

# *SONY 360 REALITY AUDIO/ MPEG-H 3D AUDIO*

Tonseminar (Wahlfach CR7)

*Louis Göllner*

**Dozent: Prof. Oliver Curdt**  
**Eingereicht am: 29.02.2024**

1 Einleitung .....	2
2 Immersives Audio.....	2
3 MPEG-H.....	3
3.1 MPEG-H 3D Audio .....	4
3.2 Metadaten von MPEG-H 3D Audio Codec .....	4
3.2.1 Rendering .....	7
4 Creation, Streaming, Playback .....	9
4.1 MPEG-H 3D Audio in der Produktion .....	9
4.2 MPEG-H 3D Audio in der Distribution .....	11
4.3 MPEG-H 3D Audio im Playback .....	11
5 Weitere Anwendungen von MPEG-H 3D Audio.....	11
Abbildungsverzeichnis .....	13
Literaturverzeichnis.....	14
Eidesstattliche Erklärung.....	15

# 1 Einleitung

Bei Sonys 360 Reality Audio handelt es sich um ein immersives hochauflösendes Audioformat, welches auf dem Audiostandard MPEG-H 3D Audio beruht. Im Folgenden soll eine kurze Einführung immersives Audio sowie den Audiostandard MPEG-H 3D Audio erklären. Anschließend wird sich der Anwendung von Sony 360 Reality Audio in der Praxis gewidmet, bevor abschließend spezielle Praxisbeispiele zur Anwendung von MPEG-H 3D Audio veranschaulicht werden.

## 2 Immersives Audio

Immersives Audio bezieht sich auf eine audiotecnologische Erfahrung, bei der Hörer\*innen vollständig in die klangliche Umgebung eintauchen können. Der Begriff "Immersion" leitet sich vom lateinischen Wort "immersio" ab, was "Eintauchen" bedeutet<sup>1</sup>. Nach der allgemeinen Definition von Bilandzic<sup>2</sup> bedeutet Immersion, dass Rezipient\*innen vollständig in die durch Medien vermittelte Welt eintauchen und die darin präsentierten Gegebenheiten und Erlebnisse unmittelbar und direkt nachvollziehen können, fast so, als würden sie diese selbst in der realen Welt erleben.

Wilfried van Baelen prägte den Begriff "Immersive Sound" im Jahr 2010 als einen Oberbegriff für Mehrkanalformate, die zusätzliche Höhenlautsprecher verwenden<sup>3</sup>. Diese Technologie ermöglicht eine erweiterte räumliche Wiedergabe von Audioinhalten, die den Hörern ein Gefühl der Präsenz und Tiefe vermittelt. Immersives Audio wird häufig auch als "3D Audio" bezeichnet, da es den Klang in drei Dimensionen reproduziert, was eine realistischere und immersivere Hörerfahrung für Nutzer\*innen ermöglicht. Bei Sony 360 Reality Audio handelt es sich um immersives 3D-Audio.

Immersives Audio, wie 360 Reality Audio wird in der Unterhaltung schon heute regelmäßig genutzt. Neben der Filmindustrie wird es vermehrt auch in der Gaming-Welt, gerade bei Virtual-Reality-Anwendungen eingesetzt und eignet sich für viele Formen der digitalen Unterhaltung. Es birgt den Vorteil, dass Nutzer\*innen neben den visuellen Elementen (etwa bei VR) auch auf der visuellen Ebene in die ihnen präsentierte Welt eintauchen können.

---

<sup>1</sup> Meier (2004), S. 15, Eintrag „Eintauchen“ (zitiert nach de.wiktionary.org, Zugriff am 29.02.2024)

<sup>2</sup> Vgl. Bilandzic (2014), S. 273 (zitiert nach Bielecki (2019) S. 3)

<sup>3</sup> Vgl. Van Baelen (2016), S. 19 (zitiert nach Bielecki (2019) S. 3)

### 3 MPEG-H

MPEG-H ist ein multimedialer Standard, der von der Moving Picture Experts Group (MPEG) in Zusammenarbeit mit dem Fraunhofer IIS und anderen Partnern entwickelt wurde. Dieser Standard dient der Übertragung digitaler Medien und umfasst sowohl Video-Kompression als auch einen speziellen Audio-Standard, bekannt als MPEG-H 3D Audio.

