HOCHSCHULE DER MEDIEN

FAKULTÄT ELECTRONIC MEDIA

Untersuchung zur Optimierung der automatisierten Anpassung und Konvertierung von NGA Inhalten

Masterarbeit am Institut für Rundfunktechnik

Eingereicht von: Michael Zimmermann

Eingereicht am: 08.04.2019

Studiengang: Audiovisuelle Medien

Matrikelnummer: 33946

Erstprüfer: Prof. Oliver Curdt

Zweitprüfer: Michael Meier (M. Eng)

Ehrenwörtliche Erklärung

Hiermit versichere ich, Michael Zimmermann, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Untersuchung zur Optimierung der automatisierten Anpassung von NGA-Inhalten“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), §24 Abs. 2 Bachelor-SPO (7 Semester), §23 Abs. 2 Master-SPO (3 Semester) bzw. §19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Stuttgart, den 08.04.2019

Kurzfassung

Die hier vorgestellte Arbeit befasst sich mit dem Prozess der automatisierten Anpassung von Next Generation Audioinhalten und erläutert die Vorteile von objektbasiertem Audio, sowie verschiedene Aspekte der Vorverarbeitung.
Next Generation Audio Inhalte (szenebasiert, objektbasiert, kanalbasiert) bieten viele innovative Möglichkeiten Audiosignale zu beschreiben und wiederzugeben. Interaktivität, Immersivität und die Benutzung von Metadaten sind im Zeitalter von Next Generation Audio von grundlegender Bedeutung. Im Rundfunk müssen viele Endgerätetypen und Übertragungsstrecken mit teils sehr unterschiedlichen Anforderungen und Möglichkeiten bedient werden. Eine Option ist die Produktion für den kleinsten gemeinsamen Nenner, welche jedoch viel Potenzial verschenkt und aus diesem Grund nicht ausreichend ist. Eine andere Möglichkeit wird durch die Konvertierung und Anpassung der Inhalte mit Hilfe einer Vorverarbeitung gegeben. Anhand von Metadaten, deren Gewichtung und der Bewertung nach ausgewählten Kriterien, kann der Prozess der Anpassung optimiert werden. Unter Berücksichtigung von Richtlinien und Anforderungen des Rundfunks, wird in einem Vorrenderprozess für jedes Wiedergabemedium- und Format, ein ideales Signal nach psychoakustischen Gesichtspunkten erstellt.
Unter Verwendung des EBU ADM Renderers werden Signale vorgerendert und mit Hilfe von Daten zur Lokalisation und zur Lokalisationsunschärfe bewertet und gewichtet. Mit diesem Vorgang wird eine Verbesserung des wiederzugebenden Signals durch eine sinnvolle Vorverarbeitung angestrebt.

Schlagwörter: Fehlerberechnung, Vorverarbeitung, NGA, Next Generation Audio, ADM, EAR, 3D, VBAP, MAA, Minimum Audible Angle

Abstract

This thesis is about the process to automatize converting next generation audio (NGA) files. Advantages of objectbased audio and different options about preprocessing will be illuminated.
Object-based audio and next generation audio provides a large set of new features. Furthermore the bit rate increases fast and mostly the full potential of NGA is not tapped. Interactivity, immersive aspects and the use of metadata is essential in times of NGA. Due to the diversity of end user devices and transmission channels it is necessary to automate the process of rendering with useful metadata. The rating and weighting of different criteria to describe sound files is used to pre-render the best signal for every individual format.
The aim of this work is to create a model that can pre-render a superb signal for every reproduction format. With the EBU ADM Renderer signals will be pre-rendered on VBAP base to get by on localization data and localization blur data. With this procedure an improvement of the reproduction signal due to a reasonable pre processing is aspired.

Keywords: Error calculation, pre-processing, NGA, Next Generation Audio, ADM, Audio Definition Model, EAR, EBU ADM Renderer, spatial audio, VBAP, Vector Based Amplitude Panning, MAA, Minimum Audible Angle

Inhaltsverzeichnis

Abkürzungsverzeichnis

ADM Audio Definition Model.

DAW Digital Audio Workstation.

DVB Digital Video Broadcasting.

EAR EBU ADM Renderer.

EBU European Broadcast Union.

HOA Higher Order Ambisonics.

ILD Interaural Level Difference.

ITD Interaural Time Difference.

ITU International Telecommunication Union.

MAA Minimum Audible Angle.

NGA Next Generation Audio.

RMU Rendering and Mastering Unit.

UHD Ultra High Definition Television.

VBAP Vector Base Amplitude Panning.

WFS Wellenfeldsynthese.

Vorwort

Die folgende Masterarbeit entstand im 5. Semester meines Studiums der audiovisuellen Medien an der Hochschule der Medien in Stuttgart. Ich möchte mich an dieser Stelle bei allen Personen bedanken, die mich tatkräftig unterstützt haben und mir mit Rat und Tat zur Seite standen. Allen voran geht ein Dank an meinen Betreuer Michael Meier, für die gute Zusammenarbeit, die Unterstützung, die Geduld und die vielen spannenden Diskussionen. Ein besonderer Dank geht auch an meine anderen Kollegen des Instituts für Rundfunktechnik. An Robert Schwering und Benjamin Weiss konnte ich mich jederzeit wenden und alle meine Fragen beantworten lassen. Letzterer unterstützte mich tatkräftig bei der Einarbeitung in LaTeX. Außerdem geht ein Dank an meinen Chef Sebastian Goossens, der mir den Besuch der Tonmeistertagung ermöglichte. Zu guter Letzt möchte ich einen besonderen Dank an meinen betreuenden Professor Oliver Curdt aussprechen. Er stand mir bei Fragen und Problemen jederzeit zur Verfügung und gestaltete außerdem die Studienzeit sehr interessant. Neben stets aktuellen Vorlesungsthemen und vielen Exkursionen und Ausflügen zu attraktiven Unternehmen möchte ich Prof. Curdt darüber hinaus für sein Engagement im Tonbereich während meines gesamten Studiums danken.

1. Einleitung

In den letzten Jahren gab es auf dem Gebiet der Audiotechnik einige neue Errungenschaften. Ein ganz klarer Trend zeigt sich: Objektbasierte Audioformate. Seit Ende des 19. Jahrhunderts ist die Audiowiedergabe kanalbasiert. Was anfangs mit der Monowiedergabe begann, schließlich auf Stereo überging, entwickelte sich weiter zu Surround und ist in der dreidimensionalen Ebene angekommen. Der nächste Schritt ist die Verbindung von Audio mit Metadaten. Der sogenannte objektbasierte Ansatz bringt die Unabhängigkeit vom Zielwiedergabesystem mit sich. Ein Objekt besteht aus einem Audiosignal und Metadaten, welche das Objekt näher beschreiben. Mittlerweile kann Ton somit nicht nur rechts-links, vorne-hinten und oben-unten, sondern auch objektbasiert an jeder beliebigen Position im dreidimensionalen Raum dargestellt werden. Die objektbasierte Herangehensweise erlaubt dabei neue Methoden zur Entwicklung und Anwendung von interaktiven, personalisierten, immersiven und skalierbaren Inhalten.

1.1. Problemstellung

Neben dem Trend des objektbasierten Audios haben sich auch Wiedergabegeräte, Nutzungsumgebung der Medien und Transportwege weiterentwickelt. Das Internet übernimmt eine wichtige Rolle bei der Übermittlung von Daten. Das Smartphone ermöglicht den Abruf von Hörfunk und Fernsehen auch unterwegs. Apps bieten neue Interaktionsmöglichkeiten. Dabei führt die Vielfalt an Formaten und Lautsprecherlayouts zu verschiedenen Technologien, welche untereinander kompatibel sein sollten. Gleichzeitig entstehen dadurch Probleme bei der Konvertierung und dem Up- und Downskalieren. Auch Produzenten und Rundfunkanstalten müssen sich an diese Vielzahl an Innovationen anpassen, um die optimale Nutzung von neuen objektbasierten Audioformaten zu ermöglichen. Am Institut für Rundfunktechnik GmbH (IRT) wird seit einigen Jahren an Standardisierungen gearbeitet,

um die Anwendbarkeit von objektbasiertem Audio im Rundfunkumfeld zu gewährleisten. Das IRT ist ein Forschungsinstitut für Rundfunk- und Multimediatechnologien mit audiovisuellem Hintergrund.

1.2. Ziel der Arbeit

In Kooperation mit dem Institut für Rundfunktechnik wird in dieser Arbeit ein Modell entwickelt, das zur Optimierung von objektbasierten Audioinhalten auf Basis des Audio Definition Models beiträgt. Das Audio Definition Model (ADM) ist ein frei verfügbares Metadatenmodell zur Beschreibung und Erfassung von Audiodateien mit ihren dazugehörigen Metainformationen (vgl. 2.2). Das Hauptanliegen des erarbeiteten Modells ist die Entwicklung eines Vorrenderprozesses. Dabei sollen Quelldateien verschiedenster Anzahl an Kanälen und Objekten auf ein möglichst universelles Ausgabesetup angepasst werden. Trotzdem soll die vollständige Information der Aufnahme erhalten bleiben. Der EBU ADM Renderer (EAR), ein von der EBU standardisierter Renderer zur Interpretation des ADM Metadatenformats (vgl. 2.3), wird somit um ein Programm erweitert, welches automatisch immer die beste Auswahl der Objekte trifft und die Anpassung steuert. Ziel dieser Arbeit ist die Optimierung der automatischen Anpassung und die damit verknüpfte Konvertierung von objektbasierten bzw. Next Generation Audio (NGA) Inhalten. Dazu gehört auch die Entwicklung und Erprobung von Methoden zur quantitativen Schätzung des wahrnehmbaren Unterschieds zwischen Varianten von NGA Inhalten verschiedenster Komplexität. Das Hauptziel ist die Erstellung eines Vorrenderprozesses.

1.3. Objektbasierte Wiedergabeverfahren im Kino

Objektbasierte Audioformate sind im Rundfunk eine neue Errungenschaft. Nie dagewesene Möglichkeiten stehen dem Nutzer zur Verfügung. Die Problematik be-

steht darin, das Format für jedermann zugänglich zu machen. Was im Rundfunk aktuell in der Anfangsphase ist, hat im Kino bereits Einzug gehalten. Seit den 1990er Jahren ist in Kinosälen der Standard des Surroundformats vertreten. Erst fünf Jahre später wurde das Format für den Heimgebrauch eingeführt [40]. Das Surroundformat erlaubt Mischungen in 5.1 und liefert somit die Kanäle Links, Center, Rechts, Linker Surround, Rechter Surround und einen LFE Kanal für Subwoofer. Dieses Format wurde auch stetig erweitert. Als nächstes wurden für Kinosäle ein innerer linker und ein innerer rechter Lautsprecherkanal hinzugefügt, die auch von vorne, zwischen dem Center und den äußeren Lautsprechern positioniert wurden. Um die Umhüllung weiter zu verbessern, wurden die Surroundlautsprecher mit einem Backsurroundkanal ergänzt, welcher das Publikum von hinten beschallt. Erst 2010 folgte im nächsten Schritt die Erweiterung zu 7.1. Auf das Kinoformat gesehen findet eine Aufteilung der Surroundkanäle auf jeweils zwei getrennte Surroundkanäle statt [33]. Dies bot den Produzenten und Sounddesignern mehr Möglichkeiten in der Gestaltung. Zusätzlich konnten Positionen von Audioelementen genauer gesetzt werden. Ein erstes Problem, was sich zeigte, war die Inkompatibilität vom Produktionsformat mit dem Wiedergabesetup. Obwohl zusätzliche Kanäle ein besseres Panning und eine exaktere Platzierung erlaubten, blieb 5.1 ein beliebtes Zielformat. Bei der Produktion in 5.1 blieben die zusätzlichen Lautsprecher eines 7.1 Kinos somit häufig unbenutzt. Vom Surroundsetup kommend, wurden die ersten Kinosäle 2012 auf dreidimensionale Formate aufgerüstet, um dem Zuschauer Situationen noch realer darstellen zu können und ihm somit die Möglichkeit zu geben, direkt ins Geschehen einzutauchen. Die Hauptunterschiede gegenüber der Surroundformate sind:

- eine verbesserte Audio Qualität und Klangfarbenabstimmung
- Hörereignisse, die von oben kommen
- eine größere räumliche Kontrolle und Auflösung [33]

Neben Auro3D, Iosono und Ambisonics ist Dolby Atmos aktuell eines der verbreitesten objektbasierten Audioformate für die Wiedergabe immersiver Audioinhalte. Am Beispiel von Dolbys Kinoformat werden Audiosignale als Objekte betrachtet,

welche dynamisch auf Lautsprecher im kompletten Raum verteilt werden. Beim sogenannten Dolby Atmos stehen bis zu 64 individuell ansteuerbare Wiedergabekanäle, sowie 128 Quellkanäle zur Verfügung [33]. Hier stellt sich schnell die Frage, inwieweit es Probleme mit der Kompatibilität gibt und wie Formatunterschiede am besten angeglichen werden können. Ein entscheidender Vorteil ist der Umstieg vom kanalbasierten- auf das objektbasierte Audio (siehe Unterschied in 2.1.1). Dieser revolutionäre Denkansatz kann die Erstellung und Entwicklung von Audio mit zusätzlichen Informationen verbessern. Dabei werden über interaktive, immersive, und skalierbare Elemente neue Möglichkeiten der Gestaltung geschaffen, um das Nutzererlebnis zu steigern [61]. Ein objektbasiertes Wiedergabeformat ergibt sich standardmäßig aus mindestens einem Kanalbett auf der einen Seite und den Objekten mit potenziellen Metadaten auf der anderen Seite. Der Audioinhalt wird als Zusammenstellung von individuellen Eigenschaften in Verbindung mit Metadaten, welche die Beziehungen beschreiben, repräsentiert. Sobald die Recheneinheit (Renderer) beispielsweise über die Position der Lautsprecher und die Art des Schallereignisses (z. B. Dialog, Musik, Effekt) informiert ist, kann diese berechnen mit welchem Pegel bestimmte Lautsprecher angesteuert werden müssen, um an einer ausgewählten Position gehört zu werden (vgl. hierzu Abschnitt 3.9). Audio kann auf diese Weise personalisiert, immersiv und formatunabhängig überliefert werden.

2. Grundlagen

Objektbasiertes Audio bzw. Next Generation Audio bietet viele neue Möglichkeiten in der Audioproduktion. Eine entscheidende Rolle spielen hierbei z. B. Interaktivität oder Personalisierung. Die aktuelle Entwicklung weist eine steigende Anzahl an Ausspielwegen und Endgeräten auf, welche auf die neue Mediennutzung zurückzuführen ist. Ziel der neuen Audioformate ist somit die Bedienung dieser vielen Zielplatformen so weit wie möglich zu vereinheitlichen. Eine natürliche Folge des Angleichens sind Einschränkungen, die beispielsweise durch den Rundfunk oder verschiedene Codec-Systeme aufgestellt werden. Dazu gehört z. B. die Eingrenzung der maximalen Audiobjektzahl aufgrund von fehlender Rechenleistung mobiler Endgeräte, sowie das Bedürfnis redundante Daten zu beseitigen. Unter Berücksichtigung der genannten Restriktionen soll durch eine Vorverarbeitung ein optimales Ausspielergebnis erzielt werden. Der Umgang mit dynamischen Metadaten für das jeweilige Codec-System stellt eine neue Herausforderung dar. Neue Herangehensweisen sind unabdingbar für die Kontrolle von objektbasierten Audioinhalten. Um den Endnutzer nicht mit unnötig vielen, komplexen Möglichkeiten zu überfordern, ist eine automatische, optimierte Anpassung der Inhalte auf eine beschränkte Kanal- und Objektzahl sinnvoll. Bei der Betrachtung aktueller Aufnahmen auf der Produktionsseite liegt nahe, dass es an bestimmten Stellen essentiell ist, Daten zu sparen. Für diverse Codecs und Transportverfahren ist es sogar notwendig, die Datenrate zu reduzieren.
Im Folgenden wird das Vorgehen dieser automatischen Anpassung beschrieben, nachdem grundlegende Begriffe erklärt wurden.

2.1. Next Generation Audio

Unter *Next Generation Audio* werden alle möglichen, neuen Eigenschaften für die Definition der Klangmedien bzw. des Fernsehtons in der Zukunft zusammengefasst.

Ein NGA Format umfasst weitere Merkmale, die in einem einzigen Datenstrom erfasst werden. Darüber hinaus wird die Anpassung der Audiowiedergabe an die Präferenzen des Endnutzers ermöglicht. Im Wesentlichen erweitert NGA die Tondarstellung für den Zuhörer mit interaktivem 3D-Sound und/oder personalisierter Wiedergabe. Aufgrund der Trennung der Metadaten vom Audioinhalt, kann das Signal immer noch beim Endnutzer flexibel angepasst werden. Somit muss sich in der Produktion nicht auf ein Format wie z. B. Stereo oder 5.1 festgelegt werden. Die Basis einer NGA Audioszene ist vorerst nicht bestimmt. Bisher gibt es drei verschiedene Herangehensweise NGA Systeme aufzubauen:

- kanalbasiertes Audio
- objektbasiertes Audio
- szenebasiertes Audio.

Des Weiteren besteht die Möglichkeit diese zu kombinieren [54]. Zu einer kanalbasierten Mischung können z. B. zusätzlich Objekte hinzugefügt werden. Diese werden zur Basis dazugemischt und lassen sich vom Zuhörer individuell benutzen. Häufig sind diese Objekte verschiedene Sprachspuren, wie z. B. eine Hörfassung mit Audiodeskription. In anderen Fällen können Objekte der dreidimensionalen Klanggestaltung dienen und beispielsweise immersive Inhalte sein [46]. Die Begriffe Immersion und immersive Audio sind in der heutigen Zeit keine Seltenheit mehr und dennoch stellt sich des Öfteren die Frage, was die Bedeutung dahinter ist. Übersetzt wird es mit den Worten *umfassen* oder *eintauchen*. Aus geometrischer Sichtweise kann Immersive Audio als Erweiterung der Dimensionen von Stereo über Surround bis hin zu 3D gesehen werden. Bei Surround konnte eine Ebene dargestellt werden, bei 3D wird auf einen kompletten Raum erweitert. Immersion ist die Umhüllung des Zuhörers mit einem realistisch wirkenden Raumklang [43]. Häufig wird Immersion auch mit Virtual Reality in Verbindung gebracht. Virtual Reality wird definiert durch die gleichzeitige Wahrnehmung der Wirklichkeit und einer virtuellen Umgebung [53]. Immersive Audio unterstützt somit den Betrachter beim Eintauchen in diese Mischwelt und versucht diese so real wie möglich zu gestalten [13]. Nun gehören nicht nur visuelle Eindrücke zur Virtual Reality, sondern

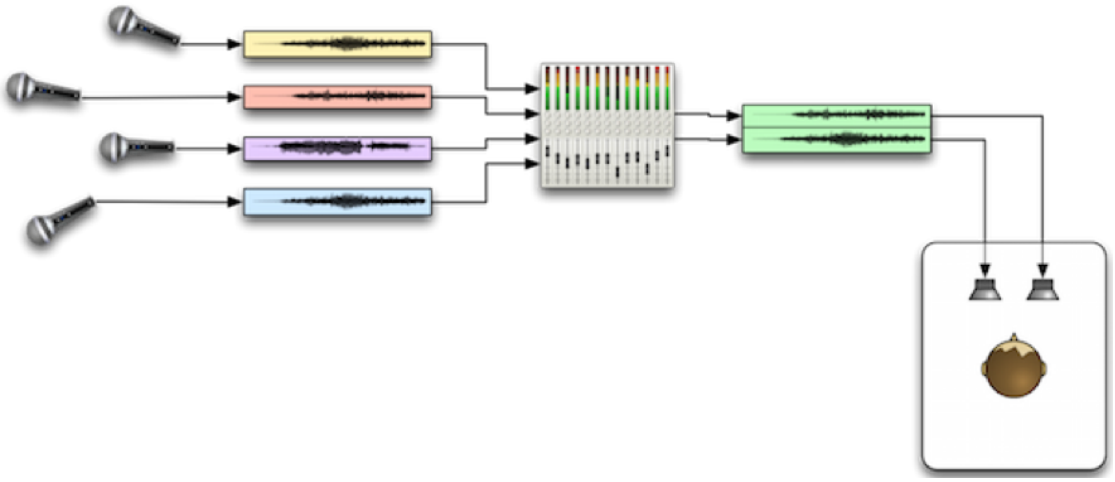


Abbildung 2.1.: Workflow kanalbasiertes Audio [44].

auch auditive Reize. Das entstandene Format wird immersive Audio genannt und kann als zugehöriges Audiosignal zur virtual Reality gesehen werden. Es ermöglicht dem Hörer die Horizontalebene des Hörens zu verlassen und Hörereignisse aus allen Richtungen wahrzunehmen. Immersive Audio wird gerne auch mit den Begriffen *3D-* oder *Spatial Audio* bezeichnet.

2.1.1. Kanalbasiertes Audio

Bei kanalbasierten Audioformaten ist jeder Kanal des Mixes fest einem Lautsprecher zugeordnet. Jede Eingangsgeräuschquelle (in der Abbildung 2.1 dargestellt mit Mikrofonen) entspricht einer Audiospur bzw. einem Kanal. Die verschiedenen Audiospuren werden in einer Digital Audio Workstation (DAW) zusammen gemischt und somit für das Ziellautsprechersetup in einem kanalbasierten Mix kreiert. Im nächsten Schritt folgt die Speicherung auf einem Medium bzw. die Wiedergabe auf Lautsprechern. Es ergibt sich für jeden Lautsprecher eine diskrete Tonspur, also ein fertig gemischtes Signal. Im Fall der Abbildung 2.1 bleiben nach dem Mischen zwei Kanäle erhalten, welche dann auf zwei Lautsprechern wiedergegeben werden können. Neben Stereo haben sich in den vergangenen Jahren zunehmend mehr Surroundformate wie beispielsweise 5.1 oder 7.1 verbreitet und

wurden standardisiert [4]. Mit Hilfe des Phänomens der Summenlokalisation können Phantomschallquellen gebildet werden (vgl. Abschnitt 3.2.2). Auf diese Weise können, je nach Lautsprechersetup, eine Vielzahl an Positionen auf der Horizontalebene dargestellt werden. Das einfachste Beispiel der Summenlokalisation ist die Stereomitte beim 2.0 Stereoformat. Die Stereonorm beschreibt die Aufstellung der Lautsprecher in einem gleichseitigen Dreieck, mit der Abhörposition an der dritten Ecke des Dreiecks. Sobald diese Lautsprecherpositionen garantiert sind, kann über gleiche Pegel des selben Signals auf beide Lautsprecher eine Phantomschallquelle mittig zwischen den Lautsprechern wahrgenommen werden. Dieses Phänomen ist bei Surroundformaten auf anderen Positionen vorhanden, jedoch nicht so ideal wie bei der Stereomitte. Nachdem die Mischungen speziell für ein bestimmtes Format und deren Lautsprecherstellung angefertigt werden, kann es zu Einschränkungen bei der Wiedergabe kommen. Die Lautsprecheraufstellung des Endnutzers wird selten exakt gleich zur Lautsprecheraufstellung des Produzenten sein. Kleine Abweichungen der Positionen der Lautsprecher bewirken bereits große Änderungen in der Klangwahrnehmung [59], [44].

