



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

FAKULTÄT ELEKTROTECHNIK UND INFORMATIONSTECHNIK

DIPLOMARBEIT

**Untersuchungen zur subjektiven und
objektiven Bewertung und Beurteilung von
Musikinstrumenten anhand von
Solomusikstücken**

vorgelegt von:

Sebastian Merchel

geboren am:

31.07.1980 in Grimma

zum Erlangen des akademischen Grades

Diplomingenieur (Dipl.-Ing.)

Betreuer:

Dipl.-Ing. M. Eichner

Verantwortlicher Hochschullehrer:

Prof. Dr.-Ing. habil. R. Hoffmann

22. November 2005

INSTITUT FÜR AKUSTIK UND SPRACHKOMMUNIKATION

THEMA:

Untersuchungen zur subjektiven und objektiven Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken

BEARBEITUNGSZEITRAUM:

23. Mai bis 22. November 2005

BEARBEITET VON:

Sebastian Merchel

BETREUER:

Dipl.-Ing. M. Eichner

VERANTWORTLICHER

HOCHSCHULLEHRER:

Prof. Dr.-Ing. habil. R. Hoffmann

KOPIEN: 6

SEITENZAHL:

- IM HAUPTTEIL: 110

- IM ANHANG: 41

GESAMT: 151

Kurzfassung:

Diese Arbeit untersucht die subjektive Qualitätswahrnehmung des Menschen, als Grundlage für die Entwicklung objektiver Algorithmen zur qualitativen Differenzierung von Musikinstrumenten. Dazu wurden drei Hörversuche mit zwei unterschiedlichen Methoden (serielle und blockweise Darbietung der Stimuli) entworfen, durchgeführt und ausgewertet. Es war die klangliche Gesamtqualität von ausgewählten klassischen Gitarren, anhand von aufgezeichneten Tonsequenzen (Tonleiter und Melodie), zu beurteilen.

Auf Grund der Verwendung von binauralen Aufnahmen in den Hörversuchen war es notwendig, den Frequenzgang des Kopfhörers, mit Hilfe digitaler Filter, zu kompensieren. Zur Bestimmung des Frequenzgangs wurden drei Messverfahren verglichen und eines als geeignet befunden.

Parameter wie der Musiker, der Raum, die Musiksequenz und das wiederholte Anspiel wurden in den Hörversuchsentwurf einbezogen. Die Komplexität der menschlichen Qualitätswahrnehmung verdeutlichte sich durch eine Vielzahl von Interaktionen der genannten Einflussgrößen. Es wurde geschlussfolgert, dass den Hörversuchsaufwand reduzierende, hierarchische Strukturen daher im Weiteren nur sehr beschränkt anwendbar sind. Es existierte kein signifikanter Unterschied zwischen dem Qualitätsurteil der Hörergruppen Laien und Gitarristen.

Laboratory of Acoustics and Speech Communication**TITLE:**

Subjective and objective evaluation of musical instruments on the basis of solo pieces of music

TIME PERIOD:

May 23rd to November 22nd, 2005

GRADUAND:

Sebastian Merchel

SUPERVISOR:

Dipl.-Ing. M. Eichner

PROFESSOR IN CHARGE:

Prof. Dr.-Ing. habil. R. Hoffmann

COPIES: 6**NO OF PAGES:**

- IN MAIN REPORT: 110

- IN APPENDIX: 41

TOTAL: 151

Abstract:

This report examines the human quality perception. This provides the background for future development of objective algorithms to distinguish between musical instruments on a quality basis. Three listening tests have been developed, using two different methods (serial and block by block presentation of stimuli). They have been realized and evaluated. It was asked, to judge the acoustical overall quality of selected classical guitars, by listening to recorded tone sequences (scale and melody). Due to the use of binaural recordings in the listening experiments, it was necessary to compensate for the characteristic of the used headphones, using digital filters. Three methods for measuring the frequency response of the headphone have been evaluated and one was found to be suitable.

Parameters like the musician, the room, the played sequence and the repetition have been included into the experimental design. The complexity of human quality perception can be seen from a multitude of interactions between the mentioned factors. It was concluded, that hierarchical plans, which would reduce the necessary effort in listening experiments, can only be applied very restricted. There was no significant difference between the quality judgement of guitarists and non guitarists.

TECHNISCHE UNIVERSITÄT DRESDEN

FAKULTÄT ELEKTROTECHNIK UND INFORMATIONSTECHNIK

Institut für Akustik und Sprachkommunikation

Thesen zur Diplomarbeit

Thema: Untersuchungen zur subjektiven und objektiven Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken

Tag der Einreichung: 22. November 2005

Verfasser: Sebastian Merchel

Zur Differenzierung von Musikinstrumenten wurden bisher hauptsächlich Einzeltöne verwendet. Es hat sich gezeigt, dass dies für die qualitative Unterscheidung von Musikinstrumenten nicht ausreicht. Die vorliegende Arbeit verfolgt deshalb das Ziel, durch subjektive Untersuchungen an ausgewählten Gitarren mit Solomusikstücken, eine Referenz für neue objektive Methoden zur Qualitätsbeurteilung zu schaffen. Dazu wurden Hörversuche entworfen und durchgeführt.

- **These 1:** Zur natürlichen Wiedergabe von binauralen Aufnahmen ist es nötig, die Charakteristik des Kopfhörers zu kompensieren. Die Messung der Kopfhörerübertragungsfunktion auf dem Cortex-Kunstkopf (MK-2) oder einem künstlichen Ohr von B&K (Typ 4153) ist dazu nicht geeignet. Messungen an realen Probanden lieferten zufriedenstellende Ergebnisse.
- **These 2:** Zur Durchführung der subjektiven Experimente wurden zwei Hörversuchsmethoden (serielle Darbietung der Stimuli und Darbietung in Blöcken) verglichen. Wenn die Stimuli auf mehrere Blöcke aufgeteilt werden müssen, ergeben sich keine Vorteile gegenüber der seriellen Methode. Ist es möglich, die Anzahl der Stimuli zu reduzieren, so dass sie in einen Block passen, ist diese Methode vorteilhafter.
- **These 3:** In den meisten Fällen wird in Hörversuch eins und zwei die Referenzgitarre G1 etwas besser bewertet als die Referenzgitarre G5. Die beobachteten klanglichen Qualitätsunterschiede zwischen den untersuchten Gitarren sind aber relativ gering und es bleibt zu zeigen, ob sie praktisch relevant sind.
- **These 4:** Die subjektiven Beurteilungen streuen stark. Die Probanden haben oft Probleme, konkrete Qualitätsmerkmale zu benennen oder zu beschreiben. Durch Training und gezieltere Fragen nach speziellen Qualitätsmerkmalen, könnte die Streuung verringert werden. Allerdings verringert sich dadurch auch die externe Validität des Experiments.
- **These 5:** Die statistische Auswertung der Ergebnisse zeigt, dass sowohl Laien wie auch Gitarristen die akustische Gesamtqualität der hier untersuchten Instrumente im Mittel gleich beurteilen. Es scheint also nicht nötig, Gitarristen für die Beurteilung der Gesamtqualität zu engagieren.

- **These 6:** Die beschriebenen Interaktionen verdeutlichen die Komplexität der menschlichen Qualitätswahrnehmung. Objektive Methoden zur Abschätzung der wahrgenommenen qualitativen akustischen Differenzierung zwischen Instrumenten können daher relativ umfangreich werden. Sie sind immer durch sorgfältige subjektive Experimente zu bestätigen.
- **These 7:** Auf Grund einer Vielzahl von Interaktionen zwischen den unabhängigen Variablen, sind in zukünftigen Versuchen hierarchische Strukturen nicht oder nur sehr beschränkt möglich.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die von mir am heutigen Tage dem Prüfungsausschuss der Fakultät Elektrotechnik eingereichte Diplomarbeit zum Thema "Untersuchungen zur subjektiven und objektiven Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken" vollkommen selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie Zitate kenntlich gemacht habe.

Dresden, den 22. November 2005

Sebastian Merchel

Vorwort

Diese Diplomarbeit wurde vorgelegt von Sebastian Merchel, sie ist an Mitarbeiter und Studenten am Institut für Akustik und Sprachkommunikation der TU Dresden gerichtet. Die Arbeit ist in mehrere Kapitel unterteilt:

- **Kapitel 1** leitet in das Thema ein und beinhaltet eine Problembeschreibung.
- **Kapitel 2** beschreibt die Grundlagen der Qualitätsbeurteilung. Praktische Fragen zur Durchführung von Hörversuchen werden analysiert und statistische Methoden zur Auswertung der Daten dargestellt.
- **Kapitel 3** untersucht relevante Verzerrungen auf dem Signalweg und zeigt Kompensationsmöglichkeiten auf. Es werden verschiedene Messverfahren zur Bestimmung der Kopfhörercharakteristik verglichen und schließlich digitale Filter entworfen.
- **Kapitel 4** enthält die Beschreibung von Ziel und Methoden des Hörversuchentwurfs. Die zu untersuchenden Parameter werden spezifiziert und die Auswahl der Probanden erläutert.
- **Kapitel 5** wertet die erhobenen Daten statistisch aus und vergleicht die Ergebnisse.
- **Kapitel 6** fasst die Arbeit abschließend zusammen.

Der anschließende Anhang enthält ergänzende Erläuterungen, Abbildungen und Tabellen. Quellen werden in eckigen Klammern angegeben, welche den Namen des Autors und das Jahr der Veröffentlichung enthalten.

Die Nummerierung von Abbildungen, Tabellen und Gleichungen folgt der Kapitelnummerierung, zum Beispiel wird die zweite Abbildung in Kapitel fünf, Abbildung 5.2 genannt. Verweise auf Gleichungen sind durch runde Klammern gekennzeichnet, z.B. Gleichung (3.11) für Gleichung Nummer elf in Kapitel drei.

Beigefügt ist eine CD-ROM, die die Hörversuche mit den verwendeten Hörbeispielen, die Messergebnisse, den MATLAB Quellcode, Literatur aus dem Internet im PDF-Format, sowie eine Kopie der Diplomarbeit im PDF/PS-Format enthält.

Ich möchte mich bei Herrn Eichner für die freundliche Betreuung, bei Herrn Ziegenhals für informative Gespräche und bei Frau Jäger für die sachkundige Unterstützung bedanken. Danke an alle Mitarbeiter, Gitarristen und Studenten, die unentgeltlich an den Hörversuchen teilgenommen haben.

Inhaltsverzeichnis

Inhaltsverzeichnis	XI
1 Einleitung	1
1.1 Problembeschreibung	1
2 Grundlagen	3
2.1 Beurteilung von Qualität	3
2.1.1 Einflüsse auf die Qualitätsbeurteilung von Musikinstrumenten . . .	6
2.1.2 Wahrnehmungsbeeinflussende Faktoren	8
2.2 Praktische Gesichtspunkte	9
2.3 Merkmalsarten und Skalen	12
2.4 Deskriptive Statistik	13
2.4.1 Häufigkeitsverteilung und Stem-and-Leaf-Plot	13
2.4.2 Statistische Kennwerte	15
2.4.3 Error-Bar	16
2.4.4 Box-Plot	16
2.5 Analytische Statistik	17
2.5.1 Transformation der Datenverteilung	18
2.5.2 Kolmogorov-Smirnov-Anpassungstest auf Normalverteilung	19
2.5.3 Hypothesen	20
2.5.4 α - und β - Fehler	21
2.5.5 Varianzanalyse	21
2.5.6 Relevanz eines signifikanten Ergebnisses	30
3 Kompensation der Kopfhörercharakteristik	33
3.1 Modell für Schallübertragung im Ohr	33

3.2	Herleitung des Korrekturfilters	35
3.3	Vereinfachungen	36
3.4	Frequenzgangmessung des Kopfhörers auf dem Kunstkopf	37
3.4.1	Messsystem 1: MLSSA	37
3.4.2	Messsystem 2: ONO SOKKI	40
3.5	Frequenzgangmessung des Kopfhörers mit realen Probanden	42
3.5.1	Frequenzgangmessung des Miniaturmikrofons durch Vergleich mit Messmikrofon im Freifeld	49
3.6	Frequenzgangmessung des Kopfhörers mit künstlichem Ohr	51
3.7	Vergleich der drei unterschiedlichen Messverfahren	52
3.8	Entwurf inverser Filter	54
4	Entwurf Hörversuche	59
4.1	Wahl der unabhängigen Variablen	60
4.1.1	Auswahl der Gitarren	60
4.1.2	Auswahl der Sequenzen	62
4.1.3	Aufnahme der Klangbeispiele	62
4.2	Entwurf Hörversuch 1	63
4.3	Entwurf Hörversuch 2	67
4.4	Entwurf Hörversuch 3	69
4.5	Wahl repräsentativer Stichproben	70
5	Auswertung	73
5.1	Hörversuch 1	73
5.2	Hörversuch 2	87
5.2.1	Statistische Auswertung ohne Ausrichtung am Ankerbeispiel . . .	87
5.2.2	Statistische Auswertung nach Ausrichtung am Ankerbeispiel . . .	94
5.3	Hörversuch 3	100
6	Zusammenfassung und Ausblick	105
	Literaturverzeichnis	107
A	Die Messmethode MLSSA	A1

A.1	Eigenschaft 1: Hoher Energiegehalt	A1
A.2	Eigenschaft 2: Flaches Frequenzspektrum	A2
A.3	Eigenschaft 3: Einfache Berechnung der Impulsantwort	A2
A.4	Nichtlinearitäten	A5
B	Kopfhörmessung	B7
C	Notenbilder der drei verwendeten Sequenzen	C11
D	Screenshots der Benutzeroberflächen für Hörversuche	D15
E	Graphen und Tabellen zu Hörversuch 1	E21
F	Graphen und Tabellen zu Hörversuch 2	F27
G	Graphen und Tabellen zu Hörversuch 3	G37

Einleitung

Die vorliegende Diplomarbeit wurde im Rahmen des Verbundprojekts "Bewertung und Beurteilung von Musikinstrumenten anhand von Solomusikstücken" zwischen der TU Dresden, Institut für Akustik und Sprachkommunikation und dem Vogtländischen Förderverein für Musikinstrumentenbau und Innovation e.V., Institut für Musikinstrumentenbau (IfM) bearbeitet. Im Rahmen des Projektes sollen beispielhaft vier Instrumentengruppen untersucht werden: Klarinetten, Trompeten, Gitarren und Geigen. Diese Arbeit greift die Gruppe der Gitarren heraus. Da sich in bisherigen Untersuchungen gezeigt hat, dass die Auswertung von Einzeltönen zur Beurteilung eines Instrumentes nicht ausreicht, sollen im Rahmen dieses Projektes Tonsequenzen verwendet werden.

Um Instrumente einzuordnen, existieren zur Zeit einige psychoakustisch motivierte Maße, wie zum Beispiel Schärfe, Rauigkeit oder Volumen, für die zum Teil Berechnungsvorschriften existieren. Sie beschreiben die klanglichen Eigenschaften eines Instrumentes. Fasst man solche Merkmale zusammen, so kann versucht werden, daraus die klangliche Gesamtqualität eines Instrumentes abzuschätzen. Stimmt die Schätzung mit dem tatsächlichen Qualitätsempfinden überein, so ist die Beschreibung des Instruments gelungen. Zur Auswertung objektiver Ergebnisse, dient hier die subjektive Empfindung als Referenz. Die genaue Kenntnis der subjektiven Bewertung und Beurteilung eines Instrumentes ist, anders gesagt, die Grundlage für Untersuchungen zur objektiven Klassifizierung. Dazu ist es nötig, die Wahrnehmung des Menschen, speziell in Bezug auf die akustische Qualität von Gitarren, näher zu untersuchen. Dabei soll die klangliche Gesamtqualität hier im Mittelpunkt stehen. Es ergibt sich folgende Problembeschreibung für diese Arbeit.

1.1 Problembeschreibung

Eine Auswahl von Musikinstrumenten soll auf ihre akustische Qualitätsunterschiede hin untersucht werden. Da bisherige Untersuchungen zeigen, dass Messungen an Einzeltönen die akustische Gesamtqualität eines Instrumentes nicht ausreichend charakterisieren, sollen verschiedene Sequenzen gespielter Musik verwendet werden. Zum Vergleich mit objektiven Tests sind dazu subjektive Untersuchungen notwendig. Daraus ergeben sich die im Rahmen dieser Arbeit zu lösenden Aufgaben:

- Entwurf und Durchführung von Hörversuchen zur qualitativen Beurteilung der Instrumente anhand von geeigneten Melodiesequenzen.
- Wahl der zu berücksichtigenden Randparameter.
- Entwurf eines geeigneten Systems zur Aufnahme und natürlichen Reproduktion der Hörbeispiele.
- Statistische Auswertung der Versuchsdaten unter besonderer Berücksichtigung des Einflusses der Randparameter.
- Einteilung der verwendeten Instrumente nach deren klanglicher Qualität.

Grundlagen

Diese Kapitel beschäftigt sich mit den theoretischen Grundlagen, die zur Bearbeitung der gegebenen Aufgabenstellung notwendig sind. Zu Beginn sollen in Abschnitt 2.1 Modelle zur Beurteilung von Qualität erläutert und in Abschnitt 2.1.1 für Musikinstrumente angepasst werden. Abschnitt 2.1.2 beschäftigt sich mit wahrnehmungsbeeinflussenden Faktoren. Grundlegende praktische Überlegungen zu Hörversuchen sind in Abschnitt 2.2 dargestellt. Abschnitt 2.3 gibt eine Übersicht der Merkmalsarten und der damit verbundenen Skalentypen. Danach folgt in Abschnitt 2.4 und 2.5 die Darstellung benötigter statistischer Methoden.

2.1 Beurteilung von Qualität

Um die akustische (klangliche) Qualität von Musikinstrumenten beurteilen zu können, muss zunächst geklärt werden, was der Begriff Qualität bedeutet. Dazu sind einige Begriffserklärungen notwendig. Nach [Jekosch, 2000] gelten folgende Definitionen:

- **Qualität:** Ergebnis der Beurteilung der wahrgenommen Beschaffenheit einer Einheit im Hinblick auf die erwünschte Beschaffenheit.
- **Wahrgenommene Beschaffenheit:** Gesamtheit der wahrgenommenen Merkmale einer Einheit.
- **Einheit:** Materieller oder immaterieller Gegenstand der Betrachtung. [DIN 55350, 1987]
- **Erwünschte Beschaffenheit:** Gesamtheit der Merkmale von individuellen Erwartungen und/oder sachgerechten Erfordernissen und/oder gesellschaftlichen Forderungen.
- **Merkmal:** Erkannte und benennbare Eigenschaft einer Einheit.

Im vorliegenden Fall entspricht die Einheit dem Klang eines Musikinstrumentes. Beim Hören eines Instrumentes wird die **Gesamtheit der wahrgenommen klanglichen Merkmale** mit der **Erwartung bezüglich dieser Merkmale** verglichen.

Die sensorische Wahrnehmung ist nach [Meilgaard et al., 1999] kein einstufiger Vorgang vom physikalischen Stimulus zum wahrgenommenen und formulierten Merkmal der Einheit, sondern vielmehr wenigstens dreistufig. Dies ist in Abbildung 2.1 dargestellt. Der Stimulus trifft das Sinnesorgan und wird über die Nerven zum Gehirn transportiert. Mit vorangegangenen Erfahrungen im Gedächtnis wird er dort eingeordnet, verarbeitet und schließlich durch Formulierung zum Ausdruck gebracht.

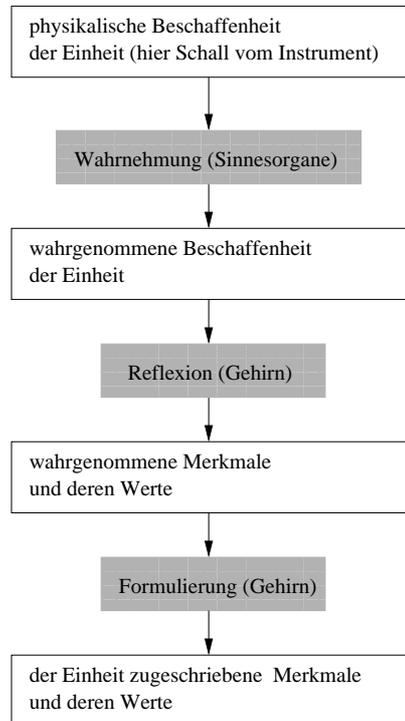


Abbildung 2.1: Schematische Darstellung der Wahrnehmung und Beschreibung von Merkmalen einer Einheit.

Das verschiedene Personen auf ein und denselben Stimulus unterschiedlich reagieren beziehungsweise ihn unterschiedlich beschreiben, kann nach [Meilgaard et al., 1999] mehrere Ursachen haben. Zum einen kann die wahrgenommene Beschaffenheit von Person zu Person schwanken, da ihre Sinnesorgane unterschiedliche Empfindlichkeit aufweisen. Es ist daher zu untersuchen, ob Probanden, die an einem Hörversuch teilnehmen, normalhörig sind oder Gehörschäden aufweisen. Zum anderen kann die Verarbeitung im Gehirn abweichen. Dies kann zum Beispiel durch variierendes Vorwissen bezüglich des wahrgenommenen Merkmals begründet werden oder durch mangelndes Training um das empfundene Merkmal in Zahlen oder Worte zu fassen. Es macht daher Sinn, Hörversuchsprobanden in Gruppen bezüglich ihres Vorwissens einzuteilen. Vor einem Hörversuch kann eine Übung (Training) durchgeführt werden, um den jeweiligen Probanden mit der Verwendung der anzuwendenden Skale vertraut zu machen.

Die Erwartung an den Klang eines Instrumentes ist subjektiv. Sie hängt vom gesellschaftlich musikalischen Hintergrund des jeweiligen Hörers ab. Darunter zählen die musikalischen Vorlieben, die musikalische Erfahrung und die eigene musikalische Tätigkeit des Hörers. Durch den Vergleich von wahrgenommener und erwünschter Beschaffenheit ist eine Qualitätsbeurteilung möglich. In Abbildung 2.2 ist dieser Vorgang nach [Jekosch, 2000] speziell für Klangqualitätsbeurteilung von Musikinstrumenten schematisch darstellt. Für

die Qualitätsbeurteilung als relevant empfundene Merkmale sollen im weiteren als Qualitätsmerkmale bezeichnet werden.

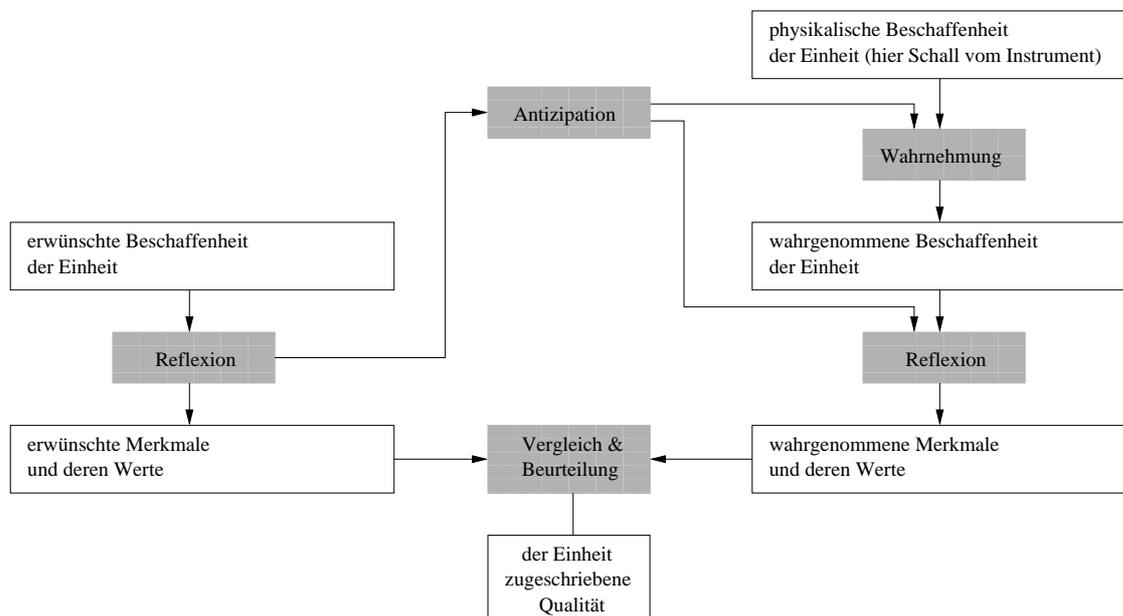


Abbildung 2.2: Schematische Darstellung des Qualitätsereignisses nach [Jekosch, 2000].

Als Beispiel soll die Beurteilung des Klanges einer Klassikgitarre betrachtet werden. Es ist anzunehmen das ein Liebhaber von Heavy-Metal-Musik, der selbst nie ein Instrument gespielt hat, die klangliche Qualität der Gitarre anders empfinden wird als ein klassikliebender Gitarrist. Noch größere Beurteilungsunterschiede könnte es geben, wenn die Personen unterschiedlichen Kulturkreisen angehören oder starke Altersunterschiede auftreten. Es stellt sich nun die Frage, ob dennoch Aussagen über das vorwiegende Qualitätsempfinden bestimmter Personengruppen gefunden werden können. In dieser Arbeit sollen zwei Hörergruppen unterschieden werden:

- Instrumentalisten, die das untersuchte Instrument (Gitarre) erlernen, langjährig spielen oder studieren, werden in eine Gruppe zusammengefasst: **Gitarristen**.
- Instrumentalisten anderer Instrumentengruppen und Personen, die kein Instrument spielen, sollen als **Laien** zusammengefasst werden.

Laut Definition gehen in die Qualitätsbeurteilung alle Eigenschaften des Klanges eines Instrumentes ein, die **erkannt und benennbar** sind. Zuerst soll betrachtet werden, ob und wodurch die Erkennbarkeit von Qualitätsmerkmalen beeinflusst wird. Danach wird nach deren Benennbarkeit gefragt.

Wie viele und welche dieser Merkmale erkannt bzw. als qualitätsbestimmend wahrgenommen werden, hängt nicht nur von den Wahrnehmungsfähigkeiten des Hörers ab, sondern auch von der Führung während der Wahrnehmung. Ein Hörer kann gezielt auf Qualitätsmerkmale hingewiesen werden oder nicht. Man unterscheidet **gelenkte und un gelenkte Wahrnehmung** und entsprechend **gelenkte und un gelenkte Qualitätsbeurteilung**.

Als gelenkte Wahrnehmung versteht man das "Ergebnis der Reflexion über ein extern initiiertes (extrinsisches), im künstlichen Zusammenhang evoziertes Wahrnehmungsereignis. Ungelenkte Wahrnehmung bezeichnet das Ergebnis der Reflexion über ein ungezwungenes (intrinsisches), im natürlichen Zusammenhang evoziertes Wahrnehmungsereignis." [Jekosch, 2000] Auch die Qualitätsbeurteilung kann ein extrinsisch bzw. intrinsisch initiiertes Vorgang sein. Es kann also sein, dass bei gelenkter Wahrnehmung bestimmte, im un gelenkten Fall erkennbare, Qualitätsmerkmale nicht erkannt oder anders bewertet werden. Das beeinflusst natürlich die Beurteilung der Qualität und erschwert deren Interpretierbarkeit. Wenn also gelenkte Wahrnehmung nötig ist, so muss darauf geachtet werden, ein Merkmalbeschreibungssystem zu wählen, welches möglichst alle für die Beurteilung der Qualität relevanten Merkmale enthält. Daraus leitet sich die Vermutung ab, dass es nicht unbedeutend ist, an welcher Stelle in einem Versuch die Testperson nach einem Gesamtqualitätsurteil befragt wird. Wenn diese Frage zu Beginn eines Hörversuchs gestellt wird, so entgeht man der Gefahr durch fehlerhafte Lenkung (z.B. durch ein vorgegebenes ungeeignetes Merkmalsystem) das Urteil zu verfälschen. Wird nach der Gesamtqualität gefragt nachdem schon etliche vorgegebene Qualitätsmerkmale beurteilt wurden, so könnte eine signifikant verschiedene Antwort die Folge sein. Im Weiteren dieser Arbeit soll nur nach der akustischen Gesamtqualität eines Instrumentes gefragt werden, ohne ein Merkmalsystem vorzugeben. Dadurch ist der Proband gezwungen die für ihn relevanten Qualitätsmerkmale selbst zu finden. Eine Verfälschung der Ergebnisse durch ungeeignete Lenkung wird vermieden.

Ob die Testperson in der Lage ist, ihre Merkmalsempfindung auszudrücken oder mitzuteilen, beeinflusst ihre Qualitätsbeurteilung. "Durch die Benennung des Typischen und Eigentümlichen einer Einheit oder ihrer Untereinheit(en) tritt das sie kennzeichnende Besondere hervor, das ihre Distinktivität zu anderen Einheiten oder Untereinheiten begründet." [Jekosch, 2000] Verschiedene Instrumente unterscheiden sich anhand ihrer Merkmale. Soll die Qualität beurteilt werden, so ist es nötig die individuell empfundene Ausprägung der Qualitätsmerkmale zu reflektieren und mit den erwarteten oder erwünschten Merkmalen zu vergleichen.

2.1.1 Einflüsse auf die Qualitätsbeurteilung von Musikinstrumenten

Bei der Qualitätswahrnehmung von Musikinstrumenten spielen viele Faktoren eine Rolle. Sie ist mehrdimensional. Natürlich ist der Klang wichtig, aber auch das Erscheinungsbild (z.B. Farbe, Form, Oberfläche) eines Instrumentes oder dessen mechanische Eigenschaften können die Qualitätsbeurteilung beeinflussen. Auch das Vorwissen über z.B. den Hersteller ist ein Einflussfaktor.

Für unterschiedliche Personengruppen können bestimmte Faktoren unterschiedliche Wichtigkeit haben. So ist für den Spieler sicher der Spielkomfort eines Instrumentes wichtig. Im Gegensatz dazu interessiert den Hörer nur, ob der Spieler mit dem jeweiligen Instrument umgehen kann, beziehungsweise ob man seine Spielanstrengung wahrnimmt oder nicht. Es könnte also den Spieler große Mühe kosten ein Musikstück gut zu spielen, dennoch merkt man ihm seine Anstrengung nicht an. Die Qualitätsbeurteilung des Instrumentes von Spieler und Hörer wäre vermutlich sehr unterschiedlich. Aus Gründen der Praktikabilität soll sich diese Arbeit auf die qualitativ akustischen Eigenschaften eines Instrumentes beschränken. Interessiert ausschließlich die rein klangliche Qualität eines Instrumentes,

ist die Verwendung einer Tonaufnahme von Vorteil. Es entfallen alle optisch und taktil wahrgenommenen Einflussfaktoren.

Doch nicht nur die *klanglichen Eigenschaften des Instrumentes* selbst beeinflussen die Qualitätswahrnehmung. Zum Beispiel die *akustischen Eigenschaften des Raumes*, in dem musiziert wird, können zum Qualitätsempfinden beitragen. Es könnte sein, dass ein Instrument in einem hallarmen Raum besser klingt und ein anderes in einem halligeren. Verschiedene Instrumente wurden sogar speziell für bestimmte akustische Raumsituationen konstruiert (zum Beispiel Orgeln in Kirchen). Es könnte also sein, dass die Qualitätswahrnehmung eines Instrumentes von dem Faktor Raum abhängig ist. Der Autor nimmt weiterhin an, dass *verschiedene Musikstücke* zu einer unterschiedlichen Qualitätsbeurteilung eines Instrumentes führen. Instrumente werden oft speziell für bestimmte Musikrichtungen gefertigt. Eine klassische Konzertgitarre klingt zum Beispiel anders als eine Flamenco-Gitarre. Spielt man nun ein Flamenco-Stück und ein klassisches Werk auf beiden Gitarren, so wird die qualitative Bevorzugung stark vom gespielten Musikstück abhängen. Doch selbst innerhalb einer Musikrichtung (es werden nur klassische Werke auf klassischen Gitarren dargeboten) ist solch ein Effekt vorstellbar. Zudem können der Stil oder die Fähigkeiten des *Instrumentalisten* vom Hörer mit der Qualität der Gitarre verwechselt werden. Hört man ein Musikstück, gespielt von einem Spieler mehrmals auf dem selben Instrument, ist annehmbar, dass er mit fortschreitender Zeit besser eingespielt oder an das Instrument gewöhnt ist. Bei wiederholtem *Anspiel* sind also bessere Leistungen des Spielers erwarten, was wiederum Einfluss auf die Qualitätsbeurteilung des Instrumentes haben kann. Fasst man die eben erläuterten möglichen Einflussgrößen auf die Qualitätsbeurteilung von Instrumenten zusammen, erhält man folgende **Einflussfaktoren**:

- Die klanglichen Eigenschaften des Instrumentes (**Gitarre**).
- Die akustischen Eigenschaften des Raumes (**Raum**).
- Der Einfluss unterschiedlicher Musikstücke (**Sequenz**).
- Der Stil oder die Fähigkeiten des Musikers (**Spieler**).
- Veränderungen im Musizierverhalten des Musikers bei Wiederholungen (**Anspiel**).

Wenn ein Instrumentalist eine bestimmte musikalische Stilrichtung bevorzugt und für diese Beispiele bessere spielerische Leistungen erbringt als ein anderer und diese Leistungsverbesserung sich im Qualitätsurteil des Hörers über das Instrument widerspiegelt, so existiert eine Abhängigkeit zwischen Spieler und Sequenz. Solche Abhängigkeiten können theoretisch zwischen allen Faktoren existieren.

Es bleiben nun folgende Fragen: Gibt es hörbare Qualitätsunterschiede zwischen den Instrumenten? Wenn ja, wie groß sind diese Unterschiede? Welchen Einfluss haben Faktoren wie der Raum, die gespielte Sequenz, der Spieler und wiederholte Anspiele? Gibt es Interaktionen zwischen diesen Einflussgrößen? Es sollen Experimente durchgeführt werden, um diese Fragen zu beantworten. Dazu müssen einige Betrachtungen zu Hörversuchen und den dazugehörigen statistischen Methoden dargestellt werden.

2.1.2 Wahrnehmungsbeeinflussende Faktoren

Für allgemeine Hörversuche sollten Probanden mit gesundem Gehör befragt werden, da die Wahrnehmungskette nach Abbildung 2.1 mit dem sensorischen Erfassen des Stimulus beginnt. Doch nicht nur physiologische sondern auch psychologische Faktoren beeinflussen das Urteil einer Versuchsperson. Nach [Meilgaard et al., 1999] sollen hier einige psychologische Faktoren genannt und diskutiert werden, die Einfluss auf die Wahrnehmung und somit auf die Ergebnisse in einem Hörversuch haben können:

- Die Tendenz, ein gleichbleibendes Urteil zu geben, auch wenn sich die Beispiele in einer Serie langsam verbessern oder verschlechtern, nennt man **Gewöhnungsfehler**.
- Wenn eine Versuchsperson schon bevor sie ein Hörbeispiel beurteilen soll, etwas über die Quelle des Hörbeispiels weiß, so kann dieses Hintergrundwissen ihr Urteil beeinflussen. Dieser **Erwartungsfehler** kann umgangen werden, indem Probanden im Voraus so wenig Information bezüglich des Ursprungs der Hörbeispiele wie möglich erhalten. Dazu sollte die Bezeichnung der Hörbeispiele verschlüsselt sein und die Präsentationsreihenfolge zufällig.
- Wahrnehmung ist mehrdimensional und so können selbst nichtakustische Variablen (zum Beispiel Farbe eines Instrumentes) unterbewusst auf die Beurteilung akustischer Größen Einfluss haben. Dieser **”Stimulus Error”** ist zu vermeiden, indem Probanden nur die für die spezielle Fragestellung relevanten Stimulusmerkmale angeboten bekommen.
- Als **logischen Fehler** bezeichnet man die unterbewusste Verknüpfung eines Merkmals mit einem anderen. Um solche Effekte zu untersuchen ist es notwendig, einzelne Merkmale eines Untersuchungsobjektes getrennt zu beeinflussen.
- Wenn mehr als ein Merkmal für jeden Stimulus zu beurteilen ist, so können sich die Merkmale gegenseitig beeinflussen. Dieser sogenannte **”Halo Effect”** ist vermeidbar, indem für jedes zu beurteilende Merkmal eine eigene Reihe von Stimuli bereitgestellt wird.
- Die **Reihenfolge der Präsentation** hat einen großen Einfluss auf das Urteil des Probanden. Zum Beispiel könnte ein qualitativ gutes Beispiel nach einem sehr schlechten Beispiel als besser eingestuft werden, wohingegen es nach einem extrem guten Beispiel schlechter eingestuft wird. Reihenfolgeeffekte vermeidet man durch **”balanced”** oder zufällige Präsentationsreihenfolge. Für eine ausführliche Einführung in **”balanced block designs”** soll hier auf [Cochran and Cox, 1957] verwiesen werden.
- Probanden können sich auch **gegenseitig beeinflussen**. Es ist also vorteilhaft, sie voneinander optisch bzw. akustisch zu separieren. Dies kann zum Beispiel durch einzelne Sitzungen geschehen.
- **Fehlende Motivation** bei Probanden kann die Ergebnisse stark beeinflussen. Interessierte Probanden sind bessere Versuchsteilnehmer. Probanden werden motiviert, indem ihr Interesse für die Untersuchung geweckt wird und sie das Gefühl haben ein wichtiger Bestandteil der Untersuchung zu sein. Es ist äußerst wichtig,

dass der Teilnehmer gut definierte und gut verständliche Untersuchungsbedingungen vorfindet.

- Die **Lautstärke** eines Hörbeispiels kann das Urteil eines Probanden stark beeinflussen. So kann ein lauterer Beispiel vor einem leiseren bevorzugt werden, selbst wenn die Stimuli sich tatsächlich nur in der Lautstärke unterscheiden. Es ist also wichtig die Lautstärke konstant zu halten.

All diese psychologischen Effekte sollten bei dem Entwurf der Hörversuche in Kapitel 4 berücksichtigt werden.

2.2 Praktische Gesichtspunkte

Bei dem Entwurf eines Hörversuchs sind zudem viele praktische Aspekte zu berücksichtigen, besonders wenn er unter Laborbedingungen stattfindet. Der Autor geht davon aus, dass auch im Labor angestrebt wird der Realität so nahe wie möglich zu kommen. Darauf soll näher in Kapitel 4 eingegangen werden. Die Einteilung von Hörversuchsmethoden kann nach unzähligen unterschiedlichen Gesichtspunkten erfolgen. Es können zum Beispiel absolute Schwellen oder gerade noch wahrnehmbaren Unterschiede ("just noticeable differences" = JND) untersucht werden. Der dargebotene Stimulus kann konstant sein oder sich vom Probanden verändern lassen. Es kann immer nur ein Hörbeispiel gespielt oder es soll paarweise oder zwischen mehreren Stimuli verglichen werden. Je nach Art der abhängigen Variablen und Anzahl der unabhängigen Variablen eignen sich bestimmte Versuchsentwürfe besonders um eine günstige statistische Auswertung zu ermöglichen. Es müssen zudem die Rahmenbedingungen für den Versuch definiert werden.

Konzertsituation

Bei der Beurteilung von Instrumenten wäre eine typische Hörsituation zum Beispiel während eines Konzertes gegeben. Solch eine Situation ließe sich für einen Hörversuch nachstellen, indem ein Auditorium in einem Konzertsaal zur Qualität live gespielter Instrumente befragt würde. Wenn, wie in Abschnitt 2.1.1 entschieden, nur die akustischen Eigenschaften eines Instrumentes interessieren, so müsste man den Musiker mit einem akustisch transparenten Vorhang vom Publikum trennen um optische Einflüsse auf die Wahrnehmung auszuschließen. Diese Methode hat einige Vor- aber auch Nachteile. Vorteile:

- Viele Probanden sind auf einmal befragbar.
- Das gespielte Musikstück wird nicht aufgenommen, bearbeitet und reproduziert und dadurch möglicherweise der natürliche Eindruck des Instruments verfälscht.

Nachteilig ergeben sich Fragen wie:

- Sind alle Probanden, Instrumente und Musiker zum selben Termin verfügbar?

- Hat der Raum an verschiedenen Hörpositionen unterschiedlichen Einfluss auf das Urteil des Probanden?
- Spielfehler des Musikers sind sicher nicht vermeidbar. Haben sie Einfluss auf das Urteil?
- Wie hoch sind die entstehenden Kosten?
- Beeinflussen sich die Probanden gegenseitig?
- Ist es statistisch günstig, allen Probanden die gleichen Hörbeispiele in der selben Reihenfolge vorzustellen?

Stereoaufnahme und Reproduktion über Lautsprecher

Alternativ können die Klangbeispiele im Voraus aufgezeichnet und zu einem späteren Zeitpunkt wiedergegeben werden. Dafür stehen verschiedene Methoden zur Verfügung. Es soll hier nur auf zwei davon näher eingegangen werden, die Zweikanalstereoaufnahme- und Reproduktion über Lautsprecher, sowie die binaurale Technik mit einem künstlichen Aufnahmekopf und Reproduktion über Kopfhörer. Beide Methoden können in unzähligen Varianten ausgeführt werden. Weitere Möglichkeiten wären Surroundtechniken oder aufwendigere Methoden wie Klangfeldsyntheseverfahren.

Im Falle eines Hörversuchs wird versucht, die Realität, also die tatsächliche akustische Situation während der Aufnahme, in der Wiedergabesituation zu reproduzieren. Man spricht von einer virtuellen Realität. Im Idealfall kann der Proband eine reelle Klangquelle nicht von der reproduzierten oder simulierten unterscheiden. Es soll hier angemerkt werden, dass der qualitätsbestimmende Faktor nicht die geschmackliche Bevorzugung bestimmter Wiedergabemethoden ist, sondern die Natürlichkeit der Reproduktion.

Eine klassische Methode ist die **Stereotechnik**. Die aufgenommenen Klangbeispiele können individuell und zu beliebigen Zeitpunkten wiedergegeben werden. Einige weitere Vorteile sollen hier genannt werden:

- Einfache Aufnahme- und Wiedergabetechnik.
- Die Aufnahmesituation kann nach Belieben modifiziert werden. Zum Beispiel können unterschiedliche Räume für die Aufnahme genutzt werden.
- Schneller und individueller Wechsel zwischen unterschiedlichen Hörbeispielen möglich.
- Wiedergabereihenfolge der Hörbeispiele kann individuell und zufällig gewählt werden.

Aus statistischer Sicht sind damit Verbesserungen denkbar. Die nicht speziell untersuchten Parameter können für jede Versuchsperson konstant gehalten werden. Der Hörversuch kann individuell an den jeweiligen Probanden angepasst werden, zum Beispiel indem der Abbruch von längeren Sequenzen erlaubt ist. Die Wiedergabemethode lässt aber auch nachteilig einige Fragen offen:

- Wie verändert sich das reale Musiksinal durch den Stereomix?
- Entspricht das Stereobild dem realen räumlichen Eindruck in der Aufnahmesituation?
- Welchen Einfluss hat der Wiedergaberaum und das Wiedergabesystem?

Binaurale Aufnahme und Reproduktion über Kopfhörer

Um den Einfluss des Wiedergaberaumes zu minimieren, bietet sich die Verwendung von Kopfhörern an. Wird eine Stereoaufnahme über Kopfhörer wiedergegeben, so treten unterschiedliche Effekte auf. Oft wird die Position der Schallquelle innerhalb des Kopfes wahrgenommen und der Hörer ermüdet rasch [Moore, 2003].

Eine weitere Methode ist die **binaurale Aufnahmetechnik**. Nach [Møller, 1992] ermöglicht sie den realen Höreindruck zu reproduzieren, wenn es gelingt den Schalldruck am Trommelfell eines Hörers aufzuzeichnen und ihn dort später wieder exakt herzustellen. Dabei sind Reflektionen an Pinna, Kopf und Torso in der Aufnahme enthalten. Diese Reflektionen ergeben zusammen mit Laufzeitunterschieden zwischen den Ohren die für jede Richtung und Position im Raum typische kopfbezogene Transferfunktion (head related transfer function, HRTF). Sie enthält alle akustischen Informationen, die der Mensch zur realen räumlichen Schallwahrnehmung benötigt.

Binaurale Aufnahmen sind speziell für Kopfhörerreproduktion geeignet, da die Wiedergabekanäle an den Ohren vollständig getrennt werden können (zum Beispiel im Gegensatz zur Lautsprecherwiedergabe). Es ergeben sich einige Vorteile für einen Hörversuch:

- Unter idealen Bedingungen entspricht der Höreindruck während des Hörversuchs dem in der Aufnahmesituation.
- Der Einfluss des Wiedergaberaumes ist reduziert.
- Hintergrundgeräusche werden zusätzlich durch den Kopfhörer gedämpft.
- Das gesamte Wiedergabesystem ist relativ klein und damit transportabel.
- Die Wiedergabereihenfolge der Hörbeispiele kann individuell, zufällig gewählt werden.
- Wechsel zwischen unterschiedlichen Hörbeispielen sind leicht möglich.
- Probanden können sich nicht gegenseitig beeinflussen.
- Es können zum Beispiel unterschiedliche Räume für die Aufnahme verwendet werden.

Auch hier stellen sich einige Fragen:

- Hat die Verwendung von Kopfhörern psychologische Auswirkung auf das Urteil des Probanden?
- Gibt es Einflüsse die den realistischen Eindruck während der Wiedergabe zerstören?

Unrealistische Schallwahrnehmung während der Wiedergabe, zum Beispiel das Vertauschen von Vorn und Hinten, kann aus unterschiedlichen Gründen auftreten:

- In [Møller, 1992] wird erläutert, dass Pinna, Kopf und Torso bei jedem Individuum unterschiedliche Formen haben und somit auch die Filterung des Schalls durch diese individuell ist. Die besten Ergebnisse erreicht man mit Aufnahmen, die mit dem eigenen Ohren gemacht wurden. Wird eine andere Person oder gar ein künstlicher Kopf verwendet, verschlechtert sich der realistische Eindruck [Minnaar et al., 2001]. Trotzdem werden in der Praxis oft solche künstlichen Aufnahmeköpfe, sogenannte "head and torso simulator" (HATS), verwendet.
- Solch ein HATS ist während der Aufnahme fixiert. Kopfbewegungen während der Wiedergabe mit Kopfhörern haben, im Gegensatz zur Situation im realen Schallfeld, keinen Einfluss auf das Signal an den Ohren.
- Zudem können Verzerrungen auf dem Signalweg, zum Beispiel durch die Transfercharakteristik des Kopfhörers, die Wahrnehmung beeinflussen. Ideal wäre ein Kopfhörer mit komplett flachem Frequenzgang.

Die Qualität von Kunstköpfen wurde bisher hauptsächlich nach deren räumlicher Abbildung eingeordnet. Beispiele dafür sind [Møller et al., 1999] und [Minnaar et al., 2001]. Die räumliche Abbildung eines Aufnahmekopfes ist auch für die Qualitätsbeurteilung eines aufgenommenen Instrumentes wichtig, da dessen räumliche Wirkung ein wesentlicher Einflussfaktor auf das Qualitätsurteil sein könnte. In [BS1116-1, 1997] wird darauf hingewiesen, dass bestimmte Qualitätsmerkmale besser bei Kopfhörerwiedergabe und andere besser bei Lautsprecherwiedergabe zu erkennen sind.

Auf Grund der Vorteile binauraler Aufnahmen in Kombination mit Reproduktion über Kopfhörer sollen die Hörversuche in dieser Arbeit mit einem solchen System durchgeführt werden. Einige Grundlagen zur Binauraltechnik und die Kompensation für das Übertragungsverhalten des Kopfhörers sind in Abschnitt 3 dargestellt. Zur Auswertung der aus einem Hörversuch erhaltenen Daten sind einige Betrachtungen zu Merkmalsarten und Skalentypen (Abschnitt 2.3) und statistischen Methoden notwendig, die der Autor in den Abschnitten 2.4 und 2.5 darstellt.

2.3 Merkmalsarten und Skalen

In der Psychoakustik kommen je nach zu beurteilendem Merkmal verschieden Skalen zum Einsatz. Ist das Merkmal nominal einer Kategorie zuzuordnen, verwendet man eine sogenannte Nominalskala. Kann man zudem die Merkmalsstufen in eine Rangordnung bringen, verwendet man eine Ordinalskala. Sind die Abstände zwischen den Merkmalsstufen äquidistant, bezeichnet man den Skalentyp als Intervallskala. Haben zudem die Verhältnisse zwischen den Merkmalsstufen empirischen Sinn, liegt eine Verhältnisskala vor. Wenn ein Merkmal intervall- und verhältnisskaliert ist, nennt man es auch **kardinalskaliert** bzw. **metrisch skaliert**. Eine Übersicht der vier Merkmalsarten und der dazugehörigen Skalentypen ist nach [Jekosch, 2000] und [DIN 55350, 1987] in Tabelle 2.1 dargestellt. Es ist zu erkennen, dass sich die statistische Auswertung je nach untersuchtem Merkmal unterscheidet. Bei der Betrachtung von Qualitätsurteilen soll von einem informationsreichen,

kardinalskalierten Merkmale ausgegangen werden. Es kommen parametrische Methoden zur statistischen Auswertung zum Einsatz. Der Autor blendet daher nichtparametrische Methoden im Folgenden aus.

2.4 Deskriptive Statistik

Statistik kann man grob in zwei Bereiche einteilen. Die **deskriptive Statistik** beschränkt sich auf das Zusammenfassen und Darstellen von Daten. Mit der **analytischen** oder **Inferenz - Statistik** können Hypothesen objektiv bekräftigt oder verworfen werden. Beide Teile sind notwendig und haben ihre Berechtigung. Zuerst ist es günstig, sich einen Überblick über die gesammelten Daten zu verschaffen, indem diese dargestellt und verdichtet werden. Die häufigsten Kennwerte und Methoden dazu werden in diesem Abschnitt erläutert. Allgemeine Erklärungen, warum die Daten so verteilt sind, bleiben dabei spekulativ. Lässt sich im Voraus eine theoretisch begründete Hypothese formulieren, so kann diese anhand der gesammelten Daten empirisch untersucht und deren Brauchbarkeit getestet werden. Abschnitt 2.5 behandelt ausgewählte statistische Methoden die speziell für die in dieser Arbeit zu analysierenden Daten von Nutzen sind.

Einige üblichen Methoden der tabellarischen und grafischen Darstellung seien hier beispielhaft erläutert. Sie dienen dazu, einen besseren Überblick über die Daten zu erhalten. Weiterhin wird versucht, durch Kennwerte Daten zusammenzufassen. Für eine umfassende Einführung soll auf [Bortz, 2005] verwiesen werden.

2.4.1 Häufigkeitsverteilung und Stem-and-Leaf-Plot

Um einen ersten Überblick über die erhobenen Daten zu erhalten, lassen sich zur grafischen Veranschaulichung verschiedene Darstellungsformen wählen. Eine davon ist die Häufigkeitsverteilung (auch Histogramm genannt). Sie ist geeignet, um die Verteilung der Messwerte einer diskreten Variablen darzustellen, aber auch kontinuierliche Variablen lassen sich für diese Darstellungsform aufbereiten. Dazu wird die kontinuierliche Messskala in Intervalle eingeteilt und die in ein Intervall fallenden Messwerte gezählt. Ein Beispiel für solch eine Häufigkeitsverteilung ist in Abbildung 2.3a) zu sehen. Man erkennt das 10 Messwerte im Bereich $3,00 \leq \text{Messwert} < 3,50$ liegen. Um die Verteilung der Variablen mit einer Normalverteilung zu vergleichen, kann man eine solche überlagern. In Abschnitt 2.5.2 soll eine analytische Methode zur Überprüfung auf Normalverteilung erläutert werden und 2.5.5 zeigt das solch eine Überprüfung sinnvoll und notwendig ist.

Eine weitere Darstellungsform ist der sogenannte Stem-and-Leaf-Plot. Er enthält mehr Information als die Häufigkeitsverteilung, da die Verteilung der Variablen innerhalb eines Intervalls sichtbar ist. In Abbildung 2.3b) ist solch ein Plot für die schon in der Häufigkeitsverteilung dargestellte Variable zu sehen. Jeder Messwert wird auf eine Dezimalstelle gerundet. Der Stamm (stem) gibt den ganzzahligen Teil eines Messwertes an. Die dazugehörige erste Dezimalstelle wird als Blatt (leaf) angehängt. Der gerundete Wert 3,0 kommt im vorliegenden Fall also fünf mal vor, da an den Stamm 3 fünf Nullen als Blätter angehängt sind. Je nach Anzahl der Messwerte kann der Stamm weiter geteilt werden. Im vorliegenden Fall ordnet man jeder ganzen Zahl zwei Äste zu, wobei zum Beispiel Werte im Bereich $3,0 \leq \text{gerundeterMesswert} < 3,5$ dem ersten Ast und im Bereich $3,5 \leq \text{gerundeterMesswert} < 4,0$ dem zweiten Ast zuzuordnen sind. Man erkennt das

Tabelle 2.1: Übersicht der Merkmalsarten und Skalentypen nach [Jekosch, 2000] und [DIN 55350, 1987]. *)Die statistischen Kennwerte in jeweils allen links von einer betrachteten Tabellenspalte stehenden Tabellenspalten sind weitere Beispiele für statistische Kennwerte.

Merkmalsart	Qualitative Merkmale		Quantitative	
	Nominalmerkmal	Ordinalmerkmal	(kontinuierliche oder diskrete) Merkmale	
Skalentyp	Topologische Skalen		Kardinalskalen (Metrische Skalen)	
	Nominalskala	Ordinalskala	Intervallskala	Verhältnisskala
Definierte Beziehungen	= ≠	= ≠	= ≠	= ≠
		< >	< >	< >
			+ -	+ -
				...
Interpretation der hinzukommenden Beziehungen	Unterscheidung gleich / ungleich möglich	Unterscheidung kleiner / größer möglich	Differenzen haben empirischen Sinn	Verhältnisse haben empirischen Sinn
Zugelassene Transformationen	umkehrbar eindeutige (bijektive)	monoton steigende (isotone)	lineare $y = ax + b(a > 0)$	
Beispiel für Merkmale	Postleitzahlen Autokennzeichen	Schulnoten Dienstgrade	Celsius-Temperatur Kalenderdatum	Einkommen Fehleranzahl
Beispiele für statistische Kennwerte	Modalwert Häufigkeiten	Quantile *) (Median, Quartile,...)	arithmetischer Mittelwert *) Standardabweichung	geometrischer Mittelwert *) Variationskoeffizient
Anzuwendende statistische Verfahren		nichtparametrische		parametrische (unter Beachtung der Modellvoraussetzungen)
Informationsgehalt	sehr gering	gering	hoch	sehr hoch
Empfindlichkeit gegenüber Ergebnisabweichung	sehr gering	gering	hoch	sehr hoch

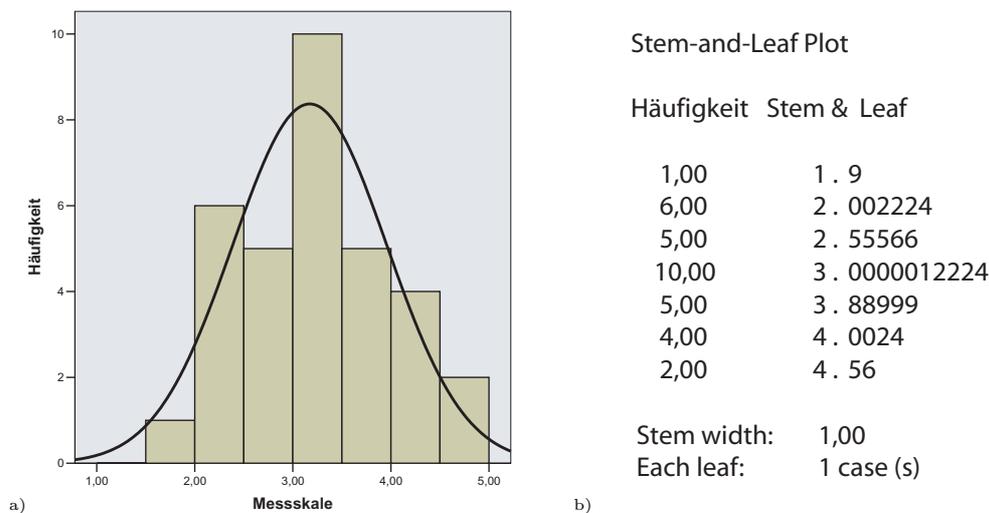


Abbildung 2.3: Alternative Darstellung der Verteilung einer Variablen durch a) Häufigkeitsverteilung (Histogramm) mit überlagelter Normalverteilungskurve, b) Stem-and-Leaf-Plot.

10 Messwerte im Bereich $3,00 \leq \text{gerundeter Messwert} < 3,50$ liegen und fünf im Bereich $3,5 \leq \text{gerundeter Messwert} < 4,0$.

2.4.2 Statistische Kennwerte

Welche Kennwerte für den speziellen Fall geeignet sind, hängt von der Verteilung der Daten und der Fragestellung ab. In [Bortz, 2005] werden solche Kennwerte für die zentrale Tendenz einer Merkmalsverteilung und deren Unterschiedlichkeit oder Variabilität erläutert. Sie sollen hier kurz zusammengefasst werden.

Für Kennwerte der zentralen Tendenz unterscheidet man unter anderen das arithmetische Mittel (oder kurz Mittelwert AM), den Modalwert (Mo) und den Medianwert (Me).

- **arithmetische Mittel:** Das AM berechnet sich klassisch als Summe aller Werte geteilt durch deren Anzahl. Diese Berechnung setzt voraus, dass die Daten kardinalskaliert sind.
- **Modalwert:** Der Mo ist der am häufigsten auftretende Wert eines kategorialen Merkmals oder einer Nominalskala. Daten mit Intervall- oder Verhältnisscharakteristik müssen kategorisiert werden, wenn der Mo gesucht ist. Er ist dann der Mittelwert, der am häufigsten vorkommenden Kategorie.
- **Medianwert:** Der Me halbiert eine Häufigkeitsverteilung. Es ist derjenige Wert, über und unter dem gleich viele Werte liegen. Dies funktioniert ohne Probleme für eine ungerade Anzahl von Messwerten. Bei einer geraden Anzahl wird das arithmetische Mittel der zwei in der Mitte liegenden Messwerte als Me bezeichnet. Auch für kategorisierte Daten lässt sich der Me berechnen [Bortz, 2005, S. 37].

Für spezielle Aufgaben kann man weiterhin das geometrische Mittel, das harmonische Mittel und das gewichtete Mittel errechnen [Bortz, 2005].

