

Technische Universität Graz

Universität für Musik und Darstellende Kunst Graz

Diplomarbeit

Optimierung eines Blind Upmix Binaural Systems für die
Kopfhörerwiedergabe

Durchgeführt am :

Fraunhofer Institut für Integrierte Schaltungen Erlangen

Abteilung Audio/Forschung Multimedia

Durchgeführt von: Andreas Wörle

Geboren am : 24.10.1977 in Weilheim i.Obb

Studiengang: F750 Elektrotechnik Toningenieur

Verantwortlicher Professor:

o.Univ-Prof Dr. Robert Höldrich

Betreuung am Fraunhoferinstitut durch :

Dipl.Ing. Oliver Hellmuth

Dipl.Ing Jan Plogsties

Dipl.Ing Harald Mundt

Dipl.Ing Andreas Walther

Zusammenfassung

In dieser Arbeit wird ein bereits existierendes Upmixverfahren mit einem bereits existierenden Verfahren zur virtuellen Mehrkanalwiedergabe so kombiniert, dass maximale Wiedergabequalität für die Mehrheit der untersuchten Testmusikstücke erzielt wird. Genaue Qualitätskriterien, also eine Antwort auf die Frage, was eine als gut empfundene virtuelle Mehrkanalwiedergabe über Kopfhörer ausmacht, werden zu Beginn der Arbeit erarbeitet. Diese Erkenntnisse werden anschließend auf die Parameter beider Systeme zurückgeführt, welche dann im Sinne der Qualitätskriterien entsprechend optimiert werden. Die Arbeit wird mithilfe der Implementierung der Systeme in MATLAB ausgeführt. Zusätzlich wird in der Echtzeitumgebung Pure Data experimentiert. Eine spezielle Hörtestprozedur wird benutzt um die Auswirkungen auf subjektive Bevorzugung zu messen. Abschließend werden die erzielten Verbesserungen mit Hörtests im Vergleich zum unoptimierten System belegt.

Abstract

During this work an already existing technique for stereo to multichannel upmix is combined with a method creating virtual surround sound on a headphone gaining a maximum of quality for the majority of test songs. Criteria, that describe good virtual surround sound for headphones are set up at the beginning of the studies. The cognitions are ascribed to the parameters of both systems and optimized regarding the criteria of quality. All the work is done, using the existing code, written in MATLAB. Additionally the Systems are combined in the realtime environment of Pure Data for experimenting. A special listening test procedure is used to test the parameters of the system for subjective preference. To support the achieved improvement, the new system is to be listening tested versus the unoptimized system.

Inhaltsverzeichnis

1. Einführung.....	1
2. Grundlagen.....	4
2.1. Räumliches Hören.....	4
2.1.1. Interaurale Zeit- und Pegeldifferenz.....	4
2.1.2. Kopfbezogene Außenohr-Übertragungsfunktionen.....	6
2.1.3. Richtungsbestimmende Frequenzbänder.....	7
2.1.4. Raumklang.....	8
2.1.5. Gesetz der ersten Wellenfront.....	11
2.1.6. Distanzwahrnehmung.....	12
2.1.7. Cocktailpartyeffekt.....	14
2.2. Produktevaluierung.....	17
2.2.1. Methoden zur Evaluierung von Audiosystemen.....	17
2.2.2. Subjektive Qualitätsbewertung.....	18
2.2.2.1. Präferenztest.....	18
2.2.2.2. Wahrnehmungsbetonte Tests.....	19
2.3. Surroundverfahren.....	22
2.3.1. Historisches.....	22
2.3.2. Blind Upmix.....	24
2.4. Binauraltechnik.....	25
3. Stand der Technik.....	26
3.1. Blind Upmix nach IIS.....	26
3.2. Kopfhörerwiedergabesystem nach IIS.....	31
3.2.1. Messung der binauralen Raumimpulsantworten (BRIA).....	32
3.2.2. Realisierung des binauralen Wiedergabesystems.....	35
4. Verfahren.....	36
4.1. Das Modell.....	36
4.2. Problemdiagnose.....	39
4.2.1. Finden verbaler Beschreibung.....	39
4.2.2. Ähnliche Arbeiten und Ergebnisse.....	41
4.3. Vorbereitungsphase.....	42
4.3.1. Physikalische Variablen und informelle Hörtests.....	42

4.3.2. Pilothörtests.....	44
4.3.2.1. Spektrale Entzerrung.....	47
4.3.2.2. Upmix Strategien.....	53
4.3.2.3. Filterlängen.....	56
4.3.2.4. Virtuelle Lautsprecher Position	58
4.3.2.5. M-S Dekorrelation der hinteren Kanäle des „In the Band“ Upmix Ansatzes.....	62
4.3.3. Kombinationstest.....	65
5. Verifikation.....	72
6. Zusammenfassung und Ausblick.....	78
7. Literaturverzeichnis.....	80
8. Anhang.....	86

1 Einführung

Kopfhörerwiedergabe ist ein Thema, das in letzter Zeit in steigendem Maß an Wichtigkeit gewinnt. Neben der Beobachtung im täglichen Leben, wo portable Musikwiedergabesysteme vermehrt an Bedeutung gewinnen, kann auch im wissenschaftlichen Bereich verfolgt werden, dass in Form von Studien aktuelleren Datums, der Fokus auf Kopfhörerwiedergabe Systeme verstärkt wird[1],[2],[3].

Es ist nicht zuletzt die Existenz neuer, Platz sparender Audioformate und deren Verfügbarkeit über Internet, die diese Entwicklung in den letzten Jahren förderte. So wird eine Studie von Morgan Stanley aus dem Jahr 2005 zitiert [1], dass im Sommer-Ferienzeitraum des Jahres in den USA mehr Geld für tragbare Musikwiedergabegeräte ausgegeben wurde als für andere Unterhaltungselektronikprodukte. Auch sieht man sich bei der genaueren Betrachtung eines Mobiltelefons der neueren Generationen oftmals einem multifunktionalen Gerät gegenüber, das als tragbares Musikwiedergabegerät benutzt werden kann und wird. Mobiler Musikgenuss gewinnt also mehr und mehr an Bedeutung wobei nicht auf gewohnte Qualität und Funktionalität verzichtet werden soll. Man möchte alle Möglichkeiten der Musikwiedergabe haben, und das unabhängig vom Aufenthaltsort. Vorherrschende mobile Audiowiedergabe erfolgt über Kopfhörer.

Seit dem Einzug neuartiger Surround Heimkinoanlagen als Standard Unterhaltungselektronik in moderne Wohnzimmer und neben der Verfügbarkeit von Filmen mit Surround Ton ist der Markt der Audio Aufnahmen in Surround ein ständig wachsender. Wird Qualitätserweiterung für Musikwiedergabe konstatiert? Eine realgetreue Abbildung musikalischer Klangszenerien und Erweiterung der Hörebene im Vergleich zu Stereo sind Tatsachen. Menschen gewöhnen sich daran vom Klang des Geschehens umgeben zu sein, mehr noch, mitten drin zu sein. Musik hören wird zum Erlebnis. Ein Erlebnis, unabhängig von Konzertterminen, das auf Knopfdruck wiederholbar ist.

Eine andere Art der Realität will der Mensch, der sich mit Computerspielen beschäftigt und sich dabei durch virtuelle Welten bewegt, erleben. Die Möglichkeiten von Computerspielen werden mit virtuellen 3-D Klangszenerien um die Hördimension erweitert. Nun gibt es diverse Möglichkeiten eine virtuelle Klangwelt zu erschaffen, und alle haben Vor- und Nachteile. Entscheidender Vorteil für breitenwirksame Produkte ist kein zusätzlicher Aufwand für den Kunden. Dieser Aufwand lässt sich mit einem Kopfhörerwiedergabesystem sehr gut kontrollieren. Der Kunde benötigt, um den vollen Genuss der dritten Dimension zu

erleben, nichts weiter als einen Kopfhörer. Viele der Ideen und Ansätze zur Realisierung solcher 3-D Systeme kommen aus eben diesem Bereich der Computerspiele. Es stellt sich die Frage, ob es nicht auch einen Mehrwert für die Audiowiedergabe darstellen kann, eine qualitativ hochwertige räumliche Surroundwiedergabe mit dem Kopfhörer zu simulieren. Stand der Technik für portable Audiowiedergabegeräte ist ein Kopfhörer mit zwei Kanälen. Der Großteil der konsumierten Musik ist ebenfalls in Stereo, also mit zwei Kanälen produziert, wenn gleich wie bereits erwähnt das Angebot für Surroundaufnahmen steigt. Für die Wiedergabe über Lautsprecher gibt es bereits Systeme, die zwei Kanäle zu fünf konvertieren, um Stereoaufnahmen für Film oder Musik auf den mehr oder weniger standardisierten 5.1 Heimkinoanlagen wiederzugeben. Wie also das Klangbild, dass die Vorteile einer 5-Kanalwiedergabe hat, auf zwei Kanälen im Kopfhörer reproduzieren? Die Antwort für alle Stereoaufnahmen könnte eine Kombination aus einem intelligenten Blind Upmix- und einem Binauralwiedergabesystem sein (Abbildung 1). Viele Möglichkeiten gibt es, einen Blind Upmix aus einer Stereoaufnahme zu erstellen [4]. Die meisten sind in ihren Möglichkeiten begrenzt und somit in ihrer Wirkung. Im Rahmen dieser Arbeit wird eine neuartiger Ansatz des Blind Upmix betrachtet.

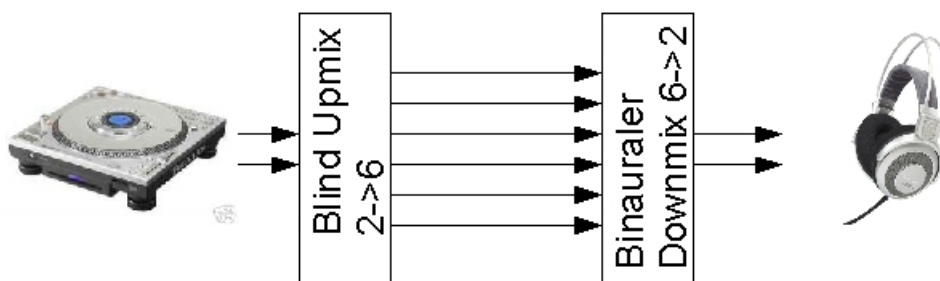


Abbildung 1: Funktionskette Upmix- Binaural Downmix

Binauraltechnik basiert auf Erkenntnissen, die bereits Anfang des 20. Jahrhunderts bis in die 80'er Jahre hinein gemacht wurden. Erste Aufnahmen und Rundfunksendungen in dieser Technik wurden in den späten 60'er Jahren veröffentlicht. Erst seit den 90'er Jahren ist es möglich, mittels schneller Prozessoren und Speicher, die einen Bruchteil ihrer bisherigen Größe und Preis haben und in Kombination mit Datenkompressionsverfahren diese Technologie auch für bestehendes Aufnahmematerial massenwirksam ein zusetzen. Erfolgreich kann man diskrete mehrkanalig produzierte Aufnahmen so bearbeiten, dass in der 2-Kanal Binaural Version der Vorteil dieser Mehrkanalvarianten gegenüber einem Standard

Stereo Downmix nachempfunden werden kann. Der Großteil der Musik jedoch wird nach wie vor in 2-Kanal Stereo produziert, woraus die Motivation für Hörer und vielleicht auch Musikproduzenten resultiert, jede dieser Aufnahmen durch eine mit Blind Upmix hergestellte 5-Kanal Version aufzuwerten. Auch bei dieser Aufgabe werden respektable Ergebnisse erzielt, hört man sich derartiges in einem Lautsprecheresetup an. Beide Technologien in ihrem Kontext funktionieren. Die Frage, die in dieser Arbeit beantwortet werden soll ist, ob ihre Kombination auch dahingehend optimiert werden kann, dass sie als Alternative, oder mehr noch, als Qualitätserweiterung zur Stereowiedergabe betrachtet werden kann. Dazu wird in Kapitel 2 ein Grundlagenüberblick über räumliche Wahrnehmung gegeben, bevor ein Einstieg in Methoden der Produktevaluierung und Grundsätzliches zu Surround und Binauraltechnik behandelt wird. Kapitel 3 beschäftigt sich mit der genaueren Behandlung der vorliegenden Technologien, bevor Kapitel 4 das durchgeführte Verfahren zur Auswertung und Optimierung des kombinierten Wiedergabesystems beschreibt. In Kapitel 5 wird die Effektivität der Vorgehensweise gezeigt, bevor in Kapitel 6 eine Zusammenfassung und ein Ausblick gegeben wird.

2 Grundlagen

Das Kapitel behandelt zunächst die Grundlagen der räumlichen Schallwahrnehmung. Einen guten Überblick über bekannte Effekte und elementare Veröffentlichungen bieten Arbeit von Kapralos et al. [5], Jakka [6] und Walther [4]. Ein Einblick in Hörtestmethoden zur subjektiven Bewertung von Wiedergabesystemen wird gegeben, bevor Grundsätzliches zur Entwicklung von Surround- und Binauralwiedergabetechniken besprochen wird.

2.1 Räumliches Hören

Dieses Kapitel beschäftigt sich mit den Grundideen des räumlichen Hörens. Schwerpunkt ist die Lokalisation eines Klangereignisses und Raumwahrnehmung. Dabei werden einige essentielle Eigenschaften der auditiven Wahrnehmung behandelt. Von den grundlegenden Ideen der interauralen Zeit- und Pegeldifferenzen bis zu der komplexen Theorie der kopfbezogenen Außenohr-Übertragungsfunktionen, des Raumklangs und der Distanzwahrnehmung wird hier ein Überblick über die wichtigsten Höranker gegeben.

2.1.1 Interaurale Zeit- und Pegeldifferenz

Bereits im frühen 20. Jahrhundert wurde von Lord Raleigh[5] die unter dem Namen Duplex Theorie bekannte Beschreibung zur Lokalisierung eines Hörereignisses formuliert. Sie beschreibt Wahrnehmung eines Klangereignisses an den beiden Ohren unter der Voraussetzung, dass sich diese nicht an der selben Stelle am Kopf befinden. In der offensichtlichen Realität liegen diese etwa 17 cm auseinander. Dies führt dazu, dass Schallereignisse unterschiedliche Wege zu beiden Ohren zurück legen müssen. Je nach Ort des Schallereignisses relativ zum Kopf, können sich diese Wege in ihrer Länge mehr oder weniger deutlich unterscheiden. Die Differenz der zurückgelegten Wege führt zu zwei Effekten:

1. **Interaurale Zeit Differenz** (eng.: Interaural Time **D**ifference): Der Schall erreicht ein Ohr früher (ipsilateral)als das andere (kontralateral).
2. **Interaurale Pegel Differenz** (eng.: Interaural Level **D**ifference): Der Schall erreicht ein Ohr (ipsilateral) mit höherem Pegel als das andere (kontralateral)

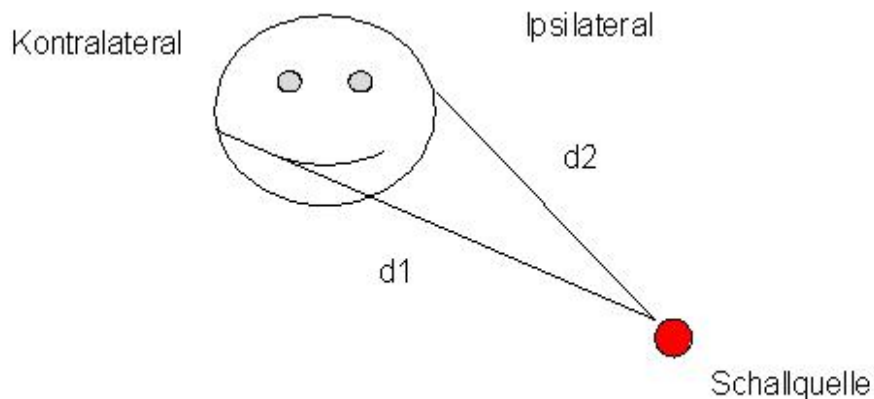


Abbildung 2: ITD und ILD

Man nennt diese beiden Effekte binaurale Höranker, da sie auf der Annahme von zwei nicht am selben Ort befindlichen Senken (Ohren) basieren. Nehmen wir also ein Schallereignis wahr, so führt ITD zu zeitverzögerter Wahrnehmung, den Phasenunterschieden, und ILD zu Lautstärkeunterschieden zwischen beiden Ohren. ITD hat vorherrschende Wirkung für Frequenzen unter 1500Hz, ein Bereich, in dem die Wellenlänge größer ist, als die 17 cm zwischen den beiden Ohren. Über 1500 Hz ist aus diesem Grund die Phasendifferenz zwischen beiden Ohren nicht mehr eindeutig. Umgekehrt für tiefe Frequenzen ist der Kopf klein im Gegensatz zur Wellenlänge, so dass er für sie kein Hindernis darstellt. Die Schallwellen beugen sich um den Kopf zum kontralateralen Ohr. Die ILD ist somit gering. In einem Frequenzbereich, indem die Wellenlänge kleiner als der Kopf ist, kommt es zu Abschattungen und somit zu Dämpfung. Die ILD wird größer. Die beiden Effekte ergänzen sich also. Nun hat die Duplex Theorie ihre Grenzen. Nehmen wir an, eine Schallquelle befindet sich in einem festen Abstand d_1 vom linken Ohr und einem festen Abstand d_2 vom rechten Ohr. Es ist nun möglich eine Menge von Punkten zu finden, so dass $d_1 - d_2 = const$ gilt. Alle diese Punkte befinden sich auf der Oberfläche eines Kegels. Die Differenz der beiden Wege bleibt für jeden von ihnen konstant, somit auch ITD und ILD. Als Sonderfall dieses Effekts stelle man sich ein Schallereignis direkt vor oder direkt hinter einem Hörer vor. Nach der Duplex Theorie ist nicht eindeutig bestimmbar, an welchem Ort sich das Schallereignis befindet. Man nennt dies den *Cone of Confusion* (siehe Abbildung 3). Eine Studien mit Menschen, die auf einem Ohr taub sind [7] zeigen, dass es ihnen möglich ist, eine

Schallquelle zu lokalisieren. Es gibt also weitere Höranker, um die Position einer Schallquelle zu bestimmen. Im Zusammenhang mit der ernsthaften Simulation einer dreidimensionalen Realität ist es unerlässlich auf einen der beiden Effekte zu verzichten-

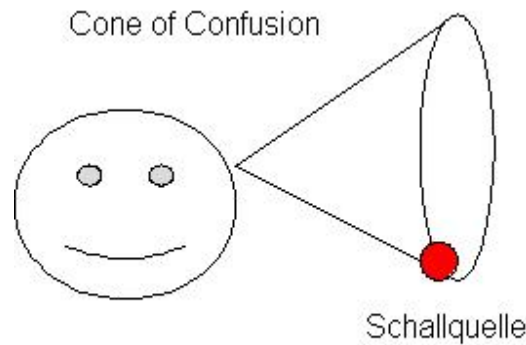


Abbildung 3: "Cone of Confusion"

2.1.2 Kopfbezogene Außenohr-Übertragungsfunktionen

Vorheriges Kapitel zeigte mit der Duplex Theorie ein Modell zur Beschreibung der Schallquellenlokalisierung mit seiner weitreichenden Bedeutung, aber auch Grenzen. Unter normalen Hörvoraussetzungen ist es jedem Hörer möglich den Ort einer Schallquelle relativ eindeutig zu identifizieren. Die Unvollständigkeit der Duplex Theorie liegt also nahe. In den späten 60'er Jahren wurde der Filtereffekt des Außenohrs genauer erforscht. Diese Arbeiten werden als eine Art Revolution für die Modelle der Schallquellenlokalisierung beschrieben [5]. Als kopfbezogene Außenohr-Übertragungsfunktionen (eng: *Head Related Transfer Functions*, kurz: *HRTF*) bezeichnet man den gemeinsamen Filtereffekt, den Kopf, Torso und im Speziellen die Pinna, also das Außenohr, auf die Frequenzverteilung eines einfallenden Schallsignals haben. Die typische Struktur der Ohrmuschel erzeugt Verzögerungen des einfallenden Schallsignals und verstärkt oder schwächt somit gewisse Frequenzbänder im Mitten- und Höhenbereich in charakteristischer Weise, abhängig von der Einfallsrichtung. Die entstehenden Verzögerungen bewegen sich im Bereich von 0-300 μs [5]. Der Filtereffekt ist ein monoauraler Höranker, das bedeutet, dass die Schallwahrnehmung an einem Ohr unabhängig vom Geschehen am zweiten Ohr entscheidend für eine Richtungswahrnehmung

ist. Eine mathematische Beschreibung der HRTF wird in der Literatur [5] als Verhältnis des frequenzabhängigen Schalldrucks am Trommelfell des jeweiligen Ohres (rechts oder links) zum richtungsunabhängigen Schalldruck am Mittelpunkt des Kopfes (natürlich ohne dass sich der Kopf an dieser Stelle befindet).

$$H_L = \frac{\Phi_L(\omega, \theta, \phi, d)}{\Phi_f(\omega)} \quad 2.1$$

$$H_R = \frac{\Phi_R(\omega, \theta, \phi, d)}{\Phi_f(\omega)} \quad 2.2$$

Dabei beschreibt ω die Abhängigkeit von der Winkelfrequenz, θ und Φ sind Azimuth- und Elevationswinkel der Schalleinfallrichtung und d die Entfernung der Schallquelle vom Hörer. HRTFs liefern dem Hörer Information um horizontale und vertikale Richtungen zu bestimmen und um Mehrdeutigkeiten der Hörereignisrichtung zu eliminieren. Bekannt ist, [5] dass bei ausgegossener Ohrmuschel (z.B. mit Wachs, der Gehörgang bleibt offen) der Grad der Verwechslung von vorderen und hinteren Hörereignisrichtungen im Vergleich zur offenen Pinna ansteigt. Andere Arbeiten[8] zeigen, dass dabei nicht nur der Filterungsprozess alleine von großer Wichtigkeit ist. Zusätzlich ist die spektrale Beschaffenheit der Schallquelle von entscheidender Bedeutung. Daraus zu schließen ist, dass für hohe Lokalisationsgenauigkeit ein breitbandiges Signal notwendig ist. Interessant im Zusammenhang mit der vorliegenden Arbeit ist, dass ein binaurales Wiedergabesystem wie das vorliegende sich in hohem Maße auf diese Theorie stützt und ein grundlegendes Verständnis ihrer Anwendung von elementarer Wichtigkeit ist.

2.1.3 Richtungsbestimmende Frequenzbänder

In der Medianebene, also die Ebene, die den Kopf in eine linke und eine rechte Hälfte teilt, sind binaurale Höranker einer einzeln auftretenden Schallquelle irrelevant, weil zu gleichen Zeiten das gleiche Signal an beiden Ohren eintrifft. Bekannt ist jedoch das Phänomen, dass bestimmte Signale zu räumlich erhöhter Wahrnehmung von Schallereignissen (Hörereignisrichtung) im Vergleich zur tatsächlichen Schallquellenposition (Schallereignisrichtung) führen. Blauert [8] hat diesen Effekt genauer untersucht, indem er eine Gruppe von Hörern mit Terzbandsignalen aus einer Lautsprecheranordnung, die garantierte, dass an beiden Ohren der gleiche Schalldruck auftritt, beschallte. Verschiedene Mittenfrequenzen des Terzbandrauschens führten zu verschiedenen Hörereignisrichtungen. Diese Feststellung gilt

unabhängig vom Filtereffekt des Torso und der Ohrmuschel: der Versuch wurde zusätzlich mit Messingröhrchen im Ohr durchgeführt, um den Effekt des Außenohrs auszuschalten. So wurde gezeigt, dass alleine die Mittenfrequenz eine Lokalisation in der Medianebene hervorruft. Abbildung 4 zeigt die Abhängigkeit der relativen Häufigkeit der wahrgenommenen Hörereignisrichtung von der Mittenfrequenz des Terzbandrauschens, wobei Diagramm (a) für hinten, (b) für oben und (c) für vorne wahrgenommene Ereignisse gilt.

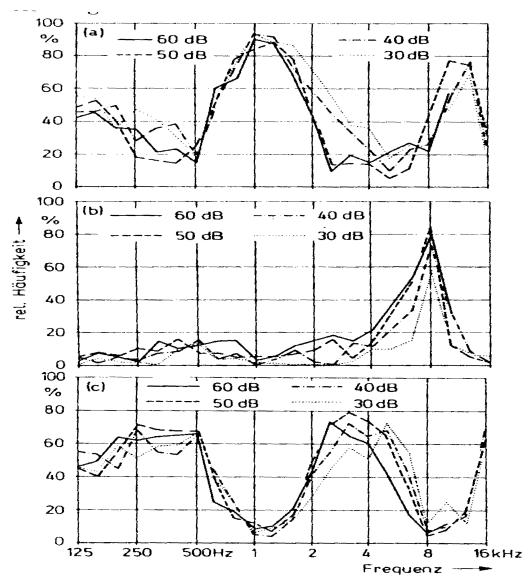


Abbildung 4: Richtungsbestimmende Bänder nach Blauert aus [8]

Dieser Effekt wird oftmals für Musik-, Film-, und Fernsehproduktionen ausgenutzt, um Klänge ihren Eigenschaften nach zu betonen, zum Beispiel präsenter erscheinen zu lassen. Durch spezielle Filterung der Signale in den entsprechenden Bändern wird dieser Effekt erreicht. Zusätzlich kann ein ganzes Klanggeschehen bezüglich einem Referenzklanggeschehen in bestimmter Weise frequenzverzerrt sein, so dass der Eindruck entsteht, beide Szenarien sind zwar identisch, spielen sich aber an verschiedenen Orten in der Medianebene ab. Ein interessanter Aspekt für die vorliegende Arbeit, wie sich im späteren Verlauf zeigen wird.

2.1.4 Raumklang

Viele Faktoren beeinflussen die Ausbreitung einer Schallwelle. Lässt man Eigenschaften unbetrachtet, die das Ausbreitungsmedium (Luftfeuchte, Temperatur, etc.) betreffen, so wird in einer typischen Hörsituation der Schall auf Hindernisse treffen. Sowohl bevor er beim

Hörer ankommt als auch nachher. Die auftretenden Effekte sind Absorption und Reflexion. Situationen in denen keiner dieser Effekte auftritt, lassen sich in der Natur nur auf dem Gipfel eines Berges, in tief verschneiter Landschaft, oder künstlich in einem schalltoten Raum annähern. In allen anderen Situationen wird der Schall auf dem Weg von der Quelle zum Hörer und darüber hinaus reflektiert und absorbiert. Beim Hörer kommt also nicht nur der direkte Schall an, sondern auch eine Anzahl von verzögerten und abgedämpften Versionen des Direktschalls aus unterschiedlichsten Richtungen. Alle diese verzögerten Versionen in Summe ergeben ein Klangbild, das man als Raumklang oder Nachhall bezeichnet. Der Raumklang kann als Funktion des Raumes gesehen werden, der ihn produziert. Er ist abhängig von der Geometrie und der Oberflächenbeschaffenheit des Raumes und nicht zuletzt von den Objekten und deren Beschaffenheit im Inneren des Raumes.

Schallquellen können also mehrfach reflektiert werden. Die Ordnung einer Reflexion zeigt an, wie oft die Welle bereits reflektiert wurde. Dabei verliert sie jedesmal einen Teil ihrer Energie durch Absorption, so dass die Ordnung eine endliche Zahl ist. Reflexionen niedriger Ordnung werden als frühe Reflexionen bezeichnet. Sie kommen direkt von den Wänden, der Decke und dem Boden. Typischer Weise werden reflektierte Schallwellen, die in den ersten 80 ms nach dem Schallereignis vom Hörer wahrgenommen werden, als frühe Reflexionen bezeichnet. Späte Reflexionen werden als aus allen Richtungen gleichmäßig kommend angenommen und auch als diffus bezeichnet (Abbildung 5). Ein wichtiger Beschreibungsparameter des Raumklanges ist die Nachhallzeit. Die Nachhallzeit definiert sich über die Zeit, die die Energie in einem eingeschwungenen Quelle-Raum-System nach Abschalten der Quelle benötigt um 60 dB abzufallen. Der Energieverlauf im Raum nach Abschalten der Schallquelle mit der Schallenergie E_0 kann wie folgt beschrieben werden:

$$E(t) = E_0 \cdot e^{\left[\frac{S \cdot c}{4V} \cdot \ln(1 - \bar{\alpha}) - \alpha_L \cdot c \right] \cdot t} \quad 2.3$$

Wobei E_0 die Energie ist, mit der der Raum im eingeschwungenem Zustand angeregt wird, so dass gilt: $E_{\text{zugeführt}} = E_{\text{absorbiert}}$

S , V und $\bar{\alpha}$ sind die Raumparameter, also Gesamtoberfläche, Volumen und mittlerer Absorptionskoeffizient. c bezeichnet die Schallgeschwindigkeit und α_L die Dämpfungskonstante der Luft.

Für den Zeitpunkt der Nachhallzeit gilt

$$\frac{E(t)_{(t=T_N)}}{E_0} = 10^{-6} = e^{\left[\frac{S \cdot c}{4V} \cdot \ln(1-\bar{\alpha}) - \alpha_L \cdot c\right] \cdot T_N} \quad 2.4$$

Dabei wurde die Absorbtion der Luft nicht vernachlässigt. In der Praxis kann sie bei bekannten Materialdaten durch eine Formel angenähert werden (Absorbtion der Luft diesmal vernachlässigt):

$$T_{60} \approx 0,161 \cdot \frac{V}{-A \cdot \ln(1-\bar{\alpha})} \cdot \left[\frac{s}{m} \right] \quad (\text{Eyring}) \quad 2.5$$

Die Nachhallzeit ist also stark abhängig vom Raum und den Materialien, aus denen die Wände, Decke und Boden sind. Auch der Inhalt des Raumes ist relevant. So absorbiert ein mit Zuschauern gefüllter Konzertsaal wesentlich mehr Schallenergie als der selbe Raum, wenn er leer ist. Es existieren verschiedene Arten der Messung für die Nachhallzeit. Eine oft benutzte Möglichkeit für die Messung der Nachhallzeit ist eine Pseudozufall-Sequenz mit geeigneter Länge um die Raumimpulsantwort zu messen. Aus der so erhaltenen Impulsantwort wird mittels Schröder-Rückwärts-Integration die Nachhallzeit berechnet. Die frequenzabhängige Nachhallzeit wird aus Effizienzgründen meist über die Abschätzung des Energieabfalles um 30dB bzw 20dB heraus gemessen. Anschließend wird linear interpoliert um den Zeitwert für 60 dB Abfall zu erhalten. Üblicherweise wird für Räume die Nachhallzeit von 500 Hz angegeben und als Vergleichswert benutzt. Andere Methoden benutzen die EDT (engl.: **E**arly **D**ecay **T**ime), sie beschreibt die Zeit für einen Abfall der Schallenergie um 10 dB nach dem Abschalten der Quelle eines eingeschwungenen Quelle-Raumsystems. Die EDT ist eine genauere Schätzung der Nachhallzeit.

Der Hallradius hängt direkt mit der Nachhallzeit zusammen und beschreibt die Entfernung zwischen Schallquelle und Senke (Mikrofon, Hörer), bei der der direkte Schalleinfall am Ort der Senke die gleiche Schallintensität hat wie der Nachhall. Für die Berechnung des Hallradius sind die Abmessungen des Raumes nötig. Es ist nicht möglich ihn aus gemessenen Impulsantworten zu extrahieren. Er kann über folgende Formel abgeschätzt werden:

$$d_{reverb} = 0.25 \cdot \frac{\alpha \cdot S}{\pi} \quad 2.6$$

$$d_{reverb} = 0.006 \cdot \sqrt{\frac{V}{T_{60}}}$$

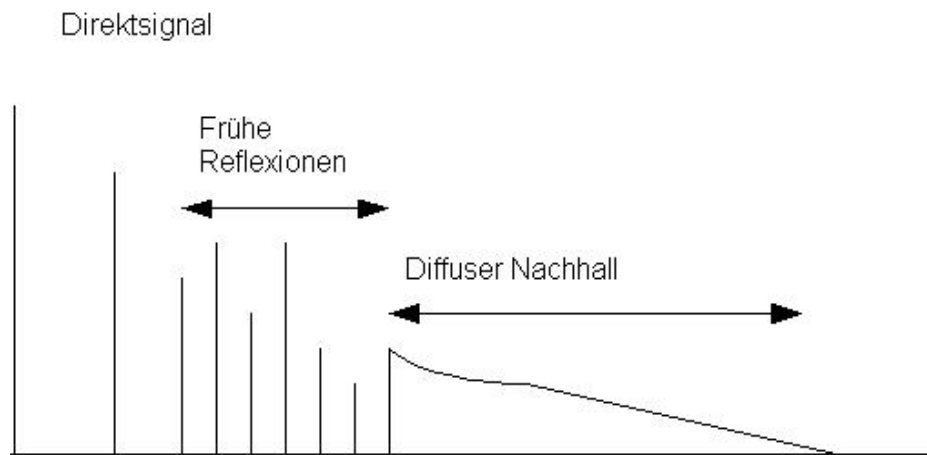


Abbildung 5: Nachhall Diagramm

Die Betrachtungen des Charakteristikums Nachhallzeit eines Raumes ist für die Arbeit von Wichtigkeit, da für die binaurale Aufbereitung der Signale im vorliegenden Blind Upmix Binaural Wiedergabesystem nicht nur die in vorhergehenden Kapiteln besprochenen Höranker benutzt werden, sondern auch die Verarbeitung von Rauminformation berücksichtigt wird.

2.1.5 Gesetz der ersten Wellenfront

Wie in vorhergehenden Kapiteln gezeigt, besteht eine typische Hörsituation also darin, einerseits das direkte Schallereignis über den kürzesten Weg zwischen Quelle und Senke wahrzunehmen, andererseits die reflektierten, abgedämpften und verzögerten Versionen des Originalsignals aus allen möglichen Richtungen als Raumklang wahrzunehmen. Dennoch ist es dem Gehör möglich genau den Ort der Schallquelle zu bestimmen. Das Gehör ist in der Lage all die Reflektionen mit dem Direktschall zu einer Einheit zu kombinieren und genau der Richtung zu zuordnen, woher der direkte Anteil kommt. Dies ist auch möglich, wenn die Schallintensität des Raumanteils höher ist als die des Direktanteils. Diese Fähigkeit wird als das Gesetz der ersten Wellenfront (oder auch Precedence Effect, Haas Effekt) bezeichnet und wurde von Wallach 1949 aufgestellt [5]. Zwicker und Fastl [9] beschreiben den Versuch mit Lautsprechern einer Stereoaufstellung: Zwei identische Signale werden durch linken und um $>1\text{ms}$ verzögert durch den rechten Lautsprecher geschickt. Selbst bei Verstärkung des rechten (verzögerten) Lautsprechersignals gegenüber dem linken wird das Schallereignis aus dem rechten Lautsprecher kommend wahrgenommen. Der Effekt ist besonders ausgeprägt bei

transienten Signalen. Ab einer Verzögerung des rechten Signals über etwa 50 ms (abhängig vom Signal) wird es als störendes Echo wahrgenommen. Die Zeitpunkt, ab dem das Signal als Echo wahrgenommen wird hängt sehr stark vom Signalinhalt ab.

