolby Laboratories, 2022). Im Gegensatz zum DAMF handelt es sich bei einem ADM BWF um eine einzige Datei, welche sämtliche Audiospuren in Form von linearem PCM-Audio enthält, ergänzt um einen Header, welcher alle Metadaten beinhaltet (Dolby Laboratories, 2025d). Ein ADM BWF lässt sich in kompatible DAWs importieren, um daraus wieder eine DAW-Session mit allen Audiospuren und objektbezogenen Metadaten generieren zu können (Dolby Laboratories, 2025d).

3.2.1.3 IMF IAB

Bei einem IMF IAB-File handelt es sich um eine MXF-Datei, welche sämtliches Audio und begleitende Metadaten in einem Immersive Audio Bitstream (IAB), spezifiziert durch die Society of Motion Picture and Television Engineers (SMPTE, 2022) in SMPTE ST 2098-2:2022, enthält (Dolby Laboratories, 2025c). IMF IAB-Dateien können mit begleitendem Bild und Untertiteln in einem Interoperable Master Format-Container (IMF) verpackt werden (Dolby Laboratories, 2025c), welcher in SMPTE ST 2067 standardisiert ist und sich insbesondere für den Austausch audiovisueller Inhalte anbietet (SMPTE, 2025).

3.2.2 Consumer-orientierte Formate

Bei Master Files handelt es sich mitunter um sehr große Dateien, deren Wiedergabe eine hohe Bandbreite erfordert. Durch die Verwendung von unkomprimiertem, linearem PCM-Audio ergeben sich bei einer Sampling Rate von 48 kHz, einer Bittiefe von 24 bit und bis zu 128 diskreten Kanälen Datenraten von bis zu 147 Mbit/s allein für die beinhalteten Audiosignale – sämtliche Metadaten sind in dieser Rechnung noch nicht berücksichtigt.

Dies verdeutlicht, weshalb insbesondere für Streaming-Anwendungen, aber auch für die Speicherung auf Blu-ray Disks mit begrenztem Speicherplatz, eine Reduktion der Datenrate notwendig ist, wobei verschiedene Encodierungsverfahren zum Einsatz kommen (Purnhagen et al., 2016).

Da die alleinige Komprimierung von 128 Kanälen mit perzeptuellen Encodierungsverfahren noch keine ausreichende Reduktion der Datenrate bewirkt, muss die tatsächliche Anzahl an übertragenen Kanälen ebenfalls reduziert werden. Das hierfür verwendete Verfahren nennt sich *Spatial Coding* (Tsingos, 2017) und wird im Folgenden erläutert. Dem schließt sich eine Vorstellung relevanter, Consumer-orientierter Codecs an.

3.2.2.1 Spatial Coding

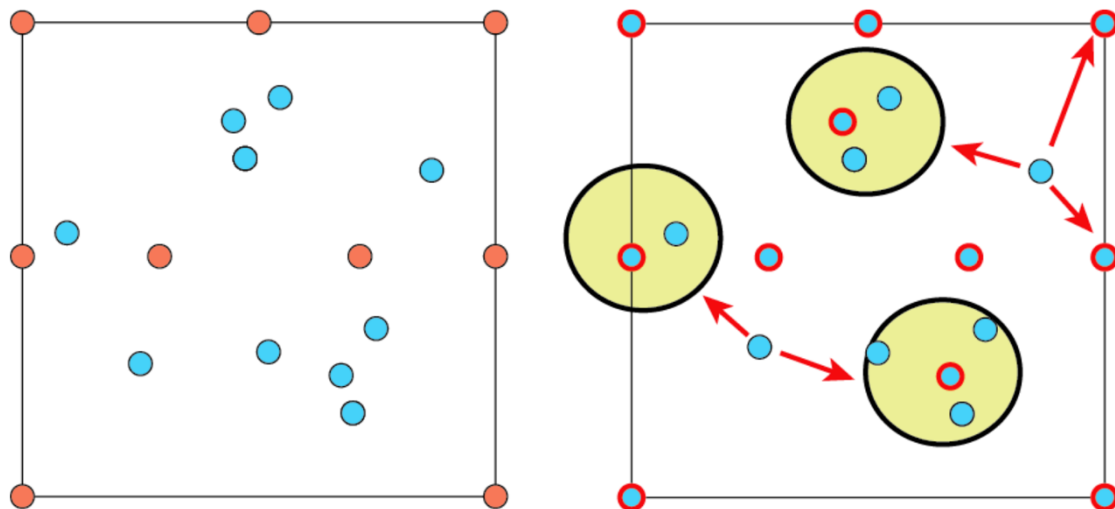
Breebart & Faller (2007) beschreiben Spatial Coding als Möglichkeit zwei oder mehr diskrete Audiokanäle in Form eines um Metadaten ergänzten Downmixes zu übertragen. Diese Metadaten beinhalten

Parameter, welche die räumlichen Eigenschaften der ursprünglichen, höheren Anzahl an Kanälen beschreiben. Wiedergabeseitig können diese Metadaten nun genutzt werden, um die ursprünglichen Kanäle aus dem Downmix zu extrahieren und so das ursprüngliche räumliche Abbild zu rekonstruieren (Breebart & Faller, 2007). Im Kontext von objektbasiertem Audio beschreibt Tsingos (2017) Spatial Coding als die Überführung einer ursprünglichen Anzahl an Audioobjekten in eine geringe Anzahl an Audioobjekten, welche perzeptuell gleichwertig sind. Dies erfolgt durch ein sog. *Clustering* von Audioobjekten (Tsingos, 2017).

Im konkreten Fall von Dolby Atmos findet das sog. Joint Object Coding (JOC) Verwendung (Purnhagen et al., 2016). Hierbei wird zuerst die Anzahl der Audioobjekte reduziert. Dies erfolgt in Form von dynamischem Clustering in sog. *spatial object groups*. Elemente einer solchen Gruppe werden summiert und teilen sich die räumlichen Koordinaten, d.h. es findet eine Reduktion der räumlichen Auflösung statt (siehe Abb. 18).

Abbildung 18:

Audio-Objekt Clustering

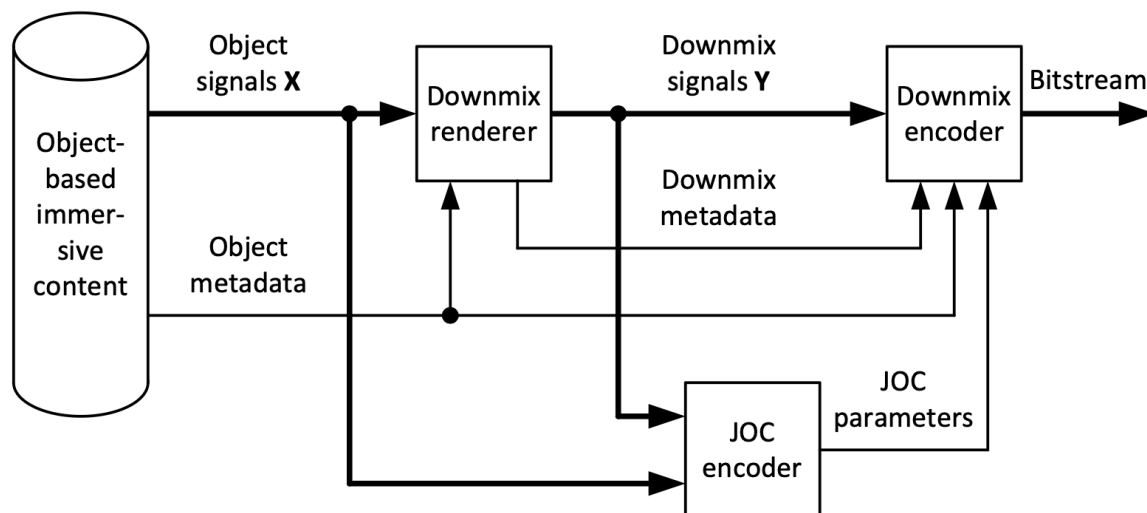


Anmerkung. Visualisierung der Reduktion der räumlichen Auflösung durch dynamische Gruppierung von Audioobjekten im JOC-Prozess. Links dargestellt sind die ursprünglichen 19 simultanen Audioobjekte, unterteilt in die hinsichtlich ihrer Position statischen Bed-Kanäle (rot) und die dynamischen Audioobjekte (blau). Rechts – dargestellt durch die roten Kreise – sind die nach der Gruppierung verbleibenden 11 Audioobjekte. Adaptiert aus *Immersive Audio Delivery Using Joint Object Coding* [Konferenzbeitrag], von H. Purnhagen, T. Hirvonen, L. Villemoes, J. Samuelsson & J. Klejsa, 4.–7. Juni 2016, S. 5. 140th Convention of the Audio Engineering Society (AES), Paris, France.

Ein Audioobjekt kann dabei auch Element mehrerer Cluster sein, z.B. bei Nutzung des Object Size-Parameters. Das Objekt-Clustering erfolgt dabei priorisierend anhand der Berücksichtigung von Attributen der einzelnen Audioobjekte (z.B. deren Lautheit), um den wahrgenommenen Verlust der räumlichen Auflösung zu minimieren. In Abhängigkeit von der verfügbaren Datenrate wird die Anzahl der bis zu 128 simultanen Audio-Objekte so auf 11–15 dynamische Audioobjekte mit begleitenden Metadaten und den LFE-Kanal reduziert (Purnhagen et al., 2016). Darauf folgt das eigentliche JOC-Encoding (siehe Abb. 19).

Abbildung 19:

JOC-Encoding



Anmerkung. Blockdiagramm eines Encoders zur Anwendung von Joint Object Coding. Adaptiert aus *Immersive Audio Delivery Using Joint Object Coding* [Konferenzbeitrag], von H. Purnhagen, T. Hirvonen, L. Villemoes, J. Samuelsson & J. Klejsa, 4.–7. Juni 2016, S. 3. 140th Convention of the Audio Engineering Society (AES), Paris, France.

Die 11–15 Signale der Audioobjekte und ihre begleitenden Metadaten dienen dabei einerseits als Input für einen Downmix-Renderer, welcher aus diesen einen kanalbasierten Downmix erstellt, andererseits dienen sie als Input für den eigentlichen JOC-Encoder, welcher zusätzlich die Downmix-Signale als Input erhält und aus den Differenzen sog. JOC-Parameter generiert. Diese werden vom Downmix-Encoder gemeinsam mit den Downmix-Signalen, den Metadaten der 11–15 Audioobjekte und den Downmix-spezifischen Metadaten in den encodierten Bitstream eingegliedert. Sie können anschließend wiedergabeseitig von einem kompatiblen Decoder interpretiert werden, um die 11–15 Audioobjekte approximativ aus den Downmix-Signalen zu extrahieren (Purnhagen et al., 2016). Ist kein derartiger Decoder

vorhanden, kann der Downmix wiedergegeben werden, womit eine Abwärtskompatibilität zu kanalbasierten Systemen hergestellt wird (Tsingos, 2017).

Frühere Versionen des DAR boten die Möglichkeit die Auswirkungen von Spatial Coding in Form einer aktivierbaren Emulation für das Monitoring zu überprüfen (Dolby Laboratories, 2018c; Rothermich, 2021). Diese Funktion wurde jedoch unter Version 5.2 des DAR entfernt (Dolby Laboratories, 2024c). Für eine Überprüfung ist somit die Encodierung in ein Consumer-orientiertes Format mit anschließender Wiedergabe über ein Dolby Atmos-fähiges Endgerät erforderlich.

3.2.2.2 Dolby Digital Plus Joint Object Coding

Dolby Digital Plus Joint Object Coding (DD+JOC) basiert auf dem Codec Dolby Digital Plus (DD+, alternativ Enhanced AC-3), spezifiziert durch die EBU in ETSI TS 102 366 (EBU, 2017), welcher um die JOC-Funktionalität zur Repräsentation objektbasierten Audios erweitert wurde (EBU, 2018a). DD+ wiederum basiert auf Dolby Digital (AC-3) und stellt eine Abwärtskompatibilität zu diesem Codec sicher (Fielder et al., 2004). Es handelt sich hierbei um verlustbehaftete, perzeptuelle Codecs (Dolby Laboratories, 2025e). DD+JOC ermöglicht die Übertragung von 11-15 Audioobjekten und einem LFE-Kanal in Form eines DD+-Bitstreams, welcher einen um Metadaten ergänzten 5.1-Downmix enthält, der von Endgeräten ohne Dolby Atmos-Kompatibilität decodiert werden kann. Ein Dolby Atmos-fähiges Endgerät enthält einen Decoder, welcher die Audioobjekte aus dem 5.1-Downmix extrahiert, die dann im Anschluss vom OAR auf die Ausgangskanäle des jeweiligen Endgerätes gerendert werden (Dolby Laboratories, 2023a; Dolby Laboratories, 2025e).

DD+JOC wird für viele Streaming- und Broadcast-Anwendungen genutzt (Rothermich, 2021), wobei die niedrigste zu erzielende Datenrate laut Dolby (2023b) bei 384 kbit/s liegt. In diesem Fall ist die Anzahl der encodierten Audioobjekte auf 11 begrenzt (Dolby Laboratories, 2023b). Datenraten ab 448 kbit/s nutzen 15 Audioobjekte (Rothermich, 2021). DD+JOC ist darüber hinaus auch derjenige Codec, der vom DAR beim Export des für die Qualitätskontrolle bestimmten .mp4-Files Verwendung findet (Dolby Laboratories, 2021c). Bei der Encodierung dieses DD+JOC-Bitstreams werden die jeweils aktuellen Downmix- und Trim-Einstellungen des DAR für den 5.1-Downmix berücksichtigt und angewendet (Dolby Laboratories, 2024a; Dolby Laboratories, 2025f).

3.2.2.3 Dolby TrueHD

Dolby TrueHD ist ein durch die Dolby Laboratories (2018e) spezifizierter, verlustfreier Codec, welcher auf dem Kompressionsalgorithmus Meridian Lossless Packing (MLP) der Firma Meridian Audio (Gerson et al., 2004) basiert. Der Codec findet laut Dolby (2016b) primär für die Repräsentation von Dolby Atmos-Inhalten auf Blu-ray Disks Verwendung. Für diesen Anwendungsfall ist eine Übertragung von bis zu 8 Kanälen mit einer Samplerate von bis zu 96 kHz und einer Wortlänge von 24 bit je Kanal möglich, woraus sich maximale Datenraten von 18 Mbit/s ergeben (Slavik & Weinzierl, 2008). Zur

Übertragung objektbasierten Audios wurde der Bitstream um einen Substream für Metadaten ergänzt (Dolby Laboratories, 2016b), welcher die JOC-Parameter enthält. Für Dolby Atmos-Anwendungen kann bei der Encodierung – mit Rücksicht auf den verfügbaren Speicherplatz – zwischen 11, 13 oder 15 Audioobjekten gewählt werden, welche wiedergabeseitig aus dem Downmix extrahiert werden (Rothermich, 2021).

3.2.2.4 Dolby AC-4

Bei Dolby AC-4 handelt es sich um einen von Grund auf neu entwickelten, für die Übertragung von NGA optimierten, perzeptuellen Codec (Dolby Laboratories, 2021d), welcher durch die EBU (2018b; 2025) in ETSI 103 190-1 und ETSI 103 190-2 spezifiziert ist. Der Codec zeichnet sich durch eine durchschnittlich 50 % höhere Effizienz gegenüber E-AC-3 aus und bietet mit Advanced Joint Object Coding (A-JOC) ein optimiertes Spatial Coding-Verfahren (Dolby, 2021d). A-JOC sieht neben erweiterten JOC-Parametern Dekorrelatoren im Decoder vor, welche die aus dem Downmix extrahierten Audio-Signale der Audio-Objekte dekorrelieren und so die Präzision der Approximation an die ursprünglichen Signale gegenüber dem bei DD+JOC verwendeten Verfahren erhöhen (Purnhagen et al., 2016). Hinsichtlich der Distribution von Dolby Atmos-Inhalten, die für die Wiedergabe über Lautsprecher bestimmt sind, findet Dolby AC-4 obgleich der genannten Vorteile aktuell noch wenig Anwendung, mutmaßlich aufgrund der nicht gegebenen Rückwärtskompatibilität zu AC-3 (Devine, 2022; Dolby Laboratories, 2024d).

3.3 Decodierung und Prozessierung durch die Soundbar

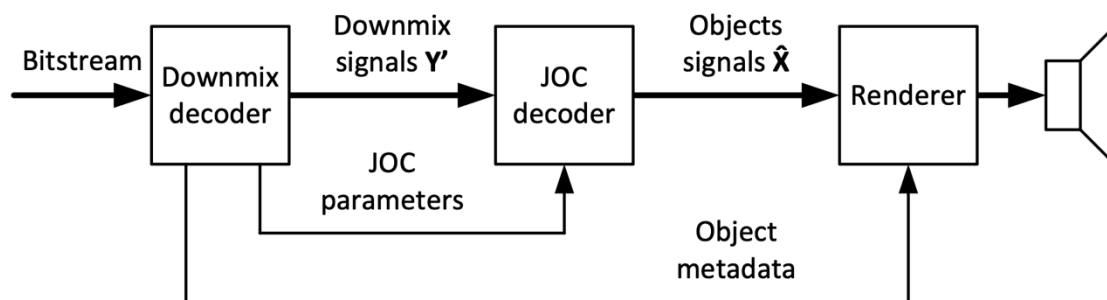
Moderne Soundbars verfügen in der Regel über einen HDMI-Eingang zum Empfang von Audio-Signalen über den Enhanced Audio Return Channel (eARC). Damit die Decodierung des Bitstreams in der Soundbar erfolgt, muss sichergestellt werden, dass die Quelle (z.B. Smart-TV oder Over-the-Top-Gerät) diesen Bitstream unverändert an die Soundbar weiterreicht. Hierzu muss die Option HDMI Passthrough aktiviert werden (Apple, 2025b; Sennheiser, 2023).

3.3.1 Decodierung und Rendering

Im nächsten Schritt wird der durch die Soundbar empfangene Bitstream von einem integrierten Decoder decodiert (siehe Abb. 20).

Abbildung 20:

JOC-Decoding



Anmerkung. Blockschaltbild eines Decoders zur Dekodierung von JOC-encodiertem objektbasiertem Audio. Adaptiert aus *Immersive Audio Delivery Using Joint Object Coding* [Konferenzbeitrag], von H. Purnhagen, T. Hirvonen, L. Villemoes, J. Samuelsson & J. Klejsa, 4.–7. Juni 2016, S. 3. 140th Convention of the Audio Engineering Society (AES), Paris, France.

Nach Purnhagen et al. (2016) generiert dabei ein Downmix-Decoder anhand der mitgeführten Downmix-spezifischen Metadaten die Downmix-Signale aus dem Bitstream und übergibt diese gemeinsam mit den JOC-Parametern an einen JOC-Decoder, welcher anhand der Differenzbeschreibung die 11-15 Audiosignale der Audioobjekte aus den Downmix-Signalen extrahiert. Diese werden an den in die Soundbar integrierten OAR übermittelt, welcher anhand der vom Downmix-Decoder erhaltenen Positionsdaten der Audioobjekte das Rendering für die Ausgangskanäle der Soundbar vollzieht (Purnhagen et al., 2016).

3.3.2 Zusätzliches Processing

Viele Soundbars verwenden zusätzliches Processing, um die decodierten, grundsätzlich für die Wiedergabe über diskrete Mehrkanalsysteme bestimmten Audiosignale für die Wiedergabe über dieses in seiner Funktionsweise abweichende Wiedergabesystem zu optimieren. Dies umfasst neben etwaigem Bass-Management sogenanntens *Virtualizing*, welches laut Dolby unter anderem die Nutzung von HRTFs und CTC umfasst (Dolby Laboratories, 2018b). Einige Hersteller verfolgen darüber hinaus

eigene Ansätze (Fraunhofer IIS, 2023). Die meisten Aspekte dieses Processings sind in der Regel herstellerspezifisch und nicht transparent. Im Folgenden soll daher nur ein Teilaspekt behandelt werden.

3.3.2.1 Automatische Raumkorrektur

Soundbars nutzen häufig, wie in Kapitel 2.5 erläutert, gezielt Raumreflexionen, um virtuelle Lautsprecher an den Begrenzungsflächen des Raumes zu generieren. Der hier relevante Reflexionspfad der Schallwellen für diese indirekt wiedergegebenen Ausgangskanäle – in der Regel die Surround- und Top Surround-Kanäle – ist demzufolge länger als der Pfad des für die Front-Kanäle L, R und C relevanten Direktschalls. Diese Distanzdifferenz ist dabei abhängig von den Raumdimensionen, sowie der Anordnung von Soundbar und Abhörpunkt innerhalb des Raumes und resultiert in Verzögerungen der Ausgangskanäle zueinander. Die akustische Übertragungsstrecke des Raums bewirkt, wie in Kapitel 2.6 erläutert, dass sich die Schalldruckverläufe am Abhörpunkt von den Schalldruckverläufen direkt vor den Lautsprechern hinsichtlich Amplitude und Phase unterscheiden. Um diesen individuellen, mitunter unerwünschten Einfluss des Raumes in Teilen zu kompensieren, bieten viele der erhältlichen Soundbar-Modelle automatische Funktionalitäten zur Raumkorrektur. Häufig wird dazu ein im Lieferumfang enthaltenes Messmikrofon an die Soundbar angeschlossen, am Abhörpunkt platziert und ein automatisiertes Programm gestartet (Sennheiser, 2023; Yamaha Corporation, 2015). Im Folgenden soll in Grundzügen erläutert werden, wie eine derartige Raumkorrektur für einen Ausgangskanal der Soundbar realisiert werden kann.

Initial muss hierzu die jeweilige Raumimpulsantwort $h(t)$ für den Ausgangskanal ermittelt werden. Die Soundbar gibt hierzu ein Messsignal, im Folgenden als Eingangssignal $x(t)$ bezeichnet, über den Ausgangskanal wieder, welches von der Soundbar mithilfe des angeschlossenen Messmikrofons aufgezeichnet wird und dann als Ausgangssignal $y(t)$ vorliegt. Hierzu werden meist Sinus-Sweeps verwendet, vorrangig weil die Wiedergabe von Impulsen über Lautsprecher aufgrund der kurzen Impulsdauer und der begrenzten Membranauslenkung nicht die nötige Schallenergie einbringen kann, was in einem niedrigen Signal-Rausch-Abstand für die Messung resultiert (Ahnert & Goertz, 2008). Das Ausgangssignal $y(t)$ lässt sich nach Farina (2000) unter der Annahme, dass das System linear, zeit-invariant und rauschfrei ist, als das durch die Raumimpulsantwort $h(t)$ gefaltete Eingangssignal $x(t)$ beschreiben, in der Form

$$y(t) = x(t) * h(t) , \quad (11)$$

wobei $h(t)$ unbekannt ist. Um $h(t)$ zu ermitteln, muss eine Entfaltung der Funktionen $x(t)$ und $h(t)$ erfolgen:

$$y(t) = F^{-1} \left\{ \frac{F\{y(t)\}}{F\{x(t)\}} \right\}. \quad (12)$$

Dies wird durch Division der Fouriertransformierten des Ausgangssignals durch die Fouriertransformierte des Eingangssignals mit anschließender inverser Fouriertransformation realisiert, um die Raumimpulsantwort $h(t)$ in der Zeitdomäne zu erhalten (Farina, 2000). Diese Messung erfolgt für alle Ausgangskanäle. Anhand der so ermittelten Raumimpulsantworten der einzelnen Ausgangskanäle können unter Anderem individuelle Delays ermittelt werden, um die Ausgangskanäle auf eine zeitgleiche Ankunft des Schalls am Abhörpunkt zu verzögern (Ahnert & Goertz, 2008), wohingegen die Übertragungsfunktionen genutzt werden können, um die Amplitudenfrequenzgänge der Ausgangskanäle mithilfe digitaler Filter zu glätten oder auf individuelle Zielkurven zu korrigieren (Cecchi et al., 2017; Gerzon, 1991).

3.4 Akustische Wiedergabe über die Soundbar

Im nächsten Schritt erfolgt die akustische Wiedergabe der Ausgangskanäle über die Lautsprecher der Soundbar. Hierbei ist der Wiedergaberaum aufgrund des direkten Einbezugs in Form von Schallreflexionen an dessen Begrenzungsflächen von hoher Relevanz. Dieses letzte Glied der Signalkette wird im folgenden Kapitel 4 anhand eines praktischen Versuchsaufbaus behandelt.

4. Analyse der akustischen Soundbar-Wiedergabe und Vergleich mit diskreter Lautsprecherwiedergabe

4.1 Methodik

Die im Rahmen der Untersuchung gewählte Methodik verfolgt einen gemischten Ansatz unter Einbezug subjektiver und objektiver Kriterien.

4.1.1 Subjektive Kriterien – Hörbeispiele

Im Rahmen der Arbeit wurden Hörbeispiele erstellt, welche das Ziel verfolgen die Wiedergabeeigenschaften der untersuchten Soundbar mit den Wiedergabeeigenschaften des untersuchten diskreten Mehrkanalsystems im Hinblick auf verschiedene Aspekte der Wiedergabe möglichst isoliert vergleichbar zu machen. Insbesondere die Illustration von Gemeinsamkeiten und Unterschieden hinsichtlich der Attribute Lokalisation, ASW und LEV lag bei der Erstellung und Auswahl der Hörbeispiele im Fokus. Auf diese Hörbeispiele wird innerhalb des Textes zur Illustration der Ausführungen des Autors verwiesen und sie sind in den Formaten DAMF, ADM BWF und .mp4 Bestandteil des digitalen Anhangs dieser Arbeit. Der Leser sollte diese im Idealfall über die nachfolgend vorgestellte Wiedergabekonstellation abspielen, um die subjektiven Höreindrücke des Autors überprüfen und ggf. nachvollziehen zu können. Das Hörbeispielverzeichnis liefert zudem eine Auflistung aller Hörbeispiele mit einer kurzen Beschreibung zu deren Inhalt.

