Technische Universität Berlin
Fachgebiet Audiokommunikation

Magisterarbeit

Vergleichende Untersuchung live aufgeführter und medial reproduzierter Musik mittels einer immersiven Hörversuchsumgebung

vorgelegt von Michael Horn

[redacted]

Abgabe: 4. November 2013

Erstgutachter: Prof. Dr. Stefan Weinzierl
Zweitgutachter: Alexander Lindau, M.A.

Hiermit versichere ich gegenüber der Fakultät I der Technischen Universität Berlin, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Ausführungen, die anderen veröffentlichten oder nicht veröffentlichten Schriften wörtlich oder sinngemäß entnommen wurden, habe ich kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Fassung noch keiner anderen Prüfungsbehörde vorgelegen.

Berlin, den 4. November 2013 _____

Michael Horn

Inhaltsverzeichnis

1. Motivation und Inhalt der Arbeit **1**

2. Stand der Forschung **5**

2.1. Einschlägige Vorstudien zu medial rezipierter Musik 5
2.2. Versuchsdesign 8
2.3. Methoden der Erhebung deskriptiver Vokabulare 10
2.4. Grundlagen des räumlichen Hörens 11
2.5. Dynamische Binauralsynthese 11
2.5.1. Richtcharakteristik der Messlautsprecher 15
2.5.2. Individualisierung der BRIRs 16
2.5.3. Datensatzdiskretisierung 18
2.5.4. Ansatz zur effizienten Darstellung späten Nachhalls - perceptual mixing time 18
2.5.5. Systemlatenz 19
2.5.6. Wiedergabesystem 20
2.5.7. Qualitätskriterien für virtuelle akustische Umgebungen 21
2.5.8. Zusammenfassung 22
2.6. Aufnahmenbasierte Synthese akustischer Szenen 24
2.7. Binaurale Datenakquise 25
2.8. Räumliches Sehen und stereoskopische Projektion 26
2.8.1. Anforderungen an eine optische Simulation 29
2.8.2. Stereoskopische Aufnahmesysteme 30
2.8.3. Stereoskopische Wiedergabesysteme 32
2.8.4. Zusammenfassung 34
2.9. Raum- und Stimulusauswahl 34
2.10. Zusammenfassung 37

3. Audiosimulation **41**
3.1. Erzeugung der musikalischen Inhalte 41
3.2. Erhebung und Aufbereitung der Daten für binaurale Raumsimulationen 47
3.2.1. Messung Tonstudio 51
3.2.2. Messung WFS-Labor 52
3.2.3. Messung Konzertsaal 54
3.2.4. Kopfhörersimulation 55
3.2.5. Pseudobinaurale Hintergrundgeräusche 56
3.3. Akustische Simulationsumgebung 58
3.3.1. Parametrierung der Softwaremodule der Audiosimulation 59
3.3.2. Kopfhörer BK211 68
3.3.3. Akustische Kalibrierung 68

4. Videosimulation **75**
4.1. Bildaufnahmen (konventionell und 3D) 75
4.1.1. Parameter der Bildaufnahmen 75
4.1.2. Durchführung der Bildaufnahmen 79
4.2. Bildbearbeitung 79
4.3. Optische Simulationsumgebung 82
4.3.1. Geometriekalibrierung 86
4.3.2. Farb- Helligkeits- und Dynamikkalibrierung 88

5. AV-Umgebung **93**
5.1. Cavebeschreibung 93
5.2. Synchronisierung 93

6. Methoden - Hörversuch **101**
6.1. Triadengenerierung und Stichprobenbetrachtung 101
6.2. Umsetzung 104

7. Resultate **109**

8. Zusammenfassung und Ausblick **115**

9. Literaturverzeichnis **117**

A. Technikspezifikationen des verwendeten Equipments und Messaufbauten 123

B. Pluginparameter der MTB-Entrauschung 131

1. Motivation und Inhalt der Arbeit

„Warum hinterläßt ein eindrucksvolles Konzert meist stärkere Spuren als eine Schallplatte? Weil der Hörer etwas Leibhaftiges erlebt hat“[1]

Der heutige Musikkonsum ist überwiegend medial bestimmt und sein Verhältnis zum Livekonzert unklar. Die Medienwahrnehmungsforschung interessiert sich daher besonders für den Unterschied zwischen live aufgeführter und medial produziert und wiedergegebener Musik.

Typische Thesen bisheriger Arbeiten lauten:

- Die Klanggestaltung medialer Musik beeinflusst das Gefallensurteil [1].
- Abhängig von einer rein akustischen oder audiovisuellen Darbietung von medialer Musik werden unterschiedliche Klanggestaltungen präferiert [1].
- Die mediale Produktion ist bestrebt den fehlenden Bildeindruck durch ein real eigentlich nicht existierendes Klangbild zu kompensieren, 'artifizielle Natürlichkeit' genannt [2].

Während die ersten beiden Thesen in Laboruntersuchungen überprüft werden konnten, entzog sich die letztere bisher eine Verifizierung.

Der Wechsel zwischen dem Hören einer Liveaufführung und dem Hören einer produzierten Aufnahme desselben Stückes kann nicht instantan erfolgen. Ebenso ändert sich hierbei im Normalfall nicht nur die Wiedergabesituation und die Nachbearbeitung, die Produktion der medialen Darbietung, sondern auch der 'musikalische Inhalt', die Einspielung. Erst durch das kontrollierte Verändern der unabhängigen Variablen und der Möglichkeit instantaner Wechsel zwischen den akustischen Szenen (Abbildung 1.1), können methodisch sicher Daten zu den Wahrnehmungsqualitätsunterschieden

[1] Alfred Brendel: „Die überlistete Einmaligkeit“, in: *Die Zeit* vom 20.1.1984, `http://www.zeit.de/1984/04/die-ueberlistete-einmaligkeit`

erhoben und Rückschlüsse auf signalverarbeitende Komponenten gezogen werden. Während Hörversuche zum Unterschied medialer Wiedergabesysteme relativ einfach zu realisieren sind, ist die Möglichkeit einen Konzertsaal plausibel in einem Labor zu synthetisieren erst in den letzten Jahren durch dynamische Binauralsynthese gegeben. Diese ermöglicht *plausible* virtuelle akustische Umgebungen, d.h. Simulationen, die gegen eine innere Referenz von der Realität nicht mehr unterscheidbar sind.

Eine musikalische Darbietung (live oder medial) erzeugt nicht nur einen akustischen Eindruck. So existieren immer auch nicht-akustische Sinneseindrücke oder es treten komplexe Prozesse wie beispielsweise soziale Interaktionen auf. Stereoskopische Videoaufnahmen der medialen und live aufgeführten Situationen und deren Visualisierung über ein dreidimensionales Projektionssystem ermöglichen heutzutage zumindest eine immersive audiovisuelle Reproduktion.

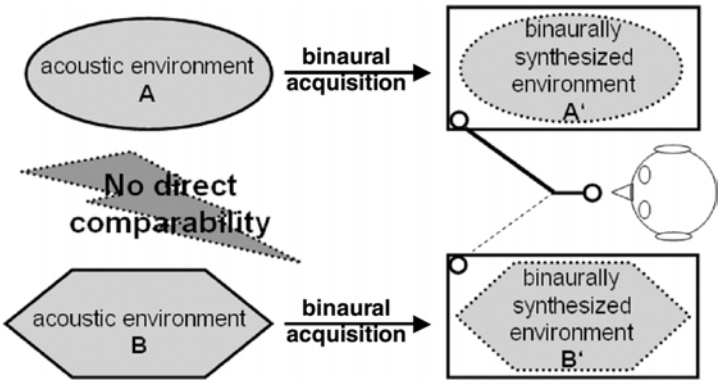


Abbildung 1.1.: Vergleich akustischer Umgebungen, aus [3, S.2]

Welche Wahrnehmungsqualitäten sollen aber abgefragt werden? Sind diese überhaupt relevant und voneinander unabhängig? Diese Fragen können mit einem qualitativen Versuch geklärt werden, der späteren systematisch-quantitativen Untersuchungen vorangestellt wird. Die Realisierung dieses qualitativen Versuchs sowie erste Ergebnisse werden in dieser Arbeit vorgestellt.

Diese ist in das DFG Forschungsprojekt *Konzertsaal und mediale Rezeption* eingebettet und beschäftigt sich primär mit der Akquise der benötigten Daten, deren Aufbereitung und der Schaffung eines möglichst realitätsnahen Simulationssystems.

Folgende Ziele sollen erreicht werden:

1. Es soll ein Messinstrument zur experimentellen Bestimmung von Wahrnehmungsunterschieden zwischen live und medial rezipierter Musik geschaffen werden.

2. Es soll ein Versuchsverfahren implementiert werden, das es ermöglicht, die gesuchten Unterschiedsqualitäten möglichst frei von Konzept-Skalen-Interaktion zu erheben. Eine hohe interne Validität wird dabei erreicht durch kontrollierte und unabhängige Variation der unabhängigen Variablen (Livekonzert vs. mediale Produktion, Art der medialen Produktion & Wiedergabe, Art des visuellen Eindrucks) unter konstanten Störvariablen (z.B. Bildrauschen, Lautheitsunterschiede, identische musikalische Einspielung). Die externe Validität soll dabei, soweit im Labor möglich, durch eine technisch optimale Wiedergabe akustischer und optischer Inhalte, deren originalgetreue Resynthetisierung und der Wahl einer geeigneten Stichprobe gesichert werden.

2. Stand der Forschung

2.1. Einschlägige Vorstudien zu medial rezipierter Musik

Jürgen Meyers Arbeit „Klangliche Unterschiede zwischen realem Orchester und der Einspielung nachhallfreier Musik in Sälen“ [4] von 1992 beschäftigt sich mit dem unterschiedlichen Höreindruck eines realen Orchesters gegenüber einer Stereo-Lautsprecherwiedergabe nachhallfreier Aufnahmen in einem Konzertsaal. Die den unterschiedlichen Höreindruck evozierenden Ursachen sollen hier kurz aufgezählt werden.

1. Während ein Orchester aus so vielen Klangquellen wie Instrumenten besteht, werden diese bei der Synthese durch nur wenige Lautsprecher repräsentiert. Dadurch werden die individuellen, positionsspezifischen Ankunftszeiten des Direktschalls und der ersten Reflexionen durch weniger komplexe Reflexionsmuster der Lautsprecher ersetzt.

2. Die Schallausbreitung von Lautsprechern wird kaum behindert, da diese sich meist auf einer leeren Bühne befinden. In einer realen Situation mit sich auf der Bühne befindlichen Musikern würde es durch Abschattung zu Pegelverlusten (vor allem zwischen 1 kHz und 3 kHz) kommen.

3. Die Schallabstrahlung von Instrumenten kann ab 500 Hz als gerichtet angenommen werden. Diese frequenzabhängigen Richtcharakteristiken bedingen frequenzabhängige Hallradien und damit frequenzabhängige Direkt-Diffusschallverhältnisse.

4. Musiker passen ihre Spielweise an die akustischen Gegebenheiten des Raumes an. Da auf bereits angefertigte nachhallfreie Aufnahmen zurückgegriffen wird, entfällt diese Rückkopplung.

5. Wurden die nachhallfreien Aufnahmen mit wenigen, nahen Hauptmikrofonen angefertigt, ist die Balance zwischen den Instrumenten unnatürlich. Die Instrumente mit der geringsten Entfernung haben einen überproportional hohen Einfluss auf den Summenpegel.

Die hier aufgezählten Ursachen können so auch auf die Binauralsynthese übertragen werden. Daher bietet es sich, an schon erste Anforderungen an die akustische Datenakquise in dieser Arbeit zu stellen.

- Es sind so viele Lautsprecher wie Quellen zu verwenden.
- Die Ausrichtung der Lautsprecher in Relation zum Hörerort ist durch Experten so anzupassen, dass ein möglichst natürlicher Höreindruck entsteht.
- Während der nachhallfreien Aufnahmen sind die Instrumente einzeln zu mikrofonieren.

Die Arbeiten von Hans-Peter Reinecke in den Jahren 1978 und 1980 [5] [6] [7] mit dem übergreifenden Thema „Musik im Original und als technische Reproduktion“ beschäftigen sich mit der auch hier gestellten Forschungsfrage nach dem Einfluss einer Livesituation auf die Musikrezeption.

In seiner Arbeit „Musik im Original und als technische Reproduktion“ [7] wird zunächst der Wandel des musikalischen Hörens aus der Sicht des Musikwissenschaftlers nachvollzogen. Bevor es technisch möglich war Schall zu speichern und zu reproduzieren, war Musik immer eine Einheit aus musikalischem Inhalt und Aufführung. Letzteres subsumiert die Interaktion mit Musiker und Publikum, die musikalische Interpretation und natürlich den visuellen Aspekt einer Liveaufführung. Durch die technische Evolution der darauffolgenden Jahrhunderte ist mittlerweile die Einheit von Ort, Zeit und handelnden Personen aufgehoben und der ursprüngliche Normalfall eines Konzertbesuchs ist heute ein Sonderfall. Reinecke zufolge gilt es zu klären, „ob das, was er [der Rezipient] durch Platte oder Band über eine hochwertige Stereoanlage hören kann, überhaupt 'das Gleiche' bleibt wie im Falle der originalen Darbietung“ und „inwieweit und ggf. in welche Richtung die unmittelbare Atmosphäre einschließlich des mimischen und pantomimischen Verhaltens der beteiligten Personen den Höreindruck beeinflusst oder gar modifiziert.“

In [5, S.19] zieht Reinecke den Vergleich zwischen dem Erleben einer Konzertaufführung, mit der ihr eigenen Spannung, Erwartung und den Publikumsgeräuschen, und

dem Anhören dieser Aufführung zu einem späteren Zeitpunkt. Werden die in der Reproduktion vorhandenen 'Artefakte' einer Liveaufnahme wie eben Publikumsgeräusche wahrgenommen, dienen Reinecke zufolge diese als Symbole für die Erlebnissituation und evozieren unter Umständen dieselbe Atmosphäre.

Er folgert daraus, dass der Wahrnehmungsunterschied nicht primär dem Objekt, dem musikalischen Inhalt, geschuldet ist, sondern einer Interaktion zwischen den an einem Konzert teilnehmenden Publikum und Musikern. Seine Hypothese lautet, „daß die Beziehungsaspekte von Situationen musikalischen Verhaltens eine erlebnismodifizierende Funktion besitzen können.“[7, S.91]

Im Rahmen der Internationalen Funkausstellung 1977 in Berlin fand eine Untersuchung statt, die von Reinecke als „Zwitter zwischen experimentellen Anspruch und lebendiger Musikdarbietung“ mit „Befunden mittlerer theoretischer Reichweite“ bezeichnet wurde. In einem 60 Personen fassenden Raumteil auf dem Messegelände wurden kleinere musikalische Darbietungen aufgeführt. Ein Teil der Besucher sollte während der Aufführung einen Fragebogen ausfüllen, um somit Daten für die 'reale' Rezeptionssituation zu generieren. Dieselben Aufführungen wurden gleichzeitig mit Mikrofonen und Kunstkopf aufgenommen und bildeten die Basis für die mediale Rezeptionssituation, wiedergegeben in einem tonstudioähnlichen Raum. Dieselben Personen sollten auch hier den gleichen Fragebogen ausfüllen und generierten so die Daten für den 'medialen' Fall.

Der Fragebogen bestand aus vorher ausgewählten, in Abbildung 2.1 aufgeführten Gegensatzpaaren, die auf einer sechsstufigen Skala bewertet werden konnten. Es wurden verschiedene Stücke mit verschiedenen Besetzungen dargeboten, unter anderem eines aus dem Biedermeier und ein eher atonales, zeitgenössisches.

Nach Auswertung der Polaritätsprofile und einer Faktorenanalyse kommt Reinecke zu dem Ergebnis, die auf das Live-Erlebnis bezogenen Urteile wiesen „fast durchweg eine emotionalere, 'lebensvollere' oder eher von Wohlklang geprägte Komponente auf, während die Wiedergabe-Urteile entweder distanzierter oder aber aggressiver ('männlicher') empfunden [...] werden“. Ebenso sei es unabhängig von der signalverarbeitenden Qualität der Wiedergabetechnik nicht möglich ein 'naturgetreues' Liveerlebnis zu reproduzieren, da die dazugehörige 'Atmosphäre' fehlt und somit das Urteil über die mediale Wiedergabe abwertender ausfällt.

Der Einfluss des musikalischen Inhalts auf die Rezipientenurteile wird am Beispiel des Biedermeier- und des zeitgenössischen Stücks gezeigt. Während das 'zutiefst vertraute'

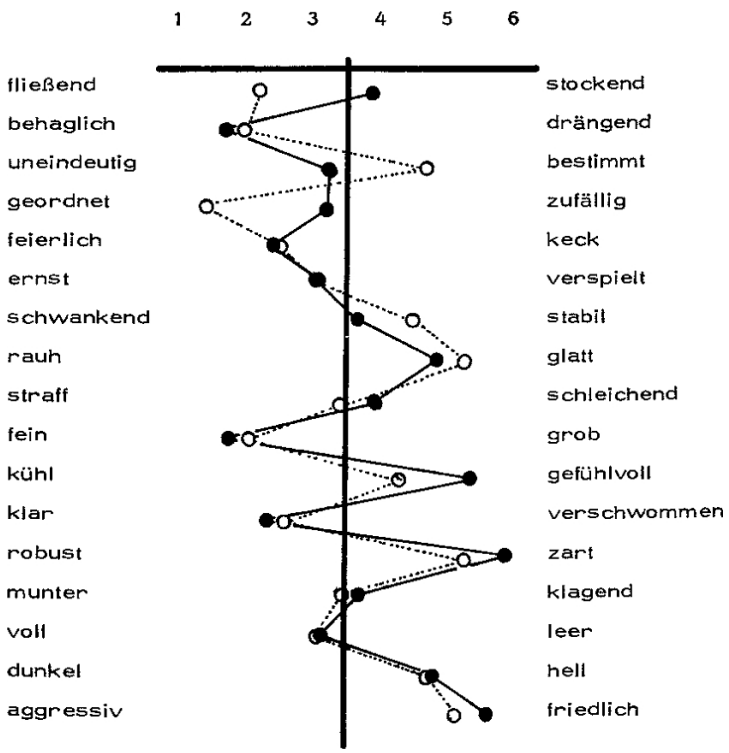


Abbildung 2.1.: „Polaritätsprofile des Musikbeispiels, gespielt auf dem Biedermeier-Reiseklavier (Reichhardt, Rondo) Durchgezogene Linie = Original, punktierte Linie = Reproduktion“[7, S.95]

Biedermeierstück ein Symbol für tief verankerte musikalische Erlebnisse sei und dadurch wohl auch in der medialen Wiedergabesituation die 'Atmosphäre' evoziert, sind die Unterschiede zwischen der realen und medialen Situation sehr gering. Die atonale Musik wurde im Gegensatz dazu in der Livesituation insgesamt als weniger fremd und atonal empfunden, vermutlich durch die Anwesenheit und Beobachtung der handelnden Personen.

2.2. Versuchsdesign

Der im vorherigen Kapitel vorgestellte Versuch von Reinecke besitzt eine sehr hohe externe Validität, lässt sich methodisch allerdings kritisch betrachten. Die Probanden bewerteten eine mediale Darbietung, deren Liveaufführung sie bereits gesehen haben. Es ist nicht ersichtlich, wie groß der zeitlich Abstand zwischen beiden Bewertungseinheiten war und inwiefern die Erinnerung an die Livedarbietung das Antwortverhalten

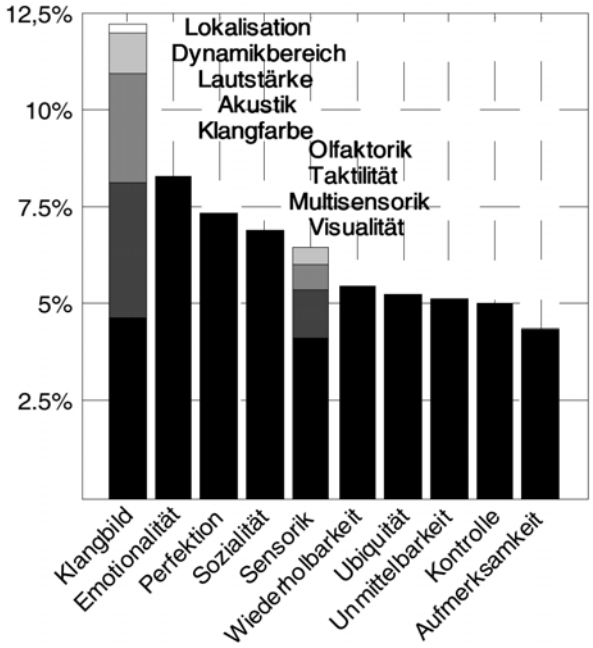


Abbildung 2.2.: Auftretungshäufigkeit der 10 wesentlichsten Hauptkategorien (Klangbildliche und sensorische sind in weitere Unterkategorien aufgeschlüsselt), aus [8]

beeinflusst hat. Ebenso ist unklar durch wen und wie die medialen Produktionen angefertigt wurden und ob sie dem Anspruch einer 'kommerziell üblichen' Produktion gerecht wurden. Insbesondere aber wurde der Fragebogen aus Gegensatzpaaren gebildet, die einer anderen Untersuchung mit anderer Forschungsfrage entnommen wurden. Beispielsweise kann das Gegensatzpaar 'voll-leer' als soziale Komponente verstanden werden, oder aber auf das Signalspektrum bezogen sein. Das Vokabular ist somit nicht frei von Konzept-Skalen-Interaktion und muss auch nicht zwangsläufig die von Versuchspersonen empfundenen Wahrnehmungsunterschiede beinhalten.

Mutmaßlich relevante Aspekte wurden in einer Studie [8] mit Hilfe einer Online-Befragung aufgedeckt. Deren Teilnehmer sollten drei inhaltliche Fragen zur Wahrnehmung von „Live-“, „Medienmusik“ und deren Unterscheidungsmerkmalen beantworten. Die 10 dabei durch Inhaltsanalyse und quantitative Auswertung am häufigsten gefundenen Kategorien sind in Abbildung 2.2 dargestellt. Sie zeigen eine „[...]breite Kombination klanglicher, emotionaler, ästhetischer, sozialer, sensorischer und kognitiver Aspekte[...]“ [8, S.2]. Die Hauptfrage ist nun, welche dieser Aspekte sich in der zu realisierenden Laborsimulation (detaillierte Beschreibung in den nachfolgenden Kapiteln) unter wel-

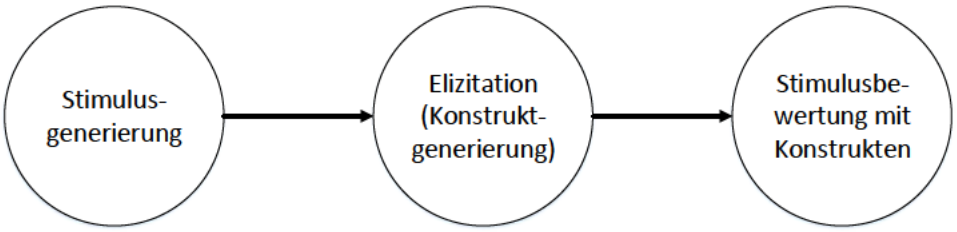


Abbildung 2.3.: RGT-Ablauf, nach [9, fig. 1]

chem Aufwand berücksichtigen lassen und welche im Hinblick auf die Forschungsfrage von Interesse sind. Die technischen Möglichkeiten erlauben eine befriedigende akustische Simulation und eine immersive optische Repräsentation, wohingegen soziale und andere sensorische Aspekte sich kaum nachbilden lassen. Inwieweit sich eine der Realität vergleichbare Emotionalität einstellt, kann nur gemutmaßt werden.

2.3. Methoden der Erhebung deskriptiver Vokabulare

Neben der Schaffung einer möglichst realistischen Simulation realer und medialer Rezeptionssituationen, ist die Erhebung von Wahrnehmungsqualitäten ebenfalls Bestandteil der Magisterarbeit. Wie bei Reinecke beispielhaft gezeigt, werden in quantitativen Untersuchungen die abzufragenden Beurteilungsskalen oftmals durch den Untersuchungsleiter, basierend auf Vorwissen aus anderen oder selbst durchgeführten Studien oder nach Literaturrecherche festgelegt. Da Qualitäten aber immer eine Interdependenz zwischen Stimulus und Versuchsperson aufweisen, kann es sein, dass einige Qualitäten übersehen werden oder anderen zu viel Bedeutung beigemessen wird. Moderne Ansätze (z.B. [9]) verwenden daher eine stimulusbasierte Erhebung individueller qualitativer Beschreiber. Durch die, auf der Konstruktpsychologie von George Kelly [10] basierende, RGT-Methodik (siehe Abb. 2.3) sollen persönliche Konstrukte der Versuchspersonen verbal erhoben werden, die die wahrgenommenen Unterschiede zwischen den Stimuli beschreiben.

Voraussetzung für eine erschöpfende Beschreibung ist die Generierung eines Stimuluspools mit einer möglichst großen Stimulusvariation.

Bei Anwenden der RGT-Methodik wird im Allgemeinen so vorgegangen, dass aus dem Pool drei verschiedene Stimuli gewählt werden, eine sogenannte Triade. Die Versuchsperson soll urteilen, welche zwei der drei Stimuli zueinander ähnlich sind und

was sie vom dritten unterscheidet. In diesem sogenannten Elizitationsschritt werden die persönlichen, bipolaren Konstrukte der Versuchsperson 'entlockt'.

An die Erhebungsphase schließt sich meist eine Bewertung aller Stimuli auf Basis des erhobenen Vokabulars an. Diese so erhobenen quantitativen Daten können zur inter- und intraindividuellen Prüfung zur Differenziertheit und Übereinstimmung des Vokabulars benutzt werden. Hinzu kommen Verfahren der Datenreduktion wie Cluster- oder Faktorenanalyse zum Einsatz. Alternativ kann die Analyse aber auch mit qalitativen Verfahren, wie beispielsweise der Inhaltsanalyse durchgeführt werden.

2.4. Grundlagen des räumlichen Hörens

Die Fähigkeit zum räumlichen Hören basiert überwiegend auf den Ohrsignalen, d.h. auf dem Schalldruckverlauf an den Trommelfellen eines Hörers. Diese unterscheiden sich in Abhängigkeit von Azimut (θ)- und Elevationswinkel (ϕ), Entfernung der Schallquelle vom Hörer (r) sowie dessen individueller Morphologie. Befindet sich die Schallquelle außerhalb der Medianebene (Abbildung 2.4), erreicht der Schall ein Ohr, wird um den Kopf gebeugt und erreicht das quellenabgewandte Ohr später. Abhängig von Einfallswinkel und individueller Kopfgröße kommt es zu einem interauralen Zeitunterschied, ITD *interaural time difference* genannt. Durch die Abschattungswirkung des Kopfes kommt es zu einem frequenzabhängigen Pegelunterschied zwischen den Ohren, bezeichnet als *interaural level difference* - ILD. Befindet sich die Quelle in der Medianebene, sind keine interauralen Pegel- und Zeitunterschiede vorhanden. Spektrale Veränderungen, verursacht durch die Form des Außenohres und Schulterreflexionen, werden zur Lokalisation ausgewertet, aber auch minimale Kopfpositionsänderung und die damit einhergehende, postionsabhängige Schallquellentranslation.

2.5. Dynamische Binauralsynthese

Die Reproduktion realer Schallfelder kann prinzipiell über einen raum- oder ohrbezogenen Ansatz erfolgen. Raumbezogene Sytheseverfahren (z.B. WFS oder Ambisonics [12]) sollen das Schallfeld hörerpositionsunabhängig möglichst korrekt innerhalb eines Wiedergaberaums wiederherstellen, sind allerdings technisch aufwändig und können Schallfelder momentan nur bis zu einer für verfärbungsfreie Wiedergabe untauglichen

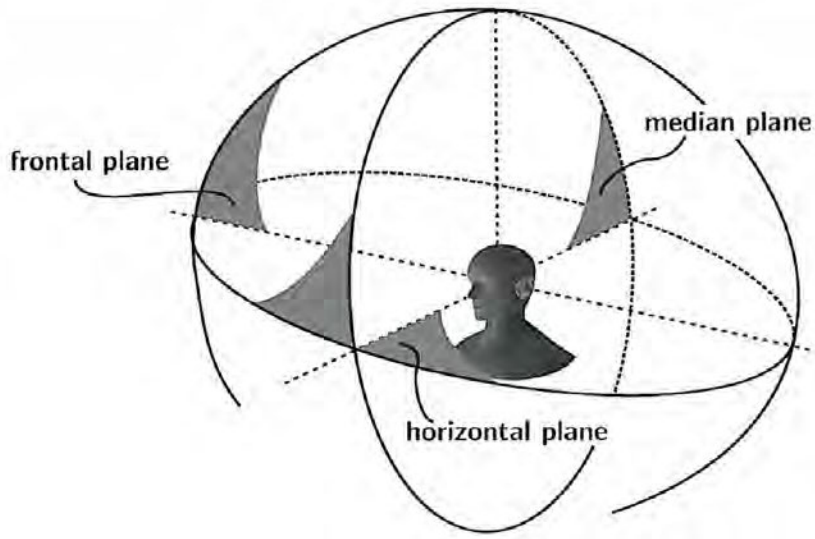


Abbildung 2.4.: Kopfbezogenes Koordinatensystem, aus [11, S.87]

niedrigen Grenzfrequenz synthetisieren.

Ohrbezogene Schallfeldsynthese nimmt als Ansatz, dass alle Informationen über den zu simulierenden Raum und die sich darin befindliche Primärquelle in den Signalen am Trommelfell enthalten sind und somit auch nur diese am Trommelfell des Hörers wiederhergestellt werden müssen. Der große Aufwand eines Lautsprechersystems entfällt hierbei, da für binaurale Synthese im Prinzip nur ein Kopfhörer benötigt wird[1].

Die Resynthese kann hierbei eine einfache Wiedergabe einer Kunstkopfaufnahme sein, oder aber die Faltung eines beliebigen Eingangssignals mit kopfbezogenen Übertragungsfunktionen. Eine im Freifeld gemessene Übertragungsfunktion, *head related transfer function - HRTF* genannt, beschreibt alle orts- und anatomieabhängigen Eigenschaften der Schallquelle und des Kunstkopfs.

Y bezeichnet das Mikrofoneingangssignal, X das Anregesignal, $j^2 = -1$ und ω die Kreisfrequenz. Im folgenden wird bei allen kopfbezogenen Funktionen nicht zwischen linkem und rechtem Ohr unterschieden, die Signale sind somit immer zweikanalig zu

[1] auch wenn lautsprecherbasierte, ohrbezogene Syntheseverfahren existieren, werden diese hier nicht näher behandelt und der Leser sei auf *transaurale Binauralsynthese* verwiesen, z.B. in W. G. Gardner (1997) „3-D Audio Using Loudspeakers.“ PhD thesis, Cambridge, MIT Media Laboratories,

verstehen.

$$HRTF(j\omega, \phi, \theta, r) = \frac{Y(j\omega)}{X(j\omega)} \tag{2.1}$$

Unter Annahme konstanter Temperatur und Druckverhältnisse in einem Raum kann von einem linearen, zeitinvarianten System ausgegangen werden und die binaurale Raumübertragungsfunktion (*binaural room transfer function - BRTF*) setzt sich zusammen aus richtungs- und frequenzabhängigen Reflexionen sowie Direktschallanteilen und den korrespondierenden HRTFs.

$$BRTF = \sum_{i=1}^{N} e^{-j\omega t_i} \omega_i HRTF_i \tag{2.2}$$

Die zurückgelegte Weglänge, Dämpfung durch materialspezifische Reflexionsfaktoren und Luftabsorbtion werden im Gewichtungsfaktor ω_i berücksichtigt. Durch eine inverse Fouriertransformation erhält man die binaurale Raumimpulsantwort (*binaural room impulse response - BRIR*), mit der beliebige Eingangssignale $x_s(t)$ gefaltet werden können.

$$BRIR = F^{-1}\{BRTF\} \tag{2.3}$$

$$y_s(t) = x_s(t) * BRIR \tag{2.4}$$

Wird das synthetisierte Signal y_s über Kopfhörer wiedergegeben, stellt sich annähernd der ursprüngliche Höreindruck der Primärquelle wieder ein. Es ist intuitiv ersichtlich, dass sich die akustische Szene bei Kopfbewegungen nicht anpassen kann und somit mitgeführt wird. Befinden sich Quellen im Bereich der Medianebene, erhält man keine Laufzeit- oder Pegeldifferenzen und wertet somit ausschließlich die hochindividuellen, spektralen Unterschiede aus. Dadurch kommt es bei solch statischer Binauralsynthese oftmals in eben diesen Bereichen zu vorne-hinten Fehllokalisationen.

Diese Einschränkungen werden überwunden, indem BRIRs für mehrere Orientierungen des Kopfes über dem Torso gemessen werden[2]. Auf der Syntheseseite kann dann mit Hilfe eines head-trackers die momentane Kopfposition des Hörers an eine Faltungssoftware übermittelt werden, worauf wiederum die entsprechende BRIR aus dem vorliegenden Datensatz ausgewählt und zur Faltung benutzt wird (Abbildung

[2]Das am Fachgebiet entwickelte Messsystem FABIAN [13] wird in Kapitel 2.7 vorgestellt.

2.5). Diese dynamische Binauralsynthese verringert Lokalisationsunsicherheiten signifikant ([14]), erfüllt in dieser Form aber noch nicht das in der Motivation aufgestellte *Plausibilitätskriterium*.

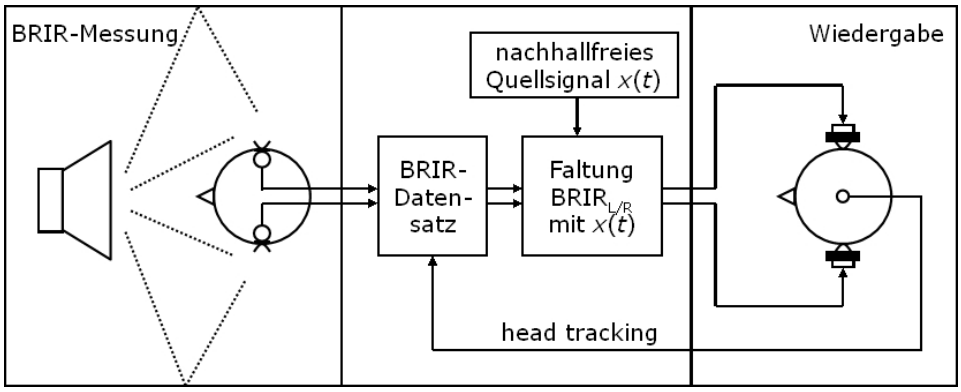


Abbildung 2.5.: schematische Darstellung dynamischer Binauralsynthese

Hierzu betrachten wir zunächst den Einfluss einzelner ausgewählter Komponenten auf das am Ohrkanaleingang ankommende synthetisierte Signal y_{TF}.

$$y_{TF}(t) = \underbrace{h_Q(t) * h_R(t) * h_{mic}(t) * h_{KK}(t) * h_{rec}}_{BRIR} * h_{WS}(t) * x_s(t) \tag{2.5}$$

Mit

- der den Raum anregenden Quelle $h_Q(t)$ mit Richtcharakteristik und Frequenzgang
- dem Raumeinfluss $h_R(t)$
- dem Frequenzgang der benutzen Mikrofone $h_{mic}(t)$
- dem Einfluss des nicht-individuellen Kunstkopfes $h_{KK}(t)$
- dem Aufnahmesystem $h_{rec}(t)$
- dem Wiedergabesystem $h_{WS}(t)$ auf der Syntheseseite, bestehend aus Verstärker, Kopfhörer und der Übertragungsstrecke zwischen Kopfhörertreiber und Ohrkanal

Diese einzelnen Komponenten können nun grob in drei Kategorien aufgeteilt werden:

1. die Rauminformationen $h_R(t)$ und das Quellsignal $x_S(t)$ als gewünschter Einfluss.

2. erwünschte aber möglicherweise nicht-ideale Einflüsse, wie die Quellenlautsprecherrichtcharakteristik und insbesondere der HATS mit nicht-individueller Anatomie.

3. üblicherweise unerwünschte, zu kompensierende Einflüsse $h_{mic}(t)$, $h_{rec}(t)$, $h_{WS}(t)$ und $|H_Q(j\omega)|$.

Das unter 1. aufgeführte Quellsignal wird später im Methodenabschnitt näher spezifiziert. $h_{mic}(t)$, $h_{rec}(t)$ und $|H_Q(j\omega)|$ sind durch vorherige Messungen bekannt und können somit im Postprocessing durch Faltung mit den inversen Anteilen jeweils kompensiert werden. Die nicht-individuelle Morphologie des Kunstkopfes, die Richtcharakteristik der Anregelautsprecher und das Wiedergabesystem sollen kurz im Kontext der entstehenden Artefakte Fehllokalisierung, Systemträgheit und spektrale Verfärbung näher betrachtet werden.

2.5.1. Richtcharakteristik der Messlautsprecher

Die den Raum anregende Quelle sollte idealerweise den Eigenschaften des gewünschten simulierten Schallerzeugers entsprechen. Soll beispielsweise eine Studioumgebung auralisiert werden, werden die vorhandenen Studiolautsprecher als Quelle für die BRIR-Messung benutzt, da diese Teil der zu auralisierenden Szene sind. Ist das Ziel allerdings die Hörbarmachung anderer Quellen, beispielsweise von Musikinstrumenten, sollte idealerweise ein Lautsprecher benutzt werden, der in seiner Richtcharakteristik der zu simulierenden Quelle gleicht. In [15] wurde der Einfluss der Berücksichtigung korrekter Richtwirkungen modellbasiert untersucht. Die Autoren kamen zu dem Schluss, dass die Richtcharakteristik einen Einfluss auf die Simulation hat. In Kapitel 2.1 wurde bereits die Arbeit von Meyer vorgestellt, der ebenfalls die Richtcharakteristik als wichtigen Faktor bezüglich Direkt-/Diffusschallfeldverhältnis identifiziert hat.

Datenbanken zu Richtcharakteristiken[3] sind mittlerweile verfügbar und werden von Verfahren der modellbasierten Raumakustiksimulation verwendet. Auch wenn die Anregung eines realen Raums mit Instrumentenrichtcharakteristiken für eine datenbasierte Impulsantwort prinzipiell mit einem mehrkanaligen Lautsprecher über Beamforming

[3] Beispielsweise von der Physikalisch-Technischen Bundesanstalt `http://www.ptb.de/` oder Daten von 41 historischen und modernen Musikinstrumenten aufgenommen mit einem sphärischen 32 Kanalarray im RAR der TUB[16]

möglich ist, liegen trotz einschlägiger Forschung derzeit noch keine praktikablen Ansätze zur messtechnischen Realisierung vor.

In dieser Arbeit wurden daher Beschallungslautsprecher verwendet, deren Unzulänglichkeiten durch eine perzeptiv motivierte Ausrichtung minimiert wurden (siehe Kapitel 3.1)

2.5.2. Individualisierung der BRIRs

Aufgrund morphologischer Unterschiede, vor allem bei Kopf und Außenohren, vom Hörer und dem zur BRIR-Messung verwendeten Kunsktopf, kommt es zu Verfälschungen in den lokalisationsrelevanten Cues, der ILD und ITD und den monauralen klangfarblichen Cues (SC spectral cues). Die Pegeldifferenzen sind nahezu vernachlässigbar, da spektrale Verfärbungen sehr schnell adaptiert werden, insbesondere wenn keine reale Referenz vorgegeben wird. Kritischer verhält es sich mit Verfälschungen der ITD, wie sie schon bei typischen Abweichungen der individuellen Kopfdurchmesser auftreten können.

