

Bachelorarbeit im Studiengang Audiovisuelle Medien

**Workflow-Analyse von immersiven
Audioproduktionen mit Produktionsmischpulten
für moderne Standards
wie MPEG-H und Dolby Atmos**

vorgelegt von

Jan Bieletzki

Matr.-Nr.: 30379

am 06. April 2019

an der Hochschule der Medien Stuttgart
zur Erlangung des akademischen Grades
Bachelor of Engineering

Erstprüfer: Prof. Oliver Curdt

Zweitprüfer: Dipl.-Tonmeister Christian Struck

Ehrenwörtliche Erklärung

„Hiermit versichere ich, Jan Bieletzki, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „*Workflow-Analyse von immersiven Audioproduktionen mit Produktionsmischpulten für moderne Standards wie MPEG-H und Dolby Atmos*“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. 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 (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.“

Stuttgart, den 06.04.2019

Jan Bieletzki

Kurzfassung

Im Zuge der Verbreitung von Immersive Audio und der Etablierung von Next Generation Audio Standards, müssen sowohl die Produktionsworkflows als auch die Tools angepasst werden. Produktionsumgebungen benötigen dafür zusätzlich externe Authoring- und Monitoring-Systeme.

Die vorliegende Bachelorarbeit beschäftigt sich mit der Analyse von immersiven Audioproduktionen mit Produktionsmischpulten im Live-Broadcast. Dabei wird berücksichtigt, dass auch objektbasierte Produktionsansätze mit Metadaten unterstützt werden müssen. Ziel ist es, den Workflow für MPEG-H und Dolby Atmos darzustellen und Anforderungen an moderne Mischpulte abzuleiten. Berichte, Analysen und Signalfluss-Diagramme von Testimplementierungen sowie Produktionen unter Realbedingungen werden dafür ausgewertet.

Aus der Analyse geht hervor, dass immersive Audioproduktionen mit geringem Mehraufwand durchführbar sind. Die Workflows für beide vorgestellte Systeme sind ähnlich, allerdings nicht kompatibel. Aufgrund dessen ist es empfehlenswert, dass Mischpulthersteller einerseits möglichst unabhängig von modernen Standards bleiben. Gleichzeitig sollte die Bedienbarkeit der Authoring Units von Produktionsmischpulten aus verbessert werden.

Abstract

With the spread of immersive audio and the establishment of next generation audio standards, both production workflows and tools need to be adapted. Production environments require additional external authoring and monitoring systems.

This bachelor thesis deals with the analysis of immersive audio productions with production consoles in live broadcasts. This takes into account that object-based production approaches must be supported with metadata. The aim is to present the workflow for MPEG-H and Dolby Atmos and to derive requirements for modern consoles. Reports, analyses and signal flow diagrams of test implementations as well as productions under real conditions are evaluated.

The analysis shows that immersive audio productions can be carried out with little additional effort. Both workflows of the presented systems appear to be similar, but not compatible. Therefore, it is recommended that mixing console manufacturers remain as independent as possible from modern standards. At the same time, the usability of the authoring units from production mixing consoles should be improved.

Danksagung

An dieser Stelle möchte ich allen Personen danken, die mich im Verlauf meiner Bachelorarbeit unterstützt haben. Besonders herzlich bedanken möchte ich mich bei Prof. Oliver Curdt und Christian Struck für die Betreuung meiner Thesis.

Ein herzlicher Dank geht an die Firma Lawo AG für die Bereitstellung der Räumlichkeiten und die Unterstützung in jeglicher Form. Weiterhin möchte ich mich beim Fraunhofer Institut für Integrierte Schaltungen IIS in Erlangen bedanken, ganz besonders bei Christian Simon und Yannik Grewe, für die wertvollen Anregungen und die zur Verfügung gestellten Materialien. Ein großes Dankeschön geht auch an Tobias Habann, Prof. Felix Krückels, Lucas Zwicker und Tom Ammermann. Ohne sie wäre die Arbeit in dieser Form nicht möglich gewesen.

Ganz besonders möchte ich mich an dieser Stelle bei meiner Familie und meinen Freunden für die Unterstützung und den Rückhalt während meines gesamten Studiums bedanken.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung	ii
Kurzfassung	iii
Abstract.....	iii
Danksagung	iv
Inhaltsverzeichnis.....	v
Abbildungsverzeichnis	vii
Abkürzungsverzeichnis	viii
1. Einleitung	1
2. Immersive Audio.....	3
2.1 Definition Immersive Audio	3
2.2 Grundlagen räumliches Hören.....	4
2.3 Entwicklung des Mehrkanaltones.....	5
2.4 Anwendungsbereiche und Perspektiven	6
3. Audio-Wiedergabeverfahren	7
3.1 Grundkonzepte	7
3.1.1 Kanalbasiert.....	8
3.1.2 Objektbasiert.....	8
3.1.3 Szenenbasiert.....	9
3.1.4 Hybrider Ansatz.....	10
3.2 3D-Audioformate	10
3.2.1 Auro-3D	11
3.2.2 Dolby Atmos.....	12
3.2.3 DTS:X	13
3.2.4 NHK 22.2	13
3.3 Alternative Wiedergabekonzepte.....	14
3.4 Problematik bei unterschiedlichen Audioformaten.....	15
4. MPEG-H.....	16
4.1 Grundlagen MPEG-H 3D Audio.....	16
4.1.1 Metadaten	17
4.1.2 Rendering	19
4.1.3 Lautheit und Dynamik	19
4.2 MPEG-H Audio Stream.....	21
4.3 MPEG-H Produktionstools.....	21
4.3.1 MPEG-H Audio Monitoring and Authoring Unit	22
4.3.2 Tools für die Postproduktion.....	23
5. Dolby Atmos.....	24

5.1 Grundlagen.....	25
5.2 Codecs für einen Dolby Atmos Workflow	26
5.2.1 Dolby Digital Plus with Dolby Atmos	26
5.2.2 AC-4.....	26
5.2.3 Dolby ED2.....	27
5.3 Dolby Atmos Produktionstools.....	28
5.3.1 DP580.....	28
5.3.2 DP590.....	29
5.3.3 DP591.....	30
6. Workflow-Analyse	31
6.1 Broadcast-Workflow im Allgemeinen	32
6.2 Workflow für MPEG-H.....	33
6.2.1 Signalfluss von MPEG-H im Live-Broadcast	33
6.2.2 Anpassung der Produktionsumgebung für MPEG-H	35
6.2.3 French Open 2018	36
6.3 Workflow für Dolby Atmos	37
6.3.1 Signalfluss von Dolby Atmos im Live-Broadcast.....	38
6.3.2 Anpassung der Produktionsumgebung für Dolby Atmos	39
6.4 Ergebnisse der Analyse.....	41
7. Anforderungen an die Produktionsmischpulte.....	41
7.1 3D-Panning und Bussystem.....	42
7.2 Authoring und Monitoring.....	44
7.2.1 Bedienung der AMAU vom Produktionsmischpult aus	44
7.2.2 Integration der AMAU in das Produktionsmischpult.....	46
7.3 Weitere Anforderungen.....	48
8. Fazit	49
Literaturverzeichnis	ix

Abbildungsverzeichnis

Abb. 1 Kopfbezogenes Koordinatensystem	4
Abb. 2 Kanalbasierte Produktion und Wiedergabe	8
Abb. 3 Objektbasierte Produktion und Wiedergabe	9
Abb. 4 Auro 9.1	11
Abb. 5 Dolby Atmos 5.1.4	12
Abb. 6 NHK 22.2 Lautsprechersystem	14
Abb. 7 Aufbau einer Audio-Szene	18
Abb. 8 Lautheitskompensation bei Anhebung der Dialoglautstärke	20
Abb. 9 Webapplikation der MMA von Jünger Audio	22
Abb. 10 Dolby Professional Reference Decoder DP580	28
Abb. 11 Webapplikation des DP590	29
Abb. 12 Dolby Professional Audio Encoder DP591	31
Abb. 13 Allgemeiner Broadcast-Workflow	32
Abb. 14 Blockdiagramm MPEG-H Testimplementierung	34
Abb. 15 Einbindung eines Produktionsmischpultes für den MPEG-H Audio Broadcast	35
Abb. 16 Blockdiagramm eines Dolby Atmos Broadcast-Workflows	38
Abb. 17 Signalfluss in einem Ü-Wagen der Firma Telegenic	40
Abb. 18 User Interface des DP590 an einer Lawo Konsole	46

Abkürzungsverzeichnis

ADM	Audio Definition Model
AMAU	Audio Monitoring and Authoring Unit
AMS	Authoring and Monitoring System von Linear Acoustic
AR	Augmented Reality
ATSC	Advanced Television Systems Committee
AVC	Advanced Video Coding
DAW	Digital Audio Workstation
DRC	Dynamic Range Control
DVB	Digital Video Broadcasting
EBU	European Broadcasting Union
ETSI	European Telecommunications Standards Institute
GPIO	General Purpose Input/ Output
HDR	High Dynamic Range
ISOBMFF	ISO Base Media File Format
LC	Low Complexity Profile
LFE	Low Frequency Effects Kanal
LIME	Lawo Immersive Mixing Engine
MADI	Multichannel Audio Digital Interface
MHAPI	MPEG-H Authoring Plug-in
MHAS	MPEG-H Audio Stream
MHAT	MPEG-H Authoring Tool
MMA	Multichannel Monitoring and Authoring System von Jünger Audio
MPEG	Moving Picture Experts Group
NGA	Next Generation Audio
PCM	Puls-Code-Modulation
SDI	Serial Digital Interface
STB	Set-Top-Box
UHD	Ultra High Definition
USAD	Unified Speech and Audio Coding
VR	Virtual Reality

1. Einleitung

Mit der Einführung von Ultra High Definition (UHD) im TV, möchte man dem Nutzer nicht nur hochwertigere Bilder, sondern auch ein besseres Hörerlebnis liefern. In diesem Zuge wurden Audio-Konzepte entwickelt, die unter dem Begriff *Next Generation Audio* (NGA) zusammengefasst und standardisiert wurden. Audio-Codecs für diese NGA-Systeme sind zum einen effizienter als die bisherigen im Broadcast, haben zum anderen aber auch neue Funktionen integriert, wie die Unterstützung von Immersive Audio.¹ Im Kino haben sich immersive Audioproduktionen mit Formaten wie Dolby Atmos und Auro-3D mancherorts bereits etabliert. Dagegen steht man im Live-Broadcast noch am Anfang. In den letzten Jahren wurde daher viel Entwicklungs- und Forschungsarbeit betrieben sowie zahlreiche Testimplementierungen durchgeführt, um die vorhandene Infrastruktur zu überprüfen. Die Fußball-Weltmeisterschaft 2018 in Russland, gehört beispielsweise mit zu den ersten großen Live-Übertragungen, bei der den internationalen Sendeanstalten ein immersiver Audio-Mix auch für die Ausstrahlung zur Verfügung stand. Tatsächlich wurde diese Mischung allerdings nur von sehr wenigen Sendern übertragen und kam dort überwiegend für Testzwecke zum Einsatz.²

Diese Situation ist ein Beispiel dafür, dass sich Immersive Audio bei Liveübertragungen noch in einer frühen Phase der Integration befindet. Es lohnt sich deshalb gerade jetzt, die bisherigen Ergebnisse zusammenzutragen und einen umfassenden Blick auf die Produktions- und Übertragungskette zu werfen. Im Zuge der Kooperation mit der Firma Lawo AG, einem Hersteller der unter anderem für seine Digitalmischpulte im Live-Broadcast bekannt ist, werden bei der vorliegenden Bachelorarbeit auch die Schnittstellen mit Produktionsmischpulten behandelt.

Der Titel der Arbeit *„Workflow-Analyse von immersiven Audioproduktionen mit Produktionsmischpulten für moderne Standards wie MPEG-H und Dolby Atmos“* mag auf den ersten Blick zunächst verwundern. Schließlich handelt es sich bei MPEG-H um einen multimedialen Standard inklusive 3D-Audio-Codec für den Einsatz in Echtzeit-Anwendungen³, wohingegen Dolby Atmos ursprünglich als 3D-Audioformat aus dem Kino bekannt ist. Dolby verwendet den Begriff *Dolby Atmos* jedoch für die gesamte Produktpalette mit immersiven Hörerlebnis und bezieht sich dementsprechend auch auf die 3D-Audioproduktion von Dolbys Produktionstools und Codecs im Live-Broadcast. Der Vergleich mit Dolby-AC4 würde demnach nicht den gesamten Produktionsworkflow abdecken. Eine Gegenüberstellung von MPEG-H und Dolby Atmos ist dagegen sinnvoll.

¹ Vgl. Deutsche TV-Plattform (2016)

² Vgl. Slavik (2018)

³ Vgl. Fleischmann (2017)

Ziel ist die Beantwortung der Frage, wie aktuelle Workflows von immersiven Audioproduktionen mit Produktionsmischpulten für diese beiden Standards aussehen. Dabei soll herausgefunden werden, welche Rolle das Mischpult einnimmt, wenn kein internes Monitoring von Immersive Audio und kein Authoring der NGA-Funktionen unterstützt wird. Wie wird diese Problematik aktuell gelöst? Welche Tools werden dafür benötigt? Hierbei ist genau zu klären, wie weit die Aufgabenbereiche der beteiligten Komponenten jeweils reichen und ob es möglicherweise Einschränkungen oder Überschneidungen an den Schnittstellen im Produktionsworkflow gibt. Der Fokus wurde hierbei explizit auf die Produktion und Übertragung im Live-Broadcast gelegt. Diese Eingrenzung der Thematik lag zum einen an dem begrenzten Umfang einer Bachelorarbeit. Des Weiteren ist die Post- und Offlineproduktion von Immersive Audio, beispielsweise für Kinofilme, mit Echtzeit-Anwendungen nur bedingt zu vergleichen. Produktionsmischpulte spielen dort eine untergeordnete Rolle und der Workflow profitiert meist von DAW-Controllern.

Zur Klärung der forschungsleitenden Fragen, erfolgt eine empirische Datenerhebung, bei der vorhandene Berichte von Testimplementierungen, Datenblätter sowie verschiedenste Internetquellen systematisch zusammengetragen und ausgewertet werden. Zunächst werden in Kapitel 2 grundlegende Informationen zum Thema *Immersive Audio* vermittelt. Es folgt in Kapitel 3 eine Zusammenfassung der Audio-Grundkonzepte bezüglich Produktion, Übertragung und Wiedergabe. Im Hinblick auf die Wiedergabeseite beim Verbraucher, werden anschließend etablierte 3D-Audioformate vorgestellt. Gleichzeitig wird aber auch die Kompatibilitätsproblematik der Formate thematisiert. Des Weiteren werden alternative Wiedergabekonzepte für Immersive Audio präsentiert. In den darauffolgenden Kapiteln 4 und 5 wird intensiv auf die neuen Standards MPEG-H und Dolby Atmos eingegangen. Hier werden auch die wesentlichen Produktionstools erklärt. In der Workflow-Analyse (Kapitel 6) wird zunächst der Signalfluss von Broadcasts im Allgemeinen dargestellt. Weitergehend wird im Detail auf die Produktions- und Übertragungskette der beiden Standards eingegangen. Zusammen mit den Informationen aus den vorherigen Kapiteln, sollen die Funktionen und Aufgabenbereiche der jeweiligen Produktionstools voneinander abgrenzt werden. Neben Testimplementierungen, beispielsweise von der MPEG-H TV Audio Alliance, werden auch Live-Broadcasts, wie die French Open 2018 oder die BT-Sports Produktionen, behandelt. Die Ergebnisse werden ausgewertet und dienen mit als Grundlage für die in Kapitel 7 definierten Anforderungen an moderne Produktionsmischpulte.

2. Immersive Audio

Dieses Kapitel dient als Einstieg in die Thematik von immersiven Audioproduktionen. Zunächst sollte ein einheitliches Verständnis über die Bedeutung der Begriffe *Immersion* und *Immersive Audio* hergestellt werden, da im Allgemeinen sehr viele Definitionen und Interpretationen kursieren. Auch die Wortherkunft wird hierbei geklärt. Anschließend folgt ein kurzer Überblick über die Grundlagen des räumlichen Hörens und der Lokalisation von Schallquellen. Daraus wird ersichtlich, wie die menschliche auditive Wahrnehmung funktioniert, weshalb die dreidimensionale Audiowiedergabe ein paar Vorteile mit sich bringt. Neben den wesentlichen Entwicklungsschritten der Mehrkanal-Tontechnik hin zu Immersive Audio, wird auch ein Überblick über die Anwendungsbereiche und möglichen Perspektiven dieser aktuellen Thematik gegeben.

2.1 Definition Immersive Audio

Der Begriff *Immersion* entstammt ursprünglich dem lateinischen Wort *immersio* und bedeutet Eintauchen. In der Medientheorie liefert Bilandzic dafür eine recht allgemeine Definition. Demnach lässt Immersion *„Rezipientinnen und Rezipienten völlig in die medial vermittelte Welt eintauchen und die Gegebenheiten und Erlebnisse unmittelbar und direkt nachvollziehen, fast so, als würden sie diese selbst in der realen Welt erleben“*.⁴

Übertragen auf den Bereich Audio, entspricht Immersion dem natürlichen Hören, bei dem Schall aus allen Richtungen wahrgenommen werden kann.⁵ Häufig wird in diesem Zusammenhang neben Immersive Audio auch von *3D-Audio* gesprochen. Wilfried van Baelen, der Entwickler der Auro-3D-Formate, definierte *Immersive Sound* erstmals im Jahr 2010 als Oberbegriff für Mehrkanalformate mit zusätzlichen Höhenlautsprechern.⁶ Mittlerweile wird dieser Begriff universal im Zusammenhang mit umhüllendem Klang verwendet. Nach Lasse Nipkow, Gregor Zielinsky und Tom Ammermann, können Schallquellen, die über ein solches Lautsprechersystem wiedergegeben werden, vor allem deutlich natürlicher klingen und besser lokalisiert werden. Deshalb sei 3D-Audio auch *„wesentlich mehr als Surround with Height, wie es manchmal auch genannt wird.“*⁷

Zusammenfassend betrachtet, handelt es sich bei Immersive Audio also um eine Audio-Wiedergabetechnik, bei der Schallquellen aus allen Richtungen wiedergegeben werden können. Das natürliche Hören (siehe Kapitel 2.2) des Menschen funktioniert ebenfalls dreidimensional und kann Schall aus allen Richtungen lokalisieren. Dem Hörer fällt es dadurch möglicherweise leichter in die Audio-Szenerie einzutauchen, da die Grenze zwischen realem und künstlich erzeugtem Raum kleiner wird.

⁴ Bilandzic (2014), S. 273

⁵ Vgl. Roginska, Geluso (2018), S. 1

⁶ Vgl. Van Baelen (2016), S. 19

⁷ Nipkow, Zielinsky, Ammermann (2019), S. 22

2.2 Grundlagen räumliches Hören

Das natürliche Hören basiert auf verschiedenen Lokalisationsmechanismen, mithilfe derer das menschliche Gehirn in der Lage ist, die Richtung und Entfernung einer Schallquelle zu bestimmen.

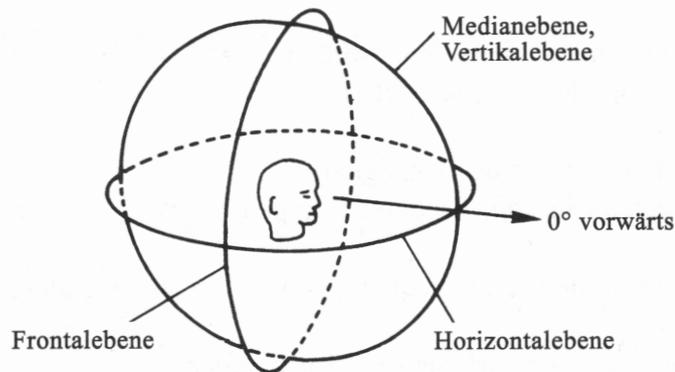


Abb. 1 Kopfbezogenes Koordinatensystem⁸

In der Horizontalebene (siehe Abb. 1) lässt sich die Richtung im Wesentlichen über die Pegel- und Laufzeitunterschiede zwischen dem linken und rechten Ohr bestimmen. Ein Signal von links kommt nicht nur zeitlich später, sondern durch die frequenzabhängige Abschattung des Kopfes auch mit geringerem Pegel am rechten Ohr an. Problematisch sind Audiosignale, welche direkt von vorne oder von hinten kommen und an beiden Ohren sowohl gleiche Pegel, als auch gleiche Laufzeiten aufweisen. Kompensieren lässt sich das beispielsweise durch leichte Drehbewegungen des Kopfes.⁹

Unterschiedliche Positionen von Schallquellen auf der Medianebene führen zu Veränderungen im Frequenzgang. Diese sogenannten *richtungsbestimmenden Frequenzbänder* wurden erstmals 1983 von Jens Blauert nachgewiesen. Demnach tendieren Signale bei deutlichen Ausprägungen im Bereich 250-500Hz und 2-7kHz zu einer Lokalisation von vorne. Betonungen bei 1kHz oder 12-16kHz werden eher von hinten und bei 8kHz verstärkt von oben wahrgenommen. Zusätzlich dazu sind Reflexionen an den Ohrmuscheln für schmalbandige Frequenzeinbrüche verantwortlich, die wiederum zur Bestimmung der Höhe beitragen können.¹⁰

Die Entfernungswahrnehmung einer Schallquelle hängt dagegen ebenfalls von mehreren Aspekten ab. Zum einen lassen die frühen Erstreflexionen Rückschlüsse auf die Position einer Schallquelle innerhalb eines Raumes zu. Größere Entfernungen führen aufgrund der Luftabsorption zu einem Pegelabfall in den hohen Frequenzen. Des Weiteren gibt auch

⁸ Bildquelle: Dickreiter et al. (2014), S. 128

⁹ Vgl. Friesecke (2014), S. 134-135

¹⁰ Vgl. ebd., S. 135-136

das Verhältnis von Direkt- zu Diffusschall Aufschluss über die Entfernung einer Schallquelle.¹¹

2.3 Entwicklung des Mehrkanaltones

Heutige immersive Audio-Formate sind Teil einer kontinuierlichen Entwicklung der Mehrkanaltontechnik. Sie erweitern die bisherigen Surround-Formate um zusätzliche Audiokanäle für die Höhenwiedergabe. Deshalb ist für ein Grundverständnis von Immersive Audio auch das Wissen um ebendiese bisherigen Formate unerlässlich. Gleichzeitig müssen die 3D-Audioformate stets eine bestmögliche Kompatibilität zu den bisherigen Formaten gewährleisten. In diesem Zuge ist die Monokompatibilität immer noch ein wichtiger Punkt, der bei immersiven Audioproduktionen beachtet werden muss. Beim Downmixing von Höhen- und Surroundkanälen besteht immer die Gefahr von Auslöschungen.

