Technische Universität Berlin
Fakultät I

Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

Über die raumakustische Wahrnehmung durch Musiker

– eine experimentelle Studie –

*

Masterarbeit

im Studiengang
Audiokommunikation und -technologie
zur Erlangung des akademischen Grades Master of Science (M. Sc.),

vorgelegt von

Erik Stahnke

[redacted]

Abgabe: 8. Juli 2013

Betreuer: Prof. Dr. Stefan Weinzierl
Zora Schärer Kalkandjiev M. A.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig, eigenhändig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Arbeiten direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Ort, Datum *Erik Stahnke*

Danksagung

Mein herzlicher Dank gilt allen, die zur Durchführung und Fertigstellung dieser Arbeit beigetragen haben: Zora Schärer Kalkandjiev, Stefan Weinzierl, Steffen Lepa, Alexander Lindau, Hans-Joachim Maempel, Michael Horn und im Besonderen den Musikern, die an der in dieser Arbeit vorgestellten Studie teilgenommen haben. Außerdem möchte ich mich selbstverständlich bei meinen Freunden, meiner Familie und Eva für die immer fortwährende Unterstützung bedanken. //

Inhaltsverzeichnis

1. Einleitung **1**

2. Literaturüberblick und Forschungsstand **4**

2.1. Erste Annäherungen an die Wahrnehmung durch Musiker 4

2.1.1. Beranek (1962) 4

2.2. Experimentelle Untersuchungen 5

2.2.1. Barron (1978) 5

2.2.2. Marshall et al. (1978) 6

2.2.3. Gade (1989a,b, 1992) 7

2.2.4. Chiang et al. (2003) 10

2.2.5. Ueno und Tachibana (2003, 2005) 11

2.2.6. Dammerud (2009) 14

2.2.7. Van den Braak und van Luxemburg (2008) 16

2.3. Erkenntnisse der Auditoriumsakustik 17

2.4. Zusammenfassung 18

3. Experimentelle Studie **21**

3.1. Fragestellung und Forschungsmethode 21

3.2. Simulationsumgebung und Versuchsaufbau 23

3.2.1. Pegelkalibrierung 24

3.2.2. Kopfhörer- und Außenohrentzerrung 25

3.3. Bedingungsvariationen 26

3.3.1. Auralisierte Räume 27

3.3.1.1. BRIR-Generierung 27

3.3.1.2. Raumakustische Parameter 32

3.3.2. Instrumente 37

3.3.3. Musikalische Passagen 37

3.4. Stichprobenumfang und -beschreibung 38

3.5. Messinstrument 39

3.5.1. Repertory-Grid-Technik 40

3.5.2. RGT-Vorversuch zur Fragebogenerstellung 42

3.5.3. Zusammenstellung der Ratingitems 44

3.6. Versuchsablauf 57

3.7. Statistische Auswertung des Ratingversuchs 57
3.7.1. Darstellung der Ratingergebnisse 58
3.7.2. Überprüfung des Messinstruments 63
3.7.3. Untersuchung der Urteilerreliabilität bzw. -konsistenz 65
3.7.4. Untersuchung des Instrumenteneinflusses 73
3.7.5. Zusammenhänge zwischen Bewertungsitems 78
3.7.6. Korrelationen zwischen raumakustischen Parametern und Items . . . 83
3.7.7. Untersuchung der Teststärke (Power) 88
3.8. Diskussion 92
3.8.1. Zusammenfassung und Interpretation der Ergebnisse 92
3.8.2. Technische Fehlerquellen 95

4. Ausblick **96**

Abbildungsverzeichnis **98**

Tabellenverzeichnis **100**

Literaturverzeichnis **102**

A. Versuchsanleitung **108**

B. Quellcode **110**

C. Eigenschaften der auralisierten Raummodelle **116**
C.1. Schroeder-Plots 116
C.2. Raumakustische Parameter 118

D. Daten der statistischen Auswertung **123**
D.1. Varianzanalysen mit Messwiederholung 123
D.1.1. MANOVA mit z-transformierten Daten; I1–I20 123
D.2. ANOVAs mit z-transformierten Daten; I1-I20 123

1. Einleitung

Bezüglich der Zuhörerperspektive in Konzertsälen wurde die Wahrnehmung und Bewertung raumakustischer Begebenheiten seit Mitte des zwanzigsten Jahrhunderts intensiv untersucht. Modelle, welche der persönlichen Erfahrung entstammen (siehe Beranek, 1962), wurden durch experimentelle Untersuchungen verifiziert und angepasst (siehe Hawkes und Douglas, 1971). Zusammenhänge zwischen Wahrnehmungsdimensionen und physikalisch messbaren Parametern wurden hergestellt (siehe Schroeder et al., 1974; Lehmann und Wilkens, 1980) und Empfehlungen anzustrebender Wertebereiche dieser Parameter für verschiedene Aufführungsstätten in Richtlinien für die raumakustische Planung festgehalten (siehe DIN EN ISO 3382-1:2009; DIN 18041:2004-05). Nach DIN EN ISO 3382-1:2009 besteht beispielsweise ein Zusammenhang zwischen dem subjektiv wahrgenommenem Schallpegel und dem Stärkemaß G, zwischen der wahrgenommenen Halligkeit und der frühen Abklingzeit EDT sowie der wahrgenommenen Transparenz des Schallfelds und dem Klarheitsmaß C_{80}. Diese hergestellten Zusammenhänge beziehen sich dabei einzig und allein auf die Wahrnehmung durch das Publikums im Auditorium.

Um die Wahrnehmung der Raumakustik durch Musiker bzw. Musikerinnen[1] auf der Bühne nachvollziehen zu können, sind separate Untersuchungen notwendig, welche die Eigenheiten der Aufführungssituation miteinbeziehen. Im Vergleich zur Zuhörersituation unterscheidet sich die Balance zwischen direktem und reflektiertem Schall aufgrund der geringeren Distanz zum Instrument sehr deutlich. Zudem befinden sich Musiker im Gegensatz zu den Rezipienten in einem Schaffensprozess, bei welchem das akustische Ereignis erst erzeugt wird. Der Aufführungsraum kann in diesem Prozess als Erweiterung des Instruments betrachtet werden, durch welchen Musiker ihr Spiel und das ihrer Kollegen wahrnehmen (vgl. Gade, 1989a). Aufgrund persönlicher Berichte (vgl. Beranek, 1962) und experimenteller Untersuchungen (vgl. Schärer Kalkandjiev und Weinzierl, 2013) kann ein Einfluss der Raumakustik auf die Interpretation der Musiker und Dirigenten als gegeben angenommen werden. Um diese Effekte en détail untersuchen zu können, sind Kenntnisse über die Wahrnehmung und Bewertung und die Zusammenhänge zu physikalisch messbaren Prädiktoren hilfreich. Notwendig erscheinen diese sogar für die raumakustische Planung von Konzertsälen, um möglichst optimale Bedingungen für Musiker herstellen zu können, die schlussendlich wiederum zu einer Verbesserung der musikalischen Aufführung führen.

[1] Die weibliche Form wird aus Gründen der vereinfachten Darstellung und einem besseren Lesefluss im Folgenden unterschlagen. In jedem Fall ist dabei jedoch implizit auch die entsprechende weibliche Person gemeint.

Wie sich die Wahrnehmung und Bewertung der raumakustischen Verhältnisse durch Musiker im Einzelnen gestaltet, wurde seit den späten 70er Jahren des zwanzigsten Jahrhunderts untersucht. Nach ersten Publikationen von Marshall et al. (1978) und Barron (1978) präsentierte Gade (1989a,b) eine ausführliche Forschungsarbeit, welche sowohl Interviews, Versuche im Labor mit Solisten und Kammermusikgruppen als auch Studien im Feld mit zwei Orchestern beinhaltete. Zwei im Laufe dieser Arbeit entwickelte und in der Folge überarbeitete (siehe Gade, 1992) akustische Parameter ST_{early} und ST_{late} wurden daraufhin in die DIN EN ISO 3382-1:2009 als Prädiktoren für die Ensemblebedingungen (das gegenseitige Hören der Instrumente) und die wahrgenommene Halligkeit durch Musiker aufgenommen. Weitere bühnenakustische Studien folgten, im Besonderen hervorzuheben sind die Untersuchungen von Ueno und Tachibana (2003, 2005, 2010), welche unter anderem eine sehr aufwendige Simulationsumgebung für die raumakustische Umgebung für Musiker vorstellten.

Es konnte gezeigt werden, dass neben dem Höreindruck des Rezipienten natürlich auch der des Musikers durch raumakustische Veränderungen beeinflusst wird. Die Raumakustik bedingt unter anderem die Hörbarkeit des eigenen Instruments sowie der Instrumente eventueller Ensemblekollegen und hat somit entscheidenden Anteil an der Qualität der aufgeführten Musik. Bei der Frage, welche physikalischen Eigenschaften eines Raums sich auf welche Bewertungsaspekte von Musikern auswirken, kommen die verschiedenen Studien der Bühnenakustik allerdings nicht zu einheitlichen Ergebnissen. Beispielsweise zeigt sich bei Gade (1989a) ein Einfluss der Stärke früher Reflexionen auf die von den Musikern bewertete Unterstützung des Raums, während sich bei Ueno und Tachibana (2003) vor allem späte Reflexionen auf diesen subjektiven Aspekt auswirken.

Mögliche Defizite dieser Studien könnten zum einen eine zu geringe Varianz der zu bewertenden raumakustischen Umgebungen, zum anderen aber auch die Verwendung von Bewertungsaspekten sein, welche von den Musikern nicht richtig verstanden bzw. nicht wahr genommen wurden. Um die bisherigen Ergebnisse des Forschungsfeldes Bühnenakustik besser einordnen und hinterfragen zu können, wurde eine experimentelle Studie durchgeführt, in welcher Musiker die raumakustische Umgebung während des Spiels bewerten sollten. Die Bewertungsattribute wurden dabei in einem Vorversuch durch die *Repertory Grid Technique* eliziert und nicht vom Versuchsleiter vorgegeben. Im Hauptversuch wurden anschließend 10 raumakustische Umgebungen anhand einer Auswahl der elizierten Ratingskalen von professionellen Musikern bewertet. Durch den Einsatz dynamischer Binauralsynthese und akustischer Simulationssoftware konnten die Versuche im Labor stattfinden, wodurch ein schneller Wechsel zwischen den zu bewertenden Räumen ermöglicht werden konnte. Auf diese Weise sollte es den Musikern ermöglicht werden, auch geringe Unterschiede zwischen den Raumsimulationen in ihre Bewertungengen miteinfließen zu lassen. Die über Kopfhörer dargebotenen „Räume“ stellten dabei bewusst Simulationen möglicher prototypischer Konzertsäle dar. Die Stimuligenerierung durch Variation verschiedener Teilbereiche der zur Faltung genutzten Impulsantworten wie bei Ueno und Tachibana (2003) wird als proble-

matisch erachtet und daher nicht verwendet. Es wird vermutet, dass durch diese Methode keine ausreichend große Varianz möglichst vieler raumakutsischer Merkmale gegeben ist und somit auch nicht von Probanden bewertet werden kann.

Da aus bisherigen Studien deutliche Unterschiede zwischen den Aufführungssituationen von Solisten, kleinen und großen Ensemblen hervorgingen, muss die raumakustische Wahrnehmung von Musikern für jede dieser Situationen getrennt untersucht werden. Diese Studie konzentriert sich auf die raumakustische Wahrnehmung durch Solisten. Welche Wahrnehmungsdimensionen in Kammermusikgruppen oder gar in einem Orchester entscheidend sind, kann daher durch das vorliegende Versuchsdesign nicht beantwortet werden.

Bevor die im Rahmen einer Masterarbeit durchgeführte experimentelle Studie vorgestellt wird, soll in Kapitel 2 ein Überblick über die Forschung und Literatur zum Thema Bühnenakustik gegeben werden. Der anschließenden Beschreibung der statitsischen Auswertung und Darstellung der Ergebnisse folgt eine kritische Auseinandersetzung mit der eigenen Arbeit und daraus abgeleitete Empfehlungen für zukünftige Untersuchungen.

2. Literaturüberblick und Forschungsstand

Die Untersuchung akustischer Bedingungen für Musiker auf der Bühne stellt ein komplexes Forschungsfeld dar. Zu den Aspekten, welche die Wahrnehmung jedes einzelnen Musikers beeinflussen, gehören unter anderem dessen Positionierung auf der Bühne, die Richtcharakteristik des Instruments, die Gestalt des Musiksignals sowie die Maskierung des Nachhalls durch das eigene Instrument (vgl. Marshall et al., 1978). Von grundlegender Bedeutung für die Ausprägung des Schallfelds ist ferner die Größe und Zusammenstellung des aufführenden Ensembles.

Experimentelle Studien sowie theoretische Abhandlungen zur raumakustischen Wahrnehmung durch Musiker sollen in diesem Abschnitt zusammengestellt und hinterfragt werden, um eine Grundlage und den Ausgangspunkt für die eigene Untersuchung aufzuzeigen. Dabei ist nicht nur die Darstellung reiner Ergebnisse von Interesse, gleichzeitig sollen die Untersuchungen bezüglich ihres Versuchsdesigns und des technischen Aufbaus in Augenschein genommen werden. Im Mittelpunkt der Betrachtung stehen Solisten-Situationen, sollten jedoch auch Erkenntnisse bezüglich Ensemblebedingungen als nützlich erscheinen, werden auch solche vorgestellt.

2.1. Erste Annäherungen an die Wahrnehmung durch Musiker

2.1.1. Beranek (1962)

Beranek (1962, S. 53–56) widmete sich in seinem Standardwerk *Music, Acoustics & Architecture* in einem relativ kurzen Abschnitt der Beziehung von Künstlern zur Akustik, indem er von Aussagen mehrerer Musiker und Dirigenten berichtet, Zitate aufführt und diese kommentiert. Die Raumakustik habe prinzipiell nicht nur direkten Einfluss auf das musikalische Signal, sondern gleichzeitig einen indirekten, da sie den Musiker und folglich dessen Spielweise beeinflusse. Durch die Aussagen verschiedener Instrumentalisten werden die unterschiedlichen Bedürfnisse deutlich. Der Violinist Issac Stern erachtet neben einem klaren Klang ebenso eine Vermischung sukzessiver Töne durch den Raum als wünschenswert. Ein Pianist hingegen sei durch das ihm zur Verfügung stehende Haltepedal in der Lage, jeweils selbst über die Bindung von Tönen zu entscheiden. Da das Piano außerdem ein an sich laut und nachhallendes Instrument darstelle, sei der Pianist nicht so sehr wie andere Instrumentalisten auf den Raum angewiesen, um die gewünschten Effekte zu erzeu-

gen. Bei der Betrachtung der Musikerzitate ist neben der jeweiligen Aussage das verwendete Vokabular von Interesse, da ein solches für die Bewertung der Raumakustik unabdingbar ist. Stern hat bei wahrgenommenem Anhalten eines Tons durch Raumreflexionen gleichzeitig das Gefühl, dass jede Note umgeben ist von Stärke („surrounded by strength“), sein Spiel nicht nackt („bare or naked“) und jede Note von einer freundlichen Aura umgeben ist („friendly aura surrounding each note“). Beranek beleuchtet ferner die unterschiedlichen Stile der Dirigenten Leopold Stokowski und Sergej Kussewiṭzky und führt diese auch auf die Heimspielstätten in Philadelphia und Boston[1] zurück. Nach Anpassung der Spielweisen ihrer Musiker anhand des eigenen Konzertsaals, empfinden sie die jeweils eigene Spielstätte auch als die jeweils bessere.

Durch Beraneks Abhandlung werden bereits mehrere Aspekte thematisiert: Die Raumakustik beeinflusse das Spiel des Künstlers, verschiedene Instrumente bedürfen unterschiedlicher akustischer Eigenschaften und die Wahrnehmung sei abhängig von der eigenen Erfahrung bzw. Gewöhnung wie im Falle von Stokowski und Kussewiṭzky. Die Richtigkeit und Ausprägung dieser Aussagen gilt es in experimentellen Studien zu überprüfen.

2.2. Experimentelle Untersuchungen

2.2.1. Barron (1978)

In einer Studie von Barron (1978) wurde die Variabilität der Bühnenelemente im Gulbenkian Konzertsaal in Lissabon dazu genutzt, die subjektive raumakustische Wahrnehmung sowohl von Musikern eines kleinen Orchesters als auch von Zuhörern in neun unterschiedlichen Bühnenkonfigurationen abzufragen. Das gespielte musikalische Motiv wurde bei Variation der Orchesterposition, der Höhe des Deckenreflektors, der Ausrichtung der Seitenreflektoren, der Beschaffung der Bühnenrückseite sowie der Neigung des Bühnenbodens nicht verändert. Die Musiker bewerteten die Konfigurationen hinsichtlich der drei Hauptkategorien Gesamteindruck der Akustik (*general impression of the acoustic situation*), Hörbarkeit von anderen und sich selbst (*ability to hear themselves and others*) und Erleichterung für das Spiel (*facility for playing*). Die visuelle Analyse der Ergebnisse suggerierte jedoch in den meisten Fällen die Bewertung auf lediglich einer einzigen Dimension. Über die Instrumentengruppen hinweg wurde die Konfiguration mit tiefer Deckenhöhe präferiert. Wenn auch statistisch nicht signifikant deuten die Ergebnisse darauf hin, dass Streicher prinzipiell eine offene Bühne ohne zu viele frühe Reflexionen bevorzugen. Bläser hingegen favorisierten die Konfigurationen mit nahen reflektierenden Flächen.

Barrons Studie bietet zwar erste Erkenntnisse bezüglich der raumakustischen Wahrnehmung von Musikern, allerdings birgt der Versuch im Feld einige Gefahren. Die Varianz der akustischen Eigenschaften innerhalb der neun verschiedenen Bühnenkonfigurationen ist

[1] Leopold Anthony Stokowski war von 1912 bis 1936 Musikdirektor an der Acadamy of Music in Philadelphia (Beranek, 1962, S. 55), Sergej Alexandrowitsch Kussewiṭzky von 1924 bis 1949 am Symphonieorchester in Boston mit der Spielstätte Symphony Hall tätig (Riemann, 1959, S. 985).

möglicherweise nicht groß genug, um bestehende Effekte aufzudecken. Durch die unterschiedlichen Positionen der Musiker auf der Bühne erleben und bewerten diese zudem nicht die gleiche Akustik. Während diese Herangehensweise für die Bewertung raumakustischer Umgebungen in der Planungsphase sinnvoll ist, ist die Variation der Musikerposition in der Grundlagenforschung als Störvariable einzuschätzen.

2.2.2. Marshall et al. (1978)

Statt einer Untersuchung im Feld führten Marshall et al. (1978) Versuche im reflexionsarmen Raum durch. Dabei spielte ein Streicher gemeinsam mit der über Lautsprecher dargebotenen Musik zweier Kollegen im Trio, während gleichzeitig Raumreflexionen über Lautsprecher befindlich über, hinter, rechts und links des Probanden simuliert wurden. Allerdings berücksichtigte man bei dieser Studie nicht die Raumreflexionen des jeweils eigenen Instruments des untersuchten Musikers, wodurch die Ergebnisse von vornherein in Frage gestellt werden müssen. Im ersten Experiment sollten vier verschiedene Situationen, in welchen je eine Reflexion mit unterschiedlichen Verzögerungszeiten und Pegeln dargeboten wurde, bezüglich der Leichtigkeit des Zusammenspiels (*ease of ensemble*) bewertet werden. Da Verzögerungszeiten von 10 ms und 40 ms abgelehnt, von 20 ms sehr und von 80 ms präferiert wurden, schließen Marshall et al. (1978) auf eine multidimensionale Bewertungsgrundlage bezüglich der Ensemblebedingungen. Als erster Planungsrichtwert für als gut eingeschätzte frühe Reflexionen wird der Bereich zwischen 17 und 35 ms angegeben.

Das zweite Experiment sollte die Hypothese testen, dass hochfrequente Anteile der Raumreflexionen am wichtigsten für Ensemblemusiker sind, da die meiste für Musiker notwendige musikalische Information in den Anschlagstransienten beinhaltet ist, welche Hinweise für Rhythmus und Ausdruck darstellen. Dazu durchliefen die simulierten Raumreflexionen mit einer Verzögerungszeit aus dem zuvor genannten Bereich Hoch- und Tiefpassfilter verschiedener Grenzfrequenzen, die Musiker sollten wiederum ihre Präferenz bezüglich der Ensemblebedingungen angeben. Bei den Bewertungen der hochpassgefilterten Reflexionen zeigte sich, dass diese bei den Grenzfrequenzen von 500 Hz und 1 kHz nahezu identisch ausfielen. Bei der tiefpassgefilterten Raumsimulation war erst ab Grenzfrequenzen von 1 kHz und niedriger eine deutliche Verringerung der Präferenz bezüglich der Ensemblebedingungen zu erkennen. Dass trotzdem Unterschiede der Klangfarbe (*timbre*) wahrgenommen wurden, lasse sich laut Marshall et al. aus Kommentaren der Versuchspersonen schließen.

Da in dieser Untersuchung die Reflexionen des jeweils eigenen Instruments der Musiker nicht simuliert wurde und sich die Fragestellung grundsätzlich auf die Ensemblesituation bezog, sind die Ergebnisse nicht direkt auf die Situation von Solomusikern übertragbar. Außerdem wurden nur einzelne frühe Reflexionen, jedoch kein später Nachhall simuliert und somit war auch keine möglichst authentische raumakustische Umgebung Grundlage der Musikerurteile. Allerdings deutet sich an, dass auch die Dimension der Klangfarbe für Musiker auf der Bühne relevant sein kann.

2.2.3. Gade (1989a,b, 1992)

Weiterführende Untersuchungen zur speziellen akustischen Situation von Musikern auf der Bühne führte Gade (1989a,b) durch. Ausgangspunkte waren eine angenommene Bedeutung des Raums für die musikalische Interpretation wie auch Beschwerden von Musikern bezüglich der Raumakustik solcher Säle, welche vom Publikum als gut empfunden wurden. Gades Studien sind bislang die einzigen, deren Ergebnisse in Normen als Planungsgrundlage für die Bühnenakustik dienen (siehe DIN EN ISO 3382-1:2009).

Um ein gemeinsames Vokabular zu bilden und die Bedürfnisse der Musiker abzufragen, führte Gade (1989a) Interviews mit 32 Künstlern der klassischen Musik (Dirigenten, Pianisten, Sängern und Instrumentalisten verschiedener Orchesterinstrumente) durch. Es wurde nach raumakustisch bedeutenden Elementen gefragt, welche anschließend bezüglich ihrer Wichtigkeit in eine Rangfolge gebracht werden sollten. Mithilfe der Antworten wurde ein Set von sechs subjektiven Aspekten aufgestellt. Die Halligkeit (*Reverberance*) wird vor allem in Spielpausen wahrgenommen, bindet nacheinander folgende Töne, kann Details der Aufführung verwaschen und ein Gefühl der Raumreaktion geben. Die Unterstützung des Raums (*Support*) ermöglicht den Musikern das Hören des eigenen Instruments. Da diese auch während des Anspielens der Töne wahrgenommen werden kann, nimmt Gade an, dass sie mit anderen physikalischen Eigenschaften in Verbindung steht als die wahrgenommene Halligkeit. Die Klangfarbe des Raums (*Timbre*) beeinflusst die wahrgenommene Klangfarbe des Instruments und die Balance zwischen den Registern. Die Dynamik des Raums (*Dynamics*) beschreibt die Spannweite der wahrgenommenen Lautstärke zwischen *ppp* und *fff*. Das Gegenseitige Hören (*Hearing each other*) stellt die Eigenschaft dar, welche eine Gruppe von Musikern für das Zusammenspiel benötigt. Dabei ist auch die Balance zwischen dem eigenen und allen anderen beteiligten Instrumenten entscheidend. Die Zeitverzögerung von verschiedenen Instrumenten (*Time Delay*) ist bedingt durch die Geschwindigkeit des Schalls und kann bei großen Orchestern und demzufolge weit entfernt sitzenden Musikern zu Problemen führen.

Da Gade eine Verbindung zwischen der Stärke früher Reflexionen und dem wahrgenommenen Effekt der Unterstützung (*Support*) vermutete, wurde desweiteren die Wahrnehmungsschwelle einer frühen Deckenreflexion des eigenen Instruments, welche durch einen Lautsprecher in drei Metern Höhe simuliert wurde, bei Solisten gemessen. Bei je sechs verschiedenen Verzögerungszeiten bestimmte man die Wahrnehmungsschwelle als denjenigen Pegel, bei welchem die Probanden in einem Two-Alternative-Forced-Choice-Test (2AFC) zu einer Wahrscheinlichkeit von 50 % angaben, eine Reflexion wahrgenommen zu haben. In einem zweiten Experiment wurde anstatt einer einzelnen Reflexion die Wahrnehmungsschwelle eines Bündels von sechs frühen Reflexionen aus unterschiedlichen Richtungen mit Verzögerungszeiten unter 100 ms gesucht. Wie Abbildung 2.1 zeigt, wurde bei Streichern eine um etwa 5 dB höhere Wahrnehmungsschwelle als bei Flötisten festgestellt. Gleichzeitig traten zwischen der Wahrnehmung von Cellisten und Violinisten keine signifikanten Unter-

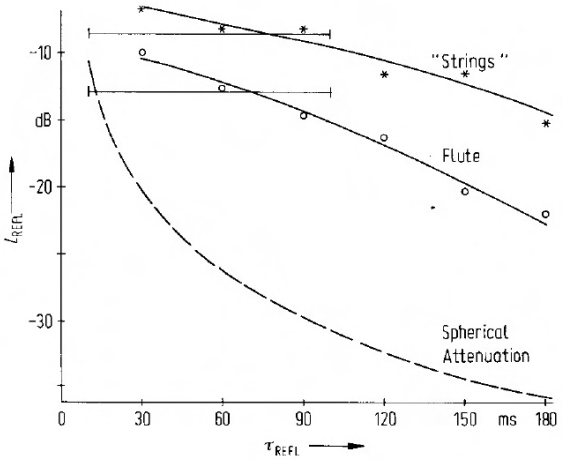


Abbildung 2.1.: Gefundene Wahrnehmungsschwelle einer Deckenreflexion des eigenen Instruments in Abhängigkeit von der Verzögerungszeit für Streicher und Flötisten. Horizontale Linien: Wahrnehmungsschwelle für eine Gruppe von sechs frühen Reflexionen zwischen 20 und 100 ms. Gestrichelte Kurve: Pegel einer einzelnen kugelförmig gedämpften Reflexion an einer flachen, harten Oberfläche. Die Ordinate beschreibt den Pegel in dB relativ zum Pegel des Direktschalls in 1 m Entfernung zur Quelle (Gade, 1989a).

schiede auf, weshalb diese in der Auswertung zusammengefasst wurden. An welcher Stelle der Pegel des Direktschalls sowie der frühen Reflexionen gemessen wurde, geht aus den Beschreibungen des Versuchs (siehe Gade, 1989a) leider nicht hervor. Somit ist nicht sicher, ob der unterschiedliche Abstand zwischen Ohren und Schallquelle bei den verschiedenen Instrumenten berücksichtigt wurde.

Nach einem Vergleich der gefundenen Schwellen mit gemessenen Werten in Konzerthäusern zeigte sich, dass die Energie der frühen Reflexionen die gefundene Wahrnehmungsschwelle in manch großem Saal nicht überschreitet, weshalb Gade in solchen Sälen an einem hörbaren, bzw. unterstützenden Effekt früher Reflexionen zweifelt. In großen Orchestern werde zudem der Aspekt der Unterstützung von Musikern überhaupt nicht wahrgenommen, da die frühen Reflexionen durch den Direktschall der anderen Instrumente stark verdeckt würden. Dies gelte insbesondere für Musiker von Streichinstrumenten, deren gefundene Wahrnehmungsschwelle bezüglich früher Reflexionen sehr hoch ist und deren Instrumente einen vergleichsweise geringen Schallpegel erzeugen.

In einem weiteren Versuch wurden vier Schallfelder, welche im Pegel früher Reflexionen sowie im Pegel des späten Nachhalls (ab 100 ms) variierten, mithilfe von Paarvergleichen bezüglich der Präferenz des Klangs des eigenen Instruments bewertet. Dabei spielten die Musiker gemeinsam in einem Trio (Flöte, Violine und Cello). Ergebnissen einer MDPREF-Analyse[2] zufolge, deutete die erste Dimension (*consensus dimension*) auf eine Präferenz der

[2]MDPREF stellt eine interne Analyseform des Präferenz-Mappings dar, welche anhand von Präferenzdaten gewonnen aus Paarvergleichen, Rangfolgen oder Ratings durchgeführt wird. Dabei werden die Stimuli als Punkte in einem mehrdimensionalen Raum dargestellt, während die Präferenz der Versuchspersonen

jeweils höchsten der untersuchten Pegel sowohl für frühe als auch für späte Reflexionen hin.

Um die raumakustische Wahrnehmung durch Musiker im Ensemble zu untersuchen, ließ Gade (1989b) Musiker eines britischen und eines dänischen Symphonieorchesters Konzertsäle auf einem semantischen Differential, welches die zuvor aufgestellten Aspekte beinhaltete (vgl. Gade, 1989a), bewerten. Eine anschließend durchgeführte Faktorenanalyse elizierte in beiden Stichproben zwei Faktoren, welche in der dänsichen Stichprobe 90 % und in der britischen 97 % der Gesamtvarianz erklärten. Die Bewertungsitems *Support*, *Dynamics*, *Reverberance*, *Hearing others in ensemble*, *Ensemble generally*, *Hearing oneself in ensemble* zeigten hohe Ladungen auf Faktor 1. *Timbre* lud als einziges Item hoch auf dem zweiten Faktor.

Desweiteren definierte Gade raumakustische Maße, welche einen Anhaltspunkt für die Planung von Bühnen darstellen sollten. *ST1* und *ST2* (Gade, 1989b) wurden in experimentellen Studien wie auch in der beruflichen Praxis getestet und in ST_{early}, ST_{total} und ST_{late} überführt. Diese drei Supportmaße beziehen sich auf die gemessene Impulsantwort auf der Bühne bei einer Entfernung von einem Meter zwischen omnidirektionaler Quelle und Empfänger (Gade, 1992).

ST_{early} (anfangs als *ST1* bezeichnet) beschreibt das Verhältnis der Energie früher Reflexionen zur Energie des Direktschalls und sollte nach Gade (1989a) ein Maß für die wahrgenommene Unterstützung des eigenen Spiels darstellen. In der Praxis erwies sich der Parameter jedoch als guter Prädiktor für das Hören anderer Orchestermitglieder (Gade, 1992), weshalb ST_{early} in der DIN EN ISO 3382-1:2009 in Bezug auf die frühe Unterstützung des Ensembles genannt wird. Die Ergebnisse von Gade (1989b) lassen auf einen optimalen Bereich von −13 bis −11 dB für ST_{early} schließen (siehe auch Barron, 2010, S. 61). ST_{late} soll die von Musikern wahrgenommene Halligkeit vorhersagen und ging als zweite Bühnen-Messgröße in die DIN EN ISO 3382-1:2009 ein. ST_{total} beschreibe desweiteren die Unterstützung des eigenen Instruments (Gade, 1992), die Aufgabe die Gade (1989a) noch ST_{early} zukommen ließ. ST_{total} findet in der Norm jedoch keine Berücksichtigung.

$$ST1 = ST_{early} = 10 \cdot \lg \left[\frac{\int_{20}^{100} p^2(t)\,\mathrm{d}t}{\int_{0}^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB} = 10 \cdot \lg \frac{E(20, 100\,\mathrm{ms})}{E(0, 10\,\mathrm{ms})}\ \mathrm{dB} \tag{2.1}$$

$$ST_{late} = 10 \cdot \lg \left[\frac{\int_{100}^{1000} p^2(t)\,\mathrm{d}t}{\int_{0}^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB} = 10 \cdot \lg \frac{E(100, 1000\,\mathrm{ms})}{E(0, 10\,\mathrm{ms})}\ \mathrm{dB} \tag{2.2}$$

$$ST_{total} = 10 \cdot \lg \left[\frac{\int_{20}^{1000} p^2(t)\,\mathrm{d}t}{\int_{0}^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB} = 10 \cdot \lg \frac{E(20, 1000\,\mathrm{ms})}{E(0, 10\,\mathrm{ms})}\ \mathrm{dB} \tag{2.3}$$

bezüglich dieser Stimuli durch Vektoren repräsentiert wird. Je kleiner der Abstand zwischen Stimulus und Versuchspersonenvektor ist, desto besser wurde der Stimulus von der Versuchsperson bewertet bzw. eingestuft (vgl. Marketo et al., 1994).

mit	$p(t)$	Momentanwert des Schalldrucks der Impulsantwort
	$t = 0$	Zeitpunkt des Eintreffens des Direktschalls
	$E(t_1, t_2)$	Energie der Impulsantwort zwischen t_1 und t_2, äquivalent zum Integral aus quadriertem Momentanschalldruck

2.2.4. Chiang et al. (2003)

Chiang et al. (2003) untersuchten die raumakustische Wahrnehmung auf der Bühne in fünf verschiedenen Konzertsälen mit je vier verschiedenen raumakustischen Konfigurationen. In jeder dieser 20 Situationen spielten je drei Solisten (Piano, Violine, Trompete) und drei Ensembles (Violine-Piano-Duett, Violine-Cello-Piano-Trio, Bläser-Quintett) je drei musikalische Exzerpte. Nach Spielen jedes Exzerpts bewerteten die Kammermusikensembles mithilfe von Ratingskalen die Hörbarkeit des eigenen Instruments (*hearing oneself*), das Hören der anderen Instrumente (*hearing others*), die Leichtigkeit des Zusammenspiels (*ease of ensemble*) sowie den Gesamteindruck (*overall impression*). Die Solisten wurden nur bezüglich des Gesamteindrucks befragt, da vorangegangene Studien auf die Bewertung auf einer einzelnen Dimension schließen ließen. Die Möglichkeit zur Überprüfung dieser Annahme wurde somit leider von vornherein verhindert.

Um die Probandenstruktur zu untersuchen, wurde ausgehend von den Bewertungen des Gesamteindrucks eine Hauptkomponentenanalyse durchgeführt, welche nach Varimax-Rotation zwei Faktoren mit einem Eigenwert größer 1 zum Ergebnis hatte. Die Streicher konnten dabei dem ersten, die Bläser dem zweiten Faktor zugeordnet werden. Die geringe Varianzaufklärung der beiden Komponenten von lediglich 53 % deutet für Chiang et al. auf eine geringe Übereinstimmung der Probanden hin.

Bei Betrachtung der Korrelationen der über die Musiker gemittelten Bewertungen zeigte sich ein hoher Zusammenhang zwischen Gesamteindruck der Solisten und der Hörbarkeit des eigenen Instruments in den Ensemblesituationen ($r = 0,82$). Der Gesamteindruck im Ensemble und der Solisten korrelierte hingegegen nur mit einem Wert von $r = 0,61$. Dies ist eventuell darauf zurückzuführen, dass der Gesamteindruck im Ensemble von der Kommunikationsfähigkeit zwischen den Musikern (Hören anderer Musiker $r = 0,92$, Leichtigkeit des Zusammenspiels $r = 0,87$) dominiert wurde. Bei der Betrachtung der Zusammenhänge des Gesamteindrucks der Solisten mit raumakustischen Parametern zeigten sich zwar mittlere signifikante Korrelationen, allerdings lassen diese aus Sicht des Autors eben aufgrund der Beschränkung auf eine einzige Bewertungsdimension nur schwierig Deutungen zu.

Bezüglich der Zusammenstellung der Ensembles ist anzumerken, dass Violine und Piano jeweils von den gleichen Musikern gespielt wurden, wodurch diese mehrfach getestet wurden. Das Versuchsdesign ist daher grundsätzlich in Frage zu stellen, weil der personenspezifische Fehler nicht als zufällig angenommen werden kann.

2.2.5. Ueno und Tachibana (2003, 2005)

Ueno und Tachibana (2003) untersuchten die raumakustische Wahrnehmung durch Musiker in Laborversuchen mithilfe einer sechskanaligen, dreidimensionalen Schallfeldsimulation im reflexionsarmen Raum. Das durch ein gerichtetes Mikrofon abgenommene Instrumentensignal wurde nach Faltung mit einer Impulsantwort über die 2 m vor, hinter, über, unter, links und rechts der Hörposition befindlichen Lautsprecher wiedergegeben. Die Impulsantworten wurden auf Konzertsaalbühnen mithilfe eines bis 2 kHz omnidirekionalen Dodekaeders und eines gerichteten und nacheinander zu sechs verschiedenen Seiten ausgerichteten Mikrofons in kurzer Entfernung gemessen. Der Direktschall wurde aus der Impulsantwort entfernt, da dieser aufgrund des eigenen Musizierens der Probanden nicht mehr simuliert werden sollte. Auch die erste Reflexion des Bühnenbodens wurde nicht simuliert, da die Verzögerung von $8 - 9$ ms durch den 2 m entfernten Lautsprecher nicht simuliert werden konnte. Zur Berechnung raumakustischer Parameter wurde an gleicher Stelle eine Impulsantwort mit omnidirektionalem Empfänger gemessen. Durch eine weitere Messung einer monauralen Impulsantwort am Hörpunkt der Simulationsumgebung konnten die realen und simulierten Schallfelder anhand der raumakustischen Parameter verglichen werden.

Um die vorgestellte akustische Simulationsumgebung zu überprüfen, wurden Musiker in einem Vorversuch bezüglich ihres Höreindrucks abgefragt. Neben allgemeiner Akzeptanz des Systems und vereinzelter Kritik bezüglich der Klangqualität bei hohen Frequenzen beobachteten Ueno und Tachibana, dass die Musiker zu einer detaillierten Beschreibung des Schallfelds fähig waren. Die beschreibenden Formulierungen fielen dabei allerdings sehr individuell aus.

Im Hauptversuch wurde ausgehend von einer sechskanaligen Impulsantwort eines Konzertsaals der Pegel früher Reflexionen (im Bereich 10–100 ms), die Nachhallzeit sowie die Stärke hinzugefügter später Reflexionen mit einer Verzögerung im Bereich von 250 ms aus Richtung des Auditoriums variiert. Zwölf Musiker von Streich- und Holzblasinstrumenten spielten frei wählbare Stücke, um die dargebotene Akustik kennen zu lernen, sie in eigenen Worten zu beschreiben, mehrere Items auf einer 5-Punkt-Skala zu bewerten und die Stufen der veränderten Variablen der Impulsantwort zu vergleichen. Die Ergebnisse der Itembewertungen waren geprägt von starken interindividuellen Abweichungen, weshalb Ueno und Tachibana eine statistische Auswertung dieser Daten zur Interpretation der Probandenbewertungen nicht für sinnvoll erachteten. Welche Items abgefragt wurden oder wie sehr die Bewertungen der einzelnen Probanden voneinander abwichen, wurde nicht angegeben. Die Auswertung der Präferenzbewertungen ergab, dass die Streicher längere Nachhallzeiten als die Holzbläser präferierten, die schwächsten frühen Reflexionen ($ST\,1 = -23{,}7$ dB) von den meisten Musikern als am besten eingestuft wurden und die Holzbläser hörbare späte Reflexionen bevorzugten. Die Gründe für die beobachteten Abweichungen in den Bewertungen der Streicher vermuteten Ueno und Tachibana in der unterschiedlichen Zugehörigkeit der Versuchspersonen zu den Stimmen im Orchester („soloist, ‘top,’ or ‘tutti’ “(Ueno und Ta-

chibana, 2003, S. 137)). Im Gegensatz zu den Erkenntnissen von Gade (1989a) erfuhren die Versuchspersonen bei Ueno und Tachibana eine unterstützende Wirkung eher durch späte als durch frühe Reflexionen. Bei angemessender Stärke der späten Reflexionen wurden diese nicht als störend sondern als vorteilhaft aufgefasst. Ein möglicher Grund für diese Unterschiede kann darin liegen, dass sich bei Gade die unterstützende Wirkung früher Reflexionen im Ensemble herausstellte, Ueno und Tachibana hingegen die Solisten-Situation untersuchten.

Obwohl die raumakustische Simulationsumgebung von Ueno und Tachibana im Bereich der Bühnenakustikforschung eine herausragende Stellung bezüglich des Aufwands und der Simulationsergebnisse darstellt, sollen hier gerade im Hinblick auf die eigene Arbeit auch etwaige Probleme aufgezeigt werden. Durch Verwendung einer omnidirektionalen Quelle bei der Messung der Impulsantworten wird die jeweilige Richtcharakteristik der Instrumente vernachlässigt. Inwieweit diese Vereinfachung problematisch ist, soll in einem späteren Abschnitt dieser Arbeit diskutiert werden. Bei Wiedergabe des simulierten Schallfelds über Lautsprecher kommt es desweiteren zu einer Rückkopplung. Ueno und Tachibana nehmen an, dass die gemessenen Pegel im simulierten Schallfeld in den Oktavbändern 125 Hz und 250 Hz aus diesem Grund und aufgrund der Richcharakteristik[3] des verwendeten Mikrofons bis zu 3 dB höher waren als auf den realen Bühnen. Hier sei angemerkt, dass eine Rückkopplung durch das dreidimensionale Lautsprechersetup nicht nur durch Ansprache des Mikrofons aus 180°-Richtung erfolgt sein müsste, sondern auch aus den jeweiligen Richtungen der sechs Lautsprecher. Aus diesem Grund sind weitere hörbare Artefakte dieser Rückkopplungen nicht auszuschließen. Ein weiteres mögliches Problem des Systems stellt die fixe Positionierung des Direktschall–Mikrofons dar. Bei Bewegungen des Musikers und seines Instruments verändert sich der Abstand zwischen Quelle und Empfänger und somit der Eingangspegel des Systems. Die Variabilität des Eintreffwinkels des akustischen Instrumentensignals kann gleichzeitig aufgrund der Nieren-Richtcharakteristik des Mikrofons Veränderungen der Klangfarbe bedingen. Bei hohen Frequenzen um die 10 kHz kommt es bei einem Eintreffwinkel von 90° im Vergleich zum 0°-Winkel zu einer Dämpfung von 10 dB.

Basierend auf Interviews mit 13 professionellen Musikern stellten Ueno und Tachibana (2005) ferner ein kognitives Modell zur Wahrnehmung in Konzertsälen durch Musiker vor. In offenen Gesprächen wurden die Musiker zu ihrer musikalischen Erfahrung, ihrer Wahrnehmung während des Konzerts und bezüglich des Konzertsaals sowie der Kommunikation mit anderen Musikern über Konzerthäuser befragt. Sie gaben an, neben ihrem Instrument auch die Eigenschaften eines Konzertsaals für den musikalischen Ausdruck zu nutzen. Je nach vorhandener Akustik stellten sie ihr Spiel auf diese ein – manche gaben an, die notwendigen Anpassungen im Laufe jahrelanger Erfahrung erlernt zu haben. Gleichzeitig berichteten die

[3] Ueno und Tachibana (2003) geben an, dass das Instrumentensignal mit dem unidirektionalen Mikrofon *Sony C48* abgenommen wurde. Es handelt sich dabei um ein Doppelmembran-Kondensatormikrofon mit umschalbarer Richtcharakteristik (siehe im Literaturverzeichnis unter „Datenblätter und Handbücher“). Den Angaben zufolge wurde die Nierencharakteristik genutzt.

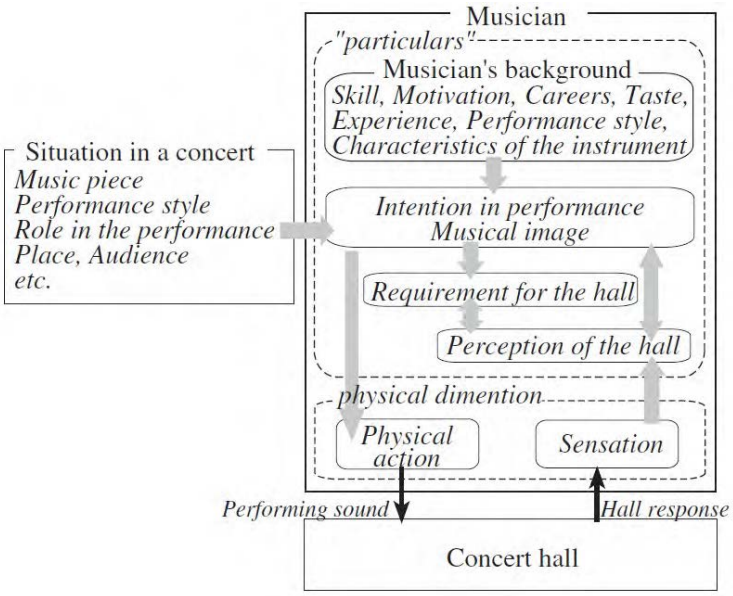


Abbildung 2.2.: Kognitives Modell des impliziten Wissens bei der Musikaufführung (Ueno und Tachibana, 2005).

Musiker, dass sie nicht auf die Akustik eines Raums als ganzen achten, sondern vielmehr darauf, wie sie in diesem mit ihrem jeweiligen Instrument zu spielen haben. Bei der Frage, ob dieser Prozess logisch oder bewusst erfolgt, kommt es zu Uneinigkeiten.

Die Kommunikation der Musiker bezüglich der Akustik bezieht sich folglich ebenfalls meist auf die Art und Weise, wie in einem Saal gespielt werden muss – eine direkte Beschreibung der Raumakustik findet daher meist nicht statt. Folgendes Zitat beschreibt die Schwierigkeiten bei der Abfrage wahrgenommmener raumakustischer Qualitäten von Musikern sehr präzise:

> Here, it should be noted that knowing is different from telling. For example, musicians cannot necessarily express their perception of acoustics of a concert hall by words even if they have clear perception.
>
> (Ueno und Tachibana, 2005, S. 158).

Ausgehend von den Interviews und mithilfe der Theorie des impliziten Wissens wird illustriert, wie ein Musiker beeinflusst durch den jeweiligen Hintergrund, die individuelle Erfahrung und die Konzertsituation die Raumakustik eines Raums wahrnimmt und sein Spiel an diesen anpasst. Wie Abbildung 2.2 zeigt, stehen die *Einzelheiten* (*particulars*) dieses Lernprozesses dabei in komplexer wechselseitiger Beziehung. Aus dem Ganzen der *Einzelheiten* resultiert die ausgeübte Handlung. Neben dem Instrument wird somit auch die Raumakustik des Konzertsaals als Werkzeug begriffen, das es zu beherrschen gilt.

