

Förderpreis 2019

Evaluation des Einflusses linearer und kompressiver Hörgeräteverarbeitungen auf die musikalische Wahrnehmung und Szenenanalyse von Schwerhörigen

Masterarbeit

Verfasser: Josef Schröder
Erstgutachter: Dr. Kai Siedenburg
Zweitgutachter: Prof. Dr. Volker Hohmann
Datum der Abgabe: 15.04.2019

EUHA

Europäische Union der
Hörakustiker e.V.

Herausgeber: Europäische Union der Hörakustiker e. V.
Neubrunnenstraße 3, 55116 Mainz, Deutschland
Tel. +49 (0)6131 28 30-0
Fax +49 (0)6131 28 30-30
E-Mail: info@euha.org
Internet: www.euha.org

Alle hier vorhandenen Daten, Texte und Grafiken sind urheberrechtlich geschützt. Eine Verwertung über den eigenen privaten Bereich hinaus ist grundsätzlich genehmigungspflichtig.

© EUHA 2019

Zusammenfassung

In der vorliegenden Masterarbeit wurde der Einfluss unterschiedlicher Hörgeräteverarbeitungen auf die wahrgenommene Qualität von Musik, die musikalische Szenentransparenz und das Sprachverstehen im Rauschen untersucht. Dafür wurden drei Versuche mit zwölf normalhörenden und elf schwerhörenden Versuchspersonen durchgeführt. Die Musikqualität wurde mithilfe eines MUSHR-Tests mit Musikexzerpten unterschiedlicher Genres, die musikalische Szenentransparenz durch einen adaptiven Melodie-Erkennungstest und die Sprachdiskrimination mit dem Oldenburger Satztest quantifiziert. Es wurde ein signifikanter Einfluss der Hörgeräteverarbeitungen auf die Musikqualität und die Sprachdiskrimination gefunden, jedoch kein Zusammenhang zwischen der wahrgenommenen Musikqualität, der musikalischen Szenentransparenz und dem Sprachverstehen.

Abstract

The purpose of the present master thesis was to further investigate the effect of different types of hearing aid processing on perceived quality of music, musical scene transparency, and speech intelligibility in noise. Twelve normal hearing and eleven hearing impaired listeners did three experiments. The music quality was measured with a MUSHR for different musical excerpts from different genres. The musical scene transparency was adaptively measured with a melody recognition test and the speech intelligibility was quantified using the Oldenburg sentence test. We found a significant effect of the hearing aid processing on perceived music quality and speech intelligibility, but no relation between music quality, musical scene transparency, and speech intelligibility.

Inhaltsverzeichnis

Impressum	i
Zusammenfassung	ii
Abstract	iii
Inhaltsverzeichnis	iv
1. Einleitung	1
2. Grundlagen	3
2.1 Musikwahrnehmung	3
2.1.1 Bedeutung und Aufbau von Musik	3
2.1.2 Qualität von Musik	3
2.1.3 Auditorische Szenenanalyse	6
2.1.4 Klangfarbe	8
2.1.5 Tonhöhe	9
2.1.6 Gegenüberstellung von Sprache und Musik	9
2.2 Hörgeräte	10
2.2.1 Geschichte und Intention	10
2.2.2 Hörgerätetechnik	12
2.3 Open Master Hearing Aid	13
3. Methodik	15
3.1 Fragestellung und Ziele	15
3.2 Probanden	15
3.3 Stimuli	17
3.3.1 Auswahl	17
3.3.1.1 Musikqualität	17
3.3.1.2 Musikalische Szenentransparenz	18
3.3.2 Versuchsaufbau	19
3.3.3 Konfiguration openMHA	21
3.4 Versuch 1: Musikqualität	24
3.5 Versuch 2: Musikalische Szenenanalyse	25
3.6 Versuch 3: Sprachverständlichkeit	27
4. Ergebnisse	29
4.1 Datenanalyse	29
4.2 Musikqualität	29
4.2.1 Musikauswahl	29
4.2.2 Ergebnisse der normal- und schwerhörenden Versuchspersonen	30
4.2.3 Diskussion normal- und schwerhörende Probanden	32
4.2.4 Einfluss des Genres auf normalhörende Probanden	33
4.2.5 Einfluss des Genres auf schwerhörende Probanden	34
4.3 Musikalische Szenentransparenz	34
4.3.1 Vergleich von normal- und schwerhörenden Probanden	34
4.3.2 Diskussion normal- und schwerhörende Probanden	36
4.4 Sprachverständlichkeit	36
4.4.1 Vergleich von normal- und schwerhörenden Probanden	36
4.4.2 Diskussion normal- und schwerhörende Probanden	39
4.5 Korrelationen	39
4.5.1 Zusammenhang von Qualität und musikalischer Szenenanalyse	39

4.5.2 Zusammenhang von Qualität und Sprachverstehen	40
4.5.3 Zusammenhang von musikalischer Szenenanalyse und Sprachverstehen	41
5. Diskussion	42
6. Zusammenfassung und Ausblick	44
7. Verzeichnisse	46
7.1 Abbildungsverzeichnis	46
7.2 Tabellenverzeichnis	48
7.3 Literaturverzeichnis	49

1. Einleitung

Musik ist für viele Menschen, ob bewusst oder unbewusst, ein täglicher Begleiter. Sie erschafft oder nimmt Emotionen, ist seicht oder harsch, in jedem Fall aber individuell. Dabei ist der Pfad der Musik vom initialen Empfinden des Komponisten bis zum Perzept durch den Zuhörer länger und diffiziler als man im Allgemeinen vermuten möchte. Neben den unendlichen Möglichkeiten der Umsetzung der kompositorischen Idee ist es letztlich auch die technische Darbietung, welche das Empfinden des Rezipienten prägt. Der technische Aspekt einer Musikwiedergabe kann simplifizierend zunächst auf wenige Segmente heruntergebrochen werden: Schallgeber, Raum, kognitive Verarbeitung und Perzept. In den meisten Fällen nehmen ein Paar Lautsprecher oder Kopfhörer die Rolle der Schallgeber ein. Dementsprechend kann der Raum in seiner Größe zwar um mehrere Potenzen variieren, schlussendlich gelangt die Musik jedoch an das Ohr des Adressaten. Auf dieser Stufe muss die erste Differenzierung bezüglich der Musikhörer geschehen. Im Fall eines normalhörenden Menschen gelangt das Schallereignis nun mehr oder weniger direkt an das Trommelfell beider Ohren, wird mithilfe multipler Verarbeitungsprozesse entschlüsselt und bildet abschließend ein Perzept. Tritt hingegen eine Hörschädigung auf, nutzt man im besten Fall Hörgeräte, um möglichst geringe Einschränkungen in der alltäglichen Kommunikation zu erfahren. Einige Hörgeräteträger berichten jedoch von einer eingeschränkten Qualität beim Hören von Musik mithilfe ihrer Hörgeräte.

Die vorliegende Arbeit will die empfundene Qualität beim Hören von Musik mit verschiedenen Hörgeräteverarbeitungen in den Kontext der musikalischen Szenenanalyse setzen. Dabei soll der Fokus jedoch nicht vollständig von der Diskriminationsfähigkeit von Sprache abgewandt werden. Vielmehr soll eine initiale Betrachtung verschiedener Hörgeräteverarbeitungen im Hinblick auf mögliche Interaktionen der drei Attribute Qualität, musikalische Szenentransparenz und Sprachverständlichkeit erfolgen.

Dazu wurden fünf verschiedene Hörgeräteeinstellungen ausgewählt. Es wurden eine breitbandig, zwei kompressibel und zwei linear verstärkende Einstellungen untereinander sowie mit einer Referenzkondition ohne Signalverarbeitung verglichen. Die Verarbeitung der verwendeten Stimuli erfolgte dabei durch das open Master Hearing Aid (openMHA). Die Stimuli wurden im Vorhinein in einer typischen stereophonen Abhörsituation durch Gehörgangsmikrofone eines Kunstkopfes sowie omnidirektionale Hörgerätemikrofone aufgenommen und im Anschluss der Hörgeräteverarbeitung summiert. Diese Stimuli wurden in drei verschiedenen Experimenten zwölf normalhörenden und elf schwerhörenden Versuchspersonen präsentiert. Die schwerhörenden Probanden besitzen alle einen ähnlichen mittelgradigen, sensorineuralen Hochtonhörverlust.

Für die Bewertung der Musikqualität fand ein *multi-stimulus test with hidden reference* Anwendung. Bei diesem bewerteten die Versuchspersonen die sechs Verarbeitungskonditionen bei neun Musikexzerpten unterschiedlicher Genres hinsichtlich ihrer empfundenen Klangqualität auf einer Skala von 0 („am schlechtesten“) bis 100 („am besten“).

Das Experiment zur musikalischen Szenentransparenz forderte das Heraushören und Wiedererkennen einer einfachen Melodie aus einer simplen Begleitung. Dabei wurde der Pegel der Melodie nach der 1-up-2-down-Methode adaptiert. Die verwendeten Stimuli in Gestalt einer Klarinette (Melodie) und eines Cellos (Begleitung/Maskierer) wurden ebenfalls mit dem openMHA verarbeitet.

Die Sprachverständlichkeitsmessung der Versuchspersonen bei den verschiedenen Hörgeräteverarbeitungen wurde mit dem Oldenburger Satztest (OLSA) durchgeführt.

Kapitel 2 gibt einen Überblick über die theoretischen Grundlagen, die für das Verständnis der Zusammenhänge der betrachteten Aspekte der Musikwahrnehmung als wichtig erachtet werden. Kapitel 3 schildert die Intentionen bestimmter methodischer Entscheidungen und ihre Umsetzungen in den einzelnen Experimenten. Kapitel 4 widmet sich der Betrachtung und Auswertung der erfassten Daten, welche in Kapitel 5 zusammenfassend diskutiert werden. Abschließend fasst Kapitel 6 die Arbeit zusammen und gibt einen Ausblick auf mögliche und sinnvolle Folgebetrachtungen.

2. Grundlagen

2.1 Musikwahrnehmung

2.1.1 Bedeutung und Aufbau von Musik

Musik ist seit Langem ein fester Bestandteil unserer Gesellschaft, und Funde von Musikinstrumenten reichen bis in die Bronzezeit zurück (Keil, 2018). Die Verfügbarkeit von Musik ist durch mp3-Dateien, Streamingdienste und Smartphones heute umfangreicher als je zuvor, und das Musikhören ist für viele Menschen zu einem täglichen Begleiter geworden. Dabei stellt die Qualität der wahrgenommenen Musik einen wichtigen Faktor dar, welcher in Kapitel 2.1.2 aus verschiedenen Sichtweisen betrachtet und definiert wird.

Beim Musikhören nehmen wir ein Musikstück zwar als ein einheitliches Konstrukt wahr, fokussieren uns in vielen Fällen aber auf einen bestimmten Aspekt, beispielsweise die tragende Melodie oder den zugrunde liegenden Rhythmus. Diese musikalischen Objekte, welche oft parallel auftreten, aber dennoch separat erfasst werden können, werden anhand bestimmter Faktoren gruppiert und verschiedenen auditorischen Strömen zugeordnet (siehe Kapitel 2.1.3). Die Separation und das Verfolgen dieser Ströme ist Teil der musikalischen Szenenanalyse und wird nachfolgend als musikalische Szenentransparenz bezeichnet.

Befragungen von schwerhörenden Menschen bezüglich des Einflusses ihres Hörverlustes auf den Genuss von Musik zeigten, dass ca. ein Drittel der Befragten negative Auswirkungen ihres Hörverlustes auf die Musikwahrnehmung empfinden (Leek, Molis, Kubli und Tufts, 2008). Zur Quantifizierung dieser Auswirkungen eines Hörverlustes auf die Musikwahrnehmung gab es bereits Bestrebungen, einen standardisierten funktionalen Test zu entwickeln, welcher verschiedene niedrigdimensionale Aspekte musikalischer Wahrnehmung thematisiert (Kirchberger und Russo, 2015). Dieser *Adaptive Music Perception*-Test konzentriert sich allerdings vorwiegend auf grundlegende Faktoren wie den Pegel, die Dauer oder die wahrgenommene Tonhöhe (*pitch*) eines musikalischen Ereignisses und zielt deshalb nicht konsequent auf die Quantifizierung der musikalischen Szenentransparenz ab.

Aus diesem Grund sollen die musikalische Szenenanalyse und -transparenz hier thematisiert und ihre Relation zu der wahrgenommenen Audioqualität untersucht werden. Außerdem wird die Tonhöhen- und Klangfarbenwahrnehmung definiert und dargestellt, wie sie für das Design eines Experiments zur Quantifizierung der Szenentransparenz nutzbar gemacht werden kann.

2.1.2 Qualität von Musik

Ziel dieser Arbeit ist es, eine Aussage über den Zusammenhang von Qualität und Szenentransparenz bei verschiedenen Hörgeräteverarbeitungen zu treffen. In diesem Kapitel wird der Begriff Qualität allgemein und im Kontext der Musikwahrnehmung beschrieben.

Der Begriff Qualität hat seinen Ursprung im lateinischen Wort *qualitas* (Beschaffenheit, Merkmal, Eigenschaft) und wird in der DIN EN ISO 9000 beschrieben als:

„Vermögen einer Gesamtheit inhärenter Merkmale eines Produkts, eines Systems oder eines Prozesses zur Erfüllung von Forderungen von Kunden und anderen interessierten Parteien“ (DIN EN ISO 9000:2015-11)

Im Kontext der Musik lassen sich dagegen andere Definitionen heranziehen, z. B. eine ursprünglich für Sprachqualität formulierte Definition, welche durch (Blauert und Jekosch, 2007) auf den Begriff der Klangqualität (*sound quality*) abgeändert wurde (durch das Ersetzen des Wortes *Speech* durch *Sound*):

„Sound Quality is the result of an assessment of the suitability of a sound item, considering all of its recognized and nameable features and feature values. The assessment specifies as to which amount the sound item under consideration complies with a reference arising from aspects such as individual expectations and/or social demands and/or pragmatic necessities – again considering all recognized and nameable features and feature values.“ (Jekosch, 2006)

Blauert und Jekosch beschreiben das Phänomen der Klangqualität als komplex und mehrdimensional und entwickelten ein Modell, mithilfe dessen der Prozess der Bildung eines Qualitätseindrucks strukturiert werden kann (Blauert und Jekosch, 2012). Dabei teilen sie die beteiligten Prozesse nach dem Grad der beteiligten Abstraktion ein und identifizieren vier verschiedene Qualitätsebenen: auditive Qualität, aurale Szenenqualität (*aural-scene quality*), akustische Qualität und die aural-kommunikative Qualität (*aural-communication quality*). Wie das Modell als Ganzes folgt auch die Kennzeichnung und Reihenfolge dieser unterschiedlichen Ebenen dem konstruktivistischen Modell der Realität. Dieses beruht grundsätzlich auf der Idee der prozessualen Entstehung von Wirklichkeit (Pörksen, 2011).

Die auditive Qualität behandelt dabei auditorische Ereignisse und ihre Attribute (z. B. Lautheit, *pitch*, *timbre*), welche durch klassische psychoakustische Experimente quantifiziert werden können. Die Beurteilung dieser Attribute durch Probanden geschieht immer auf der Basis eines Bias, also der durch Erfahrung entwickelten subjektiven kognitiven Verzerrung. Die Beurteilung von Klangqualität benötigt jedoch eine Referenz, welche nach dem Modell von Blauert und Jekosch womöglich in der Analogie zu den Gefühlen, welche bei der Wahrnehmung eines auditorischen Ereignisses entstehen, zu suchen ist. So würde ein Klang, der für ein positives Gefühl steht, hoch, ein Klang mit einer negativen Assoziation dagegen eher niedrig bewertet werden.

Eine Abstraktionsebene höher wird die aurale Szenenqualität am Beispiel der Musik beschrieben. Dies geschieht mithilfe (für die empfundene Qualität wichtiger) Attribute wie dem räumlichen Eindruck, der Immersion und der perzeptuellen Plausibilität. Die Bewertung dieser Merkmale bedarf (verglichen mit einem einzelnen auditorischen Ereignis) einer höheren Abstraktion und ist kontextabhängig. Sie wird als die Qualität dargestellt,

die beispielsweise bei der Evaluation von Audioequipment oder -codecs bewertet wird. Auf dieser Ebene wird die Form eines auralen Objektes oder einer auralen Szene, die durch diese technischen Maßnahmen geprägt werden, bewertet, nicht aber der Inhalt (z. B. einer musikalischen Komposition). Es wird deutlich, dass die Bewertung verschiedener Musikausschnitte und Verarbeitungskonditionen durch Versuchspersonen (wie in dieser Studie geschehen) nach diesem Modell auf ebendieser Ebene der auralen Szenenqualität liegt.

Die akustische Qualität wird im Modell von Blauert und Jekosch durch messtechnisch erfassbare Parameter wie Schalldruckpegel, Impulsantworten oder Übertragungsfunktionen beschrieben und meint damit nicht die Bewertung einer akustischen Szene, z. B. Musik, als Perzept.

Auf der höchsten Abstraktionsebene wird die aural-kommunikative Qualität gesehen, bei der auditorische Objekte als Träger von Informationen begriffen werden. Diese Informationen formen in dem Hörer ein Bild der Welt, geben ihr eine Bedeutung und erzeugen Gefühle. Beispielhaft können die Klangqualität eines Produkts und die Emotionen, die es in einem Subjekt auslöst, genannt werden.

Abseits des Konstruktivismus stellen auch Möller und Raake (2014) heraus, dass Qualität eine Wahrnehmung und einen Bewertungsprozess durch einen Rezipienten voraussetzt. Dabei vergleicht dieser ein wahrgenommenes Ereignis mit einer bis dahin unbekanntem Referenz. Dabei ist die wahrgenommene Komposition nicht zwingenderweise eine „permanente Charakteristik“ eines Objektes, sondern wird durch die Referenz maßgeblich beeinflusst. Durch diesen permanenten Vergleich jeglicher Perzepte ist Qualität relativ und geschieht als ein „qualitatives Ereignis“ in einem bestimmten räumlichen, zeitlichen und funktionalen Kontext.

Neben dem Begriff der Qualität findet sich auch der Ausdruck *Quality of Experience* (QoE) – hauptsächlich im Bereich der medialen Übertragungssysteme und -dienstleistungen.

„Quality of Experience (QoE): Degree of delight of the user of a service. In the context of communication services, it is influenced by content, network, device, application, user expectations and goals, and context of use.“

Diese Definition wurde durch Teilnehmer aus der Industrie und dem akademischen Bereich bei dem Dagstuhl-Seminar 09192 im Mai 2009 entwickelt und wird seitdem (obwohl nicht offiziell publiziert) häufig als Alternative zu der Definition nach der Norm DIN EN ISO 9000 verwendet (Möller, 2017).

Geht man dem Qualitätsbegriff hinsichtlich der Audioqualität beispielsweise von Übertragungsanlagen nach, finden sich bereits früher aufgestellte generelle Aussagen im Hinblick auf favorisierte Übertragungsfunktionen. Eine „glatte“ spektrale Einhüllende einer breitbandigen Übertragungsfunktion wurde mehrfach als erstrebenswert herausgestellt

(Bücklein, 1962; Gabrielsson, Lindström und Till, 1991; Toole, 1986a, 1986b; Toole und Olive, 1988).

Moore und Tan (2003) untersuchten den Einfluss verschiedener spektraler Verläufe, Unebenheiten (*Welligkeit*) und Bandbreiten. Dabei zeigte sich, dass eine höhere Dichte und Tiefe spektraler „Riffel“, die Kombination dieser mit zunehmender Bandbreite sowie schräg verlaufende Spektren die wahrgenommene Qualität der Versuchspersonen negativ beeinflussen. Für die Wahrnehmung von Musik (Jazz: Piano, Bass, Schlagzeug) stellte sich die maximale Bandbreite (55-16854 Hz) als qualitativ am höchsten bewertet heraus. Eine Anhebung der unteren und eine Verringerung der oberen Grenzfrequenz führten zu einer deutlichen Qualitätssenkung. Die perzipierte Qualität von Sprache hingegen litt erst bei größeren Beschränkungen der Bandbreite (208-10869 Hz). Die Stimuli wurden dabei über Kopfhörer wiedergegeben, und der Proband bewertete nach jeder Präsentation (Sprache: 3,1 Sekunden, Musik: 7,3 Sekunden) anhand einer 10-Punkte-Skala (1 = „sehr unnatürlich – stark gefärbt“, 10 = „sehr natürlich – ungefärbt“) die wahrgenommenen Qualität.

Überträgt man diese Beobachtungen (der erweiterten Bandbreite) auf die empfundene Klangqualität von Schwerhörenden bei der Nutzung ihres Hörsystems oder Cochlea-Implantats (CI), erklärt sich die Steigerung des Qualitätsempfindens durch bimodale Stimulation (kombinierte Stimulation durch CI und Hörgerät) im Vergleich zu der alleinigen elektrischen Stimulation durch ein CI (Sucher und McDermott, 2009). Eine bimodale Stimulation erweitert das Spektrum vorwiegend in den niedrigen Frequenzen und führt unter anderem deshalb zu einem qualitativ hochwertigeren Perzept.

In einer Online-Umfrage durch Madsen und Moore (2014) stellten sich als Gründe für eine als nicht ausreichend empfundene Musikqualität Verzerrungen, akustische Rückkopplungen, zu geringe oder zu hohe Verstärkungen der Hörgeräte sowie eine unausgeglichene spektrale Gewichtung und generell reduzierte Klangqualität heraus.

Untersuchungen zur Bewertung der rezipierten Qualität von Musik in Abhängigkeit verschiedener Hörgeräteverarbeitungen durch Arehart, Kates und Anderson (2010) zeigten generelle Einflüsse durch die Signalverarbeitung (nicht-lineare Verarbeitung, lineare Filterung und Kombinationen dieser mit Rauschen). Auch zeigten sie einen signifikanten Effekt verschiedener Musikgenres auf und rieten in Ergänzung zu Davies-Venn, Souza und Fabry (2007) dazu, bei Testverfahren zur Musikqualität nicht nur einen einzigen Stimulus zu betrachten. Aus diesem Grund finden in dieser Studie verschiedene Musikausschnitte aus unterschiedlichen Genres Verwendung (siehe Kapitel 3.3.1).

