



Objekt- und kanalbasierte Audioformate

Vor- und Nachteile

Tonseminar WS17/18

Max Utke

Matrikel-Nr.:32238

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Abkürzungsverzeichnis	3
1. Grundlagen	4
2. Kanalbasierte Audioformate	5
2.1. IFF.....	5
2.2. RIFF WAVE.....	5
2.3. BWF.....	5
2.4. MPEG.....	6
2.5. AC3.....	6
3. Objektbasierte Audioformate	7
3.1. AES-31.....	7
3.2. ASF.....	7
3.3. SDIF.....	8
3.4. MPEG-4.....	8
3.5. AC-4	8
4. Vor- und Nachteile der Audioformate	9
Quellenverzeichnis	11

Abkürzungsverzeichnis

AAC	Advanced Audio Coding
AAF	Advanced Authoring Format
AC3	Adaptive Transform Coder 3
AES	Audio Engineering Society
AIFF	Audio Interchange File Format
ASF	Advanced Systems Format
BWF	Broadcast Wave Format
BWFF	Broadcast Wave Format File
DAB	Digital Audio Broadcast
DTS	Digital Theater Systems
DVB	Digital Video Broadcast
EA	Electronic Arts
EBU	European Broadcast Union
IFF	Interchange File Format
HRTF	Head-Related Transfer Function
MADI	Multi Channel Audio Digital Interface
MIDI	Musical Instrument Digital Interface
MPEG	Moving Pictures Experts Group
MXF	Material Exchange Format
OMFI	Open Media Framework Interchange
PCM	Pulse Code Modulation
RIFF	Resource Interchange File Format
RTF	Rich Text Format
SDIF	Sound Description Interchange Format
WMA	Windows Media Audio

1 Grundlagen

Bei einem Dateiformat handelt es sich um ein Regelsystem von Zeichen zur Datenanordnung. So können physikalische Werte (z.B. Spannung) in Binärdaten gespeichert werden. Audio-Dateiformate beschreiben dabei hauptsächlich Containerformate. Dies sind „logische Behälter, in denen aus einem gemeinsamen Ursprung stammende und daher zusammengehörige Daten in einer definierten Struktur abgelegt und angeordnet sind.“ (Dickreiter, 2014, S. 704)

Basis aller Audioformate ist die PCM zur Wandlung des analogen Signals in eine Abfolge von Zahlen, sogenannten Codeworten. Unterschieden wird dabei in einer linearen Quantisierung (alle Wertebereiche werden gleichmäßig aufgelöst), einer nichtlinearen/dynamischen Quantisierung (Bereiche mit kleinen Signalauslenkungen werden genauer aufgelöst um Quantisierungsrauschen zu unterdrücken)² und in der verlustfreien, beziehungsweise verlustbehafteten Kompression. Bei letzterer wird zugunsten kleiner Datenmengen (geringer Speicherplatz/hohe Übertragungsraten) die Wiedergabequalität im Vergleich zum Original verringert.

Die Abbildungsvorschrift nennt man Kodierung. Ein Containerformat kann mehrere Kodierungen und damit auch unterschiedliche Audioformate beinhalten. Daneben werden weitere Parameter wie der Adressbereich der eigentlichen Audiodatei gespeichert, aber auch Video- oder Textformate und damit auch Metadaten (Zusatzinformationen). Zum Beispiel kann das MPEG-Format aus WAVE-Dateien bestehen und ID3-Tags Angaben zu Genre, Künstler und/oder Herstellungsjahr enthalten.