Die Kennwerte für zentrale Tendenz charakterisieren eine Verteilung mit einem sie repräsentierenden Wert. Selbst wenn zwei Verteilungen einen ähnlichen AM , Mo oder Me haben, können sie noch stark voneinander abweichen. Diese Abweichung wird versucht mit sogenannten Dispersionsmaßen zu beschreiben. Davon sind die Varianz (s^2) und die Standardabweichung (s) die am häufigsten benutzten. Die **Varianz** (s^2) wird wie folgt berechnet:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (2.1)$$

Dabei ist \bar{x} das arithmetische Mittel AM und x_i der i 'te Messwert. Die quadrierte Summe der Abweichungen aller Messwerte vom AM dividiert durch deren Anzahl, ergibt die Varianz. Wieder werden durch die Quadrierung große Abweichungen stärker berücksichtigt als kleine. Wenn daraus die Wurzel gezogen wird, so erhält man ein leichter zu interpretierendes Maß, die **Standardabweichung** oder **Streuung** (s):

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (2.2)$$

Ein weiteres Maß, das die in Absolutwerten gemessene mittlere Abweichung aller Messwerte vom AM angibt, ist die sogenannte **AD-Streuung**:

$$AD = \sqrt{s^2} = \frac{\sum_{i=1}^n (|x_i - \bar{x}|)}{n} \quad (2.3)$$

2.4.3 Error-Bar

Wenn von mindestens intervallskalierten Daten (siehe Abschnitt 2.3) ausgegangen wird, lässt sich die Verteilung einer Variablen graphisch durch einen "Error-Bar" darstellen. Solch eine Darstellung ist in Abbildung 2.4 zu sehen. Es wird der Mittelwert ME der Verteilung durch einen Kreis markiert. Davon ausgehend zeichnet man zum Beispiel eine Standardabweichung s in positive und negative Richtung.

2.4.4 Box-Plot

Man kann die Streuung einer Verteilung allerdings auch noch anders beschreiben. Im einfachsten Fall durch die **Variationsbreite**. Sie ist die Differenz aus größtem und kleinstem Wert und somit nur abhängig von den Extremen der Verteilung. Wenn nur ein Teil der Werte berücksichtigt wird, die in der Mitte der Verteilung liegen, so erhält man eingeschränkte Streubereiche. Diese werden durch sogenannte Perzentile begrenzt, die sich ähnlich wie der Medianwert Me berechnen. Das x -te Perzentil schneidet $x\%$ der Verteilungsfläche ab. Da der Me die Verteilungsfläche halbiert, also 50% der Verteilungsfläche trennt, entspricht er dem P_{50} . Betrachtet man den Streubereich der mittleren 50% einer Verteilung, so wird er begrenzt von dem 25-ten Perzentil (P_{25}) und dem 75-ten Perzentil (P_{75}). Diesen Streubereich bezeichnet man auch als **Interquartilbereich**. Ein Quartil umfasst je 25% der Verteilungsfläche und wird durch die Perzentile P_0 , P_{25} , P_{50} , P_{75} und P_{100} begrenzt. Der **mittlere Quartilabstand** ist definiert durch $(P_{75} - P_{25})/2$.

Zur grafischen Veranschaulichung ist es möglich einen sogenannten "Box-Plot" zu erstellen. Diese Art der Darstellung wird häufig für ordinalskalierte Daten eingesetzt. In Abbil-

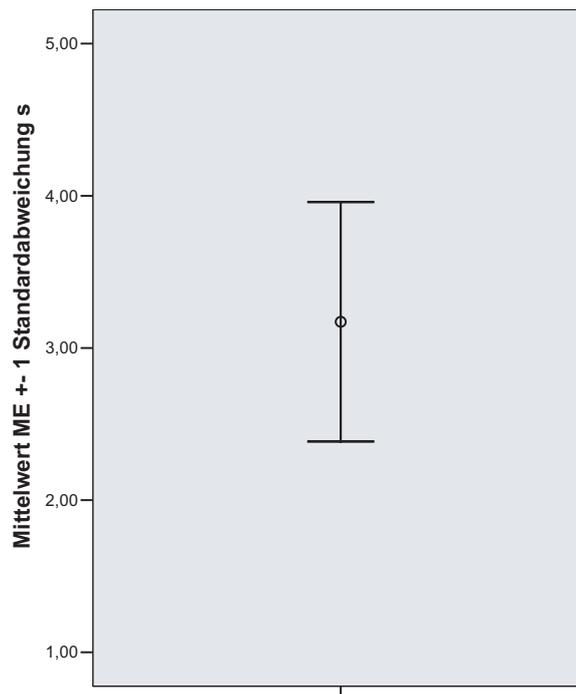


Abbildung 2.4: Error-Bar für die bereits in Abbildung 2.3 dargestellte Häufigkeitsverteilung.

Abbildung 2.5 ist ein Box-Plot dargestellt. Man zeichnet den Medianwert und um ihn eine Box, die von dem 25-ten und dem 75-ten Perzentile begrenzt wird. Liegt der Medianwert in der Mitte der beiden Perzentile, so spricht das für eine symmetrische Verteilung. Die Grenzen für die gesamte Verteilung werden durch Striche über und unter der Box markiert.

Man erkennt gut, ob und wie stark eine Verteilung nach einer Seite verschoben (schief) ist. Man unterscheidet rechtssteile, linkssteile und symmetrische Verteilungen. Die Verteilung ist nur symmetrisch, wenn der Medianwert (ME) dem Mittelwert (AM) gleicht. Die Differenz zwischen beiden Werten ist ein Indikator für die Stärke der Verschiebung einer Verteilung.

Im Weiteren dieser Arbeit sollen Daten die mindestens intervallskaliert sind mit Hilfe ihrer Häufigkeitsverteilungen, des Error-Bars und dem Vergleich von AM und ME betrachtet werden.

2.5 Analytische Statistik

Wenn eine Verschiebung der Daten festgestellt wird, so kann diese eventuell korrigiert werden. Warum dies nötig sein kann und wie eine solche Korrektur angewendet wird, ist in Abschnitt 2.5.1 erläutert. Es soll zudem in Abschnitt 2.5.2 beispielhaft ein analytischer Test zur Überprüfung der Verteilung von Daten, der Kolmogorov-Smirnov-Anpassungstest, beschrieben werden. Es wurde bereits erwähnt, dass sich theoretisch aufgestellte Hypothesen mathematisch überprüfen lassen. Das Vorgehen zum Aufstellen solcher Hypothesen wird in Abschnitt 2.5.3 behandelt. Durch empirische Forschung lässt

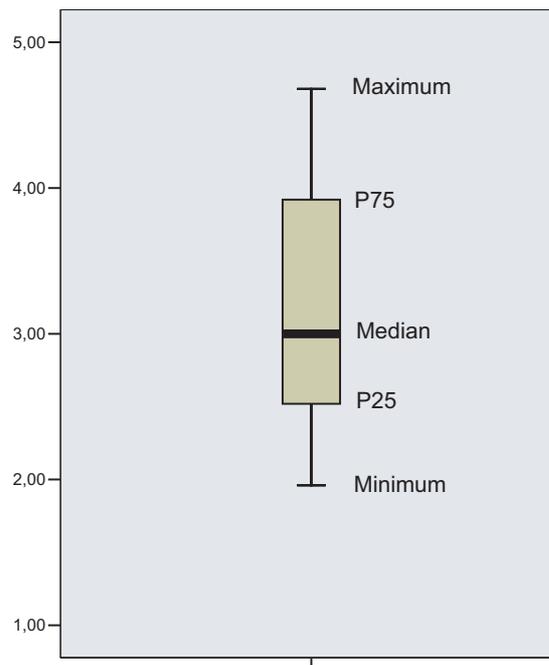


Abbildung 2.5: Box-Plot für die bereits in Abbildung 2.3 dargestellte Häufigkeitsverteilung.

sich allerdings nie die absolute Richtigkeit einer Theorie nachweisen. Dafür wären unendlich viele Tests notwendig. In der Praxis wird immer eine Fehlerwahrscheinlichkeit festgelegt. Welche Fehler auftreten können und wie hoch gängige Fehlerwahrscheinlichkeiten sind, zeigt Abschnitt 2.5.4. Anschließend wird in Abschnitt 2.5.5 eine Einführung in varianzanalytische Methoden, von der einfaktoriellen Varianzanalyse bis zur mehrfaktoriellen Varianzanalyse mit Messwiederholungen, gegeben.

2.5.1 Transformation der Datenverteilung

Bei der Anwendung statistischer Methoden zur Auswertung von Testdaten, wird von bestimmten Verteilungen der Daten bzw. Antworten ausgegangen. Liegt eine Normalverteilung vor, können sogenannte parametrische Tests (z.B. der t-Test nach Student oder eine Varianzanalyse) zum Einsatz kommen. Ist dies nicht der Fall, müssen nichtparametrische Tests (z.B. der Mann-Whitney-U-Test) verwendet werden.

Es kann vorkommen, dass einige Variablen die für die jeweilige Auswertestatistik nötige Verteilung nicht hinreichend widerspiegeln. Dies ist zum Beispiel graphisch zu erkennen, wenn die erhaltene Verteilung stark von einer Glockenkurve abweicht. Nach [Everitt and Hay, 1992] unterscheidet man drei typische Fälle, in denen keine Normal- bzw. Gaußverteilung angenommen werden kann:

- Es wird eine starke Asymmetrie beobachtet.
- Es treten Ausreißer auf.

- Daten unter unterschiedlichen Merkmalsstufen haben stark unterschiedlich gestreute Verteilungen.

Die ersten beiden Punkte beziehen sich auf die Form der Verteilung einer Variable unter einer bestimmten Merkmalsstufe. Der letzte Punkt bezieht sich auf den Vergleich der Daten einer Variable unter unterschiedlichen Merkmalsstufen. Dies ist von Bedeutung, wenn statistische Methoden (wie die Varianzanalyse) von einer gleichen Streuung der Daten ausgehen.

Wenn Abweichungen vorhanden sind, kann versucht werden, die Daten durch eine mathematische Transformation der von der statistischen Methode geforderten Verteilung anzunähern. Dabei wird jeder Datenwert der Datenverteilung mit derselben Operation behandelt. Solch eine Glättung der Datenverteilung ist nach [Jekosch, 2000] nur möglich, wenn die Verteilungsauffälligkeit eindeutig auf eine fehlende Intervallskalierbarkeit zurückzuführen ist. In [Everitt and Hay, 1992] werden solche Transformationen wie folgt charakterisiert:

- Die Rangfolge der Samples wird erhalten, auch wenn sich der Abstand zwischen benachbarten Samples verändert.
- Es bleiben zusammenfassende statistische Kennwerte erhalten, die auf der Rangfolge der Messwerte basieren (Median, Quartile).
- Es sind kontinuierliche Funktionen.

Es werden einige Faustregeln gegeben, welche Funktionen geeignet sind um die Symmetrie einer Verteilung herzustellen:

- Wenn es wichtig ist Symmetrie im Mittelteil der Verteilung zu erhalten und Verschiebung an den Rändern relativ unwichtig ist, wird eine logarithmische Transformation bevorzugt.
- Wenn Symmetrie an den Rändern der Datenverteilung wichtig ist, kommt eine Transformation mittels Quadrat-Wurzel in Frage.
- Wenn eine Verschiebung im Hauptteil und an den Rändern der Datenverteilung ausgeglichen werden soll, könnte man die vierte Wurzel als Transformationsfunktion anwenden.

Die Verwendung und Auswahl solcher Transformationen erfordert einige Erfahrung. In [Everitt and Hay, 1992] wird für weiteres Studium auf [Emerson and Stoto, 1983] verwiesen.

2.5.2 Kolmogorov-Smirnov-Anpassungstest auf Normalverteilung

Die bisher dargestellten Methoden sind mehr oder weniger grafisch. Es ist sicher hilfreich, zu Beginn der Auswertung die aus dem Hörtest erhaltenen Ergebnisse grafisch darzustellen um Abweichungen festzustellen. Wie stark dürfen solche Abweichungen nun aber sein

um noch als Gauß- bzw. Normalverteilung approximiert werden zu können? Zur Überprüfung stehen dafür Testmethoden zur Verfügung. Um zu überprüfen ob eine Stichprobe aus einer bekannten Grundgesamtheit kommen kann, verwendet man bei kleinen Stichproben häufig den Kolmogorov-Smirnov-Anpassungstest (KSA-Test). Es soll hier eine kurze Einführung nach [Bortz et al., 1990] gegeben werden. Man vergleicht also eine zu prüfende empirische Verteilung mit einer bekannten oder theoretischen Vergleichsverteilung. Dazu müssen die Parameter der Vergleichsverteilung bekannt sein. Sind diese nicht bekannt, so müssen diese aus der Stichprobe geschätzt werden. Man verwendet dann den KSA-Test mit Lilliefors-Schranken.

Hier soll das Vorgehen zur Überprüfung auf Normalverteilung erläutert werden. Man berechnet dazu den Mittelwert \bar{x} und die Standardabweichung s einer Stichprobe. Zu prüfen ist nun ob die Messwerte der Stichprobe aus einer normalverteilten Grundgesamtheit stammen, welche den Mittelwert $\mu = \bar{x}$ und die Standardabweichung $\sigma = s$ hat. Dazu berechnet man die maximale Abweichung \hat{D} der Verteilungsfunktionen von normalverteilter Grundgesamtheit ($F_{\mu,\sigma}(x)$) und Stichprobe ($F(x)$):

$$\hat{D} = \max |F_{\mu,\sigma}(x) - F(x)|. \quad (2.4)$$

Der ermittelte \hat{D} -Wert wird mit einem tabellierten kritischen Wert auf einem vorher festgelegten Signifikanzniveau (z.B.: $\alpha = 0.2$) verglichen und darf diesen nicht überschreiten, wenn eine Normalverteilung vorliegt. Die tabellierten kritischen Werte nach Lilliefors (Lilliefors-Schranken) entstammen empirisch einer Monte-Carlo-Studie und sind für verschiedene Signifikanzniveaus und Stichprobengrößen angegeben. Eine solche Tabelle für Stichprobengrößen von 4 bis 30 Werten findet man bei [Bortz et al., 1990].

2.5.3 Hypothesen

Es können nun Hypothesen aufgestellt und anhand der erhobenen Daten überprüft werden. Hypothesen müssen immer im Voraus auf Grund theoretischer Überlegungen oder vorangegangener Experimente formuliert werden. Nach [Bortz, 2005] ist die Aufstellung und anschließende Überprüfung von einer Hypothese an denselben Daten auf keinen Fall zulässig. In [Bortz, 2005] findet sich auch eine Übersicht und eine Einführung in die Formulierung und Überprüfung von Hypothesen. Man unterscheidet in Unterschiedshypothesen (Es existiert ein Unterschied zwischen ... und) und Zusammenhangshypothesen (Wenn ... dann). Zudem kann eine Hypothese gerichtet und ungerichtet sein, das heißt es wird eine bestimmte Richtung des Unterschieds oder Zusammenhangs prognostiziert oder nicht. Kann zudem die Stärke des Unterschieds oder Zusammenhangs angegeben werden, spricht man von einer spezifischen Hypothese (Gegenteil: unspezifische Hypothese).

Untersucht man einen neuen Sachverhalt, stellt man meist eine Behauptung auf, die die bisherige Theorie erweitern oder ersetzen kann. Die aufgestellte Hypothese heißt daher Alternativhypothese und wird mit H_1 bezeichnet. In der Statistik betrachtet man nun die der Alternativhypothese komplementäre Aussage, die Nullhypothese H_0 . Kann die Nullhypothese H_0 verworfen werden, wird die Alternativhypothese H_1 akzeptiert.

2.5.4 α - und β - Fehler

Wird nun H_0 verworfen (und H_1 akzeptiert) obwohl H_0 korrekt war, begeht man einen Fehler erster Art, einen sogenannten α -Fehler. Kann H_0 auf Grund der Untersuchung nicht verworfen werden, obwohl H_1 richtig ist, begeht man einen Fehler zweiter Art, einen β -Fehler. Je nach aufgestellter Hypothese unterscheidet sich, welcher Fehler gravierendere Auswirkungen hat.

Die Wahrscheinlichkeit für einen α -Fehler heißt Irrtumswahrscheinlichkeit. Es ist die Wahrscheinlichkeit H_0 fälschlicherweise zu verwerfen. "Beträgt die Wahrscheinlichkeit des gefundenen oder eines extremeren Untersuchungsergebnisses unter der Annahme, die H_0 sei richtig, höchstens 5%, wird dieses Ergebnis als signifikant bezeichnet. Beträgt diese Wahrscheinlichkeit höchstens 1%, ist das Ergebnis sehr signifikant." [Bortz, 2005]

2.5.5 Varianzanalyse

Es soll die Unterschiedshypothese geprüft werden, dass sich die Mittelwerte aus mehreren Stichproben signifikant unterscheiden. Handelt es sich um zwei Stichproben, verwendet man zum Beispiel den t-Test (siehe [Bortz, 2005]). Werden mehrere Stichproben bezüglich deren Mittelwerte verglichen, kann unter bestimmten Voraussetzungen die Varianzanalyse (ANalysis Of VAriance = ANOVA) zum Einsatz kommen.

"Das Gemeinsame aller varianzanalytischen Versuchspläne ist darin zu sehen, dass sie die Unterschiedlichkeit von Versuchspersonen in Bezug auf ein Merkmal (abhängige Variable) auf eine oder mehrere unabhängige Variablen zurückführen." [Bortz, 2005]

Die häufigsten Begriffe und das Analyseverfahren sollen an einem Beispiel erläutert werden, in dem es um die klangliche Qualitätsbeurteilung von mehreren Gitarren geht.

Die **abhängige Variable** ist das Merkmal, dessen Varianz untersucht werden soll. In diesem Beispiel sollen Versuchspersonen nach einem "klanglichen Qualitätsurteil für eine Gitarre" (abhängige Variable) gefragt werden. Das Urteil wird unterschiedlich ausfallen und es stellt sich die Frage, woher diese Unterschiedlichkeit oder Varianz kommt. Um diese Frage zu lösen, sucht man mögliche Einflussfaktoren, sogenannte **unabhängige Variablen**. Nehmen wir an, dass zwei verschiedene Gitarren bei dem Hörversuch zum Einsatz kommen. Die unabhängige Variable "Gitarre" hat hier zwei **Stufen**. Man bezeichnet sie auch als zweifach gestuften **Faktor**. Werden die Stufen eines Faktors zufällig ausgewählt, spricht man von einem Faktor mit zufälligen Effekten. Wählt man die Stufen nicht zufällig, spricht man von einem Faktor mit festen Effekten. Die statistische Auswertung unterscheidet sich. In dieser Arbeit wird immer von festen Effekten ausgegangen, so dass man letztlich nur über die ausgewählten Stufen Aussagen formulieren kann.

Zu Beginn sollen die Voraussetzungen für eine Varianzanalyse diskutiert werden. Wird der Einfluss *einer* unabhängigen Variable (mit p Stufen) auf *eine* abhängige Variable untersucht, spricht man von einer einfaktoriellen Varianzanalyse. Dieser einfachste Fall soll nach [Bortz, 2005] demonstrativ erläutert werden. Befragt man dabei jeden Probanden mehrmals, spricht man von einer einfaktoriellen Varianzanalyse mit Messwiederholung. Wenn mehrere unabhängige Variablen gefunden werden, wird das Analyseverfahren modifiziert und es kommt die mehrfaktorielle Varianzanalyse mit Messwiederholung zum Einsatz. Zum Abschluss soll noch die praktische Relevanz eines signifikanten Unterschieds betrachtet werden.

Voraussetzungen für die Varianzanalyse

Um die erhobenen Daten mittels der Varianzanalyse zu untersuchen, müssen einige Voraussetzungen erfüllt sein [Bortz, 2005]. Sie sollen hier nur von der praktischen Seite aus beleuchtet werden. Es werden Methoden aufgezeigt, um die Erfüllung dieser Bedingungen am Datensatz zu prüfen.

- Die Messwerte innerhalb jeder Stichprobe müssen normalverteilt sein. Um diese Voraussetzung zu prüfen, gibt es mehrere Testverfahren. Häufig wird der Kolmogorov-Smirnov-Test verwendet. Er wurde bereits in Abschnitt 2.5.2 näher erläutert.
- Die Varianzen innerhalb der Stichprobe dürfen sich nicht signifikant unterscheiden. Es wird häufig der Bartlett-Test oder der Levene-Test auf homogene Fehlervarianzen verwendet. Nach [Bortz, 2005] ist die obige Forderung unerheblich, wenn gleich große Stichproben verwendet werden. Dieses ist im Folgenden anzunehmen.
- Die Messwerte müssen voneinander unabhängig sein. Das heißt an jeder Versuchsperson darf nur ein Messwert erhoben werden. Diese Bedingung ist für die nächsten Abschnitte theoretisch anzunehmen. Die Verletzung dieser Bedingung folgt, wenn Messwiederholungen an einer Versuchsperson in das Versuchsdesign aufgenommen werden. In diesem Fall ändert sich die genannte Bedingung und ist in der sogenannten Zirkularitätsannahme zusammengefasst. Bei Verletzung der Zirkularitätsannahme lassen sich eventuell dennoch verwertbare Ergebnisse berechnen, indem Korrekturfaktoren zum Einsatz kommen. Informationen dazu findet man bei [Bortz, 2005].

Die Zirkularitätsannahme soll im Folgenden als erfüllt betrachtet werden.

Einfaktorielle Varianzanalyse

Als unabhängige Variable soll im Folgenden ein Faktor A (Gitarren) untersucht werden, der in p Stufen eingeteilt ist (p unterschiedliche Gitarren). Die Faktorstufen werden durch $i = 1, 2, \dots, p$ markiert. Unter jeder Faktorstufe, also für jede Gitarre, werden n Versuchspersonen zu einem Qualitätsurteil befragt. Die resultierenden Messwerte x_{mi} werden durch die jeweilige Faktorstufe i und die befragte Person $m = 1, 2, \dots, n$ indiziert. Der Messwert des Urteils der dritten Versuchsperson für die zweite Gitarre wird somit als x_{32} gekennzeichnet. Ein allgemeines Datenschema ist in Tabelle 2.2 zu sehen.

Summiert man die Messwerte unter einer Faktorstufe i , so erhält man die Summe $A_i = \sum_{m=1}^n x_{mi}$. Der Mittelwert ergibt sich dann zu $\bar{A}_i = \frac{A_i}{n}$. Die Gesamtsumme über alle Messwerte berechnet sich als $G = \sum_{m=1}^n \sum_{i=1}^p A_i$. Damit folgt für das arithmetische Mittel über alle Messwerte $\bar{G} = \frac{G}{p \cdot n}$.

Mit der Varianzanalyse überprüft man eine Nullhypothese. Diese besagt, dass die Mittelwerte unter allen p Stufen des Faktors A identisch sind bzw. das Qualitätsurteil sich für alle p Gitarren nicht ändert. Die Alternativhypothese heißt dann, dass sich mindestens zwei der Mittelwerte unterscheiden. Die H_0 würde also verworfen und die H_1 bestätigt, wenn mindestens zwei der Gitarren zu signifikant unterschiedlichen Qualitätsurteilen führen.

Tabelle 2.2: Allgemeines Datenschema für eine einfaktorielle Varianzanalyse nach [Bortz, 2005].

	Faktor A					
	1	2	...	i	...	p
	x_{11}	x_{12}	...	x_{1i}	...	x_{1p}
	x_{21}	x_{22}	...	x_{2i}	...	x_{2p}
	\vdots	\vdots		\vdots		\vdots
	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mp}
	\vdots	\vdots		\vdots		\vdots
	x_{n1}	x_{n2}	...	x_{ni}	...	x_{np}
Summen:	A_1	A_2	...	A_i	...	A_p G

Die Gesamtvarianz aller Messwerte quantifiziert im Beispiel, in welchem Maß sich die von den Versuchspersonen abgegebenen Qualitätsurteile unterscheiden. Es wird nun gefragt, in welchem Ausmaß diese Unterschiede auf die Unterschiede zwischen den Gitarren rückführbar sind.

Dazu muss als *erstes* die Gesamtvarianz aller Messwerte berechnet werden. Sie wird in diesem Zusammenhang aber nicht wie in Gleichung 2.1 als deskriptives Maß betrachtet, sondern stellt eine Schätzung der Populationsvarianz σ^2 dar aus der die Versuchspersonen stammen. Für eine ausführliche Erläuterung soll hier auf [Bortz, 2005] verwiesen werden. Die Schätzung der Varianz $\hat{\sigma}^2$ wird oft auch als "MS" = "mean square" bezeichnet und berechnet sich zu:

$$\hat{\sigma}_{tot}^2 = \frac{\sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{G})^2}{n \cdot p - 1} = \frac{QS_{tot}}{df_{tot}}. \quad (2.5)$$

Sie ergibt sich aus der Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert, welche oft als **totale Quadratsumme** (QS_{tot}) bezeichnet wird, dividiert durch die **Freiheitsgrade** der Varianz für alle Messwerte (df_{tot}). Die Quadratsumme wird oft auch als "sum of squares" (SS) bezeichnet.

Als *zweites* wird der Anteil der Unterschiedlichkeit aller Messwerte bestimmt, der auf die p verschiedenen Gitarren rückführbar ist. Dazu berechnet man die Schätzung der Varianz, die nur von den p Stufen der unabhängigen Variablen stammt. Dabei geht man davon aus, dass innerhalb einer Stufe des Faktors A kein Unterschied zwischen den n Versuchspersonen mehr besteht. Damit lässt sich für $m = 1, 2, \dots, n$ jeder Messwert x_{mi} durch den Mittelwert der Stufe \bar{A}_i ersetzen. Der modifizierte Datensatz ist in Tabelle 2.3 zu sehen.

Eine Stufe des Faktors A bezeichnet man als Treatmentstufe. Die Varianz, welche auf die p Treatmentstufen zurückzuführen ist, berechnet sich zu:

$$\hat{\sigma}_{treat}^2 = \frac{n \cdot \sum_{i=1}^p (\bar{A}_i - \bar{G})^2}{p - 1} = \frac{QS_{treat}}{df_{treat}}. \quad (2.6)$$

QS_{treat} bezeichnet man als **Treatmentquadratsumme** und df_{treat} sind die **Freiheitsgrade** der QS_{treat} . Die Anzahl der Freiheitsgrade ergibt sich aus der Anzahl frei variierbarer Messwerte bei festgelegtem \bar{G} . Da die Messwerte innerhalb einer Treatmentstufe auf den Mittelwert der Stufe festgelegt worden, können sie nicht variieren. Bei bekann-

Tabelle 2.3: Allgemeines Datenschema für eine einfaktorielles Varianzanalyse nach [Bortz, 2005].

		Faktor A					
		1	2	...	i	...	p
	\bar{A}_1	\bar{A}_2	...	\bar{A}_i	...	\bar{A}_p	
	\bar{A}_1	\bar{A}_2	...	\bar{A}_i	...	\bar{A}_p	
	\vdots	\vdots		\vdots		\vdots	
	\bar{A}_1	\bar{A}_2	...	\bar{A}_i	...	\bar{A}_p	
	\vdots	\vdots		\vdots		\vdots	
	\bar{A}_1	\bar{A}_2	...	\bar{A}_i	...	\bar{A}_p	
Summen:	A_1	A_2	...	A_i	...	A_p	G

tem Gesamtmittelwert \bar{G} , sind von den p Treatmentstufenmittelwerten noch $p - 1$ frei variierbar.

Als *drittes* betrachtet man den Anteil der Unterschiedlichkeit aller Messwerte, der nicht auf Treatmenteffekte zurückzuführen ist. Der Einfluss von störenden, nicht in die Untersuchung aufgenommenen, Faktoren (sogenannten Störvariablen) wird als **Fehlervarianzanteil** bezeichnet. Er beinhaltet alle störenden Effekte, die Treatmenteffekte überlagern. Solche Effekte können unterschiedliche Motivation, unterschiedliches Vorwissen, Messfehler, etc. sein. Je größer der Einfluss der Störvariablen, desto ungenauer werden Aussagen über Treatmenteffekte.

Um den Fehlervarianzanteil zu berechnen, entfernt man den Anteil in den Messwerten, der auf Treatmenteffekte zurückgeht. Das heißt von jedem Messwert x_{mi} wird der jeweilige Mittelwert der Stufe \bar{A}_i abgezogen. Nach [Bortz, 2005] ergibt sich dann:

$$\hat{\sigma}_{Fehler}^2 = \frac{\sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{A}_i)^2}{p \cdot (n - 1)} = \frac{QS_{Fehler}}{df_{Fehler}} \quad (2.7)$$

Er zeigt weiterhin, dass aus Gleichung 2.5, 2.6 und 2.7 folgt:

$$QS_{treat} + QS_{Fehler} = QS_{tot} \quad (2.8)$$

$$df_{treat} + df_{Fehler} = df_{tot} \quad (2.9)$$

Man erkennt, dass die totale Quadratsumme in einen Treatmentanteil und einen Fehleranteil zerlegbar ist. Äquivalentes gilt für die Freiheitsgrade.

Um die Frage zu klären, wie stark die Unterschiedlichkeit in den Messwerten auf die verschiedenen Gitarren zurückzuführen sind, berechnet man die sogenannte **Varianzaufklärung** η^2 :

$$Varianzaufklärung = \frac{QS_{treat}}{QS_{tot}} \cdot 100\% \quad (2.10)$$

$$\eta^2 = \frac{QS_{treat}}{QS_{tot}} \quad (2.11)$$

Anschließend wird überprüft, ob die Varianzaufklärung nur zufällig durch die Auswahl der Stichprobe entstanden ist, obwohl die H_0 gilt und gar keine Unterschiede zwischen den Faktorstufen (Gitarren) existieren. Ist die Wahrscheinlichkeit dafür geringer als eine

im Voraus festgesetzte Irrtumswahrscheinlichkeit (z.B.: 5%), wird die H_0 zugunsten der H_1 verworfen und man kann von mindestens zwei signifikant verschiedenen Faktorstufen (Gitarren) sprechen.

Es lässt sich zeigen, dass bei Gültigkeit von H_0 (Die Mittelwerte in den einzelnen Faktorstufen sind gleich.) gilt:

$$\hat{\sigma}_{treat}^2 = \hat{\sigma}_{Fehler}^2. \quad (2.12)$$

Lässt sich also zeigen, dass die beiden Varianzen identisch sind, muss die H_0 akzeptiert werden. Das hieße, es gibt keine signifikanten Unterschiede zwischen den Faktorstufen (Gitarren). Lässt sich zeigen, dass $\hat{\sigma}_{treat}^2 > \hat{\sigma}_{Fehler}^2$, dann kann die H_0 verworfen werden. Ob zwei Varianzen identisch sind, prüft der F-Test.

Der **F-Test** überprüft die Nullhypothese, dass zwei Stichproben aus normalverteilten Grundgesamtheiten mit gleichen Varianzen stammen. Hier soll nur die grundlegende Vorgehensweise beschrieben werden. Eine ausführliche Herleitung findet man bei [Bortz, 2005]. Der F-Wert berechnet sich wie folgt:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}. \quad (2.13)$$

Wenn die Stichproben normalverteilt sind, lässt sich zeigen, dass die soeben definierte Prüfgröße mit $df_Z = n_1 - 1$ Zählerfreiheitsgraden und $df_N = n_2 - 1$ Nennerfreiheitsgraden F-verteilt ist. Gibt man einen α -Fehler (Irrtumswahrscheinlichkeit) an, so lässt sich ein kritischer F-Wert bestimmen. Ist der aus Gleichung 2.13 errechnete F-Wert größer als der kritische F-Wert, wird die Nullhypothese auf dem vorgegebenen α -Niveau verworfen.

Im oben beschriebenen Fall errechnet man den F-Wert und die bereits erläuterten Freiheitsgrade zu:

$$F = \frac{\hat{\sigma}_{treat}^2}{\hat{\sigma}_{Fehler}^2} \quad (2.14)$$

$$df_Z = df_{treat} = p - 1 \quad (2.15)$$

$$df_N = df_{Fehler} = p \cdot (n - 1). \quad (2.16)$$

Ist der berechnete F-Wert größer als der erwartete F-Wert mit den genannten Freiheitsgraden auf einem $\alpha = 5\%$ -Niveau, bezeichnet man die errechnete Varianzaufklärung als signifikant. Zum A-posteriori-Vergleich auf Unterschiede zwischen einzelnen Faktorstufen, verwendet man häufig den Scheffé-Test. Ist der Unterschied zwischen zwei Mittelwerten größer als eine kritische Differenz $Diff_{crit}$, bezeichnet man den Unterschied als auf einem bestimmten α -Niveau signifikant. Eine gute Einführung ist in [Bortz, 2005] zu finden.

Einfaktorielle Varianzanalyse mit Messwiederholung

Wenn jede Versuchsperson nicht, wie bisher angenommen, nur einen Messwert liefert, entsteht eine Versuchsstruktur mit Messwiederholung. Es ist leicht ersichtlich, dass solch eine Struktur effektiver ist als die bisherige, da mehr Messwerte mit weniger Versuchspersonen gesammelt werden können. Wenn also eine Person gebeten wird, zu allen p Faktorstufen ein Qualitätsurteil zu fällen, kann die Auswertung mit der einfaktoriellen

Varianzanalyse mit Messwiederholung stattfinden. Der Treatmentfaktor (Gitarren) wird als **Messwiederholungsfaktor** bezeichnet, da jede Versuchsperson unter allen Faktorstufen untersucht wird. Es wird wieder die Nullhypothese überprüft, dass es keinen Unterschied zwischen den einzelnen Faktorstufen (Gitarren) gibt. Es ergibt sich ein Datenschema welches in Tabelle 2.4 zu sehen ist.

Tabelle 2.4: Allgemeines Datenschema für eine einfaktorielles Varianzanalyse mit Messwiederholung nach [Bortz, 2005].

		Faktor A						Summen
		1	2	...	i	...	p	
Vpn	1	x_{11}	x_{12}	...	x_{1i}	...	x_{1p}	P_1
	2	x_{21}	x_{22}	...	x_{2i}	...	x_{2p}	P_2
	\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
	m	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mp}	P_m
	\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
	n	x_{n1}	x_{n2}	...	x_{ni}	...	x_{np}	P_n
Summen:		A_1	A_2	...	A_i	...	A_p	G

A_i ist wieder die Summe aller n Messwerte der Treatmentstufe i . P_m ist die Summe über alle Messwerte einer Versuchsperson m . G ist wieder die Gesamtsumme aller Messwerte.

Im vorigen Abschnitt wurde die Unterschiedlichkeit der Messwerte auf Unterschiede zwischen den Faktorstufen (Treatments) und Fehleranteilen zurückgeführt. Im vorliegenden Fall kommen noch Unterschiede zwischen den Versuchspersonen hinzu. So lässt sich die totale Quadratsumme in Unterschiede zwischen Versuchspersonen und Unterschiede innerhalb der Versuchspersonen zerlegen:

$$QS_{tot} = QS_{zwVpn} + QS_{inVpn}. \quad (2.17)$$

Die Unterschiede innerhalb der Versuchspersonen (QS_{inVpn}) gehen wiederum auf Treatmenteffekte (QS_{treat}) und einen zweiten Anteil (QS_{res}) zurück, der sich aus Interaktionen ($Vpn \times Treatment$) sowie Fehlereffekten zusammensetzt:

$$QS_{inVpn} = QS_{treat} + QS_{res}. \quad (2.18)$$

Da nur ein Messwert für jede Kombination aus Versuchsperson und Treatment erhalten wird, sind Interaktionseffekte und Fehlereffekte untrennbar konfundiert. Später soll noch auf eine mehrfaktorielle ANOVA mit Messwiederholung eingegangen werden, bei der auch Interaktionseffekte separierbar sein werden.

Es sollen wieder Varianzschätzungen $\hat{\sigma}^2$ errechnet werden, mit deren Hilfe dann wie oben ein F-Wert zur Signifikanzprüfung aufgestellt wird. Die Varianzschätzungen ergeben sich wieder als Teiler aus Quadratsumme und entsprechendem Freiheitsgrad. Die einzelnen

Quadratsummen errechnen sich nach [Bortz, 2005] zu:

$$QS_{tot} = \sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{G})^2 \quad (2.19)$$

$$QS_{zwVpn} = p \cdot \sum_{m=1}^n (\bar{P}_m - \bar{G})^2 \quad (2.20)$$

$$QS_{inVpn} = \sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{P}_m)^2 \quad (2.21)$$

$$QS_{treat} = n \cdot \sum_{i=1}^p (\bar{A}_i - \bar{G})^2 \quad (2.22)$$

$$QS_{res} = \sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{A}_i - \bar{P}_m + \bar{G})^2 \quad (2.23)$$

$$(2.24)$$

wobei \bar{P}_m der Mittelwert aller Messwerte der Versuchsperson m ist.

Für die Freiheitsgrade ergibt sich entsprechend der Zerlegung der Quadratsummen:

$$df_{tot} = p \cdot n - 1 = df_{zwVpn} + df_{inVpn} \quad (2.25)$$

$$df_{zwVpn} = n - 1 \quad (2.26)$$

$$df_{inVpn} = n \cdot (p - 1). \quad (2.27)$$

Die Freiheitsgrade innerhalb der Versuchspersonen lassen sich noch weiter zerlegen in:

$$df_{inVpn} = df_{treat} + df_{res} \quad (2.28)$$

$$df_{treat} = p - 1 \quad (2.29)$$

$$df_{res} = (n - 1) \cdot (p - 1). \quad (2.30)$$

Der gesuchte F-Wert zur Überprüfung der Nullhypothese ergibt sich wieder, indem man die Varianzschätzung für das Treatment mit der Varianz des Restfehlers vergleicht:

$$F = \frac{\hat{\sigma}_{treat}^2}{\hat{\sigma}_{res}^2}. \quad (2.31)$$

$$(2.32)$$

Das weitere Vorgehen entspricht dem bei der einfaktoriellen Varianzanalyse ohne Messwiederholung, nur dass $\hat{\sigma}_{Fehler}^2$ durch $\hat{\sigma}_{res}^2$ ersetzt wird. Nach [Bortz, 2005] lässt sich eine einfaktorielle Varianzanalyse mit Messwiederholung auch als einfaktorielle Varianzanalyse ohne Messwiederholung darstellen. Dazu zieht man von jedem Messwert einer Versuchsperson den Mittelwert über alle Messwerte der Versuchsperson ab und erhält sogenannte ipsativierte Daten. Die Unterschiedlichkeit zwischen den Versuchspersonen wurde eliminiert. Es können dann unter Berücksichtigung der Voraussetzungen die Methoden der einfaktoriellen Varianzanalyse ohne Messwiederholung angewendet werden.

Mehrfaktorielle Varianzanalyse mit Messwiederholung

In Abschnitt 2.1.1 wurde erläutert, dass nicht nur ein unabhängiger Faktor existiert, sondern mehrere. Bisher ist hier aber nur ein Messwiederholungsfaktor (Gitarren) aufgetreten. Der Einfluss dieses Faktors auf die Varianz aller Messwerte wurde untersucht. Zusätzlich gab es Störfaktoren, deren Einfluss auf die Unterschiedlichkeit der Messwerte in der Residualquadratsumme versteckt war. Ist der Einfluss der Störfaktoren zu groß, erschwert oder verhindert das die signifikante Varianzaufklärung für den Messwiederholungsfaktor. Gelingt es nun, einige Störfaktoren theoretisch zu identifizieren und als unabhängige Variablen in den Versuchsentwurf einzubinden, so ist eine bessere Varianzaufklärung möglich. Man erhält mehrere **Messwiederholungsfaktoren**.

Zusätzlich kann es interessant sein, die Versuchspersonen in Gruppen einzuteilen, von denen man im Voraus annimmt, unterschiedliche Messwerte zu erhalten. Auch dadurch kann eine differenziertere Varianzaufklärung erfolgen. Man spricht von **Gruppierungsfaktoren**.

Im Beispiel könnte man vermuten, dass das Qualitätsurteil für ein Hörbeispiel nicht nur vom Instrument selber (Faktor A = Gitarre), sondern auch vom Spieler abhängt. Lässt man also jede Gitarre von mehreren Spielern spielen, so erhält man einen zusätzlichen Messwiederholungsfaktor (Faktor B = Spieler). Auch das Urteil der Versuchspersonen könnte von ihrem musikalischen Vorwissen abhängen. Das heißt die Versuchspersonen lassen sich in mehrere Hörergruppen einteilen (z.B.: Laie, Gitarrist für einen zweistufigen Faktor). Man spricht von einem Gruppierungsfaktor (Faktor C = Hörergruppe).

An dieser Stelle soll nur auf den später relevanten Sonderfall eines Gruppierungsfaktors und mehrere Messwiederholungsfaktoren eingegangen werden. Eine ausführliche Abhandlung findet sich bei [Bortz, 2005]. Das in Tabelle 2.5 dargestellte Datenschema ist für den Sonderfall eines Gruppierungsfaktors und zweier Messwiederholungsfaktoren konstruiert. Dabei wird eine vereinfachte Darstellungsform gewählt, in der jeder Faktorstufe des Gruppierungsfaktors C eine **Stichprobe** S_i von n Versuchspersonen zugeordnet wird.

Tabelle 2.5: Allgemeines Datenschema für eine mehrfaktorielle Varianzanalyse mit einem Gruppierungsfaktor und zwei Messwiederholungsfaktoren.

	Faktor A	A_1				A_2				A_i				A_p			
	Faktor B	B_1	B_2	B_j	B_q												
Faktor C	C_1	S_1															
	C_2	S_2															
	C_k	S_k															
	C_r	S_r															

Die totale Quadratsumme beinhaltet wieder QS_{zwVpn} und QS_{inVpn} :

$$QS_{tot} = QS_{zwVpn} + QS_{inVpn}. \quad (2.33)$$

Die QS_{zwVpn} setzt sich zum einen aus den Unterschieden zwischen den Stichproben S_k und den Stufen des Faktors C (QS_C) und zum anderen aus den Unterschieden zwischen

den Versuchspersonen innerhalb einer Stichprobe (QS_{inS}) zusammen:

$$QS_{zwVpn} = QS_C + QS_{inS}. \quad (2.34)$$

Innerhalb einer Versuchsperson können Unterschiede mehrere Ursachen haben. Sie können als direkter Effekt eines Faktors auftreten (QS_A oder QS_B), man spricht vom **Haupteffekt** des Faktors. Sie können auf der Interaktion mehrere Faktoren beruhen ($QS_{A \times B}$, $QS_{B \times C}$, $QS_{C \times A}$ oder $QS_{A \times B \times C}$). Oder sie können konfundiert mit Fehleranteilen als Interaktion zwischen den Faktoren und der Versuchsperson auftreten ($QS_{A \times Vpn}$, $QS_{B \times Vpn}$ oder $QS_{A \times B \times Vpn}$). Die letztgenannten entsprechen der QS_{res} in der einfaktoriellen Varianzanalyse mit Messwiederholung. Sie wurden unter anderem benötigt, um nach Teilung durch die entsprechenden Freiheitsgrade, den F-Wert zur Signifikanzbestimmung zu berechnen. Man sieht nun, dass je nach Kombination von Faktorstufen eine andere Residualvarianz verwendet werden muss, mit der der untersuchte Effekt zu vergleichen ist. Die Zerlegung der Varianz innerhalb einer Versuchsperson als Formel lautet:

$$\begin{aligned} QS_{inVpn} = & QS_A + QS_{C \times A} + QS_{A \times Vpn} \\ & + QS_B + QS_{C \times B} + QS_{B \times Vpn} \\ & + QS_{A \times B} + QS_{A \times B \times C} \\ & + QS_{A \times B \times Vpn}. \end{aligned} \quad (2.35)$$

Die Errechnung der einzelnen Quadratsummen folgt dem bisherigen Schema. Sie ist sehr schreibaufwendig und daher soll hier für eine ausführliche Erläuterung wiederum auf [Bortz, 2005] verwiesen werden. Er beschreibt unter anderem die Berechnung von Hand mittels Kennziffern, es wird allerdings schon hier deutlich das für höherfaktorielle Varianzanalysen die Benutzung eines Rechenprogramms unabdingbar wird.

Es lassen sich wiederum die entsprechenden Freiheitsgrade ermitteln und daraus die Schätzungen der Varianzen errechnen mit denen dann ein F-Test zur Signifikanzprüfung ausgeführt werden kann. Man kann anschließend zu jedem F-Wert nach [Bortz, 2005] eine partielle Varianzaufklärung η_p^2 errechnen:

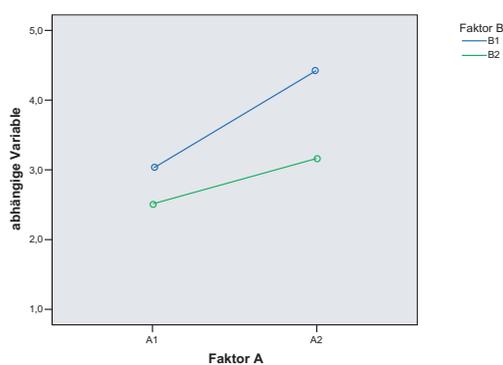
$$\eta_p^2 = \frac{QS_{Effekt}}{QS_{Effekt} + QS_{Prüf}}. \quad (2.36)$$

Interaktionsdiagramme

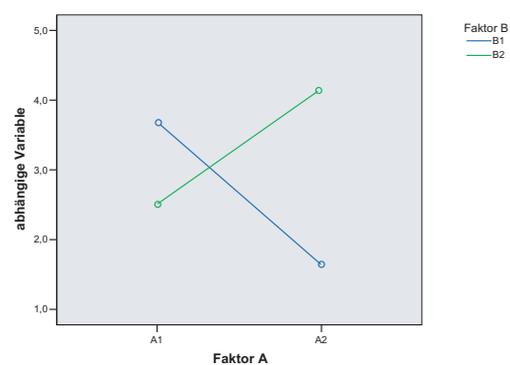
Gibt es nur signifikante Haupteffekte und keine Interaktionen, ist die Interpretation relativ einfach. Werden signifikante Interaktionen festgestellt, müssen die Haupteffekte daran relativiert werden. Es lassen sich zur besseren Interpretation sogenannte Interaktionsdiagramme zeichnen. Die Ordinate bezeichnet dabei immer die abhängige Variable (Mittelwerte). Bei einer Interaktion ersten Grades (Interaktion zwischen zwei unabhängigen Faktoren), wird einer der Faktoren auf der Abszisse abgetragen und der zweite als Parameter eingeführt. Beispiele sind in Abbildung 2.6 dargestellt. Man unterscheidet dabei drei Klassen von Interaktionen. **Ordinale Interaktionen** treten auf, wenn die Linienzüge den gleichen Trend aufweisen (in Abbildung 2.6a) monoton steigend) und sich nicht überschneiden. Beide Haupteffekte lassen sich eindeutig interpretieren. Divergieren die Linienzüge wie in Abbildung 2.6 b) und überschneiden sich, spricht man von einer **disordinalen Interaktion**. Keiner der Haupteffekte hat für sich genommen Bedeutung.

Nur im Zusammenhang der einzelnen Faktorstufen lassen sich inhaltlich richtige Aussagen formulieren. Haben die Linienzüge einen gegenläufigen Trend (Abbildung 2.6 c)) oder überschneiden sich mit gleichem Trend (Abbildung 2.6 d)), spricht man von einer **hybriden Interaktion**. Im ersten Fall c) lässt sich nur der Haupteffekt für Faktor B interpretieren, da für alle Stufen des Faktors A die Rangfolge der Stufen des Faktors B erhalten bleibt. "B1" wird immer besser bewertet als "B2". Im zweiten Fall d) kann nur der Haupteffekt für Faktor A eindeutig interpretiert werden. "A1" wird immer besser bewertet als "A2". Vertauscht man die beiden Faktoren in Abbildung 2.6 d), so erhält man eine Darstellung wie in c).

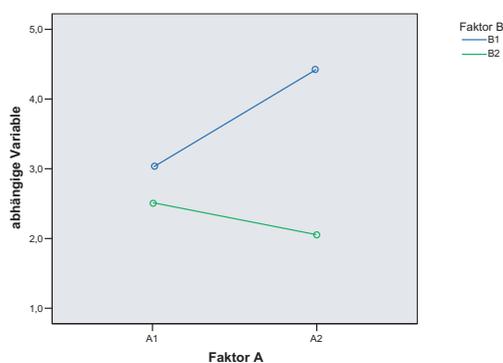
a)



b)



c)



d)

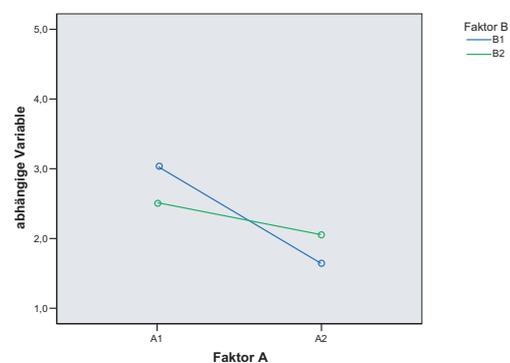


Abbildung 2.6: Beispiele für Interaktionen: a) ordinale Interaktion, b) disordinale Interaktion, c) und d) hybride Interaktionen zwischen dem zweifach gestuften Faktor A und B.

2.5.6 Relevanz eines signifikanten Ergebnisses

Es wurden bisher Unterschiede zwischen den Mittelwerten mehrerer Stichproben untersucht. Wenn nun solch ein Unterschied als signifikant geprüft wurde, stellt sich die Frage ob er auch von praktischer Bedeutung ist. Dies hängt davon ab, wie groß der Unterschied zwischen den Mittelwerten ist. Weichen die Mittelwerte nur geringfügig voneinander ab, kann die praktische Bedeutung möglicherweise sehr gering sein. (Es stellt sich zum Beispiel

die Frage, ob ein erhöhter Produktionsaufwand die minimale klangliche Qualitätsverbesserung einer Gitarre lohnen würde oder nicht.) Um zu untersuchen, ob sich die Mittelwerte von zwei unabhängigen Stichproben unterscheiden, definiert man nach [Bortz, 2005] eine Effektgröße ε in der der Unterschied zwischen zwei Mittelwerten μ_1 und μ_2 und die Streuung des untersuchten Merkmals σ eingeht:

$$\varepsilon = \frac{\mu_1 - \mu_2}{\sigma}. \quad (2.37)$$

Die Effektgröße ε lässt sich auch über die Varianzaufklärung η^2 berechnen:

$$\varepsilon = \sqrt{\frac{\eta^2}{1 - \eta^2}}. \quad (2.38)$$

Für mehrfaktorielle varianzanalytische Untersuchungen mit Messwiederholung soll für die Herleitung der entsprechenden Effektgrößen auf [Bortz and Döring, 2001] verwiesen werden. Man berechnet die partielle Effektgröße ε_p für jeden Haupteffekt zu:

$$\varepsilon_p = \sqrt{\frac{\eta_p^2}{1 - \eta_p^2}}. \quad (2.39)$$

Man spricht in der Varianzanalyse von einem kleinen, mittlerem bzw. großen Effekt, wenn die Effektgröße $\varepsilon > 0,10; 0,25$ bzw. $0,40$ oder die Varianzaufklärung $\eta^2 > 0,01; 0,06$ bzw. $0,14$ ist. Welcher Effekt als praktisch relevant betrachtet wird, kann man im Voraus einer Untersuchung festlegen. Nach dem Versuch lassen sich die in der untersuchten Stichprobe tatsächlich auftretenden Effektgrößen anhand der erhobenen Daten schätzen. In weiteren Untersuchungen soll zur Abschätzung der Effektgröße immer die geschätzte partielle Varianzaufklärung $\hat{\eta}_p^2$ angegeben werden.

Kompensation der Kopfhörercharakteristik

Nach [Møller et al., 1995] ist es für binaurale Systeme ausschlaggebend, dass die Übertragung von der Aufnahme- zur Wiedergabesituation korrekt erfolgt, um nicht räumliche Information hinzuzufügen oder zu entfernen und die Klangfarbe nicht zu verändern. Dies ist in dieser Arbeit von besonderer Bedeutung, da mit dem System Hörversuche zur Beurteilung von Qualität durchgeführt werden sollen. Das Qualitätsurteil wird von vielen Parametern beeinflusst, deren Bedeutung und Interaktion nicht bekannt sind. Es ist also wichtig, den Einfluss des Aufnahme- und Wiedergabesystems so gering wie möglich zu halten um das Ergebnis nicht unbewusst zu verfälschen. Damit wird die Grundlage für interpretierbare Hörversuche gelegt. Es folgt, dass der Kopfhörer einen flachen Frequenzgang besitzen sollte, entweder direkt oder in Kombination mit einem Kompensationsfilter. Dieses Kapitel soll sich ausführlich mit der Kompensation für die Kopfhörercharakteristik beschäftigen. In [Møller et al., 1995] wird weiterhin festgestellt, dass interindividuelle Variationen für Kopfhörer auf verschiedenen Personen relativ groß sind und so im Prinzip eine individuelle Kompensation nötig ist.

Dazu muss der Frequenzgang des Kopfhörers auf dem Ohr bekannt sein. Abschnitt 3.1 soll ein einfaches Modell für die Schallübertragung im Außenohr nach [Møller, 1992] erläutern. Daraus kann in Abschnitt 3.2 und 3.3 die Messvorschrift zur Bestimmung der Kopfhörercharakteristik hergeleitet werden. Die Messung ist in Abschnitt 3.4, 3.5, 3.6 und 3.7 beschrieben. Abschnitt 3.8 behandelt den Entwurf eines digitalen Kompensationsfilters für den Kopfhörer.

3.1 Modell für Schallübertragung im Ohr

Das menschliche Ohr kann nach [Møller, 1992] durch ein einfaches elektrisches Thevenin-Modell approximiert werden. In Abbildung 3.1 ist das Außenohr anatomisch dargestellt und in Abbildung 3.2 das äquivalente elektrische Modell. P_{offen} ist der Schalldruck am Eingang des nicht verschlossenen Gehörganges, $P_{\text{geschlossen}}$ ist der Druck am verschlossenen Eingang des Gehörganges, $P_{\text{Trommelfell}}$ ist der Druck vor dem Trommelfell, $Z_{\text{Gehörgang}}$ ist die Impedanz in den offenen Gehörgang hinein gemessen, $Z_{\text{Abstrahlung}}$ ist die Impedanz

am selben Ort, aber nach außen in die freie Umgebung hinein gesehen und $Z_{\text{Trommelfell}}$ ist die Impedanz zwischen Gehörgang und Trommelfell.

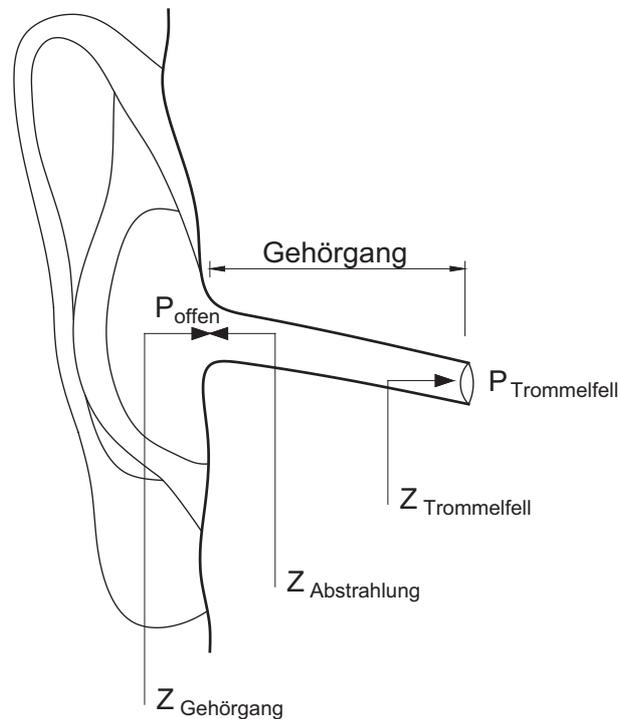


Abbildung 3.1: Schallübertragung durch das Außenohr nach [Møller et al., 1995] anatomisch dargestellt.

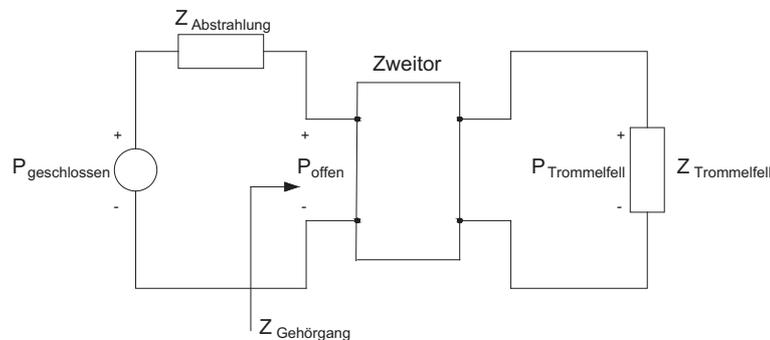


Abbildung 3.2: Schallübertragung durch das Außenohr nach [Møller, 1992] als analoges elektrisches Thevenin-Modell.

Nach [Møller, 1992] ergibt Thevenins Theorem am Eingang des Gehörganges eine Aufteilung in den Leerlaufdruck $P_{\text{geschlossen}}$ und den Quellwiderstand $Z_{\text{Abstrahlung}}$. Der Leerlaufdruck $P_{\text{geschlossen}}$ ist der Schalldruck welcher an der Stelle P_{offen} , also am Gehörkanaleingang, existieren würde, wenn keine Schallschnelle (Strom) durch $Z_{\text{Abstrahlung}}$ fließt. Keine Schallschnelle ist nur für den Fall eines physikalisch verschlossenen Gehörkanals möglich, woher die Bezeichnung $P_{\text{geschlossen}}$ stammt.

Der Druck $P_{\text{geschlossen}}$ existiert also nicht in einer natürlichen Hörsituation. Um ihn zu messen, kann der Gehörgang verschlossen werden. In solch einer Situation existieren dann natürlich P_{offen} und $P_{\text{Trommelfell}}$ nicht.

Das Verhältnis zwischen $P_{\text{Trommelfell}}$ und $P_{\text{geschlossen}}$ ergibt sich zu:

$$\frac{P_{\text{Trommelfell}}}{P_{\text{geschlossen}}} = \frac{P_{\text{offen}}}{P_{\text{geschlossen}}} \cdot \frac{P_{\text{Trommelfell}}}{P_{\text{offen}}} \quad (3.1)$$

Aus dem analogen Ersatzschaltbild ergibt sich das Verhältnis zwischen P_{offen} zu $P_{\text{geschlossen}}$ als Druckteiler zwischen $Z_{\text{Abstrahlung}}$ und Z_{Ohrkanal} :

$$\frac{P_{\text{offen}}}{P_{\text{geschlossen}}} = \frac{Z_{\text{Gehörgang}}}{Z_{\text{Abstrahlung}} + Z_{\text{Gehörgang}}} \quad (3.2)$$

Wenn ein Kopfhörer über das Ohr gesetzt wird, verändert sich die Impedanz $Z_{\text{Abstrahlung}}$, da vom Ohrkanal aus nicht mehr in die freie Umgebung hinaus gesehen werden kann. Wenn es sich um ohrumschließende Kopfhörer handelt, wird der Ohrkanal durch ein kleines Volumen und die Kopfhörermembran nach außen hin abgeschlossen. Die vom Eingang des Gehörganges in den Kopfhörer hinaus gesehene Impedanz soll $Z_{\text{Kopfhörer}}$ heißen. Wenn sich nun $Z_{\text{Abstrahlung}}$ zu $Z_{\text{Kopfhörer}}$ ändert, so ändert sich auch das Druckverhältnis zwischen P_{offen} und $P_{\text{geschlossen}}$ aus Gleichung 3.2 zu:

$$\frac{P_{\text{offen}}^*}{P_{\text{geschlossen}}^*} = \frac{Z_{\text{Gehörgang}}}{Z_{\text{Kopfhörer}} + Z_{\text{Gehörgang}}} \quad (3.3)$$

Bei aufgesetztem Kopfhörer werden die Schalldrücke von nun an mit einem Stern gekennzeichnet.