2.1.3 Auditorische Szenenanalyse

Als auditorische Szene kann das gesamte Bild aller akustischen Ereignisse beschrieben werden, die in einer bestimmten Zeit ein auditorisches Perzept bilden. Die Bestandteile dieses Bildes können unterschiedlichen Quellen entspringen, wobei der Rezipient die Segmente entflechtet und entsprechend seiner Intention klassifiziert. Die Bestrebung

dieser Klassifizierung kann je nach aktueller Situation unterschiedlich sein und verschiedene sogenannte auditorische Ströme erzeugen. Ein auditorischer Strom ist eine Verarbeitungsstufe im Prozess der vollständigen Beschreibung eines auditorischen Events (Bregman, 1990). Stellt man sich eine auditorische Szene beispielsweise in einem Café vor, ist die Hauptintention des Rezipienten höchstwahrscheinlich die Wahrnehmung und das Verstehen des Gesprächspartners.

Die Sprache bildet ein Objekt der gesamten Szene, steht dabei allerdings in Konkurrenz zu Objekten, welche das Wahrnehmungsverhältnis zwischen Nutz- und Störschall benachteiligen. Die Deklarierung des Schalls als nützlich oder maskierend geschieht durch den Empfänger, der Erfolg einer Zuordnung unterliegt jedoch nicht in jedem Fall seiner vollständigen Kontrolle. Entscheidend für die Separation der eintreffenden Informationen sind zum Beispiel das Pegelverhältnis der Schallereignisse, ihr Informationsgehalt sowie ihre temporale Separation.

Als ein Modell der Zuweisung bestimmter akustischer Objekte zu bestimmten Strömen kann das Prinzip der *exklusiven Allokation eines Hinweises* (engl. *exclusive allocation of evidence*) herangezogen werden. In den visuellen Kontext gesetzt, kann es mit einer mehrdeutigen Grafik beschrieben werden (siehe Abb. 1).

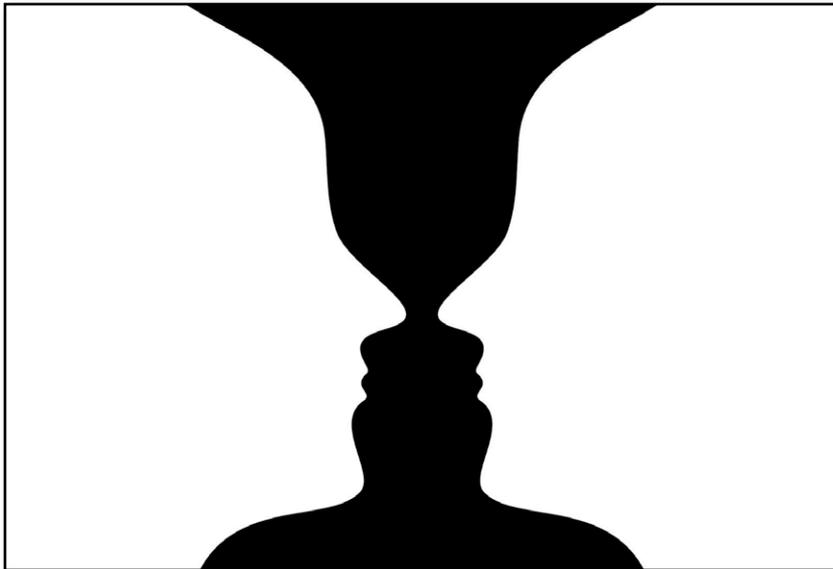


Abb. 1: Optische Vase-Gesicht-Illusion zur Veranschaulichung des Prinzips der „exklusiven Allokation“

Die in Abb. 1 erkennbaren Formen können als eine Vase oder zwei gegeneinander gerichtete Gesichter interpretiert werden. Die Abgrenzung zwischen der dunklen und den hellen Flächen und die Interpretation der Zugehörigkeit dieser werden durch das Prinzip der exklusiven Allokation, also exklusiven Zuweisung, beschrieben. Sieht man eine Vase, gehört die Abgrenzung der Flächen zu der Vase und beschreibt deren Form. Sieht man die Gesichter, verhält sich dieselbe Grenze reziprok. In keinem Fall wird die Abgrenzung beiden Formen, Vase und Gesichter, gleichzeitig zugesprochen. Das Prinzip der exklusiven Allokation sagt also aus, dass jedes sensorische Element nur einer Beschreibung gleich-

zeitig zugeordnet werden sollte. Übertragen auf akustische Ereignisse beginnt beispielsweise in den meisten Fällen kein Schallobjekt zur selben Zeit, zu der ein anderes endet. Wenn sich dementsprechend die spektrale Struktur zu einem gesetzten Zeitpunkt ändert, kann das auditorische System daraus schließen, dass wahrscheinlich nur ein bestimmtes Element des eintreffenden Schalls geendet oder seine Charakteristik verändert hat.

Bezogen auf die Szenenanalyse von Musik macht Bregman (1990) darauf aufmerksam, dass Musik oft versucht, unser auditorisches System zu täuschen, indem es uns fiktive Ströme hören lässt. In einer natürlichen Umgebung ist die auditorische Szenenanalyse darauf bedacht, präzise perzeptuelle Repräsentationen der einzelnen Klangerzeuger zu schaffen. Die gibt es auch in der Musik. Dazu gehören beispielsweise einzeln schwingende Saiten einer Geige oder eines Pianos. Dennoch bestehen viele als einzelner Strom wahrgenommene musikalische Ereignisse eigentlich aus einer Ansammlung verschiedener harmonisierender Instrumente. Um dabei als kohärentes Gebilde wahrgenommen zu werden, muss die Komposition dieses Stroms die Prozesse der Szenenanalyse überlisten, welche versuchen, das Geflecht zu entwirren und die einzelnen Klangquellen zu separieren. Dabei helfen ihnen Eigenschaften der auftretenden Schallereignisse wie die Klangfarbe (siehe Kapitel 2.1.4) und die Tonhöhe (siehe Kapitel 2.1.5). Dabei ist nicht nur eine parallele, sondern auch eine sequenzielle Gruppierung verschiedener Ströme denkbar. Eine sequenzielle Organisation ist verantwortlich für die Kohärenz von Melodien (Bregman, 1990).

2.1.4 Klangfarbe

Timbre, also die Klangfarbe eines auditorischen Objektes, fasst eine komplexe Ansammlung auditorischer Attribute sowie eine Vielzahl psychologischer und musikalischer Aspekte zusammen. Es ist eine facettenreiche Ansammlung von abstrakten sensorischen (teilweise variierenden) Eigenschaften, z. B. Anschlagschärfe, Helligkeit, Nasalität und Reichtum/Fülle. Außerdem ist es einer der primären Informationsträger für die Wiedererkennung, Identifizierung und das Verfolgen von Klangquellen und beinhaltet so die Kategorisierung eines Klanges (McAdams und Giordano, 2009).

Es sollte deshalb vermieden werden, den Begriff *Timbre* mit einem Klangereignis oder komplexen Ton gleichzusetzen (wie in einigen Veröffentlichungen geschehen), da zwischen einem akustischen Stimulus und seiner spezifischen Wahrnehmung in einem bestimmten *Timbre* komplexe Abfolgen von Informationsverarbeitungsschritten im menschlichen auditorischen System stattfinden (Siedenburg und McAdams, 2017).

Zusammenfassend kann man die 150 Jahre alte (aber nach wie vor aktuelle) Definition von Helmholtz heranziehen: „By the quality of a tone [*Klangfarbe*] we mean that peculiarity which distinguishes the musical tone of a violin from that of a flute or that of a clarinet or that of the human voice, when all these instruments produce the same note at the same pitch“ (Helmholtz, zitiert nach Siedenburg und McAdams, 2017).

2.1.5 Tonhöhe

Ein wichtiges Attribut eines akustischen Objektes, besonders in der Musikwahrnehmung, ist die Tonhöhe (*pitch*). Sie lässt sich nach Plack, Oxenham und Fay (2006) grundsätzlich von zwei verschiedenen Perspektiven betrachten: aus der Perspektive, die eine Assoziation zwischen dem *pitch* und der Tonleiter herstellt, und der, die einen generellen Bezug zur Musik vermeidet. Erstere soll hier im Kontext der Musikwahrnehmung betrachtet werden.

Die American Standard Association (ASA) beschrieb 1960 *pitch* als „that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale.“ Dabei sollte herausgestellt werden, dass *pitch* nicht als physikalisches Attribut eines Klanges, sondern als eine Empfindung beschrieben wird. Burns und Viemeister (1976) erzeugten beispielsweise ein Breitbandrauschen, das den Eindruck eines *pitch* (und sogar einer Melodie) durch sinusförmige Amplitudenmodulation hervorrief, obwohl das Spektrum eines Breitbandrauschens invariant mit einer Änderung der Modulationsfrequenz ist und damit kein Melodieperzept durch spektrale Änderungen erzeugen sollte. Der *pitch* kann also als perzeptuelles Attribut eines Klanges gesehen werden, welches zur Melodiebildung befähigt (Plack et al., 2006).

Chasin (2012) schreibt, dass der *pitch* sich nicht allein durch seine Grundfrequenz definiert, sondern auch durch den Abstand seiner harmonischen Obertöne. Eine Grundfrequenz von beispielweise 50 Hz erzeugt harmonische Obertöne bei 100, 150, 200 Hz, Dies erklärt auch, warum am Telefon zwischen einer männlichen und weiblichen Stimme unterschieden werden kann, obwohl der Übertragungsbereich üblicherweise zwischen 340 Hz und 3400 Hz liegt. Der wahrgenommene Ton bzw. die Tonhöhe wird Residualton genannt (Schouten, Ritsma und Cardozo, 1962).

2.1.6 Gegenüberstellung von Sprache und Musik

Hörgeräte sind grundsätzlich für die Verbesserung der Sprachdiskrimination konzipiert. Es ist daher sinnvoll, Schnittmengen und Unterschiede zwischen Sprache und Musik zu betrachten. Chasin (2012) schreibt des Weiteren (in Ergänzung zu Kapitel 2.1.5), dass sie sich hinsichtlich ihrer spektralen Verteilungen ähneln, sie aber auch Unterschiede besitzen. Die Grundfrequenz von Sprache, also die niedrigste vorkommende Frequenz, beträgt bei Männern ca. 100 Hz bis 125 Hz (Johnson, 2011). Die Vibration der Stimmlippen im Vokaltrakt bei tonalen Sprachanteilen (Vokale) determiniert die Grundfrequenz. Darunter ist keine Sprachenergie vorhanden. In der Musik liegt die niedrigste Frequenz, abhängig vom Genre bzw. den verwendeten Instrumenten, weit unterhalb von 100 Hz. In vielen Musikrichtungen finden sich große Energieanteile im Frequenzbereich von 50 Hz bis 60 Hz und in bestimmten Richtungen noch darunter. Diese Genres definieren sich teilweise über diese Frequenzanteile.

Die Wahrnehmung des korrekten *pitch* ist demnach auch mit einer Übertragungskette, welche die niedrigen Frequenzen nicht adäquat wiedergeben kann, möglich. Die wahrgenommene Qualität des Klanges hängt jedoch wahrscheinlich von der Art der Musik und den individuellen Präferenzen des Hörers ab. Die diesbezüglichen Zusammenhänge werden im ersten Versuch untersucht.

Betrachtet man die Spektren von Musik und Sprache im oberen Frequenzbereich, wird deutlich, dass Sprache und Musik ähnlich hohe Energieanteile über 10 kHz besitzen. Sprache besitzt hochfrequente Sibilanten (Zischlaute, z. B. „s“ wie in „Haus“), Affrikate (Verschlussreibelaute, z. B. „ts“ in „Zeh“) und Frikative (Reibelaute, z. B. „ç“ in „Ich“), wohingegen in der Musik eine Vielzahl von Instrumenten (z. B. Gesang, Gitarre, Violine, Piano, Synthesizer) hochfrequenten Schall erzeugen können.

Bezogen auf die Bewertung von Qualität und Klarheit von weiblicher und männlicher Sprache sowie Musik (Perkussionsinstrument, klassische Orchestermusik, Jazz-Trio) beschreiben Moore, Füllgrabe und Stone (2011) eine obere Grenze bei dem Einsatz eines Tiefpassfilters. Diese Grenze ist gebunden an die Steigung des Hörverlustes bei hohen Frequenzen und stellt sich in einer Präferenz für breitere Bandbreiten (7,5 kHz versus 5 kHz) bei flacheren Verläufen dar. Eine Anhebung der Bandbreite auf 10 kHz wird im Hinblick auf die Sprachverständlichkeit als nicht weiter förderlich erachtet (Moore, Füllgrabe und Stone, 2010; Moore et al., 2011).

2.2 Hörgeräte

2.2.1 Geschichte und Intention

Mit abnehmendem Hörvermögen leidet die Sprachverständlichkeit zunehmend und kann betroffene Personen aus dem sozialen Leben drängen. Das Leben mit einer Schwerhörigkeit bedarf eines starken Willens und Kraft, um weiterhin in gewohntem Maß in der Gesellschaft zu stehen und einer Vereinsamung entgegenzutreten (Richtberg, 1989).

Systeme bzw. Gerätschaften zur Verbesserung des Hörens finden seit Jahrhunderten Anwendung. Die ursprüngliche Form des trichterförmigen Hörrohres wurde ab 1800 industriell hergestellt und war bis Anfang des 20. Jahrhunderts die einzige Möglichkeit, einem Hörverlust zumindest ansatzweise entgegenzusteuern. Die darauffolgende Radiotechnik wurde mit der Erfindung des Transistors (1948) und der *integrated circuits* (1950er Jahre) zunehmend miniaturisiert, sodass eine erhöhte Mobilität die Verbreitung von tragbaren Hörhilfen vorantrieb (Mills, 2011).

Die grundsätzliche Intention von aktuellen Hörsystemen liegt in der Verbesserung der Sprachverständlichkeit. Die verbreitetste Form der Hörschädigung ist der sensorineurale Hochtonhörverlust (Cruickshanks et al., 1998). Allerdings treten insbesondere die für die Sprachverständlichkeit notwendigen Konsonanten, aber auch Formanten der Vokale (siehe Tabelle 1), in diesem Frequenzbereich auf.

Vokal	F ₁ (Hz)	F ₂ (Hz)
i:	250	2400
ɪ	363	2088
e:	331	2163
ɛ	487	1755
u:	250	668
ʊ	369	750
o:	369	713

Vokal	F ₁ (Hz)	F ₂ (Hz)
ɔ	487	975
y:	250	1563
ʏ	400	1440
ø:	375	1463
œ	500	1449
ɑ:	686	1213
a	646	1301

Tabelle 1: Die ersten beiden Formanten einer Auswahl deutscher Vokale aus artikulatorischen Kontexten (gerundete Mittelwerte nach Rausch (1972))

Abb. 2 zeigt die Hörfläche des Menschen, also den Pegel in dB SPL in Abhängigkeit der Frequenz, bei dem der Großteil aller Schallereignisse auftritt und wahrgenommen wird. Schon ein mittelgradiger Hörverlust (N3-Hörverlust nach DIN EN 60118-15 (Bisgaard, Vla-ming und Dahlquist, 2010), siehe Kapitel 3.2) beschneidet die Musik- und Sprechschallfläche und führt zu einer unvollständigen Wahrnehmung.

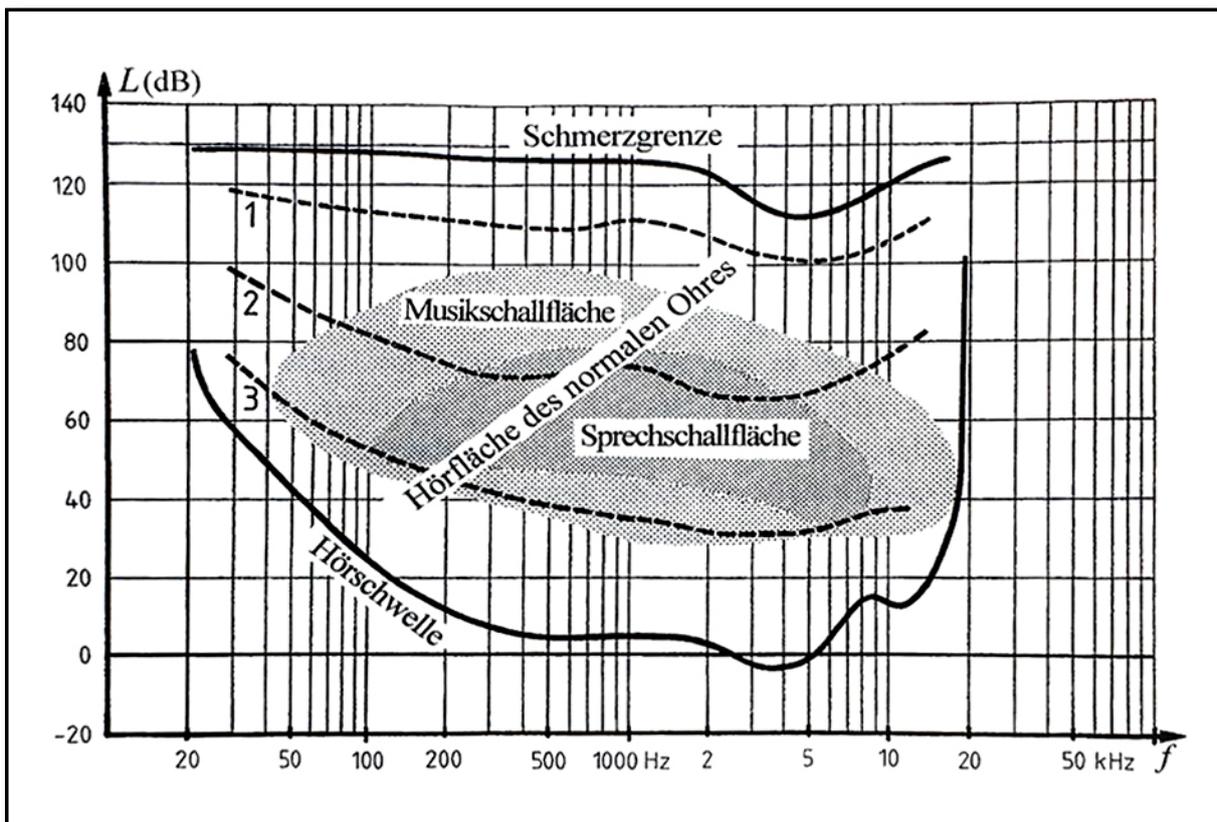


Abb. 2: Die Hörfläche des menschlichen Ohres mit der Unbehaglichkeitsschwelle (1), Isophonen angenehm lauten Hörens (2) und dem durchschnittlichen Störschallpegel (3) in Abhängigkeit der Frequenz (Pétursson und Neppert, 2002)

2.2.2 Hörgerätetechnik

Moderne Hörgeräte können eintreffende Schallereignisse frequenzspezifisch verstärken und so einen Ausgleich des Hörverlustes anstreben. Dazu ist eine Signalverarbeitung in mehreren Kanälen mit individuellen Verstärkungswerten notwendig. Ausgehend vom Audiogramm eines Hörgeschädigten verfolgen verschiedene Hersteller unterschiedliche Anpassstrategien, welche die Konfiguration eines Hörsystems bezüglich seiner Verstärkung und Kompression definieren.

Die Kompression des Schalls ist notwendig, da durch die zu höheren Pegeln verschobene Hörschwelle und die dabei gleichbleibende Unbehaglichkeitsschwelle ein geringerer Dynamikbereich zur Verfügung steht. Dieser Effekt wird als *Recruitment* bezeichnet (Fowler, 1937). Es müssen deshalb geringe Eingangspiegel stärker als hohe Eingangspiegel verstärkt werden. Ein beispielhaftes Diagramm der Ein- und Ausgangsverhältnisse ist in Abb. 3 dargestellt. Zu sehen ist die relative Verringerung des Ausgangspiegels (*output level (dB)*) bei steigendem Eingangspiegel (*input level (dB)*) ab einem festgesetzten Eingangspiegel (*threshold/Kniepunkt*). Das Kompressionsverhältnis gibt die Relation der Eingangspiegel zu den Ausgangspiegeln an. Ein Kompressionsverhältnis von 2 : 1 impliziert also bei einer Erhöhung des Eingangspiegels um 2 dB eine Ausgangspiegelerhöhung von 1 dB.

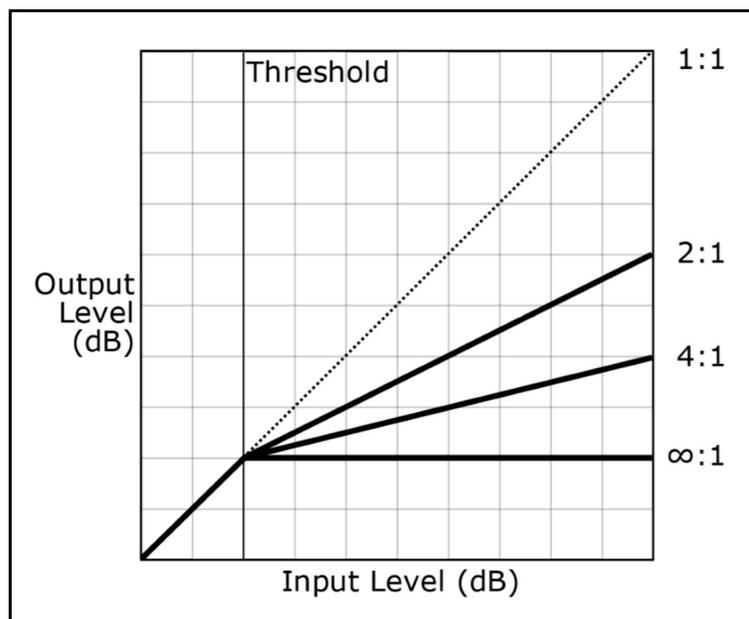


Abb. 3: Ein- und Ausgangspiegeldiagramm für verschiedene Kompressionsverhältnisse (Grafik von Fergusson (2007))

Diese Pegel Eingriffe unterliegen bestimmten Regelzeiten. Erfolgt ein steiler Pegelanstieg über den Kniepunkt, definiert die Einregelzeit (*attack time*) die Dauer bis zur Anpassung der Verstärkungswerte auf üblicherweise 63 % der angestrebten Dynamikreduktion. Die Ausregelzeit (*release time*) hingegen gibt an, innerhalb welchen Zeitraums nach dem Abfall des Eingangspiegels unter den Kniepunkt das Hörgerät mit der ursprünglichen linearen Verstärkung (keine Kompression) arbeitet (Huber und Runstein, 2013).