Sollte der Hörer einen größeren Kopf als der HATS haben, entspricht die synthetisierte ITD einer eigentlich geringeren Auslenkung des Hörerkopfes. Perzeptiv bewegen sich die Quellen somit in die Bewegungsrichtung. Ist der Hörerkopf im Gegensatz dazu kleiner, entspricht die ITD einer größeren Auslenkung und die Quellen bewegen sich entgegengesetzt der eigentlichen Kopfbewegung.

Ein Verfahren zur Individualisierung der ITD wurde in [18] vorgestellt. Grundgedanke hierbei ist die nicht individuelle Kunstkopf-ITD über einen Skalierungsfaktor an den Hörer anzupassen. Hierzu werden mittels einer Schwellwertdetektion zunächst die akustischen Laufzeiten in den BRIRs bestimmt und diese dann, nach Berechnung der ITD, mit einer Diskretisierungsauflösung von 2,2 μs entfernt. Ein der dynamischen Faltung nachgeschalteter Prozess fügt den nun skalierbaren Zeitunterschied zwischen dem linken und rechten Signal subsamplegenau wieder ein (Abbildung 2.7). Als guter Prädiktor für den Skalierungsfaktor hat sich der Kopfdurchmesser erwiesen, gemessen als Distanz zwischen den *incisurae anteriores* beider Außenohren (auch 'Intertragusdistanz'), welche einfach mit einem Außentaster gemessen werden kann (Abbildung 2.6). Ein Nebenprodukt des Verfahrens ist ein quasi minimalphasiger Datensatz, wodurch bisher bestehende Artefakte bei Impulsantwortüberblendungen eliminiert wurden.

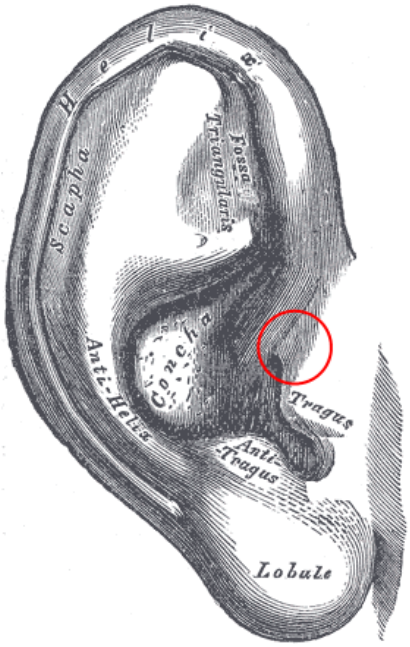


Abbildung 2.6.: Außenohr, Markierung zeigt Messpunkt für die Bestimmung der Intertragusdistanz, aus [17]

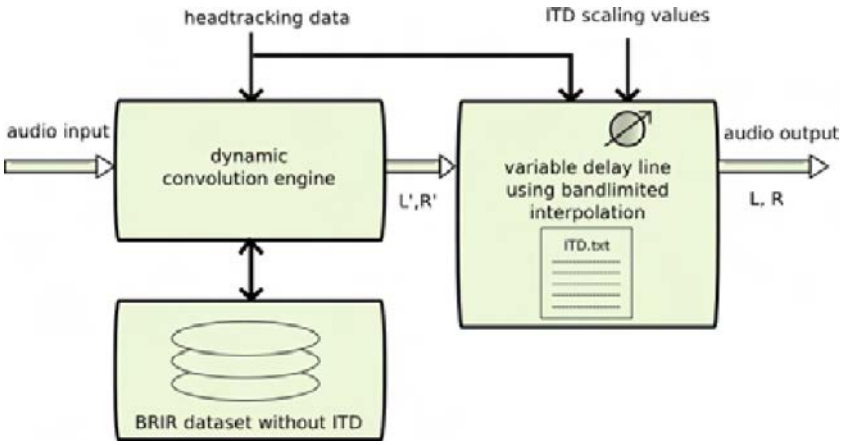


Abbildung 2.7.: Schematische Darstellung der dynamischen Binauralsynthese mit nachfolgender ITD-Indivdualisierung, aus [18, S.7]

2.5.3. Datensatzdiskretisierung

Systemlatenz, d.h. die Reaktion des gesamten Synthesesystems auf eine veränderte Kopfposition, hängt in starkem Maße von der Last des Rechnersystems ab und somit auch von perzeptiv nicht wahrnehmbarer Datenreduktion. Daher werden in diesem und den nächsten beiden Kapiteln Untersuchungen zur Ortsdiskretisierung und Verkürzung der dynamischen Impulsantwort sowie Latenzwahrnehmung kurz vorgestellt.

Dynamische Binauralsynthese basiert auf einem BRIR-Datensatz diskreter Kopf-über-Torso Positionen. Hohe Auflösungen erhöhen die Messdauer bei der Akquise und den Rechenaufwand bei der Synthese, während zu geringe Auflösungen zu wahrnehmbaren Sprüngen zwischen den Impulsantworten führen.

In [19] wurde die Wahrnehmungsschwelle (JND *just noticable difference*) für alle drei rotatorische Freiheitsgrade untersucht. Es wurden mit dem FABIAN HATS[4] Datensätze mit 1° Auflösungen in verschiedenen Räumen, vom reflexionsarmen Raum bis zu einem Vorlesungssaal mit 2,1 Sekunden Nachhallzeit, aufgenommen und Versuchspersonen sollten in einem adaptiven 3AFC Versuch den gröber diskretisierten Datensatz erkennen. Es zeigte sich für horizontale und vertikale Bewegungen eine starke Abhängigkeit von dem verwendeten Teststimulus (rosa Rauschen oder Gitarre), eine geringere von der Raumakustik. Probanden waren empfindlicher in Bezug auf Rauschen, mit zwei Testpersonen, die 2° Diskretisierung reliabel detektieren konnten. Die Schwellen für laterale Bewegungen bei frontaler Schallquelle liegen durch nicht vorhandene ITD und ILD Änderungen über den anderen beiden Freiheitsgraden, für kritische Quellenanordnung (d.h. über dem Kopf) ergeben sich jedoch ähnlich niedrige Schwellwerte, wie bei den anderen rotatorischen Bewegungen. Tabelle 2.1 zeigt die JNDs für verschiedene Perzentile.

2.5.4. Ansatz zur effizienten Darstellung späten Nachhalls - perceptual mixing time

Eine Raumimpulsantwort kann unterteilt werden in einen initialen, durch Direktschall und erste Reflexionen dominierten Teil und einen statistisch zufällig verteilten Nachhall. Während bei ersterem noch diskrete Reflexionsmuster erkannt werden können, zeichnet sich letzterer durch eine gleichverteilte Energiedichte und gleichverteilte Einfallsrich-

[4] Ein Messroboter, der auf Seite 25 vorgestellt wird.

...war hörbar für	**Rauschen** hor/ver/lat1/lat2	**Gitarre** hor/ver/lat1/lat2
50%	6° x 5° x 16° x 4°	9° x 12° x 16° x 7°
25%	4° x 4° x 12° x 3°	7° x 9° x 12° x 6°
5%	4° x 3° x 8° x 2°	5° x 4° x 8° x 5°
0%	2° x 1° x 3° x 1°	3° x 2° x 3° x 4°

Tabelle 2.1.: Rasterauflösung gerade hörbar für verschiedene Perzentile der Stichprobe für horizontale, vertikale und laterale Kopfbewegungen (lat1 Quelle frontal/lat2 Quelle über HATS)

tung der Schallreflexionen aus. Nach dem Übergangspunkt zum stochastischen Teil, der sogenannten *physical mixing time* [20], ist die Impulsantwort somit positionsunabhängig. Daher muss bei der dynamischen binauralen Synthese nicht der komplette Datensatz mit den vollen BRIRs im Speicher vorgehalten werden, sondern nur der kopfpositionsabhängige initiale Teil, während der stochastische Teil in einem separatem Prozess statisch gefaltet wird. In [21] wurden binaurale Datensätze in neun (in Absorptionsgrad und Volumen systematisch variierten) Schuhkartonräumen aufgenommen. In einem Hörversuch wurde eine Simulation mit voll gerendertem Datensatz einer Simulation mit statischem Nachhall und dynamischen initialen Teil gegenübergestellt. Der Übergang zwischen initialem und statischen Teil konnte variiert werden. In einem adaptiven 3AFC Versuch wurde daraufhin die perzeptive Schwelle ermittelt. Interessanterweise zeigten die Ergebnisse keine Abhängigkeit von der Menge diffuser Energie im Raum, d.h. von der relativen Halligkeit, sondern nur eine Abhängigkeit vom Raumvolumen. So steigt die perzeptive mixing time mit dem Volumen. Das ist ein Indikator dafür, dass eine gewisse Mindestanzahl an Wandreflexionen den Aussschlag für perzeptives 'mixing' gibt (anhand der Studienergebnisse zu 2-3 Reflexionen geschätzt). Detaillierte Betrachtungen sind in der zugehörigen Magisterarbeit zu finden [22].

2.5.5. Systemlatenz

Auf der Suche nach Möglichkeiten die Systemlast zu verringern, erscheint es auch sinnvoll die Schwelle zu finden und als Ziel zu formulieren, ab der Systemlatenz wahrnehmbar wird. Frühe Studien benutzten Lokalisierungsleistung als Prädiktor für Latenzwahrnehmung, mit sehr unterschiedlichen und teilweise unplausibel hohen Resultaten. Andere Untersuchungen schlugen als Prädiktor die minimal wahrnehmbare Winkel-

änderung bewegter Quellen vor (*minimal audible movement angle - MAMA*), letzteres abhängig von Quellengeschwindigkeit und Stimulusbandbreite. In [23] wurde die Wahrnehmbarkeitsschwelle mit Hilfe eines 3AFC-Versuchs untersucht, wobei die Kopfbewegungsdaten mit aufgezeichnet wurden, um die Ergebnisse zum MAMA - Prädiktor vergleichen zu können. Ergebnisse zeigen keinen Zusammenhang zwischen Latenzwahrnehmung, Nachhallzeit, Kopfrotationsgeschwindigkeit und Stimulusbandbreite, obwohl letztere nicht systematisch variiert wurde. Die Ergebnisse der Studie sind in Abbildung 2.8 dargestellt.

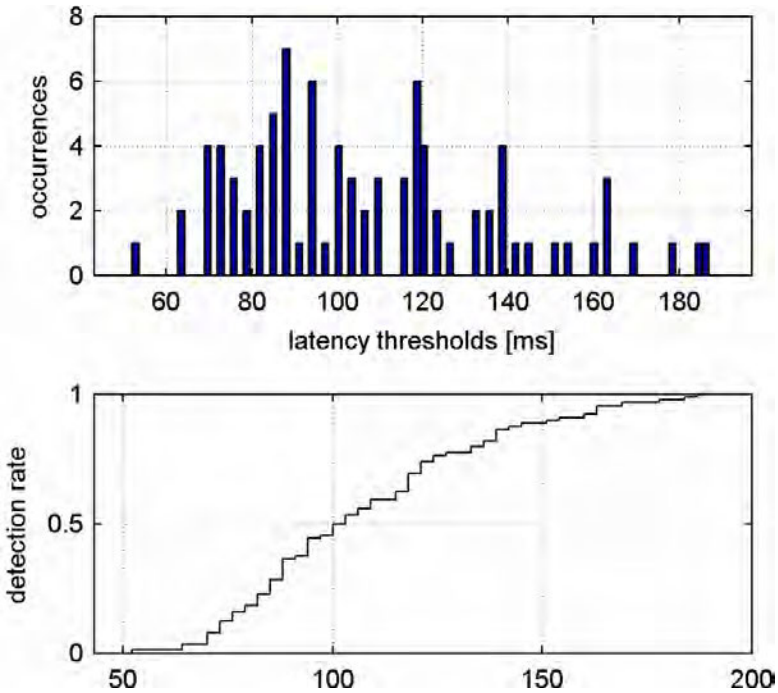


Abbildung 2.8.: Oben: Histogramm aller Latenzschwellen. Unten: Kumulierte Detektionsrate der Latenzwahrnehmbarkeitsschwellen, aus [23]

2.5.6. Wiedergabesystem

Einflüsse von elektronischen Geräten, Mikrofonen sowie in Einzellfällen Lautsprechern (bzw. zumindest der axialen Übertragungsfunktion letzterer beider) können relativ einfach kompensiert werden, während die Entzerrung der Wiedergabekette etwas genauer betrachtet werden sollte. Nach der Behandlung der zuvor genannten Artefakte Latenz und Lokalisierung, verbleiben nahezu ausschließlich spektrale Verfärbungen

[3] als Fehler, hauptsächlich verursacht durch das Wiedergabesystem. Wir betrachten hierzu die Kopfhörerübertragungsfunktion (HpTF - headphone transfer function) sowie Anforderungen an den Kopfhörer inklusive Verstärker.

Eine optimale Wiedergabe binauraler Signale wird nur mit sogenannten *free field equivalent coupling* (FEC) Kopfhörern erreicht ([24] und [25]), deren akustische Abstrahlungsimpedanz bei Betrachtung des Ohrkanals nach außen, dem des Freifelds entspricht. In [26] wurden verschiedene Entzerrungsmethoden der mit dem FABIAN HATS gemessenen HpTFs perzeptiv untersucht, insbesondere die des dem FEC-Kriterium besonders gut entsprechenden circumauralen STAX Kopfhörers.

Bei der Untersuchung wurde eine binaurale Simulation gegen ein reales Schallfeld mittels eines ABC/HR Versuchs verglichen, wobei die Realität immer von der Simulation unterschieden werden konnte. Die hochpassregularisierte LMS-Entzerrungsmethode erreichte jedoch die klanglich geringsten Abweichungen zur Realität.

Weiterführende Untersuchungen [27] mit einer am Fachgebiet entwickelten Otoplastik ermöglichten den Vergleich zwischen nicht-individuellen HpTFs (gemessen mit dem FABIAN HATS), individuellen HpTFs (gemessen an der Versuchsperson) und generischen HpTFs (Mittelung aus Messungen vieler VPs). Perzeptiv bevorzugt wurde die nicht-individuelle Entzerrung vom FABIAN HATS, von dem auch der zur Auralisierung benutzte BRIR Datensatz verwendet wurde. Die individuelle Entzerrung wurde nicht mit individuellen, d.h. durch die VP aufgenommenen BRIRs, untersucht, stellt in diesem Fall aber vermutlich die beste Entzerrungsmethode dar.

Beide Untersuchungen beobachteten große Variabilitäten im Frequenzgang bei geringfügig veränderten Aufsetzpositionen der circumauralen Kopfhörer. Um diese zu minimieren, wurde ein extraaurales Kopfhörersystem [28] [29] mit hohem SNR, ausreichendem Schalldruckpegel und Frequenzgangsrobustheit gegenüber Neupositionierungen entwickelt. Als Nebeneffekt können auch beispielsweise (3D-)Brillen ohne Einfluss auf die akustische Übertragung getragen, in-ear Messsysteme eingesetzt oder AB-Vergleiche gegen die Realität durchgeführt werden.

2.5.7. Qualitätskriterien für virtuelle akustische Umgebungen

Der höchste Anspruch, den man an eine akustische Umgebung stellen kann, ist die Ununterscheidbarkeit von Simulation und Realität. Ist das der Fall, sprechen wir von 'Authentizität'. Ein Simulationssystem kann hinsichtlich seiner Authentizität nur *in situ*

evaluiert werden, dass heißt in dem realen Raum, der auralisiert wird. Versuchspersonen können somit instantan zwischen der Simulation und der Realität wechseln. Dieser Fall tritt aber eigentlich nie ein, sofern das Simulationssystem nicht selbst Untersuchungsgegenstand ist. Die Referenz, auf die sich Versuchspersonen im Simulator beziehen, besteht aus Erinnerungen multimodaler Eindrücke und Erwartungen an vergleichbare reale Räume. Stimmt die Simulation mit der aus Erwartungen und Erfahrungen gebildeten inneren Referenz überein, kann von einer 'plausiblen' Simulation gesprochen werden.

Dieses Plausibilitätskriterium wurde in [30] aufgestellt und das in dieser Untersuchung benutzte Simulationssystem diesbezüglich mit Hilfe eines *in situ* Hörversuchs evaluiert. Da Forced-Choice-Verfahren immer die externe Referenz als Vergleichsstimulus enthalten würden, wurde mit Hilfe der Signalentdeckungstheorie und einem Ja/Nein Paradigma ein Messverfahren erstellt, dass die Bestimmung kriterienfreier Urteile gegen eine innere Referenz zulässt. Das Verfahren ermöglicht Aussagen über die Antwortneigung der Versuchspersonen (β) und die davon losgelöste sensorische Diskriminierbarkeit (d') der Stimuli. Von fünf in einem großen Hörsaal verteilten Lautsprechern wurden BRIR-Datensätze aufgenommen, während der HATS bereits den später im Hörversuch genutzten, nahezu akustisch transparenten, Kopfhörer trug. Versuchspersonen saßen im Hörversuch an der HATS-Position bei identischem Lautsprecheraufbau der BRIR-Messungen und mit aufgesetzten Kopfhörer. Nach Präsentation eines kurzen Audiosamples, zufallsgesteuert entweder über das Simulationssystem oder die Lautsprecher wiedergegeben, sollte die Entscheidung getroffen werden, ob das Audiosample durch den Lautsprecher abgespielt wurde, d.h. ob es 'real' war oder nicht.

Das auf dynamischer Binauralsynthese basierende Simulationssystem wurde in zwei Ausbaustufen getestet, einem rudimentären, basierend auf dem in [3] genutzten Stand und in einem verbesserten, mit den in diesem Kapitel aufgeführten Optimierungen.

Für die optimierte Simulation zeigen die Ergebnisse keinen statistisch signifikanten Unterschied zwischen dem Detektionsverhalten des realen oder simulierten Stimulus, während die rudimentäre Simulation tendenziell erkannt wurde (Abbildung 2.9).

2.5.8. Zusammenfassung

Ein direkter Vergleich zwischen realem und simulierten Schallfeld zeigt zeitliche und spektrale Unterschiede auf, verursacht durch unterschiedliche Kopfmorphologien der Versuchsperson und des benutzten Kunstkopfes. Wird ein hoch aufgelöster BRIR-Datensatz

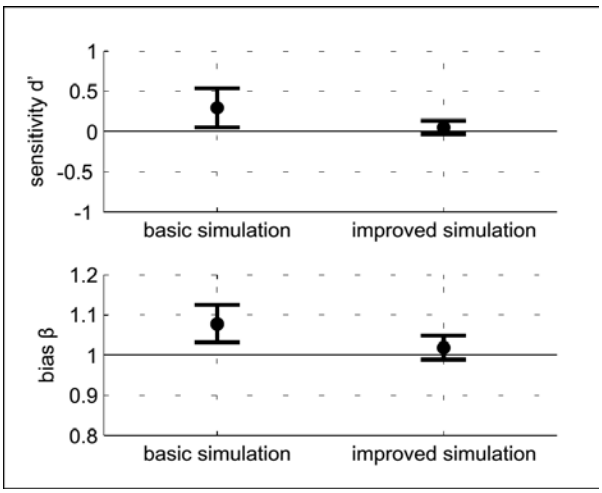


Abbildung 2.9.: Mittelwerte mit 90% Konfidenzintervallen der individuellen sensorischen Diskriminierbarkeit d' und der Antwortneigung β für die zwei Ausbaustufen des Simulators

verwendet, die nicht individuelle ITD des HATS an den Rezipienten angepasst, das Wiedergabesystem linearisiert und die Auralisation perzeptiv latenzfrei gehalten, ist die Simulation gegen eine innere Referenz von der Realität nicht mehr unterscheidbar. Die vorangegangenen Kapitel geben zudem Schwellwerte an, die jeweils eingehalten werden sollten:

Nutzung korrekter Lautsprecherrichtcharakteristiken
2° horizontale Winkelauflösung des BRIR-Datensatzes
Individualisierung der ITD
korrekt eingestellte *mixing time*
Kopfhörerentzerrung nicht-individuell auf dem FABIAN-Kunstkopf
Kopfhörerentzerrung nach der hochpassregularisierten LMS-Entzerrungsmethode
Nutzung eines extraauralen Kopfhörers
Totale Systemlatenz von unter 100 ms

Tabelle 2.2.: Anforderungen an die Audiosimulation, grün hinterlegte Anfoerderungen wurden erreicht, gelb hinterlegte angenähert

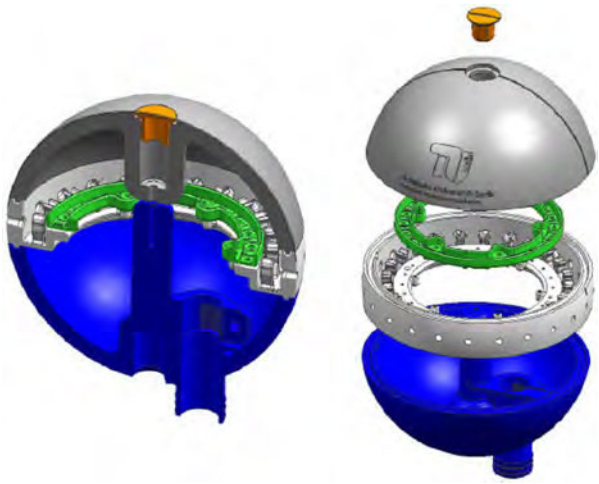


Abbildung 2.10.: CAD Modell des MTB-Arrays, aus [31]

2.6. Aufnahmenbasierte Synthese akustischer Szenen

Da soziale Interaktion als eine wesentlich kennzeichnende Qualität für reale Konzerterlebnisse in [18] genannt wurde, wurde beschlossen Hintergrundgeräusche bei Wiedergabe der Livesituation als unabhängige Variable miteinzubeziehen. Impulsantwortbasierte Syntheseverfahren erscheinen weniger geeignet für die Simulation einer Konzertatmosphäre, da für jede mögliche Quellen-Empfängerposition ein eigener Datensatz aufgenommen und mit jeweils eigenen Hörereignissen, wie Papierrascheln oder Räuspern, getrennt auralisiert werden müsste.

Im Rahmen einer Abschlussarbeit [32] am Fachgebiet wurde, basierend auf dem 'Motion-Tracked Binaural'-Verfahren von Algazi et al. [33], das sogenannte MTB-Array entwickelt. Dieses Mikrofonarray besteht aus 24 Mikrofonen, horizontal eingelassen in eine feste Kugel mit dem Durchmesser eines durchschnittlichen Kopfes (Abbildung 2.10). Akustische Szenen können hiermit gleichzeitig von mehreren möglichen Positionen aus aufgenommen werden. Bei Wiedergabe wird abhängig von der momentanen Kopfposition des Hörers auf Basis der 24 Audiospuren und eines definierbaren Interpolationsalgorithmus ein dynamischer, pseudobinauraler Audiostream generiert. Nach Hörversuchen, in denen Mikrofonanzahl, Interpolationsalgorithmus und Signaltyp unabhängig voneinander variiert wurden, zeigten sich signifikante Interaktionen erster Ordnung zwischen den unabhängigen Variablen (Abbildung 2.11). Je nach Anwendung und verfügbarer Rechenzeit können somit verschiedene Kombinationen aus Interpolati-

onsalgorithmus und Mikrofonanzahl ein optimales Ergebnis erzielen.

Da Konzertsaalhintergrundgeräusche ein Gemisch aus rauschhaften und tonalen Signalen darstellen, wurde der aufwändigste aber plausibelste *Two Band Spectral-Interpolation Restoration (TB-SI)* Algorithmus genutzt.

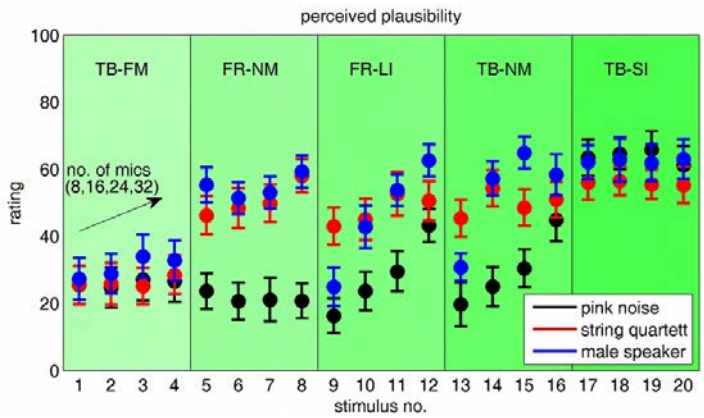


Abbildung 2.11.: Dargestellt sind Mittelwerte mit 95% Konfidenzintervallen auf einer die Plausibilität darstellenden Skala über den Faktoren Mikrofonanzahl, Interpolationstyp und Signaltyp (aus [31]).

2.7. Binaurale Datenakquise

Verschiedene kommerziell verfügbare Kunstkopf-Torso-Simulatoren können auf Drehtellern komplett rotiert werden, wobei wichtige, vor allem durch Schulterreflexionen bedingte Cues entfallen, oder die Kopfposition kann nur manuell verändert werden. Der am Fachgebiet Audiokommunikation der TU Berlin entwickelte [13] HATS FABIAN (**F**ast and **A**utomatic **B**inaural **I**mpulse response **A**cquisitio**N**) (Abbildung 2.12) ermöglicht eine automatisierte BRIR-Akquise für drei rotatorische Freiheitsgrade der Kopfbewegung. Basierend auf einem individuellen Abguss eines Kopfes ist dieser mit einem, nach anthropometrischen Mittelwerten entwickelten, Torso über ein präzises servomotorisches Halsgelenk verbunden. Dadurch ist eine räumliche Diskretisierung in 1°-Schritten bis auf einen Fehler von 0,002° möglich.

Die dazugehörende rauscharme und nichtlinearverzerrungsfreie Signalkette und in Matlab implementierte Messroutinen ermöglichen eine schnelle und zuverlässige Messung von BRIR-Datensätzen. Durch messtechnische Bestimmung von Mikrofon- und

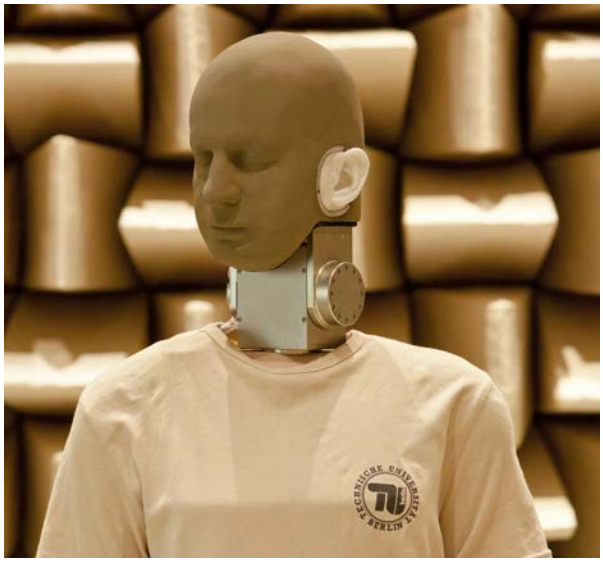


Abbildung 2.12.: Der Messroboter FABIAN

Messhardwareübertragungsfunktionen, kann die gesamte aufnahmeseitige Signalkette linearisiert werden. Die Messmikrofonemembran ist bündig in die Cavum Conchae des Kunstkopfs eingelassen, womit es sich um eine Messung an einem geblockten Ohrkanal handelt[5].

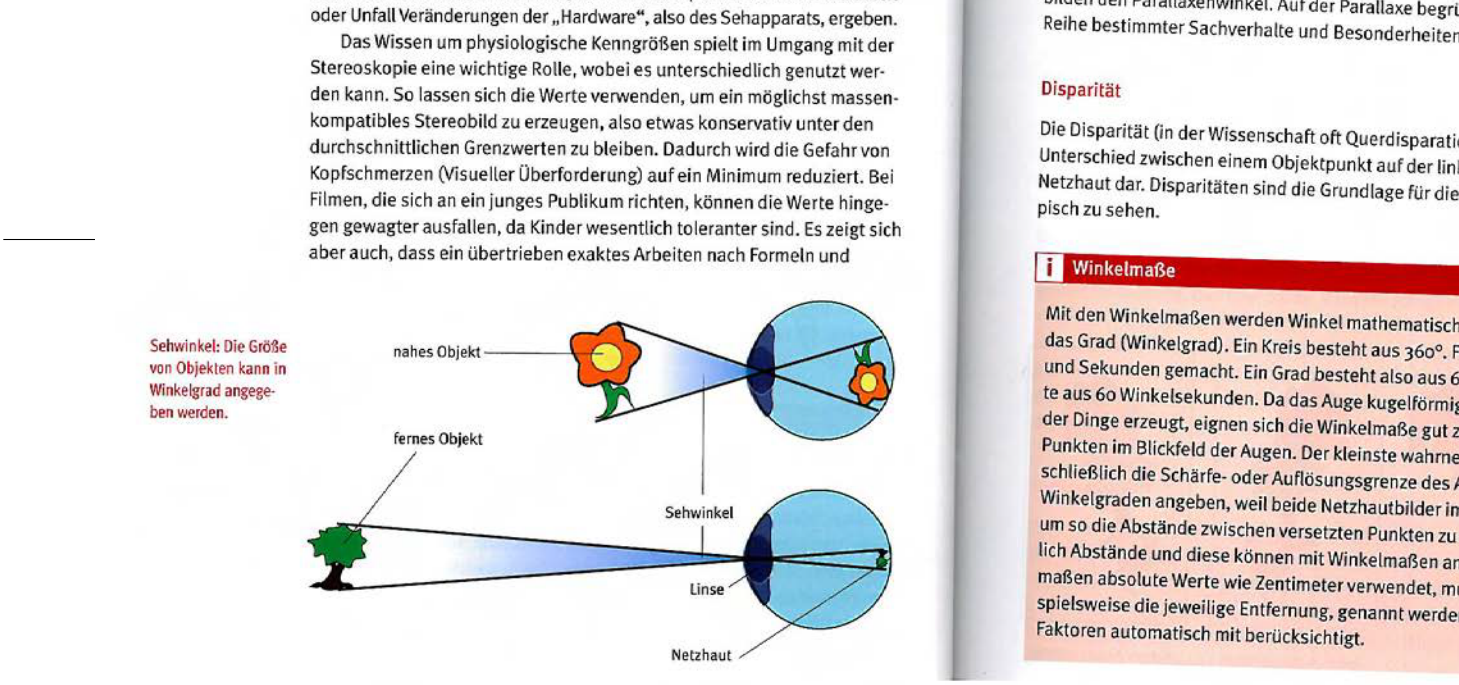
2.8. Räumliches Sehen und stereoskopische Projektion

Im folgenden werden kurz grundlegende, für eine visuelle Simulation relevante physiologische Voraussetzungen des Sehens wiedergegeben. Dabei handelt es sich im Wesentlichen um eine kurze Zusammenfassung der ersten beiden Kapitel aus [34].

Durch die Pupille tritt Licht in den Augapfel hinein, wird durch die Linse gebrochen und trifft auf die mit Sinneszellen besetzte Retina auf (Abbildung 2.13). Der sich in der Augenhöhle befindliche Augapfel ist an sechs Muskeln elastisch befestigt und ermöglicht Bewegungen auf allen Drehachsen. Die Bewegungen werden unterschieden in *Versionen*, bei denen beide Augen in die gleiche Richtung eine Bewegung ausführen, beispielsweise bei der Zuwendung auf ein Objekt, und *Vergenzen*, bei denen sich beide Augen aufeinander zu- oder voneinander Weg bewegen, auch Schielbewegungen genannt.

Das menschliche Sehfeld umfasst vertikal ca. 120° und horizontal ca. 180°, wobei

[5] Zu Vor- und Nachteilen sei auf S.30ff. in [13] verwiesen



oder Unfall Veränderungen der „Hardware", also des Sehapparats, ergeben.

Das Wissen um physiologische Kenngrößen spielt im Umgang mit der Stereoskopie eine wichtige Rolle, wobei es unterschiedlich genutzt werden kann. So lassen sich die Werte verwenden, um ein möglichst massenkompatibles Stereobild zu erzeugen, also etwas konservativ unter den durchschnittlichen Grenzwerten zu bleiben. Dadurch wird die Gefahr von Kopfschmerzen (Visueller Überforderung) auf ein Minimum reduziert. Bei Filmen, die sich an ein junges Publikum richten, können die Werte hingegen gewagter ausfallen, da Kinder wesentlich toleranter sind. Es zeigt sich aber auch, dass ein übertrieben exaktes Arbeiten nach Formeln und

Sehwinkel: Die Größe von Objekten kann in Winkelgrad angegeben werden.

Reihe bestimmter Sachverhalte und Besonderheiten

Disparität

Die Disparität (in der Wissenschaft oft Querdisparati
Unterschied zwischen einem Objektpunkt auf der lin
Netzhaut dar. Disparitäten sind die Grundlage für die
pisch zu sehen.

i Winkelmaße

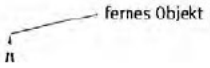
Mit den Winkelmaßen werden Winkel mathematisch
das Grad (Winkelgrad). Ein Kreis besteht aus 360°. F
und Sekunden gemacht. Ein Grad besteht also aus 6
te aus 60 Winkelsekunden. Da das Auge kugelförmig
der Dinge erzeugt, eignen sich die Winkelmaße gut z
Punkten im Blickfeld der Augen. Der kleinste wahrne
schließlich die Schärfe- oder Auflösungsgrenze des A
Winkelgraden angeben, weil beide Netzhautbilder in
um so die Abstände zwischen versetzten Punkten zu
lich Abstände und diese können mit Winkelmaßen ar
maßen absolute Werte wie Zentimeter verwendet, m
spielsweise die jeweilige Entfernung, genannt werde
Faktoren automatisch mit berücksichtigt.

Abbildung 2.13.: aus [34, S.20]

im letzteren Fall nur 120° von beiden Augen gleichzeitig erfasst und stereoskopisch ausgewertet werden können. Durch sogenannte *Akkomodation* oder auch *Scharfstellung*, wird über eine Änderung der Linsenkrümmung die Brennweite verändert, so dass die Lichtstrahlen des fixierten Objekts auf der Netzhaut konvergieren. Die Pupille reguliert schnelle Lichtintensitätsänderungen durch eine Veränderung des Lichteintrittsdurchmessers. Große Helligkeitsunterschiede, wie beispielsweise zwischen Tag und Nacht, werden langsamer über Rezeptorenempfindlichkeiten in einem Zeitbereich von bis zu 20 Minuten adaptiert. Der Ort der höchsten Rezeptorendichte wird Sehgrube genannt und kann als Fokalpunkt des optischen Systems bezeichnet werden. Außerhalb der Sehgrube projizierte Lichtstrahlen werden unscharf wahrgenommen. Durch ständige, unbewusste, schnelle Augenbewegungen, sogenannte *Sakkaden*, wird ein größerer Bereich abgetastet und aus den einzelnen, punktuellen Eindrücken entsteht ein flächenmäßig größerer, scharfer Eindruck.

Physiologische Voraussetzungen für stereoskopisches Sehen ist der Vergleich zweier perspektivisch unterschiedlicher Bilder, gegeben durch zwei im Mittel 6,5 Zentimeter voneinander entfernte Augen. Abhängig von der Objektentfernung konvergieren beide Augen, so dass zwei unterschiedliche Sehachsen zum Objekt entstehen. Der von den Sehachsen am Fixationspunkt eingeschlossene Winkel wird als Parallaxe bezeichnet. Der daraus entstehende Unterschied eines Objektpunkts zwischen beiden Augen wird Disparität genannt. Während das fixierte Objekt im Nullpunkt liegt, treffen davor und dahinter liegende Objekte an unterschiedlichen Stellen auf der Netzhaut auf (Abbildung 2.14a). Die beiden Teilbilder mit den enthaltenen Disparitäten können in einem bestimmten Bereich zu einem Gesamtbild fusionieren. Dieses ist begrenzt auf einen

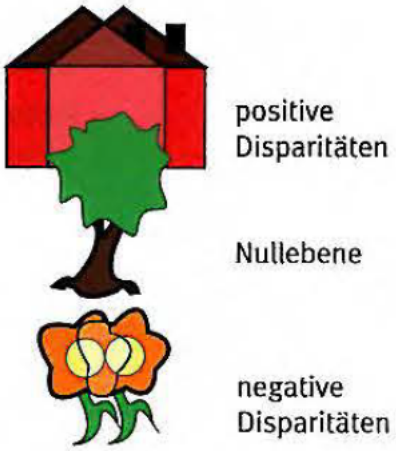
Bereich um den anvisierten Fixationspunkt. Sich vor dem Fixationspunkt befindliche Objekte weisen eine andere Disparität auf als dahinter liegende. Durch diesen Vergleich zum disparitätsfreien Fixationspunkt entsteht der relative Eindruck von „davor“ und „dahinter“. Fusionieren beide Teilbilder nicht zu einem Gesamteindruck, wird dies binokulare Rivalität genannt und tritt unter anderem auf, wenn die Disparitätsunterschiede größer als 1,5° sind.



Disparitätsgrenzen

In den Hyperkolumnen des Sehzentrums werden die beiden Netzhautbilder punktweise miteinander verglichen und auf Disparitäten analysiert. Wenn die Disparitäten zu groß sind, befinden sie sich nicht mehr in dem Bereich, der sich vergleichen lässt. Stereoneurone mit derart weit voneinander entfernten rezeptiven Feldern sind nicht vorhanden.

Der gesamte Vergleichsbereich des Auges umfasst etwa 1,5 Grad Sehwinkel, also 90 Winkelminuten. Dieser Wert ist eine grundlegende Kenngröße für die Stereoskopie. Bei stereoskopischen Bildern stellt er die ungefähre Grenze zwischen angenehmer und unangenehmer Betrachtung dar.



(b) Netzhautbilder unterschiedlich entfernter Objekte

Abbildung 2.14.: Disparitäten, aus [34]

Flexibilität

Stereozellen sprechen vorwiegend auf horizontale Disparitäten an, denn die Augen liegen neben- und nicht übereinander. Da diese Programmierung in der frühesten Kindheit eines jeden Menschen stattfindet, ist die Fähigkeit binokularen Sehens weitestgehend erlernbar. Kleine Kinder sind noch in der Lernphase und weisen deshalb eine größere Toleranz gegenüber großen Disparitäten oder vertikalen auf. Das zeigt sich auch bei der Vorführung stereoskopischer Filme. Je älter ein Mensch wird, desto inflexibler reagiert er auf Abweichungen.

▶ *Im angloamerikanischen Sprachgebrauch wird der Begriff der Parallaxe (parallax) etwas großzügiger ausgelegt und kann nicht direkt mit dem deutschen Begriff gleichgesetzt werden.*

Abgesehen von den binokularen Tiefeninformation Vergenz und Disparität, werden auch monokulare (einäugige) und bewegungsinduzierte Informationen ausgewertet. Zu den monokularen Informationen gehören

- Kriterien der relativen Größe, gleich große Objekte erscheinen größer, je näher sie am Betrachter sind.
- geometrische Abbildungsgesetze, parallele Objektkanten laufen in einem Fluchtpunkt zusammen.

