Analyse des Audio Definition Model hinsichtlich der Funktionalität auf Kreationsebene

Bachelorarbeit im Studiengang Audiovisuelle Medien

vorgelegt von

Daniel Strübig

Matr.-Nr.: 28143

am 01. Februar 2018

an der Hochschule der Medien Stuttgart

zur Erlangung des akademischen Grades

Bachelor of Engineering

Erstprüfer/in:	Prof. Oliver Curdt
Zweitprüfer/in:	Dipl.-Ing. Daniel Deboy

Erklärung an Eides statt

Hiermit versichere ich, Daniel Strübig, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel "Analyse des Audio Definition Model hinsichtlich der Funktionalität auf Kreationsebene" 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. 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.

Auszug aus dem Strafgesetzbuch (StGB):

(§ 156 StGB) Falsche Versicherung an Eides Statt

"Wer von einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft."

Stuttgart, 31.01.2018,

_____

Daniel Strübig

Kurzfassung

Mit der Markteinführung der Systeme DTS:X und Dolby Atmos® haben sich objektbasierte Audiosysteme für die Wiedergabe immersiver Audioinhalte zu Bewegtbild in Kinosälen etabliert. Mithilfe von Rendering Units und proprietären Metadatensystemen wird die Kompatibilität von Mehrkanalmischungen ermöglicht. Durch die wachsende Popularität von VR-Inhalten, immersiven Heimkino-Systemen und Kopfhörerwiedergabe durch Binauralsynthese steigt allerdings auch der Bedarf für ein transkompatibles Metadatenformat auf Kreativen- und Konsumentenseite. Um einen effizienten Workflow für immersive Audioinhalte im Studio zu entwickeln, sollte ein Metadatenformat existieren, das möglichst systemagnostisch wiedergegeben werden kann. Die folgende Thesis beschäftigt sich mit dem Metadatenformat Audio Definition Model. Nachdem das Konzept, die Zielsetzung und die Bestandteile des Audio Definition Model herausgearbeitet wurden, folgt anschließend die exemplarische Auswertung verschiedener Interviews mit Mischtonmeistern für immersive Audioinhalte. Kernfrage des Interviews ist: Welche Anforderungen stellen Kreateure an Metadatenformate? Abschluss der Arbeit bildet die Beantwortung der Frage, in welchem Umfang das Audio Definition Model die Bedürfnisse der Inhaltsentwickler erfüllt und gibt einen Ausblick, welche technischen Hürden bewältigt werden müssen, um das Audio Definition Model als Metadatenformat zu etablieren.

Schlagwörter: Immersive Audio, Metadaten, Spatial Audio

Abstract

With the introduction of DTS:X and Dolby Atmos®, object-based audio systems for playing back immersive audio content for motion picture in cinemas have been established. With the help of rendering units and proprietary metadata systems, downscaling of multichannel mixes is now possible. Due to the growing popularity of VR content, immersive home cinema and headphone playback using binaural synthesis, content creators as well as consumers demand transcompatible metadata. In order to develop an efficient workflow for creating immersive audio content, there should be a metadata format with transcompatible and system-agnostic playback possibilities. This thesis deals with the metadata format Audio Definition Model. After the concept, goals and components of the Audio Definition Model are described, interviews with mixing engineers are held and analysed. Which specific demands do mixing engineers have with regard to metadata? The thesis concludes with answering the question, to which degree the Audio definition model fulfils the content creator's needs.

Keywords: Immersive Audio, Metadata, Spatial Audio

Inhaltsverzeichnis

Abkürzungsverzeichnis

VR	Virtual Reality
AR	Augmented Reality
MR	Mixed Reality
IA	Immersive Audio
ADM	Audio Definition Model
ILD	Interaurale Intensitätsunterschiede
ITD	Interaurale Laufzeitunterschiede
HRTF	Head Related Transfer Function
HRIR	Head Related Impulse Response
BRIR	Binaural Room Impulse Response
FFT	Fast Fourier Transform
CBA	Channel based Audio
OBA	Object based Audio
SBA	Scene based Audio
DAW	Digital Audio Workstation
BWF	Broadcast Wave File
RMU	Rendering and Mastering Unit
PCM	Pulse-Code-Modulation
chna	Channel Allocation
MXF	Material Exchange Format
HMD	Head-Mounted Display
VBAP	Vector based amplitude panning
ODV	Omnidirectional Video
OSC	Open Sound Control
LTC	Local Timecode

1 Einleitung

Mit der Markteinführung neuer Bewegtbildmedien steigt parallel dazu der Bedarf neuer immersiver Hörmedien. Durch die steigende Popularität virtueller Realitäten wurden Technologien wie beispielsweise die Binauralsynthese für Kopfhörer immens weiterentwickelt. VOD-Plattformen wie Netflix bieten seit kurzer Zeit Dolby Atmos® im Stream an[1]. Staaten wie Südkorea haben den UHD-TV-Standard ATSC 3.0 bereits im laufenden Betrieb eingesetzt[2]. Immersive Heimkino-Formate finden immer größeren Anklang bei Konsumenten und spätestens mit der Umstellung auf die neuen UHD DV-Standards DVB-UHD in Europa wird immersives Audio eine breite Masse an Endkonsumenten erreichen. Egal ob Virtual Reality, herkömmliches TV, Streaming oder Kinofilm: Immersives Audio gewinnt an Relevanz.

Diese Entwicklung birgt für Konsumenten ein neues Hörerlebnis, für Sounddesigner, Film- und Mischtonmeister und Content Creators hingegen eine neue Palette an kreativen Werkzeugen, die es zu erlernen gilt. Mit in der Entwicklung eher alten, in der Anwendung hingegen völlig neuen Wiedergabeformaten wie Ambisonics und Binauralsynthese durch HRTF-Faltung drängen auch kreationsseitig Soft- und Hardwarelösungen auf den Markt, die bei der Erstellung dieser neuen Audioformate helfen sollen. Da viele dieser neuen Formate in einer klassischen Stereo- oder Surround-Mischung kaum oder gar nicht Anwendung gefunden haben, gibt es ebenso kaum etablierte Workflows für beispielsweise VR-Mischungen. Darüber hinaus bildet sich seit einigen Monaten der Trend, dass Endgerätentwickler für VR auf ein proprietäres Format setzen[3], sodass für jede Plattform ein eigener Mix erstellt werden muss. Die Etablierung eines plattformübergreifenden Standards für Next Generation Audio scheint demnach wünschenswert und die Kreation von VR-Inhalten ist hierbei nur ein Teilmetier, welches davon profitieren würde.

Das von der Europäischen Union geförderte Forschungsprogramm ORPHEUS beschäftigt sich seit Dezember 2015 mit der Entwicklung solcher Audioformate. Ein Teilbereich dieses Forschungsprogramms ist die Entwicklung des Metdatenformats Audio Definition Model. Die vorliegende Arbeit beschäftigt sich mit zunächst mit den Grundlagen, Ansätzen und Funktionsweisen bestehender Wiedergabesysteme, namentlich: kanalbasierte Wiedergabe, objektbasierte Wiedergabe, szenenbasierte Wiedergabe. Im Anschluss wird das Audio Definition Model hinsichtlich seiner Zielsetzung, Technologie und Bestandteile analysiert. Hierauf folgt die Auswertung eigens geführter Interviews mit praktizierenden Toningenieuren.

[1] Nico Jurran, „3D-Sound: Netflix bietet nun auch Dolby-Atmos®-Ton“.
[2] Jan Fleischmann, „MPEG-H -- ein Audioformat der nächsten Generation (NGA)“.
[3] Rieger, „Virtual Reality Audio Formats - Pro und Contra“.

Exemplarisch wird untersucht: Was wünschen sich Kreateure? Welche Anforderungen haben Toningenieure an Audioformate der nächsten Generation? Die Befragten arbeiten alle in unterschiedlichen Bereichen der Audiokreation (VR, Spatial Music, Filmmischung, Klanginstallationen im Raum), sie alle eint jedoch der unmittelbare Gebrauch neuer Audioformate. Abschluss der Arbeit bildet die Beantwortung der Frage, ob und in welchem Umfang das Audio Definition Model die Anforderungen der Toningenieure abdeckt und gibt anschließend einen Ausblick, mit welchen Mitteln eine Markteinführung des Audio Definition Model als standardisiertes Metadatenformat erreicht werden kann. Ziel der vorliegenden Arbeit ist der Erkenntnisgewinn, welche Ansprüche Toningenieure zur Erstellung von immersive Audio haben und ob diese mit dem hier vorgestellten Metadatenmodell erreicht werden kann.

2 Grundlagen

Das folgende Kapitel bietet Definitionen relevanter Fachtermini und Praxisbeispiele für die hier erklärten Wiedergabeformate. Aus Gründen der Übersichtlichkeit wird hierbei auf die Definitionen von Schalldruck, Schallpegel, Schallfeld, Spezifikationen für Surround-Sound und anderen grundlegenden Größen verzichtet.

2.1 Räumliches Hören

Der folgende Abschnitt gibt einen Überblick über die Grundlagen des räumlichen Hörens. Da das Forschungsfeld des räumlichen Hörens in der Psychoakustik ein komplexer Themenblock in sich ist und in seiner Gänze den Rahmen dieser Arbeit sprengen würde, werden nur die elementaren Aspekte behandelt.

2.1.1 Lokalisation

Bei der Lokalisation eines Schallereignisses wird durch die Einflussnahme diverser Faktoren, bedingt durch die Anatomie des menschlichen Ohrs, das besagte Schallereignis einer Position im Raum zugewiesen. Diese Position wird in der Regel durch das kopfbezogene Koordinatensystem oder Polarkoordinatensystem beschrieben. Ursprung dieses Koordinatensystems ist die Mitte der interauralen Achse zwischen beiden Gehöreingängen des menschlichen Kopfes. Die in diesem Zuge definierte Horizontalebene verläuft durch die interaurale Achse. Die Medianebene verläuft orthogonal zur Horizontalebene durch die interaurale Achse. Die Frontalebene verläuft orthogonal zu beiden zuvor genannten Ebenen und "teilt" den gedachten Kopf hälftig. Es bedarf drei Größen, um die Position eines Schallereignisses im Polarkoordinatensystem eineindeutig zu beschreiben: Der Azimut-Wert beschreibt den Winkel auf der Horizontalebene, der Elevation-Wert beschreibt den Winkel auf der Frontalebene und die Distanz den relativen Abstand zum Ursprung an.[4] Alternativ kann diese Position auch im kartesischen Koordinatensystem über die Koordinaten XYZ angegeben werden.

2.1.1.1 Binaurale Faktoren

Binaurale Faktoren beschreiben Signalunterschiede, die beim Vergleich der wahrgenommenen Signale beider Ohren bei der Lokalisation eines Schallereignisses auftreten. Grundsätzlich wird zwischen interauralen Laufzeitunterschieden (ITD) und interauralen Intensitätsun-

[4] Weinzierl und Verband Deutscher Tonmeister, *Handbuch der Audiotechnik*. S. 88

terschieden (ILD) verglichen. ITD treten auf, wenn das wahrgenommene Signal an beiden Gehöreingängen an verschiedenen Zeitpunkten auftrifft. Hierbei ist der Kopfdurchmesser für die frequenzabhängige Lokalisation von Schallereignissen relevant: Bei einem Kopfdurchmesser von durchschnittlich 17 Zentimeter ist eine Lokalisation bis circa 1,5 kHz möglich, da unterhalb die resultierende Wellenlänge größer als der Kopfdurchmesser ist.[5]

ILD beeinflussen die Wahrnehmung des Schallereignisses bei seitlich eintreffendem Schall. Durch die Zuwendung eines Ohres zum Emitter des Schallereignisses entsteht an diesem Ohr ein Druckstau, am anderen Ohr durch die Masse des Kopfes ein Schallschatten.[6]

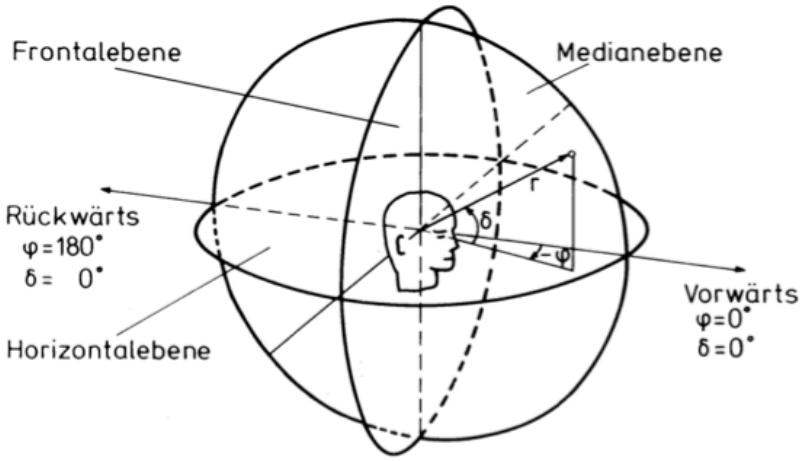

Abbildung 2-1 Polarkordinatensystem[7]

2.1.1.2 Monaurale Faktoren

Monaurale Faktoren treten auf, sobald keine ITD und ILD messbar sind. Dies ist beispielsweise der Fall, wenn der Azimut-Winkel 0 beträgt, wenn also Schallereignisse innerhalb der Medianebene auftreten. Die hierbei auftretenden Effekte sind im Frequenzspektrum zu finden und geschehen durch die Beugung des Schalls durch Torso und Pinna. Die resultierenden Frequenzbänder bestimmen maßgeblich die Lokalisation auf der Medianebene. Gemäß dieser Bänder (Blauertsche Bänder, benannt nach den Untersuchungen durch den Elektrotechniker Jens Blauert) ergeben sich erhebliche Defizite im räumlichen Hören beim Men-

[5] Dickreiter u. a., *Handbuch der Tonstudiotechnik*. S. 106 ff

[6] Görne, *Tontechnik*. S. 118 ff

[7] aus Weinzierl und Verband Deutscher Tonmeister, *Handbuch der Audiotechnik*. S. 88

schen bei von oben eintreffenden Schallereignissen.[8]

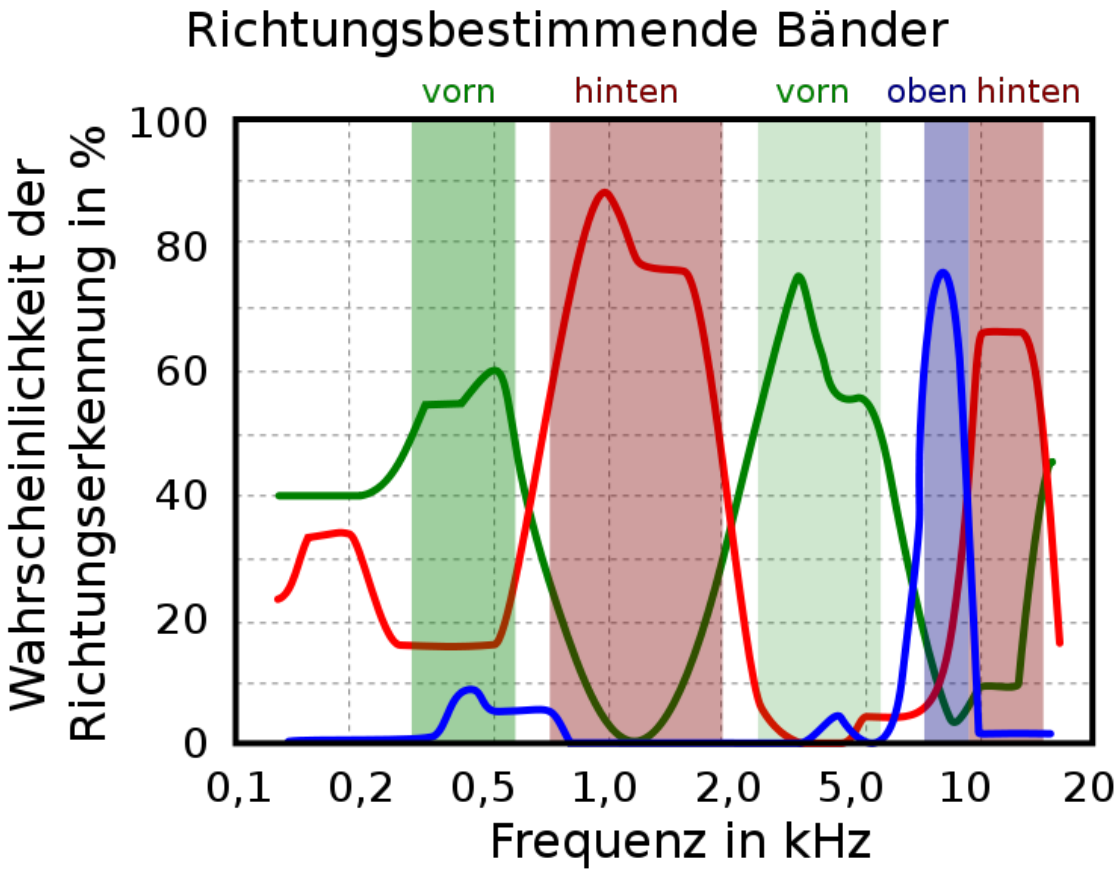

Abbildung 2-2 RIchtungsbestimmende Bänder nach Blauert[9]

2.1.2 Head-Related Transfer Functions

Wird ein Schallereignis vom menschlichen Gehör wahrgenommen, so tritt eine Vielzahl an Beugungen, Brechungen und linearen Verzerrungen auf. Anatomische Bestandteile des Kopfes und des Torsos, namentlich Torso, Schulter, Kopf, Pinna, Gehöreingang und der Gehörkanal selbst sind für diese verantwortlich. Diese Verzerrungen lassen sich durch eine kopfbezogene Impulsantwort (HRIR) messen und mittels FFT in eine kopfbezogene Transferfunktion (HRTF) zerlegen, welche die Frequenzabweichungen durch die erwähnten Beugungen beschreibt. Hierbei gilt zu beachten, dass sich in Abhängigkeit der Rotation und Neigung des Kopfes auch die dazugehörige HRTF ändert.[10] Nomenklatorisch wird hier zwischen HRIRs und binauralen Raumimpulsantworten (BRIR) unterschieden. Während HRIRs in re-

[8] Dickreiter u. a., *Handbuch der Tonstudiotechnik*. S. 95 ff
[9] Blauert, *Räumliches Hören*.
[10] Weinzierl und Verband Deutscher Tonmeister, *Handbuch der Audiotechnik*. S. 568 ff

flexionsarmen Räumen aufgenommen werden, beziehen BRIRs die Beugungen, Brechungen und Reflexionen des gemessenen Raumes mit in die Impulsantwort ein.[11]

2.2 Immersion

Im Wortsinne bezeichnet der Begriff der Immersion das Eintauchen eines Gegenstands in ein flüssiges Umgebungsmedium. Curtis beschreibt in "Immersion und Einfühlung" zwei Begriffe, die in einer mediatisierten Welt unmittelbar mit Immersion in Verbindung stehen: Telepräsenz und Virtual Reality (VR).[12] Die Medientheoretikerin Marie-Laure Ryan greift beide Begriffe in "Facetten räumlicher Immersion in technischen Medien" erneut auf und stellt fest, dass an den Begriff der Immersion stets die virtuelle Reise an einen anderen Ort geknüpft ist. Dazu definiert Ryan die zu bereisende Welt als *medialen Raum*, die physische Welt als *Rezeptionsraum*. Maxime für totale Immersion sei nach Ryan, durch gezielte Sinnesreize dem Konsumenten zu suggerieren, er habe den Rezeptionsraum verlassen und sei gänzlich im medialen Raum versunken.[13]

2.3 Immersive Audio

Folgt man der Begriffsdefinition von Marie-Laure Ryan, so gehören zur vollständig gelingenden Immersion neben der visuellen auch die auditiven Reize. Das folgende Kapitel beschäftigt sich mit Definitionen des Begriffs "Immersive Audio" sowie wiedergabesystematische Ansätze, um ein immersives Klangbild reproduzierbar machen zu können.

Altman et.al. definieren den Begriff "Immersive Audio" vorwiegend durch die Umhüllung des Konsumenten durch ein Schallfeld[14], was sich rein nomenklatorisch Ryans Definition von Immersion deutlich annähert. Shivappa et.al. schreiben besagtem Begriff eine große Relevanz im Bezug auf das Immersionserlebnis bei VR-Inhalten zu[15]. Immersion setzt nach diesen Definitionen eine Anzahl von virtuellen oder realen Schallquellen voraus, die diskret auf allen Achsen um den Hörer und durch den Hörer verortbar sind. Diese Ortung kann auf Basis der Lautsprecherwiedergabe durch die richtige Positionierung der Lautsprecher im Raum, auf Basis der Kopfhörerwiedergabe hingegen durch die Spatialisierung von Schallquellen

[11] Hammershøi und Møller, „Binaural Technique — Basic Methods for Recording, Synthesis, and Reproduction“.

[12] Curtis, „Immersion und Einfühlung“. S. 89 ff

[13] Ryan, *Narrative as Virtual Reality*. S. 93 ff

[14] Susal u. a., „Immersive Audio for VR“. S.3

[15] Shivappa u. a., „Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics“. S. 2

durch die Faltung des Ausgabesignals mit einer kopfbezogenen Transferfunktion (HRTF) erzielt werden[16].

Zusammenfassend lässt sich sagen, dass sich immersive Audio ein den Konsumenten in allen Achsen umhüllendes Schallfeld auszeichnet. Relevant ist dies für zwei-und dreidimensionales Bewegtbild sowie VR-, AR- und MR-Inhalte. Es wurden jedoch auch zahlreiche 3D-Produktionen im Musikbereich ohne Bewegtbildanteil produziert[17]. Kombiniert man diese Erkenntnis mit dem ursprünglichen Wortsinn des Begriffs "Immersion" (zur Gänze eintauchen), so ist zur vollständigen Immersion ein alle Dimensionen einnehmendes Schallfeld unabdingbar.

2.4 Wiedergabe von Audiodateien

2.4.1 Kanalbasierte Wiedergabe

Messonier et.al. schreiben dem Begriff des kanalbasierten Wiedergabesystems folgende Eigenschaften zu:

1. Mischungsseitig: Ein Signal auf einer Spur in einer DAW wird diskret einem Kanal zugeordnet.
2. Wiedergabeseitig: Jeder Kanal der abzuspielenden Mischung wird diskret einem Lautsprecher zugeordnet.[18]

So bildet das kanalbasierte Wiedergabesystem die Grundlage unter anderem für herkömmliche Mono-, Stereo- und 5.1-Wiedergabe.

In diesen Eigenschaften erkennen Messonier et.al. einen immensen Nachteil: Die Abspielkompatibilität auf verschiedenen Wiedergabesystemen ist nicht gewährleistet. Eine Mischung im Format 7.1 ist nicht in einem 5.1-Lautsprechersystem optimal reproduzierbar. Ein geläufiges Beispiel geben Messonier et.al. mit den mangelnden Abbildungen von Stereo-Mischungen auf Kopfhörern: Auf Lautsprechern wird das Signal des linken Lautsprecher in einem gleichseitigen Dreieck zwischen Hörer und Lautsprechern in einem Winkel von +-30 Grad wiedergegeben. Auf Kopfhörern hingegen wird es in einem Winkel von +-90 Grad wiedergegeben[19]

[16] Shivappa u. a. S.3
[17] Wisse, „IAN – Immersive Audio Network“.
[18] Messonnier u. a., „Object-Based Audio Recording Methods“.
[19] Messonnier u. a.

2.4.2 Anwendungsbeispiele für Immersive kanalbasierte Wiedergabesysteme

Bezieht man die in Abschnitt 2.1 gewonnene Definition von IA auf kanalbasierte Wiedergabesysteme, bieten die Surround-Formate 5.1 und 7.1 aufgrund fehlender Höhenlautsprecher keine vollständige auditive Immersion. Das Mehrkanalsystem AURO 3D hingegen stützt sich auf folgenden Ansatz: Die bestehenden 5 Kanäle werden zusätzlich zum bestehenden 5.1-System auf einem sogenannten Top-Layer angebracht. Optional ist ein "Voice-Of-God"-Lautsprecher, der direkt über dem Konsumenten angebracht und vertikal nach unten strahlt.

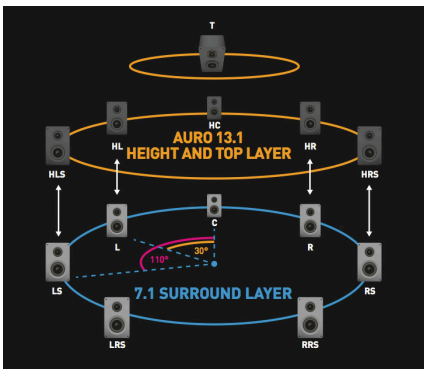

Abbildung 2-3: Anordnung eines Auro3D 11.1-Setup[20]

2.4.3 Objektbasierte Wiedergabe

Ruiz et.al. stellen in "A Description of an Object-Based Audio Workflow for Media Productions" klare Unterschiede zwischen der objektbasierten und der kanalbasierten Wiedergabe heraus. Objektbasierte Wiedergabe gelingt mit den folgenden Komponenten:

1. Eine (mehrkanalige) Audiodatei
2. Metadaten, die diverse Parameter wie Lautstärke oder Koordinaten im Raum für jeden Kanal der Audiodatei zu einem bestimmten Zeitpunkt beinhalten. Diese Metadaten können sowohl in der Audiodatei selbst als auch als Asset gespeichert sein.
3. Einen Prozessor (Rendering and Mastering Unit oder RMU), der die Metadaten lesen kann, sodass die Lautsprecher mit den entsprechenden Signalen angefahren werden können.[21]

Durch das "intelligente" Auslesen einer Mehrkanaldatei und dessen Metadaten mithilfe des Prozessors ist es also möglich, systemagnostisch den gleichen Immersionsgrad zu erzielen.

[20] http://www.bluray-disc.de/blu-ray-news/filme/45077-red-tails-erster-film-auf-blu-ray-disc-mitauro-3d-111-technologie Abgerufen am: 26.10.2017

[21] Gasull Ruiz, Sladeczek, und Sporer, „A Description of an Object-Based Audio Workflow for Media Productions“. S.1

2.4.4 Anwendungsbeispiele für immersive objektbasierte Wiedergabesysteme

Die beiden prominentesten objektbasierten Wiedergabesysteme sind Dolby® Atmos der Firma Dolby® und IOSONO der Firma BARCO. Beide liefern systemproprietäre Software zur Erstellung der Metadaten (in der Regel als Standalone-Software oder als Plugin-Einbindung in eine DAW) sowie Hardware-Prozessor. Als Panorama-Metadaten werden Parameter im kopfbezogenen Polarkoordinatensystem (Horizontaler Winkel, vertikaler Winkel und Entfernung) oder im kartesischen Koordinatensystem über die Koordinaten XYZ definiert[22].

2.4.5 Szenenbasierte Wiedergabe

Das Schallfeldrepräsentationsverfahren "Ambisonics" wurde im Jahr 1970 von Michael Gerzon entwickelt und bietet einen alternativen Ansatz zur Aufnahme und Wiedergabe dreidimensionaler Schallfelder. Es handelt sich dabei um ein koinzidentes Verfahren und teilt ein dreidimensionales Schallfeld in vier Komponenten auf:

1. Omnidirektionaler Anteil (W)
2. Nach X gerichteter Anteil (X)
3. Nach Y gerichteter Anteil (Y)
4. Nach Z gerichteter Anteil (Z)

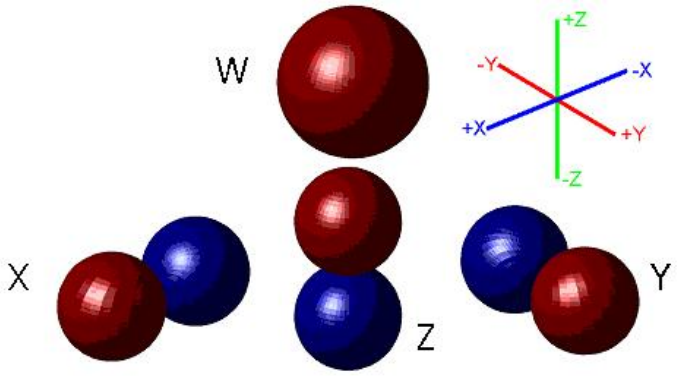

Abbildung 2-4 Aufteilung eines dreidimensionalen Schallfelds in sphärische Komponenten (FOA)[23]

Aufnahmeseitig kann die Aufteilung nach Gerzon mit einem omnidirektionalen (Entspricht Signal W) und drei Mikrofonen mit Achtcharakteristik umgesetzt werden, die in die jeweiligen Raumachsen X, Y und Z zeigen. Da sich jedoch die Membranen der Mikrofone zum gleichen Zeitpunkt im Koordinatenursprung befinden müssten, wird in der Praxis meist auf vier Mikrofone mit Nierencharakteristik in tetraedischer Anordnung zurückgegriffen. Aus diesen Signalen werden die Komponenten wieder herausberechnet.[24]

[22] Weinzierl und Verband Deutscher Tonmeister, *Handbuch der Audiotechnik*. S.88
[23] aus: „True-Multichannel-Mixing - Ambisonics“.
[24] Kronlachner, „Spatial transformations for the alteration of Ambisonic recordings“.