Basiseigenschaften eines Audioformats sind die Byte-Reihenfolge (*Big/Little Endian*), Zahlendarstellung (feste Werte, oder Gleitkommastellen), Ausrichtung/Adressierung der Datenblöcke (*Alignment/Padding*) und bei kanalbasierten Formaten noch die Mehrkanalspeicherung (*interleaved/seperated*). Diese Eigenschaften wirken sich auf die Rechengeschwindigkeit aus. So ist ein 24-bit Format aufwendiger als ein 16-bit kodiertes, da statt 2 Byte pro Datenwort 3 Byte, also das 1,5-fache an Rechenaufwand benötigt wird.³

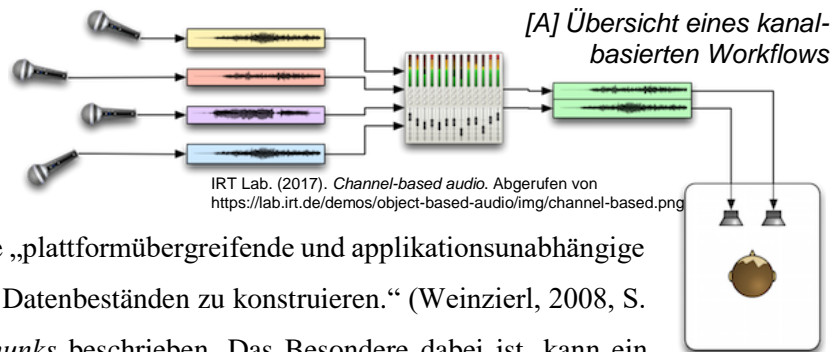
Da die Grenzen zwischen reinen Audioformaten und Containern fließend sind und neuere Formate nur noch aus Containern bestehen, wird weitergehend nicht explizit darunter unterschieden.

²vgl. Dickreiter, 2010, S.658

³vgl. Weinzierl, 2008, S.688-689

2 Kanalbasierte Audioformate

Bei kanalbasierten Audioformaten wird jeder Kanal einem Lautsprecher zugeordnet⁴. Ziel ist die Speicherung und Wiedergabe einer Präsentation (z.B. ein Konzert), beziehungsweise eines Klangphänomens (z.B. Sound-Design). Nicht vorhandene Lautsprecher können mittels Summenlokalisierung als Phantomschallquelle generiert werden, bewegen sich dabei aber immer auf der Ebene der vorhandenen Lautsprecher. Trotz Up- oder Downmixes (z.B. Stereo ↔ Mono) beziehen sich alle Kanäle auf die Speicherung der selben Präsentation⁵



2.1 IFF

Mit dem IFF versuchte EA eine „plattformübergreifende und applikationsunabhängige Vorschrift zur Weitergabe von Datenbeständen zu konstruieren.“ (Weinzierl, 2008, S. 695) Die Daten werden in *Chunks* beschrieben. Das Besondere dabei ist, kann ein Dekoder diese nicht entschlüsseln, werden sie beim Lesen übersprungen, beim Neuschreiben aber blind mitkopiert. So ist dieses Format vielfältig anwendbar und bildet das Grundgerüst für viele weitere Audioformate wie RIFF, WAVE, AIFF aber auch für Steuerungsdaten wie MIDI oder Textformate wie RTF. Apple nutzt bei ihrem AIFF als einziger Unterschied eine lineare PCM.⁶

2.2 RIFF WAVE

Bei dieser Weiterentwicklung von Microsoft und IBM kann zur Mono-Speicherung jedes Kanals auch eine Mehrkanalspeicherung *interleaved* (ineinander verschachtelt) stattfinden. Im Header werden die Chunks beschrieben, welche in 2 oder 4 Byte die Art der Audiokodierung, die Kanalzahl, Abtastrate und die Blockgröße der Audiodaten (Kanalzahl*Bytes/s) enthalten.

WAVE kann bei 44,1 kHz theoretisch bis zu 147 Kanäle speichern und wäre damit 3D-fähig⁷, unterliegt jedoch der 4GB-Grenze. Ein maximaler 4GB-Datenblock errechnet sich aus der *Anzahl der Spuren* multipliziert mit der *Auflösung in Byte*, der *Abtastrate* und der *Laufzeit*. Um die 4GB-Grenze zu Umgehen wurde das RIFF WAVE zu Wave64 weiterentwickelt.