3.2 Herleitung des Korrekturfilters

Die Aufnahme der Hörbeispiele wurde mit einem Kunstkopf mit verschlossenen Gehörgängen durchgeführt. Eine umfassend Beschreibung der Aufnahmen befindet sich in Abschnitt 4.1.3. Die Übertragungsfunktion des verwendeten Mikrofons soll M_1 heißen. Am Ausgang des Mikrofons ergibt sich somit die Spannung:

$$U_{\text{Mikrofon}} = P_{\text{geschlossen}} \cdot M_1 \quad (3.4)$$

Vor der Wiedergabe soll ein Korrekturfilter G_K zur Entzerrung der Kopfhörercharakteristik verwendet werden. Die Wiedergabesituation lässt sich durch die Übertragungsfunktion vom elektrischen Eingang des Kopfhörers ($E_{\text{Kopfhörer}}$) zum Schalldruck am Trommelfell ($P_{\text{Trommelfell}}^*$) beschreiben. Gewünscht ist der Schalldruck der am Trommelfell des Probanden in der Aufnahmesituation existiert hätte ($P_{\text{Trommelfell}}$).

$$\underbrace{P_{\text{Trommelfell}}}_{\text{gewünschter Druck}} = \underbrace{P_{\text{geschlossen}} \cdot M_1}_{\text{Aufnahme}} \cdot \underbrace{G_K}_{\text{Korrekturfilter}} \cdot \underbrace{\frac{P_{\text{Trommelfell}}^*}{E_{\text{Kopfhörer}}}}_{\text{Wiedergabe}} \quad (3.5)$$

Nach G_K aufgelöst ergibt sich:

$$G_K = \frac{\frac{P_{\text{Trommelfell}}}{P_{\text{geschlossen}}}}{M_1 \cdot \frac{P_{\text{Trommelfell}}^*}{E_{\text{Kopfhörer}}}} \quad (3.6)$$

Die Übertragungsfunktionen lassen sich nun wie in Gleichung 3.1 zerlegen und es ergibt sich:

$$G_K = \frac{\frac{P_{\text{Trommelfell}}}{P_{\text{offen}}} \cdot \frac{P_{\text{offen}}}{P_{\text{geschlossen}}}}{M_1 \cdot \frac{P_{\text{Trommelfell}}^*}{P_{\text{offen}}^*} \cdot \frac{P_{\text{offen}}^*}{P_{\text{geschlossen}}^*} \cdot \frac{P_{\text{geschlossen}}^*}{E_{\text{Kopfhörer}}}} \quad (3.7)$$

Die Übertragungsfunktionen vom Druck am offenen Gehörkanaleingang zum Trommelfell sind nicht abhängig von $Z_{\text{Abstrahlung}}$ bzw. $Z_{\text{Kopfhörer}}$ und können somit gekürzt werden. Das Druckverhältnis vom offen zum geschlossenen Gehörgang verändert sich, wie in Gleichung 3.2 und 3.3 gesehen, in Abhängigkeit vom Quellwiderstand. Die zwei entsprechenden Druckverhältnisse werden daher zum PDR, dem so genannten "Pressure Division Ratio" ([Møller et al., 1995]), zusammengefasst:

$$\text{PDR} = \frac{P_{\text{offen}}}{P_{\text{geschlossen}}} \cdot \frac{P_{\text{geschlossen}}^*}{P_{\text{offen}}^*} = \frac{Z_{\text{Kopfhörer}} + Z_{\text{Gehörgang}}}{Z_{\text{Abstrahlung}} + Z_{\text{Gehörgang}}} \quad (3.8)$$

Gleichung 3.7 lässt sich damit umschreiben zu:

$$G_K = \frac{\text{PDR}}{M_1 \cdot \frac{P_{\text{geschlossen}}^*}{E_{\text{Kopfhörer}}}} \quad (3.9)$$

Es ist ersichtlich, dass der Korrekturfaktor aus einem Korrekturfaktor (PDR), der inversen Übertragungsfunktion des Kopfhörers (gemessen auf einem Kopf mit verschlossenen Ohrkanälen), sowie der Übertragungsfunktion des verwendeten Mikrofons besteht.

3.3 Vereinfachungen

Gleichung 3.9 vereinfacht sich für den Fall das der PDR-Wert gleich Eins ist. Dies ist nach Gleichung 3.8 der Fall, wenn

$$Z_{\text{Kopfhörer}} \approx Z_{\text{Abstrahlung}} \quad (3.10)$$

$Z_{\text{Abstrahlung}}$ ist nach [Møller, 1992] wesentlich kleiner als $Z_{\text{Gehörgang}}$ für Frequenzen unter 1 kHz und damit vereinfacht sich diese Bedingung zu

$$Z_{\text{Kopfhörer}} \ll Z_{\text{Gehörgang}} \quad (3.11)$$

Ein Kopfhörer für den diese Bedingung gilt, bezeichnet man nach [Møller, 1992] als "open". Diese Eigenschaft wird auch FEC (Free air Equivalent Coupling) genannt. Charakteristisch ist, dass der Kopfhörer den Quellwiderstand vom Eingang des Gehörganges nach außen gesehen nicht verändert. Dies muss nicht dasselbe sein wie ein offener Kopfhörer, da darunter meist nur Verstanden wird, das Schall von außen gut hörbar ist.

Der PDR-Wert ist Abhängig von $Z_{\text{Gehörgang}}$ und damit von der Versuchsperson. Er ist schwer zu messen, da der Schall im Gehörgang durch den eingebrachten Messaufnehmer gestört wird und die Positionierung desselben genau erfolgen muss. In [Møller et al., 1995] wurde unter anderem der PDR-Verlauf von 14 Kopfhörern auf 40 Probanden gemessen. Diese Messung war nur verlässlich bis etwa 7 kHz. Mit einer erlaubten Mittelwertabweichung des PDR von maximal 2 dB konnten einige der Kopfhörer als "FEC" eingestuft

werden. Bei einer erlaubten Mittelwertabweichung von 4 dB alle. Darunter befand sich auch der elektrostatische Kopfhörer "SR lambda professional" der Firma STAX zusammen mit seiner Verstärkereinheit (SRM-1/MK-2), der dem hier verwendeten Kopfhörer, SR-202 von STAX, ähnelt. Daher wird weiterhin angenommen, dass auch dieser Kopfhörer als "FEC" einstuftbar ist und das PDR zu eins genähert werden kann.

Die Übertragungsfunktion des Kopfhörers zusammen mit dem Messmikrofon ist nun bei verschlossenem Gehörgang zu messen, um danach einen digitalen Kompensationsfilter für G_K zu entwerfen (Abschnitt 3.8).

3.4 Frequenzgangmessung des Kopfhörers auf dem Kunstkopf

Der Hersteller des Kopfhörers gibt im dazugehörigen Datenblatt [STAX, 2005] einen Frequenzgang an, der mit einem künstlichen Ohr der Firma B&K gemessen wurde (Anhang B, Abbildung B.1). Mit dieser Messmethode wird versucht die Übertragungsfunktion von Kopfhörereingang zum Schalldruck am Trommelfell zu erfassen.

Wie in Abschnitt 3.2 beschrieben, ist es für die Kopfhörerkompensation günstig eine Kopfhörmessung mit geschlossenem Gehörgang durchzuführen, wenn die Hörbeispiele mit geschlossenem Gehörgang aufgenommen wurden. Dabei bietet es sich an die Messung der Kopfhörer direkt auf dem zur Aufnahme verwendeten Kunstkopf MK-2 der Firma Cortex durchzuführen. Dieser ist mit zwei Kondensatormikrofonen ausgestattet. Zur Kalibrierung müssen diese aus dem Kunstkopf ausgebaut werden, was in regelmäßigen Abständen erfolgte.

Die für den Hörversuch bestimmten Kopfhörer (SR-202, STAX) wurden zusammen mit der zugehörigen Verstärkereinheit (SRM-212, STAX) verwendet. Die Lautstärkeeinstellung am Verstärker wurde beliebig in einer mittleren Position (Drehknopfmarkierung senkrecht nach oben) gewählt und über den gesamten Versuch konstant gehalten. Die Kopfhörer sollen weiterhin mit "KH1" (Inv.-Nr. 135930) und "KH2" bezeichnet werden.

Da sich der Frequenzgang des Kopfhörers in Abhängigkeit von seiner Position ändert, wurde über fünf Positionen gemittelt. Nacheinander fanden fünf Messungen statt zwischen denen der Kopfhörer vom Kunstkopf genommen und wieder zurückgesetzt wurde. Es wurde darauf geachtet, dass der Kopfhörer das Ohr bestmöglich umschließt. Dabei wurde festgestellt, dass der Kopfhörer KH2 nicht so straff sitzt wie KH1 und dadurch manchmal kleine Öffnungen zwischen Kopfhörer und Kopf bleiben.

Zur Ermittlung der Impulsantwort der Kopfhörer wurden zwei Messsysteme verwendet. So können später die Ergebnisse zu Kontrollzwecken verglichen werden.

3.4.1 Messsystem 1: MLSSA

Das Messsystem MLSSA (Maximum-Length Sequence System Analyzer) basiert auf einem speziellen MLS-Signal. Das generelle Funktionsprinzip ist in Anhang A beschrieben. Der Messaufbau ist in Abbildung 3.3 dargestellt. A/D und D/A-Wandler sind bereits in das Kunstkopfsystem MK-2 integriert.

Die Abtastfrequenz sollte zu 96 kHz gewählt werden um ein einfaches Verringern der Abtastfrequenz auf 48 kHz zu ermöglichen. Der interne Taktgeber des MLSSA-Systems zeigte

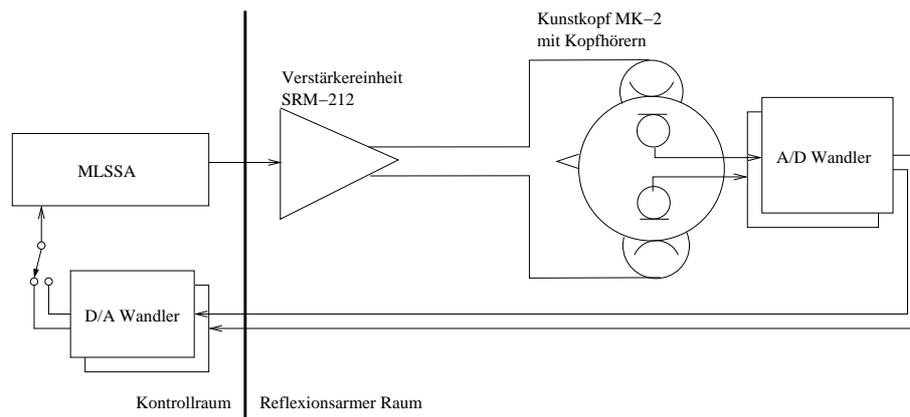


Abbildung 3.3: Messaufbau mit MLSSA-System um die Impulsantwort des Kopfhörers zu bestimmen.

an, das nur 95, 24 kHz erzeugbar sind und so wurde diese Abtastfrequenz gewählt. Es wurde ein MLS-Signal 16. Ordnung verwendet. Das entspricht einer Stimuluslänge von 65535 Abtastwerten. Die Länge des Stimulus sollte nach [Rife and Vanderkooy, 1998] mindestens so lang sein wie die zu messende Impulsantwort. Je länger der Stimulus gewählt wird desto höher ist jedoch die Frequenzauflösung. Dies ist wichtig, wenn der Frequenzgang bei tiefen Frequenzen interessiert. Die Stimulusamplitude betrug etwa $\pm 0,5$ V.

Die Messungen wurden mit einer Mittelung über 160 Impulse durchgeführt (16 pre-averages \cdot 10 averages = 160). Dadurch verbessert sich der Signal-Rausch-Abstand um etwa 20 dB. Eine Aufzeichnung des Grundrauschens wurde nicht angefertigt. Der Signal-Rausch-Abstand der Messung ist allerdings nicht von großer Bedeutung, da die gemessenen Frequenzgänge in der weiteren Verarbeitung noch geglättet werden.

Die erhaltenen Impulsantworten wurden gespeichert und später in Matlab ausgewertet. Als Beispiel soll hier die in Abbildung 3.5 dargestellte Impulsantwort für den Kopfhörer KH1, linke Seite, betrachtet werden. Es ist zu erkennen das die Impulsantwort bereits nach 50 ms (4762 Abtastwerten) weitgehend abgeklungen ist. Die gewählte Länge des MLS-Signals (65535 Abtastwerte) beträgt damit ein vielfaches der Impulsantwortlänge.

Die gemessenen Frequenzverläufe für den Kopfhörer KH1 sind in Abbildung 3.6 und 3.7 dargestellt. In Abbildung 3.6a) und 3.7a) sind schwache Abweichungen des Frequenzgangs oberhalb von 2 kHz in Abhängigkeit von der Kopfhörerposition erkennbar. Bei etwa 100 Hz und 2 kHz treten zwei charakteristische Resonanzen auf. Unterhalb von 100 Hz fällt der Frequenzverlauf stark ab. Oberhalb von 2 kHz wird der Amplitudengang komplexer, bis er bei etwa 14 kHz stark einbricht. Bildet man den Mittelwert der Messungen, so kann dies zum Beispiel auf dB-Basis, auf Druck-Basis oder auf Energie-Basis erfolgen. In [Møller et al., 1995] wird argumentiert, dass bei einer Mittelung auf Energie-Basis, im Vergleich zur dB- oder Druck-Basis, Spitzen im Frequenzverlauf stärker bedämpft werden als Einbrüche. Da scharfe Spitzen als tonale Komponente störender wahrgenommen werden als Einbrüche, wäre nach [Møller et al., 1995] aus psychoakustischer Sicht ein Mittelung des Energiespektrums sinnvoll, was durch Hörversuche noch zu bestätigen ist. Die Unterschiede zwischen den Methoden sind allerdings nicht sehr groß, so dass der Autor sich für eine Mittelung auf dB-Basis entscheidet. Dadurch werden in der gängigen graphischen dB-Darstellung, Spitzen und Einbrüche gleich stark bewertet.

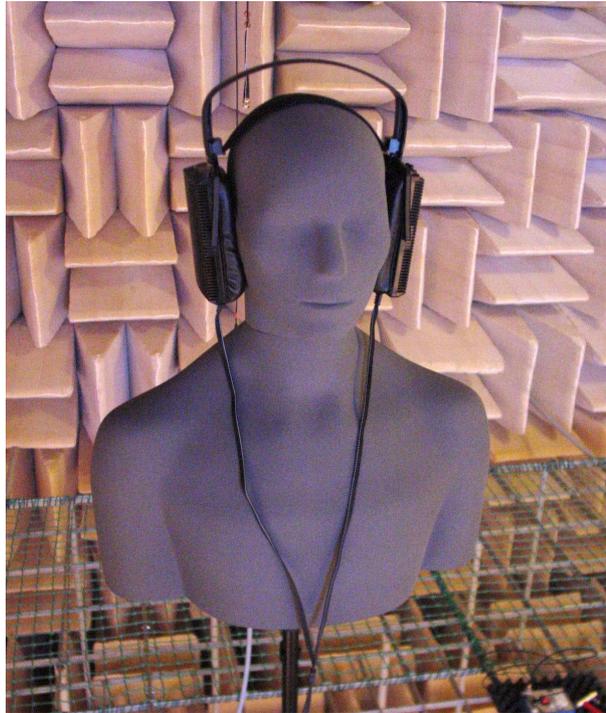


Abbildung 3.4: Messung des Kopfhörers KH1 auf dem Cortex Kunstkopf MK-2 im reflexionsarmen Raum des Instituts für Musikinstrumentenbau Zwota.

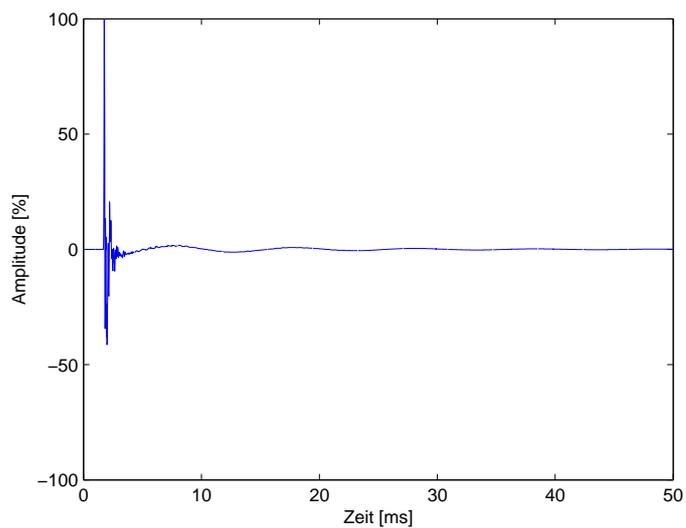


Abbildung 3.5: Impulsantwort (MLSSA) für die linke Seite des Kopfhörers KH1, beschränkt auf die ersten 50 ms (4762 Abtastwerte).

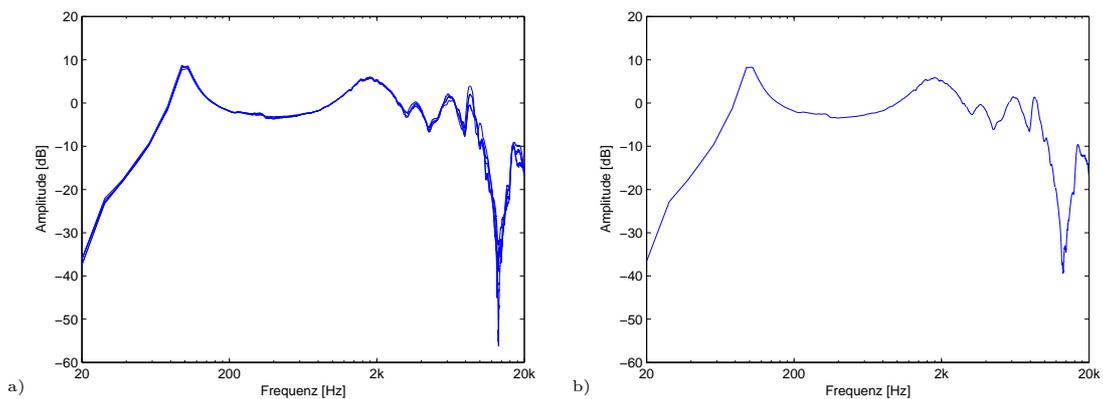


Abbildung 3.6: Frequenzgang (MLSSA) für die linke Seite des Kopfhörers KH1 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

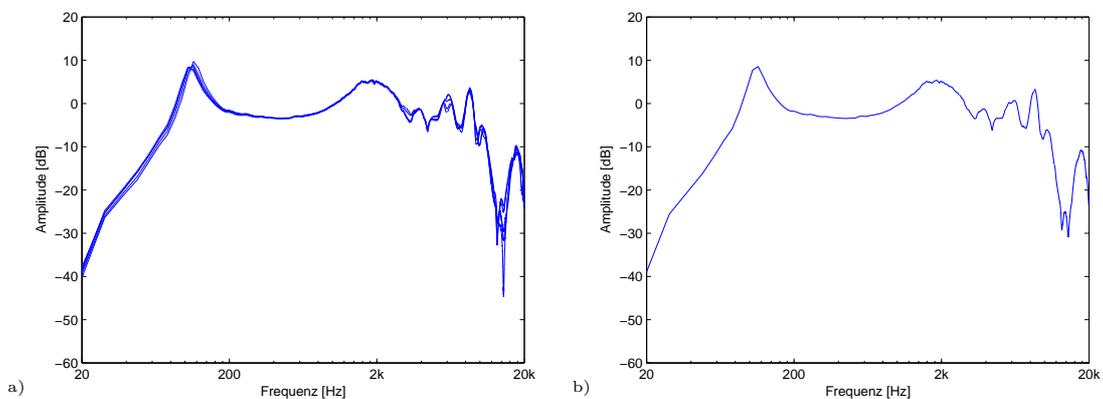


Abbildung 3.7: Frequenzgang (MLSSA) für die rechte Seite des Kopfhörers KH1 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

Die Frequenzverläufe für den Kopfhörer KH2 verlaufen ähnlich und sind in Anhang B dargestellt.

3.4.2 Messsystem 2: ONO SOKKI

Der vierkanalige FFT-Analysator CF-6400 von ONO SOKKI wurde für eine Vergleichsmessung eingesetzt. Es wurden Gleitsinusse als Stimuli gewählt. Die Stimulusamplitude betrug wiederum etwa 0,5 V. Der Analysator errechnet aus der Systemantwort mit einer FFT den Frequenzgang und speichert ihn zusammen mit der Impulsantwort ab. Die FFT-Länge wurde zu 8192 Abtastwerten gewählt. Der Versuchsaufbau war wie in Abbildung 3.3, nur das MLSSA-System wurde durch den FFT-Analysator ersetzt.

Die damit erhaltenen Amplitudenverläufe für linke und rechte Seite des Kopfhörers KH1 können in Abbildung 3.8 und 3.9 gesehen werden. Es sind wieder unter a) die Verläufe für fünf Kopfhörerpositionen und unter b) der gemittelte Verlauf zu sehen. Die Frequenzgänge für Kopfhörer KH2 finden sich im Anhang B.

Vergleicht man die gemittelten Kurven der Messungen mit dem System MLSSA und

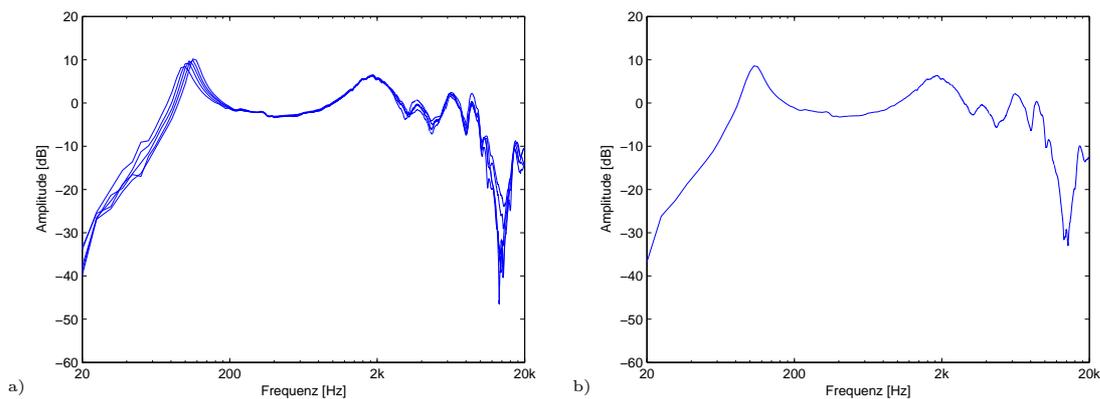


Abbildung 3.8: Frequenzgang (ONO) für die linke Seite des Kopfhörers KH1 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

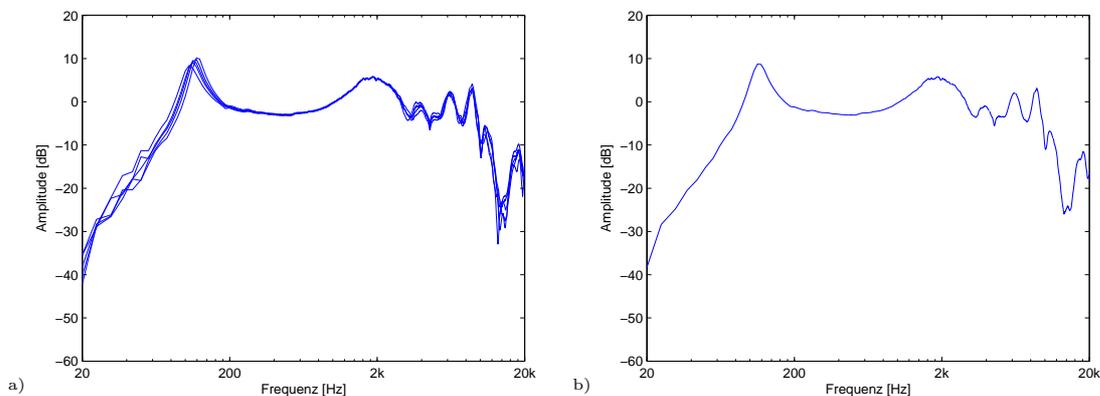


Abbildung 3.9: Frequenzgang (ONO) für die rechte Seite des Kopfhörers KH1 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

CF-6400 von ONO SOKKI für Kopfhörer KH1, so ergibt sich der in Abbildung 3.10 dargestellte Verlauf. Man erkennt das sich die Ergebnisse beider Messmethoden nur minimal unterscheiden. Für zukünftige Messungen soll nur noch das MLSSA-System verwendet werden. Der Vergleich für Kopfhörer KH2 ist wieder im Anhang B zu finden.

Es stellt sich die Frage, ob die Kopfhörmessung auf dem Kunstkopf die Kopfhörercharakteristik auf realen Köpfen korrekt widerspiegelt. Ein Vergleich von Kunstkopfmessung und Messung des Kopfhörerherstellers auf einem Kunstohr ist nur bedingt möglich, da unterschiedliche Übertragungsfunktionen nachgebildet wurden ($\frac{P_{\text{geschlossen}}^*}{E_{\text{Kopfhörer}}}$ sowie $\frac{P_{\text{Trommelfell}}^*}{E_{\text{Kopfhörer}}}$). Um zu überprüfen ob der Kunstkopf das Verhalten des Kopfhörers auf einem durchschnittlichen Kopf repräsentiert, soll eine Vergleichsmessung auf realen Köpfen durchgeführt werden (Abschnitt 3.5). Zum Vergleich mit den Herstellerangaben soll der Kopfhörer zusätzlich mit einem Ohrsimulator der Firma B&K vermessen werden (Abschnitt 3.6).

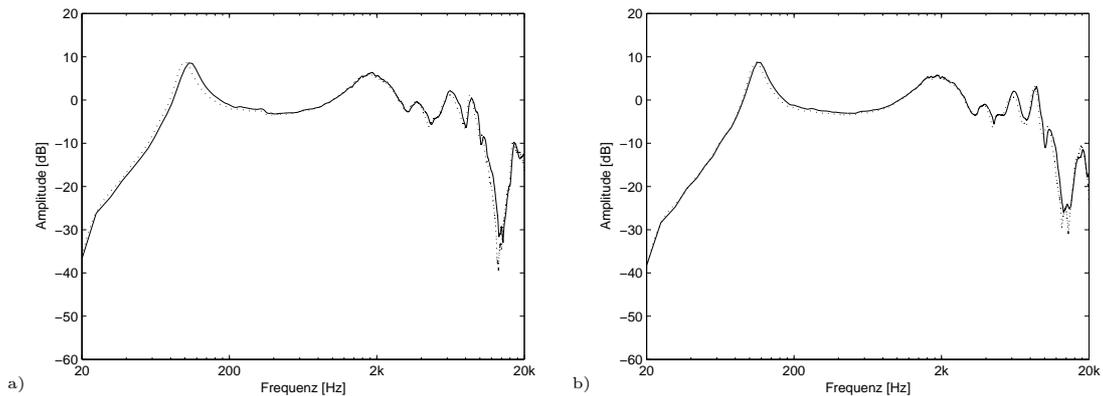


Abbildung 3.10: Vergleich der über fünf Positionen gemittelten Frequenzgänge gemessen mit dem System MLSSA (schwarz) und ONO (gestrichelt) für den Kopfhörers KH1 a) linke Seite, b) rechte Seite.

3.5 Frequenzgangmessung des Kopfhörers mit realen Probanden

Es sollte wieder eine Messung mit geblockten Ohrkanaleingang durchgeführt werden. Der Messaufbau ist in Abbildung 3.11 zu sehen.

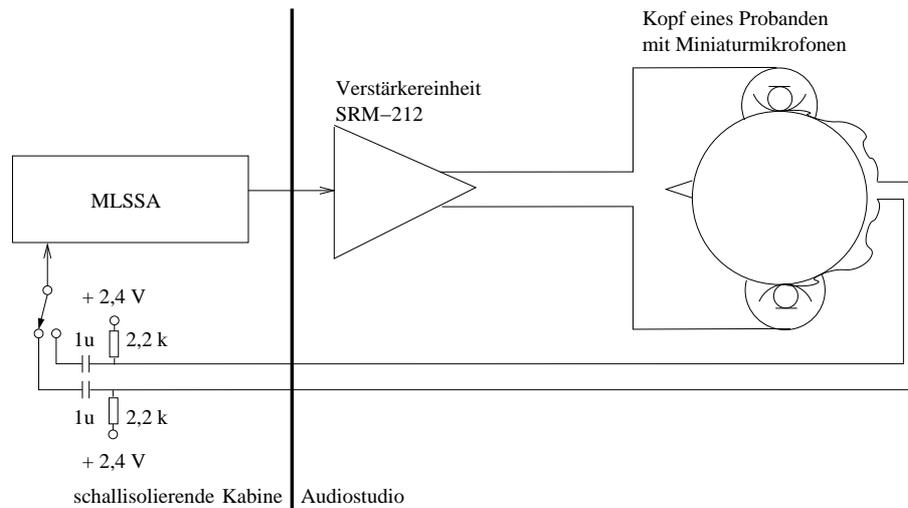


Abbildung 3.11: Messaufbau mit MLSSA-System um die Impulsantwort des Kopfhörers auf realen Probanden zu bestimmen.

Diesmal sollten an Stelle des Kunstkopfes reale Probanden zum Einsatz kommen. Dafür war es notwendig Miniaturmikrofone in den Gehörgang einzusetzen. Es wurden Ohrstöpsel aus Schaumstoff (ähnlich dem Ohrstöpsel "E-A-R classic") mit einem Loch versehen und das Miniaturmikrofon eingesteckt, so dass es bündig mit der Oberfläche abschloss. Durch ein zweites seitliches Loch führten die dünnen Anschlussdrähte nach außen. Der Ohrstöpsel wurde (je nach Proband) um etwa 4 mm verkürzt, zusammengedrückt und in das Ohr eingesetzt, so dass er bündig mit dem Gehörgang abschloss. Ein entsprechendes Foto ist in Abbildung 3.12 zu sehen.



Abbildung 3.12: Mit Miniaturmikrofon im geblockten Ohrkanal präpariertes Ohr zur Messung des Kopfhörerfrequenzgangs.

Die Messung wurde ähnlich wie die Kunstkopfmessung durchgeführt. Allerdings kam nur der Kopfhörer KH1 zum Einsatz. Das MLS-Signal wurde auf $2^{14} - 1$ Werte (MLS-Signal 14. Ordnung) verkürzt, wodurch sich die Frequenzauflösung verringert und die Messzeit verkürzt. Jeder Proband platzierte jeweils zwischen den Messungen den Kopfhörer neu, so dass wieder für jede Seite fünf Positionen gemessen wurden. Acht Probanden nahmen am Versuch teil. Die ermittelten Frequenzgänge für linke und rechte Seite des Kopfhörers KH1 sind in Abbildung 3.13 und 3.14 zu sehen. Unterhalb von 500 Hz sind Verläufe zunehmend verrauscht. Dies soll in Abschnitt 3.5.1 näher erläutert werden. Zwischen 500 Hz und 5 kHz ist eine einheitliche Struktur erkennbar. Oberhalb von 5 kHz ist die Übertragungsfunktion durch individuell charakteristische Einbrüche geprägt. Man kann teils große Abweichungen zwischen den Kopfhörerpositionen innerhalb einer Versuchsperson erkennen.

Die Messungen sollen wieder auf dB-Basis gemittelt werden. Die gemittelten Verläufe sind in Abbildung 3.15 und 3.16 zu sehen. Durch die Mittelung ergibt sich eine "natürliche" Glättung des Verlaufs und scharfe Einbrüche werden reduziert. Es sind trotzdem noch Unterschiede zwischen den Versuchspersonen erkennbar. In [Møller et al., 1995] wird eine individuelle Kompensation für den Frequenzgang des Kopfhörers zur Wiedergabe binauraler Aufnahmen empfohlen, eine allgemeine Kompensation aber nicht ausgeschlossen. Da eine individuelle Kopfhörerkompensation sehr aufwendig erscheint, wird der Mittelwert für jede Kopfhörerseite über alle Probanden gebildet. Die Ergebnisse sind in Abbildung 3.17 und 3.18 zu sehen. Die Einbrüche im Frequenzgang werden durch die Mittelung weiter bedämpft. Die Kurven für linke und rechte Seite (KH1) ähneln sich sehr. Der Vergleich zur vorangegangenen Messung soll in Abschnitt 3.7 folgen. Zunächst soll im nächsten Abschnitt 3.5.1 das starke Rauschen unterhalb von 500 Hz näher untersucht werden.

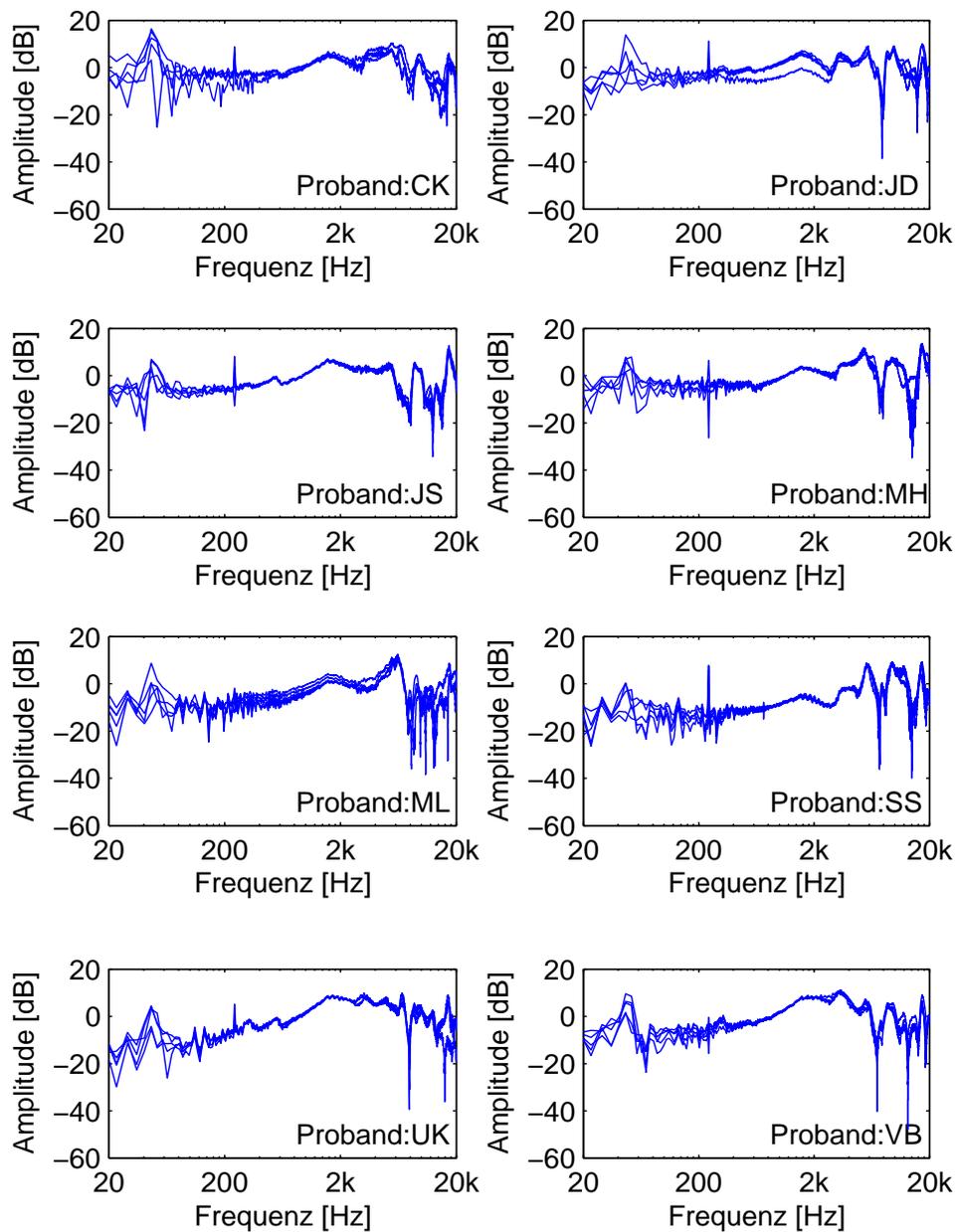


Abbildung 3.13: PTFs Kopfhörer KH1 (linke Seite) für acht Probanden und jeweils fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

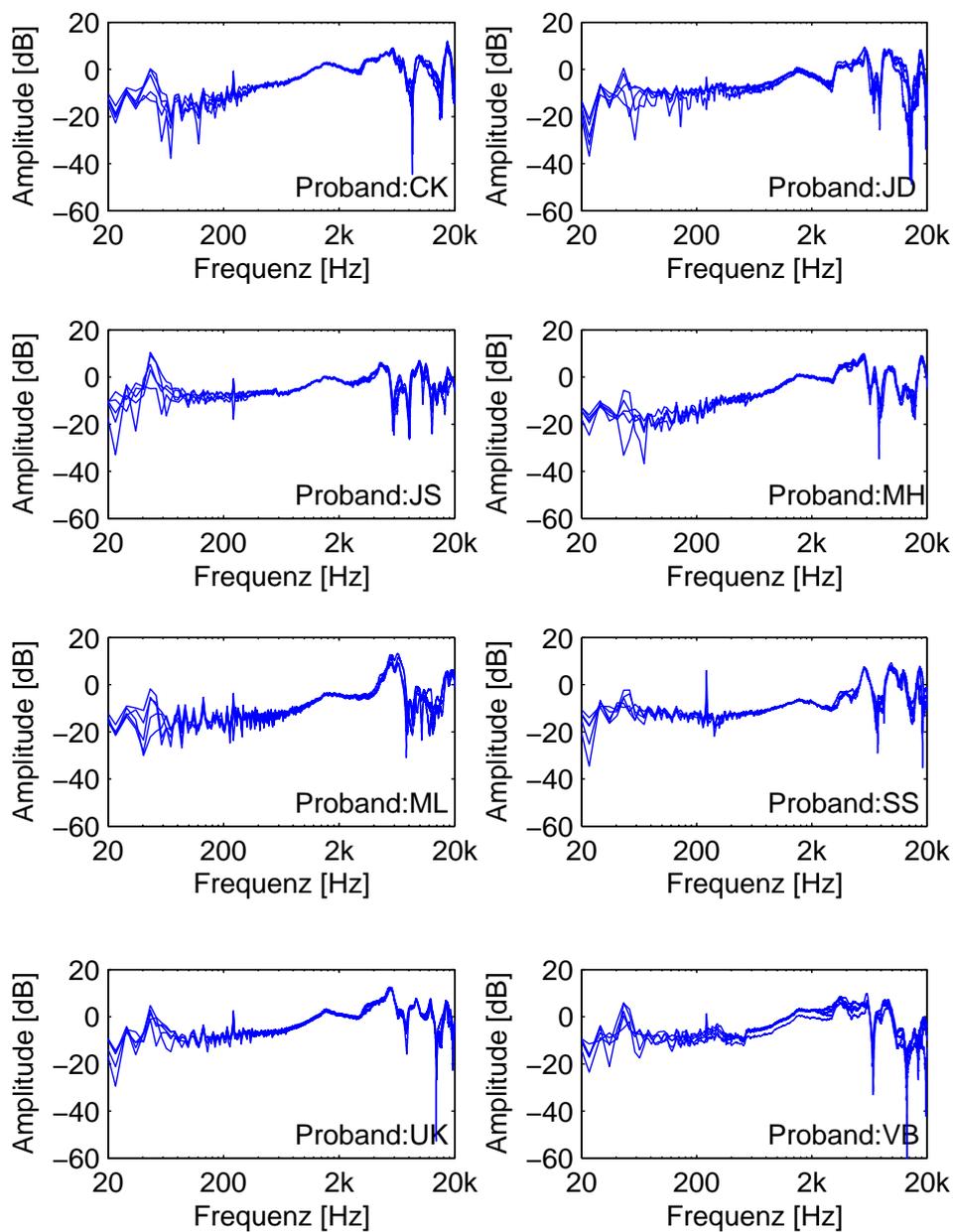


Abbildung 3.14: PTFs Kopfhörer KH1 (rechte Seite) für acht Probanden und jeweils fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

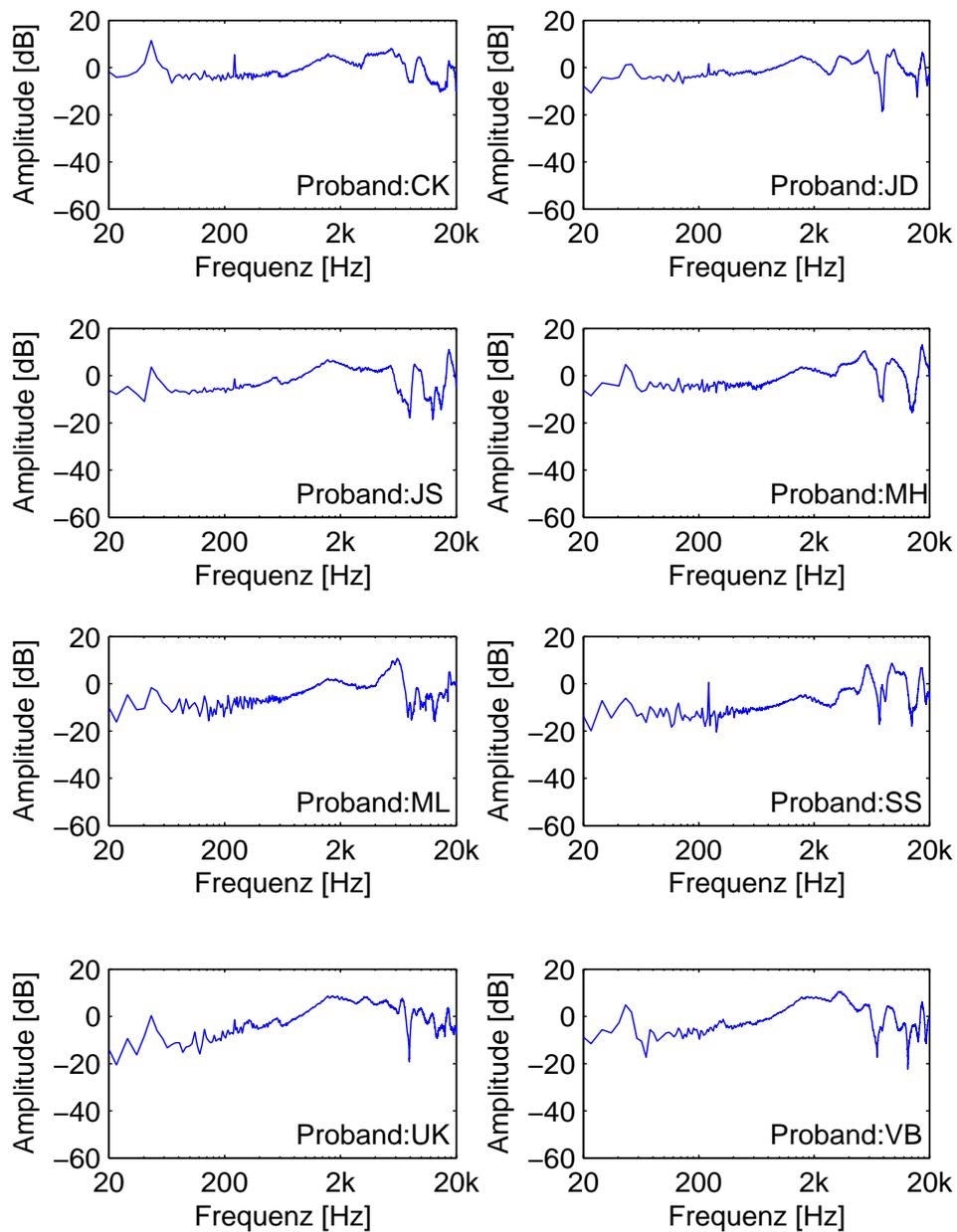


Abbildung 3.15: PTFs Kopfhörer KH1 (linke Seite) für acht Probanden, gemittelt über fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

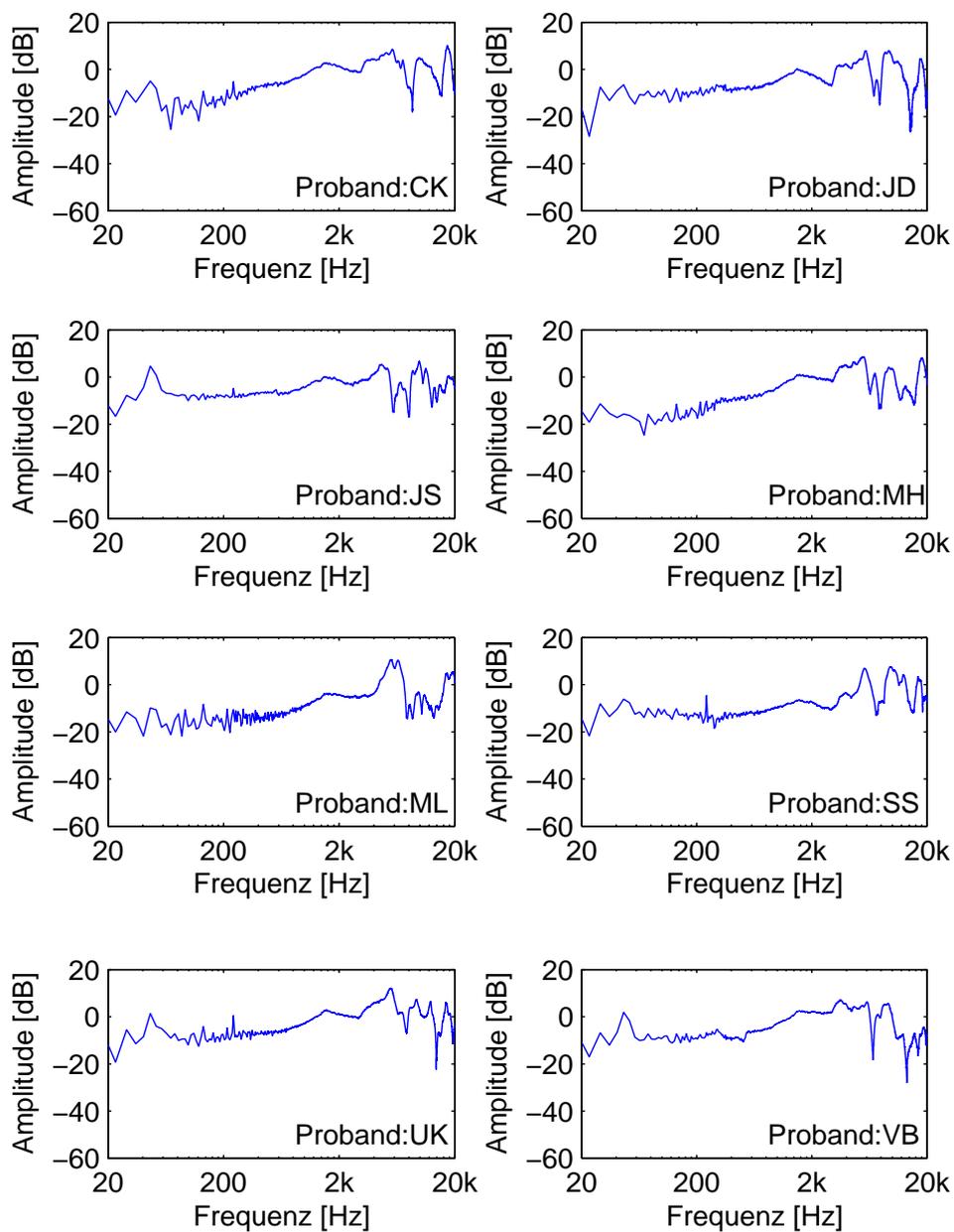


Abbildung 3.16: PTFs Kopfhörer KH1 (rechte Seite) für acht Probanden, gemittelt über fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

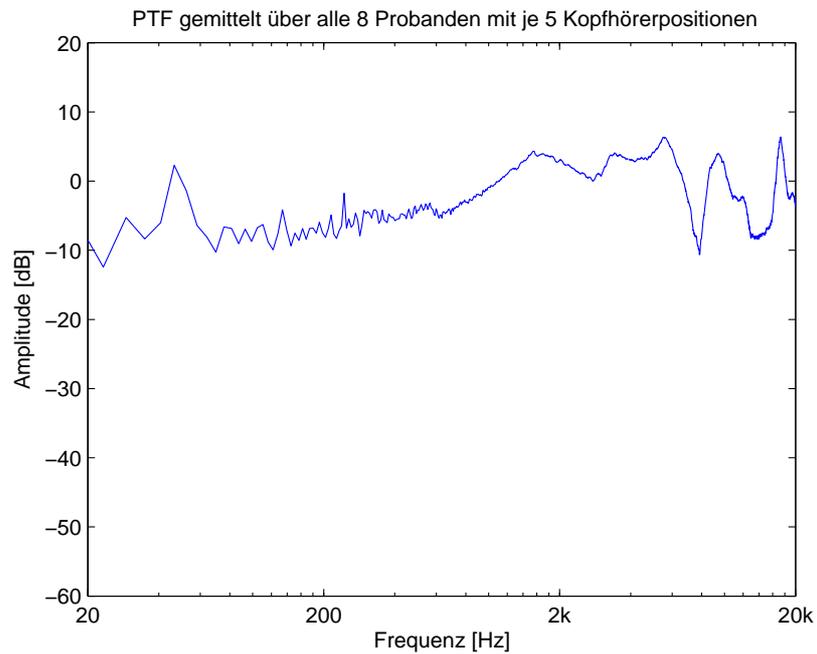


Abbildung 3.17: PTF Kopfhörer KH1 (linke Seite), gemittelt über acht Probanden und jeweils fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

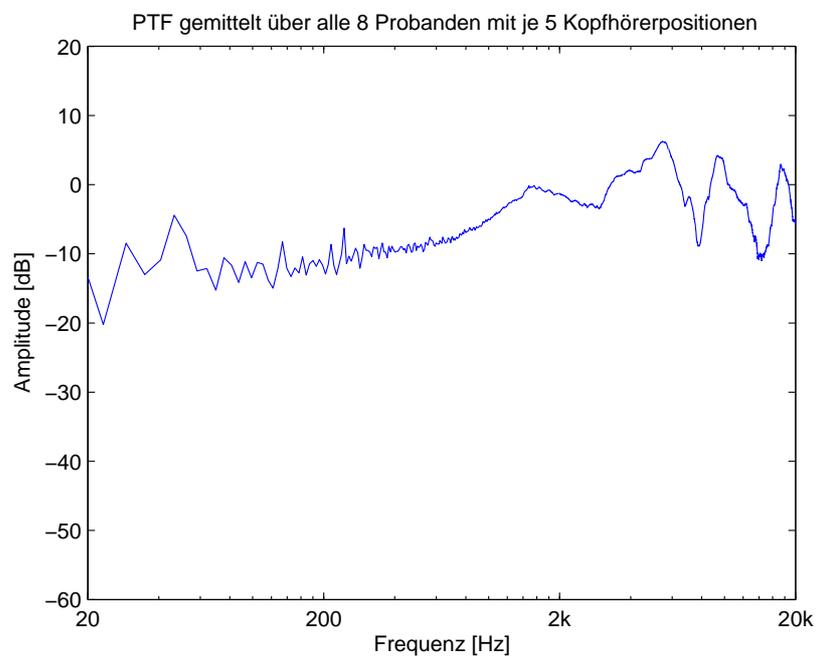


Abbildung 3.18: PTF Kopfhörer KH1 (rechte Seite), gemittelt über acht Probanden und jeweils fünf Messungen mit unterschiedlichen Kopfhörerpositionen.

3.5.1 Frequenzgangmessung des Miniaturmikrofons durch Vergleich mit Messmikrofon im Freifeld

Für den oben beschriebenen Versuch wurden zwei Miniaturmikrofone unbekanntes Herstellers aus Lagerbeständen verwendet. Die Miniaturmikrofone (MINMIC) hatten einen Durchmesser von etwa 6 mm. Um den Frequenzgang zu bestimmen, wurde jedes Mikrofon im schalltoten Raum dicht neben ein Messmikrofon (B&K, 4234) in etwa 2 m Abstand vor einen Lautsprecher montiert. Eine Skizze des Messaufbaus ist in Abbildung 3.19 zu sehen. Die Versorgungsspannung von 2,4 V wurde mit zwei Akkus erzeugt. Durch die Beschaltung ergibt sich ein Hochpass zwischen Koppelkondensator ($1 \mu\text{F}$) und Eingangswiderstand des MLSSA-Systems. Der Eingangswiderstand wurde mit einem Multimeter zu $8,4 \text{ k}\Omega$ bestimmt. Die Grenzfrequenz beträgt $\frac{1}{2 * \pi * R * C} \approx 19 \text{ Hz}$. Zur Abschätzung des Tiefpasses zwischen Widerstand ($2,2 \Omega$) und der Kapazität des Mikrofonkabels, wurde eine Kabelkapazität von 100 pF/m angenommen. Bei einer Kabellänge von 3 m ergibt sich die Grenzfrequenz zu $\frac{1}{2 * \pi * R * C} \approx 80 \text{ kHz}$. Die resultierende Filterung ist für den zu messenden Bereich vernachlässigbar.

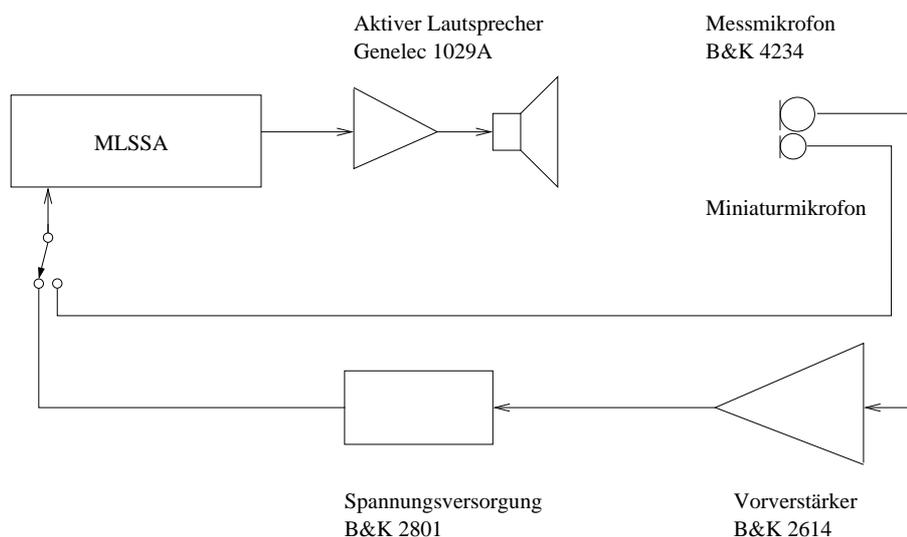


Abbildung 3.19: Messaufbau mit MLSSA-System um die Übertragungsfunktion zwischen Miniatur- und Messmikrofon zu bestimmen.

In Abbildung 3.20 sind die gemessenen Frequenzverläufe dargestellt. Für die Miniaturmikrofone ist ein starker Abfall zu tiefen Frequenzen hin sichtbar.

Berechnet man die Übertragungsfunktion zwischen Mess- und Miniaturmikrofon erhält man Abbildung 3.21. Der Autor geht von einem Messmikrofon mit ideal flachem Frequenzgang aus. Daraus folgt, dass alle Messungen mit dem in Abbildung 3.21 dargestellten Frequenzgang zu korrigieren sind. Dies ist in allen bisher dargestellten Messungen bereits geschehen. Tiefe Frequenzen wurden also mit einem geringeren Signalpegel aufgezeichnet, was das in Abschnitt 3.5 beobachtete starke Rauschen begründet. In der weiteren Signalverarbeitung (Abschnitt 3.8) soll das berücksichtigt werden, indem nur der Verlauf oberhalb von 500 Hz verwendet wird.

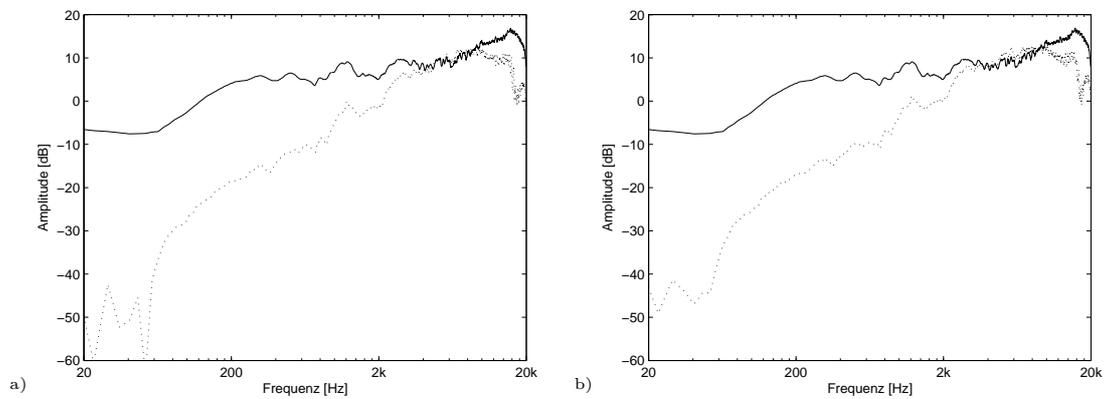


Abbildung 3.20: Frequenzgänge des Genelec Lautsprechers gemessen im Freifeld mit Messmikrofon (durchgezogene Linie) und a) Miniaturmikrofon 1 (gestrichelt), b) Miniaturmikrofon 2 (gestrichelt).

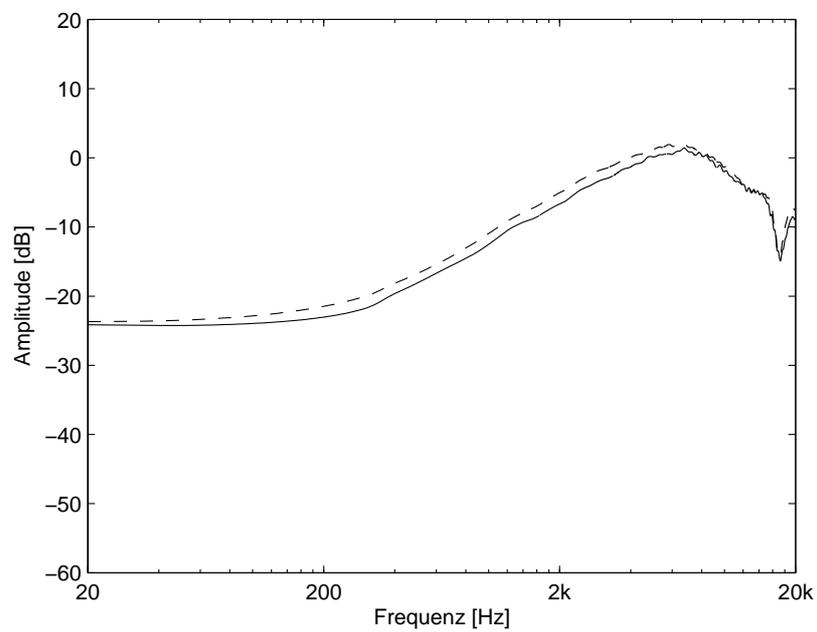


Abbildung 3.21: Übertragungsfunktion vom Messmikrofon zum Miniaturmikrofon 1 (durchgezogene Linie) und zum Miniaturmikrofon 2 (gestrichelte Linie).