Bedeutung

Die Disparität ist ein sehr wichtiger Begriff und für die Stereoskopie von großer Bedeutung. Das liegt daran, dass die Fähigkeit der Augen, räumlich zu sehen, auf der Disparität beruht. Daher findet sich der Begriff Disparität als Unterschied zwischen linkem und rechtem Teilbild auch bei stereoskopischen Abbildungen, Wiedergabemethoden, Aufnahmetechniken, sowie der Kodierung und Übertragung wieder. Im Verlaufe dieses Buches taucht

- Objektverdeckungen hintereinander positionierter Körper.
- atmosphärische Perspektive, weit entfernte Objekte erscheinen unschärfer durch sich in der Luft befindliche Partikel.
- Bewegungsaspekte die auf Entfernung deuten: Für einen bewegten Beobachter erscheinen nahe Objekte eine schnelle Bewegung in die entgegengesetzte Richtung auszuführen, während entfernte Objekte sich langsam mit dem Betrachter bewegen. Befinden sich zwei Objekte in unterschiedlicher Entfernung zueinander, kann aus dem resultierenden Verdeckungsprozess bei Betrachterbewegung die Entfernung bestimmt werden.

2.8.1. Anforderungen an eine optische Simulation

Eine der auditiven Simulation adäquate, transparente und immersive visuelle Simulation sollte sich an den gerade eingeführten physiologischen Voraussetzungen orientieren und die Reize und Reizänderungen liefern, die das visuelle System auch im realen Pendant der simulierten Umgebung erhalten würde. Dies bedeutet im Einzelnen:

1. ein *field of view* von mindestens 120° x 120°, besser 180° x 120°
2. stereoskopische Aufnahme und Wiedergabe, hierbei korrekte Abbildung der Objektformen und Objektgrößen
3. dynamische, auf *eye tracking* basierte Vergenz auf den Fixpunkt
4. entweder dynamische, auf *eye tracking* basierte Akkomodation und Vergenz auf den Fixpunkt oder Rückgriff auf mit kleiner Blende aufgenommenes Bildmaterial, um einen möglichst großen Schärfebereich zu erhalten
5. eine Bildauflösung, deren Pixeldichte keine wahrnehmbare Diskretisierung zulässt
6. artefaktfreie Aufnahme und Wiedergabe

2.8.2. Stereoskopische Aufnahmesysteme

Abbildung 2.15.: 3d One CP31 Kamera[6]

Die Videoaufnahmetechnik kann momentan keine Systeme liefern, die die oben genannten Punkte erfüllen würde. Kommerzielle Kameras mit sogenannten Fischaugenobjektiven und einem *fov* von 180° werden im künstlerisch-medialen Bereich genutzt, in dem physikalisch korrekte Szenenwiedergabe nicht notwendig ist. Vollpanoramische Multikamera- oder Multiobjektivsysteme, vornehmlich aus der Robotik, haben wiederum geringe Anforderungen an Bildauflösung, Farbechtheit und sind schwierig in der Anschaffung. Diese technische Lücke wird mittelfristig geschlossen werden, beispielsweise wird ein vollpanoramisches, hochauflösendes, stereoskopisches Kamerasystem derzeit am Fraunhofer Heinrich-Hertz-Institut entwickelt[7]. Seit diesem Jahr ist eine kommerzielle Lichtfeldkamera erhältlich[8], deren Akkomodationspunkt im Nachhinein, interaktiv verändert werden kann, allerdings wiederum in verhältnismäßig schlechter Auflösung und nur für Fotografien.

Momentan kann einzig ein auf Fotos basierendes Verfahren hochauflösende stereoskopische Vollpanoramen generieren [35], allerdings mit dem Nachteil das Bewegtbild aufwändig über Bluescreen-Technik im Nachhinein in das statische Bild zu integrieren.

Sowohl aus zeitlichen als auch ökonomischen Gründen, wurde daher auf ein handels-

[6] Abbildung von, `http://www.testberichte.de/p/3d-one-tests/cp-31-testbericht.html` abgerufen Oktober 2013

[7] `http://www.hhi.fraunhofer.de/de/kompetenzfelder/image-processing/solutions/capture/2d-3d-omnicam.html`

[8] www.lytro.com

übliches stereoskopisches Kamerasystem zurückgegriffen, dessen mögliche Artefakte jetzt beschrieben werden sollen.

Die meisten stereoskopischen Aufnahmesysteme basieren im Prinzip auf zwei Kameras, die in einem Gehäuse oder auf einer Schiene untergebracht sind. Im Ausgangszustand sind diese geometrisch identisch ausgerichtet, bis auf einen optional veränderlichen Abstand entlang der horizontalen Achse. Alle Funktionen wie Zoom, Schärfe, Belichtungszeiten, Aufnahmefrequenz, etc. müssen zwischen den Kameras möglichst perfekt synchronisiert sein. Ebenso sollten identische Optiken verwendet werden, um Unterschiede zwischen linkem und rechten Bildmaterial minimal zu halten. Zusätzlich können die meisten Kameras über Motoren um den Fokalpunkt nach innen gedreht werden, so dass die ursprünglich parallelen Sehachsen sich in einem Fixationspunkt treffen.

Die möglichen Artefakte sollen nun kurz aufgeführt werden. Überschreiten Abweichungen zwischen den Augen die Kompensationsgrenzen, entsteht binokulare Rivalität und die Bilder können nicht fusioniert werden.

- **Farbunterschiede** mit einem Wellenlängenabstand von bis zu 100 nm können als Mischfarbe kompensiert werden.

- **Schärfeunterschiede** von bis zu einem Prozent werden kompensiert. Allerdings wirkt ein summiertes Stereobild immer schärfer als eines der Teilbilder monoskopisch betrachtet. Perzeptiv orientiert sich das Schärfeempfinden innerhalb der Kompensationsgrenzen daher am besseren Auge.

- **Geometrische Unterschiede** werden entweder durch nicht-ideale mechanische Aufhängungen oder durch die unter Umständen gewünschte Vergenz auf einen Fixationspunkt verursacht. Durch Ersteres können die Teilbilder zueinander verdreht sein (0,5° Toleranz), vertikal versetzt (1% Toleranz) oder unterschiedlich skaliert (1% Toleranz) werden. Konvergente Ausrichtung verursacht eine Trapezverzerrung und somit neben den horizontalen auch vertikale, inhaltsabhängige Disparitäten, die perzeptiv eine Tiefenkrümmung verursachen (Abbildung 2.16). Dieser Effekt wird verstärkt durch eine steigende Kamerabasis, kürzere Brennweiten und größere Konvergenzwinkel.

(a) Horizontalversatz bei parallel ausgerichteten Kameras

(b) Trapezförmige Verzerrung und daraus resultierende vertikale Fehldisparitäten bei vergenten Kameras

Abbildung 2.16.: Einfluss der Kameravergenz, aus [34]

Für Orthostereoskopie[9] werden Kameras daher parallel ausgerichtet, um Trapezverzerrungen zu vermeiden. Der Vergenzpunkt kann im Nachhinein durch ein horizontales Verschieben der Bilder frei gewählt werden.

Ein weiteres qualitätsbestimmendes Merkmal ist die genutzte Bildkodierung, die meist mit einem Schärfeverlust einhergeht.

2.8.3. Stereoskopische Wiedergabesysteme

Die bei der Aufnahme angefertigten separaten Bildinhalte müssen wiedergabeseitig ebenso getrennt dem linken und rechten Auge dargeboten werden. Verbreitete Verfahren sind:

- räumliche Trennung der Wiedergabeflächen, entweder über Head-Mounted-Displays, Konstruktionen, bei denen vor jedem Auge eine Wiedergabefläche positioniert wird oder Autostereoskopieverfahren, bei denen abhängig von der Sehachse unterschiedliche Bildschirmbereiche sichtbar sind. Letzteres Verfahren benötigt keine Hilfsmittel in Form von Brillen.

- passive Stereoskopie in Form von gleichzeitig projizierten, unterschiedlich polarisierten Bildern und einer Aufteilung durch ebenfalls polarisierte Brillen.

- aktive Stereoskopie durch zeitlich versetzte Wiedergabe der Bildinformationen und dazu synchronisierten Shutterbrillen.

[9] form- und größentreue Projektion

Da letzteres in der fachgebietseigenen Simulationsumgebung zum Einsatz kommt, soll nur dieses Verfahren kurz betrachtet werden.

Die Flüssigkeitskristallgläser der Shutterbrillen werden abwechselnd abgedunkelt, synchron zu alternierend dargebotenen Bilder für das linke und rechte Auge. Um ein flimmerfreies Bild zu gewährleisten, muss die gesamt Bildwiedergabefrequenz auf mindestens 120Hz erhöht werden, so dass jedes Auge einem 60Hz Bild exponiert wird. Ein Nachteil dieses Verfahrens ist ein hoher Helligkeitsverlust und möglicher *crosstalk*, auch *ghosting* genannt. Letzteres bezeichnet schwach wahrnehmbare Doppelkonturen, verursacht durch entweder nicht ganz synchron laufende Brillen oder ein Nachleuchten des jeweils vorherigen Bildes, wodurch jedes Auge noch teilweise dem Bild des anderen Auges ausgesetzt wird.

Die Bilddaten können auf zwei verschiedene Arten übertragen werden. Die stereoskopischen Bilder können schon aus dem Rechner sequentiell an den Projektor weitergegeben werden. Dies setzt sowohl Grafikkarten mit einer möglichen Bildfrequenz von 120Hz voraus, als auch Projektoren, die eingangsseitig diese Bildfrequenz verarbeiten und weitergeben können. Hierbei wird die Brille auf den Grafikkartentakt synchronisiert.

Handelsübliche Grafikkarten arbeiten mit einer maximalen Bildwiederholfrequenz von 60Hz, Blue-Ray-Player nur mit 24Hz. Um hier die Informationen für das linke und rechte Auge übertragen zu können, werden das linke und rechte Bild gleichzeitig, in einem gegebenenfalls größerem Bild übertragen (Abbildung 2.17). Die Aufteilung des Gesamtbilds und sequentielle Wiedergabe der Einzelbilder geschieht hier im Beamer, die Brillen müssen demzufolge über diesen synchronisiert werden.

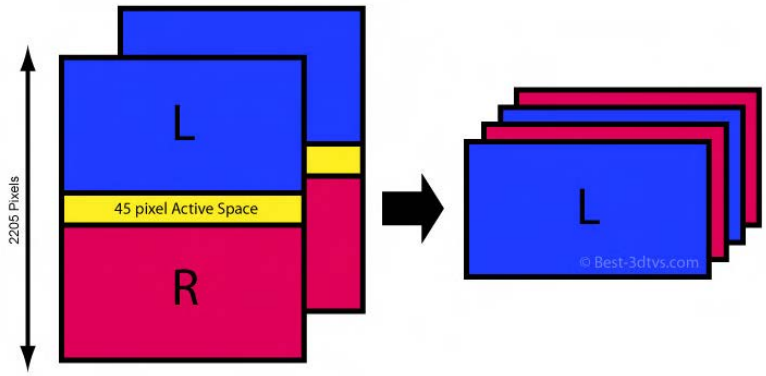


Abbildung 2.17.: Full HD frame packing Verfahren ohne Auflösungsverlust[10]

2.8.4. Zusammenfassung

Während die Projektion von Videomaterial mittels mehrerer Projektoren heutzutage eine qualitativ hochwertige Wiedergabe ermöglicht, stellt die Datenakquise noch ein großes Problem dar. Insbesondere der geringe *fov* der momentan verfügbaren Kamerasysteme führt zu einer starken Diskrepanz zwischen natürlichem Sehen und der optischen Simulation. In den letzten Kapiteln wurden Schwellwerte und Anforderungen an die Aufnahmesysteme gestellt, die es, sofern möglich, einzuhalten gilt. Tabelle 2.3 stellt diese zusammenfassend noch mal dar.

fov	180° x 120°
Farbunterschied zwischen den Kameras	< 100 nm
Schärfeunterschied zwischen den Kameras	< 1%
Zueinander verdrehte Bilder	< 0,5°
Vertikaler Versatz zwischen den Kameras	< 1%
Unterschiedliche Skalierung zwischen den Bildern	< 1%
Hohe Tiefenschärfe	möglichst große Blendenzahl
Form und größentreue Projektion	
Stereoskopische Aufnahme und Wiedergabe	
Vermeiden von Trapezverzerrungen durch parallele Kameraausrichtung	
Auf eye tracking basierte Vergenzanpassung	

Tabelle 2.3.: Anforderungen an die Videosimulation, grün hinterlegte Anfoerderungen wurden erreicht, rot hinterlegte nicht

2.9. Raum- und Stimulusauswahl

Die Stimuli sollen jeweils typische Rezeptionssituationen wiederherstellen und eine möglichst große Bandbreite medial 'artifizieller' bis realer Wiedergabe abdecken. Während die Wahl der *realen* Stimuli kaum Freiheitsgrade zulässt (abgesehen von der Sitzplatzwahl), lassen sich leicht diverse Kombinationen von Aufnahme/Wiedergabeverfahren finden, die sich deutlich bezüglich ihrer Klangfarbe und ihres Räumlichkeitseindrucks

[10] Abbildung von `http://www.best-3dtvs.com` abgerufen Oktober 2013

unterscheiden. Da man vermuten konnte, dass mehrkanalige Verfahren im Hinblick auf Quellenlokalisierung und Umhüllung eine höhere empfundene Ähnlichkeit zu Livemusik evozieren, wurden aus denselben nachhallfreien Aufnahmen unterschiedliche Mikrofonspuren erzeugt. Dazu wurden unterschiedliche Hauptmikrofonieverfahren, Stütz- und Raummikrofone mit unterschiedlichen Wiedergabeformaten kombiniert. Auf deren Basis wurden daraufhin von einem Tonmeister mehrere verschiedene Mischungen erzeugt.

Tabelle 2.4 stellt die zu realisierenden unabhängigen Bedingungsvariationen vor, die jetzt, von links nach rechts gelesen, kurz erläutert werden sollen.

Musikstück

Das 'moderne' Livekonzert wird oftmals über eine Beschallungsanlage verstärkt, nachdem Instrumente, meist nah, mikrofoniert und deren Signale einer mehr oder weniger komplexen Bearbeitung unterzogen wurden. Dieser Ablauf entspricht im Prinzip bereits einer medialen Produktion. Daher muss der musikalische Inhalt so gewählt werden, dass eine Livesituation ohne Verstärkung typisch ist. Um eventuelle Inhaltsabhängigkeiten aufzuzeigen, soll der Inhalt zumindest zweistufig variiert werden. Daher wurde zum einen ein als 'klassisch' bezeichnetes Streichquartett von Mozart [11], zum anderen ein als 'populär' bezeichnetes Stück mit Streichquartett und Stimme [12] produziert. Alle später erzeugten Stimuli basieren damit auf den identischen nachhallfreien Aufnahmen (Tabelle 2.18).

Bildstimulus

- *Rein akustisch* bezeichnet eine Darbietung ohne zusätzlichen optischen Stimulus.
- *Optoakustisch* bezeichnet die Darbietung mit optischer Wiedergabe entweder des Livekonzerts oder einer Ansicht des Abhörraums.
- *Optoakustisch-medial* bezeichnet die Darbietung mit optischer Wiedergabe entweder des Livekonzerts oder einer Ansicht des jeweiligen Abhörraums und nun inklusive der Darbietung eines passend produzierten Musikvideos.

[11] W.A. Mozart, KV 80, Streichquartett in G-Dur, 4. Satz (1:20 min) – Reinhold Quartett
[12] T. Zaufke, Tango: 'Chantey' (1:20 min) – Reinhold Quartett, Yamil Borgés, Gesang

Die technische Umsetzung muss möglichst nah an die originalen Gegebenheiten angepasst werden. Absolute Winkelverhältnisse und damit Größen als auch Farbtemperatur werden kalibriert. Alle Aufnahmen werden stereoskopisch mit möglichst weitem Bildwinkel vorgenommen und identischer Technik aufgenommen, um mögliche Artefakte konstant zu halten.

Medialität
Allgemeine Kategorisierung der Rezeptionssituationen in *real* oder *medial*.

Rezeptionssituation
Die 'Live'-Bedingung soll von zwei unterschiedlichen Sitzplätzen rezipiert werden können. Von einem allgemein als akustisch 'guten' bezeichneten Platz mittig und nah vor der Bühne, zum anderen von einem seitlichen, entfernten Platz aus.

Die medialen Rezeptionssituationen auf Basis simulierter Mikrofonaufnahmen können nach der Wiedergabe über Kopfhörer (KH) oder Lautsprecher (LS) unterschieden werden. Die Lautsprecherverfahren können weiterhin nach der Kanalzahl (1, 2, 5, 5, 56.1) klassifiziert werden sowie über die Mikrofonierungsverfahren, die für verschiedene Mischungen herangezogen wurden.

Sozialer Kontext
Der Besuch eines Konzertes geht immer mit zwischenmenschlicher Interaktion einher, die sich im Labor besonders schwierig nachstellen lässt. Technisch möglich ist aber die Wiedergabe der akustischen Hintergrundgeräusche eines besetzten Konzertsaals, welche zusätzlich zum eigentlichen Musikinhalt wiedergegeben werden können und hier als sozialer Kontext bezeichnet werden.

Szenengestaltung
Moderne Mehrkanalsyntheseverfahren ermöglichen eine größere Freiheit in der Gestaltung akustischer Szenen. Schallquellen müssen jetzt nicht mehr zwingend in der Frontalebene zwischen Lautsprechern in einer Stereoanordnung platziert werden, sondern können nahezu frei im Raum verteilt werden. Daher wurden für die Wellenfeldsynthese je Musikstück zwei Szenen erstellt, eine konventionelle mit sich vorne befindlichen Musikern und eine 'kreative' mit räumlich unüblicher Schallquellenverteilung.

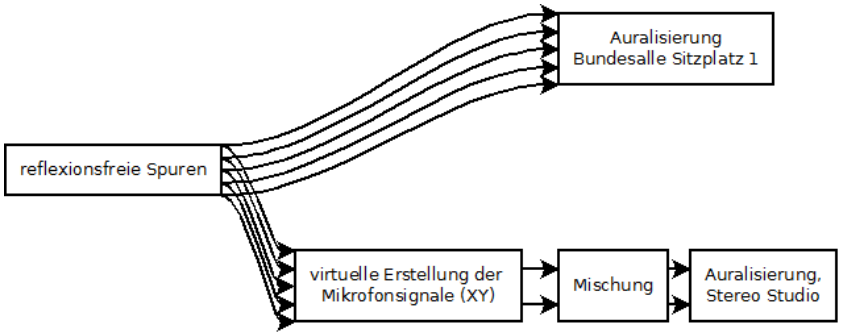


Abbildung 2.18.: Nutzung derselben reflexionsfreien Spuren für alle Stimuli. Hier gezeigt für eine 'Livesituation' sowie eine 'Tonstudiosituation' mit einer auf XY-Hauptmikrofonie basierenden Abmischung.

Raumwahl und BRIR-Datensätze

Der Raum muss ein prototypischer Aufführungsort für den gewählten musikalischen Inhalt sein und aus eher praktischen Gründen eine längere Nutzung für Impulsantwortakquise und Videoaufnahmen ermöglichen. Hierzu wurde der üblicherweise für Kammerkonzerte genutzte Joseph-Joachim-Konzertsaal der UDK-Berlin als akustischer und visueller Raum, sowohl für die 'Live'-Stimuli als auch alle Mikrofonaufzeichnungen, gewählt. Die optoakustischen Daten für eine moderne mediale Wiedergabe wurden im WFS-Labor der Telekom-Laboratories Berlin erhoben, die für alle verbleibenden haushaltsüblichen medialen Situationen im kleinen Elektronischen Studio des Fachgebiets Audiokommunikation (TU Berlin). Zusätzlich wurden Übertragungsfunktionen von In-Ear Kopfhörern gemessen und ebenfalls auralisiert. Insgesamt standen somit somit 76 verschiedene Stimuli zur Verfügung.[13]

2.10. Zusammenfassung

Es wurden die methodischen und technischen Voraussetzungen zur Schaffung eines individuell verständlichen, relevanten und weitestgehend validen Vokabulars zur Beschreibung von Wahrnehmungsunterschieden zwischen live und medial rezipierter Musik vorgestellt. Da von einem bedeutenden Einfluss des optischen Eindrucks auf das

[13] 2 Inhalte x 2 Bildinhalte x 2 Sitzplätze x 2 soziale Kontexte = 16 'Live'-Stimuli
2 Inhalte x 3 Bildinhalte x 8 mediale Rezeptionssituationen = 42 mediale Stimuli
2 Inhalte x 3 Bildinhalte x 1 moderne mediale Rezeptionssituationen x 2 Szenengestaltungen = 12 moderne mediale Stimuli

Musikstück	Bildstimulus	Medialität	akustische Rezeptionssituation	sozialer Kontext	Szenen-gestaltung	BRIR-Datensätze
Klassisch oder Populär	rein akustisch oder optoakustisch	Live	Konzertsaal Platz 1, vorne	ohne		Konzertsaal
				mit		
			Konzertsaal Platz 2, hinten	ohne		
				mit		
	rein akustisch oder optoakustisch oder optoakustisch-medial	Medial	LS 1.0 & Kugel			Studio
			LS 2.0 & Groß-AB			
			LS 2.0 & XY			
			LS 2.0 & ORTF			
			LS 5.0 & Decca Tree + Surround			
			LS 5.0 & INA-3 + Surround			
			KH 2.0 & Groß-AB			
			KH 2.0 & XY			
			LS 56.1 & WFS		konventionell	WFS-Labor
					kreativ	

Tabelle 2.4.: Übersicht aller möglichen Stimuli

Wahrnehmungsurteil ausgegangen werden kann, wurde vereinbart, auch den visuellen Aspekt zu simulieren. Während die Audiosimulation nach heutigem Stand der Technik dem aufgestellten *Plausibilitätskriterium* gerecht werden kann, müssen bei der optischen Simulation Abstriche, insbesondere im Bezug auf das Sichtfeld, in Kauf genommen werden. Wahrnehmungsschwellen wurden als Indikatoren für Anforderungen an technische Parameter betrachtet, die es in den folgenden Kapiteln einzuhalten gilt.

Tabelle 2.5 fasst die technischen Qualitätsanforderungen im Kontext der Versuchmethodik zusammen.

Anforderungen an		
interne Validität	konstante Störvariablen	Bildartefakte
		Stimuluslautstärke
		identische musikalische Einspielung
		synchrone A/V Wiedergabe
	ausschließliche Variation der unabhängigen Variablen	
externe Validität	Allgemeingültigkeit	prototypische Situationen
		plausible Audiosimulation
		immersive visuelle Simulation
		repräsentative Stichprobe
allgemein	flexibles System	einfache Zuweisung von Bild/Ton Situationen

Tabelle 2.5.: Übersicht verschiedener Qualitätsanforderungen der zu realisierenden Versuchsmethodik

3. Audiosimulation

Das Erreichen der zwei Hauptziele dieser Arbeit (immersive und flexible Hörversuchsumgebung, Generierung eines deskriptiven Vokabulars) machte eine Vielzahl methodischer Schritte notwendig. Da diese Arbeit in ein längerfristiges Projekt mit mehreren Bearbeitern eingebettet war, konnte der Autor nicht an allen Schritten selbst beteiligt sein, beziehungsweise wurde, neben den in Kapitel zwei genannten Vorarbeiten, teilweise auf Zuarbeiten anderer Beteiligter zurückgegriffen. Im Sinne der Verständlichkeit wurde jedoch entschieden, alle methodischen Schritte zu dokumentieren. Die vom Autor selbst bearbeiteten Teile (Synthetisierung der Mikrofonsignale, BRIR-Messungen, 3D-Videoaufnahmen, 3D-Videobearbeitung, Implementierung und Kalibrierung der Versuchsumgebung, Betreuung von Vorversuchen) wurden jedoch besonders gekennzeichnet.

In diesem Kapitel werden alle Schritte vorgestellt, die zur Erzeugung der gewünschten akustischen Simulation erforderlich waren. Angefangen mit der Aufnahme der nachhallfreien Quellsignale (nicht originärer Teil dieser Arbeit), über die Erzeugung der musikalischen Inhalte, Akquise binauraler Raumimpulsantworten in den zu auralisierenden Räumen, 'Studio', 'WFS-Labor' und 'Konzertsaal' bis hin zum Postprocessing der erhobenen Daten und letztendlich die Implementierung und Kalibrierung in einer interaktiven Hörversuchsumgebung.

3.1. Erzeugung der musikalischen Inhalte

Ausgangsmaterial für die Auralisation waren nachhallfreie Aufnahmen von sechs Stücken, die im Rahmen einer Abschlussarbeit ([36, S.18ff]) am Fachgebiet entstanden sind. Von diesen sechs Stücken wurden letztendlich zwei für die vorliegende Arbeit genutzt. Zum einen das Streichquartett G-Dur KV 80 'Lodi Quartett' von W.A. Mozart, zum anderen 'Chantey' für Streichquartett und Stimme von Thomas Zaufke. Die Streichinstrumen-

te wurden jeweils vom 'Reinhold-Quartett', bestehend aus Mitgliedern des Leipziger Gewandhausorchesters, eingespielt. Der Gesang stammt von Yamil Borges, einer in Berlin lebenden Sängerin. Die so erhaltenen vier beziehungsweise fünf übersprechfreien Audiospuren wurden auf verschiedene Arten verwendet, um die verschiedenen 'live erlebten' und medialen Rezeptionssituationen zu synthetisieren. Diese wurden

- für die Livesituation ohne weitere Bearbeitung mit den Konzertsaal-BRIRs auralisiert.

- für die Erstellung der medialen Produktionen (Abmischungen) in Form synthetisch erzeugter, mehrkanaliger Mikrofonaufnahmen (s.u.) einem Tonmeister zur Verfügung gestellt. Dieser fertigte im kleinen Elektronischen Studio der TU Berlin mehrere Abmischungen an, die nachfolgend mittels der am selben Ort aufgenommenen BRIRs zur Auralisation der medialen Situation 'Tonstudio' dienten. Die Abmischungen sollten mittels typischer marktgängiger Produktionsweisen und damit unter Verwendung von Bearbeitungsmöglichkeiten wie Pegelung, Panoramapositionierung, Dynamik- und Klangbearbeitung angefertigt werden.

- im WFS-Labor mit dem Sound Scape Renderer (SSR) [1] von demselben Tonmeister spatialisiert und klanglich bearbeitet. Die WFS-Szenen wurden ebenfalls unter Verwendung von BRIRs des gesamten WFS-Arrays, der WFS-Treiberfunktionen der jeweiligen vom Tonmeister eingerichteten spatialen Szene und den final bearbeiteten Einzelspurausspielungen reproduziert.

Konventionelle Aufnahmeverfahren - Synthese der Mikrofonsignale
Mikrofonaufnahmen klassischer Streichquartette erhält man üblicherweise durch Aufnahmen in einem Konzertsaal mit je nach Zielmedium und persönlichen Präferenzen des Tonmeisters verschiedenen Mikrofonierungsverfahren. Um aus den gegebenen nachhallfreien solche virtuellen Mikrofonsignale zu erhalten, wurden im Joseph-Joachim-Konzertsaal der UDK Berlin fünf, mit einem externen Lautsprechercontroller linear entzerrte (Abbildung 3.1), Lautsprecher in Position, Höhe und Abstrahlwinkel einem echten Streichquartett plus Stimme entsprechend aufgebaut und durch einen Tonmeister mikrofoniert.

[1] abgerufen am 12.09.2013 http://www.tu-berlin.de/?ssr

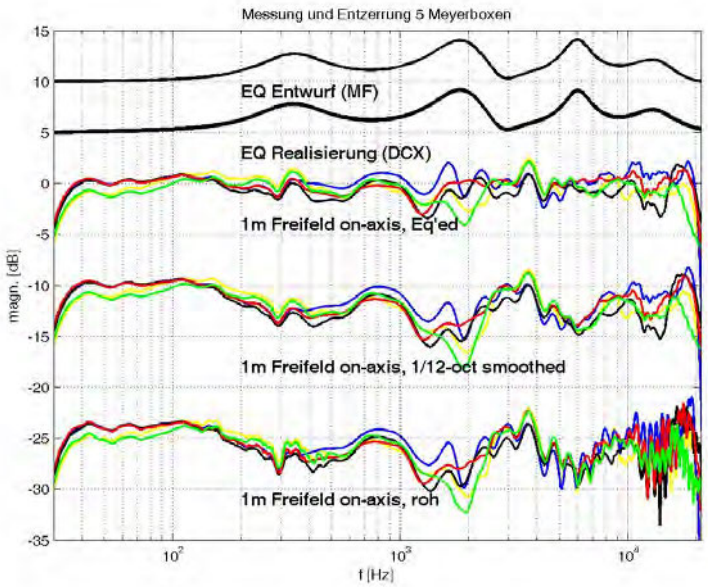


Abbildung 3.1.: Entzerrung der benutzten Meyer UPL-1 Lautsprecher.

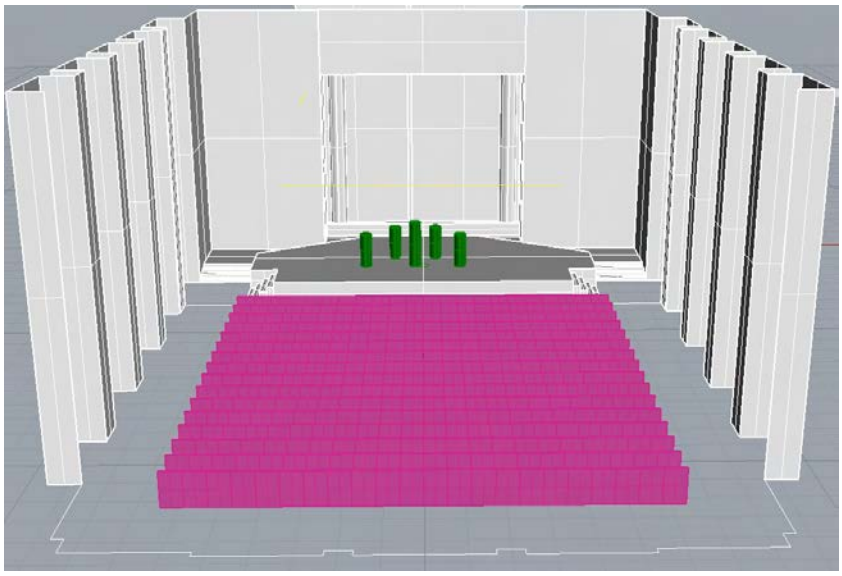


Abbildung 3.2.: CAD-Modell des Konzertsaals, die grünen Zylinder visualisieren die Lautsprecher-/Musikerpositionen

Die Richtcharakteristik der Lautsprecher wird als zumindest dem realen Instrument ähnlich angenommen und aufgrund zu dem Zeitpunkt mangelnder Alternativen als hinreichend gut akzeptiert. Während des Mikrofon- und Lautsprecheraufbaus wurden zu Testzwecken die nachhallfreien Aufnahmen wiedergegeben und Winkel und Positionen der Lautsprecher nach Vorgaben des Tonmeisters angepasst, um den Klang eines realen Instruments zu erhalten. Folgende Mikrofonaufbauten wurden verwendet:

- Stützmikrofone vor jeder Schallquelle
- Einkanalige Verfahren: Kugel und Nierenmikrofon
- Zweikanalige Verfahren: XY, ORTF, Groß-AB
- Mehrkanalige Verfahren: INA 3, INA 5, Decca Tree und ein Soundfieldmikrofon
- Raumsignale: IRT-Kreuz und neun Diffusfeldmessungen für Diffusfelsynthese in der WFS-Produktion

Detaillierte Aufbaupläne befinden sich ab Seite 125 im Anhang.

(a) Lautsprecheraufbau für die Messung der Mikrofonübertragungsfunktionen und BRIRs

(b) Lautsprecheraufbau mit aufgebauten Stützmikrofonen

Abbildung 3.3.: Konzertsaalmessung

Nach der Einrichtung aller Mikrofone und Lautsprecher war es die Aufgabe des Autors, die Übertragungsfunktionen zwischen allen Lautsprecher- und jeweils allen Mikrofonpositionen zu messen. In einem Nachbearbeitungsschritt erzeugte der Autor über die

Faltung der so erhaltenen Impulsantworten mit den nachhallfreien Instrumentenspuren virtuelle Mikrofonsignale. Diese entsprachen Signalen einer klassischen Aufnahme unter Berücksichtigung aller natürlich vorkommender 'Unzulänglichkeiten', wie beispielsweise Instrumentenübersprechen[2]. Die virtuelle Mikrofonspur für beispielsweise das Stützmikrofon der ersten Violine setzte sich also zusammen aus:

$$y_{v1} = x_{v1} * h_{LS1-M1} + x_{v2} * h_{LS2-M1} + x_{vl} * h_{LS3-M1} + x_{vc} * h_{LS1-M14} \tag{3.1}$$

x bezeichnet die nachallfreien Spuren, die Indizes das jeweilige Instrument, h bezeichnet die Übertragungsfunktion, die Indizes jeweils die Übertragungsstrecke nach Abbildung 3.4.

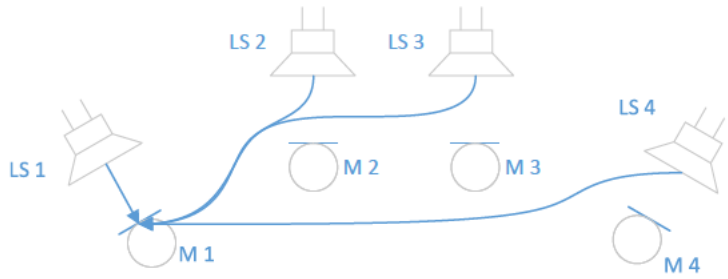


Abbildung 3.4.: Genutzte Übertragungswege für ein virtuelles Stützmikrofonsignal

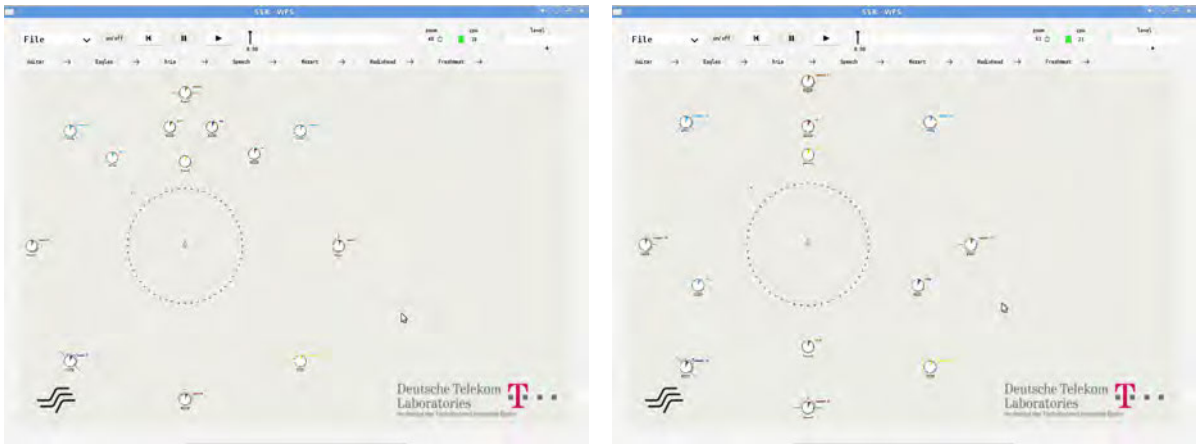
Die so erzeugten virtuellen Mikrofonspuren wurden dann von einem Tonmeister[3] im Kleinen Elektronischen Studio des Fachgebiets über eine modifizierte ITU-Anordnung [37] der Lautsprecher gemischt. Alle Bearbeitungsschritte wie beispielsweise Entzerrungen oder Dynamikbearbeitungen waren zugelassen, jedoch keine Eingriffe in den zeitlichen Ablauf der Stücke, um mit den anderen Stimulipräsentationen konsistent zu bleiben. Das Endprodukt waren Abmischungen für verschiedene Lautsprecherkonfigurationen, erstellt mit verschiedenen Hauptmikrofonierungsverfahren, Stütz- und Raummikrofonen. Insgesamt wurden 11 Abmischungen erzeugt, von denen letztendlich die folgenden sechs benutzt wurden:

- eine Monomischung mit einem omnidirektionalen Hauptmikrofon
- eine Stereomischung mit Groß-AB als Hauptmikrofonverfahren

[2]Maltabcode auf Daten-CD
[3]Eckhard Glauche

- eine Stereomischung mit XY als Hauptmikrofonverfahren
- eine Stereomischung mit ORTF als Hauptmikrofonverfahren
- eine 5.0 Surroundmischung mit DeccaTree + IRT-Kreuz
- eine 5.0 Surroundmischung mit INA 3 + IRT-Kreuz

Die Abmischungen für die moderne mediale Wiedergabe wurden im WFS-Labor (Details S.52) durch denselben Tonmeister erstellt. Als Quellsignale dienten die fünf nachhallfreien Instrumentenspuren und acht Nachhallspuren, realisiert über eine Faltung mit den Diffusfeldimpulsantworten des Konzertsaals. Die insgesamt 13 Audiospuren wurden aus einem PC-basierten Multitracker mit der DAW 'Sequoia' in das Wiedergabesystem eingespeist und mit der SSR-Software spatialisiert (Abbildung 3.5). Für die Audiobearbeitung galten die Einschränkungen der Tonstudioabmischungen, die Spatialisierung sollte allerdings das Potential des modernen Wiedergabeverfahrens ausschöpfen. Daher wurden zwei Szenen erstellt, eine mit konventioneller, frontaler Schallquellenverteilung und eine mit 'kreativen' Quellenpositionen. Es wurde ebenfalls eine Mischung für ein weiteres Spatialisierungsverfahren (VBAP) erzeugt, das aber nicht in das endgültige Simulationssystem implementiert wurde und hier somit nicht weiter betrachtet werden wird.



(a) 'konventionelle' Spatialisierung des 'klassischen' Stücks

(b) 'kreative' Spatialisierung des 'modernen' Stücks

Abbildung 3.5.: Benutzeroberfläche des SSR mit Positionen der Punktquellen und ebenen Wellen.

Durch die Bearbeitung wurden teilweise DAW bedingte Grunddelays eingefügt, welche im Nachhinein kompensiert wurden, um identische Startzeitpunkte wiederherzustellen und dadurch artefaktfreies, instantanes Wechseln zwischen den Mischungen zu ermöglichen.