2.1.6 Distanzwahrnehmung

Bisher wurden Höranker gezeigt, die das Gehör zur Schallquellenlokalisierung auf einer gedachten Kugel mit festem Radius um den Kopf herum benutzt. Befindet man sich jedoch im Fernfeld einer Schallquelle, ist die Entfernung zum Hörer also größer als zirka 1 m, so geben diese Höranker vorwiegend Information über die Richtung, aus der das Schallereignis kommt. Unberücksichtigt bleibt die Entfernung zwischen Schallquelle und Senke. In der Realität ist es Menschen aber durchaus möglich Entfernungen von Objekten anhand von Geräuschen fest zu stellen. Ein Beispiel sei gegeben durch blinde Menschen. Sie senden Signale aus (z.B. Hiss-Laute oder das Klicken des Blindenstockes) und anhand der Reflexionen von Gegenständen können sie Entfernungen zu diesen einschätzen. Diese Fähigkeit wird Echolokation genannt und bekanntlich auch von Tieren (Fledermaus, Delphin) zur Orientierung benutzt. Auch nicht sehbeeinträchtigte Menschen haben diese Fähigkeit [5].

Welche Höranker sind also entscheidend für die Distanzwahrnehmung? In der Literatur sind die folgenden fünf als die wichtigsten Höranker zu finden[10]:

- Schallintensität
- Direkt- zu Raumklang Verhältnis
- Frequenzspektrum der Schallquelle
- Binaurale Differenzen (ILD,ITD)
- Art des Stimulus

Eine Hierarchie dieser Höranker kann nur schwer angegeben werden, da sie von der Hörsituation abhängig ist. Man geht davon aus, dass Intensität und Raumklang die wichtigsten Parameter sind. Dennoch kann die Vertrautheit mit einem Geräusch oder einem Raum für den Hörer der wichtigste Anker sein. Zusätzlich muss eine Änderung eines dieser Parameter nicht notwendigerweise auch zu einer Änderung des Distanzeindrucks führen. Zu groß ist die Abhängigkeit untereinander. Somit scheint es wenig zielführend zu sein nur einen Parameter zu verändern [5]. Die Berücksichtigung aller dieser wichtigen Anker führt zum Ziel um eine bestimmte Distanzwahrnehmung zu erzeugen. Die Anker können nach zwei Konzepten kategorisiert werden: relative Anker und absolute Anker. Erstere beurteilen den

Unterschied eines Klangs bezüglich eines anderen Klangs. Letztere beurteilen einen Klang absolut, ohne Referenzklang.

Die Schallintensität im Fernfeld fällt nach dem $1/r^2$ Gesetz. Der Abfall kann mit der Formel

$$L_{Abfall} = 20 \times \log_{10} \left(\frac{r_d}{r_0} \right) \quad 2.7$$

angenähert werden. r_d bezeichnet hierbei die Entfernung zur Quelle, r_0 die Referenzentfernung. Jede Verdoppelung der Entfernung zieht also einen Pegelabfall von 6 dB nach sich. Somit kann Lautstärke als relativer Höranker bezeichnet werden. Ein Versuch dahingehend funktioniert nur unter bestimmten Voraussetzungen. So wird dabei nur der direkte Weg zwischen Quelle und Senke betrachtet und jeglicher Raumklang vernachlässigt. Gewöhnliche Hörsituationen finden nie ohne Raumklang und Reflexionen statt. Raumklang ist ein absoluter Höranker. Der elementarste Anker ist das Direkt- zu Raumklang Verhältnis zur Schätzung der Entfernung. Die Größe des Verhältnisses

$$10 \cdot \log_{10} \left(\frac{\int_0^a p^2(t) \cdot dt}{\int_a^\infty p^2(t) \cdot dt} \right) \quad 2.8$$

bei einer Raumimpulsantwort gemessen für verschiedene Entfernungen zwischen Schallquelle und Mikrofon ist umgekehrt proportional zur dieser Entfernung. Die Werte für a in Arbeiten, die diesen Effekt untersuchen liegen im Bereich zwischen 5 und 15 ms [10]. Aus der Konzertsaal-Akustik kennt man dieses Verhältnis aus der Analyse der Raumimpulsantwort mit den Integrationsgrenzen 0 und 50 ms als das Deutlichkeitsmaß und mit den Intergrationsgrenzen 0 und 80 ms als Klarheitsmaß. Beide dienen als objektive Einflussgrößen zur Beschreibung eines Raumes.

Anders verhält es sich mit dem Frequenzspektrum der Schallquelle. Die Entfernung zur Quelle kann nur geschätzt werden, wenn ein Referenzklang vorhanden ist oder der Klang dem Hörer sehr gut bekannt ist. Durch Erhöhung der Entfernung treten vermehrt Dämpfungseffekte im Übertragungsmedium (Luft) auf. Natürlich wird das Frequenzspektrum auch von Raumklanganteilen beeinflusst und ist somit nicht unabhängig. Binaurale Anker sind hauptsächlich dann interessant, wenn sich die Schallquelle im Nahfeld (zirka $d < 1m$) befindet. Einfallende Schallwellen können nicht mehr als plane Wellenfront bezeichnet werden. Die gekrümmte Natur der Schallwellen hat Einfluss auf ILD. Je geringer die

Entfernung zwischen Hörer und Schallquelle ist, umso größer ist der Effekt der Abschattung durch den Kopf. Die ILD steigt, ITD bleibt weitgehend unberührt davon. Abbildung 6 aus [5] verdeutlicht diesen Zusammenhang. Gezeigt wird die ILD für 3 verschiedene Einfallrichtungen (90° , 45° , 15°) als Funktion der Entfernung (0,12 m - 1 m) zwischen Kopf und Schallquelle.

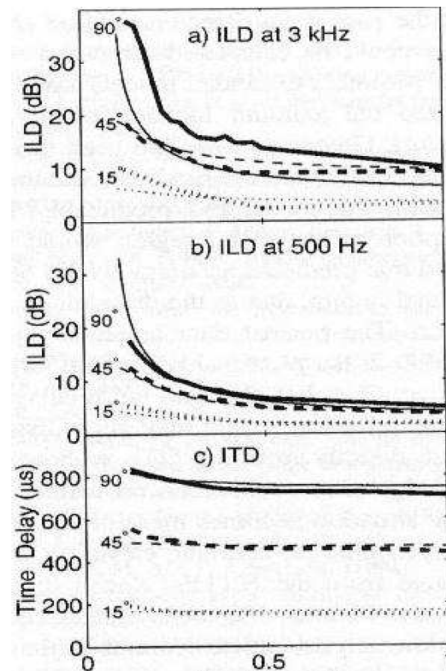


Abbildung 6: ILD Abhängig von Entfernung im Nahfeld [5]

Im Hinblick auf ein binaurales Wiedergabesystem ist das Thema Distanzwahrnehmung ein höchst interessantes. Einer der am angenehmsten empfundenen Effekte einer binauralen Wiedergabe ist die so genannte Außer-Kopf-Lokalisation. Er entsteht durch die Modellierung aller Höranker, die in den Messungen der binauralen Raumimpulsantworten (BRIA) enthalten sind. Die zusätzliche Information über den Raum in der gemessenen BRIA lässt sich dazu benutzen, Entfernung in einem gewissen Bereich einzustellen.

2.1.7 Cocktailpartyeffekt

Der Cocktailpartyeffekt beschreibt die Fähigkeit des menschlichen Gehörs aus einem Klanggemisch bestimmte Aspekte zu extrahieren und subjektiv hörbarer zu machen. Er wurde 1953 von Cherry entdeckt[8]. Zum Beispiel ist es Menschen möglich ein Gespräch aus einem Stimmengewirr gezielt zu verfolgen, es aus dem Umgebungsgeräusch zu extrahieren, ohne den Kopf drehen zu müssen, oder visuellen Kontakt herzustellen. Der Cocktailpartyeffekt ist

ein binauraler Höranker. Bei einseitigem Abdecken eines Ohres ist der Mensch eingeschränkt diese Fähigkeit der neuronalen Signalverarbeitung zu benutzen. Grundsätzlich gilt, dass ein Nutzsinal von einem Störsignal weniger verdeckt wird, wenn man mit beiden Ohren hören kann. Zudem verfügen die meisten hörgeschädigten Menschen bei Einzelschallquellen (die mit entsprechender Lautstärke wiedergegeben werden) über gute Sprachverständlichkeit, besitzen aber im Stimmengewirr große Probleme sich auf einzelne Gespräche zu konzentrieren. Zurück geführt werden kann dies einerseits auf weniger hohe Zeitauflösung bei der neuronalen Verarbeitung von Hörbeeinträchtigten [9]. Das heißt, dass der aus der Psychoakustik bekannte Maskierungseffekt bei Hörgeschädigten größere zeitlich Ausdehnung besitzt. Unterschiede zwischen Hintergrundgeräusch und interessantem Gespräch können nicht mehr erkannt werden. Diesbezügliche Versuche können bei Zwicker/Fastl [9] nachgelesen werden. Andererseits verringert sich die Fähigkeit Klanggeschehen nach Frequenzen zu separieren. Die sogenannten auditorischen Filter verbreitern sich [11] bei Hörgeschädigten, wie Abbildung 7 zeigt. Das bedeutet, dass auch Geräusche, die in Frequenzbereichen außerhalb der kritischen Bandbreite des Nutzsignals liegen, nur geringen Pegel benötigen, um das Nutzsinal zu verdecken.

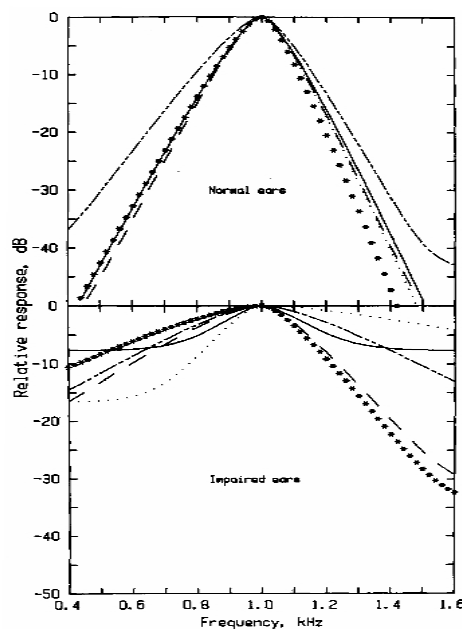


Abbildung 7: Verbreiterung der auditorischen Filter bei Hörgeschädigten aus [11]

2.2 Produktevaluierung

Das Kapitel bietet einen Einstieg in die Bewertung von Systemen zur Audiowiedergabe. Unterschieden wird dabei zwischen objektiven und subjektiven Methoden. Subjektive Methoden werden genauer besprochen, aus der Motivation heraus eine Evaluierungsprozedur für Änderungen am System zu erstellen.

2.2.1 Methoden zur Evaluierung von Audiosystemen

Für den Entwicklungsingenieur ist es wichtig eine verlässliche Methode zur Beurteilung seiner Arbeit zu haben, da Audioprodukte einer täglichen Güteprüfung durch das menschliche Ohr ausgesetzt sind. In einigen Bereichen der Produktevaluierung gibt es objektive Modelle (z.B.[12] PEAQ), die die Bewertung eines Systems erlauben, ohne auf menschliche Testhörer zurückgreifen zu müssen und ein verlässliches Maß der Güte eines Systems liefern. Die Modelle jedoch sind begrenzt in ihrer Anwendbarkeit. So wird der benannte PEAQ Standard dazu herangezogen, auf der Basis eines psychoakustischen Modells die Güte von Audiocodierung anhand der spektralen Verteilung von Verzerrungen im Signal zu beurteilen. Wenn es um objektive Beurteilung räumlicher Eigenschaften geht, stellt sich gleich zu Beginn das Problem, dass psychoakustische Modelle bei weitem nicht hoch genug entwickelt sind, um die Wahrnehmung von Raum eindeutig mit Signaleigenschaften und in Folge Wahrnehmungsgüte zu korrelieren. Daraus resultiert die Notwendigkeit subjektiver Testmethoden. Zum Einen, um Zusammenhänge dahingehend zu etablieren und Modelle weiter zu entwickeln, zum Anderen um Systeme, die für Konsumenten zum täglichen Gebrauch gedacht sind, zu evaluieren und weiter zu entwickeln. Etabliert haben sich Standardverfahren [13],[14],[15](ohne Anspruch auf Vollständigkeit), die auf der Basis eines mittleren Meinungsquerschnitts Aussagen über die Güte einer Signalbearbeitung machen. Hierbei wird die Gesamtqualität eines oder mehrerer Stimuli von erfahrenen Hörern beurteilt, optional im Vergleich zu einer Referenz, in jedem Fall aber untereinander. Meist werden sie benutzt, um die Qualität von Audiocodierung zu testen, wobei je nach Aufwand für die Codierung die Qualitätsunterschiede differieren und somit die Testmethoden. Sie gehören zur Gruppe der affektbetonten Bewertungsmethoden im Gegensatz zu den wahrnehmungs-
betonten Bewertungsmethoden [16]. Bei den affektbetonten Bewertungsmethoden kann unterschieden werden zwischen Präferenz-, Akzeptanz- und Angebrachtheitstest, wobei die bereits erwähnten Standardverfahren zu den letzteren zu zählen sind. Allen drei Arten von

Tests ist gemein, dass sie je nach Fragestellung eine Gesamtbewertung des Hörers generieren, aber keine detaillierte Information über das Entscheidungskriterium, das der Hörer für sich hatte, liefern. Dies ist einerseits sehr hilfreich, um zu sehen das ein System gut oder schlecht akzeptiert ist, andererseits jedoch erhält man keine Information über das Warum. Im Falle der Bewertung von Audiocodierung ist diese Information eventuell ausreichend, da die physikalische Variable (Bitrate) direkt mit der Wahrnehmung von Güte zusammenhängt. Für die Entwicklung eines räumlichen Wiedergabesystems ist diese Information nicht ausreichend, da viele Variablen Einfluss haben, die je nach Ausprägung eine Präferenz, oder Akzeptanzbewertung beeinflussen können. Es liegt also ein multidimensionales Problem vor von hohem Informationspotential für die Weiterentwicklung eines Systems.

Bei wahrnehmungsbetonten Bewertungsmethoden wird bei der Befragung der Hörtestteilnehmer genaue analytische Information über die Wahrnehmung gesammelt. Dies kann verbal oder nonverbal geschehen, wobei die Methoden jeweils ihre Vor- und Nachteile haben. Eine ausführliche Diskussion findet man bei Mason in [17]. Wegen ihrer statistischen Auswertbarkeit werden verbale Methoden den nonverbalen meist vorgezogen.

2.2.2 Subjektive Qualitätsbewertung

In den letzten Kapiteln wurden grundlegende Modelle zur Beschreibung räumlicher Hörerfahrung vorgestellt. Um die tatsächlichen Auswirkungen eines Wiedergabesystems auf die Wahrnehmung feststellen und bewerten zu können, ist eine geeignete Testprozedur notwendig. Wie bereits angemerkt gibt es derzeit keine aussagekräftigen objektiven Methoden um räumliche Attribute von Wiedergabesystemen zu kategorisieren und zu beurteilen. Forschung in diese Richtung wird betrieben, Ergebnisse sind aber nicht weit genug fortgeschritten um auf ihrer Grundlage Voraussagen über die Akzeptanz eines Systems zu machen [2]. Dieses Kapitel beschäftigt sich detaillierter mit der Spezialform der affektbetonten Bewertung, dem Präferenztest, und den Grundzügen der wahrnehmungsbetonten verbalen Bewertung. In der Literatur wird eine Kombination aus diesen Tests als wirkungsvolle Methode zur Produktevaluierung betrachtet [18], [2], [19].

2.2.2.1 Präferenztest

Präferenztests werden auf 2 Arten durchgeführt: Entweder als Paarvergleich mit zwei Antwortmöglichkeiten, oder mit einer größeren Anzahl an zu vergleichenden Stimuli (Multi

Stimuli) als Rangfolgetest (*Indirect Scaling*). Es wird dabei abgefragt, in welcher Reihenfolge die Stimuli präferiert würden. Vorteil ist, dass er einfach und schnell durchgeführt werden kann. Eine exzessive Trainingsphase für die Hörer muss nicht stattfinden. Dabei wird eine Ordinalskala benutzt, die nicht vorsieht, dass Stimuli äquivalent sein können. In Arbeiten der jüngeren Vergangenheit [2],[20] wird meist eine kontinuierliche Verhältnisskala benutzt, auf der Hörer ihre Wahrnehmung direkt repräsentieren können (*Direct Scaling*). Die Möglichkeiten die Auswertungen mit bekannten statistischen Methoden ist dabei in höherem Maße gegeben als bei einem Rangfolgetest. Dieser Test jedoch gibt keine Information über das Entscheidungskriterium des Hörers. Vorteil eines solchen Test ist, dass es allgemein als nicht notwendig betrachtet wird geübte Hörer zu benutzen, um eine Aussage zu gewinnen [16]¹.

2.2.2.2 Wahrnehmungsbetonte Tests

Wahrnehmungsbetonte Tests geben die Antwort, die affektbetonte Tests schuldig bleiben, sind jedoch weitaus aufwendiger und schwieriger durchzuführen. Im Falle der verbalen Bewertung kann der Prozess in mehrere Phasen gegliedert werden: Um Wahrnehmungen zwischen den verschiedenen Individuen anzugleichen muss sicher gestellt werden, dass jede Person im Test den selben Wortschatz benutzt um ihre Wahrnehmungen zu beschreiben, bevor in der zweiten Phase die Bewertung der Stimuli hinsichtlich dieser gefundenen Attribute stattfinden kann. Dieser Teil des Tests wird allgemein für notwendig befunden, da auf diesem Gebiet nicht von objektivem Wissen gesprochen werden kann, vielmehr existiert unter den unterschiedlichen Experten nicht unbedingt ein Konsens in den Konzepten und den dafür benutzten verbalen Beschreibungen räumlicher Wahrnehmung [21], so dass durch eine geeignete Prozedur Einklang geschaffen werden muss (Abbildung 8).

¹ Kapitel 5.4 Seite 112

		Terminology	
		Same	Different
Attributes	Same	Consensus Experts use terminology and concepts in the same way	Correspondence Experts use different terminology for the same concepts
	Different	Conflict Experts use same terminology for different concepts	Contrast Experts differ in terminology and concepts

Abbildung 8: Konzepte und verbale Beschreibung aus[21]

Alleine dieser Prozess nimmt einen großen Teil des Tests in Anspruch. Bech [16] beschreibt verschiedene Methoden, wie zum Beispiel *Quantitative Descriptive Analysis* oder *Flash Profile* die zu diesen Attributen führen. In Veröffentlichungen wissenschaftlicher Arbeiten von Berg und Rumsey, Lorho, Zacharov finden sich auch dokumentierte Anwendungen von diesen Methoden, zum Beispiel die *Repetory Grid Technique* [21] oder *Descriptive Analysis* [2]. Allen beschriebenen Verfahren ist gemein, dass sie auf der Basis von gegebenen Stimuli ein Set an beschreibenden Attributen finden. Dieses Set wird auf spezifische Weise mit Hilfe klassischer Statistik auf eindeutige Beschreibungen reduziert, um zu einem endgültigen und für alle Testteilnehmer gültigen Set zu gelangen. Innerhalb der Prozeduren werden auch die Skalen mitdefiniert, was ein entscheidender Vorteil für die Testhörer ist. Den Testhörern ist es nun möglich zwischen den Stimuli in konsistenter Weise bezüglich den konstatierten Eigenschaften auf den Skalen zu unterscheiden. Es gibt eine Einschränkung, die noch nicht ausgeräumt werden kann: Es ist nicht unbedingt möglich, die erhobenen Attribute automatisch für ein anderes Set an Stimuli zu verwenden als für die, die im Attributerhebungsprozess integriert waren. Zwar gibt Arbeiten[22], die Attribute räumlicher Wiedergabe auf verschiedene Stimuli mit verschiedenen Testhörern erfolgreich benutzen, jedoch andererseits auch Arbeiten in denen in Betracht gezogen wird, dass hohe Korrelation zwischen den Attributen genau aus diesem Grund resultiert [2].

Die Voraussetzung für eine Testprozedur dieser Art ist ein Hörerplenum, das sich aus Hörexperten zusammen setzt und während der Prozedur keine personellen Veränderungen erfährt. Die Definition von Hörexperten ist vielfältig über verschiedene Standards, jedoch kann eine allgemeine Definition von Kategorien von Teilnehmern an wahrnehmungsbetonten

Tests bei ISO 8586-2 (dem Autor leider nur als Tabelle in [16]² vorliegend) nachgeschlagen werden. Die Voraussetzung für Hörexperten begründet sich aus dem Testdesign: Naive Hörer könnten Eigenschaften eines Systems nicht feststellen und somit könnte wichtige Information verloren gehen. Des Weiteren lässt sich statistisch argumentieren: Bech [23] und Olive[24] zeigen in ihren Arbeiten, die wesentlich höhere Signifikanz von Testergebnissen, die von Experten geliefert werden. Dies impliziert auch, dass ein Test mit wesentlich weniger Testpersonen auskommt. Bech ([16] und [23]) nennt die Zahl 7 als Reduktionsfaktor, beziehungsweise eine Zahl von 5-15 Hörern als ausreichend für Tests mit Hörexperten.

2.3 Surroundverfahren

2.3.1 Historisches

Klangerlebnisse ähnlich einem Konzert auch zuhause zu erleben, erfordert die Abbildung räumlicher Komponenten einer live erlebten Wiedergabe. Ein erster Schritt in diese Richtung ist das Stereoformat mit 2 Kanälen. Stereo in seiner wörtlichen Übersetzung (griech.: fest, räumlich) bedeutet in keinsten Weise eine Limitierung auf 2 Kanäle. Vielmehr ist dies eine Auswirkung des technischen Entwicklungsstandes der Zeit, in der die Einführung von Stereo passierte (1958)[5]. Das hauptsächlich für Konsumenten verfügbare Wiedergabeformat damals war die Schallplatte. Auf ihr ist es physikalisch nur möglich 2 Kanäle zu speichern. Im Vergleich zum damaligen Standard Mono (1 Kanal) eine unglaubliche Erweiterung aber nicht das Ende der Skala. Neue Wiedergabemedien und Speichertechniken, die nun seit einigen Jahren verfügbar und vor allem bezahlbar sind, ermöglichen einen erneuten Fortschritt und Erweiterung von Hörqualität. Ein Mehrkanalsystem kann also aus einer nicht näher festgelegten Anzahl an Wiedergabekanälen bestehen, mit dem Ziel eine möglichst realistische Darbietung zu erreichen. Der Wunsch möglichst authentische Wiedergabeszenarien zu produzieren kommt zum Einen für Reproduktion von Konzerten zum Anderen aus der Filmbranche. Beide Ansätze sind eng miteinander verknüpft, wenngleich sie manchmal unterschiedliche Ansätze verfolgen. Eines der ersten Systeme für eine mehrkanalige Wiedergabe, das „Wall of Sound“[5], wurde von der Fletcher Gruppe an den Bell Laboratorys entwickelt. Dabei wurde ein Orchester mit 80 horizontal vor dem Orchester angebrachten Mikrofonen aufgenommen und über ebenso viele Boxen wiedergegeben. Das Resultat war eine große Hörfläche (Sweetspot) und eine originalgetreue Wiedergabe des Konzertes. Nun ist in der Konsumentenrealität so ein System unpraktisch. Man reduzierte das Setup auf 3 Kanäle. Noch heute ist dies gängige Praxis bei der Reproduktion des Frontbildes in kommerziellen Mehrkanalsystemen. Mit drei Kanälen ist eine Positionierung der Phantomschallquellen einfacher und zufriedenstellender möglich als mit zwei. Die erste öffentliche Demonstration eines Mehrkanal Systems kam von Walt Disney mit der Präsentation seines Filmes *Fantasia* 1940[5]. Man benutzte 8 Wiedergabekanäle für Musik und Klangeffekte. Keines der damaligen Kinos war mit einem Wiedergabesystem für diesen Film ausgestattet, und so ging man mit eigener Technik auf Tour damit.

Weiter verbreitete Systeme für Konsumenten kamen Ende der 60' er und sind zum Teil bis heute aktuell: Quadraphonic, Ambisonic, diverse Dolbyformate, DTS.

Quadraphonic benutzt vier Kanäle zur Wiedergabe, wobei es keinen Standard für die Lausprecheraufstellung gibt. Aufgrund fehlender Akzeptanz für das neue Format, niemand wollte sich neue und vor allem zusätzliche Wiedergabegeräte kaufen, entwickelte man die Methode der Matrizierung: Die 4-Kanalversionen wurden durch Superposition auf zwei Kanäle gemischt. Aus den so gewonnenen 2-Kanalinhalten wurde durch inverse Matrizierung der Kanäle wieder ein 4-Kanalsignal erzeugt. Die Kompatibilität auf vorhandene Wiedergabegeräte und Medien wurde somit gewährleistet. Das Verfahren wurde bekannt unter dem Namen Scheibermatrix (Abbildung 9 aus [5]). Auch heute noch gibt es Systeme, die diese Techniken in abgewandelter Form benutzen und entscheidend zum Erfolg und zur Akzeptanz von heutigen Surroundsystemen beigetragen haben.

$$\begin{aligned}L_{Stereo} &= 0,924 \cdot L_F + 0,924 \cdot L_R + 0,383 \cdot R_F + 0,383 \cdot R_R \\R_{Stereo} &= 0,924 \cdot R_F + 0,924 \cdot R_R + 0,383 \cdot L_F + 0,383 \cdot L_R\end{aligned}$$

$$\begin{aligned}L_F &= 0,924 \cdot L_{Stereo} + 0,383 \cdot R_{Stereo} \\R_F &= 0,383 \cdot L_{Stereo} + 0,924 \cdot R_{Stereo} \\L_R &= 0,924 \cdot L_{Stereo} - 0,383 \cdot R_{Stereo} \\R_R &= 0,383 \cdot L_{Stereo} - 0,924 \cdot R_{Stereo}\end{aligned}$$

Abbildung 9: Scheibermatrix

$L_{F,R}$ =Links vorne/hinten

$R_{F/R}$ =Rechts vorne/hinten

In den 70'er Jahren stellte *Dolby* ein neues Filmtoneformat vor (*Dolby Stereo*), welches die Übertragung des Filmtones auf optischem Wege ermöglichte[4]. Die nur 2 Tonspuren wurden zusätzlich auf dem 35 mm Filmcelluloid untergebracht. In den meisten Filmtheatern wurden jedoch aufgrund ihrer Größe meist mehr als 2 Kanäle benutzt, zumal auch die Filme üblicherweise mit vier Kanälen produziert wurden (Links, Mitte, Rechts, Surround). Somit bediente man sich bei den Ideen der Quadrophonie und benutzte seine Matrixtechnik. Der Filmtone war also ein Downmix des original Mehrkanaltones, und aus diesem wurde beim Abspielen mittels Matrixtechnik wieder ein 4 Kanal Signal erstellt. Später, 1982, wurde dann *Dolby Surround* als Heimkino Variante von *Dolby Stereo* auf den Markt gebracht [4]. Der Erfolg blieb nicht aus, da das System rückwärtskompatibel blieb, also auch mit Stereo Equipment abgespielt werden konnte, und das bereits existierende Tonmaterial einfach auf die

Konsumentensysteme übertragen werden konnte. Viele Varianten von solchen Matrix Surround Systemen haben sich etabliert und auch weiterentwickelt, so dass sie dynamisch auf Signalinhalte reagieren können und somit weit mehr Möglichkeiten und Qualität hervorbringen. Neueste etablierte Technik nennt sich *Spatial Audio Coding* und wird zum Beispiel für MPEG-Surround verwendet[4]. Auch hier kann aus einem Mehrkanaldatenstrom eine parametrische Beschreibung gewonnen werden und als Metadaten zusätzlich zu einem Downmix übertragen werden, aus dem dann wieder ein Mehrkanalsignal gewonnen werden kann[25]. Die parametrische Beschreibung besteht aus Intensitätsdifferenzen, Kohärenz- und Korrelationswerten zwischen den Kanälen. Ein Vorteil ist, dass keine Beschränkung auf Stereodownmixe gegeben ist, es kann auch zum Beispiel aus einem Monodownmix mittels Metadaten ein Mehrkanalsignal erzeugt werden.

2.3.2 Blind Upmix

Im letzten Kapitel wurden Surroundformate behandelt, die als Mehrkanalsignal mit M Kanälen produziert werden und dann nach einer Vorschrift auf N Kanäle ($N < M$) konvertiert (Downmix) werden, um zum Beispiel erforderliche Kanalkapazitäten bei der Übertragung gering zu halten und zusätzlich Rückwärtskompatibilität, also auch Abspielbarkeit mit nur N Kanälen, zu gewährleisten. Bei der M Kanalwiedergabe wird das N Kanalsignal wieder rückkonvertiert (Upmix) auf M Kanäle mittels Dekodierung der Downmixvorschrift. Diese Art des Upmix wird in der Literatur oft parametrischer Upmix genannt. Im Gegensatz, oder als Erweiterung, dazu steht der so genannte der Blind Upmix. Mit einem Blind Upmix ist es möglich ein N Kanalstück in ein M Kanalstück zu konvertieren (Upmix), wobei $N < M$ gilt, ohne spezielles Vorwissen über die N-Kanal Version und ohne zusätzliche Metadaten. Verschiedene Möglichkeiten existieren zur Verwirklichung eines solchen Upmixsystems. Durch die Integration aktiver Steuerelemente in ein System mit Matrixtechnik kann auch aus diesem ein Blind Upmix fähiges System gemacht werden. Ein Beispiel dafür ist Dolby Pro Logic II. Ein anderer Ansatz basiert auf eine Idee von Avendano und Jot [26] und wird im Späteren genauer beschrieben. Dadurch ist die Möglichkeit gegeben, aus jeder Art von 2-Kanal Ausgangsmaterial eine, in diesem speziellen Fall, 5-Kanalversion herzustellen und die Qualität des Hörerlebnisses entscheidend zu erweitern. Die Möglichkeiten, die durch diesen Ansatz gegeben sind lassen sich einfach in das Konzept des *Spatial Audio Coding* integrieren.

2.4 Binauraltechnik

Binauraltechnik versucht eine akustische Realität herzustellen, indem die Effekte des räumlichen Hörens ausgenutzt werden. Den wichtigsten Höranker stellt dabei die Verwendung der kopfbezogenen Außenohr-Übertragungsfunktionen (eng.:HRTF) und die Modellierung der ITD und ILD dar. Von der Aufnahme bis zur Wiedergabe gibt es verschiedene Varianten: So kann bereits die Aufnahme des Materials binaural geschehen mit einem Kunstkopf oder der Jeklinscheibe. Alle wichtigen Höranker (ILD, ITD, HRTF, Raumklang) sind in der Aufnahme vorhanden um die originalgetreue Wiedergabe der speziellen Aufnahmesituation zu reproduzieren. Zusätzliche binaurale Bearbeitung ist nicht erforderlich. Diese Lösung ist mit relativ einfachen Mitteln zu erreichen, keine leistungsfähigen Rechner oder Speicher sind notwendig um eine Aufnahme zu erhalten. Wird ein Signal nicht bereits in der Aufnahmephase binaural aufgenommen, ist es dennoch möglich dieses binaural aufzubereiten. Hierfür ist neben leistungsfähigen Rechnern eine Datenbank aus gemessenen HRTFs, beziehungsweise dem Zeitequivalent kopfbezogene Außenohr-Impulsantworten (eng.: HRIR) notwendig. Eine (statische) Quelle kann dann mit dem zugehörigen Paar HRTFs oder HRIR gefiltert werden und man erhält somit bei Kopfhörerwiedergabe den Richtungseindruck, als ob die Quelle aus dieser Richtung einfallen würde. Auch hierbei sind ILD und ITD durch die Messung der HRIR berücksichtigt. Zusätzlich kann Raum in der Messung der HRIRs so genannte binaurale Raumimpuls Antwort (BRIR) Realitätsempfindung deutlich steigern. Vorteil ist, dass die Aufnahme sofort über Kopfhörer auf Stereo Equipment wiedergegeben werden kann. Anders verhält es sich, wenn binaurale Lautsprecherwiedergabe ermöglicht werden soll. Dann ist ein Encoder notwendig, der das Kanalübersprechen der beiden Lautsprecher verhindert. Die erforderlichen Daten können bereits encodiert auf das Wiedergabemedium gespeichert werden, oder in der Wiedergabeanlage befindet sich ein zusätzliches Gerät, das diese Arbeit übernimmt.

Die erste in Deutschland öffentliche Sendung, die mit Binauraltechnik über Radio gesendet wurde war 1973 ein Hörspiel zum Anlass der Funkausstellung in Berlin³.

Kommerzielle binaurale Klangwiedergabe findet man heute am häufigsten in Computerspielen, die für einen Benutzer eine möglichst reale Umgebung erzeugen wollen, oder als Stereoverbreiterungs Ansatz in Software- und Hardware- DVD-Playern, der versucht Mehrkanalwiedergabe über Lautsprecher für Kopfhörer zu simulieren.

3 http://de.wikipedia.org/wiki/Binaurale_Tonaufnahme

3 Stand der Technik

Fokus in diesem Kapitel ist die Konversion von 2-Kanal- auf 5-Kanalwiedergabe, dem so genannten Blind Upmix, und die Rückkonversion von 5-Kanal- auf 2-Kanalwiedergabe zur Erzeugung von virtuellem Surround mittels Binauraltechnik. Beide Techniken sind am Fraunhofer Institut für Integrierte Schaltungen (IIS) bereits funktionsfähig implementiert und getestet worden.

3.1 *Blind Upmix nach IIS*

Die implementierte Technik basiert auf einem Ansatz, der von Avendano und Jot [26] vorgeschlagen wurde. Es handelt sich um eine Quellseparationstechnik, die die Positionierung der einzelnen Schallquellen im Stereopanorama als Grundlage benutzt. Der Ansatz bezieht sich dabei vorwiegend auf Positionierungen, die mit Pegelunterschieden zwischen linkem und rechtem Kanal (Intensitätspositionierung) hergestellt wurden, zeitliche Positionierung wird vernachlässigt. Die getrennten Signale aus der 2-Kanal Stereoverversion werden auf fünf Kanäle für die Wiedergabe in einer Lautsprecheraufstellung nach ITU-R BS 775[27] konvertiert. Zusätzlich wird ein sechster Kanal für einen eventuell vorhandenen Subbasslautsprecher generiert, dieser ist jedoch im Rahmen der vorliegenden Arbeit nicht von größerem Interesse. Im Zuge einer Diplomarbeit [4] am IIS wurde ein System aufbauend auf dieser Technik implementiert. Mit dem Verfahren ist es möglich zwei verschiedene Blind Upmix Philosophien zu verfolgen:

- „Ambient“ Ansatz: Natürlicher Upmix, Positionen der Instrumente vor dem Hörer bleibt erhalten, in die hinteren Kanäle kommen keine Direktschall-, nur Ambientanteile
- „In the Band“ Ansatz: Quellen werden um den Hörer herum verteilt

Ausgegangen wird dabei von einem in der Stereomischung auftretenden Signalmodell, gültig für Intensitätspositionierung:

$$x_i(t) = \alpha_{ij} \cdot \left(\sum_{j=1}^N s_j(t) * d_{ij}(t) + \sum_{j=1}^N s_j(t) * r_{ij}(t) + n_i(t) \right) \quad 3.1$$

x_i = Signal aus Kanal links ($i=1$) oder rechts ($i=2$)

s_j = Schallquelle bei Präsenz von N Schallquellen

d_{ij} = Direktschallkomponente für Kanal i und Schallquelle j

r_{ij} = Nachhallkomponente für Kanal i und Schallquelle j

n_i = Zusätzliche Nebengeräusche

Eine Schallquelle s_j in einem Kanal x_i der Stereowiedergabe wird beim Abmischprozess als Summe des Direktanteiles $s_j \cdot d_{ij}$, Raumanteiles $s_j \cdot r_{ij}$ und von der Schallquelle unabhängigen Grundgeräusch n_i aufgefasst. Beim Raumanteil wird davon ausgegangen, dass er für linken und rechten Kanal wenig korreliert und wenig kohärent ist. Weitere Annahme für den Raumanteil und auch das Grundgeräusch ist, dass in linkem und rechtem Kanal vergleichbare Energien vorhanden sind. Um der Quelle eine Position im Stereobild zu zuweisen, wird dann jeder Kanal mit den Verstärkungsfaktoren α_{ij} nach dem Sinus-Cosinus-Gesetz versehen.

$$\frac{\sin(\rho_p)}{\sin(\rho_0)} = \frac{\alpha_1 - \alpha_2}{\alpha_1 + \alpha_2} \quad 3.2$$

ρ_p ist der Winkel der Phantomschallquelle ρ_0 ist der Basiswinkel der Lautsprecheranordnung.