4.1.2 Objektive Kriterien – Messungen

Die subjektiven Kriterien werden durch objektive Kriterien in Form von Messungen ergänzt. Hierzu wurden Raumimpulsantworten aller Ausgangskanäle der betrachteten Soundbar und aller Ausgangskanälen des betrachteten diskreten Wiedergabesystems mit Ausnahme des LFE am gemeinsamen Abhörpunkt des vorgestellten Versuchsaufbaus gemessen. Zusätzlich wurden mit einem Kunstkopf (Neumann KU-100) binaurale Impulsantworten (BRIRs) aller Ausgangskanäle beider Systeme am gemeinsamen Abhörpunkt gemessen. Hierfür wurde die Software Room EQ Wizard (REW) verwendet. Einerseits dienen diese Messungen der Visualisierung von Eigenschaften der Schallfelder beider Wiedergabesysteme, andererseits ermöglichen sie die Berechnung objektiver Kriterien, die mit relevanten Attributen der auditiven Wahrnehmung korrelieren. Die erhobenen Messungen dienen primär der Ergänzung und Unterstützung der dargelegten subjektiven Höreindrücke des Autors und sind in Form von REW-Messdateien ebenfalls Bestandteil des digitalen Anhangs dieser Arbeit.

4.2 Versuchsaufbau

Der im Folgenden beschriebene Versuchsaufbau ermöglichte im Rahmen der Arbeit den direkten Vergleich zwischen Soundbar und diskretem Mehrkanal-System hinsichtlich ihrer Wiedergabeeigenschaften von immersivem Audio.

Als Abhörraum wurde hierfür die Regie B des Tonstudios U54 der Hochschule der Medien in Stuttgart genutzt. Der Raum verfügt über eine umfassende raumakustische Optimierung und ermöglicht eine verlässliche Beurteilung der reproduzierten Inhalte. Die in Kapitel 3.1.3.1 beschriebenen Anforderungen an Tonregionen können durch den Raum als erfüllt betrachtet werden.

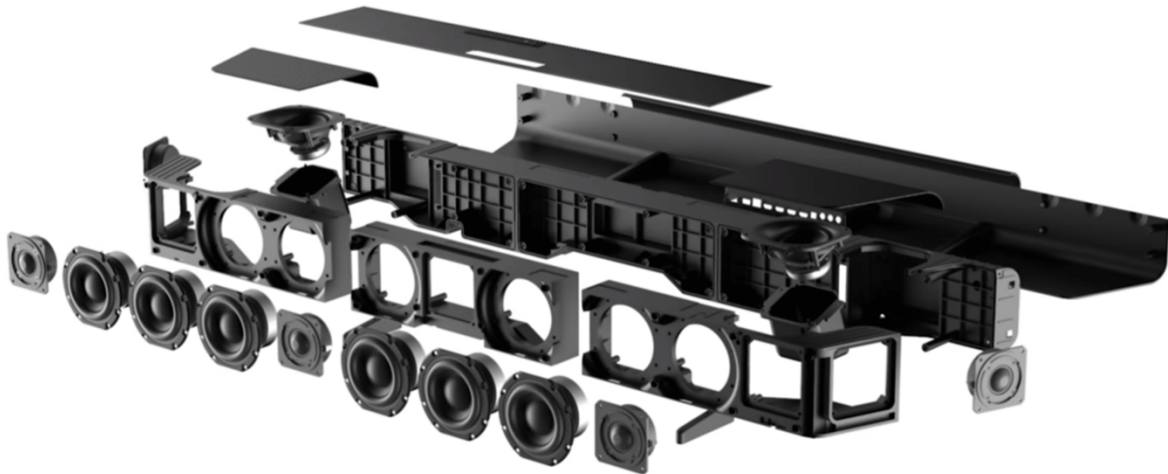
Der Raum verfügt über ein diskretes Monitoring mit 13 Lautsprechern. Dieses umfasst sieben Lautsprecher im Middle Layer, einen Subwoofer für Bass-Management und die Wiedergabe des LFE-Kanals, sowie vier Lautsprecher im Upper Layer. Mit Ausnahme des Subwoofers handelt es sich dabei um koaxiale Studiomonitore des Herstellers Genelec, welche auf identische Schalldruckpegel und einen annähernd linearen Frequenzgang am Abhörpunkt eingemessen sind, sowie durch die Anwendung individueller Delays auf die Ausgangskanäle eine virtuelle Äquidistanz zum Abhörpunkt aufweisen. Die in Kapitel 3.1.3.1 erläuterten Anforderungen an Referenz-Lautsprecher werden erfüllt, ebenso die in Kapitel 2.3.1.2 erläuterten Richtlinien zur Platzierung von Lautsprechern für Dolby Atmos 7.1.4.

Zusätzlich verfügt der Raum über eine vor dem Hörer auf 137 cm über dem Boden positionierte Soundbar. Hierbei handelt es sich um eine Sennheiser Ambeo Max, welche um einen auf die Soundbar abgestimmten Subwoofer für Bass-Management und Wiedergabe des LFE-Kanals ergänzt wird. Der Hersteller (Sennheiser, 2023) gibt für dieses Modell eine Ausgangskonfiguration im Format 5.1.4 an. Die Soundbar verfügt über insgesamt 13 im Gehäuse verbaute Lautsprecher (exkl. externem Subwoofer). Dabei handelt es sich um sechs frontal angebrachte Tief- und Mitteltöner, fünf Hochtöner, wovon drei auf die frontal abgestrahlten Kanäle L, R und C entfallen und zwei auf die seitlich abgestrahlten

Surround-Kanäle Ls und Rs, sowie zwei rechts und links außen integrierte, nach oben abstrahlende Breitband-Lautsprecher zur Wiedergabe der Kanäle Ltf, Rtf, Ltr und Rtr (Sennheiser, 2023). Abbildung 21 zeigt eine Explosionsdarstellung der Soundbar.

Abbildung 21:

Explosionsdarstellung der Soundbar



Anmerkung. Explosionsdarstellung der im Rahmen der Arbeit untersuchten Soundbar Sennheiser Ambeo Max. Hier ist die Anordnung der 13 Lautsprecher im Gehäuse der Soundbar ersichtlich. Adaptiert aus *High-End TV-Sound mit der Sennheiser Ambeo Soundbar*, von HiFi im Hinterhof GmbH, 2019.

Um dem Funktionsprinzip dieser direkt abstrahlenden Soundbar Rechnung zu tragen, wurden an den Seitenwänden der Tonregie mobile Stellwände mit schallharter Oberfläche platziert, welche die Schallwellen der seitlich abstrahlenden Lautsprecher reflektieren. Außerdem sind die Reflexionspunkte an der Decke des Raumes für die nach oben abstrahlenden Lautsprecher der Soundbar mit schallharten Oberflächen versehen. Die Rückwand des Abhörtraumes ist zudem mit einem schallharten QRD-Diffraktor ausgestattet, welcher auftreffende Schallwellen breitbandig horizontal zerstreut. Die mobilen Stellwände lassen sich drehen und sind auf der gegenüberliegenden Seite mit Breitbandabsorbern bestückt. Die Einmessung der Soundbar, sowie die Messungen der RIRs und BRIRs wurden mit reflektierenden Stellwänden vorgenommen. Die Messungen der RIRs und BRIRs des diskreten Systems erfolgten hingegen mit absorbierenden Stellwänden, um einen Vergleich zwischen reflexionsarmer Studioumgebung und einem typischen Anwendungsumfeld der Soundbar in den Messdaten repräsentieren zu können. Darüber hinaus wurde die reflektierende Oberfläche des Studiotisches für sämtliche

Messungen und Hörvergleiche mit porösen Absorbieren abgedeckt – dies gilt sowohl für die Soundbar als auch für die diskrete Wiedergabe. Der Versuchsaufbau ist Abbildung 22 zu entnehmen.

Abbildung 22:

Verwendeter Versuchsaufbau für Hörvergleiche und Messungen



Anmerkung. Visualisierung des im Rahmen der Arbeit verwendeten Versuchsaufbaus für Hörvergleiche und Messungen von diskretem Mehrkanalsystem und Soundbar mit gemeinsamem Abhörpunkt in Regie B des Tonstudios U54 der Hochschule der Medien in Stuttgart. Eigene Darstellung.

Die Hörbeispiele wurden mit der DAW Steinberg Nuendo unter Verwendung des externen DAR erstellt. Die Ausgänge des DAR wurden dazu via Dante mit den physischen Eingängen der Lautsprecher des diskreten Mehrkanalsystems verbunden. Um einen direkten Vergleich von Hörbeispielen auf beiden Wiedergabesystemen zu ermöglichen, wurden alle Hörbeispiele zunächst als DAMF gesichert. Dieses Format konnte direkt über den DAR wiedergegeben werden, wobei die im DAR konfigurierten Sub-Layouts zusätzlich für das Monitoring in 7.1.2, 5.1.4 und 5.1.2 genutzt werden konnten. Aus den geöffneten DAMFs der Hörbeispiele wurden anschließend .mp4-Container exportiert, welche jeweils ein

encodiertes DD+JOC mit einer Datenrate von 768 kbp/s enthalten. Diese sind für die Wiedergabe über die Soundbar geeignet. Die Einstellungen hinsichtlich Downmix und Trims sind zwischen den DAMF- und .mp4-Versionen der Hörbeispiele identisch, um eine Vergleichbarkeit gewährleisten zu können. Um die .mp4-Dateien auf der Soundbar wiedergeben zu können, wurden diese auf einen via HDMI eARC an die Soundbar angebundenen Media-Player (Nvidia Shield) übertragen. Wird das .mp4-File über dessen GUI angewählt, wird der Bitstream via HDMI Passthrough an die Soundbar geleitet, von dieser decodiert, auf die Ausgänge der Soundbar gerendert und wiedergegeben. Alternativ ließe sich das .mp4-File auch über das diskrete System wiedergeben. Hierfür war ursprünglich ein zwischen Media-Player und Soundbar eingebundener, externer Decoder/Renderer (Arvus H1-D) vorgesehen, welcher innerhalb der Bearbeitungszeit dieser Arbeit jedoch nicht rechtzeitig zur Verfügung stand. Der vom Decoder/Renderer durch den Media-Player empfangene Bitstream kann wahlweise via HDMI Passthrough an die Soundbar durchgereicht oder auf die gewünschte Lautsprecher-Konfiguration des diskreten Systems gerendert werden. Die Ausgangskanäle des Decoders/Renderers können dann via Dante auf die physischen Kanäle der Lautsprecher geroutet werden. Durch diese Herangehensweise ließe sich ausschließen, dass etwaige Differenzen in den Höreindrücken in der verlustbehafteten Encodierung und im angewendeten Spatial Coding begründet liegen. Da im Rahmen der Arbeit insbesondere der Transfer zwischen Monitoring in der Tonregie und der Wiedergabe über die Soundbar beim Endnutzer von Interesse war, ist das Vorgehen mit der Wiedergabe über die zwei verschiedenen Ausgabeformate repräsentativ für das untersuchte Szenario.

4.3 Überblick

Für einen ersten Überblick über die Ausgangskanäle der beiden Wiedergabesysteme wurden zwei dedizierte Hörbeispiele erstellt.

Hörbeispiel 1 ist für die Wiedergabe über das diskrete System vorgesehen und liefert eine sequenzielle Ansage der Ausgangskanäle der 7.1.4-Anordnung zu deren Identifikation.

Hörbeispiel 2 ermöglicht die Identifikation der 5.1.4-Ausgangskanäle der Soundbar nach dem gleichen Prinzip.

4.4 Wiedergabe von Einzelkanälen und Kanalgruppen

Zunächst wird behandelt, wie die Wiedergabe einzelner Kanäle durch die beiden Systeme realisiert wird. Zur Strukturierung wird eine Unterteilung in die drei Kanalgruppen Front-Kanäle, Surround-Kanäle und Top Surround-Kanäle vorgenommen. Der LFE-Kanal wird aufgrund des hinsichtlich beider Wiedergabesysteme identischen Wiedergabepinzips über einen externen Subwoofer nicht gezielt behandelt.

4.4.1 Front-Kanäle

Die untersuchte Soundbar verfügt auf der Vorderseite über drei Hochtöner, welche jeweils für die Wiedergabe der Front-Kanäle L, R und C bestimmt sind. Diese sind über die Breite der Soundbar verteilt angeordnet und frontal ausgerichtet. Wie bei der diskreten Wiedergabe wird der aus der Wiedergabe resultierende Höreindruck dabei maßgeblich vom Direktschall der Lautsprecher bestimmt. Die Wiedergabe über diese Kanäle zeichnet sich durch eine scharfe Lokalisierbarkeit breitbandiger Signale an den Positionen der Hochtöner aus. Dieser Eindruck deckt sich in diesem Hinblick weitgehend mit dem Höreindruck über das diskrete System.

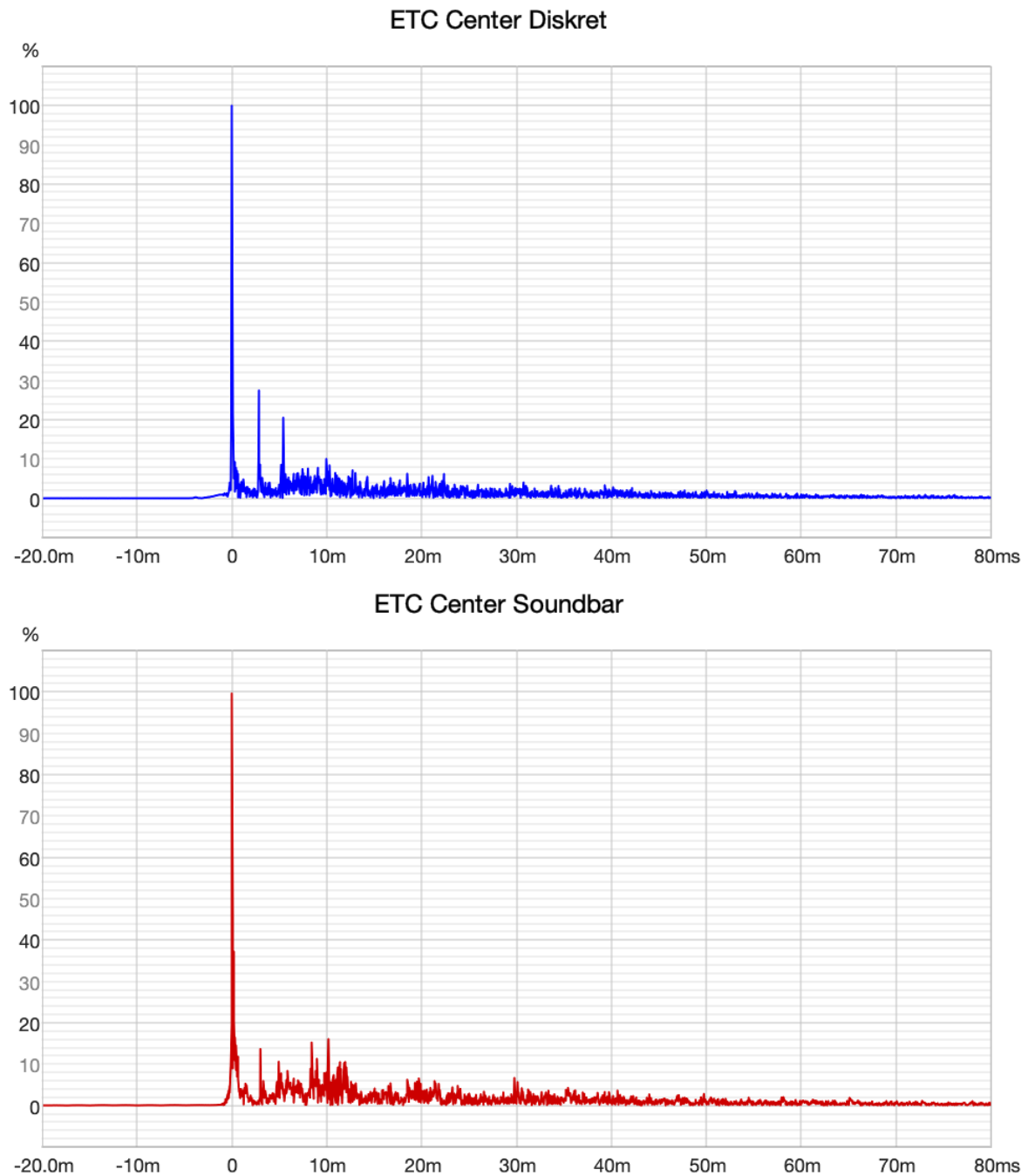
Hörbeispiel 3 demonstriert dies anhand der kanal-sequenziellen Wiedergabe pinken Rauschens über die Kanäle L, C und R. Die Wiedergabe von transienten Signalen ermöglicht ebenfalls eine scharfe Lokalisation und die wahrgenommene Impulstreue der Soundbar-Wiedergabe ist vergleichbar mit der Wiedergabe über das diskrete System.

Hörbeispiel 4 ermöglicht diesbezüglich den direkten Vergleich anhand der kanal-sequenziellen Wiedergabe von Snare-Samples über die Kanäle L, C und R.

Abbildung 23 ergänzt diese Höreindrücke anhand einer Gegenüberstellung der normierten Energy Time Curves (ETC) von diskretem System und Soundbar für die Wiedergabe über den Center-Kanal, wobei sich ein ähnlicher Verlauf der beiden Kurven erkennen lässt. Die ETC stellt nach Definition die Hüllkurve des quadrierten Schalldrucks über die Zeit dar und visualisiert aufgrund der direkten Proportionalität zwischen Schalldruck und Schallintensität so den relativen Verlauf der Schallenergie über die Zeit am Messpunkt (Görne, 2015). Die hier genutzte, aus der Software REW übernommene Darstellung der ETC ist jedoch nicht quadriert, sondern visualisiert die normierte Hüllkurve des Amplitudenverlaufs des Messsignals. Um das aus der Hüllkurve ersichtliche Verhältnis zweier Spitzen in einem Pegel auszudrücken, muss daher der zwanzigfache Logarithmus dieses Verhältnisses berechnet werden. Im Folgenden wird diese Form der Darstellung in Übereinstimmung mit der Benennung innerhalb der Software dennoch als ETC bezeichnet.

Abbildung 23:

ETCs des Center-Kanals von diskretem System und Soundbar



Anmerkung. Normierte ETCs des Center-Kanals von diskretem Mehrkanalsystem (blau) und Soundbar (rot) im Vergleich. Deutlich zu erkennen ist jeweils der initiale Impuls bei $t_0 = 0$ ms. Eigene Darstellung.

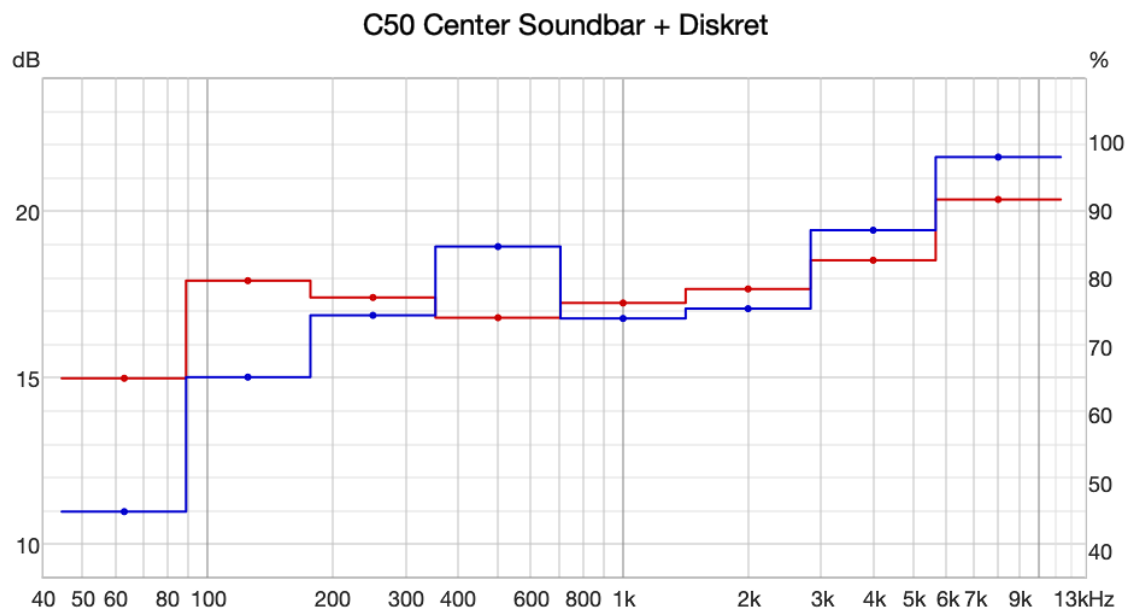
Für die Verständlichkeit von Dialogen ist der Center-Kanal von hoher Bedeutung. Anhand der RIRs beider Systeme lässt sich das Klarheitsmaß C_{50} berechnen, welches als Indikator für die Sprachverständlichkeit einer Tonwiedergabe gilt (Ahnert & Tennhardt, 2008). Hierbei wird die am Messpunkt eintreffende Schallenergie der ersten 50 ms ins Verhältnis zur Schallenergie der verbleibenden Impulsantwort gesetzt, in der Form

$$C_{50} = 10 \cdot \log_{10} \left(\frac{\int_0^{50 \text{ ms}} p^2(t) dt}{\int_{50 \text{ ms}}^{\infty} p^2(t) dt} \right) dB. \quad (13)$$

Abbildung 24 ermöglicht den Vergleich der aus den RIRs ihrer jeweiligen Center-Kanäle berechneten und über Oktavbänder gemittelten Werte für C_{50} für beide Wiedergabesysteme. Mit Ausnahme der Frequenzen unterhalb von 200 Hz zeigen sich hier nur geringe Abweichungen von weniger als 3 dB, was als Indiz für eine vergleichbare Sprachverständlichkeit beider Wiedergabesysteme zu deuten ist.

Abbildung 24:

Klarheitsmaß C_{50} für den Center-Kanal von diskretem System und Soundbar



Anmerkung. Darstellung der über Oktavbänder gemittelten Klarheitsmaße C_{50} für die jeweiligen Center-Kanäle von diskretem Mehrkanalsystem (blau) und Soundbar (rot). Eigene Darstellung.

Unterschiede zwischen diskreter Wiedergabe und Soundbar-Wiedergabe der Front-Kanäle offenbaren sich insbesondere hinsichtlich der Azimuth-Auslenkung von Hörereignissen, die über die Kanäle L und R wiedergegeben werden. Während die entsprechenden Lautsprecher beim diskreten Wiedergabe-

system für einen Öffnungswinkel von 60° am Abhörpunkt bei $\pm 30^\circ$ auf der Horizontalebene positioniert sind, ist der Öffnungswinkel der Soundbar beim gewählten Hörabstand von 1,55 m und einem Abstand der Hochtöner der Kanäle L und R von ca. 1 m auf ca. 36° beschränkt. Ein auf Kanal L positioniertes Signal wird folglich mit Abbildungsverzerrungen gegenüber dem diskreten Wiedergabesystem wahrgenommen.

Hörbeispiel 5 demonstriert dies anhand der Wiedergabe eines breitbandigen und transienten Monosignals (Schlagzeug & Bass) über den Kanal L. Bei einem Wechsel zwischen den beiden Wiedergabesystemen ist dabei ein deutlicher Unterschied hinsichtlich der Lokalisationsrichtung auszumachen. Da Soundbars in der Regel mit begleitenden Bildinhalten verwendet werden, wird diese Eigenschaft jedoch in Teilen relativiert. Ist die Soundbar direkt über oder unter einem TV-Bildschirm mit einer Bilddiagonale von ca. 55 Zoll platziert, so decken die Kanäle L und R die Randbereiche des Bildes ab, während der Center-Kanal die Bildmitte repräsentiert. Der Bildschirm dient hier als optische Referenz. Insofern überträgt sich die Lokalisation von bildbegleitenden Audio-Inhalten akkurat auf die Soundbar-Wiedergabe und ist dahingehend mit der Funktion der LCR-Kanäle im Kino vergleichbar.

4.4.2 Surround-Kanäle

Die in der 5.1.4-Repräsentation der Soundbar enthaltenen Surround-Kanäle Ls und Rs werden über zwei dedizierte, seitlich abstrahlende Lautsprecher wiedergegeben. Da es sich dabei um Hochtöner handelt, erfolgt die Wiedergabe tiefer und mittlerer Frequenzen zusätzlich über die frontal angeordneten Tief-/Mitteltöner. Die seitlich abgestrahlten Signalanteile erreichen den Hörer dabei primär indirekt über Wandreflexionen.

Beim diskreten 7.1.4-System erfolgt die Wiedergabe der Surround-Kanäle über die bei ca. $\pm 110^\circ$ Azimuth platzierten Surround-Lautsprecher Ls und Rs, sowie die zusätzlichen Rear Surround-Lautsprecher Lrs und Rrs bei ca. $\pm 135^\circ$ Azimuth.

Hörbeispiel 6 liefert einen ersten Eindruck hinsichtlich der Wiedergabe breitbandiger Signale über die Kanäle Ls und Rs in Form von sequenziellem pinkem Rauschen. Dabei fällt auf, dass die Lokalisation der Hörereignisse zwischen den beiden Wiedergabesystemen hinsichtlich der Azimuth-Auslenkung von starken Diskrepanzen geprägt ist. Bei der diskreten Wiedergabe erfolgt die Lokalisation jeweils an der Position des Lautsprechers. Bei der Soundbar ist keine derart scharfe Lokalisation der Hörereignisse möglich. Die beiden Rauschsignale wirken im Vergleich räumlich stark ausgedehnt und deren Mitte wird tendenziell links seitlich bei ca. $\pm 40\text{--}60^\circ$ Azimuth lokalisiert.

Hörbeispiel 7 ermöglicht den Vergleich hinsichtlich der Wiedergabe transienter Signale in Form von sequenziellen Snare-Samples über die Kanäle Ls und Rs. Die Azimuth-Auslenkung wird bei der Soundbar-Wiedergabe in diesem Beispiel geringer wahrgenommen als beim Rauschsignal und erfolgt bei ca. $\pm 30\text{--}45^\circ$ Azimuth. Das Hörereignis ist dabei präziser zu lokalisieren als das Rauschsignal, wirkt aber dennoch räumlich ausgedehnt. Dies fällt insbesondere im Vergleich mit der diskreten Wiedergabe auf.