3.6 Frequenzgangmessung des Kopfhörers mit künstlichem Ohr

Wieder kam das MLSSA-System zum Einsatz. An den B&K Ohrsimulator 4153 mit dem Kondensatormikrofon 4134 wurde der Schallpegelmesser B&K 2231 angeschlossen [BK4153, 2005]. Eine Skizze des Messaufbaus ist in Abbildung 3.22 zu sehen. Auf das künstliche Ohr wurde der Adapter DB0843 für zirkumaurale Hörer und die Ohrkappe YJ0305 montiert. Anschließend positionierte der Autor den Kopfhörer, wie in Abbildung 3.23 dargestellt.

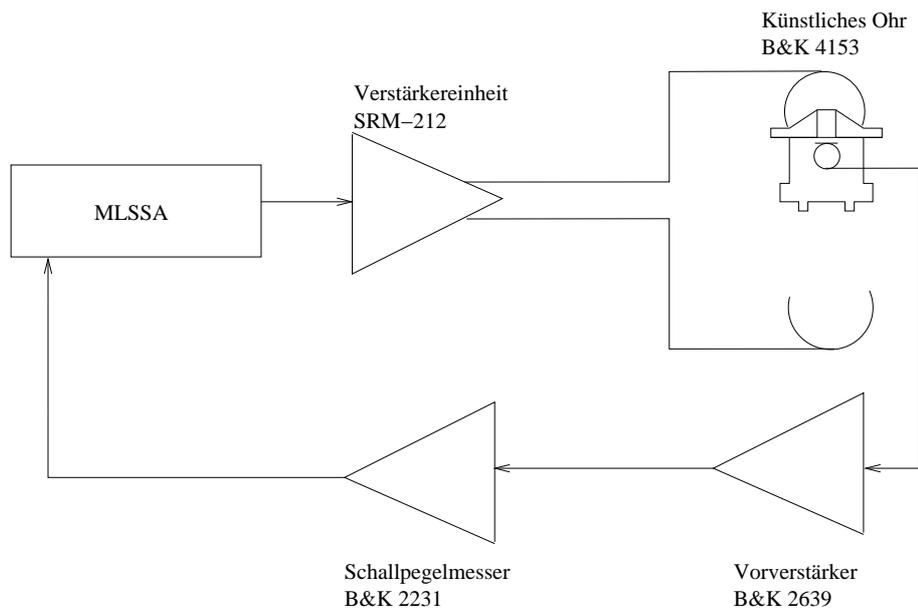


Abbildung 3.22: Messaufbau mit MLSSA-System um die Impulsantwort des Kopfhörers auf einem künstlichen Ohr zu bestimmen.



Abbildung 3.23: Kopfhörer KH1 auf Ohrsimulator B&K 4153 mit Schallpegelmesser B&K 2231.

Mit einem MLS-Signal 15. Ordnung wurde die Impulsantwort des Systems für linke und rechte Seite, sowie je zwei Kopfhörerpositionen gemessen. Während der Messung wurde das Signal durch den Schallpegelmesser A-bewertet. Da keine Frequenzbewertung er-

wünscht ist, hätte der Versuch wiederholt werden müssen. Da das MLSSA-System nicht mehr zur Verfügung stand, wurde die A-Bewertung in MATLAB kompensiert. Die resultierenden Frequenzverläufe sind für Kopfhörer KH1 in Abbildung 3.24 und für Kopfhörer KH2 im Anhang B, Abbildung B.7 dargestellt. Der Frequenzverlauf ändert sich nur geringfügig in Abhängigkeit von der der Kopfhörerposition. Daher soll im Folgenden nur eine Messung ohne Mittelung über mehrere Positionen verwendet werden.

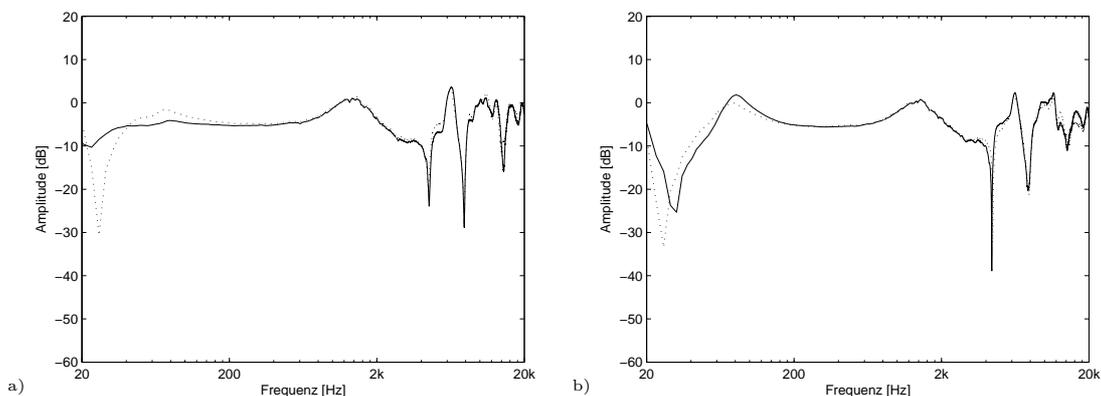


Abbildung 3.24: PTFs gemessen mit dem künstlichen Ohr B&K 4153, wobei jeweils zwei Messungen mit Neupositionierung zu sehen sind (durchgezogen, gestrichelt) für Kopfhörer KH1 a) linke Seite, b) rechte Seite.

3.7 Vergleich der drei unterschiedlichen Messverfahren

Zum besseren Vergleich der drei Messungen sind die Frequenzverläufe in einem Diagramm darzustellen und so zu verschieben dass sie übereinander liegen. Die resultierenden Diagramme sind für Kopfhörer KH1 in Abbildung 3.25 und 3.26 zu sehen. Es ist zu beachten das unterschiedliche Transferfunktionen gemessen wurden, da mit dem B&K-Ohr nicht wie bei den anderen zwei Methoden der Schalldruck am geblockten Gehörkanaleingang, sondern am Trommelfell aufgenommen wurde.

Die Frequenzverläufe zeigen nicht viele Gemeinsamkeiten. Die mit dem Kunstkopf gemessene Resonanzüberhöhung bei etwa 100 Hz findet sich in den anderen Kurven nicht wieder. Mit dem B&K-Ohr wurde für die rechte Seite eine schwächere Überhöhung bei etwa 80 Hz festgestellt. Allen drei Messungen gemeinsam ist die unterschiedlich starke Überhöhung um 2 kHz. Oberhalb variieren die Kurven stark. Die Messung mit dem B&K-Ohr zeigt drei starke Einbrüche bei etwa 4, 8 und 14 kHz. Die Messung auf dem Kunstkopf zeigt nur den Einbruch bei 14 kHz, dafür um 20 – 30 dB stärker. Die Messung mit dem Miniaturmikrofon bestätigt schwächer ausgeprägt die Einbrüche bei 8 und 14 kHz. Es zeigt sich, dass der Kunstkopf nicht für die Kopfhörerkompensation geeignet ist, da er die tatsächlichen Verhältnisse auf einem realen Kopf nicht widerspiegelt. Die Messungen mit dem B&K-Ohrsimulator zeigt geringere Abweichungen von der Messung mit Miniaturmikrofonen. Sie ist von charakteristischen scharfen Einbrüchen geprägt. In der Miniaturmikrofonmessung hat sich gezeigt das solche Einbrüche auftreten, aber sehr individuell sind. Durch die Mittelung über mehrere Probanden verschwinden diese. Wenn sie berücksichtigt werden soll, wären individuelle Kompensationsfilter unabdingbar.

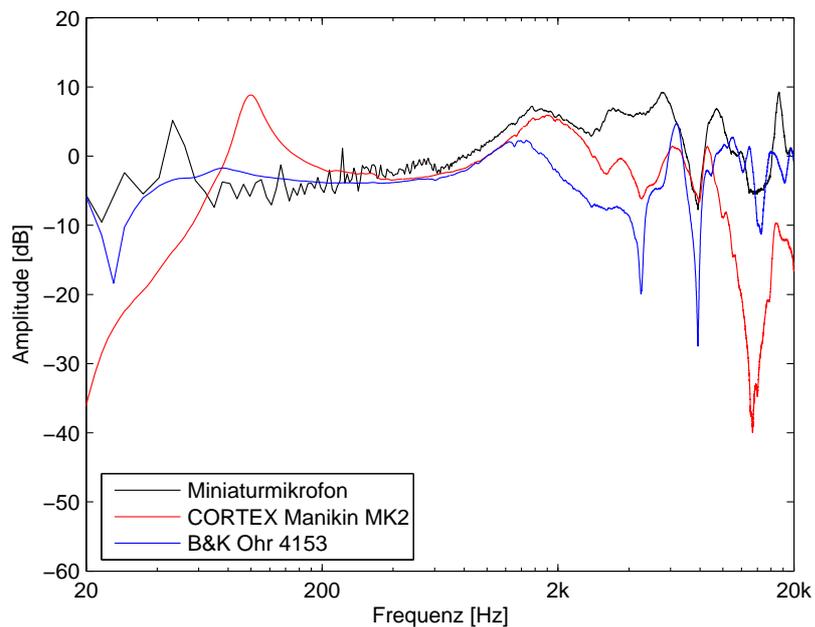


Abbildung 3.25: Vergleich der über sämtliche Messungen gemittelten PTFs (Kopfhörer KH1, links), für die drei unterschiedlichen Messmethoden.

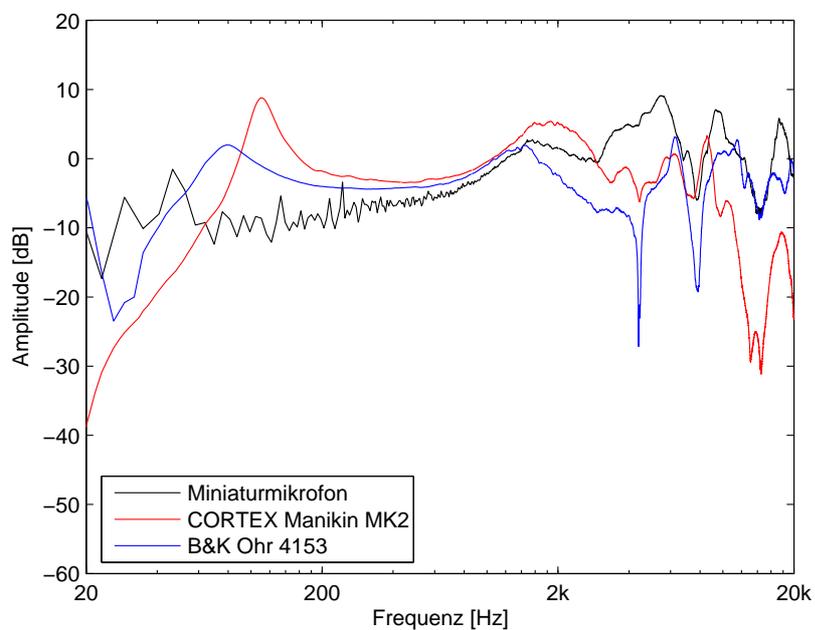


Abbildung 3.26: Vergleich der über sämtliche Messungen gemittelten PTFs (Kopfhörer KH1, rechts), für die drei unterschiedlichen Messmethoden.

Für den Entwurf des Kompensationsfilters soll (für Frequenzen größer 500 Hz) die über acht Probanden und je fünf Kopfhörerpositionen gemittelte Messung mit Miniaturmikrofonen verwendet werden. Es ist jeweils ein Filter für linke und rechte Seite zu entwerfen.

3.8 Entwurf inverser Filter

Im vorangehenden Abschnitt wurde geschlussfolgert, dass ein inverser Filter zur Kompensation für die Kopfhörercharakteristik nötig ist. Zuerst ist die Zielfunktion des Filters zu bearbeiten. Danach sollen zwei MATLAB-Methoden zum Filterentwurf von einem praktischen Standpunkt aus diskutiert und verglichen werden.

Ein Digitalfilter kann nach [Oppenheim and Schaffer, 1999] allgemein durch folgende Differenzgleichung dargestellt werden:

$$\begin{aligned} a_1 y[n] &= b_1 x[n] + b_2 x[n-1] + \dots + b_{n_b+1} x[n-n_b] \\ &\quad - a_2 y[n-1] - \dots - a_{n_a+1} y[n-n_a] \end{aligned} \quad (3.12)$$

wobei x das Eingangs-, y das Ausgangssignal und \mathbf{a} und \mathbf{b} Polynome mit den Koeffizienten $a_1 \dots a_{n_a}$ bzw. $b_1 \dots b_{n_b}$ darstellen. Wenn außer $a_1 = 1$ keine a -Koeffizienten existieren, spricht man von einem FIR (Finite Impulse Response) Filter. Wenn durch die Existenz von weiteren a -Koeffizienten eine Rückkopplung entsteht erhält man einen IIR (Infinite Impulse Response) Filter. Die Wurzeln der b -Koeffizienten bezeichnet man als Nullstellen und die der a -Koeffizienten als Pole des Filters. Jeder Filter ist in einen Allpass- und eine Minimalphasen-Anteil zerlegbar. Der Minimalphasen-Anteil hat den selben Amplitudengang wie der Originalfilter. Ist der Phasengang nicht wichtig, so kann der Originalfilter durch seinen Minimalphasen-Anteil ersetzt werden. Soll ein Filter stabil sein, so müssen alle a -Koeffizienten innerhalb des Einheitskreises der z -Ebene ($|z| = 1$) liegen. Dies ist für einen minimalphasigen Filter immer der Fall. In dieser Arbeit sollen Filterentwurfalgorithmen verwendet werden, die versuchen einen vorgegebenen Amplitudengang zu approximieren und minimalphasige Filter zu entwerfen.

Invertiert man den Amplitudengang aus Abbildung 3.17 und 3.18 erhält man Abbildung 3.27a) und 3.28a). Dieser Verlauf soll, wie in Abschnitt 3.5.1 erläutert, nur oberhalb von 500 Hz als Vorgabe dienen. Unterhalb wird er linear zu 0 dB bei 0 Hz ausgeblendet. Dies verhindert auch eine mögliche Beschädigung des Kopfhörers durch zu starke Verstärkung bei sehr niedrigen Frequenzen. Oberhalb von 18 kHz soll der Frequenzgang ebenfalls linear zu 0 dB bei der halben Abtastfrequenz ausgeblendet werden um einer zu großen Verstärkung bei hohen Frequenzen vorzubeugen. Da der generelle Amplitudengang von Interesse ist und scharfe Veränderung im Frequenzverlauf den Filterentwurf erschweren, soll ein moving-average-Filter mit einer Fensterbreite von fünf Abtastwerten den Kurvenverlauf glätten. Zusätzlich ist der Amplitudengang so zu verschieben, dass seine Gesamtverstärkung etwa 0 dB beträgt. Der Autor wählt eine etwas niedrigere Verstärkung um Begrenzungseffekte (clipping) beim Schreiben der Wav-Dateien zu vermeiden. Die bearbeiteten Frequenzverläufe für Kopfhörer KH1 sind in Abbildung 3.27b) und 3.28b) zu sehen.

Die MATLAB-Funktion *yulewalk* verwendet die modifizierten Yule-Walker-Gleichungen und einen iterativen Algorithmus im Zeitbereich um den Filter zu optimieren. Als zweites soll die MATLAB-Methode *invfreqz* betrachtet werden. Sie approximiert den Amplitudengang durch Lösung eines linearen Gleichungssystems im Frequenzbereich und minimiert

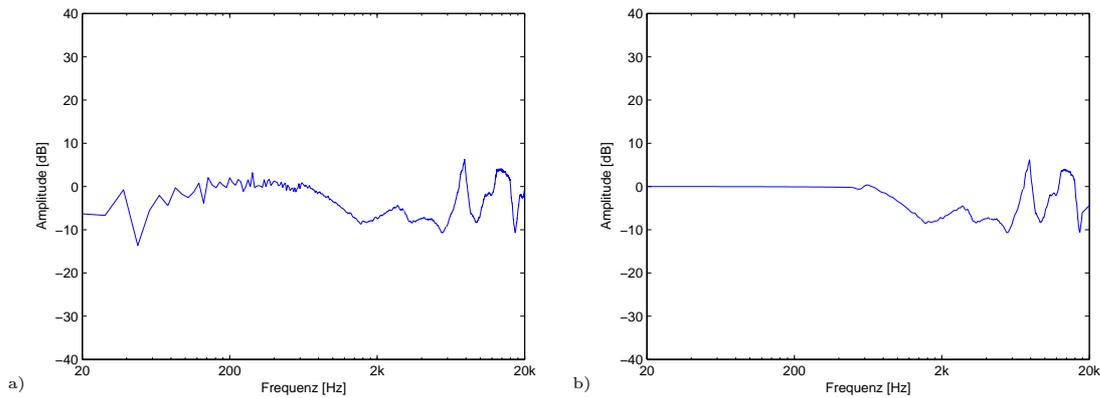


Abbildung 3.27: Vorgabe für den Entwurf des digitalen Filters für den Kopfhörer **KH1** (linke Seite) **a)** invertierter Amplitudengang unbearbeitet, **b)** invertierter Amplitudengang linear ausgeglichen unterhalb von 500 Hz zu 0 dB bei 0 Hz und oberhalb von 18 kHz auf 0 dB bei 24 kHz. Zusätzlich wurde der Kurvenverlauf geglättet.

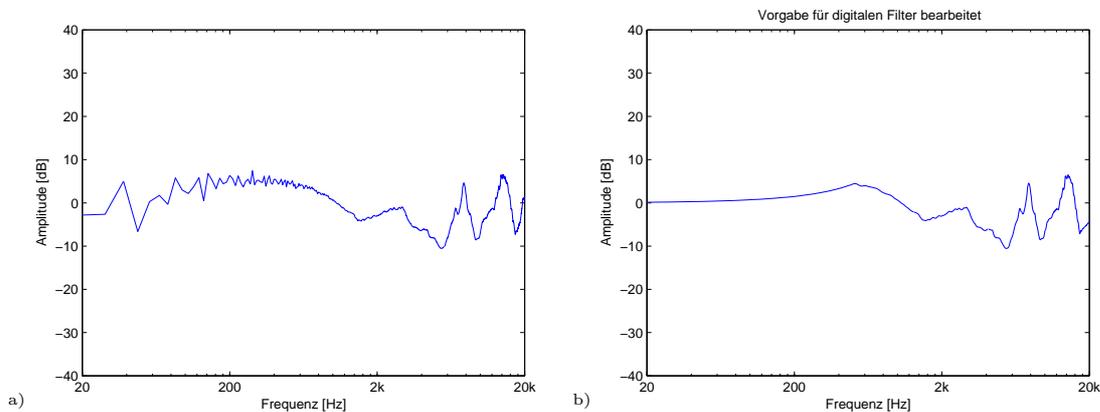


Abbildung 3.28: Vorgabe für den Entwurf des digitalen Filters für den Kopfhörer **KH1** (rechte Seite) **a)** invertierter Amplitudengang unbearbeitet, **b)** invertierter Amplitudengang linear ausgeglichen unterhalb von 500 Hz zu 0 dB bei 0 Hz und oberhalb von 18 kHz auf 0 dB bei 24 kHz. Zusätzlich wurde der Kurvenverlauf geglättet.

die quadratische Abweichung durch einen iterativen Gauss-Newton-Algorithmus. Es sollen mit beiden Methoden für unterschiedliche Ordnungen Filter entworfen und beispielhaft für die linke Kopfhörerseite des Kopfhörers KH1 grafisch dargestellt werden. Zum numerischen Vergleich soll die jeweilige mittlere quadratische Abweichung nach folgender Gleichung berechnet werden:

$$Abweichung = \frac{1}{N} \sum_{n=1}^N \left(20 \log \frac{T(n)}{F(n)} \right)^2 \text{ dB} \quad (3.13)$$

Der *yulewalk*-Algorithmus benötigt drei Eingaben, die Filterordnung und zwei Vektoren die den gewünschten Amplitudengang und die entsprechenden Frequenzen enthalten. In Abbildung 3.29 ist das Ergebnis des Filterentwurfs mit 10ter und 60ter Ordnung im Vergleich zur Zielfunktion dargestellt. Mit steigender Filterordnung nähert sich die Filterkurve zunehmend der Zielfunktion und es verkleinert sich die mittlere quadratische

Abweichung. Für höhere Filterordnungen hat die Methode bei der gegebenen Zielfunktion numerische Probleme.

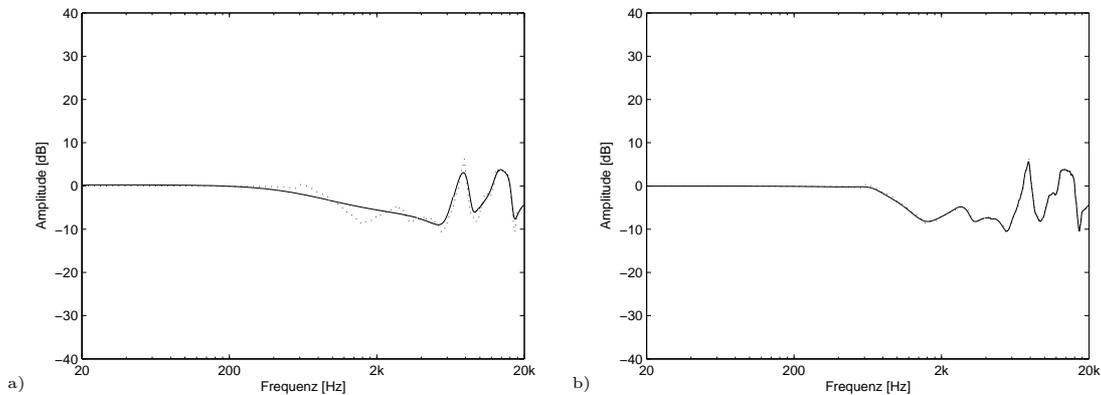


Abbildung 3.29: Verlauf der Vorgabe für den Filterentwurf (gestrichelt) und dem resultierenden FIR-Filter (durchgezogen) unter Verwendung des MATLAB *yulewalk*-Algorithmusses (**Kopfhörer KH1, linke Seite**) a) mit **Filterordnung 10** (mittlere quadratische Abweichung = 1,507 dB), b) mit **Filterordnung 60** (mittlere quadratische Abweichung = 0,028 dB).

Zum Vergleich werden Filter 10ter, 20ter, 30ter und 60ter Ordnung mit der *invfreqz*-Methode entworfen und in Abbildung 3.30 und 3.31 dargestellt. Bei dieser Methode sind mehrere Parameter wählbar. Zum Beispiel kann eine Frequenzwichtung vorgenommen werden, was hier aber zu keiner wesentlichen Verbesserung der Ergebnisse führt. Es zeigt sich das bei gleicher Filterordnung eine geringere mittlere quadratische Abweichung auftritt (zum Beispiel bei 10ter Ordnung: 0,908 dB < 1,507 dB). Erhöht man die Filterordnung weiter treten auch hier numerische Probleme auf.

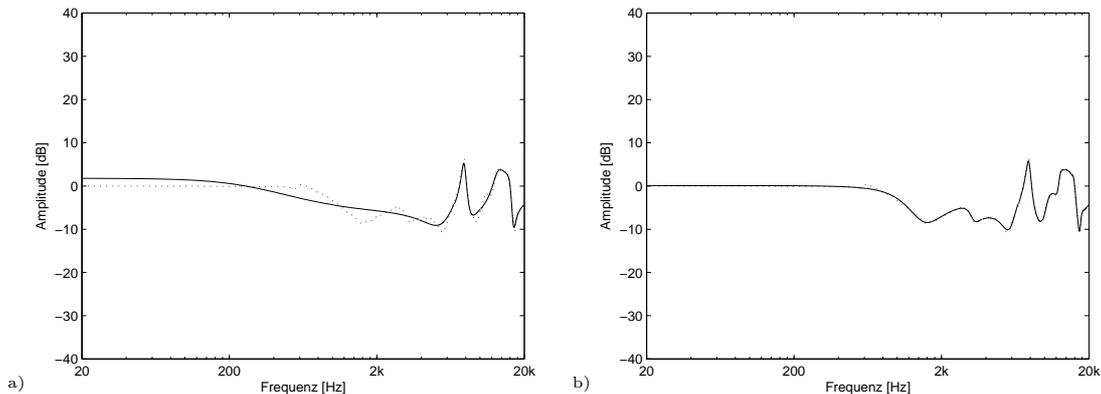


Abbildung 3.30: Verlauf der Vorgabe für den Filterentwurf (gestrichelt) und dem resultierenden FIR-Filter (durchgezogen) unter Verwendung des MATLAB *invfreqz*-Algorithmusses (**Kopfhörer KH1, linke Seite**) a) mit **Filterordnung 10** (mittlere quadratische Abweichung = 0,908 dB), b) mit **Filterordnung 20** (mittlere quadratische Abweichung = 0,050 dB).

Allerdings kann man für die *invfreqz*-Methode die Ordnung für a- und b-Koeffizienten getrennt bestimmen. Erhöht man die Anzahl der b-Koeffizienten bis zur 90ten Ordnung verkleinert sich die mittlere quadratische Abweichung weiter (Abbildung 3.32) bis auf 0,005 dB. Die Verbesserung ist nicht mehr bedeutend, es soll trotzdem der zuletzt dar-

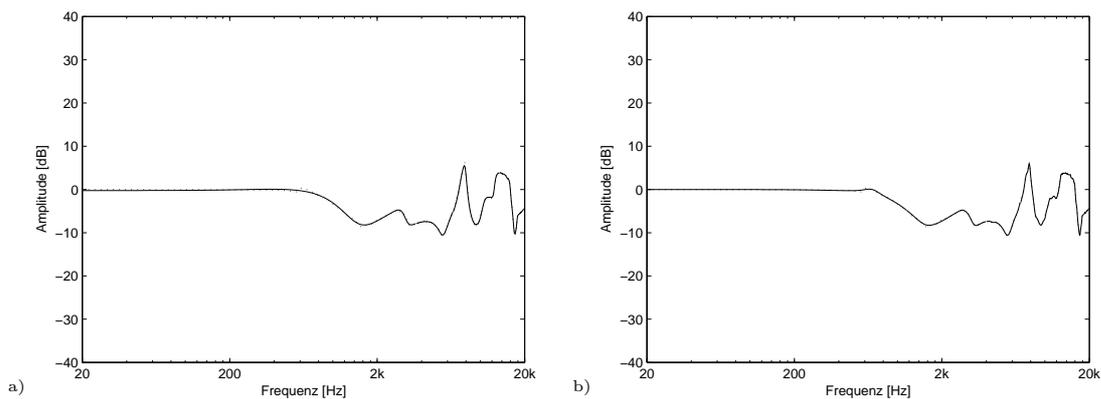


Abbildung 3.31: Verlauf der Vorgabe für den Filterentwurf (gestrichelt) und dem resultierenden FIR-Filter (durchgezogen) unter Verwendung des MATLAB *invfreqz*-Algorithmusses (**Kopfhörer KH1, linke Seite**) a) mit **Filterordnung 30** (mittlere quadratische Abweichung = 0,030 dB), b) mit **Filterordnung 60** (mittlere quadratische Abweichung = 0,008 dB).

gestellte Filter im Folgenden verwendet werden. Der Filter wird auf Minimalphasigkeit überprüft und es wird in Abbildung 3.33 ein Pol-Nullstellen-Plan dargestellt. Die Pole und Nullstellen liegen teils sehr eng am, aber alle innerhalb des Einheitskreises. Es wird ein zweiter Filter selbiger Ordnung für die rechte Seite des Kopfhörers entworfen.

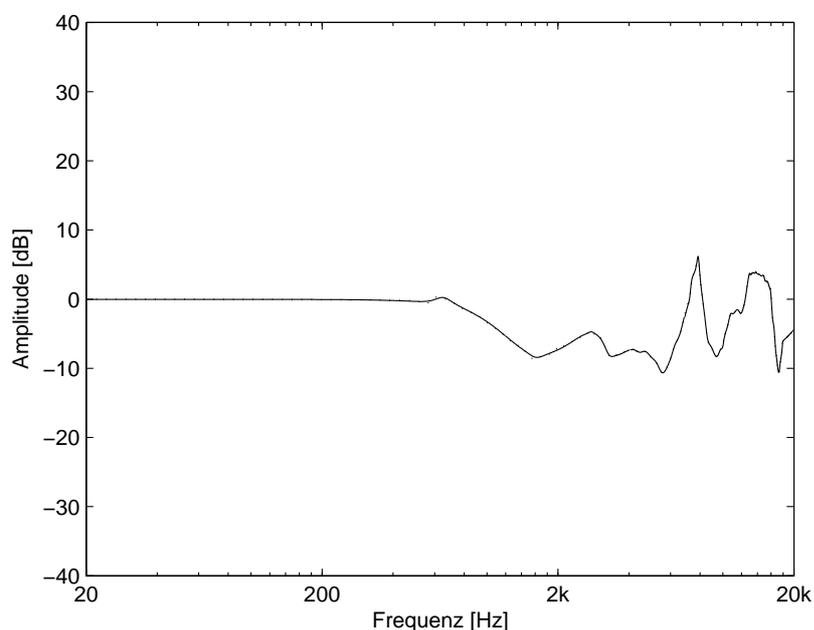


Abbildung 3.32: Verlauf der Vorgabe für den Filterentwurf (gestrichelt) und dem resultierenden FIR-Filter (durchgezogen) unter Verwendung des MATLAB *invfreqz*-Algorithmusses und **60 a** sowie **90 b Koeffizienten** (**Kopfhörer KH1, linke Seite**, mittlere quadratische Abweichung = 0,005 dB).

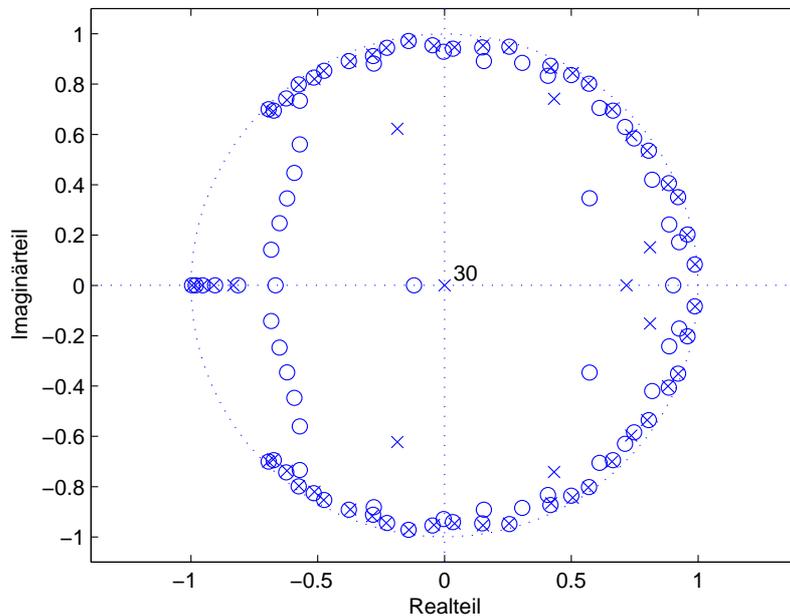


Abbildung 3.33: Pol-Nullstellen-Plan des FIR-Filters unter Verwendung der Matlab "invfreq"-Algorithmus und 60 a sowie 90 b Koeffizienten (**Kopfhörer KH1, linke Seite**).

Zur Veranschaulichung soll in **Abbildung 3.34** die original gemessene gemittelte Übertragungsfunktion des Kopfhörer KH1 für linke und rechte Seite im Vergleich zur Übertragungsfunktion nach Kompensation mit Digitalfilter (*yulewalk*-Algorithmus, 60 a- und 90 b-Koeffizienten) dargestellt werden.

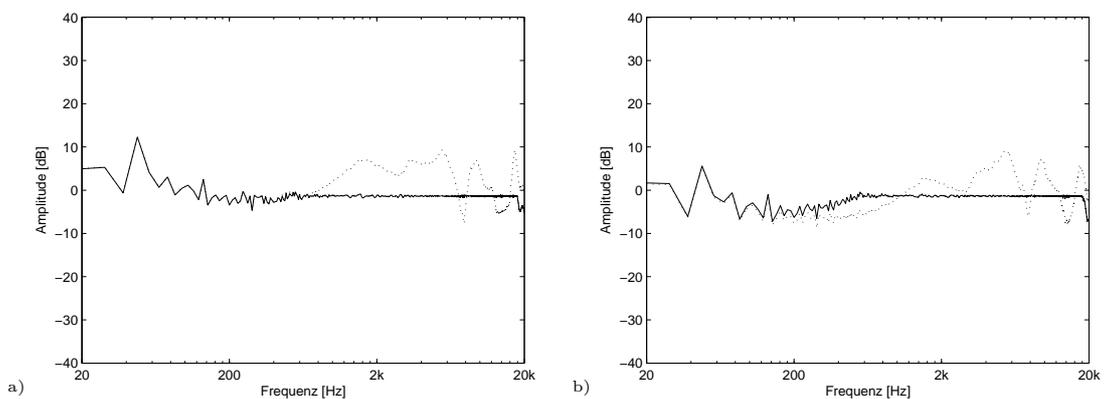


Abbildung 3.34: Der originale Frequenzgang des **Kopfhörer KH1** (gestrichelt) im Vergleich mit dem Frequenzgang nach der Kompensation durch einen digitalen Filter (*yulewalk*, 60 a- und 90 b-Koeffizienten, durchgezogen) a) linke Seite, b) rechte Seite.

Entwurf Hörversuche

Psychoakustische Forschung kann Labor- oder Feldcharakter besitzen, je nachdem ob der Forscher die Randbedingungen des Versuchs stark oder schwach beeinflusst. Beide Extreme haben Vor- und Nachteile. Im Labor können störende Einflüsse kontrolliert werden. Bei Wiederholung über mehrere Testpersonen oder Sitzungen, können die Rahmenbedingungen weitestgehend konstant gehalten werden. Die Kontrolle von untersuchungsbedingten Störvariablen führt nach [Bortz, 2005] zu einer eindeutigeren Interpretierbarkeit und hohen **internen Validität** der Ergebnisse: "Die interne Validität sinkt mit wachsender Anzahl plausibler Alternativerklärungen für das Ergebnis auf Grund nicht kontrollierter Störvariablen."

Die Durchführung der Untersuchung im Labor hat auch Nachteile. So ist der künstliche Versuchsaufbau im Labor nur bedingt auf natürliche Alltagssituationen übertragbar. Dies bezeichnet man als geringe **externe Validität**. Ein extern valides Ergebnis ist angestrebt, aber ohne interne Validität nicht möglich. In [Bortz, 2005] wird externe Validität wie folgt definiert: "Eine Untersuchung ist extern valide, wenn ihr Ergebnis über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinausgehend generalisierbar ist. Die externe Validität sinkt mit wachsender Unnatürlichkeit der Untersuchungsbedingungen bzw. mit abnehmender Repräsentativität der untersuchten Stichproben." Es müssen also zwei Dinge angestrebt werden:

- Die Bedingungen im Labor sollten eine natürliche Situation so gut wie möglich reproduzieren.
- Die untersuchten Stichproben sollten repräsentativ sein. Die Wahl der Stichproben ist in 4.5 beschrieben.

Die Beurteilung der Qualität eines Instrumentes wird vom Autor als sehr störvariablenanfällig eingestuft. Es ist also notwendig die Rahmenbedingungen zu kontrollieren, indem sie konstant gehalten oder als unabhängige Variablen in das Versuchsdesign eingebunden werden. Die Wahl der unabhängigen Variablen ist in Abschnitt 4.1 beschrieben.

Zudem sind praktische Aspekte zu berücksichtigen. So ist es nicht möglich für jeden Hörversuch Musiker zu engagieren. Anstelle dessen sollen, wie in Abschnitt 2.2 erläutert, einmalig Aufnahmen mit Hilfe eines Kunstkopfes angefertigt werden. Diese können

dann jedem Probanden per Kopfhörer wiedergegeben werden. Es wurde erläutert, dass unter bestimmten Voraussetzungen eine natürliche Reproduktion des Klanges möglich ist. Die Voraussetzungen dafür wurden in Kapitel 3 geschaffen, indem Filter entwickelt wurden, mit denen die Kopfhörercharakteristik kompensiert werden kann. In Abschnitt 4.1.3 ist die Aufnahme der Klangbeispiele dokumentiert. Danach können die Hörversuche zur Qualitätsbeurteilung der Instrumente entworfen werden. Abschnitt 4.2 beschreibt die Realisierung des sequentiellen Hörversuchs 1 mit den notwendigen Beschränkungen auf Grund der großen Anzahl unabhängiger Variablen. Abschnitt 4.3 beschäftigt sich mit den Vor- und Nachteilen eines Versuchsentwurfs in Blöcken. Es ist ein zweiter Hörversuch mit dieser Methode zu entwerfen. In einem dritten Hörversuch (Abschnitt 4.4) sollen zehn Instrumente bei konstanten Umgebungsbedingungen miteinander verglichen werden.

4.1 Wahl der unabhängigen Variablen

Einige wichtige Einflussgrößen auf die Qualitätsbeurteilung wurden bereits in Abschnitt 2.1.1 erläutert. Die dort genannten fünf Faktoren sollen in den Entwurf eingebunden werden. Dazu ist für jeden Faktor die Anzahl der Faktorstufen zu wählen:

- 10 Gitarren,
- 2 Räume,
- 3 Sequenzen,
- 5 Spieler,
- 2 Anspiele.

Die spezielle Auswahl der Gitarren ist in Abschnitt 4.1.1 erläutert. Die Beschreibung der gewählten Sequenzen folgt in Abschnitt 4.1.2. Es sollen zwei relativ extreme Räume (halliger Konferenzraum und reflexionsarmer Raum) im Institut für Musikinstrumentenbau (IfM) in Zwota für die Aufnahmen verwendet werden. Sie sind näher in Abschnitt 4.1.3 beschrieben. Fünf professionelle Gitarristen (Spieler) wurden für die Aufnahmen engagiert. Jede Aufnahme wurde wiederholt, um zwei Anspiele für jede Kombination aus Gitarre, Raum, Sequenz und Spieler zu erhalten.

4.1.1 Auswahl der Gitarren

Es wurden zehn klassische Gitarren gewählt, die ein breites Spektrum, sowohl von der Bauart, als auch vom Klang abdecken sollen. Die Instrumente entstammen aus dem Referenzbestand des IfM und wurden am 08.06.2005 einheitlich mit D'Addario Classic Guitar J45 Normal Tension besaitet. Eine Beschreibung der Instrumente durch Herrn Ziegenhals vom IfM ist in Tabelle 4.1 zu sehen.

Die Gitarren G1, G2, G4, G5, G22, G23, sind nach Angaben von Herrn Ziegenhals von der Bauart her sehr verschieden, die Instrumente G24, G25, G26 und G27 eher ähnlich, heben sich aber andererseits als Gruppe von den ersten sechs ab.

Tabelle 4.1: Übersicht der für die Aufnahmen gewählten Gitarren aus dem Referenzbestand des IfM Zwota.

Instrument	Bezeichnung	Charakterisierung
Referenz 1	G1	Japanisches Industrieinstrument, welches in früheren Tests stets sehr gut beurteilt wurde. Takamine Modell C-128 Baujahr: 10/1979
Referenz 2	G2	Vogtländisches Meisterinstrument mit im Verhältnis größerem Korpus. Baujahr: 9/1977; Hersteller: Armin Gropp
Referenz 4	G4	Japanisches Industrieinstrument, welches in früheren Untersuchungen stets mittelmäßig beurteilt wurde. Marlin Modell MC 315
Referenz 5	G5	Finnische Industriegitarre mit im Verhältnis größerem Korpus, und auffallend dunkler Klangfärbung. Die Urteile früherer Tests waren sehr differenziert. Landola SL 3 Nr.: 151472
Referenz 22	G22	Session C 425 Sehr billiges Instrument Laminierte Decke, wahrscheinlich Indonesien
Referenz 23	G23	Prototyp Doppelbodengitarre Kreul Entwickelt von Zupfinstrumentenmachermeister Eberhard Kreul mit Unterstützung durch IfM ca. 1975
Referenz 24	G24	Versuchsmodell IfM (Resokrit II) Decken-Nr.: 248 entspricht Höfner Modell HF 12 Baujahr: 2002; Hersteller: Höfner
Referenz 25	G25	Versuchsmodell IfM (Resokrit II) Decken-Nr.: 616 entspricht Höfner Modell HF 12 Baujahr: 2002; Hersteller: Höfner
Referenz 26	G26	Versuchsmodell IfM (Resokrit II) Decken-Nr.: 228 entspricht Höfner Modell HGL 50 SE Baujahr: 2001; Hersteller: Höfner Mustergitarre Instrumentenpass
Referenz 27	G27	Versuchsmodell IfM Projekt „Modifiziertes Holz“ Instr.-Nr.: F03304 entspricht Höfner Modell HGL 50 Baujahr: 2005; Hersteller: Höfner

4.1.2 Auswahl der Sequenzen

Die verwendeten Musikstücke müssen einige Kriterien erfüllen, um für den Vergleich von Instrumenten günstig zu sein. Die auszuwählenden Stücke sollten das Instrument hinreichend charakterisieren. Es sollten aber auch möglichst einfache Stücke sein, da sie vielfach wiederholt und gleichmäßig von verschiedenen Gitarristen zu spielen sind. In [BS1116-1, 1997] wird vor der Verwendung von Hörbeispielen mit mitreißenden oder beängstigenden Inhalten gewarnt, die von der eigentlichen Aufgabe des Probanden ablenken können. Im vorliegenden Fall sollten die Stimuli zum Beispiel nicht zu emotional sein, damit das Instrument nicht in den Hintergrund tritt. Die Auswahl erfolgte in Zusammenarbeit mit dem Gitarristen Sveto Popov. Die gewählten Stücke sind in Tabelle 4.2 zusammengefasst. Notenbilder der Stücke sind im Anhang C zu sehen. Klangbeispiele können auf der beigelegten CD gehört werden.

Tabelle 4.2: Gewählte Sequenzen für die Aufnahmen mit zehn Gitarren.

Stück 1	Chromatische Tonleiter E ... e2 Spielweise Doppelschlag, jeder Ton Notenwert $\frac{1}{8}$ gespielt E ... ab1 in erster Lage, dann nur noch e1 - Saite Tempo 80 $\frac{1}{4}$ pro Minute
Stück 2	kleiner Blues Tempo 80 $\frac{1}{4}$ pro Minute
Stück 3	Para Gitarra Tempo 70 $\frac{1}{4}$ pro Minute

4.1.3 Aufnahme der Klangbeispiele

Für den Hörtest werden Klangbeispiele benötigt, die so unverfälscht wie möglich die Charakteristik des Instrumentes reproduzieren. Es wurde deswegen in Abschnitt 2.2 entschieden die Aufnahmen mit einem Kunstkopf durchzuführen.

Eine Skizze der Aufnahmesituation im reflexionsarmen Raum des IfM in Zwota ist in Abbildung 4.1 zu sehen. Die Aufnahmesituation im Konferenzraum ist in Abbildung 4.2 dargestellt. Der Gitarrist saß dem Kunstkopf frontal gegenüber, welcher sich ebenfalls in Höhe einer sitzenden Person befand. Das verwendete Equipment ist in Tabelle 4.3 zusammengefasst.

Begonnen wurden die Aufnahmen für jede Gitarre mit Stück 1, welches einmal wiederholt wurde. Danach wurden die Stücke 2 und 3 mit Wiederholung gespielt. Stücke mit Verspielern wurden abgebrochen und neu begonnen. Teilweise waren Musiker unzufrieden mit Anspielen und wiederholten die Stücke selbstständig. Hier wird im weiteren Anspiel 1 und 3 verwendet. Um ein annähernd konstantes Tempo zwischen Anspielen und Musikern zu gewährleisten, wurde dem Gitarristen per Kopfhörer ein Metronom eingespielt.

In Abschnitt 2.1.2 wurde erläutert das die Lautstärke ein wichtiger Einflussfaktor bei der Beurteilung von Hörbeispielen ist. Daher ist es normalerweise notwendig die wahrgenommene Lautheit aller Aufnahmen anzugleichen. Da die Gitarren aber auch in einer natürlichen Situation je nach Bauart und Spielweise unterschiedlich laut sind, sollen diese Lautheitsunterschiede hier beibehalten werden. Lediglich zwischen den beiden Aufnahmeräumen besteht ein extremer Lautheitsunterschied, der ausgeglichen werden soll,

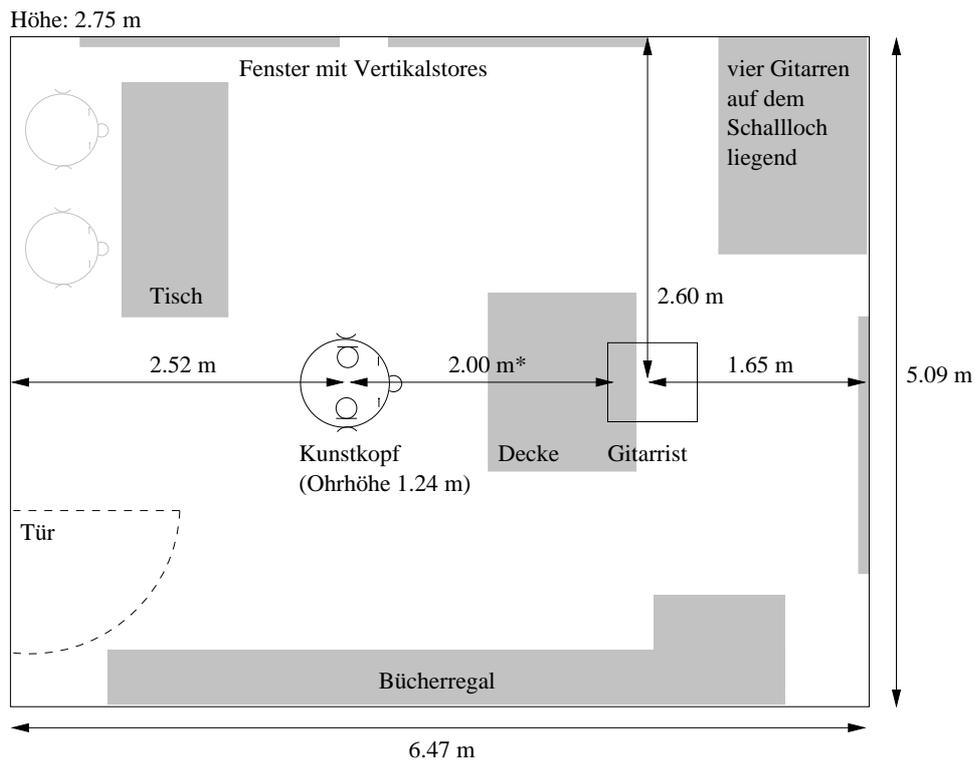


Abbildung 4.2: Aufbau für die Aufnahmen im Konferenzraum des IfM in Zwota. Der Gitarrist sitzt auf einem Stuhl frontal gegenüber dem Kunstkopf. *Der Abstand zwischen Kunstkopf und Gitarre wurde von den Ohren zum Schalloch gemessen. Die Decke wurde nur bei Bedarf verwendet um das störende Mitklöpfen mit dem Fuß zu bedämpfen.

30 s langes Hörbeispiel anhören muss, obwohl sie ihr Urteil schon nach fünf Sekunden gefällt hat, schnell demotiviert und möglicherweise unkonzentriert. In diesem Fall wird die Motivation und Konzentration der Versuchsperson als wichtiger eingeschätzt als der konstante Randparameter und die Möglichkeit zum Abbruch gewählt. Zusätzlich verkürzt sich dadurch die benötigte Gesamtzeit des Hörversuchs.

Auf Grund der begrenzten Anzahl von Versuchspersonen soll ein Entwurf mit vollständiger Messwiederholung zum Einsatz kommen. Das heißt jeder Versuchsperson werden alle Hörbeispiele zur Beurteilung angeboten. Es ist trotzdem nicht praktikabel jedem Probanden alle möglichen Kombinationen der Wiederholungsfaktorstufen zur Beurteilung vorzuspielen (2 Anspiele x 2 Räume x 10 Instrumente x 5 Musiker x 3 Musiksequenzen = 600 Aufnahmen). Ein solcher vollständiger Versuchsaufbau wäre aus statistischer Sicht am einfachsten auszuwerten. Reduziert man den Versuchsaufwand, indem bestimmte Faktorstufenkombinationen ausgelassen werden, erhält man einen hierarchischen oder teilhierarchischen Entwurf. Durch solch einen Entwurf ist es möglich den Versuchsaufwand beträchtlich zu reduzieren. Für eine ausführliche Einführung in hierarchische und teilhierarchische Versuchspläne soll hier auf [Bortz, 2005] verwiesen werden. Um einen hierarchischen Versuchsaufbau zu verifizieren müssen die unabhängigen Variablen tatsächlich voneinander unabhängig sein. Ob diese Bedingung gilt, soll in dieser Arbeit untersucht werden. Um die Abhängigkeiten zwischen den Variablen zu untersuchen, ist ein vollständiger Entwurf (Verwendung aller Kombinationen der Faktorstufen) unabdingbar. Um den Umfang des Hörversuchs trotzdem realisierbar zu machen, wird die Anzahl der

Faktorstufen auf je zwei pro unabhängige Variable reduziert (2 Anspiele x 2 Räume x 2 Instrumente x 2 Musiker x 2 Musiksequenzen = 32 Aufnahmen). Das entsprechende Datenschema ist in Tabelle 4.4 zu sehen. Unter jeder Kombination der fünf zweistufigen Wiederholungsfaktoren wird eine Stichprobe S_1 bzw. S_2 zu je n Probanden untersucht. Die Probanden werden in zwei Hörergruppen eingeteilt und es entsteht dadurch ein sechsfaktorieller Entwurf (fünf Wiederholungsfaktoren und ein Gruppierungsfaktor).

Eine klassische Hörversuchsmethode ist das sequentielle Darbieten eines Stimuli mit anschließender Beurteilung der gewünschten abhängigen Variablen. Der Proband vergleicht dabei die gehörte Gitarre mit einer "inneren Referenzgitarre" und bildet daraufhin sein Urteil. Es wurde darauf geachtet, dass die Fragestellung nicht beeinflussend formuliert ist: "Bitte bewerten sie die allgemeine Qualität des gehörten Instrumentes." In [BS1284-1, 2003] wird empfohlen eine quasikontinuierliche Skala zu verwenden. Es zeigt sich in informellen Vorexperimenten, dass die Beurteilung so relativ anspruchsvoll ist. In vorangegangenen Experimenten am IfM wurden gute Erfahrungen mit fünfstufigen Skalen gemacht. Alternativ wäre zum Beispiel eine sechsstufige Skala denkbar. Sie hätte den Vorteil, dass es keine Skalenmitte gibt. Der Proband wird dadurch zu einer Entscheidung gezwungen. Leider gibt es keine einheitliche verbale Bezeichnung für solch eine Skala zur Qualitätsbeurteilung. Der Autor entschied eine fünfstufige Skala zu verwenden, die zur einheitlichen Orientierung mit einer verbalen Bezeichnung zu versehen ist. Für die Qualitätsbeurteilung gibt es eine standardisierte Urteilsskala in englischer Sprache, die der Autor [BS1284-1, 2003] entnimmt:

- excellent (5),
- good (4),
- fair (3),
- poor (2),
- bad (1).

In [NOKIA-Slides, 2003] wird diese Skala als "ITU MOS scale" bezeichnet und die Varianzanalyse zur Auswertung der Daten empfohlen. Die deutsche Übersetzung lautet:

- ausgezeichnet (5)
- gut (4),
- ordentlich (3),
- dürftig (2),
- schlecht (1).

Die spätere statistische Auswertung mit parametrischen Tests fordert, dass die Abstände zwischen den einzelnen Skalenstufen äquidistant sind, also eine Intervallskala vorliegt. Dafür ist keine Verifizierung bekannt. Es gibt allerdings einige Untersuchungen, die darauf hinweisen, andere Skalen zu verwenden, welche speziell als Intervallskalen konstruiert sind.

Tabelle 4.4: Datenschema des vollständigen sechsfaktoriellen Hörversuches 1 mit fünf Wiederholungsfaktoren und einem Gruppierungsfaktor. Alle Faktoren haben je zwei Faktorstufen (z.B.: zwei verschiedene Instrumente I_1 und I_2 , sowie zwei Musiker M_1 und M_2 , ...). S_1 und S_2 sind Stichproben mit je n Probanden.

Instrument I		I_1																I_2 ...	
Musiker M		M_1								M_2								M_1 ...	
Anspiel A		A_1				A_2				A_1				A_2				A_1 ...	
Raum R		R_1		R_2		R_1		R_2		R_1		R_2		R_1		R_2		R_1 ...	
Sequenz Se		Se_1	Se_2	Se_1	...														
Hörergruppe	HG_{Git}	S_1	...																
	HG_{Laien}	S_2	...																

(zum Beispiel die sogenannte CR-10 Kategorieurteilsskala nach Borg). Für eine Einleitung zum Thema soll hier auf [Jekosch, 2000] verwiesen werden. In [BS1284-1, 2003] wird darauf hingewiesen, dass die Verwendung einer verbalen Skalenbezeichnung ("pre-defined intermediate anchor points") das Ergebnis verzerren kann. Wenn nur eine numerische Skala ohne verbale Bezeichnung verwendet wird, so sind die erhobenen Daten jeder Versuchsperson für Mittelwert und Standardabweichung zu normalisieren. Solch ein Verfahren wäre unter anderem vorteilhaft für den internationalen Vergleich von Testergebnissen. Der Autor geht im Folgenden davon aus, dass die verwendete Skala Intervallcharakter besitzt, beziehungsweise die Abweichungen tolerierbar sind.

Zu Beginn des Hörversuchs wurden für jeden Probanden zehn ausgewählte Einführungsbeispiele gespielt, um den Hörer an das Spektrum verwendeter Stimuli zu gewöhnen und ihn mit der Skala vertraut zu machen. Die ersten zehn Urteile sollen im weiteren Verlauf nicht mit in die Bewertung einbezogen werden, was der Proband nicht wusste. Dies kann man als Training der Versuchsperson betrachten. In [BS1534-1, 2003] wird die Wichtigkeit einer Trainingsphase unterstrichen. Man kann vermuten, dass sich nach den zehn Einführungsbeispielen die "innere Referenz" der Versuchsperson stabilisiert hat. Danach wurden sequentiell alle 32 Aufnahmen zwei mal, in für jede Versuchsperson zufälliger Reihenfolge, gespielt. Durch diese Wiederholungen, in Kombination mit der Randomisierung, sollte nach Mittelung über alle Probanden ein möglichst von der Reihenfolge unabhängiges Urteil erhalten werden. In Abschnitt 2.1.2 wurden bereits psychologische Effekte erläutert und geschlussfolgert, dass eine zufällige Präsentation der Hörbeispiele oft vorteilhaft ist.

[BS1284-1, 2003] empfiehlt keinen Hörversuch länger als 15 - 20 Minuten ohne Unterbrechung durchzuführen. Nach 40 Beispielen (etwa 20 Minuten) wurde der Hörversuch für eine fünfminütige Pause unterbrochen, nach der erneut fünf ausgewählte Einführungsbeispiele gespielt wurden, welche nicht in die Auswertung einzubeziehen sind.

Der Hörversuch wurde mit der Skriptsprache TclTk programmiert und auf einem PC im Audio-Studio (Willkomm-Studio, Raum 71, Barkhausenaubau, TU-Dresden) durchgeführt. Vor Beginn des Hörversuchs wurde eine schriftliche und eine mündliche Einweisung in das System gegeben, sowie die Lautstärke den Wünschen des Probanden angepasst. Screenshots der Einweisung zu Beginn des Hörversuchs und der Benutzeroberfläche sind im Anhang D, Abbildung D.1 und D.2 zu sehen. Die mündliche Einweisung diente zur Erhöhung der Motivation der Versuchspersonen, indem allgemeine Hintergründe zum Projekt und die Wichtigkeit von Hörversuchen dargestellt wurden. Dabei erhielten die Probanden keine detaillierten Informationen zum Ursprung der Hörbeispiele, um den in Abschnitt 2.1.2 erläuterten Erwartungsfehler gering zu halten. Es war jeweils nur ein Proband im Raum anwesend, wodurch eine gegenseitige Beeinflussung der Versuchspersonen verhindert wurde. Die Lautstärke kann individuell gewählt werden und ist innerhalb einer Sitzung konstant zu halten.

4.3 Entwurf Hörversuch 2

Mit dem sequentiellen Hörversuch 1 ist kein direkter Qualitätsvergleich zwischen zwei oder mehreren Instrumenten möglich, durch den eventuell genauere Ergebnisse erzielt werden können. Ein hierarchischer *paarweiser Vergleich* zur Qualitätsbeurteilung wird zum Beispiel in [Boullosa, 1997] beschrieben, allerdings nur für vier Gitarren, zwei Spieler und

zwei Musiksequenzen (je etwa 10 s). Die Versuchspersonen im Hörversuch werden nicht in Hörergruppen eingeteilt und beurteilen die Aufgabe als schwierig. Es wird berichtet das Probanden dazu neigen, die zweite Gitarre in einem Paar zu bevorzugen. Würde man einen vollständigen paarweisen Vergleich der 32 Aufnahmen aus Hörversuch 1 durchführen, müsste man 696 Paare beurteilen. Selbst wenn nur 16 Aufnahmen zu vergleichen wären, blieben 160 Paare zu beurteilen. Dies würde den zeitlichen Rahmen des Projektes sprengen und so wird die Methode des paarweisen Vergleichs verworfen.

Eine effektive Methode soll, angelehnt an [BS1534-1, 2003], entwickelt werden. Dort wird die sogenannte MUSHRA-Methode zum Vergleich der mittleren Klangqualität von Audio-codierungssystemen verwendet. Dabei kann der Proband beliebig zwischen einer Referenz (bestes Hörbeispiel) und verschiedenen schlechteren Hörbeispielen wechseln. Er vergleicht nicht nur zwei Beispiele, wie bisher, sondern gleich einen ganzen Block. Damit werden die Vorteile eines vollständigen Paarvergleichs erhalten. Es wird empfohlen nicht mehr als 15 Beispiele in einem Block anzubieten. Die Erfahrung zeigt, dass Probanden dabei mit einer groben Qualitätsschätzung beginnen, um dann die Stimuli zu sortieren oder in eine Reihenfolge zu bringen und abschließend absolute Urteile zu fällen. Die zu verwendende quasikontinuierliche Qualitätsskala sollte in fünf gleich große Intervalle eingeteilt werden, welche mit der oben beschriebenen verbalen Bezeichnung zu versehen sind. Nur für das zuletzt gehörte Beispiel sollte eine Beurteilung möglich sein. Die Methode verwendet unter anderem eine direkte Referenz und geht davon aus, dass alle zu untersuchenden Stimuli in einen Block passen.

