

“Der hörende Tourist: Eine Expedition in immersive Reisemedien mit MPEG-H-Integration”

Masterarbeit zur
Erlangung des
akademischen Grades

Master of Arts

Communication, Media, Sound and Interaction Design, FH JOANNEUM

Vorgelegt von - Florian Schaubmaier BA
Thesis Tutor - DI. Astrid Drechsler
23.01.2024, Graz



FH | JOANNEUM
University of Applied Sciences

**"The aural traveler: A journey into immersive
travel media with MPEG-H integration"**



**"Der hörende Tourist: Eine Expedition in
immersive Reisemedien mit MPEG-H-
Integration"**



Danksagung

Ich danke allen, die mich in dieser Zeit unterstützt haben und besonders Madeleine.

Weiters möchte ich mich gerne bei der UPV Valencia für den Aufenthalt und beim Fraunhofer IIS sowie dem IEM, für die hilfreiche Auskunft bedanken.



Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt,

- dass ich die vorliegende Bachelorarbeit/Masterarbeit selbstständig angefertigt und die mit ihr verbundenen Tätigkeiten selbst erbracht habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe;
- dass ich mich bei der Erstellung der Arbeit an die Richtlinie der FH JOANNEUM zur Sicherung Guter Wissenschaftlicher Praxis und zur Vermeidung von Fehlverhalten (kurz Richtlinie GWP) gehalten habe;
- dass ich alle aus gedruckten oder ungedruckten Werken sowie aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Regeln für gutes wissenschaftliches Arbeiten (Richtlinie GWP) zitiert und durch genaue Quellenangaben gekennzeichnet habe;
- dass ich in der Methodendarstellung oder einem Verzeichnis alle verwendeten Hilfsmittel (Assistenzsysteme der Künstlichen Intelligenz wie Chatbots [z.B. ChatGPT], Übersetzungsapplikationen [z.B. DeepL], Paraphrasierapplikationen [z.B. Quill bot]), Bildgeneratorapplikationen [z.B. Dall-E] oder Programmierapplikationen [z.B. Github Copilot] deklariert und ihre Verwendung bei den entsprechenden Textstellen angegeben habe;
- dass die vorliegende Originalarbeit in dieser Form zur Erreichung eines akademischen Grades noch keiner anderen Hochschule vorgelegt worden ist.

Ich wurde darüber aufgeklärt, dass meine Arbeit auf Plagiate und auf Drittautor:innen-schaft menschlichen (Ghostwriting) oder technischen Ursprungs (Assistenzsysteme der künstlichen Intelligenz) überprüft werden kann.

Ich bin mir darüber im Klaren, dass eine wahrheitswidrige Erklärung rechtliche Folgen wie eine negative Beurteilung meiner Arbeit, die nachträgliche Aberkennung des dadurch erlangten Titels und Strafverfolgung nach sich ziehen kann.

Abstract

This thesis investigates and implements the Next-Generation Audio (NGA) standard MPEG-H 3D audio in unconventional media formats beyond broadcasting and streaming. Developed by the *MPEG Group* in collaboration with *Fraunhofer IIS*, MPEG-H aims to provide end users with a barrier-free and personalized experience while streaming and broadcasting. Dedicated tools are used to add metadata to audio content, allowing end users to customize the audio mix.

The relevance of NGA in a historical context is analyzed, the extended applications of MPEG-H are discussed, and a suitable production workflow is developed.

The linked work focuses on the integration of MPEG-H 3D audio in travel media, specifically using a customized audio tour of the city of Valencia to provide users with an immersive experience and thus enable a better understanding of their travel destination.

Diese Arbeit untersucht und implementiert den Next-Generation-Audio (NGA) Standard MPEG-H 3D-Audio in unkonventionellen Medienformaten abseits von Broadcasting und Streaming. Entwickelt von der *MPEG-Gruppe* in Zusammenarbeit mit dem *Fraunhofer IIS* strebt MPEG-H an, Endnutzer*innen ein barrierefreies und individuelles Erlebnis bei der Wiedergabe von Streaming- und Broadcasting-Inhalten zu bieten. Durch dedizierte Tools werden Audioinhalte mit Metadaten versehen, was Endnutzer*innen ermöglicht, die Audiomischung individuell anzupassen.

Es wird die Relevanz von NGA im historischen Kontext untersucht, auf die erweiterten Anwendungen von MPEG-H eingegangen und ein geeigneter Produktionsworkflow entwickelt.

Das verknüpfte Werkstück konzentriert sich auf die Integration von MPEG-H 3D-Audio in Reisemedien, konkret anhand einer individualisierten Audiotour durch die Stadt Valencia, um Nutzer*innen ein immersives Erlebnis zu bieten und ihnen so ein besseres Verständnis für ihre Reisedestination zu ermöglichen.

Inhaltsverzeichnis

1 Einleitung.....	1
<i>Hintergrund, Motivation und Ziele.....</i>	<i>2</i>
<i>Umfang, Forschungsfrage und Beschränkungen.....</i>	<i>3</i>
2. Theoretischer Hintergrund	4
2.1 <i>Entwicklung der Audiotechnologien und deren Einfluss auf das Hörverhalten der Anwender*innen</i>	5
Schallplatten.....	6
Radios.....	6
Tonfilm	7
Tonband.....	7
Stereo	8
Quadrophonie.....	9
Ambisonics.....	9
Surround-Sound	10
CD.....	11
DAW.....	11
MP3	12
Streaming.....	12
ITU-R BS.1770	13
Objektbasierte Audiosysteme.....	13
KI-gestützte Systeme	14
2.3 <i>Grundsätze von Spatial Audio</i>	<i>15</i>
Head related transfer Functions (HRTF).....	15
Räumliche Faktoren.....	16
Präzedenz, Haas- und Trading Effekt	17
2.4 <i>Spatial Audioformate und Standards.....</i>	<i>19</i>
Kanalbasiert	19
Matrixbasiert.....	20
Szenenbasiert.....	20
Binauralbasiert.....	22
Objektbasiert.....	23
Audio Definition Model (ADM).....	24

3. Theoretischer Rahmen	26
3.1 <i>MPEG-H: Merkmale und Fähigkeiten.....</i>	27
Grundfunktionen.....	27
Produktion	29
Lautheit und Dynamik	31
Rundfunk und Streaming	31
Interaktivität in MPEG-H Produktionen	33
3.2 <i>Alternative immersive Audio-Anwendungen.....</i>	34
Hörspielproduktion	34
Interaktive Medien.....	35
Videospiele	35
Automobilindustrie.....	35
Immersive Audio Experience in Reisemedien und Tourismus.....	36
4. Erstellung eines MPEG-H Audio-Teaser anhand der Beispielstadt Valencia.....	39
4.1 <i>Design der Audio-Reise durch die Stadt.....</i>	40
Sprecher	41
Standort.....	42
4.2 <i>Mikrofonierung.....</i>	45
Front Arrays	45
Decca-Tree	45
Optimized Cardioid Triangle (OCT)	46
Nachteile von drei-Mikrofon-Techniken	46
Ambience Arrays.....	47
Doppel ORTF.....	47
IRT-Cross	47
Optimized Cardioid Triangle (OCT) Surround.....	48
Tetraeder-Mikrofon.....	49
Zylia.....	50
ORTF-3D	50
Mikrofonwahl	51
4.3 <i>Produktion, Postproduktionsprozess und zusätzliche Werkzeuge</i>	56
Produktionsaufbau:	58
Spuraufbau.....	60
Encoder	60
Multiencoder.....	61
Stereoencoder	61
Monitoring	62

SoundParticles.....	62
Dekodierung.....	65
MPEG-H Authoring.....	67
4.4 Visuelles.....	70
Conclusio	73
Literaturverzeichnis.....	75
Abbildungsverzeichnis	93
Anhang.....	96



Abkürzungsverzeichnis:

ACN	-	Ambisonics Channel Order
ADM	-	Audio Definition Model
BBC	-	British Broadcast Union
BWF	-	Broadcast Wave Format
CBS	-	Columbia Broadcasting System
CD	-	Compact Disc
DRC	-	Dynamic Range Control
EBU	-	European Broadcast Union
EMI	-	Electric and Musical Industries
FOA	-	First Order Ambisonics
HOA	-	Higher Order Ambisonics
IEM	-	Institut für elektronische Musik
IRT	-	Institut für Rundfunktechnik
ITU	-	International Telecommunication Union
KI	-	Künstliche Intelligenz
LFE	-	Low Frequency Effect
LP	-	Longplay
MPEG	-	Moving Picture Experts Group
MPF	-	MPEG-H Production Format
MS	-	Millisekunden
NGA	-	Next Generation Audio
NRDC	-	National Research and Development Council
ORTF	-	Office de Radiodiffusion Télévision Française
PCM	-	Pulse-Code-Modulation
UHD	-	Ultra High Definition
XML	-	Extensible Markup Language

1 Einleitung

Diese Masterarbeit befasst sich mit der Untersuchung und Implementierung von MPEG-H 3D-Audio in dafür unkonventionelle Medienformate. MPEG-H ist ein von der *Moving Picture Entertainment Group* (MPEG) und in Zusammenarbeit mit dem *Fraunhofer IIS* entwickelter Standard, der Benutzer*innen ein barrierefreieres und individuelles Erleben von 3D-Audio- sowie Streaming- und Broadcasting-Medieninhalten bieten soll. Über einen Authoringprozess werden Audiodateien mit einer Metadatenspur versehen, die Endnutzer*innen einen individuellen Eingriff in die Audiomischung ermöglicht (Herre et al., 2015). Beispielsweise können sinnvolle Elemente, wie bei einer Fußballspiel-Übertragung die Kanäle, welche den Live-Ton aus dem Stadion übertragen, stummgeschaltet, beziehungsweise lauter und leiser gestellt werden (Simon et al., 2019). Bisweilen beschränken sich die Medieninhalte jedoch hauptsächlich auf Sport- und Musikveranstaltungen und so versucht diese Arbeit herauszufinden, welche weiteren Anwendungsmöglichkeiten sich anhand eines Modellwerkstücks offenbaren könnten. Das konkret bearbeitete Fallbeispiel konzentriert sich darauf, wie MPEG-H 3D-Audio sinnvoll in ein Medienformat der Tourismusbranche integriert werden kann

und welche Potenziale sich dadurch eröffnen. So wird ein geeigneter Produktionsworkflow mit umfassender audiorelevanter Berücksichtigung und eine individualisierbare Audiotour durch die spanische Stadt Valencia geschaffen, die beispielsweise das Angebot von Reise-Printmedien durch immersives Erleben der Stadt erweitern könnte.

Hintergrund, Motivation und Ziele

Next-Generation-Audio-Standards sind der nächste technologische Schritt in der Audioentwicklung (Pöres, 2016) und werden somit auch in einem weiten Teil der Medienformate immerzu präsenter (Internationalisierungszentrum Steiermark, 2021). Im Zuge einer Exkursion zum *Fraunhofer IIS* in Erlangen, Deutschland und dem *Schoeps Mikroforum 2022* in Karlsruhe, Deutschland wurde eine unbedingte Relevanz dieser Technologien erkannt und es entstand Neugierde, diese genauer zu untersuchen, sowie den entsprechenden Workflow zu verstehen. Etwa zeitgleich war der Ausbau und die technische Umsetzung eines Minivans samt Oldtimer-Kleinwohnwagens ein allgegenwärtiges Projekt des Autors. In Arbeitspausen im Fahrzeug kamen immer wieder alte Reiseführer zum Vorschein. Dabei fiel auf, dass die Beschreibungen und Bilder vieler Orte kaum die Erlebnisse widerspiegelten, die subjektiv empfunden wurden, als der Ort physisch besucht wurde.

Die offensichtliche Verknüpfung beider Themen ergab sich durch die, häufig in Werbetexten versprochenen, immersiven Erlebnisse von Next-Generation-Audio (*AURO-3D*, 2023; *Dolby Atmos*, 2023; Fraunhofer IIS, 2023; Zeiler, 2020).

Das, infolge des 4. Semesters absolvierte ERASMUS+ Semester in Valencia bot eine einzigartige Gelegenheit, ein 3D-

Audioerlebnis zu erstellen, um die Hypothese zu überprüfen und mehr über den Workflow und die Produktion von 3D-Audioinhalten zu erfahren.

Die Arbeit will nicht nur eine Analyse dieses spezifischen Anwendungsfalls durchführen, sondern auch technologische Hintergründe und gestalterische Aspekte betrachten. Darüber hinaus sollen praxisrelevante Leitlinien für die Integration von MPEG-H 3D Audio in unkonventionellen Anwendungen herausgearbeitet und ein möglicher Arbeitsablauf beschrieben werden.

Umfang, Forschungsfrage und Beschränkungen

Die Arbeit gliedert sich in drei Hauptabschnitte, die sich jeweils Kernthemen widmen.

Der erste Abschnitt spiegelt eine theoretische Auseinandersetzung mit dem Thema wider. Ein historischer Überblick beleuchtet die Entwicklungen der Audiotechnik seit Ende des 19. Jahrhunderts und deren Einfluss auf die Hörgewohnheiten der Bevölkerung. Darüber hinaus werden fundamentale Themen um Spatial Audio erläutert und eine Zusammenfassung von relevanten Spatial Audio Standards präsentiert.

Der zweite Abschnitt setzt sich mit der MPEG-H Technologie auseinander und betrachtet dessen Implementierung genauer. Weiters werden Überlegungen zur innovativen Implementierung von MPEG-H angestellt und der Einsatz von immersiver Audiotechnologie in Reisemedien besprochen.

Im dritten Abschnitt wird ein konkretes Fallbeispiel umgesetzt. Dabei wird ein 3D-Audio-Reiseerlebnis in der spanischen Stadt Valencia für ein MPEG-H-fähiges Wiedergabegerät konzipiert. Der Aufnahme-, Produktions- und Postproduktionsablauf wird dokumentiert und zusätzlich Distributionsmöglichkeiten in Betracht gezogen.

Die zentrale Frage dieser Arbeit ist, inwiefern die Implementierung von MPEG-H 3D-Audio in alternative Medienformate, abseits der traditionellen Broadcasting- und Streaming-Anwendungen, eine sinnvolle Erweiterung darstellt. Der Fokus der Arbeit liegt darauf, einen geeigneten Arbeitsverlauf zu entwickeln, um Medieninhalte für objektbasierte Audiosysteme wie MPEG-H zu erstellen und dessen interaktive Fähigkeiten zu testen. Das konkret bearbeitete Fallbeispiel bezieht sich auf die Integration von MPEG-H 3D-Audio in ein Medienformat für Tourismus, dessen Ziel darin besteht, eine individualisierbare Audiotour zu schaffen, die den Nutzer*innen ein informationsreiches und barrierefreies Vorab-Erlebnis der Stadt ermöglicht.

Wenn in dieser Arbeit von "MPEG-H" gesprochen wird, bezieht es sich primär auf MPEG-H 3D Audio, sofern nicht anders gekennzeichnet.

Technische Angaben werden in der Form der folgenden Beispiele beschrieben: 2 m (zwei Meter), 90° (neunzig Grad).

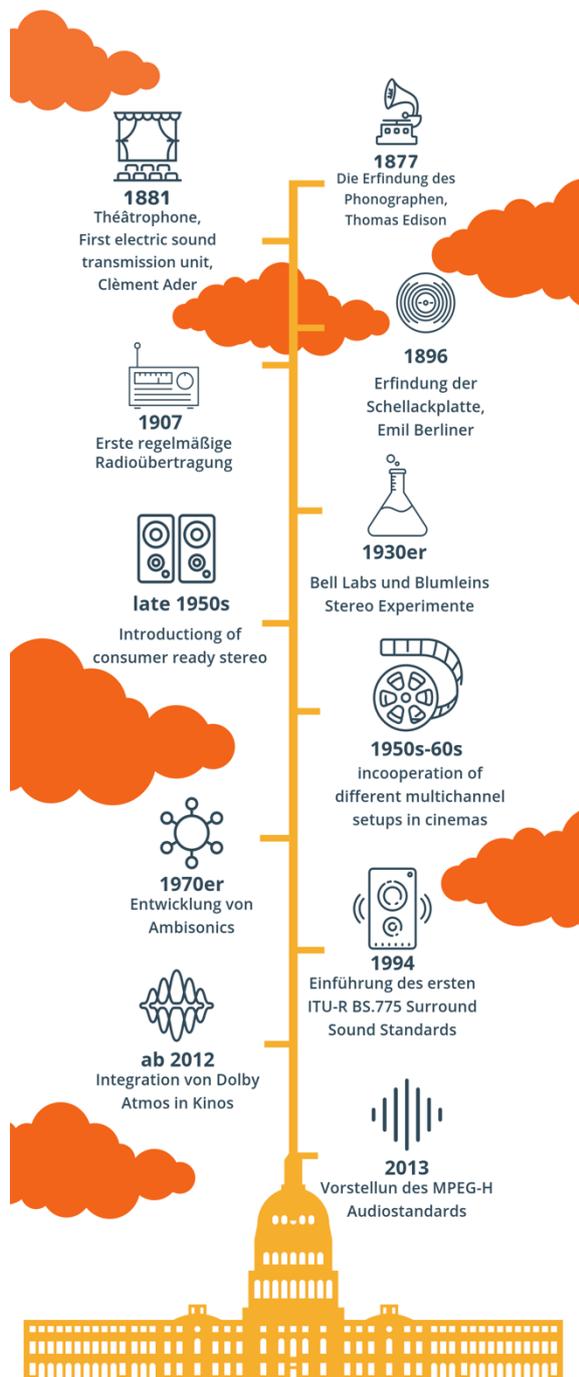
Beim Verfassen dieser Masterarbeit wurden zwei KI-Assistenzsysteme genutzt: DeepL für eine bessere Verständlichkeit zur Übersetzung von englischen Texten ins Deutsche. Und ChatGPT (Version 3.5) ausschließlich für eine optimierte Formulierung und Vereinfachung eigens verfasster Texte.

2. Theoretischer Hintergrund

"The art challenges the technology, and the technology inspires the art."

- John Lasseter

2.1 Entwicklung der Audiotechnologien und deren Einfluss auf das Hörverhalten der Anwender*innen



Seit Beginn des 19. Jahrhunderts gab es immer wieder markante Wendepunkte in der Tonwiedergabetechnik. Im folgenden Abschnitt wird eine nicht vollständige, chronologische Aufzählung von Erfindungen beschrieben, die im weitesten Sinne in Verbindung zur Thematik dieser Arbeit stehen.

Allen voran trug Alexander Graham Bell maßgeblich zur Gestaltung dieses technologischen Umfelds bei, indem er im Jahr 1867 das Telefon erfand (Bellis, 2018). Wenig später, 1877, prägte Thomas Edison die Tonwiedergabetechnik nachhaltig durch die Einführung des Phonographen (Thompson, 1995). Obwohl die grundlegenden Überlegungen zu Klängen Jahrhunderte älter sind, legten diese Erfindungen den Grundstein für eine fortschreitende Evolution der Audiotechnologie (Ampel & Uzzle, 1993).

Nach den Entwicklungen von Bell und Edison häuften sich die Innovationen um den Klang. Besonders Erfindungen wie Magnettonbänder und Schallplatten trugen dazu bei, die technische Leistungsfähigkeit und vor allem die Qualität der Audiotechnik kontinuierlich zu verbessern (Vaičiūtė, 2021).

Abbildung 1, Geschichte der Audiotechnik

Diese Innovationen ermöglichten die Aufzeichnung sowie Wiedergabe von Tönen und ganzen Musikstücken.

Schallplatten

Die Ära der Schallplatte hielt von Ende des 19. bis etwa Ende des 20. Jahrhunderts an, bis sie weitgehend von der Digitalisierung verdrängt wurde (Stephan & Selle, 2013). Dennoch behält die Schallplatte, besonders unter Sammlern und Audiophilen, bis ins 21. Jahrhundert eine gewisse Nischenstellung bei (Mrozek, 2011). Vor allem die späteren Vinyl-Schallplatten ermutigten Hörer*innen zu einer bewussten Auseinandersetzung mit der Musik und machte Plattenläden zu kulturellen Zentren (Ananda, 2023). Die Einführung der *Longplay* (LP) veränderte die Musikaufnahme und ermöglichte es Künstlern, zusammenhängende Alben zu veröffentlichen. Mit späteren Techniken wie dem Scratching im Hip-Hop, wechselte die Schallplatte sogar von einem reinem Wiedergabemedium zu einem musikalischen Performanceinstrument (Ananda, 2023).

Radios

Die Entwicklung des Radios begann mit den Arbeiten von Heinrich Hertz und Guglielmo Marconi (Garratt, 1994). Die ersten Übertragungen von drahtlosen Signalen schufen ein Fundament für die

Radiotechnologie und so wurde im Laufe der Zeit das Radio zu einem Massenmedium, das die Kommunikation revolutionierte und Menschen weltweit miteinander verband (Riegler & Handwerk, 2006).

Nachdem die Radiotechnik seit den ersten regelmäßigen Übertragungen 1907 aus einem New Yorker Hochhaus schnell an Beliebtheit gewonnen hat, markierte vor allem der erste Weltkrieg eine Zeit, in der Medienübertragung in den verschiedensten Bereichen unglaublich an Bedeutung gewann (Kittler, 1996). Es konnten im Kriegsgeschehen Nachrichten ohne große Verzögerungen übertragen werden, was den individuellen Parteien gewissermaßen Vorteile und Nachteile brachte (Kittler, 1996).

Für das gemeine Volk war der erste Weltkrieg aber ein Hemmnis, da die Entwicklung der Ziviltechnologie in dieser Zeit stagnierte und revolutionsgerichtete Ereignisse Ende des ersten Weltkriegs, wie etwa der "Funkerspuk" (Bleek, 2023) einen weitreichenden Einfluss auf die Verbreitung und Zensur des Rundfunks hatten (Dussel, 2022).

Als nach dem ersten Weltkrieg die zivilen und staatlichen Radiostationen errichtet wurden, waren die 1920er und 1930er Jahre die Blütezeit des Radios, in der es zu einem unverzichtbaren Bestandteil des täglichen Lebens wurde (Hilmes & Loviglio, 2002).

In den 1930er Jahren erreichte das Radio seinen Höhepunkt als Informationsquelle, Unterhaltungsmedium und politisches Werkzeug (Smith, 2014). Die Menschen versammelten sich um ihre Radiogeräte, um Nachrichten, Musik und dramatische Darbietungen zu hören.

Inmitten des aufkommenden Nationalsozialismus spielte das Radio eine entscheidende Rolle als Propagandainstrument (Adena et al., 2015). Adolf Hitler erkannte die Macht dieses Mediums und förderte die Produktion von erschwinglichen Radiogeräten für die breite Bevölkerung. Der Volksempfänger, ein kostengünstiges Radio, wurde ab 1933 in Deutschland eingeführt (König, 2003).

Der Begriff "Volksempfänger" verdeutlichte die nationalsozialistische Ideologie der Einheit des Volkes.

Entgegen eines verbreiteten Glaubens waren die Geräte nicht dafür konzipiert nur staatliche Sender zu empfangen (König, 2003), dennoch war die Grundlegende Idee von Joseph Goebbels die Nationalsozialistische Ideologie zu verbreiten, die Bevölkerung zu beeinflussen und eine scheinbare Einheit in der Gesellschaft zu fördern (Dempf, 2021). Goebbels sprach vom Radio als „Der Rundfunk als achte Großmacht“ (Goebbels, 1940).

Die Manipulation dieses Mediums für politische Zwecke im Kontext des Nazi-Regimes illustriert die dunkle Seite der

Radiogeschichte und erinnert daran, wie mächtig audiobasierte Medien in der Gestaltung der öffentlichen Meinung sein können.

Tonfilm

Die Einführung des Tons in Filmen revolutionierte nicht nur die Branche, sondern veränderte auch die Art und Weise, wie das Publikum mit visuellen Erzählungen umgeht (Qizi & Kazimbekovich, 2023). Der Übergang von der Live-Musikbegleitung und gelegentlich, gesprochenen Prologen in Stummfilmen hin zum synchronisierten Ton in Tonfilmen, bedeutete eine wesentliche Veränderung der audiovisuellen Einflüsse auf Blick und Aufmerksamkeit (Chion, 2019).

Carlo Cenciarelli behauptet, dass diese Entwicklungen ebenfalls einen Einfluss darauf hatten, wie wir allgemein hören. Die Untersuchungen zum Beitrag des Kinos zu einer breiteren Hörkultur sind allerdings noch nicht aussagekräftig (Cenciarelli, 2021).

Tonband

Der deutsche Ingenieur Fritz Pfeumer entwickelte das erste Magnettonband in den 1920er Jahren, indem er magnetische Partikel auf einem Kunststoffstreifen fixierte (Pernè, 2014). Die Einführung des Magnettonbandes in den 1930er und 1940er Jahren brachte mehrere Vorteile

im Vergleich zu bis dato genutzten Schellackplatten. Das Tonband konnte länger bespielt und neu überschrieben werden. Es eröffnete zudem Bearbeitungsmöglichkeiten des Inhalts, was zu kreativen Experimenten in der Musikproduktion führte (Kimizuka, 2012).

Pierre Schaeffer, ein französischer Toningenieur und Radiopionier, spielte eine entscheidende Rolle in der Nutzung des Magnettonbandes für experimentelle Musik. Zusammen mit Pierre Henry entwickelte er die sogenannte "Musique concrète" (Albiez, 2003). Diese avantgardistische Musikrichtung nutzte aufgezeichnete Klänge, Geräusche und Stimmen, die auf dem Magnettonband bearbeitet wurden, um neue Klanglandschaften zu schaffen (Schaeffer et al., 2012). Pierre Schaeffer wird auch als der "Godfather of Sampling" bezeichnet und hat durch seine Arbeit einen bedeutenden Teil zur heutigen Art und Weise der Musikproduktion beigetragen. (Patrick, 2016)

Stereo

Ein signifikanter Wendepunkt in der Geschichte war der gesellschaftliche Übergang zu Stereo-Systemen. Der Einsatz von Stereophonie revolutionierte die Art und Weise, wie Klangreproduktion wahrgenommen wurde (Katz, 2010). In den 1930er Jahren wurde bei Bell Lab versucht, die Schallwellenfront, die sich

aus einer unendlichen Anzahl von Mikrofon-/Lautsprecherkanälen ergeben würde, durch die Verwendung einer geringeren Anzahl von Kanälen zu approximieren (Rumsey, 2001). Die durchgeführten Experimente beinhalteten die Aufnahme von Klängen auf einer Bühne, gefolgt von einer Wiedergabe auf einer anderen Bühne (Rumsey, 2001). Die Ergebnisse dieser Experimente zeigten, dass bei Verwendung von drei Kanälen überzeugende Resultate erzielt wurden. Hingegen führte eine Reduktion von drei auf zwei Kanäle zu einer Vergrößerung der Breite der reproduzierten Klangbühne, wobei sich die zentralen Quellen räumlich in den hinteren Teil der Bühne verschoben (O'Brien, 2018). Später wurde klar, dass die Ursache dafür der psychoakustische Präzedenzeffekt ist, der im Abschnitt 2.3 genauer beschrieben wird.

Alan Blumlein reichte Anfang der 1930er ein Patent ein, das die Umwandlung von Signalen aus räumlich angeordneten Mikrofonen für die Lautsprecherwiedergabe ermöglichte. Durch Amplitudenunterschiede zwischen Lautsprechern werden Phasenunterschiede erzeugt, um eine bessere Schallquellenortung zu erzielen. Die Arbeit präsentierte, wie ein Zwei-Lautsprecher-System durch die Kontrolle von Signalamplituden eine präzise Beziehung zwischen der Originalposition

einer Schallquelle und ihrer wahrgenommenen Position in der Wiedergabe schaffen kann (Blumlein, 1931).

Blumleins Arbeit wurde lange Zeit nicht kommerziell umgesetzt aber in den 1950er Jahren von *Electric and Musical Industries Ltd.* (EMI) wieder aufgegriffen (Vanderlyn, 1978) und Methoden ähnlich dem vom Blumlein vorgeschlagenen Prinzip, zur Aufzeichnung von Zweikanal-Stereoton auf Schallplatten eingeführt.

Viele Aufnahmen waren monophone Aufnahmen, die mit Kammfilter und Bandsplittingtechnologien nachbearbeitet wurden, um einen stereophonen Effekt auf Stereo-Endgeräten zu erzielen (Rumsey, 2001).

Durch die Einführung des Ultra-Kurzwellen-Radios (UKW), der nach anfänglichen Monogeräten in den 1960ern hauptsächlich stereophon vertrieben wurde, hatte eine breite Masse plötzlich Zugang zur Stereotechnik (Riegler & Handwerk, 2011).

Frühe Pop-Stereomischungen waren teilweise extrem in der Anwendung von Panning-Effekten und wurden deswegen oft als Ping-Pong-Stereo bezeichnet (Rumsey, 2001).

Quadrophonie

In weiterer Folge wurden diverse weitere Systeme, wie die Quadrophonie auf dem

Markt vorgestellt. Da diese aber, trotz einer angedachten breiten Markteinführung, teils fehlerbehaftet waren, konnte sich der Ansatz bei Verbraucher*innen nie durchsetzen (Flied, 1978). Die 4.0 Lautsprecherkonfiguration war nicht mit einer Zweikanalwiedergabe kompatibel und lieferte ein suboptimales akustisches Bild (Schabbing, 2005). Ein Scheitern des Formates liegt vermutlich einerseits an der Vielzahl an Codierungsformen, unter denen sich nie ein einzelner Standard etabliert hat. Die Tatsache, dass Benutzer*innen des Systems im Gegensatz zu Stereo die doppelte Anzahl an Lautsprecher benötigen sowie grobe Marketing und Distributionsfehler sind laut Schoder ebenfalls dafür verantwortlich (Schoder, 2013).

Ambisonics

Die Grundlage für die Entwicklung von Ambisonic Sound lieferte das zuvor erwähnte Blumlein Patent, wobei die ersten tatsächlichen Anwendungen erst in den 1970er Jahren geschahen (Malham, 2019). Es war als umfassender Ansatz für eine „full sphere“ Klangwiedergabe gedacht (Gerzon, 1974).

Michael Gerzon, einer der Pioniere auf dem Gebiet, war von der Quadrophonie nicht überzeugt und verfasste einen umfangreichen Artikel¹ über deren

¹https://audiosignal.co.uk/Resources/What_is_wrong_with_quadraphonics_A4.pdf14.01.2024

Probleme. Folglich entwickelte er das Soundfield-Mikrofon, ein erste Ordnung Ambisonics-Mikrofon im A-Format (Thornton, 2010). Über dessen Entwicklung sagte Gerzon folgendes:

„from the observation that the effects one would like to produce include a continuum of directions around the listener. Such systems imagine a limited number of channels being used to convey the sound to the listener, but are designed to recreate a continuous range of directions around the listener approximating [to] the original. The mathematics used is not ‘matrix’ algebra (which is used only to describe transformations of a finite number of variables) but ‘kernel’ algebra (which is the corresponding mathematics used when one has an infinite continuum of variables).“ (Thornton, 2010)

Das System ist agnostisch gegenüber Lautsprecherkonfigurationen und ermöglicht die Darstellung von räumlichem Klang.²

Auf Ambisonics wird in Abschnitt 2.4 ausführlicher eingegangen.

² Ein interessanter Artikel mit Bildern zu den ersten Soundfield-Mikrofonen kann hier gefunden werden: <https://www.michaelgerzonphotos.org.uk/ambisonics.html> 14.1.2024

Surround-Sound

Obwohl der Disney-Film *Fantasia* bereits 1941 Gebrauch von einem Surround-Soundsystem machte (Klapholz, 1991), dauerte es noch Jahrzehnte bis Surround-Sound im Konsumentenbereich erschien.

Dolby Surround wurde 1982 eingeführt und ermöglichte ein Surround-Sound-Erlebnis, insbesondere für den Heimkino-Bereich (Teufel Editorial, 2015).

Durch die Einführung der AC-3 (Dolby Digital) Audiocodingtechnologie wurde der 5.1 Standard (drei Frontkanäle, zwei Surround-Kanäle und ein optionaler Low-Frequency-Enhancement (LFE)) vorgestellt (Davis, 1993), der auch heute noch Anwendung findet.

Die ITU-Empfehlung für Surroundsound wurde bereits im selben Jahr erarbeitet (ITU Radiocommunication Assembly, 1992). Sie beinhaltet eine 5.1

Kanalkonfiguration³, wobei der linke und der rechte Frontkanal die Kompatibilität mit Zweikanal-Stereo beibehalten (Holman, 2014).

Bei Bedarf kann eine separate Zweikanalmischung zusammen mit einer Surroundmischung desselben Materials übertragen werden. Obwohl es keine genaue Richtlinie zur Schallfeldwiedergabe gibt, wurden die drei

³ Das System besteht aus drei Frontkanälen, zwei Surround-Kanälen und ein optionaler Low-Frequency-Enhancement (LFE)-Kanal, der wegen seiner begrenzten Bandbreite, er gibt üblicherweise Frequenzen zwischen 20-120 Hz wieder, als "0.1"-Kanal bezeichnet wird.

vorderen Kanäle für primäre Signalquellen mit klaren Richtungsmerkmalen vorgesehen, während die hinteren/seitlichen Kanäle immer nur als unterstützende Effektkanäle zur Verstärkung der räumlichen Wirkung gedacht waren.