Bis in die 1950er Jahre waren Audio-Aufnahmen und deren Wiedergabe überwiegend monophon, also einkanalig (1.0).¹² Eine räumliche Abbildung lässt sich damit nicht wirklich erzeugen. Dafür ist die Lokalisation der Schallquelle unabhängig von der genauen Abhörposition stabil.¹³ Auch wenn monophone Tonträger schon lange von stereophonen (2.0) verdrängt wurden, müssen die Mehrkanalformate weiterhin monokompatibel sein.

Stereo ermöglicht mit zwei Lautsprechern bereits eine bessere räumliche Wiedergabe da Audioquellen auch zwischen den beiden Lautsprechern verteilt werden können. Eine korrekte Lokalisierung dieser Phantomschallquellen ist allerdings nur innerhalb des Sweet-Spots möglich.¹⁴ Aus diesem Grund wurde gerade im Kinobereich das 2.0-System um einen zusätzlichen Center-Kanal erweitert, damit auch weit außen sitzende Zuschauer vorhandene Phantomschallquellen in der Mitte der Leinwand lokalisieren können.¹⁵ Das Stereoformat ist bis heute in vielen Anwendungsbereichen Standard (z.B. in der Musikindustrie).

Weitere Entwicklungen stellten beispielsweise ein zusätzlicher LFE-Kanal oder Surround-Lautsprecher für Effekte und mehr räumliche Umhüllung dar. Anfänge der Surround-Tonformate gab es bereits um 1940 mit dem Film *Fantasia*, bei dem verschiedene Mehrkanalformate getestet wurden. Das 5.1-Format, wie man es kennt (Left, Right, Center, LFE, Left-Surround, Right-Surround), hat sich spätestens ab 1987 als Standard im Bereich Filmtone etabliert.¹⁶ Bei diesen Surround-Formaten, trifft man häufig auch das erweiterte 7.1 an. Grund dafür ist der vergleichsweise große Sprung zwischen vorderen

¹¹ Vgl. Friesecke (2014), S. 143-144

¹² Vgl. Scholz (2015), S. 4

¹³ Vgl. Friesecke (2014), S. 785

¹⁴ Vgl. Scholz (2015), S. 4

¹⁵ Vgl. ebd., S. 227-228

¹⁶ Vgl. Wikipedia (2018) 5.1

und hinteren Lautsprechern. Zwei zusätzliche Lautsprecher an den Seiten sorgen für eine homogenere Wiedergabe.¹⁷ Laut Friesecke kann nur „eine hohe Anzahl an Surround-Lautsprechern [...] ein wirklich diffuses, sitzplatzunabhängiges Surround-Signal produzieren.“¹⁸

Die Erweiterung der Surround-Formate um zusätzliche Höhenkanäle ermöglicht ein komplett umhüllendes Wiedergabesystem. Den Anfang dieser Entwicklungsgeschichte, bildete das 2.2.2-Format von Werner Dabringhaus, der speziell für Musikaufnahmen die Funktion des Center- und LFE-Kanal im 5.1-Format abwandelte, und durch zwei Höhenkanäle über dem linken und rechten Lautsprecher ersetzte.¹⁹ Mittlerweile gibt es eine Vielzahl an unterschiedlichsten 3D-Audio-Formaten bis hin zu 22.2, dem derzeitigen Produktionsstandard der japanischen Rundfunkgesellschaft NHK. Eine Studie des Fraunhofer Institutes zeigt, dass bereits mit vier zusätzlichen Höhenlautsprecher (5.1.4)²⁰ eine Verbesserung der Audioqualität im Vergleich zu einem normalen Surround-Setup (5.1) erzielt werden kann.²¹

2.4 Anwendungsbereiche und Perspektiven

„Immersive sound can give the listener an experience of being there through sound. Compared to vision, sound provides a fully immersive experience and can be perceived from all directions simultaneously.“²²

(Agnieszka Roginska und Paul Geluso, New York University)

Demnach kann Immersive Audio dem Zuhörer das Gefühl geben, in eine Klangwelt einzutauchen. Im Gegensatz zu der rein bildlichen Wahrnehmung, soll ein vollständig immersives Erlebnis möglich sein.²³ Roginska und Gelusa beschreiben damit einen Aspekt, der für zahlreiche Anwendungsbereiche einen interessanten Mehrwert bieten kann. Dieses Gefühl durch Klang zu vermitteln, ermöglicht es beispielsweise in Rauminstallationen bei Events oder in Museen mit neuen dramaturgischen Möglichkeiten Inhalte unmittelbar erfahrbar zu machen und Storytelling in eine neue Dimension zu führen. Gleiches gilt auch für Virtual Reality (VR) - oder Augmented Reality (AR) -Games.

In der Beschallung gibt es bereits Systeme, welche die Vorteile von Immersive Audio nutzen um das Publikum mit einem homogeneren Klang unabhängig von der Abhörposition zu versorgen. Die Technologie L-ISA von L-Acoustics beispielsweise,

¹⁷ Vgl. Theile (2014), S. 311

¹⁸ Friesecke (2014), S.790

¹⁹ Vgl. Theile (2014), S. 311-312

²⁰ Sofern es sich nicht um feste Bezeichnungen handelt (z.B. NHK 22.2, Auro-Formate), richtet sich die Bezeichnung der Lautsprecherformate in dieser Bachelorarbeit nach folgendem System: a.b.c, a = Anzahl der Lautsprecher auf Ohrhöhe, b = Anzahl der LFE Kanäle, c = Anzahl der Höhen-Lautsprecher

²¹ Vgl. Silzle et al. (2011), S. 340

²² Roginska, Geluso (2018), S. 1

²³ Vgl. ebd.

verspricht ein natürlicheres Klangbild sowie eine präzisere Lokalisierung über den gesamten Beschallungsbereich.²⁴ Beispiele für 3D-Audio-Systeme finden sich auch in der Automobilindustrie. Für den Porsche Panamera gibt es beispielsweise ein Sound-System mit 21 Lautsprechern von Burmester und dem Upmixing-Algorithmus AuroMatic von Auro-Technologies, welche zusammen ein 3D-Audio-Erlebnis im Auto versprechen.²⁵

Ganz neue Klangmöglichkeiten ergeben sich bei Filmen für Kino, Heimkino und Streaming. In diesem Bereich sind Technologien von Dolby Laboratories, Inc., einem der Marktführer digitaler Mehrkanal-Tonformate, stark vertreten. Seit 31. August 2018 bietet beispielsweise Prime Video, der Video-on-Demand Service von Amazon, auch Filme und Serien zum Streaming in Dolby Atmos an.²⁶

Gerade auch im Live-Broadcast hat man erkannt, welche Möglichkeiten Immersive Audio bietet. Den Zuschauern mehr das Gefühl zu vermitteln dabei zu sein und Programminhalte mit Interaktionsfunktionen anzubieten, hat viel Potenzial. Zudem kann unmittelbar ein großes Publikum erreicht werden. Anwendungsszenarien gibt es dafür genügend. Bisherige Testimplementierungen fanden überwiegend an großen Sportveranstaltungen statt. Beispiel dafür ist unter anderem die Fußball-Weltmeisterschaft 2018 in Russland²⁷ oder auch das Grand-Slam Turnier French Open. Weiterhin konnte Immersive Audio vergangenes Jahr auch an dem internationalen Musikwettbewerb Eurovision Song Contest getestet werden.²⁸

3. Audio-Wiedergabeverfahren

Die Produktions- und Wiedergabeverfahren für Audio im Allgemeinen, basieren auf drei Grundkonzepten. Dieses Kapitel gibt einen Überblick über deren Funktionsweise. Anschließend werden bekannte Audioformate genauer vorgestellt, die in Bezug auf eine Erweiterung von Audio um die Höheninformation am Markt vorzufinden sind. Zudem werden auch alternative Wiedergabekonzepte präsentiert und auf die Probleme bei diesen unterschiedlichen Wiedergabeverfahren hingewiesen.

3.1 Grundkonzepte

Audioformate funktionieren in der Regel nach einem kanalbasierten, objektbasierten oder szenenbasierten Prinzip. Hiervon ist das kanalbasierte, das am häufigsten anzutreffende. Zunehmend findet man aber auch hybride Ansätze, bei denen diese Verfahren sowie

²⁴ Vgl. L-Acoustics (2019)

²⁵ Vgl. Fidelity Media (2016)

²⁶ Vgl. Westphal (2018)

²⁷ Vgl. Slavik (2018)

²⁸ Vgl. Simon et al. (2018), S. 1

deren jeweilige Vorteile miteinander kombiniert werden können. Welches Verfahren tatsächlich angewendet wird, hat Auswirkung auf die gesamte Produktionskette, das Übertragungsformat sowie den Renderingvorgang. Um nachvollziehen zu können, worin im Einzelnen die Unterschiede bestehen, werden zunächst alle drei Verfahren im Folgenden kurz erläutert. Anschließend werden hybride Ansätze vorgestellt.

3.1.1 Kanalbasiert

Bei der kanalbasierten Wiedergabe entspricht jedes Audiosignal genau einem Lautsprecherkanal in einer spezifischen Lautsprecherkonfiguration. Wie in Abbildung 2 zu sehen, wird bei der Produktion aus allen einzelnen verfügbaren Audio-Elementen ein fertiger Mix erzeugt, der für ein vorab definiertes Wiedergabe-Setup vorgesehen ist. Für den Endverbraucher gibt es folglich keinerlei Möglichkeiten, diese Mischung an persönliche Vorlieben oder eine abweichende Lautsprecherkonfiguration anzupassen.

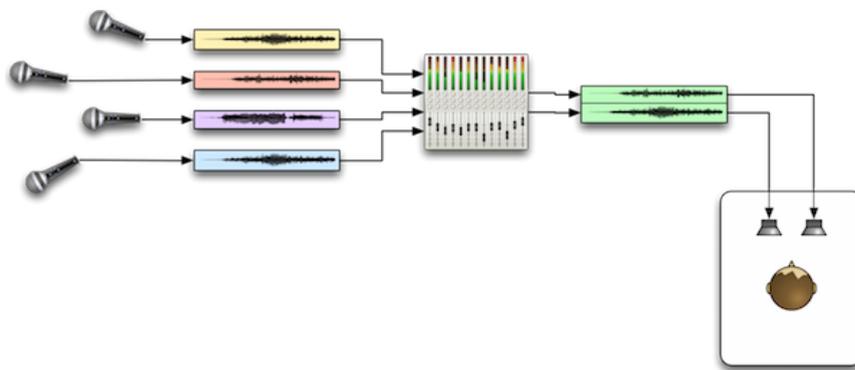


Abb. 2 Kanalbasierte Produktion und Wiedergabe²⁹

Die am häufigsten anzutreffenden, kanalbasierten Audioformate sind Stereo (2.0) und Surround (5.1), welche standardisiert sind und eine feste Anordnung der Lautsprecher sowie der genauen Abhörposition vorgeben. Bei den immersiven Audioformaten variieren die Lautsprecherkonfigurationen dagegen oft in Bezug auf die Anzahl und die Positionierung der zusätzlichen Höhenlautsprecher. Dementsprechend sind gerenderte Mischungen untereinander nur bedingt kompatibel. Abweichungen führen in der Regel zu einer Minderung der Klangqualität.³⁰

3.1.2 Objektbasiert

Das objektbasierte Verfahren bietet deutlich mehr Flexibilität als der zuvor beschriebene kanalbasierte Ansatz. Indem die einzelnen Audio-Elemente getrennt voneinander übertragen werden und die Mischung erst am Endgerät final gerendert wird, ist es möglich bis zuletzt individuelle Anpassungen vorzunehmen.

²⁹ Bildquelle: IRT GmbH (2019) Conceptual overview of channel-based audio production and consumption

³⁰ Vgl. Herre et al. (2015), S. 770-771

Die Audio-Elemente werden dafür, wie in Abbildung 3 zu sehen ist, mit zusätzlichen Metadaten versehen. Vor der Wiedergabe werden diese *Audio-Objekte* von einem Renderer in diskrete Audiokanäle umgerechnet. Dieser Renderer besitzt wiederum Informationen über die Anzahl, die Positionen der zur Verfügung stehenden Lautsprecher sowie beispielsweise die Art der Abhörumgebung oder des Wiedergabemediums. Dadurch lässt sich die Mischung auch auf andere Wiedergabesysteme adaptieren und bietet dem Endverbraucher zusätzlich die Möglichkeit nach eigenen Vorstellungen Änderungen daran vorzunehmen. In welchem Umfang dies realisierbar ist, lässt sich auf der Produktionsseite ebenfalls über die Metadaten festlegen.³¹

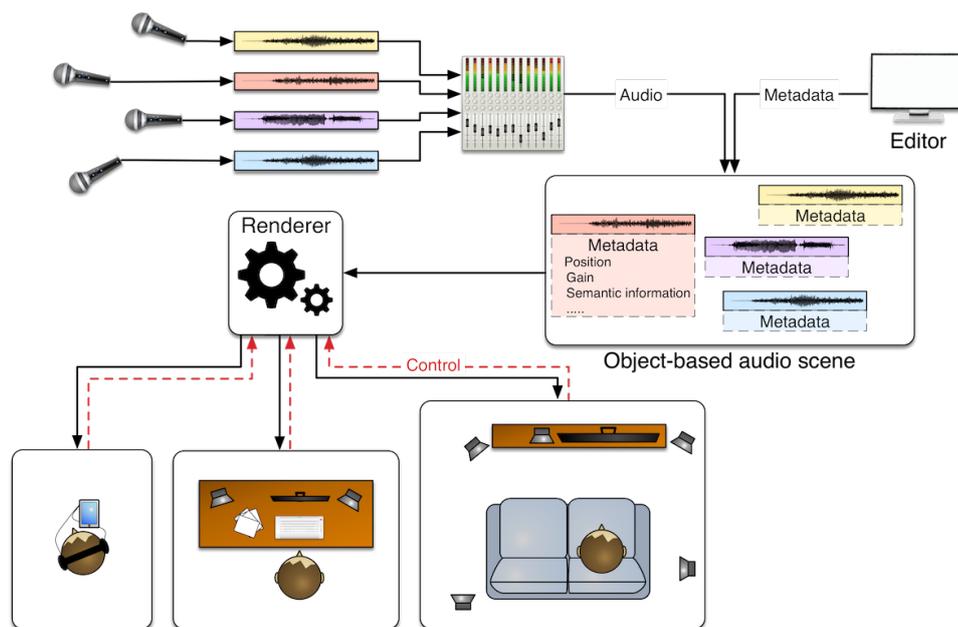


Abb. 3 Objektbasierte Produktion und Wiedergabe³²

Mit diesem Ansatz lässt sich zudem viel Bandbreite einsparen, da nicht mehr für jede Wiedergabekonfiguration ein separater Mix erstellt und übertragen werden muss.

3.1.3 Szenenbasiert

Bei einem szenenbasierten Format wird 3D-Audio über die Reproduktion eines Schallfeldes abgebildet. Dieses Verfahren findet sich häufig in Audioformaten für 360°-Video und VR-Anwendungen. Zu selbiger Kategorie zählen auch die Ambisonics-Formate. Dabei wird versucht ein dreidimensionales Schallfeld von einem Punkt aus möglichst genau zu beschreiben. Bei Ambisonics 1. Ordnung lässt sich bereits mit vier Signalen ein solches Schallfeld aufspannen. Dafür werden ein omnidirektionaler Anteil und drei gerichtete Anteile, einer für jede Raumachse X, Y und Z, benötigt. Vor allem die High-

³¹ Vgl. IRT GmbH (2019) Object-based audio - The future of audio production, delivery and consumption

³² Bildquelle: IRT GmbH (2019) Conceptual overview of object-based audio production and consumption

Order-Ambisonics-Formate (HOA) ermöglichen durch die feinere Auflösung des Schallfeldes eine hohe Wiedergabequalität.³³

Szenenbasierte Verfahren sind unabhängig von der genauen Lautsprecherkonfiguration auf der Wiedergabeseite, da die einzelnen Lautsprecherkanäle erst vor Ort über eine Decodiermatrix berechnet werden.³⁴ Die Lautsprecheranzahl sollte bei der Wiedergabe von Ambisonics-Formaten allerdings mindestens so groß sein, wie die Anzahl der einzelnen Schallfeld-Komponenten.³⁵ Ebenso, wie bei den objektbasierten Audio-Wiedergabeverfahren, wird am Ende ein Decoder und Renderer benötigt.

Auf der Produktionsseite die Audio-Elemente gemischt und anschließend in einer Schallfeldbeschreibung gespeichert. Hierbei gibt es zwei Möglichkeiten: entweder wird die akustische Szene künstlich konstruiert oder die einzelnen Schallfeldkomponenten werden beispielsweise über ein Ambisonics-Mikrofon aufgenommen. Bestimmte Manipulationen des Schallfeldes sind noch vor der Wiedergabe möglich. Szenenbasierte Formate verfolgen demnach, im Gegensatz zu den objektbasierten Verfahren, einen eher physikalischen Ansatz.³⁶

3.1.4 Hybrider Ansatz

Neben den Grundkonzepten finden sich mittlerweile einige Audioformate, welche auf einer Kombination von diesen drei Grundkonzepten basieren.

Die Mischung aus einem kanal- und objektbasierten Ansatz findet sich beispielsweise bei Dolby Atmos. Der Hauptmix wird dabei weiterhin kanalbasiert erstellt. Parallel dazu werden ausgewählte Audio-Elemente als Objekte übertragen. Räume mit vielen Lautsprechern, wie beispielsweise im Kinobereich, können bewegte Audio-Objekte dadurch besser wiedergeben, da jeder Lautsprecher einzeln adressiert werden kann.

Für den Einsatz im Broadcast wurden in den letzten Jahren die Audio-Standards überarbeitet. Die neuen NGA-Formate unterstützen die Speicherung und Übertragung von kanalbasiertem, objektbasiertem und szenenbasiertem Audio wie HOA.

3.2 3D-Audioformate

Als 3D-Audioformate werden Audio-Wiedergabesysteme bezeichnet, welche nicht nur die Surround-Ebene, sondern auch die Höhenebene abbilden und somit ein räumlicheres Klangerlebnis erzeugen können (vgl. Kapitel 2.1). Es ist unwahrscheinlich anzunehmen, dass sich Verbraucher ihre Wiedergabesysteme extra für immersive TV-Broadcasts

³³ Vgl. Herre et al. (2015), S. 771

³⁴ Vgl. ebd.

³⁵ Vgl. Slavik, Weinzierl (2008), S. 663

³⁶ Vgl. Remy (2018), S. 21

umrüsten lassen. Stattdessen wird die Audiowiedergabe von Liveübertragungen eher auf bereits vorhandenen Heimkino-Anlagen erfolgen.

3.2.1 Auro-3D

Das Auro-3D-Format ist ein kanalbasiertes 3D-Audioformat, welches im Jahre 2006 von Wilfried van Baelen vorgestellt wurde, mit dem Ziel, die bisherigen Surround-Formate um eine authentische Abbildung der Höheninformation zu erweitern. Dabei sollte gleichzeitig eine möglichst hohe Kompatibilität zu vorhandenen Standards, wie Stereo und 5.1, gewährleistet sein.³⁷

In der Basiskonfiguration Auro 9.1 wird dafür das Standard 5.1-Setup um vier zusätzliche Höhenlautsprecher erweitert (siehe Abb. 4). Diese befinden sich in einem 30°-Winkel zur Abhörposition oberhalb der Lautsprecher des Surround-Setups. Grund dafür sieht Wilfried van Baelen darin, dass wenn die Winkel in der vertikalen Achse größer seien als 35 Grad, wie es bei Dolby Atmos und DTS:X der Fall sei, gehe die vertikale Kohärenz verloren, die für ein natürliches räumliches Klangerlebnis so entscheidend sei.³⁸ Bei Auro 10.1 wird die Basiskonfiguration noch um einen zusätzlichen, zentralen Decken-Lautsprecher (sog. *Voice of God*) erweitert. Auro 11.1 enthält einen Center-Kanal für die vorderen Höhenlautsprecher und Auro 13.1 zwei zusätzliche Center-Kanäle für die Surrounds. Als kanalbasiertes Verfahren müssen die Anzahl und Positionen der Lautsprecher genau definiert sein. Das geht auch aus der nachfolgenden Abbildung hervor.

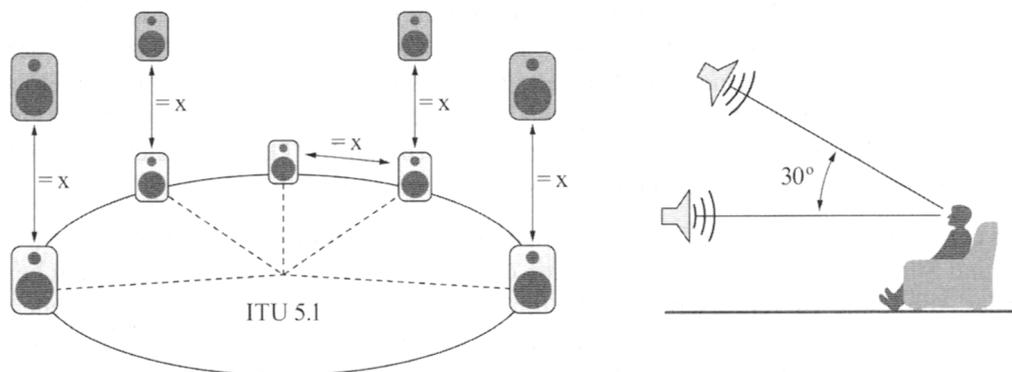


Abb. 4 Auro 9.1³⁹

Die Firma Auro-Technologies liefert neben verschiedenen Tools für die Produktion, wie dem Auro-Panner, einem 3D-Hall Plugin *Auro-Verb* und einem Upmixer *Auro-Matic*, auch einen herstellereigenen, auf PCM basierenden Codec, der die Höhenkanäle quasi-verlustfrei ohne zusätzliche Bandbreite mit überträgt.

³⁷ Vgl. Theile (2014), S. 312-313

³⁸ Van Baelen (2016), S. 18

³⁹ Bildquelle: Theile (2014), S. 312

Auro-3D zielt hauptsächlich auf Anwendungen im Bereich Kino, Live-Events, Car-Hifi oder Games ab. Aufgrund der guten räumlichen Umhüllung eignet es sich besonders für die Wiedergabe von Musik. Die kanalbasierte Funktionsweise und Kompatibilität zu einem herkömmlichen Surround-Setup, ermöglicht es zudem auf das Wissen von bisherigen Mikrofonierungs- und Produktionstechniken zurückzugreifen.