Das Vorliegen dieser Kanäle erlaubt Transformationen wie Rotation, Stauchung, Streckung und Translation des Schallfeldes. Wohingegen eine dreidimensionale kanalbasierte Wiedergabe wie das bereits beschriebene Auro 3D vierzehn Kanäle benötigt, sind für die Wiedergabe von Ambisonics erster Ordnung (FOA) nur vier Kanäle vonnöten. Darüberhinaus bietet Ambisonics alle Vorteile eines koinzidenten Aufnahmeverfahrens wie Monokompatibilität und optimale Phasenkohärenz. Die Nachteile eines koinzidenten Verfahrens wie die eingeschränkte optimale Hörposition bleiben jedoch bestehen.

1990 wurde unter anderem gezeigt, dass Ambisonics in seiner Ordnung mittels Reihenentwicklung von sphärischen Harmonischen skalierbar ist[25]. Abgesehen von Kugelflächenfunktionen der nullten und ersten Ordnung können weitere Kugelflächenfunktionen zur Beschreibung des Schallfelds hinzugefügt werden. Dies resultiert in einer feineren Auflösung des Schallfeldes und einer Vergrößerung der optimalen Hörerposition, aber auch in der Vervielfachung der Lautsprecher bei der Wiedergabe.

2.4.6 Anwendungsbeispiele für immersive szenenbasierte Wiedergabesysteme

Spätestens seit der Akquirierung des Unternehmens "Two Big Ears" durch das Social-Media-Unternehmen "Facebook" ist Ambisonics ein nahezu standardisiertes Container-Format für immersives Audio zu VR.[26] Facebook unterstützt hierbei eine hybride Version aus Ambisonics zweiter Ordnung und kanalbasierter Wiedergabe. Die Ambisonics-Komponente besteht aus acht Kanälen und kann abhängig vom Blickfeld des Zuschauers rotiert werden. Zwei kanalbasierte Signale bieten die Möglichkeit, nicht-positionierte Spuren ("Head-Locked") zu verwenden. Die Video-Plattform "YouTube" unterstützt ebenfalls Ambisonics, allerdings nur in erster Ordnung[27].

Lautsprecherinstallationen sind in der Ambisonics-Domäne eine Ausnahme. Beispiele sind dafür sind das "Loudspeaker Sphere Observatory"[28] an der Universität Vilnius oder die mobile Anwendung "Ambisonics Klangdome"[29].

[25] Sontacchi, „Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen“.

[26] Moore, „Facebook 360 Spatial Workstation, Jibo SDK, and Twilio updates developer console—SD Times news digest“.

[27] Roland, „Ambisonic Mastering: The challenges of sound design in 360° videos“. S.4

[28] Kronlachner, „Loudspeaker Sphere Observatory Vilnius University | matthiaskronlachner.com“. S.1

[29] Jauch und Romanov, „Ambisonic Hits The Road!“ S.1

2.4.7 Hybride Formate

Seit einiger Zeit gibt es Containerformate, die eine simultane Speicherung von kanalbasierten, objektbasierten und szenenbasierten Audiodaten erlauben. Beispielsweise arbeitet Dolby®-Atmos mit einer Mischform aus herkömmlichen Spuren und Objekten mit Metadaten. Ein 7.1 Mix wird hierbei ohne Metadaten, weitere Spuren mit proprietären Panning-Plugin generierten Metadaten an die RMU geschickt.[30]

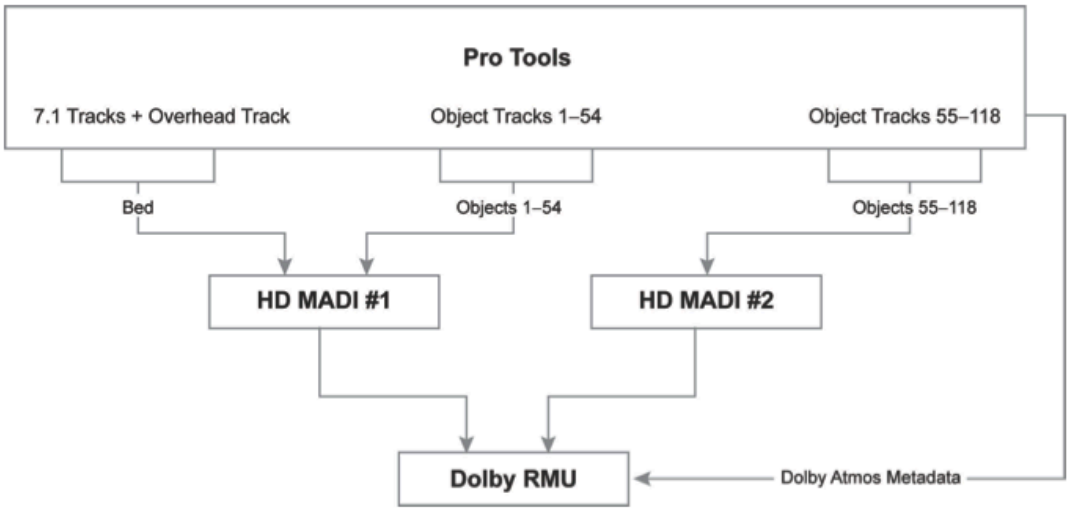

Abbildung 2-5 Signalflussdiagramm für Dolby®-Atmos[31]

Ein weiteres hybrides Format bietet der Container MPEG-H. Als spezifizierter Audio-Container erlaubt er die Übertragung von bis zu 128 Kanälen, 128 Objekten und HOA der 29. Ordnung. Bezeichnend für MPEG-H ist unter anderem das Senden sogenannter Switch Groups, die dem Konsumenten erlaubt, verschiedene Gruppen von Kanälen an- und auszuschalten, die Laustärke einzelner Kanäle zu regeln sowie die Implementierung eines Renderers zum Nutzer-Tracking für VR. Es ist davon auszugehen, dass das Audio Definition Model hier als Träger der Metadaten und als Austauschformat dienen wird[32].

[30] *Authoring for Dolby Atmos® Cinema Sound Manual*.
[31] *Authoring for Dolby Atmos® Cinema Sound Manual*.
[32] Füg u. a., „Design, Coding and Processing of Metadata for Object-Based Interactive Audio“.

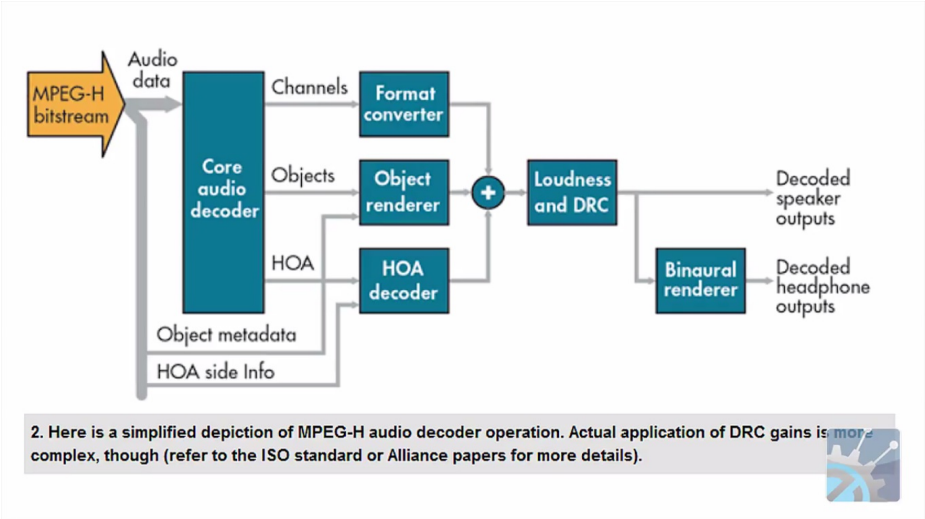

Abbildung 2-6 Signalflussdiagramm des MPEG-H-Containers[33]

2.5 Zusammenfassung

Ein Überblick über die Grundlagen des menschlichen Gehörs und verschiedene Wiedergabeverfahren wurde gegeben. Zusammenfassend lässt sich sagen, dass sich kanalbasierte Wiedergabe durch die diskrete Zuordnung eines Kanals an einen Lautsprecher auszeichnet, was keinerlei Processing benötigt, allerdings auch dementsprechend inkompatibel auf verschiedenen Lautsprecher-Setups ist. Im objektbasierten Wiedergabeverfahren werden zusätzlich zu den Audiodaten Metadaten übertragen, die bestimmte Eigenschaften die Position im Raum zu einem bestimmten Zeitpunkt bestimmen. Diese Metadaten werden von einem separaten Renderer ausgelesen, der das Audiosignal gemäß den Metadaten anpasst und so eine größere Abwärtskompatibilität ermöglicht. Szenenbasiertes Audio verfolgt einen anderen Ansatz. Hierbei wird ein Schallfeld in vier Komponenten aufgeteilt, die sich aus einem omnidirektionalen Anteil und drei Raumachsenanteilen zusammensetzen. Das Vorliegen dieser Signale erlaubt einfache Transformationen mittels einfacher Berechnungen. Im Folgenden wird untersucht, ob das Audio Definition Model diese Wiedergabeverfahren erlaubt und wie diese in der Datenstruktur umgesetzt wurden.

3 Das Audio Definition Model (ADM)

Das folgende Kapitel beschäftigt sich mit der Metadaten-Spezifikation Audio Definition Model. Nach einem kurzen Überblick und ein Einordnung in aktuelle Forschungsentwicklungen werden die einzelnen Komponenten und Parameter des ADM gesichtet, analysiert und auf Anwendbarkeit geprüft.

[33] Jan Fleischmann, „MPEG-H -- ein Audioformat der nächsten Generation (NGA)“.

3.1 Geschichte und Definition

Das Audio Definition Model (ADM) ist Teil des Projektes Orpheus, einer Kooperation internationaler Multimedia-Entwickler, Rundfunkunternehmen und Forschungseinrichtungen, das sich mit der Workflow-Standardisierung von objektbasierten Audioproduktionen beschäftigt. Ziel des von der EU im Rahmen des Horizon 2020-Programms unterstützen Forschungsprogramms ist die Stromlinienformung objektbasierter Audiokreationen und den Aufbau einer Infrastruktur, die eine Ende-zu-Ende-Lösung für immersive Audioinhalte bereitstellt. Teil des Projektes Orpheus ist unter anderem die Entwicklung des bereits erwähnten MPEG-H Containers durch das Fraunhofer IIS.[34]

Das Audio Definition Model ist eine spezifizierte Repräsentation von Metadaten auf XML-Basis und wurde im Rahmen des Projektes Orpheus am R&D-Center der BBC entwickelt[35]. Die Einbettung der Metadaten in einer nach der ITU-BS-2088-0 spezifizierten BW64-Datei[36] erfolgt durch die Erweiterung durch einen sogenannten "Channel Allocation Chunk" (”<chna>").[37] Dort befinden sich Referenzen auf alle einzelnen Kanäle der Audiodatei sowie deren Metadaten.

Erwähnenswert ist hier, dass das ADM ausschließlich die formalisierte Beschreibung der Audiodatei enthält, niemals aber Datenträger des Audiosignals selbst ist.[38] Es dient als Ergänzung zu bereits bestehenden Audiocontainern, welches von designierten RMU's gelesen werden kann. Bestehende, in XML-Sprache eingebettete Metadaten können laut Spezifikation mittels Remapping auf andere Sprachen wie JSON umcodiert werden.[39]

[34] Andreas Silzle, „Orpheus Audio Project: Piloting an End-to-End object-based audio broadcasting chain“.

[35] „Audio Definition Model Software - BBC R&D“.

[36] „BS.2088 : Long-form file format for the international exchange of audio programme materials with metadata“.

[37] „BS.2076 : Audio Definition Model“. S. 2

[38] „BS.2076 : Audio Definition Model“. S. 2

[39] „BS.2076 : Audio Definition Model“, S. 3.

3.2 BW64 und der <chna>-Chunk

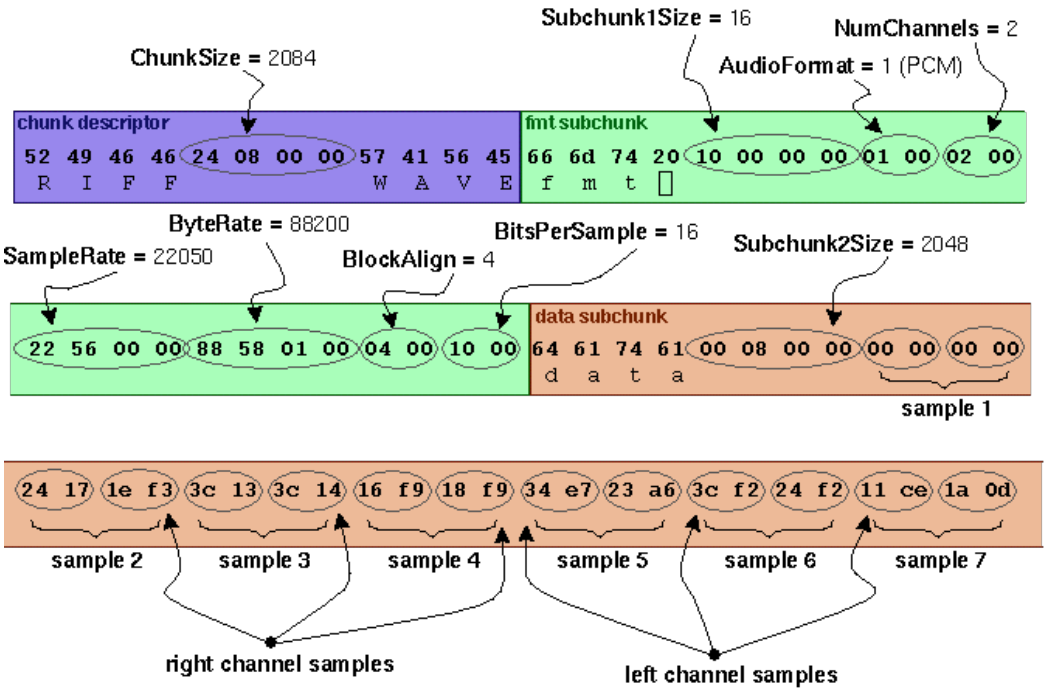

Abbildung 3-1 Die Datenstruktur eines Wav-Files

Gemäß der ITU BS.2088 basiert das dort spezifizierte BW64-Format auf dem WAVE-Format. Als Behältnis von Daten innerhalb eines .wav-Datei dienen sogenannte RIFF-Container. Jeder RIFF-Container beinhaltet sogenannte chunks, die sich durch die Definierung von drei Datentypen auszeichnen:"chunkID", einen Identifikator "ChunkSize", einen ganzzahligen Wert zur Repräsentation der Länge in Bytes und "riffType", die Information selbst. Um als Datenträger für ADM-Metadaten funktionieren zu können, wird die BW64-Datei um zwei chunks erweitert:

- <axml> chunk: Erlaubt die Speicherung und den Transfer von Metadaten in XML-Sprache
- <chna> chunk: Enthält alle Referenzen einer Spur innerhalb der BW64-Datei mittels einmaliger IDs.[40]

Somit sind alle Anforderungen für eine standardisierte Speicherung von ADM-Metadaten in einem Containerformat erfüllt. Wichtig sei hier zu erwähnen, dass das WAVE-Format seit Jahren als Standard zur Speicherung von Audiodaten gilt. Es ist davon auszugehen, dass jeder Tonschaffende damit vertraut ist. Codierung und Speicherung sollten demnach keinen Mehraufwand oder Einarbeitung erfordern.

[40] „BS.2088 : Long-form file format for the international exchange of audio programme materials with metadata“. S. 2

Anhand des SML-Diagramms in Abbildung 3-1 lässt sich die Datenstruktur des ADM in zwei Sektionen aufteilen: Format und Content. Während alle Format-Daten Informationen über bereits vollzogene oder noch nötige Decodierungen, Encodierungen und Renderings enthalten, tragen alle Content-Parameter Informationen über den Inhalt selbst wie beispielsweise die Sprache des Dialogs oder der Voice-Over-Elemente.[41] Beide Sektionen werden im folgenden analysiert.

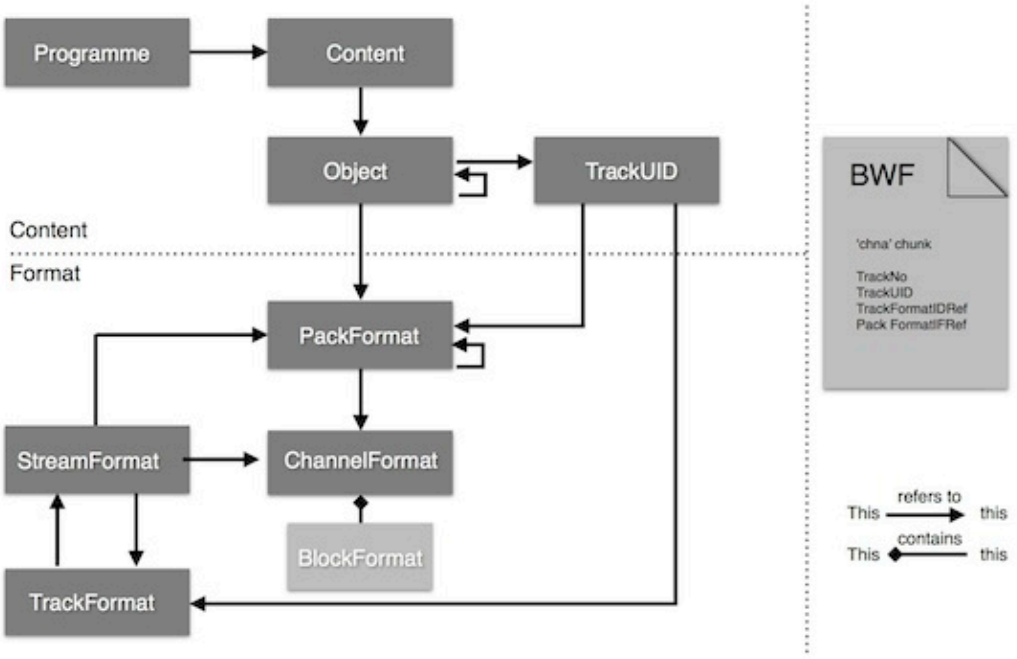

Abbildung 3-2 SML Diagramm des Audio Definition Model[42]

3.3 Sektion: Format

Bei genauerer Betrachtung des SML-Diagramms fällt auf, dass die einzelnen XML-Blöcke in etwa die Verbindungen verschiedener Komponenten einer DAW repräsentieren. TrackFormat repräsentiert eine Sour, ChannelFormat einen Kanal, BlockFormat Automationsparameter, StreamFormat die Codierungsart der Audiodaten und PackFormat repräsentieren Gruppenkanäle. Die Verweise im SML-Diagramm sind in der XML-Datenstruktur mit dem Suffix "IDRef" versehen. Darüberhinaus enthält jeder XML-Block eine eineindeutige ID.

[41] „BS.2076 : Audio Definition Model“, S. 3

[42] Quelle: http://www.bbc.co.uk/rd/publications/audio-definition-model-software Abgerufen am: 31.10.2017

3.3.1 Maßeinheiten für Koordinaten

Sofern Lautsprecher- oder Objektkoordinaten definiert werden, lässt das ADM zwei verschiedene Koordinatensysteme zu. Das Polarkoordinatensystem mit den Parametern "azimuth", "elevation" und "distance" oder das kartesische Koordinatensystem mit den Parametern "x", "y" und "z".

3.3.2 audioTrackFormat

Alle Metadaten im XML-Block audioTrackFormat beschreiben die Daten auf einer einzelnen Spur. Dieser Block beinhaltet folgende Parameter. Der Parameter **audioTrackFormatID** dient zur Cross-Referenzierung von anderen XML-Blöcken. Unter **audioTrackFormatName** wird der Name der Spur gespeichert. **formatLabel** speichert den Namen des Stream-Formates als Zahl, **formatDefinition**: Name des Stream-Formates als Text. Bei **audioStramFormatIDRef** handelt es sich um einen Booleschen Wert, der beantwortet, ob auf ein audioStreamFormat referenziert wird (Siehe Kapitel 3.3.2).

Die Daten innerhalb des audioTrackFormat können zeitvariant sein und mittels Referenzierungen von audioBlockFormat verändert werden.[43]

3.3.3 audioStreamFormat

Falls mehrere Spuren benötigt werden, um ein Signal vollständig zu decodieren, so wird dies in diesem XML-Block gespeichert. Auch hier sind die Namen, Definitionen und Labels durch die Felder **audioStreamFormatName**, **formatLabel** und **formatDefinition** repräsentiert. Zusätzlich sind noch weitere Verweisungs-IDs enthalten:

- **audioChannelFormatIDRef**: Verweis zu audioChannelFormat
- **audioPackFormatIDRef**: Verweis zu audioPackFormat
- **audioTrackFormatIDRef**: Verweis zu audioTrackFormat[44]

3.3.4 audioChannelFormat

Rein nomenklatorisch ähnelt audioChannelFormat dem Datenblock audioTrackFormat. Das Äquivalent zu diesem Datenblock wäre in der DAW-Domäne ein Kanal. Wie die beiden bereits beschriebenen XML-Datenblöcke, beinhaltet audioChannelFormat Informationen über den Namen (**audioChannelFormatName**), eine eindeutige ID (**audioChannelFormatID**) sowie den Typ des Kanals als Text und als Ganzzahl (**typeLabel, typeDefinition**).

[43] „BS.2076 : Audio Definition Model“, S.7 f
[44] „BS.2076 : Audio Definition Model“, S. 9

In der aktuellen Fassung der BS.2076-1 sind fünf verschiedene Werte für typeDefinition verfügbar:

- **Direct Speakers**: Für kanalbasiertes Audio
- **Matrix**: Für matriziertes Audio, beispielsweise MS-Stereofonie
- **Objects**: Für objektbasiertes Audio mit Positionsdaten
- **HOA**: Für szenenbasiertes Audio höherer Ordnung
- **Binaural**: Für binauralisiertes Audio für Kopfhörerwiedergabe
- User Custom: Für eigens spezifizierte Wiedergabetypen (nicht näher definiert)

Darüberhinaus wurden folgende Unterelemente definiert:

- **audioBlockFormat**: Zeitintervall, das zeitvariante Metadaten enthält (siehe dazu Kapitel: 3.3.4 audioBlockFormat)
- **frequency**: Hoch-, Tief-, oder Bandpassfilter für den jeweiligen Kanal, geeignet für den LFE-Kanal in einem Surround-Setup[45]

Hier lässt sich erkennen, dass die ADM-Datenstruktur sowohl kanalbasierte als auch objektbasierte und szenenbasierte Wiedergabe zulässt.

3.3.5 audioBlockFormat

Im XML-Container audioBlockFormat werden einzelne, durch audioChannelFormat beschriebene Samples in einem definierten Zeitintervall repräsentiert. Dies lässt sich grob mit Automationsdaten in einer DAW vergleichen. Neben einer eindeutigen ID (**audioBlockFormatID**) werden in diesem Block der Startzeitpunkt (**rtime**) und die Dauer (**duration**) der Automation festgehalten. Als Maß für rtime und duration gilt: hh:mm:ss:samples.

In Abhängigkeit von der typeDefinition des beschriebenen Kanals werden die nötigenParameter hinzugefügt, die eine kanalbasierte, objektbasierte oder szenenbasierte Wiedergabe ermöglichen. Ist die typeDefinition vom beschriebenen audioChannelFormat "**Direct Speakers**", enthält audioBlockFormat folgende Unterelemente:

- **speakerLabel**: Name des angesteuerten Lautsprechers
- **coordinate = "azimuth"**: Exakte Lautsprecherposition auf der Horizontalebene; Dazu korrespondierend sind zwei Max- und Min-Parameter zum minimalen beziehungsweise maximalen Winkel enthalten.
- **coordinate = "elevation"**: Exakte Lautsprecherposition auf der Frontalebene; Dazu korrespondierend sind zwei Max- und Min-Parameter zum minimalen beziehungsweise maximalen Winkel enthalten.

[45] „BS.2076 : Audio Definition Model“, S. 10 f

- **coordinate = "distance"**: Exakte Distanz von Lautsprecherposition zu Ursprung; Dazu korrespondierend sind zwei Max- und Min-Parameter zur minimalen beziehungsweise maximalen Distanz enthalten.[46]

Ist die typeDefinition vom beschriebenen audioChannelFormat "**Matrix**", enthält audioBlockFormat folgende Unterelemente:

- **gain / gainVar**: Verstärkung eines Kanals mittels Multiplikation mit einer Konstanten / Variablen als Dezimalzahl
- **phase / phaseVar**: Phasenverschiebung eines Kanals mittels Addition mit einer Konstanten / Variablen in Grad
- **delay / delayVar**: Verzögerung eines Kanals mit einer Konstanten / Variablen in Millisekunden[47]

Ist die typeDefinition vom beschriebenen audioChannelFormat "**Object**", enthält audioBlockFormat folgende Unterelemente:

- **cartesian**: boolescher Wert, der über das verwendete Koordinatensystem entscheidet. Ist cartesian = true, wird das kartesische Koordinatensystem, andernfalls das Polarkoordinatensystem. Je nach Wert werden die Paramter azimuth, elevation und distance bzw. x, y und z verfügbar.
- **gain**: Funktioniert analog zu "gain" im Block "Matrix"
- **diffuse**: Beschreibt den Diffus-/Direktanteil eines Objekts. Nähere Angaben werden dazu in der Spezifikation nicht gemacht.
- **channelLock & MaxDistance**: Boolescher Wert, der der RMU mitteilt, ein Objekt am nächsten Lautsprecher zu positionieren. MaxDistance gibt die maximale Distanz zwischen Lautsprecher und virtueller Position an.
- **objectDivergence & azimuthRange**: Erlaubt die positionelle Abweichung eines Objekts auf der Horizontalebene. objectDivergence ist ein Gleitkommawert zwischen 0 und 1, **azimuthRange** der erlaubte Abweichungswert in Grad
- **objectDivergence & positionRange**: Siehe oben, allerdings wird hier bei die Abweichungsspezifizierung entlang der X-Achse angegeben.
- **jumpPosition: & interpolationLength**: jumpPosition ist ein boolescher Wert, der angibt, ob die Position eines Objektes interpoliert werden soll. Falls jumpPosition 1 ist, wird durch den in interpolationLength enthaltenen Gleitkommawert angegeben, über welche Laufzeit die Interpolation erfolgt.

[46] „BS.2076 : Audio Definition Model“. S.12-15
[47] „BS.2076 : Audio Definition Model“. S.15-16

- **zone**: Beschreibt eine bestimmte Zone im virtuellen Raum. Die Unterelemente von **zone** können entweder im kartesischen Koordinatensystem, im Polarkoordinatensystem oder durch vordefinierte Zeichenketten angegeben werden (minX, minY, minZ, maxX, maxY, maxZ oder minAzimuth, minElevation, maxAzimuth, maxElevation oder "Rear Half"). So wird die RMU nicht durch unnötige Prozesse belastet.
- **zoneExclusion**: Boolescher Wert, der die in **zone** definierte Zone vom Rendering ausschließt.
- **screenRef**: Boolescher Wert, der angibt, ob ein Objekt im Verhältnis zum Bildschirm / zur Leinwand gerendert werden soll.
- **importance**: Rating-Wert zwischen 0 und 10. Objekte mit höherem Wert erhalten mehr RMU-Ressourcen.[48]

Ist die typeDefinition vom beschriebenen audioChannelFormat "**HOA**", enthält audioBlockFormat folgende Unterelemente:

- **equation**: Formel zur Berechnung der einzelnen HOA-Komponente
- **order**: Ordnung der HOA-Komponente
- **degree**: Grad der HOA-Komponente
- **normalization**: Definiert die Normalisierung der HOA-Komponente. Es gibt verschiedene berechnungsarten zur Normalisierung von Schallfeldern, die geläufigsten sind FuMa und Ambix.
- **nfcRefDist**: Gibt die relative Distanz eines Lautsprechers zur Nahfeldkompensation an.
- **screenRef**: Analog zu screenRef, wenn audioChannelFormat = "Object" ist.[49]

3.3.6 audioPackFormat

In audioPackFormat sind alle Informationen enthalten, die zusammengehörende Spuren gruppieren. Es lässt sich mit einem Gruppenkanal in einer DAW vergleichen. Hierfür referenziert audioPackFormat durch das Element **audioChannelFormatIDRef** auf die Parameter von audioChannelFormat. Wie alle bereits genannten Blöcke verfügt audioPackFormat über ID und Namen (**audioPackFormatID, audioPackFormatName**) sowie Label und Definition des Gruppenkanals (**typeLabel, typeDefinition**). Analog zu audioBlockFormat in Kapitel 3.3.5 wird auch hier ein importance-Parameter gespeichert, der eventuell bestimmte Signale vom Rendering ausschließt. In Abhängigkeit von der **typeDefinition** des referenzierten audioChannelFormat-Objekts werden weitere Parameter verfügbar. So können beispielsweise

[48] „BS.2076 : Audio Definition Model“. S. 17-20
[49] „BS.2076 : Audio Definition Model“. S. 21-22

alle Komponenten eines HOA-Signals in einem Gruppenkanal zusammengefasst werden. Sie entsprechen in etwa den Parametern in Kapitel 3.3.5[50].

Hier sei erwähnt, dass audioPackFormats ineinander verschachtelbar sind, der XML-Container also sich selbst referenzieren kann. Dies ist beispielsweise nützlich bei HOA-Inhalten, die Ambisonics-Inhalte niedrigerer Ordnung enthalten.