Der Einfluss auf die wahrgenommene Qualität von Musik der Ein- und Ausregelzeiten und der Kanalanzahl wurden durch Croghan, Arehart und Kates (2014) anhand von klassischer und Rockmusik untersucht. Dabei führten schnelle Regelzeiten bei mehrkanaligen Hörgeräten oft zu schlechter Musikqualität, wohingegen langsame Regelzeiten bei einkanali-ger Hörgeräteverarbeitung generell präferiert wurden. Die Verarbeitungskonditionen der Experimente in der vorliegenden Arbeit besitzen konstante Regelzeiten (siehe Kapitel 3.3.3) und werden daher nicht als Faktor betrachtet.

2.3 Open Master Hearing Aid

Der Vergleich verschiedener Verarbeitungskonditionen bei unterschiedlichen Versuchspersonen erfordert eine kontrollierte Anpassung von Hörsystemen bzw. der Verarbeitungskette. Da es wahrscheinlich ist, dass die Versuchspersonen im Alltag Hörsysteme unterschiedlicher Hersteller mit jeweils individuellen Einstellungen verwenden, muss eine homogenere Alternative Einsatz finden. Da zusätzlich die qualitative und komparable Beurteilung der Audioqualität verschiedener Verarbeitungen durch die Probanden ein unmittelbares Umschalten der gesamten Übertragungskette voraussetzt, ist der „klassische“ Weg der Hörgeräteeinpassung eines handelsüblichen Hörsystems keine Option. Stattdessen werden die Stimuli im Vorfeld durch das open Master Hearing Aid (openMHA) verarbeitet, gespeichert und im Probandenversuch lediglich wiedergegeben (siehe Kapitel 3.3).

Das openMHA ist Teil des Projektes „open community platform for hearing aid algorithm research“, welches durch das National Institute of Health (NIH Subvention 1R01DC015429-01) finanziert wird. Ziel des Projektes ist die Förderung und Realisierung von Forschung, welche sich der Entwicklung neuer und verbesserter Hörsysteme annimmt. Das openMHA ist eine Open-Source-Software, welche Signalverarbeitung in Echtzeit (Verzögerung < 10 ms) erlaubt. Dabei stehen verschiedene Algorithmen und Signalverarbeitungsstrategien zur Verfügung. (Quelle: openmha.org)

Die in Kapitel 2.2.2 beschriebenen für ein Hörsystem essenziellen Verarbeitungsschritte vom Ein- bis zum Ausgangssignal sind Bestandteil des openMHA und können mithilfe der hierarchisch aufgebauten Konfigurationssprache frei konzipiert werden.

Der grundlegende Aufbau des openMHA setzt sich aus vier Komponenten zusammen: Die *openMHA command line application (MHA)*, die *signal processing plugins (plugins)*, die *audio input-output modules (IO)* und die *openMHA toolbox library (libopenmha)*.

Die „command line application“ dient der Verwaltung der verschiedenen Plugins. Diese können der Signalverarbeitung dienen oder Ein- und Ausgangsmodule sein, welche unter anderem Echtzeitverarbeitung oder Dateiverarbeitung zulassen.

Jedes Plugin beinhaltet in den meisten Fällen einen Algorithmus, welcher bei Bedarf mit anderen kombiniert werden kann. Die Plugins umfassen neben anderen verschiedene

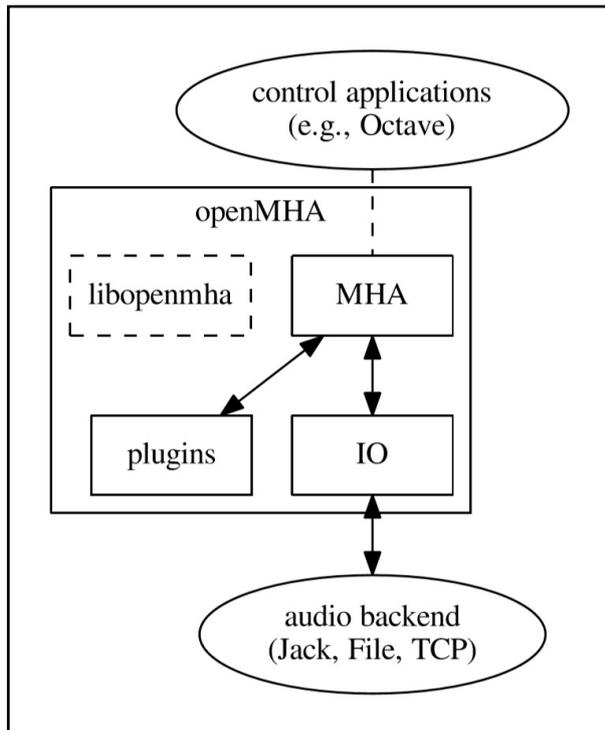


Abb. 4: Genereller, struktureller Aufbau des openMHA (Quelle: *openMHA_application_manual.pdf*)

Filterbanken, Fensterfunktionen, Algorithmen der Rückkopplungsunterdrückung, Dynamikkompressoren und Richtmikrofonalgorithmen. Das openMHA kann mithilfe von Matlab oder Octave gesteuert und konfiguriert werden. Konfigurationsdateien ermöglichen das vollständige Wiederherstellen der Verarbeitungskette. Die in den Experimenten verwendeten Signalverarbeitungen werden in Kapitel 3.3.3 beschrieben.

Die pegelabhängige Verstärkung (siehe Kapitel 2.2.2) der verwendeten Anpassstrategien wird für jede Verarbeitungskondition einmalig anhand des eingegebenen Audiogramms berechnet und in eine *gain table* geschrieben. Diese stellt eine Tabelle dar, in der für jeden Eingangspegel jedes Kanals ein definierter Verstärkungswert in dB eingetragen ist. Das openMHA ermittelt anschließend anhand der *gain table* bei der Echtzeitverarbeitung in jedem Verarbeitungsblock den Verstärkungswert des aktuellen Eingangspegels.

3. Methodik

3.1 Fragestellung und Ziele

Die Intention der vorliegenden Studie war es, grundsätzliche Aspekte des Musikhörens unter dem Einfluss verschiedener Hörgeräteverarbeitungen zu eruieren. Ein inhärentes Merkmal von Musik ist sicherlich ihre Fähigkeit, Freude und Genuss bei einem Hörer hervorzubringen. Da viele Hörgeräteträger dieses Merkmal der Musik mit ihren Hörsystemen nicht mehr in vollem Umfang wahrnehmen können, sollen hier grundsätzlich verschiedene Hörgeräteeinstellungen hinsichtlich der empfundenen Klangqualität verschiedener Musikstücke verglichen werden. Dies soll einen Anhaltspunkt darüber geben, welche Signalverarbeitungen für diesen Zweck vorteilhaft sind und welche vermieden werden sollten.

In den letzten Jahren ist ein wachsendes Interesse an dem Thema der musikalischen Szenenanalyse bzw. Szenentransparenz zu erkennen. Diese Fähigkeit, einem bestimmten musikalischen Objekt, z. B. einer Melodie, innerhalb einer musikalischen Begleitung zu folgen, soll ebenfalls Gegenstand der vorliegenden Arbeit sein und ebenfalls unter dem Faktor der Hörgeräteverarbeitung betrachtet werden.

Hörsysteme sind so konzipiert, dass sie dem hörgeschädigten Nutzer ein besseres Sprachverstehen ermöglichen können. Um zu prüfen, ob es grundlegend unterschiedliche Anforderungen an Hörsysteme für die Verarbeitung von Sprache und Musik gibt, soll auch die Diskriminationsfähigkeit von Sprache bei verschiedenen Hörgeräteverarbeitungen untersucht werden.

Diese drei Aspekte des Hörens sollen Gegenstand der Studie sein und mithilfe verschiedener Experimente mit normal- und schwerhörenden Versuchspersonen quantifiziert werden. Abschließend soll die Frage beantwortet werden, welcher Zusammenhang zwischen perzipierter Musikqualität und musikalischer Szenentransparenz besteht und welchen Einfluss verschiedene Hörgeräteverarbeitungen auf ihn haben. Da es über das Sprachverstehen mit Hörsystemen einen großen empirischen Wissensstand gibt, wird es gewissermaßen als Referenz ebenfalls mit betrachtet.

3.2 Probanden

Es werden je Experiment zwölf junge normalhörende Personen (NH) und elf ältere hörgeschädigte Personen (SH) getestet. Die hörgeschädigten Versuchspersonen besitzen alle mindestens ein Jahr Hörgeräteerfahrung und einen mittelgradigen sensorineuralen Hochtonhörverlust. Diese schwerhörenden Probanden wurden aufgrund ihres Hörverlustes ausgewählt, welcher nahe dem N3-Hörverlust nach DIN EN 60118-15 (Bisgaard et al., 2010) ist. Die maximale mittlere Abweichung des Hörverlustes der Probanden zum N3-Hörverlust beträgt links $\pm 11,8$ dB und rechts $\pm 13,6$ dB. Dieser ist in Abb. 5 und Abb. 6 als

grüne Linie dargestellt. Dabei ist zu beachten, dass der N3-Hörverlust keinen Wert bei 8 kHz angibt. Dieser wurde daher in den folgenden Darstellungen dem Trend folgend auf 70 dB *Hearing Loss* (dB HL) gesetzt.

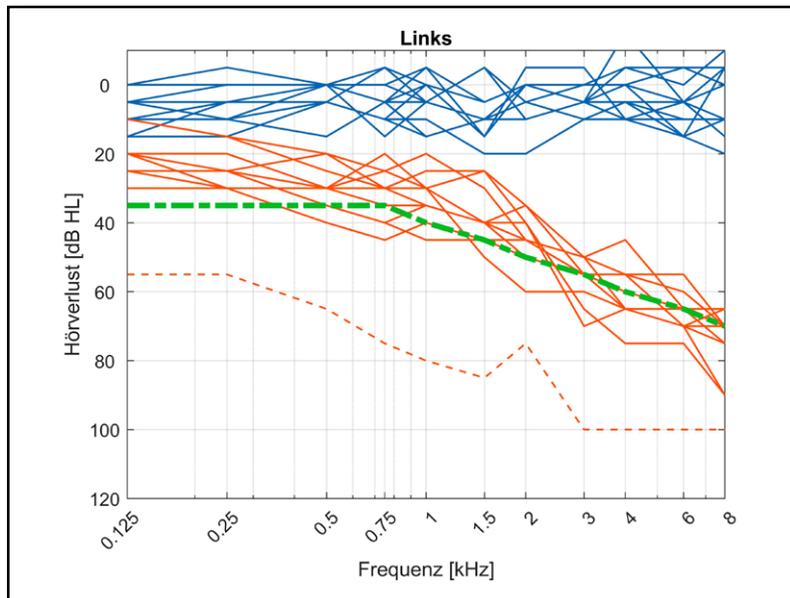


Abb. 5: Tonaudiogramm des linken Ohres der normal (blau)- und schwerhörenden (rot) Probanden sowie der N3-Hörverlust nach DIN EN 60118-15 (grün) in dB HL

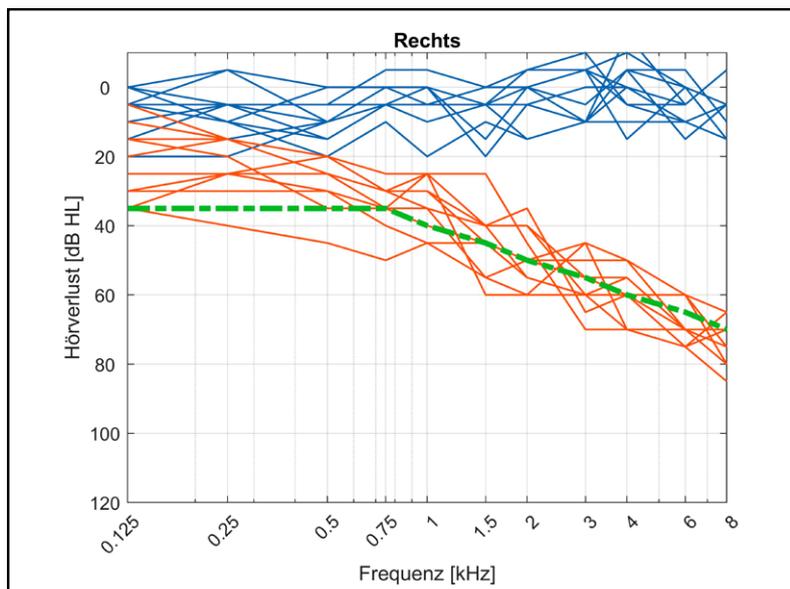


Abb. 6: Tonaudiogramm des rechten Ohres der normal (blau)- und schwerhörenden (rot) Probanden sowie der N3-Hörverlust nach DIN EN 60118-15 (grün) in dB HL

Der durchschnittliche Sinustonpegel (PTA) bei 0,5, 1, 2 und 4 kHz beträgt bei den NH 3 dB (HL) und bei den SH 48 dB (HL). Während der Tonaudiometrie ist bei einer Versuchsperson rechtsseitig ein deutlich höherer Hörverlust gemessen worden. Nach Auswertung der Studie zeigten sich bei dieser Person allerdings keine auffälligen Ergebnisse in Form von Ausreißern. Deshalb wird dieser Proband hinsichtlich der Auswertung äquivalent zu den anderen Versuchspersonen behandelt. Keiner der normalhörenden Probanden hat eine musikalische Ausbildung oder mehrjährige Erfahrung mit dem Spielen eines Instrumen-

tes. Zwei der hörgeschädigten Probanden gaben an, in der Vergangenheit über einen längeren Zeitraum ein Musikinstrument gespielt zu haben. Das durchschnittliche Alter der NH beträgt 25,4 Jahre und das der SH 75,3 Jahre. Die Teilnahme der Probanden an der Studie wurde vergütet.

3.3 Stimuli

3.3.1 Auswahl

3.3.1.1 Musikqualität

Die Auswahl der Musikausschnitte für den Versuch zur Musikqualität besteht aus neun Musikstücken. Die Musikstücke stammen größtenteils aus der ‚*Mixing Secrets’ Free Multi-track Download Library* von *Cambridge Music Technology*, welche freie Musikstücke für Bildungszwecke zur Verfügung stellt. Ausnahmen dazu bilden Musikausschnitt Nummer 2 (Lepa, Weinzierl, Maempel und Ungeheuer, 2014) sowie Nummer 4 und 6 (freemusic-archive.org). Die entsprechenden Dateien werden von unterschiedlichen Musikern und Musikgruppen verschiedener Genres hochgeladen. Jeweils drei Musikausschnitte der neun sind den Genres Jazz, Klassik und Pop/Rock zuzuordnen. Dadurch werden unterschiedliche temporale und spektrale Aspekte der Musik bedient. Innerhalb der Genres wird versucht, eine möglichst breite Auswahl zu treffen, wobei je Genre genau ein Ausschnitt Gesang enthält. Jeder Musikausschnitt ist 10 Sekunden lang.

Nr.	Titelname	Interpret	Genre	Kurzbeschreibung
1	Queens Light	Dino On The Loose	Jazz	Jazz/Funk mit Schlagzeug, E-Gitarre, E-Bass und Synthesizer
2	Nick Drake: Riverman	(Lepa et al., 2014)	Jazz	Kontrabass, Streicher und männlicher Gesang
3	Much Too Much	Selwyn Jazz: Bill Ashton	Jazz	Live-Big-Band
4	W. A. Mozart: Laudate Dominum	Advent Chamber Orchestra	Klassik	Arie (weiblicher Gesang)
5	W. A. Mozart: Piano Concerto No. 12 K414 (Allegro)	Andrey Yaroshinky & The Gnessin Academy Chamber Orchestra	Klassik	Piano mit leiser Begleitung
6	Arcangelo Corelli: Concerto in D Major Op. 6 No. 4	Advent Chamber Orchestra	Klassik	Streicher-Ensemble
7	Devil Words	Actions	Pop / Rock	Rocksong mit weiblichem Gesang und verzerrten Gitarren
8	Amber Skye	Howlin	Pop / Rock	Rhythmuslastiger Elektro-Pop
9	Scott Holmes	Inspiring Corporate	Pop / Rock	Gitarren-Pop mit Schlagzeug

Tabelle 2: Liste der verwendeten Musikexzerpte des Experiments zur Musikqualität

Die in Abb. 7 gezeigten Spektren der unverarbeiteten Stimuli sind nach Genre geordnet und wurden mithilfe eines Terzbandfilters erzeugt.

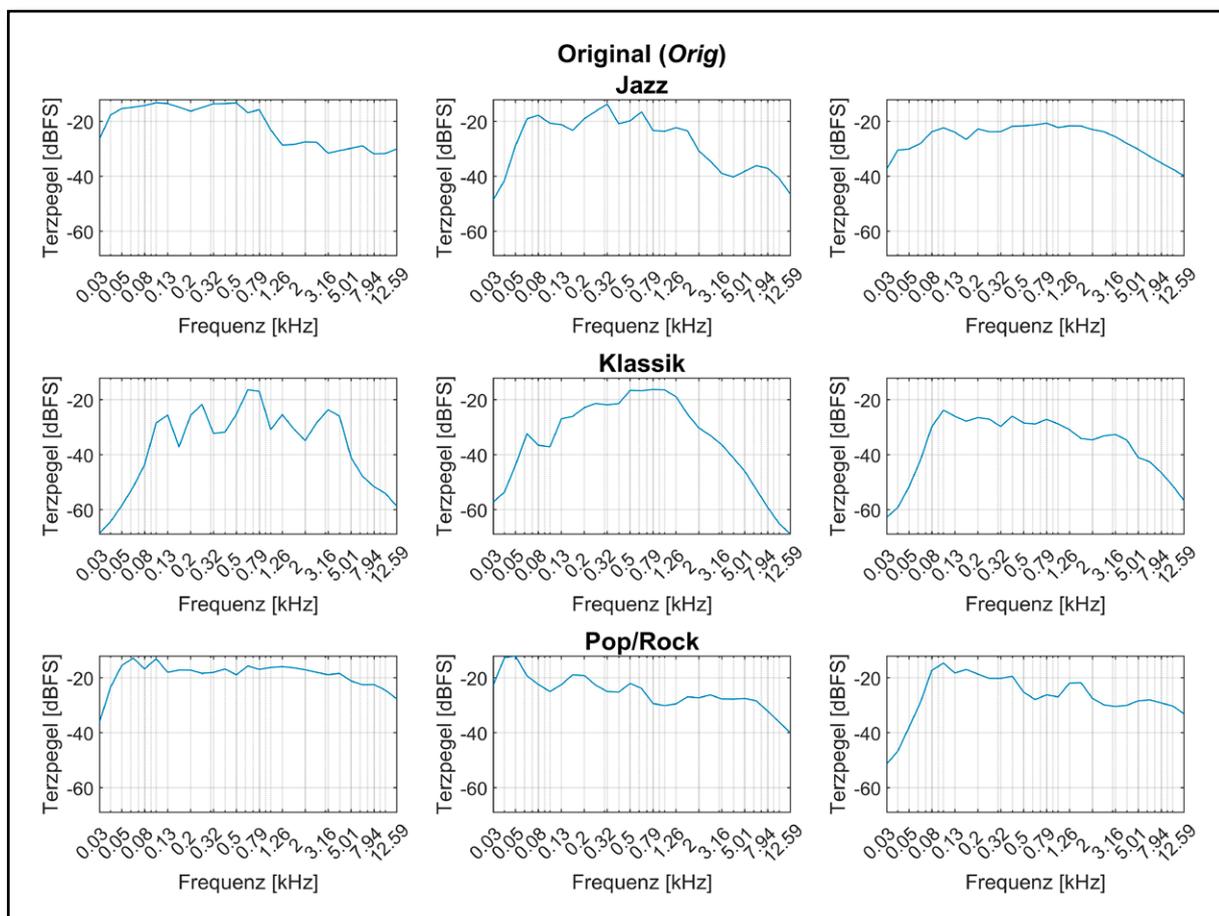


Abb. 7: Terzbandpegel in dB FS der neun unverarbeiteten Musikausschnitte (drei je Musikgenre)

3.3.1.2 Musikalische Szenentransparenz

Die Erzeugung der Stimuli für die musikalische Szenenanalyse geschieht in Anlehnung an frühere Studien der Arbeitsgruppe (Siedenburg, Röttges, Wager und Hohmann, 2019). Bei der Durchführung des Tests durch die Probanden soll eine Melodie aus einer musikalischen Begleitung herausgehört werden. Im Gegensatz zu den früheren Studien findet in dieser Arbeit eine Beschränkung der Musikinstrumente statt, sodass die Melodie von einer Klarinette und der Maskierer in Form der Begleitung von einem Cello gespielt werden. Die Generierung der Melodien und des Maskierers geschieht im Vorfeld der Messungen durch einen Algorithmus. Dieser erzeugt Melodien mit jeweils vier Einzeltönen (Viertelnoten), indem er randomisierte Variationen der Attribute *pitch*, Melodiesatz, SMR und Ordnung erstellt. Außerdem erzeugt er die Akkorde für den Maskierer (zwei halbe Noten). Die Probanden haben für die Separation der beiden Ströme (Melodie und Maskierer) neben dem Pegelunterschied und den unterschiedlichen Klangfarben der beiden dafür verwendeten Musikinstrumente also auch den temporalen Aspekt (Viertel- und halbe Noten) zur Verfügung.

Der *pitch* ist dabei als *centre pitch* zu verstehen und bildet die Basis für die Akkorde des Maskierers und die Töne der Melodie. Er reicht von D 4 (ca. 294 Hz) bis Fis 4 (ca. 370 Hz), sodass in Halbtonabständen fünf Töne als *centre pitch* zur Verfügung stehen. Für die Erzeugung des Maskierers in Form einer Begleitung werden ausgehend von jedem dieser fünf Töne sechs Akkordtypen angewandt (Grundstellung Dur/Moll, erste und zweite Umkehrung Dur/Moll).

