

Bachelorarbeit im Studiengang Audiovisuelle Medien

Konzeption und Entwicklung eines Audio- Setups für eine Live TV-Show in Dolby Atmos am Beispiel „The Masked Singer“



Vorgelegt von Leonie Thomaschewski
An der Hochschule der Medien Stuttgart
Am 13. März 2025

Zur Erlangung des akademischen Grades eines
Bachelor of Engineering

Erstprüfer: Prof. Oliver Curdt
Zweitprüfer: Dipl.-Ing. Andreas Albert

Danksagung

Hiermit möchte ich mich bei allen bedanken, die mich bei der Erstellung dieser Bachelorarbeit unterstützt haben. Mein besonderer Dank gilt meinen Betreuern und Prüfern, Prof. Oliver Curdt und Dipl. Ing. Andreas Albert für die Unterstützung und die hilfreichen Anregungen während der Bearbeitungszeit. Herzlichsten Dank auch an Dipl. Ing. Constantin Gabrysch, der durch die fachliche Betreuung und Bereitstellung relevanter Materialien maßgeblich zum Gelingen dieser Arbeit beigetragen hat und mir zu jeder Zeit Antworten auf meine Fragen gegeben hat. Danke auch an alle anderen Kollegen der Firma MMC, insbesondere Emilio Bonta und Martin Antulov, die mit mir im Austausch standen, zudem möchte ich mich bei der MMC GmbH bedanken, die mir die Möglichkeit der Zusammenarbeit geboten hat. Darüber hinaus bedanke ich mich bei all meinen Freunden und vor allem meiner Familie, die mich während des gesamten Studiums unterstützt haben.

In dieser Bachelorarbeit wird aus Gründen der besseren Lesbarkeit das generische Maskulinum verwendet. Weibliche und anderweitige Geschlechteridentitäten sind dabei stets eingeschlossen.

Ehrenwörtliche Erklärung

„Hiermit versichere ich, Leonie Thomaschewski, ehrenwörtlich, dass ich die vorliegende

Bachelorarbeit mit dem Titel: „Konzeption und Entwicklung eines Audio-Setups für eine Live-TV Show in Dolby Atmos am Beispiel ‚The Masked Singer‘ “ 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, § 23 Abs. 2 Master-SPO (Vollzeit)) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.“



Leonie Thomaschewski, 09.03.2025

Zusammenfassung

Das Ziel dieser Arbeit ist es, ein funktionierendes Audio-Setup zu entwickeln und eine stimmige 3D-Audio Mischung des ProSieben Formats „The Masked Singer“ anzufertigen, die auch in Wiedergabeformaten mit weniger Audiokanälen funktioniert und somit abwärtskompatibel ist. Außerdem können durch die neu gewonnene Expertise und den entwickelten Arbeitsablauf auf dem Firmengelände der MMC zukünftig weitere Formate in Dolby Atmos produziert werden. Auf Grundlage von recherchierten Inhalten zu Lautsprecheranordnungen und räumlichem Hören wurde eine kanalbasierte 5.1.4 Mischung produziert, aus der man gleichzeitig eine 5.1 Surround Mischung und eine 2.0 Stereomischung als Downmix generieren kann. Dabei wird nur eine Mischung angefertigt, aus der zwei weitere Mischungen hervorgehen, ohne den Produktionsaufwand erheblich zu erhöhen.

Die Gestaltung einer TV-Show in 3D-Audio in technischen und gestalterisch-kreativen Aspekten wird insbesondere durch die Downmixkompatibilität bestimmt. Es kann nicht jede Art von Quelle frei im Raum platziert werden, da sie beim Downmixvorgang möglicherweise durch die Pegelabsenkungen untergehen. Schallquellen, die dramaturgisch relevant sind, können somit nicht in der oberen Ebene oder den hinteren Lautsprechern platziert werden. Publikumsatmo setzt sich aus Publikumsreaktionen und Umgebungsgeräuschen zusammen, sie hingegen ist ein Bestandteil einer TV-Show, welcher den Zuschauer einhüllen kann und deren Pegelabsenkung im Downmix nicht wesentlich die Qualität der Mischung beeinträchtigt.

Über die Relevanz dieser Arbeit für zukünftige Entertainment-Shows und den Endnutzer wurde mit Toningenieur Emilio Bonta in einem Interview gesprochen.

Abstract

The aim of this work is to create a functioning Audio-setup and a consistent 3D audio mix of the ProSieben format "The Masked Singer", which also works in playback formats with fewer audio channels and is therefore downward compatible. In addition, the newly acquired expertise and the developed workflow will open up the possibilities, that future formats can be produced in Dolby Atmos by the MMC company. A channel-based 5.1.4 mix was produced based on researched content on loudspeaker arrangements and spatial hearing, from which a 5.1 surround mix and a 2.0 stereo mix can be generated simultaneously as a downmix. Only one mix is produced, from which two further mixes are created without significantly increasing the production effort.

The design of a TV show in 3D audio in terms of technical and creative aspects is determined particularly by downmix compatibility. Not every type of source can be placed freely in the room, as they may be drowned out by the level reductions during the downmix process. Sound sources that are dramaturgically relevant can therefore not be placed in the upper level or at the back. Audience ambience signals, on the other hand, are components of a TV show that can envelop the viewer and whose level reduction does not significantly affect the quality of the mix.

The relevance of this work for future entertainment shows and the consumer was discussed with sound engineer Emilio Bonta in an interview.

Inhaltsverzeichnis

Abkürzungsverzeichnis	vi
Abbildungsverzeichnis	x
Tabellenverzeichnis	xi
1. Einleitung	1
2. Terminologie – Grundlagen	2
2.1 Surround	2
2.2 Immersives Audio	4
2.3 Binaurales Audio und binaurale Aufnahmetechnik	5
2.4 Lokalisierung von Hörereignissen	7
2.4.1 Menschliches Hören.....	7
2.4.2 Lokalisation von Schallereignissen.....	9
2.5 Lautsprecheranordnungen	16
2.6 Verwendung der Höhenlautsprecher	18
2.7 Kanalbasierte 3D Mikrofonanordnungen.....	22
2.7.1 2L-Cube.....	25
2.7.2 OCT-3D	26
2.7.3 ORTF-3D.....	27
3. Dolby Atmos	29
3.1 Anwendungsfälle	31
3.2 Technische Richtlinien	32
4. 3D-Audio im Fernsehbereich und objektbasiertes Audio (Next Generation Audio).....	36
4.2 Tontechnische Positionen	40
4.4 Zuspielsysteme	43
4.5 Aktuelle Produktionsstandards im Fernsehen – Audiobereich	44
5. Konzeption.....	47
5.1 Aufzeichnung der Live-Show “The Masked Singer”	47
5.2 Allgemeine Informationen zur Tonregie 11	48
5.3 Anpassung der Mikrofonierung für 3D-Audio.....	48
5.4 Lautsprecherformat.....	51
5.5 Abhöre	51
5.6 Upmix	53

5.7	Downmix	53
5.8	Mischpulteinrichtung	55
5.9	Bild-Ton-Synchronität	57
5.10	Lautheitsmessung	58
6.	Mischung	58
6.1	Sprache.....	58
6.2	Musik	59
6.3	Atmo.....	60
6.4	Gesamte Sendung	62
6.5	Codierung	65
7.	Fazit.....	66
8.	Literaturquellenverzeichnis.....	68
9.	Abbildungsquellenverzeichnis	72
10.	Tabellenquellenverzeichnis	72
11.	Anhänge.....	73
11.1	Audiobelegungsplan The Masked Singer S11	73
11.2	MAZ Belegungsplan.....	74
11.3	Grundriss Tonregie 11	75
11.4	Transkription Interview mit Emilio Bonta	76

Abkürzungsverzeichnis

3D	-	dreidimensional
ARD	-	Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland
AV	-	Audio/Video
BBC	-	British Broadcast Corporation
C	-	Center
DAW	-	Digital Audio Workstation
EBU	-	European Broadcasting Union
ER	-	Early reflections (frühe Reflexionen)
HD	-	High definition
HDMI	-	High definition multimedia interface
HRIR	-	Head related impulse response
HRTF	-	Head related transfer function
HSVC	-	Horizontally spaced, vertically coincident
HVC	-	Horizontally and vertically coincident
HVS	-	Horizontally and vertically spaced
ICXT	-	Interchannel crosstalk
IID	-	Interaural intensity difference
ILD	-	Interaural level difference
IMS	-	Immersive stereo
IRT	-	Institut für Rundfunktechnik
ITD	-	Interaural time difference
ITU	-	International telecommunication union
JOC	-	Joint object coding
L	-	Links / left
LFE	-	Low frequency effects
LR	-	Late reflections
Lrs	-	Left rear surround
Ls	-	Left surround
Ltf	-	Left top front
Ltr	-	Left top rear
MADI	-	Multichannel Audio Digital Interface
MAZ	-	Magnetische Aufzeichnung (- Einspieler)
MMC	-	Magic Media Company GmbH Köln

NGA	-	Next generation Audio
NHK	-	Nippon Hōsō Kyōkai (japanische Rundfunkgesellschaft)
OBA	-	Object-based Audiosystem
OBS	-	Olympic Broadcasting Service
OCT	-	Optimized cardioid triangle
ORF	-	Österreichischer Rundfunk
ORTF	-	Office de Radiodiffusion Télévision Française
PC	-	Personal computer
PCM	-	Pulse code modulation
R	-	Right / rechts
Rs	-	Right surround
Rrs	-	Right rear surround
Rtf	-	Right top front
Rtr	-	Right top rear
SHV	-	Super HiVision
TMS	-	The Masked Singer
TV	-	Television, Fernsehen
UHD	-	Ultra high definition
UHF	-	Ultra High Frequency
VCA	-	Voltage Controlled Amplifier
VR	-	virtual reality
ZDF	-	Zweites Deutsches Fernsehen

Abbildungsverzeichnis

Abbildung 1: Lautsprecheraufstellung 5.1 Surround mit Winkelangaben.....	3
Abbildung 2: Neumann KU8li Kunstkopf	7
Abbildung 3: Anatomie des menschlichen Ohrs (Außenohr, Mittelohr, Innenohr) .	8
Abbildung 4: Demonstration der ITD einer Schallquelle seitlich kommend	10
Abbildung 5: Kopfbezogenes Koordinatensystem, Definition von Medianebene, Frontalebene und Horizontalebene.....	12
Abbildung 6: Relative Häufigkeit der Beobachter, die eines der Urteile "v", "o" oder "h" häufiger abgeben als die beiden andern zusammen	13
Abbildung 7: Zeitliche Folge der Reflexionen in einem Raum bei Impulsschall ..	15
Abbildung 8: Morten Lindberg ´s 2L-Cube Mikrofonanordnung	25
Abbildung 9: OCT-3D Mikrofonanordnung	27
Abbildung 10: ORTF-3D Mikrofonanordnung	28
Abbildung 11: Detail der Kapselausrichtung eines ORTF-3D Mikrofonarrays	28
Abbildung 12: Äquidistantes Lautsprecherlayout 5.1.4	34
Abbildung 13: Positionen der Surround- und Deckenlautsprecher	35
Abbildung 14: Atmo-Mikrofonplan The Masked Singer Staffel 11, 2024	42
Abbildung 15: ORTF-3D Positionierung im Studio bei The Masked Singer	50
Abbildung 16: vorher, Tonregie 11 in 5.1	53
Abbildung 17: nacher, Tonregie 11 in 5.1.4	53
Abbildung 18: Panning der Atmos in beiden Ebenen	61

Tabellenverzeichnis

Tabelle 1: Lautsprecherlayout für Sound System D (4+5+0) - 5.1.4.....	17
Tabelle 2: Lautsprecherkonfiguration für Sound System B (0+5+0) - 5.1	17
Tabelle 3: Lautsprecherkonfiguration für Sound System D (4+5+0) - 5.1.4	18
Tabelle 4: Empfohlene Mindest- und Maximalabmessungen des Layouts	33
Tabelle 5: Downmixparameter von 5.1 Surround zu 2.0 Stereo	47
Tabelle 6: Downmixmatrix von 5.1 zu 2.0	54
Tabelle 7: Vorläufige Downmixmatrix von 5.1.4 zu 5.1	54
Tabelle 8: Downmixmatrix von 5.1.4 zu 5.1	64
Tabelle 9: Kanalzuweisungen der gesamten Mischung	64

1. Einleitung

In der Medienwelt gibt es eine immer größer werdende Vielfalt an Formaten, sowohl im Video- als auch im Audiobereich. Jede Erweiterung definiert eine andere Anzahl an Kanälen und verschiedensten Positionierungen der Lautsprecher, das hat auch Einfluss auf den Workflow beim Mischen. 3D-Audio erreicht den Zuhörer aus allen Ebenen und Richtungen und ermöglicht ein immersives Klangerlebnis. Dolby Atmos ist eines dieser immersiven Mischformate, welches bisher vor allem im Film- und Gaming Bereich eingesetzt wird. Im Fernsehbereich werden bisher vor allem Sportsendungen in diesem 3D-Audio Format gemischt. TV-Shows im Entertainment Bereich wurden bisher nicht ausgestrahlt. Aus diesem Grund soll in dieser Arbeit die Umsetzung einer Live TV-Show in Dolby Atmos dokumentiert werden, damit zukünftige Sendungen bei der Planung und Umsetzung darauf zurückgreifen können. Ziel ist die Entwicklung eines Audio-Setups und in diesem Zusammenhang die Anfertigung einer funktionierenden Mischung für die ProSieben Show „The Masked Singer“. Im Jahr 2023 wurde im Rahmen von „Let´s Dance“ bereits eine Testaufzeichnung in Dolby Atmos aus der Remote Regie in Darmstadt durchgeführt, ohne dass sie an den Endnutzer gesendet wurde.

In den folgenden Kapiteln wird zunächst das Surround-Format 5.1 vorgestellt. Außerdem werden Definitionen für immersives und binaurales Audio aufgestellt. Das darauffolgende Kapitel beschreibt die Funktionsweise des menschlichen Hörens mit dem Schwerpunkt auf der Lokalisierung von Schallereignissen in der Medianebene. Des Weiteren werden die verschiedenen Lautsprecheranordnungen und ihre möglichen Schreibweisen vorgestellt. Basierend auf der Funktionsweise des menschlichen Gehörs und der Lokalisation wird beschrieben, wie die Höhenlautsprecher eingesetzt werden. Das darauf aufbauende Kapitel beschäftigt sich mit den vielfältigen kanalbasierten Mikrofonanordnungen, die für den 3D-Audio Einsatz geeignet sind, wovon drei Systeme vorgestellt werden. Anschließend wird sich mit der Definition von Dolby Atmos und den technischen Vorgaben bezüglich der Lautsprecheraufstellungen auseinandergesetzt. Außerdem wird erläutert, inwiefern 3D-Audio bisher im Fernsehbereich eingesetzt wird. Im darauffolgenden Kapitel erfolgt eine Vorstellung der „The Masked Singer“ Show in technischer Hinsicht. Bevor die Mischung in den verschiedenen

Showbestandteilen betrachtet wird, wird zunächst ein Konzept für die Aufzeichnung der Show in Dolby Atmos entwickelt, welches auf den vorherigen Recherchen basiert. Zuletzt werden im Fazit die gewonnenen Erkenntnisse zusammengefasst und Auskunft über die zukünftige Relevanz des 3D-Audio Formats für Endnutzer und Produktionsstätten gegeben.

2. Terminologie – Grundlagen

2.1 Surround

Surround Sound bringt das Standardaudioformat Stereo, das aus den zwei Kanälen L (links) und R (rechts) besteht, auf die nächste Ebene. Während bei Stereo der Klang nur von vorne und aus zwei Lautsprechern kommt, werden beim Surround Sound mehrere Lautsprecher um den Zuhörer auf einer Kreisebene platziert, um ihm das Gefühl zu geben, vom Klang umgeben zu sein. Ziel ist es, ein räumliches, immersives Hörerlebnis zu erzeugen, bei dem der Klang wie im echten Leben aus verschiedenen Richtungen wahrnehmbar ist.¹ Es gibt keine einheitliche Kanalkonfiguration für den Surround Sound, sondern mehrere unterschiedliche Anordnungen. Jedoch gibt die ITU (International Telecommunication Union) einige Lautsprecheranordnungen vor. Diese beziehen sich vor allem auf die Musikproduktion in einem konventionellen Tonstudio und die Produktion von Radio- und TV-Inhalten.

Am häufigsten kommen die Kanalordnungen 5.1 oder 7.1 vor. Dabei bestimmt die vordere Zahl, wie viele Lautsprecher sich um den Zuhörer herum befinden. Neben den Lautsprechern L und R gibt es außerdem einen Center Kanal (C) und abhängig von der Anzahl der Lautsprecher noch jeweils einen bzw. zwei weitere Kanäle hinten links und rechts, die die Surround-Kanäle abbilden (Left-Surround – Ls und Right-Surround – Rs). Diese werden nicht in der Höhe verändert, sondern befinden sich in derselben Ebene wie die Front-Lautsprecher. Je nachdem ob es fünf oder sieben Lautsprecher sind, verändert sich der Winkel zur Mitte. Um das gewollte Surround-Sound-Feld wiederzugeben, müssen die Lautsprecher in präzise definierten Positionen in Relation zum Zuhörer aufgestellt

¹ Hugh Robjohns, „Surround Sound Explained: Part 4“, *Sound on Sound/Techniques/Monitors-Monitoring*, November 2001, <https://www.soundonsound.com/techniques/surround-sound-explained-part-4>.

Der LFE-Kanal wird primär für intensive Tieffrequenzeffekte mit hohem Pegel genutzt und ist bei normalen Audiosignalen in der Regel nicht nötig.³

Der Grund, warum es nur einen Tieftonlautsprecher in einer Mehrkanalton-Anordnung gibt, ist, dass der Mensch zwar die hohen Töne lokalisieren kann, die tiefen Basstöne jedoch nicht. Daher reicht ein LFE-Kanal, um die tiefen Töne zu übertragen.⁴ Breitband- oder Fullrange Lautsprecher sind vorteilhaft, da sie die tiefen Frequenzen im gesamten Raum gleichmäßig verteilen können.⁵ Man spricht hier von Bassmanagement.

Mittlerweile findet Surround-Sound in vielen Bereichen Anwendung. Das schließt Kino, aber auch Heimkino Systeme und Streaming-Dienste ein, sowie Gaming und Virtual Reality (VR).

2.2 Immersives Audio

Der Begriff immersiv kommt von dem Wort „immersio“ aus dem Lateinischen, was übersetzt „Eintauchen“ oder „Hineinversetzen“ bedeutet. Wenn man von immersivem Audio oder auch spatial Audio spricht, lässt man den Zuhörer vollständig in die Klangumgebung eintauchen. Man vermittelt ihm hierbei ein Gefühl von Präsenz, indem man ihn aus allen möglichen Richtungen mit gerichteten oder ungerichteten Schallquellen umhüllt. Gerichtete Klänge sind Punktschallquellen, die dem Zuhörer nah sind und die man eindeutig lokalisieren kann. Ungerichtete Schallquellen sind weiter entfernt, sie wirken mit der Akustik des Raumes zusammen und stellen die Umgebung dar, zudem sind sie nicht eindeutig lokalisierbar. Ein umhüllendes Klangbild erreicht man, indem man beide Arten von Schallquellen kombiniert. Zum Beispiel sind Schritte oder Gespräche an einem Bahnhof gerichtete Klangquellen, einfahrende Züge oder Durchsagen sind dagegen ungerichtete Klangquellen. Dadurch entsteht eine hallende und umhüllende Atmosphäre.⁶

³ Florian C. Scholz, „10 - Surround und Mehrkanalton“, in *Audiotechnik für Mediengestalter* (De Gruyter Saur, 2015), S. 228.

⁴ Andreas A. Reil und Wilim Brucker, „Vom Einkanal- zum Surroundton“, in *Praxisbuch: Ton - perfekter Ton für News, TV-Beiträge, Industriefilme und Eigenproduktionen* (mediabook Verlag, 2007). S. 240.

⁵ Günther Theile, „Kapitel 5.4.1.3 - Tieftonwiedergabe“, in *Handbuch der Tonstudientechnik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 304.

⁶ Agnieszka Roginska und Paul Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel-Audio* (New York: Taylor & Francis, Routledge, 2017). S. 1.

Das Ziel von immersivem Audio ist also, dem Zuhörer ein möglichst realistisches und intensives Erlebnis zu ermöglichen.⁷

Um beim Zuhörer dieses Gefühl zu erreichen, völlig von den Klängen aus allen Richtungen eingenommen zu werden, kann man die bestehende horizontale Surround-Ebene aus fünf bzw. sieben Lautsprechern durch Höhenlautsprecher in der vertikalen Ebene ergänzen. Somit erreicht man ein dreidimensionales Klangfeld.⁸ Die Positionen der Schallquellen, des Zuhörers und der Raumbegrenzungen zueinander erzeugen auditive Hinweise, die einen räumlichen Eindruck vermitteln.⁹

Die Nutzung von objektbasierter Audiotechnologie ermöglicht es, einzelne Klangobjekte unabhängig im Raum zu platzieren und dynamisch zu bewegen. Dabei enthalten die Metadaten Information darüber, wo im Raum sich das Klangobjekt befindet, also wohin es zu welchem Zeitpunkt gepannt ist.¹⁰ Objektbasiertes Audio wird im Kapitel „4. 4. 3D-Audio im Fernsehbereich / Objektbasiertes Audio (Next Generation Audio)“ beschrieben.

2.3 Binaurales Audio und binaurale Aufnahmetechnik

Den Ursprung findet „binaural“ im Lateinischen, was übersetzt „mit beiden Ohren“ bedeutet. Wenn man von binauralem Audio spricht, handelt es sich um eine Technologie, die dem Zuhörer beim Nutzen von Kopfhörern einen realistischen Eindruck des dreidimensionalen Raums vermittelt. Dabei werden die Klänge im Ohr so präzise nachgestellt, wie sie auch in der echten Welt auf die beiden Ohren treffen würden. Die binaurale Audiotechnologie simuliert, wie akustische Signale am Höreingang links und rechts wahrgenommen werden und stellt somit das natürliche räumliche Hören nach. Es werden komplexe 3D Klangwelten mit nur zwei Audiokanälen erzeugt. Am häufigsten werden Kopfhörer verwendet, um den natürlichen dreidimensionalen Raum im Ohr zu simulieren. Dennoch gibt es auch Lautsprechertechniken, die binaurale Wiedergabe ermöglichen.

⁷ Chris Pike, „Chapter 1 - 3D Audio in Broadcasting“, in *3D Audio - Perspectives on Music Production*, hg. von Hyunkook Lee & Justin Paterson, 1. Auflage (New York: Taylor & Francis, Routledge, 2022), <https://doi.org/10.4324/9780429491214>, S.1.

⁸ Edwin Pfanzagl-Cardone, „4.3 Multichannel Speaker-Layout: Improved Audio Quality and Timbre Matching“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 151.

⁹ Roginska und Geluso, *Immersive Sound: The Art and Science of Binaural and Multi-Channel-Audio*. S. 1.

¹⁰ Pfanzagl-Cardone, „4.3 Multichannel Speaker-Layout: Improved Audio Quality and Timbre Matching“. S. 151.

Insbesondere Rundfunkanstalten können von dieser Technologie profitieren. Man benötigt keine neuen oder speziellen Geräte, sondern kann die bestehenden Infrastrukturen, die man auch für Stereo nutzt, beibehalten und trotzdem Teil eines immersiveren 3D Audioerlebnisses werden. ¹¹

Denn binaurale Audiosysteme können Teil von immersiven Systemen sein, um 3D Audio über nur zwei Kanäle zu erleben.

Binaurale Aufnahmetechnik bezieht sich darauf, dass die Schalldrücke in den beiden Trommelfellen aufgenommen und unverändert wiedergegeben werden. Das ermöglicht eine vollständige Reproduktion des Hörerlebnisses und schließt räumliche Aspekte sowie die Klangfarbe ein. ¹² Die Klangfarbe beinhaltet die spektrale Zusammensetzung und Qualität des Klangs, die auch von Reflexionen beeinflusst wird.

Für die Aufnahme von binauralem Audio kann man kleine Mikrofone in den Gehörgängen eines menschlichen Zuhörers platzieren. Häufiger verwendet man allerdings einen Kunstkopf, also eine exakte Nachbildung des durchschnittlichen menschlichen Kopfes mit Nase, Augenhöhlen, Ohrmuscheln und Gehörgängen. In einigen Fällen simuliert man sogar den akustischen Widerstand des Trommelfells. Es soll mit dieser Nachbildung eines menschlichen Kopfes sichergestellt werden, dass die Schallwellen den Kopf und die Gehörgänge auf die Weise erreichen, wie sie es bei einem echten Zuhörer tun. Die Kunstkopfaufnahme bietet eine präzise Abbildung aller Schalleinfallrichtungen und gibt dem Zuhörer in Bezug auf Raumeindruck und Umhüllung das Gefühl, selbst am Ort der Aufnahme zu sein.

13

¹¹ Pike, „Chapter 1 - 3D Audio in Broadcasting“, S.2.

¹² Henrik Møller, „Fundamentals of binaural technology“, *Applied Acoustics* 36, Nr. 3 (1992): 171–218, [https://doi.org/10.1016/0003-682X\(92\)90046-U](https://doi.org/10.1016/0003-682X(92)90046-U), S. 171-172.

¹³ Gerhard Spikofski, „Kapitel 5.5.5.1 - Kopfbezogene Übertragung“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Band 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 374.



Abbildung 2: Neumann KU8li Kunstkopf

Quelle: Pfanzagl-Cardone, 2023, S. 18

Um die Aufnahmen wiederzugeben, verwendet man Kopfhörer, damit der in einem Ohr aufgenommene Schall auch nur in diesem entsprechenden Ohr des Zuhörers wiedergegeben wird. Eine Lautsprecherwiedergabe würde dazu führen, dass der Schall jedes Lautsprechers auf beiden Ohren ankommt. Dieses unerwünschte Übersprechen würde den binauralen Effekt aufheben.¹⁴

Um die Überlagerung der Lautsprechersignale und das daraus folgende Übersprechen zu verhindern, kann man Crosstalk-Cancellation-Filter oder Transaural-Filter in den Lautsprechern vorschalten. Jedoch gelten diese Filter nur für eine bestimmte Kopfposition und -ausrichtung. Wenn sich der Zuhörer in diesem Umfeld bewegen möchte, kann man head tracking anwenden. Dabei ermittelt man laufend die Ausrichtung und Position des Kopfes und dementsprechend werden Filter kontinuierlich angepasst.¹⁵

2.4 Lokalisierung von Hörereignissen

2.4.1 Menschliches Hören

Um zu verstehen, wie Schall beim menschlichen Zuhörer ankommt und wahrgenommen wird, ist es sinnvoll sich auch auf biologischer Ebene mit dem menschlichen Ohr auseinanderzusetzen. Es ist als Sinnesorgan für die Wahrnehmung von Schall und Gleichgewicht verantwortlich. Das Gehör schließt

¹⁴ Møller, „Fundamentals of binaural technology“, S. 172.

¹⁵ Günther Theile, „Kapitel 5.5.5.3 - Lautsprecherwiedergabe“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. Band 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 377-378.

auch die Verarbeitung von Sinneseindrücken im Nervensystem ein und sorgt dafür, dass ein Schallereignis im Hirn auch als Hörereignis ankommt.