2.3 BWF

Das BWF wurde zur Qualitätssicherung innerhalb der Rundfunkanstalten entwickelt und 1997 von der EBU standardisiert. Es ist ähnlich zu WAVE, wird aber mit einer Abtastrate von 48 kHz und einer

⁴vgl. Böhm, 2017, S.33

⁵vgl. Weinzierl, 2008, S.705

⁶vgl. Bourke, 1996 3. Absatz

⁷vgl. Stahl, 2013, S.72

⁸vgl. EBU – TECH 3285, 2011, S.20

Auflösung von 16 Bit empfohlen⁸. Die Chunks enthalten neben Materialbeschreibung und Referenznummern für Abspiellisten auch Urheberangaben. Die Dokumentation der Nutzungsrechte wurde vorher in dedizierten Datenbanken geführt. 2001 kam mit dem 2. Standard eine *Coding-History* hinzu, um Bearbeitungen zu dokumentieren und gegebenenfalls das Ursprungsmaterial wiederherzustellen.

RF64 entstand aus dem BWF ohne 4GB-Grenze und wurde 2004 standardisiert⁹. Es ermöglicht zwischen Audiostream und Lautsprecherkanal unabhängige Mischungen.

2.5 MPEG

Hier handelt es sich um ein ursprüngliches Streaming-Format aus verschiedenen Ebenen (Layer genannt), die aufsteigend nach Komplexität aufeinander aufbauen. Während Layer 1 kaum genutzt wird, ist der zweite Layer wichtig für DAB und DVB (da mit hoher Komplexität die Delay-Zeiten zunehmen). Der dritte Layer ist unter der Dateierdung mp3 bekannt. Die Besonderheit am MPEG ist die Framebasis¹⁰ (normalerweise wird bei Audio auf Samplebasis gearbeitet). MPEG1 Layer 3 komprimiert verlustbehaftet und verschwindet aufgrund der heutigen Speicherkosten und Bandbreiten immer mehr. Trotzdem bildet er noch die Basis für weitere Formate wie AC3, DTS und ist noch im aktuellsten MPEG-4 Format implementiert¹¹. In ID3-Tags können Angaben zu Genre, Herstellungsjahr, Titelnummer/-name, Albuminfos und noch weitere Angaben gespeichert werden. Durch die erwähnte Frameaufteilung ist MPEG auch als objektorientiertes Audioformat geeignet.¹²

2.6 AC3

Der AC3 beschreibt eigentlich einen Kodieralgorithmus. Auf MPEG1-Basis wurde er jedoch 1994 in den USA als Standard-Format für digitales Fernsehen eingeführt. Im Consumer-Segment ist er unter dem Namen *Dolby Digital* von den *Dolby Laboratories* bekannt und ist DVD-Standard¹³. Er besteht aus einer verlustbehafteten konstanten Kompressionsrate von 348 Kbit/s und kann bis zu 6 diskrete Audiokanäle enthalten¹³. Durch *Adaptive Bit Allocation* und *Perceptual Coding* (Berechnung und Neuzuteilung des Bit-Budgets) werden aus 50 Frequenzbändern, verteilt über einen Frequenzbereich von 0 Hz bis 24 kHz, so viele Bits entfernt um die Datenmenge zu reduzieren, solange trotzdem kein Qualitätsunterschied zum Original hörbar wird. Bei den weiterentwickelten Formaten wurden teilweise die Kanalzahlen erhöht oder durch Matrizierungen ein Upgrade (z.B. 4 Tonkanäle auf 2 Spuren) möglich.¹⁴