Die Melodien basieren ebenfalls auf den fünf verschiedenen *centre pitches* und werden anhand von sechs Tonvorräten zusammengesetzt. Die Tonvorräte bestehen aus vier Tönen, welche je Tonvorrat einzigartig sind, sodass sich insgesamt 24 verschiedene Kombinationsmöglichkeiten (Ordnungen) ergeben (sechs Tonvorräte mit je vier Tönen). Die Töne der Tonvorräte sind Intervalle relativ zum *centre pitch*, wobei das größte Intervall eine Quarte ist, also fünf Halbtöne umfasst. Das Intervall 0, -2, 1, 3 besteht also beispielsweise aus dem *centre pitch* + 0, *centre pitch* - 1 Ganzton, *centre pitch* + 1 Halbton und *centre pitch* + kleine Terz.

Aus den entstandenen Tonvorräten werden nun die Melodien herausgesucht, bei denen die Kontur der Melodie maximal eine Umkehr besitzt. Eine Umkehrung der Kontur tritt auf, wenn auf eine aufsteigende Tonfolge ein tieferer Ton folgt oder ein abfallender Tonverlauf durch einen höheren Ton „unterbrochen“ wird.

Anschließend werden je SMR (-40 dB bis +20 dB) vier Ordnungen randomisiert ausgewählt und die Melodie mithilfe eines Synthesizers erzeugt. Für jeden *centre pitch* werden nun ebenfalls die sechs Akkordtypen als Maskierer mithilfe des Synthesizers erzeugt und entsprechend dem SMR einzeln mit den vier Melodien summiert und gespeichert. Diese Dateien werden anschließend durch das openMHA in allen fünf Einstellungen verarbeitet (siehe Kapitel 3.3.2).

3.3.2 Versuchsaufbau

Die in den Versuchen verwendeten verarbeiteten Stimuli werden in einem Messraum (H x B x L = 3,3 x 3,58 x 4,54 m) im diffusen Schallfeld aufgenommen. Dabei werden die unverarbeiteten Musikbeispiele und Melodien über zwei Studiomonitore (ESI aktiv08) als Stereosignal wiedergegeben und durch einen Kunstkopf aufgenommen. Die Positionierung der Lautsprecher und der beiden Mikrofone erfolgt nach Rec. ITU-R BS.775-1, sodass die Anordnung ein gleichschenkliges Dreieck der Kantenlänge 2,7 m ergibt. Die Mikrofone des Kunstkopfes und die Lautsprecher befinden sich in 1,5 m Höhe.

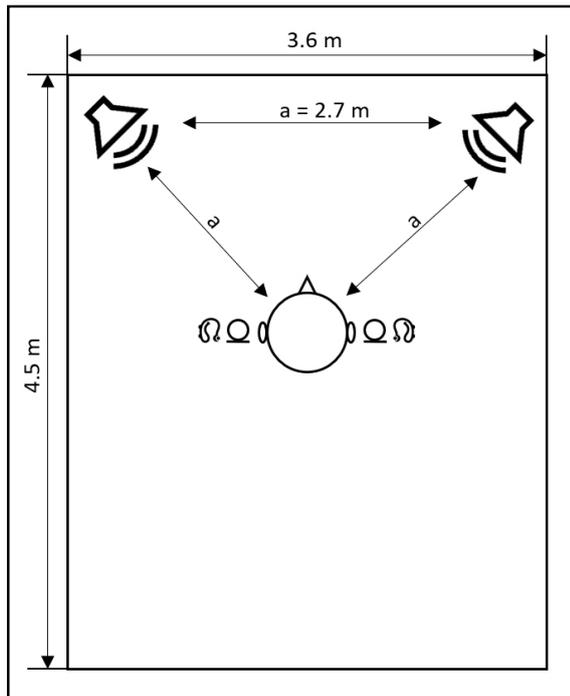


Abb. 8: Schematische Darstellung des Messaufbaus

Der Kunstkopf ist eine Nachbildung eines durchschnittlichen menschlichen Kopfes. Er besitzt zwei aus Silikon gefertigte Ohren, deren Gehörgangseingänge 13 cm Abstand zueinander haben. 2,7 cm innerhalb des Gehörganges ist jeweils ein Mikrofon positioniert. Hinter jedem Ohr befindet sich außerdem ein Hörgerät, dessen vorderes, omnidirektionales Mikrofon zusätzlich den Schall der Lautsprecher aufnimmt. Beide Mikrofone werden mit einem RME Fireface UFX Audiointerface digitalisiert.

Die Stimuluserzeugung findet in drei Phasen statt. In der ersten Phase werden alle unverarbeiteten Stimuli mit 70 dB SPL wiedergegeben und über die Gehörgangsmikrofone des Kunstkopfes in Stereo aufgenommen und gespeichert. Da keine Otoplastiken zum Einsatz kommen, wird eine offene Versorgung simuliert. Dies ist seit einigen Jahren für den in Kapitel 3.2 beschriebenen Hörverlust Standard. In Phase 2 werden alle unverarbeiteten Stimuli mit 70 dB SPL wiedergegeben, über die Mikrofone der Hörgeräte aufgenommen, mit dem open Master Hearing Aid (openMHA) verarbeitet und anschließend gespeichert. Dies wird für jede der fünf Verarbeitungsbedingungen automatisiert durchgeführt, ohne Veränderung der Messumgebung und ohne Anwesenheit einer Person im Messraum.

Die dritte Phase ist die Verarbeitungsphase. In dieser werden die per Gehörgangsmikrofon aufgenommenen Signale und die durch das openMHA verarbeiteten Signale auf ihre ursprüngliche Länge geschnitten und mithilfe einer Kreuzkorrelation zeitlich korrekt summiert. Die Pegeldifferenzen zwischen den Signalen des offenen Ohres und der verarbeiteten Aufnahmen bleiben dabei erhalten. Diese Unterschiede treten durch die frequenzspezifische Verstärkung und die Kompression des openMHA auf und finden sich so auch in einer realen Hörgeräteversorgung wieder. Um den äquivalenten Pegel in dB FS des openMHA ohne Verarbeitung, also Verstärkung oder Kompression, zu ermitteln, wird

weißes Rauschen mit 70 dB SPL über die Lautsprecher wiedergegeben, mithilfe der Hörgerätemikrofone aufgenommen, durch das neutral eingestellte openMHA geleitet und gespeichert. Aus dem Root-Mean-Square(RMS)-Wert des weißen Rauschens (WN) RMS_{WN} des so erhaltenen Rauschens $WN_{neutral}$ wird mit

$$L_{dBFS} = 20 \log_{10}(RMS_{WN} \cdot \sqrt{2}) \quad (3.1)$$

der Pegel L_{dBFS} in dB-Full-Scale (dB FS) berechnet.

Alle offenen Kunstkopfaufnahmen werden daraufhin auf den Pegel L_{dBFS} skaliert. Dafür wird mithilfe des RMS-Wertes des Rauschens RMS_{WN} und des RMS-Wertes des jeweiligen Signals (Sig) RMS_{Sig} ein Faktor α_{Sig} durch

$$\alpha_{Sig} = \left(\frac{RMS_{Sig}}{RMS_{WN}} \right)^{-1} \quad (3.2)$$

berechnet. Jedes Signal wird mit seinem individuellen Faktor multipliziert, um den Pegel L_{dBFS} zu erreichen. Durch diese Normierung der Pegel kann die oben beschriebene Pegeldifferenz zwischen dem unverarbeiteten Schall, welcher über die Kunstkopfmikrofone aufgenommen wird, und dem durch das openMHA verarbeiteten Schall konstant und realitätsnah gehalten werden.

Die auf diese Weise summierten Signale werden anschließend durch eine Cosinus-Rampe ein- und ausgeblendet (*fade in* und *fade out*), um artifizielles Knacken am Beginn und Ende der Signale zu vermeiden.

Bestimmte Verarbeitungskonditionen weisen aufgrund ihrer hohen Verstärkung im Tieftonbereich einen entsprechend hohen tieffrequenten Rauschanteil auf. Für den Vergleichstest der Melodie-Stimuli besitzt dieses Rauschen eine starke Maskierungswirkung und wirkt sich höchstwahrscheinlich ebenfalls auf die Konzentrationsfähigkeit der Probanden aus. Der tiefste vorkommende Ton im Maskierer ist ein G3, weshalb die niedrigste vorkommende Frequenz ca. 196 Hz beträgt. Jeder Stimulus des Vergleichstests wird aus den genannten Gründen deshalb durch einen Hochpass gefiltert, dessen Grenzfrequenz bei 187,5 Hz liegt. Dadurch kann das Rauschen stark eingeschränkt werden, ohne den allgemeinen Klangcharakter der verschiedenen Verarbeitungskonditionen zu verändern.

Die in allen Versuchen als Referenz genutzten originalen, unverarbeiteten Stimuli weisen im Vergleich zu den Kunstkopfaufnahmen der fünf Verarbeitungskonditionen keinen Raumanteil auf.

3.3.3 Konfiguration openMHA

Das in Kapitel 2.3 einleitend beschriebene openMHA simuliert bei der Stimuluserzeugung ein Hörsystem in verschiedenen Verarbeitungseinstellungen. In den durchgeführten Versuchen werden sechs verschiedene Konditionen verglichen, wovon eine die originalen,

unverarbeiteten Stimuli beschreibt und die anderen hinsichtlich ihrer openMHA-Konfiguration differieren.

Die fünf Signalverarbeitungskonfigurationen unterscheiden sich hinsichtlich der Parameter Verstärkung und Kompression, welche untereinander in Wechselwirkung stehen.

Die erste Konfiguration (*BB*) verstärkt breitbandig und linear in jedem Kanal um 23 dB. Dieser Wert berechnet sich durch die Verstärkungsregel $HV/2$, welche zu Zeiten der ein-kanaligen, linearen Hörsysteme gängig war. Als Hörverlust (*HV*) wurde hier der PTA aus 0,5, 1, 2 und 4 kHz angenommen. Mithilfe dieser Kondition soll der Einfluss der grundlegenden Verarbeitung des openMHA und des Messaufbaus betrachtet werden, da keine Aliasingeffekte aufgrund unterschiedlicher Verstärkungswerte in benachbarten Bändern sowie keine Effekte durch intendierte Kompression zu erwarten sind.

Die zweite und vierte Konfiguration beruht auf First-Fit-Berechnungen durch die Anpassstrategie Camfit. Die vierte Konfiguration (*LinFF*) nutzt dabei die Variante für die Berechnung von Verstärkungswerten für lineare Hörgeräte. Die Zielsetzung von Moore und Glasberg ist dabei der Lautheitsausgleich aller sprachrelevanten Kanäle, sodass in jedem Band dieselbe Lautheit erreicht wird. Dabei soll die generelle Lautheit über den gesamten Frequenzbereich komfortabel bleiben. Das Lautheitsmuster, welches durch die Anregung durch normale Sprache mit einem Pegel von 65 dB SPL erzeugt wird, ist dementsprechend flach, und die gesamte Lautheit entspricht dem Modell nach der Lautheit, die ein Normalhörender durch diesen Sprachpegel erfahren würde (ca. 23 Sone). Diese sogenannte *Cambridge formula* führt zu einem höheren, berechneten *Articulation Index (AI)* verglichen mit den Anpassregeln NAL(R), FIG6 und DSL (Moore und Glasberg, 1998). Im openMHA wurde anhand dieser Anpassstrategie ein First Fit ausgeführt, welcher gemäß der Konvention der Camfit-Berechnung ausschließlich auf der Hörschwelle (siehe Kapitel 3.2) beruht.

Die zweite Konfiguration (*CompFF*) richtet sich nach der Camfit-Variante für komprimierende, mehrkanalige Hörgeräte. Die Intentionen bezüglich der spezifischen Lautheit je kritischem Band und der allgemeinen Lautheit sind mit der zweiten Konfiguration kongruent. Zusätzlich soll Sprache mit einem Gesamtpegel von 45 dB SPL in allen Frequenzbändern von 500 Hz bis 4 kHz hörbar sein, währenddessen allerdings Kompressionsverhältnisse über 3 vermieden werden sollen. Wie in Konfiguration 3 basiert die Berechnung der Verstärkungswerte vollständig auf der Hörschwelle (Moore, Alcántara, Stone und Glasberg, 1999). Die Einregelzeit des Kompressors beträgt 20 ms und die Ausregelzeit 1 s. Die Regelzeiten sind in beiden kompressiven Hörgerätekonfigurationen kongruent.

Ausgehend von den beiden oben genannten Camfit-Berechnungen wird zusätzlich jeweils eine modifizierte Variante getestet. Variante 5 (*LinMod*) basiert auf der linearen First-Fit-Einstellung und ist empirisch auf die Verstärkung von Musik ausgerichtet. Ausgehend von der Annahme, dass die Verstärkungswerte der First-Fit-Berechnungen dazu führen, dass ein für die Hörbarkeit von Sprache ausreichender Pegel auch mindestens ausrei-

chend für die Wahrnehmung von Musik ist, wurden verschiedene additive Verstärkungswerte durch die Versuchsleitung getestet. Die Einstellungen beruhen auf den in Kapitel 2.1.6 beschriebenen Annahmen zur unterschiedlichen Hoch- und Tieftonpräferenz zwischen Sprache und Musik. Anschließend wurde durch die Versuchsleitung die Einstellung für die durchzuführenden Probandenversuche ausgewählt, welche die für diese subjektiv beste Musikqualität aufwies. Es wurde daraufhin mithilfe der grafischen Bedienoberfläche des openMHA der Tieftonbereich bei 250 Hz und 500 Hz um 15 bzw. 10 dB mehr verstärkt. Der Kanal mit der Mittenfrequenz von 8 kHz wurde um 5 dB zusätzlich verstärkt. Einstellung 3 (*CompMod*) teilt sich die Modifizierung bezüglich der zusätzlichen Verstärkung bei 250, 500 und 8000 Hz. Zusätzlich wurde hier das Compressionsverhältnis im Vergleich zur First-Fit-Einstellung (*CompFF*) halbiert.

Die in Abb. 9 dargestellten Verstärkungswerte in dB der Terzbänder zwischen 40 und 12589 Hz zeigen exemplarisch die absoluten Verstärkungen der Verarbeitungskonditionen bei einem Eingangspegel von 70 dB SPL. Es ist zu beachten, dass alle Stimuli, wie in Kapitel 3.3.2 beschrieben, nach ihrer Verarbeitung und Summierung mit den Aufnahmen der Gehörgangsmikrofone des Kunstkopfes normalisiert wurden. Deshalb geben die dargestellten Verstärkungswerte ausschließlich die Relationen der einzelnen spektralen Bereiche wieder.

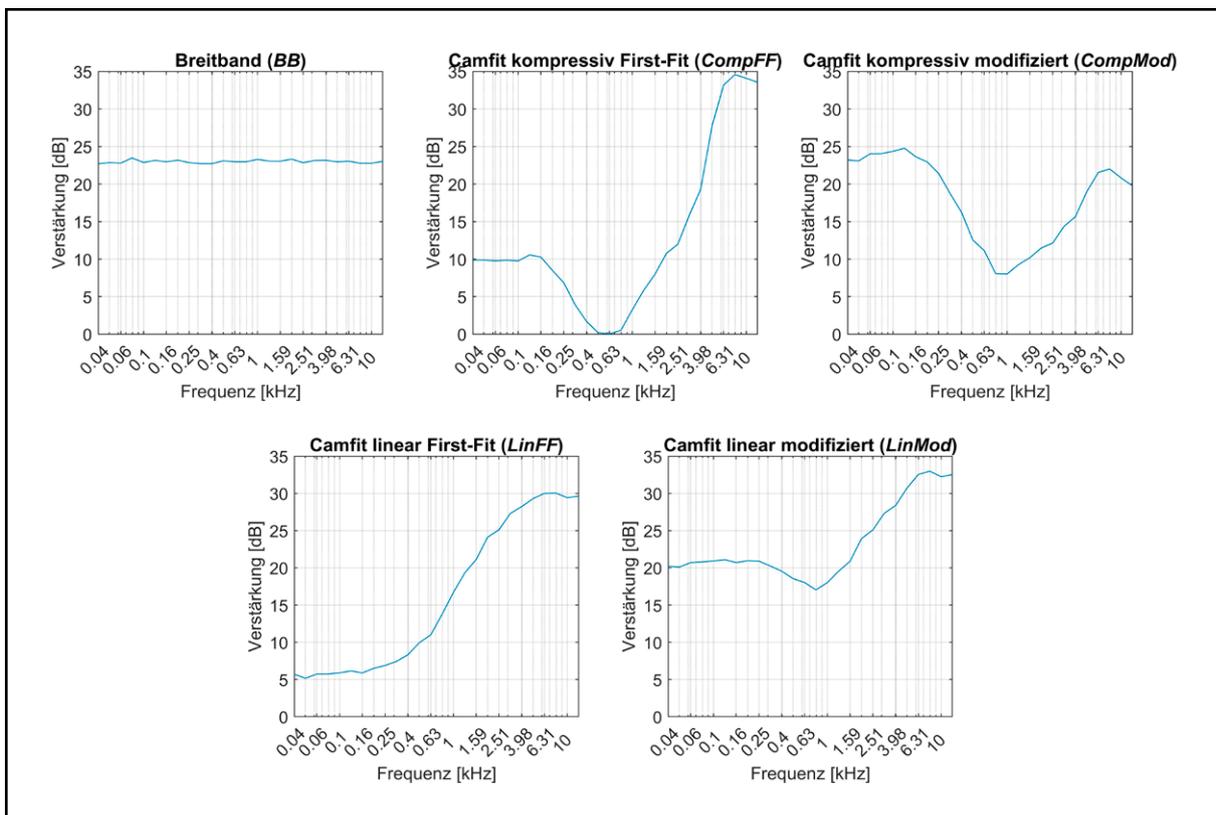


Abb. 9: Verstärkung in dB je Kondition bei 70 dB Eingangspegel

Abb. 10 stellt die Spektren der sechs Verarbeitungskonditionen nach ihrer Summierung mit den Signalen der Gehörgangsmikrofone am Beispiel des siebten Musikexzerpts (Genre: Rock) gegenüber. Dabei sind vor allem die lokalen Maxima und Minima interessant, welche bei der Betrachtung der originalen, unverarbeiteten Kondition und der Breitband-Einstellung bei ca. 150 Hz (Minimum) sowie 600 und 4000 Hz (Maxima) auffallen. Diese lassen sich auch bei den restlichen Verarbeitungskonditionen beobachten. Da bis auf die Original-Kondition alle Konditionen auf der Verarbeitung von Raumaufnahmen durch die Hörgerätemikrofone beruhen, können mögliche Gründe für die Maxima und das Minimum die Übertragungsfunktion der Hörgerätemikrofone und Raummoden sein.

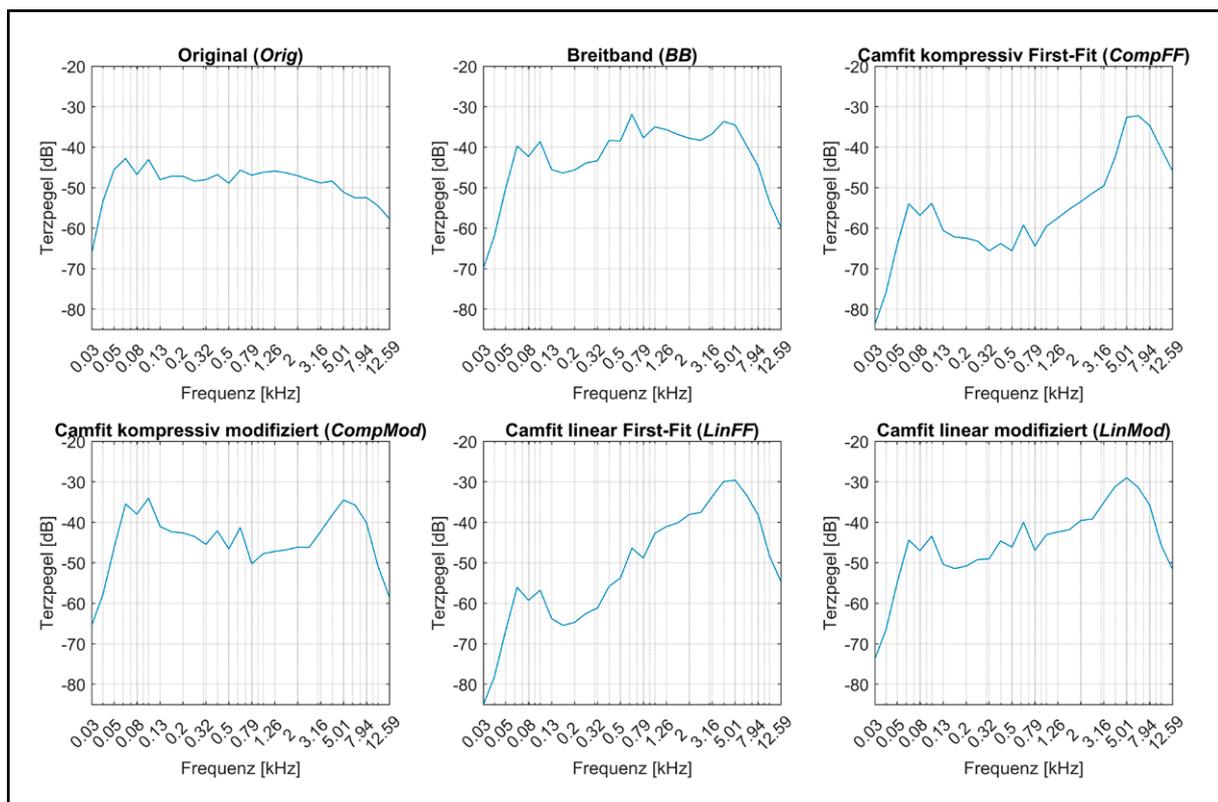


Abb. 10: Spektren der Verarbeitungskonditionen am Beispiel des siebten Musikausschnitts (Rock/Pop)

3.4 Versuch 1: Musikqualität

Für die Bewertung der Qualität der verschiedenen Verarbeitungskonditionen findet als Erstes ein Test in Anlehnung an einen MUSHRA (*multi-stimulus test with hidden reference and anchor recommendation*, ITU-R BS.1534-3 (10/2015)) Anwendung. Hierbei wird den Versuchspersonen eine grafische Bedienoberfläche zur Verfügung gestellt, mithilfe derer sie die unterschiedlichen Verarbeitungskonditionen (Sound A bis F) hinsichtlich ihrer empfundenen Qualität ordnen können. Dafür steht je Kondition ein Schieberegler zur Verfügung, der von 0 (am schlechtesten) bis 100 (am besten) in 100 Stufen verstellt werden kann. Unter jedem Schieberegler befindet sich ein Button zur beliebigen Wiedergabe des jeweiligen Stimulus. Bei jedem neuen Musikfragment müssen zunächst die Stimuli

aller Verarbeitungskonditionen durch das Drücken des Start-Buttons in voller Länge (10 Sekunden) angehört werden. Die Lautheit jedes Musikstücks und jeder Kondition wurde im Vorfeld durch drei trainierte Versuchsleiter auf das jeweilige Originalsignal (Referenz) eingestellt (nach Recommendation ITU-R BS.1534-3 (10/2015)).