Grundlegend besteht das Ohr aus dem Außenohr, Mittelohr und Innenohr. ¹⁶

Schall trifft auf den menschlichen Kopf, dabei rufen Schultern, Kopf und Brust sowie die Ohrmuschel Veränderungen des Klangcharakters hervor, bevor der Schall im Eingang des Gehörgangs ankommt. ¹⁷

Die Ohrmuschel, also das Außenohr des Menschen besteht aus Vertiefungen und Erhebungen, die akustische Resonatoren bilden. Diese Resonatoren werden bei Schalleinfall aus verschiedenen Richtungen jeweils unterschiedlich angeregt¹⁸ und haben somit den größten spektralen Einfluss auf die Klangfarbe. ¹⁹ Auch im Gehörgang selbst werden die Frequenzantwort und andere Parameter noch beeinflusst und verändert. ²⁰

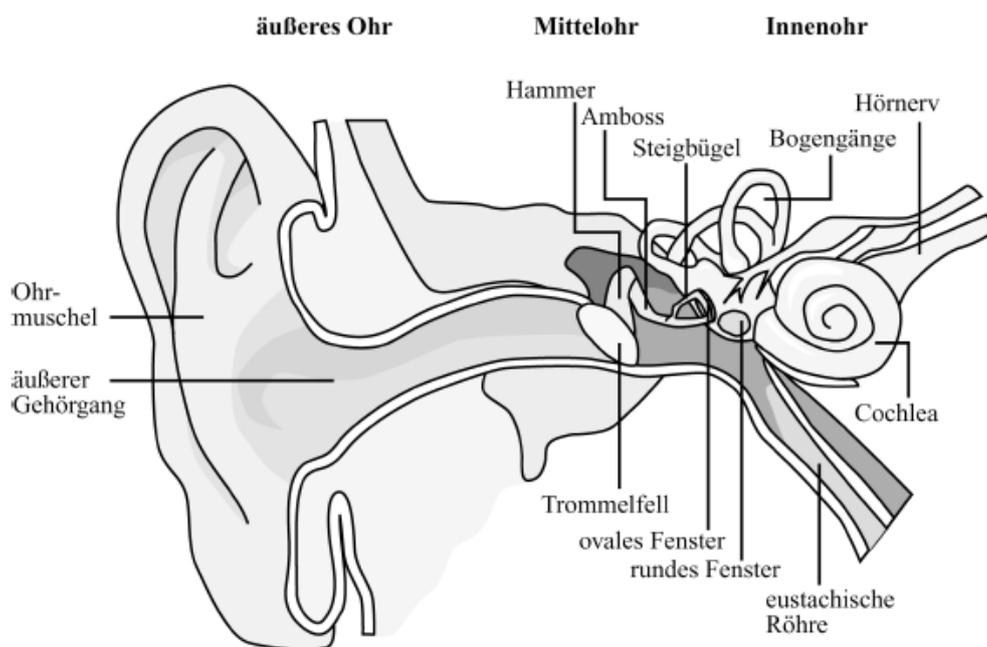


Abbildung 3: Anatomie des menschlichen Ohrs (Außenohr, Mittelohr, Innenohr)

Quelle: Dickreiter, 2023, S. 117

¹⁶ Jürgen Goeres-Petry, „Kapitel 3.1 - Das Gehör“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. Band 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 117.

¹⁷ Edwin Pfanzagl-Cardone, „2.2 Basics of Sound Perception in Humans“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 55.

¹⁸ Jürgen Goeres-Petry, „Kapitel 3.1.1 - Außenohr“, in *Handbuch der Tonstudioteknik*, 9. aktualisierte und erweiterte Auflage, Bd. Band 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 117.

¹⁹ Pfanzagl-Cardone, „2.3 Mechanisms of Localization“, S. 64.

²⁰ Edwin Pfanzagl-Cardone, „2.3 Mechanisms of Localization“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 64.

Das Mittelohr besteht aus dem Trommelfell und den Gehörknöchelchen Hammer, Amboss und Steigbügel. Das Trommelfell wandelt dabei, ähnlich wie eine Mikrofonmembran, den Luftschall in mechanische Schwingungen um.

Über das ovale Fenster, das sich zwischen dem Mittelohr und Innenohr befindet, gelangt diese mechanische Energie in das Innenohr, wo es zu Druckänderungen in der Flüssigkeit kommt. Abhängig von der Frequenz erzeugt dieser Druck verschiedene Schwingungsmuster auf der sogenannten Basilarmembran. Die feinen Härchen bzw. die Haarzellen, die in die Flüssigkeit hineinreichen, reagieren auf die Druckänderungen und lösen somit Nervenimpulse aus. Aus den Nervenimpulsen werden elektrische Aktionspotenziale in den Neuronen des Hörsystems erzeugt und schließlich zusammen mit den Informationen aus dem anderen Ohr im Hirn verarbeitet und kombiniert.²¹

2.4.2 Lokalisation von Schallereignissen

Bei der Lokalisierung von Schallereignissen handelt es sich um eine möglichst präzise Bestimmung des Ortes, von dem ein Geräusch herkommt.

Anhand bestimmter Merkmale stellt man fest, aus welcher Richtung und Entfernung das Schallereignis den Zuhörer erreicht.²²

Wie eine Schallquelle im Raum wahrgenommen wird, hängt von der Lokalisation in der horizontalen Ebene (auch Azimuth genannt) und der vertikalen Ebene (auch Elevation) ab, außerdem spielt auch der räumliche Eindruck eine Rolle, der sich aus der Impulsantwort des Raumes ableiten lässt.²³

Ein unmittelbares Schallereignis, welches in einem realen Raum stattfindet und nicht elektronisch erzeugt wurde, zum Beispiel durch Lautsprecher, enthält einige frequenzabhängige Informationen.

Die Interaural Time Difference (ITD) beschreibt Laufzeitunterschiede zwischen den beiden Ohren.²⁴ Schallwellen treffen nicht gleichzeitig auf das linke und rechte Ohr auf. Sollte die Lokalisation ungenau sein, kann der Zuhörer mit

²¹ Pfanzagl-Cardone, „2.2 Basics of Sound Perception in Humans“, S. 55-56.

²² Jens Blauert, „2.1 Localization and Localization Blur“, in *Spatial Hearing: The Psychophysics of Human Sound Localization* (The MIT Press, 1996), <https://doi.org/10.7551/mitpress/6391.001.0001>, S. 37.

²³ Pfanzagl-Cardone, „2.2 Basics of Sound Perception in Humans“, S. 56.

²⁴ Nate Mars, „Producing Spatial Audio in Dolby Atmos, Binaural, and Surround“, LinkedIn, 28. Juli 2022, <https://www.linkedin.com/learning/producing-spatial-audio-in-dolby-atmos-binaural-and-surround/>.

kleinen Kopfbewegungen die Zeitdifferenzen verändern und somit die Position der Schallquelle bestimmen. ²⁵

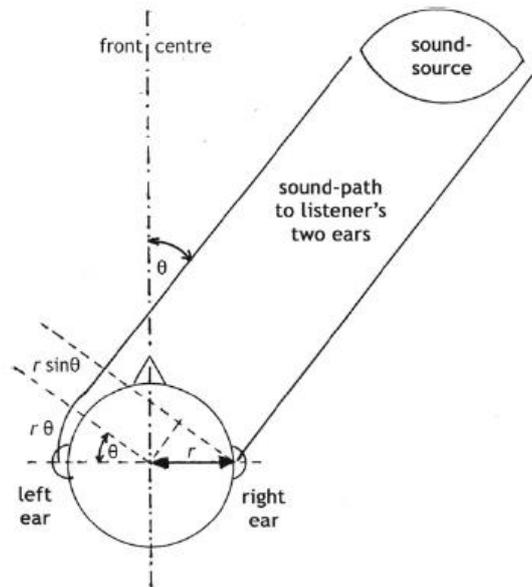


Abbildung 4: Demonstration der ITD einer Schallquelle seitlich kommend

Quelle: Pfanzagl-Cardone, 2023, S. 58

Es gibt neben der ITD auch interaurale Pegelunterschiede (ILD bzw. IID) zwischen den beiden Ohren, basierend auf denen der Zuhörer ebenfalls orten kann, woher ein Schallereignis kommt und wie weit es entfernt ist. ²⁶

Unterschiedliche Frequenzen haben unterschiedlich lange Schallwellen und erreichen daher auch auf unterschiedliche Art den menschlichen Kopf. Tiefe Frequenzen haben entsprechend lange Wellenlängen, im Verhältnis dazu ist der Durchmesser des menschlichen Kopfes ziemlich gering. Das kann dazu führen, dass der Schall bei niedrigen Frequenzen um den Kopf gebogen wird. Der Pegelunterschied zwischen dem linken und rechten Ohr ist dann eher gering und das hat zur Folge, dass die Ortung des Schalls und die räumliche Wahrnehmung für den Zuhörer erschwert werden. ²⁷

²⁵ Lord Rayleigh, „XII. On our perception of sound direction“, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13, Nr. 74 (1. Februar 1907): 214–32, <https://doi.org/10.1080/14786440709463595>.

²⁶ Mars, „Producing Spatial Audio in Dolby Atmos, Binaural, and Surround“.

²⁷ Edwin Pfanzagl-Cardone, „1.5 Inter Aural Cross-Correlation (IAAC) and the Binaural Quality Index of Reproduced Music (BQIrep)“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 22.

Schallwellen von hohen Frequenzen zeichnen sich dagegen durch kurze Wellenlängen aus, die kleiner sind als der Durchmesser des menschlichen Kopfes. Sie werden für das andere Ohr effektiv abgeschattet. Daraus entsteht ein großer Pegelunterschied zwischen den beiden Ohren.

Die beiden Informationen ITD und ILD bilden die Basis, auf der das menschliche Gehör entscheidet, wo die Schallquelle positioniert ist. Bei niedrigen Frequenzen treten keine erwähnenswerten interauralen Pegeldifferenzen (ILD) auf, daher basiert die Lokalisierung in diesem Bereich vor allem auf den interauralen Zeitunterschieden (ITD).²⁸ Die Zeit, mit der sich der Schall von einem Ohr zum anderen ausbreitet, nimmt bei niedrigen Frequenzen zu. Schallwellen werden an glatten, konvexen Oberflächen wie Kopf und Ohrmuschel reflektiert und gebeugt. Dadurch kommt es zu einer Phasenverschiebung und verlängerter Laufzeit. Die ITD ist also aufgrund der Reflexionen am Kopf frequenzabhängig und funktioniert im niedrigen Frequenzbereich besser.²⁹

Die Lokalisierung von niedrigen Frequenzen erfolgt vor allem durch Zeitunterschiede, während hohe Frequenzen dagegen durch Pegelunterschiede (ILD) lokalisiert werden. Diese beiden psychoakustischen Mechanismen werden etwa bei 1,5kHz Übergangsfrequenz getrennt.³⁰ Die genannten Merkmale zur Lokalisation in der Horizontalebene sind allerdings eher individuell, da sie von der Größe und Form des Kopfes, der Ohrmuscheln sowie Schultern und Brust abhängig sind.³¹

Man unterscheidet zwischen dem Richtungshören in der Horizontalebene und dem Richtungshören in der Medianebene. Die Lokalisierung in der Horizontalebene geschieht vor allem über die interauralen Pegel- oder Zeitdifferenzen. In der Medianebene, die den Bereich direkt vor, hinter und über dem Kopf einschließt³², existieren diese Bestimmungsparameter kaum, da das Signal durch die Symmetrie des Kopfes an beiden Ohren nahezu identisch ankommt.³³

²⁸ Rayleigh, „XII. On our perception of sound direction“.

²⁹ Edwin Pfanzagl-Cardone, „2 - ‚3D‘ or ‚Immersive‘ Audio - The Basics and a Primer on Spatial Hearing" in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 54.

³⁰ Rayleigh, „XII. On our perception of sound direction“.

³¹ Pfanzagl-Cardone, „2.3 Mechanisms of Localization“, S. 61.

³² Brad Rakerd, William Hartmann, und Timothy McCaskey, „Identification and localization of sound sources in the median sagittal plane“, *The Journal of the Acoustical Society of America* 106 (1. Dezember 1999): 2812–20, <https://doi.org/10.1121/1.427615>.

³³ Blauert, „2.1 Localization and Localization Blur“, S. 44.

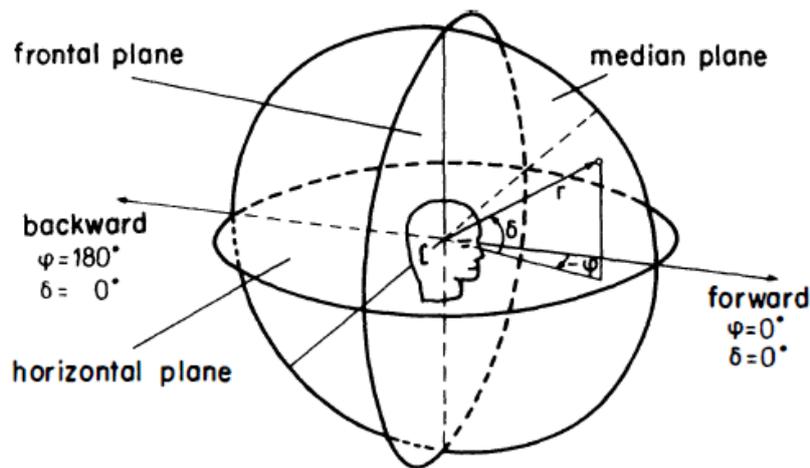


Abbildung 5: Kopfbezogenes Koordinatensystem, Definition von Medianebene, Frontalebene und Horizontalebene

Quelle: Blauert, 1996, S. 14

Die Entscheidung, ob sich eine Schallquelle vor oder hinter dem Zuhörer befindet und in welchem Erhebungswinkel, trifft der Zuhörer in der vertikalen Ebene durch spektrale Hinweise und Veränderungen.³⁴ Eine Schallwelle trifft auf den Kopf des Zuhörers, dabei rufen der Kopf, die Ohrmuschel und der Gehörgang Beugungs-, Resonanz- und Abschattungseffekte hervor, die das Signal linear verzerren. Je nach Einfallswinkel variiert das Spektrum des Signals, was vom Zuhörer als Klangfarbenveränderung wahrgenommen wird. Diese spektralen Hinweise haben auch Einfluss auf die Lokalisierung des Hörereignisses.³⁵

Die sogenannte Head-related Transfer Function (HRTF) oder Außenohr-Übertragungsfunktion wird verwendet, um die spektralen Veränderungen einer Schallwelle vom Ursprungsort bis zum Trommelfell mathematisch zu berechnen und messtechnisch zu erfassen. Eine Messung erfolgt in reflexionsfreien Räumen, indem man der Versuchsperson in jeden Ohrkanal je ein Miniaturmikrofon einsetzt. Man sendet über einen Lautsprecher ein Signal aus, das anschließend von den Miniaturmikrofonen als transformierter Schall aufgenommen wird. Daraus lassen sich die spektralen Veränderungen zwischen dem abgesendeten

³⁴ Michael Dickreiter, „Kapitel 3.4 - Räumliches Hören natürlicher Schallquellen“, in *Handbuch der Tonstudioteknik* hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Band 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 130.

³⁵ Jens Blauert, „Sound Localization in the Median Plane“, *Acta Acustica united with Acustica* 22 (30. November 1969).

Schallsignal und dem Signal, das am Trommelfell ankommt, bestimmen.³⁶ Eine HRTF kann als Frequenzspektrum der kopfbezogenen Impulsantwort (HRIR) definiert werden. Bei komplexen Schallereignissen ist die HRTF der wichtigste Anhaltspunkt für das Gehirn, um Verwechslungen zwischen vorne und hinten aufzulösen und eine erhöhte Schallquelle zu lokalisieren.³⁷

Der Zuhörer lokalisiert eine Schallquelle korrekt, wenn sein Richtungsempfinden des Schallsignals mit dem tatsächlichen Ort der Schallquelle übereinstimmt. Es ist keine Voraussetzung, dass das Signal dem Zuhörer bereits bekannt ist. Das Schallsignal muss lediglich eine ausreichende Bandbreite besitzen.³⁸ Die Lokalisation von schmalbandigen Signalen in der Medianebene ist nicht möglich.

³⁹ Blauert führte im Jahr 1969 mehrere Experimente mit Probanden durch. Er stellte fest, dass es Frequenzbänder gibt, die fest mit der Lokalisation in bestimmten Bereichen der Medianebene zusammenhängen. Er führte den Testpersonen Schmalbandrauschen von verschiedenen Lautsprechern vor, aufgrund dessen beurteilt werden sollte, ob sich das Signal vorne („v“), hinten („h“) oder oben („o“) befindet. Abbildung 6 zeigt auf der Y-Achse, mit welcher relativen Häufigkeit die Beobachteraussagen auf die einzelnen Urteile „v“, „h“ oder „o“ fiel. Eines der beiden Urteile wurde öfter angegeben als die beiden anderen Urteile zusammen.

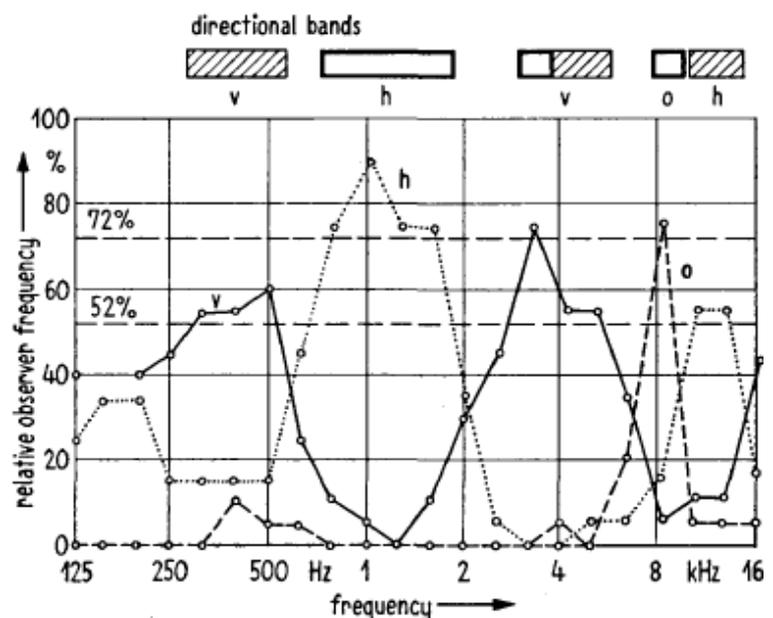


Abbildung 6: Relative Häufigkeit der Beobachter, die eines der Urteile "v", "o" oder "h" häufiger abgeben als die beiden andern zusammen (Signifikanzniveau 95%)

Quelle: Blauert, 1969, S. 209

³⁶ Jens Blauert und Jonas Braasch, „Räumliches Hören“, in *Handbuch der Audiotechnik*, hg. von Stefan Weinzierl (Berlin, Heidelberg: Springer Berlin Heidelberg, 2008), https://doi.org/10.1007/978-3-540-34301-1_3, S. 90.

³⁷ Hyunkook Lee, „Chapter 5 - Psychoacoustics of Height Perception in 3D Audio“, in *3D Audio*, 1. Auflage, Perspectives on Music Production (New York: Taylor & Francis, Routledge, 2022), S. 83.

³⁸ Blauert, „Sound Localization in the Median Plane“.

³⁹ Michael Dickreiter, „Kapitel 3.4.1.2 - Vertikale Ebene“, in *Handbuch der Tonstudioteknik* hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 134.

Auf der X-Achse ist die Frequenz dargestellt. Der obere Rand zeigt die sogenannten richtungsbestimmenden Bänder auf, es handelt sich dabei um Bereiche, in denen die absolute Mehrheit der Probanden eines der beiden Urteile häufiger abgibt als die beiden anderen zusammen. Die Umrandung des Blocks bedeutet ein Signifikanzniveau von 90%, der schattierte Bereich bedeutet höchstwahrscheinlich. ⁴⁰

Der Mensch lokalisiert Schallereignisse in der Medianebene insbesondere durch das Außenohr, welches eintreffende Schallsignale filtert. Bestimmte Frequenzbänder eines Breitbandsignals werden dabei verstärkt, andere werden abgesenkt, abhängig von der Schalleinfallrichtung. In der Regel stimmen die verstärkten Bereiche mit den richtungsbestimmenden Bändern überein, die mit der Schalleinfallrichtung verknüpft sind. Man nimmt an, dass diese richtungsbestimmenden Bänder, die den größten Anteil der Signalstärke enthalten, die Richtung der Schallempfindung bestimmen. Im Falle einer etwa gleichen Signalstärke im „v“- und „h“-Bereich wird die Schallempfindung im Kopf lokalisiert oder trennt sich in einen vorderen und hinteren Teil auf. Wenn ein ausreichender Anteil der Signalstärke in das „o“-Band fällt, empfindet der Zuhörer den Schall aus einem bestimmten Elevationswinkel. ⁴¹

Der Frequenzbereich um etwa 8kHz wird vor allem im Richtungsband „oben“ bestimmt, während „vorne“ insbesondere durch Signalteile im Spektrum von 300-600Hz sowie von 3000-6000 Hz bestimmt wird. „Hinten“ lokalisiert man vor allem die Frequenzbänder um 1200Hz und 12kHz. ⁴²

Eine Schallquelle strahlt in einem Raum einen Schallimpuls in alle Richtungen aus. Dieser Impuls wird von den Wänden und Gegenständen im Raum reflektiert, absorbiert, gebeugt, zerstreut oder auch gebündelt, es handelt sich dabei um die Impulsantwort des Raumes. Was genau mit der Schallwelle passiert, hängt von dem Material, der Form und Größe der Gegenstände ab. Neben dem Direktschall, der den Hörer auf dem kürzesten Weg erreicht, treffen zunächst einige einzelne Reflektionen, die sogenannten ersten oder frühen Reflexionen (ER – early reflections) ein. Die Verzögerung, Stärke und Einfallrichtung dieser ersten

⁴⁰ Blauert, „Sound Localization in the Median Plane“.

⁴¹ Blauert und Braasch, „Räumliches Hören“, S. 94-95.

⁴² Edwin Pfanzagl-Cardone, „2.3.2 Localization and HRTFs“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 65.

Reflexionen haben eine entscheidende Bedeutung für den Hörer und seine Lokalisation des Hörereignisses. Je mehr Zeit vergeht, desto mehr verdichten sich die Reflexionen, sie werden späte Reflexionen genannt (LR – late reflections). Es handelt sich dabei um den Nachhall, der mit einer Verzögerung zum Direktschall einsetzt. ⁴³

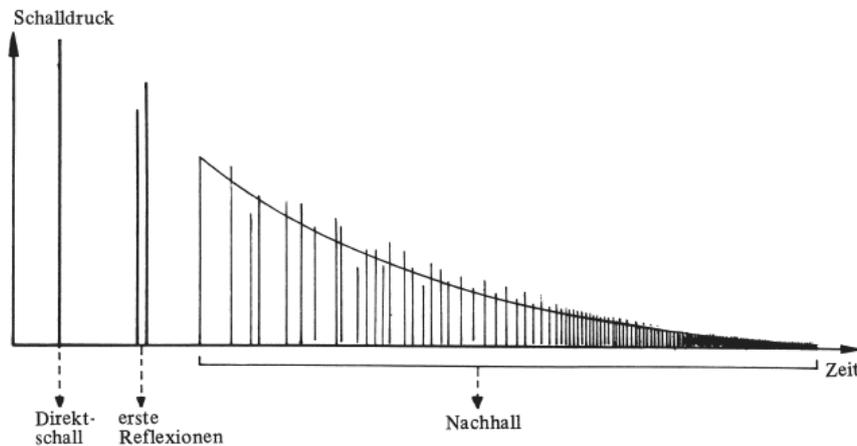


Abbildung 7: Zeitliche Folge der Reflexionen in einem Raum bei Impulsschall

Quelle: Dickreiter, 2023, S. 31

Die ersten Reflexionen treten in den ersten 80ms nach dem Schallereignis auf. Sie helfen dem Zuhörer, einen Eindruck von seiner Position im Raum zu gewinnen aber auch, das Hörereignis zu lokalisieren. Die späten Reflexionen, also der Nachhall, werden als deutlich diffuser und mit schwächerem Pegel wahrgenommen, was aus den mehrfachen Reflexionen an den Raumbegrenzungen und Objekten resultiert. ⁴⁴

Beim Zuhörer treffen dann durch die Reflexionen innerhalb kurzer Zeit zwei gleiche Schallereignisse nacheinander ein. Solange diese nur durch eine Verzögerung von 1-30ms getrennt sind, bestimmt hier die erste Wellenfront die scheinbare Lokalisation. Beim „Gesetz der ersten Wellenfront“, auch Haas-Effekt oder Präzedenz-Effekt genannt, entscheidet die erste eintreffende Wellenfront, aus welcher Richtung der Zuhörer das Hörereignis wahrnimmt. Die nachfolgenden Reflexionen haben für den Richtungseindruck kaum eine Bedeutung und können vernachlässigt werden. Der Pegel des nachfolgenden

⁴³ Michael Dickreiter, „Kapitel 1.3.1 - Zeitlicher Aufbau des Schallfeldes“, in *Handbuch der Tonstudiotchnik* hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 30.

⁴⁴ Edwin Pfanzagl-Cardone, „2.5 Spatial Impression“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 69.

Schalls bei Laufzeitdifferenzen zwischen 5 und 30ms kann sogar bis zu 10dB über dem Pegel des Primärschalls liegen, das Gesetz der ersten Wellenfront verliert dennoch nicht an Bedeutung.^{45 46 47}

In der vertikalen Ebene sind interaurale Zeitdifferenzen unwirksam, daher funktioniert auch der Präzedenz-Effekt nicht bei vertikal angeordneten Lautsprechern, unabhängig davon welche Zeitdifferenz man zwischen den Signalen der Lautsprecher anwendet.⁴⁸

2.5 Lautsprecheranordnungen

Die ITU-R BS.2051-3 – Advanced Sound System for Programme Production aus dem Jahr 2022 stellt unter anderem eine Definition für Lautsprecheraufstellungen bei der Produktion von kanalbasierten, objekt- und szenenbasierten sowie hybriden Audiomischungen auf.

Die ITU definiert „o + m + u“ als Schreibweise für Lautsprecheraufstellungen. O beschreibt dabei die Anzahl der Lautsprecher in der oberen Ebene, m die Anzahl der Lautsprecher in der mittleren Ebene und u die Anzahl der Lautsprecher in der unteren Ebene. Der LFE-Kanal wird bei dieser Schreibweise nicht berücksichtigt.

⁴⁹ Diese Arbeit nutzt aufgrund der Leserlichkeit nicht die angegebene Schreibweise der ITU, sondern die häufig genutzte Schreibweise x.y.z, wobei x die Anzahl der Lautsprecher in der horizontalen Ebene beschreibt, y die Anzahl an LFE-Kanälen und z die Anzahl der Surround-Lautsprecher in der oberen Ebene. Die Anordnung 5.1.4 beschreibt also eine Aufstellung mit 5 Hauptkanälen in der mittleren Ebene, einem LFE-Kanal und 4 Höhenkanälen.⁵⁰

⁴⁵ Michael Dickreiter, „Kapitel 3.4.1.1 - Horizontale Ebene“, in *Handbuch der Tonstudiotchnik* hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 133.

⁴⁶ Pfanzagl-Cardone, „2.3 Mechanisms of Localization“, S. 63.

⁴⁷ Blauert und Braasch, „Räumliches Hören“, S. 103.

⁴⁸ Pfanzagl-Cardone, „2 - ‚3D‘ or ‚Immersive‘ Audio - The Basics and a Primer on Spatial Hearing“ in *The Art and Science of 3D Audio Recording.*“, S. 53.

⁴⁹ International Telecommunication Union, „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production‘“ (Radiocommunication Sector of ITU, Mai 2022).

⁵⁰ Dolby Laboratories, Inc., „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, 6. Mai 2021, S. 9.