3.3.7 Zusammenfassung

Hernach lässt sich folgende Hierarchiekette ableiten: audioTrackFormat beschreibt das Codierformat der korrespondierenden Samples auf der Spur in der BWF-Datei. audioStream-Format besteht aus den Informationen mehrerer audioTrackFormat-Container. So kann mit einer Referenz auf audioChannelFormat und/oder audioPackFormat der Format- und Codierungstyp der Spuren bestimmt werden. audioChannelFormat beschreibt den gesamten Inhalt einer Audiospur. audioPackFormat fasst mehrere Spuren zu Gruppen zusammen, lässt sich also mit einem Summenkanal in einer DAW vergleichen. Die Konstruktion dieses Datenmodells lässt vermuten, dass während der Konzipierung des ADM auf Anwenderzentrierung fokussiert wurde.

3.4 Sektion: Content

3.4.1 audioTrackUID

Das Element audioTrackUID dient als eineindeutiger Identifikator einer Spur. Auf dieses Element wird beispielsweise im in Kapitel 3.2 beschriebenen <chna>-chunk verwiesen, darüberhinaus stellt es Beziehungen zwischen den Format-XML-Blöcken her.

Die grundlegenden Elemente des Blockes audioTrackUID sind:

- **UID**: Die eineindeutige ID selbst
- **sampleRate**: Die Abtastrate der Spur in Hertz
- **bitDepth**: Die Quantisierungstiefe der Spur in Bit

3.4.1.1 MXF-Integration

Außerdem bietet das ADM an dieser Stelle eine erweiterte Funktionalität mit Dateien im Material Exchange Format (MXF), welches selbst eine erweiterte, bereits vorhandenen Metadatenstruktur aufweist. Diese dienen der Katalogisierbarkeit, Verschlagwortung und Archivie-

[50] „BS.2076 : Audio Definition Model“. S. 23-27

rung in Sendeanstalten. Das MXF kann sowohl Video- als auch Audiodaten enthalten.[51] Dadurch ergibt sich für MXF eine spezielle, dem ADM entgegengesetzte Nomenklatur: Spur bzw. "track" beschreibt ein reines Containerformat, welches Kanäle bzw. "channels" verschiedener Arten enthalten kann. Daher beinhaltet audioTrackUID folgende Unterelemente:

- **audioMXFLookUp**: Gibt Auskunft, ob die ADM-Daten in einer MXF-Datei geschrieben sind.
- **audioTrackFormatIDRef**: Referenz zu Informationen im Block audioTrackFormat
- **audioPackFormatIDRef**: Referenz zu Informationen im Block audioPackFormat

Enthält eine MXF-Datei ADM-Daten, wird diese um die folgenden Unterelemente erweitert:

- **packageUIDRef**: Referenz an ein MXF-Paket
- **trackIDRef**: Referenz an eine Spur nach MXF-Nomenklatur
- **channelIDRef**: Referenz an einen Kanal nach MXF-Nomenklatur

Durch den vielfachen Gebrauch des MXF-Formats im Broadcasting-Bereich ist es von Vorteil, vorgefertigte Schnittstellen im ADM zu haben, die eine lückenlose Integration in MXF-Dateien erlauben.[52]

3.4.2 audioObject

Der XML-Block audioObject verknüpft audioPacks und andere, inhaltlich relevante Daten. Es lässt sich DAW-seitig mit den Daten eines verschachtelten Projektes vergleichen. Dafür werden die in Kapitel 3.4.1 beschriebenen audioTrackUIDs verwendet. audioObject gibt keinerlei Auskunft darüber, ob bestimmte Spuren als Objekte gerendert werden oder nicht, auch wenn dies der Name implizieren könnte. Hier werden inhaltliche Informationen verknüpft. Es können verschiedene audioObject-Blöcke ineinander verschachtelt werden. Neben ID und Name des Blocks werden außerdem die Startzeit und die Laufzeit des Blocks definiert(**audioObjectID, audioObjectName, start, duration**). Der Parameter **importance** gibt die Rendering-Priorität an, ähnlich zu audioPackFormat. Außerdem kann durch den Wahrheitswert im Parameter **interact** die Interaktionsmöglichkeit durch den Zuschauer an- oder abgeschaltet werden. audioObjects können andere audioObjects dynamisch in der Lautstärke verändern (bekannt als Sidechaining). Dies kann mit dem Parameter disableDucking aktiviert oder deaktiviert werden.

[51] Oezturan, „Austausch von Metadaten in der Broadcasting Branche“.
[52] „BS.2076 : Audio Definition Model“. S. 39-40

3.4.2.1 Interaktionsparameter

Das ADM bietet Parameter, die Auskunft über die Interaktionsmöglichkeit zwischen dem Konsumenten und Objekten geben. Hierfür sind zunächst folgende boolesche Werte definiert:

- **onOffInteract**: Zur Stummschaltung verschiedener Objekte durch den Zuschauer
- **gainInteract**: Zur Lauststärkeanpassung verschiedener Objekte durch den Zuschauer
- **positionInteract**: Zur Positionsveränderung verschiedener Objekte durch den Zuschauer

Je nach Einstellung der oben genannten Elemente, ergeben sich folgende Sub-Parameter:

- Falls gainInteract = 1:
  - min: Mindestwert, den der Endkonsument als Faktor zur Lautstärkeanpassung einstellen kann
  - max: Maximalwert, den der Endkonsument als Faktor zur Lautstärkeanpassung einstellen kann
- Falls positionInteract = 1 und falls das Polarkoordinatensystem aktiviert ist, sind folgende Parameter zuweisbar:
  - **azimuth min / max**: Minimaler / Maximaler Versatz auf der Horizontalebene, den der Zuschauer an einem Objekt in Relation zur eigentlichen Position einstellen kann.
  - **elevation min / max**: Minimaler / Maximaler Versatz auf der Vertikalebene, den der Zuschauer an einem Objekt in Relation zur eigentlichen Position einstellen kann.
  - **distance min / max**: Minimaler / Maximaler Distanzversatz, den der Zuschauer an einem Objekt in Relation zur eigentlichen Position einstellen kann.
- Falls positionInteract = 1 und falls das kartesische Koordinatensystem aktiviert ist, sind die oben genannten Daten als X, Y oder Z-Koordinaten verfügbar.

3.4.3 audioContent

Der XML-Block audioContent speichert rein inhaltliche Informationen und gibt Auskunft über Titel (**audioContentName**) oder Sprache (**audioContentLanguage**) in einem Subelement des gesamten Programms (Eine Spur, Gruppe von Spuren etc.). Darüberhinaus sind Lautheitsparameter zur Lautheit des Subelements gespeichert.

- **audioObjectIDRef**: Referenzvariable zum korrespondierenden audioObject-Block
- **loudnessMetadata**: Siehe dazu Kapitel 3.4.3.1

- **dialogue**: Variable, die den Auskunft über den Dialoginhalt des Programmelements enthält. Ist dialogue=0, ist kein Dialog im Element enthalten. Ist dialogue=1, ist in dem Element ausschließlich Dialog enthalten. Ist dialogue=2, besteht das Element aus Dialog und anderen Elementen (Musik, Sound-Effekte etc.). In Abhängigkeit dieses Parameters werden andere Parameter verfügbar, die Auskunft darüber, welche Art von (nicht-) stimmanteiligen Inhalten enthalten sind (Voice-Over, szenischer Dialog, Musik, Sound-Effekte etc.)[53]

3.4.3.1 Lautheitsparameter

In Abhängigkeit des Ausstrahlungslandes bestehen für verschiedene Kontinente unterschiedliche Lautheitsnormen. Seit 2012 gilt in der EU die Norm EBU R128, andere Staaten folgen hingegen der ATSC A/85.[54] Zur Beschreibung der Lautheit einzelner Programmelemente wurden folgende Parameter deklariert:

- **loudnessMethod**: Die verwendete Lautheitsmessmethode
- **loudnessRecType**: Die empfohlene Lautheitsnorm (bspw. "R128")
- **loudnessCorrectionType**: Die verwendete Lautheitskorrekturmethode
- **integratedLoudness**: Der integrierte Lautheitswert in LUFS
- **loudnessRange**: Die Lautstärkedynamik des Elements in LU
- **maxtruePeak**: Der maximale Peak-Wert des Elements in dB
- **maxMomentary**: Der maximale Lautheitswert, gemessen über ein Intervall von 400 ms in LUFS
- **maxShortTerm**: Der maximale Lautheitswert, gemessen über ein Intervall von 3 s in LUFS
- **dialogueLoudness**: Die Lautheit des Sprachanteils eines Elements in LUFS[55]

3.4.4 audioProgramme

Der XML-Block audioProgramme vereint mehrere audioContent-Blöcke und beschreibt den Inhalt eines gesamten Ausstrahlungsprogrammes. Auch dieser Block enthält die in Kapitel 3.4.3.1 genannten Lautheitsparameter, hier beziehen sich jedoch auf die Lautheit des gesamten Programms und nicht einzelner Unterelemente (**loudnessMetadata**). Die Parameterstruktur ähnelt der von audioObject und beinhaltet ID, name, Startwert, Endwert und Lauf-

[53] „BS.2076 : Audio Definition Model“, S. 31-33.
[54] Tischmeyer, „EBU-Norm R128 – Die leise Revolution der Pegelmessung“.
[55] „BS.2076 : Audio Definition Model“. S. 34-35

zeit des Programms (**audioProgrammeID, audioProgrammeName, start, end**). Darüber wird hier die Sprache des Programmblocks angegeben (**audioProgrammelanguage**).[56]

3.4.4.1 audioProgrammeReferenceScreen

Im XML-Block audioProgramme können außerdem Daten festgehalten werden, auf welchem Gerät der zu vertonende Bildinhalt während der Mischung angeschaut wurde. Diese Metadaten können beispielsweise bei der Reproduktion ausgelesen und die Mischung beim Endkonsumenten dementsprechend angepasst werden.

- **aspectRatio**: Bildseitenverhältnis des Abspielgeräts
- **screenCentrePosition**: Wie bei den meisten Positionsparametern im ADM kann die Bildmittenposition durch die Parameter azimuth, elevation und distance im Polarkoordinatensystem oder durch XYZ-Koordinaten im kartesischen Koordinatensystem angegeben werden.
- **screenWidth**: Bildbreite, angegeben entweder durch azimuth (in Grad) im Polarkoordinatensystem oder eine Distanz auf der X-Achse im kartesischen Koordinatensystem.[57]

3.5 audioFormatExtended

Beim XML-Block audioFormatExtended handelt es sich um die Elternklasse aller in Kapitel 3 beschriebenen Elemente. Außerdem wird hier die aktuelle Version des ADM deklariert.[58]

3.6 Zusammenfassung

Anhand von Kapitel 3 lassen sich folgende Erkenntnisse gewinnen. Mittels der ADM-Metadaten ist man in der Lage, immersives Audio innerhalb einer BW64-Datei eine detaillierte Informationen zu vermitteln, die in ihrer Struktur alle Daten einer DAW enthalten. Spuren können kanalbasiert wiedergegeben werden, zu einem Objekt mit Metadaten gemacht werden oder Teil einer HOA-Komponente werden (audioTrackFormat, audioStreamFormat, audioChannelFormat). Spuren lassen sich zu einem Bed oder einer HOA-Komponente gruppieren (audioPackFormat). Spuren und ihre Parameter können entlang der Zeitachse verändert, umgemünzt und anders verwendet werden (audioBlockFormat) und gleicht so Automationsdaten. All diese Funktionen sind der XML-Sektion "Format" zugeordnet.

Neben der rein technischen Umschreibung von Audiodaten durch die Sektion Format, lässt das ADM auch die inhaltliche Umschreibung durch die Sektion "Content" zu. Die beiden Sek-

[56] „BS.2076 : Audio Definition Model“. S. 36-37
[57] „BS.2076 : Audio Definition Model“. S. 39-40
[58] „BS.2076 : Audio Definition Model“. S. 40-41

tionen werden durch Parameter des XML-Blocks audioObject verlinkt. Der Block audioContent beschreibt den Inhalt einzelner Komponenten einer Audiodatei, der Block audioProgramme hingegen den Inhalt der gesamten Audiodatei. Zusätzliche Parameter erlauben die komponentenweise und programmweise Lautheitsanalyse und geben Auskunft über die Sprache des Sendematerials. Weitere Schnittstellen erlauben die Verheiratung des ADM mit einer MXF-Datei, was die Funktionalität für Broadcasting und Archivierung erheblich erweitert.

4 Konzipierung und Durchführung der Experteninterviews

Das folgende Kapitel beschäftigt sich mit der Konzeption der empirischen Erhebung, worauf Audio Designer, die täglich mit immersive Audio arbeiten, Wert legen. Als empirische Erhebungsmethode wurde hierfür das Experteninterview gewählt. Das folgende Kapitel gibt einen Einblick in das Experteninterview als empirische Erhebungsmethode selbst, die Vorgehensweise bei der Auswahl der Experten, und die Konzipierung des Fragebogens als Leitfaden für die Interviews.

4.1 Das Experteninterview als empirische Erhebungsmethode

Im Rahmen eines Experteninterviews wird vorausgesetzt, dass der Interviewte Expertise in einem bestimmten Arbeitsfeld vorweisen kann. Dieses Wissen, beziehungsweise diesen erheblichen Wissensvorsprung gegenüber konkurrierender Mitbewerber in diesem Metier kann er durch Erfahrung, Didaktik und Reflexion erworben haben.[59] Bogner zufolge zeichnet sich das Expertenwissen durch die Bestätigung und Anerkennung in der Praxis aus. Ihr durch ihr Wissen gestütztes Handeln verfügt also über eine positive Realweltbestätigung.[60]

Im Rahmen dieser Untersuchung ergaben sich folgende Umstände, die es zu berücksichtigen galt: Zum einen kann immersive Audio an eine Vielzahl von Medien geknüpft sein, die sich sowohl in ihrer Darreichungsform als auch in ihrer grundlegenden Funktionsweise unterscheiden. So ist die 3D-Musik in ihrer Darreichungsform als separat funktionierendes Medium hervorzuheben. Sie muss in der Regel ohne visuellen Anteil immersiv wirken.

Im Bereich der Bewegbildmedien ist, wie in Kapitel 1 herausgearbeitet, immersives Audio für die Immersion des gesamten Mediums unabdingbar. Hier stellen sich jedoch folgende Unter-

[59] Kaiser, *Qualitative Experten- Interviews Konzeptionelle Grundlagen und praktische Durchführung*. S. 36

[60] Alexander Bogner, *Interviews mit Experten - Eine praxisorientierte Einführung*. S. 14

scheidungsfragen: Ist der Film Teil eines größeren Werks, wie beispielsweise einer Rauminstallation oder einer Ausstellung? Ist der Film VR oder auf einer Projektionsfläche wiedergegeben? Gibt es 3D-Anteile? Gibt es interaktive Elemente?

In Anbetracht dieser Vielzahl an Unterscheidungsfragen lässt sich zusammenfassen, dass je nach Kontext des Mediums die Expertise im Bereich Immersive Audio unterschiedlich stark ausfällt. Aus diesem Grund wurde sich dazu entschieden, aus möglichst jedem dieser Bereiche einen Experten zu befragen. In der anschließenden Auswertung werden die voraussichtlichen Schnittmengen und Divergenzen herausgearbeitet und im abschließenden Kapitel mit den Funktionsweisen des ADM verglichen.

4.2 Konzipierung des Leitfadens

Gemäß Bogner erfüllt ein Leitfaden während eines Experteninterviews zwei Funktionen. Zunächst dient er während der Erhebungssituation als Stütze und beugt der Tendenz, in ein loses und unstrukturiertes Gespräch abzudriften, vor. Darüber hinaus dient er während der Konzeptionsphase als Strukturierungswerkzeug.[61] Hier ist zu erwähnen, dass der Leitfaden nicht die Funktion eines starren Korsetts beim Interview erfüllen soll und eine gewisse Flexibilität stets gewahrt bleibt. Vor diesem Hintergrund ist es von höchster Wichtigkeit, das gesamte Thema der Erhebung in einzelne Blöcke zu unterteilen. Dieser Vorgang ist rekursiv zu wiederholen, bis sich besagte Themenblöcke sich nicht mehr unterteilen lassen und sogenannte Forschungsfragen entstehen.[62]

Im Hinblick auf die Zielsetzung dieser Arbeit dienen zur Konzipierung des Leitfadens die gewonnenen Erkenntnisse aus Kapitel 2. Hieraus lassen sich folgende Themenblöcke ableiten:

1. Zusammensetzung der Mix-Elemente hinsichtlich der verwendeten Wiedergabeformate
2. Finales Datenformat
3. Transkompatibilität
4. Immersionserlebnis des Konsumenten

4.2.1 Konkretisierung der Forschungsfragen

Zielsetzung der Arbeit ist, die Bedürfnisse der Experten, in diesem Fall Audio Designern, hinsichtlich der Funktionalität bei der Kreation von immersiven Audioinhalten zu untersuchen.

[61] Alexander Bogner. S. 27-28

[62] Kaiser, *Qualitative Experten- Interviews Konzeptionelle Grundlagen und praktische Durchführung*. S. 54

Hier gilt es, für einen möglichst umfangreichen Erkenntnisgewinn, die Forschungsfragen als Interviewfragen zielführend zu formulieren. Der Leitfaden befindet sich im Anhang.

Im Themenblock 1 wird gezielt nach den verwendeten Elementen während des Kreationsprozesses gefragt. Die zugrundeliegende Erkenntnis soll sein, ob es sinnvoll ist, momentan noch nischenhaft verwendete Ambisonics-Decoder oder Objekt-Renderer zu verwenden. Die Frage nach Interaktivität bleibt optional.

Themenblock 2 fragt nach dem aktuellen Exportformat, in dem der Audio Designer seinen Mix exportiert. Durch tiefergehende Fragen soll herausgefunden werden, ob es dabei standardisierte Prozeduren gibt oder ob je nach Projektaufwand und Projektanforderungen das Endformat variiert.

Themenblock 3 fragt nach der Transkompatibilität der immersiven Mischungen. Da die Kreation von immersiven Inhalten in aller Regel von mehr als einer Person und mehreren Departments (Video, Audio, Konzeption etc.) abhängt, geht dieser Themenblock auch auf Einflüsse auf die Transkompatibilität ein, die möglicherweise nicht beeinflussbar sind, sobald die Endmischung weitergereicht wurde. Ein Szenario wäre folgendes: Ein externer Toningenieur mischt Ton für einen Film im Auftrag einer Sendeanstalt. Er schickt eine WAV-Datei an die Sendeanstalt, welche die gelieferte Datei sendefertig konvertiert. In diesem Konvertierungsprozess ist der Toningenieur nicht mehr involviert.

Themenblock 4 dient als Abschluss des Interviews und geht auf die der Mischung zugrundeliegende Information ein. Hierbei sind vor allem persönliche Erfahrungen von Interesse, beispielsweise, wovon die zugrundeliegende Immersion abhängig ist. es ist davon auszugehen, dass die Experten durch ihre tägliche Arbeit mit immersivem Audio für diesen Themenkomplex sensibilisiert sind.

Es sei erwähnt, dass der konzipierte Leitfaden während des Interviews nicht als starres Korsett und Abarbeitungsliste diente, sondern eine Möglichkeit der losen Gesprächsführung bot. Sobald der Interviewte das Gespräch in eine Richtung lenkte, die interessant erschien, wurde bei Bedarf vom Leitfaden abgewichen. Weiterhin wurden je nach Interview die verschiedenen Themenblöcke hinsichtlich des Zeitumfangs unterschiedlich gewichtet. Sobald ein Interviewter in einem Themenblock mehr zu sagen hatte, wurde dies gestattet, da der Erkenntnisgewinn größer erschien.

4.3 Auswahl der Experten

Insgesamt wurden vier Interviews durchgeführt. Ein Interview erfolgten im persönlichen Gespräch, drei mittels einer Skype-Aufnahme.

4.4 Aufbereitungsmethode

Alle Interviews wurden digital als Audiodatei aufgenommen und gespeichert. Im Anschluss wurden alle Interviews transkribiert. Hierbei wurde auf die Transkription unvollständiger und im Nachhinein neu angesetzter Aussagen verzichtet. Füllwörter, die im Gespräch entstanden, wurden ebenfalls nicht transkribiert, sofern sie inhaltlich keinen Mehrwert beitrugen. Ferner wurden längere Pausen und Räuspern nicht transkribiert. Lediglich Lachen wurde mit "(lacht)" abgekürzt. Falls Namen genannt wurden oder über Sachverhalte gesprochen wurde, über die der Interviewte im Vorhinein mündlich darum gebeten hat, dass sie im Rahmen der Thesis nicht veröffentlicht werden, so sind betroffenen Stellen in der Audiodatei zensiert und in der Transkription geschwärzt. Falls der Betroffene selbst im Rahmen des Interviews anonymisiert werden wollte, so wurde sein Name durch ein zufällig generiertes Kürzel und die Informationen bezüglich seiner Beschäftigung unkenntlich gemacht.

4.5 Auswertungsmethode der Experteninterviews

Im Rechercheverlauf der Arbeit wurde deutlich, dass es kein eineindeutiges Auswertungsverfahren für Experteninterviews gibt. Dies liegt einerseits daran, dass das Material, in diesem Fall die durchgeführten Interviews in transkribierter Form, in der Regel von Forschungsarbeit zu Forschungsarbeit divergiert und aus diesem Grund kein allumfassender Algorithmus abgeleitet werden kann. Andererseits hängt die Analyse grundsätzlich davon ab, welche Forschungsfrage gestellt wurde und ob der Erkenntnisgewinn zur Informationsverdichtung oder zur Theoriegewinnung dient.[63] Daher ist es ratsam, für jede Forschungsarbeit mit anteiligen Experteninterviews eine individuelle Herleitung zur Auswertungsmethode anhand bereits bewährter Leitfäden zu formulieren.

Im Hinblick auf die Forschungsfrage dieser Arbeit liegt die Informationsgewinnung im Vordergrund. Eine Theorie soll nicht abgeleitet werden. Der Auswertungsleitfaden zur Informationsgewinnung nach Bogner legt folgende Arbeitsschritte nahe:

1. Fragestellung und Materialauswahl
2. Aufbau eines Kategoriensystems
3. Extraktion
4. Aufbereitung der Daten
5. Auswertung[64]

Kaisers Begriff der Kodierung legt eine ähnliche Vorgehensweise nahe. Im Hinblick auf die Auswertung unter dem Dogma der qualitativen Inhaltsanalyse empfiehlt Kaiser, die während

[63] Alexander Bogner, *Interviews mit Experten - Eine praxisorientierte Einführung*. S. 72 ff
[64] Alexander Bogner. S. 73-75

der Konzipierung des Leitfadens entwickelten Kategorien zu nutzen, um die Aussagen verschiedener Experten zu einem eindeutigen Sachverhalt gegenüberzustellen.[65]

Hierbei sei jedoch erwähnt, dass im Rahmen dieser Arbeit die Gegenüberstellung der Expertenaussagen nicht zur Polarisierung oder zur Beschreibung der Komplexität des Themas immersive Audio dient. Sobald sich Aussagen widersprechen, ist dies kein Zeichen dafür, dass sie unterschiedlicher Meinung sind. Es geht in bestimmten Themenblöcken nicht um die Meinung der Experten, sondern um ihre Herangehensweise an immersives Audio. Sofern sich Aussagen konterkarieren, sollte dies der übergreifenden Fragestellung nicht die Integrität nehmen, sondern die Notwendigkeit besagter Fragestellung nur bestärken.

5 Auswertung der Experteninterviews

Zur besseren Einordnung der Interviews folgt eine Darstellung der Experten und ihrer momentanen Tätigkeit im Bereich des immersiven Audios.

Interview	Kürzel	Name	Tätigkeit
I1	FS	Felipe Sanchez	Inhaber von klingklangklong GbR, Toningenieur für immersive Klanginstallationen im Raum
I2	MR	Martin Rieger	Freier Toningenieur für VR
I3	BM	Benedikt Maile	Freier Toningenieur für Pop-Musik mit Produktionserfahrung für Musikmischungen in Auro 3D
I4	AM	Andreas Mühlschlegel	Freier Mischtonmeister für Spielfilme und immersive Klanginstallationen

5.1 Themenblock 1: Zusammensetzung der Mix-Elemente

Im Hinblick auf die verschiedenen Mischungselemente ist innerhalb der befragten Expertengruppe klar zu konstatieren, dass vorwiegend mit Mono-, Stereo-Schallquellen oder Ambisonics-codierten Szenen gearbeitet wird. FS macht den Bedarf an Objekten von dem Bedarf an Interaktion abhängig.

[65] Kaiser, *Qualitative Experten- Interviews Konzeptionelle Grundlagen und praktische Durchführung*. S. 102

> *"FS: Vor allem, was bei sehr wichtig ist: Haben wir einen interaktiven Anteil oder nicht? Weil: Normalerweise, wenn es linear ist, wenn wir ein Multikanal-immersives Raumerlebnis haben, für ein Museum oder einen Showroom oder was auch immer, wenn es linear ist, sehen wir keinen Grund, anders als kanalbasiert auszuspielen. Das heißt, da gibt es ein interleaved WAV-File, der einfach die X Kanäle hat, und die werden im Wiedergabesystem ausgegeben."*[66]

Sobald Interaktion gefordert ist, wird ein eigener Player und Renderer entwickelt. Metadaten und Audiodaten werden dabei von zwei separaten Geräten übertragen.

> *"FS:[…] Also diese objektbasierten Informationen sind immer sehr cool, denn das sind bisher 2 verschiedene Informationen, die von 2 verschiedenen Orten kommen. Einmal macht der MAX die Orientierung, einmal macht der Player die Audio-Verteilung. Schön wärs, wenn es von einer Quelle kommen würde.*
>
> *DS: Also ihr fragmentiert momentan noch die Koordinaten und das eigentliche Audio.*
>
> *FS: Ja genau. Und ich kann mir auch so ein paar Momente vorstellen, wo ich denke: Ach, wär cool wenn das drin ist, aber das kann ich dir gerade im Augenblick nicht sagen. Wir sind da zu sehr eingegrooved in unser System.*[67]*"*

Da MR in der Regel sowohl in während des Drehs als auch in der Postproduktion für Tonaufnahme und -mischung verantwortlich ist, hat sich bei ihm das folgende Setup herauskristallisiert:

> *"MR: Also ich will es jetzt nicht mein eigenes Setup nennen, aber ich habe es mir überlegt und hab gemerkt, das funktioniert sehr gut, so wie ich auch in der Postproduktion arbeite, dass ich einerseits versuche, möglichst isolierte Mono-Schallquellen aufzunehmen[…] Aber ich versuche, am Set Signale zum Beispiel mit Ambisonics Mikrofonen mitzunehmen, die bieten sich einfach an, weil die kann ich gut unter der Kamera verstecken."*[68]

Im Falle der Produktionen von BM sind die Auro-3D-Versionen der verschiedenen Stücke ausnahmslos bereits bestehende Stücke, die für die Stereo-Wiedergabe produziert wurden. Bei einer Auro 3D-Mischung wurden die bestehenden Signale lediglich im zusätzlich gewonnenen Raum verteilt.

---

[66] Sanchez, I1. S.2 Z. 27-37
[67] Sanchez. S. 6, Z. 15-25
[68] Rieger, I2. S. 2 Z. 33- S. 3 Z. 9

> *"BM: Also bisher war es so, alle Sachen die ich gemacht habe, waren kanalbasiert. Das waren Mono-Aufnahmen, ganz wenig Stereo-Aufnahmen und habe die einfach hochgemischt auf 9.1 in den allermeisten Fällen. Und auch nur so angelegt, dass im Endeffekt eine Verteilung der Signale stattfindet, ein ähnliches Bewusstsein wie im Stereo."*[69]

Gründe für diese Wahl liegen in der Funktionalität des kanalbasierten und szenenbasierten Audios, da es maximal einen Ambisonics-Decoder gibt, der nicht dynamisch Signale prozessieren muss. Weiterhin wurde angegeben, dass eine objektorientierte Kreationsumgebung mit RMU nicht erschwinglich sei.[70]

> *"AM: Es war eine Mischung aus kanalbasiert und szenenbasiert. […]Das war letztendlich einfach eine Wahl der Mittel, weil es einfach geht mit dem szenen- und kanalbasiert."*[71]

Es lässt sich erkennen, dass objektbasierte Technologien durch Preis- und Aufbauhürde nicht zu den hier befragten Kreativen durchgedrungen ist. Weiterhin ist auch bei objektbasierten Setups eine Lernkurve, die es zu überwinden gilt. Kanalbasierte Produktionen haben stets Vorrang, der Wille (oder bei VR-Produktionen eher die Notwendigkeit) szenenbasiertes Audio zu verwenden, ist jedoch da. bei der Notwendigkeit wird entweder auf eigens konzipierte oder durch einen Hardware- oder Software-Hersteller vorgegebene Kanalbelegungen zurückgegriffen:

> *"AM: Und zwar handelt es sich dabei um ein absolut proprietäres Custom-Setup bei einem Auto. Oft ist es so, dass keine ideale Speaker-Verortung möglich ist, aufgrund des begrenzten Platzes. Das heißt die Speaker sind da, wo sie gut verbaut werden können, aufgrund ihrer eigenen Dimensionierung."*[72]

Weiterhin könnte die fehlende Verbreitung ein Grund für die Nicht-Nutzung von Objekten sein. Eine objektbasierte Produktion lässt sich in ihrem vollen Funktionsumfang nur auf einem anderen Gerät abspielen, wenn eine entsprechende RMU verbaut ist. Das ist sie in den meisten Fällen nicht. Darüberhinaus sind kanalbasierte Formate so etabliert, dass die Kreateure sich darauf am besten auskennen. Aufgrund der über die Jahre erworbenen Expertise wurde im Interview mit BM erkennbar, dass er seine inhaltliche Herangehensweise dem System unterordnet, nicht umgekehrt:

> *"BM: […]Ich mag am kanalbasierten, dass ein ganz klares Format ist. Das ist super einfach, du stellst deine Boxen auf im richtigen Winkel und in der richti-*

[69] Maile, I3. S. 2, Z. 24-28
[70] Maile. I3 S. 2, Z. 10. ff
[71] Mühlschlegel, I4. S. 2, Z. 25-28
[72] Mühlschlegel. S. 3, S, Z. 12-16

gen Positionierung. Mann kann Dinge breiter und schmaler machen, weil es halt einen anderen ästhetischen Effekt hat. Ich verstehe am objektbasierten natürlich, dass es in einen Raum gerendert werden kann, exakt so, wie man es gehört hat. Das wäre schon praktisch, ja. Da habe ich aber zu wenig Erfahrung."[73]

"BM: Für mich heißt es immer: Form follows function. Und in dem Zusammenhang heißt es für mich, dass es niemals um das System geht, sondern es geht immer nur um den Inhalt der Produktion.[…] Das heißt für mich geht es immer nur um den Inhalt der Musik, dass man den darstellt, und das System einen Mehrwert für die Musik darstellt, nicht umgekehrt.[74]*"*

Abschließend lässt sich also sagen, dass sich aus der finanziell erschwerten Zugänglichkeit und der fehlenden Marktetablierung objektbasierte Produktionen noch nicht durchgesetzt haben. Szenenbasierte Anwendungen sind laut MR in der VR-Produktion unabdingbar, bei allen anderen Teilnehmern werden diese jedoch nur innerhalb proprietärer Kanalsysteme eingesetzt.