2.1.2. Objektbasiertes Audio

Objektbasiertes Audio ist ein revolutionärer Ansatz um interaktive, personalisierte und immersive Inhalte zu erstellen und bereitzustellen [60]. Das kann ermöglicht werden durch die Repräsentation des Audioinhalts als eine Anzahl an individuellen Objekten in Kombination mit Metadaten, die Beziehungen und Verbindungen der verschiedenen Objekte beschreiben. Als Metadaten bzw. Metainformationen werden strukturierte Daten bezeichnet, die Informationen über andere Informationsressourcen enthalten [1]. Von entscheidender Bedeutung bei objektbasierten Audioformaten sind Metadaten wie z. B. Eigenschaften über die Position des Schallereignisses im Raum, Pegel, Frequenzgang, Schallausbreitung, Bewegungsenergie, Renderinfos, Diffusität, Lautstärke, Bewegung des Objekts, Entfernung und jede weitere Information, die einem Objekt mitgegeben werden möchte. Beim objektbasierten Audio werden die erstellten Spuren statt zu Kanälen, zu Objekten verarbeitet. Damit wird beabsichtigt, dass das Audiosignal gehört werden kann, wie vom

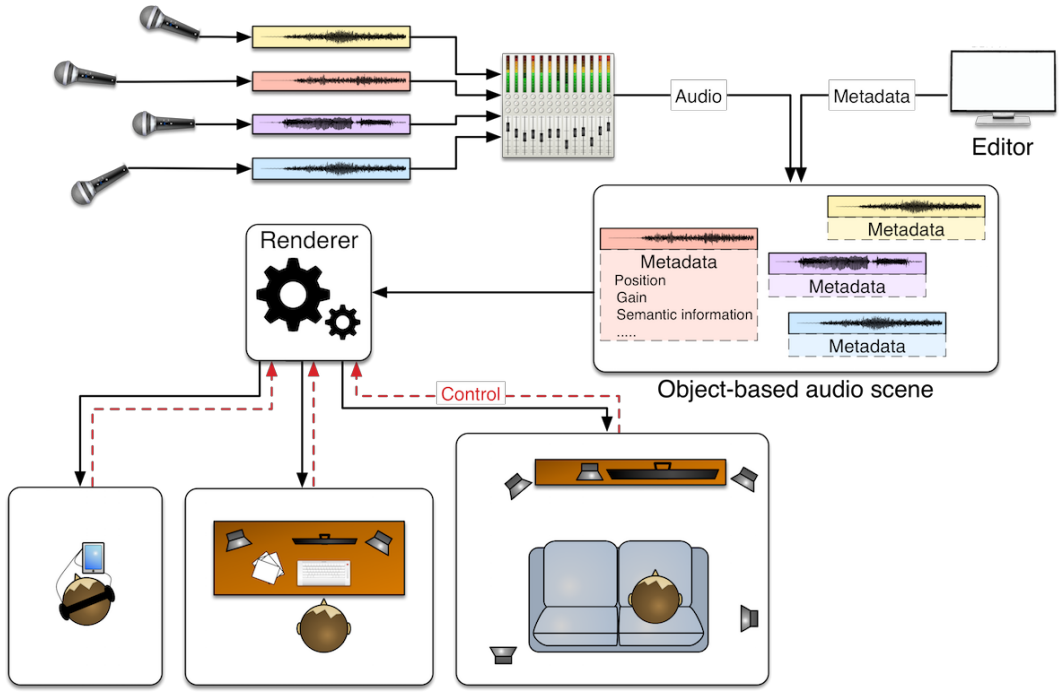


Abbildung 2.2.: Workflow objektbasiertes Audio [44].

Produzent beabsichtigt. Das Ziel ist ein unverfälschtes Hörerlebnis, unabhängig von Endgerät und Umgebung [45]. Es entsteht eine Audioszene aus Objekten ohne ein fest definiertes Zielformat wie z. B. Stereo oder 5.1 Surround. Sobald die Eigenschaften eines Raums bekannt sind, können Lautsprechersignale berechnet werden und die Objekte somit auf ein Kanalbett, als auch individuell im Raum positioniert werden. Um eine objektbasierte Audioszene hörbar zu machen, müssen aus den Metadaten und Audiosignalen erst Lautsprechersignale für ein Lautsprechersystem generiert werden. Der Prozess wird beschrieben durch den Fachterminus *rendern* [44]. Dieser Vorgang kann vor der Ausstrahlung oder im besten Fall direkt zu Hause beim Enddnutzer mit einem entsprechenden Endgerät erfolgen. Ein Produzent ist bei objektbasierten Audioformaten somit frei in seiner Gestaltung, da die Mischung über Objekte auf jedes Zielformat gerendert werden kann. Durch verändern der Metadaten kann eine neue Mischung erstellt werden. In den nächsten Kapiteln wird darauf eingegangen, welche Daten bearbeitet werden dürfen, und bei welchen der Eingriff eher nicht sinnvoll ist. Der entscheidende Vorteil am objektbasierten Audio ist die bleibende Flexibilität in der Nutzung der verschiedenen Audioobjekte und somit auch die Unabhängigkeit des Reproduktionssystems. Es wäre wünschenswert, dass der Endnutzer die Möglichkeit für sich behält, auf Inhalte zuzugreifen, sich die Mischung seinen Wünschen anzupassen und schließlich auf jedem beliebigen Endgerät und Lautsprechersetup wiederzugeben. Hier besteht wiederum eine Abhängigkeit vom Vorrenderprozess, bzw. von der bereitgestellten Datei[54].

2.1.3. Szenenbasiertes Audio

Bei kanalbasierten Audioformaten bildet jeder Kanal auf einen einzelnen Lautsprecher ab. Szenebasierte Audioformate unterscheiden sich von kanalbasierten durch die Darstellung der Audiokanäle auf eine lautsprecherunabhängige Abbildung des Schallfelds. Generell umfasst der Begriff *Ambisonics* und *Higher Order Ambisonics* (*HOA*). Bei Ambisonics werden dreidimensionale Schallfelder bzw. virtuelle Schallquellen übertragen. Je mehr Übertragungskanäle bei der Aufnahme und je mehr Lautsprecher zur Wiedergabe zur Verfügung stehen, desto genauer ist die Dar-

stellung von Schallquellen und die dreidimensionale Auflösung des Schalls. Beim szenebasierten Audio muss das Schallfeld zur Wiedergabe auf ein gewähltes Lautsprecherformat dekodiert werden [59].

2.1.4. Produktion

Die Vielzahl der Möglichkeiten von NGA wird in der Praxis schnell eingeschränkt. Für die Anwendung von NGA ist MPEG-H ein aktueller Standard, der in Unterhaltungselektronik und professionellen Geräten integriert ist. Im Zusammenhang mit zukünftigen TV Standards wurde die maximale Objektzahl für die gleichzeitige Wiedergabe mit MPEG-H auf 16 beschränkt. Wie bereits angedeutet, wird die Objektanzahl reduziert, um auf der einen Seite Daten zu sparen und auf der anderen Seite auch für weniger rechenstarke Endgeräte objektbasiertes Audio zu ermöglichen [23]. Diese besagten 16 Objekte können beispielsweise ein 5.1 oder 7.1 Surroundbett in der normalen Abhörebene beinhalten. Kommen nun weitere vier Kanäle in der Höhe dazu, bleiben im Bezug auf das 7.1 Surround nur noch vier freie Plätze für zusätzliche Objekte. Diese werden bei der Wiedergabe mit dem Audiobett zusammen gerendert. Wie der Decoder das Signal darzustellen hat, wird hauptsächlich über die mitgelieferten Metadaten bestimmt. Weitere Details zu Einschränkungen finden sich in Abschnitt 2.4.1. Problematisch ist die Übertragung mittels neuer Codec-Systeme. Es kann passieren, dass die Metadateninformation nicht zwingend kompatibel zwischen den verschiedenen NGA Codecs ist. Aus diesem Grund wurde das Audio Definition Model von der ITU entwickelt.

2.2. Audio Definition Model

Das Audio Definition Model (ADM) ist ein, in der ITU-R BS.2076 [2], standardisiertes Metadatenmodell, um die technischen Eigenschaften von Audio zu beschreiben. Dabei werden Metadaten zu einer Audiodatei hinzugefügt, um dem Renderer die korrekte Handhabung mit der Audiospur zu überreichen. Das ADM

kann sowohl kanal-, szene-, als auch objektbasierte Audiosignale für immersive Audioformate repräsentieren [55]. Die Notwendigkeit dieses Modells lässt sich aus den immer größeren, komplexeren Lautsprecherformaten und der Personalisierbarkeit, und den damit verbundenen alternativen Mischungen begründen. Hinzu kommt, dass das ADM als offener Standard zum Austausch und zum Archivieren von NGA Inhalten, ohne proprietäre Formate, dient. Wie bereits zu Beginn dieser Arbeit definiert, entsteht das objektbasierte Audioformat über Audiospuren, die mit Metadaten versehen werden. In den Metadaten kann verankert werden, dass ein Objekt z. B. an einer bestimmten Stelle im Raum auftritt oder dass die Sprachen deutsch und französisch zur Verfügung stehen. Aktuell können die ADM Metadaten mit XML beschrieben werden. Darüber hinaus kann das XML jederzeit auf andere Auszeichnungssprachen erweitert werden. In der ITU-R Empfehlung BS.2088 wird auf das Broadcast Wave 64 (*BW64*) Format eingegangen [32]. Es basiert auf dem WAVE File Format und umgeht Einschränkungen des Vorgängerformats *BWF*[30]. Die BS.2076 über das ADM, enthält eine Definition für sogenannte *chunks*, die das Speichern und Übermitteln von Metadaten als XML ermöglichen. Die Hauptabsicht dieser *chunks* ist die Verknüpfung von Audiospuren einer BW64 Datei mit den IDs in den ADM Metadaten. Auf diese Weise gehen notwendige Metadaten nicht mehr verloren. Audio und Metadaten können über ADM und BW64 gespeichert werden. Ein zu beachtender Aspekt ist die Nutzbarkeit dieses Formats. Es muss abgedeckt werden, dass jedes Abspielgerät und jedes Abspielformat bedient werden- und eine mit ADM-Metadaten beschriebene Audioszene verarbeiten kann. Im Idealfall kann jedes Endgerät die mitgegebenen Parameter lesen, Beziehungen untereinander erstellen und diese auf das vorliegende System korrekt rendern. Es ist das empfohlende Format für alle Anwendungsfälle für die Produktion von NGA Inhalten [44].

2.3. EBU ADM Renderer

Das objektbasierte Audio verbindet Metadaten mit Audiokanälen. In diesem Fall wird ein Vorgang benötigt, um die Metadaten und die Audiodatei zu Audiosignalen zu konvertieren, welche dann wiederum an konventionelle Kanäle übertragen

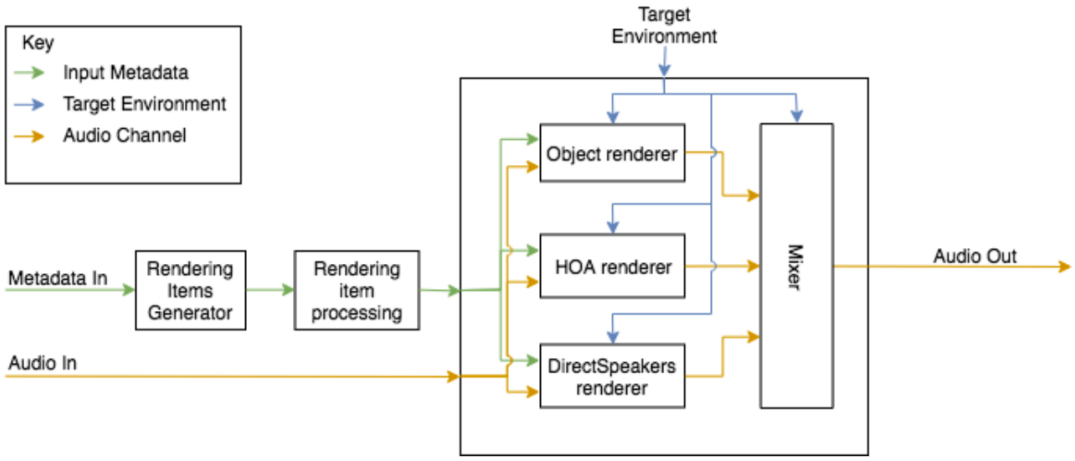


Abbildung 2.3.: Ablauf eines Renderings mit dem EAR [54]

werden können, um Lautsprecher anzusteuern. Dieser Prozess wird Rendering genannt. Sowohl das objektbasierte, als auch das szenebasierte Format sind auf einen Renderer angewiesen. Beim objektbasierten Audio beschreiben Metadaten die räumlichen Positionen der Audiosignale, welche ohne Renderer nur als Monosignale behandelt würden. Beim szenebasierten Audio kann das Rendering durch einen simplen HOA Decoder oder einem anspruchsvolleren Decodierprozess durchgeführt werden. Nachdem das ADM nicht die Reproduktion der Signale übernimmt wird ein Renderer benötigt, der die ADM Metadaten korrekt interpretiert und folglich aus den Inputsignalen und den dazugehörigen Metadaten ein angemessenes Ausgabesignal für das gewünschte Reproduktionssystem erstellt. Der EBU ADM Renderer liefert eine vollständige Interpretation des ADM Formats [55]. Seine Implementierung ist fähig Audiosignale auf alle gängigen Reproduktionssysteme zu rendern. Darin enthalten sind alle Reproduktionsformate der ITU Empfehlung *Advanced sound system for programme production* [31]. Der EBU ADM Renderer kann alle drei NGA Technologien (objektbasiert, szenebasiert, kanalbasiert) verarbeiten. Der Renderer ist ein offener Standard. Funktionalitäten und Algorithmen werden im EBU Tech 3388 weiter referenziert [55].

Abbildung 2.4.: NGA Codec Systeme ermöglichen dem Zuschauer die Audiopräsentation individuell anzupassen. Im Bild auf Basis von MPEG-H [46].

2.4. Distribution

Die vorangegangenen Abschnitte haben die Produktion eines objektbasierten Audioformats behandelt. Dieses Format muss im nächsten Schritt an den Endnutzer geliefert werden. Aktuell werden viele Übertragungskanäle in Betracht gezogen. Um über diese Kanäle Daten zu schicken, muss das Produktionsformat mit der Sendeinfrastruktur (ADM + Audio), in ein effizienteres, streamingfähiges Format konvertiert werden. Zu guter Letzt muss sich das Format für den Transport und die Reproduktion für Empfangsgeräte eignen. Für die Distribution werden eine Reihe von Standards vorgesehen. Als Standard in *DVB* in *UHD* werden Dolbys *AC-4* und das vom Fraunhofer Institut entwickelte *MPEG-H* gehandhabt [34, 7]. Beide besitzen insgesamt gleiche Fähigkeiten. MPEG-H wurde speziell für objektbasierte Formate entwickelt und erlaubt den einfachen Transport von Audio mit Metadaten. Nachfolgend wird näher auf das MPEG-H Format eingegangen.

2.4.1. MPEG-H

MPEG-H ist primär ein Audiokompressionsformat, dass für die Anforderungen des NGAs entworfen wurde. Fraunhofer, Qualcomm and Technicolor haben den Standard MPEG-H entwickelt, um hochqualitative, bitrateneffiziente Überlieferungen und flexiblen 3D-Sound zu reproduzieren [7]. Dieser Standard erlaubt den universellen Transport von enkodiertem 3D-Sound von NGA Formaten. Die Reproduktion wird unterstützt für viele Wiedergabesysteme, angefangen bei Stereo, über 5.1, bis hin zu 22.2 und binauraler Wiedergabe [28]. Das Fraunhofer Institut beschreibt das MPEG-H Format als ein neues TV System für personalisiertes und immersives Audio. Dabei werden die Eigenschaften:

- Interaktivität
- Immersion und
- universelle Übertragung und Wiedergabe

aufgeführt [23]. Zusammengefasst wird der präsentierte Audioinhalt personalisiert, das heißt, dass der Nutzer z. B interaktiv bestimmen kann, welche Sprache das Format hat oder auch, wie die Balance zwischen Sprache und Effekten sein soll. Hinzu kommt die Immersion, also die Ausgabe von Hörereignissen aus allen Richtungen. Der wichtigste Aspekt im Bezug auf dieses Projekt ist jedoch der universelle Transport an Endgeräte. MPEG-H ermöglicht die Wiedergabe auf jedem Abspielgerät mit der bestmöglichen Klanggestaltung. Ein großer Vorteil von MPEG-H ist die Kompatibilität zu den drei genannten NGA Technologien. Unterstützt werden nicht nur Objekte, sondern auch die Kombination aus kanal-, objekt-, und szenebasierten Audioformaten [23]. Dieser neue Standard ist nicht nur ein Codec. Das Fraunhofer Institut spricht sogar von einem Audiosystem, da er Anpassungen zu Lautheit, Rendering und Downmix integriert und somit Stereo, 5.1, 7.1+4H und weitere Formate unterstützt. Wie bereits angesprochen werden mehrere Sprachen, sowie Audiodeskription und Dialogerweiterung als zusätzliche Eigenschaft angeboten. Trotz all dieser Möglichkeiten wird die Audio Datenrate, im Vergleich zu

Tabelle 2.1.: MPEG-H Low Complexity Profile Level 3 [23]

MPEG-H LC Level 3	
Max. Abtastrate	48000
Max. Anzahl an Signalen in einem Datenstrom	32
Max. Anzahl an Decoder verarbeiteten Signalen	16
Max. HOA Ordnung	6
Max. Anzahl an Lautsprecher Ausgabe Kanäle	12

mehreren Komplettmischungen, reduziert.

2.4.2. Einschränkungen und Profile

Das MPEG-H *Low Complexity Profile Level 3* (siehe Tabelle 2.1) wird als Standard bei Rundfunkanwendungen und Fernsehgeräten verwendet [23].

Wie in der Tabelle 2.1 zu erkennen, ist die maximal Anzahl an Objekten für die Wiedergabe beschränkt auf 16. Diese 16 Objekte können sich beispielsweise ergeben aus:

- 16 Objekten oder
- 12 Kanälen + 4 Objekten oder
- 6ter Ordnung HOA + 4 Objekten

Eine häufig verwendete Konfiguration ist die Kombination aus einem oder mehreren Kanalbetten, aufgefüllt mit einzelnen Objekten. So können sich 16 Objekte auch zusammenstellen aus z. B. einem 5.1 Bett, einem Stereo 2.0 Bett und acht Audioobjekten. Ein entscheidender Vorteil der Reduzierung auf 16 Objekten ergibt sich bei der Verarbeitung auf mobilen Endgeräten. Die geringe Rechenleistung von Smartphones und Tablets ist für ein komplexeres Rendering mit mehr als 16

Objekten nicht immer ausreichend [23]. Aus diesem Grund sollte der Prozess der Verabeitung von ADM Metadaten zu MPEG-H Metadaten einfach gestaltet werden bzw. über eine Vorverarbeitung verbessert werden.

3. Vorverarbeitung von NGA Inhalten

Aufgrund von Einschränkungen, wie beispielsweise durch das *MPEG-H Low Complexity Profile* und dem Anliegen diese zu beseitigen, wird im folgenden Kapitel ein Ansatz zur Optimierung durch Vorverarbeitung erläutert. Audioszenen, die mit ADM Metadaten beschrieben sind und vor allem deren Transport bringen Einschränkungen bzw. besondere Funktionalitäten mit sich. Hierzu zählt z. B. der Aspekt der Interaktivität. Eine Audioszene kann beispielsweise interaktive Elemente enthalten, für welche schlussendlich entschieden werden muss, inwieweit diese wichtig sind und ob diese in der Vorverarbeitung erhalten bleiben sollen. Der MPEG-H Codec erlaubt maximal 32 Objekte zu transportieren und nur 16 gleichzeitig wiederzugeben. Von grundlegender Bedeutung ist, die Anzahl der Objekte in einer Audioszene zu verringern, indem ein Teil der Objekte in ein Kanalbett vorgerendert wird. Die Konfigurationen und die Anzahl der Kanalbetten, sowie die Zuordnung von Objekten zu bestimmten Kanalbetten, erlaubt vergleichsweise viele mögliche Kombinationen. Die Auswahl der *besten* Kombination ist schlussendlich ein Optimierungsproblem. Zur Lösung dieses Optimierungsproblems ist eine Kostenfunktion nötig: Eine Funktion, welche die wahrnehmbare Verschlechterung beim Vorrendern eines Objektes in ein Kanalbett berechnet. Erstes Ziel ist daher der Entwurf einer solchen Kostenfunktion. Darauf aufbauend könnte anschließend beispielsweise die Suche nach einem idealen Optimierungsverfahren entstehen.

3.1. Fehlerberechnung von Objektpositionen beim Vorrendern

Eine Kostenfunktion kann den entstehenden Fehler bei der Neustrukturierung von Objektgruppen und Positionen berechnen. Die Kostenfunktion ergibt sich aus dem Vorrenderprozess. Dieser ist eine Bearbeitung des produzierten Signals, bevor dieses im nächsten Schritt an ein Endgerät übergeben wird oder auch anderweitig übertragen wird. Am Anfang steht der Gedanke, dass Audioobjekte zusammengefasst werden sollen, um die Datenrate gering zu halten und dennoch wenig klangliche Verluste zu erzeugen. In der Bearbeitung selbst kann beispielsweise eine Objektposition verändert werden. Folglich kann es während des Vorrenderprozesses zu einer Verschiebung der Objektpositionen kommen. Besagte Objekte befinden sich nach der Bearbeitung z. B. nicht mehr an den vom Produzenten vorgegebenen Originalpositionen. In anderen Fällen können Objekte auch komplett entfernt werden. Dieser Sachverhalt tritt beispielsweise bei unbenutzten Sprachen oder stillen, interaktiven Elementen auf. Ein Audiosignal ohne Inhalt ist redundant und soll dem Endnutzer im Bestfall nicht zur Verfügung stehen. Aus diesem Grund können ungenutzte Signale während der Vorverarbeitung vollständig gelöscht werden. Im nächsten Schritt wird aus einer alten Objektposition eine neue Position für ein bestimmtes Objekt berechnet. Nachdem diese neu errechnete Position meist nicht mit der alten Position übereinstimmt, ergibt sich hieraus eine Abweichung, bzw. ein Fehler. Ein Fehler errechnet sich aus der Differenz aus alter- und neuer Position und gibt die Darstellbarkeit im gewünschten Zielwiedergabesystem wieder. Um dem entstandenen Fehlerwert einen Bezug zu geben, wird dieser im Zusammenhang mit psychoakustischen Erkenntnissen gewichtet (siehe Kapitel 3.7). Zusätzlich spielt die menschliche Richtungswahrnehmung eine entscheidende Rolle. Die räumliche Wahrnehmung des Schalls ergibt sich aus den Grundlagen des räumlichen Hörens.

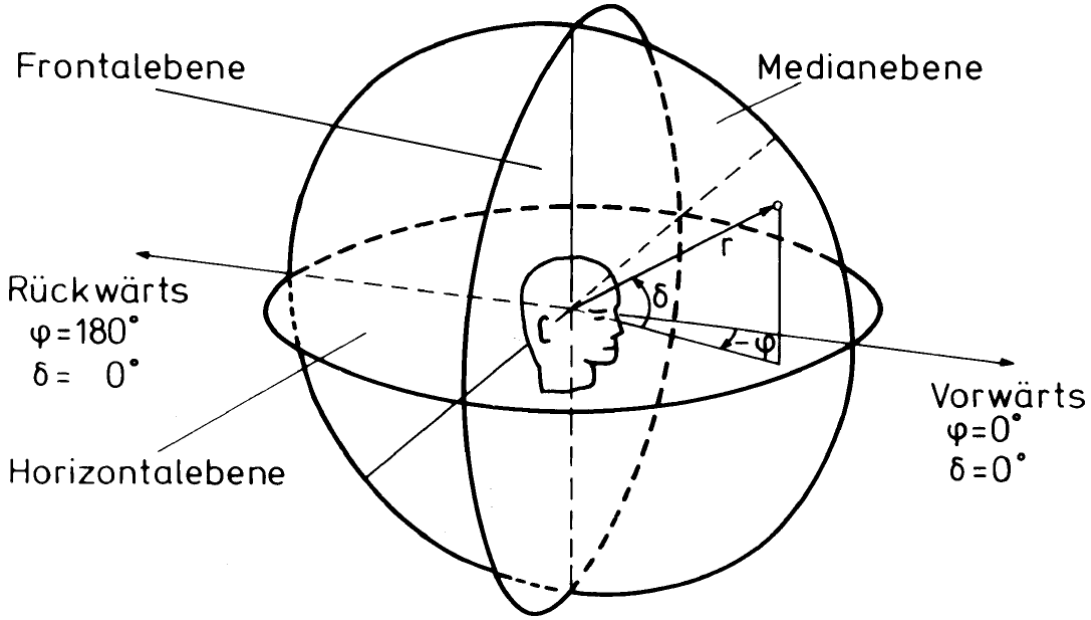


Abbildung 3.1.: Kopfbezogenes Koordinatensystem [8].