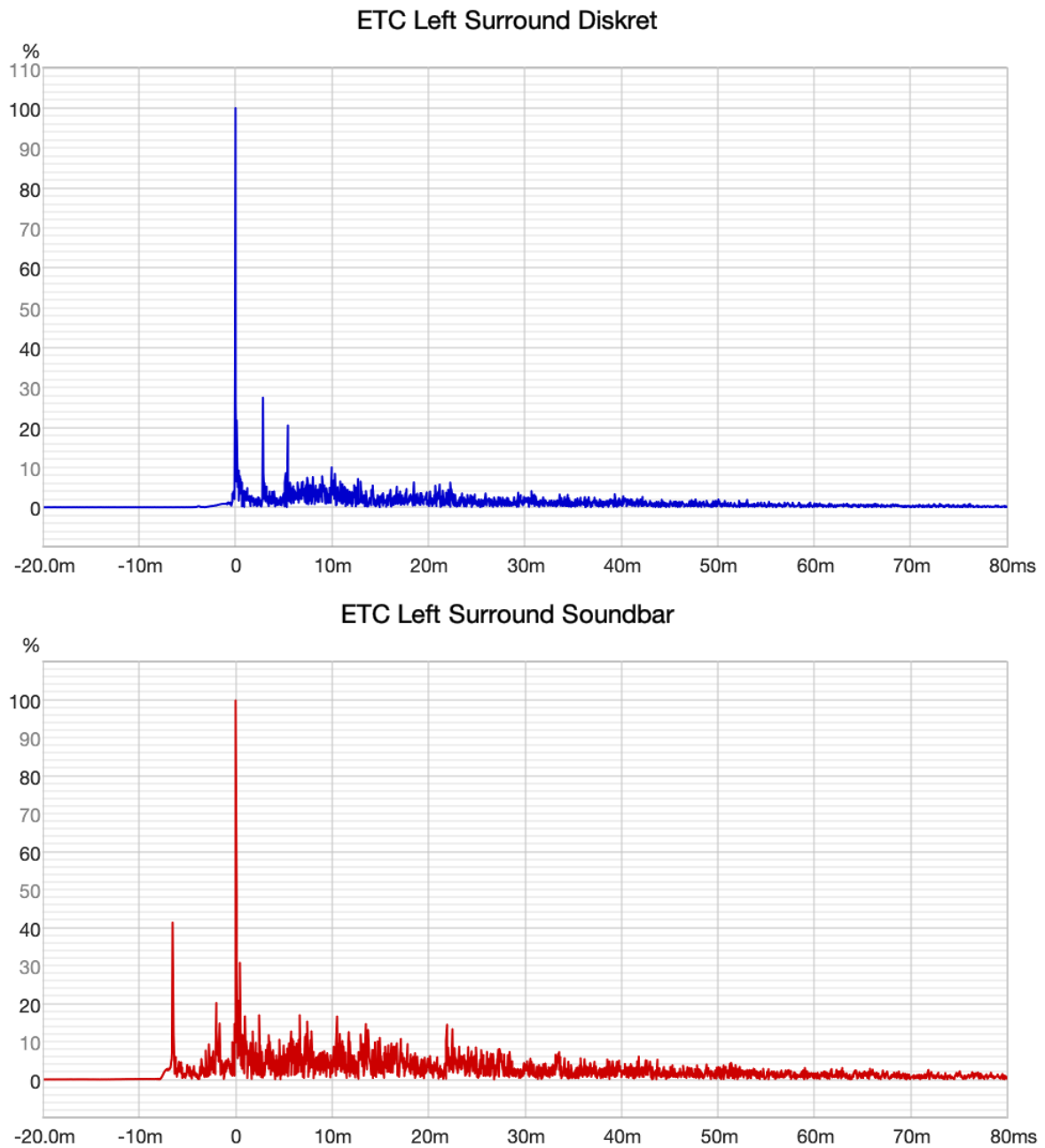
Hier sind die Signale deutlich schärfer an den Positionen der Lautsprecher lokalisierbar. Die Abbildungsverzerrung zwischen den beiden Wiedergabesystemen ist in diesem Beispiel mit einem Betrag von ca. 65-80° Azimuth besonders auffällig.

Darüber hinaus fällt auf, dass der aus der Soundbar-Wiedergabe resultierende Klangeindruck diffuser ist und die Transienten weit weniger präzise wirken als bei der diskreten Lautsprecherwiedergabe. Auch im direkten Vergleich zwischen den Front- und Surround-Kanälen der Soundbar mit den gleichen Signalen bestätigt sich dieser Höreindruck, was durch Wiedergabe von *Hörbeispiel 8* und *Hörbeispiel 9* überprüft werden kann.

Abbildung 25 zeigt die aus den RIRs ihrer jeweiligen Ls-Kanäle generierten, normierten ETCs von Soundbar und diskretem Wiedergabesystem. Hier lässt sich erkennen, dass die zum jeweils energiereichsten Impuls ins Verhältnis gesetzten frühen Reflexion bei der Soundbar-Wiedergabe stärker ausgeprägt sind. Der Raum wird für eine gegebene Schallintensität am Messpunkt folglich stärker angeregt als bei der diskreten Wiedergabe.

Abbildung 25:

ETCs des Left Surround-Kanals von diskretem System und Soundbar



Anmerkung. Normierte ETCs des Left Surround-Kanals von diskretem Mehrkanalsystem (blau) und Soundbar (rot) im Vergleich. Eigene Darstellung.

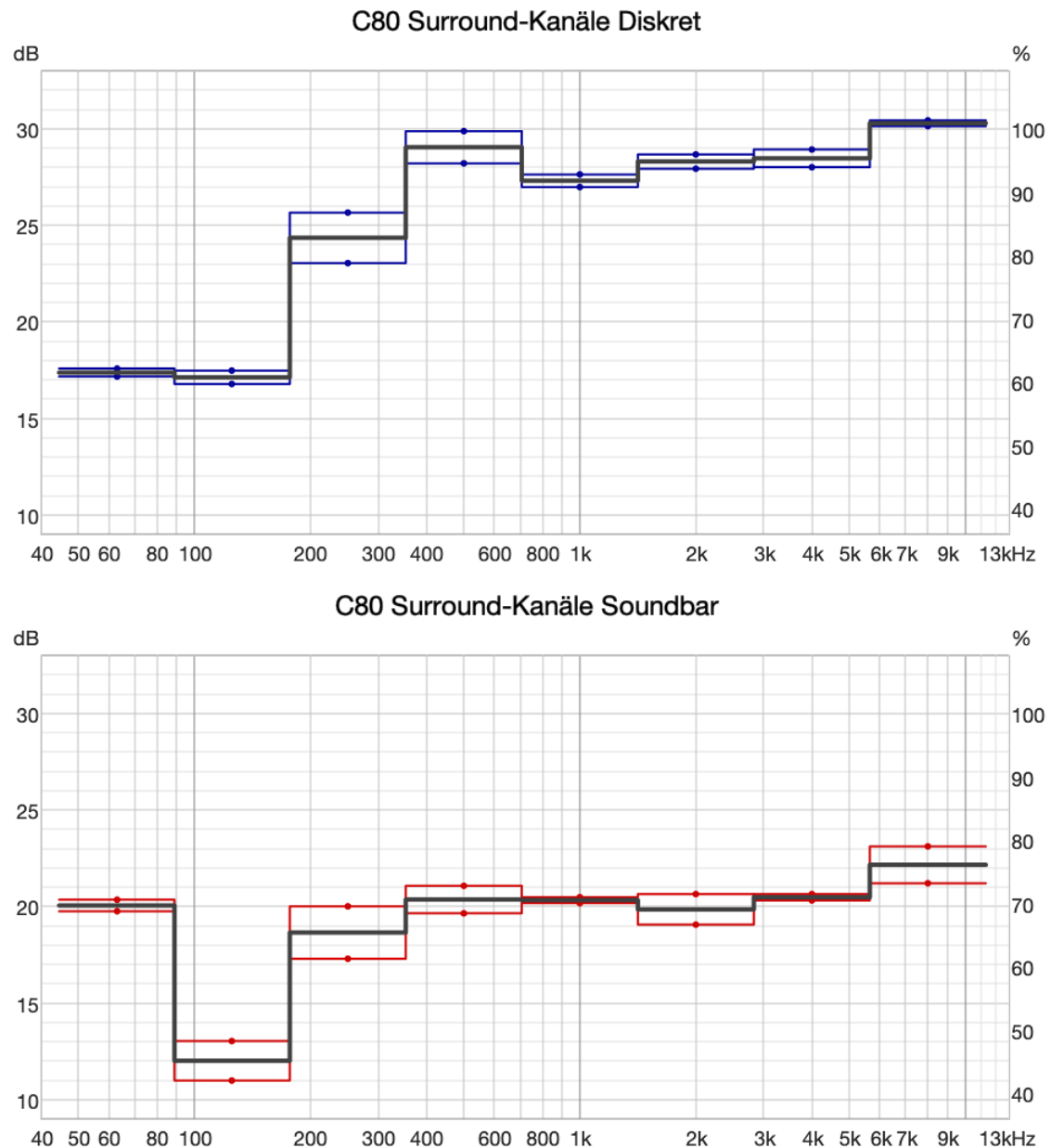
Auch die späten Reflexionen sind bei der Soundbar-Wiedergabe im Verhältnis zur Summe aus Direktschall und frühen Reflexionen gegenüber der diskreten Lautsprecherwiedergabe offenbar stärker ausgeprägt, was sich anhand der Ermittlung des Klarheitsmaßes für Musik C_{80} (Ahnert & Tennhardt, 2008) demonstrieren lässt. C_{80} berechnet sich dabei wie C_{50} , mit dem Unterschied, dass die Integrationsgrenze bei 80 ms liegt. Es gilt:

$$C_{80} = 10 \cdot \log_{10} \left(\frac{\int_0^{80 \text{ ms}} p^2(t) dt}{\int_{80 \text{ ms}}^{\infty} p^2(t) dt} \right) dB \quad (14)$$

C_{80} setzt demnach die Schallenergie der ersten 80 ms ins Verhältnis zur gesamten Schallenergie, die danach eintrifft (Beranek, 2004; Dickreiter & Hoeg, 2014). Abbildung 26 zeigt eine Gegenüberstellung der innerhalb von Oktavbändern gemittelten Werte für C_{80} von Soundbar und diskretem Wiedergabesystem, berechnet aus ihren jeweils gemittelten RIRs der Kanäle Ls und Rs. Hier lässt sich erkennen, dass die energetische Differenz zwischen Direktschall plus frühen Reflexionen gegenüber den späten Reflexionen ab $t = 80$ ms bei der Soundbar-Wiedergabe – mit Ausnahme der tiefen Frequenzen unter 100 Hz – über alle Frequenzen hinweg weniger ausgeprägt ist als bei der Wiedergabe über das diskrete System. Dies zeigt, dass eine Wiedergabe über die Surround-Kanäle der Soundbar eine im Verhältnis höhere Anregung des Raumes bewirkt als die Wiedergabe der korrespondierenden Kanäle des diskreten Wiedergabesystems.

Abbildung 26:

Klarheitsmaß C_{80} der Surround-Kanäle von diskretem System und Soundbar



Anmerkung. Gegenüberstellung der über Oktavbänder gemittelten Klarheitsmaße C_{80} für die Surround-Kanäle Ls und Rs von diskretem Mehrkanalsystem (oben) und Soundbar (unten). Die jeweils schwarze Kurve illustriert dabei jeweils das arithmetische Mittel des für beide Surround-Kanäle ermittelten Kriteriums (farbige Kurven). Eigene Darstellung.

Für die korrekte Einordnung dieser Messungen ist jedoch wichtig an dieser Stelle noch einmal zu betonen, dass die Messbedingungen unterschiedlich waren und die Messungen der RIRs der Soundbar aufgrund der reflektierenden Stellwände in einem reflektiveren Umfeld erfolgten.

Hörbeispiel 10 illustriert eine weitere Eigenschaft der Soundbar, welche bei der Wiedergabe breitbandiger Signale über die Surround-Kanäle in Erscheinung tritt. Wiedergegeben wird ein breitbandiges und transientes Monosignal, bestehend aus Schlagzeug und Bass über den Kanal Ls. Es lässt sich hier eine frequenzabhängige räumliche Ausdehnung des Hörereignisses beobachten. Die tiefen und tiefmittigen Frequenzen des auf Ls wiedergegebenen Signals werden vom Hörer tendenziell mehr in Richtung der Schallquelle lokalisiert, während die Lokalisation hoher Frequenzen zunehmend in Richtung des Reflexionspunktes bei ca. 60° Azimuth erfolgt. Insbesondere das China-Becken des Schlagzeugs kann dabei mit höherer lateraler Auslenkung wahrgenommen werden. Diese Ausdehnung wirkt jedoch nicht statisch, sondern scheint auch abhängig vom aktuellen Spektrum des Signals zu sein. Der Effekt ist insbesondere dann wahrzunehmen, wenn man den Kopf in Richtung des Reflexionspunktes dreht oder den Abhörpunkt verlässt. Bei der Wiedergabe über das diskrete System ist ein solcher Effekt nicht zu beobachten, hier erfolgt die Lokalisation aller Frequenzen ohne auffällige räumliche Ausdehnung stets scharf an der Position des Lautsprechers. Dies gilt auch bei Verlassen des Sweetspots.

Man könnte nun schlussfolgern, dass dieser Effekt darin begründet liegt, dass die betrachtete Soundbar ausschließlich seitlich abstrahlende Hochtöner verwendet und die Lokalisation tiefer Frequenzen deshalb an Position der frontal angeordneten Tief-/Mitteltöner erfolgt. In Konsequenz könnte man weiter annehmen, dass der Effekt bei anderen Soundbar-Modellen nicht zu beobachten wäre.

Vielmehr kann jedoch ein komplexes Zusammenwirken verschiedener Faktoren, die sowohl in der Physik als auch in der Psychoakustik begründet liegen, als Ursache für den beobachteten Effekt angesehen werden.

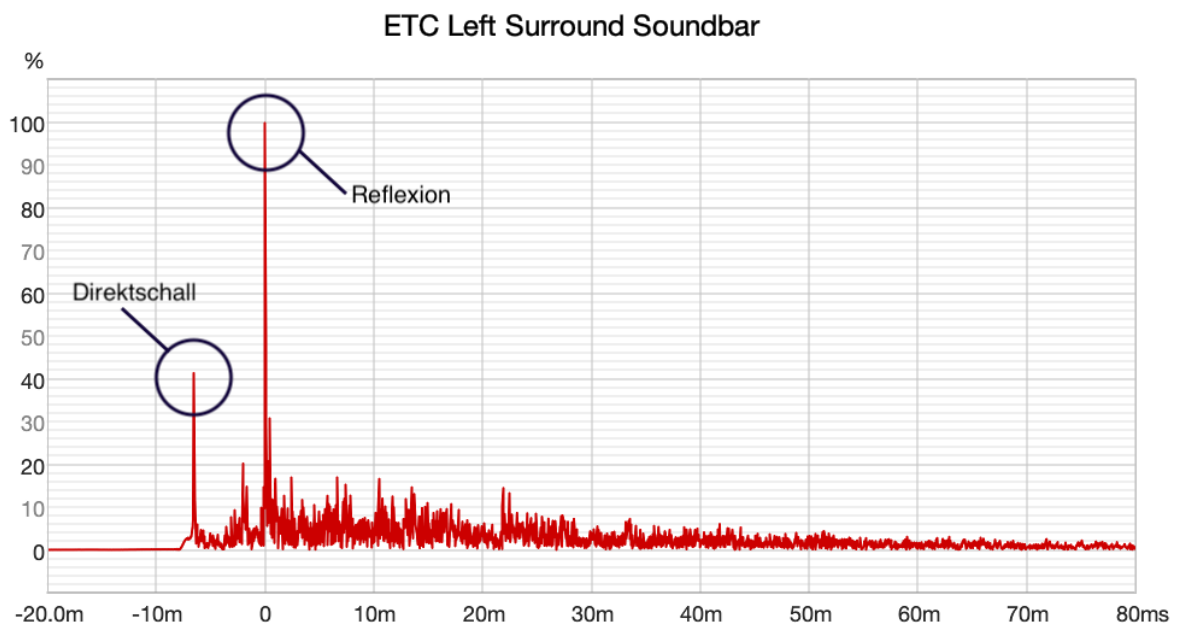
Die indirekte Wiedergabe der Surround-Kanäle über Wandreflexionen bewirkt, dass der starken frühen Reflexion stets ein Direktschallanteil vorausgeht, welcher den Hörer auf kürzestem Wege erreicht. Primärschall und Reflexion lassen sich dabei auch als zwei getrennte Schallquellen betrachten (Blauert, 1996). Zwischen diesen beiden Schallquellen liegen in diesem Fall entgegengesetzt wirkende ICTD und ICLD vor. Der Primärschall eilt der Reflexion voraus (ICTD), die Reflexion weist jedoch einen höheren Schalldruckpegel auf (ICLD). Diese Merkmale können beim Hörer in widersprüchlichen ITD und ILD resultieren, welche kompensatorisch auf die Hörereignisauslenkung einwirken. Die ICTD hängt dabei von der Wegdifferenz ab und ist umso größer, je weiter die reflektiven Wände von der Soundbar entfernt sind. Die ICLD ist primär von den Absorptionseigenschaften am Reflexionspunkt, der Wegdifferenz, dem Abstrahlverhalten der Lautsprecher und dem Auftrittswinkel der Schallwellen auf die Wand abhängig. So lässt sich erklären, weshalb die Hersteller von Soundbars häufig explizite Empfehlungen zu maximalen Abständen der Seitenwände geben (Sennheiser, 2023; Yamaha, 2015). Einerseits muss sichergestellt werden, dass die Wegdifferenz zwischen Primärschall und Reflexion nicht so groß ist, dass die daraus resultierende ICTD im Wirkungsbereich des Haas-Effekts liegt (ca. 10-20 ms, entspricht Δd von ca. 3,4–6,9 m), andererseits bewirkt eine höhere Wegdifferenz aufgrund des entfernungsbedingten Pegelabfalls auch eine Verringerung der ICLD zwischen Primärschall und

Reflexion, welche für eine laterale Auslenkung des Hörereignisses in Richtung Reflexionspunkt von hoher Bedeutung ist.

Die ETC des Ls-Kanals der Soundbar weist für die im Rahmen der Arbeit genutzte Anordnung eine ICTD von ca. 6,5 ms und eine im Schalldruckpegel ausgedrückte ICLD von ca. 8 dB (siehe Abb. 27) zwischen Primärschall und Reflexion auf.

Abbildung 27:

Annotierte ETC des Kanals Ls der Soundbar



Anmerkung. Normierte ETC des linken Surround-Kanals der Soundbar mit erkennbar vorausgehendem Direktschallanteil ($t = -6,5$ ms) und der um ca. 8 dB höheren Erstreflexion ($t = 0$ ms). Eigene Darstellung.

Effektiv findet hier also ein ITD/ILD-Trading (Blauert, 1996; Blauert & Braasch, 2008) statt, damit das Hörereignis aufgrund des hier wirksamen Präzedenzeffekts nicht an der Position der Primärschallquelle lokalisiert wird. Aufgrund der Beugungseffekte um den Kopf können sich die ICLD zwischen Primärschall und Wandreflexion jedoch nur bis zu einer Grenzfrequenz von ca. 2000 Hz unmittelbar auf die ILD auswirken (Blauert, 1996; Görne 2015; Wenzel et al., 2017). Unterhalb dieser Grenzfrequenz wird die ITD zum dominierenden Faktor für die Hörereignisauslenkung, insbesondere jedoch für Frequenzen unterhalb von ca. 1400-1600 Hz (Blauert, 1996; Wenzel et al., 2017).

Dazu kommen physikalische Faktoren: Einerseits ist die für eine hohe ICLD zwischen Primärschall und Reflexion notwendige Schallbündelung mit Konuslautsprechern einfacher für hohe Frequenzen zu

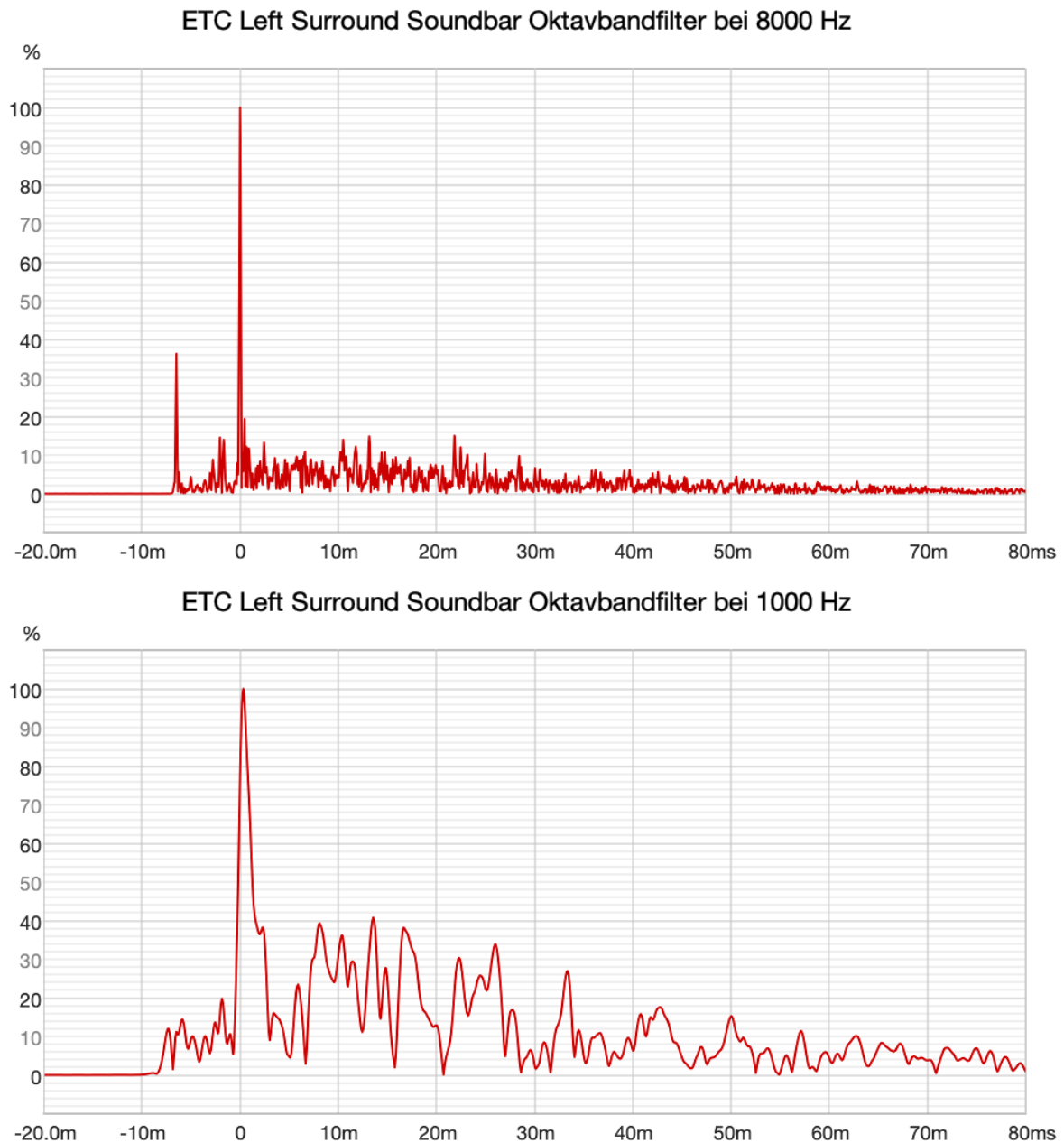
realisieren, deren Wellenlängen kurz in Relation zum Membrandurchmesser sind (Frisch, 2004; Schullan et al., 2014). Andererseits ist die in Kapitel 2.6.1.2 unter Formel (8) notierte Reflexionsbedingung für Totalreflexion (Dickreiter & Hoeg, 2014) zu beachten. Die im Rahmen der Untersuchung verwendeten mobilen Stellwände an den Reflexionspunkten weisen einen umgerechneten Kreisdurchmesser von ca. 2 m auf, wobei der Schall der seitlich abstrahlenden Lautsprecher diese mit einem Aufttrittswinkel von ca. $60\text{--}65^\circ$ erreicht. Zusammen mit dem Abstand der Soundbar zur linken Stellwand von 1,60 m und dem Abstand dieser Stellwand zum Abhörpunkt von ca. 2,10 m lässt sich für die Versuchsumgebung so eine untere Grenzfrequenz von ca. 280 Hz für eine vollständige Reflexion annehmen. Unterhalb dieser Frequenz würde der Schall zunehmend gebeugt und der Schalldruckpegel des reflektierten Schalls fiel mit abnehmender Frequenz um 6 dB/Oktave (Dickreiter & Hoeg, 2014).

Eine Kombination dieser Faktoren zeigt, dass es nicht zielführend ist, tiefe Frequenzen über einen seitlich abstrahlenden Lautsprecher wiederzugeben, um an den Reflexionspunkten lokalisierbare Hörereignisse zu erwirken. Das resultierende Hörereignis würde aufgrund der in diesem Frequenzbereich unterhalb von 1600–2000 Hz dominanten ITD zu einer Lokalisation an der Primärschallquelle führen und somit einen Zerfall des Hörereignisses bei breitbandigen Signalen begünstigen.

Hooley (2006) beschreibt daher eine prototypische Soundbar, bei welcher die tiefen Frequenzen direkt und ohne Wandreflexion wiedergegeben werden, jedoch im Zeitbereich so verzögert werden, dass diese zeitgleich mit den indirekt über Wandreflexionen wiedergegebenen hohen Frequenzen beim Hörer eintreffen. So würde das resultierende Hörereignis an der Stelle des Reflexionspunktes wahrgenommen werden, da das menschliche Gehirn die einzelnen Komponenten wieder zu einem gemeinsamen Hörereignis zusammenfügen könne (Hooley, 2006). Ein Blick auf die mit phasenlinearen Oktavbandfiltern mit Mittenfrequenzen von 1000 Hz und 8000 Hz beaufschlagte ETC des linken Surround-Kanals der Soundbar zeigt, dass die im Rahmen der Arbeit untersuchte Soundbar ebenfalls eine derartige Verzögerung der über die frontal angeordneten Tief-/Mitteltöner wiedergegebenen Frequenzen vornimmt, damit diese zeitgleich mit den Reflexionen des Hochtöners am Abhörpunkt eintreffen. Andernfalls würden die tiefen Frequenzen aufgrund der nahen Anordnung der akustischen Treiber zueinander in etwa zeitgleich mit dem Direktschall der Hochtöner am Abhörpunkt eintreffen.

Abbildung 28:

Gefilterte ETC des Left Surround-Kanals der Soundbar



Anmerkung. Darstellung der mit zwei verschiedenen phasenlinearen Oktavbandfiltern beaufschlagten ETC des Kanals Ls der Soundbar. Die obige Darstellung wurde unter Verwendung eines Oktavbandfilters mit einer Mittenfrequenz von 8000 Hz erstellt, die untere mit einem Oktavbandfilter mit einer Mittenfrequenz von 1000 Hz. Eigene Darstellung.