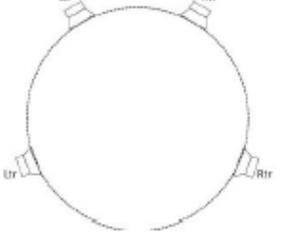
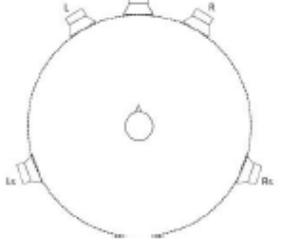
D (4+5+0)			
Upper layer 2/0/2			
Middle layer 3/0/2			
Bottom layer 0/0/0.1			

Tabelle 1: Lautsprecherlayout für Sound System D (4+5+0) - 5.1.4

Quelle: ITU, 2022, S. 15

Die linken und rechten Lautsprecher befinden sich in der mittleren Ebene bei einem Winkel von $\pm 30^\circ$ zum Zuhörer. Der Center-Lautsprecher befindet sich frontal direkt vor dem Zuhörer. Die hinteren Surround-Kanäle Ls und Rs werden zwischen ± 100 und 120° aufgestellt. Die Höhenlautsprecher orientieren sich ungefähr an den Positionen der seitlichen Surround-Lautsprecher in der mittleren Ebene. ⁵¹

**Loudspeaker configuration for Sound System B (0+5+0)
(from Recommendation ITU-R BS.775)**

SP Label	Channel		Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	-	-
M+110	Ls	Left surround	+100 .. +120	0 .. +15
M-110	Rs	Right surround	-100 .. -120	0 .. +15

Tabelle 2: Lautsprecherkonfiguration für Sound System B (0+5+0) - 5.1

Quelle: ITU, 2022, S. 9

⁵¹ International Telecommunication Union, „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production“.

Loudspeaker configuration for Sound System D (4+5+0)

SP Label	Channel		Loudspeaker location, Polar	
			Azimuth	Elevation
	Label	Name	Range	Range
M+030	L	Left	+30	0
M-030	R	Right	-30	0
M+000	C	Centre	0	0
LFE1	LFE	Low frequency effects	-	-
M+110	Ls	Left surround	+100 .. +120	0
M-110	Rs	Right surround	-100 .. -120	0
U+030	Ltf	Left top front	+30 .. +45	+30 .. +55
U-030	Rtf	Right top front	-30 .. -45	+30 .. +55
U+110	Ltr	Left top rear	+100 .. +135	+30 .. +55
U-110	Rtr	Right top rear	-100 .. -135	+30 .. +55

Tabelle 3: Lautsprecherkonfiguration für Sound System D (4+5+0) - 5.1.4

Quelle: ITU, 2022, S.10

Normalerweise werden Lautsprecher um den Zuhörer auf einer Kreisbahn platziert, wobei dieser sich mittig im Kreis befindet. Dieser Platz wird auch als „sweet spot“ bezeichnet und stellt die ideale Hörzone dar. Im Kino befindet sich der Centerkanal aus Platzgründen nicht auf einer Kreisbahn, sondern bildet mit dem L und R Kanal eine Reihe. Somit wird der Centerkanal um den jeweiligen Zeitwert verzögert, den das Signal hätte, man simuliert hier also die größere Entfernung vom Centerkanal für den Zuschauer. ⁵²

2.6 Verwendung der Höhenlautsprecher

Die bereits bekannte 5.1 Konfiguration wird um vier Lautsprecher für die Höhenkanäle erweitert. Sie sind jeweils über den seitlichen Left- und Right-Kanälen sowie über den Left-Surround und Right-Surround Kanälen angeordnet. Die Höhenlautsprecher haben in Bezug auf die Kopfhöhe des Zuhörers einen horizontalen Winkel von 30° und sind somit leicht erhöht angebracht. Der Einsatz von Lautsprechern in der oberen Ebene oder in der Decke ermöglicht es, die Wiedergabe von Audiosignalen in die dritte Dimension auszuweiten und die Lokalisation zu verbessern. Ziel ist es, das natürliche Schallfeld zusammen mit der Räumlichkeit und einer klanglichen Umhüllung zu erleben. Optional kann man

⁵² Scholz, „10 - Surround und Mehrkanalton“, S. 230.

die 5.1.4 Anordnung um weitere Kanäle ergänzen, der 3D-Audio Bereich bietet viele verschiedene Möglichkeiten, Lautsprecher anzuordnen. ⁵³

Die Lautsprecher in der oberen Ebene verfügen über dieselben Abbildungseigenschaften wie die der Horizontalebene, allerdings ohne den Center-Lautsprecher. Im oberen Bereich ergänzen die beiden stereobasierten Kanäle Ltf und Rtf die Abbildung L, C, R, außerdem können die oberen hinteren Surround-Lautsprecher Ltr und Rtr auf gleiche Weise Stereo abbilden wie die unteren. Der Einsatz von Höhenlautsprechern erweitert somit den Gestaltungsspielraum und erlaubt neue Möglichkeiten der Abbildung von Audioquellen, Reflexionen und diffusem Schall. ⁵⁴

In der vertikalen Ebene erfolgt die Lokalisierung nicht über interaurale Pegel- und Laufzeitdifferenzen, wie es in der horizontalen Ebene geschieht. Stattdessen bilden hier spektrale Hinweise die Grundlage. Das bedeutet, dass die Prinzipien der horizontalen Stereophonie nicht explizit auch für vertikale Stereophonie gelten. ⁵⁵

Eine Phantomschallquelle zeichnet sich durch eine stabile stationäre Hörereignisrichtung aus, was bedeutet, dass ihre wahrgenommene Richtung sich nicht verändert und eindeutig aus einer bestimmten Richtung kommt.

Die Erzeugung einer Phantomschallquelle zwischen den oberen und unteren Lautsprechern und die Nutzung von Panning ist aus diesem Grund nicht sinnvoll, wenn man eine möglichst große Hörzone realisieren will. Die Hörzone bildet den Hörbereich innerhalb einer Lautsprecheraufstellung ab, in dem das wiedergegebene Klangbild unabhängig von der Position annähernd identisch wahrgenommen wird. ⁵⁶ Eine Lokalisierung dieser Phantomschallquelle zwischen der oberen und unteren Ebene ist in Bezug auf Laufzeitdifferenzen unzuverlässig und außerdem vom Frequenzspektrum abhängig.

⁵³ Michael Dickreiter, „Kapitel 17.6.2.2 - Weitere Mehrkanalton-Formate“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 8. überarbeitete und erweiterte Auflage, Bd. 1 (Berlin/Boston: De Gruyter Saur, 2014), S. 1208, <https://doi.org/10.1515/9783110316506>.

⁵⁴ Günther Theile, „Kapitel 5.4.5.2 - Anwendung der Höhenlautsprecher“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 335.

⁵⁵ Pfanzagl-Cardone, „2 - ‚3D‘ or ‚Immersive‘ Audio - The Basics and a Primer on Spatial Hearing“ in *The Art and Science of 3D Audio Recording*, S 53.

⁵⁶ Günther Theile, „Kapitel 5.4.5.1 - Nutzen der Höhenlautsprecher“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 334.

Bereits geringe Lautzeitdifferenzen bewirken, dass sich die Phantomschallquelle nach unten oder oben verlagert. Mit einem Delay von 0,5ms lässt sich das Hörereignis bereits in einen anderen Lautsprecher verschieben, dazu kommen außerdem Klangverfärbungen.⁵⁷

In einem 5.1 Surroundsystem wird das Signal auf insgesamt fünf Breitbandlautsprecher verteilt, die maximal 20% des Gesamtsignals tragen. Ein 3D-Audio Setup hat deutlich mehr Wiedergabekanäle, weshalb ein einzelner Lautsprecher entsprechend einen geringeren Teil des Gesamtsignals trägt. Mit steigender Lautsprecheranzahl wird die Beziehung zwischen den Kanälen komplexer, die Bedeutung eines einzelnen Lautsprechers verringert sich allerdings, da sein Beitrag zum Gesamtsignal sich verringert.⁵⁸

Das akustische Ambiente einer Szene setzt sich aus zahlreichen Schallquellen zusammen, die im Raum verteilt sind, jedoch können sie nicht einzeln lokalisiert werden. Beispiele für diesen sogenannten richtungsdiffusen Direktschall, auch Atmo, sind Blätterrauschen im Wald, Verkehrslärm oder Publikumsreaktionen. Der diffuse Schallanteil beeinflusst beim natürlichen Hören besonders das Gefühl eines intensiven, immersiven Hörerlebnisses. Alle diffusen Signale auf den Lautsprechern müssen ausreichend unähnlich, also im gesamten Frequenzbereich dekorreliert sein, damit der aufgenommene Diffusschall auch diffus beim Hörer ankommt.⁵⁹ Die Wahrnehmung des Gefühls, von den Audiosignalen umhüllt zu sein oder sich mitten im Aufnahmeraum zu befinden, hängt davon ab, wie unähnlich sich die aufgenommenen Signale sind. Je mehr Wiedergabekanäle im System integriert sind, desto wichtiger wird die akustische Kanaltrennung bei der Aufnahme.⁶⁰ Um einen überzeugenden räumlichen Eindruck zu vermitteln, verwendet man für das 3D Multikanal-Lautsprechersystem dazu passende 3D Mikrofonarrays bei der Aufnahme. Sie gewährleisten eine hohe Trennung der Kanäle und gleichzeitige Signaldifferenzierung, insbesondere im niedrigen Frequenzbereich.⁶¹ Der Mikrofonaufbau vergrößert sich also im Vergleich zu

⁵⁷ Theile, „Kapitel 5.4.5.2 - Anwendung der Höhenlautsprecher“, S. 335.

⁵⁸ Pfanzagl-Cardone, „1.5 Inter Aural Cross-Correlation (IAAC) and the Binaural Quality Index of Reproduced Music (BQIrep)“, S. 26.

⁵⁹ Lasse Nipkow, „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“, in *Handbuch der Tonstudioteknik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 338.

⁶⁰ Theile, „Kapitel 5.4.5.2 - Anwendung der Höhenlautsprecher“, S. 336.

⁶¹ Pfanzagl-Cardone, „1.5 Inter Aural Cross-Correlation (IAAC) and the Binaural Quality Index of Reproduced Music (BQIrep)“, S. 25.

einer Stereo-Aufnahme, denn man erreicht eine ausreichende Dekorrelation nur durch Pegel- und Laufzeitdifferenzen. ⁶²

Bei einer Zweikanal-Stereo Anordnung der Lautsprecher sind die Möglichkeiten der Gestaltung begrenzt, direkter und diffuser Schall kann ausschließlich über die beiden Frontlautsprecher wiedergegeben werden. Im Gegensatz dazu bietet es sich bei Multikanal-Lautsprechersystemen an, die Einfallrichtungen von diffusen Reflektionen und Direktschall für den Zuhörer zu trennen. Den Frontlautsprechern L, C, R wird normalerweise der Direktschall zugeteilt. Seitliche oder hintere Lautsprecher strahlen vor allem den Diffusschall bzw. die Atmo ab. In geringem Ausmaß wird der Diffusschall auch auf die Frontlautsprecher verteilt, wenn eine akustische Umhüllung abgebildet werden soll. ⁶³

Aufgrund von vielen unterschiedlichen Laufzeitdifferenzen der einzelnen Reflexionen, die an den Mikrofonen ankommen, gibt es die Möglichkeit, die frühen Reflexionen in die obere Ebene zu verteilen. In einem natürlichen Umfeld fallen Reflexionen aus allen Richtungen ein, unter anderem ebenfalls von oben. Indem man die Reflexionen auf die Gesamtzahl der Wiedergabekanäle aufteilt, wird die räumliche Dichte verringert. Das Gehör kann also die räumlichen Informationen besser wahrnehmen und identifizieren. ⁶⁴

Wenn alle Lautsprecher eines 3D Audio Setups diffusen Raumschall wiedergeben und dabei eine Korrelation von etwa 0 haben, sich die Signale also kaum ähnlich sind, empfindet der Zuhörer eine akustische Umhüllung, bei der die einzelnen Lautsprecher nicht als Einzelschallquellen wahrnehmbar sind. Vor allem wenn Frequenzen unter 200Hz von allen Lautsprechern mit einer Korrelation um null wiedergegeben werden, wird die Wahrnehmung der Umhüllung unterstützt. Das Gleichgewicht der Pegel zwischen Diffusschall und Direktschall muss stimmen, damit eine auditive Verbindung zwischen oben, unten und den seitlichen Lautsprechern gewährleistet werden kann. Die Frontlautsprecher werden mit Direktschall und einem kleinen Teil Diffusschall bespielt. Die hinteren, oberen und

⁶² Nipkow, „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“, S. 338.

⁶³ Edwin Pfanzagl-Cardone, „1.7 Some Thoughts on Psychoacoustic Signal Interaction in Multichannel Microphone Array Systems“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 32.

⁶⁴ Theile, „Kapitel 5.4.5.2 - Anwendung der Höhenlautsprecher“, S. 336.

seitlichen Lautsprecher sollen ein vollständig umhüllendes Klangbild abbilden und geben deshalb den Diffusschall zum größten Teil wieder. Wenn der Pegelunterschied zwischen vorne und hinten allerdings zu groß ist, wird der Direktschall vorne von der diffusen Schallenergie überstrahlt. Die auditive Verbindung ist für den Zuhörer sehr schwach.⁶⁵ Der Zuhörer sollte nicht in der Lage sein, Klanginformationen einem einzelnen bestimmten Lautsprecher zuzuordnen, in diesem Fall wurde das Ziel einer natürlichen Wiedergabe nicht erreicht.⁶⁶ Strahlen die Frontlautsprecher gar keinen Raumschall (Atmo) ab, identifiziert der Zuhörer zwei voneinander getrennte Hörereignisse. Die auditive Verbindung und das Gefühl, vom Klang umhüllt zu sein, geht somit verloren. Zwischen der oberen und mittleren Ebene können die Pegel variieren, ohne dass diese Effekte auftreten. Es besteht die Möglichkeit, bestimmte Klänge zu verstärken, um sie von den anderen Schallereignissen abzuheben.⁶⁷

2.7 Kanalbasierte 3D Mikrofonanordnungen

Um 3D-Audio Content für die neuen 3D-Audioformate zu produzieren, sind angepasste Techniken notwendig. 3D-Audioinhalte nutzen grundsätzlich mehrere Tonspuren und können entweder kanalbasiert, objektbasiert, szenenbasiert oder durch die Kombination dieser Techniken erstellt werden.⁶⁸ Kanalbasiertes Audio wird für eine bestimmte Lautsprecherkonfiguration aufgenommen und wiedergegeben, bei der jede Tonspur einem festen Lautsprecher zugewiesen wird. Bei objektbasiertem Audio wird jeder Audiostream als eigenes Klangobjekt beschrieben. Metadaten machen die Audioobjekte unabhängig vom Wiedergabesystem und geben an, auf welcher Position sich ein Objekt befindet und ob es sich im Verlauf der Zeit verändert. Szenenbasiertes Audio bezieht sich auf die Erstellung einer richtungsabhängigen Kodierung einer Klangszene. Ebenfalls unabhängig von bestimmten Wiedergabekonfigurationen wird das Schallfeld in einem Raum mathematisch beschrieben, auf der Wiedergabeseite muss es bis zu einem gewissen Grad mithilfe von Matrizen rekonstruiert werden,

⁶⁵ Nipkow, „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“, S. 341.

⁶⁶ Edwin Pfanzagl-Cardone, „1.10 Distribution of Direct- and Diffuse-Sound Reproduction in Multichannel Loudspeaker Systems“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 43.

⁶⁷ Nipkow, „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“, S. 341.

⁶⁸ Hyunkook Lee, „Multichannel 3D Microphone Arrays: A Review“, *Journal of the Audio Engineering Society*, Applied Psychoacoustics Lab (APL), University of Huddersfield, Huddersfield, United Kingdom, 69, Nr. January/February 2022 (6. Januar 2021), S. 5, <https://doi.org/10.17743/jaes.2020.0069>.

denn Quelle und Richtung sind nicht direkt an die Lautsprecherposition gebunden. ⁶⁹

Im Kontext der physischen Konfiguration lassen sich 3D-Mikrofonanordnungen grob in drei verschiedene Kategorien einteilen: horizontal und vertikal beabstandet (HVS – horizontally and vertically spaced), horizontal beabstandet und vertikal koinzident (HSVC – horizontally spaced and vertically coincident) und horizontal und vertikal koinzident (HVC- horizontally and vertically coincident). Bei HVS-Arrays (horizontally and vertically spaced) haben alle Mikrofone einen bestimmten Abstand zueinander. In horizontaler und vertikaler Ebene entstehen so Zeitdifferenzen zwischen den Kanälen. Mit größerem horizontalem Abstand zwischen den einzelnen Mikrofonen einer 3D-Mikrofonanordnung wird der räumliche Eindruck bei der Wiedergabe verstärkt, da die Signale eine stärkere Dekorrelation aufweisen. ⁷⁰ Allerdings führten Lee und Gribben Hörtests durch, die zeigten, dass mehr Abstand zwischen den Haupt- und Höhenmikrofonen eines 3D-Mikrofonarrays nur geringen Einfluss auf den 3D-Raumeindruck beim Zuhörer hat. Zwischen den Abständen 0,5m, 1,0m und 1,5m gab es im wahrgenommenen Raumeindruck keine signifikanten Unterschiede. Ein Abstand von 0m erzeugte bei der Hälfte der Versuche einen erheblich größeren räumlichen Eindruck. ⁷¹ Auf Grundlage der Erkenntnis, dass vertikale Dekorrelation durch größere vertikale Mikrofonabstände kaum Auswirkungen auf den räumlichen Eindruck einer dreidimensionalen Klangwiedergabe hat, wurden Mikrofonarrays in Anlehnung an das HSVC-Konzept entwickelt. Horizontale Mikrofone haben einen gewissen Abstand zueinander, um den räumlichen Eindruck zu verstärken. Zwischen den Mikrofonen der mittleren und oberen Ebene gibt es keine oder nur geringe Abstände. Die Verwendung von Richtmikrofonen ist in diesem Konzept sinnvoll, um die Kanäle ausreichend voneinander zu trennen und möglichst dekorrelierte Signale aufzunehmen. ⁷² Richtmikrofone in der oberen Ebene müssen gezielt positioniert werden, entweder in größerem Abstand oder mit einer präzisen Ausrichtung nach

⁶⁹ Pfanzagl-Cardone, „2 - ‚3D‘ or ‚Immersive‘ Audio - The Basics and a Primer on Spatial Hearing“ in *The Art and Science of 3D Audio Recording*, S. 52.

⁷⁰ Lee, „Multichannel 3D Microphone Arrays: A Review“, S. 8.

⁷¹ Hyunkook Lee und Christopher Gribben, „Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array“, *Journal of the Audio Engineering Society* 62 (5. Januar 2015): 870–84, <https://doi.org/10.17743/jaes.2014.0045>.

⁷² Lee, „Multichannel 3D Microphone Arrays: A Review“, S. 8.

oben oder unten. Sie nehmen denselben Direktschall auf, wie die Hauptmikrofone in der mittleren Ebene. Um zu verhindern, dass sich das Quellenbild unerwünscht nach oben verschiebt, sollte der mit dem Höhenmikrofon aufgenommene Direktschall etwa um 7dB im Vergleich zum selben Signal aus der mittleren Ebene gedämpft werden. ⁷³

Mikrofonarrays, die der HVC-Kategorie zugeordnet werden, verfügen über keinen oder nur minimalen Abstand zwischen allen Mikrofonen. Für die Richtungsabbildung der Klangquellen sind Pegeldifferenzen zwischen den Kanälen entscheidend. ⁷⁴ Die meisten HVS- und HSVC-Mikrofonarrays lassen sich als wahrnehmungsorientierte Mikrofonanordnungen kategorisieren. Das Ziel ist hierbei nicht, das aufgenommene Klangerlebnis exakt zu reproduzieren, sondern dem Hörer stattdessen ein klanglich und räumlich glaubhaftes und angenehmes Klangerlebnis zu bieten. Mithilfe von Manipulation der Beziehungen zwischen den einzelnen Mikrofonsignalen erreicht man bestimmte Eigenschaften in Zusammenhang mit der Lokalisierung von Phantomschallquellen und dem Raumeindruck. Jedes einzelne Mikrofon einer wahrnehmungsorientierten 3D-Mikrofonanordnung wird von jeweils einem zugewiesenen Lautsprecher wiedergegeben. Die Arrays wurden gezielt für eine Lautsprecherkonfiguration konzipiert, die Anzahl der Mikrofone stimmt mit der Anzahl der Wiedergabekanäle im Zielformat überein, beispielsweise verwendet man neun Mikrofone für eine 5.1.4 Konfiguration. Viele wahrnehmungsorientierte Mikrofonanordnungen ergänzen bestehende Surround-Mikrofonarrays um die Höhenkanäle. ⁷⁵

Im Folgenden werden 3D-Mikrofonanordnungen vorgestellt, die mit empfohlenen Lautsprecheraufstellungen der ITU-R BS.2051-3 kompatibel sind. ⁷⁶ Die binaurale Kunstkopfaufnahme (siehe Kapitel „2.3 Binaurales Audio und binaurale Aufnahmetechnik“) eignet sich ebenfalls für eine immersive 3D-Audio Aufnahme, lässt sich allerdings nicht in die kanalbasierte 3D-Aufnahmetechnik einordnen.

⁷³ Edwin Pfanzagl-Cardone, „1.8 A Few Thoughts on Microphone Pattern Choice and Capsule Orientation in 3D Audio“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 36.

⁷⁴ Lee, „Multichannel 3D Microphone Arrays: A Review“, S. 8.

⁷⁵ Lee, S. 8.

⁷⁶ International Telecommunication Union, „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production‘“.

2.7.1 2L-Cube

Morten Lindberg entwickelte den sogenannten 2L-Cube, der sich in die Kategorie der HVS-Mikrofonarrays einordnen lässt. In einer Würfelanordnung werden neun Mikrofone mit Kugelcharakteristik für die Wiedergabe in einem 5.1.4 System angebracht. Die Wahl der Richtcharakteristik steht hier nicht im Vordergrund, sondern die Klangqualität der Mikrofone mit Kugelcharakteristik. Sie haben meist im Vergleich zu unidirektionalen oder bidirektionalen Mikrofonen einen erweiterten Tiefenfrequenzgang. Die Breite und Tiefe der Würfelanordnung kann je nach Größe des aufzunehmenden Schallkörpers zwischen 0,4m und 1,2m variieren. Die Höhe des Würfels und somit der Abstand der Höhenmikrofone zur mittleren Ebene bleibt konstant bei 1m. Das Mikrofon, das für den Centerkanal vorgesehen ist, wird dabei im Gegensatz zum linken und rechten Mikrofon etwas weiter vorne platziert, ein genauer Abstand ist allerdings nicht definiert.^{77 78}



Abbildung 8: Morten Lindberg's 2L-Cube Mikrofonanordnung

Quelle: Lindberg Lyd, 2022

Wenn das Wiedergabeformat weniger als 9 Lautsprecher hat, werden die Mikrofonensignale nicht zusammengefasst, sondern schlichtweg die Mikrofone weggelassen, die keinen direkt zugeordneten Lautsprecher haben. Beispielsweise sind bei 5.1 als angestrebtem Wiedergabeformat nur die fünf unteren Mikrofone

⁷⁷ Lee, „Multichannel 3D Microphone Arrays: A Review“, S. 10.

⁷⁸ Edwin Pfanzagl-Cardone, „9.11 The Morton Lindberg ‚2L‘ Technique“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 316.

in 2L-Cube aktiv, für Stereo nur vorne links und rechts. Es geht vor allem darum, die Arrays so einzurichten, dass die Ankunftszeit des Schalls in der Reihenfolge natürlich aufgenommen und wiedergegeben wird.⁷⁹

2.7.2 OCT-3D

Die OCT-3D Mikrofonanordnung ist Teil der HVS-Kategorie. Grundlage des OCT-3D Mikrofonarrays ist die OCT-Surround Anordnung (optimized cardioid triangle surround) mit fünf Kanälen, die zunächst auf Klangaufnahmen für die Wiedergabe in 5.1 Systemen ausgerichtet war. Durch die Ergänzung von vier nach oben gerichteten Supernieren in 1m Höhe über dem L, R, Ls und Rs Kanal wird diese Anordnung dreidimensional.⁸⁰

Die vorderen drei Mikrofone dieser Anordnung, also L, C und R, sollen eine klare und stabile Phantombilddarstellung erreichen, indem der Anteil an Übersprechen zwischen den Kanälen (interchannel-crosstalk, ICXT) auf ein Minimum reduziert wird. Mikrofonsignale, die nicht vorrangig für die Darstellung des Phantombildes verantwortlich sind, werden als unerwünschter Übersprechanteil betrachtet. Um dem Übersprechanteil im vorderen Dreieck entgegenzuwirken, setzt man vorne Mikrofone mit Supernierencharakteristik ein, die seitlich ausgerichtet werden. Allerdings kann der Übersprechanteil zwischen den Kanälen sich auch positiv auf die Wahrnehmung auswirken, da der Zuhörer möglicherweise eine größere Breite des Hörereignisses wahrnimmt. In der oberen Ebene verwendet man ebenfalls Mikrofone mit Supernierencharakteristik, um den Übersprechanteil zu minimieren.

Ein vertikales Panning zwischen der mittleren und oberen Ebene zur Erzeugung einer Phantomschallquelle ist nicht stabil, da der Zuhörer mithilfe von interauralen Laufzeit- oder Pegeldifferenzen keine eindeutige Richtung identifizieren kann. Stattdessen soll der Raum zwischen den mittleren und oberen Lautsprechern bei der Wiedergabe mit Reflexionen und Nachhall ausgefüllt werden. Die Supernieren in der oberen Ebene können nach oben ausgerichtet werden und nehmen auf diese Weise nur reduziert Direktschall auf,

⁷⁹ Edwin Pfanzagl-Cardone, „9.11.1 Use of Center Speaker and LFE-Channel“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 318.

⁸⁰ Edwin Pfanzagl-Cardone, „9.12 The OCT-3D Technique“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 319.

das ermöglicht eine Abbildung hauptsächlich diffuser Klangbilder in der vertikalen Ebene.⁸¹ Zusammenfassend nehmen die Frontmikrofone L, C und R insbesondere direkten Schall auf, die hinteren und oberen Mikrofone sind so positioniert, dass sie primär Umgebungsatmosphäre und Diffusschall einfangen.

82

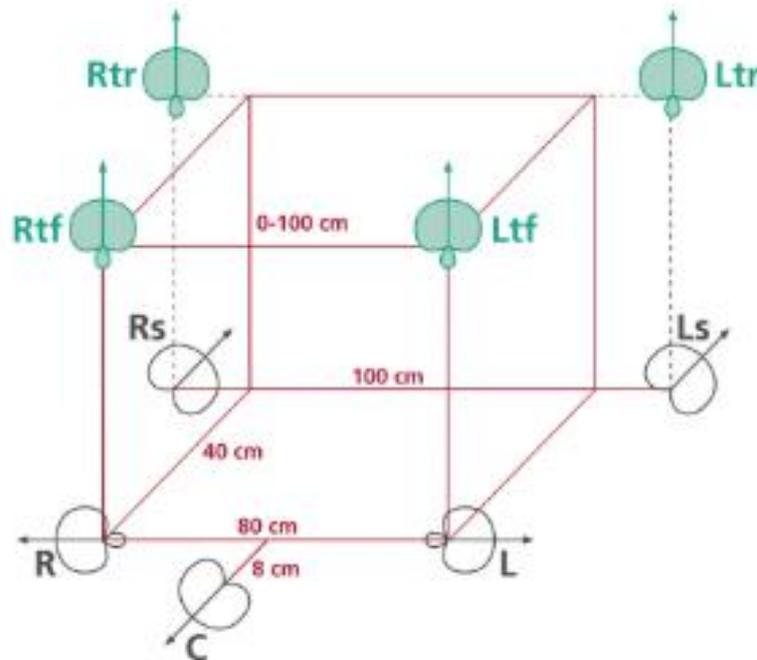


Abbildung 9: OCT-3D Mikrofonanordnung

Quelle: Schoeps, 2025

2.7.3 ORTF-3D

Das ORTF-3D Mikrofonarray wurde von Helmut Wittek und Günter Theile entwickelt und lässt sich der HSVC-Kategorie zuordnen. In der unteren und oberen Ebene sind jeweils vier Supernieren in einem Rechteck mit einer Seitenlänge von 10x20 cm übereinander positioniert.⁸³ Die acht Kanäle, L, R, Ls und Rs in der unteren Ebene sowie Ltf, Rtf, Ltr und Rtr in der oberen Ebene werden ohne Matrixbildung direkt auf einen diskreten Wiedergabekanal geleitet. In Umgebungsaufnahmen ist ein Centerkanal selten bespielt und erwünscht,

⁸¹ Lee, „Multichannel 3D Microphone Arrays: A Review“, S. 9.