3.2.2 Dolby Atmos

Dolby Atmos ist die Weiterentwicklung und Ergänzung von Dolby Surround um die *Overhead*-Ebene.⁴⁰ Dabei baut Dolby Atmos in der einfachsten Heimkino-Konfiguration ähnlich wie Auro-3D zunächst auf einem Standard 5.1-System auf (siehe Abb. 5). Dieses wird einfach um zwei oder vier Lautsprecher über der Abhörposition erweitert. Audiosignale können dadurch auch von oben wiedergegeben werden. Die nachfolgende Abbildung stellt exemplarisch eine 5.1.4-Konfiguration für Heimkino-Anwendungen dar.

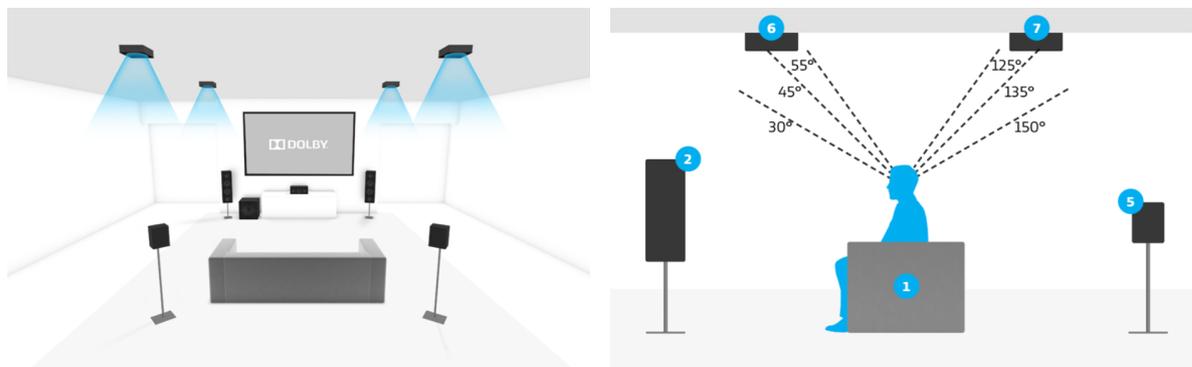


Abb. 5 Dolby Atmos 5.1.4⁴¹

Wie in Abbildung 5 auf der rechten Seite zu sehen ist, gibt es bei der Positionierung der Höhenlautsprecher allerdings einen bestimmten Spielraum. Im Fall von dieser 5.1.4-Heimkinokonfiguration zwischen 30° und 55° zur horizontalen Ebene. Die optimale Position hängt dabei von der Größe und Höhe des Raumes ab. Im Unterschied zu Auro-3D werden die Lautsprecher nicht direkt über den einzelnen Surround-Lautsprechern, sondern an der Decke positioniert. Ein Dolby-Atmos-System lässt sich flexibel um zusätzliche Lautsprecher erweitern. Dolby-Konfigurationen für den Heimkino-Bereich reichen von 5.1.2 bis hin zu 11.1.8.

Ein wesentlicher Unterschied zu den Auro-Formaten ist, dass Dolby Atmos nicht rein kanalbasiert arbeitet. Neben einem statischen, kanalbasiertem *Audio-Bed* können außerdem dynamische Audio-Objekte passend auf das zur Verfügung stehende Setup gerendert werden. Gerade bei Konfigurationen mit einer größeren Anzahl an Lautsprechern hat das Vorteile.⁴² Ausführlichere Informationen über die Funktionsweise,

⁴⁰ Vgl. Dolby Laboratories, Inc. (2014), S. 5

⁴¹ Bildquelle: Dolby Laboratories, Inc. (o.D.) Home Theater Speaker Guide. 5.1.4 Overhead Speakers

⁴² Vgl. Dolby Laboratories, Inc. (2014), S. 4-7

Produktionstools und den Workflow für Dolby Atmos, mit Fokus allerdings auf den Bereich Live-Broadcast, folgen in den Kapiteln 5 und 6.3.

3.2.3 DTS:X

Unter diesem Namen vertreibt DTS, Inc. weltweit eine ganze Reihe von Audiolösungen, die für Immersive Audio konzipiert sind. Anwendung finden diese im Kino, Heimkino, Automobilindustrie, Gaming oder auch in mobilen Endgeräten.

DTS:X ist ein objektbasiertes Format, welches laut Hersteller auf jeder Lautsprecheranstellung wiedergegeben werden kann.⁴³ Eine komplett beliebige Positionierung kann sich aber negativ auf die Wiedergabequalität auswirken, da der Decoder über Optimierungsalgorithmen in die Audiosignale eingreifen muss. Insofern ist zu empfehlen auf die Vorschläge anderer Formate (z.B. Auro-3D, Dolby Atmos) zurückzugreifen, wodurch gleichzeitig eine bessere Kompatibilität zu diesen erreicht werden kann.⁴⁴

Um die flexible Wiedergabe auf unterschiedlichen Systemen überhaupt erst möglich zu machen, kommen in den jeweiligen Endgeräten verschiedene Rendering-Engines zum Einsatz. Für den Heimkino-Bereich beispielsweise, lässt sich mit DTS Virtual:X ein bestehendes 2.0- oder 5.1-System mithilfe von psychoakustischen Effekten um die Surround- und Höhenwiedergabe erweitern.⁴⁵ Bei DTS Neural:X handelt es sich um einen Upmixer, der räumliche Information in Stereo- oder Surround-Audiomaterial erkennt und bei vorhandenen Höhenlautsprechern die jeweiligen Lautsprechersignale generiert.⁴⁶ Mit Headphone:X bietet DTS, Inc. zudem auch eine binaurale Lösung für Kopfhörer an.⁴⁷

3.2.4 NHK 22.2

Dieses kanalbasierte, immersive Audioformat wurde von der Entwicklungsabteilung des japanischen Rundfunks entwickelt und erstmals 2005 auf der EXPO in Japan präsentiert. Seit 2018 ist 22.2 zusammen mit Ultra High Definition (UHD)⁴⁸ neuer TV-Standard in Japan und dort per Satellit empfangbar.

Hintergrund für diese neuartige Entwicklung, war die Suche nach einem neuen Broadcast-Standard, der die Übertragung und Wiedergabe von UHD-TV unterstützt und gleichzeitig ein immersiveres Klangerlebnis ermöglicht. Mit der höheren Bildauflösung ist bei identischem Abstand ein deutlich größerer Blickwinkel möglich. Für die Synchronisation von bewegten Bild- und Audio-Objekten reichen drei Lautsprecher (Links, Center, Rechts) dadurch nicht mehr aus. Deshalb wurde die gesamte Frontseite um acht zusätzliche

⁴³ Vgl. DTS, Inc. (2018) DTS:X Homepage

⁴⁴ Vgl. Schuller (2015)

⁴⁵ Vgl. Denon, DTS Virtual:X

⁴⁶ Vgl. Griffin (2017)

⁴⁷ Vgl. DTS, Inc. (2018) Home Solutions

⁴⁸ In Japan: Super Hi-Vision

Lautsprecher erweitert (siehe Abb. 6). Bisher ist 22.2 das einzige Wiedergabeverfahren, welches auch Lautsprecher unterhalb der Ohr-Höhe, das sogenannte *Lower layer*, enthält.⁴⁹

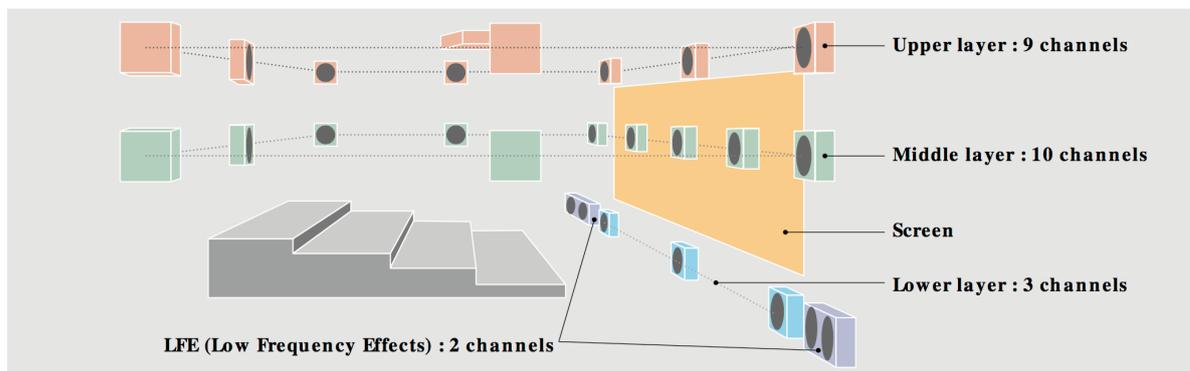


Abb. 6 NHK 22.2 Lautsprechersystem⁵⁰

Insgesamt betrachtet, ermöglicht das 22.2-Format einen größeren Sweet-Spot und gute Kompatibilität zu den bisherigen Produktionsformaten.⁵¹ Allerdings stellt die große Anzahl von Lautsprechern und Audiokanälen extreme Anforderungen an die Produktions- und Wiedergabeumgebungen, sowie die Übertragungsstrecke. Gleichzeitig fehlen die Vorteile von objektbasierten Verfahren für Live-Broadcast-Anwendungen.

3.3 Alternative Wiedergabekonzepte

Der Installations- und Kostenaufwand für die vorgestellten 3D-Audioformate ist verhältnismäßig hoch. Deutlich bequemer und vermutlich kostengünstiger wird beispielsweise der Einsatz von Soundbars mit einem integrierten Lautsprecher-Array sein. Diese arbeiten mithilfe von Decken- und Wandreflexionen und ermöglichen dadurch eine akzeptable Umhüllung ohne zusätzliche, diskrete Lautsprecher im Raum. Mit Sennheisers AMBEO Soundbar, ist ab Mai 2019 eine Alternative zu aufwendigen Lautsprecherkonfigurationen auf dem Markt verfügbar.⁵²

Weiterhin bieten Audiotechnologiehersteller den Verbrauchern, die bereits über ein Surround- oder 7.1-System verfügen, die Möglichkeit an, dieses über *Upfiring-Speaker* um die Höhenwiedergabe zu ergänzen. Dolby entwickelte dafür in Zusammenarbeit mit mehreren Herstellern sogenannte *Dolby Atmos Enabled Speaker*. Diese werden einfach auf die bereits verfügbaren Surround-Lautsprecher gestellt oder sind bereits integriert und in Richtung Decke ausgerichtet. Upfiring-Speaker arbeiten also ebenfalls über

⁴⁹ Vgl. Hamasaki, Hiyama (2006), S. 9-10

⁵⁰ Bildquelle: ebd., S. 10

⁵¹ Vgl. Hamasaki (2011), S. 15

⁵² Vgl. Sennheiser electronics GmbH & Co. KG (2019)

Reflexionen. Vorausgesetzt wird allerdings eine einigermaßen ebene und reflektierende Decke. Das gleiche gilt auch für den Einsatz der bereits genannten Soundbars.

Die Wiedergabe von Immersive Audio ist zudem auch über handelsübliche Kopfhörer möglich. Dafür bieten die Hersteller unterschiedlichste Binauralisierungs-Algorithmen an, welche einen umhüllenden Klang erzeugen sollen. Diese Art der Audiowiedergabe ist besonders praktisch, weil sie sehr verbreitet und zudem vergleichsweise günstig ist.

3.4 Problematik bei unterschiedlichen Audioformaten

Wie bereits in Kapitel 3.1.1, gerade in Bezug auf die kanalbasierten Formate angedeutet wurde, liegt eine Hauptproblematik bei der Nutzung von unterschiedlichen Audioformaten in deren Inkompatibilität untereinander. Produktionen, die für ein bestimmtes Format erstellt wurden, sind zu abweichenden Lautsprecherkonfigurationen möglicherweise nur bedingt kompatibel. Gleichzeitig kann die Wiedergabesituation beim Verbraucher, im Gegensatz zu den Qualitätsanforderungen in kontrollierten Umgebungen, durchaus gewissen Einschränkungen unterworfen sein. Die Aufstellung und Hüllkurven der Lautsprecher im Kino sind weitestgehend standardisiert. Dagegen werden im Heimbereich die Lautsprecher aufgrund der vorhandenen Räumlichkeiten oder auch aus optischen Gründen nicht immer optimal positioniert. Zudem ist die Bandbreite an Wiedergabeendgeräten beim Endverbraucher sehr groß. Ob es sich um eine Heimkino-Anlage, die TV-Lautsprecher, eine Soundbar, das Tablet oder handelsübliche Kopfhörer handelt, bleibt offen.⁵³

Es ist klar, dass nicht für jedes Audioformat und Wiedergabeendgerät eine separate Mischung erstellt werden kann. Deshalb ist es notwendig, dass die jeweiligen Produktionen untereinander transkodiert werden können und die vorhandenen Decoder in den Empfangsgeräten akzeptable Up- und Downmixing-Algorithmen integriert haben um das Audiomaterial bestmöglich wiederzugeben. Objektbasierte Formate sind in dieser Hinsicht sicherlich im Vorteil, da die Audiokanäle grundsätzlich erst am Endgerät final gerendert werden und dadurch das Wiedergabesystem im Renderingvorgang mit berücksichtigt werden kann.

Gerade im Kino und Live-TV-Broadcast bietet sich hier großes Potenzial. Allerdings werden Audio-Objekte in einer Filmmischung anders verwendet wie in TV-Übertragungen. Während im Kino Audio-Objekte meist für eine präzisere räumliche Wiedergabe sowie dynamische Spezialeffekte eingesetzt werden, ermöglicht eine objektbasierte Signalkette im Broadcast dem Verbraucher mehr Kontrolle über die Sprach-Lautstärke zu überlassen.⁵⁴ Und das Ganze, ohne dabei aufwendig verschiedene Mixversionen zu

⁵³ Vgl. Bleidt (2014), S. 9

⁵⁴ Vgl. ebd.

generieren und zu übertragen. Dafür werden allerdings neue Audio-Codex und Produktionstools benötigt, welche diesen bisher rein kanalbasiert geprägten Audio-Workflow um die neuen Funktionen erweitern.

4. MPEG-H

Bei MPEG-H handelt es sich um einen multimedialen Standard, der von der Moving Picture Experts Group (MPEG) unter maßgeblicher Beteiligung des Fraunhofer IIS und deren Partner entwickelt wurde. Neben einem Standard für den Transport digitaler Medien und für Video-Kompression, ist auch der für die vorliegende Arbeit relevante Audio-Standard enthalten. Dessen Anwendung zielt vor allem auf die Bereiche UHD-TV, Streaming und VR ab.⁵⁵

Dieses Kapitel liefert grundlegende Informationen zum Aufbau, der Funktionsweise und der Handhabung des MPEG-H Audio Systems. Es werden Produktionstools erklärt und in den Kontext einer Produktionsumgebung eingeordnet. Ausführlichere Informationen über die Einbindung in moderne Produktionssysteme sind in der Workflow-Analyse in Kapitel 6 dargestellt.

4.1 Grundlagen MPEG-H 3D Audio

Im Jahr 2015 wurde MPEG-H in der ISO/IEC-Norm 23008 standardisiert. Es folgten weitere Spezifizierungen für UHD-Fernsehen von DVB und ATSC. Südkorea ist mit MPEG-H bereits seit 2017 auf Sendung und war somit das erste Land, welches ausschließlich diesen NGA-Standard als neuen Audio-Broadcast-Standard verwendet hat.⁵⁶ Zusammen mit Dolby AC-4 (siehe Kapitel 5.2.2) ist MPEG-H 3D Audio als NGA-Codec im ATSC 3.0 Standard für US-UHD TV enthalten. Von beiden Codex wurden dementsprechend ähnliche Funktionalitäten gefordert, weshalb in Kapitel 5.2.2 die nachfolgenden Punkte 4.1.1 bis 4.1.3 nicht erneut tiefergehend betrachtet werden.

MPEG-H 3D Audio ermöglicht die Übertragung von Immersive Audio und unterstützt ein formatunabhängiges Rendering. Kanalbasierte, objektbasierte und szenenbasierte Audioinhalte werden hierbei gleichermaßen unterstützt. Des Weiteren können diese zudem auch miteinander kombiniert werden (vgl. Kapitel 3.1.4). Im Kern basiert der Audio-Codec auf *MPEG Unified Speech and Audio Coding*, kurz USAC, welcher im Hinblick auf die Codierungseffizienz sowie die Unterstützung immersiver Audioformate erweitert wurde. Theoretisch ermöglicht ein MPEG-H Container bis zu 128 Audiokanäle, 128 Audio-Objekte und HOA bis zur 29. Ordnung. Für die meisten Anwendungsbereiche und deren

⁵⁵ Vgl. Fleischmann (2017)

⁵⁶ Vgl. Simon et al. (2018), S. 1

Infrastruktur ist dies jedoch gar nicht erforderlich, weshalb sogenannte *Low Complexity Profile* (LC) spezifiziert wurden.⁵⁷

Das Profil MPEG-H LC-3, zielt auf ein Anwendungsszenario innerhalb einer SDI-Infrastruktur ab. Aufgrund der bereits vorhandenen SDI-Infrastruktur in den meisten Broadcast-Produktionsumgebungen, ist die Relevanz dieses Komplexitätsprofils daher besonders hoch. Mithilfe des MPEG-H LC-3 können insgesamt 32 Audio-Elemente innerhalb eines Bitstreams transportiert und bis zu 16 Audiokanäle gleichzeitig decodiert werden. Dadurch wird sogar die Übertragung von HOA bis zur 6. Ordnung möglich. Das in ATSC 3.0 spezifizierte MPEG-H TV Audio System für UHD-TV basiert ebenfalls auf MPEG-H LC-3, auch wenn das beispielsweise für einen 22.2-Broadcast nicht ausreichend wäre. Dafür müsste ein höheres Komplexitätsprofil verwendet werden. In diesem Fall mindestens LC-4, welches das Decodieren von bis zu 28 Audiokanälen unterstützt.⁵⁸

Über die Verwendung von Audio-Objekten bekommt der Endverbraucher zusätzliche Steuerungsmöglichkeiten an die Hand. Beispielsweise lassen sich im Live-TV-Broadcast verschiedene Sprachen oder Kommentatoren als einzelne Objekte definieren und separat von dem internationalen Mix übertragen. Dadurch kann die Lautstärke der Objekte unabhängig vom Audio-Bed verändert werden, ohne dass die Gesamtlautheit verändert wird. Im Vergleich zu den bisherigen Audio-Codecs im TV- und Rundfunk bietet MPEG-H 3D Audio dementsprechend einerseits mehr Funktionalität, andererseits aber auch eine effizientere Datenkompression. Bei gleicher wahrgenommener Qualität können zu aktuell gängigen Bitraten doppelt so viele Audiokanäle komprimiert werden.

4.1.1 Metadaten

Der MPEG-H 3D Audio-Codec unterstützt ein objektbasiertes Übertragungskonzept (siehe auch Kapitel 3.1.2). Dementsprechend können Audiosignale zusammen mit Metadaten encodiert, transportiert und erst später am Decoder final gerendert werden. Diese Metadaten enthalten zum einen strukturelle Informationen über die Audiosignale, definieren zum anderen aber auch die Art und den Umfang der User Interaktion sowie dynamische Objekteigenschaften. Wird im folgenden von MPEG-H gesprochen, bezieht sich das auf MPEG-H 3D Audio.

Die Metadaten werden in Offline-Workflows oder zur statischen Konfiguration eines MPEG-H Encoders in Form eines XML-Files gespeichert. In Echtzeit-Anwendungen, wie Live-Broadcasts, wird dagegen ein sogenannter *Control-Track* generiert. Dabei handelt es sich um ein PCM-Signal, ähnlich einem Timecode-Signal, welches dem MPEG-H Kontributions- oder Emission-Encoder zugeführt wird. Innerhalb eines SDI-Workflows ist dieser Control-Track mit der jeweiligen Videoframerate synchronisiert und übersteht auch

⁵⁷ Vgl. Grewe, Simon, Scuda (2018), S. 336

⁵⁸ Vgl. ebd., S. 335-336

einfache Signalmanipulationen, wie Lautstärkeveränderungen, Filterung oder sogar Resampling. Neu generiert werden muss der Control-Track dagegen, wenn Eingriffe direkt an den Metadaten einer MPEG-H Audio-Szene vorgenommen wurden.⁵⁹ Solche Änderungen am Control-Track sind nicht zu verwechseln mit Eingriffen in den *MPEG-H Audio Stream* (MHAS) (siehe Kapitel 4.2).