3.2. Erhebung und Aufbereitung der Daten für binaurale Raumsimulationen

Alle Rezeptionssituationen sollten mit dynamischer Binauralsynthese auralisiert werden, dazu erhob der Autor in allen in Frage kommenden Umgebungen entsprechende BRIR-Datensätze. Der prinzipielle Aufbau einer BRIR-Messungen ist in Grafik 3.6 dargestellt, die verwendeten Geräte in Tabelle 3.1.

alle Messungen	Lenovo T61 Messlaptop RME HDSP ExpressCard RME Multiface RME Quadmic FABIAN
Studio	Genelec 1031a
Konzertsaal	Behringer DCX2496 Ultradrive Pro Meyer UPL-1 Neumann KM140, KM130, KM 184 Microtech Gefell M930 Behringer ECM 8000
WFS-Labor	56-kanaliges Lautsprecherarray Genelec 7060A

Tabelle 3.1.: Verwendete Geräte für die akustische Datenakquise

Alle Messungen wurden mit dem Messroboter FABIAN und der dazugehörigen Matlab-Messsoftware [13] durchgeführt. Als Anregesignal diente ein gefärbter Sinussweep mit Bassemphase, in Länge angepasst an die Nachhallzeit des Raumes in vielfachen von 2^n Samples, um den Anforderungen an die nachfolgende FFT gerecht zu werden. Vor Beginn einer Messreihe wurde über eine Loopbackmessung zwischen DA-Wandler und Mikrofoneingang ein zweikanaliges Referenzspektrum generiert. Das bei der Messung dann aufgenommene binaurale Signal wurde durch die Referenz komplex dividiert

und kompensiert somit die Einflüsse des Anregesignals, der AD/DA-Wandler und der Mikrofonvorverstärker. Die resultierende BRIR enthielt dann nur noch Informationen über den Raum, die Schallquelle und den Empfänger. Eine korrekte Positionierung des Messroboters (in Lotrichtung und definierter Ohrhöhe) wurde immer mit Hilfe von Laserdistrometern und Wasserwaagen vorgenommen.

Nach Einrichtung der Lautsprecher und der Messtechnik wurde das System auf den vermutlich maximalen Schalleinfall ausgepegelt, das heißt auf die nächstliegende Quelle mit dem Ohr in Quellrichtung. Alle hier aufgenommenen Datensätze wurden für horizontale Kopfrotation über dem Torso in einem Winkelbereich von -80° bis 80° mit einer Diskretisierung von einem Grad und einer Samplefrequenz von $44,1kHz$ aufgenommen.

Die nun vorliegenden Rohdatensätze wurden in mehreren Postprocessingschritten aufbereitet[4], um vor allem die in den Unterkapiteln 2.5.2 und 2.5.3 eingeführte Datenreduktion, BRIR-Individualisierung und Latenzminimierung vorzunehmen. Die relativen Pegelunterschiede zwischen den Quellen wurden beibehalten, davon ausgehend, dass diese korrekten realen Ausbreitungdämpfungen entsprechen und sich bei Anregung mit für alle Quellen in gleicher Entfernung aufgenommenen nachhallfreien Ausgangssignalen, auch wieder korrekt einstellen.

Wie in Kapitel 2 bereits beschrieben, erhält man durch die Extraktion von Laufzeit und ITD aus den Impulsantworten quasi-minimalphasige Impulsantworten, wodurch bei der Synthese Überblendartefakte und Latenzen minimiert werden und eine nachträgliche individualisierte ITD-Anpassung ermöglicht wird. Dabei ist jedoch zu beachten, dass der *initial time gap*, also die Zeit zwischen Raumanregung und dem Auftreffen des Direktschalls auf die Mikrofone, hierbei ebenfalls aus allen Quellen entfernt wird, wodurch wiederum die Latenz auf Kosten natürlich auftretender Laufzeitverzögerungen zwischen den Quellen minimiert wird. Als Kompromiss zwischen Latenz und Abbildungstreue wurden die natürlichen Laufzeitdifferenzen zwischen den Quellen in den BRIRs beibehalten.

Die Extraktion wird über einen sogenannten *onset detection* Algorithmus vollzogen, der den Punkt in der Impulsantwort detektiert, der N dB unter dem Amplitudenmaximum liegt (Schwellwertdetektion). Zum Schutz des Einschwingvorgangs wird zudem ein festes Predelay von einigen 10 samples berücksichtigt. Das Fenster wird daher k Samples

[4] Die Bearbeitung wurde nicht vom Autor vorgenommen.

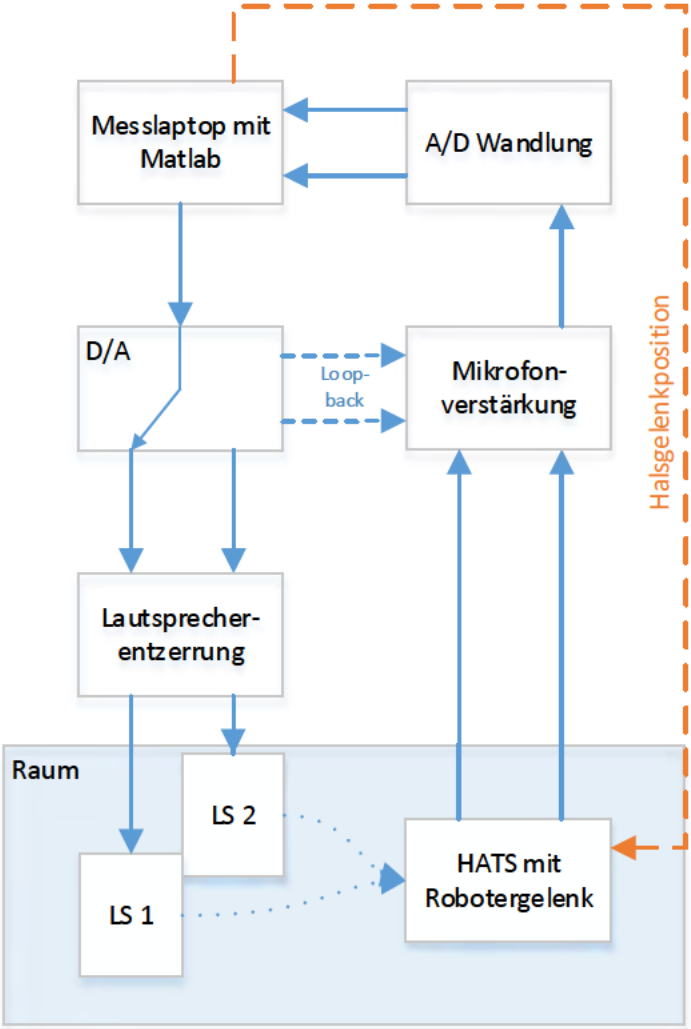


Abbildung 3.6.: Grundprinzip aller BRIR-Messungen, die Lautsprecherentzerrung wurde nur im Konzertsaal genutzt

vor dem Maximum gesetzt, bezeichnet als *peak secure margin*. Der optimale Schwellwert des Onset-Algorithmus ist iterativ zu bestimmen. So kann es vorkommen, dass starke Seitenreflexionen lauter als der Direktschall sind und somit ein fehlerhaftes Ergebnis verursachen. Daher wird jeder Datensatz optisch überprüft und gegebenenfalls händisch optimiert. Das Ergebnis einer solchen Laufzeitextraktion ist in Abbildung 3.7 dargestellt.

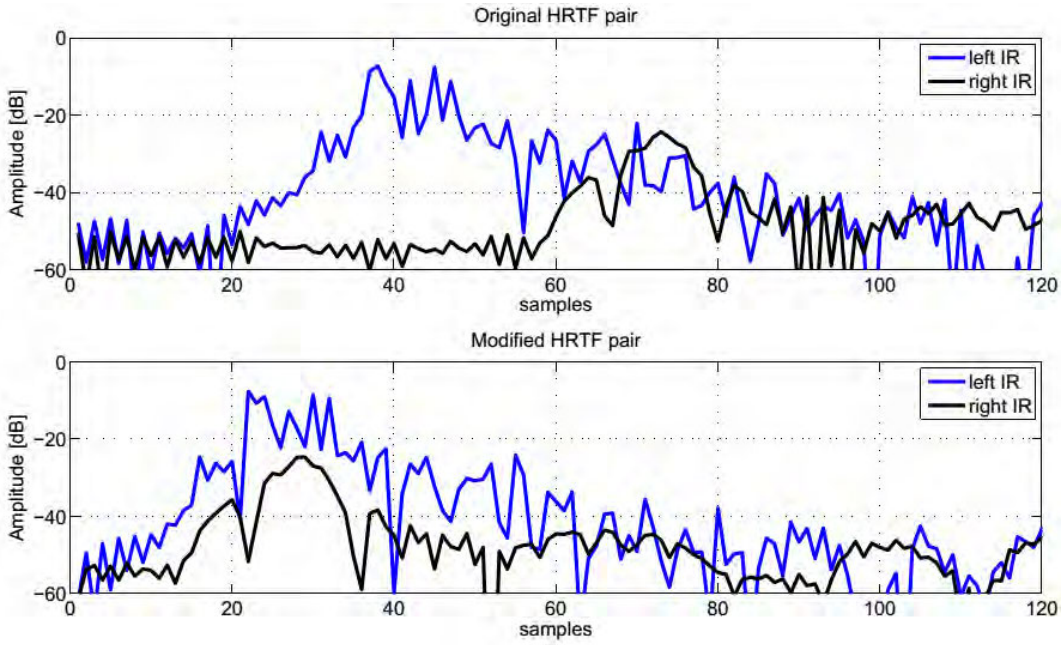


Abbildung 3.7.: Extraktion quasiminimalphasiger Impulsantworten aus einer HRTF über die '*onset detection*' Methode, die beobachteten Hüllkurvenänderungen haben keine perzeptiven Auswirkungen, aus [38, S.24]

Eine weitere Optimierung basiert auf der in Kapitel 2.5.4 eingeführten *perceptual mixing time*, die den Übergang zwischen statischem und positionsabhängigen Teil der Impulsantwort angibt. In [21] wurde ein Schätzer vorgestellt, der basierend auf einem empirischen Diffusitätsmaß [39] den Moment in BRIRs schätzt, bei dem der Übergang für 95% der Probanden unhörbar ist, Formel 3.2 (aus [22]). Die so erhaltenen Übergangszeiten wurden über alle BRIR-Datensätze gemittelt und auf das nächste 2^n Sample aufgerundet.

$$t_{mp95\%} = 1,77 \cdot t_{mix_{Abel_I}} - 38 \tag{3.2}$$

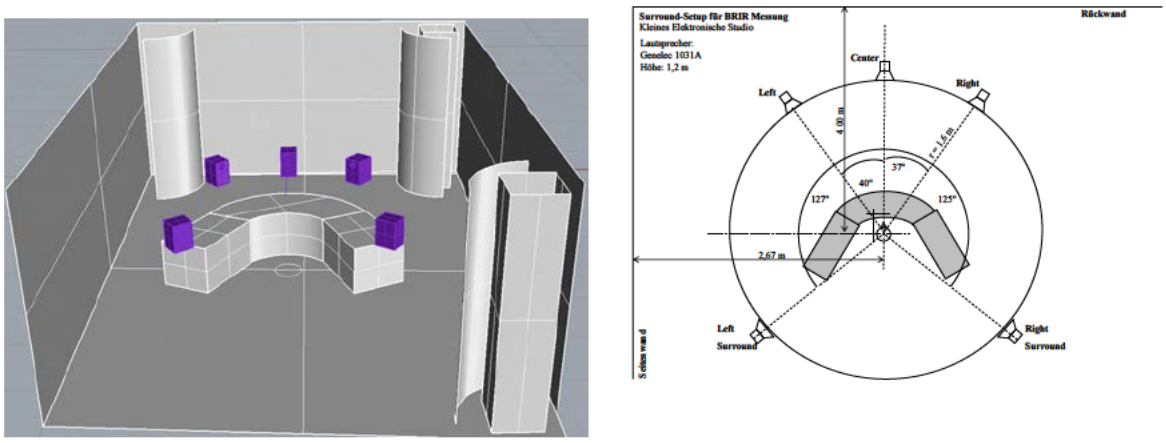
Daraus resultierend erhält man einen initialen Teil der BRIR, der kopfpositionsabhängig ausgewählt wird sowie einen statischen Nachhall, *tail* genannt, der aus der BRIR-Messung der neutralen Kopforientierung extrahiert und für alle Kopfpositionen benutzt wird.

3.2.1. Messung Tonstudio

Die Messung fand im Kleinen Elektronischen Studio der TU Berlin statt und war Grundlage der medialen Situationen Monolautsprecher, Stereolautsprecher und 5.0 Surround. Der Raum ist akustisch behandelt und mit einer Nachhallzeit von $t_{30} = 0,32s$ relativ trocken.

Der zu auralisierende Gegenstand ist hier also der Lautsprecher selbst in einem Raum, so dass keine Frequenzgangskorrekturen vorgenommen wurden. Als Quellen wurden Genelec 1031A Studiomonitore benutzt, Midfieldlautsprecher mit einem, nach Herstellerangaben [40], linearen Frequenzbereich von 47 Hz bis 22kHz.

Die fünf Quellen wurden nach in einer vom Tonmeister gewünschten leichten Modifikation der 5.0 Surround Aufstellung[5] positioniert, mit den Hochtönern jeweils auf 1,20 m Höhe und jeweils 1,60 m Entfernung zum FABIAN, welcher sich somit im 'sweet spot' befand.



(a) CAD-Modell

(b) Aufbauplan, minimaler Centerversatz von -1° nicht eingezeichnet

Abbildung 3.8.: Kleines Elektronisches Studio der TU Berlin

[5] 80° statt 60° Öffnungswinkel zwischen linkem und rechtem Lautsprecher

Quelle	**L**	**C**	**R**	**LS**	**RS**
LS-Azimutwinkel relativ zu 0°	-40	-1	37	-127	125
T_{30} in s	0,32				
FFT Ordnung des Sweeps	16				
Distanz Quelle-FABIAN in m	1,6				
peak secure margin in Samples	100				
ITD-*onset-threshold* in dB	-20				
Mittelwert $t_{mp95\%}$ in ms	48,2				
initialer BRIR-Teil in Samples	2^{12}				
Länge des synth. Nachhalls in Samples	20000				

Tabelle 3.2.: Studiomessung

3.2.2. Messung WFS-Labor

Das WFS-Labor der Telekom Laboratories besteht aus einem cirkulären, 56-kanligen Lautsprecherarray mit einem Durchmesser von 3 m. Der Lautsprecherring kann in der Höhe stufenlos variiert werden und wurde somit, um die medialen Lautsprechersituationen konstant zu halten, ebenfalls auf 1,20 m Tweeterhöhe eingestellt. FABIAN wurde ebenfalls mittig in dem 'sweet spot' positioniert. Daraufhin wurde ein BRIR-Datensatz von jedem Lautsprecher aufgenommen, resultierend in dementsprechend 56 Datensätzen.

Die Synthetisierung einer Primärquelle wird bei der Wellenfeldsynthese durch Zeit- und Pegelveränderungen aller Sekundärquellen (d.h. der 56 individuellen Lautsprechersignale) realisiert, wodurch jedem Lautsprecher eine sogenannte *Treiberfunktion* zugeordnet wird. Nach Erstellung der WFS-Szenen durch den Tonmeister, wurden die Treiberfunktionen $D(Quelle, LS)$ der Lautsprecher für jede simulierte Schallquelle (Primärquelle, d.h. Instrumente, Gesang, Nachhallquellen) bestimmt. Durch die Faltung dieser mit den entsprechenden BRIR-Datensätzen $BRIR(LS)$ und nachfolgender Superposition reduzieren sich die 56 Ausgangs-BRIR-Datensätze auf die Anzahl der genutzten Primärquellen. Demzufolge stehen am Ende der Bearbeitung je Szene 13

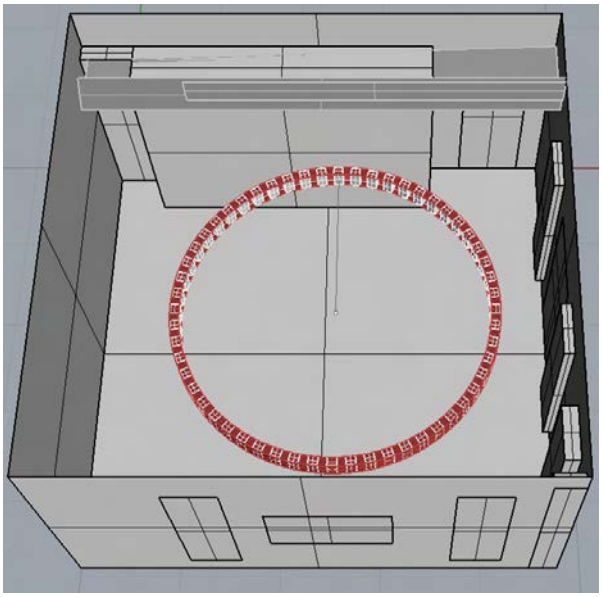


Abbildung 3.9.: CAD-Modell des Holophonielabors

BRIR-Datensätze zur Verfügung.

$$BRIR(Quelle) = \sum_{LS=1}^{56} D(Quelle, LS) * BRIR(LS) \tag{3.3}$$

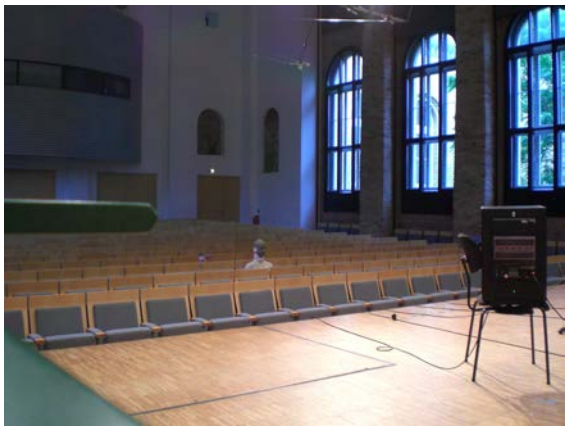
Auf den so erhaltenen Datensätzen wurde dann ebenfalls eine ITD-Entfernung und Bestimmung der *perceptual mixing time* durchgeführt.

T_{30} in s	0,22
FFT Ordnung des Sweeps	16
ITD-*onset-threshold* in dB	individuell je Quelle[6]
peak secure margin in Samples	individuell je Quelle[7]
Mittelwert $t_{mp95\%}$ in ms	134
initialer BRIR-Teil in Samples	2^{13}
Länge des synth. Nachhalls in Samples	20000

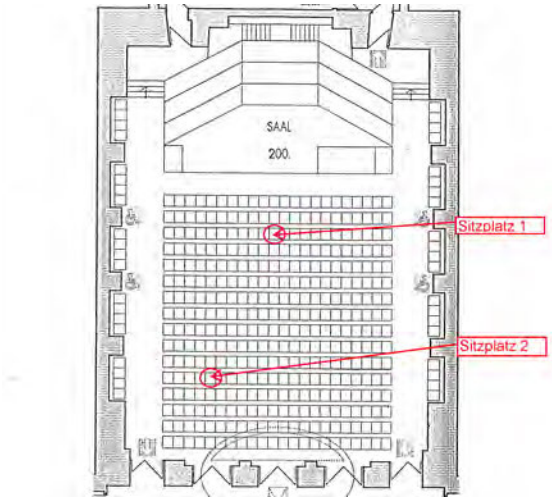
Tabelle 3.3.: Holophonielabormessung, die Daten beziehen sich auf die generierten BRIR-Primärquellen

Zusätzlich zu den 56 Sekundärquellen wird ein GENELEC 7060A Subwoofer verwendet, mit einem Frequenzgang von 29-120Hz (-6dB Punkte, nach Datenblatt). Der Pegel wurde schon bei der Szenengenerierung vom Tonmeister mehrmals angepasst und wurde daher in der Simulation ebenfalls nach subjektivem Urteil hinzugemischt, korrekte Pegelverhältnisse werden daher nicht benötigt. Die Subwoofer-WFS-Treiberfunktion wurde ebenfalls mit der BRIR komplex multipliziert, das Ergebnis normalisiert und die BRIR auf 2^{14} Samples gekürzt.

3.2.3. Messung Konzertsaal



(a) Foto des Saals mit FABIAN auf Platz 1



(b) Grundriss des Konzertsaals mit eingezeichneten FABIAN-Positionen

Abbildung 3.10.: BRIR-Messung

Im Joseph-Joachim-Konzertsaal der UdK Berlin wurden die BRIRs für die reale Situation an zwei unterschiedlichen Plätzen aufgenommen. Als ein 'guter/teurer' Platz vorne mittig in 5,47 m, als eher 'schlechter/billiger' Platz hinten links in 13,5 m Entfernung von der Stimmenquelle (Abbildung 3.10b). Im Gegensatz zu der Studio und WFS-

[7] ITD-*onset-threshold* für die konventionelle und kreative Szene, je Einzelinstrumente und ebene Wellen [-18 -20 -19 -20 -19 -19 -18 -19 -19 -19 -20 -18 -19] [-19 -21 -18,5 -21 -19 -22 -18,5 -19 -19 -21 -19,5 -18,5 -19,5]

[7] *peak secure margins* für die konventionelle und kreative Szene inklusive natürlicher Laufzeitunterschiede, je Einzelinstrumente und ebene Wellen [225 225 225 225 100 350 350 350 350 350 350 350 350] [180 120 180 180 100 360 360 360 360 360 360 360 360]

Laborsituation sollten diesmal nicht primär Lautsprecher, sondern reale Instrumente auralisiert werden.

Wie schon erwähnt, ist die Richtwirkung eines realen Instruments über Beamforming prinzipiell möglich, effiziente Messverfahren für mehrkanalige Lautsprecher liegen jedoch derzeit noch nicht vor.

Daher wurde der Aufbau mit den auf Achse linear entzerrten Meyer UPL-1 Lautsprechern aus Kapitel 3.1 beibehalten. Die dadurch verursachten Abweichungen vom Verhalten realer Instrumente müssen in der Ergebnisdiskussion kritisch beleuchtet werden. Im Zuge der Videoaufnahmen (s.u.) wurden Testaufnahmen mit realen Instrumenten vorgenommen und mit der Simulation verglichen. Die Unterschiede wurden hierbei als tolerierbar eingeschätzt. Die BRIR-Messung wurde nach Aufnahme der Mikrofonübertragungsfunktionen angefertigt und der Lautsprecheraufbau übernommen (S.125). FABIAN wurde auf dem vorderen als auch hinteren Sitzplatz jeweils mit der 0° Richtung auf die Position der Sängerin ausgerichtet.

Quelle	**vl1**	**vl2**	**vla**	**vc**	**voc**
T_{30} in s	1,81				
FFT Ordnung des Sweeps	18				
Distanz Quelle-FABIAN in m					5,47
ITD-*onset-threshold* in dB	-30	-30	-30	-18,5	-40
peak secure margin in Samples	110	200	210	120	100
Mittelwert$t_{mp95\%}$ in ms	104,9				
initialer BRIR-Teil in Samples	2^{13}				
Länge des synth. Nachhalls in Samples	100000				

Tabelle 3.4.: Konzertsaal Platz 1, *peak secure margin* inklusive natürlicher Laufzeitunterschiede

3.2.4. Kopfhörersimulation

Eine der heutzutage häufigsten Rezeptionssituationen stellt die Musikwiedergabe über Kopfhörer dar. Diese zeichnet sich durch eine statische, kopfpositionsunabhängige Wahrnehmung und eine Lokalisation der Schallquellen im Kopf aus. Dazu wurden Übertragungsfunktionen von typischen in-ear-Kopfhörern (Apple iPhone) vermessen,

Quelle	**vl1**	**vl2**	**vla**	**vc**	**voc**
T_{30} in s	1,81				
FFT Ordnung des Sweeps	18				
Distanz Quelle-FABIAN in m					13,5
ITD *onset-threshold* in dB	-25	-25	-25	-15	-30
peak secure margin in Samples	100	200	240	210	140
Mittelwert $t_{mp95\%}$ in ms	104,9				
initialer BRIR-Teil in Samples	2^{13}				
Länge des synth. Nachhalls in Samples	100000				

Tabelle 3.5.: Konzertsaal Platz 2, *peak secure margin* inklusive natürlicher Laufzeitunterschiede

die in die Ohrmuschel FABIANs eingesetzt wurden. Die so erhaltene HpTF wurde auf 2^{11} Samples gekürzt.

3.2.5. Pseudobinaurale Hintergrundgeräusche

Abbildung 3.11.: Das MTB-Array während der Konzertsaalaufzeichnung, aus [32].

Wie bereits erläutert wurde, stellt die soziale Interaktion eine wesentliche Qualität eines Konzertbesuches dar. Mit dem in Kapitel 2.6 vorgestellten MTB-Array wurden daher im Rahmen einer Magisterarbeit [32] 24-kanalige Aufnahmen eines Konzerts im Joseph-Joachim-Konzertsaal erzeugt (Abbildung 3.11).

Die 24 Audiospuren wurden vom Autor in eine DAW geladen, um zunächst Abschnitte relativer Ruhe herauszuschneiden. Relative Ruhe bedeutet in diesem Kontext, dass keine Musiker, Schritte oder sehr dominante Störgeräusche zu hören waren, sondern lediglich

die Atmosphäre eines besetzten, aber ruhigen Saals. Die so erhaltenen, meist wenige Sekunden langen Ausschnitte wurden mittels Überblendungen mehrmals verschachtelt aneinander gefügt, bis eine zusammenhängende, ca. 120 Sekunden lange Hintergrundatmosphäre entstanden ist. Anfang und Ende der Spuren wurden so gewählt, dass eine problemlose, artefaktfreihe Wiedergabe in einer Schleife möglich ist.

Nach der Bearbeitung entstanden Zweifel an der Aufnahmequalität bezüglich Signal-Rausch-Abstand. Durch das Ins-Verhältnis-Setzen von Grundrauschen zum lautesten Nutzsignal (ein Husten) wurde ein geringer SNR von ca. 20dB bestätigt. Mittels Anwendung von Filter- und Entrauschungsalgorithmen konnte allerdings der SNR um 9dB erhöht werden, ohne den Originalklang wesentlich zu beeinträchtigen. Ein weiteres Problem entstünde nun, wenn das Rauschen der Atmo-Aufnahmen in der kalibrierten Wiedergabeumgebung noch hörbar wäre. Dies konnte jedoch erst final in der Hörversuchsumgebung überprüft werden. Im Falle der Hörbarkeit hätte man im Sinne der Konsistenz aller Bedingungen dasselbe Grundrauschen ohne Publikumsgeräusche allen Stimuli beimischen müssen. Die Signalkette mit den Programmparametern ist im Anhang ab Seite 131 aufgeführt.

3.3. Akustische Simulationsumgebung

Dieses Kapitel gibt eine Übersicht über die am Fachgebiet bestehende dynamische binaurale Simulationsumgebung. Bestehend aus einem Rechnersystem mit mehreren Softwaremodulen und einer Wiedergabesystem, bestehend aus extraauralen Kopfhörern und einem dazugehörigen rauscharmen Verstärker. Zunächst werden die einzelnen Softwaremodule kurz vorgestellt, um danach einen Einblick in die vom Autor entworfene Implementierung zu geben.

Die Softwareumgebung benötigt ein Linuxsystem mit dem Jack Audio Framework [41] für ein flexibles Audiorouting. Die dann erforderlichen Module sind:

- **fWonder**. Eine Software zur zeitvarianten schnellen Faltung, die Impulsantworten in Abhängigkeit der momentanen Kopfposition austauschen kann.
- **stretcher**. Eine Software, die zuvor extrahierte ITDs in Echtzeit und in Abhängigkeit vom Kopfdurchmesser der Versuchsperson wieder einfügen kann.
- **tracker3**. Eine Software, die Positionsdaten eines Headtrackers an fWonder und stretcher weiterleiten kann.
- eine statisches FIR-Filter zur Entzerrung des Kopfhörers, genutzt wurde hierfür **fWonder**
- zwei weitere statische FIR-Filter zur Realisierung der Rezeptionssituation mit In-Ear Kopfhörern und zur Faltung der Subwoofer-BRIR, genutzt wurde **jconvolver**.
- **jack-rack**. Eine Sammlung von Audioplugins, genutzt als Raum-EQ für das Subwoofersignal
- **MTB-Renderer** zur Generierung eines pseudobinauralen Audiostreams aus den 24-kanaligen MTB-Aufnahmen.
- **Pure Data (Pd)** für die Audiowiedergabe der noch nicht verräumlichten Audiosignale, als Kanalmatrix und Kommunikationsschnittstelle für stretcher, den Videorechner und das Hörversuchssystem.

Abbildung 3.12 zeigt die prinzipielle Signalkette anhand einfacher Beispiele. Das einkanalige Eingangssignal, beispielsweise in der Situation 'Livekonzert' einer nachhallfreien Instrumentenaufnahme, wird durch Pd abgespielt und in fWonder mit einer

kopfpositionsabhängigen BRIR gefaltet. Diesem so entstehenden binauralen Signal wird dann im *stretcher*, wiederum kopfpositionsabhängig, die ITD hinzugefügt. Das Wiedergabesystem wird dann letztendlich über eine statische Kopfhörerübertragungsfunktion linearisiert. Die *tracker3*-Software sendet die Positionsdaten an die *stretcher*- und *fWonder*-Module über OSC [42], einem Netzwerkprotokoll zum einfachen Austausch von Steuerdaten, an einen oder mehrere Rechner. Daher kann die Audiosimulation aus Lastgründen gegebenenfalls auf mehrere Rechner verteilt werden.

3.3.1. Parametrierung der Softwaremodule der Audiosimulation

Der folgende Abschnitt erläutert kurz die Prametrierung aller Module der Audiosimulationsumgebung.

Routingsoftware Jack Audio Connection Kit

Das Jack Audiosystem ist eine Software, welche Audiorouting zwischen voneinander unabhängigen Audioprogrammen ermöglicht und somit eine virtuelle Patchbay bereitstellt. Die Blockgröße wurde auf 256 Samples gesetzt.

Faltungssoftware fWonder

FWonder ist eine interaktive, echtzeitfähige Faltungssoftware, die dazu dient beliebige Audiosignale mittels gemessener oder simulierter BRIR-Datensätze zu auralisieren und die auf über OSC ankommende Positionsdaten reagiert. Auch wenn ein fWonder prinzipiell mehrere Quellen gleichzeitig rendern kann, wird intern bereits im Frequenzbereich ein Downmix auf ein binaurales Stereosignal vollzogen. Da die noch einzufügende ITD aber quellenspezifisch ist, muss jede Quelle mit einer individuellen Instanz der Faltungssoftware auralisiert werden und somit müssen so viele Instanzen geöffnet werden wie Quellen vorhanden sind.

Wie bereits erwähnt, kann die Impulsantwort in einen kopfpositionsabhängigen initialen Teil mit den ersten Reflexionen und einen unabhängigen statischen Nachhall aufgetrennt werden. Hier muss die früher ermittelte *perceptual mixing time* als Übergangswert in Samples entsprechend angegeben werden.

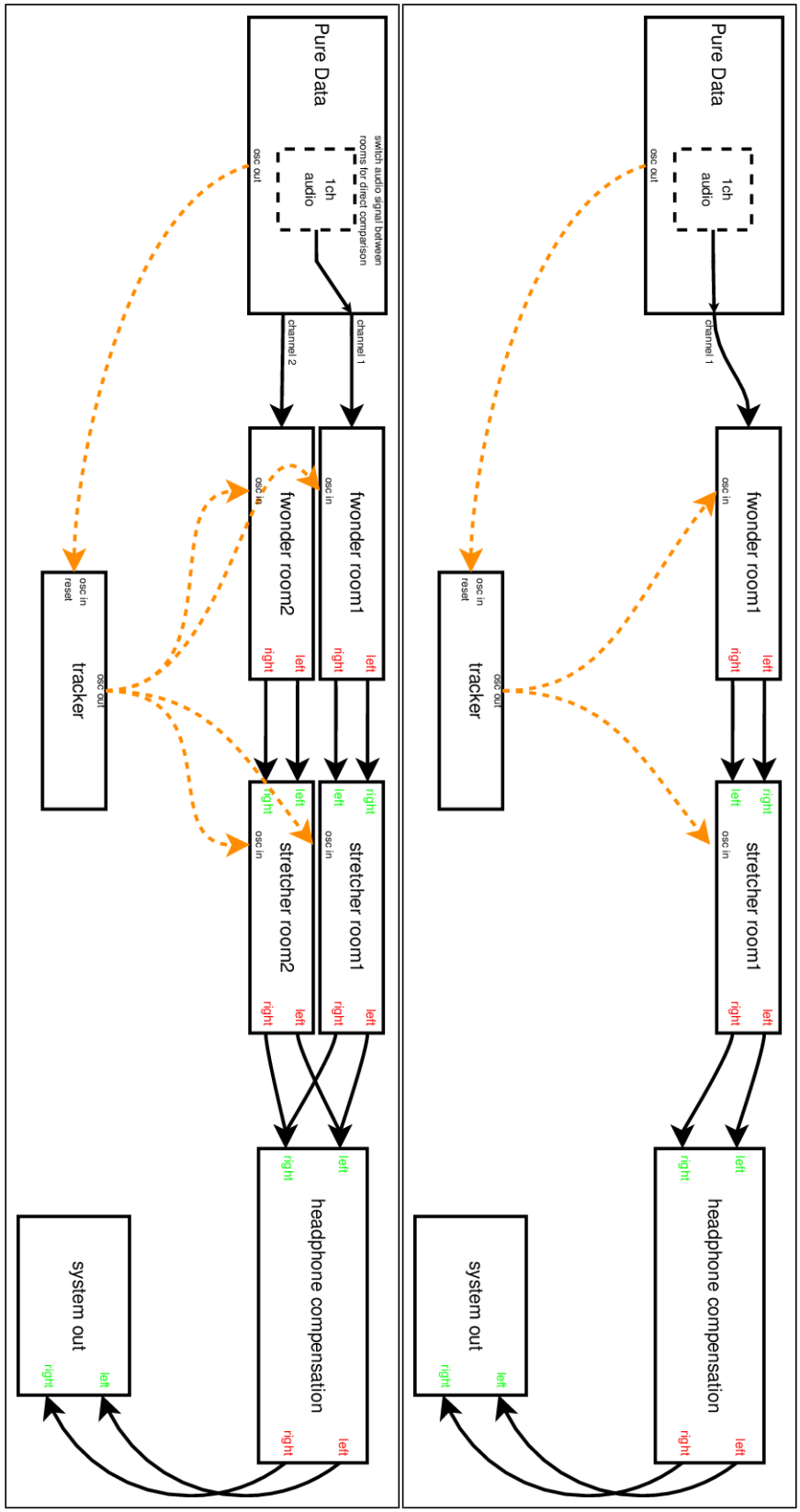


Abbildung 3.12.: Das obere Diagramm zeigt die Signalkette für einen Raum mit einer Schallquelleupper, das untere die Signalkette für zwei Räume mit jeweils einer Schallquelle. Durchgezogene Linien kennzeichnen Audiosignale, gepunktete OSC-Daten.

Die horizontale Auflösung des BRIR-Datensatzes wurde aus Gründen der Rechnerbelastung auf 2° festgelegt, auch wenn der Datensatz in 1° Auflösung vorlag. Laut der in Kapitel 2 zitierten Forschungsergebnisse ist bei einer Schrittweite von 2° nicht mit hörbaren Artefakten zu rechnen (siehe auch Tabelle 2.1).

ITD-Individualisierungssoftware *stretcher*

In diesem Softwaremodul wird die im Postprocessing aus dem BRIR-Datensatz entfernte ITD in Echtzeit und Abhängig von der individuellen Kopfposition wieder eingefügt. Dazu wird für jede Versuchsperson mit Hilfe eines Außentasters die Entfernung zwischen den in Abbildung 2.6 gezeigten Traguspunkten bestimmt. Diese Entfernung kann in Millimetern entweder dem *stretcher* bei Programmaufruf als Argument übergeben oder über eine OSC-Nachricht als Skalierungsfaktor gesendet werden. Die Umrechnung von Intertragusdistanz d_i in Millimetern auf Skalierungsfaktor S ist in Formel 3.4 gegeben (aus [18, Gl.6]).

$$S = 0.00304 \cdot d_i + 0,5792 \tag{3.4}$$

Die ITD wird über eine *sample rate conversion* als zeitvariable Verzögerung in das Audiosignal des schallquellenabgewandten Ohres eingebracht. Für die Abtastratenänderung wurde das Verfahren mit der größten Bandbreite und somit auch dem größten Rechenaufwand ausgewählt.

Kompensation der Kopfhörerübertragungsfunktion

Idealerweise sollte das Wiedergabesystem, bestehend aus Verstärker, Kopfhörer und der Strecke zwischen dem Kopfhörertreiber und dem Ohrkanaleingang, keinen Einfluss auf die gesamte Übertragungsstrecke ausüben, sofern auf individuelle BRIRs zurückgegriffen wird. Dies würde individuelle Messungen der versuchspersonenspezifischen HpTFs bedingen. Bei Verwendung nicht individueller, mit einem HATS aufgenommener BRIR-Datensätze stellte sich die größte perzeptive Übereinstimmung mit der Realität nach der Entzerrung des Aufnahmekopfes ein [27]. Daher wird auf dem verwendeten Kunstkopf die Übertragungsfunktion des Wiedergabesystems ermittelt. Auf Basis dieser Übertragungsfunktion wird über eine hochpassregularisierte LMS-Entzerrungsmethode ein Filter erzeugt, das eben diese Strecke linearisiert. Gleichzeitig werden die auch bei der Akquise benutzten Mikrofone kompensiert.

Dieses Filter wird in der Signalkette als Letztes mit Hilfe einer Faltungssoftware implementiert. Hierfür wird ebenfalls die fWonder-Software benutzt, konfiguriert als kopfpositionsunabhängiges, statisches FIR-Filter.

Abbildung 3.13.: FABIAN HATS mit aufgesetztem BK211.

Wiedergabe von Konzertsaalatmosphäre: MTB-Renderer

Zur Untersuchung des sozialen Kontextes wurden 24-Kanal Atmos aufgenommen, geschnitten und nachbearbeitet. Die Anwendung MTB-Renderer erwartet am Eingang 24 Audiokanäle und OSC-Nachrichten mit Kopfpositionsdaten und erzeugt einen zweikanaligen, pseudobinauralen Audiostream. Als Interpolationsalgorithmus wurde der in [31] perzeptiv am besten Beurteilte *Two Band Spectral-Interpolation Restoration* verwendet.

Audiowiedergabe und 'Kommunikationszentrale': Pure Data

Pd ist eine frei erhältliche, objektorientierte, visuelle Programmiersprache, mit Fokus auf Audio- und Bildsignalverarbeitung. Die Anwendung dient zur Wiedergabe der noch nicht auralisierten Audiospuren, als Kanalmatrix, zur Weitergabe von Zeitinformationen an das Videosystem zur Bild/Ton-Synchronisierung und Kommunikation mit den stretcher-Modulen zur Anpassung der ITDs an die momentane Versuchsperson. Die Software

unterstützt mehrere Netzwerkprotokolle und kann somit fernbedient werden.

Implementierung

Alle Audioinhalte zusammengenommen, erhält man je Musikstück 72 verschiedene Audiospuren, aufgeteilt auf 53 musikstückspezifische Spuren und zusätzlich 24 Spuren für das Rendern der Konzertsaalatmosphäre. Diese 72 Audiokanäle müssen auf 68 Eingangskanäle der individuellen Faltungsprozesse aufgeteilt werden. Schon zu Beginn der Simulatorkonzeption war ersichtlich, dass diese Arbeitslast nicht von einem Rechnersystem bewältigt werden würde. Trotzdem wurde der Simulator zunächst auf nur einem PC erstellt. Durch den modularen Ansatz konnte das gesamte System auf einen zweiten PC kopiert und nicht benötigte Module deaktiviert werden. Daher wird zunächst auch nicht zwischen verschiedenen Audiorechnern unterschieden werden. Es wird ebenfalls nicht mehr unterschieden zwischen dem vierkanaligen klassischen Stück und dem fünfkanaligen populären, sondern immer von der maximalen Kanalanzahl ausgegangen. Abbildung 3.15 zeigt den gesamten Signalverlauf.

