Hochschule der Medien Stuttgart
Studiengang Audiovisuelle Medien

Objektbasierte Musikproduktion

Entwicklung eines kombinierten Workflows für Dolby Atmos Music und 360 Reality Audio auf Basis einer bereits bestehenden Stereo-Mischung

Masterarbeit

zur Erlangung des akademischen Grades
Master of Engineering (M.Eng.)

Vorgelegt von:	Daniela Rieger, <daniela.rieger(at)gmx.de>
Matrikelnummer:	37030
Datum:	09.10.2020
Erstprüfer:	Prof. Dr. Frank Melchior
	Hochschule der Medien, Stuttgart
Zweitprüfer:	Prof. Oliver Curdt
	Hochschule der Medien, Stuttgart
Betreuer:	Dr. Ulli Scuda, M.Eng. Philipp Eibl
	Fraunhofer-Institut für Integrierte Schaltungen IIS, Erlangen

Ehrenwörtliche Erklärung

Hiermit versichere ich, Daniela Rieger, ehrenwörtlich, dass ich die vorliegende Masterthesis mit dem Titel: „Objektbasierte Musikproduktion: Entwicklung eines kombinierten Workflows für Dolby Atmos Music und 360 Reality Audio auf Basis einer bereits bestehenden Stereo-Mischung“ 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.

Ort, Datum *Unterschrift*

Zusammenfassung

In der Stereo-Musikproduktion haben sich über die Jahre verschiedene Vorgehensweisen und damit Workflows zur Produktion, Distribution und Wiedergabe von Musik etabliert. Dieser Prozess steht bei objektbasierter Musikproduktion noch am Anfang. Es existiert eine Vielzahl an unterschiedlichen Begrifflichkeiten, Formaten und Codecs, welche sich auf alle Schritte der Produktionskette auswirken. Aktuell entspricht objektbasierte Musik der Produktion für einen Codec, was in Diskrepanz zur Codec-agnostischen Stereo-Musikproduktion steht. Mit Dolby Atmos Music und 360 Reality Audio sind seit Ende 2019 zwei objektbasierte, immersive Technologien über Musik-Streaming-Dienste kommerziell verfügbar. Für diese Arbeit wurden aktuelle objektbasierte Produktionsketten am Beispiel von Dolby Atmos Music und 360 Reality Audio betrachtet, um technologische Unterschiede zu ermitteln. Anhand dieser Erkenntnisse wurde ein Workflow entwickelt, welcher auf einer bereits angefertigten Stereo-Mischung basiert und durch Nutzung einer gemeinsamer Produktionsumgebung eine kombinierte Vorgehensweise zur Produktion von Dolby Atmos Music und 360 Reality Audio ermöglicht. In der praktischen Anwendung hat sich herausgestellt, dass objektbasierte Musikproduktion bewährte Workflows aufgreift und weiterentwickelt, diese jedoch durch die objektbasierte Produktionsweise ab der Mischung bis zur Wiedergabe von etablierten Stereo-Produktionsprozessen abweichen. Der praktische Vergleich von Dolby Atmos Music und 360 Reality Audio hat gezeigt, dass insbesondere im Export-, Encoding-, Distributions- und Wiedergabeprozess Unterschiede herrschen und Merkmale objektbasierten Audios wie Personalisierung oder flexibles Wiedergabe-Rendering noch nicht durchgehend implementiert wurden. Außerdem wurde deutlich, dass speziell die Wiedergabe von technischen Hintergrundprozessen wie beispielsweise Binaural-Rendering beeinflusst wird, die bei der Mischung nicht gesteuert werden können. Weiterhin wurde in der praktischen Arbeit ersichtlich, dass der aktuelle Stand objektbasierter Produktionsprozesse weiterer Entwicklung sowie der Einführung formatübergreifender Konventionen bedarf, beispielsweise im Bereich des Masterings.

Keywords: 3D Audio, Musikproduktion, immersiv, objektbasiertes Audio, Sony, 360 Reality Audio, Dolby Atmos Music, Workflow, NGA, MPEG-H, AC-4, DD+JOC

Abstract

In stereo music production, different procedures and thus workflows for the production, distribution and playback of music have been established and refined over time. This process is still in its infancy in object-based music production. There are a multitude of different terms, formats, and codecs each effecting every step of the production chain and, currently, mixing object-based audio means producing for one specific object-based format, such as Dolby Atmos Music or 360 Reality Audio, which have been commercially available in music streaming services since the end of 2019. This is far removed from the universal stereo mix with one simple output format. This thesis examines these two currently available object-based production chains to identify technological differences and develop an object-based workflow that expands upon an existing stereo mix. This workflow allows a combined approach to the production of Dolby Atmos Music and 360 Reality Audio by using a common production environment. Object-based music production systems further develop legacy workflows, mostly differing from stereo workflows after the mixing stage up until playback, and allows for addition object-specific features. Practically, it has shown that Dolby Atmos Music and 360 Reality Audio differ in their export, encoding, distribution and playback processes, and that features of object-based audio such as personalization or flexible playback rendering have not yet been implemented consistently. Especially playback on these systems is heavily influenced by background processes that cannot be controlled during the mixing stage, such as binaural rendering which results in spectral colorations.This thesis concludes that further work is required in the development of object-based production systems to introduce cross-format conventions and improve technical implementations, beneficial for both mixing and mastering.

Keywords: 3D-Audio, music production, immersive, object based audio, Sony, 360 Reality Audio, Dolby Atmos Music, workflow, NGA, MPEG-H, AC-4, DD+JOC

Danksagung

Zuerst möchte ich Prof. Dr. Frank Melchior von der Hochschule der Medien Stuttgart für die Betreuung dieser Arbeit danken. Die zahlreichen virtuellen Meetings haben mir sehr geholfen, den roten Faden dieser Thesis zu finden und in verschiedene Richtungen aufzuspannen. Vielen Dank an meinen Zweitprüfer Prof. Oliver Curdt, welcher bereits während des gesamten Masterstudiums stets mit wertvollen Ratschlägen zur Seite stand.

Diese Masterarbeit wurde vom Fraunhofer-Institut für Integrierte Schaltungen IIS in Erlangen gefördert. Insbesondere möchte ich meinem Praxisbetreuer Dr. Ulli Scuda für seine zahlreichen Ideen, Anregungen und Kommentare sowie seine Unterstützung während der Masterarbeitszeit danken. Ebenso danke ich meinem zweiten Praxisbetreuer M.Eng. Philipp Eibl für das jederzeit offene Ohr. Ein besonderer Dank gilt an dieser Stelle meinem Kollegen M.Eng. Yannik Grewe, welcher über den gesamten Zeitraum der Masterarbeit ein offenes Ohr für meine zahlreiche Fragen hatte. Sein technisches Fachwissen hat diese Arbeit unterstützend vorangetrieben. Ferner bedanke ich mich bei meinen Kollegen vom Fraunhofer IIS SoundLab Team: Dipl.-Tonmeister Christian Simon, Dipl.-Ing. Thomas Mayenfels, Dipl.-Ing.(FH) Andreas Turnwald, B.A. Valentin Schilling und Wolfgang Hörlbacher.

Ein besonderer Dank gilt Marcel Remy für die unzähligen Gespräche, Ideen sowie all die persönliche Unterstützung und Ermutigung während meines gesamten Masterstudiums und dieser Arbeit. Danke an meine weiteren Kommilitonen Tobias Kurzweg, Leon Hofmann, Jonas Kieser, Jona Eisele und Bastian Kilper. Danke für diese wunderbare Zeit und die Realisierung unseres Masterprojektes, welches das Highlight meiner gesamten Studienzeit bleiben wird.

Vielen Dank an meine Familie für die Unterstützung während meiner Studienjahre.

Schließlich möchte ich Davide Straninger meinen herzlichen Dank ausdrücken – für all die Ermutigung, Motivation und Unterstützung zu jeder Zeit. Danke, dass du da bist.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

360 RA	360 Reality Audio
AAC	MPEG-4 Advanced Audio Coding
AC-4 IMS	AC-4 Immersive Stereo
ADM	Audio Definition Model
A-JOC	Advanced Joint Object Coding
AES	Audio Engineering Society
AVR	Audio Video Receiver
BRIR	Binaural Room Impulse Response
BWF	Broadcast Wave File
C	Center
DAM	Dolby Atmos Music
DAMF-Set	Dolby Atmos Master File Set
DAW	Digital Audio Workstation
DBAP	Distance-Based Amplitude Panning
DD+	Dolby Digital Plus
DD+JOC	Dolby Digital Plus Joint Object Coding
DRC	Dynamic Range Control
eARC	Enhanced Audio Return Channel
EBU	European Broadcasting Union
EQ	Equalizer
Fraunhofer IIS	Fraunhofer-Institut für Integrierte Schaltungen IIS
HDMI	High Definition Multimedia Interface
HOA	Higher Order Ambisonics
HRIR	Head-Related Impulse Response

HRTF	Head-Related Transfer Functions
IC	Interaural Coherence
ILD	Interaural Level Difference
ISRC	International Standard Recording Code
ITD	Interaural Time Difference
JOC	Joint Object Coding
L	Left
LFE	Low-Frequency Effects
Lrs	Left-Rear-Surround
Lss	Left-Side-Surround
Ltf	Left-Top-Front
Ltr	Left-Top-Rear
LUFS	Loudness Units relative to Full Scale
MPEG-H	MPEG-H 3D Audio Standard
MPS	MPEG-D MPEG Surround
NGA	Next Generation Audio
R	Right
Rrs	Right-Rear-Surround
Rss	Right-Side-Surround
Rtf	Right-Top-Front
Rtr	Right-Top-Rear
SAC	Spatial Audio Coding
S-ADM	Serial-ADM
SAOC	MPEG-D Spatial Audio Object Coding
USAC	MPEG-D Unified Speech and Audio Coding
VBAP	Vector Base Amplitude Panning
XML	Extensible Markup Language

Glossar

Authoring

Der Begriff „Authoring“ (engl. to author = verfassen) bezieht sich auf einen neuen Schritt in der Produktionskette objektbasierter Audioinhalte. Insbesondere wird damit auf das Generieren der Metadaten, das Monitoring der objektbasierten Produktion sowie die Exporteinstellungen referenziert. Beim Authoring werden diese unterschiedlichen Komponenten in Beziehung zueinander gesetzt und verknüpft.

Bett

Als „Bett“ soll nachfolgend eine Kombination aus verschiedenen Klangelementen beschrieben werden. Das Bett kann als Basis betrachtet werden, auf der sich die Musikmischung aufbaut und wird auch als „IT-Stem“ (International Tape Stem) oder „M&E-Stem“ (Music & Effects Stem) bezeichnet. Es gilt anzumerken, dass der Begriff „Bett“ aktuell je nach Anwendungsbereich oft unterschiedlich definiert wird.

DAW

„DAW“ steht für Digital Audio Workstation und beschreibt ein Programm, mit dem Audio aufgenommen, bearbeitet und abgespielt werden kann. DAWs bestehen meist aus einer horizontalen Zeitachse (Timeline) und vertikal untereinander angeordneten Spuren.

Immersives Audio bzw. 3D-Audio

Der Begriff „3D-Audio“ bezeichnet ein Hörerlebnis, bei dem der Schall aus allen Richtungen (= dreidimensional) um den Zuhörer herum wahrgenommen werden kann, einschließlich oben und unten. „Immersives Audio“ (lat. immersio = Eintauchen) beschreibt das emotionale und räumliche „Eintauchen“ des Hörers in die Klanglandschaft, was durch dreidimensionale Wiedergabe verstärkt werden kann. Im Folgenden soll der Begriff „immersiv“ gleichbedeutend mit „3D-Audio“ verwendet werden, hierbei soll in keinster Weise

impliziert werden, dass andere Produktionsformate (Stereo, 5.1 Surround) nicht auch immersiv wahrgenommen werden können.

Kanal

Als „Kanal“ soll im Folgenden ein Audiosignal bezeichnet werden, das mit der Absicht produziert wurde, an einer vorbestimmten Lautsprecherposition wiedergegeben zu werden.

Objekte

Als „Objekt“ soll im Folgenden ein Klangelement bezeichnet werden, das aus einem Audiosignal mit zugehörigen Metadaten besteht und im dreidimensionalen Raum positioniert wird. Es wird unterschieden zwischen *statischen* und *dynamischen* Objekten. Im Gegensatz zu statischen Objekten verändern sich die Positionen dynamischer Objekte über die Zeit.

Workflow

Der Begriff „Workflow“ (engl. work = Arbeit; flow = Fluss) soll im Folgenden eine Reihung an Arbeitsabläufen beschreiben, die zur Musikproduktion nötig sind. Im Speziellen werden die Schritte der objektbasierten Produktionsketten von Dolby Atmos Music und 360 Reality Audio betrachtet. Ein Workflow stellt dar, wie, in welcher Reihenfolge und mit welchen technischen Hilfsmitteln die einzelnen Arbeitsschritte ausgeführt werden.

1 Einleitung

„While an engineer familiar with the complications of sound reproduction may be amazed at the tens of thousands of trouble-free performances given daily, the public takes our efforts for granted and sees nothing remarkable about it.[...] The public has to hear the difference and then be thrilled by it [...]. Improvements perceptible only through direct A-B comparisons have little box-office value.“
Garity & Hawkins (Garity & Hawkins, 1941, S. 127)

1.1 Motivation & Zielsetzung

3D-Audio oder *immersives Audio* übt zweifelsohne einen Reiz aus – auf Tonschaffende möglicherweise noch mehr als auf Konsumenten. Dies zeigt sich auch durch die Vielzahl an Beiträgen zu „3D-Audio“ auf den Audio-Konferenzen der letzten Jahre (Tonmeistertagung 2018, Prolight+Sound 2019, AES Virtual Vienna 2020). 3D-Audio bringt eine neue Komplexität mit sich – Codecs und Formate, verschiedenste Technologien und Software, neue, kreative Gestaltungsmöglichkeiten und erweiterte Wiedergabesysteme. Doch eben diese Komplexität birgt die Gefahr, dass der Endkonsument aus dem Fokus verdrängt wird. Wie bereits die beiden Fantasound-Entwickler William E. Garity und John N.A. Hawkins im Jahre 1941 feststellten, so kann eine Technologie noch so kompliziert und komplex sein – der Konsument interessiert sich großteils nicht für all die Hürden, die überwunden wurden oder all die Technik, die dahintersteckt. Der Endkonsument möchte die Inhalte möglichst unkompliziert hören und einen Mehrwert erleben. Denn nur so kann eine neue Technologie funktionieren – nur so entsteht Begeisterung und Interesse, nur so wird das neue Format konsumiert und nur so kann es sich für alle beteiligten Partner durchsetzen. Auch dies haben Garity und Hawkins bereits vor knapp 80 Jahren festgestellt und auch dies hat in der heutigen Zeit nicht an Relevanz verloren.

Hat sich 3D-Audio im Filmbereich bereits seit längerem etabliert (Dolby Atmos, DTS:X, Auro-3D), so war die Verfügbarkeit von (kanalbasierten) 3D-Musikproduktionen bisher

noch recht gering. Objektbasierte, immersive Musik wurde Ende 2019 erstmals eingeführt, indem die beiden Technologien *Dolby Atmos Music* und *360 Reality Audio* vorgestellt wurden. Wie bei der Quadrophonie ab den 1960er Jahren und der darauffolgenden Einführung von 5.1 Surround – was sich beides kommerziell nicht durchsetzen konnte – handelt es sich bei den objektbasierten Formaten um neue Technologien zur Produktion, Distribution und Wiedergabe von Musik. Im Unterschied zu damals ist die Ausgangslage heutzutage jedoch eine andere: Während 5.1 Surround durch die kanalbasierte Produktionsweise auf ein festes Lautsprecher-Layout und die Distribution über physische Medien abzielte, so kann objektbasiertes Audio flexibel auf verschiedene Wiedergabesysteme (Lautsprecher und Kopfhörer) gerendert und über Musik-Streaming-Dienste vertrieben werden.

Seit den ersten Versuchen der Schallaufzeichnung gegen Ende des 19. Jahrhunderts hat sich der Bereich „Musik“ stetig gewandelt: Die Techniken zur Aufnahme und Produktion (Mono, Stereo, 5.1 Surround, 3D) ebenso wie die der Distribution (Schallplatte, CD, Digital, Streaming-Dienste) und Wiedergabe (Plattenspieler, Radio, Lautsprecher, Kopfhörer, Soundbar) unterliegen einer fortlaufenden Entwicklung. Somit haben sich über die Jahre bei der Musikproduktion, abhängig von Genre und Kontext, viele einzelne Produktionsschritte etabliert sowie auf Seiten der Distribution und Wiedergabe einheitliche Konventionen gefestigt. Dieser Prozess der Entwicklung eines Workflows steht bei objektbasiertem Audio generell, sowie insbesondere der objektbasierten Musikproduktion noch am Anfang. Es existiert eine Vielzahl an unterschiedlichen Begrifflichkeiten, Formaten und Codecs, welche sich auf alle Schritte der Produktionskette von der Mischung bis zur Wiedergabe auswirken.

Das Ziel dieser Thesis ist, aktuelle objektbasierte Produktionsketten am Beispiel von Dolby Atmos Music und 360 Reality Audio zu betrachten. Darauf basierend wird ein Workflow entwickelt und angewendet, welcher durch Nutzung einer einheitlichen Produktionsumgebung eine kombinierte Vorgehensweise zur Produktion von Dolby Atmos Music und 360 Reality Audio ermöglicht. Der Begriff „Workflow“ soll im Folgenden eine Reihung an Arbeitsabläufen beschreiben, die zur Produktion objektbasierter Musik nötig sind. Insbesondere bei der Verwendung neuartiger Technologien wie Dolby Atmos Music und 360 Reality Audio gilt es, eigene und bereits etablierte Workflows aus herkömmlichen Produktionsformaten mit neuen Methoden zu kombinieren, um eine möglichst effiziente Produktionsweise zu erreichen. Gegeben durch technologische Unterschiede (wie

Produktions-Software, Objekt-Rendering, Binauralisierung und Wiedergabeformate) soll ein Ziel dieser Arbeit sein, diese formatspezifischen Differenzen aufzuzeigen. Der Fokus dieser Thesis liegt auf der Entwicklung eines Workflows für immersive objektbasierte Musikproduktion auf Basis einer bereits angefertigten Stereo-Mischung. Die Vorgehensweise einer von Beginn an immersiven und objektbasierten Produktionsweise wird an dieser Stelle außer Acht gelassen, mögliche Workflow-Abweichungen dadurch werden in Kapitel 6 kurz betrachtet.

1.2 Aufbau der Thesis

Nachfolgend soll zur besseren Orientierung ein kurzer Überblick über diese Masterarbeit gegeben werden.

Kapitel 2 beschäftigt sich mit den für diese Thesis relevanten Grundlagen. Zum einen werden die produktionsseitig etablierten Konventionen des Stereo-Musikproduktions-Workflows dargestellt, um eine Basis für die Betrachtung des objektbasierten Workflows zu schaffen. Zum anderen wird das Thema Next Generation Audio (NGA) und die damit einhergehenden Produktions-, Distributions- und Wiedergabearten von kanalbasiertem, szenenbasiertem und objektbasiertem Audio betrachtet. Des weiteren wird kurz auf binaurales Audio eingegangen, da dieses Thema bei der Wiedergabe objektbasierter Musik über Streaming-Dienste in den Vordergrund rückt. Abschließend erfolgt eine nähere Beschreibung des Audio Definition Model (ADM).

Kapitel 3 stellt die objektbasierten Produktionsketten Dolby Atmos Music und 360 Reality Audio im Detail dar. Hierbei wird auf alle Produktionsschritte eingegangen und sowohl Software als auch Exportformate, zugrunde liegende Codecs und technische Vorgänge bei der Distribution beschrieben. Aktuelle Wiedergabemöglichkeiten beider Formate werden ebenfalls dargelegt. Zur besseren Darstellung wurden für beide Technologien Blockdiagramme erstellt, welche die gesamte Produktionskette beschreiben. Abschließend wird auf das Thema Mastering von Audioobjekten eingegangen.

Kapitel 4 beschäftigt sich mit der Entwicklung eines kombinierten Produktions-Workflows für Dolby Atmos Music und 360 Reality Audio durch Verwendung einer gemeinsamen Produktionsumgebung, basierend auf einer bereits angefertigten Stereo-Mischung.

Kapitel 5 dient der praktischen Anwendung des zuvor entwickelten kombinierten Workflows. Zuerst wird das vorhandene Stereo-Material (Pro Tools Projekt der Stereo-Mischung mit Einzelspuren) konzeptionell analysiert. Ferner werden die Produktionsschritte detailliert dokumentiert sowie auf die Punkte Export und Encodierung eingegangen.

Kapitel 6 stellt als Diskussion die Erkenntnisse der vorherigen Kapitel dar. Es werden zum einen technische Hintergründe wie Encodierung, Übertragung und Wiedergabemöglichkeiten kritisch betrachtet, zum anderen die Ergebnisse der praktischen Anwendung dargelegt. Insbesondere erfolgt eine Unterscheidung zum Stereo-Workflow sowie Erkenntnisse in Bezug auf objektbasierte Produktionsmethoden. Ferner wird auf die Bedeutung von Objekten in der Musikproduktion eingegangen.

Kapitel 7 bildet eine Zusammenfassung der Arbeit und ordnet die Erkenntnisse in den Gesamtzusammenhang ein.

Kapitel 8 dient als Ausblick und zeigt einige zukünftig relevante Forschungsthemen für den Bereich der objektbasierten Musikproduktion auf.

Aus Gründen der besseren Lesbarkeit wird in dieser Masterarbeit die Sprachform des generischen Maskulinums verwendet. Es wird an dieser Stelle ausdrücklich darauf hingewiesen, dass sämtliche Personenbezeichnungen gleichermaßen für alle Geschlechter gelten.

2 Grundlagen

„Populäre Musik [nimmt] zumeist nicht auf vorbestehende oder klanglich klar definierbare Aufführungssituationen Bezug. Sie entsteht nicht vor, sondern in dem technischen Medium, das hier neben der vermittelnden auch eine Produktionsfunktion besitzt. [...] Die Freiheit, nicht einem Ideal aufführungsbezogener Natürlichkeit genügen zu müssen, [erlaubt] eine besonders wirksame Unterstützung der Wahrnehmung musikalischer Strukturen.“
(Weinzierl, 2008, S. 779)

2.1 Workflow für Stereo-Musikproduktion

„Although most engineers ultimately rely on their intuition when doing a mix, they do consciously or unconsciously follow certain mixing procedures“ (Owsinski, 2006, S. 7). Jeder Tonschaffende verfügt durch Erfahrungswerte, Präferenzen und eine Vielfalt an weiteren Einflüssen über individuelle Workflows, die zahlreiche unbewusste, zeitliche und materialbezogene Entscheidungen bewirken. Nichtsdestotrotz gibt es eine Schnittmenge an allgemeingültigen Vorgehensweisen. Eine übliche Produktionskette bei der Stereo-Musikproduktion ist in Abb. 2.1 dargestellt. Insbesondere auf den Schritt der Mischung soll im Folgenden Bezug genommen werden.

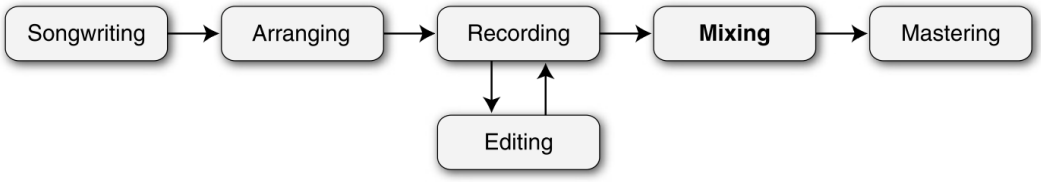


Abbildung 2.1: Übliche Produktionskette bei Stereo-Musikproduktion (Izhaki, 2008, S. 28).

2.1.1 Klangästhetik

Weinzierl (2008) unterscheidet beim Vorgang der Audiobearbeitung zwischen physikalischer und psychologischer Ebene. Während des Prozesses der Mischung werden „physikalische Maße des Audiosignals zuverlässig verändert: Amplituden- und Zeitwerte oder andere hieraus konstruierte Signalmaße wie Frequenz und Phasenlage, und zwar zumeist in Abhängigkeit voneinander“ (Weinzierl, 2008, S. 776). Der Klang, der durch diese Bearbeitung entsteht, wird schließlich auf psychologischer Ebene durch Merkmale wie Lautstärke, Tonhöhe, Positionierung und Räumlichkeit sowie der Klangfarbe wahrgenommen. Durch das Hören werden weitere Prozesse wie Erinnerungen oder Emotionen ausgelöst (Weinzierl, 2008). Als übergeordnete klangästhetische Prinzipien gelten „Symmetrie der Schallverteilung über die Stereobasis, Transparenz der Klangebenen, Richtungs- und Tiefenstaffelung der Schallinformationen“ (Dickreiter, Dittel, Hoeg & Wöhr, 2014, S. 335). Dies kann erreicht werden, indem die Audiobearbeitung am Inhalt orientiert wird:

> *„Sowohl die Ermittlung der musikalischen bzw. inhaltlichen Intention als auch die konkrete Festlegung von Klangeigenschaften ist jeweils ein klarer interpretatorischer Vorgang. Audiobearbeitung ist im Rahmen der Musikübertragung daher nicht nur ein technischer, sondern auch und gerade ein musikalischer Prozess“* (Weinzierl, 2008, S. 778).

Arrangement und Konzeption

Der Schritt des Arrangements (oder der Instrumentierung) erfolgt sowohl im Vorfeld der Aufnahme als auch während der Mischung selbst. Hier wird entschieden, ob bereits ein dichtes Klangbild vorliegt und somit Platz für einzelne Instrumente geschaffen werden muss, oder ob klangliche Lücken im Frequenzbereich oder der räumlichen Darstellung gefüllt werden müssen. Durch Analyse der einzelnen Spuren kann außerdem die Musikalität, Stimmung und emotionale Intention des Stückes festgestellt und darauf basierend ein klangliches Konzept erarbeitet werden (Owsinski, 2006; Izhaki, 2008).

2.1.2 Mischung

„A mix can, and should enhance the music, its mood, the emotions it entails, and the response it should incite“ (Izhaki, 2008, S. 4). Der Prozess einer Musikmischung lässt sich grob in drei Ebenen einteilen: Eine ausgeglichene Balance (Relation der Lautstärke der einzelnen Elemente), Frequenzdarstellung und Dynamik, sowie Räumlichkeit und Tiefe.

Alle dieser Elemente bauen aufeinander auf und beeinflussen sich gegenseitig, weshalb man in diesem Zusammenhang von einem iterativen Prozess sprechen kann (Owsinski, 2006; Izhaki, 2008). Eine übliche Reihenfolge der Arbeitsschritte bei Stereo-Mischungen ist in Abb. 2.2 dargestellt. Es gilt anzumerken, dass es sich hierbei um eine mögliche Vorgehensweise handelt, Abweichungen hiervon sind nicht unüblich. Dennoch liegt eine grundlegende Logik hinter diesem Workflow, gegeben durch den iterativen Prozess der Mischung. Izhaki (2008) beschreibt die Vorgehensweise des „dry-wet approach“, wobei zuerst alle Spuren „trocken“ gemischt werden (Lautstärke, Panning, EQ, Kompression) und erst danach die prozessierten Signale an Effekte gesendet werden, die klangliche Elemente hinzufügen (z.B. Hall oder Delay).

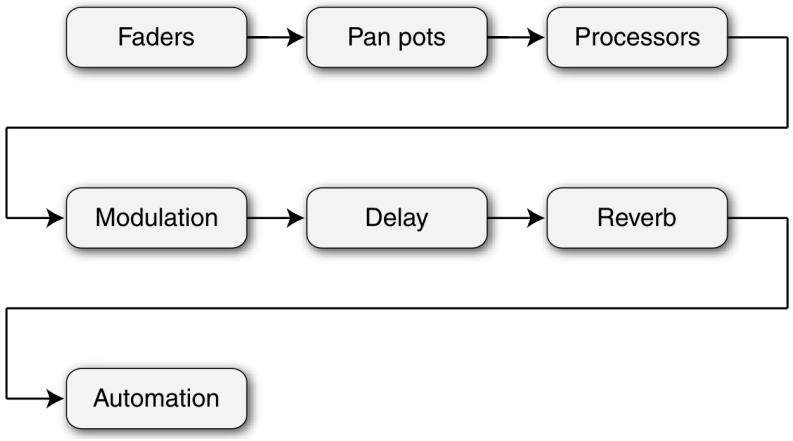


Abbildung 2.2: Übliche Reihenfolge der Arbeitsschritte bei Stereo-Mischungen (Izhaki, 2008, S. 37).

Die Erstellung von Gruppen, Aux-Spuren sowie damit einhergehende Routing-Einstellungen ermöglicht, dass sowohl Einzelspuren als auch Gruppen bearbeitet werden können. Besonders beim Einsatz von Effekten wird dieser meist nicht direkt auf die entsprechende Audiospur gelegt, sondern über einen Bus an eine Effekt-Aux-Spur gesendet, um das Ursprungssignal weiterhin trocken vorliegen zu haben (Izhaki, 2008). Für den Einsatz von Audiobearbeitungsmethoden werden meist Plugins verwendet.

> *„Plug-ins sind spezialisierte Software-Komponenten, die über standardisierte Schnittstellen (z. B. AU, VST) in integrierte Audioproduktionsumgebungen (digital audioworkstations, DAWs) eingebunden werden können. Sie bestehen aus einem Algorithmus zur Berechnung der Audiodaten und einer grafischen Benutzeroberfläche“* (Weinzierl, 2008, S. 720).

Equalizer (EQ)

Equalizer (EQ) sind Filter, die in allen Bereichen der Audiobearbeitung eingesetzt werden, um beispielsweise bestimmte Frequenzen zu betonen, zu entfernen oder Klänge zu verfremden (Weinzierl, 2008). Die Verwendung von EQs ermöglicht außerdem, das Frequenzspektrum einzelner Elemente so einzugrenzen, dass jedes Klangelement Platz in der Mischung bekommt und alle Elemente zusammen eine ausgewogenen Frequenzdarstellung erzielen (Owsinski, 2006; Izhaki, 2008).

Panning

Durch Einstellung des Panoramas (Panning) werden die Positionen von Klangelementen im Raum festgelegt. In der Musikproduktion werden Instrumente und Gesangsstimme häufig nach dem Kontrastprinzip gegenübergestellt: „Wichtige Monosignale mittig, sonstige Signale außen“ (Weinzierl, 2008, S. 729). Durch Panning entsteht sowohl Bewegung als auch eine Klarheit der Mischung – indem Elemente an unterschiedlichen Positionen abgebildet werden, kann somit Verdeckungseffekten entgegengewirkt werden (Owsinski, 2006; Izhaki, 2008).

Kompression

Der Einsatz von Kompression in der Musikproduktion dient der Anpassung von Lautstärke und Dynamikumfang und sorgt für eine Verdichtung des Gesamtklangs.

> *„Kompressoren dienen der Verringerung [...] der technischen Dynamik. Ein Kompressor ist ein Verstärker, dessen Verstärkungsfaktor sich bei über einer definierbaren Schwelle (threshold) liegenden Eingangssignalen in einem vorbestimmt festen oder variablen Verhältnis (ratio) verringert“* (Weinzierl, 2008, S. 730–731).

Eine spezielle Einsatzmöglichkeit ist außerdem die Multibandkompression: Hierbei wird das entsprechende Signal in verschiedene Frequenzbänder eingeteilt, welche einzeln komprimiert und schließlich wieder zusammengemischt werden. Somit wird ermöglicht, dass die Kompressionseinstellungen für verschiedene Frequenzbereiche variieren können. Ziel davon ist unter anderem eine höhere Lautheit (Owsinski, 2006; Weinzierl, 2008).

Delay

Die grundlegendste Anforderung an ein Delay ist die Verzögerung des Eingangssignals um eine bestimmte Zeitspanne. Die Wahrnehmung von verschiedenen Verzögerungszeiten variiert hierbei (Izhaki, 2008). Der Einsatz von Delay-Effekten bietet verschiedene Möglichkeiten, stereofone Effekte zu erzielen und wird in der Musikproduktion häufig an das Tempo des Stückes angepasst (Owsinski, 2006; Weinzierl, 2008).

> *„Je nach Kombination verschiedener Verzögerungszeiten, verschiedener Pegeldifferenzen, ein- und mehrmaliger Signalwiederholung, zeitlich regelmäßiger oder unregelmäßiger Signalwiederholung sowie räumlich identischer oder getrennter Wiedergabe von Original- und verzögerten Signalen kommen verschiedene psychoakustische Effekte wie Klangverfärbung, Raumeindruck, Echowahrnehmung und Lokalisationswirkungen (Summenlokalisation, Präzedenzeffekt oder Haas-Effekt) zum Tragen“* (Weinzierl, 2008, S. 748).

Hall

Durch den Einsatz von künstlichem Hall wird zum einen ermöglicht, die bereits aufgenommene Nachhalldauer und den räumlichen Klang von Musikaufnahmen zu verstärken oder zu verändern – dies kommt insbesondere in der Klassik-Produktion zum Einsatz. Zum anderen bietet sich die Möglichkeit, Instrumente im Studio trocken aufzunehmen und in der Postproduktion künstlichen Hall hinzuzufügen. Es lässt sich unterscheiden zwischen Hall, welcher eine reale bzw. virtuelle Räumlichkeit suggerieren soll und Hall, der die Klangfarbe verändern soll. Hierdurch entsteht eine Vielzahl an Optionen, Räumlichkeit und Charakter der Musikproduktion zu beeinflussen (Weinzierl, 2008; Izhaki, 2008).

2.1.3 Mastering

Technisch gesehen ist Mastering der Schritt zwischen Abnahme einer Mischung und deren Distribution. Mastering gilt als der letzte Schritt im kreativen Prozess, bei welchem die Mischung optimiert und vollendet sowie dafür gesorgt wird, dass die Produktion so gut wie möglich auf verschiedensten Wiedergabesystemen klingt. Es gilt als der Prozess, der Kunst und Technik vereint, bei welchem eine Kollektion an Titeln in ein Album verwandelt wird, indem Klang, Dynamik, Reihenfolge und Abstand zwischen einzelnen Liedern angepasst wird (Owsinski, 2006; Romanowski, 2020).

2.1.4 Lautheit

Durch das Maxim, immer lautere Musik zu produzieren, entstand der sogenannte „Loudness War“, ein Wettstreit der Musikindustrie, immer höhere Lautheitspegel zu erreichen. Die Anfänge sind auf die Digitalisierung der Musik Mitte der 1980er Jahre zurückzuführen, den Höhepunkt erreichte der Lautheits-Trend in den Jahren 2006–2008 (Robjohns, 2014). Das Kriterium der erhöhten Lautheit führt Weinzierl unter anderem darauf zurück, dass „das Klangbild (...) möglichen Unzulänglichkeiten der medialen Übertragungskette in dem Sinne Rechnung tragen [sollte], dass das inhaltlich oder musikalisch Wesentliche deutlich erkennbar übertragen wird“ (Weinzierl, 2008, S. 781). Dass lauter oft als „besser“ empfunden wird, ist auf die *Kurven gleicher Lautstärkepegel* zurückzuführen. Diese legen dar, dass die Wahrnehmung der Lautstärke unterschiedlicher Frequenzen bei gleichem Pegel variiert, sich jedoch bei erhöhter Lautstärke anpasst. Je lauter, desto linearer die Wahrnehmung des Frequenzspektrums (Weinzierl, 2008; Dickreiter et al., 2014).