⁸² Pfanzagl-Cardone, „9.12 The OCT-3D Technique“, S. 319.

⁸³ Edwin Pfanzagl-Cardone, „9.13 The ORTF-3D Technique“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 319.

deshalb wird er hier ebenfalls außer Acht gelassen, da er das Gleichgewicht zwischen vorne und hinten stören würde.⁸⁴



Abbildung 10: ORTF-3D Mikrofonanordnung

Quelle: Wittek und Theile, 2017

Die Grundlage für diese dreidimensionale Aufnahmetechnik bildet die Verdoppelung und Modifizierung des IRT-Kreuzes in zwei Ebenen. Eine maximale Signaltrennung zwischen der oberen und unteren Ebene erreicht man dadurch, dass die Mikrofonkapseln um jeweils 45° nach oben oder unten ausgerichtet sind. Zusammen ergibt diese Neigung eine 90° Anordnung, die in Kombination mit der Nutzung von Supernieren eine sehr hohe Dekorrelation der Signale in der vertikalen Ebene hervorruft. Die Supernierencharakteristik führt außerdem dazu, dass einfallender Direktschall ausreichend unterdrückt wird und so der ambiente Diffusschall realitätsnah abgebildet werden kann.

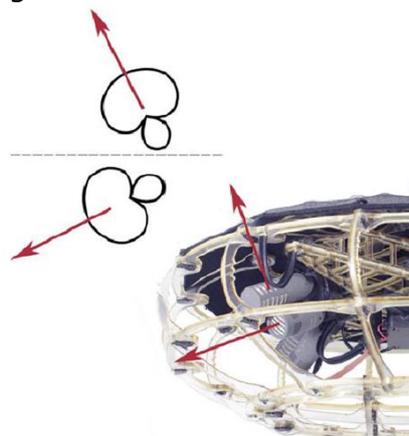


Abbildung 11: Detail der Kapselausrichtung eines ORTF-3D Mikrofonarrays

Quelle: Pfanzagl-Cardone, 2023, S. 322

⁸⁴ Helmut Wittek und Günther Theile, „Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR“, *Journal of The Audio Engineering Society*, 2017, S. 8, <https://api.semanticscholar.org/CorpusID:116160580>.

Die Vorteile des ORTF-3D Mikrofonarrays ergeben sich in der hohen Signaltrennung mit ausreichend hoher Dekorrelation sowie einer optimalen immersiven 3D Abbildung mit hoher Stabilität bei der Wiedergabe.⁸⁵ Zudem ist es im Vergleich zu den vorgestellten Alternativen kompakt und ermöglicht dabei dennoch räumliche Offenheit.⁸⁶

3. Dolby Atmos

Im April 2012 stellten die Dolby Laboratories ein neues, immersives 3D-Audioformat vor. Dabei ist zum ersten Mal die Anzahl der Lautsprecher mit bis zu 64 getrennten Wiedergabekanälen variabel. Es können bis zu 128 gleichzeitige und verlustfreie Audioeingänge übertragen werden,⁸⁷ die sich aus einem zehnkanaligen Bed (9.1) und bis zu 118 Audioobjekten zusammensetzen können.⁸⁸ Dieses Audioformat verfügt neben den Lautsprechern auf der Horizontalebene über weitere Kanäle, die in der Decke angebracht werden und den Zuhörer somit auch von oben umhüllen können.⁸⁹ Klänge, die von oben kommen, sorgen für eine realistischere Wahrnehmung, in der echten Welt erreichen den Zuhörer die Klänge ebenfalls aus verschiedenen Richtungen, die sich nicht nur auf die horizontale Ebene beschränken.⁹⁰

In vergangenen Dolby-Surround Systemen wurde mehreren Lautsprechern ein identisches Signal zugeordnet, dabei entsprechen die Kanäle jeweils einer spezifischen vorbestimmten Lautsprecherposition.⁹¹ Ein Atmos-System ermöglicht es, jedem einzelnen Lautsprecher in der jeweiligen Konfiguration ein individuelles dynamisches oder statisches Signal zu geben.

⁸⁵ Nipkow, „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“, S. 339.

⁸⁶ Pfanzagl-Cardone, „9.13 The ORTF-3D Technique“, S. 321.

⁸⁷ Dolby Laboratories, Inc., „Dolby Atmos Is the Future of Entertainment Sound“, *Business Wire (English)*, 23. April 2012, <https://www.businesswire.com/news/home/20120423005868/en/Dolby-Atmos-Is-the-Future-of-Entertainment-Sound>.

⁸⁸ Stefan Meltzer, „Kapitel 13.4.7.3 - Dolby Atmos“, in *Handbuch der Tonstudiotchnik*, hg. von Michael Dickreiter, 9. aktualisierte und erweiterte Auflage, Bd. 1 u. 2 (Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023), S. 895.

⁸⁹ Dolby Professional, „Dolby Atmos Documentation“, o. J., <https://professional.dolby.com/gaming/gaming-getting-started/dolby-atmos-documentation/#fmod>.

⁹⁰ Edwin Pfanzagl-Cardone, „4.2 Dolby Atmos - An Overview“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 146.

⁹¹ Dolby Professional, „Dolby Atmos Documentation“.

Klangereignisse werden nicht mehr ausschließlich kanalbasiert,⁹² sondern auch objektbasiert, also nach Koordinaten gemischt.⁹³ Statische Objekte haben eine fest zugeordnete Position, während dynamische Objekte sich über die Zeit im Raum bewegen und somit ihre Koordinaten verändern. Der Toningenieur platziert die jeweiligen Objekte in einem virtuellen dreidimensionalen Raum. Jedes Objekt enthält neben den Audioinformationen, die mit den Metadaten übertragen werden, auch die Panoramainformationen zur aktuellen Position im 3D Raum und wie sie sich über die Zeit verändert.⁹⁴ Die Koordinaten des Audio-Objektes werden mit jedem Frame aktualisiert.

Die Hintergrundgeräusche und -musik werden als „Bed“ bezeichnet und bilden die Basis, sie werden weiterhin im bekannten kanalbasierten Verfahren abgemischt⁹⁵, da sie keine konkrete räumliche Zuordnung besitzen und somit aus mehreren Lautsprechern gleichzeitig in gleichem Ausmaß wiedergegeben werden können.⁹⁶ Da für das Audioformat Dolby Atmos sowohl statische oder dynamische Audioobjekte als auch kanalbasierte „Beds“ verwendet werden, handelt es sich nicht ausschließlich um ein objektbasiertes Verfahren, sondern um ein hybrides Audioformat.⁹⁷

Beim Endnutzer werden die Metadaten im Dolby Atmos-fähigen AV-Receiver dekodiert. Anhand der Position im dreidimensionalen Raum entscheidet das Gerät, welcher Lautsprecher aus dem Setup sich für die Abbildung des jeweiligen Audio-Objektes am besten eignet. Durch diese individuelle Signalverteilung für die eigene Lautsprecherkonfiguration wird eine authentische Abbildung der Klangszene ermöglicht. Um Dolby Atmos im Heimkino-Setup zu erleben, benötigt man lediglich einen AV-Receiver, der Atmos unterstützt sowie ein geeignetes und leistungsfähiges Lautsprecher-System.⁹⁸

⁹² Detlev Schnick und Philipp Kind, „Fakten rund um Dolby Atmos“, 21. November 2024, https://www.hifi-regler.de/wissenswertes_und_kaufberatung/kino-tonformate/dolby_atmos.php?srsltid=AfmBOoqz_KZeafLy5NjJ1r-OUqsaSXdAkoxURXL-4dg4ly5FRvXNeLu.

⁹³ Florian C. Scholz, „10.4.1 - Mordere digitale Surroundformate“, in *Audiotechnik für Mediengestalter* (De Gruyter Saur, 2015), S. 236.

⁹⁴ Schnick und Kind, „Fakten rund um Dolby Atmos“.

⁹⁵ Dolby Professional, „Dolby Atmos Documentation“.

⁹⁶ Meltzer, „Kapitel 13.4.7.3 - Dolby Atmos“, S. 894.

⁹⁷ Dolby Laboratories Inc., „Dolby Atmos Is the Future of Entertainment Sound“.

⁹⁸ Schnick und Kind, „Fakten rund um Dolby Atmos“.

Die Rückwärtskompatibilität des Atmos-Systems bietet einen großen Vorteil, auch die Integration in bestehende Produktionsabläufe werden minimiert.⁹⁹ Zudem ermöglicht das 3D-Audioformat eine verbesserte Raumentzerrung in Kombination mit präziserem Bassmanagement und Klangfarbenabstimmung.¹⁰⁰

Zunächst sollte Dolby Atmos insbesondere das Kinoerlebnis klanglich erweitern. Seit 2014 wurde es mit geeigneten AV-Receivern auch für den Heimbereich zugänglich gemacht. Im Jahr 2017 war das Audioformat erstmals im Video-on-Demand-Streaming bei Netflix verfügbar, weitere Plattformen folgten. 2019 wurde Dolby Atmos zuerst bei Amazon Music HD eingeführt und in der folgenden Zeit bei weiteren Musikanbietern integriert.¹⁰¹ Im Gaming-Bereich hat das immersive 3D-Audioformat in den letzten Jahren ebenfalls an Bedeutung gewonnen, da dem Spieler hier durch die Audioobjekte ein präzises Eintauchen ermöglicht wird.¹⁰²

3.1 Anwendungsfälle

Zu Beginn wurde Dolby Atmos für den Kinosaal herausgebracht, mit der Zeit erweiterten sich die Anwendungsbereiche. Die Einführung von Audioobjekten und Deckenlautsprechern im Kino macht die Wiedergabe anpassungsfähig an die individuellen räumlichen Bedingungen. Gleichzeitig mit den Audioobjekten wird ein weiterhin kanalbasiertes Bed für die diffuse Geräuschkulisse eingesetzt.¹⁰³ Der Einsatz der Dolby Atmos Technologie im Heimbereich unterscheidet sich jedoch vom Kinobereich, es bezeichnet die Verwendung des Audiocodierverfahren Dolby E-AC-3 mit einer Joint Object Coding (JOC) Erweiterung. Dieses parametrische Verfahren vereint sowohl die Audioobjekte als auch die Höhenkanäle in einer E-AC-3 codierten, kanalbasierten 5.1 oder 7.1 Mischung, es ist demnach auch eine Wiedergabe des Signals von einem E-AC-3 Decoder möglich, wenn keine JOC Erweiterung vorhanden ist. Ein Decoder mit JOC-Erweiterung kann durch die JOC-Daten Audioobjekte und Höhenkanäle aus der

⁹⁹ Pfanzagl-Cardone, „4.2 Dolby Atmos - An Overview“, S. 145.

¹⁰⁰ Pfanzagl-Cardone, „4.3 Multichannel Speaker-Layout: Improved Audio Quality and Timbre Matching“, S. 147.

¹⁰¹ Vrtonung und Daniela Rieger, „Dolby Atmos Music - Was ist dieses 3D Sound Erlebnis genau?“, o. J., <https://www.vrtonung.de/dolby-atmos-music/>.

¹⁰² Dolby Professional, „Dolby Atmos Documentation“.

¹⁰³ Meltzer, „Kapitel 13.4.7.3 - Dolby Atmos“, S. 894-895.

kanalbasierten Mischung extrahieren. Aus den bis zu 16 PCM-Signalen einschließlich Metadaten kann der Renderer am Ausgang des Decoders je nach verwendeter Konfiguration des Wiedergabesystems Signale für die einzelnen Lautsprecher erzeugen. Eine zu niedrige Datenrate für das E-AC-3 codierte Signal oder die JOC-Daten kann unbeabsichtigt zu Artefakten bei der Extrahierung führen, Dolby Atmos verwendet aus diesem Grund Datenraten bis zu 768 kbps, die erreichbare Audioqualität ist begrenzt. ¹⁰⁴

Im Gaming-Bereich kann Dolby Atmos in Spielen für die gesamte Xbox-Familie sowie die PlayStation 5 und auch auf dem PC und Mobilgeräten genutzt werden. Häufig spielen Gamer mit Kopfhörern, bei deren Verwendung besitzt jedes Audio-Objekt eine eigene HRTF (head related transfer function). Bei der Wiedergabe über HDMI am PC oder auf der Xbox One sind 20 Atmos-Objekte verfügbar, Dolby Atmos for Headphones ermöglicht die Wiedergabe von 16 Atmos-Objekten. Zusätzliche Audio-Objekte, die die verfügbare Anzahl übersteigen, werden in den Bed-Kanälen wiedergegeben. Bei der Wiedergabe über Kopfhörer bewegen sie sich weiterhin im 3D-Raum, besitzen allerdings keine individuell zugehörige HRTF-Verarbeitung. ¹⁰⁵

Mobiltelefone bilden einen weiteren Anwendungsbereich, bei dem Dolby AC-4 in Kombination mit der Immersive Stereo Erweiterung (IMS) verwendet wird. Ein Stereosignal ist in diesem Zusammenhang im Binaural Renderer Mode encodiert und enthält die zugehörigen Metadaten. Abhängig davon, ob die Wiedergabe über Kopfhörer oder über den integrierten Lautsprecher des Geräts erfolgt, generiert der Decoder im Mobiltelefon entweder ein binaurales Stereosignal oder ein virtuelles immersives Signal. ¹⁰⁶

3.2 Technische Richtlinien

Die Dolby Laboratories haben technische Richtlinien für ein Dolby Atmos Home Entertainment Studio entwickelt, in dem sich Dolby Atmos Inhalte bei der Produktion ideal abhören lassen. Diese Richtlinien schließen unter anderem

¹⁰⁴ Meltzer, S. 895.

¹⁰⁵ Dolby Professional, „Dolby Atmos Documentation“.

¹⁰⁶ Meltzer, „Kapitel 13.4.7.3 - Dolby Atmos“, S. 895.

Raumgeometrie, Raumakustik und die Positionierung der Lautsprecher ein.¹⁰⁷ Die Abmessungen für die Lautsprecheranordnung stehen allerdings im Vordergrund, nicht die Raummaße. Die von Dolby bevorzugte und empfohlene Anordnung für die Lautsprecher ist 7.1.4, eine 5.1.4 Konfiguration wird allerdings ebenfalls akzeptiert, mit dem Hinweis, dass die fehlenden hinteren Surround-Lautsprecher zu einer unvollständigen räumlichen Abbildung führen.

Es gibt drei verschiedene Parameter, die bei einer Lautsprecheranordnung relevant sind: Höhe, Breite und Länge des Raumes oder alternativ der Lautsprecherkonfiguration. Die Höhe bezieht sich auf den Abstand zwischen Boden und der Unterkante des höchsten Surround-Lautsprechers und sollte mindestens 2,4m betragen. Die seitlichen Lautsprecherpaare bestimmen die Breite der Anordnung und sollten mindestens 3m voneinander entfernt sein. Die Länge der Anordnung ist abhängig vom Abstand zwischen den Front- und hinteren Lautsprechern und beträgt idealerweise mehr als 3,5m. In einer Kreisordnung erfolgt die Orientierung für die Länge und Breite am Durchmesser des Kreises. Zwischen den Lautsprechern und der Mischposition liegen im Optimalfall 5m oder weniger. Es wird zudem ein Raumvolumen von mindestens 50m³ empfohlen.¹⁰⁸

Dimension	Specification
Minimum layout height	2.4 m
Minimum layout width	3 m
Minimum layout length	3.5 m
Recommended room volume	>50 m ³
Speaker distance to mix position	≤5 m (<4 m recommended) ^[a]

T

Tabelle 4: Empfohlene Mindest- und Maximalabmessungen des Layouts

Quelle: Dolby Laboratories, Inc., 2021, S. 9

¹⁰⁷ Dolby Laboratories, Inc., „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, S. 8.

¹⁰⁸ Dolby Laboratories, Inc., S. 9.

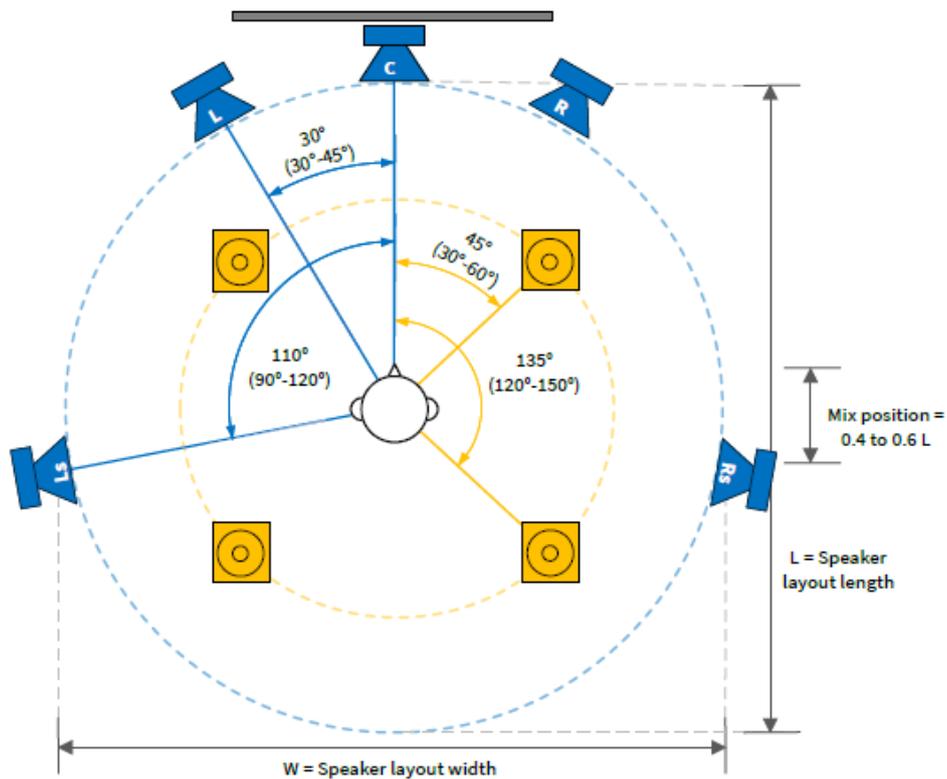


Abbildung 12: Äquidistantes Lautsprecherlayout 5.1.4

Quelle: Dolby Laboratories, Inc., 2021, S. 23

Die Position, von der gemischt wird, dient als Maßstab für die Aufstellung der Lautsprecher, man rechnet dabei mit 1,2m Höhe in direkter Linie mit dem Center-Lautsprecher. Wenn möglich, sollten sich alle Lautsprecher in gleichem Abstand und optimal ausgerichtet zur Mischposition befinden. Unterscheiden sich die Lautsprecherdistanzen zur Mischposition, kann man das mit Verzögerungen von bis zu 10ms ausgleichen oder die Amplitude anpassen. In der horizontalen Ebene werden die fünf bzw. sieben üblichen Lautsprecher L, C, R, Ls, Rs, (Lrs, Rrs) aufgestellt. Die obere Ebene wird mit Ltf, Rtf, Ltr und Rtr definiert.¹⁰⁹ Lautsprecher können entweder alle gleich weit von der Mischposition entfernt sein oder in einer rechteckigen Anordnung, bei der eine Seite länger ist.¹¹⁰ Die Lautsprecher der oberen Ebene sollten im Optimalfall einen Winkel von 45° zur Horizontalen haben. Sofern die Lautsprecher L, C, R, Ls und Rs den Hörer

¹⁰⁹ Dolby Laboratories, Inc., S. 10.

¹¹⁰ Dolby Laboratories, Inc., S. 11.

unter leicht erhöhtem Winkel erreichen, werden die Deckenlautsprecher ebenfalls zusätzlich geneigt. ¹¹¹

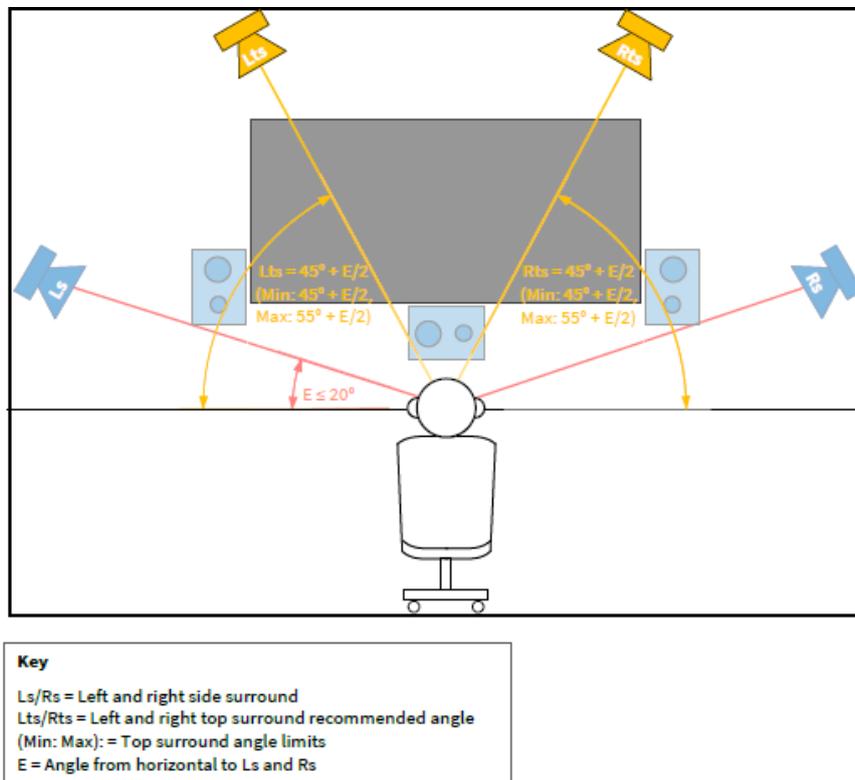


Abbildung 13: Positionen der Surround- und Deckenlautsprecher

Quelle: Dolby Laboratories, Inc., 2021, S. 20

Beim Einsatz eines LFE-Kanals wird dieser nicht in gleicher Linie mit dem Center-Lautsprecher auf dem Boden platziert, sondern seitlich. ¹¹²

Der Dolby Atmos Renderer kann sowohl mit einer Samplerate von 48kHz als auch 96kHz arbeiten. Arbeitet der Renderer mit 48kHz, stehen 128 Eingangskanäle zur Verfügung, 118 davon können Audioobjekte sein, die übrigen 10 bilden das Bed. Für die Kodierung werden die Dateien einheitlich in 48kHz konvertiert. ¹¹³

Alle eingesetzten Lautsprecher müssen den vollen Dynamikumfang von (Kino-) Inhalten wiedergeben können und sollten nach Möglichkeit vom selben Hersteller und Modell sein. Im Frontbereich bzw. im Leinwandbereich sollten mindestens

¹¹¹ Dolby Laboratories, Inc., S. 19.

¹¹² Dolby Laboratories, Inc., S. 15.

¹¹³ „What Sample Rates Does Dolby Atmos Support?“, FAQ in Knowledge, *Dolby Atmos FAQs - Mixing in Atmos*, 4. Januar 2021, https://professionalsupport.dolby.com/s/dolby-atmos-faqs?language=en_US&tabset-c803d=3.

drei Lautsprecher platziert sein, bei einer Kinoleinwand breiter als 12m kommen weitere Lautsprecher hinzu. ¹¹⁴

4. 3D-Audio im Fernsbereich und objektbasiertes Audio (Next Generation Audio)

Das Prinzip des Rundfunks besteht darin, Geschichten, Ideen, Informationen oder Unterhaltung zu vermitteln. Ein Sender kann dabei über Radio oder Fernsehen an Millionen Menschen gleichzeitig dieselben Informationen übertragen, weshalb er eine besondere Bedeutung für die Gesellschaft hat. Die Art, wie Menschen auf audiovisuelle Medien und Informationen zugreifen, hat sich in den letzten Jahren insbesondere durch das Internet in eine andere Richtung entwickelt.

Fortschrittliche Codierungstechnologien ermöglichen den Konsum von Inhalten über das Internet oder mobile Datennetze. Der Nutzer kann zuhause auf einem Fernseher mit Internetzugang Inhalte konsumieren, aber auch unabhängig davon mit dem mobilen Endgerät jederzeit unterwegs audiovisuelle Medieninhalte aufrufen. Durch die veränderte Rolle der Medien im Leben der Menschen sollten auch Rundfunkanstalten andere, erweiterte Dienste anbieten.

3D-Audio beabsichtigt, dem Zuhörer einen realistischeren und immersiveren Höreindruck zu bieten. Für die Rundfunkanstalten könnte das eine Gelegenheit sein, fesselndere, angenehmer und einnehmende Hörerlebnisse zu schaffen und auch neue Möglichkeiten in der kreativen Programmgestaltung wahrzunehmen.

¹¹⁵

Die British Broadcast Corporation Research and Development (BBC R&D) veröffentlichte im Jahr 1998 einen Bericht über Produktionsexperimente mit Fünfkanal-Surround-Sound. ¹¹⁶ Nachdem dieses Format für Heimkinoanwendungen bereits eingeführt war, wurde es im Jahr 2006 auch zum ersten Mal im kostenlosen Rundfunk des Vereinigten Königreichs etabliert. Bezüglich angebotener Audioformate veränderten sich die Rundfunkanstalten seitdem kaum. In anderen Bereichen der Unterhaltung, wie Kino und Gaming, wurde 3D-Audio ein immer größerer Aspekt und etablierte sich zunehmend, mit

¹¹⁴ „Dolby Atmos Specifications“ (Dolby Laboratories, Inc., April 2024), S. 1.

¹¹⁵ Pike, „Chapter 1 - 3D Audio in Broadcasting“, S. 1.