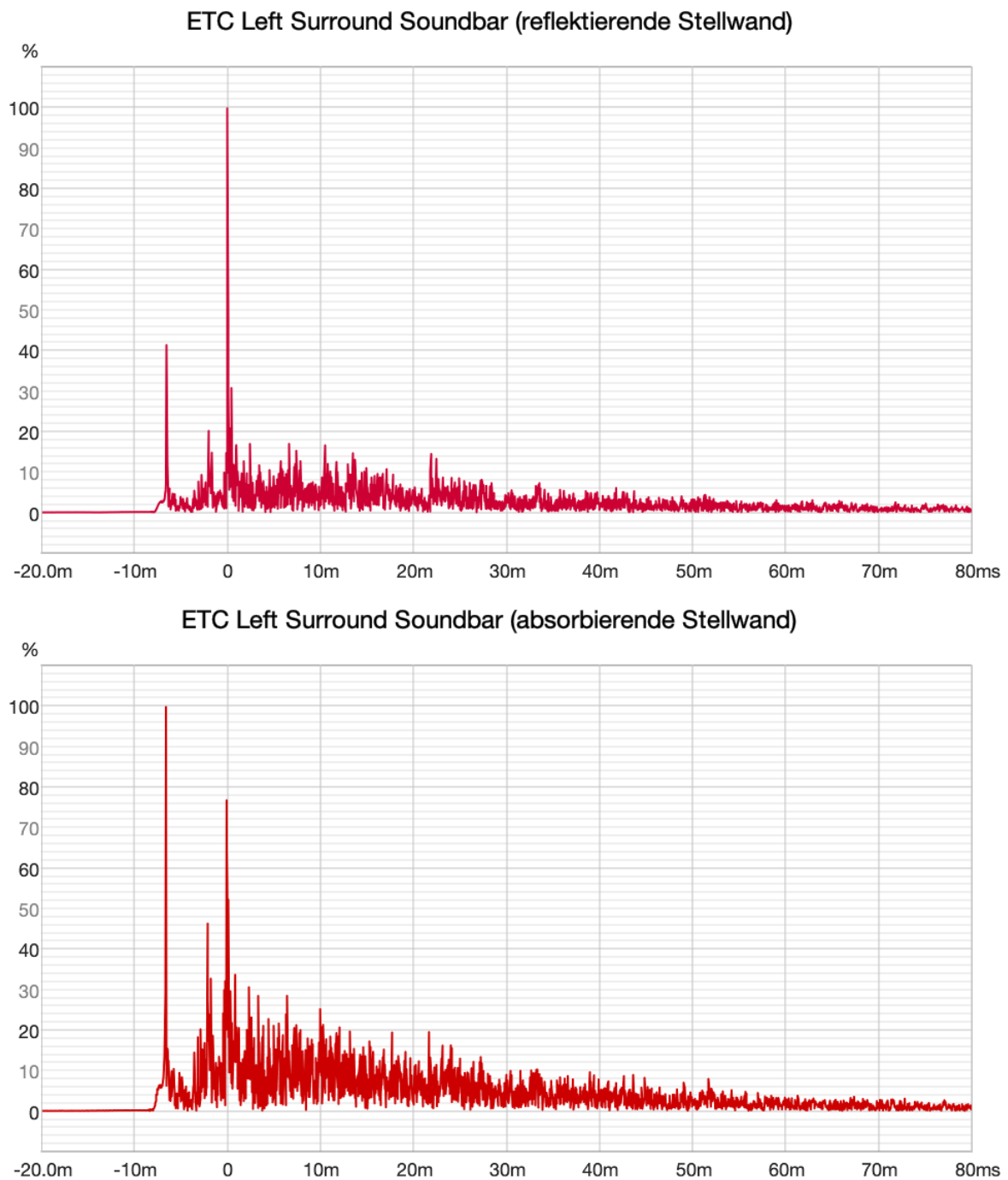
Dieses Vorgehen sorgt möglicherweise dafür, dass ein Zerfall des Hörereignisses in zwei scharf voneinander getrennte Hörereignisse verhindert werden kann.

Es kann folglich angenommen werden, dass der beobachtete Effekt der horizontalen, dynamischen Ausdehnung in ähnlicher Form auch bei anderen Soundbar-Modellen zu beobachten ist. In diesem Hinblick wären jedoch dringend weitere Untersuchungen notwendig, welche einen systematischen Vergleich verschiedener Soundbar-Modelle beinhalten, sowie verschiedene Distanzen zu den seitlichen Reflexionspunkten berücksichtigen. Im Folgenden wird auf die zeitgleich am Abhörpunkt eintreffende Überlagerung aus Reflexionsanteil des Hochtöners und Direktschallanteil der darauf verzögerten Tief-/Mitteltöner gesamtheitlich vereinfachend weiterhin als „Reflexion“ referenziert.

Hörbeispiel 6 und *Hörbeispiel 7* bieten sich an, die Notwendigkeit schallharter Oberflächen an den Reflexionspunkten zu illustrieren. Die Wiedergabe der Signale erfolgt hier alternierend zwischen Ls und Rs. Dreht man nun eine der Stellwände in der Versuchsumgebung auf die absorbierende Seite, so lässt sich auf dieser Seite ein diffus wirkendes Hörereignis am äußeren Rand neben der Soundbar lokalisieren. Die Azimuth-Auslenkungen der Hörereignisse sind gegenüber der Seite mit reflektierender Stellwand erheblich reduziert. Abbildung 29 illustriert die Ursache dieses wahrnehmbaren Effekts anhand zweier, jeweils normierter ETCs für den Kanal Ls der Soundbar – einmal mit schallhartem Reflexionspunkt, einmal mit Breitbandabsorber am Reflexionspunkt. Hier ist deutlich zu erkennen, dass die jeweils bei $t = 0$ ms liegende Erstreflexion bei absorbierender Stellwand energetische Verluste erfahren hat und in diesem konkreten Fall einen geringeren Schalldruckpegel am Abhörpunkt liefert als der Direktschall des seitlich abstrahlenden Hochtöners, was aufgrund des Präzedenzeffekts in einer deutlichen Verschiebung des Hörereignisses in Richtung der Primärschallquelle resultiert.

Abbildung 29:

ETCs des Ls-Kanals der Soundbar mit reflektivem vs. absorbierendem Reflexionspunkt



Anmerkung. Illustration des Einflusses des Reflexionspunktes auf die ICLD zwischen Direktschall und Erstreflexion. Im obigen Beispiel war eine schallharte Stellwand am Reflexionspunkt und der Schalldruckpegel der Reflexion lag ca. 8 dB über dem des Direktschalls. Im unteren Beispiel war ein Breitbandabsorber am Reflexionspunkt platziert. Der Schalldruckpegel des Direktschalls war nun um ca. 2 dB höher als der der Reflexion. Eigene Darstellung.

Die im Vergleich mit der diskreten Lautsprecherwiedergabe als höher empfundene räumliche Ausdehnung von Hörereignissen, welche bei einer Wiedergabe über einen Surround-Kanal der Soundbar wahrgenommen werden kann, lässt sich durch eine Gegenüberstellung des aus den BRIRs der beiden Wiedergabesysteme ermittelten Kriteriums $IACC_{\text{Early}}$ nachvollziehen. Dieses Kriterium korreliert nach Beranek (2004) mit dem subjektiven Attribut der ASW, wobei insbesondere die Oktavbänder mit Mittenfrequenzen von 500 Hz, 1000 Hz und 2000 Hz eine hohe Relevanz haben. Tabelle 2 stellt hierfür das arithmetische Mittel der jeweils aus den BRIRs der Kanäle Ls und Rs ermittelten $IACC_{\text{Early}}$ für Soundbar und diskrete Wiedergabe gegenüber.

Tabelle 2:

$IACC_{\text{Early}}$ von Soundbar und diskretem System für die Wiedergabe der Surround-Kanäle Ls und Rs

	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Soundbar	0,360	0,301	0,207	0,283	0,198	0,179	0,462	0,236
Diskret	0,998	0,906	0,769	0,796	0,542	0,330	0,733	0,133

Anmerkung. Gegenüberstellung der arithmetischen Mittel der $IACC_{\text{Early}}$ für die Wiedergabe der Surround-Kanäle Ls und Rs von Soundbar und diskreter Lautsprecherwiedergabe. Eigene Darstellung.

Hier lässt sich erkennen, dass das Kriterium für die Soundbar-Wiedergabe eines Surround-Kanals – mit Ausnahme des Oktavbands mit Mittenfrequenz von 8000 Hz – durchweg niedrigere Werte annimmt. Dies unterstützt die Beobachtung der bei der Soundbar-Wiedergabe gegenüber der diskreten Wiedergabe als höher empfundene ASW bei Ausgabe eines Signals über einen Surround-Kanal.

4.4.3 Top Surround-Kanäle

Für die Wiedergabe der Top Surround-Kanäle verfügt die Soundbar über zwei dedizierte, nach oben abstrahlende Breitband-Lautsprecher. Die Wiedergabe erfolgt – wie auch bei den Surround-Kanälen – indirekt und in diesem Fall über Deckenreflexionen.

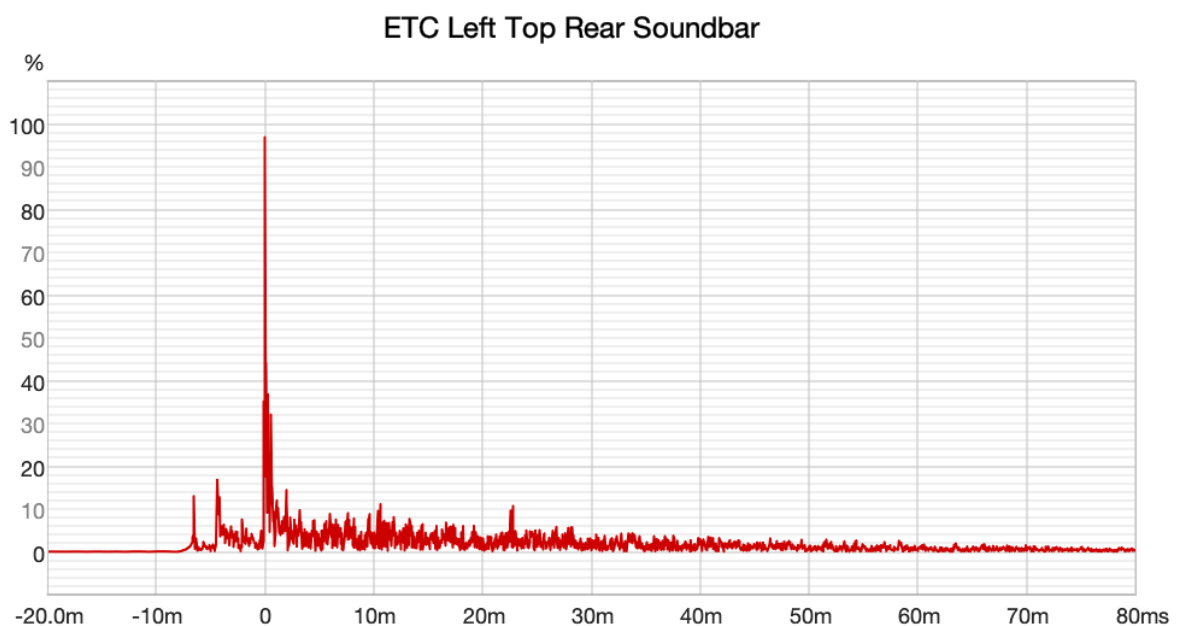
Das diskrete Wiedergabesystem verfügt im Vergleich über vier dedizierte Top Surround-Lautsprecher, deren Anordnung mit den in Kapitel 2.3.1.2 erläuterten Positionen für Ltf, Rtf, Ltr und Rtr übereinstimmt.

Hörbeispiel 11 liefert einen ersten Höreindruck anhand der kanal-sequenziellen Wiedergabe von pinkem Rauschen über die Kanäle Ltf, Ltr, Rtr und Rtf. Während bei der diskreten Wiedergabe eine scharfe Lokalisation an den Positionen der Lautsprecher möglich ist, die eine starke Differenzierung räumlicher Tiefe zwischen den Top Front- und Top Rear-Kanälen ermöglicht, ist dies bei der Soundbar-Wiedergabe so nicht zu beobachten. Dennoch ist ein Unterschied zwischen den Top Front- und Top Rear-

Kanälen der Soundbar festzustellen. Die Wiedergabe der Top Rear Surround-Kanäle erfolgt zusätzlich über die seitlich abstrahlenden Hochtöner, was beim Abspielen des Hörbeispiels mit geringem Pegel und einer Untersuchung der Soundbar aus nächster Nähe leicht überprüft werden kann. Diese Tatsache ist in der RIR von Kanal Ltr zudem in Form der zwei Direktschallanteile ersichtlich. Der erste ist dabei auf den seitlich abstrahlenden Lautsprecher zurückzuführen, dessen Reflexionspfad länger ist als der Reflexionspfad des nach oben abstrahlenden Lautsprechers über die Decke (siehe Abb. 30).

Abbildung 30:

ETC des Left Top Rear-Kanals der Soundbar



Anmerkung. Plot der ETC des Left Top Rear-Kanals der Soundbar. Zu erkennen sind die Direktschallanteile des Side-firing Speakers ($t = -6,5$ ms) und des Up-firing Speakers ($t = -4,5$ ms), sowie der die zeitgleiche Ankunft derer Reflexionen am Messpunkt bei $t = 0$ ms. Eigene Darstellung.

Hörbeispiel 12 ermöglicht einen Vergleich der Eigenschaften beider Systeme hinsichtlich der Wiedergabe transienter Signale (Snare-Samples). Ähnlich wie bei der Wiedergabe der Surround-Kanäle im Middle Layer lassen sich die Signale hier bei der diskreten Wiedergabe schärfer lokalisieren und die wahrgenommene Impulstreue ist höher als bei der Soundbar-Wiedergabe. Zudem wirkt die wahrgenommene Elevation bei der Soundbar bei Wiedergabe transienter Signalen geringer als bei breitbandigen Signalen zu sein. Die mit der Soundbar-Wiedergabe assoziierten Hörereignisse wirken darüber hinaus diffuser und räumlich ausgedehnter. Dieser Eindruck stellt sich ebenfalls bei einem Vergleich der Top-Surround-Kanäle der Soundbar mit den Front-Kanälen der Soundbar ein.

Hörbeispiel 13 und *Hörbeispiel 14* ermöglichen den direkten Vergleich anhand der alternierenden Wiedergabe von pinkem Rauschen bzw. Snare-Samples über die Kanäle L und Ltf.

Eine Berechnung des mit der ASW korrelierenden Kriteriums $IACC_{\text{Early}}$ für die Top Front-Kanäle von Soundbar und diskreter Lautsprecherwiedergabe unterstützt auch hier die Beobachtung der hinsichtlich der Soundbar-Wiedergabe als größer wahrgenommenen Ausdehnung von Hörereignissen, welche aus der Wiedergabe über diese Ausgangskanäle resultieren. Tabelle 3 stellt das arithmetische Mittel der jeweils aus den BRIRs der Kanäle Ltf und Rtf ermittelten $IACC_{\text{Early}}$ für Soundbar und diskrete Wiedergabe gegenüber.

Tabelle 3:

$IACC_{\text{Early}}$ von Soundbar und diskretem System für die Wiedergabe der Kanäle Ltf und Rtf

	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Soundbar	0,967	0,273	0,599	0,518	0,516	0,404	0,532	0,673
Diskret	0,972	0,944	0,817	0,635	0,604	0,562	0,639	0,334

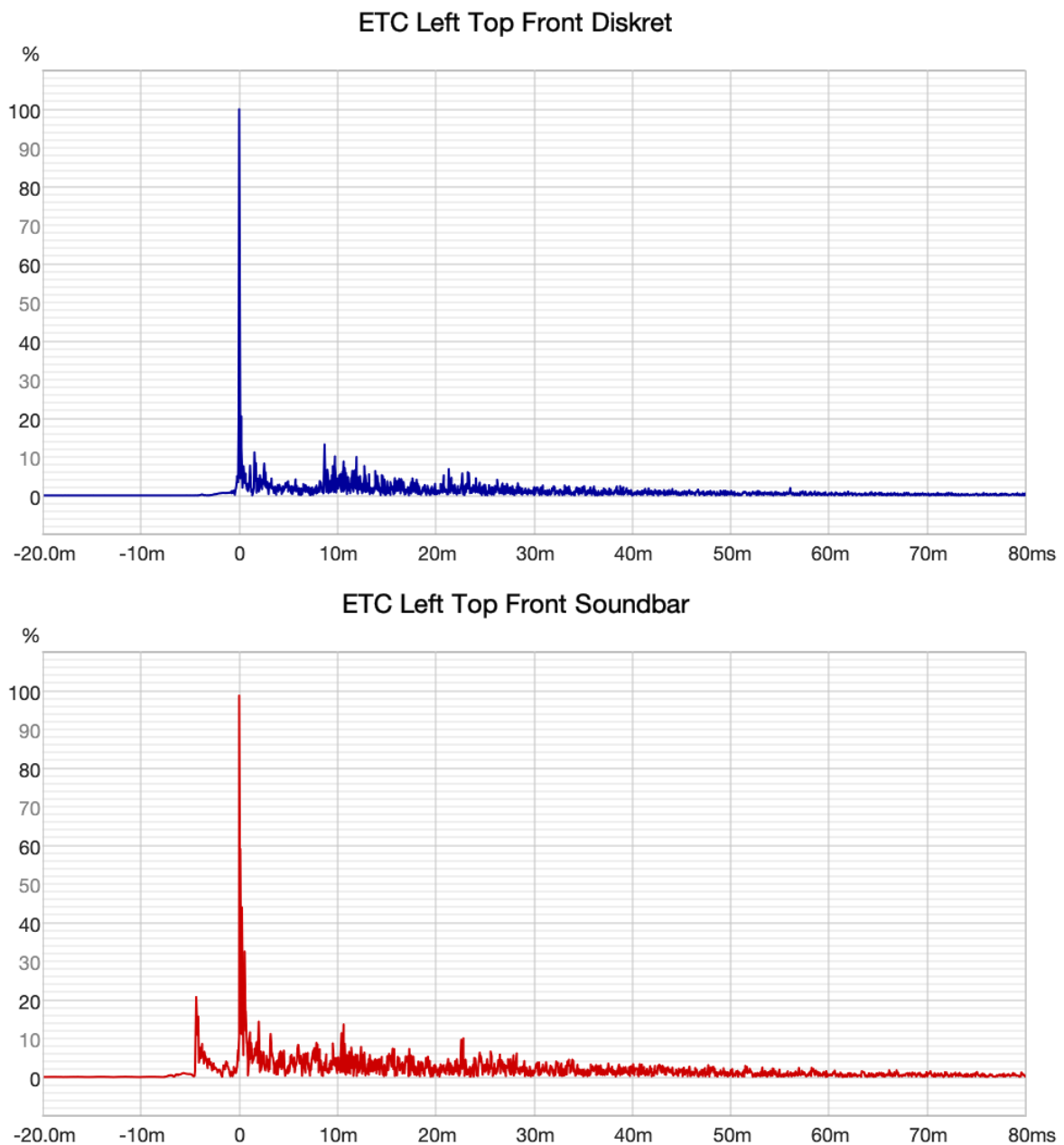
Anmerkung. Gegenüberstellung der arithmetischen Mittel der $IACC_{\text{Early}}$ für die Wiedergabe der Top Surround-Kanäle Ltf und Rtf von Soundbar und diskreter Lautsprecherwiedergabe. Eigene Darstellung.

Auch hier zeigt sich, dass das Kriterium hinsichtlich der Soundbar-Wiedergabe – mit Ausnahme des Oktavbands mit einer Mittenfrequenz von 8000 Hz – niedrigere Werte annimmt, was auf eine höher wahrgenommene ASW hindeutet. Die Unterschiede zwischen den Wiedergabesystemen in den nach Beranek (2004) besonderes relevanten Oktavbändern mit Mittenfrequenzen von 500 Hz, 1000 Hz und 2000 Hz fallen dabei jedoch geringer aus als bei den ermittelten Werten für die Surround-Kanäle.

Abbildung 31 visualisiert die normierten ETCs von Soundbar und diskretem System für die Wiedergabe ihres jeweiligen Kanals Ltf. Wie auch bei den Surround-Kanälen lässt sich hier erkennen, dass bei der Soundbar die frühen Reflexionen bis $t = 80$ ms im Verhältnis zur energiereichen Erstreflexion stärker ausgeprägt sind als die frühen Reflexionen im Verhältnis zum Direktschall bei der diskreten Wiedergabe. Die Unterschiede fallen hier jedoch geringer aus als bei der Wiedergabe der Surround-Kanäle. Zudem ist auch die ICLD zwischen Direktschall und Erstreflexion der nach oben abstrahlenden Lautsprecher der Soundbar mit ca 14 dB im Vergleich zur Wiedergabe der Surround-Kanäle stärker ausgeprägt.

Abbildung 31:

ETCs des Left Top Front-Kanals von diskretem System und Soundbar

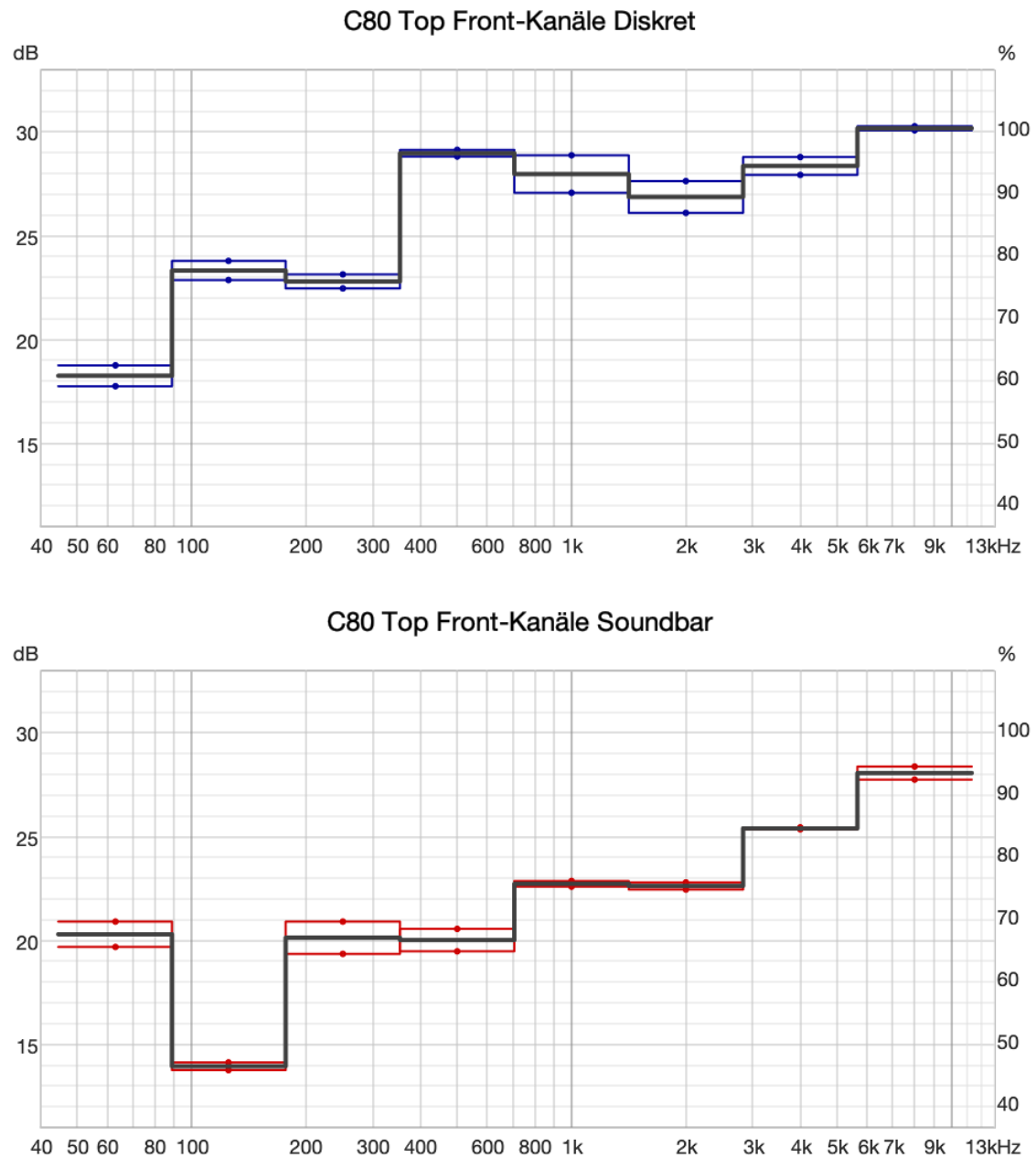


Anmerkung. Visualisierung der jeweils normierten ETCs für den Kanal Left Top Front von diskretem Mehrkanalsystem (blau) und Soundbar (rot) im Vergleich. Eigene Darstellung.

Auch die späten Reflexionen ab $t = 80 \text{ ms}$ sind bei der Soundbar-Wiedergabe der Top Surround-Kanäle im Verhältnis stärker ausgeprägt, was sich durch die Gegenüberstellung der über Oktavbänder gemittelten Klarheitsmaße C_{80} der Kanäle Ltf und Rtf von Soundbar und diskretem Wiedergabesystem nachvollziehen lässt (siehe Abb. 32). Insbesondere in den Mitten sind dabei deutliche Differenzen von bis zu 9 dB ablesbar.

Abbildung 32:

Klarheitsmaß C_{80} für die Top Front-Kanäle von diskretem System und Soundbar



Anmerkung. Gegenüberstellung der über Oktavbänder gemittelten Klarheitsmaße C_{80} für die Top Front-Kanäle Ltf und Rtf von diskretem Mehrkanalsystem (oben) und Soundbar (unten). Die jeweils schwarze Kurve illustriert dabei jeweils das arithmetische Mittel des für beide Top Front-Kanäle ermittelten Kriteriums (farbige Kurven). Eigene Darstellung.

Wie auch bei der Wiedergabe der Surround-Kanäle können der aus der ETC ersichtliche Primärschallanteil und die darauffolgende Deckenreflexion als zwei separate Schallquellen betrachtet werden, welche kohärente Signale abstrahlen (Blauert, 1996). Auch hier bestehen zwischen diesen Signalen gegensätzlich wirkende ICTD und ICLD: Der Primärschall erreicht den Hörer früher (ICTD), die Reflexion weist jedoch einen höheren Schalldruckpegel auf (ICLD).