Der MPEG-H-Standard bietet eine breite Palette von Funktionen, die die Übertragung und Wiedergabe von Audio- und Videodaten verbessern. Dabei liegt der Fokus nicht nur auf der Kompression von Bild- und Tondaten, sondern auch auf der Bereitstellung immersiver und interaktiver Audioerlebnisse. MPEG-H 3D Audio ermöglicht es den Produzenten, Audioinhalte mit einer erweiterten räumlichen Dimension zu versehen, was zu einer realistischeren und immersiveren Wiedergabe für Endverbraucher\*innen führt. Die Integration von Audio- und Videokomprimierungstechnologien ermöglicht zudem eine effiziente und hochwertige Übertragung von Inhalten über verschiedene Medienkanäle.

In Südkorea wurde MPEG-H schon früh nach der Veröffentlichung in Fernsehformaten zur Übertragung genutzt. Heute nutzen die Südkoreaner\*innen MPEG-H als den multimedialen Standard ihres gesamten UHD-Fernsehservices. Brasilien hat zudem angekündigt, MPEG-H 2024 als Teil ihres revolutionären TV 3.0 Fernsehservices an den Start zu bringen<sup>4</sup>. Fernsehzuschauer\*innen können dann etwa bei linearen Sportveranstaltungen in Echtzeit Audiokommentare wechseln und mit unterstützten Soundbars immersiv in den Stadionsound eintauchen. Ein erster Belastungstest bei der Fußball-WM in Katar im Jahr 2022 war ein Erfolg.<sup>5</sup>

Neben MPEG-H 3D Audio gibt es auch andere immersive Audiottechnologien wie Dolby Atmos und DTS:X, die in verschiedenen Anwendungsbereichen eingesetzt werden und den meisten Endverbrauchenden wohl zumindest namentlich bekannter sind. Durch den Einsatz von MPEG-H 3D Audio in Sonys 360 Reality Audio, welches wiederum in Sonys neuesten Kopfhörern<sup>6</sup> verbaut und intensiv beworben wird erhält MPEG-H auch immer mehr Aufmerksamkeit von Endverbraucher\*innen.

---

<sup>4</sup> Vgl. Pressemeldung Fraunhofer IIS <https://www.audioblog.iis.fraunhofer.com/de/globo-mpeg-h-fussball>

<sup>5</sup> Vgl. CSI-Magazine (2023), Ausgabe *spring 2023*, Abgerufen von [https://www.csimagazine.com/eblast/Digital\\_Editions/Spring\\_2023/CSIspring2023.pdf](https://www.csimagazine.com/eblast/Digital_Editions/Spring_2023/CSIspring2023.pdf) am 29.02.2024

<sup>6</sup> Z.B. Sony WH-1000MX5

### **3.1 MPEG-H 3D Audio**

MPEG-H 3D Audio umfasst bis zu 128 Audiokanäle, 128 Audio-Objekte und unterstützt Higher Order Ambisonics bis zur 29. Ordnung. Obwohl solche enormen Kapazitäten theoretisch vorhanden sind, sind sie in den meisten Anwendungsfällen nicht erforderlich und überschreiten die Anforderungen aus der Praxis. Aus diesem Grund wurden sogenannte Low Complexity Profile (LC) eingeführt, um den Rahmen für typische Anwendungsfälle zu setzen und keine aufwändige Rechenarbeiten anwenden zu müssen.

Ein Beispiel dafür ist das „Low Complexity 3-Profil“ (LC3-Profil), das für gewöhnlich etwa in einer SDI-Infrastruktur<sup>7</sup> Anwendung findet. Ein solches SDI wird in professionellen Fernsehstudios eingesetzt und dient der unkomprimierten und unverschlüsselten Übertragung von Videodaten. Bei dem LC3-Profil können 32 Audio-Elemente innerhalb eines Bitstreams transportiert werden, wobei gleichzeitig 16 Audiokanäle decodiert werden können. Es ermöglicht auch die Verwendung von Higher Order Ambisonics bis zur 6. Ordnung.

MPEG-H 3D Audio ist dabei nicht auf spezifische Wiedergabegeräte beschränkt. Eine Wiedergabe von MPEG-H 3D Audio, und damit auch von Sony 360 Reality Audio kann auf einer Vielzahl an unterschiedlichsten Geräten gewährleistet werden: darunter Heimkinosysteme, Smart Speaker, Kopfhörer, Laptops, andere mobile Endgeräte und VR-Geräte.