Ausgehend von der bereits dargestellten Versuchsumgebung (Ueno und Tachibana, 2003) übertrugen Ueno und Tachibana (2005) das Modell auf die Situation des Musikers im Experiment. Dabei werden zusätzlich zum Prozess des impliziten Wissens des Musikers das

komplexe System des Versuchsleiters und die Wechselwirkungen zwischen beiden Personen beschrieben. Demzufolge müssen bei Abfrage der raumakustischen Wahrnehmung alle *Einzelheiten* berücksichtigt werden. Die Aufführungsssituation, welche im Experiment nicht besteht, müsse den Versuchspersonen durch den Versuchsleiter erklärt werden, wodurch bei den Musikern die Vorstellung eines Konzerts erreicht werden könne. Da je nach Ausprägung der verschiedenen *Einzelheiten* die Wahrnehmung eines Konzertsaals und dessen Beschreibung durch Musiker variieren würden, befragten Ueno und Tachibana die Probanden in Einzelgesprächen, um auf alle für sie wichtigen persönlichen Aspekte der Musiker eingehen zu können. Bei der Abfrage der Wahrnehmung mithilfe einheitlicher Skalen müsse hingegen überprüft werden, ob die benutzten Begriffe unter den befragten Musikern auch die gleiche Bedeutung besäßen.

Das kognitive Modell von Ueno und Tachibana stellt ein hilfreiches Werkzeug dar, das bei der eigenen Versuchsplanung Beachtung finden sollte. Die bereits in mehreren Studien genannten interindividuellen Unterschiede zwischen den Bewertungen der Wahrnehmung der raumakustischen Umgebung durch Musiker finden durch das Modell und die darin aufgenommenen *Einzelheiten* Erklärungsansätze. Die notwendige Transformation von implizitem Wissen der individuellen Musiker in explizites Wissen, welches sich durch ein gemeinsames Vokabular auszeichnet und gleichzeitig unabhängig von den jeweiligen persönlichen *Einzelheiten* funktioniert, stellt vermutlich die größte Herausforderung für die empirische Bühnenakustik dar.

2.2.6. Dammerud (2009)

Dammerud widmete sich in seiner Doktorarbeit den akustischen Bedingungen von Symphonieorchestern auf Konzertsaalbühnen, den Auswirkungen der Akustik auf die Situation der Musiker und den daraus abzuleitenden Schlüssen für die Raumakustikplanung. Im Zuge der Arbeit wurde eine empirische Studie durchgeführt, in welcher die Musiker des Bournemouth Symphony Orchestra in einem Feldversuch acht britische Konzertsäle auf einem semantischen Differential bewerteten (Dammerud, 2009, S. 151–172). Die Bewertungen der Musiker auf den folgenden sieben Ratingskalen gingen in die finale Auswertung ein:

- Physical comfort (Co): Uncomfortable – Comfortable
- Hearing self (HS): Insufficiently – Sufficiently
- Hearing others (HO): Severe problem – Not a problem
- Clarity (Cl): Struggle to hear details – Easy to hear details
- Hall reverberance (Rev): Dead, no response – Live, reverberant
- Other instruments not too loud (LNP): Severe problem – Not a problem
- Overall impression (OAI): Very poor – Very good

Neben diesen Items wurden weitere Aspekte bezüglich der wahrgenommmen Klangfarbe, des Nachhallcharakters, der Leichtigkeit eines zu spielenden pianissimo und fortissiomo sowie der Wahrnehmbarkeit einzelner später Reflexionen abgefragt. Da diese Aspekte jedoch

sehr hohe Korrelationen mit den oben genannten zeigten, wurden sie nicht in die Analyse miteinbezogen. An der Studie nahmen insgesamt 24 Musiker teil, darunter 16 Streicher, drei Holzbläser, vier Blechbläser und ein Perkussionist. Sie bewerteten, wie bereits erwähnt, jeweils acht Räume mithilfe der oben genannten fünfstufigen Ratingskalen. Das Orchester, welchem die Versuchspersonen angehörten, trat regelmäßig in den zu bewertenden Konzertsälen auf, die Musiker hatten ihre Eindrücke daher womöglich über eine längere Zeitspanne gesammelt. Dies kann neben dem Vorteil der intensiven Spielpraxis in den Räumen auch zu Problemen führen. Zwar waren die Fragen des semantischen Differentials eindeutig auf die akustische Situation bezogen, doch könnten beispielsweise der Austausch mit anderen Orchestermitgliedern, die allgemeine Resonanz des Publikums in den Konzerthäusern, einzelne besonders gute oder schlechte Aufführungen und somit nicht-akustische Aspekte Einfluss auf die Bewertungen genommen haben.

Die Unterschiede zwischen den Bewertungen der acht Konzertsäle wurden anhand von t-Tests auf Basis der Mittelwerte aller Versuchspersonen je Ratingskala ermittelt. Zwischen den vier meistpräferierten Sälen zeigten sich bezüglich der Items HS (*Hearing self*), HO (*Hearing others*), Cl (*Clarity*) und Rev (*Hall reverberance*) keine signifikanten Unterschiede ($\alpha = 0,05$), gleiches wurde zwischen den zwei am wenigsten präferierten beobachtet. Die Ergebnisse legten nahe, dass sich der Gesamteindruck der Akustik (OAI) am besten zur Unterscheidung der Räume eignete, da sich bei diesem Item die meisten signifikanten Unterschiede zwischen den Bewertungen der Säle ergaben.

Um die Unterschiede zwischen den Instrumentengruppen bezüglich der Bewertungen des Gesamteindrucks zu untersuchen, wurden wiederum t-Tests durchgeführt. Dabei zeigten sich außer beim Perkussionisten keine signifikanten Effekte. Nach Dammerud wies dieses Ergebnis auf eine allgemeine Erfahrung des Gesamteindrucks der Musiker über die Instrumentengruppen hinweg hin. Dementsprechend waren die Standardabweichungen innerhalb der Instrumentengruppen mit der Standardabweichung der gesamten Stichprobe vergleichbar.

Bei der statistischen Auswertung wurden leider keinen Angaben gemacht, welche Art von t-Test durchgeführt wurde[4] und inwieweit der Akkumulierung des Alpha-Fehlers durch mehrere t-Tests mit einer Bonferroni-Korrektur begegnet wurde.

Bei der Untersuchung der Zusammenhänge zwischen Musikerbewertungen und raumakustischen Parametern errechnete sich eine hohe Korrelation zwischen G_l und den Bewertungen des Nachhalls (Rev) mit $r = 0,82$ (sig. bei $\alpha = 0,05$) bei Berücksichtigung aller Räume. Nach Ausschluss der Bewertungen von zwei Theaterbühnen zeigten sich hingegen keine signifikanten Korrelationen mit raumakustischen Parameteren, sondern nur mit architektonischen Maßen.

G_l ist das Stärkemaß des Zeitintervalls von $80\,\mathrm{ms} - \infty$ und wurde von Dammerud (2009, S. 131) neben G_e, dem frühen Stärkemaß, vorgestellt. Die Gleichungen 2.4 und 2.5 zeigen die

[4] Schätzungsweise handelte es sich um Paardifferenztests zwischen den einzelnen Raumbewertungen und zwischen den Bewertungen der einzelnen Instrumentengruppen.

Berechnungsvorschriften für die beiden akustischen Parameter. G_l stellt den Pegel des Nachhalls, der akustischen Antwort aus dem Auditorium, dar. Die Integrationsgrenzen wurden auch aus dem Grund gewählt, da somit eine Berechnung anhand der Oktavbandmittelwerte über 500 Hz und 1000 Hz von G und C_{80}, wie in den Formeln angegeben, möglich ist und da G_l dadurch auch anhand des Volumens und der Nachhallzeit berechnet werden kann. Dadurch werden nachträgliche Berechnungen anhand von Literaturangaben, in welchen die Nachhallzeit und das Volumen häufige Kennzahlen darstellen, ermöglicht (Dammerud, 2009, S. 50).

$$G_e = G_{0-80} = 10 \cdot \lg \left[\frac{\int_{0\,\mathrm{ms}}^{80\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{\infty} p_{10}^2(t)\,\mathrm{d}t} \right] \mathrm{dB} = 10 \cdot \lg \left[\frac{10^{C_{80}/10} \cdot 10^{G/10}}{1 + 10^{C_{80}/10}} \right] \mathrm{dB} \tag{2.4}$$

$$G_l = G_{80-\infty} = 10 \cdot \lg \left[\frac{\int_{80\,\mathrm{ms}}^{\infty} p^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^{\infty} p_{10}^2(t)\,\mathrm{d}t} \right] \mathrm{dB} = 10 \cdot \lg \left[\frac{10^{G/10}}{1 + 10^{C_{80}/10}} \right] \mathrm{dB} \tag{2.5}$$

mit	$p(t)$	Momentanwert des Schalldrucks der Impulsantwort an der Empfängerposition
	$p_{10}(t)$	Momentanwert des Schalldrucks im Freifeld bei gleicher Schallquelle in einer Entfernung von 10 m zum Empfänger
	$t = 0$	Zeitpunkt des Eintreffens des Direktschalls

Bezüglich der Festlegung eines optimalen Wertebereichs gibt Dammerud an, keinen optimalen Wert für den Pegel des späten Nachhalls formulieren zu können (Dammerud, 2009, S. 172). Es scheine neben dem Wert von G_l auch von Bedeutung zu sein, dass die späte Energie aus Richtung des Auditoriums kommt.

2.2.7. Van den Braak und van Luxemburg (2008)

Um die Feinabstimmung der Bühnenakustik in der Praxis besser bewerkstelligen zu können, stellten van den Braak und van Luxemburg (2008) den akustischen Parameter LQ_{7-40} vor, welcher das Verhältnis der Energie der frühen frühen Reflexionen (7–40 ms) sowie der späten frühen Reflexionen und des Nachhalls beschreibt. Ausgangspunkt für die Bestimmung des LQ_{7-40} war die Beratung bezüglich der Raumakustik des Casa da Musica in Porto. Bei Variation des hängenden Deckenreflektors zeigte sich ein hörbarer Unterschied für den Dirigenten (nicht für die Musiker), welcher jedoch nicht mit einer Veränderung des ST_{early} einherging.

$$LQ_{7-40} = 10 \cdot \lg \left[\frac{\int_{7\,\mathrm{ms}}^{40\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{40\,\mathrm{ms}}^{\infty} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB} \tag{2.6}$$

mit	$p(t)$	Momentanwert des Schalldrucks der Impulsantwort an der Empfängerposition
	$t = 0$	Zeitpunkt des Eintreffens des Direktschalls

Zur Validierung des Parameters überprüften van den Braak und van Luxemburg die Zusammenhänge der Werte von LQ_{7-40} und der Wahrnehmung von Musikern und Dirigenten in einem weiteren Konzertsaal (Muziekgebouw aan de A2 in Leidsche Rijn), in welchem durch Variation von Diffusoren, Deckenreflektoren und Podesten zwei verschiedene Bühnenkonfigurationen getestet wurden. Vom Dirigenten und dem Orchester wurden in Konfiguration 2, in welcher höhere Werte für LQ_{7-40} gemessen wurden, eine deutliche Verbesserung des Zusammenspiels erfahren.

2.3. Erkenntnisse der Auditoriumsakustik

Der Zusammenhang zwischen subjektiver Bewertung von Konzertsälen im Auditorium und physikalisch messbaren Parametern wurde seit den 1970er Jahren hergestellt. Die Studien dieses Forschungsfeldes sind sowohl aufgrund der angewandten Methoden als auch den erbrachten Ergebnissen hilfreich für die empirische Bühnenakustik.

Tabelle 2.1.: Subjektive Aspekte bei der Bewertung der Raumakustik im Auditorium

Subjektive Aspekte	**Englische Synonyme**	**Akustische Parameter**	**Quellen**
Subjektiver Schallpegel	loudness	Stärkemaß G	DIN EN ISO 3382-1:2009, Barron (2010), Gade (2007)
Wahrgenommene Halligkeit	reverberant response, reverberance	Frühe Abklingzeit EDT	DIN EN ISO 3382-1:2009, Barron (2010), Gade (2007)
Wahrgenommene Transparenz des Schalls	clarity	Klarheitsmaß C_{80}, Deutlichkeit D_{50}, Schwerpunktzeit T_S	DIN EN ISO 3382-1:2009, Barron (2010), Gade (2007)
Scheinbare Quellenbreite	apparent source width (ASW)	Früher seitlicher Energieanteil J_{LF} oder J_{LFC}	DIN EN ISO 3382-1:2009, Barron (2010), Gade (2007)
Zuhörer-Einhüllung	listener envelopment (LEV)	Später seitlicher Schallpegel L_J	DIN EN ISO 3382-1:2009, Barron (2010), Gade (2007)
Intimität	intimacy	Initial Time Delay Gap $ITDG$	Barron (2010), Gade (2007)
Spektrale Klangfärbung	spectral coloration, timbre, tone colour	Bass Ratio BR, Treble Ratio TR, spectral coloration index	Gade (2007), Rubak (2004), Barron (2010)
Zeitliche Klangfärbung	temporal coloration	temporal coloration index	Rubak (2004)
Echo			Gade (2007), Kuttruff (2009)

Lehmann und Wilkens (1980) ließen 40 Versuchspersonen 60 verschiedene Aufnahmen (3 Musikbeispiele, 6 Säle, 3 bzw. 4 Plätze pro Saal) auf 19 Bewertungsskalen durch Gegensatzpaare bewerten. Die gewonnenen Daten weisen auf drei Beurteilungsfaktoren hin, die *Empfindung der Stärke und Ausdehnung der Schallquelle*, die *Empfindung der Deutlichkeit des Gesamtklanges* und die *Beurteilung des Gesamtklanges bezüglich der Klangfarbe*. Für die drei Beurteilungsfaktoren wurden anhand von Korrelationsanalysen akustische Parameter ermittelt, welche am höchsten mit den Variablenbewertungen korrelieren und somit am ehesten die Möglichkeit bieten, die subjektiven Beurteilungen durch physikalische Maße vorherzusagen. Bezüglich der *Gütebeurteilung* konnte die Probandengruppe in zwei Gruppen unterteilt werden, in welchen jeweils der gleiche akustische Parameter Auswirkungen auf die subjektive Gütebeurteilung eines Raums hatte. Aufgrund „der Unterschiedlichkeit der Beurteilung durch verschiedene Versuchspersonen (-gruppen) sowie bei verschiedenen Musiken“ müsse „von dem Gedanken der Formulierung eines Optimums von raumakustischen Eigenschaften [...] abgegangen werden“ (Lehmann und Wilkens, 1980, S. 267).

Schroeder et al. (1974) fanden in ihren Untersuchungen, in welchen die Präferenz bezüglich unterschiedlicher Konzertsäle abgefragt wurde, neben einem *consensus preference factor* drei Faktoren, bei welchen die individuellen Unterschiede sehr hoch sind. Große interindividuelle Bewertungsunterschiede sind somit auch in der Auditoriumsakustik zu finden, in welcher, anders als bei der empirischen Bühnenakustik, die Möglichkeit der Darbietung absolut gleicher Stimuli gegeben ist.

Nachdem Lehmann und Wilkens drei subjektive Bewertungskriterien bezüglich der Raumakustik finden konnten, wurden diese in folgenden Untersuchungen erweitert. In der DIN EN ISO 3382-1:2009 sind insgesamt fünf subjektive Zuhöreraspekte genannt (siehe Tabelle 2.1). In weiterer Literatur finden sich zusätzliche Wahrnehmungsdimensionen bezüglich der Intimität und der Klangfarbe (vgl. Barron, 2010, S. 40: *intimacy*, *tone colouration*). Rubak (2004) unterteilt den Begriff der *Klangfärbung* weiter in *spectral coloration* und *temporal coloration*. Die *spektrale Klangfärbung* meint dabei die Klangfärbung durch frühe Reflexionen bis 30 ms, welche wahrnehmungsbedingt nicht vom Direktschall zu trennen sind. Die *zeitliche Klangfärbung* wird hingegen durch spätere Reflexionen bewirkt und durch Begriffe wie *whooshing*, *fluttering* und *echo disturbance* beschrieben.

2.4. Zusammenfassung

Anhand der Auflistung subjektiver Wahrnehmungsaspekte der Auditoriumsakustik in Tabelle 2.1 wird die Mehrdimensionalität der Raumakustikwahrnehmung deutlich. Obwohl für Musiker auf der Bühne aufgrund der Nähe zum eigenen Instrument und der aktiven Aufführungssituation andere Aspekte der Raumakustikwahrnehmung entscheidend sein können, ist eine eindimensionale Wahrnehmung schwer vorstellbar. Zwar wurde diese von Chiang et al. (2003) im Fall von Solisten angenommen und von Barron (1978) in einer Untersuchung bei Ensemblemusikern beobachtet, jedoch deuten die von Gade (1989a) durchgeführten In-

terviews auf eine mehrdimensionale Bewertung hin. In zwei Feldstudien zeigten sich bei Gade (1989b) desweiteren bereits zwei Dimensionen bei der Bewertung der Raumakustik durch Orchestermusiker. Bei der Faktorenanalyse anhand der Mittelwerte der Musikerurteile wurde ein Faktor als allgemeines Qualitätsurteil, welchem bis auf die Klangfarbe alle Bewertungsaspekte zuzuordenen waren, und ein zweiter als Bewertung ebendieser Klangfarbe interpretiert. Ueno und Tachibana (2003) untersuchten zwar Solisten im Labor und ließen diese auch die raumakustische Umgebung auf Ratingskalen bewerten, jedoch eigneten sich die Ergebnisse Ueno und Tachibana zufolge nicht zu einer statistischen Auswertung. Ein Grund für diese aufgetretenen Probleme kann die zu geringe Varianz der raumakustischen Umgebungen sein. Eine mögliche multidimensionale Wahrnehmung bzw. Bewertung wird erst zu beobachten sein, wenn die Räume sich innerhalb der Untersuchung bezüglich der zu bewertenden Aspekte auch möglichst stark unterscheiden. Die dargebotenen raumakustischen Umgebungen sollten daher bezüglich ihrer physikalischen Eigenschaften, welche als ausschlaggebend für die Wahrnehmung angenommen werden, eine hohe Varianz aufweisen. Bei dem von Barron (1978) durchgeführten Feldversuch in einem einzelnen Konzertsaal, wenn auch ausgestattet mit variablen Akustikelementen, ist solch eine Variation nicht in ausreichendem Maße gegeben.

In den meisten der vorgestellten Studien wurden Unterschiede zwischen Instrumentengruppen beobachtet. Bei Ueno und Tachibana (2003) zeigte sich bei der Untersuchung von Solisten, dass Streicher längere Nachhallzeiten als Holzbläser präferierten. Bei Barron (1978) unterschied sich im Ensemble bei Streichern und Bläsern die präferierte Aufstellung akustischer Bühnenelemente, welche in einer unterschiedlichen Stärke der frühen Reflexionen resultiert. Gade (1989a) wies unterschiedliche Wahrnehmungsschwellen früher Raumreflexionen von Streichern und Flötisten nach. Bei der Bewertung der Raumakustik von Konzertsälen konnte Dammerud (2009) jedoch statistisch unter Verwendung von t-Tests keine Unterschiede zwischen den verschiedenen Instrumentengruppen nachweisen. Vor allem die großen interindividuellen Unterschiede, welche bei den meisten Studien beobachtet wurden, erschweren den statistischen Nachweis des Instrumenteneinflusses. Diese interindividuelle Varianz wurde allerdings bislang in keiner der betrachteten Untersuchungen beziffert.

Bei der Untersuchung des Aspekts der Unterstützung durch den Raum kam es zu unterschiedlichen Forschungsergebnissen. Von Gade (1989a) wurde zunächst ein Zusammenhang zwischen wahrgenommener Unterstützung des eigenen Instruments und der Stärke früher Reflexionen bis 100 ms festgestellt, woraufhin der Parameter ST_{early} definiert wurde. In der Praxis stellte sich dieser Parameter jedoch als erfolgreicher Prädiktor für die Hörbarkeit anderer Ensemblemitglieder heraus (Gade, 1992). In der Studie von Ueno und Tachibana (2003) wurde desweiteren ein Zusammenhang der Unterstützung mit der Stärke später Reflexionen gefunden, wobei in diesem Fall die Musiker alleine und nicht im Ensemble spielten. Gleichzeitig gaben die Musiker an, dass ein starker später Nachhall sich bei der Aufführung im Ensemble möglicherweise störend auswirken könnte. Daraus ist abzuleiten, dass sich die Bewertung in verschiedenen Aufführungssituationen unterscheidet und diese daher ge-

trennt voneinander untersucht werden müssen. Diese Unterscheidung wird nach Auffassung des Autors in bisherigen Publikationen nicht deutlich genug formuliert. Desweiteren zeigt sich, dass der semantische Begriff der Unterstützung in verschiedenen Situationen unterschiedliche physikalische Phänomene beschreibt.

Außer in den Studien von Gade (1989a), welcher wichtige Bewertungsdimensionen durch Interviews elizierte, und Ueno und Tachibana (2003), welche die Form eines offenen Interviews zur Bewertung der Raumakustik wählten, wurden die Musiker mit vom Versuchsleiter vorgegebenen Bewertungskriterien konfrontiert. Da es sich bei der raumakustischen Wahrnehmung während des eigenen Spiels um einen sehr komplizierten psychologischen Prozess handelt, müssen auch die angewandten Methoden selbst hinterfragt werden (vgl. Ueno und Tachibana, 2003, S. 137–138).

Auf Grundlage der in diesem Abschnitt zusammen getragenen Erkenntnisse und Mängel bisheriger Studien zur raumakustischen Wahrnehmung durch Musiker wurde ein Versuchsdesign für eine eigene Untersuchung entwickelt. Im nächsten Kapitel soll die Durchführung im Detail berschrieben und die gewonnenen Ergebnisse statistischer Auswertungen vorgestellt werden.

3. Experimentelle Studie

3.1. Fragestellung und Forschungsmethode

Die grundlegenden Forschungsfragen, welchen sich diese Arbeit widmet, stellte bereits Gade (1989a) auf. Sie fragen nach dem Zusammenhang subjektiver Bewertungsmerkmale mit physikalisch messbaren Parametern und folgen somit der Tradition der empirischen Auditoriumsakustik.

1. Welche subjektiven Aspekte sind bei Musikern bezüglich der Bewertung (Wahrnehmung) raumakustischer Bedingungen vorhanden?
2. Welche Eigenschaften der Impulsantwort (objektive Parameter) erfassen jeden dieser subjektiven Aspekte?

Auch wenn Gade (2010) die Beantwortung dieser Fragen im Rückblick mit dem Erreichen „Elysions“ gleichsetzte, sind sie grundlegend für unsere Forschungsbemühungen. Die in Kapitel 2 hervorgehobenen interindividuellen Unterschiede bei der Beurteilung von Räumen durch Musiker stellen jedoch ein beachtliches Problem für die Annäherung an die Problematik dar, weshalb die Unterschiede im Detail untersucht und quantifiziert werden sollten. Auch der Einfluss des Faktors Instrument, welcher erheblichen Anteil an der Wahrnehmungssituation des Musikers hat, sollte in diese Betrachtungen miteingebunden werden.

3. Wie gestalten sich interindividuelle Unterschiede bei der Bewertung der Raumakustik durch Musiker?

Geschlossene Hypothesen, welche bestimmte Ergebnisse erwarten, wurden bewusst nicht formuliert, um den explorativen Charakter der Studie zu manifestieren. Die Zusammenhänge der Musikerbewertungen und der raumakustischen Parameter wurden durch einfache Korrelationen, die interindividuellen Bewertungsunterschiede durch die Anwendung der Intraklassenkorrelation geprüft. Um die bei Musikern vorhandenen Wahrnehmungsdimensionen zu elizieren, wurde in einem Vorversuch die Repersory-Grid-Technique (RGT) eingesetzt. Anhand der Ergebnisse wurden 20 Items (abhängige Variablen) ausgewählt, mithilfe derer eine Bewertung raumakutischer Umgebungen aus Sicht der Musiker stattfinden konnte. Die variierte unabhängige Variable stellte ein simuliertes Schallfeld eines Aufführungsraums dar, welches neben der Bewertung durch die Musiker anhand seiner Impulsantwort und den daraus zu errechnenden raumakustischen Parametern definiert werden konnte. Die raumakustischen Umgebungen wurden den Musikern während des eigenen Spiels durch dynamische

Binauralsynthese dargeboten. Dabei wurden am Instrument abgenommene Mikrofonsignale mit binauralen Raumimpulsantworten[1] gefaltet und über Kopfhörer wiedergegeben. Dabei wurden nur die Raumreflexionen simuliert, durch das aktive Musizieren der Probanden wurde der Direktschall von diesen selbst produziert. Bei der Binauraltechnik wird das Schallfeld direkt vor den Ohren des Hörers synthetisiert. Im dynamischen Fall werden die zur Faltung verwendeten binauralen Raumimpulsantworten abhängig von der Kopforientierung, welche mittels Headtracker ermittelt wird, ausgewählt. Durch die Berücksichtigung der Kopfbewegungen können Vorne-Hinten-Vertauschungen bei der Lokalisation von Schallquellen eliminiert werden (vgl. Slavik und Weinzierl, 2008, S. 671–676). Auch wenn die Schallquelle selbst nicht simuliert wird, ist es doch von großer Bedeutung für den Musiker, beispielsweise Reflexionen aus dem Auditorium vor und nicht hinter sich zu orten.

Die musikalische Aktivität der Musiker bedingt zwar Unstetigkeiten im Anregungsstimulus, ist für die naturgetreue Simulation der Aufführungssituation jedoch unabdingbar. Nur auf diese Weise kann der Wahrnehmungsprozess, wie er auf der Bühne stattfindet, nachgebildet und somit valide untersucht werden. Die binauralen Raumimpulsantworten wurden mithilfe der Akustiksimulationssoftare *EASE* und einem am Fachgebiet Audikommuniaktion entwickelten Matlab-Skriptpaket erstellt. Eine detaillierte Darstellung der Generierung folgt in Abschnitt 3.3.1.1. Insgesamt wurden im Hauptversuch 10 verschiedene raumakustische Umgebungen anhand der 20 Items von jeweils 14 Musikern auf Ratingskalen bewertet. Die unabhängige Variable Raum kann durch die Bewertung aller 10 Räume durch jede Versuchsperson als Messwiederholungs- bzw. Innersubjektfaktor aufgefasst werden.

Bis zum heutigen Zeitpunkt existieren keine Publikationen bezüglich eines mehrdimensionalen Ratings von Solo-Musikern in einer Laborumgebung.[2] Die in Feldversuchen verwendeten subjektiven Bewertungsaspekte wurden von den Versuchsleitern festgelegt (Sanders, 2003; Dammerud, 2009), die Ergebnisse sind zudem aufgrund der Fokussierung auf Ensembles nicht auf die akustische Situation von Solisten übertragbar. Aus diesen Gründen soll in dieser Arbeit eine Annäherung mit Hilfe einer experimentellen Studie im Labor stattfinden, in welcher die entscheidenden Wahrnehmungsdimensionen von Solisten in einem Vorversuch eliziert werden.

Bevor in Abschnitt 3.7 die Ergebnisse dargestellt und die angewandten statistischen Auswertungsverfahren erläutert werden, wird das Untersuchungsdesign inklusive Simulationsumgebung und dargebotener Räume vorgestellt. Die Einzelheiten des RGT-Vorversuchs werden in Abschnitt 3.5 beschrieben.

[1] Im Folgenden BRIR genannt, als Abkürzung für *binaural room impulse response*.

[2] Ueno und Tachibana (2003) führten ein solches durch, geben aber an, dass sich die Ergebnisse des Ratingversuchs aufgrund interindividueller Unterschiede nicht statistisch auswerten lassen. Die Angabe der abgefragten Items fehlt.

3.2. Simulationsumgebung und Versuchsaufbau

Die experimentelle Studie fand im reflexionsarmen Raum der Technischen Universität Berlin statt.[3] Während die Musiker auf einem Stuhl sitzend musizierten, wurde ihnen im gleichen Moment über Kopfhörer die Akustik eines Raumes dargeboten. Die Versuchsdurchführung im Freifeld ist daher notwendig, da neben dem simuliertem Nachhall keine Raumreflexionen ausgebildet werden dürfen, welche Einfluss auf die Wahrnehmung und somit die Bewertung der Probanden nehmen können.

Abbildung 3.1.: Musiker im reflexionsarmen Raum

Das jeweilige Instrument wurde mit einem Mikrofon des Typs *4099 T* von *DPA*[4] abgenommen. Dieses wurde direkt am Instrumentenkorpus angebracht, wodurch ein konstanter Abstand zwischen Instrument und Mikrofon, auch bei Bewegungen des Musikers, gewährleistet werden konnte. Die von Ueno und Tachibana (2003) gewählte Mikrofonierung mithilfe eines fest stehenden Mikrofons hätte hingegen zu unerwünschten Pegeländerungen geführt. Problematisch bei der Anbringung des Mikrofons am Instrument selbst war jedoch die Übertragung von Körperschall, welche auch durch die spezielle Aufhängung des *DPA 4099 T* nicht vollständig unterdrückt werden konnte.

Das Mikrofonsignal wurde nach Vorverstärkung mittels *Motu Audio Express*, welches auch die Phantomspeisung für das Mikrofon lieferte, im PC analog-digital gewandelt und mithilfe der Echtzeit-Faltungssoftware *fwonder* mit BRIRs gefaltet, entzerrt sowie im Pegel angepasst, wodurch die in Abschnitt 3.3.1 beschriebenen Konzertsäle simuliert wurden. Die Wahl der zur Faltung verwendeten BRIRs war dabei abhängig von der Kopfausrichtung der

[3] Der reflexionsarme Vollraum des Fachgebiets für Technische Akustik verfügt zu allen Richtungen hin über Absorber. Durch ein Gitternetz ist der 1070 m^3 und 126 m^2 große Raum begehbar. Die untere Grenzfrequenz wird mit 63 Hz angegeben (siehe im Literaturverzeichnis unter „Datenblätter und Handbücher“).

[4] Das *DPA 4099 T* ist ein Kondensatormikrofon mit Richtcharakeristik Superniere (siehe „Datenblätter und Handbücher“ im Literaturverzeichnis).

Versuchsperson, welche durch den Headtracker *Polhemus Patriot* registriert und an *fwonder* übermittelt wurde.

Das gefaltete Signal wurde nach DA-Wandlung und Verstärkung durch die Endstufe *Yamaha P3500S* über Kopfhörer wiedergegeben. Die Bauweise des verwendeten Kopfhörers *K1000* der Firma *AKG* ist extraaural, die Treiber befinden sich somit frei vor den Ohren. Dadurch hörten die Musiker das eigene Insrument ohne Einschränkungen im Frequenzbereich. Eine Rückkopplung des simulierten Nachhalls auf das Instrumenten-Mikrofon wurde aufgrund der dichten Mikrofonierung sowie der kurzen Entfernung zwischen Kopfhörer und Ohren und den daraus resultierenden großen Pegelunterschieden am Mikrofon verhindert. Eine schematische Darstellung des technischen Versuchsaufbaus findet sich in Abbildung 3.2.

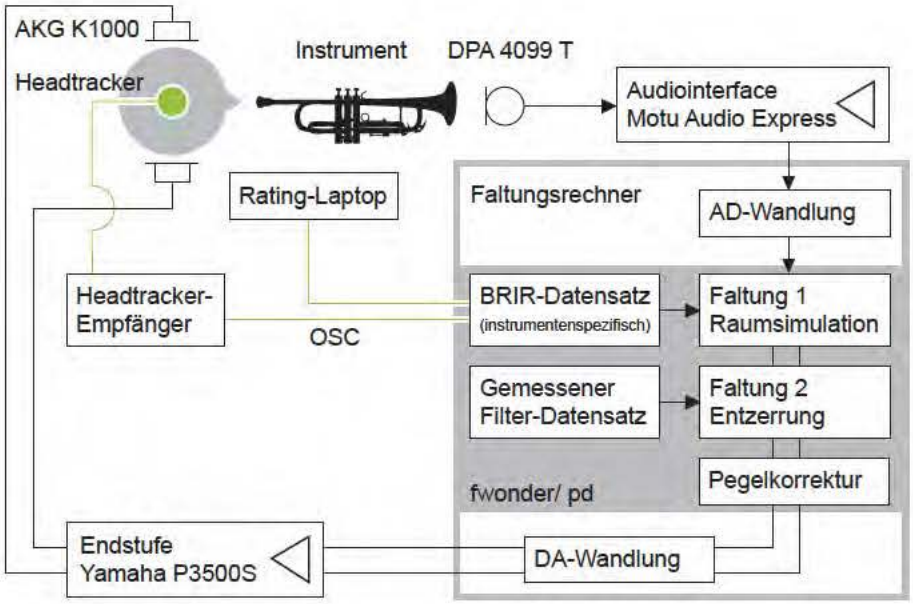


Abbildung 3.2.: Technischer Versuchsaufbau

3.2.1. Pegelkalibrierung

Ein wichtiger Faktor bei einer naturgetreuen Raumakustiksimulation ist die Darbietung der Raumreflexionen mit dem korrekten, d. h. in der Realität auftretenden Schallpegel. Durch einen Vergleich des real im reflexionsarmen Raum vorhandenen Schallfelds mit einer Simulation desgleichen durch die bereits beschriebene Simulationsumgebung wurde diese grundlegende Vorraussetzung erfüllt. Durch die BRIR-Generierung (siehe Abschnitt 3.3.1.1) war zudem die korrekte Pegelrelation zwischen den einzelnen Räumen gegeben, wodurch eine Kalibrierung ohne Kenntnis des absoluten Schallpegels möglich war. Da der Abstand des Mikrofons zum Instrument und der Eingangspegel bei der Vorverstärkung des Instrumentensignals bei jeder Versuchsperson variierten, musste die Anpassung des Pegels bei jedem Versuchsdurchgang nach Anbringung des Mikrofons und dem Einpegeln neu erfolgen. Um

ein Anregungssignal für den Pegelvergleich zu erhalten, wurden die Versuchspersonen darum gebeten, auf ihrem Instrument einen möglichst lauten, tiefen Ton oder falls möglich einen lauten, tiefen, dissonanten Akkord für ca. fünf Sekunden zu halten. Das Signal wurde sowohl über das am Instrument angebrachte Mikrofon *DPA 4099 T* als auch über das 5 m entfernte Kunstkopfmikron *KU 81i* von *Neumann* aufgenommen. Um die gleiche akustische Situation aus Perspektive des Kunstkopfes zu simulieren, wurde diesem der Kopfhörer aufgesetzt und das zuvor reale Schallfeld simuliert. Dazu wurde das aufgenommene Signal des Instrumentenmikrofons in *fwonder* mit einer BRIR der gleichen simulierten Situation im reflexionsarmen Raum gefaltet,[5] wodurch einerseits die Pegelabnahme über 5 m sowie der bei allen Berechnungen in *EASE* verwendete Quellpegel berücksichtigt wurde. Anschließend wurden die RMS-Pegel der über den Kunstkopf aufgenommenen, tiefpassgefilterten[6] Signale des realen und des simulierten Schallfelds verglichen, der Quotient der Pegel diente als Verstärkungsfaktor für das Ausgangssignal in *fwonder*. Da den in *EASE* simulierten Quellen in allen anderen Raummodellen die gleichen Pegel zugewiesen wurden, stellt dieses Verfahren ohne Kenntnis des absoluten Schalldruckpegels ein geeignetes Verfahren zur Pegelkalibrierung dar. Abbildung 3.3 zeigt eine schematische Darstellung des Vorgangs.

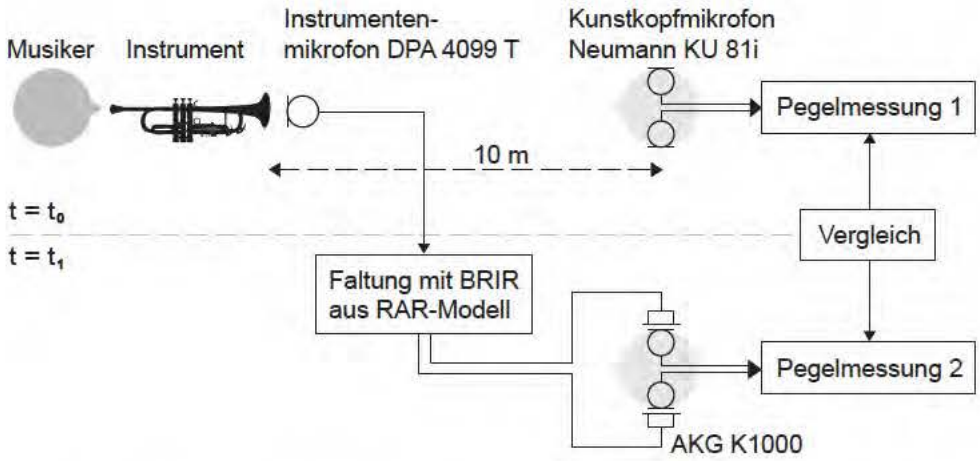


Abbildung 3.3.: Schematische Darstellung der Pegelkalibrierung

3.2.2. Kopfhörer- und Außenohrentzerrung

Beim Vergleich realer Schallfelder mit deren binauralen Simulationen stellen spektrale Unterschiede einen wesentlichen Aspekt der Unterscheidbarkeit dar (Lindau et al., 2007). Maßgeblich beeinflusst werden diese durch die Übertragungsfunktionen des verwendeten Kopf-

[5] Auch die Richtcharakteristik der Instrumente wurde bei der Erstellung der BRIRs beachtet. Genaueres zu der Vorgehensweise findet sich in Abschnitt 3.3.2.

[6] Das Tiefpassfilter hatte eine Grenzfrequenz von 500 Hz, eine Flankensteilheit von 48 dB/Oktave und einen Gütefaktor von 0,71

hörers, der vorliegenden HRTF[7] bei der Messung oder Generierung der BRIRs sowie die HRTF der jeweiligen Versuchsperson. Wie in Abschnitt 3.3.1.1 noch genauer beschrieben, wurden bei Generierung der in der Studie verwendetetn BRIRs aus Gründen der Praktikabilität keine individuellen HRTFs der Versuchspersonen, sondern ein winkelabhängiger HRTF-Datensatz eines Kunstkopf-Torso-Systems verwendet. Da nicht-individuelle binaurale Aufnahmen, die mit der Simulation in dieser Studie vergleichbar sind, bei Kompensation mit individuellen HpTFs[8] der Probanden in einer Studie als am meisten realistisch wahrgenommen wurden (Lindau und Brinkmann, 2010), wurde eine Entzerrung der Endstufe, des Kopfhörer sowie des Außenohrs des jeweiligen Musikers vorgenommen. Dazu wurde die Übertragunsfunktionen je Proband per In-Ear-Mikrofone[9] gemessen. Über den aufgesetzten Kopfhörer wiedergegebene Sinus-Sweeps wurden in zehn Wiederholungen, zwischen welchen der Proband den Kopf leicht bewegte, aufgenommen. Durch die Kopfbewegungen sollte ein etwaiges Verrutschen des Kopfhörers während des Versuchs berücksichtigt werden. Die mit Variation der Kopfhörerposition einhergehende Variation der Übertragungsfunktion, insbesondere schmalbandige Einbrüche in oberen Frequenzbereichen, flossen somit in die Messungen mit ein und konnten noch vor der Filtererstellung herausgemittelt werden (vgl. Schärer und Lindau, 2008, S. 520). Ausgehend von der mittleren Übertragungsfunktionen wurde durch Inversion[10] ein entsprechendes Filter erstellt, welches während des Versuchs mit dem Hallsignal gefaltet wurde (siehe Abbildung 3.2, Faltung 2).

3.3. Bedingungsvariationen

Um die raumakustische Wahrnehmung durch Musiker zu untersuchen, ist in erster Linie eine Variation der raumakustischen Umgebung sicher zu stellen. Beschränkt man sich ferner nicht auf bestimmte Instrumentalisten, sondern ist interessiert an einer allgemeinen Wahrnehmung von Orchstermusikern, wird zwischen den Probanden auch das Instrument variiert. Beim vorgestellten Versuchsaufbau, bei welchem der Musiker selbst das Anregungssignal durch sein Spiel bestimmt, verändert sich auch die musikalische Passage sowohl von Proband zu Proband als auch von einer Wiederholung zur nächsten bei unterschiedlichen simulierten Räumen. Auf alle drei vorliegenden Bedingungsvariationen sowie mögliche Probleme soll in diesem Abschnitt eingegangen werden.

[7] HRTF als Abkürzung für *head related transfer function* (kopfbezogene Übertragungsfunktion).

[8] HpTF als Abkürzung für *headphone transfer function* (Kopfhörerübertragungsfunktion). Sie beschreibt die Übertragungsfunktion des Kopfhörers und des Außenohrs, da sie am Trommelfell gemessen wird.

[9] Die verwendeten In-Ear-Mikrofone sowie die Messprozedur wurden im Rahmen der Magisterarbeit von Fabian Brinkmann am Fachgebiet Audiokommunikation entwickelt (Brinkmann, 2011).

[10] Die Inversion wurde durch die high-pass-regularized least-mean-square (LMS) Methode durchgeführt, welche sich sich bezüglich der Wahrnhemung als geeignet gezeigt hat (vgl. Brinkmann und Lindau, 2010).

3.3.1. Auralisierte Räume

Die zu bewertenden Räume wurden durch Faltung der Instrumentensignale mit dynamischen binauralen Impulsantworten simulierter Schallfelder dargeboten. Eine Untersuchung der wahrgenommenen Ähnlichkeit von Auralisationen durch in Räumen gemessene und durch Raytracing gewonnene BRIRs von Lokki und Järveläinen (2001) ergab, dass bei Anregungssignalen ohne starke Transienten wie Klarinette, Gitarre und der menschlichen Stimme beide Methoden im Median in den meisten Dimensionen als ziemlich ähnlich, höchstens aber als geringfügig unterschiedlich bewertet wurden. Neben den geringen perzeptiven Differenzen bieten sich durch die raumakustische Simulation verschiedenste Anpassungsmöglichkeiten der architektonischen und raumakustischen Eigenschaften der Modelle. Bei Zusammenstellung eines geeigneten Stimulisets ist diese Flexibiltät sehr hilfreich. Der erhebliche Mehraufwand durch Messungen eines vergleichbaren Sets an dynamischen BRIRs in realen Konzertsälen steht zudem in keiner Relation zur gewonnenen wahrnehmbaren Steigerung der Authentizität der Auralisation. Aus diesen Gründen wurden die zur Auralisation verwendeten BRIRs mithilfe der Akustiksimulationssoftware *EASE* sowie eines am Fachgebiet Audiokommunikation entwickelten Matlab-Skriptpakets erstellt.

Um eine möglichst umfassende Abfrage von Wahrnehmungsaspekten zu ermöglichen, ist eine hohe Varianz bezüglich der auf der Bühne gemessenen raumakustischen Parameter der dargebotenen Raumsimulationen vonnöten. Aus Gründen der Akzeptanz der Musiker und der externen Validität sollten sich die Modelle gleichzeitig an realen Aufführungsstätten orientieren. Ausgehend von bestehenden Datensätzen zu raumakustischen Parametern von Konzertsälen wurden sieben architektonische Prototypen ausgewählt und anhand von real existierenden Vorbildern als 3D-Modelle in *EASE* nachgebildet, darunter zwei Kammersäle sowie zwei Konzertsäle unterschiedlicher Bauweise, eine Kirche, eine Oper und ein Teater.

Durch Variation der Absorptionsgrade der Oberflächenmaterialien im Kirchenmodell, einem der Kammer- und einem der Konzertsäle wurde die Anzahl der Räume auf 10 erhöht und eine größere Streuung bezüglich der raumakustischen Parameter erreicht. In den variierten Modellen wurden vor allem kürzere Nachhallzeiten sowie ein unterschiedlicher Frequenzverlauf der Oktavbandwerte von $T_{30\,1\mathrm{m}}$ angestrebt (siehe Abschnitt 3.3.1.2). Die Raumsimulationen wurden den Versuchspersonen randomisiert dargeboten, damit sich mögliche Lern- bzw. Transfereffekte (siehe Bortz und Döring, 2006, S. 549) zufällig verteilen.

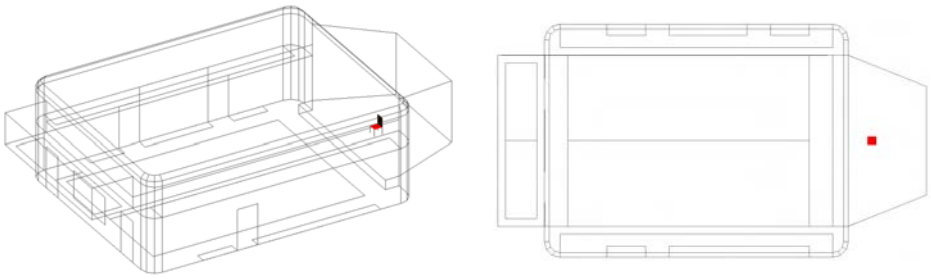
3.3.1.1. BRIR-Generierung

Zur Erstellung der 3D-Modelle und der raumakustischen Simulation wurde das Softwarepaket *EASE* verwendet. Die aus der Simulation gewonnenen Reflektogrammdaten wurden in Matlab zur Erstellung dynamischer BRIR-Datensätze genutzt, welche eine Faltung der Instrumentensignale abhängig von der Kopfposition der Musiker erlaubten. Der späte diffuse Anteil der BRIR enthält keine perzeptiv relevanten Informationen zur Lokalisation (vgl. Lindau et al., 2010) und wurde deshalb nicht bezüglich der Kopfbewegung variiert.