In zukünftigen IP-basierten Workflows sollen die einzelnen Komponenten Audio, Video und Metadaten, als separate IP-Streams übertragen werden. In der Folge würde die gesamte Produktions- und Übertragungskette vereinfacht und Kanalbeschränkungen, wie es noch bei SDI der Fall ist, aufgehoben werden.⁶⁰

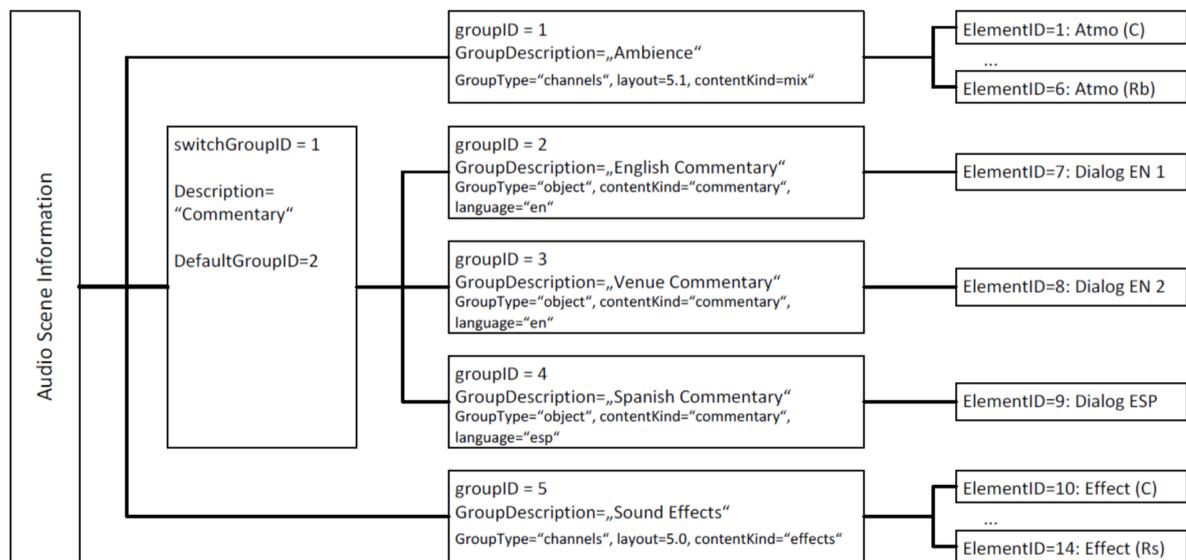


Abb. 7 Aufbau einer Audio-Szene⁶¹

In der MPEG-H Terminologie werden einzelne Audiospuren als sogenanntes *Audio-Element* bezeichnet, welche in einer gemeinsamen *Audio-Szene* zusammengefasst und in Form von Gruppen⁶² oder Switch-Groups⁶³ organisiert werden können. Dadurch entsteht eine hierarchische Struktur (siehe Abb. 7), die in den Metadaten der Audio-Szene gespeichert ist. Darin wird auch festgelegt ob, und inwiefern Interaktivität für den User erlaubt sein soll.⁶⁴

⁵⁹ Vgl. Bleidt et al. (2017), S. 225-226

⁶⁰ Vgl. Simon et al. (2018), S. 3

⁶¹ Bildquelle: MPEG-H Audio Scene Information example (2017)

⁶² Einzelne Audio-Elemente lassen sich gruppieren und dadurch gleichzeitig bearbeiten. Praktisches Anwendungsbeispiel dafür ist die Gruppierung der beiden Audiokanäle eines Stereo-Tracks

⁶³ In einer Switch-Group ist immer nur ein Audio-Element aktiv. Praktisches Anwendungsbeispiel ist die Zuordnung mehrerer Audio-Sprachen zu einer gemeinsamen Switch-Group. Der User hat die Möglichkeit eine Sprache auszuwählen ohne dass alle gleichzeitig wiedergegeben werden (siehe in Abb. 7 Audio Elemente 7-9)

⁶⁴ Vgl. Bleidt et al. (2017), S. 211

Bei dynamischen Objekten werden die in der Produktion erstellten Metadaten zusammen mit einem Zeitstempel verknüpft und an einen MPEG-H Encoder weitergeleitet. Unabhängig von der jeweiligen Produktionsumgebung werden diese Audio-Objekte anschließend auf die jeweilige Wiedergabekonfiguration gerendert.⁶⁵

4.1.2 Rendering

Bevor ein MPEG-H Audio Stream wiedergegeben werden kann, muss dieser zunächst decodiert werden. Anschließend müssen die einzelnen Ausgabesignale berechnet werden. Dafür kommen je nach Art des Audiomaterials verschiedene Rendering-Module zum Einsatz. Vor diesen Modulen verarbeitet ein *Element Metadata Processor* jedoch die Metadaten sowie die User Interaktivität. Dieser verwaltet auch die Informationen über das lokale Wiedergabesystem.⁶⁶

Kanalbasierte Audiosignale werden über den *Format Converter* (auch Format Renderer genannt) flexibel auf die zur Verfügung stehenden Lautsprecher gerendert, unabhängig davon, ob deren Anzahl und Positionierung mit dem Wiedergabesystem auf der Produktionsseite übereinstimmt. Zudem bietet dieser Renderer intelligente Downmix-Algorithmen an, welche mögliche Klangverfärbungen oder Auslöschungen vermeiden sollen. Der *Object Renderer* erzeugt die Lautsprechersignale abhängig von den mitgelieferten Metadaten und dem jeweiligen Lautsprechersystem. Sind beispielsweise zu wenige Lautsprecher verfügbar, wird versucht diese über sogenannte *Virtuelle Lautsprecher* nachzubilden. Auch die User Interaktivität wird im Object Renderer verarbeitet. Für HOA werden im *HOA Renderer* die decodierten Signale mit einer Rendering-Matrix verrechnet, um daraus die einzelnen Lautsprechersignale zu generieren. Diese berücksichtigt unter anderem die Anzahl der HOA-Koeffizienten und ist unabhängig vom Wiedergabesystem.⁶⁷

Im Standard MPEG-H 3D Audio Decoder ist auch ein binaurales Rendering Modul vorgesehen, welches kanal-, objekt- oder szenenbasierte 3D-Audioinhalte direkt auf einen Stereo Ausgang für Kopfhörer ausgeben kann.⁶⁸

4.1.3 Lautheit und Dynamik

Die NGA-Standards, wie MPEG-H, müssen viele unterschiedliche Tools für die Kontrolle der Lautheit (eng. Loudness) und Anpassung der Programmdynamik innerhalb der Metadaten unterstützen, weil die benötigten Werte je nach Wiedergabeumgebung und Endgerät variieren. Dazu gehört zum einen die Messung und Überwachung der gesamten Programmlautheit, die Anpassung der Dynamik einzelner Audio-Elemente (*Dynamic*

⁶⁵ Vgl. Bleidt et al. (2017), S. 212

⁶⁶ Vgl. ebd., S. 213

⁶⁷ Vgl. ebd., S. 213-217

⁶⁸ Vgl. ebd., S. 217

Range Control), eine Peak- und Clipping-Kontrolle aber auch Duckingfunktionen für Voice-Over und Audiodeskriptionen.⁶⁹

Damit am Decoder stets eine gleichmäßige Programmlautheit gewährleistet werden kann, muss die Lautheit zunächst normalisiert (*Loudness Normalization*) und bei Veränderungen adaptiv angepasst werden (*Loudness Compensation*). MPEG-H Audio bietet dafür Messungen nach zahlreichen internationalen Broadcast-Standards, wie ITU-R BS.1770-4, EBU R-128 oder ATSC A/85, an. Neben der Normalisierung unterstützt MPEG-H zudem auch eine Lautheitskompensation.⁷⁰ Wird beispielsweise durch einen User die Sprachlautstärke angehoben, würde sich das auch auf die gesamte Programmlautheit auswirken. Mit der Kompensation verändert sich dann aber lediglich das Verhältnis zwischen Sprache und dem Grundmix, sodass die angestrebte *Target Loudness* erhalten bleibt (vgl. Abb. 8).



Abb. 8 Lautheitskompensation bei Anhebung der Dialoglautstärke⁷¹

Mit der *Dynamic Range Control* (DRC) lässt sich die Dynamik von Audiosignalen einschränken. Bei NGA-Codecs sind Gain-Veränderungen allerdings nicht destruktiv, sondern werden parallel zum unbearbeiteten Signal in den Metadaten mit übertragen. Erst der Decoder entscheidet, ob die jeweilige DRC angewendet wird.⁷² Für verschiedene Umgebungsszenarien (z.B. Wohnzimmer, Flugzeug, etc.) oder Endgeräte (z.B. Heimkino-Anlage, TV, Smartphone, etc.) lassen sich jeweils eigene DRC-Profile übertragen. Im Flugzeug ist der umgebende Geräuschpegel in der Regel lauter wie in einem Wohnzimmer. Leise Parts in einem sehr dynamischen Audiosignal setzen sich gegebenenfalls nicht durch. Hier ist eine Mischung mit geringerer Dynamik von Vorteil. In der Regel werden schon im Vorfeld Audio-Peaks oder Clipping im Programm durch die beschriebenen Lautheitsregulierungen verhindert. Ausgeschlossen werden soll das aber spätestens durch einen Peak-Limiter am Ausgang des MPEG-H Decoders.⁷³

⁶⁹ Vgl. Bleidt et al. (2017), S. 217

⁷⁰ Vgl. ebd., S. 218-219

⁷¹ Bildquelle: ebd., S. 219

⁷² Vgl. ebd., S. 219-220

⁷³ Vgl. ebd., S. 220

4.2 MPEG-H Audio Stream

Unter *MPEG-H Audio Stream* (MHAS) versteht man ein paketbasiertes Bitstream-Format, welches für den Transport von MPEG-H encodierten Audiosignalen und Metadaten entwickelt worden ist. MHAS kann außerdem Bestandteil eines MPEG-2 Transport Streams oder auch eines ISO Base Media File Formates (ISOBMFF) als Container sein. Auf letzterem basieren viele aktuelle adaptive Streaming- und Broadcasting-Technologien.⁷⁴

Seinen Einsatz in der Praxis findet MHAS auf der Übertragungstrecke zwischen MPEG-H Encoder und Decoder. Hierbei muss berücksichtigt werden, dass MHAS auf der Produktionsseite, im Kontributionsbereich, nicht gleichbedeutend mit dem ist, welches final ausgestrahlt wird. Der MPEG-H Audio Stream eines Kontributions-Encoders kann nicht von einem normalen MPEG-H Decoder verarbeitet werden, wie er in handelsüblichen Endgeräten zu finden ist. Dazu wird explizit ein professioneller Kontributions-Decoder benötigt, der aus MHAS die einzelnen Audiosignale wieder decodiert und die Metadaten des Bitstreams z.B. in einen neuen Control-Tack gewandelt. Contribution-MHAS ersetzt, ähnlich wie Dolby E und ED2, den SDI-basierten Transport überall dort, wo eine transparente Codierung erforderlich ist um die Datenrate für die Übertragung zu reduzieren ohne die Qualität einzuschränken.⁷⁵ Erst bei der Generierung des MPEG-H Audio Streams am Emission-Encoder für die Übertragung zum Verbraucher findet eine deutliche Datenreduktion statt.

4.3 MPEG-H Produktionstools

Für die Produktion von MPEG-H muss die Infrastruktur entsprechend adaptiert werden. Dabei reicht die Produktpalette der MPEG-H Produktionstools von Hardwarekomponenten über Stand-Alone Software bis hin zu Plugins oder der nativen Integration in moderne DAWs. Mithilfe dieser Tools ist es möglich immersive Audioinhalte und Objekte zu erstellen, ohne dabei die Infrastruktur und den Workflow grundlegend umzubauen.

Dabei ist zu berücksichtigen, dass Echtzeit-Anwendungen andere Anforderungen stellen als beispielsweise dateibasierte Offline-Workflows. Da der Fokus primär auf dem Workflow von Live-Broadcast-Produktionen liegt, werden die Postproduktionstools an dieser Stelle zwar thematisiert, jedoch nicht tiefergehend behandelt. Letztendlich lassen sich auch mit diesen Tools MPEG-H Audioinhalte für die Ausstrahlung im Broadcast erstellen und Encoder vorkonfigurieren. Produktionsworkflows mit einer Beteiligung dieser Postproduktionstools, werden innerhalb dieser Bachelorarbeit allerdings nicht thematisiert. Das schließt aber nicht aus, dass nicht auch hybride Broadcast-Workflows denkbar sind.

⁷⁴ Vgl. Bleidt et al. (2017), S. 221-222

⁷⁵ Vgl. The MPEG-H TV Audio System. Use Cases and Workflows (o.D.), S. 3

4.3.1 MPEG-H Audio Monitoring and Authoring Unit

Die *Audio Monitoring und Authoring Unit* (AMAU) wurde speziell für den Einsatz in Produktionsumgebungen konzipiert, bei denen das Monitoring, Authoring und Rendering von MPEG-H 3D Audio in Echtzeit erfordert wird. Derzeit gibt es auf dem Markt bereits zwei kommerziell erhältliche Systeme: die *MMA* (Multichannel Monitoring and Authoring) von Jünger Audio und die *AMS* (Authoring and Monitoring System) von Linear Acoustic.⁷⁶

Über eine Webapplikation (siehe Abb. 9) erhält der Toningenieur Zugriff auf sämtliche Parameter, Metadaten und Statusinformationen. Hierüber lassen sich Audio-Objekte definieren und Presets für die unterschiedlichen Audioinhalte erstellen. Auch welcher Umfang von User Interaktivität möglich sein soll, lässt sich dort präzise festlegen. Alle Konfigurationsdaten können sowohl als Presets gespeichert, als auch zu einem späteren Zeitpunkt erneut importiert werden. Beispielsweise lassen sich damit MPEG-H Encoder statisch konfigurieren, was je nach Produktionsszenario die Verwendung einer AMAU hinfällig macht. Daneben bietet die Applikation zusätzlich umfassende Tools für die Messung und das Monitoring der einzelnen Audio-Pegel sowie der Programmlautheit.⁷⁷

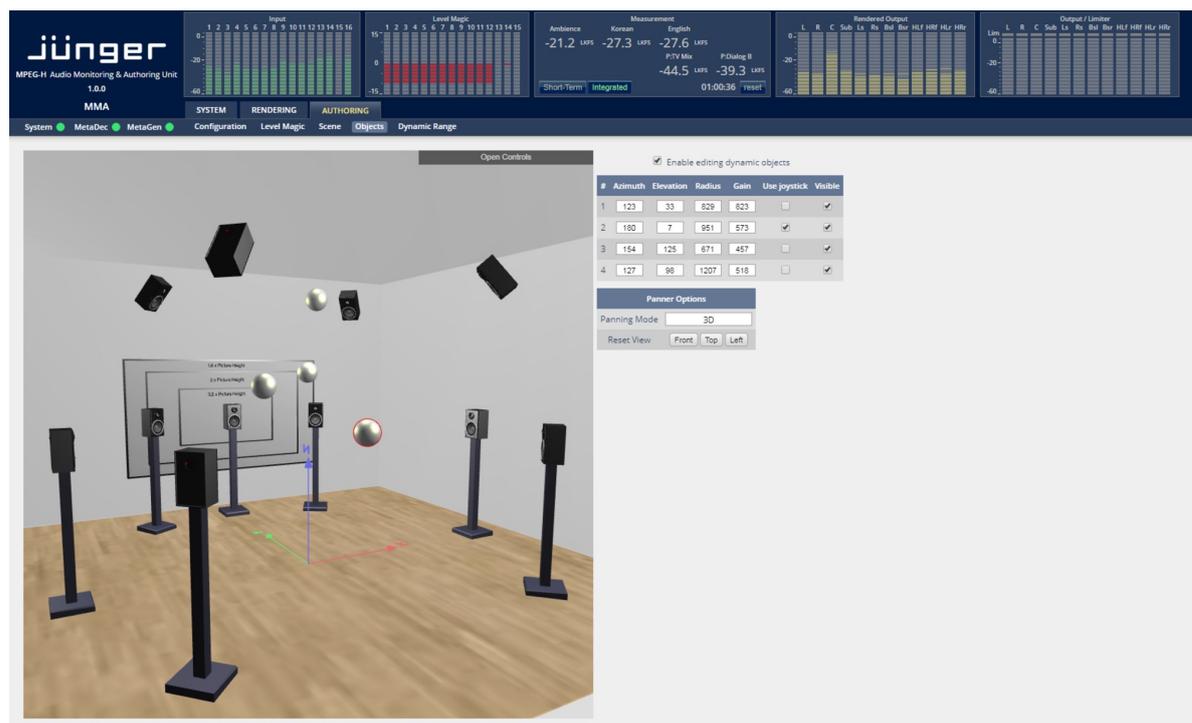


Abb. 9 Webapplikation der MMA von Jünger Audio⁷⁸

Eine virtuelle 3D-Raumansicht in der Webapplikation der MMA (siehe Abb. 9) ermöglicht ein graphisches Feedback zur Positionierung der einzelnen Audio-Objekte. Diese erfolgt

⁷⁶ Vgl. Grewe, Simon, Scuda (2018), S. 338

⁷⁷ Vgl. ebd., S. 338-339

⁷⁸ Bildquelle: Jünger Audio GmbH (o.D.)

entweder manuell durch Eingabe der jeweiligen Koordinaten oder durch die Einbindung eines zusätzlichen, externen Controllers.⁷⁹

Ausgabeseitig können verschiedenste Lautsprecheraufstellungen von Mono bis hin zu 7.1.4 mit der AMAU verbunden werden. Eine binaurale Kopfhörerausgabe direkt am Hardwaregerät, wie bei Dolbys DP590, ist nicht integriert. Dafür bieten diese Geräte zusätzlich auch eine Set-Top-Box Emulation an, womit sich die Einstellungen der MPEG-H Audio-Szene aus der Verbraucherperspektive leicht überprüfen lassen. Parallel zur Ausgabe von Immersive Audio nach ATSC 3.0, also bis zu 15 Audiokanäle plus einem Control-Track mit Metadaten, können auch ein gerenderter 5.1- und Stereo-Down-Mix generiert werden.⁸⁰ Alle unkomprimierten PCM-Audiosignale sowie die von der AMAU erstellten Metadaten und Konfigurationsdaten landen an einem MPEG-H Encoder, der daraus letztendlich den MPEG-H Audio Bitstream (MHAS) generiert.⁸¹ Eingesetzt werden kann die AMAU auch mehrmals innerhalb der Produktionskette. Beispielsweise für die Qualitätskontrolle vor dem Emission-Encoder oder für die Wiedergabe von bereits verfügbaren MPEG-H Audioinhalten.

Keines der derzeit verfügbaren Mischpulte besitzt die notwendigen Mixing- und Authoring-Tools, welche bei der Produktion von Immersive Audio für NGA-Standards wie MPEG-H benötigt werden. Ein Austausch der vorhandenen Produktionsmischpulte wäre für die Rundfunkanstalten und die Außenübertragungsdienstleister auch nicht ohne erheblichen Kostenaufwand möglich. Die praxistaugliche Lösung ist die Verwendung der AMAU-Systeme. Diese lassen sich einfach in die bisherige SDI-Infrastruktur und Produktionskette einbinden. Mit der Erweiterung können Mischpulte, die nur über Stereo- und Surround-Busse verfügen, weiterhin genutzt werden.

4.3.2 Tools für die Postproduktion

Die im nachfolgenden Abschnitt vorgestellten Tools sind für die Offline- und Postproduktion gedacht. Zu deren Aufgaben gehört zum einen die Erstellung von Presets und Metadaten einer MPEG-H Audio-Szene für Encoder, zum anderen das Generieren des Control-Tracks. Eine ausführlichere Beschreibung der einzelnen Tools ist im Rahmen dieser Bachelorarbeit allerdings nicht möglich.

In einem effizienten Postproduktionsworkflow wären in modernen DAWs nicht nur die Produktionstools für Immersive Audio integriert, sondern auch die notwendigen Authoring-Funktionen für NGA-Standards. Das setzt allerdings voraus, dass die Software einen objektbasierten Produktionsansatz unterstützen. Bei Pyramix ist das bereits möglich. Dort

⁷⁹ Vgl. Jünger Audio GmbH (2017), S. 2

⁸⁰ Vgl. Grewe, Simon, Scuda (2018), S. 339

⁸¹ Vgl. The MPEG-H TV Audio System. Use Cases and Workflows (o.D.), S. 3

können mittlerweile ab Version 11.1 MPEG-H Audio-Szenen direkt von der Session aus exportiert werden.⁸²

Wird MPEG-H Audio nicht nativ innerhalb der DAW unterstützt, gibt es die Möglichkeit Plugins zu erwerben, welche in alle gängigen DAWs eingebunden werden können. Beispielsweise das *MPEG-H Authoring Plugin* (MHAPi) des Fraunhofer IIS, ist ein frei erhältliches Tool für das Authoring von MPEG-H Audio-Szenen. Mit MHAPi erstellte Presets sollen auch für die Konfiguration von AMAU-Systemen und MPEG-H Encodern verwendet werden können.⁸³ Neben der Plugin-Variante gibt es auch eine Stand-Alone Applikation, das *MPEG-H Authoring Tool* (MHAT).⁸⁴

Eine weitere Alternative bietet der Hersteller New Audio Technology seit Januar 2019 mit einem Upgrade für den *Spatial Audio Designer* (SAD) auch das Authoring und Monitoring von MPEG-H an.⁸⁵ Der SAD ist ein gängiges Tool für die Produktion von Mehrkanal-Audioformaten. Aus der Zusammenarbeit von New Audio Technology mit Lawo, entstand die sogenannte *Lawo Immersive Mixing Engine* (LIME), bei der es sich um eine adaptierte Stand-Alone Variante des SAD handelt. Diese lässt sich in alle Mischpulte der mc²-Serie einbinden und von dort problemlos steuern. Mit Lawo Mischpulten ist es dadurch möglich, immersive Audioproduktionen in allen gängigen kanalbasierten Formaten zu erstellen.⁸⁶ Naheliegender wäre demnach die Kombination von LIME und einer mc² Konsole zu einem fertigen System für die Produktion von Immersive Audio in MPEG-H. Dies wurde bislang jedoch nicht realisiert, da mit LIME, ebenso wie mit dem SAD, nur ein Offline-Authoring möglich ist. Im Live-Broadcast wird deshalb weiterhin eine AMAU benötigt.

5. Dolby Atmos

Dolby Atmos ist eine Immersive Audio-Technologie, die im Jahr 2012 von den Dolby Laboratories vorgestellt wurde und ursprünglich vor allem auf die Anwendung im Bereich Kino abzielte. Dort erweitert es die bisherigen Dolby Surround-Formate um zusätzliche Höhenlautsprecher und die Unterstützung von Audio-Objekten.⁸⁷ Ziel ist es aber offensichtlich auch, dieses Erlebnis einem breiteren Publikum zugänglich zu machen. Mittlerweile findet man zahlreiche Dolby-Technologien, welche unter dem Begriff *Dolby Atmos* geführt werden, im Equipment von Heimkino-Anlagen, Live-TV-Broadcast und Streaming, aber auch in mobilen Endgeräten. Laut Slavik setzen Streaming-Anbieter und

⁸² Vgl. Merging Technologies (2017)

⁸³ Vgl. Fraunhofer IIS (2019)

⁸⁴ Vgl. Grewe, Simon, Scuda (2018), S. 340

⁸⁵ Vgl. Fraunhofer Audio Blog (2019)

⁸⁶ Vgl. Lawo AG (2017) Lawo kooperiert mit New Audio Technology: Präsentation von LIME Immersive Mixing Engine für mc² auf der IBC 2017

⁸⁷ Vgl. Dolby Laboratories, Inc. (2014), S. 1

TV-Sender wohl recht häufig auf Dolby Atmos, da es in Heimgeräten am weitesten verfügbar sei.⁸⁸

5.1 Grundlagen

Bei Dolby Atmos wird ein kanalbasierter Produktionsansatz mit einem objektbasierten kombiniert. So gesehen handelt es sich hierbei also um ein hybrides Audioverfahren (siehe Kapitel 3.1.4).

Ein Dolby Atmos Mix besteht insgesamt aus zwei Komponenten: das *Audio-Bed* ist ein statischer, kanalbasierter Mix. Dieses Bed kann für verschiedene kanalbasierte Lautsprecherkonfigurationen, wie 5.1, 7.1 oder 9.1⁸⁹ erstellt werden. Dagegen handelt es sich bei den einzelnen *Objects* um separate Audiosignale mit Metadaten, die zusammen mit dem Audio-Bed übertragen und erst am finalen Renderer berechnet werden.⁹⁰ Somit unterstützt Dolby Atmos grundlegende Voraussetzungen, auf denen die modernen NGA-Standards im Live-Broadcast aufbauen: Immersive Audio und die objektbasierte Produktion.