Oft wird der Eindruck vermittelt, dass mit einem Surround-Layout eine Rundum-Lokalisierung ermöglicht wird, was aber technisch schwierig realisierbar ist (Rumsey, 2001). Es können zwar seitliche und hintere Klangbilder erzeugt werden, diese sind aber weit nicht so stabil wie Klangbilder der vorderen Lautsprecher (Blauert & Rabenstein, 2012). Aufgrund des großen, hinteren Winkels, ist auch eine Abbildung durch die hinteren Kanäle problematisch (Rumsey, 2001).

CD

1982 wurde die Compact Disc (CD) kommerziell eingeführt und löste die damaligen Schallplatten und Kassetten schnell ab (Bolin, 2012). Die CD bot mit 74 Minuten und 33 Sekunden viel Platz als digitaler Speicher, eine höher aufgelöste Klangqualität und ein robustes Medium (Immink, 2009). Die CD ebnete den Weg in die digitale Audiowelt für die Bevölkerung. Mittlerweile stirbt die CD durch die Einführung von digitalen Audio-Distributionswegen fast gänzlich aus (Goodwin, 2020).

Geräte wie der Walkman und später tragbare CD-Player haben regelrecht Kultstatus erlangt oder wie Paul du Gay es ausdrückt:

“The Sony Walkman is not only part of our culture. It has a distinct ‘culture’ of its own.” (Du Gay, 2013)

Die Geräte ermöglichten einem breiten Publikum plötzlich, ihre Musik überall hin mitzunehmen.

DAW

Etwa gleichzeitig revolutionierten die Digital Audio Workstations (DAW) die Musikproduktion (Bell, 2018). Computer ermöglichten, mit DAWs die digitale Aufnahme, Bearbeitung und Mischung von Audiomaterial in einem wie Bell beschreibt: „space-less studio“ (Bell, 2018). Durch die flexible und effiziente Arbeitsweise ist es möglich Audio in beinahe Echtzeit zu bearbeiten und komplexe Routingnetzwerke zu erstellen (Levine, 2019). Der Zugang zu einer breiten Palette von virtuellen Instrumenten und die verschiedenen Möglichkeiten von Sampling, eröffneten Musikern neue Produktionsmöglichkeiten (Levine, 2019). Durch DAWs wurde es möglich, dass auch Künstler*innen, die sich bis dato keine Studioaufnahmen leisten konnten, zu Hause aufzunehmen begannen. Dies resultierte zum einen in einer extremen Zunahme der verbreiteten Musikedichte und zur anderen Extreme in der Schließung von Studios wie dem

legendären „Olympic Studios“ in London (Homer, 2009).

MP3

MP3-Audio wurde als Teil des MPEG-1 Standards 1991 das erste Mal öffentlich vorgestellt (Chiariglione, 2011) und ist ein Verfahren zur verlustbehafteten Kompression digital gespeicherter Audiodateien. Bei der Kodierung werden dabei psychoakustische Wahrnehmungseffekte genutzt, um so weit wie möglich, nur die für den Menschen hörbaren relevanten Signalanteile zu speichern (Brandenburg, 1999). Der Erfolg von MP3 hängt eng mit dem 1999 veröffentlichten Peer-to-Peer Netzwerk *Napster* zusammen (Hunter et al., 2001). In einer Zeit, in der man sich mit 56k-Modems in das Internet eingewählt hat, war eine radikale Reduktion der Datenmengen nötig. Mithilfe von MP3 lassen sich Audiodateien um bis zu fünfundachtzig Prozent, ohne drastischen Qualitätsverlust, komprimieren (Haug & Weber, 2002). Der Service nahm seinen Lauf und in kurzer Zeit war es für jeden mit Internetzugang möglich, seine Musiktitel zu teilen und anzubieten. Mit den Titeln verbreitete sich auch das Format weltweit und wird bis heute genutzt (Hunter et al., 2001).

⁴ Nicht nur Napster trug mit nicht-rechtskonformen Mitteln zur Verbreitung von MP3 bei. Nicht lange nach Beginn des Verkaufs der MP3-Software kaufte ein australischer Student mit einer gestohlenen

Napster wurde aufgrund rechtlicher Copyright-Klagen vom Netz genommen (FAZ, 2002), ebnete aber durch den innovativen Ansatz den Weg für internetbasierte Vertriebsmodelle wie iTunes und Ähnliche (Gomes, 2016).⁴ Die Patente für MP3 sind 2017 weltweit ausgelaufen und die Lizenzierung wurde eingestellt (Fraunhofer IIS, 2017).

Streaming

Streaming und On-Demand Audiodienste, sind mittlerweile weltweit die führende Art und Weise, wie Musik konsumiert wird. Laut *Music Consumer Insight Report 2018* konsumierten im Jahr 2018 bereits 86 Prozent aller Hörer*innen Musik auf Streaming und On-Demand Audiodiensten wie etwa *Spotify*, *Youtube* oder *Apple Music* (ifpi, 2018).

Der Erfolg dieser Dienste beruht in Teilen darauf, dass Benutzer*innen eine bequeme, individuelle und personalisierbare Musikerfahrung geboten wird (Maftei et al., 2016).

Haberer behauptet, dass die Entwicklung des Musikkonsums durch Streaming nicht nur ökonomische Auswirkungen hat, sondern auch neue Rezeptionsformen schafft, die Praktiken, Kulturen und

Kreditkarte die Software und stellte sie zur freien Verfügung ins Internet.
<https://www.mp3-history.com/en/marketing.html>, 08.01.2024

Ästhetiken des Musikhörens hervorbringen (Haberer, 2020).

Verschiedene Mechanismen, wie selbstorganisierende Playlists, dem Qualitätsverlust durch Datenkompression auf Kosten schnellerer Übertragung⁵ und der Kopplung von sozialen Netzwerken und den dadurch einhergehenden algorithmischen Empfehlungssystemen, verändern die bis dahin übliche Art und Weise wie Musik konsumiert wird⁶ (Papenburg, 2013).

Anhand von *Spotify* ist ebenfalls erkennbar, dass Genres mittlerweile von vielen Hörer*innen als überflüssig empfunden werden und Hörer*innen eher Fans von Künstler*innen und "Vibes" (Gefühlen) sind, da viele Top-Künstler*innen auch mittlerweile genreübergreifend performen (Brooks, 2023).

ITU-R BS.1770

Auch der "Lautheitskrieg" der 2000er Jahre hatte die Gewohnheiten der Hörer*innen sowohl auch der Produzent*innen verändert (Vickers, 2010). Die Einführung von Dynamik und Lautheitsstandards wie dem ITU-R BS.1770⁷ beendeten diesen Konflikt (Camerer, 2010) und es folgte langsam

⁵ Dies trifft insbesondere auf die 2000er Jahre zu, als die Kompressionsverfahren noch nicht derart fortgeschritten waren und hörbare Artefakte verursachten.

⁶ für eine ausführliche Beobachtung, wie Streamingdienste mit Nutzerdaten arbeiten,

eine Renaissance des qualitativen Hörens, die jedoch eher schleppend verlief.

Erst durch die Einbindung von verlustfreien Streaming-Möglichkeiten, könnte eine höhere qualitative Audioauflösung wieder an breitbandigem Interesse gewinnen (Coffey, 2016).

Objektbasierte Audiosysteme

Dolby Atmos wurde 2012 im Kino präsentiert und führte die objektbasierte Audiotechnologie damit in den Konsumentenbereich ein.

Dolby Atmos wurde seit der Einführung im Kino auf verschiedene Unterhaltungsplattformen ausgeweitet.

Im Gegensatz zu herkömmlichen Surround-Sound-Systemen, die auf kanalbasiertem Audio aufbauen, nutzt Dolby Atmos eine objektbasierte Audiotechnologie (Thomas, 2023).

Diese Technologie ermöglicht es Toningenieuren, nicht nur Klänge auf bestimmten Kanälen zu platzieren, sondern auch die Position von Audioobjekten im dreidimensionalen Raum zu definieren.

Jedes Audioelement wird als unabhängiges Objekt betrachtet, das in Bezug auf seine Position im Raum und

wird folgender Artikel empfohlen:
https://www.researchgate.net/publication/355657452_Versuch_uber_Spotify_oder_Musikstreaming_als_Arbeit_am_Subjekt, 14.01.2024
⁷<https://www.itu.int/rec/R-REC-BS.1770>, 23.12.2023

weiteren Parametern gesteuert werden kann. Die Technologie und verschiedene Ansätze werden in Abschnitt 2.4 genauer ausgeführt.

Dementsprechend gewinnt auch der MPEG-H Standard an Bedeutung. Die Entwicklung von MPEG-H begann als Antwort auf die wachsenden Anforderungen an hochwertigen und flexiblen Raumklang in verschiedenen Anwendungsbereichen, einschließlich Rundfunk, Streaming und Virtual Reality. MPEG-H wird in Abschnitt 3.1 detaillierter beschrieben.

KI-gestützte Systeme

Die letzten technologischen Fortschritte in der Musiktechnologie sind Künstliche-Intelligenz (KI)-unterstützte Prozesse. Dies umfasst Deep-Fake Musik⁸, automatisierte Masteringprozesse⁹, Audibearbeitungswerkzeuge wie dynamische Equalizer und De-Reverberation Tools¹⁰ sowie viele weitere Plug-Ins, die selbstständig analysieren und dadurch zur Klängaufwertung beitragen.

Die traditionelle Vorstellung, dass kreative Berufe wie Audio-Mastering aufgrund ihrer Abhängigkeit von Intuition, affektivem Wissen und menschlichen „Eigenschaften“ immun gegen KI-Einflüsse sind, wird durch Technologien von Firmen wie *LANDR*¹¹ oder *Izotope*¹² in Frage gestellt (Birtchnell & Elliott, 2018). Unternehmen wie *LANDR* nutzen KI-gesteuerte Audio-Mastering-Systeme, die auf maschinellem Lernen basieren, um die von menschlichen Toningenieuren durchgeführten Prozesse zu replizieren.

Birtchnell und Elliott beschreiben das Verhältnis von künstlich intelligenten Masteringtools und menschlicher Expertise trotzdem als synergetische Beziehung, die der vorherrschenden Vorstellung widerspricht, dass KI menschliche Arbeitsplätze ersetzen wird (Birtchnell & Elliott, 2018).

Allgemein werden KI-gestützte Prozesse die Medienwelt vermutlich verändern. Durch die unvorhersehbare Dichte an Innovationen in den letzten Jahren, ist die Zukunft dieser Technologien derzeit aber rein spekulativ

⁸ Im folgenden Link findet sich ein Bericht der Deepfake Musik genauer analysiert: <https://ceur-ws.org/Vol-3528/paper3.pdf>, 04.01.2024

⁹ beispielsweise <https://www.izotope.com/en/products/ozone.html>, 05.01.2024

¹⁰ beispielsweise <https://www.soundtheory.com/home>, <https://product.supertone.ai/clear>, 05.01.2024

¹¹ <https://www.landr.com/de/>, 06.01.2024
¹² <https://www.izotope.com/en/products/ozone.html>, 08.1.2023

2.3 Grundsätze von Spatial Audio

Räumliches Hören ist ein komplexer Prozess, der es ermöglicht, die Richtung, Entfernung und Position von Schallquellen in der Umgebung wahrzunehmen. Diese Fähigkeit ist von entscheidender Bedeutung für die Orientierung, Navigation und dem sozialen Austausch (Friesecke, 2014).

Head related transfer Functions (HRTF)

Das räumliche Hören beginnt bereits mit der Erfassung von Schallwellen durch das äußere Ohr, wo diese durch den Gehörgang zum Trommelfell weitergeleitet werden (Hudspeth, 1997). Die Ohrmuschel und die Form des Kopfes erzeugen Unterschiede in den Schallintensitäten und -laufzeiten zwischen den beiden Ohren, die als binaurale Höreindrücke bekannt sind (Kohlrausch et al., 2013).

Das menschliche Gehirn verarbeitet diese Unterschiede, und kann so die Richtung und Entfernung von Schallquellen bestimmen.

In der Horizontalen erfolgt die Richtungslokalisierung durch Unterschiede in Pegel und Laufzeit zwischen dem linken und rechten Ohr. Signale von links erreichen das rechte Ohr verzögert. Diese Verzögerungen werden als interaurale Zeitunterschiede (ITD) bezeichnet und spielen vor allem im Frequenzbereich von

80-800 Hz eine Rolle (Friesecke, 2014; D. R. Moore et al., 2010).

Im rechten Ohr treten die Schallwellen dann ebenfalls mit einem geringeren Pegel, aufgrund der frequenzabhängigen Abschattung des Kopfes, ein. Diese Laufzeitunterschiede sind als interaurale Pegelunterschiede (ILD) bekannt. Mit steigender Frequenz wird die ILD durch die zunehmende Abschattung immer ausgeprägter und gewinnt daher an Bedeutung für die Lokalisation (D. R. Moore et al., 2010). Dennoch ist die Laufzeit- und Pegelinformation nicht eindeutig, da es mehrere Positionen eines Schallereignisses mit gleichem Abstand zu beiden Ohren gibt. Dieses Phänomen, als „cone of confusion“ bezeichnet, wird von Menschen grundsätzlich durch eine leichte Drehung des Kopfes behoben, wobei das Gehirn in der Lage ist, falsche Positionen aufgrund von Bewegungsrichtung und -geschwindigkeit auszuschließen (Letowski & Letowski, 2012).

Eine zusätzliche wichtige Lokalisationsinformation ist die frequenzabhängige Filterung des Schalls durch Torso, Kopf und Außenohr. Abhängig von der Einfallsrichtung des Schalls werden verschiedene Resonanzen angeregt, wodurch jeder Richtung ein anderes frequenzabhängiges Profil

zugeordnet wird. Diese komplexe akustische Filterung trägt dazu bei, die Richtungserkennung bereits ohne Kopfbewegungen deutlicher zu gestalten, insbesondere bei Klängen die sich nach oben bewegen (Lan et al., 2022).

Zusammenfassend werden diese Mechanismen als Head-Related-Transfer-Functions (HRTF) bezeichnet (Rumsey, 2001).

Da jeder Körper individuell geformt ist, sind auch die HRTF's verschieden.

HRTFs werden als Filter für die Erzeugung virtueller 3D-Akustik über Kopfhörer verwendet. Dabei werden Signale mit den Impulsantworten (Head-related Impulse Responses, HRIRs) für die gewünschte Richtung gefaltet (Hammershøi & Møller, 2005).

Um binaurale Wiedergabesysteme breit verfügbar zu machen, werden standardisierte HRTFs verwendet.

Allerdings zeigen sich im frequenzabhängigen Verhalten der HRTFs starke individuelle Unterschiede, weshalb individuelle HRTFs eine wesentlich genauere Lokalisation ermöglichen als standardisierte HRTFs (Hammershøi & Møller, 2005).

Für eine akkurate Messung der persönlichen HRTFs sind hohe Anforderungen an die räumliche Verteilung der Messpositionen notwendig (Li & Peissig, 2020).¹³

Räumliche Faktoren

Natürliche Klanglandschaften zeichnen sich durch dreidimensionale Merkmale (Breite, Höhe, Tiefe) aus, die zu einem Gefühl der räumlichen Wahrnehmung beitragen. Hörer*innen verlassen sich bei der räumlichen Wahrnehmung sowohl auf das Seh- als auch auf das Hörvermögen. Wobei das Sehvermögen häufig für nach vorne gerichteten Szenen und das Gehör für Objekte dahinter und darüber verwendet wird (Rumsey, 2001).

Rumsey behauptet, dass Quellen im Freien in der Regel weit entfernt sind, und minimale Reflexionen aufweisen. Im Gegensatz dazu werden Innenräume, die von Oberflächen umschlossen sind, von Reflexionen beeinflusst, die die räumlichen Eigenschaften von Schallquellen verändern.

Die Reflexionen in Räumen sind in der Regel kurz nach dem Direktschall und Hörer*innen können dadurch Distanzen

¹³ Für einen ausführlichen Review verschiedener Messmethoden wird folgender Bericht empfohlen: Li, S., & Peissig, J. (2020). Measurement of Head-Related Transfer Functions: A review. *Applied*

Sciences, 10(14), 5014.
<https://doi.org/10.3390/app10145014>,
<https://www.mdpi.com/2076-3417/10/14/5014>,
08.01.2024

und Eigenschaften des Raumes besser wahrnehmen (Rumsey, 2001)

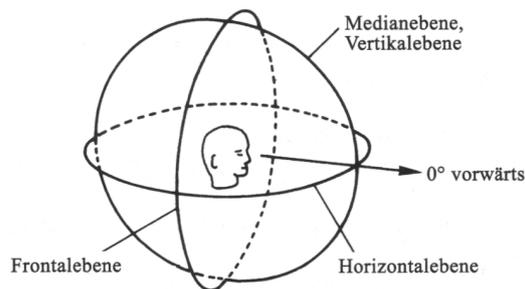


Abbildung 2, Hörebenen, Dickreiter et al. (2014), S. 128

Die Position von Schallquellen auf der Vertikale wird durch Frequenzgangveränderungen ermittelt. Die, von Jens Blauert erforschten, richtungsbestimmenden Frequenzbänder zeigen, dass Signale im Bereich von 250-500Hz und 2-7kHz eine Lokalisation von vorne fördern. Betonungen bei 1kHz oder 12-16kHz werden eher von hinten wahrgenommen, wobei der 8kHz-Bereich verstärkt von oben wahrgenommen wird (Sengpiel, 1995).

Reflexionen an den Ohrmuscheln tragen zu schmalbändigen Frequenzeinbrüchen bei, die wiederum die Höhenbestimmung beeinflussen können (B. C. J. Moore, 1995).

Moore behauptet, dass die Wahrnehmung der Entfernung einer Schallquelle von verschiedenen Aspekten abhängig ist. Frühe Erstreflexionen ermöglichen Rückschlüsse auf die Position innerhalb eines Raumes. Größere Entfernungen führen aufgrund der Luftabsorption zu einem Pegelabfall in den hohen

Frequenzen. Das Verhältnis von Direkt- zu Diffusschall liefert ebenfalls Informationen über die Entfernung einer Schallquelle (B. C. J. Moore, 1995).

Präzedenz, Haas- und Trading Effekt

Der Präzedenz- und dessen spezielle Form, dem Haas-Effekt, sind psychoakustische Effekte, die die Lokalisation von Schallquellen beeinflussen.

Brown, Stecker und Tollin beschreiben den Präzedenz-Effekt als ein psychoakustisches Phänomen. Wenn gleiche Schallsignal zeitverzögert aus verschiedenen Richtungen bei Hörer*innen eintrifft, nehmen diese nur die Richtung des zuerst eintreffenden Schallsignals wahr. Verzögerte Schallsignale werden in der Richtung des ersten Signals lokalisiert (Brown et al., 2015).

Haas beschreibt 1951 folgenden Effekt: Bei Verzögerungszeiten zwischen 10 ms und 30 ms (spezifisch $20 \text{ ms} \pm 10 \text{ ms}$) ist der zuerst eintreffende Schall maßgeblich für die Lokalisation der Schallquelle, unabhängig von der Richtung des verzögerten Schalls.

Trotz unterschiedlicher Einfallrichtungen wird nur eine Schallquelle wahrgenommen. Bei Laufzeitdifferenzen über 40 ms wird allmählich das Vorhandensein separater Schallreflexionen bemerkbar, jedoch bleibt

die Lokalisierung der Schallquelle weiterhin in Richtung des zuerst eintreffenden Schalls, sofern der Reflexionspegel dem Direktsignalpegel entspricht (Haas, 1951).

Laut Sengpiel kann dieser Effekt sinnvoll in der PA (Public Address) Beschallung eingesetzt werden. Um in großen Räumen eine optimale Schallversorgung für entfernte Hörer*innen sicherzustellen, werden anstatt eines Lautsprechers mehrere Lautsprecher im Raum verteilt. Die synchronisierte Abstrahlung aller Lautsprecher würde jedoch aufgrund des Präzedenz-Effekt dazu führen, dass Hörer*innen den Eindruck haben, die Schallquelle sei der nächstgelegene Lautsprecher. Um dies zu verhindern,

werden für weiter entfernte Lautsprecher Verzögerungszeiten eingestellt, sodass sie etwa 20 ms später abstrahlen als die erste Wellenfront. Der Haas-Effekt ermöglicht, dass der verzögert abstrahlende Lautsprecher den Direktschallpegel um bis zu 10 dB übertreffen kann, ohne den Eindruck einer separaten Schallquelle zu erzeugen. Dieser Effekt wird als Trading-Effekt bezeichnet, bei dem das Direktsignal "leise" und "früh" ist, während das verzögerte Signal "laut" und "spät" ist. Im Gegensatz dazu steht die Äquivalenz bei der Stereoaufnahme, bei der "laut" und "früh" in einem Kanal "leise" und "spät" im anderen Kanal entspricht (Sengpiel, 2003).

2.4 Spatial Audioformate und Standards

“SPATIAL AUDIO denotes the attempt to capture the salient parts of a sound field and reproduce it in some form at other, possible distant places (and times), such that a human listener perceives the spatial characteristics of the original sound scene to a large extent during reproduction.”

(Herre et al., 2015)

Spatial Audio hat sich auch als Überbegriff für eine Sammlung von Technologien etabliert, die gemeinsam eine authentische Illusion räumlicher Darstellungen von Tonereignissen und Szenen ermöglichen (Gowan, 2023). Als Grundlage dazu dienen die räumlichen Dimensionen von natürlichen Klängen und ihre Auswirkungen auf die menschliche Wahrnehmung, die in Abschnitt 2.3 genauer beschrieben werden.

Die *European Broadcast Union* (EBU) beschreibt grundsätzlich fünf verschiedene Arten von Audio (EBU ADM, 2023).

Kanalbasiert

Bei kanalbasiertem Audio sind dedizierte Kanäle dedizierten Lautsprechern zugewiesen, ergo die Lautsprecherkonfiguration ist im voraus festgelegt und kann auch nicht mehr verändert werden (Olivieri et al., 2019). Bei einer Stereo-Audiodatei beispielsweise ist, unabhängig vom Endgerät ein linker

Kanal für den linken Lautsprecher und ein rechter Kanal für den rechten Lautsprecher vorhergesehen.

Die Wiedergabe in kanalbasiertem Audio erfolgt durch die Ansteuerung der entsprechenden Lautsprecher, um die Schallquellen aus den vorgesehenen Richtungen zu reproduzieren. Durch Kombination von Klängen aus den verschiedenen Lautsprechern wird ein räumliches Klangbild erzeugt, das es den Hörer*innen ermöglicht, den Klang aus verschiedenen Richtungen zu hören (Olivieri et al., 2019).

Eine Herausforderung bei kanalbasierten Ansätzen ist, dass eine feste Lautsprecherkonfiguration besteht, die möglicherweise nicht immer ideal für die spezifische Raumakustik ist. Dies kann zu akustischen Problemen führen, wie zum Beispiel Phasenauslöschungen oder Überlagerungen, bei denen die Klangquellen nicht genau an der gewünschten Position wiedergegeben werden können (Amato, 2023).

Kanalbasiertes Audio kann ein befriedigendes Hörerlebnis bieten und findet weiterhin Anwendung in verschiedenen Bereichen wie Heimkino und Konzertsälen.

Es gibt Situationen, in denen kanalbasiertes Audio in andere Konfigurationen umgewandelt werden kann und soll. Eine häufige Anwendung ist

das Downmixing von 5.1 auf Stereo. (EBU ADM, 2023)

Matrixbasiert

Matrixbasiertes Audio bezeichnet die Kombination von Audiokanälen über Matrixgleichungen, um andere Kanäle zu erzeugen. Ein einfaches Beispiel dafür ist Mid/Side-Audio (MS)-Processing, bei dem der Mid-Kanal die Summe des linken und rechten Kanals in einem Stereopaar ist und der Side-Kanal die Differenz zwischen links und rechts darstellt („Mid/Side Processing – Das Wichtigste kompakt – Update 2018“, 2018).

Mid/Side Audio kann auch durch eine geeignete Aufnahmemethode erstellt werden. Für eine M/S-Aufnahme werden zwei Mikrofone verwendet. Ein Mikrofon, das auf die Schallquelle gerichtet ist und das Mono-Signal aufnimmt. Dafür wird in der Regel ein Mikrofon mit einer Kugel- oder Nierencharakteristik verwendet. Dieses wird als Mid-Mikrofon bezeichnet. Das Side-Mikrofon hat eine Achtercharakteristik und wird direkt über dem Mid-Mikrofon positioniert und um 90° gedreht, ohne es zu berühren. Dadurch nimmt das Mid-Mikrofon mehr direkte Schallanteile auf als das Side-Mikrofon. Aufgrund der räumlichen Nähe entsteht kein signifikanter Laufzeitunterschied zwischen den beiden Mikrofonen (Keller, 2017; Sonible, 2018).

Mid-Side wird auch bei der FM-Radio-Stereocodierung verwendet, sodass bei Nichtempfang des schwächeren Side-Kanals immer noch ein Mono-Signal (Mid) wiederhergestellt werden kann (EBU ADM, 2023).

Szenenbasiert

Szenenbasiertes Audio ist ein allgemeiner Begriff, der Ambisonics und Higher-Order-Ambisonics (HOA) einschließt. Im Gegensatz zur kanalbasierten Audioreproduktion, bei der jeder Kanal einen einzelnen Lautsprecher repräsentiert, ermöglichen Ambisonics und HOA eine räumlich unabhängige Darstellung eines Schallfeldes durch ein matrixbasiertes System (EBU ADM, 2023).

Ambisonics wird aufgrund ihrer Auflösung in Ordnungen unterteilt, wobei grundsätzlich zwischen First-Order-Ambisonics (FOA) und Higher-Order-Ambisonics unterschieden wird.

Ambisonics erster Ordnung besteht aus vier Komponentenkanälen. Die erste Komponente repräsentiert ein omnidirektionales Signal, das den Gesamtklang des Schallfeldes einfängt, während die nächsten drei Komponenten die Schallinformationen in den X (vorne/hinten)-, Y(links/rechts)- und Z(oben/unten)- Dimensionen des Raums kodieren (Zotter & Frank, 2019).

Diese Darstellung ermöglicht es, die Richtung, die Entfernung und die Höhe der

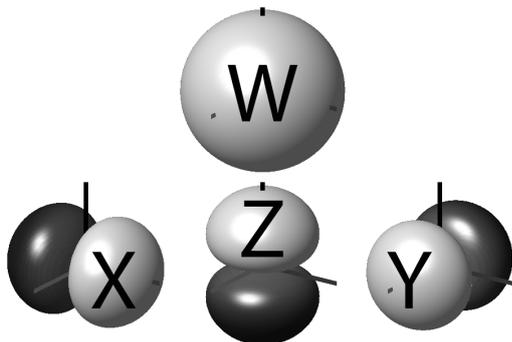


Abbildung 3, First-Order-Ambisonics, bearbeitet, by Dr Franz Zotter zotter@iem.at CC BY-SA 3.0 or GFDL, via Wikimedia Commons

Schallquellen zu erfassen und sie so zu reproduzieren, dass Hörer*innen ein umhüllendes Klangerlebnis erhalten.

Trotz der Möglichkeiten von First-Order-Ambisonics gibt es Einschränkungen, welche die räumliche Auflösung betreffen. Der Sweetspot von FOA hat ungefähr die Größe eines Tennisballes (0.08m) (Frank, 2014). Bei komplexen Klangumgebungen oder stark lokalisierten Schallquellen kann die räumliche Genauigkeit begrenzt sein, was zu einer weniger präzisen Wiedergabe führt. Ambisonics höherer Ordnung bietet durch die höhere Kanalzahl eine weitaus definiertere Abbildung.

Es ist wichtig zu beachten, dass die Anzahl der Komponentenkanäle mit der Ordnung wächst. Als Faustregel gilt, dass die Anzahl der Kanäle sich aus dem Quadrat der Ordnung addiert mit eins ergibt (Zotter & Frank, 2019)

$$K_{(\text{Kanalanzahl})} = (n_{(\text{Ordnung})} + 1)^2$$

Higher Order Ambisonics (HOA) erweitert die Ambisonics-Technologie durch die Verwendung von zusätzlichen Komponentenkanälen. Bei der 2. Ordnung werden demnach fünf zusätzliche Komponenten zu den ursprünglichen vier hinzugefügt, und bei der dritten Ordnung gibt es weitere sieben zusätzliche Komponente, wodurch insgesamt sechzehn Komponentenkanäle entstehen. Durch diese Erweiterung steigt die räumliche Auflösung erheblich an, und es wird möglich, Schallquellen genauer im Raum zu platzieren und komplexe Klangumgebungen realistischer

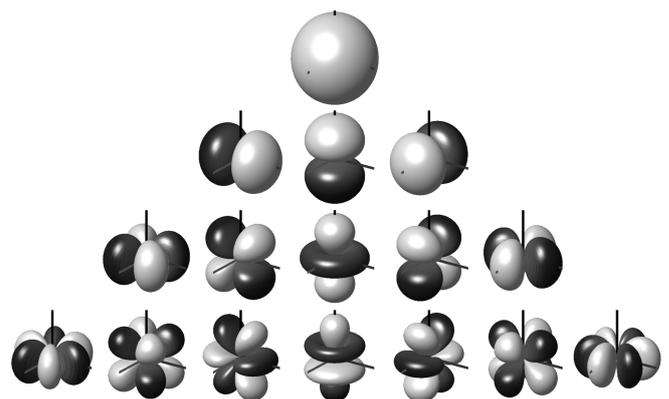


Abbildung 4, Third-Order-Ambisonics, by Dr Franz Zotter zotter@iem.at CC BY-SA 3.0 or GFDL, via Wikimedia Commons

wiedergegeben (Zotter & Frank, 2019).

Für die Wiedergabe wird ein Decoder verwendet, um die Tonszene auf dem vorhandenen Tonsystem korrekt wiedergeben.

Mit der Einführung der All-Round Ambisonic Decoding (AllRAD)-Methode (Zotter & Frank, 2012), die durch Sampling zu einer von Zotter, Frank und Sontacchi vorgeschlagenen, optimalen virtuellen *t-design*-Lautsprecheranordnung dekodiert (Frank et al., 2015), eröffnet sich ein flexibles Verfahren, das sehr gut mit unregelmäßigen Anordnungen umgehen kann (Zotter & Frank, 2012). Die Signale der virtuellen Lautsprecher werden mit Hilfe von Vector-Base Amplitude Panning (VBAP) auf den realen Lautsprecher abgebildet (Frank et al., 2015).

Da die Spurenzahl von Ambisonics dritter und siebter Ordnung genau durch acht teilbar ist, sind diese Ordnungen im Hinblick auf die Hardwarewiedergabegeräte effizienter, da Audioschnittstellen meist Ein- und Ausgänge in 8er-Gruppen besitzen (Bernsteiner, 2019).

Als Resultat der fortschreitenden Entwicklungen um Ambisonics wurde ein Standardisierungsvorschlag präsentiert, die AmbiX Konvention, um ein möglichst barrierefreies Arbeiten zwischen Ambisonic-Softwaretools zu ermöglichen (Nachbar et al., 2011). Die Konvention legt die Kanal Order *Ambisonics Channel Number* (ACN) und den SN3D Standard fest (Nachbar et al., 2011).

In der Praxis haben viele Plattformen, die Ambisonics in ihren Dienstleistungen

anbieten, sich bisher auf die erste Ordnung beschränkt (beispielsweise *YouTube* und *Facebook*) (Rieger, 2019). Besonders für ambiente Aufnahmen und Virtual-Reality-Anwendungen kann Ambisonics erster Ordnung ein ausreichendes räumliches Hörerlebnis bieten, da räumliches Hören stark mit visuellen Reizen verknüpft ist (Gil-Carvajal et al., 2016).

Dennoch bleibt die Implementierung höherer Ordnungen eine Herausforderung für die Audiobranche, um die Grenzen von Spatial Audio-Anwendungen zu erweitern und realitätsgetreuere 3D-Audioerfahrungen zu ermöglichen.

Binauralbasiert

Binaurales Audio ist dafür vorgesehen, räumliches Audio über Kopfhörer wiederzugeben. Binaurales Audio nutzt die Eigenschaften der Klangverarbeitung im Gehirn, die sich über die Physiologie des menschlichen Ohres und dem Torso ergeben, siehe Abschnitt 2.3, um den Eindruck eines immersiven Klangs über zwei Kanäle (das linke und das rechte Ohr) zu vermitteln.

Binaurales Audio kann mit einem binauralen Renderer generiert werden, der jede der anderen Audioformate interpretieren kann, oder es wird direkt mit einem binauralen Mikrofon (entweder

einem Dummy-Kopf¹⁴ oder In-Ear-Mikrofonen¹⁵) erzeugt.

Ein Nachteil eines binauralen Audiosignals besteht darin, dass es schwierig ist das Signal anders als auf Kopfhörern wiederzugeben. Es gibt zwar verschiedene Ansätze zum Upmixing (Jukka, 2005), diese eignen sich aber kaum in einem Produktionsumfeld.