Im Zuge des Loudness-War wurde durch Hyperkompression die Lautheit der Titel immer höher und die Dynamik somit immer kleiner. Als Reaktion darauf wurde schließlich von der European Broadcasting Union (EBU) die *EBU-Empfehlung R 128* veröffentlicht, wodurch Audioprogramme auf die wahrgenommene Lautstärke (Lautheit) normalisiert werden sollten (EBU, 2014). Während die *EBU-Empfehlung R 128* mit einer Normalisierung von –23 LUFS auf den Rundfunkbereich abzielt, hat die Audio Engineering Society (AES) eine Empfehlung für Lautheitsnormalisierung speziell im Streaming-Bereich veröffentlicht, wobei eine Normalisierung auf –16 LUFS empfohlen wird (Byers et al., 2015).

2.2 Next Generation Audio

Next Generation Audio (NGA), definiert in *DVB ETSI TS 101 154* (EBU, 2018), ermöglicht eine Vielzahl neuer Konzepte und Techniken zur Bereitstellung immersiver Audioinhalte und bietet eine größere Flexibilität in der Produktion und Distribution eben dieser (Olivieri, Peters & Sen, 2019). Audioinhalte können durch *kanalbasierte, szenenbasierte* oder *objektbasierte* Herangehensweisen produziert, übertragen und wiedergegeben werden. Audioobjekte sind hierbei einzelne Klangelemente, die sowohl in horizontaler als auch vertikaler Ebene mittels positionsbezogener Metadaten angeordnet und wiedergabeseitig flexibel auf das entsprechende System gerendert werden können.

Dieser neue Schritt im Produktions-Workflow zur Erstellung von Audio + Metadaten nennt sich *Authoring*. Sowohl der 360 Reality Audio zugrunde liegende Codec MPEG-H als auch AC-4, einer der Codecs auf denen Dolby Atmos Music basiert, gelten als NGA-Codecs und werden in Abschnitt 3.2.4 sowie Abschnitt 3.3.4 beschrieben.

2.2.1 Immersives Audio

Um dem Hörer ein möglichst realistisches und/oder immersives Hörerlebnis zu ermöglichen, gewann in den letzten Jahren die Höhendimension an Bedeutung – einige typische Beispiele für solche sogenannten „3D"-Lautsprechersysteme sind 7.1 mit zwei Höhenkanälen, 9.1 und 22.2. Zwar hat sich gezeigt, dass „3D"-Lautsprechersysteme eine räumliche Qualität liefern, die diejenige herkömmlicher „2D"-Systeme übertrifft, doch gibt es derzeit keinen „gemeinsamen Nenner" in der Vielzahl möglicher Lautsprechersysteme, der eine Interoperabilität zwischen Produzenten, Geräteherstellern und Verbrauchern in der gleichen Weise gewährleisten könnte, wie das 5.1-System bisher als gemeinsamer Nenner für den Surround-Sound gedient hat (Herre, Hilpert, Kuntz & Plogsties, 2015). Mögliche Lösungsansätze hierfür können sowohl objektbasierte Produktionen als auch NGA-Standards wie MPEG-H sein. Diese bieten die Möglichkeit einer qualitativ hochwertigen Wiedergabe für eine Vielzahl von Ausgabeformaten, jeweils angepasst an das Wiedergabesystem (Herre et al., 2015).

Es gilt herauszustellen, dass immersives Audio (oder 3D-Audio) nicht gleichzusetzen ist mit objektbasiertem Audio. Ersteres beschreibt die räumliche Dimension einer Produktion, letzteres beschreibt die Art und Weise, wie die Audioinhalte produziert, übertragen und wiedergegeben werden. Im Rahmen dieser Thesis soll es um die Betrachtung objektbasierter Herangehensweisen gehen, die immersiv produziert werden.

2.2.2 Kanalbasiertes Audio

Der Prozess der Positionierung von Klangelementen hat sich seit den 1930er Jahren mit der Entwicklung der Stereofonie etabliert. Aufgrund technischer Begrenzungen der Distributions- und Wiedergabemöglichkeiten wurden diese räumlich positionierten Klangelemente jedoch lange in die diskreten Signale der einzelnen Lautsprecherkanäle eingerechnet, welche direkt auf den jeweils zugehörigen Lautsprechern abgespielt werden.

Kanalbasiertes Audio (z.B. Stereo, 5.1 Surround) erfordert somit für jedes Wiedergabesystem eine an das entsprechende Lautsprecher-Layout angepasste Mischung. Eine Ausnahme hierzu stellen beispielsweise kanalbasierte Systeme wie *MPEG-Surround* oder *Dolby ProLogic* dar, welche mittels parametrischer Audiocodierung auf verschiedenen Lautsprecher-Layouts wiedergegeben werden können (siehe Abschnitt 2.2.4).

Bereits während der Mischung werden die Positionen der Klangelemente impliziert, unter Annahme spezifischer Lautsprecherpositionen beim Konsumenten – sofern diese nicht den angenommenen Positionen entsprechen, kann dies den angestrebten räumlichen Klang negativ beeinflussen (Coleman et al., 2016; Tsingos, 2017; Rumsey, 2018). Um kanalbasierte Inhalte auf einem anderen Lautsprecher-Setup als dem bei der Produktion gewählten wiederzugeben, sind zusätzliche Schritte erforderlich. Die Wiedergabe von Audioinhalten auf Systemen mit einer geringeren Anzahl an Lautsprechern erfordert *Downmixing*[1], während die Wiedergabe auf Systemen mit einer höheren Anzahl an Lautsprechern durch *Upmixing*[2] erreicht werden kann. In beiden Fällen kann es durch diese Konvertierung zu Qualitätsverlusten kommen, beispielsweise kann Downmixing bei kohärenten Signalkomponenten einen Kammfiltereffekt verursachen (Herre et al., 2015).

2.2.3 Szenenbasiertes Audio

Szenenbasiertes Audio basiert auf Ambisonics und bildet die Audioszene auf einer Reihe von orthogonalen Basisfunktionen ab, welche im Wiedergabegerät des Konsumenten decodiert und auf das entsprechende Wiedergabesystem angepasst werden (Hestermann, Seideneck & Sladeczek, 2018; Olivieri et al., 2019). Diese sogenannten *räumlichen Harmonischen* repräsentieren Schalldruck- und Schallgeschwindigkeitsvektoren einer Szene an einem bestimmten Punkt. Je nach Anzahl dieser räumlichen Harmonischen kann die Richtungsauflösung der Szene höher oder niedriger sein, und die Ergebnisse werden in unterschiedlicher räumlicher Darstellung wiedergegeben. Je höher die *Ordnung*, desto genauer die Auflösung. Auf diese Weise erstellte Audioinhalte können problemlos modifiziert und gedreht werden. Im Gegensatz zu Objekten werden räumliche Informationen in HOA jedoch nicht in speziellen geometrischen Metadaten, sondern in den Koeffizien-

[1] In der Audiotechnik beschreibt der Vorgang des „Downmixings“ ein manuelles oder automatisches Verfahren, bei dem die Audiosignale eines Mehrkanal-Formates auf ein Format mit einer geringeren Anzahl an Kanälen zusammengefasst werden.

[2] Der Begriff „Upmixing“ beschreibt das Gegenteil von Downmixing. Bei einem Upmix wird eine gewisse Anzahl an Audiokanälen in eine größere Anzahl an Kanälen umgewandelt.

tensignalen selbst übermittelt. Die szenenbasierte Produktionsweise ist insbesondere für Inhalte passend, die als zusammengesetzte räumliche Audioszenen unter Verwendung einer begrenzten Anzahl von Audiostreams dargestellt werden müssen, wie beispielsweise die Aufnahme des räumlichen Tons für VR oder 360 Grad Videos (Oldfield, Shirley & Spille, 2014; Herre et al., 2015; Rumsey, 2018).

2.2.4 Objektbasiertes Audio

In den letzten Jahren hat insbesondere *objektbasiertes Audio* stark an Bedeutung gewonnen, und findet nun mittels objektbasierter Technologien wie Dolby Atmos Music[3] und 360 Reality Audio auch Zutritt zum Musik-Streaming-Markt (Fraunhofer IIS, 2019; Sony Corporation, 2019b; Kenny, 2020).

Im Vergleich zum kanalbasiertem Audio, bei dem die Lautsprechersignale als Produktions-Endformat gespeichert und übertragen werden, werden bei objektbasiertem Audio die Audioinhalte durch *Audioobjekte* beschrieben. Ein Audioobjekt kann dabei als eine virtuelle Klangquelle betrachtet werden, die aus einem Audiosignal mit zugehörigen Metadaten wie beispielsweise Positionen oder Lautheit besteht. Audioobjekte können unabhängig der Lautsprecherpositionen im Raum platziert werden – hierbei gilt zu unterscheiden zwischen *dynamischen* und *statischen Objekten*. Im Gegensatz zu statischen Objekten verändern sich die Positionen dynamischer Objekte über die Zeit. Auch bei objektbasierten Produktionen können Kanäle verwendet und übertragen werden, die an festen Lautsprecherpositionen platziert sind. Ein Beispiel hierfür ist Dolby Atmos (Herre et al., 2015; Tsingos, 2017; Hestermann et al., 2018).

[3]Hierbei gilt anzumerken, dass es sich nur bei 360 Reality Audio um eine rein objektbasierte Technologie handelt. Dolby Atmos ist eine hybride Technologie (kanal- und objektbasiert), kann jedoch auch ausschließlich kanalbasiert übertragen werden. Dies wird im Folgenden bei allen Nennungen von Dolby Atmos Music impliziert, zur Vereinfachung wird jedoch sowohl auf 360 Reality Audio als auch Dolby Atmos Music als objektbasierte Technologie referenziert.

Spatial Audio (Object) Coding

Eine Möglichkeit der flexiblen Anpassung von kanalbasierten Signalen stellen Systeme wie *MPEG-Surround* oder *Dolby ProLogic* dar, welche mittels parametrischer Audiocodierung auf flexiblen Lautsprecher-Layouts wiedergegeben werden können. Diese Art der Codierung wurde mit *Parametric Stereo* (parametrische Darstellung von stereophonem Audio basierend auf Mono-Downmixes) eingeführt, und mit *MPEG Surround* als ISO-Standard für die parametrische Mehrkanalerweiterung monophoner oder stereophoner Audiocodecs fortgeführt. Dieses Konzept der Parametrisierung wird als Spatial Audio Coding (SAC) bezeichnet. Hierbei wird ein Mono/Stereo-Downmix der Eingangssignale (z.B. Surround Sound) erstellt, sowie akustische Parameter wie Interaural Level Difference (ILD), Interaural Time Difference (ITD) und Interaural Coherence (IC) als Nebeninformationen extrahiert, um eine parametrische Rekonstruktion des Eingangssignals zu ermöglichen (Herre et al., 2012; Jia, Zhang, Bao & Zheng, 2017; Tsingos, 2017).

MPEG-D Spatial Audio Object Coding (SAOC) ist der nächste Schritt in der parametrischen Audiocodierung und weicht vom traditionellen Ansatz ab, mehrere Audiokanäle zu codieren, die für die Wiedergabe auf einem bestimmten Lautsprecher-Setup bestimmt sind. Stattdessen wird eine gemeinsame Codierung mehrerer Audioobjekte durchgeführt. Basierend auf der SAC-Technologie bietet SAOC neue Eigenschaften: Zum einen wird die Darstellung von Audioinhalten mit einer Vielzahl gleichzeitig vorhandener Objekte zu Bitraten bewirkt, die bisher für die Codierung von Mono- oder Stereo-Signalen verwendet wurden. Zum anderen ermöglicht es die rechnerisch effiziente und flexible Wiedergabe solcher Audioinhalte auf allen gängigen Lautsprecheranordnungen oder als binaurale Ausgabe über Kopfhörer. Außerdem kann der Wiedergabeprozess vollständig vom Benutzer auf Seite des Decoders gesteuert werden, was Personalisierung und Nutzerinterktivität ermöglicht (Herre et al., 2012).

Rendering

Um objektbasierte Produktionen wiederzugeben, wird ein sogenannter Audio-Renderer benötigt. Das Audio-Rendering stellt den Prozess der Generierung von Lautsprecher- oder Kopfhörersignalen dar, die auf Grundlage der jeweiligen Wiedergabesysteme der Endkonsumenten (Lautsprechersysteme, Soundbars, Kopfhörer, mobile Endgeräte usw.) erstellt werden und die Objekte innerhalb dieser Umgebung passend berechnen und wiedergeben (Herre et al., 2015; Tsingos, 2017; Hestermann et al., 2018).

Aktuelle objektbasierte Systeme verwenden Metadaten, die sich auf die jeweiligen Eigenschaften jedes Objekts beziehen. Um Lautsprechersignale für das spezifizierte Wiedergabesystem zu erzeugen, werden Rendering-Techniken wie Vector Base Amplitude Panning (VBAP) oder Distance-Based Amplitude Panning (DBAP) verwendet (Lossius, Baltazar & de la Hogue, 2009). Vector Base Amplitude Panning spielt insbesondere beim Panning von Objekten auf 3D Lautsprecher-Layouts eine Rolle – hierbei werden jeweils drei (je nach System physische oder virtuelle) Lautsprecher verwendet, um ein Audioobjekt mit einer bestimmten Einfallsrichtung auf den Hörer an die entsprechende Position zu rendern. Da vektorbasiertes Panning nur die Richtung einer Schallquelle relativ zur Referenzposition verwendet, kann es nicht unterscheiden zwischen mehreren Objekten an unterschiedlichen Positionen desselben Richtungsvektors. Richtungsabhängiges Panning kann außerdem hörbare Übergänge zwischen den einzelnen Lautsprechern erzeugen, wenn sich Objekte der Raummitte nähern, wo eine kleine Bewegung der Position eines Objekts nicht immer zu einer kleinen Variation der Lautsprecher-Gains führt. Um diese Probleme zu beheben, ist eine entfernungsabhängige Überblendung nötig, bei welcher der Panning-Algorithmus vom ursprünglichen richtungsabhängigen Verhalten dazu übergeht, beispielsweise alle Lautsprecher gleichmäßig anzusteuern, wenn sich das Objekt dem Ursprung der Referenzposition – d.h. dem Sweet Spot oder der Raummitte – nähert (Tsingos, 2017).

Bei bewegten Objekten werden die Gain-Werte über kurze zeitliche Abschnitte ausgewertet und interpoliert. Dies geschieht entweder direkt oder unter Verwendung einer rekonstruierten segmentierten Faltung des Ausgangs-Audiosignals. Audio-Renderer tasten üblicherweise die eingehenden Objektkoordinaten auf eine feste Audio-Framerate (z.B. 100 Hz) ab, mit welcher die Auswertung der Panning-Werte erfolgt. Diese Werte werden anschließend weiter pro Sample interpoliert, passend zur Frame Rate (z.B. 48 kHz) (Tsingos, 2017).

Da beim Hörer die Lautsprechersignale für das entsprechende Wiedergabesystem gerendert werden, wird die objektbasierte Audioübertragung oft als formatagnostisch bezeichnet (Shirley, Oldfield, Melchior & Batke, 2013; Woodcock et al., 2018).

Eine objektbasierte Produktion und Distribution ermöglicht insbesondere: verstärkte Immersion durch Hinzufügen der Höhen-Ebene in Kombination mit flexiblem Wiedergabe-Rendering wie beispielsweise Binauralisierung; erweiterte Möglichkeiten zur Personalisie-

rung der Audioinhalte für den Konsumenten durch Auswahl von Presets (Dialogverbesserung, Auswahl mehrerer Sprachen, etc.) oder Anpassung ausgewählter Parameter der Mischung (z.B. Lautstärke oder Positionen einzelner Klangobjekte); Rendering des Inhaltes auf das jeweilige Wiedergabesystem; verbesserte Barrierefreiheit durch Dialogverbesserung und Audiodeskription; die Möglichkeit für zukunftssichere Produktions-Workflows, die es ermöglichen, aus einem einzigen objektbasierten Mix weitere Mischungen zu erstellen (Tsingos, 2017).

2.3 Binaurales Audio

> *„Binaural sound refers to the two-channel sound that enters a listener's left and right ears. Although many could argue that all stereo sound is binaural, the term binaural is reserved for sound where the two-channel sound entering listeners ears has been filtered by a combination of time, intensity, and spectral cues intended to mimic human localisation cues.“*
>
> (Roginska, 2017, S. 88)

Die Wiedergabe von Musik über Kopfhörer hat im Zeitalter der mobilen Endgeräte stetig an Bedeutung gewonnen. Neben herkömmlichem Stereo-Ton existiert eine weitere Wiedergabemethode bestehend aus zwei Kanälen: Binaurales Audio, welches sich auf den Zweikanalton bezieht, der in das linke und rechte Ohr des Hörers gelangt. Im Gegensatz zur Stereo-Wiedergabe über Kopfhörer werden binaurale Signale durch eine Kombination von Zeit-, Intensitäts- und Spektralfunktionen gefiltert, um die Lokalisierung nachzuahmen. Die Wahrnehmung von binauralem Ton beruht auf der Interaural Time Difference (ITD) und Interaural Level Difference (ILD). Zusätzlich dazu liefern spektrale Werte dem Hörer weitere Informationen über die Position einer Schallquelle. Die Zusammensetzung aus ITD, ILD und den spektralen Eigenschaften wird in Head-Related Transfer Functions (HRTF) erfasst. Die HRTF ist die Frequenzbereich-Darstellung der Head-Related Impulse Response (HRIR) (Roginska, 2017). Detaillierte Ausführungen zum Thema binaurales Hören finden sich unter Anderem bei Blauert (1974).

Binaurales Rendering für die Kopfhörerwiedergabe unter Verwendung eines Satzes binauraler Raumimpulsantworten (Binaural Room Impulse Response (BRIR)) ist somit eine weitere Methode zur Darstellung einer immersiven, räumlichen 3D-Audioszene. BRIR

beschreiben die akustische Übertragung von einem Punkt des Raumes zu den Ohren des Hörers. BRIR werden üblicherweise mit Mikrofonen im Gehörgang gemessen oder mittels Modellen erstellt. Der direkte Anteil der BRIR wird als kopfbezogene Impulsantwort (HRIR) bzw. als deren Fourier-Transformation HRTF bezeichnet, die für eine präzise Lokalisierung unerlässlich ist und von Person zu Person variiert.

Die beim natürlichen Hören vorkommenden psychoakustischen Effekte werden beim Binaural-Rendering durch digitale Signalverarbeitungstechniken simuliert. Dadurch sollen realistische Hörereignisse durch präzise Steuerung der Schalldrucksignale, die das Trommelfell erreichen, erzeugt werden. Basierend auf dem Verständnis des räumlichen Hörens wurden binaurale Wiedergabeverfahren entwickelt, die in der Lage sind, Simulationen zu erzeugen, die von realen Schallereignissen nicht unterscheidbar sind. Dieser Grad an Realismus erfordert jedoch hochgradig kontrollierte Bedingungen, einschließlich akustischer In-situ-Messungen (im Gehörgang), die speziell auf den einzelnen Hörer zugeschnitten sind. Da die Praktikabilität und Flexibilität fehlt, die für Verbraucheranwendungen nötig wäre, wird die Individualisierung aktuell außerhalb von Hörversuchen nur selten durchgeführt. Auch nicht individualisierte binaurale Rendering-Systeme können in der Lage sein, einen überzeugenden räumlichen Eindruck zu vermitteln. Es gab jedoch bisher keine eindeutigen Hinweise darauf, dass dies das gesamte Hörerlebnis mit Kopfhörern in Multimedia-Anwendungen im Vergleich zu herkömmlichen Stereo-Inhalten verbessern kann (Pike, 2019).

Weiterführend wurde in einem von Pike (2019) vorgestellten Hörversuch die Qualität verschiedener repräsentativer Ansätze zur binauralen Wiedergabe räumlicher Audiosignale in Verbraucheranwendungen mit *Head-Tracking* ermittelt. Head-Tracking ermöglicht es einer Anwendung, die Kopfbewegungen eines Benutzers zu erkennen. Dies kann mittels spezieller Soft- oder Hardware erfolgen, beispielsweise durch Einbauen eines Headtrackers in Kopfhörer. Da Kopfbewegungen ein wichtiger Teil der auditiven Wahrnehmung sind und Änderungen der Kopfposition die Signale am Trommelfell verändern, kann Head-Tracking den räumlichen Eindruck verbessern. „The aim is to keep the virtual auditory space fixed while the head is moving, rather than it moving with the head in an unrealistic way“ (Pike, 2019, S. 77–78). In o.g. Studie wurde ein objektbasierter Ansatz mit verschiedenen kanal- und szenenbasierten Darstellungen unterschiedlicher Komplexitätsstufen verglichen. Um verschiedene Qualitätsmerkmale zu untersuchen, wurden bei der Wiedergabe dieser Formate sowohl einzelne musikalische Klangquellen als

auch komplexe 3D-Audioszenen verwendet. Das binaurale Rendering der objektbasierten Darstellung durch separate HRTF-Faltung für jede Klangquelle erwies sich als wesentlich hochwertiger als alle Lautsprecher-Virtualisierungsmethoden, welche zur binauralen Wiedergabe kanalbasierter Formate verwendet werden. Der Ambisonics-Ansatz erster Ordnung zeigte eine besonders schlechte Qualität. Der objektbasierte Ansatz ergab einen klaren, natürlichen Eindruck mit guten räumlichen Eigenschaften (Pike, 2019).

Weitere Ausführungen zum Bereich der Binauraltechnik würden über den Rahmen dieser Thesis hinausgehen, weshalb an dieser Stelle auf weitere Publikationen verwiesen werden soll (Blauert, 1974; Roginska, 2017; Wenzel, Begault & Godfroy-Cooper, 2017; Pike, 2019).

2.4 Audio Definition Model (ADM)

Durch die vermehrte Produktion immersiver und interaktiver Audioinhalte steigen auch die Anforderungen an flexiblere Audioformate: Insbesondere szenen- und objektbasierte Herangehensweisen erfordern neue Wege, um die verschiedenen Arten von Audio auszutauschen, zu übertragen und zu archivieren. Eine zentrale Voraussetzung hierfür ist es, dass unabhängig der verwendeten Produktionsweise Metadaten zur Beschreibung der Audioinhalte existieren: „Each individual track within a file or stream should be able to be correctly rendered, processed or distributed according to the accompanying metadata“ (ITU-R, 2015, S. 6). *ITU-R BS.2076-2* beschreibt weiterhin, dass es insbesondere für die flexible Wiedergabe von objektbasiertem Audio erforderlich ist, dass Metadaten zur Beschreibung der Merkmale und Beziehungen der einzelnen Objekte mitgeliefert werden. Ein Metadatenmodell, das diesen Anforderungen entspricht, ist das Audio Definition Model (ADM), spezifiziert in *Recommendation ITU-R BS.2076* (ITU-R, 2015). Das Audio Definition Model beschreibt verschiedene Arten von Audioinhalten (einschließlich kanal-, objekt-, und szenenbasierten Darstellungen) durch Verwendung der Extensible Markup Language (XML), was ermöglicht, dass ADM von Menschen gelesen und editiert werden kann. ADM wurde insbesondere zur Nutzung in Kombination mit RIFF/WAVE-basierten Workflows entwickelt, um eine Methode zum Austausch immersiver Audioformate zu bieten. RIFF/WAVE-Dateien bestehen aus mehreren „chunks“, wobei ein „chunk“ aus einer Vielzahl an Informationen besteht. Um ADM-Metadaten zu übertragen, werden spezielle „chunks“ zur Unterstützung von XML-Metadaten benötigt (Füg, Marston & Norcross, 2016).

Das Audio Definition Model ist in zwei Bereiche unterteilt – den *Format-Teil* sowie *Content-Teil*. Ersterer beschreibt technische Details des Audios, die für korrekte Decodierung oder Rendering relevant sind (beispielsweise, ob es sich um Kanäle oder Objekte handelt), letzterer beschreibt die Audioinhalte und verknüpft diese mit dem Format. Im Format-Teil wird jeder Audiokanal einem der vordefinierten Typen „Objects“, „DirectSpeakers“, „Binaural“, „HOA“ oder „Matrix“ zugeordnet. Ist im Format-Teil etwas als „Object“ definiert, so werden die objektbasierten Inhalte mittels weiterer Attribute und Sub-Elementen (z.B. Position, Gain, Size) im Detail beschrieben. Diese Attribute definieren außerdem, ob das Objekt beispielsweise als Phantomschallquelle oder auf den nächsten Lautsprecher gerendert werden soll (Füg et al., 2016; Geier, Carpentier, Noisternig & Warusfel, 2017).

BW64 Format

Ein Beispielformat, welches ADM-Metadaten unterstützt, ist *BW64*, ein WAVE-basiertes Dateiformat zur Übertragung kanal-, objekt-, und szenenbasierter Audioinhalte. Es ist in *Recommendation ITU-R BS.2088* spezifiziert (ITU-R, 2019a) und wurde insbesondere zum Transport von ADM-Metadaten entwickelt. BW64 gilt als Nachfolger des Broadcast Wave File (BWF) und kann im Vergleich zu diesem mehr Daten ($>$ 4 GByte) umfassen, was insbesondere in der Anwendung größerer, objektbasierter Produktionen einen Vorteil bietet. Außerdem wurden zwei ADM-spezifische „chunks“ definiert: *<chna> (channel allocation)* verweist jeden Kanal der BW64-Datei auf die entsprechenden ADM-Metadaten Bezeichner. Die Speicherung und Übertragung der ADM-Metadaten ist im *<axml>* „chunk“ definiert. Die Kombination aus ADM und BW64 ermöglicht somit Workflows, bei welchen sowohl mit herkömmlichen, auf WAVE-Dateien basierten Inhalten sowie neueren, immersiven Inhalten gearbeitet werden kann (Füg et al., 2016; ITU-R, 2019a).

2.4.1 ADM Profile

Im Produktions-Workflow können die ADM-Metadaten, bestehend aus Parametern wie Abtastraten, Bittiefen und Anzahl der Spuren sowie Kanal-IDs, Lautheit und Objektpositionen, sowohl automatisch als auch manuell generiert werden, abhängig der verwendeten Hard- und Software (Füg et al., 2016). Dieser Prozess der Metadatengenerierung nennt sich Authoring und kann softwareseitig beispielsweise mittels des Fraunhofer MPEG-H Authoring Plugins (MHAPI) oder Dolby Atmos Renderers erfolgen.

Die ADM-Spezifikationen beschreiben sowohl unterstützte Elemente, Unterlemente und Attribute sowie Einschränkungen für Datentypen als auch Anforderungen an Verhältnisse der Elemente untereinander. Zusätzliche Anforderungen an ADM-Metadaten werden durch anwendungsspezifische ADM-Profle erstellt – so ermöglicht beispielsweise das MPEG-H-ADM-Profil eine Interoperabilität mit etablierten NGA Produktions- und Distributionssystemen für MPEG-H 3D Audio (Fraunhofer IIS, 2020a). Dies ist besonders hinsichtlich der Distribution immersiver und interaktiver Audioinhalte relevant, die es dem Nutzer ermöglichen, sich das Audio an persönliche Bedürfnisse anzupassen (Wiedergabesprache, Dialogverbesserung etc.). Eine Übereinstimmung mit den ADM-Spezifikationen (und denen der verwendeten ADM-Profile) ist sicherzustellen, da sonst bei ADM-Tools beim Einlesen der Metadaten Probleme auftreten können. Um diesen Prozess zu vereinfachen wurde vom Fraunhofer-Institut für Integrierte Schaltungen IIS das Fraunhofer ADM Info-Tool entwickelt, welches automatisierte Tests von ADM-Inhalten durchführen kann, basierend auf *ITU-R BS.2076-2, ITU-R BS.2088-1* und *ITU-R BS.2125-0* sowie aller ADM-Profile die vom MPEG-H 3D Audio Standard unterstützt werden. Weitere ADM-Profile sind das *Atmos*-Profil und das *EBU ADM Broadcast Production* Profil, die Kompatibilität der Profile untereinander ist abhängig davon, ob jeweils eine Konvertierung implementiert wurde oder nicht (Fraunhofer IIS, 2020a).

2.4.2 S-ADM

Während mit ADM-Metadaten Austausch und Übertragung immersiver kanal-, szenen- und objektbasierter Audioinhalte möglich ist, so eignen sich die in *Recommendation ITU-R BS.2076* definierten Spezifikationen nicht für lineare Anwendungen wie (Musik)-Streaming. Da bei Audio-Streaming-Anwendungen die Audiodatei entweder in Frames zerlegt wird oder Frames generiert werden, welche in Echtzeit über Schnittstellungen wie AES, MADI, SDI und IP-Netzwerke bereitgestellt werden, ist ein serielles ADM-Format erforderlich, welches das Slicing von Audio und zugehörigen Metadaten ermöglicht. Das in *Recommendation ITU-R BS.2125-0* definierte Serial-ADM (S-ADM) wurde somit speziell für den Einsatz in linearen Arbeitsabläufen wie Live- oder Echtzeit-Produktion für Broadcast- und Streaming-Anwendungen konzipiert.

S-ADM hat die gleichen Strukturen, Attribute und Elemente wie ADM, sowie zusätzliche Attribute zur Beschreibung des Frame Formats. Ein Frame von S-ADM-Metadaten enthält ein Set an Metadaten, die das Frame über den zugehörigen Zeitraum (und darüber

hinaus) beschreiben. Wie in Abb. 2.3 dargestellt, überlappen sich die S-ADM-Frames nicht, sondern sind mit einer festen Startzeit und Dauer direkt aneinander angrenzend (ITU-R, 2019b).

S-ADM hat neben der Kompatibilität mit ADM-Elementen weitere Spezifikationen wie eine unlimitierte Anzahl an Audiospuren, Unabhängigkeit von Transport oder Schnittstellen Methoden, Handhabung aller Kombinationen von kanal-, objekt-, und szenenbasierten Audioinhalten, keine Größenbeschränkung der Frames oder Unterstützung von Direktzugriff. Letzterer Punkt bedeutet, auf jedes Frame des Flows (eine Sequenz von S-ADM Frames, Äquivalent zu einer Datei im ADM) zugreifen und es vollständig decodieren zu können (ITU-R, 2019b).

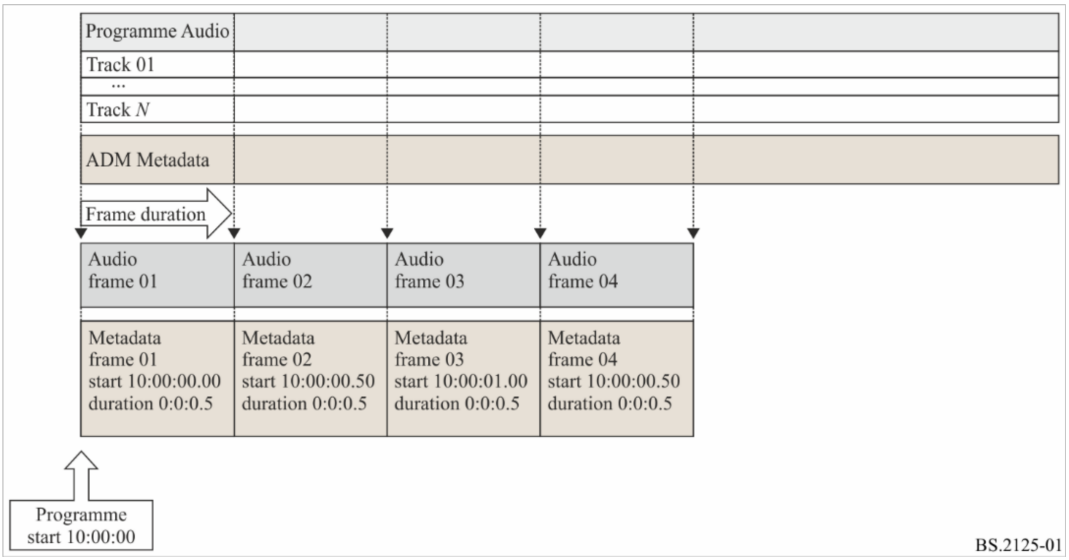


Abbildung 2.3: Aufbau des S-ADM (ITU-R, 2019b, S. 3).

3 Objektbasierte Produktionsketten

> *„The success of the immersive audio industry relies upon not one but at least three key concerns: first, improving the technical delivery from concept to listener; second, creating new ideas for aesthetic uses of height channels and binaural applications; third, developing sufficient experience to establish a workflow that allows delivery in all of the main formats simultaneously.“*
> Theile und Wittek, Van Baelen, zitiert nach (Lawrence, 2019, S. 152)

Im Folgenden werden die Produktions- und Distributionsketten für Dolby Atmos Music (DAM) und 360 Reality Audio (360 RA) betrachtet, sowie darauf basierend ein Vorschlag eines Workflows zur kombinierten Produktion erstellt (siehe Kapitel 4), welcher in Kapitel 5 praktisch erprobt wird.

3.1 Immersive Konzeption

Generell gilt bei immersiven Musikmischungen, dass das räumliche Klangfeld an die Art der Musik angepasst werden muss. Durch die Dreidimensionalität bieten sich viele neue Möglichkeiten, Klangquellen zu gestalten, platzieren und zu bewegen. Insbesondere letzteres sollte mit Vorsicht bedacht werden – auch wenn die Möglichkeiten gegeben sind, so muss der Hörer und die Art der Musik im Vordergrund stehen:

> *„Every piece of music is unique and needs to be approached as such. Let the music and production dictate how aggressively you use the system and the space. Some music lends itself to elements moving around the room, or elements hard-panned to the rear or height channels, and some does not. [...] Having all this new space to work with is fantastic, but you can definitely get caught up and overdo it very easily. I always try to remember that at some point, someone is going to want to sit in their house and simply listen to music. It's still our job to deliver that experience, not to show off our cool new speaker system“* (Genewick, 2020, S. 42).

Je nachdem wie ein Titel aufgebaut ist, kann es sich anbieten, das horizontale und vertikale räumliche Feld erst nach und nach einzuführen und auszuweiten. Das so entstehende Wechselspiel von nahen Klängen und einhüllender Räumlichkeit bietet eine vielfältige Möglichkeiten, die Mischung aufzubauen. Allerdings kann durch einen fortwährend dezenten Einsatz der Höhenkanäle (beispielsweise durch Hall, Delays oder geringe Anteile von Signalen der Surround-Ebene) ein natürlicheres Gefühl der Umhüllung geschaffen werden, wenn letztendlich einzelne Klangquellen oben positioniert werden (Schultz, 2020a; Genewick, 2020). Ein wichtiger Aspekt bei immersiven Musikmischungen ist außerdem, bei der Verteilung der Klangquellen darauf zu achten, dass der Gesamtklang nicht zerfällt: „This modern music, a sort of wall of sound, sometimes doesn't translate as well when you pick it apart and move things. Are you collapsing the imaging, are you collapsing the vibe of the song?“ (Harvey, 2020, S.30).

Bei der Konzeption der objektbasierten immersiven Mischung spielen psychoakustische Effekte eine bedeutende Rolle, da die Wahrnehmung räumlicher Informationen dadurch maßgeblich beeinflusst wird. Phantomschallquellen werden auf der horizontalen Achse durch Zeit- und Pegelunterschiede zwischen beiden Ohren erzeugt - dadurch, dass die Ohren seitlich am Kopf positioniert sind, ist diese Wahrnehmung auf vertikaler Achse beeinträchtigt. Ferner ist bekannt, dass hohe Frequenzen tendenziell eher als „oben“ wahrgenommen werden und dass die Lokalisation einer Schallquelle deutlich erschwert wird, je tiefer die Frequenzanteile sind. Daher kann bei der Verwendung von Höhenkanälen die Reduzierung tieffrequenter Anteile sowie eine Verstärkung hoher Frequenzen zu einer verbesserten „oben“ Lokalisation führen (Lee, 2017). Außerdem gilt in Bezug auf Höhenkanäle, dass ab einem vertikalen Elevations-Winkel von mehr als 40 Grad die vertikale Kohärenz verloren geht – diese ist jedoch entscheidend, um für den Hörer ein natürliches immersives Klangerlebnis zu schaffen (Lee, 2017; Lawrence, 2019).