¹¹⁶ D. G. Kirby, N. A. F. Cutmore, und J. A. Fletcher, „Programme Origination of 5-channel surround sound“ (Tadworth, UK: British Broadcast Corporation Research and Development, 1. April 1998), <https://downloads.bbc.co.uk/rd/pubs/reports/1998-17.pdf>.

diesen Entwicklungen und Erwartungen kann der Rundfunk aktuell noch nicht mithalten. Ultra-High Definition (UHD) ist eine aktuelle Entwicklung im Fernsehen und strebt eine höhere Bildqualität an, 3D Audio wird als das Äquivalent zu UHD-Video gesehen.¹¹⁷

Binaurales Audio ist eine Möglichkeit, 3D-Audioszenen mit nur zwei Kanälen darzustellen, der Zuhörer benötigt nur Kopfhörer und kein zusätzliches spezielles Equipment, um immersives Audio zu erleben.¹¹⁸ Einige Programminhalte im Rundfunk sind bereits in binauralem Audio verfügbar, die Produktion ist allerdings etwas komplexer als herkömmliches Stereo.¹¹⁹

Für die Produktion von 3D-Audio für Lautsprecher sind weitere technologische Veränderungen in der Produktions- und Verteilungskette des Rundfunks notwendig, weshalb diese Technik bisher noch nicht weit verbreitet ist.¹²⁰ Produktionsstätten und Studios benötigen eine entsprechende technische Ausrüstung, wie beispielsweise 3D-Lautsprecheranordnungen für das Monitoring sowie Mischpulte, die Multikanal-Busse und 3D-Panning unterstützen. Außerdem können bisherige Produktionswerkzeuge aus dem Stereo- und Surroundbereich nicht problemlos bei einer 3D-Audio Mischung angewendet werden, denn fähige Multikanal-Dynamik- und Hallprozessoren sind erst in der Entstehung. Um die erforderlichen Kenntnisse für die 3D-Audioproduktion zu erlangen, muss das Personal geschult werden. Auch auf Endnutzerseite gibt es Herausforderungen für die Lautsprecherwiedergabe, da neue Receiver und Multikanal-Lautsprechersysteme benötigt werden, deren Installation komplex ist. Je größer die Anzahl der Lautsprecher im System wird, desto komplizierter ist die Einrichtung für den Nutzer. Codecs, die die Übertragung von Multikanal-3D-Audio unterstützen, müssen auch in den Geräten der Endnutzer verfügbar sein. Viele neue Fernseh- und Mobilgeräte unterstützen mindestens einen dieser Codecs, durch die zunehmende Verfügbarkeit wird die Übertragung in naher Zukunft umsetzbar. Die Produktion von binauralem Audio und 3D-Audio erhöht jeweils den Zeitaufwand und die Kosten, da mindestens zwei separate Audiomischungen zur Verfügung gestellt werden müssen.¹²¹

¹¹⁷ Pike, „Chapter 1 - 3D Audio in Broadcasting“, S. 1.

¹¹⁸ Pike, S. 2.

¹¹⁹ Pike, S. 9-10.

¹²⁰ Pike, S. 10.

¹²¹ Pike, S. 12.

Im Bereich des Live-Broadcastings kommt zunehmend die objektbasierte Audiotechnologie zum Einsatz, die nicht ausschließlich nur für die Übertragung von Multikanal-3D-Audio genutzt wird, sondern auch für die Personalisierbarkeit. Der Endnutzer erhält die Möglichkeit, zwischen verschiedenen Audiostreams zu entscheiden.¹²² Um beispielsweise die Dialogverständlichkeit zu steigern, kann das Verhältnis zwischen Audioobjekten individuell angepasst werden. Innerhalb der Mischung können Objekte ergänzt werden, zum Beispiel eine Audiobeschreibung für Menschen mit Sehbeeinträchtigung.¹²³ Alternativ kann der Endnutzer zwischen verschiedenen Sprachen oder Perspektiven bei Sportkommentaren auswählen.¹²⁴

Das Ziel von objektbasierten Audiosystemen (OBA) ist es, nur eine Mischung zu erstellen, die dann auf dem Empfangsgerät für das eingesetzte Wiedergabesystem optimiert und angepasst wird. Das hebt die vermehrte Produktionsarbeit auf und ist unabhängig davon, ob die Wiedergabe dreidimensional über Kopfhörer oder Lautsprecherkonfigurationen jeglicher Art erfolgt, auch Stereo- und Surroundformate sind hier eingeschlossen. Die Elemente der Audiomischung werden getrennt voneinander geliefert, die zugehörigen Metadaten beschreiben dabei, wie und aus welcher Position die Wiedergabe der einzelnen Objekte erfolgen soll. Am Empfangsgerät entscheidet der Renderer, welche Kombination und Konfiguration der Audiosignale eine optimale Wiedergabe erzeugt. Der Begriff „Next-generation Audio“ (NGA) beschreibt in der Rundfunkindustrie Codec-Systeme, die objektbasierte Audiosysteme unterstützen, bisher gibt es drei standardisierte NGA-Codex: MPEG-H 3D Audio¹²⁵, Dolby AC-4¹²⁶ und DTS-UHD Audio¹²⁷.

Beispielsweise übertrug Sky im Jahr 2017 die Fußballspiele der englischen Premier League und verwendete dafür den Dolby Digital Plus Codec mit einer Joint Object Coding Erweiterung, die Spiele wurden also mit Dolby Atmos

¹²² Edwin Pfanzagl-Cardone, „4.7.4 The Future of Object-Based Audio“, in *The Art and Science of 3D Audio Recording* (Switzerland: Springer, 2023), S. 163.

¹²³ Pike, „Chapter 1 - 3D Audio in Broadcasting“, S. 13.

¹²⁴ Pfanzagl-Cardone, „4.7.4 The Future of Object-Based Audio“, S. 163.

¹²⁵ „ISO/IEC 23008-3:2022 Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D Audio“, August 2022.

¹²⁶ „ETSI TS 103 190-2 V.1.1.2.1 Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio“, Februar 2018.

¹²⁷ „ETSI TS 103 491 V.1.1.1 DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields“, April 2017.

übertragen. ¹²⁸

Der nationale japanische Rundfunkanbieter NHK startete im Dezember 2018 als Erster sein öffentliches 8K UHD-TV Format Super HiVision (SHV) mit dem 22.2 Multikanal 3D Audiosystem, bei dem sich neun Lautsprecher in der oberen Ebene befinden, 10 in der mittleren und 3 in der unteren Ebene (9+10+3) ¹²⁹ Dieses System wurde 2005 vorgestellt und erzeugt ein besseres Präsenzgefühl und Realismus beim Zuhörer auch außerhalb der zentralen Hörposition. ¹³⁰

France Télévisions testete im Jahr 2019 bei den Roland Garros Tennisturnieren eine UHD-Übertragung mit MPEG-H Audio. ¹³¹

Der Olympic Broadcasting Service (OBS) produziert seine Inhalte aktuell im neunkanaligen System 5.1.4 mit vier erhöhten Lautsprechern, dieses System wird in der ITU-R BS. 2051-3 als System D (4+5+0) standardisiert. ^{132 133}

4.1 Vorstellung der Live-TV Show „The Masked Singer“

„The Masked Singer“ ist eine Live-Unterhaltungsshow, die saisonal jeweils im Frühjahr und im Winter auf ProSieben ausgestrahlt wird. Alternativ kann man diese Show auch auf Joyn streamen. In der Regel wird The Masked Singer über einen Zeitraum von sechs Wochen jeden Samstag um 20:15 live auf ProSieben übertragen. Die erste Staffel wurde im Jahr 2019 im Sommer produziert. ¹³⁴ Die elfte Staffel ist die aktuelle und hatte ihren Auftakt am 23.11.2024.

Ausnahmsweise wurde die zweite Folge von sechs direkt am nächsten Tag, Sonntag, dem 24.11.2024 ausgestrahlt. Am 21.12.2024 endete die Staffel mit dem Finale. Eine Folge hat im Durchschnitt eine Dauer von 120-160 Minuten, bei der die Werbung noch hinzugerechnet werden muss. ¹³⁵ Die Produktion erfolgt in den MMC-Studios in Köln.

¹²⁸ Fergal Ringrose, „Next Generation Audio Summit: Inside Sky’s journey through Dolby Atmos production“ (Europe: SVG Europe, 6. November 2018).

¹²⁹ Kimio Hamasaki u. a., „A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)“, *SMPTÉ Motion Imaging Journal* 117 (1. April 2008): 40–49, <https://doi.org/10.5594/J151119>.

¹³⁰ Pike, „Chapter 1 - 3D Audio in Broadcasting“, S. 10-11.

¹³¹ France Télévisions, „Roland Garros 2019: An UHD event channel with MPEG-H Audio“, 21. Mai 2019.

¹³² Pike, „Chapter 1 - 3D Audio in Broadcasting“, S. 11.

¹³³ International Telecommunication Union, „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production“.

¹³⁴ „Gewinner:innen aus allen vergangenen Staffeln“, ProSieben, aufgerufen am 9. Januar 2025, <https://www.prosieben.de/serien/the-masked-singer/gewinner-innen>.

¹³⁵ „The Masked Singer“, Streaming-Anbieter, Joyn, aufgerufen am 9. Januar 2025, <https://www.joyn.de/serien/the-masked->

Im Format „The Masked Singer“ verkleiden sich zehn Prominente aus unterschiedlichsten Lebensbereichen in bunten Kostümen, durch die sie nicht identifizierbar sind. Einzeln oder im Duell performen sie Songs und verleihen der Maske einen Charakter. Das Publikum und die zwei Rateduos, von denen eins jede Woche mit anderen Prominenten besetzt wird, versuchen anhand von Indizien aus einem Einspieler herauszufinden, welche Person sich unter der Maske verstecken könnte. In jeder Folge wird mindestens eine Maske enthüllt. Welche Maske demaskiert wird und wer am Ende das Finale gewinnt, entscheidet das Publikum in mehreren Votings über die Joyn App. Matthias Opdenhövel moderiert diese Unterhaltungsshow.

Die Publikumsreaktionen, Musikeinspielungen in Kombination mit Gesang, Einspieler und Sprache bilden die tontechnischen Hauptelemente der Show. Zusätzlich werden die Stimmen der Prominenten unter der Maske nach ihrem Auftritt verzerrt, sodass man sie nicht erkennt, wenn sie sprechen. Die Proben beginnen freitags, am Tag vor der Show. An diesem Tag finden heiße Proben und sämtliche Soundchecks mit Kandidaten und möglichen Gastauftritten statt. Am Samstag wird eine Generalprobe durchgeführt, bis der Einlass für das Publikum im Studio startet. Ab 20:15 beginnt die Live-Übertragung der Show auf ProSieben und Joyn.

4.2 Tontechnische Positionen

Für die Produktion einer The Masked Singer-Folge gibt es im Bereich Ton standardmäßig folgende Positionen: Ein Toningenieur ist für den Sendeton verantwortlich und übernimmt dabei die Einrichtung des eigenen Mischpults und die Betreuung der Interkom-Anlage. Für die Musik- und Effektzuspielungen wird ein Sample-Operator eingesetzt. Im Studio befinden sich drei Drahtlostechniker, die für die Verwaltung der Audiofunkstrecken, Platzierung von Atmo-Mikrofonen und korrekte Verteilung und Verkabelung der Hand- und Taschensender an die Kandidaten zuständig sind. Außerdem betreut ein Beschallungsingenieur im Studio die Beschallungsanlage für das Publikum. Ein zusätzlicher Beschallungstechniker kümmert sich um mögliche Havariefälle oder bühnentechnische Umbauten. Im Falle von The Masked Singer kommt das vor,

wenn Gäste aus dem Rateduo einen eigenen Song performen und dafür beispielsweise zusätzliche Stative für eine Gitarre oder Adapter für den Gitarrenausgang auf Klinken benötigen. Aufgrund der hohen Anzahl an musikalischen Acts wird außerdem eine Position für das Monitoring besetzt. Diese Person erstellt für jeden Soundcheck und jede Person eine individuell angepasste Mischung.

4.3 Mikrofonierung

Die Show von The Masked Singer findet grundsätzlich nur auf der Bühne statt. Jeweils links und rechts befindet sich ein Gang, durch den die maskierten Prominenten auf die Bühne gelangen, um ihre Songs zu performen. Frontal zur Bühne sind die vier Personen der zwei Rateduos an einem breiten Pult platziert. In einem Halbkreis um die Bühne ist Platz für das Publikum. Es gibt keine weiteren Schauplätze außerhalb des Studios, wie es beispielsweise für Außenspiele der Fall wäre. Anhand des Bühnenplans kann man erkennen, dass im Studio insgesamt 14 Atmo-Mikrofone von der Decke aufgehängt werden. Sie hängen einzeln möglichst mittig über dem Publikum, um dessen Reaktionen einzufangen. Bei The Masked Singer wird unter anderem ein Kamerakran eingesetzt, der sich durch das Studio bewegt. Die Höhe der Atmo-Mikrofone beträgt etwa 2m von der Stuhlsitzfläche. Aufgrund des Krans kann man die Höhe einiger Atmo-Mikrofone nicht beliebig bestimmen, sondern muss sich mit dem Kran- und Lichtpersonal absprechen. Die Atmo-Mikrofone müssen dann etwas weiter nach oben gezogen werden. Vier von den 14 eingesetzten Mikrofonen haben also statt einer Nierencharakteristik eine Supernierencharakteristik, um die Publikumsreaktionen trotz des vergrößerten Abstands zwischen Mikrofon und Publikum optimal einfangen zu können. Es ist keine erweiternde Drahtlos-Option für die Atmo-Mikrofonierung vorgesehen. Zum Finale der elften Staffel von The Masked Singer wurde zusätzlich zu den Atmos von der Decke für Testzwecke ein ORTF-3D Mikrofonarray mittig über dem Publikum installiert. Weitere Informationen dazu folgen im Kapitel „5.3 Anpassung der Mikrofonierung für 3D-Audio“.

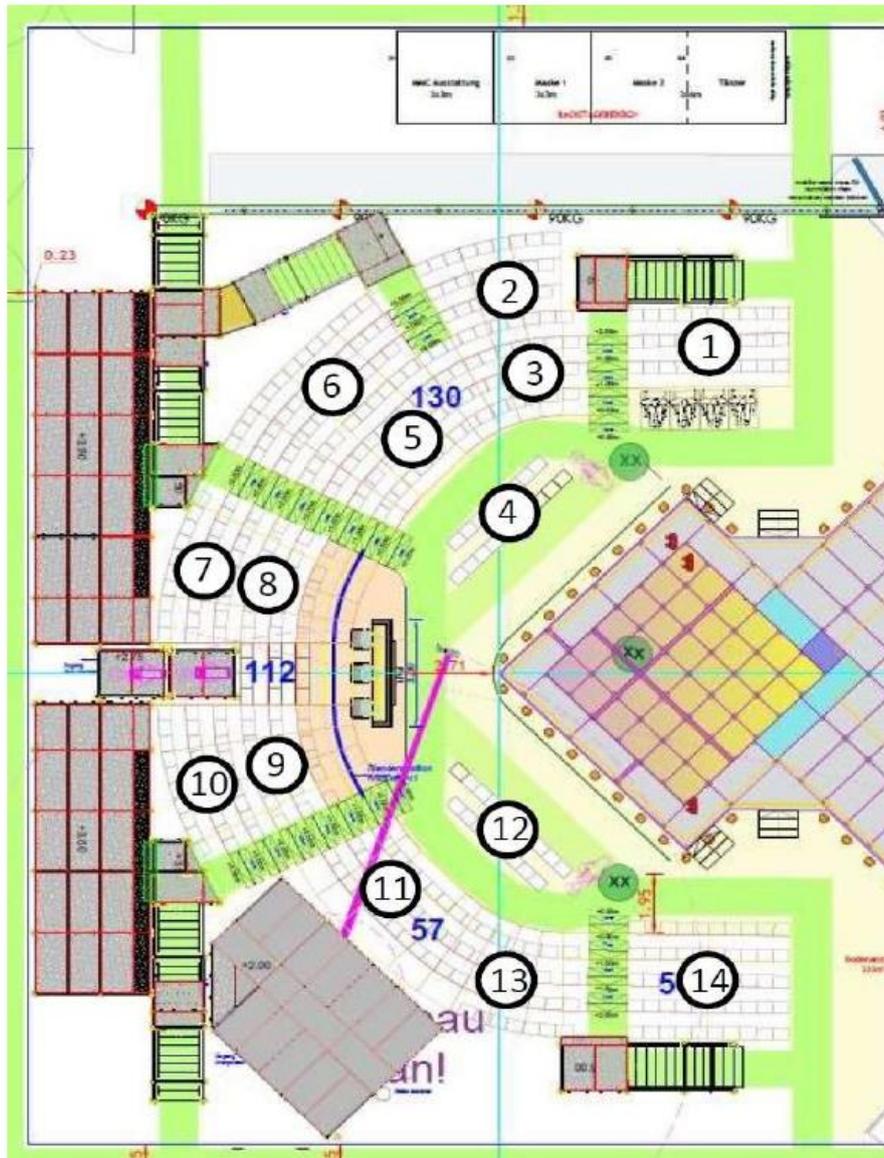


Abbildung 14: Atmo-Mikrofonplan The Masked Singer Staffel 11, 2024

Quelle: MMC interne Kommunikation

Für die drahtlose Audioübertragung sind insgesamt 20 Empfangswege vorhanden. Zehn davon sind für die zehn verschiedenen Masken der Show reserviert, eine ist zusätzlich für den Havariefall oder alternativ Gastauftritte eingeplant. Die vier Personen aus den beiden Rateteams nutzen weitere vier Funkstrecken. Die restlichen stehen für die Moderation der Show und Red-Aftershow zur Verfügung, außerdem gibt es einen Handsender für die Regie und ein weiteres Mikrofon für den Vocal Coach. Insgesamt dienen 19 Stereosendewege als In-Ear-Monitoring für die Masken und Personen aus dem Rateteam.

Um die Audiosignale über Funk zu empfangen, werden Sennheiser EM 9046 Empfänger der Digital 9000er Reihe verwendet. Diese stellen als Handsender SKM 9000 Mikrofone zur Verfügung, alternativ kann ein Taschensender der SK 9000 Reihe verwendet werden. Im Normalfall ist vorgesehen, dass die Masken ihre Gesangsauftritte mit einem Handsender performen. Teilweise sind die Kostüme allerdings ungünstig, sodass die Person in der Maske entweder mit der Hand nicht an ihren Mund kommt, weil das Kostüm im Weg ist oder die Akustik durch den Hohlraum und die Beschallung nicht ganz optimal ist. In diesen Fällen werden die Masken mit einem DPA-4066 Headset verkabelt. Der Audiobelegungsplan ist unter 11.1 als Anhang zu finden.

Für das In-Ear Monitoring werden Sennheiser A SR 2050 Anlagen eingesetzt, die Signale über drei A 2003 UHF Richtantennen versendet. Die Antennen werden oben hinter dem Publikum am Geländer befestigt. Direkt daneben befindet sich außerdem eine zirkular polarisierte Antenne in Form einer Halbkugel. Man kann sie sowohl als Sende- als auch als Empfangsantenne einsetzen, bei The Masked Singer wird sie jedoch als Sende-Antenne eingesetzt und versorgt die In-Ears mit Signalen, damit diese auch hinter der Bühne die Abläufe und ihren Einsatz mitbekommen. Alle eingesetzten Empfangsfrequenzen haben Festfrequenzen in äquidistanten Abständen, die Sendefrequenzen werden intern von den Anlagen berechnet und verwaltet.

4.4 Zuspielsysteme

Für die Produktion von The Masked Singer werden zwei Zuspielsysteme eingesetzt: Die Show wird über ein EVS-System aufgezeichnet, dabei wird temporär eine LSM-Einheit in die bestehende Infrastruktur der Firma MMC eingebunden. Mit diesem Gerät kann man das Programm einschließlich Audiospuren aufzeichnen, aus dem das Material für Einspieler der nächsten Folge entnommen wird. Die EVS dient auch als Wiedergabegerät für vorproduzierte Einspieler. Zusätzlich können auch Slow-Motions produziert werden, im Falle von The Masked Singer wird diese Option allerdings nicht benötigt. Stattdessen bekommen die Zuschauer Highlightschnitte von allen Auftritten der Masken, damit sie beim Voting ihre Stimme abgeben können.

Sämtliche Audiozuspielungen wie Musiken, Jingles oder Effekttöne erfolgen über die R.O.S. Software Teatro in der Version 5.7.0. Hier können in 64 Wiedergabekanälen jeweils mehrere hundert Sounds abgespeichert werden. Es besteht auch die Möglichkeit, Presets zu erstellen und sie in einer Playlist abzuspeichern. Mehrere Kanäle können als Play- oder Stopppgruppe zusammengefasst und gleichzeitig gestartet oder gestoppt werden. Bei The Masked Singer läuft die Zuspielung der Halbplaybacks für die Masken über die vorprogrammierten Presets. Weitere Sounds wie Indikative und Effekte sind hier eingeschlossen. Alle Musiken und Sounds werden standardmäßig in Stereo angeliefert.

4.5 Aktuelle Produktionsstandards im Fernsehen – Audiobereich

Für die Produktion von Audio- und Videomaterial im Rundfunkbereich gelten einige Vorgaben, um die ausgestrahlten Sendungen auf technischer Ebene einheitlich umzusetzen. Die ARD, der ORF und das ZFD folgen ihren technischen Produktionsrichtlinien, die zuletzt im November 2016 aktualisiert wurden. Darin ist unter anderem festgelegt, dass das Mischungsverhältnis zwischen den einzelnen Audioelementen möglichst ausgeglichen sein sollte, die Verständlichkeit der Sprache steht dabei immer im Vordergrund. Auf Grundlage dessen soll eine geeignete Mikrofonierung verwendet werden, die zur jeweiligen Aufnahmesituation passt. Damit können auch Hintergrundgeräusche reduziert werden. Um festzustellen, ob das Mischungsverhältnis auch für den Endnutzer stimmt, wird empfohlen, die Mischung auch in einer Konsumersituation abzuhören und unter realistischen Wiedergabebedingungen zu überprüfen.¹³⁶ Das Standardverfahren in der Medienproduktion und -distribution ist die zweikanalige Stereophonie. In manchen Kategorien, beispielsweise Sport, Unterhaltung, Spielfilm, Musik oder Dokumentationen, ist es sinnvoll, Center- und Surroundkanäle zu ergänzen, der Einsatz von Mehrkanalton ist allerdings nicht zwingend.¹³⁷ Die Codierung von Audiosignalen im Rundfunk erfolgt mit

¹³⁶ Institut für Rundfunktechnik GmbH (IRT), Hrsg., „TPRF-HDTV 2016 Technische Richtlinien zur Herstellung von Fernsehproduktionen in HDTV für ARD, ZDF und ORF“, November 2016, S. 16, <https://www.ard.de/die-ard/TPRF-HDTV-2016-Technische-Richtlinien-zur-Herstellung-von-Fernsehproduktionen-in-HDTV-100.pdf>.

¹³⁷ Institut für Rundfunktechnik GmbH (IRT), S. 17.

einer Abtastfrequenz von 48kHz und einer Bittiefe von 24 Bit.¹³⁸ Bild und Ton sollten synchronisiert sein, sodass kein Versatz entsteht. Ist der Ton dem Bild mehr als 40ms voraus oder mehr als 60ms hinterher, stimmen die Bilder nicht mehr mit dem dazugehörigen Ton überein.¹³⁹ Um die Sprachverständlichkeit für den Zuschauer optimal zu gestalten, gibt es weitere Empfehlungen vom ZDF und der ARD. Sprache und Dialoge können vom Zuschauer klarer abgetrennt werden, wenn Musik und Hintergrundgeräusche einen Abstand von mindestens 7 LU haben. Die Sprachverständlichkeit wird außerdem beeinträchtigt, wenn andere Signale sich im selben Frequenzspektrum befinden.¹⁴⁰

In der Produktion befinden sich Redaktion und Bildregie meist an einem abgesetzten Arbeitsplatz von der Tonregie. Für die Bildregie und Redaktion soll nach Möglichkeit eine eigene Abhöreinheit bereitgestellt werden, die bezüglich Pegel und Laufzeit einen ähnlichen Höreindruck liefert wie der Sweet-Spot des Toningenieurs.¹⁴¹ Auch die Beschallungslautsprecher im Studio können die Sprache möglicherweise negativ beeinflussen, die Platzierung und auch der Pegel der Beschallung sind daher auch für den Sendeton relevant.

Die Qualität des Tons hängt auch von der Kleidung einer Person ab. Grundsätzlich trifft der Toningenieur die Entscheidung darüber, welche Art von Mikrofon eingesetzt wird. Je nach Aufnahmesituation handelt es sich dabei entweder um Handmikrofone, Headsets oder auch Anstecker. In Absprache mit der Redaktion und dem Darsteller wird entschieden, welche Art von Mikrofon für das beabsichtigte Bild eingesetzt wird.¹⁴²

Zu guter Letzt gibt es einige Empfehlungen von der European Broadcast Union (EBU) bezüglich der Lautheitsaussteuerung des Programms. Die Programmlautheit (programme loudness) beschreibt dabei den Durchschnittswert der Lautheit des gesamten Programms in LUFS. Alle Programme sollen laut EBU R128 auf -23,0 LUFS normalisiert werden, sodass zwischen den Sendern und Programmen keine stetigen Wechsel zwischen laut und leise herrschen. In einer

¹³⁸ Institut für Rundfunktechnik GmbH (IRT), S. 18.

¹³⁹ Institut für Rundfunktechnik GmbH (IRT), S. 22.

¹⁴⁰ ARD / ZDF Netzwerk Produktion, Hrsg., „ARD ZDF - Sprachverständlichkeit im Fernsehen Empfehlungen für Programm und Technik zur Sprachverständlichkeit im Fernsehen“, Juli 2023, S. 9, <https://www.ard.de/die-ard/ARD-ZDF-Sprachverstaendlichkeit-im-Fernsehen-102.pdf>.

¹⁴¹ ARD / ZDF Netzwerk Produktion, S. 11.

¹⁴² ARD / ZDF Netzwerk Produktion, S. 12.

Live-Situation lässt sich das Programm nicht exakt auf den Zielwert normalisieren, daher ist eine Abweichung von $\pm 1,0$ LU erlaubt.

Der Lautheitsbereich (loudness range) ist ein weiterer Parameter, um ein Audiosignal zu beschreiben und definiert, wie die Lautheitspegel innerhalb eines Programms verteilt sind. Der exakte maximale Spitzenpegel (maximum true peak level) bezieht sich auf den maximalen Wert in einer stetigen Audiosignal-Wellenform innerhalb des Zeitbereichs im Programm. Dabei sollte der Spitzenpegel den Wert -1dbTP (dB True Peak) nicht überschreiten. Ein Audiosignal soll laut EBU-Empfehlung nach diesen drei Parametern gemessen und bewertet werden, wobei keine besondere Gewichtung auf bestimmten Elementen wie Sprache oder Musik liegt. ¹⁴³

In der MMC werden hauptsächlich Sendungen in 2.0 Stereo und HD umgesetzt, gelegentlich gibt es auch Formate in UHD HDR oder in 5.1 Surround. Beim ZDF ist das 5.1 Surround Format der Standard. In der „Production Guideline“ für Mehrkanalton im Fernsehen ist die Kanalreihenfolge L, R, C, LFE, Ls, Rs festgelegt. Die Kanäle Links und Rechts sind dabei der Abbildung von Stereosignalen wie Musik vorenthalten, trotzdem steht es dem Toningenieur frei, diese Kanäle auch für andere Signale zu nutzen. Der Center-Kanal erweitert die beiden Frontkanäle für eine bessere Lokalisation, diesem Kanal wird hauptsächlich Sprache zugewiesen. Geringe Teile der Atmo und Musik können auch über den Center-Kanal wiedergegeben werden. Die Surroundkanäle Ls und Rs sind insbesondere für die Erzeugung von Räumlichkeit vorgesehen, die durch Publikumsatmo, Musiken und Effekte erreicht wird. ¹⁴⁴

Die Produktion einer Sendung im 5.1 Surround Format setzt voraus, dass die Mischung ein Wiedergabeformat mit weniger Kanälen nicht negativ beeinträchtigt. Bei einem Downmix zu Stereo verändert sich die Räumlichkeit, da Informationen zum Klangbild anders verteilt werden und somit Bewegungen verloren gehen. Auch die Lautstärkeverhältnisse der einzelnen Signale zueinander sind möglicherweise nicht optimal aufeinander abgestimmt. Es steht

¹⁴³ G. Spikofski und F. Camerer, Übers., „EBU - Empfehlung R128 - Lautheitsaussteuerung, Normalisierung und zulässiger Maximalpegel von Audiosignalen“ (Genf: European Broadcast Union, September 2011), https://tech.ebu.ch/docs/r/r128_2011_DE.pdf.