### **3.2 Metadaten von MPEG-H 3D Audio Codec**

Der MPEG-H 3D Audio Codec zeichnet sich durch die Unterstützung objektbasierter Übertragungskonzepte aus. Dies ermöglicht eine höhere Flexibilität im Vergleich zu kanalbasierten Ansätzen, da die einzelnen Audio-Elemente separat übertragen und erst am Endgerät final gerendert werden. Dieser Ansatz ermöglicht es den Endnutzer\*innen, individuelle Anpassungen bis zum letzten Schritt vorzunehmen, was einen großen Vorteil darstellt und mehr Interaktivität verspricht. Das ist sicher auch ein Grund dafür, dass er von Sony genutzt wurde, um das 360 Reality Audio zu schaffen.

Die Audio-Elemente werden bei MPEG-H 3D Audio mit zusätzlichen Metadaten versehen, die vor der Wiedergabe von einem Renderer in Audiokanäle umgerechnet werden können. Diese Metadaten enthalten strukturelle Informationen über die Audiosignale sowie Angaben zur Art und zum Umfang der Benutzerinteraktion und dynamische Objekteigenschaften, wie etwa

---

<sup>7</sup> Serial/ Standard Digital Interface

Position und Bewegung im dreidimensionalen Raum. Diese Metadaten sind entscheidend für die präzise und immersive Wiedergabe von 360 Reality Audio<sup>8</sup>.

Der objektbasierte Ansatz bietet zudem eine effiziente Nutzung der Bandbreite, da die Audio-Mischung sich auf verschiedene Wiedergabesysteme adaptieren lässt, ohne dass für jede Konfiguration ein separater Mix erstellt und übertragen werden muss. Dies optimiert die Übertragung von Audioinhalten, insbesondere in Live-Broadcasts und Echtzeitanwendungen.

In Offline-Workflows werden die Metadaten üblicherweise in Form einer XML-Datei gespeichert, die strukturierte Informationen über die Audio-Elemente enthält. Im Gegensatz dazu wird in Live-Broadcasts und anderen Echtzeitanwendungen ein Control-Track generiert, der Informationen über die gewünschten Objekteigenschaften und strukturelle Angaben zu den verschiedenen Audiosignalen enthält<sup>9</sup>. Diese Flexibilität ermöglicht es, dass die Audioproduktion je nach Anwendungsfall und Workflow effizient angepasst werden kann. Die Verwendung von standardisierten Metadatenformaten ermöglicht zudem die Interoperabilität zwischen verschiedenen Systemen und Plattformen. Dies erleichtert die Integration von MPEG-H 3D Audio in bestehende Produktions- und Wiedergabesysteme.