Objektbasiert

Objektbasiertes Audio bedeutet, dass jedem Audiokanal Positions- (und möglicherweise andere räumliche oder signalbezogene) Metadaten zugeordnet sind. Jeder Kanal, beziehungsweise jedes Audio-Objekt repräsentiert einzelne Klänge in einer gesamten Szene, sodass es viele verschiedene Objekte geben kann, die gemeinsam die gesamte Klangszene aufbauen (EBU ADM, 2023). Audio-Objekte müssen nicht in der gesamten Klangszene existieren und sie können über die Zeit individuell ihre Position und ihre Eigenschaften ändern. Die ursprünglichen Audiosignale im objektbasierten Audio können als einfache Monosignale ohne Metadaten wiedergegeben und dadurch auch mit herkömmlichen Audiobearbeitungswerkzeugen individuell bearbeitet werden. Die mitgelieferten Metadaten geben einem Renderer ausreichend Informationen, um

den Klang räumlich richtig zu positionieren und mit anderen räumlichen Merkmalen (wie Diffusität und Größe) anzupassen (Bleidt et al., 2017).

Ein großer Vorteil von objektbasiertem Audio ist, dass die Metadaten unabhängig von der tatsächlichen Lautsprecheranordnung sind und es ermöglichen, dasselbe Audioobjekt beziehungsweise dieselbe Klangszenerie in verschiedenen Wiedergabesystemen zu dekodieren, ohne dass eine Neumischung erforderlich ist.

Die meisten objektbasierten Audiosysteme zielen darauf hin, auf eine Lautsprecherkonfiguration mit Höhenkanälen wiedergegeben zu werden. Dabei werden klassische Surround-Konfigurationen wie 5.1. oder 7.1, zwei bzw. vier Höhenlautsprecher hinzugefügt. Eine 7.1 Aufstellung mit vier Höhenkanälen würde demnach in der Regel als 7.1+4H bezeichnet werden (*Dolby Atmos Speaker Setup*, 2023).

Grundsätzlich gibt es neben anwendungsspezifischen Systemen vier „große“ kommerzielle, objektbasierte 3D-Audiosysteme, die je nach Anwendung Vor- und Nachteile mit sich bringen.

Dolby Atmos ist der aktuelle Marktführer in der objektbasierten Surround-Sound-Technologie und wurde 2012 eingeführt (Thomas, 2023). Atmos hat im Heimkino

¹⁴ <https://www.neumann.com/de-de/produkte/microphones/ku-100/>

¹⁵ <https://www.dpamicrophones.de/immersiv/45-60-binaurales-kopfbuegelmikrofon>

schnell an Popularität gewonnen und ist auf einem großen Teil von AV-Receiver, sowie Fernsehern und Soundbars, mittlerweile standardisiert verfügbar. Es ermöglicht klassische Konfigurationen wie 5.1., 7.1 und 9.1 Setups, die zwei bis acht Höhenlautsprecher zu dem traditionellen Setup hinzufügen. Für *Dolby Atmos Home* können maximal 34 Lautsprecher verwendet werden (Dolby, 2018) und für die professionelle Anwendung in Kinos 64 (Dolby, 2023a).

DTS:X wurde 2015 eingeführt und unterstützt bis zu 32 verschiedene Lautsprecherkonfigurationen. Es konkurriert mit Atmos, ist jedoch auf physischen Medien und Streaming-Diensten weniger verbreitet, obwohl *DTS:X* durch die Vielzahl verschiedener Lautsprecherkonfigurationen flexibler einsetzbar ist (Westphal, 2021). *DTS Virtual:X* bietet eine vereinfachte Version von *DTS:X* für Soundbars und Systeme mit begrenzter Lautsprecheranordnungsmöglichkeit (DTS, 2023).

Auro-3D wurde bereits 2006 vorgestellt und bietet ein dreistufiges Sound-System das in Surround, Height und Top Lautsprechern unterteilt wird. *Auro-3D* erfordert eine höhere Anzahl von Lautsprechern, was zu höheren Kosten führt, und ist aufgrund von Unterschieden in der Überkopfkannalnutzung nicht ideal

mit anderen Systemen kompatibel (*AURO-3D*, 2023).

MPEG-H 3D Audio ist Teil des Audio- und Video-Standards MPEG-H. Ähnlich wie Dolby Atmos ermöglicht es Entwickler*innen, Audioobjekte im 3D-Raum zu platzieren. Obwohl MPEG-H noch nicht weit in Hardware verbreitet ist, wird es bereits von südkoreanischen und brasilianischen Übertragungsinstitutionen genutzt, sowie in einigen Lautsprecher- und Receiverprodukten.

Als vielseitiger Standard könnte MPEG-H in Zukunft eine begehrte Option für 3D-Audio im Live-TV werden und wurde deshalb im Werkteil als Arbeitsmittel genutzt. Im folgenden Kapitel 3.1 wird genauer auf die Eigenschaften von MPEG-H Audio eingegangen.

Audio Definition Model (ADM)

Das Audio Definition Model (ADM) ist ein Metadatenmodell, das von der *European Broadcasting Union* (EBU) entwickelt wurde, um eine umfassende technische Beschreibung von Audioinhalten zu ermöglichen. Marston beschreibt ADM wie folgt:

ADM basiert auf XML und ist darauf ausgelegt, verschiedene Audioformate abzudecken, einschließlich traditionellem Stereo und komplexerem objektbasiertem Audio. ADM umfasst mehrere verschiedene Elemente. Obwohl es das Format von Audio-Tracks, -Streams,

-Kanälen, -Blöcken und -Packs beschreibt, gibt es keine Vorgaben, wie der Ton wiedergegeben werden sollte. Es zielt stattdessen darauf ab, ausreichend Informationen für Renderprogramme bereitzustellen, um ihren Anforderungen gerecht zu werden. Die wesentliche Bedeutung von ADM liegt darin, auf die sich entwickelnde Landschaft von Audioerlebnissen einzugehen,

einschließlich höherer Auflösung, Interaktivität und immersivem Audio. „Gute“ Metadaten, wie sie durch das ADM ermöglicht werden, sind entscheidend, um eine korrekte Verarbeitung von Audio während der Produktion, Verteilung und Ausstrahlung sicherzustellen (Marston, 2014).

3. Theoretischer Rahmen

"The ear tends to be lazy, craves the familiar, and is shocked by the unexpected; the eye, on the other hand, tends to be impatient, craves the novel and is bored by repetition."

- W. H. Auden

3.1 MPEG-H: Merkmale und Fähigkeiten

MPEG-H ist ein multimedialer Standard, der von der *Moving Picture Experts Group* (MPEG) gemeinsam mit dem *Fraunhofer IIS* und deren Partner*innen entwickelt wurde (Fraunhofer IIS, 2023). MPEG-H ist ein Standard in mehreren Abschnitten, der Teile für den Transport digitaler Medien und für Video-Kompression beinhaltet, aber auch den für die Arbeit relevanten Audiostandard MPEG-H 3D Audio umfasst.

Grundfunktionen

MPEG-H zeichnet sich durch seine Personalisierungsfunktionen aus (siehe Abbildung 5, 6), die die Erstellung individuell einstellbarer Dialogebenen, anpassbare Audiobeschreibungen, mehrere Sprachen und die interaktive Positionierung von Objekten unterstützt. So können Nutzer*innen ihr Erlebnis auf die individuellen Vorlieben und Bedürfnisse abstimmen und auf allen Arten von Wiedergabegeräten abspielen. (MPEG-H Audio Tools, 2023)

Die innovativen Eigenschaften von MPEG-H lässt sich in drei grundlegenden Funktionen zusammenfassen:

1. Die Möglichkeit, dreidimensionale Klangszene zu schaffen und zu broadcasten.
2. Universelle Übertragung, die eine optimierte Wiedergabe über verschiedene Klassen von Wiedergabeoptionen sicherstellt. Beispielsweise über verschiedene Lautsprecherkonfigurationen aus einem einzigen Audiostrom.
3. Erweiterte Benutzerinteraktivität und Barrierefreiheit, die es Benutzer*innen ermöglicht, die endgültige Audiomischung an die individuellen Bedürfnisse und den Geschmack anzupassen. Beispielsweise durch Personalisierung des Pegels und der Position des Dialogs. Die interaktiven Optionen können über eine Fernbedienung oder eine sprachgesteuerte Benutzeroberfläche gesteuert werden (Simon et al., 2019).

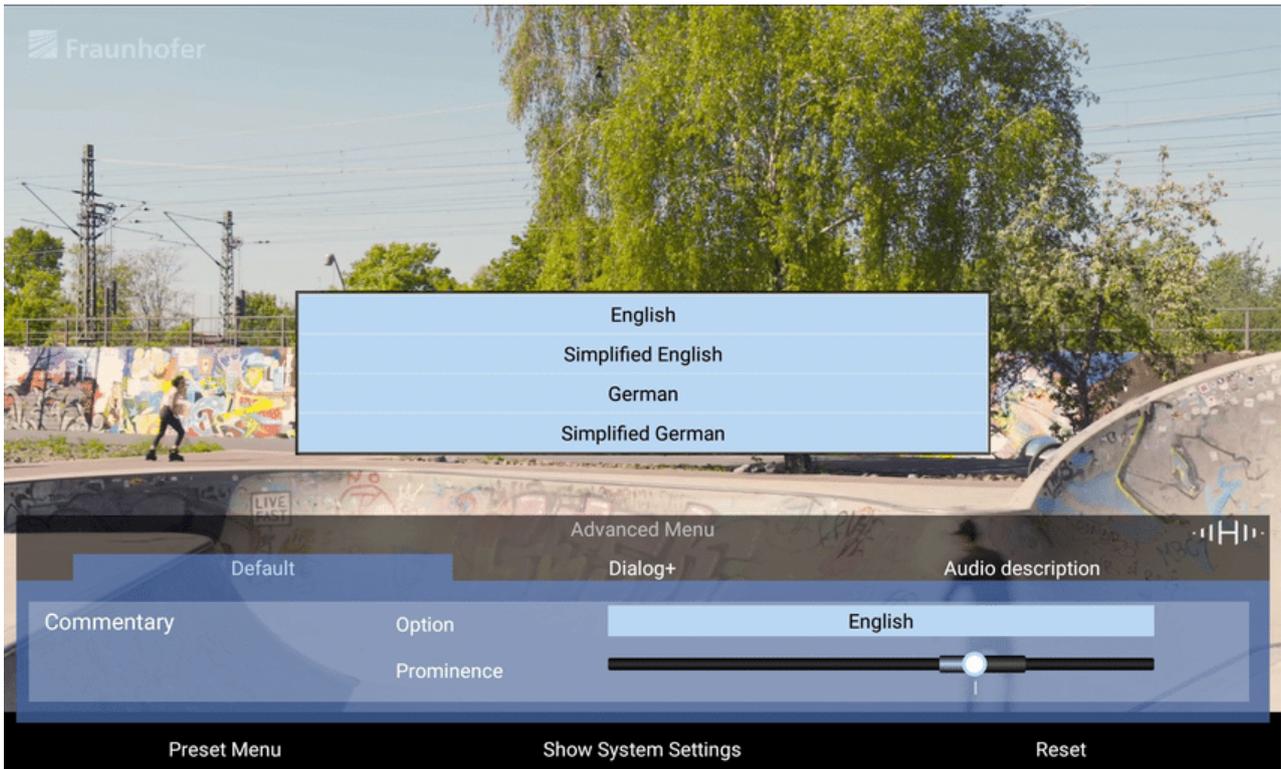


Abbildung 5, Benutzerdefinierte Auswahl, MPEG-H Audio for Improving Accessibility in Broadcasting and Streaming - Scientific Figure on ResearchGate. Available from: <https://www.researchgate.net/figure/Screenshot-of-an-accessible-MPEG-H-Audio-audio-scene-with>

MPEG-H 3D Audio basiert auf einem objektbasierten Übertragungskonzept. Die Audiosignale werden zusammen mit einer Metadatenspur encodiert, die nach der Übertragung am Endgerät decodiert wird. Die Metadaten enthalten Informationen über die Audiosignale und definieren die Presets sowie den Umfang der User-Interaktionen. Sie werden als XML-Datei gespeichert.

MPEG-H ermöglicht theoretisch bis zu 128 Kanäle, 128 Audioobjekte und Ambisonics bis zur 29. Ordnung (Fleischmann, 2017).

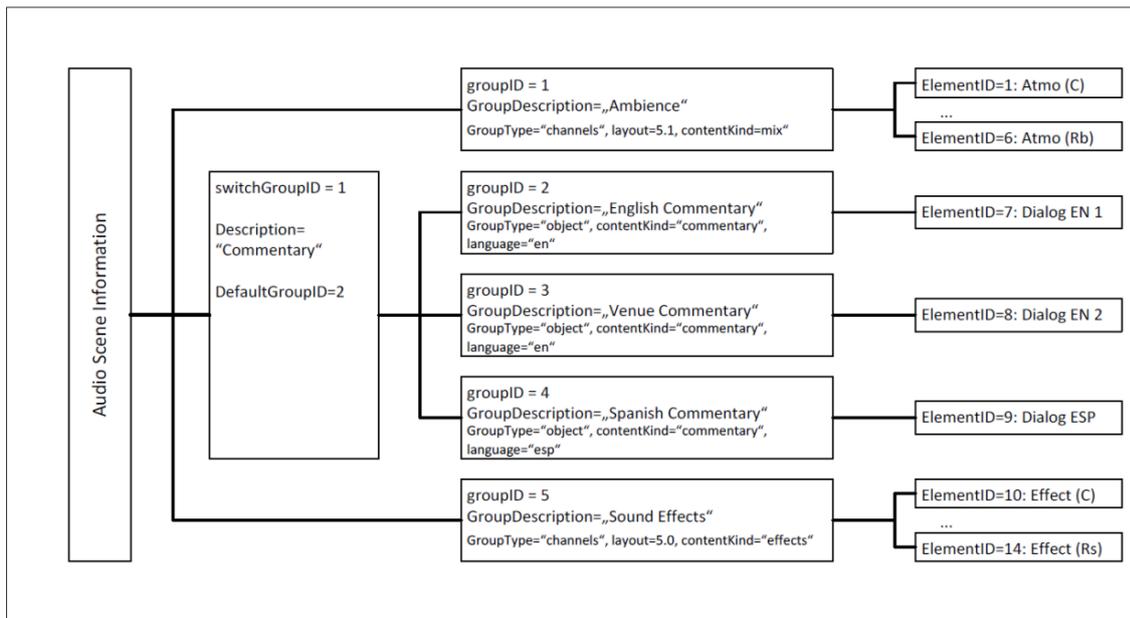


Abbildung 6, Bild: MPEG-H Audio Scene Information example (2017) figure 4.1, ATSC Standard: A/342 Part 3, MPEG-H System, Doc. A/342-3:2017, ATSC, S. 5

Im MPEG-H Workflow werden Audiospuren als Audio-Elemente bezeichnet.

Diese können zu einer gemeinsamen Audio-Szene zusammengefasst und in Gruppen oder Switchgroups organisiert werden.

Switchgroups ermöglichen den nahtlosen Wechsel zwischen zwei Audiospuren.

Wird eine Spur der Switchgroup aktiviert, wird eine andere automatisch deaktiviert.

Die so entstehende Hierarchie ist in den Metadaten der Audio-Szene gespeichert (siehe Abbildung 6), (Bleidt et al., 2017).

Dynamische Objekte werden mit einem Zeitstempel versehen, der an den Encoder übergeben und anschließend unabhängig der Produktionsumgebung auf das jeweilige Wiedergabeformat gerendert wird (Bleidt et al., 2017).

Produktion

Die Produktionstools für das MPEG-H Audio System sind zentral bei der Erstellung von Metadaten. Die Tools ermöglichen den Content-Ersteller*innen, im Authoring-Prozess interaktive Funktionen und Optionen für die 3D-Klangwiedergabe festzulegen.

Die Standalone-Software und DAW-Plugins sind lokal auf Maschinen installierte Tools, die die Erstellung, Bearbeitung und Handhabung von Dateien mit MPEG-H-Szenen ermöglichen. Sie können nahtlos in Host-Anwendungen integriert oder als dedizierte Authoring-Tools verwendet werden. Diese Tools exportieren Audio und Metadaten in das MPEG-H Audio Master Format (MPF oder BWF/ADM) (Grewe et al., 2021), das für die

Archivierung, weitere Verarbeitung und Codierung geeignet ist.

Für Live-Produktionen sind Hardware-basierte Lösungen in Form von Audio Monitoring and Authoring Units relevant. Diese Geräte ermöglichen eine schnelle Metadatenerstellung in Echtzeit, sowie die Überwachung von Audio- und Metadatenrendering (Grewe et al., 2021).

Das Unternehmen *LZE.Innovations* stellt die MPEG-H Authoring Suite (MAS)¹⁶, die eine Palette an Tools bereitstellt, zum derzeitigen Zeitpunkt (Ende 2023) frei zur Verfügung stehen. Mit den Werkzeugen können Anwender*innen AV- oder reine Audio-Inhalte verfassen, konvertieren, erstellen, überwachen, analysieren, codieren, Authoringprozesse abwickeln und Produktions- sowie gerenderte Dateien abspielen.

Die MAS umfasst sieben eigenständige Tools sowie eine exemplarische MPEG-H-Playback-App für Android OS Geräte und wird von den Entwickler*innen wie folgend beschrieben (MPEG-H, 2023):

MPEG-H Authoring Tool (MHAT)

Ein umfassendes Software-Tool zum Definieren spezifischer MPEG-H Audio-Parameter. Es ist ein direktes Monitoring der gewählten Lautsprecher-Konfigurationen möglich und Mischungen können exportiert werden.

MPEG-H Authoring Plug-in (MHAPi)

Das Plug-In ermöglicht es, mit einer bevorzugten DAW eine MPEG-H Audio-Produktion auf Objekt- oder Kanalbasis zu erstellen und hat grundlegend dieselben Funktionen wie das MHAT, kann aber als Plug-In geladen werden.

MPEG-H Production Format Player (MPF)

Der MPF-Player ermöglicht die Überprüfung der A/V-Synchronisation, die Bewertung der Szenenerstellung, das Prüfen von Renderlayouts und das Vorschauen von Interaktivität und kann so zur Quality Control eingesetzt werden.

MPEG-H Info Tool (MHIT)

Das MHIT bietet Informationen zu den Details eines MPEG-H Projekts. Es ermöglicht einen Einblick in dateibezogene Informationen sowie eine Szenenübersicht eines MPEG-H Masters mit sämtlichen Metadaten. Darüber hinaus kann es automatisierte Tests für ADM basierende Inhalte durchführen.

MPEG-H Encoding & Muxing Tool (MHEX)

Das Programm ermöglicht es, Audio-Daten einfach zu encodieren und in mehrere Ausgabeformate zu multiplexen,

¹⁶ Die einzelnen Tools werden auf der Website veranschaulicht präsentiert und können dort

direkt bezogen werden: <https://mpeg-h.com/mas/> , 10.12.2023

wahlweise mit oder ohne Hinzufügen einer bereits vorab codierten Videospur.

MPEG-H VVPlayer (MHVP)

Der VVPlayer ermöglicht das Abspielen von codierten MPEG-H Audio MP4-Dateien und die Anpassung sämtlicher benutzerdefinierbarer Funktionen während der Wiedergabe.

MPEG-H Conversion Tool (MCO)

Das Programm ermöglicht das Konvertieren von MPEG-H-konformen Content-Masters zwischen verschiedenen Dateiformaten. Es erlaubt die Integration von Dateien, die von anderen Tools erstellt wurden, und fügt MPEG-H Authoring in bestehende Workflows ein.

Die konkrete Verwendung der Programme sowie der Authoringprozess wird in Abschnitt 4.3 genauer erläutert.

Lautheit und Dynamik

NGA-Standards wie MPEG-H unterstützen verschiedene Werkzeuge zur Kontrolle der Lautstärke und Anpassung der Programmdynamik in den Metadaten, um sich an unterschiedliche

Wiedergabeumgebungen und Endgeräte anzupassen.

Dazu gehören die Messungen und Überwachung der Programmlautheit, Anpassung der Dynamik einzelner Audio-Elemente (Dynamic Range Control), Peak- und Clipping-Kontrolle sowie

Duckingfunktionen für Voice-Over und Audiodeskriptionen (Bleidt et al., 2017).

Um eine gleichmäßige Programmlautheit am Decoder sicherzustellen, erfolgt zunächst eine Normalisierung, die bei Änderungen adaptiv angepasst wird (Kuech et al., 2015).

MPEG-H Audio bietet Messungen nach internationalen Broadcast-Standards wie ITU-R BS.1770-4, EBU R-128 oder ATSC A/85 sowie Lautheitskompensation (Bleidt et al., 2017).

Die Dynamic Range Control (DRC) ermöglicht die Einschränkung der Dynamik von Audiosignalen, wobei Gain-Veränderungen nicht destruktiv sind, da sie mit den Metadaten übertragen werden und anschließend der Decoder über deren Anwendung entscheidet (Bleidt et al., 2017).

Unterschiedliche DRC-Profile können für verschiedene Umgebungsszenarien oder Endgeräte übertragen werden. Ziel ist es, eine angestrebte Target-Loudness beizubehalten und Peaks oder Clipping durch Lautheitsregulierungen zu verhindern, notfalls durch einen Peak-Limiter am Ausgang des MPEG-H Decoders (Kuech et al., 2015).

Rundfunk und Streaming

Das Gesamtsystem ist bereits standardisiert und wird vom größten kommerziellen TV-Netz Lateinamerikas *TV GLOBO, Rede Amazonica* in Brasilien

genutzt und wurde als einziges obligatorisches Audiosendesystem der neuen digitalen Fernsehstandartisierung TV 3.0 und dem brasilianischen Kommunikationsministerium sowie dem brasilianischen Forum für digitales terrestrisches Fernsehen (SBTVD Forum) ausgewählt (*Brazil*, 2023).

In Südkorea wurde der MPEG-H Standard bereits 2017 mit dem dort im Dauerbetrieb gestarteten *4K UHD TV* Streamingbetrieb als erster Next-Generation-Audio Standard eingeführt (*South Korea Launches UHD TV with MPEG-H Audio*, 2017).

Im Jahr 2019 führte das *Fraunhofer IIS* gemeinsam mit Technologiepartnern und der EBU einen Test beim Songcontest in Tel-Aviv durch (*Eurovision Song Contest (ESC) trifft Next Generation Audio (NGA)* - , 2020).

Dieses getestete Übertragungskonzept wurde auf der *International Broadcast Convention (IBC) 2019* nachgebildet. Da der Produktionsablauf für Broadcasting im Vergleich zu einer Offline-Produktion anders gestaltet ist, wird in folgendem Abschnitt nicht näher auf die Produktionsdetails eingegangen aber ein Überblick über die verschiedenen Stationen geschaffen.

¹⁷<https://solidstatellogic.com/products/s500>, 14.01.2024

¹⁸ <https://www.telosalliance.com/on-air-tv-audio-processing/monitoring/linear-acoustic-ams>, 14.01.2024 oder <https://junger-audio.com/en/legacy->

Im konkreten Fall wurde die *System T S500*¹⁷ Broadcastkonsole genutzt um einen 5.1+4H Mix der Show, inklusive fünf verschiedener Sprachspuren als Audioobjekte, zu erstellen.

Als nächsten Schritt wurden die Metadaten nun mithilfe eines Hardware MPEG-H Authoringsystems¹⁸ in Echtzeit geauthored und überwacht. Als letzte Übertragungsinstanz generierte ein geeigneter Broadcastencoder¹⁹ einen Rundfunk-Transportstream sowie Internetstreams in adäquaten Formaten wie DASH oder HLS.

Die Wiedergabe geschah in diesem Fallbeispiel über eine MPEG-H fähige Soundbar und als binauraler Stream über ein Tablet mit Kopfhörern (*MPEG-H Audio System Live-Produktionskette vom Eurovision Song Contest 2019 auf der IBC – präsentiert vom Fraunhofer IIS – Fraunhofer Audio Blog*, 2019).

In Broadcast oder Streaming-Anwendungen wird für die Metadaten ein sogenannter *Control-Track* generiert. Es handelt sich dabei um ein PCM-Signal, ähnlich dem eines Timecodes, das dem MPEG-H Kontributions- oder Emission-Encoder zugeführt wird.

[product/flexai/multichannel-monitoring-and-authoring-mma](https://www.flexai.com/multichannel-monitoring-and-authoring-mma) 14.01.2024, 14.01.2024

¹⁹ Beispielsweise <https://www.ateme.com/press/ateme-titan-powers-a-new-4k-uhd-channel/>, 14.01.2024

Der Control-Track ist mit der jeweiligen Videoframerate synchronisiert und ist resistent gegenüber einfachen Signalmanipulationen (Bleidt et al., 2017).

Interaktivität in MPEG-H Produktionen

Wie zu Beginn dieses Abschnitts beschrieben, können im MPEG-H Authoring Switchgroups festgelegt werden. Im konkreten Fall sind die Sprecher*innen als Switchgroups zu behandeln, da diese ja nie gleichzeitig spielen sollten.

Da Produktionen mit MPEG-H Implementierung immer noch lineare Medien sind, gibt es Grenzen bezüglich der Interaktivität. Lineare Medien zeichnen sich durch einen vordefinierten und festgelegten Ablauf aus. Dies bedeutet, dass der Inhalt bereits im Voraus strukturiert abgespielt wird, ohne dass Benutzer*innen aktiv in den zeitlichen Verlauf eingreifen können (Hoepner, 2019). Diese Strukturiertheit des Mediums begrenzt die Interaktionsmöglichkeiten erheblich.

Bucher unterscheidet zudem zwischen Interaktivität und Selektivität.

Selektivität bezieht sich auf die Reduzierung der Komplexität von Informationsangeboten durch Rezipient*innen. Die Selektion erfolgt nach

bestimmten Kriterien, um die Information auf ein verträgliches, nützliches oder angenehmes Maß für die Weiterverwendung zu reduzieren. Interaktivität dagegen, ermöglicht eine größere Handlungs- und Entscheidungsfreiheit in der Rezeption digitaler Medien. Diese Freiheiten sind auf die hypertextuelle und hypermediale Struktur zurückzuführen, die es den Nutzer*innen ermöglicht, ihre Lektüre, Programme und Nutzungspfade eigenständig zu gestalten (Bucher, 2001). Laut Bucher ist der Begriff Selektion aber zu eng definiert, um alle Aspekte des selektiven Auswählens mit einzubeziehen (Bucher, 2001).

In Medien, in denen sich die Interaktivität, wie in dieser Arbeit, auf vorprogrammierte Optionen oder vordefinierte Pfade beschränkt, zeichnet sich so eher ein selektiver Charakter ab.

Ein weiterer Faktor ist die begrenzte Rückkopplungsmöglichkeit in linearen Medien. Die Reaktionen auf Benutzerinteraktionen sind im Voraus festgelegt und bieten wenig Spielraum für dynamische Anpassungen. Dies führt dazu, dass das Medium wenig flexibel auf spontane Handlungen der Benutzer*innen reagieren kann.

3.2 Alternative immersive Audio-Anwendungen

Objektbasierte Audioformate wie Dolby Atmos finden in verschiedenen Bereichen wie dem Kino, Videospielen oder in der Musikproduktion Anwendung (Dolby, 2023b), wobei MPEG-H bisher vor allem bei Sportveranstaltungen und Konzerten getestet wurde (DWHI Sao Paulo, 2019), (Fraunhofer IIS, 2018). Für MPEG-H gestützte Medien gibt es außerhalb dieser Produktion noch kaum Anwendung. Objektbasiertes Audio im Allgemeinen wird dennoch bereits in verschiedenen Bereichen angewendet.

Hörspielproduktion

Durch binaurale Aufnahmetechniken sind 3D-Hörbücher immer wieder am Markt zu finden²⁰, stellen aber ein Nischenprodukt dar. Michael Lafrance beschreibt in einem Interview jedoch die Zunahme an neuen Produktionen mit innovativen Techniken: *„I do think it's important and publishers have certainly been criticised in the past for being slow to adapt to new technologies, but having attended FutureBook in December, I was impressed with a lot of the new technologies and approaches publishers are using to appeal to younger, more tech-driven audiences. Hopefully this embracing of new formats and technologies continues.“* (Abma, 2016)

Binaural produzierte Hörspiele gibt es zwar schon seit den Anfängen der binauralen Audiotechnik, beispielsweise die von *BBC* produzierten Radiohörspiele „The Revenge“ 1978 oder „Bravo November“ 2008 (Pike, 2013), wurden aber bereits in den 1980ern aufgrund ihrer Anwendungseinschränkung als „failed technology“ bezeichnet (Krebs, 2017). Trotzdem werden „3D Hörspiele“ von großen Produktionen, wie den „Drei ???“ immer wieder als Spezialfolge veröffentlicht. In diesem Fall sogar in Dolby Atmos (*Die drei ??? in besonderem Klang*, 2023).

Das Studio *Owl Field* bietet auch nicht-lineare 3D-Hörspiele. Das bedeutet, dass die Hörer*innen die Geschichte beeinflussen können, indem sie am Ende eines jeden Kapitels Entscheidungen treffen (*Productions · Owl Field*, 2022).

Ein Grund dafür, dass es am Markt vergleichsweise wenige immersive Hörspiel-Produktionen im Gegensatz zu Stereo-Produktionen gibt, ist vermutlich unter anderem mit dem hohen Produktionsaufwand und den damit verbundenen Kosten verbunden (Abma, 2016).

²⁰ Eine Empfehlung an dieser Stelle sind die Owl Field Studios, die hochqualitative 3D-

Hörbücher produzieren: <https://www.owlfield.com/> , 10.11.2023

Interaktive Medien

Die potenzielle Anwendung von objektbasiertem Audio geht jedoch über diese Bereiche hinaus. Ein möglicher Einsatz wäre beispielsweise der interaktive Film, in dem die dynamische Natur von objektbasiertem Audio das Eintauchen der Benutzer*innen in die Welt des Films verbessern kann. Adaptive Audioerlebnisse auf der Grundlage von Benutzerentscheidungen über Presets in einer MPEG-H Produktion können dies ermöglichen. Ebenso erdenklich sind Produktionen in Bildungsinhalten, bei denen Schüler*innen gamifizierte, visuelle Inhalte gezeigt werden und über die optionalen Audioobjekte Spuren hinzu- und weggeschaltet werden können. Interaktivität fördert das Lernverhalten (Abma, 2016), (Sims, 2000) und durch das Wegschalten der Information lässt sich der gelernte Inhalt direkt überprüfen.

Videospiele

Videospiele, als Art Symposium vieler innovativer multimedialer Disziplinen, profitieren ebenfalls durch Implementierung von objektbasiertem Audio (Audiokinetic, 2021). Audio Middleware-Programme wie *WWise*²¹ verwenden objektbasiertes Audio und der Inhalt kann direkt in dem Programm geauthored werden (Kastbauer, 2021). Eine Studie von Gallacher beschreibt die Rolle von Audio bei der Immersion in Videospiele.

Die Studie umfasste ein Testspiel, das von 20 Teilnehmern mit und ohne Audio gespielt wurde. Die Ergebnisse deuten darauf hin, dass eine adaptive Hintergrundbeschallung die Leistung der Spieler und das Gefühl der Immersion verbessert hat (Gallacher, 2013).

Automobilindustrie

In der Automobilindustrie optimiert objektbasiertes Audio das akustische Erlebnis im Fahrzeug (Kaprocki et al., 2018). Die Audiotechnologie wird genutzt, um nicht nur eine verbesserte Klangqualität für Musik und Unterhaltung zu bieten, sondern auch, um wichtige Warnsignale, Navigationshinweise und Verkehrsinformationen auf eine klare und zielgerichtete Weise zu vermitteln (Duprè et al., 2021). Duprè argumentiert auch,

21

<https://www.audiokinetic.com/en/products/wwise/>, 14.01.2024

dass objektbasiertes Audio dazu beiträgt, dass Fahrer*innen und Mitfahrer*innen relevante akustische Informationen erhalten, ohne von anderen Geräuschen im Fahrzeugumfeld abgelenkt zu werden (Dupré et al., 2021).

In einer Umgebung, in der visuelle Ablenkungen vermieden werden sollten, spielt Audio eine grundlegende Rolle bei der Bereitstellung von Sicherheits- und Kommunikationsinhalten im Auto (Kaprocki et al., 2018).

Immersive Audio Experience in

Reisemedien und Tourismus

Statt sich auf herkömmliche visuelle Darstellungen zu beschränken, eröffnen moderne Audio-Technologien, Reisemedien eine neue Plattform. Audioguides sind seit ihrer Einführung 1952 ein fester Bestandteil in Reisemedien (Moserrat Narvaez, 2019). Wo Bücher vor allem im Vorhinein einen visuellen Eindruck eines Ortes schaffen können, aber im Moment als Informationsquelle durch das Abwenden des Blickes störend wirken, kann man sich durch eine auditiv geführte Tour vollkommen auf das Erleben einlassen. Die Firma *Detour*²² ist ein Unternehmen im Bereich der Audioguides, das innovative Konzepte für immersives Erzählen und Audioführungen entwickelt.

Die App nutzt die GPS-Technologie von Mobiltelefonen, um Benutzer*innen durch Audioführungen von verschiedenen Stadtvierteln zu führen. Die Idee ist, den Nutzer*innen Hintergrundinformationen und Anweisungen zu geben, während sie verschiedene Sehenswürdigkeiten oder "narrative Triggerpunkte" in der Umgebung erreichen (Locke, 2015).