> *„From an audience perspective, sounds perceived in space are not only subjective to each individual listener; the listener's perceived position and the process in which the brain processes, and focuses in on, specific sounds is a complex one. The way we experience sound is beyond the mechanics of simply hearing a sound in the space we perceive“* (Lawrence, 2019, S. 149).

Die wahrgenommene räumliche Tiefe einer Mischung, auch „true envelopment, engulfment, or immersion“ (Lawrence, 2019, S. 144) (wobei der Begriff „Engulfment“ erstmals von Sazdov (2007) eingeführt wurde) erhält durch eine Wiedergabe auf mehr als zwei

Lautsprechern eine größere Detailschärfe, weshalb der Aspekt der Räumlichkeit im immersiven Produktions-Workflow größere Beachtung finden sollte. Ziel eines immersiven Mixes kann sein, eine klangliche Klarheit und damit verbundene Verbesserung der räumlichen Darstellung zu schaffen, beispielsweise durch EQ, Kontrolle des Dynamikumfangs, Panning, Delay, Hall etc. Ein weiterer Parameter ist die spektrale Balance, welche die Ausgewogenheit der gesamten Frequenzverteilung beschreibt (Lawrence, 2019).

Der Einsatz von Hall kann auf verschiedene Arten erfolgen: Zum einen durch Platzierung eines Mehrkanal-Mikrofonarrays bei der Musikaufnahme, zum anderen über Hall-Plugins. Hierbei bietet sich sowohl die Möglichkeit, 3D Hall-Plugins einzusetzen, oder auch herkömmliche Stereo- oder Surround-Raumsimulationen zu verwenden, beispielsweise durch Kombination von Hall-Kanälen.

Die Verwendung von Bett und Objekten spielt bei objektbasierter Musikproduktion ebenfalls eine Rolle – werden bei Dolby Atmos Music tatsächlich Kanal-Betten verwendet, so bestehen diese bei 360 Reality Audio aus statischen Objekten. Hier gilt zu beachten, dass bei der Verwendung von Kanal-Betten Klangquellen beim Hörer an Positionen erklingen können, die in der Mischung nicht beabsichtigt wurden. Beispielsweise durch Verwendung eines 5.1.4 Kanal-Bettes in der Mischung und Wiedergabe auf einem 5.0 Lautsprecher-Setup des Konsumenten – da es sich um Kanäle handelt, werden diese nicht wie Objekte flexibel gerendert, sondern mit einem Downmix auf das entsprechende Wiedergabe-Layout angepasst, wodurch sich Positionen von Klangquellen verschieben können. Bei statischen Objekten werden diese Klangquellen nach wie vor als Objekt übertragen und gerendert, weshalb sich hier die beabsichtigte Position des Bettes beim Hörer an das entsprechende Wiedergabesystem anpasst.

3.2 Dolby Atmos Music

„Dolby Atmos Music is an immersive music experience that adds more space, clarity, and depth to your music. Instead of just hearing your music – with Dolby, it feels like you're inside the song."
(Dolby Laboratories, 2020e)

3.2.1 Kurzbeschreibung

Der Begriff „Dolby Atmos" ist ein Überbegriff für das immersive Klangerlebnis von Dolby, welches eine Produktion und Wiedergabe von Musik und audiovisuellen Inhalten ermöglicht. Diese können über verschiedene Codecs übertragen werden (Dolby Digital Plus, Dolby TrueHD, AC-4). Es wurde im Jahr 2012 als Wiedergabesystem in Kinos eingeführt, über die Jahre auf den Heimkino-Bereich ausgeweitet und ist seit 2017 auch über Streaming-Dienste wie Netflix abrufbar.

Mit der Einführung von Dolby Atmos Music Ende 2019 erfolgte schließlich der Eintritt in den Musik-Streaming-Markt (Kjörling et al., 2016; Dolby Laboratories, 2017b; Thomas, 2020d). Trotz des Aufkommens von Musik-Streaming-Diensten wie Spotify, Tidal, Deezer oder Amazon Music, welche den Zugang zu Musik maßgeblich verändert haben, ist die Art und Weise der Musikproduktion zumeist noch klassisch kanalbasiert. DAM hingegen ist eine objektbasierte Technologie zur Produktion und Wiedergabe immersiver Audioinhalte, basierend auf den AC-4 und Dolby Digital Plus (DD+) Codecs.

Im Musik-Streaming-Bereich kann Dolby Atmos Music über Tidal HiFi (in Kombination mit Dolby Atmos fähigen Geräten über Kopfhörer und Lautsprecher) oder Amazon Music HD (über den Amazon Echo Studio Lautsprecher) wiedergegeben werden, Dolby arbeitet hierfür mit verschiedenen Musikunternehmen zusammen. Infolge dieser Kooperationen werden sowohl bereits in Stereo bzw. 5.1 Surround publizierte Alben und Titel neu im Dolby Atmos Music Format veröffentlicht, als auch neu dafür produziert. Vereinzelt finden sich auch DAM Titel auf Blu-ray (Cohen, 2020b).

Das von Dolby für die Mischung von Dolby Atmos Music Inhalten empfohlene Lautsprecher-Layout (dargestellt in Abb. 3.1) ist 7.1.4[1], wobei die Lautsprecher-Konfiguration in diesem Falle nicht dem in *ISO/IEC 23001-8* definierten CICP-19-Layout entspricht, sondern leicht abgeändert wurde. Es gilt die Reihenfolge L, C, R, LFE, Lss, Rss, Lrs, Rrs, Ltf, Rtf, Ltr, Rtr (Dolby Laboratories, 2020d).

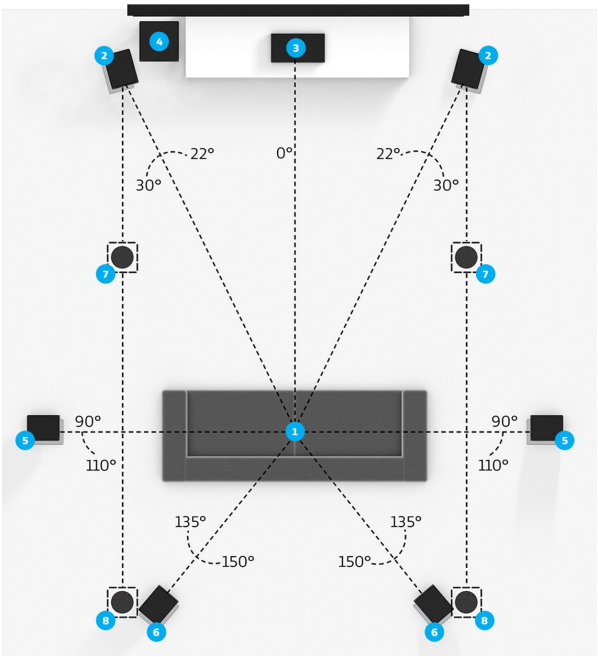


Abbildung 3.1: Zur Dolby Atmos Music Mischung empfohlenes 7.1.4 Lautsprecher-Layout (Dolby Laboratories, 2020a).
1 = Sitzposition, 2 = L und R, 3 = C, 4 = LFE,
5 = Lss und Rss, 6 = Lrs und Rrs, 7 = Ltf und Rtf, 8 = Ltr und Rtr

Die aktuelle Produktions- und Distributionskette für Dolby Atmos Music ist in Abb. 3.2 dargestellt. Die Erstellung der Grafik erfolgte neben einer Literaturanalyse aus Mangel an öffentlich verfügbaren Informationen auch mittels Recherche und persönlichem Mailkontakt mit Dolby-Mitarbeiter Ceri Thomas (siehe Anhang A.1 und Anhang A.2). Darauf basierend wurde ein Blockdiagramm erstellt, welches den gesamten Prozess darstellt (Dolby Laboratories, 2018; Nvidia, 2019; Grüner, 2019; Amazon, 2019; Dolby Laboratories, 2020b, 2020d, 2020g; Thomas, 2020a, 2020b, 2020c; Tidal, 2020b; Cohen, 2020a, 2020b; Roberts, 2020; Apple, 2020; Serck, 2020; Kaczmarek, 2020).

[1]Für Lautsprecher-Layouts mit Höhenebene existieren zwei Schreibweisen: 7.1+4 und 7.1.4. Im Folgenden wird Letztere gewählt.

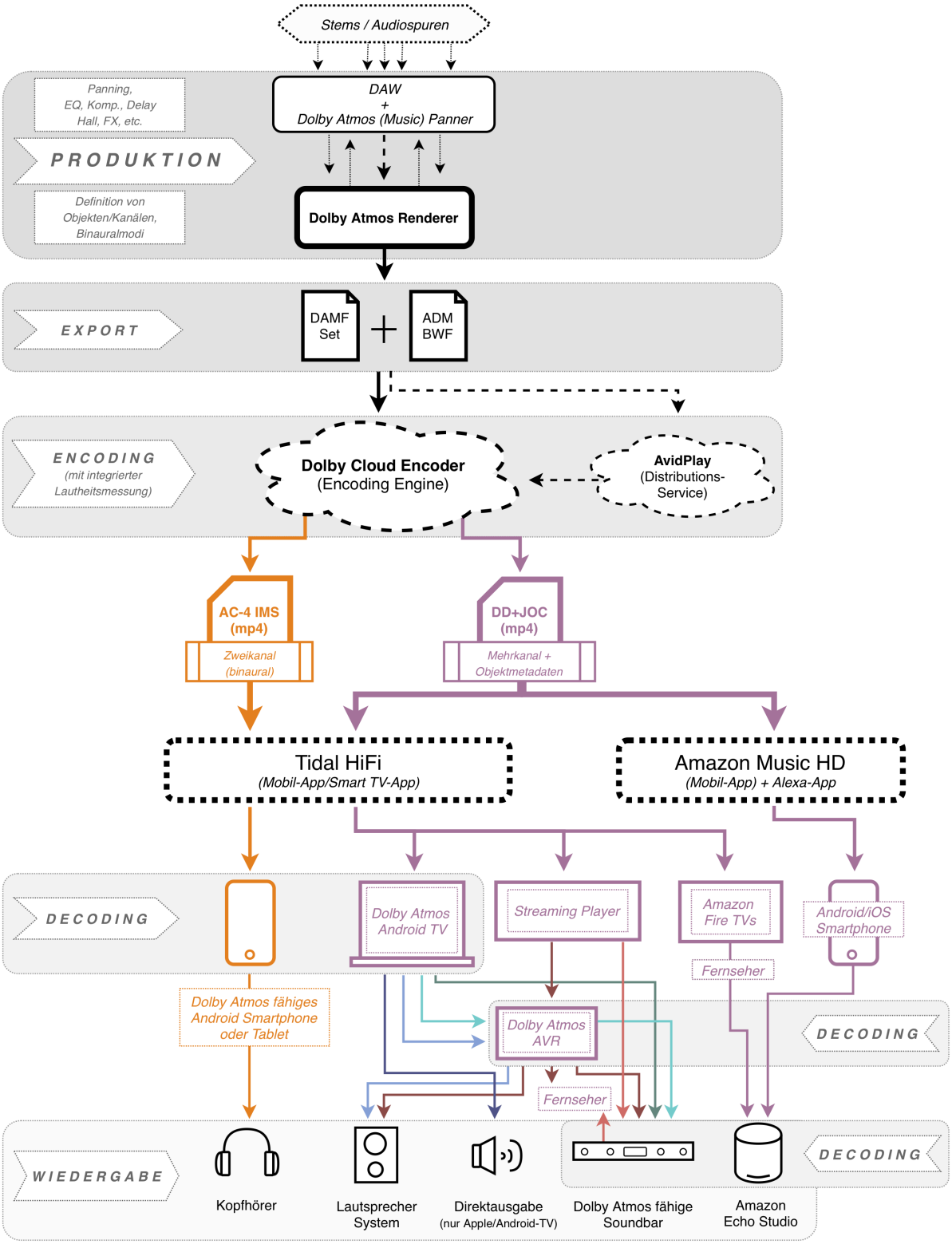


Abbildung 3.2: Aktuelle Produktions- und Distributionskette von Dolby Atmos Music. Stand: 08/2020

3.2.2 Produktion: Dolby Atmos Software

Für die Erstellung einer Dolby Atmos-Mischung ist sowohl der Dolby Atmos Renderer notwendig, der in der Dolby Atmos Production Suite und der Dolby Atmos Mastering Suite enthalten ist, sowie das DAM Panner-Plug-in oder die nativ in Pro Tools[2] und Nuendo[3] integrierten Dolby Atmos Panner.

Dolby Atmos Music Panner

Das *Dolby Atmos Music Panner-Plugin* (AAX-, AU- und VST3-Format) dient zur dreidimensionalen Positionierung von Audioobjekten einer DAM Mischung in einer unterstützten DAW. Zusätzlich ermöglicht der Panner die Synchronisierung der Objektpositionen mit dem DAW-Tempo. Der auf Objektspuren insertierte Panner liefert dem Renderer beispielsweise Positionierungs- und weitere Metadaten, zur Verwendung ist die Dolby Atmos Production oder Mastering Suite erforderlich (Dolby Laboratories, 2020d; Kenny, 2020).

Dolby Atmos Production/Mastering Suite

Die *Dolby Atmos Production Suite* enthält die Version des Renderers, welche auf dem selben Produktions-Rechner wie die DAW läuft. Hierbei gilt anzumerken, dass ein MacOS-Betriebssystem notwendig ist. Die *Dolby Atmos Mastering Suite* bietet zusätzlich zu den Merkmalen der Dolby Atmos Production Suite noch einige weitere Funktionen, wobei die Verwendung des Renderers in diesem Fall auf einem dedizierten Mac- oder Windows-Rechner (auch *Rendering- und Mastering-Workstation* genannt) vorgesehen ist, während die DAW auf einem separaten Rechner läuft. Außerdem ist zusätzlich noch eine Dolby Atmos Renderer Remote-Applikation enthalten, welche den Renderer auf der Rendering- und Mastering-Workstation über *Dante* oder *MADI* von einem Rechner im selben Netzwerk aus steuert (Dolby Laboratories, 2020d).

Dolby Atmos Renderer

Der *Dolby Atmos Renderer* ist Teil der *Dolby Atmos Production/Mastering Suite*. Es handelt sich hierbei um eine Software, die für das Rendering von Audio und Dolby

[2]Software „Avid Pro Tools“, im Folgenden als „Pro Tools“ verwendet
[3]Software „Steinberg Nuendo“, im Folgenden als „Nuendo“ verwendet

Atmos Metadaten aus einer DAW nötig ist und ermöglicht, eine Dolby Atmos Mischung zu monitoren. Der Renderer ab Version v3.4 verfügt über eine Realtime- und Offline Lautheitsmessung (Dolby Laboratories, 2018).

Audiospuren können aus der DAW *Pro Tools* mittels des *Dolby Renderer Send Plugins* an den Renderer gesendet werden. Hierbei muss jede Kanalbett- und Objektspur einen eigenen Renderer-Inputkanal verwenden. Das Kanalbett besteht in einem Standard Dolby Atmos Setup aus einer 7.1.2 Mehrkanal Spur, Objekt-Spuren werden an ein Renderer-Objekt geschickt. Mit dem *Dolby Renderer Return Plugin* wird Audio vom Renderer empfangen und die gerenderte Mischung an Pro Tools Output-Kanäle gesendet. Auch hier gilt, dass jeder Renderer-Output eine entsprechende Pro Tools Input-Spur benötigt (Dolby Laboratories, 2018).

Für die Verwendung des Renderers mit anderen DAWs ist die *Dolby Audio Bridge* notwendig. Diese sendet jede Audiospur der DAW (Kanalbett und Objekte) an die entsprechenden Renderer-Inputs. Sowohl die DAW als auch der Renderer müssen die Dolby Audio Bridge verwenden. Die Signale der Renderer-Outputs werden schließlich an ein Audio Interface geschickt (Dolby Laboratories, 2018). Im Renderer werden bis zu 128 Input-Kanäle unterstützt, die entweder als Kanalbett oder Objekte definiert werden können. Diese Input-Konfiguration ist vollständig unabhängig vom Wiedergabesystem – der Renderer fungiert hierbei als Verbindung beider Umgebungen. Der Dolby Atmos Renderer enthält auch einen binauralen Echtzeit-Renderer, mit dem binaurale Mischungen über Kopfhörer erstellt und abgehört werden können (Dolby Laboratories, 2018, 2020d).

Binaural-Einstellungen

Der Dolby Atmos Renderer unterstützt Binaural-Rendering, welches für die Verwendung zur Codierung von Inhalten als Dolby AC-4 Immersive Stereo (AC-4 IMS) ausgelegt ist. Die binauralen Einstellungen werden ebenfalls im Renderer aufgenommen und in das Dolby Atmos Master File Set (DAMF-Set) gespeichert (Dolby Laboratories, 2018). Um während der immersiven Mischung eine ausgeglichene Wiedergabe auf Lautsprechern und Kopfhörern zu erreichen, wird empfohlen, ein Kopfhörer-Monitoring zu integrieren. Speziell für die binaurale Mischung bietet der Renderer verschiedene Binaural-Modi: *Off, Near, Mid, Far*. Diese verschiedenen Einstellungen dienen dazu, verschiedene Arten und Grade von Räumlichkeit auf die entsprechenden Objekte bzw. Bett-Kanäle anzuwenden.

Near, Mid, Far beschreibt hierbei den Abstand zwischen Klangquelle und Kopfposition des Hörers, was mittels Metadaten gespeichert wird. Die Einstellungen der Binauralmodi werden während der Mischung beim Abhören, Aufnehmen oder Abspielen auf das Kopfhörersignal angewandt, bei der Lautsprecherwiedergabe spielen sie keine Rolle. Es gilt außerdem zu beachten, dass bei Verwendung von mehreren Betten jeder Bett-Kanal nur einen binauralen Rendermodus haben kann – bei Verwendung von drei Betten hätte somit jeder L-Kanal die selbe binaurale Einstellung (Dolby Laboratories, 2018, 2020d; Genewick, 2020).

Laut Dolby haben bisherige Erfahrungen gezeigt, dass es den immersiven Produktions-Workflow für die Wiedergabe auf unterschiedlichen Systemen vereinfachen kann, wenn zuerst für Kopfhörer gemischt und schließlich beim Abhören auf Lautsprechern Anpassungen vorgenommen werden. Eine weitere Empfehlung ist, Objekte nicht exakt in der Mitte des Raumes zu platzieren, da es an dieser Position keine Binauralisierung am Kopfhörerausgang gibt, und die Einstellungen des binauralen Rendermodus somit keine Wirkung haben. Um eine trockene Center-Abbildung zu erhalten, wird empfohlen, das Objekt mittig an der Vorderwand zu positionieren und den binauralen Rendermodus auf *Off* oder *Near* einzustellen (Dolby Laboratories, 2020d).

Bett und Objekte

Dolby Atmos Inhalte bestehen aus Bett und Objekten, sowie zugehörigen Metadaten. Ein Bett ist ein kanalbasierter Premix bzw. Stems mit bereits enthaltenem Mehrkanal-Panning. Das Bett benötigt kein dediziertes Panning über Dolby Atmos-Metadaten – man kann sich ein Bett als traditionell kanalbasierte Konfiguration (wie 2.0, 5.1 und 7.1) vorstellen. Dabei handelt es sich um feste Positionen im Raum, die für traditionelle Lautsprecher-Layouts (einschließlich Lautsprecher-Arrays) verwendet werden.

Ein Objekt ist ein Mono- oder Stereo-Audioinhalt, der über ein dediziertes Dolby Atmos-Panning verfügt und an einer beliebigen Stelle im dreidimensionalen Klangfeld platziert werden kann. Ein Audioobjekt kann – je nach Definition durch Positions- und Größen-Metadaten – eine unterschiedliche Anzahl an Lautsprechern zur Wiedergabe verwenden. Objekte können statisch oder dynamisch (beweglich) sein.

Der Dolby Atmos Renderer unterstützt bei 48kHz insgesamt 128 Input-Kanäle (mindestens zehn Bettkanäle und bis zu 118 Objekte). Wie viele Objekte in der Mischung letztendlich verwendet werden, hängt von klanglichen und kreativen Entscheidungen ab. Für die Überlegungen zur Verwendung der beiden Elemente empfiehlt Dolby, der Frage nachzugehen, ob ein Klangelement besser als Bett oder Objekt funktionieren kann. Ein Aspekt, der hierbei bedacht werden sollte, ist, dass Objekte standardmäßig keinen Zugriff auf den LFE-Kanal haben. Von Dolby wird die Empfehlung ausgesprochen, sich nicht auf den LFE zu verlassen, sondern die tieffrequenten Signale auch in der Hauptmischung einzubauen und den LFE als klangliche Ergänzung zu sehen (Dolby Laboratories, 2018, 2020d).

Je nach Positions-Metadaten eines Objektes können Objekte und Bett-Kanäle klanglich identisch sein. Ein Objekt, das beispielsweise vorne links platziert wird und dessen Größe auf Null gesetzt ist, ist klanglich identisch zum linken Teil des Kanalbettes. Die Übertragung des Bettes im Renderer wird durch die Anzahl der Lautsprecher im Raum definiert. Wenn z.B. ein Raum ohne Center-Lautsprecher für die Wiedergabe konfiguriert ist, wird bei Bettkonfigurationen ab 3.0 versucht, den Inhalt im Center-Kanal des Bettes auf die verfügbaren Positionen links und rechts als Phantomschallquellen abzubilden. Ebenso im Upper-Layer: Elevierte Lautsprecher in einer x.y.4-Konfiguration erzeugen ein Phantom-Center (vorn/hinten) der x.y.2-Komponente, während in einer x.y.6-Upper-Konfiguration die Punktschallquelle verwendet wird. Die Verwendung eines Objektes ermöglicht hier erweiterte Wiedergabekontrolle. Da die Lautsprecherkonfigurationen der Konsumenten sehr unterschiedlich sind, kann eine Verwendung von Klangquellen als Teil eines Bettes (insbesondere an den Seiten und der Höhenebene) dazu führen, dass diese an Positionen hörbar sind, wo sie in der Mischung nicht platziert wurden (Kenny, 2020; Genewick, 2020, S. 25). In Bezug auf die Verwendung des Center-Kanals in der Musikproduktion empfiehlt Dolby, die Stimme nicht zu sehr zu extrahieren und bei der Mischung zu beachten, dass viele Wiedergabesysteme (wie traditionelle Stereo-Systeme) keinen Center-Lautsprecher haben (Dolby Laboratories, 2020d).

Gestalterische Möglichkeiten

Durch die Integration des DAM-Panners in die DAW sowie der Ansteuerung des Renderers aus der DAW wird ermöglicht, dass auch die objektbasierte Produktion innerhalb der DAW-Umgebung erfolgen kann. Grundlage hierfür kann somit sowohl das originale Stereo- bzw. 5.1 Surround-DAW-Projekt sein, in das die Dolby Atmos Tools integriert

werden, als auch ein neues Projekt, in welches die zuvor exportierten Stereo- bzw. 5.1 Surround-Stems und Audiospuren importiert oder neue Spuren aufgenommen und die Dolby Atmos Tools eingebaut werden. Wie in Abb. 3.2 dargestellt, können klangliche Anpassungen (wie Hall, EQ, Effekte) sowie die objektbasierte Mischung (wie 3D-Panning, Definition von Bett und Objekten, Binauraleinstellungen) somit parallel erfolgen.

Qualitätskontrolle

Um Fehler in Mischung, Export oder Wiedergabe zu vermeiden – und um den fertigen Titel so zu hören, wie ihn auch der Endkonsument hören wird – ist die Qualitätskontrolle am Ende des Produktions-Workflows ein wichtiger Aspekt. Hierzu kann aus dem Dolby Atmos Renderer eine encodierte .mp4-Datei exportiert werden, welche einen e-AC-3-Bitstream (DD+JOC) enthält und sich zur Qualitätskontrolle der erstellten Mischung bzw. zur Wiedergabe über Dolby Atmos fähige AVRs, BluRay-Player, Fernsehgeräte oder Soundbars eignet (Dolby Laboratories, 2020f).

3.2.3 Exportformate: DAMF und ADM BWF

Dolby schreibt in der Richtlinie zu den *Delivery Specifications*, dass die Mischung in einem Raum mit mindestens einem 7.1.4 Lautsprecher-Layout abgehört werden sollte. Die Lautheit für Streaming muss –18 LUFS betragen, für Blu-ray –31 LUFS (Dolby Laboratories, 2020c). Laut Dolby muss die Dolby Atmos Music Auslieferung zur Distribution über Streaming-Dienste eine BWF/ADM-Datei sein, welche aus dem Renderer exportiert wird (Dolby Laboratories, 2020c). Hierzu wird nach Fertigstellung der Produktion zuerst ein Dolby Atmos Master File Set erstellt. Hierbei wird im Renderer ein Dolby Atmos Master aufgenommen, wobei Metadaten in das Dolby Atmos Master File Set gespeichert werden. Weitere Details sind Tabelle 3.1 zu entnehmen.

Ein *.atmos*-Master kann im Dolby Atmos Renderer schließlich als BWF/ADM *.wav*-Master exportiert werden. Dolby Atmos Master Dateien können somit im ADM mitgeführt werden, jedoch sind dafür nicht alle ADM-Elemente erforderlich. Aus diesem Grund wurde hierfür das ADM Atmos-Profil entwickelt. Details hierzu sind in (Dolby Laboratories, 2019a) zu finden. Zusätzlich zur Übertragung von Objekt-Metadaten über den ADM-„chunk“ enthalten DAMF-Sets den Dolby Audio Metadaten-„chunk“ in der umschließenden BWF-Datei.

Dateiendung	enthaltene Informationen
.atmos	Auskunft über Präsentation im Master Dateiset
.atmos.audio	interleaved PCM-Datei mit allen Audiosignalen
.atmos.dbmd	zusätzliche Parameter, z.B. für Dolby Digital Plus
.atmos.metadata	3D-Positionskoordinaten für statische und dynamische Signale der .audio-Datei

Tabelle 3.1: Im Dolby Atmos Master File (DAMF) Set enthaltene Informationen (Dolby Laboratories, 2018, S. 263).

Der Begriff „Dolby Audio“ wird hierbei für Formate wie AC-4 oder Dolby Digital Plus verwendet. Der Dolby Audio Metadaten-„chunk“ enthält zusätzliche Metadaten-Elemente mit relevanten Informationen über die Werkzeuge, mit denen die Audioinhalte erstellt wurden, das Programm und die Encoder-Konfiguration für die nachfolgende Verarbeitung (Dolby Laboratories, 2019a). Es ist außerdem möglich, im Renderer mehrere kanalbasierte Wiedergabeformate anzugeben um die Mischung so beispielsweise auch in Stereo zu exportieren. Wenn aus dem Dolby Atmos Master-Format kanalbasierte Mischungen abgeleitet werden sollen, empfiehlt Dolby, während des Misch-Prozesses auch in diesen Konfigurationen abzuhören (Dolby Laboratories, 2020d).

3.2.4 Encodierung: AC-4 IMS und DD+JOC

Dolby Atmos Music basiert sowohl auf dem neuen AC-4 Immersive Stereo (AC-4 IMS) Codec (bei der Kopfhörer-Wiedergabe über Musik-Streaming-Dienste wie Tidal) als auch auf Dolby Digital Plus Joint Object Coding (DD+JOC) (bei der Lautsprecher-Wiedergabe via Tidal oder Amazon Music HD). AC-4 IMS wurde laut Dolby speziell für die Distribution von DAM für mobile Endgeräte optimiert (Thomas, 2020d; Dolby Laboratories, 2018). Bei der AC-4 IMS encodierten Datei handelt es sich um eine Zweikanal-Datei, wobei die Binauralisierung bereits während des Encodiervorgangs erfolgt (C.Thomas, persönliche Kommunikation, 17.07.2020, siehe Anhang A.1). Im folgenden Distributionsprozess werden letztendlich zwei Dateien übertragen: AC-4 IMS (binaural mit zwei Kanälen) und DD+JOC (Mehrkanal mit Objektmetadaten). Aus Mangel an öffentlich einsehbaren Informationen ist es zum jetzigen Zeitpunkt jedoch nicht ersichtlich, ob bei der Übertragung von AC-4 IMS nicht doch Metadaten irgendeiner Form

mitgeliefert werden. Im Folgenden wird daher davon ausgegangen, dass es sich um eine binaurale Zweikanal-Datei handelt. Weitere Informationen zu den Distributionsformaten sind in Abb. 3.3 dargestellt (Thomas, 2020a).

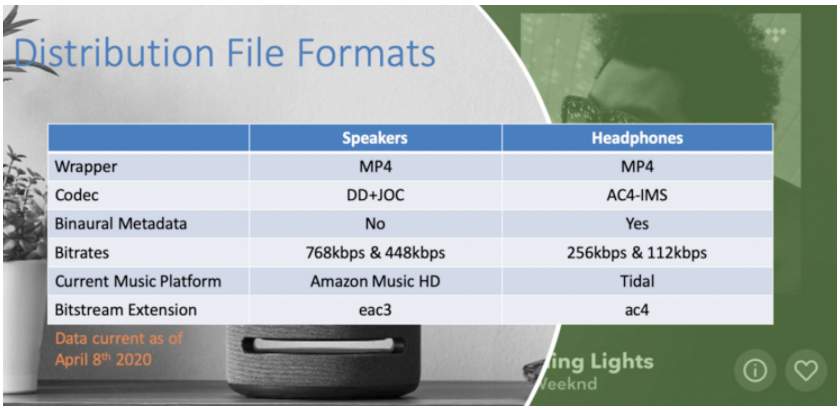


Abbildung 3.3: Dolby Atmos Music Distributionsformate. Die geringere der beiden Bitraten bei AC-4 IMS und DD+JOC wird jeweils für den Fall einer niedrigen Bandbreite mitübertragen (Thomas, 2020a, 2:40 min), siehe Anhang A.2.

AC-4

AC-4 ist ein Next Generation Audio Codec der Dolby Laboratories Inc., standardisiert in *ETSI TS 103 190* und *ETSI TS 103 190-2* und gilt als Nachfolger von Dolby Digital Plus und Dolby Digital. AC-4 wurde insbesondere zur Anwendung in aktuellen und zukünftigen Multimedia Unterhaltungsdiensten wie beispielsweise Streaming entwickelt und unterstützt NGA Merkmale wie immersives und personalisiertes Audio, erweiterte Lautheits- und DRC Steuerung für verschiedene Gerätetypen und Anwendungen, Dialogverbesserung, Programm-IDs, flexibles Wiedergabe-Rendering etc. Der AC-4 Codec erlaubt die Übertragung von kanalbasierten und objektbasierten Audioinhalten sowie eine Kombination dieser mit hoher Audioqualität bei geringen Bitraten. AC-4 bietet eine durchschnittlich 50% höhere Komprimierungseffizienz als Dolby Digital Plus (Dolby Laboratories, 2015; Kjörling et al., 2016) und ist nicht rückwärtskompatibel zu Dolby Digital (AC-3) oder Dolby Digital Plus (E-AC-3).

Genaue Bitraten verschiedener Qualitätsstufen bei gängigen Kanalformaten sind in Tabelle 3.2 dargestellt. Hierbei gilt anzumerken, dass die Bezeichnungen der MUSHRA-Skala nach *ITU-R BS.1534-2* (ITU-R, 2014) entsprechen.

Format	gut	exzellent
Mono	24 kbp/s	40 kbp/s
Stereo	40 kbp/s	64 kbp/s
5.1	96 kbp/s	160 kbp/s
7.1.4	224 kbp/s	288 kbp/s

Tabelle 3.2: AC-4 Bitraten zweier Qualitätsstufen (Dolby Laboratories, 2015, S. 8).

AC-4 ermöglicht, dass innerhalb eines Audio Bitstreams verschiedene Substreams übertragen werden. Diese Substreams können beispielsweise verschiedene Kommentare (Mono), ein 5.1 Bett oder ein 7.1.4 Bett sein. Der Audio Renderer kann hierbei verschiedene Substreams kombinieren (z.B. 5.1 Bett + englischer Kommentar oder 7.1.4 Bett + spanischer Kommentar), welche in den sogenannten „Presentations“ definiert werden. Welche dieser „Presentations“ letztendlich wiedergegeben wird, wird im Decoder abhängig des entsprechenden Wiedergabesystems entschieden (Dolby Laboratories, 2015).

Lautheit

Um die Vielzahl an Wiedergabegeräten und -systemen bedienen zu können, ist in AC-4 ein flexibles Dynamic Range Control (DRC) und Lautheits-Management implementiert, wobei vier DRC Decoder Betriebsarten standardmäßig definiert sind: Heimkino, Flachbildfernseher, tragbare Lautsprecher und Kopfhörer. Das Lautheitsmanagement in AC-4 umfasst eine adaptive Echtzeit-Lautheitsverarbeitung (Kjörling et al., 2016). Der Dolby AC-4-Encoder verfügt über ein integriertes Lautheitsmanagement. Der Encoder ermittelt die Lautheit des eingehenden Audiosignals und kann die Lautheits-Metadaten auf den korrekten Wert anpassen oder mittels einer Multibandverarbeitung das Programm auf den Ziel-Lautheitspegel bringen. Anstatt das Audio im Encoder zu verarbeiten, wird diese Information dem Bitstream in den DRC-Metadaten hinzugefügt, so dass die Verarbeitung im Endgerät entsprechend dem Wiedergabeszenario erfolgen kann. Der Prozess ist daher nicht-destruktiv; das Original-Audio wird im Bitstream mitgeführt und steht für zukünftige Anwendungen zur Verfügung. Um eine mehrfache Lautheitsverarbeitung in der Produktions- und Wiedergabekette zu vermeiden, verwendet Dolby AC-4 das erweiterte Metadaten-Framework, das in *ETSI 102 366 Anhang H* standardisiert ist. Dieses Framework enthält Informationen über die bisherige Lautheitsverarbeitung der Inhalte (Dolby Laboratories, 2015).

Dolby Digital Plus (DD+)

Dolby Digital Plus (DD+) stellt die Erweiterung von Dolby Digital dar und basiert auf dem bestehenden Mehrkanal Standard AC-3. DD+ sollte diesen um ein größeres Spektrum an Datenraten (> 640 kbp/s) und Kanalformaten (mehr Kanäle als das von AC-3 unterstütze Maximum 5.1) erweitern. Aus diesem Grund wurde ein flexibles, AC-3 kompatibles Codiersystem entwickelt, welches als Dolby Digital Plus bzw. *enhanced AC-3 (e-AC-3)* bekannt ist. Dolby Digital Plus ist abwärtskompatibel zu Dolby Digital und kann – zusätzlich zu erweiterten Kanalformaten mit bis zu 14 diskreten Kanälen – herkömmliche Mono, Stereo- oder 5.1 Surround-Formate mit etwa halber Datenrate im Vergleich zu Dolby Digital übertragen (Fielder et al., 2004).

Joint Object Coding (JOC)

Joint Object Coding (JOC) ermöglicht es, objektbasierte und immersive Audioinhalte wie Dolby Atmos mit geringen Bitraten zu übertragen, und somit auch Produktionen mit einer Vielzahl an Objekten auf die entsprechenden Wiedergabesysteme der Konsumenten zu rendern. Hierbei werden nah beieinander liegende Objekte in räumlichen Objektgruppen zusammengefasst, dargestellt in Abb. 3.4. Die orangenen und blauen Punkte der linken Darstellung entsprechen 19 originalen Objekten, die roten Kreise der rechten Darstellung 11 räumlichen Objektgruppen. Objektgruppen können aus Objekten oder einer Kombination aus originalen Objekten und Bett-Kanälen bestehen, außerdem können Objekte auch in mehreren Objektgruppen vertreten sein (Purnhagen, Hirvonen, Villemoes, Samuelsson & Klejsa, 2016; Kjörling et al., 2016).