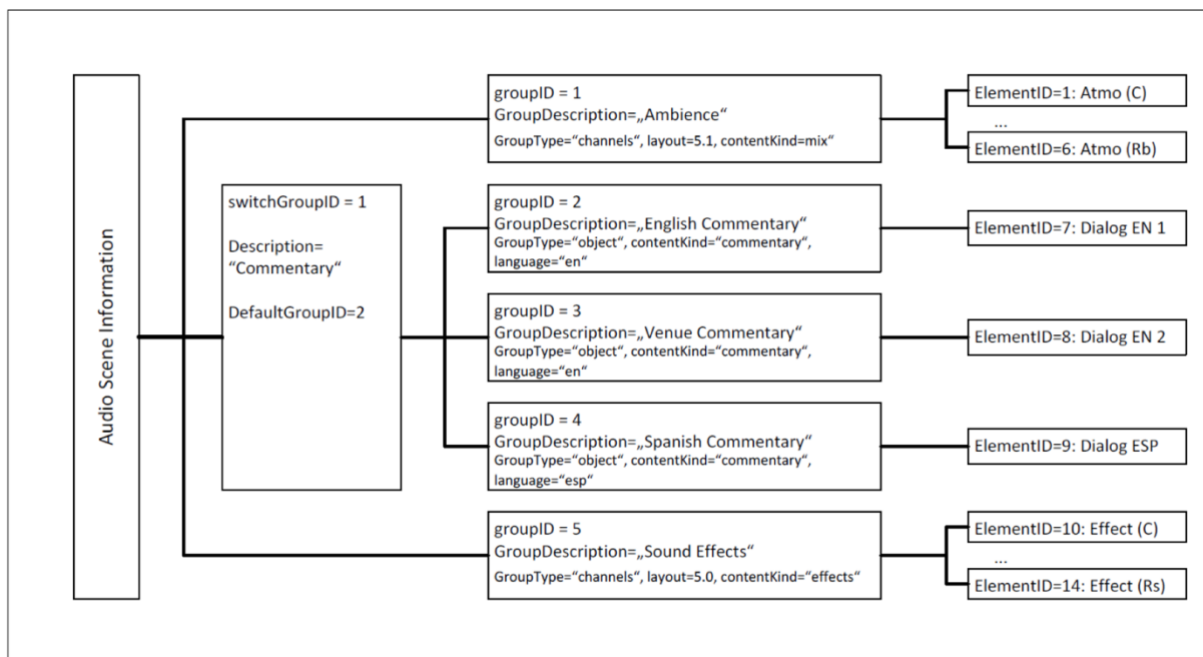


Abbildung 1 Beispiel der Metadaten einer MPEG-H Audio Szene | ATSC (2017)

Im Folgenden sollen einige Bestandteile der Abbildung 1 näher erklärt werden.

- **ElementID oder groupID**

<sup>8</sup> Füg & Kuntz (2015)

<sup>9</sup> Bleidt (2015)

- Die ElementID ist ein zentraler Bestandteil des MPEG-H Audio-Systems und fungiert als eindeutige Kennung für jedes einzelne Audioelement innerhalb der Audioszene. Anhand dieser kann jedes Audioobjekt in der Szene gesteuert werden.
- ***GroupDescription***
  - Die GroupDescription dient dazu, komplexe Audioszenen effizient zu organisieren und zu steuern. Durch die Gruppierung von Elementen zu einzelnen Gruppen können Rendering-Parameter und Verhaltensweisen auf einer höheren Ebene festgelegt werden und auf eine gesamte Gruppe innerhalb einer Szene angewandt werden. Dies trägt zur Flexibilität und Effizienz des MPEG-H 3D Audio-Codecs bei, insbesondere bei der Handhabung komplexer Audioinhalte mit mehreren Elementen. Die switchGroupID 1 in der Abbildung 1 umfasst beispielhaft die vorhandenen Sprachen des Audiokommentars. Den Metadaten kann man entnehmen, dass in diesem Fall ein Default vorbestimmt wurde: per default spielt immer der englische Kommentar, soweit nicht anders vom Nutzenden eingestellt. Die GroupID kann ebenfalls zur eindeutigen Identifizierung von verschiedenen Arten von Gruppen verwendet werden.
- ***contentKind***
  - Der contentKind definiert die Art des Audiomaterials innerhalb der Audioszene. Durch diese Klassifizierung kann zwischen verschiedenen Arten von Audiodaten wie Sprache, Musik, Effekten oder anderen unterschieden werden. Diese Information ist wichtig, um unterschiedliche Verarbeitungs- und Rendering-Parameter anzuwenden, die spezifisch auf die jeweilige Art des Audiomaterials zugeschnitten sind.
- ***GroupType***
  - In MPEG-H 3D Audio wird zwischen kanalbasiertem und objektbasiertem Audio unterschieden. Kanäle repräsentieren traditionelle Lautsprecherpositionen, während Objekte frei im Raum platziert werden können. Die Entscheidung, ob ein Audioelement als Kanal oder als Objekt behandelt wird, hängt von den Anforderungen des Szenarios ab. Diese Flexibilität ermöglicht eine präzise und immersive Wiedergabe von Audioinhalten, die den individuellen Bedürfnissen und Vorlieben entspricht.

- **Layout**

- Bei der Angabe des Layouts handelt es sich um die Beschreibung der Anordnung der Lautsprecher in einem bestimmten Audiosystem. Die Bezeichnung "5.1" oder "5.0" gibt dabei Auskunft über die Anzahl der Hauptlautsprecher und die Anordnung des Subwoofers. Zum Beispiel bezieht sich "5.1" auf ein System mit fünf Hauptlautsprechern (vorne links, vorne rechts, hinten links, hinten rechts und mittig) sowie einem Subwoofer für tiefe Frequenzen, während "5.0" darauf hinweist, dass kein separater Subwoofer verwendet wird.