Was bei MPEG-H als Audio-Szene bezeichnet wird, sind bei Dolby Atmos im Live-Broadcast die sogenannten *Presentations*. Eine Presentation besteht aus einer Kombination von verschiedenen Audio-Elementen, wie beispielsweise einem 5.1-Bed und einem zusätzlichen Objekt für den Dialog oder Kommentar. Die mögliche Bezeichnung für eine solche Presentation bei einem englischsprachigen Kommentar, wäre z.B. 5.1-English.⁹¹ Stehen mehrere Sprachen oder auch ein Immersive Audio-Mix zur Verfügung, können bei identischem Bed-Mix auch unterschiedliche Presentations erstellt werden, beispielsweise 5.1.4-English, 5.1.4-German oder 5.1.4-Spanish.

Theoretisch betrachtet kann Dolby Atmos auf unterschiedlichsten Wiedergabesystemen genutzt werden. Es kommen aber je nach Anwendungsfall verschiedene Technologien zum Einsatz.⁹² Eine große Stärke von Dolby Atmos ist daher mitunter die End-to-End Skalierbarkeit.⁹³ Immersive Audio im Bereich Postproduktion stellt zweifellos andere Anforderungen an die Infrastruktur wie die Live-Produktion im TV-Broadcast. Es muss berücksichtigt werden, dass Endgeräte, wie AV-Receiver, Smartphones oder Laptops, nicht über die gleiche Rechenleistung verfügen wie ein Dolby Prozessor im Kino. Damit die bisherige Infrastruktur im Broadcast weiterhin genutzt werden kann, müssen Dolbys Audio-Codecs daher noch effizienter werden.

⁸⁸ Slavik (2018)

⁸⁹ Also auch Konfigurationen mit Höhenkanälen, z.B. 7.1.2 oder 5.1.4.

⁹⁰ Vgl. Dolby Laboratories, Inc. (2014), S. 6-7

⁹¹ Vgl. Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. product specification, S. 6

⁹² Vgl. Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 1 - Understanding the concept

⁹³ Vgl. Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 3 - Movie Theaters

5.2 Codecs für einen Dolby Atmos Workflow

Immersive Audioformate benötigen grundsätzlich eine höhere Anzahl an Audiokanälen. Zwar gibt es mit *Dolby Digital Plus* (DD+) bereits einen Dolby-Codec, der mehr als sechs Kanäle zulässt, für Dolby Atmos müssen aber auch die Metadaten von Audio-Objekten unterstützt werden. Neben den Audiosignalen müssen daher auch deren Daten mit integriert werden. Je mehr Objekte, desto mehr Metadaten und somit Bandbreite wird benötigt. Bei Liveübertragungen ist die zur Verfügung stehende Bandbreite allerdings stark begrenzt.⁹⁴ Des Weiteren kann davon ausgegangen werden, dass viele Wiedergabegeräte beim Endverbraucher nicht in der Lage sind aufgrund der hohen Komplexität eine größere Anzahl von Audio-Objekten zu decodieren und zu rendern.

Nicht alle Audio-Codecs sind deshalb für die Produktion und Übertragung von Dolby Atmos Inhalten geeignet. Folgende Codecs (siehe Kapitel 5.2.1-5.2.3) sind speziell für den Workflow im Live-Betrieb ausgelegt und zeichnen sich durch effizientere Bitraten, sowie umfassenderes Loudness- und Metadaten-Management aus.⁹⁵

5.2.1 Dolby Digital Plus with Dolby Atmos

Bei *Dolby Digital Plus with Dolby Atmos* handelt es sich um eine neue Version von Dolbys DD+ Mehrkanal-Surround-Technology.⁹⁶ Dolby Atmos Inhalte können demnach durch eine effiziente Codierung mit DD+ übertragen werden. Dafür werden die zusätzlichen Audiosignale innerhalb der Audiokanäle des normalen 5.1-Mixes transportiert. *Dolby Digital Plus with Atmos*-fähige Decoder sind anschließend in der Lage, den ursprünglichen originalen Dolby Atmos Mix wieder zu rekonstruieren. Dieser Vorgang wird auch als *Joint Object Coding* bezeichnet⁹⁷ und wurde durch das European Telecommunications Standards Institute (ETSI) in TS 103 420 standardisiert.⁹⁸ Dementsprechend häufig findet sich an Stelle von *DD+ with Dolby Atmos* auch der eher technischere Begriff *Dolby Digital Plus with Joint Object Coding* (DD+JOC).⁹⁹

5.2.2 AC-4

AC-4 ist Dolbys neuester Audio-Codec und der logische Nachfolger von AC-3. Zusammen mit MPEG-H ist er Teil des UHD TV-Standards ATSC 3.0. Standardisiert wurde die Technologie bereits im April 2014 in ETSI TS 103 190.¹⁰⁰ Derzeit wird AC-4 zwar noch nicht ausgestrahlt, dafür wurden bereits mehrere Testimplementierungen im Live-

⁹⁴ Vgl. Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 1 - Tools for Transport

⁹⁵ Vgl. Dolby Laboratories, Inc. (2018) Dolby Professional Reference Decoder DP580. User's guide, S. 7

⁹⁶ Vgl. ebd.

⁹⁷ Vgl. Dolby Laboratories, Inc. (2017) What is Dolby Digital Plus JOC? (Joint Object Coding)

⁹⁸ Vgl. European Telecommunications Standards Institute (2016)

⁹⁹ Vgl. Dolby Laboratories, Inc. (2017) What is Dolby Digital Plus JOC? (Joint Object Coding)

¹⁰⁰ Vgl. Dolby Laboratories, Inc. (2015), S. 29

Broadcast durchgeführt. Als Beispiel können hier das UEFA Champions League Finale oder das French Open im Jahr 2015 genannt werden.¹⁰¹

Der Codec unterstützt Immersive Audio mit Objekten, Personalisierung und Dialog Enhancement. Verbraucher können so zwischen verschiedenen Dolby Presentations auswählen. Des Weiteren bietet AC-4 ebenso wie MPEG-H 3D Audio, eine integrierte Loudness- und Dynamic Range Kontrolle, sowie adaptives Rendering auf das zur Verfügung stehende Wiedergabesystem. Für die Wiedergabe von 3D-Audio auf Kopfhörern ist zudem auch ein binauraler Renderingalgorithmus integriert.¹⁰²

Einsatz findet dieser Audio-Codec hauptsächlich im Bereich Broadcast. Ebenso wie *DD+ with Dolby Atmos* soll er dort als Distributions-Codec verwendet werden. Im Vergleich zu DD+ ist der neuere AC-4-Codec allerdings deutlich effizienter. Laut Dolby kann damit, ähnlich wie bei MPEG-H 3D Audio, die gleiche Audioqualität bei halber Bitrate erreicht werden.¹⁰³

5.2.3 Dolby ED2

Dolby ED2 ist die Weiterentwicklung von Dolbys Broadcast Kontributionsformat Dolby E und unterstützt die zusätzlich benötigten Metadaten für die Übertragung von Dolby Atmos Inhalten über SDI oder Satellit.¹⁰⁴

ED2 ermöglicht den Transport von Immersive Audio inklusive Metadaten innerhalb der existierenden Broadcast Infrastruktur. Dolby E-fähige Komponenten können trotzdem weiter genutzt werden um Dolby Atmos Inhalte durchzureichen. Auch wenn Dolby E bereits grundlegende Metadaten bezüglich Lautheit und Downmixing unterstützt, benötigt Dolby Atmos deutlich mehr Audiokanäle und einen umfangreicheren Support von Metadaten. *Professional Metadata* heißt das neue Metadaten-System in Dolby ED2. Es beinhaltet sowohl Metadaten für die Handhabung von Immersive Audio, als auch für dynamische Objekte oder Personalisierung. Ermöglicht wird zudem auch eine parallele, synchrone Übertragung mehrerer ED2 Sub-Streams, womit eine höhere Anzahl an Audiokanälen möglich wird. Die Metadaten werden entweder direkt innerhalb eines ED2-Streams oder als separater Datenstrom transportiert.¹⁰⁵

Eingesetzt wird ED2 ausschließlich auf der Produktionsseite, zwischen den einzelnen Produktionsbereichen. So zum Beispiel auf der Verbindungsstrecke zwischen Übertragungswagen und Broadcast Center, bei dem die Audiosignale in hoher Qualität übertragen werden müssen. Für eine effizientere Codierung vor der Distribution, muss der

¹⁰¹ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 5 - Preparing for Dolby AC-4

¹⁰² Vgl. Dolby Laboratories, Inc. (2018) Dolby Professional Reference Decoder DP580. User's guide, S. 8

¹⁰³ Vgl. ebd., S. 7

¹⁰⁴ Vgl. Dolby Laboratories, Inc. (o.D.) Dolby ED2, S. 1

¹⁰⁵ Vgl. ebd.

ED2-Stream in einen stärker komprimierten Audio-Codec, nach DD+JOC oder AC-4 transcodiert werden.¹⁰⁶

5.3 Dolby Atmos Produktionstools

Aktuell bietet Dolby drei zusätzliche Hardwaregeräte an, die in Kombination miteinander für die Erstellung und Übertragung von Dolby Atmos Inhalten im Broadcast entwickelt wurden. Sie ähneln in ihrer Funktionalität den früheren 5.1-Broadcast-Tools. Zu ihren Aufgaben zählt Authoring und Monitoring, aber auch die Encodierung, Decodierung oder Transkodierung verschiedener Dolby-Codecs.

5.3.1 DP580

Abbildung 10 zeigt einen Dolby Professional Reference Decoder DP580. Bei diesem handelt es sich um einen Referenz Decoder der zur Qualitätskontrolle innerhalb der Produktionskette von Dolby Atmos eingesetzt werden kann. Einen vergleichbaren Zweck im Surround-Broadcast erfüllte der frühere Reference Decoder DP564.¹⁰⁷



Abb. 10 Dolby Professional Reference Decoder DP580¹⁰⁸

Der DP580 kann einen Großteil der aktuellen Dolby-Codecs (siehe Kapitel 5.2) decodieren. Dazu gehören neben DD+, DD+ with Dolby Atmos, Dolby E und ED2 auch der neueste AC-4 NGA-Codec.¹⁰⁹

In der Broadcast Umgebung wird er auch eingesetzt, um das Verhalten der in den Verbraucher-Endgeräten verbauten Dolby Atmos Decoder zu emulieren. Möglicherweise beim Decoding- oder Rendering-Prozess auftretende Artefakte können so bereits frühzeitig auf der Produktionsseite erkannt werden. Aufgrund der Latenz zwischen Encoding und Decoding, ist ein grundsätzliches Monitoring über dieses Tool nicht zu empfehlen. Stattdessen eignet sich der DP580 vor allem für stichprobenartige Qualitätskontrollen.¹¹⁰

¹⁰⁶ Vgl. Dolby Laboratories, Inc. (o.D.) Dolby ED2, S. 2

¹⁰⁷ Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby's Content Creation Tools

¹⁰⁸ Bildquelle: Dolby Laboratories, Inc. (2018), Dolby DP580 Front Panel

¹⁰⁹ Vgl. Dolby Laboratories, Inc. (2018) Dolby Professional Reference Decoder DP580. User's guide, S. 7

¹¹⁰ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 2 - DP580-Reference Decoder

5.3.2 DP590

Der DP590 ist ein Echtzeit Authoring-Tool von Dolby. In der Regel hat der zuständige Toningenieur mit diesem Gerät am meisten zu tun, da hier die wesentlichen Elemente einer Dolby Atmos Mischung im Live-Broadcast definiert werden.¹¹¹ Die Hardware Unit ist dabei identisch zu der des DP591 Encoders (vgl. Abb. 12 in Kapitel 5.3.3). Im Prinzip ist der DP590 ein logischer Nachfolger des DP570 Metadata Authoring Tools.¹¹²

Über eine MADI Verbindung mit dem Produktionsmischpult werden die Audiosignale zum Authoring-Tool gesendet, dort verarbeitet und die gerenderten Mischungen wieder an die Monitoring Sektion der Konsole zurückgeschickt. Die Metadaten werden gleichzeitig per Ethernet an einen oder optional zwei DP591 Encoder übermittelt.¹¹³ Zwingend notwendig ist, dass die Geräte untereinander korrekt synchronisiert werden. Als Referenz dient dafür meist das vorhandene Produktionsmischpult. Der Clock-Status lässt sich entweder in der Webapplikation überprüfen oder am Front-Panel der Hardware Unit ablesen.¹¹⁴

Zugriff auf die Konfigurationseinstellungen und Authoring-Funktionen des DP590 erhält man über ein webbasiertes User Interface (siehe Abb. 11).



Abb. 11 Webapplikation des DP590¹¹⁵

¹¹¹ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 2 - DP590-Monitoring&Authoring

¹¹² Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby's Content Creation Tools

¹¹³ Vgl. Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. product specification, S. 12-13

¹¹⁴ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Audio Connectivity

¹¹⁵ Bildquelle: Dolby Atmos - Live TV Production Training, Webapplikation des DP590

Dieses lässt sich einfach über die IP-Adresse des DP590 in einem normalen Webbrowser aufrufen, vorausgesetzt beide Geräte befinden sich im gleichen Netzwerk.¹¹⁶ Alle DP590 Konfigurationsdaten werden in einer Session gesammelt und können gespeichert oder wieder aufgerufen werden.¹¹⁷

In der Webapplikation können die Eingangs- und Ausgangspegel der Audiosignale, sowie die Programm-Lautheit der aktuell angewählten Presentation überwacht werden. Eingangsseitig stehen bis zu 64 Audiokanäle zur Verfügung, deren Pegel jeweils in 16er Blöcken auf vier schaltbaren Ebenen dargestellt werden (siehe Abb. 11 oben links). Ausgangsseitig sind nur die ersten 20 MADi-Kanäle als Monitoring Wege in Verwendung. Diese sind für den Dolby Atmos Mix (1-10), einen Surround-(11-16) und Stereo-Downmix (17-18) sowie einer binauralen Version (19-20) für Kopfhörer reserviert. Die jeweiligen Mischungen sind voneinander unabhängig und können individuell angepasst werden.¹¹⁸

Presentations und Audio-Objekte werden ebenfalls über das User Interface erstellt. Eingangssignale werden dort als kanalbasiertes Bed oder Objekt definiert. Pro Session gibt es immer nur ein Audio-Bed.¹¹⁹ Bei den Objekten gibt es unterschiedliche Auswahlmöglichkeiten mit denen zusätzliche Features verbunden sind. Ein Audiosignal kann beispielsweise als *Mono Dialogue* Kanal deklariert werden, womit Dialog Enhancement-Funktionen unterstützt werden.¹²⁰ Damit ein Audiosignal als Objekt verarbeitet werden kann, muss dieses jedoch einzeln an der Authoring Unit anliegen. Die Audio-Objekte lassen sich im Übrigen auch frei im 3D-Raum positionieren. Deren Positionsdaten können in den Metadaten mit übertragen werden.

Grundsätzlich ist es möglich, mit dem DP590 mehrere Presentations innerhalb einer Session zu erstellen. Im Verbund mit Dolbys ED2 und AC-4 in der Produktions- und Übertragungskette ist es somit möglich, Immersive Audio sowie personalisierte Inhalte an den Verbraucher zu übertragen.

5.3.3 DP591

Bei dem DP591 (siehe Abb. 12) handelt es sich um Dolbys aktuellen Audio Encoder für den professionellen Audio-Broadcast. Auch dieser basiert auf ehemaligen Surround-Prozessoren, wie dem DP569, DP571 oder DP572. In diesem Tool sind allerdings Encodierungs-, Decodierungs- und Transcodierungsfunktionen vereint.¹²¹

¹¹⁶ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Network Connectivity

¹¹⁷ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Clearing Session Data

¹¹⁸ Vgl. Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. product specification, S. 31-32

¹¹⁹ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Preparing the Bed

¹²⁰ Vgl. Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 5 - Creating Dialogue Channels

¹²¹ Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby's Content Creation Tools



Abb. 12 Dolby Professional Audio Encoder DP591¹²²

Eingesetzt wird der DP591 überall dort, wo die Audiosignale vom Mischpult zusammen mit den Metadaten des DP590 in einen gemeinsamen ED2-Stream en- bzw. decodiert werden müssen, wie es beispielsweise im Kontributionsbereich zwischen Ü-Wagen und Sendezentrum der Fall ist. Weitere Aufgaben betreffen die Transkodierung von ED2 in DD+JOC oder PCM-Audio und Dolby E in DD+ in der Distribution.¹²³

Der DP591 erhält vom Mischpult die gleichen Audiosignale wie der DP590. Beide Geräte sind zudem in einem gemeinsamen Netzwerk miteinander verbunden. Das bedeutet allerdings nicht, dass die Hardware Units notwendigerweise am gleichen Produktionsort sein müssen. Häufig werden im Live-Broadcast zwei Encoder parallel eingesetzt, um bei einem Ausfall gegebenenfalls auf den zweiten DP591 zurückzugreifen.¹²⁴

6. Workflow-Analyse

In diesem Kapitel wird der Workflow von immersiven Live-TV-Broadcasts mit Produktionsmischpulten analysiert. Zunächst wird dafür ausgehend von einem sehr allgemeinen Model der Workflow von Live-Broadcasts schematisch erläutert. Anschließend wird die gesamte Produktionskette für MPEG-H und Dolby Atmos Produktionen im Detail betrachtet.

Durch die Analyse sollen alle wesentlichen Aspekte entlang der Infrastruktur untersucht und die Aufgabenbereiche der beteiligten Komponenten geklärt werden. Ausgehend davon sollen Rückschlüsse gezogen werden, an welchen Stellen gegebenenfalls noch Optimierungsbedarf besteht oder welche Anforderungen an die Produktionsmischpulte in zukünftigen Produktionsszenarien gestellt werden. Letzteres wird in Kapitel 7 ausführlich thematisiert.

¹²² Bildquelle: Dolby Atmos - Live TV Production Training, DP591

¹²³ Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - DP591 Broadcast Processor

¹²⁴ Vgl. Slavik (2018)

6.1 Broadcast-Workflow im Allgemeinen

In der folgenden Abbildung sind die wesentlichen Bestandteile einer Broadcast-Übertragungskette schematisch dargestellt.

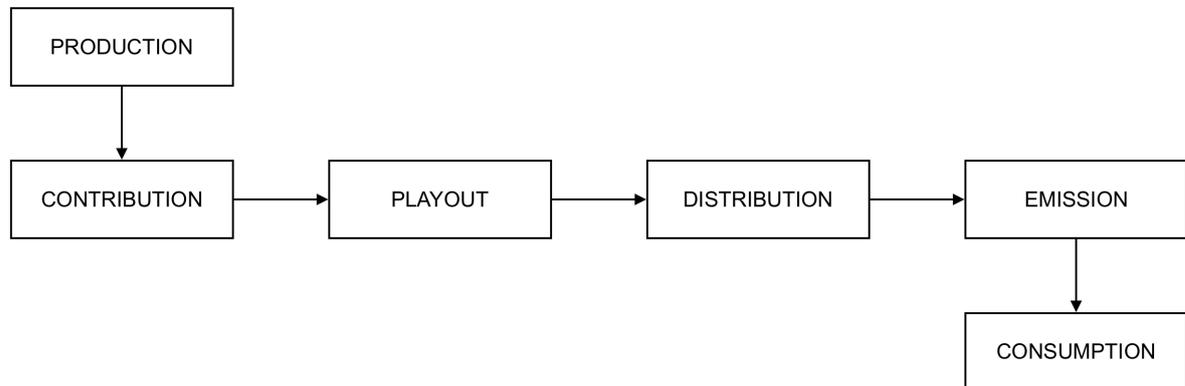


Abb. 13 Allgemeiner Broadcast-Workflow¹²⁵

Jeder Broadcast-Workflow beginnt zunächst an einem Produktionsort, in der die ersten Programminhalte generiert werden. Im Bereich der Außenübertragung wäre das beispielsweise der Ü-Wagen.¹²⁶ Von dort werden die Signale direkt an das Sendezentrum, auch Broadcast- oder Playout-Center, weitergeleitet. Diesen Vorgang bezeichnet man allgemein als Kontribution (eng. Contribution). Das Sendezentrum mit dem *Master Control Room* ist im Wesentlichen verantwortlich für einen koordinierten Sendebetrieb und stellt das Programmsignal zusammen. Anschließend wird dieses Signal auf unterschiedlichen Distributionswegen verteilt. Bei TV- und Radio-Netzwerken erfolgt diese Bereitstellung in der Regel über Satellit.¹²⁷

Im nordamerikanischen Mediensystem wird das Sendesignal erst an sogenannte *Affiliates* übertragen. Dabei handelt es sich um lokale Rundfunk-Sendeanstalten, die das Distributionssignal vom Broadcast-Center erhalten und um lokale Inhalte (z.B. Lokalnachrichten, Werbung) ergänzen.¹²⁸ Die Emission ist die letzte Station in der Übertragungskette zu den Verbraucher-Endgeräten. Im Gegensatz zu Kontributions- und Distributionskanälen handelt es sich bei der Emission nicht um eine Punkt-zu-Punkt-, sondern um eine Punkt-zu-Mehrpunkt-Verbindung. Des Weiteren sind die Anforderungen an die Audioqualität vor der Emission deutlich höher, da in diesem Bereich gegebenenfalls noch Eingriffe in die Produktionssignale erfolgen.¹²⁹

¹²⁵ Bildquelle: Eigene Abbildung

¹²⁶ Vgl. The MPEG-H TV Audio System. Use Cases and Workflows (o.D.), S.4

¹²⁷ Vgl. Pizzi, Kean (2018), S. 911

¹²⁸ Vgl. Wikipedia (2018) Affiliate (Rundfunk)

¹²⁹ Vgl. Pizzi, Kean (2018), S. 911

6.2 Workflow für MPEG-H

„The MPEG-H 3D Audio standard describes a state of the art audio codec and renderer, but does not address how that codec is used in actual end-to-end systems.“¹³⁰

Zur Überprüfung des MPEG-H Systems wurde deshalb von der MPEG-H Audio Alliance, eine Kollaboration von Fraunhofer Institut, Qualcomm und Technicolor, ein Testmodell entwickelt, welches 2015 auf der NAB und einer ATSC-Veranstaltung präsentiert wurde. Dieses sollte ein praktikables TV Audio System in einem kompletten MPEG-H End-to-End Workflow mit der zur Verfügung stehenden Infrastruktur abbilden.¹³¹ Der gesamte Aufbau und Signalfluss sowie die daran beteiligten Komponenten, sind im folgenden Kapitel 6.2.1 beschrieben.