Für die Verstärkungsfaktoren gilt:

$$\alpha_{2,j} = \sqrt{1 - \alpha_{1j}^2} \quad 3.3$$

Mathematische Grundlage für die Upmixtechnik ist die Betrachtung der Signale im Zeit-Frequenzbereich. Die Signale werden in kurzen Zeitblöcken frequenztransformiert und analysiert. Ein Positionsindex wird berechnet durch die Einführung von Maßen für die Ähnlichkeit der Kanäle:

Korrelation:

$$KOR = \Re \left\{ \frac{KK_{12}}{\sqrt{E_{11} \cdot E_{22}}} \right\} \quad 3.4$$

Kohärenz:

$$KOH = \left| \frac{KK_{12}}{\sqrt{E_{11} \cdot E_{22}}} \right| \quad 3.5$$

Wobei gilt:

$$\begin{aligned} E_i(d, f, \alpha) &= \{\alpha \cdot [X_i(d, f) \cdot X_i^*(d, f)]\} + \{(1 - \alpha) \cdot E_i(d - 1, f, \alpha)\} \\ KK_{12}(d, f, \alpha) &= \{\alpha [X_1(d, f) \cdot X_2^*(d, f)]\} + \{(1 - \alpha) \cdot KK_{12}(d - 1, f, \alpha)\} \end{aligned} \quad 3.6$$

E_i = Energie im Kanal, wobei $i = l, r$

d = Index des Zeitblocks

f = Frequenz

α = Eingefügter Faktor um Zeitabhängigkeiten nicht zu vernachlässigen

Ähnlichkeitsfunktion:

$$\psi(d, f) = 2 \cdot \frac{|KK_{12}(d, f, 1)|}{[E_1(d, f, 1) + E_2(d, f, 1)]} \quad 3.7$$

Partielle Ähnlichkeitsfunktion

$$\psi_i(d, f) = \frac{|KK_{ij}(d, f)|}{E_i(d, f)} \quad i, j = 1, 2 \wedge i \neq j \quad 3.8$$

Ausgehend vom Signalmodell (Gl. 3.1) wird nun eine Quelle im Stereobild positioniert (Verstärkungsfaktoren für linken und rechten Kanal nach Gleichung 3.3), und so liefert die Ähnlichkeitsfunktion (Gl. 3.7) den Wert

$$\psi(m, k) = 2\alpha \cdot \sqrt{1 - \alpha^2}$$

Dieser Wert ist maximal (=1) wenn eine Quelle in die Mitte positioniert ist und minimal (=0) für links oder rechts positionierte Quellen. Die quadratische Abhängigkeit vom Verstärkungsfaktor lässt eine Unterscheidung zwischen links und rechts noch nicht zu, somit wird als zusätzlicher Wert die Differenz der partiellen Ähnlichkeiten (Gl. 3.8) berechnet:

$$\Delta(d, f) = \psi_1(d, f) - \psi_2(d, f)$$

Dabei ist nur das Vorzeichen interessant:

$$\hat{\Delta}(d, f) = \begin{cases} 1 & \forall \Delta(d, f) \geq 0 \\ 0 & \forall \Delta(d, f) = 0 \\ -1 & \forall \Delta(d, f) \leq 0 \end{cases} \quad 3.9$$

Wobei positive Werte linken und negative rechten Signalen entsprechen.

Der Positionsindex ergibt sich nun zu

$$\Psi(d, f) = [1 - \psi(d, f)] \hat{\Delta}(d, f) \quad 3.10$$

Weiterer Schritt ist die Trennung von Ambient- und Direktsignalen. Mit der Voraussetzung der Intensitätspositionierung der Schallquellen im Stereobild gilt nur für Direktsignale eine höchstmögliche Korrelation (Gl. 3.4) mit dem Wert von 1. Das gleiche gilt für die Kohärenz der beiden Kanalsignale. Daher wird ein Maßwert ζ aus dem Korrelationswert und dem Kohärenzwert gebildet.

Aus berechneten ζ Werten lässt sich ein Übergangsbereich festlegen, ab wann ein Signal per Definition zum Ambient- oder Direktanteil gehört. Die Trennung wird in Form von Gewichtungsfunktionen $f_{direkt/ambient}(d, f)$ realisiert, wobei immer gilt:

$$f_{direkt}(d, f) = 1 - f_{ambient}(d, f) \quad 3.11$$

Die so getrennten Signale werden auf die 3 neuen Front Kanäle (L,R,C) verteilt. Der Front Ambient Anteil des linken und rechten Kanals verbleibt zu gleichen Teilen im linken und rechten Kanal. Der Centerkanal wird ohne Ambientanteil erzeugt.

$$\begin{aligned} L_{amb} &= f_{ambient}(d, f) \cdot |X_1(d, f)| \\ R_{amb} &= f_{ambient}(d, f) \cdot |X_2(d, f)| \end{aligned} \quad 3.12$$

Bei der Neuverteilung des Direktanteiles wird neben den Gewichtungsfunktionen zusätzlich der Positionsindex betrachtet: Zunächst werden seine Werte $[-1,1]$ auf das Intervall $[0, \pi/2]$ abgebildet und die Gewichtungsfaktoren $\gamma_{1,2}$ berechnet:

$$\begin{aligned} \Psi_{deg} &= \frac{\Psi + 1}{2} \cdot \frac{\pi}{2} \\ \gamma_1 &= \cos(\Psi_{deg}) \\ \gamma_2 &= \sin(\Psi_{deg}) \end{aligned} \quad 3.13$$

Die Umrechnung der Positionen von Phantomschallquellen in Verstärkungsfaktoren wird mit dem Tangensgesetz vorgenommen.

$$\Omega_s = \arctan\left(\tan(\Omega_{0s}) \cdot \frac{\gamma_{2s} - \gamma_{1s}}{\gamma_{2s} + \gamma_{1s}}\right), \quad s = L, R \quad 3.14$$

Um Energieerhaltung zu garantieren werden die Verstärkungsfaktoren normiert:

$$\begin{aligned} G_{links} &= \frac{\gamma_{links}}{\sqrt{\gamma_{links}^2 + \gamma_{center}^2}} \\ G_{center} &= \frac{\gamma_{center}}{\sqrt{\gamma_{links}^2 + \gamma_{center}^2 + \gamma_{rechts}^2}} \\ G_{center} &= \frac{\gamma_{rechts}}{\sqrt{\gamma_{rechts}^2 + \gamma_{center}^2}} \end{aligned} \quad 3.15$$

Für Ambient Signale gilt zusätzlich die Annahme, dass sie vergleichbare Energien in den Kanälen haben. Somit bietet sich ein Energievergleich als weiteres Kriterium für die Trennung von Ambient- und Direktsignalen beider Kanäle an. Ein Schwellwert kann definiert werden, ab dem Signale als Ambient- oder Direktsignale definiert werden.

$$\Delta E_{dB}(d, f) = \left| 10 \cdot \log_{10} \left[\frac{X_1(d, f) \cdot X_1^*(d, f)}{X_2(d, f) \cdot X_2^*(d, f)} \right] \right| \quad 3.16$$

Ist der Pegel unterhalb der Schwelle, so wird der Anteil im Ambientsignal voll gewichtet, überhalb wird er je nach Höhe des Pegels gedämpft.

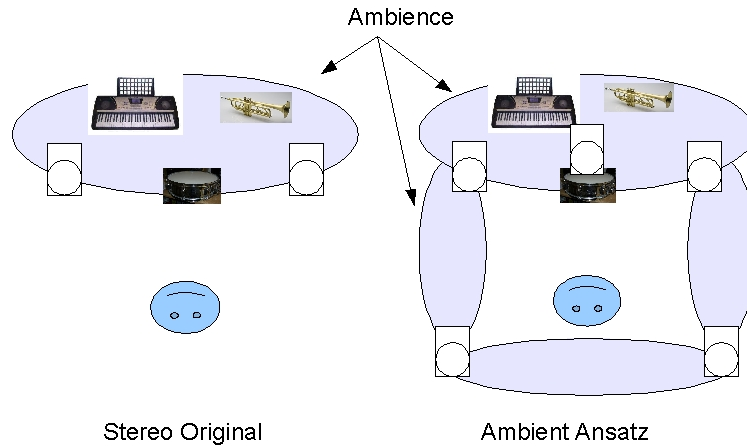


Abbildung 10: Blind Upmix mit Ambientansatz

Diese Art der Trennung von Signalen wird verwendet um beim Ambientansatz Signale für die hinteren Kanäle (L_{surround} , R_{surround}) zu generieren (Abbildung 10). Für den „In the Band“ Ansatz werden die gleichen Vorgehensweisen zur Signaltrennung benutzt, jedoch werden die hinteren Kanäle nicht ausschließlich zur Reproduktion von Ambientanteilen benutzt. Stattdessen wird der berechnete Positions-Index nicht nur umgerechnet auf die 3 Front Kanäle, sondern es werden alle Lautsprecher verwendet. So entsteht eine Art Stereo Basis Verbreiterung um den Hörer (Abbildung 11). Eine Schallquelle mit einem Positions-Index im Intervall $[-1, 0]$ würde auf die Lautsprecher L_{surround} , L und C verteilt werden. Problem für den „In the Band“ Ansatz ist die Tatsache, dass die Positionierung von stabilen Phantomschallquellen seitlich vom Hörer nicht möglich ist [28]. Zusätzlich ist die Qualität der Separation nicht hoch genug um die Quellen eindeutig zu positionieren [4]. So wird eine lineare Methode angewandt um Verstärkungsfaktoren für die Direktsignale in den jeweiligen Kanälen zu berechnen. Ambientanteile werden auf allen Lautsprechern wiedergegeben außer auf dem Centerlautsprecher.

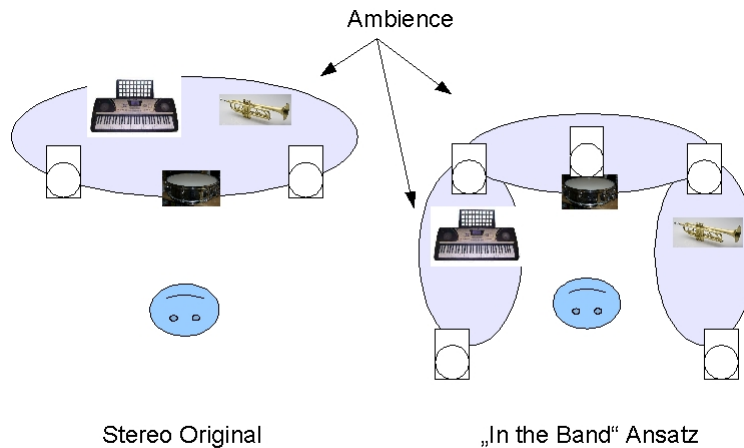


Abbildung 11: "In the Band"- Ansatz

3.2 Kopfhörerwiedergabesystem nach IIS

Hierbei handelt es sich um ein System, dass auf den Prinzipien der binauralen Wiedergabe über Kopfhörer mit Außenohr-Übertragungsfunktionen (HRTFs), beziehungsweise der im Zeitbereich äquivalenten Außenohr Impulsantwort (HRIR) aufbaut. Ziel dieses Systems ist die virtuelle Nachbildung der Surroundwiedergabe über Lautsprecher in einem Raum für die Kopfhörerwiedergabe. Ausgegangen wird vom Modell (Abbildung 12) für den Signalweg vom Lautsprecher zum Ohr. Im Falle der Kopfhörerwiedergabe existiert nur der direkte Übertragungsweg von einem Kanal an ein Ohr, ohne Kanalübersprechen, wie es etwa bei Wiedergabe über Lautsprecher der Fall ist. Das System modelliert die Übertragungswege im Fall von Lautsprecherwiedergabe von jedem Lautsprecher zu jedem Ohr und summiert das entstehende Schallfeld für das jeweilige Ohr auf. Dabei werden die in Kapitel 2 besprochenen Höranker einer realen Wiedergabe benutzt um ein möglichst reales Abbild der Wiedergabesituation zu erzeugen. Wesentlichster Schritt ist die Messung der binauralen Raumimpulsantworten (BRIA). Sie dienen anschließend als Filter, in denen die Höranker (HRIR, ILD, ITD, Raumanteil) enthalten sind.

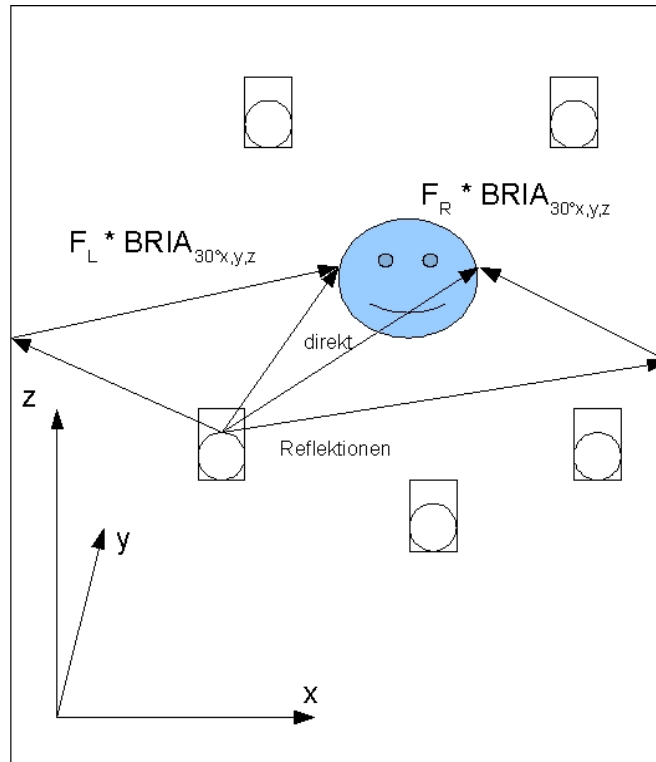


Abbildung 12: Signalmodell Surround Erlebnis

3.2.1 Messung der binauralen Raumimpulsantworten (BRIA)

Die BRIA können typischer Weise am Kunstkopf oder am individuellen Kopf mit Ohrmikrofonen gemessen werden. Für die Arbeit wurden ausschließlich Kunstkopfmessungen verwendet, da dies den realen Anwendungsfall dieses Systems besser widerspiegelt. Für breitenwirksamen Einsatz dieses Systems ist derzeit keine Möglichkeit der effizienten, individuellen BRIA Messung möglich. Die BRIA wurden gemessen mit einem *Cortex MK2* Kunstkopf, der sich mittig in einer typischen nach der Empfehlung ITU-R BS 775-1 [27] aufgestellten 5.0 Surroundaufstellung befindet (Abbildung 13). Der Radius der Kreisauflistung beträgt 2 Meter in deren Mitte sich der Kunstkopf befindet. Der Kunstkopf und die Lautsprecher befinden sich auf einer Höhe von zirka 1.5 Meter. Die Raummaße betragen 6m x 4m x 3m. Der Raum hat Teppich Fußboden, absorbierendes Material an der Decke, eine Fensterfront und ansonsten normale Wände. Für die Messung wurden 2 verschiedene Arten von Lautsprechern benutzt: für den Center wurde ein *Genelec Biamp 1032A*, für alle anderen ein *Genelec Biamp 1031A* Lautsprecher verwendet.

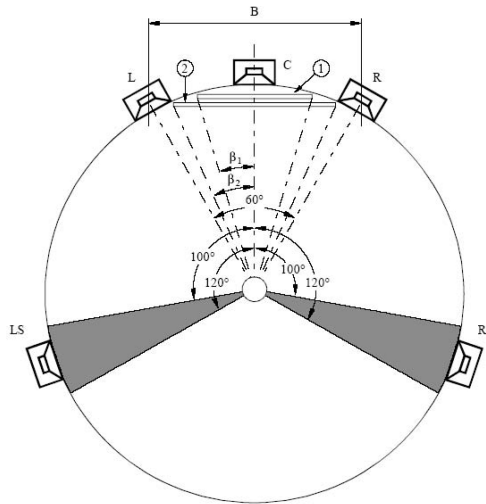


Abbildung 13: Messung der BRIA im Surround Setup aus [27]

Die Messung wird mit der MLS Methode durchgeführt: an jeweils einem Lautsprecher wird ein 16 bit Pseudoranschsignal angelegt, und der Schalldruck an den beiden Ohren gemessen. Die 2^{16} Werte lange Impulsantwort erhält man durch Korrelation des Meßsignals mit dem Mikrophonsignal. Die so erhaltene BRIA beschreibt die Übertragung von einer Schallquelle am Ort des Lautsprechers zum jeweiligen Ohr des Kunstkopfes. Aus Effizienzgründen bezüglich Implementierbarkeit werden nur die ersten 4096 Werte (ca 92 ms) der Messungen benutzt. Die Nachhallzeit des Raumes liegt bei zirka 290 ms (T_{20} -Schätzung aus dem Mittel mehrerer Impulsantworten) für das Oktavband um 500 Hz. Für die Implementierung des Systems wurde jedoch angenommen, dass genug Rauminformation im Signal vorhanden ist, um ihn ausreichend zu simulieren.

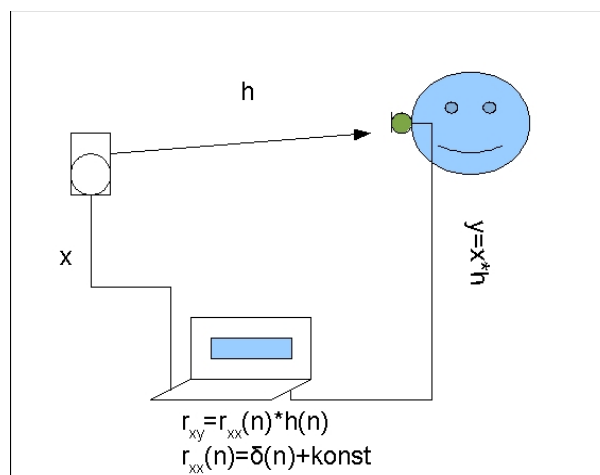


Abbildung 14: MLS Methode

Aus dem geometrischen Modell (Abbildung 15) errechnen sich erste Reflexionen im Raum, welche das Ohr erst nach etwa 4 ms ($\equiv 176$ Samples bei 44,1 kHz) erreichen. Signalausschläge, die unmittelbar nach dem Eintreffen des Direktschalls am Ohr gemessen werden, stellen die im Zeitbereich gegebene Außenohr Impulsantwort (HRIR) dar. Erst danach treffen erste Reflexionen von Boden und Decke ein. Dieser Zusammenhang wird deutlich bei Betrachtung der Messungen in Abbildung 16.

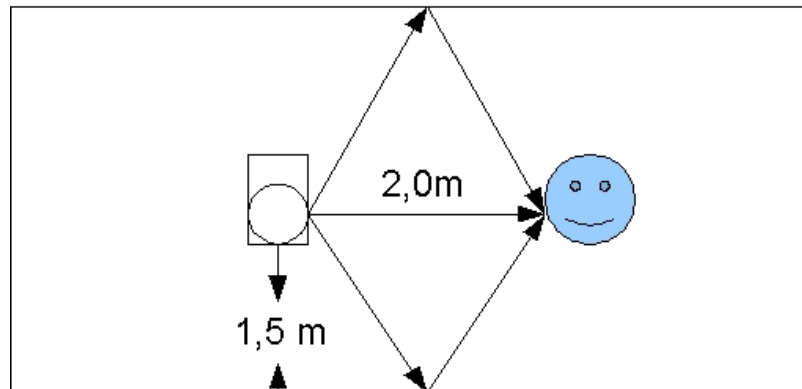


Abbildung 15: Richtungen der ersten Reflexionen

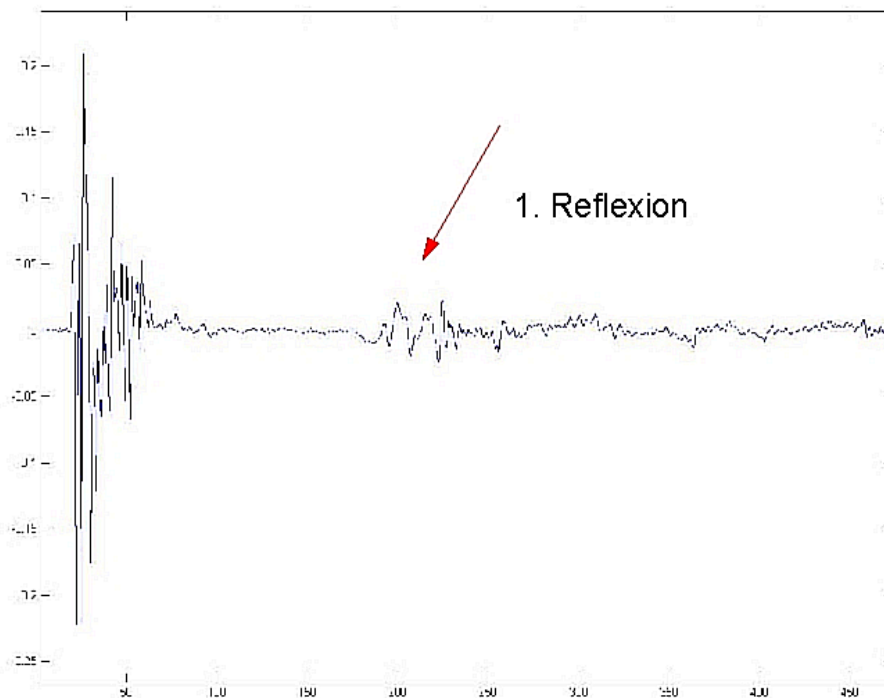


Abbildung 16: Ausschnitt aus gemessener BRIA

Um auch Kopfdrehungen simulieren zu können werden Messungen für alle 5 Lautsprecher des Surroundsetups in 4.5° Auflösung der Kopfdrehrichtungen gemacht. Anschließend wird eine spezielle Diffusfeldentzerrung auf das Set der gemessenen BRIA angewendet um unter anderem die Effekte der Meßkette zu kompensieren.

3.2.2 Realisierung des binauralen Wiedergabesystems

Das Herz des Systems ist eine Bibliothek in der die gemessenen binauralen Raumimpulsantworten in Tabellen für das linke und rechte Ohr gespeichert liegen. Da das System dafür konzipiert ist 5.1 Surround Signale, oder 2.0 Stereo Signale zu bearbeiten liegen in jeder dieser Tabellen 6 BRIA die mit den Übertragungswegen vom jeweiligen Lautsprecher zum Ohr korrespondieren. Die Codierung der Lautsprecher hält sich dabei an die gängige Abfolge, die auch in 6 Kanal .wav-Dateien eingehalten wird:

- Links: BRIA von 30°
- Rechts BRIA von -30°
- Center BRIA von 0°

- Low Frequency Effects
- Linkssurround BRIA von 120°
- Rechtssurrounds BRIA von -120°

Die Ohrsignale entstehen aus Superposition der entsprechend gefilterten Lautsprechersignale.

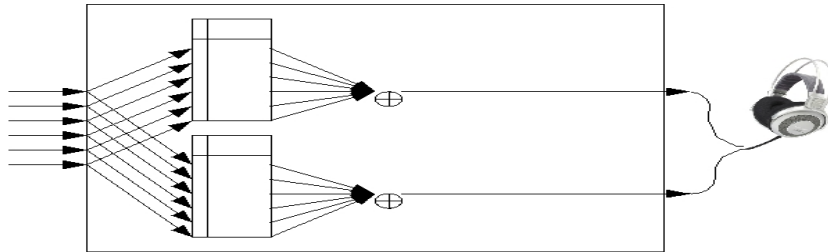


Abbildung 17: Prinzip des binauralen Downmixes

4 Verfahren

Die vorgestellten Systeme aus Kapitel 3 erreichen hohe Wiedergabequalität, wenn sie als Einzelsysteme arbeiten, da sie für spezielle Wiedergabesituationen optimiert sind (Lautsprecherwiedergabe für Blind Upmix und diskretes 6 Kanal Eingangssignal für binaurale Wiedergabe). Offensichtliche Nachteile für das Kombinationssystem sind, dass der 5 Kanal Upmix auf 2 Kanälen wiedergegeben wird und das binaurale Wiedergabesystem Signale zur Verarbeitung bekommt, die nicht der Natürlichkeit einer diskreten 6 Kanalproduktion entsprechen. Die Erwartungen an das Kombinationssystem sind hoch: das Ergebnis muss eine echte Alternative, im Sinne von Hörerbevorzugung zur Stereowiedergabe sein, da es in der Realität dem ständigen Vergleich zu dieser ausgesetzt ist. Als Ziel wird gesetzt, dass die Aufwertung, die ein 2 Kanal Stereo Signal durch den Blind Upmix für Lautsprecherwiedergabe erfährt, mit Binauraltechnik reproduzierbar gemacht werden kann. Beide Systeme bieten aufgrund ihrer Implementierung eine Vielzahl von Parametern an, die Änderungen an der Verarbeitung der Eingangssignale bestimmen. Die Zuordnung zu Hörerlebnissen ist nicht gegeben, jedoch in höchstem Maße interessant. Hinweise über Einflüsse auf Höreindrücke bestimmter Signalverarbeitung im System helfen ein hohes Maß an Wiedergabeästhetik zu schaffen. Das Kapitel beschreibt Ausgangspunkt und Vorgehensweise, mit der physikalische Variablen des Systems mit Wahrnehmungsattributen verknüpft und mittels Variation zu höchstmöglicher Hörerbevorzugung geführt werden.

4.1 Das Modell

Ein Modell wird benötigt um Zusammenhänge zwischen physikalischen Parametern des Systems und den Hörerbewertungen herzustellen. Auf der Grundlage von Bechs [18] Arbeit entsteht die Vorgehensweise für die gestellte Anforderung. Abbildung 18 zeigt sein Modell für das Zusammenspiel zwischen den physikalischen Parametern eines Klanggeschehens und deren Auswirkung auf den letztendlichen individuellen Gesamteindruck. Dieses Modell berücksichtigt alle Faktoren auf dem Weg vom subjektiven Wahrnehmungsprozess zum Entscheidungsprozess.

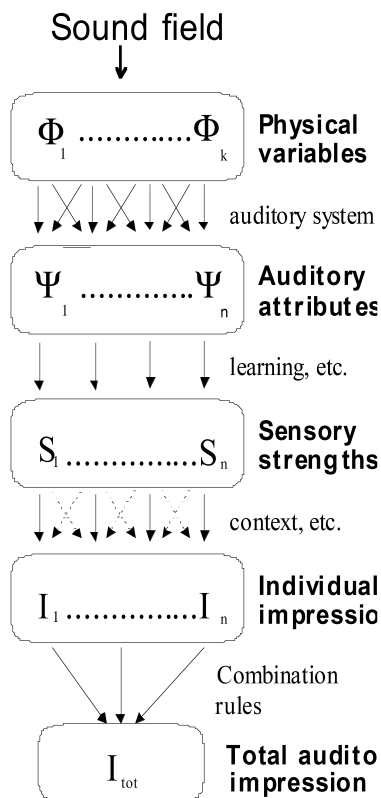


Abbildung 18: Vom Klanggeschehen zur subjektiven Bewertung aus [18]

Dem Klanggeschehen sind also eine Menge physikalischer Variable eigen, die es charakteristisch und eindeutig beschreiben. Als bekannte Beispiele können genannt werden:

- Schalldruck
- ILD, ITD
- spektrale Färbungen der Einzelquellen, bzw des Gesamtgeschehens
- Korrelationsgrad zwischen den Ohrsignalen

Nicht notwendiger Weise sind alle diese Variablen bekannt. Somit ist die objektive Analyse nur eingeschränkt möglich. Es werden Hörer zur subjektiven Bewertung benötigt. Nach der Transformation der physischen Realität des Schallfeldes durch das Hörsystem in psychische Wahrnehmung lässt sich das Geschehen beschreiben durch verbale Attribute, die individuell Gültigkeit haben. Auf der Basis gemeinsamer Erarbeitung, im Sinne wahrnehmungsbetonter

Tests (Kapitel 2.2.2.2) kann damit Gültigkeit für eine Mehrzahl an Individuen erreicht werden. Diese Beschreibung ist dann ebenfalls charakteristisch und eindeutig. Die Eindrücke werden in ihrer Summe zu einem individuellen Gesamteindruck im Sinne einer Bevorzugung führen. Das Gewicht der einzelnen Summanden gerät hierbei zu einem Problem. So zeigt sich konkret in informellen Hörtests spektrale Verfärbung von bearbeiteten Musikstücken in hohem Maße als präferenzbestimmend. Positive Eigenschaften räumlicher Natur spielen ihr gegenüber eine untergeordnete Rolle. Am deutlichsten wird diese Tatsache, wenn Monoversionen von Musikstücken besser empfunden werden als binaural bearbeitete. Selbst die als angenehm beschriebene Außer-Kopf-Lokalisation führt dann zu keiner Bevorzugung mehr. In der Literatur wird diese Unterscheidung nach räumlichen und klangfarblichen neben weiteren Aspekten oft vorgenommen [22],[2]. Somit wird versucht für diese Art der Wiedergabe, die in hohem Maße auf räumliche Wahrnehmung baut, bewertungsverzerrende klangfarbliche Attribute soweit als möglich zu kompensieren um vornehmlich räumliche Aspekte beurteilen zu können.

Die additive Komponente 'Kontext Effekte' ist ein wichtiger Faktor, der für das Design der Testprozedur zu betrachten gilt. Ihre Bedeutung ist nicht unumstritten: Sollen sie in das Design explizit eingebunden werden, oder durch, zum Beispiel eine Zufallsstrategie, herausgemittelt werden? In [29] wird der Einfluss des Kontext Effekts 'Abfolge der Testkonditionen' als Einflussfaktor auf das Testergebnis diskutiert. Das Hinzufügen von zusätzlichen Variablen zum statistischen Modell erlaubt eine höhere Sensibilität des Tests. Die Befürchtung, dass der Einfluss des Kontext Effekts mit dem Effekt der unabhängigen Variable verwechselt werden könnte, wird kontrollierbar. Zu berücksichtigen ist, das dabei Freiheitsgrade der Fehlervarianz verloren gehen, was zu reduzierter Aussagekraft der Teststatistik führen kann. In der Literatur wird vorgeschlagen, dass bei nicht signifikanten Effekten eines Faktors dieser aus genanntem Grund aus dem Modell genommen werden sollte. Im Verlauf der Testprozeduren zur Evaluierung des vorliegenden Systems werden Kontext Effekte nicht explizit in das Testdesign eingebunden, da angenommen wird, dass sie keinen signifikanten Einfluss auf das Testsystem haben. Sie werden durch Konstanthalten der Testbedingungen oder Anwendung von Zufallsstrategien kontrolliert, beziehungsweise herausgemittelt. (Zur näheren Beschreibung siehe Kapitel 4.3.2)

Im Rahmen dieser Arbeit wird zunächst am vorliegenden Kombinationssystem eine Analyse des Systems vorgenommen, und seine Probleme bei der Wiedergabe konstatiert. Dies geschieht mit einem Ansatz, basierend auf einem wahrnehmungsbetonten Testdesign.

Physikalische Parameter am System werden gesucht, die Einfluss auf die festgestellten Wiedergabeeigenschaften haben. Mit Hilfe einer Echtzeitumgebung können Einflüsse schnell beurteilt werden. Mit informellen Hörtests wird die Wirkung bestätigt. Die einflussreichsten Parameter werden in Pilothörexperimenten auf ihren Einfluss bezüglich Hörerbevorzugung getestet. Die Kombination präferenzfördernder Parameter wird getestet. Schlussendlich wird ein System ausgewählt und zur Verifikation gegen das ursprüngliche System in einem abschließenden Hörexperiment getestet.

4.2 Problemdiagnose

In diesem Kapitel wird die Durchführung informeller Hörtests beschrieben und ein Blick in ähnliche Arbeiten der jüngsten Vergangenheit geworfen. Die Antwort auf die Frage nach präferenzfördernden Attributen räumlicher Wiedergabesysteme und den diesbezüglichen Einschränkungen des vorliegenden Systems soll gefunden werden. Das System im Test ist das aus Blind Upmix und binauraler Wiedergabe kombinierte System. Das Design möchte die Stärken der parallelen Bewertung eines Systems nach affektbetonten und wahrnehmungsbetonten Tests (Kapitel 2.2.2) nutzen um Informationen über die Qualität des vorliegenden Systems zu erhalten. Wiedergabeeigenschaften des Systems werden anhand einer freien verbalen Beschreibungsmethode bestimmt. Mittels Diskussion über subjektive Präferenz und Erfahrungswerten anderer Arbeiten zum Thema werden wichtige Diagnosen über das System gestellt. In der Annahme, dass Veränderung bestimmter Attribute Veränderungen in der Präferenzbewertung zur Folge haben, werden physikalische Variablen am System zur Veränderung gesucht.

4.2.1 Finden verbaler Beschreibung

Dieses Kapitel beschreibt ein von den benannten Methoden (Kapitel 2.2.2.2) inspiriertes Experiment zur Aufstellung von Beschreibungsformen für Hörereignisse mit dem Blind Upmix Binauralwiedergabesystem. Eine große Anzahl von Musikstücken aller Genres wurde dafür herangezogen. Es wurde darauf geachtet, dass bestimmte Kriterien bezüglich der Aufnahmen mit dieser Auswahl abgedeckt werden konnten. Im Anhang zeigt eine Übersicht über benutzte Musikstücke.

- Stücke, die mit deutlich wahrnehmbarer Position von Schallquellen (Amplitudenstereophonie)
- Stücke mit deutlichen (überlagerten) Raumeffekten und Zeitstereophonie

- Stücke mit hohem Transientanteil
- Live Aufnahmen mit hohem Ambientanteil

Aus jedem dieser Stücke wurde ein repräsentativer Part herausgeschnitten zu jeweils einer Maximaldauer von 1 Minute. Jede der beteiligten Personen hatte über das Betriebsinterne Netzwerk jederzeit Zugriff auf die Sammlung von Musikstücken. Um gegenseitige Beeinflussungen aus zu schließen, war jede Person dazu angehalten sich in einem begrenzten Zeitraum (2 Wochen) eine Liste mit persönlichen Beschreibungseigenschaften zu erstellen, die Unterschiede zur Stereowiedergabe markierten. In mehreren Diskussionsrunden wurde das System im Vergleich zur Stereowiedergabe blind getestet. Jeder hatte nun die Möglichkeit seine Wahrnehmung mit seiner Liste zu verifizieren und mit den anderen Teilnehmern zu diskutieren. Somit erfolgt eine Selektion und gleichzeitig Anpassung der wichtigsten Beschreibungsformen, die für dieses System im Vergleich zur Stereowiedergabe generisch sind. Ein statistisches Modell und eine dahingehende Durchführung eines wahrnehmungsbetonten Tests nach einer der bekannten Methoden (vgl Kapitel 2.2.2.2) wurde nicht angewandt. Die gefundenen beschreibenden Attribute werden als unabhängig angenommen. Am Ende stand eine Liste der nach Meinung der Teilnehmer und des Autors wichtigsten Beschreibungsattribute für das Blind Upmix Binaural System die es entscheidend im negativen Sinne von einem Stereowiedergabesystem unterscheiden.