Im vorliegenden Fall ist die Anzahl der Stimuli, selbst mit reduziertem Versuchsdesign, zu hoch. Das Versuchsdesign ist deswegen noch weiter zu reduzieren und es soll nur noch eine Sequenz, die Melodie "Para Guitarra", verwendet werden. Das resultierende Datenschema mit 16 Stimuli ist in Tabelle 4.5 zu sehen.

Tabelle 4.5: Datenschema des vollständigen fünffaktoriellen Hörversuches 2 mit vier Wiederholungsfaktoren und einem Gruppierungsfaktor. Alle Faktoren haben je zwei Faktorstufen. Es wurde nur eine Melodiesequenz (Melodie Para Guitarra) verwendet. S_1 und S_2 sind Stichproben mit je n Probanden.

Instrument I		I_1				I_2				...	
Musiker M		M_1				M_2				M_1	...
Anspiel A		A_1	A_2		A_1	A_2		A_1	...		
Raum R		R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2	R_1	...
Hörergruppe	HG_{Git}	S_1	...								
	HG_{Laien}	S_2	...								

Es sollen zudem mehrere Blöcke zu je sechs Beispielen gebildet werden. Da kein Beispiel im Voraus als bestes eingeordnet werden kann, verzichtet der Autor auf die Verwendung einer benannten Referenz. Um die Blöcke später aneinander ausrichten zu können, werden zwei ausgewählte Ankerbeispiele in jedem Block wiederholt. Diese können auch als versteckte Referenz betrachtet werden. Sie sind so gewählt, dass ein gutes und ein relativ schlechtes Beispiel aus den vorläufigen Ergebnissen von Hörversuch 1 verwendet wird.

Dies soll eine gleichbleibende Verwendung der Skala bewirken. Nehmen wir zum Beispiel an, dass in einem ersten Block zufällig nur Stimuli eines schmalen Qualitätsspektrums enthalten wären (zum Beispiel nur relativ schlechte oder nur relativ gute Qualität). In einem zweiten Block könnten Stimuli eines breiten Qualitätsspektrums enthalten sein. Die Skala würde vermutlich völlig anders verwendet und ein Vergleich zwischen den Blöcken wäre nicht möglich. Darum soll jeder Proband in jedem Block als erstes Beispiel den Stimulus "G1IBMR1" hören, welcher nach einer vorläufigen Auswertung des Hörversuchs 1 oft gut bewertet wurde. An fünfter Stelle folgt in jedem Block der, als wesentlich schlechter angenommene, Stimulus "G5SPMK2". Alle 16 Hörbeispiele werden für jeden Probanden über vier Blöcke zufällig auf die Plätze zwei, drei, vier und sechs verteilt. Ein fünfter Block wird angehängt, der eine exakte Wiederholung des ersten Blockes ist um zu untersuchen, ob die Ankerbeispiele im selben Kontext die gleichen Urteile erhalten.

Später sind die Blöcke mit Hilfe der Ankerbeispiele aneinander auszurichten, um die absoluten Urteile zwischen den Blöcken vergleichen und auswerten zu können. Dabei können ein oder beide Ankerbeispiele verwendet werden. Ausrichtung an einem Ankerbeispiel entspricht einer konstanten Verschiebung der Urteile innerhalb eines jeden Blockes. Bei Ausrichtung an zwei Ankerbeispielen wird die Urteilsskalen zusätzlich gedehnt oder gestaucht.

Zu Beginn eines jeden Blockes werden alle Stimuli des Blockes nacheinander gespielt und es ist eine erste Bewertung der Gesamtqualität möglich. Der weitere Hörversuchsentwurf hält sich überwiegend an die oben dargestellten Richtlinien. So kann jedes Hörbeispiel abgebrochen werden. Der Proband kann beliebig oft zwischen den Beispielen innerhalb eines Blockes wechseln. Das aktuell gespielte Beispiel ist durch eine farbliche Hinterlegung eindeutig zu kennzeichnen. Ein Urteil ist nur für das jeweilig zu letzt gehörte Beispiel möglich. Die Skala soll nicht nur fünf Stufen enthalten, sondern quasikontinuierlich sein, wobei sie von eins bis vier in 80 Stufen eingeteilt wird. Dies ist im Gegensatz zu Hörversuch 1 notwendig, da zu vermuten ist, dass Probanden unzufrieden werden, wenn sie einen geringen Qualitätsunterschied zwischen zwei Stimuli wahrnehmen, ihn aber nicht auf einer fünfstufigen Skala ausdrücken können. In diesem Fall wird wieder der zufriedene und damit motiviertere Proband bevorzugt und die Skala quasikontinuierlich gewählt. Jeder Proband wird wieder schriftlich und mündlich eingewiesen. Screenshots der Einweisung und der Benutzeroberfläche sind in Anhang D, Abbildung D.3 und Abbildung D.4 zu sehen.

In [BS1534-1, 2003] wird empfohlen, dass nur jeweils ein Attribut untersucht werden soll. Dies ist hier der Fall, da nur die Gesamtqualität des Instrumentes gesucht ist. Die Lautstärke soll wieder individuell eingestellt und innerhalb einer Sitzung konstant sein.

4.4 Entwurf Hörversuch 3

Bisher wurden immer nur zwei Gitarren unter verschiedenen Randbedingungen untersucht. Um einen Überblick der zehn Gitarren zu erhalten, sollen sie in einem letzten Hörversuch 3 mit ausgewählten konstanten Randbedingungen untersucht werden. Alle zehn Instrumente wurden mit der Melodiesequenz Se_1 (Para Guitarra) im Raum R_1 (Konferenzraum) vom Gitarristen M_1 (IB) gespielt. Das zweite Anspiel soll verwendet werden. Ein entsprechendes Datenschema ist in Tabelle 4.6 dargestellt. Für diesen Versuch bietet es sich an, die im Hörversuch 2 entwickelte Methodik in Blöcken zu verwenden. Es wurden

alle zehn Stimuli in einem Block präsentiert. Der Versuch fand im direkten Anschluss an Hörversuch 2 statt. Es war daher keine erneute Einweisung nötig. Wieder wurden alle Beispiele zu Beginn nacheinander gespielt, so dass der Proband mit ihnen vertraut werden konnte. Ein Screenshot der Benutzeroberfläche ist in Anhang D, Abbildung D.5 zu sehen.

Tabelle 4.6: Datenschema des zweifaktoriellen Hörversuches 3 mit einem zehnstufigen Wiederholungsfaktor (Instrument) und einem zweistufigen Gruppierungsfaktor (Hörergruppe). Alle Aufnahmen wurden mit der Melodiesequenz Se_1 (Para Guitarra) im Raum R_1 (Konferenzraum) vom Gitarristen M_1 (IB) gespielt. Es wurde immer das zweite Anspiel A_2 verwendet. S_1 und S_2 sind Stichproben mit je n Probanden.

	Instrument	I	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
Hörergruppe	HG_{Git}		S_1									
	HG_{Laien}		S_2									

4.5 Wahl repräsentativer Stichproben

In [BS1284-1, 2003] wird zur subjektiven Beurteilung von Klangqualität empfohlen "Experten" zu verwenden. "Experten" sind hier trainierte Probanden, die sich mit den zu beurteilenden Variablen und der Verwendung der Skala auskennen. Allerdings sollte im Zweifelsfall untersucht werden, ob es einen Unterschied zwischen Experten und Nicht-Experten gibt. Die minimale Anzahl für Experten wird zu zehn angegeben und die Anzahl von Nicht-Experten sollte normalerweise mindestens 20 sein. In der vorliegenden Arbeit kommen als Experten nur "Gitarristen" in Frage und als Nicht-Experten "Laien". Beide Gruppen haben keine Erfahrung mit der Verwendung von Skalen und müssen damit vertraut gemacht werden. In [BS1534-1, 2003] wird empfohlen, vor Durchführung eines Tests zur Qualitätsprüfung von Audiocodierung, Probanden auszuwählen, die Erfahrungen im kritischen Hören von Audiosignalen und keine Hörbeeinträchtigung haben. Als Argument dafür wird die erhöhte Effizienz des Hörversuchs angeführt. Allerdings beeinflusst dies die externe Validität des Versuchs negativ. Deswegen sollen in dieser Arbeit nur Probanden mit Gehörschäden ausgeschlossen werden. Es wird in [BS1534-1, 2003] weiterhin die Möglichkeit zum nachträglichen Ausschluss von Probanden erläutert, die zum Beispiel die Skala nicht konsistent verwenden. Da dadurch die Gefahr besteht, Daten im Nachhinein stark zu manipulieren, verzichtet der Autor darauf. Dies ist wiederum günstig für die externe Validität des Versuchs.

Es wurden Mitarbeiter des Instituts für Akustik und Sprachkommunikation an der TU-Dresden, Studenten der TU-Dresden, Studenten der Musikhochschule Carl-Maria-von-Weber (Studienrichtung Gitarre), Mitglieder Dresdner Bands und Bekannte des Autors für die Hörversuche geworben. Die Auswahl war damit nicht zufällig. Alle Probanden nahmen unentgeltlich an den Hörversuchen teil. Es wurden jeweils 32 Probanden für jeden Hörversuch befragt. Am ersten Hörversuch nahmen 24 Männer und 8 Frauen am zweiten und dritten Hörversuch 28 Männer und 4 Frauen teil. Im ersten Hörversuch waren davon

4 professionelle Gitarristen (Studenten) und 12 Hobbygitarristen sowie 16 Laien. Am zweiten und dritten Hörversuch nahmen 6 professionelle Gitarristen, 10 Hobbygitarristen und wieder 16 Laien teil. Davon hatten 12 bereits am 1. Hörversuch mitgewirkt, der an einem separaten Tag stattfand. Alle Probanden gaben an keine bekannten Gehörschäden zu besitzen.

Für die spätere Auswertung sollten die Probanden in zwei Hörergruppen eingeteilt werden. Als "Gitarristen" zählten dabei sowohl Profis als auch Hobbygitarristen. Alle anderen wurden als "Laien" zusammengefasst.

Kapitel 5

Auswertung

In diesem Kapitel sollen die erhobenen Daten der in Kapitel 4 beschriebenen Hörversuche mit statistischen Methoden ausgewertet werden.

Man kann folgende allgemeine Untersuchungshypothese aufstellen, die anhand der Daten zu prüfen ist:

H_0 : Die Qualitätsbeurteilung von Musikinstrumenten anhand von gespielten Musiksequenzen ist nicht abhängig von Faktoren wie dem Musiker, dem Musikstück, dem Instrument, dem Raum, dem Anspiel oder der Hörergruppe.

Die dazugehörige Alternativhypothese lautet:

H_1 : Die Qualitätsbeurteilung von Musikinstrumenten anhand von gespielten Musiksequenzen ist abhängig von mindestens einem Faktor wie dem Musiker, dem Musikstück, dem Instrument, dem Raum, dem Anspiel oder der Hörergruppe.

Es sind weiterhin Interaktionen zwischen den einzelnen Faktoren zu überprüfen. Für jeden Faktor und jede Interaktion kann eine spezielle Hypothese aufgestellt werden. Existieren wenige oder keine Interaktionen, können, wie in Abschnitt 4.2 erläutert, in nachfolgenden Untersuchungen mit mehr Faktorstufen (zum Beispiel zehn verschiedenen Instrumenten) vorteilhafte hierarchische Strukturen zum Einsatz kommen. Dies wäre sehr von Vorteil, denn der Versuchsaufwand reduziert sich dadurch drastisch.

5.1 Hörversuch 1

Es wurde bereits erläutert, dass die fünf Stufen der Qualitätsskala als äquidistant angenommen werden, was nicht unkritisch ist. Jeder Stufe wird ein Zahlenwert von eins bis fünf zugeordnet, wobei eins das schlechteste und fünf das beste Qualitätsurteil abbildet. Da die Auswertung sehr rechenaufwendig ist, soll das statistische Softwarepaket "SPSS" ("Statistical Package for the Social Sciences") Version 13 verwendet werden. Alternativ hätten Programme wie z.B. "STATISTICA" oder "SAS" zum Einsatz kommen können. Auch im Internet findet sich eine Reihe kostenloser Programme, zum Beispiel "R" (www.r-project.org).

Deskriptive Statistik

Im Hörversuch 1 wurden für jede mögliche Faktorstufenkombination zwei Messwerte pro Versuchsperson aufgenommen. Der Autor bildet zur weiteren statistischen Betrachtung den jeweilige Mittelwert aus beiden Messwerten. Die gemittelten Messwerte aller Versuchspersonen für eine Faktorstufenkombination werden im Folgenden als Stichprobe bezeichnet. Das in Tabelle 5.1 dargestellte Schema veranschaulicht dies und teilt jeder Stichprobe einen Namen zu. Zum Beispiel sind die Messwerte für das Hörbeispiel mit Instrument 1, Musiker 2, Sequenz 1, Raum 1 und Anspiel 2 unter der Stichprobe "G1SPMK2" zusammengefasst. Die praktische Bedeutung der Faktorstufen ist in Tabelle 5.2 aufgelistet. Eine tabellarische Übersicht der erhaltenen Messwerte, sortiert nach den beiden Hörergruppen "Laie" und "Gitarrist", findet sich in Anhang E, Abbildung E.1. Es werden zudem Mittelwerte AM und Mediane Me angegeben. Der Unterschied beider Werte gibt einen Hinweis auf die Form der Verteilung der Messwerte einer Stichprobe.

Dies ist auch aus den Häufigkeitsverteilungen der Variablen ersichtlich. In Abbildung 5.1 ist beispielhaft die Häufigkeitsverteilung für die Stichprobe "G1IBMK1" dargestellt. Man erkennt, dass für diese Stichprobe die Verteilung bei Gitarristen schmalgipfelig und für Laien leicht rechtssteil ist. Einen ähnlichen Hinweis hätte man auch aus der Tabelle in Abbildung E.1 durch Vergleich von AM und Me entnehmen können. Für Laien ist das AM (Mean) mit 3,59 kleiner als der Me (Median) mit 3,75 und man vermutet somit eine etwas rechtsteile Verteilung. Für Gitarristen ist das $AM = 3,53$ und der $ME = 3,50$ nahezu identisch und man kann von einer symmetrischen Verteilung ausgehen. In Anhang E, Abbildung E.2, E.3 und E.4 sind die Häufigkeitsverteilungen für alle Stichproben aus Hörversuch 1 graphisch dargestellt.

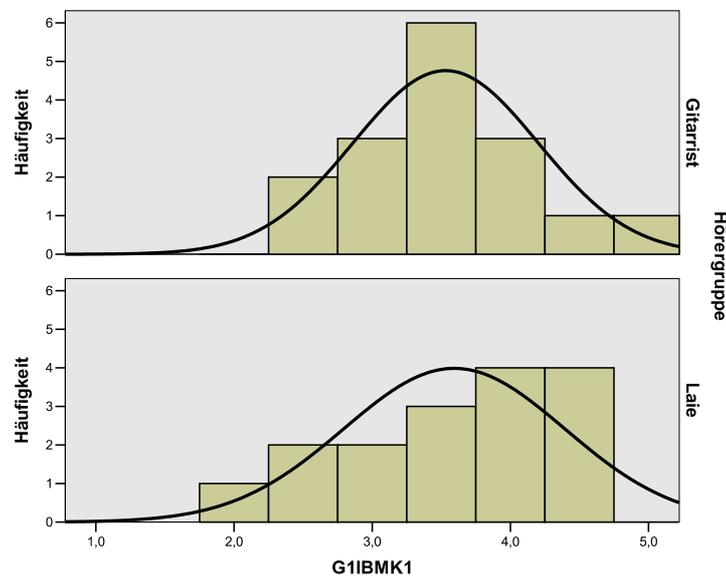


Abbildung 5.1: Häufigkeitsverteilung der Stichprobe "G1IBMK1" mit überlagerter Normalverteilungskurve getrennt dargestellt für die zwei Stufen des Gruppierungsfaktors: Gitarristen (oben) und Laien (unten).

Die Daten lassen sich weiterhin als Error-Bar (Abbildung 5.2) darstellen. Es ist ersichtlich, dass sich die Mittelwerte zwischen dem Qualitätsurteil "ordentlich" und "gut", also

Tabelle 5.1: Bezeichnung der Stichproben in Abhängigkeit der Kombinationen von Stufen der unabhängigen Faktoren.

Instrument I	Musiker M	Sequenz Se	Raum R	Anspiel A	Stichprobe		
1	1	1	1	1	G1IBMK1		
				2	G1IBMK2		
				2	G1IBMR1		
			2	1	G1IBMR2		
				1	G1IBTK1		
				2	G1IBTK2		
	2	1	1	1	1	G1SPMK1	
					2	G1SPMK2	
					2	G1SPMR1	
				2	1	G1SPMR2	
					1	G1SPTK1	
					2	G1SPTK2	
		2	2	1	1	1	G5IBMK1
						2	G5IBMK2
						2	G5IBMR1
					2	1	G5IBMR2
						1	G5IBTK1
						2	G5IBTK2
2	2	2	1	1	G5IBTR1		
				2	G5IBTR2		
				2	G5SPMK1		
			2	1	G5SPMK2		
				1	G5SPMR1		
				2	G5SPMR2		
	2	2	2	1	1	G5SPTK1	
					2	G5SPTK2	
					2	G5SPTR1	
				2	1	G5SPTR2	
					2	G5SPTR2	
					2	G5SPTR2	

Tabelle 5.2: Praktische Bedeutung der Faktorstufen.

Instrument	I_1 : Referenzgitarre G1	I_2 : Referenzgitarre G2
Musiker	M_1 : Gitarrist IB	M_2 : Gitarrist SP
Anspiel	A_1 : erstes Anspiel	A_2 : Wiederholung
Raum	R_1 : Konferenzraum IfM	R_2 : reflexionsarmer Raum IfM
Sequenz	Se_1 : Melodie Para Gitarra	Se_2 : Tonleiter
Hörergruppe	HG_{Git} : Gitarristen	HG_{Laien} : Laien

zwischen drei und vier konzentrieren. Die Urteile fallen eher positiv aus. Zudem streuen die Messwerte relativ stark um den Mittelwert. Es entsteht der Eindruck, dass die Mittelwerte bei Gitarristen etwas mehr streuen als bei Laien. Die Extreme der Skale werden vermieden.

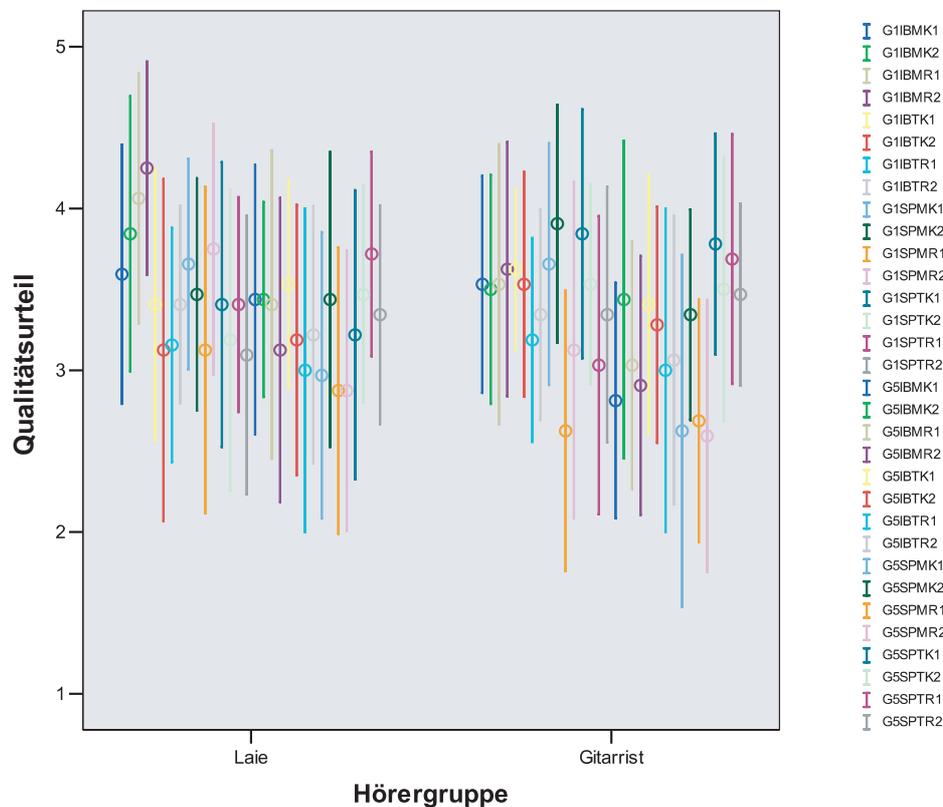


Abbildung 5.2: Error-Bar der Messwertverteilungen aus Hörversuch 1 für 32 Personen, 16 Laien und 16 Gitarristen. Es wird der jeweilige Mittelwert einer Stichprobe \pm eine Standardabweichung abgetragen.

Analytische Statistik

Vor Durchführung einer Varianzanalyse soll überprüft werden, ob die Stichproben als normalverteilt angesehen werden können. Dazu verwendet der Autor den in Abschnitt 2.5.2 bereits erläuterten Kolmogorov-Smirnov-Anpassungstest. Die Ergebnisse sind in Anhang E, Tabelle E.1 zusammengefasst. Geht man von einem Schwellenwert der α -Fehlerwahrscheinlichkeit von 5% aus, werden alle Stichproben als normalverteilt getestet. Betrachtet man einen Schwellenwert von 20% so werden sechs der 32 Stichproben als nicht normalverteilt getestet. Da der zu verwendende Schwellenwert nicht verbindlich festgelegt ist, soll im weiteren von normalverteilten Daten ausgegangen werden. Die Daten werden weiterhin mit SPSS auf Zirkularität überprüft und es wird für alle Stichproben die Zirkularitätsannahme bestätigt.

Mit diesen Voraussetzungen kann nun eine mehrfaktorielle Varianzanalyse mit Messwiederholungen (wie in Abschnitt 2.5.5 beschrieben) verwendet werden. Die Ergebnisse der Varianzanalyse für Effekte zwischen den Versuchspersonen ist in Tabelle 5.3 zu sehen. Der

Tabelle 5.3: Ergebnistabelle der sechsfaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 1 zwischen den Versuchspersonen. *HG* bezeichnet den Faktor Hörergruppe.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
<i>HG</i>	1,722	1	1,722	0,65	0,426	0,021
Error	79,52	30	2,650			

Effekt für die Hörergruppen *HG* ist nicht signifikant. Damit ist das Qualitätsurteil für ein Instrument nicht abhängig davon, ob ein Hörer Laie oder Gitarrist ist. Die Ergebnisse der Varianzanalyse für Effekte innerhalb der Versuchspersonen sind in Tabelle 5.4 zu sehen. Es wird auch die geschätzte Varianzaufklärung $\hat{\eta}_p^2$ angegeben.

Tabelle 5.4: Ergebnistabelle der sechsfaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 1 innerhalb der Versuchspersonen. Die einzelnen Faktoren werden abgekürzt als: Instrument *I*, Musiker *M*, Anspiel *A*, Raum *R*, Sequenz *Se* und Hörergruppe *HG*. Ist ein Haupteffekt oder eine Interaktion sehr signifikant (Irrtumswahrscheinlichkeit < 1%) wird sie mit ** gekennzeichnet, ist sie signifikant (Irrtumswahrscheinlichkeit < 5%) mit *.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
<i>I</i> **	16,000	1	16,000	19,61	0,000	0,395
<i>I</i> * <i>HG</i>	0,098	1	0,098	0,12	0,732	0,004
Error(<i>I</i>)	24,480	30	0,816			
<i>M</i>	1,266	1	1,266	1,46	0,237	0,046
<i>M</i> * <i>HG</i>	1,129	1	1,129	1,30	0,263	0,042
Error(<i>M</i>)	26,059	30	0,869			
<i>Se</i>	0,391	1	0,391	0,13	0,725	0,004
<i>Se</i> * <i>HG</i>	9,379	1	9,379	3,02	0,093	0,091
Error(<i>Se</i>)	93,184	30	3,106			
<i>R</i>	7,910	1	7,910	3,47	0,072	0,104
<i>R</i> * <i>HG</i>	5,063	1	5,063	2,22	0,147	0,069
Error(<i>R</i>)	68,355	30	2,279			
<i>A</i>	0,712	1	0,712	2,29	0,141	0,071
<i>A</i> * <i>HG</i>	0,353	1	0,353	1,13	0,295	0,036
Error(<i>A</i>)	9,326	30	0,311			
<i>I</i> * <i>M</i>	2,066	1	2,066	4,10	0,052	0,120
<i>I</i> * <i>M</i> * <i>HG</i>	0,016	1	0,016	0,03	0,861	0,001
Error(<i>I</i> * <i>M</i>)	15,121	30	0,504			
<i>I</i> * <i>Se</i> **	18,063	1	18,063	24,83	0,000	0,453
<i>I</i> * <i>Se</i> * <i>HG</i>	0,191	1	0,191	0,26	0,612	0,009
Error(<i>I</i> * <i>Se</i>)	21,824	30	0,727			
<i>M</i> * <i>Se</i> **	13,598	1	13,598	19,81	0,000	0,398
<i>M</i> * <i>Se</i> * <i>HG</i>	0,016	1	0,016	0,02	0,881	0,001
Error(<i>M</i> * <i>Se</i>)	20,590	30	0,686			
<i>I</i> * <i>M</i> * <i>Se</i>	1,000	1	1,000	2,44	0,129	0,075
<i>I</i> * <i>M</i> * <i>Se</i> * <i>HG</i>	0,473	1	0,473	1,15	0,291	0,037
Error(<i>I</i> * <i>M</i> * <i>Se</i>)	12,293	30	0,410			
<i>I</i> * <i>R</i>	0,004	1	0,004	0,01	0,924	0,000
<i>I</i> * <i>R</i> * <i>HG</i> **	2,641	1	2,641	6,31	0,018	0,174

Fortsetzung auf nächster Seite

Tabelle 5.4 – Fortsetzung von letzter Seite

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
Error($I * R$)	12,559	30	0,419			
$M * R^*$	2,066	1	2,066	5,44	0,027	0,154
$M * R * HG$	0,562	1	0,562	1,48	0,233	0,047
Error($M * R$)	11,387	30	0,380			
$I * M * R^{**}$	4,516	1	4,516	10,45	0,003	0,258
$I * M * R * HG$	0,098	1	0,098	0,23	0,638	0,007
Error($I * M * R$)	12,965	30	0,432			
$Se * R$	0,063	1	0,063	0,06	0,802	0,002
$Se * R * HG$	0,004	1	0,004	0,00	0,950	0,000
Error($Se * R$)	29,262	30	0,975			
$I * Se * R$	0,766	1	0,766	1,64	0,210	0,052
$I * Se * R * HG$	0,473	1	0,473	1,01	0,323	0,033
Error($I * Se * R$)	14,027	30	0,468			
$M * Se * R^{**}$	5,348	1	5,348	22,24	0,000	0,426
$M * Se * R * HG$	0,016	1	0,016	0,06	0,801	0,002
Error($M * Se * R$)	7,215	30	0,240			
$I * M * Se * R$	0,063	1	0,063	0,15	0,704	0,005
$I * M * Se * R * HG$	0,191	1	0,191	0,45	0,506	0,015
Error($I * M * Se * R$)	12,699	30	0,423			
$I * A$	0,118	1	0,118	0,43	0,518	0,014
$I * A * HG$	0,001	1	0,001	0,00	0,953	0,000
Error($I * A$)	8,271	30	0,276			
$M * A$	0,079	1	0,079	0,32	0,573	0,011
$M * A * HG$	0,001	1	0,001	0,00	0,950	0,000
Error($M * A$)	7,311	30	0,244			
$I * M * A$	0,024	1	0,024	0,11	0,744	0,004
$I * M * A * HG^*$	1,196	1	1,196	5,33	0,028	0,151
Error($I * M * A$)	6,732	30	0,224			
$Se * A^{**}$	4,649	1	4,649	12,78	0,001	0,299
$Se * A * HG$	0,079	1	0,079	0,22	0,644	0,007
Error($Se * A$)	10,912	30	0,364			
$I * Se * A$	0,001	1	0,001	0,00	0,969	0,000
$I * Se * A * HG$	0,938	1	0,938	1,49	0,232	0,047
Error($I * Se * A$)	18,889	30	0,630			
$M * Se * A^{**}$	1,642	1	1,642	6,31	0,018	0,174
$M * Se * A * HG$	0,001	1	0,001	0,00	0,952	0,000
Error($M * Se * A$)	7,811	30	0,260			
$I * M * Se * A$	0,001	1	0,001	0,00	0,959	0,000
$I * M * Se * A * HG$	0,048	1	0,048	0,13	0,719	0,004
Error($I * M * Se * A$)	10,904	30	0,363			
$R * A$	0,024	1	0,024	0,09	0,761	0,003
$R * A * HG$	0,048	1	0,048	0,19	0,670	0,006
Error($R * A$)	7,756	30	0,259			
$I * R * A^{**}$	5,204	1	5,204	17,14	0,000	0,364
$I * R * A * HG$	0,079	1	0,079	0,26	0,613	0,009

Fortsetzung auf nächster Seite

Tabelle 5.4 – Fortsetzung von letzter Seite

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
Error($I * R * A$)	9,107	30	0,304			
$M * R * A$	0,165	1	0,165	0,45	0,508	0,015
$M * R * A * HG$	0,517	1	0,517	1,41	0,245	0,045
Error($M * R * A$)	11,021	30	0,367			
$I * M * R * A$	1,337	1	1,337	6,15	0,019	0,170
$I * M * R * A * HG$	0,220	1	0,220	1,01	0,323	0,033
Error($I * M * R * A$)	6,521	30	0,217			
$Se * R * A^*$	1,806	1	1,806	5,34	0,028	0,151
$Se * R * A * HG$	0,938	1	0,938	2,77	0,106	0,085
Error($Se * R * A$)	10,146	30	0,338			
$I * Se * R * A$	1,337	1	1,337	3,40	0,075	0,102
$I * Se * R * A * HG$	0,024	1	0,024	0,06	0,805	0,002
Error($I * Se * R * A$)	11,779	30	0,393			
$M * Se * R * A$	1,337	1	1,337	3,94	0,056	0,116
$M * Se * R * A * HG$	1,806	1	1,806	5,32	0,028	0,151
Error($M * Se * R * A$)	10,186	30	0,340			
$I * M * Se * R * A$	0,009	1	0,009	0,03	0,869	0,001
$I * M * Se * R * A * HG$	0,165	1	0,165	0,52	0,475	0,017
Error($I * M * Se * R * A$)	9,467	30	0,316			

Man erkennt das nur **ein Haupteffekt** signifikant ist. Das Qualitätsurteil für ein **Instrument** unterscheidet sich sehr signifikant für die zwei untersuchten Instrumente. Instrument I_1 wird mit einem AM von 3,47 besser bewertet als Instrument I_2 mit einem AM von 3,22. Dabei entspricht Instrument I_1 Referenzgitarre G1 der Firma Takamine und I_2 der Referenzgitarre G5 der Firma Landola. Es existiert also ein hörbarer Qualitätsunterschied zwischen den beiden Instrumenten. Die geschätzte partielle Varianzaufklärung beträgt 0,395. Dieser Wert ist größer als 0,14 und man kann nach Abschnitt 2.5.6 von einem großen Effekt sprechen, obwohl die Differenz der beiden Mittelwerte nur 0,25 beträgt. Diese Interpretation ist wegen der signifikanten Interaktion zwischen Instrument und Sequenz $I * Se$ und weiterer Interaktionen höherer Ordnung mit dem Faktor Instrument jedoch zu relativieren (s.u.).

Im Folgenden sollen signifikante Interaktionen grafisch dargestellt und interpretiert werden. Die Mittelwerte der abhängigen Variablen (klangliche Qualität des Instrumentes) liegen dabei meist in einem engen Bereich zwischen drei und vier, so dass nur dieser Abschnitt in den folgenden Diagrammen auf der Abszisse abgetragen werden soll. Es wird ausdrücklich darauf hingewiesen, dass dadurch ein falscher Eindruck über die Stärke der Interaktionen entstehen kann. Betrachtet man jedoch die ebenfalls geringe Mittelwertdifferenz für den Haupteffekt Instrument im Vergleich zu den Interaktionseffekten, so stellt man fest, dass diese in derselben Größenordnung auftreten. Es ist die Frage zu klären, ob solch kleine Unterschiede praktische Relevanz besitzen oder nicht. Dies soll nicht Gegenstand dieser Arbeit sein.

Als erstes sei in Abbildung 5.3 die bereits erwähnte signifikante disordinale Interaktion zwischen **Instrument** und **Sequenz** dargestellt. Man erkennt, dass Gitarre 1 bei

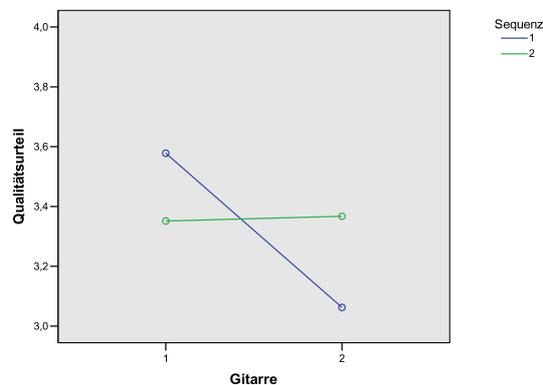


Abbildung 5.3: Interaktionsdiagramm für Instrument und Sequenz.

Sequenz 1 bevorzugt wird und Gitarre 2 bei Sequenz 2. Die Qualität der Gitarre von Takamine wird bei Verwendung der Melodie besser eingeschätzt als bei Verwendung der Tonleiter. Die Gitarre von Landola schneidet für die Tonleiter etwas besser ab als mit der Melodie. Anders formuliert werden beide Gitarren mit der Tonleiter etwa gleich gut eingeschätzt, für die Melodie allerdings unterschiedlich. Diese Interaktion ist nicht unerwartet, da die Bauart eines Instrumentes vom späteren musikalischen Gebrauch mitbestimmt wird. Schlussfolgernd wird dadurch der oben beschriebene Haupteffekt inhaltlich bedeutungslos. Unterschiede zwischen den Instrumenten sind nur in Verbindung mit den Stufen des Faktors Sequenz sinnvoll interpretierbar. Dazu sei vermerkt, dass hier nur zwei ausgesuchte Instrumente und Melodiesequenzen untersucht wurden.

Eine weitere Interaktion, in welcher der Faktor Instrument vorkommt, ist die zweiten Grades zwischen **Instrument, Raum und Hörergruppe** (Abbildung 5.4). Man zeichnet mehrere Diagramme, um die Interaktion von zwei Faktoren mit einem dritten zu überprüfen. Es wird die Interaktion Instrument*Raum in jeweils einem Diagramm für die Hörergruppe Laien bzw. Gitarristen dargestellt. Für Laien ist eine hybride und für Gitarristen eine ordinale Interaktion zu verzeichnen. In allen Fällen wird Gitarre 1 besser bewertet als Gitarre 2, so dass der Haupteffekt eindeutig interpretierbar bleibt. Gitarristen bevorzugen allerdings für beide Gitarren den Konferenzraum (Raum 1), wohingegen Laien bei Gitarre 1 den reflexionsarmen Raum (Raum 2) vorziehen. Auch die Interaktion zweiten Grades zwischen **Instrumenten, Raum und Anspiel** ist für den Haupteffekt zu berücksichtigen. Sie ist in Abbildung 5.5 dargestellt. In beiden Fällen ist die Interaktion ordinal. Auch hier bleibt der Haupteffekt eindeutig interpretierbar. Gitarre 1 wird in allen Fällen besser eingeschätzt als Gitarre 2. Diese Interpretation ist wiederum an der Interaktion **Instrument * Musiker * Raum * Anspiel** zu relativieren. Darauf soll hier verzichtet werden.

Eine weitere signifikante Interaktion ergibt sich zwischen **Musiker und Raum** (Abbildung 5.6). In Raum 1, einem normal halligen Konferenzraum, ändert sich das Qualitätsurteil für ein Instrument kaum, wenn die beiden Musiker das Instrument spielen. In Raum 2, dem reflexionsarmen Raum, hat der Musiker einen größeren Einfluss auf das Qualitätsurteil. Die Interaktion ist hybrid. Man erkennt, dass Urteile mit Aufnahmen aus

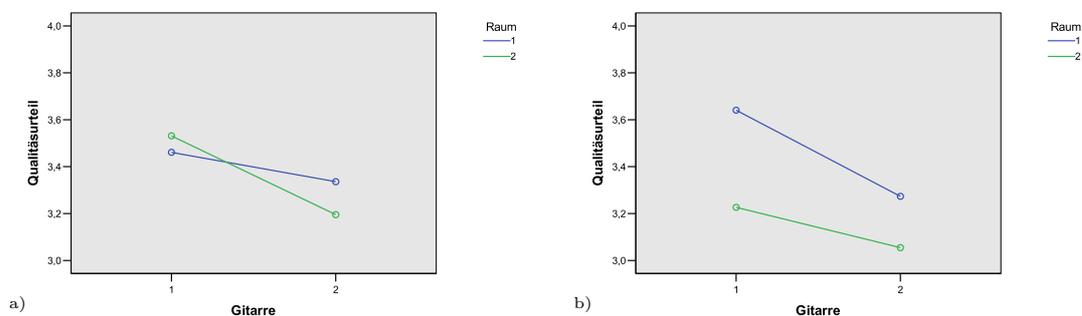


Abbildung 5.4: Interaktionsdiagramm für Instrument, Raum und Hörergruppe a) Laien, b) Gitarristen.

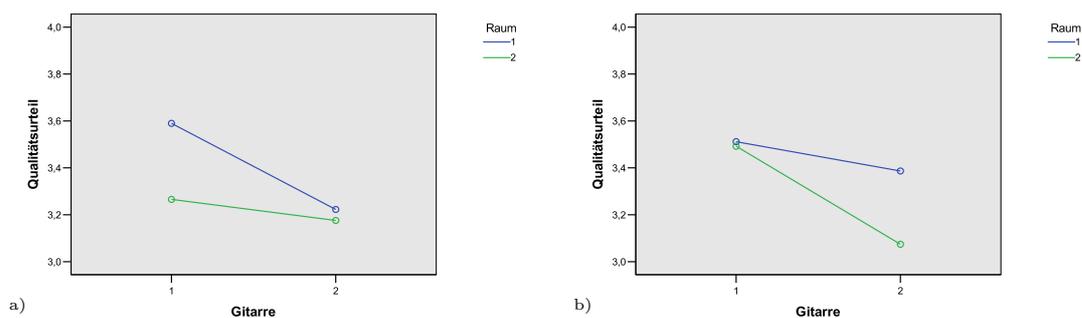


Abbildung 5.5: Interaktionsdiagramm für Instrument, Raum und Anspiel a) Anspiel 1, b) Anspiel 2.

dem Konferenzraum besser ausfallen als Urteile mit Aufnahmen aus dem reflexionsarmen Raum. Dieser Effekt muss wieder relativiert werden an einer signifikanten Interaktion zweiter Ordnung zwischen **Musiker, Raum und Instrument**. Sie ist in Abbildung 5.7 zu sehen. Wieder wird die Interaktion Musiker, Raum betrachtet, allerdings aufgeteilt in zwei Diagramme für Instrument 1 und Instrument 2. Für Instrument 1 wird die bereits beschriebene Interaktion disordinal, für Instrument 2 bleibt sie hybrid. Es ist zu sehen, dass die oben beschriebene Interaktion: Urteile in Raum 1 besser als in Raum 2, nur für Instrument 1 gilt, aber nicht für Instrument 2. Der Einfluss des Musikers auf das Urteil stammt allerdings von Instrument 1, so dass die obige Interpretation inhaltlich bedeutungslos wird. Interaktionen zwischen Musiker und Raum sind nur bei Angaben zum Instrument sinnvoll zu interpretieren. Diese Interpretation wäre noch an der Interaktion zwischen **Instrument, Musiker, Raum und Anspiel** zu relativieren.

Signifikant ist zudem die disordinale Interaktion zwischen **Musiker und Sequenz** (Abbildung 5.8). Die Musiker unterscheiden sich zum Beispiel in Griffsauberkeit und Anschlagstärke. Es kann nun sein, dass diese Merkmale des Musikers in Verbindung mit einer bestimmten Melodiesequenz Einfluss auf das Qualitätsurteil der Probanden haben.

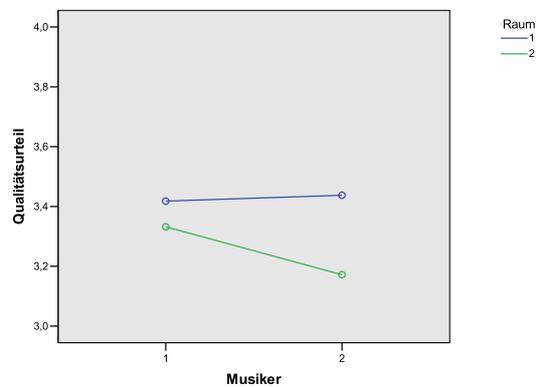


Abbildung 5.6: Interaktionsdiagramm für Musiker und Raum.

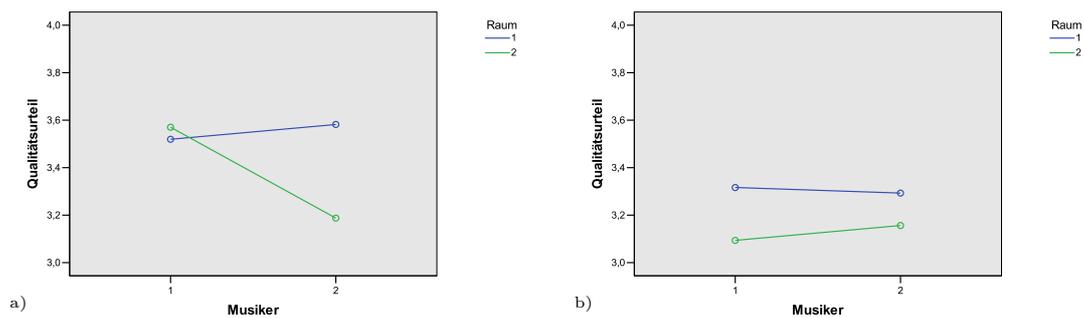


Abbildung 5.7: Interaktionsdiagramm für Musiker, Raum und Instrument a) Gitarre 1, b) Gitarre 2.

So fällt das Urteil besser aus, wenn Musiker 1 die Sequenz 1 (Melodie) spielt, oder wenn Musiker 2 die Sequenz 2 (Tonleiter) spielt. Diese Interpretation ist wieder an den signifikanten Interaktionen höherer Ordnung **Musiker * Sequenz * Raum** und **Musiker * Sequenz * Anspiel** zu relativieren. Die Interaktion zwischen **Musiker, Sequenz und Raum** ist in Abbildung 5.9 dargestellt. Der disordinale Charakter der Interaktion bleibt erhalten. In Raum 1 (Konferenzraum) ist die Interaktion zwischen Musiker und Sequenz weniger stark ausgeprägt als in Raum 2 (reflexionsarmer Raum). Auch in der Interaktion zweiter Ordnung **Musiker * Sequenz * Anspiel** (Abbildung 5.10) bleibt die beobachtete disordinalität erhalten, so dass der Interaktionseffekt erster Ordnung **Musiker * Sequenz** eindeutig interpretierbar bleibt.

Die disordinale Interaktion zwischen **Anspiel und Sequenz** ist in Abbildung 5.11 zu sehen. Es ergibt sich wieder eine disordinale Interaktion. Das Urteil scheint bei Wiederholung für Sequenz 1 (Melodie) besser zu werden und für Sequenz 2 (Tonleiter) schlechter. Es müssen wieder signifikanten Interaktionen höherer Ordnung zur Relativierung des Effekts herangezogen werden. Die varianzanalytische Auswertung hat gezeigt, dass der

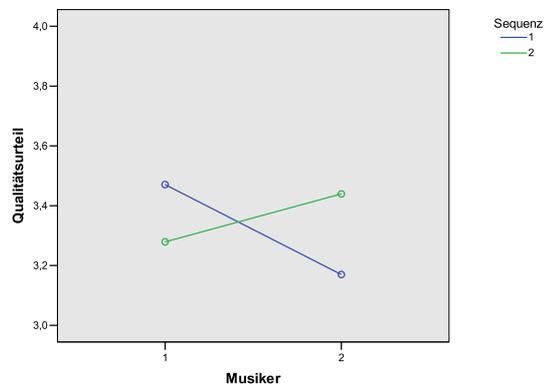


Abbildung 5.8: Interaktionsdiagramm für Musiker und Sequenz.

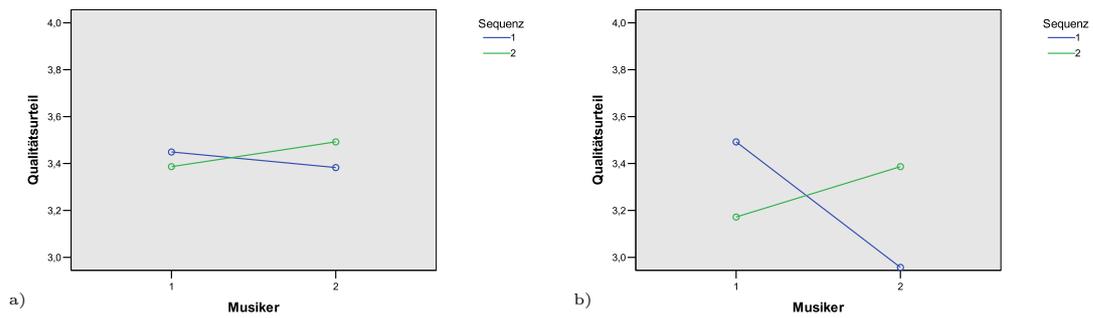


Abbildung 5.9: Interaktionsdiagramm für Musiker, Sequenz und a) Raum 1, b) Raum 2.

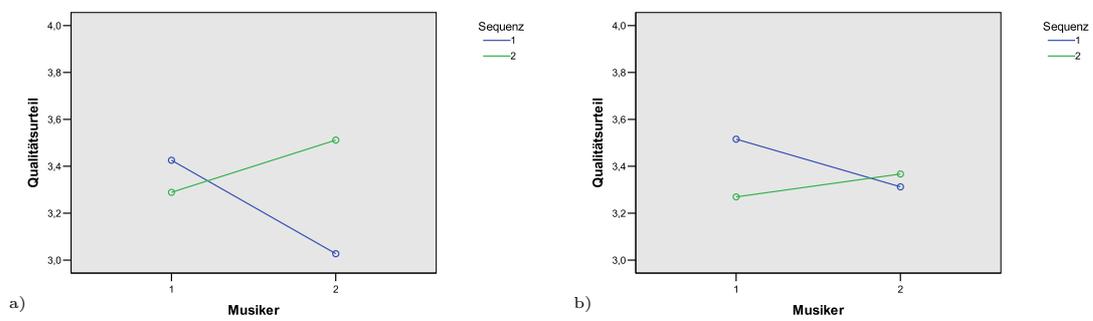


Abbildung 5.10: Interaktionsdiagramm für Musiker, Sequenz und a) Anspiel 1, b) Anspiel 2.

Faktor **Raum** Einfluss auf **Sequenz und Anspiel** hat. Die entsprechende Interaktion ist in Abbildung 5.12 dargestellt. Der disordinale Charakter bleibt nur für Raum 1 (Konferenzraum) erhalten. In Raum 2 (reflexionsarmer Raum) verbessert sich das Urteil zwar etwas für Sequenz 1 (Melodie), bleibt aber annähernd konstant für Sequenz 2 (Tonleiter). Damit ist die Interaktion zwischen Anspiel und Sequenz nur in Zusammenhang mit dem Raum interpretierbar. Weiter oben wurde bereits die Interaktion **Musiker * Sequenz * Anspiel** dargestellt (Abbildung 5.10). Sie soll an dieser Stelle erneut gezeichnet werden, allerdings mit vertauschten Faktoren. Abbildung 5.13 zeigt die Interaktion **Anspiel * Sequenz * Musiker**. Die Interaktion zwischen Anspiel und Sequenz aus Abbildung 5.11 wird in beiden Fällen hybrid und damit eine Aussage ohne Beachtung des Spielers bedeutungslos. Betrachtet man das Qualitätsurteil in Abhängigkeit von Anspiel und Sequenz für Spieler 1, so scheint es sich bei Wiederholungen nicht stark zu ändern und Sequenz 1 (Melodie) wird besser bewertet als Sequenz 2 (Tonleiter). Für Spieler 2 sieht das anders aus. Es existiert eine starke Abhängigkeit vom Anspiel, außerdem wird Sequenz 2 vor Sequenz 1 bevorzugt.

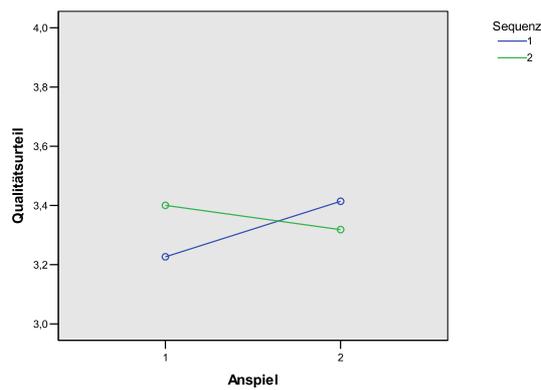


Abbildung 5.11: Interaktionsdiagramm für Anspiel und Sequenz.

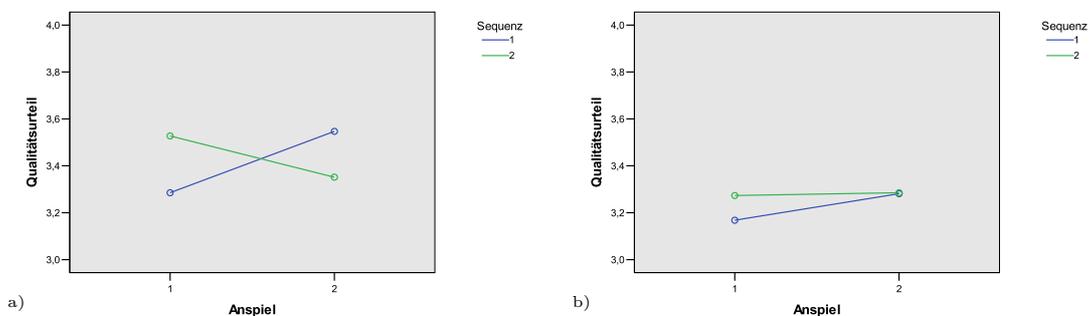


Abbildung 5.12: Interaktionsdiagramm für Anspiel, Sequenz und Raum a) Raum 1, b) Raum 2.

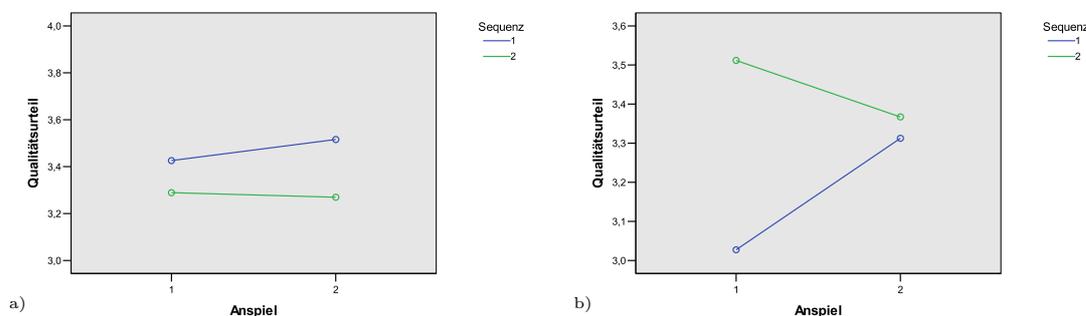


Abbildung 5.13: Interaktionsdiagramm für Anspiel, Sequenz und Musiker a) Musiker 1, b) Musiker 2.

Zu diesen Interpretationen sei nochmals vermerkt, dass eine signifikante Interaktion dritten Grades zwischen **Instrument, Musiker, Anspiel und Hörergruppe**, sowie eine Interaktion vierten Grades zwischen **Musiker, Sequenz, Raum, Anspiel und Hörergruppe** existiert. Auf die aufwendige Darstellung wird an dieser Stelle verzichtet. Es wird schon hier deutlich, dass die Interpretation der Interaktionen komplexe Ergebnisse liefert. Der Autor nimmt an, dass dies an der ebenso komplexen Wahrnehmung des Menschen liegt. Besonders die Beurteilung von Qualität scheint sehr vom Zusammenwirken vieler Einflussfaktoren abzuhängen. Die beschriebenen signifikanten Interaktionen sind zur besseren Übersicht noch einmal in Tabelle 5.5 zusammengefasst. Es werden für jede Interaktion die Qualitätsurteile in Abhängigkeit des Abszissenfaktors verglichen. Als Beispiel ist die Interaktion zwischen Instrument und Sequenz zu betrachten. Das Qualitätsurteil der Probanden war für Aufnahmen der Sequenz 2 (Tonleiter) im Mittel etwa gleich für Instrument 1 und Instrument 2. Bei Verwendung der Sequenz 1 (Melodie), wurde Instrument 1 besser bewertet als Instrument 2. Eine allgemeine Aussage, wie Instrument 1 wurde besser bewertet als Instrument 2, ist daher nicht möglich.

Einige Interaktionen kommen in der Tabelle mehrfach vor, je nachdem welcher Faktor auf der Abszisse abgetragen wurde. Zum Beispiel $I * R * M$, wenn das Qualitätsurteil in Abhängigkeit vom Instrument verglichen werden sollte oder $M * R * I$, wenn Aussagen über die Interaktion von Musiker und Raum zu prüfen sind.

Es ist notwendig, zur Beurteilung der Qualität eines Instrumentes auch die verwendete Sequenz, den Raum und den Musiker einzubeziehen. Es treten zudem viele Interaktionen auf, in denen das Anspiel eine Rolle spielt. Im vorliegenden Fall wird dadurch allerdings die Interpretierbarkeit des wichtigsten Faktors Instrument nicht gestört. Es zeigt sich, dass die Qualitätsbeurteilung von dem Zusammenwirken vieler Faktoren abhängig ist. Leider sind mit diesen Ergebnissen hierarchische Versuchsstrukturen nicht vertretbar. Berücksichtigt man den vernachlässigbaren Einfluss der Hörergruppe, so sind nicht unbedingt Experten (Gitarristen) für weitere Hörversuche notwendig. Dies bedeutet eine erhebliche Erleichterung der Versuchsdurchführung.

Unter Beachtung der großen Streuung der Messwerte, scheinen die Mittelwertsunterschiede der Interaktionen und des Haupteffekts relativ gering. Es stellt sich die Frage der praktischen Relevanz. Die große Streuung der Messwerte kann mehrere Ursachen haben. Zum

Tabelle 5.5: Zusammenfassung der Auswertung von Haupteffekt und Interaktionen in Hörversuch 1. Die einzelnen Faktoren werden abgekürzt als: Instrument I , Musiker M , Anspiel A , Raum R , Sequenz Se und Hörergruppe HG . Bedeutungslose Aussagen sind grau geschrieben.

Interaktion	Beurteilte Qualität des Instrumentes für	Bedingung
I	$I_1 > I_2$	
$I * Se$	$I_1 \approx I_2$	Se_2
	$I_1 > I_2$	Se_1
$I * R * M$	$I_1 > I_2$	$(R_1 \vee R_2)M_1 \vee R_1M_2$
	$I_1 \approx I_2$	R_2M_2
$I * R * HG$	$I_1 > I_2$	
$I * R * A$	$I_1 > I_2$	
$I * R * A * M$		
$M * R$	$M_1 \approx M_2$	R_1
	$M_1 > M_2$	R_2
$M * R * I$	$M_1 \approx M_2$	$(R_1 \vee R_2)I_2 \vee R_1I_1$
	$M_1 > M_2$	R_2I_1
$M * R * I * A$		
$M * Se$	$M_1 > M_2$	Se_1
	$M_1 < M_2$	Se_2
$M * Se * R$	$M_1 > M_2$	$Se_1(R_1 \vee R_2)$
	$M_1 < M_2$	$Se_2(R_1 \vee R_2)$
$M * Se * A$	$M_1 > M_2$	$Se_1(A_1 \vee A_2)$
	$M_1 < M_2$	$Se_2(A_1 \vee A_2)$
$A * Se$	$A_1 < A_2$	Se_1
	$A_1 > A_2$	Se_2
$A * Se * R$	$A_1 > A_2$	Se_1R_1
	$A_1 \approx A_2$	Se_2R_2
	$A_1 < A_2$	$Se_2R_1 \vee Se_1R_2$
$A * Se * M$	$A_1 > A_2$	Se_2M_2
	$A_1 \approx A_2$	Se_2M_1
	$A_1 < A_2$	$Se_1(M_1 \vee M_2)$
$I * M * A * HG$		
$M * Se * R * A * HG$		

einen variieren die verwendeten Randbedingungen in den Hörbeispielen relativ stark, so dass die Urteilsfindung dadurch erschwert wird. Zum anderen könnte der individuelle Geschmack stark variieren. Dies war sowohl bei Gitarristen als auch bei Laien der Fall. Einige übersprangen die vorgesehene Pause und waren gegen Ende des Versuchs relativ erschöpft, was Einfluss auf die Urteilsfähigkeit haben kann. In Abschnitt 2.1 wurde erläutert, dass erkannte und benennbare Qualitätsmerkmale in das Qualitäturteil einfließen. In Gesprächen nach den Hörversuchen zeigte sich, dass die Probanden auf sehr unterschiedliche Merkmale des Instrumentenklangs achteten. Vielen fiel die Benennung von spezifischen Qualitätsmerkmalen schwer und sie gaben an, "nach Gefühl" geurteilt zu haben. Detailliertere Ergebnisse sind also denkbar, wenn gezielt nach bestimmten einzelnen Qualitätsmerkmalen gefragt wird, welche im Voraus zu erläutern sind. Dadurch würde die Wahrnehmung gelenkter, aber die externe Validität des Versuchs sinkt.

5.2 Hörversuch 2

Im zweiten Hörversuch wurde die Hörversuchsmethode geändert. Statt einer sequentiellen Darbietung der Stimuli wurden Blöcke mit mehreren Stimuli angeboten. Zwischen den Beispielen konnte beliebig gewechselt werden. Es wurde außerdem eine quasikontinuierliche Skale verwendet. Der Faktor Sequenz wurde auf eine Stufe reduziert und damit entfernt. Es kamen nur Aufnahmen mit Sequenz 1 (Melodie Para Gitarra) zum Einsatz. Dadurch entfallen die entsprechenden Interaktionen.