⁹vgl. EBU – TECH 3306, 2009, S.8

¹⁰vgl. Weinzierl, 2008, S.701

¹¹vgl. Stahl, 2013, S.76

¹²vgl. Quackenbush, 2005, 4. Absatz

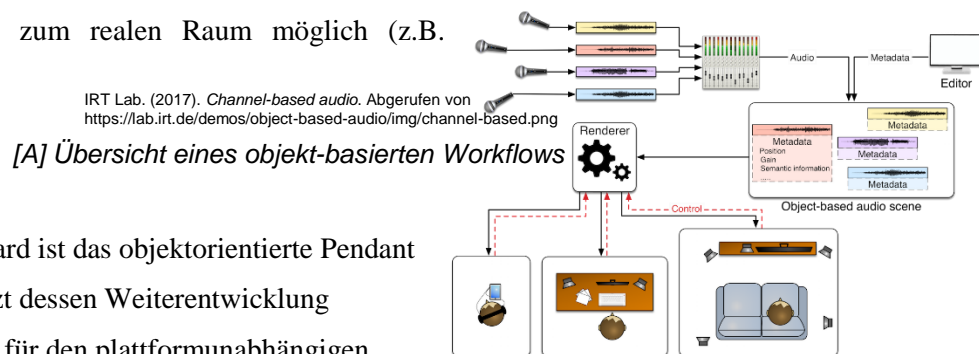
¹³vgl. ATSC A/52:2015, 2015, S.48

¹⁴vgl. Razmowa, 2010, S.4-5

3 Objektbasierte Audioformate

„In der Objektorientierung entsteht ein komplexes System aus mit- und zueinander kooperierender Objekte. Eigenschaften und Funktionen von Objekten werden dabei zusammengefasst, um diese einfacher auf reale Beispiele übertragen.“ (Stahl, 2013, S.40) Dabei können diese unabhängig voneinander manipuliert werden. Statt wie bei der Kanalbasis steht dabei das generative Moment im Vordergrund. Es fordert ein neues Medienverständnis und ordnet den Zusammenhang von Abbildung und Produktion neu. Für die objektorientierte Audioumgebung werden vielfältige Eigenschaften gebraucht, um die Abhängigkeiten, aber auch Unabhängigkeiten der Objekte im komplexen System zu definieren. Zunächst wird in der Produktion unabhängig vom Endmedium produziert (z.B. Positionierung einer Schallquelle im Raum), da moderne Wiedergabemedien nicht immer kanaldefiniert sind. Auf der Wiedergabeseite wandelt ein Dekoder dann die Audioszene auf das jeweilige Medium um (Lautsprecheranlage, Handy, Kopfhörer, usw.) Dadurch wird auch die Wiedergabe in einem virtuellen Raum unabhängig zum realen Raum möglich (z.B.

HRTF).



4.1 AES-31

Der AES-31 Standard ist das objektorientierte Pendant zum BWF und nutzt dessen Weiterentwicklung (BWFF). Es wurde für den plattformunabhängigen

Austausch entwickelt¹⁵. Es enthält in 3 Teilen statische Angaben wie Spurnamen (Part 2), timecode-bezogene Events (Schnitte, Blenden, Steuersignale) und verhält sich ähnlich einer *Audio Decision List* (Part 3). Im ersten Part ist der Header enthalten (macht das Format unabhängig der Plattform lesbar). Es arbeitet samplegenau und ist mehrkanalfähig¹⁶. Während der AES31-1 Standard noch im Low-Level-Diskformat (FAT32) spezifiziert ist wurde die 4GB-Grenze inzwischen durchbrochen. Aktuell wird der AES31-4 Standard entwickelt. Dieser soll noch mehr Metadaten wie beispielsweise Mischautomatiken, Filtereinstellungen und Effektparameter enthalten.¹⁷

4.2 ASF

Schon 1998 führt Microsoft ein objektorientiertes Audioformat ein. Es ist ähnlich wie RIFF aufgebaut und eine WMA-Weiterentwicklung. Da es für Streaming optimiert ist, werden die Datenpakete in Medien-Objekte aufgeteilt, welche aus den Audioobjekten und den Metadaten wie beispielsweise *File*- und *Streamproperties*, einer Inhaltsbeschreibung (für Hörbücher), Marker und dem Timecode bestehen¹⁸. Letztere machen das Fortsetzen der letzten Sitzung und eine Kapitelwahl innerhalb des Files möglich.