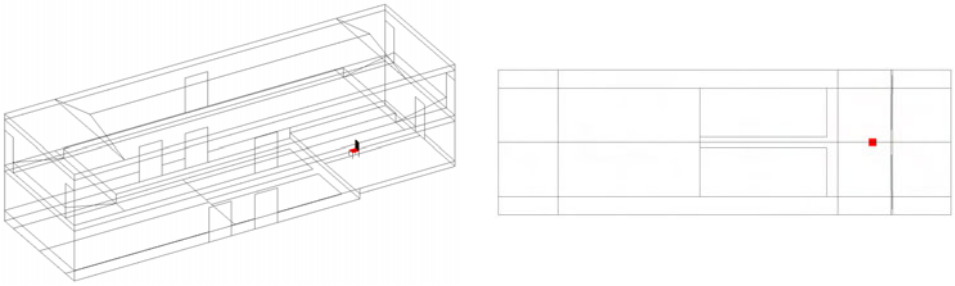
Die Modellierung der Aufführungsstätten und die Anpassung der Materialeigenschaften in *EASE* wurden von Zora Schärer Kalkandjiev im Rahmen ihrer Promotion durchgeführt und für diese Untersuchung zur Verfügung gestellt. Abbildung 3.4 und Tabelle 3.1 zeigen die sieben architektonisch bzw. zehn raumakustisch unterschiedlichen Modelle. Die Empfängerpositionen sind in den Drahtgittermodellen markiert. Sie befinden sich jeweils in der Bühnenmitte, 1,5 m entfernt von der Bühnenkante. Die Höhe wurde als mittlere Höhe der Ohren in Sitzposition je 1,2 m über dem Bühnenboden gewählt. Die Spielposition wurde als sitzend festgelegt, da sich etwaige Größenunterschiede der Probanden dadurch nicht so gravierend auswirkten wie sie es bei einer stehenden Position getan hätten. Die raumakustische Simulation wurde in *EASE 4.3* durch das Modul *Aura-Response* durchgeführt. Die frühen bis zu einer bestimmten Ordnung (*Cut Off Order*) eintreffenden geometrischen Reflexionen werden durch ein hybrides Spiegelschallquellenmodell simuliert. Durch Strahlenverfolgungen, welche eine deterministische, ungleichmäßg gerichtete Verteilung aufweisen, wird der Raum dabei nach Spiegelschallquellen abgesucht. Zur Simulation früher gestreuter sowie später Reflexionen höherer Ordnung wird ein energetischer Strahlenverfolgungsalgorithmus mit einer gleichmäßigen Dichteverteilung der Strahlen angewendet.[11] Die Simulationen wurden jeweils mit besetztem Auditorium durchgeführt. Da für die Untersuchung nur die Solistensituation von Interesse war, wurde die Bühne als leer angenommen. Der Solist selbst wurde als Absorptionsfläche nicht modelliert, da dessen Auswirkung auf die Raumakustik zu vernachlässigen ist. Die in *EASE Aura-Respone* gewählten Optionen sind in den Tabellen 3.1 und 3.2 angegeben. Der Parameter *Length* legt die Länge des Reflektogramms fest und wurde bei den Simulationen so gewählt, dass er höher als T_{30} war und damit bei der Auralisation keine plötzlichen Abbrüche des Nachhalls wahrnehmbar waren. *Particles* gibt die Anzahl der Strahlen sowohl für den ersten als auch den zweiten Teil der Simulation an. Über den Parameter *Max. Diameter after 1 s* wird der Durchmesser des Suchballons nach 1 s festgelegt, wodurch die Präzision des frühen Teils in Relation zum späten variiert werden kann. Es wird empfohlen, den Durchmesser der kleinsten Fläche zu wählen, Werte < 1 m seien jedoch nicht notwendig. Durch die bereits angesprochene *Cut Off Order* wird die Reflexionsordnung, bei welcher der Übergang zwischen Spiegelschallquellenmethode und Raytracing erfolgt, angegeben. Der *Density Factor* definiert die Anzahl der Impulse pro 1 ms, welche die Detailstruktur des späten Teils des Reflektogramms erzeugen. Die zeitliche Auflösung des späten Teils wird durch die *Tail Resolution* definiert. Dies hat Einfluss auf die Bin-Breite, welche für die Energieverteilung des späten Teils genutzt wird. Die Option *Diffuse Rain* kann in sehr halligen Räumen angewendet werden, um die Anzahl der *Particles* und somit gleichzeitig die Berechnungszeit zu reduzieren. Dies war im Fall dieser Studie nicht notwendig.

Die berechneten Response-Files (*.rsp) wurden mithilfe von *EASE Probe* als Post-Files gespeichert, in welchen alle an der Empfängerposition eintreffenden Reflexionen mit Verzöge-

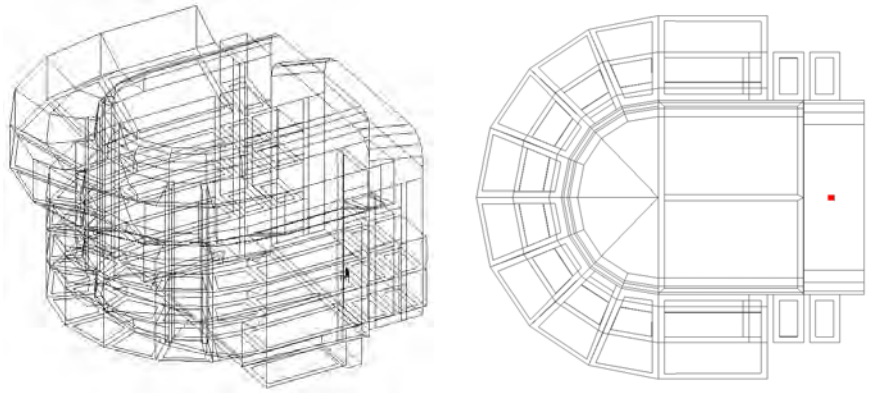
[11] Siehe unter „Datenblätter und Handbücher“ im Literaturverzechnis, S. 249.



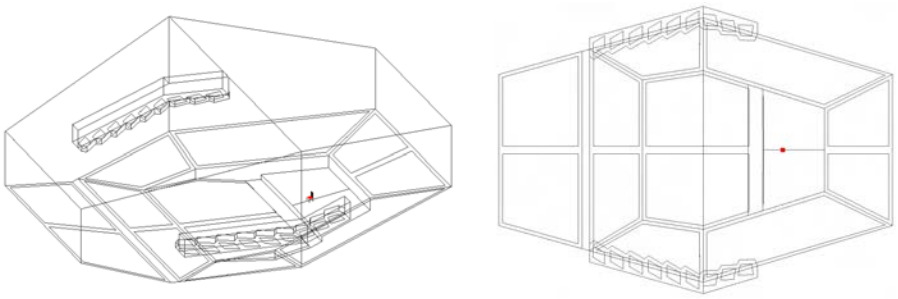
(a) R1: Kammersaal1, $V = 2380{,}63\,\mathrm{m}^3$



(b) R2: Kammersaal2, R8: Kammersaal2Short, $V = 3232{,}76\,\mathrm{m}^3$

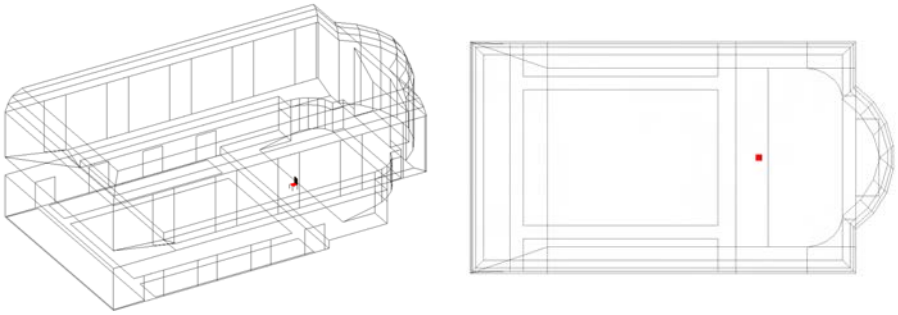


(c) R3: Oper, $V = 14\,862{,}23\,\mathrm{m}^3$

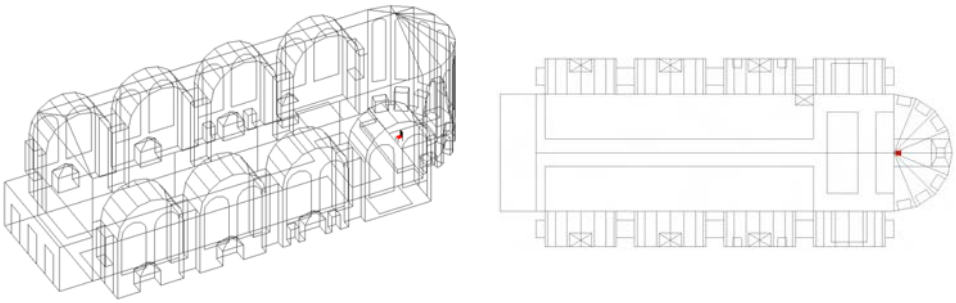


(d) R4: Konzertsaal1, R9: Konzertsaal1Short, $V = 21\,661{,}07\,\mathrm{m}^3$

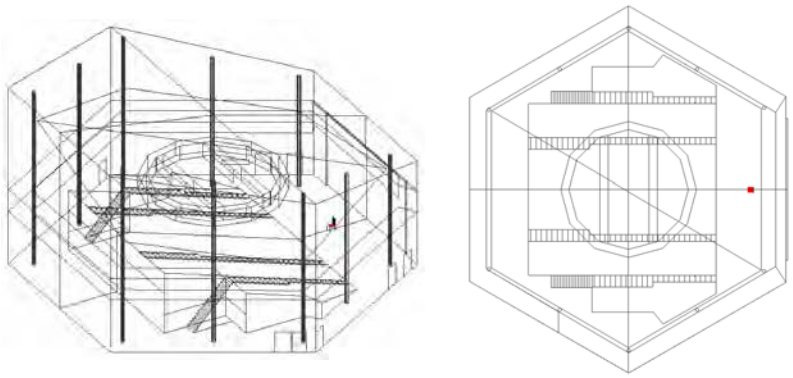
Abbildung 3.4.: 3D-Modelle und Grundrisse der architektonisch unterschiedlichen Räume mit Markierung der Musikerposition



(e) R5: Konzertsaal2, $V = 10\,261{,}58\,\mathrm{m}^3$



(f) R6: Kirche, R7: KircheShort, $V = 12\,529{,}96\,\mathrm{m}^3$



(g) R10: Theater, $V = 11\,176{,}06\,\mathrm{m}^3$

Abbildung 3.4.: (Fortsetzung) 3D-Modelle und Grundrisse der architektonisch unterschiedlichen Räume mit Markierung der Musikerposition

Tabelle 3.1.: Verwendete Raumakustikmodelle

	Bezeichnung	Vorbild	Kommentar	V m^3	*EASE*-Optionen Particles	Length ms
R1	Kammersaal1	Konzerthaus kl. Saal, Berlin		2380	2 000 000	2200
R2	Kammersaal2	Brahmssaal, Wien		3233	2 000 000	3100
R3	Oper	Opera Garnier, Paris		14 862	5 500 000	2200
R4	Konzertsaal1	De Doelen Concertgebouw, Rotterdam		21 661	6 500 000	3000
R5	Konzertsaal2	Stadtcasino, Basel		10 262	3 500 000	3100
R6	Kirche	Dom St. Stephan, Passau		12 530	3 500 000	3000
R7	KircheShort	siehe R6	R6 mit kürzerer T_{20} & variiertem Frequenzverlauf	12 530	3 500 000	3000
R8	Kammersaal2 Short	siehe R2	R2 mit kürzerer T_{20} & variiertem Frequenzverlauf	3233	2 000 000	3000
R9	Konzertsaal1 Short	siehe R4	R4 mit kürzerer T_{20} & variiertem Frequenzverlauf	21 661	6 500 000	3000
R10	Theater	Théâtre Jean Vilar, Vitry-sur-Seine		11 176	4 000 000	1000

Tabelle 3.2.: Raytracing-Optionen in *EASE Aura* für alle Modelle

Max. Diameter after 1 s in m:	1
Cut Off Order:	10
Density Factor:	10
Tail Resolution in ms:	5
Diffuse Rain:	off

rungszeit, Eintreffrichtung und Pegeln in Terzbändern von 100 Hz bis 10 kHz aufgelistet sind. Diese Informationen stellten die Grundlage für die Erstellung der dynamischen BRIR-Sets mithilfe eines Matlabskriptpakets dar, welches am Fachgebiet Audiokommunikation entwickelt wurde.[12] Nach Auslesen des jeweiligen Post-Files und der Daten-Umstrukturierung wird in den Matlab-Skripten des BRIR-Exports für jede Reflexion anhand der Terzbandpegel eine normalisierte FIR[13] generiert und als *.wav-Datei gespeichert. Mithilfe der dazugehörigen Informationen über Verzögerung, Dämpfung und Richtung wird aus den einzelnen Reflexionen das dynamische BRIR-Set erstellt. Der erste Strahl, welcher den Direktschall darstellt, wurde im Fall dieser Studie nicht in den BRIRs gespeichert, da eine Direktschallsimulation aufgrund des Versuchsdesigns nicht erwünscht war. Der die Räume anregende

[12] Die Skripte wurden von Max Röhrbein, Frank Schultz und Alexander Lindau erstellt. Die Vorgehensweise wird von Röhrbein (2011) beschrieben.

[13] FIR: *finite impulse response.*

Direktschall wurde von den Musikern während des Versuchs selbst erzeugt. Eine zusätzliche Simulation des Direktschalls hätte eine plausible Auralisation verhindert. Beim Export des BRIR-Sets wurde ein winkelabhängiger HRTF-Datensatz, gemessen mit dem Kunstkopf-Torso-System von Christoph Moldrzyk, verwendet (siehe Moldrzyk et al., 2004). Für die bewegungsabhängige Simulation wurde ein Kopfazimuth von ±70° mit einer Auflösung von 1° und einer Elevation von −30° bis +20° mit einer Auflösung von 2° festgelegt. Neben den dynamischen frühen Teilen der BRIRs wurde eine BRIR in kompletter Länge in 0°-Richtung extrahiert, welche den diffusen Nachhall simulierte. Die beiden Bestandteile der BRIR aus frühen Reflexionen und diffusem Nachhhall wurden in *fwonder* durch einen Crossfade zusammengefügt. Die Dauer zwischen Direktschall und der Übergangszeit, nach welcher keine perzeptiv relevanten Lokalisationsinformationen mehr in den Reflexionen vorhanden sind, nennt man *Mixing-Time* (vgl. Lindau et al., 2010). Der diffuse Nachhall bleibt aus Gründen der begrenzten Rechenkapazität sowie des geringeren Zeitaufwands der BRIR-Exporte konstant. Ein Verfahren, welches die perzeptiv nicht wahrnehmbare Übergangszeit von frühen Reflexionen zu diffusem Nachhall, die sogenannte *Mixing Time*, anhand einer empirischen Untersuchung berechnet, wurde von Lindau et al. (2010) entwickelt. Folgende Gleichung beschreibt die konservative Berechnungsmethode anhand des Volumens des zu simulierenden Raums, welche eine exakte Simulation gewährleisten soll. In dieser Studie wurde jeweils die nächsthöhere Zweierpotenz in ms gewählt, da *fwonder* nur an solchen Stellen einen Crossfade ermöglicht.

$$t_{mp95\%} = 0,0117 \cdot V + 50,1 \qquad [\text{ms}] \tag{3.1}$$

3.3.1.2. Raumakustische Parameter

Um Zusammenhänge zwischen den Bewertungen der Musiker und den physikalischen Eigenschaften der simulierten Räume herstellen zu können, wurden die Werte einer Auswahl raumakustischer Parameter berechnet. Diese stammen sowohl aus der Auditoriums- als auch der Bühnenakustik. Die Nachhallzeit T_{30}, die frühe Abklingzeit EDT, das Stärkemaß G, die Schwerpunktzeit T_S, das Deutlichkeitsmaß C_{50}, das Klarheitsmaß C_{80}, die frühe Unterstützung ST_{early} und die späte Unterstützung ST_{late} sind in der DIN EN ISO 3382-1:2009 festgehalten. Neben früher und später Unterstützung, welche in die Norm aufgenommen wurden, definierte Gade den Parameter ST_{total} (Gade, 1992) sowie $EDTF$, welcher ein Verhältnis verschiedener Oktavbandwerte der frühen Abklingzeit darstellt und eine grobe Beschreibung der tonalen Eigenschaften eines Raums bieten soll (Gade, 1989b). In der Norm DIN EN ISO 3382-1:2009 wird D_{50}[14] als Deutlichkeit definiert und kann durch $C_{50} = 10 \cdot \lg\left(\frac{D_{50}}{1-D_{50}}\right)$ dB in das Deutlichkeitsmaß, wie es Ahnert und Tennhardt (2008, S. 192) nennen, umgerechnet werden. Die Messung und Verwendung beider Größen ist nicht nötig, weshalb D_{50} nicht separat berechnet und in die statistische Auswertung miteinbezogen wurde.

Wie bereits in Abschnitt 2.2.6 geschildert, wurden das frühe Stärkemaß G_e und das späte

[14] $D_{50} = \int_0^{50\,\text{ms}} p^2(t)\,\mathrm{d}t \Big/ \int_0^{\infty} p^2(t)\,\mathrm{d}t$ (DIN EN ISO 3382-1:2009, S. 19).

Tabelle 3.3.: Definitionen der in dieser Arbeit verwendeten raumakustischen Parameter

Bezeichnung	Berechnungsvorschrift
Nachhallzeit	T_{30}: Zeitliche Auswertung der integrierten Impulsantwort im Bereich von $-5\,\mathrm{dB}$ bis $-35\,\mathrm{dB}$, Extrapolation bis $-60\,\mathrm{dB}$ (siehe DIN EN ISO 3382-1:2009 und Matlabskript in B)
Frühe Abklingzeit	EDT: Zeitliche Auswertung der integrierten Impulsantwort im Bereich von $0\,\mathrm{dB}$ bis $-10\,\mathrm{dB}$, Extrapolation bis $-60\,\mathrm{dB}$ (siehe DIN EN ISO 3382-1:2009 und Matlabskript in B)
Stärkemaß	$G = 10 \cdot \lg \left[\frac{\int_0^\infty p^2(t)\,\mathrm{d}t}{\int_0^\infty p^2_{10\,\mathrm{m}ff}(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Deutlichkeitsmaß	$C_{50} = 10 \cdot \lg \left[\frac{\int_0^{50\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{50\,\mathrm{ms}}^\infty p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Klarheitsmaß	$C_{80} = 10 \cdot \lg \left[\frac{\int_0^{80\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{80\,\mathrm{ms}}^\infty p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Schwerpunktzeit	$T_S = \frac{\int_0^\infty t\ p^2(t)\,\mathrm{d}t}{\int_0^\infty p^2(t)\,\mathrm{d}t} \mathrm{s}$
Frühe Unterstützung	$ST_{early} = 10 \cdot \lg \left[\frac{\int_{20}^{100} p^2(t)\,\mathrm{d}t}{\int_0^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Gesamtunterstützung	$ST_{total} = 10 \cdot \lg \left[\frac{\int_{20}^{1000} p^2(t)\,\mathrm{d}t}{\int_0^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Späte Unterstützung	$ST_{late} = 10 \cdot \lg \left[\frac{\int_{100}^{1000} p^2(t)\,\mathrm{d}t}{\int_0^{10} p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
	$LQ_{7-40} = 10 \cdot \lg \left[\frac{\int_{7\,\mathrm{ms}}^{40\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{40\,\mathrm{ms}}^\infty p^2(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Frühes Stärkemaß	$G_e = 10 \cdot \lg \left[\frac{\int_{0\,\mathrm{ms}}^{80\,\mathrm{ms}} p^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^\infty p^2_{10}(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Spätes Stärkemaß	$G_l = 10 \cdot \lg \left[\frac{\int_{80\,\mathrm{ms}}^\infty p^2(t)\,\mathrm{d}t}{\int_{0\,\mathrm{ms}}^\infty p^2_{10}(t)\,\mathrm{d}t} \right] \mathrm{dB}$
Gewichtetes Stärkemaß	$G_w = 10 \cdot \lg \left[10^{G_e(125\,\mathrm{Hz})/10} + 3 \cdot 10^{G_l(125\,\mathrm{Hz})/10} + 0{,}5 \cdot (10^{G_e(250\,\mathrm{Hz})/10} + 3 \cdot 10^{G_l(250\,\mathrm{Hz})/10})\right]$
Bass Ratio	$BR_T = \frac{T_{30}(125\,\mathrm{Hz})+T_{30}(250\,\mathrm{Hz})}{T_{30}(500\,\mathrm{Hz})+T_{30}(1\,\mathrm{kHz})}$
	$BR_{EDT} = \frac{EDT(125\,\mathrm{Hz})+EDT(250\,\mathrm{Hz})}{EDT(500\,\mathrm{Hz})+EDT(1\,\mathrm{kHz})}$
	$BR_G = 0{,}5 \cdot [G(125\,\mathrm{Hz}) + G(250\,\mathrm{Hz}) - G(500\,\mathrm{Hz}) - G(1\,\mathrm{kHz})]$
Treble Ratio	$TR_T = \frac{T_{30}(2\,\mathrm{kHz})+T_{30}(4\,\mathrm{kHz})}{T_{30}(500\,\mathrm{Hz})+T_{30}(1\,\mathrm{kHz})}$
	$TR_{EDT} = \frac{EDT(2\,\mathrm{kHz})+EDT(4\,\mathrm{kHz})}{EDT(500\,\mathrm{Hz})+EDT(1\,\mathrm{kHz})}$
	$TR_G = 0{,}5 \cdot [G(2\,\mathrm{kHz}) + G(4\,\mathrm{kHz}) - G(500\,\mathrm{Hz}) - G(1\,\mathrm{kHz})]$
EDT-Verhältnis	$EDTF = \frac{EDT(250\,\mathrm{Hz})+EDT(250\,\mathrm{Hz})}{EDT(1\,\mathrm{kHz})+EDT(2\,\mathrm{kHz})}$

Stärkemaß G_l, welches die akustische Antwort aus dem Auditorium beschreiben soll, von Dammerud (2009) definiert. Van den Braak und van Luxemburg (2008) schlugen desweiteren den Parameter LQ_{7-40} vor, bei welchem die sehr frühe Energie der Impulsanrwort zur späten Energie ins Verhältnis gesetzt wird und welcher von ihnen zur Feinabstimmung der Bühnenakustik vewendet wurde. Vor allem die Unterstützung quer über die Bühne („transversely support over stage“, van den Braak et al., 2009, S. 9) und somit die Hörbarkeit der einzelnen Orchestergruppen für den Dirigenten soll durch LQ_{7-40} ermittelt werden können. Um zu testen, ob sich dieser Parameter auch als Prädiktor für die Solisten-Situation eignet, wurde er in dieser Arbeit berücksichtigt.

Neben *EDTF* wurden weitere raumakustische Parameter berechnet, welche Verhältnisse verschiedener Oktavwerte von Parametern darstellen und somit unter Umständen mit klangfarblichen Bewertungsaspekten der Musiker in Verbindung stehen. Das Bassverhältnis bzw. die Bass Ratio *BR* definierte Beranek (2004, S. 512) als erster als Summe der Nachhallzeiten der Oktavbänder von 125 Hz und 250 Hz geteilt durch die Summe von 500 Hz und 1 kHz. In ähnlicher Form lässt sich eine Treble ratio bilden, indem die Nachhallzeitwerte hoher Frequenzbänder (2 kHz, 4 kHz) durch die der mittleren (500 Hz, 1 kHz) geteilt werden (Gade, 2007, S. 310). Desweiteren wurde sowohl die Verwendung der frühen Abklingzeit *EDT* zur Bildung der Treble Ratio *TR* vorgeschlagen (Chiang, 1994, zit. in Siebein und Kinzey, Jr., 1999, S. 262) als auch die Verwendung des Stärkemaßes G zur Berechnung beider Verhältnismaße (Gade, 2007, S. 310). Für eine Unterscheidung der Maße werden in dieser Arbeit die Parameter, aus welchen die Treble und die Bass Ratio hervorgehen im Index genannt. Als weitere Prädiktoren für die wahrgenommene Bass-Lautstärke nennen Bradley (2011) den 125 Hz-Oktavbandwert des Stärkemaßes G_{125} und G_w, eine gewichtete Kombination der 125 und 250 Hz-Oktavbänder aus frühem und spätem Stärkemaß. Alle verwendeten raumakustischen Parameter und ihre Berechnungsvorschriften sind in Tabelle 3.3 aufgetragen.

Zur Berechnung wurden monaurale Impulsantworten benötigt, weshalb erneut für alle zehn Modelle Akustiksimulationen mit *EASE Aura*, nun allerdings mit einer Kugelquelle 1 m entfernt von der Musikerposition, durchgeführt wurden. Die geringe Entfernung von Quelle und Empfänger wurde gewählt, um die Situation des Musikers nachzubilden, wie es bereits Gade (1989a) im Falle der Supportmaße praktizierte. In einem Matlabskript wurden die raumakustischen Parameter anhand der mit *EASE Probe* gespeicherten monauralen Impulsantworten berechnet. Zur Korrektur von durch Filter bedingten Verzögerungen wurden die frühen und späten Zeitfenster bei der Berechnung der Energierelationsmaße getrennt gefiltert. Diese Methode wird in der DIN EN ISO 3382-1:2009 unter A.3.4 als die am besten geeignete bezeichnet. Für die Berechnung des Stärkemaßes wurde zur Herstellung des Referenzpegels in *EASE* eine Freifeld-Simulation mit einem Quell-Empfänger-Abstand von 10 m durchgeführt, dessen Impulsantwort wiederum in *EASE Probe* als Audiodatei exportiert wurde. Da im Zuge dieses Exports eine nicht gewünschte Normalisierung der Impulsantwort statt fand, wurde der Pegelunterschied von 20 dB zwischen den Raumimpulsantworten

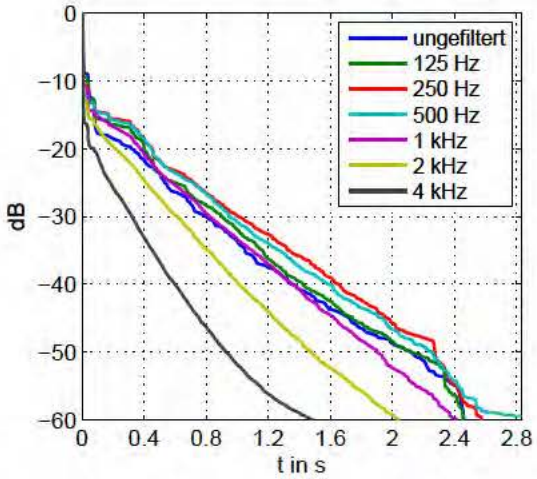


Abbildung 3.5.: Schroeder-Plot – R4 Konzertsaal1

und der Freifeldreferenz im Matlabskript wiederhergestellt. Das Matlabkript ist Anhang B zu entnehmen.

Um die Varianz der raumakustischen Parameter zwischen den Raumsimulationen zu visualisieren, sind in Abbildung 3.6 die Werte einiger Parameter in Scatterplots aufgetragen. Der Zusatz *1m* im Index gibt an, dass bei Parametern aus der Auditoriumsakustik wie auch bei den Supportmaßen ein Abstand von 1 m – und nicht wie in der Norm empfohlen von mindestens 1,5 m – zwischen Quelle und Empfänger vorlag. Es handelt sich jeweils um die Mittelwerte über Oktavbänder nach DIN EN ISO 3382-1:2009.

Während die erreichten Werte von $T_{30\,1\mathrm{m}}$ eine relativ gleichmäßige Streuung zwischen 0,6 und 2,6 s aufweisen, zeigen sich bei den Räumen R3–R5 und R9–R10 *EDT*-Werte kaum über 0 s. Aufgrund der geringen Entfernung zwischen Quelle und Empfänger ist der Direktschallanteil der Impulsantworten sehr hoch. Wenn gleichzeitig nicht auch die Energie der frühen Reflexionen besonders stark ist, bedingt dies einen sehr schnellen Abfall der Schroeder-Kurven[15]. Ein extremes Beispiel bietet Konzertsaal1 (R4), bei welchem gleichzeitig eine relativ hohe Nachhallzeit $T_{30\,1\mathrm{m}}$ von 2,64 s und eine sehr geringe *EDT* von 0,02 s berechnet wurde. Abbildung 3.5 zeigt den dazugehörigen Schroeder-Plot, in welchem der schnelle Abfall von etwa 15 dB innerhalb der ersten 10 ms und der anschließende stetige flache Abfall ersichtlich wird. Dieses Phänomen entsteht durch die Bühnenposition in der Saalmitte und der damit verbundenen großen Entfernung zu Saalwänden und -decke (siehe Abbildung 3.4d). In Konzertsaal1 Short (R9), welcher über die gleiche Architektur bei höher absorbierenden Wandmaterialien verfügt, errechnet sich neben der geringen *EDT* eine kürzere Nachhallzeit. ST_{early} als Prädiktor für die frühe Unterstützung eines Musikers unterscheidet hingegen genauer zwischen den Räumen mit schwachen frühen Reflexionen. Gleichzeitig errechnen sich bei R2 sowie R6-R8 bezüglich dieses Parameters jedoch beina-

[15] Auch Energiezerfallskurven (engl.: energy decay curve, EDC) genannt.

he gleiche Werte. Auffällig ist weiterhin, dass sich die Räume bezüglich des Stärkemaßes $G_{1\mathrm{m}}$ und $G_{e\,1\mathrm{m}}$ kaum unterscheiden. Dies lässt sich wiederum mit dem geringen Abstand von Quell- zu Empfängerposition und der dadurch bedingten Dominanz des Direktschalls erklären. Eine ausführliche Auflistung der berechneten Parameter über Oktavbänder findet sich im Anhang in den Tabellen C.1 bis C.13. Ebenfalls im Anhang zeigt Abbildung C.2 die Schroeder-Plots aller monauraler Impulsantworten, welche zur Berechnung der raumakustischen Parameter genutzt wurden.

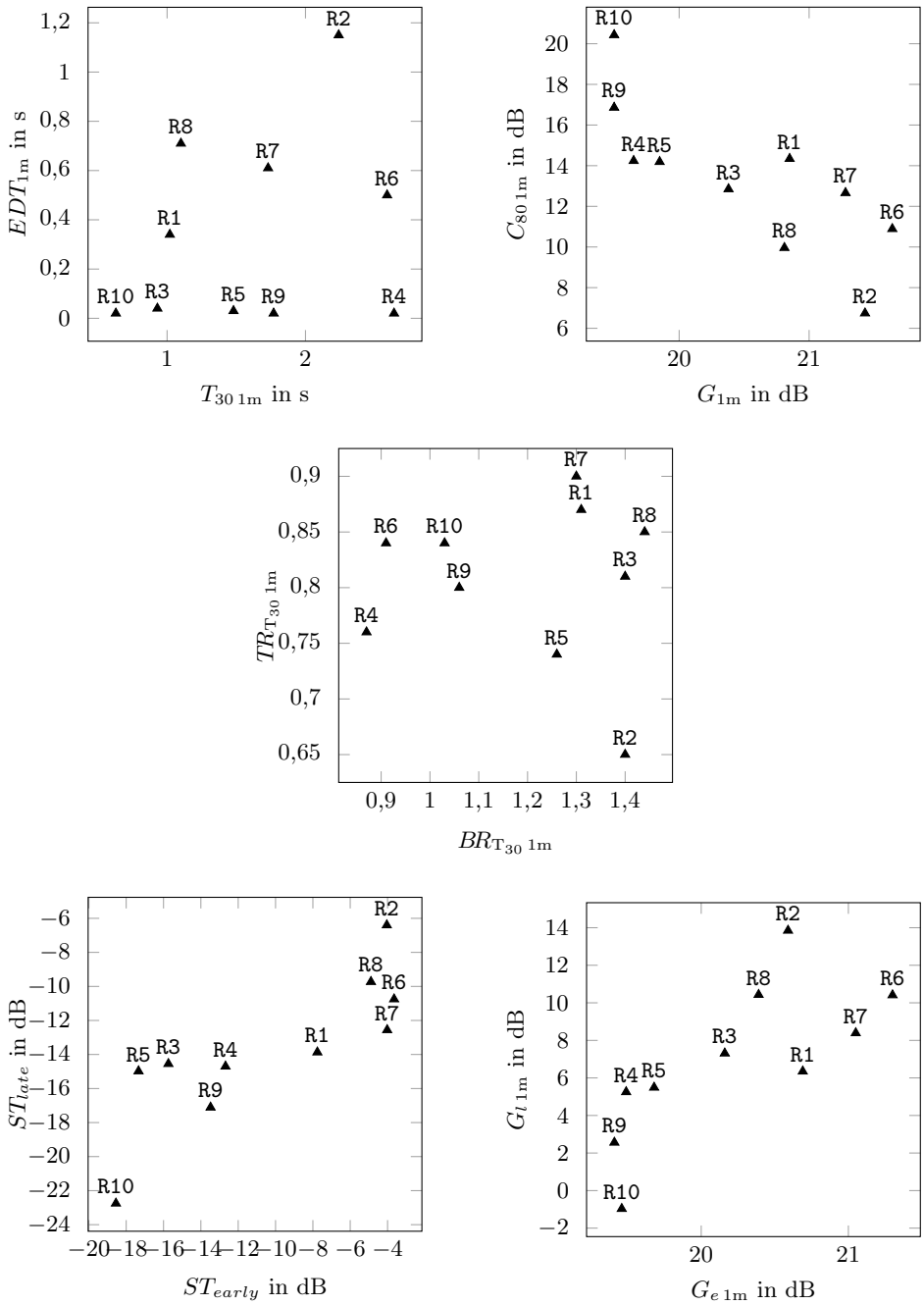


Abbildung 3.6.: Scatterplots raumakustischer Parameter der auralisierten Räume

3.3.2. Instrumente

Neben den klanglichen Eigenschaften unterscheiden sich die im Versuch gespielten Instrumente vor allem in ihrer Richtcharakteristik, ihren Maßen und dem Abstand zu den Ohren des Musikers. Um diese Unterschiede auch in der akustischen Simulation zu berücksichtigen, wurden für jedes Instrument separate Akustiksimulationen mit jeweils angepasster Quellposition sowie simulierter Instrumentenrichtcharakteristik durchgeführt. Die gewählten Positionen relativ zur Musikerpositon sind Tabelle 3.4 zu entnehmen und stellen Abschätzungen dar. Die Richtcharakteristiken wurden durch die in der Magisterarbeit von Schneider (2011) gemessenen *.spk-Dateien in *EASE* simuliert. Die Mikrofonierung der Instrumente erfolgte entsprechend des in der Arbeit angegebenen Polarwinkels (Schneider, 2011, S. 105). Diese Anpassungen wurden wie die Kanalentzerrung und die Pegelkalibrierung vorgenommen, um ein als möglichst real wahrgenommenes Schallfeld zu simulieren.

Die Unterschiedlichkeit der Instrumente erschwert jedoch gleichzeitig die Auswertung der Bewertungen raumakustischer Umgebungen durch Musiker. Dass unterschiedliche Instrumentengruppen unterschiedliche Tendenzen bezüglich der raumakustischen Bewertung aufweisen, zeigen die in Kapitel 2 bereits vorgestellten Studien von Chiang et al. (2003) sowie Ueno und Tachibana (2003).

Die Instrumente verfügen über unterschiedliche Tonumfänge und Richtcharakteristika, bewirken Signale unterschiedlicher Schallpegel, Dynamik, Hüllkurven und Klangfarben. Dementsprechend erfährt der Raum eine instrumentenspezifische Anregung, welche die Raumreflexionen bestimmt. Die unterschiedlichen Richtcharakteristika und Abstände zwischen Instrument und Ohren des Musikers haben zudem großen Einfluss auf die Ausprägung von Maskierungseffekten. Dies kann beispielsweise in unterschiedlichen Wahrnehmungsschwellen früher Reflexionen von verschiedenen Instrumentalisten resultieren, wie sie Gade (1989a) beobachtete. Es bestehen nicht nur große Unterschiede bezüglich des erzeugten Schalldruckpegels sondern eine große Spanne der Variabilität des Schalldruckpegels gemessen am linken und rechten Ohr des Musikers. Während bei einem Cello der Schalldruckpegel am linken Ohr im Frequenzbereich über 4 kHz etwa 2,5 dB höher ist als am rechten Ohr, sind es bei einer Violine mehr als 15 dB (vgl. Meyer, 2009, S. 18). Für eine ausführliche Beschreibung von Instrumentenqualitäten sei auf Meyer (2009) verwiesen, da eine detaillierte Darstellung dieses Themenbereichs an dieser Stelle nicht stattfinden kann.

3.3.3. Musikalische Passagen

Die Musiker wurden darum gebeten, zwei in Tempo, Dynamik, Artikulation sowie Intonation möglichst abwechslungsreiche musikalische Passagen für den Versuch vorzubereiten. Die Dauer der beiden Passagen wurde vor dem Versuch auf insgesamt 90 Sekunden beschränkt, so dass der gesamte Versuch möglichst nicht länger als eine Stunde dauerte und die Musiker bei ihrer Bewertung einfacher Vergleiche zwischen den nacheinander dargebotenen Räumen anstellen konnten. Die Unterschiede der gespielten Passagen sollten bei der Bewertung der

Tabelle 3.4.: Instrumentenpositionen im Modell relativ zum Musiker in m. Negative z-Werte beschreiben eine Verschiebung nach unten, positive x-Werte eine Verschiebung nach links, negative y-Werte eine Veschiebung in Blickrichtung der Musiker.

Instrument	x	y	z
Cello	0	−0,4	−0,6
Violine	0,2	−0,2	0
Oboe	0	−0,2	−0,2
Klarinette	0	−0,2	−0,2
Fagott	0	−0,2	−0,2
Trompete	0	−0,5	0
Posaune	0,2	−0,3	0
Flöte	−0,3	−0,1	0

raumakustischen Umgebung keine Probleme darstellen, da angenommen werden kann, dass Musiker einen Raum unabhängig von den Einzelheiten der jeweiligen Passage bewerten können. Tabelle 3.5 gibt einen Überblick über die Probanden und zeigt, welchen Werken die Passagen entnommen wurden.

3.4. Stichprobenumfang und -beschreibung

Die Stichprobe bestand aus 14 professionellen Musikern im Alter von 22 bis 46 Jahren ($\bar{x}_{atithm} = 27{,}43$; $\sigma_x = 6{,}62$), darunter 13 Studenten der Universität der Künste Berlin. Alle Versuchspersonen verfügten über eine langjährige Aufführungserfahrung ($\bar{x}_{atithm} = 15{,}86$; $\sigma_x = 6{,}84$), sieben waren weiblich, sieben männlich. Gespielt wurden folgende Instrumente, in Klammern ist die Anzahl der Musiker angegeben: Violine (4), Cello (2), Klarinette (1), Trompete (2), Posaune (1), Oboe (2) und Fagott (1). Die Musiker wurden für ihre Teilnahme am Versuch finanziell entschädigt.

Durch die kleine Stichprobengröße ist die Übertragung der Studienergebnisse auf Musiker im Allgemeinen problematisch. Da sowohl Blechbläser, Holzbläser als auch Streicher an der Untersuchung teilnahmen, wurden zumindest die verschiedenen Instrumentengruppen eines Orchesters bis auf die Perkussionisten berücksichtigt.

Tabelle 3.5.: Informationen zur Stichprobe bezüglich Geschlecht, gespieltem Instrument, Alter, Konzerterfahrung und den gespielten musikalischen Passagen

		Instrument	Alter	Erfahrung	Musikalische Passagen
			in Jahren		
M1	♂	Violine	26	8	Bach, Partita II, 1. & 2. Satz
M2	♂	Violine	23	15	Bach, Sonate Nr. 1, g-moll für Violine Solo, BWV 1001, 3. Satz: Siciliano – Zimmermann: Sonate für Violine solo, 3. Satz: Toccata
M3	♂	Klarinette	23	17	Edison Denissow, Sonate für Klarinette 1. Satz
M4	♀	Violine	35	20	Bach, d-moll Gigue für Violine, Sarabande Solo
M5	♂	Cello	46	30	Bach, 1. Suite, 1. & 4.Satz
M6	♂	Trompete	33	27	Jean-Baptiste Lully, Marche – Georg Friedrich Händel, Sonate As-Dur (Original F-Dur), 3. Satz: Siciliano
M7	♀	Fagott	26	6	Mozart, Konzert K191, 1. Satz (Anfang bis Durchführung)
M8	♂	Trompete	29	16	Bach, Cello-Suiten, Sarabande, Suite 2 – Bach, Cello-Suiten,Gigue, Suite 1
M9	♀	Posaune	24	14	Launy Gröndahl, Konzerte für Posaune mit Klavier oder Orchester, 1. Satz Anfang
M10	♂	Cello	25	14	Bach, Suite Nr4 für Cello Solo, Ausschnitte
M11	♀	Violine	22	16	Bach, Sonata N2 für Violine Solo 1. Adagio – Sibelius, Konzert für Violine 1. Satz
M12	♀	Cello	22	9	Bach, Suite für Cello Solo, S3, Sarabande – Bach, Suite für Cello Solo, S3, Guige
M13	♀	Oboe	26	20	Tschaikowsky, 4. Sinfonie, 2. Satz – C. M. von Weber, Der Freischütz
M14	♀	Oboe	24	10	Dorati, Cinq Pieces pour le Itantbois, I. La cigale et la foumie – Telemann, aus 12 Fantasien für Querflöte ohne Bass, TWV 40:2-13, Nr.7

3.5. Messinstrument

Da eine unmittelbare Messung der menschlichen auditiven Wahrnehmung bislang nicht möglich ist, werden in den meisten Fällen Hörtests durchgeführt, in welchen die Versuchspersonen ihre Wahrnehmung durch Fragebögen quantifizieren (vgl. Bech und Zacharov, 2006, S. 1). Dass Musiker prinzipiell in der Lage sind, Auskunft über ihre auditive Wahrnehmung in Bezug auf die Raumakustik zu geben, haben die vorgestellten Studien in Kapitel 2 gezeigt. Um eine statistische Auswertung durch Produkt-Moment-Korrelationen zwischen raumakustischen Parametern und Musikerbewertungen und somit darauf aufbauende Methoden wie eine Faktorenanalyse zu ermöglichen, müssen die Bewertungen ferner als intervallskalierte Daten erfasst werden (Bortz und Schuster, 2010, S. 397). Da Ratingskalen als intervallskaliert angenommen werden können (vgl. Bortz und Döring, 2006, S. 176), galt es, zunächst einen Ratingfragebogen zur Abfrage raumakustischer Wahrnehmung zu erstellen, welcher die vorhandenen Wahrnehmungsdimensionen inklusive der für Musiker verständlichen Bezeichnungen beinhalten sollte. Anstatt die in anderen Studien wie beispielsweise von Gade (1989a) gefundenen Dimensionen zu verwenden, wurde ein Vorversuch zur Elizierung von Wahrnehmungsdimensionen durch die Musiker selbst durchgeführt. Anhand der Ergeb-

nisse wurden Ratingitems für den Hauptversuch ausgewählt. Auf diese Weise bestand die Möglichkeit zur Elizierung „neuer“ Dimensionen sowie die Überprüfung der bisher von anderen aufgestellten. Da zudem eine Beschränkung auf die Solistensituation vorlag, erschien eine Untersuchung der vorhandenen Wahrnehmungsdimensionen speziell für diese Situation als besonders sinnvoll.

3.5.1. Repertory-Grid-Technik

Eine Methode, welche sich zur Elizierung von Bewertungsdimensionen eignet, ist die *Repertory-Grid-Technik* (RGT), welche Kelly 1955 im Rahmen seiner Theorie der persönlichen Konstrukte vorstellt. Darin betrachtet er den „*Menschen als Wissenschaftler* [...], der immer den Lauf der Dinge, in die er verwickelt ist, vorherzusagen und zu kontrollieren sucht“, Hypothesen bildet, prüft und somit ein Weltbild beziehungsweise ein Konstruktsystem entwickelt (Kelly, 1986, S. 18–19; Bortz und Döring, 2006, S. 187). Die Konstrukte sind dabei als individuell, mit der Zeit veränderbar und handlungsleitend zu verstehen und können gleichzeitig gegensätzlicher Natur sein.

Um die bestehenden Konstrukte eines Probanden bezüglich eines Untersuchungsgegenstandes zu elizieren, vergleicht dieser bei der RGT normalerweise Triaden von Elementen miteinander und gibt an, inwiefern zwei dieser Elemente sich ähneln und gleichzeitig vom dritten unterscheiden. Elemente stellen die zu beurteilenden Objekte dar, in der Psychologie sind dies oft Personen wie Familienmitglieder oder Arbeitskollegen. Bei Übertragung der Methode auf die Audiokommunikation sind Wiedergabesysteme (Berg und Rumsey, 1999) oder wie in dieser Studie raumakustische Umgebungen als Elemente vorstellbar. Beim Vergleich der Elemente werden Gegensatzpaare, meist in Form von Adjektiven, gebildet wie bspw. hoch/tief, leise/laut usw. Obwohl man für gewöhnlich in seiner persönlichen Bewertung keine Gegensatzpole bildet, wird die bipolare Struktur gewählt, da der Mensch die Welt durch das gleichzeitige Registrieren von Ähnlichkeiten und Unterschieden versteht. Dem Probanden werden nach und nach alle möglichen Kombinationen der Elemente präsentiert, woraufhin wiederum Konstrukte gebildet werden. Anschließend wird ein sogenanntes Grid erstellt, mit welchem der Proband jedes Element anhand seiner zuvor selbst aufgestellten Konstrukte bewertet. Dies kann entweder durch dichotomische Zuordnung, Einstufung bzw. Rating oder Rangfolgenbildung geschehen (vgl. Berg und Rumsey, 1999, S. 5–10.).

Bei dieser Methode sind somit nur solche Kriterien für die Konstruktbildung entscheidend, welche der Proband selbst als entscheidend einstuft. Auch die Benennung der Konstrukte erfolgt allein durch den Probanden, weshalb davon ausgegangen werden kann, dass dieser im Stande ist, sie zu verstehen und zu bewerten.

Berg und Rumsey (1999, 2000, 2006) befassten sich mit der Konstitution und den Dimensionen subjektiver Qualitätsmerkmale bezüglich räumlicher Audiowiedergabe. Zur Identifizierung und Definition der Dimensionen stellten sie neben anderen Methoden die RGT und eine Studie vor, in welcher diese zum Einsatz kam. Durch die Ähnlichkeit der behandelten

Forschungsfragen wurde somit ein gutes Vorbild zur Anwendung in dieser Studie gegeben.