3.2. Räumliches Hören

Seit vielen Jahrzehnten sind sich Toningenieure und Künstler über die räumliche Dimension bewusst. Dennoch wurde sich erst in den letzten Jahren aktiver an den Möglichkeiten des menschlichen Hörens zu Entwicklungen in der Rundfunktechnik bedient. Der Mensch hat die Fähigkeit, sein Umfeld wahrzunehmen und mit diesem zu interagieren. Das menschliche Gehhör spielt dabei eine wichtige Rolle. Dabei ist es vollkommen natürlich, Geräusche von überall wahrzunehmen und diese sogar lokalisieren zu können. Visuell betrachtet kann der Mensch nur nach vorne schauen. Alles, was hinter und über ihm passiert, wird über die Ohren wahrgenommen [51]. Den wissenschaftlichen Fachbegriff *räumliches Hören* beschreibt Blauert als „die Beziehung zwischen den Orten sowie die räumliche Ausdehnungen der Hörereignisse untereinander und zu den korrelierten Merkmalen anderer Ereignisse – vorwiegend Schallereignisse“ [8]. Das menschliche Gehör liefert Informationen über Entfernung, Richtung und Räumlichkeit eines Hörereignisses.

Mit dem in 3.1 dargestellten kopfbezogenen Polarkoordinatensystem lassen sich

Hörereignisorte gut darstellen. Der Ursprung des Koordinatensystems liegt auf der interauralen Achse, in der Mitte bei den Gehörkanaleingängen. Die Horizontalebene wird durch die interaurale Achse und die Unterkanten der Augenhöhlen aufgespannt und verläuft orthogonal zur Horizontalebene [8]. Die drei räumlichen Ebenen (Horizontalebene, Frontalebene und Medianebene) verlaufen alle durch den Ursprung. Wie in Abbildung 3.1 zu sehen ist, wird die Position eines Schallereignis, bzw. die Richtung aus der es kommt, mit Polarkoordinaten angegeben. Der Azimuthwert φ gibt die Position auf der Horizontalebene an, der Elevationswert θ die Höhe auf der Medianebene. Um aus psychoakustischer Sicht Hörereignisorte zu bestimmen, wird zwischen zwei Ohrsignal-Merkmalsklassen unterschieden. Bei monauralen Ohrsignalen reicht ein Ohr um Schallwellen zu empfangen. Folglich sind bei interauralen Ohrsignalen beide Ohren notwendig [8]. Diese werden in drei verschiedene Arten untergliedert:

- Monotisch: wenn nur ein Ohr beschallt wird.
- Diotisch: wenn zwei Ohren identisch beschallt werden und
- Dichotisch: wenn beide Ohren unterschiedlich beschallt werden.

3.2.1. Lokalisation

Auch beim objektbasierten Audio spielt die Lokalisation eine wichtige Rolle. Sie ist Bestandteil des räumlichen Hörens und wird durch den physischen Aufbau des menschlichen Schädels bestimmt. Das menschliche Gehör greift auf verschiedene Verfahren zurück um Schallquellen zu lokalisieren. Auf akustischer Ebene arbeitet das Ohr quasi als Druckempfänger. Somit kann es keine Informationen über die Richtung erfassen. Nur durch die Form von Rumpf, Kopf und Ohrmuscheln kommt es zu Ablenkungen von Schallsignalen, wodurch über unterschiedliche Frequenzverzerrungen eine grobe Lokalisation ermöglicht wird. Die Lokalisation ist laut Definition das Gesetz oder die Regel, für welche die Lokalisation eines auditiven Ereignisses zu einem bestimmten Attribut eines Schallereignisses zugeord-

net wird [9]. Lokalisation erfolgt auf psychoakustischer Ebene durch interaurale Zeitdifferenzen und interaurale Pegeldifferenzen. Wie die Namen bereits vermuten lassen, werden hier die Unterschiede der beiden Ohrsignale im Bezug auf einerseits den Schalldruckpegel und andererseits den Zeitpunkt des Eintreffens betrachtet [8]. Laufzeitdifferenzen (engl. Interaural Time Difference) entstehen, wenn Schallquellen das eine Ohr, zeitlich vor dem anderen Ohr erreichen. Das ist der Fall, wenn der Schall zum abgewandten Ohr einen längeren Weg zurücklegen muss. Das menschliche Gehirn kann aus dieser Zeitdifferenz den Ursprung des Schalls errechnen. Signale von vorne sind sehr gut zu orten, da berechnet wird, wann das Geräusch beim linken und wann beim rechten Ohr ankommt. Die Berechnung der Laufzeitdifferenz basiert auf einem Ohrabstand von 17cm. Bei einem Einfallswinkel von 90° entspricht das einem maximalen Schallumweg von 21cm. Die Laufzeitdifferenz zwischen den Ohren beträgt somit zwischen 0μs und 650μs. Die kleinste wahrnehmbare Laufzeitdifferenz beläuft sich auf 10μs. Das entspricht einer Richtungsänderung von ungefähr 1° [26]. Neben Laufzeitdifferenzen treten Pegeldifferenzen (engl. Interaural Level Difference) auf. Wird auf einem Ohr ein stärkerer Pegel als auf dem anderen wahrgenommen, wird dem Hirn automatisch suggeriert, dass der Ton auch aus dieser Richtung kommt. Auf der einen Seite des Schädels entsteht ein Druckstau, auf der anderen Seite ein sogenannter Schallschatten. Folglich erreicht das schallzugewandte Ohr höhere Pegel, als das andere. Bei hohen Tönen mit kleiner Wellenlänge können jedoch Pegelunterschiede von bis zu 35 dB auftreten, die die interaurale Pegeldifferenz für die Lokalisation prädestinieren [42]. Pegeldifferenzen sind für Frequenzen unter 2kHz schwierig festzustellen. Dies wird erklärbar dadurch, dass niederfrequente Töne eine große Wellenlänge besitzen und ein Gegenstand von der Größe eines Kopfes kaum ein Hindernis für die Ausbreitung dieser Wellenlänge darstellt, sondern um das Hindernis gebeugt werden kann [20]. Insgesamt ist das menschliche Gehör in der Lage Pegeldifferenzen zu erkennen und in Richtungsinformationen umzusetzen [17]. Pegeldifferenzen kommen insgesamt kaum zum Tragen. Zeitdifferenzen sind ohne Probleme erkennbar.

Jens Blauert unterscheidet räumliches Hören zwischen einer und mehrerer Schallquellen. Bei einer Schallquelle und Schalleinfall aus der Medianebene ist die Lokalisation nicht einfach, was auf die Symmetrie des Kopfes zurückzuführen ist. Es gibt kaum interaurale Zeitdifferenzen (ITD) und interaurale Pegeldifferenzen (ILD). Es

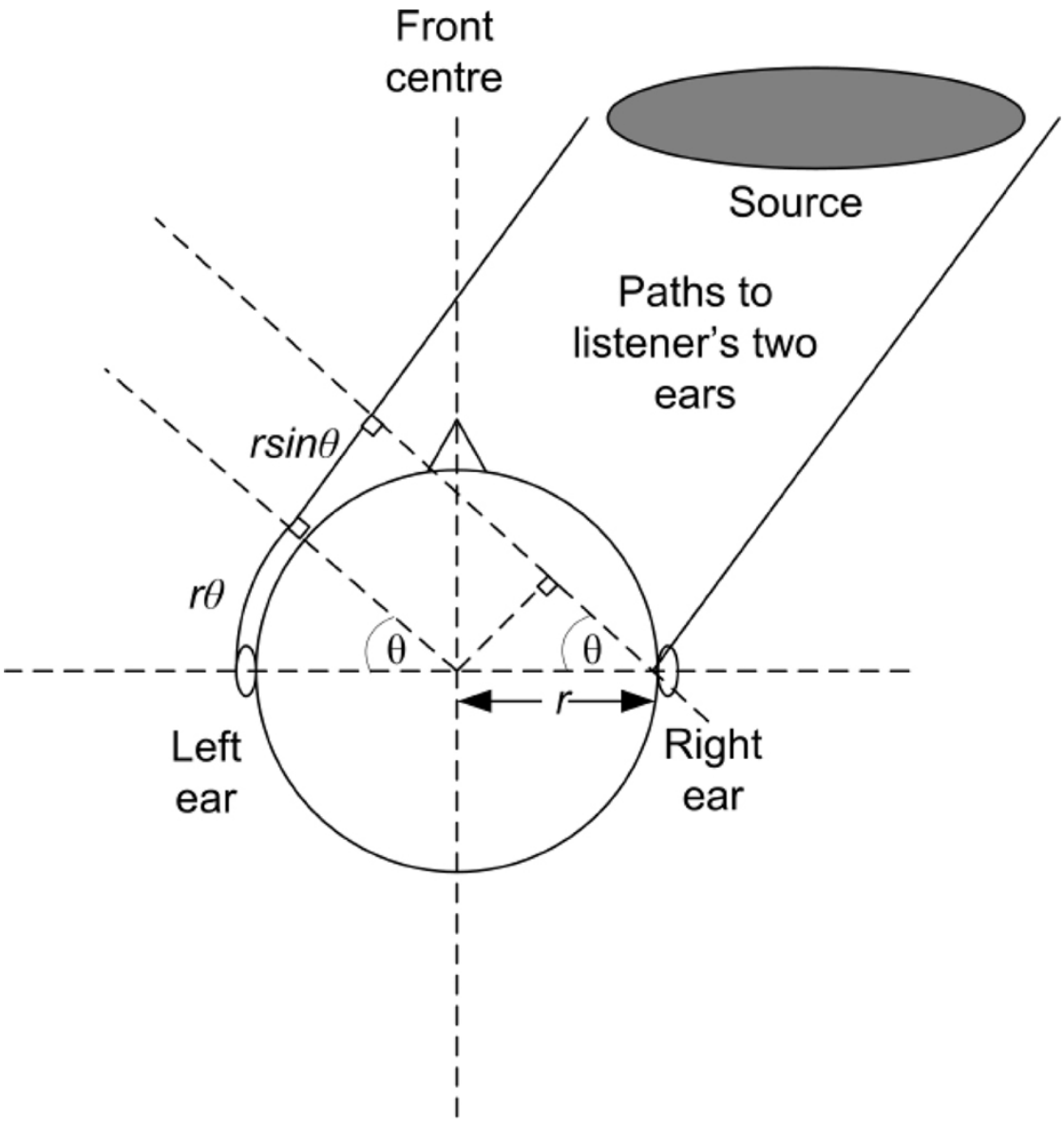


Abbildung 3.2.: Die Laufzeitdifferenz (engl. interaural time difference (ITD)) hängt vom Einfallswinkel des Schallereignisses ab. Dieser ist der Grund für eine zusätzliche Distanz, welche die Schallwelle zum weiter entfernten Ohr zurücklegen muss. In der Abbildung ergibt sich der ITD aus $r(\theta + sin\theta)/c$ [51].

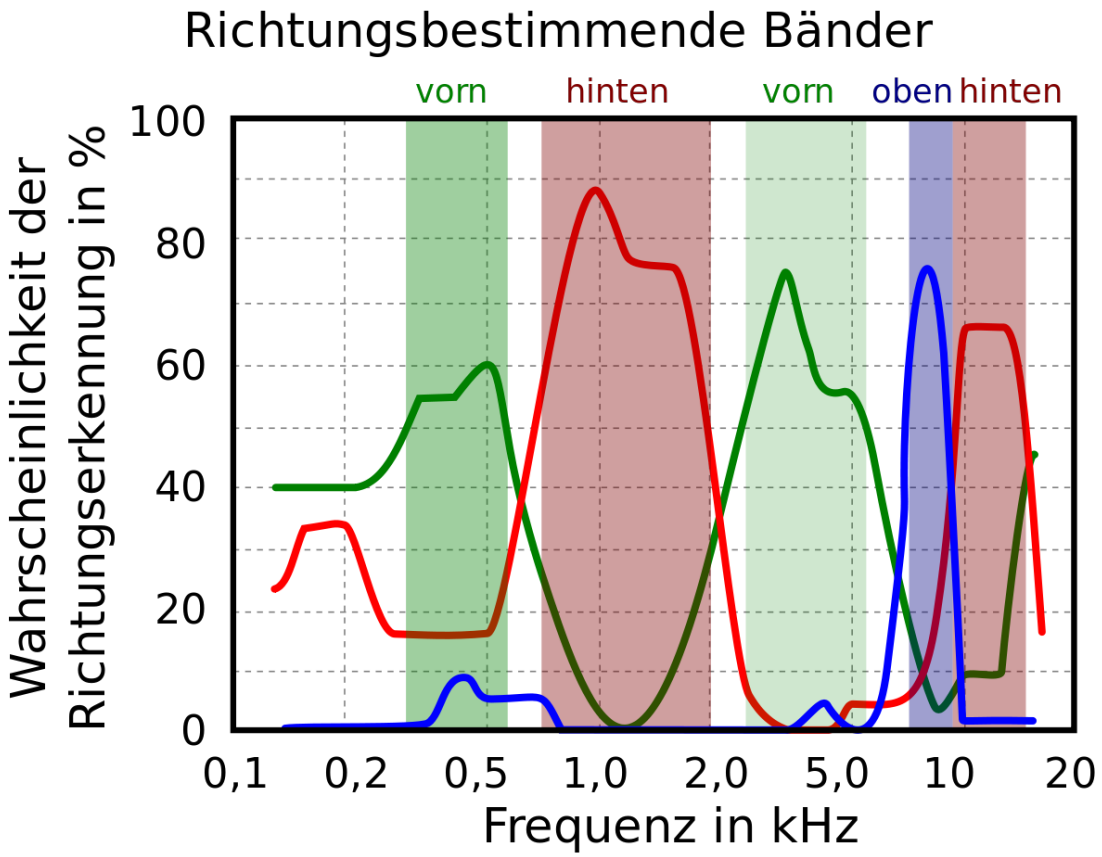


Abbildung 3.3.: Richtungsbestimmende Bänder nach Blauert. Die Abbildung gibt die Wahrscheinlichkeit für die Richtungserkennung in Abhängigkeit zur Frequenz an [8].

kann von einem diotischen Schalleinfall gesprochen werden, da beide Ohren sehr ähnlich beschallt werden.

Die Richtungswahrnehmung ist auf die richtungsbestimmenden Bänder zurückzuführen (siehe Abbildung 3.3). Sobald ein Schallsignal auf das menschliche Ohr trifft, wird es aufgrund der Beschaffenheit des Außenohrs und des Gehörgangs gefiltert. Es entstehen Filter, die von der Schalleinfallsrichtung abhängig sind.

Blauert fand heraus, dass die Hörereignisrichtung nicht von der Schalleinfallsrichtung, sondern von der Terzmittenfrequenz abhängt. Bei Breitbandsignalen ist dies

Tabelle 3.1.: Mittelwerte des absoluten Azimuth- und Elevationfehlers. (Schallquelle fixiert, keine Störgeräusche [27].)

	Azimuth	Elevation
Pink noise	6.6	17.5
Pink Floyd	8.3	12.1
Frog croaks	7.8	13.5
All signals	7.9	15.0

jedoch auch meist die Schalleinfallsrichtung. Erklären lässt sich das durch die Filterwirkung der Außenohren, bei der bei einem Breitbandsignal bestimmte Spektralanteile angehoben bzw. abgesenkt werden. Bei Ohrsignalen um die 8 kHz, wird das Hörereignis oberhalb der Horizontalebene wahrgenommen [8].Fällt der Schall aus Richtungen seitlich zur Medianebene ein, resultieren Unterschiede zwischen beiden Ohrsignalen [8]. Die Folge für das Gehör sind monaurale und auch interaurale Ohrsignalmerkmale, die wahrgenommen werden können. Das führt wiederum zu einer vergleichsweise guten Lokalisation (vgl. hierzu Absatz 3.6).

3.2.2. Phantomschallquellen

Unter einer Phantomschallquelle wird ein Schallereignis vestanden, welches aus einer Richtung lokalisiert wird, an der kein Lautsprecher vorhanden ist [17]. Dieses Phänomen tritt auf, wenn zwei Lautsprecher mit einem bestimmten Abstand zueinander ein exakt gleiches Signal abstrahlen und der Zuhörer anstatt zwei getrennten Schallquellen eine einzige fiktive Quelle zwischen den beiden Lautsprechern wahrnehmen kann (vgl. hierzu Abbildung 3.4). Je nach Pegel- und Laufzeitdifferenzen zwischen den beiden Lautsprechersignalen ändert sich auch die Position der Phantomschallquelle entlang der Lautsprecherbasis. Folglich sind Phantomschallquellen abhängig von Pegel- oder Laufzeitdifferenzen zwischen zwei oder mehreren Lautsprechern. Präzision und Lokalisierbarkeit sind verbunden mit der Art der Entstehung, sowie der Signal- und Wiedergabekonfiguration [25].

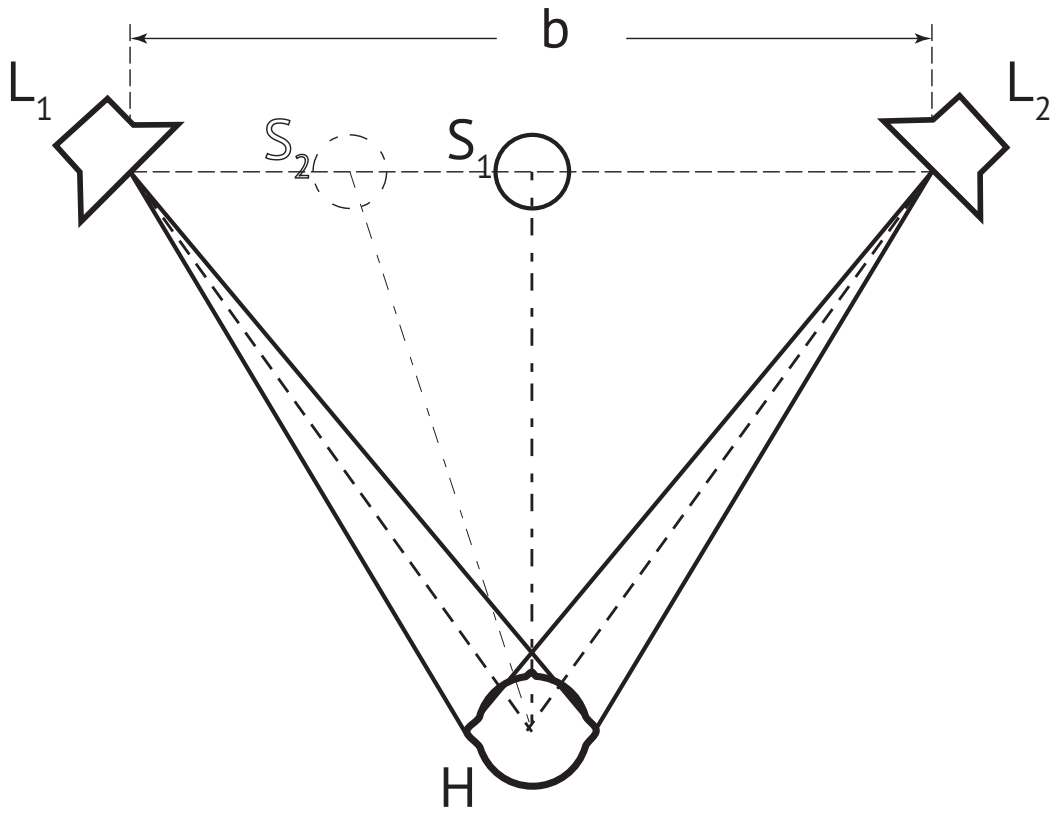


Abbildung 3.4.: Lautsprecheranordnung in einem gleichschenkligen Dreieck mit dem Hörer H, Basisbreite b und Schallereignis S_1 bzw. S_2 für Zweikanal-Stereowiedergabe [17].

Tabelle 3.2.: Mittelwerte des absoluten Azimuth- und Elevationfehlers. (Schallquelle bewegt, Störgeräusche [27].)

	Azimuth	**Elevation**
Pink noise	17.4	28.3
Pink Floyd	18.5	24.9
Frog croaks	15.3	22.8
All signals	17.0	25.4

Tabelle 3.3.: Ergebnisse eines Experiments zur Lokalisationsgenauigkeit nach Gröhn [27]

	Azimuth	Elevation	Error angle
Real Source	4.5	5.7	8.5
VBAP	4.9	11.6	15.5
Dynamic experiment	13.4	11.5	20.8

Phantomschallquellen lassen sich einfach bilden, doch sind sie für den Menschen nicht an jeder Stelle des Raumes gleich gut lokalisierbar. Die Lokalisation wird z. B. bei einohrigem Hören gestört. Ein ähnlicher Fall tritt bei einer seitlichen Phantomschallquelle auf, bei welcher der Schall sich laut Theile „im binauralen Korrelationsmuster nicht ausreichend unterscheiden lässt“ [57] (vgl. hierzu Abbildung 3.5).

In diesem Zusammenhang erreichen die Ohren zwei unterschiedliche Signale. Speziell bei breitbandigen Signalen tritt dieses Phänomen auf, wenn eine zu große oder zu kleine Zeitdifferenz der überlagerten Signale vorliegt. Zur Veranschaulichung wird ein klassisches Stereo Setup aufgestellt, nur die Versuchsperson ist um 90° zur Seite gedreht (vgl. hierzu den Fall 90° in Abbildung 3.5). In diesem Fall ist die Zeitdifferenz der überlagerten Signale zu gering, und somit sind die Signale nicht unterscheidbar. Folglich gibt es hier keine seitlichen Phantomschallquellen. Je weiter der Kopf sich der Ausgangsstellung nähert, desto größer wird auch die Zeitdifferenz. Die Bildung von Phantomschallquellen wird wieder möglich [57]. Ein weiterer Gesichtspunkt der Lokalisation, ist das Entfernungshören. Dieses hat für die durchgeführten Versuche keine besonders große Rolle gespielt und wird aus diesem Grund an dieser Stelle vernachlässigt. Mit der Entfernung eines Schallereignisses zum Ohr, verändert sich auch deren Klang, bzw. die Klangwahrnehmung.