¹⁴⁴ Michael Eberhard u. a., „Production Guideline - Mehrkanalton im Fernsehen“ (ARD / ZDF, September 2013), S. 8.

immer die Sprachverständlichkeit im Vordergrund. ¹⁴⁵ Die Abwärtskompatibilität wird durch einfaches Matrizieren erreicht, bei dem die Kanäle umverteilt werden. Der Center-Kanal wird dabei mit jeweils -3dB auf Links und Rechts verteilt. Der Ls-Kanal wird ebenfalls mit -3dB auf Links umverteilt, identisch ist es mit Rs auf Rechts. ¹⁴⁶

	L	R	C	LS	RS
L' =	1,0 L	0,0 R	0,7 C	0,7 LS	0,0 RS
R' =	0,0 L	1,0 R	0,7 C	0,0 LS	0,7 RS

Tabelle 5: Downmixparameter von 5.1 Surround zu 2.0 Stereo

(0,7 entspricht einer Absenkung von -3dB)

Quelle: IRT, 2016, S. 23

5. Konzeption

5.1 Aufzeichnung der Live-Show "The Masked Singer"

Um 3D-Mischversuche vorzunehmen, wurden mehrere Folgen von „The Masked Singer“ aufgezeichnet, die Show wurde nicht live in 3D-Audio ausgestrahlt, sondern dient als experimentelles Material für mögliche zukünftige Ausstrahlungen von Unterhaltungsshows in 3D-Audio. Teil der Aufzeichnung sind das Programmbild mit der Stereosumme und 60 Audiospuren. Von sechs Folgen der elften Staffel aus dem Jahr 2024 wurden vier Folgen aufgezeichnet, wobei Folge 3, 4 und 5 bei der Mikrofonierung keine Besonderheiten aufweist, für das Finale in Folge 6 wurde zusätzlich zur bestehenden Atmo-Mikrofonierung ein ORTF-3D Mikrofonarray eingesetzt, um mögliche Vergleiche zu ziehen. Das Programmbild wurde im DNxHD Codec aufgezeichnet und hatte eine Auflösung von 1920x1080i bei 25fps, es wurde das Dateiformat mxf verwendet. Zugehörig zum Programmbild wurden ebenfalls Timecode und die Stereosumme des Programms aufgezeichnet. Für jeden Videokanal stehen bei der Aufzeichnung je acht Audiokanäle zur Verfügung, die beliebig belegt werden können. Im Fall von The Masked Singer sind alle acht Kanäle von Gruppen belegt, die der Toningenieur im Mischpult festgelegt hat. Die genaue Belegung ist im Anhang 11.2 zu finden.

¹⁴⁵ Eberhard u. a., S. 11-12.

¹⁴⁶ International Telecommunication Union, „Recommendation ITU-R BS.775-4, Multichannel Stereophonic Sound System with and without Accompanying Picture“, S. 8.

Außerdem wurden die 60 Audiospuren für höhere Flexibilität bei den geplanten Mischversuchen auch von der DAW (Digital Audio Workstation) Nuendo 11 von Steinberg aufgezeichnet, die auf einem Windows PC lief. Insgesamt wurden 60 Audiospuren vom Lawo-Mischpult über eine MADI-Glasleitung an den aufzeichnenden Rechner übertragen. Die Quellen wurden dabei direkt und ohne weitere Bearbeitung weitergeleitet und aufgezeichnet. Für die Audioaufnahmen wurde eine Samplingrate von 48kHz und eine Bittiefe von 24 Bit verwendet.

5.2 Allgemeine Informationen zur Tonregie 11

Zum Zeitpunkt dieser Arbeit besitzt die Tonregie 11 auf dem Produktionsgelände der MMC die technische Ausstattung für Produktionen in 5.1 Surround. Ihr Grundriss hat die Form eines Rechtecks, welches die Breite von etwa 6,25m und die Länge von etwa 5,75m besitzt. Der Raum hat eine Höhe von 2,93m. Die Tonregie bietet einen abgesetzten Arbeitsplatz für den Toningenieur, den Sample-Operator und gegebenenfalls (Musik-)Redakteure. Das ermöglicht es, sich ungestört auf das Abhören und Mischen zu konzentrieren.

Im Raum sind zwei Reihen für die Arbeitsplätze aufgebaut, vorne im Sweet-Spot sitzt der Toningenieur und rechts neben ihm der Sample-Operator. Die Musikredakteure sind dahinter an einem erhöhten Holztisch platziert. Frontal mittig im Raum befindet sich ein Regiefenster mit doppelter Verglasung, dabei sind die Scheiben jedoch nicht parallel zueinander, sondern angewinkelt angeordnet. Die Schallwellen, die auf diese Scheibe treffen, werden nicht in dieselbe Richtung zurück reflektiert. Außerdem sind die Wände des Raums mit Schallabsorbern ausgestattet, die den Nachhall im Raum und nach außen reduzieren.

An der rechten Seite des Raums befindet sich im hinteren Bereich, neben dem hinteren Holztisch, ebenfalls eine Glasscheibe, allerdings nur mit einfacher Verglasung. Der Grundriss im Anhang 11.3 zeigt die frühere Raumausstattung der Tonregie 11.

5.3 Anpassung der Mikrofonierung für 3D-Audio

In Kapitel „4.3 Mikrofonierung“ wird die Mikrofonierung der Künstler unter den Masken sowie die Atmo-Mikrofonierung beschrieben. Je nach Kostüm verwenden die Personen entweder ein Headset-Mikrofon oder einen Handsender. Unter

Umständen kann die eingesetzte Mikrofonart nicht nur vom Kostüm abhängen, sondern auch vom performten Song. Es ist möglich, dass die Künstler nicht kontinuierlich die gleiche Art von Mikrofon erhalten, um ihre Songs zu performen, sondern dass es sich in den Folgen unterscheidet.

Allgemein ist es für die Produktion einer 3D-Audio Fernsehsendung nicht erforderlich, die Mikrofonierung der Sprache zu erweitern oder zu verändern. Sprache lässt sich als Direktschall kategorisieren und wird somit vor allem auf die Frontlautsprecher L, C, R verteilt. Sämtliche Moderationen, Jurybewertungen und Gesang bilden Einzelschallquellen, die für den Zuschauer lokalisierbar sein sollen. Headsetmikrofone und Handmikrofone nehmen Monosignale auf und sind somit weiterhin geeignet für eine 3D-Audio Produktion. Im Falle einer Band wird ebenfalls die Einzelmikrofonierung der Instrumente beibehalten, damit man die Instrumente in beliebigem Mischverhältnis verwenden und im Raum verteilen könnte. Diese Einzelschallquellen ergeben ebenfalls klar lokalisierbare Monosignale.

Die einzige Anpassung der Mikrofonierung erfolgt bei der Atmo. Als diffuses Audioelement nimmt sie Publikumsreaktionen jeglicher Art auf und stellt eine räumliche Beziehung zu den Einzelschallquellen her. Die Atmo-Mikrofone nehmen im Fall von The Masked Singer auch ungewollt und unvermeidbar einen Teil der Musik aus der Beschallung auf, auch wenn die von der Decke hängenden Mikrofone mit Nieren- und Supernierencharakteristik nicht primär auf die Lautsprecher ausgerichtet sind. Auch wenn die Aufnahme von Musik ursprünglich nicht beabsichtigt ist, lässt sich damit gegebenenfalls die Umhüllung des Zuschauers fördern. Durch die Kombination aus diffusem Schallsignal und direkten Einzelschallquellen erhält der Zuschauer den Raumeindruck einer lebendigen Atmosphäre. Es bietet sich an, den diffusen räumlichen Eindruck bei dreidimensionalem Audio auch in die vertikale Ebene zu verteilen, um ein zusammenhängendes Klangbild hervorzurufen. Ziel ist es, den Zuschauer vollständig mit Klang zu umhüllen, sodass man die Mikrofonensignale eines diffusen Signals nicht eindeutig lokalisieren kann, während Direktschall klar verständlich und eindeutig identifizierbar von vorne kommt. Zudem soll eine möglichst große Hörzone gebildet werden, innerhalb der das Klangbild unabhängig von der Hörposition nahezu unverfälscht wahrgenommen wird.

Um die Hörzone ideal auszuweiten und Kammfiltereffekte bei einem Downmix in Formate mit weniger Lautsprechern zu vermeiden, ist es notwendig, möglichst dekorrelierte Signale aufzunehmen. Bei einem Downmix wird das Mischungsverhältnis verändert, dies ist allerdings bei Atmo aufgrund der Diffusität nicht relevant. Die Publikumsreaktionen sind keine Einzelschallquelle, deren klar verständliches Signal möglicherweise untergehen könnte. Das Panning von einzelnen Atmo-Signalen auf diskrete Lautsprecher vergrößert die Hörzone, da sie stabile Quellenpositionen bilden, die auch beim Downmix leichter umverteilt werden können.

Drei der vier aufgezeichneten Folgen der elften Staffel von The Masked Singer wurden mit der Standard-Mikrofonierung für die Atmo durchgeführt. Für das Finale wurde ein ORTF-3D Mikrofonarray von Schoeps installiert. Es besteht aus insgesamt acht Supernieren, die in zwei Ebenen angeordnet sind und jeweils eine Neigung von 45° nach oben oder unten besitzen. Die Platzierung wurde insbesondere durch das Showlicht und den Kamerakran eingeschränkt, die durch die Installation des Mikrofonarrays nicht beeinträchtigt werden sollen. Dennoch war es möglich, die ORTF-3D Anordnung fast symmetrisch über dem Publikum und im Raum zu platzieren. Sowohl die unteren Mikrofone als auch die oberen nehmen somit sinnvolle Signale mit hoher Dekorrelation auf, da auch Reflexionen vom gesamten Studio das Array erreichen.



Abbildung 15: ORTF-3D Positionierung im Studio bei The Masked Singer

Quelle: eigene Darstellung

Das ORTF-3D bietet durch die kompakte HSVC-Anordnung ausreichend Laufzeit- und Pegelunterschiede und somit dekorrelierte Signale in der horizontalen Ebene. In der vertikalen Ebene ist das nicht notwendig, stattdessen sorgen die unterschiedlich ausgerichteten Supernieren für ausreichende Unterschiede zwischen den aufgenommenen Signalen.

Alle weiteren Atmo-Mikrofone weisen durch ihre Entfernung zueinander und die unterschiedliche Höhe über dem Publikum ebenfalls ausreichend dekorrelierte Signale auf.

5.4 Lautsprecherformat

Die Tonregie 11 kann bis zum Zeitpunkt dieser Arbeit Audioinhalte in 5.1 produzieren. Die Erweiterung auf eine 5.1.4 Lautsprecheraufstellung lässt sich mit den bestehenden 5 Hauptlautsprechern leicht realisieren. Dolby präferiert die 7.1.4 Aufstellung für die Produktion von 3D-Audio Inhalten, die 5.1.4 Formation wird von Dolby jedoch auch für die Produktion dieser Inhalte anerkannt.¹⁴⁷ Eine aktuelle Methode der 3D-Audio-Produktion im Fernsehbereich ist die kanalbasierte 5.1.4 Produktion, die anschließend mit einer Dolby Digital Plus JOC codiert wird. Sie bildet die Grundlage für diese Arbeit. Die von der ITU vorgegebene Bezeichnung, Kanalreihenfolge und Positionen der einzelnen Lautsprecher sind im Kapitel „4.5 Aktuelle Produktionsstandards im Fernsehen – Audiobereich“ aufgelistet und werden eingehalten.¹⁴⁸

5.5 Abhöre

Für die Abhöre in 5.1 Surround stehen in der Tonregie 11 zwei unterschiedliche Lautsprecherarten zur Verfügung. Die drei Frontlautsprecher L, R, C sind KS-digital ADM 20 Lautsprecher. Die beiden Surround-Lautsprecher Ls und Rs sind hingegen vom Typ KS-digital-ADM3. Es ist üblich, dass alle Lautsprecher einer Ebene vom selben Typ sind, allerdings ist es ebenso sinnvoll, die Front mit einem größeren Typ zu besetzen. Für den LFE-Kanal wird ein KS digital ADM B2 eingesetzt, der anders als die meisten LFE-Kanäle nach oben abstrahlt. Es wird von Dolby empfohlen, den LFE-Kanal nicht in einer Linie mit dem Center-Kanal zu

¹⁴⁷ Dolby Laboratories, Inc., „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, S. 9.

¹⁴⁸ International Telecommunication Union, „Recommendation ITU-R BS.2051-3 „Advanced Sound System for Programme Production““.

platzieren ¹⁴⁹, was in der Tonregie 11 nicht gegeben ist. Für eine homogenere Abstrahlung der tieffrequenten Effekte ist der LFE-Kanal jedoch phasengedreht. Für eine 5.1.4 Konfiguration, mit der 3D-Abhören möglich ist, werden vier Höhenlautsprecher vom Typ Genelec 8030A ergänzt.

Dolby empfiehlt, die Höhenlautsprecher in der Decke zu platzieren. In der Tonregie 11 gibt es aus Sicherheitsgründen keine Möglichkeiten, dies realistisch und kurzfristig umzusetzen. Stattdessen wurden die Höhenlautsprecher auf Stativen im Raum installiert. Sie befinden sich nicht in direkter Linie über den L, R, Ls und Rs Kanälen, sondern etwas weiter innen. Die Ausrichtung und Platzierung erfolgt innerhalb der Toleranzbereiche, die von der ITU vorgegeben werden. Die Positionen der Lautsprecher aus der mittleren Ebene wurden dabei nicht verändert.

Die neue 5.1.4 Abhöre wurde mit dem bestehenden 5.1 System eingemessen und aneinander angepasst. Beim Einmessen fiel auf, dass zwischen der oberen und mittleren Ebene ein Ungleichgewicht des Pegels auftritt, die obere Ebene war im Verhältnis etwas zu laut. Es ist allerdings nicht sinnvoll, die jeweiligen diskreten Summen, die auf die oberen Lautsprecher geroutet sind, um den gemessenen Pegelwert zu verringern, da dies Auswirkungen auf den Downmix hat. Zunächst sollte auch vermieden werden, den Pegel direkt am Lautstärkeregler des Lautsprechers anzupassen. Es wurde versucht, ein DSP-Offset in die oberen Kanäle einzufügen. Allerdings stellte sich heraus, dass dieser Parameter beim Erstellen eines neuen Snapshots im Lawo-Mischpult nicht übernommen wird und daher leider nicht praktikabel ist. Aus Mangel an Alternativen wurde der Pegel der oberen Lautsprecher schließlich doch am Lautstärkeregler selbst angepasst.

¹⁴⁹ Dolby Laboratories, Inc., „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, S. 15.



Abbildung 16: vorher, Tonregie 11 in 5.1

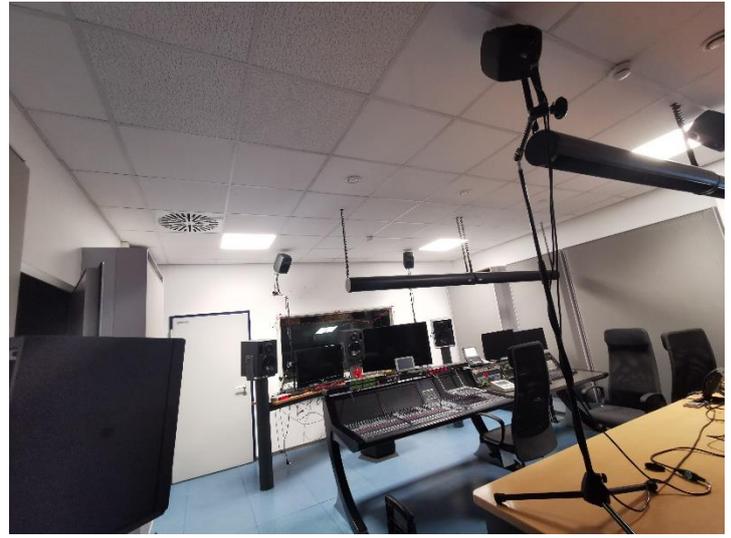


Abbildung 17: nacher, Tonregie 11 in 5.1.4

Quelle: Eigene Darstellung

5.6 Upmix

Der Upmix wandelt Audiosignale mit wenig Kanälen in ein Format mit mehreren Kanälen um, dabei können entweder zusätzliche Signalanteile hinzugefügt werden oder das vorhandene Signal umverteilt werden. Bei The Masked Singer werden in Stereo angelieferte Musiken von 2.0 auf 5.1 hochgemischt. Dafür wird der interne Upmixer „Ambit“ (AMBience IT) des Lawo-Mischpults verwendet, der laut Hersteller vollständig downmixkompatibel ist. Man kann mit ihm sowohl Input aus 2 Kanälen in ein 5.1 Signal konvertieren als auch ein bestehendes 5.1 Signal räumlicher klingen lassen (spatialize).¹⁵⁰ Einspieler und Musiken werden als Stereogruppen zum Ambit geroutet, hochgemischt und räumlich verteilt.

5.7 Downmix

Das Gegenteil vom Upmix ist der Downmix, der den Vorgang beschreibt, bei dem Audioinhalte mit mehreren Kanälen in ein Format mit weniger Kanälen umgewandelt werden, beispielsweise von 5.1 Surround zu 2.0 Stereo. Dabei werden einzelne Kanäle nicht einfach weggelassen, sondern mit einer vorgegebenen Pegelabsenkung auf andere Kanäle verteilt. Der entscheidende Aspekt bei der Erstellung eines Downmixes ist, dass die Audiosignale der verschiedenen Formate sich nicht wesentlich unterscheiden dürfen, sondern auch

¹⁵⁰ „mc²56 - AMBIT (Upmix)“, o. J., aufgerufen am 09.02.25, <https://docs.lawo.com/audio-production-mc-mixing-consoles/mc56/mc56-operation/mc56-stereo-and-surround/mc56-ambit-upmix>.

im Format mit weniger Kanälen (z.B. 2.0 Stereo) alle relevanten Inhalte vorhanden sind und nicht beim Downmix untergehen. Aus einer Mischung im 5.1.4 Format gehen eine 5.1 Surround Mischung und eine 2.0 Stereo Mischung hervor. Alle drei Formate können zwar gleichzeitig und ohne erheblichen Mehraufwand produziert werden, allerdings kann man sie nicht gleichzeitig abhören. Um also zu überprüfen, ob aus dem Dolby Atmos Format 5.1.4 zwei korrekte Downmixe entstehen, muss der Toningenieur zwischen den Formaten umschalten.

Der Downmixparameter bestimmt, um wie viel dB der Pegel eines Kanals abgesenkt wird, um dann auf einen anderen Kanal verteilt zu werden. Für den Downmix von 5.1 Surround auf 2.0 Stereo gibt es bereits vorgegebene Richtwerte. Die Surroundkanäle Ls und Rs werden mit einer Pegelabsenkung von -3dB auf den L bzw. R Kanal verteilt. Bei diesem Vorgang wird die im Lawo-Mischpult vorhandene Downmixmatrix verwendet.

	L	R	C	LFE	Ls	Rs
L	0		-3		-3	
R		0	-3			-3

Tabelle 6: Downmixmatrix von 5.1 zu 2.0

Quelle: eigene Darstellung

Beim Downmix von 5.1.4 auf 5.1 liegen keine konkreten Vorgaben für eine Pegelabsenkung oder vorgefertigte Matrizen im Mischpult vor. Deshalb wird zu Beginn mit einem Downmixparameter von -6dB gearbeitet, der im folgenden Kapitel „Mischung“ überprüft und nach Bedarf angepasst wird. Die Höhenkanäle Ltf, Rtf, Ltr und Rtr werden dabei auf die zugehörigen Kanäle der mittleren Ebene verteilt.

	L	R	C	LFE	Ls	Rs	Ltf	Rtf	Ltr	Rtr
L	0						-6			
R		0						-6		
C			0							
LFE				0						
Ls					0				-6	
Rs						0				-6

Tabelle 7: Vorläufige Downmixmatrix von 5.1.4 zu 5.1

Quelle: eigene Darstellung

Durch die Pegelabsenkungen der Kanäle, die beim Downmixvorgang umverteilt werden, ist der Toningenieur bei gestalterischen Entscheidungen etwas eingeschränkt. Es sollte sichergestellt werden, dass dramaturgisch relevante Signale nicht auf die Höhenkanäle oder Surroundkanäle verteilt werden, da sie beim Downmixvorgang möglicherweise untergehen könnten und das im Vorfeld angestrebte Mischverhältnis nicht mehr stimmig ist. Die drei angefertigten Mischungen sind in diesem Fall nicht mehr nahezu identisch und somit wäre das Ziel einer abwärtskompatiblen Mischung verfehlt.

Die Downmixkompatibilität ist außerdem von der Dekorrelation der Signale abhängig. Bei einer hohen Korrelation können im Downmixvorgang durch die Addition von zwei Signalen unerwünschte Kammfiltereffekte entstehen.

5.8 Mischpuleinrichtung

Die im Folgenden beschriebene Mischpuleinrichtung bezieht sich ausschließlich auf das Abmischen der Sendung „The Masked Singer“. Kommando und weitere Elemente, die für das Durchführen einer Live-TV Show relevant sind, werden hier nicht berücksichtigt.

Die Tonregie 11 ist mit dem Lawo Mischpult mc² 56 mkII ausgestattet, welches nicht nativ über 5.1.4 Subgruppen oder Summen verfügt. Um mit diesem Mischpult 3D-Audio mischen und abhören zu können, muss mit einer alternativen Lösung gearbeitet werden. Dafür gibt es die Möglichkeit, das 5.1.4 Format zu simulieren, indem zwei Surround Summen kombiniert werden. Eine der beiden Summen bedient dabei die mittlere Ebene, die andere Summe ist für die Höhenkanäle vorgesehen. In der oberen Ebene sind kein Center und LFE-Kanal vorhanden, daher werden diese nicht bespielt. Mithilfe von horizontalem Panning innerhalb der zwei Summen können jedoch Phantommitten erstellt werden. Wie die Recherchen in Kapitel „2.6 Verwendung der Höhenlautsprecher“ zeigten, ist vertikales Panning zwischen der oberen und unteren Ebene wenig sinnvoll, da die Lokalisierung in der Medianebene insbesondere über spektrale Veränderungen geschieht und somit nur die Lautsprecher selbst als stabile Quelle eingesetzt werden können. Pegelunterschiede zwischen der oberen und mittleren Ebene können bei identischem Signal ein vertikales Panning hervorrufen. Das pegelbasierte Panning wird über die Subgruppen gesteuert, die auf die Summen gehen. Ein zu großer Pegelunterschied führt jedoch dazu, dass das Klangbild sich

überwiegend in eine der beiden Ebenen verschiebt oder die Schallquelle eindeutig lokalisiert werden kann, somit wäre das Ziel des umhüllenden Klangerlebnisses verfehlt.

Insgesamt wurden drei 5.1 Surround Gruppen eingerichtet, zwei davon sind für die Atmo vorgesehen, eine für die mittlere Ebene und eine für die obere Ebene. Eine weitere 5.1 Surround Gruppe beinhaltet alle Elemente der Acts.

Die Gruppe Mod enthält das Mikrofon des Moderators, außerdem gibt es die Gruppe Rateteam, welche alle vier Headsetmikrofone der Mitglieder aus dem Rateteam umfasst. Die Line-Gruppe ist in Stereo angelegt und durchläuft keinen Upmix, sie beinhaltet sämtliche Jingles und Effekte sowie die EVS-Highlights und Auftrittsmusiken. Für die Acts wurden zwei Gruppen angelegt, eine davon in Stereo und eine als Resultat des Upmixes in 5.1 Surround. Über diese Gruppe werden alle Halbplaybacks der Masken und Einspieler, die sogenannten „Indizien“ zugespielt. Die Vocal Gruppe ist als Teil der Act 2.0 Gruppe in Stereo angelegt und wird somit ebenfalls hochgemischt, sie umfasst die Backing-Vocals, falls vorhanden sowie den Gesang der Masken. Zuletzt gibt es noch die Talk Gruppe, die dafür vorgesehen ist, die Masken nach einer Enthüllung unverzerrt sprechen zu lassen. Die Talk FX Gruppe enthält die Voice Pitches der Masken, wenn sie nach dem Auftritt mit verzerrter Stimme sprechen.

Für die vereinfachte Steuerung der Gruppen wurden VCAs angelegt, die übergreifend mehrere Signale über einen einzigen Fader bedienen. Die beiden VCAs Atmo Top und Atmo Bottom steuern in der oberen und mittleren Ebene die Atmos, damit kann das Verhältnis zwischen oben und unten bestimmt werden. Sobald dieses Verhältnis feststeht, steuert der übergeordnete VCA Atmo-all die beiden VCAs. Außerdem wurde ein eigener VCA für das ORTF-3D Mikrofonarray angelegt, welches alle acht Mikrofone der beiden Ebenen einschließt, dieser VCA wird ebenfalls übergeordnet vom Atmo-All VCA bedient. Moderation und Rateteam werden von den zwei weiteren VCAs Mod und Rate gesteuert.

Für die erste Stufe des Downmixes von 5.1.4 auf 5.1 gibt es intern im Lawo-Mischpult der Tonregie 11 keine vorprogrammierte Matrix, daher wird dieser Downmix manuell durchgeführt. Die 5.1.4 Summe wird im Mischpult als Input angelegt, um aus ihr den 5.1-Downmix zu generieren. Dabei werden die

einzelnen Bestandteile der 5.1.4 Summe (L, R, C, LFE, Ls, Rs, Ltf, Rtf, Ltr, Rtr) auf die jeweilig gewünschten Pegel gebracht, welcher am Fader eingestellt wird.

5.9 Bild-Ton-Synchronität

Bild und Ton sollten zueinander synchron sein, es sollte also kein zeitlicher Versatz entstehen. Es ist nicht ausreichend, nur die Stereosumme des Programms anzugleichen. Die unterschiedlichen Arten von Quellen wie Sprachsignale, Musik und Atmo haben jeweils unterschiedliche Laufzeiten. Da die Sprache und Musiken eine Laufzeit von nahezu null haben, werden sie um die volle Zeit verzögert, die es braucht, damit Bild und Ton synchron sind. In einer Fernsehproduktion wird die Latenz zwischen Bild- und Tonsignal vorher mit einem Testsignal gemessen. Der Toningenieur stellt Verzögerungen (Delays) ein, die sich in den Subgruppen unterscheiden. Je nach Sendung wird der Ton um etwa 1-2 Frames verzögert. Bei The Masked Singer beträgt die Verzögerung der Sprache etwa 80ms, bei allen anderen Quellenarten wie Atmo und Musik sind es etwa 100ms.

Die Verzögerung der Atmo lässt sich allerdings nicht einheitlich bestimmen. Durch die unterschiedliche Höhe und Position der Atmomikrofone treffen Beschallungssignale nicht an allen Atmos gleichzeitig ein. Bei der Wiedergabe der Atmo kann es deshalb zu einer Asynchronität kommen, bei der die Musik mehrfach und nicht gleichzeitig wiedergegeben wird. Der Takt, in dem das Publikum klatscht oder andere Reaktionen zeigt, ist ebenfalls nicht mehr synchron und stimmt weder mit den anderen Atmomikrofonen noch mit dem Takt der Musik überein. Deshalb werden Atmomikrofone individuell gemessen, indem man ein Signal wie Pink Noise (Rauschen) über die Beschallung abspielt und mithilfe von Software berechnen lässt, wie lange es dauert, bis das Rauschen bei den Atmos wieder ankommt. Die daraus errechnete Differenz ist die nötige Verzögerungszeit. Gegebenenfalls wird der Durchschnitt aus den unterschiedlichen Werten berechnet, um die Atmos ungefähr zu vereinheitlichen. Bei The Masked Singer erhält die Standard-Atmo Mikrofonierung ein Delay von etwa 30ms. Die akustische Laufzeit addiert sich somit zu einer Gesamtverzögerung von 100ms, das ORTF-3D Mikrofon ist hier nicht eingeschlossen. Der Grund dafür ist, dass beim Messen festgestellt werden konnte, dass die Laufzeiten des ORTF-3D Mikrofons wesentlich kürzer waren,

weshalb sie in einer eigenen Subgruppe eine andere Verzögerung zugewiesen bekommen haben. Die akustische Laufzeit des 3D-Mikrofonarrays beträgt etwa 60-70ms.

