



Bachelorarbeit im Studiengang Audiovisuelle Medien

Lautheitsmessung von objektbasierten Audioszenen

Maike Richter

zur Erlangung des akademischen Grades eines Bachelor of Engineering

an der Hochschule der Medien Stuttgart

18. November 2019

Erstprüfer:

Zweitprüfer:

Praxisbetreuer:

Prof. Oliver Curdt

Michael Weitnauer, M.Eng.

Robert Schwering, B.Eng.

Eidesstattliche Erklärung

Hiermit versichere ich, Maike Richter, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel ***Lautheitsmessung von objektbasierten Audioszenen*** eigenständig und ohne fremde Hilfe verfasst habe. Ich habe keine anderen als die angegebenen Hilfsmittel verwendet und Stellen der Arbeit, die im Sinn oder Wortlaut aus anderen Werken übernommen wurden, sind in jedem Fall unter Angabe der entsprechenden Quelle kenntlich gemacht. Diese Arbeit wurde bislang in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Stuttgart, den 18.11.2019

(Maike Richter)

Abkürzungsverzeichnis

ADM Audio Definition Model

BW64 Broadcast Wave 64 Bit

BWF Broadcast Wave Format

DAW Digital Work Station

dBTP Dezibel True Peak

EAR EBU ADM Renderer

EBU European Broadcasting Union

IIR Infinite Impuls Response

IRT Institut für Rundfunktechnik

ITU International Telecommunication Union

LFE Low Frequency Effects

LKFS Loudness, K-weighted, relative to full scale

LU Loudness Units

LUFS Loudness Units relative to Full Scale

NGA Next Generation Audio

OBA Object-Based Audio

QPPM Quasi Peak Programme Meter

RLB Revised Low-frequency B-curve

VBAP Vector Base Amplitude Panning

XML Extensible Markup Language

Kurzfassung

Diese Arbeit beschäftigt sich mit der Lautheitsmessung objektbasierter Audioszenen in Zusammenhang mit der Entwicklung des EBU ADM Renderers. Ziel der Arbeit ist die Entwicklung und Implementierung eines Verfahrens zur Lautheitsberechnung von objektbasierten Audioszenen. Das Verfahren soll eine möglichst hohe Kompatibilität zur kanalbasierten Lautheitsmessung nach Rec. ITU-R BS.1770-4 aufweisen und unabhängig vom Wiedergabesystem sein. Hierzu wird der EBU ADM Renderer mit seinen Funktionen vorgestellt und die Umsetzung der Objekteigenschaften beim Rendering-Prozess auf einen möglichen Einfluss auf die Lautheit einer Audioszene messtechnisch untersucht. Auf Basis der Untersuchungsergebnisse erfolgt die Entwicklung und anschließende Evaluierung eines objektbasierten Lautheitsmessverfahrens.

Abstract

This thesis deals with the loudness measurement of object-based audio scenes in conjunction with the development of the EBU ADM renderer. The aim is to develop and implement a method for the loudness calculation of object-based audio scenes. The method should yield the highest possible compatibility with channel-based loudness measurement defined by the Recommendation ITU-R BS.1770-4. Furthermore it is supposed to be independent of any used playback system. For this purpose, the core functions of the EBU ADM renderer are presented. A metrological investigation is carried out in order to determine if the object properties have any influence on the loudness of an audioscene. Based on the results of these measurements, a method for object-based loudness measurement is developed. Subsequently the created method is tested and evaluated.

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen zur Lautheit und Lautheitsmessung	3
2.1. Richtlinie der kanalbasierten Lautheitsmessung nach EBU/ITU	5
2.1.1. Richtlinie der EBU/ITU	5
2.1.2. Das kanalbasierte Messverfahren nach Rec. ITU-R BS.1770	6
3. Produktionarten	9
3.1. Kanalbasierte Produktion	9
3.2. Szenenbasierte Produktion	9
3.3. Objektbasierte Produktion	10
4. Stereophone Wiedergabesysteme	11
4.1. Panningverfahren bei stereophonen Wiedergabesystemen	12
4.1.1. Sine/Cosine Pan Law	12
4.1.2. Vector Base Amplitude Panning (VBAP)	13
4.2. Stereophone Lautsprecheranordnungen für Soundsysteme nach Rec. ITU-R BS.2051	15
5. Rendering von objektbasierten Audioszenen	17
5.1. Das Audio Definition Model (ADM)	17
5.1.1. Aufbau	17
5.1.2. Parameter von Audioblöcken bei objektbasierter Audiorepräsentation nach Rec. ITU-R BS.2076-1	19
5.2. Der EBU ADM Renderer (EAR)	20
5.2.1. Polar Point Source Panner	21
5.2.2. Cartesian Point Source Panner	22
5.2.3. Block Processing	23
5.2.4. Interpretation Objektbezogener ADM Parameter	23
6. Objektbasierte Lautheitsmessung	29
6.1. Aktueller Forschungsstand zur Lautheitsmessung von objektbasierten Audioszenen	29
6.2. Amplitudenaddition kohärenter Audioobjekte	30

6.3.	Einfluss der ADM-Parameter auf die Lautheit von Audioszenen	31
6.3.1.	Veranschaulichung der Lautheitswerte mithilfe einer Heatmap	33
6.3.2.	Lautheitsmessung ohne Parameter	34
6.3.3.	Extent	37
6.3.4.	Divergence	38
6.3.5.	Diffuse	39
6.3.6.	Zusammenfassung der Messergebnisse	39
6.4.	Implementierung eines Messverfahrens für objektbasierte Audioszenen	41
6.5.	Evaluierung des entwickelten Verfahrens	42
6.5.1.	Messmaterial	43
6.5.2.	Messtechnische Evaluierung	44
6.5.3.	Plausibilitätsprüfung des entwickelten Verfahrens	48
6.5.4.	Auswertung der messtechnischen und subjektiven Evaluierung	50
7.	Zusammenfassung und Ausblick	53
A.	Auszug aus Recommendation ITU-R BS.2051	57
B.	Messtechnische Überprüfung der ADM-Parameter in Bezug auf die Lautheit einer Audioszene	61
B.1.	Lautheitsabweichungen in Verbindung mit <i>extent</i> und <i>divergence</i>	62
B.2.	Lautheitsabweichungen in Verbindung mit <i>diffuse</i>	64
C.	Messtechnische Evaluation des entwickelten Verfahrens	67
C.1.	Differenzwerte aller Messgruppen	68
C.2.	Differenzwerte der Messgruppe <i>Hörspiele</i>	68
C.3.	Differenzwerte der Messgruppe <i>Musik</i>	69
C.4.	Differenzwerte der selbst erstellten Audioszene in Abhängigkeit der Objektparameter	69
	Literaturverzeichnis	71

Abbildungsverzeichnis

2.1. Die Kurven gleicher Lautstärkepegel[1, S.121]	3
2.2. Die Richtungsabhängigkeit der empfundenen Lautstärke[26, S.63]	4
2.3. Aussteuerung nach Pegel vs. Aussteuerung nach Lautheit[15]	5
2.4. Das kanalbasierte Lautheitsmessverfahren nach Recommendation ITU-R BS.1770[19]	6
2.5. K-Filter Stufe 1: High Shelving Kurve[19]	7
2.6. K-Filter Stufe 2: Revised Low-frequency B-curve[19]	7
4.1. 2D Vector Base Amplitude Panning (VBAP)[12]	14
4.2. 3D Vector Base Amplitude Panning (VBAP)[12]	15
5.1. ADM Metadatenstruktur nach Rec. ITU-R BS.2127-0[22]	18
5.2. Polare Koordinaten[21]	20
5.3. Kartesische Koordinaten[21]	20
5.4. Blockweise Verarbeitung im <i>EAR</i> nach Rec. ITU-R BS.2127[22, S.30]	23
5.5. Polare Gewichtungsfunktion zur Umsetzung des <i>extent</i> -Parameters mit <i>width</i> = 90° und <i>height</i> = 30° nach Recommendation ITU-R BS.2127[22, S.51]	25
5.6. Die Struktur des Objekt-Renderers nach Rec. ITU-R BS.2127[22, S.34]	27
6.1. Objektbasiertes Messverfahren mit Gewichtung nach Objektposition[8, S.20] . .	30
6.2. Stapelverarbeitung zur kanalbasierten Lautheitsmessung mit unterschiedlichen Ob- jektparametern	32
6.3. Heatmap der Lautheitswerte einer 4+9+0-Lautsprecheranordnung bei Lautheits- messung mit positionsabhängiger Gewichtung	34
6.4. Heatmap zur Veranschaulichung der Lautheitswerte eines Audioobjekts bei Rende- ring auf ein 2.0-Lautsprecher-setup mit polaren Positionsangaben	35
6.5. Gewichtete Lautheitsmessung bei den Lautsprecheranordnungen 0+5+0 (l) und 9+10+3 (r)	36
6.6. Struktur des objektbasierten Lautheitsmessverfahrens	41
6.7. Boxplots der Lautheitswerte für alle Audioszenen	44
6.8. Lautheitswerte der Gruppe <i>Hörspiele</i>	45
6.9. Lautheitswerte der Gruppe <i>Musik</i>	46

6.10. Lautheitswerte der selbst erstellten Audioszene in Abhängigkeit unterschiedlicher Objektparameter	47
B.1. Lautheitsabweichungen für ein Rauschobjekt mit den <i>extent</i> -Werten <i>height</i> = 40.0 und <i>width</i> = 40.0 bei Rendering auf 0+5+0	62
B.2. Lautheitsabweichungen für ein Rauschobjekt mit den <i>extent</i> -Werten <i>height</i> = 40.0 und <i>width</i> = 40.0 bei Rendering auf 9+10+3	63
B.3. Lautheitsabweichungen für ein Rauschobjekt mit den <i>divergence</i> -Werten <i>value</i> = 0.5 und <i>azimuthRange</i> = 45.0° bei Rendering auf 4+7+0	63
B.4. Lautheitsabweichungen für ein Rauschobjekt mit <i>diffuse</i> = 0.5 bei Rendering auf 0+5+0	64
B.5. Lautheitsabweichungen für ein Rauschobjekt mit <i>diffuse</i> = 0.5 bei Rendering auf 4+5+1	65
B.6. Lautheitsabweichungen für ein Rauschobjekt mit kartesischen Positionsangaben und <i>diffuse</i> = 0.5 bei Rendering auf 0+5+0	65
B.7. Lautheitsabweichungen für ein Rauschobjekt mit <i>diffuse</i> = 0.5 bei Rendering auf 4+9+0	66
B.8. Lautheitsabweichungen für ein Rauschobjekt mit <i>diffuse</i> = 0.5 bei Rendering auf 9+10+3	66

1. Einleitung

Seit einigen Jahren werden Forschungen im Bereich *Next Generation Audio (NGA)* angestellt. *Next Generation Audio* ist ein Sammelbegriff für alternative Technologien, Formate und Wiedergabesysteme im Audibereich. *NGA* soll unter anderem immersives Audio in hoher Qualität bieten und Interaktion mit dem Hörer ermöglichen. Hierzu werden neue Ansätze zur Produktion und Wiedergabe von Audioinhalten untersucht. Ein wichtiger Teil dieser Forschungen ist die *objektbasierte* Audiorepräsentation. Diese ermöglicht unter anderem die Wiedergabe von nur einer Anfertigung einer Tonmischung auf verschiedenen Lautsprecheranordnungen.

2018 entwickelte die European Broadcasting Union einen offenen Standard zum Rendering von objektbasierten Audioszenen, den sogenannten *EBU ADM Renderer (EAR)*. Ziel war es dabei, einen einheitlichen Renderer für die Produktion, Qualitätssicherung und die Evaluation bereit zu stellen.

Allerdings ergeben sich aus dem objektbasierten Ansatz auch weitere Herausforderungen. Gegenstand dieser Arbeit ist die Frage nach der Lautheitsmessung von objektbasierten Audioszenen. *Lautheit* spielt vor allem in Broadcastingumgebungen eine wichtige Rolle, da wahrnehmbare Sprünge der Lautstärke zwischen zwei Sendeeinheiten zu Zuschauerbeschwerden führen. Richtlinien zur Lautheitsmessung bestehen bislang aber nur für kanalbasiertes Audio.

Das Ziel der vorliegenden Thesis ist daher die Entwicklung eines Lautheitsmessverfahrens für objektbasierte Audioszenen. Das Verfahren soll einen Lautheitswert pro Audioszene generieren, der für alle Wiedergabesysteme gilt, und möglichst kompatibel zum kanalbasierten Lautheitsmessverfahren nach Rec. ITU BS.1770-4 sein.

In diesem Rahmen werden bestehende Ansätze zur objektbasierten Lautheitsmessung aufgezeigt und im Kontext der Entwicklung des *EBU ADM Renderers* weitere Untersuchungen angestellt. Ein Schwerpunkt liegt hier auf der Untersuchung der Objektparameter und ihren eventuellen Einfluss auf die Lautheit einer Audioszene. Auf Basis der Untersuchungsergebnisse soll daraufhin ein Lautheitsmessverfahren für objektbasierte Audioszenen implementiert werden. Zuletzt folgt eine Evaluation des entwickelten Verfahrens, welche messtechnische Untersuchungen und einen informellen, subjektiven Hörtest beinhaltet.

Das im Rahmen dieser Arbeit entwickelte generische Lautheitsmessverfahren soll bei objektbasierten Produktionsketten einsetzbar sein. Beispielsweise könnte das Messverfahren einen Lautheitsrichtwert zur Aussteuerung im Produktionsablauf generieren. Zudem kann der Lautheitswert einer

1. Einleitung

Audioszene in dessen Metadaten gespeichert werden, sodass die Möglichkeit einer Pegelanpassung durch einen Renderer besteht. Damit können Lautheitssprünge zwischen verschiedenen Audioinhalten vermieden werden.

2. Grundlagen zur Lautheit und Lautheitsmessung

Der Begriff *Lautheit* beschreibt die subjektiv empfundene Lautstärke eines Schallereignisses und ist im Gegensatz zum Pegel eines Audiosignals keine objektiv messbare Größe [2, S.46]. Modelle zur Lautheitsberechnung einer vorliegenden Audioszene stellen immer eine Näherung der empfundenen Lautstärke dar. Neben dem Schalldruckpegel wirkt sich auch das Frequenzspektrum eines Audiosignals auf seine Lautheit aus [26, S.60]. Abbildung 2.1 zeigt die im Jahre 1933 von Fletcher und Munson erstellten Kurven gleicher Lautstärke. Sie veranschaulichen, dass das menschliche Ohr Sinustöne verschiedener Frequenzen bei gleichem Pegel als unterschiedlich laut empfinden kann. So wird beispielsweise ein 1 kHz Sinuston mit einem Schalldruckpegel von 70 dB als gleich laut empfunden wie ein 20 Hz Sinuston mit 105 dB Schalldruckpegel. Beide Töne besitzen einen Lautstärkepegel von 70 phon. Der Lautstärkepegel in phon ist eine objektiv messbare Größe, welche sich lediglich auf Pegel und Frequenz bezieht. Sie ist nicht mit der tatsächlich empfundenen Lautstärke gleichzusetzen.

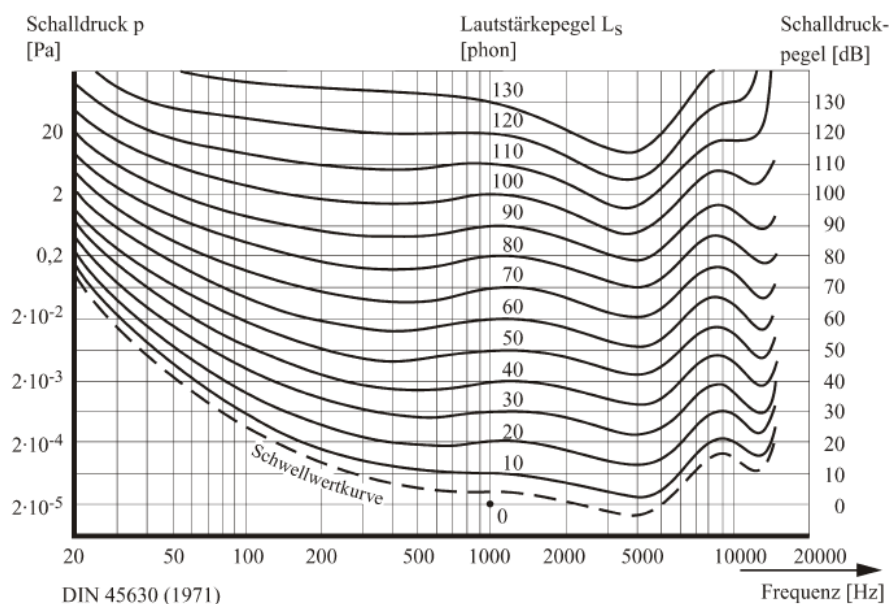


Abbildung 2.1.: Die Kurven gleicher Lautstärkepegel [1, S.121]

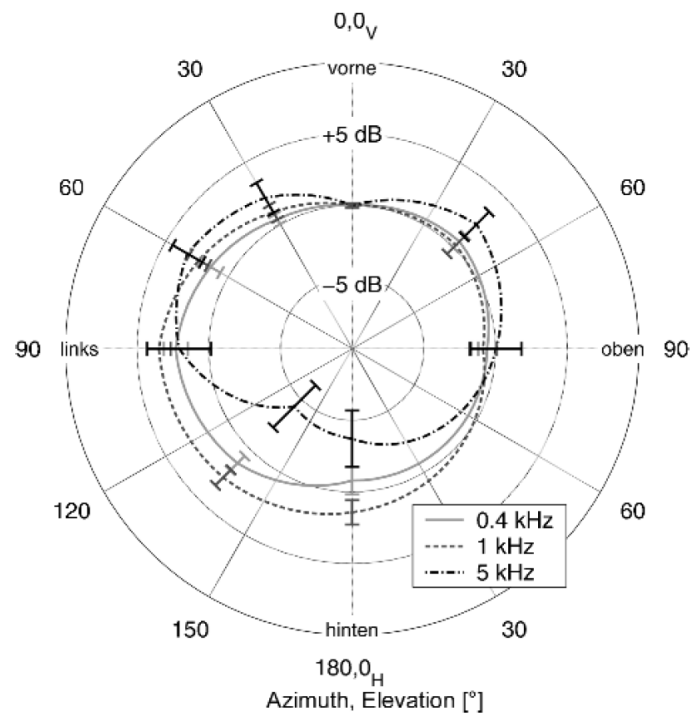


Abbildung 2.2.: Die Richtungsabhängigkeit der empfundenen Lautstärke[26, S.63]

1975 entwickelte Eberhard Zwicker ein Modell zur Lautheitsberechnung, welches auf psychoakustischer Forschung basiert. Nach Zwicker beeinflussen die Rauheit, Schärfe, Signallänge und Frequenzbandbreite eines Audiosignals ihre Lautheit, ebenso wie psychoakustische Effekte, unter anderem Maskierung und Verdeckung [3].

Bei der Wiedergabe auf erweiterten Lautsprecher-Systemen, die Lautsprecherpositionen hinter der hörenden Person beinhalten, spielt zudem die Wirkung des Schalleinfallswinkels eine Rolle. Stefan Weinzierl beschreibt den Zusammenhang zwischen Schalleinfallswinkel und empfundener Lautstärke von Schallsignalen. In Abbildung 2.2 ist zu erkennen, dass ein seitlich auf den Hörer einfallender Sinuston der Frequenz 400 Hz oder 1 kHz als subjektiv lauter wahrgenommen wird als ein frontal eintreffendes Signal. Die Richtungsabhängigkeit höherer Frequenzen weicht von diesem Schema ab. Da jedoch praxisrelevante Audiosignale, wie Sprachsignale, nur einen geringen Frequenzanteil über 4 kHz besitzen[2, S.75] und sich Grundtöne ebenfalls im unteren Frequenzbereich befinden, sind diese Abweichungen für die Entwicklung eines Lautheitsmessverfahrens vernachlässigbar.

2.1. Richtlinie der kanalbasierten Lautheitsmessung nach EBU/ITU

2.1.1. Richtlinie der EBU/ITU

Die International Telecommunication Union (ITU) veröffentlichte 2006 die erste internationale Richtlinie zur Bestimmung der subjektiv empfundenen Lautstärke von Audioprogrammen. Ein *Audioprogramm* wird in diesem Zusammenhang definiert als eine in sich geschlossene Audioproduktion oder audiovisuelle Produktion, welche in Radio, Fernsehen oder anderen elektronischen Medien präsentiert wird[14]. Die Motivation zu dieser einheitlichen Richtlinie besteht darin, Lautheitssprünge zwischen aufeinanderfolgenden Audioprogrammen zu vermeiden.

Abbildung 2.3 zeigt den Vergleich zwischen der bis dahin angewandten Aussteuerung nach Pegelspitzen und der nun empfohlenen Aussteuerung nach Lautheit.

Besonders bei Audioprogrammen mit unterschiedlichen Dynamikumfängen führt die Aussteuerung nach Pegel zu Sprüngen der empfundenen Lautstärke. Dies führte zu Unzufriedenheit der Konsumenten, da diese die Lautstärke manuell anpassen mussten. Ziel der ITU-R ist es, mit den erforschten Algorithmen zur Lautheitsmessung echte Vergleichbarkeit der empfundenen Lautstärke verschiedener Audioproduktionen zu ermöglichen. Im Zusammenhang mit der Richtlinie R128[14] der European Broadcasting Union (EBU)) aus dem Jahr 2011 führten deutsche Fernsehsender im Jahr 2012 die Aussteuerung nach Lautheit ein. Die neu eingeführten Einheiten sind *Loudness Units (LU)* und *Loudness Units relative to Full Scale (LUFS)*¹. *LU* ist eine relative Einheit und beschreibt den Lautheitsunterschied eines Audiosignals in Bezug auf ein anderes Signal oder einen bestimmten Lautheitswert. Eine Lautheitsabsenkung um -3 LU entspricht dabei einer Pegelabsenkung um -3 dB. Werte der Einheit *LUFS* sind absolut und beziehen sich auf die digitale Vollaussteuerung. Die EBU R128 definiert für Rundfunkprogramme einen Zielwert von -23 LUFS, wobei eine Toleranz von 0.5 LU tragbar ist. Bei Live-Übertragungen und Streams dürfen Schwankungen von bis zu 1 LU auftreten. Die EBU definiert weiterhin, dass neben der Programmlautheit auch der Lautheitsumfang

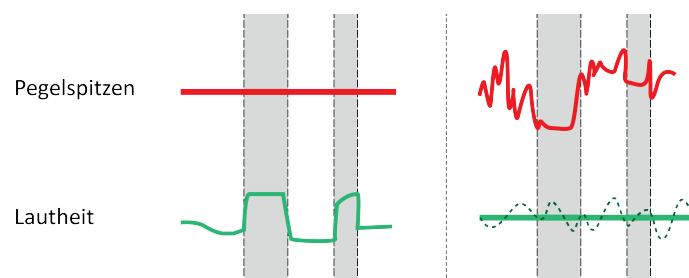


Abbildung 2.3.: Aussteuerung nach Pegel vs. Aussteuerung nach Lautheit[15]

¹Die International Telecommunication Union definiert diese Einheit als *LKFS (Loudness, K-weighted, relative to full scale)*. *LUFS* und *LKFS* sind austauschbare Begriffe. In dieser Arbeit wird die Einheit *LUFS* verwendet.

2. Grundlagen zur Lautheit und Lautheitsmessung

(*Loudness Range*) und der absolute Spitzenwert (*Maximum True Peak Level*) für die Charakterisierung von Audiosignalen angewandt werden soll. Zur Vermeidung von Übersteuerungen sei der absolute Spitzenwert eines Audioprogramms stets geringer als -1 Dezibel True Peak (dBTP). Weiterhin empfiehlt die EBU die Erhebung temporärer Lautheitswerte (*Maximum Momentary Loudness* und *Maximum Short-term Loudness*). Diese geben Auskunft über die aktuelle Lautheit eines Audiosignals und können dadurch die Zielwerterreichung einer Programmlautheit erleichtern[17].