Hörbeispiel 15 gibt ein transientes, breitbandiges Monosignal (Schlagzeug und Bass) über den Kanal Ltf wieder. Bei der Wiedergabe des gleichen Signals über den Kanal Ls (*Hörbeispiel 10*) war eine frequenzabhängige Ausdehnung des Hörereignisses festzustellen. Dies ist bei Wiedergabe über einen Top Surround-Kanal nicht zu beobachten.

Die Primärschallquelle und der Reflexionspunkt sind in Bezug auf den Abhörpunkt nahezu vertikal übereinander angeordnet und liegen so annähernd in der Medianebene. Aufgrund der hinreichenden Symmetrie des Kopfes und der Anordnung der Ohren nebeneinander, resultieren die ICTD und ICLD zwischen Primärschall und Reflexion daher nicht in vergleichbaren ITD und ILD, die binauralen Ohrsignalmerkmale sind in der Medianebene nur wenig ausgeprägt (Blauert, 1996; Wenzel et al., 2017). Lee (2011) kam in einer Studie zu der Erkenntnis, dass in der Vertikalen keine vergleichbare Beziehung zwischen ICTD und ICLD besteht, wie in der Horizontalen. Es konnte gezeigt werden, dass ein eleviertes Schallereignis in der Medianebene, welches um bis zu 5 ms gegenüber eines primären Schallereignisses in der Medianebene verzögert ist, zwischen 6–7 dB im Schalldruckpegel reduziert sein muss, damit das damit assoziierte Hörereignis vollständig an der Stelle des primären Schallereignisses lokalisiert wird. Der Präzedenzeffekt zeigt in der Vertikalen demnach keine Wirkung und die Auslenkung des Hörereignisses wird ausschließlich von der ICLD der beiden Quellen bestimmt (Lee, 2011). Ein Blick auf die in Abbildung 31 dargestellte ETC des Kanals Ltf der Soundbar zeigt eine ICTD zwischen Primärschall und Erstreflexion von weniger als 5 ms, weshalb sich diese Erkenntnis auch auf die hier betrachtete Situation übertragen lässt.

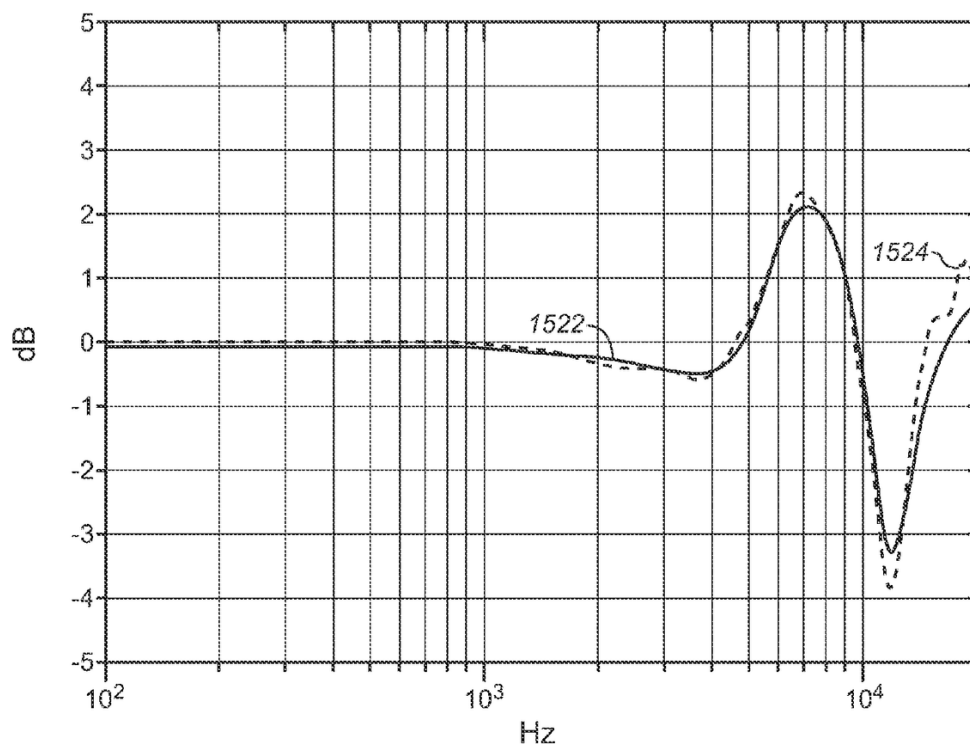
Entscheidend für die Lokalisation in der Medianebene sind primär die monauralen Ohrsignalmerkmale, d.h. die spektralen Differenzen der beiden Ohrsignale (Blauert, 1996). Hierbei hat auch die Art des Signals einen starken Einfluss auf dessen Lokalisation innerhalb der Medianebene, was durch *Hörbeispiel 16* demonstriert wird. In Anlehnung an die von Blauert (1996) ermittelten richtungsbestimmenden Bänder und den von Lee et al. (2014) beschriebenen *Pitch-Height Effect* erfolgt hier die alternierende Wiedergabe von Terzbandrauschen mit Mittenfrequenzen von 800 Hz und 8000 Hz über den Kanal Ltf. Für die Lokalisationsrichtung „oben“ ist nach Blauert (1996) insbesondere der Frequenzanteil um 8 kHz von Bedeutung. Lee et al. (2014) beschreiben darüber hinaus, dass die Lokalisation von Hörereignissen in der Vertikalen insbesondere dann an der Position der Schallquelle erfolgt, wenn das Signal breitbandig ist und Frequenzen oberhalb von 7 kHz enthält. Bei schmalbandigen Signalen hingegen sei die Lokalisation primär vom Signal selbst abhängig und weitgehend unabhängig von der tatsächlichen Position der Schallquelle (Lee et al., 2014).

Diese Erkenntnis scheint gleichermaßen für die Soundbar-Wiedergabe zutreffend zu sein. Es kann festgestellt werden, dass das tieffrequente der beiden Rauschsignale eine deutlich geringer wahrgenommene Elevations-Auslenkung aufweist als das hochfrequente. Dieser Effekt ist sowohl bei der Wiedergabe über die Soundbar als auch bei der Wiedergabe über den Deckenlautsprecher des diskreten Wiedergabesystems zu beobachten.

Dolby (2016a; 2018b) beschreibt in den Ausführungen zu Up-firing Speakern für die Wiedergabe der Top Surround-Kanäle einen sog. Pinna-Filter, welcher auf die darüber wiedergegebenen Signale angewendet wird. Diese Filterung soll demnach dazu beitragen können, die Auswirkung des Primärschallanteils auf die Hörereignisauslenkung zu reduzieren und den Eindruck von Elevation zu verstärken (Dolby Laboratories, 2016a; Dolby Laboratories, 2018b). Ein erteiltes Patent von Crockett et al. (2017) beschreibt einen derartigen, hier *Virtual Height Filter* genannten Filter mit einer charakteristischen Anhebung der Frequenzen um 7 kHz und einer Absenkung von Frequenzen um 12 kHz (siehe Abb. 33).

Abbildung 33:

Virtual Height Filter

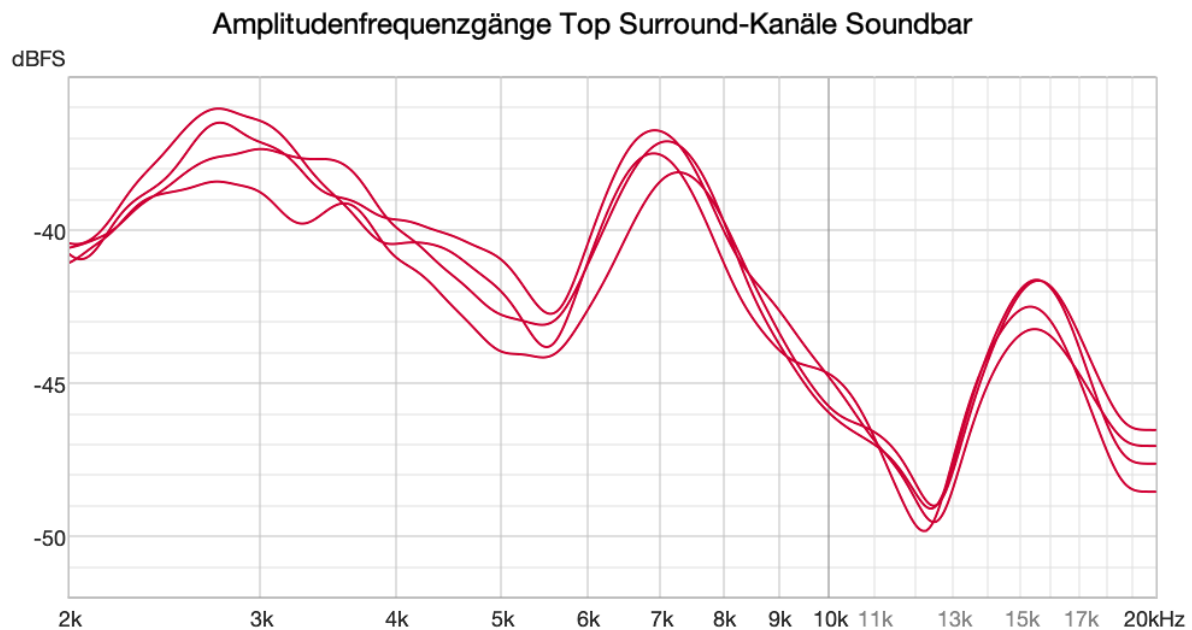


Anmerkung. Exemplarische Übertragungsfunktion eines sog. Virtual Height Filters. Adaptiert aus *Virtual Height Filter for Reflected Sound Rendering Using Upward Firing Drivers* (US 9,648,440 B2), von B. G. Crockett, C. Chabanne, M. Tuffy, A. J. Seefeldt, C. P. Brown & P. Turnmire, 2017, S. 12. United States Patent and Trademark Office.

Die im Rahmen der Arbeit gemessenen und geglätteten Amplitudenfrequenzgänge der Kanäle Ltf und Rtf der Soundbar weisen einen sehr ähnlichen Verlauf im entsprechenden Frequenzbereich auf (siehe Abb. 34).

Abbildung 34:

Ausschnitt der Amplitudenfrequenzgänge der Top Surround-Kanäle der Soundbar



Anmerkung. Darstellung der gemessenen und geglätteten Amplitudenfrequenzgänge der Kanäle Ltf, Rtf, Ltr und Rtr der Soundbar im Frequenzbereich zwischen 2 kHz und 20 kHz. Zu erkennen ist eine Anhebung der Frequenzen um 7 kHz und einer Absenkung der Frequenzen um 12 kHz. Eigene Darstellung.

Dies legt nahe, dass die verwendete Soundbar eine vergleichbare Filterung auf die Ausgangssignale der Top Surround-Kanäle anwendet, um die Lokalisation elevierter Hörereignisse zu unterstützen. Folglich ist eine Lokalisation von Hörereignissen in Richtung der Reflexionspunkte an der Decke insbesondere dann zu erwarten, wenn das dem Hörereignis zugrunde liegende Signal hinreichende Anteile in diesem Frequenzbereich aufweisen kann.

4.5 Mehrkanal-Wiedergabe – kohärente Signale

Als nächstes wird die simultane Wiedergabe kohärenter Signale über mehrere Ausgangskanäle betrachtet. Zuerst werden dabei statische Phantomschallquellen behandelt, im Anschluss Pan-Automationen.

4.5.1 Statisches Panning

Hinsichtlich des Pannings auf dem Middle Layer müssen die verschiedenen Downmix-Algorithmen für Dolby Atmos bedacht werden.

Für *Hörbeispiel 17* wurde ein pinkes Rauschsignal als Audioobjekt auf die Mitte der Lautsprecherbasis zwischen L und Ls gepannt. Damit sich dies gleichermaßen auf den für die Soundbar-Wiedergabe relevanten 5.1-Downmix überträgt, wurde der Algorithmus *Direct render with room balance* ausgewählt. Die Wiedergabe erfolgt so mit identischen Ausgangspegeln über die Kanäle L und Ls. Dabei ist ein interessanter Effekt zu wahrnehmbar: Bei der Soundbar-Wiedergabe ist erneut eine horizontale Ausdehnung des Hörereignisses festzustellen, wobei tieffrequente Signalanteile tendenziell mehr in Richtung der Primärschallquelle und hochfrequente Anteile eher in Richtung des Reflexionspunktes zu verorten sind. Aufgrund der zusätzlichen Wiedergabe des Signals über den linken Front-Kanal sind die hochfrequenten Anteile dabei aber mit geringerer Azimuth-Auslenkung zu verorten als bei der alleinigen Wiedergabe über den Kanal Ls – das ausgedehnte Hörereignis ist entsprechend horizontal „gestaucht“. Darüber hinaus sind bei Bewegung des Kopfes und Abweichungen vom Sweetspot Kammfiltereffekte wahrzunehmen.

Bei der diskreten Wiedergabe ist hingegen ein Zerfall des Hörereignisses zu beobachten, der Effekt ist hier jedoch umgekehrt. Die höheren Frequenzen werden dabei eher in Richtung des linken Front-Lautsprechers lokalisiert, während tiefere Frequenzen tendenziell mittig auf der Basis oder hinter dem Hörer verortet werden.

Die Wiedergabe eines breitbandigen und transientenbetonten Signals liefert einen ähnlichen Höreindruck, nachvollziehbar durch *Hörbeispiel 18*. Diese Beobachtung kann auf die rückseitige Abschottung der hohen Frequenzen des diskreten Surround-Lautsprechers durch die Pinna (Blauert, 1996) zurückgeführt werden. Eine Drehung des Kopfes und Bewegung um den Abhörpunkt offenbart dabei zudem eine Instabilität dieser seitlichen Phantomschallquelle. Das mit der Soundbar-Wiedergabe assoziierte Hörereignis wirkt im direkten Vergleich stabiler, was darauf zurückgeführt werden kann, dass sowohl Reflexionspunkt als auch der Lautsprecher des linken Front-Kanals bei Blickrichtung geradeaus vor der Frontal-Ebene des Hörers liegen und somit keine seitliche Phantomschallquelle vorliegt.

Hörbeispiel 19 demonstriert eine vertikale Phantomschallquelle nach dem gleichen Prinzip. Hierbei erfolgt die Wiedergabe monofonen pinken Rauschens mit identischen Pegeln über die Ausgangskanäle L und Ltf. Es fällt auf, dass das daraus resultierende Hörereignis bei der Soundbar-Wiedergabe nur eine geringe wahrnehmbare Elevation aufweist, sich nicht scharf lokalisieren lässt, und Bewegungen des Kopfes sich durch wahrnehmbare Kammfiltereffekte bemerkbar machen. Dieser Effekt ist auch bei der

diskreten Wiedergabe zu beobachten, wobei das Hörereignis auch hier nicht mittig zwischen den Lautsprechern für L und Ltf wahrgenommen wird, sondern tendenziell mit geringerer Elevation in Richtung des linken Front-Lautsprechers. Untersuchungen legen nahe, dass vertikale Phantomschallquellen im Hinblick auf die diskrete Wiedergabe insbesondere in der Medianebene interindividual lokalisiert werden (Pulkki, 2001) und dabei häufig auch nicht eindeutig lokalisierbar sind (Barbour, 2003). Zudem sei die Lokalisation statischer Phantomschallquellen in der Vertikalen nach Theile & Wittek (2011) signalabhängig und häufig von tonalen Färbungen begleitet, weshalb sie konstatieren, dass die Lautsprecher selbst die einzigen stabilen Positionen seien (Theile & Wittek, 2011). Im Hinblick auf die Soundbar-Wiedergabe muss dabei bedacht werden, dass auch die ausschließliche Wiedergabe über einen Top-Surround-Kanal aufgrund der Überlagerung von Direktschall und Reflexion eines Up-firing Speakers eine Abhängigkeit vom Abhörpunkt aufweist.

4.5.2 Dynamisches Panning

4.5.2.1 Dynamisches Panning im Middle Layer

Hörbeispiel 20 demonstriert eine langsame Pan-Automation zwischen den Objekt-Koordinaten $(-1, 1, 0)$ und $(-1, -1, 0)$ anhand der Wiedergabe pinken Rauschens. Dies äußert sich bei der diskreten Wiedergabe in 7.1.4 durch ein kontinuierliches Panning auf den Lautsprecher-Basen zwischen L, Ls und Lrs. Damit bei der Soundbar-Wiedergabe ebenfalls ein kontinuierliches Panning zwischen den Kanälen L und Ls erfolgt, wurde für den relevanten 5.1-Downmix die Downmix-Einstellung *Direct render* gewählt.

Bei der diskreten Wiedergabe fällt dabei auf, dass eine Lokalisation auf der gesamten Strecke zwischen vorne und hinten möglich ist, wobei im Bereich zwischen L und Ls ein sprunghaftes und instabiles Verhalten wahrgenommen werden kann. Dies lässt sich auf die steilen Lokalisationskurven seitlicher Phantomschallquellen zurückführen (Blauert & Braasch, 2008; Theile et al. 2014). Bei der Soundbar-Wiedergabe ist ebenfalls eine kontinuierliche Bewegung festzustellen, die von einem zunehmend diffuser werdenden Klangbild und einer steigenden räumlichen Ausdehnung des Hörereignisses begleitet wird, bevor es bei der Bewegung nach vorne wieder kompakter wird, um am Ende der Trajektorie wieder scharf lokalisierbar auf dem linken Front-Lautsprecher der Soundbar zu liegen.

Hörbeispiel 21 demonstriert eine Automation nach dem gleichen Prinzip für ein breitbandiges und transientes Monosignal (Schlagzeug und Bass), wobei die Transienten im vorderen, vom Direktschall geprägten Bereich präzise wiedergegeben werden und mit zunehmender Wiedergabe über den von der lateralen Reflexion bestimmten Kanal Ls merklich verwaschen.

Über die Positionierung von Quellen hinaus, bekommt der Panner so im Hinblick auf die Soundbar-Wiedergabe eine weitere klanggestalterische Bedeutung, da er aktiv das Verhältnis zwischen Direktschall und frühen Reflexionen plus Diffusschall beeinflusst. Der Klangeindruck wird dabei zunehmend ausgedehnter und diffuser, je weiter die Basis der Front-Lautsprecher verlassen wird.

Hörbeispiel 22 demonstriert eine Pan-Automation für ein monofones pinkes Rauschsignal mit kreisförmiger Trajektorie um den Hörer im Middle Layer. Für ein kontinuierliches Panning wurde auch hier ein Audioobjekt verwendet, sowie der Downmix-Algorithmus *Direct render* für den 5.1-Downmix. Bei der diskreten Wiedergabe in 7.1.4 ist das Hörereignis wie erwartet zu lokalisieren und es stellt sich der nachvollziehbare Eindruck einer kreisförmigen Bewegung um den Hörer ein. Die Bewegung wirkt kontinuierlich mit einer Tendenz zu sprunghaftem Verhalten zwischen den Front- und Surround-Lautsprechern, welches aufgrund der schnelleren Bewegung jedoch weniger prägnant wahrnehmbar ist. Die Wiedergabe des gleichen Hörbeispiels über die Soundbar führt zu einem erheblich abweichenden Höreindruck. Das Hörereignis ist auf den Front-Kanälen stets sehr scharf lokalisierbar, bekommt dann eine seitliche Auslenkung und wird dabei zunehmend diffuser und räumlich ausgedehnter. Bevor das Hörereignis jedoch hinter dem Hörer wahrgenommen werden kann, kippt es von rechts nach links über. Zu keinem Zeitpunkt der abgelaufenen Trajektorie, kann das Hörereignis hinter dem Hörer lokalisiert werden. Anschließend nimmt die räumliche Ausdehnung wieder kontinuierlich ab, bevor das Hörereignis wieder scharf lokalisierbar an der räumlichen Position der Soundbar verortet werden kann.

Hörbeispiel 23 demonstriert die gleiche Automation der Objekt-Koordinaten unter Verwendung des 5.1-Downmix-Algorithmus *Direct render with room balance*, wobei das Signal hier für je 1/8 der Periodendauer der kreisförmigen Bewegung auf dem Middle Layer statisch über Ls, respektive Rs wiedergegeben wird und das Amplitudenpanning zwischen den Surround- und Front-Kanälen in doppelter Geschwindigkeit erfolgt. Die Wahl des 5.1-Downmix-Algorithmus sollte daher stets an Kontext und Intention angepasst werden, wobei bedacht werden muss, dass der verwendete Algorithmus global wirksam ist und nicht selektiv für einzelne Audioobjekte bestimmt werden kann.

Hörbeispiel 24 illustriert abschließend eine Automation nach dem gleichen Prinzip beim Panning im kanalbasierten 7.1.2-Bed unter Verwendung des passiven *Lo/Ro*-Downmix mit statischen Koeffizienten. Hier sind deutlich wahrnehmbare Pegelsprünge des breitbandigen und annähernd pegelkonstanten monofonen Ausgangssignals (pinkes Rauschen) bei der Soundbar-Wiedergabe wahrnehmbar, welche auf diesen passiven Downmix zurückzuführen sind.

Alle drei Beispiele zeigen, dass bei der Soundbar-Wiedergabe keine vergleichbare Lokalisation für „hinten“ erfolgt, wie bei der diskreten Wiedergabe. Bei der diskreten Wiedergabe stellt sich dieser Eindruck insbesondere beim Monitoring in der 7.1.4-Konfiguration mit den zusätzlichen Rear Surround-Lautsprechern ein, scheint jedoch auch im 5.1-Downmix stärker ausgeprägt als bei der Soundbar-Wiedergabe. Bei dieser kippt das diffus in Richtung Reflexionspunkt des rechten Surrounds ausgelenkte Hörereignis auf die linke Seite über, bevor es in seiner wahrnehmbaren Ausdehnung bei der Bewegung nach vorne wieder kompakter wird und zwischen den Kanälen L, C und R scharf zu lokalisieren ist.

4.5.2.2 Dynamisches Panning zwischen Middle Layer und Upper Layer

Theile & Wittek (2011) konstatieren, dass statisches Panning in der Vertikalen keine verlässliche Lokalisierung gewährleiste, Bewegungen in gewissen Grenzen jedoch realisierbar seien.

Hörbeispiel 25 dient dazu diese Aussage im Hinblick auf die Soundbar-Wiedergabe zu überprüfen. Dabei wird ein monofones pinkes Rauschsignal zuerst langsam vom Kanal L des Middle Layers zum Kanal Ltf des Upper Layers gepannt, sowie anschließend wieder zurück. Bei der Wiedergabe des Beispiels über das diskrete System trifft die Aussage zu. Die vertikale Trajektorie des Rauschsignals kann nachvollzogen werden, wird dabei jedoch von kammfilterartigen Färbungen begleitet. Bei der Soundbar-Wiedergabe stellt sich ein ähnlicher Eindruck ein. Hier scheint die Bewegung jedoch weniger linear zu verlaufen. Die Elevations-Auslenkung bleibt zunächst gering, bevor im weiteren Verlauf der Automation eine später einsetzende vertikale Bewegung wahrgenommen werden kann. Die kammfilterartige Färbung wird hier zudem von einer zunehmenden Diffusität und einer höheren Ausdehnung des Hörereignisses im oberen Teil der Trajektorie begleitet.

Hörbeispiel 26 illustriert ein dynamisches Panning in der Medianebene, ebenfalls anhand eines monofonen pinken Rauschsignals. Hierbei wird der *spherical mode* des Dolby Atmos Music Panners genutzt, um eine Trajektorie in der Medianebene an den Randbereichen des Panners zu realisieren. Das Signal liegt zunächst auf dem Center-Kanal, wird zunehmend auf die Surround- und vorderen Top Surround-Kanäle verteilt, im Upper Layer zwischen vorderen und hinteren Top Surround-Kanälen gepannt und anschließend zunehmend auf die beiden Surround-Kanäle verteilt, bevor es mit identischen Ausgangspegeln über Ls und Rs im 5.1.x-Downmix, oder über Lrs und Rrs im Rendering auf ein 7.1.4-Layout wiedergegeben wird. Bei der diskreten Lautsprecherwiedergabe lässt sich die im Panner visualisierte Trajektorie wie erwartet nachvollziehen. Das Hörereignis liegt zunächst vor dem Hörer, wandert über ihn hinweg, bis es am Ende der Bewegung hinter dem Hörer zu lokalisieren ist und über den Hörer hinweg zurück auf den Center-Lautsprecher wandert. Auch hier stellt sich dieser Höreindruck insbesondere für die 7.1.4-Wiedergabe ein. Doch auch die Wiedergabe des 5.1.4-Downmix mit nur fünf Lautsprechern im Middle Layer ermöglicht eine Lokalisation hinter dem Hörer, wenn auch weniger scharf. Die Wiedergabe dieses Beispiels über die Soundbar kann als Grenzfall für die darstellerischen Möglichkeiten des Wiedergabesystems bezeichnet werden. Zunächst ist der Höreindruck mit der diskreten Lautsprecherwiedergabe vergleichbar, ab einer Elevation von ca. 60° scheint sich das Hörereignis jedoch in zwei Stränge aufzuteilen, welche am Ende der Trajektorie als zwei separierte und nicht eindeutig lokalisierbare Teil-Hörereignisse auf der rechten und linken Seite des Hörers wahrnehmbar sind. Sie können auch als die voneinander unabhängigen Randbereiche eines extrem ausgedehnten, frontal vor dem Hörer lokalisierbaren Hörereignisses wahrgenommen werden.

Es kommt demnach zu keiner vergleichbaren Phantomschallquellenbildung im hinteren Bereich des Panners. Wie auch bei der kreisförmigen Automation im Middle Layer stellt sich hier kein Höreindruck ein, welcher zu einer Lokalisation eines zusammenhängenden Hörereignisses hinter dem Hörer führt. Da die Wiedergabe im hinteren Teil der Bewegung zu gleichen Pegeln über die Ausgangskanäle Ls und Rs erfolgt, kann an dieser Stelle angenommen werden, dass die starken Reflexionen an den Seitenwänden eine derart hohe Dekorrelation der beiden Ohrsignale bewirken, dass die pegelidentische

Wiedergabe eines kohärenten Ausgangssignals über beide Surround-Kanäle der Soundbar keine scharfe Lokalisation von Phantomschallquellen zwischen den Surround-Kanälen bewirken kann, insbesondere nicht solche, die bei 180°-Azimuth hinter dem Hörer zu verorten sind. Dies deckt sich mit den Ausführungen von Toole (2008), wonach insbesondere die lateralen Reflexionen zu einer Dekorrelation der Ohrsignale beitragen. Die frühen lateralen Reflexionen bewirken dabei eine Erhöhung der subjektiven ASW, die späten sind relevant für die wahrgenommene LEV (Toole, 2008).