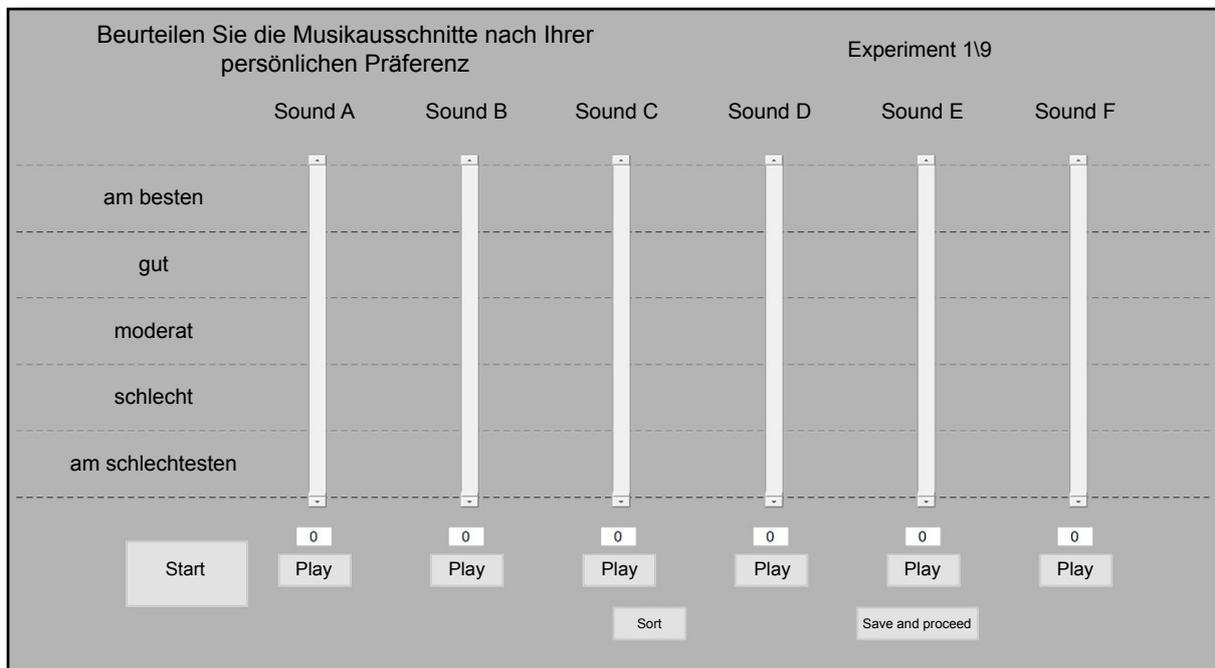


Abb. 11: Grafische Benutzeroberfläche des MUSHR-Tests

Durch die nebeneinanderliegende Anordnung aller Konditionen ist eine gute relative Bewertung möglich. Zusätzlich können die Konditionen absteigend geordnet werden. Um ein Musikstück abzuschließen, muss mindestens eine Kondition mit 100 bewertet werden. Der größte Unterschied dieser Methode zum MUSHRA ist das Fehlen eines versteckten Ankers, also eine in der Qualität deutlich schlechtere Kondition verglichen mit den übrigen. Dies wurde aufgrund der ohnehin teilweise deutlich schlechteren Qualität einzelner Verarbeitungskonditionen entschieden. Die in dieser Studie genutzte Methode wird daher im Folgenden MUSHR genannt. Die versteckte Referenz ist das originale, unverarbeitete Musikexzerpt.

Die Reihenfolge der Musikfragmente sowie der Verarbeitungskonditionen wird bei jeder Versuchsperson randomisiert dargeboten. Dieser Versuch wird als Erstes durchgeführt, damit der Proband ohne Bias bezüglich der verschiedenen Verarbeitungskonditionen Angaben zur Qualität machen kann.

3.5 Versuch 2: Musikalische Szenenanalyse

Zur Evaluation der Einflüsse der Verarbeitungskonditionen auf die musikalische Szenenanalyse wird ein 2-A 3-IFC-Test (XAB-Test) durchgeführt. Dabei werden dem Probanden drei Intervalle mit jeweils einer Melodie dargeboten. Diese werden von einer Klarinette

gespielt. Anschließend muss entschieden werden, ob das zweite oder das dritte Intervall gleich dem ersten ist. Die erste Melodie wird gleichzeitig mit einem Maskierer in Form eines Cellos (siehe Kapitel 3.3.1.2) dargeboten. Die Intervalle haben jeweils eine Dauer von einer Sekunde und werden durch Stille derselben Zeitspanne getrennt. Zwischen der Auswahl für eine Melodie durch die Versuchsperson und dem Abspielen der nächsten Intervalle liegen ca. 6 Sekunden. Diese Zeit variiert um ungefähr eine halbe Sekunde und entsteht aufgrund des Rechenaufwandes, der für das Auswählen und Zusammensetzen der Intervalle nötig ist.

Der Pegel des Maskierers bleibt konstant bei einem Pegel von 65 dB bei den normalhörenden Versuchspersonen und wird bei den schwerhörenden Probanden entsprechend des Pegels bei 25 CU (*categorical units*) der Lautheitsskalierung angepasst. 25 CU entsprechen einem mittleren/komfortablen Lautheitsperzept (Oetting, Brand und Ewert, 2014). Bei Letzteren beträgt der maximal auszugebene Ausgabepegel des Maskierers 80 dB SPL, und der maximal mögliche Signal-Maskierer-Abstand (SMR) soll nicht größer als 10 dB gewählt werden.

Die Anpassung des SMRs geschieht adaptiv nach dem 1-up-2-down-Verfahren. Dabei wird bei zwei richtigen Antworten hintereinander der Pegel des Zielstimulus verringert und bei einer falschen Antwort erhöht. Dabei beträgt die initiale Schrittweite 8 dB, welche sich nach der zweiten Umkehrung halbiert. Als Umkehrung versteht sich die Unterbrechung eines Aufwärts- oder Abwärtstrends des Zielpegels (siehe grüne Kreise in Abb. 12). Nach der vierten Umkehrung halbiert sich die Schrittweite final auf 2 dB. Nach sechs Umkehrungen beginnt die für die Auswertung verwendete Phase, welche acht Umkehrungen andauert. Der mittlere SMR dieser acht Umkehrungen wird für die Auswertung genutzt.

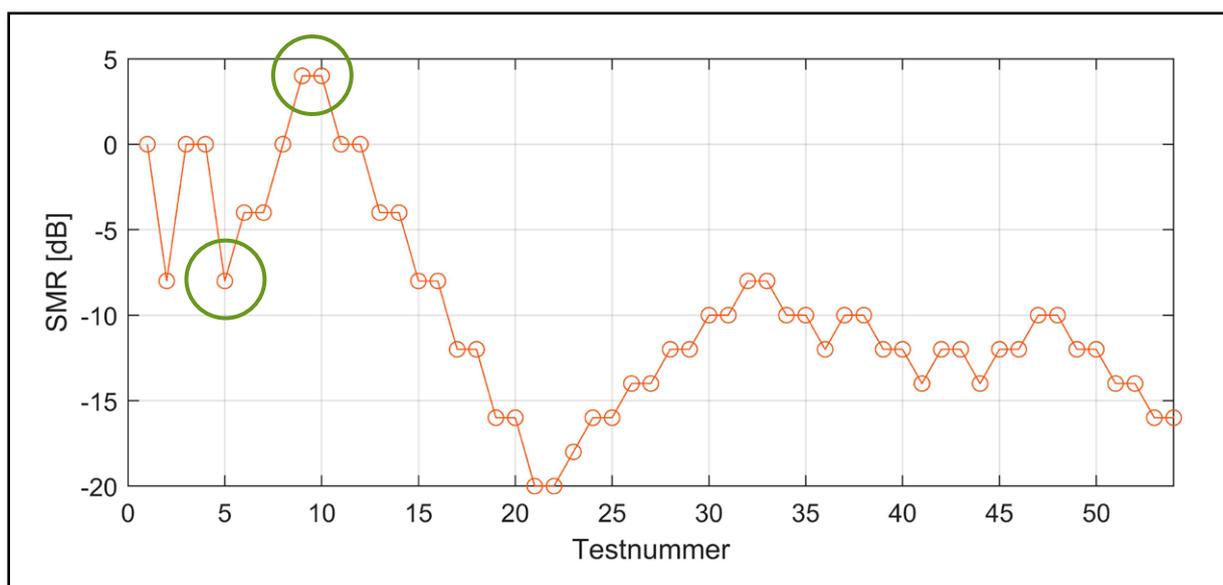


Abb. 12: Beispielhafter Antwortverlauf einer Versuchsperson für eine bestimmte Verarbeitungskondition. Adaptive SMR-Anpassung in Abhängigkeit der Antworten (rote Kreise) auf die Melodiedarbietungen als prinzipielle Darstellung des 1-up-2-down-Verfahrens. Zwei Umkehrungen sind beispielhaft grün markiert.

Auf diese Weise wird für jede Verarbeitungskondition und die originalen, unverarbeiteten Stimuli der SMR ermittelt, an dem die Wahrscheinlichkeit einer SMR-Erhöhung gleich der Wahrscheinlichkeit einer SMR-Erniedrigung ist. An dieser Schwelle sind ca. 71 % der Antworten korrekt. Dies lässt sich folgenderweise zeigen: Die Wahrscheinlichkeit einer SMR-Erhöhung P_{up} ist die Wahrscheinlichkeit, falsch zu antworten P_F (1-up), und damit gleich der entgegengesetzten Wahrscheinlichkeit, richtig zu antworten $1-P_C$. Dagegen ist die Wahrscheinlichkeit einer SMR-Verringerung P_{down} gleich der Wahrscheinlichkeit, zweimal hintereinander korrekt zu entscheiden, P_C^2 . Der zur ermittelten Schwelle korrespondierende Prozentsatz an richtigen Antworten berechnet sich also durch

$$P_{up} = P_{down} \Leftrightarrow 1 - P_C^2 = P_C^2 \Rightarrow P_C = \frac{1}{2}\sqrt{2} \approx 70\% \quad (3.3)$$

(Levitt, 1971).

3.6 Versuch 3: Sprachverständlichkeit

Um den Einfluss der verschiedenen Verarbeitungskonditionen auf die Sprachverständlichkeit zu untersuchen, wird der Oldenburger Satztest (OLSA) in der offenen Version eingesetzt. „Offen“ bedeutet in diesem Kontext, dass der Proband nicht aus einer Auswahl von Antwortmöglichkeiten auswählt. Je Verarbeitungskondition (in der Reihenfolge randomisiert) wird eine zufällige Liste von 20 Sätzen zur Ermittlung des SRTs, also des Signal-Rausch-Abstandes, bei dem 50 % der Wörter verstanden werden, durchgeführt. Als Referenz wird außerdem eine zufällige Liste von 20 unverarbeiteten (*Orig*-Kondition) Sätzen gemessen.

Die Versuchsperson wiederholt die verstandenen Wörter, und der Versuchsleiter zählt die korrekten Antworten. Basierend auf dieser Quote wird der Pegel der Sätze adaptiv verändert, während der Rauschpegel konstant bleibt.

Jeder Satz besteht aus fünf Wörtern der deutschen Sprache und weist dieselbe Satzstruktur auf (Name, Verb, Zahlwort, Adjektiv, Objekt). Jedes Element eines Satzes liegt in zehn verschiedenen Varianten vor, welche zufällig je Satz automatisiert ausgewählt werden. Durch diese zufällige Auswahl ergeben die Sätze nur in begrenztem Maße Sinn, wodurch nur eine geringe Adaption der Versuchsperson an den Satztest erfolgt und später präsentierte Listen mit den vorangegangenen komparabel sind (Wagener, Kühnel und Kollmeier, 1999).

Der Oldenburger Satztest zeigt einen Lerneffekt, welcher den SRT in sechs Listen mit je 20 Sätzen um ungefähr 2 dB absenkt. Da es in dieser Studie genau sechs Listen mit je 20 Sätzen gibt, also je Verarbeitungskondition eine Liste, wird davon ausgegangen, dass sich die Leistung der Versuchspersonen zu den späteren Konditionen hin verbessert. Deshalb wird die Reihenfolge der Konditionen über die Probanden hinweg gewürfelt, und es ent-

fällt eine Trainingsphase. Der Lerneffekt ist bei der geschlossenen Version des OLSA mit der angewandten offenen Version vergleichbar, während der SRT bei der offenen Version geringfügig niedriger liegt (Brand, Wittkop, Wagener und Kollmeier, 2004).

4. Ergebnisse

4.1 Datenanalyse

Die Auswertung der Messdaten erfolgte für alle Versuche konsistent nach nachfolgendem Schema. Für die generelle Betrachtung der Faktoren wurde eine Varianzanalyse mit Messwiederholung durchgeführt. Wenn nicht anders angegeben, liegt Sphärizität vor (Validierung durch den Mauchly-Test) und werden die p -Werte unkorrigiert angegeben. Ergibt der Mauchly-Test ein signifikantes Ergebnis, werden die p -Werte durch die Anpassung der Freiheitsgrade korrigiert.

Die *Post-hoc*-Tests werden (wenn nicht anders vermerkt) in Form gepaarter (Levenes-Test ergab in keinem Fall ein signifikantes Ergebnis) t -Tests durchgeführt. Bei multiplen Vergleichen auf den gleichen Daten müssen die resultierenden p -Werte korrigiert werden, um einer Alphafehler-Kumulierung (Fehler 1. Art) entgegenzuwirken. Basierend auf der Anzahl n der paarweisen Vergleiche werden die p -Werte mit der Bonferroni-Holm-Methode entsprechend korrigiert.

4.2 Musikqualität

4.2.1 Musikauswahl

Um die nachfolgenden Ergebnisse zu der Musikqualität und ihre möglichen Implikationen einordnen zu können, ist eine initiale Betrachtung der Songauswahl nötig. In Abb. 13 sind die Mittelwerte der Qualitätsbewertungen durch den MUSHR-Versuch über alle Konditionen für jeden Musikausschnitt dargestellt. Dabei symbolisieren die blauen Säulen die Bewertungen der NH, die roten Säulen die der SH und die grünen Säulen den Mittelwert

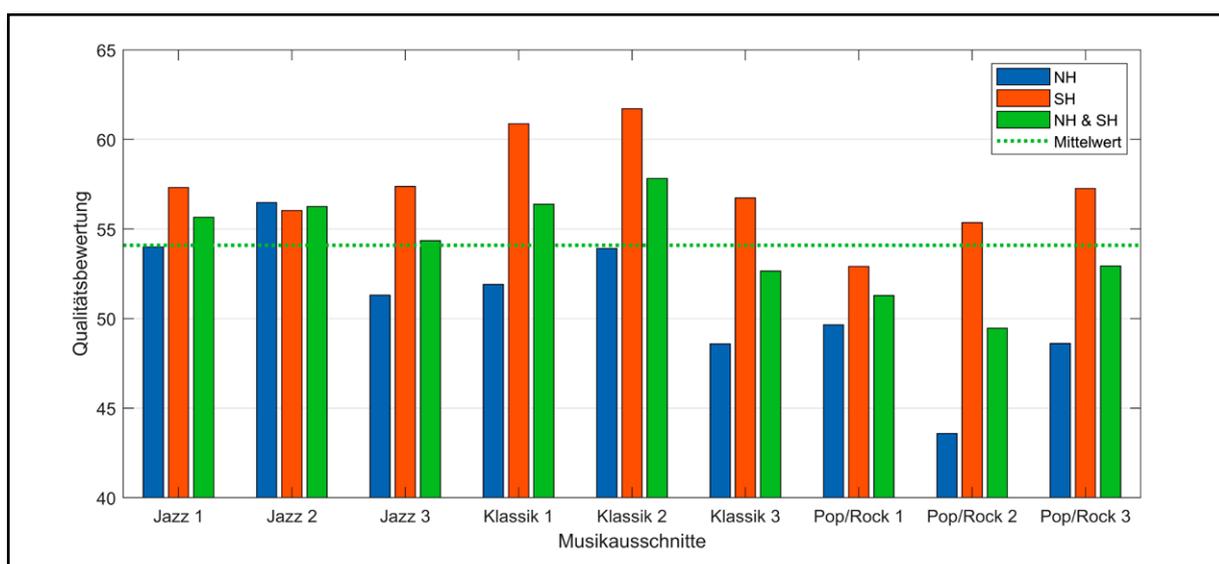


Abb. 13: Mittlere Qualitätsbewertungen der einzelnen Musikausschnitte durch die NH (blau) und SH (rot) über alle Konditionen. Der Mittelwert beider Gruppen je Musikausschnitt (grün) und der Mittelwert über alle Musikausschnitte und beide Gruppen (grüne Linie) werden angegeben.

beider Probandengruppen. Zusätzlich gibt die grüne gestrichelte Linie das globale Mittel von 54,1 Punkten über beide Probandengruppen und alle neun Musikausschnitte an.

Bis auf das zweite Musikexzerpt wurden alle Stimuli im Mittel durch die schwerhörenden Probanden um 6,5 Punkte höher bewertet als durch die normalhörenden Versuchspersonen. Die Standardabweichung der NH beträgt über alle Musikausschnitte 3,8 Punkte, die der SH 2,7 Punkte. Die Mittelwerte beider Gruppen haben eine gemeinsame Standardabweichung von 2,7 Punkten.

4.2.2 Ergebnisse der normal- und schwerhörenden Versuchspersonen

Die durchschnittlichen Ergebnisse der Qualitätsbewertungen der NH und SH sind in Abb. 14 dargestellt. Abgebildet sind die Mittelwerte aller Probanden der jeweiligen Gruppe für alle neun Musikausschnitte mit den dazugehörigen 95%-Konfidenzintervallen. Die blauen Kreise symbolisieren dabei die NH und die roten Dreiecke die SH. Ohne schwarze Kontur sind zu jedem Mittelwert die zugrunde liegenden Mittelwerte jedes Probanden über alle neun Musikausschnitte aufgetragen. Die Abszisse unterteilt die Messdaten in die Referenzkondition (*Orig* - ohne Verarbeitung durch das openMHA) und die fünf Verarbeitungskonditionen.

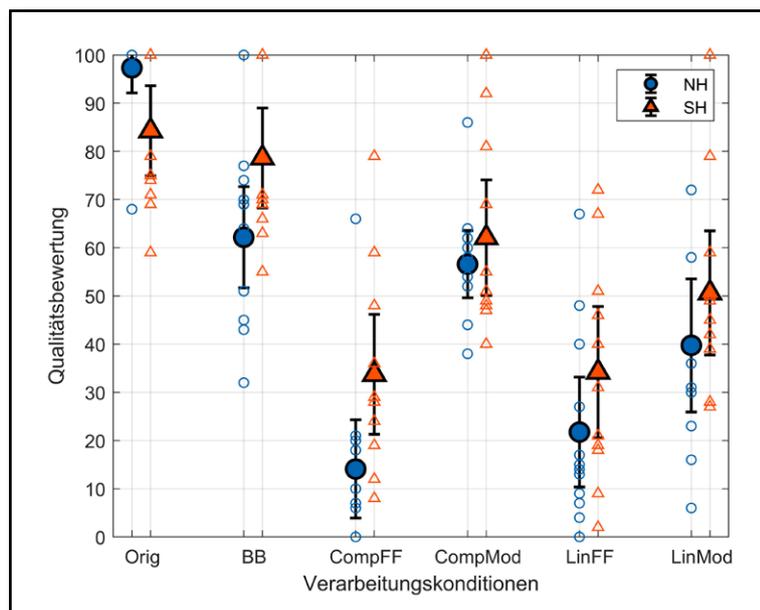


Abb. 14: MUSHR-Qualitätsbewertungen je Verarbeitungskondition, gemittelt über alle neun Musikausschnitte. Mittelwerte mit 95%-Konfidenzintervallen und den zugrunde liegenden Mittelwerten über alle Musikausschnitte je Proband für junge Normalhörende (NH) und ältere Schwerhörende (SH)

Die Mittelwerte der schwerhörenden gegenüber den normalhörenden Versuchspersonen zeigen gemittelt über alle neun Musikausschnitte in allen fünf Verarbeitungskonditionen eine durchschnittlich um 13 Punkte höhere Bewertung. Die originalen, unverarbeiteten Stimuli (*hidden reference* – versteckte Referenz) wurden dagegen von den NH im Mittel um 13 Punkte höher bewertet. Die schwerhörenden Probanden haben die Qualität dieser Kondition durchschnittlich mit 84,3 Punkten bewertet und damit 5,7 Punkte höher als die Breitband-Kondition (78,6 Punkte), was verglichen mit der Standardabweichung der Mittelwerte von 21,6 Punkten relativ gering ist. Die Gruppe der SH tendiert also dazu, die

verschiedenen Konditionen ähnlich zu bewerten, während die NH eine eher differenziertere Aussage bezüglich der Qualität der Konditionen tätigen.

Die Streuung der Mittelwerte über alle Konditionen ist bei den NH höher als bei den SH. Die Standardabweichung der Mittelwerte der NH beträgt 30,4 und bei den schwerhörnden Versuchspersonen 21,6 Punkte.