5.2 Themenblock 2: Finales Datenformat

Beim Export einer finalen Ausgabedatei wird im Augenblick bei allen Experten auf eine Ausspielung aller Kanäle als Mono-Dateien oder in einem Interleaved-File gesetzt. Im Falle FS's wird auf eine eigens entwickelte Kombination aus einer Abspiel-Software und zusätzlichen MAX MSP-Patches zum Tracking oder anderen Renderings mit Interaktions-Input genutzt:

"FS: […]die Signale werden hin und hergeschickt, was auch immer da gebraucht wird. Ein UDP-Port, kann man auch in TCP schicken, aber im Endeffekt machen wir OSC-Signale in OSC-Paketen, wo wir per OSC die Koordinaten vom Raum schicken, von wo der Sound kommen kann. Nehmen wir einen Raum, und ein Sound muss einer Person folgen, das heißt, es wird getrackt, das heißt, wir kriegen von dem System wo diese Person ist in XYZ-Koordinaten. Und dann schicken wir auf unsere eigenen gebauten Player: An diesen Koordinaten, spiel diesen Sound."[75]

Im Falle MR's hängt das Exportformat stark von dem finalen Abspielgerät ab. Je nach gewählter Applikation und Brillentyp wird ein gesamtes Dateienpaket von gerendertem Video

[73] Maile, I3. S. 6, Z. 33 - S. 7, Z. 5
[74] Maile. S. 3, Z.19 ff
[75] Sanchez, I1. S. 3, Z. 15-25

inklusive Audio exportiert. Dies liegt, laut MR, an der fehlenden Sensibilisierung auf Kundenseite (bspw. Filmproduktionsfirma) und Konsumentenseite:

> *"MR: Dann mache ich es so, dass ich das Video halt selber encodier, denen schicke und sage: Ja hier, ihr müsst das mit dem Player mit dem Device anschauen und dann funktioniert's. Weil dann habe ich die Garantie: Das Video, was ich gemacht habe, das funktioniert."*[76]

> *"MR: Also da ist eben das Problem, dass das Thema so neu ist für alle, dass ich sie da lieber bei der Hand nehme und halt wirklich ein ganz finales File gebe, was bedeutet, dass sich halt immer ein Gigabyte große oder zwei Videos hochladen muss. […] Aber das ist gerade für mich so die beste Lösung, weil dann kann ich halt wie gesagt sicherstellen, dass das alles so läuft wie ich mir das vorstelle."*[77]

Wie in Kapitel 6.1 bereits zusammengefasst, finden Exporte mit Objekt-Metadaten bisher noch keinerlei Verwendung. In diesem Zusammenhang sind die Experten einer Zentralisierung von Metadaten in einer Datei nicht zwingend wohlgesonnen. Im Falle FS wurde bereits eine eigene Lösung entwickelt, die beispielsweise ein dynamisches Rendering erlaubt:

> *"FS: Also es gibt schon Momente, wo wir sagen, wir liefern einfach nur Content, und der Player, das wird halt irgendwie gespielt. Meine Erfahrung hat gezeigt, dass, wenn wir die komplette Kontrolle haben, von der Kreation bis zur Abspielung, dass die Immersion, und das Endergebnis genau so machen, wie wir es wollen. Natürlich sprichst du mit jemandem, der eigentlich Player baut, deswegen verlier ich die Kontrolle darüber nur sehr ungern.[…] Denn ich will am Ende genau wissen, wie es klingt und warum es so klingt."*[78]

Nichtsdestotrotz ist in anderen Bereichen eine objektorientierte Audioumgebung mit audiodeskriptiven Metadaten explizit erwünscht:

> *"MR: Mein Traumszenario wäre irgendwie, dass man so eine Arbeitsumgebung hat, wo man einfach irgendwie im 3d-Raum seine Schallquellen platzieren könnte, was also irgendwie objektbasiert wäre, weil bei Ambisonics ist man immer irgendwie an Kanäle beschränkt und kannst das aber nicht höher skalieren, wenn du das mal tun wolltest. Und so kann man kann seine Töne irgendwie im Raum platzieren, hat zum Beispiel aber zusätzlich seine Ambisonics-Atmo.[…] Dass ich sie mir einfach in alle Formate rendern kann, dass ich sag:*

[76] Rieger, I2. S. 6, Z. 7-11
[77] Rieger. S. 6, Z. 18-25
[78] Sanchez, I1. S. 7, Z. 9-19

> *Exportieren ich will das jetzt als dritter Ordnung Ambisonics, oder ich will das jetzt als mpeg-H oder…[…] wie beim 3D-Modelling, da kannst du es dir auch in unendlich hohen Auflösungen herausrendern, so wie du magst und genauso was fehlt irgendwie beim Ton."*[79]

Zusammenfassend lässt sich also sagen, dass alle der befragten Teilnehmer Audioinhalte aus der DAW ohne Objekt-Metadaten ausspielen. Sobald diese gefragt sind, beispielsweise aufgrund von dynamischem Rendering von Interaktion, wird diese mit einem eigenen Setup, bestehend aus separaten Player und Renderer-Instanzen gelöst.

5.3 Themenblock 3: Transkompatibilität

Wie in Kapitel 6.1 bereits erläutert, finden Objekte in immersiven Audio-Produktionen bisher kaum Verwendung, da die technische Infrastruktur noch nicht verfügbar genug ist. In den Gesprächen hat sich herausgestellt, dass momentan kaum bis gar keine Transkompatibilität für die Mischungen der Befragten besteht. Vor allem im Bereich VR scheinen die Faktoren, die die Transkompatibilität beeinflussen, am komplexesten zu sein.

> *"MR: Das war es aber auch schon, weil wenn ich halt jetzt ein Video habe und das würde ich irgendjemandem schicken wollen, dann müsste ich sagen: Ja du musst jetzt den und den Player nehmen und keine Ahnung ob das in einem Monat noch funktioniert. Also vielleicht hat der Player auch einen Bug, der selbständig ein Problem macht. […]Ich weiß nicht, ob Youtube überhaupt noch in fünf Jahren funktionieren wird. Das ist diese unangenehme Unsicherheit, was passiert überhaupt mit meinem Content."*[80]

Im weiteren Gespräch kam heraus, dass die korrekte Wiedergabe einer Mischung von zwei Faktoren abhängt. Dem Wiedergabegerät (HMD, Smartphone oder PC-gebunden) und dem Abspielprogramm (Player). Die am häufigsten genannten Abspielfehler sind falsche Lokalisation der Schallquellen und Frequenzverschlucken:

> *"BM: Das Problem ist, dass die Phantommitten nicht mehr stimmen, dass Distanzen nicht passen, dass die Höhen nicht richtig stimmen, denn eine Box war ein wenig niedriger angeordnet als die andere, die eine ein wenig weiter außen als die andere."*[81]

Die fehlende Transkompatibilität sehen die meisten Experten als Nachteil und als Behinderung im Kreationsprozess:

[79] Rieger, I2. S. 4, Z. 25 ff
[80] Rieger. S. 10, Z. 27 ff
[81] Maile, I3. S. 6, Z. 11-14

> *"AM: Man schränkt sich ja auch ein, wenn man auf einmal Stereo-Kompatibilität herstellen muss. Eigentlich will man ganz viel ausprobieren, aber man kommt halt ganz schnell zu dem Punkt wo man denkt: 'Haargh, das gibt bestimmt irgendwelche Phasen-Probleme, das lassen wir lieber.' Das ist ein totales Hindernis."*[82]

In bestimmten Fällen ist die Mischung jedoch ausschließlich für ein Wiedergabeszenario konzipiert. In diesem Fall wird nach Angaben von FS die Mischung vor Ort finalisiert:

> *"DS: Stimmst du dann die Mischungen vor Ort ab?*
>
> *FS: Ja, jedes Mal. Sehr wichtiger Schritt, sehr sehr wichtiger Schritt. Egal, was du davor machen kannst, den Raum kann man nicht simulieren. Egal, wie gut dein Testaufbau ist, den Raum kann man nicht simulieren, das wird schwierig."*[83]

Auch in szenenbasierter Wiedergabe, die prinzipiell systemagnostisch funktioniert, sieht FS einige Fehlerquellen und beeinflussende Faktoren, wie Raum und Kalibrierung.

> *"FS: Und wo der Punkt ist, ist, ob das System richtig kalibriert ist. Denn im Endeffekt bleiben die Koordinaten und wenn das System richtig gebaut ist, sollte es keinen großen Unterschied geben. Ein Raum bleibt aber immer ein Raum, und wird immer anders klingen. Und es ist immer gut, vor Ort eine Anpassung zu machen."*[84]

Vor diesem Hintergrund würden es zum großen Teil die Experten begrüßen, ein selbstanpassendes System mit der Zuhilfenahme von Metadaten zu verwenden, sofern diese standardmäßig gelesen werden können.

> *"BM: Klar, wenn es dann einen Objekt-Renderer gäbe, der die Anordnung der Boxen und den Raum kennt, wird natürlich alles ausgeglichen und sinnvoller angeordnet in dem Moment. Das sorgt dafür, dass inhaltlich das präsentiert wird, wie es präsentiert werden sollte. Das wäre natürlich schon schön, definitiv."*[85]

Im Bereich VR stellt MR klar die Forderung nach individualisierbarer HRTF's:

> *"MR: Das ist halt immer das Ding, dass es halt nicht personalisiert ist. Und genau deswegen wär es cool, wenn jetzt jeder Player Ambisonics abspielen kann, aber selbst dann klingt es immer noch nicht so, wie das für einen ir-*

[82] Mühlschlegel, I4. S. 10, Z. 13-18
[83] Sanchez, I1. S. 9, Z. 4-8
[84] Sanchez. S.9, Z. 18-25
[85] Maile, I3. S. 6, Z. 24-28

> *gendwie personalisiert cool wäre, sondern man hat immer so eine standardisierte Kunstkopfform, von der man auch überhaupt nicht weiß, wie sie aussieht, was für Maße sie hat und ob man die verändern kann oder so. Es gibt halt einfach einen Button "Binauralisieren" und mehr man auch nicht wirklich einstellen. Das kann sehr unbefriedigend sein, weil dann natürlich auch jeder Hersteller anders klingt. Wie gesagt, dass es dann schon auf Facebook anders klingt als über Youtube."*[86]

An anderer Stelle werden Bedenken geäußert, es würde zu viel Kontrolle abgegeben. Eine Blackbox ohne Eingriffsmöglichkeiten würde als störend empfunden werden:

> *"FS: Meine Erfahrung hat gezeigt, dass, wenn wir die komplette Kontrolle haben, von der Kreation bis zur Abspielung, dass die Immersion, und das Endergebnis genau so passt, wie wir es wollen."*[87]

Gleichzeitig werden Entwicklungswerkzeuge gefordert, die erschwinglich im Preis, einfach in der Bedienung und schnell und intuitiv zu verstehen sind:

> *"BM: Bei mir kommt da immer eine finanzielle Komponente hinzu. Wenn es gescheit integriert ist und kein Technik-Krampf, dann finde ich das alles toll. Ich bin das beste Beispiel für einen Anwender, der einfach mit Technik arbeiten möchte, ich will Arbeitstiere und keine Technikschlachten, wo ich denke, ich muss erst mal 10 Tage Kabel stecken oder Plug-Ins aufmachen, damit ich endlich mal an den Punkt komme, um Musik zu machen."*[88]

Im Abschluss dieses Themenblocks lässt sich sagen, dass auf Seite der technischen Infrastruktur für die befragten Teilnehmer kaum bis keine Transkompatibilität auf anderen Wiedergabesystemen gewährleistet ist. Ist diese jedoch erforderlich, wird auf Work-Arounds zurückgegriffen, wie beispielsweise die Neu-Mischung für andere Zielmedien. Bei einer Filmmischung werden zum beispielsweise Mischungen für Surround-Formate und eine Mischung für Stereo-Wiedergabe erstellt.[89] Im Bereich VR-Audio werden hingegen abhängig von HMD und Abspielapplikation zusätzliche Mischungen erstellt.[90] Auch hier wären Metadaten, die eine Mischung einer Applikation anpassen, von Vorteil. Der Bereich der Concept Cars ist ebenfalls ohne einheitlichen Standard und die Kanalbelegung variiert von Hersteller zu Her-

[86] Rieger, I2. S: 7, Z. 28 ff
[87] Sanchez, I1. S. 7, Z. 11-14
[88] Maile, I3. S. 9, Z. 11-17
[89] vgl. dazu: Mühlschlegel, I4. S. 9, Z. 29 ff
[90] Rieger, I2. S. 12, Z. 10 ff

steller. Für die Zukunft wäre auch hier ein Metadatenformat, das die Kanäle der Audiodatei den entsprechenden Kanälen im Auto zuweist, wünschenswert.[91]

5.4 Themenblock 4: Immersionserlebnis

In diesem Themenblock wurden subjektive Eindrücke der Experten gesammelt, was für sie technisch eine gelungene auditive Immersion auszeichnet und ob ein Datenformat wie das ADM zu einem höheren Immersionsgrad führen würde. Kennzeichnet dieser Themenblock nicht zwingend technische Eineindeutigkeit, sind diese Informationen dennoch hilfreich, da davon ausgegangen wird, dass die Experten durch ihre Arbeit mit immersivem Audio sensibilisiert für dieses Thema sind und außerdem Auskunft darüber gibt, wie sie die Immersionswahrnehmung des Konsumenten und dessen Faktoren einschätzen.

Laut BM steht die inhaltliche Komponente einer immersiven Audioproduktion klar im Vordergrund:

> *"BM: Also gab es selten die Grenze, wo ich offensiv gedacht habe: Oh, jetzt wird's blöd, weil jetzt ist es sau eingeschränkt und die Immersion findet nicht mehr statt. Es ist auch so, dass die Immersion in der Popmusik, die wir gemacht haben, gar nicht so wahnsinnig entscheidend ist. Immersion ist für mich auch immer eine Art von Natürlichkeit die stattfindet, dass man halt da massiv eintauchen kann. Wenn wir ein Klassik- oder Jazz-Konzert haben, was als Konzert stattgefunden hat, da finde ich es toll, wenn es immersiv ist und ich eintauchen kann, weil ich mich wie ein Zuschauer, der zuhört und es spürt."*[92]

Unter egal welchen Umständen sollte die eingesetzte Technik dem Inhalt dienen, nicht umgekehrt.[93]

Im Fall der binauralen Reproduktion sprechen sich AM und MR stark für die Personalisierung einer HRTF aus. Gleichzeitig sehen beide die größte Herausforderung darin, die Konsumenten für die Vorteile einer personalisierten HRTF zu sensibilisieren, da den meisten Konsumenten die technische Komplexität dieser Thematik nicht bewusst ist.

> *AM: "Die andere Sache ist das mit den HRTFs. Es gibt Leute, bei denen funktionieren die generischen total gut, bei manchen halt nicht. Aber letztendlich, wenn es um de Kompatibilität ginge, könnte es zumindest das Hörerlebnis bei*

[91] Mühlschlegel, I4. S. 9, Z. 1 ff
[92] Maile, I3. S. 5, Z. 12-21
[93] vgl. Maile. S. 3, Z. 19 ff

allen Menschen auf die gleiche Erfahrungsebene bringen. Weil nicht bei jedem die Mischung in der bestmöglichen Qualität abspielseitig ankommt!"[94]

MR: " Jein. Ich habe es letztens ausprobiert im Klangstüberl und das ist schon nice. Also wenn es mal richtig gut funktioniert, macht das Spaß. Aber ich glaube, da haben wir echt noch einen langen Weg hin, dass den Leuten überhaupt bewusst wird, das klingt jetzt so, wie ich das irgendwie kenne. Also muss halt schon echt diesen AHA-Effekt haben, dass es den Leuten bewusst wird. Und es kann sein dass es das aktuell genau das Problem ist, dass das was würde und die Leute denken: Ja okay es ist irgendwie räumlich aber auch dass jetzt vorne oder hinten ist, ach keine Ahnung[…]"[95]

Auf die Frage, ob ein standardübergreifendes Metadatenmodell das Immersionserlebnis steigern würde, reagierten alle Experten durchweg positiv. Es würde Kompatibilitätsprobleme lösen, eine vereinheitlichte Produktionskette erlauben, die tägliche Arbeit erleichtern und Hardware- oder softwareproprietäre Restriktionen lösen.

Weiterhin könnte flexibler und kreativer mit Ideen gearbeitet werden.

"AM: Weil es ja natürlich nervig ist, gerade für mich, dass jeder Hersteller mit seiner Super-Lösung ankommt, wie das denn jetzt zu machen sei, aber das funktioniert dann auch nur in diesem einen Ökosystem. Aber kombinieren kann ich diese Ökosysteme nicht. Ich kann nicht ein tolles Feature vom einen Format und ein tolles Feature vom anderen Format nehmen und in einen Topf schmeißen."[96]

Nichtsdestotrotz stünde für alle Beteiligten das Inhaltliche im Vordergrund. Würde der Inhalt mithilfe von Metadaten besser zum Konsumenten transportiert werden, stelle das eine Bereicherung auf Kreations- und Konsumentenseite dar.

"BM: Aber grundsätzlich finde ich so was alles spannend, also objektbasierte Sachen, wenn die mit den richtigen Metadaten im richtigen Containerformat übermittelt werden, dass man den Inhalt, den man produziert hat, besser und passender zum Endkonsumenten bringen kann: Super! Das ist ja das beste, was es gibt. Das ist ja das, was man sich wünscht. Dass jeder, der die Musik hört, so hört, wie man sie selber gehört hat. Das ist ja das Ziel eigentlich davon."[97]

[94] Mühlschlegel, I4. S. 11, Z. 30-35
[95] Rieger, I2. S. 12, Z. 28 ff
[96] Mühlschlegel, I4. S. 11, Z. 2-7
[97] Maile, I3. S. 9, Z. 17-21

"FS: Du kannst das beste Format der Welt haben, wenn am Ende Inhalt und Hardware nicht stimmen, ist es völlig egal. Was da natürlich helfen kann, ist ein bestimmte Optimierung des Prozesses, dahin zu kommen. Ich glaube, dass es nicht wirklich direkt einen Einfluss hat auf das Erlebnis, was man da hat, sondern einen Einfluss hat in den Prozess, den man hat, um dahin zu kommen."[98]

5.5 Zusammenfassung der Ergebnisse

Im Ergebnis lässt sich festhalten, dass objektbasierte Audioproduktionen unter den Projekten der Experten kaum geschehen. In den meisten Fällen liegt das an der finanziellen Verfügbarkeit, an der mangelnden Infrastruktur auf Wiedergabeseite und an der komplexen Einrichtung. Sobald Interaktion gefragt ist, werden proprietäre Lösungen mithilfe eines externen Renderers verwendet. Transkompatibel sind diese jedoch nicht. Szenenbasierte Audiodaten sind hingegen ein integraler Bestandteil für VR-Produktionen. Die Kombination aus kanalbasierten und szenenbasierten Elementen ist mithilfe des proprietären .tbe-Formates möglich, allerdings stößt dieses Format schnell an seine Grenzen, sobald ein HMD oder ein oder eine Wiedergabe-Software dieses Format nicht unterstützt. Das Interesse an objektbasierten Elementen ist jedoch bei allen Experten vorhanden und würde, sofern verfügbar, stark begrüßt werden. Solange dies nicht der Fall ist, werden weiterhin kanalbasierte Elemente im Mono- und Stereo-Format genutzt. Die Verheiratung von verschiedenen Audiotypen ist im ADM durch die Definition verschiedener Kanäle durch den XML-Datenblock "audioChannelFormat" möglich (vgl. dazu Kapitel 3.3.4). Zusätzlich kann definiert werden, ob der Inhalt binaural gerendert werden soll. In diesem Punkt wäre das ADM, sofern es ausgelesen werden könnte, für die Experten hilfreich.

Als finales Datenformat dient bei allen Experten eine kanalbasierte Ausspielung. Weiterhin werden hybride Exportformate aus kanal- und szenenbasiertem Audio verwendet. Diese sind jedoch ebenfalls auf ein proprietäres Format gemünzt und die Transkompatibilität zu anderen Formaten ist nicht automatisch gewährleistet. Des Weiteren ist besteht im Bereich VR momentan keine konstante Transkompatibilität. Dies hängt zum größten Teil vom verwendeten HMD und der Wiedergabe-Applikation ab. Updates dieser Applikationen können dafür sorgen, dass Mischungen unter Umständen zukünftig nicht mehr funktionieren. Der Impuls an Soft- und Hardwareentwickler im Bereich VR, kein proprietäres Rendering sondern einen offenen Standard zu verwenden, wäre erforderlich, würde Audio Designern Zeit und Energie sparen und möglicherweise zu einem Qualitätsschub der Inhalte führen. Das ADM würde sich auch hier durch die möglichen Kombination aus Kanälen, Szenen und Objekten sowie binauralem Rendering anbieten. Das Falten des Audiosignals mit einer individualisierten

[98] Sanchez, I1. S. 10 Z. 21-26

HRTF ist jedoch eine Aufgabe, die das ADM alleine nicht erfüllen kann, da es nur den Träger der Metadaten repräsentiert. Die nötige Schnittstelle müsste in der Applikation geschaffen werden.

Im Überblick haben alle Experten signalisiert, dass die Verwendung mit einem extensiven Metadatenmodell eine Bereicherung für ihre Projekte darstellen würde. Geknüpft daran sind jedoch die Bedingungen, dass die Verwendung der Metadaten intuitiv geschehen, auf den Geräten auf Konsumentenseite auslesbar und in der Anschaffung erschwinglich sein sollte. Letzterer Grund erklärt beispielsweise, warum objektbasierte Audioproduktionen noch immer ein Nischenprodukt darstellen. Sind diese drei Bedingungen gewährleistet, könnten die Experten sich vorstellen, auch objektbasierte Produktionen umzusetzen. Explizite Wünsche hinsichtlich des Metadatenformats konnten aus den bisherigen Problemen abgeleitet werden und bestehen im Wesentlichen aus dem Laden individualisierter HRTFs, dem dynamischen Rendering von Positionsdaten und dem Erstellen von Downmixes mithilfe von Objektdaten.

6 Aktuelle technische Implementierungen des ADM

Im Folgenden werden bisher implementierte Soft- und Hardware-Lösungen vorgestellt, die das Lesen und Schreiben von ADM-Metadaten ermöglichen.

6.1 IRCAM: ADMix und TOSCA

Die ADM Authoring Tools "ADMix" wurden vom französischen Forschungsinstitut IRCAM entwickelt und bieten eine Lösung zum Lesen von WAV-Dateien mit ADM-Inhalt als Standalone-Lösung. Das Programm "ADM-Player" spielt nur ins Programm geladene Dateien ab. "ADM-Renderer" bietet zusätzlich diverse Rendering-Funktionen wie beispielsweise Binaural-Rendering mit der Faltung zuvor generierter HRTFs sowie Lautsprecher-Rendering für bis zu 64 Lautsprecher.[99]

Die momentane Einbindung dieser Tools in eine DAW sind bisher nicht möglich, somit erlaubt das ADMix Toolset kein direktes Recording von Metadaten in einer DAW. Um diese Funktionalität indirekt zu ermöglichen wurde vom IRCAM das dezidierte Plug-In "Tosca" entwickelt. Ist eine Instanz von Tosca auf einer Spur in der DAW insertiert, sendet das Plugin OSC-Pakete zum Standalone-Programm "ADM Recorder". Automationsdaten können so bidirektional gesendet und empfangen werden, zudem besteht die Synchronisationsmöglich-

[99] Matthias Geier, „Software tools for object-based audio production using the Audio Definition Model (ITU-R Recommendation BS.2076)“.

keit beider Programme durch Timecode. Es wird allerdings betont, dass bisher nur die grundlegenden Funktionen und Parameter des ADM durch die ADMix-Tools abgedeckt sind.

6.2 MAGIX: Sequoia

Als Teil des Projekts OPRHEUS hat der deutsche Software-Hersteller "Magix" seine DAW "Sequoia" dahingehend umgestaltet und erweitert, dass sie das Lesen und Schreiben von ADM-Metadaten erlaubt. Da die ADM-Container in etwa den Komponenten und Parametern einer DAW entsprechen (Audiospur, Audioclip, Audiokanal etc.), werden die ADM-Parameter beim Import einer Datei den Einstellungen in Sequoia neu zugeordnet. Ordner-Spuren, Audiospuren, Clips und Kanäle und Automationsdaten werden automatisch erstellt. Umgekehrt werden beim Export die Parameter in der DAW in die XML Datenstruktur der zu speichernden Datei geschrieben. Zusätzliche Parameter wie beispielsweise "importance", also die Priorität eines Clips während des endgerätseitigen Renderings, wird über ein separates Editor-Fenster gelöst. Weiterhin wurde ein 3D-Panner mit VBAP-Algorithmus integriert. Es wird allerdings betont, dass längst nicht alle Funktionen und Parameter des ADM in der DAW abgedeckt sind.[100]

6.3 Merging Technologies: Pyramix

Der schweizerische Software-Hersteller "Merging Technologies" hat ebenfalls ADM-Unterstützung in die elfte Version ihrer DAW "Pyramix" implementiert. Nach Angaben des Herstellers können bestimmte Audioclips als Objekte definiert werden. Darüberhinaus ist es möglich, Spuren zu Gruppen zusammenzufügen, die sich gegenseitig ausschließen, um so beispielsweise das Umschalten verschiedener Sprachen zu ermöglichen.

6.4 AVID: Pro Tools

Der US-amerikanische Software-Hersteller "AVID Technologies" bietet ADM-Unterstützung in Version 12.7.1 ihrer DAW "Pro Tools". Das Programm gilt als Industriestandard für Filmproduktionen und beinhaltet unter anderem die native Integration von Dolby Atmos. Wird ein Dolby Atmos Master File in eine Session importiert, können Metadaten importiert werden. Als Datenträger dieser Information dient das ADM. Die XML-Informationen werden den entsprechenden Automationsspuren zugewiesen. in Version 12.8.2 können bisher nur Panning- und Routing-Informationen importiert werden.[101]

[100] Herberger, „D3.6: Implementation and documentation of object-based editing and mixing“.
[101] Avid Technologies, *Pro Tools® Reference Guide Version 12.8.2*. S. 1222

6.5 New Audio Technologies: Spatial Audio Designer

Der deutsche Software-Hersteller "New Audio Technologies" bietet Plug-Ins fürs Mischen auf Kopfhörern mittels Binauralisierung an. Das Plug-In "Spatial Audio Designer" stellt dafür eine separate Mischumgebung innerhalb der DAW bereit, die objektbasiertes Mischen und die Definierung von virtuellen Lautsprechern ermöglichen. Deren Signale anschließend binaural wiedergegeben. Mit der Vorstellung der zweiten Version des Plug-Ins wurde die Unterstützung von MPEG-H angekündigt. Ob im exportierten MPEG-H-Container ADM-Metadaten enthalten sind, ist vom Hersteller nicht näher definiert.[102]

7 Aktuelle und zukünftige Anwendungsfälle des ADM

Im Folgenden werden Produktionen präsentiert, die in ihrer Produktionskette das Audio Definition Model als Metadatenformat verwenden.

7.1 The Turning Forest

Der lineare VR-Film "The Turning Forest" wurde 2016 federführend vom Research and Development-Department der BBC produziert und für das HMD Oculus Rift veröffentlicht. Es ist Teil des "S3A Future Spatial Audio"-Programms, was sich mit der Inhaltskreation von Medien mit binauralem Ton beschäftigt. Das Projekt begann ursprünglich als Hörspiel, zu dem später korrespondierende Visuals kreiert wurden.[103] Während der Audio-Produktion wurden Mix-Elemente mit ADM-Metadaten exportiert. Diese Exportdateien wurden anschließend in die Game Engine Unity importiert. Durch das Erstellen diverser Skripte in der Game Engine wurde ermöglicht, die ADM-Metadaten der zu importierenden Dateien zu extrahieren. Basierend auf den Metadaten wurden in Unity korrespondierende Audio-Objekte an den richtigen Koordinaten im virtuellen Raum generiert.[104] Das ADM dient hier also als Schnittstelle zur Datenübermittlung zwischen DAW und Game Engine.