¹⁵vgl. Audio Engineering Society, 2013, S.1

¹⁶vgl. Audio Engineering Society, 2013, S.17

¹⁷vgl. Weinzierl, 2008, S.709

¹⁸vgl. Microsoft Corporation, 2004, S.5

4.4 SDIF (Handbuch Audiotechnik/oder Dickreiter)

SDIF nimmt bei den Audioformaten eine Sonderstellung ein. Es dient zwar der Speicherung und dem Austausch von Audiodaten, greift dabei aber auf Repräsentationen zurück. Das bedeutet, es wird nicht das direkte Audiosignal/-objekt übertragen, sondern durch die *schnelle Fourier-Transformation* in *sinusoidale Partialwellen* zerlegt¹⁹. Zusammen mit den Metadaten (u.a. Energiekoeffizient, Hüllkurven, Marker, usw.) in Matrizen werden die Objekte zum Ausgangsmaterial wieder zusammengesetzt. Für die Streaming-Anwendung können noch weitere Informationen hinterlegt werden, welche zur Abwärtskompatibilität „überlesen“ werden können.¹⁹

4.5 MPEG-4

Eine Weiterentwicklung von Quicktime ist der MPEG-4 Standard. Es handelt sich dabei um einen Container, der zwar aus einem objektorientierten Konstrukt besteht, aber kanalbasierte (z.B. alle vorherigen MPEG-Layer) sowie objektorientierte Audioinhalte beinhalten kann. Bei der Objektorientierung werden mittels *Spectral Band Replication* Differenzen zwischen den Audiokanälen analysiert und separat als kompakte Werte gespeichert¹⁹. So werden Metadatenströme zusammen mit dem Audiomaterial zu Szenen zusammengefasst. Durch das *Delivery Multimedia Integration Framework* wird das Format broadcastfähig. Trotz der kompakten Datenübertragung bleiben die Objekte dabei manipulierbar. Theoretisch können bis zu 65.536 Kanäle untergebracht werden, welche aber zugunsten von weiteren Metadaten zweckentfremdet werden. So entsteht bei *MPEG-SAOL* die Szene aus Mono-Stereo-Downmixes zusammen mit Meta-Informationen zur Rückgewinnung (relative Lautstärke zum Objekt, Ähnlichkeit zwischen Objekten, Lautstärkepegel des Objekts und absolutes Energieniveau des lautesten Objekts).²⁰

4.7 AC-4

Im neuen 3D-Audiobereich konnte sich Dolby seit 2012 mit ihrem *Dolby Atmos* System etablieren. Dieser beruht auf dem *Advanced Codec 4*. Es kann kanalbasiert arbeiten (für Downmixes und Surroundanwendungen), ist aber auch objektorientiert nutzbar (für Dolby Atmos). Damit handelt es sich um ein Hybridsystem, welches vom *Europäischen Institut für Telekommunikationsnormen* standardisiert wurde. Aufgrund der Kopplung an 2 MADI-Schnittstellen können derzeit entweder 128 Objekte oder 128 Kanäle verarbeitet werden²¹. Für Die Ausgabe des Formats speziell für eine Lautsprecheranwendung wird eine Rendering Unit zur Umsetzung benötigt.²²

¹⁹vgl. Weinzierl, 2008, S.710

²⁰vgl. Quackenbush, 2005, 4. Absatz

²¹vgl. Dolby Laboratories, 2015, S.1

²²vgl. Dolby Laboratories, 2015, S.4

4 Vor- und Nachteile der Audioformate

Kanalbasiert

Der große Vorteil von kanalbasierten Audioformaten ist die Etablierung. Durch Standardisierungen sind sie bekannt und vereinheitlicht. Dies macht den einfachen Austausch innerhalb Produktions- und Signalketten möglich.