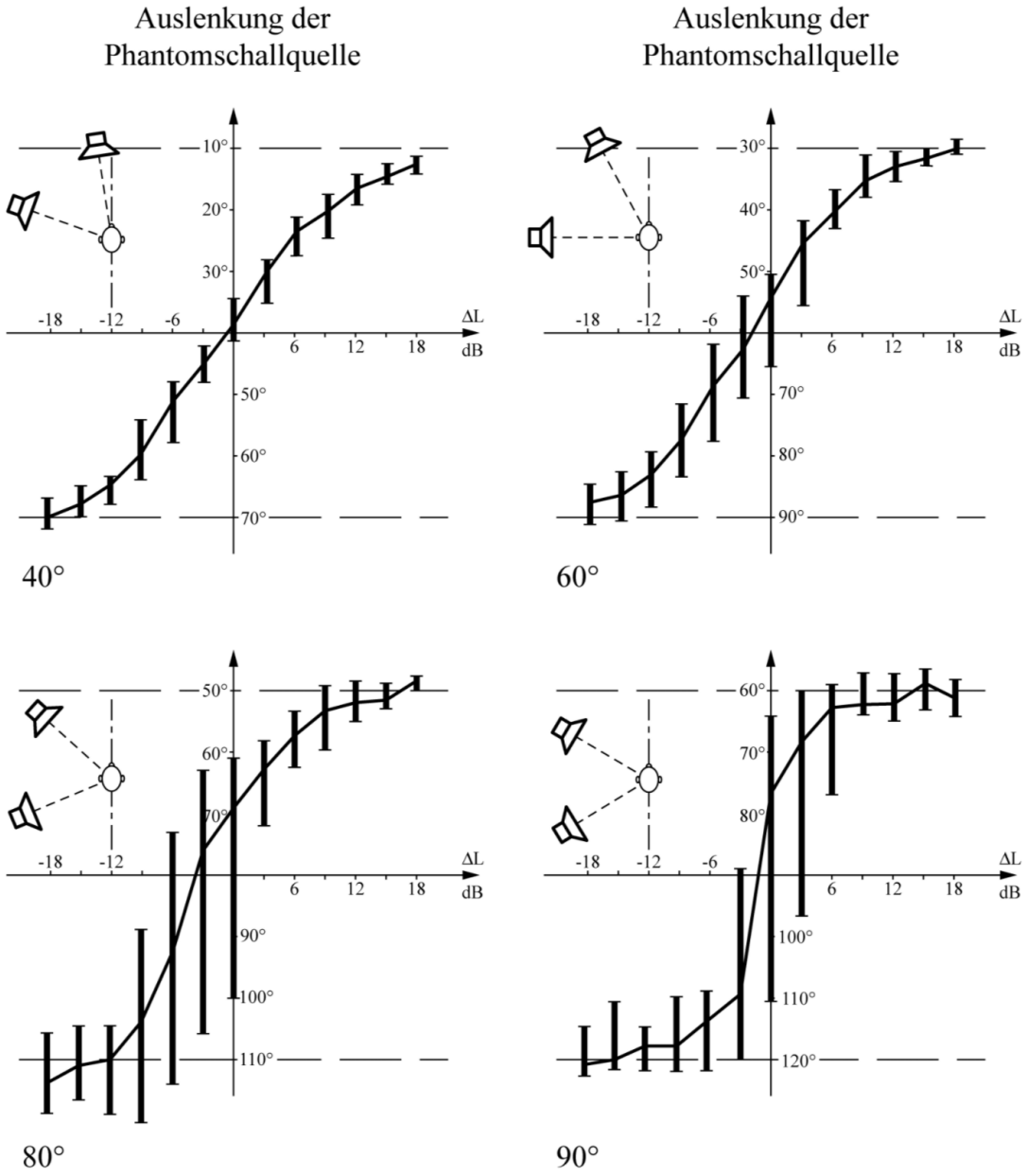


Abbildung 3.5.: Auslenkung der Phantomschallquellen mit ihren Unschärfebereichen bei Pegeldifferenzen in Abhängigkeit vom Ausrichtungswinkel zum Hörer [56].

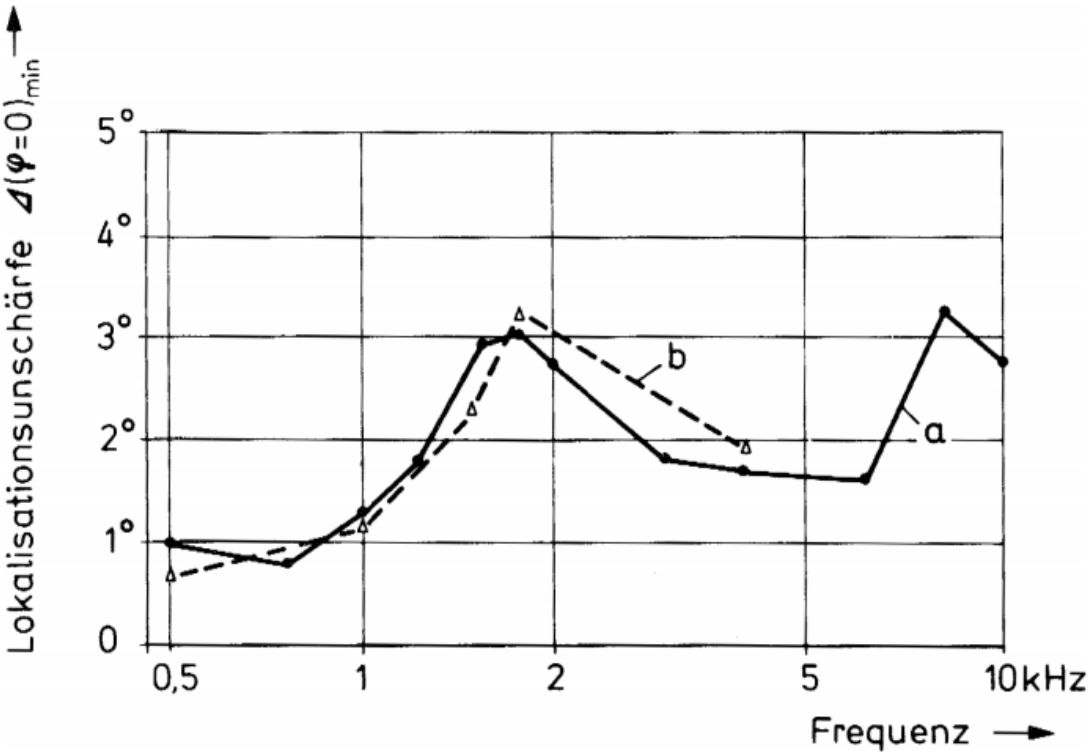


Abbildung 3.6.: Lokalisationsschärfe in Abhängigkeit von der Frequenz. Kurve a: Dauertöne, Kurve b: Gauß-Töne [8].

3.3. Fehlergewichtung

Aus Absatz 3.2 lässt sich schlussfolgern, dass Schallsignale in der Horizontalebene besser lokalisierbar sind als in der Medianebene. Der physische Aufbau des menschlichen Körpers und seine räumliche Wahrnehmung ist für die Orientierung in der Horizontalebene optimiert. Hier wird die maximale Lokalisationsschärfe erreicht [25]. Eine direkte Abhängigkeit von der spektralen Zusammensetzung des Signals bleibt aus. Jedoch werden Signale mit spektralen Anteilen über 6 kHz besser lokalisiert [8].

Um auf den Prozess der Vorverarbeitung von NGA Inhalten zurückzukommen, wird im Folgenden anhand eines Beispiels die, bereits in 3.1 angeführte, Gewichtung erläutert. Das Ziel dieser Gewichtung ist hauptsächlich die Bewertung von

Abweichungen mit einer sinnvoll gewählten Wertung. Ohne Gewichtung könnte keine Beziehung zwischen den berechneten Abweichungen hergestellt werden. Von einem verschobenen Lautsprecher oder sogar einem fehlenden Lautsprecher im Wiedergabesystem ausgehend, wird der Ablauf des Vorrenderns bestimmt. Ein fehlender Lautsprecher rechts vorne gibt in der Konsequenz ein verschobenes Klangbild wieder. Rechts vorne platzierte Objekte werden nicht wiedergegeben. In der Vorverarbeitung besteht die Möglichkeit einzugreifen. Umliegende Lautsprecher können so angesteuert werden, dass über Phantomschallquellen Signalanteile aus der besagten Richtung erzeugt werden können. Die Voraussetzung ist, dass eine Mitteilung an den Renderer über den fehlenden Lautsprecher erfolgt. In der Fehlerberechnung entsteht in der Konsequenz ein geringerer Fehler für die vorverarbeitete-, im Vergleich zur nicht bearbeiteten Version, da Objekte trotz des fehlenden Lautsprechers, dargestellt werden können. Im nächsten Schritt werden die Fehler gewichtet. Nach psychoakustischen Erkenntnissen ist die menschliche Wahrnehmung und Lokalisation in der Horizontalebene, besonders bei frontalem Hörereignis, sehr gut. Folglich sollten die Fehler aus frontaler Richtung auch stärker gewichtet werden. Ein fehlender Lautsprecher in der Front würde somit sofort auffallen und einen großen Fehler ausgeben.

3.4. Räumliche Klangwiedergabe - Das Vector Base Amplitude Panning

Der Begriff räumliche Klangwiedergabe (engl. *spatial sound reproduction*) bezeichnet Methoden um räumliche Klangelemente beim Zuhörer abzubilden [50]. Rückblickend auf die Geschichte der Mehrkanalstereofonie, wurden bereits im Mono-Zeitalter erste Versuche unternommen, um Schallquellen im Raum zu verteilen. Mit Hilfe des traditionellen Panorama Potentiometers wurden Monosignale auf Stereokanäle aufgeteilt. Das Vector Base Amplitude Panning kann als dreidimensionale Erweiterung des Panorama Potentiometers angesehen werden [59]. Wie der Begriff VBAP bereits andeutet, handelt es sich auf der einen Seite um eine vektorbasierte Methode. Auf der anderen Seite enthält der Name VBAP *Amplituden*

Panning. Beim Amplituden Panning werden Audiosignale mit unterschiedlichen Amplituden Lautsprechern zugeordnet. Ausgangspunkt für das VBAP ist eine unbegrenzte Anzahl an Lautsprechern im dreidimensionalen Raum. VBAP kann diese Lautsprecher nutzen, um Schallquellen im Raum darzustellen. Das Hauptziel des Amplituden Pannings, ist die Kontrolle über die Phantomschallquellenlokalisation. Ausgehend vom zweidimensionalen Amplituden Panning, welches auch Intensitätspanning genannt wird, entwickelte Pulkki durch mathematische Formalisierung das VBAP [50]. Bei der Methode des zweidimensionalen Amplitudenpannings können Hörereignisse zwischen zwei Lautsprechern durch die Veränderungen ihrer Signalamplituden platziert werden. Die Signalamplituden können durch die Anpassung der Verstärkungsfaktoren (in Formel 3.1 mit g_1, g_2 gekennzeichnet) kontrolliert und verändert werden. Durch eine Veränderung bzw. Normierung der Verstärkungsfaktoren wird die Summe ihrer Leistungen und somit die Lautheit konstant gehalten.

$$g_1^2 + g_2^2 = C \tag{3.1}$$

Die vorliegende Abbildung 3.7 zeigt ein zweikanalstereo Setup. Der stereophone Aufbau ist die am häufigsten genutzte Abhörsituation [49]. Dabei sind zwei Lautsprecher vor dem Hörer platziert. Der Winkel zwischen den Lautsprechern beträgt 60°. Bei Vernachlässigung der Entfernung der virtuellen Quelle, kann ein virtueller Bogen gespannt werden (gestrichelte Linie), auf der sich die virtuelle Schallquelle befinden kann. Der Radius entspricht dem Abstand des Hörers zum Lautsprecher. Der Bereich, auf dem sich die Quelle befinden kann, nennt Pulkki aktiver Bogen [50].

Panning laws definieren die Beziehung zwischen der gewünschten Position und dem Pegel eines jeden Lautsprechers.

$$\frac{\sin\theta}{\sin\theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{3.2}$$

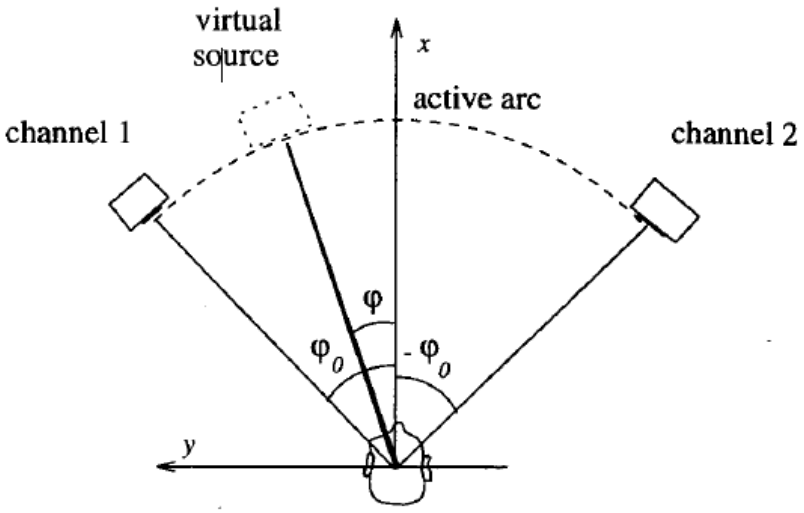


Abbildung 3.7.: Zweikanalige stereophone Konfiguration. Zwischen den beiden Lautsprechern bildet sich ein aktiver Bogen, auf dem sich die virtuelle Quelle befinden kann [50].

Das erste Model für frontales, paarweise Amplituden Panning war der *Sinussatz* [21]. Dieser Satz bzw. diese Gleichung aus der Trigonometrie zeigt nach Bauer und Blumlein eine Verbindung zur stereophonen Richtungswahrnehmung einer virtuellen Schallquelle auf [5, 10]. Dabei wird die Phantomschallquellenposition über die Lautsprecherpegel vorhergesagt. Der Sinussatz beruht auf einem geometrischen Kopfmodell, ohne die Schallausbreitung um den Kopf zu berücksichtigen. Der Winkel zwischen der x-Achse und der Richtung der virtuellen Quelle in 3.2 entspricht θ. Beide Lautsprecher sind symmetrisch zu X und befinden sich bei $\pm\theta_0$. θ_0 beschreibt den Winkel zwischen der x-Achse und den Lautsprechern. Es wird angenommen, dass der Hörer mit dem Kopf nach vorne schaut. Folgt der Hörer den Schallereignissen liefert das *tangent law* genauere Werte [50].

$$\frac{\tan\theta}{\tan\theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{3.3}$$

Das *tangent law* basiert auf einem akustischen Modell von Leakey und Bennett, welches „den Hörereignisort der Phantomschallquelle [...] voraussagt“ [59]. Dieses einfache, geometrische Kopfmodell ist das beliebteste panning Gesetz für paarwei-

ses panning [22]. Im Vergleich zum Sinussatz wird beim *tangent law* der Weg der Schallausbreitung um den Kopf miteinbezogen. Es liefert etwas präzisere Ergebnisse, was in der Praxis jedoch vernachlässigt werden kann [35]. Aus dem *tangent law* entwickelte Pulkki mit VBAP eine Vektorform, die auch die Berechnung nichtsymmetrischer Lautsprecherrichtungen integriert. Durch Gleichung 3.3 lässt sich der Hörereignisort der Phantomschallquelle, mit Hilfe der Verstärkungsfaktoren g_1, g_2 der benachbarten Lautsprecher und dem Winkel θ_0 der Lautsprecherbasis ermitteln. Aus Gleichung 3.3 geht für jede Hörereignisrichtung nur ein Verhältnis der Verstärkungsfaktoren g_1, g_2 hervor. Um die Lautheit der Phantomschallquellen konstant zu halten und konstante Energie für alle Panningrichtungen zu garantieren, kann zusätzlich über die Gesamtenergie normiert werden (siehe Gleichung 3.4).

$$g_1^2 + g_2^2 = 1 \tag{3.4}$$

Die gleiche lineare Summation der gewichteten Lautsprecherrichtungen wird angewendet um den sogenannten Velocity Vector $\boldsymbol{r}_V$ (vgl. Gleichung 3.5 zu berechnen [36, 24]). Die Vektorberechnung funktioniert am besten für Signale mit Spektralanteilen unter 700 Hz.

$$\boldsymbol{r}_V = \frac{\sum_{l=1}^{L} g_l \theta_l}{\sum_{l=1}^{L} g_l} \tag{3.5}$$

Bei der vektorbasierten Betrachtung wird aus der zweikanaligen Stereolautsprecherkonfiguration eine zweidimensionale Vektorbasis [50]. Im übertragenen Sinne zeigt der Einheitsvektor $\boldsymbol{p}$ auf die virtuelle Schallquelle. Die Basis wird aufgestellt über die Einheitsvektoren $\boldsymbol{l}_1 = [\boldsymbol{l}_{11}\ \boldsymbol{l}_{12}]^T$ und $\boldsymbol{l}_2 = [\boldsymbol{l}_{21}\ \boldsymbol{l}_{22}]^T$, welche auf die beiden Lautsprecher zeigen (vgl. Abb. 3.8). Daraus ergibt sich für $\boldsymbol{p} = [\boldsymbol{p}_1\ \boldsymbol{p}_2]^T$.

$$\boldsymbol{p} = g_1 \boldsymbol{l}_1 + g_2 \boldsymbol{l}_2 \tag{3.6}$$

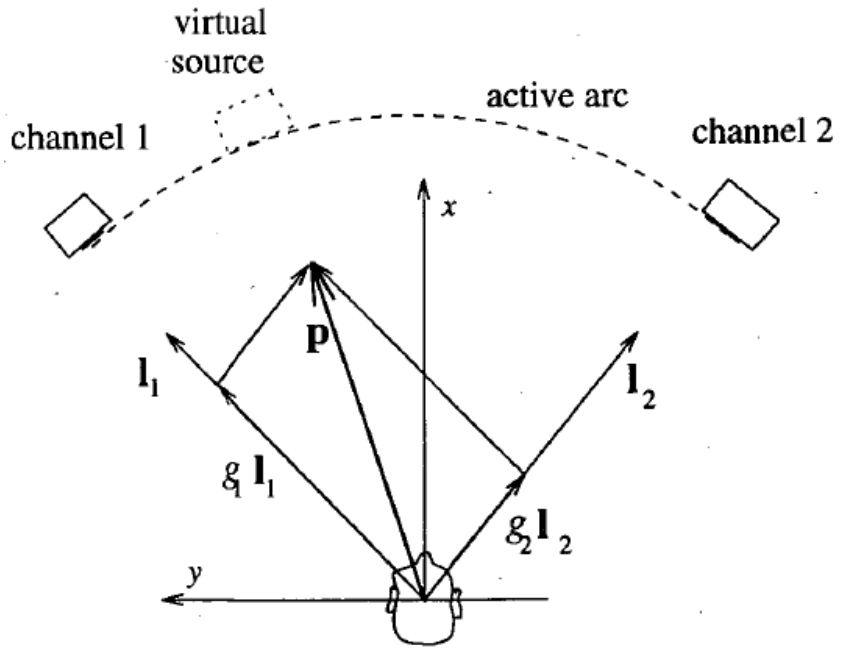


Abbildung 3.8.: Zweikanalige stereophone Konfiguration mit Vektoren dargestellt. Die Einheitsvektoren $\boldsymbol{l_1}$ und $\boldsymbol{l_2}$ zeigen auf die Lautsprecher. Der Einheitsvektor $\boldsymbol{p}$ entspricht der Linearkombination der Lautsprechervektoren und zeigt auf die virtuelle Quelle. [50].

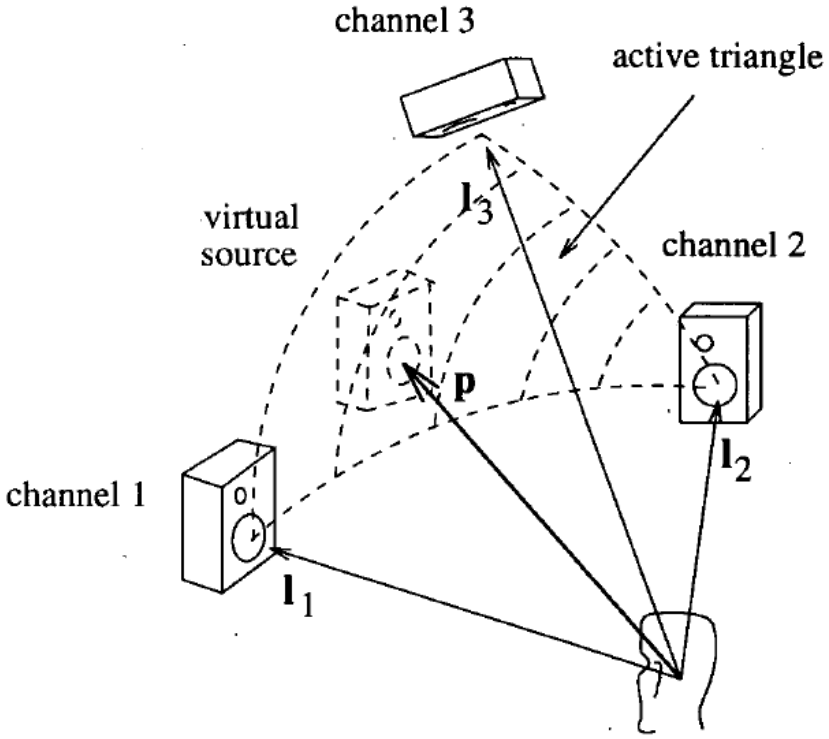


Abbildung 3.9.: Das sogenannte aktive Dreieck wird aufgestellt durch die drei Einheitsvektoren $\boldsymbol{l}_1$, $\boldsymbol{l}_2$ und $\boldsymbol{l}_3$. Der Einheitsvektor $\boldsymbol{p}$ zeigt auf die virtuelle Quelle. Diese befindet sich auf dem aktiven Dreieck. Das aktive Dreieck wird aus drei Lautsprechern gebildet, welche vom Zuhörer aus in einer Dreiecksformation zu erkennen sind [50].

Bisher wurde nur von maximal zwei Lautsprechern ausgegangen. Das Prinzip kann jedoch auch auf größere Lautsprechersetups angewendet werden. Bei einer größeren Anzahl an Lautsprechern, wird das Lautsprechersetup in Dreiecke aufgeteilt. Es wird ein Dreieck ausgewählt und schließlich wie beim zweidimensionalen Amplituden Panning vorgegangen [49]. Für die dritte Dimension kommt lediglich ein weiterer Verstärkungsfaktor hinzu. Auch aus dem aktiven Bogen, wird ein aktives Dreieck.

$$g_1^2 + g_2^2 + g_3^2 = C \tag{3.7}$$

Der Begriff *dreidimensionales Amplitudenpanning* ist eine Methode, um eine virtu-

elle Schallquelle in einem Dreieck darzustellen. Hierfür werden drei reale Schallquellen benötigt, welche mit unterschiedlichen Signalamplituden ein Dreieck formen. Das Verhältnis der drei Verstärkungsfaktoren definiert die, vom Zuhörer wahrgenommene, virtuelle Quellrichtung. Auf diese Weise wird die virtuelle Quelle auf die dreidimensionale Sphäre platziert. Im dreidimensionalen Raum werden Signalpegel über VBAP für benachbarte Lautsprechertripel berechnet. Diese sind abhängig von der Abstrahlrichtung der Lautsprecher und werden über die Richtungsvektoren $\boldsymbol{l}_m$, $\boldsymbol{l}_n$ und $\boldsymbol{l}_k$ gegeben. Es wird sowohl beim zwei- als auch beim dreidimensionalen VBAP versucht, im Hinblick auf den Stereoaufbau, eine Summenlokalisation durchzuführen und somit Phantomschallquellen (im 3D-Raum) zu generieren. Ähnlich zum zweidimensionalen VBAP werden im 3D-Raum aus dem Vektortripel $\boldsymbol{L} = \{\boldsymbol{l}_1\ \boldsymbol{l}_2\ \boldsymbol{l}_3\}$ und der zugehörigen Gewichtung der Lautsprecher mit $g = [g_1\ g_2\ g_3]^T$, die Verstärkungsvektoren für eine gegebene Position einer Phantomschallquelle berechnet (vgl. Gleichung 3.8).

Mit der Linearkombination aus den Verstärkungsfaktoren g_1, g_2, g_3 und den drei Lautsprechervektoren $\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3$, lässt sich der virtuelle Quellvektor p ausdrücken (3.8).

$$\boldsymbol{p} = g_1\boldsymbol{l}_1 + g_2\boldsymbol{l}_2 + g_3\boldsymbol{l}_3 \tag{3.8}$$

Neben diesem Anwendungsfall kann das dreidimensionale VBAP auch auf Systeme mit mehr als drei Lautsprechern angewendet werden (vgl. hierzu Abbildung 3.10). Es sollte beachtet werden, dass sich die aktiven Dreiecke der Basen nicht kreuzen.