Die Stichprobe bei Berg und Rumsey bestand aus 18 Toningenieur-, Musik- und Medienstudenten, welche jeweils angaben, mehrmals in der Woche aktiv[16] Musik zu hören und somit von Berg und Rumsey im Vergleich zur Durschnittsbevölkerung als Expertenhörer eingestuft werden. Die zu vergleichenden Elemente waren sieben Wiedergabesysteme mit unterschiedlicher Anzahl von Kanälen (1 – 5) und Variationen bezüglich der Phase und des Anteils künstlichen Halls, welche als Triaden mit einem von sechs Programminhalten dargeboten wurden. Während der Konstruktelizierung wurden alle gebildeten Konstrukte aufgenommen, auch solche, welche sich aus den Unterschieden der beiden ähnlichen Elemente ergaben. Beschrieb ein Konstrukt die persönliche Präferenz, nahm der Versuchsleiter dieses nicht auf, sondern fragte, auf welche Art und Weise ein Element gegenüber dem anderen bevorzugt wurde. Durch dieses sogenannte *Laddering* konnten neue Konstrukte bestimmter Qualitätsdimensionen eliziert werden. Im Ratingteil bewerteten die Testpersonen neun Stimuli bezüglich ihrer eigenen Konstrukte auf 5-Punkt-Skalen. Um alle Konstrukte bewerten zu können, wurde den Probanden eine mehrmalige Wiederholung der präsentierten Stimuli ermöglicht.

Von allen 18 Versuchspersonen wurden insgesamt 342 Konstrukte gebildet. Die Hälfte der Stichprobe, welche die zusätzliche Instruktion erhalten hatte, nur auf die Unterschiede der dreidimensionalen Gestalt der Schallquellen und ihrer Umgebung zu achten, bildete im Median 26 Konstrukte. Der Medianwert der anderen Testpersonen lag bei 30. Für Berg und Rumsey deutete dieser Unterschied darauf hin, dass Probanden durch Instruktionen fokussierter urteilen. Im ersten Schritt der Auswertung wurden die Konstrukte in semantische Gruppen eingeteilt, welche sofern räumlichen Charakters mit den Oberbegriffen Authentizität, Quellbreite, Umhüllung und Tiefe beschrieben wurden. Zudem wurde die *Verbal Protocol Analysis* (VPA) angewendet, um die Konstruktbezeichnungen aufgrund logischer Bedeutung, Stimuluszugehörigkeit und semantischer Gesichtspunkte in beschreibend (*descriptive*) oder einstellungsbezogen (*attitudinal*) einzuteilen. Somit konnte sich in der weiteren Auswertung auf die 228 deskriptiven Konstrukte konzentriert werden. Die bislang lediglich auf den Bezeichnungen der Konstrukte basierenden Gruppierungen ergänzten Berg und Rumsey durch die Anwendung einer Clusteranalyse, welche die Daten der Ratings zur Grundlage hat. Die Clusteranalyse sollte aufzeigen, welche deskriptiven, subjektiven Konstrukte bezüglich der Stimulibewertung ähnlich benutzt wurden und somit möglicherweise ein ähnlich wahrgenommenes Ereignis beschreiben. Mithilfe der Daten aus der Clusteranalyse, dargestellt in Dendogrammen, wurden Gruppen gebildet, welche anhand ihrer Konstruktbezeichnungen durch die Versuchsleiter benannt wurden.

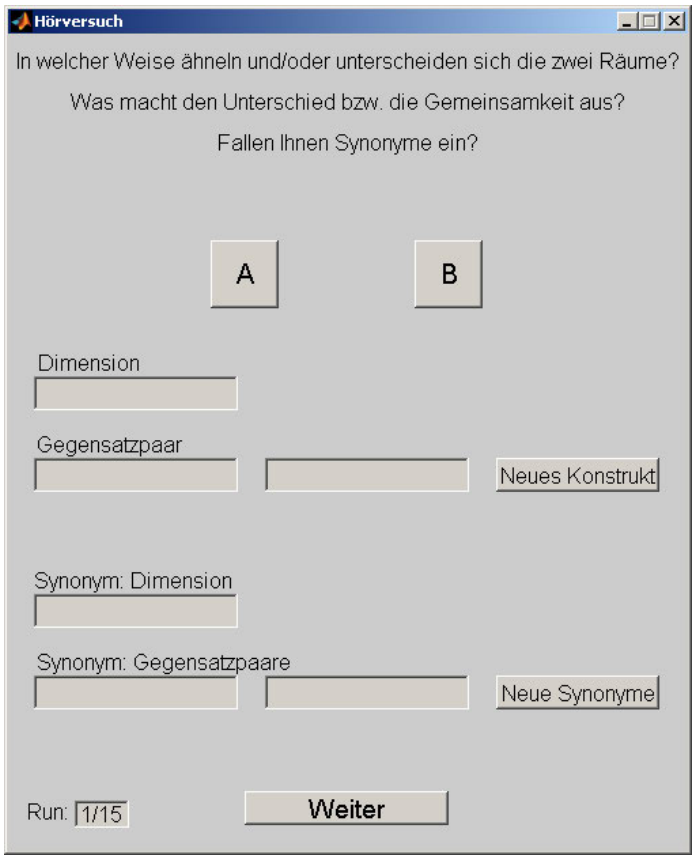


Abbildung 3.7.: Bedienoberfläche, mit welcher der Versuchsleiter die Konstrukte während des RGT-Versuchs notierte

3.5.2. RGT-Vorversuch zur Fragebogenerstellung

Um einen geeigneten Rating-Fragebogen zu entwickeln, wurden in einem Vorversuch mithilfe der RGT Gegensatzpaare zur Beschreibung der während des Musizierens wahrgenommenen Raumakustik eliziert. Die untersuchte Stichprobe bestand aus sechs Probanden, welche eine Woche später auch am Hauptversuch teilnahmen: M5, M6, M7, M9, M11 und M14 (siehe Tabelle 3.5). Dabei spielten sie die gleichen musikalischen Passagen wie im Hauptversuch.

Auch wenn es sich bei den Versuchspersonen um professionelle Musiker handelte, kann nicht davon ausgegangen werden, dass diese innerhalb ihrer regulären Praxis mit der Bewertung raumakustischer Umgebungen vertraut sind. Daher ist die RGT bei der Elizierung der Wahrnehmungsdimensionen offenen Interviews bezüglich der Erfahrungen der Musiker vorzuziehen (vgl. Berg, 2006, S. 3).

Die Untersuchung zweier Streicher (Cello,Violine), zweier Blechbläser (Trompete, Posaune) sowie zweier Holzbläser (Fagott, Oboe) sollte eine möglichst breite Variation bezüglich der Instrumente darstellen. Als Elemente wurden die Raumsimulationen R1–R6 (siehe Tabelle 3.1) verwendet, wodurch das von Danielsson als geeignet angegebene Minimum ein-

[16] Aktives Musikhören besteht, wenn neben dem Musikhören keine weitere Tätigkeit ausgeführt wird.

gehalten wurde (Danielsson, 1991, zit. in Berg und Rumsey, 1999, S. 6). Wesentlich für die Elizierung eines Konstrukts ist der Unterschied zwischen den Stimuli bezüglich der Wahrnehmungsdimension. Verhalten sich alle Elemente im Hinblick auf ein bestimmtes Phänomen gleich, besteht die Gefahr, dass das dazugehörige Konstrukt nicht eliziert wird, auch wenn nach Gemeinsamkeiten gefragt wird. Aus diesem Grund wurde bei der Auswahl der Elemente solche gewählt, welche sich anhand der raumakustischen Parameter möglichst stark unterschieden. Während der Elizierung wurden den Musikern die Räume in Diaden dargeboten, anhand derer die Konstrukte gebildet wurden. In einer Studie von Choisel und Wickelmaier (2006) zeigte sich, dass bei Verwendung von Diaden im Vergleich zur Darbietung von Triaden eine größere Anzahl von Konstrukten gebildet wurde. Deshalb wurde angenommen, dass bei der Verwendung von Diaden eine geringere Konstruktmaskierung (*construct masking*) besteht, bei welcher dominantere Konstrukte die Bildung weniger dominanter verhindern. Bech und Zacharov (2006, S. 56) räumten zudem ein, die Nutzung von Triaden würde nicht zwangsweise zur Bildung bipolarer Konstrukte führen, was wiederum einen Vorteil von Diaden darstelle. Desweiteren wird durch Nutzung von Diaden die Versuchsdauer verkürzt, da einerseits bei jeder Kombination nur zwei Räume angespielt werden und sich andererseits weniger Kombinationen der Elemente ohne Berücksichtigung der Reihenfolge ergeben.[17]

Nachdem die Musiker in beiden Raumsimulationen des jeweiligen Paares ihre gesamte Passage gespielt hatten, wurden ihnen folgende Fragen gestellt.

- Auf welche Weise ähneln und/ oder unterscheiden sich die beiden Räume?
- Was macht den Unterschied bzw. die Gemeinsamkeit aus?
- Bilden Sie ein Gegensatzpaar aus beschreibenden Adjektiven, Substantiven oder Beschreibungen.
- Geben Sie die Dimension an, in welcher sich die Unterschiede oder Gemeinsamkeiten zeigen.
- Fallen Ihnen Synonyme zum gebildeten Gegensatzpaar oder der genannten Dimension ein? Wenn ja, geben Sie diese bitte an.

Die Antworten wurden vom Versuchsleiter festgehalten und in eine Ratingumgebung in Matlab integriert. Die Bildung mehrerer Gegensatzpaare während eines Vergleichs war möglich und wurde vom Versuchsleiter erfragt, um wiederum den Effekt des (*construct masking*) zu minimieren (vgl. Berg, 2006, S. 3–4). Bei Nennung von Konstrukten bezüglich der Präferenz wurde die Versuchsperson zusätzlich gefragt, in welcher Dimension sich die Präferenz ausdrücke, um ein beschreibendes Konstrukt bilden zu können (vgl. Berg und Rumsey, 1999, S. 17). Für die Musiker bestand weiterhin die Möglichkeit kein neues Konstrukt zu bilden, falls sie entweder nur Unterschiede bereits gebildeter Konstrukte wahrnahmen oder keine Unterschiede erkennen konnten. Falls gewünscht, konnten die Musiker Teile der Passagen in den jeweilgen Raumsimularionen mehrfach spielen.

[17] Bei Verwendung von Triaden ergeben sich bei 6 Elementen (Räumen) $\binom{6}{3} = 20$ mögliche Kombinationen, bei Diaden sind es $\binom{6}{2} = 15$.

Die Aufgabe des Versuchsleiters war es, die Fragen an die Probanden zu richten sowie die dadurch gewonnen Daten zu dokumentieren. Zur Eingabe der gebildeten Konstrukte wurde die in Abbildung 3.7 gezeigte Bedienoberfläche genutzt. Da die Raumsimulationen aufgrund der Wiedergabe über Kopfhörer für den Versuchsleiter nicht hörbar waren und dieser zudem aufgrund der randomisierten Reihenfolge der Diaden nicht wusste, welche Räume die Musiker hören, war eine Beeinflussung in eine von ihm gewünschte Richtung nicht möglich. Verständnisfragen können natürlich auch eine Beeinflussung der Probanden darstellen, waren allerdings aufgrund des nicht einheitlichen Vokabulars dringend erforderlich. Zur Simulation der raumakustischen Umgebung wurde der gleiche technische Versuchsaufbau wie im Hauptversuch verwendet (siehe Abschnitt 3.2).

Nachdem die Elizierungsphase nach 15 Diaden abgeschlossen war, bewerteten die Musiker die Räume R1–R9 anhand ihrer eigenen Konstrukte auf kontinuierlichen Ratingskalen, wozu sie wiederum in allen Räumen die gewählte Passage spielten. Auch hier war es ihnen möglich, Teile der Passagen während der Bewertung noch einmal zu spielen, um auf die einzelnen Aspekte genauer achten zu können. Die neun Raumsimulationen wurden randomisiert dargeboten. Je nach Bedarf legten die Musiker zwischen Elizierungsphase und Rating eine Pause ein. Die Elizierungsphase dauerte zwischen 30 und 45 Minuten, der Ratingteil beanspruchte eine ähnlich lange Zeitspanne, wodurch ein kompletter Durchgang des RGT-Versuchs inklusive Kalibrierungsmessungen 90 bis 120 Minuten in Anspruch nahm.

3.5.3. Zusammenstellung der Ratingitems

Der RGT-Versuch diente vornehmlich dem Zweck, anhand der subjektiven Konstrukte einen allgemeinen Fragebogen für den Hauptversuch erstellen zu können. An dieser Stelle soll daher ein Überblick über die elizierten Konstrukte und das zur Zusammenstellung der Ratingitems verwendete Verfahren gegeben werden.

Von den sechs Probanden wurden insgesamt 90 Konstrukte gebildet, die maximale Konstruktanzahl einer Versuchsperson war 18, die minimale 10. Der Median der Konstruktanzahl über die sechs Versuchspersonen errechnete sich zu $Md = 15{,}5$; der arithmetische Mittelwert betrug $\bar{x} = 15$. Da Probandin M7 nach dem Rating angab, drei ihrer elizierten Konstrukte ergeben für sie keinen Sinn mehr, wurden diese nicht in die Auswertung miteinbezogen. Es handelte sich um die Gegensatzpaare *groß und breit – quadratisch*, *Kirche – gleichmäßiges Echo* und *Direktklang klein, Echo groß – Direktklang groß, Echo nicht viel.*

Berg und Rumsey (1999, S. 7) zufolge sind Konstrukte sowohl als individuell als auch als allgemein zu begreifen. Der Mensch reagiere nicht auf einen physikalischen Stimulus, sondern seine Wahrnehmung des Stimulus, welche wiederum durch die bereits bei der Versuchsperson bestehenden Konstrukte bestimmt wird. Selbst die allgemeinsten Konzepte werden einzigartig verstanden. Werden Erfahrungskonstrukte ähnlich verwendet, seien allerdings auch die psychologischen Prozesse als ähnlich anzusehen. Die Verwendung der individuellen

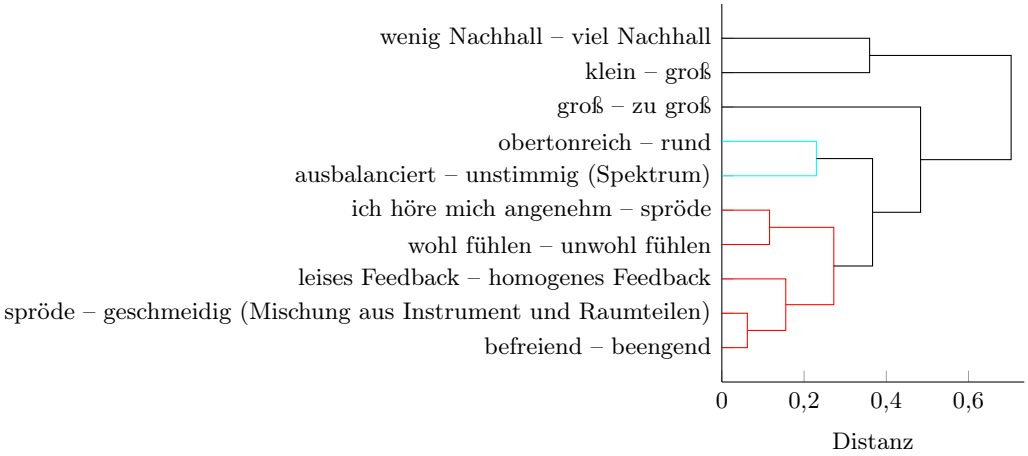


Abbildung 3.8.: Dendrogramm der Clusteranalyse der von M5 (Cello) elizierten und bewerteten Gegensatzpaare

Konstrukte kann aus den Bewertungen der Elemente, im Fall dieser Studie der Räume, abgeleitet werden.

Um neben den semantischen Bedeutungen der Konstrukte über eine weitere Entscheidungsgrundlage für die Zusammenstellung der Ratingitems zu verfügen, wurde je Versuchsperson sowie über die Konstrukte aller Probanden anhand der Raumbewertungen eine Clusteranalyse durchgeführt. Durch die Einzelanalysen sollte die Intra-Rater-, in der globalen Analyse auch die Inter-Rater-Ähnlichkeit der Konstrukte beschrieben werden (vgl. Berg und Rumsey, 1999, S. 20). Die Distanzen wurden anhand der Korrelationen der Bewertungen errechnet, da „weniger der Abstand der Objektprofile, sondern deren Ähnlichkeit aufgrund der Profilverläufe“ (Bortz und Schuster, 2010, S. 458) von Interesse war. Um die Forderung gleicher Mittelwerte und Streuungen der Merkmale zu erfüllen, wurden die Raumbewertungen vor Berechnung der Korrelationen z-transformiert, wodurch zudem bei unterschiedlicher Nutzung der Skalen eine Vergleichbarkeit der Urteile hergestellt werden konnte (vgl. Bortz und Döring, 2006, S. 185). Da die Ähnlichkeit der Konstrukte unabhängig von deren Polarität in die Analyse eingehen sollte, wurden die Vorzeichen der Korrelationswerte vernachlässigt. Als Fusionskriterium wurde die UPGMA-Methode (*Unweighted Pair Group Method with Arithmetic Mean*) verwendet, da sich Average-Linkage-Methoden bei Korrelationsdistanzen als präzise erwiesen haben (vgl. Scheibler und Schneider, 1985).

Die explorative Clusteranalyse ist eine Methode, welche Vorschläge möglicher Zusammenhänge anbietet, ihre schematische Prozedur allein ist frei von jeglicher Bedeutung. Die Nutzung und Auswertung ist bedingt durch den Kontext des Problems, das Wissen der Untersuchenden bezüglich dieses Kontexts sowie den Forschungsgegenstand, die Ergebnisse sind in erster Linie als Anregung zu verstehen (vgl. Anderberg, 1971, S. 34–35). Bei der Auswahl der Ratingsitems boten die Dendrogramme, welche die Distanzen zwischen den Konstrukten visualisieren, neben den Konstruktbezeichnungen sowie den bisherigen Erkenntnissen des Forschungsgebiets eine Hilfestellung. Die Clustereinteilungen wurden dabei

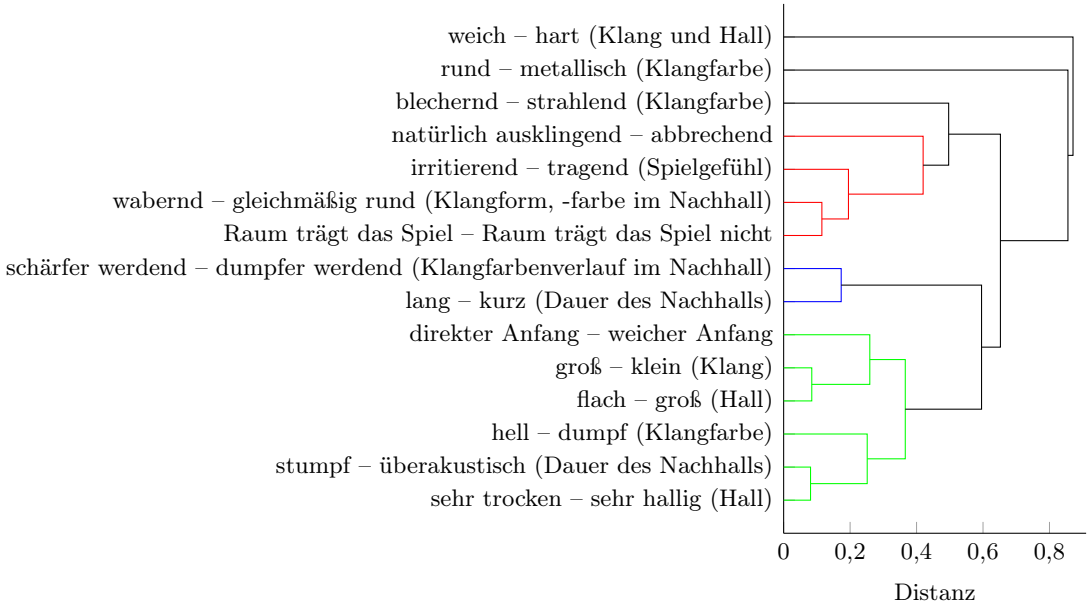


Abbildung 3.9.: Dendrogramm der Clusteranalyse der von M6 (Trompete) elizierten und bewerteten Gegensatzpaare

jedoch nicht als stringente Vorgaben betrachtet.

Während des Auswahlprozesses der Ratingitems für den Hauptversuch wurden die elizierten Konstrukte jeder einzelnen Versuchsperson anhand der individuellen Dendrogramme begutachtet. Die Zugehörigkeit der semantischen Bezeichnungen zu übergeordneten Wahrnehmungsdimensionen wurde hergestellt, anhand derer wiederum ein Vergleich zwischen den Versuchspersonen stattfand. Für Konstrukte, deren Bedeutung die Versuchsleiter als ähnlich einschätzten, wurde eine möglichst für alle Musiker verständliche Bezeichnung gewählt. Neben der genutzten Semantik der Gegensatzpaare bezog diese Methode die Kenntnisse vorangegangener Studien der Bühnen- und Auditoriumsakustik mit ein. Bei der Zusammenstellung der Ratingitems wurde darauf geachtet, dass möglichst alle von den Versuchsleitern als plausibel eingestuften Konstrukte der sechs Versuchspersonen repräsentiert wurden. Das Dendrogramm der globalen Clusteranalyse wurde als zusätzliches Hilfsmittel herangezogen. Es bot vor allem einen guten Überblick über alle genannten Konstrukte, die gebildeten Cluster stellten sich jedoch selten als plausibel heraus. Die Zusammenstellung der Items fand in Zusammenarbeit mit Prof. Dr. Stefan Weinzierl und Zora Schärer Kalkandjiev M.A. statt, wodurch ein fundiertes Expertenwissen vorausgesetzt werden kann.

Die Methode der Itemzusammenstellung beinhaltet durch die Interpretation der Versuchsleiter selbstverständlich eine gewisse Verzerrung (Bias), welche jedoch bei der Generalisierung individueller Daten nicht zu verhindern ist (vgl. Berg und Rumsey, 2000, S. 9). Die Verzerrung kann als sehr viel geringer angenommen werden als solche, die bei Nutzung vorgegebener Skalen auftritt, welche ausschließlich auf dem Vokabular des Forschers fußen.

Die Ergebnisse der Clusteranalysen der individuellen Grids sind durch die Dendrogram-

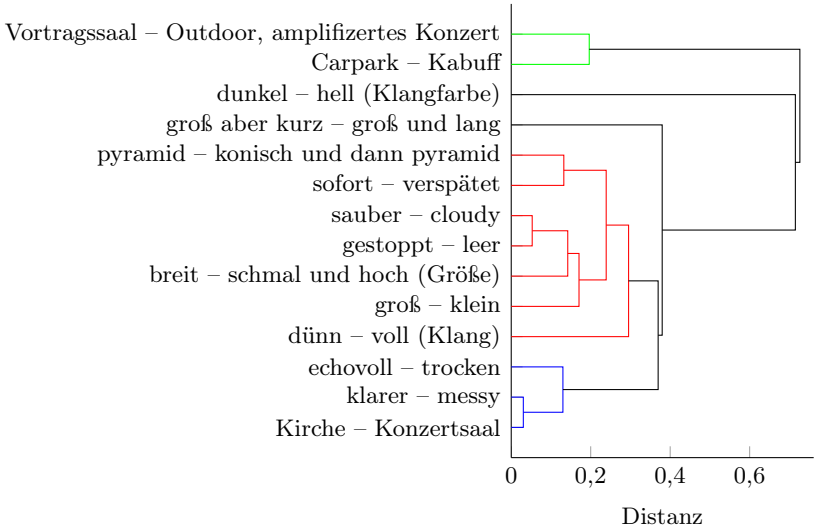


Abbildung 3.10.: Dendrogramm der Clusteranalyse der von M7 (Fagott) elizierten und bewerteten Gegensatzpaare

me in den Abbildungen 3.8–3.13 dargestellt. Ausgehend von semantischen Gesichtspunkten nannten alle Versuchspersonen Gegensatzpaare bezüglich der Länge des Nachhalls und der Klangfarbe. Auch die beiden anderen von Gade (1989a) bekannten Solisten-Dimensionen, Unterstützung und Dynamik, können in den Antworten wieder gefunden werden. Die Benennungen der Gegensatzpaare lassen darauf schließen, dass die Versuchspersonen zwar nicht über ein präzises gemeinsames Vokabular bezüglich der wahrgenommenen Raumakustik verfügen, es aber durchaus Überschneidungen gibt. Konstrukte, welche beispielsweise die Länge des Nachhallvorgangs beschreiben, werden nicht vollkommen gleich benannt, beinhalten jedoch häufig den Konstruktpol *trocken*:

- echovoll – trocken (M7)
- trocken – warm (M11)
- Raum gibt zurück – immer trocken (M11)
- verspätet, Echo – trocken (M11)
- natürlich trocken – verstärkt tiefe Töne (M11)
- trocken – langer Nachhall (M14)
- sehr trocken – sehr hallig (M6)
- trocken – mehr Nachklang (M9)

Aufgrund der jeweiligen Gegensatzpole kann vermutet werden, dass *trocken* nicht in jedem Fall das gleiche Phänomen beschreibt. Vor allem die Mehrfachnennung durch Versuchsperson M11 deutet darauf hin, dass Schwierigkeiten bei der präzisen Benennung der gedachten Konstrukte bestanden und sich die aufgeladene Bedeutung des Begriffs *trocken* von Konstrukt zu Konstrukt untescheidet. Ferner fällt auf, dass Probandin M11 für das Gegensatzpaar *hohe Dynamik – piano forte* die Dimension *Länge* angab. Diese Auffälligkeiten können unter Umständen darin begründet liegen, dass die Muttersprache dieser Versuchspersonen

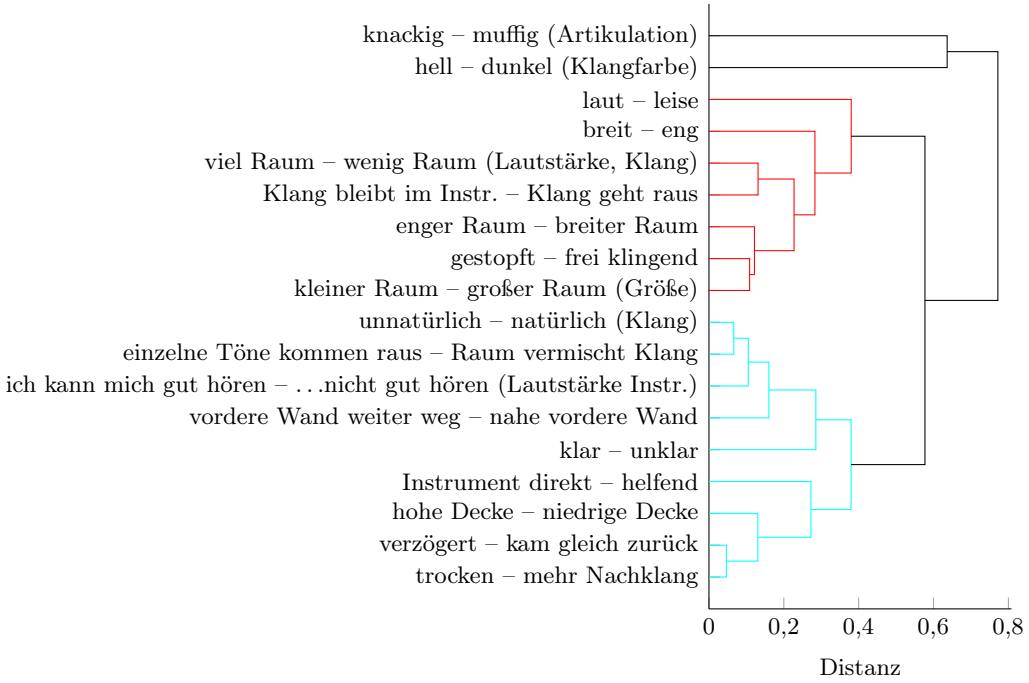


Abbildung 3.11.: Dendrogramm der Clusteranalyse der von M9 (Posaune) elizierten und bewerteten Gegensatzpaare

nicht die deutsche Sprache war.

Die Ergebnisse der Clusteranalyse mit den Konstrukten aller Versuchspersonen sind in Abbildung 3.15 mithilfe eines Dendrogramms visualisiert. Neben den Gegensatzpaaren ist die Nummer der Versuchsperson, welche das Konstrukt nannte, sowie die Dimensionbezeichnung, falls vorhanden, angegeben.

Beim Blick auf das Dendrogramm in Abbildung 3.15 zeigt sich keine konsistente Gruppierung der Konstrukte. Beispielsweise werden die Gegensatzpaare *laut – leise* von M9 und M14 zwar gleich benannt, zeigen in den Ergebnissen der Clusteranalyse jedoch keine große Ähnlichkeit. Bei *groß – klein* bzw. *klein – groß* eliziert von M5, M6 und M7 zeigt sich ein ähnliches Bild. M6 gab als Dimension *Klang* an, während M5 *groß* als *positives Merkmal* beschrieb und M7 *leer* als Synonym für *groß* und *rund* als Synonym für *klein* nannte. Außerdem werden *großer Raum – kleiner Raum* und *großer Klang – kleiner Klang* jeweils von M14 zu keinem der drei vorher genannten gruppiert, sondern sind in einem gemeinsamen Cluster zu finden. Konstrukte gleicher semantischer Bezeichnungen beschreiben somit nicht zwangsweise gleiche Phänomene bzw. ergeben sich anhand der Bewertungen keine hohen Korrelationen.

Auffallend ist zudem, dass Konstrukte geringer Distanz häufig von derselben Versuchsperson stammen, obwohl sie nach semantischen Gesichtspunkten nicht immer Ähnlichkeiten aufweisen, sondern unterschiedliche Dimensionen zu beschreiben scheinen:

- wabernd – gleichmäßig rund / Raum trägt das Spiel – Raum trägt das Spiel nicht

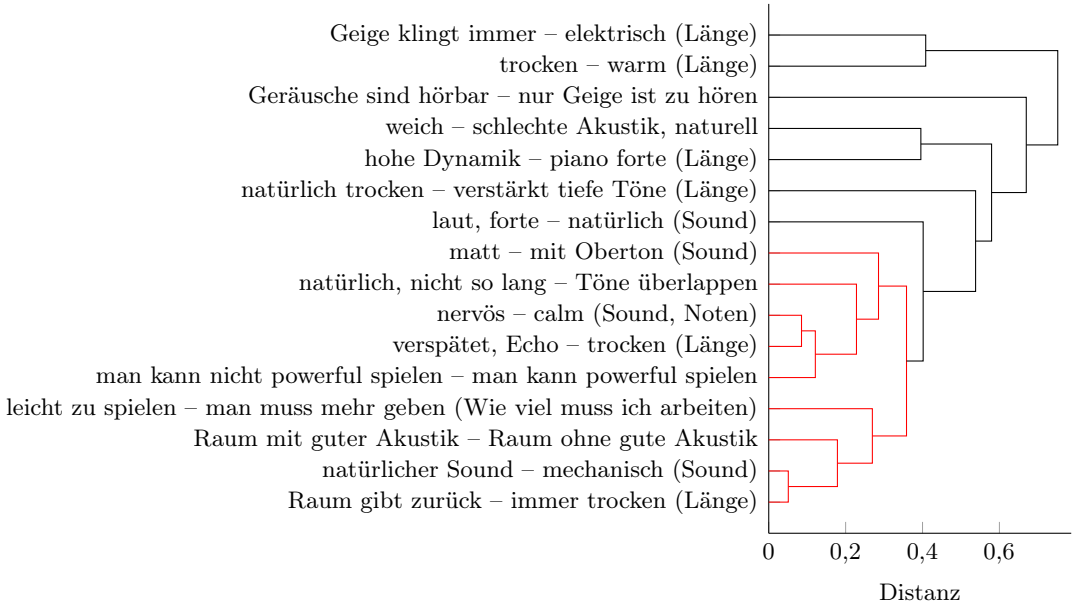


Abbildung 3.12.: Dendrogramm der Clusteranalyse der von M11 (Violine) elizierten und bewerteten Gegensatzpaare

(M6)
- dunkel – grell / direkte Ansprache – weiche Ansprache (M14)
- natürlicher Sound – mechanisch / Raum gibt zurück – immer trocken (M11)
- Klang wird diffus – Nachhall abgegrenzt / diffus – klar (M14)
- ich höre mich angenehm – spröde / wohl fühlen – unwohl fühlen (M5)
- spröde – geschmeidig / befreiend – beengend (M5)
- sauber – cloudy / gestoppt – leer (M7)
- nervös – calm / verspätet, Echo – trocken (M11)
- verzögert – kam gleich zurück / trocken – mehr Nachklang (M9)
- klarer – messy / Kirche – Konzertsaal (M7)
- unnatürlich – natürlich / Einzelne Töne kommen raus – Raum vermischt Klang (M9)
- starke Tragfähigkeit – weniger starke Tragfähigkeit / trocken – langer Nachhall / großer Raum – kleiner Raum / wenig Resonanz – viel Resonanz / großer Klang – kleiner Klang (M14)

Das Struktogramm in Abbildung 3.16, welches die Clusteranzahl gegenüber der mit den Fusionstufen bedingten Distanzdifferenz zwsichen den Clustern plottet, zeigt einen beinahe stetigen Abfall. Sprünge, welche auf die Anzahl der zu wählenden Clusteranzahl hindeuten (vgl. Bortz und Schuster, 2010, S. 464), finden sich in relativ kleiner Ausprägung bei Fusionsstufe 6 und 16. Eine eindeutige Festlegung auf die Anzahl der zu bildenden Cluster kann daher nicht erfolgen.

Diese Beobachtungen deuten die Problematik der Zusammenstellung der Ratingitems an. Überschneidungen der Konstrukte sind auf semantischer Ebene zwar vorhanden, die Zusammenhänge werden jedoch durch die Ergebnisse der Clusteranalysen nicht unbedingt bestätigt. Diese Beobachtungen decken sich mit denen von Berg und Rumsey (2000, S. 9)

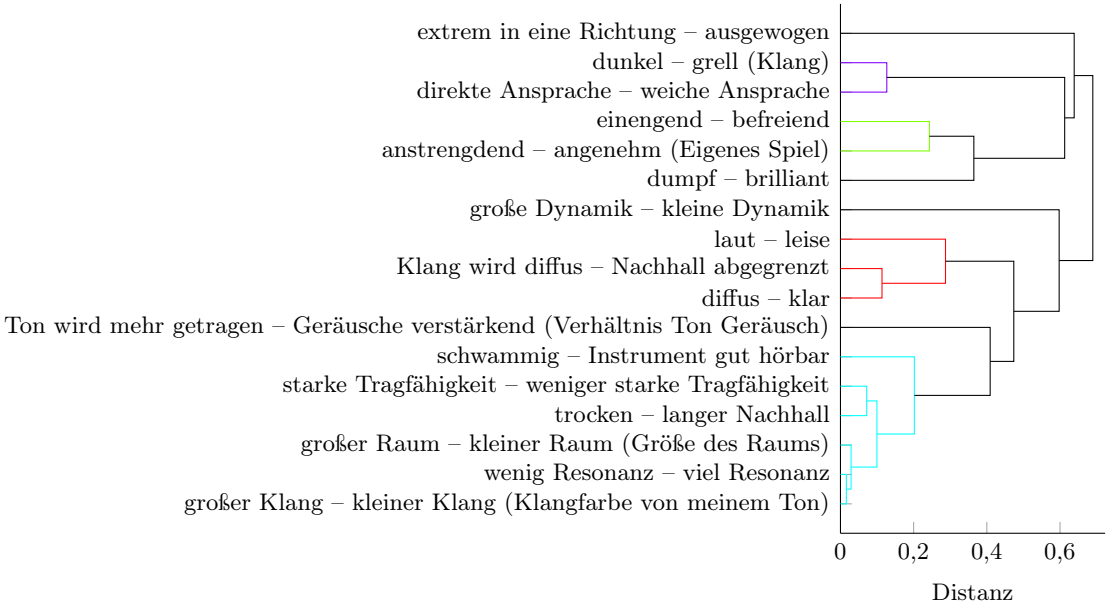


Abbildung 3.13.: Dendrogramm der Clusteranalyse der von M14 (Oboe) elizierten und bewerteten Gegensatzpaare

im Bereich der Wahrnehmung räumlicher Dimensionen bei der Audiowiedergabe.

Tabelle 3.6 listet die zusammengestellten Ratingitems sowie die subjektiven Konstrukte, aus welchen sie gebildet wurden, auf. Bei der Verwendung semantischer Differentiale gibt Nunnally (1978, S. 611) zu bedenken, dass die Bedeutung der Skalen abhängig von den zu bewertenden Konzepten seien. Da raumakustische Umgebungen sehr komplexe Stimuli darstellen, wurden die Gegensatzpaare durch Erläuterungen der zu bewertenden Dimensionen ergänzt, welche den Versuchspersonen vorgelegt wurden. Da Berg und Rumsey (2006) schildern, dass unpräzise Definitionen der verwendeten Attribute zu geringer Urteilerkonsistenz führten, sollten die Erklärungen mögliche Irritationen verhindern. Zusätzlich wurden die Versuchpersonen gefragt, ob ihnen die vorgelegten Gegensatzpaare und ihre Erläuterungen verständlich seien, um etwaige Unklarheiten vor dem Rating klären zu können.

Die Liste der Ratingitems deckt alle von Gade (1989a) genannten Solo-Dimensionen Halligkeit, Unterstützung, Dynamik und Klangfarbe ab. Da diese einzelnen Dimensionen im RGT-Versuch von den Versuchspersonen noch einmal in verschiedene Aspekte unterteilt wurden, wurden zur Beschreibung der Halligkeit vier (I3, I4, I10, I17) der Unterstützung drei (I1, I2, I11) und der Klangfarbe wiederum drei (I9, I16, I19) Gegensatzpaare in den Fragebogen aufgenommen. Zur Beschreibung der Dynamik wurde hingegen nur ein Aspekt genannt und somit berücksichtigt (I12). Die Einteilung ist dabei jedoch nicht trennscharf zu verstehen, da sich die verschiedenen Aspekte innerhalb des komplexen Stimulus der Raumakustik überschneiden und auch in Abhängigkeit zueinander stehen. I14, I15 und I20 können von vornherein nicht in eine von Gade genannte Dimension eingeteilt werden, sie beinhatlten sowohl Aspekte der Klangfarbe als auch der Halligkeit. Durch I13 wird ein Gegensatzpaar

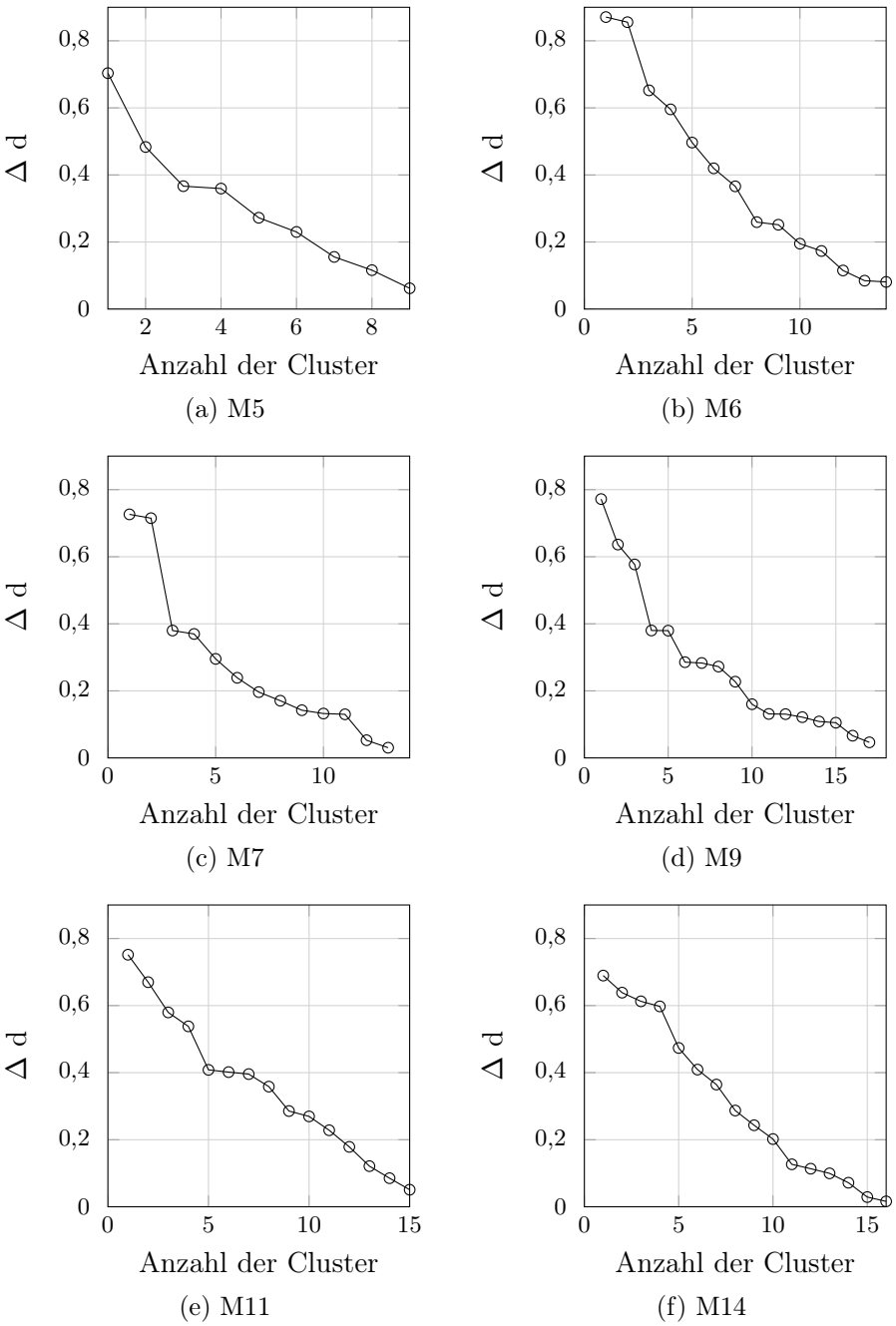


Abbildung 3.14.: Struktogramme der individuellen Clusteranalysen

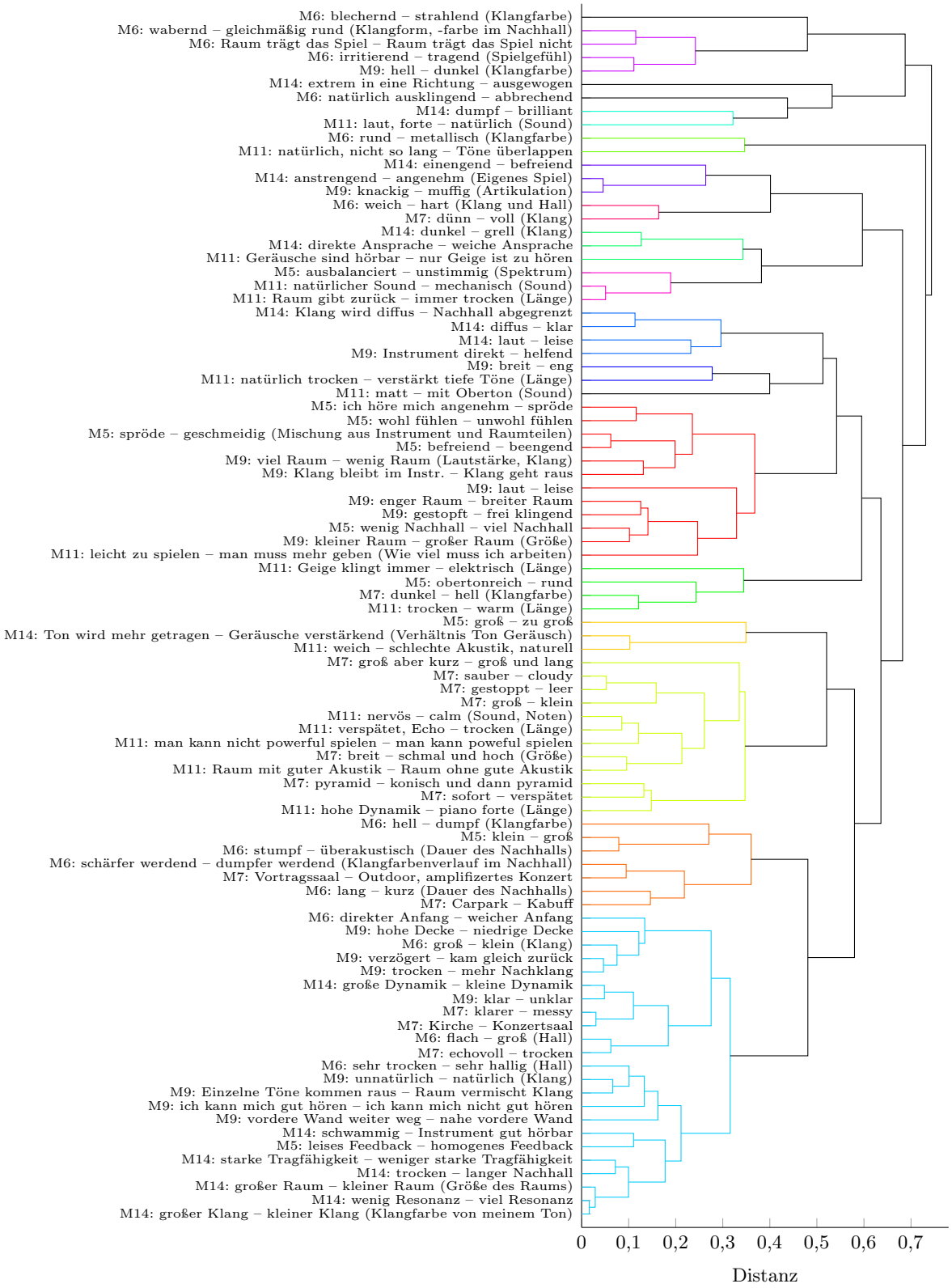


Abbildung 3.15.: Dendrogramm einer Clusteranalyse anhand der Bewertungen der individuellen Konstrukte aller Probanden. Die farbliche Einteilung in Cluster erfolgt ab der Häfte der maximalen Distanz ($0{,}5 \cdot 0{,}74 = 0{,}37$).