- Frequenzgang: Gegenüber dem Stereo Original entsteht eine hörbare Frequenzverzerrung -> Unangenehme Verzerrung, Verfälschung
- Entfernungswahrnehmung: Klanggeschehen findet außerhalb des Kopfes statt in bestimmter Entfernung. -> Schöner Effekt. Aber Entfernung evtl zu groß
- Raumwahrnehmung: Klanggeschehen findet in einem mehr oder weniger wahrnehmbaren Raum statt. -> Raumwahrnehmung störend bei trockenen impulsiven Signalen (Badezimmereffekt)
- Lokalisation: Einzelne Quellen sind klar bestimmbar und entsprechen in Richtung und Relation der Stereowiedergabe. -> Aber: Klanggeschehen nimmt insgesamt weniger Breite ein
- Einhüllung: Klang passiert um den Hörer herum. -> Zu wenig wahrnehmbar

Einigkeit bestand darüber, dass eine Beeinflussung zu findender physikalischer Parameter unmittelbar zur Änderung der beschreibenden Attribute führen kann. Änderungen von

Attributen folgt eine Veränderung in der Präferenzbewertung, welche Gegenstand genauerer Untersuchung sein sollte.

4.2.2 Ähnliche Arbeiten und Ergebnisse

Ein Blick in die einschlägige Literatur zeigt ähnliche Arbeiten im Zusammenhang mit Präferenzbewertung von Lautsprecher- und Kopfhörerwiedergabesystemen. Erfahrungen und Erkenntnisse sollen in dieser Arbeit nicht unberücksichtigt bleiben. Marui und Martens [1] bewerten verschiedene Signalverarbeitungsalgorithmen für Kopfhörerwiedergabe in subjektiven Hörtests. Kriterien sind dabei einmal der Gesamteindruck (Präferenz) und die Breite des Klanggeschehens (als einzelnes Attribut). Der Test wird außerdem mit verschiedenen Kopfhörertypen (Zirkumaural und „Im Ohr“ Kopfhörer) durchgeführt. Die Ergebnisse zeigen signifikante Effekte unabhängig von der Art des Kopfhörers, dass eine hohe Bewertung der Breite des Klanggeschehens mit hoher Präferenzbewertung korreliert. Diese Erkenntnis hat großen Einfluss auf das weitere Vorgehen innerhalb dieser Arbeit. Weniger offensichtlich und erfreulich sind die Ergebnisse die Lorho in seinen Arbeiten macht. Er testet parallel in affektbetonten Präferenztests und wahrnehmungsbetonten Attributtests diverse kommerzielle Algorithmen zur virtuellen Surroundwiedergabe, beziehungsweise Stereoerweiterung, gegen eine Stereoreferenz [2]. Die gleichen Algorithmen werden in einem Rangfolgetest untereinander ohne explizite Referenz bewertet [3]. In einer dritten Veröffentlichung [20] bewertet er Heimkinosysteme in subjektiven A/B Vergleichen für Lautsprecher und Kopfhörerwiedergabe gegen eine Stereo-Downmix Referenz. Leider ist dabei nicht ganz offensichtlich, welcher Algorithmus welche Bewertung erfährt, aber dennoch: Es gibt in [2] und [3] kein System, das signifikant dem unverarbeiteten Stereo-Ausgangsmaterial vorgezogen wird, meist ist das Gegenteil der Fall. Aus den Attributtests in [2] ist zu folgern, dass keines der Systeme mit hoher Präferenzwertung sich in den Attributwertungen signifikant von der Stereoersion nach oben absetzt. Dies führt zu dem Schluss, dass Eigenschaften am vorliegenden System, wenn sie nicht die erweiterte räumliche Darstellung betreffen, auf jeden Fall so nahe am Stereo sein sollten wie möglich. Dennoch hätte man erwartet, dass bei dem Signalverarbeitungsaufwand sich das ein oder andere System deutlich von einer Mehrzahl an Hörern gegenüber dem Stereo-Ausgangsmaterial bevorzugt wird.

Weitere Anhaltspunkte lassen sich finden in Arbeiten zu subjektiven Bewertungen im Bereich Surroundwiedergabe mit Lautsprechern: Um eine Methode zur objektiven Messung der

Einhüllung zu finden, führen Souloudre et al [30] 4 subjektive Tests durch, in denen das Einhüllungsgefühl (keine Präferenz) von Hörern gemessen wird. Abhängig soll dieses von der Nachhallzeit, dem Level der späten Reflexionen, der Verteilung, aus welchen der Lautsprecher die späten Reflexionen kommen und dem zeitlichen Übergangspunkt zwischen frühen und späten Reflexionen gemessen werden. Souloudre argumentiert die Wichtigkeit seiner Arbeit mit der Notwendigkeit, dass Wissen aus der Konzertsaalakustik in den Bereich räumlicher Wiedergabesysteme übertragbar ist. Gute Raumwahrnehmung und führt zu einem als angenehm empfundenen Klangfeld. Raumwahrnehmung setzt sich nach Kenntnisstand der letzten 30 Jahre aus zwei Faktoren zusammen: Quellenbreite und Einhüllung des Klanggeschehens.

4.3 Vorbereitungsphase

In diesem Kapitel wird die Vorgehensweise erläutert, mit der Veränderungen am System vorgenommen und bewertet werden. Beginnend mit der Suche nach geeigneten physikalischen Parametern am Kombinationssystem, die in informellen Hörtests als aussichtsreich eingestuft werden, um eine Veränderung von Wahrnehmungen bei der Wiedergabe hervor zurufen. In ausgedehnteren Pilothörtests werden Hörer auf ihre Bevorzugung bestimmter Parametereinstellung befragt. Das Ziel ist die höchst mögliche Bevorzugung für die Mehrzahl an Musikstücken.

4.3.1 Physikalische Variablen und informelle Hörtests

Geeignete Variablen zur Beeinflussung des physikalischen Klanggeschehens, erzeugt durch das aus Blind Upmix und binauraler Wiedergabe kombinierten Wiedergabesystem wurden gesucht. Eine Zuordnung der physikalischen Parameter auf die in Kapitel 4.2.1 diagnostizierten Wahrnehmungseigenschaften und die damit einhergehenden Probleme musste vorgenommen werden, um irrelevante physikalische Parameter auszuschließen. Aus dem Set der relevanten Parameter wurden die wichtigsten in die engere Testprozedur aufgenommen. Beide Systeme liegen als erweiterbare VST-Plugins vor, so dass es möglich ist beide in einer Echtzeitumgebung zu kombinieren. Dies wurde realisiert mit PD und zum Experimentieren benutzt. Tabelle_A II Im Anhang zeigt die im Plugin verfügbaren Parameter. Zusätzlich war es möglich mit dem Softwarepaket von MATLAB in die verschiedenen Implementierungen der Systeme einzugreifen. Die Notwendigkeit dafür bestand, um Möglichkeiten bezüglich der Parameter zu erweitern. Ein genauer Blick auf Tabelle_A II

zeigt, dass ein Großteil der Parameter sich aus Verstärkungsfaktoren für die jeweiligen Kanäle und Übergangsschwellwerten für die Trennung in Ambient- und Direktsignale zusammensetzt. Der Einsatzbereich dieser ist begrenzt, so dass neue Möglichkeiten erschlossen werden mussten (Tabelle 4.1). Zusätzlich ist es möglich mit den MATLAB Implementierungen Versionen mit fest eingestellten Parametern beliebiger Musikstücke in größerer Anzahl zu generieren. Diese Versionen dienten für die ständig durchgeführten Hörtests. Es wird vereinfacht der Zusammenhang zwischen physikalischer Variable und Wahrnehmung angenommen. Diese Annahmen des Autors werden über kurze Blindhörtests mit 4 Experten-Hörern überprüft. Als Ausgangspunkt dienen hierbei wieder die in Kapitel 4.2.1 gefundenen fünf Attribute. Die Tests finden in einem gemeinsamen Raum statt, jeder Hörer hat einen eigenen Kopfhörer. Die Attribute sind bekannt. Die Präsentation der Stimuli erfolgt über Steuerung durch den Autor. Dabei wird ein einfaches A-B Verfahren angewendet. Die unveränderte Version des Wiedergabesystems wird aufgrund dieser Attribute gegen das veränderte System bewertet. Die gegenseitige Unabhängigkeit der physikalischen Variablen und ihrer Auswirkung auf die Wahrnehmung werden vorausgesetzt und aufgrund der Bewertungen überprüft. Tabelle 4.1 zeigt einen Überblick der in Erwägung gezogenen technischen Parameter. Anschließend wurden die Parameter nach ihrer Effektivität sortiert und in das nächste Stadium des Tests aufgenommen.

Physikalisch Veränderung	Potentieller positiver Effekt	Tatsächlich wahrnehmbarer Effekt
Verzögerung der Surround Kanäle	Spektrale Färbung, Änderung der Raumwahrnehmung durch Erhöhung der Dekorrelation zu den Frontkanälen	Kaum hörbar, Kammfiltereffekt
Vertauschen der Surround Kanäle	Eventuell Verbesserung der Raumwahrnehmung da potentiell höhere Dekorrelation zu vorne	Kaum hörbar, auch Verstärkung der hinteren Kanäle bringt keinen Effekt
Phasendrehung 180° hinten	Eventuell Verbesserung der Raumwahrnehmung da potentiell höhere Dekorrelation untereinander	Kaum hörbar, auch Verstärkung der hinteren Kanäle bringt keinen Effekt
Spreizung des Centersignales	Änderung des Frontbildes	Nicht hörbar
Kein Centersignal und zusätzlich kommt Stereo Original in linken und rechten Kanal	Änderung der Raumwahrnehmung	Klanggeschehen spielt sich erhöht ab gegenüber der Version mit Center
Frontkanäle werden nicht binaural bearbeitet	Einstellbarkeit der Entfernungswahrnehmung	Keine Außer Kopf Lokalisation wahrnehmbar
„In the Band“ Upmix	Mehr Einhüllungsempfindung durch mehr Energie in den Surroundkanälen	Klanggeschehen passiert auch hinten, weniger Raumgefühl
Kürzen der Impulsantworten	Weniger Raumanteil	Tatsächlich weniger Raumanteil
Dämpfung der kontralateralen Impulsantworten	Verbreiterung des Frontbildes	Frontbild wird verbreitert, Klangfärbung
Verstärkung der hinteren Kanäle	Mehr Einhüllung durch mehr Energie in den	Klangfärbung, Artefakte durch Rearchannel

Physikalisch Veränderung	Potentieller positiver Effekt	Tatsächlich wahrnehmbarer Effekt
	Surroundkanälen	Processing
Anhebung des dekorrelierten Anteiles in hinteren Kanälen von „In the Band“ Upmix durch SVD	Mehr Einhüllung durch Erhöhung der Dekorrelation in den Surroundkanälen	Tatsächlich mehr Einhüllung, aber auch weniger Lokalisierbarkeit
Position der vorderen virtuellen Lautsprecher	Mehr Stereobreite	Ensemblebreite vergrößert, weniger Entfernung, weniger Raum

Tabelle 4.1: Physikalische Variablen

4.3.2 Pilothörtests

Um Anhaltspunkte bezüglich Hörerbevorzugung verschiedener Parametereinstellungen (Konditionen) zu bekommen wurden Pilot Tests mit einer geringen Zahl an Hörern durchgeführt. Diese Tests sollten zur Präferenzbewertung der Parameter dienen. Es wurde eine modifizierte Form der Multi Stimulus Methode mit verstecktem Anker und Referenz (MUSHRA) gemäß ITU-R BS 1534[14] aufgrund ihrer Effizienz ausgewählt. Eine A-B Vergleichsmethode wurde verworfen, da ihr die Referenz fehlt. In Anbetracht der Tatsache, dass in der Realität des fertigen Produkts der Kunde die Möglichkeit haben wird zwischen Original Stereo und Blind Upmix Binauralwiedergabe zu wählen wurde auch für den Test unbedingt der Vergleich zur Stereoreferenz vorausgesetzt. In [13] vorgeschlagene A-B-C Test mit versteckter Referenz wurde wegen seinem ausgedehnten Testdesign, das nötig gewesen wäre für diese Anzahl an Tests verworfen. Um Aussagen bezüglich der Wirkung einer Parametereinstellung machen zu können sollten 3-4 Stufen dieses Parameters getestet werden. Dabei ist neben dem Vergleich zur Stereoreferenz auch der Vergleich zwischen den Stufen von höchstem Interesse, um Tendenzen ablesen zu können. Mit dem A-B-C Test wären im Falle von 3 Stufen pro Musikstück 6 Einzelvergleiche notwendig, um alle Möglichkeiten gegeneinander abzuwägen. Auch die vollen paarweisen Vergleichsmöglichkeiten, die die Multi Stimulus Methode bietet sind nicht gegeben. Es wurde in Anlehnung an ITU-R 1284 [15] für Vergleichstests eine 7 Punkt Verhältnisskala benutzt (Abbildung 19).

Comparison	
3	Much better
2	Better
1	Slightly better
0	The same
-1	Slightly worse
-2	Worse
-3	Much worse

Abbildung 19: Skala für Vergleichstests

Um die Interpretation der Ergebnisse mit den bekannten statistischen Methoden zu gewährleisten wurde sie mit eine Dezimalstelle (somit 61 Stufen) zur kontinuierlichen Verhältnisskala [16]⁴ erhoben. Im Haus existierte bereits eine in C/C++ programmierte Plattform (*Wavswitch* siehe Abbildung 20), die für Codierungs- Vergleichstests benutzt wurde. *Wavswitch* ermöglicht das Umschalten (mit 200 ms *Crossfades*) zwischen den verschiedenen Konditionen (Versionen eines Musikstückes) und stellt eine Bewertungsskala zur Verfügung, auf der ein Cursor per Tastendruck verschoben werden kann. Die Möglichkeit sich kurze Ausschnitte in den Musikstücken zu definieren (Loops) ist in diesem Programm ebenso gegeben wie die Möglichkeit den Test jederzeit durch Drücken der Pausetaste zu unterbrechen. Damit wird der Hörer unterstützt auch bei größeren Anzahlen von Konditionen ähnlich hoher Qualität, durch Anwahl kurzer Musikphrasen sein Tongedächtnis nicht zu überstrapazieren. Jedem Teilnehmer war es somit möglich den Test ohne Anwesenheit des Testleiters (Doppelblindtest) durch zu führen. Die Abfolge der Konditionen wie auch der Musikstücke wurde für jeden Teilnehmer neu und zufällig gemischt um eventuelle Lern- oder Ermüdungseffekte (oben genannte Kontexteffekte) zu minimieren. Die Testteilnehmer hatten keine Zeitbegrenzung bei der Durchführung der Tests.

4 Kapitel 4.2 Seite 72

```

Anwendungen Aktionen
Terminal
Datei Bearbeiten Ansicht Terminal Beiter Hilfe
Item 2/5: HipHop

Press:

<1> for Reference ( 0 dB)
<2> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<3> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<4> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<5> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<6> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<7> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<8> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<9> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<0> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
<8> for Testitem ( 0 dB) [+-----+-----+-----#-----+-----+-----+]
                                Mu_Wo   Worse   Sl_Wo  0  Sl_Be   Better   Mu_Be

<a> to set loop start (<A> to reset): START
<b> to set loop end  (<B> to reset): END

<+> to increase rating of playing item
<-> to decrease rating of playing item

<space> for stop/restart

<shift + q> for next item

6.76 s

```

Abbildung 20: Wavswitch

Mit der Anzahl der Konditionen wurde dem Standard [14] entsprochen und höchstens 8 Systeme (+ versteckte Referenz und Anker) untereinander verglichen. Laut Standard [14] wären bis zu 12 Konditionen pro Musikstück in Ordnung. Die Erfahrung zeigte, dass bereits die Zahl von 8 Konditionen, auch bei reduzierter Anzahl von Musikstücken zu erheblichen Problemen und exorbitantem Zeitaufwand bei manchen Hörern bis zum Füllen einer Präferenzentscheidung führte. Die Hörer benötigten bis zu 60 Minuten um den Test mit nur 5 Musikstücken zu durchlaufen. Dabei ist zu bedenken, dass im Standard von sogenannten „*intermediate impairments*“ also mittleren Beeinträchtigungen des Systems ausgegangen wird. Da im Rahmen der Arbeit jedoch nicht auf Beeinträchtigungen des Systems getestet wird sondern auf Bevorzugung qualitativ gleichwertiger Systeme, ist es möglicherweise eine obere Grenze an zu vergleichenden Systemen, die ein Hörer, auch wenn er geübt sein mag noch beurteilen kann. Um Ermüdungserscheinungen zu verhindern wurde die Anzahl von 8 Konditionen nicht überschritten. Zusätzlich wurde der Hörer ermutigt, wenn nötig, kurze Pausen einzulegen. In vergleichbaren Arbeiten wird eine Zahl von 7 zu vergleichenden Systemen als obere Grenze genannt, jedoch wenn es die Situation erfordert auch überschritten (vgl [2]). Alle Konditionen wurden zeitlich an die Stereoversion angeglichen. Mittels Berechnung der Kreuzkorrelation wurden positive oder negative Verzögerungen ermittelt. Somit war stufenloses Umschalten zwischen verschiedenen Versionen eines Teststückes möglich. Lautstärken wurden mittels RMS Abgleich und zusätzlicher individueller Kontrolle angeglichen. Als Anker wurde anders als im Standard

beschrieben nicht mit einer 3.5 kHz Variante des jeweiligen Musikstückes gedient. Es erschien nicht sinnvoll ein frequenzverzerrtes Signal als Anker zu benutzen, wenn räumliche Aspekte beurteilt werden sollten. Der Entscheidungsraum sollte somit auf räumliche Aspekte fokussiert werden und so wurde ein Mono-Downmix als Anker benutzt.

Die Einteilung der Musikstücke wurde wiederum wie in Kapitel 4.2.1 vorgenommen. Die vorgenommene Zuordnung zeigt im Anhang.

Der Abhörraum, der für alle Tests benutzt wurde, war ein extra dafür bereitgestellter reflexionsarmer Raum mit 40 cm langen Schaumstoffkeilen an den Wänden und Decke. Die Maße des Raumes betragen 4,90m x 3,50m x 3,40m. Im Raum befindet sich ein Tisch, ein Stuhl, Bildschirm und Tastatur für die Steuerung des Pilothörtests. Die Signale kommen aus einem Rechnersystem, das unter Linux läuft und über eine Soundkarte in SPDIF an einen Wandler (Lake People) die Signale an die Wiedergabestufe leitet. Das Rechnersystem befindet sich in einem separaten Raum. Die Wiedergabe findet über offene elektrostatische Kopfhörer (Lamda Nova) der Firma Stax, die aufgrund ihrer Bauweise immer mit angepasstem Verstärker (SRM-1 MK2) zur Verfügung stehen, statt. Die Lautstärke der Wiedergabe kann somit individuell variiert werden. Die Raumtemperatur wurde konstant über alle Testreihen gehalten. Die Hörer wurden in allen Tests nach ihrer persönlichen Präferenz gegenüber der Referenz befragt. Zusätzlich sollten sie auf der Skala in eine Position gebracht werden, die ihrer Meinung nach dem Grad ihrer Bevorzugung entspricht. Keine weitere Information über die Art der Bearbeitung der Stimuli und dem Ziel des Tests wurde gegeben.

4.3.2.1 Spektrale Entzerrung

Die erste Überlegung war, die dominante Frequenzverzerrung als Parameter soweit als möglich aus den Tests auszuschließen, da es aus der Erfahrung der informellen Vorhörtests als schwer galt räumliche Parameter gegen klangfarbliche abzuwägen und somit zu einer Präferenzbewertung zu gelangen. Dies bedeutete eine Einschränkung der Annahme, dass die Parameter unabhängig voneinander sind. Fakt, der in den Vorexperimenten zu Tage trat war, dass eine Veränderung räumlicher Eigenschaften oftmals eine Veränderung der Klangfarbe mit sich brachte, so dass die Güte der räumlichen Veränderung verschleiert wurde. Aus den Erkenntnissen der Kapitel 4.2.1 und 4.2.2 heraus wurde die Voraussetzung aufgestellt, dass Eigenschaften am System, wenn sie nicht räumlicher Natur sind so nah am Stereo Original sein sollen wie möglich. Für das vorliegende System wurde die spektrale Entzerrung nach dem Prinzip in Abbildung 21 durchgeführt.

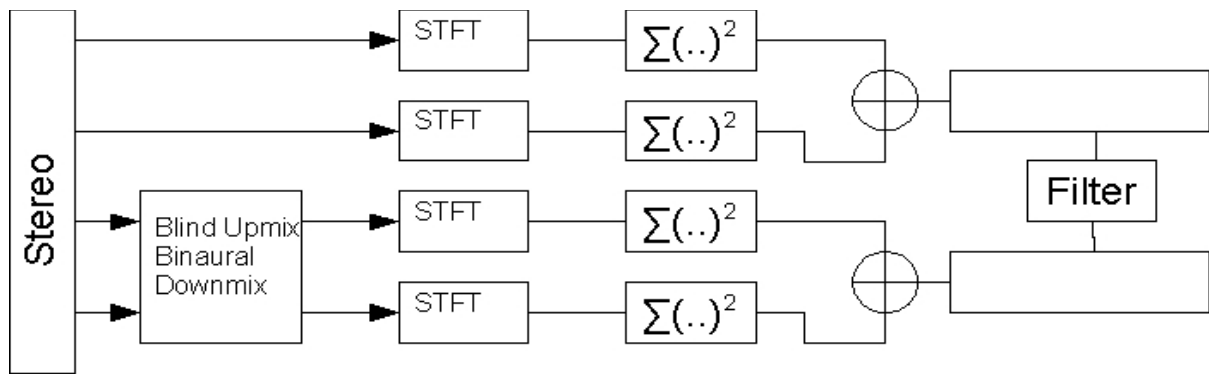


Abbildung 21: Prinzip der spektralen Entzerrung

Diese Art des Eingriffes wurde als nicht ganz unproblematisch betrachtet, war es doch möglich, dass eine Veränderung des Spektrums in den Bereichen die wichtig sind für Richtungswahrnehmung in der Medianebene, die so genannten Richtungsbestimmenden Bänder nach Blauert (Kapitel 2.1.3) entscheidende Veränderungen bewirken. Durch Betrachtung der Betragsfrequenzgänge der im folgenden Test verwendeten Entzerrungsfilter (Abbildung 24) wurde diese Befürchtung bestätigt: Alle Frequenzgänge der Filter zeigen einen charakteristischen Einbruch bei 2-4 kHz was im Vergleich zum Stereo mit einer Betonung dieses Bereichs durch dominante Signalanteile in den Frontkanälen und Faltung mit BRIA aus frontalen Richtungen zu tun hat. Diese Frequenzbetonung wird nun ausgeglichen. Auch der Anstieg im Betragsfrequenzgang zwischen 8 kHz 12 kHz deutet auf Neutralisierung des richtungsbestimmenden Effektes hin. Dennoch zeigen die Betragsfrequenzgänge der Filter, dass es in hohem Maße Frequenzverzerrungen gibt, die ausgeglichen werden müssen. Um die Qualität der Signale durch die zusätzliche Filterung nicht zu beeinträchtigen, wurden die BRIA in ihrer Rohform, also ohne die hauseigene spezielle Diffusfeldentzerrung benutzt. Stattdessen wurde eine auf die Verwendung als Stereoerweiterungsansatz zugeschnittene Art der Entzerrung implementiert mit dem entscheidenden Vorteil, dass neben der Kompensation der Frequenzgänge etwaiger Meßkomponenten auch die Einflüsse des Downmixes von fünf auf zwei Kanäle und die damit einhergehenden Frequenzverzerrungen minimiert werden. Dabei bleiben die Richtungs- und Rauminformationen, die die Lautsprechersignale durch Filterung mit den BRIA aufgeprägt bekommen untereinander erhalten, die räumliche Auflösung bleibt erhalten. Problem, dass nicht außer Acht gelassen wird ist, dass das Frontbild aufgrund der besprochenen Neutralisierung des richtungsbestimmenden Effektes nach oben rutschen könnte. Für die nun verwendete Art der Entzerrung (Abbildung 21)

wurde, im Sinne der späteren Verwendbarkeit, über eine große Anzahl von Stereoverionen verschiedener Musikstücke ein mittleres Periodogramm H_{stereo} nach der Welchmethode erstellt. Über das gleiche Set von Stücken in der bearbeiteten Version wurde ebenfalls ein Periodogramm H_{param} erstellt. Die Länge des Hammingfensters zur Berechnung des Periodogramms beträgt 8192 Punkte, bei doppelter Überlappung. Anschließend wurde die so erhaltene Liniendarstellung des Leistungsdichtespektrums mit einer zur Hälfte überlappenden Terzfilterbank abgetastet und die pro Terzband enthaltenen Energien aufgemittelt. Das Leistungsdichtespektrum wird somit geglättet und ist dadurch einfacher als Filter zu entwickeln. Das Differenzspektrum H_F wurde nach folgender Gleichung ermittelt:

$$H_F = \frac{H_{\text{Stereo}}}{H_{\text{Param}}}$$

Durch die große Anzahl von Mittelungen (28 Musikstücke) wurde davon ausgegangen sich an die dem System eigene Frequenzverzerrungsfunktion anzunähern und diese somit betragsmäßig wirkungsvoll umzukehren. Das resultierende Differenzspektrum wurde mittels Frequenz Abtast Methode mit der MATLAB Funktion *fir2* als linearphasiges 2049 Punkte FIR Filter (2048ter Ordnung) realisiert. Abbildung 22 zeigt beispielhaft die Darstellung des gefilterten Stereo Referenz Periodogramms und des gefilterten Blind-Upmix-Binaural Periodogramms mit dem als FIR Filter entworfenem Differenzspektrum. In Abbildung 23 wird die Filterentwurfskurve und die Differenzkurve dargestellt, um zu zeigen, dass die Auflösung des Filters groß genug ist, eine höchst mögliche Genauigkeit bei der Annäherung an die Differenzkurve zu erreichen.

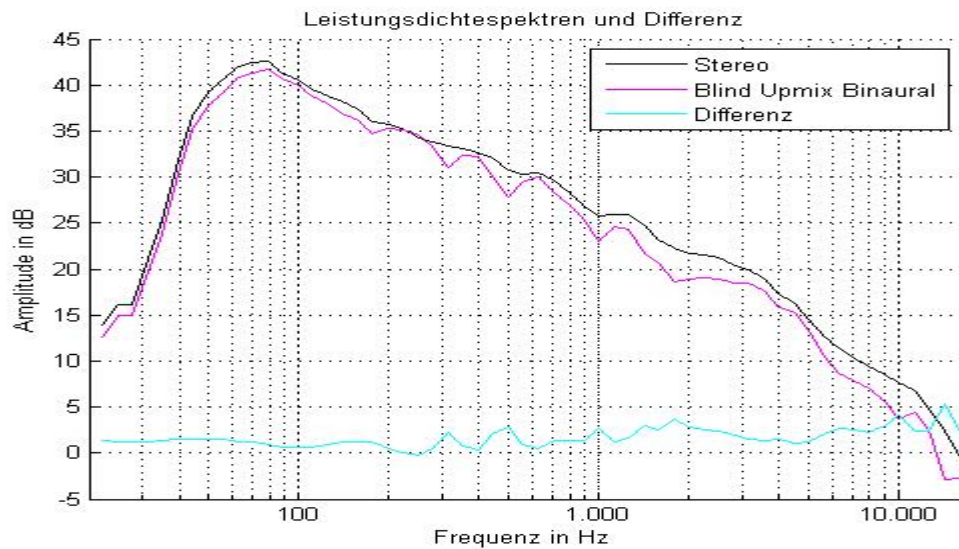


Abbildung 22: Differenzspektrum

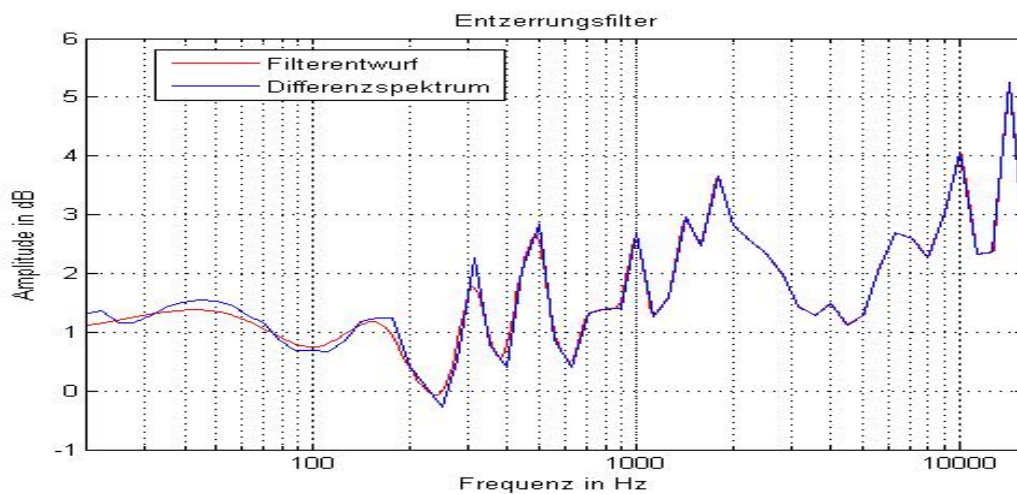


Abbildung 23: Entzerrungsfilter

Abbildung 24 zeigt die Entzerrungsfilter, die verwendet wurden, um im folgenden Test die Anpassung an das Stereo Original vorzunehmen. Für die unterschiedlichen Testsysteme wurde jeweils eine Anpassungsfiler entwickelt nach der beschriebenen Methode.

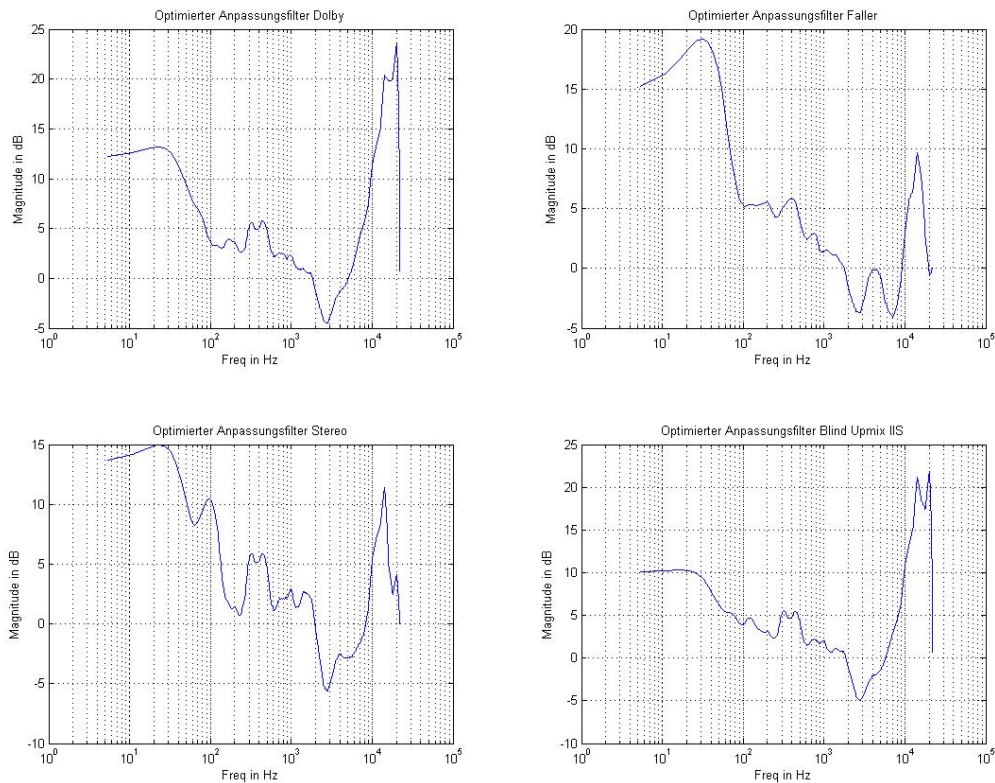


Abbildung 24: Betragfrequenzgang der optimierten Anpassungsfilter

Um zu zeigen, dass diese Art der Entzerrung zielführend und notwendig ist, wurde ein Pilottest aufgesetzt. Der Pilottest war aufgeteilt in 4 Sitzungen, zu 6 Musikstücken und jeweils 4 Konditionen. Die Sitzungen waren aufgeteilt nach verschiedenen Upmixvarianten. Vier Upmixvarianten wurden in den Test aufgenommen:

- Kein Upmix (Stereo Datei wurde binauralisiert)
- Upmix in „Ambient“ Konfiguration wie in Kapitel 3.1 beschrieben
- „In the Band“ Upmix (extern, erweiterter Upmix auf Basis der beschriebenen Technik von Kapitel 3.1)
- Dolby Pro Logic Upmix

Es sollte die Annahme Bestätigung finden, dass eine systemangepasste Entzerrung einer Standardanpassung vorzuziehen ist. Für die Standardanpassung wurde ein denkbar einfaches Upmixsystem benutzt, das folgende Vorschrift für die Berechnung der Kanäle benutzt:

$$\begin{aligned}
L &= L \\
R &= R \\
C &= 0,7 \cdot L + 0,7 \cdot R \\
Ls &= L * TP_{7\text{kHz}} \\
Rs &= R * TP_{7\text{kHz}}
\end{aligned}
\tag{4.1}$$

Diese 5 erzeugten Kanäle des Upmixsystems wurden binaural aufbereitet mit der in Kapitel 3.2 beschriebenen Standardkonfiguration des Ensonidosystems. Einzig der interne EQ wurde ausgeschaltet, in dem die roh gemessenen BRIAs benutzt wurden. Der Filter, der zur Anpassung dieses Systems an die Stereo Originalversion entworfen wurde, fungierte als Filter für die Standardanpassung aller Systeme im Test. Als Anker wurde das jeweilige System im Test ohne jegliche Entzerrung benutzt.

Systeme im Test:

- Unbearbeitete Stereooriginalversion als versteckte Referenz
- Upmix Binauralwiedergabe mit Standard EQ
- Upmix Binauralwiedergabe mit auf spektrale Verfärbung des Systems angepasstem EQ
- Upmix Binauralwiedergabe ohne EQ als Anker

Die Ergebnisse in Abbildung 25 zeigen die 95% Konfidenzintervalle für die Gesamtwertung der 4 Sitzungen mit 7 Personen. Information, die aus diesem Test gezogen werden kann ist folgende: Eine Entzerrung ist auf jeden Fall immer notwendig, das Ankersignal wurde immer als signifikant schlechtestes gewertet. Die angepasste Entzerrung wird im Mittel nie schlechter empfunden als die Standardanpassung. In zwei Fällen unterscheidet sich die systembezogene Anpassung signifikant im positiven Sinne zur Standardanpassung. Der Test wurde aus Aufwandsgründen lediglich mit 7 Personen durchgeführt, wird daher nur benutzt um Tendenzen abzulesen. Als Fazit aus dem Test wurde gezogen, dass für alle Tests in Zukunft die systemangepasste spektrale Entzerrung bezüglich dem Stereo Original durchgeführt wird.