2.1.2. Das kanalbasierte Messverfahren nach Rec. ITU-R BS.1770

Das Lautheitsmessverfahren der ITU[19] ermittelt aus den Audiosignalen der einzelnen Lautsprecherkanäle einen Lautheitswert. Es handelt sich um ein elektroakustisches Messverfahren, welches eine Alternative zum psychoakustisch motivierten Modell nach Zwicker darstellt. Der Algorithmus besteht aus vier Grundkomponenten: *Filterung*, *Mittelwertbildung*, *gewichtete Summation* und *Gating*. Abbildung 2.4 illustriert die einzelnen Elemente des Algorithmus.

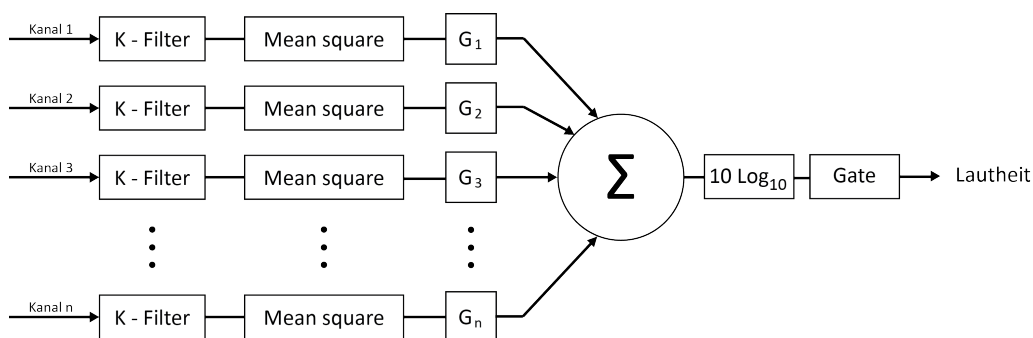


Abbildung 2.4.: Das kanalbasierte Lautheitsmessverfahren nach Recommendation ITU-R BS.1770[19]

Die Signale der einzelnen Kanäle durchlaufen zunächst das Modul *K-Filter*. Dieses entspricht einer zweistufigen Filterung bestehend aus zwei *Infinite Impuls Response (IIR)*-Filtern zweiter Ordnung: einem *Shelving*-Filter und einem *Hochpassfilter*. In Abbildung 2.5 ist die *High Shelving*-Kurve des ersten Filters zu sehen. Diese gleicht den akustischen Einfluss des Kopfes auf Schallsignale aus. Forschungen von Soenke Mehrgardt und Volker Mellert belegen eine Pegelerhöhung von Sinustönen im Frequenzbereich 2–5 kHz, welche auf Abschattungseffekte durch den Kopf des Hörers zurückzuführen sind[9]. Der sogenannte *Revised Low-frequency B-curve (RLB)*-Filter (Abbildung 2.6) dient zum Ausgleich der Frequenzbewertung des menschlichen Hörorgans. Durch den *RLB*-Filter werden tiefe Frequenzen in Bezug auf die Lautheit weniger stark bewertet als hohe Frequenzen, da der Mensch, wie in Kapitel 2 durch die Kurven gleicher Lautstärkepegel veranschaulicht wurde, tiefe Frequenzen als weniger laut empfindet als hohe Frequenzen. Der *RLB*-Filter ist demnach eine stark vereinfachte Umkehrung der Kurven gleicher Lautstärke und beruht auf der Frequenzbewertungskurve B der DIN Norm 60651.

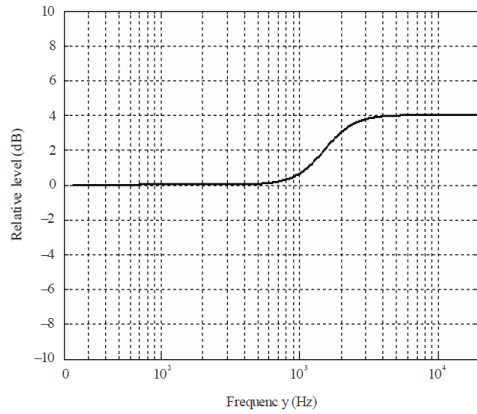


Abbildung 2.5.: K-Filter Stufe 1: High Shelving Kurve[19]

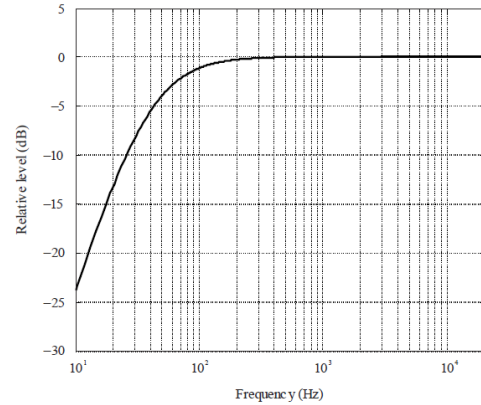


Abbildung 2.6.: K-Filter Stufe 2: Revised Low-frequency B-curve[19]

Im Modul *Mean Square* wird jeweils über ein definiertes Zeitintervall der quadratische Leistungsmittelwert des vorliegenden Signals ermittelt. Auf diese Weise entstehen Audioblöcke mit einer Länge von 400 ms. 100 ms nach dem Beginn eines Audioblocks startet der darauffolgende Audio-block, sodass eine 75-prozentige Überschneidung stattfindet. Formel 2.1 zeigt die mathematische Beschreibung des *Mean Square*-Moduls.

z_{ij} entspricht hierbei der Leistung des j -ten Audioblocks des Signals y_i im Kanal i über das Zeitintervall $T_g = 400$ ms. *Step* entspricht dem Zeitabstand zwischen den Anfängen zweier aufeinanderfolgender Audioblöcke und beträgt 100 ms.

$$z_{ij} = \frac{1}{T_g} \int_{T_g \cdot j \cdot \text{step}}^{T_g \cdot (j \cdot \text{step} + 1)} y_i^2 dt \quad (2.1)$$

In Kapitel 2 wurde aufgezeigt, dass Schallsignale mit Einfallsrichtung von der Seite des Kopfes als lauter empfunden werden als Signale aus anderen Richtungen. Dieser Umstand wird durch das Modul *Gewichtung* in die Lautheitsmessung einbezogen. Es handelt sich hierbei um eine positionsabhängige Gewichtung der Leistungswerte mit dem Gewichtungsfaktor $G_i = 1.41$. Dies entspricht einer Pegeladdition von ungefähr 1.5 dB. Laut Vorschrift der ITU-R betrifft die Gewichtung Lautsprecherkanäle, für deren Position gilt:

Azimuth (θ): $60^\circ < \theta < 120^\circ$ und Elevation (ϕ): $|\phi| < 30^\circ$.

Nach der *Gewichtung* erfolgt die *Summation* aller Kanäle. Anschließend durchläuft das summierte Signal das *Gating*-Modul. Dieses besteht aus einem *absoluten Gate* und einem *relativen Gate* und dient der Berücksichtigung von Stille und leisen Stellen im Signal, welche den Lautheitswert verfälschen können. Die *absolute Threshold* Γ_a filtert alle Signalanteile unter -70 LUFS heraus,

2. Grundlagen zur Lautheit und Lautheitsmessung

sodass lediglich Audioblöcke j übrig bleiben, für die gilt:

$J_g = \{j : l_j > \Gamma_a\}$ (mit J_g = gegatete Audioblöcke, l_j = Lautheit der Audioblöcke, Γ_a = absolute Threshold)

Die Threshold des *relativen Gates* Γ_r berechnet sich aus der Lautheit des aus dem *absoluten Gate* resultierenden Signals. Um die *relative Threshold* zu ermitteln, werden von diesem Lautheitswert 10 LU abgezogen. Die vollständige Gleichung zur Berechnung der Programmlautheit lautet:

$$\text{Gated loudness } L_{KG} = -0.691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{j \in J_g} z_{ij} \right) LKFS \quad (2.2)$$

Mit:

$$J_g = \{j : l_j > \Gamma_r \text{ and } l_j > \Gamma_a\} \quad (2.3)$$

Der Korrekturwert $-0,691$ dient der Herstellung einer Kompatibilität zum ehemaligen Aussteuerungsverfahren mit dem *Quasi Peak Programme Meter (QPPM)*. Ein 1 kHz Sinuston mit einem Pegel von -9 dBFS entspricht durch den Korrekturwert einer Lautheit von -9 LUFS.

3. Produktionarten

Zur Erstellung einer Tonmischung existieren drei grundlegende Ansätze: *kanalbasiert*, *szenenbasiert* und *objektbasiert*. Im Folgenden werden die verschiedenen Ansätze in ihren Grundzügen beschrieben.

3.1. Kanalbasierte Produktion

Die *kanalbasierte* Audiorepräsentation ist bis heute der Standard. Hierbei sind die einzelnen Kanäle einer Multikanal-Audiodatei jeweils einer bestimmten Lautsprecherposition zugeordnet. Eine 2.0-Stereomischung besteht beispielsweise aus einer zweikanaligen Audiodatei, wobei der erste Kanal bei der Wiedergabe auf dem linken und der zweite Kanal auf dem rechten Lautsprecher des genutzten Stereosystems zu hören ist. Die Anfertigung einer *kanalbasierten* Audiomischung erfolgt durch anteilige Verteilung der vorliegenden Audiosignale auf die vorhandenen Kanäle des Zielformats. Außer dem Stereo-Format sind die Surround-Formate 5.1 und 7.1 verbreitet. Sie werden vorwiegend bei Filmmischungen fürs Kino angewandt, es gibt aber auch teilweise Heimkinosysteme, auf welchen Surround-Mischungen abgespielt werden können.

Angefertigte Audiomischungen sind grundlegend auf das jeweilige Zielformat beschränkt. Sollen unterschiedliche Formate unterstützt werden, so muss für jedes Format eine eigene Mischung angelegt werden. Algorithmen für *Up-* bzw. *Downmixing* können bei der Übertragung einer Mischung von einem Format in ein anderes helfen. Sie stellen jedoch meist einen Kompromiss dar, bei dem Abstriche in einer für das Zielsystem optimalen Abbildung gemacht werden müssen.

3.2. Szenenbasierte Produktion

Eine *szenenbasierte* Audiorepräsentation beruht auf dem sogenannten *Ambisonics*-Verfahren, das die Aufnahme und Wiedergabe ganzer Schallfelder ermöglicht.

Ziel ist es, die Eigenschaften des Schallfeldes in einem Raum zu erfassen und diese über ein Lautsprechersystem realistisch in einen anderen Raum zu übertragen. Die Aufnahme der Signalkomponenten geschieht mithilfe spezieller Mikrofonarrays, welche die reine Schalldruckkomponente, sowie die Druckgradienten der Raumachsen erfassen. Auf Wiedergabeseite erhält ein entsprechender Renderer

die Audiosignale sowie die Signalkoeffizienten als Metadaten und überträgt sie auf das vorhandene Wiedergabesystem[23, S.3]. Auf diese Weise kann die ursprüngliche Audioszene nachgebildet werden.

Eine Alternative zur Herstellung einer *szenenbasierten* Audioproduktionen durch Live-Capturing sind künstlich erstellte Szenen durch Platzierung einzelner Audioobjekte im Raum.

In jedem Fall sind die Audioszenen unabhängig vom spezifischen Wiedergabesystem produzierbar[11].

3.3. Objektbasierte Produktion

Der *objektbasierte* Ansatz zur Erstellung einer Tonmischung beruht auf der Darstellung der einzelnen Elemente der jeweiligen Mischung als sogenannte Audioobjekte. Diese bestehen aus einem Audiosignal und Metadaten mit Informationen zu den Eigenschaften des Objekts, wie beispielsweise seine Position, Signaldauer, Höhe und Breite oder sein Pegel. Eine sogenannte OBA-Szene beinhaltet alle Audioobjekte der Mischung und kann entweder über einen Audiostream und einem zugehörigen Metadatenstream übertragen werden oder in einer Audiodatei mit eingebundenen Metadaten gespeichert werden. Für die Wiedergabe wird ein Renderer benötigt, welcher die Audioobjekte unter Berücksichtigung ihrer Eigenschaften auf die vorliegenden Lautsprecherkanäle überträgt.

Eine *objektbasiert* generierte Tonmischung ist somit auf allen gewünschten Lautsprecheranordnungen abspielbar. Im Gegensatz zum *kanalbasierten* Workflow muss nicht für jedes Wiedergabeformat eine separate Mischung angelegt werden.

Weiterhin entstehen durch die Flexibilität der objektbasierten Audiostruktur neue Möglichkeiten für den Nutzer durch Interaktion mit dem Material. Der Nutzer kann die Mischung individuell auf seine Bedürfnisse anpassen. Hörgeschädigte können z.B. die Sprachverständlichkeit verbessern, indem sie die Hintergrundgeräusche und -musik absenken. Zudem können beispielsweise bei Sportübertragungen Kommentare in verschiedenen Sprachen auswählbar sein[18].

4. Stereofone Wiedergabesysteme

Stereofone Wiedergabeverfahren beruhen auf dem psychoakustischen Phänomen der Summenlokalisation. Wird dasselbe Audiosignal mit gleicher Leistung von beiden Lautsprechern eines Stereo-Systems emittiert, so lokalisiert der Hörer die Schallquelle in der Mitte zwischen den Lautsprechern[3, S.310]. Durch Pegel- oder Laufzeitunterschiede der Signale kann jede beliebige Position zwischen den Lautsprechern dargestellt werden. Das 2.0 Stereo-Format besteht aus zwei Lautsprechern L und R , welche zu jeweils 30° links und rechts vom Zentrum ausgelenkt sind. Das in erster Linie fürs Kino vorgesehene 5.1 Surround-Format erweitert dieses um einen Mittenlautsprecher C , sowie Seitenlautsprecher Ls und Rs zwischen 100° und 120° auf der Azimuth-Ebene und einen Low Frequency Effects (LFE)-Kanal[25, S.4]. Die Surround-Systeme 5.1 und 7.1 können lediglich Positionen auf einer Ebene darstellen. Umfangreichere Formate wie z.B. das 2006 eingeführte Auro 9.1-Format sehen Lautsprecher in mehreren Elevationsebenen vor und sind daher in der Lage, Phantomschallquellen in der horizontalen, sowie in der vertikalen Ebene zu erzeugen. Bei stereofonen Wiedergabesystemen gilt in Bezug auf die Lokalisierbarkeit von Phantomschallquellen generell, dass eine größere Anzahl an Lautsprechern eine bessere räumliche Abbildung einer Audioszene ermöglicht als beschränkte Lautsprechersysteme[1, S.311].

Abseits der stereofonen Wiedergabesysteme existieren die Binauraltechnik und die Schallfeldsynthese.

Bei der binauralen Wiedergabe wird eine sogenannte *kopfbezogene Übertragungsfunktion* angewandt, welche die Situation des natürlichen Hörens simuliert[1, S.348].

Die Schallfeldsynthese strebt dagegen die physikalische Synthese von realen und virtuellen Schallfeldern an[26, S.657], wodurch Schallquellen im Raum auch an Randpositionen des Hörers für diesen lokalisierbar sind.

Bis heute konnten sich weder binaurale Wiedergabe noch Schallfeldsynthese kommerziell behaupten. Nach wie vor sind stereofone Wiedergabesysteme der Standard.

4.1. Panningverfahren bei stereofonen Wiedergabesystemen

Das *Panning* bezeichnet in der Audioproduktion die Platzierung von Audioobjekten oder -kanälen auf möglichen Positionen des Wiedergabeformats.

Am weitesten verbreitet ist das *Amplitude Panning*, welches Phantomschallquellen zwischen Lautsprechern durch Pegelunterschiede erzeugt[4]. Im zweidimensionalen Fall emittieren zwei Lautsprecher kohärente Audiosignale mit gegebenenfalls unterschiedlichen Amplituden. Der Hörer empfindet dabei die Illusion einer einzigen Schallquelle. Die Position, auf welcher die Phantomschallquelle lokalisiert wird, hängt von der Pegeldifferenz zwischen den Signalen und den Positionen der Lautsprecher und des Hörers ab[12].

Die gewünschte Position einer Phantomschallquelle kann über sogenannte *Panner* eingestellt werden. Ein *Panner* kalkuliert Verstärkungsfaktoren, über die das Audiosignal zu unterschiedlichen Anteilen auf die Lautsprecherkanäle verteilt werden kann.

Für jene Kalkulation existieren unterschiedliche Ansätze. Im Folgenden werden *Panningverfahren* mithilfe trigonometrischer Funktionen und solche basierend auf Vektoren näher beschrieben.

4.1.1. Sine/Cosine Pan Law

Die *Sine/Cosine Pan Law* berechnet die Verstärkungsfaktoren eines Lautsprecher-Paares zur Erstellung einer Phantomschallquelle mithilfe der trigonometrischen Funktionen *Sinus* und *Kosinus* und findet Anwendung in einem Großteil der Digital Work Stations (DAWs)[5].

Der Algorithmus arbeitet leistungserhaltend, sodass für Verstärkungsfaktoren g_1 und g_2 gilt:

$$g_1^2 + g_2^2 = 1 \quad (4.1)$$

Für die Darstellung einer Schallquelle außerhalb der Mittenposition wird der Signalpegel eines Lautsprechers angehoben, während das Signal des zweiten Lautsprechers analog dazu abgesenkt wird. Das Verhältnis der Verstärkungsfaktoren entspricht der Beziehung zwischen Sinus - und Kosinusfunktion[6, S.113].

$$\cos^2(x) + \sin^2(x) = 1 \quad (4.2)$$

Die Gewichtungsfaktoren der Lautsprecher lassen sich somit aus dem Kosinus- bzw. Sinus-Wert eines normierten Positionswertes bestimmen[5]. Für die Position in der Mitte zwischen den Lautsprechern gilt: $g_1 = g_2$.

Daraus resultierend berechnen sich die Verstärkungsfaktoren g_1 und g_2 wie folgt:

$$g_1 = \cos\left(\frac{\pi \cdot (p + 1)}{4}\right) \quad (4.3)$$

$$g_2 = \sin\left(\frac{\pi \cdot (p + 1)}{4}\right) \quad (4.4)$$

Hierbei entspricht p einem Positionswert zwischen -1 (ganz links) und 1 (ganz rechts).

Die Anwendung der *Sine/Cosine Pan Law* im dreidimensionalen Raum ist ebenfalls möglich. Hierzu erfolgt die Berechnung der Verstärkungsfaktoren jeweils für die Koordinatenachsen X, Y und Z. Für jede der drei Positionskoordinaten werden Ebenen, Reihen oder ein Paar an Lautsprechern ermittelt, die die gesuchte Position abbilden können. Daraufhin werden die Verstärkungsfaktoren der gewählten Lautsprecher berechnet.

Um die Signalleistung des Eingangssignals konstant zu halten, erfolgt abschließend eine Normalisation der kalkulierten Verstärkungsfaktoren.

4.1.2. Vector Base Amplitude Panning (VBAP)

Das 1997 von Ville Topias Pulkki eingeführte *VBAP*-Verfahren[12] erzeugt die Verstärkungsfaktoren der Audiosignale mithilfe der vektoriellen Beziehungen zwischen Lautsprecherpositionen und der gewünschten Position der virtuellen Schallquelle.

Das *VBAP*-Verfahren ist eine vektorielle Formulierung der *tangent law* (Gleichung 4.5) und kann die Position der Phantomschallquelle (ϕ) anhand des Lautsprecherbasiswinkels (ϕ_0) und der Verstärkungsfaktoren g_1 und g_2 ermitteln.

$$\frac{\tan \phi}{\tan \phi_0} = \frac{g_1 - g_2}{g_1 + g_2} \quad (4.5)$$

Abbildung 4.1 veranschaulicht die vektoriellen Beziehungen bei zweidimensionalem *VBAP*. In diesem Beispiel wird eine zweikanalige Lautsprecherkonfiguration als zweidimensionale Vektorbasis behandelt. Die Einheitsvektoren $\vec{l}_1 = [l_{11} \ l_{12}]^T$ und $\vec{l}_2 = [l_{21} \ l_{22}]^T$ zeigen in Richtung der Lautsprecher. Der Einheitsvektor $\vec{p} = [p_1 \ p_2]^T$, welcher in Richtung der virtuellen Schallquelle zeigt, kann durch eine lineare Kombination von $\vec{l}_1$ und $\vec{l}_2$ dargestellt werden. g_1 und g_2 sind dabei die richtungsbestimmenden Gewichtungsfaktoren.

$$\vec{p} = g_1 \vec{l}_1 + g_2 \vec{l}_2 \quad (4.6)$$

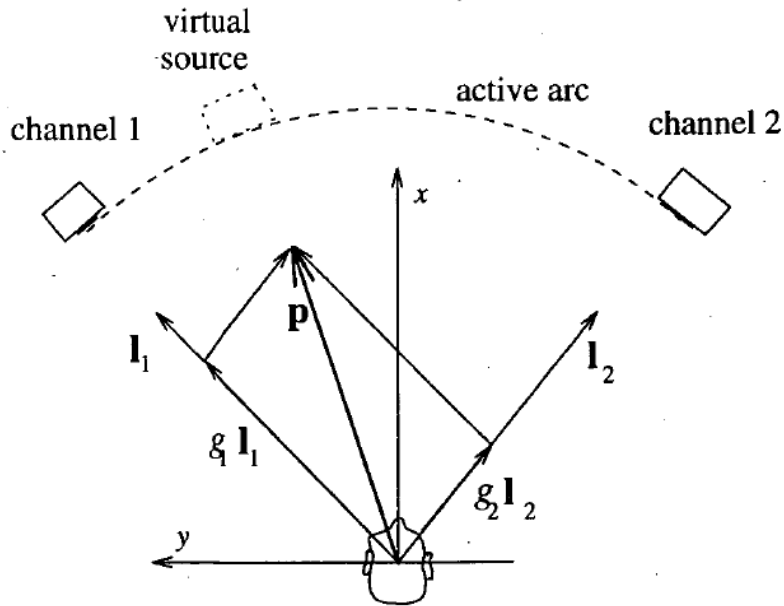


Abbildung 4.1.: 2D Vector Base Amplitude Panning (VBAP)[12]

Als Matrixgleichung mit $g = [g_1 \ g_2]$ und $L_{12} = [l_1 \ l_2]^T$ lautet die Formel wie folgt:

$$p^T = g \cdot L_{12} \quad (4.7)$$

Mithilfe einer invertierten Matrix kann die Gleichung nach g aufgelöst werden:

$$g = p^T \cdot L_{12}^{-1} \quad (4.8)$$

Um die Signalenergie beim Panning zu erhalten, ist eine anschließende Normalisierung der Gewichtungsfaktoren notwendig. Gleichung 4.9 zeigt die zu erfüllende Bedingung. Der Leistungswert C ist konstant zu halten.

$$g_1^2 + g_2^2 = C \quad (4.9)$$

Für das Panning auf dreidimensionale Wiedergabesysteme werden Signalpegel für Lautsprechertripel berechnet[26, S.657].

Die drei Lautsprecher befinden sich auf einer Sphäre und haben somit den gleichen Abstand zur Hörerposition. Die Position der virtuellen Schallquelle wird nun durch den dreidimensionalen Vektor $\vec{p} = [p_1 \ p_2 \ p_3]$ bestimmt. Dieser wird aus einer linearen Kombination der Lautsprechervektoren l_1 , l_2 und l_3 gebildet.