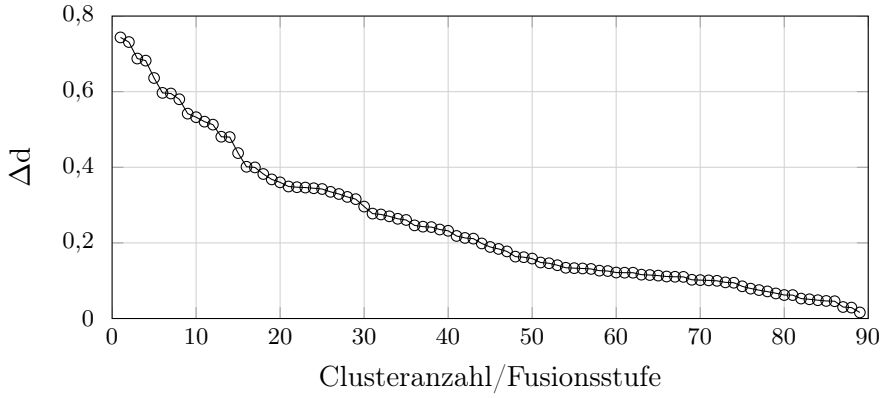


Abbildung 3.16.: Struktogramm der Clusteranalyse über die gebildeten Konstrukte aller Probanden.

der Dimension der Deutlichkeit aufgenommen, welche bereits von Wilkens (1977) in der empirischen Auditriumsakustik verwandt wurde. I5, I7, und I8 stellen Items dar, welche die wahrgenommene Raumgröße abfragen, welche anhand der dargebotenen Raumakustik wahrgenommen wird. Außerdem sollte durch I6 und I18 die Präferenz der Probanden abgefragt werden, da diese in vorrangegangenen Studien oft die einzig abgefragte Dimension darstellte. Die 20 zusammengestellten Items stellen somit einen rein aus semantischer Sicht umfassenden Abfragekatolog zur Bewertung raumakustischer Umgebungen aus Musikerperspektive dar.

Bei der Datenerhebung durch Ratingskalen können systematische Urteilsfehler auftreten, welche die Verwendbarkeit der Urteile unter Umständen einschränken. Ist eine Versuchsperson entweder nicht imstande oder nicht bereit dazu, auf die unterschiedlichen Ausprägungen der Merkmale eines Objekts bei der Bewertung einzugehen, spricht man vom sogenannten Halo-Effekt. Er kann zu einer ähnlichen Bewertung vieler Items, beispielsweise anhand eines Pauschalurteils, führen. Der Halo-Effekt tritt vor allem bei schwer zu fassenden, schlecht definierten Merkmalen auf. Um den Effekt im Fall dieser Studie zu minimieren, wurden die Items möglichst genau beschrieben. Durch die Komplexität der zu bewertenden Stimuli lassen sich jedoch mögliche globale Bewertungstendenzen anhand dominanter Merkmale nicht ausschließen. Eine skalenweise Beurteilung, die zu einer Minimierung von Halo-Effekten führen soll, wurde aus Gründen der dadurch erhöhten Versuchsdauer nicht in Betracht gezogen (vgl. Bortz und Döring, 2006, S. 183). Milde-Härte-Fehler treten auf, wenn die Elemente pauschal zu gut oder zu schelcht bewertet werden. Mögliche Gründe eines Fehlers in dieser Studie wäre beispielsweise die fehlende Akzeptanz der Simulationsumgebung. Allerdings würden nur solche Items betroffen sein, bei welchen überhaupt eine Präferenzabfrage vorgenommen wird. Beim Fehler der zentralen Tendenz werden die Extremwerte der Ratingskalen nicht genutzt. Versuchspersonen lassen diese Bereiche beispielsweise bewusst frei, um Bewertungsspielraum für nachfolgende Stimuli zu reservieren. Die Auswirkungen von Milde-Härte-Fehlern sowie zentralen Tendenzen können durch die Standardisierung bzw. z-

Transformation der gemessenen Skalenwerte je Versuchsperson minimiert werden. Primacy-Recency-Effekte beziehen sich auf die Darbietungsreihenfolge der Elemente, welche in die Bewertung miteinfließen kann. Um systematische Fehler zu vermeiden, wurde die Reihenfolge der Raumsimulationen zwischen den Versuchspersonen variiert.

Die Items wurden auf kontinuierlichen Skalen mit einem Wertebereich von ±5 bewertet. Der Schieberegler befand sich vor der Bewertung in Mittelstellung, wo sich zudem die einzige visuelle Marke befand. Die Darbietung der Items wurde je Versuchsperson randomisiert, um mögliche Effekte ausgehend von der Bewertungsreihenfolge zu verhindern.

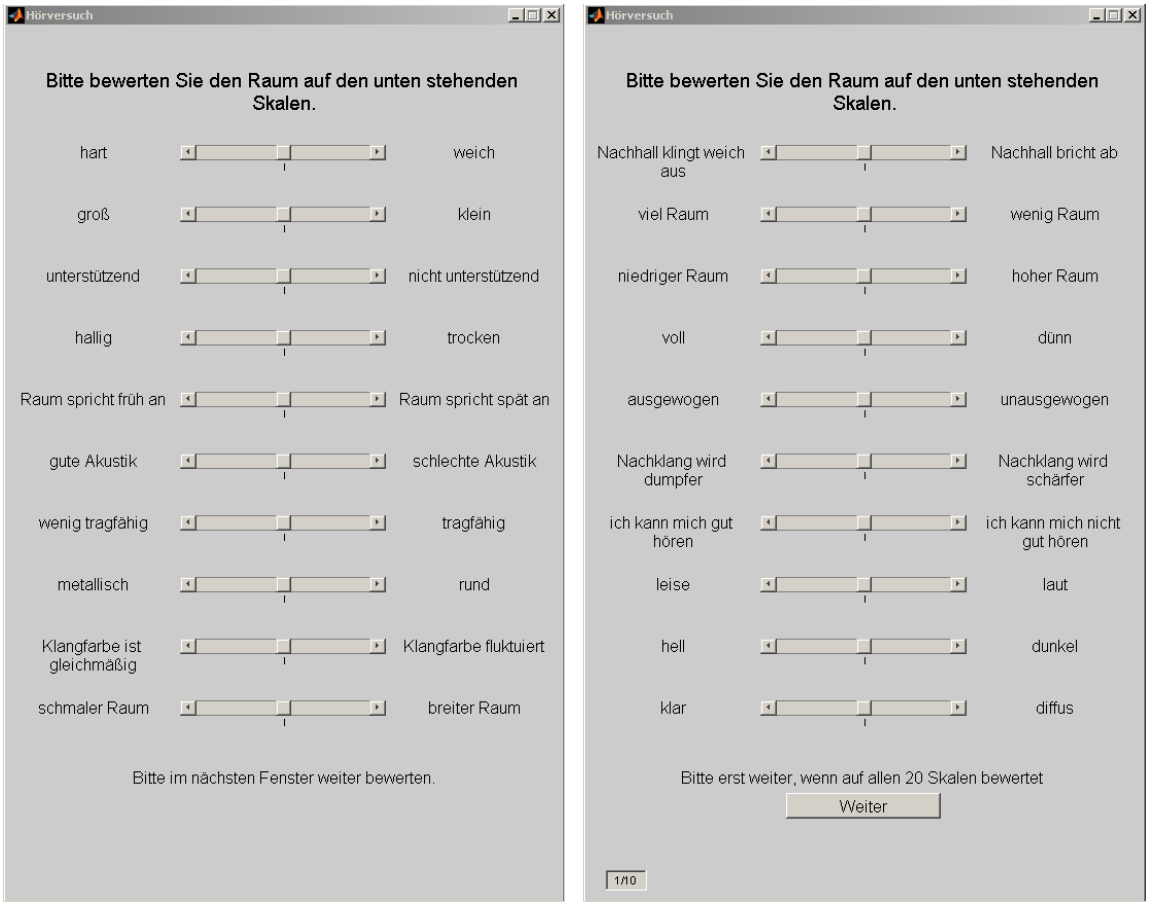


Abbildung 3.17.: Bedienoberfläche des Ratingversuchs mit beispielhafter Itemreihenfolge

Tabelle 3.6.: Auswahl der Ratingitems für den Hauptversuch

	Gegensatzpaar		Beschreibung	Abgedeckte Konstrukte
I1	tragfähig	wenig tragfähig	Trägt der Raum das Spiel oder nicht?	M14: starke Tragfähigkeit – weniger starke Tragfähigkeit, M6: Raum trägt das Spiel – Raum trägt das Spiel nicht
I2	ich kann mich gut hören	ich kann mich nicht gut hören	Können Sie Ihr Instrument gut hören?	M9: ich kann mich gut hören – ich kann mich nicht gut hören
I3	trocken	hallig	Hat der Raum einen langen oder kurzen Nachhall?	M6: sehr trocken – sehr hallig, M11: Raum gibt zurück – immer trocken, M5: wenig Nachhall – viel Nachhall, M6: lang – kurz, M9: trocken – mehr Nachklang, M7: echovoll – trocken, M14: trocken – langer Nachhall
I4	Raum spricht früh an	Raum spricht spät an	Kommt der Raumklang gleich oder verzögert?	M9: verzögert – kam gleich zurück, M6: direkter Anfang – weicher Anfang, M14: direkte Ansprache – weiche Ansprache, M7: sofort – verspätet, M9: knackig – muffig
I5	groß	klein	Handelt es sich um einen großen oder kleinen Raum?	M5: klein – groß, M7: groß – klein, M9: kleiner Raum – großer Raum, M5: groß – zu groß, M14: großer Raum – kleiner Raum, M6: groß – klein
I6	gute Akustik	schlechte Akustik	Gefällt die Akustik oder nicht?	M11: Raum mit guter Akustik – Raum ohne gute Akustik, M5: wohl fühlen – unwohl fühlen, M5: ich höre mich angenehm – spröde, M11: weich – schlechte Akustik, naturell
I7	hoher Raum	niedriger Raum	Wirkt der Raum niedrig oder hoch?	M7: breit – schmal und hoch, M9: hohe Decke – niedrige Decke, M6: flach – groß
I8	breiter Raum	schmaler Raum	Wirkt der Raum breit oder schmal?	M9: breit – eng, M9: enger Raum – breiter Raum, M5: befreiend – beengend, M14: einengend – befreiend
I9	dunkel	hell	Wie ist die Klangfarbe des Raums, des Nachhalls?	M9: hell – dunkel, M14: dunkel – grell, M7: dunkel – hell, M6: hell – dumpf, M14: dumpf – brilliant, M11: matt – mit Oberton, M5: obertonreich – rund
I10	viel Raum	wenig Raum	Wie ist die Balance zwischen Instrument und Raumklang?	M9: viel Raum – wenig Raum

Tabelle 3.6.: (Fortsetzung) Auswahl der Ratingitems für den Hauptversuch

	Gegensatzpaar		Beschreibung	Abgedeckte Konstrukte
I11	unterstützend	nicht unterstützend	Hilft der Raum beim Spielen?	M9: Instrument direkt – helfend, M11: leicht zu spielen – man muss mehr geben, M11: man kann nicht powerful spielen – man kann poweful spielen
I12	laut	leise	Nehmen Sie Ihr Instrument laut oder leise wahr?	M9: laut – leise, M11: laut, forte – natürlich, M14: laut – leise
I13	diffus	klar	Ist der Raumklang durchmischt oder klar?	M14: Klang wird diffus – Nachhall abgegrenzt, M14: diffus – klar, M14: schwammig – Instrument gut hörbar, M7: sauber – cloudy, M7: klarer – messy, M9: Einzelne Töne kommen raus – Raum vermischt Klang, M9: klar – unklar
I14	dünn	voll	Hat der Raum einen dünnen oder vollen Klang?	M7: dünn – voll
I15	weich	hart	Wirkt der Raum weich oder hart?	M6: weich – hart
I16	rund	metallisch	Hat der Raum einen dumpfen oder eher einen metallischen Klang?	M6: rund – metallisch
I17	Nachhall klingt weich aus	Nachhall bricht ab	Wie klingt der Nachhall aus?	M6: natürlich ausklingend – abbrechend
I18	ausgewogen	unausgewogen	Bietet der Raum ein in sich stimmiges Gesamtbild?	M14: extrem in eine Richtung – ausgewogen, M5: ausbalanciert – unstimmig
I19	Klangfarbe ist gleichmäßig	Klangfarbe fluktuiert	Bleibt die Klangfarbe im Nachklang gleichmäßig oder schwankt sie?	M6: wabernd – gleichmäßig rund
I20	Nachklang wird schärfer	Nachklang wird dumpfer	Wie ändert der Nachklang im Verlauf seine Klangfarbe?	M6: schärfer werdend – dumpfer werdend

3.6. Versuchsablauf

Da bei der durchgeführten Studie die Wahrnehmung von Musikern in Solo-Situationen von Interesse war, wurden die Versuche mit den Probanden einzeln durchgeführt. Während die Versuchspersonen sich an den reflexionsarmen Raum gewöhnen konnten, wurde ihnen der grobe Ablauf des Einpegelns des Instrumentenmikrofons, der Kalibrierungsmessungen und des Ratingversuchs erklärt.

Das Mikrofon wurde am Instrument angebracht, die Versuchsteilnehmer spielten die vorbereitete Passage, währenddessen der Eingangspegel am *Motu Audio Express* angepasst, der Kopfhörerweg von einem Versuchsleiter überprüft sowie die Dauer der Passage gemessen wurde. Im Anschluss wurde die Passage mit dem Versuchsteilnehmer angepasst, falls diese zu lang oder nicht abwechslungsreich genug war. Anschließend folgte die Pegelkalibrierung sowie die Entzerrungsmessung des Kopfhörers und der Außenohren mithilfe von In-Ear-Mikrofonen, bei welcher der Versuchsperson der Kopfhörer aufgesetzt wurde. Dieser erste Teil des Versuchs dauerte in etwa zwanzig Minuten.

Am Anfang des eigentlichen Versuchs las die Versuchsperson eine Liste mit den zwanzig zu bewertenden Gegensatzpaaren mit zusätzlichen Erläuterungen durch (siehe Abbildung A.1) und wurde gebeten, bei Unklarheiten nachzufragen. Nach einem ersten Probedurchlauf, bei welchem nach gespielter Passage auf einem bereitgestellten Laptop die Bewertung eines über Kopfhörer gehörten Raums vorgenommen wurde, verließ der Versuchsleiter nach Klärung aller entstandenen Fragen den Raum und die Versuchsperson konnte nun eigenhändig die zehn Räume bewerten. Dabei war es ihr gestattet, die Passage während des Ratings öfter anzuspielen, um auf die einzelnen Bewertungsdimensionen genauer achten zu können. Die Versuchsleiter konnten über Video- und Audiosignal den Versuch mitverfolgen und bei Problemen eingreifen, bzw. weiterhelfen. Die Reihenfolge der Items des Fragebogens wurde von Versuchsperson zu Versuchsperson randomisiert, um etwaige im letzten Abschnitt genannte Messfehler zufällig zu verteilen. Die Reihenfolge in Abbildung 3.17 stellt somit nur ein Beispiel dar. Das Rating an sich nahm zwischen 30 und 45 Minuten in Anspruch.

3.7. Statistische Auswertung des Ratingversuchs

Die gewonnenen Daten des Ratingversuchs bestehen aus einer dreidimensionalen Datenmatrix mit den Dimensionen Ratingitems (20), Versuchspersonen (14) und bewerteten Räumen (10), welche die Grundlage der statistischen Auswertung darstellt. Nach Visualisierung der Raumbewertungen folgt in diesem Abschnitt eine detaillierte Beschreibung der Auswertungsmethoden sowie der resultierenden Ergebnisse. Schlüsse und Erkenntnisse aus den statitsischen Methoden werden erarbeitet und abschließend in Abschnitt 3.8 unter Berücksichtigung des zu Anfang dieser Arbeit vorgestellten Forschungsstandes diskutiert.

Zur Prüfung der Unterscheidbarkeit der dargebotenen Raumsimulationen durch die Ratingskalen werden Varianzanalysen durchgeführt. Die Reliabilität der Versuchspersonen be-

ziehungsweise die Konsistenz der Urteile wird mithilfe der Intraklassenkorrelation geprüpft. Durch die Anwendung eines Mehrebenenmodells wird auch der Einfluss der Instrumente auf die Intraklassenkorrelation berücksichtigt. Die Zusammenhänge zwischen den einzelnen Bewertungsitems werden durch Faktorenanalysen untersucht. Sowohl die Mittelwerte über alle Probanden als auch über die einzelnen Instrumentengruppen sowie die individuellen Ratings sollen in diesem Zusammenhang in die Analyse miteingehen. Um mögliche Beziehungen zwischen den verwandten Items und raumakustischen Parametern aufdecken zu können, werden die entsprechenden Korrelationskoeffizienten berechnet, bei welchen lineare Zusammenhänge berücksichtigt werden. Durch die Ergebnisse von Powerberechnungen können schließlich Empfehlungen bezüglich der Stichprobengröße für nachfolgende ähnliche Studien gegeben werden.

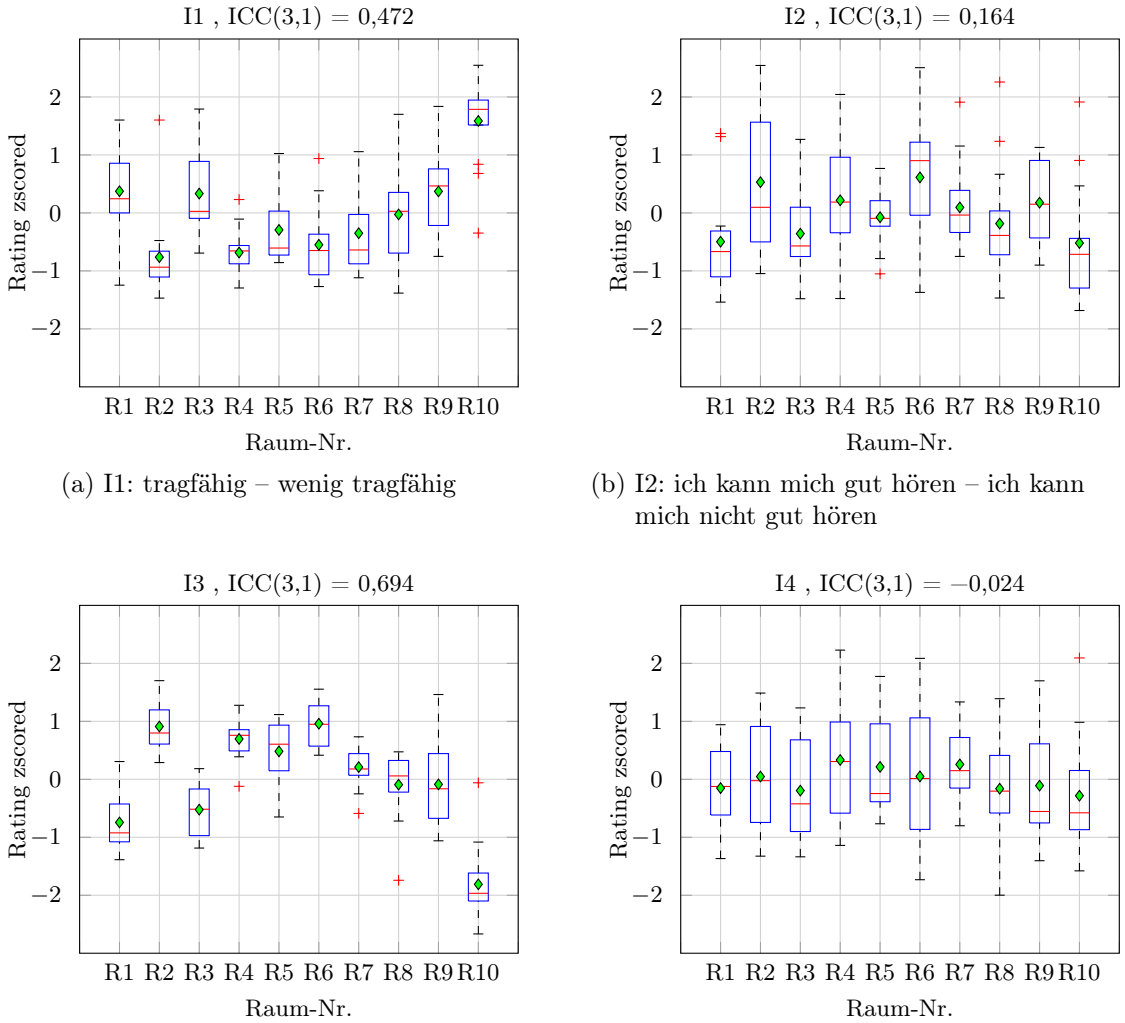
3.7.1. Darstellung der Ratingergebnisse

Zur Visualisierung der Ergebnisse des Hauptversuchs werden die z-transformierten Ratings pro Gegensatzpaar durch Boxplots dargestellt (siehe Abbildung 3.18), dabei sind zusätzlich die Werte des in Abschnitt 3.7.3 beschriebenen Reliabilitätsmaßes ICC(3,1) aufgetragen. Neben dem Median (−), den Quartilen (□), dem Wertebereich ($\overset{\top}{\perp}$) und den Ausreißern (+) sind auch die arithmetischen Mittelwerte in grün (♦) geplottet. Die Ausreißergrenzen sind um den 1,5-fachen Interquartilsabstand größer bzw. kleiner als der jeweilige Angelpunkt des ersten und dritten Quartils. Negative Werte tendieren zum erstgenannten Pol der Ratingskalen, postive zum zweitgenannten.

Die z-Transformation ermöglicht die Vergleichbarkeit unterschiedlicher Verteilungen (vgl. Benninghaus, 1998, S. 157). Um sowohl die Bewertungen zwischen den Versuchspersonen sowie die einzelnen Items innerhalb der Versuchspersonen vergleichbar zu machen, erfolgt die Standardisierung im vorliegenden Fall anhand der Mittelwerte und der Standardabweichungen der 10 Raumbewertungen je Item und je Versuchsperson. Zwischen den Probanden kann sich beispielsweise die Ausnutzung der Skalen prinzipiell unterscheiden. Als Beispiel sei hier noch einmal auf den Fehler der zentralen Tendenz verwiesen, welcher besteht, wenn Versuchspersonen extreme Bewertungen vermeiden. Durch die Standardisierung wird angenommen, dass trotz der Definition der Ratingskalen und der gegebenen Anweisungen die gleichen Skalenpositionen für unterschiedliche Versuchspersonen verschiedene Bedeutungen haben (vgl. Osgood et al., 1967, S. 236). Die dadurch entstehende Verlust von Informationen bezüglich der individuellen Bewertungsunterschiede, welche nach Osgood et al. unter anderem Rückschlüsse auf die Persönlichkeit, den Intellekt oder die Situation von Versuchspersonen zulassen, wird an dieser Stelle in Kauf genommen.

Bei Betrachtung der Boxplots fallen bei einigen Items bereits große Streuungen der Bewertungen auf, welche ähnlich hoch erscheinen wie die Streuungen über die Merkmalsvariation der Räume. Die größte Konsistenz der Beurteilungen scheint bei *I3: trocken – hallig* (siehe Abb. 3.18c), *I10: viel Raum – wenig Raum* (siehe Abb. 3.18j) und den größenbeschreibenden

Items (I5: Abb. 3.18e, I7: Abb. 3.18g, I8: Abb. 3.18h) zu bestehen, während bei Attributen der Klangfarbe oder der Präferenz große Quartile zu beobachten sind. Diese ersten Beobachtungen gilt es in den folgenden Abschnitten statistisch zu prüfen.



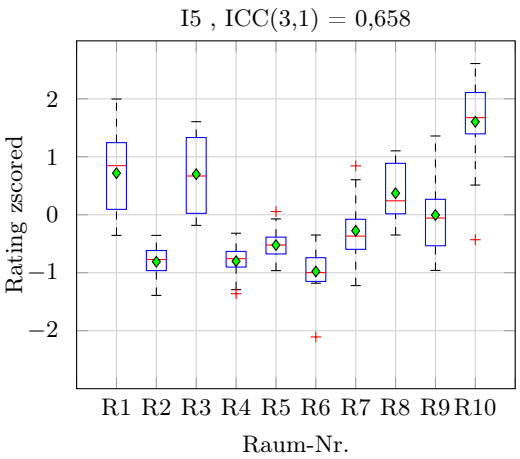
(a) I1: tragfähig – wenig tragfähig

(b) I2: ich kann mich gut hören – ich kann mich nicht gut hören

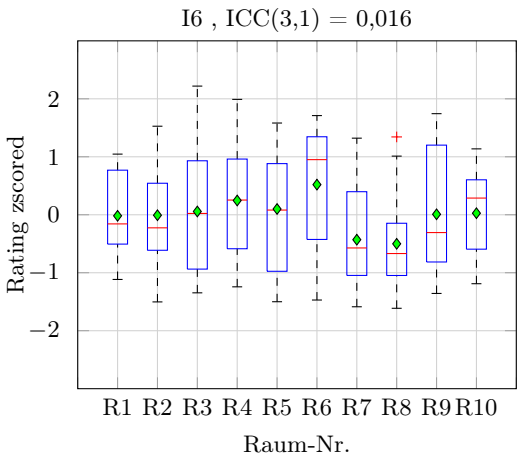
(c) I3: trocken – hallig

(d) I4: Raum spricht früh an – Raum spricht spät an

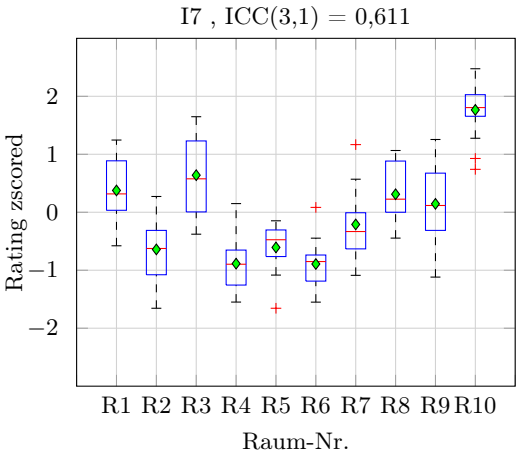
Abbildung 3.18.: Boxplots der z-transformierten Ratings mit aufgetragenen Medianen (–), Quartilen (□), verwendeten Wertebereichen zur Quartilsbestimmung (⊤ ⊥), Ausreißern (+) und arithmetischen Mittelwerten (♦).



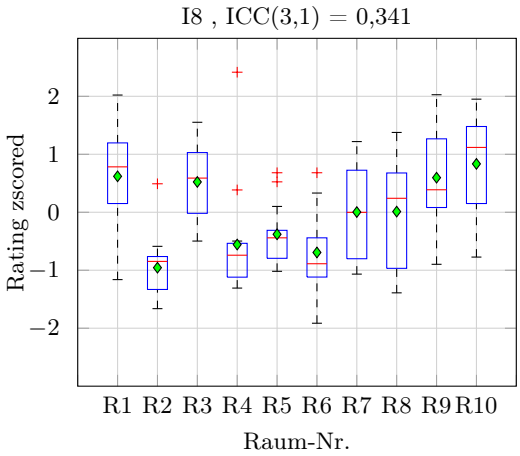
(e) I5: groß – klein



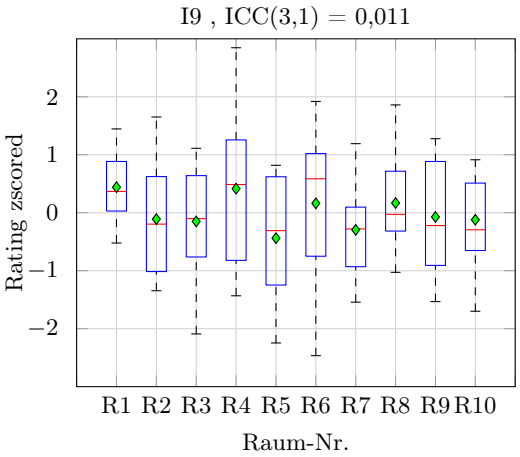
(f) I6: gute Akustik – schlechte Akustik



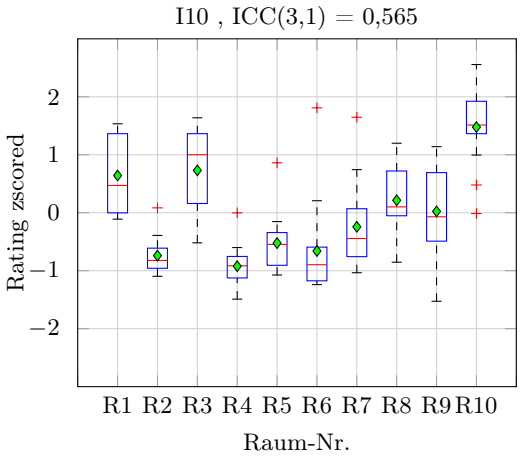
(g) I7: hoher Raum – niedriger Raum



(h) I8: breiter Raum – schmaler Raum

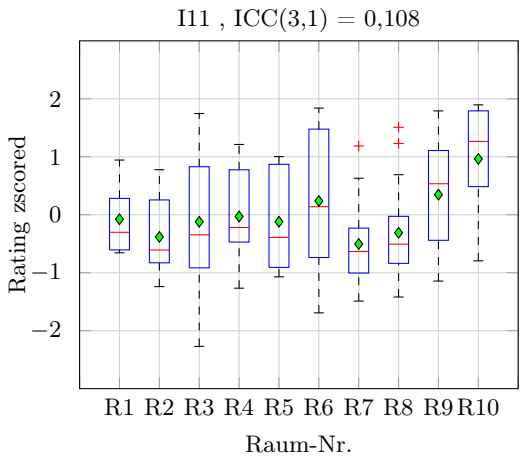


(i) I9: dunkel – hell

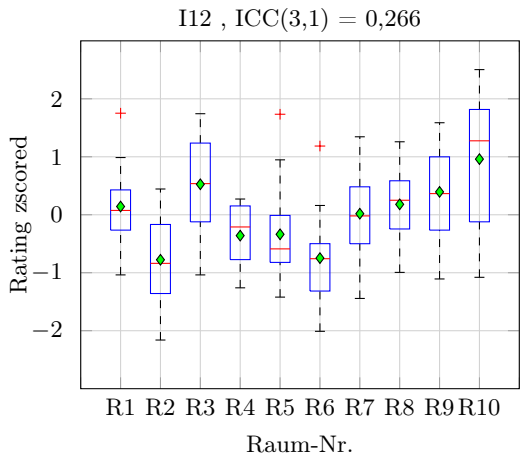


(j) I10: viel Raum – wenig Raum

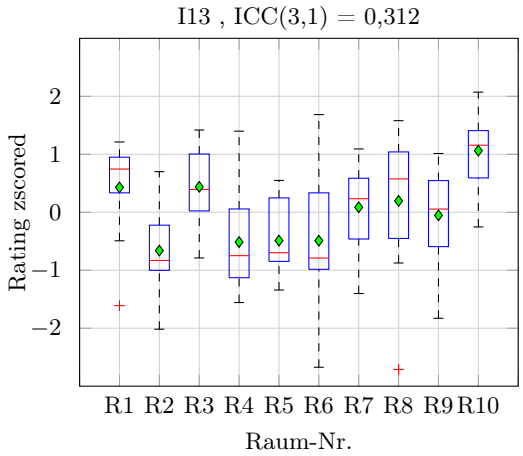
Abbildung 3.18.: (Fortsetzung 1) Boxplots der z-transformierten Ratings



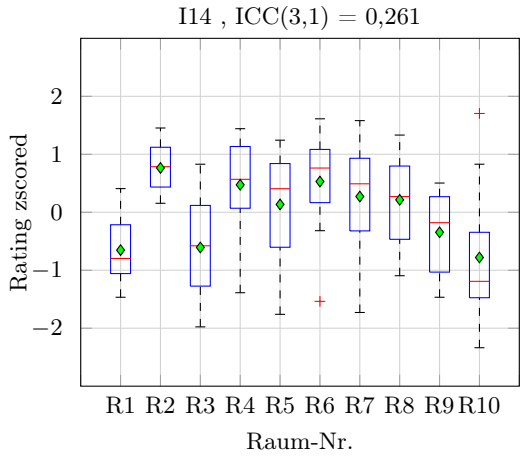
(k) I11: unterstützend – nicht unterstützend



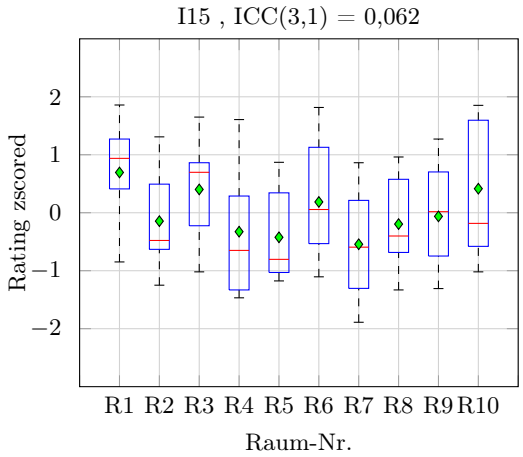
(l) I12: laut – leise



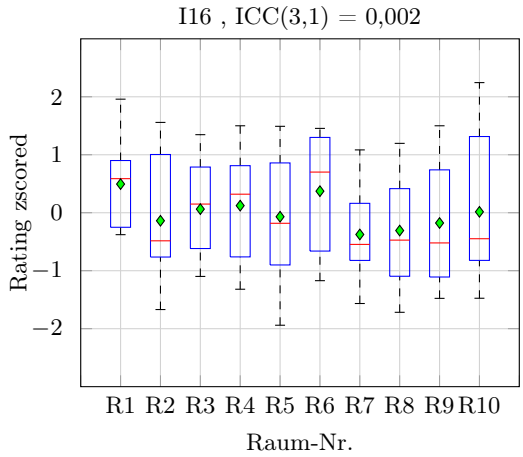
(m) I13: diffus – klar



(n) I14: dünn – voll

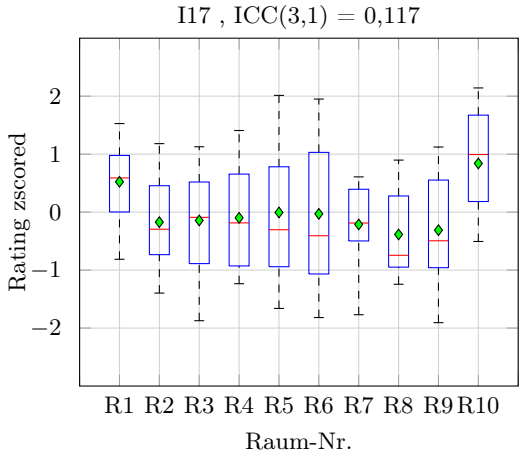


(o) I15: weich – hart

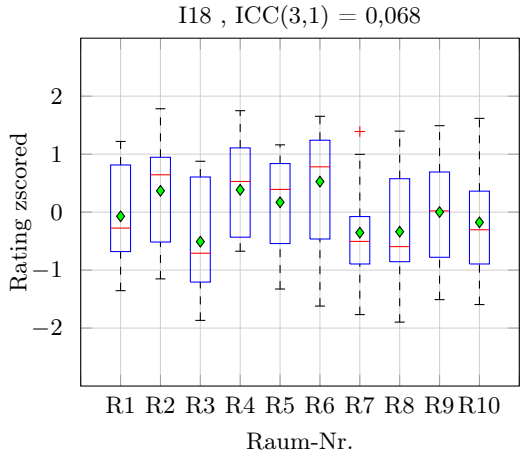


(p) I16: rund – metallisch

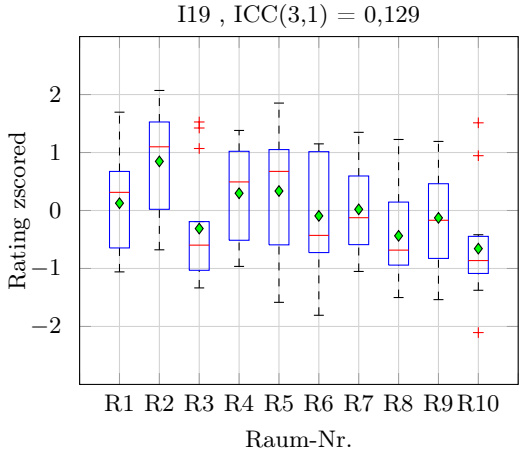
Abbildung 3.18.: (Fortsetzung 2) Boxplots der z-transformierten Ratings



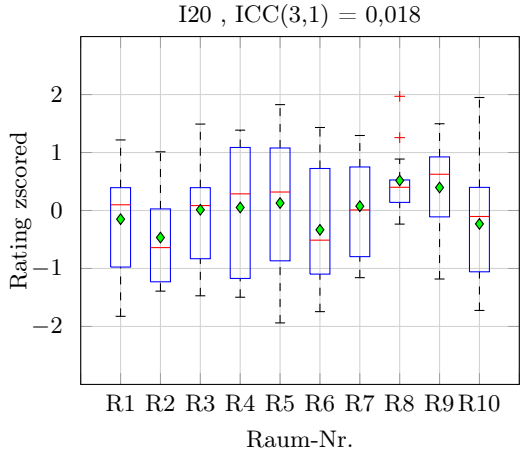
(q) I17: Nachhall klingt weich aus – Nachhall bricht ab



(r) I18: ausgewogen – unausgewogen



(s) I19: Klangfarbe ist gleichmäßig – Klangfarbe fluktuiert



(t) I20: Nachklang wird schärfer – Nachklang wird dumpfer

Abbildung 3.18.: (Fortsetzung 3) Boxplots der z-transformierten Ratings

3.7.2. Überprüfung des Messinstruments

Um zu testen, ob der Fragebogen bestehend aus den zwanzig durch den RGT-Versuch elizierten Gegensatzpaaren zur Unterscheidung der zehn dargebotenen Räume prinzipiell geeignet war, wurde in *SPSS* durch Benutzung des Befehls *GLM* eine multivariate einfaktorielle Varianzanalyse durchgeführt. Der Faktor Raum, welcher innerhalb der Versuchspersonen variiert wurde, wurde dabei als Messwiederholungsfaktor bzw. Innersubjektfaktor aufgefasst. Die Fragestellung an die MANOVA mit Messwiederholung kann ebenso lauten, ob die Varianz innerhalb der dargebotenen Raumsimulationen zu einer signifikanten Unterscheidung der Bewertungen durch die Versuchspersonen geführt hat und sich dementsprechend die Auswahl der Stiumuli als geeignet herausstellt.

Da geprüft werden soll, ob sich die Messwiederholungen „insgesamt, also in Bezug auf alle berücksichtigten abhängigen Variablen unterscheiden, ist ein multivariater Mittelwertvergleich durchzuführen“ (Bortz und Schuster, 2010, S. 472), durch welchen eine Kontrolle des globalen Alpha-Niveaus ermöglicht wird. Desweiteren werden die Korrelationen zwischen den Itembewertungen, welche in Abschnitt 3.7.5 genauer betrachtet werden, anders als im univariaten Fall berücksichtigt (vgl. Bray und Maxwell, 1985, S. 10–11). Bei der Berechnung in *SPSS* wurden Quadratsummen des Typs III verwendet.

Ein Signifikanztest der MANOVA mit Messwiederholung testet die Nullhypothese

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k,$$

wobei μ_k jeweils den Mittelwertsvektor der k-ten Messwiederholung über alle gemessenen abhängigen Variablen darstellt (vgl. Bray und Maxwell, 1985, S. 15; Bortz und Schuster, 2010, S. 285). Die Widerlegung der Nullhypothese kann daher nur den Schluss zulassen, dass sich die Bewertungen mindestens eines Raums im Mittel bezüglich mindestens eines Ratingitems von denen eines anderen Raums so deutlich unterscheiden, dass sich auf Grundlage des gewählten Signifikanzniveaus ein signifikanter Effekt zeigt.

Annahmen bzw. Vorrausseztungen einer einfaktoriellen MANOVA mit Messwiederholung sind eine multivariate Normalverteilung der Daten sowie die Unabhängigkeit der Innersubjektkontraste (vgl. O'Brien und Kaiser, 1985), welche durch Randomisierung der Raumdarbietungen angenommen werden kann. Laut O'Brien und Kaiser sind Hypothesentests der Annahmen nicht empfehlenswert, da diese normalerweise nicht aufzeigen, ob eine Annahme in solchem Maße verletzt wurde, dass Probleme bei der eigentlichen Prozedur auftreten. Trotzdem wurden zur Verifizierung in *Matlab* zwanzig Henze-Zirkler-Tests[18] über alle Items durchgeführt. Sie zeigten bei den z-transformierten Daten in keinem der Fälle eine multivariate Normalverteilung. Die MANOVA gilt jedoch wie die ANOVA im Allgemeinen als sehr robust in Bezug auf die Verletzung dieser Annahme (vgl. Bray und Maxwell, 1985, S. 33).

[18] Zur Durchführung des Henze-Zirkler-Tests wurde das Matlab-Skript *HZmvntest* von Antonio Trujillo-Ortiz verwendet. Quelle: http://www.mathworks.de/matlabcentral/fileexchange/17931-hzmvntest; letzter Zugriff: 5. Juni 2013.

Tabelle 3.7.: Teststatistiken des Innersubjektfaktors Raum aus der MANOVA mit Messwiderholungen über alle Items I1 –I 20 mit z-transformierten Daten

	Wert	F	Hypothese df	Fehler df	Sig./p	part. η^2
Pillai-Spur	2,063	1,576	180	954,000	0,000	0,229
Wilks-Lambda	0,043	2,136	180	826,725	0,000	0,295
Hotelling-Spur	6,438	3,442	180	866,000	0,000	0,417
Größte charakteristische Wurzel nach Roy	4,885	25,891	20	106,000	0,000	0,83

Nach den vier in SPSS durchgeführten Signifikanztests Pillai-Spur, Wilks-Lambda, Hotelling-Spur und der größten charakteristischen Wurzel nach Roy kann die Nullhypothese bei $\alpha = 0{,}05$ jeweils mit $p = 0{,}000$ verworfen werden (siehe Tabelle 3.7). Dieses eindeutige Ergebnis ist aufgrund der Formulierung der Nullhypothese und der Anzahl der Messwiederholungen bzw. Räume und der Anzahl der abgefragten Items nicht verwunderlich. Unterschiede zwischen den Teststatistiken werden vor allem durch die partiellen η^2-Werte deutlich. Pillai-Spur gilt als robusteste Teststatistik bezüglich Verletzungen der Annahmen (Bray und Maxwell, 1985, S. 28). Gemäß des aus dieser Teststatistik berechneten partiellen η^2 werden knapp 23 % der Gesamtvarianz durch den Innersubjektfaktor Raum erklärt. Der Test der größten charakteristischen Wurzel nach Roy errechnet hingegen ein partielles η^2 von 0,83, gilt allerdings auch als weniger robust gegen Verletzungen der Annahmen (vgl. O'Brien und Kaiser, 1985).

Um die Ergebnisse der globalen MANOVA über alle zwanzig abhängigen Variablen besser deuten zu können, wurde je Item eine ANOVA mit Messwiederholung durchgeführt. Dadurch konnte die Unterscheidung der Räume im Bezug auf die einzelnen Gegensatzpaare getestet werden. Eine zu erfüllende Annahme für ANOVAs mit Messwiederholungen ist die Sphärizität, bzw. Zirkularität der Daten. Sie „besagt, dass die Varianz der Differenzen der Messungen von jeweils zwei Treatmentstufen homogen sein muss“ (Bortz und Schuster, 2010, S. 477) und kann mithilfe des Mauchly-Tests je abhängiger Variable in SPSS getestet werden (siehe D.1). Ergibt sich ein signifikanter Effekt, kann die Annahme der Sphärizität nicht beibehalten werden. Eine Greenhouse-Geisser-Korrektur, welche die Freiheitsgrade des kritischen F-Werts bei Verstoß gegen die Sphärizitätsannahme nach unten korrigert (Eid et al., 2011, S. 463), erlaubt jedoch eine Interpretation der Ergebnisse.

Die Werte für p und das partielle η^2, welche sich aus den zwanzig ANOVAs ergaben, sind in Tabelle 3.9 aufgeführt. Im Falle der Items I4 (*Raum spricht früh an – Raum spricht spät an*), I6 (*gute Akustik – schlechte Akustik*), I9 (*dunkel – hell*), I17 (*Nachhall klingt weich aus – ... bricht ab*), I18 (*ausgewogen – unausgewogen*) und I20 (*Nachklang wird schärfer – ... wird dumpfer*) zeigen sich keine signifikanten Unterschiede zwischen den 10 verschiedenen Räumen ($\alpha = 0{,}05$). Die Nullhypothese, dass sich die Mittelwerte der Versuchspersonen über die Räume hinweg gleichen, kann daher bezüglich dieser abhängigen Variablen nicht verworfen

werden. Diese Ergebnisse kommen zustande, da die 14 Versuchspersonen die besagten Items sehr unterschiedlich bewertet haben, was bereits durch die großen Wertebereiche in den Boxplots deutlich wurde (siehe Abbildung 3.18). Dies wiederum kann darin begründet sein, dass wirkliche intraindividuelle Unterschiede der Wahrnehmung der Aspekte bestehen oder die Musiker keine Unterschiede zwischen den Räumen bezüglich der Aspekte wahr genommen haben, wodurch eine beliebige Bewertung der Items aufgetreten wäre. Bei der Bewertung der Präferenz (I6) sind beispielsweise intraindividuelle Unterschiede sehr gut vorstellbar. Obwohl den Versuchspersonen die Bedeutungen der Gegensatzpaare durch die zusätzlichen Beschreibungungen (siehe Tabelle 3.6) nahe gebracht wurden, könnten ferner die semantischen Begriffe unterschiedlich interpretiert und somit auch unterschiedlich bewertet worden sein. Eine ausführlich Diskussion der Ergebnisse folgt in Abschnitt 3.8.

Bei I1 (*tragfähig – wenig tragfähig*), I3 (*trocken – hallig*), I5 (*groß – klein*), I7 (*hoher Raum – niedriger Raum*) und I10 zeigt sich ein partielles $\eta^2 > 0{,}4$ und somit eine Varianzaufklärung durch den Messwiederholungsfaktor von mehr als 40 %. Die visuelle Analyse der Boxplots, bei welcher die Gegensatzpaare I3, I5, I7, I8 und I10 als am konsistentesten bewertet bezeichnet wurden, wird somit teilweise gestützt. Bei Anwendung von mehreren univariaten Varianzanalysen steigt das Risiko, Fehler 1. Art zu begehen, also die Nullhypothese fälschlicherweise zu verwerfen (Bortz und Schuster, 2010, S. 230). In unserem Fall ist die Wahrscheinlichkeit mindestens eines solchen Fehlers bei 20 univariaten Varianzanalysen und einem Signifikanznveau von 0,05 % bereits $P = 1 - (1 - \alpha')^{20} = 0{,}64$.