- **Element Position und Bewegung**

- Diese Metadaten beschreiben die genaue Position und Bewegung eines Audioelements im dreidimensionalen Raum. Sie umfassen Parameter wie Azimut (horizontale Position), Elevation (vertikale Position) und Distanz zum Hörer. Durch die Erfassung dieser Daten wird eine präzise Platzierung der Audioobjekte im Raum ermöglicht, was zu einem realistischen und immersiven

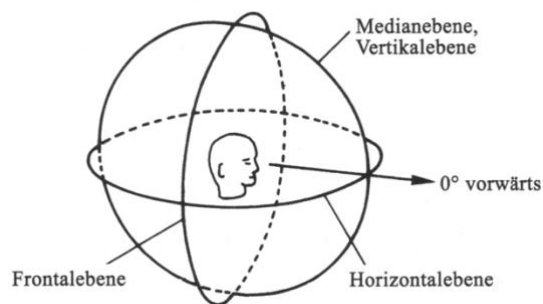


Abbildung 2 Kopfbezogenes Koordinatensystem | Dickreiter et al. (2014)

Klangerlebnis beiträgt. Zum Beispiel kann die Position eines Instruments im virtuellen Raum festgelegt werden, um einen bestimmten räumlichen Effekt zu erzeugen oder eine bestimmte Hörperspektive zu simulieren.

### 3.2.1 Rendering

Vor der Ausgabe des Audio-Streams an das Endgerät der Nutzer müssen mehrere Schritte durchlaufen werden. Zunächst muss der Audio-Stream decodiert werden, um die darin enthaltenen Daten zu extrahieren. Anschließend müssen die einzelnen Ausgabesignale berechnet werden, um ein immersives und hochwertiges Klangerlebnis zu gewährleisten.

Für diesen Prozess werden bei MPEG-H Audio verschiedene Rendering-Module eingesetzt. Diese Module sind dafür verantwortlich, die decodierten Audiodaten in die spezifischen Ausgabesignale umzuwandeln, die auf dem jeweiligen Wiedergabesystem wiedergegeben werden können. Während des Rendering-Prozesses verarbeitet der Decoder die Position, Bewegung, Eigenschaften und Einstellungen der Audioobjekte, was in den Metadaten gespeichert ist.



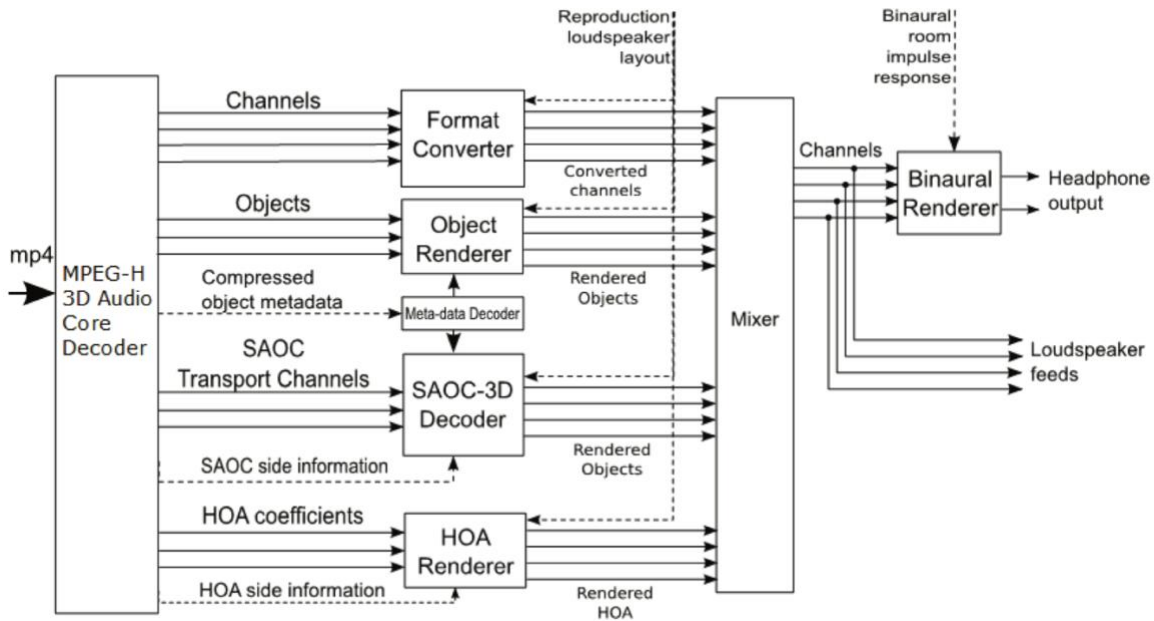


Abbildung 3 Aufbau eines MPEG-H 3D Audio Decoder | Füg & Kuntz (2015)