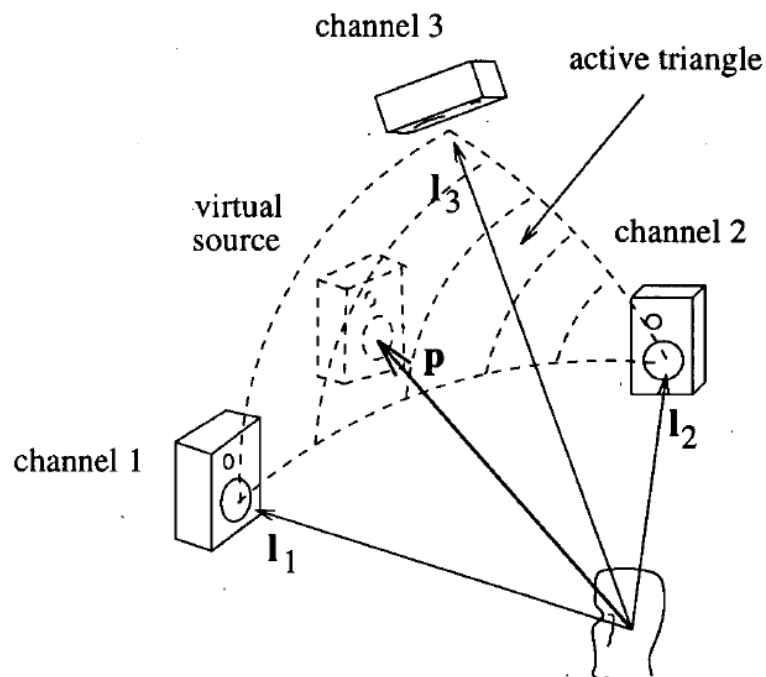


Abbildung 4.2.: 3D Vector Base Amplitude Panning (VBAP)[12]

Um die Gesamtleistung konstant zu halten, gilt für Gewichtungsfaktoren und Leistungswert analog zur Anwendung im zweidimensionalen Raum:

$$g_1^2 + g_2^2 + g_3^2 = C \quad (4.10)$$

Abbildung 4.2 veranschaulicht die vektoriellen Beziehungen bei VBAP im dreidimensionalen Raum.

4.2. Stereofone Lautsprecheranordnungen für Soundsysteme nach Rec. ITU-R BS.2051

In der Empfehlung ITU-R BS.2051[23] definiert die International Telecommunication Union zehn Lautsprecheranordnungen zur Wiedergabe von kanalbasierten, objektbasierten und szenenbasierten Audioszenen. Sie ist eine Weiterführung der Empfehlung ITU-R BS.775[25], welche lediglich Mono-, Stereo-, und Surround-Systeme mit Lautsprechern auf einer Horizontalebene beinhaltet. Die Lautsprecher der in Rec. ITU-R BS.2051 beschriebenen Systeme befinden sich auf einer Kugeloberfläche mit jeweils gleichem Abstand zur Hörerposition und können drei Ebenen zugeordnet werden: *Upper Layer*, *Middle Layer* und *Bottom Layer*.

4. Stereophone Wiedergabesysteme

Die *Middle Layer* befindet sich auf der Horizontalebene, auf Höhe des Hörerkopfes. Die *Bottom Layer* liegt 30° unterhalb der *Middle Layer* und die *Upper Layer* liegt 30° in der Höhe. Letztere enthält außerdem den *Top*-Lautsprecher, welcher sich direkt über dem Hörer befindet und auch als *Voice of God* bezeichnet wird.

Die Bezeichnung einer Lautsprecheranordnung besteht aus drei Zahlen und gibt die Anzahl der Lautsprecher auf den drei Ebenen an. Beispielsweise hat die Auro 3D Aufstellung 9.1 die Bezeichnung $4+5+0$, da sie aus vier Höhenlautsprechern und fünf Lautsprechern in der mittleren Horizontalebene besteht. Die Bezeichnungen der Lautsprecherpositionen enthalten die Ebene, auf der sie sich befinden, und den Horizontalwinkel. Der Lautsprecher $M+030$ befindet sich beispielsweise auf der mittleren Ebene und liegt, gesehen aus der Blickrichtung des Hörers, auf 30° . Die genauen Horizontalwinkel und Toleranzbereiche der einzelnen Lautsprecher können den entsprechenden Tabellen der Rec. ITU-R BS.2051 entnommen werden.

Anhang A enthält einen Auszug der ITU-R Empfehlung, der einen Überblick über alle unterstützten Lautsprecheranordnungen gibt.

5. Rendering von objektbasierten Audioszenen

5.1. Das Audio Definition Model (ADM)

Das *Audio Definition Model* ist ein in der Rec. ITU-R BS.2076-1 definiertes Metadaten-Modell, welches zur Beschreibung von objektbasierten, szenenbasierten und kanalbasierten Audiomischungen genutzt werden kann[18]. Es enthält die den Audioobjekten zugehörigen Metadaten, welche sowohl technische als auch inhaltliche Aspekte der Audioszene beinhalten. Für die Wiedergabe interpretiert ein Renderer die *ADM*-Audioszene und produziert die Ausgangssignale für ein Wiedergabesystem, sodass eine kanalbasierte Repräsentation der Szene erstellt werden kann. Bei Übertragung von objektbasierten Audioszenen definiert das *ADM*, auf welchen Positionen sich die Audioobjekte befinden und welche weiteren Eigenschaften ihnen zugeordnet sind. Die Umsetzung dieser Informationen auf ein bestimmtes Wiedergabeformat ist Aufgabe des Renderers.

Das *ADM* kann in wave-file basierte Formate wie *Broadcast Wave 64 Bit (BW64)* (nach Rec. ITU-R BS.2088[24]) und *Broadcast Wave Format (BWF)* (nach Rec. ITU-R BS.1352[20]) eingebettet werden. Außerdem lässt sich das *ADM* auch in andere Formate, welche zusätzliche *chunks* zur Einbindung von Metadaten unterstützen[21], integrieren.

Zudem ist das Metadaten-Modell als offener Standard deklariert, sodass ein *Lock-in-Effekt* vermieden werden kann.

Aufgrund seiner recht hohen Komplexität und der Tatsache, dass es sich um einen offenen Standard handelt, lässt sich das *ADM* auch als Austauschformat zwischen unterschiedlichen *NGA*-Systemen (z.B. Dolby AC4 bzw. Dolby Atmos oder MPEG-H) nutzen.

5.1.1. Aufbau

Das Modell nutzt grundlegend *Extensible Markup Language (XML)* als Auszeichnungssprache, wobei bei Bedarf die Möglichkeit der Übertragung in anderen Formaten, z.B. JSON (Java Script Object Notation), besteht.

Bei Anwendung in Verbindung mit *BW64*-Dateien kann die *XML*-Datei im *<axml>*-Tag der Datei eingebettet werden.

Die Parameter eines *ADM*-Elements werden durch Attribute und Sub-Elemente definiert.

5. Rendering von objektbasierten Audioszenen

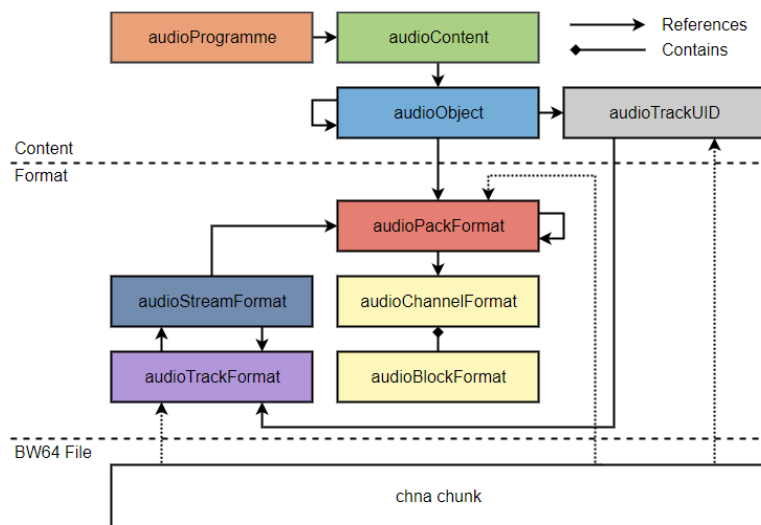


Abbildung 5.1.: ADM Metadatenstruktur nach Rec. ITU-R BS.2127-0[22]

Das *ADM* ist in zwei Sektionen strukturiert: *content part* und *format part*[21]. Abbildung 5.1 veranschaulicht diese Unterteilung.

Der *content part* beschreibt inhaltliche Aspekte des zugehörigen Audiosignals.

Beispielsweise wird spezifiziert, ob Dialogsignale enthalten sind und in welcher Sprache diese vorhanden sind. Weiterhin können hier Lautheitsinformationen eingetragen werden. Referenzen zwischen den Elementen eines *ADMs* geschehen mithilfe von *IDs*.

Das Element *audioContent* bezieht sich auf eine Komponente eines Programms, z.B. die Hintergrundmusik, und verweist auf die zugehörigen Audioobjekte (*audioObjects*). *audioProgramme* referenziert *audioContent*-Elemente, welche zusammen ein Programm bilden[21].

Der *format part* beschreibt technische Informationen zu den Audiodaten.

Das *audioBlockFormat* enthält beispielsweise die Art der vorliegenden Audiorepräsentation als *typeDefinition*.

Weiterhin enthält der *format part* Informationen zu den Kanälen, und wie diese gruppiert sind, z.B. welche Kanäle ein Stereo-Paar bilden. *audioTrackFormat* enthält Format-Metadaten, die über die *audioTrackFormatID* den entsprechenden Audiospuren zugeordnet werden können. *audioStreamFormat* verweist auf *audioTrackFormat*-Elemente und wird zur Identifizierung einer Kombination an Audiospuren verwendet. *audioPackFormat* umfasst ein oder mehrere *audioChannelFormat*-Elemente und *audioChannelFormat* repräsentiert eine Sequenz an Audiosamples. Diese ist zeitlich unterteilt in Audioblöcke, die durch das *audioBlockFormat* beschrieben sind und die für ein bestimmtes Zeitintervall konstante Parameter festlegen. Durch die zeitliche Unterteilung einer Audiospur in Blocks, können die Eigenschaften eines Audioobjekts zeitlich geändert werden.

Die *audioTrackUIDs* sind zur Identifizierung der Audiospuren vorgesehen und sind individuell für jedes Audioobjekt. Aufgrund dieser Art der individuellen Zuordnung können die Spuren einer Datei in beliebiger Reihenfolge vorhanden sein.

Der *Chna Chunk* der *BW64*-Datei dient der Verbindung von *ADM*-Metadaten und Audiokanälen anhand einer Lookup-Tabelle. So kann mithilfe des *Chna Chunks* unter Angabe der *audioTrackUID* eines Objekts der zugehörige Audiokanal identifiziert werden.

5.1.2. Parameter von Audioblöcken bei objektbasierter Audiorepräsentation nach Rec. ITU-R BS.2076-1

audioChannelFormat besitzt unter anderem das Attribut *typeDefinition*, welches die Art der Audiorepräsentation beschreibt. Je nach vorliegendem Audiotypen können für die Audioblöcke unterschiedliche Sub-Elemente und Attribute angegeben werden.

Zu den wichtigsten Parametern, welche die Eigenschaften eines Objekts beschreiben, gehören unter anderem die Position, die räumliche Ausdehnung (*extent*), die Divergenz (*objectDivergence*), die Diffusion (*diffuse*) und der Pegel (*Gain*).

Die Position eines Objektes kann in polarer oder kartesischer Form angegeben werden.

Primär ist laut der ITU-R das polare Koordinatensystem vorgesehen[21, S.17]. Dabei wird eine Position durch Azimuthwinkel, Elevationswinkel und Distanz definiert. Der Azimuthwinkel θ entspricht der horizontalen Auslenkung einer Schallquelle. Es gilt: $-180^\circ \leq \theta \leq 180^\circ$, wobei 0° der Position in Blickrichtung des Hörers entspricht und sich positive Winkel auf der linken und negative auf der rechten Seite befinden. Der Elevationswinkel ϕ bestimmt die vertikale Position und entspricht einem Wert zwischen -90° und 90° .

Die Distanz bestimmt den Abstand der Schallquelle zum Ursprung. Als Distanzwert kann eine Fließkommazahl zwischen 0 und 1 gewählt werden. Ist kein Wert gesetzt, wird eine Distanz von 1.0 angenommen. Alle im polaren Koordinatensystem möglichen Punkte bilden eine Kugel, in deren Mittelpunkt die Position des Hörers vorgesehen ist.

Weiterhin sind Positionsangaben in kartesischen Koordinaten mit den Achsen X, Y und Z möglich. Hierzu muss in den Metadaten das Attribut *cartesian flag* gesetzt sein. X, Y und Z enthalten normierte Werte und sind als Fließkommazahlen zwischen -1 und 1 anzugeben. X-Werte bestimmen die seitliche Position, positive Werte sind rechts von der Hörerposition zu finden, negative Werte befinden sich links. Der Y-Wert zeigt an, ob das Objekt sich vor oder hinter dem Hörer befindet und Z beschreibt die Höhenposition der Schallquelle.

Die Raumform des kartesischen Koordinatensystem entspricht der Form eines Würfels.

Die Abbildungen 5.2 und 5.3 veranschaulichen die Koordinatensysteme, wie sie für das *Audio Definition Model* definiert sind.

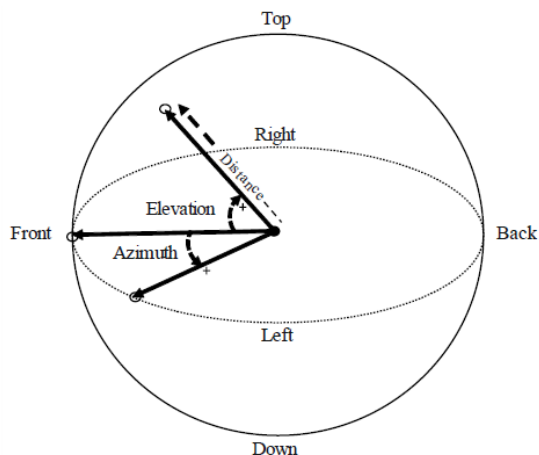


Abbildung 5.2.: Polare Koordinaten[21]

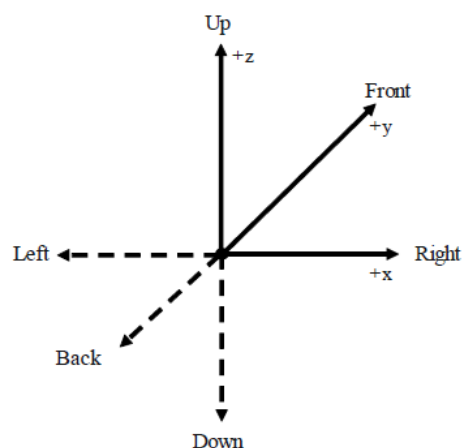


Abbildung 5.3.: Kartesische Koordinaten[21]

Um Bewegungen eines Audioobjekts darzustellen, kann dieses zeitlich in Audioblöcke mit unterschiedlichen Positionsdaten unterteilt werden. Durch Interpolation wird die Illusion einer fließenden Bewegung kreiert. Mit dem *ADM*-Parameter *jumpPosition* wird angegeben, ob der Übergang von einer Position zur nächsten sprunghaft oder fließend geschehen soll. Für letztere Option besteht die Möglichkeit, eine individuelle Interpolationszeit anzugeben. Diese spezifiziert, wie lange die Interpolation zwischen zwei Positionen andauert.

5.2. Der EBU ADM Renderer (EAR)

Zur Wiedergabe von objektbasierten Audioszenen wird ein Renderer benötigt. Dieser liest die Informationen zu den Audioobjekten ein und berechnet aus den einzelnen Objektsignalen und unter Berücksichtigung der in den Metadaten spezifizierten Eigenschaften den Pegelverlauf jedes Lautsprechers des gegebenen Wiedergabesystem.

Der *EBU ADM Renderer (EAR)* wurde zur Erstellung eines einheitlichen Standards zum Rendering von Audioszenen entwickelt. Ursprünglich wurde der *EAR* von der European Broadcasting Union in EBU TECH 3388[16] vorgestellt. Später folgte dann die Veröffentlichung der ITU-R Recommendation BS.2127-0 zur Spezifikation des Renderers. Zusätzlich wurde eine Implementierung des *EAR* in Python erstellt, die es ermöglicht, *ADM*-Szenen dateibasiert zu rendern.

Der Renderer enthält Implementierungen zur Umsetzung von kanalbasierten, objektbasierten, szenenbasierten und matrizierten Audiodaten sowie Mischformen dieser. Zum Zeitpunkt der Verfassung dieser Arbeit sind Implementierungen zum Rendern von binauralem Audio vorgesehen, sie liegen jedoch noch nicht vor.

Als Wiedergabesystem unterstützt der *EAR* die in der Rec. ITU-R BS.2051[23] definierten Lautsprecheranordnungen. Die ITU-R empfiehlt für jeden Lautsprecher eines Systems eine Position

und zum Teil Toleranzbereiche für die Positionen. Nach Wahl der vorliegenden Anordnung kann der Anwender die realen Positionen der Lautsprecher angeben. Der *EAR* überprüft, ob sich die Lautsprecher im Toleranzbereich befinden, sodass die Voraussetzungen für eine akkurate Abbildung der Audioszenen erfüllt sind. Ist dies nicht der Fall, wird dem Anwender eine Fehlermeldung angezeigt[22, S.7].

Sowohl polare als auch kartesische Positionsangaben werden vom *EAR* unterstützt. Die Signale durchlaufen je nach Koordinatensystem zu Teilen unterschiedliche Render-Module, um den jeweiligen Herausforderungen der einzelnen Systeme gerecht zu werden. Da die ITU-R das polare Koordinatensystem als primäres System definiert[21, S.17], liegt der Fokus dieser Arbeit auf der Arbeitsweise des *EAR* bei Angabe polarer Objektpositionen.

Im Folgenden werden die bezüglich dieser Arbeit relevanten Teile des Renderingprozesses in ihren Grundzügen erläutert.

5.2.1. Polar Point Source Panner

Der *Polar Point Source Panner* ist Teil des *EAR* und berechnet mit Informationen zur vorhandenen Lautsprecheranordnung die Signalpegel aller Lautsprecher zur Darstellung eines Audioobjekts als Punktquelle, sodass der Hörer dieses auf der gewünschten Position im Raum ortet. Außerdem ist der *Polar Point Source Panner* unter anderem Teil des Renderingprozesses zur Umsetzung des *extent*-Parameters. Um eine akkurate Positionierung von Audioobjekten zu ermöglichen, muss das Lautsprecher-Setup die Vorgaben nach Rec. ITU-R BS.2051[23] erfüllen, sodass sich alle Lautsprecher auf einer Kugel befinden.

Die Algorithmen des *Polar Point Source Panners* basieren auf dem in Kapitel Abschnitt 4.1.2 vorgestellten *Vector Base Amplitude Panning (VBAP)*. Diese wurden erweitert und zur Nutzung in Broadcastumgebungen angepasst[22, S.22].

Um alle Positionen abbilden zu können, sind im *Polar Point Source Panner* sogenannte *RegionHandler*-Objekte implementiert, welche einen bestimmten Teilbereich des 3D-Raums durch Dreieck - oder Viereckbildung einzelner realer und gegebenenfalls auch virtueller Lautsprecher darstellen können. Die Triangulation von Lautsprechern einer gegebenen Lautsprecheranordnung erfolgt zunächst mit den nominalen Positionen und wird daraufhin auf die real vorliegenden Lautsprecher angepasst. Weiterhin können *RegionHandler*-Objekte auch Dreiecke aus jeweils einem realen Lautsprecher-Paar und einem virtuellen Lautsprecher enthalten. Nach Berechnung der Verstärkungsfaktoren werden die Signalanteile des virtuellen Lautsprechers mithilfe einer *Downmixing*-Funktion auf umliegende reale Lautsprecher verteilt.

Regionen, welche aus vier Lautsprechern gebildet sind, werden im *EAR* als *QuadRegions* bezeichnet und produzieren an den Randpositionen der Region dieselben Gains wie Triplet-Regionen, weshalb Kompatibilität zwischen den Regionentypen besteht.

Aufgrund der Beschränkungen, die mit einem $0+2+0$ Stereo-Layout einhergehen, behandelt der *Polar Point Source Panner* jenes als Spezialfall. Das Panning einer gegebenen Objektposition erfolgt zunächst mit dem Algorithmus für die Lautsprecheranordnung $0+5+0$. Dieser produziert fünf Verstärkungsfaktoren als Vektor g' in der Reihenfolge $M+030, M-030, M+000, M+110, M-110$. Aus diesem Vektor werden durch Anwendung einer Konvertierungsmatrix (Gleichung 5.1) die Stereo-Gains erzeugt.

$$g'' = \begin{bmatrix} 1 & 0 & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{2}} & 0 \\ 0 & 1 & \sqrt{\frac{1}{3}} & 0 & \sqrt{\frac{1}{2}} \end{bmatrix} \cdot g' \quad (5.1)$$

Die anschließende Leistungsnormalisation hängt vom Verhältnis zwischen den Signalanteilen auf den vorderen Lautsprechern zu jenen auf den hinteren Lautsprechern ab, sodass Schallquellen zwischen den Lautsprecherpositionen von $M+030$ und $M-030$ konstant bleiben, während Signale mit Position zwischen $M+110$ und $M-110$ um 3 dB abgesenkt werden[22, S.25]. Dies hat zur Folge, dass der Panning-Algorithmus für den Stereo-Fall insgesamt nicht lautheitserhaltend arbeitet.

5.2.2. Cartesian Point Source Panner

Der *EBU ADM Renderer* wendet den *Cartesian Point Source Panner* zur Berechnung der einzelnen Lautsprecher-Gains zur Abbildung einer Schallquelle an, wenn ihre Position als kartesische Koordinaten X, Y und Z vorliegt. Grundlage des Algorithmus ist eine 3D Erweiterung der in Abschnitt 4.1.1 beschriebenen *Sine/Cosine Pan Law*. Der *Cartesian Point Source Panner* setzt voraus, dass alle Lautsprecher sich an den Wänden, der Decke oder am Boden des Raumwürfels befinden. Um die Verstärkungsfaktoren der Lautsprecher zu berechnen, werden zunächst zwei Lautsprecher-Ebenen bestimmt: eine Ebene oberhalb und eine Ebene unterhalb der gewünschten Objektposition. Aus den Z -Positionen dieser zwei Ebenen und der Position der Quelle erfolgt die Berechnung der z -Verstärkungsfaktoren. Dann werden auf beiden Ebenen zwei Reihen an Lautsprechern ermittelt, welche sich vor bzw. hinter der gewünschten Objektposition befinden und die zugehörigen y -Gains kalkuliert. Als letztes sucht der *Panner* in jeder dieser Reihen zwei Lautsprecher, welche sich rechts und links der Objektposition befinden und berechnet deren x -Gains.

Auf diese Weise werden bis zu acht Lautsprecher ausgewählt. Der Verstärkungsfaktor jeder dieser Lautsprecher berechnet sich aus den jeweiligen Faktoren für x, y und z als $g = x \cdot y \cdot z$.

Die Gains aller anderen Lautsprecher beträgt 0[22, S.53].

Im Gegensatz zum *Polar Point Source Panner* behandelt der *Cartesian Point Source Panner* das Stereo-Layout nicht als speziellen Fall. Der kartesische Panning-Algorithmus des *EAR* soll für alle Lautsprecheranordnungen leistungserhaltend arbeiten[22, S.55].

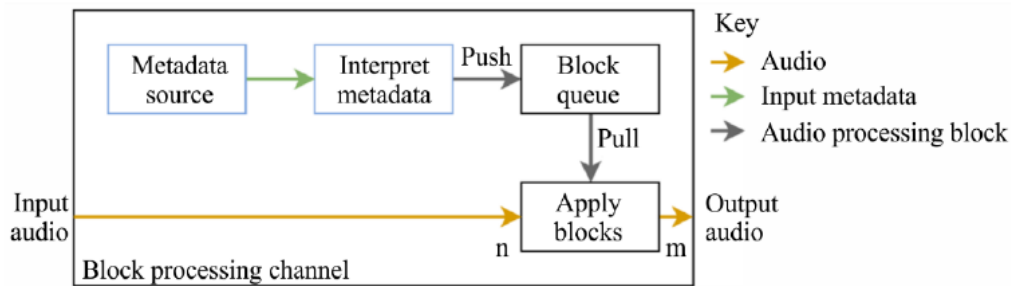


Abbildung 5.4.: Blockweise Verarbeitung im *EAR* nach Rec. ITU-R BS.2127[22, S.30]

5.2.3. Block Processing

In den Metadaten definierte Attribute von Audioobjekten können sich auf das ganze Objekt beziehen, sie können aber auch nur zeitliche Teilabschnitte innerhalb des Audioobjekts betreffen. Um die Umsetzung dieser zeitgebundenen Parameter zu ermöglichen, erfolgt das *Processing* im *EAR* blockweise. Abbildung 5.4 veranschaulicht die Funktionsweise der Blockverarbeitungen. Durch Aufruf der Methode *get_next_block* der Klasse *Metadata Source* erhält der Interpreter den nächsten vorliegenden Metadatenblock und wandelt diesen durch Interpretation der Metadaten in *ProcessingBlocks*. Die *ProcessingBlocks* werden in einer Schlange zwischengespeichert und nacheinander mit der *process*-Methode auf die Audiosamples angewandt. Aufgrund der im *ProcessingBlock* enthaltenen Angaben zum ersten und letzten Sample des Blocks ist dem Renderer bekannt, wann der nächste Block aufgerufen werden muss[22, S.31].