Im Detail teilen sich die Audiospuren auf in:

- Fünf Spuren für die beiden Konzertsaalsituationen. Jeder Sitzplatz wird durch jeweils fünf fWonder mit jeweils eigenem, ein Instrument simulerenden BRIR-Datensatz auralisiert.

- Insgesamt 17 Spuren für die mediale Wiedergabe im Studio, auralisiert durch fünf fWonder mit jeweils einem, einen Lautsprecher simulierenden Datensatz.

- 13 Spuren für die konventionelle WFS-Abmischung, bestehend aus fünf Instrumentensignalen und acht Nachhallsignalen, auralisiert über 13 fWonder.

  Ein fWonder simuliert hierbei die Summe der 56 Lautsprecher, konfiguriert für jeweils eine instrumentenspezifische Primärquellenposition (Punktquelle) und acht ebene Wellen, über die der Nachhall eingespielt wird. Zusätzlich wird eine statische Subwoofer-BRIR mit einer Monomischung der 13 Spuren und nachfolgender Filterung zur Raummodenkompensation via *jack-rack* bespielt.

- 13 Spuren für die kreative WFS-Abmischung, bestehend aus fünf Instrumentensignalen und acht Nachhallsignalen, auralisiert über 13 fWonder. Der prinzipielle Signalverlauf gleicht dem der konventionellen WFS-Abmischung.

Die Situation *Kopfhörer* greift auf zwei der drei Stereospuren der medialen Studiowiedergabe zurück. Die auf den Hauptmikrofonierungsarten XY und Groß-AB basierenden Abmischungen wurden hierfür an ein statisches FIR-Filter mit der in-ear-Kopfhörer-HpTF geschickt, anstatt an die Stereolautsprecher der Studioszene.

Da, wie bereits erwähnt, alle produzierten Studio- und WFS-Abmischungen auf den Beginn der nachhallfreien Aufnahmen synchronisiert wurden, können diese nun in Pd gleichzeitig abgespielt und instantan artefaktfrei umgeschaltet werden.

Die Kommunikation mit der Hörversuchsumgebung *WhisPER*[8] erfolgte ebenfalls über OSC-Daten unter Verwendung einer die Bedingungsvariationen beschreibenden Codetabelle. 14 verschiedene akustische Situationen und zwei musikalische Inhalte resultieren in insgesamt 28 verschiedenen Stimuli, in *WhisPER* und Pd repräsentiert durch die Zahlen 1-28 auf dem OSC-Pfad `/situation`. Des weiteren wird über den Pfad `/stop` ein Starten oder Stoppen des Stimulus ausgelöst.

Wird beispielsweise die Situation *Konzertsaal Platz 2* gewünscht, werden die fünf nachhallfreien Audiosignale an die Pd-Ausgangskanäle 6-10 geschickt, die wiederum verbunden sind mit den einzelnen fWonder-Prozessen, die den BRIR-Datensatz für Konzertsaal Platz zwei geladen haben. Die Ausgänge aller anderen Mischungen werden stumm geschaltet, die Wiedergabe läuft aber trotzdem synchron weiter. Wird der andere Sitzplatz gewünscht, werden dieselben nachhallfreien Audiospuren an die Pd-Ausgänge geschickt, die mit den entsprechenden fWonder verbunden sind. Wird eine Studiosituation mit der XY-Stereomischung gewünscht, wird ebendiese Stereomischung 'laut gezogen' und an die Pd-Ausgangskanäle geschickt, die mit den Studio-fWonder verbunden sind, während die nachhallfreien Signale stummgeschaltet werden.

Die gewünschten Routing- und Verstärkungsoperationen werden durch die OSC-Eingangsbefehle ausgelöst, der prinzipielle Ablauf ist in Abbildung 3.14 dargestellt. Ein Wechsel innerhalb der Studiosituationen wird über zusätzliche crossfades realisiert, um hörbare Artefakte zu vermeiden.

Sowohl die Pegel der Pd-Ausgänge und damit die Lautstärke der jeweils simulierten Wiedergabelautsprecher, als auch die Pegel der Audiodateien und somit die Lautstärken der einzelnen Abmischungen, konnten separat über interne Variablen gesetzt werden. Die Werte wurden bei der später beschriebenen Lautheitskalibrierung festgesetzt.

[8] zu finden unter *Digitale Ressourcen*, *Research Tools* auf `http://www.ak.tu-berlin.de` (Stand September 2013)

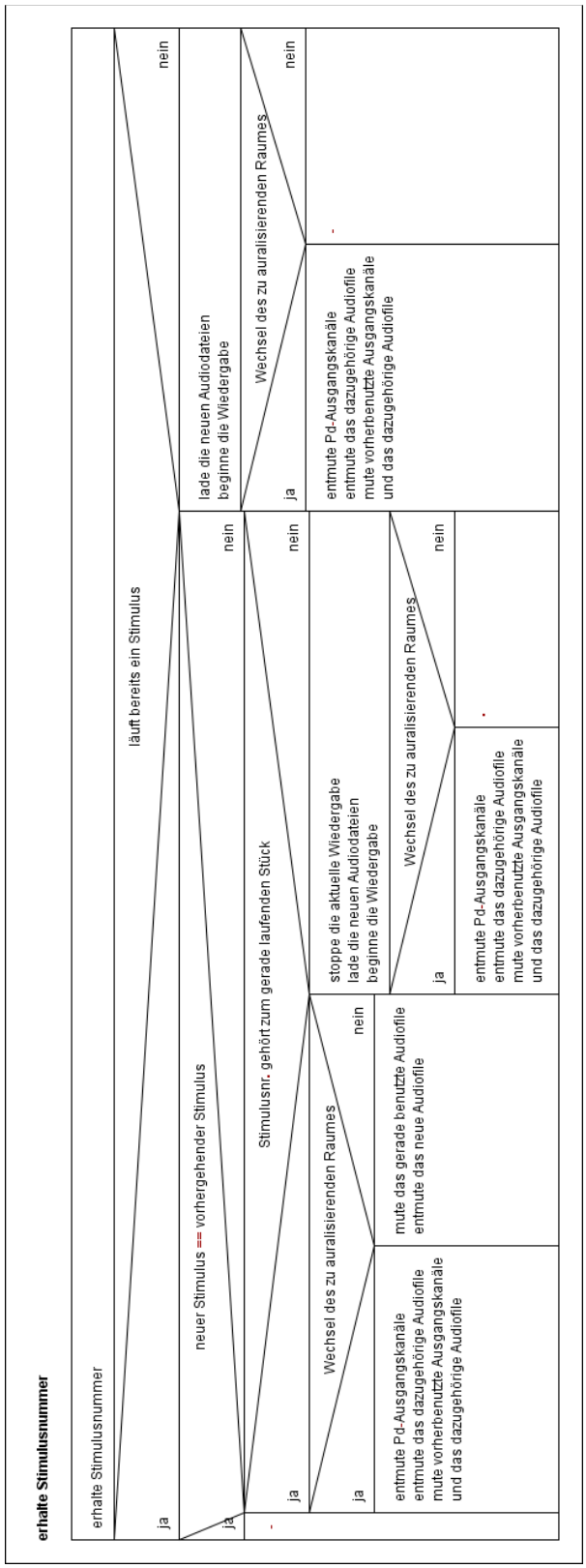


Abbildung 3.14.: Struktogramm Stimuluswechsel

Der Start aller Programme sowie aller Routingkonfigurationen wurde über Linux-Shellskripte realisiert.

Wie schon eingangs erwähnt, musste die Audiosignalverarbeitung letztendlich auf zwei PCs aufgeteilt werden. Dabei war einer der Rechner für die Auralisierung der Studio- und Konzertsaalsituation mit dem MTB-Renderer zuständig, der andere für die WFS-Labor Situationen. Der mit dem Wiedergabesystem verbundene Hauptrechner war mit einem RME Multiface als Audiointerface ausgestattet. Es verfügt über acht analoge symmetrische Ein- und Ausgänge sowie weitere Adat und SMTP/AES-EBU Schnittstellen und einen DSP, auf dem Ein- und Ausgangskanäle nahezu latenzfrei (<1ms) zusammengemischt werden können. Der zweite, nahezu baugleiche Rechner verfügt über eine M-Audio Audiophile 192 Audiokarte mit je zwei symmetrischen Ein- und Ausgängen. Die binaurale Summe des Holophoniesystems wurde mit symmetrischen Kabeln analog in das Multiface geführt. Dort wurde auf dem RME-DSP das Signal sofort auf die Ausgangskanäle zum Wiedergabesystem geroutet. Die im Hauptrechner ankommenden Steuerbefehle des Hörversuchssystem wurde sofort an den 'Holophonierechner' weitergeleitet.

Latenz

Zur Verifizierung der totalen Systemlatenz (TSL), das heißt der Reaktionszeit des Systems auf eine Kopfpositionsänderung, führte der Autor mit dem Simulationssystem im Endzustand die in [23] gezeigte Messung durch.

Dazu wurde der Sensor des Trackingsystems auf einem mechanischen Schwingarm montiert. Bewegte man diesen aus seiner Ausgangsposition, wurde ein elektrischer Gleichstromkreis unterbrochen, der zu einem Spannungsabfall über einem Stromteiler führte und als Beginn der Kopfbewegung mit einem Speicheroszilloskop millisekundengenau erfasst werden konnte. Die resultierende Systemreaktion wurde mit einem zweiten Eingang des Speicheroszilloskops erfasst.

Ein BRIR-Datensatz wurde durch einen Testdatensatz ersetzt, unter Beibehaltung der gesamten, finalen Signalkette. In neutraler Kopfposition wurde nur Stille gerendert, während die Impulsantworten aller anderen Kopfpositionen einen Impuls enthielten. Wurde das System nun mit einem Rechtecksignal bespielt, wurde dieses ausschließlich bei von der Ausgangsstellung abweichenden Schwingarmpositionen wiedergegeben. Wurde der Schwingarm nun bewegt, fiel die Gleichspannungsflanke instantan und kurz

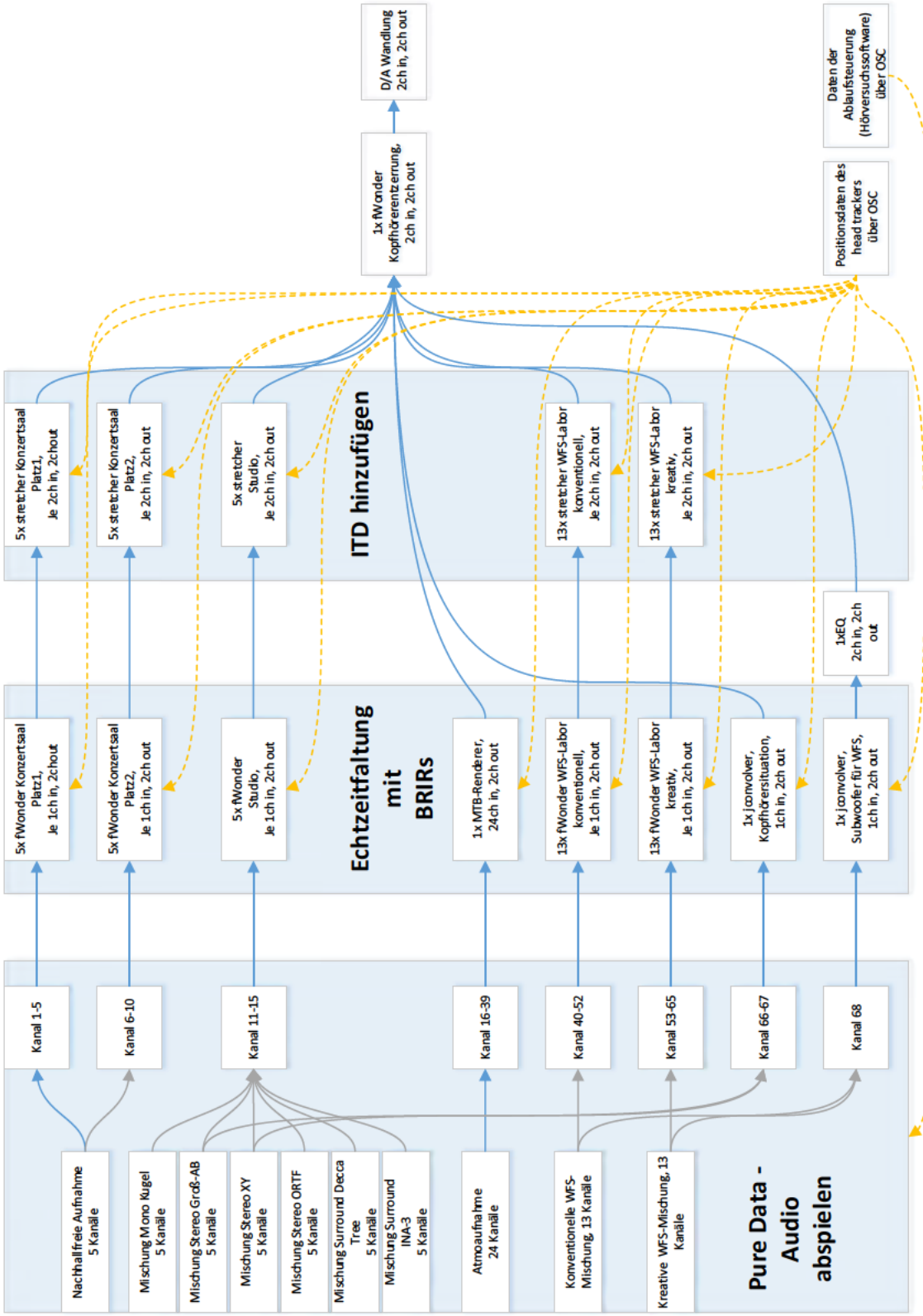


Abbildung 3.15.: Darstellung des Signalflusses des Audiosystems für die Situation Konzertsaal Platz 1 inklusive Einspielung der pseudobinauralen Konzertatmosphäre. Blaue Pfeile zeigen aktive Verbindungen an, graue deaktivierte, gelbe OSC Daten.

darauf reagierte das System mit der Wiedergabe des Rechtecksignals. Der Zeitraum zwischen beiden Ereignissen ist die totale Systemlatenz. Dieses Experiment wurde 30 Mal wiederholt. Zu dem erhaltenen Mittelwert (83,13 ms) wurde daraufhin jeweils situationsspezifisch die maximale Laufzeitverzögerung und die Latenz des Wiedergabesystems (2,52 ms) addiert. Die Latenz betrug damit für die Livesituationen 91,1 ms, für die Studiosituationen 87,92 ms und für die WFS-Situationen 93,82 ms.

3.3.2. Kopfhörer BK211

Die binaurale Wiedergabe sollte über ein möglichst transparentes System erfolgen, das sich durch einen linearen Frequenzgang und einen hohen Signal-Rausch Abstand auszeichnet, des weiteren das FEC-Kriterium erfüllt und robust gegenüber HpTF-Änderungen durch Neupositionierungen des Kopfhörers ist. Das am Fachgebiet entwickelte und gebaute Kopfhörersystem BK211, bestehend aus einem extraauralen Kopfhörer, einem Dsp-System für IIR-, FIR-Filterung, sowie Thermo- und Clip-Limiting und einer rauscharmen Endstufe, erfüllt all diese Eigenschaften. So liegt sein Eigenrauschen unter der Hörschwelle, mit einem Frequenzgang von $\pm 3dB$ zwischen 60Hz und 18kHz (Abbildung 3.16), einem maximalen Schalldruckpegel von ca. $109dB_{SPL}$ am Ohrkanaleingang und der Erfüllung des FEC-Kriteriums.

Aufgrund eines Fehlers in der Firmware des DSPs konnte es vorkommen, dass der Masterfader unter Umständen bei einem Neustart des Systems nicht mehr auf der 0dB Position war. Daher wurde bei späteren Hörversuchen nach jedem Neustart die DSP-Konfiguration überprüft.

3.3.3. Akustische Kalibrierung

Der Hörversuch fand im sogenannten *Medialab* des Fachgebiets statt, bestehend aus einem Maschinenraum und einem komplett abgedunkelten Raum mit einer halbzylindrischen Projektionsfläche, fünf Projektoren, dem BK211 System und einem Polhemus FASTRAK Tracker zur Verfolgung der Kopfbewegungen(im folgenden Cave genannt). Die Kalibrierung wurde mit folgenden Zielen durchgeführt:

- Linearisierung des Wiedergabesstems.
- Lautheitsnormierung der medialen Bedingungen.

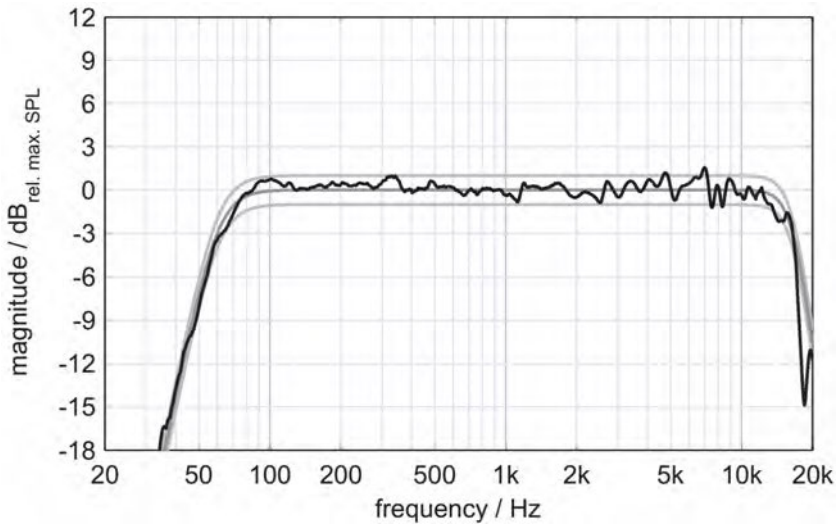


Abbildung 3.16.: Betragsfrequenzgang der Kopfhörer BK211 mit der BKamp Endstufe im Freifeld, aus [43]

- Absolutpegel der Livesituation und Sitzplatzunterschiede kalibrieren.
- Pegel der Hintergrundatmosphäre kalibrieren.
- Subwooferpegel der modernen medialen Situationen wählen.

Für die akustische Kalibrierung wurde vom Autor zunächst der Störpegel im Raum ermittelt, das Wiedergabesystem linearisiert und letztendlich eine Lautheitskalibrierung der Stimuli durchgeführt.

Grundstörgeräusch

Die Messung des Grundstörgeräusches wurde mit einem NTI Acoustilyzer AL-1, einem Schallpegelmesser mit einem kalibrierten Klasse 2 NTI MiniSPL 1/2“ Elektretmikrofon durchgeführt. Die Messung wurde im Endzustand der Cave bei geschlossenen Fenstern, aktivierten Beamern im lauteren 3D-Modus, gleichzeitiger Projektion eines weißen Bildes und in Versuchspersonenohrhöhe über dem später benutzten Stuhl durchgeführt (Abbildungen 3.6 und 3.7). Das Grundstörgeräusch war somit zwischen den Stimulusdarbietungen unkritisch leicht wahrnehmbar.

Leq_{slow}/dB_{SPL} unbewertet	56,6
Leq_{slow}/dB_{SPL} (C)	49,3
Leq_{slow}/dB_{SPL} (A)	33,2

Tabelle 3.6.: Grundstörgeräusch in der Cave.

Frequenz/Hz	**31,5**	**63**	**125**	**250**	**500**	**1k**	**2k**	**4k**	**8k**	**16k**
SPL/dB unb.	53	38,9	36,7	37,3	31,9	29,4	25,5	24,3	25,8	28,2
SPL/dB(C)	47,9	38	36,4	34,9	29,8	27,3	24,6	23,7	23,7	21,5
SPL/dB(A)	18,9	20,1	21,7	25,8	26,5	26,7	25,3	24,9	25	22,1

Tabelle 3.7.: Grundstörgeräusch in der Cave in Oktavauflösung.

Lautheitskalibrierung der Simulation

Bei den Videoaufnahmen für die 'Live'-Situation im Konzertsaal nahm ein reales Streichquartett die Positionen der Lautsprecher der BRIR-Messung ein. Diese Situation wurde genutzt, um kalibrierte binaurale Aufnahmen zu erzeugen, die später zur Einmessung der Simulation in der Cave dienen konnten. Zu diesem Zweck wurde das FABIAN-Messsystem auf der Empfängerposition 1 aufgebaut. Die Musiker wurden gebeten, jeweils das klassische und populäre Stück ausschnittsweise zu spielen, während die Mikrofonsignale des HATS in 0°-Position mit aufgenommen wurden. Nach messtechnischer Bestimmung der Pegel des Mikrofonvorverstärkers und unter Berücksichtigung der bekannten Mikrofonempfindlichkeiten verfügte man nun über eine Referenzaufnahme mit echten Instrumenten im echten zu auralisierenden Raum.

In der Cave wurde der FABIAN HATS mit dem BK211 Wiedergabesystem[9] am VP-Sitzplatz aufgebaut. Nach Wiederherstellung derselben Mikrofonvorverstärkungspegel wie bei der Referenzaufnahme wurde über das Simulationssystem das auralisierte Pendant wiedergegeben und aufgenommen. Mit Hilfe eines Matlabskripts wurde der SPL-Wert des aufgenommenen Ausschnitts berechnet und zum Wert der Referenzaufnahme verglichen. Die sich so bildende Differenz wurde in Pure Data jeweils stückspezifisch kompensiert.

[9] BKamp #001 und BK211 #001 mit Preset 'fullrange, analog in'

Die so erhaltene final entzerrte und bezüglich Kunstkopf lautheitskalibrierte Auralisation des Konzertsaals wurde daraufhin von zwei Expertenhörern subjektiv überprüft, ob noch weitere klangfarbliche oder pegelbezogene Korrekturen vorgenommen werden müssen. Da die Referenzaufnahme im Konzertsaal mit anderen Musikern als auch anderen Instrumenten im Vergleich zur nachhallfreien Einspielung getätigt wurde, sind Unterschiede in Klangfarbe, Dynamik und Tempo unvermeidlich. Hinzu kommt die unterschiedliche Richtcharakteristik der Messlautsprecher, dadurch eine andere Raumanregung und unterschiedliche Halleindrücke.

Aufgrund der schwierigen Gewichtung der einzelnen Unterschiede und der je nach nach Ausschnitt teilweise kongruenten, teilweise unterschiedlichen Lautheitseindrücken zwischen Referenz und Auralisation, wurde die berechnete Verstärkung beibehalten. Auch wurden keine Klangfarbenkorrekturen für nötig erachtet. Einzig die Balance zwischen den einzelnen Instrumenten wurde geringfügig korrigiert und die gesamte Kalibrierung mit den neuen Instrumentenverhältnissen wiederholt.

Zur Kalibrierung des hinteren Konzertsaalplatzes wurde die Lautstärke hier zunächst ebenfalls messtechnisch (durch Pegelvergleich) und perzeptiv (durch zwei Expertenhörer) ebenso laut eingestellt, wie bei der Situation am vorderen Sitzplatz. Als Stimulus wurde rosa Rauschen gewählt, abgespielt von der Position des Violoncellos. Dann wurde zuerst der Ausgangspegel der RME-Karte abgelesen und der Pegel der hinteren Quelle angepasst. Im anschließenden perzeptiven Vergleich mit dem klassischen Musikstück zwischen den beiden Plätzen wurde die Lautstärke am hinteren Sitzplatz noch als gering lauter beurteilt und daraufhin leicht abgesenkt. Das moderne Stück erschien nun hinten etwas leiser, es wurden aber keine weiteren Anpassungen mehr vorgenommen. Danach wurde der Lautheitsunterschied zwischen den beiden Sitzplätzen rekonstruiert. Dazu lagen aus raumakustischen Messungen Stärkemaße beider Sitzplätze vor. Die Differenz von $6,45dB$ zuungunsten des hinteren Sitzplatzes wurde dann mit Pegelkorrekturen in Pd realisiert.

Die Lautstärken der beiden medialen Umgebungen wurden in perzeptiver Beurteilung durch zwei Expertenhörer an die Lautstärke der vorderen Konzertsaalsituation angepasst, um mögliche Präferenzentscheidungen zugunsten lauterer Stimuli zu unterbinden. Die Lautheiten der einzelnen Abmischungen wurden beibehalten, da sie bereits (nach Anweisung) vom Tonmeister lautheitsangeglichen ausgespielt wurden, ein üblicherweise im Mastering stattfindender Prozess.

Die Lautheit des Subwoofers im simulierten WFS-Labor wurde nach Geschmack der

Expertenhörer eingestellt. Dabei fiel eine starke Verfärbung (monofrequentes Dröhnen) negativ auf, offensichtlich durch ungünstig ausgeprägte Raummoden am Hörerort im relativ kleinen WFS-Labor verursacht. Daher wurde ein geeignetes Raumkorrekturfilter in der Software Monkey Forest[10] erstellt und über einen parametrischen EQ im jack-Plugin *jack-rack* [11] realisiert (Tabelle 3.8).

Frequenz [Hz]	**Q**	**Gain [dB]**
36	4	-7,5
60	4	-8,5

Tabelle 3.8.: Parameter des Subwooferfilters

Da die benutzten BRIRs laufzeitbefreit sind (Delay ca. 3 ms), erschien das Subwooferhörereignis (Delay ca. 25ms) stark verzögert. Im die Subwooferfaltung realisierenden Plugin jconvolver[12] wurde dieser Unterschied durch eine neue Parametrierung ausgeglichen.

Am Ende musste der globale Pegel aller Situationen allerdings um 4 dB abgesenkt werden, da es im Ursprungszustand in *fortissimo* gespielten Passagen der Situation 'Konzertsaal Platz 1, populäres Stück' zu Übersteuerung bei der D/A-Wandlung kam.

Lautheitskalibrierung der MTB-Wiedergabe

Um die Umgebungsgeräusche lautheitskorrekt in die binaurale Simulation einzubetten, wurde der absolute Schalldruckpegel eines kurzen Ausschnitts (Husten) von zwei gegenüberliegenden Mikrofonen (±90° zur Frontalrichtung liegend) bestimmt. Die Messungen wurden durchgeführt mit dem MTB-Renderer im finalen Renderingmodus, neutraler Kopfposition und kompensierter HpTF. Zunächst wurden die Vorverstärkungswerte und die Mikrofonempfindlichkeiten bestimmt und dann der Schalldruckpegel für eine schmales Frequenzband im mutmaßlich noch linearen Übertragungsbereich des MTB-Arrays (Oktavbandpassfilter, 1kHz, 4.Ordnung) berechnet. Die erhaltenen Werte von $42,8/47,3dB_{SPL}$ für das rechte, bzw. linke Ohr. wurden erneut als plausibel eingestuft. Zur lautheitsrichtigen Wiedergabe wurde dasselbe Verfahren mit dem FABIAN

[10] `http://www.four-audio.com/de/produkte/monkey-forest.html` abgerufen September 2013
[11] `http://jack-rack.sourceforge.net/`
[12] `http://kokkinizita.linuxaudio.org/linuxaudio/index.html`

HATS durchgeführt, wie zur Kalibrierung der vorderen Konzertsaalsituation.

Final wurde die Verstärkung des Atmo-Signals trotzdem abermals um insgesamt 13.5dB angehoben, da insbesondere auf dem vorderen Sitzplatz die Hintergrundgeräusche von den Musikinstrumenten komplett maskiert wurden. Nichtsdestotrotz war die Atmo auch nach dieser Korrektur nur sehr dezent wahrnehmbar. Hierbei erwies sich außerdem die Effizienz der oben beschriebenen Entrauschung der MTB-Aufnahmen. So lag das Grundrauschen der MTB-Aufnahmen bei ca. 25 dB SPL und damit nach Anhebung bei etwa 38,5 dB SPL. Vergleicht man diesen Wert mit dem Grundstörgeräusch in der Cave, zeigt sich, dass das Grundrauschen der Atmo nun nur knapp überschwellig war. Im 'unentrauschten' Fall hätten die Atmo-Signale nicht angehoben werden können.

4. Videosimulation

Dieses Kapitel stellt die zur Aufnahme, Nachbearbeitung und Projektion von stereoskopischen Videos gewählten Verfahren dar. Hauptziele bei der Erstellung der optischen Stimuli waren eine möglichst originalgetreue Abbildung der verschiedenen Situationen in Geometrie, Farbtemperatur und Helligkeit unter konstanten Störvariablen. Die Erstellung optischer Stimuli teilt sich auf in:

- stereoskopische Aufnahmen und photometrische Messungen der Livesituation von beiden Plätzen aus für beide Musikstücke (vom Autor durchgeführt).
- Aufnahme und Produktion eines konventionellen Musikvideos mit Mehrkameraperspektiven für beide Musikstücke (von externer Firma durchgeführt).
- stereoskopische Aufnahmen und photometrische Messung der medialen Situationen mit und ohne Darstellung eines Fernsehers. Die konventionellen Videos wurde im Nachhinein in die stereoskopischen Aufnahmen mit Fernseher gesetzt (vom Autor durchgeführt).
- festlegen des Fixpunkts, Korrektur von Bildfehlern und -verzerrungen sowie Wahl eines geeigneten Wiedergabecodecs (vom Autor durchgeführt).
- Geometrie-, Farb- und Helligkeitskalibrierung (vom Autor durchgeführt).

4.1. Bildaufnahmen (konventionell und 3D)

4.1.1. Parameter der Bildaufnahmen

Die Videoaufnahmen fanden zeitlich gesehen nach den akustischen Messungen statt, Empfänger- und Quellenpositionen waren daher von vornherein festgelegt. Das optische Zentrum der genutzten Kamera entsprach somit jeweils der FABIAN-Kopfposition. Das

Sichtfeld (*field of view - fov*) der zu wählenden Kameras sollte es ermöglichen alle szenenrelevanten Objekte ohne Anschnitt zu beinhalten. Zur Ermittlung des benötigten Aufnahmewinkels wurden von einem weiteren Bearbeiter angefertigte CAD-Modelle verwendet, mit deren Hilfe Bildausschnitte für verschiedene Objektivtypen und Sichtfelder erstellt wurden. Die vordere Konzertsaalsituation hätte demnach einen Öffnungswinkel von 60° benötigt, während die Studiosituation sogar 80° benötigt hätte, um die vorderen Stereolautsprecher darstellen zu können. Auch wenn die medialen Szenen zunächst statisch sind und man hier auf Fotografien und größere Bildausschnitte zurückgreifen hätte können [35], sollten diese trotzdem mit derselben Technik wie die Livesituation aufgenommen werden, um Bildausschnitte und Aufnahmeartefakte über die verschiedenen Stimuli konstant zu halten.

Alle 3D-Aufnahmen wurden mit dem integrierten stereoskopischen digitalen Kamerasystem CP 31 von *3D One* angefertigt (siehe Abbildung 2.15). Dieses besteht aus zwei Optiken fester Basisbreite und einer synchronen Ansteuerung der Vergenz, Fokussierung und des Zooms. Die Aufnahmen wurden in für das linke und rechte Auge separaten Videodateien mit 1920 x 1080 Pixeln im Motion JPEG Format abgespeichert. Obwohl der aufgrund technischer Angaben theoretisch bestimmte *fov* bei 67,8° x 38,2° liegen sollte, wurde der tatsächliche *fov* nach Erhalt der Kamera zu 48° x 27° bestimmt. Eine Simulation des Bildausschnitts (Abbildung 4.1) zeigte einen Anschnitt der Musiker in der Livesituation und die Stereolautsprecher der Studiosituation außerhalb des Aufnahmebereiches. Aus organisatorischen und finanziellen Gründen konnte die Kamera nicht mehr kurzfristig ausgetauscht werden. Stattdessen wurde die Kamera im Konzertsaal um 2° nach oben geneigt, um einen Anschnitt der dargestellten Personen zu vermeiden. Aus Konsistenzgründen wurden die anderen Szenen ebenfalls mit dieser Elevation aufgenommen. Um den Versatz bei der Projektion berücksichtigen zu können, wurde jede Szene auch mit 0° Ausrichtung kurz gefilmt, um markante Objektgeometrien später in der Nachbearbeitung als Ankerpunkt zu benutzen.

Um Trapezverzerrungen zu vermeiden (siehe Kapitel 2.8.2), wurden die Objektive parallel ausgerichtet und der Fokuspunkt erst im Nachhinein durch eine horizontale Bildverschiebung in Deckung gebracht. Auf diesen Fokuspunkt (jeweils Gesichtsebene Sängerin, bzw. Musiker, bzw. TV-Bildschirm, bzw. Lautsprecher) wurde die Kamera jeweils scharf gestellt. Minimales Bildrauschen und hohe Tiefenschärfe wurden durch maximale Beleuchtung vor Ort und Nutzung möglichst großer Blendenzahlen erreicht. Um die absolute Helligkeit wie auch Farbtemperatur der Simulation kalibrieren zu

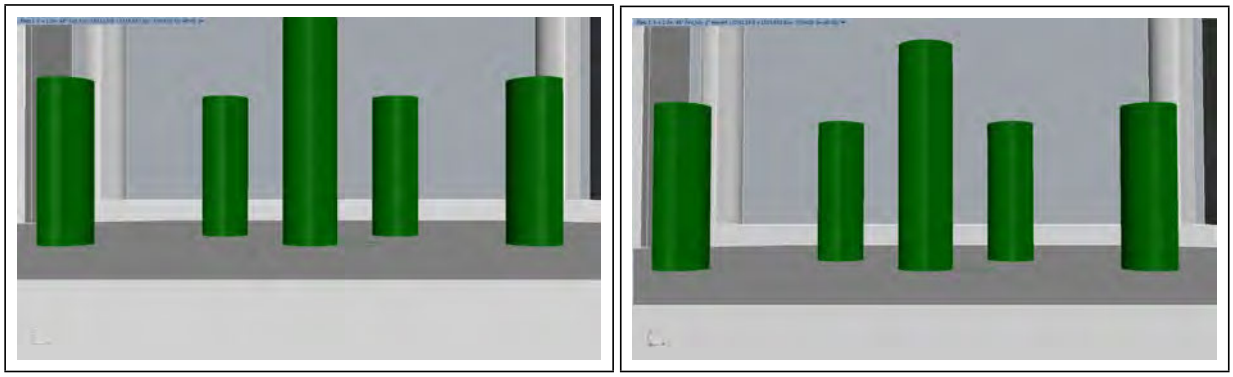


Abbildung 4.1.: Simulierte Kameraabbildung für die Konzertsaalszene. links: Kamera ohne Elevation, rechts: Kamera um 2° eleviert

können, wurde im Szenenzentrum jeweils eine Weißkarte positioniert und mit der stereoskopischen Kamera aufgenommen. Mit Hilfe einer digitalen Spiegelreflexkamera (DSLR - Canon EOS 450D) wurde, indirekt über die Belichtungszeit, die Beleuchtungsstärke der Weißkarte gemessen und die Szene ebenfalls fotografiert.

Zur Bildüberwachung wurde ein professioneller, stereoskopischer Monitor jeweils vor Ort aufgebaut und mit dem Echtzeitsignal der Kamera gespeist. In den medialen Situation mit Fernseher diente dieser ebenfalls als Szenenelement.

Konzertsaal

Die Aufnahmen wurden Anfang April 2012 im Joseph-Joachim-Konzertsaal durchgeführt und teilten sich auf in die Aufnahme der konventionellen Musikvideos unter Leitung der „Hot & Cheesy“ Filmproduktionsfirma und der Aufnahme der stereoskopischen Videos durch Mitarbeiter des Fachgebiets.

Während der Musikvideoaufnahmen wurden die Musiker frei auf der Bühne platziert und in mehreren Durchläufen mit mehreren Kameras gefilmt. Zusätzliche Lichtquellen wurden nach Geschmacksentscheidungen ebenfalls frei verteilt. Aufgrund terminlicher Probleme handelte es sich bei den Musikern nicht um das Reinhold-Quartett, stattdessen konnte das Berliner 3plus1-Quartett für die Aufnahmen gewonnen werden. Das Playbackspiel der beiden Stücke wurde vom Quartett über mehrere Vorproben eingeübt. Die Musiker spielten zu vorbereiteten, mit einem Clicktrack unterlegten nachhallfreien Aufnahmen der beiden Stücke. Zugespielt wurden diese über mehrere, außerhalb der Bildbereiche aufgestellte PA-Lautsprecher.

Für die stereoskopischen Aufnahmen wurden die Musiker auf den Lautsprecherpositionen der akustischen BRIR-Aufnahmen platziert und die stereoskopische Kamera jeweils auf Position und Höhe des FABIAN HATS installiert (siehe Abbildung 4.2). Um spätere Anschlussfehler bei direktem Umschalten zwischen den nacheinander aufgenommenen Sitzplatzvarianten zu minimieren, wurden die Musiker angehalten in ihrer Körpersprache über die verschiedenen Aufnahmen hinweg möglichst konstant zu agieren. Die Kamera wurde jeweils mit Hilfe von Wasserwaagen austariert. Die Schärfe wurde für das populäre Stück auf die Sängerin fokussiert, für das klassische Stück ohne Gesang auf die erste Violine. Vor Beginn der eigentlichen Aufnahme wurden (bei Bedarf je Sitzplatz)

- mit der DSLR Belichtungszeitmessungen durchgeführt und Referenzfotos der Weißkarte an der Position der Sängerin angefertigt,
- ein Weißkartenvideo mit der stereoskopischen Kamera angefertigt,
- ein interner Weißabgleich der stereoskopischen Kamera durchgeführt, die Weißkarte wiederholt aufgenommen.
- ein Video mit horizontaler Ausrichtung (0°-Elevation) mit der stereoskopischen Kamera angefertigt.

Wie bereits erläutert, wurde die Kamera während der eigentlichen Aufnahme um 2° eleviert. Die Synchronität von Ton und Bild während der eigentlichen Aufnahme wurde von einer Tonmeisterin überwacht. Nach jeweils mehreren Takes je Musikstück und zumindest einer zufriedenstellenden Einstellung, wurde der Kameraaufbau an den hinteren Sitzplatz verlegt. Da die Lichtszene konstant blieb, wurde auf eine wiederholte Belichtungsmessung mit der DSLR-Kamera verzichtet, alle weiteren Punkte wurden wiederholt. Die Kamera war ebenfalls um 2° geneigt und auf die Sängerin ausgerichtet. Auch hier wurde die Aufnahme bis zum Erreichen eines vom Ausdruck her befriedigenden und synchronen Ton/Bild-Take wiederholt.

Tonstudio und WFS-Labor

Die prinzipielle Verfahrensweise entsprach auch hier der 'Live'-Situation, das heißt zu Beginn wurde jeweils eine Photometrie durchgeführt, Weißkartenvideos wurden angefertigt und die Kamera jeweils an der Position des HATS aufgestellt. Der Abstand

Abbildung 4.2.: Kombination von Fotografien des Lautsprecheraufbaus und der Musiker, jeweils vom Sitzplatz 1 aus

des Fernsehers vom Betrachter aus wurde durch die Holophonieszene bestimmt und dort rein durch die praktisch realisierbare Entfernung zu dem Lautsprecherarray begrenzt. Da bei Szenenwechseln innerhalb der Simulation das Fernsehbild in Betrachtungswinkel und Position konstant sein soll, wurde diese Entfernung auch in der Studioszene genutzt. Aus Gründen der Konsistenz des Betrachtungswinkels war auch hier die Kamera um 2° gegenüber der neutralen Kopfhaltung eleviert. Der Fixations- und Schärfepunkt war jeweils der Fernseher beziehungsweise der frontale Lautsprecher.