7.2 ADM Object-Based Audio Player

In einem weiteren Forschungsprojekt des Research and Development-Departments der BBC wurde ein web-basierter Audio-Player entwickelt, der objektbasierte Produktionen auf Client-Seite binaural rendern kann. Der Nutzer kann auf die Website zugreifen, eine Datei mit inte-

[102] Ammermann, „The Spatial Audio Designer Version 2 – New Audio Technology“.

[103] Pike u. a., „Object-Based 3D Audio Production for Virtual Reality Using the Audio Definition Model“. S. 2

[104] Pike u. a. S. 5

grierten ADM-Metadaten hochladen, der Player kann diese Daten lesen und Client-seitig binaural rendern. Hierzu wurde sich der 3D Audio Engine der Web Audio API bedient. Serverseitig werden die ADM-Metadaten extrahiert, von ihrer XML-Darstellung in die Web-Sprache JSON konvertiert, in einer separaten Datenbank gespeichert, und während der Wiedergabe dynamisch gerendert. Die Anwendung verfügt weiterhin über eine visuelle Repräsentation aller Audio-Objekte im 3-Raum und die Schnittstelle für die Datenübermittlung eines externen Head-Trackers.[105] Es wird gezeigt, dass Client-seitig binaurales Rendering mithilfe des ADM möglich ist.

8 Fazit und Ausblick

Ein Überblick über die verschiedenen Charakteristika des ADM wurde präsentiert. Ebenso wurden die Grundlagen des räumlichen Hörens, die Reproduktion von Schallereignissen durch kanalbasierte, objektbasierte und szenenbasierte Wiedergabe erläutert und Beispiele zur heutigen Nutzung gegeben. Anschließend wurde durch die Durchführung von Experten-Interviews mit Toningenieuren empirisch ein Überblick gegeben, mit welchen Herausforderungen sie sich während der Inhaltskreation sie sich konfrontieren. Ergebnis dieser Befragung war, dass in aller Regel kanalbasiert gearbeitet wird und teils hybride Formen aus kanalbasierter und szenenbasierter Schallfeldreproduktion genutzt werden. Objektbasierte Produktionen sind aus finanziellen und infrastrukturellen Gründen nicht nutzbar und fristen ein Nischendasein. Einen großen Streitpunkt stellt die Wiedergabe von immersivem Audio mittels Binaural-Rendering dar, da personalisierte HRTFs momentan noch nicht unterstützt werden. Wann dies der Fall sein wird und den Konsumentenmarkt erreicht, ist fraglich. Abgesehen davon begrüßen alle befragten Experten die Möglichkeit, bestimmte Kanäle mit Metadaten zu versehen, sodass sie kanalbasiert ausgespielt oder objekt- bzw. szenenbasiert ausgespielt werden. Auf Kreationsseite ist ein offenes Metadatenmodell also gutzuheißen, zumal die Erweiterung und Modellierung mittels XML-Code möglich ist.

Das Fraunhofer hat bereits bestätigt, dass das ADM als Metadatenmodell im MPEG-H-Container Verwendung finden wird.[106] Dieser wurde wiederum als standardisierter Audio-Container für den DVB-UHD-Standard und den ATSC-3.0-Candidate.[107] Im Umkehrschluss müssen sich mindestens Audio Designer und Mischtonmeister, die im Broadcasting-Bereich arbeiten, auf eine Umstellung ihrer Workflows vorbereiten, sofern sie MPEG-H-Inhalte produzieren wollen.

[105] Chris Pike, „Delivering Object-Based 3D Audio Using The Web Audio API And The Audio Definition Model“. S.3 ff

[106] Bleidt, „Development of MPEG-H TV Audio-System for ATSC-3-0.pdf“.

[107] IIS, „MPEG-H jetzt auch in DVB-Spezifikation“.

Der Zugang zu einem offenen Metadatenmodell ist auf Kreationsseite allerdings nur dann wünschenswert, wenn konsumentenseitig die nötige Infrastruktur geschaffen wird, die eine Auslesung der Metadaten erlaubt. Dieser Teil wurde in dieser Arbeit nicht behandelt, ist jedoch nicht minder interessant und im Auge zu behalten. Vorwiegend liegt dies in den Händen der Gerätehersteller, die entscheiden, welche Chips in beispielsweise TV-Geräten verbaut werden. Im Zuge der IBC 2017 in Amsterdam wurden bereits Chips und Fernseher vorgestellt, die das Auslesen der in einem MPEG-H Container enthaltenen Metadaten unterstützen.[108] Wann diese Geräte allerdings in Serie, die Multimedia-Geschäfte und somit auch den Konsumenten erreichen, bleibt abzuwarten. Auch dann wird aller Erwartung nach eine erneute Akklimatisierungsphase stattfinden, in der Audio Designer das Potenzial und die Risiken hinter dem neuen Standard herausfinden und Konsumenten die neuen Einstellungsmöglichkeiten von MPEG-H herausfinden. Nichtsdestotrotz bieten MPEG-H als Container und ADM als Metadatenmodell einiges Potential, um immersives Audio zu produzieren, was in dieser Arbeit gezeigt wurde.

[108] IIS, „Neue Produkte mit MPEG-H-Support vorgestellt“.

9 Literaturverzeichnis

Alexander Bogner, Wolfgang Menz, Beate Littig. *Interviews mit Experten - Eine praxisorientierte Einführung*. Springer-Verlag, 2014.

Ammermann, Tom. „The Spatial Audio Designer Version 2 – New Audio Technology“. Zugegriffen 23. Januar 2018. https://www.newaudiotechnology.com/en/the-spatial-audio-designer-version-2/.

Andreas Silzle, Uwe Herzog Michael Weitnauer, Olivier Warusfel, Werner Bleisteiner, Tilman Herberger, Nicolas Epain, Benjamin Duval, Niels Bogaards, Chris Baume. „Orpheus Audio Project: Piloting an End-to-End object-based audio broadcasting chain“. In *IBC Conference*, 2017.

„Audio Definition Model Software - BBC R&D“. Zugegriffen 30. Oktober 2017. http://www.bbc.co.uk/rd/publications/audio-definition-model-software.

Authoring for Dolby Atmos® Cinema Sound Manual. 100 Potrero Avenue San Francisco, CA 94103-4813 USA: Dolby Laboratories, Inc., 2014.

Avid Technologies. *Pro Tools® Reference Guide Version 12.8.2*. Zugegriffen 16. Januar 2018. http://resources.avid.com/SupportFiles/PT/Pro_Tools_Reference_Guide_12.8.2.pdf.

Blauert, Jens. *Räumliches Hören*. Stuttgart; Leipzig: Hirzel, 1997.

Bleidt. „Development of MPEG-H TV Audio-System for ATSC-3-0.pdf“. Zugegriffen 17. Januar 2018. https://www.iis.fraunhofer.de/content/dam/iis/en/doc/ame/Conference-Paper/BleidtR-IEEE-2017-Development-of-MPEG-H-TV-Audio-System-for-ATSC-3-0.pdf.

„BS.2076 : Audio Definition Model“. Zugegriffen 30. Oktober 2017. https://www.itu.int/rec/R-REC-BS.2076/en.

„BS.2088 : Long-form file format for the international exchange of audio programme materials with metadata“. Zugegriffen 30. Oktober 2017. https://www.itu.int/rec/R-REC-BS.2088-0-201510-I/en.

Chris Pike, Frank Melchior Peter Taylor. „Delivering Object-Based 3D Audio Using The Web Audio API And The Audio Definition Model“. In *IRCAM Web Audio Conference*, 2015.

Curtis, Robin. „Immersion und Einfühlung“. *montage AV*, Nr. 17/2/2008 (2008).

Dickreiter, Michael, Volker Dittel, Wolfgang Hoeg, und Martin Wöhr. *Handbuch der Tonstudiotechnik*. 8., überarbeitete und erweiterte Auflage. De Gruyter Saur reference. Berlin: De Gruyter Saur, 2014.

Füg, Simone, Andreas Hölzer, Christian Bor\s s, Christian Ertel, Michael Kratschmer, und

Jan Plogsties. „Design, Coding and Processing of Metadata for Object-Based Interactive Audio“. In *Audio Engineering Society Convention 137*, 2014. http://www.aes.org/e-lib/browse.cfm?elib=17420.

Gasull Ruiz, Alejandro, Christoph Sladeczek, und Thomas Sporer. „A Description of an Object-Based Audio Workflow for Media Productions“. In *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology – Cinema, Television and the Internet*, 2015. http://www.aes.org/e-lib/browse.cfm?elib=17611.

Görne, Thomas. *Tontechnik: Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis*. 4., aktualisierte Aufl. Medien. München: Hanser, 2015.

Hammershøi, Dorte, und Henrik Møller. „Binaural Technique — Basic Methods for Recording, Synthesis, and Reproduction“. In *Communication Acoustics*, herausgegeben von Jens Blauert, 223–54. Berlin/Heidelberg: Springer-Verlag, 2005. http://link.springer.com/10.1007/3-540-27437-5_9.

Herberger, Tilman. „D3.6: Implementation and documentation of object-based editing and mixing“. Object-based broadcasting – for European leadership in next generation audio experiences, 30. November 2017. https://orpheus-audio.eu/wp-content/uploads/2017/12/orpheus-d3.6_impl.doc-of-ob-editing-and-mixing.pdf.

IIS, Fraunhofer. „MPEG-H jetzt auch in DVB-Spezifikation“. *Fraunhofer Audio Blog* (blog), 31. Januar 2017. http://www.audioblog.iis.fraunhofer.de/mpeg-h-dvb-specification/.

„Neue Produkte mit MPEG-H-Support vorgestellt“. *Fraunhofer Audio Blog* (blog), 23. Oktober 2017. http://www.audioblog.iis.fraunhofer.de/new-products-supporting-mpeg-h-audio-hitting-market/.

Jan Fleischmann. „MPEG-H -- ein Audioformat der nächsten Generation (NGA)“. Zugegriffen 26. Oktober 2017. http://tech-magazin.de/2017/05/mpeg-h-ein-audioformat-der-naechsten-generation-nga/.

Jauch, Jo, und Michael Romanov. „Ambisonic Hits The Road!“ In *ICSA PAPERS*, 2017.

Kaiser, Robert. *Qualitative Experten- Interviews Konzeptionelle Grundlagen und praktische Durchführung*. Springer-Verlag, 2014.

Kronlachner, Matthias. „Loudspeaker Sphere Observatory Vilnius University | matthiaskronlachner.com“, 20. Oktober 2013. http://www.matthiaskronlachner.com/?p=1774.

Kronlachner, Matthias, und Franz Zotter. „Spatial transformations for the enhancement of Ambisonic recordings“. In *ICSA 2014*. Graz, o. J. https://ambisonics.iem.at/Members/zotter/publications/2014_KronlachnerZotter_AmbiTransfo

rmationEnhancement_ICSA.pdf.

Maile, Benedikt. I3, 04. Januar 2017.

Matthias Geier, Olivier Warusfel Thibaut Carpentier, Markus Noisternig. „Software tools for object-based audio production using the Audio Definition Model (ITU-R Recommendation BS.2076)“. In *ICSA PAPERS*, 2017.

Messonnier, Jean-Christophe, Jean-Marc Lyzwa, Delphine Devallez, und Catherine De Boisheraud. „Object-Based Audio Recording Methods“. In *Audio Engineering Society Convention 140*, 2016. http://www.aes.org/e-lib/browse.cfm?elib=18268.

Moore, Madison. „Facebook 360 Spatial Workstation, Jibo SDK, and Twilio updates developer console—SD Times news digest: May 24, 2016“. *SD Times* (blog), 24. Mai 2016. https://sdtimes.com/facebook-360-spatial-workstation-jibo-sdk-twilio-updates-developer-console-sd-times-news-digest-may-24-2016/.

Mühlschlegel, Andreas. I4, 03. Januar 2018.

Nico Jurran. „3D-Sound: Netflix bietet nun auch Dolby-Atmos-Ton“. Zugegriffen 26. Oktober 2017. https://www.heise.de/newsticker/meldung/3D-Sound-Netflix-bietet-nun-auch-Dolby-Atmos-Ton-3758814.html.

Oezturan, Altan. „Austausch von Metadaten in der Broadcasting Branche“, 2018.

Pike, Chris, Richard Taylor, Tom Parnell, und Frank Melchior. „Object-Based 3D Audio Production for Virtual Reality Using the Audio Definition Model“. In *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, 2016. http://www.aes.org/e-lib/browse.cfm?elib=18498.

Rieger, Martin. I2, 30. November 2017.

Rieger, Martin. „Virtual Reality Audio Formats - Pro und Contra“. *vr-tonung.de* (blog). Zugegriffen 26. Oktober 2017. https://www.vrtonung.de/virtual-reality-audio-file-format-pro-contra/.

Roland, Abel. „Ambisonic Mastering: The challenges of sound design in 360° videos“. In *ICSA PAPERS*, 2017.

Ryan, Marie-Laure. *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media.* Baltimore; Boulder: Johns Hopkins University Press NetLibrary, Inc. [distributor, 2003. http://public.eblib.com/choice/publicfullrecord.aspx?p=3318161.

Sanchez, Felipe. I1, 22. November 2017.

Shivappa, Shankar, Martin Morrell, Deep Sen, Nils Peters, und S. M. Akramus Salehin. „Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics“. In *Audio Engineering Society Conference: 2016 AES International Con-*

ference on Audio for Virtual and Augmented Reality, 2016. http://www.aes.org/e-lib/browse.cfm?elib=18493.

Sontacchi, Alois. „Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhörer-anwendungen“. Technische Universität Graz, 2003.

Susal, Joel, Kurt Krauss, Nicolas Tsingos, und Marcus Altman. „Immersive Audio for VR“. In *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, 2016. http://www.aes.org/e-lib/browse.cfm?elib=18512.

Tischmeyer, Friedemann. „EBU-Norm R128 – Die leise Revolution der Pegelmessung“. delamar.de, 14. Juni 2016. https://www.delamar.de/mastering/r128-14870/.

„True-Multichannel-Mixing - Ambisonics“. Zugegriffen 29. Januar 2018. https://uod-true-multi-channel-mixing.wikispaces.com/Ambisonics.

Weinzierl, Stefan, und Verband Deutscher Tonmeister, Hrsg. *Handbuch der Audiotechnik*. Berlin: Springer, 2008.

Wisse, Elke. „IAN – Immersive Audio Network“. *VDT-Magazin*, Juli 2017, 18–22.

10 Abbildungsverzeichnis

11 Anhang

Einstieg	
Einleitung: Danksagung. Einverständniserklärung für die Aufnahme. **Wie ist Ihr Name, wo arbeiten Sie und was sind Ihre Aufgaben?**	
Hauptfragen	**Themen zum Nachfragen**
Zusammensetzung der Mix-Elemente	
Welche verschiedenen Audiowiedergabetypen (Objektbasiert, Kanalbasiert, Szenenbasiert) nutzt du für ein immersives Erlebnis deiner Produktionen?	- Nutzt du objektbasiertes Audio? Wenn ja, warum? Wenn nein, warum nicht? - Nutzt du Ambisonics? Wenn ja, warum? Wenn nein, warum nicht?
Worauf legst du in der Phase der Kreation besonders Wert?	- Ist diese Priorisierung aus persönlichen Motiven entstanden oder gibt es bestimmte Normen, die dich dazu zwingen?
Wie sehr beeinflussen dich dein finales Datenformat und das Endabspielgerät während der Kreation?	- Wärst du kreativer und produktiver, wenn du dir um das finale Datenformat keine Gedanken machen müsstest?
Arbeitest du mit Protokollen wie beispielsweise OSC?	-
Arbeitest du mit interaktiven Elementen?	- Wie setzt du diese um?
Finales Datenformat	
Wie sieht deine Abgabedatei aus? Wovon ist die Abgabedatei abhängig?	-
In welchem Umfang schränkt dich dein aktuelles Datenformat ein?	- Wodurch könnte das Immersionserlebnis schwächer werden? Wie wirkt sich das Abgabeformat auf die Immersion aus?
Welche Anforderungen stellst du an ein Abgabeformat, die es momentan nicht erfüllen kann?	- Hast du dir schon mal ein proprietäres Format programmiert oder programmieren lassen, das nur einen Zweck erfüllt?
Transkompatibilität	
Auf welchem Abspielgerät wird dein Inhalt in der Regel wiedergegeben?	- Welche Faktoren kannst du dabei beeinflussen? Welche nicht?
Welche technischen Schritte geschehen zwischen der Abgabe deiner Datei und der Veröffentlichung?	- Bei welchen dieser Schritte bist du (nicht) involviert?
In welchem Umfang schränkt dich die Funktionalität des Abspielgerätes ein?	- Stimmst du Mischungen auf bestimmte Wiedergabegeräte ab?

Würdest du als Kreateur von höherer Transkompatibilität profitieren?	-
Immersionserlebnis	
Von welchen technischen Parametern hängt das Immersionserlebnis der Kunden ab?	- Welche Parameter kannst du beeinflussen?
Erfüllt die heutige technische Infrastruktur deine Anforderungen an ein immersives Klangerlebnis?	-
Würde ein übergreifendes Open-Source Datenmodell das Immersionerlebnis steigern?	- Würde ein übergreifendes Open-Source Datenmodell ein größeres Immersionserlebnis generieren?
Abschluss	
I.	

12 Transskripte

Transkription - Interview I1

Interview vom 22.11.2017

Interviewter: Felipe Sanchez

Abkürzungen: FS = Felipe Sanchez ; DS = Daniel Strübig

DS: Hallo Felipe.

FS: Hallo Daniel.

DS: Vielen Dank, dass du mit mir das Interview machst im Rahmen meiner Bachelor-Arbeit. Und, wir haben es eben schon besprochen, mit der Aufnahme gibst du mir quasi auch die Einverständniserklärung, dies im Rahmen meiner Bachelorarbeit niederzuschreiben.

FS: Genau.

DS: Gut. Thematisch geht es ja um die Funktionalität von immersiven Audiosystemen. Und wie lassen sich mehr Daten, oder Metadaten in ein einem WAV-File speichern. Da gibt es von der BBC das Audio Definition Model. Und, darüber werden wir gar nicht so viel reden, sondern eher: Was machst du überhaupt in deiner täglichen Arbeit, wenn du Klangszenographie machst oder immersives Audio. Also, ich habe das Interview in 4 Themenblöcke eingeteilt: Die Zusammensetzung einer Mix-Elemente oder deiner Kreationselemente, das finale Datenformat, was du ausspielst, dann die Transkompatibilität und abschließend das Immersionserlebnis. Und dann würde ich sagen, fange ich einfach mal an.

FS: Jo.

DS: Wenn du jetzt an eine Klangszenographie oder an immersives Audio, was für verschiedene Wiedergabetypen nutzt du da? Nutzt du kanalbasierte Elemente, objektbasierte Elemente, oder sogar szenenbasierte Elemente wie zum Beispiel Ambisonics? Wo setzt du da deinen Focus?

FS: Das hängt immer von den Nutzung ab. Wir haben verschiedene Möglichkeiten für verschiedene Endprodukte und der Fokus liegt darauf: Was funktioniert für die verschiedenen Formate? Vor allem, was bei sehr wichtig ist: Haben wir einen interaktiven Anteil oder nicht? Weil: Normalerweise, wenn es linear ist, wenn wir ein Multikanal-immersives Raumerlebnis haben, für ein Museum oder einen Showroom oder was auch immer, wenn es linear ist, sehen wir keinen Grund, anders als kanalbasiert auszuspielen. Das heißt, da gibt es ein interleaved WAV-File, der einfach die X Kanäle hat, und die werden im Wiedergabesystem ausgegeben. Wenn es linear ist, dann brauchen wir nicht unbedingt etwas anderes, auch wenn es Wellenfeldsynthese, eine kleine Version von Wellenfeldsynthese oder Ambisonics ist: Wenn wir

da nicht eine performative oder interaktive Geschichte sehen, gibt es nicht den Grund, einen Renderer vor Ort zu haben. Sondern einfach auf kanalbasiert alles auszuspielen. Der Punkt, wo es interaktiv werden könnte, da fängt es an, zu fragen: Was benutzen wir für die verschiedenen Formate? Sollen wir Ambisonics machen, wollen wir Wellenfeldsynthese machen, oder wollen wir ein eigenes System machen, wo wir dann objektbasiert arbeiten? Wir arbeiten vor allem objektbasiert, weniger szenenbasiert, Ambisonics machen wir nicht so viel. Erst mit den VR-Geschichten kommt es in die Gegenwart. Ehrlich gesagt machen wir eher objektbasierte Sachen à la Wellenfeldsynthese oder mit einem eigenen Rendersystem.

DS: Ok. Also Stichwort Interaktion: Wie genau implementiert ihr das? Nutzt ihr beispielsweise OSC?

FS: Genau. Mittlerweile hat sich das eingespielt, eingegrooved, in der Kombination Max MSP, Ableton Live. Und, die Signale werden hin und hergeschickt, was auch immer da gebraucht wird. Ein UDP-Port, kann man auch in TCP schicken, aber im Endeffekt machen wir OSC-Signale in OSC-Paketen, wo wir per OSC die Koordinaten vom Raum schicken, wo der Sound kommen kann. Nehmen wir einen Raum, und ein Sound muss einer Person folgen, das heißt, es wird getrackt, das heißt, wir kriegen von dem System wo diese Person ist in XYZ-Koordinaten. Und dann schicken wir auf unsere eigenen gebauten Player: An diesen Koordinaten, spiel diesen Sound. Und die komplette Kommunikation läuft über OSC.

DS: Und inwiefern beeinflusst dich dieses System schon während der Kreation?

FS: Viel. Also ich bin sowieso der Meinung, dass jegliche DAW, jegliches Interface, wenn du arbeitest, beeinflusst deinen kreativen Prozess. Man kann sogar ab und zu hören: Das wurde in Pro Tools gemacht, dies wurde in Ableton gemacht. Die Farben, die Begrenzungen, die du hast: Sobald du anfängst zu denken, ok: Wir haben XY interaktive Kanäle, dann fängst du an zu überlegen: Haben wir genug Kanäle für die Objekte? Jeder Kanal ist ja ein Objekt. Kann der Rechner das mit 8 oder 9 Objekten? Wir haben neulich ein Projekt gemacht, wo der Rechner schon mit 10 verschiedenen Quellen an der Kante war, denn er hat auch das ganze Licht gemacht und so. Also, das musst du von vornherein überlegen, wie das System gebaut ist, wie die Architektur ist, und da, da limitiert man sich ein bisschen im Vergleich zu dem,

was man ursprünglich geplant hat. Wir versuchen, ohne Limitierungen des Systems zu denken, und dann fangen wir an zu begrenzen, wo müssen wir uns limitieren durch die Begrenzung von Kanälen, Begrenzung des Systems, was der Rechner kann, wenn wir keine super-krassen Renderer haben, dann muss der Player das auch leisten. Also ja: Das beeinflusst stark!

DS: So wie ich das jetzt herausgehört habe: Baut ihr dann spezielle Rendering-Systeme?

FS: Genau. Also Rendering-Systeme ist ein bisschen übertrieben, denn ein Rendering-System wird erst gebracht, wenn wir Ambisonics oder Wellenfeldsynthese verwenden, also wo wirklich Algorithmen verwendet werden, und wir haben keine richtigen Algorithmus-Renderer, sondern es sind wirklich nur Positionsdaten. Und wo das Objekt ist, dann gehen wir von den Lautsprechern aus, und je näher das Objekt an den Lautsprechern ist, desto lauter ist der und der Lautsprecher. So wie man das kennt von der Wellenfeldsynthese, nur das es eben keine Wellenfeldsynthese, sondern eine ganz normale Laufzeitdifferenz gibt, oder eine Entfernung zu den Lautsprechern gibt. Und ja, mehr und mehr von unserem Tagesgeschäft entwickeln wir und wir verwenden gerne unsere eigenen Player, das heißt wir lösen das über Software, die genau angepasst ist, für was auch immer wir machen, also für die Interaktivität, für den Play, alles mögliche, von 2 Kanälen bis zu X Kanälen und das bauen wir selber, damit es wirklich gebaut ist für dieses System und nicht irgendeinen Player, sodass wir am Ende nicht wissen, wie es eigentlich klingt.

DS: Also würdest während der Kreation davon profitieren, wenn es schon einen Renderer gäbe, der alles erfüllt und du musst eigentlich nur sagen: Ich hätte gern das, Ich hätte gern das, Ich hätte gern das?

FS: Ich weiß nicht, ob es mir lieber wäre: Ein Renderer, der alles kann, oder ein weißes Blatt Papier, was für mich Max MSP ist, und ich genau das brauche, nicht mehr und nicht weniger. Das ist mir viel lieber, ich weiß genau, wie es funktioniert, und ganz viele Dinge sind maßgeschneidert für das, was es ist. Also, es kommt ein OSC-Signal, diese Tür geht auf, dieses Licht geht an, dieses System fängt an, diese Musik fängt an, an dieser Position, in diesem Raum, mit dieser Volume, und das alles in einem System zu machen, das schon existiert, habe ich noch nicht irgendwas gefunden, das mir das

gibt. Wir arbeiten sehr customed, für die Installationen, die wir machen. Natürlich, viel ist auch einfach ein 16-Kanal-Player, der einfach linear spielt, aber, hm, das kann jeder Player.

DS: Eine letzte Frage noch zu diesem Themenblock: Worauf legst du in der Kreationsphase wert? Denkst du da beispielsweise schon an die finale Spielstätte, weil du eben meintest, dass ihr viel mit Licht und Installationen arbeitet?

FS: Viel. Also wir versuchen soviel wie möglich… Das Schwierige mit Audio ist ja diese Abstraktion, was wie am Ende klingen kann. Und man geht in diesem Prozess auf verschiedene Arten vor. Zwischendurchgibt es immer Feedback-Runden, man kann dem Kunden nicht einfach sagen: So wird es in Stereo klingen, Surround klingt aber viel besser, denn die haben diese Abstraktion nicht. Und deswegen muss man in verschiedenen Schritten vorgeht. Also wenn man produziert, zum Beispiel: Man produziert den Inhalt, wie man es haben will, normal in stereo, und das würde man mit dem Kunden abstimmen, und man weiß sowieso es im Multikanal sein wird, das heißt, es muss mehr Material kommen. Man muss mehr Elemente einbauen. Aber man kann nicht erst mal alles in stereo machen, denn das wäre zu voll. Man denkt die ganze Zeit schon: Was wird am Ende sein. Man versucht sich die ganze Zeit zu überlegen, während man arbeitet, man macht auch Bilder von dem Raum, legt sich die auf den Bildschirm, sodass man ein Gefühl hat, wie fühlt sich das an. Also es ist sehr wichtig, dass man sich überlegt, wie das am Ende sein wird. Und es ist schon sehr wichtig, diese Abstraktionsmöglichkeiten im Kopf zu haben: Wie arbeite ich in einem Raum, der eigentlich nicht existiert, für den ich aber komponiere, und deswegen ist es immer super wichtig, am Ende des Tages in den RAum zu mischen. Und man hat schon eine starke Beeinflussung vom Endformat, man weiß schon vorher: Das ist eine X-Kanal-Experience: Wir komponieren dafür. Wir komponieren nicht in stereo und machen dann irgendeinen Upmix, sondern wir komponieren schon für den Raum. Unsere Canvas ist ein Raum, Und nicht einfach die Lautsprecher.

DS: Das ist ein Punkt, da gehen wir in Themenblock 3 nochmal drauf ein. Dann würde ich sagen, reden wir über deine Exportdatei. Was genau exportierst du nach der Kreation?

FS: Das Übliche sind MONO-WAV-Formate. 44,1 kHz und 24 Bit. Die Bits sind sehr wichtig, weil im Raum hat man viel mehr Möglichkeiten tie-

fer zu gehen, das ist wichtiger als man denkt, 48 kHz, wenn es filmgekoppelt ist, aber normalerweise 44,1 kHz und 24 Bit. Und normalerweise in WAV-Dateien und je nach Art der Player entweder Mono-Files oder Interleaved-Files. Interleaved-Files machen wir immer um sicherzugehen, dass alles synchron läuft.

DS: Das macht Sinn. Wünscht du dir von dem Export-Format manchmal mehr Features? Dass du sagst, ok, ich hätte gerne, dass bestimmte Tracks schon in Ambisonics B-Format konvertiert sind?

FS: Ich kann mich erinnern, dass ich manchmal dachte: Wär geil, wenn man da richtig Metadaten hätte. Was ein großes Ding wäre, wenn du zum Beispiel sagen könntest, du hast einen Mono-File, der auf 24 Kanälen abgespielt wird, und du machst eine Bewegung, die hast du schon im Studio gemacht, das heißt du hast XYZ-Koordinaten, und dann hast du ein File, und in der File weiß genau, wo er spielen muss. Also diese objektbasierten Informationen sind immer sehr cool, denn das sind bisher 2 verschiedene Informationen, die von 2 verschiedenen Orten kommen. Einmal macht der MAX die Orientierung, einmal macht der Player die Audio-Verteilung. Schön wär's, wenn es von einer Quelle kommen würde.

DS: Also ihr fragmentiert momentan noch die Koordinaten und das eigentliche Audio.