Bei Objektdaten liegen zeitliche Informationen des Objekts als *objekt_start* und *object_duration* vor. Sind diese nicht definiert, wird der Start des Objekts als 0 angenommen und die Dauer als unendlich. Start- und Endzeiten der einzelnen Blocks ergeben sich aus den Block-Parametern *rtime* und *duration*. Sind diese nicht angegeben, entspricht die zeitliche Ausdehnung des vorliegenden Blocks der des zugehörigen Objekts. Ansonsten entspricht der absolute Blockstart *object_start* plus *rtime*.

5.2.4. Interpretation Objektbezogener ADM Parameter

Bei der objektbasierten Audiorepräsentation besteht die Möglichkeit, weitere Objekteigenschaften in den *ADM*-Parametern zu definieren. Aufgabe des Renderers ist die Umsetzung der vorgegebenen Blockparameter. Im Folgenden werden die *EAR*-Implementierungen zu den für diese Arbeit relevanten Parametern *Extent*, *Divergence*, *Diffuse* und *Gain* grundlegend beschrieben.

Position

Positionsangaben zu den Audiosignalen sind für die *typeDefinitions DirectSpeaker* und *Objects* vorgesehen. Bei der objektbasierten Audiorepräsentation besteht außerdem die Möglichkeit, die Bewegung eines Objekts im Raum darzustellen. Dies geschieht durch die zeitliche Unterteilung eines Objekts in Blöcke. Die Positionsparameter dieser Blöcke können so bestimmt werden, dass eine räumliche Bewegung erzielt wird. Mit Setzung des Parameters *jumpPosition* kann außerdem die Interpolationszeit zwischen den einzelnen Blöcken festgelegt werden. Sollte dies nicht der Fall sein, so wird über die gesamte Blockdauer interpoliert.

Extent

Der Parameter *extent* wird durch die Elemente Höhe (*height*), Breite (*width*) und Tiefe (*depth*) repräsentiert und bestimmt die räumliche Ausdehnung eines Audioobjekts. Bei polaren Positionskordinaten sind Höhe und Breite als Winkel anzugeben, während die Tiefe als Fließkommazahl vorliegt. Wird mit kartesischen Positionsdaten gearbeitet, entsprechen Höhe, Breite und Tiefe der jeweiligen Ausdehnung auf den entsprechenden Achsen und sind, wie die Positionsangaben, normalisierte Einheiten.

Die in *core.objectbased.gain_calc* implementierte *PolarExtentHandler*-Klasse berechnet mit Angaben zu Breite (*width*), Höhe (*height*), Tiefe (*depth*) und Position eines Objekts die Verstärkungsfaktoren aller vorhandenen Lautsprecher. Zunächst erfolgt eine Modifizierung der Parameter *width*, *height* und *position*, um *depth* und die Distanz-Komponente der Position umzusetzen. Ein Distanzwert von 1 lässt Höhe und Breite des Objekts unverändert. In anderen Fällen gilt: Je größer der Distanz-Wert, desto kleiner ist die Ausdehnung des Objekts.

Zur Umsetzung eines *depth*-Werts ungleich null produziert der *EAR* zwei ausgedehnte Signalquellen mit unterschiedlicher Distanz. Die konkreten Distanzwerte dieser Quellen werden in Abhängigkeit der Ausdehnungstiefe berechnet. Anschließend werden die Gain-Vektoren zu beiden Signalquellen ermittelt, welche letztlich mit einem leistungserhaltenden Algorithmus zu einem Verstärkungsfaktor pro Lautsprecher verrechnet werden.

Die modifizierten Werte werden dem *Spreading Panner* übergeben, der eine Gruppe von virtuellen Schallquellen enthält und durch Zugriff auf den *Point Source Panner* die Lautsprecher-Verstärkungsvektoren zu jeder dieser Positionen berechnet. Auf die virtuellen Schallquellen wird eine der Objektausdehnung entsprechende Gewichtungsfunktion angewandt. Abbildung 5.5 zeigt ein Beispiel einer solchen Gewichtungsfunktion.

Innerhalb eines Rechtecks mit den Maßen *width* x *height* betragen die Werte der Funktion 1.

Die Ecken des Rechtecks sind abgerundet und der Übergang von 1 auf 0 am Rand des gewichteten Bereichs ist fließend.

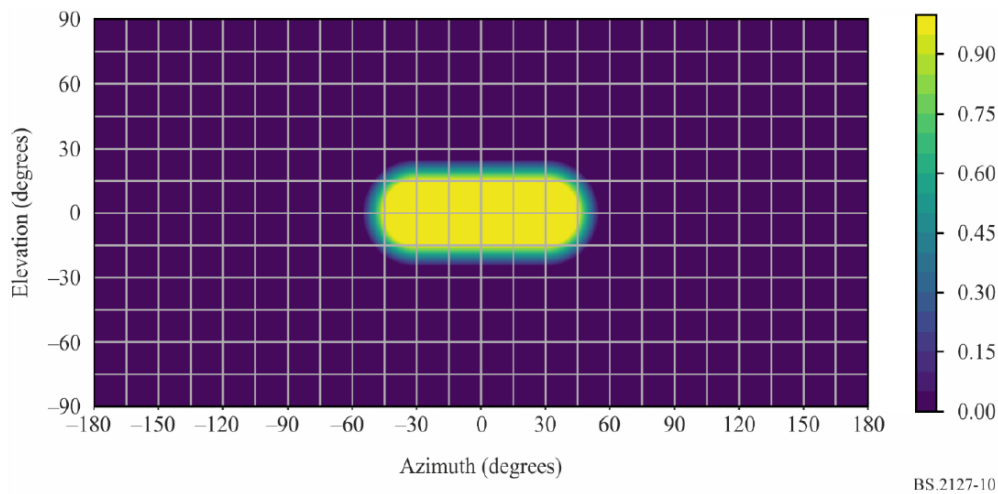


Abbildung 5.5.: Polare Gewichtungsfunktion zur Umsetzung des *extent*-Parameters mit $width = 90^\circ$ und $height = 30^\circ$ nach Recommendation ITU-R BS.2127[22, S.51]

Nach Anwendung der Gewichtungsfunktion liegen Verstärkungsfaktoren für die virtuellen Schallquellen vor. Mit den bereits ermittelten Verstärkungsfaktoren der einzelnen Lautsprecher zur Darstellung der Schallquellenpositionen wird schließlich jeweils ein Gain-Wert pro Lautsprecher berechnet. Zuletzt findet eine Leistungsnormalisation statt.

Im Falle von kartesischen Parameterangaben werden andere Klassen und Methoden des Renderers herangezogen. Die Implementierung ähnelt in ihren Grundzügen der beschriebenen Vorgehensweise bei polaren Positionsangaben. Zunächst werden virtuelle Schallquellen erzeugt. Diese werden in Abhängigkeit der Objektposition und der zugehörigen *extent*-Werte gewichtet und letztlich werden die Verstärkungsfaktoren der virtuellen Schallquellen mithilfe des *Cartesian Point Source Panners* auf die real vorliegenden Lautsprecher übertragen.

Aufgrund der Kubusform des kartesischen Koordinatensystems sind weitere Anpassungen erforderlich. Aus ästhetischen Gründen soll beispielsweise verhindert werden, dass sich gegenüber stehende Lautsprecher gleichzeitig dasselbe Signal wiedergeben. Durch Algorithmen der Methode *core.objectbased.allo_extent.get_gains* wird diese Situation verhindert.

Divergence

Mit dem Parameter *objectDivergence* kann ein Objekt symmetrisch in virtuelle Objekte geteilt werden, sodass eine Phantomschallquelle auf der originalen Objektposition entsteht. Durch Nutzung des *divergence*-Parameters ändert sich die räumliche Abbildung der Audioszene nicht, es werden lediglich gegebenenfalls andere Lautsprecher zur Bildung der Phantomschallquelle verwendet. Außerdem ist eine leistungserhaltende Arbeitsweise vorgesehen.

Der Parameter besitzt die Attribute *value* und *azimuthRange*.

value bestimmt das Leistungsverhältnis zwischen den zwei symmetrischen virtuellen Schallquellen und der Originalschallquelle und entspricht einer Fließkommazahl zwischen 0 und 1, wobei bei *value* = 0.0 nur die Originalschallquelle und bei *value* = 1.0 ausschließlich die virtuellen Schallquellen verwendet werden.

azimuthRange wird bei Nutzung des polaren Koordinatensystems in Grad angegeben und definiert die Positionen der virtuellen Schallquellen relativ zur Position der Originalschallquelle. Wird eine *azimuthRange* von 45° festgelegt, so befinden sich die virtuellen Schallquellen 45° links und rechts von der Objektposition.

Falls mit kartesischen Positionsangaben gearbeitet wird, ist an Stelle der *azimuthRange* die *positionRange* als Float-Wert zwischen 0 und 1 anzugeben.

Die Objektdivergenz ist im *EAR* wie folgt implementiert:

Zur ursprünglichen Position p_c werden in Abhängigkeit der angegebenen *azimuthRange* bzw. *positionRange* links und rechts zwei weitere Punktquellen p_l und p_r generiert. Die *objectDivergence*-Methode des Renderers nimmt eine Position und die Divergenz-Daten eines Audioobjektes entgegen und produziert die Gains g_c , g_l und g_r der drei Punktquellen p_c , p_l und p_r . Für einen *objectDivergence*-Wert x berechnen sich die drei Verstärkungsfaktoren, wie folgt:

$$g_c = \frac{1-x}{x+1} \quad \text{und} \quad g_l = g_r = \frac{x}{x+1} \quad (5.2)$$

Positionen und Gains werden dem *Extent Panner* zum Rendern übergeben[22, S.45].

Diffuse

Der Objektparameter *diffuse* kann als beliebiger Wert zwischen 0.0 und 1.0 gewählt werden und bestimmt, wie direkt oder diffus ein Signal klingen soll. Der Default-Wert 0.0 bedeutet keine Unschärfe im Klang, während 1.0 für die Erstellung vollständig diffuser Klänge vorgesehen ist.

Liegt der *diffuse*-Wert eines Audioobjekts über 0, wird der in Abbildung 5.6 abgebildete Dekorrelationspfad des Objekt-Renderers verwendet. Bei den Dekorrelationsfiltern handelt es sich um Allpass-Filter mit zufälliger Phase. Liegt ein *diffuse*-Wert von 1 vor, durchlaufen alle Audiosamples den Dekorrelationspfad. Bei einem Wert zwischen 0 und 1 werden die Audiosamples anteilig de-korreliert, wobei die nicht-dekorrelierten Signalanteile verzögert werden, um das Delay des Filters auszugleichen.

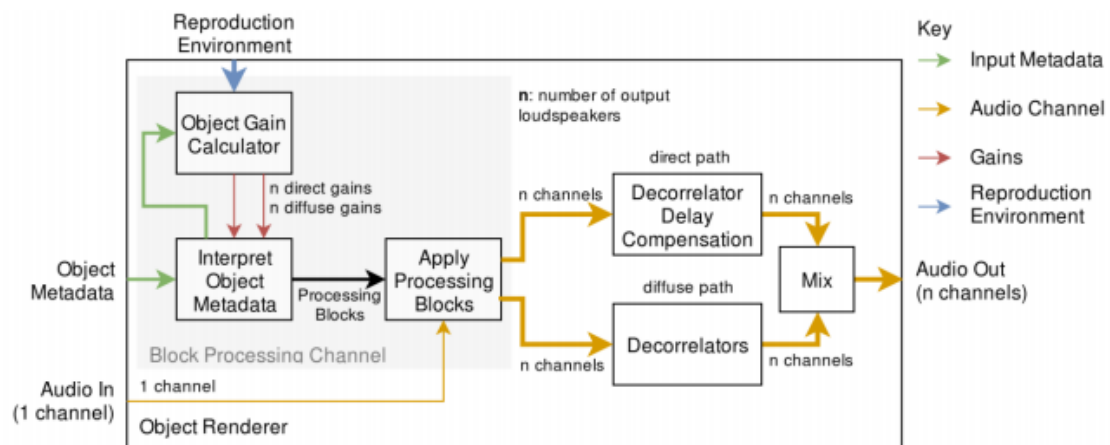


Abbildung 5.6.: Die Struktur des Objekt-Renderers nach Rec. ITU-R BS.2127[22, S.34]

Gain

Der *gain*-Parameter kontrolliert den Pegel des Objektsignals. Um Übersteuerungen zu verhindern, können Audiosignalpegel lediglich abgesenkt und nicht angehoben werden. Als Gain-Wert kann eine Fließkommazahl zwischen 0 und 1 festgelegt werden. Diese wird durch Multiplikation auf den Signalpegel des Audioblocks angewandt.

Beim Rendering wird der *gain*-Wert mit den einzelnen Leistungswerten der Audiosamples innerhalb des Audioblocks multipliziert. Zur Verhinderung von Pegelsprüngen und damit verbundenen Artefakten findet zwischen zwei Audioblöcken eine Interpolation der Gain-Werte statt. Dies ist implementiert in der Klasse *core.renderer_common.Interp_Gains*. Liegt kein *gain*-Wert vor, so wird dieser als 1.0 angenommen und beeinflusst somit nicht den ursprünglichen Signalpegel eines Objektblocks.

6. Objektbasierte Lautheitsmessung

Bisher existiert noch kein Standard zur Lautheitsmessung von objektbasierten Audioszenen. Projekte, in deren Rahmen mit objektbasierten Audioszenen gearbeitet wurde, mussten auf das kanalbasierte Messverfahren nach Rec. ITU-R BS.1770-4 zurückgreifen[27, S.30]. Zur Messung der kanalbasierten Lautheit muss eine OBA-Szene jedoch erst gerendert werden. Da OBA-Szenen aber unabhängig vom Wiedergabesystem und der Lautsprecheranordnung sind, lässt sich mit der kanalbasierten Messmethode kein einheitlicher Lautheitswert für eine objektbasierte Audioszene generieren. Aus diesem Grund ist die Entwicklung eines generischen also vom Wiedergabesystem unabhängigen Lautheitsmessverfahren interessant. Ein derartiges Messverfahren würde die Aussteuerung einer Audioszene auf einen gewünschten Lautheitswert ermöglichen.

6.1. Aktueller Forschungsstand zur Lautheitsmessung von objektbasierten Audioszenen

Im Jahre 2017 beschäftigte sich das Institut für Rundfunktechnik (IRT) mit verschiedenen Ansätzen zur Lautheitsmessung objektbasierter Audioszenen auf Basis des kanalbasierten Messverfahrens nach Rec. ITU-R BS.1770-4. Unter anderem wurde das in Abbildung 6.1 vereinfacht dargestellte Modell entwickelt. Anstatt der einzelnen Kanäle einer Datei oder eines Streams werden hier Audioobjekte als Eingangssignale verwendet, welche anschließend die Bearbeitungsschritte nach Rec. ITU-R BS.1770-4 durchlaufen. Zur Durchführung der positionsabhängigen Gewichtung wurde die jeweilig vorliegende Objektposition berücksichtigt.

Ein standardisierter Renderer existierte 2017 noch nicht. Das Modell wurde mit damals vorhandenen Renderer Proposals getestet und erzielte dabei zufriedenstellende Ergebnisse[8, S.23]. Die Möglichkeit, Objekteigenschaften, wie Ausdehnung, Pegel oder Divergenz zu rendern, bestand damals noch nicht. Aus diesem Grund konnten Auswirkungen der Objektparameter auf die Lautheit nicht untersucht werden. Auf Basis der verschiedenen Renderer Proposals wurde der *EBU ADM Renderer (EAR)* entwickelt, der nun den vollständigen *ADM*-Parametersatz für Audioobjekte unterstützt.

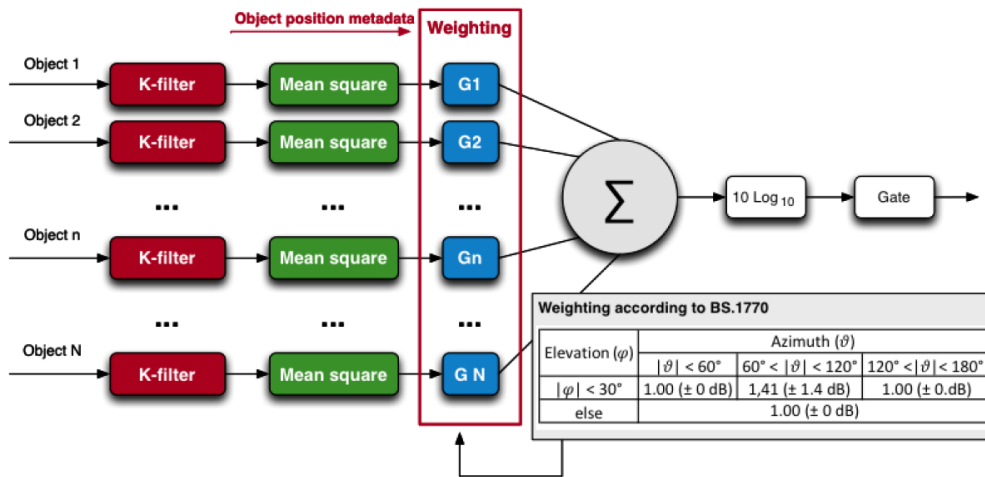


Abbildung 6.1.: Objektbasiertes Messverfahren mit Gewichtung nach Objektposition[8, S.20]

In Zusammenhang mit der Fertigstellung des *EAR* besteht nun die Notwendigkeit, die vorhandenen Ansätze zu überprüfen und die Auswirkungen der nun implementierten Objekteigenschaften auf die Lautheit zu untersuchen. Darüber hinaus sind in der Zwischenzeit weitere objektbasierte Audioszenen produziert worden. Diese nutzen unter anderem mehrere Audioprogramme, die vom Konsumenten gewählt werden können. Für derartige Audioszenen existiert bislang keine Möglichkeit zur Lautheitsmessung.

6.2. Amplitudenaddition kohärenter Audioobjekte

Bisherige Forschungen zur Lautheitsmessung von objektbasierten Audioszenen zeigten eine Problematik in Verbindung mit der Kohärenz von Audioobjekten auf, welche zu Abweichungen der Lautheit führen kann. Der Hintergrund dieser Problematik liegt im Unterschied zwischen Amplituden- und Leistungsaddition von Audiosignalen. Bei der Amplitudenaddition werden zu jedem Zeitpunkt die Momentanwerte mehrerer Signale addiert. Das Ergebnis ist abhängig von Verlauf und Phasenlage der Signale. Liegen zwei kohärente und phasengleiche Audiosignale *A* und *B* mit gleichem Pegel vor, so gilt zur Berechnung des resultierenden Pegels:

$$L_{A+B} = 20 \log_{10} \left(10^{\frac{L_A}{20}} + 10^{\frac{L_B}{20}} \right) \text{ dB} \quad (6.1)$$

Der Pegel des summierten Signals ist um 6 dB höher als eines der Ursprungssignale.

Liegt aber keine Kohärenz vor, so gilt die Gleichung:

$$L_{A+B} = 10 \log_{10} \left(10^{\frac{L_A}{10}} + 10^{\frac{L_B}{10}} \right) \text{ dB} \quad (6.2)$$

Die Anhebung des Gesamtpegels beträgt in diesem Fall lediglich +3 dB.

Die Addition von Leistungswerten ist unabhängig von Verlauf und Phasenlage der zu addierenden Signale. Daher gilt für die Leistungsaddition die Gleichung (6.2) [13].

Enthält eine Audioszene Objekte mit einem Korrelationsgrad $\neq 0$ und werden diese Objekte durch ihre räumliche Nähe auf denselben Lautsprecherkanal gerendert, so findet eine Mischung aus Amplituden- und Leistungsaddition statt. Der resultierende Pegel ist höher als bei einer Addition von Leistungswerten, wie sie bei den Ansätzen der Lautheitsmessung von OBA-Szenen erfolgen würde (siehe Abschnitt 6.1). Eine Abweichung des Lautheitswerts ist die Folge. Die Höhe der Abweichung hängt zum einen vom Korrelationsgrad der Objekte und deren Position ab, zum anderen hängt sie mit der vorliegenden Lautsprecheranordnung, auf die die OBA-Szene gerendert wird, zusammen. Insbesondere Anordnungen mit einer geringen Lautsprecheranzahl sind problematisch, da es hier häufiger vorkommt, dass zwei kohärente Objekte auf denselben Lautsprecher gerendert werden.

Weiterhin ist auch eine negative Korrelation von Audioobjekten möglich. Diese führt zu einer Absenkung der Lautheit.

Aufgrund der Abhängigkeit vom Wiedergabesystem ist eine Lösung des Problems im Rahmen eines generischen Messverfahrens nicht möglich. Ein solches Verfahren ermittelt einen einzigen Lautheitswert unabhängig vom Wiedergabesystem. Bei korrelierenden Objekten entstehen beim Rendern auf unterschiedliche Lautsprecheranordnungen jedoch Lautheitsabweichungen. Es ist zu prüfen, ob sich diese Abweichungen der Lautheit in der Praxis in einem akzeptablen Bereich befinden.

6.3. Einfluss der ADM-Parameter auf die Lautheit von Audioszenen

Für die Entwicklung eines objektbasierten Lautheitsmessverfahrens ist es essentiell, den Einfluss der ADM-Parameter auf die Lautheit einer Audioszene zu erforschen. Aus diesem Grund wurde von der Autorin ein Messsystem eingerichtet (siehe Abbildung 6.2). Als Testsignal diente ein weißes Rauschen, da dieses alle Frequenzen in gleicher Intensität beinhaltet. Mithilfe der im *EAR* implementierten Kommandozeilenfunktion *make_test_bwf* wurden ADM-Dateien erstellt, die jeweils aus einem 20-sekündigen Audioobjekt mit besagtem Testsignal bestanden. Für jede dieser ADM-Dateien wurden anschließend kanalbasierte Dateien entsprechend aller nach Rec. ITU-R BS.2051[23] definierten Lautsprecheranordnungen gerendert. Hierfür wurde die *render*-Funktion

6. Objektbasierte Lautheitsmessung

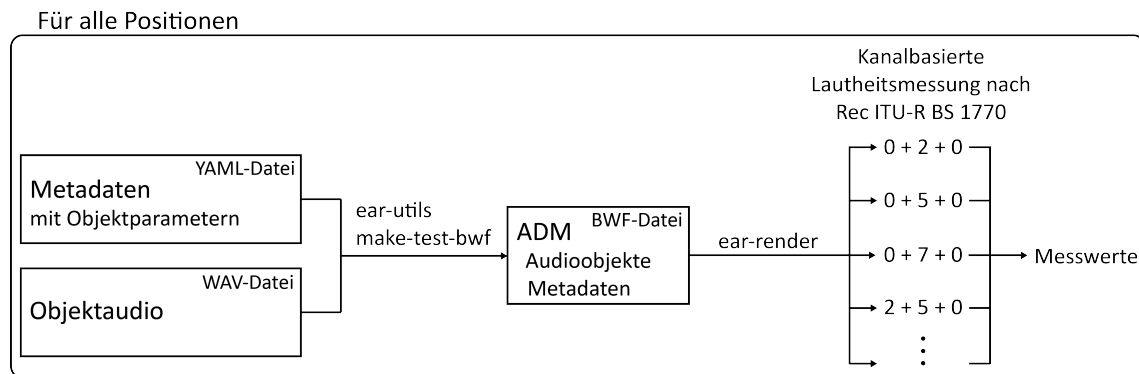


Abbildung 6.2.: Stapelverarbeitung zur kanalbasierten Lautheitsmessung mit unterschiedlichen Objektparametern

des *EAR* in der Kommandozeile ausgeführt.