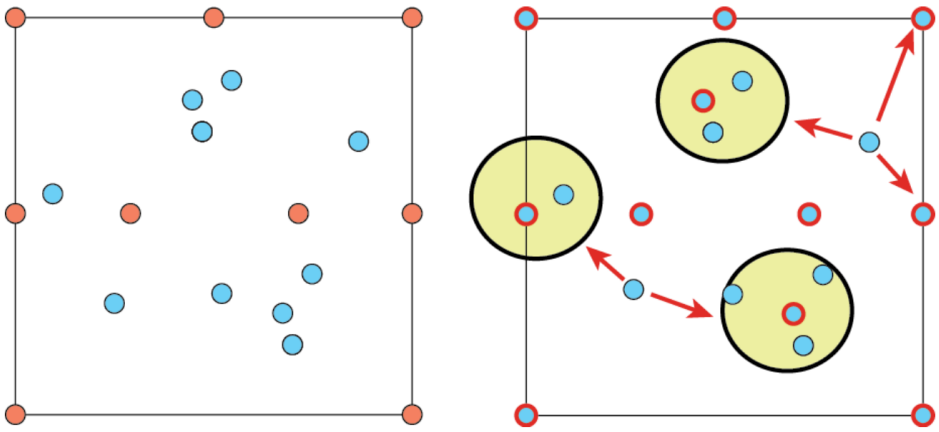


Abbildung 3.4: Objektbasiertes Audio vor (links) und nach (rechts) der räumlichen Joint Object Codierung, (Dolby Laboratories, 2018, S. 164–265).

Weiterführend wird ein Mehrkanal-Downmix der Audioinhalte zusammen mit parametrischen Informationen übertragen, welche letztendlich im Decoder die Rekonstruktion der Audioobjekte aus dem Downmix ermöglichen. Die parametrischen Informationen umfassen sowohl JOC-Parameter als auch Objektmetadaten. Die JOC-Parameter übertragen in erster Linie die zeit- und frequenzvariablen Elemente einer Upmix-Matrix (Purnhagen et al., 2016; Kjörling et al., 2016).

Joint Object Coding wird derzeit in zwei verschiedenen Anwendungen implementiert. Diese unterscheiden sich sowohl in der verwendeten Downmix-Codierung als auch in der Konfiguration des JOC-Systems selbst. Eine erste Version von JOC wurde entwickelt, um eine rückwärtskompatible Erweiterung des Dolby Digital Plus-Systems bereitzustellen und die Übertragung von immersiven Dolby Atmos-Inhalten mit Bitraten wie 384 kbp/s zu ermöglichen, die häufig in bestehenden Rundfunk- oder Streaming-Anwendungen verwendet werden. Um eine direkte Rückwärtskompatibilität mit vorhandenen Dolby Digital Plus-Decodern zu gewährleisten, wird im DD+ JOC-System ein 5.1 (oder 7.1) Downmix verwendet. Die Decodierung für DD+ JOC ist z.B. bereits in AVRs verfügbar, die Dolby Atmos unterstützen. Ein erweitertes Advanced Joint Object Coding (A-JOC) Tool, einschließlich adaptiver Downmix-Optionen und Dekorrelationseinheiten wurde im AC-4 Decoder implementiert. Darüber hinaus unterstützt das AC-4 A-JOC-System auch alle anderen Merkmale des AC-4-Codecs (Kjörling et al., 2016; Purnhagen et al., 2016).

Möglichkeiten zur Encodierung

Die räumliche Codierung erfolgt an zwei verschiedenen Punkten des Dolby Atmos Authoring-Workflows: Beim Abhören der Mischung in Echtzeit mittels des Renderers und bei der Encodierung des Dolby Atmos Master File Sets. Das DAMF-Set enthält bis zu 128 Audiosignale und wurde noch nicht räumlich prozessiert – dies geschieht beim Encodierungsprozess durch eine Software (*Dolby Media Encoder*). Diese Software liest die *.atmos*-Datei ein, wendet räumliche Codierung durch Erstellung von Objektgruppen an und encodiert diese schließlich in das Ausgabeformat. Die räumliche Dolby Atmos Codierung wurde insbesondere zur Reduzierung der Bitraten und Komplexität für die Bereitstellung von Inhalten auf mobilen Endgeräten entwickelt. Dadurch soll auch die Übertragung durch Online-Streaming verbessert werden, indem die benötigte Rechenleistung der Wiedergabegeräte durch eine reduzierte Objektanzahl verringert wird (Dolby Laboratories, 2018).

Es gilt jedoch zu beachten, dass der Dolby Media Encoder nur folgende Formate unterstützt: Dolby TrueHD (mit und ohne Dolby Atmos), Dolby Digital Plus (mit und ohne Dolby Atmos), Dolby Digital und MLP Lossless (Dolby Laboratories, 2017a, 2019b). Hieraus lässt sich schließen, dass der Dolby Media Encoder nicht zur Codierung von Dolby Atmos Music verwendet werden kann, da diese für Musik-Streaming auf AC-4 basiert, was der Media Encoder nicht unterstützt. Aus der *Dolby Atmos Music Webinar-Serie* geht hervor, dass die Encodierung für DAM mittels des sog. *Cloud Encoders* passiert, welcher auf der Dolby Atmos Encoding Engine basiert und die zuvor exportierte BWF/ADM-Datei sowohl als AC-4 IMS als auch als DD+JOC Datei encodiert und in einen .mp4-Container verpackt. Diese Dateien werden dann für die weitere Distribution an die Musik-Streaming-Dienste geliefert (Thomas, 2020a). Laut Dolby steht dieser Cloud-Encoder aktuell nur Labels- und Streaming-Partnern zur Verfügung, seit Juli 2020 wurde die Möglichkeit zur Encodierung jedoch ausgeweitet (siehe Abschnitt 3.2.5) (Thomas, 2020b).

3.2.5 Distribution und Wiedergabe

Die gesamte Produktions- und Distributionskette für Dolby Atmos Music ist in Abb. 3.2 dargestellt. In Abb. 3.5 wird zur detaillierteren Darstellung der Decodierung und Wiedergabe ein Ausschnitt daraus abgebildet (Dolby Laboratories, 2018; Nvidia, 2019; Grüner, 2019; Amazon, 2019; Dolby Laboratories, 2020b, 2020d, 2020g; Thomas, 2020a, 2020b, 2020c; Tidal, 2020b; Cohen, 2020a, 2020b; Roberts, 2020; Apple, 2020; Serck, 2020; Kaczmarek, 2020).

Distributionsservice AvidPlay

Im Juli 2020 wurde bekanntgegeben, dass nun insbesondere für Independent-Labels und -Künstler eine Distribution von DAM über AvidPlay möglich ist. Hierbei werden sowohl Metadaten (Label, Künstler, Titel etc.) als auch die BWF/ADM-Datei hochgeladen, welche im Dolby Cloud Encoder in die unterschiedlichen Formate AC-4 IMS und DD+JOC encodiert und auf den entsprechenden Streaming-Diensten veröffentlich wird. Außerdem erfolgt die Distribution einer Stereo-Datei zur Wiedergabe auf nicht-Atmos-fähigen Geräten oder – sofern bereits eine Stereo-Version veröffentlicht wurde – die Verknüpfung mit eben dieser durch Verwendung des selben International Standard Recording Code (ISRC). Diese Codes werden mit der Musikproduktion als Metadaten abgespeichert und dienen zur Identifikation der Titel (Dolby Laboratories, 2020b; Thornton, 2020).

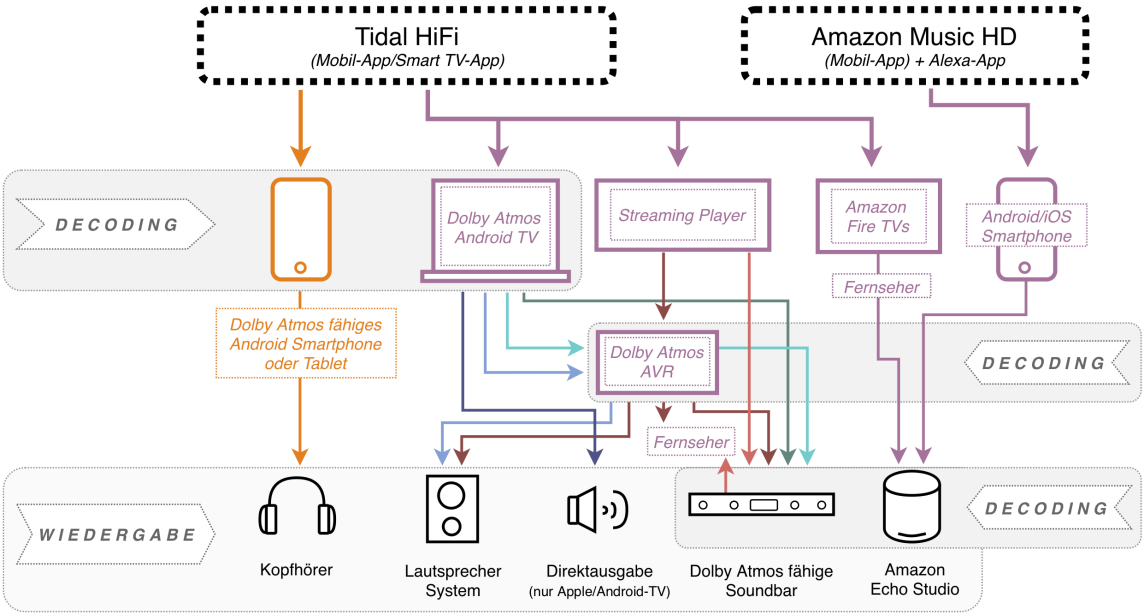


Abbildung 3.5: Decodierung und Wiedergabe von Dolby Atmos Music.
Stand: 08/2020

Amazon Music

Wie in Abb. 3.2 dargestellt, kann Dolby Atmos Music aktuell über die Streaming-Dienste Tidal (HiFi-Abonnement) und Amazon Music (HD-Abonnement) abgerufen werden. Aufgrund der Distribution zweier unterschiedlich encodierter Dateien (AC-4 IMS und DD+JOC) gilt zu beachten, dass via Amazon Music HD nur auf dem Amazon Echo Studio wiedergegeben werden kann. Ob Amazon auch die AC-4 IMS-Datei erhält, und somit eine Kopfhörerwiedergabe theoretisch implementieren könnte, ist nicht bekannt.

Der Amazon Echo Studio ist ein Smart-Speaker, in welchen drei 51-Millimeter-Mitteltöner, ein 25-Millimeter-Hochtöner und ein nach unten gerichteten 133-Millimeter-Woofer mit Bassöffnung integriert sind sowie ein 24-Bit-DAC. Der Echo Studio „unterstützt die Audioformate Proband FLAC, MP3, AAC, Opus, Vorbis, Dolby Digital, Dolby Digital Plus, Dolby Atmos, Sony 360 Reality Audio/MPEG-H sowie 16- und 24-Bit-Audiokodierung“ (Floemer, 2019). Wie bei einer 3D-Soundbar werden im Echo Studio eingebaute Mikrofone zur räumlichen Kalibrierung genutzt. Ferner besteht die Möglichkeit, ein oder zwei Echo Studio Geräte zusammen mit einem Fire TV Stick 4K oder Fire TV Cube drahtlos zu koppeln (Kenny, 2020).

Tidal

Der Musik-Streaming-Dienst Tidal bietet im Gegensatz zu Amazon Music etwas mehr Möglichkeiten, um Dolby Atmos Music Inhalte zu hören. Mittels der Mobil-App eines Dolby Atmos fähigen Android Smartphones oder Tablets kann über Kopfhörer gehört werden – da die Decodierung des AC-4 IMS-Streams in diesem Fall im Endgerät passiert, ist eine Dolby Atmos Unterstützung zwingend notwendig, weshalb momentan kein Streaming über iOS-Smartphones möglich ist. Wenn ein Gerät oder angeschlossenes Wiedergabesystem nicht Dolby Atmos fähig ist, wird dies von der Tidal-App erkannt und die Stereo-Version des Titels abgespielt. Aufgrund der zwei-Codec-Distributionskette ist es außerdem nicht möglich, DAM über Kopfhörer zu hören, die an Dolby Atmos fähige Heimkino-Geräte (AVR, TV-Geräte, Soundbar etc.) angeschlossen werden, da hierüber ausschließlich die DD+JOC-Datei übertragen wird (Tidal, 2020b). Seit Frühjahr 2020 ermöglicht es Tidal außerdem, DAM Titel auf Lautsprechern wiederzugeben. Dadurch, dass eine Vielzahl an Dolby Atmos fähigen Geräten für Endnutzer zur Verfügung stehen, ergeben sich hierdurch eine Reihe an Geräte-Kombinationen zur Wiedergabe, wobei die Verbindung über HDMI erfolgt und die Decodierung im letzten Dolby Atmos fähigen Gerät der Wiedergabekette passiert (siehe Anhang A.2).

Eine Option ist die Wiedergabe über einen Dolby Atmos fähigen Streaming-Player (Nvidia Shield TV (Pro), Fire TV (Cube, Stick 4K) oder Apple TV 4K) oder Dolby Atmos fähigen Android-Fernseher (Sony, Philipps). Diese beiden Geräte steuern jeweils die Tidal-App. Streaming-Player und Fernseher können zum einen direkt mit einer Dolby Atmos fähigen Soundbar und zum anderen mit einem Dolby Atmos AVR verbunden werden. Letzterer steuert zur Tonwiedergabe entweder ein Heimkino Lautsprecher-System oder eine Dolby Atmos fähige Soundbar an. Dolby Atmos Music Titel können außerdem auch nativ über die im Fernseher integrierten Lautsprecher wiedergegeben werden (Zehden, 2018; Nvidia, 2019; Grüner, 2019; Roberts, 2020; Kaczmarek, 2020; Tidal, 2020b; Cohen, 2020a, 2020b; Roberts, 2020; Serck, 2020; Apple, 2020). Auch Tidal ermöglicht die Wiedergabe über einen Amazon Echo Studio, wobei die Steuerung über einen an den Fernseher angeschlossenenen Amazon FireTV (Cube, Stick 4K) geschieht, welcher mit dem selben Netzwerk wie der Echo Studio verbunden sein muss (Amazon, 2019; FireTV-Blog, 2020, 2018).

Blu-ray Discs in Kombination mit einem Dolby Atmos AVR und/oder Soundbar können ebenfalls zur Wiedergabe von Dolby Atmos Music verwendet werden (Cohen, 2020b).

3.3 360 Reality Audio

„Immerse yourself in sound all around you. As real as if you are there at a live concert or with the artist recording in a studio. With 360 Reality Audio, music has never been so immersive and so real.“
(Sony Corporation, 2020a)

3.3.1 Kurzbeschreibung

360 Reality Audio ist – neben Dolby Atmos Music – eine weitere kommerzielle objektbasierte Technologie zur Wiedergabe immersiver Audioinhalte, welche ebenfalls Ende 2019 vorgestellt wurde. Anders als DAM wird bei 360 RA rein objektbasiert produziert und übertragen.

360 RA ist über die Musik-Streaming-Dienste Deezer HiFi (360 by Deezer Mobil-App), Tidal HiFi (Mobil-App), Nugs.net HiFi (Mobil-App) für die Kopfhörerwiedergabe verfügbar sowie über Amazon Music HD zur Wiedergabe auf dem Amazon Echo Studio Lautsprecher. Sony arbeitet hierfür mit verschiedenen Musikunternehmen zusammen. Infolge dieser Kooperationen werden sowohl bereits in Stereo bzw. 5.1 Surround veröffentlichte Alben und Titel im 360 RA Format herausgebracht, als auch neu dafür produziert (Sony Corporation, 2019a; Cohen, 2020b). 360 Reality Audio basiert auf dem MPEG-H Codec.

Das von Sony für die Mischung von 360 Reality Audio Inhalten empfohlene Lautsprecher-Layout (dargestellt in Abb. 3.6) ist 13ch(Music)[4]. Die Konfiguration besteht aus zwei 5.0-Setups (nach *ISO/IEC 23001-8* CICP-5 Layout), die übereinander angeordnet werden, sowie einem nach unten versetzten LCR-Setup (nach *ISO/IEC 23001-8* CICP-3 Layout). Die genauen Positionen der Lautsprecher werden von Sony als vertraulich behandelt und wurden aktuell nicht an die Öffentlichkeit kommuniziert (Fedak, 2020).

[4] „13ch(Music)“ ist die von Sony offiziell verwendete Bezeichnung für deren proprietäres Lautsprecher-Layout und wird im Folgenden verwendet.

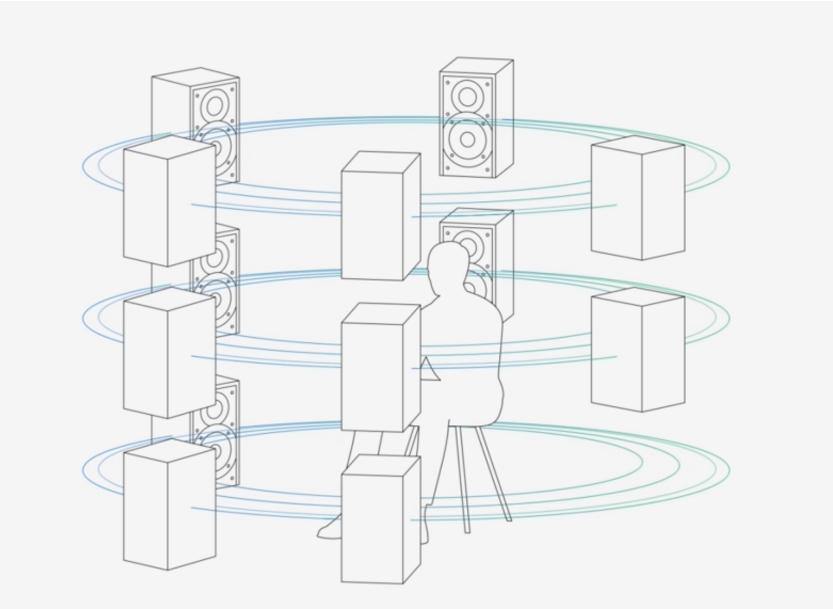


Abbildung 3.6: Zur 360 Reality Audio Mischung empfohlenes 13ch(Music) Lautsprecher-Layout (Sony Corporation, 2020a).

Die aktuelle Produktions- und Distributionskette für 360 Reality Audio ist in Abb. 3.7 dargestellt (Amazon, 2019; Sony Corporation, 2020a, 2020b; Tidal, 2020a).

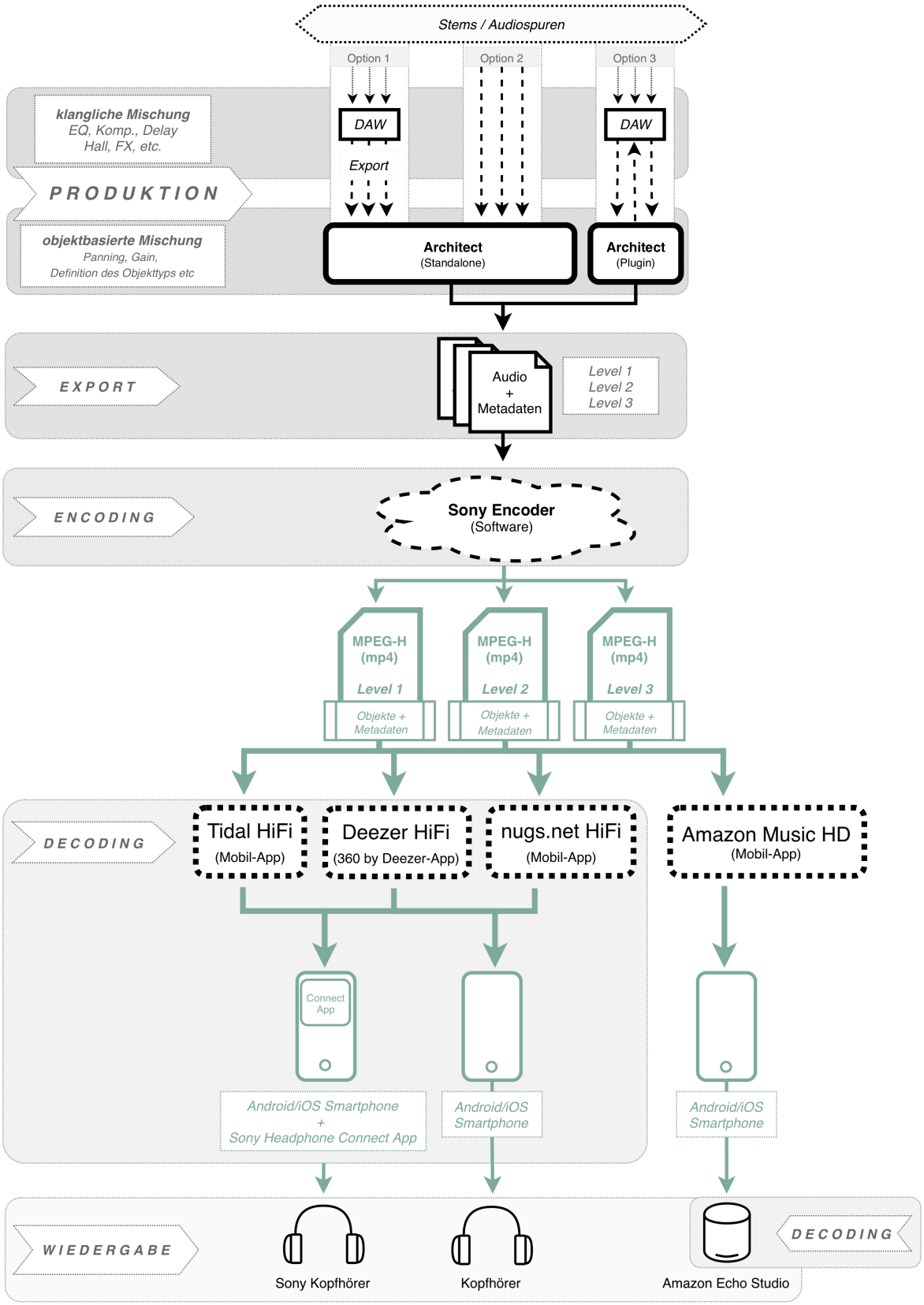


Abbildung 3.7: Aktuelle Produktions- und Distributionskette von 360 Reality Audio. Stand: 08/2020

3.3.2 Produktion: Architect Software

Zur Produktion von 360 RA Inhalten stellt Sony ausgewählten Partnern die firmeneigene Software *Architect* zur Verfügung. Die Software ist, anders als die Dolby Atmos Production Suite, derzeit nicht kommerziell zu erwerben sondern steht nur ausgewählten Partnern zur Verfügung und wird von Sony aktuell – wie nahezu alle Schritte des Workflows zwischen Anlieferung der Audiospuren und Wiedergabe über Musik-Streaming-Dienste – als vertraulich eingestuft. Es handelt sich um eine Standalone-Applikation bzw. ein AAX-Plugin zur Verwendung mit der DAW Pro Tools. Sonys Architect bietet – abgesehen von Pegelveränderungen – keine Möglichkeiten, den Klang zu bearbeiten, sondern dient rein zur räumlichen Positionierung der Audioobjekte. Bei Verwendung der Standalone-Version beginnt der objektbasierte Produktions-Workflow somit erst nach der klanglichen Mischung, wobei alle benötigten Elemente als Audiospuren exportiert werden müssen, um im Architect platziert werden zu können (Fedak, 2020).

Obwohl im Architect das Hinzufügen von Kanälen (darunter auch der LFE-Kanal) möglich ist, wird keines dieser beiden Elemente in der 360 Reality Audio-Distribution unterstützt. Der Grund hierfür ist die Entscheidung von Sony, ein MPEG-H-Profil zu verwenden, welches nur über einen eingeschränkten Funktionsumfang verfügt. Alle Elemente der objektbasierten Mischung werden somit als Objekte in der Architect-Applikation definiert. Es werden maximal 128 Objektspuren unterstützt, wobei eine Unterscheidung zwischen *statischen* und *dynamischen* Objekten erfolgt. Statische Objekte werden an festen Positionen im Raum platziert (und können, bei Bedarf, als Bett eingesetzt werden), dynamische Objekte können sich hingegen räumlich bewegen und werden als einzelne Objekte encodiert, was eine präzisere Lokalisation und Bewegungsschärfe ermöglicht (Fedak, 2020).

Gestalterische Möglichkeiten

Durch Bereitstellung des Architects als AAX-Plugin wurde ein vereinfachter Workflow ermöglicht: Muss in der Standalone-Applikation noch für jede klangliche Veränderung erneut das DAW-Projekt aufgerufen und die Spur neu ex- und im Architect importiert werden, so kann das AAX-Plugin direkt in Pro Tools integriert werden. Grundlage hierfür kann somit sowohl das originale Pro Tools Stereo- bzw. 5.1 Surround-Projekt als auch ein neues Projekt sein, in welches die zuvor exportierten Stereo- bzw. 5.1 Surround-Stems und Audiospuren importiert und das AAX-Architect-Plugin eingebaut werden. Wie in

Abb. 3.7 dargestellt, können klangliche Anpassungen (wie Hall, EQ, Effekte) sowie die objektbasierte Mischung (wie 3D-Panning, Definition des Objekttyps) somit parallel erfolgen. Für jede Spur, auf der das Plugin eingefügt wurde, wird eine Objektspur im Architect erzeugt (Fedak, 2020). Sony empfiehlt für die 360 RA-Produktion, alle Elemente einer Mischung so weit wie möglich als separate Spuren zu exportieren, damit im Architect beispielsweise Instrumente und zugehöriger Hall separat vorliegen und platziert werden können. Insgesamt wird die Empfehlung ausgesprochen, nahezu jede Spur der im Vorfeld passierten Mischung separat bereitzustellen. Zur Verwendung eines LFE-Kanals kann die zugrundeliegende Audiodatei mit einem Tiefpass-Filter bearbeitet werden und entweder als Mono-Objekt oder dupliziert als Stereo-Objekt importiert werden (Sony Corporation, 2020b; Fedak, 2020).

Binaural-Einstellungen

Es empfiehlt sich, die objektbasierte Mischung sowohl über Lautsprecher (empfohlenes Setup: 13ch(Music)) als auch über Kopfhörer abzuhören. Das Binaural-Rendering für letzteres Wiedergabemedium geschieht in Echtzeit. Zur Kopfhörerwiedergabe sollte das von Sony bereitgestellte Profil verwendet werden, welches im Architect geladen werden kann. Außerdem besteht die Möglichkeit, ein eigenes, personalisiertes HRTF-Set in die Produktions-Software zu laden (Sony Corporation, 2020b; Fedak, 2020).

3.3.3 Exportformat: Audio + Metadaten

Da bei der 360 RA Produktion mit bis zu 128 Objekten gearbeitet werden kann, muss die Vielzahl an Objekten für das Musik-Streaming verringert werden. Deshalb werden im 360 RA Format drei Export-Level definiert (Level 1, 2, 3), sodass maximal 24 Objekte übertragen werden. Der drei-Level-Export muss für jede Produktion erfolgen und resultiert für jedes Level in einer Metadaten-Datei und den Audiodateien (Anzahl der Audiodateien entspricht maximaler Objekt-Anzahl des Levels) (Fedak, 2020).

Pre-Rendering

Im ersten Schritt des Exportvorgangs muss die Anzahl der Input-Objekte auf die passende Anzahl an Output-Objekten gerechnet und angepasst werden. Die Zahl der Output-Objekte setzt sich zusammen aus statischen Output-Objekten mit fester Positionierung

im Raum und dynamischen Output-Objekten. Für jedes Input-Objekt wird somit eine der beiden Output-Optionen eingestellt – dies kann sowohl manuell geschehen als auch automatisiert mittels der *Pre Analyze* Funktion. Diese wählt Objekte mit hohem Schalldruckpegel als dynamisch aus und verteilt alle weiteren so, dass sie statisch gerendert werden. Die Anordnung der statischen Output-Objekte ist definiert und abhängig von der ausgewählten Anzahl der Output-Objekte (2, 5, 7, 8, 9, 10, 11, 12, 13, 22) sowie des *Position Type* (Fedak, 2020). Eine Lautheitsnormalisierung ist aktuell kein Bestandteil des 360 RA Formates, jedoch sollte vor dem Export im Architect das *Max Peak Level* gemessen und auf < 0.00dB eingestellt werden.

Qualitätskontrolle

Zur Kontrolle der exportierten Dateien und um zu überprüfen, ob die Mischung wie erwartet reproduziert wird, ist es möglich, die zuvor exportierten Dateien im Architect zu importieren und nochmals abzuhören. Sofern dies den Erwartungen entspricht, können die Dateien encodiert werden.

3.3.4 Encodierung: MPEG-H 3D-Audio

360 Reality Audio basiert auf dem MPEG-H 3D Audio Standard (MPEG-H), spezifiziert als *ISO/IEC 23008-3*. MPEG-H ist ein Next Generation Audio Codec und unterstützt NGA-Merkmale wie immersives Audio (kanal-, objekt-, und szenenbasierte Übertragung sowie deren Kombinationen), Interaktivität und Personalisierung, sowie eine flexible und universelle Bereitstellung der produzierten Inhalte (Grewe, Simon & Scuda, 2018). Der MPEG-H 3D Audio Standard wurde u.A. speziell für die Integration in Streaming-Anwendungen entwickelt. 360 Reality Audio nutzt ein MPEG-H-Profil mit einem eingeschränkten Funktionsumfang. Es basiert auf dem *MPEG-H Low Complexity Profil* (dargestellt in Abb. 3.8) und unterstützt die Übertragung von bis zu 24 Objekten. Interaktivität, Kanäle und Lautheits-Normalisierung sind allerdings nicht im Funktionsumfang enthalten (Sony Corporation, 2020b).

Audio wird bei MPEG-H als eine Kombination aus Audio Komponenten und zugehörigen Metadaten beschrieben, wobei zwischen *statischen* und *dynamischen* Metadaten unterschieden wird: Statische Metadaten bleiben konstant und beschreiben beispielsweise Informationen zur Art des Audioinhaltes. Dynamische Metadaten hingegen verändern

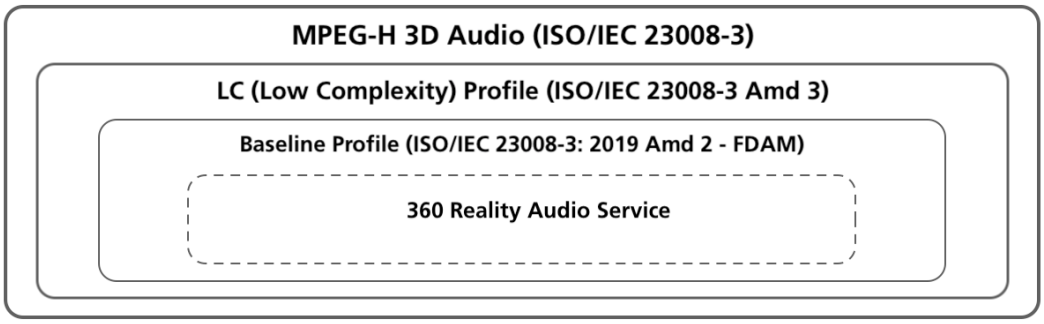


Abbildung 3.8: Struktur des 360 Reality Audio Service, basierend auf dem MPEG-H Low Complexity Profil, angelehnt an (Sony Corporation, 2020b).

sich über die Zeit (z.B. in den Positionsinformationen). Der Bitstream des MPEG-H Standards unterstützt bis zu 128 Kanäle oder Objekte, die gleichzeitig auf bis zu 64 Lautsprecher übertragen werden können. Da es je nach Anwendungsfall aus technischen Gründen nicht praktikabel ist, einen solch komplexen Bitstream an handelsübliche Geräte zu übertragen, wurden verschiedene MPEG-H Komplexitätsprofile (*MPEG-H Low Complexity Profiles*) eingeführt, um die Decoderkomplexität zu limitieren. Basis für den ATSC 3.0 Standard ist beispielsweise das *Low Complexity Profile Level 3*, mit welchem 32 Audio Elemente innerhalb eines Bitstreams übertragen werden und 16 davon gleichzeitig decodiert werden können (Herre et al., 2015; Grewe et al., 2018).

Audioobjekte, welche mit (zeitvariablen) Positionsmetadaten übertragen werden, werden im MPEG-H 3D-Audiodecoder durch einen VBAP Algorithmus gerendert. Dies ermöglicht die Wiedergabe desselben Audioinhalts auf einer Vielzahl unterschiedlicher Wiedergabesysteme. Hierfür wurde eine umfangreiche Dynamic Range Control (DRC) sowie ein Formatkonverter implementiert, welcher den MPEG-H Bitstream für das ensprechende Wiedergabeformat adaptiert. Aufgrund dieses Konzeptes wird somit im größten Wiedergabeformat produziert, die Anpassung für kleinere Formate erfolgt schließlich über einen Downmix-Algorithmus im Renderer des Wiedergabegerätes. Die Downmix-Faktoren sind hierbei während der Produktion flexibel definierbar. Jeder MPEG-H Decoder enthält außerdem einen binauralen Renderer, welcher die immersive Wiedergabe kanal-, objekt- und szenenbasierter Mischungen über Kopfhörer ermöglicht. DRC passt den Gesamtpegel an die geräteabhängige Ziel-Lautheit an, basierend auf der integrierten Lautheitsmessung jedes Elements der Audioszene. Für Mobile Endgeräte liegt diese Ziel-Lautheit beispielsweise bei –5 bis –12 LUFS (Jax, Meltzer, Neuendorf & Sen, 2014; Herre et al., 2015). Darüber hinaus normalisiert das MPEG-H System automatisch die Lautheit

gemäß gängiger Standards (z.B. EBU R-128, ITU-R BS.1770.4, ATSC A/85 etc.) (Jax et al., 2014). Da bei der Entwicklung von MPEG-H eine Rückwärtskompatibilität mit vorherigen Codecs wie MPEG-4 Advanced Audio Coding (AAC) keine Voraussetzung war, konnten neue Codierungstechnologien angewendet werden, welche – für eine möglichst hohe Codiereffizienz – auf vorherigen Codecs wie MPEG-D Unified Speech and Audio Coding (USAC), MPEG-D Spatial Audio Object Coding (SAOC) und MPEG-D MPEG Surround (MPS) basieren (Jax et al., 2014; Herre et al., 2015). Genaue Bitraten für verschiedene Audioformate sind in Tabelle 3.3 dargestellt. Hierbei gilt anzumerken, dass die Bezeichnungen der MUSHRA-Skala nach *ITU-R BS.1534-2* (ITU-R, 2014) entsprechen.

Format	**gut**	**empfohlen**	**transparent**
Stereo	48 kbp/s	64 kbp/s	96 kbp/s
5.1	128 kbp/s	192 kbp/s	256 kbp/s
7.1.4 (+ 2 Obj.)	256–288 kbp/s	384–420 kbp/s	512–576 kbp/s
22.2	512 kbp/s	768 kbp/s	1024 kbp/s

Tabelle 3.3: MPEG-H Bitraten verschiedener Qualitätsstufen (Fraunhofer IIS, 2020b).