Die Audioguides von *Detour* haben das Ziel, die Geschichte und Bedeutung von Gebäuden, Straßen und Stadtvierteln zu vermitteln, die für gewöhnlich von Tourist*innen oder sogar von Einheimischen unbemerkt bleiben könnten. Die Führungen dauern in der Regel etwa eine Stunde, passen sich jedoch an die Gehgeschwindigkeit der Benutzer*innen an (Epstein, 2017). Dadurch bleibt die Erfahrung nahtlos und interaktiv.

Die Audioführungen werden von Personen erstellt, die mit den jeweiligen Stadtvierteln gut vertraut sind. Diese Erzähler*innen können lokale Persönlichkeiten, Geschichtenerzähler*innen oder Expert*innen sein, die ihre persönlichen Einblicke und Erfahrungen teilen. Es kann sogar vorkommen, dass die App von Benutzer*innen verlangt, mit ihrer Umwelt zu interagieren oder in bestimmte Einrichtungen zu besuchen. Das Ziel ist es, den Benutzer*innen eine Variante der

²²

<https://play.google.com/store/apps/details?id=c>

[om.detour.detour&hl=de&gl=US&pli=1](https://play.google.com/store/apps/details?id=com.detour.detour&hl=de&gl=US&pli=1),
10.01.2024

bekanntesten touristischen Attraktionen zu bieten (Velasco-Puffleau, 2023). Obwohl *Detour* in den Anfangsjahren auf großes Interesse stieß und als Pionier in der audio-geführten Touristik galt (Everson, 2015), hatte die Firma später Schwierigkeiten, den Erfolg aufrechtzuerhalten (May, 2018).

Einen innovativen Ansatz bietet die Anwendung *Sonictraces*. Die App-Anwendung ermöglicht den Benutzer*innen, sich in einer virtuellen Umgebung mittels ihrer eigenen Smartphones und Kopfhörer durch den realen Raum zu bewegen. Dabei werden sie Teil einer Geschichte und der dazugehörigen Klangwelt. Mithilfe von dritter Ordnung Ambisonics werden Klänge so erzeugt und manipuliert, dass eine dreidimensionale Umgebung simuliert wird. Die Verwendung von realen Orten verstärkt die Authentizität der Geschichten, indem ein Gefühl von Realismus vermittelt wird. *Sonictraces* nutzt GPS und optisches AR-Tracking im Freien, um die Position der Personen zu bestimmen.²³

Die Anwendung bietet nichtlineare Inhalte, bei denen Benutzer*innen durch ihre

Bewegungen den Verlauf der Storyline beeinflussen können, und ermöglicht es, Klangräume relativ zu positionieren, um begehbare Orte virtuell zu erleben (*AR Audio. Der next level 3D Audio Guide.*, 2020).

Als immersives Klangerlebnis sind die Klangspaziergänge von Hildegard Westerkamp eine Erweiterung ihrer Arbeiten im Bereich der akustischen Umgebung und des ökologischen Hörens (Velasco-Puffleau, 2023). Westerkamp entwickelte die Soundwalks als künstlerisches Mittel, um Menschen zu einer bewussteren Wahrnehmung ihrer akustischen Umgebung zu führen. Die Soundwalks von Westerkamp sind oft in spezifischen Orten verankert, wobei sie Klänge der Umgebung aufnimmt und diese dann künstlerisch bearbeitet und montiert. Aufnahmen, die während dieser Soundwalks entstehen, sind sorgfältig kuratierte akustische Collagen, die die Teilnehmer*innen in eine poetische Klangwelt eintauchen lassen und sich mit ihrer akustischen Ökologie auseinandersetzen.

Im Werk „Kits Beach Soundwalk (1989)“²⁴ führt Westerkamp Hörer*innen als

²³ Die Entwickler*innen geben keinen vollen Einblick in den Produktionsprozess, aber ein Screenshot auf der Website zeigt die Verwendung der Game-Engine *Unity*. Daraus lässt sich schließen, dass die räumlichen Aufnahmen angefertigt und in der Gamewelt platziert werden. So kann der Ambisonicstream direkt auf die Position der „Spieler*innen“ gerendert werden. <https://sonictraces.com/>, 18.01.2024

Für weiteres Interesse von Ambisonicseinbindungen in Unity wird folgende Anwendung empfohlen:

<https://resonance-audio.github.io/resonance-audio/>, 18.01.2024

²⁴ <https://www.youtube.com/watch?v=hg96nU6fLk>, 12.01.2024

erzählende Stimme durch den Mikrokosmos des Kitsilano Beach in Vancouver (Westerkamp, 1989). Mithilfe von audiotecnischen Bearbeitungswerkzeugen wie Equalizern entfernt sie nach und nach die Stadt aus dem Klangbild und wechselt in traumhafte Sequenzen. Auffällig ist auch die, von

Westerkamp in diesem Stück erwähnte Nähe zur „Musique concrète“ (siehe Abschnitt 2.1). Durch geschicktes Arrangement schafft sie so eine künstlerische Betrachtung von Orten und lässt Hörer*innen in Details eintauchen.

4. Erstellung eines MPEG-H Audio-Teaser anhand der Beispielstadt Valencia

"A methodical regularity is visible in the apparent irregularities of nature."

- Mary Somerville

4.1 Design der Audio-Reise durch die Stadt

Um ein geeignetes Design zu erlangen, müssen zuerst die Rahmenbedingungen festgelegt werden. Die Audiotour sollte als Fallbeispiel für eine alternative Inhaltserstellung eines MPEG-H gestützten Mediums dienen.

Wie in Abschnitt 3.1 beschrieben, bietet die Wiedergabe von MPEG-H Dateien eine individuelle Gestaltungsmöglichkeit des Erlebten. Dies wird durch vorprogrammierte Presets und Audioobjekte, die eine Interaktivität der Endbenutzer*innen in linearen Grenzen erlauben, ermöglicht.

Um die interaktiven Möglichkeiten von MPEG-H zu nutzen wurde entschieden, zwei unterschiedliche Charaktere mit verschiedenen Altersgruppen und Interessensprofil zu wählen, die Hörer*innen Informationen zu den vorgestellten Plätzen der Tour geben. Dadurch sollte ein möglichst breites Publikum angesprochen und eine vielseitige Benutzererfahrung geschaffen werden.

Im Gegensatz zu einem Soundwalk wie in Abschnitt 3.2 beschrieben, der auf künstlerisch gestaltete Klanglandschaften und „vergrößerten“ Details setzt, verfolgt das Design dieser Audiotour einen werbeorientierteren Ansatz. Es setzt auf ein einfaches Narrativ, das durch modernes Sounddesign verstärkt wird.

Die Integration von MPEG-H 3D Audio und gezielten Sound-Designelementen, sollte eine dynamische und lebhaft Atmosphäre erschaffen.

Viele Menschen treffen ihre Reiseentscheidungen basierend auf visuellen Grundlagen in Büchern, Prospekten und dem Internet (Weitenthaler, 2017). Das vorherige Erleben des Ortes und Einfühlen auf die Situation ist aber durch eine eindimensionale Bildfläche beschränkt und die tatsächliche Wahrnehmung des Ortes kann von der Vorstellung abweichen (Tercia et al., 2020). Dimitris und Vasiliki beschreiben drei unterschiedliche Aspekte, die zufriedenstellende Reiseerfahrungen prägen. Der Erste ist die Vorerfahrung, die durch Nacherzählungen oder Informationen vor der Reise definiert ist. Der zweite Aspekt ist die eigentliche Erfahrung. Wie etwa Sehenswürdigkeiten bestaunen, das Essen genießen, Freizeitaktivitäten ausüben und ähnliches. Der letzte Aspekt ist die Erfahrung nach der Reise: die Erinnerungen, die Gespräche mit Freunden über das Erlebte (Dimitris & Vasiliki, 2013). Weiters behaupten Puerta et al. in einem Bericht über ein immersives Reiseprojekt, dass Tourismusagenturen virtuelle Tourguides im Allgemeinen gut verwenden können,

um aktuelle Reiseziele zu bewerben. Durch immersive Technologien können sich Menschen so vorab ein Bild von ihrer Wunschdestination machen. Auf diese Weise würden die Agenturen den erwähnten ersten Aspekt einer zufriedenstellenden Reiseerfahrung erfüllen (Puerta et al., 2016).

Studien zeigen, dass sich der Konsum von Medien von traditionellen Printmedien in Richtung webbasierten und sozialen Medienplattformen verändert hat (Bhuller et al., 2023).

Trotzdem sind die meisten konkreten Reisemedien immer noch Printmedien (Sheivachman, 2017).

Durch die zunehmende Nutzung sozialer Medien durch Reisende für die Suche, Organisation und den Austausch reisebezogener Informationen (Ruiz-Mafe et al., 2016), etabliert sich so eine indirekte Art von Reisemedien.

Diese beiden Tatsachen müssen bei dem Design eines Tourguides für innovative Medien in Betracht gezogen werden, darum wurde der in Abschnitt 4.4 gezeigte Flyer mit QR-Code entwickelt, um diese Aspekte zu erfüllen.

Sprecher

Da die geplante Audiotour mit zwei Sprechern im selben Audiobett verwendet wurde, war es wichtig, einen ähnlichen Text zu erstellen. Dieser sollte zwar grundlegend dieselben Informationen

vermitteln, sich jedoch in Punkten vom Anderen unterscheiden.

Um diese Aufgabe zu bewältigen, wurde entschieden, eine eher einfache Sprache und Wortwahl zu benutzen und den Text für die Kapitel relativ kurz und sinnweisend zu halten, sodass das Endergebnis eine Länge von jeweils fünf Minuten nicht überschreitet.

Der jüngere Charakter im Alter von 25 Jahren wird als abenteuerlustig beschrieben und repräsentiert eine Zielgruppe, die offen für neue Erfahrungen ist und aktive Aktivitäten bevorzugt. Dynamische Kommentare eine lebendige und aufregende Atmosphäre vermitteln.

Der ältere Charakter im Alter von 55 Jahren, als kulturbegeistert beschrieben, repräsentierte eine Zielgruppe mit Interesse an kulturellem Erbe. Kommentare von diesem Charakter geben Informationen zu historischen und kulturellen Empfehlungen.

Im Anhang können die erarbeiteten Texte durchgelesen werden.

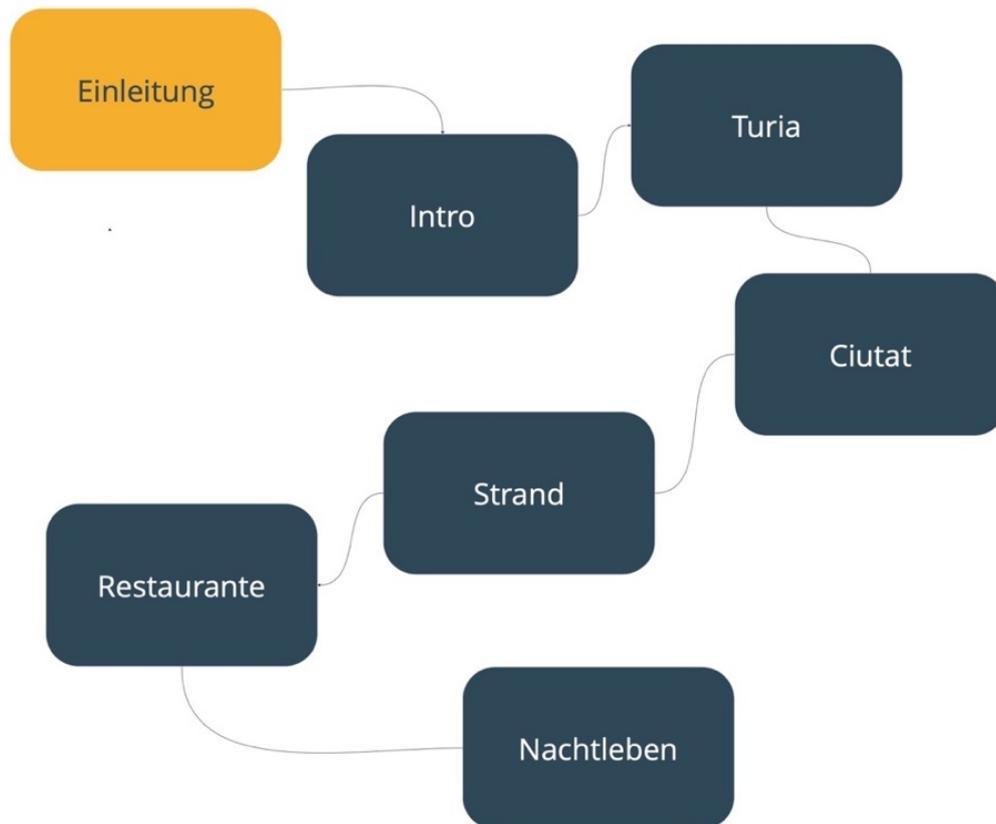


Abbildung 7, Beschreibung des Pfades durch Valencia

Standort

Die Auswahl spezifischer Standorte für die Audiotour erfolgte strategisch, um eine möglichst große Bandbreite an verschiedenen Orten zu ermöglichen.

Es wurden in Valencia sieben Menschen verschiedener Alters- und Geschlechtsgruppen befragt, welche Orte für sie persönlich die wichtigsten sind und welche Orte sie Tourist*innen empfehlen würden.

Fünf empfahlen die Turia Gärten, drei hielten sich am liebsten Strand auf und andere fünf davon würden *Russafa* und *El*

Carmen für die Abendgestaltung empfehlen.

Die *Turia*-Gärten (Abbildung 8, 1) dienen als Ausgangspunkt für eine dynamische Atmosphäre, wo die *Ciutat de les Arts i les Ciències* (2) kulturelle und architektonische Aspekte repräsentiert. Der Strand von Valencia (3) zeigt den charakteristischen Lebensstil der Stadt am Meer, und Orte wie das *Russafa*-Viertel (5) und *El Carmen* (4, 5) spiegeln die kulinarische und nächtliche Szene wider.

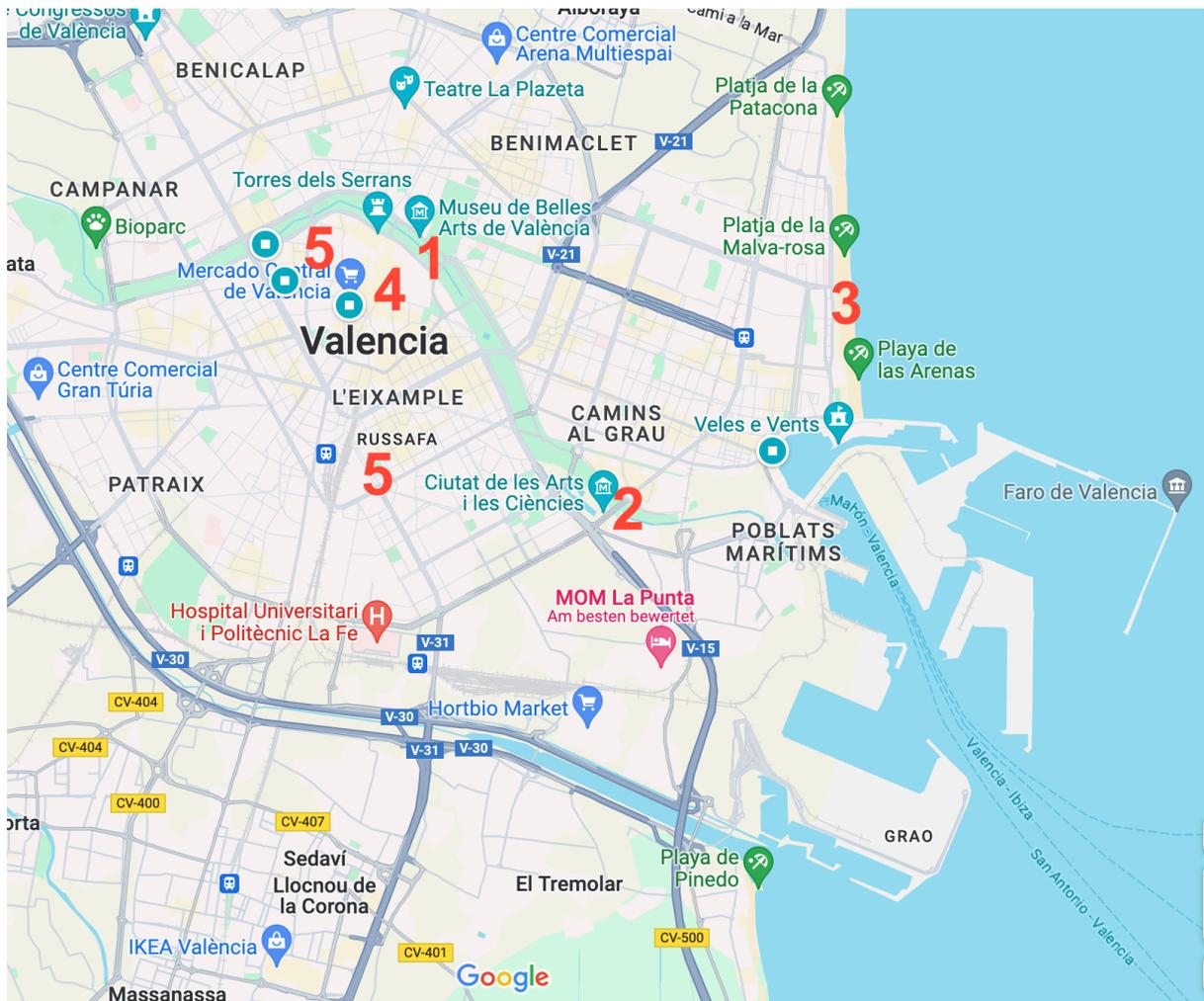


Abbildung 8, Die Standorte auf der Karte, <https://www.google.at/maps/@39.407669,-0.5263212,11z?entry=ttu>, 12.01.2024

Um die angedachten Schauplätze zu bestätigen, wurden Mikrofonierungstests an den Standorten durchgeführt. Diese Tests beeinflussten den Designprozess, indem sie die Klanganforderungen an die verschiedenen Umgebungen identifizierten, wie etwa unerwünschte Hintergrundgeräusche, um eine kohärente und lebhaftere Audioerfahrung zu gestalten.

Die Auswahl der richtigen Klangelemente und die Optimierung der räumlichen Audiodarstellung erfolgte ebenfalls durch mehrere Testläufe.

Um die Immersion der Audiotour zu verbessern, wurden an den Schauplätzen Notizen gemacht, welche Sounddesignelemente die Szenerie mitgestalten können. Im Turia-Garten wurde im Produktionsprozess zum

Beispiel das Geräusch von Radfahrern hinzugefügt, um die lebhafte Atmosphäre des Parks einzufangen und den Eindruck einer aktiven Freizeitgestaltung zu verstärken.

In den Restaurants wurden gezielt Geräusche von klirrenden Gläsern und gedämpftem Gesprächslärm eingefügt, um die kulinarische Erfahrung authentischer zu gestalten.

Die Weite wurde durch Musik aus der Umgebung betont, insbesondere in den Restaurants und Nachtclubs im Ruzafa-Viertel.

Um Benutzer*innen einen einfachen Einstieg in das Erlebnis zu bieten, wurde ein physischer Flyer mit einer Anwendungsbeschreibung erstellt (siehe Abschnitt 4.4).

4.2 Mikrofonierung

Die Mikrofonierung von räumlichen Aufnahmen ist ebenso vielfältig wie deren Abspielmöglichkeiten.

Das *Los Senderos Studio* in Texas, USA hat eine Auflistung verschiedener Mikrofonierungstechniken erstellt und unterscheidet grundsätzlich in Front-Arrays, Ambient-Arrays und Spatial Microphones (*Los Senderos Studio*, 2022). Eine ganzheitliche Erfassung aller existenten Aufnahmetechniken würde an dieser Stelle den Rahmen sprengen, deshalb wurde eine relevante Auswahl getroffen.

Front Arrays

Front Arrays repräsentieren im Wesentlichen stereophone Mikrofontechniken, die dazu dienen, das Hauptstereobild abzubilden. (*Los Senderos Studio*, 2022)

Decca-Tree

Die Decca-Tree-Methode ist eine stereophone Mikrofontechnik, die in den 1950er Jahren in den *Decca Studios* in London entwickelt wurde und seit jeher hauptsächlich zur Stereoaufnahme von Orchestern verwendet wird (Hannah, 2021).

In Zeiten der Surroundaufnahme hat die Decca-Tree Anordnung eine ungeahnte Rückkehr erlebt. Sie kann in

verschiedenen Variationen aufgestellt werden, wobei eine populäre Version wie folgend dargestellt wird.

Hierbei kommen drei omnidirektionale Mikrofone in einer Dreiecksanordnung zum Einsatz.

Zwei Mikrofone werden links und rechts etwa 2 m voneinander entfernt platziert, wobei das dritte Mikrofon in der Mitte etwa 1 m vorne und etwa 2,5 m bis 3 m hoch angeordnet wird. Die seitlichen Mikrofone werden hart links und rechts gepannt, während das dritte Mikrofon genau in der Mitte gemischt wird. Diese Technik erzeugt ein ausgeprägtes Stereobild und wird gelegentlich auch als A-B-C-Stereophonie bezeichnet (Gernemann, 2002).

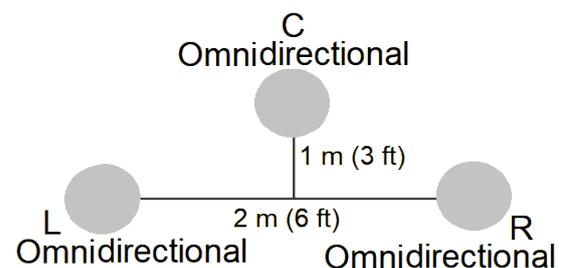


Abbildung 9, Decca Tree, <https://lossenderosstudio.com/article.php?subject=17>, 18.08.2023

Optimized Cardioid Triangle (OCT)

Die OCT-Technik, konzipiert von Günther Theile und Helmut Wittek, für das Front-Array, besteht aus Dreiecksanordnung aus drei Mikrofonen. Dabei befindet sich das zentrale Nieren-Mikrofon etwa 8 cm vor den beiden seitlichen Supernierenmikrofonen. Die seitlichen Mikrofone sind nach außen ausgerichtet und weisen Abstände von etwa 40 cm bis 100 cm zueinander auf, wobei die Aufnahmewinkel zwischen 160° und 90° variieren.²⁵ Diese Struktur resultiert in einer effektiven Trennung zwischen den Sektoren L/C und R/C (*OCT Sets | Sets für Surround-Aufnahmen | SCHOEPS Mikrofone*, o. J.).

Schall, insbesondere von der rechten Seite, wird von der linken Superniere nur minimal erfasst. Im Extremfall, wenn der Schall exakt von rechts kommt, erfolgt die Hauptübertragung durch die rechte Superniere, begleitet von einer geringeren Übertragung durch das Nierenmikrofon, jedoch mit einem um 6 dB abgeschwächten Pegel und zeitlicher Verzögerung. Das Signal erreicht die linke Superniere noch später mit etwa 10 dB weniger Pegel. Zusätzlich überträgt die rückwärtige Empfindlichkeitskeule der Superniere gegenphasige Signale. Diese Aspekte eliminieren die Entstehung irritierender Phantomschallquellen einer

rechten Schallquelle im linken Wiedergabesektor. Das Ergebnis dieser Anordnung ist ein klarer Center-Kanal, da frontaler Schall vorwiegend von der zentralen Niere aufgezeichnet wird, während die seitlich gerichteten Supernieren ihn unter einem Winkel von 90° aufnehmen und folglich bei gleicher Empfindlichkeit etwa 10 dB weniger Pegel liefern. (Wittek & SCHOEPS, 2012)

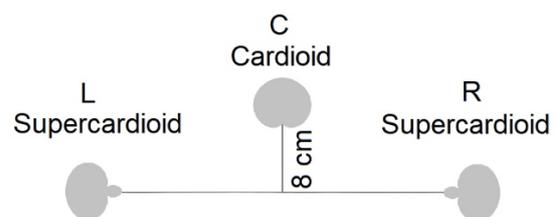


Abbildung 10, OCT-Mikrofonierung, <https://lossendersstudio.com/article.php?subject=17>, 18.08.2023

Nachteile von drei-Mikrofon-Techniken

Während Drei-Mikrofon-Techniken sich für traditionelles Zwei-Kanal-Stereosound gut eignen, entsteht ein Problem bei der Umwandlung in die drei Kanäle (L-C-R), die im Bereich von Surround-Sound verwendet werden. Bei der Wiedergabe einer Dreimikrofonaufnahme über Lautsprecher kann eine "dreifache Phantom-Schallquelle" entstehen. Aufgrund von kanalübergreifendem Crosstalk erzeugt jedes Zwei-Kanal-Paar (C-L, C-R und L-R) ein unterschiedliches Phantom-Bild, das sich mit der Änderung der Hörposition zu bewegen scheint, was

²⁵ Aufnahmewinkel können hiermit berechnet werden: <https://schoeps.de/wissen/image-assistant.html>. 10.01.2024

zu einem Verlust von Fokus und Klarheit führt (Theile, 2001).²⁶

Ambience Arrays

Diese Arrays wurden entwickelt, um die Atmosphäre in einem Raum einzufangen. Sie werden normalerweise in Verbindung mit einem, der oben beschriebenen Front-Arrays verwendet (Los Senderos Studio, 2022).

Doppel ORTF

Doppel-ORTF ist eine Raumklangtechnik, die auf der ORTF (*Office de Radiodiffusion-Television Francaise*) Stereotechnik basiert. Es handelt sich um zwei ORTF-Konfigurationen, die Rücken an Rücken angeordnet sind.

Das Doppel-ORTF besteht aus vier Mikrofonen mit Nierencharakteristik, die in einer rechteckigen Formation angeordnet sind, wobei das vordere und die hintere Anordnung in einem Winkel von 110° zueinanderstehen und 17 cm voneinander entfernt sind. Die hintere Anordnung befindet sich 25 cm hinter der vorderen Anordnung. (Czyzewski et al., 2001).

Eine Doppel-ORTF-Aufnahme eignet sich als Stützmikrofon und für Atmoaufnahmen. (Wittek, 2013)

Wie später in diesem Kapitel beschrieben, wurde die Doppel-ORTF Anordnung als Hauptmikrofonaufstellung für dieses Projekt gewählt.

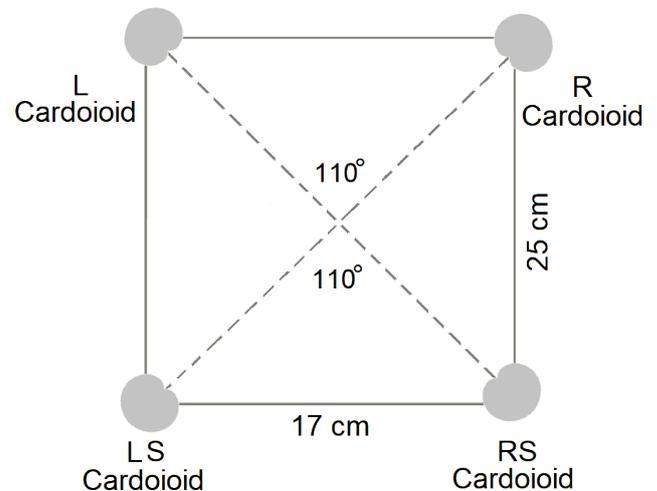


Abbildung 11. Doppel-ORTF, <https://lossenderosstudio.com/article.php?subject=17>, 18.08.2023

IRT-Cross

Der IRT-Cross (auch "Atmo-Cross") ist eine, vom deutschen IRT (Institut für Rundfunktechnik) entwickelte Raumklangtechnik. Es besteht aus vier Mikrofonen mit Nierencharakteristik, die in einer quadratischen Anordnung in einem Winkel von 90° zueinander und 45° von der Schallquelle entfernt in einem Abstand von 20 cm bis 25 cm aufgestellt sind. Der Hauptnachteil dieses Ansatzes ist, dass ein erheblicher Anteil an Direktschall vorhanden ist. Die Technik ähnelt hintereinander geschalteten, nahezu

²⁶ weitere interessante Mikrofonierungstechniken an dieser Stelle wären das INA-3 oder Ideal Cardioid (ICA), das Doppel-XY und das Doppel-AB Verfahren. Zu finden zum Beispiel in Pfanzagl-Cardone,

E. (2020). Surround Microphone Techniques. In: The Art and Science of Surround and Stereo Recording. Springer, Vienna. https://doi.org/10.1007/978-3-7091-4891-4_4

deckungsgleichen Stereopaaren. Wie der Name schon ahnen lässt, wird das Atmo-Cross hauptsächlich für Atmo-Aufnahmen verwendet (Theile, 2001).

Optimized Cardioid Triangle (OCT)

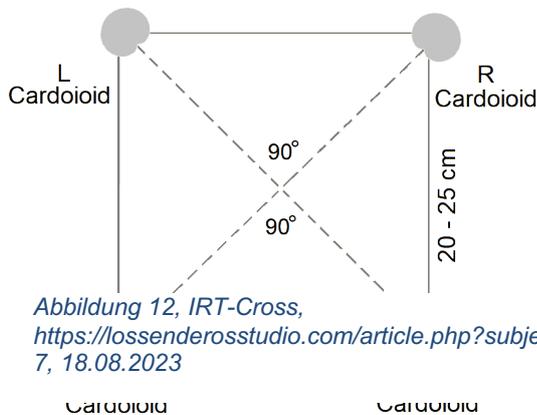


Abbildung 12, IRT-Cross, <https://lossenderosstudio.com/article.php?subject=17>, 18.08.2023

Surround

Die OCT-Surround-Technik verwendet die OCT-Technik für die vordere Anordnung sowie zusätzliche Surround-Kanal-Mikrofone.

Die vordere Anordnung ist darauf ausgelegt, Übersprechen zu minimieren, indem für die vorderen linken und rechten Mikrofone Supernieren im Abstand von 40 cm bis 100 cm verwendet werden, unter einem Winkel von 90° zum zentralen Nieren-Mikrofon platziert, das etwa 8 cm nach vorne versetzt ist.

Zwei zusätzliche hinter dem vorderen Bereich platzierte Nieren-Mikrofone befinden sich etwa 40 cm hinter der

vorderen Linie und weitere 20 cm dahinter, gerichtet zur Rückseite.

Die Benutzer*innenanleitung beschreibt zu diesen folgendes Verhalten:

„Die Nieren weisen nach hinten, um die Aufnahme von Direktschall zu unterdrücken. Die Laufzeit- und Intensitätsunterschiede zwischen Superniere und Niere sind so bemessen, dass der Seitenschall in ähnlicher Weise wie der Schall von vorne jeweils stereofon aufgenommen wird. Das heißt, wenn sich die Zuhörer*innen zur Seite drehen, stimmt die Abbildung auch in den Sektoren zwischen L und LS bzw. R und RS. Diese korrekte Wiedergabe der seitlichen Reflexionen führt zu einer guten Reproduktion der räumlichen Perspektive.“

(OCT-Sets | Sets für Surround-Aufnahmen | SCHOEPS Mikrofone, o. J.)²⁷

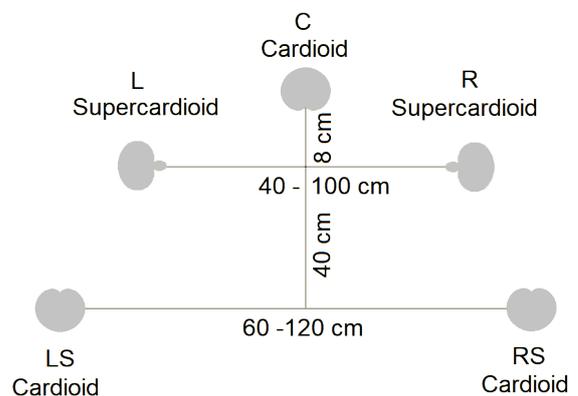


Abbildung 13, OCT-Surround, <https://lossenderosstudio.com/article.php?subject=17>, 18.08.2023

²⁷ Weitere nennenswerte Mikrofonsysteme wären an dieser Stelle der Hamasaki Square, der Fukada Tree, Double MS, das Polyhymnia Pentagon, welches mit 5 Kugelmikrofonen in

ITU-R BS775 Surround Norm aufgestellt ist und das Perception Control Microphone Array (PCMA).

Diese kombinierten Systeme versprechen teilweise äußerst authentische Reproduktionsergebnisse, doch konnte aufgrund der im vorhinein begrenzten Anzahl von vier Mikrofonen und dem erheblichen logistischen Aufwand in dieser Arbeit nicht näher damit experimentiert werden.

Tetraeder-Mikrofon

Tetraeder oder erste Ordnung Ambisonics Mikrofone sind Mikrofone, die in Form eines Tetraeders gestaltet sind und zeichnen sich durch ihre Fähigkeit aus, Schallsignale in einer sphärischen Form zu erfassen. Die alternative Bezeichnung, "First Order Ambisonics Microphone" verweist auf die erste Ordnung Ambisonics wie in Abschnitt 2.4 beschrieben. Tetraeder-Mikrofone bieten eine räumliche Darstellung von Schallereignissen und ermöglichen eine immersive Wiedergabe von Audioinhalten.