3.5. Fehlerberechnung

Die Vorverarbeitung im Allgemeinen wurde bereits in Kapitel 3.1 behandelt. Im folgenden Abschnitt wird erläutert, wie die Bestimmung und Abweichung von Positionen genutzt werden können um die Vorverarbeitung zu beeinflussen. Um den wahrnehmbaren Unterschied bzw. den Fehler zu berechnen, wird auf den EAR, den EBU ADM Renderer, zurückgegriffen. Der EAR wird mit einer Klasse er-

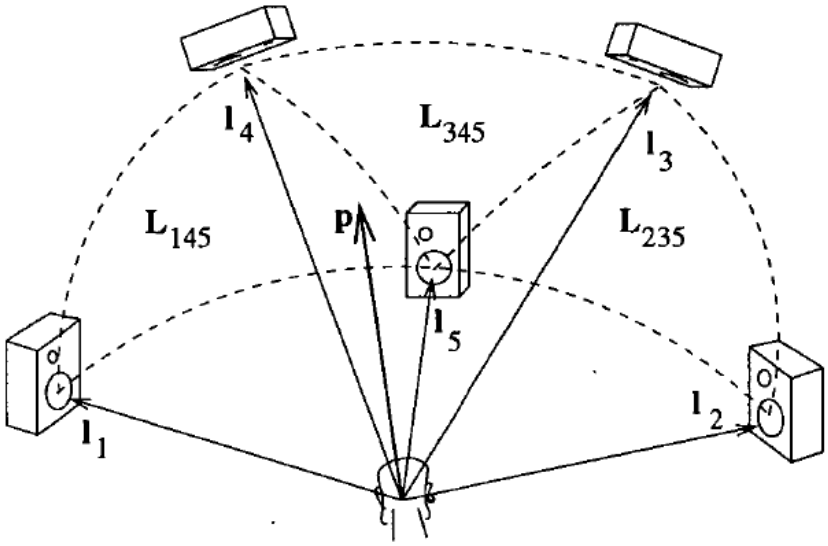


Abbildung 3.10.: Dreidimensionales VBAP mit fünf Lautsprechern. Die Vektoren $\boldsymbol{l}_n$ zeigen auf die Lautsprecher. $\boldsymbol{p}$ kann aufgestellt werden durch die drei Lautsprecherbasen: $\boldsymbol{l}_{145}$, $\boldsymbol{l}_{235}$ und $\boldsymbol{l}_{345}$ [50].

weitert, die sich mit der Fehlerschätzung bzw. Fehlerberechnung von Objekten in unterschiedlichen Wiedergabelayouts befasst. Die Abbildungen zeigen Fehlerberechnungen für die Formate 2.0 Stereo, 5.1 Surround und 22.2 (siehe Abbildungen 3.11, 3.12 und 3.13). Die dargestellten Fehlerberechnungen zeigen die allgemeine Darstellbarkeit von Objekten im 3D-Raum. Selbstverständlich lassen sich Objekte in einem größeren Wiedergabesetup auch besser im Raum verteilen bzw. darstellen. In den folgenden Absätzen werden verschiedene Möglichkeiten der Fehlerschätzung beschrieben.

3.5.1. Einfache Fehlerberechnung über VBAP

Der Grundgedanke der Fehlerberechnung beruht auf Abweichungen in der Darstellbarkeit von Objektpositionen. Für die Objektplatzierung wird von einem Ziellayout (Stereo, Surround, 22.2) ausgegangen. Dieses bildet ein Bett als Basis. Über das Vector Base Amplitude Panning (3.4) werden die Lautsprecherpegel für ein bestimmtes Wiedergabelayout berechnet. In der Berechnung werden für eine be-

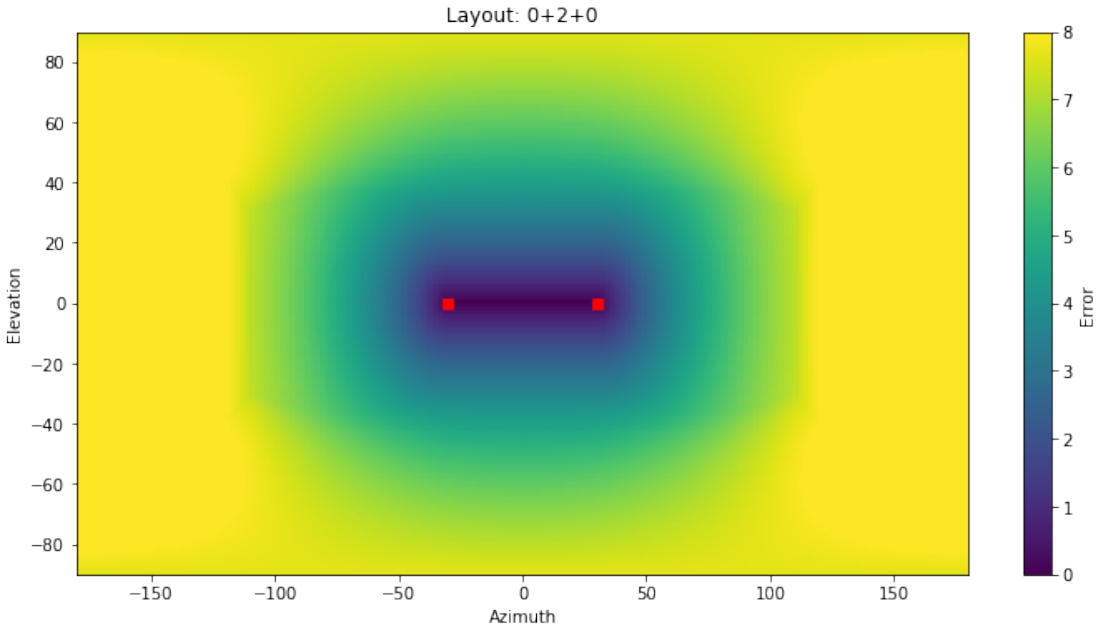


Abbildung 3.11.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten bei einem 2.0 Stereosetup. Die roten Quadrate repräsentieren die Lautsprecher.

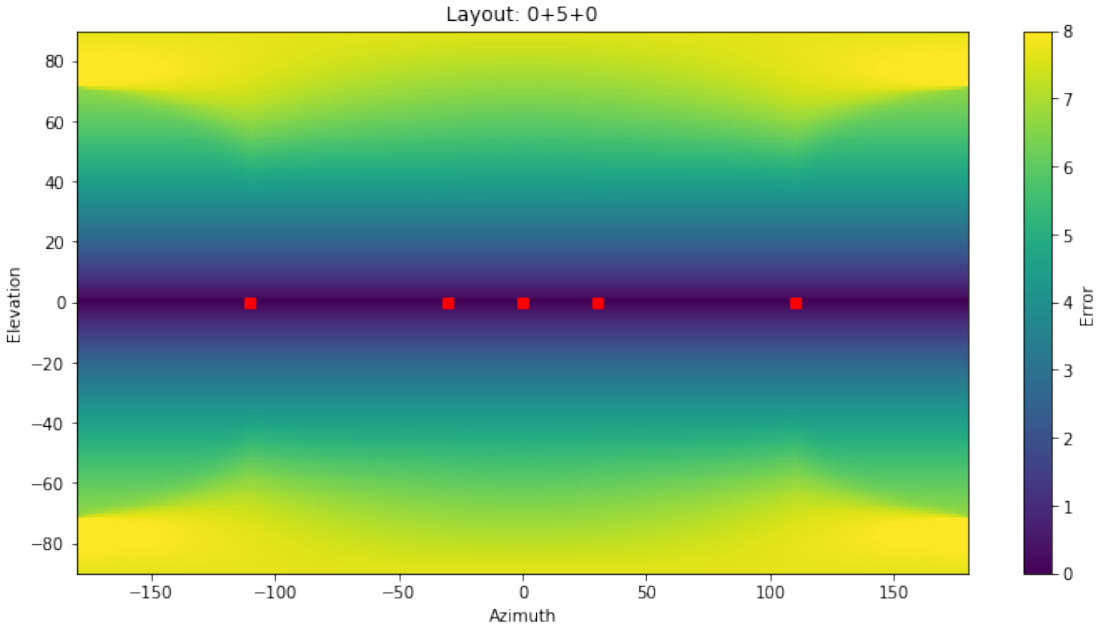


Abbildung 3.12.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten bei einem 5.1 Surroundsetup.

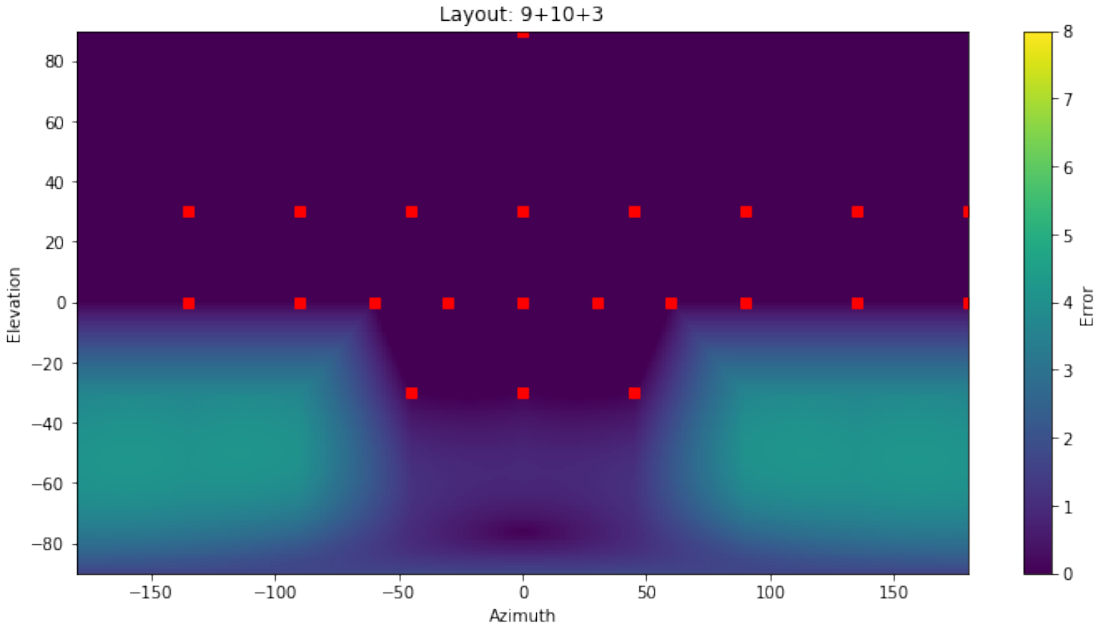


Abbildung 3.13.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten bei einem 22.2 Setup. Die roten Quadrate repräsentieren die Lautsprecher.

stimmte Position über den EAR Signalanteile für jeden Lautsprecher ermittelt. Je nach Wiedergabelayout kann die realisierbare Abbildung der Positionen variieren. In einem Stereolayout ist die Abbildung außerhalb der Medianebene beispielsweise unmöglich. Dieses Verfahren wird im nächsten Schritt umgekehrt, um die mögliche Darstellung aller Positionen im dreidimensionalen Raum im Bezug auf das gewünschte Wiedergabelayout vorherzusagen. Bei der Rückrechnung wird aus der Lautsprecherposition und den Gainwerten die eigentliche Zielposition mit Hilfe eines invertierten Vector Based Amplitude Pannings geschätzt. Die Zielposition kann von der Eingabeposition abweichen. Die Position der Objekte wird in Polarkoordinaten zurückgegeben. Zusammengefasst werden zwei Positionen berechnet. Eine Eingabeposition über das VBAP und die tatsächliche Position, an der das Objekt im besagten Wiedergabelayout über das invertierte VBAP dargestellt wird. Die Differenz bzw. die Distanz dieser Positionen gibt den Fehlerwert an, der an dieser Stelle gewichtet werden kann. Für die vorliegende Abbildung 3.11 wurde für jede Position im Raum die Differenz bzw. der Fehler von jeder Quellposition im Raum berechnet. Ebenso bietet diese Methode die Möglichkeit den Fehler für eine

bestimmte Position, mit Azimuth- und Elevationswert als Eingabe, zu berechnen. Aus einer Inputposition werden auch über den Point Source Panner neue Lautsprecherpegel berechnet. Mit der Angabe des gewünschten Ausgabelayouts werden neue Objektpositionen bestimmt. Die Berechnung läuft wie folgt ab.

$$Error = Position_{alt} - Position_{neu} \tag{3.9}$$

Das Ergebnis ist der Fehlerwert für die entsprechende Position im entsprechenden Layout und kann gewichtet werden. Die erhaltenen Fehlerwerte können dann für jedes einzelne Lautsprecherformat verglichen werden. Eine große Abweichung bzw. ein großer Fehler bedeutet ein schlechte Darstellbarkeit. Der Ansatz zur Berechnung des Fehlers über invertiertes VBAP kann mit der Berechnung des Velocity Vectors $\boldsymbol{r}_V$ gleich gesetzt werden, da die Vektoren aufgrund des *tangent laws* in dieselbe Richtung zeigen (siehe Gleichung 3.10 von Gerzon und Makita)[24, 36]. Dabei kann über die Richtung des *Velocity Vektors* $\boldsymbol{r}_V$ eine Position im Raum vorhergesagt werden [21].

$$\boldsymbol{r}_V = \frac{\sum_{k=1}^{K} G_k \theta_k}{\sum_{k=1}^{K} G_k} \tag{3.10}$$

Der Vektor wird über lineare Summation der gewichteten Lautsprecherrichtungen berechnet.

3.5.2. Fehlerberechnung über den Energy Vector

Ein ähnliches Ergebnis zu VBAP erzielt die Berechnung des *Energy Vectors* $\boldsymbol{r}_E$. Dieser beschreibt die Richtung und die Verteilung der Energie des Schallfelds an der Abhörposition und versucht somit die Position auf eine andere, ähnliche Art und Weise zu berechnen [24, 11]. Analog zum Velocity Vector gibt es keine Begrenzung,

was die Anzahl der aktiven Lautsprecher angeht. Der Hauptunterschied in der Berechnung liegt nach Gerzon und Makita im Quadrat der Gainwerte. Richtungen für höhere Frequenzen und Breitbandsignale werden besser abgebildet[21].

$$\boldsymbol{r}_E = \frac{\sum_{k=1}^{K} G_k^2 \boldsymbol{u}_k}{\sum_{k=1}^{K} G_k^2} \tag{3.11}$$

Der Vektor $\boldsymbol{u}_k$ der Gleichung beschreibt den Richtungsvektor des k-ten Lautsprechers. Die Richtung des Energy Vectors $\boldsymbol{r}_E$ wird als Voraussage der Quellrichtung bei hohen Frequenzen benutzt (siehe auch: Quellcode zur Berechnung des Energy Vectors: [1]) [11]. Die Norm $\|\boldsymbol{r_E}\|$ des Energy Vectors $\boldsymbol{r}_E$ (siehe Gleichung 3.12[16]) hingegen korreliert mit der wahrgenommenen Quellbreite. Zusammenfassend lässt sich sagen: Je kürzer der Energy Vector, desto breiter ist die Energie auf die Lautsprecher verteilt [22, 62]. Somit kann die Stärke des Energy Vectors $\boldsymbol{r}_E$ neben der räumlichen Verteilung der Energie, auch die wahrgenommene Breite von Phantomschallquellen beschreiben [21].

$$\|\boldsymbol{r}_E\| = \begin{cases} \frac{2\sum_{n=1}^{N} g_n g_n - 1}{g_0^2 + 2\sum_{n=1}^{N} g_n^2} & \text{in 2-D} \\ \frac{2\sum_{n=1}^{N} n g_n g_n - 1}{\sum_{n=0}^{N} (2n+1) g_n^2} & \text{in 3-D} \end{cases} \tag{3.12}$$

Es wird deutlich, dass der Velocity Vector identisch zur gewünschten Panningrichtung ist. Der Hauptgrund dafür ist eine reguläre Lautsprecheranordnung. Tabelle 3.4 zeigt die gemittelte, absolute Abweichung der Positionsvorhersagen für VBAP, welche durch subjektive Versuche ermittelt wurden. Obwohl der Velocity- und der Energy Vector die einfachsten Methoden zur Vorhersage des Schallereignisses sind, sind sie gleichzeitig auch die besten [21]. Im Vergleich zu anderen Panning Methoden, schneiden die vektorbasierten Modelle besser ab [49].

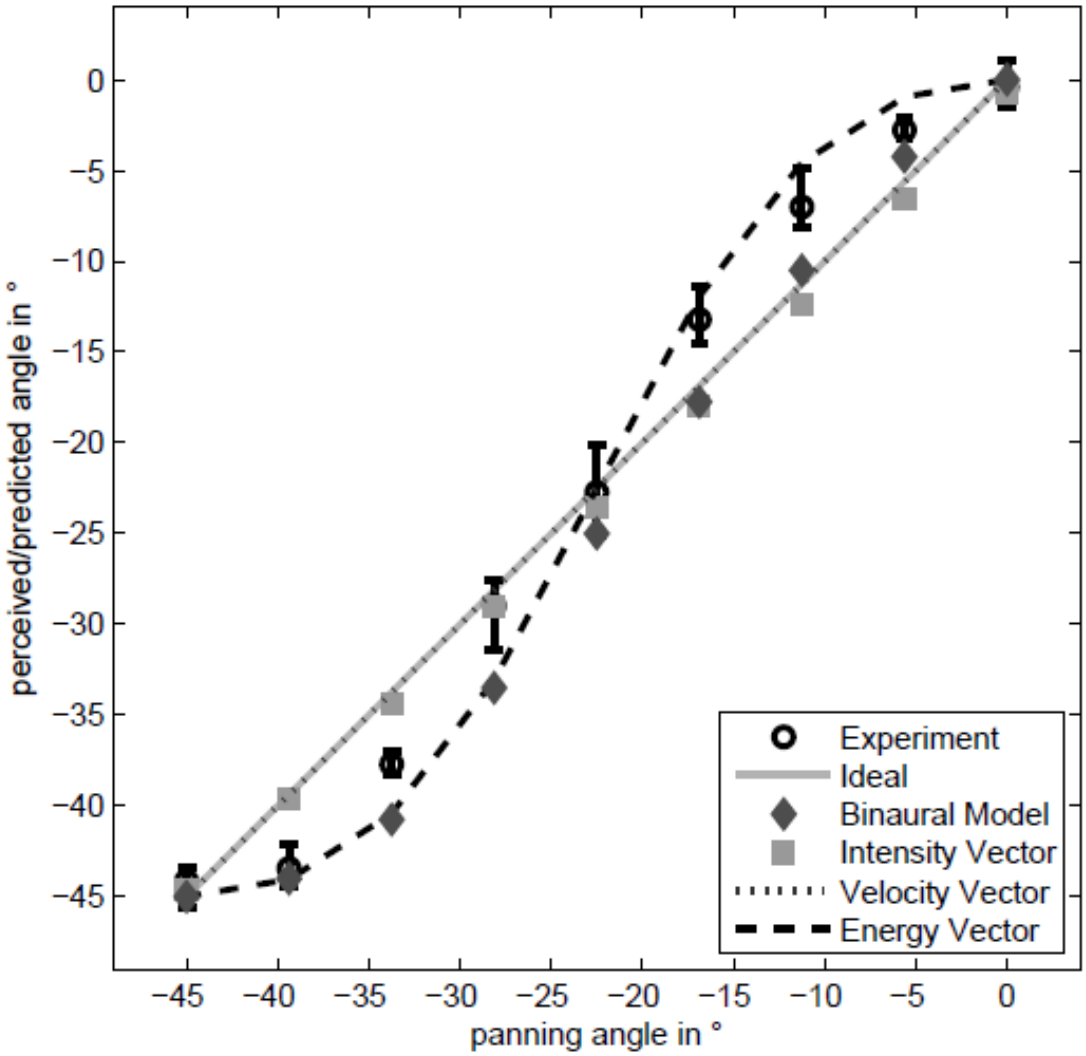


Abbildung 3.14.: Lokalisationskurve für VBAP: experimentelle Ergebnisse für die Lokalisation von Phantomschallquellen bei VBAP. Die Abbildung zeigt Mittelwerte und die dazugehörigen Konfidenzintervalle der wahrgenommenen Winkel für verschiedene Panningwinkel $[0° - 45°]$ an der zentralen Abhörposition [21].

Tabelle 3.4.: Durchschnittliche absolute Abweichung von Positionsvorhersagen für die Panning Methoden VBAP mit Velocity Vector und Energy Vector [21].

	VBAP
Velocity Vector $\boldsymbol{r}_V$	2.35°
Energy Vector $\boldsymbol{r}_E$	1.60°

Insgesamt werden Hörereignisse beim Rendern hauptsächlich über Phantomschallquellen gebildet, weil überwiegend kein direkter Lautsprecher zur Verfügung steht. Statt einer genauen Position wird das Schallereignis, aufgrund der Verteilung der Lautsprecher, unschärfer abgebildet.

Nachdem der Energy Vector $\boldsymbol{r}_E$ in verschiedenen Hörversuchen nach Frank (vgl. Tabelle 3.5) minimal bessere Ergebnisse erzielt hat, wurde dieses Verfahren für die Fehlerberechnung bevorzugt [22]. In der Wahrnehmung kann kein Unterschied zwischen den beiden Methoden festgestellt werden. Diese Feststellung ist auf den Renderer (EAR) zurückzuführen, da dieser auf VBAP Basis rendert und somit auch den Energy Vector über Intensitäten ausgibt.

3.6. Räumliche Unschärfe

Wie bereits in Absatz 3.5.2 angedeutet, werden Schallereignisse häufig unscharf abgebildet. Es wird unterschieden zwischen räumlicher Unschärfe (*Spatial Blur*) und Lokalisationsunschärfe (*Localization Blur*). Unter Lokalisationsunschärfe wird „die kleinste Änderung von Schallereignismerkmalen, die zu einer Änderung des Hörereignisortes führt“ verstanden [8]. Nach Daniel, Nicol und McAdams ist die räumliche Unschärfe die zusätzliche Lokalisationsunschärfe, welche entsteht, wenn zur Lokalisationsunschärfe in stiller Umgebung, abgelenkte Schallquellen zur eigentlichen Lokalisationsunschärfe dazukommen [15]. In Verbindung mit der Lokalisationsunschärfe wird oftmals auch der Begriff *Minimum Audible Angle (MAA)* verwendet. Als Minimum Audible Angle wird der kleinste Winkel zwischen zwei

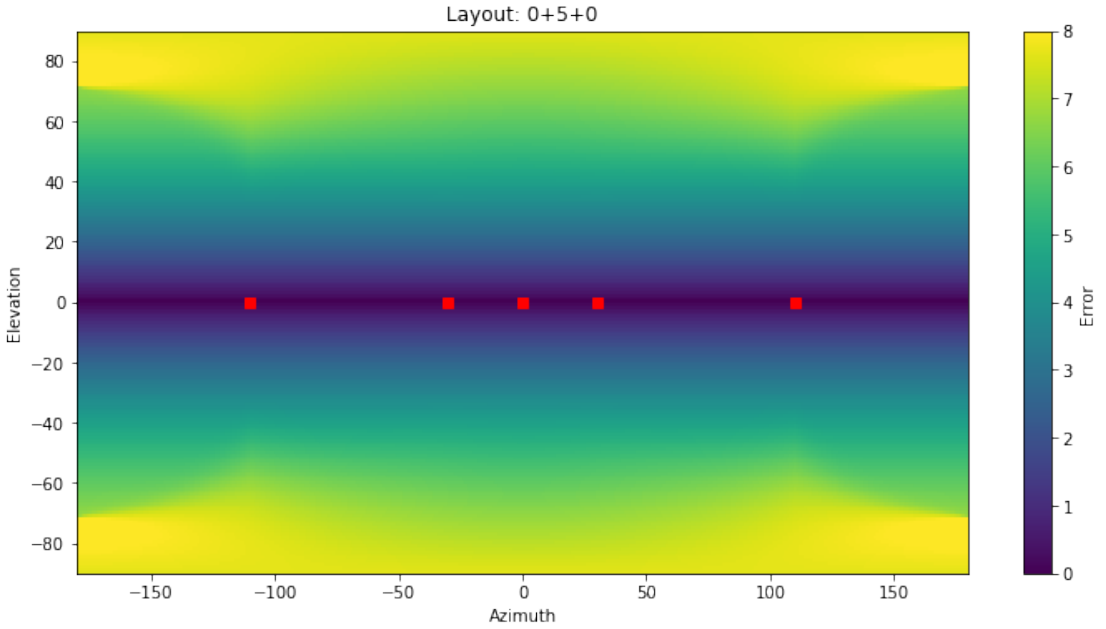


Abbildung 3.15.: Ermittelte Fehler für die Darstellbarkeit von Objekten über den Energy Vector als Plot dargestellt. Die roten Quadrate repräsentieren die Lautsprecher. Das Ergebnis ist ungewichtet.