Die für beide Wiedergabesysteme aus der Überlagerung der jeweiligen gemessenen BRIRs der Ausgangskanäle Ls und Rs berechnete $IACC_{Early}$ illustriert die Dekorrelation der beiden Ohrsignale bei simultaner Wiedergabe über beide Surround-Kanäle. Die entsprechenden Werte sind in Tabelle 4 eingetragen und nehmen für die Soundbar-Wiedergabe durchweg niedrigere Werte an. In Kombination mit den Erkenntnissen von Faller & Merimaa (2004), nach denen eine Lokalisation nur dann möglich ist, wenn der Betrag der IC hoch genug ist, lässt sich die unterschiedliche Wahrnehmung von *Hörbeispiel 26* zwischen den beiden Wiedergabesystemen nachvollziehen.

Tabelle 4:

$IACC_{Early}$ von Soundbar und diskretem System für simultane Wiedergabe über die Kanäle Ls und Rs

	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Soundbar	0,982	0,251	0,712	0,371	0,291	0,184	0,448	0,410
Diskret	0,999	0,989	0,972	0,789	0,791	0,609	0,742	0,832

Anmerkung. Gegenüberstellung des aus der Überlagerung der jeweiligen BRIRs der Surround-Kanäle Ls und Rs berechneten Kriteriums $IACC_{Early}$ für Soundbar und diskrete Lautsprecherwiedergabe zur Illustration der durch simultane Wiedergabe über die Kanäle Ls und Rs unterschiedlich erfolgenden Dekorrelation der Ohrsignale. Eigene Darstellung.

Hörbeispiel 27 illustriert als nächstes eine Pan-Automation in der Frontalebene anhand des gleichen Prinzips unter Nutzung des *spherical mode*. Das breitbandige pinke Rauschsignal liegt dabei zuerst auf Ls, wird dann zusätzlich über Rs und die linken Top Surround-Kanäle wiedergegeben, wird im Upper Layer zwischen den linken und rechten Top Surround-Kanälen gepannt, bevor das Signal von den rechten Top Surround-Kanälen anteilig auf Ls und Rs verteilt wird und am Ende der Trajektorie ausschließlich über Rs wiedergegeben wird. Nach Barbour (2003) können elevierte Phantomschallquellen in der Frontalebene aufgrund der hier stärker ausgeprägten interauralen Ohrsignalmerkmale verlässlicher lokalisiert werden als in der Medianebene. Im Hinblick auf die diskrete Lautsprecherwiedergabe ist das Hörereignis zuerst scharf an Position des linken Surround-Lautsprechers bei 110° Azimuth lokalisierbar. Anschließend bewegt es sich mittig über den Hörer hinweg und kommt mit wieder zunehmender

Azimuth-Auslenkung an Position des rechten Surround-Lautsprechers scharf lokalisierbar zum Stehen. Anhand der kontinuierlichen Bewegung ist die aktuelle Position des Hörereignisses zu jedem Zeitpunkt nachvollziehbar. Bei der Wiedergabe des gleichen Hörbeispiels über die Soundbar ist das Hörereignis zunächst unscharf auf der linken Seite lokalisierbar. Mit Start der Bewegung sind merkliche kammfilterartige Klangfärbungen wahrnehmbar, die jeweils aktuelle Position lässt sich anhand der kontinuierlichen Bewegung jedoch ebenfalls nachvollziehen. Tendenziell wird das Hörereignis dabei insbesondere bei Wiedergabe über die Surround-Kanäle mit weniger Azimuth-Auslenkung lokalisiert. Dennoch lässt sich festhalten, dass die intentionierte Bewegung des Hörereignisses bei Wiedergabe über die Soundbar weitgehend erhalten bleibt, auch wenn die diskrete Lautsprecherwiedergabe eine schärfere Lokalisation ermöglicht.

4.5.2.3 Dynamisches Panning im Upper Layer

Hörbeispiel 28 dient der Illustration einer kontinuierlichen Pan-Automation im Upper Layer. Dabei wird ein mit monofonem pinkem Rauschen bespieltes Audioobjekt zuerst mit identischen Pegeln als Phantomschallquelle zwischen Ltf und Rtlf abgebildet, bewegt sich danach nach rechts auf Rtlf, wird anschließend zwischen Rtlf und Rtr in die hintere Hälfte des Upper Layers bewegt, zwischen Rtr und Ltr auf die linke Seite des Raums gepannt und bewegt sich dann wieder in die vordere Hälfte des Raumes in Richtung Ltf, bevor es am Ende einer Periode wieder mittig vorne im Upper Layer liegt. Bei der diskreten Lautsprecherwiedergabe lässt sich diese Trajektorie nachvollziehen und das resultierende Hörereignis ist jederzeit lokalisierbar. Insbesondere bei Verfolgung des Hörereignisses durch Bewegungen des Kopfes ist eine sehr scharfe Lokalisation auf den Basen der benachbarten Lautsprecher möglich.

Bei der Soundbar-Wiedergabe ist im direkten Vergleich eine Ambiguität des Hörereignisortes festzustellen. Während im vorderen Teil der Bewegung, obgleich der größer wahrgenommenen Ausdehnung, eine mit der diskreten Wiedergabe vergleichbare Lokalisation möglich ist, ist das Hörereignis insbesondere im hinteren Teil der Bewegung nicht eindeutig lokalisierbar. Die wahrgenommene Elevation nimmt dabei im hinteren Teil der Trajektorie ab und ähnlich wie bei den kreisförmigen Pan-Automatiken im Middle Layer lässt sich auch hier ein Kippen des Hörereignisses von rechts nach links feststellen. Auch im Upper Layer erfolgt demnach – im Gegensatz zur diskreten Wiedergabe – keine Lokalisation des Hörereignisses hinter der Frontalebene.

4.6 Mehrkanal-Wiedergabe – teilkohärente und inkohärente Signale

4.6.1 Object Size

Wie zuvor in Kapitel 3.1.1 erläutert, verfügt jedes Audioobjekt in Dolby Atmos über einen Object Size-Parameter, dessen Erhöhung zu einer zunehmenden Ausdehnung des Signals – mit vorheriger Dekorrelation durch den OAR – auf benachbarte Ausgangskanäle führt. Er wird in den Metadaten durch Werte zwischen 0 und 1 repräsentiert.

In *Hörbeispiel 29* wird der Object Size-Parameter eines auf den Center-Kanal gepannten, mit pinkem Rauschen bespielten Audioobjekts linear automatisiert. In den ersten vier Sekunden bleibt der Parameter statisch auf dem Wert 0 (Wiedergabe nur über den Center-Kanal), wird anschließend innerhalb von fünf Sekunden linear auf den Wert 0,5 erhöht, verbleibt vier Sekunden auf diesem Wert und wird anschließend wieder auf den Ausgangswert von 0 reduziert. Bei der Wiedergabe über das diskrete Mehrkanalsystem ist eine zunehmende Ausdehnung des initial scharf auf dem Center-Lautsprecher lokalisierbaren Hörereignisses zu vernehmen, der Parameter zeigt dementsprechend die erwartete Wirkung. Bei der Wiedergabe über die Soundbar zeigt der resultierende Höreindruck eine hohe Übereinstimmung mit der diskreten Wiedergabe. Auch hier ist eine kontinuierliche Ausdehnung des anfangs scharf in der Mitte der Soundbar lokalisierbaren Hörereignisses bemerkbar. Die Mitte des ausgedehnten Hörereignisses ist dabei in beiden Fällen frontal vor dem Hörer zu verorten. Die Intention überträgt sich demnach verlässlich auf die Wiedergabe über die Soundbar.

Hörbeispiel 30 demonstriert eine Automation nach dem gleichen Prinzip, hier ist das Signal jedoch auf den Kanal Ls gepannt. Bei der diskreten Lautsprecherwiedergabe äußert sich dies in einer zunehmenden Ausdehnung des initial scharf lokalisierbaren Hörereignisses hinten links, wobei die Lokalisation bei Erhöhung des Object Size-Parameters weniger eindeutig wird und das Hörereignis zunehmend zerfällt. Bei der Wiedergabe des gleichen Hörbeispiels über die Soundbar ist die Lokalisation des Signals von Beginn an nicht eindeutig, da hier bereits vorab eine horizontale Ausdehnung wahrgenommen werden kann. Mit Erhöhung des Object Size-Parameters scheint sich die Azimuth-Auslenkung des Hörereignisses zu reduzieren, sodass es innerhalb des Raums mehr in Richtung der Position der Soundbar verortet wird. Die durch die GUI des Dolby Atmos Music Panners suggerierte Wirkung der Ausdehnung eines statischen, lateral lokalisierbaren Hörereignisses überträgt sich demnach nicht entsprechend auf die Wiedergabe über die Soundbar.

4.6.3 Teilkohärente Signale

Als nächstes sollen die Eigenschaften der beiden Systeme im Hinblick auf die Wiedergabe von teilkohärenten Signalen über mehrere Kanäle verglichen werden. Hierzu wurden mehrere Spuren teilkohärenten pinken Rauschens durch Anwendung der Überlagerungsmethode erstellt (Blauert & Braasch, 2008). Jeder verwendete Ausgangskanal erhält dabei ein individuelles Rauschsignal, welches zu gleichen Teilen mit einem für alle Ausgangskanäle identischen Rauschsignal überlagert wurde. Somit

ergibt sich zwischen den einzelnen Kanälen im Mittel ein Kohärenzgrad von 0,5 (eng. inter-channel coherence, ICC; Pulkki & Karjalainen, 2015).

Hörbeispiel 33 illustriert nach diesem Prinzip die Wiedergabe teilkohärenter Signale über die Ausgangskanäle einer 5.1.4-Anordnung (ohne LFE). Dabei werden die Kanäle L, R, C, Ls, Rs, Ltf, Rtf, Ltr und Rtr mit jeweils einem individuellen Signal bespielt, welches zu allen anderen Kanälen eine mittlere ICC von 0,5 aufweist. Bei der Wiedergabe über das diskrete System lässt sich eine Umhüllung wahrnehmen, das Hörereignis lässt sich dabei initial keiner spezifischen Richtung zuordnen und der Schall scheint aus allen Richtungen auf den Hörer einzufallen. Bewegungen des Kopfes und eine Entfernung vom Sweetspot führen jedoch zu wahrnehmbaren Kammfiltereffekten. Diese sind auf die durch die Bewegung variabel veränderten Phasenbeziehungen der kohärenten Signalanteile zum Hörer zurückzuführen und liegen in den dadurch eingeführten Laufzeitdifferenzen begründet.

Die Wiedergabe des gleichen Hörbeispiels über die Soundbar führt ebenfalls zu einer hohen wahrgenommenen Umhüllung. Die bei der diskreten Wiedergabe wahrnehmbaren Kammfiltereffekte wirken jedoch weniger auffällig, was auf eine zusätzliche Dekorrelation der Signale durch die Wand- und Deckenreflexionen zurückgeführt werden kann. Darüber hinaus scheint der Höreindruck jedoch mehr auf den Bereich vor dem Hörer fixiert: Einerseits beschränkt sich der Höreindruck weitgehend auf den Bereich vor der Frontalebene, andererseits scheint er von den räumlich nah beieinander angeordneten, vom Direktschall bestimmten Kanälen L, R und C der Soundbar dominiert zu sein, was sich in Form eines diffus wirkenden, jedoch lokalisierbaren Teil-Hörereignisses frontal vor dem Hörer ausdrückt.

Hörbeispiel 32 illustriert vor diesem Hintergrund die Wiedergabe teilkohärenter Signale über die Ausgangskanäle Ls, Rs, Ltf, Rtf, Ltr und Rtr, d.h. die Wiedergabe erfolgt hier ohne Nutzung der Front-Kanäle. Bei der Soundbar-Wiedergabe ist der resultierende Höreindruck bei Wiedergabe des Hörbeispiels nicht von den Front-Kanälen dominiert, es stellt sich der Eindruck einer homogenen Umhüllung ein, welche jedoch auf den Bereich vor der Frontalebene des Hörers beschränkt ist. Ein „Loch“ im Bereich der räumlichen Position der Hochtöner der Front-Kanäle ist dabei nicht als solches wahrnehmbar, was unter anderem auf den von der Position der Soundbar auf den Hörer einfallenden Direktschallanteil und den zuvor erläuterten Effekt der horizontalen Ausdehnung von breitbandigen Hörereignissen, welche über die Surround-Kanäle wiedergegeben werden, zurückgeführt werden kann. Die diskrete Lautsprecherwiedergabe des Hörbeispiels führt zu einem vergleichbaren Höreindruck, dabei sind jedoch einerseits wieder verstärkt Kammfiltereffekte bei Bewegungen um den Abhörpunkt wahrnehmbar, andererseits ist im direkten Vergleich mit der Soundbar-Wiedergabe eher eine Unterrepräsentation der Einfallsrichtungen um die Positionen der hier weiter auseinander angeordneten Front-Lautsprecher feststellbar.

Hörbeispiel 33 soll die anhand vorheriger Hörbeispiele festgestellte Beschränkung der Soundbar-Wiedergabe auf den Bereich vor der Frontalebene des Hörers gezielter untersuchen. Hierzu wurden die teilkohärenten Signale auf Audioobjekte geroutet, deren Koordinaten mit den Positionen der Kanäle Lrs, Rrs, Ltr und Rtr einer 7.1.4-Anordnung korrespondieren. Diese liegen allesamt hinter der

Frontalebene eines im Sweetspot sitzenden Hörers. Im für die Soundbar-Wiedergabe relevanten 5.1.4-Downmix erfolgt dabei die Wiedergabe der Rear Surround-Signale über die Surround Kanäle Ls und Rs. Die Wiedergabe des Hörbeispiels über die diskrete 7.1.4-Lautsprecheranordnung bewirkt einen Höreindruck, der vom Bereich hinter der Frontalebene des Hörers dominiert wird. Dabei ist die unscharfe Lokalisation eines ausgedehnt wirkenden Hörereignisses hinter dem Hörer möglich, was insbesondere durch Drehbewegungen des Kopfes bestätigt wird. Die Wiedergabe des 5.1.4-Downmixes über das diskrete Wiedergabesystem lässt eine derartige Lokalisation nicht mehr zu, es bleibt jedoch der Eindruck eines überwiegenden Schalleinfalls aus dem Bereich hinter der Frontalebene bestehen.

Die Wiedergabe über die Soundbar führt zu einem erheblich abweichenden Höreindruck. Es stellt sich ein stärkerer Eindruck von räumlicher Umhüllung ein, dieser beschränkt sich aber auch hier auf den Bereich vor der Frontalebene, was sich insbesondere durch Drehbewegungen des Kopfes bestätigt.

4.6.2 Inkohärente Signale

Im abschließenden Teil der praktischen Untersuchung wird die Wiedergabe inkohärenter Signale über mehrere Ausgangskanäle betrachtet. Hierfür wurde für jeden verwendeten Ausgangskanal ein individuelles, zufälliges pinkes Rauschsignal erzeugt. Die ICC zwischen den Wiedergabekanälen der hier vorgestellten Hörbeispiele beträgt demnach im Mittel den Wert 0.

Hörbeispiel 34 illustriert die Wiedergabe zweier individueller und inkohärenter Signale über die Ausgangskanäle Ls und Rs. Bei der Wiedergabe über die Soundbar lässt sich das daraus resultierende Hörereignis als horizontal weit ausgedehnt und diffus beschreiben. Während insbesondere für hohe Frequenzen zwei räumlich separierte, jedoch jeweils sehr unscharf lokalisierbare Hörereignisse in Richtung der beiden Reflexionspunkte wahrgenommen werden können, wirken tiefere Frequenzen tendenziell frontal verankert. Dies äußert sich jedoch nicht in einem wahrgenommenen Zerfall des Hörereignisses in drei Teile, sondern in einer breiten frequenzabhängigen Ausdehnung vor dem Hörer. Diese wahrgenommene Ausdehnung scheint dabei auch nicht auf den Sektor zwischen den beiden Reflexionspunkten beschränkt zu sein, sondern deckt annähernd den gesamten Bereich zwischen $\pm 90^\circ$ Azimuth ab.

Die diskrete Lautsprecherwiedergabe des Hörbeispiels bewirkt einen abweichenden Höreindruck. Die beiden Signale sind scharf voneinander getrennt an den Positionen der Lautsprecher lokalisierbar und in ihrer räumlichen Ausdehnung deutlich kompakter. Während bei der Wiedergabe über die zwei Ausgangskanäle der Soundbar eine starke Ausdehnung wahrnehmbar war, scheint das mit der diskreten Wiedergabe assoziierte Hörereignis in seiner Ausdehnung auf die Verbindungslinie zwischen den beiden Surround-Lautsprechern beschränkt zu sein.

Hörbeispiel 35 illustriert die Wiedergabe inkohärenter pinker Rauschsignale über die Kanäle Ltf und Rtf nach dem gleichen Prinzip. Bei der Wiedergabe über die Soundbar wird dabei ein eleviertes, diffus wirkendes Hörereignis wahrgenommen, welches horizontal auf den Bereich zwischen ca. ± 30 – 45°

Azimuth ausgedehnt zu sein scheint. Es ist hier initial keine scharfe Differenzierung zwischen zwei separaten Hörereignissen möglich. Bewegungen des Kopfes und eine räumliche Bewegung um den Abhörpunkt begünstigen zwar eine Unterscheidung der beiden Kanäle, die Fusion der beiden wiedergegebenen Signale zu einem gemeinsamen Hörereignis kann dadurch jedoch nicht gänzlich aufgehoben werden.

Die Wiedergabe des Hörbeispiels über das diskrete Wiedergabesystem lässt im direkten Vergleich umgehend eine Identifikation von zwei voneinander unabhängigen Hörereignissen an den Positionen der beiden Deckenlautsprecher zu, welche jeweils scharf lokalisierbar sind. Die beiden Hörereignisse wirken dabei kompakter und nicht vergleichbar ausgedehnt.

Hörbeispiel 36 stellt eine Kombination der beiden vorherigen Hörbeispiele dar und gibt simultan individuelle, zueinander inkohärente Rauschsignale über die Kanäle Ls, Rs, Ltf und Rtf aus. Bei der Soundbar-Wiedergabe reichen diese vier Signale aus, um den subjektiven Eindruck einer Umhüllung des Hörers zu bewirken. Zwar bestätigt sich auch in diesem Beispiel – insbesondere durch Drehung des Kopfes – der Eindruck, dass die Einfallsrichtungen des Schalls auf den Bereich vor der Frontalebene beschränkt sind. Innerhalb dieses Bereiches scheint es jedoch keine auffällige Über-oder Unterrepräsentation bestimmter Schalleinfallsrichtungen zu geben.

Bei der diskreten Lautsprecherwiedergabe kann keine derartige Umhüllung festgestellt werden. Die Lautsprecher sind hier – insbesondere durch Peilbewegungen des Kopfes – eher als Quellen der Hörereignisse zu identifizieren. Dies fällt jedoch schwerer als in den vorherigen Beispielen, in welchen die Wiedergabe der Surround- und Top Surround Kanäle isoliert betrachtet wurde. Der Bereich um die räumliche Lage der Front-Lautsprecher im Middle Layer wirkt im direkten Vergleich mit der Soundbar jedoch unterrepräsentiert und auch der Bereich zwischen ca. 60–90° Elevation über dem Hörer wirkt im direkten Vergleich mit der Soundbar-Wiedergabe weniger ausgefüllt.

Die $IACC_{Late}$ mit den Integrationszeitgrenzen bei $t_1 = 80$ ms und $t_2 = 500$ ms korreliert nach Banarek (2004) mit dem subjektiven Attribut der Umhüllung bzw. LEV. In diesem Zusammenhang wurde durch Überlagerung der im Rahmen der Arbeit gemessenen BRIRs für die Kanäle Ls, Rs, Ltf und Rtf des jeweiligen Wiedergabesystems versucht, den beschriebenen Höreindruck der hinsichtlich der Soundbar-Wiedergabe als stärker wahrgenommenen Eindruck der auditiven Umhüllung bei simultaner Wiedergabe inkohärenter Signale über die entsprechenden Kanäle anhand dieses Messkriteriums nachvollziehen zu können. Tabelle 5 zeigt eine Gegenüberstellung des berechneten Kriteriums für Soundbar und diskretes Mehrkanalsystem.

Tabelle 5:

IACC_{Late} von Soundbar und diskretem System für simultane Wiedergabe über die Kanäle Ls, Rs, Ltf und Rtf

	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Soundbar	0,983	0,566	0,761	0,152	0,338	0,178	0,177	0,121
Diskret	0,990	0,963	0,472	0,383	0,264	0,182	0,191	0,174

Anmerkung. Gegenüberstellung des aus der Überlagerung der jeweiligen BRIRs der Surround- und Top Surround-Kanäle Ls, Rs, Ltf und Rtf für Oktavbänder berechneten Kriteriums IACC_{Late} für Soundbar und diskrete Lautsprecherwiedergabe. Eigene Darstellung.

Die ermittelten Werte zeigen insbesondere für die hohen Oktavbänder eine jeweils hohe Dekorrelation der Ohrsignale und nur geringe Differenzen zwischen diskreter Wiedergabe und Soundbar-Wiedergabe. In den Oktavbändern mit Mittenfrequenzen von 250 Hz und 1000 Hz nimmt die IACC_{Late} für die diskrete Wiedergabe geringere Werte an. Die als höher empfundene LEV bei der Wiedergabe der entsprechenden Kanäle über die Soundbar lässt sich anhand der Berechnung dieses Kriteriums daher nicht stützen. Beranek (2004) nennt die IACC_{Late} zwar als Indikator für die LEV, betont aber zugleich, dass dieser nicht verlässlich ist. Griesinger (1999) nennt interaurale Fluktuationen der ITD und ILD ab $t = 160$ ms als maßgebliche Faktoren für die LEV und empfiehlt insbesondere für kleine Räume die Ermittlung der Diffusfeldtransferfunktion (engl. Diffuse Field Transfer Function, DFT) als Messgröße für die LEV. Diese wurde im Rahmen der Arbeit nicht ermittelt.

Hörbeispiel 37 illustriert abschließend die simultane Wiedergabe individuellen pinken Rauschens über alle Ausgangskanäle des diskreten 7.1.4-Wiedergabesystems mit Ausnahme des Center-Kanals und des LFE. Dies resultiert bei der diskreten Wiedergabe in einem stark umhüllenden Höreindruck, welchem kein singulärer Ursprung zugewiesen werden kann. Bewegungen des Kopfes lassen allerdings die Erkenntnis zu, dass der Schalleinfall aus dem Sektor zwischen den Front- und Surround-Kanälen im Middle Layer (± 30 – 110° Azimuth) reduziert ist. Um bei der Wiedergabe über die Soundbar eine Dominanz der Front-Kanäle zu verhindern, welche in *Hörbeispiel 31* attestiert wurde, wurde für dieses Beispiel der Algorithmus *Direct render with room balance* für den 5.1-Downmix verwendet. Dies bewirkt, dass die Anteile der in dieser Konfiguration nicht vorhandenen Rear Surround-Kanäle Lrs und Rrs im Downmix auf den Surround-Kanälen Ls und Rs wiedergegeben werden, die Anteile der Kanäle Ls und Rs jedoch ebenfalls auf Ls und Rs verbleiben. Bei Anwendung von *Direct render* würden diese als Phantomschallquelle zwischen den Front- und Surround-Kanälen wiedergegeben werden, was den Ausgangspegel der Front-Kanäle folglich erhöht hätte. Die Wiedergabe von *Hörbeispiel 37* über die Soundbar zeigt im direkten Vergleich mit der diskreten Wiedergabe eine ebenfalls hohe Umhüllung, wobei

der angesprochene Sektor zwischen ± 30 – 110° Azimuth hier homogener ausgefüllt scheint. Im direkten Vergleich mit der diskreten Lautsprecherwiedergabe ist auch hier erneut auffällig, dass Schalleinfallrichtungen aus dem Bereich hinter der Frontalebene durch die Soundbar nicht überzeugend repräsentiert werden können.

5. Diskussion

5.1 Zusammenfassung

Die Arbeit konnte verschiedene Gemeinsamkeiten und Unterschiede hinsichtlich der Wiedergabe von immersiven Audioinhalten zwischen diskreter Lautsprecherwiedergabe in der Tonregie und der Wiedergabe über eine Soundbar identifizieren. Ein Verständnis für derartige Differenzen ist nützlich, um den Transfer der klanggestalterischen Intention auf das Wiedergabesystem Soundbar besser antizipieren zu können und auf dieser Grundlage gezielte Entscheidungen innerhalb des Postproduktionsprozesses treffen zu können.

Einerseits bieten sich Tonschaffenden verschiedene Möglichkeiten, um im Postproduktionsprozess die eigene Monitoring-Situation auf Bedingungen der endnutzerbezogenen Wiedergabe anzupassen: Eine Dolby Atmos-Mischung kann so beispielsweise über verschiedene Sub-Layouts der verwendeten diskreten Lautsprecheranordnung abgehört werden, um so in Abhängigkeit vom produzierten Inhalt diejenigen Downmix- und Trim-Einstellungen auszuwählen, die bei Wiedergabe über kompaktere Ausgangskonfigurationen zur höchsten Übereinstimmung mit der ursprünglichen Intention führen. Insbesondere der 5.1.x-Downmix hat dabei eine hohe Relevanz für die Soundbar-Wiedergabe. Hinsichtlich der Encodierung von Dolby Atmos-Inhalten in die für die Distribution erforderlichen Bitstream-Formate bieten sich zudem verschiedene Optionen mit verschiedenen Eigenschaften. Dabei muss jedoch beachtet werden, dass Tonschaffende häufig nicht gezielten Einfluss auf den Distributionsweg der erstellten Inhalte nehmen können und sich ggf. an externen Richtlinien zu orientieren haben. Ein Verständnis für die im Rahmen der Arbeit vorgestellten Eigenschaften der verschiedenen Codecs kann jedoch hilfreich sein, um deren Einfluss auf die produzierten immersiven Audioinhalte besser abschätzen zu können. So könnte beispielsweise die Anzahl der simultan verwendeten Audioobjekte bewusst reduziert werden, um etwaigen Auswirkungen von Spatial Coding und der damit verbundenen Reduktion der räumlichen Auflösung entgegenzuwirken.