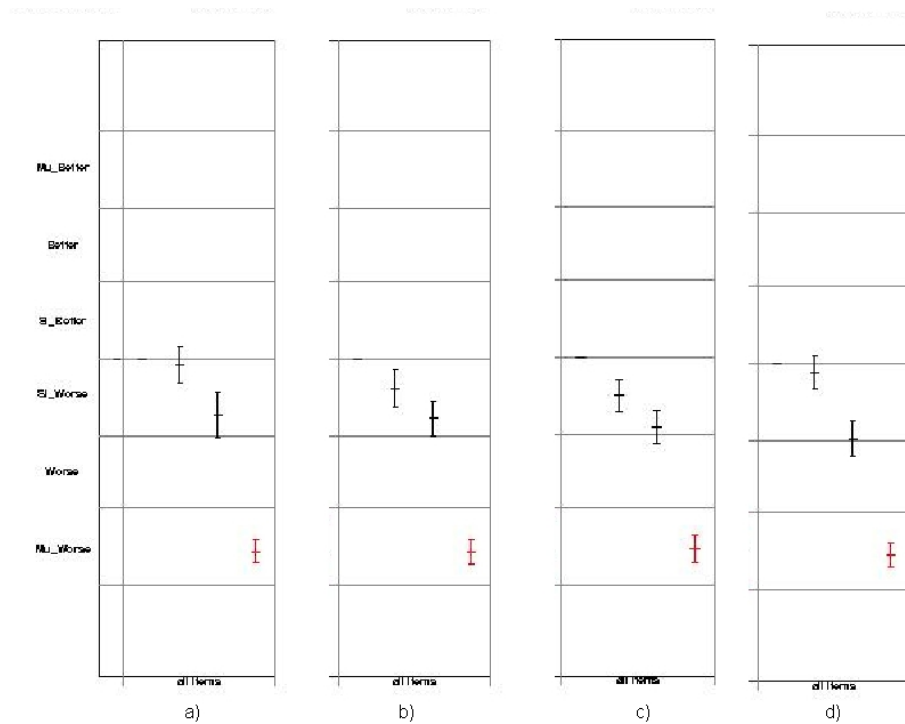


Abbildung 25: 95% Konfidenzintervalle spektrale Entzerrung

- a) Dolby Pro Logic 2 b) „In the Band“ Blind Upmix (extern) c) Stereo
d) „Ambient“ Blind Upmix(Fraunhofer)

4.3.2.2 Upmix Strategien

Wie in Kapitel 4.3.1 gezeigt boten die aus der Implementierung des Upmix (Kapitel 3.1) zur Verfügung gestellte Parameter nur wenig Spielraum um Veränderung in der Wahrnehmung für die binaurale Wiedergabe zu bewirken. Die Möglichkeiten die am Algorithmus gegeben waren führten entweder zum besprochenen Ambient Ansatz, einem „In the Band“ Ansatz, oder einer Trivillösung (Surroundkanäle liefern das selbe Signal wie Frontkanäle) für den Upmix. Nun sollte sich die Frage stellen welche Art des Upmix am meisten Einfluss auf Hörerbevorzugung bei Binauralwiedergabe hat. Diese Frage diente als Anlass zu einem Strategievergleich um bei unveränderter Parametereinstellung am Binauralsystem die Wirkung verschiedener Upmix-Strategien zu bewerten. Dabei wird davon ausgegangen, dass die Upmix-Parameter nicht mit den Binaural-Parametern bezüglich der Hörerpräferenz

interagieren. Die zweite Antwort auf die Frage, ob der Aufwand eine 2-Kanalversion in eine 5-Kanalversion zu konvertieren notwendig ist, um Hörerbevorzugung zu erreichen sollte gefunden werden. Sollte die Antwort zugunsten der Kanalkonvertierung ausfallen, so kann eventuell auch eine Richtung angezeigt werden, die eingeschlagen werden muss. Ist es notwendig Rechenleistung zu investieren für hochqualitative Kanalcodierung, oder genügt es mit wenig Aufwand ein gleichwertiges Ergebnis im Sinne von Hörerbevorzugung zu erreichen. Auch Konkurrenzprodukte wurden in dieser Umschau nicht ausgeschlossen, bei denen Strategien zum Erstellen des Upmix mehr oder weniger bekannt sind. Die Upmixvarianten lauten:

- Blind-Upmix (Fraunhofer) in der Standardkonfiguration (siehe Kapitel 3.1)
- Einfachster Upmix siehe Kapitel 3.2.2.1
- „In the Band“ Upmix (extern, erweiterter Upmix auf Basis der beschriebenen Technik von Kapitel 3.1)
- Kein Upmix, Stereoverversion wird binauralisiert
- Dolby Pro Logic 2 (Standardkonfiguration Music, Centerspread 20°)

Der Test wurde mit 6 Personen, 6 Musikstücken und in einer Sitzung durchgeführt. Ergebnisse zeigen, dass aufwendigere Kanalkonvertierung auf jeden Fall zu einer Hörerbevorzugung führt.

Überraschendes Ergebnis aus dem Test ist, dass der umgebende Upmix weit zurück liegt. In informellen Vorhörttests die von den 4 Stammhörern blind gemacht wurden erhielt dieser Upmix Attribute die ihn als räumlich, mit angenehm warmen, transparenten Klangbild beschreiben. Lokalisationseigenschaften jedoch werden schlechter attestiert als im original Stereo, was zu der Annahme führt, dass die Richtungsempfindung bzw die Veränderung dieser im Vergleich zum original Stereo minimal bleiben muss. Die Annahme gilt unter der nach wie vor wichtigen Voraussetzung, dass der Hörer die Möglichkeit in einem Gerät zwischen Stereo und Upmix-Binauralwiedergabe umzuschalten hat. Interessant ist auch die Bewertungsregion, in denen sich die unterschiedlichen Ansätze aus dem Hause Dolby und Fraunhofer befinden. Diese Varianten des Upmix führen klar das Feld an. Gespräche im Kollegium zeigen, dass dieser Sachverhalt im Hause bekannt ist aus ausgedehnten 5.1 Surround Tests mit Lautsprechern. Die Wege die zu einem ähnlichen Wahrnehmungsziel führen sind sehr unterschiedliche.

Anschließend wurde in kleinerem Stil am hauseigenen Upmix System 3 verschiedene aus informellen Tests des Autors sinnvoll erscheinende Parametereinstellungen getestet. Zu dem im ersten Test bereits bewerteten Lösungen „Ambient“ Ansatz und Trivial Lösung gesellte sich nun ein „In the Band“ Ansatz. Die „In the Band“ Lösung stellte sich als nicht so einfach her zustellen heraus als gedacht, die Stereoverbreiterung führte bei Lautsprecherwiedergabe zu „Löchern“ im Schallfeld, und aufgrund fehlender Dekorrelation zwischen den Kanälen oftmals zu unangenehmen Phaseneffekten bei Bewegungen des Kopfes aus dem Sweetspot. Dennoch wurde für die binaurale Aufbearbeitung des Signals die in Kapitel 3.1 erklärte „In the Band“ Variante benutzt, da keine Bewegung ausserhalb des Sweetspots stattfindet. Ein Test zeigt, dass auch die neue Variante sich von einer Triviallösung abzusetzen vermag. Jedoch zeigt der Test auch, dass wenig Unterschiede zum „Ambient“ Ansatz wahrgenommen werden. Der Test wurde mit 7 Personen durchgeführt.

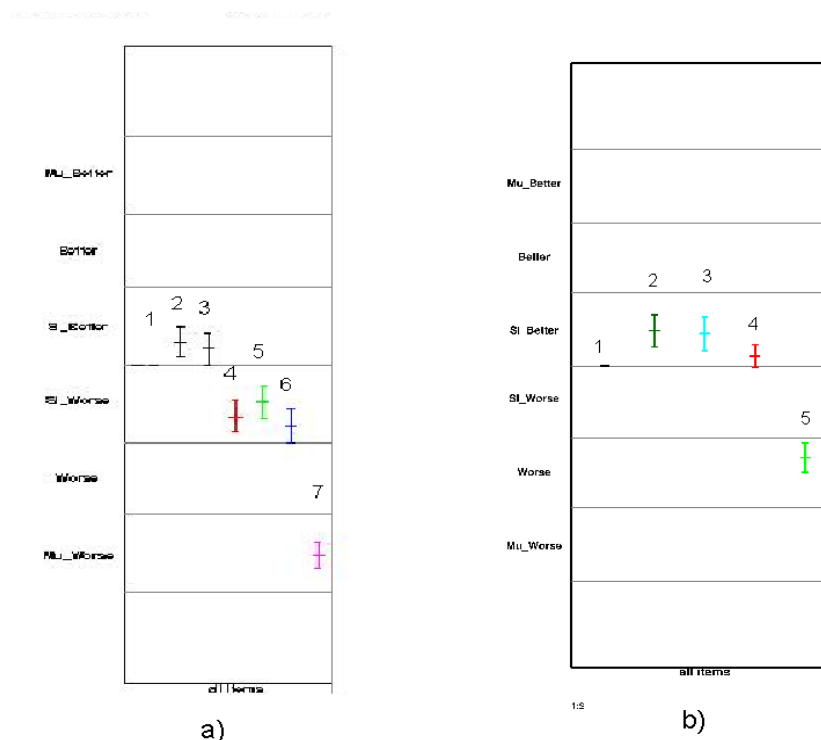


Abbildung 26: 95% Konfidenzintervalle Upmix Strategien

a) mit externen, b) mit internen Algorithmen

a: 1)versteckte Stereo Referenz 2) „Ambient“ Upmix (Fraunhofer) 3)Dolby Pro Logic II
4)Umgebend (extern) 5)Binaural Stereo 6)Triviallösung 7)Mono

b: 1) versteckte Stereo Referenz 2) „Ambient“ Upmix 3) „In the Band“ Upmix 4)Triviallösung
5) Mono

4.3.2.3 Filterlängen

In den informellen Hörtests aus Kapitel 4.3.1 zeigte sich die Länge der verwendeten BRIA als wirkungsvoller Parameter, um die Wahrnehmung des virtuellen Raumes zu beeinflussen. Alle wichtigen Elemente, die der Wiedergabe Information über den virtuellen Raum aufprägen sind in ihnen enthalten. Aus Kapitel 3.2.1 ist bekannt, dass die Impulsantwort zu kurz ist um die ganze Nachhallzeit zu enthalten. Bereits Berechnungen der Nachhallzeit über den T_{30} Wert durch die Schröder Integration sind nichtmehr möglich mit dieser Impulsantwort. Das führt zu der Annahme, dass der Raumklang durch den Fensterungseffekt einerseits unnatürlich klingen könnte, da noch zuviel Energie am Zeitpunkt des Abschneidens enthalten ist, beziehungsweise die Fensterung zu Filtereffekten im Frequenzbereich führen. In einem weiteren informellen Vorversuch werden daher die Raumimpulsantworten $h(t)$ mit Exponentialkurven $w(t)$ mit variierter Abklingkonstante τ im Zeitbereich multipliziert um das Abklingen der Raumimpulsantwort zu beschleunigen und so das Realitätsempfinden der vorhandenen Impulsantworten zu steigern.

$$h'(t) = h(t) \cdot w(t)$$

$$w(t) = w_0 \cdot e^{-\frac{t}{\tau}}$$

Die erhoffte Wirkung ist den als störend empfundenen „Badezimmereffekt“ (Kapitel 4.2.1) zu mindern. Dieser Vorgang sollte eine erhöhte Absorption des virtuellen Raumes simulieren. Zur mathematischen Darstellung des zusätzlichen Absorptionseffektes wird das Abklingverhalten der BRIA Energie mit Formel 2.3 dargestellt und zusätzlich der Faktor $e^{2(-\tau t)}$ für die Energie der Abklingkurve eingeführt (ohne Berücksichtigung der Luft-Dissipation).

$$E(t) = E_0 e^{\left[\frac{S \cdot c}{4V} \cdot \ln(1-\bar{\alpha})\right] \cdot t} \cdot e^{2(-\tau t)}$$

Daraus ergibt sich für die veränderte Nachhallzeit:

$$T_N = \frac{\ln(10^{-6})}{-\frac{S \cdot c}{4 \cdot V} \cdot \ln(1-\bar{\alpha}) - 2 \cdot \tau}$$

Diese Versionen die den natürlichen Verlauf eines Schallfeldes annähern, wurden verglichen mit Impulsantworten, deren reduzierte Länge erhalten wurde durch Rechteck Fensterung. Als Ausgangspunkt galt dabei, dass die Fensterlänge des Rechteckes gleich der Abklingkonstante τ der Abklingkurve ist. Die Frage stellte sich, ob sich die aus der Signalverarbeitung

bekannten Effekte (Verschmierung von Spektralkomponenten durch den *Leakage Effect*, Glättung des Spektrums) unterschiedlich auf die Wahrnehmung auswirken.

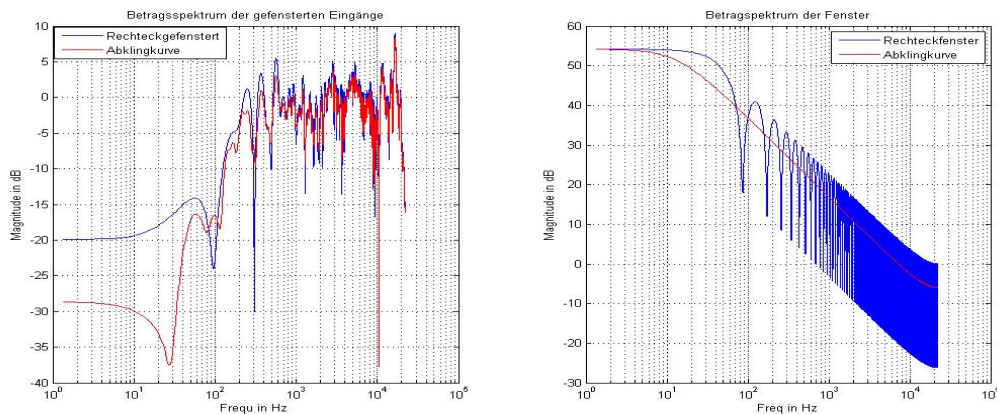


Abbildung 27: Spektrale Verläufe der Fensterfunktionen

Aufgrund in informellen Hörtests mit 4 Stammhörern festgestellter geringer Wahrnehmungsunterschiede zwischen Rechteck gefensterter BRIR und den äquivalenten exponentiellen Abklingkurven wurden im weiteren die Rechteck gefensterter Versionen benutzt. Eine Erklärung dafür ist, dass etwaige Klangfarbenunterschiede die entstehen könnten durch die immer benutzte spektrale Entzerrung ausgeglichen. Im Hinblick auf sparsame Implementierbarkeit wurden die Impulsantworten also mit einem Rechteckfenster auf Längen gekürzt, die sich auf Prozessoren gut in *Look Up Tables* speichern lassen, also die Länge in 2^n er Potenzen variiert. Verglichen wurden folgende Längen:

- 64 Samples (= 1,5ms entspricht nicht ganz der Länge der HRIR ohne Raumanteil)
- 512 Samples (= 11ms enthalten sind frühe Reflektionen von Boden und Decke)
- 1024 Samples (= 22 ms)
- 4096 Samples (Volle Länge der Impulsantwort)

Ein kurzer Pilottest im voraus zeigte, dass diese Einteilung sinnvoll war, da bei zuerst zusätzlich eingefügten Zwischenschritten es den teilnehmenden 5 Hörern nicht möglich war konsistente Unterscheidungen zu treffen. Dieser Fakt war überraschend, da zum Beispiel der Schnitt bei 1024 Punkten aus dem Grund gewählt wurde, weil kurz dahinter für die vorderen Lautsprecher Richtungen eine letzte ausgeprägte Reflexion zu erkennen ist. So wurde vermutet, dass es zu deutlichen Wahrnehmungsunterschieden im Vergleich zur Version mit 2048 Punkten kommt, was nicht der Fall war.

Der Test reduzierte wurde mit 8 Personen und 5 Musikstücken durchgeführt, wobei wiederum nur eine Präferenzentscheidung von Interesse war. Die Ergebnisse zeigen deutliche Tendenzen der Präferenzbewertung von 64 Punkten nach oben bis zur Länge von 1024 Punkten der BRIA bevor Stagnation bei 4096 Punkten eintritt.

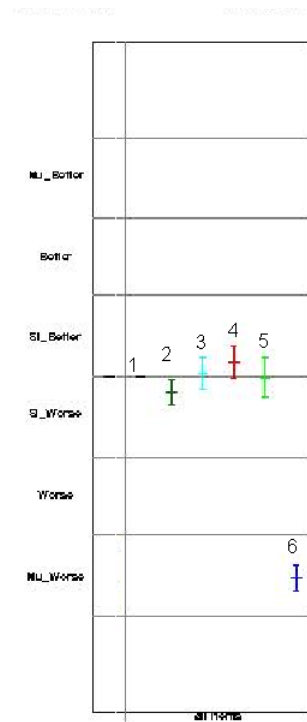


Abbildung 28: 95% Konfidenzintervalle BRIA Längen

- 1) Stereo Referenz 2)64 Punkte 3)512 Punkte
4)1024 Punkte 5)4096 Punkte 6)Mono

4.3.2.4 Virtuelle Lautsprecher Position

Dem Problem der verringerten Stereobreite gegenüber der original Stereo Version wurde mit zwei verschiedenen Strategien begegnet: Die erste Idee basiert darauf, dass reduzierte Breite einer Schallquelle durch das bei der binauralen Wiedergabe zusätzlich eingefügte Kanalübersprechen zwischen ipsi- und kontralateralem Ohr verursacht wird. Kanalübersprechen bei BRIA für einen Einfallswinkel von 30° ist relativ ausgeprägt, ILD ist gering. Somit schien es zielführend zu sein, die kontralaterale Impulsantwort zu dämpfen, wodurch sich die ILD erhöht, um so eine Aufweitung des wahrgenommenen Stereobilds zu erreichen. Der Vollständigkeit der Modellierung wegen müsste man ebenfalls die ITD verändern und schließlich die spektrale Färbung durch den Filtereffekt des Außenohres und

Torso. Das Ergebnis ist die zweite Idee: Aus dem vorhandenen Set aus gemessenen BRIA werden andere virtuelle Lautsprecherpositionen ausgewählt. Einziges Problem das sich stellte war der Fakt, dass bei den Messungen der BRIA die Position der Lautsprecher nicht verändert wurde und die BRIA nur im statischen 5.0 Setup durch Kopfdrehung gemessen wurden. Das bedeutet, dass unter Einbeziehung des Raumes für eine andere virtuelle Position als 30° keine korrekten BRIA vorliegen. Natürlich gab es Kopfpositionen, an denen der Azimutwinkel zwischen Blickrichtung Lautsprecherrichtung zum Beispiel 90° betrug (Beispiel siehe Abbildung 29), die offensichtliche Problematik ist, dass ab dem Zeitpunkt des Eintreffens der ersten Reflexionen keine authentische Rauminformation anhängt. Eine Arbeit der Universität Aalborg [31] beschäftigte sich unter anderem mit diesem Thema um Ressourcen bei der Implementierung von Kopfdrehungen in einem binauralen Wiedergabesystem mit Rauminformation zu sparen. Dabei wurde festgestellt, dass es ausreichend ist die HRIR aus einer BRIA bei einer Kopfdrehung zu aktualisieren. Es ist nicht notwendig für jede Drehrichtung eigens Rauminformation zu speichern um hohe Lokalisation zu erreichen. Die Frage die diese Studie beantworten sollte war, ob die korrekte Richtungswahrnehmung bei Kopfhörerwiedergabe auch mit nicht originaler Rauminformation bei den Hörern ausgelöst wird. Indem Lautsprecher (nur visuell verfügbar) benannt werden sollten, denen der Hörer ein Klangerlebnis, dass er mit Kopfhörer hört, zuordnen soll wurde die Güte der Annäherung bestimmt. Es wird jedoch keine subjektive Bewertung der Präferenz des Gehörten gegeben, wie es für diese Arbeit von Interesse wäre. Dennoch wurden aufbauend auf diesem Ergebnis benötigte virtuelle Lautsprecherpositionen einfach nach dem in Abbildung 29 gezeigten System aus dem gemessenen Set entnommen. Zusätzliche Optimierung, wie etwa das Anfügen der korrekten Rauminformation wurde nicht durchgeführt. Möglich ist, wenn auch unwahrscheinlich, dass durch diese Anpassung ein natürlicheres Klangbild erzeugt werden kann.

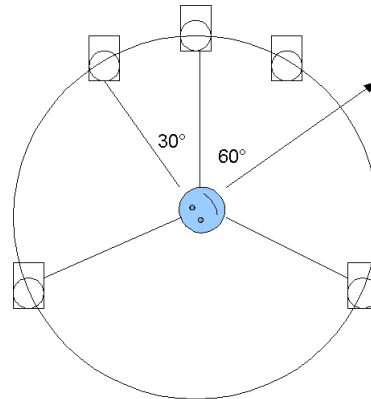


Abbildung 29: 90 ° BRIA

Zunächst wurde die Manipulation der kontralateralen Impulsantwort getestet. Die Dämpfung der kontralateralen Impulsantwort wurde im Test mit 3 Parametereinstellungen gerüft:

- Dämpfung der kontralateralen Impulsantwort um 6dB
- Verstärkung der kontralateralen Impulsantwort um 6dB
- Keine Änderungen

Die Ergebnisse (Abbildung 30) zeigen, dass für alle Musikstücke die unbehandelte Version tendenziell bevorzugt wird. Zwar lässt sich aus den Ergebnissen ebenfalls ablesen, dass Dämpfung gegenüber Verstärkung der Vorzug gegeben wird, dennoch sind die Ergebnisse eindeutig. Attribute, die den bearbeiteten Versionen zugeschrieben werden sind unter dem Oberbegriff Unnatürlichkeit zusammenfassbar. Der gewünschte Effekt, dass Stereoverbreiterung im Falle der Dämpfung eintritt ist wenn überhaupt nur sekundär wahrnehmbar, auf jeden Fall dominieren andere Effekte.

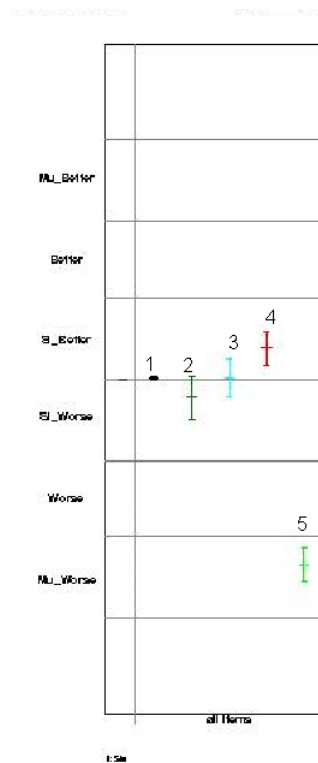


Abbildung 30: 95% Konfidenzintervalle Dämpfung der kontralateralen BRIA

- | | |
|--------------------|---------------------|
| 1) Stereo Referenz | 2) Verstärkung 6 dB |
| 3) Dämpfung 6dB | 4) Keine Änderung |
| | 5) Mono |

Anders für die erweiterte Variante des Tests. Hier wurden jetzt nach oben erklärtem Verfahren verschiedene virtuelle Lautsprecherpositionen ausgewählt und getestet.

- Vordere Kanäle Stereo Basis: 12°
- Vordere Kanäle Stereo Basis: 30° (Standardkonfiguration)
- Vordere Kanäle Stereo Basis: 62°
- Vordere Kanäle Stereo Basis: 89°

Die Ergebnisse zeigen tendenziell Vorlieben für breitere Frontlautsprecher Aufstellungen. Aus dem Test wird angenommen, dass über die 89° in der Horizontalebene hinaus keine positivere Auswirkung mehr zu messen sein wird. Interessant ist hier zum ersten mal auch die Betrachtung der Musikstücke, bei denen am meisten Variation in den Antworten steckt. So erreichen Stücke, die in der Stereo Version der Einteilung nach dem Signalmodell der Intensitäts Positionierung (Formel 3.1) in hohem Maße entsprechen (z.B. Tied & Tickled

Trio, Tom Waits) eindeutiger Wertungen als Stücke, in denen sich räumliche oder dekorrelierte Effekte (Queens of the Stoneage, Sarah McLachlan) hörbar machen.

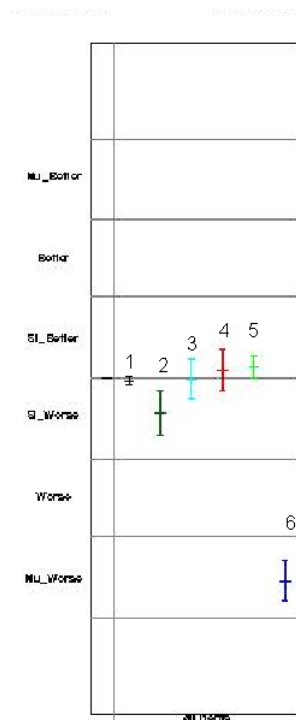


Abbildung 31: 95% Konfidenzintervall virtuelle Lautsprecherposition

1) Stereo Referenz

2) 12°

3) 30° 4) 62° 5) 89° 6) Mono

4.3.2.5 M-S Dekorrelation der hinteren Kanäle des „In the Band“ Upmix Ansatzes

Aufgrund bisheriger fehlender Dekorrelation der hinteren Kanäle vorwiegen beim „In the Band“ Ansatz des Upmix wurde, inspiriert durch die Veröffentlichung von Martens [1], ein Dekorrelations Verbreiterungs Algorithmus für die hinteren Kanäle in Erwägung gezogen. Aufgrund der Erzeugung der Signale für die hinteren Kanäle beim „Ambient“ Ansatz (Kapitel 3.1) machte diese Vorgehensweise nur Sinn für den „In the Band“ Ansatz. Bestätigt wurde die Annahme durch informelle Hörtests, keine Änderung bei Versionen mit „Ambient“ Ansatz war hörbar. Da aus der Idee wie der „In the Band“ Ansatz funktioniert das Problem hervortritt, dass ein Signal in mindestens 3 Lautsprechern wiedergegeben wird, stellt sich die Frage nach der großen Ähnlichkeit der Signale in den Kanälen. Als angemessene, weil sehr einfache, Lösung bot sich so eine Erhöhung der Dekorrelation in den hinteren Kanälen mittels Seitenkanal Verstärkung an. Das Frontbild würde erhalten bleiben, ebenso Quellen, die

aufgrund des Positionsindex verstärkt in hintere Kanäle geraten würden. Zusätzlich würde aber erhöhtes Gefühl der Einhüllung erzeugt werden, weil weniger korrelierte Information zwischen den hinteren Kanälen und zwischen den hinteren und vorderen Kanälen betont würde. Die Methode die Martens vorschlägt basiert auf einer Singularwertzerlegung des Stereosignals, die auch als M-S Decoder betrachtet werden kann. Ein in Matrixschreibweise vorliegendes Signal A kann wie folgt zerlegt werden:

$$A = QSP^T$$

Im Falle einer Stereoaufnahme wird das 2-Kanalsignal zerlegt in zwei Komponenten, die in Form von orthonormalen Eigenvektoren in der Matrix P stehen, ein Mittensignal und ein Seitensignal. Die Matrix S enthält die Singularwerte (=Wurzel der zugehörigen Eigenwerte) der Matrix A, die Verstärkungsfaktoren für die jeweilige Komponente. Martens benutzt den Algorithmus und vergrößert die Seiteninformation auf den Pegel der Mitteninformation (Abbildung 32). Dies geschieht indem die sparsame Version der Singularwertzerlegung (SVD) berechnet wird[32], die Singularwerte angleicht und daraus wieder ein Stereosignal durch Rückberechnung der veränderten Matrixgleichung macht.

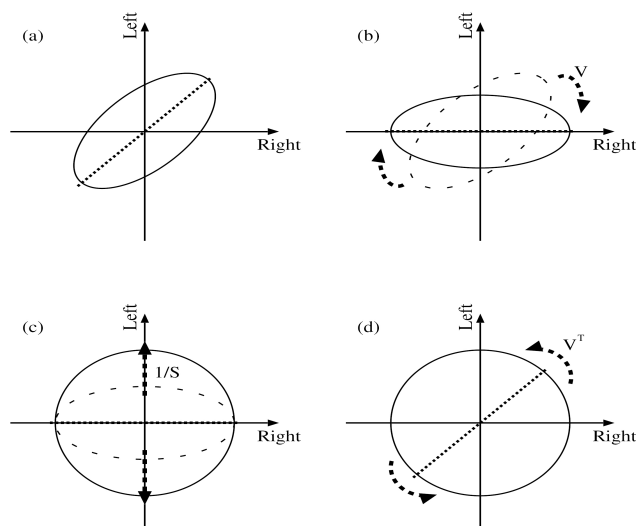


Abbildung 32: Funktionsweise der SVD

- Keine Änderung

- Seitenkanal aus
- Seitenkanal um 6dB verstärkt
- Seitenkanal um 12 dB verstärkt

Die Ergebnisse (Abbildung 33) zeigten leider nicht den erwarteten Effekt. Die Varianz in den Antworten wird hauptsächlich von anderen Effekten hervorgerufen, nicht jedoch von den verschiedenen Algorithmen. Weiteres Problem war das der Test nicht nach der anfänglich aufgestellten Einteilung der Musikstücke vorgenommen wurde. Bei Betrachtung der Ergebnisse fällt auf, dass alle verwendeten Stücke sehr deutlich mit zeitlichen (weil Liveaufnahme eines Orchesters) Positionierungen, oder künstlichen Raumeffekten versehen waren, und somit kein repräsentativer Schnitt durch die Aufnahmen betrachtet wurde. Dem Autor jedoch scheint dieser oder ein ähnlicher Ansatz zur Erhöhung der Dekorrelation in den hinteren Kanälen für den „In the Band“ Ansatz als sinnvoll und als Möglichkeit diesen zu verbessern. Dazu laufen am Hause parallel noch andere Arbeiten. Aus Überschneidungs- und Zeitgründen wurde an dieser Stelle nicht weiter vertieft geforscht.

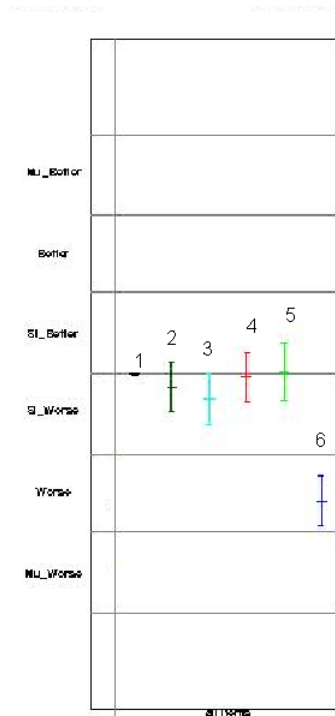


Abbildung 33: 95% Konfidenzintervalle MS-Dekorrelation der hinteren Kanäle

- | | |
|---------------------|---------------------|
| 1) Stereo Referenz | 2) Keine Änderung |
| 3) Nur Monoanteil | 4) 1x Mono 2x Seite |
| 5) 1x Mono 4x Seite | |

4.3.3 Kombinationstest

In den letzten Kapiteln wurden, die aus informellen Hörtests am interessantesten erscheinenden, physikalischen Parameter des Blind Upmix Binaural Wiedergabesystems in Pilothörtests auf ihre Wirkung getestet. Ausschlaggebendes Kriterium war die subjektive Bevorzugung. Entscheidend und wichtig ist nun die Bewertung der Kombination aus verschiedenen Parametern. Bewirken die jeweiligen Parameterstufen ein für den Hörer unterscheidbares Klangbild? Ist seine Präferenzbewertung zu beeinflussen durch Kombination der ausschlaggebenden Einzelparameter? Ein Versuch dieser Art nimmt an Ausmaßen zu, je mehr Parameter in ihren Ausprägungen vorkommen. Zusätzlich sollte das Feld an Hörern und Musikstücken erhöht werden, um in hohem Maße allgemeingültige Aussagen treffen zu können. Zu testende Parameter wurden ausgewählt:

- Upmix
 - „Ambient“

- „In the Band“
- Filterlänge
 - 1024
 - 4096
- Virtuelle Lautsprecherposition
 - 89°
 - 62°

Daraus resultierten 8 mögliche Kombinationen. In einem Test mit versteckter Referenz und Anker würde dies eine Anzahl von 10 Konditionen bedeuten. Zusätzlich wurden aus der vorgenommenen Kategorisierung für Musikstücke 10 ausgewählt. Dies bedeutet pro Musikstück alle Konditionen untereinander bezüglich Bevorzugung zu bewerten, womit letztendlich jeder Hörer 100 Antworten produziert. Dies stellt ein großes Problem für die Testteilnehmer dar: Anker und Referenz nicht betrachtet muss für 8 Konditionen eine Meinung über Bevorzugung gebildet werden. Ermüdung oder Ärger über das Testdesign könnten Ergebnisse verfälschen. Die Lösung ist die Teilung des Tests. Zunächst wurde überlegt nach Konditionen zu teilen.

Die Verwendung eines *Latin Square Designs*⁵ wurde verworfen, da das *Latin Square Design* die Unabhängigkeit der Faktoren im Test voraussetzt und somit keine Interaktionstests zulässt. Die einzige akzeptable Lösung für dieses Problem ist, dass alle Konditionen mit allen verglichen werden. So wurde in 2 Tests mit jeweils 5 Musikstücken geteilt. Um Abhängigkeit von der Tagesform der Hörer zu minimieren, wurden verschiedene Gruppen gebildet, in denen verschiedene Gruppierungen der Musikstücke vorkamen. Dieser Fall der Teilung ist am wenigsten problematisch, da der Einfluss der Musikstückwahl aus den Pilottests und Vorhörttests zwar als wichtig angesehen wurde, jedoch nicht die Stücke untereinander verglichen werden sollten. Bei der Gruppierung der Musikstücke wurde darauf geachtet, dass aus der anfänglich getroffenen Einteilung (Kapitel 4.2.1) aus jeder Gruppe mindestens 1 Stück dabei war. So wird der Tagesformeffekt der Teilung minimiert durch Zufallsverteilung. Am Test nahmen 12 Hörerprobt Personen Teil, einige davon bereits vertraut mit der

5 Festlegung der Abfolge von Konditionen pro Hörer für geblockte Experimente. Anzahl der Hörer n muss gleich der Anzahl der Musikstücke n , muss gleich der Anzahl der Algorithmen n sein, und n muss gerade sein dann gilt die Bildungsregel:

1. Hörer: MUSK1*ALG1, MUSK2*ALG2, MUSK3*ALGn, MUSK4*ALG3, MUSK5*ALGn-1....
2. Hörer: Addiere 1 zur Nummer des ALG der 1. Reihe und rechne modulo n und so weiter..

Testprozedur aus den Vorversuchen. Jeder Hörer wurde ermutigt im Falle von Überanstrengung den Test zu unterbrechen und zu Pausieren.

Die Interpretation der Ergebnisse wurde mit der gewohnten graphischen Darstellung der 95% Konfidenzintervalle durchgeführt. Zusätzlich wurden die Ergebnisse mit ANOVA und T-TEST auf Interaktionen, Bestimmtheit und Signifikanzniveaus überprüft. Tabelle 4.2 zeigt die Ergebnisse paarweiser T-Tests der Algorithmen gegenüber der Stereoreferenz um zu prüfen, ob sich die Mittelwerte der Antwortverteilungen bei 95% Signifikanzniveau unterscheiden. Farblich hinterlegte Felder zeigen signifikante Unterschiede an, wobei rot in negativem und türkis in positivem Sinne bedeutet.