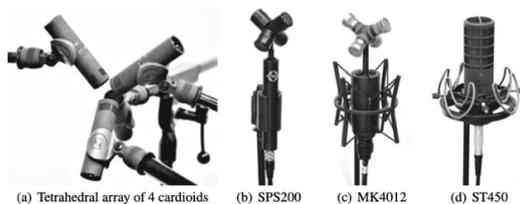


Abbildung 14, Verschiedene First-Order Ambisonics Mikrofone, Zotter, F., & Frank, M. (2019). *Ambisonics: A practical 3D audio theory for*

Das Mikrofon erzeugt vier Signale, 4x Niere, ausgeführt als Ambisonics A-Format (Sennheiser, 2023):

- 1: Front Left Up (FLU)
- 2: Front Right Down (FRD)
- 3: Back Left Down (BLD)
- 4: Back Right Up (BRU)

Mithilfe einer geeignete Soft- oder Hardware²⁸, wird das A-Format über eine Matrix in das sogenannte B-Format, umgewandelt (Sennheiser, 2023):

- W = FLU + FRD + BLD + BRU
- X = FLU + FRD – BLD – BRU
- Y = FLU – FRD + BLD – BRU
- Z = FLU – FRD – BLD + BRU

First-Order Mikrofone stehen oft in Kritik, da wie in Abschnitt 2.4 beschrieben durch den kleinen Sweetspot kaum eine hochauflösende Darstellung möglich ist und das Klangbild gerne als „phasig“ und unscharf beschrieben wird.

²⁸ die oft dem Produkt mitgeliefert wird, wie in etwa bei <https://www.sennheiser.com/de-de/catalog/products/mikrofon/ambeo-vr-mic/ambeo-vr-mic-507195>), 11.01.2024

Zylia

Zylia ist eine Ambisonics-Mikrofonanordnung und nimmt Ambisonics in bis zur dritten Ordnung auf. Das Mikrofon besitzt laut Herstellern 19 Micro-Electromechanical-System (MEMS) Mikrofone mit einer digitalen Schnittstelle (*ZYLIA PRO - 3rd Order Ambisonics Microphone Array with Software*, 2023). MEMS ist eine Miniatur-Mikrofontechnik, die direkt auf Platinen platziert werden kann und ist mittlerweile standardisiert in den meisten Konsumentenprodukten wie etwa Smartphones verbaut wird (Walser, 2017).

Das Zylia wäre, allein schon aufgrund der handlichen Abmessungen von 155mm Höhe mal 103mm Länge (*ZYLIA PRO - 3rd Order Ambisonics Microphone Array with Software*, 2023) eine optimale Wahl zum einfachen Transport durch die Stadt gewesen und da das Projekt in der dritten Ordnung Ambisonics angelegt wurde eine barrierefreie Lösung. Aufgrund von Verfügbarkeit und einer Budgetbarriere war es nicht möglich darauf zuzugreifen.



Abbildung 15, Zylia PRO, <https://www.zylia.co/zylia-pro.html>, 13.09.2023

ORTF-3D

ORTF-3D ist eine 3D Mikrofon-Anordnung, die 8 Studiomikrofone mit supernierenförmiger Richtcharakteristik verwendet. Es handelt sich im Wesentlichen um eine Verdoppelung des "ORTF Surround"-Systems auf zwei Ebenen. Jede Ebene besitzt vier Supernieren (oben und unten), die Rechtecke mit Seitenlängen von 10 und 20 cm bilden. Die beiden "ORTF Surround"-Anordnungen werden direkt übereinander platziert. Die Mikrofone sind zusätzlich nach oben oder unten geneigt, um eine Signalabtrennung in der vertikalen Ebene zu erzeugen. Auf diese Weise entsteht eine 8-Kanal-Anordnung mit einer Abbildung in der horizontalen Ebene, die in etwa dem "ORTF Surround"-System entspricht. Die Mikrofonsignale werden diskret auf vier Kanäle für die untere Ebene (L, R, LS, RS) und vier für die obere Ebene (Lh, Rh,

LSh und RSh) geleitet (Wittek, 2016).²⁹

3031



Abbildung 16, ORTF-3D,
<https://schoeps.de/produkte/surround-3d/ortf-3d/ortf-3d-indoor-set.html>, 20.11.2023

Mikrofonwahl

Für die Mikrofonierung des Projekts wurde ein Doppel-ORTF-Anordnung gewählt.

Diese Entscheidung war eine Folge mehrerer Faktoren. Da die Aufnahmen während eines Auslandsaufenthaltes in Valencia, in einem Campingbus wohnend geschahen, musste der Aufbau platzsparend zusammenpackbar sein.

Zum anderen ist bei längeren Aufnahmesitzungen, bei denen man sich oft mehrere Kilometer zu Fuß zu verschiedenen Aufnahmeorten bewegt, ein geringes Gewicht von Vorteil. Eine

weitere Einschränkung war die Verfügbarkeit von Mikrofonen.

Das finale Mikrofondepot bestand aus 4 Kleinmembrankondensator, einem eigengebautem Binaural-Kopf mit *Roland CS-10 EM in-Ear* Mikrofonen und einem *Sennheiser VR Mic Tetraeder*-Mikrofon. Als Aufnahmegerät wurde das *Zoom F6*³² benutzt.

Es wurde ein Gerüst konzipiert, dass aus drei Schienen besteht, wobei jeweils zwei identische Mikrofonpaare *Neumann KM 184*³³ für den vorderen und *Lewitt 040*³⁴ für den hinteren Bereich verwendet wurden.

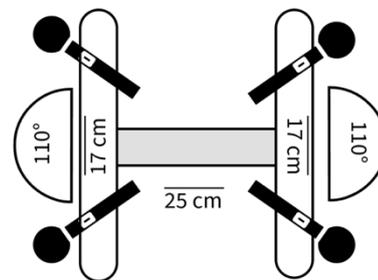


Abbildung 17, Doppel-ORTF Aufbau

Auf den nächsten Seiten folgt eine Auswahl von Bildern, die den Konzeptionsprozess und Aufnahmeprozess dokumentieren.

²⁹ Für weiteres Interesse an 3D Mikrofonierungstechniken wird ein Blick auf folgendes öffentliches Paper empfohlen in dem verschiedene Techniken verglichen werden: <https://www.aes.org/e-lib/browse.cfm?elib=21013> , 11.11.2023

³⁰ Für einen Aussagekräftigen Vergleich einiger dieser Techniken empfiehlt sich die Internetseite und folglich Berichte: <https://www.hauptmikrofon.de/>, 14.01.2024

³¹ Auch wenn es scheint, dass die Mikrofone in Abbildung 16 in die gleiche Richtung deuten,

richtet sich die obere Ebene vorrangig nach oben und ist bei genauerer Betrachtung an der Spitze geschlossen.

³² <https://www.zoom-europe.com/de/field-recorder/zoom-f6> , 13.11.2023

³³ <https://www.neumann.com/de-de/produkte/microphones/km-184-series-180/> , 07.09.2023

³⁴ <https://www.lewitt-audio.com/de/mikrofone/lct-recording/lct-040-match-stereo-paar> , 07.09.2023



Abbildung 18, Links oben: ORTF-Test, oben mitte: Sennheiser AMBEO VR Tetraeder-Mikrofon, rechts oben: DIY Binauralkopf, unten: Kombination aus Doppel-ORTF und Binauralkopf



Abbildung 19, links oben: Mikrofonierung in Granada, mitte oben: Mikrofonierung in Valencia (las Arenas), rechts oben: Test mit Höhenkanälen in der Markthalle von Valencia, rechts unten: Mikrofonierung des Osterumzugs in Granada, mitte unten Mikrofontest in der Caló d'en Monjo, Mallorca, rechts unten: Mikrofonierung im botanischen Garten Valencia,

In den Abbildungen 18,19, wird deutlich, dass verschiedene Mikrofon-Aufbauten getestet wurden. Unter den Optionen befand sich auch ein selbstgebauter binauralen Kopf, bestehend aus einem Styropor-Puppenkopf und Silikon-Ohrabdrücken, in denen zwei Elektretmikrofone verbaut sind. Dieser wurde jedoch aufgrund der, als zu unflexibel empfundenen, binauralen Stereofilen, ausgeschlossen. Ebenso wurde das *Sennheiser AMBEO VR* aufgrund der undeutlichen Auflösung (siehe Abschnitt 2.4) von erste Ordnung Ambisonics nicht verwendet. Zur Aufnahme der Atmosphären blieben schließlich der IRT-Cross und eine Doppel-ORTF-Anordnung über. Die beiden Aufstellungen unterscheiden sich nur gering, aber vor allem im Aufnahmewinkel. Der Mikrofonwinkel von 90° auf einer Seite des IRT-Cross bildet laut dem *Schoeps Image-Assistant*³⁵ einen maximalen Aufnahmewinkel von 91° ab (Abbildung 20,21). Was laut der IRT-Anordnung etwa vier aneinandergereihten Aufnahmen im etwa 90° Winkel entspricht.

Im Gegensatz dazu bildet das ORTF-Verfahren mit einem Öffnungswinkel von 110° einen Aufnahmewinkel von 101° mit einer etwas flacheren Dezibel-Abfall-Kurve zu den Seiten hin, ab (Abbildung 22,23).

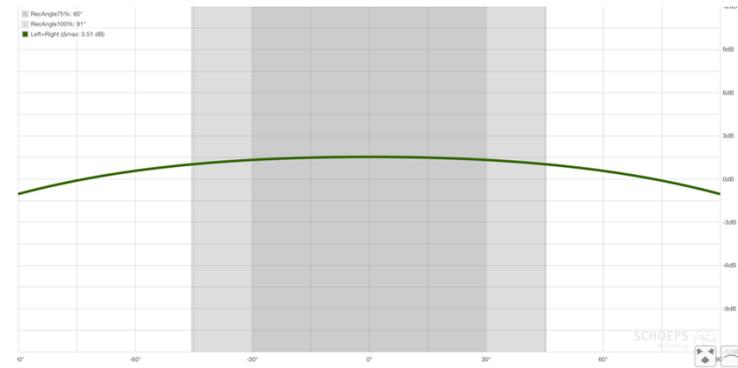


Abbildung 20, IRT-Teil, <http://ima.schoeps.de/>, 28.09.2023

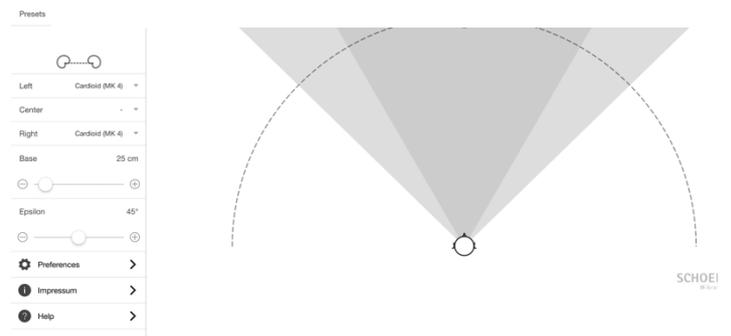


Abbildung 21, IRT-Teil, <http://ima.schoeps.de/>, 28.09.2023

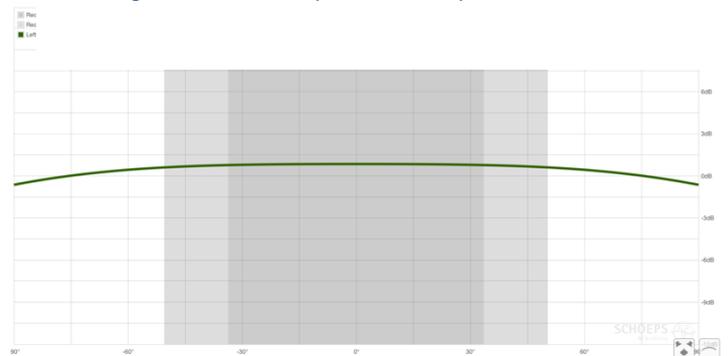


Abbildung 22, ORTF-Teil, <http://ima.schoeps.de/>, 28.09.2023

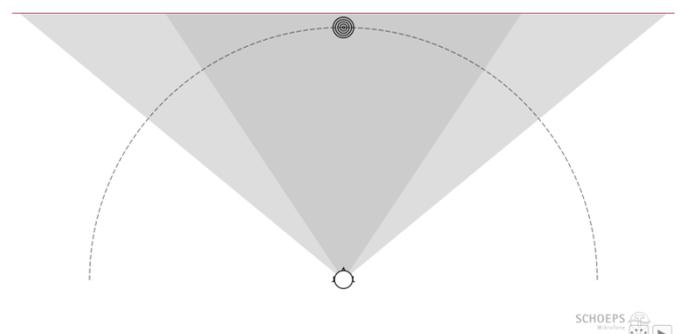


Abbildung 23, ORTF-Teil, <http://ima.schoeps.de/>, 28.09.2023

³⁵ <http://ima.schoeps.de/>, 11.01.2024

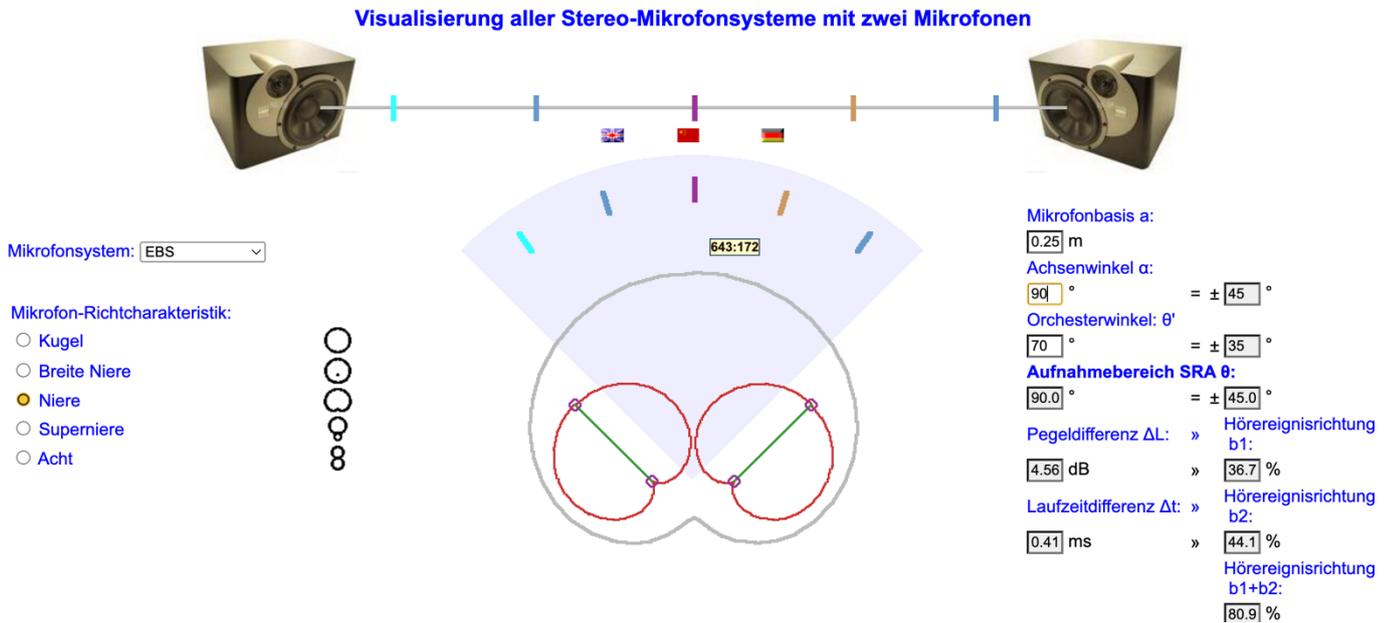


Abbildung 25, IRT-Teil, <http://www.sengpielaudio.com/HejiaD.htm>, 25.10.2023

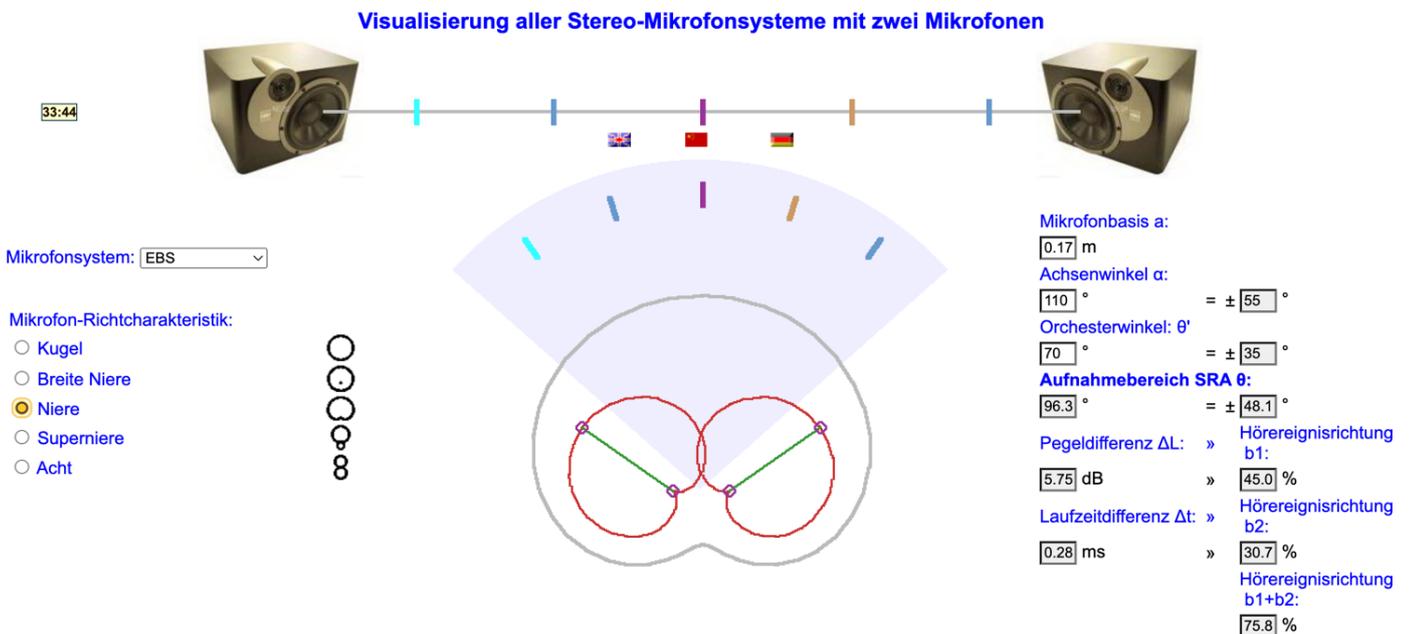


Abbildung 24, ORTF-Teil, <http://www.sengpielaudio.com/HejiaD.htm> 25.10.2023

Laut der Visualisierung von Sengpiel (Abbildung 24,25) bilden sowohl der 90° Stereo und der 110° ORTF-Aufbau ein relativ getreues Bild in der Stereowiedergabe ab.

Da beide Aufbauten im subjektiven Hören über Kopfhörer ebenfalls ein breites und offenes Klangbild wiedergaben, dass sich passend für Atmo-Aufnahmen eignet (Wuttke, 2002), fiel die Entscheidung schließlich aus praktischen Gründen

aufgrund der beiden verfügbaren Mikrofonreihen. Diese waren mit 20 cm zu kurz für die IRT-Anordnung, aber genau passend für den 17cm Abstand der ORTF-Aufnahme waren.

4.3 Produktion, Postproduktionsprozess und zusätzliche Werkzeuge

In einem 3D-Audioprojekt, das mit verschiedenen Audioformaten arbeitet, ist die Einrichtung einer geeigneten Arbeitsumgebung notwendig. Als Digital Audio Workstation (DAW) erweist sich für dieses Vorhaben das Programm *Reaper*³⁶, aufgrund der vielfältigen Multichannel-Routingmöglichkeiten, als geeignet. Die Anwendung von Higher-Order-Ambisonics im Produktionsprozess dieses Projekts birgt mehrere vorteilhafte Aspekte. Einerseits ermöglicht es die Enkodierung sämtlicher Formate in den Ambisonic-Stream, andererseits erfolgt die Dekodierung des „Ambisonic-Betts“ im Nachhinein, mittels geeigneter Decoder in die gewünschte, höchstmögliche Lautsprecherkonfiguration. Des Weiteren ermöglichen Werkzeuge, wie die frei verfügbare *IEM-Plugin-Suite*³⁷, eine Bearbeitung und Positionierung von Elementen im Raum

Für den abschließenden Authoring-Prozess sollten die verschiedenen

vorgesehenen Elemente im Vorhinein identifiziert werden. Das in Abschnitt 4.1 beschriebene Design sieht vor, dass es zwei eigenständige Sprecher*innen und ein umhüllendes Klangbett geben soll. Zusätzlich kann der Musiktrack bei Bedarf weg- oder hinzugeschaltet werden.

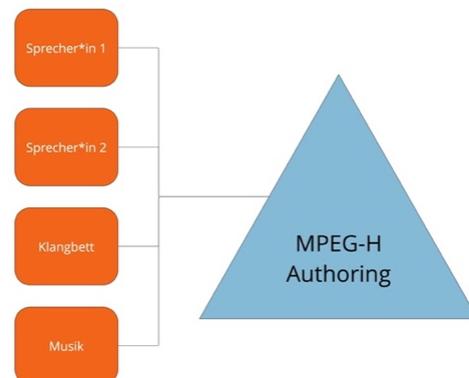


Abbildung 26, Übersicht über die Gruppen für den Authoringprozess

Der Produktionsablauf bestimmt, dass das Klangbett in einem dedizierten Ambisonics-Workflow erstellt und dann im Authoring-Prozess als Audioobjekt gehandhabt wird. Dadurch ist es Endbenutzer*innen möglich, die gesamte Szene zu rotieren, als würden sie sich in ihr bewegen.

³⁶ <https://www.reaper.fm/>, 14.09.2023

³⁷ <https://plugins.iem.at/>, 12.11.2023

Die Sprecher*innen werden separat als Audioobjekte in Switchgroups behandelt. Weiters wird die Musik auch als Audioobjekt integriert, um Endbenutzer*innen ein möglichst barrierefreies Erlebnis zu ermöglichen.

Im Nächsten Schritt erfolgt die Sichtung und Auflistung des Materials, um die vorhandenen Audioformate zu identifizieren und eine sinnvolle Gruppierung vorzunehmen.

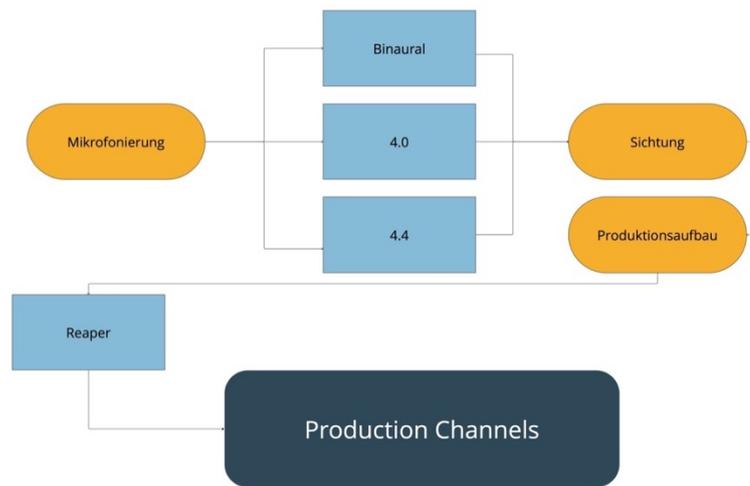


Abbildung 27, Produktionsablauf

Name	Konfiguration	Anzahl Kanäle
Mono	1.0	1
Stereo	2.0	2
Doppel ORTF	4.0	4
Doppel ORTF + Höhe	4.0+4H	8
ORTF 3D	4.0+4H	8
Ambisonics 3. Ordnung	/	16

Tabelle 1, Aufnahmen und ihre Konfiguration

Das Material aus den Aufnahmen besteht aus Doppel-ORTF-Aufnahmen (4.0) beziehungsweise Doppel-ORTF-Aufnahmen mit Höhenkanälen (4.0+4H). Ergänzend können ORTF 3D-Aufnahmen zum Einsatz kommen, die ebenfalls die 4.0+4H Konfiguration aufweisen. Das Projekt wird in Ambisonics dritter Ordnung

festgelegt, um einerseits die Endbenutzerfreundlichkeit der Kanalausgänge gemäß Abschnitt 2.4 zu bewahren und andererseits die Vorgaben einer später im Text beschriebenen Drittsoftware einzuhalten. Die Wahl der dritten Ordnung würde weiters auch die unkomplizierte Nutzung kommerzieller

HOA-Mikrofone wie dem *Zylia* ermöglichen (*ZYLIA PRO - 3rd Order Ambisonics Microphone Array with Software*, 2023). Wenngleich eine höhere Auflösung wünschenswert wäre, wird darauf verzichtet, da es sich hauptsächlich um ein diffuses, umhüllendes Bett handelt, bei dem eine präzise Lokalisierung, wie sie bei einer höheren Ambisonics Ordnung der Fall ist, keinen zwingenden Vorteil darstellt. (Rudrich, 2017)

Produktionsaufbau

In *Reaper* werden sinngemäße Gruppen erstellt und diese nach dem Schema in Abbildung 28 geroutet.

Für die Gestaltung von klangrelevanten Sounddesign-Objekten wie Foleys oder Effekten ist es wichtig, dass diese einzelne Spuren bilden und somit einzeln an einen Encoder auf ihrer Spur übermittelt werden, um die maximale Freiheit in der Automation zu gewährleisten.

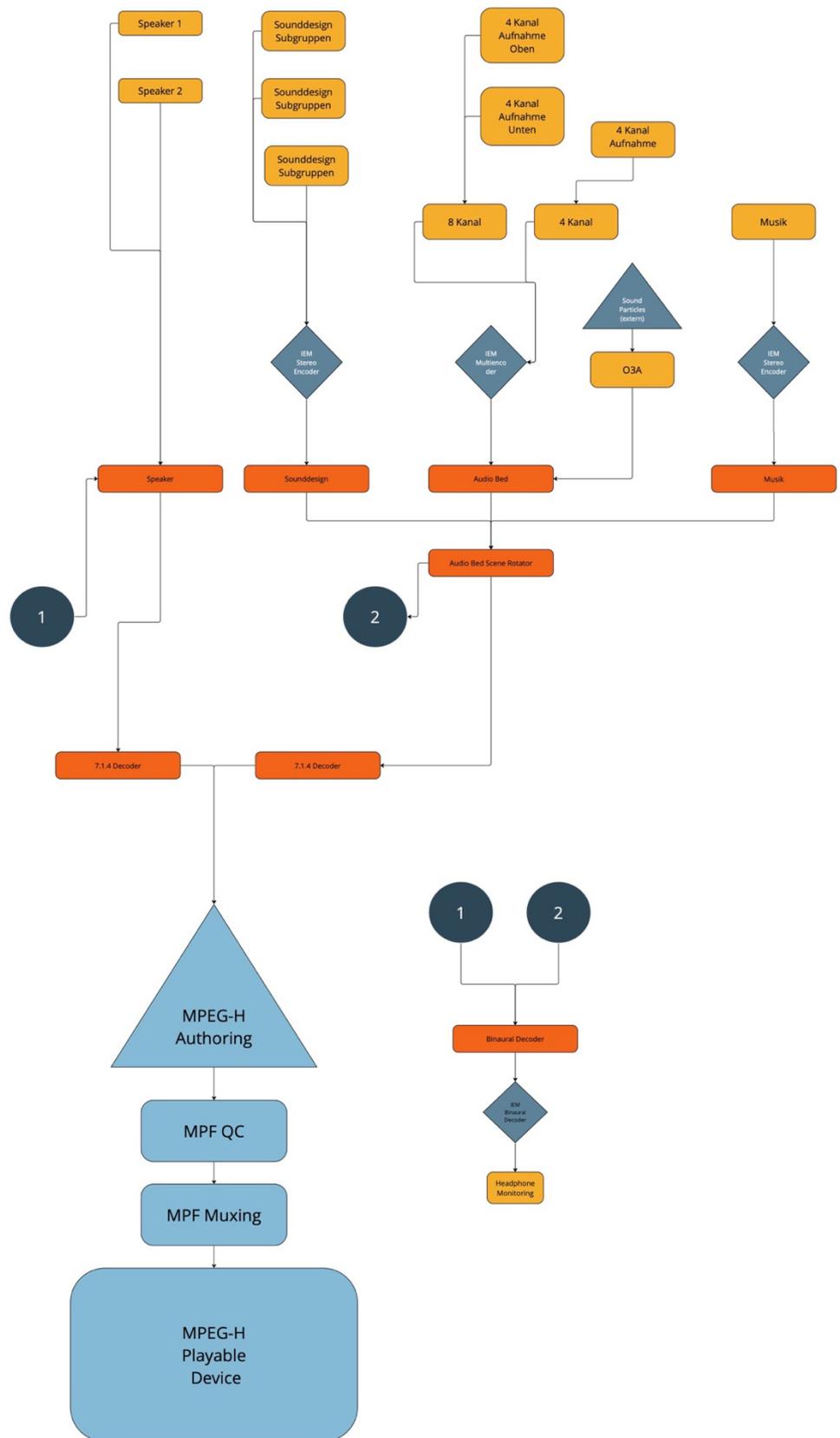


Abbildung 28, Produktionsaufbau

Wie in Abbildung 28 ersichtlich gibt es abseits des Authoring-Busses noch einen zweiten Kanalweg (gekennzeichnet mit 1, 2), der zum *Binaural Decoder* führt.

Dieser dient als Kopfhörermonitor während des Produktionsprozesses und binauralisiert die Ausgabe um einen einfachen Monitoring-Ausgang zu ermöglichen (Rudrich, 2018a).

In einem professionellen Produktionsumfeld wäre es wünschenswert, die Produktionsarbeit in einem Kontrollraum auszuführen, der mit einem adäquaten Lautsprechersystem ausgestattet ist.

Aus logistischen Gründen war dies für die vorliegende Masterarbeit jedoch nicht realisierbar.

Zusätzlich kann vor dem *Binaural Decoder* noch eine *Scene Rotator* Einheit gesetzt werden, um die Szene akustisch zu drehen oder beispielsweise Drehungen des Kopfes mithilfe eines Headtrackers zu simulieren (Rudrich, 2018d).³⁸

Fallweise kann es nötig sein, große und komplexe Audioszenen zu gestalten. Dies hat durch eine große Anzahl von Spuren und den damit einhergehenden Encodern auf den Effekt-Spuren einen erheblichen Rechenaufwand zur Folge.

In diesem Projekt wurde zur Bewältigung dieses Problems, beispielsweise bei einer

Szene, in der Schritte von mehreren Personen in einem großen Gebäude simuliert wurden, die Software *SoundParticles*³⁹ verwendet.

SoundParticles repräsentiert ein 3D-Framework, das die visuelle Manipulation von 3D-Objekten im Raum ermöglicht und die Aufnahme dieser Objekte durch individuell festlegbare, virtuelle Mikrofon-Konfigurationen ermöglicht. Hierdurch können komplexe räumlich-akustische Szenen geschaffen werden, bei deren Produktion der visuelle Vorteil einer 3D-Software zur Verfügung steht.

SoundParticles bietet weiters die Möglichkeit, verschiedene Ein- und Ausgabeformate zu bearbeiten, wobei auch individuelle Formate erstellt werden können, um eine optimale Kompatibilität im Produktionsworkflow sicherzustellen.(Fonseca, 2020).

Der Arbeitsvorgang wird im folgenden Abschnitt genauer beschrieben. Für die Gestaltung von klangrelevanten Mono-Objekten wie Foleys oder Effekten ist es wichtig, dass diese eigenständigen Gruppen bilden und einzeln an einen Encoder auf ihrer Spur übermittelt werden, um so die maximale Freiheit in der Automation gewährleisten zu können.

³⁸ für einen Überblick über die verschiedenen Plug-ins der *IEM Plugin Suite* wird folgender Link empfohlen: <https://plugins.iem.at/docs/plugindescriptions/>, 12.01.2024

³⁹<https://soundparticles.com/products/soundparticles/overview>, 14.12.2023
Die Software ist derzeit (Stand Dezember 2023) im akademischen Kontext frei nutzbar

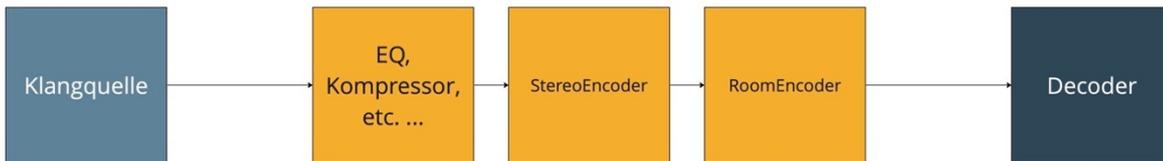


Abbildung 29, Beispiel einer möglichen Effektkette

Spuraufbau

Da für das Projekt die dritte Ordnung Ambisonics festgelegt wurde, müssen die Spuren nun in *Reaper* dementsprechend vorbereitet werden.