Tabelle 4.1 zeigt alle optischen Situationen in ihrem Ausgangszustand.

4.1.2. Durchführung der Bildaufnahmen

4.2. Bildbearbeitung

Das hier dargestellte Postprocessing beinhaltet nur die finale Generierung der stereoskopischen Stimuli. Die Produktion der konventionellen Videos wurde durch die Filmproduktionsfirma und zwei Mitarbeiter des Fachgebiets durchgeführt. Die Auswahl der finalen stereoskopischen Takes war durch den aufnahmeleitenden Tonmeister nach Gesichtspunkten der Bild/Ton-Synchronität ebenfalls schon durchgeführt und wurde nur noch verifiziert. Das erhobene Ausgangsmaterial ist in Tabelle 4.2 zusammengefasst.

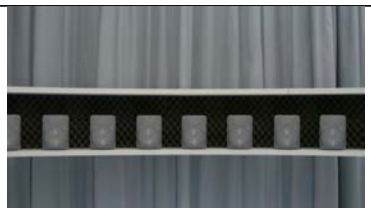
Konzertsaal		
Studio		
WFS-Labor		

Tabelle 4.1.: Beispielhafte Standbilder der erzeugten Bildstimuli (im unbearbeiteten Ausgangszustand), Konzertsaal nur für klassisches Musikstück abgebildet

Alle Bearbeitungen zur Erzeugung der finalen 3D-Bildstimuli wurden vom Autor mit dem Videoschnittprogramm Adobe Premiere CS5.5 durchgeführt.

Schnitt und Vergenz

Um Trapezverzerrungen zu vermeiden, wurden die stereoskopischen Aufnahmen mit parallel ausgerichteten Optiken durchgeführt. Die Konvergenzebene kann hierbei im Nachhinein durch horizontales Verschieben der Bilder festgelegt werden und soll mit dem Fixationspunkt übereinstimmen. Der Nachteil dieses Verfahrens ist, dass nach der Verschiebung jeweils Randbereiche der Bilder nur monoskopisch vorliegen und nicht verwendet werden können und der horizontale *fov* damit weiter eingeschränkt wird. Die Verschiebung muss umso größer ausfallen, je näher die Konvergenzebene am Betrachter liegt. Im vorliegenden Fall betrifft das die Bildsituationen mit Fernseher. Aus Konsistenzgründen wurden alle anderen Videos ebenfalls um den gleichen Betrag horizontal beschnitten, auch wenn damit stereoskopische Informationen verloren gingen. Abildung 4.3 zeigt dies beispielsweise für die Studiosituation. Zum Festlegen des Bildversatzes wurden in Premiere Videospuren jeweils für das linke und rechte Auge angelegt und eine

	Auflösung	**fps**	**Codec**
Konzertsaal Platz 1 klassisch*	1920x1080	30	MJPEG
Konzertsaal Platz 2 klassisch*	1920x1080	30	MJPEG
Konzertsaal Platz 1 populär*	1920x1080	30	MJPEG
Konzertsaal Platz 2 populär*	1920x1080	30	MJPEG
Studio mit TV*	1920x1080	30	MJPEG
Studio ohne TV*	1920x1080	30	MJPEG
Holophonie mit TV*	1920x1080	30	MJPEG
Holophonie ohne TV*	1920x1080	30	MJPEG
Musikvideo klassisch	1920x1080	25	ProRes V0
Musikvideo populär	1920x1080	25	ProRes V0

Tabelle 4.2.: Übersicht der Ausgangsvideodaten, die mit * gekennzeichneten Videos liegen stereoskopisch, also jeweils ein Mal für das linke und rechte Auge vor

Spur auf 50% Transparenz gesetzt. Die Bilder wurden danach so lange verschoben, bis der gewünschte Fixationspunkt größtmögliche Kongruenz aufwies. Als Fixationspunkte dienten bei den Livesituationen die Sängerin, beziehungsweise die erste Geige und für die medialen Situationen der Fernseher, beziehungsweise der frontal gelegene Lautsprecher. Hierbei wurden auch noch mehrere, kleinere, aufnahmebedingte Artefakte ersichtlich, die sich als Kanalunterschiede potentiell negativ auf die stereoskopische Information hätten auswirken können:

1. Ein minimaler Höhenversatz zwischen den Augen. Dieser wurde durch das Verschieben des linken Auges um einen Pixel nach oben kompensiert.

2. Ein minimaler Schärfeunterschied zwischen den beiden Augen der nicht kompensiert werden konnte.

3. Minimale Farbunterschiede zwischen den beiden Augen, die ebenfalls unkompensiert belassen wurden.

Weitere Artefakte, die jedoch identisch auf beiden Kanälen auftraten, waren:

4. Eine tonnenförmige Verzeichnung der Bilder (auch Cushion oder Barrel-Effekt). Die in Abbildung 4.3 im Hintergrund zu sehende schwarze Traverse ist leicht nach

unten gebogen, obwohl sie gerade sein sollte. Diese Verzerrung ist ein typischer Fehler der optischen Abbildung und fällt vor allem bei größeren Bildwinkeln und zunehmend zu den Rändern hin auf. Premiere hält für diesen Fall ein Effektplugin 'Linsenverzerrung' bereit, mit dem diese Krümmung kompensiert werden kann. In allen Situationen wurde jeweils eine eigentlich gerade geometrische Form (Traversen, Bühnenkanten, Rand des Lautsprecherrings) gesucht und mit Hilfe dieses Effekts begradigt. Damit einhergehend kam es allerdings zu einem weiteren, algorithmusbedingten *fov*-Verlust.

5. Vor allem auf den großen weißen Wandflächen der hinteren Konzertsaalsituationen war deutliches Lichtflimmern [1] zu sehen. Da es sich um statische Bereiche des Bildes handelte, wurde ein Standbild der betroffenen Flächen über das Bildmaterial gelegt. Das Manipulationsergebnis war befriedigend und dabei unauffällig.

2D-3D-Montage der konventionellen Musikvideos

In den bereits beschnittenen und verschobenen medialen Situationen mit Fernseher wurden zunächst die noch sichtbaren Reflexionen des Messaufbaus im Fernseher geschwärzt. Daraufhin wurden das monoskopische Bildmaterial der konventionellen Videos in die stereoskopisch dargestellte Ebene des Fernsehers eingepasst. Dazu waren Positionsverschiebungen und leichte Bildtransformationen nötig. Dies musste separat für das linke und rechte Auge vollzogen werden, da auch die Fernseher in den stereoskopischen Aufnahmen nicht absolut kongruent übereinanderlagen.

Nach allen Bearbeitungsschritten ergaben sich 1920x1080px große Videos (inklusive schwarzen Rändern) mit einer Bildrate von 30fps für alle Situationen, jeweils für das linke und rechte Auge. Das endgültige Video musste nun noch bildschirmfüllend an das Projektionssystem ausgegeben werden. Somit wurden die FullHD Videos nahezu mittig (s.u.) in ein 5400x1920px großes Schwarzbild eingebettet.

4.3. Optische Simulationsumgebung

Die optische Wiedergabeumgebung besteht aus einer halbzylinderförmigen Projektionsfläche mit 5 m Durchmesser und 2,5 m Höhe (Abbildung 4.5). Fünf um 90° gedrehte

[1] vermutlich verursacht durch Interaktion von Zündfrequenz der HQI-Lampen vor Ort und der Shutterfrequenz der Kamera

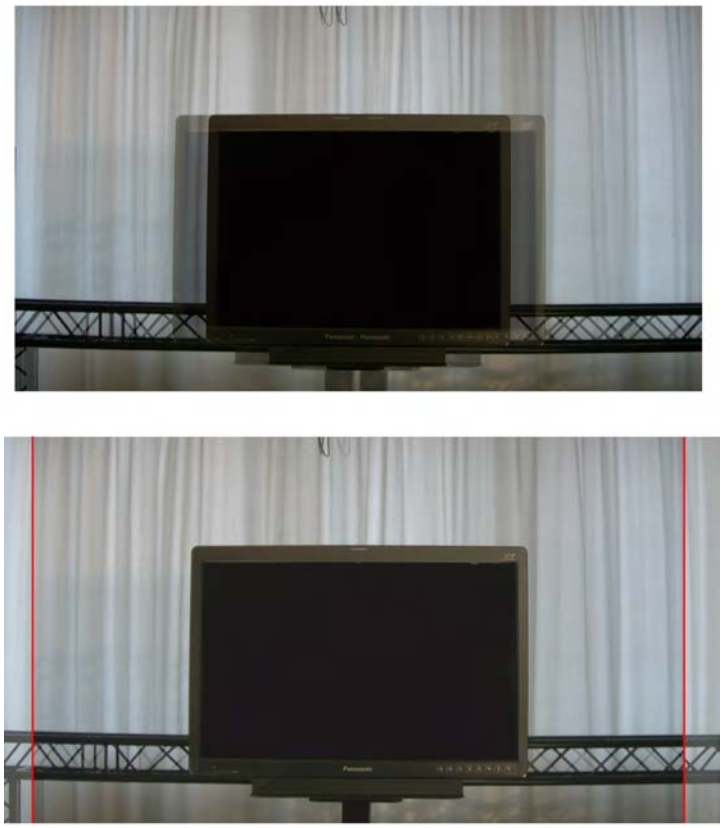
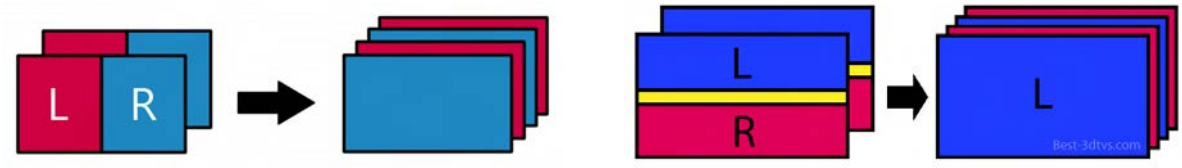


Abbildung 4.3.: Tonstudio mit Fernseher. Oben das linke und rechte Bild übereinander, halb transparent, unten das linke und rechte Bild jeweils horizontal verschoben mit Fernseher als Fixationspunkt. Außerhalb der eingezeichneten roten, vertikalen Linien liegt nur ein Bild vor und muss beschnitten werden. Ebenfalls ersichtlich sind Tonnenverzerrung anhand der eigentlich geraden Traverse im Hintergrund.

Epson EH-TW9000 Beamer, mit jeweils einer Auflösung von 1920 x 1080 Pixeln erzeugen eine theoretische Gesamtauflösung von 5400 x 1920 Pixeln. Durch benötigten Overlap zwischen den einzelnen Beamerprojektionen, verringert sich allerdings die effektiv genutzte Auflösung auf 5144 x 960px und man erhält einen nutzbaren *field of view* von 160° x 56°. Die Beamer werden von einem Windows-PC über eine ATI Radeon HD 7870 Eyefinity 6 betrieben. Die 90°-Drehung der einzelnen Beamer wird treiberseitig kompensiert und die sechs separaten Bildschirme werden im System zu einem virtuellen 5400 x 1920 Pixel großen Bildschirm zusammengefasst. Das Bild wurde mit Hilfe der Warpingsoftware 'Pixelwarp Evo Wide Screen Warper' für eine

korrekte Zylinderprojektion unter Benutzung von Testbildern und Linienlaser geometrisch entzerrt, Farbabweichungen zwischen den Projektoren wurden nach perzeptivem Kriterium minimiert.

Das verwendete Stereoskopieverfahren ist *side by side* (beamerseitig, da diese um 90° gedreht sind) beziehungsweise *top and bottom* (rechnerseitig, da die Beamerdrehung kompensiert wurde). Bei diesen Verfahren werden die Bilder für das linke und rechte Auge nicht sequentiell, sondern gleichzeitig, über- oder nebeneinanderliegend, in einem Bild übertragen. Erst im Beamer wird dieses Bild aufgeteilt und sequentiell dargeboten. Die Synchronisierung zwischen Bild und Shutterbrillen wird somit auch von den Beamern vorgenommen. Damit alle Beamer zueinander synchron sind, müssen die Einzelbilder und somit der Takt der einzelnen Grafikkartenausgänge ebenfalls synchronisiert sein. Dies wird ausschließlich durch den ATI-eigenen Eyefinitymodus und der Bildung eines großen virtuellen Desktops ermöglicht.



(a) Side-by-Side-Übertragungsverfahren mit vertikalem Auflösungsverlust

(b) Top-and-Bottom-Übertragungsverfahren mit horizontalem Auflösungsverlust

Abbildung 4.4.: 3D Übertragungsverfahren[2]

Die höchstmögliche stereoskopische Auflösung wird im *frame packing 3D* Modus mit, ausgehend von einem Beamer, 1920 x 1080 Pixeln pro Auge erreicht. Die Beamer werden hierfür, dem *top and bottom* Verfahren ähnelnd, mit einem 1920 x 2205 Frame angesprochen, bestehend aus den übereinanderliegenden Bildern für das linke und rechte Auge, mit 45 Pixel *blanking space* dazwischen. Auch wenn die Beamer prinzipiell diesen Modus und somit die genannte Auflösung von 1920 x 1080 Pixeln pro Auge unterstützen, ist es momentan, vermutlich treiberbedingt, nicht möglich, allen Beamern über den Eyefinitymodus dieses Format zur Verfügung zu stellen. Betrachtet man die gesamte Projektionsfläche, werden die Bilder für das linke und rechte Auge gleichzeitig in einem 5400 x 1920px großen Bild übertragen, jedem Bild steht daher entweder nur die halbe horizontale (*side-by-side*-Modus) oder vertikale (*top-and-bottom*-Modus)

[2]Abbildung von `http://www.best-3dtvs.com` abgerufen Oktober 2013

Auflösung zur Verfügung. Der dann sequentiell angebotene Bilderstream wird gestreckt und interpoliert. In der Cave wird daher jedem Auge nur die halbe theoretisch mögliche Pixeldichte angeboten und die Auflösung beträgt somit pro Auge 5400 x 960 Pixel am Beamereingang.

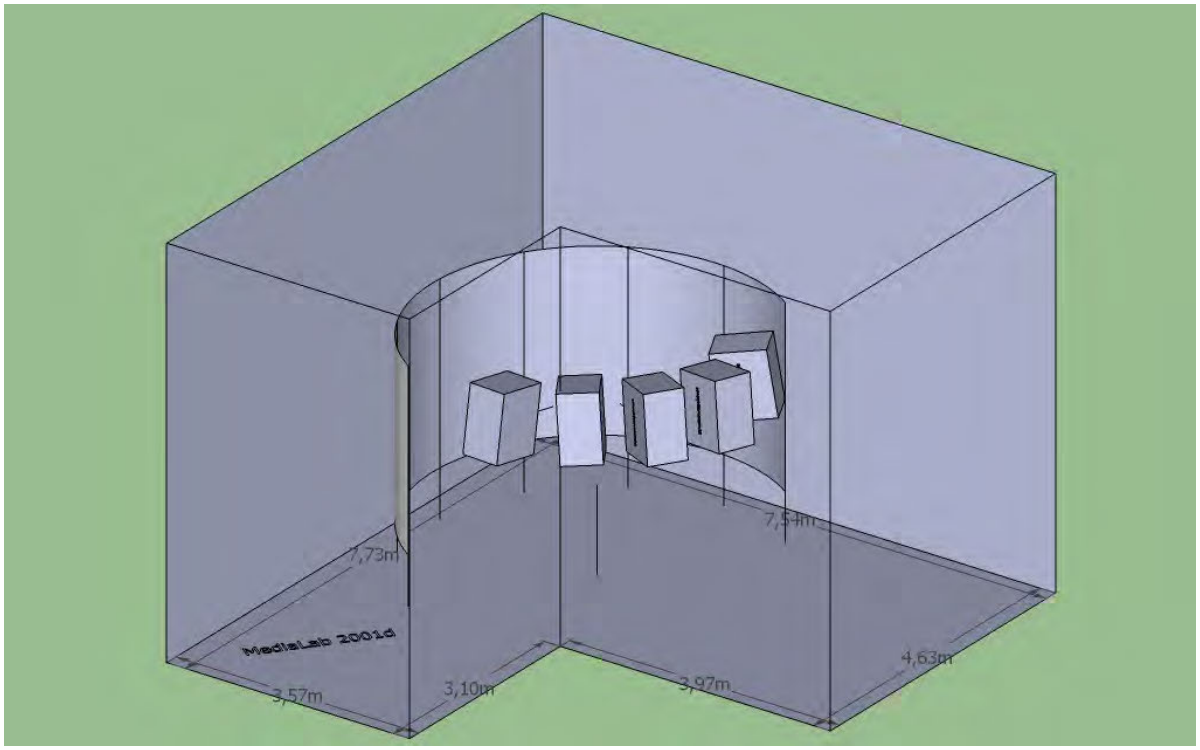


Abbildung 4.5.: Cave-Modell (nicht maßstabsgetreu)

Die Anwendung *Stereoscopic Player* in der Version 1.8.1[3] ist explizit für die Wiedergabe von 3D-Formaten entwickelt worden. Unabhängig vom Eingangsformat der dem linken und rechten Auge entsprechenden Videos, können diese in unterschiedliche Wiedergabeformate umgewandelt und wiedergegeben werden. Weiterhin kann der Player über VB-Skripte automatisiert werden, wodurch erst eine Synchronisierung zum Audiostream ermöglicht wurde.

Die Wahl des Videocodecs wurde sehr stark durch zwei Anforderungen eingeschränkt. So begrenzte zum einen die unüblich hohe Auflösung von 5400 x 1920 Pixeln je Auge die Auswahl, zum anderen der Anspruch instantan zwischen den Szenen wechseln zu können. Ein schon früher genutztes Verfahren, in dem alle Videosituationen gleichzei-

[3] das Bildmaterial konnte nicht mit neueren Versionen wiedergegeben werden

tig abgespielt wurden und nur der Fokus des Betriebssystems auf die je gewünschte Playerausgabe geändert wird, ließ sich technisch nicht realisieren. Daher wurden alle Situationen nacheinander in einem längeren Video vereint und der Wechsel von Situationen durch Sprünge im Video realisiert. Alle Codecs mit inter-frame-Kodierung basieren auf Bewegungsprädiktionen und nur ein, meist definierbarer, Bruchteil der Videobilder wird voll übertragen. Setzt die Wiedergabe nun zwischen den voll übertragenen Bildern ein, kommt es zu sehr starken Bildartefakten, die je nach Codec erst nach bis zu mehreren Sekunden verschwinden können. Der einzige Codec mit annehmbarer Datenreduktion ohne inter-frame-Kodierung war *Motion JPEG*. Da ein Großteil unseres Bildes aus schwarzer Fläche besteht, ist der im Vergleich zu anderen Codecs geringe Kodierungsgewinn ausreichend groß, um eine flüssige Wiedergabe zu ermöglichen[4]. Auch wenn der Stereoscopic Player eigentlich über einen internen *MJPEG* Decoder verfügt, wurden sowohl Umschaltzeiten als auch Videoruckeln durch die Nutzung des externen Decoders von *ffdshow*[5] minimiert.

Die finale Kodierung der aus Premiere verlustlos exportierten Videodaten wurde mit ffmpeg[6] durchgeführt. Da aufgrund eines Fehlers in ffmpeg komplett schwarze Frames nicht kodiert werden konnten, wurde in diese ein kleiner dunkelroter, sich in einer Bildecke befindlicher Kreis gesetzt.

Als Eingangsdaten für den Player fungierten somit zwei Videodateien (je eine für das linke/rechte Auge), kodiert mit *MJPEG* in einer Auflösung von 5400 x 1920 Pixeln und 30 Bildern pro Sekunde.

4.3.1. Geometriekalibrierung

Die optische Kalibrierung soll eine möglichst realitätsnahe Reproduktion der realen Geometrie- und Farbverhältnisse sichern. Geometrische Verzerrungen durch Abbildungsfehler der Kamera und der Zylinderprojektion wurden bereits in den beiden vorhergehenden Unterkapiteln erläutert. Nun verbleibt die korrekte Skalierung des Bildes auf das zu erzielende Sichtfeld und die korrekte Höhenpositionierung im Projektionssystem.

Der gemessene *fov* der stereoskopischen Kamera betrug 48° x 27° und reduzierte

[4] Als überlegene Alternative mit sehr hohem Kodierungsgewinn hat sich mittlerweile der *libx264*-Codec erwiesen. Die damit kodierten Videos in unüblich großen Auflösungen lassen sich allerdings nicht mit dem Stereoscopic Player wiedergeben

[5] `http://ffdshow-tryout.sourceforge.net/`

[6] `http://www.ffmpeg.org/`

sich nach der Vergenzanpassung auf 45,54° x 27°. Die nachfolgende Kompensation der Tonnenverzerrung ging wiederum mit einem Sichtfeldverlust von 3% einher, resultierend in einem finalen *fov* von 44,17° x 26,12°.

Die Versuchspersonen werden im Zentrum der Cave positioniert und sowohl Bildposition als auch Skalierung müssen so lange angepasst werden, bis der *fov* der Versuchsperson mit dem *fov* des Bildmaterials übereinstimmt (siehe Abb 4.6).

Ausgehend von $r = 2,50$ m Entfernung von der Leinwand, können über trigonometrische Funktionen die für die zu erreichenden Betrachtungswinkel notwendigen Bildhöhen und Breiten bestimmt und mit Hilfe von Entfernungsmessern in der Cave eingestellt werden. Genau genommen wurde nur die korrekte Bildbreite eingestellt, während sich die Höhe aus dem festen Seitenverhältnis von selbst ergab und nur zur Verfizierung nachgemessen wurde.

$$\begin{aligned} Bild_{ver} &= r \cdot tan(fov_{ver}) = 1,16m \\ Bild_{hor} &= r \cdot sin(fov_{hor}) = 1,88m \end{aligned} \tag{4.1}$$

Die Höhe der Kamera betrug, übereinstimmend mit der auditiven Simulation, in jeder Situation 1,20 m über dem Boden. Die letztendliche Sitzhöhe der Versuchsperson in der Cave war demgegenüber frei wählbar, so lange die Projektionshöhe mit angepasst wurde. Die Größen- und somit Positionsunterschiede der Versuchspersonen sollten mit einem höhenregulierbaren Sessel ausgeglichen werden. Mit diesem hätten, nach den Ergebnissen von Vorversuchen, kleinere Personen aber vermutlich nicht auf 1,20 m erhöht werden können, daher wurde 1,15 m als Referenzhöhe gewählt. Diese Anpassung hat keinerlei Auswirkung auf die Kongruenz der auditiven und visuellen Stimuli.

Für die Höhenkorrektur wurde zunächst ein bei der Akquise mit aufgenommenes Bild mit 0° Kameraneigung projiziert[7] und die horizontale Achse eines eingefügten Mittenkreuzes auf die Referenzhöhe 1,15 m eingestellt. Dieses Bild wurde wiederum auf halbtransparent gesetzt, und ein darunterliegendes Bild mit 2° Kameraneigung wurde so lange verschoben, bis Kongruenz an einem zentral gewählten Fixpunkt erreicht wurde. Diese Prozedur wurde für alle Situationen wiederholt, wobei ein leichter Winkelfehler der Aufnahme im WFS-Labor zutage trat. Die Kamera wurde dort anscheinend nur um 1,5° geneigt, was vor allem zu Musikvideobildsprüngen im Vergleich zur Studiosituation

[7] alle Postpocessingschritte wurde ebenfalls auf dieses Bild angewandt

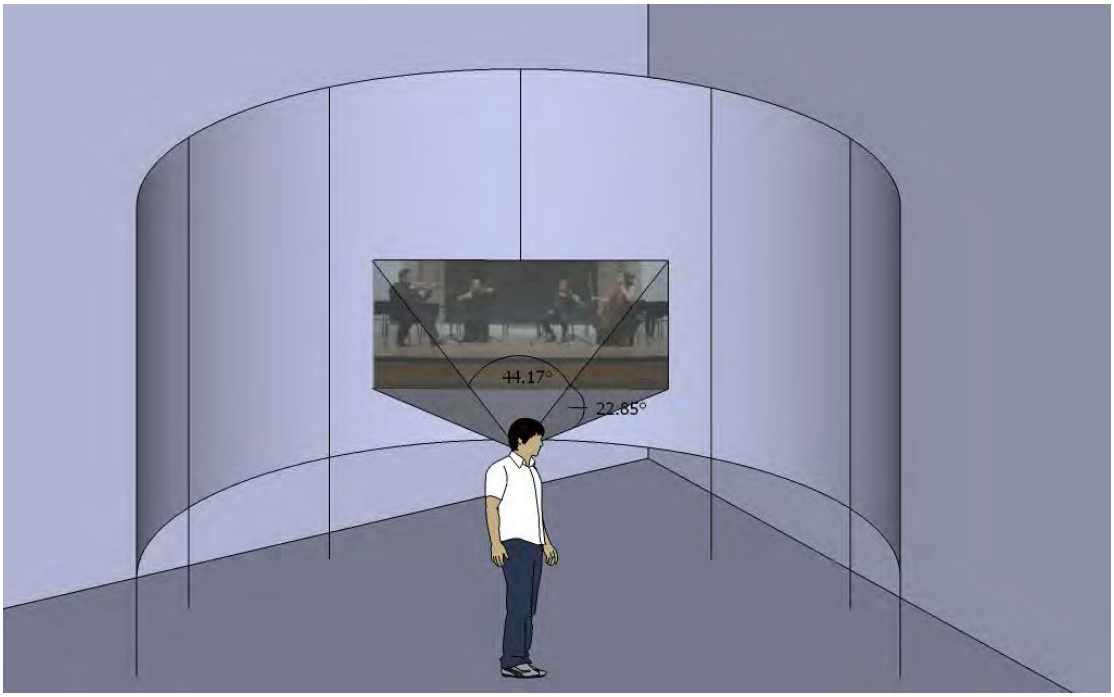


Abbildung 4.6.: Nicht maßstabsgetreues Modell der Cave mit Versuchsperson und eingezeichneten Betrachtungswinkeln

führen würde. Glücklicherweise konnte ein einsekündiges Testvideo mit korrekter Neigung gefunden und in Premiere durch kopieren auf die Stimuluslänge verlängert werden. Noch vorhandene Klebebandmarkierungen am Fernseher wurden retuschiert und alle Postprocessingschritte wiederum angewandt. Die Situation ohne Fernseher wurde belassen, da sich hier aufgrund der im Hörversuch getrennt untersuchten AV-Bedingungen (s.u.) kein direkter inhaltlicher Bezug zu den anderen Situationen ergab.

4.3.2. Farb- Helligkeits- und Dynamikkalibrierung

Kameras und Wiedergabesysteme benutzen an die Wahrnehmung des Menschen angepasste Helligkeitsübertragungsfunktionen mit sogenannten Gammakorrekturen. Werden in der Wiedergabekette durchgängig dieselben Korrekturwerte benutzt, sollte eine nachträgliche Kalibrierung theoretisch nicht nötig sein. Durch Produktstreuung, Produktqualität oder beispielsweise Reflexionseigenschaften von Projektionswänden, können die projizierten Bilder von den realen abweichen. In dem stereoskopischen Fall muss außerdem der Shutterbrilleneinfluss mit betrachtet werden.

Farbkalibrierung

Für jede Raumsituation wurde mit der stereoskopischen Kamera ein Weißkartenvideo aufgenommen und im Nachhinein in der Cave projiziert. Vor Ort wurde ebenfalls von derselben Szene mit einer digitalen Spiegelreflexkamera ein Foto im RAW-Format aufgezeichnet. Die Idee des Vorgehens war nun, dass, wenn die Projektion nun ebenfalls mit derselben DSLR-Kamera mit identischen Einstellungen und durch die stereoskopische Brille fotografiert wird, die Differenzen zwischen Projektion und Realität ausgelesen werden können, ohne auf absolute Messwerte angewiesen zu sein.

Zunächst wurde empirisch verifiziert, dass die Projektionshelligkeit auf die Farbtemperatur der Weißkarte keine Auswirkung hat und somit auch noch nach der Farbkalibrierung frei gewählt werden kann. Eine Aufnahme durch die Shutterbrille verändert die Farbtemperatur um ca. 100-150°K und liegt somit innerhalb des, ebenfalls empirisch ermittelten, Messgenauigkeitsfehlers der DSLR-Kamera. Dabei wurden für das verwendete Shutterbrillenmodell systematische Differenzen zwischen dem linken und rechten Brillenglas festgestellt und daher alle folgenden Messungen nur durch das linke Brillenglas durchgeführt und der dadurch induzierte Fehler als gegeben hingenommen.

Die Weißkartenszene der stereoskopischen Kamera wurde mit der Spiegelreflexkamera durch das Brillenglas aufgenommen. Mit dem 'White Balance Tool' des 'Camera RAW' Plugins für Adobe Photoshop wird die Farbtemperatur auf der Weißkarte ausgelesen. Die Farbtemperaturen der vor-Ort- und der Caveaufnahme werden dann in RGB-Werte umgerechnet[8] und zueinander ins Verhältnis gesetzt. Die so erhaltenen Korrekturfaktoren für die dei R, G, und B-Kanäle konnten mit dem 'RGB-Farbkorrektur' Plugin in Premiere auf das Video angewendet und mit einer erneuten Messung verifiziert werden.

Während der Kalibrierung der vorderen Konzertsaalszene war vor Erreichen des Zielwerts der Rotkanal schon voll ausgesteuert und die Szene unplausibel gelbstichig.

Für das Verständnis dieser Befunde sind folgende Betrachtung über die Farbwahrnehmung hilfreich. Das von nicht selbst leuchtenden Körpern emittierte Farbspektrum hängt ab von den Materialeigenschaften des Körpers und der spektralen Zusammensetzung der ihn anstrahlenden Lichtquelle. Trotz teilweise großer Farbtemperaturunterschiede zwischen möglichen Lichtquellen, wie beispielsweise Sonnenlicht und durch Leuchtstoffröhren emittiertes Licht, verändert sich die subjektive Farbwahrnehmung derselben

[8] dazu wurde der 'CIE Color Calculator' auf der Seite `http://www.brucelindbloom.com/` verwendet, mit den Einstellungen Ref.White=D65, RGB-Model=AdobeRGB(1989), Gamma=1.0 und Adaption=none

Objekte jedoch nur geringfügig, ein weißes Blatt bleibt immer annähernd weiß. Diese sogenannte chromatische Adaption ist ähnlich dem technischen Weißabgleich, so dass aktuelle lichtquellenbedingte Farbtemperaturunterschiede adaptiert werden. Bei Präsentation der optischen Stimuli werden in einem relativ geringen Sichtfeld der Versuchsperson in kurzen Abständen unterschiedliche Räume und damit Lichtquellensituationen präsentiert, die das visuelle System mutmaßlich nicht adaptieren kann. Auch wenn die Szenen korrekt projiziert werden, erscheinen sie subjektiv farbstichig und unplausibel. Daher wurde entschieden, nicht mehr die absoluten Farbtemperaturen herzustellen, sondern nur die relativen Unterschiede zwischen den Räumen beizubehalten.

Da die Korrekturmöglichkeit in dem Bereich des Rotanteils schon begrenzt erschien, wurde die röteste Szene (Konzertsaal) als Referenz gewählt und alle anderen Szenen in ihren Farbtemperaturdifferenzen an diese angepasst. Die Konzertsaalszene wurde zunächst perzeptiv so angepasst, dass vor allem die Hauttöne der Musiker plausibel erschienen. Der letztendlich so erhaltene Korrekturfaktor von -1000°K wurde in alle Bildsituationen zuzüglich der Szenendifferenzen von ca. 500°K eingerechnet. Die Anpassungen wurden mit dem eingangs erwähnten Verfahren durchgeführt und jeweils verifiziert. In einer finalen Geschmackskorrektur mussten die beiden Konzertsaalplätze jeweils noch leicht angepasst werden. Die einzelnen Werte sind in Tabelle 4.3 aufgeführt.

Die in den Szenen befindlichen Musikvideos wurden nicht angepasst, da der Einfluss des Fremdlichts auf eine aktiv leuchtende Quelle geringer ist als auf eine passive.

Raum/Position	**Konzertsaal Platz 1**	**Konzertsaal Platz 2**	**Studio**	**WFS-Labor**
	Farbtemp/°K			
Referenzfoto vor Ort	2600	-	3050	2950
In Cave gemessen	6400	-	4900	6100
Sollwerte in Cave	6400	-	6850	6750
Istwerte nach Korrektur	6400	-	6900	6800
nach Geschmackskorrektur	5300	5650	5850	5650

Tabelle 4.3.: Farbkorrektur der einzelnen Szenen.

Helligkeitskalibrierung

Für die Wiederherstellung der identischen Helligkeiten in Simulation und Aufnahme wurde bei der Fotografie der im Szenenmittelpunkt stehenden Weißkarten die Verschlusszeit der DSLR-Kamera mitgeführt. Das Ziel wäre die Wiederherstellung der in Tabelle 4.4 aufgeführten Verschlusszeiten, während die Weißkartenaufnahme der stereoskopischen Kamera mit allen bereits vorgestellten Bearbeitungsschritten in der Cave projiziert wird. So würde man dieselbe Belichtungsstärke (d.h. Helligkeit) erreichen, ohne einen absoluten Wert zu benötigen.

Auch wenn die benutzten Beamer vom Datenblatt her in der Lage wären die realen Helligkeitsverhältnisse wiederzugeben, beträgt die brillenverursachte Helligkeitsdämpfung nach eigenen Messungen ca. (typische) 80%. Daher wurde die Zielstellung auf die Wiederherstellung der Helligkeitsunterschiede zwischen den Szenen reduziert.

Hierzu wurde zunächst die hellste Szene (Studio) in Premiere bis kurz unter die Sättigungsgrenze um 30% verstärkt und projiziert. Die Helligkeitsdifferenzen zu den anderen Szenen wurden implementiert und jeweils durch Messungen verifiziert. Diese wurden nicht mehr durch eine Shutterbrille vorgenommen, da diese Messwertschwankungen verursachte und absolute Werte nicht mehr von Interesse waren.

Die perzeptive Beurteilung zeigte allerdings, dass die Helligkeitsunterschiede zwischen den Bildszenen jetzt unplausibel stark erschienen und vermutlich einen sehr starken Einfluss auf das Antwortverhalten der Versuchsperson gehabt hätten. Die Unterschiede zwischen den Tageslicht- (Studio) und Kunstlichtszenen (Konzertsaal) könnten zwar prinzipiell korrekt gewesen sein, erschienen aber vermutlich durch mangelnde Adaption unplausibel. Da die Szenen inhaltsbedingt den gewünschten Abstufungen von hell nach dunkel (Studio, Holophonielabor, Konzertsaal) entsprachen, wurden letztendlich alle Szenen um 25% verstärkt.

Ort	vor Ort bestimmte Verschlusszeiten	in Cave bestimmte Verschlusszeiten	in Cave Zielverschluss-zeiten	Skalienrungsfaktoren für die Erreichung des Sollwerts (erreichte Verschlusszeiten)	benutzte Skalierung
Studio	1/125	1/30	1/30	1,3 (1/40)	1,25
Holo	1/50	1/30	1/12	0,65 (1/10)	1,25
Konz.	1/15	1/30	1/3,6	0,4 (1/5)	1,25

Tabelle 4.4.: Verschlusszeiten der DSLR-Kamera für Weißkartenfotos vor Ort und in der Cave sowie in Premiere benutzte Skalierungsfaktoren. Alle Aufnahmen mit ISO 200 und 5,6 Blende angefertigt.

5. AV-Umgebung

In diesem Kapitel wird das Gesamtsystem aus bereits eingeführten Audio- und Videosimulationskomponenten vorgestellt, einschließlich der benötigten AV-Synchronisierung zwischen zwei Audio- und einem Videorechner. Ausgangspunkt sind eine Audiosimulation mit je Musikstück 14 verschiedenen auditiven Rezeptionssituationen, verteilt auf zwei Rechner und eine Videosimulation mit sechs verschiedenen visuellen Rezeptionssituationen.

5.1. Cavebeschreibung

Der Versuch fand im sogenannten Medialab des Fachgebiets statt, bestehend aus der Cave (siehe auch S.3.3.3) und einem danebenliegenden Maschinenraum. Abbildung 5.1 zeigt den Signalfluss aller beteiligter Systeme. Hierbei befinden sich die drei blau unterlegten Rechner im Maschinenraum. Die verwendeten Hard- und Softwarekonfigurationen sind in Tabelle A.1 zusammengefasst.

5.2. Synchronisierung

Aufgrund der Menge an situationsspezifischen Rohspuren war es nicht möglich diese zusammen mit dem Video in einem Videocontainerformat zusammenzufassen[1]. Video- und Audioinhalte wurden daher von zwei unabhängigen Systemen wiedergegeben und mussten synchronisiert werden.

Die Anforderungen an das Gesamtsystem waren dabei:

1. Ausgabe einer gewünschten Audio/Video-Stimuluskombination nach Erhalt einer OSC-Nachricht von einem Hörversuchssystem.

[1] in einem Videocontainerformat, beispielsweise avi oder mp4, werden die parallelen Video- und Audiodatenströme nach formatspezifischer Vorschrift in einen seriellen gewandelt, gemuxt.