5.10 Lautheitsmessung

Das Mischpult Lawo Mc² 56 mkII besitzt interne Möglichkeiten, um die Lautheit zu messen. In der Tonregie 11 wird jedoch das externe Gerät RTW TM9 verwendet, um eine einheitliche Lautheit feststellen zu können. Das RTW erhält unter normalen Umständen alle nötigen Signale über vier AES-Wege, also insgesamt acht digitale Kanäle, alternativ stehen ebenfalls acht analoge Wege zur Verfügung. Da acht Kanäle allerdings für eine 5.1.4 Mischung nicht ausreichend sind, bot sich an, die Audiosignale stattdessen über die SDI-Schnittstelle an das RTW zu leiten.

Für die Lautheitsmessung von dreidimensionalem Audio wird die Software Erweiterung „Immersive Sound Analyzer“ (ISA) benötigt, die im Rahmen dieser Arbeit auf dem Gerät installiert wurde. Bis zu 16 Kanäle können somit gleichzeitig gemessen werden. Eine Audiogruppe auf dem RTW kann allerdings nur 8 Kanäle verwalten, weshalb mit einer Kombination aus einer 5.1 Gruppe und einer 4.0 Gruppe gearbeitet wurde.

6. Mischung

6.1 Sprache

Sprachanteile der Sendung können entweder vom Center-Lautsprecher, der Phantommitte oder durch eine Kombination der beiden Möglichkeiten wiedergegeben werden. Bei der Anfertigung der Mischung wurden alle Optionen angewendet, allerdings wurde bevorzugt, Sprache ausschließlich über den Center wiederzugeben.

In Kombination mit Atmo können die Moderation und das Rateteam besser vom Applaus differenziert werden. Wenn man Sprache über die Phantommitte aus L und R wiedergibt, wäre sie mit in den Applaus eingebettet und würde etwas weniger präsent sein, da von L und R ebenfalls Applaus wiedergegeben wird. Headset- und Handmikrofone der Moderation und des Rateteams nehmen unvermeidlich auch Applaus auf und haben somit Auswirkungen auf das

räumliche Klangbild. Bei Headsetmikrofonen tritt dieser Effekt aufgrund der Kugelcharakteristik verstärkt auf. Möglicherweise ist ein zu lautes Headsetmikrofon während eines Applauses das lauteste Atmo-Mikrofon und verzerrt dadurch das Atmo-Bild, sodass es nach vorne in sich zusammenfällt. Der Zuschauer bekommt so das Gefühl, als wäre der Applaus zu nah und die Räumlichkeit geht verloren. Bei einer Wiedergabe der Sprache aus der Phantommitte über L und R könnte dieses Phänomen in noch größerem Ausmaß auftreten.

Im Stereo-Downmix Vorgang wird der Center-Kanal auf L und R umverteilt, weshalb Sprachanteile aus der Phantommitte abgebildet werden. Die beschriebene Problematik lässt sich dabei nicht vollständig vermeiden, aus diesem Grund sollte in höheren Formaten der verfügbare Center-Kanal genutzt werden.

Für die Mischung von Moderation und Rateteam wurde zur Vereinfachung ein Automixer eingesetzt, der positive Auswirkungen auf die Sprachverständlichkeit hat.

Die eingesetzten Funkmikrofone wurden nicht einzeln komprimiert und limitiert, sondern übergreifend in den Gruppen Mod, Rateteam und Talk.

6.2 Musik

Die musikalischen Elemente der Sendung „The Masked Singer“ sind insbesondere Halbplaybacks und der Gesang der Masken. Als Teil der Act 2.0 Gruppe werden diese Bestandteile in den Upmixer geschickt und bespielen somit die mittlere Ebene als 5.1 Surround. Im Gegensatz zur Sprache ist es jedoch bei Musik sinnvoll, den Gesang aus dem Phantomcenter abzubilden und den Center-Lautsprecher nicht zu verwenden. Das Ziel ist hier, dass der Gesang sich einheitlich in das Halbplayback einfügt und nicht davon abhebt, wie es bei Sprache beabsichtigt ist.

Die Breitbandlautsprecher der Tonregie 11 ermöglichen Bassmanagement, wodurch tieffrequente Effekte gleichmäßig im Raum wiedergegeben und verteilt werden können, sie ergänzen den LFE-Kanal.

Auftrittsmusiken und Jingles sind ebenfalls musikalische Elemente der Sendung, durchlaufen allerdings nicht den Upmixer, da dies Auswirkungen auf die

Sprachverständlichkeit im Downmix haben könnte. Der charakteristische Soundeffekt zum Lichtwechsel kurz vor dem Auftritt der Maske bildet die Ausnahme, da keine sprachlichen Elemente dadurch gefährdet werden. Bei Gesang kann es sinnvoll sein, eine Parallelkompression einzurichten, die ein eingehendes Signal auf zwei Wege aufteilt, ein Teil davon wird sehr stark komprimiert und der andere Teil bleibt unverändert. Durch die Zusammenführung dieser beiden Signalteile erhält man ein volleres und dynamischeres Signal. Bei The Masked Singer wird jedoch keine Parallelkompression angewendet, da das Eingangssignal durch die komplex gestalteten Masken bereits teilweise unerwünschte Signalanteile aufnimmt, die bei einer Parallelkompression noch stärker hervorgehoben werden und das Signal unsauber machen würden. Für Hall- und Delay Effekte wird ein t.c. electronic 6000 in das System integriert. Mithilfe eines AUX-Weges kann bestimmt werden, welcher Pegelanteil der Gesangsmikrofone in den Hall oder Delay gesendet wird.

6.3 Atmo

Die Mischung der Atmo erfolgt aus der Perspektive eines Zuschauers, der mittig im Publikum hinter dem Rateteam sitzt und frontal zur Bühne schauen kann. Die Achssprünge zwischen Bühne und Rateteam werden nicht berücksichtigt, da Audiosprünge in diesem Zusammenhang für den Zuschauer ein unrealistisches und unangenehmes Klangbild erzeugen.

Die Atmo setzt sich aus einer Kombination des ORTF-3D Mikrofonarrays und den einzelnen Atmo-Mikrofonen zusammen. Die acht Mikrofone des ORTF-3D wurden zunächst auf die zugehörigen Lautsprecher der oberen und mittleren Ebene geroutet. Das Mikrofonarray hatte eine günstige Positionierung im Studio und musste deshalb nur minimal im Lautstärkeverhältnis untereinander angepasst werden. Sämtliche Atmo-Mikrofone erhielten einen Hochpassfilter bei 150Hz mit einer Flankensteilheit von 18dB pro Octave, außerdem wurden die Frequenzbereiche um 250 und 500Hz leicht abgesenkt. Frequenzen um den Bereich von 7000Hz wurden um 4dB angehoben. Für eine verlustfreie Rauschunterdrückung wird in allen Atmo-Mikrofonen der Cedar-DNS als Denoiser über einen Insert eingebunden.

Zu Beginn waren alle diskreten Atmos nur der mittleren Ebene zugeteilt, dabei wurde das übliche Stereo-Panning übernommen. Die Mikrofone befanden sich

gleichmäßig verteilt von Links nach Rechts in den Phantommitten, der Center selbst wurde nicht bespielt. Die untere Hälfte des ORTF-3D Mikrofonarrays ergänzte die mittlere Ebene mit vier weiteren Signalen.

In den vier Höhenkanälen waren zunächst nur die vier oberen Mikrofone des ORTF-3D geroutet.

Alle diskreten Atmo-Mikrofone wurden doppelt angelegt und jeweils miteinander verlinkt, um die einfache Nutzung in beiden Ebenen zu garantieren. Hierbei werden Panning, Filter und Equalizer identisch übernommen. Ausgewählte diskrete Atmos können ebenfalls der oberen Ebene zugeteilt werden, um ein volleres Klangbild der Atmo zu erhalten. Aus der gewählten Perspektive als Zuschauer im Publikum ergibt sich, dass die Atmos 2, 6, 7, 8, 9, 11 und 12 ebenfalls in die Höhenkanäle geroutet werden, da sie sich überwiegend hinter dem Zuschauer befinden. Es ist auch sinnvoll, diese Atmos zu wählen, da sie trotz ihrer Supernierencharakteristik in der größeren Entfernung zum Publikum ein leicht distanzierteres und diffuseres Signal aufgenommen haben und somit gut in die obere Ebene passten.

Im hinteren Bereich sind zusätzlich zu den Signalen des ORTF-3D nur die zwei Atmos 7 und 10 gepannt, da sich sonst das Raumbild nach hinten oder oben verschiebt und der Bildbezug nach vorne verloren geht.

Die Atmo wurde außerdem mit dem Richtungsmischer im Parameter „Breite“ (width) um drei erhöht, das lässt das Klangbild etwas breiter wirken.

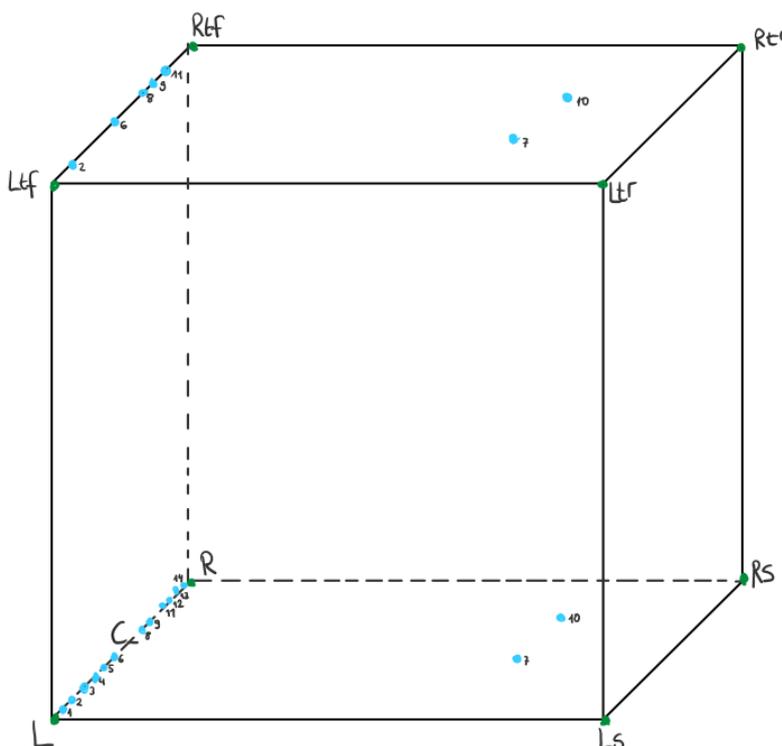


Abbildung 18: Panning der Atmos in beiden Ebenen

Quelle: eigene Darstellung

- ORTF-3D
- Diskrete Atmo-Mikrofone

Im direkten Vergleich von zwei verschiedenen Songs mit und ohne das ORTF-3D zeigte sich, dass die Abbildung eines räumlichen Klangbildes im Dolby Atmos Format auch ohne ein spezielles 3D-Mikrofonarray möglich ist. Die oben genannte Verteilung der Mikrofone zwischen oben und unten blieb unverändert, es wurde lediglich das ORTF-3D Array ausgeschlossen. Das Lautstärkeverhältnis der Atmo veränderte sich leicht durch die wegfallenden Signale, wodurch der Pegel auf 0dB stand, statt auf den vorherigen -3dB mit ORTF-3D.

Es wurde festgestellt, dass das ORTF-3D allein zwar sehr ausgewogen klingt und den Raumeindruck aus dem Studio realistisch vermittelt, jedoch keine ausreichende Publikumsatmo abbilden kann. Die Einzelatmos nehmen den Applaus sehr direkt auf und werden durch das ORTF-3D Array ergänzt. Die Kombination aus den räumlichen und distanzierteren Signalen mit den klaren und präzisen Einzelmikrofonsignalen ermöglicht eine realistische Raumabbildung. Jedes Atmo-Mikrofon nimmt unvermeidbar auch Signale der Beschallung auf, im Kontext von 3D-Audio kann das allerdings zum Vorteil der Räumlichkeit sein, wenn auch von oben in kleinen Teilen Musik wahrnehmbar ist.

6.4 Gesamte Sendung

Im Rahmen dieser Arbeit sind zwei Mischungen entstanden, die sich im digitalen Anhang befinden, beide Dateien sind in als Zip-Container angelegt. Alle diskreten Audiospuren im wav-Format und zugehörige Videodateien im mp4-Format befinden sich in diesen Containern. „*TMS_Pirat_Another_Love_Atmos_und_Downmixe*“ enthält die angefertigte Mischung für einen Ausschnitt des Piraten aus dem „The Masked Singer“ Finale der elften Staffel. In der zweiten Datei „*TMS_Qualle_Everytime_Atmos_und_Downmixe*“ wurde zum Vergleich ein etwas kürzerer Ausschnitt der Qualle aus Folge 3 gemischt, in der das ORTF-3D Mikrofonarray nicht vorhanden war.

Das Lawo-Mischpult stellt die Option zur Verfügung, Mischprozesse im zeitlichen Verlauf zu automatisieren. Die Fader-Einstellung war mit dem Timecode der aufgezeichneten Audiospuren aus den einzelnen Sendungen verknüpft und konnte mit jedem Durchlauf angepasst werden, bis die zeitliche Abstimmung und Pegel zufriedenstellend waren. Das Resultat der Aufzeichnung ist eine

zehnkanalige Summe der Mischung, bei der das Bild nachträglich hinzugefügt wurde.

Der gewählte Ausschnitt aus dem Finale enthält alle tontechnisch relevanten Elemente wie Moderation, Einspieler, Gespräche des Rateteams, ein Song der Maske, Publikumsreaktionen und den Voice Pitch der Maske. Diese Mischung ist der Schwerpunkt der vorliegenden Arbeit, weshalb der zweite Ausschnitt aus Folge 3 kürzer gewählt wurde.

Nach der kurzen Anmoderation des Piraten tritt er mit einer Auftrittsmusik auf die Bühne, gleichzeitig applaudiert und jubelt das Publikum. Die Lautstärke von Auftrittsmusiken bleibt über den gesamten Verlauf der Sendung gleich, auch wenn das Publikum durch teilweise starke und laute Reaktionen einen Teil der Musik verdecken könnte. Der Auftritt der Maske und die Freude des Publikums steht hier jedoch im Vordergrund, die Musik soll nur begleitend sein, weshalb sie nicht unbedingt isoliert zu hören sein muss.

Der Einspieler mit den Indizien des Piraten wurde versuchsweise mit dem Upmixer in 5.1 hochgemischt. Aufgrund des hohen Anteils an Musik wurde das Ergebnis als zufriedenstellend empfunden, wobei der Center-Lautsprecher trotz der auch vorhandenen Sprachanteile nicht verwendet wird.

Nach den Indizien findet ein charakteristischer Lichtwechsel mit Soundeffekt statt, das dem Rateteam Zeit gibt, kurz über die potenzielle Person unter der Maske zu sprechen. Dieser Soundeffekt wird als einzige Ausnahme auch in den Upmixer geschickt. Währenddessen spricht niemand, weshalb keine Sprachanteile durch ein Downmixverfahren in der Verständlichkeit und Präsenz gefährdet sind.

Nach einiger Zeit werden die Mitglieder des Rateteams langsam ausgeblendet und die Maske beginnt, ihren Song zu performen. Im Fall vom Piraten hat das Publikum etwa zur Hälfte des Songs angefangen, rhythmisch zu klatschen. Hier kommt es auf die individuellen Gestaltungspräferenzen an, inwiefern man Atmo zu einem Song hinzumischt. In einigen Fällen würde man Atmo vollständig weglassen, hier wurde sich jedoch entschieden, die Stimmung des Publikums einzufangen, sie aber nicht in vollem Pegel dazu zu mischen.

Sobald der Pirat seine Performance beendet hat, werden Moderation und Rateteam wieder eingeblendet und die Atmo mit vollem Pegel aufgemacht.

Während das Rateteam Feedback und Vermutungen äußert, wird die Atmo etwas lauter gemischt, damit man das Lachen im Publikum besser einfangen kann. Sobald ein Applaus einsetzt, wird der Fader wieder etwas eingezogen. Die angestrebte integrierte Programmlautheit von $-23 \text{ LUFS} \pm 1\text{LU}$ konnte im Durchschnitt eingehalten werden, an etwas lauterem Applausstellen kann die Lautheit teilweise darüber hinaus gehen. Diese Stellen halten allerdings nicht lange an und können somit als leichte Abweichung vom Durchschnitt akzeptiert werden.

Die Downmixparameter wurden im Vorfeld festgelegt und mit dieser Mischung überprüft. Bei einem Downmix von 5.1.4 auf 5.1 wurden Tests mit einem Downmixparameter von -3dB und -6dB durchgeführt. Das Ergebnis zeigte, dass -3dB in Teilen die Sprache im Zielformat verdeckte. Mit einem Parameter von -6dB ging im Downmix die angestrebte Räumlichkeit etwas verloren. Als Resultat wurde sich für den mittleren Wert von $-4,5\text{dB}$ entschieden.

	L	R	C	LFE	Ls	Rs	Ltf	Rtf	Ltr	Rtr
L	0						-4,5			
R		0						-4,5		
C			0							
LFE				0						
Ls					0				-4,5	
Rs						0				-4,5

Tabelle 8: Downmixmatrix von 5.1.4 zu 5.1

Quelle: eigene Darstellung

Die vom Lawo-Mischpult vorgegebene Downmixmatrix von 5.1 auf 2.0 mit einem Parameter von -3dB wurde als angemessen empfunden.

	L	R	C	LFE	Ls	Rs	Ltf	Rtf	Ltr	Rtr
Sprache			x							
Musik (Upmix)	x	x		x	x	x				
Einspieler (Upmix)	x	x			x	x				
Jingles, Effekte, Auftrittsmusiken	x	x								
Publikumsatmo	x	x			x	x	x	x	x	x

Tabelle 9: Kanaluweisungen der gesamten Mischung

Quelle: eigene Darstellung

6.5 Codierung

Die angefertigte Mischung wurde mithilfe von der DAW Nuendo 13 in eine 5.1.4 ADM-Datei exportiert. In dieser Version von Nuendo ist ein Plug-in für einen Dolby Atmos Renderer integriert, mit dem man ein Dolby Atmos Projekt erstellen kann. Über das Plug-in lässt sich im angelegten Projekt die Anzahl der Objekte und die Art des Bettes definieren. Die Audiokanäle werden den entsprechenden Betten oder Objekten zugewiesen. Der Renderer kann allerdings nur maximal 7.1.2 als Bed anlegen, welches für eine 5.1.4 Mischung nicht passend ist. Daher wurde mit einem 5.1 Bed gearbeitet, dem vier statische Objekte hinzugefügt wurden. Der Masterbus wurde als 5.1.4 konfiguriert.

Das angelegte Projekt wurde im ADM BWF Format exportiert, es enthält für jeden Kanal einen kurzen Audio-Test, um sicherzustellen, dass alle Kanäle am korrekten Lautsprecher ankommen. Außerdem beinhaltet die Datei den zugehörigen Videoausschnitt mit Timecode im MP4 Format und die zehnkanalige Mischung sowie 5.1 und 2.0 Downmixe.

7. Fazit

Im Rahmen dieser Arbeit wurde anhand aufgezeichneter Folgen von „The Masked Singer“ ein Audio-Setup für Produktionen in Dolby Atmos entwickelt und in der Praxis überprüft. In Anlehnung an eine Folge „Let´s Dance“ aus dem Jahr 2023, die aus der Remote-Regie Darmstadt ebenfalls in Dolby Atmos produziert wurde, beschäftigt sich diese Arbeit mit 3D-Audio Fernsehproduktionen auf dem eigenen Firmengelände. Dies schließt die Anpassung der Produktionsumgebung und Durchführung des entwickelten Konzepts ein.

Sämtliche Klangelemente einer TV-Sendung wie Publikumsreaktionen, Musik, Einspieler und Sprache wurden auf verschiedene Weisen gestalterisch umgesetzt. Daraus resultierte, dass insbesondere Publikumsatmo gut geeignet ist, dreidimensional abgebildet zu werden und den Zuschauer von allen Seiten einzuhüllen. Entscheidend für den Eindruck der Umhüllung ist die Dekorrelation der Signale.

Einige Bestandteile der Sendung lagen zu Beginn nur in Stereo vor und wurden deshalb mit dem Lawo-internen Upmixer in das 5.1 Surround Format hochgemischt. Ein Upmix in höhere Formate wäre aufgrund der erforderlichen Downmixkompatibilität nicht sinnvoll gewesen. Diese Arbeit zielte darauf ab, eine funktionierende 5.1.4 3D-Audio Mischung zu produzieren, aus der sich außerdem mithilfe von Downmixverfahren eine 5.1 Surround und 2.0 Stereo Mischung generieren lassen, ohne dass sich der Produktionsaufwand erheblich erhöht. Aufgrund der notwendigen Downmixkompatibilität sind die gestalterischen und kreativen Entscheidungen beim Mischen eingeschränkt. Dramaturgisch relevante Bestandteile einer TV-Show wie Moderation sollten durch einen Downmixprozess nicht weniger verständlich und präsent sein, eine Platzierung in der hinteren oder oberen Ebene ist nicht sinnvoll.

Als Ergebnis dieser Arbeit entstanden zwei Mischungen von verschiedenen Ausschnitten von „The Masked Singer“, wobei im Finale das ORTF-3D Mikrofonarray integriert wurde. Der Vergleich zeigte, dass theoretisch auch eine 3D-Audio Produktion ohne ein spezielles 3D-Mikrofonarray möglich ist. Das ORTF-3D ergänzt jedoch die direkten Atmo-Signale mit einer ausgeglichenen und realistischen Räumlichkeit.

Die Aufzeichnung erfolgte nachträglich mit den Audiospuren der Live-Sendung. Zukünftige Produktionen in Dolby Atmos können mit der entwickelten Vorgehensweise produziert und live gesendet werden, sofern der jeweilige Sender über ein geeignetes Distributionssystem verfügt.

Aus dem Interview mit Emilio Bonta geht hervor, dass Dolby Atmos und 3D-Audioformate auch im Fernsbereich zukünftig an Relevanz gewinnen könnten. Der Einsatz von Dolby Atmos beschränkt sich im Fernsbereich meistens auf eine kanalbasierte 3D-Audio Mischung, was jedoch im Vergleich zu den Weiterentwicklungen im Bildbereich ein angemessenes Gegenstück wäre. Die Personalisierbarkeit der Audiospuren ist für Konsument ebenfalls ein vorteilhafter Aspekt von Dolby Atmos und Next Generation Audio. Neben der Tatsache, dass teilweise die Sender selbst dieses Format noch nicht an die Endnutzer senden können, stellt es eine Schwierigkeit dar, dass die wenigsten Konsument zuhause ein korrekt eingemessenes Lautsprechersystem besitzen, mit dem eine Wiedergabe von Inhalten in Dolby Atmos optimal umgesetzt werden kann. Außerdem sind Produktionsfirmen nicht bereit, die Mehrkosten für etwas zu tragen, das die Einschaltquoten nicht erheblich verbessert. Langfristig könnten sich dennoch 3D-Audio Formate auch im Fernsbereich etablieren, da sie bereits in vielen anderen Entertainment Bereichen Anwendung finden und die Qualität auch in den niedrigeren Kanalordnungen nicht beeinträchtigt wird.

8. Literaturquellenverzeichnis

- ARD / ZDF Netzwerk Produktion, Hrsg. „ARD ZDF - Sprachverständlichkeit im Fernsehen Empfehlungen für Programm und Technik zur Sprachverständlichkeit im Fernsehen“, Juli 2023. <https://www.ard.de/die-ard/ARD-ZDF-Sprachverstaendlichkeit-im-Fernsehen-102.pdf>.
- Blauert, Jens. „2.1 Localization and Localization Blur“. In *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996. <https://doi.org/10.7551/mitpress/6391.001.0001>.
- . „Sound Localization in the Median Plane“. *Acta Acustica united with Acustica* 22 (30. November 1969).
- Blauert, Jens, und Jonas Braasch. „Räumliches Hören“. In *Handbuch der Audiotechnik*, herausgegeben von Stefan Weinzierl, 87–121. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. https://doi.org/10.1007/978-3-540-34301-1_3.
- Dickreiter, Michael. „Kapitel 1.3.1 - Zeitlicher Aufbau des Schallfeldes“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 3.4 - Räumliches Hören natürlicher Schallquellen“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 3.4.1.1 - Horizontale Ebene“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 3.4.1.2 - Vertikale Ebene“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 17.6.2.2 - Weitere Mehrkanalton-Formate“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 8. überarbeitete und Erweiterte Auflage. Bd. 1. Berlin/Boston: De Gruyter Saur, 2014. <https://doi.org/10.1515/9783110316506>.
- „Dolby Atmos Specifications“. Dolby Laboratories, Inc., April 2024.
- Dolby Laboratories, Inc. „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, 6. Mai 2021.
- . „Dolby Atmos Is the Future of Entertainment Sound“. *Business Wire (English)*, 23. April 2012. <https://www.businesswire.com/news/home/20120423005868/en/Dolby-Atmos-Is-the-Future-of-Entertainment-Sound>.
- Dolby Professional. „Dolby Atmos Documentation“, o. J. <https://professional.dolby.com/gaming/gaming-getting-started/dolby-atmos-documentation/#fmod>.
- Eberhard, Michael, Erich Ebert, Dr. Elmar Hergenröder, Max Kiefer, Daniel Matejka, Wolfgang Ort, und Askan Siegfried. „Production Guideline - Mehrkanalton im Fernsehen“. ARD / ZDF, September 2013.
- „ETSI TS 103 190-2 V.1.2.1 Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio“, Februar 2018.
- „ETSI TS 103 491 V.1.1.1 DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields“, April 2017.
- France Télévisions. „Roland Garros 2019: An UHD event channel with MPEG-H Audio“, 21. Mai 2019.
- Goeres-Petry, Jürgen. „Kapitel 3.1.1 - Außenohr“. In *Handbuch der Tonstudioteknik*, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.