Den vielen Formaten steht aber genau diese Vielfalt im Weg. Der Zusammenhang der Formate ist undurchsichtig und bei speziellen Fällen muss eine genaue und damit zeitaufwendige Einarbeitung in die Formate erfolgen. Denn durch die physischen Begrenzungen (seien es die Bandbreite oder die Kanalzahlen) gibt es keine universellen Lösungen. So erhöht sich beim Umstieg von Mono die Datenmenge und die Kanäle um das Doppelte. In Anbetracht der neuen 3D-Audiosysteme bis zu 128 möglichen Kanälen bei Dolby Atmos wird eine Betrachtung der einzelnen Kanäle nicht möglich (ganz abzusehen von den Datenmengen).

Auch innerhalb der Produktion bleibt das System begrenzt. Eine Surround-Mischung muss in Surround gemischt und abgehört werden, um sie später in Surround wiederzugeben. Dabei wird es beim Umstieg der verschiedenen Systeme (z.B. Upmix von 5.1 auf 7.1) schwierig die Klangqualität aufrecht zu erhalten. Bei Downmixes (Surround in Stereo) muss wiederum 2 Kanäle extra eingelagert werden, was die ohnehin schon großen Datenmenge nochmals vergrößert. Auch kann ein kleines Klangereignis einen kompletten Kanal belegen.

Bei der Endnutzung liegt schließlich nur lineares Material, in welches nicht mehr eingegriffen werden kann, ohne zu einem vorherigen Arbeitspunkt (z.B. Einpflegen in eine Digital Audio Workstation) zurückzukehren.

Objektbasiert

„Das bisher passive Musikhören könnte durch solche Möglichkeiten durchaus zu einer aktiven, fast spielerischen Herangehensweise an Musik bzw. Audioinhalte werden.“ (Stahl, 2013, S. 104)

Gegenüber den kanalbasierten Audioformaten überzeugen objektorientierte durch ihre hohe Kompatibilität. Schon in der Produktion können verschiedene Systeme evaluiert werden, obwohl zum Beispiel nur eine Stereo-Abhöre oder ein Kopfhörer zur Verfügung steht. Durch die Unabhängigkeit vom Endmedium kann eine einfachere Distribution erfolgen. Aufgrund der Auf- und Abwärtskompatibilität (das Rendering geschieht lokal beim Endnutzer) bleibt die Wiedergabesituation anpassungsfähig. Somit können akustisch „schwierige“ Räume besser beschallt werden oder das Hörerlebnis nimmt zu (Immersive Sound). Jedoch wird genau hierfür ein Renderer benötigt, der entweder als Software oder als dediziertes Gerät wie bei Dolby Atmos vorliegen muss. Somit ist in der Kette ein erhöhter Entwicklungsaufwand nötig, der aber in Anbetracht der Beständigkeit vernachlässigt

werden kann (beim Dolby Atmos Renderer können später einfach neue Lautsprecher eingepflegt werden).

Beim Endnutzer wird das Hörerlebnis gesteigert. Er kann aktiv auf die Inhalte zu- und eingreifen und die Mischung individuell auf sich anpassen (z.B. die Sprache hervorheben) oder das System wechseln (vom Handy/Kopfhörer auf eine Surroundanlage). Die Frage dabei ist nur, inwieweit sich der Hörer ein Eingreifen wünscht und man es zulässt. Neben der Frage, ob man in die künstlerische Tätigkeit der Mischung Eingriff gebietet, muss bei den Urheberrechten eine genaue Definition erfolgen, was Bearbeitet werden kann und was nicht. Auch eine Lautheitsbestimmung wird schwierig, wenn einzelne Klangobjekte bearbeitbar bleiben.