Um die Lautheit bei unterschiedlichen Objektpositionen zu prüfen, wurden *ADM*-Szenen für unterschiedliche Positionen im Raum erstellt. Beim Rendering ist von einer horizontal symmetrischen Arbeitsweise und Lautsprecheranordnung auszugehen. Aus diesem Grund ist lediglich die Messung einer Raumhälfte notwendig. Im polaren Koordinatensystem repräsentiert der Azimuthbereich zwischen 0° und 180° eine Hälfte des Raums. In der vertikalen Ebene ist dagegen keine Symmetrie gegeben, daher wurden Elevationswerte über den gesamten Bereich zwischen -90° und 90° berücksichtigt. Im polaren Raum wurden insgesamt 1369 Positionen untersucht, was einer vertikalen und horizontalen Abdeckung auf fünf Grad genau entspricht.

Zuletzt wurde die Lautheit aller gerenderten Dateien kanalbasiert nach Rec. ITU-R BS.1770-4 gemessen und abgespeichert.

Für die kanalbasierte Lautheitsmessung stand ein Python Script zur Verfügung, welches anhand eines Arrays aus Audiosamples den Lautheitswert der zugehörigen Audiodatei nach Rec. ITU-R BS.1770-4 bestimmen kann. Die einzelnen Module zur Lautheitsermittlung sind in Methoden implementiert. Zunächst durchlaufen die Audiosamples das Filtermodul, dann werden Leistungsblöcke gebildet, es folgt die Gewichtung und schließlich die Summenbildung.

Um Vergleiche zwischen polaren und kartesischen Objektpositionen anstellen zu können, wurde der kartesische Raum analog zum polaren Raum durch die Positionskoordinaten X, Y und Z abgetastet. Aufgrund der unterschiedlichen Raumformen der zwei Koordinatensysteme sind die so entstandenen Raumpositionen nicht identisch. Dennoch ist auf diese Weise ein Vergleich der Lautheitswerte möglich.

Da bei den Parametern *extent* und *divergence* Lautheitsabweichungen aufgrund des Gewichtungsmoduls der Lautheitsmessung zu erwarten sind, wurde vorerst eine Lautheitsmessung ohne richtungsabhängige Gewichtung durchgeführt.

So konnte überprüft werden, ob der Renderer bei der Umsetzung der einzelnen *ADM*-Parameter grundsätzlich leistungserhaltend arbeitet.

Lautheitsabweichungen in Zusammenhang mit der richtungsabhängigen Gewichtung, welche die Rec. ITU-R BS.1770-4 für die kanalbasierte Lautheitsmessung eigentlich vorschreibt, wurden mit einer zweiten Messung untersucht.

Neben den Objekteigenschaften zur Ausdehnung und Divergenz wurde der *diffuse*-Parameter bezüglich seines Einflusses auf die Lautheit untersucht. Weitere Objektparameter, wie beispielsweise die Referenz auf die vorhandene Bildschirmschirmgröße, modifizieren lediglich in manchen Fällen die Position des Objekts und sind aufgrund dessen nicht näher betrachtet worden.

Im Gegensatz zu den anderen Objektparametern ist in Verbindung mit dem *gain*-Parameter eine Pegeländerung des Objektsignals beabsichtigt. Demzufolge ist die Einflussnahme des *gain*-Parameters auf die Lautheit einer Audioszene bereits bekannt. Bei der Entwicklung eines objektbasierten Lautheitsmessverfahrens sind dementsprechend die *gain*-Werte der Audioobjekte zu berücksichtigen.

6.3.1. Veranschaulichung der Lautheitswerte mithilfe einer Heatmap

Zur Veranschaulichung der gemessenen Lautheitswerte bei allen Objektpositionen fertigte die Autorin sogenannte *Heatmaps* an. Abbildung 6.3 zeigt beispielhaft die Ergebnisse einer Lautheitsmessung mit positionsabhängiger Gewichtung. Gemessen wurden auf die Lautsprecheranordnung 4+9+0 gerenderte Dateien bestehend aus einem Audioobjekt. Als Audiosignal diente ein weißes Rauschen. Jeder Bildpixel in der gezeigten *Heatmap* repräsentiert eine Messung der entsprechenden Objektposition. Die Azimuthwerte sind auf der X-Achse und die Elevationswerte auf der Y-Achse zu finden. Die relativen Lautheitswerte werden durch die Einfärbungen der Pixel angezeigt und der Farbbalken gibt Auskunft über die Höhe der Lautheitsdifferenzen.

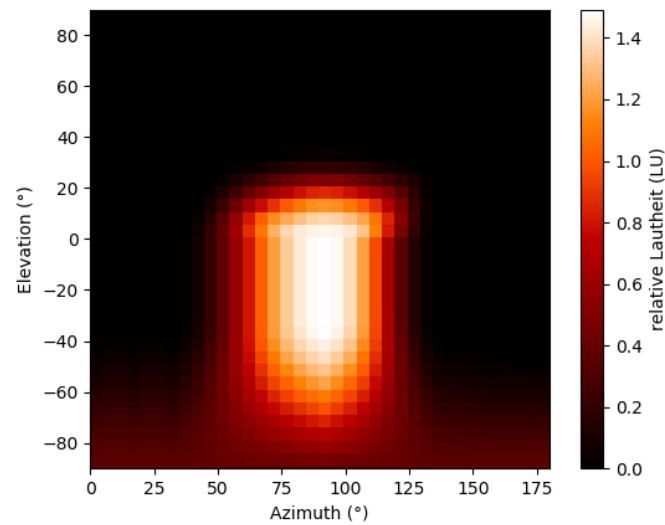


Abbildung 6.3.: Heatmap der Lautheitswerte einer 4+9+0-Lautsprecheranordnung bei Lautheitsmessung mit positionsabhängiger Gewichtung

Anhand dieser Heatmap ist z.B. zu erkennen, dass im Positionsbereich $60^\circ < |Azimuth| < 120^\circ$ und $-80^\circ < Elevation < 30^\circ$ eine Anhebung der Lautheit um bis zu 1.4 LU stattfindet. In diesem Fall sind die Lautheitsabweichungen vollständig auf die positionsabhängige Gewichtung des kanalbasierten Lautheitsmessverfahrens zurückzuführen.

6.3.2. Lautheitsmessung ohne Parameter

Um die lautheitserhaltende Arbeitsweise des *EAR* als Grundvoraussetzung für folgende Messungen sicherzustellen, wurden zunächst Audioobjekte ohne Angabe einzelner Parameter (mit Ausnahme der Objektposition) erstellt und die Lautheit der gerenderten Dateien für alle Lautsprecheranordnungen und unterschiedliche Objektpositionen gemessen. Wie erwartet, wurden bei der Lautheitsmessung ohne positionsabhängige Gewichtung keine nennenswerten Abweichungen der Lautheit festgestellt. Differenzen der Messwerte betrugen hier bis zu 0.0001 LU und sind aufgrund ihrer geringen Ausmaße in jeglicher Hinsicht vernachlässigbar.

Eine Ausnahme dieser Feststellung ist das Rendering auf eine 2.0-Stereoanordnung bei polaren Positionsangaben. Dieses wird im *Polar Point Source Panner* des *EAR* gesondert behandelt, da im Stereo-Fall Objekte mit Position hinter dem Hörer nicht an der vorgesehenen Stelle wiedergegeben werden können (siehe Abschnitt 5.2.1). Solche Objekte werden mit abgesenktem Pegel von den zwei Lautsprechern des 2.0-Setups dargestellt.

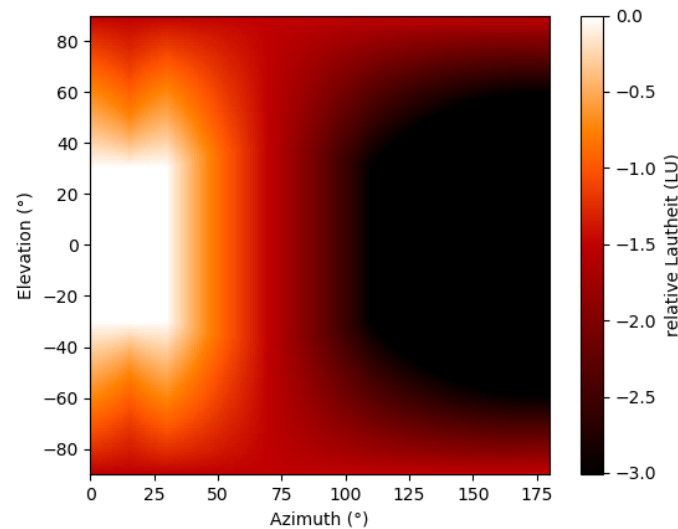


Abbildung 6.4.: Heatmap zur Veranschaulichung der Lautheitswerte eines Audioobjekts bei Rendering auf ein 2.0-Lautsprecher-setup mit polaren Positionsangaben

Abbildung 6.4 veranschaulicht die Lautheitswerte einer vom *EAR* gerenderten Stereo-Datei in Form einer *Heatmap*. Der weiß eingefärbte Bereich zwischen Elevation -30° bis 30° und Azimuthwerten bis zu 30° zeigt an, dass keine Lautheitsabweichungen entstehen. Dieser Raumbereich deckt sich mit den Lautsprecherpositionen einer 2.0-Stereoanordnung. Sobald Objekte mit Positionen außerhalb dieses Bereichs gerendert werden, entstehen Abweichungen von bis zu -3 LU. Der Differenzwert von -3 LU stimmt exakt mit den *Downmixing*-Koeffizienten des *Polar Point Source Panners* überein.

Bei Nutzung von kartesischen Koordinaten arbeitet der Renderer jedoch auch im Stereo-Fall lautheitserhaltend, da der *Cartesian Point Source Panner* im Gegensatz zum *Polar Point Source Panner* für das Rendering auf *0+2+0* keine Pegelabsenkung von Objekten mit Position seitlich oder hinter dem Hörer vorsieht.

Zur Erforschung des Einflusses der einzelnen *ADM*-Parameter auf die Lautheit in Verbindung mit der richtungsabhängigen Gewichtung nutzte die Autorin dieser Arbeit das bereits bekannte Python Script zur kanalbasierten Lautheitsmessung und erweiterte dieses um eine Methode zur Gewichtung des nach Rec. ITU-R BS.1770-4 definierten Positionsbereichs. Durch Angabe der vorgesehenen Lautsprecheranordnung einer Audiodatei konnten die betreffenden Kanalnummern identifiziert werden. Die für jene Kanäle gebildeten Leistungsblöcke wurden dann mit dem Gewichtungsfaktor 1.41 multipliziert.

Um später Vergleiche zwischen der Lautheit von Audioobjekten mit und ohne die zu untersuchenden Parameter anstellen zu können, erfolgte zunächst die Messung von Testszenen bestehend aus einem Audioobjekt ohne Angabe von Parametern. Aufgrund der unterschiedlichen Anzahl an Lautspre-

chern je nach vorliegender Lautsprecheranordnung fallen manche Objektpositionen bei kleineren Anordnungen in den zu gewichtenden Bereich, während sie bei größeren Anordnungen nicht gewichtet werden. Dieser Umstand sorgt für Lautheitsdifferenzen in Abhängigkeit des genutzten Wiedergabesystems.

Abbildung 6.5 veranschaulicht die Lautheitswerte bei gewichteter Messung für die Wiedergabesysteme 5.1 und 22.2. Die weiß eingefärbten Pixel zeigen jeweils die Objektpositionen an, bei denen durch das Gewichtungsmodul der Lautheitsmessung eine Anhebung der Lautheit entsteht.

Es ist zu erkennen, dass bei der 5.1-Lautsprecheranordnung ein deutlich größerer Positionsbereich bei der Lautheitsmessung gewichtet wird als es bei der 22.2-Anordnung der Fall ist. Als Begründung ist an dieser Stelle anzubringen, dass die Objektpositionen bei einem 22.2-Lautsprecheranordnung durch die hohe Anzahl an Lautsprechern akkurat auf der vorgesehenen Position wiedergegeben werden können, während für die 5.1-Wiedergabe lediglich fünf Lautsprecher zur Verfügung stehen, die sich zudem alle auf der gleichen Höhenebene befinden. Bei der 5.1-Lautsprecheranordnung werden daher alle Objektpositionen mit $|Azimuth| > 90^\circ$ auf die Lautsprecher L_s und R_s gerendert, wodurch sie bei der kanalbasierten Lautheitsmessung gewichtet werden.

Generell gilt: Je größer die Lautsprecheranzahl eines Systems, desto weniger Objektpositionen fallen nach dem Rendern in den zu gewichtenden Bereich.

Weiterhin ist anzumerken, dass bei der kanalbasierten Lautheitsmessung einer Stereo-Datei keine Gewichtung stattfindet. Grund dafür ist die Tatsache, dass bei der Stereo-Wiedergabe alle Objektpositionen von zwei Lautsprechern, die sich frontal zur Hörerposition befinden, wiedergegeben werden.

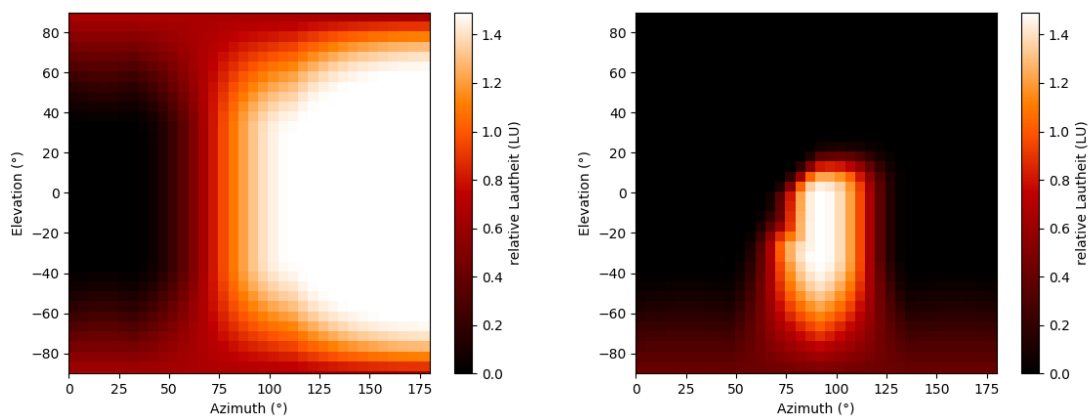


Abbildung 6.5.: Gewichtete Lautheitsmessung
bei den Lautsprecheranordnungen 0+5+0 (l) und 9+10+3 (r)

6.3.3. Extent

Um den Einfluss des *extent*-Parameters auf die Lautheit zu untersuchen, wurden *ADM*-Szenen bestehend aus jeweils einem Audioobjekt mit unterschiedlich großer Ausdehnung erstellt. Aus jenen *ADM*-Szenen wurden anschließend kanalbasierte Dateien für alle Lautsprecheranordnungen gerendert und deren Lautheit ermittelt. Auf diese Weise wurden sowohl die *extent*-Komponenten *width*, *height* und *depth* einzeln, als auch in Kombination miteinander untersucht.

Die polare Positionskomponente *distance* führt bei einem Wert $\neq 1$ zu einer Modifizierung der *extent*-Werte. Mit der Überprüfung des *extent*-Parameters ist demnach eine gesonderte Betrachtung der *distance*-Komponente nicht notwendig.

Bei Verwendung von kartesischen Positionen und *extent*-Werten durchlaufen die Audiosignale andere Bearbeitungsschritte als bei der Arbeit im polaren System. Aufgrund dessen wurden zusätzlich *ADM*-Dateien mit kartesischen Positionsangaben erstellt, zugehörige kanalbasierte Dateien gerendert und deren Lautheit überprüft.

Zunächst erfolgte die Lautheitsermittlung ohne richtungsabhängige Gewichtung der Lautsprecher-signale. Hierbei wurde festgestellt, dass die Umsetzung des *extent*-Parameters durch den *EAR* lautheitserhaltend erfolgt. Sowohl bei polaren als auch bei kartesischen Positionsangaben wurden lediglich irrelevant geringe Abweichungen des Lautheitswerts von unter 0.0002 LU verzeichnet.

Weiterhin erfolgten Messungen der Lautheit mit richtungsabhängiger Gewichtung.

Der *extent*-Parameter bewirkt eine räumliche Ausdehnung des zugehörigen Audioobjekts in Abhängigkeit der Attribute Höhe (*height*), Breite (*width*) und Tiefe (*depth*). Folglich besteht die Möglichkeit, dass Signalanteile eines Objekts mit Position außerhalb der gewichteten Zone durch die Objektausdehnung dennoch gewichtet werden und damit eine Erhöhung des Lautheitswerts verursachen. Analog dazu kann der *extent*-Parameter eines Objekts, welches sich innerhalb des Gewichtungsbereichs befindet, je nach den konkreten Werten für Breite und Höhe zu einem insgesamt geringeren Lautheitswert führen.

Zur Feststellung des Ausmaßes der erwarteten Lautheitsabweichungen wurden für *width* und *height* die polaren Winkelwerte 10.0, 20.0, 40.0 und 60.0 untersucht. Die Messungen ergaben Differenzen der Lautheitswerte von bis zu 0.3 LU. Höhere Werte für Breite und Höhe eines Objekts haben insgesamt größere Lautheitsabweichungen zur Folge. Das Attribut *depth* hat mit Abweichungen bis zu 0.06 LU einen geringeren Einfluss auf die Lautheit als Breite und Höhe.

Insgesamt sind die durch den *extent*-Parameter verursachten Lautheitsabweichungen im polaren Koordinatensystem vernachlässigbar gering.

Weiterhin wurde bewiesen, dass die Lautheit von Audioszenen bei Verwendung des *extent*-Parameters bei unterschiedlichen Lautsprecheranordnungen Varianzen aufzeigt.

Veranschaulichungen zu den Messergebnissen sind im Anhang (Abbildungen B.1 und B.2) zu finden.

Bei der Untersuchung des Einflusses der kartesischen *extent*-Parameter auf die Lautheit wurden in Verbindung mit der gewichteten Lautheitsmessung insgesamt noch geringere Lautheitsabweichungen festgestellt.

6.3.4. Divergence

Wie in Abschnitt 5.1.2 dargelegt, besteht der ADM-Parameter *objectDivergence* aus den Attributen *value* und *azimuthRange* für eine Winkelangabe bzw. *positionRange* für eine kartesische Angabe. Zur Untersuchung des Parameters in Bezug auf die Lautheit wurden Audioobjekte mit den *value*-Komponenten 0.5 und 1.0 getestet. Liegt für *value* ein Wert von 0.5 vor, so werden zur Darstellung der Phantomschallquelle sowohl die Originalquelle als auch symmetrische virtuelle Quellen verwendet. Bei *value* = 1.0 wird die Originalquelle dagegen nicht mitgenutzt.

Die folgenden Erkenntnisse gelten für beide Fälle.

Die beschriebenen Untersuchungen produzierten ähnliche Ergebnisse wie die Forschungen in Bezug auf den *extent*-Parameter.

Wenn die Lautheit ohne positionsabhängige Gewichtung gemessen wird, können bei der Anwendung des *divergence*-Parameters keine Abweichungen der Ausgangslautheit festgestellt werden. Dies gilt für beide Koordinatensysteme.

Werden Audioszenen mit gleichen Parameterangaben durch eine Lautheitsmessung mit positionsabhängiger Gewichtung gemessen, treten wie zu erwarten Abweichungen auf. Im kartesischen Koordinatensystem zeigten sich lediglich Abweichungen weit unter 0.1 LU.

Im polaren Raum kam es jedoch bei manchen Objektpositionen zu Abweichungen zwischen 0.5 und 1.5 LU. Grund dafür ist dasselbe Phänomen wie bei Nutzung des *extent*-Parameters. Durch die Darstellung einer Phantomschallquelle durch virtuelle Schallquellen links und rechts des Objektes, werden letztendlich andere Lautsprecher angesteuert als ohne Verwendung des *divergence*-Parameters. Es ist nun möglich, dass Signalanteile des Audioobjekts bei Ermittlung der Lautheit gewichtet werden, obwohl ohne Verwendung von *objectDivergence* keine Gewichtung stattgefunden hätte.

Die Differenzen der Lautheit sind zumindest bei Verwendung des polaren Koordinatensystems stärker ausgeprägt als bei den anderen untersuchten Parametern. Allerdings ist anzumerken, dass der Einsatz des *divergence*-Parameters die wahrgenommene Position eines Audioobjekts nicht ändern soll. Die Lautheitsabweichungen kommen nur aufgrund der positionsabhängigen Gewichtung der Lautsprecherkanäle zustande. Eine gebildete Phantomschallquelle kann vom Hörer demnach außerhalb der Gewichtungszone wahrgenommen werden, obwohl einer der angesteuerten Lautsprecher sich im Gewichtungsbereich befindet. Es ist daher anzunehmen, dass subjektiv kein Lautheitsunterschied zwischen einem Audioobjekt mit *divergence* und demselben Objekt ohne *divergence* empfunden wird.

6.3.5. Diffuse

Hinsichtlich des *diffuse*-Parameters wurden unter anderem die Werte 0.25, 0.5, 0.75 und 1.0 untersucht.

Die Untersuchungen haben gezeigt, dass bei einem *diffuse*-Wert von 1.0 keine nennenswerten Lautheitsabweichungen entstehen. Bei *diffuse*-Werten $\neq 1$ und $\neq 0$, wurden dagegen sowohl bei kartesischen als auch bei polaren Positionsangaben Differenzen von bis zu ± 0.3 LU festgestellt. Die richtungsabhängige Gewichtung der Lautsprecher-signale spielt dabei keine Rolle. Die ermittelten Lautheitsdifferenzen sind bei ungewichteter und gewichteter Messung identisch.

Weiterhin liegt bei der Betrachtung kanalbasierter Dateien, welche auf verschiedene Lautsprecheranordnungen gerendert wurden, eine Varianz der Lautheitswerte vor. Der Einfluss des *diffuse*-Parameters eines Audioobjekts auf die Lautheit ist demnach abhängig von der vorliegenden Lautsprecheranordnung. Aus diesem Grund ist die Berücksichtigung des *diffuse*-Parameters im Rahmen eines generischen Messverfahrens, welches unabhängig vom Wiedergabesystem sein soll, nicht realisierbar.

Insgesamt befinden sich die durch den *diffuse*-Parameter verursachten Lautheitsabweichungen in einem Bereich, der für den Menschen kaum wahrnehmbar ist.

Eine Studie von G. A. Soulodre et al. aus dem Jahre 2003[10] zeigte, dass im Vergleich zwischen gleichem Programmmaterial unter perfekten Konditionen mindestens eine Differenz von etwa 0.4 dB vorhanden sein muss, damit der Mensch einen Lautheitsunterschied wahrnehmen kann. Die erforderliche Pegeldifferenz hängt von der Signalart ab. Bei Sprachsignalen ist für die Wahrnehmbarkeit beispielsweise eine Pegeldifferenz von mindestens 0.6 dB notwendig[10].

Dennoch ist zu prüfen, ob durch eine ungünstige Addition mehrerer diffuser Audioobjekte bedeutende Diskrepanzen im Lautheitswert entstehen könnten.

Eine Auswahl der gemessenen Lautheitswerte sind in Anhang B veranschaulicht.

6.3.6. Zusammenfassung der Messergebnisse

Durch kanalbasierte Lautheitsmessungen unter Auslassung des Gewichtungsmoduls konnte die lautheitserhaltende Arbeitsweise des *EAR* in Verbindung mit der Umsetzung der Objektparameter nachgewiesen werden. Einzig bei Nutzung des *Diffuse*-Parameters sind leichte Abweichungen zu verzeichnen, die sich aber in einem tolerierbaren Rahmen befinden.

Die Messungen mit richtungsabhängiger Gewichtung zeigten zunächst, dass bei einer geringeren Lautsprecheranzahl mehr Objektpositionen in den Gewichtungsbereich fallen als bei einer höheren Lautsprecheranzahl. Dieser Umstand führt generell zu Lautheitsunterschieden in Abhängigkeit des vorliegenden Wiedergabesystems.