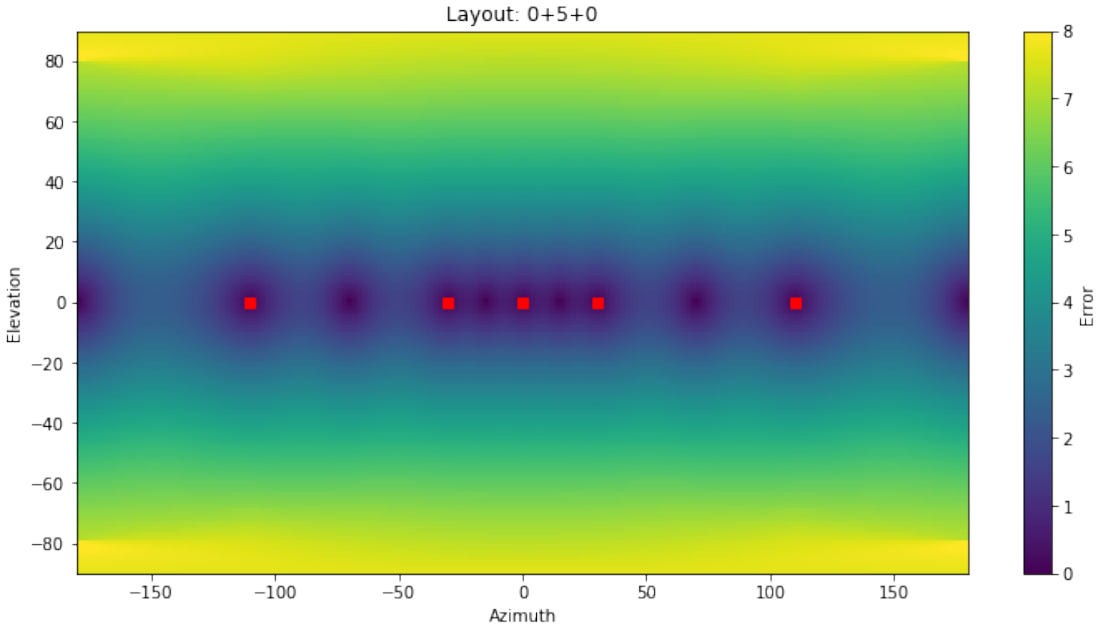


Abbildung 3.16.: Dazu im Vergleich die ermittelten Fehler für die Darstellbarkeit von Objekten über den Velocity Vector $\boldsymbol{r}_V$ (3.10) Die roten Quadrate repräsentieren die Lautsprecher. Das Ergebnis ist ungewichtet.

Tabelle 3.5.: Durchschnittliche absolute Abweichung verschiedener experimenteller Modelle für die Vorhersage von Phantomschallquellen. Die benutzten Modelle sind VBAP bzw. Velocity Vector $\boldsymbol{r}_V$, Energy Vector $\boldsymbol{r}_E$ und gewichteter Energy Vector $\boldsymbol{r}_E^w$. Abweichung vom Öffnungswinkel der Lautsprecher wird in % angegeben [22].

Experiment	**Positionen** θ_l	$\boldsymbol{r}_V$	$\boldsymbol{r}_E$	$\boldsymbol{r}_E^w$
Theile	10° 70°	6.5%	4.8%	5.5%
Theile	30° 90°	9.1%	5.3%	3.4%
Theile	50° 110°	15.8%	8.0%	4.2%
Theile	60° 120°	18.2%	7.7%	8.5%
Martin u. a.	0° 30°	9.6%	6.1%	4.9%
Martin u. a.	30° 120°	9.8%	6.8%	5.7%
Martin u. a.	-120° 120°	21.4%	11.8%	11.8%
Simon, Mason und Rumsey	0° 45°	7.5%	4.9%	6.4%
Simon, Mason und Rumsey	45° 90°	10.2%	2.9%	5.9%
Simon, Mason und Rumsey	90° 135°	16.9%	14.6%	5.1%
Simon, Mason und Rumsey	135° 180°	10.5%	10.1%	6.0%
Durchschnittliche Abweichung:		12.3%	7.5%	6.1%

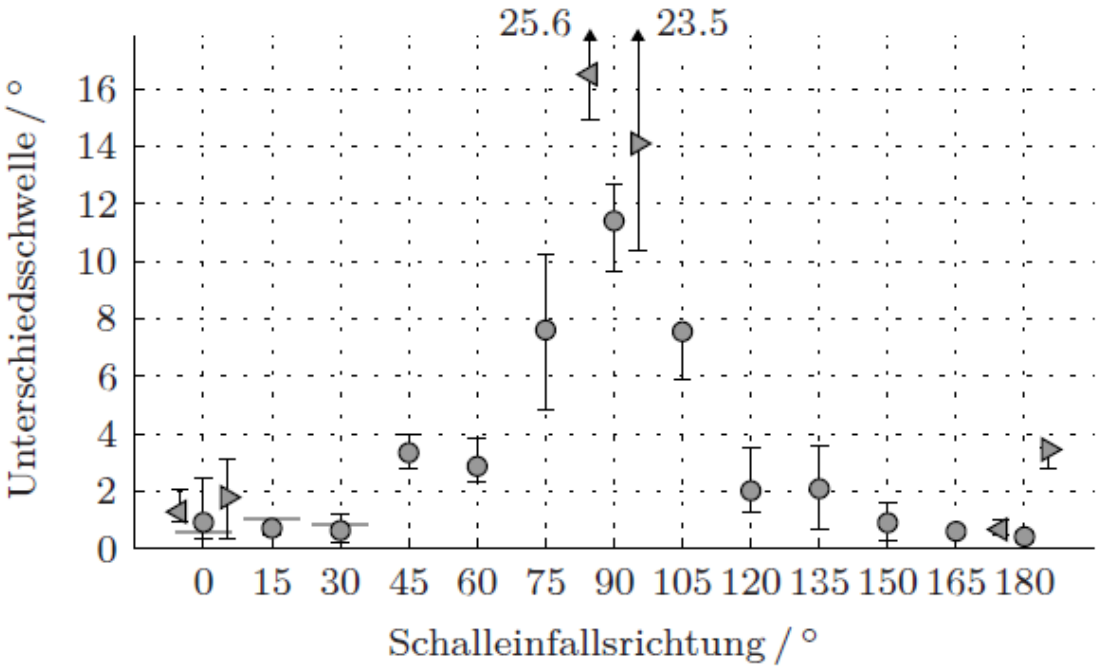


Abbildung 3.17.: Hörversuch zum MAA. WFS-Aufbau für Rauschimpulsfolgen (700 ms Puls, 300 ms Pause, 20 ms Flanke). Kugelwellen in 2 m Abstand und verschiedenen Einfallsrichtungen. Breitband-(○), Tiefpass-(◁) und Hochpassrauschen (▷). Linien: Mediane für ebene Wellen [58].

Positionen des selben Hörereignisses bezeichnet, bei dem die Hörereignisse gerade noch als unterschiedlich wahrgenommen werden können. Im Prinzip ist der MAA der Lokalisationsblur als Winkel angegeben [58]. Bereits die Ergebnisse von Mills aus dem Jahr 1958 zeigen, dass der MAA bei einer Frequenz von 500 Hz bis 750 Hz bei 1° liegt, wenn die reale Quelle direkt vor oder hinter der Abhörposition liegt. Auch Blauert führt an, dass die Lokalisationsunschärfe in Azimuthrichtung bei optimalen Bedingungen ungefähr 1° in der Front beträgt. Je weiter die Schallquelle in Richtung 90° wandert, desto größer wird auch der MAA [41]. Folglich nimmt die Richtungsauflösung für seitlichen Schalleinfall ab. Das bestätigen auch Völk und Fastl in ihrem Hörversuch (siehe Abbildung 3.17) [58].

Die Lokalisationsunschärfe ist eher signalabhängig und kann zwischen 4° (weißes Rauschen) und 17° (durchgehende Sprache von einer unbekannten Person) va-

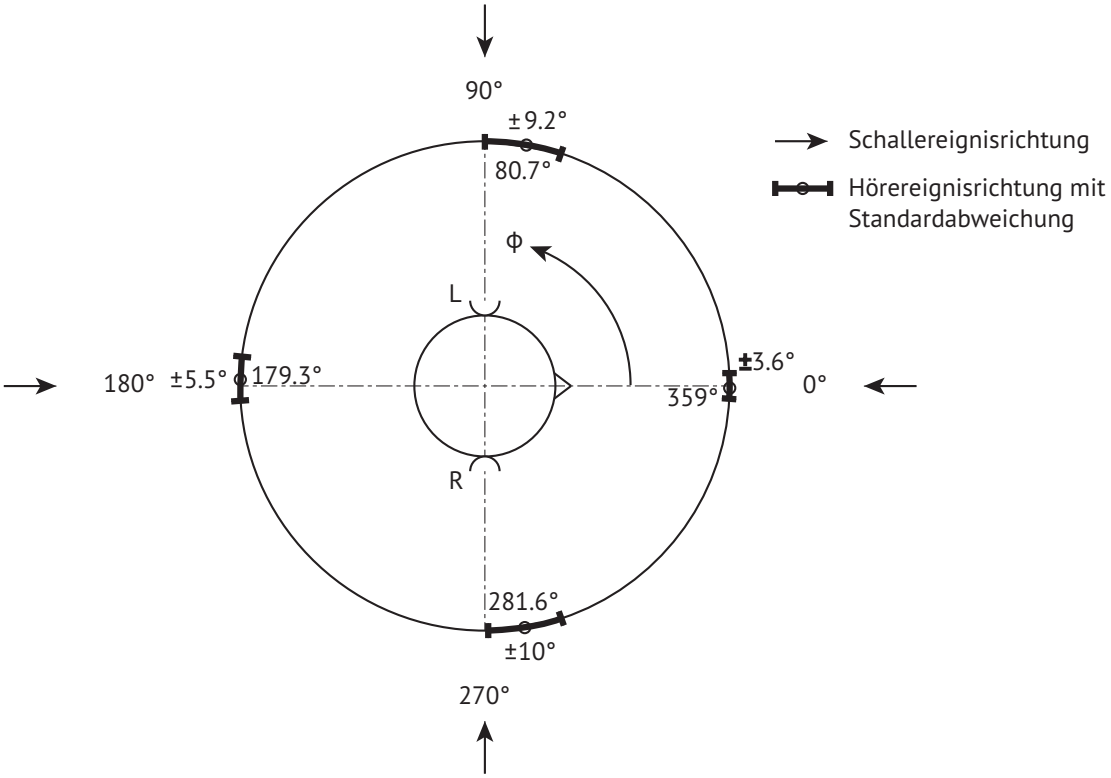


Abbildung 3.18.: Lokalisationsunschärfe und Lokalisation in der Horizontalebene nach Preibisch-Effenberger 1966. 600-900 Versuchspersonen. Impulse von weißem Rauschen mit einer Dauer von $100ms$. Kopf fixiert [8].

riieren [8]. Die Richtungswahrnehmung lässt sich nach Preibisch-Effenberger für seitliche Abweichungen der Schallquelle von vorne mit nur $\pm 4°$, für Abweichungen nach oben und unten mit $\pm 10°$ beschreiben [47].

In verschiedenen Versuchen haben Daniel, Nicol und McAdams 2012 die Lokalisationsunschärfe mit Bezug auf den MAA in Azimuthrichtung untersucht, um ein Modell zur Lokalisationsunschärfe zu erstellen, welches die räumliche Auflösung darstellt. Diese Versuchsreihe beschreibt den MAA, als den Wert für den Winkel α, bei dem der Hörer zu mindestens 80% einen Unterschied erkennt. Gleichzeitig zu den beiden Zielsignalen wurde ein Ablenkungsgeräusch von mindestens einem anderen Lautsprecher eingespielt (vgl. Abbildung 3.20) [15]. Die Ergebnisse bestätigen die Theorie von Mills. Sobald ein Ablenkungsgeräusch zugespielt wird, erhöht

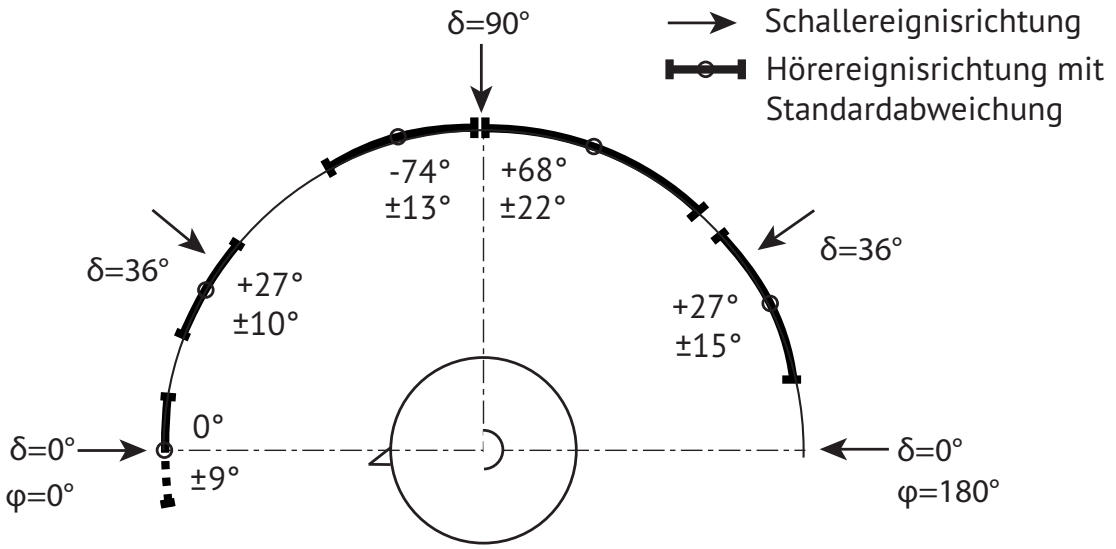


Abbildung 3.19.: Minimale Lokalisationsunschärfe und Lokalisation in der Medianebene nach Damaske und Wagener 1969. 7 Versuchspersonen. Kopf fixiert [8].

sich die Lokalisationsunschärfe. Als Beispiel wird eine Frequenz von 1400 Hz, und ein Störgeräusch von vier Rechteckbandbreiten höher genannt. Die resultierende räumliche Unschärfe liegt in diesem Fall bei 7°. Die Lautheit der verschiedenen Signale wurde angepasst. Eng mit dem Energy Vector verknüpft ist der sogenannte Angular Spread. Darunter versteht sich der Richtungsversatz bzw. der Richtungsfehler des Energy Vectors, der den Winkel zwischen der Richtung des Energy Vectors und der Richtung der gewünschten virtuellen Schallquelle angibt. Der Angular Spread lässt sich der Lokalisationsunschärfe bzw. dem Lokalisationsblur unterordnen. Die Größe des Lokalisationsblurs ist abhängig vom Winkel zwischen der Position der virtuellen Schallquelle und dem nächsten Lautsprecher. Je größer dieser Winkel ist, desto größer ist der Blur. Der Blur kann entweder als Winkel in Azimuth und Elevation oder auch als Distanz angegeben werden. Wie bereits in Kapitel 3.5.2 beschrieben, korreliert die Länge des Energy Vectors mit der wahrgenommenen Quellbreite. Der Angular Spread gibt somit laut Gleichung 3.13 die Breite der Fläche an, über welche die Schallenergie verteilt wird [19].

$$\sigma = 2\,arcos(2||\boldsymbol{r}_E|| - 1) \tag{3.13}$$

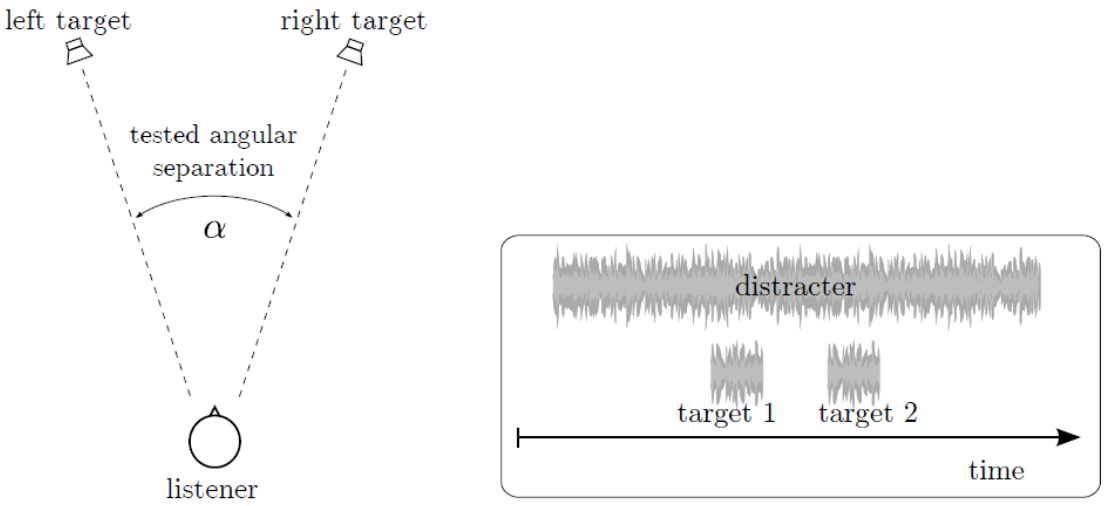


Abbildung 3.20.: Versuchsmodell zu Bestimmung des MAAs. Jede abgespielte Sequenz wird der Versuchsperson entweder links-rechts oder rechts-links präsentiert. Zusätzlich wird ein durchgehendes Ablenkunsgeräusch aus einem oder mehreren Lautsprechern von variablen Positionen zugespielt.

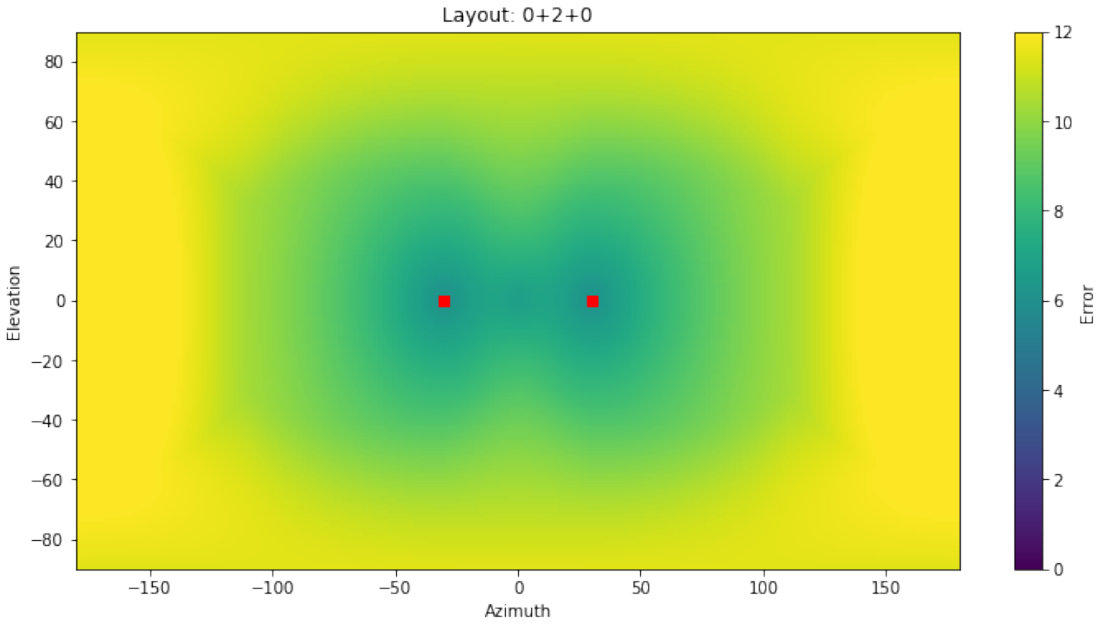


Abbildung 3.21.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten über den Spatial Blur bei einem 2.0 Setup. Die roten Quadrate repräsentieren die Lautsprecher.

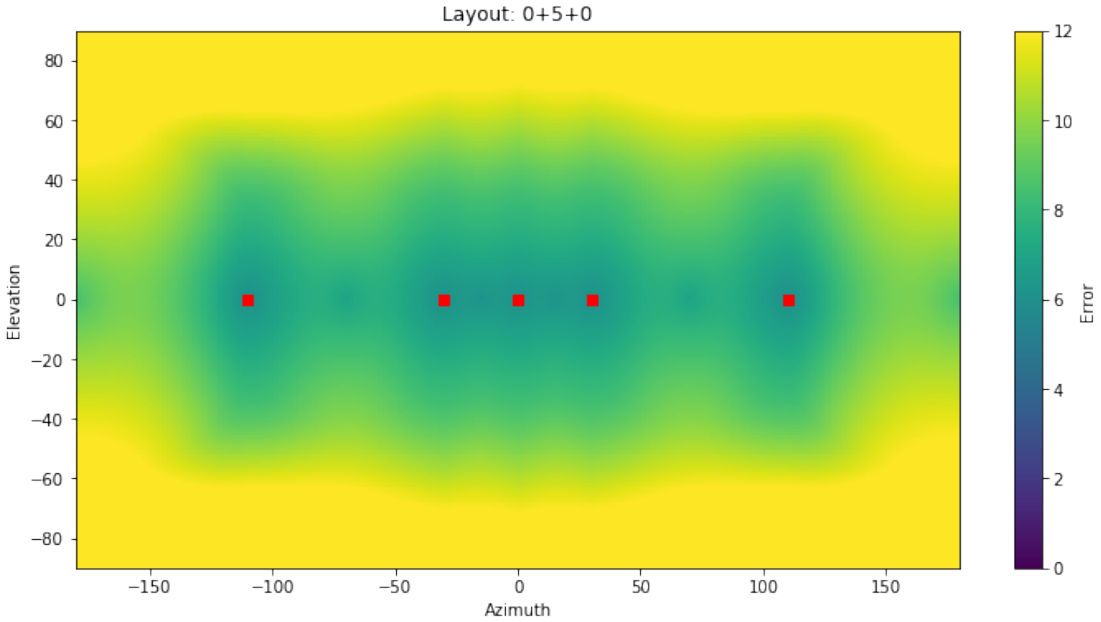


Abbildung 3.22.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten über den Spatial Blur bei einem 5.1 Setup. Die roten Quadrate repräsentieren die Lautsprecher.