Dazu wird die in Abschnitt 2.4 erläuterte Formel $(n+1)^2$ verwendet, um die mindeste Kanalanzahl für jede Spur festzulegen, bei der mit Ambisonics gearbeitet wird. Für die dritte Ordnung ergeben sich so 16 Kanäle. Die Spuren werden anschließend in ihre betreffenden Encoder-Busse weitergeleitet

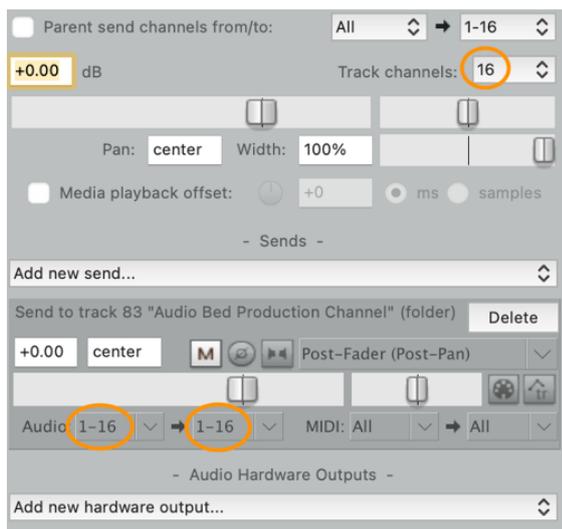


Abbildung 30, Beispiel für die Kanaleinstellung

Encoder

Je nach Audioformat gibt es verschiedene Encoder-Möglichkeiten. Für das Audiobett, das aus vier beziehungsweise acht Kanälen besteht, wurde der

MultiEncoder aus der *IEM-Plugin-Suite* verwendet.



Abbildung 31, Doppel ORTF Multiencoder

Lautsprecherposition	Azimuth	Elevation
Links	45°	0°
Rechts	-45°	0°
Surround Links	135°	0°
Surround Rechts	-135°	0°

Tabelle 2, Doppel-ORTF Positionen

Multiencoder

Die Doppel-ORTF Aufnahme mit acht Kanälen wird mithilfe des *MultiEncoder* auf die vier Aufnahmepositionen (Abbildung 31, Tabelle 2) geroutet und dadurch in den Ambisonic-Stream enkodiert.

Verwendet man ORTF-3D Aufnahmen, werden laut *Schoeps* die betreffenden Positionen wie folgt platziert (Tabelle 3) (SCHOEPS Mikrofone, 2019):

Lautsprecherposition	Azimuth	Elevation
Unten Vorne Links	45°	-45°
Unten Vorne Rechts	-45°	-45°
Unten Hinten Links	135°	-45°
Unten Hinten Rechts	-135°	-45°
Oben Vorne Links	45°	45°
Oben Vorne Rechts	-45°	45°
Oben Hinten Links	135°	45°
Oben Hinten Rechts	-135°	45°

Tabelle 3, ORTF-3D Positionen

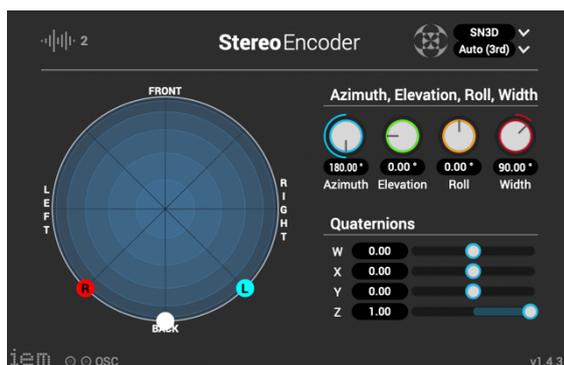


Abbildung 32, Stereo Encoder

Stereoencoder

Für Stereo-Quelldateien stellt die *IEM Plugin-Suite* den *StereoEncoder* bereit (Rudrich, 2018e) Die, in diesem Fall Stereo-encodierten Dateien, sind in dem Projekt hauptsächlich zusätzliche Sounddesignelemente.

Diese können wie in Abbildung 32 ersichtlich über die gesamte Kugelprojektion verschoben werden.

Um ein räumlicheres Gefühl zu erzeugen, befindet sich in der Effektkette einiger Sounddesign-Elemente zusätzlich der Effekt *RoomEncoder*.

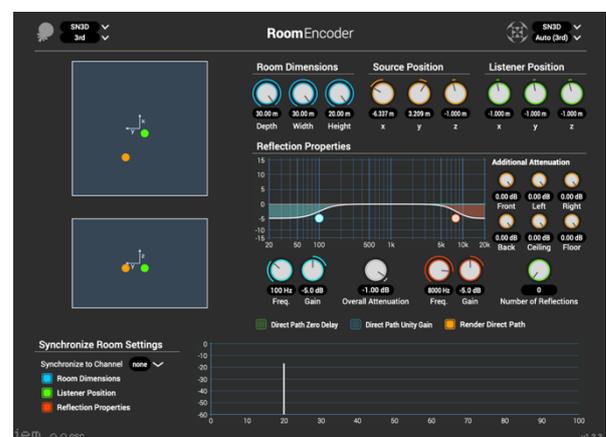


Abbildung 33, RoomEncoder

Mithilfe des *RoomEncoder* kann eine Quelle und eine Hörer*in in einen virtuellen, schuhschachtelförmigen Raum gesetzt und über 200 Wandreflexionen erzeugt werden (Rudrich, 2018c). Über den *RoomEncoder* können monophone Signale (dazu muss in der Ordnungsauswahl die nullte Ordnung gewählt werden) eine frequenzabhängige und zeitvariable Richtwirkung und den

Doppler-Effekt erhalten, was die räumliche Bewegung authentischer gestaltet (Mehra et al., 2014).

Optional kann nach den Encodern noch das Plugin *FDN Reverb* der *IEM Plugin Suite* eingefügt werden. Klänge mit Nachhall werden verstärkt externalisiert der Wiedergabe auf Kopfhörern wahrgenommen (Begault & Wenzel, 2001; Best et al., 2020).

Monitoring

Um ein Monitoring des enkodierten Signals zu ermöglichen, wird in *Reaper* eine weitere Spur erstellt und auf den Effektkanal der *BinauralDecoder* gesetzt. Anschließend können die encodierten Spuren auf den Decoder geroutet werden. Im *BinauralDecoder* kann zudem unter "Headphone Equalization" ein Equalizer-Profil für die meisten gängigen Kopfhörer ausgewählt werden (Rudrich, 2018a).

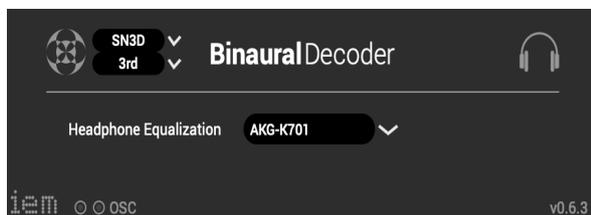


Abbildung 34, Binaurales Monitoring

SoundParticles

Mithilfe von *SoundParticles* lassen sich große Szenerien, mit vielen einzelnen Elementen, in verschiedenen Spatialaudioformaten erstellen.

Im hier beschriebenen Beispiel wurde eine Szene in einem großen Gebäude erstellt, in der verschiedene Menschen sich innerhalb dieses Gebäudes umherbewegen sollten. Die Datei wird in Ambisonics dritter Ordnung ausgegeben, um eine anschließende Einbindung und Verarbeitung in das Klangbett in *Reaper* zu gewährleisten.

Zunächst wurden einzelne Schritte aus Aufnahmen von Fußstapfen auf Betonböden geschnitten und in vier verschiedenen Gangrhythmen zusammengefügt.

Für jede einzelne Spur wird in *SoundParticles* nun ein Audiotrack angelegt (Abbildung 34, 38).

Mithilfe der sogenannten Movement Modifiers (Abbildung 36, 37) lassen sich Bewegungsmuster der einzelnen Spuren im Raum erzeugen, wodurch eine Illusion entsteht, dass sich Menschen im Raum bewegen.

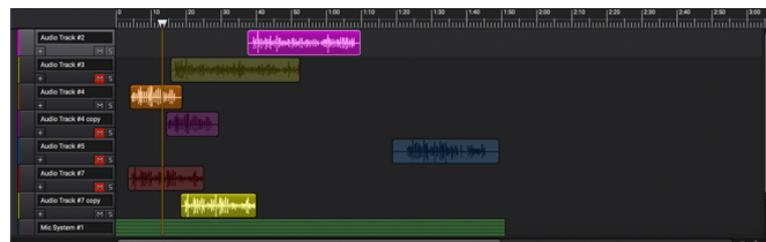


Abbildung 35, Spuransicht in SoundParticles

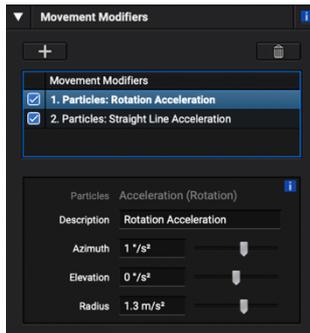


Abbildung 36, Rotationsmodulator

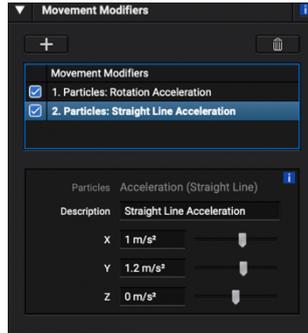


Abbildung 37, Richtungsmodulator

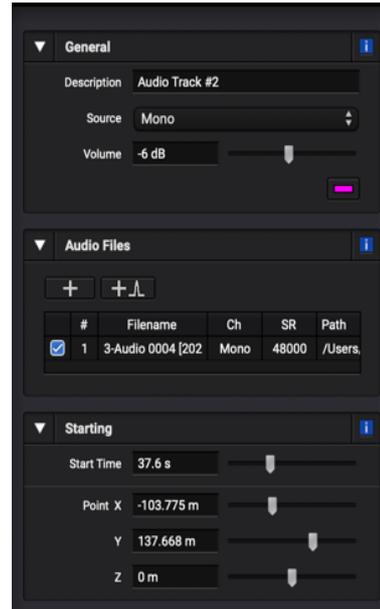


Abbildung 38, Spureigenschaften

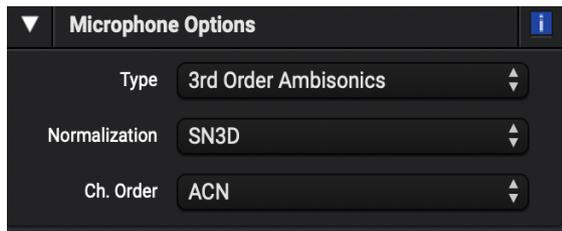


Abbildung 39, Ausgabe nach AmbiX-Konvention

Für jedes weitere Sounddesign-Element muss nun jeweils eine einzelne Mono-Spur der Szene hinzugefügt werden. Die Spuren können anschließend als 3D-Objekt im Raum verteilt und ihre Bewegung automatisiert werden.⁴⁰

Anschließend werden die aufgezeichneten Audioobjekte daraufhin mittels eines virtuellen Mikrofons ausgegeben. Hierbei ist insbesondere auf die Kanalreihenfolge zu achten (Abbildung 39).

Für dieses Projekt wurde ein virtuelles Mikrofon in Form eines Dritter-Ordnung-Ambisonics-Mikrofons ausgewählt (und die Ausgabe erfolgt nach der *AmbiX*-Konvention⁴¹).

⁴⁰ *SoundParticles* bietet, beispielsweise durch Particle-Emitter-Spuren, noch viele weitere Möglichkeiten derartige Szenerien zu erstellen, auf die an dieser Stelle nicht näher eingegangen wird. Dazu wird das offizielle Reference Manual empfohlen: https://www.soundparticles.com/assets/files/sp2_manual.pdf, 12.11.2023

⁴¹ Alle IEM-Plugins nutzen die nach der *AmbiX* Convention gängige Normalisation SN3D und die Channel Order ACN. (Nachbar et al., 2011) (Siehe Kapitel 2.3)

Dekodierung

Nach einem erfolgreichen Sounddesignprozess können alle Spuren nun in ihre sinnvollen Gruppen zusammengeführt und in die passende Lautsprecherkonfiguration dekodiert werden. Eine praktische Lösung um die Ambisonicsdateien dritter Ordnung in eine 7.1+4H Lautsprecherkonfiguration zu übersetzen, bietet der *AllRADecoder* (Rudrich, 2017).

Das Plug-in ermöglicht die Erstellung eines Ambisonic-Decoders für verschiedene Lautsprecherlayouts und Ambisonicsordnungen. Der Decoder wird auf einen Bus am Ende des Ambisonic-Audio-Stream angewendet.

Als ersten Schritt werden die einzelnen Lautsprecher im Raum platziert. Für die 7.1+4H Konfiguration ergeben sich laut Dolby (7.1.4 *Overhead Speaker Setup*, 2023) folgende Werte:

#	Kanal	Azimuth	Elevation
1	Links	-35°	0°
2	Rechts	35°	0°
3	Center	0°	0°
4	Links Surround	-90°	0°
5	Rechts Surround	90°	0°
6	Links Hinten	-135°	0°
7	Rechts Hinten	135°	0°
8	Oben Vorne Links	-35°	45°
9	Oben Vorne Rechts	35°	45°
10	Oben Hinten Links	-135°	45°
11	Oben Hinten Rechts	135°	45°
12	LFE	/	/

Tabelle 4, Lautsprecherpositionen 7.1+4H, in der Norm würde der vierte Kanal den LFE-Kanal abbilden. Da das Routing aber auch in der MPEG-H Authoring Suite frei wählbar ist, wurde dieser vorerst auf Kanal zwölf gelegt.

Die Lautsprecher werden in der 3D-Visualisierung angezeigt (Abbildung 40). Da die Bodenfläche in der Triangulation mehrdeutig interpretiert werden kann müssen imaginäre Lautsprecher hinzugefügt werden, um dieses Problem zu umgehen.

Die Behandlung imaginärer Lautsprecher erfordert bestimmte Überlegungen. Der Gain-Wert für einen imaginären Lautsprecher auf der unteren Ebene sollte auf null gesetzt werden, um zu verhindern, dass Schallquellen von unten, von allen Lautsprechern auf Bodenebene reproduziert werden (M. Frank, persönliche Kommunikation, 4. April 2023).

Für den oberen imaginären Lautsprecher wird ebenfalls empfohlen den Gain auf maximal die Hälfte zu stellen (M. Frank, persönliche Kommunikation, 4. April 2023).

Das Layout in Form einer JSON-Datei, kann für die zukünftige Verwendung oder für die Verwendung mit dem *SimpleDecoder* gespeichert werden.

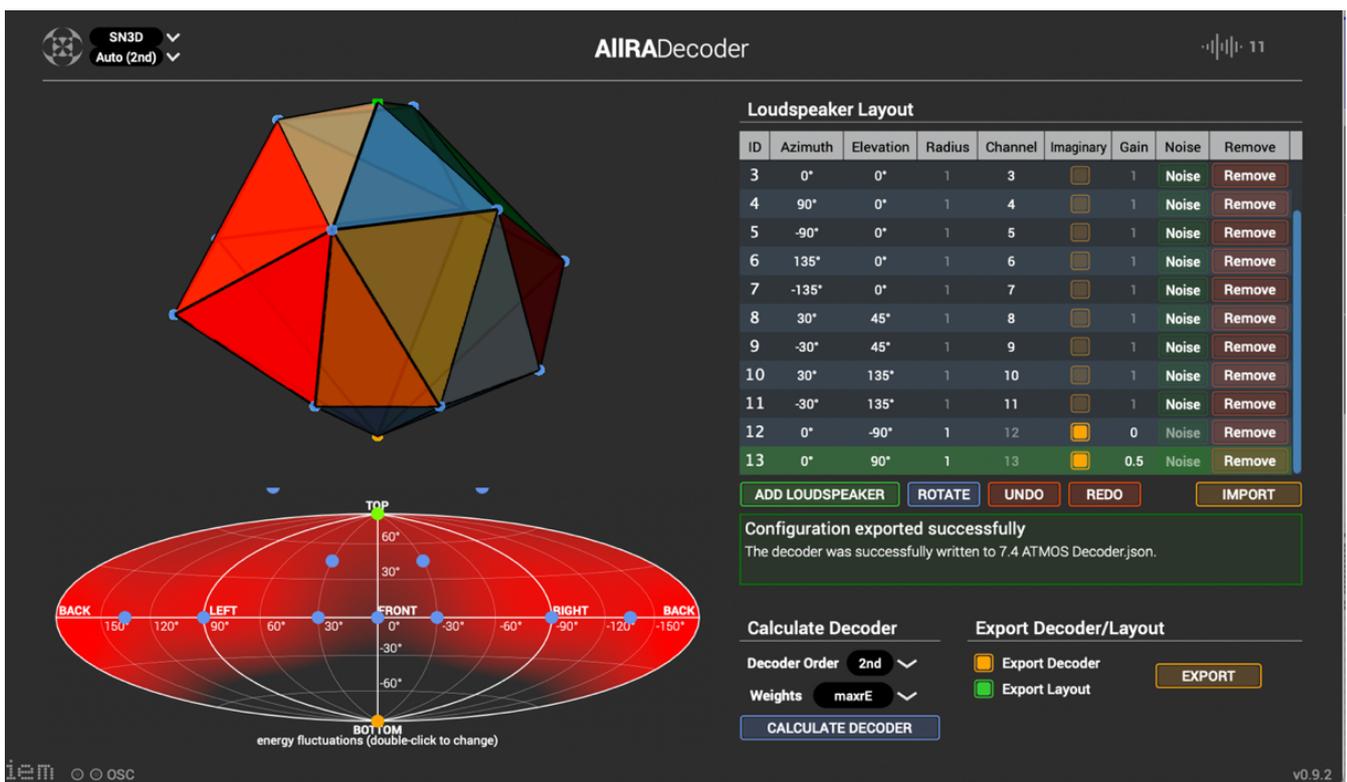


Abbildung 40, AIIRADecoder

MPEG-H Authoring

Nach dem Sounddesignprozess sind nun alle nötigen Gruppen für den Authoring-Prozess bereitgestellt.

Spätestens in diesem Schritt sollten alle Überlegungen zu

Personalisierungsmöglichkeiten und Interaktivität gemacht sein, sodass der Prozess gestartet werden kann.

Der erste Schritt im Authoring-Prozess ist die Auswahl der richtigen

Lautsprecherkonfiguration. Nach entsprechender Auswahl dieser, werden die verschiedenen Spuren gewählt.

wurden die einzelnen Kanäle als Audioobjekte gemappt.

Anschließend werden die zwei Sprecher*innen als Audioobjekt angelegt. In den Parametern kann festgelegt werden, inwieweit Benutzer*innen am Endgerät Änderungen vornehmen können (Abbildung 41). Diese Entscheidung wird hauptsächlich aus persönlicher Erfahrung und subjektivem "Geschmack" der Produzent*innen getroffen, sollte aber sinngemäß hinsichtlich Barrierefreiheit und „sinnvoller Bedienungsmöglichkeiten“ geschehen.

Spur	Konfiguration	Personalisierbar	Switchgroup
Audio-Bett	7.1+4H	Ja (Drehung)	/
Sprecher*in 1	Mono	Ja	Ja
Sprecher*in 2	Mono	Ja	Ja
Musik	Stereo / 7.1+4H	Ja	/

Tabelle 5, Authoring-Gruppen

Für das Audio-Bett wird eine Objektspur angelegt. Im konkreten Fall müssen nun die 12 Kanäle als Audioobjekte an die oben beschriebenen Positionen für eine 7.1.4 Konfiguration gebracht werden. Prinzipiell würde man für das Bett eine Kanalspur mit einer fixen Lautsprecherkonfiguration wählen. Da aber in diesem Fall die Möglichkeit offen sein sollt, das gesamte Bett zu drehen,



Abbildung 41, einstellbare Benutzerinteraktionen für die Audiobett

Da beide Sprecher*innen nicht gleichzeitig zu Wort kommen sollten, gibt es die Möglichkeit eine Switchgroup anzulegen. Switchgroups ermöglichen das einfache Wechseln zwischen Gruppen (Audioobjekten sowie Kanalspuren) ohne, dass beide jemals gleichzeitig wiedergegeben werden.

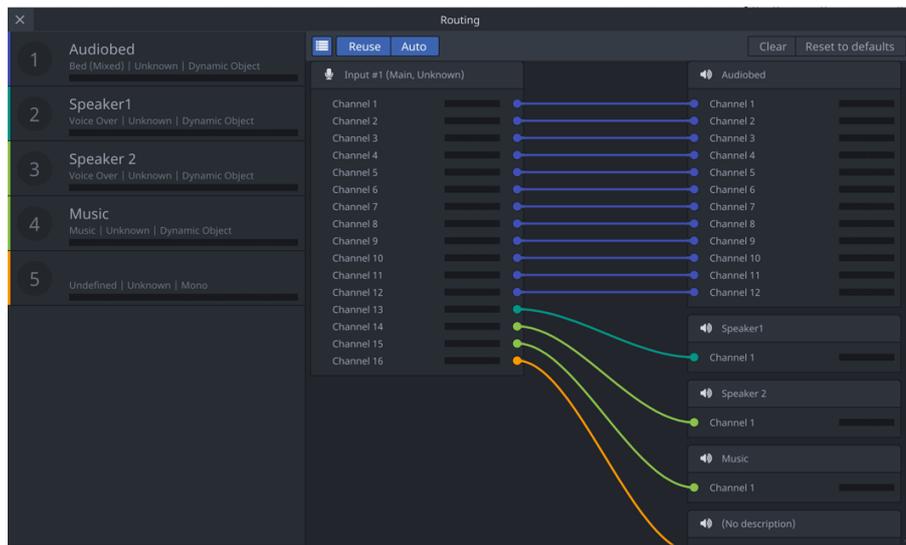


Abbildung 43, Kanalrouting



Abbildung 42, 7.1+4H Mapping auf die einzelnen Audioobjekte

Preset-Name	Spuren	Parametereigenschaften
City-Trip	Sprecher*in 1 + Audio-Bett + Musik	
Kulturreise	Sprecher*in 2 + Audio-Bett + Musik	
City-Trip Dialouge+	Sprecher*in 1 + Audio-Bett	Sprecher*in 1 + 6dB
Kulturreise Dialouge+	Sprecher*in 2 + Audio-Bett	Sprecher*in 2 + 6dB
Atmosphäre+	Audio-Bett und Musik	Atmosphäre + Musik +3dB

Tabelle 6, Preset

Anschließend werden die Presets erstellt, die Endnutzer*innen am Endgerät auswählen können (siehe Tabelle 6). Presets können mit individuellen Voreinstellungen belegt werden und geben Endnutzer*innen sinngemäß verschiedene voreingestellte Anwendungsmöglichkeiten.

In letzter Instanz kann im Exportfenster eine verschiedene Auswahl an Exportmöglichkeiten getroffen werden. Für dieses Projekt wird ein einfacher MPEG-H Production Format (MPF) Export gewählt, der ein Multichannel PCM-Wavefile mit Metadatenspur erstellt.

Im nächsten Schritt kann das MPF-File mithilfe des MPEG-H Production Format Players einer Quality Control unterzogen und anschließend archiviert werden.

Für die finale Distribution an die Endnutzer*innen muss das MPF-file im Muxing Tool encodiert und, falls ein Video vorhanden ist, mit dem Video verknüpft werden.

Als Output-Format wurde für dieses Projekt ein MP4-Render gewählt.

Nach dem Rendering-Prozess ist die Datei samt Metadaten fertig und kann in einem MPEG-H fähigen Endgerät verwendet werden.

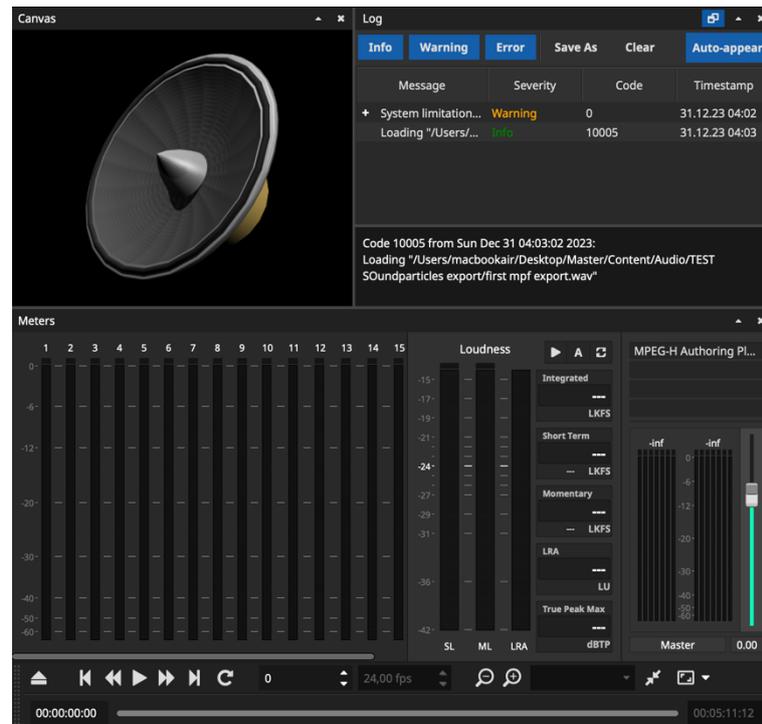


Abbildung 44, MPF-Player Quality Control

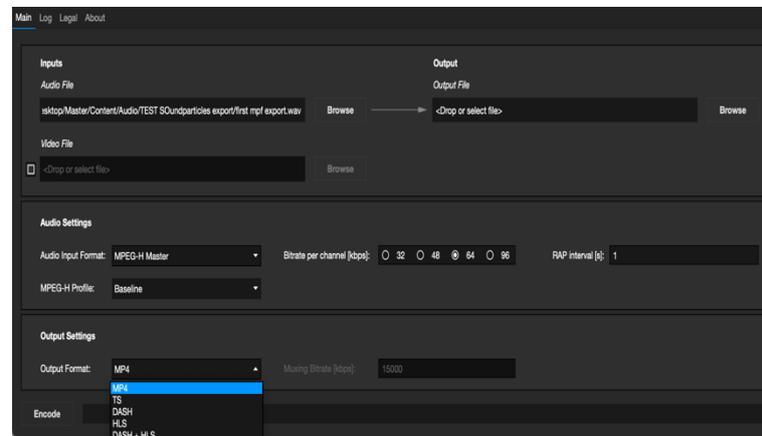


Abbildung 45, Muxing Tool

4.4 Visuelles

Als Distributionsidee wurde ein Flyer entworfen, der als Beilage zu einem Reisebuch dienen könnte. Der Flyer enthält einen QR-Code, der auf eine eigens erstellte Website verweist, auf der ein Dummy-Webplayer die Audiostreams abspielt.

Da gegenwärtig kein MPEG-H-Webplayer verfügbar ist, wurde ein binaurales File in 3 Ausführungen (Presets) exportiert und in den Dummy-

Player eingebunden, der die Vorstellung einer potenziellen Umsetzung visualisiert.

Diese Distributionsidee ermöglicht potenziellen Teilnehmern, unmittelbaren Zugriff auf die Inhalte der Audiotour. Auf der Rückseite des Flyers wird der Zugang und die Anwendung erklärt.⁴²

Um auf die Website zuzugreifen wurde im Anhang der Link und ein QR-Code beigefügt.

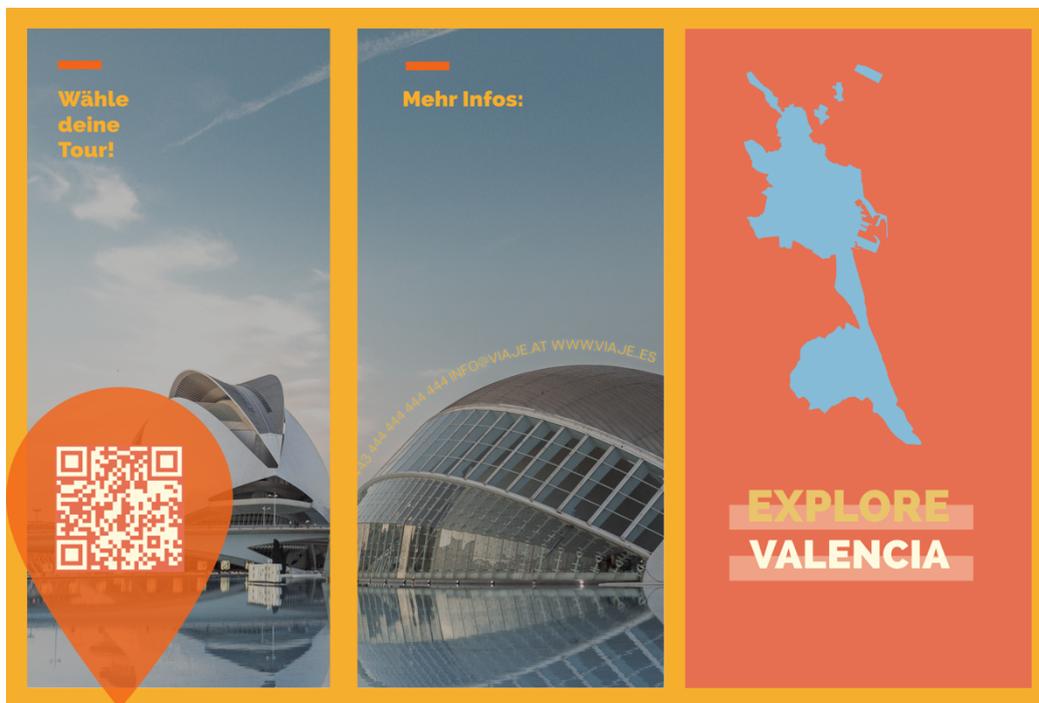


Abbildung 46, Vorderseite Flyer

⁴² Für einen barrierefreieren Zugang kann zudem eine zusätzliche Audiospur mit den Erklärungen aufgenommen werden.



Abbildung 47, Visualisierung Flyer

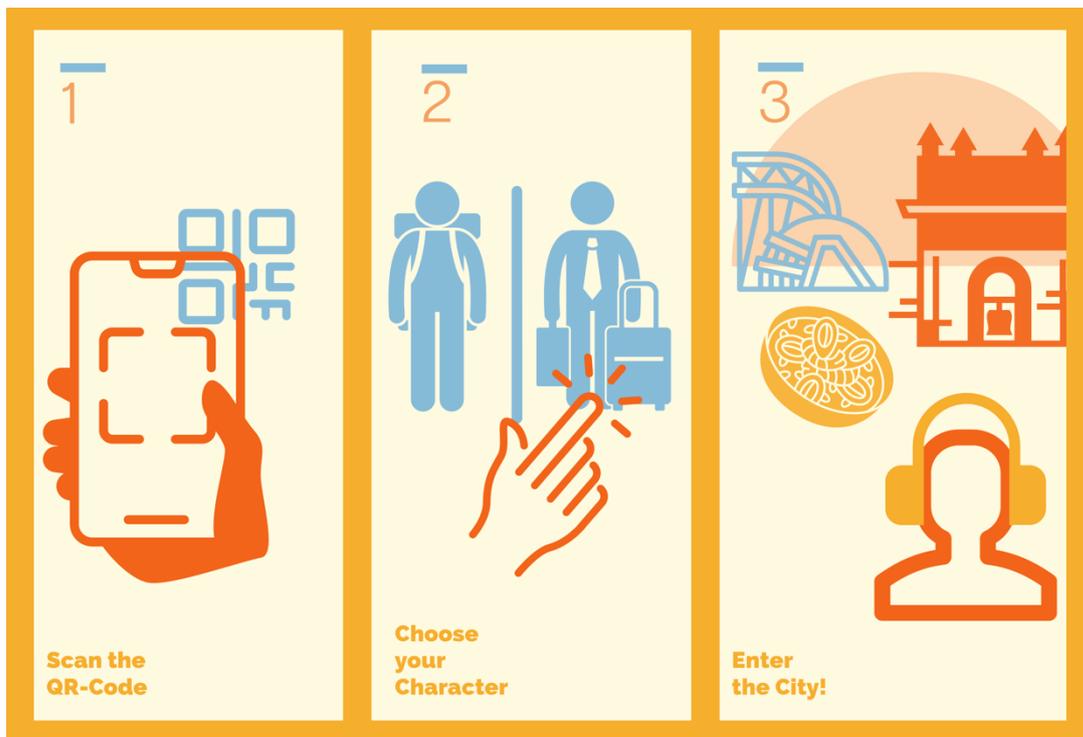


Abbildung 48, Rückseite mit Bedienungsanweisung



Abbildung 49, Screenshot der Website (1)

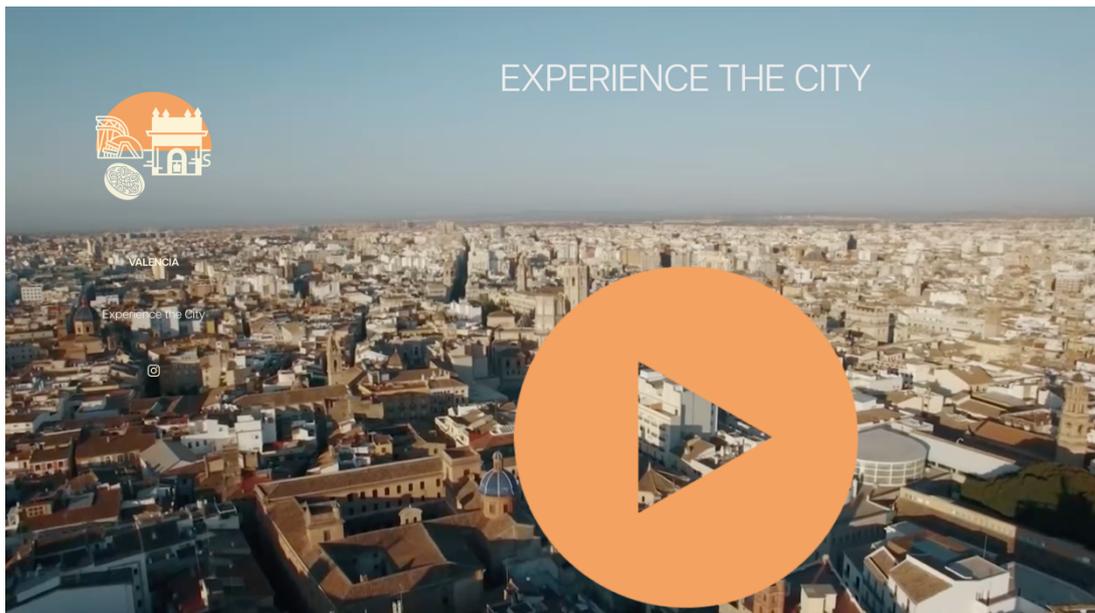


Abbildung 50, Screenshot der Website (2)

Conclusio

In einer nicht vollends definierten Welt von Next-Generation-Audio (NGA) eröffnen sich verschiedene Ansätze für die Medienproduktion. Die vorgestellte Produktionsmethode in dieser Arbeit wurde darauf ausgerichtet, einen optimierten Arbeitsablauf innerhalb der aktuellen Rahmenbedingungen zu entwickeln. Trotz der, in professionellen Produktionen üblichen elaborierten Mikrofonierungsmethoden mit einer höheren Anzahl von Mikrofonen, zeigt der hier präsentierte Ansatz eine zugängliche Möglichkeit für Jedermann, Inhalte für Next-Generation-Audio zu erstellen.