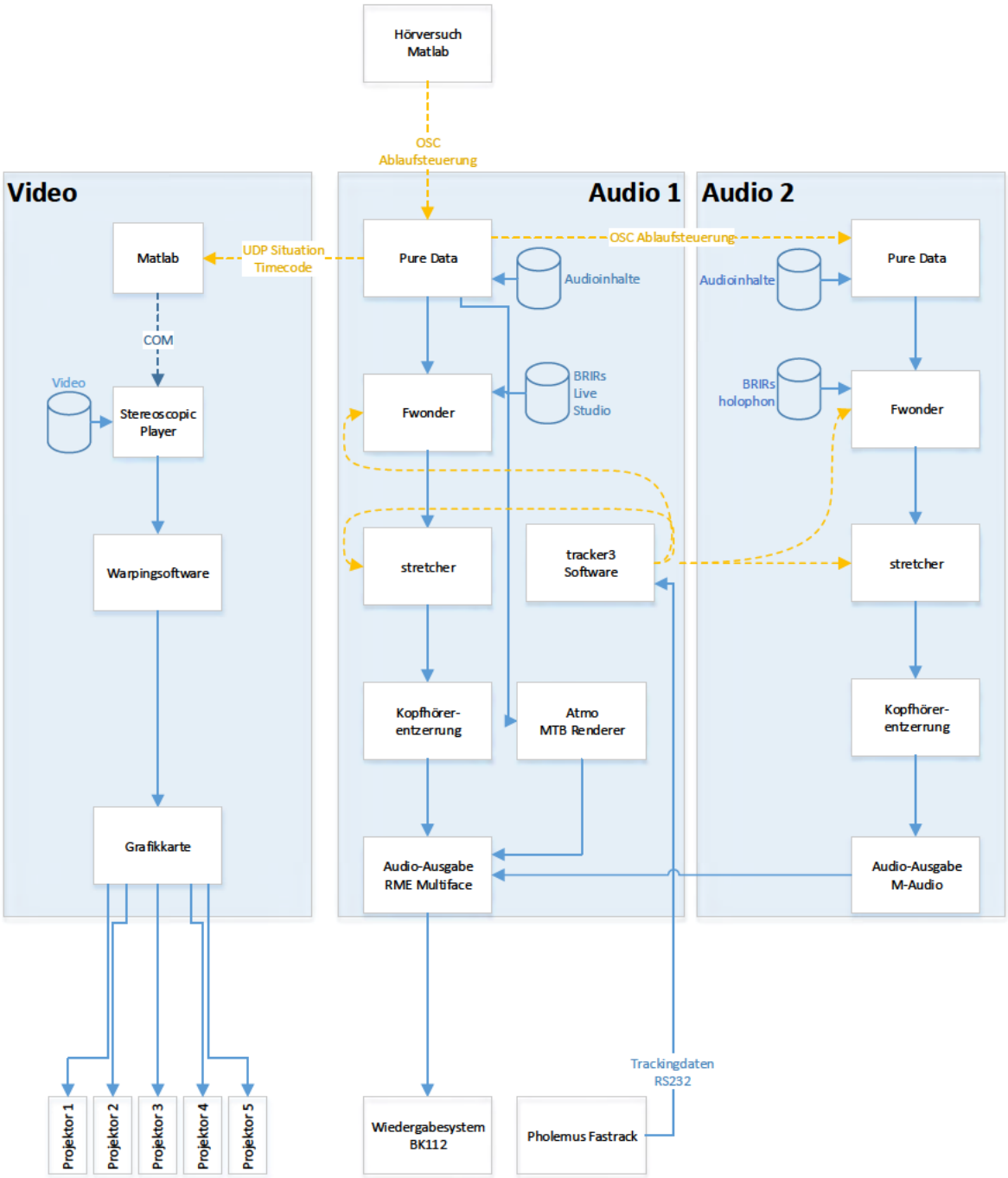


Abbildung 5.1.: Signalfluss zwischen allen beteiligten Audio-, Video- und Steuersystemen

2. Vernetzung aller beteiligten Systeme über geeignete Protokolle und synchrone Audio-Video-Wiedergabe.

3. Einfache Bedienung über ein Interface.

Um zu überprüfen, ob eine ständige Timecodeverbindung zwischen den Audio- und Video-Systemen erforderlich ist und somit neue Softwarelösungen für die Videowiedergabe gefunden werden müssten, wurde untersucht, wie stark die Asynchronität zwischen den Systemen ist.

Hierzu wurde zunächst für den Videorechner ein Matlabskript erstellt, das Daten per UDP vom Audiorechner empfängt und in die vom Stereoscopic Player unterstütze COM-Automation [2] übersetzt. Für den Synchronitätstest wurde ein Video im *MJPEG*-Format mit der später genutzten Gesamtauflösung, einem darin befindlichen 1920 x 1080 großen Bildrauschen, einem mitlaufenden Timecode und dazu synchronisierten Sinusbursts erstellt. Die genutzte Audiospur wurde ebenfalls auf den Audiorechner kopiert und in Pure Data eingebunden.

Die Audioausgänge der Video- und Audiorechner wurden von einem externen System [3] aufgezeichnet, um jeglichen konstanten Versatz und zeitabhängigen Drift im Nachhinein in einer DAW ablesen zu können.

Der Startbefehl wurde, wie auch später im Hörversuch, von einem externen System an Pure Data gegeben, dieses leitete den Befehl an das Videosystem weiter. Nach einer Aufnahme von vier Minuten ergab sich, abgesehen von einem erwarteten konstanten Offset, ein Versatz von 7ms.

Dieser ist nach [44] für Musik unkritisch. Im Hörversuch werden nur Stimuli von ca. einer Minute verwendet und es ist weiterhin zu erwarten, dass innerhalb eines Durchlaufs die Bildsituationen mehrmals gewechselt werden, wodurch immer neu synchronisiert werden muss. Daher wurde auf eine konstante Synchronisierung verzichtet. Eine weitere Messung mit dem Simulator im Endzustand wurde ebenfalls durchgeführt und soll später aufgeführt werden.

Die am Fachgebiet entwickelte Hörversuchssoftware *WhisPER* stellt der Versuchsperson eine grafische Oberfläche zur Verfügung, mit der ein Stimulus nach dem RGT-Verfahren ausgewählt und gestartet/gestoppt werden kann. Während einer laufenden

[2] Schnittstelle zur Kommunikation zwischen separaten Programmen auf einem System, vorrangig für Windows entwickelt

[3] Lenovo X60 Notebook mit Samplitude und Werksaudioeingang

Stimulusdarbietung soll die Stimulusart ebenfalls beliebig gewechselt werden. Die Software schickt über OSC somit Befehle zum starten/stoppen des Stimulus sowie zur auditiven und zur visuellen Situationswahl.

OSC-Pfad	Argument	Beschreibung
/stop	0,1	stop
/content	1	klassisches Stück
	2	modernes Stück
/situation	1-28	auditive Situationen
/bild	1	kein Bild
	2	Bild real und medial ohne Musikvideo
	3	Bild real und medial mit Musikvideo

Tabelle 5.1.: Von Pure Data empfangene OSC-Argumente. Ein Stimulusstart ergibt sich aus dem Empfangen einer 'content'-Nachricht nach vorherigem Stoppen.

Die eigentliche Kommunikation findet zwischen Pure Data auf dem Audiosystem und Matlab auf dem Videosystem statt. In der Videodatei sind die einzelnen visuellen Szenen nacheinander im Abstand von 90 Sekunden angeordnet (Abbildung 5.2), jede Situation n hat somit einen definierten Startzeitpunkt $t_{start}(n) = 1 + 90 \cdot (n - 1)$ in Sekunden. Zudem wurde ein einsekündiger Offset benutzt, um jeweils mit einem schwarzen Bild starten zu können.

Abbildung 5.2.: Abfolge der Szenen in der Videodatei für die ersten fünf Szenen

In der Ausgangssituation ist das Video geladen und pausiert. Erhält das System einen Startbefehl, springt das Video an den stimulusspezifischen Startzeitpunkt $t_{start}(n)$. Es wird ein Zeitzähler t in Pd aktiviert, die Audiowiedergabe und die Videowiedergabe gestartet. Der Zeitzähler und die Audiowiedergabe starten synchron, laufen unabhängig von Stimuluswechseln immer kontinuierlich weiter und stellen somit die zeitliche Referenz dar, auf die das Video synchronisiert wird.

Situations-nummer	Audiospuren	Raum	Atmo
1/15	nachhallfrei	Real Platz 1	
2/16	nachhallfrei	Real Platz 2	
3/17	nachhallfrei	Real Platz 1	ja
4/18	nachhallfrei	Real Platz 2	ja
5/19	Mono aus Kugel	Studio Center	
6/20	Stereo aus AB	Studio Links/Rechts	
7/21	Stereo aus XY	Studio Links/Rechts	
8/22	Stereo aus ORTF	Studio Links/Rechts	
9/23	Surround aus Decca	Studio Surround	
10/24	Surround aus INA3	Studio Surround	
11/25	Stereo aus AB	Kopfhörer	
12/26	Stereo aus XY	Kopfhörer	
13/27	WFS konventionell	WFS-Labor	
14/28	WFS kreativ	WFS-Labor	

Tabelle 5.2.: Stimulustabelle. Die erste Situationsnummer bezieht sich auf das klassische Stück, die zweite auf das moderne. In den Studiosituationen werden die genutzten Lautsprecher angegeben

Wird auf den Stimulus n_1 zum Zeitpunkt t_1 gewechselt, pausiert das Video und springt an den Zeitpunkt $t_{neu} = t_{start}(n_1) + t_1 + t_{safe}$. Im Sicherheitszeitraum t_{safe} werden alle Operationen abgeschlossen und die neue Videoposition liegt im Vergleich zur zeitlichen Referenz 'in der Zukunft'. Nach Erhalt des Stimuluswechselbefehls schickt das Audiosystem im Abstand von 1 ms den momentanen Stand des Zeitzählers t an das Videosystem. Ist $t_{neu} \leq t$, holt also die Audioposition die Videoposition wieder ein, wird die Videowiedergabe gestartet. Das Userinterface erlaubt einen Situationswechsel nur alle t_{safe} Sekunden.

Zur Verifizierung der entwickelten Synchronisierungsmethodik wurde eine weitere Messung mit folgendem Testaufbau durchgeführt:

- der Audiorechner mit der finalen Hörversuchsumgebung
- die Audiostimuli wurden durch Spuren mit Sinusbursts im Sekundenabstand ersetzt
- ein Situationswechsel wurde ebenfalls durch einen Sinusburst auf einer separaten Spur auralisiert
- das Video wurde mit derselben Audiospur unterlegt
- auf einem weiteren Rechner[4] wurde
  - eine Spur mit dem Audiorechnersignal aufgenommen
  - eine Spur mit dem Audiosignal des Videorechners aufgenommen
  - eine Spur mit den Situationswechseln aufgenommen

Nach 12 Situationswechseln und darauffolgendem Ablesen der Zeitdifferenzen, sowie Entfernung eines Gleichanteils, ergab die Messung eine Varianz von maximal $\pm 12,5ms$, einem nach [44] unkritischen Versatz. Die finale absolute Bild-Ton-Synchronisierung und damit Entfernung aller Gleichanteile wurde durch drei Expertenhörer subjektiv vorgenommen. Hierfür konnte im Pure Data Patch ein situationsspezifischer Bild-Ton-Versatz eingetragen werden.

[4] Lenovo X60 Notebook mit RME Multiface und Samplitude

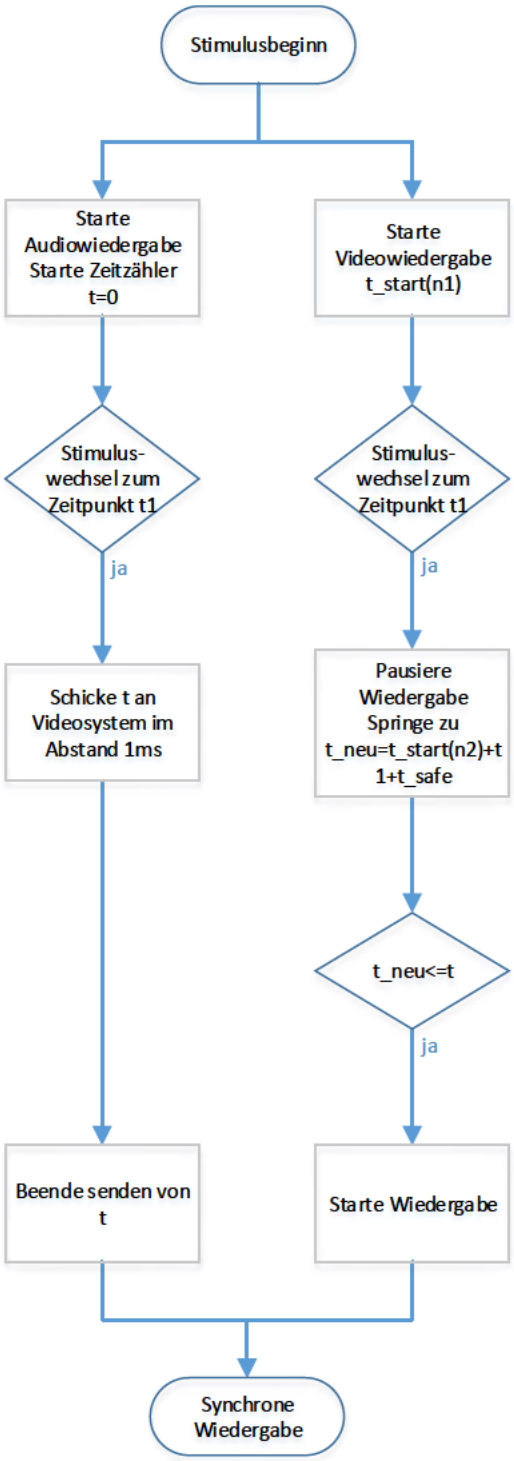


Abbildung 5.3.: Prinzipieller Ablauf eines Stimuluswechsels.

6. Methoden - Hörversuch

Abbildung 6.1.: Foto der Cave mit Versuchsperson

Die Umsetzung der in Kapitel 2.2 eingeführten RGT-Methodik wird hier vorgestellt. Dazu gehört die Auswahl und Zuordnung der Stimuli zu Triaden, die Festlegung und Akquirierung der Stichprobe und die Durchführung des Versuches. Dies wurde nicht vom Autor umgesetzt, soll hier aber aus Gründen der Vollständigkeit dokumentiert werden.

6.1. Triadengenerierung und Stichprobenbetrachtung

Zunächst müssen mit den verfügbaren Stimuli Triaden gebildet werden. Diese sollen möglichst alle wichtigen Wahrnehmungsunterschiede evozieren bei einer den Versuchspersonen zumutbaren Hörversuchsdauer. Dies resultierte in 9 Triaden mit unterschiedlichen Rezeptionssituationen und somit insgesamt 18 Triaden unter Berücksichtigung

beider Musikstücke.

Die Stimuluskombinationen wurden hierbei entsprechend der Forschungsfrage gewählt und beinhalten, bis auf einen Fall, immer mindestens eine Livesituation im Vergleich zu medialen Darbietungen. Die Livesituation mit sozialer Komponente wurde bewusst nicht direkt mit der Livesituation ohne Hintergrundgeräusche verglichen, um diese nicht zu offensichtlich werden zu lassen und so evtl. andere Wahrnehmungsunterschiede zu unterdrücken. In Tabelle 6.1 sind die finalen, im Hörversuch randomisiert angebotenen Triaden aufgeführt.

Triade	**Stimulusnr.**			**Stimulus**		
1	S1	S2	S9	Real Pl.1	Real Pl.2	Decca 5.0
2	S1	S7	S6	Real Pl.1	XY 2.0	AB 2.0
3	S3	S9	S13	Real Pl.1+MTB	Decca 5.0	WFS konv
4	S3	S6	S11	Real Pl.1+MTB	AB 2.0	AB KH
5	S1	S8	S11	Real Pl.1	ORTF 2.0	AB KH
6	S3	S4	S5	Real Pl.1+MTB	Real Pl.2+MTB	Kugel 1.0
7	S1	S12	S14	Real Pl.1	XY KH	WFS krea
8	S1	S9	S10	Real Pl.1	Decca 5.0	INA3 5.0
9	S6	S5	S8	AB 2.0	Kugel 1.0	ORTF 2.0
10	S15	S16	S23	Real Pl.1	Real Pl.2	Decca 5.0
11	S15	S21	S20	Real Pl.1	XY 2.0	AB 2.0
12	S17	S23	S27	Real Pl.1+MTB	Decca 5.0	WFS konv
13	S17	S20	S25	Real Pl.1+MTB	AB 2.0	AB KH
14	S15	S22	S25	Real Pl.1	ORTF 2.0	AB KH
15	S17	S18	S19	Real Pl.1+MTB	Real Pl.2+MTB	Kugel 1.0
16	S15	S26	S28	Real Pl.1	XY KH	WFS krea
17	S15	S23	S24	Real Pl.1	Decca 5.0	INA3 5.0
18	S20	S19	S22	AB 2.0	Kugel 1.0	ORTF 2.0

Tabelle 6.1.: Zuordnung der Stimuli zu Triaden. Die mittlere Spalte enthält Stimulusnummern entsprechend Tabelle 5.2, die rechte Spalte enthält die entsprechende Verbalisierung. Die Triaden 10-18 enthalten dieselben Rezeptionssituationen wie Triaden 1-9, benutzen aber das populäre Musikstück

Abhängig von der optischen Wahrnehmungssituation wurden die Triaden folgenden drei Konzeptgruppen zugeordnet (siehe auch Abbildung 6.4):

- **rein akustisch (A)**, ohne Bild
- **optoakustisch (B)**, akustische und optische Darbietung
- **optoakustisch-medial (C)** akustische und optische Darbietung mit Wiedergabe des Musikvideos in den medialen Szenen

Versuchspersonen wurden einer Konzeptgruppe zugeordnet und erhielten dementsprechend nur diese optische Darbietung. Das Ziel ist eine differenziertere Betrachtung der akustischen Unterschiede (A), der visuellen raumbezogenen Unterschiede (B) und der visuellen Interaktionsunterschiede (C) frei von Konzept-Skalen-Interaktion. Werden Qualitäten durch nur eine Konzeptgruppe hinzugefügt, kann davon ausgegangen werden, dass diese konzeptspezifisch sind. Über alle Gruppen konstante Qualitäten sind dann entweder auditiv oder supramodal begründet. In quantitativen Untersuchungen könnte dann ein modularer Fragebogen erstellt werden, der je nach AV-Bedingung die konzeptspezifischen Attribute enthält. Sowohl rein auditive als auch supramodale Attribute wären dann über alle Bedingungen vergleichbar.

Um möglichst ausdifferenzierte Attribute aus einem breiten Wahrnehmungsraum zu erhalten, ergeben sich Anforderungen an die Stichprobe. Die Versuchspersonen sollten über gute Artikulationsfähigkeit verfügen, um differenzierte Konstrukte kommunizieren zu können. Diese können detaillierter ausfallen, wenn der Proband bereits über Vorerfahrung verfügt. Um auch hier auf einen breiten Wahrnehmungsraum zurückgreifen zu können, soll die Vorerfahrung über die einzelnen Versuchspersonen ebenfalls möglichst weit gestreut sein, bestehend aus regelmäßigen Konzertbesuchern, aber auch an Bild- und Musikproduktion aktiv beteiligten Personen, wie beispielsweise Tonmeister, Musiker, Musiklehrer, Cutter oder Kameraleute. Je unabhängige Konzeptgruppe werden mindestens 6 VP veranschlagt (je gestuft nach 3 Altersgruppen und den zwei Geschlechtern, siehe Tabelle 6.2). Die letztendlich realisierte Stichprobe wurde in Tabelle 6.3 zusammengetragen.

Die Vorauswahl wurde durch Onlinefragebögen getroffen.

	Männlich			Weiblich		
Alter	**Bedingung (A)**	**Bedingung (B)**	**Bedingung (C)**	**Bedingung (A)**	**Bedingung (B)**	**Bedingung (C)**
<30 J.	1	1	1	1	1	1
30–55 J.	1	1	1	1	1	1
>55 J.	1	1	1	1	1	1

Tabelle 6.2.: Stichprobenplan

6.2. Umsetzung

Vor Beginn des eigentlichen Hörversuchs wurde für jede Versuchsperson die Hörschwelle mit Hilfe einer Békésy-Audiometrie im Festfrequenzverfahren bestimmt[1]. Ebenso überprüft wurde weiterhin die Stereosehfähigkeit mit Hilfe zuvor angefertigter Testbilder. Diese enthielten rauschähnliche Muster, in die nachträglich Disparitäten in Form von geometrischen Figuren und Zahlen eingefügt worden sind. Diese sind nur erkennbar bei uneingeschränkter Stereosehfähigkeit. Dieser Stereopsistest wurde in das Wiedergabevideo eingefügt, konnte über das Pure Data GUI ausgewählt werden und wurde mit den Personen vor dem Hörversuch in der Cave durchgeführt.

Die Probanden saßen auf einem für sie höhenangepassten Sessel und steuerten den Versuch über ein Notebook, das auf einem kleinen beweglichen Beistelltisch platziert wurde. Eine USB-Schwanenhalslampe beleuchtete die Tastatur (Abbildung 6.1). Zu Beginn wurde der Proband in die Versuchsmethodik eingeführt (Aufgabenstellung, Technik/3D & Kopfbewegungen, Verwendung eigener Worte, Trennung mehrerer Aspekte in einzelne, erwartete Versuchsdauer). Dann wurde der Intertragusabstand bestimmt, der Kopfhörer BK211 und die Stereobrille aufgesetzt und das Trackingsystem referenziert. Der Versuch wurde über die Hörversuchssoftware *WhisPER* gesteuert. Das Interface für den Elizitationsteil ist in Abbildung 6.2 dargestellt. Durch Drücken der großen X, Y, Z Schaltflächen konnten die Versuchspersonen instantan zwischen den Stimuli umschalten. Nach der Wahl des sich unterscheidenden Hörbeispiels sollen die qualitätsbeschreibenden Eigenschaften genannt werden. Je Triade konnten auch mehrere Attribute genannt werden, dabei lag die höchste erreicht Anzahl bei 30. Dies wird für alle 18 Triaden wiederholt und geschieht in Anwesenheit des Versuchsleiters.

[1] *ear 1.7 digital PC-Audiometer* der Firmen 'Titel++ System und Applikationsentwicklung' und 'AURITEC Medizinische Systeme GmbH' `www.ear20.de`

	Männlich			**Weiblich**		
Alter	**Bedingung (A)**	**Bedingung (B)**	**Bedingung (C)**	**Bedingung (A)**	**Bedingung (B)**	**Bedingung (C)**
<30 J.	Student Audio-kommunikation (AK)	Hobbymusiker	Student populäre Komposition & Prod.	Konzertmeisterin	Audioingenieurin Orcherstermusikerin	Studentin AK
30 – 55 J.	Musik-wissenschaftler	-	Kulturradio-moderator	Klangkünstlerin	-	TV-Regisseurin
>55 J.	Instrumenten-bauer Mischtonmeister	Musik & Technikfreund	Filmtonmeister	Klassik-liebhaberin	Musik-wissenschaftlerin	Klavier-lehrerin

Tabelle 6.3.: Realisierte Stichprobe mit Angaben über den jeweiligen Beruf, bzw. das Interessengebiet der Probanden.

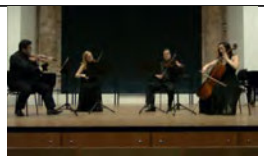
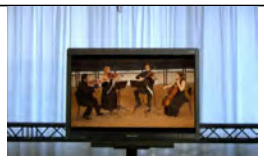
Rezeptions-situation	**Optische Bedingungsvariation**		
	Rein akustisch (A)	**optoakustisch (B)**	**optoakustisch-medial (C)**
Livekonzert			
Tonstudio			
WFS-Labor			

Tabelle 6.4.: Finale Stimuli sortiert nach getrennten optoakustischen Konzeptgruppen

Dieser sorgte für eindeutige Zuweisung von Konstrukten zu Modalitäten und begriffliche Trennschärfe (so wurden bei unspezifischen Angaben wie 'höher/tiefer' die Modalität (Bild/Klang) und das Bezugsobjekt (Musiker/Raum) erfragt) und vermied unter Umständen triviale Antworten (Konzertsaal, kein Konzertsaal) bei möglichst geringer Beeinflussung des Antwortverhaltens der Versuchspersonen. In einem zwischen Elizitation und Rating geschalteten Nachinterview wurden Doppelungen, Tippfehler und finale Missverständlichkeiten behoben.

Nach dem Elizitationsschritt werden alle in den Triaden vorkommenden Stimuli mit den zuvor generierten Konstrukten auf einer siebenstufigen Skala bewertet (Abbildung 6.3).

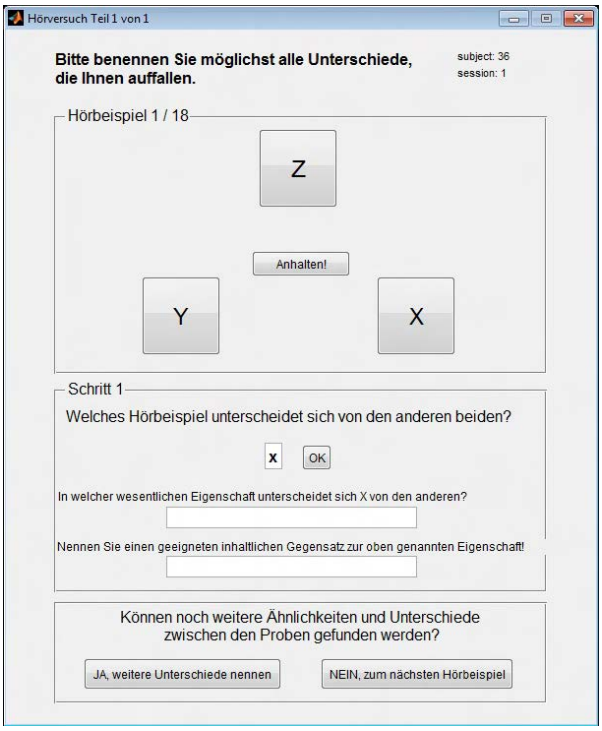


Abbildung 6.2.: Benutzerinterface während des Elizitationsschritts

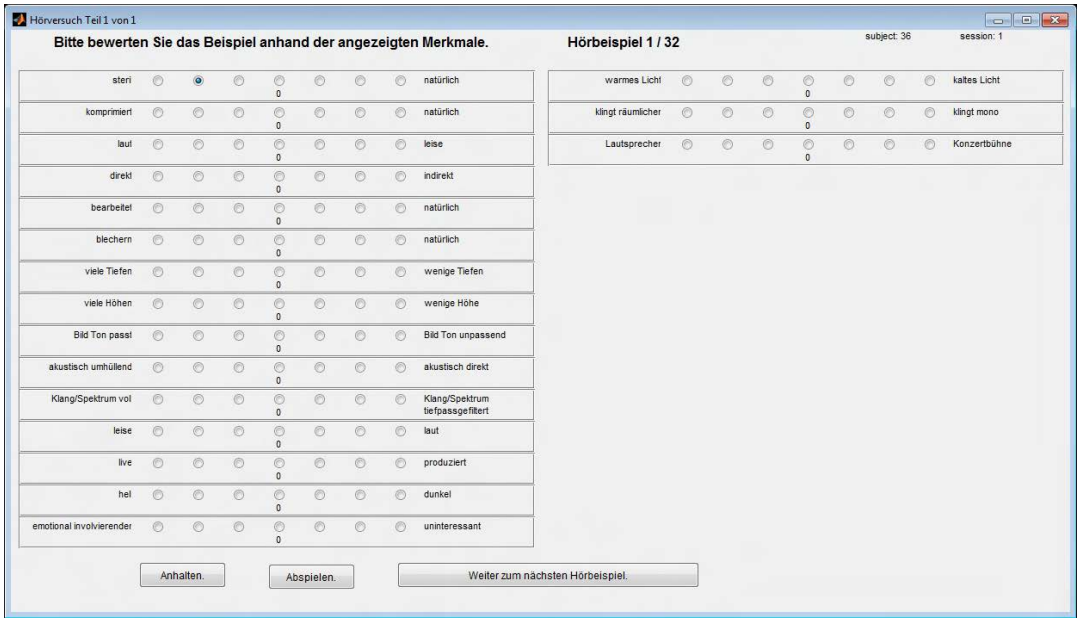


Abbildung 6.3.: Benutzerinterface während der Stimulusbewertung

7. Resultate

Die insgesamt 18 Hörversuchsteilnehmer bildeten jeweils 12-30 Konstrukte während einer mittleren Versuchsdauer von 4,1 Stunden. Die erhobenen Daten über alle unabhängigen Variablen bieten detaillierten Zugriff auf die perzeptive Wirkung von 'Musikstück', 'Rezeptionssituation', 'Abmischung', 'sozialer Kontext' und 'Bildinhalt'. In dieser Arbeit werden allerdings nur die Unterschiede zwischen den beiden Rezeptionssituationen 'medial' und 'live' anhand zweier Versuchspersonen exemplarisch betrachtet.

Abbildungen 7.1 und 7.2 zeigen Polaritätsprofile zweier Versuchspersonen (VP) aus den Konzeptgruppen (A) (rein akustische Wiedergabe) und (C) (optoakustisch-mediale Wiedergabe).

Unabhängig von der Versuchsperson lassen sich die hier erhobenen Wahrnehmungsqualitäten aufteilen in

- **Klangaspekte** (Intensität, Umhüllung, Klanggeometrie, Frequenzverteilung, Durchhörbarkeit, Instrumentenbalance, Nachhall, Dynamik),
- **Bild-/Tonaspekte** (Ton/Bild Passung, Farbintensität, Farbspektrum, Bilddramaturgie, Musikerinteraktion) und
- **supramodale Kategorien** (Natürlichkeit, Emotionalität, Sozialität, Öffentlichkeitsgefühl, Aufmerksamkeit).

Die medialen Produktionen wurden in beiden Fällen als ausgewogener bezüglich der Instrumentenbalance bewertet. Während VP 2 aber eine weniger präsente Stimme als 'Ensemblespiel' und weniger gut erkennbare Einzelinstrumente als 'symbiotisches Klanggefüge', eher positiv konnotierte Begriffe, bezeichnete, waren diese Aspekte für VP 14 Zeichen einer schlechten, technisch unbefriedigenden Aufnahme.

Die besser bewertete, eigentlich artifizielle, 'natürliche Räumlichkeit' der medialen Darbietungen, sowie deren bessere Durchhörbarkeit und Klarheit wurden von VP 14

insgesamt bevorzugt. Ein Indikator für die in der Einleitung erwähnte These von Stolla [2] über die Funktion medialer Produktionen als kompensierendes Moment des fehlenden optischen Reizes. Dahingegen liegt der Fokus der VP 2 eher auf der 'Natürlichkeit' eines Konzertbesuchs mit ihren signalverarbeitenden Unzulänglichkeiten und Umgebungsgeräuschen.

Die Wiedergabe eines Musikvideos zu einer akustisch motivierten Abmischung wirkte für VP 2 'irritierender' durch im Konflikt stehende Bild-/Ton-Quellenorte und eine unpassende Bilddramaturgie. Ein Hinweis auf die ebenfalls in der Einleitung erwähnte These, dass audiovisuelle Rezeptionssituationen andere Abmischungen erfordern können, als rein auditive [1].

Auch wenn die gerade getätigten Aussagen nur als leichte Tendenzen zu verstehen sind und erst in einem quantitativen Versuch verifiziert werden können, lassen sich schon, unter Berücksichtigung der Wahrnehmungsunterschiede aller Versuchspersonen, folgende Richtungen erkennen:

- Die Konzertsaalsituation wird oft als anstrengender und weniger 'schön' (oft aufgrund zu leiser Sängerin) bewertet. Das Gefallen geht insgesamt öfter zugunsten der medialen Rezeptionssituation.

- Es sind zwei Personengruppen erkennbar. Eine präferiert die 'natürlichere', 'ausgewogenere' Livesituation, die andere die besser 'durchhörbare', 'präzise' mediale Produktion.

- Die Bewertung von supramodalen Begriffen wie 'Natürlichkeit' hängt stark mit der gerade genannten Präferenz zusammen. So wurde unter anderem der 'mediale' Stimulus natürlicher als der 'reale' empfunden.

Betrachtet man nun die 40 elizitierten Attribute aller Versuchspersonen, lassen sich diese in 20 Kategorien bzw. Unterkategorien (Tabelle 7.1) gliedern (nicht vom Autor durchgeführt). Für den folgenden quantitativen Versuch kann nun ein semantisches Differential erstellt werden, mit konzeptspezifischen, trennscharfen Gegensatzpaaren, basierend auf diesen Kategorien. Es kann davon ausgegangen werden, dass hiermit alle möglichen Wahrnehmungsunterschiede zwischen live aufgeführter und medial rezipierter Musik erfasst wurden.

Abschließend lässt sich die Effizienz des gezeigten Verfahrens anhand eines Vergleichs zur in Kapitel 2.1 präsentierten Arbeit von Reinecke [7] darstellen. Tabelle 7.2 zeigt

dort genutzte Gegensatzpaare und eine Auswahl der Wahrnehmungsqualitäten der beiden hier vorgestellten Polaritätsprofile. Der Autor hat eine (rein subjektive) Zuordnung getroffen, wo die in dieser Arbeit elizitierten Qualitätsunterschiede sich in den 'standardisierten' Gegensatzpaaren bei Reinecke wiederfinden könnten. Vor allem im direkten Vergleich fällt bei letzterem mangelnde Trennschärfe auf, gezeigt durch Mehrfachnennungen der elizitierten Attribute und die Interaktion mehrerer Konzepte innerhalb eines Gegensatzpaares.

Kategorien	**Subkategorie**	**Kategorien**	**Subkategorie**
Klang	Intensität	Supramodal	Unvermitteltheit
	Distanz		Natürlichkeit
	Raumeindruck		Sozialität
	Klangbildgeometrie		Aufmerksamkeit
	Klangfarbe		Potenz
	Dynamik		Lebendigkeit
	Lokalisation		Kontrolle
	Störungen / Artefakte		Ausgewogenheit
Bild und Klang	Ton-Bild-Passung		Verständlichkeit
	Ton-Bild-Dramaturgie		

Tabelle 7.1.: Kategorien und Unterkategorien zur Erfassung von Wahrnehmungsunterschieden zwischen 'Live' und medial rezipierter Musik

Gegensatzpaare aus [7]	Gegensatzpaare von VP 2 & 14 (Auswahl)
fließend - stockend	
drängend - behaglich	Klangbild: höhenlastig,harsch - warm,ausgewogen lauter - leiser
uneindeutig - bestimmend	Klang: diffus, entfernt - präsent, nah leiser - lauter
geordnet - zufällig	musikalisch werktreu - musikalisch unsauber bearbeitet - natürlich
feierlich - keck	
ernst - verspielt	
schwankend - stabil	Klang: diffus, entfernt - präsent, nah
rau - glatt	Klangbild: höhenlastig,harsch - warm,ausgewogen natürlich - bearbeitet
straff - schleichend	
fein - grob	musikalisch werktreu - musikalisch unsauber
kühl - gefühlvoll	Klangbild: höhenlastig,harsch - warm,ausgewogen erreicht mich emotional kaum - erreicht mich emotional
klar - verschwommen	direkte Aufnahme - große Räumlichkeit (Klang) Klang: präsent, nah - diffus, entfernt musikalisch werktreu - musikalisch unsauber
robust - zart	klingt mächtiger - klingt intimer lauter - leiser
munter - klagend	
voll - leer	Klang voll - Klang leer
dunkel - hell	Klangbild: warm,ausgewogen - höhenlastig,harsch warmes Lampenlicht-helles Tageslicht
aggressiv - friedlich	Klangbild: höhenlastig,harsch - warm,ausgewogen lauter - leiser

Tabelle 7.2.: Subjektiver Vergleich zwischen stimulusbasiert-erhobenen Gegensatzpaaren und aus einer forschungsfragefremden Untersuchung entlehnten Gegensatzpaaren

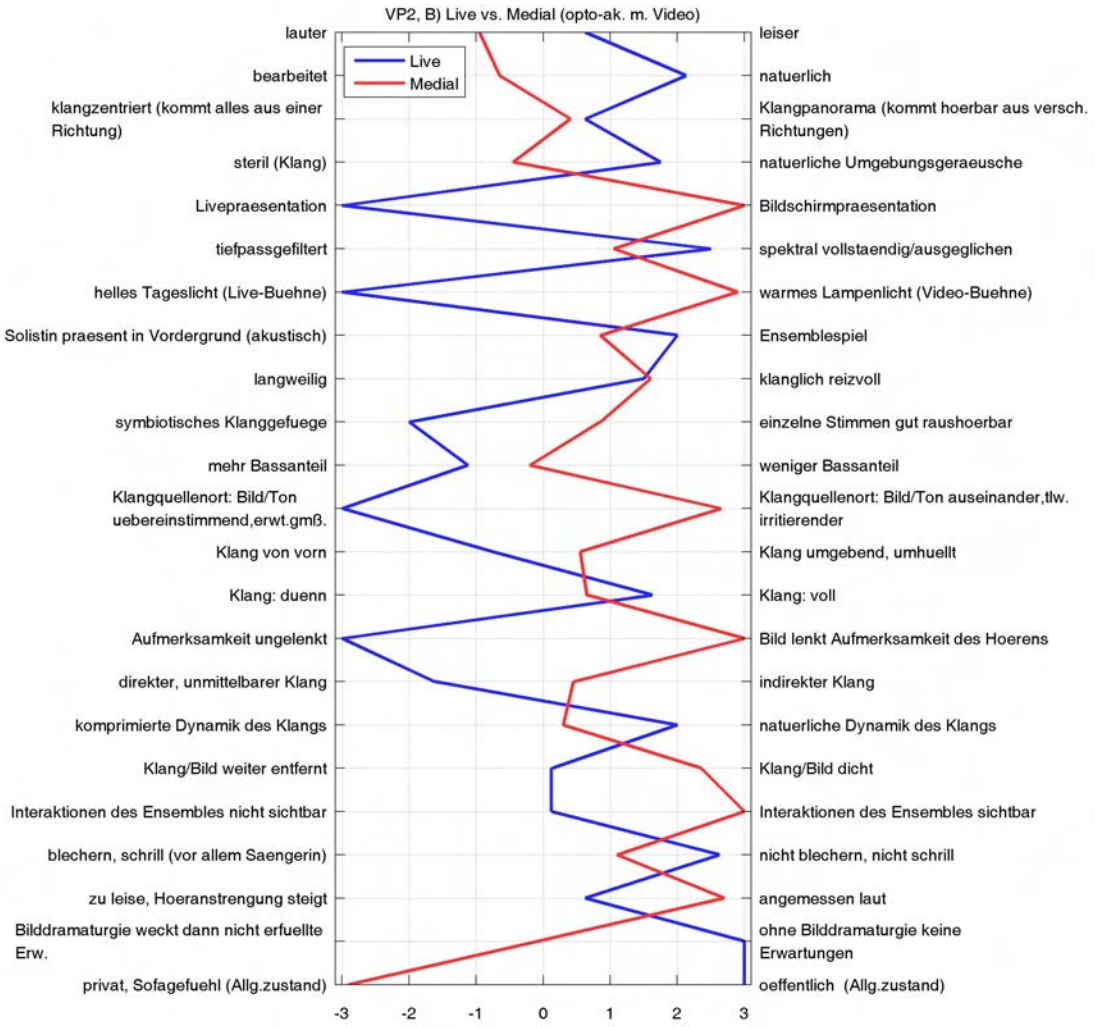


Abbildung 7.1.: Polaritätsprofil der VP 2, optoakustisch mediale Wiedergabe

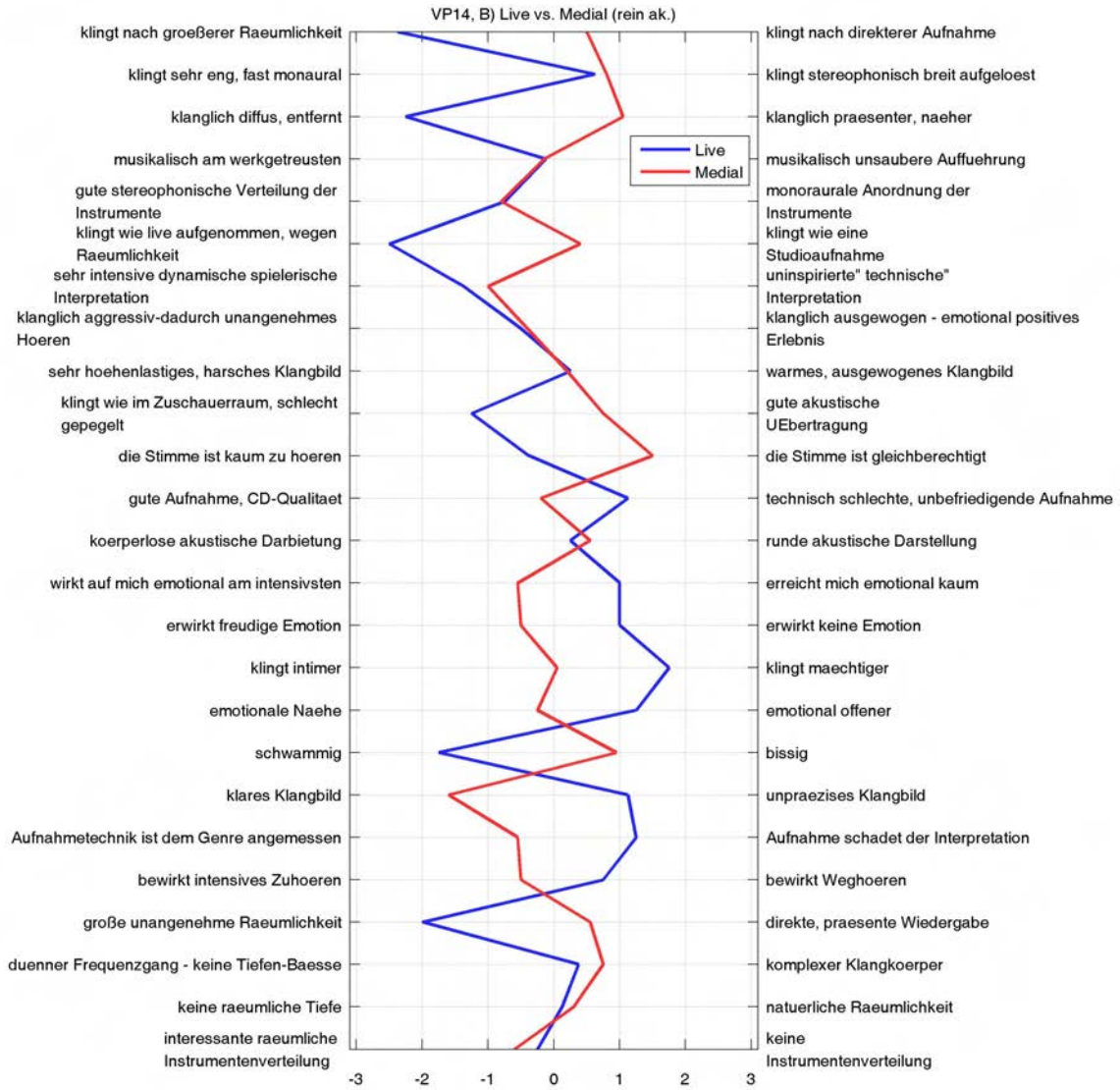


Abbildung 7.2.: Polaritätsprofil der VP 14, rein akustische Wiedergabe

8. Zusammenfassung und Ausblick

Es wurde erstmals ein Messinstrument erstellt, das ein instantanes Wechseln zwischen live aufgeführter und medial produzierter Musik in einer immersiven, audiovisuellen Versuchsumgebung erlaubt. Die akustische Simulation, basierend auf dynamischer Binauralsynthese, wurde hierbei dem *Plausibilitätskriterium* gerecht. Die optische Simulation mit form- und größentreuer stereoskopischer Projektion war *immersiv* und nah an der technisch umsetzbaren Grenze. Dies ermöglichte erstmalig eine direkte Vergleichbarkeit zwischen 'realer' und 'medialer' Musik und damit die Anwendung der *repertory grid technique* in einem kontrollierten Laborexperiment. Das resultierende, die Wahrnehmungsunterschiede beschreibende Vokabular ist frei von Konzept-Skalen-Interaktion, trennscharf und durch geeignete Versuchspersonenwahl repräsentativ. Dies kann nun die Basis für einen quantitativer Versuch innerhalb derselben Versuchsumgebung bilden, um bereits absehbare Tendenzen zu Wahrnehmungsunterschieden zu verifizieren oder zu widerlegen.