3.3.5 Distribution und Wiedergabe

Zur Wiedergabe über Musik-Streaming-Dienste wurde in die jeweiligen Mobil-Apps ein entsprechender Decoder und Renderer implementiert, welcher die objektbasierte Produktion für die Kopfhörerwiedergabe rendert. Hierfür wird ein von Sony bereitgestelltes Profil verwendet, welches den Produzenten auch bereits für die 360 RA-Mischung im Architect zur Verfügung steht. Bei der Wiedergabe auf dem Amazon Echo Studio erfolgt das Decoding auf dem Gerät. Die Format Spezifikationen sollen außerdem Hardware-Herstellern zur Verfügung gestellt werden, sodass diese auch in Endgeräten verbaut werden können. Durch die Verwendung der *Sony Headphones Connect App* in Kombination mit ausgewählten Sony Kopfhörern (eine Auflistung aller unterstützten Kopfhörer findet sich auf der Sony Website (Sony Corporation, 2020a)) kann außerdem eine Personalisierung erfolgen. Hierbei werden die Kopfhörer über Bluetooth mit der Headphones Connect App verbunden und Fotos beider Ohren aufgenommen. In der App erfolgt eine Analyse der Ohrform, worauf basierend das binaurale Rendering für den Nutzer entsprechend angepasst wird (Sony Corporation, 2020a; Fedak, 2020).

3.4 Zusammenfassende Betrachtung

Basierend auf bisherigen Erkenntnissen ist eine Auflistung einiger Spezifikationen der Dolby Atmos Music sowie 360 Reality Audio Produktion und Distribution in Tabelle 3.4 dargestellt. Eine ausführlichere Gegenüberstellung beider Technologien und deren Produktions-Workflows findet sich in Kapitel 6.

	Dolby Atmos Music	**360 Reality Audio**
Benötigte Software	Dolby Atmos Production- und Mastering Suite und Dolby Atmos fähiger Panner	Architect (Plugin oder Standalone-Applikation)
Unterstützte DAWs	Ableton Live, Logic Pro, Pro Tools, Nuendo	Pro Tools (mit Architect Plugin)
Empfohlenes Lautsprecher-Layout	7.1.4	13ch(Music)
Integrierte Lautheitsmessung	Ja	Nein
Unterstützte Objektanzahl	118	128
Produktionsweise	objekt- und kanalbasiert	objektbasiert
Exportformat (vor Encodierung)	BWF/ADM (eine Datei)	Metadaten + Audiodateien (drei Dateien, Level 1–3)
Codec der Distributionsdatei	AC-4 IMS / DD+JOC (zwei Dateien)	MPEG-H (eine Datei)
flexibles Wiedergabe-Rendering der Distributionsdatei	Nein (zwei Dateien für Binaural und Mehrkanal)	Ja
Personalisierungs-möglichkeiten	Nein	Ja (Sony Headphone Connect App)

Tabelle 3.4: Vergleich von Dolby Atmos Music und 360 Reality Audio Spezifikationen.

3.5 Mastering von Audioobjekten

„The main difference between stereo and immersive mastering is in the complexities of the channels and the delivery. [...] What gets sent to the mastering engineer? Individual tracks, ADM files, stems? A combination? Right now, there are very few tools available for immersive mastering. [...] Delivery for Immersive Audio to the manufacturers or streaming services is another challenge. Each platform has different requirements. And each format has its own software for creating those deliverables.“

(Romanowski, 2020, S. 32–33)

Hat sich das herkömmliche, kanalbasierte Mastering über die Jahre zu einem festen Teil der Musikproduktion etabliert, so erfordert objektbasiertes Mastering eine neue Herangehensweise: Dadurch, dass keine finale Mischung übertragen wird, die bei jedem Konsumenten in der gleichen Art und Weise wiedergegeben wird, sondern eine Kombination aus Audio + Metadaten, kann am Ende der Produktion nicht auf herkömmliche Masteringtechniken zurückgegriffen werden – es müssen nicht wie bei kanalbasierten Produktionen Lautsprechersignale, sondern Audioobjekte gemastert werden. Etablierte Konventionen existieren hierfür aktuell noch nicht (Hestermann et al., 2018).

Um objektbasiertes Mastering genauer zu betrachten, sollte dieser allgemeine Schritt in mehrere Einzelteile aufgeteilt werden: Klangliches Mastering (EQ, Kompression), Album-Mastering, Archivierung und Backups sowie Distributionsvorbereitungen. Insbesondere Dynamikbearbeitung (Compressor, Expander, Limiter) entscheidet bei der Musikmischung über den Gesamteindruck. Wird bei kanalbasierten Mischungen (wie Stereo oder 5.1 Surround) oft mit Bus-Kompression gearbeitet, um einzelne Instrumente, bestimmte Frequenzbereiche oder den Gesamtklang zu komprimieren, so stellt dies bei immersiven Produktionen eine erhöhte Schwierigkeit dar, da zum einen oft die Mehrkanal-Kompressoren bzw. Limiter fehlen, zum anderen bei objektbasierten Mischungen das flexible Rendering dies nicht zulässt (Lawrence, 2019; Hestermann et al., 2018). Insbesondere der Aspekt des Binaural-Renderings und damit einhergehenden klanglichen Verfärbungen (durch Verwendung verschiedenster Binaural-Renderer) stellt eine weitere Herausforderung dar, da während des Masterings kaum Einfluss darauf genommen werden kann (siehe Abschnitt 6.3.3).

Generell lässt sich feststellen, dass das klangliche Mastering – anders als bei kanalbasierten Produktionen – erfolgen muss, bevor das finale Dateiformat erreicht wurde. Bei

DAM ist es durch den Export als ADM-Datei noch möglich, an diesem Punkt einen Mastering-Schritt zu integrieren. Bei 360 RA jedoch wird aus der Software (Architect) eine Kombination aus Audio + Metadaten exportiert, welche in einem Sony spezifischen Format gespeichert werden und nur vom Encoder gelesen werden kann. Das Mastering muss somit zuvor erfolgen.

Ein Ansatz wäre, den Schritt des klanglichen Masterings nicht ans Ende zu setzen, sondern zwischen die Anlieferung der Audiodateien und den Beginn der objektbasierten Mischung – das Mastering erfolgt hierfür somit bei der Ausspielung der Stereo-Stems bzw. Spuren, welche einzeln gemastert beim Produzenten ankommen. Eine andere Vorgehensweise setzt diesen Schritt in die objektbasierte Mischung: Um die Stereo-Buskompression zu ersetzen, wird auf jedes Objekt (und Bett) ein Kompressor gelegt. Jedes Objekt wird schließlich mit dem Gesamt-Master verknüpft und gleichmäßig komprimiert (Harvey, 2020). Eine weitere interessante Herangehensweise wird von Steve Genewick (2020) beschrieben, welcher die fehlende Möglichkeit der herkömmlichen Buskompression nicht zu ersetzen versucht, sondern stattdessen mit der neu gewonnenen Räumlichkeit arbeitet:

> *„Because of the lack of effective dynamics bus processing (at this point in time at least), I've had to find ways of finishing off a mix without the use of bus compression to achieve either the apparent level I'm looking for or that 'glue' effect we all want. [...] I've found that the system can handle the wide dynamic range, and that most music benefits from it. [...] We're not trying to jam a ton of stuff into a small space anymore. If two elements are competing for frequency range, simply move them apart"* (Genewick, 2020, S. 42).

Diese dreidimensionale Räumlichkeit spielt auch beim klanglichen Mastering eine Rolle:

> *„When I am mastering a project, I am listening for the authentic nature of the presentation and the tone and space each element takes up. The art form of Immersive Mastering is in making sure that the tonal and level transition occur smoothly across the sound sources. Filling in the sonic gaps and pulling back the peaks, if needed"* (Romanowski, 2020, S. 33).

Neue Wiedergabemöglichkeiten wie Streaming führen ebenfalls dazu, dass sich der Masteringprozess wandelt: Der Schritt, verschiedene Titel als Album zusammenzuführen, gerät teilweise in den Hintergrund, da Musik-Streaming beim Hörer oftmals titelbasiert erfolgt und nicht durch Wiedergabe eines kompletten Albums. Somit erfolgt Mastering oft

nur für einzelne Titel, nicht für ein gesamtes Album – der Schritt des Album-Masterings entfällt teilweise.

> *„People just pick and choose, and I'm guessing that [...] streaming will exacerbate that. But it's something that will need to be figured out in the future to make track-to-track [...] mixes sound like an album“* (Harvey, 2020, S. 30).

Desweiteren arbeiten oftmals verschiedene Toningenieure an Titeln, die auf einem Album erscheinen. Mit Wegfallen einer allgemeinen, finalen Mastering-Instanz liegen somit die Mastering-Entscheidungen bei den einzelnen Produzenten der Mischung. Insbesondere deshalb und in Anbetracht dessen, dass noch keine etablierten objektbasierten Mastering-Konventionen vorherrschen, wird es umso wichtiger, dass sich Produzenten an vorgegebene Spezifikationen halten. Aktuell – mit Dolby Atmos Music und 360 Reality Audio als einzige, kommerzielle Technologien – und der eher eingeschränkten Verfügbarkeit der Produktions-Software kann dies noch erreicht werden. Sobald die objektbasierte Produktionsweise jedoch weithin verfügbar und der Zugang somit vereinfacht wird, könnten sich die fehlenden Mastering-Konventionen bemerkbar machen (Harvey, 2020; Romanowski, 2020).

Ein neuer Mastering Schritt ist außerdem die Überprüfung des Authoring-Prozesses (korrekte Formate und Metadaten), Archivierung objektbasierter Mischungen (beispielsweise als ADM-Dateien) sowie die Encodierung in das entsprechende Format und die damit verbundene Vorbereitung für sämtliche Distributions- und Wiedergabewege. War das Wiedergabeformat bei kanalbasierten Produktionen klar definiert, könnte bei objektbasierter Musik die Qualitätskontrolle bei Wiedergabe auf Lautsprechern (Stereo, 5.1, 5.1.4, 7.1.4), Soundbars, Smart Speakers oder Kopfhörern ein neuer Teil des Mastering-Prozesses werden (Romanowski, 2020).

Abschließend lässt sich festhalten, dass das Thema *objektbasiertes Mastering* zukünftig eine wichtige Rolle spielen wird, sollte sich objektbasierte Musikwiedergabe bei den Konsumenten etablieren. Insbesondere durch die neue Produktionsweise müssen Workflows erarbeitet werden, die die herkömmlichen Mastering-Prozesse ergänzen und umwandeln. Die Entwicklung von geeigneter Hard- und Software kann entscheidend zu diesem Schritt beitragen, Anforderungen hierzu werden durch aktuelle Produktionen definiert:

> *„As an example, if I want to use an EQ for a song, I would like to be able to apply it to any, all or some of the channels. I would also like to see a compressor/limiter that would allow me to link or trigger for any single or grouped audio, or be able to apply the same amount of dynamic reduction across all channels that are linked, based on the trigger channel. For now, we need to work around the limitations of our tools. DAWs present another challenge. It is unrealistic to try and master out of the box“* (Romanowski, 2020, S. 33).

Sobald Werkzeuge (Hard- und Software) für objektbasiertes Mastering zur Verfügung stehen, könnten dedizierte Einrichtungen mit optimierten akustischen Umgebungen und eingemessenen Abhörmöglichkeiten Mastering-Dienste anbieten. Dies würde den Markt der immersiven, objektbasierten Musikproduktion deutlich öffnen, da somit auch kleinere, weniger gut ausgestattete Studios und Künstler in diesen Formaten produzieren könnten, und sich für Kompatibilität mit anderen Wiedergabesystemen und Formaten sowie klanglicher Optimierung auf die Mastering-Studios verlassen könnten (Hestermann et al., 2018). Hestermann et al. (2018) stellen weiterführend einen Prototypen zur Implementierung einer Mastering Unit vor, bei welchem mittels sogenannter *Mastering Objekte* im Produktions-Workflow vor dem Rendering-Prozess ein Audio- und Metadaten-Mastering stattfinden kann. Eine detailliertere Betrachtung kann im Rahmen dieser Thesis nicht stattfinden, daher soll an dieser Stelle auf (Hestermann et al., 2018) verwiesen werden.

4 Entwicklung eines Produktions-Workflows

„I need to have control over individual instruments in order to imagine re-combining them in a 3-D space instead of a two-dimensional plane. [...] What you want is a really cool, evolving mix, and for that you need to have a variety of elements to work with, to create movement as well as balance.“
Romanowski in (Schultz, 2020b, S. 11)

Zu Beginn eines jeden Musikproduktions-Workflows, egal welchen Formates, steht die konzeptuelle Vision: Wie soll das Endprodukt klingen, welche künstlerische Absicht steht dahinter (Owsinski, 2006; Izhaki, 2008; Lawrence, 2019)?

Durch die erweiterten und neuen Möglichkeiten objektbasierter und immersiver Musikproduktion sowie gegeben durch die Tatsache, dass diese Produktionsweise aktuell noch am Anfang steht und es (im Vergleich zur Stereo-Produktion) noch an etablierten Konventionen mangelt, sollte vor Beginn einer objektbasierten Produktion ein Klangkonzept erstellt werden. Dazu gehören künstlerische und technische Aspekte sowie die Mittel und Ressourcen, die zur Erreichung dieser Ziele benötigt werden. Bei all diesen Überlegungen gilt, dass letztendlich der Hörer überzeugt werden muss – unter Berücksichtigung der Art und Weise, wo, wie und wann die Musik gehört werden wird (Lawrence, 2019).

Der Begriff „Workflow“ soll im Folgenden eine Reihung an Arbeitsabläufen beschreiben, die zur Produktion objektbasierter Musik – im Speziellen Dolby Atmos Music und 360 Reality Audio – nötig sind. Insbesondere bei der Verwendung neuartiger Technologien wie DAM und 360 RA gilt es, eigene und bereits etablierte Workflows aus herkömmlichen Produktionsformaten mit neuen Methoden zu kombinieren, um eine möglichst effiziente Produktionsweise zu erreichen. Für diese Arbeit liegt der Fokus auf der Entwicklung eines Workflows für immersive, objektbasierte Musikproduktion auf Basis von bereits angefertigten Stereo- bzw. 5.1 Surround-Mischungen[1].

[1] Da im praktischen Teil auf Basis einer Stereo-Mischung vorgegangen wird, wird im Folgenden auf die Nennung von „5.1 Surround“ verzichtet. Aufgrund der ähnlichen Produktionsschritte soll 5.1 bei der Nennung von Stereo impliziert werden, sofern nicht explizit ausgeschlossen.

4.1 Anlieferung des Materials

Romanowskis Aussage spiegelt sich auch in weiterer Literatur wider – so scheint sich der Ansatz der Anlieferung von einzelnen Spuren und Mehrkanal-Stems aktuell zu verbreiten, wobei das Audiomaterial meist aus einer Mischung an trockenen und bereits prozessierten Signalen besteht. Effekte werden separat zur Verfügung gestellt oder erst bei der immersiven Mischung erstellt oder nachgebaut (Schultz, 2020b, 2020a; Lawrence, 2019). Insgesamt gilt zu beachten: Je größer die Vielfalt des Ausgangsmaterials und je differenzierter es vorliegt, desto größer ist die Flexibilität bei der Mischung, da somit beispielsweise Effekte wie Hall oder Delay separat vom „trockenen“ Signal positioniert werden können. Liegen beispielsweise alle Schlagzeug- und Perkussionsinstrumente als einziges Stem vor, kann daraus nur ein einziges Bett[2] oder Objekt erstellt werden, was zu einer Einschränkung der kreativen Gestaltungsmöglichkeiten führt. Die Verwendung von Mono-Spuren bietet im Vergleich zu Stereo-Spuren mehr Kontrolle über Objektpositionen.

4.2 Kombinierte Produktionsumgebung

Für einen möglichst formatübergreifenden Workflow zur Dolby Atmos Music und 360 Reality Audio Produktion bietet sich die Möglichkeit, eine gemeinsame Produktionsumgebung zu nutzen, um die Vorgehensweisen für beide Formate zu kombinieren.

Basierend auf den bisherigen Erkenntnissen durch Analyse der aktuellen Produktions- und Distributionskette für Dolby Atmos Music (dargestellt in Abb. 3.2) und 360 Reality Audio (dargestellt in Abb. 3.7) soll im Folgenden ein Vorschlag eines solchen kombinierten Workflows beschrieben werden. Dieser reicht von der Anlieferung der Stems bzw. Audiospuren bis zur Encodierung, auf die darauffolgenden Schritte wurden in dieser Darstellung verzichtet, diese sind in den o.g. Diagrammen einsehbar. Der kombinierte Workflow ist in Abb. 4.1 dargestellt (Dolby Laboratories, 2018, 2020d; Sony Corporation, 2020b; Fedak, 2020; Thomas, 2020a). Die praktische Anwendung und Weiterentwicklung des kombinierten Produktions-Workflows findet sich in Kapitel 5.

[2] Der Begriff „Bett“ wird im Folgenden sowohl für Kanäle (Dolby Atmos Music) als auch für statische Objekte (360 Reality Audio) verwendet.

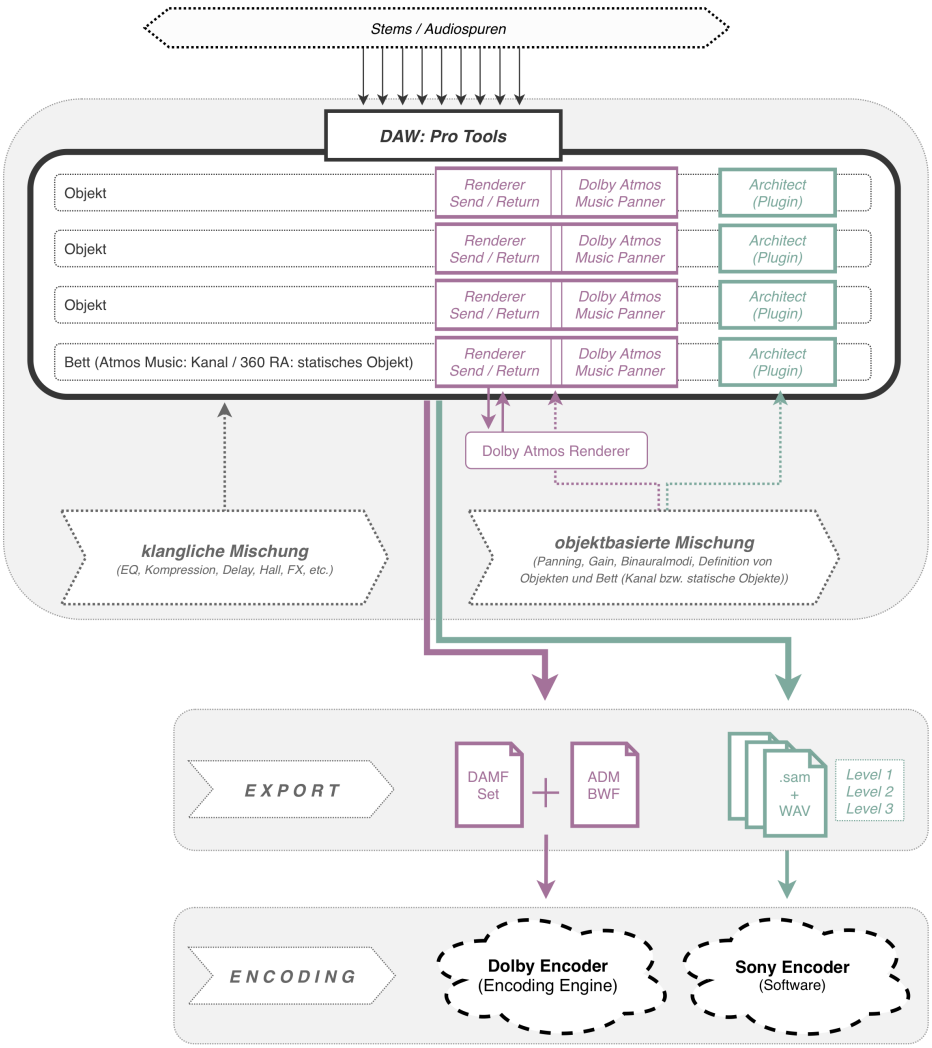


Abbildung 4.1: Kombinierter Produktions-Workflow für Dolby Atmos Music und 360 RA. Stand: 08/2020

Da sowohl für die Dolby Atmos Produktion Pro Tools die am besten unterstützte DAW ist, als auch das Architect AAX-Plugin für die Nutzung mit Pro Tools entwickelt wurde, bietet sich diese DAW als kombinierte Produktionsumgebung für beide Formate an. Es gilt an dieser Stelle anzumerken, dass eine weitere Möglichkeit wäre, die DAM-Mischung in einer der unterstützten DAWs (Pro Tools, Nuendo, Ableton Live, Logic Pro) zu erstellen, alle Audio- und Aux-Spuren mit zugehörigen Plugins als Einzelspuren zu exportieren und diese in der Architect Standalone-Applikation zu importieren. Dies ermöglicht jedoch keine kombinierte Abhörsituation innerhalb einer Produktionsumgebung, des Weiteren müssten bei klanglichen Veränderungen alle Spuren neu exportiert und im Architect importiert werden. Daher wird auf diese Produktionsweise nicht detaillierter eingegangen.

5 Anwendung des Workflows

„Learned production techniques and approaches are still relevant, since spatial audio is an extension of stereo.“
Nipkow, zitiert nach (Lawrence, 2019, S. 150)

Ziel des Praxisteils war es, eine möglichst realistische Situation zu simulieren, in welcher eine objektbasierte Produktion aus einer angelieferten Stereo-Mischung entstehen soll. Da hierfür aktuell zwei kommerziell relevante Technologien existieren (Dolby Atmos Music und 360 Reality Audio), sollten diese hierfür verwendet werden.

Die praktische Anwendung des entwickelten kombinierten Produktions-Workflows erfolgte am Beispiel des Songs *Kentia Danca* von *RIAD & J.K.Rollin'*[1]. Die Aufnahmen hierzu sind 2017 im Rahmen einer Studioproduktion des Studienganges „Audiovisuelle Medien“ an der Hochschule der Medien in Stuttgart entstanden. Das Konzept der Band „Rhythm is a Dancer (RIAD)“ ist live improvisierte elektronische Musik, die Grundlage für die Produktion war eine Jam-Session der Musiker wobei die Instrumente – abgesehen von Schlagzeug, Gesang und Saxofon – über DI-Signale aufgenommen wurden. Das Schlagzeug besteht aus hybriden Sounds (Akustik-Set und elektronische Sounds, welche über ein Pad getriggert wurden).

Basierend auf den Erfahrungen und Erkenntnissen der praktischen Anwendung wurde das in Abb. 4.1 dargestellte Workflow-Diagramm weiterentwickelt. Die Abbildung der im praktischen Teil erprobten Vorgehensweise findet sich in Abb. 5.1, dort ist der gesamte Workflow der kombinierten Produktionsweise von Dolby Atmos Music und 360 Reality Audio grafisch vereinfacht dargestellt. Im Folgenden werden die einzelnen Schritte detaillierter betrachtet und dokumentiert.

[1] Arrangement und Produktion: Jonas Kieser, Gesang: Tony Mac, Saxofon: Jakob Manz, Schlagzeug: Julian Feuchter, Bass: Brian Thiel, Gitarre: Julian Kaspar, Keyboard: Peter-Philipp Röhm.

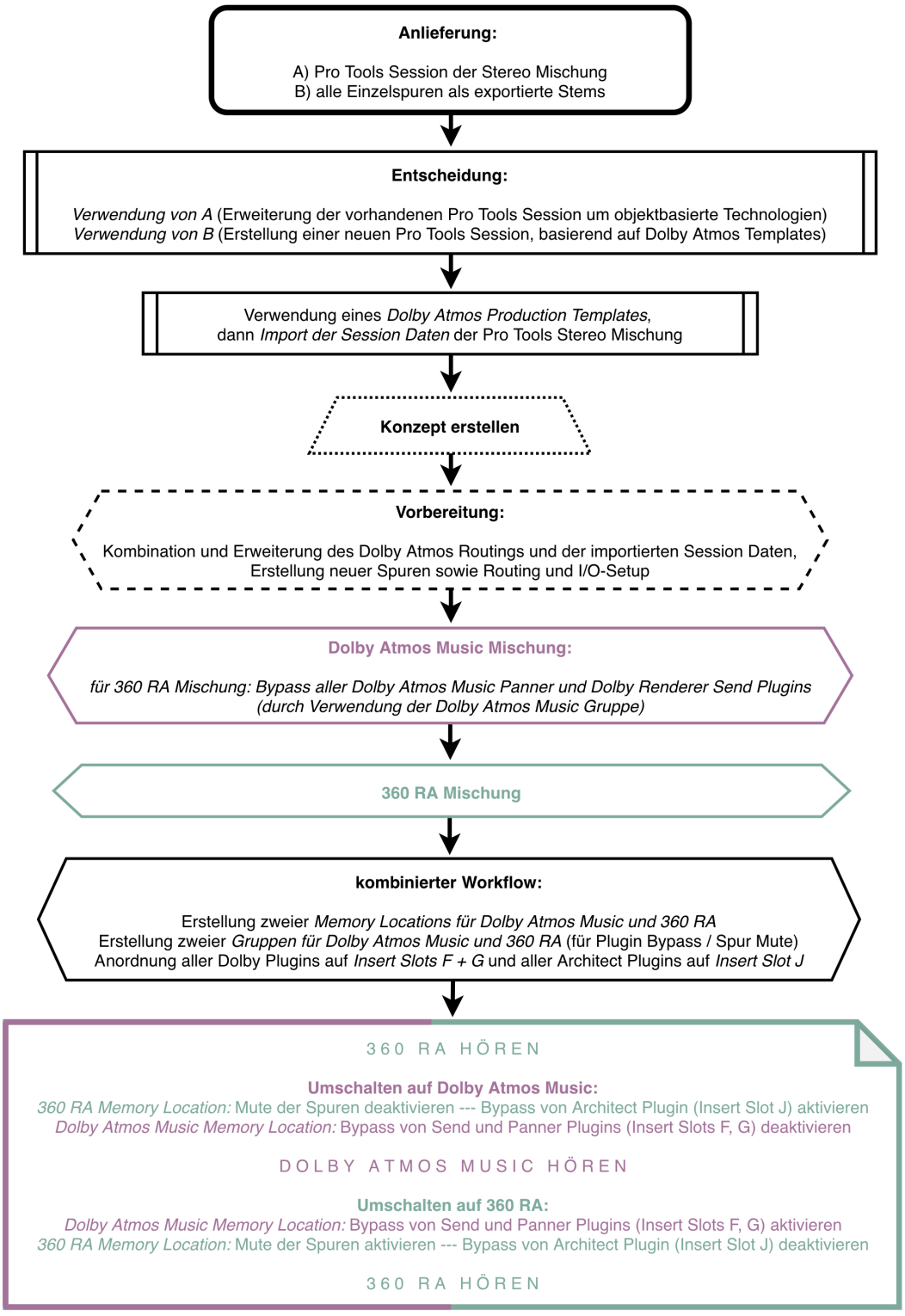


Abbildung 5.1: Darstellung eines kombinierten Workflow zur Dolby Atmos Music und 360 Reality Audio Produktion.

5.1 Vorbereitung

5.1.1 Analyse des gelieferten Stereo-Materials

Das gelieferte Material bestand aus zwei Teilen: Zum einen der Pro Tools Session der Stereo-Mischung (dargestellt in Abb. 5.2), welche sowohl alle Einzelspuren (Instrumente) als auch verschiedenen Aux-Spuren[2] (Schlagzeug-, Mix-, und Master-Bus, außerdem verschiedene Hall-, und Effekt-Spuren) beinhaltet. Die Plugins der Einzelspuren wurden bereits in die Audiodateien eingerechnet, die Plugins der Aux-Spuren sind noch im Projekt vorhanden. Zum anderen wurden Exporte aller Audiospuren bereitgestellt.

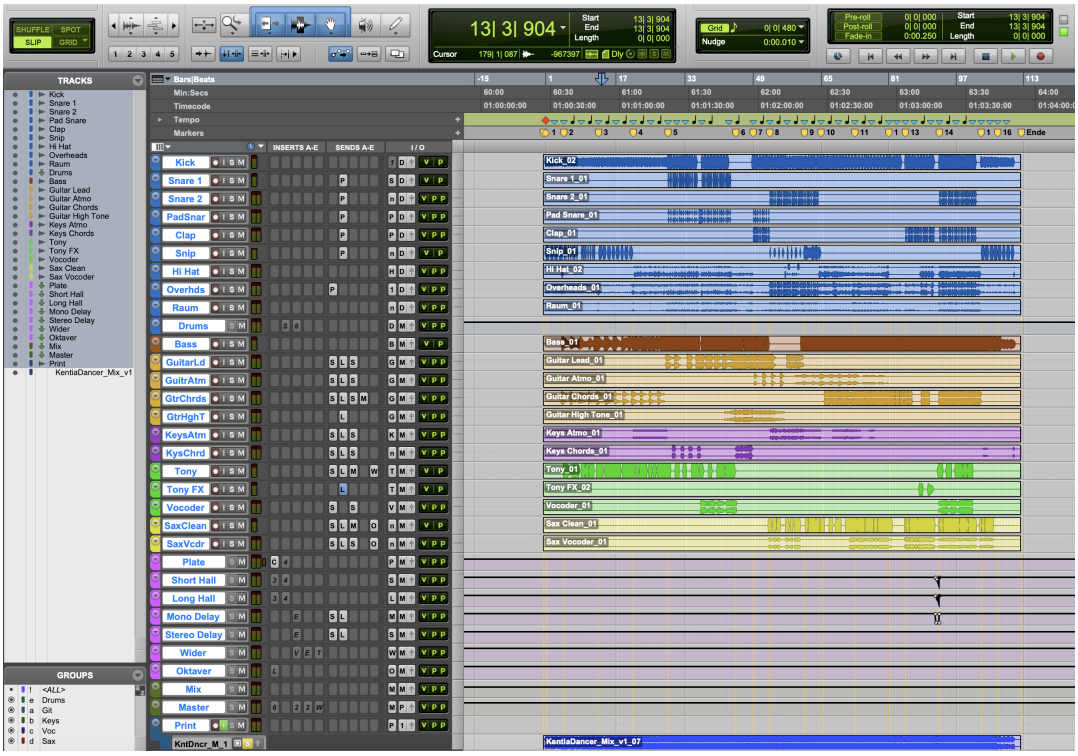


Abbildung 5.2: Anlieferung der Stereo-Mischung (Pro Tools Session).

Zu Beginn der objektbasierten Mischung stand somit die Entscheidung, welche der beiden in Abb. 5.1 genannten Möglichkeiten verwendet werden sollte. Option A würde eine Erweiterung der vorhandene Pro Tools Session um DAM und 360 RA Plugins und Routingstrukturen bedeuten. Der Vorteil hierbei wäre, dass die gelieferte Struktur der Spuren sowie das Routing der Audio- auf Aux-Spuren beibehalten und somit Hall und Effekte

[2]Der Begriff „Aux-Spur“ bezieht sich im Folgenden auf die in der DAW Pro Tools integrierte „Auxiliary-Eingangs-Spur“, welche die gleichen Möglichkeiten zum Signal-Routing bieten wie Audiospuren, außer dass die Eingangssignale von einem internen Bus oder Hardware-Eingang stammen müssen.

weiterverwendet werden könnten. Der Nachteil wäre, dass die aufwändige Dolby Atmos Bus- und Routing-Struktur manuell erstellt werden müsste. Die Anwendung von Option B würde bedeuten, dass eine neue Pro Tools Session erstellt wird, welche auf einem *Dolby Atmos Production Template* basiert. Dies würde den Vorteil mit sich führen, dass die Dolby Atmos Routing- und Bus-Struktur bereits vorhanden wäre, allerdings müssten die Zuweisungen auf Hall/Effekt/Master-Spuren sowie die Plugin Einstellungen neu erstellt werden.

Aus diesem Grund fiel die Entscheidung auf die Verwendung eines Dolby Atmos Production Templates (dargestellt in Abb. 5.3) mit anschließendem Import der Session Daten der Pro Tools Stereo-Mischung. Diese Vorgehensweise ermöglichte, dass sowohl auf vorhandenes Dolby Atmos Routing zurückgegriffen werden konnte, als auch die Routing Entscheidungen der Stereo-Mischung erhalten blieben.

Abbildung 5.3: Dolby Atmos Production Template (Pro Tools Session).

5.1.2 Entwicklung eines Konzeptes

Im nächsten Schritt folgte die Entwicklung eines Konzeptes, basierend auf der Analyse einzelner Elemente sowie der folgenden Leitfragen: Welchem Genre ist das Material zuzuordnen? Wie viele Instrumente sind enthalten? Wie ist der Titel aufgebaut? Was wird musikalisch erzählt und wie kann dies räumlich unterstützt werden? Welche Art von Elementen liegt einzeln vor? In welchem Zustand befindet sich das Material? Was bietet sich als Bett und Objekt an? Wie lässt sich Räumlichkeit erschaffen? Welches Element kann den dreidimensionalen Raum durch Panning-Automation nutzen und bereichern?

Eine der Ideen hinter dem zugrunde liegenden Material war schließlich, kein dediziertes Kanal-Bett zu verwenden. Dies würde einerseits zwar dem Vorgehen einer Dolby Atmos Mischung entsprechen, allerdings werden Kanäle bei 360 RA nicht unterstützt. Stattdessen sollten einzelne Elemente bzw. Anteile eben dieser auf ein 7.1.2 Bett geroutet werden (z.B. Kick-Drum auf C-Kanal des 7.1.2 Bettes, Overheads auf Ltf+Rtf-Bett, Schlagzeug-Raum auf Lrs+Rrs-Bett).

Abgesehen von Kick-Drum, Overheads, Schlagzeug-Raum und den Aux-Spuren wurden alle weiteren Instrumente Workflow-übergreifend als Objekte definiert (teilweise statisch, beispielsweise Schlagzeug- und Gitarren-Elemente). Instrumente und Effekte, welche nur an ausgewählten Stellen erklingen, eignen sich besonders für automatisiertes Panning im 3D-Raum, um klangliche Akzente zu setzen (z.B. Claps, atmosphärische Gitarre, Pad Snare etc.). Nach genauerer Betrachtung der Aux-Spuren wurden die unterschiedlichen Effekte und Halleinstellungen auf statische Positionen der Bett-Spuren gelegt. So erfolgte die Verteilung des Halls im Raum (L+R, Lrs+Rrs, Ltf+Rtf) sowie das Routing der Aux-Schlagzeug-Spur auf das 7.1.2 L+R-Bett. Die Master-Bus-Spur mit den Mastering Plugins wurde beibehalten, wobei jedoch – aufgrund der objektbasierten Produktionsweise – keine abschließende Zweikanalspur der Mischung darauf geroutet werden konnte. Stattdessen wurden Anteile einzelner Objekte (jene, die zusätzlich komprimiert werden sollten wie Schlagzeug, Stimme, Saxofon etc.) auf den Master-Bus gesendet, und dieser schließlich auf das L+R-Bett.

Der letzte Schritt der Vorbereitung war schließlich, das bereits vorhandene Dolby Atmos Routing des Templates mit den importierten Session Daten zu kombinieren und zu erweitern. Hierzu wurden neue Spuren erstellt, irrelevante Elemente gelöscht sowie das I/O-Setup angepasst (beispielsweise durch Erstellung von Stereo-Objekt-Bussen).