Weiterhin können in Verbindung mit der positionsabhängigen Gewichtung bei Einsatz der Parameter *extent* und *objectDivergence* Lautheitsabweichungen entstehen. Diese variieren in Abhängigkeit der

6. Objektbasierte Lautheitsmessung

vorliegenden Lautsprecheranordnung und der Objektpositionen.

Ein generischer Lautheitswert, der für alle Wiedergabesysteme gilt, kann demnach nur eine Annäherung an die kanalbasiert gemessenen Werte sein.

Es ist zu prüfen, ob die ermittelten Lautheitsabweichungen in Abhängigkeit der verwendeten Objektparameter bei der praktischen Arbeit mit objektbasierten Audioszenen relevant sind.

Da die genauen Werte immer abhängig vom Wiedergabesystem sind, können die untersuchten Parameter bei der Entwicklung eines generischen Messverfahrens nicht berücksichtigt werden.

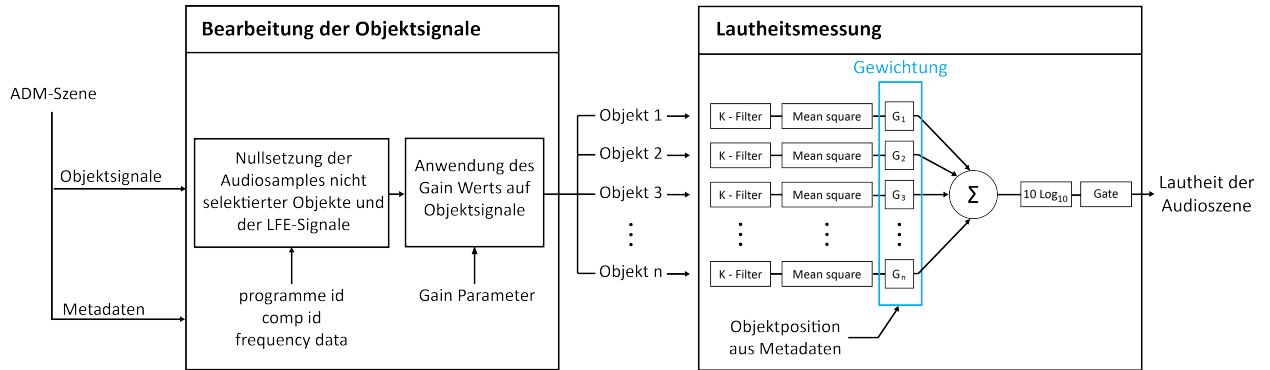


Abbildung 6.6.: Struktur des objektbasierten Lautheitsmessverfahrens

6.4. Implementierung eines Messverfahrens für objektbasierte Audioszenen

Zur Entwicklung eines generischen Lautheitsmessverfahrens für objektbasierte Audiodateien wurde das in Abschnitt 6.3 vorgestellte Python-Script zur kanalbasierten Lautheitsmessung nach ITU-R BS.1770-4 von der Autorin angepasst und erweitert.

Anstatt der einzelnen Lautsprecherkanäle werden die Audioobjekte einer *ADM*-Datei als Inputsignale festgelegt. Da diese Audiosignale aber nicht den Signalen einer gerenderten Datei entsprechen, sind einige Anpassungen der Signale unter Berücksichtigung der Metadaten vorzunehmen.

Untersuchungen bezüglich der Objektparameter *extent*, *divergence* und *diffuse* ergaben leichte Abweichungen der Lautheit einer Audioszene. Da jedoch unterschiedliche Werte in Abhängigkeit des gegebenen Wiedergabesystems entstehen und die Abweichungen sich in einem tolerierbaren Wertebereich befinden, werden diese Parameter im Rahmen dieses Messverfahrens nicht berücksichtigt.

Komponenten, deren Einbeziehung notwendig ist, sind die Objektparameter *Position* und *Gain*, sowie die *programme id*, sofern mehrere Programme vorhanden sind, und die *comp id*, falls zueinander komplementäre Objekte deklariert sind. Des Weiteren sind Audioobjekte, die LFE-Signale enthalten, zu identifizieren. Diese entsprechen effektiv einem LFE-Kanal und sollen nach Vorgabe der ITU-R BS.1770-4 keinen Einfluss auf den Lautheitswert nehmen.

Das entwickelte Messverfahren ist in Abbildung 6.6 veranschaulicht.

Die selektierten Objekte werden durch Einlesen des *ADMs* in der Klasse *openBw64ADM* und Nutzung der Methode *select_rendering_items* ermittelt. Bei Angabe einer *programme id* oder einer *comp id* gibt diese Methode nur die zum Rendern verwendeten Objekte und deren Eigenschaften zurück. Durch Iteration über alle Audioblöcke mithilfe der Methode *get_next_block* und Abfrage der Frequenzdaten werden außerdem die LFE-Signale identifiziert. Anschließend werden Audiosamples der nicht ausgewählten Objekte und der LFE-Signale auf null gesetzt.

Danach erfolgt die Modifizierung der Objektsignale in Abhängigkeit ihres *gain*-Wertes. Mithilfe der Methode *get_next_block* wird über alle Audioblöcke iteriert und deren *gain*-Parameter überprüft. Liegt ein *gain*-Wert $\neq 1$ vor, so werden Startzeit, Dauer und Kanalnummer des Objekts bestimmt und die Audiosamples innerhalb des entsprechenden Zeitintervalls mit dem *gain*-Wert multipliziert. Nach Anpassung der Objektsignale durchlaufen diese die von Rec. ITU-R 1770-4 vorgegebenen Module *Filterung*, *Mean Square*, *positionsabhängige Gewichtung*, *Summation* und *Gating*.

Die positionsabhängige Gewichtung der Audioblöcke erfolgte durch die Ermittlung der Objektpositionen aus den Metadaten. Befindet sich ein Audioobjekt im Gewichtungsbereich, so wird eine Multiplikation der entsprechenden Leistungsblöcke mit dem Faktor 1.41 vorgenommen.

Interpolationen bei der Bewegung eines Audioobjekts wurden hierbei nicht berücksichtigt, da diese bezüglich der Lautheit vernachlässigbar sind.

Kartesische Positionsangaben wurden in polare Koordinaten umgerechnet, um zu prüfen, ob sich das entsprechende Objekt im Gewichtungsbereich befindet.

Darüber hinaus unterstützt das Messverfahren auch *ADM*-Szenen mit Audio-Elementen des Typs *DirectSpeaker*. Elemente des Typs *DirectSpeaker* besitzen statische Positionskoordinaten. Diese können dem Gewichtungsverfahren, analog zu den Elementen des Typs *Objects*, hinzugefügt werden.

6.5. Evaluierung des entwickelten Verfahrens

Im Rahmen dieser Arbeit wurden mehrere Faktoren identifiziert, welche die Lautheit einer Audioszene beeinflussen können. Welches Ausmaß diese Abweichungen der Lautheit aber innerhalb von objektbasierten Audioszenen annehmen, ist nicht bekannt. Aus diesem Grund werden zur Evaluation und Bewertung des entwickelten Verfahrens Lautheitsmessungen von objektbasiert produzierten Audioszenen durchgeführt.

Zur Prüfung des entwickelten Lautheitsmessverfahrens standen OBA-Szenen mit verschiedenen Inhalten zur Verfügung, darunter Hörspiele, TV-Mitschnitte und Musikstücke. Von diesen Audioszenen erstellte die Autorin dieser Arbeit mithilfe der *ear-render*-Kommandozeilenmethode für jede der unterstützten Lautsprecheranordnungen kanalbasierte Dateien. Daraufhin wurden die Lautheitswerte der gerenderten Dateien ermittelt und die Differenzen dieser Messwerte zu dem objektbasiert gemessenen Lautheitswert der zugehörigen *ADM*-Szene berechnet.

Unter anderem sollte überprüft werden, ob die Lautheitsabweichungen, welche durch Nutzung der Parameter *extent*, *objectDivergence* und *diffuse* zustande kommen, in der Praxis relevant sind. Da keine der vorhandenen Audioszenen diffuse oder divergierte Objekte enthielt, wurde eine eigene Szene erstellt und Lautheitsmessungen dieser Szene vorgenommen.

Nach der messtechnischen Evaluation wurde die Plausibilität der Ergebnisse geprüft. Hierzu wurde ein informeller und auf subjektiver Empfindung beruhender Hörtest im 3D-Audio-Labor des IRT durchgeführt.

6.5.1. Messmaterial

Die objektbasierten Audioszenen, die zur Verfügung standen, können aufgrund ihrer unterschiedlichen Inhalte und Charakteristiken in folgende vier Gruppen eingeteilt werden:

Hörspiele, TV-Übertragungen, Musik und Testsignale.

Die Hörspiele *Family*, *Protest* und *Forest* wurden im Rahmen des S3A-Projekts der BBC produziert. Der Schwerpunkt dieser Produktionen lag darauf, ein immersives Erlebnis für den Hörer zu realisieren. Hierzu wurden einige Audioobjekte seitlich oder hinter dem Hörer positioniert, es wurden mehrere Elevationsebenen eingesetzt und viel mit Bewegung gearbeitet[28].

Weiterhin produzierte die BBC das interaktive Hörspiel *The Mermaids Tears*. Hier hat der Hörer die Möglichkeit, die Handlung aus den Perspektiven verschiedener Protagonisten zu verfolgen. Der Wechsel zwischen den Perspektiven wird durch die Definition mehrerer Audioprogramme im ADM ermöglicht. Im Gegensatz zu den Hörspielen des S3A-Projekts spielt sich bei *The Mermaids Tears* das Geschehen ausschließlich in Blickrichtung des Hörers und auf nur einer Höherebene ab.

Die Messgruppe *TV-Übertragungen* besteht aus Aufnahmen der Leichtathletik Europameisterschaft 2018 und des Eurovision Song Contests aus dem Jahre 2018.

Zur internationalen Ausstrahlung sind jeweils Moderationen auf mehreren Sprachen vorhanden, welche als Mono-Objekte vorliegen. Durch Auswahl des Audioprogramms kann die gewünschte Sprache eingestellt werden. Weiterhin liegt ein 4+7+0-Kanalbett vor, das eine bestehende Mischung der Live-Signale vor Ort enthält[7].

Die Messgruppe *Musik* umfasst ein elektronisches Musikstück, sowie zwei klassische Musikproduktionen.

Unter den *Testsignalen* befinden sich Signale unterschiedlicher Art, welche zu Testzwecken erstellt wurden und somit nur eingeschränkte Praxisrelevanz besitzen. Beispiele sind eine Applaus-Szene bestehend aus acht Audioobjekten, die um den Hörer herum positioniert sind, und ADM-Szenen mit einzelnen positions-dynamischen Objekten, die beispielsweise Sprachsignale oder Soundeffekte enthalten.

In den beschriebenen Audioszenen wurden zum Teil mehrere Audioprogramme oder komplementäre Objekte verwendet und die ADM-Parameter *gain* und *extent* eingesetzt. Als *diffuse*-Werte wurden nur die Werte 0 und 1 angewandt, *extent* wurde kaum eingesetzt und *divergence* wurde gar nicht verwendet. Eine weitere Prüfung sollte die Auswirkungen der Parameter *divergence* und *diffuse* untersuchen. Hierfür erstellte die Autorin eine eigene Audioszene. Diese Szene bestand aus 34 Audioobjekten mit Sprachsignalen, Musik, Geräuschen und einer Hintergrundatmosphäre und war strukturiert wie ein kurzes Hörspiel. Diese Audioszene wurde in vier Versionen erstellt. Eine der Szenen beinhaltete für 14 der 34 Objekte einen *diffuse*-Wert zwischen 0 und 1. Die Werte 0 und 1 wurden nicht eingesetzt, da die Untersuchungen in Kapitel 6.3.5 zeigen, dass diese *diffuse*-Werte keine Lautheitsabweichungen verursachen. Eine weitere Ausführung der Audioszene enthielt elf ausgedehnte und zwei divergierte Audioobjekte.

Darüber hinaus wurde eine Szene unter Einsatz aller genannten Parameter (*diffuse*, *extent* und

divergence) sowie eine Szene ohne diese Parameter angefertigt. Letztere diente dazu, Vergleiche zwischen der Lautheit einer Szene mit bzw. ohne die zu untersuchenden Objektparameter anstellen zu können.

6.5.2. Messtechnische Evaluierung

Zur messtechnischen Evaluierung des entwickelten Messverfahrens wurden die objektbasierte Lautheit jeder Audioszene sowie die kanalbasierten Lautheitswerte für alle Lautsprecheranordnungen mit Ausnahme der 2.0-Stereoanordnung berechnet und die Differenzen ausgewertet.

Da der *Polar Point Source Panner* des *EAR* beim Rendering auf ein 2.0-Wiedergabesystem nicht lautheitserhaltend arbeitet, kommt es immer zu Lautheitsabweichungen (bis zu -3.2 LU), welche keine Aussage über die Qualität des objektbasierten Messverfahrens liefern können. Aufgrund dessen wurden die Lautheitswerte der 2.0-Lautsprecheranordnung in der folgenden Auswertung nicht berücksichtigt.

Abbildung 6.7 veranschaulicht die Differenzen der objektbasiert ermittelten Lautheitswerte zu den kanalbasiert gemessenen Lautheiten aller Audioszenen.

Die Lautheitswerte werden durch sogenannte Boxplots repräsentiert. Hierbei zeigt jeweils der obere Whisker den Maximalwert und der untere Whisker den Minimalwert an. Zwischen den oberen und unteren Rändern der Boxen befinden sich 50 Prozent der Lautheitswerte und die horizontale Linie innerhalb einer Box zeigt den jeweiligen Median an.

Es ist zu erkennen, dass die kanalbasiert gemessenen Lautheitswerte der gerenderten Dateien um

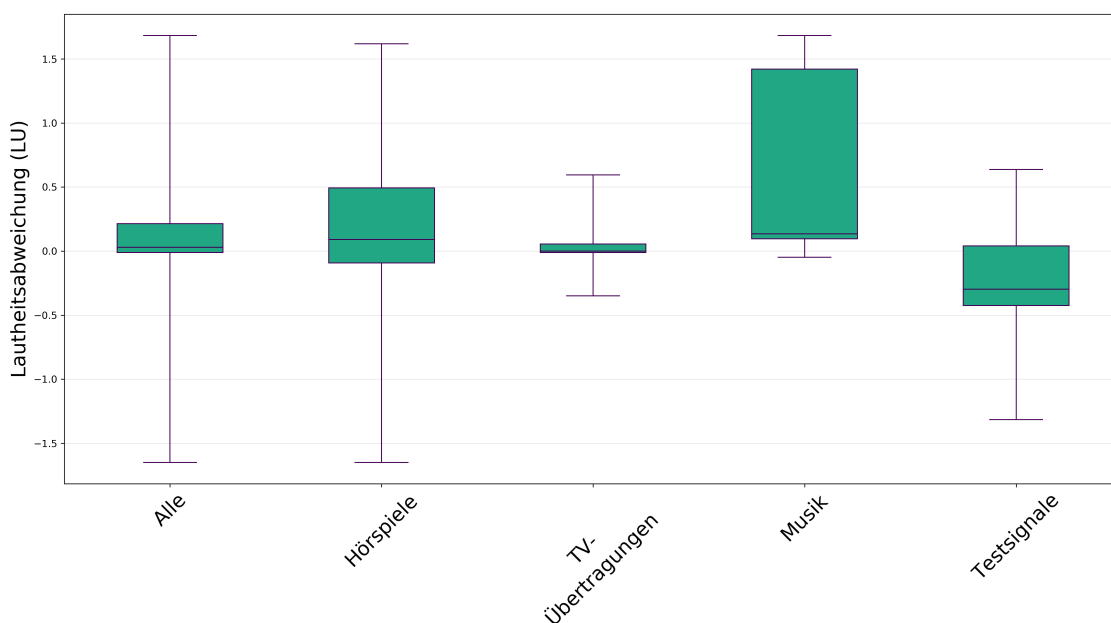


Abbildung 6.7.: Boxplots der Lautheitswerte für alle Audioszenen

bis zu 1.6 LU vom objektbasiert gemessenen Wert abweichen. Diese Differenzen hängen zum einen mit der Amplitudenaddition kohärenter Audioobjekte und zum anderen mit der positionsabhängigen Gewichtung zusammen.

In Kapitel Abschnitt 6.2 wurde dargelegt, dass kohärente Audiosignale bei Wiedergabe über denselben Lautsprecherkanal zu Abweichungen der Gesamtlautheit führen. Diese Differenzen äußern sich bei einer geringen Anzahl an Lautsprechern besonders stark.

Wie in Abschnitt 6.3 dargelegt wurde, bewirkt das Gewichtungsmodul der kanalbasierten Lautheitsmessung je nach Position eines Audioobjekts bei einer geringen Lautsprecheranzahl einen höheren Lautheitswert als bei einer Wiedergabe über eine größere Anzahl an Lautsprechern.

Die höchste im Rahmen dieser Arbeit gemessene Differenz zwischen den Lautheitswerten aller Kanalrepräsentationen einer Audioszene wurde beim immersiven Hörspiel *Protest* festgestellt und beträgt 1.4 LU.

Bei den Audioszenen der Messgruppe *TV-Übertragungen* sind die geringsten Lautheitsdifferenzen vorhanden. Dies ist darauf zurückzuführen, dass die *ADM*-Szenen hier jeweils aus einem 4+7+0-Kanalbett mit Objekten der *typeDefinition DirectSpeaker* und einem Kommentar-Objekt aufgebaut sind. Durch die bereits bestehende Mischung im 4+7+0-Kanalbett, spielt die Kohärenz einzelner Objekte zumindest bei Lautsprecheranordnungen, welche eine Wiedergabe über die vorgesehenen Lautsprecher ermöglichen, keine Rolle.

Aufgrund der erhöhten Lautheitsabweichungen in den Messgruppen *Hörspiele* und *Musik*, wurden diese näher betrachtet.

In Abbildung 6.8 ist zu erkennen, dass immersive Hörspiele deutlich höhere Lautheitsabweichungen aufweisen als nicht-immersive Hörspiele.

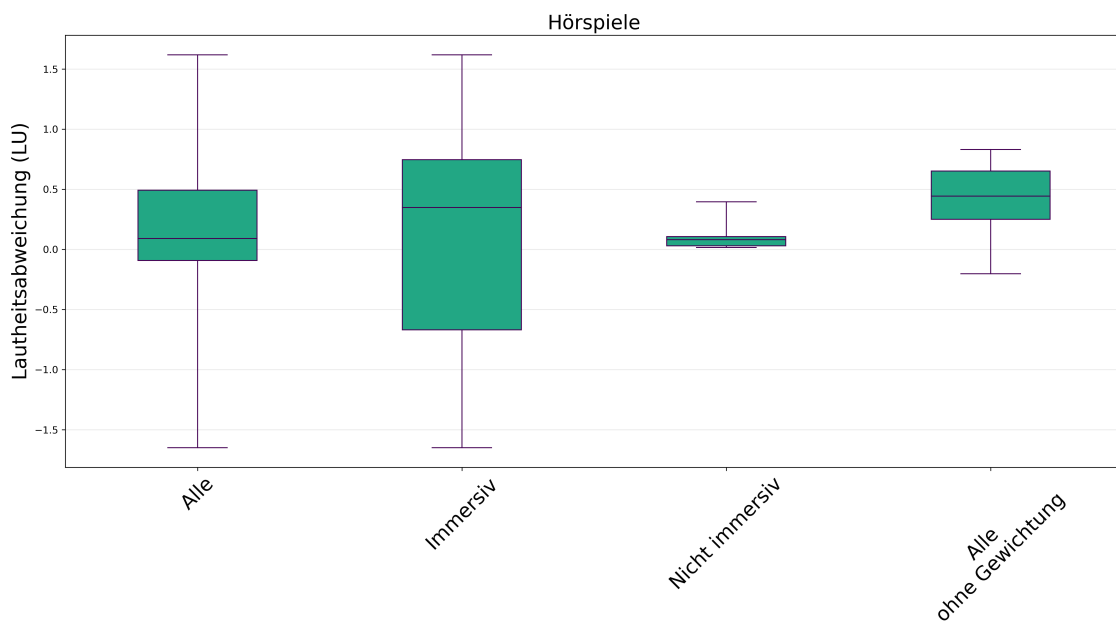


Abbildung 6.8.: Lautheitswerte der Gruppe *Hörspiele*

Ein Grund hierfür ist das Gewichtungsmodul der Lautheitsmessung.

Bei immersiven Hörspielen befinden sich einige Audioobjekte seitlich und hinter dem Hörer und werden daher bei der Lautheitsmessung gewichtet. Dies bewirkt Differenzen der Lautheitswerte je nach Lautsprecheranordnung. Um erkenntlich zu machen, inwieweit die gemessenen Lautheitsdifferenzen zwischen den Lautsprecheranordnungen mit der Gewichtung bestimmter Objektpositionen zusammenhängt, wurde zudem eine Messung ohne das Gewichtungsmodul durchgeführt. Dabei wurden weder bei der objektbasierten Messung noch bei den kanalbasierten Messungen eine richtungsabhängige Gewichtung der Leistungsblöcke vorgenommen.

Der Boxplot rechts in Abbildung 6.8 zeigt die Abweichungen der kanalbasiert gemessenen Lautheitswerte von der Lautheit der *ADM*-Szene bei Auslassung des Gewichtungsmoduls. Im Falle der gemessenen Hörspiele sinkt die maximale Lautheitsabweichung von 1.6 LU auf 0.8 LU. Es ist anzunehmen, dass die verbleibenden Lautheitsdifferenzen mit der Amplitudenaddition von kohärenten Audioobjekten zusammenhängen.

Die Messgruppe *Musik* kann in klassische und elektronische Musik unterteilt werden. Abbildung 6.9 zeigt eine Diskrepanz im Verhalten der Lautheitswerte zwischen den beiden Genres. Beim elektronischen Musikstück weichen die kanalbasiert gemessenen Lautheitswerte deutlich vom ermittelten Lautheitswert der *ADM*-Szene ab. Die Differenz beträgt je nach Lautsprecheranordnung zwischen 1.18 LU (bei $9+10+3$) und 1.68 LU (bei $0+5+0$). Dagegen beträgt die höchste Abweichung bei der klassischen Musik nur 0.45 LU.

Dieser Sachverhalt steht in Zusammenhang mit Lautheitsabweichungen durch korrelierende Audio-

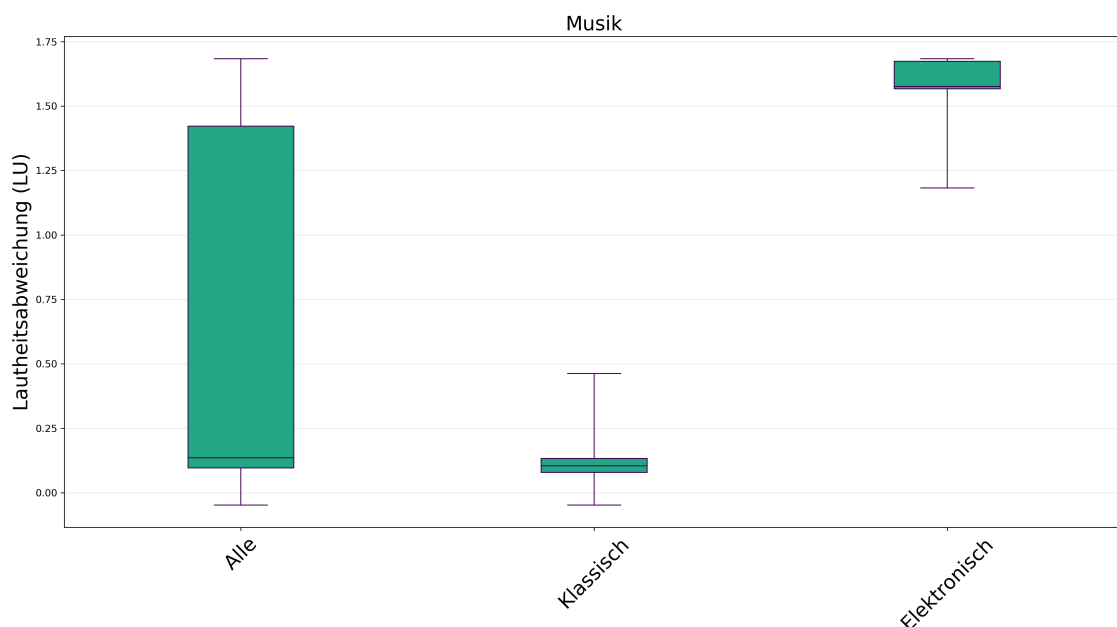


Abbildung 6.9.: Lautheitswerte der Gruppe *Musik*

signale. Aufgrund der digitalen Signalgenerierung im Rahmen der Produktion eines elektronischen Musikstücks ist anzunehmen, dass Kohärenzen zwischen Audioobjekten häufiger vorkommen und der Korrelationsgrad höher ist, als bei einer Aufnahme von analogen Signalen mit Mikrofonen.