Das größte Problem ist aktuell die fehlende Etablierung von Arbeitsumgebungen. Die objektbasierten Formate befinden sich noch in der Entwicklung und müssen bisher in die Workstations als Plugins integriert werden um objektorientierte Formate auszugeben. Und da es kaum Endformate gibt, ist das Bewusstsein beim Konsumenten noch niedrig und somit auch der Investitionswille.

Quellenverzeichnis

ATSC A/52:2015. (2015). *ATSC Standard: Digital Audio Compression (AC-3, E-AC-3)*. Abgerufen

am 15.01.2018 von [http://atsc.org/wp-content/uploads/2016/03/a_52-](http://atsc.org/wp-content/uploads/2016/03/a_52-2015.pdf?zoom_highlight=AC-3)

[2015.pdf?zoom_highlight=AC-3](http://atsc.org/wp-content/uploads/2016/03/a_52-2015.pdf?zoom_highlight=AC-3)

Audio Engineering Society. (2013, January 24). *AES standard on network and file transfer of audio -*

Audio-file transfer and exchange - File format for transferring digital audio data between

systems of different type and manufacture. Abgerufen am 15.01.2018 von

<http://www.aes.org/tmpFiles/aessc/20180216/aes31-2-2012-i.pdf>

Bourke, P. (1996, September). *Creating AIFF Audio Formatted files*. Abgerufen am 15.01.2018 von

<http://paulbourke.net/dataformats/audio/>

Böhm, M. (2017). *Aspekte der objektbasierten Rundfunkproduktion* (Bachelor's thesis,

Hochschule der Medien, Stuttgart, Deutschland). Abgerufen am 26.12.2017 von

<https://filestore.hdm-stuttgart.de/s/wSJajvfrUvvjSYJ>

Dickreiter, M., Dittel, V., Hoeg, W., & Wöhr, M. (2014). *Handbuch der Tonstudioteknik Band 2*

(8th ed.). Berlin, Deutschland: de Gruyter.

Dolby Laboratories. (2015). *Dolby AC-4: Audio Delivery for Next-Generation Entertainment*

Services. Abgerufen am 15.01.2018 von [https://www.dolby.com/us/en/technologies/ac-](https://www.dolby.com/us/en/technologies/ac-4/Next-Generation-Entertainment-Services.pdf)

[4/Next-Generation-Entertainment-Services.pdf](https://www.dolby.com/us/en/technologies/ac-4/Next-Generation-Entertainment-Services.pdf)

EBU - TECH 3285. (2011). Specification of the Broadcast Wave Format (BWF). Abgerufen am 15.01.2018 von <https://tech.ebu.ch/files/live/sites/tech/files/shared/tech/tech3285.pdf>

EBU - TECH 3306. (2009). MBWF/RF64: An extended File Format for Audio. Abgerufen am 15.01.2018 von <https://tech.ebu.ch/files/live/sites/tech/files/shared/tech/tech3306-2009.pdf>

Microsoft Corporation. (2004, December). *Advanced Systems Format (ASF) Specification*. Abgerufen am 15.01.2018 von <http://go.microsoft.com/fwlink/p/?linkid=31334>

Quackenbush, S. (2005, October). MPEG Advanced Audio Coding. Abgerufen am 15.01.2018 von <https://mpeg.chiariglione.org/standards/mpeg-4/audio>

Razmowa, J. (2010). *AC3 Dolby Digital* (Tonseminar, Hochschule der Medien, Stuttgart, Deutschland). Abgerufen am 26.12.2017 von <https://filestore.hdm-stuttgart.de/s/D1E6PYxB90E17IW>

Stahl, S. (2013). *Objektorientierte Audioumgebung für Surround-Sound* (Master's thesis, Hochschule der Medien, Stuttgart, Deutschland). Abgerufen am 26.12.2017 von <https://filestore.hdm-stuttgart.de/s/WXtrh5ADxD2IGfl>

Weinzierl, S. (2008). *Handbuch der Audiotechnik*. doi:10.1007/978-3-540-34301-1