5.2 Produktion objektbasierter Musik

5.2.1 Dolby Atmos Music

Begonnen wurde schließlich mit der objektbasierten Mischung für Dolby Atmos Music, gegeben durch die Tatsache, dass durch die Verwendung des Templates und der Vorbereitung bereits alles für Dolby Atmos aufgesetzt war. Außerdem konnte somit noch mit

Lautstärke und Einstellungen der Plugins der Aux-Spuren experimentiert werden. Da zur 360 Reality Audio Mischung die Pro Tools Spuren im Architect Plugin aufgenommen werden, sollten klangliche Entscheidungen bestenfalls vorher getroffen werden, um ein stetiges Neu-Aufnehmen zu umgehen. Als Rendering-Layout wurde – wie zur Produktion von DAM empfohlen – ein 7.1.4 Layout ausgewählt. Mit der Entscheidung, welche Spuren als Bett und Objekte definiert werden, erfolgte das Routing der Einzelspuren (Objekte) auf die entsprechenden (Dolby Atmos) Objekt-Send-Spuren, welche das Audio weiter zum Dolby Atmos Renderer leiten. Die Hall-Einstellungen wurden so beibehalten, wie bei der Stereo-Mischung gewählt. Das Routing auf die Bett-Sub-Busse wurde an den dreidimensionalen Raum angepasst. Mittels des DAM-Panners (dargestellt in Abb. 5.4) wurden die Objekte mit den Objekt-Send-Spuren verknüpft und sowohl statisch im Raum platziert als auch dynamisch automatisiert gepannt.

Abbildung 5.4: Dolby Atmos Music Panner (eines Stereo Objektes).

Zur Übersicht der Positionen und Bewegungen aller Objekte diente der Dolby Atmos Renderer (dargestellt in Abb. 5.5), welcher als Standalone-Applikation durch die Send-Plugins der Objekt-Send-Spuren angesteuert wurde. Die Anpassung der Lautheit auf –18 LUFS fand mit der integrierten Lautheitsmessung des Dolby Atmos Renderers statt (siehe Abb. 5.6). Hierzu wurde der gesamte Titel abgespielt und die Lautheit in Echtzeit gemessen.

Nach der Dolby Atmos Music Mischung wurde der Bypass aller Dolby Atmos Plugins (Music Panner und Renderer Sends) aktiviert, um mit der 360 RA Mischung fortzufahren.

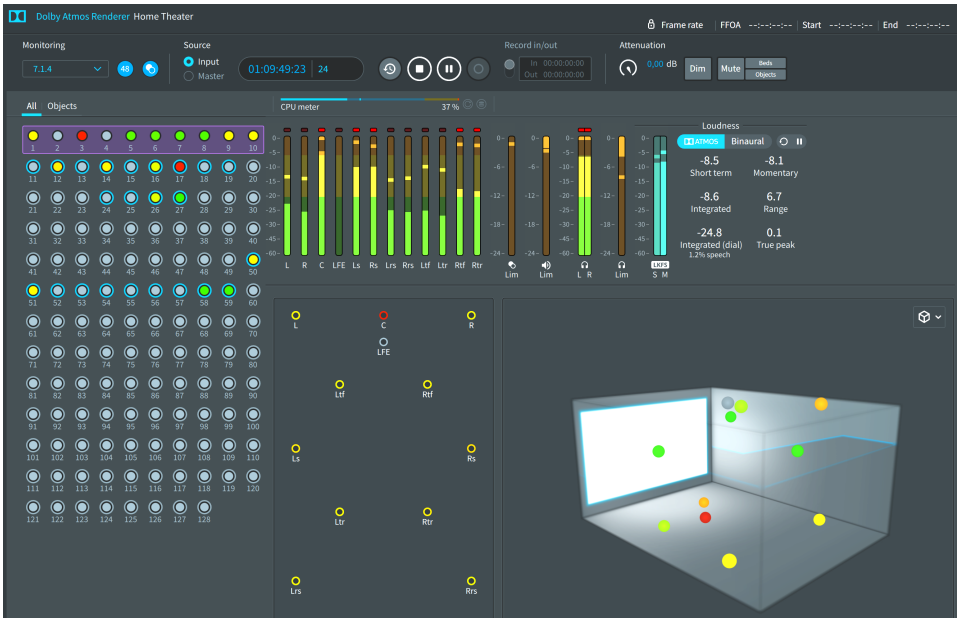


Abbildung 5.5: Dolby Atmos Renderer, Ansicht der Nutzeroberfläche.

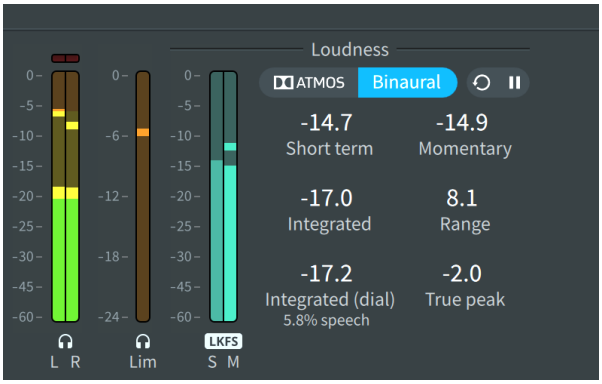


Abbildung 5.6: Integrierte Lautheitsmessung des Dolby Atmos Renderers.

5.2.2 360 Reality Audio

Zur Mischung in 360 RA wurde schließlich die Dolby Atmos Struktur um das Architect Plugin erweitert. Die empfohlene Verwendung des Plugins besteht darin, dieses in separaten Aux-Spuren statt der Objekt-Einzelspuren einzufügen, um diese Spuren weiterhin editieren oder muten zu können. Daher wurde der Architect nicht auf den einzelnen Spuren, auf denen sich auch der DAM Panner befindet, eingefügt, sondern auf die Objekt-Send-Spuren gelegt sowie auf weitere Spuren, die bei der DAM Mischung ins

Kanal Bett gehen (dargestellt in Abb. 5.12). Bei allen Objekt-Send-Spuren, auf denen das Plugin liegt, wurde schließlich „Mute“ aktiviert, damit das Audio nur aus dem Architect Plugin abgespielt wird und nicht auch aus Pro Tools.

Da zur Positionierung der Objekte die Audiospuren im Architect Plugin aufgenommen werden müssen (dargestellt in Abb. 5.7), wurde hierfür der Titel über die volle Länge einmal abgespielt.

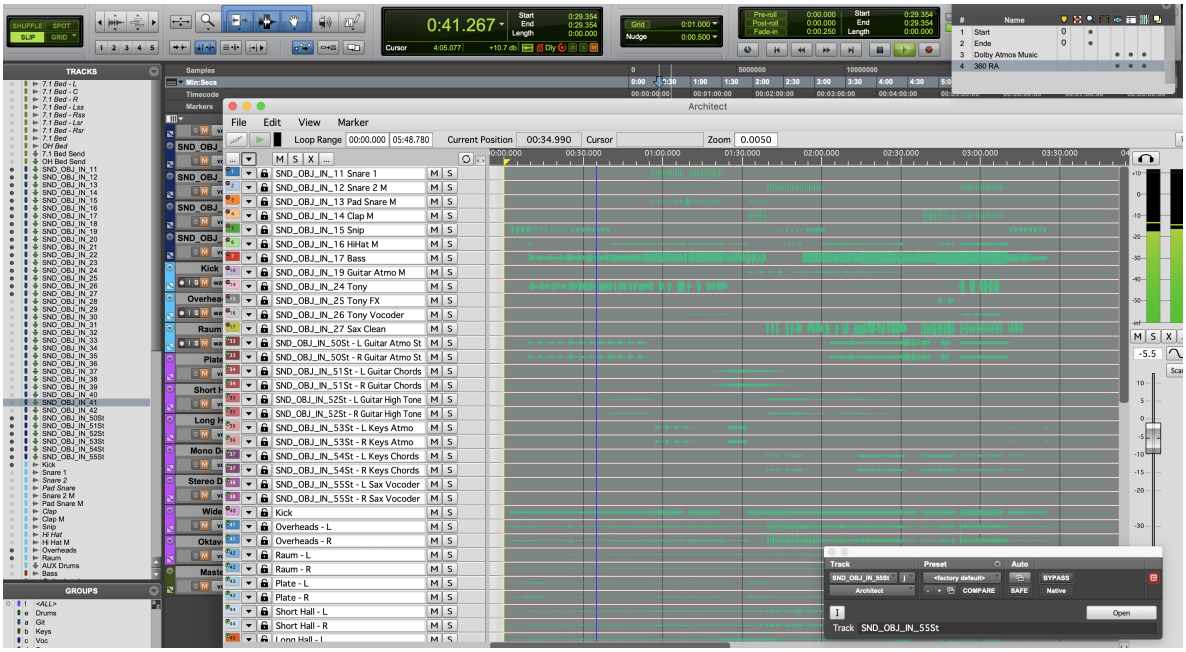


Abbildung 5.7: 360 Reality Audio (Architect Plugin).

Es hat sich jedoch in der praktischen Anwendung herausgestellt, dass Änderungen, die auf diesen Audiospuren unternommen werden – z.B. Insertion von Plugins und damit einhergehende klangliche Veränderungen, Schnitte und Verschieben der Audiodateien – dennoch korrekt abgespielt werden. Nur wird dies im Architect Plugin solange inkorrekt grafisch dargestellt, bis die Wellenformen gelöscht und neu aufgenommen werden.

In Abb. 5.8 ist die korrekte Ansicht der Wellenform abgebildet. Dies entspricht der Darstellung, nachdem das in Pro Tools bearbeitete Audio im Architect Plugin aufgenommen wurde. Die Wellenformen der beiden Audiospuren in Pro Tools sowie im Architect Plugin sind identisch. Testweise wurden anschließend in Pro Tools die beiden Audiospuren editiert und Equalizer-Einstellungen verändert. Wie in Abb. 5.9 zu sehen ist, verändert

sich die grafische Darstellungsweise im Architect Plugin bei erneutem Abspielen nicht, die klanglichen Veränderungen sind jedoch hörbar.

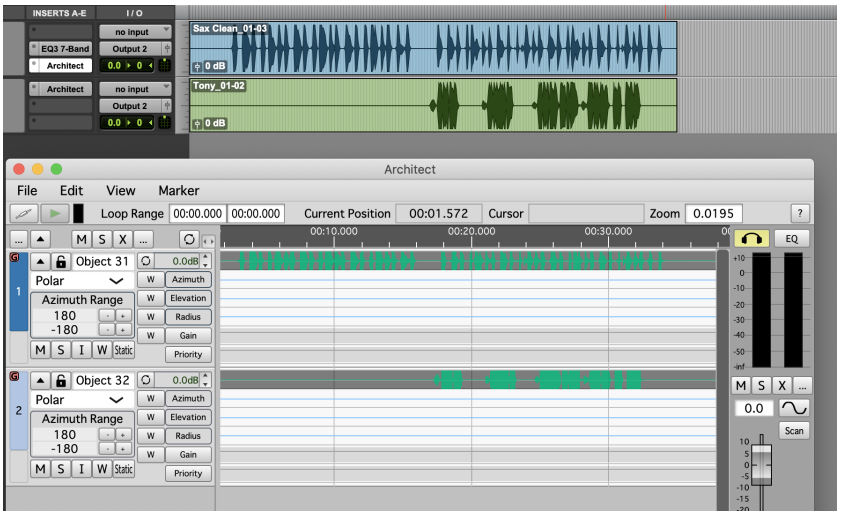


Abbildung 5.8: Korrekte Darstellung des aus Pro Tools (Hintergrund) aufgenommenen Audios im Architect Plugin (Vordergrund).

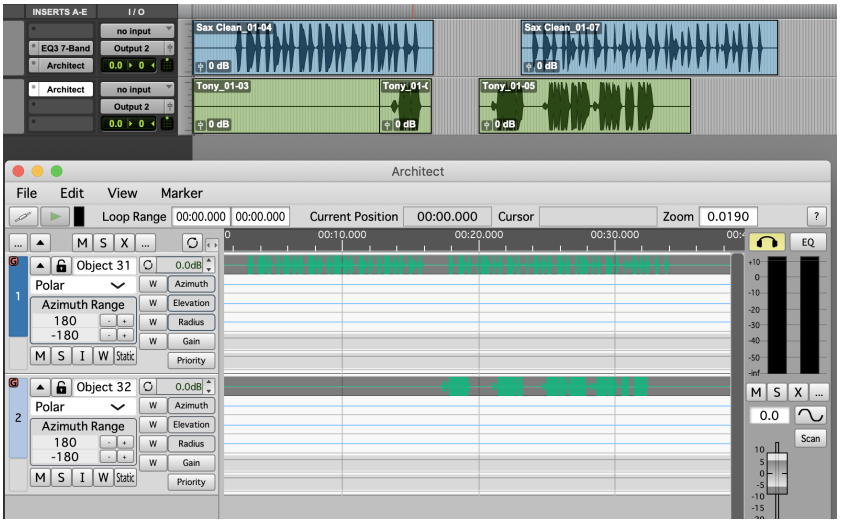


Abbildung 5.9: Inkorrekte Darstellung des aus Pro Tools (Hintergrund) aufgenommenen Audios im Architect Plugin (Vordergrund).

Ebenso wie bei der Dolby Atmos Music Produktion erfolgte anschließend – anhand des zuvor erstellten Konzeptes – die Positionierung der dynamischen und statischen Objekte. Letztere wurden in diesem Fall als Bett eingesetzt und dementsprechend auf feste Positionen im Raum verteilt. Die Automatisierung der dynamischen Objekte erfolgte

in der *POV-Ansicht* des Architects (dargestellt in Abb. 5.10). In Bezug auf Lautheit konnte im Architect ausschließlich das *True Peak Level* gemessen und angepasst werden.

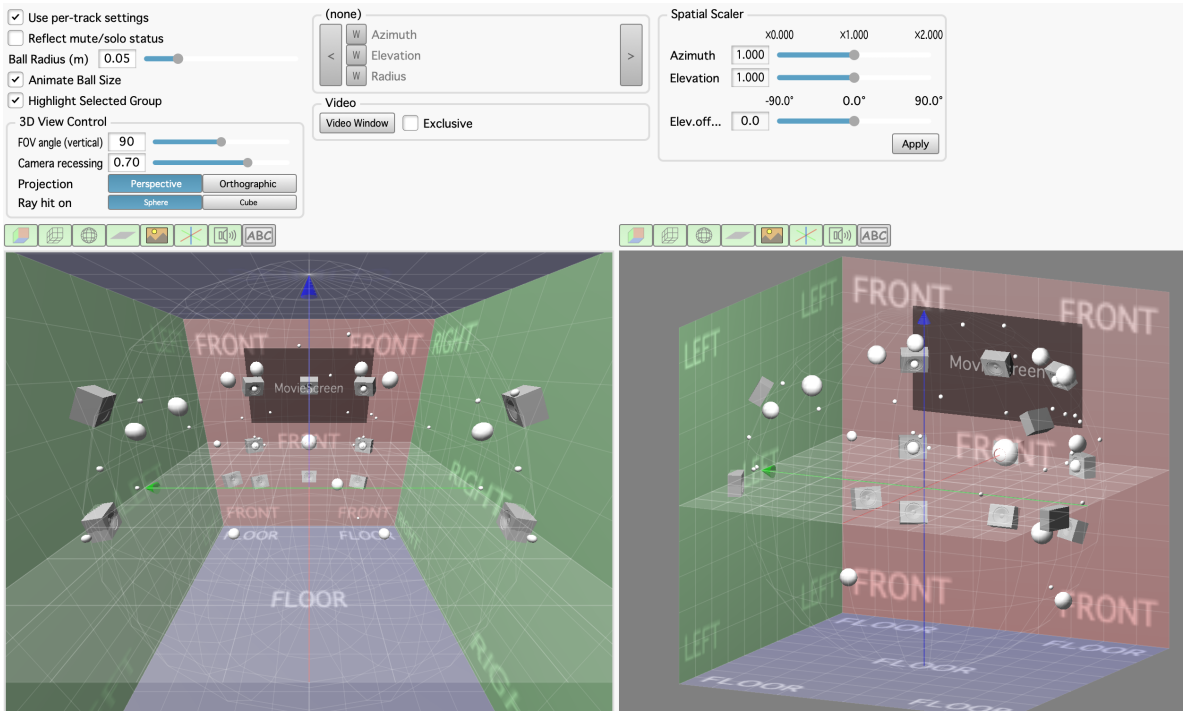


Abbildung 5.10: POV-Ansicht des Architect Plugins.

5.2.3 Kombinierte Abhörsituation

Abschließendes Ziel des praktischen Teils war es, eine kombinierte Abhörsituation zu ermöglichen, in welcher sowohl Dolby Atmos Music als auch 360 Reality Audio in der selben Pro Tools Session abgehört werden können. Hierzu wurden zwei sogenannte *Memory Locations* erstellt – eine für DAM, eine für 360 RA. Diese ermöglichen unter anderem, Spur-Sichtbarkeiten, Spur-Höhe und Gruppen zu speichern, welche dann bei Klick auf die entsprechende Memory Location wiederhergestellt werden. Dies wurde für beide Memory Locations eingestellt – hierbei wurde darauf geachtet, dass alle Spuren mit den entsprechenden Plugins sichtbar sind sowie weitere, für die Mischung relevante Spuren angezeigt werden. Ausschnitte beider Memory Locations sind in Abb. 5.11 und Abb. 5.12 dargestellt.

Außerdem wurden zwei Gruppen erstellt (für DAM und 360 RA), welche eine weitere Vereinfachung des kombinierten Abhör-Workflows mit sich führten: Durch Insertion der Dolby Plugins (Renderer-Send sowie DAM-Panner) auf den *Insert Slots F, G* und Insertion der 360 RA Plugins (Architect) auf *Insert Slot J* konnte in den jeweiligen Gruppen definiert werden, dass bei allen Plugins dieser *Insert Slots* global „Mute“ an- und ausgeschalten werden kann. Somit war es nun möglich, bei der 360 RA Produktion mit einem Klick auf allen Spuren, auf denen das Plugin liegt, gleichzeitig „Mute“ zu aktivieren, sowie beim Wechsel der beiden Workflows alle Plugins des nicht verwendeten Workflows zu deaktivieren.

Da das Audio bei der 360 RA Produktion aus dem Architect und nicht aus Pro Tools abgespielt wird, konnten alle für die DAM Mischung gesetzten Outputs und Bus-Sends bestehen bleiben.

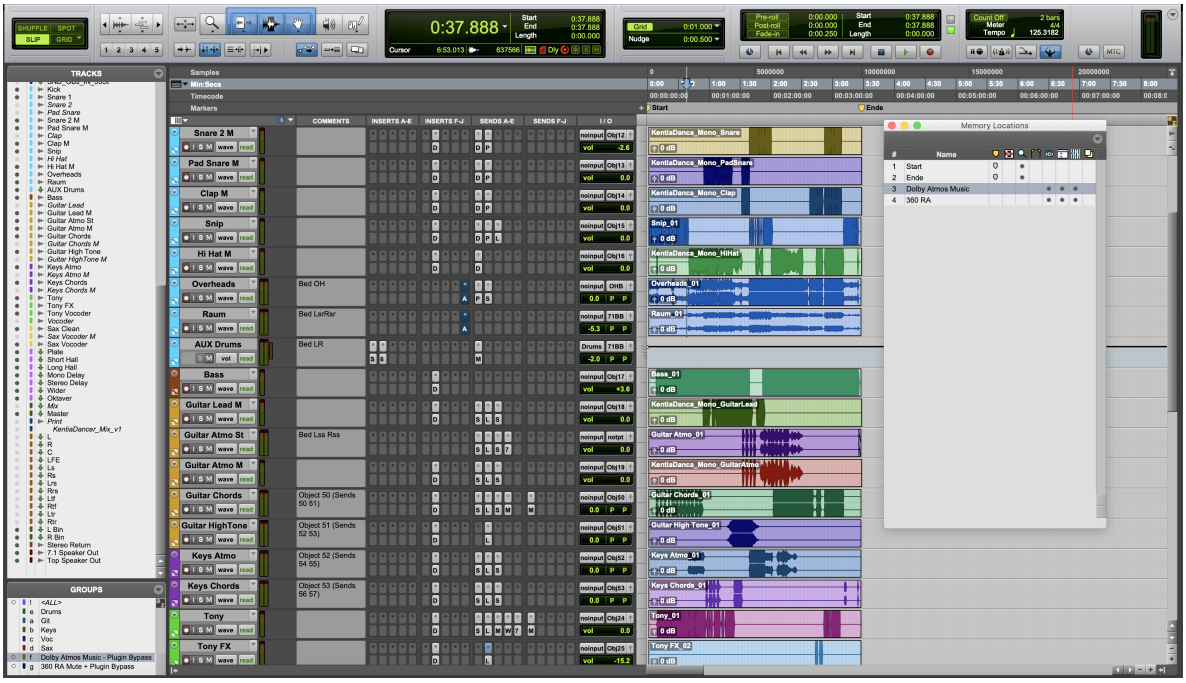


Abbildung 5.11: Kombinierte Pro Tools Session (A). Darstellung der für Dolby Atmos Music Produktion und Monitoring relevanten Spuren sowie der Gruppen (links unten), Memory Locations (rechts) und Plugin-Inserts.

Die Vorgehensweise beim Wechsel des Monitorings von Dolby Atmos Music zu 360 Reality Audio (sowie umgekehrt) ist in Abb. 5.13 dargestellt. Hierbei wird auf die im praktischen Teil erstellte Pro Tools Session referenziert.

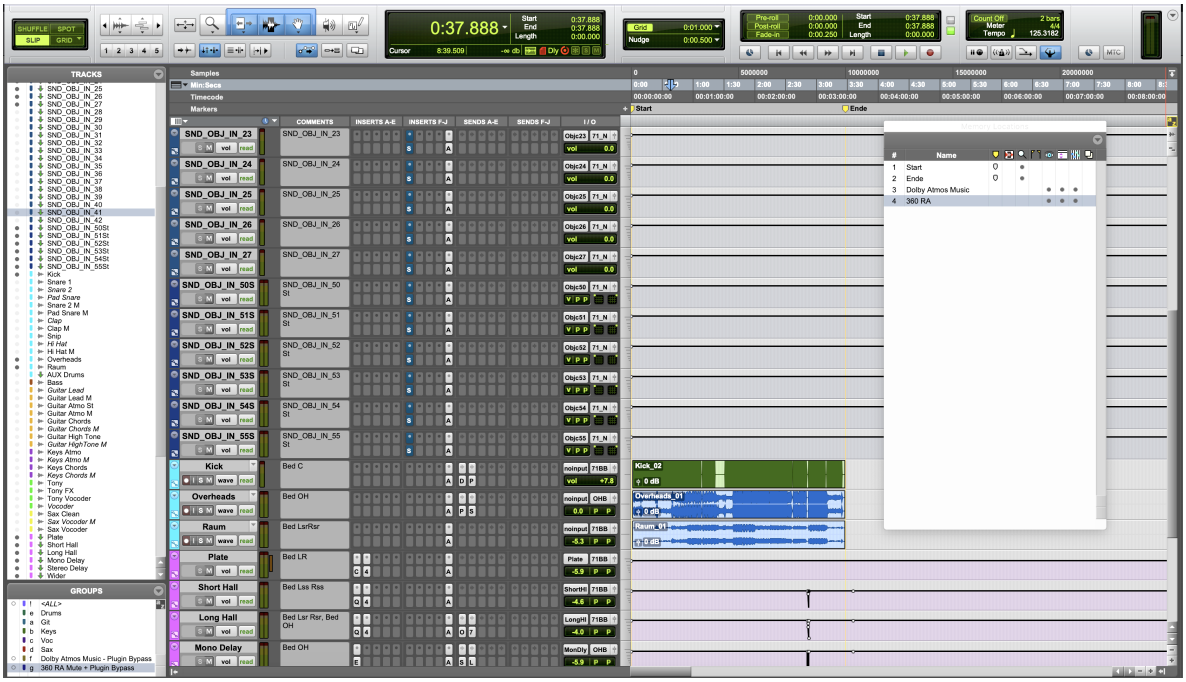


Abbildung 5.12: Kombinierte Pro Tools Session (B). Darstellung der für 360 Reality Audio Produktion und Monitoring relevanten Spuren sowie der Gruppen (links unten), Memory Locations (rechts) und Plugin-Inserts.

Abbildung 5.13: Wechsel zwischen Dolby Atmos Music und 360 Reality Audio Produktions- und Abhörsituation.

5.3 Export und Encodierung

Nach Finalisierung der Dolby Atmos Music Mischung folgte der Export. Hierzu wurde im Dolby Atmos Renderer ein DAMF-Set aufgenommen, indem die komplette Mischung in Pro Tools wiedergegeben sowie die Aufnahmefunktion der Master-Datei im Renderer aktiviert wurde. Nach der Aufnahme des DAMF-Sets konnte schließlich eine BWF/ADM *.wav*-Datei exportiert werden, welche im weiteren Distributions-Workflow an den Encoder weitergereicht wird. Im Rahmen dieser Thesis bestand jedoch kein Zugriff auf einen Encoding-Service für DAM, weshalb der Workflow an dieser Stelle endete.

Für 360 RA erfolgte der Export im Architect Plugin (siehe Abb. 5.14). Die Export-Spezifikationen von Sony besagen, dass die Mischung für die weitere Distribution in drei Levels exportiert werden muss. In diesem Prozess wird die Anzahl aller verwendeter Objekte auf eine fest definierte Anzahl an Objekten heruntergerechnet, welche für alle drei Level variiert. Dies geschieht durch automatische oder manuelle Anpassung der „Pre-Rendering“ Funktion. Bei der automatischen Vorgehensweise („Pre Analyze“) müssen alle Objekte als „Auto“ definiert werden. Der Export der drei Levels erfolgt in Echtzeit, wobei für jedes Level die gesamte Mischung abgespielt werden muss.

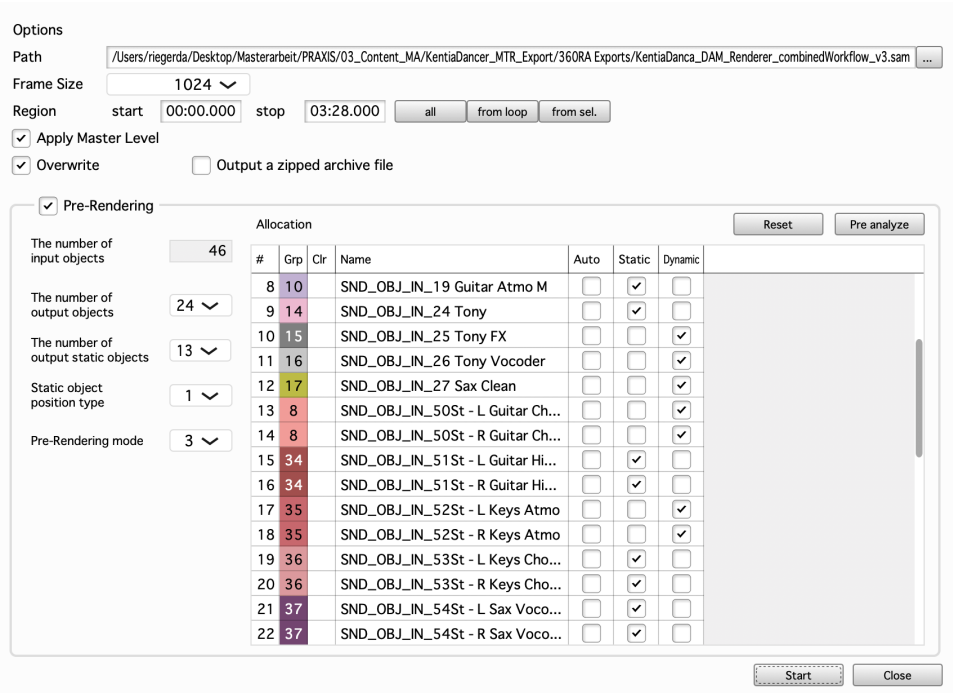


Abbildung 5.14: Export-Fenster des Architect-Plugins.

360 RA Pre-Rendering: „Number of Input Objects“

Dieser Punkt beschreibt die Anzahl der statischen und dynamischen Objekt-Spuren im Architect (Plugin), das heißt alle Spuren, welche mit statischen oder dynamischen Metadaten versehen wurden. Im Rahmen der praktischen Arbeit wurde mit „46 Input Objekten“ gearbeitet, welche im Export-Prozess auf die den jeweiligen Levels entsprechende Objekt-Anzahl heruntergerechnet wurden.

360 RA Pre-Rendering: „Number of Output Objects“

Dies referenziert auf die Anzahl der Objekte, die im jeweiligen Export-Level unterstützt werden. Die Anzahl aller Input Objekte wird während des Export-Vorgangs auf diese Zahl heruntergerechnet. Es müssen für jede Mischung alle drei Level exportiert werden, hierbei gilt: Level 1 = 10 Objekte, Level 2 = 16 Objekte, Level 3 = 24 Objekte.

360 RA Pre-Rendering: „Number of Output Static Objects“

Dies beschreibt die Anzahl der Objekte, die beim Export als statisch definiert werden. Die Anzahl der statischen Output Objekte bestimmt außerdem, welches Lautsprecher-Layout als Referenz zur Positionierung der statischen Objekte verwendet wird. Im Rahmen der praktischen Arbeit wurde manuell das 13ch(Music) Wiedergabe-Layout von Sony eingestellt, dies wird in Abb. 5.14 durch die Zahl „13“ beschrieben. Die Anzahl der statischen Output Objekte (in o.g. Beispiel „10“) subtrahiert von der Anzahl aller Output Objekte (in o.g. Beispiel „24“) definiert, wie viele dynamische Objekte exportiert werden können.

360 RA Pre-Rendering: „Static Object Position Type“

Dieser Punkt bezieht sich auf die „Number of Output Static Objects“ und differenziert zwischen verschiedenen Lautsprecher-Layouts, die die selbe Anzahl an Kanälen haben. So kann die Zahl „10“ beispielsweise ein 5.0+5H Setup oder auch ein 5.1.4 Setup beschreiben, unterschieden wird hierbei durch den „Position Type 0“ oder „Position Type 1“.

360 RA Pre-Rendering: „Pre-Rendering Mode“

Diese Funktion bezieht sich auf die automatische Einstellung, welche die Gesamtzahl an Input Objekten als statisch und dynamisch definiert. Die automatische Anpassung erfolgt mittels der „Pre Analyze“-Funktion, welche Objekte mit hohem Schalldruckpegel

als dynamisch setzt und alle weiteren Objekte so verteilt, dass sie statisch gerendert werden. Der „Pre-Rendering Mode“ gibt hierbei den Algorithmus an, mit welchem die automatische Anpassung erfolgt, d.h. mit welchem Prinzip die Objekte als statisch und dynamisch ausgewählt werden.

5.4 Lautheitsvergleich

Bei der Betrachtung der Produktionsketten sowie der praktischen Anwendung hat sich herausgestellt, dass Dolby Atmos Music über eine integrierte Lautheitsmessung verfügt, 360 Reality Audio hingegen nicht. Letzteres Format nutzt nur einen eingeschränkten Feature-Umfang von MPEG-H – so werden beispielsweise Objekte und Datenkompression unterstützt, Interaktivität, Kanäle und Lautheits-Normalisierung sind allerdings nicht im Funktionsumfang enthalten. Insbesondere letzterer Punkt führt dazu, dass beim Vergleich von Dolby Atmos Music und 360 Reality Audio Mischungen via Musik-Streaming-Diensten wie Tidal ein deutlicher Unterschied hörbar ist. Zu diesem Zweck wurden mittels eines Dolby Atmos fähigen Smartphones verschiedene, zufällig ausgewählte Songs der beiden objektbasierten Technologien via Tidal gestreamt und (bei gleichbleibenden Pegel-Einstellungen) über ein Audio-Interface aufgenommen sowie anschließend normalisiert.

Wie in Abb. 5.15 dargestellt, sind die Lautheits-Unterschiede deutlich sichtbar – auf der linken Seite befinden sich verschiedene 360 Reality Audio Mischungen, auf der rechten Seite Dolby Atmos Music Mischungen.

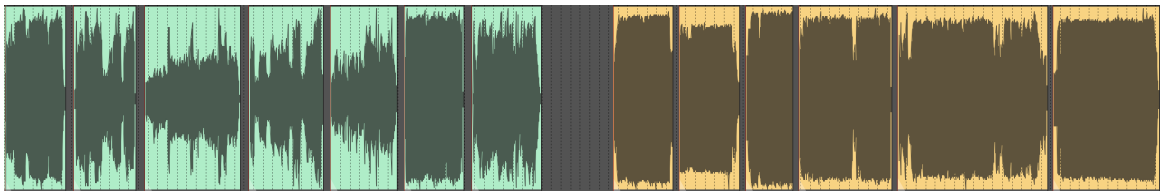


Abbildung 5.15: Vergleich der Lautheit von verschiedenen 360 Reality Audio (links) und Dolby Atmos Music (rechts) Songs bei Tidal.

Während bei DAM nahezu alle aufgenommenen Titel eine ähnliche Lautheit haben, unterscheidet sich diese bei 360 RA teils erheblich. Beispielsweise *American Tale – Paul Simon* (fünfter Song von links) wirkt zumindest grafisch erheblich leiser, die Unterschiede

im Dynamikumfang sind klar zu erkennen. Dies bestätigt sich auch in der durchgeführten Lautheitsmessung. Während bei Dolby Atmos Music, wie in Abb. 5.17 erkennbar, die integrierte Lautheit zwischen –16.2 LUFS und –13.9 LUFS liegt und somit nur geringe Unterschiede aufweist, sind diese bei 360 Reality Audio weitaus erheblicher. Dargestellt in Abb. 5.16 liegt die Lautheit hier zwischen –20.7 LUFS (o.g. Titel *American Tale*) und –12.9 LUFS. Der durchschnittliche Wert des Dynamikumfangs (Range) bei 360 Reality Audio beträgt 9.7 LU, bei Dolby Atmos Music 5.1 LU.

#	Track	Item	Integrated	Range	True peak	Maximum short-term	Maximum momentary
1	Loudness Vergleich 360 RA	02-360	-14.6 LUFS	9.8 LU	0.1 dBTP	-11.2 LUFS	-8.3 LUFS
2	Loudness Vergleich 360 RA	02-360	-16.0 LUFS	13.2 LU	0.2 dBTP	-12.1 LUFS	-10.9 LUFS
3	Loudness Vergleich 360 RA	02-360	-20.3 LUFS	11.0 LU	0.5 dBTP	-14.2 LUFS	-12.0 LUFS
4	Loudness Vergleich 360 RA	02-360	-19.8 LUFS	11.7 LU	0.2 dBTP	-13.8 LUFS	-11.2 LUFS
5	Loudness Vergleich 360 RA	02-360	-20.7 LUFS	8.1 LU	0.1 dBTP	-16.8 LUFS	-12.4 LUFS
6	Loudness Vergleich 360 RA	02-360	-12.9 LUFS	5.4 LU	0.1 dBTP	-9.6 LUFS	-8.6 LUFS
7	Loudness Vergleich 360 RA	02-360	-16.1 LUFS	9.0 LU	0.2 dBTP	-12.8 LUFS	-11.3 LUFS

Abbildung 5.16: Lautheitsmessung zufällig ausgewählter 360 Reality Audio Titel.

#	Track	Item	Integrated	Range	True peak	Maximum short-term	Maximum momentary
1	Loudness Vergleich DAM	02-360	-13.9 LUFS	4.2 LU	0.0 dBTP	-11.7 LUFS	-10.2 LUFS
2	Loudness Vergleich DAM	02-360	-16.2 LUFS	4.5 LU	0.2 dBTP	-14.0 LUFS	-12.7 LUFS
3	Loudness Vergleich DAM	02-360	-14.5 LUFS	5.3 LU	0.1 dBTP	-12.3 LUFS	-10.7 LUFS
4	Loudness Vergleich DAM	02-360	-15.0 LUFS	3.5 LU	0.2 dBTP	-12.9 LUFS	-10.5 LUFS
5	Loudness Vergleich DAM	02-360	-15.1 LUFS	7.4 LU	0.2 dBTP	-11.1 LUFS	-8.9 LUFS
6	Loudness Vergleich DAM	02-360	-15.1 LUFS	5.7 LU	0.0 dBTP	-11.1 LUFS	-10.3 LUFS

Abbildung 5.17: Lautheitsmessung zufällig ausgewählter Dolby Atmos Music Titel.