Mit dieser Definition ist der Angular Spread gleich zur Winkelöffnung der entsprechenden Schallquellenstreuung, welche in der festgestellten Energy Vector Norm resultieren würde. Beispielsweise ist der Spread 0, wenn die Energy Vector Norm eins ist. Andersherum ist der Spread gleich 2Π, wenn die Norm 0 ist. Das entspricht einer Szene, bei der die Schallquellen gleichmäßig im Raum verteilt sind. Die Berechnung des Angular Spread wurde auch in Python geschrieben (siehe Codebeispiel 2).

Die räumliche Unschärfe bzw. der MAA fließen ebenfalls in die Berechnung des Fehlers mit ein. In den Abbildungen 3.21, 3.22 und 3.23 zeigen sich die Auswirkungen der räumlichen Unschärfe. Nachdem die erstellten Ergebnisse bisher keine konkrete Wertung haben, werden die Werte im Folgenden sinnvoll gewichtet und miteinander verrechnet.

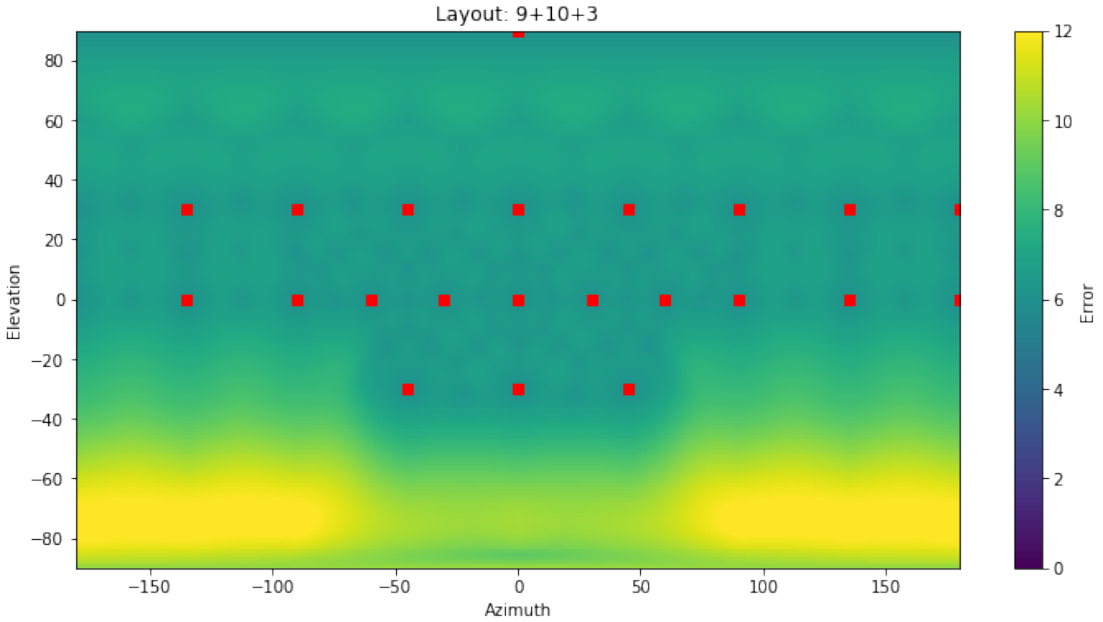


Abbildung 3.23.: Ermittelte, ungewichtete Fehler für die Darstellbarkeit von Objekten über den Spatial Blur bei einem 22.2 Setup. Die roten Quadrate repräsentieren die Lautsprecher.

3.7. Richtungsabhängige Fehlergewichtung

Um eine richtungsabhängige Fehlergewichtung zu formulieren, wurden die wichtigsten Faktoren betrachtet. Velocity Vector und Energy Vector wurden unter dem Oberbegriff Richtungsfehler zusammengefasst. Die Konsequenz aus MAA bzw. der räumlichen Unschärfe ist, dass die Lokalisation im dreidimensionalen Raum nicht überall gleich gut funktioniert. Dies führte zu der Aufspaltung des Richtungsfehlers in den Azimuth- und den Elevationsfehler. Diese werden bei der Fehlerberechnung einzeln gewichtet, um genauere Fehlerwerte zu erhalten. Dafür wird auf die Lokalisationsunschärfe (*Localization Blur*) bzw. die Lokalisation in der Horizontalebene nach Preibisch-Effenberger zurückgegriffen (Siehe Abbildung 3.18 und 3.19).

Mit diesen beiden Versuchen wurden erste Ergebnisse zur Lokalisation im dreidimensionalen Raum erstellt. Im nächsten Schritt wird zwischen den einzelnen Werten interpoliert, um für jeden Azimuth- und Elevationswert einen halbswegs

korrekten Lokalisationswert bzw. eine Abweichung zu bekommen. Die berechneten Werte entsprechen den MAAs für jede Position. Gleichzeitig wird über den Energy Vector $\boldsymbol{r}_E$ die erwartete Azimuth- und Elevationsposition ($p'_a z$ und $p'_e l$) berechnet. Von dieser Position wird jeweils die Eingabeposition p abgezogen (vgl. Gleichung 3.14). Zu guter Letzt wird der Fehler normiert, indem $p' - p$ durch den MAA geteilt wird. Das Ergebnis ist ein Fehlerwert für Azimuth und Elevation für eine gewünschte Position (vgl. Gleichung 3.15 und Codebeispiel 3).

$$\begin{aligned} \Delta_{p_{az}} &= p'_{az} - p_{az} \\ \Delta_{p_{el}} &= p'_{el} - p_{el} \end{aligned} \tag{3.14}$$

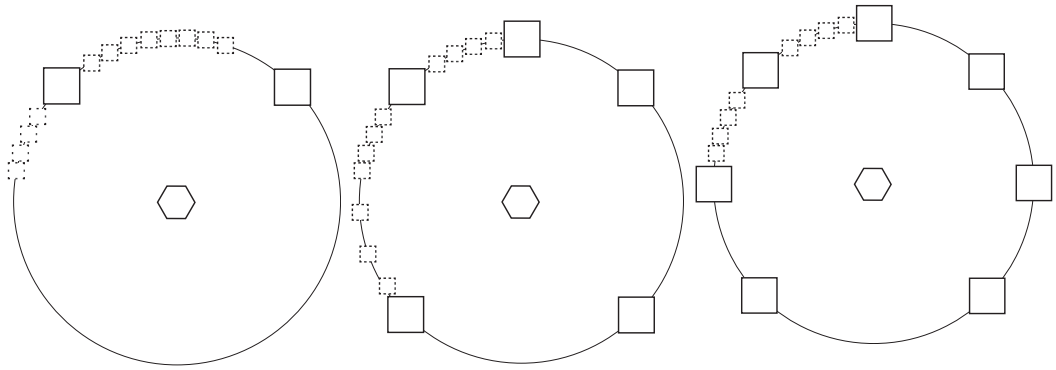
$$\begin{aligned} fehler_{az} &= \frac{\Delta_{p_{az}}}{MAA} \\ fehler_{el} &= \frac{\Delta_{p_{el}}}{MAA} \end{aligned} \tag{3.15}$$

3.8. Diffuse, Extent, Object Divergence

Im nächsten Schritt wurde versucht auf psychoakustischer Ebene herauszufinden, wie stark die Parameter *diffuse*, *extent* und *object divergence* den wahrnehmbaren Fehler beeinflussen. Die *diffuse* Funktion des ADMs gibt das Verhältnis von Direktschall zu Diffusschall an [2]. Beim Rendern mit aktivem *diffuse* Parameter werden Dekorrelationsfilter auf Lautsprecherausgaben eingesetzt. Bei einem Wert von beispielsweise 0,5 liegen genau gleiche Anteile vor. Also zu 50 Prozent Direktschall und zu 50 Prozent Diffusschall.

$$\frac{1}{N} \sum_{i=0}^{N} (\frac{1}{N} - g_i)^2 \tag{3.16}$$

Die Gleichung beschreibt die Annahme zur Berechnung des diffusen Fehlers. Dabei gibt N die Anzahl der Lautsprecher wieder und g_i deren Gainwert.

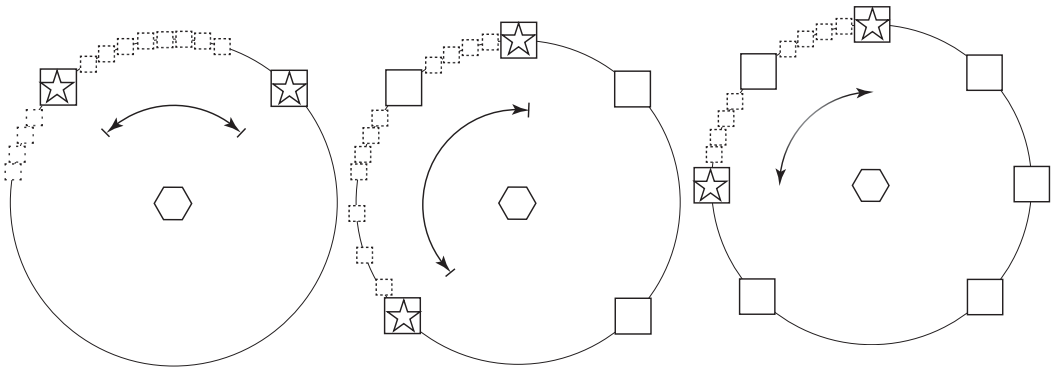


(a) 2.0 Stereo mit Extent Schallquellen **(b)** 5.1 Surround mit Extent Schallquellen **(c)** 7.1 Surround mit Extent Schallquellen

Abbildung 3.24.: Vergleich der Extent Funktion bei verschiedenen Formaten. Reale Lautsprecher werden mit großen Quadraten, die Extent Lautsprecher als kleine gestrichelte Quadrate dargestellt. Schallquellen links vom linken Stereolautsprecher sind bei 3.24a nicht darstellbar. Der *Extent* bildet zusätzliche virtuelle Schallquellen zwischen den realen Schallquellen ab. Bei Stereo können aufgrund von fehlenden Surround Lautsprechern keine virtuellen Schallquellen hinter den beiden Stereolautsprechern gebildet werden. Auf den Surroundformaten können quasi im kompletten Kreis virtuelle Schallquellen vorhanden sein. Je mehr Lautsprecher im Raum, desto besser ist die Platzierung virtueller Schallquellen möglich.

Im Zusammenhang mit dem *diffuse* Parameter, ist auch der *extent* Panner von Bedeutung. Dieser produziert zusätzliche virtuelle Schallquellen neben den existierenden, um ein Objekt breiter darzustellen [54]. Der Extent Parameter wird auf die Breite (width), Höhe (height) und Tiefe (depth) eines Objekts angewendet.
Der *extent* Wert ist nicht an den diffuse Parameter gebunden. Jedoch kann der extent Parameter in Verbindung mit dem diffuse Parameter, anliegende echte Lautsprecher direkt ansprechen.

$$\Delta ext_{az} = ext_{az} - w_{az} \tag{3.17}$$



(a) 2.0 Stereo mit Extent Schallquellen **(b)** 5.1 Surround mit Extent Schallquellen **(c)** 7.1 Surround mit Extent Schallquellen

Abbildung 3.25.: Der Hauptbestandteil der Fehlerberechnung beim Extent ist der maximale Winkelabstand zwischen der aktiven Lautsprecher. In der Abbildung sind die am weitesten voneinander entfernten, aktiven Lautsprecher mit einem Stern markiert. Dazwischen werden virtuelle Schallquellen gebildet. Das funktioniert umso besser, umso kleiner der Winkel zwischen den Lautsprechern ist. Nachdem Stereo den Fehler verfälschen würde, wird in der Formel eine zusätzliche Gewichtung für diesen Fall hinzugefügt (vgl. 3.20).

Aus der Breite des Extents ext_{az} und dem Winkelbereich zwischen den äußersten aktiven Lautsprechern w_{az}, wird der Extentfehler Δext_{az} berechnet. Je größer der Unterschied zwischen diesen beiden Werten ist, desto größer ist der Extentfehler (vgl. Gleichung 3.17). Der Extent kann nie größer als der maximale Lautsprecherabstand werden.

$$e_{extent} = \underbrace{e_{\text{distance}}}_{\text{Entfernung äußerster Lautsprecher zu Extent-Ende}} \cdot \underbrace{g_{\text{distance}}}_{\text{Extra-Gewichtung wenn Winkelbereich < Extent}} + \underbrace{e_{\text{speakercount}}}_{\text{Anzahl aktive LS innerhalb d. breiten Quelle}} \tag{3.18}$$

$$e_{\text{distance}} = |\Delta ext_{az}| \tag{3.19}$$

Mit Gleichung 3.19 wird die Entfernung zwischen den letzten bzw. äußersten Lautsprechern, welche zum Extent Rendering benutzt werden, beschrieben. Umso größer die Distanz, umso schlechter müsste die genaue Breite getroffen werden, da die virtuellen Schallquellen zum Extent-Rendering auch schlechter dargestellt werden (vgl. 3.19).

$$g_{\text{distance}} = 1 + H(\Delta ext_{az}) \cdot g_{undercover} \tag{3.20}$$

Die *Heaviside Step-Funktion* $H(x)$ sorgt dafür, dass $g_{\text{undercover}}$ nur bei $\Delta ext_{az} > 0$ verrechnet wird. Sie ist die zusätzliche Gewichtung für den Fall, dass die äußersten Lautsprecher, welche zum Extent beitragen sollen, innerhalb des gewünschten Extent-Bereichs liegen ($\Delta ext_{az} > 0$). Dadurch kann der gewünschte Extent gar nicht erreicht werden. Dies sollte wahrscheinlich deutlich stärker gewichtet werden als der andere Fall (3.20).

Der Extent-Fehler ist außerdem abhängig von der Anzahl der verwendeten Lautsprecher (vgl. Gleichung 3.21). Dabei gibt R die Anzahl der Lautsprecher innerhalb

des Extent-Winkelbereichs an.

$$e_{\text{speakercount}} = \frac{1}{R+\epsilon} \cdot g_{\text{count}} \tag{3.21}$$

oder

$$e_{\text{speakercount}} = \frac{1}{R+\epsilon} \cdot g_{\text{count}} \cdot (1 - \text{diffuse}) \tag{3.22}$$

Es wird angenommen, umso mehr Lautsprecher für die Darstellung des Extents verwendet werden, umso besser. Kein Lautsprecher zusätzlich zu den beiden äußersten ($R = 0$) dürfte gar keine Extent-Darstellung erlauben. Die Addition der Konstante ϵ verhindert ein Teilen durch 0 und liefert für $R = 0$ dadurch entsprechend große Fehlerwerte.

Die zweite Gleichung (3.22) zur Anzahl der Lautsprecher bezieht den *diffuse* Parameter mit folgender Schlussfolgerung mit ein. Umso diffuser die Lautsprechersignale sind, umso weniger dürfte die konkrete Anzahl eine Rolle spielen. Deshalb könnte man optional den diffuse Wert verrechnen. Allerdings wird die virtuelle Quelle auch nur mit einer größeren Anzahl an Lautsprechern *wirklich* diffuser.

Das ADM Attribut *object divergence* erzeugt neben der eigentlichen Objektposition noch zwei weitere virtuelle Objekte in einem bestimmten Abstand zum originalen Objekt [2]. Object divergence enthält zwei Attribute. Mit der azimuthRange wird der Winkel in Grad, also der Abstand der beiden neuen Objekte zum ursprünglichen Objekt angegeben. Diese befinden sich symmetrisch zum Objekt auf der Horizontalachse verschoben. Wenn ein Objekt z.B. bei 20° liegt und für die object divergence 30° eingestellt wird, dann liegen die beiden neuen virtuellen Quellen bei 50° und -10°. Zusätzlich wird noch ein Wert (value) mitgegeben, mit dem die Stärke der Divergenz angegeben wird. Bei 0.0 liegt keine Divergenz vor, bei 1.0 befindet sich das Signal komplett auf den neu erzeugten virtuellen Schallquellen.

Neben den bereits erarbeiteten Faktoren zur Fehlerberechnung, wurden weitere Überlegungen angestellt, wie der allgemeine Ablauf beim Vorrendern vonstatten gehen soll. Der Gesamtfehler soll sich aus verschiedenen anderen Fehlern ergeben.

Zu den Einzelfehlern zählt auch z. B der Spatial Blur oder auch der Energy Vector (Veranschaulicht dargestellt in Abbildung 3.26).

Wie in der Abbildung 3.26 zu sehen ist, fließen weitere Kriterien in die Fehlerberechnung ein. Zu Beginn stehen folgende Elemente zur Gewichtung: Richtungsfehler, Inhalt, Richtungsabhängigkeit, Spatial Blur, sowie Diffus/Extent. Diese Variablen werden selbst berechnet und schließlich gewichtet. Im nächsten Schritt findet eine Gewichtung untereinander statt. Alle in dieser Arbeit beschriebenen Fehlerschätzungen können zur Steuerung des Vorverarbeitungsprozesses genutzt werden.

3.9. Dolby Verfahren

Im Folgenden wird ein von Dolby beschriebenes, patentiertes Verfahren zur Reduktion der Komplexität von NGA Audiomaterial mit dem erarbeiteten Vorgang verglichen [12].

Das Format besteht aus den drei Komponenten *bed*, *object* und *metadata*. Das Fundament (*bed*) enthält kanalbasierte Mischungen mit bereits vorgenommenem Panning. Die objektbasierte Komponente *object* enthält Mono oder Stereo Inhalte. Über die dritte Komponente *metadata* werden die genannten Audioinhalte bzw. Objekte mit Hilfe von Metadaten gepannt, also in verschiedene Richtungen verschoben [33]. Der Rendervorgang ist generell sehr ähnlich zu dem in Kapitel 3 erarbeiteten Ablauf. Das Ziel allgemein ist Objekte und Betten zu kombinieren. Zu Beginn soll die wahrnehmbare Bedeutung von Objekten in einer Audioszene bestimmt werden. Dazu werden Objekte, Objektaudiodaten und assoziierte Metadaten zusammengefasst zu Clustern. Die Anzahl der Cluster ist folglich kleiner als die Anzahl der ursprünglich vorliegenden Audioobjekte in der Audioszene. Beim Clustern werden Schwerpunkte erstellt und Audioobjekte somit nach ihrer Wichtigkeit sortiert. Das Clustern wird hauptsächlich durchgeführt, um die Datenrate durch Audiocodierung zu reduzieren. Es ist wichtig, dass der Renderer mit dem Decoder kommunizieren kann und für jedes Ausgabeformat ein passendes Zuspielformat

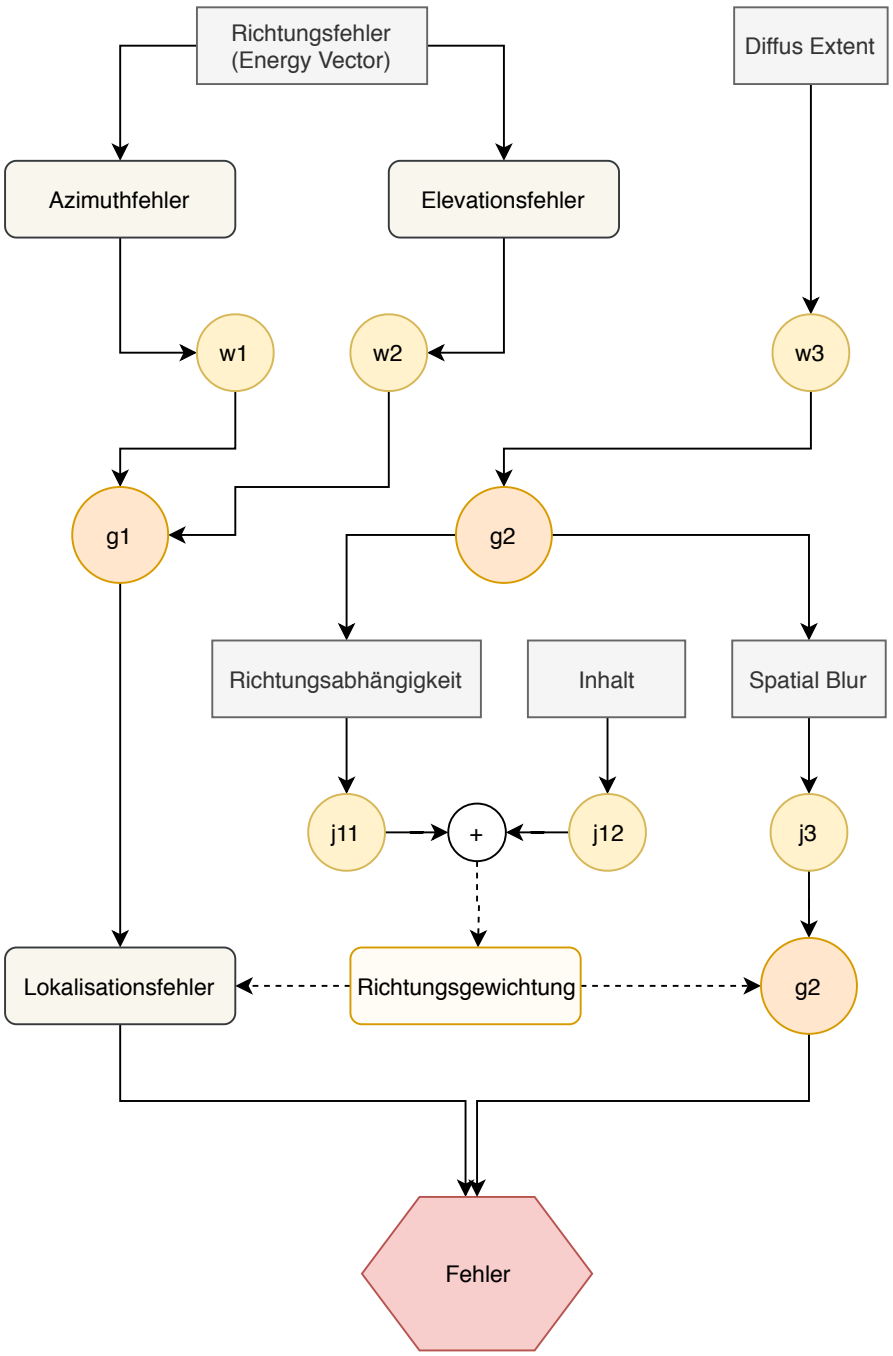


Abbildung 3.26.: Diagramm zum Ablauf einer Fehlerberechnung.