6.2.1 Signalfluss von MPEG-H im Live-Broadcast

Das Grundkonzept der MPEG-H Audio Alliance basiert auf einem gewöhnlichen nordamerikanischen TV-Broadcast. Weiterführend wurde der Ü-Wagen hier lediglich um ein 3D-Audio Monitoring-System und eine AMAU erweitert. In der Ton-Regie des Ü-Wagens werden die eingehenden Mikrofonsignale verarbeitet und von einem Toningenieur gemischt. In dem folgenden Beispiel handelte es sich allerdings um bereits aufgenommene Audiospuren. Gesteuert wird der Remote Truck von dem *Network Operations Center*. Über einen Lokalsender wird dem Internationalen Sendesignal zusätzlich passende Werbung hinzugefügt. Auf der Wiedergabeseite landet der Stream dann an der Set-Top-Box beim Endverbraucher, wo der MPEG-H Audio Stream decodiert wird und die einzelnen Lautsprechersignale gerendert werden.¹³²

Neben MPEG-H als Audio-Codec wurde für das Videobild das Advanced Video Coding-Format (AVC) verwendet. Beide Signale (MHAS und AVC) wurden in einem gemeinsamen MPEG-2 Transport Stream (TS) verpackt und zwischen den vorher beschriebenen Stationen über ein IP-Netzwerk übertragen. Der Grund, weshalb MHAS über MPEG-2-TS übertragen wurde, lag daran, da das Übertragungsprotokoll nicht in ATSC standardisiert wurde.¹³³ Seit April 2016 sind die ersten professionellen, MPEG-H-fähigen Encoder für TV-Broadcasts verfügbar.¹³⁴ Diese erzeugen aus den einzelnen PCM-Audiosignalen sowie den von der AMAU generierten Encoder-Konfigurationsdaten und Audio-Metadaten, den MPEG-H Audio Stream. In einer statischen, einfacheren Konfiguration, kann sogar ganz auf eine AMAU verzichtet werden. Die Einstellungen dafür können direkt am Encoder vorgenommen werden.¹³⁵

¹³⁰ Bleidt et al. (2017), S. 224

¹³¹ Vgl. ebd., S. 227

¹³² Vgl. ebd.

¹³³ Vgl. ebd.

¹³⁴ Vgl. Fraunhofer Audio Blog (2016)

¹³⁵ Vgl. The MPEG-H TV Audio System. Use Cases and Workflows (o.D.), S. 3

Neben den Broadcast-Encodern und -Decodern müssen auch die Verbraucher-Endgeräte MPEG-H unterstützen (siehe Abb. 14). Des Weiteren wird überall dort, wo Monitoring, Authoring oder eine Qualitätskontrolle stattfinden soll, ein AMAU-System benötigt. Für das Testmodell kommt dafür die MMA von Jünger Audio zusammen mit einem zusätzlichen Joystick Controller zum Einsatz. Ausführlichere Informationen zu diesen Authoring und Monitoring Units sind in Kapitel 4.3.1 dargestellt.

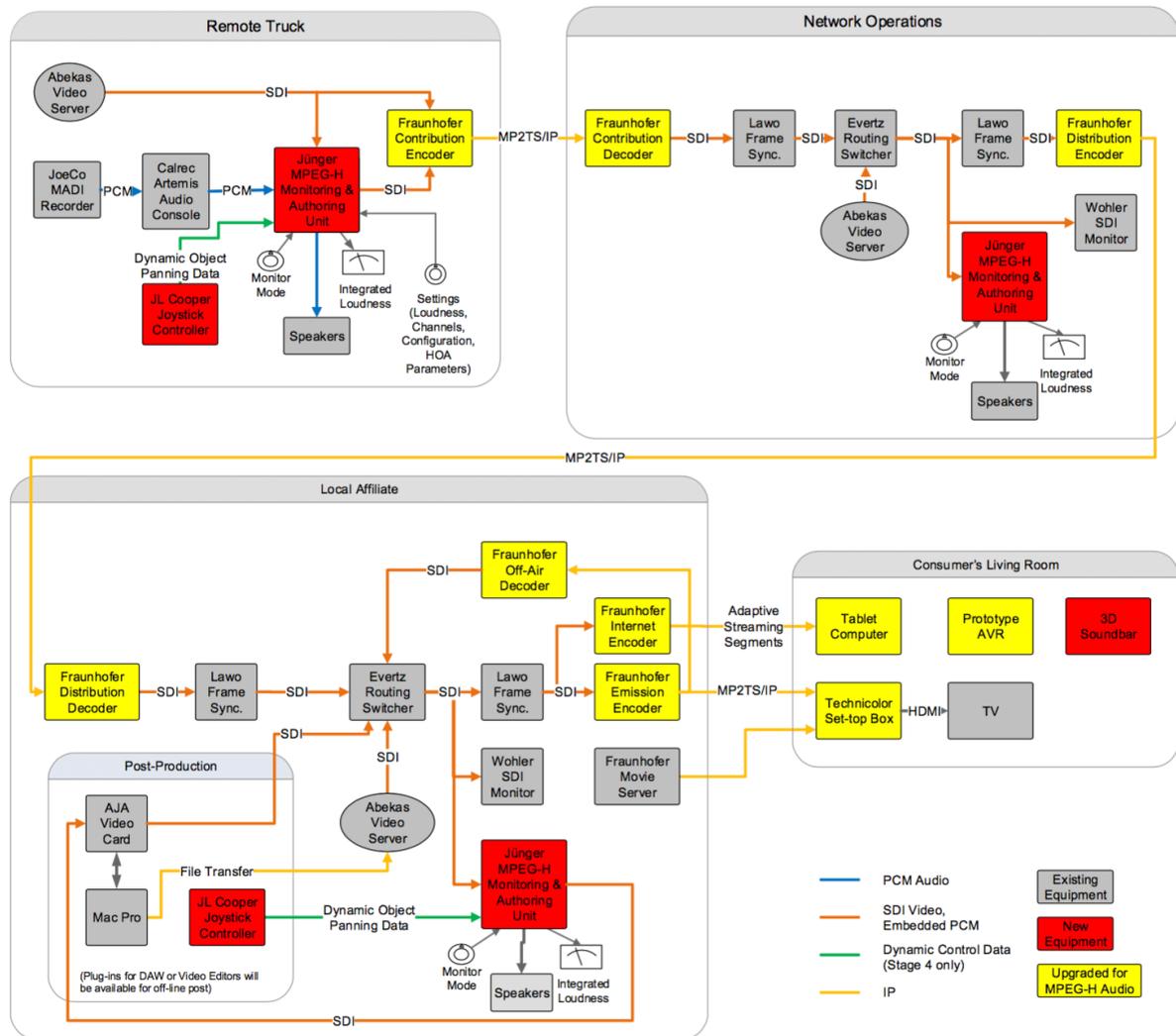


Abb. 14 Blockdiagramm MPEG-H Testimplementierung¹³⁶

Vieles deutet darauf hin, dass die Infrastruktur im TV-Broadcast sich in Zukunft mehr und mehr in Richtung IP-Vernetzung entwickelt. Derzeit basiert die Distribution hauptsächlich noch auf SDI. Damit können 16 Audiokanäle in das Signal eingebettet werden. Das ist für die meisten MPEG-H 3D Audio Produktionen ausreichend. Allerdings muss berücksichtigt werden, dass auch Metadaten transportiert oder gespeichert werden müssen. Bei einfacheren Produktionen können die Kontributions- und Distributions-Encoder in der Regel mit statischen Einstellungen arbeiten. Sobald jedoch dynamische Metadaten

¹³⁶ Bildquelle: Bleidt et al. (2016), S. 3

übertragen werden sollen, wie etwa Positionsdaten von bewegten Audio-Objekten, müssen die Metadaten in Form eines Control-Tracks direkt in das SDI-Signal eingebunden werden.¹³⁷

6.2.2 Anpassung der Produktionsumgebung für MPEG-H

Eine wichtige Voraussetzung für einen funktionierenden MPEG-H End-to-End Workflow, noch bevor die Produktionsumgebung auf MPEG-H umgerüstet wird, sind MPEG-H-fähige Decoder in den Verbraucher-Endgeräten. Dasselbe gilt für die Dolby Audio-Codecs. Sind diese Bedingungen erfüllt, können Rundfunkanstalten entscheiden, welche zusätzlichen Tools und Funktionen integriert werden sollen. In einem weiteren Schritt müssen die bisherigen Encoder zunächst durch MPEG-H-fähige Encoder ersetzt werden. Das hat zum einen den Vorteil, dass bei der Übertragung sofort Bandbreite eingespart werden kann, da die gleiche Audioqualität bereits bei halber Bitrate erreicht wird. Gleichzeitig können die neuen Features von immersiven MPEG-H-Audioproduktionen unterstützt werden.¹³⁸

Für Live-Produktions-Anwendungen von MPEG-H 3D Audio wird in der Regel eine AMAU (vgl. Kapitel 4.3.1) benötigt, welche sich einfach in bestehende Workflows integrieren lässt. Sämtliche Authoring- und Monitoring-Funktionen für Immersive Audio werden dort verarbeitet. Abbildung 15 zeigt, wie ein Produktionsmischpult zusammen mit der Authoring Unit in einen MPEG-H Live-Broadcast eingebunden wird.

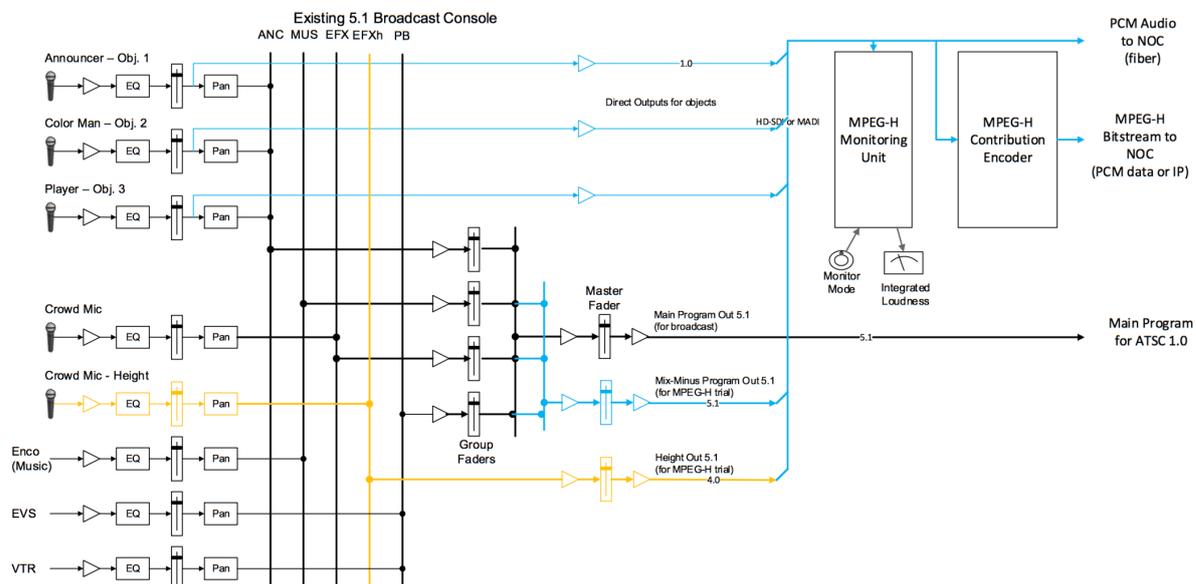


Abb. 15 Einbindung eines Produktionsmischpultes für den MPEG-H Audio Broadcast¹³⁹

¹³⁷ Vgl. Bleidt et al. (2017), S. 225

¹³⁸ Vgl. Bleidt (2014), S. 24-28

¹³⁹ Bildquelle: Bleidt, Baxter (2014), S. 15

In der Darstellung ist ebenfalls zu sehen, dass Signale, die später als Audio-Objekte gehandhabt werden sollen (z.B. Announcer), direkt über die Direct Outs an die Authoring Unit geschickt werden. Alle anderen anliegenden Audiosignale von Mikrofonen und Zuspiegeln werden zunächst in Gruppen zusammengefasst. Direkt von der Produktionskonsole aus, wird ein fertiger 5.1-Surround-Mix für den Standard ATSC 1.0 Broadcast erstellt. Die gleiche 5.1-Mischung, nur ohne Kommentator, wird zusammen mit den Höhenkanälen eines 3D-Mikrofonarrays auch für den immersiven 5.1.4-Audio-Mix verwendet.

Die Grundlage für immersive Audioproduktionen ist ein verlässliches Monitoring-System. Eine genaue Spezifikation, wie dieses in Produktionsumgebungen für MPEG-H 3D Audio Broadcasts genau konfiguriert werden muss, ist nicht vorhanden. Bleidt und Baxter empfehlen in ihrer Präsentation auf dem SVG Summit im Dezember 2014 wenn möglich auf bestehende Surround-Konfigurationen (5.1, 7.1) zurückzugreifen und diese um vier zusätzliche Höhenlautsprecher über den Ecklautsprechern zu ergänzen.¹⁴⁰ Auf einem Workshop zum Thema „MPEG-H Audio System for Broadcasting“ wurde dagegen bei der Lautsprecherkonfiguration auf ITU-R BS.2051 verwiesen.¹⁴¹ Demnach sollen alle Lautsprecher den gleichen Abstand zur Hörposition aufweisen oder die Laufzeiten und Pegel manuell angepasst werden. Das widerspricht sich mit der vorherigen Empfehlung von Bleidt und Baxter. An sich ist bei der objektbasierten Audiowiedergabe eine Flexibilität bei der Positionierung legitim. Da im 3D-Audio Live-Broadcast aber meist auch mit einem kanalbasierten Audio-Bed gearbeitet wird, sollte für die bestmögliche Kompatibilität allgemein auf eine einheitliche Spezifizierung geachtet werden. Eine gute Basis dafür sind die in CIEP ISO/IEC 23001-8 aufgeführten Lautsprecherkonfigurationen.

Kommt es aufgrund unterschiedlicher Faktoren zu der Tatsache, dass ein 3D-Audio-Monitoring-System am Produktionsort fehlt, ließe sich als Ersatzlösung auch eine Remote-Produktion von einem externen Studio durchführen. Alternativ wäre auch das binaurale Monitoring über Kopfhörer für grundlegende Qualitätskontrollen eine Option.

6.2.3 French Open 2018

Ein Test für einen MPEG-H Live-Broadcast unter realen Bedingungen, fand während den French Open 2018 statt. Das Ereignis diente unter anderem als Kulisse für einen Test-Broadcast mit den neuesten am Markt verfügbaren Video- und Audio-Technologien. In Zusammenarbeit mit dem Fraunhofer IIS übertrug der französische öffentlich-rechtliche Sender France Télévisions alle Spiele des Turniers in UHD mit *High Dynamic Range* (HDR) sowie 3D-Audio mit zahlreichen zusätzlichen interaktiven Features. Empfangen werden konnte das Sendesignal zum einen mit DVB-T2 im Pariser Sendegebiet, zum

¹⁴⁰ Vgl. Bleidt, Baxter (2014) S. 10

¹⁴¹ Vgl. Plogsties (2015), S. 16

anderen über Satellit mit DVB-S2 in ganz Frankreich. Für den internationalen Markt wurde weiterhin eine Stereo und Surround-Mischung produziert.¹⁴²

Der internationale Stereo- und 5.1-Mix lag zusammen mit den Höhenkanälen von einem 3D-Audio Mikrophonarray (ORTF-3D) und den einzelnen Signalen der Kommentatoren, über eine MADI-Verbindung an der AMAU an. Hier wurde ein kanalbasiertes 5.1.4-Audio-Bed definiert und zwei Dialog Kanäle (French, English) erstellt. Des Weiteren standen verschiedene Presets zur Auswahl, wie *Default*, *Dialog+* oder *Venue*, und bestimmte User-Parameter wurden gezielt vorkonfiguriert. Beispielsweise gab es einen Regler für *Prominence* mit dem das Verhältnis zwischen Kommentator und 5.1.4-Bed bei gleichbleibender Programm-Lautheit für eine bessere Sprachverständlichkeit individuell angepasst werden kann.¹⁴³

Von der AMAU wurden die Metadaten ausgabeseitig in einen Control-Track moduliert und zusammen mit den einzelnen Audiosignalen an einen SDI-Embedder geschickt. Dieser erzeugte zusammen mit dem Videosignal einen SDI-Stream für zwei redundante Encoder, welche ein HDR-Video-Format und MPEG-H Audio generierten. Über eine IP-Verbindung gelangten diese Signale zur Ausstrahlung an die Satellit- und Terrestrial-Sendeanstalten.¹⁴⁴

Interessant ist hierbei, dass für die Produktion kein Ü-Wagen mit 3D-Audio Abhöre zur Verfügung stand. Das Monitoring der Höhenkanäle fand zeitweise über die Surround-Abhöre statt. Lautstärkeverhältnisse zwischen oberer und unterer Ebene konnten so natürlich nicht beurteilt werden. Deshalb wurde die immersive Audio-Mischung zusätzlich in einem separaten Raum mit extra installierter 3D-Abhöre überwacht und das Feedback wiederum an den Mixing-Engineer im Ü-Wagen weitergegeben.¹⁴⁵ Das Beispiel zeigt auch, dass die Produktion von Immersive Audio in Außenübertragungen teilweise ohne 3D-Abhöre in Ü-Wagen möglich sind.

6.3 Workflow für Dolby Atmos

Eines der vielen Ziele von Immersive Audio ist es, den Zuhörer komplett in die Audio-Szenerie einzutauchen zu lassen und das Gefühl zu vermitteln mehr integriert zu sein.¹⁴⁶ Bei einigen Sport-Übertragungen wird deshalb mithilfe von Dolby Atmos versucht, die Stadion-Atmosphäre und das Publikum bestmöglich einzufangen.

¹⁴² Vgl. Simon et al. (2018), S. 4

¹⁴³ Vgl. ebd., S. 4-5

¹⁴⁴ Vgl. ebd., S. 5

¹⁴⁵ Vgl. ebd., S. 4

¹⁴⁶ Vgl. Roginska, Geluso (2018), S. 1

„For me, Dolby Atmos is all about placing you in the audience. What you’re capturing is that big stadium atmosphere around you.“¹⁴⁷

(Richard Williams, Senior Audio Engineer at Telegenic)

Die Produktion von Dolby Atmos als immersives Audioformat lässt sich daher nicht ausschließlich mithilfe zusätzlichen technischen Equipments bewerkstelligen, sondern erfordert auch neue Ansätze in der Konzeption. Wie die Umsetzung davon im Live-Broadcast aussehen kann, wird im Folgenden näher erläutert.

6.3.1 Signalfluss von Dolby Atmos im Live-Broadcast

Immersive Audio mit Dolby Atmos in Liveübertragungen wird bereits seit mehreren Jahren erprobt. Führend dabei ist unter anderem der englische Sportsender BT Sports, der auf seinem 4K UHD-Sender die Fußballspiele der Premier League seit Januar 2017 in Dolby Atmos produziert und sendet. Daran sieht man, dass die Einsatzgebiete von Dolby Atmos sehr vielfältig sein können. Neben Fußballspielen wurde Dolby Atmos bislang auch im Basketball oder beim Motorsport im NHRA Drag Racing in den USA getestet.¹⁴⁸

Eine schematische Darstellung für den Workflow bei einer Dolby Atmos Live-Produktion ist in Folgenden abgebildet (siehe Abb. 16). Ausgehend von der Produktionsumgebung vor Ort werden die Signale über Glasfaser oder Satellit an das Playout-Center übermittelt. Ein zentraler Bestandteil der Kontribution ist dabei der ED2 Codec. Vor der finalen Ausstrahlung wird dieser Stream jedoch in einen Bandbreite sparenden Codec (DD+JOC oder AC-4) encodiert und erst danach an die Set-Top-Box in den Empfängerhaushalten gesendet.

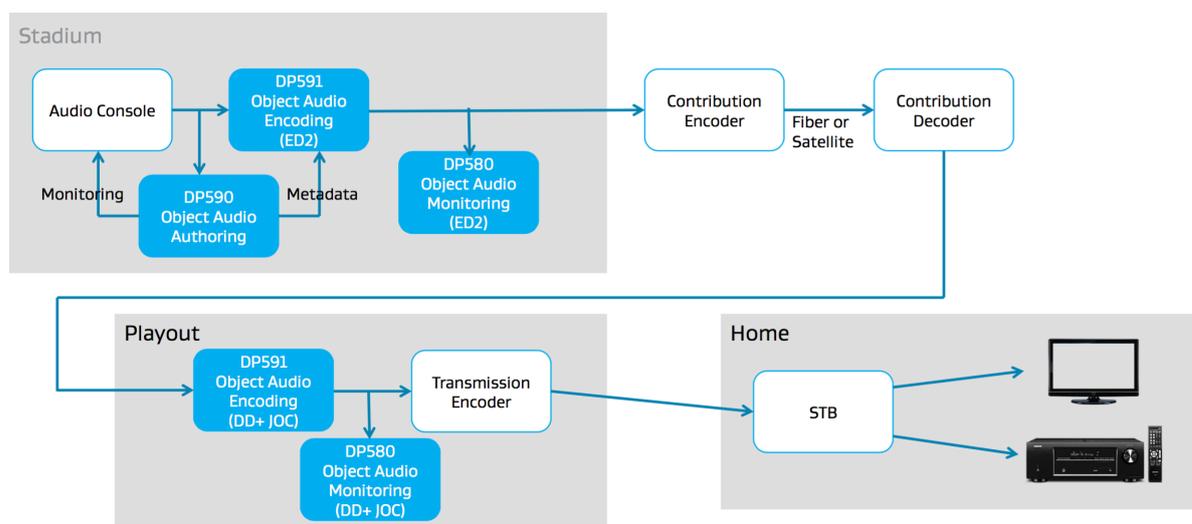


Abb. 16 Blockdiagramm eines Dolby Atmos Broadcast-Workflows¹⁴⁹

¹⁴⁷ Dolby Laboratories, Inc. (2018) How Dolby and Telegenic are bringing ringside seats into boxing fan’s homes, S. 2

¹⁴⁸ Vgl. Slavik (2018)

¹⁴⁹ Bildquelle: Dolby Laboratories, Inc. (2017) Dolby Atmos Immersive Audio From the Cinema to the Home, S. 43

Aus der Abbildung geht hervor, dass es sich um einen Dolby Atmos Workflow bei einer Außenübertragung handelt. Die Produktionsumgebung befindet sich dort häufig in einem Ü-Wagen. Alle einkommenden Mikrofonsignale liegen zunächst an dem vorhandenen Produktionsmischpult an, werden aber letztendlich einzeln oder in Subgruppen über MADI an Dolbys Authoring-Tool weitergeleitet. Dort findet der eigentliche Mixing-Vorgang statt. Die gerenderten Presentations werden für Monitoringzwecke über die MADI-Verbindung wieder zurück zum Produktionsmischpult transportiert und von dort über die Ausspielwege ausgegeben.