Eine Varianzanalyse mit Messwiederholung wurde mit den sechs Verarbeitungskonditionen (*Orig*, *BB*, *CompFF*, *CompMod*, *LinFF*, *LinMod*) als *within-subjects*-Faktor und den Gruppen (NH und SH) als *between-subjects*-Faktor durchgeführt. Die Analyse zeigt, dass es signifikante Unterschiede zwischen den Verarbeitungskonditionen ($F(5,105) = 55,1; p < 0,001; \eta_p^2 = 0,72$), nur marginal signifikante Unterschiede zwischen den Gruppen NH und SH ($F(1,21) = 3,4; p = 0,078; \eta_p^2 = 0,14$) und eine signifikante Interaktion beider Faktoren gibt ($F(5,105) = 2,8; p = 0,021; \eta_p^2 = 0,12$).

Für die *Post-hoc*-Tests dieses und der folgenden Experimente bezüglich der verschiedenen Konditionen werden diese paarweise miteinander verglichen. Es werden die beiden First-Fit-Einstellungen untereinander sowie jeweils mit ihren modifizierten Varianten verglichen. Um die Unterschiede zwischen den modifizierten Verarbeitungskonditionen zu vergleichen, werden auch diese einem *t*-Test unterzogen. Die Originalkondition fungiert in allen Fällen als Referenz, sodass mit ihr die Breitband- und kompressible First-Fit-Kondition verglichen wird. Zusätzlich wird aufgrund der ähnlichen Lage der Mittelwerte die Breitbandkondition mit der modifizierten kompressiblen Verarbeitungskondition verglichen. Insgesamt werden demnach sieben Vergleiche für jede *Post-hoc*-Analyse durchgeführt, die Anzahl wird bei der Bonferroni-Holm-Korrektur berücksichtigt. Diese Struktur und Anzahl der Vergleiche wird auch in Fällen, in denen Interaktionen anderer Faktoren mit dem Faktor der Verarbeitungskonditionen verglichen werden, beibehalten.

Betrachtet man die Konditionen über beide Gruppen hinweg, zeigt sich zunächst, dass es keine signifikanten Unterschiede zwischen den beiden First-Fit-Konditionen gibt ($t(22) = -0,98; p = 0,34$ für *CompFF* vs *LinFF*). Die Unterschiede zu ihren modifizierten Varianten sind aber jeweils signifikant ($t(22) = -8,52; p < 0,001$ für *CompFF* vs *CompMod*; $t(22) = -3,61; p = 0,008$ für *LinFF* vs *LinMod*). Die Referenzkondition *Orig* wurde bezüglich der Breitband- und der kompressiblen First-Fit-Einstellung signifikant höher bewertet ($t(22) = 3,24; p = 0,011$ für *Orig* vs *BB*; $t(22) = 10,35; p < 0,001$ für *Orig* vs *CompFF*). Die Breitband-Verarbeitungskondition wurde knapp, aber dennoch signifikant, höher bewertet als die modifizierte kompressible Kondition ($t(22) = 2,44; p < 0,047$ für *BB* vs *CompMod*). Zusammenfassend wurde festgestellt, dass es deutliche Unterschiede bezüglich der Qualitätsbewertung der Verarbeitungskonditionen gibt.

Die signifikante Interaktion der beiden Faktoren (Verarbeitungskondition und Gruppe) äußert sich darin, dass die modifizierte lineare Einstellung *LinMod* nur von den normalhörenden Probanden signifikant höher als die First-Fit-Einstellung *LinFF* bewertet worden ist

(NH: $t(11) = -4,46$; $p = 0,004$; SH: $t(10) = -1,78$; $p = 0,424$ für *LinFF vs LinMod*). Der mittlere Abstand der beiden kompressiblen Konditionen beträgt bei den normalhörenden 42,5 Punkte und bei den schwerhörenden Versuchspersonen 28,4 Punkte sowie bei den linearen Einstellungen 18 bzw. 16,5 Punkte. Die Mittelwerte der *CompMod*-Kondition und der *LinMod*-Einstellung differieren um 16,8 Punkte für die NH und 11,5 Punkte für die SH. Die Unterschiede sind für die NH signifikant und für die SH nicht signifikant (NH: $t(11) = 3,44$; $p = 0,017$; SH: $t(10) = 1,54$; $p = 0,463$ für *CompMod vs LinMod*). Weiterhin wurde die *BB*-Kondition nur von den NH signifikant niedriger bewertet als die Original-Kondition *Orig* (NH: $t(11) = 4,78$; $p = 0,003$; SH: $t(10) = 0,62$; $p = 1$ für *Orig vs BB*).

Bei der weiteren Analyse der Interaktionen der Faktoren zeigte sich jedoch auch, dass beide Probandengruppen die First-Fit-Konditionen (*CompFF* und *LinFF*) am niedrigsten bewertet haben (NH: $t(11) = 10,81$; $p < 0,001$; SH: $t(10) = 6,07$; $p < 0,001$ für *Orig vs CompFF*). Die Mittelwerte der beiden First-Fit-Konditionen betragen bei den NH 14,1 und 21,8 Punkte und bei den SH 33,7 und 34,2 Punkte. Der Unterschied zwischen diesen Konditionen ist bei beiden Probandengruppen nicht signifikant (NH: $t(11) = -1,76$; $p = 0,213$; SH: $t(10) = -0,06$; $p = 0,954$ für *CompFF vs LinFF*). Dem stehen die modifizierten Varianten dieser beiden Konditionen (*CompMod* und *LinMod*) gegenüber. Nur die kompressible Version *CompMod* wird von beiden Probandengruppen signifikant höher bewertet als ihre zugrunde liegende Basiseinstellung *CompFF* (NH: $t(11) = -10,9$; $p < 0,001$; SH: $t(10) = -3,93$; $p = 0,017$ für *CompFF vs CompMod*). Vergleicht man weiterhin die *CompMod*-Kondition mit der breitbandigen *BB*-Kondition, zeigen sich für die beiden Probandengruppen auch hier keine signifikanten Unterschiede (NH: $t(11) = 1,58$; $p = 0,143$; SH: $t(10) = 1,98$; $p = 0,377$ für *BB vs CompMod*). Es zeigt sich also, dass die schwerhörenden Probanden einige Verarbeitungskonditionen im Qualitätskontext gleichsetzen, während die Gruppe der normalhörenden Versuchspersonen eher differenziertere Bewertungen der Verarbeitungskonditionen vergeben haben.

4.2.3 Diskussion normal- und schwerhörende Probanden

Aufgrund des Testdesigns ist eine Bewertung der Originalkondition bei 100 Punkten oder knapp darunter zu erwarten. Möglicherweise war der erweiterte Bassbereich (siehe Abb. 10) der Originalkondition für die SH verglichen mit dem verstärkten (und den Hörverlust ausgleichenden) Hochtonbereich nicht relevant. Dafür spricht die ähnliche Bewertung dieser Kondition mit der Breitband-Kondition, welche einen geringeren Frequenzanteil im Tieftonbereich unter 70 Hz aufweist, ansonsten aber die neutralste Verstärkung besitzt (ausgenommen der in Kapitel 3.3.3 genannten Einflüsse). Die Einzeldaten zeigen, dass im Gegensatz zu den NH einzelne Schwerhörende auch andere Verarbeitungskonditionen als die originalen und breitbandig verarbeiteten Stimuli qualitativ als „am besten“ bewertet haben. Dies bestärkt die Vermutung, dass eine bessere Repräsentation des Hochtonbereichs für einige Schwerhörende in Bezug auf die wahrgenommene Qualität eine größere Rolle spielt als der Tieftonbereich.

Die Tendenz der SH, die verschiedenen Verarbeitungsbedingungen in einem engeren Qualitätsrahmen zu bewerten als die NH, resultiert möglicherweise aus einer durch den Hörverlust verursachten geringeren spektralen (Henry, Turner und Behrens, 2005) und temporalen (Nelson und Freyman, 1987) perceptuellen Auflösung. Dadurch könnte die Differenzierung der Verarbeitungsbedingungen eingeschränkt sein.

4.2.4 Einfluss des Genres auf normalhörende Probanden

Abb. 15 zeigt die Mittelwerte der Qualitätsbewertungen aller jungen normalhörenden Probanden in Abhängigkeit der Verarbeitungsbedingungen nach Musikgenres gruppiert. Die konturierten Symbole klassifizieren die Mittelwerte der Genres. Weiterhin sind die 95%-Konfidenzintervalle und die Mittelwerte aller Probanden je Musikausschnitt angegeben.

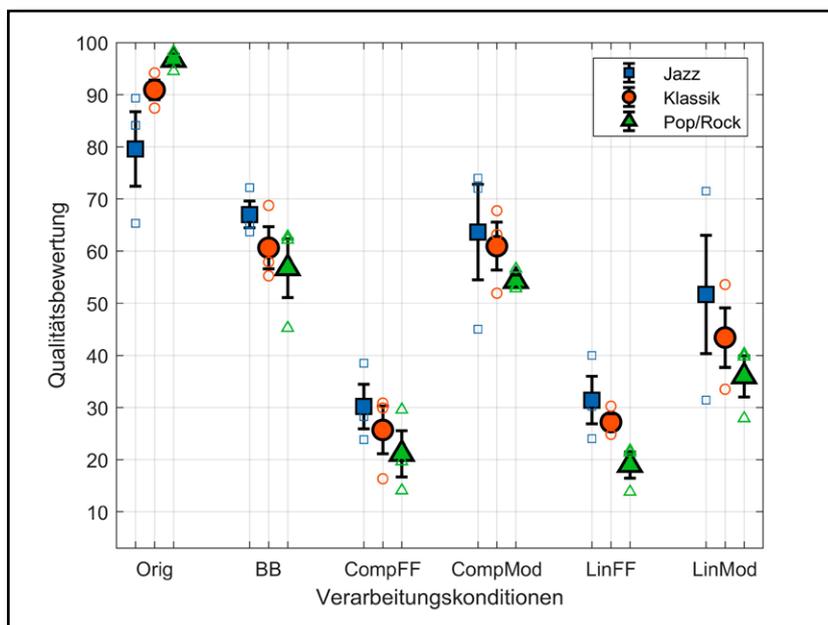


Abb. 15: MUSHR-Qualitätsbewertungen der jungen NH je Genre und Verarbeitungsbedingung. Mittelwerte mit 95%-Konfidenzintervallen und den Mittelwerten aus allen Probanden je Song

Alle fünf Verarbeitungsbedingungen zeigen den Trend, dass die Genres Jazz, Klassik und Pop/Rock eine absteigende Reihenfolge in ihrer Qualitätsbewertung aufweisen. Die Jazz-Musikbeispiele wurden durchschnittlich um 11,4 Punkte höher bewertet als die Pop/Rock-Beispiele. Dieser Trend kehrt sich bei der *Orig*-Kondition um (17,1 Punkte Abstand). Der Abstand vom niedrigsten bis zum höchsten Mittelwert je Genre beträgt im Mittel über alle Konditionen 12,3 Punkte. Die Standardabweichung der Mittelwerte der einzelnen normalhörenden Probanden beträgt im Mittel über alle Verarbeitungsbedingungen für das Genre Jazz 11,5, für Klassik 6,6 und für Pop/Rock 5,5 Punkte.

Eine Varianzanalyse mit Messwiederholung mit dem Genre (Jazz, Klassik, Pop/Rock) als Faktor zeigte, dass diese Trends nicht signifikant sind ($F(2,33) = 2,37$; $p = 0,109$; $\eta_p^2 = 0,13$). Ein Einfluss des Musikgenres auf die Qualitätsbewertung der einzelnen Verarbeitungsbedingungen ist demnach bei den normalhörenden Probanden nicht stark ausgeprägt.

4.2.5 Einfluss des Genres auf schwerhörende Probanden

Analog zu Abb. 15 zeigt Abb. 16 die Mittelwerte der älteren schwerhörigen Probanden mit den dazugehörigen 95%-Konfidenzintervallen und den Mittelwerten je Musikausschnitt. Die Daten werden nach Verarbeitungskondition und Genre gruppiert dargestellt.

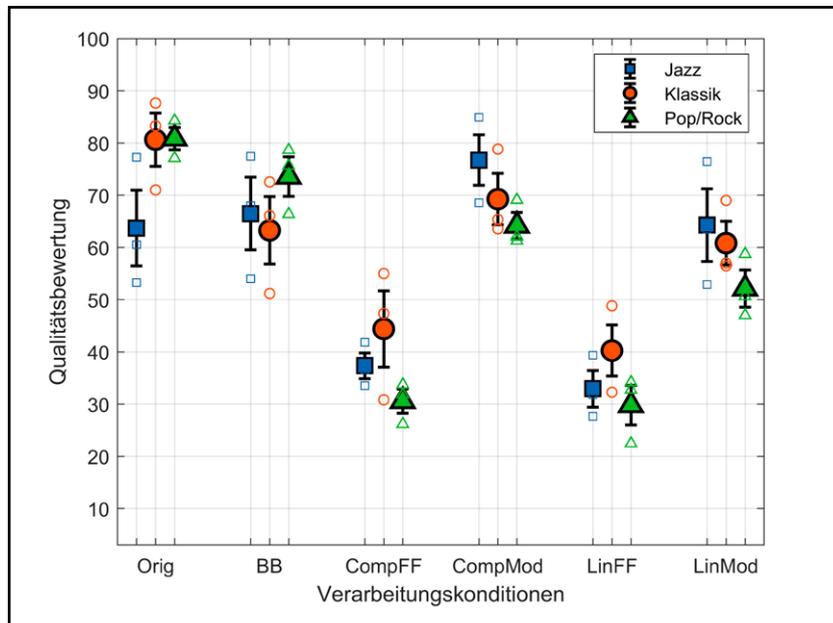


Abb. 16: MUSHR-Qualitätsbewertungen der älteren SH je Genre und Verarbeitungskondition. Mittelwerte mit 95%-Konfidenzintervallen und den Mittelwerten aus allen Probanden je Song

Anders als bei den NH zeigen die Genres keine systematische Tendenz je Kondition. Die Standardabweichung der Mittelwerte der einzelnen schwerhörenden Probanden beträgt im Mittel über alle Verarbeitungskonditionen für das Genre Jazz 9,0, für Klassik 9,3 und für Pop/Rock 5,1 Punkte. Dabei beträgt die Differenz des am niedrigsten und am höchsten bewerteten Genres im Mittel über alle Konditionen 12,7 Punkte.

Die Musikausschnitte des Genres Jazz werden in der originalen, unverarbeiteten Kondition schlechter bewertet als die der beiden anderen Genres; dies entspricht dem Niveau der *BB*-Kondition.

Jedoch zeigte auch hier (analog zu den normalhörenden Probanden) die Varianzanalyse mit Messwiederholung und dem Genre (Jazz, Klassik, Pop/Rock) als Faktor keine signifikanten Unterschiede zwischen den Genres ($F(2,30) = 2,38$; $p = 0,11$; $\eta_p^2 = 0,14$). Deshalb scheint auch für die schwerhörigen Versuchspersonen kein Einfluss der Musikgenres auf die Qualitätsbewertungen der Verarbeitungskonditionen zu bestehen.

4.3 Musikalische Szenentransparenz

4.3.1 Vergleich von normal- und schwerhörenden Probanden

Die in Abb. 17 dargestellten Signal-zu-Masker-Abstände (SMR) resultieren aus dem Experiment zur musikalischen Szenenanalyse bzw. der Transparenz. Wie in Kapitel 3.5 beschrieben, werden die SMDs der letzten acht Umkehrungen des 1-up-2-down-Verfahrens je

Proband gemittelt. Abgebildet ist das Mittel des Tests und Retests jedes Probanden für jede Kondition als Symbol ohne Kontur. Die daraus berechneten Mittelwerte und 95%-Konfidenzintervalle sind ebenfalls dargestellt. Die blauen Kreise symbolisieren die normalhörenden und die roten Dreiecke die schwerhörenden Probanden. Alle Daten werden für jede gemessene Kondition in dB angegeben.

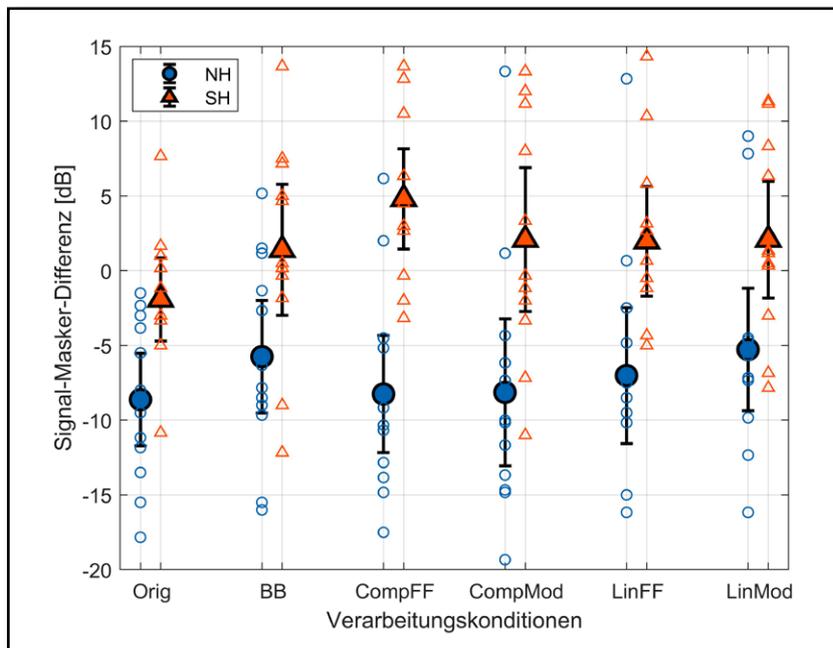


Abb. 17: Signal-Masker-Differenz (SMD) für die musikalische Szenentransparenz je Verarbeitungskondition und Probandengruppe. Mittelwerte mit 95%-Konfidenzintervalle und den zugrunde liegenden gemittelten SMDs je Proband

Mithilfe einer Varianzanalyse mit Messwiederholung mit den Verarbeitungskonditionen (*Orig*, *BB*, *CompFF*, *CompMod*, *LinFF*, *LinMod*) als *within-subjects*-Faktor und der Gruppenzugehörigkeit (NH und SH) als *between-subjects*-Faktor wurde ein signifikanter Effekt der Verarbeitungskonditionen ($F(5,105) = 2,42$; $p = 0,04$; $\eta_p^2 = 0,1$) gefunden. Der Gruppenfaktor zeigte ebenfalls einen signifikanten Einfluss ($F(1,21) = 13,46$; $p = 0,001$; $\eta_p^2 = 0,39$). Es wurde jedoch keine signifikante Interaktion der beiden Faktoren ermittelt ($F(5,105) = 1,98$; $p = 0,087$; $\eta_p^2 = 0,08$). Neben der zu dem vorherigen Experiment analogen *Post-hoc*-Analyse der Verarbeitungskonditionen wurde die Analyse des Gruppenfaktors durch ungepaarte *t*-Tests durchgeführt. Hier wurde der mögliche Unterschied beider Gruppen je Verarbeitungskondition untersucht, sodass insgesamt sechs Paarvergleiche stattfanden.

Ein marginal signifikanter Unterschied zwischen den Konditionen ließ sich nur zwischen der Referenzkondition *Orig* (NH: -8,6 dB; SH: -1,9 dB) und der Breitband- sowie der kompressiblen First-Fit-Einstellung *CompFF* (NH: -8,3 dB; SH: 4,8 dB) feststellen ($t(22) = -2,85$; $p = 0,065$ für *Orig* vs *BB*; $t(22) = -2,85$; $p = 0,057$ für *Orig* vs *CompFF*). Keine Verarbeitungskondition dieses Experiments bietet also einen signifikanten Vorteil gegenüber den anderen bezüglich der hier getesteten musikalischen Szenentransparenz.

Die Mittelwerte der SH zeigen bei jeder Verarbeitungskondition einen signifikant höheren SMR (NH vs SH: $t(21) = -3,13$; $p = 0,025$ für *Orig*; $t(21) = -2,43$; $p = 0,024$ für *BB*; $t(21) = -4,91$;

$p < 0,001$ für *CompFF*; $t(21) = -2,90$; $p = 0,026$ für *CompMod*; $t(21) = -2,98$; $p = 0,029$ für *LinFF*; $t(21) = -2,54$; $p = 0,039$ für *LinMod*). Im Mittel liegt sie 8,9 dB über denen der NH. Die geringste Differenz zwischen beiden Probandengruppen zeigt sich bei den originalen, unverarbeiteten Stimuli (*Orig*) mit 6,7 dB und die höchste Differenz bei der kompressiblen First-Fit-Einstellung (*CompFF*) mit 13,1 dB. Die Standardabweichungen der Mittelwerte über alle sechs Konditionen beträgt bei den NH 1,4 dB und bei den SH 2,2 dB. Die Standardabweichung der aus Test und Retest über alle Konditionen gemittelten SMRs ist bei den NH 7,2 dB und bei den SH 6,5 dB. Die normalhörenden sind gegenüber den schwerhörenden Probanden folglich auch bei ungünstigeren SMRs in der Lage, Zielstimulus (Melodie) und Maskierer zu separieren.

4.3.2 Diskussion normal- und schwerhörende Probanden

Die Ergebnisse des Experiments zur musikalischen Szenenanalyse bzw. der Transparenz einer musikalischen Szene zeigten vor allem eine große Diskrepanz zwischen den Gruppen der normalhörenden und der schwerhörenden Probanden bei der *CompFF*-Kondition. Eine mögliche Erklärung ist der unterrepräsentierte Frequenzbereich von ca. 200 Hz bis 1300 Hz bei dieser Kondition im Vergleich zu den anderen Konditionen, vor allem zu der originalen, unverarbeiteten Kondition (siehe Abb. 10). Dieses spektrale Minimum der kompressiblen First-Fit-Einstellung resultiert aus den Verstärkungswerten im Band um ca. 630 Hz. Es findet dort keine Verstärkung statt (siehe Abb. 9), weshalb der Ausgangspegel im oben genannten Frequenzbereich bei einigen SH möglicherweise nicht überschwellig ist und damit nicht wahrgenommen werden kann.