FS: Ja genau. Und ich kann mir auch so ein paar Momente vorstellen, wo ich denke: Ach, wär cool wenn das drin ist, aber das kann ich dir gerade im Augenblick nicht sagen. Wir sind da zu sehr eingegrooved in unser System.

DS: Das macht nichts. Wenn es dir einfällt, schreib mir einfach eine Mail.

FS: Genau (lacht).

DS: Also ist es so, dass ihr auch schon ein proprietäres System programmiert habt, in Max MSP beispielsweise.

FS: Wir programmieren eigene Player, ja. Also der Player ist für uns ein leeres Blatt Papier, da können wir machen, was wir wollen. Inklusive binauralem Rendering mit SPAT von IRCAM. Oder es gibt schon Sachen, die Wellenfeldsynthese schon in einer FAST guten Qualität machen und natürlich die ganzen Ambisonics-Geschichten, für die wir

offen sind. MAX ist dafür unsere Hauptbasis, wie wir alles dann abspielen.

DS: Ok. Würdest du sagen, wenn wir in Richtung Immersion denken, dass das Immersionserlebnis größer wäre, wenn du dir darüber keine Gedanken machen müsstest?

FS: Worüber?

DS: Beispielsweise über die Metadaten. Wenn du keinen Player noch extra dafür bauen müsstest. Das wäre alles schon da.

FS: Hmmm. Weiß ich nicht, weil… Also es gibt schon Momente, wo wir sagen, wir liefern einfach nur Content, und der Player, das wird halt irgendwie gespielt. Meine Erfahrung hat gezeigt, dass, wenn wir die komplette Kontrolle haben, von der Kreation bis zur Abspielung, dass die Immersion, und das Endergebnis genau so passt, wie wir es wollen. Natürlich sprichst du mit jemandem, der eigentlich Player baut, deswegen verlier ich die Kontrolle darüber nur sehr ungern. Da wir nicht Standard-Formate benutzen, wo wir einfach eine Custom-made Lösung brauchen, also darüber die Kontrolle zu verlieren, das mache ich nur sehr ungern. Denn ich will am Ende genau wissen, wie es klingt und warum es so klingt.

DS: Ok, vielen Dank. Dann zum Themenblock Transkompatibilität. Du hast gerade schon angedeutet, ich würde es gerne nochmals auf den Punkt bringen. Auf welchem Abspielgerät wird dein Inhalt in der Regel wiedergegeben?

FS: Normalerweise auf MAX MSP. Es gibt auch AmbiPandora, es gibt noch Sachen wie Watchout, das sind Player, die es schon gibt. Also was vor 10 Jahren einfach ein Multikanal-Player war mit einer Festplatte oder DAT. Heutzutage gibt es einfach ganz viele Media Player, die eingebaut werden, also an der IAA oder im Showroom. Da gibt es immer wieder Lösungen, die schon Standard sind. Es gibt auch Lösungen, die nicht von uns gemacht werden sondern von jemand anderem, mit v4 oder so etwas, oder es gibt auch die Möglichkeit, mit anderen Computern über Timecode irgendwas synchron zu spielen. Es gibt verschiedene Lösungen, je nachdem, was man da hat. Aber in der Regel versuchen wir unsere eigenen Player zu bauen, der auf PC oder Mac läuft, normalerweise in Max MSP. Ein Procedere, das bei uns nicht untypisch ist, dass wir im Kreationsprozess in Ableton und Max arbeiten, und dann gucken wir, dass alles, was in Ableton passiert, in

Max MSP programmiert wird, damit es keine Lizenzen und so gibt, Denn ein Standalone von Max brauchst du nicht lizenzieren. Deswegen versuchen wir normalerweise alles in Max. Aber es gibt Momente, wo wir sagen: OK, es gibt kein Budget für die Programmierung, dann wird einfach von irgendwas gespielt, es gibt zum Beispiel Synergy, die machen auch eigene Player, mit denen arbeiten wir auch oft, da laden wir dann alles rein, ich sage ok, ich weiß nicht, was passiert. Das sind dann WAV-Files, Mono oder Interleaved, und die spielen das dann ab.

DS: Wenn wir eher so in Richtung Spielstätte denken, auf welchem Lautsprechersystem werden die zum Beispiel wiedergegeben? Das ist dann einfach eine kanalbasierte Ausspielung?

FS: In der Regel ist das eine kanalbasierte Ausspielung, ja. Weil ein System wie Wellenfeldsynthese extrem teuer ist, gibt es sehr selten einen Renderer vor Ort. Wenn es eine Wellenfeldsynthese ist, dann arbeiten wir in den Renderer, und dann wird sowieso kanalbasiert ausgespielt. Es ist sehr selten, dass wir irgendwas anderes ausspielen als kanalbasiert. Nur wenn es interaktiv ist, dann muss man etwas anderes einbauen. Man baut eine bestimmte Kette von Sachen, man exportiert einen File, man exportiert die XY-Daten in einer Textdatei oder irgendwas anderem, und der Player liest das. Aber nur, wenn es irgendwas interaktives gibt, dann kommen die Koordinaten von irgendwas, das schon gemacht ist, oder durch Inputs, Sensoren. Aber alles, was linear ist, wird kanalbasiert gespielt.

schicken, wo der Sound kommen kann. Nehmen wir einen Raum, und ein Sound muss einer Person folgen, das heißt, es wird getrackt, das heißt, wir kriegen schicken, wo der Sound kommen kann.

UNTERBRECHUNG WEGEN STREAMING-PROBLEMEN.

DS: Legst du als Kreateur wert darauf, dass deine Mischung auf möglichst vielen Wiedergabesystemen funktioniert, oder mischst du für eine einzige Spielstätte zu einem einzigen Zweck?

FS: Eher Version B. Weil wir produzieren weniger Sachen, die woanders gezeigt werden. Also wir produzieren wenig Kinofilme oder eine bestimmte Art von Inhalt, der in verschiedenen Formaten funktionieren sollte. Wir produzieren für einen bestimmten Raum, für einen bestimmten Zweck und legen nicht so viel Wert darauf, dass man es wo-

anders spielen kann. Wenn jemand eine Stereo-Mischung davon braucht, dann machen wir die extra aber wir gucken nicht, dass das Ergebnis von vornherein kompatibel ist.

DS: Stimmst du dann die Mischungen vor Ort ab?

FS: Ja, jedes Mal. Sehr wichtiger Schritt, sehr sehr wichtiger Schritt. Egal, was du davor machen kannst, den Raum kann man nicht simulieren. Egal, wie gut dein Testaufbau ist, den Raum kann man nicht simulieren, das wird schwierig. Natürlich, das ist sowieso so eine Sache, damals hab ich auch eine Mischung mit Andy [Mühlschlegel] gemacht. Im Prinzip arbeitet man objektorientiert. Und wenn das System gut ist, sollte es genauso gut auf deinem System wie auf einem System klingen, was irgendwo anders ist, ohne Ambisonics. Denn du machst die gleichen Koordinaten, das gleiche System, und wenn dein System richtig abgestimmt ist, sollte man keine Probleme haben. Nichtsdestotrotz, dein Raum hat einen anderen Hall, andere Impulse Responses, und das wird immer anders klingen.

Andy [Mühlschlegel] und ich haben ein Projekt gemacht, wo wir objektorientiert gearbeitet haben. Die haben in Stuttgart ein ziemlich gutes Studio mit Ambisonics, und dann haben wir quasi eine Präsentation gemacht in Frankfurt mit einem anderen Ambisonics. Und wo der Punkt ist, ist, ob das System richtig kalibriert ist. Denn im Endeffekt bleiben die Koordinaten und wenn das System richtig gebaut ist, sollte es keinen großen Unterschied geben. Ein Raum bleibt aber immer ein Raum, und wird immer anders klingen. Und es ist immer gut, vor Ort eine Anpassung zu machen.

DS: Ok, dann kommen wir zum letzten Punkt, vielleicht auch auf persönlicher Ebene, Stichwort Immersionserlebnis. Von welchen technischen Parametern hängt denn deiner Meinung nach das Immersionserlebnis ab?

FS: Es hängt natürlich am meisten vom Inhalt ab. Aber die Hardware-Komponenten sind natürlich auch nicht unwichtig. Ein immersives Erlebnis ist mit einem Mono-Kanal natürlich Bullshit. Außer es ist genau das, was man will. Aber eine der großen Challenges ist, wenn man jetzt so einen Industrie-Job macht, ein Museum macht oder so, ist, die Balance zu finden zwischen dem Budget, das es gibt für die Lautsprecher und das immersive Erlebnis, das man machen will. Natürlich alle Kunden wollen ein Erlebnis so immersiv wie möglich haben, aber

haben ein Budget für drei Lautsprecher. Da sagen wir: Lass' uns nicht lügen, das ist nicht möglich damit. Man will mindestens einen surround-artigen Sound haben, und dann am besten noch einen Ring unten, einen Ring oben. Aber klar, man macht das beste mit dem Setup, das es gibt. Aber natürlich hängt es von der Anzahl der Kanäle ab, die man hat, und wie der Raum gebaut ist. Natürlich, wenn man eine Wellenfeldsynthese mit 4 Ringen oder 2 Ringen bauen kann, hammer geil, kann man mega viel machen, aber was kostet das? Es ist schon sehr wichtig, dass man da eine bestimmte Abstimmung von den Möglichkeiten und den Medien hat. Wir machen keine Medientechnik bei uns im Klingklangklong aber wir machen ein Consulting, was das angeht. Und bevor überhaupt anfangen zu produzieren, stimmen wir ab, welches System wir vorhaben. Um ein möglichst großes immersives Erlebnis zu haben natürlich.

DS: Ok, und wenn wir das jetzt mit meiner Fragestellung des Metadatenformats verbinden. Glaubst du, dass wenn es ein standardübergreifendes Metadatenformat gäbe, was beispielsweise auch jeder Player auslesen könnte, würde das das Immersionserlebnis steigern?

FS: Ich glaube, das hat schon eine Verbindung, nicht aber direkt, glaube ich. Du kannst das beste Format der Welt haben, wenn am Ende Inhalt und Hardware nicht stimmen, ist es völlig egal. Was da natürlich helfen kann, ist ein bestimmte Optimierung des Prozesses, dahin zu kommen. Ich glaube, dass es nicht wirklich direkt einen Einfluss hat auf das Erlebnis, was man da hat, sondern einen Einfluss hat in den Prozess, den man hat, um dahin zu kommen. Und das könnte in bestimmtem Fällen, vor allem, wenn man will, dass man den gleichen Inhalt auf verschiedenen Formaten und in verschiedenen Räumen spielen kann, kann wahnsinnig gut was bringen und den Prozess optimieren. Ich glaube aber nicht, dass dies eine direkte Verbindung hat.

DS: Ok. Dann, ja würde ich sagen, war's das. Vielen Dank. Ich schalt' jetzt gleich mal ab.

FS: Sehr gerne, sehr sehr gerne!

Transkription - Interview I2

Interview vom 30.11. 2017

Interviewter: Martin Rieger

Abkürzungen: MR = Martin Rieger; DS = Daniel Strübig

DS: So Martin Rieger. Ich sitze hier mit Martin Rieger über Skype im Rahmen meiner Bachelorarbeit und Martin, du arbeitest mit immersivem Audio in welchem Kontext genau?

MR: Also ich bin spezialisiert auf 360 grad - habe also quasi einen Schritt gemacht von einem herkömmlichen Tonmeister in Anführungszeichen auf diese Virtual Reality Schiene zu gehen. Das hat mehrere Gründe, einfach weil ich Medientechnik studiert habe und gerade bei 360 Grad gemerkt: Man braucht mehr Wissen als nur Ton, man muss auch wissen wie ich Videos encodieren muss, muss zum Teil auch selber so Programmierzeilen irgendwie hier und da mal machen, auch wissen, wie wird das hinterher implementiert und solche Geschichten, die eben ein normaler Tonmann nicht weiß. Deswegen habe ich gemerkt, das macht Sinn zu sagen, da die 360 Grad Videos jeweils nur so 3, 4, 5 Minuten lang sind: Das kann ich alleine stemmen und das kann ich am Set stemmen und in der Postproduktion, dann kann ich das eben Vorhinein planen: Wie werde ich das aufnehmen am Set und hab dann schon in Mischung im Kopf. So weiß ich, welche Töne sind nötig, wo muss ich aufpassen was es spart jetzt zum Beispiel Zeit in der Mischung zu machen und jetzt mal nicht mitzunehmen- solche Geschichten. Deswegen habe ich gemerkt: Ok, da macht das wirklich Sinn, mich zu spezialisieren auf dieses Thema, weil ich will keine Spielfilme oder so was machen. Aber wie gesagt, in diesem Kontext haben viele Faktoren einfach dafür Gestimmt dass ich das mache.

DS: Ok. Stichwort immersives Audio, das machst du für VR. Jetzt ist es ja so, dass VR in aller Regel über Kopfhörer abgespielt wird. Du musst trotzdem immersives Audio in alle Richtungen entwickeln- das muss abspielbar Kopfhörer sein- Wenn wir jetzt an einen Mix denken: was für verschiedene Audiotypen verwendest du da? Also sagst du: ich arbeite mit Objekten ich arbeite mit ganzem einem herkömmlichen kanalbasierten mono, stereo-Audio oder nutzt du szenenbasierte Lösungen wie Ambisonics beispielsweise… Wie gehst du da vor?

MR: Also aktuell bin ich mit einer Mischung ganz zufrieden. Das ist eigentlich… Also ich will es jetzt nicht mein eigenes Setup nennen, aber ich habe es mir überlegt und hab gemerkt, das funktioniert sehr gut, so wie ich auch in der Postproduktion arbeite, dass ich einerseits versuche möglichst isolierte Mono-Schallquellen aufzunehmen, also sei es durch Anstecker oder dass ich vielleicht doch irgendwie

Mikro verstecke oder das einfach als Nurton irgendwie noch aufnehmen oder dann die Foley in der Postproduktion, aber dass sich halt einerseits diese isolierten Objekte in Anführungszeichen in dem Stadium noch keine wirklichen Objekte im Kontext von objektbasiert, du weißt, was ich meine… Aber ich versuche, am Set Signale zum Beispiel mit Ambisonics Mikrofonen mitzunehmen, die bieten sich einfach an, weil die kann ich gut unter der Kamera verstecken. Da bringt mir ein 3D ORTF-System nichts, dass das dann irgendwie eine Riesenkiste oder so ist und mit Recorder und Kabelage und alles, da würde ich nur die Kameraleute nerven, und so kann ich das immer relativ schlank ans Kamerastativ riggen und kann das eben wie gesagt zusätzlich zu den Funkstrecken laufen lassen und weiß ok, so habe ich eine gute Mischung ist macht für mich nicht Sinn, alle Töne, die da sind auch mitzunehmen, weil man hat immer irgendwie Übersprechen. Aber so ist das immer ein ganz guter Kompromiss, wo ich die Szene schon mal so habe. Und dann kann ich nachher immer noch abwägen, ok, brauch ich vielleicht auch irgendwas oder nicht. Aber so ist das Setup, dass ich auch quasi für den O-Ton verwende. Und die ganzen Hall-Geschichten mache ich einfach über Impulsantworten, ich hoffe mal, dass der natürlichste ein wenig bekommen kann und es gibt auch einfach noch nicht so gute Ambisonics Hallräume. Manchmal sind es dann doch so ganz spezielle Sachen, und wenn ich mir das einfach über das Ambisonics-Mikrofon nehme, das ist eine ganz neue Art des Mischens irgendwie. Sobald man das eine Mikro irgendwie verändert, dann verändert man auch an dem anderen irgendwie was, also man kann nicht sagen, das baut irgendwie aufeinander auf, sondern das greift eher von vornherein irgendwie ineinander. Je nachdem auch, wie man das eben aufnimmt.

DS: Also arbeitest du schon während der Aufnahme mit hybriden Typen aus kanalbasierten und szenenbasierten Audiotypen. Und worauf legst du denn in der Phase der Kreation der Mischung besonders Wert?

MR: Es hängt einfach stark von dem Inhalt natürlich ab, aber ich würde generell sagen: Das wichtigste ist die Sprache und dann ist erst mal egal ob da irgendwie schon schön Hall drauf ist oder Foley oder was auch immer, es soll erst mal irgendwie die Sprache verständlich sein. Das klingt erst mal irgendwie nur durch einen Anstecker, der irgendwie sinnvoll unter der Kleidung versteckt wurde. Das ist auch schon ein Problem, dass dann eben auch mal schnell Gera-

schel oder so was drauf landet. Aber ja, deswegen würde ich erst mal sagen die Sprache, damit kann man schon viel machen, die kann man ja auch dann schön als Mono-Quelle irgendwie platzieren, sei es jetzt Ambisonics oder objektbasiert oder irgendwas. Aber die hat man dann irgendwie schon mal sauber und von dort kann ich dann immer noch schauen ok, was macht noch irgendwie alles Sinn.

DS: Und denkst du während der Postproduktion, also während des Mischens schon daran ein finales Datenformat, also was du am Ende ausspielst und eben dem Kunden gibst zur Veröffentlichung sowie an das Abspielgerät, also worauf das später abgespielt wird?

MR: Ja also das muss ich eigentlich sogar schon vor der Mischung wissen. Wenn ich mir jetzt etwas in Ambisonics, was dumm wäre, erster Ordnung… aber nehmen wir mal an, und ich eigentlich weiß ok, das Endgerät wäre irgendwas… fünfter Ordnung brauchen würde… keine Ahnung, das gibt mir jetzt wahrscheinlich gar nicht. Aber einfach vom Prinzip… Da bin ich jetzt schon auf diese vier Kanäle beschränkt, und kann das nicht hochskalieren. Und das ist eben genau die Krux an Ambisonics, deswegen, also ich mag das, das funktioniert auch gut, das hat auch seine Daseinsberechtigung, aber es ist eben nicht so schön skalierbar wie jetzt einfach objektbasiert. Und ich nehme jetzt schon ein bisschen was vorweg, weil das würdest du wahrscheinlich eh schon fragen wollen… Mein Traumszenario wäre irgendwie, dass man so eine Arbeitsumgebung hat, wo man einfach irgendwie im 3d-Raum seine Schallquellen platzieren könnte, was also irgendwie objektbasiert wäre, weil bei Ambisonics ist man immer irgendwie an Kanäle beschränkt und kannst das aber nicht höher skalieren, wenn du das mal tun wolltest. Und so kann man kann seine Töne irgendwie im Raum platzieren, hat zum Beispiel aber zusätzlich seine Ambisonics-Atmo, die ich jetzt einfach als Set habe, und dass ich weiß, diese Sound-Kugeln (so nenne ich die gerne, weil es ja irgendwie aus allen Richtungen kommt) okay die ist jetzt fertig. Dass ich sie mir einfach in alle Formate rendern kann, dass ich sag: Exportieren ich will das jetzt als dritter Ordnung Ambisonics, oder ich will das jetzt als mpeg-H oder… du wahrscheinlich mittlerweile viel mehr Formate aber dass das halt einfach so ein… wie beim 3D-Modelling, da kannst du es dir auch in unendlich hohen Auflösungen herausrendern, so wie du magst und genauso was fehlt irgendwie beim Ton.

DS: Ok und arbeitest du auch mit interaktiven Elementen oder ist das was, was auch auf dich zukommt?

MR: Gute Frage, weil da es auch so die Unterscheidung zwischen VR und 360 Grad. 360 Grad ist er so ein Spezial Teil von VR. Es gehört schon zur VR aber VR, so sagt man ist eigentlich immer interaktiv. Und 360 Grad meistens, also fast immer linear. das ist jetzt nämlich das Ding, dass da immer mehr Mischformen gibt. Also ich hab jetzt eine Produktion da ist erst ein 360 Grad Video, dann ist da so ein Quiz, wo du halt interaktiv irgendwie was auswählen kannst und dementsprechend ändern sich dann auch die Videos und so was. Und dann ist das auf einmal irgendwie Game Audio, also das ist was, wo ich so ein bisschen die Finger von lasse, weil ich weiß: Ok, sollen das lieber Leute machen, die halt Game Audio machen, aber dadurch, dass ich oft daran schramme und das auch mache, macht das Sinn da eben weiter zu schauen: Okay wie funktioniert das jetzt? Ja weißt du das ist eine ganz neue Kiste, die ich da irgendwie aufmachen muss. Zum Beispiel mit Unity irgendwie anfangen mit einzelnen Schallquellen zu arbeiten und dann wird es auch wirklich programmierlastig und wenn man dann mit Middleware irgendwie arbeiten muss und so… da bin ich immer daran vorbei gekommen, ich weiß auch, ich will jetzt nicht anfangen Spiele zu vertonen aber eben wie gesagt, wenn es dann so Kleinigkeiten sind wie so ein Quiz, und die setzt man dann in einen Raum oder gibt es den Programmierer oder so. Also das ist alles noch irgendwie im Rahmen von der Interaktion her.

DS: Also rudimentäre Interaktion, wenn überhaupt.

MR: Ja, würde ich so sagen.

DS: Okay vielen Dank. dann kommen wir jetzt zum Themenblock finales Datenformat. Das haben wir gerade schon angesprochen… Wie sieht denn deine Abgabedatei aus, also jetzt sind wir im nächsten Schritt: Der Mix ist fertig, was exportiert du dann?

MR: Meistens einfach das, wie ich es angelegt habe. Wenn ich in erster Ordnung [Ambisonics] angelegt habe, spiele ich das auch in erster Ordnung wahrscheinlich aus, oder zweiter, dritter, was auch immer. Das ist aber meistens dann nicht die Datei die ich dann raus schicke. Weil das Problem, was ich am Anfang hatte, dass die Leute halt mit dem Thema… dass das so neu für die ist, dass ich gemerkt habe: Okay es macht keinen sinn dass ich denen jetzt irgendwie eine

achtkanalige Datei schicke und die wissen überhaupt nicht, was sie damit machen sollen und wie sie das abspielen sollen. Dann mache ich es so, dass ich das Video halt selber encodier, denen schicke und sage: Ja hier, ihr müsst das mit dem Player mit dem Device anschauen und dann funktioniert's. Weil dann habe ich die Garantie: Das Video, was ich gemacht habe, das funktioniert. Weil dann so viele Sachen passieren können. Da ist eben genau das Ding, da hat jemand den Video Player genutzt und meinte, das funktioniert nicht. Beziehungsweise war so irritiert, dass das eine funktioniert hat aber das andere nicht. Aber das war nicht der Fall, der hat sich da einfach irritieren lassen, weil's eigentlich gar nicht funktioniert hat, aber er dachte, dass die Stimme funktioniert aber alles andere nicht. Und so Sachen. Also da ist eben das Problem, dass das Thema so neu ist für alle, dass ich sie da lieber bei der Hand nehme und halt wirklich ein ganz finales File gebe, was bedeutet, dass sich halt immer ein Gigabyte große oder zwei Videos hochladen muss. Das ist natürlich da auch wieder so ein bisschen nervig, wenn man eine kleine Änderung haben will, muss man halt immer solche Files schicken. Aber das ist gerade für mich so die beste Lösung, weil dann kann ich halt wie gesagt sicherstellen, dass das alles so läuft wie ich mir das vorstelle.

DS: Und in welchem Umfang schränkt dich das Datenformat ein? Also jetzt mal nur der Audio-teil, wir nehmen das Video weg. Hast du da schon manchmal gedacht: Jetzt wär's irgendwie schon geiler, noch ein bisschen mehr Funktionalität zu haben?

MR: Total, also wenn ich wenn ich die Ambisonics Datei haben kann ich mit keinem Player irgendwie abspielen. Ich muss mir jetzt erst mal wieder irgendein Player herunterladen, wo ich dann die Dateien laden kann und dann kann ich dann irgendwie so Sachen rotieren oder so. Aber dann weiß ich auch nicht was ist da jetzt für ein Decoder irgendwie drin, wie klingt der, ist der binaural, ist das einfach stereo, ist das irgendwas keine Ahnung. Also deswegen macht es für mich auch keinen Sinn, so einzelne Tondateien raus zu schicken, weil die kann ich ja zum Teil selber nicht richtig abspielen. Da muss ich immer irgendwie erst mal in der Workstation an checken und schauen, wie die klingen.

DS: Ok also welche Anforderungen willst du jetzt speziell stellen an beispielsweise in Ambisonics-Decodierung?

MR: Dass das halt so kompatibel ist wie zum Beispiel eine mp3, die kann mittlerweile jedes Gerät abspielen aber das sah vor 15 Jahren auch anders aus. Da brauchtest du du einen mp3-Player, aber Handys konnten das nicht abspielen wiederum. Ich denke, das wird in Zukunft noch alles irgendwie kommen, dass auch der Windows Media player irgendwie Ambisonics abspielen kann aber genau das kann er eben nicht. Das heißt, das muss so funktionieren, dass jeder Player erkennt: So aus den Metadaten oder was auch immer, das ist jetzt eine Ambisonics Datei, und die muss ich jetzt soundso abspielen. Und bestenfalls noch so etwas wie: Das da so das ist jetzt mit dem und dem Decoder am besten zu hören und der ist dann auch in jedem Player drin. Aber genau das ist einfach nicht der Fall, man braucht immer einen speziellen video-audio-player.

DS: Und wie soll dann der nächste Schritt aussehen deiner Meinung nach? Also um es beispielsweise auf Kopfhörern abzuspielen, ein Ambisonics-Format sollte da binauralisiert werden wie siehst du da den aktuellen Stand der Technik?

MR: Im Bezug auf was?

DS: Binauralisierung und den Zugang zu Binauralisierung.

MR: Also genau das Ding an Ambisonics ist: Ja, das ist schon nett, wobei das trifft nicht nur auf Ambisonics zu, das kann auch eine 5.1-Mischung sein, die binauralisiert werden soll… Es ist nicht personalisiert. Das ist halt immer das Ding, dass es halt nicht personalisiert ist. Und genau deswegen wär es cool, wenn jetzt jeder Player Ambisonics abspielen kann, aber selbst dann klingt es immer noch nicht so, wie das für einen irgendwie personalisiert cool wäre, sondern man hat immer so eine standardisierte Kunstkopfform, von der man auch überhaupt nicht weiß, wie sie aussieht, was für Maße sie hat und ob man die verändern kann oder so. Es gibt halt einfach einen Button "Binauralisieren" und mehr man auch nicht wirklich einstellen. Das kann sehr unbefriedigend sein, weil dann natürlich auch jeder Hersteller anders klingt. Wie gesagt, dass es dann schon auf Facebook anders klingt als über Youtube.

DS: Okay also die generische HRTF ist momentan noch die Krux für dich?

MR: Ja würde ich sagen. Ich glaube aber nicht, dass es das größte Problem ist, sondern dass das dann eher so eine Feinheit ist. Ich glaube, das viel größere Problem ist einfach, dass wenn ich so solche Dateien jetzt meinen Freunden schicken würde, die könnten damit nichts anfangen. Das ist schon mal das Problem.

DS: Also fehlt einfach der Zugang momentan?

MR: Ja irgendwie so ein bisschen das Bewusstsein, weil wenn ich erzähle was die so machen verstehen die das aber meinen dann: Och, also so das hab ich noch nie gehört oder so. Also das ist noch so ganz am Anfang, dass die Leute das einfach noch nicht kennen und auch den Unterschied nicht kennen. Also da müsste man eher so eine Aufklärung machen, dass man denen mal einen statischen Mix zeigt und dann irgendwie einen binauralisierten, der dann räumlich ist. Was ich auch gemerkt habe, ist, dass Leute einfach dieses räumlich Hören nicht gewohnt sind, wenn sie es nicht machen. Und ich glaube, das habe ich dir mal erzählt. Das war für Haman Cadon, das darf ich vielleicht nicht unbedingt sagen… Aber es war halt so. Dann ging es darum, wie Lautsprecher im Auto klingen und eigentlich muss man das nochmals viel aufwendiger machen, dass das realistisch ist und so. Aber jetzt letztendlich war es ja so, dass ich einen Song hatte, da hatte ich dann die Stereo Mischung, die dann im Radio lief. Und einmal hatte ich die 5.1 Version, die ich dann binauralisiert habe, die ich also in Ambisonics übernommen habe und die dann live binauralsiert wurde. Und das klang einfach nicht besser als das, was man gewohnt, sondern da spielen vielmehr so Kriterien rein wie: Was sind dann die Hörgewohnheiten und Leute sind so einfach gewohnt, dass sie sich ein Stereo ding auf die Ohren hauen dann klingt das halt sehr direkt. Aber bei den binauralisierten Sachen, das klingt ja immer irgendwie räumlich und damit auch so ein bisschen entfernt und da muss man sich fragen: Ist das eigentlich das, was die Leute hören wollen? Einerseits ist es irgendwo realistisch, es klingt ja dann irgendwie so als würde man in einem Raum sitzen mit Lautsprechern. Aber das macht zum Beispiel bei Musik für mich nicht so viel Sinn, also ich habe da keine guten Erfahrungen gemacht. Ich hab Kollegen machen auch viel in der Richtung, ja hier neue Binauralisierung und so. Aber ich denke mir da: ja wieso? Die Leute sind es einfach nicht gewohnt, dadurch dass ein Bruchteil der Leute kennt, wie es überhaupt gelingt räumliche Musik zu hören über Kopfhörer, dadurch ist es halt so ein

anderes Hörerlebnis, dass ich gar nicht weiß, ob sich das auch auf Dauer irgendwie durchsetzen wird. Aber dieses räumliche Hören macht wiederum für Filme Sinn, deswegen mach ich es auch, weil da ist es noch mal was ganz anderes wenn du wirklich auch klanglich im Raum drin bist, den du da siehst.