Das Videobild landet ebenso wie die einzelnen Audiosignale über einen Audio-Embedder an Dolbys DP591 Encoder, wo sie zusammen mit den vom DP590 erstellten Metadaten in einen ED2-Stream verpackt und beispielsweise über eine HD-SDI-Strecke weitergeleitet werden.¹⁵⁰ Für die Überwachung dieses Streams lässt sich optional noch vor dem Kontributions-Encoder ein zusätzliches Monitoring-System bestehend aus einem Referenz Decoder (DP580) einbauen. Auch im Playout-Center finden sich wieder Dolbys Encoding- und Monitoring-Systeme. Es können auch mehrere Sprünge in der Produktionskette, beispielsweise bei einem Wechsel des Produktions- oder Sendezentrums, über die Kontributions-Encoding- und Decoding-Systeme bewerkstelligt werden. Gleiches ist auch im MPEG-H Workflow möglich.

6.3.2 Anpassung der Produktionsumgebung für Dolby Atmos

Für die Produktion und Übertragung von Dolby Atmos müssen die Übertragungswagen entsprechend angepasst werden. Dazu werden sie um die benötigte Dolby Infrastruktur ergänzt und die in der Regel bereits vorhandenen Surround-Abhören um vier zusätzliche Höhenlautsprecher zu einem 5.1.4-System erweitert. Dieses hat sich in verschiedensten Testimplementierungen bewährt und wird von Dolby für die Live-TV-Produktion empfohlen. Genauere Spezifikationen über die Positionierung, Auswahl und Kalibrierung der Lautsprecher sowie der Raumabmessungen und Akustik werden von Dolby vorgegeben.¹⁵¹

Allerdings ist zu berücksichtigen, dass die Ton-Regien von Ü-Wagen diese Vorgaben nicht immer erfüllen können, weshalb in der Regel Kompromisse einkalkuliert werden müssen. Um weiterhin eine breite Kompatibilität zu den verschiedenen Wiedergabesystemen zu gewährleisten, ist es immer notwendig auch eine 5.1- und Stereo-Mischung anzubieten. Im Monitoring sollte man deshalb zwischen Dolby Atmos, 5.1 oder Stereo per Knopfdruck umschalten können.

¹⁵⁰ Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby Atmos OB Workflow

¹⁵¹ Vgl. Dolby Atmos - Live TV Production Training, Room Installation. Dolby Atmos Speaker Installation and Room Setup Guidelines for Live Events

6.4 Ergebnisse der Analyse

Aus der Analyse geht hervor, dass immersive Audioproduktionen für moderne Standards in Liveübertragungen zum einen technisch realisierbar und zum anderen in die bestehenden Produktionsumgebungen integrierbar sind.

Sowohl für MPEG-H als auch für Dolby Atmos wurden Produktionsszenarien in der Vergangenheit erprobt und evaluiert. Teilweise ist man damit sogar bereits auf Sendung. Neben dem klanglichen Mehrwert durch Immersive Audio, nehmen gleichzeitig die Personalisierungs- und Interaktionsmöglichkeiten im Broadcast zu. Moderne Standards wie MPEG-H und Dolby Atmos unterstützen diese Entwicklung. In einer Audioproduktionsinfrastruktur, die überwiegend kanalbasiert geprägt ist, stellt das allerdings zusätzliche Herausforderungen an die Produktion.

Noch unterstützen aktuell im Einsatz befindliche Produktionsmischpulte in der Regel keine immersive Audioproduktion mit integrierten 5.1.4-Bussen und 3D-Panning. Monitoring- sowie Authoring-Funktionen von Audio-Objekten fehlen ebenfalls. Diese Konsolen können nicht so einfach zeitnah ersetzt werden, zudem es auf dem Markt derzeit dafür kaum Alternativen gibt. Deshalb wurden für beide Standards jeweils eigene Monitoring- und Authoring Units entworfen, welche die Produktionsumgebungen einerseits mit geringem Aufwand um die notwendigen Tools erweitern. Allerdings werden dadurch viele ursprüngliche Aufgaben von Produktionsmischpulten ausgelagert, da nun zwei Systeme gleichzeitig konfiguriert, gesteuert und überwacht werden müssen. Jede Einbindung von zusätzlichem Equipment erhöht die Komplexität und birgt zudem die Gefahr einer neuen Fehlerquelle. Gerade bei Liveübertragungen möchte man das natürlich vermeiden.

7. Anforderungen an die Produktionsmischpulte

Damit die Produktion von Immersive Audio sowie die Einbindung in einen Dolby Atmos- oder MPEG-H Workflow reibungslos funktionieren kann, sollten die Produktionsmischpulte bestimmte Anforderungen erfüllen. Idealerweise würden aktuelle Mischpulte immersive Audioproduktionen bereits vollständig unterstützen. Aus der Workflow-Analyse geht allerdings hervor, dass für beide Standards die Produktionsumgebungen stattdessen mit zusätzlichen externen Authoring- und Monitoring-Systemen ausgestattet werden müssen. Leider sind diese Systeme sowie deren Workflow und die Signalverarbeitung untereinander nicht kompatibel. Solange im Produktions- und Kontributionsbereich kein einheitlicher offener Standard eingeführt wird, ist anzunehmen, dass beide Systeme am Markt ihre Koexistenz fortführen.

Wesentliche Produktionsprozesse werden also in diese Outboard-Komponenten verlegt, wodurch das Produktionsmischpult an und für sich an Bedeutung verliert. Mit Blick auf zukünftige Generationen von Mischpulten stellt sich demnach die Frage, welche Funktionen moderne Produktionsmischpulte anbieten müssen und in welche Richtung die Entwicklung gehen sollte.

7.1 3D-Panning und Bussystem

Die Produktion von Immersive Audio lässt sich mit allen 5.1-fähigen Mischpulten umsetzen. Ein Produktionsmischpult muss nicht zwingend 3D-Audio-Busse und -Panning zur Verfügung stellen. Einfache kanalbasierte Mischungen, beispielsweise für eine 5.1.4-Lautsprecherkonfiguration, können auch über zwei Subgruppen realisiert werden. Beispielsweise repräsentiert über ein 5.1-Surround-Layer und ein 4.0-Höhen-Layer (vgl. Kapitel 6.3.2, Abb. 17). Deren Ausgänge werden einfach auf die jeweiligen Monitoringwege geroutet. Die Audio-Elemente können auch zwischen den beiden Ebenen verteilt werden sollen, indem die Signale anteilig auf die jeweiligen Lautsprecher in beiden Ebenen gesendet werden. Das Panning ist demzufolge vergleichsweise umständlich. Dieser Workflow hat aber nach wie vor den Vorteil, dass die Höhenebene separat gut kontrolliert werden können. Eine Alternative ist die Verlegung der Panoramaverteilung in die Authoring- und Monitoring-Systeme für Dolby Atmos (DP590) und MPEG-H (MMA, AMS). Dort werden intern 3D-Audio-Busse unterstützt.

Die komfortabelste Lösung wäre das immersive Audio-Bed direkt am Pult zu erstellen. Dies setzt jedoch voraus, dass neben Mono-, Stereo- und Surround-Kanälen auch 3D-Audio-Busse unterstützt werden. Für den Live-Broadcast sollte mindestens die Produktion mit 5.1.4-Bussen möglich sein, um klanglich ein überzeugendes, immersives Hörerlebnis zu vermitteln (vgl. Kapitel 2.3). Audio-Busse mit einer größeren Anzahl von Audiokanälen oder auch eine flexibel konfigurierbare Bussystematik erfordert von den Mischpultherstellern einen zusätzlichen Implementierungsaufwand. Theoretisch wären Mehrkanal-Busse bis 22.2 durchaus sinnvoll, wie es im Live-Broadcast in Japan der Standard ist. Allerdings müssen diese Formate auch von den Authoring Units unterstützt werden. Der DP590 von Dolby beispielsweise bietet intern nur ein Audio-Bed mit höchstens zehn Audiokanälen an (5.1.4-Bed). Wenn am Pult daher ein immersives Bed mit mehr Audiokanälen erstellt wird, müssten die zusätzlichen Kanäle in der Authoring Unit wiederum über Objekte repräsentiert werden.

Hersteller von Produktionsmischpulten sollten den Fokus deshalb eher auf die Implementierung eines intuitiven, kanalbasierten 5.1.4-Workflows legen. Dadurch würde die Produktion für Dolby Atmos und MPEG-H im Live-Broadcast durchaus komfortabler und geradliniger machen.

Im Zuge der Integration von Immersive Audio in Mischpulte wird ein 3D-Panning Modul benötigt. Die meisten verfügbaren Pulte sind mit einfachen Stereo- oder Surround-Panning Controllern ausgestattet, welche nur die X- und Y-Achse der horizontalen Hörebene abbilden. Deshalb muss ein zusätzlicher Regler für die Steuerung der Z-Achse (Höheninformation) ergänzt werden. Hierfür gibt es verschiedene Ansätze. Eine Variante besteht darin, die Z-Achse über die bereits vorhandenen Panorama-Drehregler in den einzelnen Kanalzügen zu steuern. Allerdings kann dadurch immer nur die Position auf einer Achse gesteuert werden. Über einen Schaltvorgang muss es möglich sein, die aktuell angewählte Achse zu wechseln. Dieses Prinzip ist für statische, kanalbasierte 3D-Audio Mischungen (z.B. Audio-Bed) ausreichend, bei denen bestimmte Mikrofonsignale diskret auf die jeweiligen Lautsprecher verteilt werden. Stagetech präsentierte diese Funktion mit ihrem Software Release 4.7 für die AURUS platinum Konsole unter anderem auf der Tonmeistertagung 2018 in Köln.¹⁵³

Eine andere Alternative ist die Aufteilung der Surround-Ebene (X-/Y-Achse) und der Höheninformation (Z-Achse) auf zwei separate Controller beispielsweise in der Main Unit einer Produktionskonsole. Für die Surround-Ebene bietet es sich an weiterhin einen Joystick als Controller zu verwenden. Dieser hat sich in Konsolen für Surround-Mischungen zur Positionierung eines Audio-Elementes in der horizontalen Ebene bewährt. Daneben könnte ein Drehregler, Fader oder auch ein Wheel für die Kontrolle der Höhenposition installiert werden. Dabei spart der Drehregler sicherlich am meisten Platz, was bei der Entwicklung von Produktionsmischpulten in jedem Fall berücksichtigt werden sollte. Mit dieser Variante ist es möglich Audio-Elemente in einem Durchgang im dreidimensionalen Raum zu platzieren. Dieses Prinzip findet beispielsweise Anwendung in der mc²96 Konsole von Lawo.¹⁵⁴

Bei Produktionsmischpulten mit großem Touchbildschirm können Audio-Elemente auch über Touchgesten positioniert werden. Hierbei werden beispielsweise Gesten mit zwei Fingern zur Steuerung der Position einer Audioquelle eingesetzt. Diese Variante findet sich beispielsweise als Zoomfunktion in zahlreichen Endgeräten wie Smartphones oder Tablets. Wie gut letzteres funktioniert, ist unter anderem abhängig von der Qualität der Touchbildschirme sowie der grafischen Benutzeroberfläche. Die Positionierung von Objekten in einer dreidimensionalen Darstellung über eine zweidimensionale Oberfläche, erweist sich jedoch häufig als schwierig. Aus praktischer Erfahrung des Verfassers dieser Arbeit mit einer Avid S6-Konsole, steht diese Steuerungsvariante den physischen Controllern der vorangegangenen Beispiele etwas nach.

¹⁵³ Vgl. Stage Tec (2018)

¹⁵⁴ Vgl. Lawo AG (2017) mc² 96 Grand Production Console, S. 9

Im Bereich der Postproduktion finden sich bereits zahlreiche Funktionen für Panning in 3D-Audio, welche es im Live-Broadcast noch nicht gibt. Aber auch dort könnten diese Tools durchaus Anwendung finden. DAWs wie *Pro Tools* und *Nuendo* bieten beispielsweise für die Produktion von Dolby Atmos in ihren 3D-Pannern sogenannte *Height-Modi* an, bei denen ausgehend von verschiedenen Hüllkurven, wie Wedge, Sphere oder Ceiling, die Höhe eines Audio-Elementes abhängig von dessen Position in der horizontalen Ebene berechnet wird. Das erleichtert dynamische Panoramafahrten von Objekten, beispielsweise über den Kopf des Zuhörers. Ein weiterer Punkt ist die Möglichkeit bestimmte Achsen zu fixieren und vom Panning im Raum auszuschließen.¹⁵⁵ In der ersten vorgestellten Variante ist das hinfällig, da hiermit grundsätzlich nur eine Achse zugreifbar ist. Für alle Weiteren kann das jedoch durchaus ein hilfreiches Tool sein.

Gibt es im Pult keine für den Anwender passenden 3D-Panningfunktionen, sollten zumindest externe Controller unterstützt werden. JL Cooper Electronics bietet, beispielsweise für die Yamaha Nuage, ein zusätzliches Panning-Modul an.¹⁵⁶ Wirft man einen Blick in den Bereich CAD-Anwendungen, gibt es unter anderem die SpaceMouse von 3Dconnexion, ein Controller der explizit für die Positionierung von Objekten in 3D-Umgebungen entwickelt wurde.¹⁵⁷ Ob und inwieweit solche umfassenden 3D-Panning Tools in Produktionsmischpulte für den Live-Broadcast integriert werden sollten, hängt auch stark vom Anwendungsfall ab.

7.2 Authoring und Monitoring

Mit die wichtigsten Anforderungen an die Produktionsmischpulte für immersive Audioproduktionen, gelten dem Authoring und Monitoring. Diese werden im Live-Broadcast zum Großteil von externen Hardwaregeräten übernommen. Zur Verbesserung des Workflows, sollten moderne Produktionsmischpulte die Integration von diesen Authoring- und Monitoring-Systemen bestmöglich unterstützen und dabei gleichzeitig möglichst unabhängig von den standardspezifischen Workflows sein.

7.2.1 Bedienung der AMAU vom Produktionsmischpult aus

Die Authoring- und Monitoring-Systeme für Dolby Atmos und MPEG-H im Live-Broadcast wurden in den Kapiteln 4.3.1 und 5.3.2 bereits näher erläutert, weshalb darauf an dieser Stelle nicht noch einmal im Detail eingegangen wird. Stattdessen liegt hier der Fokus darauf, wie diese Systeme bedient werden sollten, damit der Workflow möglichst effizient und intuitiv gestaltet wird.

¹⁵⁵ Vgl. Komar (2017)

¹⁵⁶ Vgl. JLC Cooper Electronics (2019)

¹⁵⁷ Vgl. 3Dconnexion GmbH (2018)

Das beginnt damit, dass die Ansicht der Webapplikationen von MMA, AMS und des DP590 über einen ausreichend großen Bildschirm direkt im Pult integriert werden sollte. Hierfür wird eine funktionierende Netzwerkverbindung vorausgesetzt. Alle Bedien- und Kontrollmöglichkeiten in Bezug auf Level-Metering, Programm-Lautheit und Metadaten Authoring finden dadurch gebündelt an einem Interface statt. Zum einen kann dadurch in der Produktionsumgebung Platz eingespart werden, da kein zusätzlicher PC mit Bildschirm benötigt wird. Zudem ist es eine Erleichterung für den Toningenieur, welcher sich nicht von der Konsole abwenden muss, um die Einstellungen der Authoring- und Monitoring-Systeme an einem separaten Arbeitsplatz zu überprüfen.

Des Weiteren müssen die Audiosignale, in der Regel über Direct Outs, von dem Mischpult zu den Hardwaregeräten gesendet und im Fall des DP590 auch wieder zurück über die Konsole auf die Monitorwege geschickt werden. Dafür benötigen die Mischpulte eine MADI- oder SDI- Schnittstelle (vgl. Kapitel 6.2.2 und 6.3.2).

Die Authoring und Monitoring Units sollten im Idealfall unmittelbar von einem Produktionsmischpult aus gesteuert werden können. Die Interaktion zwischen Pult und Webapplikation benötigt in dieser Hinsicht am meisten Entwicklungspotenzial und hat zahlreiche Möglichkeiten, den Workflow wesentlich zu verbessern. Das 3D-Panning von Audio-Objekten, beispielsweise im DP590, sollte direkt über die Panning-Sektion im Produktionsmischpult möglich sein. Auch wenn Immersive Audio produziert wird, sollten weiterhin auch 5.1- und Stereo-Mischungen erstellt werden. Da immer nur eine Mischung über das Monitoring-System wiedergegeben werden kann, sollte zumindest per Knopfdruck zwischen verschiedenen Versionen, wie Immersive Audio, 5.1, Stereo, Mono und Binaural, sowie den jeweiligen Presentations, direkt umgeschaltet werden können. Dafür müssen Knöpfe in der Monitoring-Sektion des Produktionsmischpultes reserviert werden. Eine Mischung aus vordefinierten und frei-konfigurierbaren User Keys ist hier zu empfehlen.

Die Firma Lawo bietet dafür eine Steuerungsintegration mit dem DP590 via Netzwerk an. Über das *Ember+* Protokoll, kann die Anwahl der Presentation, der Wiedergabekonfiguration, die Aktualisierung der Metadaten für den DP591 Encoder, oder auch das Umschalten in den STB-Modus, direkt vom Pult aus erfolgen. Bis zu neun User Keys können dafür mit den genannten Funktionen belegt werden. Gleichzeitig wird auch die Anzeige des Web-Interfaces vom DP590 auf den Bildschirmen der Konsole unterstützt (siehe Abb. 18).¹⁵⁸

¹⁵⁸ Vgl. Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. user's guide, S. 21-22



Abb. 18 User Interface des DP590 an einer Lawo Konsole¹⁵⁹

Für die Steuerung der MMA, AMS oder DP590 vom Produktionsmischpult aus, müssen beide ein gemeinsames Protokoll unterstützen. Neben dem DP590 unterstützt auch die MMA das Protokoll Ember+. Dieses findet sich bereits in zahlreichen Mischpulten und Studiokomponenten. Eine Alternative stellt die *General Purpose Input und Output (GPI, GPO)* Funktionalität dar. Darüber lassen sich ebenfalls bestimmte Grundfunktionen von AMS und DP590 vom Mischpulte steuern. Letzterer benötigt dafür allerdings noch ein zusätzliches Interface, den JNIO 410 Automation Controller.¹⁶⁰ Diese Variante ist in Broadcast-Konsolen weit verbreitet, erfordert aber einen gewissen Konfigurationsaufwand. Zur MMA konnten bezüglich GPIO-Funktionalität keine Angaben gefunden werden.

7.2.2 Integration der AMAU in das Produktionsmischpult

Eine denkbare Alternative zu einem Workflow mit externen Authoring- und Monitoring-Systemen, wäre die Verlegung von deren Funktionalität in das Produktionsmischpult. Zusätzliche Hardwarekomponenten wie die MMA, AMS und DP590 werden dadurch hinfällig und der Produktionsworkflow könnte wesentlich vereinfacht werden. Warum es sich dabei derzeit nur um ein mögliches Zukunftsszenario handelt, liegt daran, da die

¹⁵⁹ Bildquelle: Murrie (o.D.)

¹⁶⁰ Vgl. Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. user's guide, S. 21

Produktionsketten für NGA-Standards, wie Dolby Atmos und MPEG-H, nicht einheitlich und deshalb inkompatibel sind. In einem Paper der EBU heißt es dazu:

„The technology for NGA is here now, but it will not achieve its potential if nothing is done to overcome the industrial fragmentation of NGA production systems, all essentially doing the same things in an incompatible way.“¹⁶¹

Was für einen Codec-unabhängigen Workflow fehlt, ist also ein offen standardisiertes Format, welches entlang der Produktionskette von allen beteiligten Geräten unterstützt wird und die Audiosignale unkomprimiert bis zu den finalen Distributions- und Emissions-Encodern überträgt.

In file-basierter Form gibt es einen solchen Standard bereits: das *Audio Definition Model* (ADM). Darin werden alle Metadaten und Parameter für die Produktion und den Austausch von NGA-Inhalten spezifiziert. Sogenannte *ADM Profile* enthalten definierte Mapping-Anweisungen für eine standardisierte Transcodierung zwischen den verschiedenen Formaten und Profilen. Ein serielles Model (sADM) für Echtzeit-Anwendungen wie Streaming und Live-Broadcast, befindet sich derzeit noch in der Entwicklung. Wie die Realisierung eines offenen End-to-End-Workflows aussehen kann, wurde in den vergangenen Jahren im Zuge des *ORPHEUS*-Projektes erarbeitet. Dabei handelt es sich um ein von der Europäischen Kommission finanziertes Forschungsprogramm, bestehend aus mehreren europäischen Forschungsinstituten, Herstellern und EBU-Mitgliedern.¹⁶² Für weiterführende Informationen zu diesem Projekt, sei an dieser Stelle auf das öffentliche Paper EBU Technical Report TR 042¹⁶³ verwiesen.

Die Integration sämtlicher Funktionen für die immersive Audioproduktion von Dolby Atmos und MPEG-H in die Produktionsmischpulte würde einen extrem hohen Implementierungsaufwand bedeuten. Audiokanäle sollten direkt im Kanalzug als Objekte deklariert werden können. Gain- und Positionseinstellungen müssen in Form von Metadaten gespeichert und auch in Echtzeit, beispielsweise in einen Control-Track, moduliert und ausgegeben werden können. Zusätzlich wird ein format- und herstellerunabhängiger Renderer benötigt, der ein standardisiertes Monitoring ermöglicht. Als Referenz kann hierfür der in EBU Tech 3388¹⁶⁴ spezifizierte EBU ADM Renderer verwendet werden.

¹⁶¹ EBU TR 045, S. 3

¹⁶² Vgl. ebd., S. 3-12

¹⁶³ Vgl. EBU TR 042

¹⁶⁴ Vgl. EBU Tech 3388

7.3 Weitere Anforderungen

Immersive Audio ist immer noch ein Nischenprodukt und wird nicht von allen Endgeräten unterstützt. Langfristig müssen neben dem immersiven Mix weiterhin reguläre 5.1- und Stereo-Mischungen erstellt werden (vgl. Kapitel 6.3.2). Des Weiteren werden im Zuge der NGA-Standards verschiedene Presentations ermöglicht. Ein gleichzeitiges Monitoring von allen Versionen ist nicht möglich. Es braucht deshalb in den Produktionsmischpulten und Authoring-Tools verlässliche visuelle Analyser und Metering-Anzeigen. Diese müssen für die höhere Anzahl an Audiokanälen bei immersiven Audioproduktionen ausgelegt sein. Grundsätzlich werden im Zuge dessen auch automatisierte Tools zunehmend an Bedeutung gewinnen. Auto-Mix und Duckingfunktionen für Voice-over, aber auch gute Up- und Downmix-Algorithmen werden essentiell, um im internationalen Broadcast einen in Teilen objektbasierten 3D-Audio-Workflow zu bewältigen.