6 Diskussion

„Object-based audio is often talked about in the same breath as immersive audio, but it is just one tool in the armory, allowing for a flexible representation of channels or mix elements that can be mapped to different spatial ocations and adapted for various replay scenarios.“
(Rumsey, 2015, S. 394)

Die detaillierte Betrachtung der Produktionsketten für Dolby Atmos Music und 360 Reality Audio sowie die praktische Anwendung beider Technologien hat gezeigt, dass objektbasierte Musikproduktion in allen Schritten des Workflows von den zugrundeliegenden technischen Spezifikationen beeinflusst wird. Aktuell bedeutet objektbasierte Musik die Produktion für einen Codec, was in Diskrepanz zur Stereo-Musik steht, da man dort Codec-agnostisch produziert.

Der Vergleich beider Technologien sollte jedoch unter dem Aspekt erfolgen, dass Dolby Atmos Music zwar ein neues Format ist, Dolby Atmos allerdings bereits seit Jahren kommerziell verfügbar ist. Dolby Atmos fähige Decoder sind in einer Vielzahl an Hardware-Geräten (Smartphones, Fernseher, AVRs, Lautsprecher) implementiert, welche bereits vor Veröffentlichung von DAM Inhalten in Konsumentenhaushalten vorhanden waren. Dolby Atmos Software konnte ebenfalls über Jahre etabliert, getestet und optimiert werden – die Produktion von DAM Inhalten war somit (zumindest bis zum Encoding-Schritt) bereits seit längerem möglich, für die Produktion von Dolby Atmos Inhalten sind außerdem zahlreiche Workflow-Tutorial-Videos verfügbar. 360 Reality Audio hingegen ist eine neue Technologie, die erst Ende 2019 veröffentlicht wurde – die Software befindet sich noch im Optimierungs-Prozess und es gibt öffentlich kaum Informationen zur Produktionsweise zu finden. Speziell in der praktischen Anwendung wird ersichtlich, dass (vor allem im Vergleich zur Dolby Atmos Software) keine jahrelangen Erfahrungswerte in der Programmierung der grafischen Benutzeroberfläche vorliegen. Dennoch sollen im Folgenden die Erkenntnisse näher betrachtet und verglichen werden.

6.1 Encodierung und Übertragung

Insbesondere bei der Encodierung zeigt sich ein deutlicher Unterschied zwischen Dolby Atmos Music und 360 Reality Audio. Bei Letzterem wird durch den MPEG-H Codec nur eine Datei übertragen, welche durch die Kombination aus Audio und zugehörigen Metadaten alle Informationen besitzt, um am Endgerät des Anwenders auf das entsprechende Wiedergabesystem gerendert zu werden. Somit kann bei dieser Technologie ein grundlegender Mehrwert objektbasierten Audios ausgespielt werden: In der Produktion wird eine Mischung erstellt, welche durch eine Distributionsdatei übertragen wird und beim Konsumenten auf vielen möglichen Wiedergabesystemen abgespielt werden kann. Bei DAM hingegen werden bei der Encodierung zwei unterschiedliche Dateien erzeugt – AC-4 IMS (Zweikanal-Datei zur binauralen Kopfhörerwiedergabe) und DD+JOC (Mehrkanal-Datei + Objektmetadaten zur Lautsprecherwiedergabe), wobei nur letztere Datei die Möglichkeit bietet, auf verschiedene Wiedergabesysteme gerendert zu werden. Durch die Distribution von zwei Dateien ist die Praktikabilität geringer und durch die Separation der Kopfhörer- und Lautsprecherwiedergabe geht ein mehrwertbringender Aspekt objektbasierten Audios verloren.

Somit zeigt sich, dass bei genauerer Betrachtung nur 360 Reality Audio als durchgehend objektbasierte Technologie bezeichnet werden kann. Hier sind sowohl die Vorgänge der Produktion als auch Wiedergabe rein objektbasiert. Bei Dolby Atmos Music hingegen wird bereits während der Produktion mit einer Mischung aus Kanälen und Objekten gearbeitet und bei der Encodierung zweier unterschiedlicher Dateien der objektbasierte Prozess verlassen. Bei der Wiedergabe von Dolby Atmos Music über Kopfhörer kann aufgrund der AC-4 IMS Zweikanal-Datei nicht von einer objektbasierten Wiedergabe gesprochen werden – durch die Vor-Binauralisierung erreichen den Endkonsumenten keine Merkmale von objektbasiertem Audio. So ist beispielsweise eine Personalisierung durch Analyse der Ohrform nicht möglich, was bei 360 Reality Audio durch die objektbasierte Übertragungsweise implementiert wurde: Da die Binauralisierung erst in der Wiedergabe-App erfolgt, können durch Verwendung der *Sony Headphone Connect App* noch personalisierte Einstellungen erfolgen.

Vor allen Dingen die Vor-Binauralisierung und Übertragung einer Zweikanal-Datei zur Kopfhörer-Wiedergabe entspricht der Distribution von Kunstkopfproduktionen aus den 1970er Jahren. Eine kommerzielle Vermarktung eben dieser Übertragungsweise 50 Jahre

später als „the next era of music“ (Dolby Laboratories, 2020e) ist fraglich – besonders unter dem Aspekt, dass die Vorteile und technischen Neuerungen von objektbasiertem Audio insbesondere im flexiblen Wiedergabe-Rendering ausgespielt werden könnten.

6.2 Wiedergabemöglichkeiten

Auch wenn der Begriff „objektbasiert“ bei der Übertragung und Wiedergabe von Dolby Atmos Music nicht gänzlich zutreffend ist, so sind die Wiedergabemöglichkeiten deutlich breiter gesät als bei 360 Reality Audio: Während die Auswahl an Streaming-Diensten bei Letzterem etwas größer ist, kann die Wiedergabe von DAM über Kopfhörer, Lautsprechersysteme und Soundbars erfolgen. Das Decoding erfolgt bei Lautsprecherwiedergabe in den hardwareseitig verbauten Decodern. Bei 360 RA passiert das Decoding bei der Wiedergabe über Kopfhörer softwareseitig in den Musik-Streaming-Apps, bei der Wiedergabe auf dem Amazon Echo Studio hardwareseitig im Gerät. Um die Inhalte jedoch auf weiteren Geräten (Android Smart-TVs, Soundbars, AVRs und Lautsprechern) wiederzugeben, müssen in diesen – je nach Gerätetyp – sowohl MPEG-H Bitstream Pass-through[1]-Möglichkeiten als auch MPEG-H Decoder und zur Übertragung eines 16-Kanal-Datenstreamss HDMI-eARC-Verbindungen implementiert werden. Hier zeigt sich, dass die technischen Voraussetzungen zur breiten Wiedergabe von MPEG-H-basierten Inhalten theoretisch vorhanden sind, allerdings noch nicht weitreichend in die Geräte eingebaut wurden.

6.3 Erkenntnisse der praktischen Anwendung

6.3.1 Unterschied zum Stereo-Workflow

In der praktischen Anwendung wurde auf Basis einer bereits vorhandenen Stereo-Mischung ein kombinierter Workflow zur Produktion von Dolby Atmos Music und 360 Reality Audio entwickelt und durchgeführt. Es hat sich gezeigt, dass sich der objektbasierte Produktions- und Distributions-Workflow in Aufnahme, Editing und einem Teil der Mischung dem in Abschnitt 2.1 dargestellten Stereo-Workflow gleichen kann. Eine Unterscheidung erfolgt ab dem Punkt, ab dem die objektbasierte Software integriert wird (siehe Abb. 6.1).

[1]Pass-through bedeutet in diesem Fall das Durchleiten (statt Decodieren) von HDMI-Signalen.

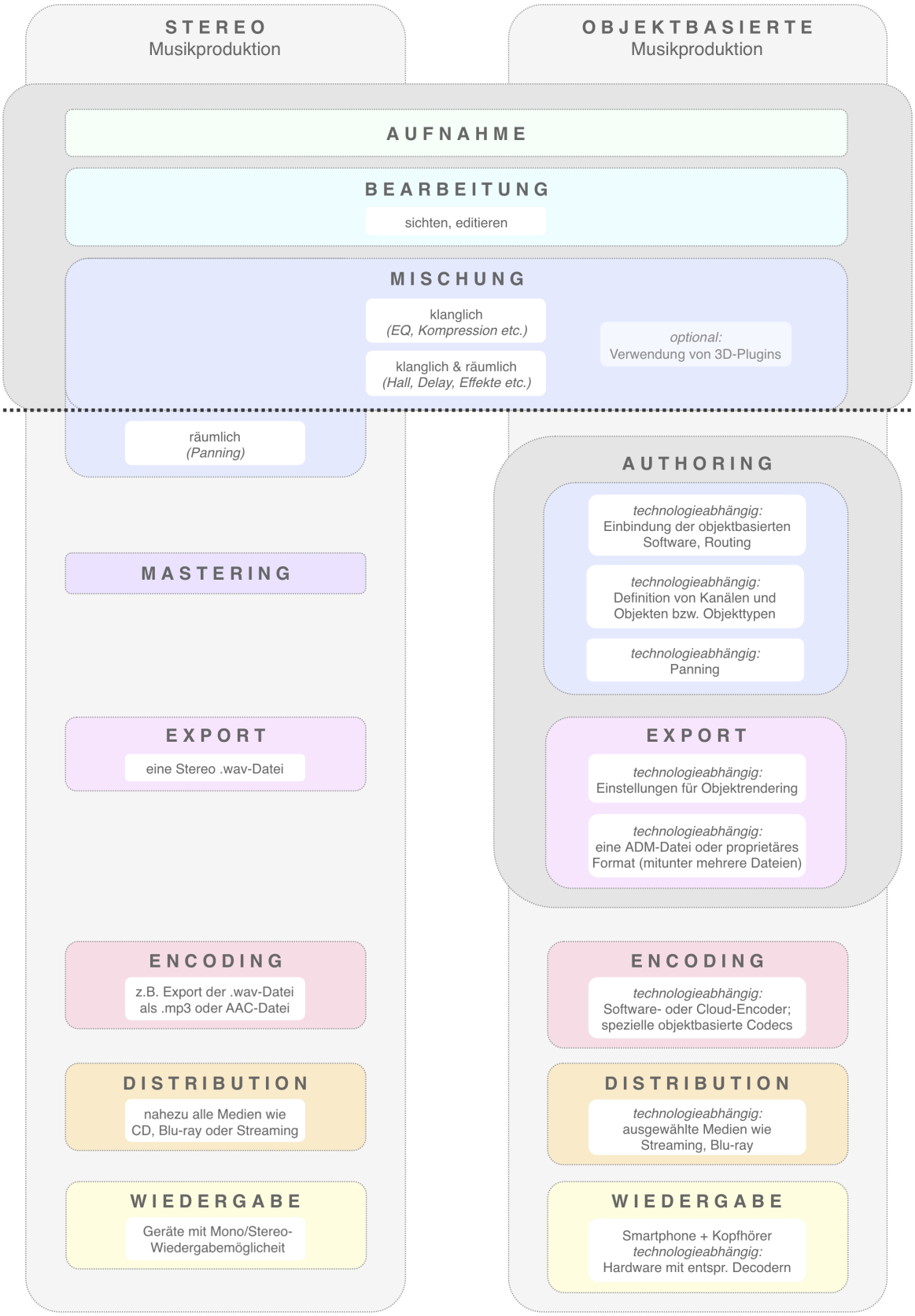


Abbildung 6.1: Stereo-Musikproduktion (links) vs. objektbasierte Musikproduktion (rechts): Vergleichende Darstellung der Produktionsschritte.

Die Relevanz der Unterscheidung zwischen den Audio-Begrifflichkeiten „immersiv“ bzw. „3D“ und „objektbasiert“ gilt es an dieser Stelle abermals herauszustellen. Während die Unterscheidung des Workflows bei objektbasierten Produktionen erst mit Integration der entsprechenden Software erfolgt, spielt ein immersiver Produktions-Workflow hingegen bereits früher eine Rolle: Während der Aufnahme kann unter Berücksichtigung von 3D-Aufnahmetechniken eine andere Räumlichkeit geschaffen werden, welche durch den Einsatz spezieller 3D-Plugins für Hall, Delay, Multibandkompression und Effekte verstärkt herausgearbeitet werden kann. Da nicht jede immersive Produktion objektbasiert, jedoch die meisten objektbasierten Produktionen immersiv sind, wird im Folgenden von letzterer Annahme ausgegangen.

Die zwei auffälligsten Abweichungen ergeben sich durch den neuen Schritt des Authorings, welcher sich mit der Erstellung und Überprüfung der Metadaten sowie Objekt-Rendering-Einstellungen beschäftigt, sowie der deutlich anderen Ausgangslage für Mastering-Prozesse. Insbesondere letzterer Punkt eröffnet (wie in Abschnitt 3.5 dargestellt) neue technische Anforderungen an mögliche Software-Entwicklungen. Im Vergleich beider Produktionsarten ist außerdem auffällig, wie sehr die Abweichungen vom herkömmlichen Produktions-Worklfow von der jeweiligen zugrundeliegenden objektbasierten Technologie abhängen. Ein Beispiel hierfür ist das Encoding und die daraus resultierende Wiedergabe: Wird bei Stereo-Produktionen meist eine .wav-Datei exportiert und beispielsweise als AAC- oder mp3-Datei encodiert (welche basierend auf den weitreichend verbauten Decodern nahezu überall wiedergegeben werden kann), so werden bei objektbasierten Produktionen entweder eine ADM-Datei mit einem speziellen, technologieabhängigen Profil oder ein proprietäres Format exportiert, welche dann im Ecoding-Prozess in abermals unterschiedliche Formate encodiert und auf unterschiedlichen Geräten wiedergegeben werden können. Dies zeigt eines der Probleme objektbasierten Audios auf: In verschiedenen Schritten des Workflows wird mit speziellen, nicht kompatiblen Formaten gearbeitet. Der Entscheidungsprozess für eines (und somit gegen alle anderen) der Formate liegt im objektbasierten Workflow deutlich früher als in der herkömmlichen Stereo oder 5.1 Surround Produktionskette.

6.3.2 Stereo-Ausgangsmaterial

Bei der Analyse und Ersteinschätzung des Materials hat sich gezeigt, dass die Anlieferung eines DAW-Projektes – sofern es sich hierbei um eine von objektbasierter Software

unterstützte DAW handelt – eine Möglichkeit bietet, gestalterische Entscheidungen der Stereo-Mischung beizubehalten. Im Falle der kombinierten Produktion von Dolby Atmos Music und 360 Reality Audio stellt die Verwendung eines Dolby Atmos Production Templates eine zeitsparende Option dar, um direkt mit dem gelieferten Material in die objektbasierte Produktion einzusteigen. Die Integration der Session-Daten der Stereo-Mischung in das Dolby Atmos Template erwies sich als unkompliziert, ein Verständnis für die Dolby Atmos Routing- und Bus-Strukturen mittels Dolby Atmos Renderer Send- und Return-Plugins zu entwickeln als deutlich komplexer. Sofern für die immersive Produktion dafür optimierte Plugins verwendet werden sollen (3D-Hall, 3D-Multibandkompressor, o.Ä.), so bietet sich die Verwendung der gelieferten Stems bzw. Audiospuren an, welche direkt im Dolby Atmos Template auf den entsprechenden Bett- oder Objektspuren importiert werden können.

6.3.3 Objektbasierte Produktion

Durch die Möglichkeit, die Produktions-Software beider Technologien in die DAW zu integrieren, kann auf bereits vorhandenen Stereo- oder Surround-Mischungen aufgebaut werden. Während der praktischen Mischung hat sich allerdings herausgestellt, dass selbst bei einer kombinierten Vorgehensweise, welche den Produktions-Workflow für zwei Technologien verbindet, die kreativen und technischen Möglichkeiten von der jeweiligen Software beeinflusst werden. Bereits die Frage, welche DAW verwendet wird, entscheidet sich durch die technischen Vorgaben der objektbasierten Software. In der praktischen Anwendung hat sich außerdem gezeigt, dass sich für einen kombinierten Workflow innerhalb einer Produktionsumgebung nur die DAW Pro Tools anbietet. Da es sich beim Architect Plugin um ein *AAX-Plugin* handelt, kann dieses ausschließlich in Pro Tools insertiert werden. Da die Software zur Produktion von DAM vielseitiger einsetzbar ist, hängt somit eine Ausweitung der verwendbaren DAWs von möglichen Erweiterungen des Architect Plugins ab.

Insbesondere in Hinblick auf die Verwendung von Aux- oder Audio-Spuren mit größeren Kanal-Konfigurationen als 5.1 (beispielsweise für die Erstellung von Mehrkanal Bett- oder Effektspuren) bieten DAWs wie Nuendo oder Reaper eine größere Auswahlmöglichkeit als Pro Tools (siehe Abb. 6.2 und Abb. 6.3).

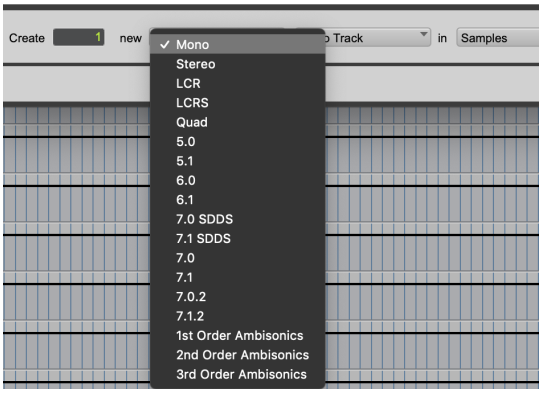
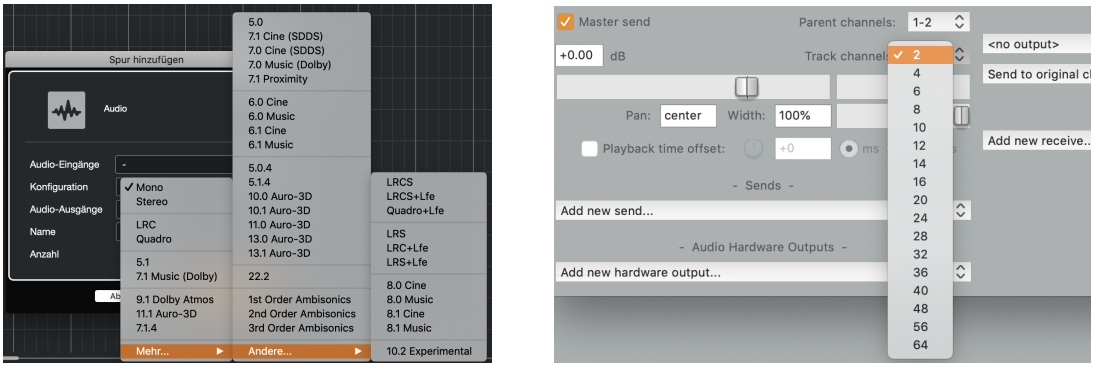


Abbildung 6.2: Mögliche Spur-Konfigurationen in Pro Tools (Version 2020.3.0).



(a) (b)

Abbildung 6.3: Mögliche Spur-Konfigurationen in Nuendo (a; Version 10.2.20) und Reaper (b; Version 6.02).

Software

Gegeben durch die komplexe Dolby Atmos Produktionsumgebung (Dolby Atmos Renderer als Standalone-Applikation, Send- und Return-Plugins des Renderers, Dolby Atmos Panner) und damit gegebene Routing-Struktur, gibt diese das Vorgehen bei der Mischung an und gewohnte Workflows müssen daran angepasst werden. Im Gegenzug dazu ermöglicht das Architect-Plugin eher, weiter an persönlich etablierten Vorgehensweisen festzuhalten, indem ausschließlich ein Plugin auf allen Spuren insertiert wird, die als Objekt definiert werden sollen. Die Eigenheit, dass die Audiospuren der DAW im Plugin aufgenommen werden und somit für eine korrekte Darstellung nach jeder klanglichen Änderung die Wellenform neu erstellt werden muss, hindert jedoch erheblich an der Entwicklung eines intuitiven Workflows – besonders, da eine Mischung iterativ erfolgt und Änderungen an einer Stelle mögliche Auswirkungen und damit verbundene Anpassungen an anderer Stelle bewirken können.

Des Weiteren hängt das finale, klangliche Ergebnis nicht nur von Faktoren ab, die vom Produzenten beeinflussbar sind (Aufnahme des Materials, Einstellungen von EQ-, Kompressions-, Effekt- und Hall-Plugins etc.) sondern in erheblicher Weise von internen, technischen Prozessen des Wiedergabe- bzw. Binaural-Renderings. Der klangliche Unterschied der Dolby Atmos Music und 360 Reality Audio Binauralisierung ist klar hörbar, und lässt sich auch in Abb. 6.4 erkennen.

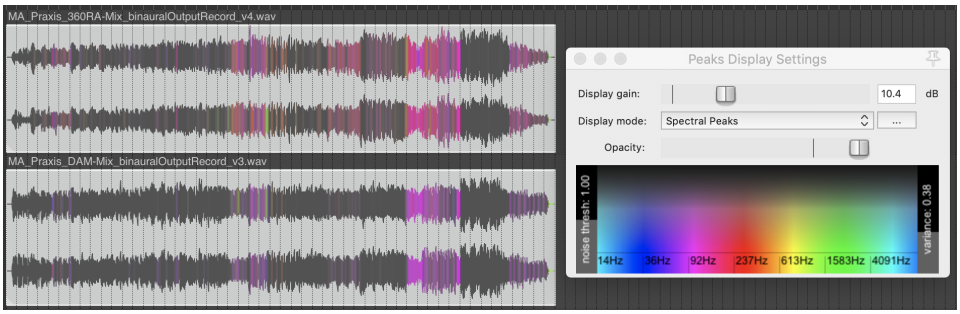


Abbildung 6.4: Darstellung spektraler Peaks der binauralen Ausgangssignale von Dolby Atmos Music (unten) und 360 Reality Audio (oben).

Diesen klanglichen Verfärbungen der binauralen Wiedergabe produktionsseitig entgegenzuwirken stellt eine Herausforderung dar, da die Wahrnehmung der binauralen Mischung bei Konsumenten variiert. Es gilt anzumerken, dass sich dies nicht nur in objektbasierten sondern allen Arten der binauralen Produktion auswirkt. Es gilt weiterhin festzuhalten, dass es bei 360 RA durch die objektbasierte Übertragung theoretisch möglich ist, einen anderen Binaural-Renderer auszuwählen oder diesen an spezifische Situationen anzupassen. Die Vor-Binauralisierung und damit verbundene Zweikanal-Übertragung von Dolby Atmos Music via AC-4 IMS ermöglicht dies (zumindest nach jetzigem Wissensstand) nicht.

Die Tatsache, dass sich Sony bei der Entwicklung von 360 Reality Audio gegen eine integrierte Lautheitsmessung im Architect entschieden hat, ist zwar durch die Wahl des zugrundeliegende MPEG-H-Profils mit eingeschränktem Funktionsumfang zu erklären, hat sich jedoch in der Anwendung als nachteilig erwiesen. Insbesondere durch die Vorschrift aktueller Lautheits-Normen, die sich im Musik-Streaming etabliert haben, sowie den direkten Vergleich mit Lautheits-normalisierten DAM Produktionen zeigt sich, dass das Thema Lautheit auch bei objektbasierter Musikproduktion relevant ist und im Workflow berücksichtigt werden muss. Bei der praktischen Anwendung hat sich für 360 RA keine Möglichkeit ergeben, die Lautheit der Mischung zu messen – durch die kombinierte Abhörsituation konnten zwar beide Mischungen stichpunktartig in der

Lautstärke verglichen werden, die Lautheit wurde jedoch nur für Dolby Atmos Music gemessen und angepasst. Der im Rahmen dieser Arbeit durchgeführte Lautheitsvergleich zufällig ausgewählter DAM und 360 RA Titel des Musik-Streaming-Dienstes Tidal hat gezeigt, dass die Unterschiede beider Produktionsweisen deutlich sichtbar und hörbar sind und bei 360 RA Mischungen eine ständige manuelle Lautstärke-Anpassung bei der Wiedergabe nötig ist.

Einbindung von Plugins

Bei der klanglichen Mischung und der damit verbundenen Verwendung von Plugins hat sich herausgestellt, dass in diesem Fall herkömmliche Produktionsprozesse annähernd beibehalten werden können. Abgesehen von den beiden Schritten Panning (hierzu bedarf es spezieller, objektbasierter Software) und Mastering (hierzu bedarf es der Entwicklung und Erprobung neuer Software oder Vorgehensweisen) können gestalterische Anpassungen auch mittels Mono- oder Stereo-Plugins erfolgen. In der praktischen Anwendung hat sich gezeigt, dass sich die Verwendung von Mono- bzw. Stereo-Plugins zum einen deshalb anbieten kann, da es sich bei den Audioobjekten um Mono- oder Stereo-Spuren handelt. Zum anderen wirken sich die von herkömmlichen Workflows abweichenden objektbasierten Produktions-Schritte insbesondere in der Positionierung der Objekte sowie im Export-, Encodierungs- und Rendering-Prozess aus – klangliche Bearbeitungsschritte, die zuvor erfolgen, bleiben annähernd gleich zu denen herkömmlicher Produktionen.

Auch wenn „objektbasiert“ nicht automatisch mit „3D-Audio“ gleichzusetzen ist, so spielt die dreidimensionale Positionierung der Audioobjekte meist eine große Rolle. Hier hat sich gezeigt, dass zur Erschaffung einer 3D-Räumlichkeit ebenfalls – mit kleineren Einschränkungen – mit Stereo-Plugins gearbeitet werden kann, gerade bei der Anwendung von Hall oder Multiband-Kompression würden spezielle (objektbasierte) 3D-Plugins den Workflow jedoch vereinfachen und erweiterte Möglichkeiten bieten.

Export

In Bezug auf den Export-Vorgang beider Technologien lässt sich ein Unterschied feststellen: Während bei Dolby Atmos Music mit der BWF/ADM-Datei bereits Audio und Metadaten kombiniert werden, erfolgt dies bei 360 RA mit Metadaten- und Audio-Dateien noch separat. Dies könnte sich auf mögliche Mastering-Ansätze auswirken: Während die

ADM-Datei theoretisch noch Möglichkeiten bietet, zukünftige objektbasierte Mastering-Techniken an dieser Stelle in den Workflow zu integrieren (sofern das aktuell anderweitig ebenfalls nicht-kompatible Atmos-ADM-Profil dies zulässt), so sind die 360 RA Export-Dateien ausschließlich von Sony's Software lesbar. In diesem Falle kann das Mastering nur vor dem Export erfolgen.

Die Vorgehensweise beim Export für 360 RA ist durch die drei verschiedenen Levels mit den jeweiligen Pre-Rendering-Einstellungen deutlich komplexer als der BWF/ADM-Export für DAM. Bei 360 Reality Audio wird dem Produzenten im Exportvorgang ein Schritt überlassen, der vollkommen auf den technischen Prozessen im Hintergrund basiert – somit werden Entscheidungsmöglichkeiten gegeben, bei denen nicht klar ersichtlich ist, welche Auswirkungen diese auf das Endprodukt haben. Die manuelle Anpassung der einzelnen Pre-Rendering-Funktionen ist zum einen komplex, zum anderen wird durch die mitgelieferte Dokumentation nicht eindeutig spezifiziert, welche Auswirkungen alle Kombinationen aufeinander haben können. Es wird beschrieben, dass die automatische Einstellung mittels der Pre-Analyze Funktion in den meisten Fällen ausreicht – hier wird jedoch nicht klar, in welchen Fällen dem nicht so ist. Außerdem muss für die automatischen Export-Einstellungen für jedes Level die Pre-Analyze Funktion durchlaufen werden, welche in Echtzeit ausgeführt wird. Gegeben mit der Tatsache, dass auch der Export für alle Levels in Echtzeit geschieht, kann dies bei etwas längeren Produktionen zeitintensiv sein.

Der Export für DAM ist deutlich weniger komplex – hier muss nur im Renderer ein DAMF-Set aufgenommen und dieses schließlich als BWF/ADM-Datei exportiert werden. Während hier keine Einstellungen vorgenommen werden müssen, die nicht-ersichtliche technisch-interne Prozesse betreffen, so erfolgt auch in diesem Fall der gesamte Export-Vorgang in Echtzeit.

6.3.4 Objekte in der Musikproduktion

Die allgemeinen Merkmale objektbasierten Audios wurden in den vorhergehenden Kapiteln detailliert betrachtet. Die Bedeutung der Objekte speziell in der Musikproduktion ist mit der Produktionsart und dem Musik-Genre verknüpft, da sich dieses auf den objektbasierten Produktions-Workflow ähnlich auswirkt wie auf den einer Stereo-Produktion. Zum einen gilt zu unterscheiden zwischen Live-Produktionen und Studio-Produktionen. Während bei einer Jazz- oder Klassikmischung (sowie bei Live-Produktionen) dem Klang-

körper im Raum (und somit der Aufnahmetechnik) eine besondere Bedeutung zukommt, so spielt bei einer elektronischen (sowie Studio-basierten) Produktion die Gestaltung und Kombination aufgenommener und synthetisch erzeugter Klangelemente eine größere Rolle. Insbesondere bei der Aufnahme eines Klangkörpers im Raum kommt es zum Übersprechen zwischen einzelnen Instrumenten, was in der Mischung berücksichtigt werden muss (Delay Compensation), und bei der Verwendung von Instrumenten als einzelne Objekte zu Problemen führen kann.

In Bezug auf objektbasierte Produktionsweisen bedeutet dies, dass sich der Fokus je nach Genre verschieben kann: Bei einer Klassik-Produktion bietet es sich an, dass insbesondere eine präzise Lokalisation einzelner Objekte sowie der Aspekt der Immersion und Räumlichkeit in den Vordergrund rückt: durch die Möglichkeit, den gesamten Raum einschließlich der Höhendimension abzubilden, kann die Akustik verschiedener Konzertsäle dargestellt werden. In Kombination mit dem flexiblen Wiedergabe-Rendering (beispielsweise der Binauralisierung) bedeutet dies, dass dem Hörer realistischere Hörerlebnisse simuliert werden können. Mögliche Nutzer-Interaktivität bei der Wiedergabe (beispielsweise durch Auswahlmöglichkeiten der Hörerposition) kann ebenfalls dazu beitragen, diesbezüglich wird an dieser Stelle auf Mäsel, Simon, Choi, Schulz & Silzle (2019) verwiesen. Bei „abstrakten“ (beispielsweise elektronischen) Produktionen hingegen können die zumeist einzeln vorliegenden, aufgenommenen oder synthetisch erzeugten Klangelemente als bewegte Objekte in Szene gesetzt und somit eine atmosphärische Klanglandschaft erzeugt werden. Bei elektronischer Musik bietet es sich außerdem besonders an, eine dreidimensionale Räumlichkeit durch kreatives (in Maßen eingesetztes) Objekt-Panning zu erschaffen, da hier – im Gegensatz zu klassischer Musik – nicht auf eine korrekte und realistische räumliche Positionierung und Reproduzierung geachtet werden muss.

Verschiedene Genres können somit von unterschiedlichen Aspekten objektbasierten Audios profitieren – wobei der Fokus stets auf dem Hörer liegen sollte. Insbesondere im Klassikbereich existieren traditionell geprägte und etablierte Hörgewohnheiten, welche auch bei objektbasierten immersiven Produktionen zumindest im Ansatz bedacht werden sollten. Techniken, die sich bei Stereo-Produktionen bewährt haben, gelten auch für objektbasierte immersive Mischungen – 3D-Audio ist eine Erweiterung von Stereo, etablierte Produktions-Workflows sind somit weiterhin relevant und können durch neue Produktionstechniken und Möglichkeiten angereichert werden – ein Beispiel hierfür sind die immersiven Klassik-Produktionen von Morten Lindbergh.

Abschließend betrachtet spielt bei objektbasierter Musikproduktion zum einen besonders der immersive Aspekt eine Rolle, welcher gleichzeitig auch eine verstärkte Emotionalität mit sich führen kann, siehe Hahn (2018). Zum anderen bewirkt das flexible Wiedergabe-Rendering, dass erstellte Produktionen vom Konsumenten über Kopfhörer, Soundbars oder Lautsprechersysteme gehört werden können, Personalisierungsmöglichkeiten gegeben sind und somit ein Mehrwert gegenüber rein binaural oder kanalbasiert produzierten immersiven Mischungen besteht.

6.4 Konvertierungsmöglichkeit der Objekt-Metadaten

Eine alternative Möglichkeit zur Produktion von Dolby Atmos Music und 360 Reality Audio – durch Übernahme von Produktionsentscheidungen des einen Workflows im anderen – ist durch Konvertierung der Objekt-Metadaten möglich. Zunächst kann die in Abschnitt 3.2 beschriebene DAM Produktionskette durchlaufen werden und eine ADM-Datei exportiert und in weiteren Verarbeitungsschritten encodiert und übertragen werden. Da es sich beim Audio Definition Model wie in Abschnitt 2.4 beschrieben um einen offenen Standard handelt, können die beinhalteten Objekt-Metadaten offengelegt werden. Die exportierte DAM ADM-Datei kann somit in einem weiteren Schritt automatisiert in eine von 360 Reality Audio interpretierbare Version konvertiert werden, was ferner eine Modifikationsmöglichkeit im Architect bietet und weiterführend im Sony eigenen Dateiformat (Metadaten + Audiodateien) exportiert und in den Encodierungs- und Distributionsprozess für 360 Reality Audio gereicht werden kann. In Anbetracht dessen, dass diese beiden objektbasierten Technologien aktuell kommerziell verfügbar sind, kann dieser Ansatz der Konvertierung den Workflow zur Produktion von Audioinhalten für beide Distributionsformen deutlich erleichtern. Somit müssen nicht von Grund auf zwei komplette Produktionsprozesse durchlaufen werden, sondern es kann ein Großteil der DAM Mischung für 360 RA übernommen werden. Dies entspricht einem wichtigen Schritt hin zu möglichst formatagnostischen Produktions-Workflows, besonders wenn Konvertierungsmöglichkeiten in beide Richtungen ermöglicht werden.

7 Fazit

„While a number of LP records were issued in various ‚quad' formats, the approach failed to capture a sufficiently large part of the consumer imagination to succeed. It seemed that people were unwilling to install the additional loudspeakers required, and there were too many alternative forms of quad encoding for a clear standard to emerge."
(Rumsey, 2017, S. 184)

Durch die vielfältigen Musik-Distributionsarten (Schallplatte, CD, Blu-ray, digitale Portale wie iTunes oder Soundcloud, Streaming-Dienste) ergeben sich insbesondere im Hinblick auf mögliche formatagnostische Produktionsweisen neue Anforderungen. Ein Aspekt ist die Produktionsart – aktuell wird neben herkömmlichen Stereo-Mischungen auch kanalbasiert, szenenbasiert und objektbasiert, immersiv sowie rein binaural produziert. Diese Vielzahl an Produktionsweisen, die auf unterschiedlichen Codecs basieren und verschiedene Endformate sowie Wiedergabeanforderungen mit sich bringen, sorgen für Unübersichtlichkeit und hindern möglicherweise daran, dass Produktions-Ressourcen aufgewendet werden. Ähnlich wie bei der *Quadrophonie* in den 1960er–1980er Jahren bergen die verschiedenen Formate und Codecs die Gefahr, dass sich nichts davon durchsetzen wird. Dennoch ist hier ein Unterschied festzustellen: Während beim Quad-Format das Problem eher auf der Konsumentenseite lag (kein Kauf der Lautsprecher für die Quad-Wiedergabe), so liegen bei objektbasierten Inhalten die Probleme vielmehr auf Seiten der Produzenten und Technologie an sich.