DS: Also wünscht du dir am besten eine Mischform? Also dass du sagst: Ich kann bestimmte Signale binauralisieren. Und manche kommen einfach ganz normal head-locked?

MR: Genau das ist nämlich das Ding an Ambsonics, deswegen ist das kein Allheilmittel. Vielleicht kennst du G-Audio Lab, die haben eine ganz clevere Herangehensweise. Ich habe es leider auch noch nicht ausprobieren können. aber die haben irgendwie bei MPEG-H mitgearbeitet und ich weiß gar nicht so genau was jetzt eigentlich MPEG-H ist. Ich weiß, das soll ja angeblich alles können. Aber anscheinend meinten die, ja es kann nicht alles gleichzeitig also es kann nicht Ambisonics und objektbasiert und noch eine statische Spur oder was. Korrigier mich, wenn ich falsch liege. Das ist jetzt meine Vermutung, deswegen haben sie ihr eigenes Format gemacht, wo man eben genau das machen kann. Und das ist eine clevere Idee, dass man sagen kann: Ok, du hast einen Ambisonics-Bed, genauso wie ich es halt eben am Set habe. Aber du hast dann Objekte und keine Ambisonics-Quellen. Und das ist schon mal ein großer Schritt, weil dann hast du eben genau das Ding, dass das eigentlich im nachhinein rendern kannst, wie du magst. Das Ambisonics-Bed kackt dann eben ein bisschen ab, aber mein Gott, wie gesagt, das ist halb eben ein bed, also immer noch okay. Und dass du zusätzlich eine statische Stereo Spur hast für Musik. Weil das ist eben genau das Ding, was ich meinte: Macht es überhaupt Sinn, Musik binaural irgendwie zu hören?

DS: Ja, nachdem Themenblock jetzt ausführlich besprochen haben…

MR: Sorry, nur eine kleine Ergänzung: Also ich hab auch tatsächlich auf meiner Homepage einen Artikel dazu geschrieben weil ich mich selber gefragt habe: Wo macht es denn Sinn, Musik irgendwie normal zu hören?

DS: Den hab ich auch gelesen, aber du darfst den gerne nochmals zusammenfassen.

MR (lacht): Ok, dann sag ich das noch schnell. Das einzige Szenario, wo ich mir denke, es macht Sinn, das ist einfach ein live Konzert.

Weil da bist ja mittendrin und das ist halt diese Schnittstelle zum O-Ton. Wo ich denke, da macht das Sinn, diese räumlichen Komponenten mitzunehmen. Aber… also 360 Grad Musikvideos sind cool. Aber muss ich da jetzt auch einmal die Musik mitdrehen, und vor allem muss ich eine neue Mischung und so was machen. Und theoretisch ist das ja auch für den Künstler komisch, zwei Mischungen von meinem Song: Ich will den einen, der kommt im Radio, fertig.

DS: Dann gehen wir einen Schritt weiter. Also du hast auch deinen Film final exportiert, die finale Version wurde abgesegnet vom Kunden und dann können die Konsumenten das genießen. Auf welchem Abspielsystem wird denn der Inhalt, den du machst in der Regel wieder gegeben?

MR: Also zum Glück einiges eben auf Youtube oder Facebook, weil das funktioniert einfach. Das ist immer sehr beruhigend zu wissen, weil dann habe ich halt ist eine Version hoch weiß: es läuft, das klingt bei mir genauso wie bei denen. Das war noch vor einem Jahr ein bisschen anders, da lief es irgendwie, allerdings nicht auf Apple Geräten. Und das kam jetzt aber Schritt für Schritt, also jetzt kann man sagen: Das wird überall unterstützt.

Das war es aber auch schon, weil wenn ich halt jetzt ein Video habe und das würde ich irgendjemandem schicken wollen, dann müsste ich sagen: Ja du musst jetzt den und den Player nehmen und keine Ahnung ob das in einem Monat noch funktioniert. Also vielleicht hat der Player auch einen Bug, der selbständig ein Problem macht. Ich kann zum Beispiel meine Videos nicht mehr auf Brille schauen, da geht es schon mal los. Weil ein Software-Update kam von Oculus, vom Oculus Home Video Player, und jetzt ist also so, dass nach einer halben Minute irgendwie auf einmal alles kollabiert, das Sound Field klingt so total phasig warum auch immer. Und das ist eigentlich, was mich jetzt seit eineinhalb Jahren irgendwie verfolgt, dass ich halt immer erst mal anchecken muss: Funktioniert eine Mischung auf MEINEM Gerät überhaupt noch?

Und dann da mal anzufangen zu sagen: Okay wir wollen für die Konsumenten und alles machen, da ist die einzige Lösung, die auch wirklich gut funktioniert, sich eine eigene App zu bauen: Also sei es halt einen eigenen Video Player, wo man halt dann irgendwie die Video und Ton reinspielt, und dann kann man sich ziemlich sicher sein. Da gibt es ja noch keine Updates, sondern das funktioniert einmal,

denn dann wird's hoffentlich auch immer so funktionieren. Aber sobald man auf irgendwelchen Plattformen unterwegs ist: Ich weiß nicht, ob Youtube überhaupt noch in fünf Jahren funktionieren wird. Das ist diese unangenehme Unsicherheit, was passiert überhaupt mit meinem Content.

DS: Ok also wird sozusagen Transkompatibilität ist für dich momentan noch nicht gewährleistet?

MR: Überhaupt nicht (lacht). Was schade ist, weil ich kann verstehen dass sich immersives Audio nicht so durchgesetzt, weil niemand 20 Lautsprecher bei sich stehen hat. Das waren immer so Spezialinstallationen. Aber jetzt ist ja gerade die Chance: Das kann jeder über Kopfhörer miterleben aber ich glaube, dass es die Menschen einfach erst seit kurzem bewusst was man damit machen kann. Aber das ist genau das Ding mit VR, man hat es halt jetzt und es gibt schon ein paar gute Ansätze. Aber was macht man denn jetzt mit der Technik?

DS: Und wenn wir jetzt mal wieder in Richtung Audio Definition Model denken: Würdest du davon profitieren, wenn alle Brillenhersteller, alle Social Media Netzwerk Hersteller wie Facebook oder Youtube, wenn die sagen würden: Okay wir greifen alle auf das gleiche Renderingsystem zurück, wir nutzen alle das gleiche Format an Metadaten.

MR: Also wär voll gut. Also ums kurz zu machen. Weil dann würden sich theoretisch genau diese Probleme lösen, dass es halt immer irgendwie anders klingt, dass man immer irgendwelche anderen Formate braucht für irgendwelche anderen Plattformen. Sondern da hat jeder dann dieses eine Format X und das funktioniert überall auf jeder Plattform. Damit könnte man arbeiten. Man sollte nur hoffen, dass es dann auch ein wirklich gutes Format, das dann eben genau diesen Ansprüchen gerecht wird. Das wird Ambisonics nicht, das wird MPEG-H "aktuell" nicht, ich kann es ja nicht mal benutzen. Da geht jetzt schon mal los: Ich weiß nicht, ich habe da keinen Zugriff drauf, was ist das warum soll ich dann damit arbeiten wollen? Ich will irgendwas, was ich nutzen kann und das auch diese nicht niedrigen Ansprüche erfüllt.

DS: Bist du schon mal so weit gegangen dass du hast du gesagt hast: Okay ich mache für verschiedene Wiedergabegeräte also für verschiedene VR-Brillen mache ich verschiedene Mischungen auch?

MR: Ja zurzeit ich versuche es zu vermeiden, weil am Ende merkt's niemand. Das klingt hart, weil das ist genau das, was ich meinte, dass… Für die Leute ist es halt so ein neues Hörerlebnis dass sie mir da vertrauen, wenn ich sag: Das passt so. Dann passt es auch meistens auch für die, was gut für mich ist. Vielleicht sieht's in fünf Jahren ganz anders sonst, dann kennt das jeder. Aber was ich zum Beispiel gemacht habe, da war so eine Installation, wo eben von vornherein klar war: Auf welchem Device ist das? Dann habe ich mich halt eben für ein Decoder entschieden, welcher klingt am besten. dann habe ich das entsprechend dem Programmierer gesagt und dementsprechend auch die Mischung angepasst, damit sie auf den Kopfhörern, die halt eben bei dieser Installation waren, auch einfach möglichst gut klingt. Wobei halt dann trotzdem mit HRTF und so… Das ist halt ein Standard-Ding gewesen und da wusste ich, es wird dann theoretisch genau so klingen wie ich das mir jetzt gedacht habe, weil die nicht irgendwie andere Kopfhörer, andere Devices oder was auch immer nutzen.

DS: Also wünschst du dir auch, so wie es bisher herausgehört habe, speziell die Möglichkeit eine eigene HRTF zu verwenden?

MR: Jein. Ich habe es letztens ausprobiert im Klangstüberl und das ist schon nice. Also wenn es mal richtig gut funktioniert, macht das Spaß. Aber ich glaube, da haben wir echt noch einen langen Weg hin, dass den Leuten überhaupt bewusst wird, das klingt jetzt so, wie ich das irgendwie kenne. Also muss halt schon echt diesen AHA-Effekt haben, dass es den Leuten bewusst wird. Und es kann sein dass es das aktuell genau das Problem ist, dass das was würde und die Leute denken: Ja okay es ist irgendwie räumlich aber auch dass jetzt vorne oder hinten, ach keine Ahnung… Wenn sie auf einmal diese diesen AHA-Effekt hätten,zur personalisierter HRTF, dann hätten sie vielleicht auch ein Bewusstsein dafür, dass haben sie eben gerade nicht. Schwer zu sagen, wo da die Gründe sind.

DS: Okay gut. Dann kommen wir jetzt noch zum letzten punkt also dem Immersionserlebnis. Also ich gehe mal davon aus dass die Immersion des Konsumenten, wenn du VR Audio machst, sehr wichtig ist. Und was würdest du sagen: Von welchen technischen Parametern hängt denn so ein Immersionserlebnis ab auf auditiver Ebene?

MR: So ganz generell?

DS: Ja. was ist so die höchste Priorität?

MR: Ja wie gesagt, HRTF, ne? Das ist schon mal eine Voraussetzung, damit es wirklich gut ist. Da müsste ich noch so ein bisschen überlegen. Klar können wir sagen: Kopfhörer und welcher Content… so vom Gefühl müsste… Was immer gut ist, sind Vergleiche. Man vergleicht als Mensch eh immer alles miteinander, und wenn man irgendwie erst mal hören würde: Okay so klingt es normal und jetzt klingt's immersiv. Dass den Leuten auch wirklich bewusst wird: Was macht es denn? Weil wenn man denen einfach was immersives zeigen würde… Selbst mit der perfekten HRTF kann ich mir nicht vorstellen, dass jeder sofort merken würde: Das ist jetzt das räumliche Hörerlebnis. Das kenn ich auch von den Kollegen, boah, das wird die Welt verändern… Ja, um es mal realistisch zu sehen: Das ist halt immer aus dem Leben eines Tonmenschen, dass es immer nicht so bewusst wahrgenommen wird außer man achtet wirklich drauf.

DS: Und willst du sagen, dass die technische Infrastruktur die du momentan nutzt, bietet das dir genug Möglichkeiten, dass du sagen kannst: Die Immersion ist gelungen?

MR: Ja also ich würde schon sagen, dass das, was ich gerade mache schon Spaß macht und das ist auch das, was ich immer wieder sehe wenn sich Leuten die Brille aufsetzen. Die erste Frage ist zwar immer die ersten Male: mit was für Kameras habt ihr gedreht? Und die zweite Frage ist dann immer: Wie ist das mit dem Ton? Aktuell ticken halt erst mal alle wegen Augen aus, aber das wird natürlich nicht langfristig so sein, irgendwann werden die Leute auch anfangen zu hören: Klingt das denn gut, ist das denn wirklich nur der Kamera Sound jetzt oder dreht sich das auch mit? Darauf achtet aktuell keiner, aber sobald man darauf achtet, merke ich schon, dass die Leute sagen: Oh das funktioniert aber gut. Und ich finde gerade mit Kombination von Bild funktioniert, selbst Ambisonics erster Ordnung zum Beispiel ziemlich gut. Weil dann hast du: Ok, du siehst jemanden, du hörst jemanden, drehst dich so ein bisschen und merkst, es macht irgendwas. Wenn man die Augen zumachen würde, wird das glaube ich schon mal viel schwerer sein, das probiere ich selber immer wieder gerne, und wie ich merke: Okay ich komm nicht immer genau da raus, vielleicht sogar um 180 Grad verdreht. Da hilft quasi das Auge so ein bisschen den Ohren und das die Erfahrung, die ich auch irgendwo gemacht habe. Deswegen kann man sagen, es macht viel Spaß und das

funktioniert schon ziemlich gut aber es ist genau wie das Bild einfach lange nicht perfekt.

DS: Würdest du denn sagen, dass jetzt für dich oder auch in Bezug auf den Konsumenten, dass der davon profitieren würde, wenn Hersteller eben Metadatenformat wie das ADM universell verwenden?

MR: Kann ich mir schon vorstellen, also ich meine, mittlerweile weiß auch jeder, was MP3 ist. Und wenn irgendwann noch jeder weiß, was ADM ist, wäre cool. Also kann ich mir schon vorstellen dass es das neue… wie eben im Bildbereich eben jetzt 4k oder Full HD oder was auch immer ist, dass das dann für Audio das ADM ist.

DS: Dann vielen Dank Martin!

Transkription - Interview I3

Interview vom 04.12.2017

Interviewter: Benedikt Maile

Abkürzungen: BM = Benedikt Maile; DS = Daniel Strübig

DS: Dann, Benedikt, vielen Dank, dass du mit mir das Interview machst. Magst du dich kurz vorstellen und erklären, in welchem Umfang, in welchem Gebiet du mit immersivem Audio arbeitest?

BM: Ich bin Bene Maile, ich bin freier Produzent und Engineer und arbeite mit Popmusik. Die Schnittmenge mit 3D ist eigentlich nur aus interessens- und hobbybedingten Gründen, weil wir halt in der Hochschule angefangen haben, 3D-Inhalte zu produzieren und zu machen und dann festgestellt haben, dass es einfach Spaß macht und total toll ist und es interessant ist, ein mehrkanaliges Format zu bedienen. Wir haben hauptsächlich mit Auro 3D, also mit kanalbasiertem Audio, und mit IOSONO. Das sind meine Schnittmengen damit. Und hauptsächlich geht's mir darum, eigentlich für die Popmusik zu transportieren und sich anzugucken: Wie kann man das machen, was macht Sinn, was macht keinen Sinn, was macht Spaß, was macht weniger Spaß, und zu checken: Warum hat Surround schon der Popmusik nicht funktioniert und warum soll dann 9.1 funktionieren? Und was wurde da alles falsch gemacht?

DS: Ok. Wenn du jetzt deine Aufnahmen fertig hast für eine 3D-Produktion, alle Spuren liegen in deiner DAW bereit. Mit welchen Audiotypen arbeitest du da?

BM: Was meinst du mit Audiotypen?

DS: Nutzt du beispielsweise Objekte oder hast du Ambisonics-Aufnahmen?

BM: Also bisher war es so, alle Sachen die ich gemacht habe, waren kanalbasiert. Das waren Mono-Aufnahmen, ganz wenig Stereo-Aufnahmen und habe die einfach hochgemischt auf 9.1 in den allermeisten Fällen. Und auch nur so angelegt, dass im Endeffekt eine Verteilung der Signale stattfindet, ein ähnliches Bewusstsein wie im Stereo. Es findet wenig Bewegung statt, ist eher statisch, bloß auf ein großformatiges Medium bezogen. Das maximale, was wir hatten, entweder Mono oder Stereo-Signale, auch keine Hallgeräte, die 9.1 konnten oder in 5.1 arbeiten, das heißt, wir hatten seltenst einen "echten" dreidimensionalen, aufgenommenen Content, außer einmal bei einer Mischung haben wir extra noch einen Hall aufgenommen in 8.0.

DS: Und gibt es bestimmte Gründe, warum du sagst, Objekte oder szenenbasiertes Audio machen für dich keinen Sinn? Warum verzichtest du auf die Features?

BM: Das hat einen praktischen Grund. Wir haben mit 9.1 angefangen an der HdM, weil einfach ein 9.1-System da stand, die man ganz popelig bedienen kann ohne Software, die dazu laufen muss, da braucht man keinen Renderer gar nichts. Hier hatte es halt den Grund, dass ich nicht mal ein Panning-Tool brauche, ich kann das ganz einfach mit 2 Surround-Circles lösen, was wesentlich kosteneffizienter ist und viel einfacher. Ich kann halt einfach meine fertigen Stereo-Mischungen nehmen und in ganz wenig Zeit einen fertigen 3D-Mix machen. Das ist ein rein praktikabler Grund und ein finanzieller Grund (lacht). Die Tools muss man halt auch kaufen, das kostet einfach viel Geld. Ich mache ja kein Einkommen damit. Es hat sich halt einfach ergeben, dass ich mit kanalbasierten Sachen arbeite. Und Dolby-Zeug ist auch nicht das billigste, was es so gibt.

DS: Das stimmt. Was ist deine höchste Priorität, wenn du dan den Mix drangehst? Was soll die fertige Datei am wichtigsten transportieren?

BM: Für mich heißt es immer: Form follows function. Und in dem Zusammenhang heißt es für mich, dass es niemals um das System geht, sondern es geht immer nur um den Inhalt der Produktion. Das sind Dinge, die in diesen Sachen sehr falsch laufen, dadurch, dass 3D-Musik in vielen Bereichen noch sehr arg in den Kinderschuhen steckt, kommt es dazu, dass die Leute sich um das System kümmern und dann den Vorteil des Systems nutzen, dass man beispielsweise Sachen kreisen lassen kann und das eher rausarbeiten, anstelle dass sie gucken dass die Musik und dessen Inhalt einen Mehrwert bekommt. Das heißt für mich geht es immer nur um den Inhalt der Musik, dass man den darstellt, und das System einen Mehrwert für die Musik darstellt, nicht umgekehrt. Das ist mir sehr wichtig. Ansonsten ist es mir auch wichtig, weil ich das einfach toll finde, das ist eine geschmacksbedingte Sache: Ich finde, 2-Kanal-Stereo ist ein gutes Format. Und ich finde, dass es bestimmte Dinge etabliert hat, die ich mag und die ich nicht so gerne breche. Ich mag keine 3D-Mischung, in der die eine Gitarre von links hinten kommt und der Bass dann von rechts hinten und das Schlagzeug von links oben. Da sehe ich ganz wenig ästhetischen Sinn dahinter. Sondern ich mag bestimmte Dinge, die sich in den letzten 50 Jahren etabliert haben und ich würde die gerne einfach weiterhin nutzen. Beispielsweise dass bestimmte Panoramen festgelegt wurden. Ich mag das Schlagzeug in der Mitte vor mir. Ich mag den Sänger in der Mitte vor mir, ich mag die Gitarre, wenn's ne

Pop-Produktion ist links und rechts, wenn eine Dopplung da ist. Das sind einfach Dinge, die ich mag und die ich ungern breche, außer es macht wieder einen inhaltlichen Sinn. Das ist auch eine natürliche Darstellung, weil wenn ich auf einem Konzert stehe, stehen die Leute vor mir. Ich will keinen Sänger, der hinter mir singt, es sei denn, es ist eine bestimmte Effekt-Passage, wo es total Sinn macht. Aber bei einer Jazz-Produktion muss ich nicht anfangen, die Sängerin hinter mich zu mischen, das finde ich doof. Das würde mir geschmacklich einfach nicht gefallen. Das sind die 2 grundlegenden Dinge, an die ich mich halte. Die Musik ist das wichtigste. Und ich möchte, dass bestimmte Dinge, die im Stereo etabliert wurden, behalten werden und nicht gebrochen werden.

DS: Und wenn du deinen fertigen Mix dann ausspielst, was nutzt du dann momentan als Container-Format?

BM: Gar keins. Also momentan immer nur WAV-Dateien. Ich bounce mir da einfach ein 9.1, also 10 Mono-Dateien und lege die jedes Mal wieder an. Weil jedes Mal war es so, dass es noch nie auf einem Medium gelandet ist, sondern wir haben die Sachen immer nur so exportiert, dass wir sie an Vorträgen vorspielen und ganz klar die Kanäle einfach zugewiesen haben. Das heißt, wir hatten immer ein Pro-Tools oder ein Cubase oder irgendein Mehrkanal-Abspielgerät direkt dabei. Wir haben nie ein Container-Format benutzt in der Form.

DS: Gab es da Momente, in denen du dachtest, die 10 Mono-Wavs schränken dich ein in bestimmten Dingen?

BM: Naja es macht halt keinen Spaß (lacht). Es wäre viel einfacher, im Projekt anklicken zu können: Bitte bounce mir ein passendes Format für ein passendes System anstatt 10 Einzelsignale ausspielen zu müssen, die man wieder anlegen muss. Es hat uns nicht eingeschränkt, aber es ist halt etwas, worüber man sich mehr Kopf machen muss. Es ist halt ein Roh-Format in dem Moment. Das eigentliche Format bedingt sich wieder nur durch die Anordnung.

DS: Und gab es schon mal Momente, wo du gesagt hast, dass deine 10 Mono-Wavs bestimmte Anforderungen nicht erfüllen können, beispielsweise an den Grad der Immersion, den du haben möchtest?

BM: Naja, es ist so: Ich gucke mir an, was das Format kann und höre zu, was das Format kann. Und wenn ich 9.1 habe, dann weiß ich halt,

was funktioniert, wo die Grenze ist und wo es dann schwieriger wird, und dann gucke ich, dass ich das Arrangement von meiner Musik so bastele, dass das Format wiederum passend dem Arrangement dient und passe das halt entsprechend an. Also gab es selten die Grenze, wo ich offensiv gedacht habe: Oh, jetzt wird's blöd, weil jetzt ist es sau eingeschränkt und die Immersion findet nicht mehr statt. Es ist auch so, dass die Immersion in der Popmusik, die wir gemacht haben, gar nicht so wahnsinnig entscheidend ist. Immersion ist für mich auch immer eine Art von Natürlichkeit die stattfindet, dass man halt da massiv eintauchen kann. Wenn wir ein Klassik- oder Jazz-Konzert haben, was als Konzert stattgefunden hat, da finde ich es toll, wenn es immersiv ist und ich eintauchen kann, weil ich mich wie ein Zuschauer, der zuhört und es spürt. Bei Produktionsmusik, von der wir hier reden, ist es so: Da geht es ja darum zu gucken, wie ich das Format nutze, um beispielsweise in einem Rock-Song noch mehr Größe und mehr Dichte zu bekommen und wie ich diesen Effekt von Strophe auf Refrain noch mehr pushen kann als bei einem Stereo-File. Da kann ich sagen, ich fahr die Strophe mono und den Refrain Stereo, und in 9.1 oder Auro kann ich sagen, ich fahr den Verse Mono und den Refrain in 9.1. Also ist für mich Immersion in diesem Zusammenhang auch einfach anders definiert. Es geht für mich da mehr um die Frage des Effektes.

DS: Du meintest schon, dass du gemeinsam mit Daniel [Schiffner] eure Produktionen auf Vorträgen präsentierst. Auf welchen Systemen wird es dort wiedergegeben?

BM: Das kommt immer drauf an. Normalerweise war es so, dass 9.1 möglich war und dass es angeboten wurde, ob wir objekt- oder kanalbasiert ausspielen. Es gab aber auch schon den Fall, wo es kein klares kanalbasiertes Format gab, was nicht richtig angeordnet war, und es klang auch dementsprechend unschön. Das war ein Ambisonics-System, was missbraucht wurde, um kanalbasiert wiederzugeben. Aber wenn die Boxen falsch aufgestellt sind, ist es wie beim schlechten Stereo. Da klingt es halt ok, aber da hat man meiner Meinung nach keinen Immersionsgrad, weil für mich da alles ein bisschen klöppelt, weil da nichts passt, denn man hat es ja nicht so gehört und nicht so angeordnet.

DS: Um es runter zu brechen: Welche technischen Parameter stehen zwischen dem, was du abspielst und dem, wie es wiedergegeben wird? Und was kannst du davon nicht kontrollieren?

BM: Das Problem ist, dass die Phantommitten nicht mehr stimmen, dass Distanzen nicht passen, dass die Höhen nicht richtig stimmen, denn eine Box war ein wenig niedriger angeordnet als die andere, die eine ein wenig weiter außen als die andere. Das kanalbasierte System bedingt ja dem, dass es eine kanalbasierte Anordnung hat und dass diese eingehalten werden soll, sonst funktioniert es ja nicht. Sonst hat man jede Menge Mono-Quellen, die irgendwo im Raum klöppeln, aber kein Stereo mehr. Wie gesagt, es stimmen Pegel nicht mehr, es stimmen Mitten nicht mehr, das Panorama stimmt nicht mehr. Wenn es dann so ist, dann braucht man die Produktion auch gar nicht mehr abspielen, weil es dann niemand verstehen wird, weil es nicht so klingt, wie es gedacht war.

DS: Und in solchen Momenten, wünschst du dir dann manchmal einen Objekt-Renderer?

BM: Definitiv. Klar, wenn es dann einen Objekt-Renderer gäbe, der die Anordnung der Boxen und den Raum kennt, wird natürlich alles ausgeglichen und sinnvoller angeordnet in dem Moment. Das sorgt dafür, dass inhaltlich das präsentiert wird, wie es präsentiert werden sollte. Das wäre natürlich schon schön, definitiv.

DS: Und würdest du als Kreateur von immersiver Musik sagen, dass dies deinen Job erleichtert?

BM: Wenn man objektbasiert arbeiten würde?

DS: Ja.

BM: Weiß ich nicht. Ich mag am kanalbasierten, dass ein ganz klares Format ist. Das ist super einfach, du stellst deine Boxen auf im richtigen Winkel und in der richtigen Positionierung. Mann kann Dinge breiter und schmaler machen, weil es halt einen anderen ästhetischen Effekt hat. Ich verstehe am objektbasierten natürlich, dass es in einen Raum gerendert werden kann, exakt so, wie man es gehört hat. Das wäre schon praktisch, ja. Da habe ich aber zu wenig Erfahrung. Ich habe noch nicht eine objektbasierte Produktion gemacht, worüber ich dir Auskunft geben könnte, ob ich es jetzt besser oder schlechter fände und ob es sich überträgt. Ich habe auch noch keine objektbasierte Produktion auf einem Medium gehört und dann auf einem

anderen Medium gehört, das anders angeordnet ist und in einem anderen Raum. Ich habe keinen Vergleichs- oder Wissenswert dazu.

DS: Funktionieren denn Binauralisierungen für dich?

BS: Das kommt sehr drauf an. Ich habe mit Daniel [Schiffner] auch schon rumprobiert mit Popmusik. Ich hatte mal Binaural-Sachen mit VR gehört, da fand ich teilweise Sachen wahnsinnig gut. Wobei da auch die visuelle Komponente dazukommt, die mich in eine gewisse Richtung zieht und den akustischen Moment besser macht. Ich habe ganz oft das psychologische Problem, wenn ich Kopfhörer aufhabe, dann spüre ich die Kopfhörer am Ohr. Und wenn ich Kopfhörer aufhabe, dann wird mir schon impliziert, dass bestimmte Sachen so stattfinden, wie sie stattfinden sollten. Dass ich beispielsweise eine bestimmte Akustik höre aber in einer U-Bahn sitze, sodass es sich nicht richtig ergänzt, weil die visuelle Komponente zu stark ist, dass es manchmal nicht immersiv ist und die Raumanordnungen nicht stimmen, sodass die Phase komisch klingt. Wie gesagt, mit VR fand ich das so eisenhart, dass ich wirklich dachte: Was ist hier los, wo bin ich? Ich fand das gut bei Sound Design, ich fand das gut bei Videospielen, bei Film ohnehin. Bei Popmusik ohne eine visuelle Komponente habe ich noch keine einzige Version gehört, die ich total geil fand. Aber auch, weil ich bestimmte Dinge schwierig finde, und weil ich den aktuellen ästhetischen Umgang mit Binauralisierung nicht richtig finde. Da wird sehr einseitig draufgeguckt und dann ist es so, da wird in der Popmusik gesagt, dass wir alles räumlich machen. Aber nehmen bestimmte Dinge weg, die in der Kopfhörer-Mischung auch total geil sind: In-Kopf-Wahrnehmung! Warum soll ich denn das killen wollen? Man muss Elemente aufbauen. Ich hab noch keine einzige Produktion gehört, wo Leute Dinge drum herum gebaut haben. Und bei den Sachen, die wir ausprobiert haben, da war immer das Problem: Wenn ich ein Mehrkanal-Boxensystem hatte, dachte ich immer: WOW! Du hast es gespürt und da war Distanz. Und auf Kopfhörern war es klein und unschön, wenn ich keinen Bildanteil hatte, der mich hineingezogen hat. Das ist aber ein momentaner Erfahrungswert. Und, das war jetzt alles auf Popmusik bezogen, bei Sound Design fand ich es auch schon wieder deutlich spannender, weil da andere räumliche Dinge stattfinden. Aber ich kann mir binaural sehr gut vorstellen, weil es das einzige Format ist, was konsumententauglich ist. Aber kein Mensch wird sich ein 9.1-System hinstellen und auch kein objektbasiertes System.

DS: Stichwort Immersion: Du hast eben schon die Ästhetik angesprochen, die du präferierst. Von welchem Parameter welcher Art auch immer hängt der Immersionsgrad deiner Meinung nach ab?