	1024_62_U	1024_89_U	4096_62_U	4096_89_U	1024_62_A	1024_89_A	4096_62_A	4096_89_A	Anker
Alle Stücke	0,738	0,491	0,173	0,453	0,044	0,831	0,012	0,676	0,000
Beatles	0,188	0,456	0,977	0,754	0,366	0,247	0,695	0,549	0,000
CHILLI	0,452	0,007	0,031	0,033	0,056	0,044	0,403	0,214	0,000
Comidian	0,104	0,033	0,983	0,001	0,086	0,300	0,732	0,061	0,000
Hermanos_ Patchekos	0,252	0,121	0,209	0,259	0,965	0,152	0,611	0,639	0,000
HipHop	0,013	0,011	0,014	0,014	0,111	0,041	0,007	0,160	0,000
Queens	0,927	0,036	0,840	0,115	0,687	0,805	0,810	0,973	0,000
babayaga	0,515	0,551	0,470	0,908	0,002	0,159	0,034	0,114	0,000
carmina	0,837	0,229	0,757	0,746	0,661	0,838	0,291	0,551	0,000
dance	0,129	0,885	0,455	0,258	0,107	0,825	0,052	0,554	0,000
opus	0,980	0,660	0,584	0,878	0,586	0,364	0,719	0,083	0,000

Tabelle 4.2: T-Test für Faktor Algorithmus, 95% Signifikanzniveau

Um mehr Aussagekraft in diese Statistik zu legen, wurde zunächst auf Normalverteilung der gegebenen Bewertungen mit dem Kolmogorow-Smirnow-Test (KS-Test) geprüft. Normalverteilung ist die Voraussetzung für die Anwendbarkeit des T-Test, wenngleich in der Literatur auf die Robustheit gegenüber der Verletzung dieser Voraussetzung hingewiesen wird [20] [11]. Auf den vorliegenden Datensatz angewendet, bedeutet dies in hohem Maße nicht normalverteilte Gruppen. Die große Variabilität unter den Hörerbewertungen durch die unterschiedliche Benutzung der Skala ließ das Bild nicht deutlich erscheinen. Somit wurde ein Normierungsverfahren für den Faktor Hörer angewandt [33]. Das Verfahren ist ein sehr einfaches und basiert darauf, dass die Verteilung aller Bewertungen pro Hörer auf den gleichen Mittelwert 0 mit der gleichen Standardabweichung 1 normiert werden. Das Verfahren wird z-Normierung genannt, der Datensatz wird dadurch standardisiert. Der Effekt dabei ist, dass der Faktor Hörer aus dem statistischen Modell verschwindet, da keine Varianz zwischen den Höermittelwerten mehr vorliegt. Erneute Überprüfung nach der Voraussetzung zeigt, dass Normalverteilung in 13 von 20 Fällen für die Faktoren Musikstück (50%) und

Algorithmus (80%) vorliegt. Mit dem Levene Test (Implementierung in MATLAB nach [34]) wurde die Voraussetzung für die Varianzanalyse geprüft: Homogenität der Varianzen zwischen den Gruppen. Varianzhomogenität liegt vor für den Faktor Algorithmus, nicht jedoch für den Faktor Musik. Aussagen aus dem statistischen Modell über den Faktor Algorithmus haben also ein hohes Maß an Gültigkeit.

Ein Varianzanalyse Modell (ANOVA, fixed) für die Faktoren Musikstück und Algorithmus mit Interaktion wurde am standardisierten Datensatz durchgeführt um weitere Anhaltspunkte für die Interpretation der Ergebnisse zu erhalten:

ANOVA							
Gruppe	SumSquare (Type III)	df	Meansquare	f	p	Bestimmtheit	
Algorithmus	576,4583496	9	64,0509277	152,31746	0,000	0,4852343	
Musik	60,69987631	9	6,7444307	16,0387146	0,000	0,05109417	
Algor.*Musik	88,28140385	81	1,08989387	2,59184171	0,000	0,07431095	
Residuum	462,5603703	1100	0,42050943				
Total	1188	1199					

Tabelle 4.3: ANOVA 2-faktoriell

Beide Faktoren und Interaktionen zeigen signifikante Effekte, das bedeutet, dass die Nullhypothese (die Streuung der Mittelwerte der Gruppen eines Faktors um den Gesamtmittelwert ist kleiner als die Streuung der Werte innerhalb der Gruppen um ihren Mittelwert) bei einem Restrisiko von 5% verworfen werden kann. Konkret bedeutet das, dass Unterschiede zwischen den verschiedenen Gruppen eines Faktors vorliegen, sich zum Beispiel die Bewertungen der Hörer für mindestens 2 betrachtete Algorithmen signifikant voneinander unterscheiden. Der Faktor Algorithmus ist zusätzlich dominant. Das Bestimmtheitsmaß zeigt an, dass 48,5% der Gesamtvariation durch die Unterschiede der Gruppen des Faktors Algorithmus verursacht wird. Wobei nur 5 % der Gesamtvariation durch den Faktor Musik und 7 % durch die Interaktion von beiden Faktoren verursacht wird. Interessant ist, welcher der Algorithmen sich von der Stereoreferenz und in welcher Weise absetzt, und welche der Musikstücke zu positiven, und welche zu tendenziell negativen Bewertungen führen. Zum Vergleich der Algorithmusgruppen untereinander werden paarweise T-Tests mit Berücksichtigung der HSD⁶ benutzt (Abbildung 34). Deutlich zu erkennen ist, dass es keine Wertungen über der Stereowertung gibt (im Gegensatz zu unstandardisiertem Datensatz

6 Beim Vergleich mehrerer Gruppen untereinander muss, um das Signifikanzniveau α des ANOVA-Modells beizubehalten, das Signifikanzniveau jedes einzelnen Tests angepasst werden. Tukey's 'honestly significant difference' (HSD) berechnet die kritische Differenz zweier Verteilungen, die bei n Paarvergleichen und einem Signifikanzniveau α nicht überschritten werden darf.

$$D_{Tukey} = q_{\alpha} \times \sqrt{\frac{MS_R}{n}}$$

q_{α} kann aus den 'Studentisierten Spannweiten'- Tabellen abgelesen werden und ist Abhängig von der Anzahl der Paarvergleiche, dem Freiheitsgrad des Fehlers (MS_R) und α . n ist die Anzahl der Paarvergleiche [16]

Tabelle 4.2). Ebenfalls in deutlicher Ausprägung zu erkennen ist, dass der Parameter virtuelle Lautsprecheranstellung zu Präferenzunterschieden führt. So fallen 3 von 4 Algorithmen in denen dieser Parameter mit der Einstellung 62° vorkommt als signifikant schlechter als die Stereoreferenz auf. 2 dieser 3 Algorithmen ist in Kombination mit dem Parameter Filterlänge 4096 Punkte. Eine Tendenz lässt sich ableiten, wenn auch in keinster Weise signifikant, für die Art des Upmix: Alle Algorithmen, die auf den „In the Band“-Upmix aufbauen sind in ihrer höchsten Bewertung immer unterhalb der Algorithmen mit „Ambient“-Ansatz.

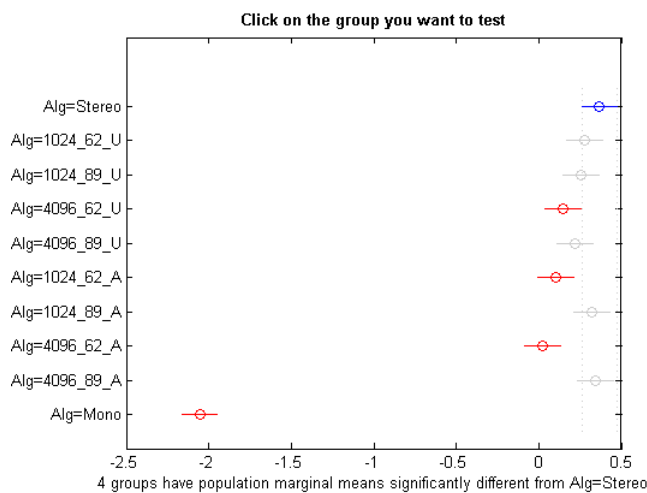


Abbildung 34: Paarvergleich der Gruppen des Faktor Algorithmus

Die Auswertung nach Musikstücken (Abbildung 35) zeigt die positivsten Auswirkungen für Stücke, die in hohem Maße dem vorausgesetzten Signalmodell des Blind Upmix (Formel 3.1) entsprechen: Stücke ohne hervortretende räumliche Effekte, mit klar definierten Positionierung von Instrumenten (Comedian Harmonists, Eve feat. Missy Elliot) die mit hoher Wahrscheinlichkeit durch Amplituden Positionierung produziert wurden. Als klarer Verlierer steht das Stück von Red Hot Chili Peppers da, das durch sehr diffuse Positionierung einer dominanten Gitarrenstimme (vermutlich durch starke dekorrelierte Anteile in den Kanälen des Stereo Originals) durch die Bearbeitung klar an Qualität verliert. Ähnliche Ergebnisse liefert die Betrachtung der Aufnahme von Mussorgskys „Bilder einer Ausstellung-Die Hütte der Babayaga“. Bei Analyse der Stereoaufnahme zeigt der Korrelationsfaktor während der Wiedergabe Werte sehr nahe an der 0 an, also auf keinen Fall eine Aufnahme die sich in hohem Maße auf Amplitudenpositionierung stützt. Aufgrund der geringen Korrelation wird beim Blind Upmix ein Großteil der Direktsignalanteile als Ambient kategorisiert und gelangt somit auch in die Surroundkanäle. Das Ergebnis ist in hohem Maße Korrelation zwischen

Front- und Surroundkanälen. Dies könnte für den binauralen Downmix bedeuten dass sich unerwünschte Überlagerungseffekte ergeben bei der Addition der Kanäle.

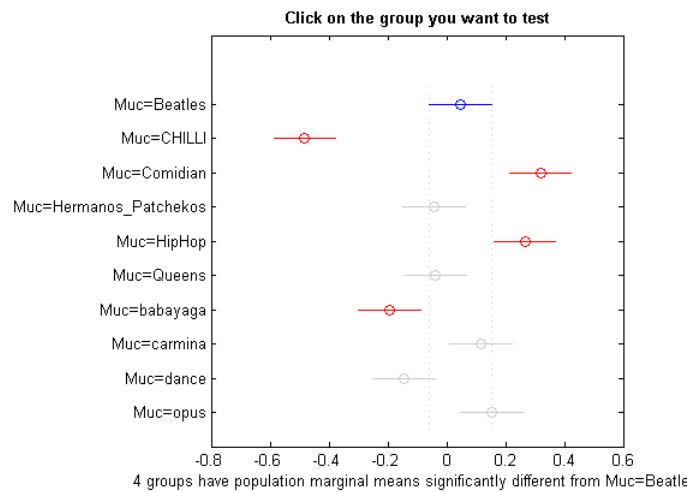


Abbildung 35: Paarvergleich Gruppen des Faktors Musikstücke

Es wurde der Test in Teilen wiederholt, so dass aussichtsreichste Kandidaten erneut getestet wurden. Effekt dieses Tests war zusätzlich die Prüfung, ob in einem erneuten Vergleich Versionen von Algorithmen gleiche Bewertungen bekamen, und ob das Verhältnis untereinander stimmt. Konsistenz einzelner Hörer konnte so überprüft werden. Abbildung 37 zeigt zunächst eine Auswertung auf der Basis von 4 Hörern, die beide Tests durchlaufen haben. Der Datensatz wurde z-normiert und auf Normalverteilung und Varianzhomogenität überprüft. Aufgrund mangelnder Normalverteilung werden die Verteilungen als Boxplot beobachtet und die darauf erkennbaren *Outliner* genauer untersucht. Sie stammen zu 80 % von einem Hörer, der aufgrund extremer Bewertungsdifferenzen zwischen beiden Tests aus der statistischen Betrachtung ausgeschlossen wird. An der Signifikanz ändert dies nur wenig, jedoch ist die Bewertung mittels T-Test und ANOVA aussagekräftiger, da nun Normalverteilung für alle Algorithmen außer Stereo vorliegt.

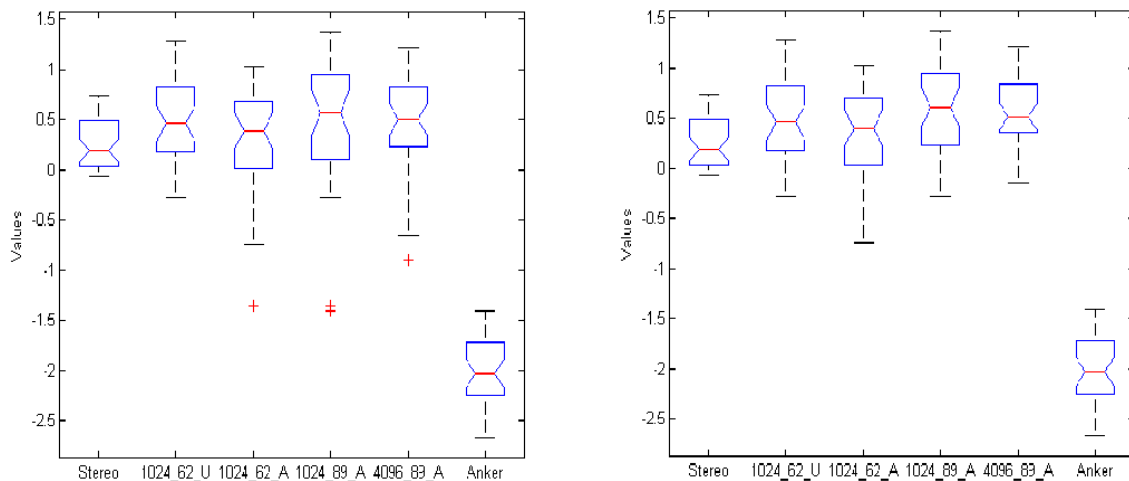


Abbildung 36: Boxplots der Bewertungen mit und ohne Outliner (z-Darstellung)

Die Ergebnisse für die Betrachtung der 95 % Konfidenzintervalle aus den paarweisen T-Tests für die verschiedenen Algorithmen (Abbildung 37) zeigen, dass die Verhältnisse der Algorithmen untereinander in etwa gleich sind, jedoch zeigt die Wiederholung (rechts) des Tests deutlichere Unterschiede zwischen den Algorithmen an. Zusätzlich kann eine Verschiebung hin zu besseren Bewertungen beobachtet werden.

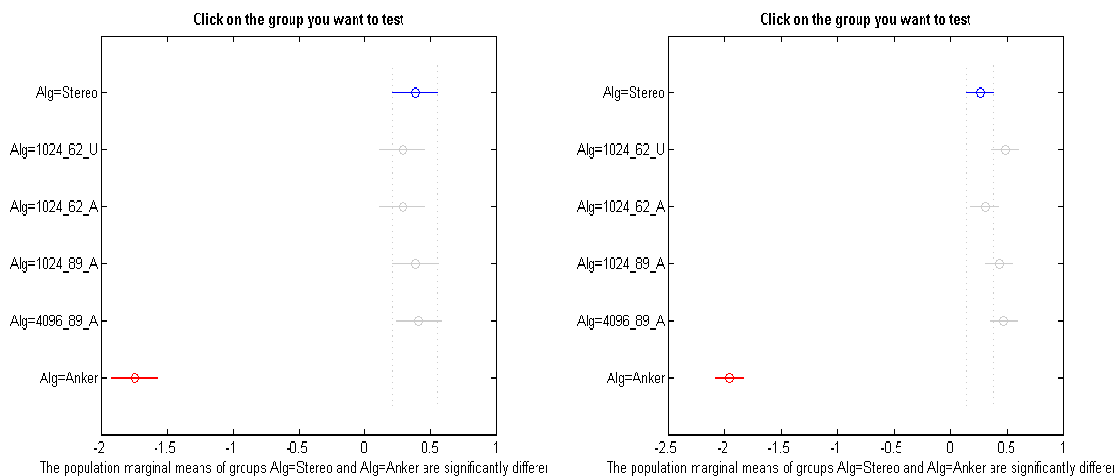


Abbildung 37: Gruppenvergleich Algorithmen: Test und Wiederholung mit 4 Hörern (z-Darstellung)

Im Gesamtergebnis dieses Tests, der mit 7 Personen durchgeführt wurde (ausschließlich oben genanntem Hörer) zeigt sich eine Algorithmusversion mit signifikant höherer Wertung als die Stereoverversion (Abbildung 38). Wenngleich dieses Ergebnis nur auf Befragung von 7 Personen beruht und die Schwankung groß zu sein scheint, wurde diese Version ausgewählt, um im

folgenden Verifikationstest zu zeigen, dass gegenüber dem ursprünglichen System signifikante Verbesserungen erzielt werden konnten. Das System beinhaltet die Parametereinstellungen

- „Ambient“ Ansatz Blind Upmix
- Virtuelle Lautsprecheraufstellung Front 89°
- Länge der BRIA 1024 Punkte

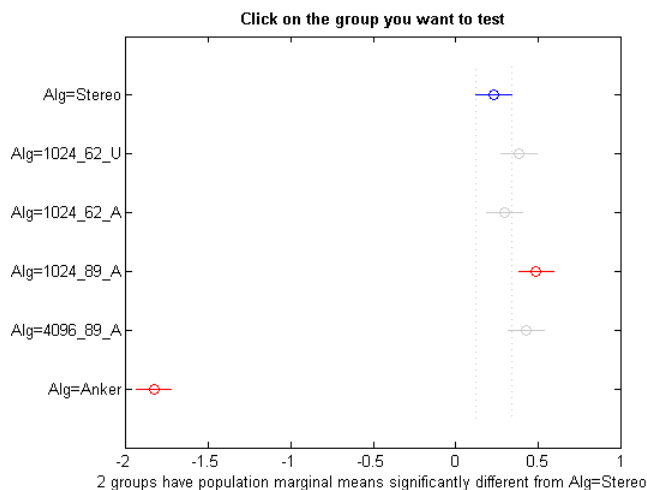


Abbildung 38: Wiederholungstest mit 7 Personen: z-Darstellung

5 Verifikation

Letztlich interessant für diese Arbeit ist die Antwort auf die Fragen: Konnte eine tatsächliche Verbesserung des Kombinationssystems durch die Optimierungen bewirkt werden? Ist diese Vorgehensweise mächtig genug eine Qualitätskontrolle für Konsumentensysteme darzustellen? Ein letzter Präferenztest wurde durchgeführt mit 18 Personen, sowohl geübten als auch ungeübten Hörern. Ungeübt in diesem Test war deswegen wichtig, weil das System ja auch in der Kundenrealität von naiven Hörern benutzt, und von ihnen beurteilt wird. Zusätzlich bedeutet ungeübt in dem Zusammenhang auch Hörer, die nicht in den vorhergehenden Entwicklungsphasen eingebunden waren. Es sollte die Gefahr ausgeschlossen werden, dass das neue System für die Hörer, deren Präferenz im Entwicklungsprozess als Kriterium diente, modelliert wurde. Aus den letzten beiden Tests wurde die Variante mit der Parameterkombination „Ambient“ Upmix, 89° virtuelle Lautsprecher Position und 1024 Punkte BRIA-Länge ausgewählt.

Wieder wurde ein repräsentativer Pool an 12 Musikstücken gewählt um die Entwicklung zu beobachten. Aus der Gesamtwertung mussten 2 von 10 ungeübten Hörern ausgeschlossen werden, da sie versteckte Referenzen in hohem Maße nicht erkennen konnten beziehungsweise ein Problem hatten den Test überhaupt in richtiger Art und Weise zu vollenden. Die Antworten der verbliebenen 16 Hörer wurden zunächst auf Normalverteilung und Varianzhomogenität geprüft. Für den Faktor Algorithmus liegen keine Normalverteilungen vor. Der Faktor Musikstück hat in seinen Gruppen ein hohes Maß an Normalverteilung. Varianzhomogenität liegt ebenfalls nur für den Faktor Musikstück vor. Dennoch wird ein ANOVA Modell mit den festen Faktoren Musikstück Hörer und Algorithmus berechnet.

Analysis of Variance						
Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F	Bestimmtheit
Algorithmus	113794,234	3	37931,4115	646,8506	0	0,43
Hörer	34993,9948	15	2332,933	39,7839	0	0,13
Music	1547,9635	11	140,724	2,3998	0,0066017	0,01
Algo*Hör	72377,599	45	1608,3911	27,4282	0	0,27
Algo*Muc	5765,2969	33	174,706	2,9793	1,55E-07	0,02
Hör*Muc	7096,5365	165	43,0093	0,73344	0,99063	0,03
Error	29026,8698	495	58,6401			0,11
Total	264602,495	767				
Constrained (Type III) sums of squares.						

Tabelle 5.1: Varianzanalyse Verifikationstest unbearbeitete Daten

Signifikante Effekte treten auf für alle Faktoren und ihre Interaktion, außer für Hörer* Musik. Das Bestimmtheitsmaß zeigt an, dass der Großteil (43 %) der Gesamtstreuung durch den Faktor Algorithmus verursacht wird. Ein nicht unerheblicher Teil wird durch den Faktor Hörer hervorgerufen, ein Indikator, dass die Hörer die Skala unterschiedlich benutzen. Die Betrachtung der Mittelwerte und 95% Konfidenzintervalle der Gruppen des Faktors Algorithmus zeigt ein optimistisches Ergebnis:

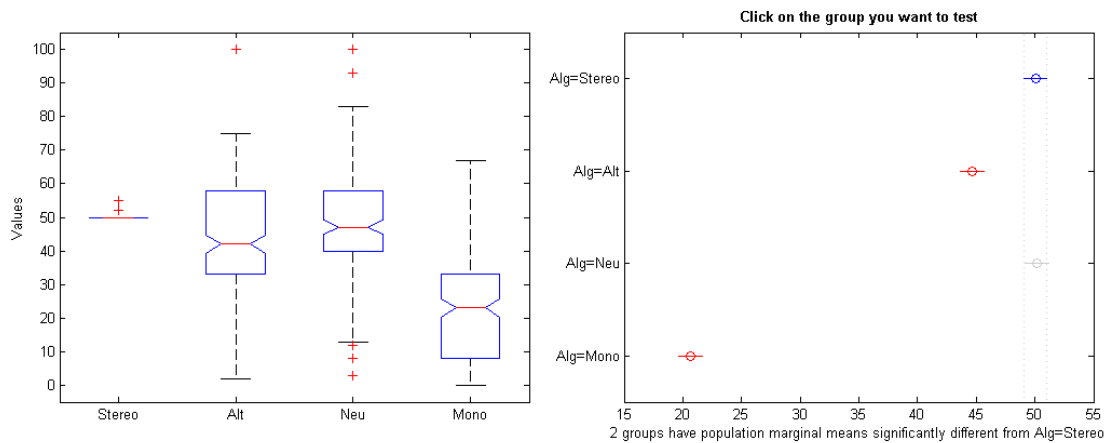


Abbildung 39: Boxplotdarstellung und 95% Konfidenzintervalle Verifikationstest
(unbearbeitete Daten)

Im rechten Bild zeigt blau die Stereoreferenz, rot die Algorithmen die signifikant zu schlechteren Bewertungen führen, grau die Bewertung des im Laufe der Arbeit veränderten Systems an. Grau bedeutet, dass keine signifikant unterschiedliche Wertung zur Stereoersion vorliegt. Dies sind Ergebnisse nach den nicht standardisierten Daten. Die Gruppen des Faktor Algorithmus zeigen weder Normalverteilung, noch Varianzhomogenität, somit hat diese Auswertung nicht allzu viel Gewicht. Die Betrachtung des Boxplot zeigt zwar signifikant unterschiedliche Medianwerte, jedoch auch breite Streuung. Die Daten werden standardisiert, um die unterschiedliche Benutzung der Skala durch die Hörer auszugleichen. Nach der Standardisierung der Daten verbessern sich die Voraussetzungen nur geringfügig für eine Aussagekräftige Interpretation der Ergebnisse.

Analysis of Variance						Bestimmtheit
Source	Sum Sq	d.f.	Mean Sq.	F	Prob>F	
Alg	425,1309	3	141,7103	600,1378	0	0,57
Lis	0	15	0	0	1	0
Muc	9,3794	11	0,85268	3,611	6,31E-05	0,01
Alg*Lis	138,5655	45	3,0792	13,0404	0	0,18
Alg*Muc	28,9735	33	0,87798	3,7182	1,11E-10	0,04
Lis*Muc	33,0665	165	0,2004	0,8487	0,89439	0,04
Error	116,8842	495	0,23613			0,16
Total	752	767				

Constrained (Type III) sums of squares.

Tabelle 5.2: Varianzanalyse Verifikationstest standardisierte Daten

Abbildung 40 zeigt die graphische Darstellung der Ergebnisse als Boxplot und 95 % Konfidenzintervall mit HSD Korrektur für Gruppenvergleiche.

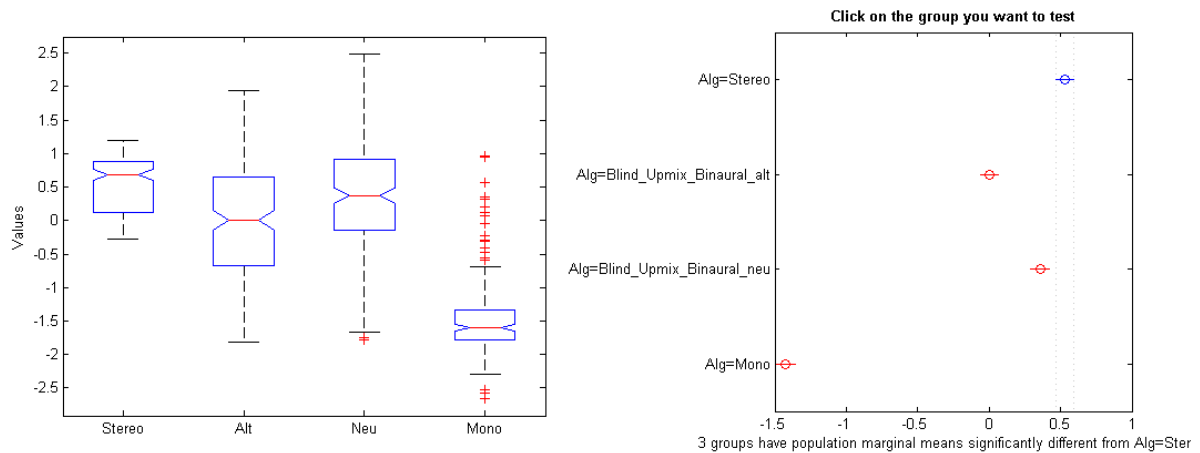


Abbildung 40: Boxplot und 95% Konfidenzintervalle des standardisierten Datensatzes

Um eine bessere Vorstellung der Datenverteilung zu bekommen werden sie mit einem MATLAB- Werkzeug zur Anpassung der Verteilung an eine Normalverteilung behandelt (Abbildung 41). Die Stufen der Datenverteilungen werden nach der Freedman Diaconis Regel⁷ berechnet. Die Parameter der Normalverteilungen wurden mit der *Maximum Likelihood* Funktion ermittelt. Die graphische Darstellung lässt vermuten, dass für eine Erhöhung der Anzahl der Hörer die angenäherten Normalverteilungen durchaus Gültigkeit besitzen. Die Betrachtung der 95% Konfidenzintervalle des standardisierten Datensatzes besitzen also mehr Aussagekraft. Die graphische Inspektion der Antwortverteilung (Abbildung 41 links oben) zeigt für die Stereoreferenz, dass es eine Gruppe Hörer gibt, die tendenziell positivere Bewertungen abgibt, und eine Gruppe, die tendenziell negativere Bewertungen abgibt. Diese Tendenzen sind jeweils auf die Stereoreferenz zu verstehen. Die Prüfung der Ergebnisse nach dieser Erkenntnis wird zusätzlich durchgeführt um genaueres zur allgemeingültigen Information, die aus diesem Test abgeleitet werden kann, zu finden. Dafür wird als Kriterium der Mittelwert der gesamten Einzelhörerwertung benutzt, so dass zwei Gruppen zu jeweils 8 Personen entstehen. Abbildung 42 zeigt die geschätzten Normalverteilungen für positivere und negativere Hörer, zusätzlich die Kurven des gesamten Datensatzes. Unabhängig davon, ob nun die Stereowiedergabe oder Blind Upmix Binauralwiedergabe bevorzugt wird, ist immer zu beobachten, dass Verbesserungen am neuen System im Vergleich zum alten System wahrnehmbar sind. Signifikanzprüfung wird mittels T-Test und KS-Test des standardisierten, aber unangepassten Datensatzes vorgenommen.

⁷ Breite der Stufen = $2 * IQR(x) N^{-1/3}$ IQR= Entfernung zwischen den Quartilen und N=Anzahl der Werte der Stichprobe x

Algorithmus	Positivere Bewerter	Negativere Bewerter	Alle
KS-Test $\alpha=0.05$	Keine Signifikanz P=0.696	Signifikante Unterschiede P=0.000	Signifikante Unterschiede P=0.000
T-Test $\alpha=0.05$	Keine Signifikanz P=0.169	Signifikante Unterschiede P=0.000	Signifikante Unterschiede P=0.001

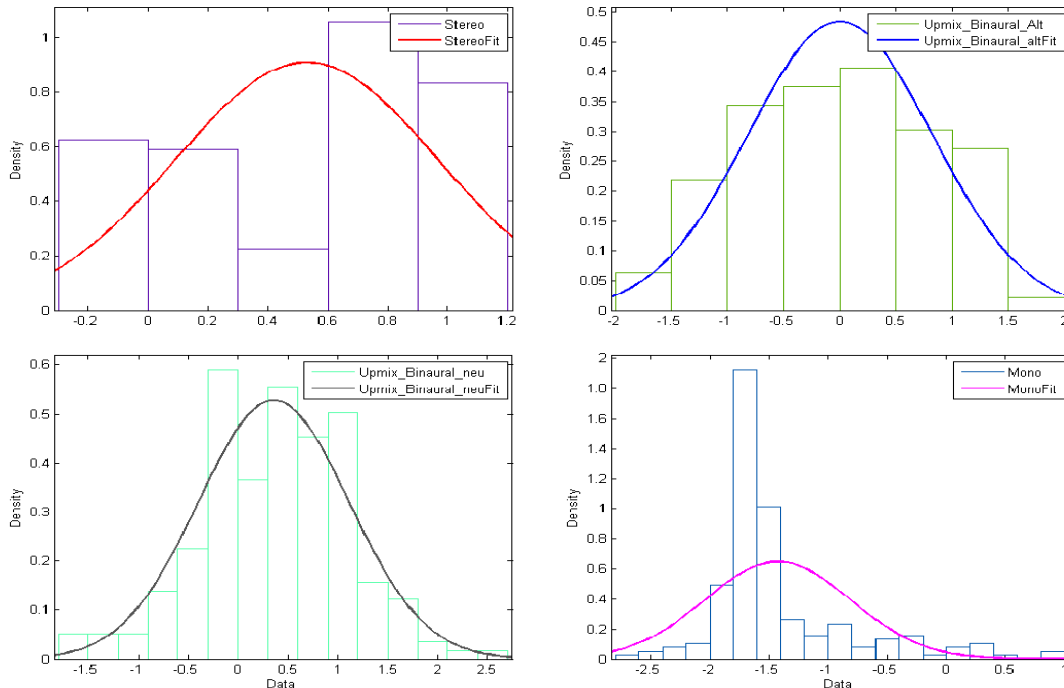


Abbildung 41: Anpassung des standardisierten Datensatzes an Normalverteilung

Für die Algorithmen (von links oben nach rechts unten) Stereo, Blind Upmix, Binaural alt, Blind Upmix Binaural neu, Mono

Interessant ist die Tatsache, dass Hörer, die nicht in den Entwicklungsprozess eingebunden waren, einen Großteil der negativ im Vergleich zur Stereoreferenz bewertenden Hörern ausmachen. Es könnte sich dabei um eine Art Gewöhnungseffekt für diese räumliche Wiedergabeform handeln, der nach oftmaligem Hören einsetzt. Dennoch zeigen sich die Relationen der Bewertungen zwischen dem Kombinationssystem vor und nach der Arbeit für beide Gruppen gleich (Abbildung 42).

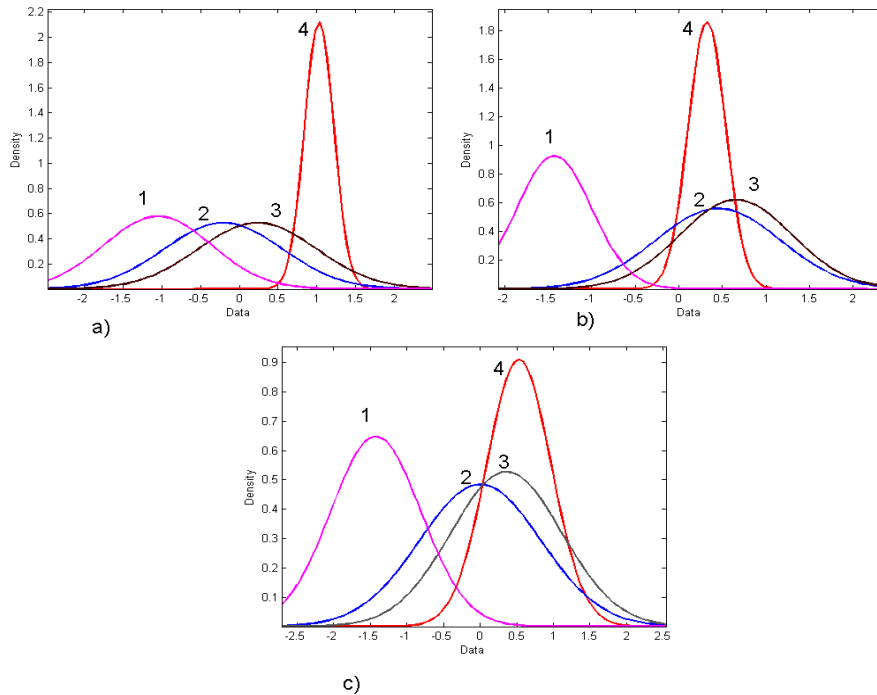


Abbildung 42: Annäherung der Normalverteilung

a) pessimistische Hörer b) optimistische Hörer

c) Gesamtergebnis

1) Monowiedergabe 2) Kombinationssystem vor der Arbeit
3) Kombinationssystem nach der Arbeit 4) Stereowiedergabe

Betrachtung der Antwortverteilungen des Faktors Musikstück zeigt ein bekanntes Bild aus den vorhergehenden Tests (Abbildung 43). Es zeichnet sich Signifikanz ab für Musikstücke, die in hohem Maße räumlich überlagerte Effekte bei der Wiedergabe aufweisen. Es zeigen sich tendenziell niedrigere Wertungen für das Mittel aller Algorithmen im Verhältnis zu Stereo. Es handelt in diesem Test wieder um die gleichen Stücke wie im Kombinationshörtest und seiner Wiederholung (Kapitel 4.3.3). Das Musikstück, und damit verbunden Aufnahme- und Produktionstechniken, spielen für die erfolgreiche Bearbeitung eine wesentliche Rolle.