Der tonale Bereich der Melodien reicht von A3 bis H4, respektive 220 Hz bis 494 Hz Grundfrequenz. Er liegt demnach im oben beschriebenen Bereich des geringsten Pegels der *CompFF*-Verarbeitungskondition. Die normalhörenden Personen können aufgrund der bei ihnen überschwelligen Präsentation der Stimuli eine bessere und über alle Konditionen hinweg relativ gleichmäßige Diskriminationsleistung erreichen. Möglicherweise ist dies der Grund für die erhöhte Differenz zwischen beiden Probandengruppen bei der *CompFF*-Kondition.

4.4 Sprachverständlichkeit

4.4.1 Vergleich von normal- und schwerhörenden Probanden

Der Einfluss der verschiedenen Verarbeitungskonditionen auf die Sprachverständlichkeit wird wie in Kapitel 3.6 beschrieben mit dem Oldenburger Satztest (OLSA) durchgeführt. Die ermittelten Sprache-zu-Rausch-Differenzen, bei denen 50% der Wörter verstanden werden (SRT), sind in Abb. 18 dargestellt. Für jede Kondition sind die Mittelwerte und 95%-Konfidenzintervalle beider Probandengruppen aufgetragen. Dabei symbolisieren die blauen Kreise die jungen normalhörenden und die roten Dreiecke die älteren schwerhörenden Probanden. Die Einzeldaten der Probanden sind je Kondition und Probandengrup-

pe aufgetragen. Aus Gründen der Übersichtlichkeit wird ein Ausreißer von 15,1 dB in der Gruppe der SH bei der *CompMod*-Kondition nicht dargestellt. Er findet jedoch Beachtung in allen Berechnungen.

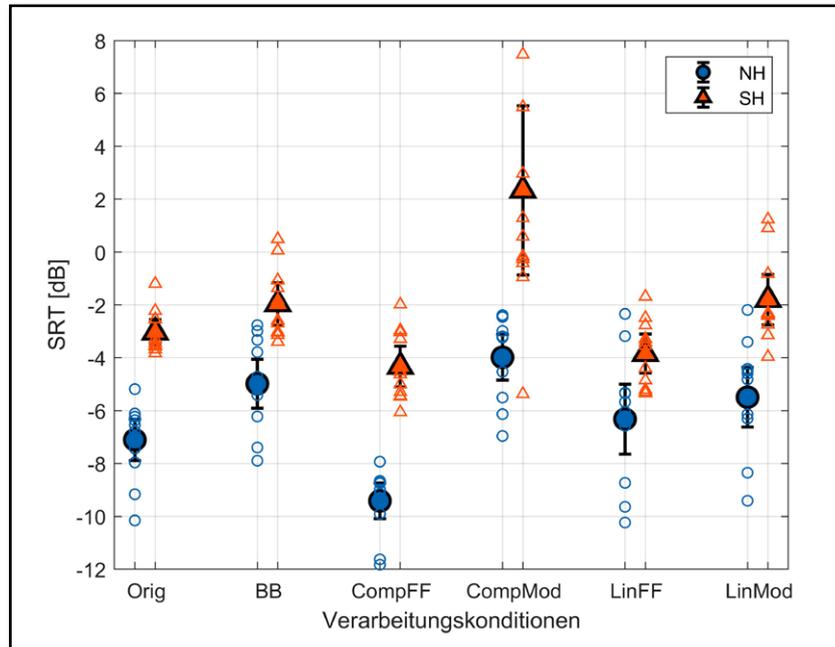


Abb. 18: Sprache-Rausch-Abstand bei 50 % Sprachverständlichkeit (SRT) im Oldenburger Satztest (OLSA) je Verarbeitungskondition und Probandengruppe. Mittelwerte mit 95%-Konfidenzintervallen und den zugrunde liegenden gemittelten SRTs je Proband

Die Varianzanalyse mit Messwiederholung mit den Verarbeitungskonditionen (*Orig*, *BB*, *CompFF*, *CompMod*, *LinFF*, *LinMod*) als *within-subjects* Faktoren und den Gruppen (NH und SH) als *between-subjects* Faktoren zeigte eine signifikante Änderung über die Verarbeitungskonditionen ($F(5,105) = 35,65$; $p(GG) < 0,001$; $\eta_p^2 = 0,63$; Mauchly-Test: $W = 0,04$; $p < 0,001$) und den Gruppen ($F(1,21) = 41,75$; $p < 0,001$; $\eta_p^2 = 0,67$) sowie eine signifikante Interaktion der beiden Faktoren ($F(5,105) = 4,23$; $p(GG) = 0,023$; $\eta_p^2 = 0,17$; Mauchly-Test: $W = 0,04$; $p < 0,001$). Der Mauchly-Test ergab ein signifikantes Ergebnis bei der Analyse der Verarbeitungskonditionen, weshalb der p -Wert mit der Greenhouse-Geisser-Korrektur angepasst wurde ($\epsilon_i = 0,38$).

Die Betrachtung der Unterschiede der Verarbeitungskonditionen zeigt, dass bei allen angestellten Vergleichen signifikante Unterschiede vorhanden sind ($t(22) = -6,05$; $p < 0,001$ für *Orig* vs *BB*; $t(22) = 7,5$; $p < 0,001$ für *Orig* vs *CompFF*; $t(22) = -3,14$; $p = 0,005$ für *BB* vs *CompMod*; $t(22) = 8,59$; $p < 0,001$ für *CompFF* vs *CompMod*; $t(22) = -4,65$; $p < 0,001$ für *CompFF* vs *LinFF*; $t(22) = 3,83$; $p = 0,002$ für *CompMod* vs *LinMod*; $t(22) = -4,32$; $p < 0,001$ für *LinFF* vs *LinMod*). Dies bedeutet, dass die Diskriminationsschwellen beider Gruppen stark von den jeweiligen Verarbeitungskonditionen abhängen.

Bei Betrachtung des Gruppenfaktors wird ersichtlich, dass innerhalb einer Kondition die Mittelwerte der SRT-Schwellen der SH gegenüber den NH generell signifikant höher sind (NH vs SH: $t(21) = -8,45$; $p < 0,001$ für *Orig*; $t(21) = -4,83$; $p < 0,001$ für *BB*; $t(21) = -9,77$; $p < 0,001$ für *CompFF*; $t(21) = -3,88$; $p = 0,002$ für *CompMod*; $t(21) = -3,12$; $p = 0,005$ für *LinFF*;

$t(21) = -4,84; p < 0,001$ für *LinMod*). Im Durchschnitt über alle Verarbeitungsbedingungen liegen sie 4,1 dB auseinander, wobei sie in der *LinFF*-Kondition mit 2,5 dB am geringsten und in der *CompMod*-Kondition mit 6,3 dB am stärksten differieren. Die gemittelten Standardabweichungen über alle Verarbeitungsbedingungen betragen bei den NH 1,7 dB und bei den SH 2 dB mit einem Maximum bei der *CompMod*-Kondition von 3,2 dB. Diese Ergebnisse zeigen, dass die in der Referenzkondition vorhandenen SRT-Differenzen zwischen beiden Gruppen in den Verarbeitungsbedingungen weiterbestehen.

Die differenziertere Betrachtung der Interaktion der beiden Faktoren (Verarbeitungsbedingung und Gruppe) zeigt Unterschiede beider Gruppen bzgl. des Einflusses der Verarbeitungsbedingungen auf. Bei beiden Gruppen ist der SRT der *CompFF*-Kondition mit -9,4 dB bei den NH und -4,3 dB bei den SH signifikant am niedrigsten (NH: $t(11) = 7,94; p < 0,001$; SH: $t(10) = 3,78; p = 0,018$ für *Orig* vs *CompFF*). Die Referenzkondition *Orig* folgt bei den NH mit -7,1 dB als zweitniedrigster Wert und bei den SH an dritter Stelle nach der *LinFF*-Kondition mit -3 dB. Diese lineare First-Fit-Einstellung hat bei den NH mit -6,3 dB einen signifikant höheren SRT gegenüber der kompressiblen First-Fit-Einstellung *CompFF*. Die Differenz dieser beiden Konditionen von 0,5 dB bei den SH (*CompFF* = -4,3 dB, *LinFF* = -3,8 dB) ist nicht signifikant (NH: $t(11) = -6,77; p < 0,001$; SH: $t(10) = -1,41; p = 0,19$ für *CompFF* vs *LinFF*).

Vergleicht man die First-Fit-Einstellungen mit ihren modifizierten Versionen (SH: *CompMod* = 2,3 dB, *LinMod* = -1,8 dB), erhält man für die schwerhörenden Versuchspersonen jeweils signifikante Unterschiede (SH: $t(10) = -4,64; p = 0,006$ für *CompFF* vs *CompMod*; $t(10) = -5,12; p = 0,003$ für *LinFF* vs *LinMod*). Bei den normalhörenden Probanden zeigen die Mittelwerte der modifizierten Einstellungen (NH: *CompMod* = -4 dB, *LinMod* = -5,5 dB) nur zwischen den beiden kompressiblen Konditionen signifikante Unterschiede (NH: $t(11) = -17,33; p < 0,001$ für *CompFF* vs *CompMod*; $t(11) = -1,80; p = 0,099$ für *LinFF* vs *LinMod*). Beide Probandengruppen erzielten mit der *LinMod*-Einstellung signifikant niedrigere SRTs als mit der *CompMod*-Einstellung (NH: $t(11) = 4,34; p = 0,004$; SH: $t(10) = 3,00; p = 0,04$ für *CompMod* vs *LinMod*).

Die breitbandig verstärkende *BB*-Einstellung ordnet sich, gemessen an den Mittelwerten (NH: -4,9 dB SH: -2 dB), für NH und SH zwischen die *Orig*- und die *CompMod*-Kondition. Die Unterschiede zwischen der *Orig*- und der *BB*-Kondition sind jedoch nur bei den normalhörenden Probanden signifikant (NH: $t(11) = -6,05; p < 0,001$; SH: $t(10) = -3,04; p = 0,050$ für *Orig* vs *BB*). Die Differenz zu der *CompMod*-Kondition ist wiederum nur für die schwerhörenden Probanden signifikant (NH: $t(11) = -2,57; p = 0,052$; SH: $t(10) = -2,80; p = 0,037$ für *BB* vs *CompMod*). Die detaillierte Ansicht der Interaktion zwischen den Faktoren der Verarbeitungsbedingungen und den Gruppen zeigt keinen eindeutigen Trend, lässt möglicherweise aber die Vermutung entstehen, dass die Diskriminationsfähigkeit der SH möglicherweise etwas stärker negativ von den Unterschieden der Verarbeitungsbedingungen beeinflusst wird als die der NH.

4.4.2 Diskussion normal- und schwerhörende Probanden

Bei den Ergebnissen des OLSA-Experiments besonders hervorzuheben sind die niedrigen SRTs der auf Sprachdiskrimination ausgerichteten First-Fit-Einstellungen *CompFF* und *LinFF*. Sie ermöglichen beiden Probandengruppen signifikant bessere Diskriminationsschwellen als die originalen, unverarbeiteten Stimuli. Für jede Kondition wurde neben den OLSA-Sätzen auch das entsprechende Rauschen durch das openMHA verarbeitet. Es kann angenommen werden, dass durch den geringeren Energieanteil im tieffrequenten Bereich (siehe Kapitel 3.3.3) weniger Energie in die Helicotrema-nahen auditorischen Filter fallen und so ein Teil der Maskierwirkung des Rauschens entfällt (Patterson, 1976).

Das oben genannte Prinzip der Aufwärtsmaskierung dürfte außerdem der Grund für die schlechteren Diskriminationsschwellen der modifizierten Einstellungen sein. Die in Kapitel 3.3.3 beschriebene Erhöhung der Verstärkung in den Kanälen mit den Mittenfrequenzen 250 und 500 Hz erhöht die Energieanteile in den niedrigen auditorischen Bändern. Auffällig ist allerdings der deutlich schlechtere Mittelwert der SH bei der *CompFF*-Kondition gegenüber der *LinMod*-Einstellung. Diese besitzt im Vergleich mehr Energieanteile im Hochtonbereich. Zusätzlich ist davon auszugehen, dass die beiden Datenpunkte bei 15,1 und 7,5 dB Ausreißer sind. Beide Diskriminationsschwellen wurden in der jeweils ersten Liste der entsprechenden Probanden gemessen und sind daher wahrscheinlich eine Folge der fehlenden Trainingsphase bei diesem Experiment.

Die höheren SRT-Schwellen der breitbandigen Verarbeitungskondition gegenüber der Referenzkondition sind möglicherweise in dem Raumanteil der *BB*-Kondition begründet. Dieser hat seinen Ursprung bei der Wiedergabe der Stimuli über die Lautsprecher und den anschließenden Kunstkopfaufnahmen im Messraum. Durch die Laufzeitunterschiede der reflektierten Schallwellen im Vergleich zu dem Direktschall entstehen temporale Verschmierungen, welche das Sprachverstehen negativ beeinflussen können (Payton, Uchanski und Braida, 1994).

4.5 Korrelationen

4.5.1 Zusammenhang von Qualität und musikalischer Szenenanalyse

Ausgehend von den vorangegangenen Ergebnissen zur musikalischen Szenenanalyse bzw. Transparenz und den Qualitätsbewertungen der verschiedenen Verarbeitungskonditionen werden im Folgenden die möglichen Zusammenhänge beider Faktoren betrachtet.

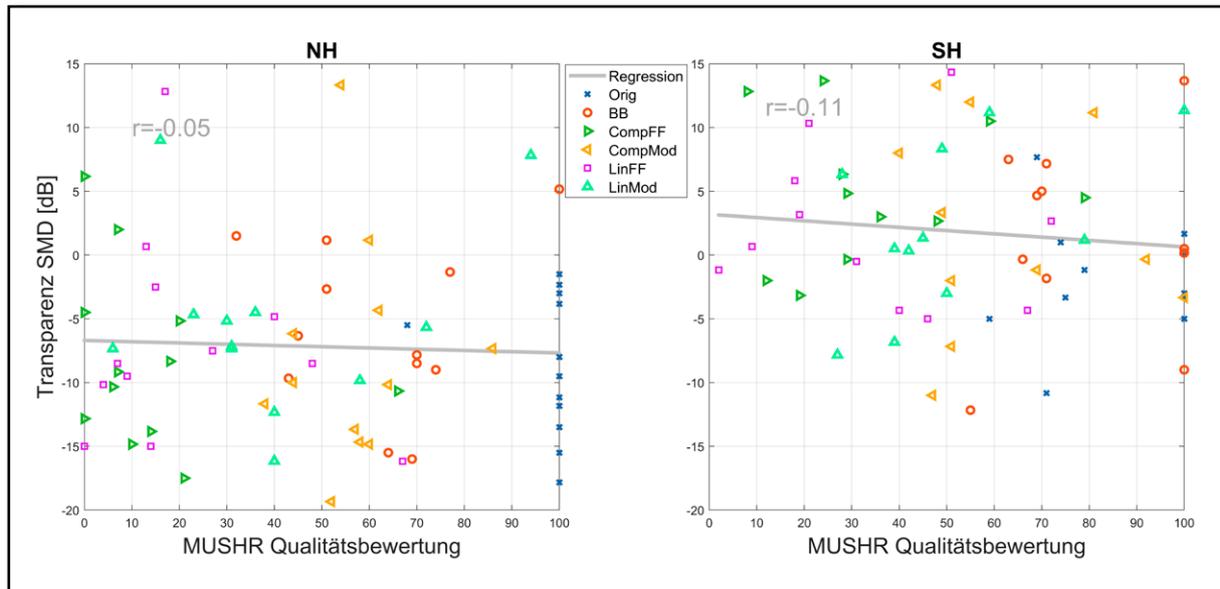


Abb. 19: Lineare Regression und Korrelation (Pearson) von musikalischer Szenenanalyse und Qualitätsbewertung für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen

Abb. 19 stellt für die jungen normalhörenden und die älteren schwerhörenden Probanden die im MUSHR-Experiment ermittelten Qualitätsbewertungen (Abszisse) den Signal-Masker-Differenzen (Ordinate) des Tests zur musikalischen Szenenanalyse gegenüber. Dafür werden die Einzelwerte beider Experimente paarweise je Kondition aufgetragen, eine Regressionsgerade wird nach dem Prinzip der kleinsten Quadrate berechnet und ebenfalls dargestellt. Der Korrelationskoeffizient nach Pearson beträgt für die NH $r = -0,05$ und für die SH $r = -0,11$ (NH: $p = 0,7064$; SH: $p = 0,3918$).

Es kann ausgehend von diesen Daten deshalb als unwahrscheinlich angesehen werden, dass ein Zusammenhang zwischen der empfundenen Qualität von Musik und der Fähigkeit zur musikalischen Szenenanalyse bei verschiedenen Verarbeitungsbedingungen besteht.

4.5.2 Zusammenhang von Qualität und Sprachverstehen

Abb. 20 zeigt den Zusammenhang zwischen der empfundenen Qualität der verschiedenen Verarbeitungsbedingungen aus dem MUSHR-Experiment und der aus dem OLSA-Experiment ermittelten Diskriminationsschwelle.

Das Verfahren zur Regression und Korrelation erfolgte wie in Kapitel 4.5.1 und ergab einen Korrelationskoeffizienten von $r = 0,25$ ($p = 0,0363$) für die NH und $r = 0,04$ ($p = 0,7426$) für die SH. Auch hier sind die Korrelationskoeffizienten zu gering, um von einem Zusammenhang der empfundenen Qualität und der Diskriminationsschwelle auszugehen.

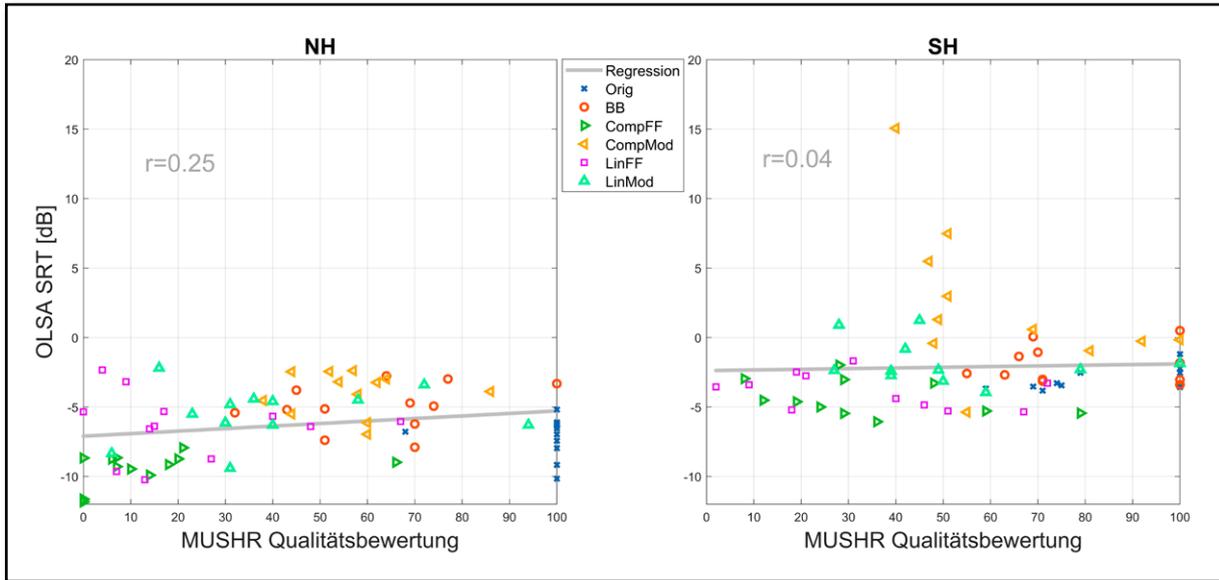


Abb. 20: Lineare Regression und Korrelation (Pearson) von Sprachverstehen und Qualitätsbewertung für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen

4.5.3 Zusammenhang von musikalischer Szenenanalyse und Sprachverstehen

Der Zusammenhang zwischen der musikalischen Szenenanalyse und der Diskriminationsschwelle über alle sechs Verarbeitungskonditionen wird in Abb. 21 dargestellt.

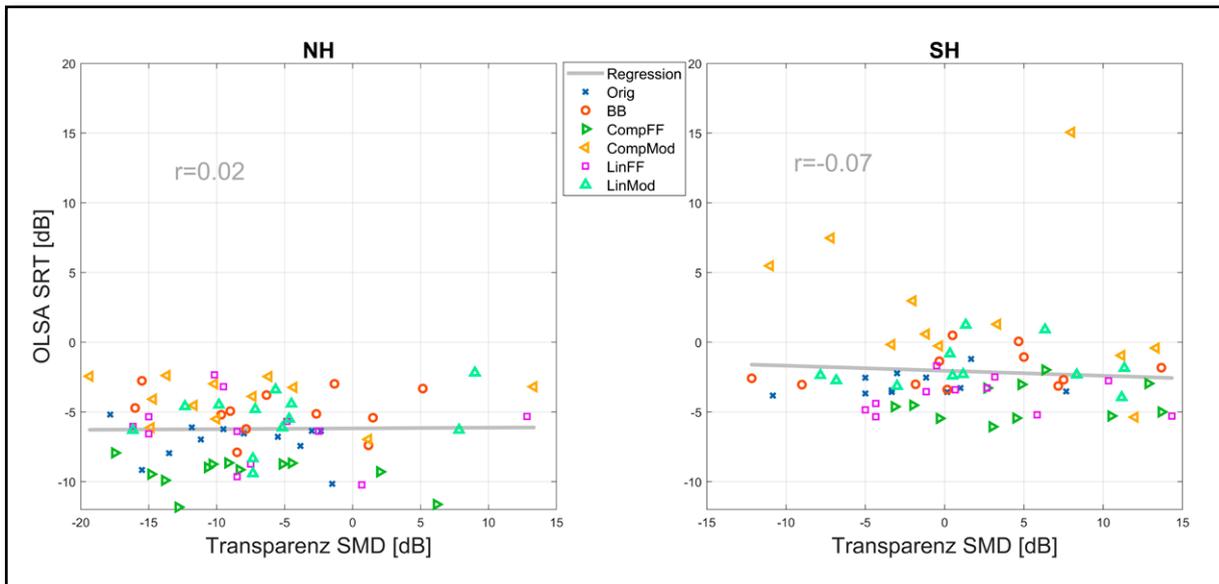


Abb. 21: Lineare Regression und Korrelation (Pearson) von Sprachverstehen und musikalischer Szenenanalyse für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen

Der Korrelationskoeffizient und die Regressionsgerade werden analog zu den vorangegangenen Kapiteln 4.5.1 und 4.5.2 berechnet. Auch die Betrachtung dieser Datenpaare ergab keinen Anlass, von einem Zusammenhang der beiden Faktoren musikalische Szenentransparenz und Sprachdiskrimination auszugehen (NH: $r = 0,02$; $p = 0,8992$; SH: $r = -0,07$; $p = 0,5598$).