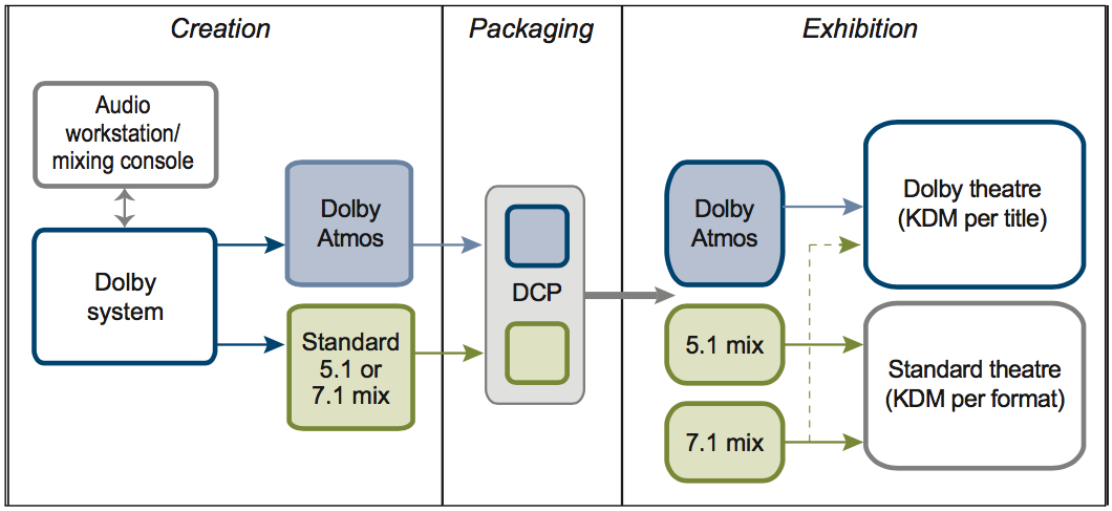


Abbildung 3.27.: Der Produktionsworkflow Dolby Atmos enthält große Teile des Patents [12]. Untergliedert wird in drei Teile. Creation behandelt die Produktion des Signals. Packaging ist für die *Verpackung* und den Transport zuständig. Exhibition betrifft die Wiedergabe [33].

bzw. Rendering ausgewählt wird. Wenn das Format nicht die Anzahl an Objekten und Kanälen unterstützt, müssen Cluster erstellt werden. Objekte werden kategorisiert: z. B. in Dialog, Musik, Effekte, und Background. Das bringt Struktur in die Sortierung und ist nützlich während des Renderprozesses. Der Clusterprozess erfolgt in regulären periodischen Abständen (einmal alle 10 Millisekunden). Dabei werden Objekte analysiert und Gruppen daraus erstellt. Bezogen auf den zeitlichen Aspekt sind zeitbasierte Grenzinfos hilfreich, da sie Beginn und Ende eines Audioobjekts definieren. Zusätzlich kommt eine auditorische Szenenanalyse hinzu, also quasi eine Audioeventerkennung. Ein wichtiger Aspekt beim Clustern ist der Spatial Error Threshold, der anfangs gesetzt wird. Jeder Wert bzw. jedes Objekt, das außerhalb dieses Thresholds liegt, wird aus dem Cluster entfernt. Es ordnet sich dann entweder einem anderem Cluster unter oder bleibt als alleiniges Objekt bestehen. Ebenfalls ähnlich gehandhabt werden feste Klangbetten. Diese können zu sich bewegenden Objekten oder geclusterten Objekten hinzugefügt werden. Es gibt verschiedene Arten Label für die verschiedenen Cluster zu vergeben. Zum Beispiel, dass ein Cluster mit der meisten Energie ein Label bekommt. Hin und wieder reicht ein Label zur vollständigen Beschreibung nicht aus. Alternativ

wird eine Gewichtung erstellt. z. B. 50 % Dialog, 30 % Musik, 20 % Effekte. Hinzu kommt die Einführung eines Fehlerwerts. Sobald dieser den Wert 0 erreicht, findet kein Clustering mehr statt. Es gibt eine *Erwartungs/Wichtigkeits-* Funktion, die eine Audio Klassifikationskomponente enthält, mit der automatisch die Inhaltstypen der Audioobjekte erwartet werden können. Diese Funktion gibt somit die Wichtigkeit eines Inhaltstyps an. Zu guter Letzt definieren die Autoren des Patents das wahrnehmungsbasierte Clustering. Dabei wird eine Clustermethode konfiguriert, um Objekte und Kanalbetten auf die aktuellen Bedingungen einzuschränken. Um räumliche Fehler zu minimieren, werden Objekte mit inhaltlich hoher Wichtigkeit favorisiert [12]. An dieser Stelle können Parallelen, aber auch abweichende Ansätze der Vorverarbeitung festgestellt werden. Der Hauptunterschied des Dolby Verfahrens spiegelt sich beim Clustern wieder. Beim Clusterprozess können Objekte im 3D-Raum zu neuen Objekten *gruppiert* werden. Objekte können bereits gruppiert betrachtet werden oder schließlich neu geclustert werden. Im Vergleich dazu ist die Methode in dieser Arbeit ein Vorrendern von einzelnen Objekten auf Kanalbetten bestimmter Lautsprechersetups. Der Hintergrund dieser Methode beruht auf der Anwendung im Rundfunk. Das erarbeitete Modell soll schlussendlich nur Lautsprechersetups der ITU Empfehlung BS.2051 [31] bedienen können. Dabei kann die hörbare Verschlechterung bzw. Verbesserung beim Vorrendern bewertet werden und aufgrund dessen die Vorverarbeitung steuern.

4. Schlussfolgerung und Ausblick

Diese Arbeit hat sich mit der Entwicklung einer automatisierten Optimierung von NGA Inhalten mit Hilfe einer Vorverarbeitung befasst. Die Vorverarbeitung beruht auf der Konvertierung von NGA Inhalten zu einem möglichst universellen Ausgabesetup. Dabei sollte über eine Berechnung von Abweichungen bzw. Fehlern der wahrnehmbare Unterschied von Audioobjekten in verschiedenen Wiedergabesetups bewertet werden.

Die im Rahmen dieser Arbeit entstandenen Untersuchungen zeigen, dass die beschriebenen Fehlerwerte bzw. Fehlerschätzungen zur Steuerung des Vorverarbeitungsprozesses genutzt werden können. Bisher konnte bei dem Prozess der Anpassung des Audiosignals keine ideale Konvertierung für das situationsabhängige Wiedergabemedium erfolgen. Das ändert sich durch das erarbeitete Modell, das die Berechnung von verschiedenen Fehlern ermöglicht. In erster Linie wurden Einschränkungen bzw. Anforderungen des Rundfunks und den damit verbundenen Codecsystemen wie z. B. MPEG-H analysiert und daraufhin für ein besseres Ausspielergebnis bearbeitet. Die Bearbeitung erfolgt auf Basis der verschiedenen Fehlerberechnungen. Der Richtungsfehler kalkuliert mittels Lokalisationsschärfe die Abweichung von Positionen verschiedener Wiedergabeformate sowohl in Azimuth- als auch in Elevationsrichtung. Die entstandene Abweichung wird über die Lokalisation an der besagten Position gewichtet. Die beiden gewichteten Fehler werden miteinander multipliziert und ergeben den Lokalisationsfehler. Die Untersuchung dieses Fehlers ergab, dass über den Richtungsfehler die Lokalisationsschärfe für jede Position im Raum bestimmt werden kann. Die Größe des Fehlers gibt somit die Darstellbarkeit von Objekten im jeweiligen Wiedergabeformat wieder. Im Bezug darauf können Anpassungen bei der Vorverarbeitung durchgeführt werden. Neben dem Lokalisationsfehler wird über die Berechnung von ausgedehnten Objekten ein weiterer Fehler eingeführt. An dieser Stelle geht als Ergebnis hervor, dass der Abstand der Lautsprecher von entscheidender Bedeutung ist. Für die Vorverarbeitung lässt sich daraus schließen, dass je weiter die äußersten Lautsprecher voneinander

entfernt sind, desto breiter ist auch die Quelle und desto schlechter können präzise Phantomschallquellen gebildet werden. In dem Zusammenhang ist darüber hinaus die Anzahl der aktiven Lautsprecher von Bedeutung. Je mehr Lautsprecher benutzt werden, desto besser können Hörereignisse dargestellt werden.

In jedem Fall wurde gezeigt, dass eine Vorverarbeitung bezogen auf das Zusammenfassen von Audioobjekten und die Bewertung von Abweichungen von Objektpositionen sinnvoll ist. Die Ergebnisse bestätigen die Notwendigkeit der Optimierung von Audioszenen über eine Vorverarbeitung. Insgesamt sind in dieser Arbeit bereits eine Reihe interessanter Möglichkeiten der Vorverarbeitung erarbeitet worden, die eine Basis für weiterführende Untersuchungen bezüglich der Anwendbarkeit von objektbasiertem Audio im Rundfunk bildet. Die erhaltenen Werte erzielen objektiv sinnvolle Ergebnisse, welche aber noch subjektiv in Hörversuchen evaluiert werden sollten.

Das Zeitalter des objektbasierten Audios im Rundfunkumfeld schreitet weiter voran. Neben objektbasierten Produktionen sind Tools für Produktion und Mischung vorhanden und Signale können über eine Reihe von Standards überliefert werden. Decoding wird über Apps, Browser und AC-Receiver bereitgestellt. Dennoch sind in nächster Zeit weitere Anpassungen, Erweiterungen und Optimierungen notwendig, um eine innovative, universelle und allgemein verfügbare Technologie für die Medienproduktion- und Wiedergabe zur Verfügung zu stellen.

Literatur

[1] *3d-audio*. 2015. URL: https://kompendium.infotip.de/3d-audio.html (besucht am 28. 12. 2018).

[2] „Audio Definition Model BS.2076“. In: (2017). URL: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2076-1-201706-I!!PDF-E.pdf (besucht am 28. 12. 2018).

[3] Hooke Audio. „The Difference Between Mono, Stereo, Surround, Binaural and 3D Sound“. In: (). URL: https://hookeaudio.com/blog/binaural-3d-audio/difference-mono-stereo-surround-binaural-3d-sound/ (besucht am 01. 04. 2019).

[4] *AUROMAX®Next generation Immersive Sound system*. Nov. 2015. URL: https://www.auro-3d.com/wp-content/uploads/documents/AuroMax_White_Paper_24112015.pdf.

[5] B. Bauer. „Phasor Analysis of Some Stereophonic Phenomena“. In: *Acoustical Society of America Journal* 33 (1961), S. 1536. DOI: 10.1121/1.1908492.

[6] J.C. Bennett. „A New Approach to the Assessment of Stereophonic Sound System Performance“. In: *J. Audio Eng. Soc. vol.33* (Mai 1985), S. 314–321.

[7] J. Berman. „Dolby AC-4 and MPEG-H Vie for ATSC 3.0 Adoption“. In: (Nov. 2017). URL: https://hometheaterreview.com/dolby-ac-4-and-mpeg-h-vie-for-atsc-30-adoption/ (besucht am 01. 04. 2019).

[8] J. Blauert. *Spatial hearing*. The Psychophysics of Human Sound Localization. MIT Press, 1999, S. 41–44, 88–96.

[9] J. Blauert. *Spatial hearing*. The Psychophysics of Human Sound Localization. MIT Press, 1999.

[10] A. D. Blumlein. „U.K. patent 394,325, 1931. Reprinted in Stereophonic Techniques“. In: *Audio Engineering Society, New York* (1986).

[11] T. Carpentier. „Ambisonic spatial blur“. In: Proceedings of the 142nd Convention of the Audio Engineering Society (Mai 2017), S. 4–5. URL: https://hal.archives-ouvertes.fr/hal-01527756.

[12] B. CROCKETT u. a. „OBJECT CLUSTERING FOR RENDERING OBJECT-BASED AUDIO CONTENT BASED ON PERCEPTUAL CRITERIA“. In: *Dolby Laboratories Licensing Corp* (2013). European Patent EP 2 936 485 B1.

[13] R. Curtis. „Immersion und Einfühlung. Zwischen Repräsentationalität und Materialität bewegter Bilder“. In: *montage AV* (2008), 91ff.

[14] P. Damaske und B. Wagener. „Directional Hearing Tests by the Aid of a Dummy Head“. In: *Acta Acustica united with Acustica, Volume 21, Number 1* (1969), S. 30–35.

[15] A. Daniel, R. Nicol und S. McAdams. „Parametric Spatial Audio Coding Based on Spatial Auditory Blurring“. In: (2012).

[16] J. Daniel. „Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia“. Diss. Université Paris 6, 2001.

[17] M. Dickreiter u. a. *Handbuch der Tonstudiotechnik*. De Gruyter, 2009. ISBN: 9783598441356. URL: https://books.google.de/books?id=srApJp3nP4wC.

[18] EBU. „next generation audio“. In: (Sep. 2018). URL: https://tech.ebu.ch/docs/factsheets/ebu_tech_fs_nga.pdf (besucht am 01. 04. 2019).

[19] N. Epain, CT. Jin und F. Zotter. In: *Acta Acustica united with Acustica* 100 (5 1. Sep. 2014), S. 928–936.

[20] T. Fedke. „Kunstkopftechnik - Eine Bestandsaufnahme“. In: *acta acustica, vol.93* (2007).

[21] M. Frank. „Localization Using Different Amplitude Panning Methods in the Frontal Horizontal Plane“. en. In: (März 2014), S. 22, 45. URL: https://depositonce.tu-berlin.de/handle/11303/165 (besucht am 02. 01. 2019).

[22] M. Frank. „Phantom Sources using Multiple Loudspeakers in the Horizontal Plane“. published: online. Diss. University of Music und Performing Arts Graz (Austria): Institut 17 Elektronische Musik und Akustik, Sep. 2013, S. 21–28.

[23] H. Fuchs. „MPEG-H AUDIO“. In: (2018), 10ff. URL: https://orpheus-audio.eu/wp-content/uploads/2018/06/03_mpeg-h_orpheus_2018-05-15.pdf.

[24] M. A. Gerzon. „General Metatheory of Auditory Localisation“. In: *Audio Engineering Society Convention 92*. März 1992. URL: http://www.aes.org/e-lib/browse.cfm?elib=6827.

[25] T. Görne. *Tontechnik. Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis*. Bd. Auflage: 3., neu bearbeitete Auflage. Carl Hanser Verlag GmbH Co. KG, 2011, S. 384. ISBN: 978-3-446-42395-4. DOI: https://doi.org/10.3139/9783446427402.

[26] T. Görne. *Tontechnik. Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis*. Carl Hanser Verlag GmbH Co. KG, 2014. ISBN: 978-3-446-42395-4. DOI: https://doi.org/10.3139/9783446427402.

[27] M. Gröhn. „Localization, accuracy, and utilization of a spatial audio in virtual environments“. In: *CSC - Scientific Computing Ltd. PO Box 405 FIN - 02101 Espoo Finland* (2004).

[28] J. Herre u. a. „MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding“. In: *J. Audio Eng. Soc* 62.12 (2015), S. 821–830. URL: http://www.aes.org/e-lib/browse.cfm?elib=17556.

[29] „Immersion virtuelle Realität“. In: (). URL: https://de.wikipedia.org/wiki/Immersion_(virtuelle_Realit%5C%C3%5C%A4t).

[30] ITU. „Recommendation ITU-R BS.1352-3“. In: (Dez. 2007). URL: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1352-3-200712-I!!PDF-E.pdf.

[31] ITU. „Recommendation ITU-R BS.2051-2“. In: (Juli 2018). URL: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2051-2-201807-I!!PDF-E.pdf.

[32] ITU. „Recommendation ITU-R BS.2088“. In: (Okt. 2015). URL: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2088-0-201510-I!!PDF-E.pdf.

[33] Dolby Laboratories. „Dolby Atmos Next-Generation Audio for Cinema“. In: (2014), S. 4, 7. URL: https://www.dolby.com/us/en/technologies/dolby-atmos/dolby-%20atmos-next-generation-audio-for-cinema-white-paper.pdf (besucht am 01. 11. 2017).

[34] Dolby Laboratories. „Setting the Record Straight on Dolby AC-4 and MPEG-H. An in-depth breakdown of what AC-4 is and how it impacts the industry.“ In: (2019). URL: https://blog.dolby.com/setting-record-straight-dolby-ac4-mpegh/ (besucht am 01. 04. 2019).

[35] D.M. Leakey. „Some Measurements on the Effect ofInterchannel Intensity and Time Difference in Two Channel Sound Systems“. In: *J. Acoust. Soc. Am. vol.31* (Juli 1959), S. 977–986.

[36] Y Makita. „On the directional localization of sound in the stereophonic sound field“. In: (1960).

[37] G. Marentakis, F. Zotter und M. Frank. „Vector-Base and Ambisonic Amplitude Panning: A Comparison Using Pop, Classical, and Contemporary Spatial Music“. In: *Acta acustica united with Acustica. the journal of the European Acoustics Association (EAA) ; international journal on acoustics* 100.5 (Okt. 2014), S. 945–955.

[38] G. Martin u. a. „Sound source localization in a five-channel surround sound reproduction system“. In: *Audio Engineering Society Convention 107* (1999).

[39] JC. Middlebrooks. „Listener weighting of cues for lateral angle: the duplex theory of sound localization“. In: *J Acoustic Society of America* (2002).

[40] M. Miller. „The History of Surround Sound“. In: *informIT* (2004). URL: http://www.informit.com/articles/article.aspx?p=337317&seqNum=2 (besucht am 01. 04. 2019).

[41] A.W. Mills. „On the Minimum Audible Angle“. In: (1958). URL: https://doi.org/10.1121/1.1909553.

[42] „MPEG Surround“. In: *Spatial Audio Processing*. Wiley, J. und Sons, 2007. Kap. 6, S. 93–125. ISBN: 9780470723494. DOI: 10.1002/9780470723494.ch6. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470723494.ch6. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470723494.ch6.

[43] L. Nipkow. „Die Bedeutung von 3D bei Immersive Audio“. In: *vdt 2019 1 S.22 Im Fokus* (Jan. 2019).

[44] „Object-based audio Demos“. In: (2018). URL: https://lab.irt.de/demos/object-based-audio (besucht am 14. 12. 2018).

[45] „Objektbasierter Ton“. In: (2018). URL: https://www.irt.de/themengebiete/av-technologien/objektbasierter-ton (besucht am 03. 04. 2019).

[46] P. Pörs. „Die Next Generation Audio Formate“. In: *vdtMagazin, 2017, Heft 3* (2017), S. 43–46.

[47] R. Preibisch-Effenberger. „Die Schallokalisationsfähigkeit des Menschen und ihre audiometrische Verwendung zur klinischen Diagnostik.“ In: (1966).

[48] V. Pulkki. „Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning“. In: *J. Audio Eng. Soc* 49.9 (2001), S. 753–767.

[49] V. Pulkki. „Uniform spreading of amplitude panned virtual sources“. In: (Feb. 1999), S. 99, 187–190. DOI: 10.1109/ASPAA.1999.810881.

[50] V. Pulkki. „Virtual sound source positioning using vector base amplitude panning“. In: *Journal of the Audio Engineering Society* 45.6 (Juni 1997), S. 456–466. URL: http://lib.tkk.fi/Diss/2001/isbn9512255324/article1.pdf.

[51] F. Rumsey. *Spatial Audio*. Focal Press, 2001.

[52] L. S. R. Simon, R. Mason und F. Rumsey. „Localization curves for a regularly-spaced octagon loudspeaker array“. In: *Audio Engineering Society Convention 127, 10* (2009).

[53] J. Steuer. „Defining Virtual Reality: Dimensions Determining Telepresence“. In: *Journal of Communication* 42.4 (1992), S. 73–93. DOI: 10.1111/j.1460-2466.1992.tb00812.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.1992.tb00812.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-2466.1992.tb00812.x.

[54] „TECH 3388 ADM RENDERER FOR USE IN NEXT GENERATION AUDIO BROADCASTING“. In: (März 2018), S. 8–12, 37–39.

[55] „TECH 3392 ADM BROADCAST PRODUCTION PROFILE“. In: (2018). URL: https://tech.ebu.ch/docs/tech/tech3392.pdf (besucht am 28. 12. 2018).

[56] G. Theile. „Localization of lateral phantom sources“. In: *J. Audio Engineering Society* 25 (1977), S. 196–200.

[57] G. Theile. „Über die Lokalisation im überlagerten Schallfeld“. In: (1980), S. 51.

[58] F. Völk und H. Fastl. „Richtungsunterschiedswellen (Minimum Audible Angles) für ein zirkulares Wellenfeldsynthesesystem in reflexionsbehafteter Umgebung“. In: *DAGA 2011, Düsseldorf* (2011).

[59] S. Weinzierl. *Handbuch der Audiotechnik*. Jan. 2008, S. 657–659. DOI: 10.1007/978-3-540-34301-1.

[60] M. Weitnauer. „Object-based audio rendering in browsers“. In: (Feb. 2017). URL: https://lab.irt.de/object-based-audio-rendering-in-browsers (besucht am 03. 04. 2019).

[61] M. Weitnauer und A. Silzle. „The Orpheus Project: Building the workflow for end-to-end object-based audio broadcasting“. In: *FKT 2018, 1-2, 72. Jahrgang* (2018).

[62] F. Zotter und M. Frank. „All-Round Ambisonic Panning and Decoding“. In: *Audio Engineering Society 60(10)* (2012), S. 807–820.

A. Quellcode

```python
def energy_vector(gains, speaker_positions):
    """ implements carpentier 2017, p4, (19)"""
    gains_squared = gains**2
    return np.dot(gains_squared, speaker_positions) / np.sum(gains_squared)

def decompose_energy_vector(r_E):
    norm = np.linalg.norm(r_E)
    direction = r_E / norm
    return norm, direction

class EnergyVectorErrorCalculator:
    def __init__(self, layout):
        self.layout_ = layout
        self.psp_ = point_source.configure(self.layout_)

    def estimate(self, azimuth, elevation):
        """ calculates energy vector """
        pos_polar = ObjectPolarPosition(azimuth=azimuth, elevation=elevation)
        pos_in = coord_trans(pos_polar, False)
        gains = self.psp_.handle(pos_in)

        vector = energy_vector(gains, self.layout_.positions)
        e_vector_norm, e_vector_dir = decompose_energy_vector(vector)

        return np.linalg.norm(pos_in - e_vector_dir)
```

Quelltext 1: Energy Vector Berechnung in Python

```python
def decompose_energy_vector(r_E):
    norm = np.linalg.norm(r_E)
    direction = r_E / norm
    return norm, direction

def angular_spread(e_vector_norm):
    return 2 * np.arccos(2 * e_vector_norm -1)

class BlurredErrorCalculator:
    def __init__(self, layout):
        self.layout_ = layout
        self.psp_ = point_source.configure(self.layout_)

    def estimate(self, azimuth, elevation):
        pos_polar = ObjectPolarPosition(azimuth=azimuth, elevation=elevation)
        pos_in = coord_trans(pos_polar, False)
        gains = self.psp_.handle(pos_in)
        indices_not_zero = np.nonzero(gains)

        vector = energy_vector(gains, self.layout_.positions)
        e_vector_norm, e_vector_dir = decompose_energy_vector(vector)
        return (np.linalg.norm(pos_in - e_vector_dir) +1) * (2-e_vector_norm)
```

Quelltext 2: Angular Spread Berechnung in Python

```python
def create_maa_az_interpolator():
    azimuth_werte = [-90,0,90,180,270,360,360+90]
    abweichung_az = [10.,3.6,9.2,5.5,10.,3.6,9.2]
    return interpolate.interp1d(azimuth_werte,
                                abweichung_az,
                                'quadratic')

def create_maa_el_interpolator():
    elevation_werte = [30,0,30,74,112,153,180]
    abweichung_el = [10,9,10,13,22,15,9]
    return interpolate.interp1d(elevation_werte,
                                abweichung_el,
                                'linear')

maa_az_estimator = create_maa_az_interpolator()
maa_el_estimator = create_maa_el_interpolator()

def estimate_energy_vector(layout, object_position_polar):
    pos_input = coord_trans(object_position_polar, False)
    point_source_panner = point_source.configure(layout)
    gains = point_source_panner.handle(pos_input)
    indices_not_zero = np.nonzero(gains)
    vector = energy_vector(gains[indices_not_zero],
                           layout.positions[indices_not_zero])
    return vector

def estimate_rendered_position(layout, position_polar):
    vector = estimate_energy_vector(layout, position_polar)
    estimated_az = ear_azimuth([vector]).round(0)
    estimated_el = ear_elevation([vector]).round(0)
    return estimated_az, estimated_el

def estimate_position_error(layout, position_polar):
    estimated_az, estimated_el = estimate_rendered_position(layout,
                                                            position_polar)
    delta_az = np.abs(estimated_az - position_polar.azimuth)
    delta_el = np.abs(estimated_el - position_polar.elevation)
    return delta_az, delta_el

def estimate_weighted_position_error(layout, position_polar):
    maa_az = maa_az_estimator(position_polar.azimuth)
    maa_el = maa_el_estimator(position_polar.elevation)
    delta_az, delta_el = estimate_position_error(layout, position_polar)
    error_az = delta_az / maa_az
    error_el = delta_el /maa_el
    return error_az, error_el
```

Quelltext 3: Gewichtete Fehlerberechnung in Python

Abbildungsverzeichnis

Tabellenverzeichnis

Quelltextverzeichnis