Zu beachten ist vorallem, dass Mpeg-H und lineare Medien nur begrenzt für interaktive Inhalte geeignet sind. Insbesondere bei der Anwendung von Spatial Audio ist die Verbindung von räumlichen Effekten mit dem Inhalt entscheidend für ein immersives Erlebnis. Schwierigkeiten wurden insbesondere bei der Integration eines nicht-diegetischen Sprechers festgestellt. Der Fokus liegt vor allem auf dem Sprecher und der räumliche Effekt, insbesondere bei diffusen räumlichen Aufnahmen, geht verloren. Möglicherweise liegt dies an psychoakustischen Effekten, die weiter erforscht werden müssen.

Trotz der Herausforderungen könnte durch einen sorgfältig durchdachten inhaltlichen Ansatz, eher im Sinne eines Hörspiels, ein Mehrwert für Reisemedien geschaffen werden.

Es hat sich herausgestellt, dass Immersive Audioanwendungen im Tourismussektor bereits existieren, auch wenn die Implementierung eher selten ist. Eine weitere, denkbare Anwendung wäre eine "augmented Roadmap" für Reisemedien, bei der Benutzer*innen die verschiedenen Orte im Voraus hören können. Die Integration von MPEG-H wäre in diesem Fall höchstens als barrierefreie Hilfe sinnvoll, um beispielsweise Blinden die Karte genauer zu beschreiben. Da MPEG-H 3D Audio prinzipiell auch ein Teil des MPEG-H-Standards ist, der vor allem die Wiedergabe mit visuellen Medien vorsieht, liegt die Stärke weiterhin vor allem bei audiovisuellen Medieninhalten.

Trotz der begrenzten Verfügbarkeit von Hardware- und Softwareapplikationen, die MPEG-H 3D Audio unterstützen, zeigt diese Arbeit einen möglichen Weg für die Umsetzung von untypischen MPEG-H Inhalten in der Zukunft auf. Vorallem aber Erweiterungen, wie der Möglichkeit HOA-Dateien im Authoringprozess einzubinden und Headtracking, etwa durch Webcams

oder Controller-Inputs für die Raumdrehung zu verwenden, würden den Arbeitsprozess erleichtern und die Immersion von Benutzer*innen steigern,

Die begrenzte Verfügbarkeit von Hardware- und Softwareapplikationen, die MPEG-H 3D Audio unterstützen, stellt derzeit noch eine Herausforderung bei der Wiedergabe dar. Trotz der noch nicht vollständigen Markteinführung der Technologie kann auf Lösungen wie die

kostenlos bereitgestellten MPEG-H Production Tools vom Fraunhofer IIS zurückgegriffen werden.

Diese Arbeit hat abschließend nicht nur Einblicke in die Integration von MPEG-H in alternative Medienformate gegeben, sondern soll auch als Anregungen für die Weiterentwicklung von Next-Generation-Audio-Inhalten dienen .

Literaturverzeichnis

7.1.4 Overhead Speaker Setup. (2023, Dezember 11).

<https://www.dolby.com/about/support/guide/speaker-setup-guides/7.1.4-overhead-speaker-setup-guide/>

Abma, J. (2016, Juli 19). Q&A with Founder of The Owl Field Michel Lafrance. *Whitefox*.

<https://wearewhitefox.com/qa-founder-owl-field-michel-lafrance/>

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of The Nazis in Prewar Germany *. *The Quarterly Journal of Economics*, 130(4), 1885–1939.

<https://doi.org/10.1093/qje/qjv030>

Albiez, S. (2003). Strands of the future: France and the birth of electronica. *Volume I, 2 : 2*, 99–114.

<https://doi.org/10.4000/volume.2274>

Alexander, R. (2013). *The Inventor of Stereo: The life and works of Alan Dower Blumlein*. Taylor & Francis. <https://books.google.at/books?id=bkl5AgAAQBAJ>

Amato, J. (2023, August 29). Importance of QC of Audio Phase issues. *Venera Technologies*.

<https://www.veneratech.com/importance-of-qc-of-audio-phase-issues/>

Ampel, F. J., & Uzzle, T. (1993, März). The History of Audio and Sound Measurement. *Audio Engineering Society Convention 94*. <http://www.aes.org/e-lib/browse.cfm?elib=6566>

Ananda, F. (2023, Oktober 20). The Influence of Vinyl Records on Music Listening and Society.

Medium. <https://medium.com/@farhanananda41192/the-influence-of-vinyl-records-on-music-listening-and-society-5057bb3fe962>

AR Audio. Der next level 3D Audio Guide. (2020). SonicTraces. <https://sonictraces.com/>

Audiokinetic (Regisseur). (2021, Januar 22). *Wwise 2021.1 Beta—Object-based Audio Features*

Hands On. <https://www.youtube.com/watch?v=HhfwlfyZLQY>

AURO-3D. (2023). <https://www.auro-3d.com/>

Ballou, G. (2015). *Handbook for Sound Engineers*. Taylor & Francis.

https://books.google.at/books?id=t__vBgAAQBAJ

- Begault, D., & Wenzel, E. (2001). *Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source*. 904–916.
- Bell, A. P. (2018). *Dawn of the DAW: The Studio as Musical Instrument*. Oxford University Press.
<https://books.google.at/books?id=ZD9FDwAAQBAJ>
- Bellis, M. (2018, Februar 27). History of the Loudspeaker Primitive Loudspeakers Were Created in the Late 1800s. *ThoughtCo*. <https://www.thoughtco.com/history-of-loudspeaker-4076782>
- Bernsteiner, M. (2019). *3D audio sound design: Creating immersive spatial audio experiences*.
- Best, V., Baumgartner, R., Lavandier, M., Majdak, P., & Kopco, N. (2020). Sound Externalization: A Review of Recent Research. *Trends in hearing*, 24, 2331216520948390.
<https://doi.org/10.1177/2331216520948390>
- Bhuller, M., Havnes, T., McCauley, J., & Mogstad, M. (2023). *How the Internet Changed the Market for Print Media* (Working Paper 30939; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w30939>
- Birtchnell, T., & Elliott, A. (2018). Automating the black art: Creative places for artificial intelligence in audio mastering. *Geoforum*, 96, 77–86. <https://doi.org/10.1016/j.geoforum.2018.08.005>
- Blauert, J., & Rabenstein, R. (2012). Providing Surround Sound with Loudspeakers: A Synopsis of Current Methods. *Archives of Acoustics*, 37. <https://doi.org/10.2478/v10168-012-0002-y>
- Bleek, T. (2023). Aufbruch inmitten der Krise. In *Im Taumel der Zwanziger: 1923: Musik in einem Jahr der Extreme* (S. 248–272). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-66602-9_8
- Bleidt, R. L., Sen, D., Niedermeier, A., Czelhan, B., Füg, S., Disch, S., Herre, J., Hilpert, J., Neuendorf, M., Fuchs, H., Issing, J., Murtaza, A., Kuntz, A., Kratschmer, M., Küch, F., Füg, R., Schubert, B., Dick, S., Fuchs, G., ... Kim, M.-Y. (2017). Development of the MPEG-H TV Audio System for ATSC 3.0. *IEEE Transactions on Broadcasting*, 63(1), 202–236.
<https://doi.org/10.1109/TBC.2017.2661258>
- Blumlein, A. D. (1931). *Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems* (EMI Ltd Electrical and Musical Industries Ltd Patent GB394325A). <https://patents.google.com/patent/GB394325A/en>
-

- Blumlein, A. D. (1936). *Sound-transmission, sound-recording, and sound-reproducing system* (United States Patent US2062275A). <https://patents.google.com/patent/US2062275A/en>
- Bolin, G. (2012). *Cultural Technologies: The Shaping of Culture in Media and Society*. Routledge. <https://books.google.at/books?id=f8lB5eT9dDYC>
- Brandenburg, K. (1999, September). MP3 and AAC Explained. *Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding*. <http://www.aes.org/e-lib/browse.cfm?elib=8079>
- Brazil. (2023). <https://mpegh.com/brazil/>
- Brookes, T., Kassier, R., Lee, H.-K., & Rumsey, F. (2005). *An Informal Comparison Between Surround-Sound Microphone Techniques*.
- Brooks. (2023, November 29). Spotify Wrapped 2023: „Music genres are now irrelevant to fans“. *BBC News*. <https://www.bbc.com/news/entertainment-arts-67111517>
- Brown, A. D., Stecker, G. C., & Tollin, D. J. (2015). The Precedence Effect in Sound Localization. *Journal of the Association for Research in Otolaryngology*, *16*(1), 1–28. <https://doi.org/10.1007/s10162-014-0496-2>
- Bucher, H.-J. (2001). Wie interaktiv sind die neuen Medien? In H.-J. Bucher & U. Püschel (Hrsg.), *Die Zeitung zwischen Print und Digitalisierung* (S. 139–171). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-14676-6_7
- Bull, M. (2005). No Dead Air! The iPod and the Culture of Mobile Listening. *Leisure Studies*, *24*(4), 343–355. <https://doi.org/10.1080/0261436052000330447>
- Camerer, F. (2010). *On the way to Loudness Nirvana*. EBU Tech.
- Camerer, F. (Regisseur). (2022, Juni 10). *Florian Camerer: Ein 9-Kanal-Location-Mikro, von Grund auf neu gebaut*. https://www.youtube.com/watch?v=wnpj00l_awo
- Cenciarelli, C. (Hrsg.). (2021). *The Oxford Handbook of Cinematic Listening* (1. Aufl.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190853617.001.0001>
- Cheung, V. S. Y., Lo, J. C. Y., Chiu, D. K. W., & Ho, K. K. W. (2023). Evaluating social media's communication effectiveness on travel product promotion: Facebook for college students in Hong Kong. *Information Discovery and Delivery*, *51*(1), 66–73. <https://doi.org/10.1108/IDD-10-2021-0117>
-

- Chiariglione, L. (2011, Mai 3). *MPEG Press Release*. INTERNATIONAL ORGANISATION FOR STANDARDISATION ORGANISATION INTERNATIONALE DE NORMALISATION ISO -. https://web.archive.org/web/20110503174827/http://mpeg.chiariglione.org/meetings/kurihama91/kurihama_press.htm
- Chion, M. (2019). *Audio-Vision: Sound on Screen*. Columbia University Press. <https://doi.org/10.7312/chio18588>
- Coffey, A. (2016). The impact that music streaming services such as Spotify, Tidal and Apple Music have had on consumers, artists and the music industry itself. *Interactive Digital Media*. University of Dublin, 195–202.
- Czyzewski, A., Kostek, B., Ody, P., & Kornacki, A. (2001, Mai). Problems Related to Surround Sound Production. *Audio Engineering Society Convention 110*. <http://www.aes.org/e-lib/browse.cfm?elib=9944>
- Davis, M. (1993, Oktober). The AC-3 Multichannel Coder. *Audio Engineering Society Convention 95*. <http://www.aes.org/e-lib/browse.cfm?elib=6459>
- DeKanter, N. (2005). Gaming redefines interactivity for learning. *TechTrends*, 49(3), 26–31.
- Dempf, P. (2021). Der Volksempfänger. *Geschichte lernen*, 2021(203), 47–50.
- Die drei ??? In besonderem Klang*. (2023). Die drei ??? <https://www.dreifragezeichen.de/3d-audio>
- Dimitris, L., & Vasiliki, K. (2013). The effect of virtual communities and social network sites on tourism destination choice. *June 27Th To The 29Th Of*. https://jotr.eu/pdf_files/V6.pdf#page=122
- Dolby. (2018). *Dolby Atmos® Home Theater Installation Guidelines* (S. 51). https://www.dolby.com/siteassets/technologies/dolby-atmos/atmos-installation-guidelines-121318_r3.1.pdf
- Dolby. (2023a). *Dolby Atmos Cinema Sound—Dolby Professional*. <https://professional.dolby.com/cinema/dolby-atmos/>
- Dolby. (2023b). *How to get Dolby Atmos for Streaming & Devices*. <https://www.dolby.com/experience/home-entertainment/articles/how-to-get-dolby-atmos/>
- Dolby Atmos: Eindringliches unvergessliches Klangerlebnis - Dolby*. (2023). <https://www.dolby.com/de/technologien/dolby-atmos/>
-

- Dolby Atmos Speaker Setup*. (2023, aufgerufen). <https://www.dolby.com/about/support/guide/dolby-atmos-speaker-setup/>
- DTS:X*. (2023). DTS. <https://dts.com/dtsx/>
- Du Gay, P. (2013). *Doing cultural studies: The story of the Sony Walkman* (Second edition). Sage Publications.
- Dupré, T., Denjean, S., Aramaki, M., & Kronland-Martinet, R. (2021). Spatial Sound Design in a Car Cockpit: Challenges and Perspectives. *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 1–5. <https://doi.org/10.1109/I3DA48870.2021.9610910>
- Dussel, K. (2022). *Deutsche Rundfunkgeschichte* (4. Auflage, mit einem neuen, umfangreichen Vorwort). Herbert von Halem Verlag.
- DWIH Sao Paulo. (2019, Herbst). *Weltpremiere: TV Globo sendete „Rock in Rio“-Konzerte live mit MPEG-H-Sound über ISDB-Tb und 5G | DWIH Sao Paulo*. <https://www.dwih-saopaulo.org/de/2019/12/03/weltpremiere-tv-globo-sendete-rock-in-rio-konzerte-live-mit-mpeg-h-sound-ueber-isdb-tb-und-5g/>
- EBU ADM, G. (2023). *Types of Audio—EBU ADM Guidelines*. Types of Audio. https://adm.ebu.io/background/audio_types.html
- Editor, D. D., Audio. (2017, Juni 27). *Fraunhofer Completes MPEG-H Training in Korea as Part of ATSC 3.0 Launch*. Sports Video Group. <https://www.sportsvideo.org/2017/06/27/fraunhofer-completes-mpeg-h-training-in-korea-as-part-of-atsc-3-0-launch/>
- Epstein, M. (2017). *Detour: Rome*. Walking Cinema. <https://www.walkingcinema.org/projects/detour>
- Eurovision Song Contest (ESC) trifft Next Generation Audio (NGA)* -. (2020, Juli 1). VRTONUNG - Virtual Reality Sound. <https://www.vrtonung.de/next-generation-audio/>
- Everson, Z. (2015, März 31). *Detour: Finally an Audio Guide That's Actually Worth Paying for*. Condé Nast Traveler. <https://www.cntraveler.com/stories/2015-03-31/detour-finally-an-audio-guide-that-s-actually-worth-paying-for>
- Farina, A., Pinardi, D., Binelli, M., Ebri, M., & Ebri, L. (2018, Mai). Virtual Reality for Subjective Assessment of Sound Quality in Cars. *Audio Engineering Society Convention 144*. <http://www.aes.org/e-lib/browse.cfm?elib=19399>
-

- FAZ. (2002, September 4). Internet: Das Ende von Napster. *FAZ.NET*.
<https://www.faz.net/aktuell/wirtschaft/internet-das-ende-von-napster-170587.html>
- Fleischmann, J. (2017, Mai 14). MPEG-H - ein Audioformat der nächsten Generation (NGA).
techmagazin. <http://tech-magazin.de/2017/05/mpeg-h-ein-audioformat-der-naechsten-generation-nga/>
- Flieh, L. (1978). HI-FI IN DER TÄGLICHEN PRAXIS: QUADROPHONIE AUF SCHALLPLATTEN, KUNSTKOPF UND ANDERE NEUHEITEN. *Österreichische Musikzeitschrift*, 33(6), 327–328.
<https://doi.org/doi:10.7767/omz.1978.33.6.327>
- Fonseca, N. (2020). *Sound Particles Reference Manual* (S. 137) [Benutzerhandbuch].
https://www.soundparticles.com/assets/files/sp2_manual.pdf
- Fotis, J., Buhalis, D., & Rossides, N. (2012). Social Media Use and Impact during the Holiday Travel Planning Process. In M. Fuchs, F. Ricci, & L. Cantoni (Hrsg.), *Information and Communication Technologies in Tourism 2012* (S. 13–24). Springer Vienna.
https://doi.org/10.1007/978-3-7091-1142-0_2
- Frank, M. (2014, September). *How to make Ambisonics sound good*.
- Frank, M. (2023, April 4). *Frage zum Stereo Encoder* [Persönliche Kommunikation].
- Frank, M., Zotter, F., & Sontacchi, A. (2015). Producing 3D Audio in Ambisonics. *Proceedings of the AES International Conference, 2015*.
- Fraunhofer IIS. (2017, Juli 28). *Alive and Kicking – mp3 Software, Patents and Licenses – Fraunhofer Audio Blog*. <https://www.audioblog.iis.fraunhofer.com/mp3-software-patents-licenses>
- Fraunhofer IIS. (2018, Juni 29). *Großes französisches Tennisturnier erfolgreich mit MPEG-H Audio über Satellit und DVB-T2 ausgestrahlt – Fraunhofer Audio Blog*.
<https://www.audioblog.iis.fraunhofer.com/de/mpeg-h-trial-france>
- Fraunhofer IIS. (2023a). *Accessibility*. <https://mpegh.com/accessibility/>
- Fraunhofer IIS. (2023b). *Benefit from Universal Delivery*. <https://mpegh.com/features/#universal>
- Fraunhofer IIS. (2023c). *MPEG-H Audio*. <https://mpegh.com/>
- Fraunhofer IIS. (2023d). *The MP3 History-Marketing*. Fraunhofer Institute for Integrated Circuits IIS.
<https://www.mp3-history.com/en/marketing.html>
-

- Friesecke, A. (2014). *Ein Nachschlagewerk für Tontechniker*. De Gruyter Saur.
<https://doi.org/doi:10.1515/9783110340181>
- Gallacher, N. (2013). Game audio—An investigation into the effect of audio on player immersion. *The Computer Games Journal*, 2(2), 52–79. <https://doi.org/10.1007/BF03392342>
- Garratt, G. R. M. (1994). *The Early History of Radio: From Faraday to Marconi*. Institution of Engineering and Technology. <https://digital-library.theiet.org/content/books/ht/pbht020e>
- Gernemann, A. (2002). „DECCA-Tree“—Gestern und heute. *Verlag K.G.Saur 2003*.
- Gerzon, M. (1974). *Studio Sound: What's wrong with quadrasonics*. 16.
<https://intothefield.music.ox.ac.uk/sites/default/files/intothefield/documents/media/whatswrongwithquadrasonics.pdf>
- Gerzon, M. (1980). *Practical Periphony: The Reproduction of Full-Sphere Sound* (AES Convention Paper 1571; Audio Engineering Society Convention Paper). Institute, University of Oxford,.
https://intothefield.music.ox.ac.uk/sites/default/files/intothefield/documents/media/practical_periphony_the_reproduction_of_full_sphere_sound_michael_gerzon_draft_for_aes_london_feb_1980.pdf
- Gil-Carvajal, J. C., Cubick, J., Santurette, S., & Dau, T. (2016). Spatial Hearing with Incongruent Visual or Auditory Room Cues. *Scientific Reports*, 6(1), 37342.
<https://doi.org/10.1038/srep37342>
- Gochermann, J. (2020). Welche Bedeutung hat die Technologie-S-Kurve? In *Technologiemanagement: Technologien erkennen, bewerten und erfolgreich einsetzen* (S. 9–13). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-28799-3_3
- Goebbels, J. (1940). *Signale der neuen Zeit* (8. Aufl.). Zentralverlag der NSDAP.
https://ia800304.us.archive.org/cors_get.php?path=/5/items/Goebbels-Joseph-Signale-der-neuen-Zeit/GoebbelsJoseph-SignaleDerNeuenZeit8.Auflage1940381S.ScanFraktur.pdf
- Gomes, R. M. (2016). Audio Quality X Accessibility How Digital Technology Changed the Way We Listen and Consume Popular Music. *Revista Vórtex*, 4(2), 1–14.
<https://doi.org/10.33871/23179937.2016.4.2.1320>
-

- Goodwin, J. (2020, September 13). *Vinyl record sales surpass CDs for the first time since the 1980s* / *CNN Business*. CNN. <https://www.cnn.com/2020/09/13/tech/vinyl-records-cd-sales-riaa/index.html>
- Gowan, M. (2023, April 10). *What is spatial audio? Here's everything you need to know*. CNN Underscored. <https://www.cnn.com/cnn-underscored/electronics/what-is-spatial-audio>
- Grewe, Y., Eibl, P., Simon, C., Torcoli, M., Rieger, D., & Scuda, U. and. (2021, Oktober). MPEG-H Audio production workflows for a Next Generation Audio Experience in Broadcast, Streaming and Music. *Audio Engineering Society Convention 151*. <http://www.aes.org/e-lib/browse.cfm?elib=21512>
- Haas, H. (1951). Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache (On the effect of a single echo on the perceptual quality of speech). *Acustica*, 1, 49–58.
- Haberer, M. (2020). *Versuch über Spotify, oder: Musikstreaming als Arbeit am Subjekt* (S. 145–162). <https://doi.org/10.1515/9783839451496-007>
- Hamasaki, K. (2003, Juni). Multichannel Recording Techniques for Reproducing Adequate Spatial Impression. *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*. <http://www.aes.org/e-lib/browse.cfm?elib=12288>
- Hammershøi, D., & Møller, H. (2005). Binaural Technique—Basic Methods for Recording, Synthesis, and Reproduction. In J. Blauert (Hrsg.), *Communication Acoustics* (S. 223–254). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-27437-5_9
- Hannah. (2021). *The Decca Tree | The Secrets Behind The Legendary Recording Technique*. United Kingdom. <https://abbeyroadinstitute.co.uk/blog/the-decca-tree-the-secrets-behind-the-legendary-recording-technique/>
- Haug, S., & Weber, K. (2002). *Kaufen, Tauschen, Teilen: Musik im Internet*. P. Lang.
- Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015a). MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio. *IEEE Journal of Selected Topics in Signal Processing*, 9(5), 770–779. <https://doi.org/10.1109/JSTSP.2015.2411578>
- Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015b). MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding. *J. Audio Eng. Soc.*, 62(12), 821–830.
-

- Hilmes, M., & Loviglio, J. (2002). *Radio Reader: Essays in the Cultural History of Radio*. Routledge.
<https://books.google.at/books?id=iSJTLDDg0XEC>
- Hoepner, G. A. (2019). Lineare Medien. In *Hoepner-Marketing-Lexicon*. https://www.wirtschaftswiki.fh-aachen.de/index.php?title=Lineare_Medien
- Holman, T. (2014). *Surround Sound: Up and Running*. Taylor & Francis.
<https://books.google.at/books?id=0qDcAwAAQBAJ>
- Homer, M. (2009). Beyond the studio: The impact of home recording technologies on music creation and consumption. *Nebula*, 6(3), 85+. Gale Academic OneFile.
- Hudspeth, A. (1997). How hearing happens. *Neuron*, 19(5), 947–950.
- Hunter, T., Smith, T., Crossan, M., Wilkinson, M. A., & Perry, M. (2001). *Napster and MP3: Redefining the Music Industry*. University of Western Ontario, Richard Ivey School of Business.
- ifpi. (2018). *Music Consumer Insight Report 2018* (Music Consumer Insight Report) [Report]. ifpi.
https://www.ifpi.org/wp-content/uploads/2020/07/091018_Music-Consumer-Insight-Report-2018.pdf
- Image Assistant | SCHOEPS Mikrofone*. (o. J.). Abgerufen 10. Januar 2024, von
<https://schoeps.de/wissen/image-assistant.html>
- Immink, K. S. (2009). *The History of the Compact Disc*.
- Inc, A. (o. J.). *Documentation | Audiokinetic*. Abgerufen 16. Januar 2024, von
https://www.audiokinetic.com/en/library/edge/?source=Help&id=ambisonics_channel_ordering
- Internationalisierungszentrum Steiermark (Regisseur). (2021, August 25). *Die Zukunft von Audio ist 3D! | Pitching-Session von Atmoky*. https://www.youtube.com/watch?v=iSa0fZdCR_I
- ITU Radiocommunication Assembly. (1992). RECOMMENDATION ITU-R BS.775-1—Multichannel stereophonic sound system with and without accompanying picture. *R BS*.
https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.775-1-199407-S!!PDF-E.pdf
- Jukka, J. (2005). *Binaural to Multichannel Audio Upmix* [HELSINKI UNIVERSITY OF TECHNOLOGY]. <http://lib.tkk.fi/Dipl/2005/urn007903.pdf>
-

- Kaprocki, N., Kovacevic, J., & Bjelica, M. (2018). Evaluation of Immersive Audio Technologies on In-Vehicle Infotainment Platforms. *2018 26th Telecommunications Forum (TELFOR)*, 420–425. <https://doi.org/10.1109/TELFOR.2018.8611863>
- Kastbauer, D. (2021, April 30). *Authoring for Audio Objects in Wwise* | Audiokinetic Blog. Ak Blog. <https://blog.audiokinetic.com/en/authoring-for-audio-objects-in-wwise/>
- Katz, M. (2010). *Capturing Sound: How Technology Has Changed Music*. University of California Press. https://books.google.at/books?id=r_p_Q6TUrqoC
- Keller, D. (2017). *Recording-Grundlagen der MS-Mikrofonierung* | Universal Audio. <https://www.uaudio.de/blog/mid-side-mic-recording/>
- Kimizuka, M. (2012). *Historical Development of Magnetic Recording and Tape Recorder* (Vol. 17; Reports on the Systemization of Technologies). National Museum of Nature and Science. https://sts.kahaku.go.jp/diversity/document/system/pdf/073_e.pdf
- Kittler, F. (1996, Juli 30). *The History of Communication Media*. The History of Communication Media. <https://journals.uvic.ca/index.php/ctheory/article/download/14325/5101>
- Klapholz, J. (1991). Fantasia: Innovations in Sound. *J. Audio Eng. Soc.*, 39(1/2), 66–68, 70.
- Kohlrausch, A., Braasch, J., Kolossa, D., & Blauert, J. (2013). An Introduction to Binaural Processing. In J. Blauert (Hrsg.), *The Technology of Binaural Listening* (S. 1–32). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37762-4_1
- König, W. (2003). Mythen um den Volksempfänger. Revisionistische Untersuchungen zur nationalsozialistischen Rundfunkpolitik. *Technikgeschichte*, 70(2), 73–102. <https://doi.org/10.5771/0040-117X-2003-2-73>
- Krebs, S. (2017). *Failure and success of dummy head recording: An innovation history of 3D*. C2DH | Luxembourg Centre for Contemporary and Digital History. <https://www.c2dh.uni.lu/projects/failure-and-success-dummy-head-recording-innovation-history-3d>
- Kuech, F., Kratschmer, M., Neugebauer, B., Meier, M., & Baumgarte, F. (2015, Oktober). Dynamic Range and Loudness Control in MPEG-H 3D Audio. *Audio Engineering Society Convention 139*. <http://www.aes.org/e-lib/browse.cfm?elib=18021>
-

- Lan, Y., Yin, T., & Yu, G. (2022). Effect of torso reflections from simplified torso models on head-related transfer function simulation and ipsilateral perception of elevation. *Applied Acoustics*, 201, 109095. <https://doi.org/10.1016/j.apacoust.2022.109095>
- Letowski, T., & Letowski, S. (2012). *Auditory Spatial Perception: Auditory Localization*. https://www.researchgate.net/publication/265033553_Auditory_Spatial_Perception_Auditory_Localization
- Levine, M. (2019, Mai 1). *Recording Basics: The History of the DAW*. Yamaha Music. <https://hub.yamaha.com/proaudio/pa-history/the-history-of-the-daw/>
- Li, S., & Peissig, J. (2020). Measurement of Head-Related Transfer Functions: A Review. *Applied Sciences*, 10(14), 5014. <https://doi.org/10.3390/app10145014>
- Locke, C. (2015, Dezember 21). Detour is an augmented reality app that gives listeners audio tours of various cities. *Wired*. <https://www.wired.com/2015/12/detour-audio-tour/>
- Los Senderos Studio. (2022). <https://lossenderosstudio.com/index.htm>
- Maftei, V. A., Gerogiannis, V. C., & Papageorgiou, E. I. (2016). Critical success factors of online music streaming services—A case study of applying the fuzzy cognitive maps method. *International Journal of Technology Marketing*, 11(3), 276. <https://doi.org/10.1504/IJTMKT.2016.077377>
- Magaudda, P. (2011). When materiality 'bites back': Digital music consumption practices in the age of dematerialization. *Journal of Consumer Culture*, 11(1), 15–36. <https://doi.org/10.1177/1469540510390499>
- Malham, D. (2019, März). The Early Years of Ambisonics at York. *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. <http://www.aes.org/e-lib/browse.cfm?elib=20404>
- Marine-Roig. (2019). Destination Image Analytics through Traveller-Generated Content. *Sustainability*, 11(12), 3392. <https://doi.org/10.3390/su11123392>
- Marston, D. (2014). *The Audio Definition Model*. Fourth W3C Web and TV Workshop, München. https://www.w3.org/2013/10/tv-workshop/papers/webtv4_submission_4.pdf
- May, K. (2018, April 23). *Groupon founder's tour app Detour goes to Bose, needs a lot more to survive*. <https://www.phocuswire.com/Detour-tour-app-Bose>
-

- Mehra, R., Antani, L., Kim, S., & Manocha, D. (2014). Source and Listener Directivity for Interactive Wave-Based Sound Propagation. *IEEE transactions on visualization and computer graphics*, 20, 495–503. <https://doi.org/10.1109/TVCG.2014.38>
- Mohanam, M., & Kushe Shekhar, S. (2021). A STUDY ON THE MEDIATING EFFECT OF FoMO ON SOCIAL MEDIA (INSTAGRAM) INDUCED TRAVEL ADDICTION AND RISK TAKING TRAVEL BEHAVIORAL INTENTION IN YOUTH. *JOURNAL OF CONTENT COMMUNITY AND COMMUNICATION*, 14(8), 57–67. <https://doi.org/10.31620/JCCC.12.21/06>
- Moore, B. C. J. (1995). *Hearing*. Elsevier Science. <https://books.google.at/books?id=OywDx9pxCMYC>
- Moore, D. R., Fuchs, P. A., Rees, A., Palmer, A. R., & Plack, C. (2010). *Oxford Handbook of Auditory Science: Hearing*. OUP Oxford. <https://books.google.at/books?id=fMHYWaOj1acC>
- Moserrat Narvaez, N. (2019, September 10). *History of Audio Guides – MuseumMate*. <https://www.museummate.com/en/audioguide-history/>
- MPEG-H. (2023). *MPEG-H Authoring Suite*. <https://mpeg.com/mas/>
- MPEG-H Audio System Live-Produktionskette vom Eurovision Song Contest 2019 auf der IBC – präsentiert vom Fraunhofer IIS – Fraunhofer Audio Blog*. (2019, September 13). <https://www.audioblog.iis.fraunhofer.com/de/mpeg-h-produktionskette-esc-ibc>
- MPEG-H Audio Tools*. (2023). <https://mpeg.lze-innovation.de/#LZE>
- Mrozek, B. (2011). *Geschichte in Scheiben: Schallplatten als zeithistorische Quellen*.
- Nachbar, C., Zotter, F., Deleflie, E., & Sontacchi, A. (2011). *AMBIX - A SUGGESTED AMBISONICS FORMAT*.
- Nicolau, A., & Murgia, M. (2023, August 8). *Google and Universal Music negotiate deal over AI 'deepfakes'*. <https://www.ft.com/content/6f022306-2f83-4da7-8066-51386e8fe63b>
- Nuendo: Premium Media Production System*. (o. J.). Abgerufen 13. Januar 2024, von <https://www.steinberg.net/de/nuendo/>
- O'Brien, G. E. (2018). *The New Age of Sound: How Bell Telephone Laboratories and Leopold Stokowski Modernized Music*.
- OCT Sets / Sets für Surround-Aufnahmen / SCHOEPS Mikrofone*. (o. J.). Abgerufen 20. Dezember 2023, von <https://schoeps.de/produkte/surround-3d/oct-surround/oct-sets.html>
-