Im Rahmen der Lautheitsmessung vorhandener Audioszenen konnte bei Szenen, welche den Objektparameter *extent* beinhalteten, keine vermehrte Lautheitsabweichung festgestellt werden. Aufgrund der begrenzten Verwendung des *extent*-Parameters wurde dieser, sowie die Objektparameter *divergence* und *diffuse* durch die Erstellung einer Audioszene, welche strukturell an ein Hörspiel angelehnt ist, näher untersucht. In Abbildung 6.10 sind die Lautheitswerte der besagten Szene veranschaulicht. Hierbei wurde festgestellt, dass ausgedehnte oder divergierte Objekte in der Praxis keinen Einfluss auf die Lautheit einer Audioszene nehmen. Die Anwendung des Parameters *diffuse* auf einige Audioobjekte der Szene bewirkte eine leichte Lautheitsdifferenz um bis zu 0.2 LU im Vergleich zur Lautheit der Szene ohne Parameter. Diese Differenz in der Lautheit ist wegen ihres geringen Ausmaßes in der Praxis nicht relevant.

Weiterhin ist anzumerken, dass mit dieser eigens erstellten Audioszene ein extremer Einsatz der Parameter erfolgte. Es ist anzunehmen, dass bei zweckmäßigen Produktionen kaum *diffuse*-Werte ungleich 0 oder 1 verwendet werden, sodass die Lautheitsabweichungen durch diesen Objektparameter eher noch geringer sind als im hier angeführten Beispiel.

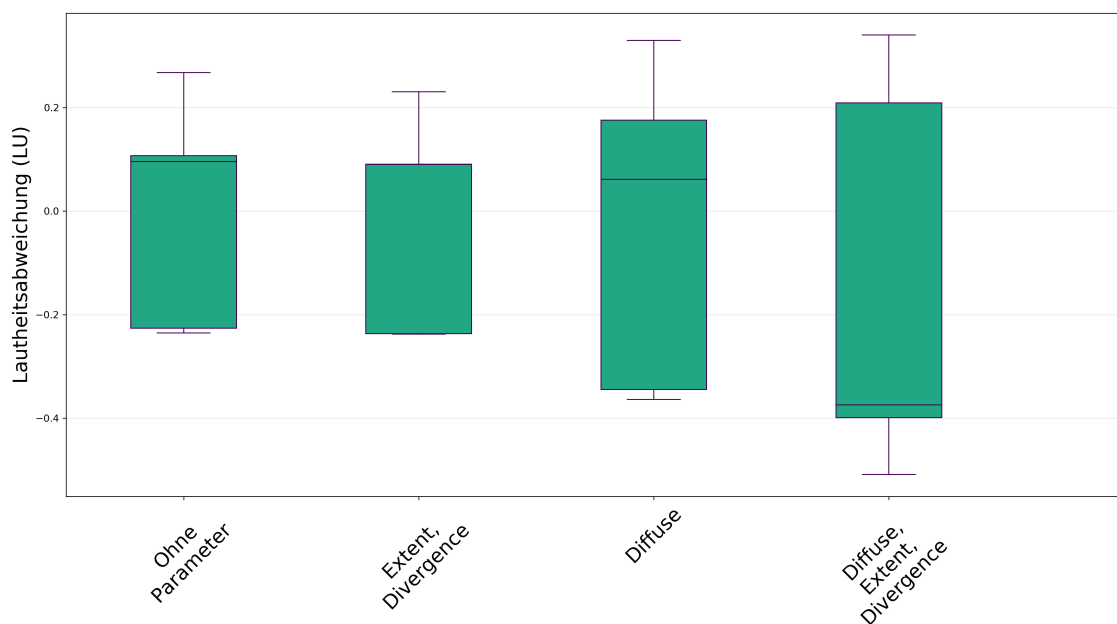


Abbildung 6.10.: Lautheitswerte der selbst erstellten Audioszene in Abhängigkeit unterschiedlicher Objektparameter

6.5.3. Plausibilitätsprüfung des entwickelten Verfahrens

Ergänzend zur messtechnischen Evaluation des entwickelten objektbasierten Lautheitsmessverfahrens wurde im Audiolabor des IRT ein subjektiver Hörtest durchgeführt.

Die Intention dieses Tests war die Plausibilitätsprüfung der messtechnisch erlangten Erkenntnisse. Dabei wurde kein Anspruch auf Wissenschaftlichkeit gestellt, da aufwändige Hörtests, welche wissenschaftlich haltbare Ergebnisse liefern, den Rahmen dieser Arbeit sprengen würden.

Ergebnisse dieses Versuchs dienen rein zur Information, sie können aber auf eventuelle Problematiken hinweisen und spielen daher im Rahmen dieser Arbeit eine essenzielle Rolle.

Weiterhin sollte in Erfahrung gebracht werden, ob die Lautheitsunterschiede einer Szene bei verschiedenen Lautsprecheranordnungen in der Praxis tolerierbar sind. Nach Soulodres[10] Studie sind bei gleichem Programmmaterial Pegeldifferenzen ab 0.4 dB möglicherweise wahrnehmbar. Diese Studie bezog sich jedoch nicht auf die empfundene Lautheit bei unterschiedlichen Lautsprecheranordnungen. Um dies zu erforschen, wurden diejenigen Audioszenen geprüft, welche bei den Messungen die größten Lautheitsdifferenzen aufzeigten.

Versuchsaufbau und Durchführung

Alle im Rahmen dieses Versuchs genutzten Audioszenen wurden mithilfe des entwickelten Lautheitsmessverfahrens auf einen objektbasierten Lautheitswert von -23 LU normalisiert. Anschließend wurde mithilfe des *ear-render*- Befehls für jede Lautsprecheranordnung eine Kanalrepräsentation der normalisierten *ADM*-Szene erstellt und die kanalbasierten Lautheitswerte aller Dateien ermittelt. Als Testszenen wurden die immersiven Hörspiele *Protest* und *Family* gewählt, da bei diesen Szenen die höchsten Lautheitsdifferenzen zwischen den gerenderten Dateien für verschiedene Lautsprecheranordnungen festgestellt wurde. Die maximalen Lautheitsunterschiede bestanden zwischen den Lautsprecheranordnungen *0+5+0* und *9+10+3*. Diese Maximalwerte betrugen 1.4 LU (*Protest*) und 1.2 LU (*Family*).

Audioszenen mit geringeren Abweichungen wurden nicht berücksichtigt, da derartig geringe Differenzen der Lautheit für den Menschen kaum wahrnehmbar sind.

Die gerenderten Dateien einer Audioszene wurden in der DAW *Reaper* sample-genau untereinander auf einzelnen Spuren angelegt, sodass durch den Solo-Button jederzeit zwischen den verschiedenen Lautsprecheranordnungen gewechselt werden konnte. Somit konnten die Lautheitsdifferenzen zwischen zwei Lautsprecheranordnungen subjektiv evaluiert werden.

Die Durchführung des subjektiven Tests fand im 3D-Audiolabor des IRT statt. Dieses unterstützt alle nach Rec. ITU-R BS.2051 definierten Lautsprecheranordnungen. Die Testpersonen waren Mitarbeiter der Audioabteilung und können daher als *professionelle Hörer* eingestuft werden. Aussagen über die Lautheitsempfindungen wurden mündlich getroffen, die Testpersonen standen in Austausch miteinander und die Äußerungen der Probanden wurden protokolliert.

Weiterhin wurde die subjektive Lautheitsempfindung bei konkreter Anwendung der Parameter *extent* und *diffuse* geprüft. Hierzu wurde eine Audioszene mit acht Audioobjekten, welche Applaus-Signale enthielten und um den Hörer herum positioniert waren, und eine Szene bestehend aus Sprachsignalen verwendet. Von diesen Szenen wurden jeweils drei Versionen erstellt: eine Version ohne Objektparameter, eine zweite Version mit einem *diffuse*-Wert von 0.5 auf allen Objekten und eine dritte Version, bei der alle Objekte die *extent*-Werte *width* 40.0 *height* 40.0 und *depth* 0.3 besaßen. Durch die Anordnung auf einzelnen Spuren der DAW untereinander wurden subjektive Lautheitsvergleiche zwischen den verschiedenen Versionen einer Audioszene ermöglicht. Die kanalbasiert gemessenen Lautheitswerte ergaben beim Vergleich zwischen Audioszenen ohne bzw. mit den untersuchten Parametern lediglich Lautheitsunterschiede bis zu 0.1 LU.

Die einzig signifikante Lautheitsabweichung wurde bei der 2.0-Lautsprecheranordnung festgestellt. Die Nutzung des *extent*-Parameters verursachte hier eine Erhöhung der Lautheit um 1.7 LU.

Ergebnisse

Beim subjektiven Vergleich der Lautheit zwischen verschiedenen Lautsprecheranordnungen war die erste Erkenntnis der meisten Testpersonen, dass ihnen die Stereo-Version der Audioszenen generell leiser vorkam als alle anderen Lautsprecher-Layouts. Es bestand die Möglichkeit, den Pegel der Stereo-Version durch bewegen des Faders der entsprechenden Spur anzugleichen. Die Höhe der Pegelangleichung auf die Lautheit des 5.1-Lautsprecheranordnung betrug beim Hörspiel *Protest* je nach Testperson zwischen 0.75 dB und 2.0 dB. Die gemessenen Lautheitsdifferenz betrug 2.1 LU. Die abweichende Lautheit des Stereo-Layouts wurde demnach subjektiv erkannt, die Höhe der Differenz wurde allerdings insgesamt eher geringer eingeschätzt als der gemessene Wert angibt. Die Wahrnehmung weiterer Lautheitsdifferenzen zwischen den Lautsprecheranordnungen fiel zunächst allen Probanden schwer. Sie behelfen sich damit, sehr kurze Abschnitte immer wieder zu hören, oder sie wählten einen Ausschnitt der Audioszene, bei dem sehr viel passiert und die Lautheit über die Länge des Ausschnitts hinweg weitgehend konstant bleibt. Nur auf diese Weise konnten weitere Lautheitsunterschiede festgestellt werden.

Im Verlauf des Hörtests wurde von einigen Testpersonen die Vermutung angestellt, dass die 5.1-Version der Audioszenen lauter sein könnte als die 22.2-Version. Die Testpersonen waren sich dabei aber nicht sicher und meinten, dass eine mögliche Lautheitsdifferenz, falls es eine geben sollte, sehr gering und kaum wahrnehmbar sei.

Die Beurteilung fiel aufgrund wechselnder Objektpositionen bei verschiedenen Lautsprecheranordnungen schwer. Außerdem unterschied sich der Klang der Audioszene je nach verwendetem Lautsprecher-Layout. Das 5.1-System wurde unter anderem als *dröhnend* und *unangenehm* beschrieben, während die größeren Lautsprecheranordnungen einen subjektiv angenehmeren Klang produzierten.

Weiterhin wurde angemerkt, dass die Hörerposition im Raum bezüglich der empfundenen Lautheit

eine wichtige Rolle spielt. Befindet sich ein Hörer beispielsweise im hinteren Teil des Raums, so kommen ihm die Lautsprecheranordnungen, welche nur die Lautsprecher direkt hinter ihm benutzen, lauter vor als solche, die Höhenlautsprecher über der Hörerposition mitverwenden.

Beim Vergleich von Audioszenen in Abhängigkeit der Nutzung von ausgedehnten oder diffusen Objekten bestätigen die Lautheitsempfindungen der Probanden die gemessenen Lautheitswerte.

Der *diffuse*-Parameter verursacht eine Klangveränderung, aufgrund derer ein Teil der Probanden keine sicher Aussage über die Lautheit treffen konnte. Tendenziell wurde jedoch bei keinem der Parameter ein Lautheitsunterschied wahrgenommen. Lediglich bei der 2.0-Stereoanordnung, bei dem der *extent*-Parameter tatsächlich einen erhöhten Lautheitswert verursacht, wurde dies auch im subjektiven Test erkannt.

Insgesamt sagten alle Teilnehmer, dass die Lautheitsunterschiede kaum wahrnehmbar seien und sich daher in einem tolerierbaren Bereich befänden. Lediglich die Lautheitsabweichungen der 2.0-Stereoanordnung sei im Vergleich zu den anderen Lautsprecheranordnungen deutlich erkennbar und stelle demnach eine relevante Problematik dar.

6.5.4. Auswertung der messtechnischen und subjektiven Evaluierung

Bei Betrachtung der Medianwerte der kanalbasierten Lautheiten fällt auf, dass diese durchgehend näherungsweise dem objektbasierten Lautheitswert entsprechen. Da ein generisches Messverfahren keine Abweichungen berücksichtigen kann, die mit den verwendeten Wiedergabesystemen zusammenhängen, sind auch höhere Abweichungen nicht auszuschließen. Im Rahmen der untersuchten Audioszenen lag die höchste Differenz zwischen objektbasierter Lautheit und einem der kanalbasiert gemessenen Lautheitswerte bei 1.68 LU. Die Höhe der Lautheitsdifferenzen hängt von den Eigenschaften der vorliegenden Audioszene ab. Die Objektpositionen haben durch die richtungsabhängige Gewichtung bei der Lautheitsmessung einen Einfluss auf die Lautheit einer Audioszene und Kohärenzen zwischen einzelnen Audioobjekten bewirken durch die Amplitudenaddition beim Rendering eine Abweichung der Lautheit.

Der subjektive Hörtest deutet jedoch an, dass die Lautheitsunterschiede zwischen verschiedenen Lautsprecheranordnungen selbst in extremen Fällen und für geübte Hörer subjektiv kaum wahrnehmbar sind. Die Ergebnisse des informellen Tests zeigten insgesamt, dass das entwickelte generische Lautheitsmessverfahren für objektbasierte Audioszenen durchaus plausible Messwerte liefert.

Beschriebene Lautheitsabweichungen in Abhängigkeit des Wiedergabesystems sind den Testergebnissen zufolge tolerierbar. Lediglich die Lautheitsdifferenzen bei Wiedergabe auf einem Stereo-System, die aufgrund der *Downmix*-Funktion des *Polar Point Source Panners* entstehen, befinden sich in einem deutlich wahrnehmbaren Wertebereich. Hier wurden messtechnisch Abweichungen zwischen -1.5 LU und -3.2 LU festgestellt und laut Aussagen der Testpersonen wurden die Lautheitsdifferenzen zwischen dem 2.0-Lautsprechersystem und den anderen Lautsprecheranordnungen in den meisten Fällen auch subjektiv erkannt.

Um jedoch wissenschaftlich haltbare Aussagen über den subjektiven Lautheitseindruck treffen zu können, wäre die Durchführung einer repräsentativen Studie mit umfangreichen Hörversuchen notwendig.

7. Zusammenfassung und Ausblick

Ziel dieser Thesis war die Entwicklung eines objektbasierten Lautheitsmessverfahrens auf Basis bereits bestehender Ansätze. Das Messverfahren wurde in Zusammenhang mit dem *EBU ADM Renderer* entwickelt. Dabei bestand die Anforderung einer möglichst hohen Kompatibilität zum kanalbasierten Messverfahren nach Rec. ITU-R 1770-4. Da der Einfluss der *ADM*-Parameter auf die Lautheit einer Audioszene bislang noch unerforscht war, wurde dieser mit entsprechenden Messungen untersucht.

Die Auswertung der Messwerte ergab geringe Lautheitsabweichungen bei Anwendung der Objektparameter *extent*, *divergence* und *diffuse*. Die Abweichungen variieren in Abhängigkeit des Wiedergabesystems, befinden sich aber insgesamt in einem akzeptablen Wertebereich.

Zur Implementierung eines objektbasierten Lautheitsmessverfahrens bedarf es aus diesem Grund keiner Beachtung der untersuchten Objektparameter. Einzig der Parameter *gain* beeinflusst den Signalpegel der Audioobjekte und wurde deshalb im entwickelten Messverfahren entsprechend berücksichtigt.

Im nächsten Schritt folgte die messtechnische Evaluation des entwickelten Messverfahrens. Dazu wurden bestehende Audioszenen auf alle Lautsprecheranordnungen gerendert und die kanalbasierten Lautheitswerte sowie der objektbasierte Lautheitswert der *ADM*-Szene berechnet. Es wurden Vergleiche angestellt und die Differenzwerte wurden in Form von Boxplots verbildlicht (siehe Kapitel 6.5.2). Abschließend wurde ein informeller Hörversuch durchgeführt, welcher der Plausibilitätsprüfung und subjektiven Evaluation diente.

Im Zuge der messtechnischen Untersuchungen wurde eine Problematik bezüglich des Stereo-Renderings einer objektbasierten Audioszene aufgedeckt. Der *Polar Point Source Panner* nutzt beim Rendering auf Stereo eine *Downmixing*-Funktion und arbeitet aufgrund dessen nicht leistungserhaltend. Daraus folgen Lautheitsdifferenzen bei der Stereo-Version einer Audioszene im Vergleich zu allen anderen Lautsprecheranordnungen (bis zu -3.2 LU). Aus Sicht der Autorin hat zur Lösung dieses Problems eine Änderung im *Polar Point Source Panner* des Renderers zu erfolgen. Im Zuge der *Downmixing*-Funktion sollte eine Leistungsanpassung durchgeführt werden, sodass die Gesamtleistung konstant bleibt und somit eine einheitliche Lautheit zwischen den Lautsprecheranordnungen gewährleistet werden kann.

Im Gegensatz zum Rendering mit polaren Koordinaten erfolgt das Rendering zur Stereo-Wiedergabe bei kartesischen Positionsangaben leistungserhaltend, da hier kein *Downmixing* stattfindet. Dies

kann zu Unterschieden im Mischungsverhältnis in Abhängigkeit des genutzten Koordinatensystems führen. Die Relevanz der Klangunterschiede, welche durch die verschiedenen Panning-Algorithmen hervorgerufen werden, sollte näher untersucht werden. Gegebenenfalls sind im Renderer entsprechende Anpassungen einzubinden.

Aufgrund der Lautheitsabweichungen beim 2.0-Rendering kann ein generisches Lautheitsmessverfahren für die Stereo-Wiedergabe bei polaren Positionsangaben aktuell keine akkuraten Lautheitswerte liefern. Hinsichtlich aller anderen Lautsprecheranordnungen erzielte das entwickelte Messverfahren aber zufriedenstellende Ergebnisse. Lautheitsunterschiede zwischen einzelnen kanalbasiert gemessenen Lautheitswerten und dem objektbasierten Messwert können durch die in Kapitel 6.2 beschriebene Amplitudenaddition durch kohärente Audioobjekte entstehen. Diese kann bei einer geringen Lautsprecheranzahl (wie z.B. bei Wiedergabe auf $0+5+0$) zu einer Lautheitsabweichung führen. Bei einer höheren Lautsprecheranzahl verursachen kohärente Audioobjekte hingegen nur geringe Lautheitsabweichungen.

Die Amplitudenaddition kohärenter Audioobjekte im Rendering-Prozess wirkt sich aber nicht nur auf die Lautheit aus, sondern kann außerdem ungewollte Klangverfärbungen verursachen. Aus Sicht der Autorin ist daher eine Vermeidung der Amplitudenaddition korrelierender Audiosignale während des Renderings anzustreben.

Zudem deuten die Ergebnisse des durchgeführten Hörtests darauf hin, dass diese Lautheitsdifferenzen subjektiv schwer wahrnehmbar sind und daher toleriert werden können. Allerdings ist anzumerken, dass zur Erlangung wissenschaftlich haltbarer Erkenntnisse, weitere Untersuchungen einschließlich einer repräsentativen Studie notwendig sind.

Weiterhin wurde bewiesen, dass die Differenzen zwischen kanalbasierten Lautheitswerten und dem objektbasierten Lautheitswert der zugehörigen *ADM*-Szene auch mit dem Gewichtungsmodule der Lautheitsmessung zusammenhängen können. Je nach Material besteht die Möglichkeit, dass beispielsweise beim Rendering auf eine 5.1-Lautsprecheranordnung mehr Signalanteile auf Lautsprecherkanäle gelangen, welche aufgrund ihrer Raumposition bei der Lautheitsmessung gewichtet werden, als beim Rendering auf eine höhere Lautsprecheranzahl (z.B. $9+10+3$). Grund hierfür ist die Tatsache, dass beim 5.1-Rendering auch Objektpositionen außerhalb des Gewichtungsbereichs auf die Lautsprecher *Ls* und *Rs* gerendert werden. Für eine akkurate Positionswiedergabe stehen zu wenige Lautsprecherpositionen zur Verfügung. Bei nicht immersiven Produktionen spielt die Gewichtung kaum eine Rolle. Im Rahmen der messtechnischen Evaluation wurde ein durchschnittlicher Differenzwert von ± 0.08 LU und ein maximaler Differenzwert von $+0.39$ LU ermittelt. Lediglich bei Verwendung entsprechender Objektpositionen entstehen Lautheitsunterschiede (durchschnittlich ± 0.85 LU und max. -1.65 LU).

Im Durchschnitt beträgt die Differenz zwischen der Lautheit einer *ADM*-Szene und den kanalbasiert gemessenen Lautheitswerten der gerenderten Szenen ± 0.33 LU. Insgesamt ergaben Messungen

und ein subjektiver Hörversuch im Zuge der Evaluation des entwickelten Messverfahrens keine kritischen Lautheitsabweichungen. Die Lautheitsdifferenzen zwischen verschiedenen Lautsprecheranordnungen (mit Ausnahme von $0+2+0$) sind akzeptabel.

Als Fazit ist festzuhalten, dass im Rahmen dieser Arbeit ein generisches Verfahren zur Lautheitsmessung von objektbasierten Audioszenen entwickelt wurde, welches zufriedenstellende Ergebnisse liefert. Bei immersiven Audioszenen und korrelierenden Audiosignalen einzelner Objekte sind Lautheitsabweichungen möglich, die jedoch toleriert werden können. Mit dem entwickelten Messverfahren kann demnach ein Richtwert zur Lautheitskontrolle einer objektbasierten Audioszene generiert werden. Zur Erstellung objektbasierter Audioproduktionen, wird derzeit an der Weiterentwicklung von DAWs gearbeitet. Auf Grundlage des entwickelten Verfahrens können Programmerweiterungen zur Lautheitsmessung von objektbasierten Audioszenen für diese DAWs erstellt werden. Weiterhin besteht die Möglichkeit, den Lautheitswert einer Audioszene in den Metadaten der *ADM*-Szene einzuspeichern. Damit ist die Lautheitsberechnung nur einmal im Produktionsablauf notwendig. Es ist allerdings noch zu bedenken, dass der objektbasierte Lautheitswert bislang nicht für das 2.0-Rendering gilt. Daher wird für Stereo-Renderings eine kanalbasierte Lautheitsmessung empfohlen. Sobald aber Anpassungen im *EAR* implementiert werden, die ein lautheitserhaltendes Stereo-Panning ermöglichen, kann das objektbasierte Lautheitsmessverfahren auch für die Stereo-Wiedergabe gelten. Zu gegebenem Zeitpunkt sollte dies durch geeignete Messungen bestätigt werden.