3.7.3. Untersuchung der Urteilerreliabilität bzw. -konsistenz

Um sich der in Abschnitt 3.1 aufgestellten Forschungsfrage 3 zu widmen, welche nach der Gestalt der interindividuellen Bewertungsunterschiede fragt, soll ein Reliabilitätsindex zur Quantfizierung dieser Unterschiede vorgestellt und die Werte pro Bewertungsitem angegeben und diskutiert werden. In Abschnitt 3.7.4 wird zudem ein Ansatz verfolgt, um die auf die Instrumentengruppen zurückzuführende Varianz bei der Berechnung der Reliabilität zu berücksichtigen.

Die Ähnlichkeit bzw. Konsistenz der Urteile ist auch daher von großem Interesse, weil die über die Versuchspersonen gebildeten Mittelwerte im Laufe der Auswertung die Grundlage statistischer Untersuchungen darstellen sollen. Eine geringe Reliabilität würde die Verwendung von Mittelwerten in Frage stellen und könnte darauf hindeuten, dass Items nicht gleich verstanden, unterschiedlich wahrgenommen wurden oder es aufgrund des Versuchsdesigns zu anderen Messfehlern kam. Die ersten beiden Punkte sprächen gegen eine weitere Verwendung der Items oder für eine intensive Schulung der Versuchspersonen, will man die Ergebnisse verallgemeinern und als Grundlage raumakustischer Planungen verwenden.

Ein Verfahren zur Berechnung der Reliabilität von intervallskalierten Ratingwerten stellt die Intraklassenkorrelation dar, das dazugehörige Maß ist der Intraklassenkorrelationskoef-

fizient[19] (vgl. Shrout und Fleiss, 1979). Um die Bewertungen einer Versuchsperson relativ zu den übrigen Werten dieser Person zu interpretieren, wurde der zweifaktorielle, justierte Reliabilitätsindex ICC(3,1) auf Basis der Rohdaten berechnet. Mittelwertsunterschiede zwischen den Versuchspersonen haben auf diesen Koeffizienten keinen Einfluss. Unterschiede zwischen den Ratern bezüglich ihres allgemeinen Bewertungsverhaltens – wie beispielsweise die Ausnutzung der Ratingskalen oder allgemein höhere oder niedrigere Bewertungen – wirken sich somit nicht auf den Index aus. Der ICC(3,1) ist allerdings nur dann als Reliabilitätsmaß im engeren Sinne aufzufassen, wenn die Stichprobe die einzig interessierende ist und nicht auf eine größere Population geschlossen werden soll, der Faktor Rater dementsprechend als fixed definiert wird. Im Random-Modell, bei zufälliger Auswahl der Versuchspersonen aus der interessierenden Gesamtpopulation, darf der Index nur als deskriptives Korrelationsmaß interpretiert werden (vgl. Wirtz und Caspar, 2002, S. 172–173). Die Gesamtpopulation besteht in unserem Fall aus professionellen Musikern der klassischen Musik. Ein unjustiertes Maß der Reliabilität stellt der ICC(2,1) dar, bei welchem die Rater als zufälliger Faktor angenommen werden und somit auf eine größere Population geschlossen werden kann (vgl. Wirtz und Caspar, 2002, S. 190). Bei der Wahl des geeigneten Reliabilitätsmaßes ist daher abzuwägen, ob die Justierung oder die Annahme zufälliger Stichproben von größerer Bedeutung ist. Wirtz und Caspar zufolge sollte bei der Wahl des ICC die Frage nach der Justierung Priorität besitzen.

> Würde die Frage nach der Justierung von der Eigenschaft des Faktors ‚Rater' als ‚fixed' oder ‚random' abhängig gemacht, so würde der vergleichsweise schwache Unterschied in der Interpretation des Koeffizienten als Korrelations- vs. Reliabilitätsmaß im Kontrast zum sehr bedeutsamen Unterschied ‚justiert' vs. ‚unjustiert' [...] zu stark gewichtet.
>
> (Wirtz und Caspar, 2002, S. 190–191)

Gemeinsam sind dem ICC(2,1) und dem ICC(3,1) die Vorraussetzungen, dass alle Rater alle Objekte bewertet haben müssen und die Objekte zufällig ausgewählt wurden, da es sich in beiden Fällen um ein zweifaktorielles varianzanalytisches Modell handelt. Beides wurde im vorliegenden Fall erfüllt, da die 10 bewerteten Räume als zufällige Auswahl von möglichen Konzertsälen betrachtet werden können. Zudem muss die Annahme erfüllt sein, dass keine Interaktion zwischen Ratern und Räumen vorliegt (Wirtz und Caspar, 2002, S. 186). Die Prüfung dieser Annahme wird auf Seite 70 vorgenommen.

Der Wertebereich der Indizes liegt per Definition zwischen 0 und 1. Die Intraklassenkorrelation beschreibt somit die „Stärke des Zusammenhangs zwischen den untersuchten Ratern" sowie im Rahmen der klassischen Testtheorie den „Anteil der Varianz der wahren Merkmalsausprägungen aufgrund der Beurteilungen durch einen Rater" (Wirtz und Caspar, 2002, S. 160). Dieser Anteil wird bei der Nutzung des ICC als Reliabilitätsmaß als Reliabilität der Ratings einer einzelnen Versuchsperson interpretiert (vgl. Shrout und Fleiss, 1979, S. 426). Werte kleiner 0 sind aufgrund negativer Korrelationen möglich, jedoch sind diese

[19] Im Folgenden mit ICC abgekürzt für *intraclass correlation coefficient*.

als Varianzaufklärung von 0 % zu interpretieren (Wirtz und Caspar, 2002, S. 161). Ein Vergleich der in dieser Arbeit vorgestellten ICCs kann anhand der Modellformeln (Wirtz und Caspar, 2002, S. 170) und Rechenvorschriften (S. 183, S. 186) in den Gleichungen 3.2 und 3.3 erfolgen.

$$\begin{aligned} \text{ICC(2,1)} = \rho_{zweifakt,unjustiert} &= \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2} = \frac{\sigma_\alpha^2}{\sigma_{Ges}^2} \\ &= \frac{MQ_{zw} - MQ_{res}}{MQ_{zw} + (k-1) \cdot MQ_{res} + \frac{k}{n} \cdot (MQ_{rat} - MQ_{res})} \end{aligned} \tag{3.2}$$

$$\begin{aligned} \text{ICC(3,1)} = \rho_{zweifakt,justiert} &= \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2} = \frac{\sigma_\alpha^2}{\sigma_{Ges}^2 - \sigma_\beta^2} \\ &= \frac{MQ_{zw} - MQ_{res}}{MQ_{zw} + (k-1) \cdot MQ_{res}} \end{aligned} \tag{3.3}$$

mit	σ_α^2	Varianz aus Variation der Objekte (hier: Räume)
	σ_β^2	Varianz aus Unterschieden zwischen Ratern
	σ_ε^2	Fehlervarianz
	σ_{Ges}^2	Gesamtvarianz
	MQ_{zw}	Mittlere Quadratsumme aus Varianzquelle Objekt (hier: Raum)
	MQ_{rat}	Mittlere Quadratsumme aus Varianzquelle Rater
	MQ_{res}	Mittlere Quadratsumme aus Residualvarianz
	k	Anzahl der Rater/ Versuchspersonen
	n	Anzahl der Objekte (hier: Räume)

Die Justierung des ICC(3,1) wird durch den Ausschluss der raterspezifischen Varianz σ_β^2 aus der Gesamt- bzw. Nennervarianz σ_{Ges}^2 erreicht. Dies hat zur Folge, dass der ICC(3,1) relativ zum ICC(2,1) umso größer wird, je größer die Mittelwertsunterschiede zwischen den Ratern ausfallen. Der ICC(3,1) stellt somit ein Maß für die Raterkonsistenz dar, „da für die einzelnen beurteilten [...] [Objekte[20]] lediglich eine ähnliche große Abweichung der Ratingwerte von den raterspezifischen Mittelwerten gefordert wird, ohne dass die absoluten Werte ähnlich sein müssen“ (Wirtz und Caspar, 2002, S. 171). Der ICC(2,1) stellt hingegen ein Ähnlichkeitsmaß dar. Für diese Studie wurde der ICC(3,1) als Prädiktor für die Urteilerkonsistenz gewählt, der ICC(2,1) wurde zum Vergleich berechnet und angegeben.

Mithilfe einer zweifaktoriellen Varianzanalyse kann die Wahl des gewählten Index überprüft werden. Zeigen sich hinsichtlich des Faktors Rater signifikante Unterschiede, wird die Entscheidung der Verwendung des justierten Maßes bekräftigt. Der justierte Index wird folglich deutlich höhere Werte als der unjustierte ergeben (vgl. Wirtz und Caspar, 2002,

[20] Bei Wirtz und Caspar (2002) handelt es sich bei den zu bewertenden Objekten um Personen. Diese sind nicht als Versuchspersonen zu betrachten. Die Versuchspersonen dieser Studie sind im Sinne von Wirtz und Caspar als Rater zu verstehen.

S. 187). Die Ergebnisse der in Matlab durchgeführten zweifaktoriellen ANOVAs zeigen bei den Items I1–I3, I4, I7, I8, I10–I14, I16 und I17 signifikante Unterschiede bezüglich des Faktors Rater (siehe Tabelle D.3), die Wahl des ICC(3,1) erscheint somit in diesen Fällen gerechtfertigt. Bei den meisten der nicht signifikanten Fälle (I4, I6, I9, I15, I20) handelt es sich desweiteren um Items, bei welchen sich auch bezüglich des Faktors Raum keine signifikanten Unterschiede zeigen. Der Faktor Raum der zweifaktoriellen ANOVA ist gleichzusetzen mit dem Messwiederholungsfaktor der einfaktoriellen ANOVA mit Messwiederholung. Aus diesem Grund sind die Ergebnisse des Faktors Raum identisch mit den Werten der Rohdaten-ANOVAs in Tabelle 3.9, welche zusätzlich gerechnet wurden. Der Rater-Effekt kann zum Teil vom Raumeffekt abhängen, da sich der F-Wert folgendermaßen berechnet: $\mathrm{F} = \frac{\mathrm{MQ}_{rater}}{\mathrm{MQ}_e}$. Bei geringen Raum-Effekten fällt die Fehlervarianz größer aus, wodurch der Rater-Effekt zwangsweise kleiner wird. Folglich wäre der Rater-Effekt bei größeren Mittelwertsunterschieden zwischen den Räumen vermutlich auch bei den entsprechenden Items signifikant ausgefallen. Aus diesem Grund und zur besseren Vergleichbarkeit der ICC-Werte zwischen den Items wird in allen Fällen der ICC(3,1) als der entscheidende gewählt. Neben der Begründung der Wahl des justierten ICC lässt sich aus den Betrachtungen des Faktors Rater auch die Notwendigkeit der z-Transformation bei weiteren Auswertungsmethoden erkennen, bei welchen keine Justierung im Modell selbst vorgenommen wird. Aufgrund der gegebenen Justierung im Fall des ICC(3,1) wurden alle ICC-Werte anhand der Rohdaten berechnet, da die Anwendung der z-Transformation in diesem Fall hinfällig wird.

Die Berechnung der ICC-Werte wurde in Matlab mithilfe des Skripts *ICC* von Arash Salarian[21] durchgeführt. Beide vorgestellten ICC-Maße sind in Tabelle 3.9 je Item aufgetragen. Die ICC-Werte beschreiben, wie bereits geschildert, den Anteil der aufgeklärten Varianz, welche auf die Variation der Raumsimulationen (σ_α^2) zurückzuführen ist. Der Restbetrag $1 - \mathrm{ICC}$ wird in den Modellen der Fehlervarianz zugesprochen. Eine konsistente, allgemeingültige Einstufung von Wertebereichen der Koeffizienten ist nicht vorhanden, da es neben den hier vorgestellten Indizes weitere Versionen gibt, deren Werte sich erheblich unterscheiden können. Zudem ist die Höhe des ICC auch immer von der Varibilität der zu bewertenden Objekte abhängig (vgl. Weir, 2005, S. 236–237). „Die Reliabilität wird im Allgemeinen um so höher ausfallen, je größer die Unterschiede zwischen den Mittelwerten der [bewerteten Objekte] sind“ (Wirtz und Caspar, 2002, S. 161). Für diese Studie bedeutet dies, dass eine höhere Variabilität der Raumbewertungen, ausgehend von größeren Unterschieden zwischen den Raumsimulationen, in höheren ICC-Werten resultiert.

Bei der Betrachtung der ICCs zeigt sich allgemein ein recht geringes Niveau der Werte. Die höchste Konsistenz der Urteile ist bei Item I3 (trocken – hallig) mit einem ICC(3,1) von 0,694 zu beobachten. Beide Begriffe des Gegensatzpaares wurden, wenn auch nicht immer in Kombination, bereits im RGT-Versuch häufig von den Versuchspersonen genannt. Neben der Ähnlichkeit der verwendeten semantischen Begriffe wird durch den im Vergleich

[21] http://www.mathworks.com/matlabcentral/fileexchange/22099; letzter Zugriff: 20. Mai 2013

zu den anderen Items hohen ICC die inhaltliche Konsistenz des Begriffspaares bestätigt. Die Bewertung der allgemeinen Raumgröße (I5, ICC$(3,1) = 0{,}658$) sowie der Raumhöhe (I7, ICC$(3,1) = 0{,}611$) wurde von den Versuchspersonen auch relativ konsistent bewertet, bei Item I8 (*breiter Raum – schmaler Raum*, ICC$(3,1) = 0{,}341$) hingegen zeigt sich ein geringerer ICC. Zu den konsistenter bewerteten Items zählen auch die Balance zwischen Raumklang und Instrument (I10, ICC$(3,1) = 0{,}565$) sowie die Tragfähigkeit des Spiels durch den Raum (I1, ICC$(3,1) = 0{,}472$). Bei der in vorherigen Studien besonders hervorgehobenen Unterstützung (I11) errechnet sich in dieser Studie lediglich ein ICC(3,1) von 0,108.

Im Zuge der ICC-Berechnung wurde die Nullhypothese ICC = 0 geprüft, welche bei den Items I4 (*Raum spricht früh an – ... spät an*), I6 (*gute Akustik – schlechte Akustik*), I9 (*dunkel, – hell*), I15 (*weich – hart*), I16 (*rund – metallisch*) und I20 (*Nachklang wird schärfer – ... dumpfer*) bei einem Signikanzniveau von $\alpha = 0{,}05$ nicht widerlegt werden konnte (siehe Tabelle D.1).[22] Es fällt auf, dass sich unter diesen bis auf zwei Ausnahmen (I4, I19) alle Items befinden, welche nach semantischen Gesichtspunkten der Beschreibung klangfarblicher Aspekte zuzuordnen sind. Item I14 (*dünn – voll*) wurde innerhalb der Stichprobe zumindest mit einer Konsistenz von ICC(3,1) = 0,261 bewertet und ist somit bezüglich des ICCs mit Item I12 (*laut – leise*) vergleichbar. I19, welches eine etwaige Veränderung der Klangfarbe im Laufe des Nachhalls abfragt, erreicht immerhin einen von 0 signifikant unterschiedlichen Wert von 0,129. Neben den vier klangfarblichen Items I9, I15, I16 und I20 besteht keine Konsistenz bezüglich der akustischen Präferenz (I6) und der Schnelligkeit der Raumantwort (I4). Wie in Tabelle 3.9 zu erkennen, decken sich die Ergebnisse der ANOVAs mit Messwiederholung unter Verwendung der Rohdaten mit den ICC-Berechnungen inklusive Signifikanztests. Diese Beobachtung verwundert nicht, da die ICC-Berechnung auf der gleichen Quadratsummenzerlegung wie die Varianzanalyse mit Messwiederholungen beruht.

Dementsprechend folgt ein Erklärungversuch für die geringe Konsistenz der Items den bereits in Abschnitt 3.7.2 bei der Auswertung der ANOVAs beschriebenen Überlegungen. Die geringen Mittelwertsunterschiede, welche in einem geringen ICC resultieren, könnten zum einen aufgrund geringer Unterschiede zwischen den Räumen betreffend der jeweiligen Beurteilungsaspekte zustande gekommen sein. Durch die Zusammenstellung der 10 Raumsimulationen anhand einer möglichst großen Varianz der raumakustischen Parameter sollte eine Unterscheidbarkeit aller wesentlichen Bewertungsmerkmale durch das Versuchsdesign ermöglicht werden. Da die Items außerdem im RGT-Versuch aufgrund von Vergleichen derselben Raumsimulationen eliziert worden sind, ist anzunehmen, dass die Räume bezüglich dieser Items auch Unterschiede aufweisen – zumindest für die entsprechende Versuchsperson, welche das Gegensatzpaar nannte. Ein weiterer möglicher Grund für einen geringen ICC ist eine unterschiedliche Interpretation der semantischen Begriffe durch die Versuchspersonen. Im Fall der Präferenz (I6) ist dies leicht nachzuvollziehen, da die Bedeutung des Begriffs *gute Akustik* von Musiker zu Musiker höchswahrscheinlich variieren wird. Auch

[22] Im Fall der Nullhypothese ICC = 0 sind die Ergebnisse der Hypothesentests für ICC(2,1) und ICC(3,1) identisch.

bei den Items I9 (*dunkel – hell*), I15 (*weich – hart*) und I16 (*rund – metallisch*) ist solch eine subjektive Interpretation möglich. Gleichzeitig ist aufgrund bisheriger Studien der Bühnenakustik anzunehmen, dass die gespielten Instrumente, mit welchen die Räume angeregt wurden, einen Einfluss auf die Bewertung der Gegensatzpaare hatten. Die beobachtete inkonsistente Bewertung von Items, welche sich auf die Klangfarbe (I9, I16, I20, I15) sowie die Schnelligkeit der Raumantwort (I4) beziehen, könnte aufgrund der Unterschiede im Tonumfang, dem Obertongehalt sowie des Hüllkurvenverlaufs der verschiedenen Instrumente aufgetreten sein. Die Prüfung systematischer Einflüsse der Instrumente auf die Bewertung der Raumakustik durch Musiker soll im nächsten Abschnitt behandelt werden.

Tabelle 3.8.: Korrigierte Trennschärfen der Musikerurteile je Item

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	σ_x
I1	0,11	0,02	0,82	0,81	0,78	0,47	0,82	0,8	0,78	0,66	0,94	0,67	0,92	0,74	0,28
I2	0,62	0,32	−0,37	0,07	−0,17	0,04	0,94	0,38	0,47	0,51	−0,46	0,52	0,52	0,7	0,38
I3	0,86	0,36	0,96	0,85	0,94	0,92	0,85	0,88	0,95	0,75	0,89	0,71	0,87	0,91	0,21
I4	−0,23	−0,18	−0,66	0	0,01	−0,47	0,07	0,08	0,84	0,01	0,24	0,31	−0,61	0,1	0,38
I5	0,83	0,46	0,78	0,86	0,88	0,94	0,83	0,84	0,77	0,59	0,9	0,85	0,88	0,96	0,2
I6	0,14	0,52	0,07	−0,06	−0,06	−0,01	0,52	−0,09	−0,54	0,18	0,28	−0,07	0,01	0,04	0,25
I7	0,93	0,68	0,87	0,84	0,79	0,86	0,9	0,91	0,73	0,76	0,85	0,4	0,8	0,85	0,2
I8	0,44	0,36	0,73	0,38	0,81	0,77	−0,12	0,75	0,63	0,51	0,8	0,44	0,66	0,68	0,25
I9	−0,55	−0,11	−0,13	0,09	−0,05	0,15	0,25	0,28	0,13	0,35	0,17	−0,15	0,1	0,07	0,22
I10	0,28	0,47	0,9	0,84	0,71	0,92	0,86	0,78	0,42	0,86	0,84	0,59	0,91	0,93	0,23
I11	−0,28	0	0,2	0,27	0,72	0,18	0,58	−0,02	0,18	0,62	0,22	0,43	0,47	0,32	0,25
I12	−0,04	0,47	0,14	0,68	−0,42	0,82	0,72	0,69	0,33	0,4	0,67	0,82	0,78	0,67	0,34
I13	0,47	0,23	−0,2	−0,19	0,7	0,75	0,93	0,86	0,52	0,64	0,86	0,36	0,03	0,88	0,36
I14	0,69	−0,07	0,05	−0,17	0,58	0,48	0,48	0,78	0,6	0,58	0,44	0,75	0,77	0,66	0,29
I15	0,33	0,41	0,65	0,06	0,36	0,14	0,27	0,57	0,59	0,32	0,03	0,23	−0,45	−0,2	0,29
I16	−0,4	0,6	0,4	−0,11	0,02	−0,04	−0,35	−0,19	−0,13	0,43	0,15	−0,23	−0,05	0,04	0,28
I17	0,49	−0,06	0,06	−0,02	−0,04	0,07	−0,27	0,51	0,56	0,77	−0,11	0,23	0,62	0,41	0,3
I18	−0,24	0,82	0,63	0,17	−0,14	0,05	0,61	0,71	−0,26	0,28	0,27	−0,33	0,04	0,03	0,36
I19	−0,58	0,56	−0,24	0,26	0,71	0,04	0,7	0,63	0,29	0,32	0,42	0,5	0,25	0,4	0,34
I20	0,3	0,48	−0,34	−0,44	0,16	0,15	0,26	0,09	−0,23	0,21	−0,04	0,18	−0,01	0,21	0,24
$\bar{x}$	0,21	0,32	0,27	0,26	0,36	0,36	0,49	0,51	0,38	0,49	0,42	0,36	0,38	0,47	0,27

Abschließend bleibt noch die Annahme zu überprüfen, dass keine Interaktion zwischen den Urteilern und den Räumen vorhanden war. Besteht eine solche Interaktion, wird also die Annahme $\sigma^2_{\alpha\times\beta} = 0$ verletzt, beinhaltet die Resudialvarianz neben der Fehlervarianz ebenso die Interaktionsvarianz. Die durch MQ_{res} geschätzte Fehlervarianz wird somit überschätzt und die ICC-Werte fallen durch die Erhöhung der Nennerwerte systematisch kleiner aus (Wirtz und Caspar, 2002, S. 186). Wirtz und Caspar (2002, S. 213) empfehlen die Durchführung von Tukey-Tests zur Überprüfung einer solchen möglichen Interkation zwischen Urteilern und Objekten. Im erhobenen Datensatz zeigt sich gemäß Tukeys Additivitätstests bei den Items I3 (*trocken – hallig*), I4 (*Raum spricht früh an – ... spät an*), I6 (*gute Akustik – schlechte Akustik*), I8 (*breiter Raum – schmaler Raum*), I9 (*dunkel – hell*), I11 (*unterstützend – nicht unterstützend*), I16 (*rund – metallisch*) und I20 (*Nachklang wird schärfer – ... dumpfer*) eine signifikante Interaktion zwischen Versuchspersonen und Räumen ($\alpha = 0{,}25$). Im Gegensatz zu Wirtz und Caspar (2002) schätzen (Asendorpf und Wallbott, 1979, zit. in Wirtz und Caspar, 2002) eine vorhandene Interaktion als grundsätzlich unproblematisch ein, da dies

nur einer Unterschätzung der wahren Reliabilität resultiere.

Da der ICC, sofern als Reliabilitätsmaß interpretiert, als Korrelationsmaß der Werte eines Raters mit denen eines beliebigen anderen Raters der Stichprobe aufgefasst wird, sollte ferner die Homogenität der Korrelationen zwischen den einzelnen Raterpaaren betrachtet werden. Ein Maß, welches sich zur Überprüfung der Homogenität der Raterurteile eignet, ist die korrigierte Trennschärfe. Diese wird durch die Korrelation der Urteile eines Raters mit den Mittelwerten aller übrigen Rater beschrieben (Wirtz und Caspar, 2002, S. 210). Im vorliegenden Fall ergibt sich somit je Versuchsperson und Bewertungsitem ein Wert für die korrigierte Trennschärfe (siehe Tabelle 3.8). Ist die Homogenität der korrigierten Trennschärfen nicht gegeben, stellt auch dies ein Indiz für eine vorhandene Interaktion zwischen Ratern und bewerteten Objekten dar.

Da die korrigierten Trennschärfen über alle Rater bei keinem der Items als homogen interpretiert werden können, sollte der ICC laut Wirtz und Caspar (2002, S. 212) nicht als Reliabilitätsmaß verwendet werden. Anhand der korrigierten Trennschärfen bei Item I3 (*trocken – hallig*) soll diese Problematik veranschaulicht werden. Ein für die gesamte Stichprobe angegebenes Reliabilitätsmaß würde eine Aussage über die Korrelation zwischen zwei beliebigen Ratern der Stichprobe angeben. M2 jedoch weist bei Item I3 mit 0,36 eine deutlich geringere korrigierte Trennschärfe als die übrigen Musiker auf, wodurch ein für dieses Item angegebender ICC nicht der Definition eines Reliabilitätsmaßes entspricht. Die Unterschiede der Trennschärfen bei den übrigen Items fallen bisweilen noch deutlich größer aus. Der Ausschluss von Ratern mit geringer korrigierter Trennschärfe empfiehlt sich aus Sicht des Autors aufgrund der kleinen Stichprobe nicht, da sich die Trennschärfen auch nicht über alle Items gleich verhalten. Die in dieser Studie ermittelten ICCs dürfen somit nicht als Reliabilitätsmaße im engeren Sinne aufgefasst werden. Da nicht im Besonderen die Reliabilität, sondern die Unterschiede zwischen den Bewertungen von Interesse waren, kann der ICC als ein Maß für die Homogenität[23] oder die Konsistenz der Musikerbewertungen dienen.

[23] McGraw und Wong (1996, S. 30) verwenden diesen allgemeineren Begriff im Zusammenhang mit der Intraklassenkorrelation.

Tabelle 3.9.: Intraklassenkorrelationskoeffizienten und ANOVA-Ergebnisse der 20 Items. Im Falle der ICC-Werte sind solche in rot dargestellt, bei welchen ein zusätzlicher Signifikanztest mit $H_0 : ICC = 0$ bei $\alpha = 0,05$ nicht verworfen werden konnte. Im Fall der ANOVAs sind die p-Werte > 0,05 rot eingefärbt.

		Intraklassenkorrelation		ANOVAs Rohdaten		ANOVAs z-scores	
	Gegensatzpaar	ICC(3,1)	ICC(2,1)	Sig./p	part. η^2	Sig./p	part. η^2
I1	tragfähig – wenig tragfähig	0,472	0,446	0,000	0,510	0,000	0,469
I2	ich kann mich gut hören – ... nicht gut hören	0,164	0,134	0,010[g]	0,224[g]	0,014	0,159
I3	trocken – hallig	0,694	0,632	0,000	0,716	0,000	0,737
I4	Raum spricht früh an – Raum spricht spät an	−0,024	−0,024	0,732	0,049	0,782	0,045
I5	groß – klein	0,658	0,638	0,000[g]	0,682[g]	0,000[g]	0,702[g]
I6	gute Akustik – schlechte Akustik	0,016	0,017	0,283	0,087	0,280	0,087
I7	hoher Raum – niedriger Raum	0,611	0,573	0,000	0,639	0,000	0,688
I8	breiter Raum – schmaler Raum	0,341	0,323	0,000	0,388	0,000	0,395
I9	dunkel – hell	0,011	0,010	0,340[g]	0,082[g]	0,307	0,084
I10	viel Raum – wenig Raum	0,565	0,515	0,000[g]	0,596[g]	0,000[g]	0,593[g]
I11	unterstützend – nicht unterstützend	0,108	0,099	0,037[g]	0,172[g]	0,004	0,182
I12	laut – leise	0,266	0,220	0,000[g]	0,318[g]	0,000	0,312
I13	diffus – klar	0,312	0,268	0,000[g]	0,361[g]	0,000	0,304
I14	dünn – voll	0,261	0,223	0,000	0,313	0,000	0,306
I15	weich – hart	0,062	0,061	0,116[g]	0,129[g]	0,009	0,166
I16	rund – metallisch	0,001	0,001	0,410[g]	0,073[g]	0,380	0,077
I17	Nachhall klingt weich aus – ... bricht ab	0,117	0,107	0,031[g]	0,180[g]	0,065[g]	0,148[g]
I18	ausgewogen – unausgewogen	0,068	0,069	0,043	0,135	0,066	0,125
I19	Klangfarbe ist gleichmäßig – ... fluktuiert	0,129	0,126	0,002	0,191	0,003	0,187
I20	Nachklang wird schärfer – ... wird dumpfer	0,018	0,017	0,272	0,088	0,212	0,095

g: Greenhouse-Geisser-korrigierte Werte, da Mauchly-Test keine Sphärizität feststellen konnte.

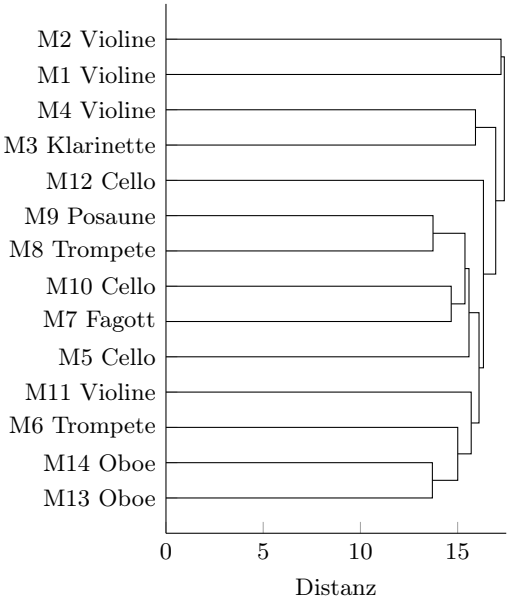
3.7.4. Untersuchung des Instrumenteneinflusses

Um im ersten Schritt die Musiker bezüglich ihrer Raumbewertungen zu klassifizieren, wurde ein clusteranalytischer Ansatz gewählt. Ergeben sich Cluster von Musikern gleicher Instrumente, kann ein Zusammenhang zwischen Instrument und Bewertung hergestellt werden. Ergeben sich keine Cluster, welche auf die gespielten Instrumente zurückzuführen sind, ist ein instrumentenspezifischer Einfluss jedoch keineswegs ausgeschlossen. Es wird lediglich gezeigt, dass der Einfluss der Instrumente nicht derart ausgeprägt ist, um andere Einflüsse bzw. Faktoren zu überdecken, da die Ähnlichkeit der Urteile die Grundlage bei der Ermittlung der Distanzen darstellt. Da sich bei der Untersuchung der Urteilerkonsistenz manche Items als problematisch erwiesen, wurden drei Clusteranalysen mit unterschiedlicher Einbeziehung der Items durchgeführt:

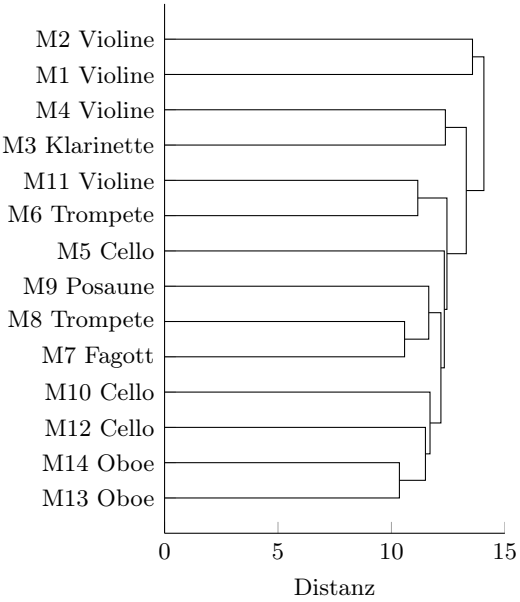
1. Inklusive der Bewertungen aller Items.
2. Ohne Items, bei welchen die Hypothese ICC = 0 nicht widerlegt werden konnte.
3. Inklusive Items, bei welchen ICC(3,1) > 0,4 gilt.

Bei den drei in Matlab durchgeführten hierarchischen Clusteranalysen wurden euklidische Distanzen ermittelt und das UPGMA-Verfahren zur Fusionierung verwendet, da sich bei dieser Kombination die höchsten Werte für den kophenetischen Korrelationskoeffizienten[24] ergaben (vgl. Handl, 2002, S. 366–380). In Abbildung 3.19 sind die ermittelten Dendrogramme zu entnehmen. Bei den ersten beiden Clusteranalysen zeigen sich im Allgemeinen sehr große Distanzen hinsichtlich der ersten Fusionsstufe. Eine Einteilung der Musiker in Cluster ist aus diesem Grund nicht sinnvoll. Zwar ergeben sich in der Clusteranalyse über alle Items zwischen M13 und M14 (jeweils Oboe) sowie zwischen M8 und M9 (Posaune und Trompete) die geringsten Distanzen, im Fall der Violinen lassen sich jedoch die größten Distanzen erkennen. Bei Ausschluss der Items I4, I6, I9, I16 und I20 ergibt sich ein ähnliches Bild mit etwas geringeren Distanzen. Bei der dritten Clusteranalyse anhand der am konsistentesten bewerteten Items I1, I3, I5, I7 und I10 sind die als erstes fusionierten Versuchspersonen in keinem Fall Musiker des gleichen Instruments. Zwar erweisen sich die Bewertungen von M3 und M7 (jeweils Holzbläser) als am ähnlichsten, allerdings lassen sich über das gesamte Dendrogramm hinweg keine klaren Cluster bezüglich der Instrumentengruppen von Streichern (Violine, Cello), Blechbläsern (Trompete, Posaune) und Holzbläsern (Oboe, Fagott, Klarinette) erkennen. Die clusteranalytische Analyse weist somit darauf hin, dass der Einfluss der gespielten Instrumente auf die Bewertung der Raumakustik nicht von übergeordneter Natur ist.

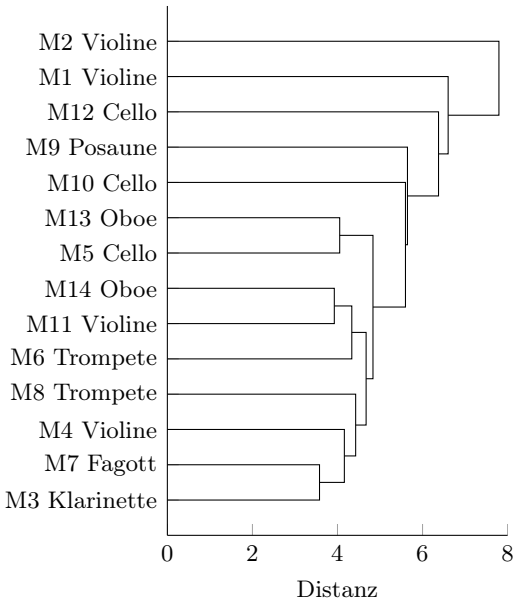
[24] Die dem Dendrogramm zugrundeliegende Distanzmatrix, welche mithilfe des gewählten Fusionierungsalgorithmus gewonnen wurde, wird als kophenetische Matrix bezeichnet. Die linerare Korrelation zwischen dieser und der Distanzmatrix im nullten Iterationsschritt bezeichnet man als kophenetischen Korrelationskoeffizienten. Dieser stellt ein Maß für die Güte einer Clusterlösung dar. Bei einem Vergleich unterschiedlicher Clusterverfahren, sollte dasjenige gewählt werden, bei welchem sich die höchste kophenetische Korrelation ergibt (vgl. Handl, 2002, S. 380).



(a) Inklusive aller Items, kophenetische Korrelation $r = 0{,}76$.



(b) Ohne Items I4, I6, I9, I16, I20, kophenetische Korrelation $r = 0{,}80$.



(c) Nur mit Items I1, I3, I5, I7, I10, kophenetische Korrelation $r = 0{,}93$.

Abbildung 3.19.: Clusteranalysen über Musiker mit unterschiedlicher Einbeziehung der Itembewertungen

Eine Methode, welche die auf die Instrumente oder die Instrumentengruppen zurückzuführende Varianz im Bezug auf alle Items bestimmen könnte, stellt eine MANOVA mit Innersubjektfaktor Raum und Zwischensubjektfaktor Instrument dar. Im vorliegenden Fall ist eine Durchführung dieser Methode jedoch aufgrund ungenühender Residuen-Freiheitsgrade nicht möglich.

Eine weitere Möglichkeit, den Einfluss der Instrumente auf die Bewertungen der Raumakustik statistisch zu prüfen, ist die Bildung eines hierarchischen linearen Modells (oder Mehrebenenmodells). Dabei ist nicht – wie bisher angenommen – eine zufällige Auswahl der Versuchspersonen aus der interessierenden Gesamtpopulation, sondern eine sogenannte mehrstufige Stichprobenziehung gegeben (vgl. Eid et al., 2011, S. 700). Im vorliegenden Fall lässt sich ein dreistufiges Modell mit folgenden Ebenen bilden, in welchem neben den Ratern und den bewerteten Räumen auch die Zugehörigkeit der Rater zu Instrumentengruppen berücksichtigt wird.

Ebene 3: Instrument (Instrumentengruppe)
Ebene 2: Rater (Versuchsperson)
Ebene 1: Raum (Messwiederholung)

Aufgrund der in Abschnitt 3.3.2 angesprochenen Unterschiede der Instrumente und der Ergebnisse bisheriger Studien zur Bühnenakustik, sind auch systematische Unterschiede zwischen den Instrumentengruppen bei der Bewertung der Raumakustik anzunehmen, weshalb die Anwendung des Mehrebenenmodells gerechtfertigt ist. Die Messwiederholung bzw. der Innersubjektfaktor wird in einem Mehrebenenmodell als unterste Ebene bzw. Ebene 1 verstanden (vgl. Hox, 2010, S. 33). Durch Berechnung der Varianzen innerhalb eines Intercept-Only-Modells, in welchem die Varianzen nicht durch unabhängige Variablen erklärt werden, sondern lediglich eine Varianzzerlegung bezüglich der Ebenen vorgenommen wird (S. Hox, 2010, 15), kann die Intraklassenkorrelation bestimmt werden. In einem Zweiebenenmodell ohne die Instrumentenebene wird der in Abschnitt 3.7.3 vorgestellte und berechnete ICC(3,1), welcher die Urteilerkonsistenz der Versuchspersonen berschreibt, durch folgende Modellgleichung bestimmt (vgl. Eid et al., 2011, S. 717). Sie entspricht der ersten Zeile in Gleichung 3.3.

$$\text{ICC}(3,1) = \frac{\sigma^2_{Raum}}{\sigma^2_{Raum} + \sigma^2_{Res}} \tag{3.4}$$

Durch Hinzunahme der Instrumentenebene lässt sich nun die auf dieser Ebene systematisch auftretende Varianz in die Berechnung der Intraklassenkorrelation miteinbeziehen. Systematische Unterschiede zwischen den verschiedenen Instrumentengruppen wirken sich somit nicht auf den ICC aus, da die Varianz des Intercepts von Ebene 3 aus der Residualvarianz ausgeschlossen werden kann. Diese Berechnungsmethode der Intraklassenkorrelation beschreibt die Korrelation der Bewertungen zweier Rater der gleichen Instrumentengruppe (vgl. Hox, 2010, S. 34). In der Summe der Varianzen von Raum- und Instrumentenebene in Modellformel 3.5 ist die Interaktion zwischen diesen beiden Ebenen miteingeschlossen, sie entspricht somit der Summe der Intercept-Varianzen von $Raum$, $Instrument$ und $Raum \times Instrument$.

$$\mathrm{ICC}(3{,}1)_{Ins} = \frac{\sigma^2_{Raum} + \sigma^2_{Ins}}{\sigma^2_{Res} + \sigma^2_{Raum} + \sigma^2_{Ins}} \tag{3.5}$$

Der unjustierte Koeffizient $\mathrm{ICC}(2{,}1)_{Ins}$ könnte zusätzlich berechnet werden, indem im Nenner die auf Raterebene auftretende Intercept-Varianz aufsummiert wird und somit durch die Gesamtvarianz geteilt wird. Die Berechnungen der Intercept-Only-Modelle wurden in SPSS mithilfe der Funktion *Mixed* durchgeführt. Zur Schätzung der Varianzkomponenten wurde die eingeschränkte Maximum Likelihood Methode (REML) verwendet, da diese Methode im Vergleich zur normalen Maximum Likelohood Methode einen erwartungstreueren Schätzer darstellt (Kim und Timm, 2007, S. 116), besonders im Fall kleiner Stichproben (Hox, 2010, S. 41). Die Instrumente wurden aufgrund der kleinen Stichprobe nicht einzeln kodiert, sondern zu den Instrumentegruppen Streicher, Blechbläser und Holzbläser zusammengefasst. Da die Personenunterschiede durch das Mehrebenenmodell wie bei der Berechnung der Intraklassenkorrelation berücksichtigt werden konnten, flossen die nicht standardisierten Rohdaten in die Berechnung ein.

In Tabelle 3.10 sind $\mathrm{ICC}(3{,}1)_{Ins}$ sowie ICC(3,1) für die Items I1, I3, I7, I10, I12, I14 und I19 angegeben. Für die übrigen Items konnten die Varianzkomponenten im Intercept-Only-Modell aufgrund einer nicht positiv definiten Hesse-Matrix nicht zuverlässig bestimmt werden. Aufgrund zu geringer spezifischer Varianz auf mindestens einer Ebene ergeben sich in SPSS somit Interceptvarianzen des Wertes 0. In solchen Fällen konnte daher kein zuverlässiger $\mathrm{ICC}(3{,}1)_{Ins}$ berechnet werden. Die Werte zeigen jeweils eine Erhöhung der Intraklassenkorrelation durch die Berücksichtigung der Instrumentengruppen. Aufgrund der zusätzlich berücksichtigten Varianzquelle der Instrumentenebene ist diese Erhöhung unumgänglich. Von größerem Interesse als die absoluten Werte sind die Differenzen zwischen $\mathrm{ICC}(3{,}1)_{Ins}$ und ICC(3,1), durch welche ein Vergleich des Instrumentengruppeneinflusses bei den verschiedenen Items erfolgen kann.

Der größte Differenzbetrag zeigt sich bei Item I14 (*dünn – voll*), der größte prozentuale Anstieg bei I19 (*Klangfarbe ist gleichmäßig – … fluktuiert*). Beide Items könnten anhand semantischer Gesichtspunkte der Bewertung klangfarblicher Aspekte zugeordent werden. Im Fall von I14 wird sich in der weiteren Auswertung jedoch zeigen, dass es von den Versuchspersonen im Mittel ähnlich interpretiert wurde wie Bewertungsitems zur Raumgröße und Halligkeit (siehe Abschnitt 3.7.5). Bei Items I3 (*trocken – hallig*) kann anhand der ICC-Werte der geringste Einfluss der Instrumentengruppen angenommen werden kann, sowohl die prozentuale als auch die absolute Differenz fallen bei diesem Item am geringsten aus.

Bei der relativ kleinen Stichprobengröße von 14 Musikern zeigte sich zwar bei keinem der Items ein signifikanter Einfluss des Instrumentengruppen-Intercepts, jedoch lassen die Ergebnisse des korrigierten $\mathrm{ICC}(3{,}1)_{Ins}$ im Vergleich zum ICC(3,1) einen Einfluss der Instrumentengruppen auf die Bewertung raumakustischer Umgebungen vermuten. Um diesen nachzuweisen bedarf es allerdings einer Untersuchung mit größerer Stichprobe. Eine weitere

Möglichkeit besteht darin, sich auf einzelne Instrumentengruppen zu konzentrieren.

Tabelle 3.10.: Um systematische Instrumentenvarianz bereinigte Intraklassenkorrelation $ICC(3,1)_{Instr}$ im Vergleich zu ICC(3,1)

	Gegensatzpaar	ICC(3,1)	$ICC(3,1)_{Ins}$	Differenz	Diff. in %
I1	tragfähig – wenig tragfähig	0,472	0,503	0,031	6,24
I2	ich kann mich gut hören – ... nicht gut hören	0,164	—[h]		
I3	trocken – hallig	0,694	0,721	0,027	3,75
I4	Raum spricht früh an – Raum spricht spät an	−0,024	—[h]		
I5	groß – klein	0,658	—[h]		
I6	gute Akustik – schlechte Akustik	0,016	—[h]		
I7	hoher Raum – niedriger Raum	0,611	0,665	0,054	8,13
I8	breiter Raum – schmaler Raum	0,341	—[h]		
I9	dunkel – hell	0,011	—[h]		
I10	viel Raum – wenig Raum	0,565	0,609	0,045	7,34
I11	unterstützend – nicht unterstützend	0,108	—[h]		
I12	laut – leise	0,266	0,290	0,025	8,54
I13	diffus – klar	0,312	—[h]		
I14	dünn – voll	0,261	0,333	0,072	21,61
I15	weich – hart	0,062	—[h]		
I16	rund – metallisch	0,002	—[h]		
I17	Nachhall klingt weich aus – ... bricht ab	0,117	—[h]		
I18	ausgewogen – unausgewogen	0,068	—[h]		
I19	Klangfarbe ist gleichmäßig – ... fluktuiert	0,129	0,198	0,068	34,63
I20	Nachklang wird schärfer – ... wird dumpfer	0,018	—[h]		

h: Endgültige Hesse-Matrix ist nicht positiv definit, obwohl Konvergenzkriterien erfüllt sind.

Bevor im nächsten Abschnitt die Zusammenhänge zwischen den Itembewertungen untersucht werden, bleibt festzuhalten, dass die Konsistenz der Urteile über die zwanzig durch den RGT-Versuch elizierten Items sehr stark variiert. Die wahrgenommene Halligkeit (I3) wurde mit einem ICC(3,1) = 0,694 von den 14 Versuchspersonen am konsistentesten bewertet. Nachdem die semantischen Begriffe *trocken* und *hallig* im RGT-Versuch am häufigsten genannt wurden, wird durch die relativ hohe Konsistenz der Urteile gezeigt, dass die Versuchspersonen mit den Begriffen auch ähnliche raumakustische Phänomene verbinden. Neben der Halligkeit fällt die Konsistenz der Items relativ hoch aus, welche im Allgemeinen mit dem späten Nachhall in Verbindung gebracht werden können: I5 (*groß – klein*), I7 (*hoher Raum – niedriger Raum*), I10 (*viel Raum – wenig Raum*) und I1 (*tragfähig – wenig tragfähig*). Eine mittlere Konsistenz ($0{,}25 < ICC(3,1) < 0{,}35$) wurde bei Aspekten der Raumbreite (I8), der Klarheit (I13), der empfundenen Lautheit (I12) sowie der Klangfülle (I14) beobachtet. Eine nicht signifikant ($\alpha = 0{,}05$) von 0 unterschiedliche Intraklassenkorrelation zeigt sich vor allem bei Klangfarben-Items (I9, I15, I16, I20), der Schnelligkeit der Raumantwort (I4) sowie der Präferenz (I6).