5.2.1 Statistische Auswertung ohne Ausrichtung am Ankerbeispiel

Die statistische Auswertung soll wieder mit der, die erhobenen Daten beschreibenden, deskriptiven Statistik beginnen. Danach sollen dieselben analytischen Auswertemethoden wie für Hörversuch 1 zum Einsatz kommen um anschließend die Ergebnisse vergleichen zu können. Zu Beginn soll davon ausgegangen werden, dass die Urteile innerhalb der Blöcke absolut und so auch zwischen den Blöcken vergleichbar sind, ohne dass eine Ausrichtung der Blöcke aneinander notwendig ist.

Deskriptive Statistik

Der Umfang der verwendeten Hörbeispiele wurde durch Verwendung nur einer Musiksequenz um die Hälfte auf 16 reduziert. Zusätzlich wählte der Autor zwei Stimuli als Ankerbeispiele, die in jedem Block (also je fünf mal) bewertet wurden. Das resultierende reduzierte Schema zur Bezeichnung der Stichproben ist in Tabelle 5.6 zu sehen.

Die erhobenen Daten sind in Anhang F, Abbildung F.1 zusammengefasst und Mittelwerte sowie Mediane für jede Stichprobe, getrennt für die beiden Hörergruppen, errechnet. Wieder interessieren die Häufigkeitsverteilungen unter den einzelnen Stichproben. Zum besseren Vergleich mit Hörversuch 1 soll eine ähnliche Darstellungsform gewählt werden. Die quasikontinuierliche Skale (80 Stufen zwischen 1 und 4) wird in Intervalle der Breite 0,5 eingeteilt und die in den jeweiligen Intervall fallenden Messwerte gezählt. Beispielhaft ist in Abbildung 5.14 das Histogramm für die Stichprobe G1IBMK1 dargestellt. Für Gitarristen ist ein starker Einbruch in der Mitte der Verteilung zu beobachten. In den Bereich zwischen 3,25 und 3,75 fallen keine Messwerte. Es ist schwer eine Normalverteilung zu erkennen. Allerdings hängt in diesem Fall der optische Eindruck stark von der gewählten Intervallbreite und Intervallpositionierung ab. Die Verteilung für Laien sieht im Gegensatz zu Experiment 1 linkssteil aus (Vergleich Abbildung 5.1). Die Häufigkeitsverteilungen für alle Stichproben sind im Anhang F, Abbildung F.2 und F.3 zu sehen.

Zeichnet man den Mittelwert und Standardabweichung als ErrorBar für alle Ankerbeispiele (R)ANKER_{xy}, erhält man Abbildung 5.15. Es sind deutlich jeweils zwei Gruppen erkennbar, wobei (R)ANKER_{x1} im Durchschnitt besser bewertet wird als (R)ANKER_{x2}. Bei Laien scheint die Differenz zwischen beiden Beispielen etwas größer zu sein als für Gitarristen. In beiden Fällen streuen die Antworten stark. Block 5 ist eine komplette Wiederholung von Block 1. Die Ankerbeispiele aus Block 1 (ANKER₁₁ und ANKER₁₂) und Block 5 (RANKER₁₁ und RANKER₁₂) sind mit gleichen Symbolen gekennzeichnet. Es wird erwartet, dass die Ankerbeispiele im selben Kontext dieselbe Wertung erhalten. Dies wird hier grafisch bestätigt. Ändert sich der Kontext, also die zufällig gewählten

Tabelle 5.6: Bezeichnung der Stichproben in Abhängigkeit der Kombinationen von Stufen der unabhängigen Faktoren für Hörversuch 2. "ANKER xy " bezeichnet die Ankerbeispiele $y = 1...2$ aus Block $x = 1...4$. "RANKER11" und "RANKER12" sind die Ankerbeispiele aus dem wiederholten Block 1.

Instrument I	Musiker M	Raum R	Anspiel A	Stichprobe		
1	1	1	1	G1IBMK1		
			2	G1IBMK2		
			1	G1IBMR1		
			2	G1IBMR2		
		2	1	G1SPMK1		
			2	G1SPMK2		
			1	G1SPMR1		
			2	G1SPMR2		
	2	1	1	1	G5IBMK1	
				2	G5IBMK2	
			2	1	G5IBMR1	
				2	G5IBMR2	
		2	1	1	G5SPMK1	
				2	G5SPMK2	
			2	1	G5SPMR1	
				2	G5SPMR2	
1	1	2	1	ANKER11		
			1	ANKER21		
			1	ANKER31		
			1	ANKER41		
	2	2	1	2	ANKER12	
				2	ANKER22	
		2	1	2	ANKER32	
				2	ANKER42	
		2	2	1	2	RANKER11
					2	RANKER12

Stichproben um die Ankerbeispiele, so kann sich auch das Urteil für die Ankerbeispiele stark ändern. Vergleicht man allerdings die Mittelwerte der Ankerbeispiele aus den Blöcken eins bis vier, so ist ein relativ konstantes Urteil für die Mittelwerte von "ANKER $x1$ " bzw. "ANKER $x2$ " zu beobachten, wobei die Werte stark streuen. Die gleichbleibenden Mittelwerte sprechen für eine Auswertung ohne Ausrichtung der Blöcke an den Ankerbeispielen. Auf Grund der starken Streuung soll trotzdem in Abschnitt 5.2.2 versucht werden, die Blöcke mit Hilfe der Ankerbeispiele aneinander auszurichten und die Skalennutzung durch Dehnung bzw. Stauchung zu vereinheitlichen.

Stellt man die unbearbeiteten Daten für die übrigen Stichproben aus den Blöcken 1 bis 4 als Error-Bar dar, erhält man Abbildung 5.16. Für Gitarristen ist eine Tendenz erkennbar, die vermuten lässt, dass Gitarre 1 "G1" besser bewertet wurde als Gitarre 2 "G5". Eine ähnliche Tendenz ist auch für Laien erkennbar, allerdings mit größerer Streuung.

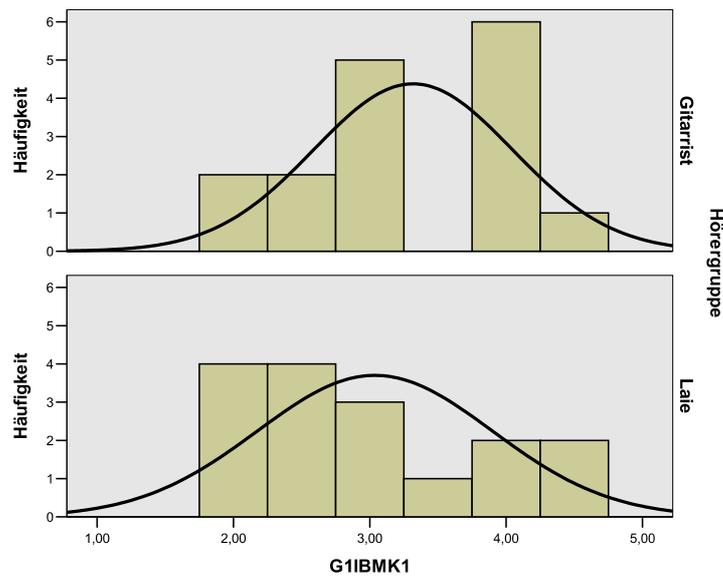


Abbildung 5.14: Häufigkeitsverteilung der Stichprobe "G1IBMK1" aus Hörversuch 2 mit überlagerter Normalverteilungskurve getrennt dargestellt für die zwei Stufen des Gruppierungsfaktors: Gitarristen (oben) und Laien (unten).

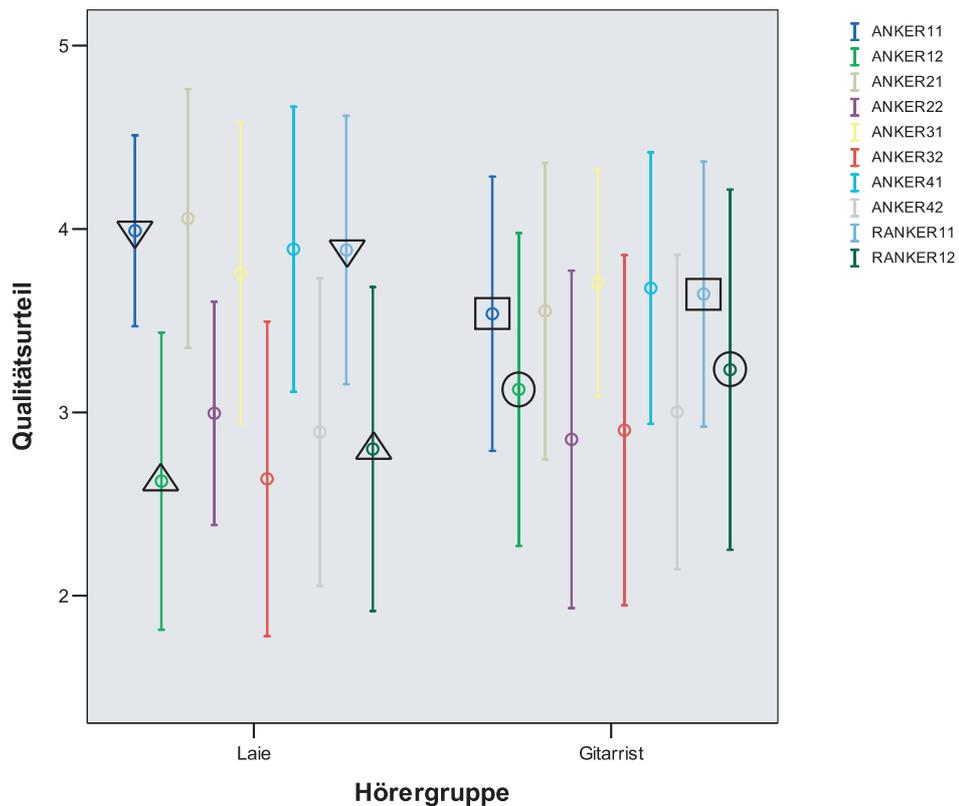


Abbildung 5.15: Error-Bar der Messwertverteilungen der Ankerbeispiele aus Hörversuch 2 für 32 Personen, 16 Laien und 16 Gitarristen. Dargestellt ist der Mittelwert \pm eine Standardabweichung.

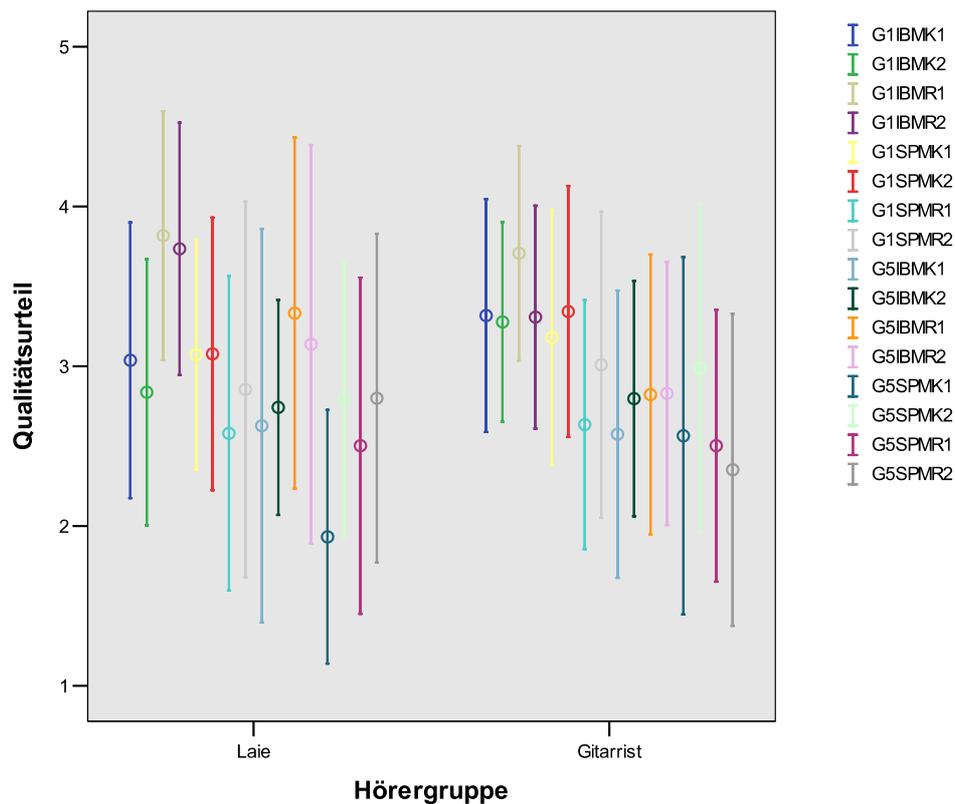


Abbildung 5.16: Error-Bar der Messwertverteilungen der erhobenen Daten aus Hörversuch 2 für 32 Personen, 16 Laien und 16 Gitarristen. Dargestellt ist der Mittelwert \pm eine Standardabweichung.

Analytische Statistik

Ob und welche Unterschiede oder Zusammenhänge existieren, zeigt erst die mathematische Analyse der Daten. Dazu wird wieder auf Normalverteilung geprüft. Die Ergebnisse des KSA-Tests sind in Anhang F, Tabelle F.1, dargestellt. Geht man von einem Schwellenwert der α -Fehlerwahrscheinlichkeit von 5% aus, so sind alle Stichproben normalverteilt. Selbst mit dem kritischeren Schwellenwert von 20% wird nur eine Stichprobe als nicht normalverteilt getestet. Dies ist eine Verbesserung der statistischen Voraussetzungen im Vergleich zu Hörversuch 1. Es wird geschlossen, dass alle Stichproben als hinreichend normalverteilt anzusehen sind. Damit lässt sich eine fünffaktorielle Varianzanalyse durchführen.

Das Ergebnis der ANOVA für Effekte zwischen den Hörergruppen ist in Tabelle 5.7 zu sehen. Es existiert wieder keine Abhängigkeit des Qualitätsurteils von der Hörergruppe.

Tabelle 5.7: Ergebnistabelle der sechsfaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 2 zwischen den Versuchspersonen. *HG* bezeichnet den Faktor Hörergruppe.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
<i>HG</i>	0,055	1	0,055	0,02	0,901	0,001
Error	104,609	30	3,487			

Die Ergebnisse für die Varianzanalyse innerhalb der Versuchspersonen sind in Tabelle 5.8 zusammengefasst. Wieder ist der Einfluss des **Instrumentes** sehr signifikant. Gitarre 1 wird mit einem Mittelwert von etwa 3,2 besser bewertet als Gitarre 2 mit 2,7. Zusätzlich wird ein zweiter Haupteffekt als sehr signifikant erkannt, der des **Musikers**. Beispiele mit Musiker 1 werden im Durchschnitt mit etwa 3,1 besser bewertet als bei Musiker 2 mit 2,8. Beide Effekte sind im Folgenden an vorhandenen Interaktionen zu relativieren. Die Qualitätsachse der relevanten Interaktionsdiagramme soll hier einheitlich von 2,4 bis 3,8 abgetragen werden. Es wird erneut darauf hingewiesen, dass durch diese Darstellung der Eindruck zu starker Effekte entstehen kann. Die Darstellung ist immer in Relation zur gesamten Qualitätsskala von 1 bis 5 zu betrachten.

Tabelle 5.8: Ergebnistabelle der fünffaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 2 innerhalb der Versuchspersonen. Die einzelnen Faktoren werden abgekürzt als: Instrument *I*, Musiker *M*, Anspiel *A*, Raum *R* und Hörergruppe *HG*. Es wurde immer die selbe Musiksequenz (Melodie Para Gitarra) verwendet. Ist ein Haupteffekt oder eine Interaktion sehr signifikant (Irrtumswahrscheinlichkeit < 1%) wird sie mit ** gekennzeichnet, ist sie signifikant (Irrtumswahrscheinlichkeit < 5%) mit *.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
<i>I</i> **	28,069	1	28,069	15,61	0,000	0,342
<i>I</i> * <i>HG</i>	0,717	1	0,717	0,40	0,532	0,013
Error(<i>I</i>)	53,931	30	1,798			
<i>M</i> **	16,345	1	16,345	15,52	0,000	0,341
<i>M</i> * <i>HG</i>	1,276	1	1,276	1,21	0,280	0,039
Error(<i>M</i>)	31,595	30	1,053			
<i>R</i>	1,562	1	1,562	0,87	0,359	0,028
<i>R</i> * <i>HG</i>	6,186	1	6,186	3,43	0,074	0,103
Error(<i>R</i>)	54,084	30	1,803			
<i>A</i>	1,399	1	1,399	3,28	0,080	0,099
<i>A</i> * <i>HG</i>	0,112	1	0,112	0,26	0,613	0,009
Error(<i>A</i>)	12,800	30	0,427			
<i>I</i> * <i>M</i>	0,363	1	0,363	0,87	0,357	0,028
<i>I</i> * <i>M</i> * <i>HG</i>	0,318	1	0,318	0,77	0,389	0,025
Error(<i>I</i> * <i>M</i>)	12,459	30	0,415			
<i>I</i> * <i>R</i>	0,287	1	0,287	0,49	0,490	0,016
<i>I</i> * <i>R</i> * <i>HG</i>	0,230	1	0,230	0,39	0,536	0,013
Error(<i>I</i> * <i>R</i>)	17,589	30	0,586			
<i>M</i> * <i>R</i> **	13,455	1	13,455	25,35	0,000	0,458
<i>M</i> * <i>R</i> * <i>HG</i>	0,203	1	0,203	0,38	0,541	0,013
Error(<i>M</i> * <i>R</i>)	15,923	30	0,531			
<i>I</i> * <i>M</i> * <i>R</i> *	2,404	1	2,404	4,93	0,034	0,141
<i>I</i> * <i>M</i> * <i>R</i> * <i>HG</i>	1,205	1	1,205	2,47	0,126	0,076
Error(<i>I</i> * <i>M</i> * <i>R</i>)	14,623	30	0,487			
<i>I</i> * <i>A</i>	1,099	1	1,099	3,46	0,073	0,104
<i>I</i> * <i>A</i> * <i>HG</i>	0,226	1	0,226	0,71	0,405	0,023
Error(<i>I</i> * <i>A</i>)	9,517	30	0,317			
<i>M</i> * <i>A</i> **	3,969	1	3,969	12,46	0,001	0,294
<i>M</i> * <i>A</i> * <i>HG</i>	0,302	1	0,302	0,95	0,338	0,031

Fortsetzung auf nächster Seite

Tabelle 5.8 – Fortsetzung von letzter Seite

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
Error($M * A$)	9,553	30	0,318			
$I * M * A$	0,034	1	0,034	0,10	0,749	0,003
$I * M * A * HG$	1,292	1	1,292	3,90	0,058	0,115
Error($I * M * A$)	9,937	30	0,331			
$R * A$	1,005	1	1,005	3,38	0,076	0,101
$R * A * HG$	0,102	1	0,102	0,34	0,562	0,011
Error($R * A$)	8,909	30	0,297			
$I * R * A^*$	1,800	1	1,800	5,65	0,024	0,159
$I * R * A * HG$	0,191	1	0,191	0,60	0,445	0,020
Error($I * R * A$)	9,556	30	0,319			
$M * R * A$	0,007	1	0,007	0,01	0,908	0,000
$M * R * A * HG$	0,049	1	0,049	0,10	0,758	0,003
Error($M * R * A$)	15,183	30	0,506			
$I * M * R * A$	0,888	1	0,888	2,92	0,098	0,089
$I * M * R * A * HG$	0,139	1	0,139	0,46	0,504	0,015
Error($I * M * R * A$)	9,121	30	0,304			

Der Haupteffekt Instrument muss an zwei Interaktionen relativiert werden. Als erstes ist die Interaktion **Instrument*Raum*Musiker** zu betrachten (Abbildung 5.17). In allen Fällen wird Gitarre 1 besser bewertet als Gitarre 2 und damit der Haupteffekt bestätigt.

Die Interaktion **Instrument*Raum*Anspiel** bestätigt ebenfalls den Haupteffekt (Abbildung 5.18). Auch hier wird Gitarre 1 immer besser bewertet als Gitarre 2. Damit wird auch das Ergebnis aus Hörversuch 1 für den Faktor Instrument bestätigt.

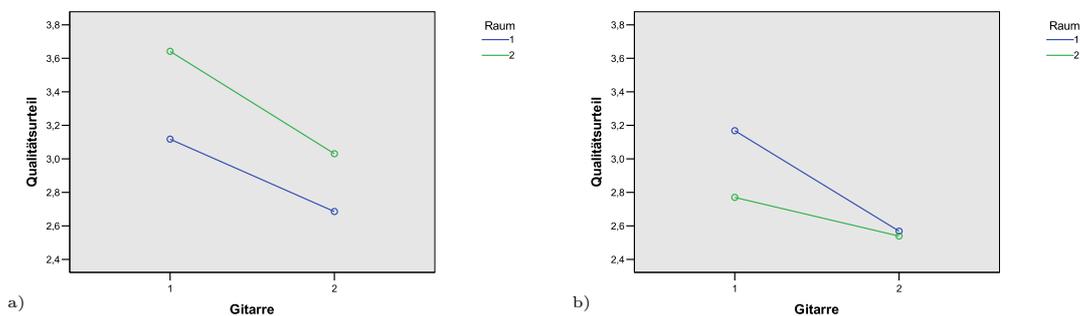


Abbildung 5.17: Interaktionsdiagramm mit den Daten aus Hörversuch 2 für Instrument, Raum und Musiker a) Musiker 1, b) Musiker 2.

Weiterhin ist der Haupteffekt Musiker zu prüfen. Eine signifikante Interaktion existiert zwischen **Musiker und Anspiel** (Abbildung 5.19). Sie lässt vermuten, dass Musiker 1 relativ gleichbleibend auf einer Gitarre gespielt hat und Musiker 2 sich vom ersten zum

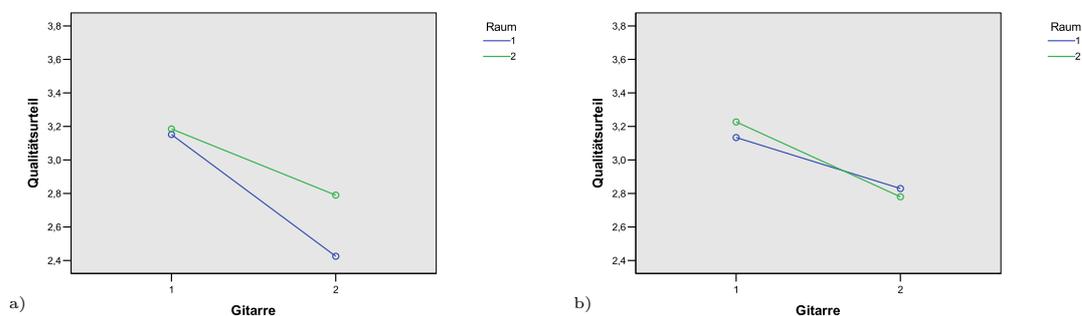


Abbildung 5.18: Interaktionsdiagramm mit den Daten aus Hörversuch 2 für Instrument, Raum und Anspiel
a) Anspiel 1, b) Anspiel 2.

zweiten Anspiel steigerte, was Auswirkung auf das Qualitätsurteil hatte. Generell ist das Urteil für Musiker 1 besser als für Musiker 2, womit der Haupteffekt bestätigt wird.

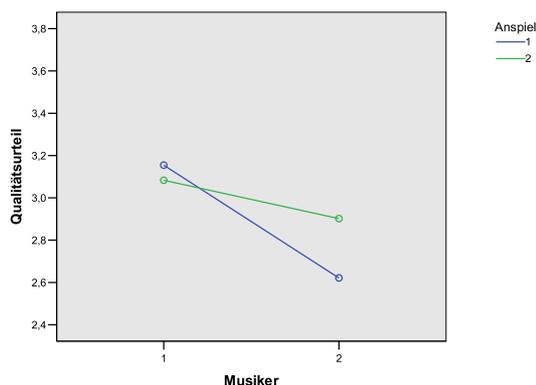


Abbildung 5.19: Interaktionsdiagramm für Musiker und Anspiel in Hörversuch 2.

Die signifikante Interaktion zwischen **Musiker und Raum** (Abbildung 5.20) ist für den Haupteffekt Musiker ebenfalls zu beachten. Sie wurde auch schon in Hörversuch 1 beobachtet. Sie ist wieder hybrid, wenn auch stärker ausgeprägt. In Raum 1 werden Hörbeispiele von beiden Musikern etwa gleich beurteilt, in Raum 2 nicht. Wie im letzten Versuch existiert eine bezogene Interaktion zweiter Ordnung zwischen **Musiker Raum und Instrument** (Abbildung 5.21). Wieder wird für Instrument 1 die Interaktion disordinal und für Instrument 2 bleibt sie hybrid. Wieder ändert sich das Urteil am stärksten zwischen Musiker 1 und Musiker 2 für Gitarre 1 und Raum 2. Etwas schwächer tritt derselbe Effekt für Gitarre 2 in Raum 2 auf. In Raum 1 ändern sich die Urteile kaum. Der komplexe Zusammenhang zwischen Musiker, Raum und Instrument wird hier bestätigt, wobei ein starker Einfluss von Musiker und Raum beobachtet wird.

Die Ergebnisse der Interaktionsauswertung für Hörversuch 2 sollen in Tabelle 5.9 zusam-

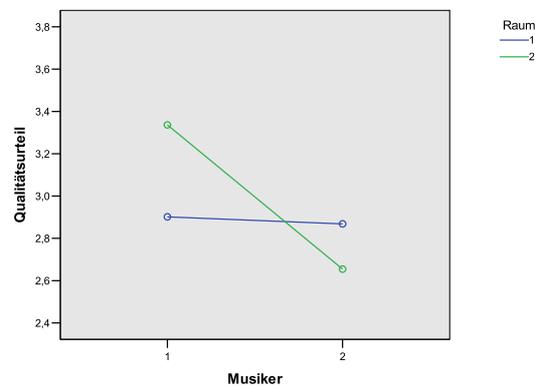


Abbildung 5.20: Interaktionsdiagramm für Musiker und Raum in Hörversuch 2.

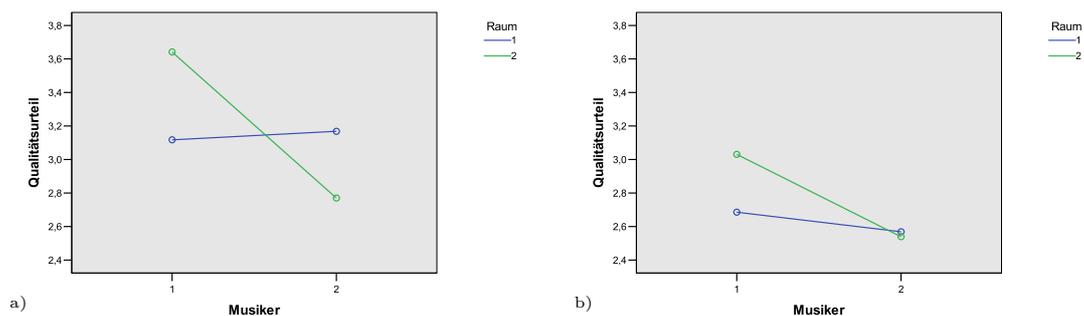


Abbildung 5.21: Interaktionsdiagramm mit den Daten aus Hörversuch 2 für Musiker, Raum und a) Instrument 1, b) Instrument 2.

mengefasst werden. Vergleicht man sie mit Tabelle 5.5, stellt man fest, dass die Interaktionen ähnlich, aber nicht gleich sind. Einige fehlen komplett (zum Beispiel: $I * R * A * M$).

5.2.2 Statistische Auswertung nach Ausrichtung am Ankerbeispiel

Die gewählten Ankerbeispiele sollten bewirken, dass die Skala möglichst vollständig ausgenutzt wird. Es konnte bereits in Abbildung 5.15 gesehen werden, dass die Ankerbeispiele ANKERx1 im Mittel wesentlich bessere Bewertungen erhielten als die Ankerbeispiele ANKERx2. Zudem ist zu sehen, dass die Differenz zwischen den jeweiligen Ankerbeispielen in einem Block für Laien größer zu sein scheint als für Gitarristen. Diese Stauchung der Skala deutet auf ein unterschiedliches Skalenverständnis hin. Es kann angenommen werden, dass durch den Vergleich der Beispiele innerhalb eines Blockes miteinander, die Skala in jedem Block in Abhängigkeit der jeweiligen Beispiele verwendet wurde. Dies wurde bisher nicht berücksichtigt. Es liegt nun nahe, die Skalen zwischen den Blöcken an den jeweiligen Ankerbeispielen auszurichten und somit die Urteile besser auswerten zu können.

Tabelle 5.9: Zusammenfassung der Auswertung von Haupteffekt und Interaktionen in Hörversuch 2. Die einzelnen Faktoren werden abgekürzt als: Instrument I , Musiker M , Anspiel A , Raum R und Hörergruppe HG . Bedeutungslose Aussagen sind grau geschrieben.

Interaktion	Beurteilte Qualität des Instrumentes für	Bedingung
I	$I_1 > I_2$	
$I * R * M$	$I_1 > I_2$	
$I * R * A$	$I_1 > I_2$	
M	$M_1 > M_2$	
$M * A$	$M_1 > M_2$	
$M * R$	$M_1 \approx M_2$	R_1
	$M_1 > M_2$	R_2
$M * R * I$	$M_1 \approx M_2$	$R_1(I_1 \vee I_2)$
	$M_1 > M_2$	$R_2(I_1 \vee I_2)$

Dazu können die Messwerte eines jeden Blockes, einer jeden Versuchsperson, zuerst mit Hilfe eines der beiden Ankerbeispiele verschoben werden, zum Beispiel indem ANKERx1 auf den konstanten Wert vier gesetzt wird. Danach sind die Skalen noch zu dehnen bzw. zu stauchen, indem ANKERx2 auf den konstanten Wert drei gezogen wird. Voraussetzung dafür ist, dass ANKERx1 in jedem Fall besser bewertet wird als ANKERx2, da sich sonst die Rangordnung der Beispiele durch die Transformation invertiert. Dies ist leider nicht der Fall. In Anhang F, Tabelle F.1 kann gesehen werden, das bei 12 von 32 Probanden (davon 7 Gitarristen) wenigstens in einem Fall ANKERx1 schlechter bewertet wird als ANKERx2. Beispielhaft sollen in Abbildung 5.22 die Qualitätsurteile für ANKER11 und ANKER12 (Block 1) verglichen werden. In diesem sogenannten Scatter-Plot sind 32 Punkte eingetragen (einer für jeden Probanden). Wird ANKER11 generell besser bewertet als ANKER12, so müssten alle Punkte oberhalb der Diagonalen liegen. Dies ist für neun Probanden nicht der Fall. Die Blöcke lassen sich also nicht mit der beschriebenen Methode ausrichten.

Es stellt sich die Frage, ob trotzdem eine Ausrichtung möglich ist. Dazu betrachtet man die Ankerbeispiele als versteckte Referenz. Die Ausrichtung (ohne Dehnung bzw. Stauchung der Skale) an einer Referenz ist denkbar. Es kann dazu jeweils ANKERx1, ANKERx2 oder ein Mittelwert beider Anker verwendet werden. Um ein Vorgehen zu wählen, soll die Beziehung der Ankerbeispiele untereinander dargestellt werden. In Abbildung 5.23 ist eine sogenannte Scatter-Plot-Matrix für die Beziehung aller Ankerbeispiele ANKERx1 der x Blöcke zu sehen. Auf Ordinate und Abszisse wird dabei jeweils das Qualitätsurteil abgetragen. In jedem Unterdiagramm werden wieder 32 Punkte für die Probanden eingezeichnet. Bewertet zum Beispiel jeder Proband ANKER11 und ANKER21 (ANKERx1 in Block 1 und Block 2) gleich, so würden alle Punkte im zweiten Feld der ersten Zeile auf einer exakten Diagonale liegen. Dies ist nicht der Fall, allerdings kann eine solche Tendenz beobachtet werden. Selbst über mehrere Blöcke hinweg bleibt das Urteil für ANKERx1 relativ konstant.

Dasselbe Diagramm kann für ANKERx2, also das zweite Ankerbeispiel, aufgestellt werden. Es ist in Abbildung 5.24 zu sehen. Auch hier ist eine diagonale Tendenz zu erkennen. Die Punkte streuen jedoch etwas mehr als für ANKERx1. ANKERx1 wurde immer zu Beginn eines Blockes gehört. Das Qualitätsurteil könnte daher gleichbleibender sein als

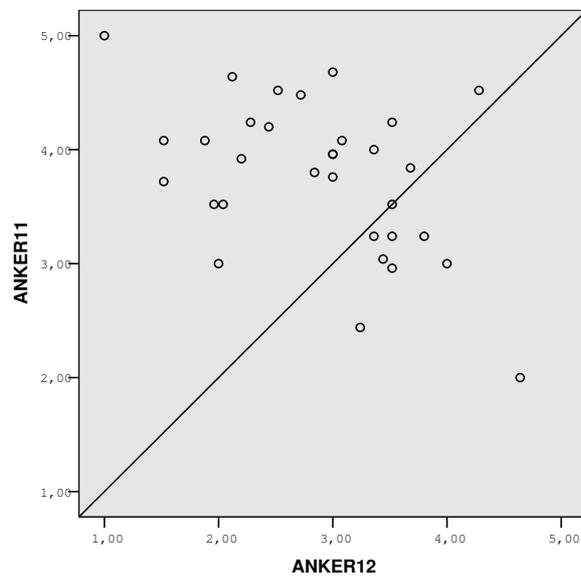


Abbildung 5.22: Beispielhafte Darstellung der Beziehung zwischen dem Urteil eines jeden Probanden zu erstem und zweitem Ankerbeispiel in Block eins. Liegt ein Punkt oberhalb der Diagonalen, so ist Ankerbeispiel eins von der Versuchsperson als qualitativ besser bewertet worden als Ankerbeispiel zwei. Liegt er unterhalb ist dies nicht der Fall.

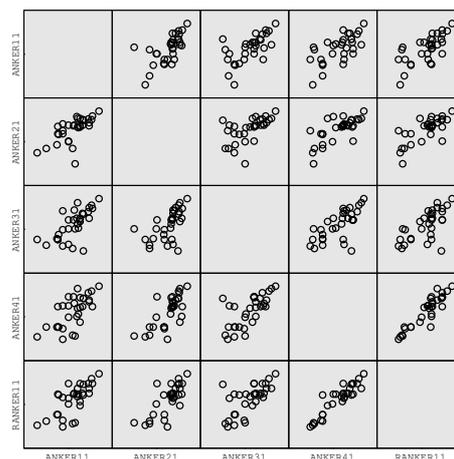


Abbildung 5.23: Eine sogenannte Scatter-Plot-Matrix, die alle möglichen Kombinationen zwischen mehreren Stichproben darstellt. Die Stichproben sind hier die jeweils ersten Ankerbeispiele aus den Blöcken eins bis vier ("ANKER11", "ANKER21", usw.) und das erste Ankerbeispiel aus dem wiederholten Block eins ("RANKER11"). Jeder Punkt repräsentiert einen Probanden.

das für ANKERx2, da dieser immer umgeben von verschiedenen Stimuli gehört wurde.

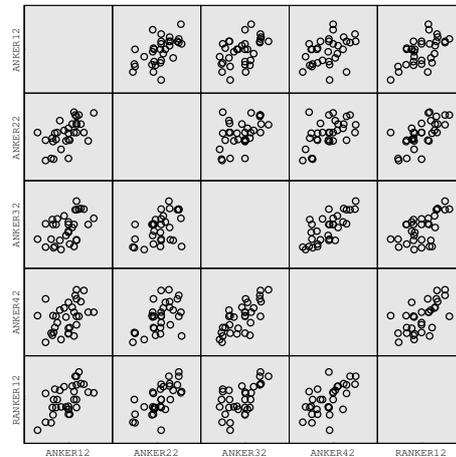


Abbildung 5.24: Eine sogenannte Scatter-Plot-Matrix, die alle möglichen Kombinationen zwischen mehreren Stichproben darstellt. Die Stichproben sind hier die jeweils zweiten Ankerbeispiele aus den Blöcken eins bis vier ("ANKER12", "ANKER22", usw.) und das zweite Ankerbeispiel aus dem wiederholten Block eins ("RANKER12").

Im Folgenden soll daher ANKERx1 als versteckte Referenz zur Ausrichtung der Blöcke verwendet werden. Alle Messwerte eines Blockes sind linear so zu verschieben, dass ANKERx1 den Wert vier annimmt. Der resultierende Error-Bar für die Ankerbeispiele ist in Abbildung 5.25 zu sehen. Es zeigt sich, dass die Streuung der Messwerte durch die Ausrichtung zusätzlich verstärkt wird. Dies wird ebenfalls bei der Darstellung des Error-Bars für die übrigen Messwerte der Blöcke eins bis vier bestätigt (Anhang F, Abbildung F.4). Die Ausrichtung scheint keine genauere Auswertung zu ermöglichen.

Die Häufigkeitsverteilungen der bearbeiteten Stichproben sind in Anhang F, Abbildung F.5 und F.6 zu sehen. Es kann nun wieder mit dem KSA-Test auf Normalverteilung geprüft werden. In Anhang F, Tabelle F.2, sind die Ergebnisse zusammengefasst. Nur zwei der Stichproben werden bei einem Schwellenwert der α -Fehlerwahrscheinlichkeit von 20% knapp nicht normalverteilt getestet. Es soll Normalverteilung angenommen werden, so dass eine varianzanalytische Betrachtung möglich ist. Die Ergebnisse der fünffaktoriellen Varianzanalyse sind in Anhang F, Tabelle F.3 zusammengefasst. Wie in der Auswertung ohne versteckte Referenz werden die beiden Haupteffekte **Instrument** und **Musiker** als sehr signifikant getestet. Der Einfluss des Instruments ist hier an der signifikanten Interaktion zwischen **Instrument, Raum und Anspiel** zu relativieren, der Einfluss des Musikers an der Interaktion **Musiker * Raum**. Zusätzlich tritt ein weiterer signifikanter Haupteffekt **Anspiel** auf. Die entsprechenden Interaktionsdiagramme sind in Anhang F, Abbildung F.7, F.8 und F.9 dargestellt. Es soll hier nur die verkürzte Auswertung in tabellarischer Form (Tabelle 5.10) dargestellt werden. Der Haupteffekt für den Faktor Instrument wird auch hier bestätigt. Die Abhängigkeit des Faktors Musiker vom Raum ist identisch. Zusätzlich tritt der Haupteffekt für den Faktor Anspiel auf, welcher allerdings an der bekannten Interaktion zwischen Anspiel, Raum und Instrument relativiert wird. Zusammenfassend ist zu sagen, dass die bisherigen Ergebnisse größtenteils bestätigt werden. Allerdings scheint die Auswertung mit Ausrichtung an ANKERx1 nicht richtigere oder genauere Erkenntnisse zu liefern.

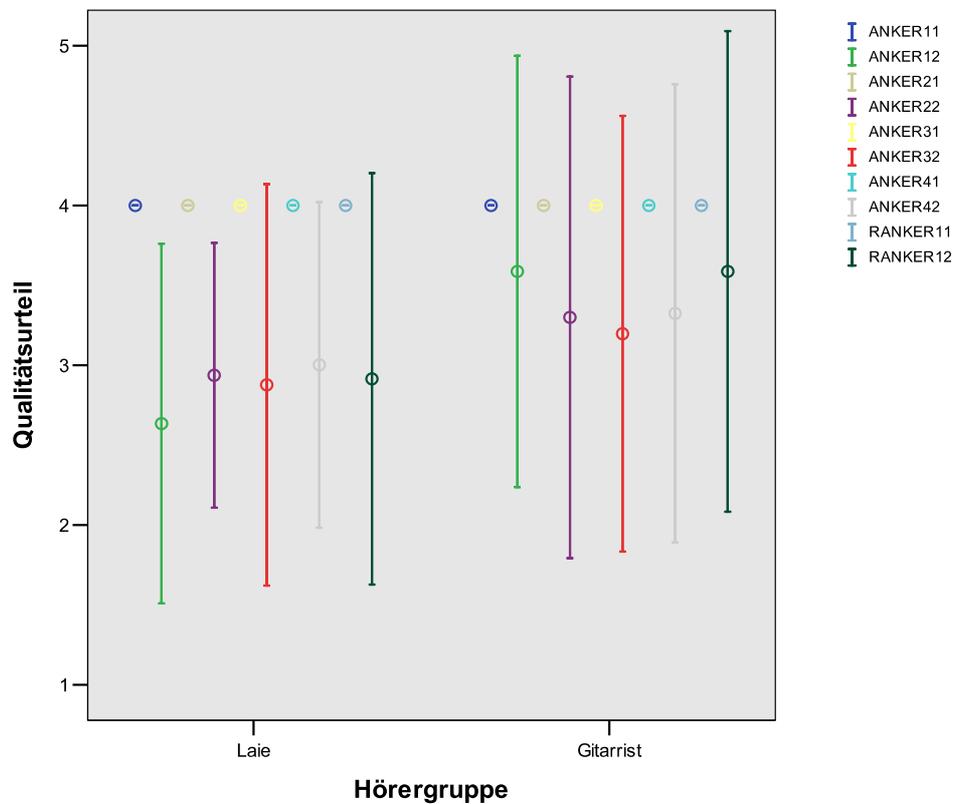


Abbildung 5.25: Error-Bar der Messwertverteilungen der Ankerbeispiele aus Hörversuch 2 für 32 Personen, 16 Laien und 16 Gitarristen nachdem die Stichproben an Ankerbeispiel "ANKERx1" ausgerichtet wurden. Dargestellt ist der Mittelwert \pm eine Standardabweichung.

Tabelle 5.10: Zusammenfassung der Auswertung von Haupteffekt und Interaktionen in Hörversuch 2 nach Ausrichtung an Referenz ANKERx1. Die einzelnen Faktoren werden abgekürzt als: Instrument I , Musiker M , Anspiel A und Raum R . Bedeutungslose Aussagen sind grau geschrieben.

Interaktion	Beurteilte Qualität des Instrumentes für	Bedingung
I	$I_1 > I_2$	
$I * R * A$	$I_1 > I_2$	
M	$M_1 > M_2$	
$M * R$	$M_1 \approx M_2$	R_1
	$M_1 > M_2$	R_2
A	$A_1 < A_2$	
$A * R * I$	$A_1 \approx A_2$	$(R_1 \vee R_2)I_1 \vee R_2I_2$
	$A_1 < A_2$	R_1I_2

Zum zusammenfassenden Vergleich sollen noch einmal die Interaktionen aller drei Verfahren in Tabelle 5.11 aufgelistet werden. Es werden zusätzlich zu den signifikanten und sehr signifikanten Interaktionen, Trends (Irrtumswahrscheinlichkeit $< 10\%$) in die Tabelle aufgenommen. Es kann zusammengefasst werden, dass das Qualitätsurteil über beide Instrumente immer abhängig vom Instrument selbst ist. Der Haupteffekt Instrument ist nur unter Berücksichtigung der Sequenz nicht eindeutig interpretierbar. Es sollten daher bei der Qualitätsbeurteilung stets Angaben zum gewählten Musikstück gemacht werden. Des Weiteren treten verschiedene Interaktionen zwischen Musiker, Anspiel, Raum, Instrument, Sequenz und Hörergruppe auf. Die Interaktionen haben oft dieselbe Größenordnung wie der Haupteffekt Instrument und müssen in weiteren Hörversuchsentwürfen berücksichtigt werden. Teilhierarchische Strukturen sind daher nur bedingt möglich. Will man mehrere Instrumente vergleichen und den Versuchsaufwand reduzieren, so sind alle Randbedingungen konstant zu halten und ausführlich zu dokumentieren. Die Ergebnisse gelten dann nur für die gewählten Randbedingungen und sollten nicht verallgemeinert werden. Im nächsten Abschnitt ist solch ein Vergleich von zehn Instrumenten unter definierten Randbedingungen dokumentiert.

Die Auswertung über mehrere Blöcke erschwert die Interpretierbarkeit der Ergebnisse und beseitigt damit den Vorteil des direkten Vergleichs von mehreren Stimuli innerhalb eines Blocks. Gelingt es, die Stimulianzahl so zu reduzieren, dass sie in einen Block passen, wäre ein genaueres Urteil denkbar.

Tabelle 5.11: Vergleich der Ergebnisse von Hörversuch 1, Hörversuch 2 und Hörversuch 2 nachdem die Stichproben an Ankerbeispiel "ANKERx1" ausgerichtet wurden. Die einzelnen Faktoren werden abgekürzt als: Instrument I , Musiker M , Anspiel A , Raum R und Hörergruppe HG . Ist ein Haupteffekt oder eine Interaktion sehr signifikant (Irrtumswahrscheinlichkeit $< 1\%$) wird sie mit ****** gekennzeichnet, ist sie signifikant (Irrtumswahrscheinlichkeit $< 5\%$) mit ***** und ist sie nicht signifikant, aber ein Trend erkennbar (Irrtumswahrscheinlichkeit $< 10\%$) wird sie ohne Stern geschrieben. Durchgestrichenen Felder können keine Interaktionen enthalten, da im Hörversuch 2 nur eine Sequenz verwendet wurde und somit der Faktor Se entfällt.

Hörversuch 1	Hörversuch 2	Hörversuch 2 nach Ausrichtung an "ANKERx1"
I^{**}	I^{**}	I^{**}
	M^{**}	M^{**}
$Se * HG$	—	—
R		
	$R * HG$	
	A	A^*
$I * M$		
$I * Se^{**}$	—	—
$M * Se^{**}$	—	—
$I * R * HG^{**}$		$I * R * HG$
$M * R^*$	$M * R^{**}$	$M * R^{**}$
$I * M * R^{**}$	$I * M * R^*$	$I * M * R$
	$I * A$	$I * A$
	$M * A^{**}$	$M * A$
$M * Se * R^{**}$	—	—
$I * M * A * HG^*$	$I * M * A * HG$	
	$R * A$	
$Se * A^{**}$	—	—
$M * Se * A^{**}$	—	—
$I * R * A^{**}$	$I * R * A^*$	$I * R * A^*$
$I * M * R * A^{**}$	$I * M * R * A$	
$Se * R * A^*$	—	—
$I * Se * R * A$	—	—
$M * Se * R * A$	—	—
$M * Se * R * A * HG$	—	—

5.3 Hörversuch 3

In einem dritten Hörversuch werden zehn Instrumente unter definierten Randbedingungen in einem Block miteinander verglichen. Wieder wurde nach der Gesamtqualität des Instrumentes gefragt.

Deskriptive Statistik

Die Bezeichnung der Stichproben lautet "AxxIBMK2", wobei an der Stelle "xx" die Nummer der Referenzgitarre aus Tabelle 4.1 einzusetzen ist. "IB" kennzeichnet den Gitarristen,

”M” die verwendete Sequenz (Melodie Para Gitarra), ”K” den Raum (Konferenzraum) und ”2” das jeweils zweite Anspiel auf der Gitarre. In Anhang G, Abbildung G.1, ist eine tabellarische Übersicht der erhobenen Messdaten zu sehen. Es werden wieder Mittelwerte und Mediane berechnet. In Abbildung 5.26 ist beispielhaft die Häufigkeitsverteilung für die Stichprobe ”A1IBMK2”, also für Referenzgitarre G1, dargestellt. Die Stichprobe liegt relativ gut auf einer Normalverteilung. Alle Häufigkeitsverteilungen sind im Anhang G, Abbildung G.2, zu finden.

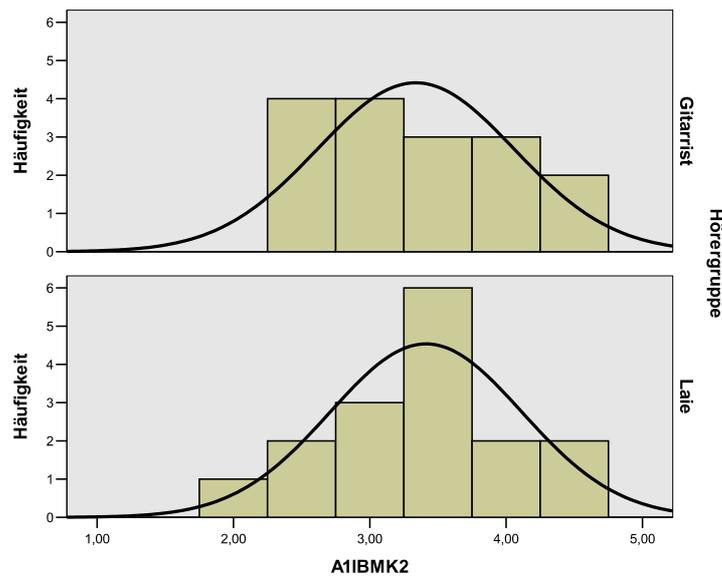


Abbildung 5.26: Häufigkeitsverteilung der Stichprobe "A1IBMK2" aus Hörversuch 3 mit überlagerter Normalverteilungskurve getrennt dargestellt für die zwei Stufen des Gruppierungsfaktors: Gitarrenisten (oben) und Laien (unten).

Die Darstellung der Daten als Error-Bar ist in Abbildung 5.27 zu sehen. Es zeigt sich, dass die Streuung wesentlich geringer als in vorangegangenen Tests ausfällt. Dass nur ein Faktor zwischen den Stimuli variiert wurde, scheint ein homogeneres Urteil der Probanden zu bewirken. Betrachtet man die bisher verwendeten Gitarren G1 und G5, in Stimuli "A1IBMK2" und "A5IBMK2", so fällt das Qualitätsurteil unter den vorliegenden Randbedingungen sehr ähnlich aus.

Analytische Statistik

Um statistisch zu überprüfen, ob Gitarren qualitativ signifikant unterscheidbar sind, soll wieder eine Varianzanalyse durchgeführt werden. Der KSA-Test bestätigt, mit einem Schwellenwert von 20% für die α -Fehlerwahrscheinlichkeit, die hinreichende Normalverteilung aller Stichproben. Das Ergebnis der zweifaktoriellen Varianzanalyse innerhalb der Versuchspersonen ist in Tabelle 5.12 zu sehen. Es besagt, dass sich mindestens ein Qualitätsurteil sehr signifikant von einem anderen unterscheidet.

In Tabelle 5.13 kann gesehen werden, dass wieder kein Einfluss des Faktors Hörergruppe existiert. Zeichnet man das Interaktionsdiagramm der nicht signifikanten Interaktion zwischen Hörergruppe und Instrument, so ergibt sich Abbildung 5.28. Die beiden Kur-

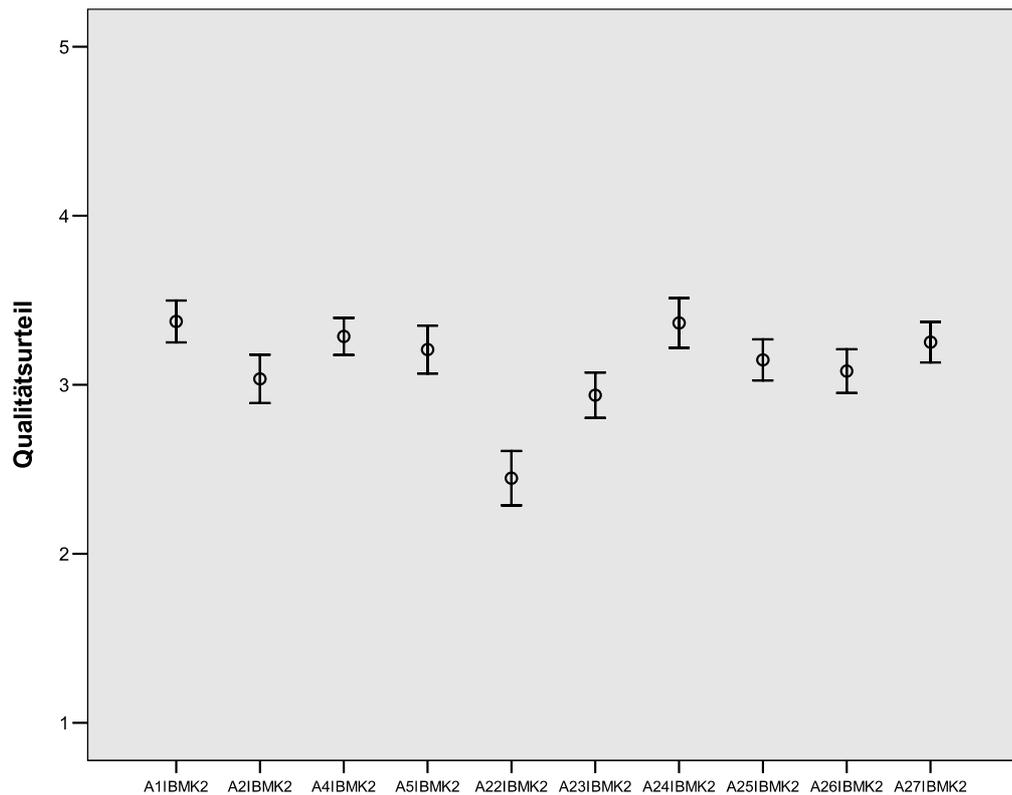


Abbildung 5.27: Error-Bar der Messwertverteilungen der erhobenen Daten aus Hörversuch 3 für 32 Personen, 16 Laien und 16 Gitarristen. Dargestellt ist der Mittelwert \pm eine Standardabweichung.

venverläufe für Laien und Gitarristen verlaufen nahezu parallel. Nur für Referenzgitarre G2 ist ein deutlicher Unterschied zu sehen.

Tabelle 5.12: Ergebnistabelle der zweifaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 3 innerhalb der Versuchspersonen. Die einzelnen Faktoren werden abgekürzt als: Instrument I und Hörergruppe HG . Es wurde immer die selbe Musiksequenz $Se1$ vom Musiker "IB" im Raum "K", das zweite Anspiel verwendet.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
I	21,534	9	2,393	5,122	0,000	0,146
$I * HG$	3,360	9	0,373	0,799	0,617	0,026
Error(I)	126,127	270	0,467			

Tabelle 5.13: Ergebnistabelle der zweifaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 3 zwischen den Versuchspersonen. HG bezeichnet den Faktor Hörergruppe.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
HG	0,068	1	0,068	0,043	0,838	0,001
Error	48,272	30	1,609			

Es interessiert nun, welche Gitarren sich in ihrem Qualitätsurteil signifikant voneinander unterscheiden. Dazu soll nach [Bortz, 2005] ein Scheffé-Test zum paarweisen Vergleich

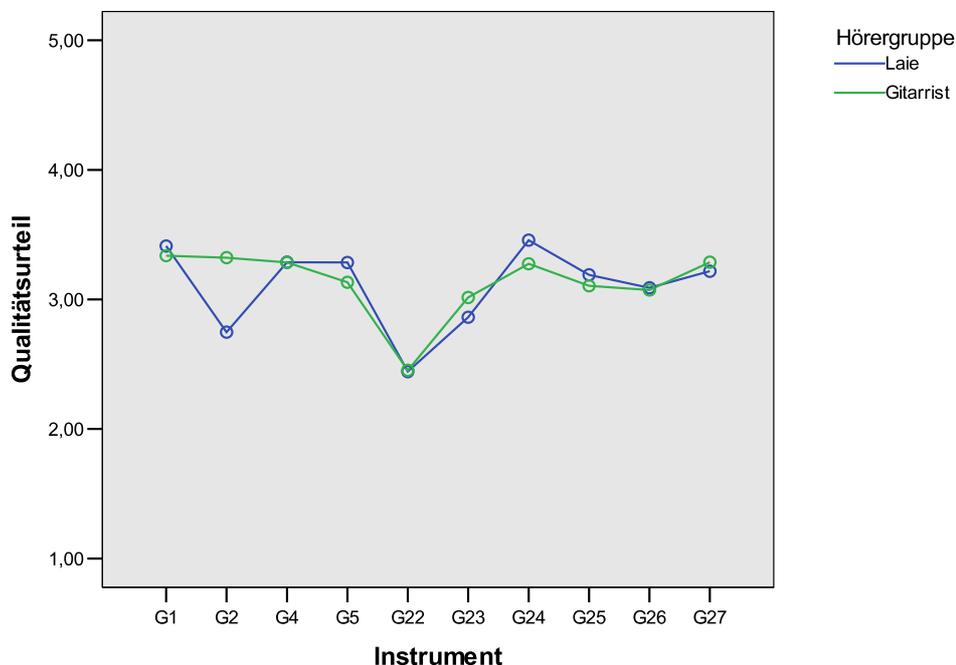


Abbildung 5.28: Interaktionsdiagramm für nicht signifikante Interaktion zwischen Instrument und Hörergruppe in Hörversuch 3.

der Gitarren durchgeführt werden. Die benötigte kritische Differenz ist auch für den Versuchsentwurf mit Messwiederholung berechenbar, allerdings nicht mit dem Programm SPSS. Da ein Versuchsplan mit Messwiederholung über einen Wiederholungsfaktor auch als Versuchsplan ohne Messwiederholung dargestellt werden kann, soll dies hier geschehen, um dann den Scheffé-Test für unabhängige Messwerte durchzuführen. Dazu müssen die Daten ipsativiert werden, das heißt, von allen Messwerten einer Versuchsperson ist der entsprechende Mittelwert abzuziehen. Der Scheffé-Test fordert außerdem homogene Fehlervarianzen. Dies soll hier mit dem Levene-Test ([Bortz, 2005] verweist für eine Beschreibung auf [Dayton, 1970]) geprüft werden. Das Ergebnis ist in Tabelle 5.14 dargestellt. Die Annahme homogener Fehlervarianzen wird bestätigt.

Tabelle 5.14: Levene-Test auf homogene Fehlervarianzen. Nach dem SPSS-Results-Coach wird bei einem Signifikanzniveau größer 0,05, wie hier der Fall, die Homogenität der Fehlervarianzen bestätigt.

F	df1	df2	Sig.
1,462	19	300	0,097

Das Ergebnis des Scheffé-Tests für unabhängige Messwerte ist in Anhang G, Tabelle G.2, dargestellt. An dieser Stelle sind zur besseren Übersicht nur signifikante und sehr signifikante Unterschiede in Tabelle 5.15 zusammengefasst. Es ist zu sehen, dass sich unter den gegebenen Randbedingungen nur Gitarre G22 von den meisten Gitarren mit

einer Mittelwertdifferenz zwischen 0,7 und etwa 0,9 unterscheidet. Die in Hörversuch eins und zwei untersuchten Gitarren G1 und G5, unterscheiden sich qualitativ unter diesen Bedingungen nicht.

Tabelle 5.15: Mit dem Scheffé Test wurde überprüft, welche Instrumente sich signifikant voneinander unterscheiden. Für signifikante Unterschiede ist die jeweilige Differenz der Mittelwerte angegeben und mit einem * versehen. Für sehr signifikanten Unterschiede ist die Mittelwertdifferenz durch ** gekennzeichnet.