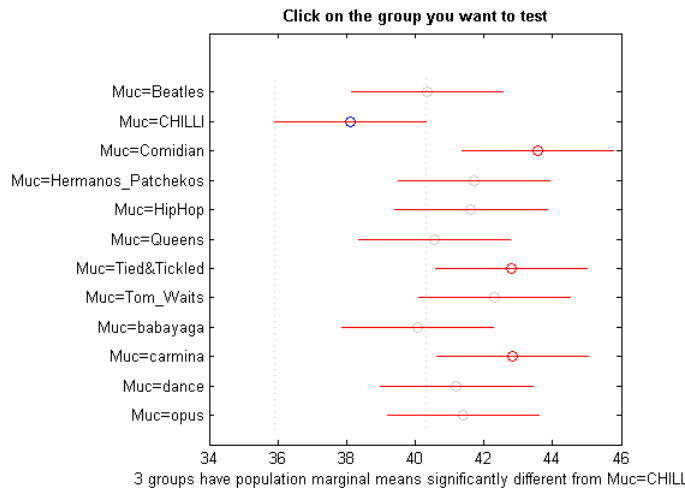


Abbildung 43: 95% Konfidenzintervalle zu Faktor Musikstück

6 Zusammenfassung und Ausblick

Diese Arbeit beschäftigt sich mit der Optimierung eines räumlichen Wiedergabesystems als Qualitätserweiterungsansatz zur Stereowiedergabe. Das System setzt sich aus zwei Einzelsystemen zusammen: Einem Blind Upmix System für 5.1 Surround Lautsprecherwiedergabe und einem Binaural Downmix System für virtuelle Surroundwiedergabe diskreter 5.1 Aufnahmen. Die erwiesene Funktionsfähigkeit der Einzelkomponenten wurde in der Kombination auf Hörerbevorzugung optimiert. Bei den Ansätzen zur Verbesserung des Systems stützt sich der Autor auf Arbeiten der jüngsten Vergangenheit und auf Hörbeurteilung des vorliegenden Systems. Die dafür benutzten Hörtestprozeduren wurden nach Stand der Wissenschaft entworfen und ausgewertet. Aus den Ergebnissen der ausgeführten Hörtests kann festgehalten werden, dass breitere virtuelle Lautsprecheraufstellung und, als Folge daraus, ein breiteres Stereobild zu erhöhter Hörerbevorzugung führt. Mittlere Längen der binauralen Raumimpulsantworten sind von Vorteil gegenüber nur geringen Raumanteilen. Eine spektrale Entzerrung im Vergleich zur Stereowiedergabe ist unverzichtbar für die Akzeptanz eines Stereoerweiterungssystems. Akzeptanz des vorliegenden Systems ist in hohem Maße abhängig von der Wahl des Musikprogramms. Damit in enger Verbindung steht die verwendete Aufnahmetechnik. Die Einstellung der Hörer zur Wiedergabeform ist, wie in Kapitel 5 angedeutet, nicht zu vernachlässigen. Es zeigt sich tendenziell (ohne eine detaillierte Betrachtung des Effekts

durchgeführt zu haben) Erhöhung der Akzeptanz des betrachteten Systems für Hörer, die mehrmals an Hörtests im Laufe der Arbeit integriert waren. Eine nicht geringe Anzahl der Hörer zeigte hauptsächlich Begeisterung für die traditionelle Stereowiedergabe. Das Etablieren als Alternative zur Stereowiedergabe im Sinne erhöhter Hörerbevorzugung konnte über ein Mittel aller Stücke nicht nachgewiesen werden. Eine signifikante Verbesserung bezüglich dem Ausgangspunkt der Arbeit konnte jedoch gezeigt werden.

Als Motivation für weitere Arbeiten könnte die Erweiterung des Signalmodells für den Blind-Upmix für zeitliche Positionierungen und Effekte von Vorteil sein. Die Wiedergabequalität könnte so auf konstantem Niveau gehalten werden. Ein weiterer Aspekt der Betrachtung finden könnte ist, dass die spektrale Anpassung im Zuge der Verarbeitung der Signale im Kombinationssystem adaptiv stattfinden könnte, und somit keine nachgeschaltete statische Entzerrung stattfinden muss. Als unbedingte Voraussetzung für ein System dieser Art ist die Verwendung von binauralen Raumimpulsantworten aus anerkannt wohlklingenden Räumen, zum Beispiel einem Aufnahmerraum für akustische Instrumente.

7 Literaturverzeichnis

- [1] A. Marui, W. L. Martens, "Spatial character and quality assessment of selected stereophonic image enhancements for headphone playback of popular music",AES 120th Convention Paris ,2006
- [2] Gaetan Lorho, "Evaluation of spatial enhancement systems for stereo headphone reproduction",AES 118th Convention Barcelona ,2005
- [3] Gaetan Lorho, D.Isherwood, N. Zacharow, J. Huopaniemi, "Round robin subjective evaluation of stereo enhancement systems for headphone",AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio ,2002
- [4] A.Walther, "Mehrkanalerweiterung von Stereo-Audioaufnahmen durch intelligente Upmix-Algorithmen",Diplomarbeit, Technische Universität Ilmenau ,2005
- [5] B.Kapralos M.R.M. Jenkin E.Milios, "Auditory Perception and Spatial (3D) Auditory Systems", University of York, DPT. of Computer Science ,2003
- [6] Julia Jakka, "Binaural to Multichannel Audio Upmix", Diploma Thesis, Helsinki University of Technology ,2005
- [7] W.H.Slattery JC Middlebrooks, "Monoaural soundlocalisation: acute versus chronic unilateral impairment",Hearing Research ,1984
- [8] Jens Blauert, "Räumliches Hören",Hirzel Verlag ,1974
- [9] E Zwicker, H. Fastl, "Psychoacoustics, Facts and Models", 2nd updated Edition ,Springerverlag ,1999
- [10] Soren H. Nielsen, "Auditory distance perception in different rooms",Journal of the AES Vol 41 No 10 ,1993

- [11] B. Laback, "Psychoakustik II: Schwerpunkt Experimentelle Audiologie",Skriptum zur Vorlesung an der KU Graz 3. Auflage ,Februar 2006

- [12] ITU-R Recommendation Bs. 1387, "PEAQ-the proposed ITU Standard for objective measurement of perceived audio quality" ,1999

- [13] Recommendation ITU-R Bs. 1116, "Method for subjektive assessment of small impairments in audio systems including multichannel sound systems",ITU-R ,1994

- [14] Recommendation ITU-R Bs 1534, "Method for the subjective assessment of intermediate audio quality MUSHRA" ,2001

- [15] Recommendation ITU-R Bs.1284, "Methods for the subjective assessment of sound quality",ITU-R ,1997

- [16] Soren Bech, Nick Zacharov, "Perceptual Audio Evaluation",John Wiley and Sons Ltd, 2006

- [17] Russel Mason, N.Ford.F.Rumsey, B. de Bruyn, "Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction",AES 109th Convention ,2000

- [18] Soren Bech, "Methods for subjective evaluation of spatial characteristics of sound",AES 16th conference on spatial sound reproduction ,1999

- [19] N. Zacharov, K. Koivuniemi, "Unravelling the perception of spatial sound reproduction: Analysis and external preference mapping",111th international AES Convention NY ,2001

- [20] Gaetan Lorho Nick Zacharow, "Subjective evaluation of virtual home theatre sound systems for loudspeakers and headphones",116th International AES Convention ,2004

- [21] Jan Berg, Francis Rumsey, "Spatial attribute identification and scaling by repertory grid technique and other methods",AES 16th International Conference on Spatial Sound Reproduction ,1999

- [22] Jan Berg, Francis Rumsey, "Systematic evaluation of perceived spatial quality",24th International AES Conference on Multichannel Audio ,2003

- [23] Soren Bech, "Selection and training of subjects for listening tests on sound-reproducing Equipment",Journal of the AES Vol 40 7/8 ,1992

- [24] Sean E. Olive, "Differences in Performance and preference of trained versus untrained listeners in loudspeaker tests: A case study",Journal of the AES V51 Number 9 ,2003

- [25] Christof Faller, "Parametric coding of spatial audio",Proceedings of the 7th international Conference on DAFX, Naples,Italy ,2004

- [26] C. Avendano J.M. Jot, "Frequency domain techniques for stereo to multichannel upmix", 22nd AES Conference on Virtual, Synthetic and Entertainment Audio ,2002

- [27] ITU-R, "Recommendation Bs 775-1 Multichannel stereophonic sound system with and without accompanying picture", ,1992-1994

- [28] Ville Pulkki, "Compensating displacement of amplitude-panned virtual sources",22nd International AES Conference on Virtual, Synthetic and Entertainment Audio ,2002

- [29] William L. Martens, "Contextual effects in sensory evaluation of spatial audio: Integral factor or nuisance?",Spatial Audio and Sensory Evaluation Techniques Guildford UK April 2006 ,2006

- [30] G.Soulodre, M.C.Lavoie, S.G.Norcross, "Objective measure of listener envelopment in multichannel surround systems",Journal of the Audio Engeniering Society ,9/2003

- [31] P Minnaar, S. K. Olesen, H. Moller, "The importance of head movements for localisation with binaural synthesis",unveröffentlicht
- [32] L. Rade B. Westergren P.Vachenauer, "Springers Mathematische Formeln",Springer Verlag Berlin Heidelberg ,1997/ 2.Auflage
- [33] J.Berg F.Rumsey, "Verification and correlation of attributes used for describing the spatial quality of reproduced sound",19 th international AES Conference ,2001
- [34] W.Ludwig-Mayerhofer, "ILMES- Internet-Lexikon der Methoden der empirischen Sozialforschung: Levene-Test auf Varianzgleichheit",2003
http://www.lrz-muenchen.de/~wlm/ilm_110.htm

Abbildungsverzeichnis

Abbildung 1: Funktionskette Upmix- Binaural Downmix.....	2
Abbildung 2: ITD und ILD.....	5
Abbildung 3: "Cone of Confusion".....	6
Abbildung 4: Richtungsbestimmende Bänder nach Blauert aus [8].....	8
Abbildung 5: Nachhall Diagramm.....	11
Abbildung 6: ILD Abhängig von Entfernung im Nahfeld [5].....	14
Abbildung 7: Verbreiterung der auditorischen Filter bei Hörgeschädigten aus [11].....	15
Abbildung 8: Konzepte und verbale Beschreibung aus[21].....	20
Abbildung 9: Scheibermatrix.....	23
Abbildung 10: Blind Upmix mit Ambientansatz.....	30
Abbildung 11: "In the Band"- Ansatz.....	31
Abbildung 12: Signalmodell Surround Erlebnis.....	32
Abbildung 13: Messung der BRIA im Surround Setup aus [27].....	33
Abbildung 14: MLS Methode.....	33
Abbildung 15: Richtungen der ersten Reflexionen.....	34
Abbildung 16: Ausschnitt aus gemessener BRIA.....	34
Abbildung 17: Prinzip des binauralen Downmixes.....	35
Abbildung 18: Vom Klanggeschehen zur subjektiven Bewertung aus [18].....	37
Abbildung 19: Skala für Vergleichstests.....	45
Abbildung 20: Wavswitch.....	46
Abbildung 21: Prinzip der spektralen Entzerrung.....	48
Abbildung 22: Differenzspektrum.....	50
Abbildung 23: Entzerrungsfilter.....	50
Abbildung 24: Betragsfrequenzgang der optimierten Anpassungsfilter.....	51
Abbildung 25: 95% Konfidenzintervalle spektrale Entzerrung.....	53
Abbildung 26: 95% Konfidenzintervalle Upmix Strategien	55
Abbildung 27: Spektrale Verläufe der Fensterfunktionen.....	57
Abbildung 28: 95% Konfidenzintervalle BRIA Längen.....	58
Abbildung 29: 90 ° BRIA.....	60
Abbildung 30: 95% Konfidenzintervalle Dämpfung der kontralateralen BRIA.....	61
Abbildung 31: 95% Konfidenzintervall virtuelle Lautsprecherposition.....	62

Abbildung 32: Funktionsweise der SVD.....	63
Abbildung 33: 95% Konfidenzintervalle MS-Dekorrelation der hinteren Kanäle.....	65
Abbildung 34: Paarvergleich der Gruppen des Faktor Algorithmus.....	69
Abbildung 35: Paarvergleich Gruppen des Faktors Musikstücke.....	70
Abbildung 36: Boxplots der Bewertungen mit und ohne Outliner (z-Darstellung).....	71
Abbildung 37: Gruppenvergleich Algorithmen: Test und Wiederholung mit 4 Hörern	71
Abbildung 38: Wiederholungstest mit 7 Personen: z-Darstellung.....	72
Abbildung 39: Boxplotdarstellung und 95% Konfidenzintervalle Verifikationstest	74
Abbildung 40: Boxplot und 95% Konfidenzintervalle des standardisierten Datensatzes	75
Abbildung 41: Anpassung des standardisierten Datensatzes an Normalverteilung	76
Abbildung 42: Annäherung der Normalverteilung.....	77
Abbildung 43: 95% Konfidenzintervalle zu Faktor Musikstück.....	78

8 Anhang

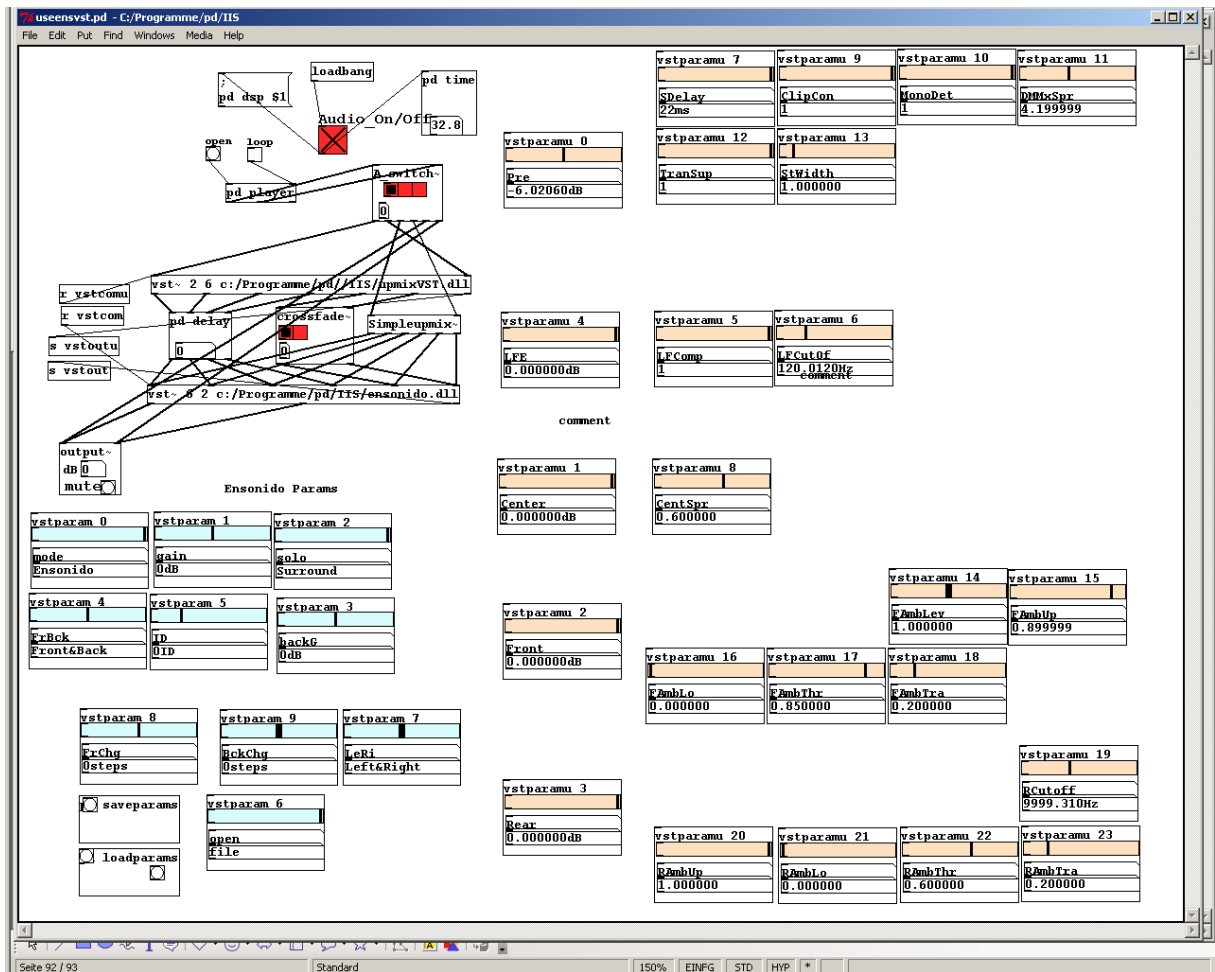
Interpret-Musikstück	Intensitäts- positionierung	Raumeffekte oder Zeitstereophonie	Stücke mit hohem Transient Anteil	Liveaufnahmen
Comedian Harmonists-“Veronika der Lenz ist da“ (Filmsoundtrack)	X			
Red Hot Chilli Peppers- „Funky Monks“		X		
Queens of the Stoneage-“No one knows“		X	X	
Mussorgsky-“Pictures at an Exhibition-The hut on fowl's legs“ Chicago Synphonie Orchestra		X		X
Carl Orff-“Carmina Burana“		X		X
Eve-“Eve ft. Missy Elliot“	X		X	
Hermanos Patchekos-“Café! Pronto!“		X		
Beatles-“All you need is Love“ live				X
Opus-“Live is Live“				X
Unknown-„Dance with me“	X			
Tied & Tickled Trio-“Revolution“	X		X	
Omara Portuondo-“Ella y Yo“		X		
Sarah Mc Lachlan-“Adia“		X		
Mogwai-“Glasgow Mega Snake“				
The Notwist-“Torture Day“				
Faith No More-“Zombie Eaters“ live				X
Sonic Youth-„Song for Karen“	X			
Bill Evans-“Pastiece“	X			X
Bill Frisell-“Shenandoah“	X			
Tom Waits-“Cold Cold Ground“	X			
Apollo 440-“Don't fear the Reaper“			X	
Michael Buble-“Quando Quando Quando“	X			
Dead can Dance- „Flowers of the sea“		X		
Gloria Gaynor- „I will	X			

survive“				
Beatles-“Help“	X			
Libero Bovio-Titel Unbekannt		X		
Björk- „New World“ (Dancer in the Dark)	X		X	
Dead can Dance-“Fortune“		X		
Hans Zimmer- „Now we are Free“ (Gladiator Soundtrack)		X		
Chorgesang, Titel und Chor unbekannt		X		X
Anthony and the Johnsons -“Hope there's someone“	X			
Triton „Vivid Dreams“				

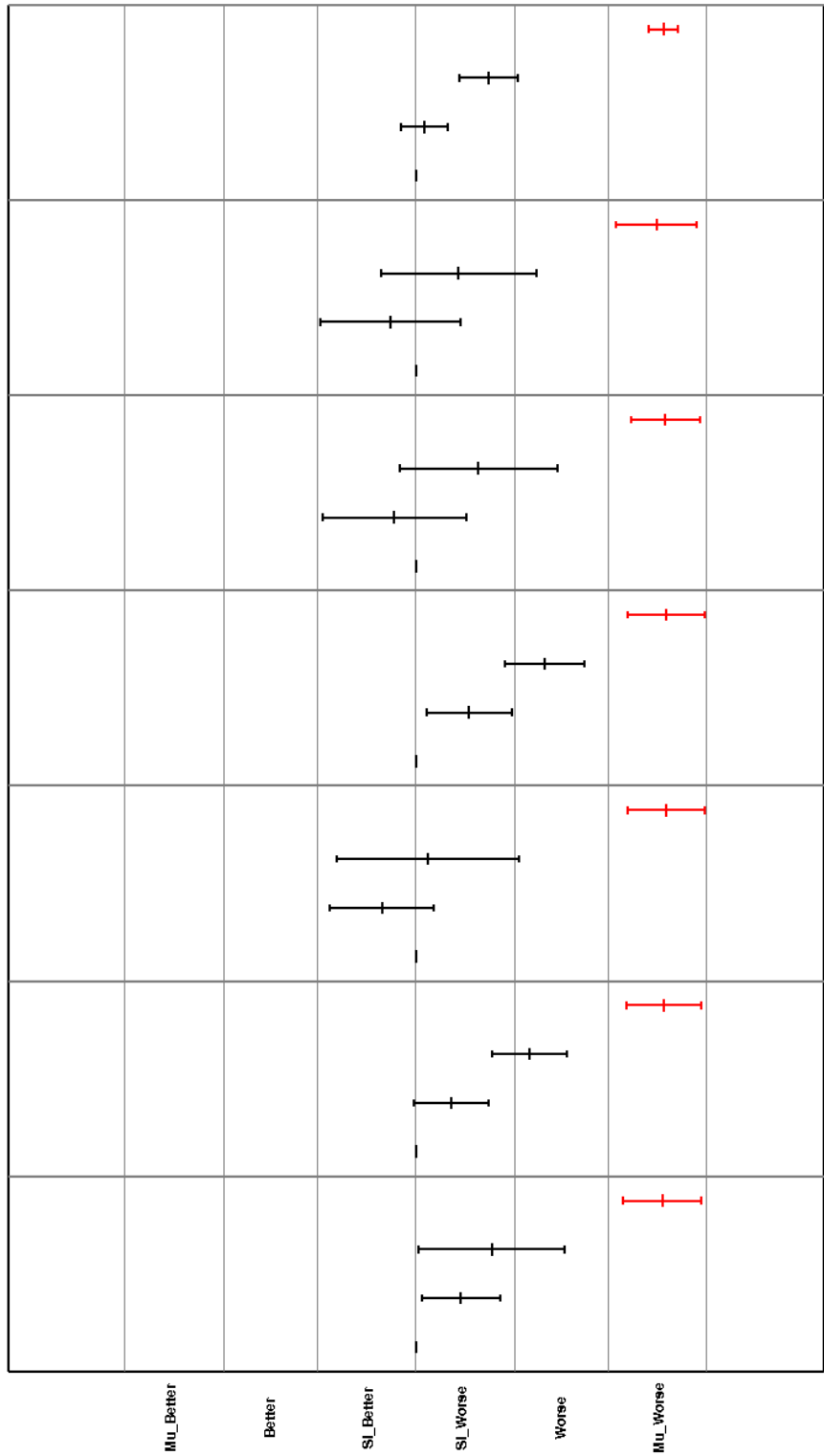
Tabelle_A I: In der Arbeit verwendete Musikstücke

Parameter		
SDelay	0-27 ms	Verzögerung der Surroundkanäle bis max 27 ms
TranSup	On/Off	
Stereowidth	0-10	Zusätzliche Stereoverbreiterung für Frontkanäle
LFE	0--inf	Lautstärkeregelung für LFE Kanal
LFECutoff		Einstellung der Grenzfrequenz für LFE
Center	0--inf	Lautstärkeregelung für Center-Kanal
CentSpr	0-1	Centerkanal wird auf links und rechts verteilt
Front	0--inf	Gemeinsamer Verstärkungsfaktor für Frontkanäle L,R,C
FAmbThr	0-1	Übergangsschwelle, ab welchem Maßwert (Kor +Koh) ein Signal als „Ambient“ kategorisiert wird
FAmbUp	0-1	Obere Grenze für Gewichtungsmaske Front Ambientsignal
FAmbLo	0-1	Unterer Wert für Gewichtungsmaske Front Ambientsignal
FAmbTra	0-1	Breite des Übergangsbereiches der Gewichtungsmaske
FAmbLev	0-1	Verstärkungsfaktor für Ambient Maske
Rear	0--inf	Gemeinsamer Verstärkungsfaktor für Rearkanäle Ls,Rs
RAmbThr	0-1	Energiedifferenz für die Kategorisierung als „Ambient“ für hintere Kanäle
RAmbUp	0-1	Obere Grenze für Gewichtungsfaktor Rear Ambientsignal
RAmbLo	0-1	Untere Grenze für Gewichtungsfaktor Rear Ambientsignal
RAmbTra	0-1	Steilheit der Gewichtungsmaske
RCutoff	0-20000	Grenzfrequenz für Rear-Kanäle
Binaural Downmix Parameter		
Mode	On/Off	Umschaltbar zwischen ITU-Downmix und Binaural Downmix
Gain	-12dB/+12dB	Verstärkung des Ausgangspegels
Solo		Einzelne Lautsprecher signale einzeln hören
FrBck	Front/Back	Nur vordere oder nur hintere Surround-Kanäle hören, oder alle.
ID	0-3	Verschieden Filtersets
backG	-12dB/+12dB	Verstärkungsfaktor für hintere Kanäle
FrChg	-8/+8	Bis zu 16 Front-Lautsprecherpositionen auswählbar.
BckChg	-8/+8	Bis zu 16 Rear-Lautsprecherpositionen auswählbar.
LeRi	0-2	Nur linke oder nur rechte Kanäle hören, oder alle.
Open	File	Eine binaurale Filterbank kann eingeladen werden.

Tabelle_A II: Parameter VST-Plugins

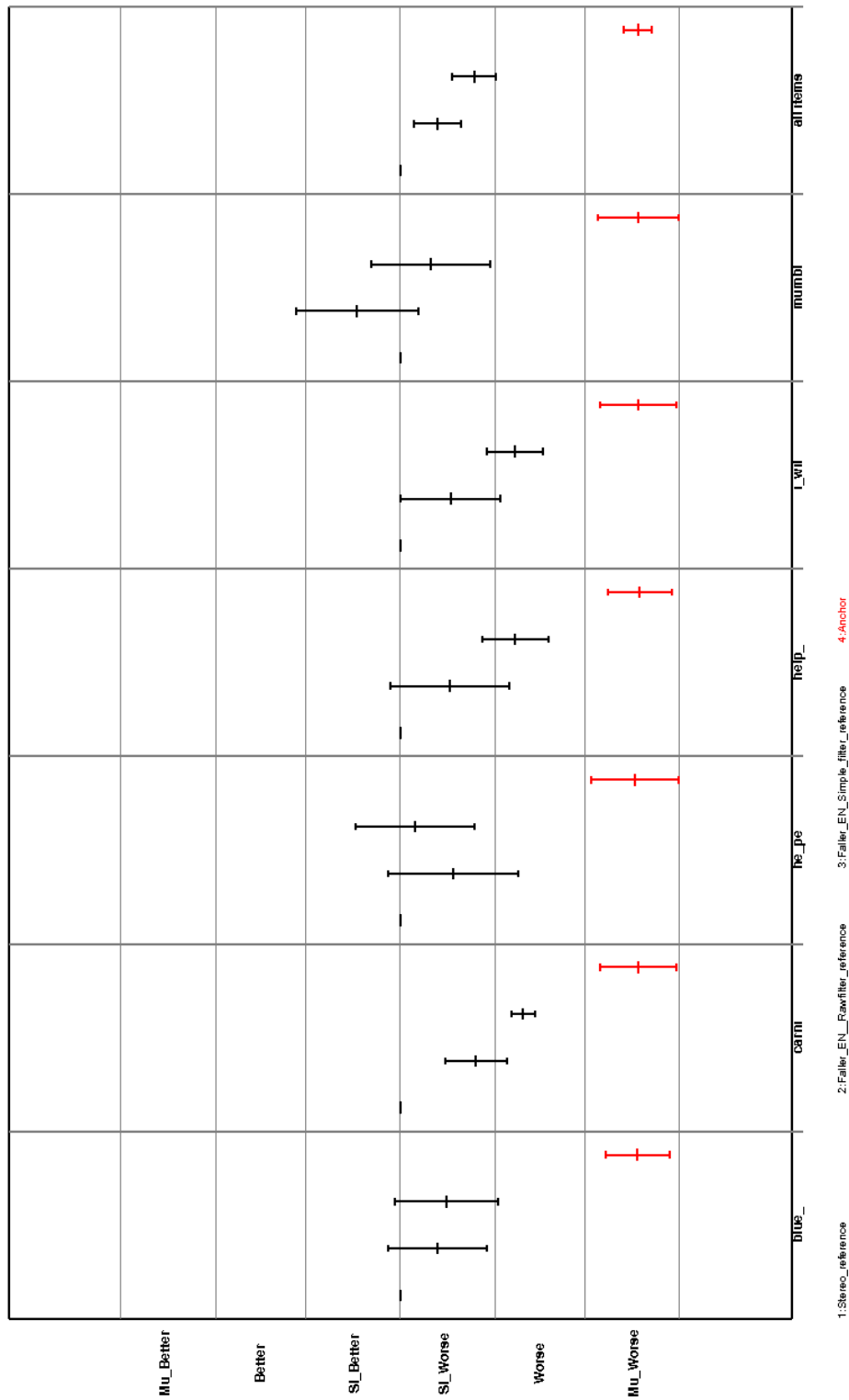


Abbildung_A I: Echtzeitumgebung für VST-Plugins

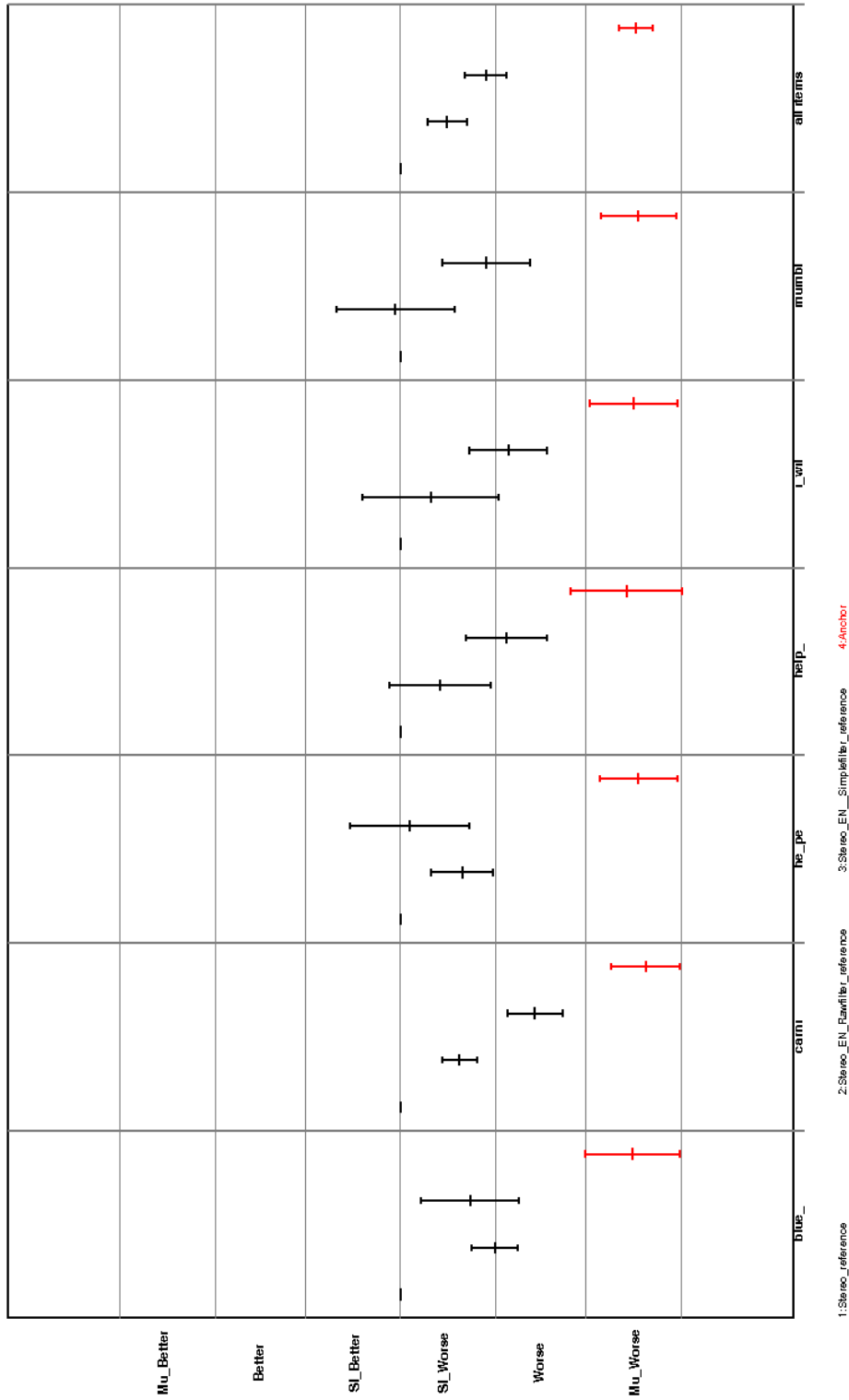


1: Stereoc_reference
 2: Dolby_EN_RawFilter_reference
 3: Dolby_EN_Simple_Filter_reference
 4: Anchor

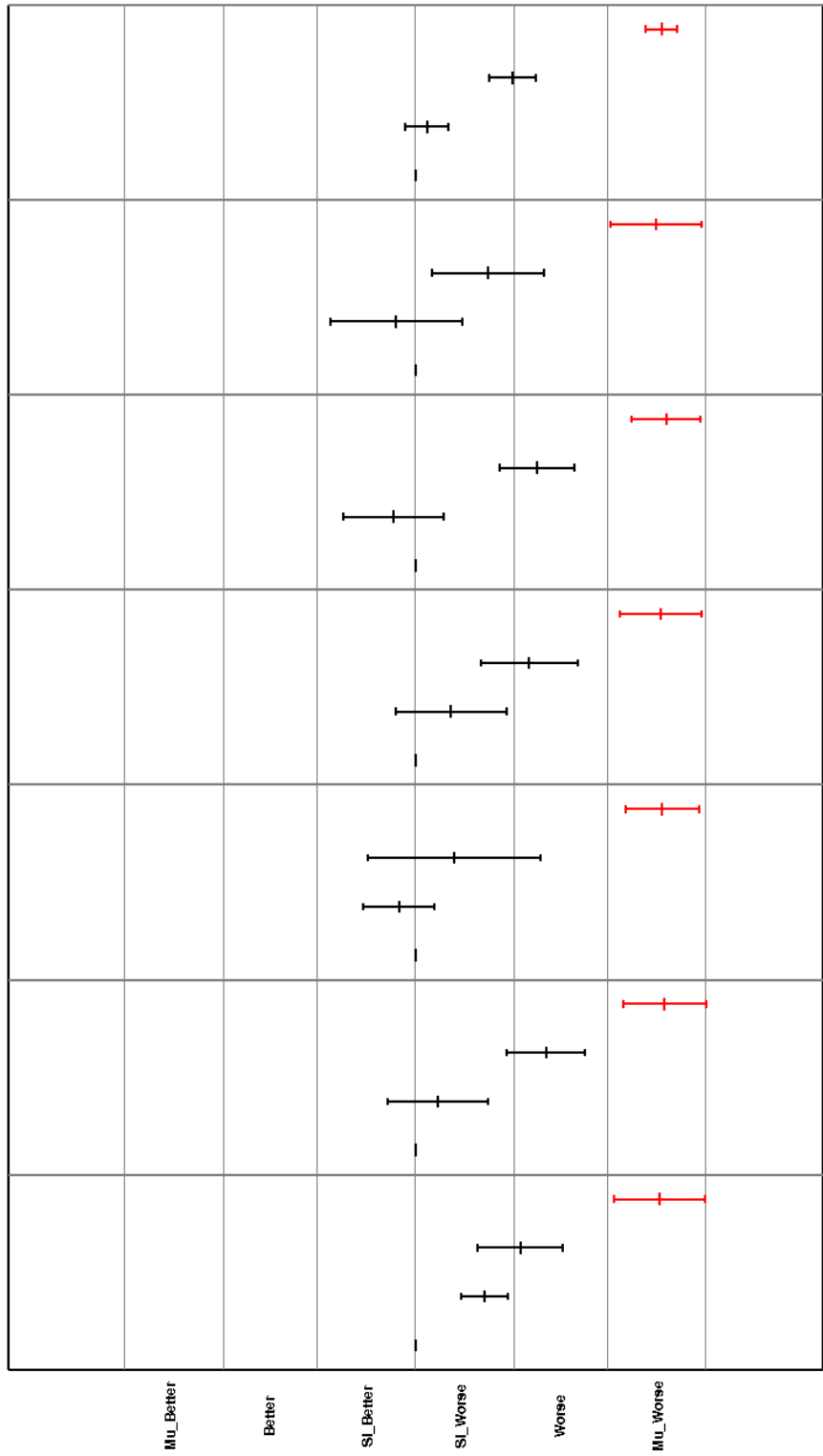
Abbildung_A II: 95% Konfidenzintervalle spektrale Anpassung System 1