A. Auszug aus Recommendation ITU-R BS.2051

TABLE 1

**List of possible loudspeaker positions for advanced sound system, identification of the loudspeaker layouts
in form of “Upper + Middle + Bottom loudspeakers”**

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H	I	J
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0
M+000	0	0		X	X	X	X	X	X	X	X	X
M+022	+22.5	0										
M-022	−22.5	0										
M+SC	Left edge of display	0							X			
M-SC	Right edge of display	0							X			
M+030	+30	0	X	X	X	X	X	X	X	X	X	X
M-030	−30	0	X	X	X	X	X	X	X	X	X	X
M+045	+45	0										
M-045	−45	0										
M+060	+60	0								X		
M-060	−60	0								X		
M+090	+90	0						X	X	X	X	X
M-090	−90	0						X	X	X	X	X
M+110	+110	0		X	X	X	X					
M-110	−110	0		X	X	X	X					
M+135	+135	0						X	X	X	X	X
M-135	−135	0						X	X	X	X	X
M+180	+180	0								X		
U+000	0	+30								X		
U+022	+22.5	+30										
U-022	−22.5	+30										
U+030	+30	+30			X	X	X					
U-030	−30	+30			X	X	X					

TABLE 1 (continued)

[illegible]

TABLE 1 (*end*)

SP Label	Azimuth	Elevation	A	B	C	D	E	F	G	H	I	J
			0+2+0	0+5+0	2+5+0	4+5+0	4+5+1	3+7+0	4+9+0	9+10+3	0+7+0	4+7+0
B-135	−135	−30										
B+180	+180	−30										
LFE1	+45	−30		X	X	X	X	X	X	X	X	X
LFE2	−45	−30						X		X		

NOTE 1 – The loudspeakers are assumed to be on a sphere. When this is not the case they should preferably be time aligned (at the central listening position) with an accuracy of 100 µs.

NOTE 2 – The reproduction configurations marked with “*” should be used with audio-related metadata for productions of advanced sound programmes.

The loudspeaker layouts shown in Table 1 are illustrated in Table 2.

B. Messtechnische Überprüfung der ADM-Parameter in Bezug auf die Lautheit einer Audioszene

B.1. Lautheitsabweichungen in Verbindung mit *extent* und *divergence*

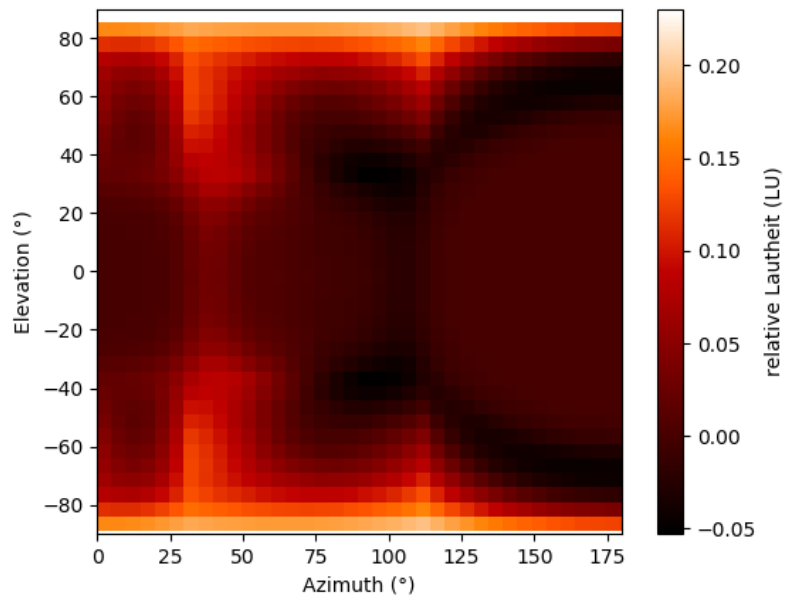


Abbildung B.1.: Lautheitsabweichungen für ein Rauschobjekt mit den *extent*-Werten *height* = 40.0 und *width* = 40.0 bei Rendering auf 0+5+0

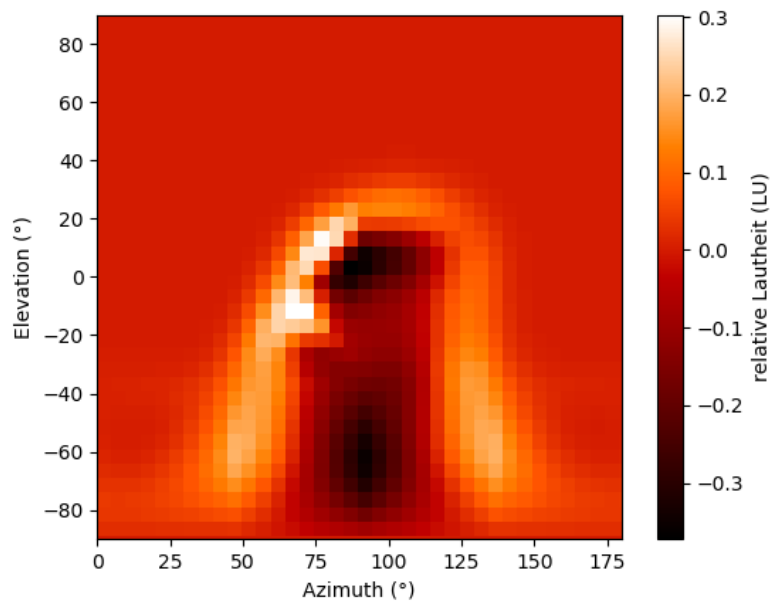


Abbildung B.2.: Lautheitsabweichungen für ein Rauschobjekt mit den *extent*-Werten *height* = 40.0 und *width* = 40.0 bei Rendering auf 9+10+3

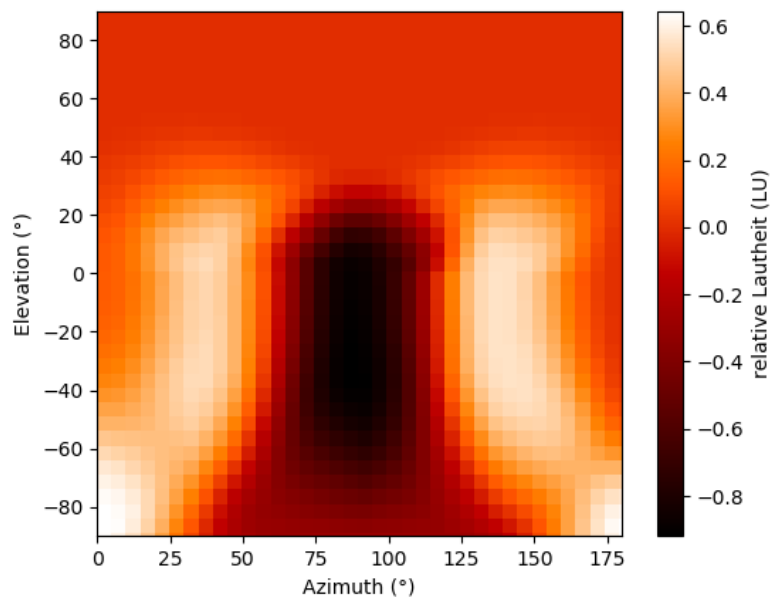


Abbildung B.3.: Lautheitsabweichungen für ein Rauschobjekt mit den *divergence*-Werten *value* = 0.5 und *azimuthRange* = 45.0° bei Rendering auf 4+7+0

B.2. Lautheitsabweichungen in Verbindung mit *diffuse*

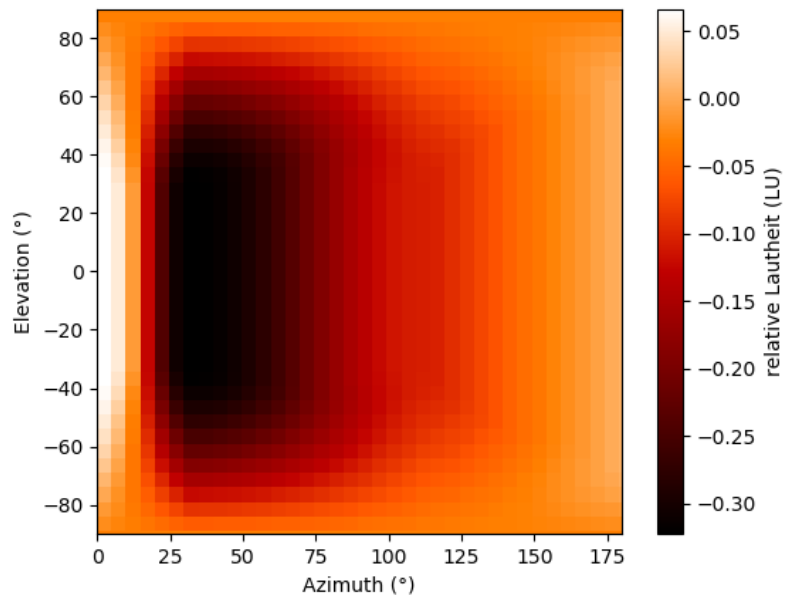


Abbildung B.4.: Lautheitsabweichungen für ein Rauschobjekt mit *diffuse* = 0.5 bei Rendering auf 0+5+0

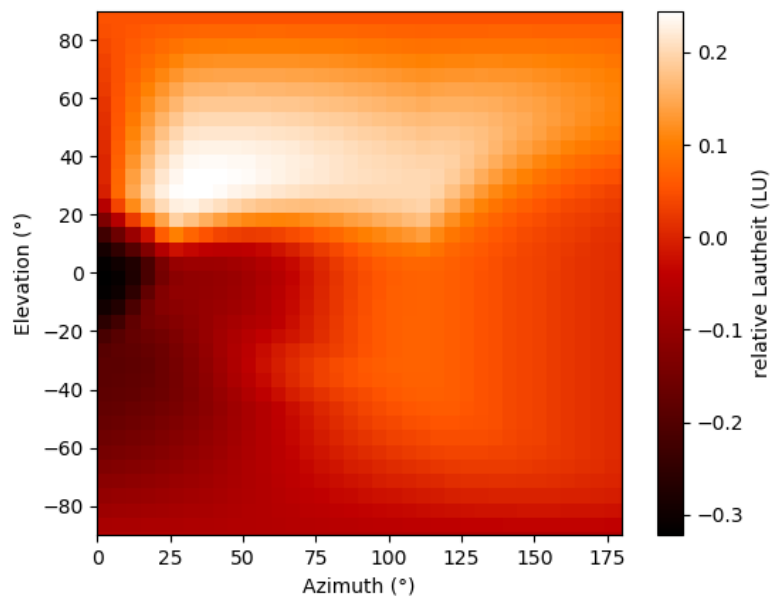


Abbildung B.5.: Lautheitsabweichungen für ein Rauschobjekt mit $diffuse = 0.5$ bei Rendering auf $4+5+1$

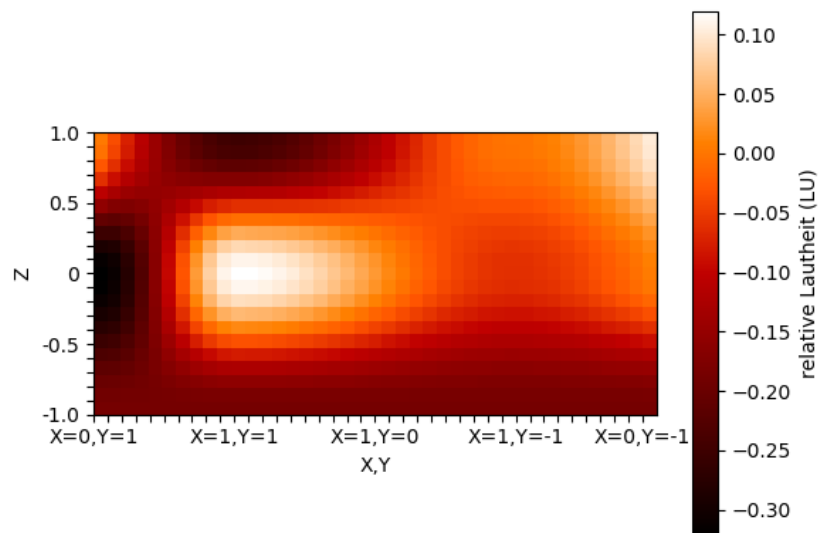


Abbildung B.6.: Lautheitsabweichungen für ein Rauschobjekt mit kartesischen Positionsangaben und $diffuse = 0.5$ bei Rendering auf $0+5+0$

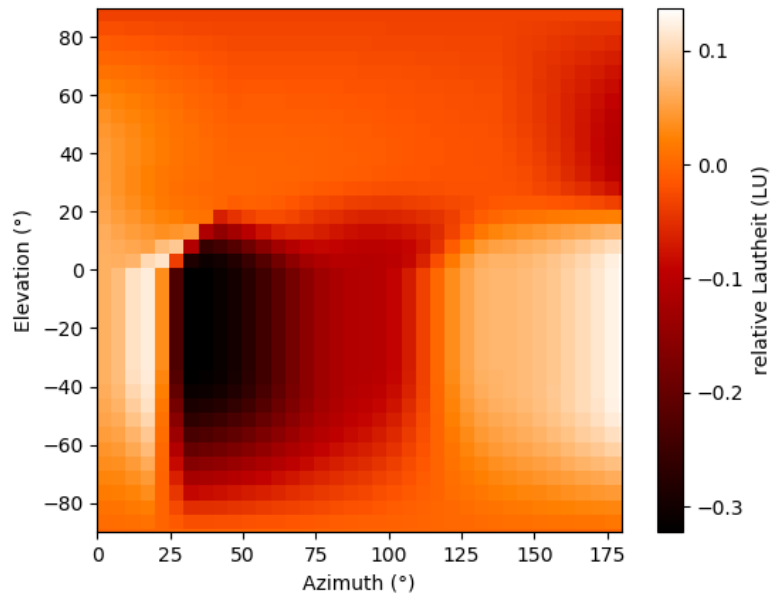


Abbildung B.7.: Lautheitsabweichungen für ein Rauschobjekt mit *diffuse* = 0.5 bei Rendering auf 4+9+0

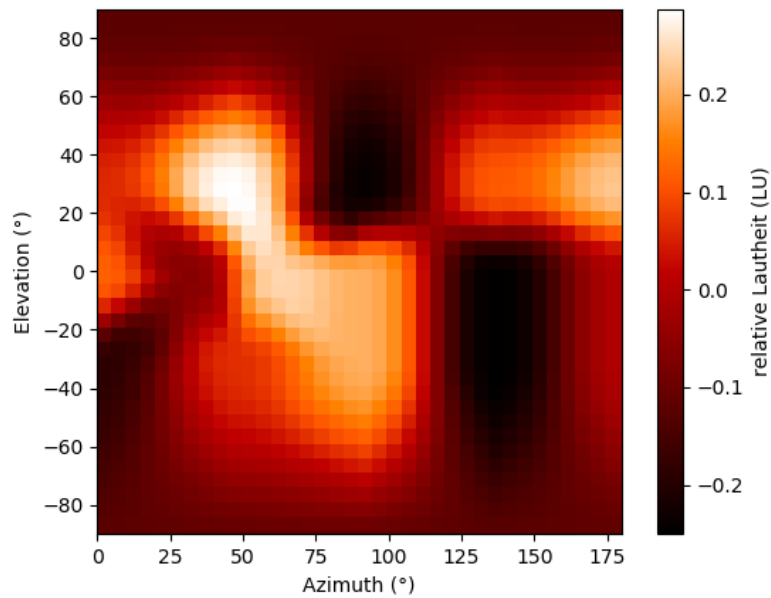


Abbildung B.8.: Lautheitsabweichungen für ein Rauschobjekt mit *diffuse* = 0.5 bei Rendering auf 9+10+3

C. Messtechnische Evaluation des entwickelten Verfahrens

C.1. Differenzwerte aller Messgruppen

	Alle Szenen [LU]	Hörspiele [LU]	TV-Übertragungen [LU]	Musik [LU]	Testsignale [LU]
Minimum	-1.65	-1.65	-0.35	-0.05	-1.3
Maximum	1.68	1.62	0.59	1.68	0.64
Unteres Quartil	-0.01	-0.09	-0.01	0.10	-0.42
Oberes Quartil	0.21	0.49	0.05	1.42	0.04
Median	0.03	0.09	0.00	0.13	0.29
Arith. Mittelwert	0.15	0.08	0.03	0.62	-0.22
Standartabweichung	0.55	0.72	0.12	0.68	0.52

C.2. Differenzwerte der Messgruppe *Hörspiele*

	Alle Hörspiele [LU]	Immersive Hörspiele [LU]	Nicht immersive Hörspiele [LU]	Alle ohne Gew. [LU]
Minimum	-1.65	-1.65	0.02	-0.20
Maximum	1.62	1.62	0.39	0.83
Unteres Quartil	-0.09	-0.66	0.03	0.25
Oberes Quartil	0.49	0.74	0.11	0.65
Median	0.09	0.35	0.08	0.44
Arith. Mittelwert	0.08	0.09	0.08	0.44
Standartabweichung	0.72	0.96	0.07	0.28

C.3. Differenzwerte der Messgruppe *Musik*

	Alle Musikstücke [LU]	Klassische Musik [LU]	Elektronische Musik [LU]
Minimum	-0.05	-0.05	1.18
Maximum	1.68	0.46	1.68
Unteres Quartil	0.10	0.07	1.56
Oberes Quartil	1.42	0.13	1.67
Median	0.13	0.10	1.57
Arith. Mittelwert	0.62	0.16	1.54
Standartabweichung	0.68	0.16	0.19

C.4. Differenzwerte der selbst erstellten Audioszene in Abhängigkeit der Objektparameter

	Ohne Parameter [LU]	Extent, Divergence [LU]	Diffuse [LU]	Extent, Divergence, Diffuse [LU]
Minimum	-0.24	-0.24	-0.36	-0.51
Maximum	0.26	0.23	0.33	0.34
Unteres Quartil	-0.23	-0.24	-0.34	-0.40
Oberes Quartil	0.11	0.09	0.17	0.21
Median	0.09	0.09	0.06	-0.37
Arith. Mittelwert	-0.02	-0.04	-0.04	-0.15
Standartabweichung	0.20	0.19	0.28	0.34

Literaturverzeichnis

- [1] M. Dickreiter, V. Dittel, W. Hoeg, M. W. (Hrsg.) *Handbuch der Tonstudientechnik*. Gruyter, de Saur, 14. Nov. 2013. 1511 S. ISBN: 3110289784 (zitiert auf S. 3, 11).
- [2] K. C. P. F. Alton Everest. *Master Handbook of Acoustics*. McGraw-Hill Education Ltd, 1. Jan. 2015. ISBN: 0071841040 (zitiert auf S. 3, 4).
- [3] H. Fastl, E. Zwicker. *Psychoacoustics: Facts and Models*. Springer, 2007. ISBN: 978-3-540-68888-4 (zitiert auf S. 4, 11).
- [4] M. Frank. „Localization Using Different Amplitude-Panning Methods in the Frontal Horizontal Plane“. In: (Jan. 2014) (zitiert auf S. 12).
- [5] D. Griesinger. „Stereo and Surround Panning in Practice“. In: *Audio Engineering Society Convention 112*. Apr. 2002 (zitiert auf S. 12).
- [6] T. Holman. *Surround Sound, Second Edition: Up and running*. Focal Press, 2007. ISBN: 9780240808291 (zitiert auf S. 12).
- [7] F. de Jong, D. Driesnack, A. Manson, M. Parmentier, P. Sunna, S. Thomson. *European Athletics Championships: Lessons from a Live, HDR, HFR, UHD and Next-Generation Audio sports event*. Paper. European Broadcasting Union (EBU), 25. Okt. 2018 (zitiert auf S. 43).
- [8] A. Mason, T. Herberger, E. Habets, M. Geier, A. Silzle, M. Volpel, N. Epain, N. Bogards, M. Meier. *D3.5: Specification and implementation of reference audio processing for use in content creation and consumption based on novel broadcast quality standards*. Report. Orpheus, 31. Mai 2017 (zitiert auf S. 29, 30).
- [9] S. Mehrgardt, V. Mellert. „Transformation characteristics of the external human ear“. In: *The Journal of the Acoustical Society of America* 61.6 (1977), S. 1567–1576 (zitiert auf S. 6).
- [10] S. G. Norcross, G. A. Soulodre, M. C. Lavoie. „The Subjective Loudness of Typical Program Material“. In: *Audio Engineering Society Convention 115*. Okt. 2003 (zitiert auf S. 39, 48).
- [11] N. Peters, D. Sen, M.-Y. Kim, O. Wuebbolt, M. Weiss. „Scene-based Audio Implemented with Higher Order Ambisonics (HOA)“. In: *SMPTE 2015 Annual Technical Conference and Exhibition*. 2015 (zitiert auf S. 10).
- [12] V. Pulkki. „Virtual Sound Source Positioning Using Vector Base Amplitude Panning, June 1997“. In: *Vol.45, No. 6* (1. Juni 1997) (zitiert auf S. 12–15).

- [13] E. Sengpiel. *Summenpegel von kohärenten Signalen*. Hrsg. von A. Sengpiel. 1. Juli 2003. URL: <https://www.sengpielaudio.com/Rechner-kohquellen.htm> (besucht am 08. 10. 2019) (zitiert auf S. 31).
- [14] E. B. Union. *EBU R128 - Loudness normalisation and permitted maximum level of audio signals*. Report. EBU, 1. Juni 2014 (zitiert auf S. 5).
- [15] E. B. Union. *EBU Tech3343 - Guidelines for production of programmes in accordance with EBU R128*. Report. EBU, 1. Jan. 2016 (zitiert auf S. 5).
- [16] E. B. Union. *EBU Tech 3388 - ADM Renderer for Use in Next Generation Audio Broadcasting*. Report. EBU, 1. Mai 2018 (zitiert auf S. 20).
- [17] E. B. Union. *R 128 s1 - Loudness Parameters for Short-form (Adverts; Promos, etc.)* Report. EBU, 1. Jan. 2016 (zitiert auf S. 6).
- [18] E. B. Union. *TR 042 - Example of an End-to-end OBA Broadcast Architecture and Workflow*. Report. EBU, 1. Mai 2018 (zitiert auf S. 10, 17).
- [19] I. T. Union. *ITU-R BS. 1770-4 Algorithms to measure audio programme loudness and true-peak audio level*. Report. ITU, 1. Okt. 2015 (zitiert auf S. 6, 7).
- [20] I. T. Union. *BWF (Recommendation ITU-R BR.1352) Metadata concerns*. Report. ITU, 30. Apr. 2004 (zitiert auf S. 17).
- [21] I. T. Union. *Recommendation ITU-R BS.2076-1*. Report. ITU, 1. Juni 2017 (zitiert auf S. 17–21).
- [22] I. T. Union. *Recommendation ITU-R BS.2127-0 - Audio Definition Model renderer for advanced sound systems*. Report. ITU, 1. Juni 2019 (zitiert auf S. 18, 21–23, 25–27).
- [23] I. T. Union. *Recommendation ITU-R BS.2051-2 - Advanced Sound System for Programme Production*. Report. ITU, 1. Juli 2018 (zitiert auf S. 10, 15, 20, 21, 31).
- [24] I. T. Union. *Recommendation ITU-R BS.2088-0 - Long-form file format for the international exchange of audio programme materials with metadata*. Report. ITU, 1. Okt. 2015 (zitiert auf S. 17).
- [25] I. T. Union. *Recommendation ITU-R BS.775-3 - Multichannel stereophonic sound system with and without accompanying picture*. Report. ITU, 1. Aug. 2012 (zitiert auf S. 11, 15).
- [26] S. Weinzierl. *Handbuch der Audiotechnik*. 1. Aufl. VDI-Buch. Springer-Verlag Berlin Heidelberg, 2008. ISBN: 9783540343004 (zitiert auf S. 3, 4, 11, 14).
- [27] M. Weitnauer, C. Baume, A. Silzle, N. Färber. *Final Reference Architecture Specification and Integration Report*. Report. ORPHEUS - Object-Based Audio Experience, 7. März 2018 (zitiert auf S. 29).

- [28] J. Woodcock, C. Pike, F. Melchior, P. Coleman, A. Franck, A. Hilton. „Presenting the S3A Object-Based Audio Drama Dataset“. In: *Audio Engineering Society Convention 140*. Mai 2016 (zitiert auf S. 43).