Gitarre	1	2	4	5	22	23	24	25	26	27
1					0,93**					
2										
4					0,84**					
5					0,76**					
22	0,93**		0,84**	0,76**			0,92**	0,70*		0,81**
23										
24					0,92**					
25					0,70*					
26										
27					0,81**					

Zusammenfassung und Ausblick

Es wurden Hörversuche zur qualitativen Beurteilung von ausgewählten Musikinstrumenten und Tonsequenzen entworfen und durchgeführt. Als Randparameter wurden dabei zusätzlich der Raum, der Musiker, das Anspiel und die Hörergruppe untersucht.

Die gewählte binaurale Aufnahmetechnik in Kombination mit der Kompensation für die Charakteristik der Wiedergabekopfhörer scheint gute Ergebnisse zu liefern. Die meisten Probanden bestätigten einen realistischen Höreindruck und zeigten keine akustischen Ermüdungserscheinungen, wenige berichteten, das Instrument manchmal hinter sich wahrzunehmen. Ob die Verwendung von Kopfhörern das Qualitätsurteil für ein Instrument beeinflusst, kann nur durch einen direkten Vergleich ohne Kopfhörer nachgewiesen werden.

Die vorliegende Arbeit umfasst zwei verschiedene Hörversuchsmethoden zur subjektiven Beurteilung von Musikinstrumenten, eine serielle und eine in Blöcken. Der jeweils gewählte Hörversuchsentwurf scheint einen Einfluss auf das Qualitätsurteil der Probanden zu haben. Die Beurteilung von Instrumenten in mehreren Blöcken bringt keine Zeiterparnis, aber einige Interpretationsprobleme mit sich, so dass bei einer großen Anzahl von Stimuli der serielle Versuchsentwurf zu bevorzugen ist. Kann man die Anzahl der Stimuli durch definierte und feste Randbedingungen begrenzen, so dass sie in einen Block passen, bevorzugen Probanden den direkteren Vergleich in Blöcken. Die gewählten Skalen (fünfstufig bzw. quasikontinuierlich) scheinen für die jeweilige Methode angemessen zu sein. Der Intervallabstand zwischen den verbalen Ankerbezeichnungen (ausgezeichnet, gut, ordentlich, dürftig, schlecht) wurde hier als äquidistant angenommen. Um diese Annahme zu unterstützen, wird empfohlen in weiteren Experimenten zusätzlich Zahlenwerte anzugeben.

Die akustische Qualitätswahrnehmung eines Instrumentes ist vom Instrument selbst, aber zusätzlich von fast allen untersuchten Randparametern abhängig. Unter den meisten Randbedingungen wird in Hörversuch eins und zwei Referenzgitarre G1 etwas besser bewertet als Referenzgitarre G5. Ob der geringe Unterschied praktische Bedeutung hat, ist offen. Die verwendete Tonsequenz hat einen großen Einfluss auf das Qualitätsurteil. Zwischen den unabhängigen Variablen existieren eine Reihe von Interaktionen. Es ist vorstellbar, dass noch weitere Einflussgrößen (wie zum Beispiel die Besaitung) existieren, die bisher konstant gehalten und damit nicht explizit berücksichtigt wurden.

Die statistische Auswertung der Ergebnisse zeigt, dass sowohl Laien wie auch Gitarristen die akustischen Gesamtqualität der hier untersuchten Instrumente im Mittel gleich beurteilen. Dennoch streuen die Messwerte stark, was auf individuelles Qualitätsempfinden hinweisen kann. Die Qualitätsbeurteilung wurde von den Probanden selten als schwierig eingestuft, allerdings variierte die Fähigkeit konkrete Qualitätsmerkmale zu benennen oder zu beschreiben. Durch gezieltere Fragen nach einzelnen Qualitätsmerkmalen ließe sich die Wahrnehmung lenken und es könnte sich die Streuung in den Messergebnissen verringern. Allerdings verringert sich dadurch auch die externe Validität und es erhöht sich der Versuchsaufwand. Da sich gezeigt hat, dass hierarchische Versuchsstrukturen auf Grund einer Vielzahl von Interaktionen zwischen den unabhängigen Variablen nicht oder nur sehr beschränkt möglich sind, ist dadurch keine wesentliche Verringerung des Versuchsaufwands möglich.

In Hörversuch drei wurden zehn Instrumente unter konstanten Rahmenbedingungen verglichen. Obwohl die Streuung der Messwerte abnimmt, hebt sich nur ein Instrument (Referenzgitarre G22) klanglich qualitativ von den meisten anderen signifikant ab. Das zwischen den anderen kein signifikanter Qualitätsunterschied getestet wird, bedeutet nicht, dass generell keine Unterschiede zwischen den Instrumenten wahrnehmbar sind. Dies wäre in Folgeuntersuchungen nachzuweisen.

Die beschriebenen Interaktionen verdeutlichen die Komplexität der menschlichen Qualitätswahrnehmung. Objektive Methoden zur Abschätzung der wahrgenommenen qualitativen akustischen Differenzierung zwischen Instrumenten können daher relativ umfangreich werden. Sie sind immer durch sorgfältige subjektive Experimente zu bestätigen.

Literaturverzeichnis

- [BK4153, 2005] *Product Data Artificial Ears — Types 4152 and 4153* (2005). Brüel und Kjaer. als PDF-Datei auf CD beigelegt.
- [Bortz, 2005] Bortz, J. *Statistik für Human- und Sozialwissenschaftler* (2005). Springer Medizin Verlag. sechste Auflage. ISBN 3-540-21271-X.
- [Bortz and Döring, 2001] Bortz, J. and Döring, N. *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler* (2001). Springer-Verlag. dritte Auflage. ISBN 3-540-41940-X.
- [Bortz et al., 1990] Bortz, J., Lienert, G. A., and Boehnke, K. *Verteilungsfreie Methoden in der Biostatistik* (1990). Springer-Verlag. ISBN 3-540-50737-X.
- [Boullosa, 1997] Boullosa, R. *Subjective Evaluation Of Classical Guitars* (1997). Centro de Instrumentos UNAM. Popular version of paper 2aMU10, 134th ASA Meeting, San Diego, CA.
- [BS1116-1, 1997] *ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems* (1997). International Telecommunications Union - Radiocommunication Assembly. als PDF-Datei auf CD beigelegt.
- [BS1284-1, 2003] *ITU-R BS.1284-1: General methods for the subjective assessment of sound quality* (2003). International Telecommunications Union - Radiocommunication Assembly. als PDF-Datei auf CD beigelegt.
- [BS1534-1, 2003] *ITU-R BS.1534-1: Method for the subjective assessments of intermediate quality level of coding systems* (2003). International Telecommunications Union - Radiocommunication Assembly. als PDF-Datei auf CD beigelegt.
- [Cochran and Cox, 1957] Cochran, W. G. and Cox, G. M. *Experimental Designs* (1957). Wiley Classics Library. zweite Auflage. ISBN 0-471-54567-8.
- [Dayton, 1970] Dayton, C. M. *The design of educational experiments* (1970). McGraw-Hill.
- [DIN 55350, 1987] *DIN 55350, Begriffe der Qualitätssicherung und Statistik. Grundbegriffe der Qualitätssicherung.* (1987). Teil 11, Nr. 1.

- [Emerson and Stoto, 1983] Emerson, J. D. and Stoto, M. A. *Transforming Data, In Understanding Robust and Exploratory Data Analysis* (1983). John Wiley and Sons.
- [Everitt and Hay, 1992] Everitt, B. and Hay, D. *Talking about Statistics* (1992). Edward Arnold. ISBN 0-470-21956-4.
- [Franco et al., 2004] Franco, A. F., Merchel, S., Pesqueux, L., Rouaud, M., and Sørensen, M. O. *Sound Reproduction By Wave Field Synthesis* (2004). Aalborg University. Faculty of Engineering and Science - Department of Acoustics.
- [Jekosch, 2000] Jekosch, U. *Sprache hören und beurteilen: Ein Ansatz zur Grundlegung der Sprachqualitätsbeurteilung* (2000). Habilitationsschrift am Fachbereich 3 Literatur- und Sprachwissenschaften der Universität/Gesamthochschule Essen.
- [Meilgaard et al., 1999] Meilgaard, M., Civile, G. V., and Carr, T. *Sensory Evaluation Techniques* (1999). CRC Press LLC. dritte Auflage. ISBN 0-8493-0276-5.
- [Minnaar et al., 2001] Minnaar, P., Olesen, S. K., Christensen, F., and Møller, H. *Localization with Binaural Recordings from Artificial and Human Heads* (May 2001). J. Audio Eng. Soc., Vol. 49, No. 5, pp. 323-336.
- [Møller, 1992] Møller, H. *Fundamentals of Binaural Technology* (1992). Applied Acoustics, Vol. 36, pp. 171-218.
- [Møller et al., 1995] Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. *Transfer Characteristics of Headphones Measured on Human Ears* (April 1995). J. Audio Eng. Soc., Vol. 43, No. 4, pp. 203-217.
- [Møller et al., 1999] Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. *Evaluation of Artificial Heads in Listening Tests* (March 1999). J. Audio Eng. Soc., Vol. 47, No. 3, pp. 83-100.
- [MLSSA, 2001] *Maximum-Length Sequence System Analyzer Reference Manual - Version 10WI-4* (2001). DRA Laboratories. als PDF-Datei auf CD beigefügt.
- [Moore, 2003] Moore, B. C. J. *An Introduction to the Psychology of Hearing* (2003). Academic Press. fünfte Auflage. ISBN 0-12-505628-1.
- [NOKIA-Slides, 2003] *Subjective Evaluation* (2003). Nokia. als PDF-Datei auf CD beigefügt.
- [Oppenheim and Schaffer, 1999] Oppenheim, A. V. and Schaffer, R. W. *Zeitdiskrete Signalverarbeitung* (1999). R. Oldenburg Verlag. dritte Auflage. ISBN 3-486-24145-1.
- [Rife and Vanderkooy, 1998] Rife, D. D. and Vanderkooy, J. *Transfer-Function Measurements with Maximum-Length Sequences* (Juni 1998). J. Audio Eng. Soc., Vol. 37, pp. 419-444.
- [STAX, 2005] *Bedienungsanleitung SR-Lambda Serie; Elektrostatischer Ohrlautsprecher SR-404 Signature, SR-303 Classic, SR-202 Basic* (2005). STAX Germany.

Anhang

Die Messmethode MLSSA

Zur Frequenzgangbestimmung der Kopfhörer wurde ein "Maximum-Length Sequence System Analyzer" (MLSSA) benutzt, mit dem Impulsantwortmessungen für lineare Systeme möglich sind. Die Messmethode soll hier angelehnt an [Franco et al., 2004] und [MLSSA, 2001] kurz erläutert werden. Eine ausführliche Diskussion der MLS-Methode findet man in [Rife and Vanderkooy, 1998]. Zur Veranschaulichung der Funktionsweise soll manuell die Antwort eines mit einer MLS-Sequenz angeregten Systems berechnet werden. Aus dieser MLS-Antwort ist dann wieder die Impulsantwort zu extrahieren.

Das Messgerät sendet ein pseudozufälliges Signal, eine sogenannte "Maximum-Length Sequence" (MLS), aus und zeichnet die Antwort des zu messenden Systems wieder auf. Das aufgezeichnete Signal kann als Faltung aus dem gesendeten MLS-Signal und der Impulsantwort des Systems betrachtet werden:

$$S[n] = MLS[n] * IR[n] \quad (\text{A.1})$$

wobei $S[n]$ das gemessene Signal, $MLS[n]$ das MLS-Signal und $IR[n]$ die Impulsantwort des Systems darstellen.

Die verwendete ML-Sequenz besteht nur aus den diskreten Werten 1 und -1 und hat keinen Gleichanteil. Sie ist konstruiert um spezielle Eigenschaften zu erfüllen, die sie für Lautsprecher oder Kopfhörermessungen vorteilhaft machen. Einige dieser Eigenschaften sind:

1. Hoher Energiegehalt
2. Flaches Frequenzspektrum außer bei 0 Hz
3. Einfache Berechnung der Impulsantwort

Im folgenden sollen diese Eigenschaften näher erläutert werden.

A.1 Eigenschaft 1: Hoher Energiegehalt

Um einen hohen Signal-Rausch-Abstand zu erreichen ist es wichtig ein Testsignal mit hohem Energiegehalt relativ zu seinem Spitzenwert zu verwenden. Ein Wert der diese

Eigenschaft kennzeichnet ist der Crestfaktor. Er bezieht den Spitzenwert eines Signals auf seinen RMS-Wert:

$$\text{Crestfaktor} = \frac{\text{Spitzenwert}}{\text{RMS}} \quad (\text{A.2})$$

Der Crestfaktor kann minimal 1 werden. Dies ist der Fall für das MLS-Signal, da es nur die Werte 1 und -1 annehmen kann. Es wird somit der gesamte Dynamikbereich maximal ausgenutzt. Zum Vergleich, ein Sinussignal hat eine Crestfaktor von $\sqrt{2}$ und ein normalverteiltes Rauschen hat in etwa einen Crestfaktor von 8. Wenn für die Messung ein Impuls $\delta[n]$ verwendet würde, welcher auf Grund seines niedrigen RMS-Wertes einen sehr hohen Crest-Faktor besitzt, so wäre der Signal-Rausch-Abstand wesentlich geringer.

A.2 Eigenschaft 2: Flaches Frequenzspektrum

Ein MLS-Signal wird konstruiert um ein komplett flaches Frequenzspektrum zu besitzen, wenn man die komplette Länge des Signals betrachtet. Es ist kein zufälliges Signal, wie etwa weißes Rauschen. Bei weißem Rauschen ergibt sich das flache Spektrum durch Mittelung.

A.3 Eigenschaft 3: Einfache Berechnung der Impulsantwort

Die Impulsantwort des Systems kann auf mehrere Arten berechnet werden. Typischerweise geschieht das mit Hilfe der Kreuzkorrelation aus dem gemessenen Signal und dem originalen MLS-Signal. Die Impulsantwort kann auch, wie aus Gleichung A.1 folgt, durch die inverse Faltung des originalen MLS-Signals vom gemessene Signal erhalten werden. Die Kreuzkorrelation gibt das selbe Ergebnis und ist schneller zu berechnen.

Da beide Rechenoperationen einer Division im Frequenzbereich entsprechen, sollte der Frequenzgang der Impulsantwort im Betrag der Fouriertransformation des gemessenen Signals zu erkennen sein.

Um das Messverfahren zu veranschaulichen, soll im folgenden ein einfacher Test in Matlab durchgeführt werden.

Eine original gemessene Impulsantwort für eine Kopfhörerkapsel ist in Abbildung A.1 zu sehen. Sie wurde zur Veranschaulichung auf 2000 Abtastwerte beschränkt.

Da in MLSSA nur die berechnete Impulsantwort abgespeichert werden kann und nicht das gemessene Signal (also die Antwort des Systems auf die Anregung mit einer ML-Sequenz), soll von dieser Impulsantwort ausgegangen werden. Normalerweise berechnet das MLS-Messsystem die Kreuzkorrelation aus MLS-Anregung $MLS[n]$ und gemessener Antwort des Systems $S[n]$ und gibt nur die resultierende Impulsantwort aus. Das gemessene Signal $S[n]$ erhält man durch erneute Faltung der Impulsantwort mit einem MLS-Signal. Dieses MLS-Signal soll in Matlab generiert werden und hat eine Länge von $2^{15} - 1 = 32767$ diskreten Werten. Man bezeichnet es dann als MLS-Signal 15ter Ordnung. Die aus der Faltung resultierende Antwort wird wieder auf $2^{15} - 1$ Abtastwerte beschränkt und entspricht der gemessenen Antwort $S[n]$ auf das MLS-Signal. Sie ist in Abbildung A.2 zu sehen.

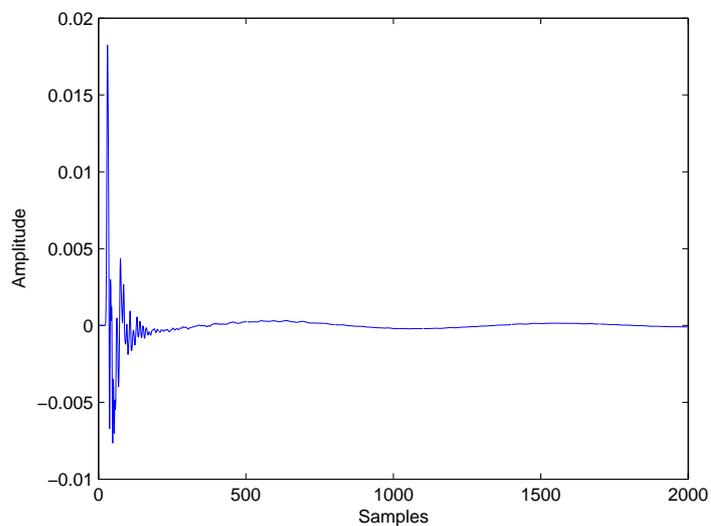


Abbildung A.1: Die originale Impulsantwort des Kopfhörers.

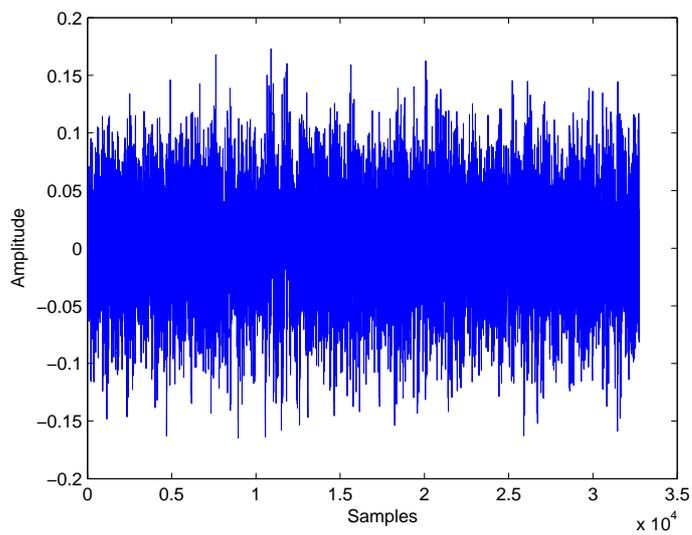


Abbildung A.2: Die gemessene Antwort $S[n]$ auf das MLS-Signals.

Um nun die Impulsantwort zu erhalten wird die Kreuzkorrelation zwischen gemessenem Signal und generiertem MLS-Signal errechnet. Sie ist in Abbildung A.3, wieder beschränkt auf 2000 Abtastwerte, zu sehen.

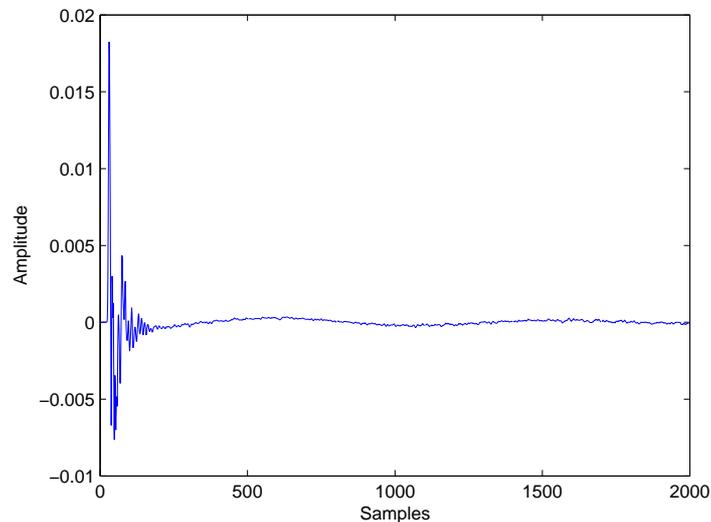


Abbildung A.3: Die berechnete Impulsantwort.

Die Fouriertransformation der originalen und der errechneten Impulsantwort sind zum Vergleich in Abbildung A.4 dargestellt. Es ist zu sehen, dass beide Frequenzgänge fast identisch aufeinander fallen, jedoch kleine Abweichungen auf Grund des verkürzten gemessenen Signals auftreten. Das zu messende Eingangssignal wird nur während der Wiedergabe des MLS-Signals aufgezeichnet, wodurch dieser Fehler entsteht.

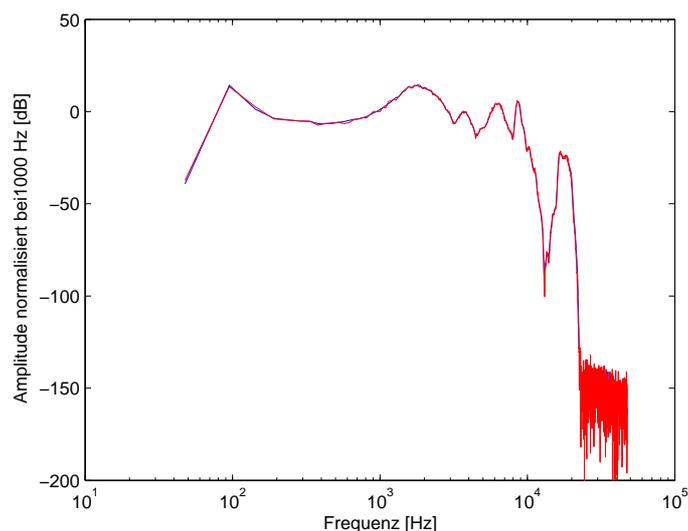


Abbildung A.4: Fouriertransformation der zwei Impulsantworten aus Abbildung A.1 und Abbildung A.3.

Betrachtet man das real gemessene Signal im Frequenzbereich so ist, wie bereits erläutert,

schon dort der Frequenzgang der Impulsantwort des zu messenden Systems erkennbar. In Abbildung A.5 ist die Fouriertransformation des MLS-Signals (flacher Frequenzgang) sowie des gemessenen Signals zu sehen.

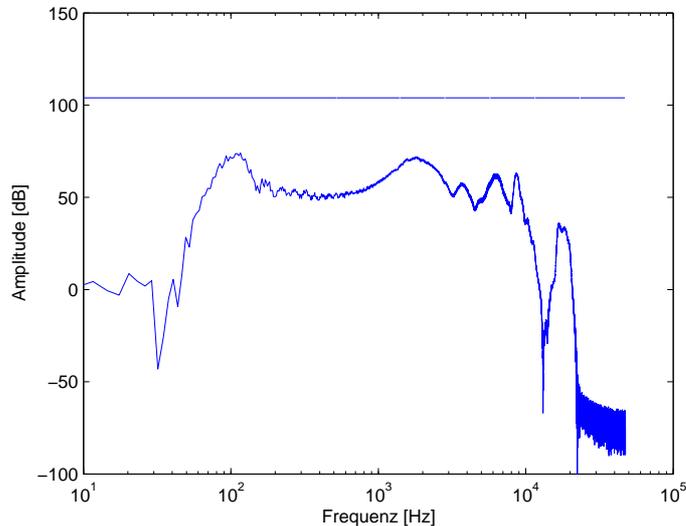


Abbildung A.5: Fouriertransformation des originalen MLS-Signals und des gemessenen Signals.

A.4 Nichtlinearitäten

Nichtlinearitäten im zu messenden System resultieren in rauschähnlichen Störungen der Impulsantwort. Nach [MLSSA, 2001] wird dieses Rauschen als "fixed-pattern noise" bezeichnet und kann für kleine Nichtlinearitäten vernachlässigt werden. Werden die Nichtlinearitäten zu groß, z.B. durch Begrenzungseffekte, Verzerrungen im Verstärker oder übersteuern der Kopfhörer, so wird der Effekt dominanter.

Durch Mittelung kann "fixed-pattern noise" nicht reduziert werden. Wenn dieser Effekt auftritt, so sollte der Versuch mit geringerer Amplitude wiederholt werden, um den elektroakustischen Wandler in seinem linearen Arbeitsbereich zu betreiben.

Wenn das zu messende System Nichtlinearitäten aufweist, so kann ein herkömmliches Messverfahren mit sinusoidalen Sweeps als Anregung besser geeignet sein, um die Impulsantwort zu bestimmen. Dazu sei vermerkt, dass eine Impulsantwort per Definition nur ein lineares System korrekt charakterisiert.

Anhang B

Kopfhörmessung

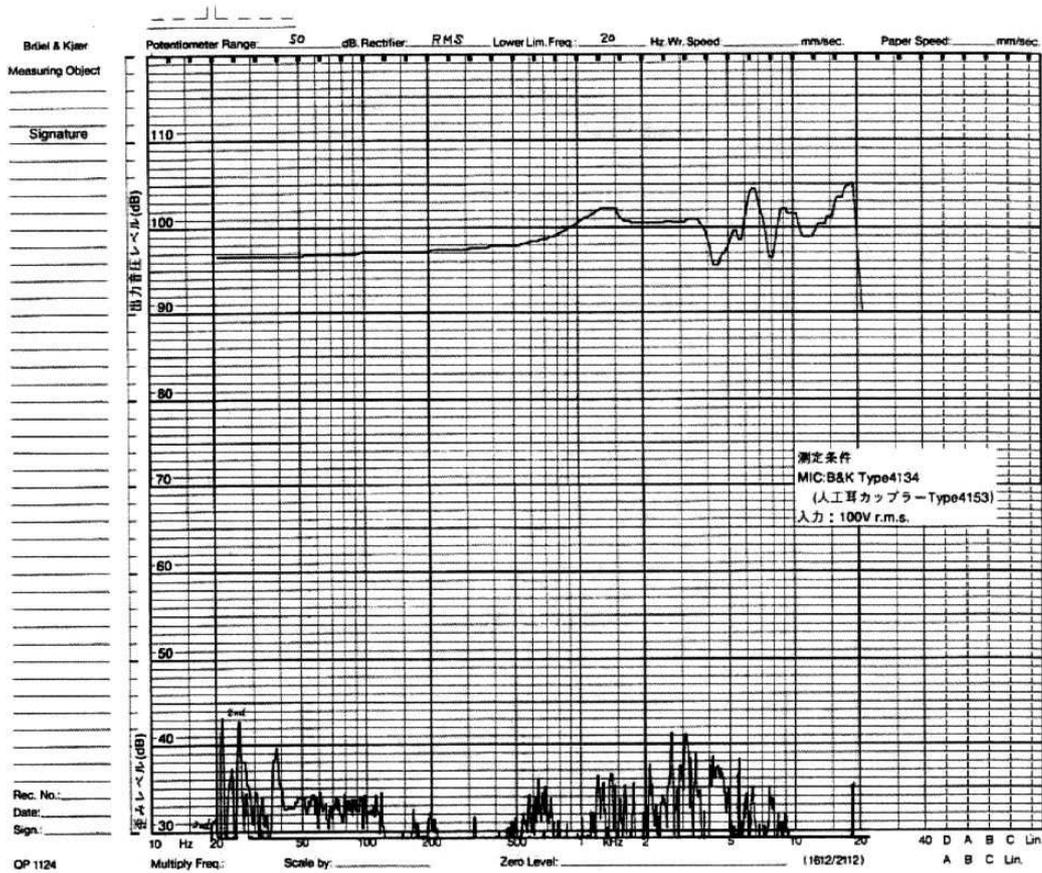


Abbildung B.1: Vom Kopfhörerhersteller STAX gemessener Frequenzgang des Kopfhörers SR-202 Basic [STAX, 2005].

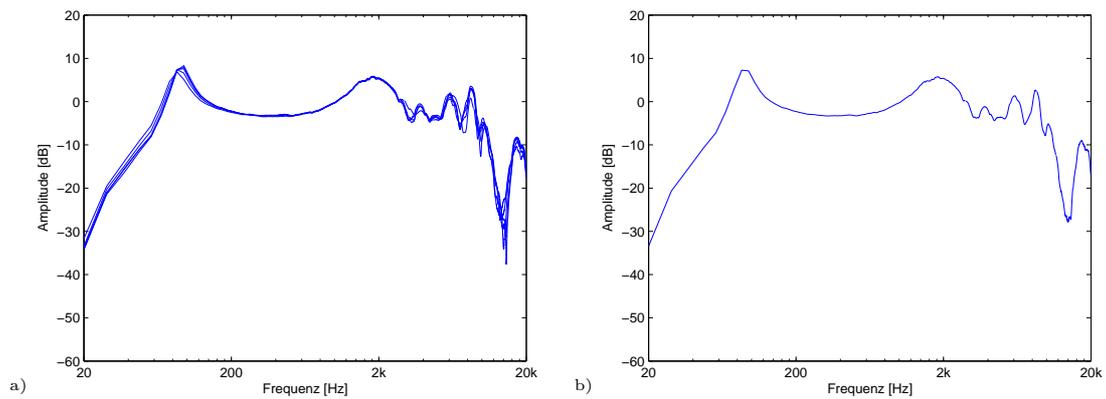


Abbildung B.2: Frequenzgang (MLSSA) für die linke Seite des Kopfhörers KH2 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

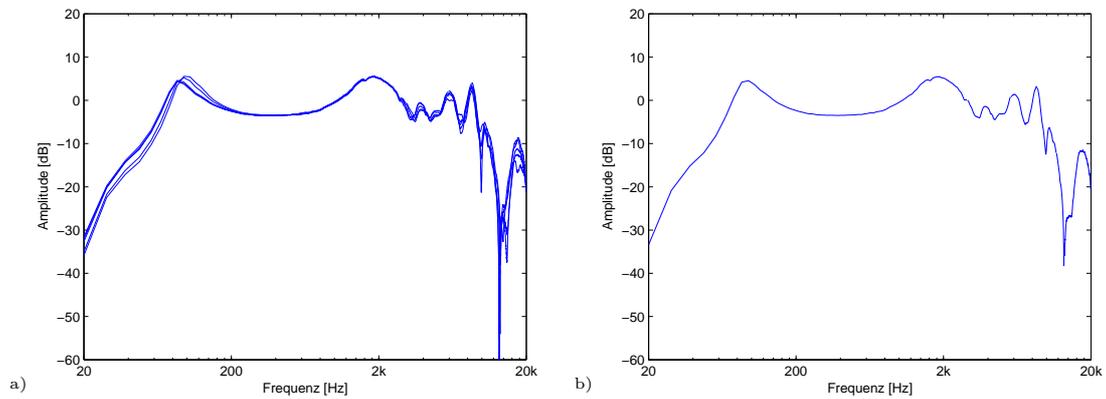


Abbildung B.3: Frequenzgang (MLSSA) für die rechte Seite des Kopfhörers KH2 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

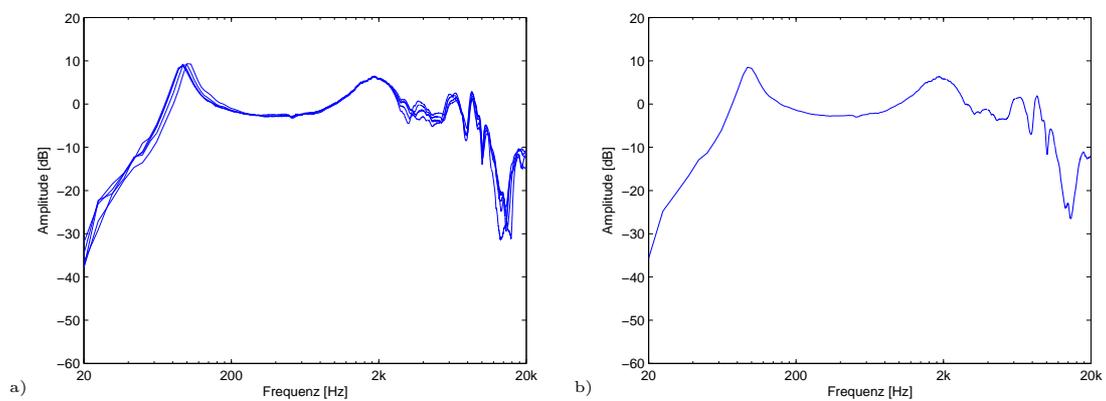


Abbildung B.4: Frequenzgang (ONO) für die linke Seite des Kopfhörers KH2 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

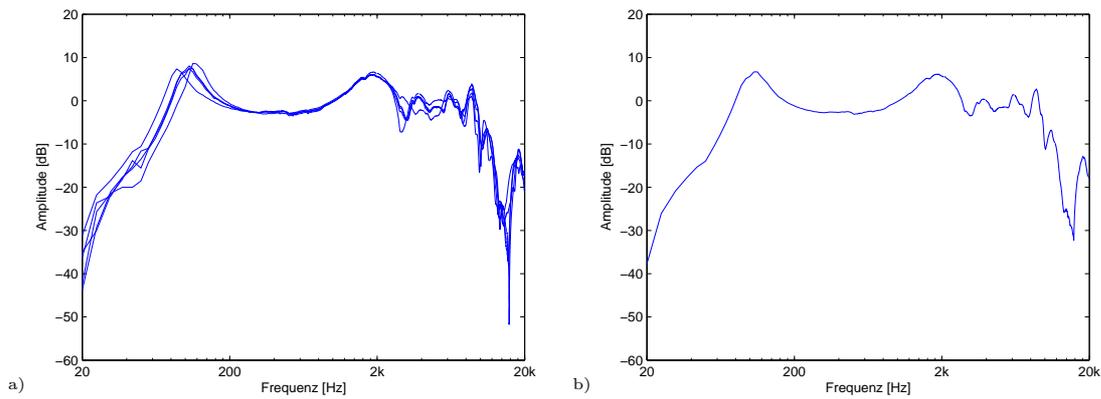


Abbildung B.5: Frequenzgang (ONO) für die rechte Seite des Kopfhörers KH2 a) gemessen für fünf Kopfhörerpositionen, b) gemittelt in dB.

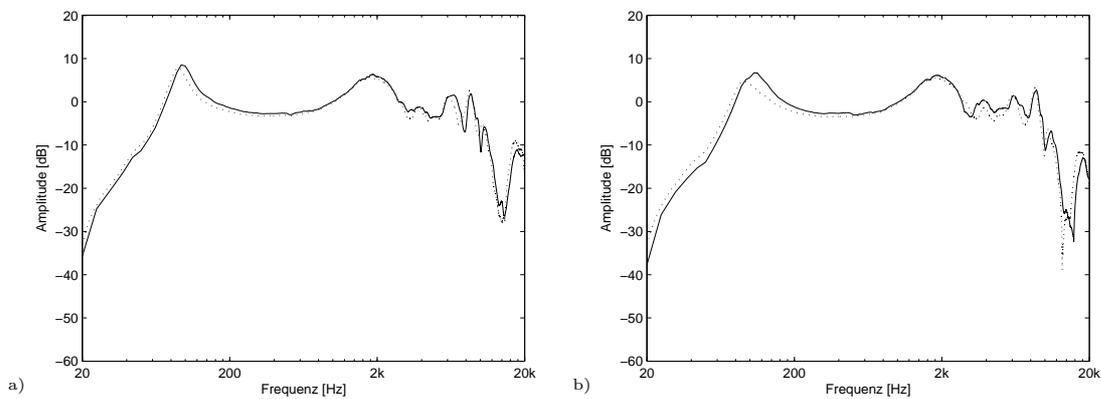


Abbildung B.6: Vergleich der über fünf Positionen gemittelten Frequenzgänge gemessen mit dem System MLSSA (schwarz) und ONO (gestrichelt) für den Kopfhörers KH2 a) linke Seite, b) rechte Seite.

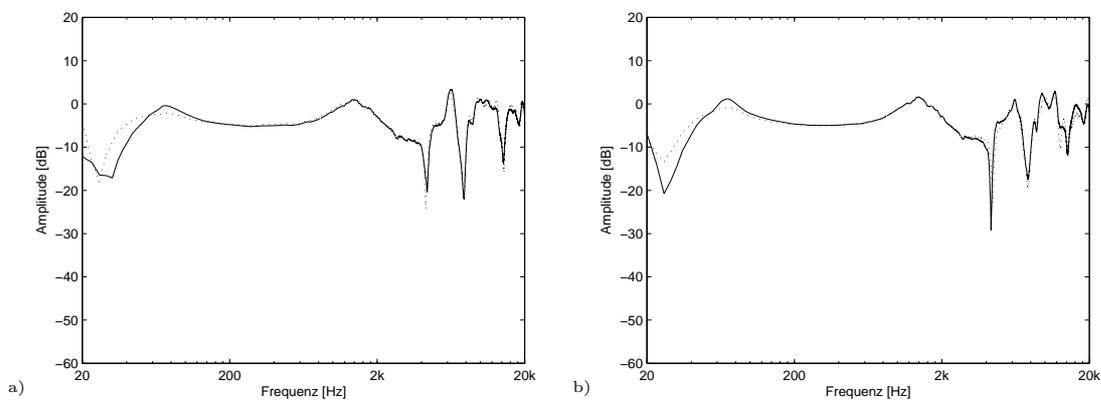


Abbildung B.7: PTFs für Kopfhörer KH2 gemessen mit dem künstlichen Ohr B&K 4153 a) linke Seite, b) rechte Seite.

Anhang C

Notenbilder der drei verwendeten Sequenzen

♩ = 80

Gtr I

T
A
B

0 1 2 3 4 0 1 2 | 3 4 0 1 2 3 4 0 | 1 2 3 0 1 2 3 4

V IX V I

T 0 1 2 3 4 5 6 7 | 8 9 10 11 12 11 10 9 | 8 7 6 5 4 3 2 1

T
A
B

0 4 3 2 1 0 3 2 | 1 0 4 3 2 1 0 4 | 3 2 1 0 4 3 2 1 0

Abbildung C.1: Notenbild der Musiksequenz 1: "Tonleiter".

Slow Blues ♩ = 80 (♩-♩-♩)

The musical score is presented in three systems, each consisting of a vocal line and a guitar accompaniment line. The tempo is marked as 'Slow Blues' with a quarter note equal to 80 beats per minute. The time signature is 4/4. The key signature has one sharp (F#).

System 1: The vocal line begins with a quarter rest, followed by a quarter note G4, a quarter note A4, and a half note B4. The guitar accompaniment (Gtr I) starts with a quarter rest, followed by a quarter note G2, a quarter note A2, and a half note B2. The guitar staff shows fret numbers 0, 1, 3, 0, 3, 5, 8 and picking notation 'H'.

System 2: The vocal line continues with a quarter note C5, a quarter note B4, and a half note A4. The guitar accompaniment continues with a quarter note G2, a quarter note A2, and a half note B2. The guitar staff shows fret numbers 5, 6, 8, 0, 1, 3, 0, 4, 3, 2 and picking notation 'H'.

System 3: The vocal line concludes with a quarter note G4, a quarter note A4, and a half note B4. The guitar accompaniment concludes with a quarter note G2, a quarter note A2, and a half note B2. The guitar staff shows fret numbers 0, 1, 3, 0, 1, 2, 0 and picking notation 'H'.

Abbildung C.2: Notenbild der Musiksequenz 2: "Blues".

♩ = 70

1

Gtr I
let ring

T 5 5 5 5 5 5 5 5
A 9 5 9 5 9 5 9 5
B 2 4 0 0 4 0 3

sl.

3

let ring

T 5 5 7 5 5
A 9 5 9 5 9 5 9 5
B 2 4 0 3 4 0 0

sl.

5

let ring

T 0 3 2 0
A 0 3 0 0 0 0
B 2 3 3 3 0 1 1

sl.

7

let ring

T 7 0 3 7 0
A 7 2 7 0 7 0
B 7 0 7 3 7 7

Abbildung C.3: Notenbild der Musiksequenz 3: "Para Guitarra".

Anhang **D**

Screenshots der Benutzeroberflächen für Hörversuche

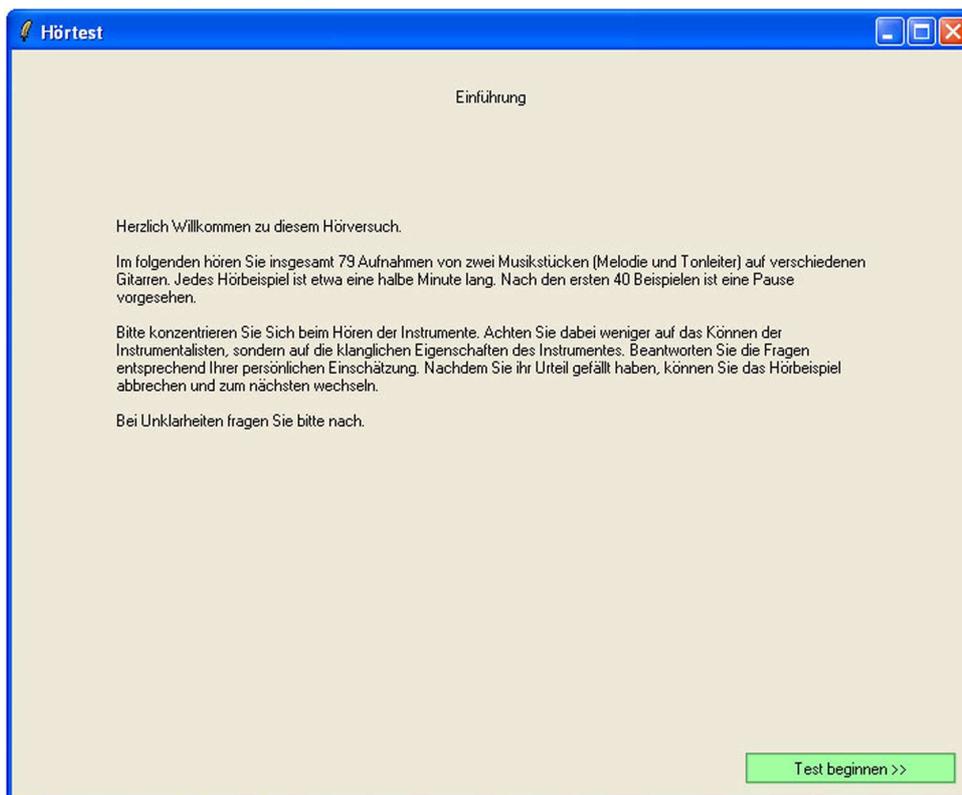


Abbildung D.1: Screenshot Einleitung für Hörversuch 1.

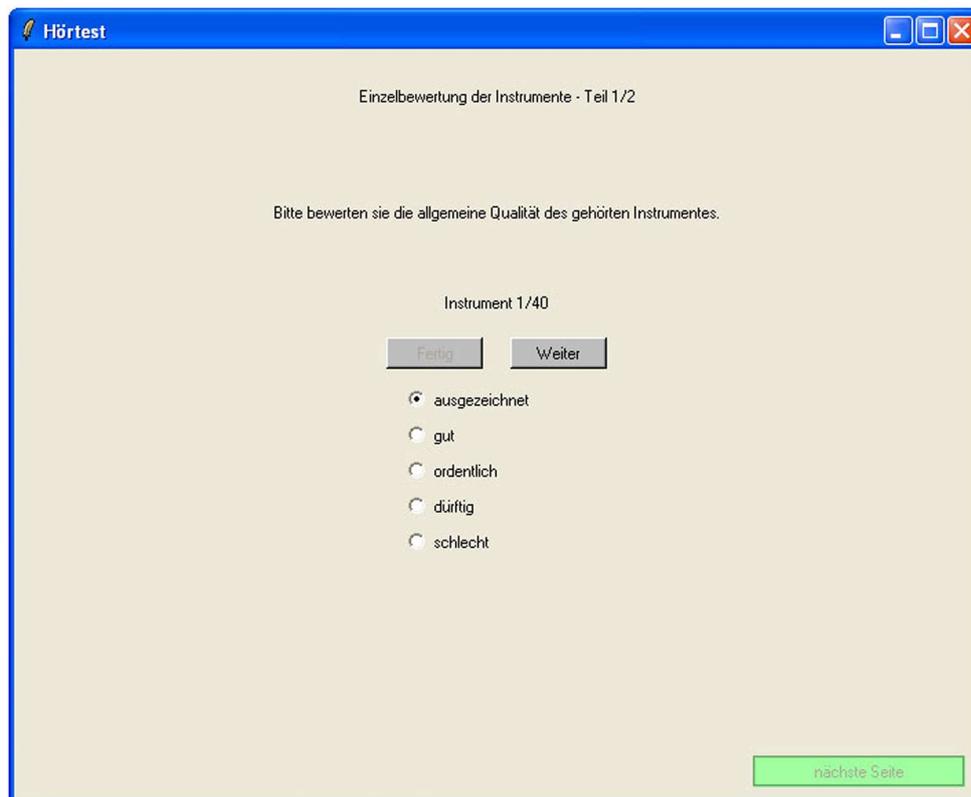


Abbildung D.2: Screenshot Benutzeroberfläche für Hörversuch 1.

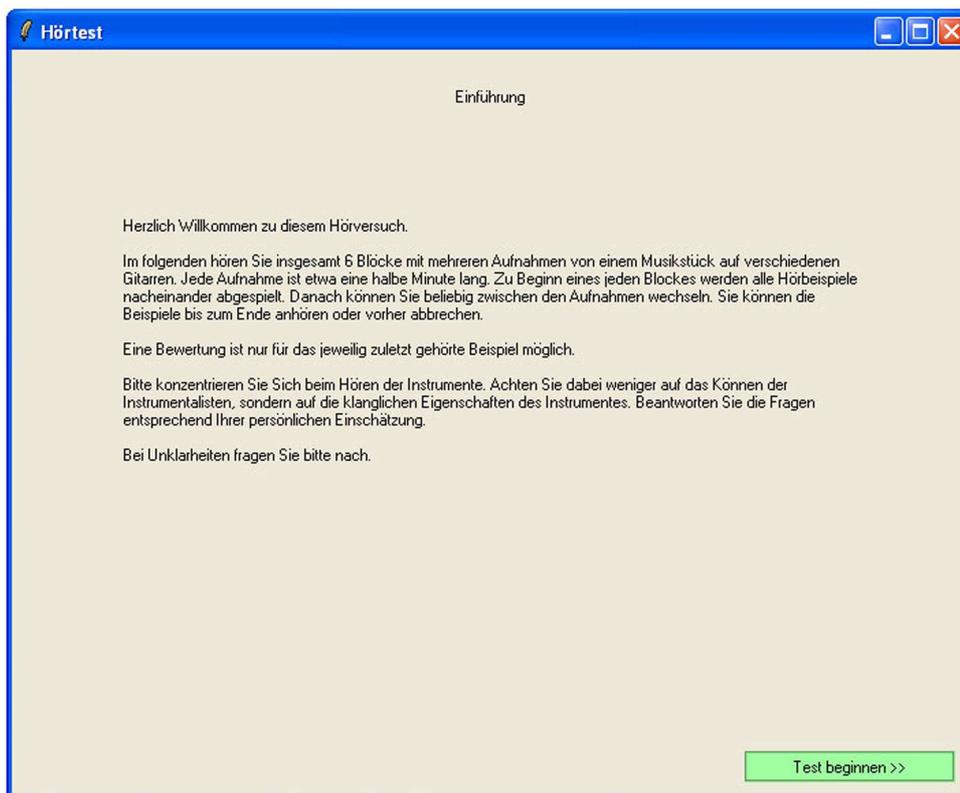


Abbildung D.3: Screenshot Einleitung für Hörversuch 2.

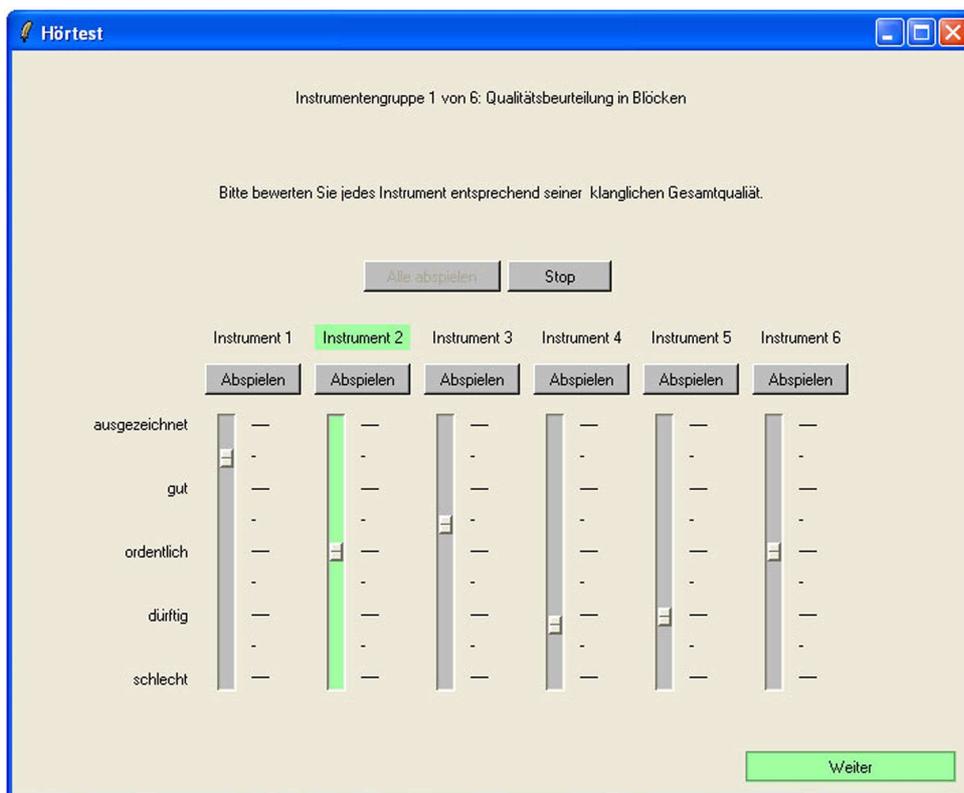


Abbildung D.4: Screenshot Benutzeroberfläche für Hörversuch 2.

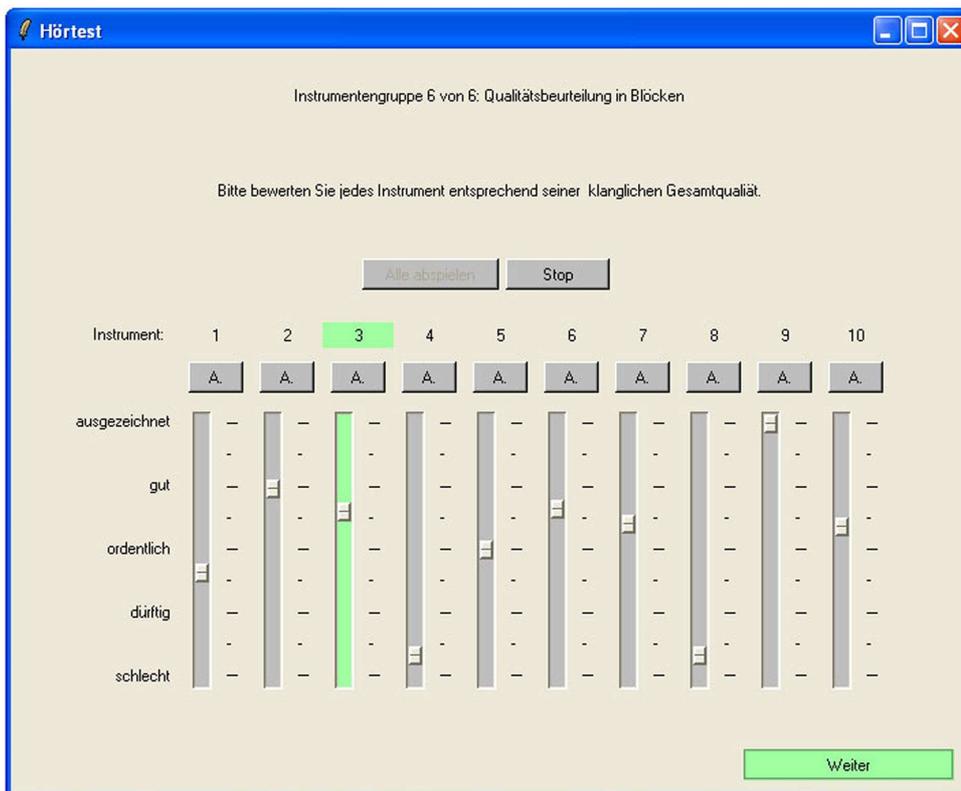


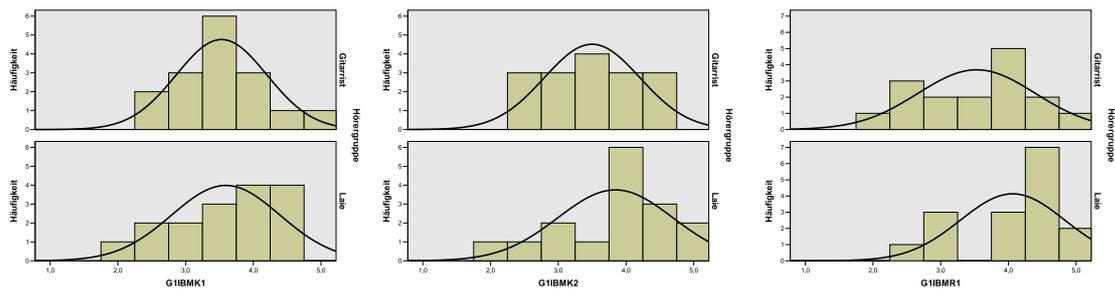
Abbildung D.5: Screenshot Benutzeroberfläche für Hörversuch 3.

Anhang E

Graphen und Tabellen zu Hörversuch 1

Tabelle E.1: Ergebnistabelle für Kolmogorov-Smirnov-Anpassungstest zur Überprüfung auf Normalverteilung der Variablen in Hörversuch 1.

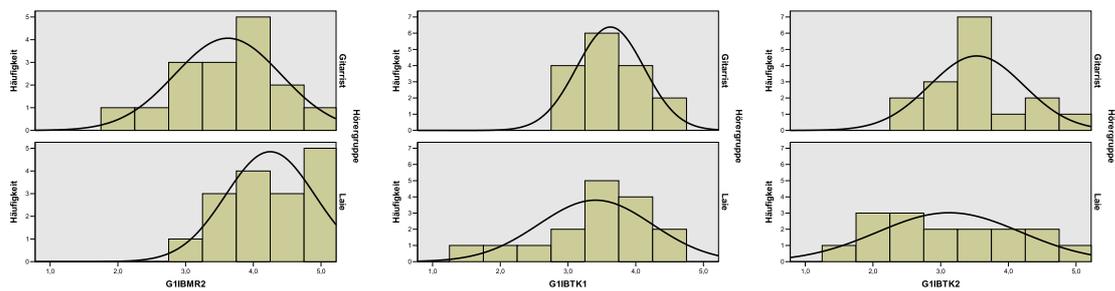
Variable	G1IBMK1	G1IBMK2	G1IBMR1	G1IBMR2	G1IBTK1
K-S Z	0,867	1,089	1,241	0,888	1,186
Sig.	0,440	0,187	0,092	0,410	0,120
Variable	G1IBTK2	G1IBTR1	G1IBTR2	G1SPMK1	G1SPMK2
K-S Z	0,811	0,928	1,157	0,861	0,915
Sig.	0,527	0,355	0,137	0,449	0,372
Variable	G1SPMR1	G1SPMR2	G1SPTK1	G1SPTK2	G1SPTR1
K-S Z	0,944	0,724	0,864	1,104	0,777
Sig.	0,335	0,671	0,444	0,175	0,582
Variable	G1SPTR2	G5IBMK1	G5IBMK2	G5IBMR1	G5IBMR2
K-S Z	0,947	0,840	1,060	0,915	0,910
Sig.	0,331	0,480	0,211	0,372	0,379
Variable	G5IBTK1	G5IBTK2	G5IBTR1	G5IBTR2	G5SPMK1
K-S Z	0,823	1,270	0,746	1,036	0,636
Sig.	0,508	0,079	0,634	0,233	0,813
Variable	G5SPMK2	G5SPMR1	G5SPMR2	G5SPTK1	G5SPTK2
K-S Z	0,846	0,814	0,869	1,061	0,932
Sig.	0,472	0,522	0,437	0,211	0,350
Variable	G5SPTR1	G5SPTR2			
K-S Z	0,937	1,050			
Sig.	0,344	0,220			



(a) G1BMK1

(b) G1BMK2

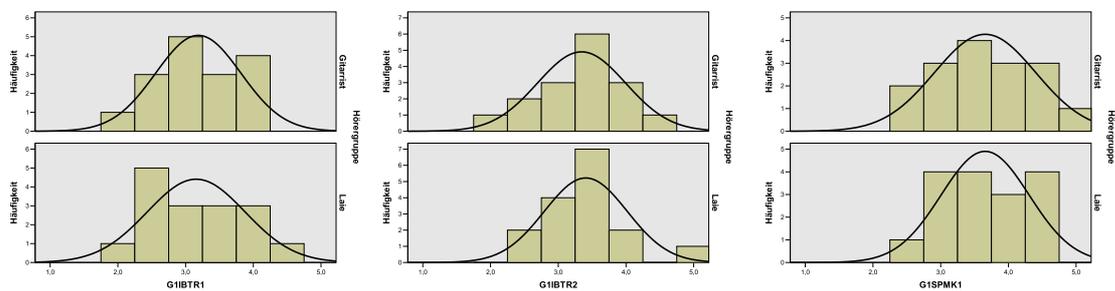
(c) G1BMR1



(d) G1BMR2

(e) G1BTK1

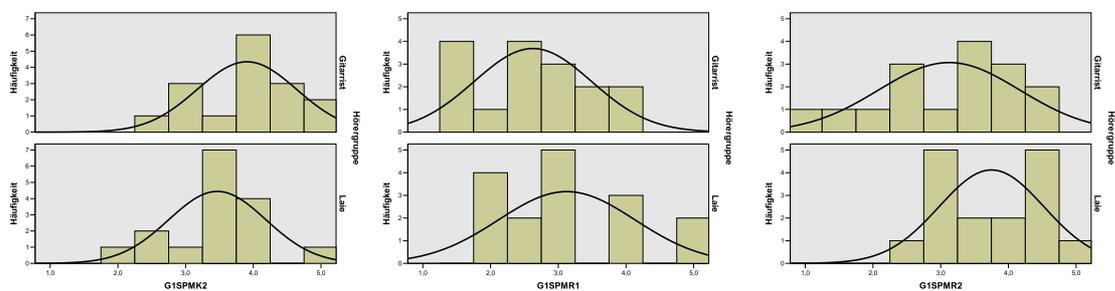
(f) G1BTK2



(g) G1BTR1

(h) G1BTR2

(i) G1SPMK1



(j) G1SPMK2

(k) G1SPMR1

(l) G1SPMR2

Abbildung E.2: Häufigkeitsverteilungen für Stichproben aus Hörversuch 1.