Die mp4 Datei trifft auf den *MPEG-H Decoder*. Dabei werden die einzelnen Kanäle, Objekte, Metadaten und auch Informationen über Higher Order Ambisonics aufgeteilt. Sie durchlaufen jeweils einzelne Module im Decoder.

Der *Format Converter* rendert flexibel die Kanäle auf die vorhandenen Lautsprecher. Er passt die Audioausgabe an die spezifischen Eigenschaften des Wiedergabesystems an und optimiert so das Hörerlebnis. Er wird mit Informationen über den Ausspielweg versorgt.

Der *Object Renderer* erzeugt die Lautsprechersignale basierend auf den mitgelieferten Metadaten, die ihm vom Metadaten-Decoder zugeliert werden. Er berücksichtigt auch die Benutzerinteraktivität. Eine interessante Funktion des Object Renderers ist die Fähigkeit, virtuelle Lautsprecher zu generieren, wenn nicht genügend physische Lautsprecher vorhanden sind. Dadurch kann das Klangerlebnis auch auf Systemen mit begrenzten Lautsprecherkonfigurationen optimiert werden<sup>10</sup>.

Der *Metadaten-Converter* dekodiert die Metadaten der mp4-File. Die Metadaten gibt er weiter an den *Object Renderer* und den *SAOC-3D Decoder*.

Der *SAOC-3D Decoder* dekodiert Audioobjekte, die mit der „Spatial Audio Object“ Coding-Technologie codiert wurden. Es ermöglicht eine präzise Wiedergabe im dreidimensionalen Raum, wobei Position, Bewegung und andere Eigenschaften der Objekte berücksichtigt werden. Diese Technologie ist spezifisch für MPEG-H 3D Audio und trägt zur Erzeugung eines

<sup>10</sup> Bleidt et al. (2017), S. 212 (zitiert nach Bielecki (2019) S. 19)

immersiven Klangbildes bei.<sup>11</sup>

Der Higher-Order-Ambisonics (HOA) Renderer verrechnet die decodierten Signale des Audio-Streams mit einer Rendering-Matrix, um daraus die einzelnen Lautsprechersignale zu generieren. Es ermöglicht eine präzise Steuerung der räumlichen Wiedergabe und trägt zur Schaffung eines naturgetreuen Klangbildes bei<sup>12</sup>.

Der gesamte Output dieser Rendering-Module wird im finalen Schritt durch einen *Mixer* geleitet, um die fertige *Output Audio Data* zu erhalten. Der Mixer kombiniert die einzelnen Signale und optimiert sie mithilfe der vorhandenen Metadaten für die Ausgabe an das Wiedergabesystem.

Bei MPEG-H 3D Audio gibt es zudem ein binaurales Rendering-Modul, das die vorhandenen Audiosignale unter Berücksichtigung aller vorherigen Schritte kanal-, objekt- oder szenenbasiert auf den Stereoausgang für Kopfhörer mit 360 Reality Audio Unterstützung ausgeben kann. Dies ermöglicht ein immersives und realistisches Hörerlebnis, selbst wenn keine Mehrkanallautsprecher verfügbar sind. Auch dafür wurde der MPEG-H Standard auf MPEG-H 3D weiterentwickelt.

## 4 Creation, Streaming, Playback

Damit 360 Reality Audio von Endnutzer\*innen wiedergegeben werden kann, muss der fertige Track einen speziellen Schaffensprozess durchlaufen, beginnend mit der Erstellung in den Musik- und Tonstudios, in denen für MPEG-H 3D Audio produziert wird. Die Verwendung von MPEG-H Audio in der Produktion eröffnet neue kreative Möglichkeiten für Toningenieur\*innen und Musikproduzent\*innen. Etwa durch die Manipulation von Audioobjekten und Metadaten können sie ein breiteres Spektrum von räumlichen und immersiven Klangerlebnissen schaffen, die mit herkömmlichen Audioformaten nicht möglich wären. Es folgt ein kurzer Einblick in die Verwendung von MPEG-H 3D Audio in der Produktion.