Die entscheidenden Diskrepanzen zwischen diskreter Lautsprecherwiedergabe in der Tonregie und der Soundbar-Wiedergabe beim Endnutzer können jedoch auf Unterschiede im Hinblick auf die akustische Wiedergabe der Audioinhalte zurückgeführt werden. Während die jeweils vom Direktschall bestimmten Front-Kanäle beider Systeme im Rahmen der Untersuchung eine hohe Übereinstimmung zeigten,

waren es insbesondere die Surround- und Top Surround-Kanäle, die in ihren Wiedergabeeigenschaften teils erhebliche Unterschiede offenbarten. Die Wiedergabe dieser Ausgangskanäle erfolgt bei der Soundbar unter der gezielten Nutzung von Wand- und Deckenreflexionen, woraus sich in Konsequenz eine große Abhängigkeit vom Wiedergaberaum ergibt. Dies steht im Gegensatz zur diskreten Lautsprecherwiedergabe, bei welcher insbesondere im Kontext des Monitorings in Tonregionen eine Unterdrückung früher Reflexionen am Abhörpunkt angestrebt wird und die Wiedergabe so weitgehend vom Direktschall der Lautsprecher bestimmt ist.

Die subjektiven Höreindrücke des Autors deuten darauf hin, dass es teilweise große Abweichungen hinsichtlich der Lokalisation von Hörereignissen zwischen den beiden Wiedergabesystemen gibt, welche tendenziell umso größer zu sein scheinen, je weiter die Basis der Front-Kanäle L, C und R verlassen wird. Diese Beobachtungen können einerseits auf die Lage der für die Soundbar-Wiedergabe relevanten Reflexionspunkte zurückgeführt werden, welche – insbesondere für die Surround- und Top Rear-Surround-Kanäle – stark von den räumlichen Positionen der damit korrespondierenden diskreten Lautsprecher abweichen. Andererseits resultiert die indirekte Wiedergabe der Soundbar über Begrenzungsflächen des Raumes auch in einer höheren Diffusität von damit assoziierten Hörereignissen. Gegenüber der vom Direktschall bestimmten diskreten Wiedergabe drückt sich dies in einer reduzierten Lokalisationsschärfe und Impulstreue, sowie einer meist höher wahrgenommenen räumlichen Ausdehnung von Hörereignissen aus.

Aus der indirekten Wiedergabe der Surround- und Top Surround-Kanäle der Soundbar ergibt sich, dass ein derartig wiedergegebener Ausgangskanal durch die Überlagerung eines vorausseilenden Direktschallanteils mit einem verzögerten, jedoch im Schalldruckpegel höheren Reflexionsanteil charakterisiert werden kann. Diese Überlagerung scheint sich dabei jedoch unterschiedlich auf die Wahrnehmung der Surround- bzw. Top Surround-Kanäle auszuwirken.

In der Horizontalen sind die interauralen Ohrsignalmerkmale stärker ausgeprägt und die gegensätzlich auf die Hörereignisauslenkung wirkenden Pegel- und Laufzeitdifferenzen zwischen Direktschall und Reflexion der seitlich abstrahlenden Lautsprecher der Surround-Kanäle der Soundbar äußerten sich in Kombination mit den auf den primär hochfrequenten Reflexionsanteil verzögerten Tiefen- und Mittenfrequenzen in einem komplexen Zusammenwirken auf die laterale Auslenkung von Hörereignissen. Dieser Zusammenhang konnte im Rahmen der Arbeit nicht vollständig erfasst werden, was eine verlässliche Antizipation der aus der Soundbar-Wiedergabe resultierenden Hörereignisauslenkung erschwert. Es konnten beispielsweise Tendenzen beobachtet werden, dass tieffrequente Signale oder Signale mit transientem Charakter mit weniger Azimuth-Auslenkung wahrgenommen werden können als Signale mit ausgeprägten hochfrequenten Anteilen und flächiger Charakteristik. Dabei schien jedoch auch die Beziehung der einzelnen Komponenten eines Frequenzgemischs zueinander deren Lokalisierbarkeit dynamisch zu beeinflussen, was sich mit Ausführungen von Hooley (2006) deckt.

Hinsichtlich der Top Surround-Kanäle schienen die Laufzeitunterschiede zwischen Direktschall und Reflexion aufgrund ihrer in der Vertikalen geringeren Auswirkung auf die interauralen Ohrsignalmerkmale einen reduzierten Einfluss auf die Lokalisation darüber wiedergegebener Signale zu haben. Es konnten Indizien ausfindig gemacht werden, die nahelegen, dass die hinsichtlich der diskreten Lautsprecherwiedergabe für die Auslenkung elevierter Hörereignisse relevanten psychoakustischen Effekte wie richtungsbestimmende Bänder (Blauert, 1996) und der Pitch-Height Effect (Lee et al., 2014) auch hinsichtlich der Soundbar-Wiedergabe hohe Relevanz haben. Da der Präzedenzeffekt in der Vertikalen nach Lee (2011) keine Wirkung zeigt, ist für ein eleviertes Hörereignis demnach auch primär die ICLD zwischen Direktschall und Reflexion der Up-firing Lautsprecher der Soundbar entscheidend. Hinsichtlich transienter Signale mit Elevation konnten ebenfalls Unterschiede zwischen Soundbar-Wiedergabe und diskreter Wiedergabe beobachtet werden. Bei Wiedergabe über die Soundbar schienen diese eine im Vergleich geringere Elevation aufzuweisen als bei Wiedergabe über die Deckenlautsprecher des diskreten Wiedergabesystems.

Zudem wurde beobachtet, dass es eine Sweetspot-Abhängigkeit für die Lokalisation von Hörereignissen zu geben scheint, welche über einen einzigen Surround- oder Top Surround-Kanal der Soundbar wiedergegeben werden. Während die Lokalisation bei der diskreten Wiedergabe in diesem Fall unabhängig von der Position des Hörers grundsätzlich an Position des Lautsprechers erfolgte, wurde bei der Soundbar-Wiedergabe die wechselseitige Abhängigkeit der Hörereignisauslenkung vom Reflexionsanteil- und Direktschallanteil für den beobachteten Effekt verantwortlich gemacht, deren Verhältnis durch eine Änderung der Abhörposition beeinflusst wird.

Der wohl prägnanteste Unterschied zwischen den beiden Wiedergabesystemen lag in deren Fähigkeiten Hörereignisse hinter der Frontalebene des Hörers zu erzeugen. Die diskrete Lautsprecherwiedergabe ermöglichte insbesondere unter Verwendung der Rear Surround-Lautsprecher scharf lokalisierbare Hörereignisse hinter der Frontalebene – sowohl statisch an den Positionen der Lautsprecher als auch in Form von statischen und dynamischen Phantomschallquellen. Mit der Soundbar war dies im Rahmen dieser Untersuchung nicht möglich. Dies steht im Widerspruch zum häufig im Kontext von Produktbeschreibungen von Soundbars anzutreffenden Marketing-Jargon. Auch die im Rahmen der Arbeit untersuchte Soundbar wird von Seiten des Herstellers mit einem Klangerlebnis beworben welches „den Eindruck vermittelt, die Lautsprecher seien auf der Rückseite des Raums aufgestellt“ (Sennheiser, 2025b). Der Umstand kann prinzipiell auf die Lage der Erstreflexionspunkte zurückgeführt werden. Die für eine Lokalisation in Richtung des Reflexionspunktes notwendige Pegeldifferenz zwischen Direktschall und Reflexion kann unter normalen Umständen und unter Beachtung des Reflexionsgesetzes nur für solche Erstreflexionen erfolgen, deren Reflexionspunkte vor der Frontalebene des Hörers liegen. So lässt sich erklären, weshalb viele der am Markt verfügbaren Soundbars um externe Rücklautsprecher ergänzt werden. Diese werden hinter dem Hörer platziert, um so auch eine Lokalisation von Hörereignissen hinter der Frontalebene ermöglichen zu können.

Auffällig war zudem die wahrgenommene Ausdehnung von Quellen bei Wiedergabe über die Surround- und Top Surround-Kanäle der Soundbar. Je mehr ein monofones Signal anteilmäßig auf diese Kanäle verteilt wurde, desto diffuser, ausgedehnter und unschärfer lokalisierbar wurde es. Diese subjektiven Beobachtungen konnten durch die Ermittlung des mit der Apparent Source Width korrelierenden Kriteriums $IACC_{Early}$ gestützt werden, welches hinsichtlich der Soundbar-Wiedergabe der Surround- und Top Surround-Kanäle geringere Werte annahm und so auf eine gegenüber den korrespondierenden Kanälen des diskreten Wiedergabesystems reduzierte Lokalisierbarkeit und erhöhte ASW hindeutet. Im Hinblick auf dynamisches Panning von Signalen abseits der scharf lokalisierbaren Front-Kanäle bekommt der Panner im Kontext der Soundbar-Wiedergabe so eine zusätzliche klanggestalterische Bedeutung, da das Panning neben der räumlichen Repräsentation auch aktiv die Diffusität, Lokalisierbarkeit und Ausdehnung des Signals beeinflusst.

Die Beobachtungen legen zudem nahe, dass die indirekte Wiedergabe der Surround- und Top Surround-Kanäle über Reflexionen an den Begrenzungsflächen des Raumes eine intendierte Erzeugung ausgedehnter Quellen und eine intendierte Umhüllung des Hörers passiv unterstützen könnte. Es konnte beobachtet werden, dass hinsichtlich der Soundbar-Wiedergabe einerseits eine geringere Anzahl an Signalen notwendig war, um einen mit der diskreten Wiedergabe vergleichbaren Eindruck von ausgedehnten Quellen und Umhüllung aus dem Bereich vor der Frontalebene zu erzeugen. Andererseits konnten die verwendeten Quellsignale auch eine höhere ICC zueinander aufweisen als bei der diskreten Wiedergabe, um eine Wahrnehmung von weiter ausgedehnten Quellen und Umhüllung zu generieren. Hierfür wurden unter anderem die frühen Reflexionen an den Seitenwänden verantwortlich gemacht. Diese fördern die Dekorrelation der Ohrsignale, was sich mit Ausführungen von Toole (2008) und Blauert & Braasch (2008) deckt.

5.2 Limitationen und Empfehlungen für zukünftige Arbeiten

Die dargelegten Erkenntnisse der Arbeit sind zwingend im Kontext der verwendeten Methodik zu bewerten. Diese weist eine Reihe von Schwächen und Einschränkungen auf.

Die geschilderten subjektiven Höreindrücke des Autors sind nicht repräsentativ und erfordern einer individuellen Überprüfung anhand der zur Verfügung gestellten Hörbeispiele. Zudem muss davon ausgegangen werden, dass sie einem erheblichen Bestätigungsfehler (Confirmation Bias) unterliegen, da der Autor die den Höreindrücken zugrunde liegenden Hörbeispiele im Rahmen dieser Arbeit selbst zu illustrativen Zwecken erstellt hat und über deren Inhalt im Vorfeld Bescheid wusste.

Zudem wurde nur ein einziges Soundbar-Modell mit einem einzigen diskreten Wiedergabesystem innerhalb eines einzigen gemeinsamen Abhörtraumes verglichen. Die Erkenntnisse sind somit nur eingeschränkt auf andere Wiedergabesysteme und -räume übertragbar. Insbesondere der Wiedergaberaum

hat hinsichtlich der Soundbar-Wiedergabe aufgrund dessen aktiver Einbeziehung in Form von Wand- und Deckenreflexionen eine hohe Relevanz. Zwischen den in Kapitel 3.1.3.1 dargelegten, kontrollierten raumakustischen Eigenschaften einer Tonregie und einem durchschnittlichen Wohnraum bestehen insbesondere hinsichtlich der frequenzabhängigen Nachhallzeiten teils deutliche Unterschiede (Jackson & Leventhall, 1972). Auch mit Blick auf die Varianz der Raumgeometrien von Wohnräumen scheint es zielführend, in zukünftigen Arbeiten verschiedene Wiedergaberäume im Hinblick auf deren Auswirkungen auf die Soundbar-bezogene Wiedergabe zu vergleichen.

Darüber hinaus war die Auswahl der Signale und Hörbeispiele begrenzt und konzentrierte sich auf die isolierte Illustration subjektiv auffälliger Eigenschaften beider Wiedergabesysteme. Die Hörerfahrung des Autors legt nahe, dass einige der beobachteten Effekte im Kontext einer komplexeren immersiven Audiomischung nicht in gleicher Art und Weise in Erscheinung treten. Auch dies müsste in nachfolgenden Arbeiten gezielter untersucht werden.

Nach Kenntnisstand des Autors gibt es zum Zeitpunkt der Einreichung dieser Arbeit keine veröffentlichte, repräsentative Studie, welche gezielt die Eigenschaften der diskreten Lautsprecherwiedergabe mit den Eigenschaften der Soundbar-Wiedergabe im Hinblick auf die Wiedergabe von immersivem Audio vergleicht. Bisherige Untersuchungen beschränkten sich auf einen Vergleich hinsichtlich der Surround-Darbietung und liefern zudem keinen Konsens in ihren Erkenntnissen (Liebl et al., 2018; Perla & Bulla, 2019; Walton et al., 2016). Da die Soundbar-Wiedergabe aufgrund ihrer weiten Verbreitung eine hohe Relevanz für einen zunehmend mit immersiven Audioinhalten bedienten Markt hat, wären weiterführende Untersuchungen der Thematik notwendig.

Aktuell ist eine Qualitätskontrolle der immersiven Soundbar-Wiedergabe aufgrund des hierfür erforderlichen Bitstream-Formats nicht in Echtzeit möglich. In diesem Zusammenhang wären Möglichkeiten den Transfer von immersivem Audio auf Soundbars bereits während der Postproduktion von immersiven Audioinhalten überprüfen zu können, von hohem Wert für Tonschaffende. Die Entwicklung einer mit diskreten Kanälen adressierbaren Referenz-Soundbar oder einer auralisierten, binauralen Repräsentation (Hoeg et al., 2014) einer Referenz-Soundbar wären hilfreiche Werkzeuge zur umgehenden Beurteilung der Kompatibilität zwischen diskreter Lautsprecherwiedergabe und Soundbar-Wiedergabe.

6. Fazit

Das Ziel dieser Arbeit war es, die spezifischen Eigenschaften der Soundbar-Wiedergabe immersiven Audios am Beispiel von Dolby Atmos zu untersuchen, Gemeinsamkeiten und Unterschiede zur diskreten Lautsprecherwiedergabe aufzuzeigen und Variablen innerhalb des Postproduktionsprozesses zu ermitteln, die einen Erhalt der ursprünglichen Intention immersiver Audioinhalte bei Wiedergabe über Soundbars begünstigen. Die zentrale Forschungsfrage lautete hierbei:

„Wie lassen sich die besonderen Eigenschaften des Wiedergabesystems Soundbar bereits bei der Produktion von Dolby Atmos-Inhalten berücksichtigen, um die Kompatibilität bei der Wiedergabe über Soundbars im Vergleich zur diskreten Lautsprecherwiedergabe zu optimieren?“.

Entscheidend für die Beantwortung dieser Frage ist die Erwartungshaltung des Tonschaffenden. Es handelt sich um zwei verschiedene Wiedergabesysteme mit individuellen, charakteristischen Eigenschaften. Ist die Erwartung, dass die erstellten, über ein diskretes Wiedergabesystem abgehörten immersiven Audioinhalte bei Wiedergabe über eine Soundbar in einem identischen Höreindruck resultieren, so wird diese Erwartung nicht erfüllt werden. Eine Anpassung von Erwartungshaltung und Intention an die Wiedergabefähigkeiten von Soundbars sind somit der Schlüssel zu einer optimierten Kompatibilität mit diesem speziellen Wiedergabesystem.

Hinsichtlich der Lokalisierbarkeit von Hörereignissen zeigten sich deutlich wahrnehmbare Unterschiede zwischen den beiden Wiedergabesystemen. Während für die Front-Kanäle L, C und R eine hohe Kompatibilität der Wiedergabe erwartet werden kann, unterscheidet sich die Repräsentation ihrer jeweiligen Surround- und Top Surround-Kanäle erheblich. Das Verlassen der vom Direktschall bestimmten LCR-Basis der Soundbar resultiert aufgrund der zum Teil indirekt über Reflexionen erfolgenden Wiedergabe in zunehmend diffuser wirkenden, weniger eindeutig lokalisierbaren Hörereignissen, welche aufgrund der Wiedergaberaum-abhängigen Lage der Reflexionspunkte und einem komplexen Zusammenwirken physikalischer und psychoakustischer Faktoren einer hohen Ungewissheit bezüglich ihrer Lokalisationsrichtung unterliegen. Auch die Art des Signals scheint darauf Einfluss zu haben. Eine Lokalisation von Hörereignissen hinter der Frontalebene war im Rahmen der Arbeit über die Soundbar-Wiedergabe nicht zu erzielen. Für Tonschaffende empfiehlt es sich daher hinsichtlich der Kompatibilität immersiver Audioinhalte zwischen beiden Wiedergabesystemen auf Intentionen und Narrative zu verzichten, die eine punktgenaue Lokalisation von Hörereignissen abseits der Front-Kanäle erfordern. Auch auf die Notwendigkeit der Lokalisierungsrichtung „hinten“ sollte in diesem Hinblick verzichtet werden. Bei der Auswahl und Erstellung von Trajektorien bei dynamischen Pannings sollte das wesentliche Funktionsprinzip einer direkt abstrahlenden Soundbar mit Reflexionspunkten vor der Frontalebene des Hörers bedacht werden. Die Visualisierung über die GUI des Panners gibt dabei häufig nur unzureichenden Aufschluss über den aus der Soundbar-Wiedergabe resultierenden Höreindruck.

Die Auswahl des für die Encodierung in Bitstream-Formate erforderlichen und die Wiedergabe über viele Soundbar-Modelle relevanten 5.1-Downmix-Algorithmus innerhalb des Dolby Atmos Renderers sollte kontextbezogen erfolgen. Selbes gilt für etwaige individuelle Einstellungen der Trims. Hier lassen sich keine pauschalen Empfehlungen formulieren.

Die indirekte Wiedergabe über Wand- und Deckenreflexionen begünstigt eine Dekorrelation der Ohrsignale. Die ermittelten Werte für das mit dem subjektiven Attribut der Apparent Source Width korrelierende Kriterium $IACC_{Early}$ implizieren, dass die Soundbar-Wiedergabe intentionierte, ausgedehnte Hörereignisse unterstützen kann.

Auf Grundlage subjektiver Höreindrücke, der zusätzlichen Dekorrelation durch Reflexionen und der im Verhältnis zur diskreten Wiedergabe in der Tonregie höheren schallenergetischen Anregung des Raums, lässt sich dies auch ansatzweise für das Attribut der auditiven Umhüllung bzw. Listener Envelopment annehmen. Im Rahmen dieser Arbeit konnte jedoch kein objektives Kriterium für die LEV ermittelt werden, welches diese Annahme stützt. Zudem blieb der subjektive Eindruck auditiver Umhüllung – ähnlich wie die Lokalisation – weitgehend auf den Bereich vor der Frontalebene beschränkt.

Die wahrscheinlich verlässlichste Methodik die Kompatibilität der Wiedergabe immersiver Audioinhalte zwischen Monitoring in der Tonregie und endnutzer-bezogener Soundbar-Wiedergabe zu überprüfen, bleibt aufgrund der vielen Variablen und Ungewissheiten bis auf weiteres der Qualitätscheck in Form der Wiedergabe der zuvor encodierten Audioinhalte über verschiedene Soundbars in unterschiedlichen Räumen.

7. Hörbeispielverzeichnis

Hörbeispiel 1

Sequenzielle Ansage der Ausgangskanäle für Dolby Atmos 7.1.4 zur Überprüfung des Output-Routings und Identifikation der einzelnen Kanäle.

Hörbeispiel 2

Sequenzielle Ansage der Ausgangskanäle für Dolby Atmos 5.1.4 zur Identifikation der Ausgangskanäle der Soundbar.

Hörbeispiel 3

Sequenzielle Wiedergabe von pinkem Rauschen über die einzelnen Front-Kanäle L, C und R.

Hörbeispiel 4

Sequenzielle Wiedergabe von Snare-Samples über die einzelnen Front-Kanäle L, C und R.

Hörbeispiel 5

Wiedergabe von breitbandigem und transientem Monosignal (Schlagzeug und Bass) über den Kanal L.

Hörbeispiel 6

Alternierende Wiedergabe von pinkem Rauschen über die einzelnen Surround-Kanäle Ls und Rs.

Hörbeispiel 7

Alternierende Wiedergabe von Snare-Samples über die einzelnen Surround-Kanäle Ls und Rs.

Hörbeispiel 8

Alternierende Wiedergabe von pinkem Rauschen über die Kanäle L und Ls.

Hörbeispiel 9

Alternierende Wiedergabe von Snare-Samples über die Kanäle L und Ls.

Hörbeispiel 10

Wiedergabe von breitbandigem und transientem Monosignal (Schlagzeug und Bass) über den Kanal Ls.

Hörbeispiel 11

Sequenzielle Wiedergabe von pinkem Rauschen über die Top Surround-Kanäle Ltf, Ltr, Rtr und Rtf.

Hörbeispiel 12

Sequenzielle Wiedergabe von Snare-Samples über die Top Surround-Kanäle Ltf, Ltr, Rtr und Rtf.

Hörbeispiel 13

Alternierende Wiedergabe von pinkem Rauschen über die Kanäle L und Ltf.

Hörbeispiel 14

Alternierende Wiedergabe von Snare-Samples über die Kanäle L und Ltf.

Hörbeispiel 15

Wiedergabe von breitbandigem und transientem Monosignal (Schlagzeug und Bass) über den Kanal Ltf.

Hörbeispiel 16

Alternierende Wiedergabe von Terzbandrauschen mit Mittenfrequenzen von 800 Hz und 8000 Hz, wiedergegeben über den Kanal Ltf.

Hörbeispiel 17

Monofones pinkes Rauschen, wiedergegeben mit identischen Ausgangspegeln über die Kanäle L und Ls.

Hörbeispiel 18

Monofones transientes und breitbandiges Signal (Schlagzeug und Bass), wiedergegeben mit identischen Ausgangspegeln über die Kanäle L und Ls.

Hörbeispiel 19

Monofones pinkes Rauschen, wiedergegeben mit identischen Ausgangspegeln über die Kanäle L und Ltf.

Hörbeispiel 20

Monofones pinkes Rauschen mit Automation der Y-Koordinate eines Audioobjekts zwischen den Koordinaten (-1, 1, 0) und (-1, -1, 0). Für den 5.1.-Downmix wird *direct render* verwendet.

Hörbeispiel 21

Monofones breitbandiges und transientes Signal (Schlagzeug und Bass) mit Automation der Y-Koordinate eines Audioobjekts zwischen den Koordinaten (-1, 1, 0) und (-1, -1, 0). Für den 5.1-Downmix wird *direct render* verwendet.

Hörbeispiel 22

Pan-Around Automation eines Audioobjekts im Middle Layer unter Verwendung von monofonem pinkem Rauschen und der 5.1-Downmix-Einstellung *direct render*.

Hörbeispiel 23

Pan-Around Automation eines Audioobjekts im Middle Layer unter Verwendung von monofonem pinkem Rauschen und der 5.1-Downmix-Einstellung *direct render with room balance*.

Hörbeispiel 24

Pan-Around Automation im Middle Layer des 7.1.2-Beds unter Verwendung von monofonem pinkem Rauschen und der 5.1-Downmix-Einstellung *Lo/Ro*.

Hörbeispiel 25

Monofones pinkes Rauschen mit Automation der Z-Koordinate eines Audioobjekts zwischen den Koordinaten (-1, 1, 0) und (-1, 1, 1).

Hörbeispiel 26

Automation der Y- und Z-Koordinaten eines mit pinkem Rauschen bespielten Audioobjekts unter Verwendung des *spherical mode* des Dolby Atmos Music Panners zur Darstellung einer Bewegung in der Medianebene. Für den 5.1-Downmix wird *direct render* verwendet.

Hörbeispiel 27

Automation der X- und Z-Koordinaten eines mit pinkem Rauschen bespielten Audioobjekts unter Verwendung des *spherical mode* des Dolby Atmos Music Panners zur Darstellung einer Bewegung in der Frontalebene. Für den 5.1-Downmix wird *direct render with room balance* verwendet.

Hörbeispiel 28

Pan-Around Automation eines Audioobjekts im Upper Layer unter Verwendung von monofonem pinkem Rauschen.

Hörbeispiel 29

Automation des Object Size-Parameters eines auf dem Center-Kanal positionierten, mit pinkem Rauschen bespielten Audioobjekts im Wertebereich zwischen 0 und 0,5.

Hörbeispiel 30

Automation des Object Size-Parameters eines auf dem Kanal Ls positionierten, mit pinkem Rauschen bespielten Audioobjekts im Wertebereich zwischen 0 und 0,5.

Hörbeispiel 31

Simultane Wiedergabe von teilkohärentem pinkem Rauschen mit einer mittleren ICC von 0,5 über alle Ausgangskanäle einer 5.1.4-Konfiguration (ohne LFE).

Hörbeispiel 32

Wiedergabe von teilkohärentem pinkem Rauschen mit einer mittleren ICC von 0,5 über die Kanäle Ls, Rs, Ltf, Rtf, Ltr und Rtr.

Hörbeispiel 33

Simultane Wiedergabe von teilkohärentem pinkem Rauschen mit einer mittleren ICC von 0,5 über die Kanäle Lrs, Rrs, Ltr und Rtr einer 7.1.4-Konfiguration, bzw. die Kanäle Ls, Rs, Ltr und Rtr einer 5.1.4-Konfiguration unter Verwendung des Downmix-Algorithmus *Direct render with room balance*.

Hörbeispiel 34

Simultane Wiedergabe von zwei inkohärenten pinken Rauschsignalen mit einer mittleren ICC von 0 über die Kanäle Ls und Rs.

Hörbeispiel 35

Simultane Wiedergabe von zwei inkohärenten pinken Rauschsignalen mit einer mittleren ICC von 0 über die Kanäle Ltf und Rtf.

Hörbeispiel 36

Simultane Wiedergabe von vier inkohärenten pinken Rauschsignalen mit einer mittleren ICC von 0 über die Kanäle Ls, Rs, Ltf und Rtf.

Hörbeispiel 37

Simultane Wiedergabe von zehn inkohärenten pinken Rauschsignalen mit einer mittleren ICC von 0 über alle Kanäle einer 7.1.4-Konfiguration mit Ausnahme von Center und LFE. Für den 5.1.x-Downmix wird der Algorithmus *Direct render with room balance* angewendet.