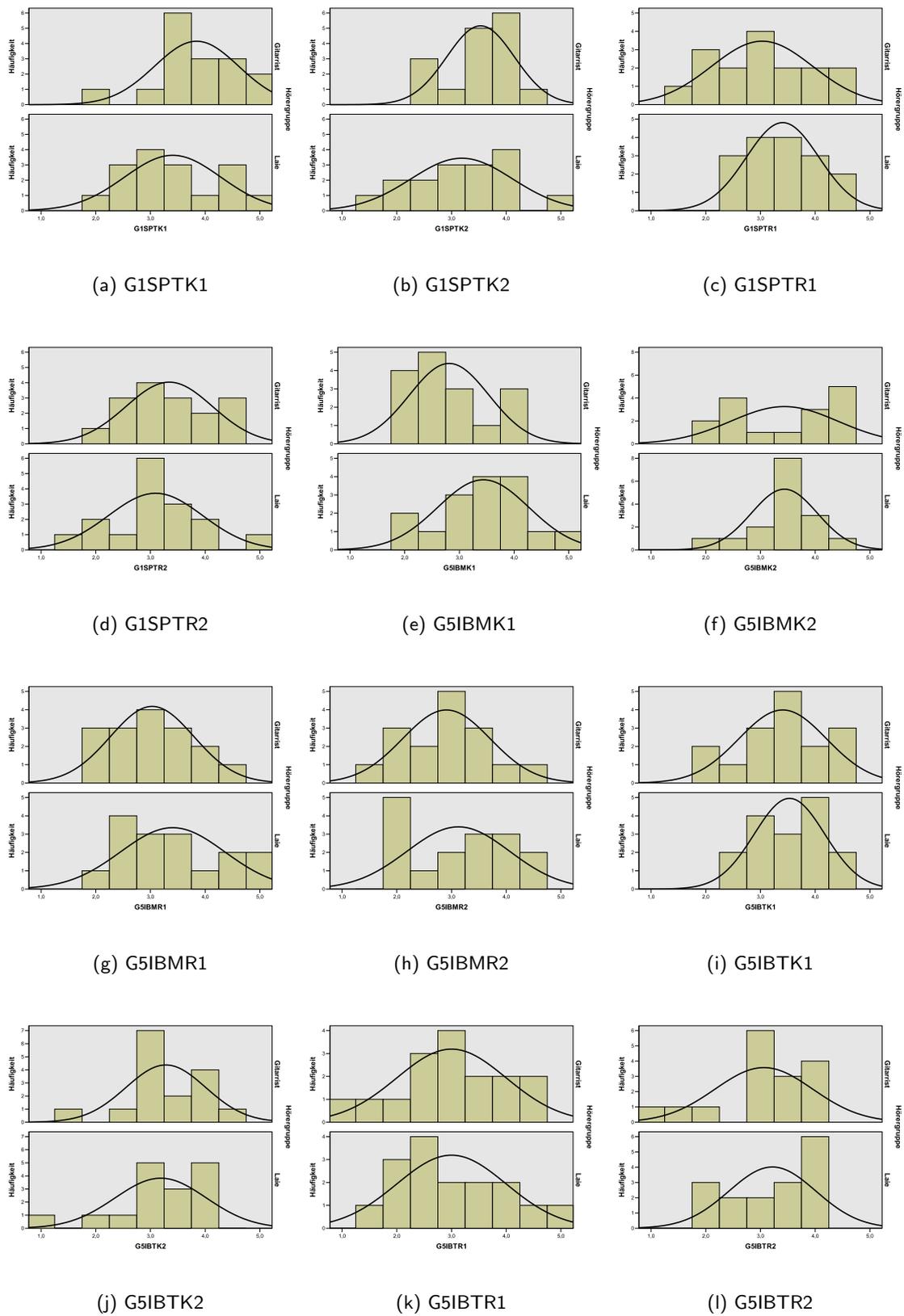
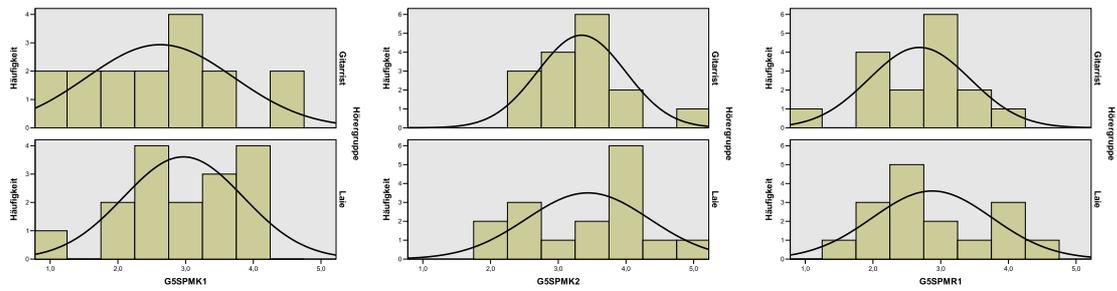


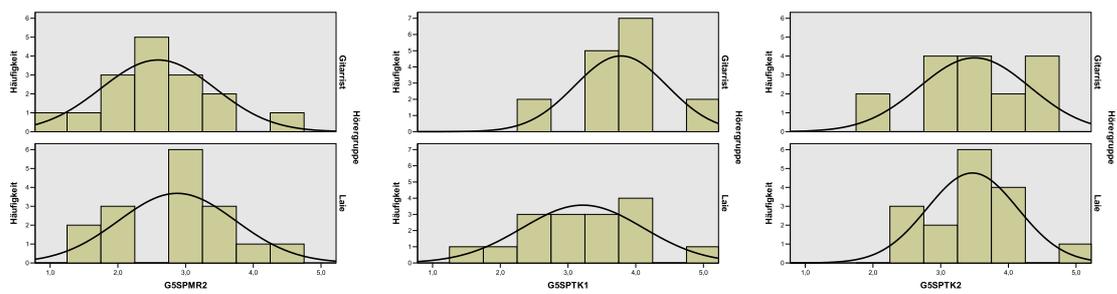
Abbildung E.3: Häufigkeitsverteilungen für Stichproben aus Hörversuch 1.



(a) G5SPMK1

(b) G5SPMK2

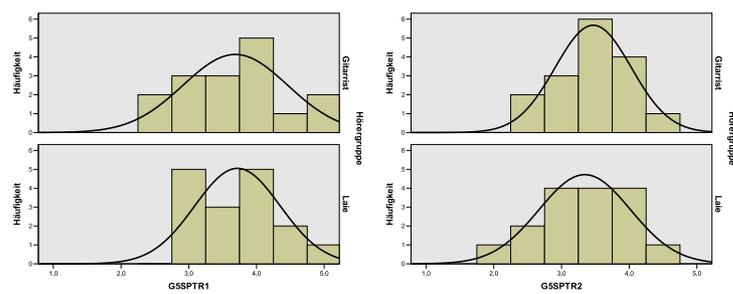
(c) G5SPMR1



(d) G5SPMR2

(e) G5SPTK1

(f) G5SPTK2



(g) G5SPTR1

(h) G5SPTR2

Abbildung E.4: Häufigkeitsverteilungen für Stichproben aus Hörversuch 1.

Anhang **F**

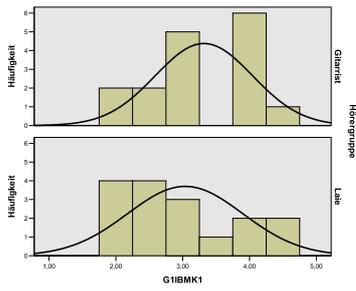
Graphen und Tabellen zu Hörversuch 2

Tabelle F.1: Ergebnistabelle für Kolmogorov-Smirnov-Anpassungstest zur Überprüfung auf Normalverteilung der Variablen in Hörversuch 2.

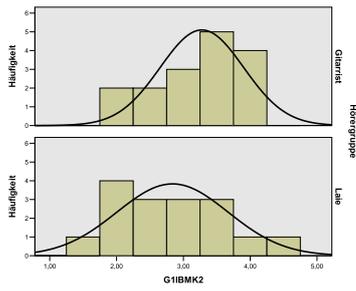
Variable	G1IBMK1	G1IBMK2	G1IBMR1	G1IBMR2	G1SPMK1
K-S Z	0,712	0,842	1,022	0,934	0,675
Sig.	0,690	0,478	0,248	0,348	0,752
Variable	G1SPMK2	G1SPMR1	G1SPMR2	G5IBMK1	G5IBMK2
K-S Z	0,880	0,550	0,530	0,795	0,849
Sig.	0,421	0,923	0,941	0,553	0,467
Variable	G5IBMR1	G5IBMR2	G5SPMK1	G5SPMK2	G5SPMR1
K-S Z	0,605	0,914	0,857	0,686	1,070
Sig.	0,858	0,374	0,455	0,734	0,203
Variable	G5SPMR2	ANKER11	ANKER12	ANKER21	ANKER22
K-S Z	0,586	0,561	0,682	1,211	0,714
Sig.	0,882	0,912	0,741	0,106	0,687
Variable	ANKER31	ANKER32	ANKER41	ANKER42	RANKER11
K-S Z	0,704	0,760	0,936	0,580	0,993
Sig.	0,705	0,611	0,345	0,890	0,278
Variable	RANKER12				
K-S Z	0,696				
Sig.	0,718				

Hörer- gruppe	Laie	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total	N	Mean	Median									
Hörer- gruppe	Laie	G11BMK1	3	2,04	3,00	2,24	3,96	3,00	2,64	2,36	2,24	1,80	2,12	2,04	2,68	1,96	2,56	3,52	1,96	4,08	2,68	2,76	3,32	4,04	1,92	3,80	2,48			
		G11BMR1	4	2,52	1,48	4,04	2,04	3,76	3,68	3,84	2,60	4,64	2,32	3,00	4,60	4,32	4,08	3,36	4,48	2,20	4,64	2,60	4,04	3,00	3,52	3,00	4,00	2,48		
		G11BMK2	6	3,24	3,08	4,00	4,00	3,52	3,84	3,52	3,00	3,56	3,68	2,56	3,48	2,52	1,96	1,64	4,08	3,08	4,04	2,56	4,04	1,72	3,96	2,36	3,64	2,04		
		G11BMR2	8	2,52	2,40	4,52	4,12	3,48	3,76	2,76	3,36	1,24	3,12	2,24	1,32	1,60	4,28	1,44	1,32	4,20	2,44	4,08	3,68	4,28	1,96	4,00	3,36	4,04	3,60	
		G11SPMK1	13	4,52	3,72	3,68	4,52	2,48	3,52	3,12	2,00	4,84	2,08	3,52	4,84	2,80	2,20	2,04	2,60	4,08	1,52	4,52	2,56	3,88	3,04	3,24	4,00	3,44	3,40	
		G11BMR1	15	4,20	3,48	2,96	2,76	4,00	4,04	1,52	1,00	3,60	2,36	2,00	2,00	2,00	3,84	2,12	2,24	3,24	3,36	3,72	4,16	2,68	1,56	2,40	1,88	2,40	4,52	
		G11SPMK2	17	3,92	3,96	4,08	4,68	3,96	3,00	4,48	3,52	3,68	4,00	3,48	4,48	2,52	2,04	3,52	2,96	4,68	3,00	4,68	2,96	4,52	2,40	2,00	2,00	4,52	2,56	
		G11BMK1	18	2,68	2,04	3,96	2,84	2,20	2,40	1,88	2,28	2,12	2,76	2,64	2,68	1,68	3,00	1,80	1,60	4,00	3,36	3,92	3,00	3,56	3,00	3,64	3,24	4,00	3,00	
		G11SPMR1	24	2,08	2,04	5,00	3,92	3,00	2,00	4,00	5,00	1,00	1,96	4,00	2,00	1,00	2,00	1,96	3,04	5,00	1,00	5,00	3,00	5,00	2,04	5,00	3,04	5,00	1,00	
		G11BMR2	25	2,56	3,64	4,04	3,84	2,56	1,84	1,24	2,00	1,72	2,40	3,64	3,40	1,44	1,36	1,48	1,44	1,52	3,72	1,52	2,56	1,40	2,56	1,52	2,60	1,52	2,56	
		G11SPMK1	27	2,24	2,20	2,56	2,52	1,84	2,24	2,52	2,32	4,80	2,96	4,80	2,88	2,48	3,16	4,48	3,88	3,96	3,00	4,52	3,24	4,80	3,16	4,88	3,80	4,28	2,36	
		G11BMK2	29	3,20	2,52	4,84	4,84	3,20	4,24	2,52	2,32	4,80	2,96	4,80	2,88	2,48	3,16	4,48	3,88	3,96	3,00	4,52	3,24	4,80	3,16	4,88	3,80	4,28	2,36	
		G11SPMR2	31	3,48	2,92	4,48	3,96	2,80	3,00	4,00	3,48	3,00	3,32	5,00	5,00	2,00	2,52	4,00	3,96	4,24	3,52	4,44	3,52	4,44	2,00	4,40	2,52	3,96	3,00	
		G11BMR1	32	1,96	2,00	3,92	3,96	2,96	3,04	2,04	2,48	2,00	2,48	3,96	3,96	2,00	2,96	2,48	4,00	3,52	3,52	3,04	3,00	3,96	1,96	4,48	3,00	4,52	2,48	
		Total	N	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	
			Mean	3,04	2,84	3,82	3,74	3,07	3,08	2,58	2,86	2,63	2,74	3,33	3,14	1,93	2,79	2,50	2,80	3,99	2,63	4,06	3,00	3,76	2,64	3,89	2,89	3,89	2,80	
	Median	2,84	2,72	3,98	3,96	3,00	3,02	2,54	2,50	2,22	2,62	3,50	2,78	1,64	2,60	2,06	2,78	4,04	2,86	4,24	3,00	4,00	2,74	4,00	3,00	4,00	2,78			
Git.	Git.	G51SPMR1	1	4,00	3,52	3,96	4,00	3,92	4,04	3,72	3,48	3,00	2,56	3,76	3,52	3,28	4,28	2,56	3,00	4,52	4,28	4,04	2,52	4,04	3,00	4,24	3,24	4,04	3,48	
		G51SPMR2	2	4,68	2,24	4,12	4,28	2,28	2,56	2,04	2,64	3,00	3,52	4,28	4,28	1,80	2,36	1,40	1,40	3,80	2,84	4,08	2,52	4,44	2,28	4,16	2,96	4,04	2,52	
		G51BMR1	3	2,44	3,00	5,00	3,52	4,48	3,00	2,00	1,48	1,52	1,92	1,52	1,52	1,00	2,04	1,00	1,00	4,08	1,88	3,96	1,52	3,96	2,04	4,44	2,04	4,52	1,48	
		G51BMR2	4	3,00	3,52	2,40	2,20	3,00	3,00	2,52	1,96	3,48	4,28	2,64	2,36	4,12	4,16	2,04	4,00	2,00	4,64	2,28	4,12	3,00	3,40	2,52	3,24	2,48	3,44	
		G51SPMK1	5	2,24	3,52	4,00	3,00	4,04	3,52	2,04	4,52	2,48	2,52	2,56	2,52	1,00	1,52	1,52	2,04	3,52	2,04	3,00	1,48	3,52	1,48	3,48	2,08	3,52	3,00	
		G51BMR1	6	4,00	4,00	4,00	3,96	3,48	3,96	2,52	3,00	2,48	2,00	2,04	2,00	2,04	2,52	2,00	1,48	3,00	2,00	4,00	2,48	3,00	3,00	4,00	3,00	4,00	2,00	2,00
		G51SPMK2	7	3,20	3,20	3,00	3,28	3,48	3,76	3,00	3,44	3,52	3,40	3,28	3,76	3,36	3,72	3,48	3,76	3,04	3,44	3,48	4,00	3,24	3,96	3,00	4,20	3,96	4,52	
		G51BMR2	8	3,92	2,56	3,92	3,64	3,00	2,52	3,40	3,00	2,36	2,64	3,96	3,40	2,76	2,60	3,00	2,48	3,96	3,00	3,96	2,60	3,76	2,56	3,92	2,32	3,88	2,36	
		G51SPMR1	9	3,00	2,12	4,12	1,96	3,00	3,00	3,00	1,60	1,92	2,12	2,56	2,52	4,00	3,52	2,76	2,20	3,84	3,68	1,56	4,00	3,52	4,00	2,56	3,00	2,60	4,80	
		G51BMR1	10	2,24	3,48	3,44	3,36	3,48	3,92	1,76	2,24	2,32	2,08	2,08	1,88	3,36	3,96	3,96	2,56	3,24	3,52	3,52	3,92	3,48	3,96	2,92	4,00	2,72	3,96	
		G51SPMK2	11	3,96	3,52	3,52	3,04	3,48	4,04	3,00	4,52	4,52	3,48	3,00	3,52	3,36	4,00	3,00	3,00	3,00	3,00	3,00	3,00	3,00	4,00	3,00	4,00	3,00	4,00	
		G51BMR2	12	3,84	3,84	3,52	3,72	4,12	4,12	3,76	4,04	3,52	3,92	3,92	3,36	2,72	3,36	4,36	2,00	2,68	3,24	3,80	4,16	3,48	4,40	4,04	3,84	3,84	4,04	
		G51SPMR2	13	3,12	4,16	3,24	3,80	1,60	3,76	1,60	3,00	1,08	2,08	1,52	2,64	1,00	1,96	3,72	1,00	4,64	2,12	4,44	2,96	4,68	1,56	4,76	2,36	4,76	3,68	
		G51BMR1	14	2,64	2,72	2,60	2,64	2,52	4,28	2,52	2,24	2,56	3,16	2,72	2,56	3,40	3,32	2,64	3,48	2,44	3,24	2,56	3,48	2,72	2,80	3,00	3,52	2,68	3,88	
		G51SPMK1	15	3,80	4,04	4,00	4,04	2,00	1,52	3,80	4,04	1,72	2,56	3,92	4,04	1,56	1,52	3,00	2,56	4,52	2,52	4,24	2,04	4,56	1,36	4,32	1,68	4,28	2,52	
		G51BMK1	16	3,00	3,00	4,48	2,48	3,00	2,48	1,48	3,00	1,72	2,52	1,96	2,04	1,52	1,96	1,96	1,00	3,76	3,00	4,04	1,52	3,96	3,00	4,48	2,04	4,00	2,04	
Total	N	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16			
	Mean	3,32	3,28	3,71	3,31	3,18	3,34	2,64	3,01	2,58	2,80	2,82	2,83	2,57	2,99	2,50	2,35	3,54	3,13	3,55	2,85	3,71	2,90	3,68	3,00	3,65	3,23			
	Median	3,16	3,50	3,94	3,44	3,24	3,64	2,52	3,00	2,48	2,56	2,68	2,60	3,02	2,96	2,60	2,52	3,64	3,12	3,96	2,78	3,64	3,00	3,96	3,02	3,92	3,46			
Total	N	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32			
	Mean	3,18	3,06	3,76	3,52	3,13	3,21	2,61	2,93	2,60	2,77	3,08	2,92	2,85	2,89	2,50	2,58	3,76	2,88	3,81	2,92	3,73	2,77	3,78	2,95	3,77	3,02			
	Median	3,06	3,04	3,96	3,78	3,00	3,28	2,52	3,00	2,42	2,56	2,94	2,68	2,00	2,64	2,10	2,56	3,88	3,00	3,92	3,00	3,92	3,00	4,00	3,00	3,92	3,00			

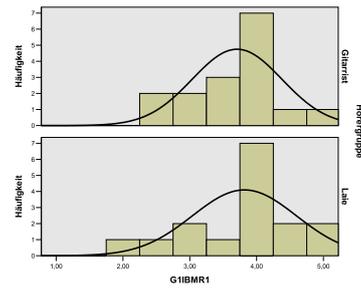
Abbildung F.1: Übersicht der Daten von Hörversuch 2 mit berechnetem Mittelwert und Median für die Gruppierungsfaktorstufen Laie und Gitarrist.



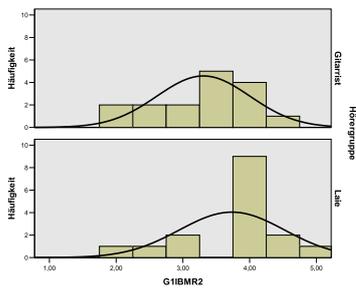
(a) G1BMK1



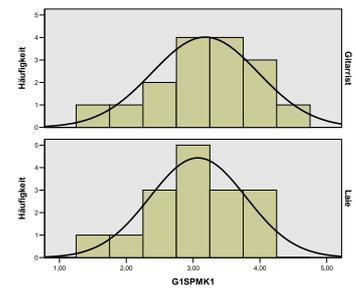
(b) G1BMK2



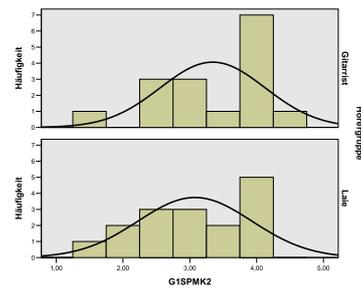
(c) G1BMR1



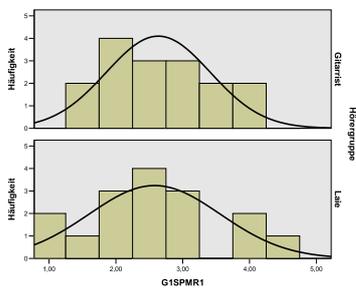
(d) G1BMR2



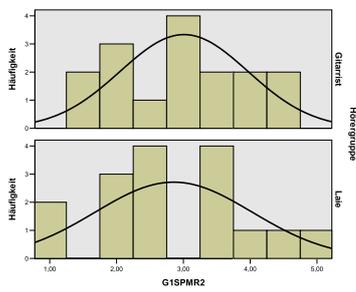
(e) G1SPMK1



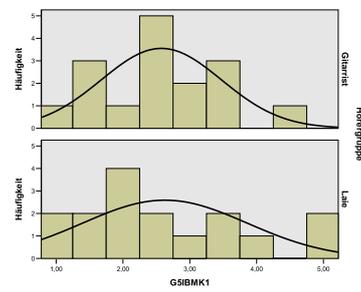
(f) G1SPMK2



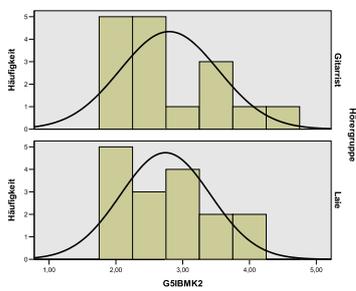
(g) G1SPMR1



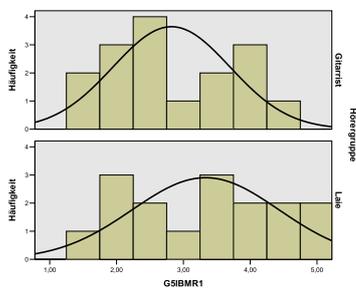
(h) G1SPMR2



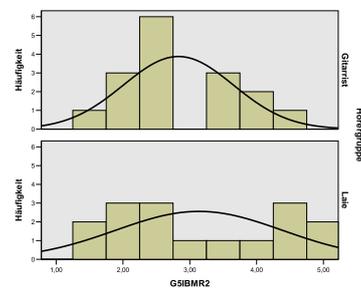
(i) G5IBMK1



(j) G5IBMK2



(k) G5IBMR1



(l) G5IBMR2

Abbildung F.2: Häufigkeitsverteilungen für Stichproben aus Hörversuch 2.

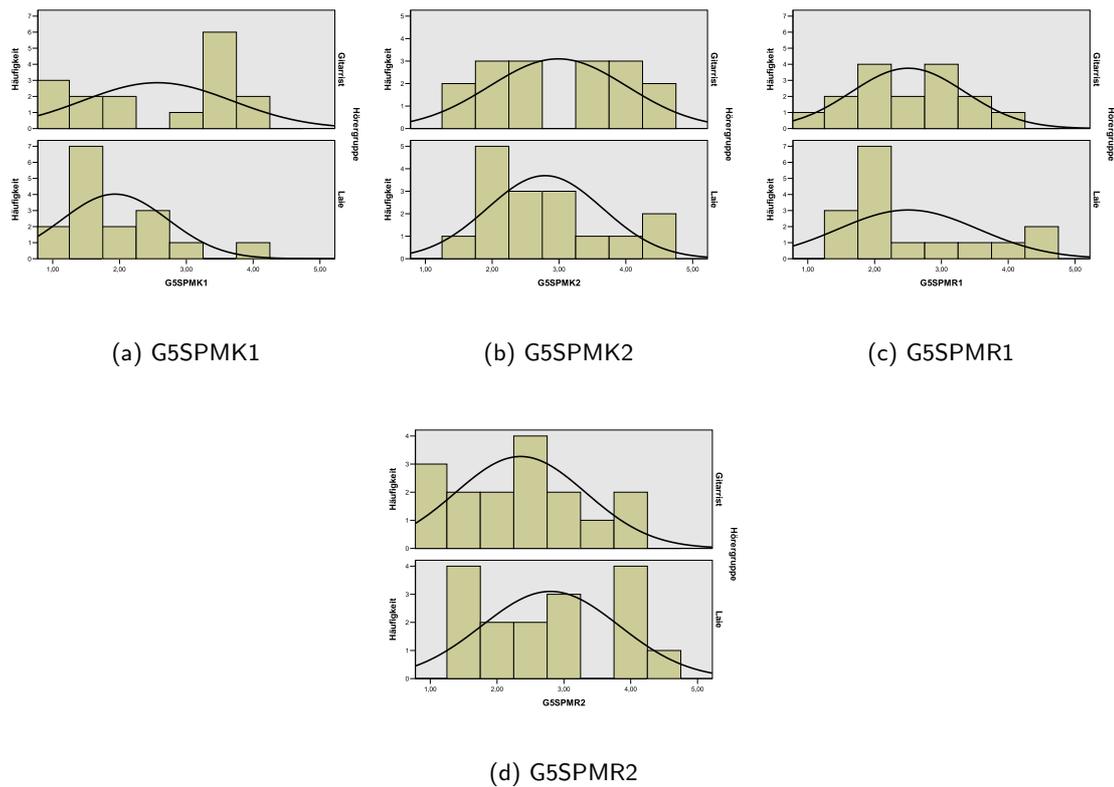


Abbildung F.3: Häufigkeitsverteilungen für Stichproben aus Hörversuch 2.

Tabelle F.3: Ergebnistabelle der fünffaktoriellen Varianzanalyse mit Messwiederholung für den Hörversuch 2 innerhalb der Versuchspersonen nachdem die Stichproben an Ankerbeispiel "ANKERx1" ausgerichtet wurden. Die einzelnen Faktoren werden abgekürzt als: Instrument I , Musiker M , Anspiel A , Raum R und Hörergruppe HG . Es wurde immer die selbe Musiksequenz $Se1$ verwendet. Ist ein Haupteffekt oder eine Interaktion sehr signifikant (Irrtumswahrscheinlichkeit $< 1\%$) wird sie mit ** gekennzeichnet, ist sie signifikant (Irrtumswahrscheinlichkeit $< 5\%$) mit *.

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
I^{**}	19,174	1	19,174	9,52	0,004	0,241
$I * HG$	1,182	1	1,182	0,59	0,450	0,019
Error(I)	60,404	30	2,013			
M^{**}	13,274	1	13,274	10,77	0,003	0,264
$M * HG$	1,144	1	1,144	0,93	0,343	0,030
Error(M)	36,978	30	1,233			
R	0,766	1	0,766	0,34	0,564	0,011
$R * HG$	4,169	1	4,169	1,85	0,184	0,058
Error(R)	67,511	30	2,250			
A^*	3,207	1	3,207	4,62	0,040	0,133
$A * HG$	0,233	1	0,233	0,34	0,567	0,011
Error(A)	20,816	30	0,694			
$I * M$	0,570	1	0,570	0,93	0,343	0,030
$I * M * HG$	0,170	1	0,170	0,28	0,603	0,009

Fortsetzung auf nächster Seite

Tabelle F.3 – Fortsetzung von letzter Seite

Quelle der Varianz	QS	df	$\hat{\sigma}^2$	F	Sig.	$\hat{\eta}_p^2$
Error($I * M$)	18,401	30	0,613			
$I * R$	0,216	1	0,216	0,31	0,582	0,010
$I * R * HG$	2,117	1	2,117	3,03	0,092	0,092
Error($I * R$)	20,949	30	0,698			
$M * R^{**}$	7,693	1	7,693	12,08	0,002	0,287
$M * R * HG$	0,077	1	0,077	0,12	0,730	0,004
Error($M * R$)	19,099	30	0,637			
$I * M * R$	2,548	1	2,548	3,70	0,064	0,110
$I * M * R * HG$	0,287	1	0,287	0,42	0,523	0,014
Error($I * M * R$)	20,634	30	0,688			
$I * A$	1,867	1	1,867	3,80	0,061	0,112
$I * A * HG$	0,142	1	0,142	0,29	0,595	0,010
Error($I * A$)	14,752	30	0,492			
$M * A$	2,404	1	2,404	3,94	0,056	0,116
$M * A * HG$	0,096	1	0,096	0,16	0,695	0,005
Error($M * A$)	18,311	30	0,610			
$I * M * A$	0,049	1	0,049	0,14	0,706	0,005
$I * M * A * HG$	0,861	1	0,861	2,55	0,121	0,078
Error($I * M * A$)	10,124	30	0,337			
$R * A$	0,729	1	0,729	1,58	0,218	0,050
$R * A * HG$	0,075	1	0,075	0,16	0,689	0,005
Error($R * A$)	13,834	30	0,461			
$I * R * A^*$	1,975	1	1,975	5,41	0,027	0,153
$I * R * A * HG$	0,000	1	0,000	0,00	0,984	0,000
Error($I * R * A$)	10,952	30	0,365			
$M * R * A$	0,001	1	0,001	0,00	0,977	0,000
$M * R * A * HG$	0,881	1	0,881	1,47	0,235	0,047
Error($M * R * A$)	17,984	30	0,599			
$I * M * R * A$	0,935	1	0,935	1,80	0,189	0,057
$I * M * R * A * HG$	0,003	1	0,003	0,01	0,940	0,000
Error($I * M * R * A$)	15,552	30	0,518			

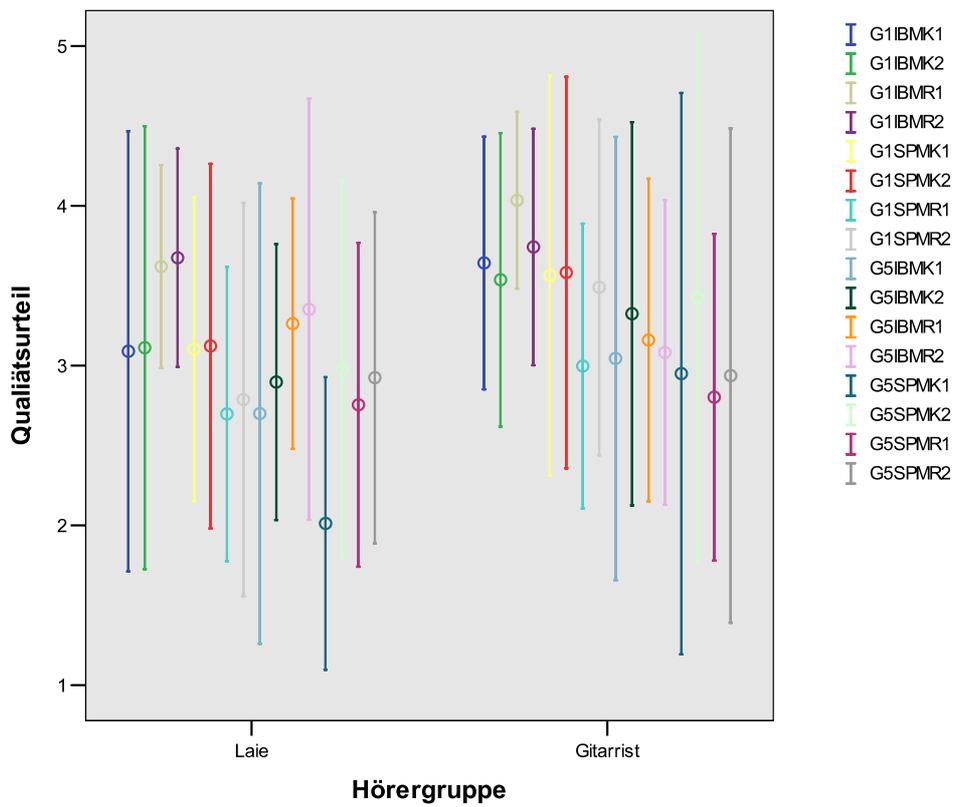
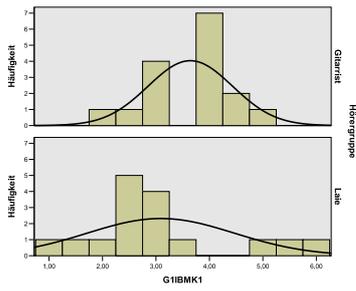
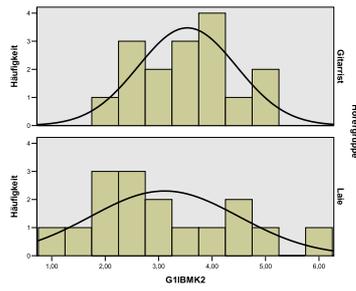


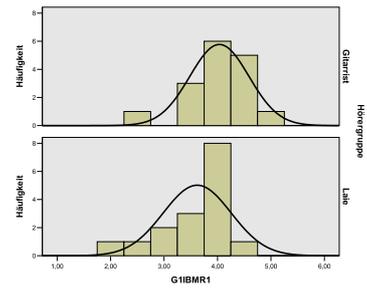
Abbildung F.4: Error-Bar der Messwertverteilungen der Stichproben aus Hörversuch 2 für 32 Personen, 16 Laien und 16 Gitarristen nachdem die Stichproben an Ankerbeispiel "ANKERx1" ausgerichtet wurden. Dargestellt ist der Mittelwert \pm eine Standardabweichung.



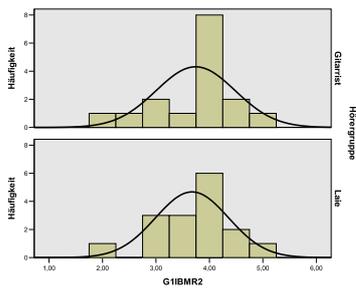
(a) G1BMK1



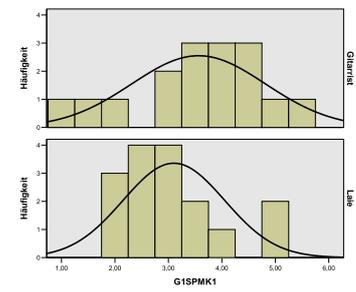
(b) G1BMK2



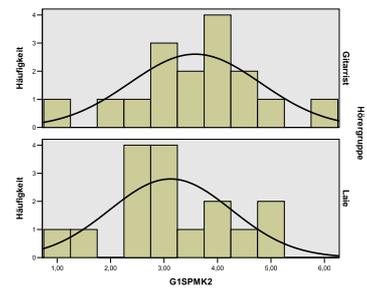
(c) G1BMR1



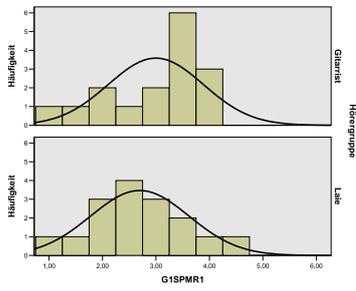
(d) G1BMR2



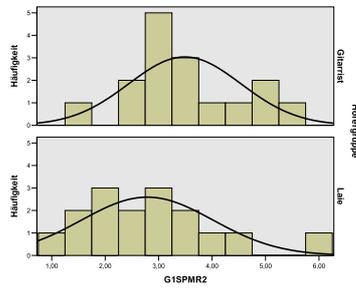
(e) G1SPMK1



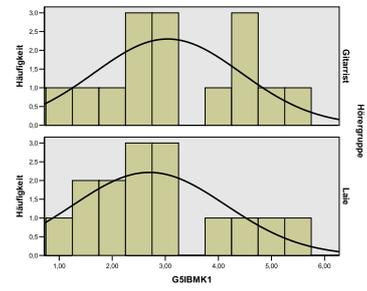
(f) G1SPMK2



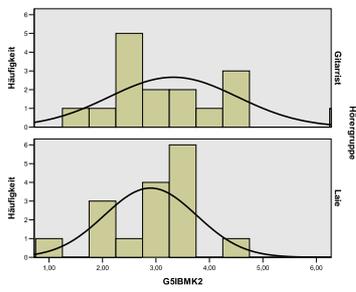
(g) G1SPMR1



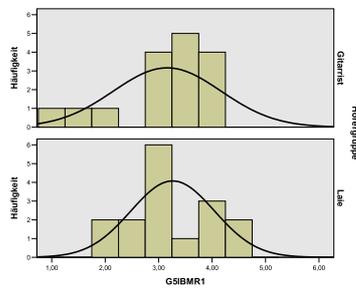
(h) G1SPMR2



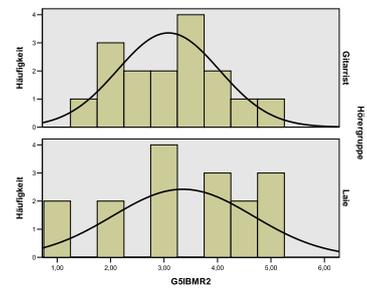
(i) G5IBMK1



(j) G5IBMK2

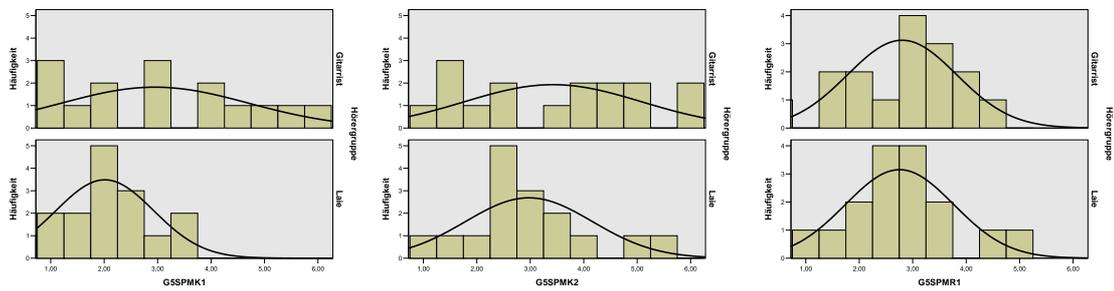


(k) G5IBMR1



(l) G5IBMR2

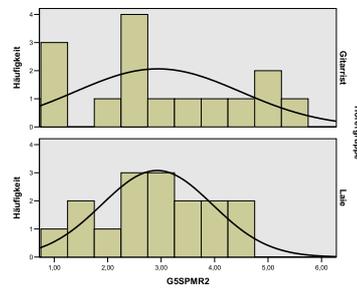
Abbildung F.5: Häufigkeitsverteilungen für Stichproben aus Hörversuch 2 nach Ausrichtung an Ankerbeispiel "ANKERx1".



(a) G5SPMK1

(b) G5SPMK2

(c) G5SPMR1



(d) G5SPMR2

Abbildung F.6: Häufigkeitsverteilungen für Stichproben aus Hörversuch 2 nach Ausrichtung an Ankerbeispiel "ANKERx1".

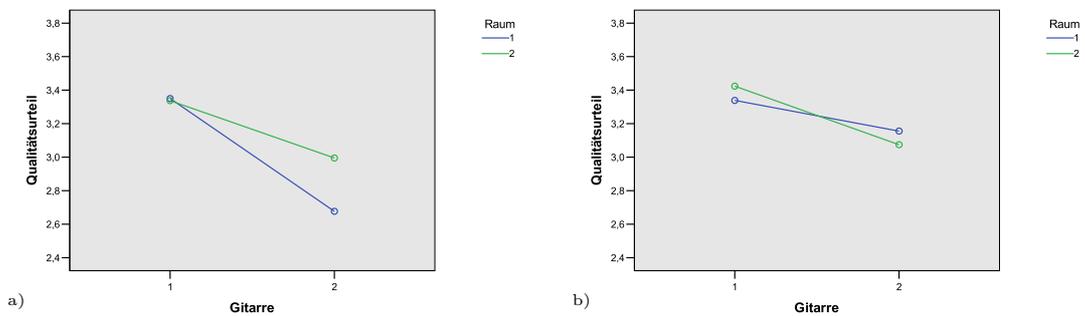


Abbildung F.7: Interaktionsdiagramm mit den Daten aus Hörversuch 2 nach Ausrichtung an versteckter Referenz ANKERx1 für Instrument, Raum und Anspiel a) Anspiel 1, b) Anspiel 2.

Tabelle F.2: Ergebnistabelle für Kolmogorov-Smirnov-Anpassungstest zur Überprüfung auf Normalverteilung der Variablen in Hörversuch 2 nach Ausrichtung an Ankerbeispiel "ANKERx1". Die entsprechenden Stichproben haben konstante Werte und können folglich nicht normalverteilt sein.

Variable	G1IBMK1	G1IBMK2	G1IBMR1	G1IBMR2	G1SPMK1
K-S Z	0,603	0,541	1,074	0,920	0,409
Sig.	0,860	0,931	0,199	0,366	0,996
Variable	G1SPMK2	G1SPMR1	G1SPMR2	G5IBMK1	G5IBMK2
K-S Z	0,419	0,751	0,594	0,856	0,656
Sig.	0,995	0,625	0,872	0,457	0,783
Variable	G5IBMR1	G5IBMR2	G5SPMK1	G5SPMK2	G5SPMR1
K-S Z	0,584	0,514	0,730	0,732	0,488
Sig.	0,885	0,954	0,660	0,657	0,971
Variable	G5SPMR2	ANKER11	ANKER12	ANKER21	ANKER22
K-S Z	0,448		0,476		1,084
Sig.	0,988		0,977		0,191
Variable	ANKER31	ANKER32	ANKER41	ANKER42	RANKER11
K-S Z		0,640		0,756	
Sig.		0,808		0,617	
Variable	RANKER12				
K-S Z	0,688				
Sig.	0,731				

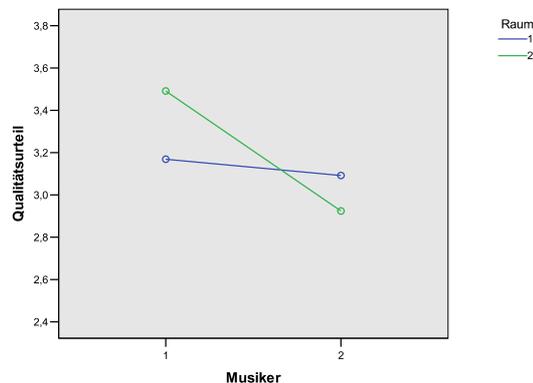


Abbildung F.8: Interaktionsdiagramm für Musiker und Raum in Hörversuch 2 nach Ausrichtung an Referenz ANKERx1.

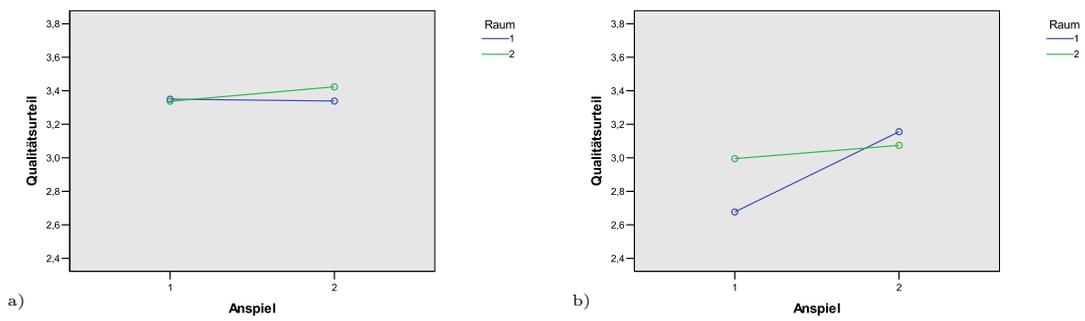


Abbildung F.9: Interaktionsdiagramm mit den Daten aus Hörversuch 2 nach Ausrichtung an versteckter Referenz ANKERx1 noch einmal für Anspiel, Instrument und Raum a) Raum 1, b) Raum 2.

Graphen und Tabellen zu Hörversuch 3

Tabelle G.1: Ergebnistabelle für Kolmogorov-Smirnov-Anpassungstest zur Überprüfung auf Normalverteilung der Variablen in Hörversuch 3.

Variable	A1IBMK2	A2IBMK2	A4IBMK2	A5IBMK2	A22IBMK2
K-S Z	0,598	0,905	0,411	0,581	0,627
Sig.	0,867	0,386	0,996	0,888	0,827
Variable	A23IBMK2	A24IBMK2	A25IBMK2	A26IBMK2	A27IBMK2
K-S Z	0,549	0,839	0,676	0,500	0,470
Sig.	0,924	0,483	0,751	0,964	0,980

Tabelle G.2: Ergebnistabelle des Scheffé Tests zur Überprüfung eines signifikanten Unterschieds zwischen jeweils zwei Instrumenten. Er setzt homogene Fehlervarianzen voraus, was zuvor mit Levene's Test geprüft wurde. Es ist besonders der Unterschied zwischen den jeweiligen Mittelwerten, sowie das Signifikanzniveau von Interesse. Ist es kleiner als 0,05 spricht man von einem signifikanten Unterschied.

Instrument (I)	Instrument (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	0,340	0,162	0,882	-0,332	1,012
	4	0,089	0,162	1,000	-0,584	0,761
	5	0,166	0,162	0,999	-0,506	0,839
	22	0,928	0,162	0,000	0,255	1,600
	23	0,436	0,162	0,612	-0,236	1,109
	24	0,009	0,162	1,000	-0,664	0,681
	25	0,228	0,162	0,992	-0,445	0,900
	26	0,294	0,162	0,951	-0,379	0,966
2	27	0,123	0,162	1,000	-0,550	0,795
	1	-0,340	0,162	0,882	-1,012	0,332
	4	-0,251	0,162	0,983	-0,924	0,421
	5	-0,174	0,162	0,999	-0,846	0,499
	22	0,588	0,162	0,162	-0,085	1,260

Fortsetzung auf nächster Seite

Tabelle G.2 – Fortsetzung von letzter Seite

Instrument (I)	Instrument (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
	23	0,096	0,162	1,000	-0,576	0,769
	24	-0,331	0,162	0,898	-1,004	0,341
	25	-0,113	0,162	1,000	-0,785	0,560
	26	-0,046	0,162	1,000	-0,719	0,626
	27	-0,218	0,162	0,994	-0,890	0,455
4	1	-0,089	0,162	1,000	-0,761	0,584
	2	0,251	0,162	0,983	-0,421	0,924
	5	0,078	0,162	1,000	-0,595	0,750
	22	0,839	0,162	0,002	0,166	1,511
	23	0,348	0,162	0,867	-0,325	1,020
	24	-0,080	0,162	1,000	-0,752	0,592
	25	0,139	0,162	1,000	-0,534	0,811
	26	0,205	0,162	0,996	-0,467	0,877
	27	0,034	0,162	1,000	-0,639	0,706
5	1	-0,166	0,162	0,999	-0,839	0,506
	2	0,174	0,162	0,999	-0,499	0,846
	4	-0,078	0,162	1,000	-0,750	0,595
	22	0,761	0,162	0,011	0,089	1,434
	23	0,270	0,162	0,972	-0,402	0,942
	24	-0,158	0,162	1,000	-0,830	0,515
	25	0,061	0,162	1,000	-0,611	0,734
	26	0,128	0,162	1,000	-0,545	0,800
	27	-0,044	0,162	1,000	-0,716	0,629
22	1	-0,928	0,162	0,000	-1,600	-0,255
	2	-0,588	0,162	0,162	-1,260	0,085
	4	-0,839	0,162	0,002	-1,511	-0,166
	5	-0,761	0,162	0,011	-1,434	-0,089
	23	-0,491	0,162	0,423	-1,164	0,181
	24	-0,919	0,162	0,000	-1,591	-0,246
	25	-0,700	0,162	0,032	-1,372	-0,028
	26	-0,634	0,162	0,089	-1,306	0,039
	27	-0,805	0,162	0,004	-1,477	-0,133
23	1	-0,436	0,162	0,612	-1,109	0,236
	2	-0,096	0,162	1,000	-0,769	0,576
	4	-0,348	0,162	0,867	-1,020	0,325
	5	-0,270	0,162	0,972	-0,942	0,402
	22	0,491	0,162	0,423	-0,181	1,164
	24	-0,428	0,162	0,642	-1,100	0,245
	25	-0,209	0,162	0,996	-0,881	0,464
	26	-0,143	0,162	1,000	-0,815	0,530
	27	-0,314	0,162	0,926	-0,986	0,359
24	1	-0,009	0,162	1,000	-0,681	0,664
	2	0,331	0,162	0,898	-0,341	1,004
	4	0,080	0,162	1,000	-0,592	0,752

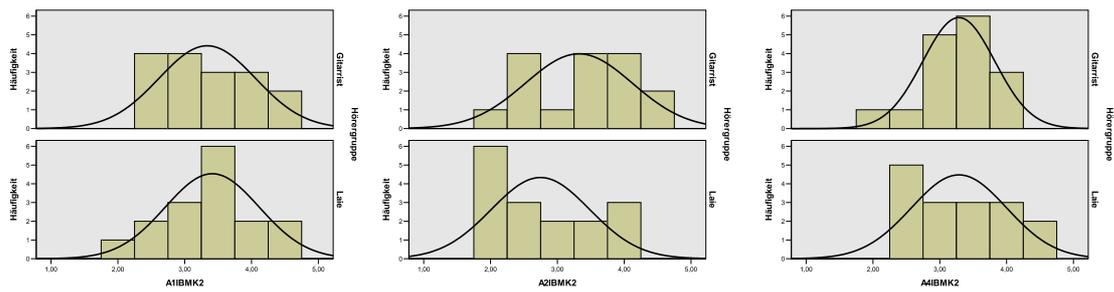
Fortsetzung auf nächster Seite

Tabelle G.2 – Fortsetzung von letzter Seite

Instrument (I)	Instrument (J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval		
					Lower Bound	Upper Bound	
25	5	0,158	0,162	1,000	-0,515	0,830	
	22	0,919	0,162	0,000	0,246	1,591	
	23	0,428	0,162	0,642	-0,245	1,100	
	25	0,219	0,162	0,994	-0,454	0,891	
	26	0,285	0,162	0,960	-0,387	0,957	
	27	0,114	0,162	1,000	-0,559	0,786	
	1	-0,228	0,162	0,992	-0,900	0,445	
	2	0,113	0,162	1,000	-0,560	0,785	
	4	-0,139	0,162	1,000	-0,811	0,534	
	5	-0,061	0,162	1,000	-0,734	0,611	
	22	0,700	0,162	0,032	0,028	1,372	
	23	0,209	0,162	0,996	-0,464	0,881	
	24	-0,219	0,162	0,994	-0,891	0,454	
	26	0,066	0,162	1,000	-0,606	0,739	
26	27	-0,105	0,162	1,000	-0,777	0,567	
	1	-0,294	0,162	0,951	-0,966	0,379	
	2	0,046	0,162	1,000	-0,626	0,719	
	4	-0,205	0,162	0,996	-0,877	0,467	
	5	-0,128	0,162	1,000	-0,800	0,545	
	22	0,634	0,162	0,089	-0,039	1,306	
	23	0,143	0,162	1,000	-0,530	0,815	
	24	-0,285	0,162	0,960	-0,957	0,387	
	25	-0,066	0,162	1,000	-0,739	0,606	
	27	-0,171	0,162	0,999	-0,844	0,501	
	27	1	-0,123	0,162	1,000	-0,795	0,550
		2	0,218	0,162	0,994	-0,455	0,890
		4	-0,034	0,162	1,000	-0,706	0,639
		5	0,044	0,162	1,000	-0,629	0,716
22		0,805	0,162	0,004	0,133	1,477	
23		0,314	0,162	0,926	-0,359	0,986	
24		-0,114	0,162	1,000	-0,786	0,559	
25		0,105	0,162	1,000	-0,567	0,777	
26		0,171	0,162	0,999	-0,501	0,844	

		A1B MK2										
		A21B MK2										
		A31B MK2										
		A41B MK2										
		A51B MK2										
		A61B MK2										
		A71B MK2										
Hörergruppe	Laie	1	3,52	3,00	2,56	2,60	2,00	3,76	4,52	3,92	3,00	4,04
		2	2,08	2,08	3,00	2,00	3,72	3,36	2,04	2,04	2,00	2,04
		3	4,60	3,44	3,80	4,28	2,76	2,48	3,28	4,48	4,24	3,40
		4	2,88	2,44	3,32	3,00	1,80	2,36	2,44	2,76	3,20	2,60
		5	4,04	3,92	2,60	3,20	1,36	2,16	3,68	3,48	2,60	3,24
		6	3,52	2,20	2,28	3,16	2,68	3,52	5,00	3,96	1,36	4,52
		7	3,48	3,80	4,36	3,44	4,00	4,04	2,12	2,88	2,48	2,64
		8	4,12	2,28	4,08	2,60	3,76	3,00	4,00	4,12	4,24	3,52
		9	4,52	2,48	4,48	3,80	2,04	3,00	3,52	3,56	4,48	3,96
		10	2,52	1,92	2,80	3,00	2,52	1,76	3,52	3,04	3,32	3,48
		11	3,52	3,04	3,48	4,52	2,04	2,56	4,52	3,52	2,56	4,00
		12	3,00	2,00	2,48	2,84	1,48	3,00	3,52	2,04	2,52	1,80
		13	2,64	2,00	3,04	2,64	1,40	2,24	3,24	2,88	3,32	2,40
		14	3,40	3,80	3,68	4,40	2,52	3,16	3,92	3,16	4,04	3,84
		15	3,72	3,48	2,64	3,52	2,48	3,36	3,48	2,64	4,12	3,00
		16	3,04	2,08	4,00	3,56	2,52	2,04	2,52	2,56	1,96	3,00
		Total	N	16	16	16	16	16	16	16	16	16
			Mean	3,41	2,75	3,29	3,29	2,44	2,86	3,46	3,19	3,09
			Median	3,50	2,46	3,18	3,18	2,50	3,00	3,52	3,10	3,10
	Gitarrist	1	3,96	3,76	3,52	4,24	3,48	4,52	4,48	2,28	3,16	3,20
		2	2,56	3,56	1,92	4,36	1,00	4,20	3,32	2,56	2,36	2,88
		3	4,56	4,00	3,00	2,92	1,00	1,92	4,04	4,00	3,00	3,56
		4	3,12	2,28	3,00	2,92	3,32	3,44	2,24	3,24	3,40	2,32
		5	4,52	3,04	4,04	3,00	2,04	4,52	3,92	3,00	3,00	4,52
		6	4,00	2,48	3,48	3,96	2,04	2,48	3,96	3,52	3,52	3,48
		7	3,32	3,48	3,00	3,20	3,52	3,20	4,00	3,96	3,48	3,44
		8	3,28	3,60	3,36	3,80	3,00	3,00	3,60	3,68	3,44	3,36
		9	2,36	4,52	3,52	2,16	2,88	3,40	2,12	2,88	2,56	3,00
		10	2,40	3,76	2,52	2,96	4,52	2,00	3,48	4,04	2,44	2,48
		11	3,00	3,96	4,00	4,52	3,00	3,00	3,56	3,00	3,00	3,00
		12	4,20	3,56	3,88	3,76	2,68	3,16	2,92	3,92	3,68	3,80
		13	3,48	4,36	3,20	1,72	1,20	1,72	2,76	2,64	3,68	4,12
		14	2,60	2,72	3,20	2,28	2,12	2,56	2,04	2,52	2,36	2,72
		15	3,00	2,28	3,44	1,80	1,92	2,76	3,96	2,44	3,36	3,72
		16	3,04	1,80	3,48	2,52	1,52	2,36	2,00	2,00	2,72	3,00
		Total	N	16	16	16	16	16	16	16	16	16
			Mean	3,34	3,32	3,29	3,13	2,45	3,02	3,28	3,11	3,07
			Median	3,20	3,56	3,40	2,98	2,40	3,00	3,52	3,00	3,08
	Total	N	32	32	32	32	32	32	32	32	32	32
		Mean	3,38	3,04	3,29	3,21	2,45	2,94	3,37	3,15	3,08	3,25
		Median	3,36	3,04	3,34	3,08	2,50	3,00	3,52	3,02	3,08	3,30

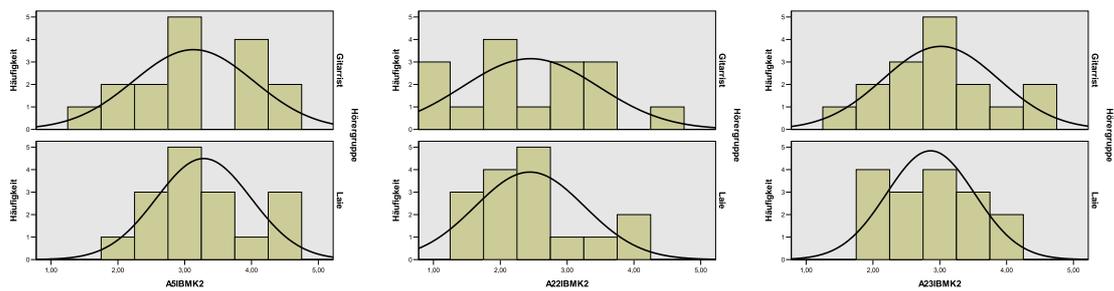
Abbildung G.1: Übersicht der Daten von Hörversuch 3 mit berechnetem Mittelwert und Median für die Gruppierungsfaktorstufen Laie und Gitarrist.



(a) A1BMK2

(b) A21BMK2

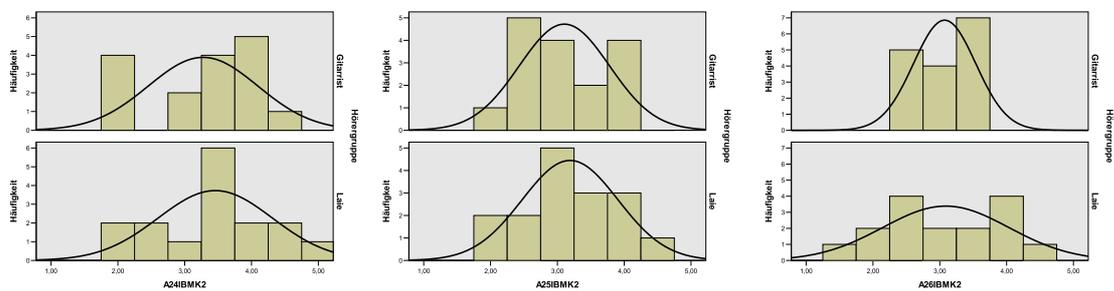
(c) A41BMK2



(d) A51BMK2

(e) A221BMK2

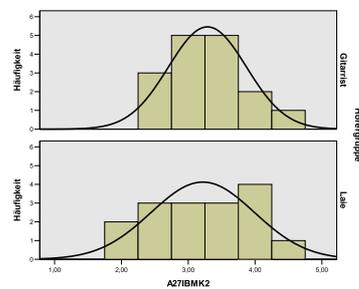
(f) A231BMK2



(g) A241BMK2

(h) A251BMK2

(i) A261BMK2



(j) A271BMK2

Abbildung G.2: Häufigkeitsverteilungen für Stichproben aus Hörversuch 3. Zum Beispiel repräsentiert "A1 IB M K 2" das Qualitätsurteil für Gitarre "G1" gespielt von Musiker "IB" mit Sequenz "M" im Raum "K", zweites Anspiel.