### 4.1 MPEG-H 3D Audio in der Produktion

Für die Produktion von MPEG-H Audio muss die Infrastruktur in den Musikstudios angepasst

---

<sup>11</sup>Engdegård et al. (o.D.)

<sup>12</sup> Bleidt et al. (2017), S. 213ff. (zitiert nach Bielecki (2019) S. 19)

werden. Dies kann durch die Verwendung von Digital Audio Workstations (DAWs), Plugins oder Hardwarekomponenten erfolgen. Ein Beispiel für eine Veränderung wäre die Anschaffung einer sogenannten **Audio Monitoring und Authoring Unit (AMAU)**. Sie wird bei der Bearbeitung von MPEG-H 3D Audio Dateien und der Überwachung bei Aufnahmen und im Mix genutzt. Es gibt verschiedene Möglichkeiten die *AMAU* in einem Studio einzusetzen:

- **Beispiel 1: MMA Multichannel Monitoring and Authoring (Jünger Audio)**

Diese Lösung ermöglicht es den Produzenten, sämtliche Parameter und Metadaten zu manipulieren und sogar Presets für Audioinhalte einer gemeinsamen Gruppe zu erstellen. Eine besonders nützliche Funktion ist die Möglichkeit, die Platzierung der einzelnen Audioobjekte im dreidimensionalen Raum anzeigen zu lassen, was es den Produzenten ermöglicht, die räumliche Anordnung der Klänge zu visualisieren und anzupassen. Diese Visualisierung kann auch an das jeweilige Lautsprecher-setup angepasst werden, was eine präzise Kontrolle über das Klangbild ermöglicht. Zudem kann das Tool von Jünger Audio zur Sprachverstärkung in Aufnahmen genutzt werden.

- **Beispiel 2: 360 WalkMix Creator – Plugin für die gängigsten DAWs**

Der 360 WalkMix Creator ist ein Plugin, das für die gängigsten Digital Audio Workstations (DAWs) verfügbar ist. Es ermöglicht eine intuitive Visualisierung und Bearbeitung der einzelnen Audioobjekte. Durch die grafische Benutzeroberfläche können Benutzer die Metadaten der Audioobjekte einfach manipulieren, indem sie die Objekte beispielsweise im Raum verschieben oder ihre Eigenschaften anpassen. Dies ermöglicht es den Produzenten, das Konzept des immersiven Audios auf spielerische Weise zu erkunden und die räumliche Anordnung der Klänge genau zu steuern. Es ist jedoch anzumerken, dass der 360 WalkMix Creator ein kostenpflichtiges Plugin ist: Der Preis liegt aktuell bei etwa 540€<sup>13</sup> für eine Lizenz. Trotz des Preises bietet dieses Tool eine effektive Möglichkeit, die Produktion von 360 Reality Audio zu erleichtern und das Potenzial dieser innovativen Audiotechnologie voll auszuschöpfen. Ein Vorteil des 360 WalkMix Creator könnte auch sein, dass es den nativsten Support für MPEG-H 3D Audio mitbringt: Es wurde von Sony selbst zu genau diesem Zweck geschaffen.

---

<sup>13</sup> Stand: Januar 2024

## 4.2 MPEG-H 3D Audio in der Distribution

Der zweite Schritt bei der Implementierung von 360 Reality Audio betrifft die Distribution oder das Streaming der Inhalte. Dabei geht es insbesondere um die Bereitstellung von 360 Reality Audio-fähigen Tracks, also von Dateien im MPEG-H-Audioformat, auf verschiedenen Streaming-Plattformen. Aktuell umfasst dieses Angebot in Deutschland folgende Plattformen: TIDAL, Deezer, nugs.net und Amazon Music Unlimited. Laut Sony soll es plattformübergreifend mehr als 7.000 Songs im 360 Reality Audio Format geben.

Nutzende der oben genannten Plattformen erhalten automatisch Zugriff auf 360 Reality Audio-Inhalte. Durch die Bereitstellung von 360 Reality Audio auf verschiedenen Streaming-Diensten wird die Reichweite dieser innovativen Audiotechnologie maximiert und ermöglicht es einer breiten Öffentlichkeit, die Vorteile von immersivem, dreidimensionalem Audio zu erleben.