3.7.5. Zusammenhänge zwischen Bewertungsitems

Die linearen Zusammenhänge zwischen den Bewertungen der Gegensatzpaare können durch die Berechnung von Korrelationen anhand der Mittelwerte über die Versuchspersonen hergestellt werden. Vor Bildung der Mittelwerte wurden die Bewertungen je Item und Versuchsperson z-transformiert. Eine Standardisierung der Mittelwerte ist aufgrund der Definition der Pearson-Korrelation, bei welcher die Kovarianz zweier Wertereihen durch das Produkt ihrer Standardabweichungen geteilt wird, nicht nötig, da die z-Transformation in der Berechnungsvorschrift inbegriffen ist (Bortz und Schuster, 2010, S. 156).

Bei Betrachtung der Korrelationsmatrix in Tabelle 3.11 zeigt sich eine Gruppe von Items mit hohen Korrelationen und eine weitere, bei welcher sich keine signifikanten oder nur mittlere Korrelationen mit den übrigen Items ergeben. Zur erstgenannten Gruppe zählen Gegensatzpaare der Tragfähigkeit (I1), der Hörbarkeit des eigenen Instruments (I2), der Halligkeit (I3), der Raumgröße (I5, I7, I8), der Balance zwischen Instrument und Nachhall (I10), der empfundenen Lautstärke (I12), der Klarheit (I13) und mit Einschränkungen der Klangfülle (I14). All diese Items haben eine relativ konsistente Beurteilung innerhalb der Stichprobe gemein, der ICC(3,1) ist jeweils größer als 0,2.

Besonders auffällig ist die negative Korrelation zwischen I2 (*ich kann mich gut hören – ... nicht gut hören*) und I12 (*laut – leise*). Durch die gleichzeitig hohe Korrelation zwischen I12 und *trocken – hallig* (I3) kann vermutet werden, dass bei der Bewertung der Lautstärke des eigenen Instruments nicht die Unterstützung des Direktschalls durch frühe Reflexionen, sondern die gesamte akustische Situation beurteilt wurde.

Die übrigen Items scheinen hingegen eigenständige Wahrnehmungsdimensionen darzustellen, darunter vor allem I9 (*dunkel – hell*) und I20 (*Nachklang wird schärfer – ... wird dumpfer*) mit keinerlei signifikanten Korrelationen und somit keinen nachweislichen Zusammenhängen mit anderen Bewertungskriterien der Raumakustik. Diese Schlussfolgerung ist jedoch aufgrund der geringen Werte des ICC(3,1) der betrachteten Items nicht möglich. Wie die Boxplots in Abbildung 3.18 zeigen, ist eine Mittelwertbildung in diesen Fällen sehr problematisch. Durch die großen Wertebereiche kommt es zu beinahe zufälligen Verteilungen der Mittelwerte um 0. Die aus den Mittelwerten inkonsistenter Items berechneteten Korrelationen können aus diesem Grund nur sehr schwer interpretiert werden. Da in bisherigen Studien im Laufe der Auswertung zwar auch meist mit Mittelwerten gerechnet, gleichzeitig jedoch keine Aussagen über die Konsistenz der Bewertungen gemacht wurden, kann an dieser Stelle kein Vergleich mit anderen Studien erfolgen. So bleibt die Beobachtung, dass die Ratings aller Items mit einem ICC(3,1) $> 0,2$ hoch miteinander korreliert sind. In diesen Fällen kann daher eine eindimensionale Bewertung der Raumakustik innerhalb der Stichprobe angenommen werden.

Diese Annahme soll mithilfe einer Faktorenanalyse überprüft werden, in welcher die Items „gemäß ihrer korrelativen Beziehungen in voneinander unabhängige Gruppen klassifiziert werden“ (Bortz und Schuster, 2010, S. 386). Ob eine Faktorenanalyse anhand der Korre-

Tabelle 3.11.: Korrelationen zwischen Bewertungsitems. Rot markierte Korrelationen unterscheiden sich nicht signifikant von $r = 0$ ($\alpha = 0{,}05$). Die grau unterlegten Items sind der ersten Gruppe zuzuordnen, welche viele gemeinsame hohe Korrelationen zu Parametern aufweisen.

Items	**I1** tragfähig – wenig tragfähig	**I2** kann mich gut hören – ... nicht gut hören	**I3** trocken – hallig	I4 Raum spricht früh an – ... spät an	**I5** groß – klein	I6 gute Akustik – schlechte Akustik	**I7** hoher Raum – niedriger Raum	**I8** breiter Raum – schmaler Raum	I9 dunkel – hell	**I10** viel Raum – wenig Raum	I11 unterstützend – nicht unterstützend	**I12** laut – leise	**I13** diffus – klar	**I14** dünn – voll	I15 weich – hart	I16 rund – metallisch	I17 Nachhall klingt weich aus – ... bricht ab	I18 ausgewogen – unausgewogen	I19 Klangfarbe ist gleichmäßig – ... fluktuiert	I20 Nachklang wird schärfer – ... dumpfer
I1	–	−0,78	−0,97	−0,78	0,94	−0,14	0,96	0,89	−0,09	0,95	0,74	0,91	0,92	−0,9	0,61	0,08	0,68	−0,55	−0,74	0,08
I2	−0,78	–	0,88	0,58	−0,9	0,37	−0,81	−0,82	−0,01	−0,86	−0,26	−0,82	−0,85	0,86	−0,5	−0,1	−0,57	0,73	0,57	−0,25
I3	−0,97	0,88	–	0,75	−0,99	0,26	−0,97	−0,91	0,01	−0,98	−0,59	−0,93	−0,97	0,91	−0,62	−0,08	−0,7	0,65	0,71	−0,08
I4	−0,78	0,58	0,75	–	−0,81	0,15	−0,82	−0,7	−0,09	−0,84	−0,49	−0,68	−0,75	0,72	−0,77	−0,15	−0,4	0,49	0,63	−0,03
I5	0,94	−0,9	−0,99	−0,81	–	−0,34	0,98	0,9	−0,01	0,99	0,49	0,92	0,97	−0,89	0,63	0,05	0,62	−0,72	−0,72	0,13
I6	−0,14	0,37	0,26	0,15	−0,34	–	−0,33	−0,29	0,22	−0,27	0,44	−0,37	−0,35	0,09	0,3	0,72	0,24	0,72	0,19	−0,53
I7	0,96	−0,81	−0,97	−0,82	0,98	−0,33	–	0,87	−0,1	0,98	0,56	0,93	0,96	−0,85	0,58	−0,05	0,59	−0,7	−0,74	0,09
I8	0,89	−0,82	−0,91	−0,7	0,9	−0,29	0,87	–	−0,01	0,9	0,48	0,95	0,9	−0,96	0,56	0,05	0,46	−0,73	−0,72	0,32
I9	−0,09	−0,01	0,01	−0,09	−0,01	0,22	−0,1	−0,01	–	−0,04	0,05	−0,12	0	0,01	0,4	0,6	0,15	0,27	−0,02	−0,05
I10	0,95	−0,86	−0,98	−0,84	0,99	−0,27	0,98	0,9	−0,04	–	0,53	0,91	0,98	−0,9	0,69	0,1	0,63	−0,71	−0,75	0,06
I11	0,74	−0,26	−0,59	−0,49	0,49	0,44	0,56	0,48	0,05	0,53	–	0,51	0,48	−0,54	0,49	0,32	0,66	0,1	−0,55	−0,16
I12	0,91	−0,82	−0,93	−0,68	0,92	−0,37	0,93	0,95	−0,12	0,91	0,51	–	0,92	−0,88	0,43	−0,15	0,42	−0,78	−0,79	0,36
I13	0,92	−0,85	−0,97	−0,75	0,97	−0,35	0,96	0,9	0	0,98	0,48	0,92	–	−0,86	0,6	0,04	0,61	−0,75	−0,79	0,11
I14	−0,9	0,86	0,91	0,72	−0,89	0,09	−0,85	−0,96	0,01	−0,9	−0,54	−0,88	−0,86	–	−0,7	−0,25	−0,59	0,62	0,63	−0,15
I15	0,61	−0,5	−0,62	−0,77	0,63	0,3	0,58	0,56	0,4	0,69	0,49	0,43	0,6	−0,7	–	0,72	0,65	−0,17	−0,38	−0,38
I16	0,08	−0,1	−0,08	−0,15	0,05	0,72	−0,05	0,05	0,6	0,1	0,32	−0,15	0,04	−0,25	0,72	–	0,54	0,39	0,04	−0,51
I17	0,68	−0,57	−0,7	−0,4	0,62	0,24	0,59	0,46	0,15	0,63	0,66	0,42	0,61	−0,59	0,65	0,54	–	−0,02	−0,28	−0,5
I18	−0,55	0,73	0,65	0,49	−0,72	0,72	−0,7	−0,73	0,27	−0,71	0,1	−0,78	−0,75	0,62	−0,17	0,39	−0,02	–	0,61	−0,47
I19	−0,74	0,57	0,71	0,63	−0,72	0,19	−0,74	−0,72	−0,02	−0,75	−0,55	−0,79	−0,79	0,63	−0,38	0,04	−0,28	0,61	–	−0,38
I20	0,08	−0,25	−0,08	−0,03	0,13	−0,53	0,09	0,32	−0,05	0,06	−0,16	0,36	0,11	−0,15	−0,38	−0,51	−0,5	−0,47	−0,38	–

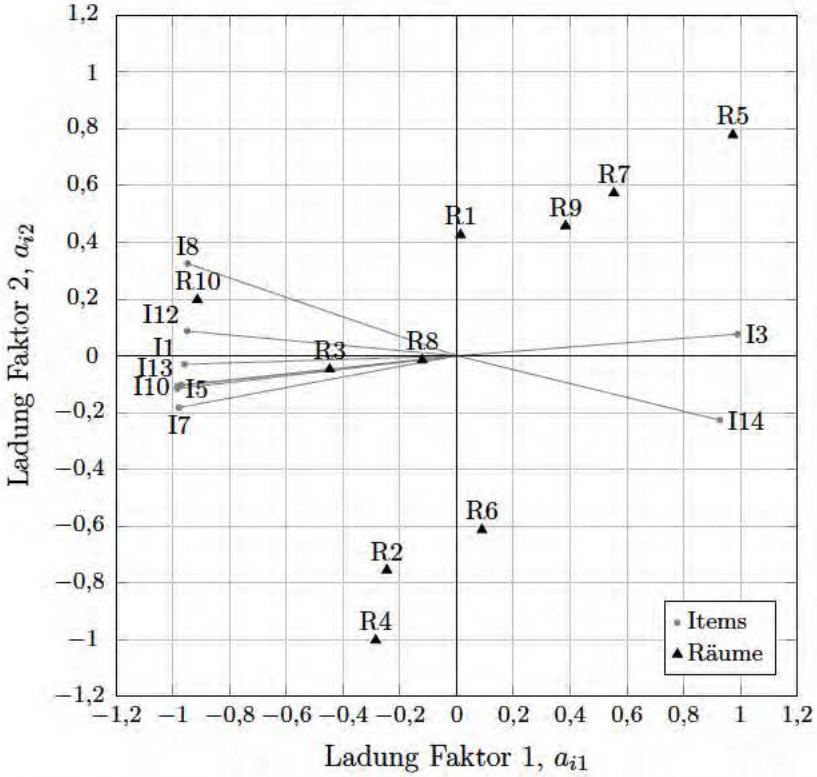


Abbildung 3.20.: Biplot aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14

lationsmatrix sinnvoll erscheint, kann durch das Kaiser-Meyer-Olkin-Kriterium (KMO)[25] geprüft werden (Backhaus et al., 2011, S. 342). Dieses gibt die Eignung der gesamten Korrelationsmatrix sowie der einzelnen Variablen in Einzahlwerten zwischen 0 und 1 an. Bei einem KMO-Wert unter 0,5 ist eine Korrelationsmatrix als Grundlage einer Faktorenanalyse ungeeignet, anhand der Einzelwerte der Variablen sollten solche mit geringen KMO-Werten aus der Analyse ausgeschlossen werden. Bei Verwendung der Items mit einem ICC(3,1) $> 0,2$ berechnet sich der KMO-Wert zu 0,503, die Eignung der Korrelationsmatrix ist daher als „kläglich", aber verwertbar, zu bezeichen (Backhaus et al., 2011, S. 343). Ein weiterer Ausschluss von Variablen soll jedoch nicht vorgenommen werden, da sich die Items bezüglich ihrer KMO-Werte nicht sehr deutlich voneinander unterscheiden. Für die Bestimmung der Anzahl der zu extrahierenden Faktoren stehen mehrere Kriterien zur Verfügung. Bei Anwendung des Kaiser-Guttmann-Kriteriums werden solche Faktoren berücksichtigt, deren Eigenwerte über 1 liegen. Eigenwerte geben den jeweils berücksichtigten Anteil an der Gesamtvarianz eines Faktors an, die Summe der Eigenwerte entspricht der Anzahl der einbezogenen Variablen. Durch das Kaiser-Guttmann-Kriterium werden demnach solche Faktoren extrahiert, welche einen höheren Anteil der Gesamtvarianz erklären als eine einzelne Variable. Eine weitere Möglichkeit, die Anzahl der Faktoren zu bestimmen, ist der Scree-Test.

[25] Das Kaiser-Meyer-Olkin-Kriterium wird auch MSA-Kriterium genannt aufgrund des Namens der Prüfgröße: *measure of sampling adequacy*.

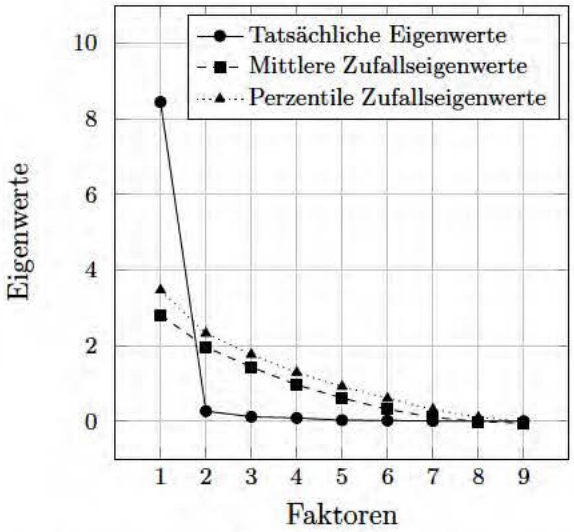


Abbildung 3.21.: Eigenwertdiagramm aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14. Erstellt mit dem Matlab-Skript *rawpar* (https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html, letzter Zugriff: 11. Juni 2013).

Bei diesem werden anhand eines Knicks im Eigenwertdiagramm solche Faktoren als wichtig betrachtet, deren Eigenwerte vor dem Knick liegen. Bei der Parallelanalyse werden die Eigenwerte der Faktoren der Korrelationsmatrix mit den Eigenwerten der Korrelationen von normalverteilten Zufallsvariablen verglichen. Solche Faktoren, deren Eigenwerte über der Eigenwertkurve der Zufallskorrelationen liegen, sollten demnach extrahiert werden (Bortz und Schuster, 2010, S. 416). Wie das Eigenwertdiagramm in Abbildung 3.21 zeigt, bildet sich gemäß der drei Kriterien jeweils nur ein zu extrahierender Faktor, der 94,88 % der Gesamtvarianz der 9 Items erklärt.

Desweiteren stellt sich die Frage nach dem Zusammenhang von Variablen und extrahierten Faktoren. Nach Fürntratt soll eine Variable einem Faktor nur zugeordnet werden, wenn der Quotient aus quadrierter Faktorladung und Kommunalität größer als 0,5 ist ($a_{ij}{}^2/h_i{}^2 > 0{,}5$), „d. h. wenn mindestens 50 % der aufgeklärten Varianz einer Variablen i auf den Faktor j entfallen“ (Bortz und Schuster, 2010, S. 423).

Die Ergebnisse der Faktorenanalyse werden durch den Biplot in Abbildung 3.20 und die Faktorladungen in Tabelle 3.12 dargestellt. Zur Berechnung wurde das Matlab-Skript *FA*[26] genutzt, welches die Hauptachsenmethode zur Faktorenextraktion verwendet. Im Gegensatz zur Hauptkomponentenanalyse (PCA) wird dabei nur der Anteil der gemeinsamen Varianz, nicht aber die spezifische Varianz oder die Fehlervarianz einer Variablen berücksichtigt. Es wird somit angenommmen, dass nicht die gesamte Varianz der Variablen durch die Faktoren aufgeklärt werden kann. Da durch die Datenerhebung mittels Ratingskalen sowie die Kom-

[26] Das Skript stammt von Lukas Malec und ist auf folgender Seite zu finden: http://www.mathworks.com/matlabcentral/fileexchange/14115, letzter Zugriff: 7. Juni 2013

Tabelle 3.12.: Faktorladungen a_{ij} und Kommunalitäten h^2 aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14

	Gegensatzpaar	a_{i1}	a_{i2}	h^2
I1	tragfähig – wenig tragfähig	−0,96	0,03	0,92
I3	trocken – hallig	0,99	−0,08	0,98
I5	groß – klein	−0,99	0,12	0,98
I7	hoher Raum – niedriger Raum	−0,98	0,18	0,99
I8	breiter Raum – schmaler Raum	−0,95	−0,33	1
I10	viel Raum – wenig Raum	−0,98	0,11	0,98
I12	laut – leise	−0,95	−0,09	0,91
I13	diffus – klar	−0,97	0,1	0,95
I14	dünn – voll	0,93	0,23	0,91

plexität der Forschungsfrage Messfehler zu erwarten sind und nicht davon auszugehen ist, dass die Variablen in einem Wiederholungsversuch exakt gleich bewertet werden würden, ist die Hauptachsenmethode der PCA vorzuziehen. Die gemeinsame Varianz einer Variablen mit den übrigen wurde durch die quadrierten multiplen Korrelationen geschätzt (Bortz und Schuster, 2010, S. 427–428).

Die Annahme einer einzigen den betrachteten Items zugrunde liegenden Bewertungsdimension, welche anhand der Korrelationsmatrix aufgestellt wurde, wird durch die Faktorenanalyse bestätigt. Die Faktorladungen der neun betrachteten Items auf Faktor 1 sind jeweils sehr hoch ($a_{i1} > 0{,}9$). Das Fürntratt-Kriterium ist bei allen Items erfüllt, sie sind eindeutig Faktor 1 zuzuordnen. Der zweite extrahierte Faktor soll vor allem verdeutlichen, dass die Items I14 (*dünn – voll*) und I8 (*breiter Raum – schmaler Raum*), welche bereits geringere Korrelationen mit den anderen Items zeigten, etwas höhere Ladungen bezüglich des zweiten Faktors aufweisen. Faktor 1 kann aufgrund der auf ihm ladenden Gegensatzpaare als *allgemein wahrgenommene Halligkeit* bezeichnet werden. Die Auswertung der Mittelwerte legt somit nahe, dass ein halliger Raum gleichzeitig als groß, hoch und breit, der wahgenommene Raumklang als diffus und voll bewertet wird, welcher das Instrument trägt und die empfundene Lautstärke des Instruments erhöht. Auch die Balance zwischen Instrument und Raum wird äquivalent bewertet. Eine Unterscheidung zwischen Nachhalllänge, der Energie früher und der Energie später Reflexionen durch Musiker ist somit anhand der Ratermittelwerte der konsistenten Items nicht zu erkennen. Beim Vergleich zwischen den im Biplot gemäß ihrer Faktorwerte verorteten Räume mit raumakustischen Parametern kann auf Anhieb kein eindeutiger Zusammenhang zu Faktor 1 hergestellt werden. Dies soll mit statistischen Methoden im nächsten Abschnitt geschehen.

Bei Einbeziehung aller Items in die Faktorenanalyse trotz geringem ICC und einer nach KMO-Kriterium (0,169) ungeeigneten Korrelationsmatrix, ergeben sich bei der Hauptachsenanalyse nach Kaiser-Guttmann-Kriterium vier und gemäß der Parallelanalyse 1 – 2 zu extrahierende Faktoren. Die Ergebnisse sollen aufgrund der genannten Kriterien an die-

ser Stelle nicht im Detail angegeben werden, eine kurze Auseinandersetzung soll aufgrund der Nähe zu Ergebnissen anderer Studien jedoch statt finden. Da gemäß des Fürntratt-Kriteriums kein Item Faktor 3 und 4 zuzuordnen wäre, wurden wiederum zwei Faktoren extrahiert. Neben den bereits untersuchten Items zeigen sich bei I2 (*ich kann mich gut hören – ... nicht gut hören*), I4 (*Raum spricht früh an – ... spricht spät an*) und I19 (*Klangfarbe ist gleichmäßig – ... fluktuiert*) hohe Ladungen auf den ersten Faktor ($a_{i1} > 0{,}75$). Faktor 2 definiert sich vor allem durch I6 (*gute Akustik – schlechte Akustik*), I20 (*Nachklang wird schärfer – ... wird dumpfer*) und I16 (*rund – metallisch*), Gegensatzpaare, durch welche die Klangfarbe und Präferenz bewertet werden. I9 (*dunkel – hell*), als weiteres Klangfarbenitem, besitzt zwar nur eine Ladung von $-0{,}39$ auf Faktor 2, ist diesem jedoch aufgrund der minimalen Ladung auf Faktor 1 zuzuordnen. Ohne Beachtung der Konsistenz der Items könnte somit der zweite Faktor als *Klangfarbe* bezeichnet werden und in Verbindung mit den von Gade (1989b) unter Ensemblebedingungen gefundenen Faktoren gebracht werden, da sich die Ergebnisse teilweise decken. Diese Interpretation soll beispielhaft zeigen, dass die Verwendung der Mittelwerte inkonsistenter Items zwar plausible Schlüsse zulassen kann, aus Sicht des Autors jedoch aufgrund der geringen Konsistenz und der kleinen Stichprobengröße nicht statistisch fundiert ist. Die bewertete Unterstützung (I11) zeigt auf beiden Faktoren in etwa gleich hohe Ladungen und bei zwei Faktoren die geringste Kommunalität aller Items. Bei zusätzlicher Extraktion eines dritten Faktors, lädt I11 auf diesen als einziges Item am höchsten, während es auch mittlere Ladungen auf den ersten und zweiten Faktor aufweist ($a_{\mathrm{I11}\,1} = -0{,}57$, $a_{\mathrm{I11}\,2} = -0{,}51$, $a_{\mathrm{I11}\,3} = -0{,}61$). Der beurteilten Unterstützung liegen in dieser Studie somit mehrere Dimensionen zugrunde.

3.7.6. Korrelationen zwischen raumakustischen Parametern und Items

Die Validierung der Bühnenakustik durch Musiker beispielsweise mithilfe von Fragebögen kann sehr zeitaufwendig und kostspielig sein. Daher ist die Möglichkeit der Vorhersage der Musikerwahrnehmung durch physikalisch messbare raumakustische Parameter für die Raumakustikplanung von großer Bedeutung. Auch für die Wahrnehmungsforschung ist von Interesse, wie sich Änderungen im Schallfeld, welche mithilfe der physikalischen Maße festgehalten werden, auf die Bewertung verschiedener Wahrnehumgsaspekte auswirken. Aus diesen Gründen werden im Folgenden die Zusammenhänge zwischen den zwanzig verwendeten Bewertungsitems und den in Abschnitt 3.3.1.2 vorgestellten raumakustischen Parametern untersucht. Zu den gewählten Parametern zählen Energierelationen verschiedener zeitlicher Abschnitte der Impulsantwort, Abklingzeiten über verschiedene Pegelbereiche der Energiezerfallskurve (EDC) sowie Verhältnisse von Oktavbandwerten dieser Maße. Sie stammen sowohl aus dem Bereich der Bühnenakustik als auch der Auditoriumsakustik und sollen die Eigenschaften des Schallfelds möglichst umfassend abbilden. Bei der statistischen Auswertung wurden die in den jeweiligen Definitionen angegebenen Oktavbandmittelwerte benutzt. Die Korrelationen der Itembewertungen mit Parameter-Mittelwerten über höhere und nied-

rigere Oktavbänder ergaben sehr ähnliche Ergebnisse, weshalb sie an dieser Stelle nicht gesondert betrachtet und aufgelistet werden.

Ähnlich wie bei der Untersuchung der Zusammenhänge der einzelnen Items, lassen sich die Korrelationen der Raumbewertungen pro Item mit den gemessenen raumakustischen Parametern bestimmen. Dazu können entweder die Mittelwerte über alle Versuchspersonen oder die individuellen Ratings verwendet werden. Wie bereits hervorgehoben, ist die Mittelwertbildung im Falle des erhobenen Datensatzes bei manchen Items aufgrund der geringen Urteilerkonsistenz problematisch. Eine statistische Auswertung auf Grundlage von Mittelwerten bei inkonsistenter Bewertung durch die Versuchspersonen erscheint nicht sinnvoll. Items, bei welchen sich der ICC(3,1) nicht signifikant von 0 unterscheidet, wurden daher an dieser Stelle aus der Auswertung ausgeschlossen (I4: *Raum spricht früh an – Raum spricht spät an*, I6: *gute Akustik – schlechte Akustik*, I9: *dunkel – hell*, I15: *weich – hart*, I16: *rund – metallisch*, I20: *Nachklang wird schärfer – ... wird dumpfer*). Die Korrelationen nach Pearson der übrigen Bewertungsitems mit den ausgewählten akustischen Parametern sind in Tabelle 3.13 angegeben. Wie auch in Abschnitt 3.7.5 wurden die zur Berechnung der Korrelationskoeffizienten genutzten Mittelwerte über die Versuchspersonen anhand der standardisierten Ratings gewonnen.

Der raumakustische Parameter mit den höchsten Korrelationen zu mehreren Items ist die Nachhallzeit $T_{30\,1\mathrm{m}}$. Wie in Abschnitt 3.7.5 gezeigt wurde, sind die Itembewertungen untereinander sehr hoch korreliert, weshalb auch die zahlreichen Korrelationen bei einem einzigen raumakustischen Parameter durchaus nachvollziehbar sind. Bei allen Items mit signifikanten Korrelationen zu $T_{30\,1\mathrm{m}}$ zeigten sich bereits bei der Faktorenanalyse, welche alle Items miteinbezog, hohe Ladungen auf den ersten Faktor. Die Polungen der Korrelationen entsprechen denen der Faktorladungen a_{i1}. Das am konsistentesten bewertete Item I3 (*trocken – hallig*) korreliert somit am höchsten mit der Nachhallzeit ($r = 0{,}88$) und nur vergleichsweise gering mit der frühen Abklingzeit $EDT_{1\mathrm{m}}$ ($r = 0{,}43$), welche in der Auditoriumsakustik als Prädiktor für die wahrgenommene Halligkeit dient. Dass auch die Bewertungen von I18 (*ausgewogen – unausgewogen*) mit 0,81 vergleichsweise hoch mit der Nachhallzeit korrelieren, verwundert, da bei diesem Item nicht unbedingt ein linearer Zusammenhang zur Nachhallzeit zu erwarten ist. An dieser Stelle muss auch der geringe ICC(3,1) von 0,07, welcher bei I18 auftrat, beachtet werden. Die Problematik der Verwendung der Mittelwerte bei diesem Item wird anhand von Abbildung 3.22 deutlich, in welcher die standardisierten Bewertungen von I18 zur Nachhallzeit $T_{30\,1\mathrm{m}}$ aufgetragen sind. Durch die geringe Konsistenz der Urteile und eine somit große Streuung der individuellen Urteile um den arithmetischen Mittelwert, ist nicht auszuschließen, dass die hohe Korrelation zur Nachhallzeit nur zufällig auftrat. Von einem eindeutigen Zusammenhang kann somit trotz eines hohen Korrelationskoeffizienten nicht die Rede sein. Bezüglich dieses Zusammenhangs müssten weitere Untersuchungen folgen. Ähnliches gilt für die hohe beobachtete Korrelation der Hörbarkeit des eigenen Instruments (I2) mit $T_{30\,1\mathrm{m}}$ ($r = 0{,}93$), obwohl dieser Zusammenhang als sehr interessant eingestuft werden könnte. Da die Urteilerkonsistenz

Tabelle 3.13.: Korrelationen nach Pearson zwischen akustischen Parametern der Raumsimulationen und Bewertungsitems, welche einen signifikant von 0 unterschiedlichen ICC(3,1) aufweisen. Es handelt sich bei den Parametern jeweils um die in den Definitionen angegebenen Frequenzbereichsmittelwerte. Nicht signifikante Korrelationen ($\alpha = 0{,}05$) sind rot eingefärbt. Zur inhaltlichen Auswertung sind in der letzten Splate die Werte des ICC(3,1) angegeben.

	Gegensatzpaar	$T_{30\,1\mathrm{m}}$	$EDT_{1\mathrm{m}}$	$G_{1\mathrm{m}}$	$C_{50\,1\mathrm{m}}$	$C_{80\,1\mathrm{m}}$	$T_{S\,1\mathrm{m}}$
I1	tragfähig – wenig tragfähig	−0,82	−0,49	−0,73	0,72	0,78	−0,71
I2	ich kann mich gut hören – ... nicht gut hören	0,93	0,47	0,78	−0,48	−0,57	0,72
I3	trocken – hallig	0,88	0,43	0,75	−0,65	−0,72	0,69
I5	groß – klein	−0,92	−0,36	−0,73	0,54	0,61	−0,62
I7	hoher Raum – niedriger Raum	−0,88	−0,32	−0,65	0,55	0,62	−0,58
I8	breiter Raum – schmaler Raum	−0,81	−0,54	−0,73	0,71	0,73	−0,78
I10	viel Raum – wenig Raum	−0,91	−0,34	−0,69	0,54	0,60	−0,60
I11	unterstützend – nicht unterstützend	−0,25	−0,57	−0,52	0,76	0,78	−0,55
I12	laut – leise	−0,83	−0,53	−0,66	0,69	0,73	−0,77
I13	diffus – klar	−0,87	−0,34	−0,61	0,58	0,62	−0,61
I14	dünn – voll	0,82	0,62	0,85	−0,71	−0,74	0,81
I17	Nachhall klingt weich aus – ... bricht ab	−0,47	−0,32	−0,64	0,56	0,62	−0,43
I18	ausgewogen – unausgewogen	0,81	0,16	0,29	−0,23	−0,24	0,45
I19	Klangfarbe ist gleichmäßig – ... fluktuiert	0,61	0,43	0,31	−0,57	−0,56	0,58

	ST_{early}	ST_{total}	ST_{late}	LQ_{7-40}	$G_{e\,1\mathrm{m}}$	$G_{l\,1\mathrm{m}}$	$G_{w\,1\mathrm{m}}$
I1	−0,59	−0,68	−0,81	0,24	−0,69	−0,83	−0,68
I2	0,53	0,57	0,59	0,07	0,75	0,67	0,61
I3	0,52	0,61	0,74	−0,22	0,72	0,78	0,63
I5	−0,46	−0,54	−0,64	0,14	−0,71	−0,69	−0,57
I7	−0,47	−0,55	−0,66	0,18	−0,64	−0,68	−0,54
I8	−0,49	−0,58	−0,72	0,23	−0,67	−0,78	−0,63
I10	−0,44	−0,52	−0,63	0,19	−0,67	−0,67	−0,52
I11	−0,58	−0,68	−0,80	0,35	−0,47	−0,77	−0,68
I12	−0,57	−0,65	−0,76	0,16	−0,61	−0,76	−0,65
I13	−0,37	−0,47	−0,64	0,32	−0,57	−0,66	−0,47
I14	0,61	0,68	0,75	−0,09	0,81	0,82	0,71
I17	−0,40	−0,49	−0,62	0,35	−0,62	−0,67	−0,5
I18	0,19	0,21	0,28	0,00	0,26	0,27	0,16
I19	0,31	0,40	0,59	−0,37	0,23	0,52	0,39

	BR_T	BR_{EDT}	BR_G	TR_T	TR_{EDT}	TR_G	$EDTF$	ICC(3,1)
I1	−0,03	0,13	0,10	0,52	0,42	0,60	−0,01	0,472
I2	−0,32	−0,06	−0,15	−0,16	−0,45	−0,38	0,22	0,164
I3	−0,10	−0,09	−0,06	−0,40	−0,48	−0,55	0,11	0,694
I5	0,23	0,08	0,03	0,28	0,47	0,43	−0,17	0,658
I7	0,18	0,08	0,06	0,34	0,42	0,44	−0,14	0,611
I8	0,10	0,33	0,11	0,41	0,59	0,62	0,12	0,341
I10	0,21	0,06	−0,06	0,25	0,49	0,39	−0,22	0,565
I11	−0,63	0,20	0,10	0,80	0,15	0,62	0,26	0,108
I12	0,10	0,33	0,28	0,47	0,50	0,59	0,10	0,266
I13	0,16	0,04	−0,05	0,26	0,62	0,46	−0,20	0,312
I14	−0,08	−0,33	−0,06	−0,40	−0,46	−0,54	−0,03	0,261
I17	−0,24	−0,25	−0,24	0,35	0,23	0,48	−0,27	0,117
I18	−0,58	−0,13	−0,13	0,12	−0,56	−0,15	0,23	0,068
I19	0,12	−0,13	−0,05	−0,33	−0,72	−0,36	0,06	0,129

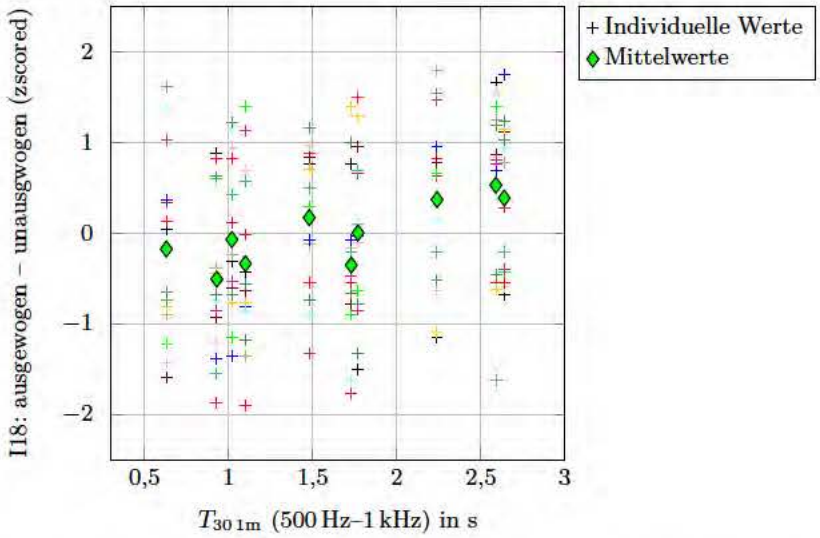


Abbildung 3.22.: Nachhallzeit $T_{30\,1\mathrm{m}}$ aufgetragen zu den standardisierten Bewertungen der Raumsimulationen bezüglich I18 (*ausgewogen – unausgewogen*)

der Items bei der Auswertung und Deutung der Korrelationen berücksichtigt werden sollte, sind neben den Korrelationskoeffizeinten auch noch einmal die jeweiligen ICC(3,1)-Werte in Tabelle 3.13 aufgelistet.

Beim Stärkemaß $G_{1\mathrm{m}}$ zeigen sich ähnliche Ausprägungen der Korrelationskoeffizienten wie bei der Nachhallzeit, wobei diese in den meisten Fällen etwas kleiner ausfallen. Eine etwas höhere positive Korrelation als bei $T_{30\,1\mathrm{m}}$ ergibt sich bei I14 (*dünn – voll*) mit $r = 0,85$. Der Klang eines Raumes wurde im Mittel folglich als umso *voller* bewertet, je höher das Stärkemaß ausfiel. Während das frühe Stärkemaß $G_{e\,1\mathrm{m}}$ allgemein geringer mit den Urteilermittelwerten korreliert, sind bei $G_{l\,1\mathrm{m}}$ vor allem die Korrelationen mit I1 (*tragfähig – wenig tragfähig*) und I11 (*unterstützend – nicht unterstützend*) auffällig. Sowohl die mittlere wahrgenommene Tragfähigkeit als auch die Unterstützung des eigenen Instruments scheinen somit vor allem mit der Stärke der späten Energie in Zusammenhang zu stehen. Außerdem ist $G_{l\,1\mathrm{m}}$ der einzige Parameter, bei welchem die Korrelation zu I17 (*Nachhall klingt weich aus – ... bricht ab*) signifikant von 0 verschieden ausfällt. Dass die Stärke später Reflexionen mit der Beurteilung des Zeitverlaufs des Ausklingvorgangs zusammen hängt, ist nachvollziehbar. Auffällig ist, dass I2 (*ich kann mich gut hören – ... nicht gute hören*) höher mit $G_{e\,1\mathrm{m}}$ korreliert als mit $G_{l\,1\mathrm{m}}$. Zwar zeigt das Stärkemaß ohne Zeitfensterung eine noch höhere Korrelation zu I2, jedoch ist anzunehmen, dass die frühe Energie einen höheren Einfluss auf die wahrgenommen Lautstärke des eigenen Instruments hat. Trotz der für den Autor plausiblen Zusammenhänge muss auch bei den Items I2, I11 und I17 der geringe ICC(3,1) von 0,1–0,16 in der Auswertung berücksichtigt werden. Ähnlich wie im Fall des Zusammenhangs der Ausgewogenheit (I18) mit der Nachhallzeit $T_{30\,1\mathrm{m}}$ gilt es, diese Zusammenhänge in zukünftigen Studien zu überprüfen.

$C_{50\,1\mathrm{m}}$ und $C_{80\,1\mathrm{m}}$ beschreiben das Verhältnis von früher zu später Energie und erreichen

umso größere Werte, je kürzer die Nachhallzeit eines Raums ausfällt. Daher unterscheiden sich auch die Korrelationen mit den Items im Vergleich zu denen der Nachhallzeit im Vorzeichen. Die Höhe der Koeffizienten ist ähnlich wie beim Stärkemaß, wobei sich der größte Unterschied bei Item I11 (*unterstützend – nicht unterstützend*) zeigt, bei welchem die Korrelation – anders als bei der Nachhallzeit und dem Stärkemaß – signifikant ist ($r = 0,78$). I11 korreliert im Betrag ähnlich hoch mit ST_{late} und $G_{late\,1\mathrm{m}}$, bei welchen die Integrationsgrenzen sehr ähnlich gewählt wurden und die Werte von ST_{late} und $G_{late\,1\mathrm{m}}$ mit denen von $C_{80\,1\mathrm{m}}$ daher über die 10 Raumsimulationen selbst mit −0,99 korrelieren. Der Parameter ST_{total}, welcher nach Gade (1992) eigentlich die Unterstützung des Raums für das eigene Instrument des Musikers beschreibt, zeigt eine geringere, jedoch noch signifikante Korrelation zu I11. Neben ST_{late} zeigt sich auch bei TR_T eine Korrelation von 0,8 zur wahrgenommenen Unterstützung I11, wobei TR_T gleichzeitig keinerlei signifikante Korrelationen zu anderen Items aufweist. Anhand der Mittelwerte der beobachteten Stichprobe bietet ein Raum eine größere unterstützende Wirkung, je geringer die Abnahme der Nachhallzeiten von mittleren zu hohen Oktavbändern ausfällt.[27] Aufgrund der Korrelationen mit anderen Parametern als der Nachhallzeit, könnte die Unterstützung eine eigenständige Wahrnehmungsdimension darstellen.

Auffallend ist, dass sich bei den Parametern $EDT_{1\mathrm{m}}$, ST_{early}, LQ_{7-40}, BR_{EDT} und $EDTF$ keinerlei signifikante Korrelationen mit Items ergeben. All diesen Parametern ist gemeinsam, dass für sie der frühe Anteil der Impulsantwort maßgebend ist, auch wenn es sich um Engergierelationen, die frühe Abklingzeit und deren Frequenzverhältnisse handelt. Als möglicher Grund für die hier nicht zu erkennenden Einflüsse der genannten Parameter auf die raumakustische Wahrnehmung durch Musiker kann die aus der geringen Distanz zwischen Ohren und Instrument resultierende Maskierung früher Reflexionen genannt werden. Durch den großen Pegelunterschied und die gleichzeitig sehr geringe Verzögerung zwischen Direktschall und frühen Reflexionen sind letztere kaum wahrnehmbar und könnten somit bei der Bewertung in der Solisten-Situation keine Rolle spielen. Zwar handelt es sich bei ST_{early} und LQ_{7-40} um zwei eigens für die Bühnenakustik vorgeschlagene Indizes, allerdings haben sich bei diesen vor allem Zusammenhänge zur Einschätzung der gegenseitigen Hörbarkeit im Ensemble gezeigt. Aus theoretischen Überlegungen heraus, ist eine höhere unterstützende Wirkung durch frühe Reflexionen für weiter entfernte Schallquellen durchaus plausibel, da sich dadurch die Pegeldifferenzen zwischen direktem und reflektiertem Schall verringern. Bei frühen Reflexionen ist daher mit einer geringeren Maskierung durch den Direktschall zu rechnen. Dass wie von van den Braak und van Luxemburg (2008) berichtet Änderungen des LQ_{7-40} in einem Teilversuch nur vom Dirigenten, jedoch nicht von den Musikern wahrgenommen wurden, kann ebenso durch die unterschiedliche Stärke des Direktschalls für Musiker und Dirigenten begründet werden. Durch die geringe Entfernung zwischen Quelle und Empfänger bei der Messung der Parameter ergeben sich zudem bei fünf Räumen sehr

[27] Durch die Luftabsorption und die Eigenschaften poröser Absorber fällt die Nachhallzeit zu hohen Frequenzen naturgemäß ab. TR_T ist daher in allen dargebotenen Raumsimulationen kleiner als 1.

ähnliche $EDT_{1\mathrm{m}}$-Werte nahe 0 s (siehe 3.3.1.2). Selbst wenn frühe Reflexionen nicht vom Direktschall maskiert würden, stellt sich die Ausprägung dieses Parameters daher als problematisch dar. Drei weitere Paramater, bei welchen die frühe Energie berücksichtigt wird und bei denen sich im Gegensatz signifikante Korrelationen zu Items zeigen, sind $G_{e\,1\mathrm{m}}$, $C_{80\,1\mathrm{m}}$ und $C_{50\,1\mathrm{m}}$. Der größte Unterschied bei der Berechnung zu den vorher genannten Parametern besteht darin, dass die späte Energie oder die Energie der Freifeld-Entsprechung der gleichen Quelle in 10 m Entfernung einbozogen wird, indem die frühe Energie zu dieser ins Verhältnis gesetzt wird.

Desweiteren zeigen sich auch bei den Parametern BR_T, BR_G und TR_G keine signifikanten Korrelationen. Diese Beobachtung könnte durch die inkonsistente Beurteilung der klangfarblichen Aspekte erklärt werden, da es sich bei den drei Parametern um Verhältnisse verschiedener Frequenzbereiche der Nachhallzeit bzw. des Stärkemaßes handelt. Auch die geringen Korrelationen von BR_{EDT} und $EDTF$ könnten durch diesen Umstand zustande gekommen sein. Bei $G_{w\,1\mathrm{m}}$, einer gewichteten Kombination der 125 und 250 Hz-Oktavbänder aus frühem und spätem Stärkemaß, berechnen sich zwar signifikante Korrelationen, allerdings sind diese im Vergleich zu anderen Parametern relativ gering. Der von Bradley (2011) im Auditorium hergestellte Zusammenhang von G_w und der Wahrnehmung der Bass-Lautstärke kann aufgrund geringern Konsistenz der entsprechenden Klangfarben-Items nicht geprüft werden. Die zwischen TR_{EDT} und der Klangfarbenfluktuation (I19) berechnte signifikante Korrelation von $-0{,}72$ ist auffällig, da sowohl das TR_{EDT} als auch I19 keine weiteren signifikanten Korrlationen aufweisen. Dass diese Korrelation auch bei einer konsistenteren Bewerung von I19 aufgetreten wäre, ist fraglich, da sich bei $EDT_{1\mathrm{m}}$ selbst keine signifikant von 0 verschiedenen Korrelationen ergaben.

Die Korrelationen der Items zu den raumakustischen Parametern bestätigen die Beobachtungen bei der Untersuchung der Zusammenhänge zwischen den Items untereinander. Die Items, welche hohe Ladungen auf den als *allgemein wahrgenommene Halligkeit* bezeichneten Faktor zeigten, korrelieren im Mittel sehr hoch mit der Nachhallzeit $T_{30\,1\mathrm{m}}$. Die übrigen Korrelationen können wie veranschaulicht aufgrund der zum Teil geringen Urteilerkonsistenz nicht als statistisch fundiert betrachtet werden und sollten in folgenden Studien mithilfe einer größeren Stichprobe überprüft werden.

3.7.7. Untersuchung der Teststärke (Power)

Durch die Berechnung von Teststärken können die Ergebnisse einer Studie abgesichert bzw. kritisch hinterfragt werden. Desweiteren ist es möglich, auf gleichem Wege Empfehlungen für optimale Stichprobengrößen für zukünftige Studien zu formulieren. Sowohl die Posthoc- als auch die Apriori-Analyse basiert auf den geschätzten Effektgrößen der interessierenden Gesamtpopulation. Werden die in einer Stichprobe beobachteten Effektgrößen zur Analyse verwendet, spricht man von einer sogenannten retrospektiven Teststärkenbestimmung, welche nicht die wahre Power sondern eine Schätzung auf Grundlage der Untersuchungsdaten

darstellt (Faul et al., 2007, S. 176). Ein Test, welcher einen Effekt anhand eines gewählten α-Niveaus als nicht signifikant bestimmt, wird nach dieser Methode zwangsläufig eine geringe Teststärke erreichen.

> Both the p-value and power are dependent upon the observed effect size and so are inversely related such that tests with high p-values tend to have low power and visa-versa. Therefore calculating power using the observed effect size and variance is simply a way of re-stating the statistical significance of the test.
>
> (Thomas, 1997, S. 3)

Der retrospektiven Teststärkenanalyse liegt gleichzeitig die fragwürdige Annahme zu Grunde, die Stichproben-Effektgröße sei identisch mit der Populations-Effektgröße. Faul et al. (2007) empfehlen daher die Bestimmung der Effektgröße aus testtheoretischen Überlegungen. Jedoch sind die zu erwartenden Effektgrößen in der Gesamtpopulation auch aufgrung der fehlenden Dokumentation bisheriger Publikationen zur raumakustischen Wahrnehmung von Musikern unbekannt. Daher soll die Teststärkenanalyse trotz allem anhand der in dieser Studie beobachteten Effektgrößen und somit retrospektiv angewendet werden.