Abbildung_A III: 95% Konfidenzintervalle spektrale Anpassung System 2

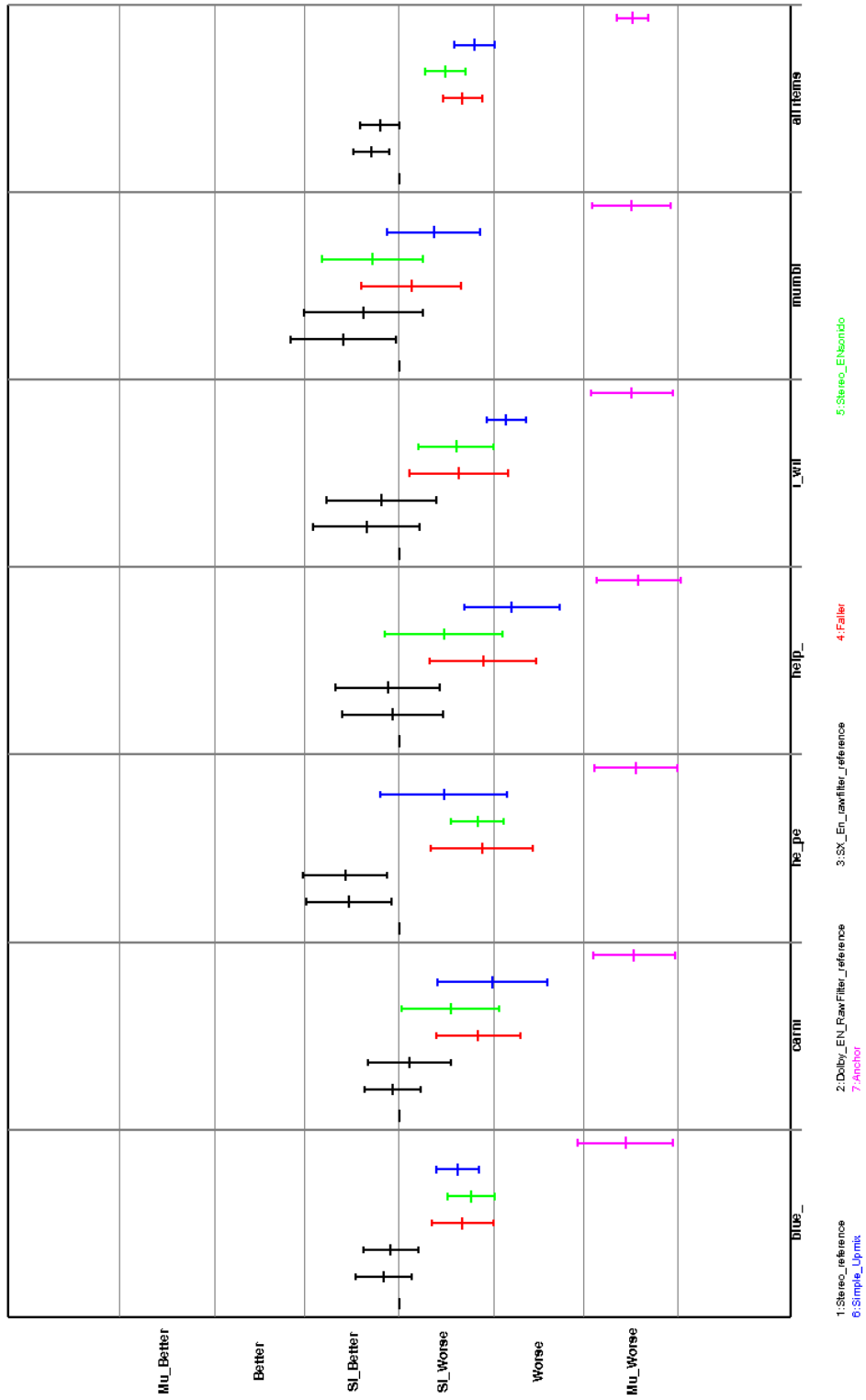


Abbildung_A IV: 95% Konfidenzintervalle spektrale Anpassung System 3

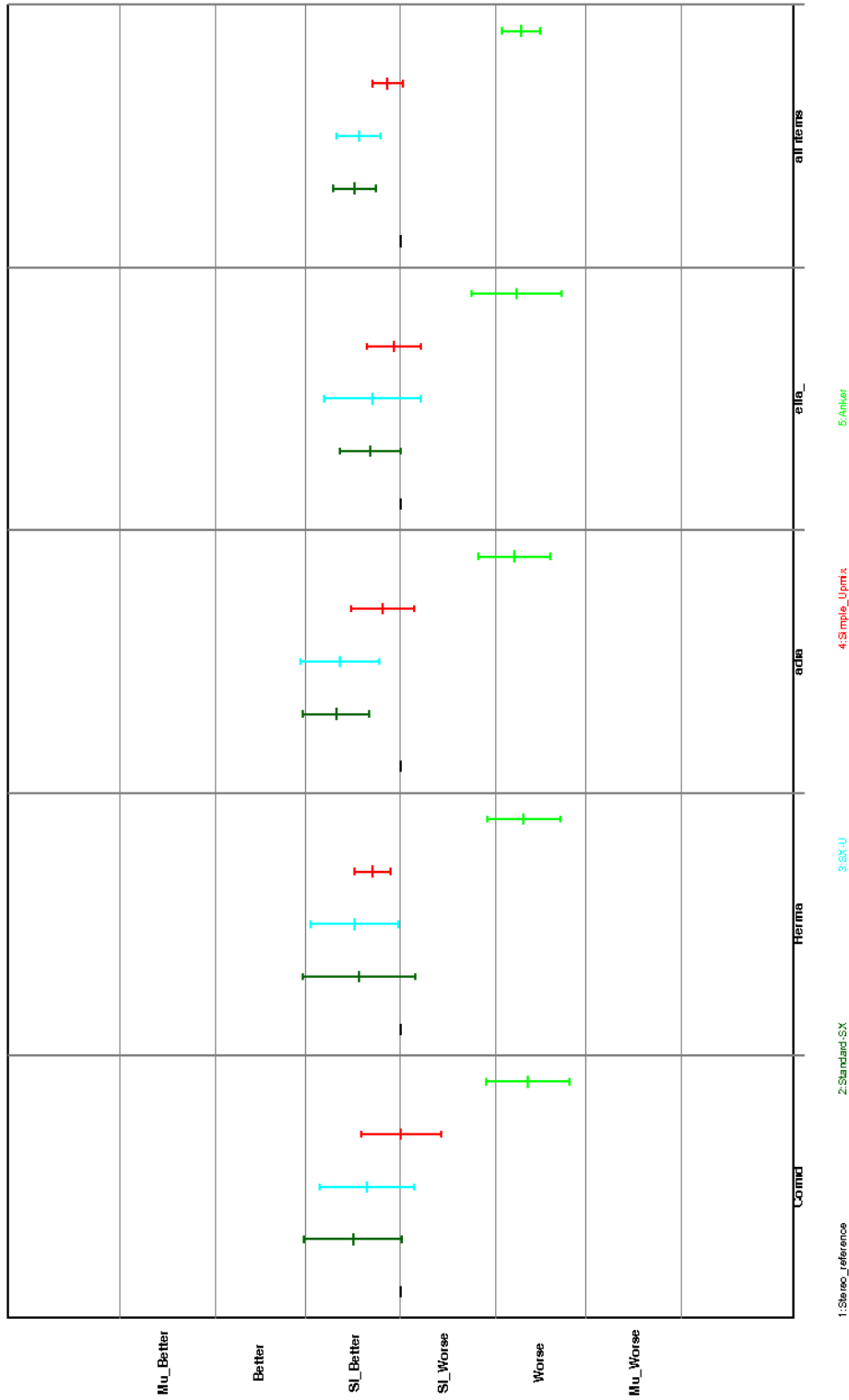


1: Stereoc_reference
 2: SX_EN_Jawfiller_reference
 3: SX_EN_Simple_filler_reference
 4: Anchor

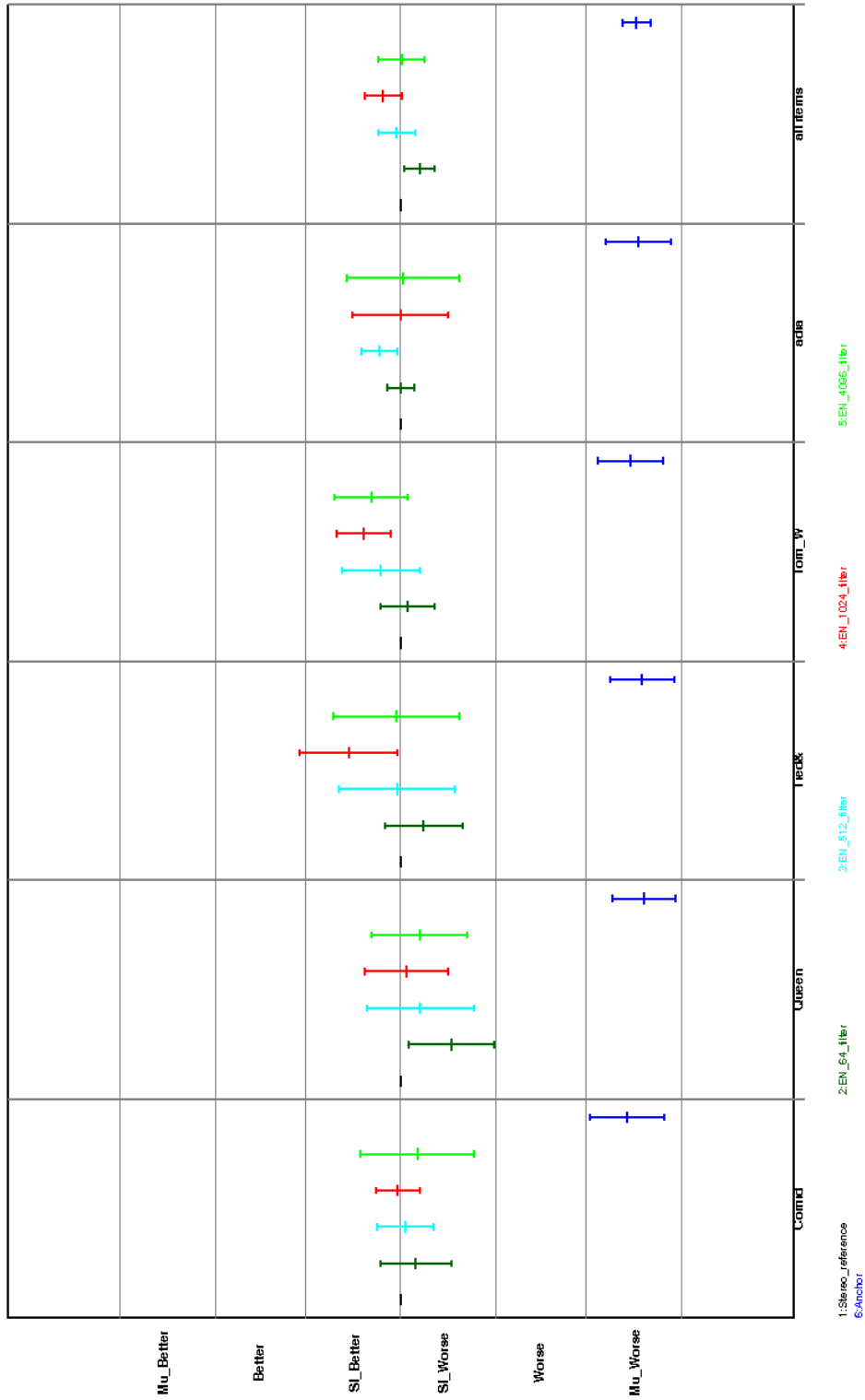
Abbildung_A V: 95% Konfidenzintervalle spektrale Anpassung System 4



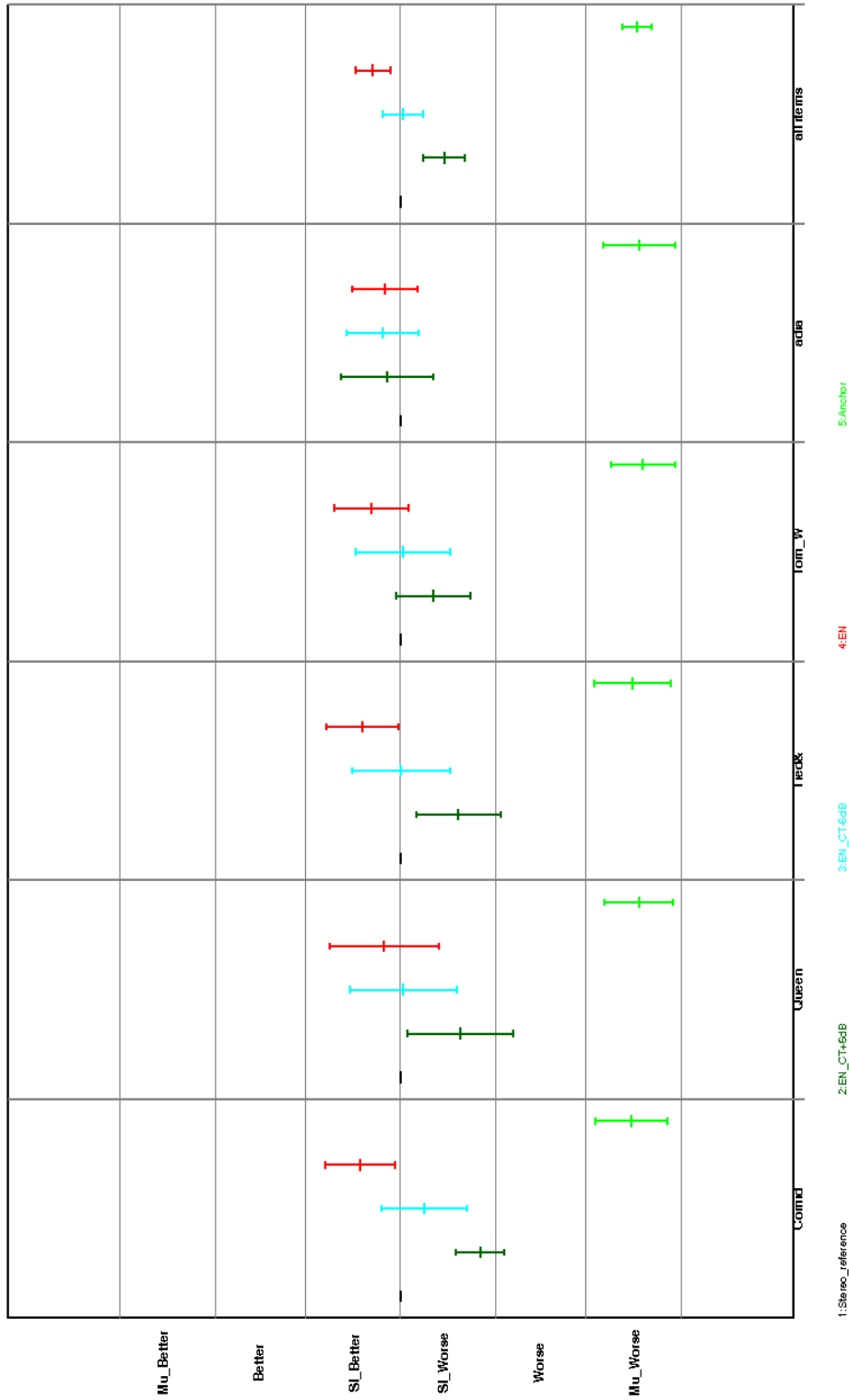
Abbildung_A VI: 95% Konfidenzintervalle Upmix Strategien (extern)



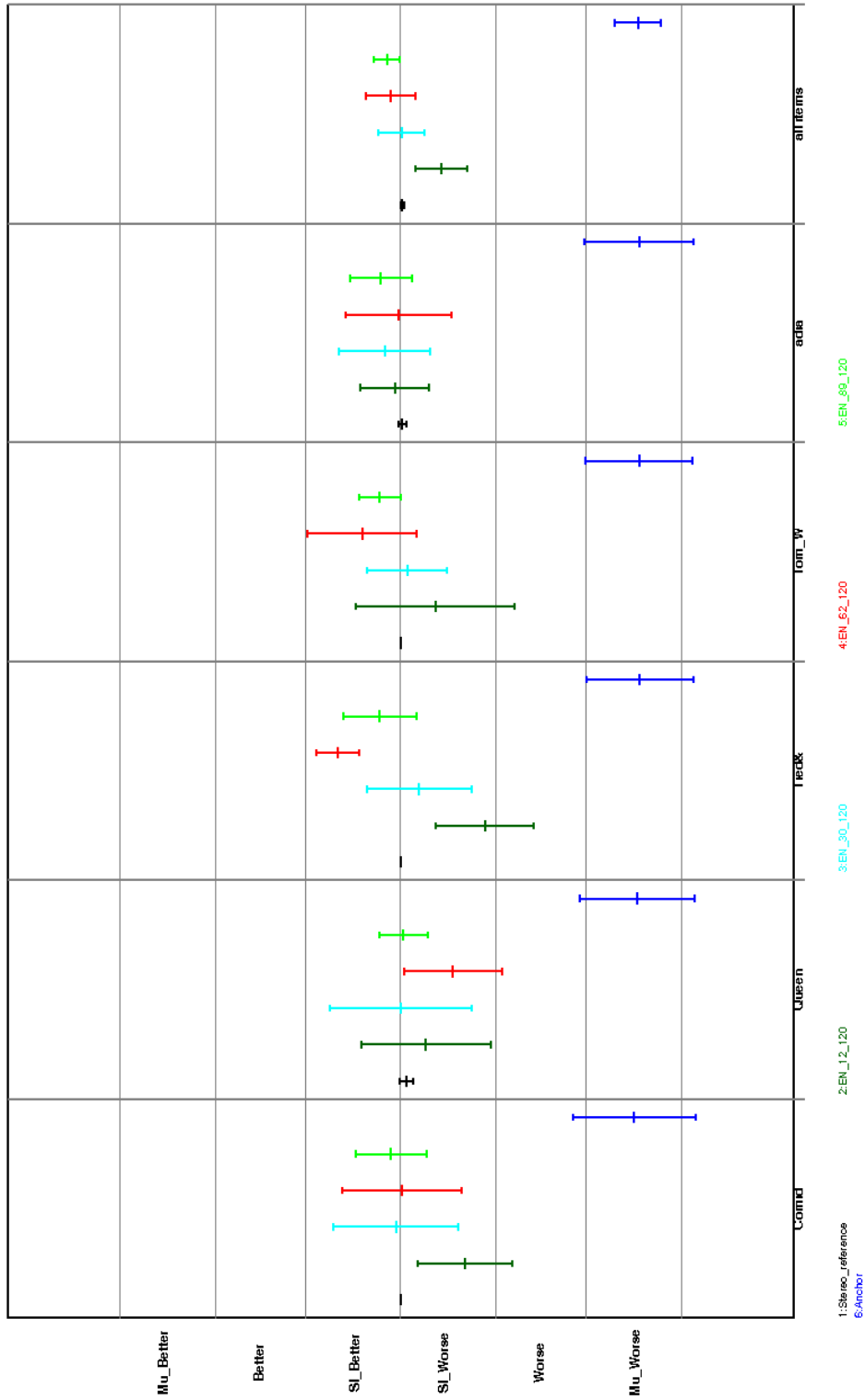
Abbildung_A VII: 95% Konfidenzintervalle Upmixstrategien (intern)



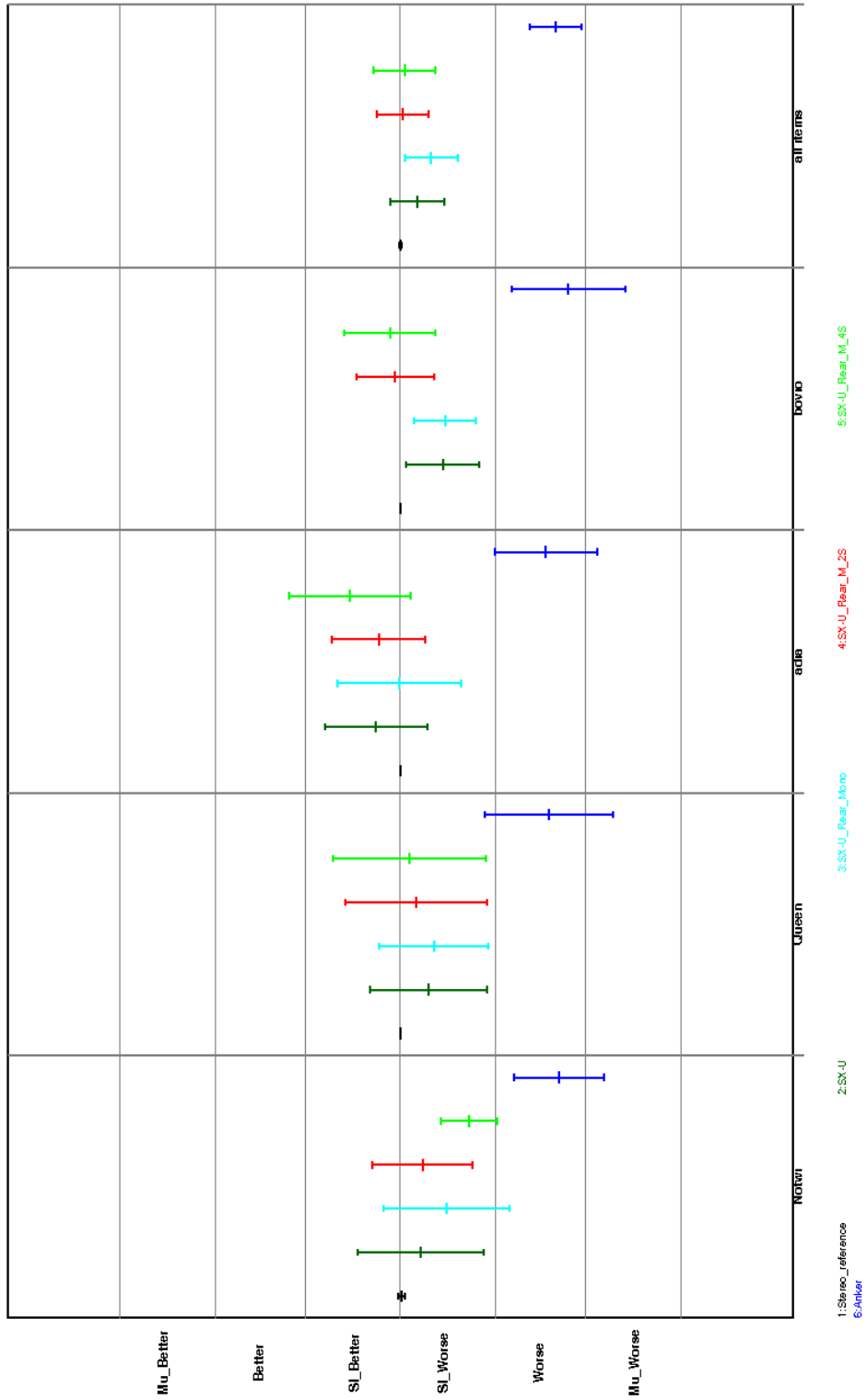
Abbildung_A VIII: 95% Konfidenzintervalle BRIA-Länge



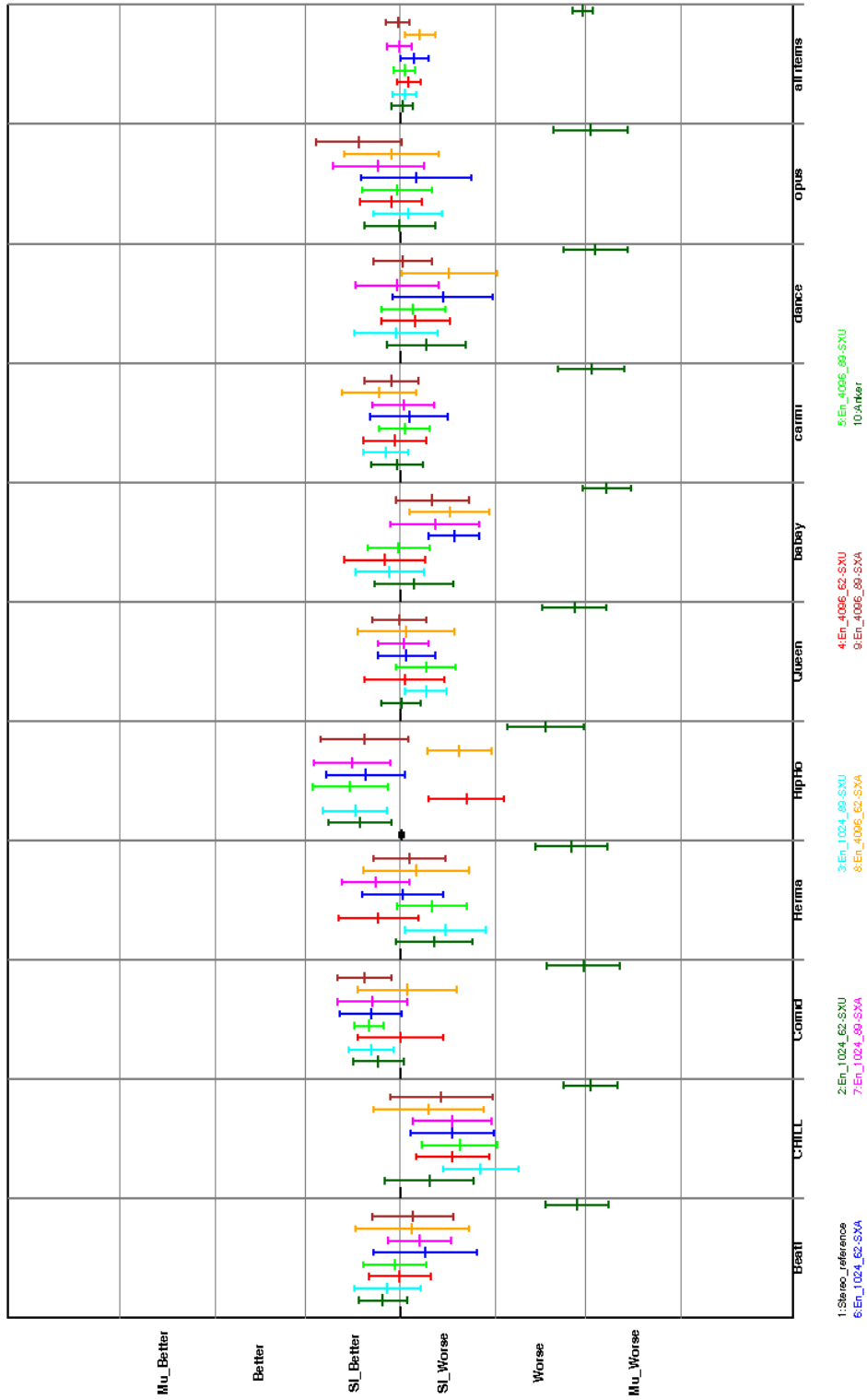
Abbildung_A IX: 95% Konfidenzintervalle Dämpfung kontralaterale BRIA



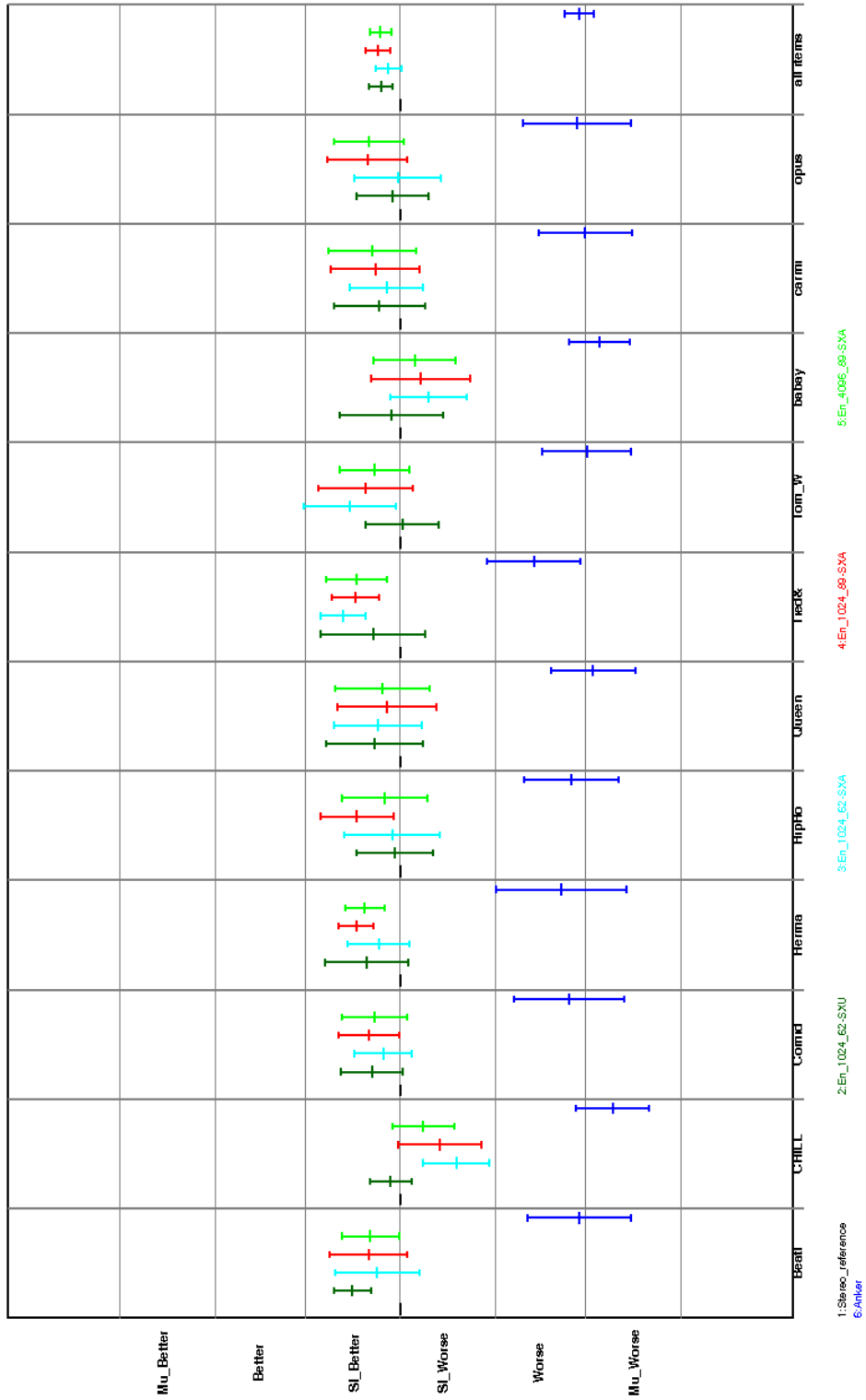
Abbildung_A X: 95% Konfidenzintervalle virtuelle Lautsprecher Position



Abbildung_A XI: 95 % Konfidenzintervalle MS-Dekorrelation

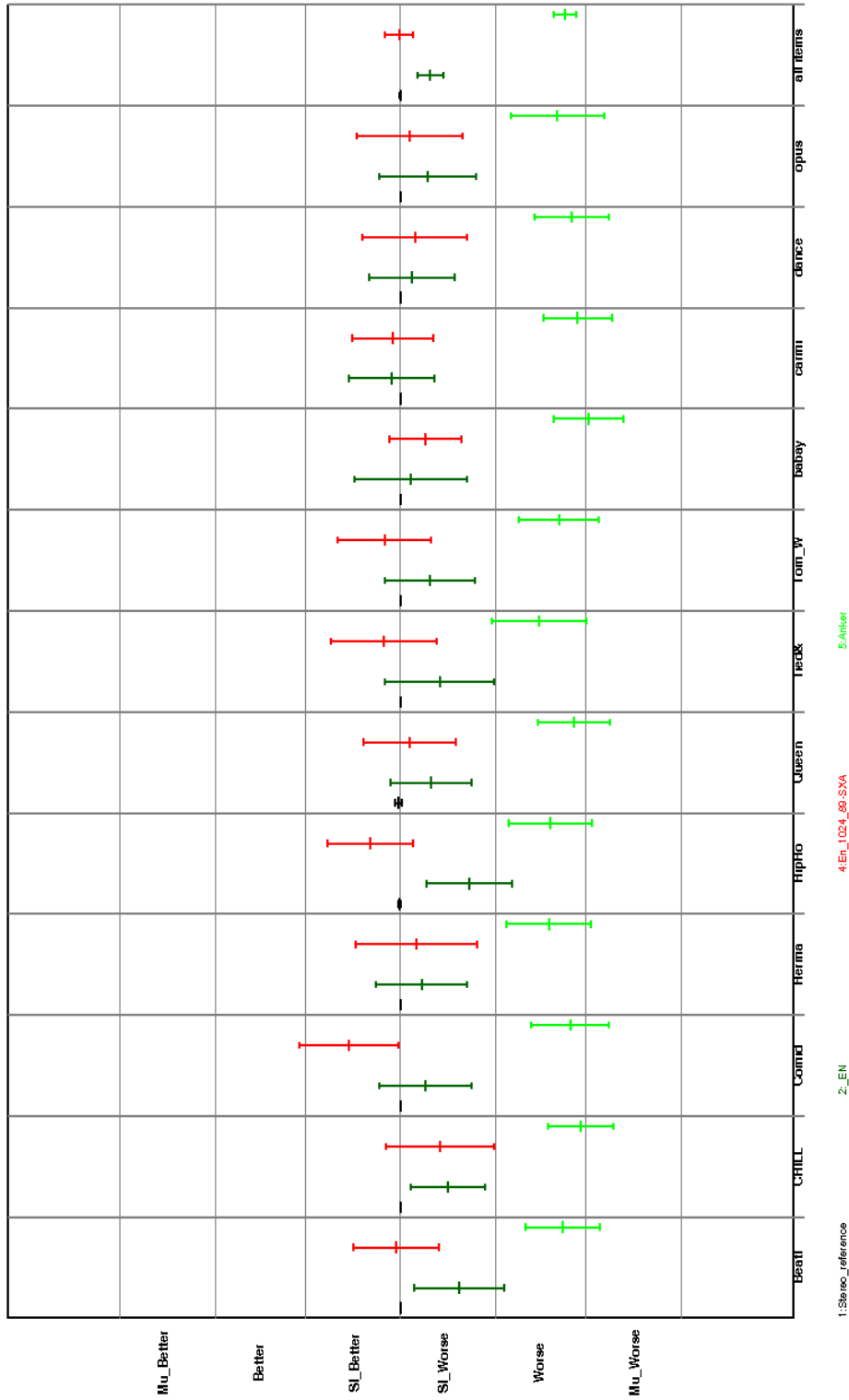


Abbildung_A XII: 95% Konfidenzintervalle Kombinationstest



1: Stereo_reference
 6: Anker
 2: En_1024_L6c_SXU
 3: En_1024_L6c_SXA
 4: En_1024_L6c_SXA
 5: En_1024_L6c_SXA

Abbildung_A XIII: 95% Konfidenzintervalle Wiederholung des Kombinationstest



Abbildung_A XIV: 95% Konfidenzintervalle Verifikationstest