## 4.3 MPEG-H 3D Audio im Playback

Sony nutzt den in 3.2 beschriebenen Encoder unter anderem schon in seinen eigenen Smartphones – die Xperia-Reihe hat eine leicht veränderte Form des Encoders an Bord. Dadurch werden Funktionen, wie Head-Tracking ermöglicht, was wiederum zu einem optimalen Hörerlebnis führt: Über Sonys App „Headphones“ und die Kopfhörer der Reihe WH1000MX5 können Nutzer\*innen seit Ende 2023 zudem dynamisches Head-Tracking auf Smartphones mit Android-Betriebssystem<sup>14</sup> einmalig ihre Ohren einscannen und das Head-Tracking und Up-Mixing der Kopfhörer passen sich dementsprechend an.

Alternativ kann Sony 360 Reality Audio (ohne Head-Tracking) aber natürlich auch weiterhin mit jedem anderen Smartphone und einem dazugehörigen Abonnement bei teilnehmenden Musikstreaming-Diensten genutzt werden.

# 5 Weitere Anwendungen von MPEG-H 3D Audio

Der Sound Codec hinter Sonys 360 Reality Audio, MPEG-H 3D Audio kann aber nicht nur im klassischen Musikstreaming genutzt werden.

Auch Autohersteller nutzen den Sound Codec MPEG-H für ihr Entertainment-System. Die Integration eines immersiven Audio-Streaming-Dienstes direkt in die Haupteinheit eines Autos ermöglicht die direkte Übertragung an einen Decoder, der MPEG-H-Audioinhalte einschließlich 360 Reality Audio umwandelt. Anschließend erzeugt ein Renderer basierend auf

---

<sup>14</sup> Ab Version 13

der vorhandenen Lautsprecheranordnung ein umfassendes Klangerlebnis für Passagiere. Gerade in High-End Modellen bietet sich eine Nutzung von MPEG-H an, hier kann mit dem immersiven Sound geworben werden. Noch findet man MPEG-H lediglich als Video-Format an Bord mancher Porsche Modelle<sup>15</sup>.

Zudem könnte MPEG-H 3D mit seiner Variabilität auch von Personen mit Höreinschränkungen genutzt werden. Ein speziell eingestellter Fernsehton mit Sprachverstärkung anhand von MPEG-H Metadaten ist durchaus vorstellbar.

---

<sup>15</sup> Laut den Handbüchern der Porsche-Modelle Macan, Panamera und Cayenne

**Nützliche Links zu MPEG-H 3D Audio/ 360 Reality Audio:**

<https://mpegh.com/demo-content/>

<https://mpegh.com/mas/>

# Abbildungsverzeichnis

Abbildung 1 Beispiel der Metadaten einer MPEG-H Audio Szene   ATSC (2017) .....	5
Abbildung 2 Kopfbezogenes Koordinatensystem   Dickreiter et al. (2014) .....	7
Abbildung 3 Aufbau eines MPEG-H 3D Audio Decoder   Füg & Kuntz (2015).....	8

# Literaturverzeichnis

*ATSC/342 Part 3, MPEG-H System.* (2017).

Bielezki, J. (o. J.). *Workflow-Analyse von immersiven Audioproduktionen mit Produktionsmischpulten für moderne Standards wie MPEG-H und Dolby Atmos.*

Bleidt, R. (2015). *Installing the MPEG-H Audio Alliance's New Interactive and Immersive TV Audio System in Professional Production and Distribution Facilities* (Technical Brief). Fraunhofer USA Digital Media Technologies.

Dickreiter, M., Dittel, V., Hoeg, W., & Wöhr, M. (Hrsg.). (2014). *Handbuch der Tonstudioteknik.*

De Gruyter Saur. <https://doi.org/10.1515/9783110316506>

Engdegård, J., Resch, B., Falch, C., Hellmuth, O., Hilpert, J., Hoelzer, A., Breebaart, J., Koppens, J., Schuijers, E., & Oomen, W. (o. J.). *Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding.*

Füg, S., & Kuntz, A. (2015). *An Introduction to MPEG-H 3D Audio.*

# Eidesstattliche Erklärung

Hiermit versichere ich, Louis Göllner, dass ich diese Arbeit selbständig und ohne fremde Hilfe verfasst habe. Die von mir verwendeten Quellen, Literatur und Hilfsmittel sind entsprechend angegeben.



Berlin, 29.02.2024