Literaturverzeichnis

- Ahnert, W. & Goertz, A. (2008). Beschallungstechnik, Beschallungsplanung und Simulation. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 491–549). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Ahnert, W., & Tennhardt, H.-P. (2008). Raumakustik. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 181–266). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Apple (2011). *Apple Core Audio Format Specification 1.0*. Permalink: https://developer.apple.com/library/archive/documentation/MusicAudio/Reference/CAFS-spec/CAF_spec/CAF_spec.html
- Apple (2025a). *Logic Pro – Benutzerhandbuch für den Mac*. Apple. Abgerufen von am 16. August 2025 von <https://support.apple.com/de-de/guide/logicpro/welcome/mac>
- Apple (2025b). *Mac User Guide. About HDMI Passthrough*. Apple. Abgerufen am 30. August 2025 von <https://support.apple.com/en-euro/guide/mac-help/mchle3a1461c/mac>
- Barbour, J. L. (2003, 26.–28. Juni). *Elevation Perception: Phantom Images in the Vertical Hemi-sphere* [Konferenzbeitrag]. 24th AES International Conference on Multichannel Audio. Banff, Canada. Permalink: <http://www.aes.org/e-lib/browse.cfm?elib=12301>
- Baxter, D. (2022). *Immersive Sound Production: A Practical Guide*. Routledge. <https://doi.org/10.4324/9781003052876>
- Beranek, L. (2004). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (2. Aufl.). Springer.
- Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization* (2. Aufl.). MIT Press. <https://doi.org/10.7551/mitpress/6391.001.0001>
- Blauert, J., & Braasch, J. (2008). Räumliches Hören. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 87–121). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Breebaart, J., & Faller, C. (2007). *Spatial Audio Processing: MPEG Surround and Other Applications*. John Wiley & Sons. <https://doi.org/10.1002/9780470723494>
- Cecchi, S., Carini, A., & Spors, S. (2017). Room Response Equalization – A Review. *Applied Sciences*, 8(1), 16. <https://doi.org/10.3390/app8010016>
- Crockett, B. G., Chabanne, C., Tuffy, M., Seefeldt, A. J., Brown, C. P., & Turnmire, P. (2017). *Virtual Height Filter for Reflected Sound Rendering Using Upward Firing Drivers* (US 9,648,440 B2). United States Patent and Trademark Office. <https://patents.google.com/patent/US9648440B2/en>
- Devine, T. (2022, 30. Juni). A deep dive into Dolby MAT. *AV Pro*. <https://www.avproglobal.com/blogs/news/a-deep-dive-into-dolby-mat>
- Dickreiter, M., & Goeres-Petri, J. (2014). Schallwahrnehmung. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudientechnik* (8. Aufl., Bd. 1) (S. 115–135). De Gruyter. <https://doi.org/10.1515/9783110316506>

- Dickreiter, M., & Hoeg, W. (2014). Grundlagen der Akustik. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudioteknik* (8. Aufl., Bd. 1) (S. 1–65). De Gruyter. <https://doi.org/10.1515/9783110316506>
- Dolby Laboratories (2012). *Dolby Atmos: Next-Generation Audio for Cinema*. Holland Film Nieuws. Abgerufen am 31. Juli 2025 von <https://www.hollandfilmnieuws.nl/files/whitepaper-dolby-atmos.pdf>
- Dolby Laboratories (2016a). *Dolby Atmos Enabled Speaker Technology*. Dolby Professional. Abgerufen am 1. August 2025 von <https://professional.dolby.com/siteassets/tv/home/dolby-atmos/dolby-atmos-enabled-speaker-technology.pdf>
- Dolby Laboratories (2016b). *Dolby Atmos for Home Theater*. Dolby Professional. Abgerufen am 31. Juli 2025 von <https://professional.dolby.com/siteassets/tv/home/dolby-atmos/dolby-atmos-for-home-theater.pdf>
- Dolby Laboratories (2018a). *Dolby Atmos Home Theater Installation Guidelines*. Dolby. Abgerufen am 24. Juli von https://www.dolby.com/siteassets/technologies/dolby-atmos/atmos-installation-guidelines-121318_r3.1.pdf
- Dolby Laboratories (2018b). *Dolby Atmos for sound bar applications*. Dolby Professional. Abgerufen am 31. Juli 2025 von <https://professional.dolby.com/siteassets/tv/home/dolby-atmos/dolby-atmos-for-sound-bar-applications.pdf>
- Dolby Laboratories (2018c). *Dolby Atmos Renderer Guide*. Dolby Professional. Abgerufen am 10. August 2025 von https://professional.dolby.com/siteassets/content-creation/dolby-atmos/dolby_atmos_renderer_guide.pdf
- Dolby Laboratories (2018d). *Dolby Atmos Studio Configuration*. Dolby Developer. Abgerufen am 12. August 2025 von <https://developer.dolby.com/dolby-games/tools-media/studio-resources/Atmos-in-studio/>
- Dolby Laboratories (2018e). *Dolby TrueHD (MLP) high-level bitstream description*. Dolby Developer. Abgerufen am 30. August 2025 von <https://developer.dolby.com/globalassets/technology/dolby-truehd/dolbytruehdhighlevelbitstreamdescription.pdf>
- Dolby Laboratories (2019). *TIDAL and Dolby are Bringing Dolby Atmos Music to TIDAL'S HiFi Members*. Dolby Newsroom. Abgerufen am 1. August 2025 von <https://news.dolby.com/en-WW/184250-tidal-and-dolby-are-bringing-dolby-atmos-music-to-tidal-s-hifi-members/>
- Dolby Laboratories (2021a). *Do I still need the Dolby Atmos Renderer if using Nuendo 11?* Dolby Professional Support. Abgerufen am 15. August 2025 von https://professionalsupport.dolby.com/s/article/Do-I-still-need-the-Dolby-Atmos-Renderer-using-Nuendo-11?language=en_US

- Dolby Laboratories (2021b). *Dolby Atmos Home Entertainment Studio Technical Guidelines*.
<https://dolby.my.salesforce.com/sfc/p/#700000009YuG/a/4u0000001FHc/UYA0IZeD632SUXVmEP-mUcr.wluhpHp6Q7bVSI4LrbUQ>
- Dolby Laboratories (2021c). *What is in the exported MP4 file and how are these encoded? Dolby Professional Support*. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/What-is-in-the-exported-MP4-file-and-how-are-these-encoded?language=en_US
- Dolby Laboratories (2021d). *Dolby AC-4: Audio delivery for next-generation entertainment services*. Dolby Professional. Abgerufen am 30. August 2025 von https://professional.dolby.com/sites-as-sets/technologies/dolby_atmos_ac-4_whitepaper.pdf
- Dolby Laboratories (2022). *Dolby Atmos ADM profile specification*. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/Dolby-Atmos-ADM-Profile-specification?language=en_US
- Dolby Laboratories (2023a). *What is Dolby Digital Plus JOC? (Joint Object Coding)*. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/What-is-Dolby-Digital-Plus-JOC-Joint-Object-Coding?language=en_US
- Dolby Laboratories (2023b). *What is the supported data rate of Dolby Digital Plus with Dolby Atmos?* Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/What-is-the-supported-data-rate-of-Dolby-Atmos?language=en_US
- Dolby Laboratories (2024a). *How do the 5.1 and Stereo downmix settings work?* Dolby Professional Support. Abgerufen am 29. August 2025 von https://professionalsupport.dolby.com/s/article/How-do-the-5-1-and-Stereo-downmix-settings-work?language=en_US
- Dolby Laboratories (2024b). *How are trims applied?* Dolby Professional Support. Abgerufen am 29. August 2025 von https://professionalsupport.dolby.com/s/article/How-are-trims-applied?language=en_US
- Dolby Laboratories (2024c, 5. Juni). *Where is spatial coding emulation?* [Forumbeitrag]. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/question/0D54u0000AUmWCqCQN/where-is-spatial-coding-emulation-?language=en_US
- Dolby Laboratories (2024d). *Dolby AC-4 FAQ*. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/Dolby-AC-4-FAQ?language=en_US
- Dolby Laboratories (2025a). *Dolby Atmos Renderer v.5.3.2 Documentation*. Dolby Customer. Abgerufen am 12. August 2025 von <https://customer.dolby.com/content-creation-and-delivery/dolby-atmos-renderer-v532/documentation>

- Dolby Laboratories (2025b). *What is the Dolby Atmos Renderer?* Dolby Professional Support. Abgerufen am 15. August 2025 von https://professionalsupport.dolby.com/s/article/What-is-the-Dolby-Atmos-Renderer?language=en_US
- Dolby Laboratories (2025c). *Overview of Dolby Atmos Master File Formats*. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/article/Overview-of-Dolby-Atmos-Master-File-Formats?language=en_US
- Dolby Laboratories (2025d). *Dolby Atmos*. Dolby Hybrik. Abgerufen am 30. August 2025 von https://docs.hybrik.com/tutorials/dolby_atmos/
- Dolby Laboratories (2025e). *Dolby Audio Overview*. Dolby Hybrik. Abgerufen am 30. August 2025 von https://docs.hybrik.com/tutorials/dolby_audio/
- Dolby Laboratories (2025f, 22. August). *Downmix Controls and their effect on encoded delivery files* [Forumbeitrag]. Dolby Professional Support. Abgerufen am 30. August 2025 von https://professionalsupport.dolby.com/s/question/0D7QQ000007JKj10AG/detail?language=en_US&fromEmail=1&slid=00D700000009YuG&slid=0DB4u000000kAJP&slid=0054u000007VGwa&slid=0&emkind=chatterCommentNotification&emtm=1756145665062
- Elliott, L. L. (1962). Backward Masking: Monotic and Dichotic Conditions. *The Journal of the Acoustical Society of America*, 34(8). 1108–1115. <https://doi.org/10.1121/1.1918253>
- Elliott, L. L. (1971). Backward and Forward Masking. *Audiology*, 10(2). 65–76. Permalink: <https://www.tandfonline.com/doi/abs/10.3109/00206097109072544>
- European Broadcasting Union (1998). *EBU Tech. 3276: Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic*. <https://tech.ebu.ch/docs/tech/tech3276.pdf>
- European Broadcasting Union (2011). *EBU Tech. 3285: Specification of the Broadcast Wave Format (BWF). A file for audio data files in broadcasting*. <https://tech.ebu.ch/docs/tech/tech3285.pdf>
- European Broadcasting Union (2017). *ETSI TS 102 366 V1.4.1: Digital Audio Compression (AC-3, Enhanced AC-3) Standard*. Permalink: https://www.etsi.org/deliver/etsi_ts/102300_102399/102366/01.04.01_60/ts_102366v010401p.pdf
- European Broadcasting Union (2018a). *ETSI TS 103 420 V1.2.1: Backwards-compatible object audio carriage using Enhanced AC-3*. Permalink: https://www.etsi.org/deliver/etsi_ts/103400_103499/103420/01.02.01_60/ts_103420v010201p.pdf
- European Broadcasting Union (2018b). *ETSI TS 103 190-1 V1.3.1: Digital Audio Compression (AC-4) Standard; Part 1: Channel based coding*. Permalink: https://www.etsi.org/deliver/etsi_ts/103100_103199/10319001/01.03.01_60/ts_10319001v010301p.pdf
- European Broadcasting Union (2019). *Technology Fact Sheet: Next Generation Audio (NGA)*. Permalink: https://tech.ebu.ch/docs/factsheets/ebu_tech_fs_nga.pdf

- European Broadcasting Union (2021). *10 things you need to know about Next Generation Audio*. Permalink: https://tech.ebu.ch/files/live/sites/tech/files/shared/other/EBU_NGA_10_Things_You_Need_To_Know.pdf
- European Broadcasting Union (2025). *ETSI TS 103 190-2 V1.3.1: Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio*. Permalink: https://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.03.01_60/ts_10319002v010301p.pdf
- Everest, F. A., & Pohlmann, K. C. (2022). *Master Handbook of Acoustics* (7. Aufl.). McGraw-Hill.
- Faller, C., & Merimaa, J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America*, 116(5). 3075–3089. <https://doi.org/10.1121/1.1791872>
- Farina, A. (2000, 19.–22. Februar). *Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique* [Konferenzbeitrag]. 108th Convention of the Audio Engineering Society (AES), Paris, France. Permalink: <https://aes2.org/publications/elibrary-page/?id=10211>
- Fielder, L. D., Andersen, R. L., Crockett, B. G., Davidson, G. A., Davis, M. F., Turner, S. C., Vinton, M. S., & Williams, P. A. (2004, 28.–31. Oktober). Introduction to Dolby Digital Plus, an Enhancement to the Dolby Digital Coding System [Konferenzbeitrag]. 117th Convention of the Audio Engineering Society (AES). San Francisco, United States of America. Permalink: <https://aes2.org/publications/elibrary-page/?id=12853>
- Fraunhofer IIS (2023). *Creating Best-in-class Audio with the upHear Technologies*. Fraunhofer Audio Blog. Abgerufen am 31 August 2025 von <https://www.audioblog.iis.fraunhofer.com/sennheiser-ambeo-uphear>
- Frisch, H. (2004). Beschallungstechnik. In G. Müller & M. Möser (Hrsg.), *Taschenbuch der Technischen Akustik* (3. Aufl.) (S. 441–458). Springer. <https://doi.org/10.1007/978-3-642-18893-0>
- Gerzon, M. (1991). *Digital room equalisation*. Gerzon Archive. Abgerufen am 31 August 2025 von https://www.audiosignal.co.uk/Resources/Digital_room_equalisation_A4.pdf
- Gerzon, M. A., Craven, P. G., Stuart, J. R., Law, M. J., & Wilson, R. J. (2004). The MLP Lossless Compression System for PCM Audio. *Journal of the Audio Engineering Society*, 52(3). 243–260. Permalink: <https://aes2.org/publications/elibrary-page/?id=12993>
- GfK SE (2021). *Guter Ton zum guten Bild: Wie wichtig ist Zuschauern die Tonqualität des Fernsehers*. Deutsche TV-Plattform. Abgerufen am 2. August 2025 von https://tv-plattform.de/wp-content/uploads/2021/12/DTVP_TV_Ton_2021-12-02.pdf
- Griesinger, D. (1999, 10–12. April). *Objective Measures of Spaciousness and Envelopment* [Konferenzbeitrag]. 16th AES International Conference on Spatial Sound Reproduction, Rovaniemi, Finland. Permalink: <https://aes2.org/publications/elibrary-page/?id=8051>
- Görne, T. (2015). *Tontechnik* (4. Aufl.). Hanser. <https://doi.org/10.3139/9783446441491>
- HiFi im Hinterhof GmbH (2019). *High-End TV-Sound mit der Sennheiser Ambeo Soundbar*. *HiFi im Hinterhof*. Abgerufen am 8. September 2025 von <https://www.hifi-im-hinterhof.de/blog/high->

- end-tv-sound-mit-der-sennheiser-ambeo-soundbar/?srsltid=AfmBOop-PZp7iR6HH1u2ihWd4zHx8pdJhMVN_eoUyho7REYk_V8txvQqI
- Hoeg, W., Nicklaus, K., Hackbarth, K., Spikofski, G., Camerer, F., & Dohlus, E. (2014). Qualitätssicherung. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudiotechnik* (8. Aufl., Bd. 1) (S. 1267–1352). De Gruyter. <https://doi.org/10.1515/9783110316506>
- Holman, T. (2008). *Surround Sound: Up and Running* (2. Aufl.). Routledge. <https://doi.org/10.1016/B978-0-240-80829-1.50007-2>
- Hooley, T. (2006) Single Box Surround Sound. *Acoustical Science and Technology*, 27(6). 354–360. <http://dx.doi.org/10.1250/ast.27.354>
- Idrovo, R. (2025). Dolby Atmos: A Breakthrough in Cinema Sound? *International Journal on Stereo & Immersive Media*, 8(1). 120–137. <https://doi.org/10.60543/ijsim.v8i1.9014>
- International Telecommunication Union (2014). *Report ITU-R BS.2159-6: Multichannel sound technology in home and broadcasting applications*. Permalink: https://www.itu.int/dms_pub/itu-r/opb/rep/r-rep-bs.2159-6-2013-pdf-e.pdf
- International Telecommunication Union (2015). *Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems*. Permalink: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf
- International Telecommunication Union (2017). *Report ITU-R BS.2399-0: Methods for selecting and describing attributes and terms, in the preparation of subjective tests*. Permalink: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BS.2399-2017-PDF-E.pdf
- International Telecommunication Union (2022a). *Recommendation ITU-R BS.775-4: Multichannel stereophonic sound system with and without accompanying picture*. Permalink: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-4-202212-I!!PDF-E.pdf
- International Telecommunication Union (2022b). *Recommendation ITU-R BS.2051-3: Advanced sound system for programme production*. Permalink: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2051-3-202205-I!!PDF-E.pdf
- International Telecommunication Union (2025). *Recommendation ITU-R BS.2076-3: Audio definition model*. Permalink: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2076-3-202502-I!!PDF-E.pdf
- Jackson, G. M., & Leventhall H. G. (1972). The Acoustics of Domestic Rooms. *Applied Acoustics*, 5(4). 265–277. [https://doi.org/10.1016/0003-682X\(72\)90030-8](https://doi.org/10.1016/0003-682X(72)90030-8)
- Kim, S. (2017). Height Channels. In A. Roginska & P. Geluso (Hrsg.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (S. 221–243). Routledge. <https://doi.org/10.4324/9781315707525>
- Lauterbach, T., Schulze, H., Färber, N., Schierbaum, T., Stoll, G., Hartmann, C., Baumgartner, H., & Hoeg, W. (2014). Rundfunksysteme. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.),

- Handbuch der Tonstudioteknik* (8. Aufl., Bd. 1) (S. 1085–1266). De Gruyter. <https://doi.org/10.1515/9783110316506>
- Lee, H. (2011, 20.–23. Oktober). *The Relationship between Interchannel Time and Level Differences in Vertical Sound Localisation and Masking* [Konferenzbeitrag]. 131st Convention of the Audio Engineering Society (AES). New York, United States of America. Permalink: <http://www.aes.org/e-lib/browse.cfm?elib=16082>
- Lee, H., Gribben, C., & Wallis, R. (2014, 20.–23. November). *Psychoacoustic Considerations in Surround Sound with Height* [Konferenzbeitrag]. 28. Tonmeistertagung (tmt 28). Köln, Deutschland. Permalink: https://eprints.hud.ac.uk/id/eprint/23151/1/2014_TMT28_Lee_et_al.pdf
- Lee, H., & Rumsey, F. (2013). Level and Time Panning of Phantom Images for Musical Sources. *Journal of the Audio Engineering Society*, 61(12). 978–988. Permalink: <https://aes2.org/publications/elibrary-page/?id=17075>
- Liebl, T., Wakan, S., & Curdt, O. (2018, 19.–22. März). *Audio quality evaluation of soundbars using the multiple stimulus ideal profile method* [Konferenzbeitrag]. DAGA 2018 – 44. Jahrestagung für Akustik. München, Deutschland. Permalink: https://pub.dega-akustik.de/DAGA_2018/data/articles/000389.pdf
- Maier, P. (2008). Studioakustik. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 267–311). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Perla, J., & Bulla, W. A. (2019, 16.–19. Oktober). A qualitative investigation of soundbar theory [Konferenzbeitrag]. 147th Convention of the Audio Engineering Society (AES). New York, United States of America. <https://aes2.org/publications/elibrary-page/?id=20674>
- Pulkki, V. (1997). Virtual Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6). 456–466. Permalink: <https://aes2.org/publications/elibrary-page/?id=7853>
- Pulkki, V. (2001). Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning. *Journal of the Audio Engineering Society*, 49(9). 753–767. Permalink: <https://aes2.org/publications/elibrary-page/?id=10179>
- Pulkki, V., & Karjalainen, M. (2015). *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. Wiley. <https://doi.org/10.1002/9781119825449>
- Purnhagen, H., Hirvonen, T., Villemoes, L., Samuelsson, J., & Klejsa, J. (2016, 4.–7. Juni) *Immersive Audio Delivery Using Joint Object Coding* [Konferenzbeitrag]. 140th Convention of the Audio Engineering Society (AES), Paris, France. Permalink: <https://www.aes.org/tmpFiles/elib/20250903/18285.pdf>
- Razer Inc. (2025). Razer Leviathan V2 Pro: The World's First Beamforming Soundbar With Head-Tracking AI. Razer. Abgerufen am 23. August 2025 von https://www.razer.com/gaming-speakers/razer-leviathan-v2-pro?srltid=AfmBOooYicf15j_DOsLf9GwxnqFwKmsGFfe-PlqfhXnJnLMorXUBpprsmp

- Rothermich, E. (2021). *Mixing in Dolby Atmos - #1 How it Works*. Abgerufen am 12. Juli 2025 von <https://dingdingmusic.com/-titles-/dolby-atmos-1.html>
- Rumsey, F. (2001). *Spatial Audio* (1. Aufl.). Routledge. <https://doi.org/10.4324/9780080498195>
- Rumsey, F. (2017). Surround Sound. In A. Roginska & P. Geluso (Hrsg.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (S. 180–220). Routledge. <https://doi.org/10.4324/9781315707525>
- Samsung (2025). *11.1.4-Kanal Soundbar Q990F*. Samsung. Abgerufen am 24. August 2025 von <https://www.samsung.com/at/audio-devices/soundbar/q990f-black-hw-q990f-en/>
- Schullan, B., Zuleeg, R., & Hoeg, W. (2014). Beschallung. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudiotechnik* (8. Aufl., Bd. 1) (S. 565–612). De Gruyter. <https://doi.org/10.1515/9783110316506>
- Sennheiser (2023). Ambeo Soundbar SB01: Bedienungsanleitung. Sennheiser. Abgerufen am 24. August 2025 von <https://eu-central-1-akqa.graphas-sets.com/AGz66yvUcQ42Ggm7CrXdgz/ZQ8WX38STga7rOFqlKNI>
- Sennheiser (2025a). *AMBEO Soundbar | Max*. Sennheiser. Abgerufen am 23. August 2023 von <https://www.sennheiser-hearing.com/de-DE/p/ambeo-soundbar/?srsltid=AfmBOoqB-kUffYXr4wj5OmAbabpcqA4OuCmvMHQ4DdMenEVPPoNBp5YI7>
- Sennheiser (2025b). *AMBEO Soundbar | Max. Häufig gestellte Fragen (FAQ)*. Sennheiser. Abgerufen am 19. September 2025 von <https://www.sennheiser-hearing.com/de-DE/max-faq/>
- Slavik, K. M., & Weinzierl, S. (2008). Wiedergabeverfahren. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 609–685). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Society of Motion Picture and Television Engineers (2022). *SMPTE ST 2098-2:2022. Immersive Audio Bitstream Specification*. Permalink: <https://pub.smpte.org/pub/st2098-2/st2098-2-2022.pdf>
- Society of Motion Picture and Television Engineers (2025). *SMPTE ST 2076: Interoperable Master Format (IMF)*. Permalink: <https://www.smpte.org/standards/st2067>
- Sonos (2024). *Sonos Reveals Arc Ultra, its New Soundbar Featuring Sound Motion*. Abgerufen am 24. August 2025 von <https://investors.sonos.com/news-and-events/investor-news/latest-news/2024/Sonos-Reveals-Arc-Ultra-its-New-Soundbar-Featuring-Sound-Motion/default.aspx>
- Steinberg Media Technologies (2025). *Nuendo 14.0.30 Benutzerhandbuch*. Steinberg. Abgerufen am 16. August 2025 von https://www.steinberg.help/v/u/nuendo_14_0_operation_manual_de.pdf
- Theile, G., Dickreiter, M., Graul, W., Camerer, F., & Spikofski, G. (2014). Tonaufnahme und Tonwiedergabe. In M. Dickreiter, V. Dittel, W. Hoeg & M. Wöhr (Hrsg.), *Handbuch der Tonstudiotechnik* (8. Aufl., Bd. 1) (S. 217–369). De Gruyter. <https://doi.org/10.1515/9783110316506>
- Theile, G., & Wittek, H. (2011). *Principles in Surround Recordings with Height*. Permalink: https://www.hauptmikrofon.de/doc/Auro3D_E_0311.pdf
- Toole, F. (2008). *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms* (1. Aufl.). Routledge. <https://doi.org/10.4324/9780080888019>

- Tsingos, N. (2017). Object-Based Audio. In A. Roginska & P. Geluso (Hrsg.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (S. 244–275). Routledge. <https://doi.org/10.4324/9781315707525>
- Universal Music Group (2025). *Dolby Atmos Audio Servicing Asset Specifications. Summary of Specifications*. Universal Music Group. Abgerufen am 30. August 2025 von <https://contentguide.universalmusic.com/dolby-atmos-audio-servicing-asset-specifications/>
- Walton, T., Evans, M., Kirk, D., & Melchior, F. (2016, 4.–7. Juni). *A subjective comparison of discrete surround sound and soundbar technology by using mixed methods* [Konferenzbeitrag]. 140th Convention of the Audio Engineering Society (AES), Paris, France. Permalink: <https://aes2.org/publications/elibrary-page/?id=18290>
- Weinzierl, S. (2008). Aufnahmeverfahren. In S. Weinzierl (Hrsg.), *Handbuch der Audiotechnik* (1. Aufl.) (S. 551–608). Springer. <https://doi.org/10.1007/978-3-540-34301-1>
- Wenzel, E. M., Begault, D. R., & Godfroy-Cooper, M. (2017). Perception of Spatial Sound. In A. Roginska & P. Geluso (Hrsg.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (S. 5–39). Routledge. <https://doi.org/10.4324/9781315707525>
- Yamaha Corporation (2015). *Digital Sound Projector YSP-5600 Owner's Manual*. Yamaha. Abgerufen am 23. August 2025 von https://usa.yamaha.com/files/download/other_assets/3/605013/web_YH433B0EN1_YSP-5600_om_UCALV_En_B0.pdf
- Ziegler, D., & Lerch, C. (2023). Dolby Atmos. In Berufsvereinigung Filmtone e.V. (Hrsg.), *Filmtone Guide 2023* (13. Aufl.) (S. 44–53). Permalink: https://www.bvft.de/epaper-Filmtonguide_2023-Teil_1/index.html#0
- Ziomek, L. J. (1995). *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing* (1. Aufl.). CRC Press. <https://doi.org/10.1201/9781003069317>

Digitaler Anhang

Digitaler Anhang 1 (DVD 1)	DAW-Projekt zur Erstellung der Hörbeispiele
Digitaler Anhang 2 (DVD 1)	REW-Messdateien
Digitaler Anhang 3 (DVD 1)	Hörbeispiele 1–37 im Dateiformat DAMF
Digitaler Anhang 4 (DVD 1)	Hörbeispiele 1–37 im Dateiformat .mp4
Digitaler Anhang 5 (DVD 2)	Hörbeispiele 1–37 im Dateiformat ADM BWF