5. Diskussion

In dieser Studie wurden drei Experimente durchgeführt, die unterschiedliche Aspekte des Hörens bei sechs verschiedenen Hörgeräteverarbeitungen quantifizieren. Mithilfe eines MUSHR-Tests wurden diese hinsichtlich ihres Einflusses auf die Musikqualität neun verschiedener Musikexzerpte untersucht. Die musikalische Szenentransparenz bei den verschiedenen Hörgeräteverarbeitungen wurde durch einen adaptiven Melodie-Vergleichstest betrachtet. Dabei sollte eine Melodie aus einer musikalischen Begleitung herausgehört und wiedererkannt werden. Zur Einordnung der Ergebnisse in den Kontext der Sprachdiskrimination wurde zusätzlich der Oldenburger Satztest in allen Verarbeitungsbedingungen durchgeführt.

Die in Kapitel 4 beschriebenen Ergebnisse dieser Tests lassen kombiniert betrachtet verschiedene Schlüsse zu. Es gibt keine Hinweise darauf, dass ein Zusammenhang zwischen musikalischer Szenenanalyse, dem Sprachverstehen im Rauschen und der empfundenen Musikqualität der in den Experimenten eingesetzten Verarbeitungsbedingungen besteht. Das bedeutet, dass eine Hörgeräteverarbeitung, welche vorteilhaft für das Sprachverstehen ist, möglicherweise nicht für das qualitativ hochwertig empfundene Musikhören geeignet ist. Dies lassen beispielsweise das gute Abschneiden der modifizierten kompressiblen Einstellung im MURSHR-Experiment zur Qualität und der hohe (also „schlechte“) SRT im Experiment zur Sprachdiskrimination erkennen. Das reziproke Ergebnis beider First-Fit-Einstellungen (niedrige Qualitätsbewertung und niedrige/„gute“ Diskriminationsschwelle) unterstreicht dies.

Die in Kapitel 4.4.2 beschriebene Aufwärtsmaskierung (Patterson, 1976), welche die höheren SRTs der Einstellungen mit verstärkten niedrigen Frequenzen erklären kann, scheint jedoch keinen signifikanten negativen Einfluss auf die musikalische Szenentransparenz zu haben. Die für das Sprachverstehen im Rauschen optimale Hörgeräteverarbeitung des OLSA-Experiments (*CompFF*) erzielte bei den schwerhörenden Probanden die geringste („schlechteste“) Szenentransparenz, wohingegen die normalhörenden Versuchspersonen aus ihr keinen signifikanten Nach- oder Vorteil ziehen konnten. Croghan et al. (2014) zeigten, dass Kompression durch Hörgeräte den Kontrast der Wellenform (temporal) reduziert, Amplitudenverteilungen im Spektrum verändert und das Modulationsspektrum glättet. Dies hat möglicherweise einen negativen Einfluss auf die Wahrnehmung der Klangfarbe (*Timbre*). Die Klangfarbe wiederum kann als Mittel zur Unterscheidung verschiedener Ströme einer auditorischen Szene genutzt werden (Smith, Hausfeld, Power und Gorta, 1982).

Looi, McDermott, McKay und Hickson (2008) beschreiben gar eine Unfähigkeit von Hörgeräteträgern, in den meisten Fällen zwischen verschiedenen Instrumenten oder Ensembles zu unterscheiden. Die Versuchspersonen aus der genannten Studie wiesen allerdings einen größeren Hörverlust auf (PTA > 55 dB HL). Zwar bestand für die Versuchspersonen

in der vorliegenden Arbeit nicht die Aufgabe, zwischen verschiedenen Instrumenten zu unterscheiden, allerdings deuten die hier vorgestellten Ergebnisse des Tests zur Szenentransparenz darauf hin, dass es einen negativen Zusammenhang zwischen Kompression durch Hörgeräte und Szenentransparenz gibt. Die kompressible First-Fit-Kondition *CompFF* besitzt das höchste Kompressionsverhältnis und hat in Bezug auf die Szenentransparenz als einzige Kondition signifikant schlechter abgeschnitten als die Referenzkondition.

Weiterhin lässt sich schließen, dass eine Person die allgemein empfundene Qualität von Musik gegenüber ihrer Transparenz präferiert. Dafür spricht die Bevorzugung der Verarbeitungskonditionen, die einen exponierten Energieanteil in den niedrigen Frequenzen besitzen, allerdings durch Aufwärtsmaskierung höherfrequente Details der Musik verdecken. Einige Probanden berichteten jedoch von hörbaren Verzerrungen der kompressiblen Konditionen. Nichtlineare Verzerrungen können die wahrgenommene Qualität und Natürlichkeit einer Schallquelle stark negativ beeinflussen (Moore und Tan, 2003). Es wurde bei der Erzeugung und Verarbeitung der Stimuli darauf geachtet, ein internes digitales Übersteuern der Audiosignale zu vermeiden, sodass davon ausgegangen wird, dass die wahrgenommenen nichtlinearen Verzerrungen durch die Kompression der entsprechenden Verarbeitungskonditionen erzeugt werden. Dabei kann die gewählte Einregelzeit des Kompressors von 20 ms möglicherweise zu kurz gewählt worden sein, und eine Erhöhung derselben könnte die nichtlinearen Verzerrungen reduzieren. Tan und Moore (2008) zeigten zwar, dass schwerhörende Menschen weniger sensitiv auf Änderungen nichtlinearer Verzerrungen reagieren bzw. es größere individuelle Differenzen als bei normalhörenden Menschen gibt, Schwerhörnde aber dennoch in der Lage sind, auch geringere nichtlineare Verzerrungen wahrzunehmen.

Zusammenfassend sollte herausgestellt werden, dass in dieser Arbeit initial der Zusammenhang zwischen perzipierter Musikqualität und musikalischer Szenentransparenz untersucht wurde. Die durchgeführten Experimente lassen vermuten, dass es keinen direkten Zusammenhang beider Größen gibt.

6. Zusammenfassung und Ausblick

In dieser Arbeit wurde der Einfluss von fünf verschiedenen Hörgeräteverarbeitungen auf die wahrgenommene Musikqualität, die musikalische Szenentransparenz und die Sprachverständlichkeit untersucht. Dafür wurden Kunstkopfaufnahmen der Stimuli über Gehörgangs- und Hörgerätemikrofone angefertigt und mithilfe des open Master Hearing Aids (openMHA) verarbeitet. Es wurden eine breitbandig verstärkende Kondition, eine kompressible und eine lineare First-Fit-Kondition (Camfit) sowie zwei darauf basierende modifizierte Verarbeitungskonditionen getestet. Als Referenzkondition kamen in allen Experimenten die originalen, unverarbeiteten Stimuli zum Einsatz. Alle Experimente wurden von zwölf normalhörenden und elf schwerhörenden Probanden durchgeführt. Die schwerhörenden Probanden hatten einen mittleren sensorineuralen Hochtonhörverlust.

Die Musikqualität wurde durch einen MUSHR-Test mit neun unterschiedlichen Musikexzerpten aus den Genres Jazz, Klassik und Rock/Pop quantifiziert. Dabei gab es keine starken Einflüsse der Genres auf die Qualitätsbewertungen. Die schwerhörenden Versuchspersonen bewerteten die originale und die breitbandige Verarbeitungskondition ähnlich. Die auf Sprache optimierten First-Fit-Einstellungen wurden von beiden Probandengruppen am niedrigsten bewertet. Die modifizierten Einstellungen wurden qualitativ höherwertig als die zugrunde liegenden First-Fit-Einstellungen eingeschätzt.

Das Experiment zur musikalischen Szenentransparenz forderte das Heraushören und Wiedererkennen einer Melodie (Klarinette) aus einer einfachen Begleitung (Cello). Die Melodien bestanden aus vier Tönen (Viertelnoten) und wurden durch die Begleitung (zwei halbe Noten) adaptiv maskiert. Der Pegel des Maskierers blieb dabei konstant, und die Pegeländerungen der Melodie erfolgten durch ein 1-up-2-down-Verfahren. Der Signal-Maskierer-Abstand (SMR), bei dem ca. 71 % der Antworten korrekt waren, wurde verglichen. Die schwerhörenden Probanden benötigten in allen Verarbeitungskonditionen durchschnittlich höhere SMRs als die normalhörenden Versuchspersonen. Ein signifikanter Einfluss der Verarbeitungskonditionen auf die musikalische Szenentransparenz konnte nur bei den schwerhörenden Probanden zwischen der originalen, unverarbeiteten und der kompressiblen First-Fit-Kondition festgestellt werden.

Die Messungen der Sprachdiskrimination erfolgten mit dem Oldenburger Satztest (OLSA) in der offenen Version. Dabei wurden das Sprachmaterial und das maskierende Rauschen ebenfalls mit dem openMHA verarbeitet. Die normalhörenden Probanden erzielten in allen Verarbeitungskonditionen geringere Signal-Rausch-Abstände (SRT), bei denen 50 % der Wörter verstanden wurden. Es zeigten sich für beide Probandengruppen niedrige SRT-Werte bei den auf Sprache optimierten First-Fit-Einstellungen und der Referenzkondition. Die modifizierten Verarbeitungskonditionen hatten durch den höheren Anteil tieferfrequenten Schalls einen negativen Einfluss auf das Sprachverstehen.

Korrelationen der drei Tests ließen keine Zusammenhänge zwischen der Musikqualität, der musikalischen Szenentransparenz und der Sprachverständlichkeit erkennen. Die ermittelten Daten zu den fünf Hörgeräteinstellungen werfen die Frage auf, ob es eine Hörgeräteinstellung geben kann, die den Bedürfnissen der verschiedenen Disziplinen des Musikhörens gerecht wird. Da die in dieser initialen Studie verwendeten Einstellungen des openMHA ein relativ grobes Raster von Hörgeräteverarbeitungen abdecken, können die Erkenntnisse verwendet werden, um weitere spezifischere Verarbeitungsprozesse zu prüfen. Dazu sollten auch Tests mit aktuelleren bzw. komplexeren Algorithmen wie beispielsweise Rückkopplungs- und Störgeräuschunterdrückung sowie binauralen Hörgeräten durchgeführt werden. Auch könnte der Prozess der Verarbeitung durch das openMHA in Echtzeit umgesetzt werden, um adaptive Hörgerätealgorithmen in die Betrachtung mit einzubeziehen.

7. Verzeichnisse

7.1 Abbildungsverzeichnis

Abb. 1:	Optische Vase-Gesicht-Illusion zur Veranschaulichung des Prinzips der „exklusiven Allokation“	7
Abb. 2:	Die Hörfläche des menschlichen Ohres mit der Unbehaglichkeitsschwelle (1), Isophonen angenehm lauten Hörens (2) und dem durchschnittlichen Störschallpegel (3) in Abhängigkeit der Frequenz (Péttursson und Neppert, 2002)	11
Abb. 3:	Ein- und Ausgangspegeldiagramm für verschiedene Kompressionsverhältnisse (Grafik von Fergusson (2007))	12
Abb. 4:	Genereller, struktureller Aufbau des openMHA (Quelle: openMHA_application_manual.pdf)	14
Abb. 5:	Tonaudiogramm des linken Ohres der normal (blau)- und schwerhörenden (rot) Probanden sowie der N3-Hörverlust nach DIN EN 60118-15 (grün) in dB HL	16
Abb. 6:	Tonaudiogramm des rechten Ohres der normal (blau)- und schwerhörenden (rot) Probanden sowie der N3-Hörverlust nach DIN EN 60118-15 (grün) in dB HL	16
Abb. 7:	Terzbandpegel in dB FS der neun unverarbeiteten Musikausschnitte (drei je Musikgenre)	18
Abb. 8:	Schematische Darstellung des Messaufbaus	20
Abb. 9:	Verstärkung in dB je Kondition bei 70 dB Eingangspegel	23
Abb. 10:	Spektra der Verarbeitungskonditionen am Beispiel des siebten Musikausschnitts (Rock/Pop)	24
Abb. 11:	Grafische Benutzeroberfläche des MUSHR-Tests	25
Abb. 12:	Beispielhafter Antwortverlauf einer Versuchsperson für eine bestimmte Verarbeitungskondition	26
Abb. 13:	Mittlere Qualitätsbewertungen der einzelnen Musikausschnitte durch die NH (blau) und SH (rot) über alle Konditionen	29
Abb. 14:	MUSHR-Qualitätsbewertungen je Verarbeitungskondition, gemittelt über alle neun Musikausschnitte	30
Abb. 15:	MUSHR-Qualitätsbewertungen der jungen NH je Genre und Verarbeitungskondition	33
Abb. 16:	MUSHR-Qualitätsbewertungen der älteren SH je Genre und Verarbeitungskondition	34
Abb. 17:	Signal-Masker-Differenz (SMD) für die musikalische Szenentransparenz je Verarbeitungskondition und Probandengruppe	35
Abb. 18:	Sprache-Rausch-Abstand bei 50 % Sprachverständlichkeit (SRT) im Oldenburger Satztest (OLSA) je Verarbeitungskondition und Probandengruppe	37

Abb. 19:	Lineare Regression und Korrelation (Pearson) von musikalischer Szenenanalyse und Qualitätsbewertung für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen	40
Abb. 20:	Lineare Regression und Korrelation (Pearson) von Sprachverstehen und Qualitätsbewertung für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen	41
Abb. 21:	Lineare Regression und Korrelation (Pearson) von Sprachverstehen und musikalischer Szenenanalyse für die jungen normalhörenden (NH) und älteren schwerhörenden (SH) Versuchspersonen	41

7.2 Tabellenverzeichnis

Tabelle 1: Die ersten beiden Formanten einer Auswahl deutscher Vokale aus artikulatorischen Kontexten (gerundete Mittelwerte nach Rausch (1972))	11
Tabelle 2: Liste der verwendeten Musikexzerpte des Experiments zur Musikqualität	17

7.3 Literaturverzeichnis

- Arehart, K. H., Kates, J. M., und Anderson, M. C. (2010). Effects of noise, nonlinear processing, and linear filtering on perceived speech quality. *Ear and hearing*, 31(3), 420-436.
- Bisgaard, N., Vlaming, M. S., und Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, 14(2), 113-120.
- Blauert, J., und Jekosch, U. (2007). *Auditory quality of performance spaces for music—The problem of the references*. Vortrag anlässlich des 19th International Congress on Acoustics (ICA 2007), Madrid.
- Blauert, J., und Jekosch, U. (2012). A layer model of sound quality. *Journal of the Audio Engineering Society*, 60(1/2), 4-12.
- Brand, T., Wittkop, T., Wagener, K., und Kollmeier, B. (2004). *Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen*. Vortrag anlässlich der Jahrestagung der Deutschen Gesellschaft für Audiologie.
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Bücklein, R. (1962). Hörbarkeit von Unregelmäßigkeiten in Frequenzgängen bei akustischer Übertragung. *Frequenz*, 16, 103-108.
- Burns, E. M., und Viemeister, N. F. (1976). Nonspectral pitch. *The Journal of the Acoustical Society of America*, 60(4), 863-869.
- Chasin, M. (2012). Music and hearing aids—An introduction. *Trends in Amplification*, 16(3), 136-139.
- Croghan, N. B., Arehart, K. H., und Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and hearing*, 35(5), e170-e184.
- Cruikshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., und Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study. *American journal of epidemiology*, 148(9), 879-886.
- Davies-Venn, E., Souza, P., und Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, 18(8), 688-699.
- DIN EN ISO 9000:2015-11, Qualitätsmanagementsysteme – Grundlagen und Begriffe (ISO 9000:2015) (2015).
- Fergusson, I. (Producer). (2007, 13.04.2019). Different compression ratios for a signal level above the threshold. [Diagram] Retrieved from https://en.wikipedia.org/wiki/Dynamic_range_compression#/media/File:Compression_ratio.svg
- Fowler, E. P. (1937). Measuring the sensation of loudness: A new approach to the physiology of hearing and the functional and differential diagnostic tests. *Archives of Otolaryngology*, 26(5), 514-521.
- Gabrielsson, A., Lindström, B., und Till, O. (1991). Loudspeaker frequency response and perceived sound quality. *The Journal of the Acoustical Society of America*, 90(2), 707-719.

- Henry, B. A., Turner, C. W., und Behrens, A. (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America*, 118(2), 1111-1121.
- Huber, D. M., und Runstein, R. E. (2013). *Modern recording techniques*. London: Routledge.
- Jekosch, U. (2006). *Voice and speech quality perception: Assessment and evaluation*. Berlin: Springer Science & Business Media.
- Johnson, K. (2011). *Acoustic and auditory phonetics*. Chichester: John Wiley & Sons.
- Keil, W. (2018). *Musikgeschichte im Überblick*. 3., aktualis. u. erw. Aufl. Paderborn: Fink/Brill.
- Kirchberger, M. J., und Russo, F. A. (2015). Development of the adaptive music perception test. *Ear & Hearing*, 36(2), 217-228.
- Leek, M. R., Molis, M. R., Kubli, L. R., und Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, 19(6), 519-526.
- Lepa, S., Weinzierl, S., Maempel, H.-J., und Ungeheuer, E. (2014). *Emotional impact of different forms of spatialization in everyday mediated music listening: Placebo or technology effects?* Vortrag anlässlich der Audio Engineering Society Convention 136.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467-477.
- Looi, V., McDermott, H., McKay, C., und Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and hearing*, 29(3), 421-434.
- Madsen, S. M., und Moore, B. C. (2014). Music and hearing aids. *Trends in hearing*, 18.
- McAdams, S., und Giordano, B. L. (2009). The perception of musical timbre. *The Oxford handbook of music psychology*, 72-80.
- Mills, M. (2011). Hearing aids and the history of electronics miniaturization. *IEEE Annals of the History of Computing*, 33(2), 24-45.
- Möller, S. (2017). *Quality Engineering: Qualität kommunikationstechnischer Systeme*. 2. Aufl. Berlin: Springer-Verlag.
- Möller, S., und Raake, A. (2014). Quality of Experience: Terminology, methods and applications. *PIK-Praxis der Informationsverarbeitung und Kommunikation*, 37(4), 255-263.
- Moore, B., Alcántara, J., Stone, M., und Glasberg, B. (1999). Use of a loudness model for hearing aid fitting: II. Hearing aids with multi-channel compression. *British journal of audiology*, 33(3), 157-170.
- Moore, B. C., Füllgrabe, C., und Stone, M. A. (2010). Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task. *The Journal of the Acoustical Society of America*, 128(1), 360-371.
- Moore, B. C., Füllgrabe, C., und Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and hearing*, 32(5), 556-568.
- Moore, B. C., und Glasberg, B. R. (1998). Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. *British journal of audiology*, 32(5), 317-335.

- Moore, B. C., und Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, 114(1), 408-419.
- Nelson, D. A., und Freyman, R. L. (1987). Temporal resolution in sensorineural hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 81(3), 709-720.
- Oetting, D., Brand, T., und Ewert, S. D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. *Hearing research*, 316, 16-27.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, 59(3), 640-654.
- Payton, K. L., Uchanski, R. M., und Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, 95(3), 1581-1592.
- Pétursson, M., und Neppert, J. (2002). *Elementarbuch der Phonetik*. 3. Aufl. Hamburg: Buske.
- Plack, C. J., Oxenham, A. J., und Fay, R. R., Hgg. (2006). *Pitch: Neural coding and perception*. New York: Springer Science & Business Media.
- Pörksen, B. (2011). Schlüsselwerke des Konstruktivismus. In: *Schlüsselwerke des Konstruktivismus*. Wiesbaden: Springer. 13-28.
- Rausch, A. (1972). Untersuchungen zur Vokalartikulation im Deutschen. *Beiträge zur Phonetik, IPK Forschungsberichte*, 30, 35-82.
- Richtberg, W. (1989). Schwerhörigkeit als psycho-soziale Behinderung. *Audiol Akust*, 28, 12-21.
- Schouten, J. F., Ritsma, R., und Cardozo, B. L. (1962). Pitch of the residue. *The Journal of the Acoustical Society of America*, 34(9B), 1418-1424.
- Siedenburg, K., und McAdams, S. (2017). Four Distinctions for the Auditory "Wastebasket" of Timbre¹. *Frontiers in psychology*, 8.
- Siedenburg, K., Röttges, S., Wagener, K., und Hohmann, V. (2019). *Musical scene analysis of hearing-impaired and normal-hearing listeners: A melody and instrument matching task*. Vortrag anlässlich des 46. Erlanger Kolloquiums, Erlangen.
- Smith, J., Hausfeld, S., Power, R. P., und Gorta, A. (1982). Ambiguous musical figures and auditory streaming. *Perception & Psychophysics*, 32(5), 454-464.
- Sucher, C. M., und McDermott, H. J. (2009). Bimodal stimulation: Benefits for music perception and sound quality. *Cochlear Implants International*, 10(S1), 96-99.
- Tan, C.-T., und Moore, B. C. (2008). Perception of nonlinear distortion by hearing-impaired people. *International Journal of Audiology*, 47(5), 246-256.
- Toole, F. E. (1986a). Loudspeaker measurements and their relationship to listener preferences: Part 1. *Journal of the Audio Engineering Society*, 34(4), 227-235.
- Toole, F. E. (1986b). Loudspeaker measurements and their relationship to listener preferences: Part 2. *Journal of the Audio Engineering Society*, 34(5), 323-348.
- Toole, F. E., und Olive, S. E. (1988). The modification of timbre by resonances: Perception and measurement. *Journal of the Audio Engineering Society*, 36(3), 122-142.

Wagener, K., Kühnel, V., und Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie*, 38(1), 4-15.