- Olivieri, F., Peters, N., & Sen, D. (2019). Scene-based audio and higher order ambisonics: A technology overview and application to next-generation audio, vr and 360 video. *EBU Tech*.
- Öz, M. (2015). Social media utilization of tourists for travel-related purposes. *International Journal of Contemporary Hospitality Management*, 27(5), 1003–1023. <https://doi.org/10.1108/IJCHM-01-2014-0034>
- Pantano, E., & Di Pietro, L. (2013). From e-tourism to f-tourism: Emerging issues from negative tourists' online reviews. *Journal of Hospitality and Tourism Technology*, 4(3), 211–227. <https://doi.org/10.1108/JHTT-02-2013-0005>
- Papenburg, J. G. (2013). *Soundfile. Kultur und Ästhetik einer Hörtechnologie*. <https://doi.org/10.25969/MEDIAREP/2253>
- Patrick, J. (2016, Februar 23). *A guide to Pierre Schaeffer, godfather of sampling*. Fact Magazine. <https://www.factmag.com/2016/02/23/pierre-schaeffer-guide/>
- Pernè, W. (2014). *Eine kleine Geschichte der Schallaufzeichnung*. Österreichische Mediathek. <https://www.mediathek.at/unterrichtsmaterialien/eine-kleine-geschichte-der-schallaufzeichnung/>
- Pike, C. (2013). *BBC - Research and Development: Listen Up! Binaural Sound*. <https://www.bbc.co.uk/blogs/researchanddevelopment/2013/03/listen-up-binaural-sound.shtml>
- Pöres, P. (2016). Monitoring and Authoring of 3D Immersive Next-Generation Audio Formats. *SMPTE Motion Imaging Journal*, 125(9), 1–7. <https://doi.org/10.5594/JMI.2016.2619459>
- Productions · Owl Field*. (2022). <https://www.owlfield.com/productions/>
- Puerta, M., Nakahira, K. T., & Kitajima, M. (2016). Collaborative Virtual Travel Experience in a Dome Augmented by a Context-Dependent Text and Audio Guide. *Poster Proceedings*, 20.
- Qizi, O. S. Q., & Kazimbekovich, B. S. (2023). THE INTRODUCTION OF SOUND INTO CINEMA: A TRANSFORMATIVE MILESTONE IN FILM HISTORY. *Новости образования: исследование в XXI веке*, 2(15), 870–872.
- Rafdinal, W., Setyawati, L., & Rachman, A. (2022). Information adoption on social Media: How does it affect travel intention? Lessons from West Java. *Journal of Tourism Sustainability*, 2(1), 36–43. <https://doi.org/10.35313/jtospolban.v2i1.33>
-

- Rieger, M. (2019, Juni 15). *Ambisonics for Virtual Reality and 360° Soundfield*. VRTONUNG - Virtual Reality Sound. <https://www.vrtonung.de/en/ambisonics/>
- Rieger, M. (2022, August 1). *MPEG-H Audio vs. „Dolby Atmos“—Es gibt einen Gewinner*. VRTONUNG - Virtual Reality Sound. <https://www.vrtonung.de/mpeg-h-audio-vs-dolby-atmos/>
- Riegler, T., & Handwerk, V. T. H. V. T. (2006). *Meilensteine des Rundfunks—Band 1*. Siebel. <https://books.google.at/books?id=QQT8DwAAQBAJ>
- Riegler, T., & Handwerk, V. T. H. V. T. (2011). *Rundfunk auf UKW*. Verlag für Technik und Handwerk. <https://books.google.at/books?id=cyD8DwAAQBAJ>
- Rosiński, A. (2022). *Microphone techniques in stereo and surround recording* (First edition). Jagiellonian University Press.
- Rudrich, D. (2017, Dezember 18). *AllRADecoder Guide*. <https://plugins.iem.at/docs/allradecoder/>
- Rudrich, D. (2018a, Juni 7). *Plug-in Descriptions—Binauraldecoder*. <https://plugins.iem.at/docs/pluginDescriptions/#binauraldecoder>
- Rudrich, D. (2018b, Juni 7). *Plug-in Descriptions—Multiencoder*. <https://plugins.iem.at/docs/pluginDescriptions/#multiencoder>
- Rudrich, D. (2018c, Juni 7). *Plug-in Descriptions—Roomencoder*. <https://plugins.iem.at/docs/pluginDescriptions/#roomencoder>
- Rudrich, D. (2018d, Juni 7). *Plug-in Descriptions—Scenerotator*. <https://plugins.iem.at/docs/pluginDescriptions/#scenerotator>
- Rudrich, D. (2018e, Juni 7). *Plug-in Descriptions—Stereoencoder*. <https://plugins.iem.at/docs/pluginDescriptions/#stereoencoder>
- Rudrich, D. (2020, November 5). *Compatibility with other Suites*. <https://plugins.iem.at/docs/compatibility/>
- Ruiz-Mafe, C., Tronch, J., & Sanz-Blas, S. (2016). The role of emotions and social influences on consumer loyalty towards online travel communities. *Journal of Service Theory and Practice*, 26(5), 534–558. <https://doi.org/10.1108/JSTP-12-2014-0294>
- Rumsey, F. (2001). *Spatial Audio*. Focal Press. <https://books.google.at/books?id=xj5AIGmS-RcC>
-

- Schabbing, B. (2005). *Musik- und Audiotechnologien zwischen Technik, Marketing und Kundenwunsch: Praxisstudien zur Einführung von Quadrophonie, Dolby Surround, Compact Disc und Music on Demand auf dem deutschen Markt*. Electronic Publ.
- Schaeffer, P., North, C., & Dack, J. (2012). *In Search of a Concrete Music*. University of California Press. <https://books.google.at/books?id=oagwDwAAQBAJ>
- Schoder, D. (2013). *Erfolg und Mißerfolg telematischer Innovationen: Erklärung der „Kritischen Masse“ und weiterer Diffusionsphänomene*. Deutscher Universitätsverlag. <https://books.google.at/books?id=xiSbBwAAQBAJ>
- SCHOEPS Mikrofone (Regisseur). (2019, September 13). *Binauralizing ORTF-3D - Tutorial*. <https://www.youtube.com/watch?v=Rd111Qjujzc>
- Schramm, H. (2019). Musik im Radio. In H. Schramm (Hrsg.), *Handbuch Musik und Medien: Interdisziplinärer Überblick über die Mediengeschichte der Musik* (S. 41–64). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21899-7_2
- Sengpiel, E. (1995, Mai). *Die Bedeutung der Blauertschen Bänder für die Tonaufnahme*. <http://www.sengpielaudio.com/DieBedeutungDerBlauertschenBaender.pdf>
- Sengpiel, E. (2003, Dezember). *Haas-Effekt und Präzedenz-Effekt (Gesetz der ersten Wellenfront)* [Tutorium]. UdK Berlin Sengpiel 12.2003 Tutorium, Berlin. <http://www.sengpielaudio.com/Haas-Effekt.pdf>
- Sennheiser. (2023). *TECHNISCHE DATEN. AMBEO VR Mic : TECHNISCHE DATEN*. https://www.sennheiser-sites.com/responsive-manuals/AMBEO_VR_MIC/DE/index.html#page/AMBEO%20VR%20Mic/VR_MIC_05_Technischdaten_DE.5.1.html
- Sheivachman, A. (2017, Mai 9). *How Travel Media Missed the Digital Leap*. Skift. <https://skift.com/2017/05/09/how-travel-media-missed-the-digital-leap/>
- Simon, C., Torcoli, M., & Paulus, J. (2019). *MPEG-H Audio for Improving Accessibility in Broadcasting and Streaming*.
- Sims, R. (2000). An interactive conundrum: Constructs of interactivity and learning theory. *Australasian Journal of Educational Technology*, 16(1). <https://doi.org/10.14742/ajet.1821>
-

- Smith, S. (2014, September 10). *Radio: The Internet of the 1930s*.
<https://www.apmreports.org/episode/2014/11/10/radio-the-internet-of-the-1930s>
- Sonible. (2018). Mid/Side Processing – Das Wichtigste kompakt – Update 2018. *sonible*.
<https://www.sonible.com/de/blog/mid-side-processing/>
- South Korea Launches UHD TV with MPEG-H Audio. (2017, Juni 1).
<https://www.businesswire.com/news/home/20170601006518/en/South-Korea-Launches-UHD-TV-with-MPEG-H-Audio>
- Spektrum. (2000). *Präzedenzeffekt*.
<https://www.spektrum.de/lexikon/neurowissenschaft/praezedenzeffekt/10222>
- Square, E. (2013, September). *Betrachtungsdauer von Werbung in verschiedenen Medien* [Statistik]. Statista. <https://de.statista.com/statistik/daten/studie/271067/umfrage/betrachtungsdauer-von-werbung-in-verschiedenen-medien/>
- Stephan, M., & Selle, M. (2013). Entwicklung von Technologien und Produkten in späten Lebenszyklusphasen. *WiSt - Wirtschaftswissenschaftliches Studium*, 42(1), 17–21.
https://doi.org/10.15358/0340-1650_2013_1_17
- Tercia, C., Teichert, T., Sirad, D., & Soehadi, A. (2020). Conveying pre-visit experiences through travel advertisements and their effects on destination decisions. *Journal of Destination Marketing & Management*, 16, 100331. <https://doi.org/10.1016/j.jdmm.2018.12.005>
- Teufel Editorial. (2015, November 27). Dolby Surround – die analogen Anfänge des Raumklangs. *Teufel Blog*. <https://blog.teufel.de/dolby-surround-die-analogen-anfaenge-des-raumklangs/>
- The impact of social networks on international tourism. A simple analysis of the romanian tourist sector. (2022). *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/IRJMETS29524>
- Theile, G. (2001a). *Multichannel Natural Music Recording Based on Psychoacoustic Principles /*.
- Theile, G. (2001b). *Natural 5.1 Music Recording Based on Psychoacoustic Principles*.
- Thomas, C. (2023, März 16). *What is Dolby Atmos and how can you get it?* SoundGuys.
<https://www.soundguys.com/what-is-dolby-atmos-25191/>
- Thompson, E. (1995). Machines, Music, and the Quest for Fidelity: Marketing the Edison Phonograph in America, 1877-1925. *The Musical Quarterly*, 79(1), 131–171. JSTOR.
-

- Thornton, S. (2010). *Michael Gerzon*. Blurb. <https://www.michaelgerzonphotos.org.uk/ambisonics.html>
- Vaičiūtė, A. (2021, Mai 26). *Evolution of Audio Technology—Focusonics®*.
<https://www.focusonics.com/evolution-of-audio-technology/>
- Vanderlyn, P. V. (1978). In Search of Blumlein: The Inventor Incognito. *J. Audio Eng. Soc.*, 26(9), 660, 662, 664, 666, 668, 670.
- Velasco-Pufleau, L. (2023). Listening is Action: A Soundwalk with Hildegard Westerkamp.
Performance Philosophy, 8(1), 86–100. <https://doi.org/10.21476/PP.2023.81430>
- Vickers, E. (2010, November). The Loudness War: Background, Speculation, and Recommendations.
Audio Engineering Society Convention 129. <http://www.aes.org/e-lib/browse.cfm?elib=15598>
- Walser, S. (2017). *Programmierbare kapazitive MEMS-Mikrofone*.
- Weitenthaler, J. M. (2017). *Der Entscheidungsprozess bei der Reiseplanung im off- und online Bereich am Beispiel vom Tourismusverband Gesäuse / von Julia Maria Weitenthaler* [Uni Graz]. <https://unipub.uni-graz.at/obvugrhs/2033220?lang=de>
- Westerkamp, H. (1989). *KITS BEACH SOUNDWALK (1989) for spoken voice and 2-channel audio*.
<https://www.hildegardwesterkamp.ca/resources/PDFs/program-notes-pdf/Kits-Beach-Soundwalk.pdf>
- Westphal, A. (2021, April 18). *DTS:X - besser als Dolby Atmos?* HIFI.DE. <https://hifi.de/ratgeber/dtsx-vergleich-dolby-atmos-56470>
- Wittek, H. (2013). *Microphone Techniques for 2.0 and 5.1 Ambience Recording*.
- Wittek, H. (2016). *ORTF-3D: Ambience recording for 3D Audio & VR*.
- Wittek, H., & GmbH, S. (2012). *Considerations on Stereo and Surround recording, reproduction and perception*.
- Wittek, H., & SCHOEPS, G. (2012). *Considerations on Stereo and Surround recording, reproduction and perception*.
- Wuttke, J. (2002). *16 Gedanken zu Surround-Aufnahmen* [16 Gedanken zu Surround-Aufnahmen].
<https://www.ingwu.de/mikrofontechnik/mikrofonaufsaetze/30-16-gedanken-zu-surround-aufnahmen.html>
- Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism Management*, 31(2), 179–188. <https://doi.org/10.1016/j.tourman.2009.02.016>
-

- Zeiler, L. (2020, November 17). *Immersive Audio: Eintauchen in 360-Grad-Technologien* / *TEDxVienna*. DER STANDARD.
<https://www.derstandard.at/story/2000121476724/immersive-audio-eintauchen-in-360-grad-technologien>
- Ziemer, T. (2023). *Psychoakustische Schallfeldsynthese für Musik: Raum schaffen für Komposition, Aufführung, Akustik und Wahrnehmung*. Springer International Publishing.
<https://doi.org/10.1007/978-3-031-26863-2>
- Zotter, F., & Frank, M. (2012). All-Round Ambisonic Panning and Decoding. *Journal of the Audio Engineering Society*, 60, 807–820.
- Zotter, F., & Frank, M. (2019). *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature.
- Zotter, F., Frank, M., & Sontacchi, A. (2010). The virtual t-design ambisonics-rig using vwap. *Proceedings of the 1st EAA-EuroRegio*.
- ZYLIA- 3D Audio Recording & Post-processing Solutions. (o. J.). Zylia Sp. z o.o. Abgerufen 11. Januar 2024, von <https://www.zylia.co/zylia-zm-1-microphone.html>
- ZYLIA PRO - 3rd order Ambisonics microphone array with software. (2023). Zylia Sp. z o.o.
<https://www.zylia.co/zylia-pro.html>
-

Abbildungsverzeichnis

Abbildung 1, Geschichte der Audiotechnik	5
Abbildung 2, Hörebene, Dickreiter et al. (2014), S. 128	17
Abbildung 3, First-Order-Ambisonics, bearbeitet, by Dr Franz Zotter zotter@iem.at CC BY-SA 3.0 or GFDL, via Wikimedia Commons	21
Abbildung 4, Third-Order-Ambisonics, by Dr Franz Zotter zotter@iem.at CC BY-SA 3.0 or GFDL, via Wikimedia Commons	21
Abbildung 5, Benutzerdefinierte Auswahl, MPEG-H Audio for Improving Accessibility in Broadcasting and Streaming - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Screenshot-of-an-accessible-MPEG-H-Audio-audio-scene-with	28
Abbildung 6, Bild: MPEG-H Audio Scene Information example (2017) figure 4.1, ATSC Standard: A/342 Part 3, MPEG-H System, Doc. A/342-3:2017, ATSC, S. 5	29
Abbildung 7, Beschreibung des Pfades durch Valencia	42
Abbildung 8, Die Standorte auf der Karte, https://www.google.at/maps/@39.407669,-0.5263212,11z?entry=tту , 12.01.2024	43
Abbildung 9, Decca Tree, https://lossenderosstudio.com/article.php?subject=17 , 18.08.2023	45
Abbildung 10, OCT-Mikrofonierung, https://lossenderosstudio.com/article.php?subject=17 , 18.08.2023	46
Abbildung 11. Doppel-ORTF, https://lossenderosstudio.com/article.php?subject=17 , 18.08.2023	47
Abbildung 12, IRT-Cross, https://lossenderosstudio.com/article.php?subject=17 , 18.08.2023	48
Abbildung 13, OCT-Surround, https://lossenderosstudio.com/article.php?subject=17 , 18.08.2023	48
Abbildung 14, Verschiedene First-Order Ambisonics Mikrofone, Zotter, F., & Frank, M. (2019). Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Springer Nature. S.11	49
Abbildung 15, Zylia PRO, https://www.zylia.co/zylia-pro.html , 13.09.2023	50
Abbildung 16, ORTF-3D, https://schoeps.de/produkte/surround-3d/ortf-3d/ortf-3d-indoor-set.html , 20.11.2023	51
Abbildung 17, Doppel-ORTF Aufbau	51

Abbildung 18, Links oben: ORTF-Test, oben mitte: Sennheiser AMBEO VR Tetraeder-Mikrofon, rechts oben: DIY Binauralkopf, unten: Kombination aus Doppel-ORTF und Binauralkopf.....	52
Abbildung 19, links oben: Mikrofonierung in Granada, mitte oben: Mikrofonierung in Valencia (las Arenas), rechts oben: Test mit Höhenkanälen in der Markthalle von Valencia, rechts unten: Mikrofonierung des Osterumzugs in Granada, mitte unten Mikrofontest in der Caló d'en Monjo, Mallorca, rechts unten: Mikrofonierung im botanischen Garten Valencia,	53
Abbildung 20, IRT-Teil, http://ima.schoeps.de/ , 28.09.2023.....	54
Abbildung 21, IRT-Teil, http://ima.schoeps.de/ , 28.09.2023.....	54
Abbildung 22, ORTF-Teil, http://ima.schoeps.de/ , 28.09.2023.....	54
Abbildung 23, ORTF-Teil, http://ima.schoeps.de/ , 28.09.2023.....	54
Abbildung 24, ORTF-Teil, http://www.sengpielaudio.com/HejiaD.htm , 25.10.2023.....	55
Abbildung 25, ORTF-Teil, http://www.sengpielaudio.com/HejiaD.htm 25.10.2023.....	55
Abbildung 26, Übersicht über die Gruppen für den Authoringprozess	56
Abbildung 27, Produktionsablauf	57
Abbildung 28, Produktionsaufbau	58
Abbildung 29, Beispiel einer möglichen Effektkette.....	60
Abbildung 30, Beispiel für die Kanaleinstellung	60
Abbildung 31, Doppel ORTF Multiencoder	60
Abbildung 32, Stereo Encoder	61
Abbildung 33, RoomEncoder	61
Abbildung 34, Binaurales Monitoring	62
Abbildung 35, Spuransicht in SoundParticles	62
Abbildung 36, Rotationsmodulator.....	63
Abbildung 37, Richtungsmodulator	63
Abbildung 38, Spureigenschaften	63
Abbildung 39, Ausgabe nach AmbiX-Konvention	63
Abbildung 40, AllRADecoder	66
Abbildung 41, einstellbare Benutzerinteraktionen für die Audiobett	67
Abbildung 42, Kanalrouting.....	68
Abbildung 43, 7.1+4H Mapping auf die einzelnen Audioobjekte	68
Abbildung 44, MPF-Player Quality Control	69
Abbildung 45, Muxing Tool	69
Abbildung 46, Vorderseite Flyer.....	70
Abbildung 47, Visualisierung Flyer.....	71
Abbildung 48, Rückseite mit Bedienungsanweisung	71
Abbildung 49, Screenshot der Website (1).....	72

Abbildung 50, Screenshot der Website (2)72

Anhang

Website

<https://travelguide.myportfolio.com/>, 22.01.2024



Text

(Ur-Text)

SPRECHER_IN 1:

(24, taucht gerne ins Leben einer fremden Stadt ein, will aus einem kurzen Urlaub das Beste rausholen, interessiert sich mehr für Erlebnisse als für historisches, Kultur im Allgemeinen findet er/sie aber spannend)

[Hintergrundgeräusche von geschäftigen Menschen und Verkehr]
(Begrüßung/Einleitung)

SPRECHER*IN 1

Hi!
Bevor es losgeht, stell dir bitte Valencia kurz vor. Die lebendige Stadt an der Südküste Spaniens ist bekannt für ihre einzigartige Architektur, ihre köstliche Küche und ihr pulsierendes Nachtleben. Valencia ist eine Stadt, in der Geschichte und Moderne nahtlos ineinander übergehen, was sie zu einem Muss für Reisende macht. Wir wollen gar nicht länger warten, also los gehts.

(Hintergrundgeräusche von Schritten auf einer Stadtstraße)
(Turia Park)

Unser erster Halt auf diesem Rundgang sind die atemberaubenden Turia-Gärten. Auf einer Länge von über neun Kilometern wurde dieses ehemalige Flussbett nach einer Flutkatastrophe in einen üppig

grünen Park verwandelt, der eine perfekte Flucht aus der geschäftigen Stadt bietet. Bei einem Spaziergang auf den verschlungenen Pfaden wirst du von einer überraschend vielseitigen Flora, beeindruckenden Skulpturen und einem Gefühl der Ruhe umgeben sein, dass man anderswo nur schwer findet. Leih dir ein Fahrrad aus, mach ein Picknick mit Freunden, schlendere ein wenig herum oder entspanne einfach in der Nähe der Springbrunnen. Der Park ist auch Treffpunkt vieler verschiedener Gruppen DIE GEMEINSAM SPORT TREIBEN ODER MUSIZIEREN.

(Ciudad de las Artes y Ciencias)

(Atmosphärenwechsel, mehr Wasser, weniger Vögel)

Wenn du ein Fan von moderner Architektur und innovativem Design bist, solltest du unbedingt in der Ciudad de las Artes y Ciencias vorbeischaun, die sich direkt in den Turia-Gärten befindet. Dieses architektonische Meisterwerk beherbergt einige der unglaublichsten kulturellen und wissenschaftlichen Einrichtungen der Stadt.

Wenn du durch diese futuristische Umgebung spazierst, wirst du auf jedenfall von den außergewöhnlichen Strukturen und Bauwerken, die die Landschaft schmücken, fasziniert sein. Das Hemisfèric, das einem riesigen Auge ähnelt, lässt einen mit Sicherheit ins Staunen versetzen, wenn du sein Planetarium betrittst oder einen Film im IMAX-Kino ansiehst. Ebenso spannend ist das Wissenschaftsmuseum mit seinen interaktiven Exponaten, die das Lernen über die Wissenschaft zu einem unterhaltsamen und spannenden Erlebnis machen.

(Geräusch von Menschen, die sich in einem Museum unterhalten)

Und schließlich das Oceanogràfic, Europas größtes Aquarium. Wenn du durch die transparenten Tunnel gehst, wirst du das Gefühl haben, unter der Meeresoberfläche zu wandeln, inmitten einer Vielzahl von Meerestieren.

(Strand)

(Strandatmosphäre, Leute unterhalten sich, Meeresrauschen, Getränke, heißer Tag)

Unser nächstes Ziel führt uns zum beliebten Valencia Beach. Mit seinem goldenen Sand, dem klaren, azurblauen Wasser und der belebten Atmosphäre ist dies der perfekte Ort, um sich zu erholen und einen Strandtag zu genießen. Egal, ob du es dir in einem Liegestuhl bequem machst, ein erfrischendes Bad im Mittelmeer nehmen oder verschiedene Wassersportarten ausprobieren möchtest, der Strand von Valencia bietet für jeden etwas.

An den Stränden von Patacona treffen sich meist jüngere Leute und man kann hier eine entspannte Atmosphäre genießen.

(Restaurants & Essen)

(Geräusch von brutzelnden Pfannen in einer Restaurantküche)

Nachdem du ein wenig Sonne getankt hast und der Appetit gewachsen ist, ist es an der Zeit, sich der lokalen kulinarischen Szene hinzugeben. Valencia ist bekannt für seine köstliche Paella, ein Gericht, das seinen Ursprung in der Region hat. Besuche unbedingt eines der vielen Restaurants, die sich auf dieses weltbekannte Reisgericht spezialisiert haben.

Von der typischen, valencianischen Paella, bei der Kaninchen und Huhn im Mittelpunkt stehen, bis hin zur Paella mit Meeresfrüchten, die mit frischen Garnelen, Muscheln und Tintenfisch zubereitet wird, hast du die Möglichkeit, deine Geschmacksnerven mit diesem kultigen valencianischen Gericht zu verwöhnen. Vergiss nicht, dazu ein Glas erfrischenden Wein aus der Region oder eine traditionelle Horchata, ein beliebtes Getränk aus Erdmandelmilch, zu trinken

Abend

(Geräusch von Schritten auf Kopfsteinpflaster)

Wenn sich der Tag langsam dem Ende zuneigt, solltest du aber unbedingt die wunderschön beleuchteten Straßen Valencias erkunden. Die historische Altstadt mit ihren engen, verwinkelten

Gassen und bezaubernden Plätzen erhält am Abend eine völlig andere Atmosphäre. Die Stadt erwacht nach Einbruch der Dunkelheit zum Leben und bietet Bars, Clubs und Live-Musik-Locations für jeden Geschmack und jede Vorliebe. Ob du nun Salsa-tanzen, elektronische Beats hören oder Live-Bands sehen willst, Valencia hat für jeden etwas zu bieten. Die lebhafteste Atmosphäre, die freundlichen Einheimischen und die elektrisierende Energie sorgen für einen Abend, den du so schnell nicht vergessen wirst.

Wenn sich der Tag langsam dem Ende zuneigt, solltest du aber unbedingt die wunderschön beleuchteten Straßen Valencias erkunden. Die historische Altstadt mit ihren engen, verwinkelten Gassen und bezaubernden Plätzen erhält am Abend eine völlig andere Atmosphäre.

(Geräusch von Schritten auf Kopfsteinpflaster)

Nach einer Nacht voller Abenteuer und Erkundungen ist es an der Zeit, unseren kleinen Trip durch Valencia zu beenden. Ich hoffe, diese Tour hat dir einen Vorgeschmack auf das pulsierende Leben in dieser Stadt gegeben. Wenn du als Reisender auf der Suche nach Geschichte, Kultur, kulinarischen Köstlichkeiten und einer lebendigen Atmosphäre bist, ist Valencia auf jedenfall der richtige Ort für dich.

[Outro-Musik verklingt]

SPRECHER_IN2:

(55, Kulturinteressiert, hat eine Vorliebe für die Museen und den historischen Hintergrund der Stadt, geht es gern ruhiger an, ist aber spontan und dynamisch)

[Hintergrundgeräusche von geschäftigen Menschen und Verkehr]

SPRECHER*IN 2:

Bevor wir beginnen, lassen Sie mich Ihnen ein Bild von Valencia malen. Die pulsierende Stadt an der Südküste Spaniens ist bekannt für ihre eindrucksvolle Architektur, ihre köstliche Küche und ihrer traditionsreichen Kultur. Valencia ist eine Stadt, in der Geschichte und Moderne nahtlos ineinander übergehen, was sie zu einer erfrischenden Abwechslung macht. Also, los gehts und rein in die Stadt.

(Hintergrundgeräusch von Schritten, Parkatmosphäre)

Unser erstes Ziel sind die bezaubernden Turia-Gärten, ein wahres verstecktes Juwel im Herzen Valencias. Diese weitläufige grüne Oase bietet eine friedliche Atempause vom städtischen Chaos. Bei einem Spaziergang über die verschlungenen Pfade treffen Sie auf malerische Brücken, charmante Brunnen und duftende Blumenbeete. Machen Sie eine Pause auf einer der Bänke und lassen Sie die Schönheit der Natur auf sich wirken. In den Turia-Gärten werden auch verschiedene Outdoor-Aktivitäten und kulturelle Veranstaltungen angeboten. Vergessen Sie nicht in den Veranstaltungskalender zu schauen, um zu sehen, was während Ihres Besuchs passiert.

(Ciudad de las Ciencias)

(Atmosphärenwechsel, mehr Wasser, weniger Vögel)

Als Nächstes führt uns unsere Reise zu einem Wunderwerk der modernen Architektur, der Ciudad de las Ciencias, der Stadt der Künste und Wissenschaften. Dieses architektonische Meisterwerk umfasst mehrere außergewöhnliche Bauwerke, von denen jedes für sich ein Kunstwerk ist. Beginnen Sie Ihre Erkundungstour im Palau de les Arts Reina Sofía, einem einzigartigen Opernhaus, das Eleganz und Erhabenheit ausstrahlt. Von dort aus können Sie sich in das beeindruckende Hemisfèric begeben, ein futuristisches Planetarium und IMAX-Kino. Erleben Sie die Wunder des Universums mit ihren atemberaubenden Projektionen und faszinierenden Shows. Und vergessen Sie nicht, dem Oceanogràfic einen Besuch

abzustatten, dem größten Aquarium Europas, in dem Sie Meereslebewesen aus aller Welt hautnah erleben können. Auf einer Gesamtfläche von 110.000 Quadratmetern kann man Meeres- aber auch Landtiere aus allen Kontinenten bestaunen und sich in ihre Welt verführen lassen.

Strand

(Strandatmosphäre, Leute unterhalten sich, Meeresrauschen, Getränke, heißer Tag)

Gestärkt und inspiriert ist es nun an der Zeit, den Charme der Küste Valencias zu genießen, indem wir uns auf den Weg zu den wunderschönen Stränden machen. Die Stadt verfügt über mehrere malerischen Sandstrände, von denen jeder seine ganz eigene Atmosphäre hat. Am beliebtesten ist der Strand Malvarrosa mit seiner belebten Promenade und einer Vielzahl von Strandrestaurants, die weitbekanntes Paella- und Meeresfrüchtespezialitäten anbieten. Wenn Sie eine ruhigere Atmosphäre bevorzugen, machen Sie sich auf den Weg zum Playa de la Patacona oder Playa de la Devesa und genießen Sie die Ruhe des offenen Meeres. Ob Sie nun schwimmen, sich Sonnen oder einfach nur gemütlich am Wasser spazieren gehen möchten, die Strände von Valencia haben für alle etwas zu bieten.

Restaurants

(Hintergrundgeräusch von klirrenden Gläsern)

Da wir uns nun Appetit geholt haben, ist es an der Zeit, die kulinarischen Köstlichkeiten Valencias kennenzulernen. Diese Stadt ist bekannt für ihre Gastronomie, in der es eine Fülle an köstlichen Gerichten zu entdecken gibt. Natürlich können wir nicht über Valencia sprechen, ohne sein typisches Gericht, die Paella, zu erwähnen. Gehen Sie in ein klassisch valenzianisches Restaurant, wo Sie den reichhaltigen Safranreis, das zarte Fleisch und die frischen Meeresfrüchte genießen können, die alle in einer traditionellen Paella-Pfanne zubereitet werden. Und nicht zu vergessen: die Tapas! Schlendern Sie durch die Straßen des historischen Zentrums und gehen Sie von Bar zu Bar, um eine Reihe von hervorragenden kleinen

Häppchen zu probieren, die mit lokalen Weinen kombiniert werden. Gönnen Sie sich die kulinarischen Schätze Valencias und lassen Sie Ihre Geschmacksknospen verzaubern.

Abend

(Hintergrundgeräusche des pulsierenden Nachtlebens)

Wenn die Sonne untergeht und die Stadt zum Leben erwacht, ist es an der Zeit, das pulsierende Nachtleben Valencias zu erkunden. Mit charmanten Tavernen bis hin zu trendigen Bars kommt hier jeder auf seine Kosten. Beginnen Sie den Abend im Ruzafa-Viertel, das für seine Hipster-Bars, Craft-Bier-Lokale und Live-Musik-Locations bekannt ist. Gönnen Sie sich einen Drink und mischen Sie sich unter die Einheimischen, während Sie in die lebhafteste Atmosphäre eintauchen. Alternativ können Sie sich auch nach El Carmen begeben, Valencias historischem Viertel, wo Sie eine Mischung aus gemütlichen Bars, Jazzclubs und Dachterrassen vorfinden. Sie haben die Wahl, aber eines ist sicher - die Nacht wird mit Lachen, guter Musik und unvergesslichen Erinnerungen gefüllt sein. Besuchen Sie Valencia und lassen Sie sich von der Stadt inspirieren!

[Outro-Musik verklingt]

E-Mail von Matthias Frank vom 04.04.2023

Hallo,

für Flexibilität würde ich die 4 Kanäle einfach auf +/-45 und +/-135 legen.

Die optimale Ordnung ist so eine Sache. Es gibt da zwei Entwurfskriterien:

1) Gleichmäßigkeit in Bezug auf die klassische Ambisonics-Idee, d.h. alle Richtungen sind gleichmäßig genau/grob aufgelöst, damit Quellbreite und Klangfarbe bei Bewegungen möglichst konstant bleiben. In diesem Fall wäre vermutlich 2. Ordnung OK.

2) maximale darstellbare Auflösung: Der kleinste Abstand ist 30° zwischen L/C und L/R, daraus ergibt sich dann 5. Ordnung. Bei 5. Ordnung sind die hinteren Richtungen aufgrund des größeren Abstands nicht mehr wirklich Ambisonisch, sondern verwenden eher VBAP, was aber kein Problem ist.

Wir verwenden immer 5. Ordnung, auch bei Systemen mit so wenig Lautsprechern.

Wichtig ist dabei noch die Platzierung und Einstellung der imaginären Lautsprecher; ohne diese lässt sich der Dekoder für 3D nicht berechnen. Ich würde je einen ganz unten und oben verwenden; den unteren unbedingt mit Gain 0 und den oberen maximal mit Gain 0.5, falls du überhaupt irgendwas von oben hast.

LG

Matthias