BS: Ich würde sagen: Panorama und räumlicher Eindruck. Und mit räumlichem Eindruck meine ich nicht Natürlichkeit darzustellender Räume. Das ist Popmusik, da geht es nur um ein ästhetisches Mittel, um bestimmte Sachen zu präsentieren. Aber sobald man das Gefühl hat, man sitzt mittendrin und ist umstellt von bestimmten Dingen, die sich schön anfühlen, auch synthetisiert natürlich, was ja auch sehr schön sein kann, dann kann man einen hohen Immersionsgrad bewirken.

DS: Und würdest du sagen, dass die heutige technische Infrastruktur, so wie du sie genutzt hast, deine Anforderungen erfüllen?

BS: Ja, aber ich habe es auch nur so popelig wie möglich genutzt. Ich hab's ja so genutzt, wie man es vor 30 Jahren schon hätte nutzen können. Ich hab ja keinen Mittelsmann zwischendrin, keinen Renderer, ich hab teilweise nicht mal einen Panner benutzt, sondern manchmal habe ich einfach hart auf die Boxen geroutet, das hat auch funktioniert. Deswegen ist mein Zeug formattechnisch so das rudimentärste, was es gibt. Weil es auch einfach ist, schnell geht und super gut umsetzbar ist, also schnell von einer Stereo-Produktion auf 9.1 umzusetzen, mit so rudimentären Sachen, das ist nicht sonderlich schwer. Da kann man sich ganz schnell auf die inhaltlichen Sachen konzentrieren und nicht auf die technischen Sachen drumherum. Und bestimmte Dinge funktionieren halt einfach nicht gut. Ich habe keine Lust, einen 9.1 Panner zu benutzen, der einfach wahnsinnig schlecht konzipiert ist. Da habe ich nichts davon, wenn es mich davon abhält, Musik zu machen.

DS: Betrachten wir mal folgendes Szenario: Es gibt standardmäßig einen Renderer, der ein bestimmtes Metadatenformat auslesen kann. Wäre das für dich ein Anreiz für dich, mit Objekten größeren Metadatenformaten zu arbeiten?

BS: Klar, wenn es gut integriert ist und Spaß macht und nicht im Weg steht, dann ist das top. Im Endeffekt würde es dann besser zum Endkonsumenten getragen, weil dann bestimmte Metadaten im File drin sind, die es dann wahrscheinlich so wiedergeben, dass es dem System angepasst ist. Bei mir kommt da immer eine finanzielle Komponente hinzu. Wenn es gescheit integriert ist und kein Technik-Krampf, dann

finde ich das alles toll. Ich bin das beste Beispiel für einen Anwender, der einfach mit Technik arbeiten möchte, ich will Arbeitstiere und keine Technikschlachten, wo ich denke, ich muss erst mal 10 Tage Kabel stecken oder Plug-Ins aufmachen, damit ich endlich mal an den Punkt komme, um Musik zu machen. Aber grundsätzlich finde ich so was alles spannend, also objektbasierte Sachen, wenn die mit den richtigen Metadaten im richtigen Containerformat übermittelt werden, dass man den Inhalt, den man produziert hat, besser und passender zum Endkonsumenten bringen kann: Super! Das ist ja das beste, was es gibt. Das ist ja das, was man sich wünscht. Dass jeder, der die Musik hört, so hört, wie man sie selber gehört hat. Das ist ja das Ziel eigentlich davon.

DS: Ok, Bene. Vielen Dank!

BS: Sehr gerne!

Transkription - Interview I4

Interview vom 03.01. 2018

Interviewter: Andreas Mühlschlegel

Abkürzungen: AM = Andreas Mühlschlegel; DS = Daniel Strübig

DS: Hallo Andreas.

AM: Hallo.

DS: Magst Du Dich kurz vorstellen? Wer du bist, was du machst und in welchen Gebieten du mit immersivem Audio arbeitest.

AM: Ja gerne. Mein Name ist Andreas Mühlschlegel und ich bin Sounddesigner, Mischtonmeister und Audio-Kreativer. Du möchtest nun also wissen, was ich so im immersiven, interaktiven Bereich schon gemacht hab.

DS: Zunächst bitte im immersiven Bereich. Also wie arbeitest Du mit immersivem Audio, die in erster Linie nicht interaktiv sind.

AM: Mehrkanaliges Audio oder binaurale Inhalte mache ich schon eine Weile. Das Immersive beginnt schon für mich bei einer Kino-Mischung, mit 5.1, also ganz „basic“. Schon hier kreiert man Immersion, also das klassische Channel-based Setup. „Von Dolbys Gnaden“. Dies muss man zunächst meistern und dann kann der gesamte Rest auch in Angriff genommen werden. Dort geht es für mich mit dem Immersiven los, man kreiert etwas, in das die Leute dann eintauchen.

DS: Also arbeitest du vorwiegend in Filmen mit Bewegtbildinhalten?

AM: Ja, so kann man es gut zusammenfassen. Stimmt.

DS: Gibt es auch Bereiche oder Projekte, in denen du nicht mit Bewegtbildinhalten gearbeitet hast, sondern mit Klangsphären Klangszenen geschaffen hast?

AM: Ja, das kommt immer wieder vor. Vor allem Bereich von Installationen, Events oder irgendwelcher Spezial-Jobs, so im Bereich Automotive. Ganz kürzlich hatten wir ein Projekt für ein Concept-Car, das wir im Jahr 2016 auch schon gemacht hatten. Dies hatte letztendlich einen integrativen Teil, einen mehrkanaligen Mischungsteil und natürlich einen Design-Teil. Dabei stellten wir uns die Frage, was wollen wir überhaupt machen, welche Sounds funktionieren hier gut in einem immersiv bzw. 3D-Kontext in einem Auto, der jetzt nicht speziell bildbezogen ist,

sondern in dem Audio für sich steht. Letztendlich ging es um ein Human-Interface, Machine-Interface, das HMI im Auto unterstützt, beispielsweise Click-Sounds. Auch beispielsweise ein Abfahren eines THX-Logos anstatt eines Anspringen des Motors beim Starten des Autos (lacht).

DS: In diesem Bereich geht es dann ja auch schon in die Interaktion rein.

AM: Genau, hier ist es dann schon interaktiv.

DS: Nehmen wir mal das Concept-Car-Beispiel. Mit welchen verschiedenen Wiedergabetypen hast du daran gearbeitet? Habt ihr da mit Objekt-Rendering gearbeitet und beispielsweise szenenbasiertem Audio?

AM: Es war eine Mischung aus kanalbasiert und szenenbasiert. Dabei handelt es sich um einen wirklich interessanten Spezialfall. Und zwar handelt es sich dabei um ein absolut proprietäres Custom-Setup bei einem Auto. Oft ist es so, dass keine ideale Speaker-Verortung möglich ist, aufgrund des begrenzten Platzes. Das heißt die Speaker sind da, wo sie gut verbaut werden können, aufgrund ihrer eigenen Dimensionierung. Hier gibt es eben auch Grenzen. Zusätzlich muss es schön aussehen und in das Design des Interieurs, des Innenraums hineinpassen. Dies sind natürlich schwierige Bedingungen, im Unterschied zum Studio, wo ich eben bestimmen kann, wo ich meine Boxen hinstelle, wo sie idealerweise stehen könnten und baue dann den Raum dazu. Dabei handelt es sich eher um Laborbedingungen im Endeffekt. Im Auto ist dies eben nicht der Fall. Hier hat man die Möglichkeit zu schauen, wie das Audio integriert werden könnte. Das heißt, dass man dies häufig channelbasiert machen muss und teilweise objektbasiert. Letzteres hab ich bisher aber noch nicht ausprobiert und kann dazu nicht viel sagen. Das war letztendlich einfach eine Wahl der Mittel, weil es einfach geht mit dem szenen- und kanalbasiert. Dies ist eben einfacher zu handeln in diesem Kontext. Zur Fragen, warum wir das so gemacht haben: man hat einen Speaker über sich, kann weitere als Ring zusammenfassen und einen Shaker im Sitz drin. Dann gibt es noch ein LFE extra. Dahinten sind dann auch nochmal zwei Speaker. Jedoch ist dann beispielsweise ein Sound nur an einer Stelle, da dieser personalisiert ist. Da-

bei gibt es viele verschiedene Szenarien bzw. Anwendungsszenarien, die dabei eine Rolle spielen. Dann muss man relativ schnell hin- und her-switchen und packt dies eben mal hier und mal dort rein. Dabei haben wir einen szenenbasierten Ambisonics Decoder verwendet und dies zudem teilweise überschneidend. Teilweise die ganze Anlage als Ambisonics decoded und den Ring, der oben in der Decke war, dann als kanalbasierte Stereo-Outputs. Es gab noch einen weiteren Ring in der Decke mit Speakern direkt über den Köpfen. Den konnten wir gleichzeitig stereo anfahren, also Stereo-Inhalte draufschieben, oder eben als Ring für den Ambisonics Hall-Return benutzen. Also verschiedene, flexible Ansätze, die dann eben alle in einer Mixing-Session funktionieren, also bedienbar sind und eine unglaubliche Flexibilität erlauben. Deswegen verwendeten wir diesen Mix aus kanal- und szenenbasiert. Dadurch, dass es in einem Auto nicht funktioniert, wegen der schlechten Bedingungen, was das Sweet-Spotting angeht, ist es eh ganz gut mit dem Ambisonics, da dieses eben etwas verschmierter ist. Das fällt dann gar nicht so auf. Deswegen haben wir hier eine ganz gute Wahl getroffen, vor allem für so einen Hall-Effekt-Return. Das klingt einfach „schön“. Es geht gar nicht so sehr um perfekte Klanglokalisation. Das war gar nicht so das Ziel, weil wir von vornherein wussten, dass dies eh nichts wird. Es musste immersiv, beeindruckend klingen, aber nicht irgendwie „wie das Vögelchen ein Grad weitergeflogen ist“.

DS: War es bei dem Projekt so, dass das Setup von vornherein stand und ihr habt inhaltlich dann darauf aufgebaut? Und wie bist du da vorgegangen? Zum Beispiel aufgrund des Umstands, dass du ein Ambisonics-Bed und punktuell kanalbasierte Ausspielungen hattest. Wie bist du auf der inhaltlichen, also auf der Design-Ebene vorgegangen? Ich meine sozusagen „form follows function“.

AM: Letztendlich trifft es „form follows function“ schon ganz gut. Das Ding war eben schon fertig designt und da war eben auch schon die Speaker-Verortung vorgenommen.

DS: Die Codierung auch?

AM: Nein. Das haben wir uns dann ausgedacht. Da hatten wir eben diese und jene Speaker zur Verfügung und diese sind auch einzeln ansteuerbar. Da ging es immer so ein bisschen hin und her. Die

waren manchmal gedoppelt, also hintereinander geschaltet, sodass wir die nicht einzeln anfahren konnten. Manchmal hieß es auch, dass jeder einzelne Subwoofer ansteuerbar war. Es handelte sich eben um ein Concept-Car, also keine fertige Entwicklung. Im Prinzip war es ein Haufen Technik zusammengebastelt, die irgendwas zeigen sollte. Kein Auto, das wirklich fährt. Um zur Vorgehensweise zurückzukommen: Diese hat sich mehrfach geändert. Da muss man eine große Flexibilität an den Tag legen, die mehrere Änderungen erlaubt. Wir haben uns zunächst theoretisch überlegt, was wir da machen könnten. Dann haben wir überlegt, welches Output-Rendering wir benötigen. Daraufhin haben wir uns eine Session aufgebaut, in die wir dann reinarbeiten konnten, die uns gleichzeitig immer erlaubt hat, flexibel zu bleiben und Änderungen vorzunehmen. Nichts war so fest installiert, dass man es nicht hätte verändern können. Wir haben von vornherein versucht, alles durchzudenken, was irgendwie sinnig werden könnte. Zum Beispiel war auch binaural head tracking geplant, was dann aber gestrichen wurde. Dies hätten wir auch machen können.

DS: Wie sehr hat Dich dann das vorgegebene Setup im weiteren (Kreations-)Prozess beeinflusst? Du hast für dieses eine Zielmedium designt, oder?

AM: Ja das stimmt. Es war relativ früh klar, dass wir diese Stereo-Outputs haben werden und hier dieses kleine 5.0-Setup. Unabhängig davon wussten wir schon, dass wir auch hier ein bisschen was machen können. Wir werden irgendwelche Kreisbewegungen haben, inhaltlich, um Menschen da im Prinzip reinzuholen. Wir werden Elemente brauchen, die sich gut gleichmäßig um Einen herumbewegen könnten. Damit erzeugt man Immersion. Es ging darum, dies zu demonstrieren. Es sollte nicht gezeigt werden, welche tolle Anlage im Auto verbaut ist und was die alles Tolles kann. Das beeinflusst einen bei der Kreation.

DS: Es sollte also für genau den Zweck funktionieren.

AM: Von diesem Standpunkt sind wir im Prinzip ausgegangen und haben uns dann überlegt, wie wir das umsetzen können und was sich in der Situation mit der Anzahl der Speaker und deren einzelner Ansteuerung an. Wir haben also schon eher vom Inhalt aus gedacht.

DS: Wir entfernen uns mal etwas von diesem Projekt. Magst du kurz mal umreißen, wie du mit interaktiven Elementen arbeitest? Also, ob du beispielsweise OSC benutzt oder wie du Interaktion einbindest.

AM: Das kann für mich relevant sein im kreativen Sounddesign-Prozess. Das kann beispielsweise OSC sein, aber auch altes Midi sein. Da geht es für mich eher darum, ob ich irgendwelche Controller an eine Kiste anschließen kann, die irgendwelche Sounds macht. Das ist für mich auch Interaktion. Ich erweitere das mal, da viele Leute das eher vom Content aus betrachten, der dann irgendwann fertig ist und es sich dann um interaktiven Content handelt. Aber auch das hat schon mit Interaktivität zu tun, also eine Performance zu machen mit irgendwelchen Controllern, die über irgendwelche Protokolle miteinander kommunizieren. Dadurch schafft man sich irgendwelche Instrumente, die es so eigentlich gar nicht gibt und verschiedene interessante Klänge dadurch machen. Speziell benutze ich da dieses Monome, das so ein bisschen ein Schattendasein fristet, aber ein interessantes Tool darstellt. Zusätzlich benutze ich aber auch Controller auf iPad-Basis, OSC, Lemur etc. Ich habe diese schon mehrfach angewendet. Ich hab zum Beispiel Touch-OSC, da es so einfach und simpel ist, für ein Theaterstück verwendet.

DS: Da war es aber dann nutzerseitige Interaktion, oder?

AM: Ja, genau. Wenn du jetzt nach interaktiven Elementen fragst, dann fragst du natürlich auch nach einem Endformat, das quasi interaktiv ist, was landläufig auch unter „Computerspiel“ läuft. Oder meinst du eine integrative Interaktion von Inhalten?

DS: Eher Zweiteres. Beispielsweise meinte Felipe Sanchez mal, dass sie mit Tracking gearbeitet haben, dessen Soundquelle, dessen Sound-Emitter dem Benutzer folgt. Hattest du damit auch schon Berührungspunkte?

AM: Nein, solche Sachen hab ich jetzt noch nicht machen dürfen. Was ich sagen kann ist, dass ich natürlich Sound gemacht hab für Virtual Reality-Sachen, sprich also wo User-Input eine große Rolle spielt. Das war einmal ein VR-Film, der dann auch in einer Version, was auch die bestklingendste war, was auf diesem Barco-IOSONO-System abspielbar war, also mit HTC-Vive-Brille. Dann al-

so quasi stehend in einem Boxen-Setup, Walking-VR in 5x5 Meter eingekreist von einem Haufen Speaker, es lief alles über Barco-IOSONO-Core-Renderer. Das Interaktive daran war letztendlich das ganze in Wwise zu bauen und dann einmal zu Barco nach Erfurt zu fahren und dann in deren Custom-Mixer-Lab zu mischen. Dies klang wirklich toll. Das ist eine ganz klar interaktive Geschichte. Audiokinetic Wwise ist das Ding, woran ich dran kleben geblieben bin. Was sich immer als sehr „handy“ erwiesen hat. Das gilt auch für das andere Projekt, die Installation von zehn verschiedenen Präsentationsstationen in 5.1, jeweils jede Station, die synchronisiert miteinander laufen mussten. Der Programmierer hat dabei mit einem Max-Patch hinbekommen, diese 13 Instanzen über Unity mit Wwise integriert, also quasi eine Audio-App. Meine Aufgabe war dabei, diese ganze Musik, die interaktiv aufeinander aufgebaut war (eine Akkord-Abfolge) zueinander anzupassen. Und trotzdem sollte es verschiedene Teile davon haben und am Ende lief alles synchron, also im Takt miteinander. Man konnte trotzdem wie in einem Menü verschiedene Phasen einer Präsentation durchlaufen, die dann jeweils immer von der richtigen Musik untermalt wurde. Das haben wir alles in Wwise gemacht.

DS: Lass uns nun mal in die Richtung vom finalen Datenformat gehen und nochmal auf das Concept-Car zu sprechen kommen. Magst du exemplarisch erläutern, wie die finale Abgabedatei aussah? Also aus was sie bestand und ob sie Metadaten enthielt.

AM: In dem Fall nicht, weil wir nur das quasi nur channelbased und scenebased gemacht haben. Letztendlich war dies ein 24-Kanal-Wave-File mit einer völlig frei gewählten, proprietären Kanalbelegung. Da haben wir gesagt, dass wir von 1 bis x das so gemacht haben und bei y überspringen wir einen Kanal und machen dann bei z weiter. Das war eine völlig frei gewählte Angelegenheit. Das lag daran, dass wir für das Concept-Car mit der Software „AudioMulch“ gearbeitet haben. Ich weiß nicht, ob dir das was sagt.

DS: Ja, das sagt mir was.

AM: Das ist eine Patching Software, so wie Bidule. Das ist wie ein vereinfachtes Max MSP, wo man nicht so tief in die Program-

mierung gehen kann, aber trotzdem patchen. Das haben die da verwendet.

DS: Ok. Würdest du sagen, dass, wenn mehr Inhalte für das Concept Car produziert werden, weil diese sich vielleicht etabliert haben im Markt. Würdest du sagen, dass Metadaten da helfen würden? Die beispielsweise sagt: Diese Kanalgruppe wird in 1st Order Ambisonics codiert, und die geht dann auf die und die Lautsprecher.

AM: Ja. Mit Sicherheit macht das Sinn. Hier kann man ja auch erkennen, was eigentlich passiert ist. Wir haben nämlich extra für diese eine Anwendung Inhalte erstellt. Es gibt ja auch kaum Inhalte, die für so ein Mehrkanalsystem funktionieren. Abgesehen davon, dass Stereo-Musik genommen wird und dann ein Upmix gemacht wird. Aber es ist halt nur ein Upmix. Also jeder weiß, dass Stereo nicht aussterben wird, weil es ein tolles Format ist und viele Musiker drauf stehen, und die dieses Mehrkanal-Thema… Naja, die Lernkurve ist jetzt nicht so steil (lacht). Die Leute müssen jedenfalls erst in das Thema reinkommen. Deswegen wird Stereo weiter überleben und Upmixe werden gemacht. Und gleichzeitig hast du dann irgendein Auto, und ein Upmix wird gemacht. Und vielleicht mal ein Klassik-Liebhaber, der seine 5.1-Aufnahme über DVD abspielt und über die Autobahn rauscht. Irgendwann müssen aber neue Formate her. Und ich weiß nicht, ob Ambisonics da technisch das richtige Format ist, zum Arbeiten in Projekten ist das toll. Aber ob es das Super-Format ist, das kann ich nicht sagen. Ich glaube, dass Stücke mit Metadaten, also object-based, das richtige ist. Mit einer Anlage, einem Renderer, und es wird so abgespielt, dass es gut klingt. Und im Auto erst recht, das ist ja eh ein Spezialfall.

DS: Also wenn du auch Produktionen am laufenden Band machen müsstest, dann müsste auf jeden Fall ein Metadatenformat her, was im Idealfall noch standardübergreifend funktioniert.

AM: Ja klar! Und ich meine, was auch immer diese Metadaten enthalten, gerade das Audio Definition Model ist ja ein relativ offenes Format. Das ist ja gar nicht schwer das zu erweitern. In Autos ist es ja naheliegend, dass man diese Bass-Shaker in den Sitzen hat. Da ist es nicht immer sinnvoll, einfach den Bass

raufzuschieben, weil das kann rumpeln, das klingt nicht gut. Viel schöner ist es doch, wenn man einen einzelnen Synthie hat, den man in den Bass-Shaker rein gibt. Dann brummt es wie ein Kätzchen (lacht). Dann müsste man sich überlegen: Eigentlich braucht der Bass-Shaker einen extra Kanal. Und da müsse es Metadaten geben, die dem System genau das mitteilen. Man kann natürlich sagen, dass dieser Kanal immer die Belegung 4 hat. Aber ob das funktioniert… ich weiß nicht. Im Grunde müssten sich die Metadaten um genau das kümmern.

DS: Gerade auch, wenn wir in Richtung Transkompatibilität denken: Deine Mischung, also deinen Master-File mit 24 Kanälen könntest du nicht in anderen Autos wiederverwenden.

AM: Nein. Wenn du den gerenderten Output, nicht die Original-Mischdaten meinst, dann nicht. Ich könnte natürlich meine Session nehmen und die Decoder verändern, mit anderen Speaker-Koordinaten. Dann würde es gehen. Klar könnte ich eine Liste mit Speaker-Koordinaten anfordern und die Decoder ändern und nochmals bouncen. Das würde ich aber niemals machen. Ich würde immer noch vor Ort mischen. Generell würde ich bei diesen Car-Sachen sagen: Egal ob kanal-, objekt-, oder szenenbasiert, ich wäre immer vorsichtig. Das sind alles Spezialanwendungen, noch. Und es war vor allem ein Concept Car. Will heißen, noch kein normaler Auto-Innenraum. Zum Beispiel eine Fahrerkabine, wo eigentlich kein Lenkrad vorgesehen ist, die also mit einer herkömmlichen Fahrerkabine nichts zu tun hat. Man sitzt sich im Grunde wie im Zug in einem Viererabteil gegenüber. Wo ist denn da jetzt die Vergleichbarkeit mit anderen Autos? Da müsste ich jetzt wieder her gehen und eigentlich einen Standard implementieren. Sogar Dolby hat ja im Grunde ein festgelegtes Boxen-Setup. Das ist zwar offiziell object-based, aber die gehen trotzdem von einer festgelegten Lautsprecherpositions-Konfigurierung aus.

DS: Dolby gibt also eine Empfehlung.

AM: Ja, da halten sich aber auch alle dran. Sonst würde Dolby auch nicht mitspielen.

DS: Oder die Rendering-Unit würde irgendwann ausflippen.

AM: Genau, oder Dolby sagt, dass man das so nicht machen kann. Die kommen sogar persönlich vorbei mit einem Consultant, wenn du

als Kino Dolby-Inhalte präsentierst oder mit einem Misch-Kino Dolby-Inhalte mischen willst. Ich muss eine Lizenz haben, abgesehen von der RMU.

DS: Und im Filmbereich machst du dann auch Downmixes aller Art.

AM: Genau, sicherlich. Aber ich gehe eher immer vom Großen ins Kleine. Ich gehe selten vom einen großen ins andere große. Das ist aber eigentlich eine wichtige Frage, auch wenn ich so drüber nachdenke, kann das schon interessant werden. Kann ich von meiner 7.1 Mischung jetzt auch in ein anderes Großformat? Gut, das Auro 3D ist relativ kompatibel zu 7.1. Aber kann ich beispielsweise mein Dolby 7.1 Bed mit Objects in ein Auro 3D hinüberreichen? Habe ich noch nie gemacht, ich kann mir aber vorstellen, dass es möglich ist.

DS: Würdest du denn als Kreativer besser arbeiten können, wenn du dir darüber keine Gedanken machen müsstest?

AM: Ja, sicherlich. Also du meinst, dass ich einfach meine Inhalte mache, die in eine Box packe und sage: Die sollen räumlich so und so gestaltet sein. Und dann entscheidet eine Technik oder ein Renderer darüber, was damit anzufangen ist, und wie der das abbildet. Ja natürlich ist das eine Befreiung. Denn es ist natürlich sehr viel Planung, und das ist es, was mich davon abhält, einfach mal Ideen reinzuwerfen. Eine super Idee, von der ich dann aber sofort abweiche, weil sie im Stereo nicht funktionieren würde. Man schränkt sich ja auch ein, wenn man auf einmal Stereo-Kompatibilität herstellen muss. Eigentlich will man ganz viel ausprobieren, aber man kommt halt ganz schnell zu dem Punkt wo man denkt: "Haargh, das gibt bestimmt irgendwelche Phasen-Probleme, das lassen wir lieber." Das ist ein totales Hindernis. Ich habe aber auch meine Zweifel, ob so ein transkompatibles Format, was am Ende alles erschlägt, ob das dann auch so etwas bieten würde. Das müsste man mir erst mal beweisen. Ob das dann auch wirklich funktioniert. Wenn ich beispielsweise das Audio Definition Model nutzen würde, aber trotzdem in Ambisonics-Decoding arbeiten würde, dann beeinflusst das ja auch wieder meine Arbeit. Und wenn jemand anderes das jetzt transferieren will auf eine andere Anlage, dann ist es ja immer noch der Ambisonics-Mix, und nicht ein object-based Mix vom Dolby Atmos. Das

ADM nimmt ja bloß all diese Sachen und schmeißt die in eine Datei rein, dann gibt es noch einen XML-File oder das ist im Header.

DS: Genau, die sind im Header vom WAV-File. Und da kannst du beispielsweise definieren: Kanäle 1-4 sind 1st Order Ambisonics, 5-8 sind kanalbasierte Ausspielungen und in 9 und 10 sind dann noch Objekte. Und die sollen Chips dann auch auslesen können, diese Formate. Das wäre für mich noch ein wichtiger Punkt, die Auslese der Chips.

AM: Ja, das stimmt. Ich will auch um Gottes Willen den Ansatz vom ADM nicht schlechtreden, ich finde das eigentlich total gut, dass es das überhaupt gibt, der versucht, das dann auch durchzusetzen. Das ist auf jeden Fall der richtige Ansatz. Weil es ja natürlich nervig ist, gerade für mich, dass jeder Hersteller mit seiner Super-Lösung ankommt, wie das denn jetzt zu machen sei, aber das funktioniert dann auch nur in diesem einen Ökosystem. Aber kombinieren kann ich diese Ökosysteme nicht. Ich kann nicht ein tolles Feature vom einen Format und ein tolles Feature vom anderen Format nehmen und in einen Topf schmeißen. Gerade im kreativen und immersiven Kontext finde ich Ambisonics mega interessant. Da gibt es tolle Elemente, die einfach zu bedienen und schnell sind. Deswegen würde ich das auch echt gerne benutzen. Es gibt da ja auch ein paar Ansätze, also kleine Öffnungen von Dolby zum Thema Ambisonics hin, dass man zumindest 1st Order Ambisonics Mischungen mit integrieren kann. Wo ich mir aber dann denke, warum nicht gleich dritte Ordnung? Ist doch nicht so schlimm, tut doch niemandem weh. Und es klingt ja auch gut. Dann wäre zumindest dazu die Brücke auch irgendwie geschlagen mit der Transkompatibilität. Die Hersteller sagen ja immer: Nimm aber bitte mein Format! Und so wäre es natürlich toll, wenn Dolby sich dem ADM anschließen würde, und auch scene-based Ambisonics unterstützen würde. Das müssten sie wahrscheinlich auch irgendwann.

DS: Und wenn wir jetzt im Abschluss nochmals konsumentenseitig denken: Glaubst du, dass ein standardübergreifendes Metadatenformat ein größeres Immersionserlebnis in sich birgt?

AM: Das glaube ich schon. Wenn man sich zum Beispiel die ganze VR-Welle anschaut, was da mit Audio so los ist. Das ist ein riesiges Durcheinander, es gibt zig verschiedene Renderer, die dies oder jenes machen, manchmal schlecht manchmal gut. Es gibt Facebook mit dem proprietären quasi De-facto Ambisonics. Dann gibt es verschiedene Audio-Engines, die dann da verwendet werden. Das ist ein totales Durcheinander. Aber man hat sich da zumindest auf Ambisonics geeinigt. Das muss aber nicht so bleiben. Die andere Sache ist das mit den HRTFs. Es gibt Leute, bei denen funktionieren die generischen total gut, bei manchen halt nicht. Aber letztendlich, wenn es um de Kompatibilität ginge, könnte es zumindest das Hörerlebnis bei allen Menschen auf die gleiche Erfahrungsebene bringen. Weil nicht bei jedem die Mischung in der bestmöglichen Qualität abspielseitig ankommt! Ich kann auch in meinem Studio in 5ter Ordnung Ambisonics mischen und mich ganz toll freuen, Und dann wird es wieder zu erster Ordnung heruntergefuhrwerkt mit irgendeiner HRTF, am besten noch mit einer schlechten HRTF durchgenudelt (lacht). Oder es landet Stereo auf irgendwelchen Speakern und wird dann mit Dolby Pro Logic mit irgendwelchen Speakern aufgeblasen. Das klingt absurd, ist aber mehr als möglich, wenn man sich mal überlegt, was alles damit passieren kann! Das ist natürlich ein Kampf, den man nie verlieren kann, aber so ein Audio Definition Model ist da ja natürlich ein hilfreicher Anfang.

DS: Ok, vielen Dank Andy!

AM: Gerne!