Aktuell bedeutet die Produktion objektbasierter Musik eine Produktion für einen spezifischen Codec – dies steht in Diskrepanz zur Stereo-Musikproduktion, welche Codec-agnostisch erfolgt. Dies hat sich durch eine detaillierte Betrachtung aktueller objektbasierter Produktionsketten – am Beispiel von Dolby Atmos Music und 360 Reality Audio – sowie der Entwicklung und Anwendung eines kombinierten Produktions-Workflows für diese beiden Technologien gezeigt. Zum einen werden herkömmliche Produktions-Workflows aufgegriffen und weiterentwickelt. Zum anderen lassen sich beide objektbasierte

Technologien zwar in herkömmliche Produktionsketten integrieren, weichen jedoch ab dem Schritt der Nutzung der objektbasierten Software bis hin zur Wiedergabe von etablierten Stereo-Produktionsprozessen ab. Der Vergleich dieser zwei objektbasierten Formate hat gezeigt, dass bereits hier große Unterschiede insbesondere im Export-, Encoding-, Distributions- und Wiedergabeprozess herrschen. Damit sich objektbasierte immersive Produktionen behaupten können, bedarf es einer Entwicklung von einheitlicheren Vorgängen in allen Bereichen der Produktionskette. Sofern sich nicht ein Format durchsetzen wird, so muss zumindest der Aspekt der kombinierten Produktionsweise und Konvertierung verschiedener Formate in den Vordergrund rücken: Für ein Album die kompletten Produktionsketten für mehrere Formate zu durchlaufen ist zeit- und kostenintensiv. Hierfür können der im Rahmen dieser Arbeit betrachtete kombinierte Produktions-Workflow sowie die Entwicklung möglicher Konvertierungs-Softwares erste Ansatzpunkte für weitere Forschungen sein.

Bei Verwendung der jeweiligen Produktions-Software wurde deutlich, dass bereits ab diesem Schritt der Workflow stark von technischen Spezifikationen abhängt und von diesen vorgegeben wird, begonnen bei der Nutzung der Produktions-DAW. Weiterhin bedarf es der Entwicklung von Plugins, die speziell für die Anwendung in objektbasierten und immersiven Mischungen ausgelegt sind (beispielsweise Hall-, Kompressions- oder Mastering-Plugins, die sowohl mit dreidimensionaler Räumlichkeit als auch mit Objekten arbeiten können). Upmix-Technologien können hierbei durch die Anwendung auf einzelnen Audiospuren interessante Möglichkeiten bieten, wurden jedoch im Rahmen dieser Arbeit nicht näher betrachtet. Weiterhin muss der gesamte Schritt des Masterings für objektbasierte Mischungen angepasst werden – insbesondere in diesem Fall ist es notwendig, dass sich Vorgehensweisen etablieren, die formatübergreifend durchgesetzt werden können. Hierzu zählt auch, den neuen Workflow-Schritt des *Authorings* (der Generierung und Überprüfung der Metadaten) in den Mastering-Prozess zu integrieren.

Weiterhin ist die konsequente Implementierung technischer Anforderungen unerlässlich. Mit 360 Reality Audio wurde ein Format kommerziell verfügbar gemacht, welches rein technisch in der Theorie weiter ist als in der Praxis. Zum einen werden durch die Verwendung eines reduzierten MPEG-H-Profiles Funktionen wie Lautheitsmessung nicht unterstützt, die der Codec bieten könnte und die den Produktionsprozess verbessern würden. Zum anderen könnten die Audioinhalte auf einer Vielzahl an Wiedergabegeräten abgespielt werden, hierfür fehlt es jedoch noch an konsequenter Implementierung der

technischen Anforderungen in den Endgeräten. Bei Dolby Atmos Music kann sich die Distribution zweier unterschiedlich encodierter Dateien negativ auf den Rendering-Prozess auswirken – da nur über Lautsprecher objektbasiertes Rendering erfolgt und bei der Kopfhörerwiedergabe eine vorbinauralisierte Datei wiedergegeben wird, kommen in diesem Fall keinerlei Vorteile der objektbasierten Produktion beim Endkunden an. Außerdem wurde im Rahmen dieser Thesis deutlich, dass objektbasierte Produktionsweisen wie Dolby Atmos Music oder 360 Reality Audio auf technisch komplexen Hintergrundprozessen basieren, auf die Produzenten wenig Einfluss haben – dies reicht von der Implementierung der Software bis zum Exportprozess und (Binaural)-Rendering bei Monitoring und Wiedergabe beim Endkonsumenten. Somit wird während der Mischung ein hohes Maß an Verantwortung an nicht beeinflussbare technische Vorgänge abgegeben, die zwischen dem Produzenten und Konsumenten liegen und dazu führen, dass nicht nur die bewusst getroffenen Entscheidungen des Produzenten den Klang bei der Wiedergabe beeinflussen.

> *„Delivering a Dolby Atmos Music project feels a little strange. I do the mix, send it out into the world and it sort of disappears. There are not many people that can hear it. [...] We're doing what we think philosophically sounds and should be right and hoping that it works. [...] We're working in the dark a little bit"* Serban Ghenea in (Harvey, 2020, S. 30).

Nicht zuletzt spielt der finanzielle Aspekt eine Rolle – während Stereo-Produktionen eine Vielzahl an Verwertungswegen bieten (Schallplatte, CD, Blu-ray, digitale Portale wie iTunes oder Soundcloud, Streaming-Dienste) und wirtschaftlich betrachtet somit breit aufgestellt sind, so beschränkt sich dies bei objektbasierten Produktionen aktuell auf einige Streaming-Dienste sowie vereinzelte Blu-rays. Abschließend gilt, wie bei jeder neuen Technologie, dass die Nachfrage über den kommerziellen Erfolg entscheiden wird. Durch die Einbindung von Dolby Atmos Music und 360 RA bei Streaming-Diensten werden die objektbasierten Inhalte der breiten Öffentlichkeit präsentiert. Wird dieses Angebot von den Endkonsumenten angenommen, so steigt der Bedarf an Produktionen sowie der Entwicklung und Optimierung von spezieller Soft- und Hardware – und damit einhergehend die Nachfrage nach einheitlicheren Workflows zur Produktion von objektbasierter Musik.

8 Ausblick

„First, do we have enough content being created? Second, is that content being distributed on services that people can get? Third, is it available on devices that people have? And then the fourth piece, is there buzz around it? Do consumers want this? Are content creators excited about creating it, and are they telling the world? If you can get all of those pieces in place and they're working positively, they start to reinforce each other. And then you have the change.“
Baker in (Kenny, 2020, S. 26)

In den aktuell verfügbaren Technologien sind die Merkmale objektbasierten Audios noch nicht durchgehend implementiert – Personalisierung ist nur zu einem geringen Teil bei 360 Reality Audio möglich, Nutzer-Interaktivität wurde nicht umgesetzt, eine integrierte Lautheitsmessung erfolgt nur bei Dolby Atmos Music und flexibles Wiedergabe-Rendering ist ebenfalls nicht konsistent möglich: 360 RA bietet die technischen Voraussetzungen, jedoch fehlt es in der Praxis an der Implementierung der entsprechenden Decoder in den Endgeräten. Dolby Atmos Music bietet flexibles Wiedergabe-Rendering für Lautsprecher, bei der Kopfhörerwiedergabe wird jedoch auf eine vorbinauralisierte Zweikanal-Datei zurückgegriffen, was bedeutet, dass hierbei wiedergabeseitig keine objektbasierten Merkmale angewendet werden können. Das, was beide Technologien aktuell auf Seite der Wiedergabe verbindet, ist die Kombination aus Musik-Streaming-Diensten und Kopfhörern.

> *„It's important to remember that immersive music is still in its infancy, artistically and technologically, from the studio down to the home. Right now many, including Sony RA360 and Dolby themselves, are betting the bank on the Binaural setting in the Renderer, knowing that a vast majority of consumers will be listening on headphones“* (Kenny, 2020, S. 26).

Damit hier ein Mehrwert gegenüber kanalbasierten, rein binauralen Produktionen ausgespielt werden kann, bedarf es zum einen der Optimierung der Binaural-Renderer bzw. der verwendeten Binaural-Profile, bei 360 RA sind beispielsweise aktuell klangliche Verfärbungen hörbar. Zum anderen könnte die Implementierung von personalisierten HRTFs

sowie Head-Tracking-Funktionen einen weiteren Vorteil sowohl produktions- als auch wiedergabeseitig bieten, da bei der binauralen Kopfhörerwiedergabe nicht unterschieden werden kann, ob ein Signal, das mit Azimuth 0° positioniert wurde, von vorne oder hinten kommt. Da die Pegel- und Laufzeitdifferenz in beiden Fällen exakt gleich wäre, und zur Lokalisation eine Kopfdrehung erfolgen müsste, kommt es hierbei zur „Vorne-Hinten-Vertauschung“ (Weinzierl, 2008, S.675). Durch die Implementierung von Head-Tracking wäre es möglich, bei der Produktion und Wiedergabe die Positionsdaten der Kopfbewegungen zu erfassen, was eine Unterscheidung von vorn und hinten ermöglichen würde. Die Implementierung einer Head-Tracking-Funktion im Apple iOS 14 Update zur Nutzung mit den Apple AirPods Pro zeigt, dass die technischen Möglichkeiten gegeben sind um Head-Tracking für Konsumenten unkompliziert erlebbar zu machen.

Für Speicherung, Austausch und Konvertierung verschiedener Formate könnte das Audio Definition Model einen gemeinsamen Nenner bieten. Aktuell sind die einzelnen ADM-Profile jedoch noch nicht durchgehend kompatibel, an diesem Punkt erfordert es weitere Forschung. So könnten sich für spezielle Anwendungsbereiche verschiedene Profile etablieren. In Bezug auf die objektbasierte Musikproduktion könnte ein solches – möglichst formatagnostisches – Profil bedeuten, dass unabhängig der Produktionstechnologie am Ende eine ADM-Datei exportiert wird, welche zum einen encodiert und zum anderen in andere objektbasierte Produktionsformate umgewandelt werden kann. Auch wenn es sich hierbei aktuell um eine Wunschvorstellung handelt, wird sich zeigen, ob sich dies über die Jahre nicht technisch implementieren ließe.

Während abzuwarten bleibt, ob weitere Firmen wie Auro oder DTS in die objektbasierte Musikproduktion einsteigen und sich diese Technologien etablieren werden, zeigt sich schon jetzt, dass immersive Musikproduktion den Weg in die Musikindustrie findet: Vor zwei Jahren änderte die *Recording Academy* den Titel ihrer GRAMMY Award Kategorie „Best Surround Sound Album“ in „Best Immersive Audio Album“. Nun hat der *Recording Academy Producers & Engineers Wing* ein neues Komitee gegründet, welches Empfehlungen zur immersiven Audioproduktion aussprechen soll. Themen hierfür sollen formatagnostische Vorgehensweisen, Namensgebungen sowie Verfahren für die Aufnahme, Distribution und Archivierung von immersiven Audioinhalten sein. Möglich also, dass zukünftig durch die „Best Immersive Audio Album“-Kategorie eine Dolby Atmos Music oder 360 Reality Audio Produktion nominiert wird und somit – zumindest kurzzeitig – einen besonderen Fokus auf objektbasierte Musikproduktion lenkt.

Literaturverzeichnis

Amazon. (2019). *What is 3D Audio?* Zugriff am 2020-07-27 auf https://www.amazon.com/gp/help/customer/display.html?nodeId=G8GRF4QSPBWNWJBZ

Apple. (2020). *Audio in Dolby Atmos oder Surround Sound auf Apple TV wiedergeben.* Zugriff am 2020-07-28 auf https://support.apple.com/de-de/HT204069

Blauert, J. (1974). *Räumliches Hören* (1. Aufl.). Stuttgart, Deutschland: Hirzel.

Byers, R., Johnston, J., Kean, J., Lund, T., Orban, R. & Wisbey, A. (2015). *Recommendation for Loudness of Audio Streaming and Network File Playback* (Bericht). New York, NY, USA: Audio Engineering Society.

Cohen, S. (2020a). *How to know if you're actually getting Dolby Atmos sound.* Zugriff am 2020-07-28 auf https://www.digitaltrends.com/home-theater/dolby-atmos-sound/

Cohen, S. (2020b). *What is Dolby Atmos Music, and how can you experience it?* Zugriff am 2020-06-17 auf https://www.digitaltrends.com/home-theater/what-is-dolby-atmos-music-and-how-to-get-it/

Coleman, P., Franck, A., Jackson, P., Hughes, R., Remaggi, L. & Melchior, F. (2016). On Object-Based Audio with Reverberation. *Proceedings of Audio Engineering Society International Conference: DREAMS*.

Dickreiter, M., Dittel, V., Hoeg, W. & Wöhr, M. (Hrsg.). (2014). *Handbuch der Tonstudiotechnik* (8. Aufl.). Berlin, Deutschland; Boston, MA, USA: Walter de Gruyter GmbH & Co KG.

Dolby Laboratories. (2015). *Dolby AC-4: Audio Delivery for Next-Generation Entertainment Services.* San Francisco, CA, USA. Zugriff auf https://www.dolby.com/us/en/technologies/ac-4/Next-Generation-Entertainment-Services.pdf

Dolby Laboratories. (2017a). *Dolby Media Encoder User's Manual* (Bericht). San Francisco, CA, USA: Dolby Laboratories Inc.

Dolby Laboratories. (2017b). *Is Dolby AC-4 the same as Dolby Atmos?* Zugriff am 2020-06-24 auf `https://developerkb.dolby.com/support/solutions/articles/16000067755-is-dolby-ac-4-the-same-as-dolby-atmos-`

Dolby Laboratories. (2018). *Dolby Atmos Renderer Guide* (Bericht Nr. August). San Francisco, CA, USA: Dolby Laboratories Inc.

Dolby Laboratories. (2019a). *Dolby Atmos Master ADM Profile v1.0 22* (Bericht Nr. July). San Francisco, CA, USA: Dolby Laboratories Inc.

Dolby Laboratories. (2019b). *Dolby Media Encoder.* Zugriff am 2020-07-15 auf `https://professional.dolby.com/product/dolby-media-encoder-client/`

Dolby Laboratories. (2020a). *7.1.4 Overhead speaker setup.* Zugriff am 2020-07-27 auf `https://www.dolby.com/about/support/guide/speaker-setup-guides/7.1.4-overhead-speaker-setup-guide/`

Dolby Laboratories. (2020b). *AvidPlay: The first DIY distribution service supporting Dolby Atmos Music.* Zugriff am 2020-07-31 auf `https://professional.dolby.com/music/avidplay/`

Dolby Laboratories. (2020c). *Dolby Atmos Music - Delivery Specifications* (Bericht). San Francisco, CA, USA: Dolby Laboratories Inc.

Dolby Laboratories. (2020d). *Dolby Atmos Music Panner Guide* (Bericht). San Francisco, CA, USA: Dolby Laboratories Inc.

Dolby Laboratories. (2020e). *Welcome to the next era of music.* Zugriff am 2020-07-31 auf `https://www.dolby.com/music/`

Dolby Laboratories. (2020f). *What is in the .mp4 export and how are these encoded?* Zugriff am 2020-07-27 auf `https://developerkb.dolby.com/support/solutions/articles/16000100315-what-is-in-the-mp4-export-and-how-are-these-encoded-`

Dolby Laboratories. (2020g). *Who is going to hear my Dolby Atmos mix?* Zugriff am 2020-06-24 auf `https://developerkb.dolby.com/support/solutions/articles/16000099189-who-is-going-to-hear-my-dolby-atmos-mix-`

EBU. (2014). *Loudness normalisation and permitted maximum level of audio signals* (Bericht). Genf, Schweiz: European Broadcasting Union.

EBU. (2018). *Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications* (Bericht). Genf, Schweiz: European Broadcasting Union.

Fedak, P. (2020). *MPEG-H Sony Architect Demo on Vimeo.* Zugriff am 05.12.2020 auf `https://vimeo.com/419028049`

Fielder, L. D., Andersen, R. L., Crockett, B. G., Davidson, G. A., Davis, M. F., Turner, S. C., ... Williams, P. A. (2004). Introduction to Dolby Digital Plus, an Enhancement to the Dolby Digital Coding System. In *Proceedings of the 117th aes convention.* San Francisco, CA, USA: Audio Engineering Society.

FireTV-Blog. (2018). *Mit dem Fire TV Dolby Atmos streamen – so geht's.* Zugriff am 2020-07-28 auf `https://firetv-blog.de/mit-dem-fire-tv-dolby-atmos-streamen-so-gehts/`

FireTV-Blog. (2020). *Tidal steamt Dolby-Atmos-Music auf den Amazon Fire TV Stick.* Zugriff am 2020-07-28 auf `https://firetv-blog.de/tidal-steamt-dolby-atmos-music-auf-den-amazon-fire-tv-stick/`

Floemer, A. (2019). *Amazon Echo Studio im Test: So gut klang Alexa noch nie.* Zugriff am 2020-07-28 auf `https://t3n.de/news/amazon-echo-studio-test-gut-klang-alexa-1218541/`

Fraunhofer IIS. (2019). *Einhüllender Sound mit „360 Reality Audio" von Sony.* Zugriff am 2020-05-16 auf `https://www.audioblog.iis.fraunhofer.com/de/sony-360-reality-audio-mpegh`

Fraunhofer IIS. (2020a). *Fraunhofer ADM Info Tool* (Bericht). Erlangen, Deutschland: Fraunhofer-Institut für Integrierte Schaltungen (IIS).

Fraunhofer IIS. (2020b). *What is the bitrate of an MPEG-H program?* Zugriff am 2020-10-05 auf `https://www.mpegh.com/en/info/`

Füg, S., Marston, D. & Norcross, S. (2016). The Audio Definition Model – A Flexible Standardised Representation for Next Generation Audio Content in Broadcasting and Beyond. In *Proceedings of the 141st aes convention.* Los Angeles, CA, USA: Audio Engineering Society.

Garity, W. & Hawkins, J. (1941). Fantasound. *SMPTE Motion Imaging Journal*, *37* (8), 127–146.

Geier, M., Carpentier, T., Noisternig, M. & Warusfel, O. (2017). *Software tools for object-based audio production using the Audio Definition Model (ITU-R Recommendation BS.2076)* (Bericht). UMR 9912 STMS IRCAM–CNRS–UPMC.

Genewick, S. (2020). 10 Things I've Learned Mixing Dolby Atmos Music. *Mix Magazine*, *44* (5), 42.

Grewe, Y., Simon, C. & Scuda, U. (2018). Producing Next Generation Audio using the MPEG-H TV Audio System. *Proceedings of Broadcast Engineering and Information Technology Conference*, 335–341.

Grüner, O. (2019). *Dolby Atmos: Alles, was du über das 3D-Klangformat wissen musst.* Zugriff am 2020-07-28 auf `https://hifi.de/ratgeber/dolby-atmos-3d-klangformat-18627`

Hahn, E. (2018). Musical Emotions Evoked by 3D Audio. In *Proceedings of the conference on spatial reproduction.* Tokio, Japan: Audio Engineering Society.

Harvey, S. (2020). After Hours On The Weeknd: Chart-Topping Stereo Mix gets Full Atmos Treatment. *Mix Magazine*, *44* (5), 28–30.

Herre, J., Hilpert, J., Kuntz, A. & Plogsties, J. (2015). MPEG-H 3D Audio - The New Standard for Coding of Immersive Spatial Audio. *IEEE Journal on Selected Topics in Signal Processing*, *9* (5), 770–779. doi: 10.1109/JSTSP.2015.2411578

Herre, J., Purnhagen, H., Koppens, J., Hellmuth, O., Engdegård, J., Hilpert, J., ... Oh, H. O. (2012). MPEG spatial audio object coding-The ISO/MPEG standard for efficient coding of interactive audio scenes. *AES: Journal of the Audio Engineering Society*, *60* (9), 655–673.

Hestermann, S., Seideneck, M. & Sladeczek, C. (2018). An Approach for Mastering Audio Objects. In *Proceedings of the conference on spatial reproduction.* Tokio, Japan: Audio Engineering Society.

ITU-R. (2014). *Method for the subjective assessment of intermediate quality level of audio systems* (Bd. BS Series; Bericht). Genf, Schweiz: International Telecommunication Union.

ITU-R. (2015). *Audio Definition Model* (Bericht). Genf, Schweiz: International Telecommunication Union.

ITU-R. (2019a). *Long-form file format for the international exchange of audio programme materials with metadata* (Bericht). Genf, Schweiz: International Telecommunication Union.

ITU-R. (2019b). *A serial representation of the Audio Definition Model* (Bericht). Genf, Schweiz: International Telecommunication Union.

Izhaki, R. (2008). *Mixing Audio: Concepts, Practices and Tools* (1. Aufl.). Burlington, MA, USA; Oxford, UK: Elsevier.

Jax, P., Meltzer, S., Neuendorf, M. & Sen, D. (2014). MPEG-H 3D Audio - The Next Generation Audio System. In *International broadcasting convention (ibc) 2014 conference* (S. 1–8). Amsterdam: Institution of Engineering and Technology. Zugriff auf `https://digital-library.theiet.org/content/conferences/10.1049/ib.2014.0011` doi: 10.1049/ib.2014.0011

Jia, M., Zhang, J., Bao, C. & Zheng, X. (2017). A psychoacoustic-based multiple audio object coding approach via intra-object sparsity. *Applied Sciences (Switzerland)*, *7* (12). doi: 10.3390/app7121301

Kaczmarek, V. (2020). *Tidal & Dolby – bringen Dolby Atmos-Musik nach Hause.* Zugriff am 2020-07-28 auf `https://play-experience.com/tidal-dolby-bringen-dolby-atmos-musik-nach-hause/`

Kenny, T. (2020). Mix special edition: Dolby Atmos Music. *Mix Magazine*, *44* (5).

Kjörling, K., Rödén, J., Wolters, M., Riedmiller, J., Biswas, A., Ekstrand, P., ... Vinton, M. (2016). AC-4 – The Next Generation Audio Codec. In *Proceedings of the 140th aes convention.* Paris, Frankreich: Audio Engineering Society.

Lawrence, R. (2019). Producing Music for lmmersive Audio Experience. In M. Hepworth-Sawyer, R., Hodgson, J., & Marrington (Hrsg.), *Producing music* (1. Aufl., S. 134–155). New York, NY, USA: Routledge.

Lee, H. (2017). Sound source and loudspeaker base angle dependency of phantom image elevation effect. *AES: Journal of the Audio Engineering Society*, *65* (9), 733–748. doi: 10.17743/jaes.2017.0028

Lossius, T., Baltazar, P. & de la Hogue, T. (2009). DBAP - Distance-based amplitude panning. In *Proceedings of the international computer music conference.* Montreal, Quebec, Kanada: Schulich School of Music of McGill University.

Mäsel, J., Simon, C., Choi, J., Schulz, A. & Silzle, A. (2019). Field Test for Immersive and Interactive Audio Production with the Gewandhausorchester Leipzig using MPEG-H. In *Proceedings of the interrnational conference on spatial audio.* Ilmenau, Deutschland: Verband Deutscher Tonmeister.

Nvidia. (2019). *Shield TV Pro Produktspezifikationen.* Zugriff am 2020-07-28 auf `https://www.nvidia.com/de-de/shield/shield-tv-pro/`

Oldfield, R., Shirley, B. & Spille, J. (2014). An object-based audio system for interactive broadcasting. *137th Audio Engineering Society Convention 2014*, 930–939.

Olivieri, F., Peters, N. & Sen, D. (2019). *Scene-Based Audio and Higher Order Ambisonics: A technology overview and application to Next-Generation Audio, VR and 360 Video* (Bericht). Genf, Schweiz: European Broadcasting Union.

Owsinski, B. (2006). *The Mixing Engineer's Handbook* (2. Aufl.). Boston, MA, USA: Cengage Learning.

Pike, C. W. (2019). *Evaluating the Perceived Quality of Binaural Technology* (Dissertation). University of York, UK.

Purnhagen, H., Hirvonen, T., Villemoes, L., Samuelsson, J. & Klejsa, J. (2016). Immersive Audio Delivery Using Joint Object Coding. In *Proceedings of the 140th aes convention.* Paris, Frankreich: Audio Engineering Society.

Roberts, B. (2020). *Tidal expands Dolby Atmos Music support to soundbars, TVs and AVRs.* Zugriff am 2020-07-25 auf `https://www.whathifi.com/news/tidal-expands-dolby-atmos-music-support-to-soundbars-tvs-and-avrs`

Robjohns, H. (2014). *The End Of The Loudness War?* Zugriff am 2020-09-18 auf `https://www.soundonsound.com/techniques/end-loudness-war`

Roginska, A. (2017). Binaural Audio Through Headphones. In A. Roginska & P. Geluso (Hrsg.), *Immersive sound: The art and science of binaural and multi-channel audio* (1. Aufl., S. 88–123). New York, NY, USA; London, UK: Taylor & Francis.

Romanowski, M. (2020). Mastering for Immersive Audio: What Is It? And Why Do We Need It? *Mix Magazine*, *4* (55), 32–33.

Rumsey, F. (2015). Immersive audio, objects, and coding. *AES: Journal of the Audio Engineering Society*, *63* (5), 394–398.

Rumsey, F. (2017). Surround Sound. In A. Roginska & P. Geluso (Hrsg.), *Immersive sound: The art and science of binaural and multi-channel audio* (S. 180–220). Taylor & Francis.

Rumsey, F. (2018). Spatial audio Channels, objects, or ambisonics? *AES: Journal of the Audio Engineering Society*, *66* (11), 987–992.

Sazdov, R., Paine, G. & Stevens, K. (2007). Perceptual investigation into envelopment, spatial clarity, and engulfment in reproduced multi-channel audio. In *Proceedings of the 31st international aes convention.* London, UK: Audio Engineering Society.

Schultz, B. (2020a). Classic Tracks Go Immersive: Elton John and Prince Get the Dolby Atmos Music Treatment. *Mix Magazine*, *44* (5), 12.

Schultz, B. (2020b). Fantastic Negrito Fills the Room With Music, Voice and Sounds. *Mix Magazine*, *44* (5), 10–11.

Serck, K. (2020). *Dolby Atmos Music von Tidal jetzt auch für AV-Receiver (Update).* Zugriff am 2020-07-28 auf `https://www.areadvd.de/news/dolby-atmos-musik-von-tidal-jetzt-auch-fuer-av-receiver/`

Shirley, B., Oldfield, R., Melchior, F. & Batke, J.-M. (2013). Platform independent audio. In O. Schreer et al. (Hrsg.), *Media production, delivery and interaction for platform independent systems: Format-agnostic media* (1. Aufl., S. 130–165). Chichester, UK: John Wiley & Sons, Ltd.

Sony Corporation. (2019a). *360 Reality Audio von Sony: Startschuss für ein neues Musikerlebnis.* Zugriff am 2020-04-01 auf `https://presscentre.sony.at/pressreleases/360-reality-audio-von-sony-startschuss-fur-ein-neues-musikerlebnis-2932190`

Sony Corporation. (2019b). *Inhalte in 360 Reality Audio von Sony über den Streamingdienst Amazon Music HD abrufbar.* Zugriff am 2020-04-01 auf `https://presscentre.sony.de/pressreleases/inhalte-in-360-reality-audio-von-sony-ueber-den-streamingdienst-amazon-music-hd-abrufbar-2924447`

Sony Corporation. (2020a). *360 Reality Audio.* Zugriff am 2020-07-20 auf `https://www.sony.de/electronics/360-reality-audio`

Sony Corporation. (2020b). *360 Reality Audio. Zusammengetragene Informationen aus Audiofachmessen, Konferenzen, Webinaren und persönlichem Kontakt.*

Thomas, C. (2020a). *Dolby Atmos Music Specifications [Webinar].* Dolby Laboratories. Zugriff auf `https://professional.dolby.com/webinars-dolby-atmos-music/`

Thomas, C. (2020b). *Dolby Atmos Music Studio to the World [Webinar].* Dolby Laboratories. Zugriff auf `https://professional.dolby.com/webinars-dolby-atmos-music/`

Thomas, C. (2020c). *Dolby Atmos Music via Headphones [Webinar].* Dolby Laboratories. Zugriff auf `https://professional.dolby.com/webinars-dolby-atmos-music/`

Thomas, C. (2020d). *Introduction to Dolby Atmos Music [Webinar].* Dolby Laboratories. Zugriff auf `https://professional.dolby.com/webinars-dolby-atmos-music/`

Thornton, M. (2020). *AvidPlay Now With Dolby Atmos - First Platform To Support Dolby Format.* Zugriff am 2020-07-31 auf `https://www.pro-tools-expert.com/production-expert-1/2020/7/27/avidplay-becomes-the-first-diy-music-distribution-service-to-support-dolby-atmos-music`

Tidal. (2020a). *360 Reality Audio.* Zugriff am 2020-07-20 auf `https://www.sony.de/electronics/360-reality-audio`

Tidal. (2020b). *Dolby Atmos Music.* Zugriff am 2020-07-17 auf `https://support.tidal.com/hc/en-us/articles/360004255778-Dolby-Atmos-Music`

Tsingos, N. (2017). Object-Based Audio. In A. Roginska & P. Geluso (Hrsg.), *Immersive sound: The art and science of binaural and multi-channel audio* (1. Aufl., S. 244–275). New York, NY, USA: Taylor & Francis.

Weinzierl, S. (Hrsg.). (2008). *Handbuch der Audiotechnik* (1. Aufl.). Berlin/Heidelberg, Deutschland: Springer. doi: 10.1007/978-3-540-34301-1

Wenzel, E. M., Begault, D. R. & Godfroy-Cooper, M. (2017). Perception of Spatial Sound. In A. Roginska & P. Geluso (Hrsg.), *Immersive sound: The art and science of binaural and multi-channel audio* (1. Aufl., S. 5–39). New York, NY, USA; London, UK: Taylor & Francis.

Woodcock, J., Francombe, J., Franck, A., Coleman, P., Hughes, R., Kim, H., ... Hilton, A. (2018). A framework for intelligent metadata adaptation in object-based audio. In *Proceedings of the conference on spatial reproduction.* Tokio, Japan: Audio Engineering Society.

Zehden, M. (2018). *Apple TV für Einsteiger: Grundlagenwissen kompakt erklärt.* Zugriff am 2020-07-28 auf `https://www.maclife.de/ratgeber/apple-tv-einsteiger-grundlagenwissen-kompakt-erklaert-10098990.html`

A Persönliche Kommunikation

A.1 Thomas Ceri (Dolby Laboratories), E-Mail

Re: [Dolby Atmos Music Webinars] Question regarding encoding
1 Nachricht

Thomas, Ceri <Ceri.Thomas@dolby.com> Fr., 17. Juli 2020, 19:11
An: Daniela Rieger <dr078@hdm-stuttgart.de>

Yes you understand correctly.

Have a great weekend.

Ceri Thomas

Dolby

Dolby Atmos Music Knowledgebase

Dolby Atmos Music Studio Onboarding Form

SXSW Masterclass

Webinar Channel - Dolby Music

From: Daniela Rieger <dr078@hdm-stuttgart.de>
Reply-To: Daniela Rieger <dr078@hdm-stuttgart.de>
Date: Friday, July 17, 2020 at 6:35 AM
To: "Thomas, Ceri" <Ceri.Thomas@dolby.com>
Subject: Re: [Dolby Atmos Music Webinars] Question regarding encoding

Hi Ceri,

thank you very much for the quick reply! That is really interesting to know and helps me a lot understanding the deliveries.

I've got one more question:
If I'd like to distribute something over e.g. Tidal, the ADM-BWAV gets encoded to both codecs AC-4 IMS (IMS means it is already binauralized) and DD+JOC. Is that correct?

Tidal then gets both deliveries, and if I'm listening over headphones, it selects the binaural AC-4 IMS and for speaker playback the DD+JOC. Did I understand that right?

Thank you so much in advance and I hope you have a nice weekend!

Best,
Daniela

"Thomas, Ceri" <Ceri.Thomas@dolby.com> hat am 16. Juli 2020 um 20:02 geschrieben:

Daniela,

Thank you for reaching out and I'm glad you enjoyed and found the webinars useful.

Your understanding of the process is correct, the single delivery package of the ADM BWAV contains all the necessary data for the encoder to create the two encoded formats, they are however two separate outputs.

We are currently in Early access for the AvidPlay integration of Dolby Atmos music. Avid have done the work to support the encoder needed for the format and also the pipeline to then deliver those encoded assets to both Amazon and Tidal.

I hope that clarifies the situation for you?

All the best,

Ceri Thomas

Dolby

Dolby Atmos Music Knowledgebase

Dolby Atmos Music Studio Onboarding Form

SXSW Masterclass

Webinar Channel - Dolby Music

From: Daniela Rieger <dr078@hdm-stuttgart.de>
Reply-To: Daniela Rieger <dr078@hdm-stuttgart.de>
Date: Thursday, July 16, 2020 at 10:00 AM
To: musicstudios <musicstudios@dolby.com>
Subject: [Dolby Atmos Music Webinars] Question regarding encoding

Hi all,

my name is Daniela and I am a sound engineering student from Stuttgart Media University (Germany). I've been focussing on 3D audio for the last two years, and I'm really impressed by Dolby Atmos and very curious about the latest and upcoming Dolby Atmos Music releases!

I've attended all Dolby Atmos Music Webinars by Ceri Thomas (they were really interesting), and now I've been rewatching some of them and came up with one question regarding the encoding process:

In the webinar, Ceri Thomas said, that Dolby Atmos Music for headphones is encoded with AC-4, and Dolby Atmos Music for speakers with DD+JOC. Now I was wondering - how does this work? Does the mentioned cloud encoding service encode AC-4 and DD+JOC in the same bitstream, and this one delivery-file is delivered to Tidal (headphones) and Amazon Music (speakers)? Or does one need to encode the mix to AC-4 (and deliver this to Tidal) and DD+JOC (and deliver this to Amazon) separately?

It would be awesome, to learn more about this!
I'm really looking forward to hearing from you!

Best regards,
Daniela Rieger

A.2 Thomas Ceri (Dolby Laboratories), E-Mail

Re: [Dolby Atmos Music Webinars] Question regarding encoding
1 Nachricht

Thomas, Ceri <Ceri.Thomas@dolby.com> Gestern um 18:52
An: Daniela Rieger <dr078@hdm-stuttgart.de>

Re Bitrates, yes the lower value is for step down for bandwidth limitations.

Re device decode: as a rule yes I would say that the last device will be doing the Atmos decode. I would urge you to find the settings on the Atmos enabled TV and set the output to bitstream. Some TVs might otherwise do the decode and then pass out PCM which defeats the purpose of your soundbar, some TVs just like to think they know better 😊

Ceri Thomas

Dolby

Cell: 310 913 8562

Dolby Atmos Music Knowledgebase

Dolby Atmos Music Studio Onboarding Form

SXSW Masterclass

Webinar Channel - Dolby Music

From: Daniela Rieger <dr078@hdm-stuttgart.de>
Reply-To: Daniela Rieger <dr078@hdm-stuttgart.de>
Date: Thursday, July 23, 2020 at 8:56 AM
To: "Thomas, Ceri" <Ceri.Thomas@dolby.com>
Subject: Re: [Dolby Atmos Music Webinars] Question regarding encoding

Hi Ceri,

thanks for replying so quickly - I've got two more questions, as I'm beginning to understand how complex Dolby Atmos really is and how little I know yet! :)

If I want to listen to Dolby Atmos Music over speakers and I've got: Atmos compatible TV and/or Atmos-AVR and then a Atmos soundbar connected to it. Is the decoding always happening in the last device of the chain (in this case soundbar)?

And: in the Webinar slides there are two bitrates for each AC4-IMS and DD+JOC. Is that (the lower one) in case the user has low-bandwith connection?

- DD+JOC = 768kbps & 448kbps
- AC4-IMS = 256kbps & 112kbps