Zusätzlich müssen Produktionsmischpulte, die ein natives Mixing von Immersive Audio unterstützen, auch die Dynamik-, EQ-, und Effekt-Sektion auf das neue Bussystem umstellen. 5.1.4-Subgruppen benötigen gegebenenfalls einen 5.1.4-Kompressor, -Limiter, oder -EQ. Hall-Algorithmen für 3D-Audio bieten für die Produktion von Immersive Audio ebenso viel Entwicklungspotenzial.

8. Fazit

Die vorliegende Bachelorarbeit behandelte die Implementierung von immersiven Audioproduktionen mit Mischpulten im Live-Broadcast. Ziel der Arbeit war es, mithilfe einer ausführlichen Analyse mögliche Schwierigkeiten im Workflow für Standards wie MPEG-H und Dolby Atmos zu erkennen.

Es konnte aufgezeigt werden, dass immersive Audioproduktionen mit Produktionsmischpulten nicht nur grundsätzlich realisierbar sind, sondern bereits teilweise bei Produktionen im Live-Broadcast Anwendung finden (z.B. French Open 2018, Premier League, etc.). Allerdings werden dafür in der Produktionskette zusätzliche Authoring- und Monitoring-Systeme benötigt. Diese erweitern die vorhandenen Mischpulte um die bislang fehlenden Funktionen. Lediglich bei statischen Konfigurationseinstellungen der Encoder, kann auf diese externen Komponenten verzichtet werden. Bei der Analyse wurde aber auch festgestellt, dass die Workflows der beiden vorgestellten Standards, MPEG-H und Dolby Atmos, einerseits zwar ähnlich, auf der anderen Seite jedoch nicht kompatibel sind. Das betrifft zum einen die Produktionsseite mit den jeweils eigenen Authoring-, Monitoring- und Encoding-Komponenten, zum anderen kommen im Kontributionsbereich mit MHAS und Dolby ED2 auch zwei unterschiedliche Codecs zum Einsatz.

Da sich bis zum Zeitpunkt der Fertigstellung dieser Arbeit noch kein offener Produktionsstandard etabliert hat, lässt sich schlussfolgern, dass die Hersteller im Bereich Broadcast-Audiotechnik vorerst die Integration beider Systeme bestmöglich unterstützen sollten. Gerade an der Schnittstelle zwischen Mischpult und Authoring Unit sollte über einen stärkeren Austausch der Steuerungs- und Monitoring-Funktionen nachgedacht werden. Wenn überhaupt, sind bisher nur Grundfunktionen der Authoring Unit, wie zum Beispiel die Anwahl der Presentation, von einem Mischpult aus steuerbar. Möglicher Ansatzpunkt für eine Optimierung, ist die Steuerung der Position von Audio-Objekten über die Panning-Sektion der Produktionskonsole.

Des Weiteren muss bei der Produktion für NGA-Formate meist nicht nur eine Mischung angefertigt, sondern auch das Authoring von zusätzlichen Presentations und Wiedergabeformaten gehandhabt werden. Ein gleichzeitiges Monitoring aller Ausgabesignale ist nur visuell möglich. Dementsprechend werden Systeme, die dem Toningenieur ein automatisiertes Mixing und zuverlässiges Monitoring ermöglichen, in Zukunft immer wichtiger. Das Beispiel French Open im Jahr 2018 zeigte exemplarisch aber auch, dass immersive Liveproduktionen im Broadcast durchaus auch mit einer eingeschränkten Produktionsumgebung möglich sind.

Eine Weiterentwicklung und Öffnung der Produktionskette zu einem gemeinsamen Standard, wäre für zukünftige Workflows sicherlich wünschenswert. Ob ein einheitlicher, standard-unabhängiger Produktionsworkflow tatsächlich realisiert wird, ist zum gegenwärtigen Zeitpunkt allerdings unklar und aufgrund zahlreicher sich gegenseitig bedingender Faktoren schwer abzuschätzen. Bei einer möglichen Ausweitung von IP-basierten Workflows und Remoteproduktionen im Broadcast, muss zudem berücksichtigt werden, dass gegebenenfalls wieder neue Anforderungen an die Produktionsumgebungen gestellt werden und mehrere Systeme von einer einzigen Konsole aus gesteuert werden müssen.

In jedem Fall werden zuverlässige Lösungen für die Produktion, das Monitoring sowie die Qualitätskontrolle von Immersive Audio in diversen Produktionsumgebungen benötigt. Hierzu wäre interessant zu betrachten, ob beispielsweise ein System wie die LIME-Mixing Engine, mit der Funktion einen Control-Track in Echtzeit zu generieren, eine unabhängige Alternative zu den bereits vorhandenen Authoring- und Monitoring-Systemen darstellen könnte.

Literaturverzeichnis

Bilandzic, Helena (2014) Immersion. In: *Handbuch Medienrezeption*. Hrsg. v. Helena Bilandzic et al., Baden-Baden: Nomos, S. 273

Bleidt, Robert (2014) Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement, 2014 Annual Technical Conference & Exhibition, Hollywood, CA., SMPTE, [<http://www.mpeg-haa.com/presentations/MPEG-H%20Fraunhofer%20SMPTE%20October%202014%2016x9%20v2.pdf>] (Zugriff am 28.02.2019)], S. 9-28

Bleidt, Robert, Baxter, Dennis (2014) Starting production in interactive and immersive sound, DTV Audio Group at SVG Summit. Dec. 2014, Fraunhofer IIS

Bleidt, Robert et al. (2016) Building The World's Most Complex TV Network. A Test Bed for Broadcasting Immersive and Interactive Audio, SMPTE Motion Imaging Journal, S. 3

Bleidt, Robert et al. (2017) Development of the MPEG-H TV Audio System for ATSC 3.0, IEEE Transactions on Broadcasting, Vol. 63, No. 1, March 2017, S. 211-227

3Dconnexion GmbH (2018) SpaceMouse Compact, [https://www.3dconnexion.de/spacemouse_compact/de/] (Zugriff am 30.03.2019)]

Denon, DTS Virtual:X. Was ist Virtual:X und wie kann ich es nutzen?, [https://denon-de.custhelp.com/app/answers/detail/a_id/5974/~/dts-virtual%3Ax] (Zugriff am 31.03.2019)]

Deutsche TV-Plattform (2016) Next Generation Audio (NGA), [<https://www.tv-plattform.de/de/next-generation-audio>] (Zugriff am 02.04.2019)]

Dickreiter, Michael et al. (2014) Handbuch der Tonstudioteknik, Band 1, Hrsg. v. Michael Dickreiter et al., 8. Auflage, Berlin/Boston: De Gruyter Saur, 2014, S. 128

Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 1 - Tools for Transport, [<http://cras-atmos.com/mod/lesson/view.php?id=6&pageid=74>] (Letzter Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 1 - Understanding the concept, [<http://cras-atmos.com/mod/lesson/view.php?id=6&pageid=72&startlastseen=no>] (Letzter Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Core Concepts. Lesson 3 - Movie Theaters, [<http://cras-atmos.com/mod/lesson/view.php?id=15>] (Letzter Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 2 - DP580-Reference Decoder, [<http://cras-atmos.com/mod/lesson/view.php?id=22&pageid=52>] (Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 2 - DP590-Monitoring&Authoring, [<http://cras-atmos.com/mod/lesson/view.php?id=22&pageid=50>] (Zugriff am 29.03.19)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Audio Connectivity, [<http://cras-atmos.com/mod/lesson/view.php?id=23&pageid=56&startlastseen=no>] (Zugriff am 29.03.19)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Network Connectivity, [<http://cras-atmos.com/mod/lesson/view.php?id=23&pageid=57>] (Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Clearing Session Data, [<http://cras-atmos.com/mod/lesson/view.php?id=23&pageid=58>] (Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 3 - Preparing the Bed, [<http://cras-atmos.com/mod/lesson/view.php?id=23&pageid=60>] (Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 5 - Preparing for Dolby AC-4, [<http://cras-atmos.com/mod/lesson/view.php?id=20&pageid=64>] (Zugriff am 29.03.2018)]

Dolby Atmos - Live TV Production Training, Dolby Atmos for Live TV. Lesson 5 - Creating Dialogue Channels, [<http://cras-atmos.com/mod/lesson/view.php?id=20&pageid=67>] (Zugriff am 29.03.2019)]

Dolby Atmos - Live TV Production Training, DP591, [http://cras-atmos.com/pluginfile.php/382/mod_lesson/page_contents/123/5.png]

Dolby Atmos - Live TV Production Training, Room Installation. Dolby Atmos Speaker Installation and Room Setup Guidelines for Live Events, [<http://cras-atmos.com/course/view.php?id=4>] (Zugriff am 30.03.2019)]

Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby's Content Creation Tools, [<http://cras-atmos.com/mod/lesson/view.php?id=47&pageid=121>] (Zugriff am 02.04.2019)]

Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - DP591 Broadcast Processor, [<http://cras-atmos.com/mod/lesson/view.php?id=47&pageid=123>] (Zugriff am 29.03.19)]

Dolby Atmos - Live TV Production Training, Room Installation. Mixing Guidelines for Live Events - Dolby Atmos OB Workflow, [<http://cras-atmos.com/mod/lesson/view.php?id=47&pageid=128>] (Zugriff am 30.03.2019)]

Dolby Atmos - Live TV Production Training, Webapplikation des DP590, [http://cras-atmos.com/pluginfile.php/62/mod_lesson/page_contents/63/41.png]

Dolby Laboratories, Inc. (2014) Dolby Atmos Next-Generation Audio for Cinema. White Paper, S. 1-7

Dolby Laboratories, Inc. (2015) Dolby AC-4: Audio Delivery for Next-Generation Entertainment Services, [<https://www.dolby.com/us/en/technologies/ac-4/Next-Generation-Entertainment-Services.pdf>], S. 29

Dolby Laboratories, Inc. (2017) Dolby Atmos Immersive Audio From the Cinema to the Home, Nick Engel, AES Melbourne Section Meeting, [http://www.aesmelbourne.org.au/wp-content/media/Dolby_Dec2017.pdf], S. 43

Dolby Laboratories, Inc. (2017) What is Dolby Digital Plus JOC? (Joint Object Coding), [<https://developerkb.dolby.com/support/solutions/articles/16000067758-what-is-dolby-digital-plus-joc-joint-object-coding->] (Zugriff am 29.03.19)]

Dolby Laboratories, Inc. (2018), Dolby DP580 Front Panel, [https://www.dolby.com/uploadedImages/wwwdolbycom/Content/Pro_Products/DP580/008-prof-ref-decoder-DP580-product-hires-front.jpg] (Zugriff am 29.03.2019)]

Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. product specification, S. 6-32

Dolby Laboratories, Inc. (2018) Dolby DP590 Object Authoring Tool. user's guide, S. 21-22

Dolby Laboratories, Inc. (2018) Dolby Professional Reference Decoder DP580. User's guide, V2.0, S. 7-8

Dolby Laboratories, Inc. (2018) How Dolby and Telegenic are bringing ringside seats into boxing fan's homes, Case Study, [<https://www.dolby.com/us/en/technologies/dolby-ac-4/dolby-telegenic-case-study.pdf>], S. 2-4

Dolby Laboratories, Inc. (o.D.) Dolby ED2. Next Generation Audio Mezzanine Codec, White Paper, S. 1-2

Dolby Laboratories, Inc. (o.D.) Home Theater Speaker Guide. 5.1.4 Overhead Speakers, [<https://www.dolby.com/us/en/guide/5.1.4-overhead-speaker-setup-guide.pdf>] (Zugriff am 16.01.2019)]

DTS, Inc. (2018) DTS:X Homepage, [<https://dts.com/dtsx>] (Zugriff am 27.03.2019)]

DTS, Inc. (2018) Home Solutions, [<https://dts.com/professional/home-solutions>] (Zugriff am 31.03.2019)]

EBU Tech 3388, ADM Renderer for use in next generation audio broadcasting, BTF Renderer Group, Version 1.0, March 2018, Geneva, [<https://tech.ebu.ch/docs/tech/tech3388.pdf>] (Zugriff am 31.03.2019)]

EBU TR 042, Example of an end-to-end OBA broadcast architecture and workflow, Orpheus Project, May 2018, Geneva, [<https://tech.ebu.ch/docs/techreports/tr045.pdf>] (Zugriff am 31.03.2019)]

EBU TR 045, Why broadcasters need an open, codec-independent workflow for NGA production deployment, January 2019, Geneva, [<https://tech.ebu.ch/docs/techreports/tr045.pdf>] (Zugriff am 31.03.2019)], S. 3-12

European Telecommunications Standards Institute (2016) ETSI TS 103 420 V1.1.1 (2016-07), Backwards-compatible object audio carriage using Enhanced AC-3, [https://www.etsi.org/deliver/etsi_ts/103400_103499/103420/01.01.01_60/ts_103420v010101p.pdf] (Zugriff am 29.03.2019)]

Fidelity Media (2016) Burmester bringt Auro-3D ins Auto, Pressemitteilung 28.07.2016, [<https://www.fidelity-magazin.de/2016/07/28/burmester-bringt-auro-3d-ins-auto/>] (Zugriff am 02.04.2019)]

Fleischmann, Jan (2017) MPEG-H - ein Audioformat der nächsten Generation (NGA), techmagazin, [<http://tech-magazin.de/2017/05/mpeg-h-ein-audioformat-der-naechsten-generation-nga/>] (Zugriff am 07.01.2019)]

Fraunhofer Audio Blog (2016) TV-Encoder mit MPEG-H Audio verfügbar, 18.04.2016, [<http://www.audioblog.iis.fraunhofer.de/tv-encoder-mpeg-h/>] (Zugriff am 30.03.2019)]

Fraunhofer Audio Blog (2019) Spatial Audio Designer integriert MPEG-H Audio, [<https://www.audioblog.iis.fraunhofer.com/de/spatial-audio-designer-integriert-mpeg-h-audio>] (Zugriff am 28.03.2019)]

Fraunhofer IIS (2019) MPEG-H Authoring Plug-in (MHAPi), [<https://www.iis.fraunhofer.de/de/ff/amm/dl/software/mhapi.html>] (Zugriff am 28.03.2019)]

Friesecke, Andreas (2014) Die Audio-Enzyklopädie, 2. Auflage, Berlin/Boston: De Gruyter Saur, 2014

Grewe, Yannik, Simon, Christian, Scuda, Ulli (2018) Producing Next Generation Audio using the MPEG-H TV Audio System, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, 2018 NAB BEITC Proceedings, S. 335-340

Griffin, Al (2017) Why does my receiver default to DTS Neural:X Mode?, Sound&Vision, [<https://www.soundandvision.com/content/why-does-my-receiver-default-dts-neuralx-mode>] (Zugriff am 31.03.2019)]

Hamasaki, Kimio, Hiyama, Koichiro (2006) Development of a 22.2 Multichannel Sound System, Broadcast Technology no.25, Winter 2006, NHK STRL, [<https://www.nhk.or.jp/strl/publica/bt/bt25/pdf/fe0025-2.pdf>] (Zugriff am 18.01.2019)], S. 9-10

Hamasaki, Kimio (2011) 22.2 Multichannel Audio Format Standardization Activity, Broadcast Technology No.45, Summer 2011, NHK STRL, [<https://www.nhk.or.jp/strl/publica/bt/bt45/pdf/fe0045-6.pdf>] (Zugriff am 18.01.2019)], S. 15

Herre, Jürgen et al. (2015) MPEG-H 3D Audio - The New Standard for Coding of Immersive Spatial Audio. In: *IEEE Journal of selected Topics in signal processing, Vol. 9, No.5*, 2015, S. 770-771

IRT GmbH (2019) Conceptual overview of channel-based audio production and consumption, [<https://lab.irt.de/demos/object-based-audio/img/channel-based.png>]

IRT GmbH (2019) Conceptual overview of object-based audio production and consumption, [<https://lab.irt.de/demos/object-based-audio/img/object-based.png>]

IRT GmbH (2019) Object-based audio - The future of audio production, delivery and consumption, [<https://lab.irt.de/demos/object-based-audio/>] (Zugriff am 07.01.2019)]

JLCooper Electronics (2019) Nuage Surround Panner. For Yamaha & Steinberg Nuage Console, [https://jlcooper.com/_php/product.php?prod=nuage] (Zugriff am 30.03.2019)]

Jünger Audio GmbH (2017) MMA - Multichannel Monitoring and Authoring, [https://junger-audio.com/uploads/brochure/598c7b9a6a612d242d1d0000/JA_broschure_MMA_web_201708.pdf] (Zugriff am 24.01.2019)], S. 2

Jünger Audio GmbH (o.D.) jpg1616_Screenshot_2-3D_Panner.jpg, [<https://jungeraudio.com/en/products/flexai/multichannel-monitoring-and-authoring-mma>] (Zugriff am 26.01.2019)]

Komar, Jeff (2017) Dolby Atmos Improvements in Pro Tools. HD 12.8.2, Avid Technology, Inc., [<http://www.avidblogs.com/dolby-atmos-improvements-in-pro-tools-hd-12-8-2/>] (Zugriff am 30.03.2019)]

L-Acoustics (2019) Live Sound, [<http://www.l-isa-immersive.com/live-sound/>] (Zugriff am 03.04.2019)]

Lawo AG (2017) Lawo kooperiert mit New Audio Technology: Präsentation von LIME Immersive Mixing Engine für mc2 auf der IBC 2017, [https://www.lawo.com/de/aktuell/news/news/translate_to_german_cooperation_with_new_audio_technology_lawo_introduced_lime_immersive_mixing.html] (Zugriff am 28.03.2019)]

Lawo AG (2017) mc2 96 Grand Production Console, [https://www.lawo.de/fileadmin/content/Products/mc296/Lawo_mc2_96_EN.pdf] (Zugriff am 30.03.2019)], S. 9

Merging Technologies (2017) Pyramix 11.1 To Include Full Object Based Audio Workflow, [<https://www.merging.com/news/news-stories/pyramix-11-1-to-include-full-object-based-audio-workflow>] (Zugriff am 28.03.2019)]

MPEG-H Audio Scene Information example (2017) In: *ATSC Standard: A/342 Part 3, MPEG-H System*, Doc. A/342-3:2017, ATSC, S. 5

Murrie, Stewart (o.D.) Real-world testing, Dolby DP590 Next-Gen Content Creation Without The Complexity, [<https://www.murrie.org/dolby-dp590>] (Zugriff am 27.03.2019)]

Nipkow, Lasse, Zielinsky, Gregor, Ammermann, Tom (2019) Die Bedeutung von 3D bei Immersive Audio. In: *VDT-Magazin*, Ausgabe 1-2019, S. 22

Pizzi, Skip, Kean, John (2018) Audio Contribution and Distribution Methods. In: *National Association of Broadcasters engineering handbook*. Hrsg. v. Garrison C. Cavell, 11th Edition, 2018, S. 911

Plogsties, Jan (2015) MPEG-H Audio System for Broadcasting. ITU-R Workshop „Topics on the Future of Audio in Broadcasting“, Fraunhofer IIS, S. 16

Remy, Marcel (2018) Audio für 360°-Video. Qualitätskriterien und Produktionsworkflow im Kontext musikalischer Aufführungspraxis am Beispiel des Hessischen Rundfunks, HdM Stuttgart, S. 21

Roginska, Agnieszka, Geluso, Paul (2018) Introduction. Immersive Sound. In: *Immersive sound: the art and science of binaural and multi-channel audio*. Hrsg. v. Agnieszka Roginska und Paul Geluso, Routledge, 2018, S. 1

Scholz, Florian C. (2015) Audiotechnik für Mediengestalter, Berlin/Boston: De Gruyter Saur, 2015

Schuller, Christian (2015) 3D Sound mit DTS X - Der aktuelle Stand der Dinge Teil 3, [<https://www.heimkinoraum.de/tests/3d-sound-mit-dts-x-der-aktuelle-stand-der-dinge-teil-3-74>] (Zugriff am 17.01.2019)]

Sennheiser electronics GmbH & Co. KG (2019) AMBEO Soundbar, [<https://de-de.sennheiser.com/ambeo-soundbar>] (Zugriff am 28.03.2019)]

Silzle, Andreas et al. (2011) Investigation on the Quality of 3D Sound Reproduction, International Conference on Spatial Audio (ICSA), Detmold, Germany, S. 340

Simon, Christian et al. (2018) Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, S. 1-5

Slavik, Karl M., Weinzierl, Stefan (2008) Kodierung. In: *Handbuch der Audiotechnik*. Hrsg. v. Stefan Weinzierl, Springer-Verlag Berlin Heidelberg, 2008, S. 663

Slavik, Karl M. (2018) Dolby Atmos in Fernseh-Workflows, [<https://www.vdt-magazin.de/news/news-detail/23847-dolby-atmos-in-fernseh-workflows>] (Zugriff am 07.02.2019)]

Stage Tec (2018) TMT: Stage Tec präsentiert neue Features für AURUS platinum und AURUS, [<https://www.stagetec.com/de/blog-hidden-2018/tmt-stage-tec-praesentierete-neue-features-fuer-aurus-platinum-und-aurus.html>] (Zugriff am 30.03.2019)]

Theile, Günther (2014) 3D Surround Sound. In: *Handbuch der Tonstudioteknik*, Band 1. Hrsg. v. Michael Dickreiter et al., 8. Auflage, Berlin/Boston: De Gruyter Saur, 2014, S. 311-313

The MPEG-H TV Audio System. Use Cases and Workflows (o.D.), Media Solutions, Fraunhofer IIS, [https://mediasolutions.ericsson.com/cms/wp-content/uploads/2018/04/Media-Solutions-MPEG-H-Brossure_160418-FINAL-FINAL-1.pdf] (Zugriff am 22.03.2019)], S. 3-4

Van Baelen, Wilfried (2016) Der Auro-3D Upmixer „AuroMatic“. In: *VDT-Magazin*, Ausgabe 2-2016, S. 18-19

Westphal, André (2018) Dolby Atmos: Amazon Prime Video bietet ab 31. August das 3D-Soundformat, [<https://www.4kfilme.de/dolby-atmos-amazon-prime-video-bietet-ab-31-august-das-3d-soundformat/>] (Zugriff am 27.03.2019)]

Wikipedia (2018) Affiliate (Rundfunk), [[https://de.wikipedia.org/w/index.php?title=Affiliate_\(Rundfunk\)&oldid=180555369](https://de.wikipedia.org/w/index.php?title=Affiliate_(Rundfunk)&oldid=180555369)] (Zugriff am 30.03.2019)]

Wikipedia (2018) 5.1, [<https://de.wikipedia.org/w/index.php?title=5.1&oldid=177222254>]
(Zugriff am 30.03.2019)]