Da die Daten in einem Labor erhoben wurden, besitzen sie unvermeidlich eine geringere ökologische Validität als Feldversuche, insbesondere im Bereich nicht-akustischer sensorischer Eindrücke und sozialer Aspekte. Die Effizienz des gezeigten Verfahrens ließ sich aber anhand eines Vergleichs zu einer einschlägigen auf Feldversuchen basierenden Vorstudie [7] darstellen.

Langfristig gesehen kann dieses Vokabular ebenfalls die Basis für einen ökologisch validen Feldversuch bieten. Die *Digital Concert Hall* der Berliner Philharmoniker[1] verfügt über die nötige Technik, Infrastruktur und musikalische Expertise, um Befragungen im Konzertsaal zu ermöglichen, aber auch audiovisuelle Produktionen derselben Musikaufführung anzufertigen und separat bewerten zu lassen. Vorstellbar wären hier noch Betrachtungen zu weiteren, bisher im medialen Kontext noch nicht angeführten Variablen, beispielsweise den möglichen Unterschied zwischen der medialen Rezep-

[1] `http://www.digitalconcerthall.com`

tion einer gerade 'live' stattfindenden Aufführung und eines bereits zurückliegenden Konzertmitschnitts.

Solch eine Kooperation würde, Reinecke zitierend, Untersuchungen ermöglichen, die ebenfalls „Zwitter zwischen experimentellem Anspruch und lebendiger Musikdarbietung“ sind, aber mit Befunden großer theoretischer Reichweite.

9. Literaturverzeichnis

[1] Maempel, H.-J. (2001): *Klanggestaltung und Popmusik - Eine experimentelle Untersuchung*. Ph.D. thesis, Technische Universität Berlin.

[2] Stolla, J. (2004): *Abbild und Autonomie: zur Klangbildgestaltung bei Aufnahmen klassischer Musik 1950 - 1994*. Marburg: Tectum Verlag.

[3] Lindau, Alexander; Torben Hohn und Stefan Weinzierl (2007): „Binaural resynthesis for comparative studies of acoustical environments." In: *Proc. of the 122nd AES Convention*. Vienna. Preprint no. 7032.

[4] Meyer, Juergen (1992): „Klangliche Unterschiede zwischen realem Orchester und der Einspielung nachhallfreier Musik in Sälen." In: Bildungswerk des VDT (Hrg.) *Bericht der 17. Tonmeistertagung*. S. 144–145.

[5] Reinecke, Hans-Peter (1978): „Kann man musikalisches Erleben messen?" In: *Kommunikationstechnik 1977. Das wissenschaftlich-technische Rahmenprogramm der Internationalen Funkausstellung 1977 Berlin*. S. 17–31. Berlin: Colloquium Verlag.

[6] Reinecke, H.-P. (1978): „Musik im Original und als Reproduktion." In: *Kommunikationstechnik 1977. Das wissenschaftlich-technische Rahmenprogramm der Internationalen Funkausstellung 1977 Berlin*. S. 140–151. Berlin: Colloquium Verlag.

[7] Reinecke, H.-P. (1980): „Musik im Original und als technische Reproduktion." In: *Jahrbuch Preußischer Kulturbesitz.*, XVI S. 85–103. Hrsg. von Werner Knopp. Berlin: Gebr. Mann.

[8] Lindau, Alexander (2010): „Zu den Dimensionen des Unterschied live aufgeführter und reproduzierter Musik: Ergebnisse einer qualitativ/quantitativen

Umfragestudie." In: *Fortschritte der Akustik: Tagungsband d. 36. DAGA*. Berlin, S. 609–610. URL http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2010/Lindau_2010a_Zu_den_Dimensionen_des_Unterschieds_live_aufgefuehrter_und_reproduzierter_Musik_Ergebnisse_einer_qualitativ-quantitativen_Umfragestudie.pdf.

[9] Berg, Jan und Francis Rumsey (2006): „Identification of Quality Attributes of Spatial Audio by Repertory Grid Technique." In: *J. Audio Eng. Soc*, 54(5) S. 365–379. URL http://www.aes.org/e-lib/browse.cfm?elib=13681.

[10] Kelly, G. (1991): *The Psychology of Personal Constructs*. The Psychology of Personal Constructs. Routledge. URL http://books.google.de/books?id=SSEZrfffTlQC.

[11] Vorländer, Michael (2008): *Auralization : fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Berlin Heidelberg: Springer. doi:http://dx.doi.org/10.1007/978-3-540-48830-9.

[12] Spors, Sascha und Jens Ahrens (2008): „A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling." In: *Audio Engineering Society Convention 125*. URL http://www.aes.org/e-lib/browse.cfm?elib=14708.

[13] Lindau, Alexander (2006): *Ein Instrument zur softwaregestützten Messung binauraler Impulsantworten in mehreren Freiheitsgraden*. Masterarbeit, TU Berlin, FG Audiokommunikation.

[14] Begault, Durand R. et al (2000): „Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source." In: *Audio Engineering Society Convention 108*. URL http://www.aes.org/e-lib/browse.cfm?elib=9204.

[15] Otondo, Felipe und Jens H. Rindel (2004): „The Influence of the Directivity of Musical Instruments in a Room." In: *Acta Acustica united with Acustica*, 90(6) S.

1178–1184. URL http://www.ingentaconnect.com/content/dav/aaua/2004/00000090/00000006/art00017.

[16] Pollow, Martin; Gottfried K. Behler und Frank Schultz (2010): „Musical Instrument Recording for Building a Directivity Database." In: *Fortschritte der Akustik: Tagungsband d. 36. DAGA, Berlin*. Dt. Ges. fÃ¼r Akustik, S. 703–704. URL http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2010/Pollow_2010_MusicalInstrumentRecordingforBuildingaDirectivityDatabase_DAGA36.pdf.

[17] Gray, Henry (1918): *Anatomy, descriptive and surgical*. URL http://upload.wikimedia.org/wikipedia/commons/c/ca/Gray904.png. Public domain, http://upload.wikimedia.org/wikipedia/commons/c/ca/Gray904.png.

[18] Lindau, Alexander; Jorgos Estrella und Stefan Weinzierl (2010): „Individualization of dynamic binaural synthesis by real time manipulation of the ITD." In: *Proc. of the 128th AES Convention*. London. Preprint no. 8088.

[19] Lindau, Alexander und Stefan Weinzierl (2009): „On the Spatial resolution of virtual acoustic environments for head movements in horizontal vertical and lateral direction." In: *Proc. of the EAA Symposium on Auralization*. Espoo. URL http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2009/Lindau_2009b_On_the_Spatial_resolution_of_virtual_acoustic_environments_for_head_movements_in_horizontal_vertical_and_lateral_direction.pdf.

[20] Polack, Jean-Dominique (1992): „Modifying Chambers to play Billiards: The foundations of Reverberation theory." In: *Acustica*, 76 S. 257–272.

[21] Lindau, Alexander; Linda Kosanke und Stefan Weinzierl (2012): „Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses." In: *J. Aud. Eng. Soc.*, 60(11) S. 887–898.

[22] Kosanke, Linda (2010): *Untersuchung des Zusammenhangs physikalische Maße zur Bestimmung der mixing tie in Raumimpulsanworten und ihrem perzeptiven Äquivalent*. Magisterarbeit, FG Audiokommunikation, TU Berlin.

[23] Lindau, Alexander (2009): „The Perception of System Latency in Dynamic Binaural Synthesis." In: *Fortschritte der Akustik: Tagungsband d. 35. DAGA*. Rotterdam, S. 1063–1066. URL http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2009/Lindau_2009_The_Perception_of_System_Latency_in_Dynamic_Binaural_Synthesis.pdf.

[24] Møller, Henrik (1992): „Fundamentals of Binaural Technology." In: *Applied Acoustics*, 36(3/4) S. 171–218.

[25] Møller, Henrik; Dorte Hammershøi; Clemen Boje Jensen und Michael Friis Sørensen (1996): „Transfer characteristics of headphones measured on human ears." In: *J. Audio. Eng. Soc.*, 43(4) S. 203–217.

[26] Schaerer, Zora und Alexander Lindau (2009): „Evaluation of Equalization Methods for Binaural Signals." In: *Proc. of the 126th AES Convention*. Munich. Preprint no. 7721.

[27] Lindau, Alexander und Fabian Brinkmann (2012): „Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings." In: *J. Audio. Eng. Soc.*, 60(1/2) S. 54–62.

[28] Erbes, V. et al (2012): „An extraaural headphone system for optimized binaural reproduction." In: *Fortschritte der Akustik: Tagungsband der 38. DAGA*. Darmstadt, S. 313–314. URL http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2012/Erbes_et_al_2012_An_extraaural_headphone_system_for_optimized_binaural_reproduction.pdf.

[29] Schultz, Frank (2011): *Elektroakustisch optimierte Wiedergabe binauraler Signale*. Masterarbeit, FG Audiokommunikation, TU Berlin.

[30] Lindau, Alexander und Stefan Weinzierl (2012): „Assessing the Plausibility of Virtual Acoustic Environments." In: *Acta Acustic u. w. Acustica*, 98(5) S. 804–810.

[31] Lindau, A. und S. Roos (2010): „Perceptual evaluation of discretization and interpolation for motion-tracked binaural (MTB-) recordings." In: *Bericht der 26. Tonmeistertagung*. S. 660–701.

[32] Roos, Sebastian (2010): *Soundscapes in der dynamischen Binauralsynthese*. Masterarbeit, FG Audiokommunikation, TU Berlin.

[33] Algazi, V. Ralph; Richard O. Duda und Dennis M. Thompson (2004): „Motion-Tracked Binaural Sound." In: *Audio Engineering Society Convention 116*. URL `http://www.aes.org/e-lib/browse.cfm?elib=12644`.

[34] Tauer, H. (2010): *Stereo-3D*. Schiele & Schön. URL `http://books.google.de/books?id=mfJ4HjD6CwEC`.

[35] Maempel, Hans-Joachim und Alexander Lindau (2013): „Opto-acoustic simulation of concert halls - a data-based approach." In: Bildungswerk des VDT (Hrg.) *Bericht der 26. Tonmeistertagung*. Köln, S. 293–309. URL `http://www2.ak.tu-berlin.de/%7eakgroup/ak_pub/2013/Maempel2013_Opto-acoustic_simulation_of_concert_halls_TMT_2012.pdf.pdf`.

[36] Klemmer, Martin (2008): *Erzeugung mehrkanaliger nachhallfreier Musiksignale für die Binauralsynthese*. Masterarbeit, FG Audiokommunikation, TU Berlin.

[37] ITU (2006): „ITU-R BS.775-2, Multichannel Stereophonic Sound System with and without Accompanying Picture. ITU Radiocommunication Assembly."

[38] Estrella, Jorgos (2011): *Real-Time Individualization of Interaural Time Differences for Dynamic Binaural Synthesis*. Masterarbeit, TU Berlin, Fachgebiet Audiokommunikation.

[39] Abel, Jonathan S. und Patty Huang (2006): „A Simple, Robust Measure of Reverberation Echo Density." In: *Proc. of the 121th AES Convention*. URL `http://www.aes.org/e-lib/browse.cfm?elib=13819`.

[40] Genelec (1997): *1031A Data sheet*.

[41] (2013): *http://jackaudio.org/*. URL `http://jackaudio.org/`.

[42] Wright, Matthew; Adrian Freed und Ali Momeni (2003): „OpenSound Control: State of the Art 2003." In: *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal Canada*.

[43] Erbes, Vera (2013): *BKsystem Reference Manual*. audio communication group, TU Berlin. [to be found on the SEACEN fileserver shared database - Proprietary Information -\15 Binaural Reference System\2 BKsystems #001-#005].

[44] Heide, Berta Luise und Hans-Joachim Maempel (2011): „Die Wahrnehmung audiovisueller Synchronität in elektronischen Medien." In: *26th VDT International Convention*. VDT, S. 525–537.

A. Technikspezifikationen des verwendeten Equipments und Messaufbauten

Die Datenblätter von folgenden Geräten befinden sich auf der Daten-CD:

- RME Multiface und Quadmic
- Meyer UPL-1 Lautsprecher
- Genelec 1031a und 7060A
- M-Audio Audiophile 192
- NTI MiniSPL Messmikrofon
- Pholemus FASTRAK
- DPA 4060

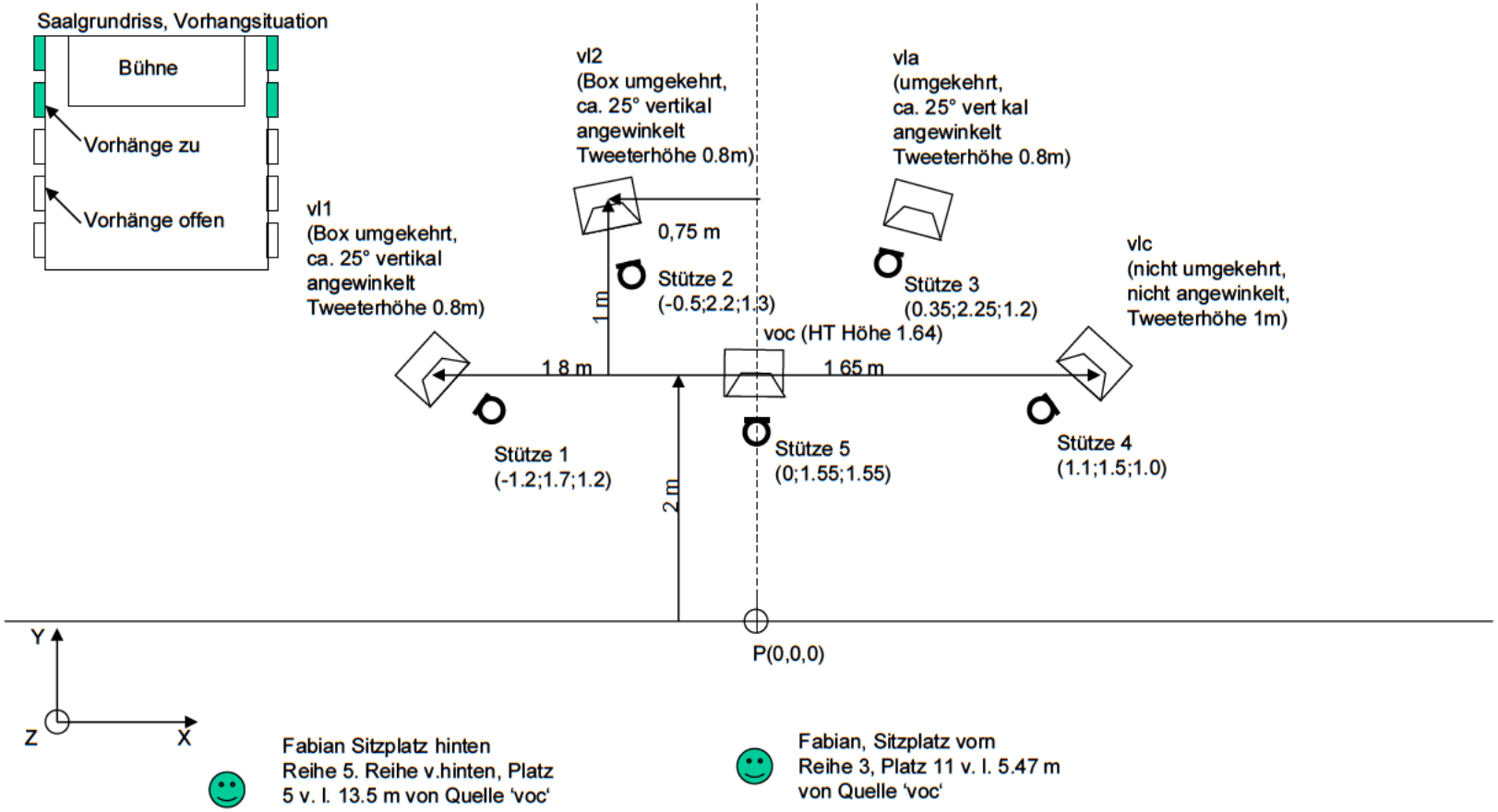
	Audio 1 - 'live', 'medial', MTB	**Audio 2 - 'WFS'**	**Video**
CPU	intel i7-970 6 Kerne@3.2GHz	intel i7-970 6 Kerne@3.2GHz	intel i7-970 6 Kerne@3.2GHz
RAM	12GB	12GB	8GB
OS	Ubuntu 10.10 64Bit Kernel 2.6.35-32-generic	Ubuntu 10.04 Kernel 2.6.32-34-generic	Windows 7 Pro 64Bit SP1
Audio	RME HDSP PCI	M-Audio Audiophile 192	
	RME Multiface		
Video			ATI Radeon HD 7870 Eyefinity 6
			Catalyst Treiberpaket 99.012-121219a-152192C-ATI
Software	Jackdmp 1.9.6	Jackdmp 1.9.7	Pixelwix Pixelwarp Evo v1.4r2
	Pure Data	Pure Data	Stereoscopic Player 1.8.1
	fWonder	fWonder	Matlab R2011a
	stretcher	stretcher	ffdshow video decoder
	tracker3	tracker3	
	jconvolver	jconvolver	
		jack-rack	

Tabelle A.1.: Übersicht der verwendeten Hard- und Software

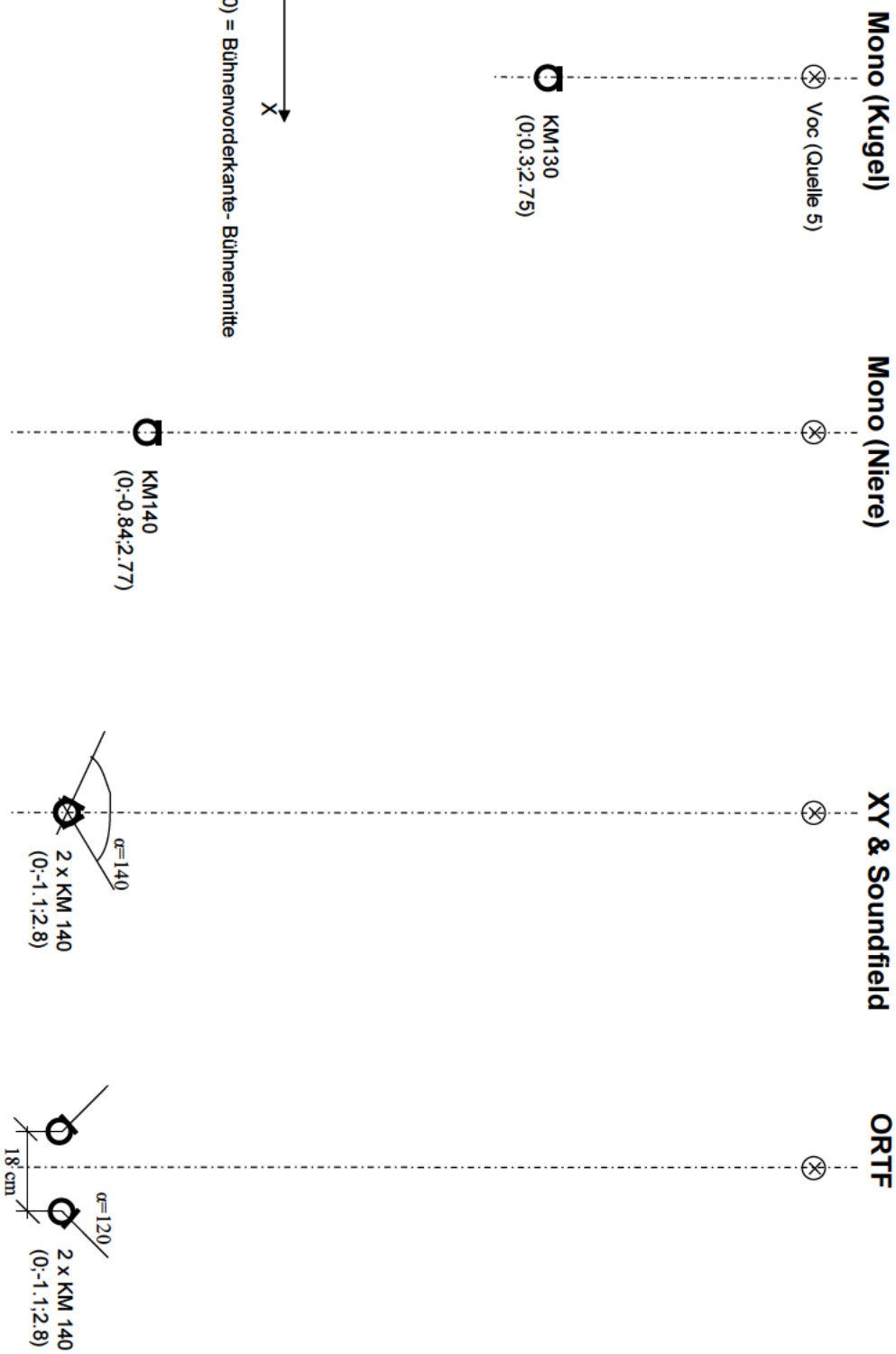
UdK Bundesallee: “Joseph-Joachim-Konzertsaal”, Messungen am 27.08.2010 und am 21.12.2010

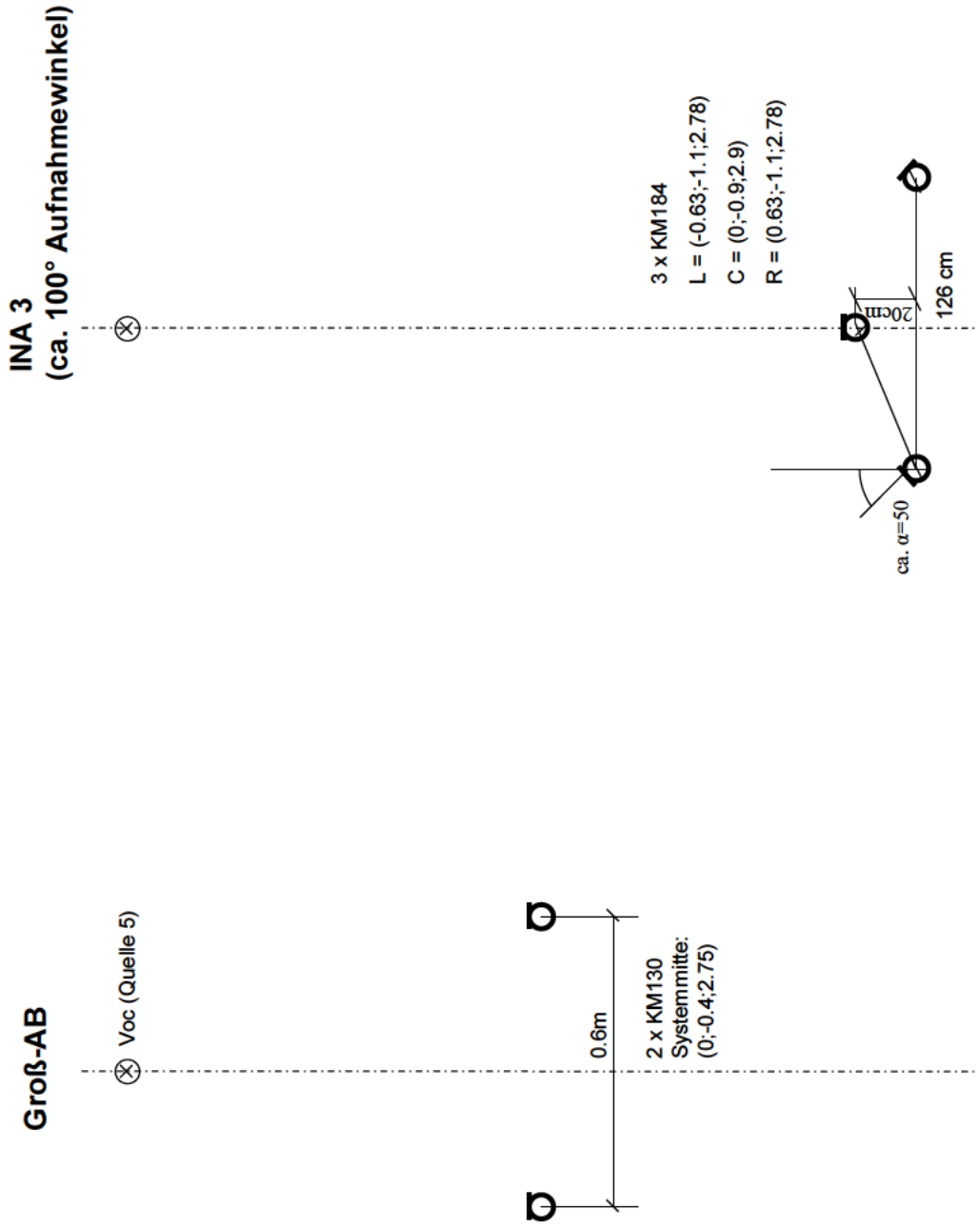
+ Stützen 1-4 an allen 4 Streichern (4 x KM 140)

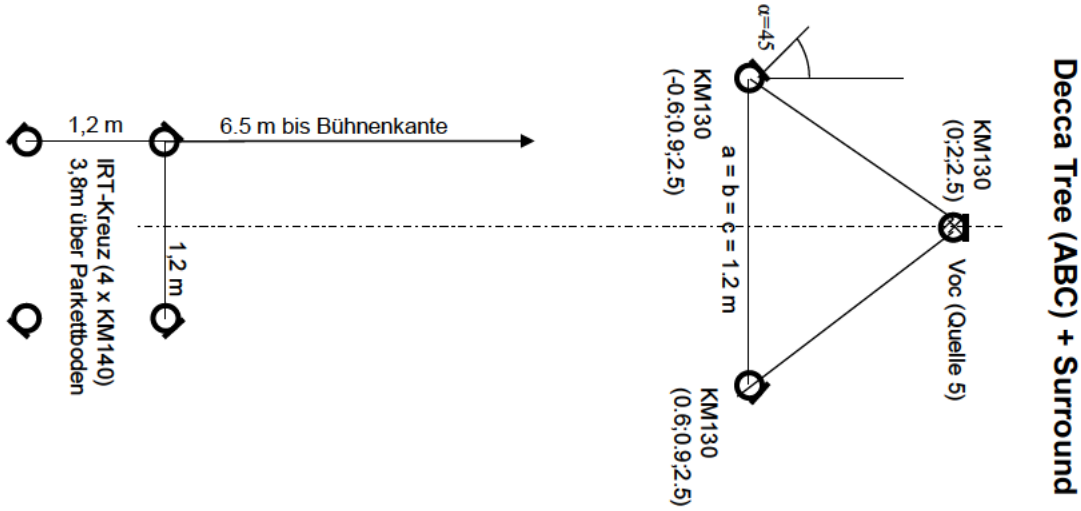
+ Stütze 5 an Sängerin (MG 930)

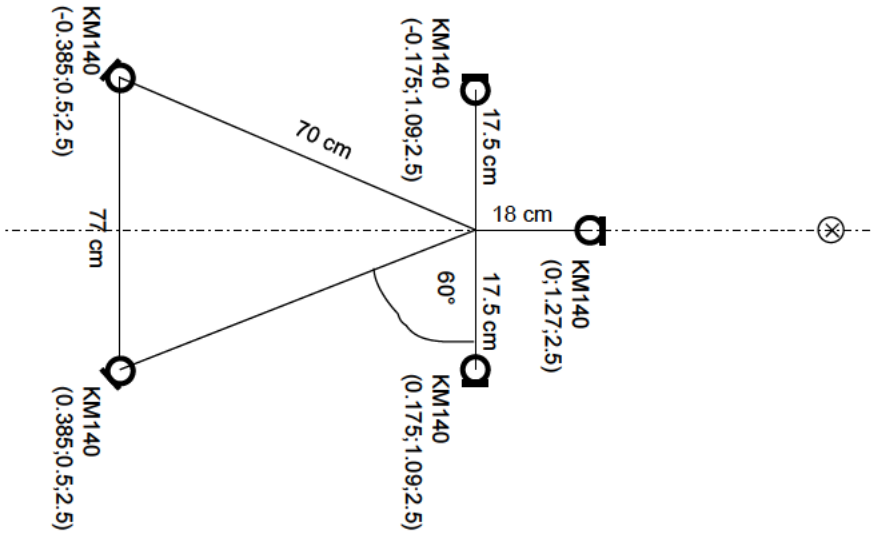


P(0,0,0) = Bühnenvorderkante- Bühnenmitte (Achtung, von FABIANs Sitzplatz aus ist die mittlere Quelle (voc) etwa 4° nach rechts versetzt !!!)
- alle Höhenangaben immer relativ zu Bühnenboden auch wenn Stative vor der Bühne im Parkett standen (Ausnahme: IRT-Kreuz)
- vl1, vl2,vla, voc: Meyer UPL auf Orchesterstühlen, Frequgg. entzerrt auf 1m vor Horn (Behringer DCX)
- voc: auf Meyer-LS Stativ, niedrigster Stufe (siehe auch Dokufotos)


Mono (Kugel)
Mono (Niere)
XY & Soundfield
ORTF
Voc (Quelle 5)
KM130
(0;0.3;2.75)
Y
Z
X
P(0,0,0) = Bühnenvorderkante- Bühnenmitte
KM140
(0;-0.84;2.77)
α=140
2 x KM 140
(0;-1.1;2.8)
α=120
18 cm
2 x KM 140
(0;-1.1;2.8)


INA 3
(ca. 100° Aufnahmewinkel)
Groß-AB
Voc (Quelle 5)
0.6m
2 x KM130
Systemmitte:
(0;-0.4;2.75)
3 x KM184
L = (-0.63;-1.1;2.78)
C = (0;-0.9;2.9)
R = (0.63;-1.1;2.78)
20cm
126 cm
ca. α=50


Decca Tree (ABC) + Surround
KM130
(0;2;2.5)
Voc (Quelle 5)
α=45
KM130
(-0.6;0.9;2.5)
a = b = c = 1.2 m
KM130
(0.6;0.9;2.5)
1,2 m
6.5 m bis Bühnenkante
1,2 m
IRT-Kreuz (4 x KM140)
3,8m über Parkettboden

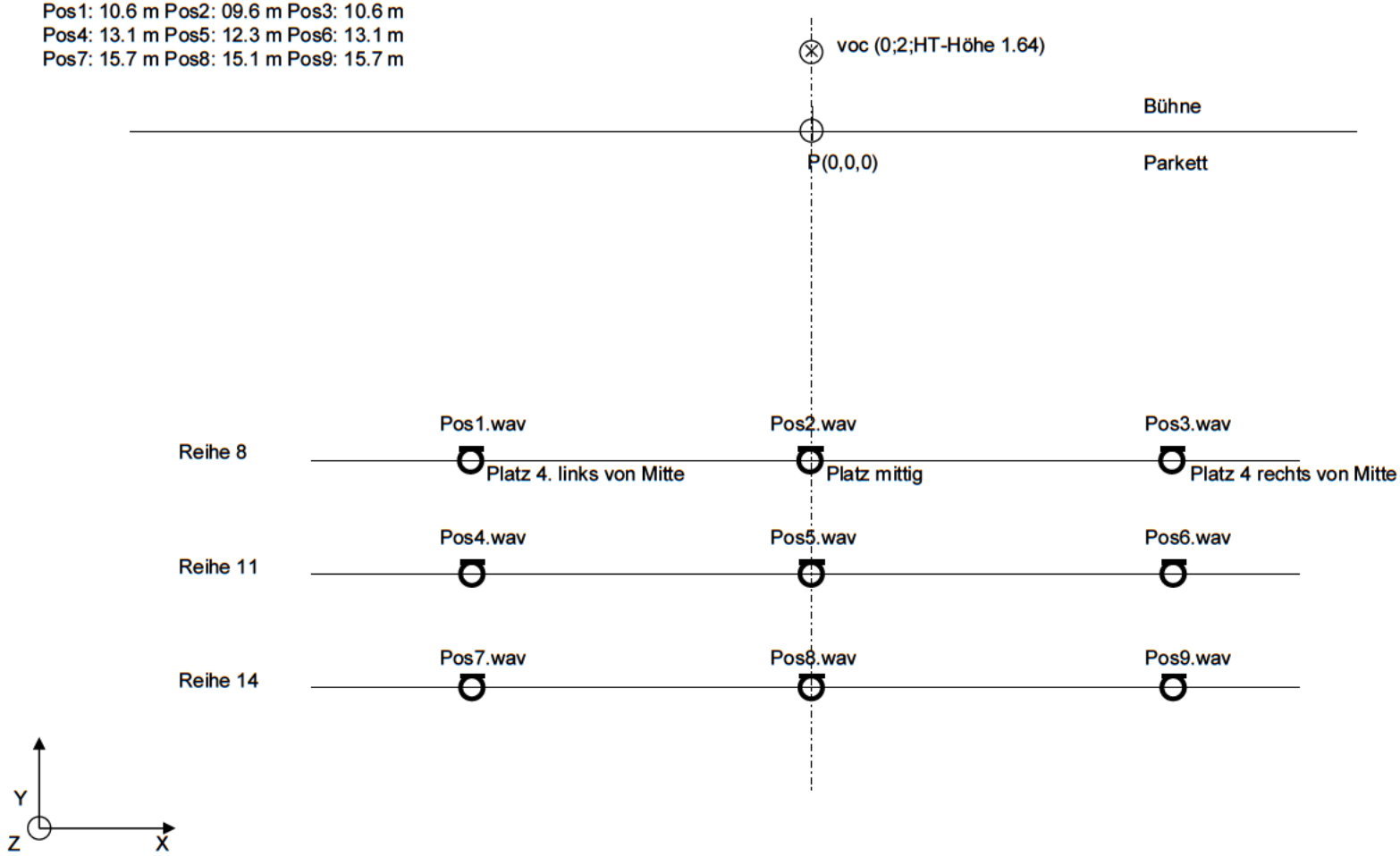

INA 5
KM140
(0;1.27;2.5)
17.5 cm
18 cm
17.5 cm
60°
KM140
(-0.175;1.09;2.5)
KM140
(0.175;1.09;2.5)
70 cm
77 cm
KM140
(-0.385;0.5;2.5)
KM140
(0.385;0.5;2.5)

Diffusfeld/RT-Messungen, gemessen für Quelle ‘ voc‘

Behringer ECM8000 senkrecht auf Stativ, Höhe ca. 1,5m,

Abstände zu Quelle ‘voc’:

Pos1: 10.6 m Pos2: 09.6 m Pos3: 10.6 m
Pos4: 13.1 m Pos5: 12.3 m Pos6: 13.1 m
Pos7: 15.7 m Pos8: 15.1 m Pos9: 15.7 m



B. Pluginparameter der MTB-Entrauschung

Aufgeführt sind die Parameter der genutzten Klangbearbeitungsprogramme zur rauschreduktion der MTB-Aufnahmen. Die Reihenfolge der Tabellen entspricht der Signalkette.

	Modus	**Flankensteilheit/Q**	**Gain**	**Frequenz**
Band 1	high pass	24dB/Oktave	-	127,86
Band 2	normal	1,95	-16,04	234,42
Band 3	normal	1	-16,6	21753,6

Tabelle B.1.: Samplitude EQ116 im linearphasigen Modus, , Gain in dB, Frequenz in Hz

Threshold	**Reduction**	**Attack**	**Release**	**Transients**	**Knee**	**Optimize**
-2	5	0,01	22,44	3,3	25	smooth

Tabelle B.2.: Waves Z-Noise stereo

	Range	**Sens**	**Duration**	**Release**
Band 1	-3,3	6,7	-	-
Band 2	5,7	6,7	-	-
Band 3	6,1	6,7	-	-
Band 4	6,1	6,7	-	-
Xover low	80	-	-	-
Xover Mid	500	-	-	-
Xover High	4000	-	-	-
Master	-	-	4,98	9,95

Tabelle B.3.: Waves TransX Multi Stereo

	Modus	**Q**	**Gain**	**Frequenz**
Band 1	normal	1	-2,1	269,4
Band 3	normal	0,1	-5	5743,6
Band 4	High-shelf	-	-5,2	-

Tabelle B.4.: Samplitude Kanal PEQ, Gain in dB, Frequenz in Hz