- . „Kapitel 3.1 - Das Gehör“. In *Handbuch der Tonstudiotchnik*, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Hamasaki, Kimio, Toshiyuki Nishiguchi, Reiko Okumura, Yasushige Nakayama, und Akio Ando. „A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV)“. *SMPTE Motion Imaging Journal* 117 (1. April 2008): 40–49. <https://doi.org/10.5594/J15119>.
- Institut für Rundfunktechnik GmbH (IRT), Hrsg. „TPRF-HDTV 2016 Technische Richtlinien zur Herstellung von Fernsehproduktionen in HDTV für ARD, ZDF und ORF“, November 2016. <https://www.ard.de/die-ard/TPRF-HDTV-2016-Technische-Richtlinien-zur-Herstellung-von-Fernsehproduktionen-in-HDTV-100.pdf>.
- International Telecommunication Union. „Recommendation ITU-R BS.775-4 ‚Multichannel Stereophonic Sound System with and without Accompanying Picture‘“. Radiocommunication Sector of ITU, Dezember 2022.
- . „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production‘“. Radiocommunication Sector of ITU, Mai 2022.
- „ISO/IEC 23008-3:2022 Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D Audio“, August 2022.
- Joyn. „The Masked Singer“. Streaming-Anbieter, 9. Januar 2025. https://www.joyn.de/serien/the-masked-singer?utm_medium=referral&utm_source=www.prosieben.de&utm_campaign=joynlinkout_The+Masked+Singer&utm_content=all_episodes_format_teaser&utm_term=format_episode_overview#folge.
- Kirby, D. G., N. A. F. Cutmore, und J. A. Fletcher. „Programme Origination of 5-channel surround sound“. Tadworth, UK: British Broadcast Corporation Research and Development, 1. April 1998. <https://downloads.bbc.co.uk/rd/pubs/reports/1998-17.pdf>.
- Lee, Hyunkook. „Chapter 5 - Psychoacoustics of Height Perception in 3D Audio“. In *3D Audio*, 1. Auflage. Perspectives on Music Production. New York: Taylor & Francis, Routledge, 2022.
- . „Multichannel 3D Microphone Arrays: A Review“. *Journal of the Audio Engineering Society*, Applied Psychoacoustics Lab (APL), University of Huddersfield, Huddersfield, United Kingdom, 69, Nr. January/February 2022 (6. Januar 2021). <https://doi.org/10.17743/jaes.2020.0069>.
- Lee, Hyunkook, und Christopher Gribben. „Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array“. *Journal of the Audio Engineering Society* 62 (5. Januar 2015): 870–84. <https://doi.org/10.17743/jaes.2014.0045>.
- Mars, Nate. „Producing Spatial Audio in Dolby Atmos, Binaural, and Surround“. LinkedIn, 28. Juli 2022. <https://www.linkedin.com/learning/producing-spatial-audio-in-dolby-atmos-binaural-and-surround/>.
- „mc²56 - AMBIT (Upmix)“, o. J. <https://docs.lawo.com/audio-production-mc-mixing-consoles/mc56/mc56-operation/mc56-stereo-and-surround/mc56-ambit-upmix>.
- Meltzer, Stefan. „Kapitel 13.4.7.3 - Dolby Atmos“. In *Handbuch der Tonstudiotchnik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Møller, Henrik. „Fundamentals of binaural technology“. *Applied Acoustics* 36, Nr. 3 (1992): 171–218. [https://doi.org/10.1016/0003-682X\(92\)90046-U](https://doi.org/10.1016/0003-682X(92)90046-U).
- Nipkow, Lasse. „Kapitel 5.4.5.3 - Höhenlautsprecher für immersiven Sound“. In *Handbuch der Tonstudiotchnik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Pfanzagl-Cardone, Edwin. „1.5 Inter Aural Cross-Correlation (IAAC) and the Binaural Quality Index of Reproduced Music (BQIrep)“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.

- . „1.7 Some Thoughts on Psychoacoustic Signal Interaction in Multichannel Microphone Array Systems“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „1.8 A Few Thoughts on Microphone Pattern Choice and Capsule Orientation in 3D Audio“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „1.10 Distribution of Direct- and Diffuse-Sound Reproduction in Multichannel Loudspeaker Systems“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2 - ‚3D‘ or ‚Immersive‘ Audio - The Basics and a Primer on Spatial Hearing“ in *The Art and Science of 3D Audio Recording*.“ In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2.2 Basics of Sound Perception in Humans“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2.3 Mechanisms of Localization“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2.3.2 Localization and HRTFs“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2.5 Spatial Impresion“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „4.2 Dolby Atmos - An Overview“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „4.3 Multichannel Speaker-Layout: Improved Audio Quality and Timbre Matching“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „4.7.4 The Future of Object-Based Audio“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „9.11 The Morton Lindberg ‚2L‘ Technique“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „9.11.1 Use of Center Speaker and LFE-Channel“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „9.12 The OCT-3D Technique“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „9.13 The ORTF-3D Technique“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- Pike, Chris. „Chapter 1 - 3D Audio in Broadcasting“. In *3D Audio - Perspectives on Music Production*, 1. Auflage. New York: Taylor & Francis, Routledge, 2022. <https://doi.org/10.4324/9780429491214>.
- ProSieben. „Gewinner:innen aus allen vergangenen Staffeln“, 9. Januar 2025. <https://www.prosieben.de/serien/the-masked-singer/gewinner-innen>.
- Rakerd, Brad, William Hartmann, und Timothy McCaskey. „Identification and localization of sound sources in the median sagittal plane“. *The Journal of the Acoustical Society of America* 106 (1. Dezember 1999): 2812–20. <https://doi.org/10.1121/1.427615>.
- Rayleigh, Lord. „XII. On our perception of sound direction“. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13, Nr. 74 (1. Februar 1907): 214–32. <https://doi.org/10.1080/14786440709463595>.
- Reil, Andreas A., und Wilm Brucker. „Vom Einkanal- zum Surroundton“. In *Praxisbuch: Ton - perfekter Ton für News, TV-Beiträge, Industriefilme und Eigenproduktionen*. mediabook Verlag, 2007.
- Ringrose, Fergal. „Next Generation Audio Summit: Inside Sky’s journey through Dolby Atmos production“. Europe: SVG Europe, 6. November 2018.
- Robjohns, Hugh. „Surround Sound Explained: Part 4“. *Sound on Sound/Techniques/Monitors-Monitoring*, November 2001. <https://www.soundonsound.com/techniques/surround-sound-explained-part-4>.

- Roginska, Agnieszka, und Paul Geluso. *Immersive Sound: The Art and Science of Binaural and Multi-Channel-Audio*. New York: Taylor & Francis, Routledge, 2017.
- Schnick, Detlev, und Philipp Kind. „Fakten rund um Dolby Atmos“, 21. November 2024. https://www.hifi-regler.de/wissenswertes_und_kaufberatung/kino-tonformate/dolby_atmos.php?srsId=AfmBOoqz_KZeafLy5NjJ1r-OUqsaSXdAkoxURXL-4dg4ly5FRvXNeLu.
- Scholz, Florian C. „10 - Surround und Mehrkanalton“. In *Audiotechnik für Mediengestalter*. De Gruyter Saur, 2015.
- . „10.4.1 - Mordernere digitale Surroundformate“. In *Audiotechnik für Mediengestalter*. De Gruyter Saur, 2015.
- Spikofski, G., und F. Camerer, übers. von. „EBU - Empfehlung R128 - Lautheitsaussteuerung, Normalisierung und zulässiger Maximalpegel von Audiosignalen“. Genf: European Broadcast Union, September 2011. https://tech.ebu.ch/docs/r/r128_2011_DE.pdf.
- Spikofski, Gerhard. „Kapitel 5.5.5.1 - Kopfbezogene Übertragung“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Theile, Günther. „Kapitel 5.4.1.3 - Tieftonwiedergabe“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 5.4.5.1 - Nutzen der Höhenlautsprecher“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 5.4.5.2 - Anwendung der Höhenlautsprecher“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- . „Kapitel 5.5.5.3 - Lautsprecherwiedergabe“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Vrtonung, und Daniela Rieger. „Dolby Atmos Music - Was ist dieses 3D Sound Erlebnis genau?“, o. J. <https://www.vrtonung.de/dolby-atmos-music/#:~:text=Dolby%20Atmos%20wurde%20im%20Jahr,Streaming%2DDienste%20wie%20Netflix%20abrufbar>.
- „What Sample Rates Does Dolby Atmos Support?“ FAQ in Knowledge. *Dolby Atmos FAQs - Mixing in Atmos*, 4. Januar 2021. https://professionalsupport.dolby.com/s/dolby-atmos-faqs?language=en_US&tabset-c803d=3.
- Wittek, Helmut, und Günther Theile. „Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR“. *Journal of The Audio Engineering Society*, 2017. <https://api.semanticscholar.org/CorpusID:116160580>.

9. Abbildungsquellenverzeichnis

- Blauert, Jens. „Sound Localization in the Median Plane“. *Acta Acustica united with Acustica* 22 (30. November 1969).
- . *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996. <https://doi.org/10.7551/mitpress/6391.001.0001>.
- Dickreiter, Michael. „Kapitel 1.3.1 - Zeitlicher Aufbau des Schallfeldes“. In *Handbuch der Tonstudioteknik*, herausgegeben von Michael Dickreiter, 9. aktualisierte und Erweiterte Auflage. Bd. 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- Goeres-Petry, Jürgen. „Kapitel 3.1 - Das Gehör“. In *Handbuch der Tonstudioteknik*, 9. aktualisierte und Erweiterte Auflage. Bd. Band 1 u. 2. Berlin/Boston: Walter de Gruyter GmbH & Co KG, 2023.
- International Telecommunication Union. „Recommendation ITU-R BS.775-4 ‚Multichannel Stereophonic Sound System with and without Accompanying Picture‘“. Radiocommunication Sector of ITU, Dezember 2022.
- „Morten Lindberg´s 2L Cube“, 2022. <https://www.lindberg.no/wp-content/uploads/2022/10/1.jpg>.
- Pfanzagl-Cardone, Edwin. „1.3 Surround Microphone Case Study: OCT-Surround“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „2.3 Mechanisms of Localization“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- . „9.13 The ORTF-3D Technique“. In *The Art and Science of 3D Audio Recording*. Switzerland: Springer, 2023.
- Schoeps. „Schoeps OCT-3D“, 2025. <https://schoeps.de/produkte/surround-3d/oct-3d.html>.
- Wittek, Helmut, und Günther Theile. „Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR“. *Journal of The Audio Engineering Society*, 2017. <https://api.semanticscholar.org/CorpusID:116160580>.

10. Tabellenquellenverzeichnis

- Dolby Laboratories, Inc. „Dolby Atmos Home Entertainment Studio - Technical Guidelines“, 6. Mai 2021.
- Institut für Rundfunktechnik GmbH (IRT), Hrsg. „TPRF-HDTV 2016 Technische Richtlinien zur Herstellung von Fernsehproduktionen in HDTV für ARD, ZDF und ORF“, November 2016. <https://www.ard.de/die-ard/TPRF-HDTV-2016-Technische-Richtlinien-zur-Herstellung-von-Fernsehproduktionen-in-HDTV-100.pdf>.
- International Telecommunication Union. „Recommendation ITU-R BS.2051-3 ‚Advanced Sound System for Programme Production‘“. Radiocommunication Sector of ITU, Mai 2022.

11. Anhänge

11.1 Audiobelegungsplan The Masked Singer S11

Produktion		The masked singer Staffel D Staffel 11						
Regie		11		Studio			53	
STB1/AK3	Was	AK1	Freespeak	(ZGR QV3)	IEM/Antenne/Wer/MHz			
Mic/Line 1	Atmo 1	1	Freespeak	AnalogExt.9-13	Combiner AC3200 I	Rate1	685,2	
2	Atmo 2	2				Rate2	683,175	
3	Atmo 3	3				Rate3	682,525	
4	Atmo 4	4				Rate4	680,925	
5	Atmo 5	5				Mod	678,9	
6	Atmo 6	6				Mod RED	668,8	
7	Atmo 7	7				Maske1	664,25	
8		8				Maske2	662,85	
9		9				Maske3	661,75	
10		10				Maske4	658	
11	Atmo 8	AK2	Backstage	(ZGR QV1)	Combiner AC3200 I	Maske5	657,25	
12	Atmo 9	1				Maske6	655,475	
13	Atmo 10	2				Maske7	638,5	
14	Atmo 11	3				Maske8	637,525	
15	Atmo 12	4				Maske9	636,775	
16	Atmo 13	5				Maske10	634,35	
17	Atmo 14	6				einzeln	Steady	631,6
18		7				einzeln	Cue	630,475
19		8				Funk		
20		9	PGM RED	ZGR Stf3		Ton	11/1	
21		10			Bild	11/2		
22			Foh/Kran	(ZGR QV2)	AL intern	11/3		
23		1	KDO < Kran	AnaExt8 Out	Gästebetreuung	11/4		
24		2	KDO Kran >	AnaExt8 In	Regiemithören nur über Sprechstelle, IEM und Freespeak			
25		3	Tally Kran	K9	ZGR Stf	Out	In	
26		4	Havarie PB < PA/Mon	QV TRG4	4	PGM VTV		
27		5	Kopfhörer 1	ZGR Stf5	5&9	Kopfhörer 1&2		
28		6	Tally Magcam	K6	12	TC>Licht/MS		
29		7	Spots	2DRaht1	QV TRG	Was	woher/wohin	
30		8	HavSum < TRG	QV TRG8	4	Havarie PB < PA	A-PC1 A-Out	
31		9	Kopfhörer 2	ZGR Stf9	8	HavSum < TRG	TRG Stf1 In	
32		10	TC < Licht/MS Havarie	ZGR Stf 12 Out				
AES In		IC Studio	Wer	wohin	IC St/Reg	Wer	wohin	
AES1-1	BR	Regie/Warmup	Mobil 1	Licht1 R1012	SCU 1/1	Mobil13	Pyro D1016	PMX Mob1
AES1-2		Vocal coach	Mobil 2	Licht2 R1012	SCU 1/2	Mobil14	MagcamD1116	PMX Mob2
AES2-1	UHF1	Maske1	Mobil 3	Licht3 R1012	SCU 1/3	Mobil15	PA D1016	PMX Mob3
AES2-2		Maske2	Mobil 4	Licht4 R1012	SCU 1/4	Mobil16	Monitor D1016A	PMX Mob4
AES3-1		Maske3	Mobil 5	Medienserver R1012	SCU 2/1	Port 3.1	CVD2 R1028	BRG
AES3-2		Maske4	Mobil 6	ServerCity D1116	SCU 2/2	Port 3.2	CVD3 R1012	BRG
AES4-1		Maske5	Mobil 7	Kran D1016+Audio Out	SCU 2/3	Port 3.3	CVD4 R1012	BRG
AES4-2		Maske6	Mobil 8	Voting	zbv	Port 3.4	Bildred. R1016	BRG
AES5-1		Maske7	Mobil 9	Autor D1016	SCU 3/1	Port 3.5	Proding	BRG
AES5-2		Maske8	Mobil 10	Vocalcoach R2016	SCU 3/2	Port 2.7	Prodbü. D1016	Raum 53.013
AES6-1		Maske9	Mobil 11	Musikred. D1016	TRG			
AES6-2		Maske10	Mobil 12	EVS RED	zbv			
AES7-1	UHF2	Maske11 (Havarie/Gast)	4-Draht	Out	In	GPI (out)	In	Was
AES7-2		RED Hand	D1	Steady	PGM	Lawo1	Riedel1	Tally VDR
AES8-1	UHF3	RED Mod	D2	IEM Mod	Pfi Mod	Lawo2	Riedel2	Tally EVS
AES8-2		Rate1	D3	IEM Mod RED	Pfi Mod RED	Lawo3	Riedel3	Tally EVS2
AES9-1		Rate2	D4	IEM Rateteam1	Pfi Rateteam1	Lawo4	Riedel4	Tally SG
AES9-2		Rate3	D5	IEM Rateteam2	Pfi Rateteam2	Riedel1	Lawo1	MT
AES10-1		Rate4	D6	IEM Rateteam3	Pfi Rateteam3	Riedel2	Lawo2	VT
AES10-2		Mod	D7	IEM Rateteam4	Pfi Rateteam4	Riedel3	Teatro1(9)	PB Start
AES11-1			D8	Studio KDO	Pfi Regie Hand	VMix1	Lawo3	SFX (Teatro Mid)
AES11-2			D9	KDO Maske1	PGM Mod	Teatro1	Summit1	Prompterclips
AES12-1			D10	KDO Maske2	PGM Mod RED	Teatro2	V-Mix1	DVE FX
AES12-2			D11	KDO Maske3	PGM Rateteam1	Lawo STB2/AK1		
Line Out		D12	KDO Maske4	PGM Rateteam2	Line Out			
Out 1	Studio KDO	D13	KDO Maske5	PGM Rateteam3	Out 1	KDO Maske1		
Out 2	PB L	D14	KDO Maske6	PGM Rateteam4	Out 2	KDO Maske2		
Out 3	PB R	D15	KDO Maske7	MT	Out 3	KDO Maske3		
Out 4	PB Maz&Jingle	D16	KDO Maske8	VT	Out 4	KDO Maske4		
Out 5	PB FX	D17	KDO Maske9	Musik Listen	Out 5	KDO Maske5		
Out 6	Stimmverzerrer (post fader)	D18	KDO Maske10	HF	Out 6	KDO Maske6		
Out 7	Click < Mon	D19		EVS Moni	Out 7	KDO Maske7		
Out 8	Click2 < Mon	D20	KDO Maske11	Steady Sprechen	Out 8	KDO Maske8		
Out 9	MusikmixL < PA/Mon	D21		EVS Red Moni	Out 9	KDO Maske9		
Out 10	MusikmixR < PA/Mon	D22		VC<ICS (post fader)	Out 10	KDO Maske10		
Out 11	AtmoL < Mon	D23		N-1<>KS	Out 11	IEM Mod		
Out 12	AtmoR < Mon	D24		VocalCoach < Masken	Out 12	IEM Rateteam1		
Out 13	KDO Maske11	A1-4/14-16/6		Sip 1-8	Out 13	IEM Rateteam2		
Out 14	TC < Licht/MS				Out 14	IEM Rateteam3		
Out 15	KDO Steady				Out 15	IEM Rateteam4		
Out 16	PGM	A9-13		Freespeak	Out 16	IEM RED Mod		

11.2 MAZ Belegungsplan

PROD-DOKU KUNDE

Regie 11

THE MASKED SINGER

D - 2024 SHOW 6

Live

Bildng: Manfred Hutmacher

Tonng: Constantin Gabrysch

AZ: 21.11.2024

MAZ-Operator: Volker Danckwortt



AZ FORMAT

Trägermaterial:	Container:	Codec/Format:	Bitrate:	Farb-Div.-Signal:	Auflösung:	Abtast-format:	Seiten-verhältnis:	Standard:	Eingangs-signal:	Timecode:
Optical Disc	MXF OP1A	XDCamHD422	50 Mbit/s	4:2:2	1080	i/25	16:9	HD	HDSDI	LTC/MITC
Audiocodec: PCM		Audiobitrate: 24 bit		Aufzeichnung-Timecode: Realtimecode						
Trägermaterial:	Container:	Codec/Format:	Bitrate:	Farb-Div.-Signal:	Auflösung:	Abtast-format:	Seiten-verhältnis:	Standard:	Eingangs-signal:	Timecode:
Hard Disc	MXF OPAtom	DNxHD	120 Mbit/s	4:2:2	1920 x1080	i/25	16:9	HD	HDSDI	LTC
Audiocodec: PCM		Audiobitrate: 24 bit		Aufzeichnung-Timecode: Realtimecode						

BELEGUNG

Maschine	VIDEO	AUDIO							
		Ch.1	Ch.2	Ch.3	Ch.4	Ch.5	Ch.6	Ch.7	Ch.8
Ingest 4	PGM	PGM	PGM	ATMO	ATMO	LINE	LINE	MOD	Jury Grp
Ingest 4	CLEANFEED	PGM	PGM	Talk Grp	Jury Grp	PGM-LINE	PGM-LINE	MUSIK L	MUSIK R
Ingest 4	KAMERA 7	MOD	JURY 1	JURY 2	JURY 3	AfterShow	OFF Spr	ATMO	ATMO
Ingest 4	KAMERA 2	MOD	JURY 1	JURY 2	JURY 3	AfterShow	OFF Spr	ATMO	ATMO
Ingest 3	KAMERA 9	PGM	PGM						
XDCAM Rec	PGM	PGM	PGM	ATMO	ATMO	LINE	LINE	MOD	Jury Grp
SendLeit HAV	PGM	PGM	PGM	ATMO	ATMO	LINE	LINE	MOD	Jury Grp
SendLeit HAV	CLEANFEED	PGM	PGM	Talk Grp	Jury Grp	PGM-LINE	PGM-LINE	MUSIK L	MUSIK R
MP4 + TC	PGM	PGM	PGM						
XDCAM Play	Havarieplayer								

TIMECODELISTE

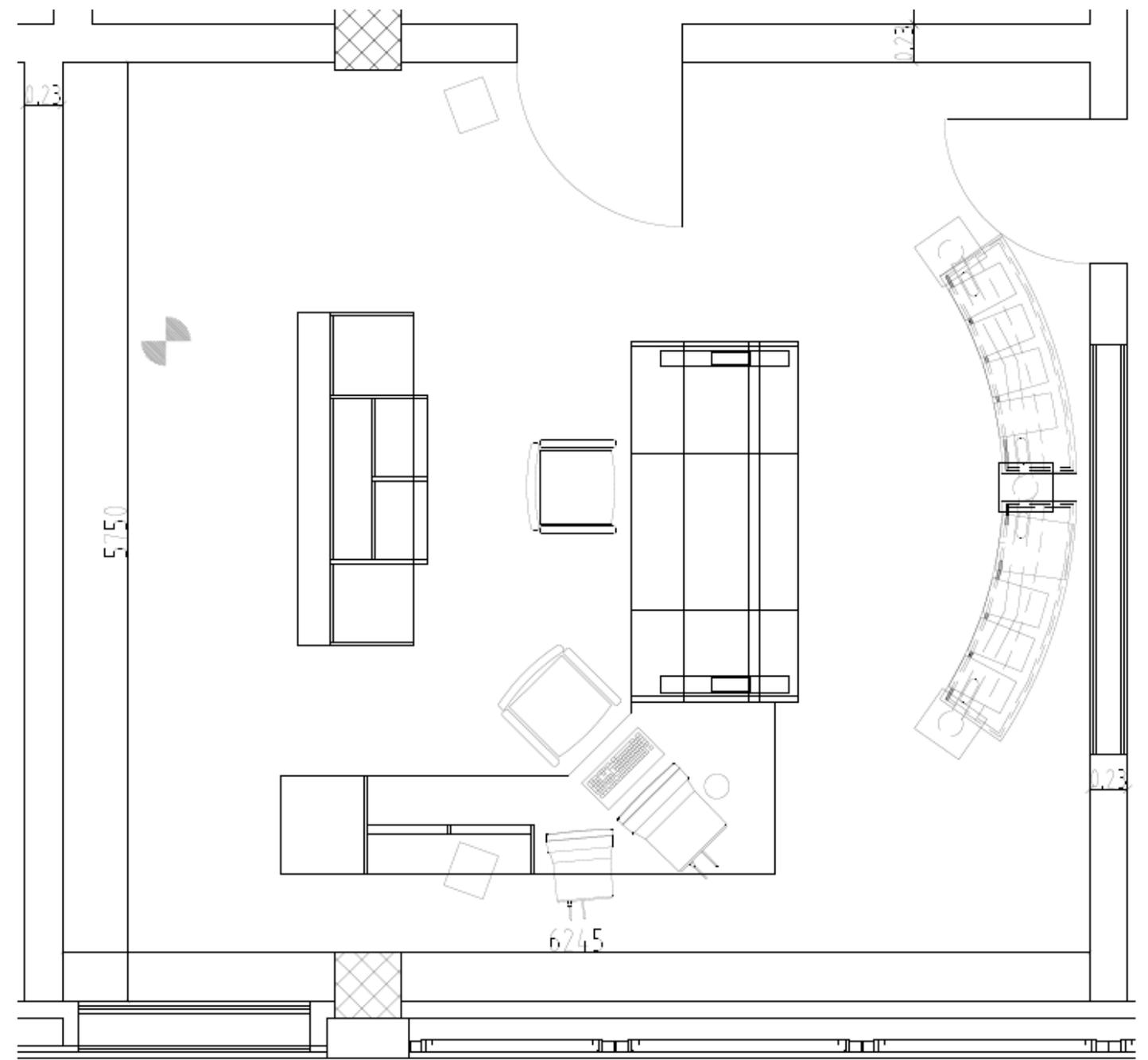
CLIP	BEITRAG	TC IN	TC OUT	DUR
1	Disc 1 Teaser			
	Disc 1 Block 1			
2	Disc 2 Block 2			
3	Disc 2 Block 3			
4	Disc 2 Block 4			
5	Disc 3 Block 5			
6	Disc 3 Block 6			
7	Disc 3 Block 7			
	Disc 3 Block 8			

... weitere Timecode's siehe Medienkarte vom PGM (1.Disc)

BEMERKUNG

RAID7 LWL1 für Post 192.168.2.150

11.3 Grundriss Tonregie 11



11.4 Transkription Interview mit Emilio Bonta (Toningenieur) – Zukünftige Relevanz von Dolby Atmos im Fernsehbereich

Frage:

Wie lange arbeitest Du bereits in dieser Firma und welche Tätigkeit übst du derzeit aus?

Emilio:

Ich habe im Jahr 1993 als Tontechniker in der MMC angefangen und habe dann irgendwann auch in der Beschallung gearbeitet und seit 1999 bin ich Toningenieur. Das sind also über 25 Jahre schon.

Frage:

In meiner Zeit hier habe ich bisher keine einzige Produktion im 5.1 Surround Format mitbekommen. Wie oft kommt das vor, dass du eine Show in 5.1 mischst?

Emilio:

Ja, es kommt tatsächlich nicht so häufig vor. Also es gab natürlich in der Vergangenheit schon einige Sendungen in 5.1 und zum Beispiel beim ZDF ist das auch schon der Standard. Beim Audio ist der Anspruch aber eher nicht, immer wieder neue Formate zu entwickeln und bessere Qualität oder ein besseres Erlebnis zu liefern, weil die bestehenden Formate auch ausreichend sind. Im Vergleich dazu bekommt man ja mit, dass es im Bildbereich von HD auf UHD und dann noch auf HDR geht. In anderen Teilen der Welt gibt es sogar noch höhere Qualitätsstandards. Beim Audio hat es im Fernsehen bei 5.1 aufgehört, sich zu verbessern und selbst das hat sich bis heute nicht einheitlich durchgesetzt, obwohl es am Anfang danach aussah.

Frage:

Siehst du also eine zukünftige Relevanz von Dolby Atmos für Fernsehen und Entertainment?

Emilio:

Ja eigentlich schon, wenn man es richtig einsetzt könnte es schon funktionieren. Dolby Atmos gibt es ja schon in vielen anderen Bereichen, vor allem bei Kinofilmen und im Gaming und VR Bereich, da macht es auch Spaß und ist sinnvoll. Da kann der Fernsehbereich bestimmt auch nachziehen. Dolby Atmos wird als das Gegenstück zu UHD HDR gesehen, da finde ich es auch angemessen, dem Zuschauer auch im Tonbereich ein immersives Erlebnis zu ermöglichen. 3D-Audio ist ja grundsätzlich was total Interessantes und ich glaube, was vor allem in Zukunft sogar noch eher eine Rolle spielen wird als das immersive Hörerlebnis, ist die Personalisierbarkeit bei den Audiospuren. Das ist für viele wahrscheinlich attraktiv, wenn man sich beispielsweise den Kommentator bei einer Sportübertragung aussuchen kann. Also langfristig denke ich schon, dass es sich im Fernsehbereich etablieren könnte. Wobei der Einsatz von dynamischen Audioobjekten im Raum im Fernsehen nur begrenzt Sinn ergibt.

Frage:

Welche Herausforderungen tauchen auf, wenn man Dolby Atmos langfristig als Möglichkeit im Fernsehbereich anbieten möchte?

Emilio:

Also aktuell ist noch nicht mal 5.1 Surround einheitlich im Fernsehen verankert, den meisten reicht schon Stereo völlig aus. Dolby Atmos scheitert schon mal daran, dass einige Sender dieses Format momentan noch nicht mal senden können. Viele Produktionsfirmen sind auch nicht bereit, mehr Geld für etwas zu zahlen, was sich nicht auch in den Einschaltquoten zeigt. Den Zuschauer interessiert das Audioformat nicht wirklich, da ist es auch nicht wichtig, ob man Stereo hört oder in Dolby Atmos umhüllt wird. Außerdem haben die wenigsten Konsumer ein korrekt eingemessenes Lautsprechersystem zuhause, darüber denkt man als Normalverbraucher auch nicht nach. Soundbars können ja

mittlerweile auch Dolby Atmos mit Reflexionen simulieren, aber das ist dann schon noch mal was anderes. Damit sich Produktionen in Dolby Atmos lohnen, müssten mehr Leute zuhause ein System haben, dass 3D-Audio auch abspielen kann. Das könnte langfristig schon funktionieren, das ist auf jeden Fall die Zukunft, da muss sich das Fernsehen eben auch mitentwickeln und mit den anderen mitziehen.