Als Grundlage der Teststärkenanalyse sollen die Effekte der multi- und univariaten Varianzanalysen genutzt werden. Denn bevor beispielsweise Zusammmenhänge zu akustischen Parametern hergestellt werden, sollte die Untescheidung der raumakustischen Umgebungen anhand der elizierten Items sicher gestellt werden. Der Einfluss der Instrumente bzw. -gruppen soll dabei unberücksichtigt bleiben. Der Einfluss dieses Faktors soll vielmehr durch die Wahl einer größeren Anzahl von Versuchspersonen ausgeglichen werden.

> Die Teststärke (*power*) eines statistischen Tests ist die Wahrscheinlichkeit, sich bei Anwendung dieses Tests für eine der Nullhypothese (H_0) entgegengestellte einfache Alternativhypothese (H_1) zu entscheiden, wenn diese in der zugrundeliegenden Population tatsächlich gültig ist.
>
> (Buchner et al., 1996, S. 123).

Sie beschreibt somit die Chance, einen Effekt bei gegebenem Signifikanzniveau statistisch nachweisen zu können. Da die Teststärke durch $1 - \beta$ definiert ist, wird durch ihr Komplement die Wahrscheinlichkeit des β-Fehlers beziffert, welcher auftritt, wenn auf Basis des errechneten α-Werts fälschlicherweise die Alternativhypothese H_1 verworfen und somit die Nullhypothese beibehalten wird.

Die Teststärkeanalysen wurden mithilfe des Compterprogramms *G*Power 3.1.5* unter Verwendung der Effektgrößen aus der durchgeführten MANOVA sowie den ANOVAs je Item (siehe Abschnitt 3.7.2) durchgeführt. Es wurden die Daten der Varianzanalysen mit standardisierten Daten verwendet.

Neben der Effektgröße f, welche durch Eingabe des partiellen η^2 berechnet wurde, sind das Signifikanzniveau ($\alpha = 0{,}05$), die Stichprobengröße (14), die Anzahl der Messwiederholungen (10), die Gruppenanzahl (1) sowie die mittlere Korrelation zwischen den Messwiederholungen anzugeben. Letztere wurde anhand der Versuchsdaten in einem Matlabskript

ermittelt, bei welchem die Korrelationen zwischen der gegenseitigen Raumbewertungen aller Versuchspersonen für die Anovas je Item und für die MANOVA global gemittelt wurden.

Abbildung 3.23 zeigt die berechneten Teststärken in Abhängigkeit von der Stichprobengröße für den Innersubjektfaktor der MANOVA mit Messwiederholung. Das partielle $\eta^2 = 0{,}229$ des Pillai-Spur-Tests ergab dabei eine Effektgröße von $f = 0{,}545$, die mittlere Korrelation zwischen den Messwiederholungen errechnete sich zu $\bar{r}_{MW} = 0{,}237$. Die in der MANOVA erreichte Teststärke beträgt bei einer Stichprobe von 14 Musikern $1-\beta = 0{,}778$. Da die Konsequenzen eines α-Fehlers nach Cohen etwa viermal so gravierend sind wie die eines β-Fehlers empfiehlt sich bei einem Signifikanzniveau von $\alpha = 0{,}05$ eine Teststärke von $1-\beta = 0,8$ Bortz und Döring (2006, S. 604). Eine solche wäre auf Grundlage der retrospektiven Analyse bei der MANOVA bereits bei einer Stichprobengröße von $n = 15$ gegeben. Wie bereits in Abschnitt 3.7.2 im Detail beschrieben wurde, wird beim Signifikanztest der MANOVA mit Messwiederholungen die Nullhypothese geprüft, dass sich die Mittelwerte je Item über alle Messwiederholungen gleichen. Eine Verringerung der zu bewertenden raumakustischen Umgebungen oder Items würde sich somit auch negativ auf die Teststärke auswirken. Die bereits ab einer Stichprobengröße von $n = 20$ einsetzende Sättigung der Power, aufgrund der Natur des Hypotehesentests der MANOVA mit Messwiederholungen nicht, dass eine solche für die Untersuchung der raumakustischen Wahrnehmung durch Musiker absolut ausreichende wäre. Um eine Empfehlung geben zu können, müssen die Items im einzelnen betrachtet werden, da sich die aufteretenden Effekte sehr deutlich unterschieden.

Aus diesem Grund wurden desweiteren Teststärkenanalysen der ANOVAs mit Messwiederholungen je Item mit *G*Power* durchgeführt, deren Ergebnisse in Tabelle 3.14 aufgelistet sind. Anhand der Effektgrößen wurden zum einen die im Fall dieser Studie bei einer Stichprobe von 14 Musikern erreichten Teststärkenwerte berechnet. Zum anderen wurden die für eine zu erreichende Power von 0,8 sowie 0,95 benötigten Stichprobengrößen ermittelt. Die Hypothesentests der ANOVAs mit Messwiederholung sowie die im Zuge der ICC-Berechnungen getestete Hypothese $ICC = 0$ zeigten ähnliche Ergebnisse, da sie auf ähnlichen Varianzkomponenten beruhen. Der signifikante Effekt der ANOVA ist somit mit einem signifikant von 0 verschiedenen ICC vergleichbar. Angenommen die beobachteten Effektgrößen entsprechen denen einer zukünftigen Studie, würden die errechneten Stichprobengrößen noch nicht in einem Datensatz resultieren, welcher sich zur Berechnung von Korrelationen mit raumakustischen Parametern eignet. Dass sich bei bei Item I3 (*trocken – hallig*) bereits bei $n = 3$ eine Power von $1 - \beta = 0,99$ ergibt, sich allerdings gleichzeitig bei 14 Musikern noch immer kein sehr hoher ICC(3,1) von 0,694 beobachten lässt, lässt die berechneten Stichproben beurteilen. Um eine ähnlich hohe Konsistenz auch bei klangfarblichen Aspekten zu erreichen, werden somit deutlich mehr als beispielsweise die berechneten 21 bei I9 (*dunkel – hell*) Versuchspersonen benötigt.

Tabelle 3.14.: Ergebnisse der retrospektiven Poweranalysen der ANOVAs je Item. Analyse 1 berechnet die Teststärken der vorliegenden Studie mit der Stichprobengröße $n = 14$. In den Analysen 2 und 3 wurden die Stichprobengrößen abhängig von der mindestens zu erreichenden Teststärke ermittelt. Vorgabe bei Analyse 2: $1 - \beta > 0{,}95$. Vorgabe bei Analyse 3: $1 - \beta > 0{,}8$.

		Eingabeparameter			Analyse 1		Analyse 2		Analyse 3	
	Gegensatzpaar	part. η^2	$\bar{r}_{MW}$	ϵ G-G	n	$\beta - 1$	n	$\beta - 1$	n	$\beta - 1$
I1	tragfähig – wenig tragfähig	0,496	0,264	0,560	14	1,000	4	0,959	3	0,814
I2	ich kann mich gut hören – ich kann mich nicht gut hören	0,159	0,229	0,560	14	0,921	16	0,956	11	0,823
I3	trocken – hallig	0,737	0,246	0,516	14	1,000	3	0,998	2	0,822
I4	Raum spricht früh an – Raum spricht spät an	0,045	0,257	0,568	14	0,334	57	0,952	37	0,800
I5	groß – klein	0,702	0,226	0,463	14	1,000	3	0,982	3	0,982
I6	gute Akustik – schlechte Akustik	0,087	0,220	0,636	14	0,655	28	0,954	19	0,817
I7	hoher Raum – niedriger Raum	0,688	0,278	0,523	14	1,000	3	0,991	3	0,991
I8	breiter Raum – schmaler Raum	0,395	0,225	0,621	14	1,000	6	0,983	4	0,848
I9	dunkel – hell	0,084	0,221	0,577	14	0,602	31	0,953	21	0,818
I10	viel Raum – wenig Raum	0,593	0,273	0,571	14	1,000	3	0,951	3	0,951
I11	unterstützend – nicht unterstützend	0,182	0,221	0,559	14	0,957	14	0,957	10	0,842
I12	laut – leise	0,312	0,233	0,590	14	1,000	7	0,953	5	0,811
I13	diffus – klar	0,304	0,255	0,531	14	1,000	8	0,966	6	0,871
I14	dünn – voll	0,306	0,223	0,582	14	1,000	8	0,972	6	0,885
I15	weich – hart	0,166	0,236	0,592	14	0,947	15	0,962	10	0,820
I16	rund – metallisch	0,077	0,238	0,567	14	0,560	34	0,955	22	0,802
I17	Nachhall klingt weich aus – Nachhall bricht ab	0,148	0,272	0,513	14	0,892	17	0,951	12	0,826
I18	ausgewogen – unausgewogen	0,125	0,209	0,674	14	0,861	19	0,959	13	0,827
I19	Klangfarbe ist gleichmäßig – Klangfarbe fluktuiert	0,187	0,195	0,707	14	0,983	12	0,958	8	0,803
I20	Nachklang wird schärfer – Nachklang wird dumpfer	0,095	0,233	0,624	14	0,707	26	0,958	17	0,809

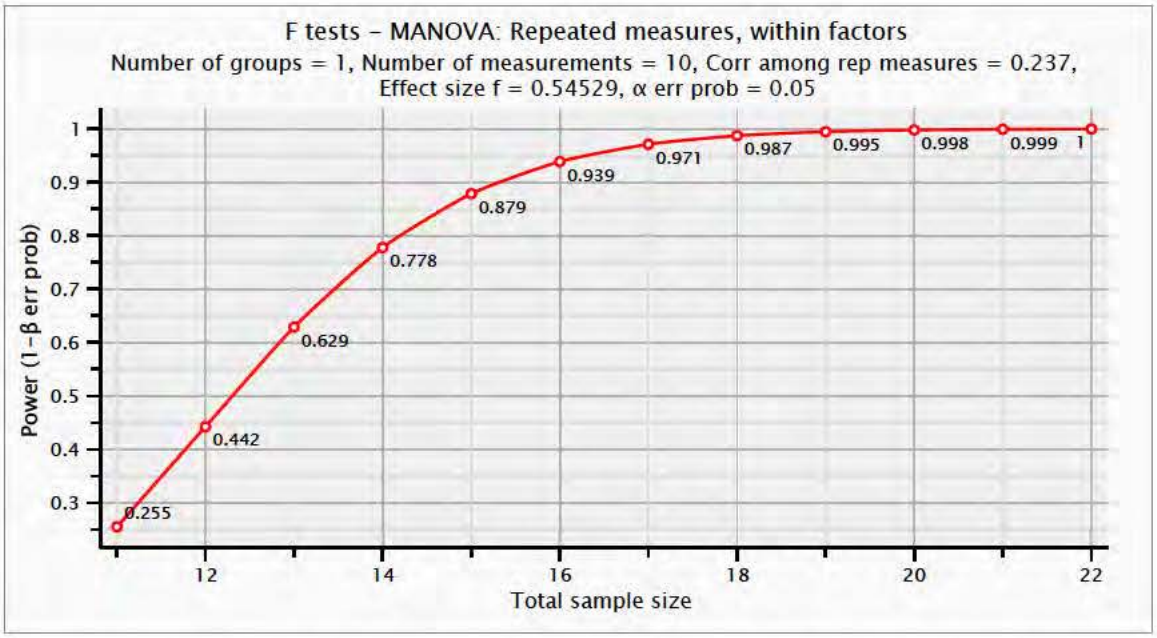


Abbildung 3.23.: Statistische Teststärke (Power) der MANOVA mit Messiederholung in Relation zur Stichprobengröße auf Grundlage der in dieser Studie ermittelten Effektgröße unter Verwendung z-transformierter Daten

3.8. Diskussion

3.8.1. Zusammenfassung und Interpretation der Ergebnisse

Die Ergebnisse der Laboruntersuchung zeigen, dass die 10 Raumsimulationen anhand der zwanzig Gegensatzpaare im Allgemeinen unterschieden werden konnten. Vor allem die Urteile bezüglich der Raumgröße und der wahrgenommenen Halligkeit fielen relativ homogen aus. Eine konsistente Bewertung über alle Items hinweg konnte allerdings durch Anwendung der Intraklassenkorrelation nicht beobachtet werden. Durch die Anwendung der Faktorenanalyse wurde gezeigt, dass die aufgrund hoher ICC(3,1)-Werte als konsistent zu bezeichnenden Items von den Versuchspersonen desweiteren sehr ähnlich bewertet wurden. Bei der Untersuchung der Zusammenhänge mit raumakustischen Parametern zeigten diese Items vor allem hohe Korrelationen mit der Nachhallzeit $T_{30\,1\mathrm{m}}$. Die zugrunde liegende Wahrnehmungsdimension lässt sich als *allgemein wahrgenommene Halligkeit*, *Raumgröße* oder im Hinblick auf den Zusammenhang mit der Nachhallzeit als *wahrgenommene Nachhalldauer* bezeichnen. Neben der Nachhallzeit zeigten sich vor allem beim späten Stärkemaß $G_{l\,1\mathrm{m}}$ hohe Korrelationen mit der wahrgenommenen Tragfähikeit (I1) sowie der Klangfülle (I14), wobei sich bei diesen Items auch mit der Nachhallzeit sehr ähnliche Zusammenhänge zeigten. Da I14 sowohl mit $G_{l\,1\mathrm{m}}$ als auch mit $G_{e\,1\mathrm{m}}$ und $G_{1\mathrm{m}}$ hoch korreliert, ist ein großer Zusammenhang der wahrgenommenen Klangfülle mit der Gesamtenergie der Raumreflexionen anzunehmen.

Der Wahrnehmungsaspekt der Unterstützung (I11), der in Untersuchungen zur Büh-

nenakustik einen besonderen Stellenwert einnimmt, weist die geringsten Korrelationen zur Nachhallzeit auf. Mit ST_{late} sowie $G_{l\,1\mathrm{m}}$ ergeben sich hingegen hohe Korrelationen, weshalb vor allem ein Einfluss der späten Reflexionen auf die wahrgenommene Unterstützung anzunehmen ist. Diese Beobachtung deckt sich mit den Ergebnissen von Ueno und Tachibana (2003), welche bei der Untersuchung von Solisten durch Musikerkommentare ebenso eine unterstützende Wirkung durch späte Raumreflexionen annahmen. Bei der Faktorenanlyse inklusive aller Items zeigte I11 zudem als einziges Item die höchsten Ladungen auf einen dritten Faktor. Dies weist daraufhin, dass für die Unterstützung neben der *allgemein wahrgenommenen Halligkeit* weitere Aspekte von Bedeutung sind.

ST_{early}, von Gade (1989a) zuerst als Prädiktor für die Unterstützung des eigenen Instruments vorgeschlagen, später zur Vorhersage der gegenseitigen Hörbarkeit im Ensemble genutzt (DIN EN ISO 3382-1:2009), zeigte zu keinem Item signifikante Korrelationen. Die von Gade (1989a, S. 197, Abb. 2) angegebenen Wahrnehmungsschwellen früher Reflexionen, welche ST_{early}-Werte angeben, wurden im Falle der Streicher durch sechs der zehn Raumsimulationen erreicht. Demzufolge müssten sie von den Musikern, sofern die Pegelkalibration[28] exakte Ergebnisse lieferte, wahrgenommen worden sein. Bei der wahrgenommenen Hörbarkeit des eigenen Instruments (I2) wurde ein größerer Zusammenhang zum frühen als zum späten Stärkemaß festgestellt. Auch aufgrund der geringen Korrelation der Items I2 und I11 (*unterstützend – nicht unterstützend*) von −0,26 zeigt sich, dass die wahrgenommene Unterstützung des Raums nicht mit der wahrgenommenen Hörbarkeit des eigenen Instruments in der Solistensituation gleichgesetzt werden darf. Die durch Musiker erfahrene Unterstützung im Ensemble, welche sich auf die Hörbarkeit des eigenen Instruments bezieht, ist somit grundlegend von der Unterstützung durch den Raums in der Solistensituation zu unterscheiden.

Auch wenn sich bei einigen Items eine geringe Urteilerkonsistenz ergab, sind diese keinesfalls als irrelevant zu bezeichnen. Da die Gegensatzpaare im RGT-Versuch durch die Musiker selbst eliziert und nicht durch die Versuchsleiter vorgegeben wurden, ist davon auszugehen, dass sie auf wahrgenommenen raumakustischen Phänomenen basieren. Mögliche Gründe für die inkonsistente Bewertung sind instrumentenspezifische Einflüsse, individuelle Wahrnehmungsunterschiede oder Unterschiede in der Deutung der semantischen Begriffe der Gegensatzpaare. Bezüglich der Präferenz, welche durch das Gegensatzpaar *gute Akustik – schlechte Akustik* (I6) abgefragt wurde, ist anzunehmen, dass keine Probleme bei der Interpretation auftraten. Das subjetktive Qualitätsurteil kann von Natur aus von unterschiedlichen physikalischen Phänomenen abhängen. Die beobachtete Heterogenität der Präferenzurteile ist daher entweder abhängig vom Individuum oder dem gespielten Instrument. Die Varianzquelle konnte in dieser Studie im Mehrebenenmodell leider nicht bestimmt werden. Obwohl durch die Auswahl der Raumsimulationen eine möglichst hohe Varianz der raumakustischen Parameter gegeben war, können sich bei manchen Items Schwierigkeiten

[28] Siehe Abschnitt 3.2.1

bei der Bewertung ergeben haben. Zum Beispiel kann vermutet werden, dass der Klangfarbenverlauf des Nachhalls (I20) bei Räumen mit sehr kurzer Nachhallzeit schwierig bewertet werden kann. Wenn sich die Versuchspersonen trotz des nicht wahr zu nehmenden Aspekts dazu verpflichtet fühlen, eine Bewertung von I20 abzugeben, fallen die Urteile zufällig aus. Eine sinnvolle Auswertung der erhobenen Daten ist in solch einem Fall bei Einbeziehung aller Räume nicht möglich.

Auch bei Bewertungen klangfarblicher Aspekte sowie der Verzögerung der Raumantwort (I4) zeigte sich eine sehr geringe Konsistenz der Bewertungen. Bei den klangfarblichen Items sind mehrere Erklärungen vorstellbar. Aufgrund der aus Sicht des Akustikers relativ unspezifischen Bedeutung der durch den RGT-Versush elizierten Klangfarbenitems, ist eine unterschiedliche Interpretation dieser gut vorstellbar. Beispielsweise könnte durch I9 (*dunkel – hell*) von einigen Musikern die durch frühe Reflexionen bedingte Klangfarbenänderung des Instrumentensignals und von anderen die Klangfarbe des späten Nachhalls bewertet worden sein. Auch eine unterschiedliche Interpretation von Raum zu Raum durch die selbe Versuchsperson ist nicht auszuschließen. Eine exakte Unterscheidung von spektraler und zeitlicher Klangfärbung (vgl. Rubak, 2004) hätte den Deutungsspielraum möglicherweise verringern können.

Ferner kann ein Einfluss des Faktors Instrument auf die Wahrnehmung klangfarblicher Aspekte angenommen werden. Durch die Anwendung eines Mehrebenenmodells, bei welchem die auf die Instrumentengruppen der Streicher, Holzbläser und Blechbläser zurückzuführende Varianz berücksichtigt wurde, ergab sich eine deutlich höhere Intraklassenkorrelation bei der Bewertung der Klangfarbenfluktuation (I19). Bei der Wahrnehmung der spektralen Klangfärbung, auf welche vor allem frühe Reflexionen Einfluss haben, spielt die Maskierung durch den Direktschall eine entscheidende Rolle. Diese hängt aufgrund der Unterschiede der Dynamik, des Spektrums, der Richtcharakteristik sowie der Ein- und Ausschwingvorgänge von den Instrumenten ab. Auf Basis der gewählten Stichprobengröße konnte der Einfluss der Instrumente bezüglich der spektralen Klangfärbung statistisch allerdings nicht nachgewiesen werden. Die Anwendung des Mehrebenenmodells kann jedoch für zukünftige Studien eine beispielhafte Methode darstellen, wie der Einfluss der Instrumente auf die Bewertung der raumakustischen Umgebungen berücksichtigt werden kann. Die geringe Konsistenz der Bewertungen von I4 (*Raum spricht früh an – ... spät an*) ist ebenso auf die unterschiedlichen Ausprägungen der Maskierungseffekte bei verschiedenen Instrumenten zurückzuführen.

Die Frage nach der Anzahl der Dimensionen, welche der raumakustischen Wahrnehmung durch Musiker zugrunde liegen, kann durch die Ergebnisse dieser Studie nicht eindeutig beantwortet werden. Es zeigte sich, dass zumindest eine Dimension von den Musikern relativ konsistent bewertet werden konnte. Dieser sind die Items der wahrgenommenen Halligkeit, der Tragfähigkeit, der Balance zwischen Raum und Instrument, der Lautstärke, der Klarheit, der Klangfülle sowie der Raumgröße zuzuordnen, es konnte ein großer Zusammenhang der betreffenden Items zur Nachhallzeit $T_{30\,1\mathrm{m}}$ beobachtet werden. Die Unterstützung ist

diesem Faktor nicht zuzuordnen. Aufgrund des geringen ICC(3,1) von 0,108 ist eine statistische Auswertung anhand der Mittelwerte jedoch problematisch. Die Korrelationen zu ST_{late} erscheinen aufgrund anderer Studien (Ueno und Tachibana, 2003) jedoch sinnvoll. Der zweite Faktor, welcher sich bei Berücksichtigung aller Items ergab und vor allem durch hohe Ladungen klangfarblicher Items geprägt ist, kann aufgrund der geringen Konsistenz der entsprechenden Items nicht als Faktor *Klangfarbe* interpretiert werden. Dieses Ergebnis würde sich zwar einerseits mit den Ergebnissen von Gade (1989b) decken, bei welchem sich bei Befragung zweier Orchester neben dem Faktor *Overall Acoustic Impression* ebenso ein Faktor *Timbre* ergab. Andererseits kann nicht davon ausgegangen werden, dass durch nicht reliablen bzw. inkonsistenten Ratings wirklich etwas gemessen wurde, was mit den dargebotenen raumakustischen Simulationen in Zusammenhang steht (vgl. Shrout und Fleiss, 1979, S. 427).

Durch die retrospektive Analyse der Teststärken unter Einbeziehung der beobachteten Effektgrößen aus den Varianzanalysen konnte gezeigt werden, dass für die Untersuchung von Items, welche nicht dem Faktor der *allgemein wahrgenommenen Halligkeit* zuzuordnen sind, deutlich größere Stichproben benötigt werden. Angenommen, die in dieser Studie beobachteten Effektgrößen entsprächen denen der Gesamtpopulation, war die Möglichkeit eines signifikanten Effekts bei der gewählten Stichprobengröße von $n = 14$ vor allem bei den klangfarblichen Aspekte nicht gegeben.

3.8.2. Technische Fehlerquellen

Neben den bereits diskutierten Einflüssen der verschiedenen Instrumente und der Interpretation der Gegensatzpaare, können auch im technischen Versuchsaufbau bzw. der Simulationsmethode mögliche Fehlerquellen beinhaltet sein. Bei der Erzeugung der BRIR-Sets wurden aus Gründen der Praktikabilität keine individuellen, sondern durch ein Kunstkopf-Torso-System gemessene HRTFs verwendet. Aufgrund der zusätzlichen Entzerrung der Endstufe, des Kopfhörers sowie des Außenohres, kann diese Fehlerquelle jedoch im Vergleich zu anderen als relativ unbedeutend eingeschätzt werden. Die Verwendung des am Instrument angebrachten Mikrofons hatte zum Vorteil, dass es während des Spiels nicht zu Änderungen des Abstands zwischen Instrument und Mikrofon und damit verbundenen Pegelschwankungen kommen konnte. Gleichzeitig zeigte sich das verwendete Mikrofon mit der verwendeten Aufhängung anfällig für die Übertragung von Körperschall, welcher somit auch verhallt wurde. Selbst wenn die Musiker mögliches auftretendes „Rumpeln“ als technische Unzulänglichkeit der Simulation interpretierten, welche sie versuchten nicht in ihren Bewertungen zu berücksichtigten, kann dieses zu Maskierungen tiefer Frequenzen der restlichen Raumsimulation führen.

4. Ausblick

Bei der Untersuchung der raumakustischen Wahrnehmung durch Solisten zeigte sich eine starke Tendenz zu einer eindimensionalen Bewertung von Konzertsälen, welche durch dynamische Binauralsynthese auralisiert wurden. Bei Items, welche dem gefundenen Faktor der *allgemein wahrgenommenen Halligkeit* zuzuordnen waren, zeigten sich hohe lineare Zusammenhänge mit der auf der Bühne in 1 m Entfernung zur Quelle gemessenen Nachhallzeit $T_{30\,1\mathrm{m}}$. Durch die geringe Stichprobengröße sowie die unterschiedlichen untersuchten Instrumentalisten stellte sich die Interpretation und Auswertung weiterer möglicher Dimensionen als problematisch dar. Durch zukünftige Studien sollten vor allem die Aspekte der wahrgenommenen Klangfarbe des Raums sowie der von Solisten erfahrenen Unterstützung, bei welcher sich ST_{late} als geeigneter Prädiktor anbot, intensiver untersucht werden. Aufgrund der in dieser Studie teilweise sehr inhomogenen Urteile, ist in jedem Fall die Angabe und Interpretation der Intraklassenkorrelation in zukünftigen Untersuchungen zu empfehlen. Die Auswertung der Mittelwerte über alle Rater ohne Berücksichtigung der Reliabilität kann nämlich, wie gezeigt wurde, zu falschen Schlüssen führen.

Um eine möglichst hohe Urteilerkonsistenz zu erreichen, gilt es, die Erfahrungen dieser Studie bei der Planung der Untersuchungsmethode zu berücksichtigen. Die Repertory-Grid-Technik erwies sich als geeignete Methode zur Elizierung subjektiver Konstrukte. Die Methode der Auswahl und Definition der Bewertungsitems, welche für alle Musiker verständlich sein sollten, sollte in Zukunft allerdings erweitert werden. Eine gemeinsame Diskussionsrunde mit den Musikern, in welcher im Anschluss an den RGT-Versuch ein allgemeiner Fragebogen zusammengestellt wird, wäre vorstellbar. Um Missverständnisse aufgrund unterschiedlicher Deutungen semantischer Begriffe zu verringern oder gar ganz auszuschließen, sollte solch eine Diskussion ferner unter Berücksichtigung der Objekte selbst, nämlich verschiedener raumakustischer Simulationen, statt finden. Eine relativ unaufwändige Variante stellt das gemeinsame Anhören einer musikalischen Darbietung aus Musikerperspektive über Kopfhörer dar, bei welcher die zur Faltung genutzten BRIRs variiert werden können. Schwieriger, jedoch in seiner Vaildität höher einzuschätzen, wäre ein gemeinsamer Vergleich der raumakustischen Umgebungen bei gleichzeitigem Spiel der Musiker. Mithilfe einer dieser Vorgehensweisen könnte genauer bestimmt werden, ob durch ein Item beispielsweise die Änderung der Klangfarbe des eigenen Instruments oder die Variation der Klangfarbe des späten Nachhalls beurteilt werden soll. Auch wenn solch eine Diskussionsrunde nicht möglich ist, sollten in jedem Fall neben den Ratings auch qualitative Daten, beispielsweise durch Interviews, erhoben werden. Somit könnte die Interpretation der Items durch die Musiker

nicht nur aus den Raumbewertungen, sondern auch aus den persönlichen Berichten abgeleitet werden. Auch der musikalische Hintergrund sollte genauer abgefragt werden. Ob die Musiker für gewöhnlich im Orchester, einem Kammerensemble oder Solo auftreten, könnte auch einen Einfluss auf deren Raumakustikbewertung haben. Die auftretende Fehlervarianz ließe sich dementsprechend verringern.

In dieser Studie wurde durch die Anwendung der Mehrebenenanalyse eine statistische Auswertungsmethode vorgestellt, mithilfe welcher die Zugehörigkeit der Musiker zu Instrumentengruppen berücksichtigt werden kann. Aufgrund der geringern Anzahl der Versuchspersonen ließ sich diese jedoch bei manchen Items nicht berücksichtigen. Sofern es die erhobenen Daten zulassen, könnte in zukünftigen Studien durch die Aufnahme raumakustischer Parameter in ein Random-Intercept-Modell der Instrumenteneinfluss auch bei der Überprüfung der Zusammenhänge von Items und möglichen Prädiktoren berücksichtigt werden.

Von großem Interesse wäre desweiteren eine Studie zur Retest-Reliabilität bzw. der intraindividuellen Konsistenz der Urteile von Musikern. Dabei könnte die grundlegende Fragestellung untersucht werden, inwieweit Musiker dazu in der Lage sind, die Raumakustik bezüglich der von ihnen genannten Gegensatzpaare auch konsistent zu beurteilen. Im Zuge solch eines Versuchs ließe sich desweiteren der Einfluss des gespielten Musikexzerpts auf die Bewertung überprüfen, indem man dieses variiert. Durch die Abfrage von Multiinstrumentalisten, welche die gleichen Räume bei Anregung mit unterschiedlichen Instrumenten bewerten, könnte sogar der Faktor Instrument losgelöst vom Faktor Indidviduum untersucht werden.

Abbildungsverzeichnis

2.1. Gefundene Wahrnehmungsschwelle einer Deckenreflexion des eigenen Instruments in Abhängigkeit von der Verzögerungszeit für Streicher und Flötisten (Gade, 1989a). 8
2.2. Kognitives Modell des impliziten Wissens bei der Musikaufführung (Ueno und Tachibana, 2005). 13

3.2. Technischer Versuchsaufbau 24
3.3. Schematische Darstellung der Pegelkalibrierung 25
3.4. 3D-Modelle und Grundrisse der architektonisch unterschiedlichen Räume mit Markierung der Musikerposition 29
3.5. Schroeder-Plot – R4 Konzertsaal1 35
3.6. Scatterplots raumakustischer Parameter der auralisierten Räume 36
3.7. Bedienoberfläche, mit welcher der Versuchsleiter die Konstrukte während des RGT-Versuchs notierte 42
3.8. Dendrogramm der Clusteranalyse der von M5 (Cello) elizierten und bewerteten Gegensatzpaare 45
3.9. Dendrogramm der Clusteranalyse der von M6 (Trompete) elizierten und bewerteten Gegensatzpaare 46
3.10. Dendrogramm der Clusteranalyse der von M7 (Fagott) elizierten und bewerteten Gegensatzpaare 47
3.11. Dendrogramm der Clusteranalyse der von M9 (Posaune) elizierten und bewerteten Gegensatzpaare 48
3.12. Dendrogramm der Clusteranalyse der von M11 (Violine) elizierten und bewerteten Gegensatzpaare 49
3.13. Dendrogramm der Clusteranalyse der von M14 (Oboe) elizierten und bewerteten Gegensatzpaare 50
3.14. Struktogramme der individuellen Clusteranalysen 51
3.15. Dendrogramm einer Clusteranalyse anhand der Bewertungen der individuellen Konstrukte aller Probanden 52
3.16. Struktogramm der Clusteranalyse über die gebildeten Konstrukte aller Probanden. 53
3.17. Bedienoberfläche des Ratingversuchs mit beispielhafter Itemreihenfolge . . . 54
3.18. Boxplots der z-transformierten Ratings 59

3.19. Clusteranalysen über Musiker mit unterschiedlicher Einbeziehung der Itembewertungen 74
3.20. Biplot aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14 . . 80
3.21. Eigenwertdiagramm aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14 81
3.22. Nachhallzeit $T_{30\,1\mathrm{m}}$ aufgetragen zu den standardisierten Bewertungen der Raumsimulationen bezüglich I18 (*ausgewogen – unausgewogen*) 86
3.23. Statistische Teststärke (Power) der MANOVA mit Messiederholung in Relation zur Stichprobengröße auf Grundlage der in dieser Studie ermittelten Effektgröße unter Verwendung z-transformierter Daten 92

A.1. Erklärung der 20 zu bewertenden Gegensatzpaare, welche den Probanden vorgelegt wurde 108

C.2. Schroeder-Plots der zur Berechnung der raumakustischen Parameter verwendeten Impulsantworten 116

Tabellenverzeichnis

2.1. Subjektive Aspekte bei der Bewertung der Raumakustik im Auditorium . . 17

3.1. Verwendete Raumakustikmodelle 31
3.2. Raytracing-Optionen in *EASE Aura* für alle Modelle 31
3.3. Definitionen der in dieser Arbeit verwendeten raumakustischen Parameter . 33
3.4. Instrumentenpositionen im Modell relativ zum Musiker in m 38
3.5. Informationen zur Stichprobe bezüglich Geschlecht, gespieltem Instrument, Alter, Konzerterfahrung und den gespielten musikalischen Passagen 39
3.6. Auswahl der Ratingitems für den Hauptversuch 55
3.7. Teststatistiken des Innersubjektfaktors Raum aus der MANOVA mit Messwiderholungen über alle Items I1 –I 20 mit z-transformierten Daten 64
3.8. Korrigierte Trennschärfen der Musikerurteile je Item 70
3.9. Intraklassenkorrelationskoeffizienten und ANOVA-Ergebnisse der 20 Items . 72
3.10. Um systematische Instrumentenvarianz bereinigte Intraklassenkorrelation ICC(3,1)$_{Instr}$ im Vergleich zu ICC(3,1) 77
3.11. Korrelationen zwischen Bewertungsitems 79
3.12. Faktorladungen a_{ij} und Kommunalitäten h^2 aus Faktoranalyse mit den Items I1, I3, I5, I7, I8, I10 und I12–I14 82
3.13. Korrelationen nach Pearson zwischen akustischen Parametern der Raumsimulationen und Bewertungsitems 85
3.14. Ergebnisse der retrospektiven Poweranalysen der ANOVAs je Item 91

C.1. $T_{30\,1\mathrm{m}}$ in s 118
C.2. $EDT_{1\mathrm{m}}$ in s 118
C.3. $G_{1\mathrm{m}}$ in dB 119
C.4. $C_{50\,1\mathrm{m}}$ in dB 119
C.5. $C_{80\,1\mathrm{m}}$ in dB 119
C.6. $T_{C\,1\mathrm{m}}$ in ms 120
C.7. ST_{early} in dB 120
C.8. ST_{total} in dB 120
C.9. ST_{late} in dB 121
C.10. LQ_{7-40} in dB 121
C.11. $G_{e\,1\mathrm{m}}$ in dB 121

C.12. $G_{l\,1\mathrm{m}}$ in dB 122
C.13. Verhältnismaße aus den angegebenen Parametern 122

D.1. ICC(3,1) 127
D.2. ICC(2,1) 128
D.3. Ergebnisse aus zweifaktoriellen ANOVAs je Item 129
D.4. Ergebnisse der Tukey-Tests 130

Literaturverzeichnis

Ahnert, W. und Tennhardt, H.-P. (2008). „Raumakustik“. In: S. Weinzierl, Hrsg., *Handbuch der Audiotechnik*. Springer. Berlin et al. Kapitel 5, S. 181–266.

Anderberg, M. R. (1971). *Cluster analysis for applications*. Dissertation. The University of Texas at Austin.

Asendorpf, J. und Wallbott, H. G. (1979). „Maße der Beobachtungsübereinstimmung: Ein systematischer Vergleich“. In: *Zeitschrift für Sozialpsychologie* 10, S. 243–252.

Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. (2011). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. 13., überarb. Auflage. Springer. Berlin et al.

Barron, M. (1978). „The Gulbenkian Great Hall, Lisbon, II: An acoustic study of a concert hall with variable stage“. In: *Journal of Sound and Vibration* 59(4), S. 481–502.

Barron, M. (2010). *Auditorium Acoustics and Architectural Design*. 2. Auflage. Spon Press. London, New York.

Bech, S. und Zacharov, N. (2006). *Perceptual Audio Evaluation - Theory, method and Application*. Wiley & Sons. Chichester, England et al.

Benninghaus, H. (1998). *Einführung in die sozialwissenschaftliche Datenanalyse*. 5., vollst. überarb. Auflage. Oldenbourg. München et al.

Beranek, L. (2004). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. 2. Auflage. Springer. New York et al.

Beranek, L. L. (1962). *Music, Acoustics & Architecture*. Wiley. New York et al.

Berg, J. (2006). „How do we determine the attribute scales and questions that we should ask of subjects when evaluating spatial audio quality?“. In: *Proc Int. Workshop on Spatial Audio and Sensory Evaluation Techniques*. Guilford.

Berg, J. und Rumsey, F. (1999). „Identification of perceived spatial attributes of recordings by repertory grid technique and other methods“. In: *Audio Engineering Society Convention 106*. München.

Berg, J. und Rumsey, F. (2000). „In search of the spatial dimensions of reproduced sound: Verbal protocol analysis and cluster analysis of scaled verbal descriptors“. In: *Proceedings of AES 108th Convention*. Paris.

Berg, J. und Rumsey, F. (2006). „Identification of qualaty attributes of spatial audio by repertory grid technique“. In: *Acustica* 54(5), S. 365–379.

Bortz, J. und Döring, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler*. 4., überarb. Auflage. Springer. Heidelberg.

Bortz, J. und Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. 7., vollst. überarb. u. erweiterte Auflage. Springer. Berlin et al.

Bradley, J. S. (2011). „Review of objective room acoustics measures and future needs“. In: *Applied Acoustics* 72, S. 713–720.

Bray, J. H. und Maxwell, S. E. (1985). *Multivariate analysis of variance.* Sage. Beverly Hills et al.

Brinkmann, F. (2011). *Individual headphone compensation for binaural synthesis.* Magisterarbeit. Technische Universität Berlin.

Brinkmann, F. und Lindau, A. (2010). „On the effect of individual headphone compensation in binaural synthesis“. In: *Fortschritte der Akustik: Tagungsband d. 36. DAGA.* S. 1055–1056.

Buchner, A., Erdfelder, E. und Faul, F. (1996). „Teststärkeanalysen“. In: E. Erdfelder, R. Mausfeld, T. Meiser und G. Rudinger, Hrsg., *Handbuch Quantitative Methoden.* Psychologie Verlags Union. Weinheim. S. 123–136.

Chiang, W., Chen, S. und Huang, C. (2003). „Subjective assessment of stage acoustics for solo and chamber music performances“. In: *Acta Acustica united with Acustica* 89, S. 848–856.

Chiang, W.-h. (1994). *Effects of Various Architectural Parameters on Six Room Acoustical Measures in Auditoria.* Dissertation. University of Florida.

Choisel, S. und Wickelmaier, F. (2006). „Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound“. In: *J. Audio Eng. Soc.* 54(9), S. 815–826.

Dammerud, J. J. (2009). *Stage acoustics for symphony orchestras on concert halls.* Dissertation. University of Bath.

Danielsson, M. (1991). „Repertory Grid Technique“. Forschungsbericht 1991:23. Luleå University of Technology.

DIN 18041:2004-05 (2004). „Hörsamkeit in kleinen bis mittelgroßen Räumen“. Beuth-Verlag, Berlin.

DIN EN ISO 3382-1:2009 (2009). „Akustik – Messung von Parametern der Raumakustik – Teil 1: Aufführungsräume (ISO 3382-1:2009); Deutsche Fassung EN ISO 3382-1:2009“. Beuth-Verlag, Berlin.

Eid, M., Gollwitzer, M. und Schmitt, M. (2011). *Statistik und Forschungsmethoden.* 2., korr. Auflage. Beltz. Weinheim et al.

Faul, F., Erdfelder, E., Lang, A.-G. und Buchner, A. (2007). „G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences“. In: *Behavior Research Methods* 39(2), S. 175–191.

Gade, A. C. (1989a). „Musicians' room acoustic conditions in concert halls I : Methods and laboratory experiments“. In: *Acustica* 69, S. 193–203.

Gade, A. C. (1989b). „Musicians' room acoustic conditions in concert halls II: Methods and laboratory experiments“. In: *Acustica* 69, S. 249–262.

Gade, A. C. (1992). „Practical aspects of room acoustic measurements on orchestra platforms“. In: *Proc. 14th ICA.* Paper F3-5. Peking.

Gade, A. C. (2007). „Acoustics in halls for speech and music“. In: T. D. Rossing, Hrsg., *Springer Handbook of Acoustics.* Springer. S. 301–350.

Gade, A. C. (2010). „Acoustics for symphony orchestras; status after three decades of experimental research“. In: *Proc. of the ISRA 2010*. Melbourne.

Handl, A. (2002). *Multivariate Analysemethoden – Theorie und Praxis multivariater Verfahren unter besonderer Berücksichtigung von S-PLUS*. Springer. Berlin et al.

Hawkes, R. J. und Douglas, H. (1971). „Subjective acoustic experience in concert auditoria“. In: *Acustica* 24, S. 235–250.

Hox, J. (2010). *Multilevel Analysis: Techniques and Applications*. 2. Auflage. Routledge. New York.

Kelly, G. A. (1986). *Die Psychologie der persönlichen Konstrukte*. Junfermann-Verlag. Paderborn.

Kim, K. und Timm, N. (2007). *Univariate and Multivariate General Linear Models – Theory and Applications with SAS*. 2. Auflage. Chapman & Hall. Boca Raton et al.

Kuttruff, H. (2009). *Room Acoustics*. 5. Auflage. Spon Press. London, New York.

Lehmann, P. und Wilkens, H. (1980). „Zusammenhang subjektiver Beurteilungen von Konzertsälen mit raumakustischen Kriterien“. In: *Acustica* 45, S. 256–268.

Lindau, A., Hohn, T. und Weinzierl, S. (2007). „Binaural resynthesis for comparative studies of acoustical environments“. In: *Audio Engineering Society Convention 122*. Wien.

Lindau, A., Kosanke, L. und Weinzierl, S. (2010). „Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses“. In: *Proc. of the 128th AES Convention*. London.

Lindau, A. und Brinkmann, F. (2010). „Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings“. In: *Proc. of the 3rd Third International Workshop on Perceptual Quality of Systems*. Dresden.

Lokki, T. und Järveläinen, H. (2001). „Subjective evaluation of auralization of physics-based room acoustic modeling“. In: *Proc. Int. Conf. on Auditory Display*. Espoo, Finnland. S. 26–31.

Marketo, C. G., Cooper, T., Petty, M. F. und Scriven, F. M. (1994). „The reliability of mdpref to show individual preference“. In: *Journal of Sensory Studies* 9, S. 337–350.

Marshall, A. H., Gottlob, D. und Alrutz, H. (1978). „Acoustical conditions preferred for ensemble“. In: *J. Acoust. Soc. Am.* 64(5), S. 1437–1442.

McGraw, K. O. und Wong, S. P. (1996). „Forming inferences about some intraclass correlation coefficients“. In: *Psychological Methods* 1(1), S. 30–46.

Meyer, J. (2009). *Acoustics and the Performance of Music*. 5. Auflage. Springer.

Moldrzyk, C., Ahnert, W., Feistel, S., Lentz, T. und Weinzierl, S. (2004). „Head-tracked auralization of acoustical simulation“. Presented at the 117th Convention of the Audio Engineering Society, San Francisco.

Nunnally, J. C. (1978). *Psychometric Theory*. 2. Auflage. McGraw-Hill. New York u.a.

O'Brien, R. G. und Kaiser, M. K. (1985). „Manova method for analyzing repeated measures designs: An extensive primer“. In: *Psychological Bulletin* 97(2), S. 316–333.

Osgood, C. E., Suci, G. J. und Tannenbaum, P. H. (1967). *The Measurement of Meaning*. University of Illinois Press. Urbana et al.

Riemann, H. (1959). *Riemann Musik Lexikon. Personenteil A-K.* 12., völlig neubearb. Auflage. Schott. Mainz et al.

Röhrbein, M. (2011). *Perzeptiv motivierte Reduktion von Schallfeldbeschreibungen für eine Resynthese in Echtzeit.* Masterarbeit. Technische Universität Berlin.

Rubak, P. (2004). „Coloration in room impulse responses". In: *Proc. Joint Baltic-Nordic Acoustics Meeting.* Mariehamn, Åland. S. 1–14.

Sanders, J. (2003). „Suitability of new zealand halls for chamber music". Online unter: http://marshallday.com/sites/marshallday.com/files/9_Suitability_of_New_Zealand_halls_for_chamber_music_%28c%29_Marshall_Day_Acoustics_1997,2003.pdf.

Schärer Kalkandjiev, Z. und Weinzierl, S. (2013). „The influence of room acoustic on solo music performance: An empirical case study". In: *Acta Acustica united with Acustica* 99, S. 433–441.

Schärer, Z. und Lindau, A. (2008). „Evaluation von Verfahren zur Entzerrung von binauralen Signalen". In: *Bericht der 25. Tonmeistertagung.* Bildungswerk des VDT. Leipzig. S. 515–532.

Scheibler, D. und Schneider, W. (1985). „Monte carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods". In: *Multivariate Behavioral Research* 20(3), S. 283–304.

Schneider, R. (2011). *Simulation einer musikalischen Aufführung durch die Auralisation von Musikinstrumenten mit korrekter Richtcharakteristik.* Magisterarbeit. Technische Universität Berlin.

Schroeder, M. R., Gottlob, D. und Siebrasse, K. F. (1974). „Comparative study of european concert halls: correlation of subjective preference with geometric and acoustic parameters". In: *J. Acoust. Soc. Am.* 56(4), S. 1195–1201.

Shrout, P. E. und Fleiss, J. L. (1979). „Intraclass correlations: Uses in assessing rater reliability". In: *Psychological Bulletin* 86(2), S. 420–428.

Siebein, G. W. und Kinzey, Jr., B. Y. (1999). „Recent innovations in acoustical design and research". In: W. J. Cavanaugh und J. A. Wilkes, Hrsg., *Architectural Acoustics. Principles and Practice.* Wiley & Sons. New York et al. Kapitel 6, S. 233–304.

Slavik, K. M. und Weinzierl, S. (2008). „Wiedergabeverfahren". In: S. Weinzierl, Hrsg., *Handbuch der Audiotechnik.* Springer. Berlin et al. Kapitel 11, S. 609–685.

Thomas, L. (1997). „Retrospective power analysis". In: *Conservation Biology* 11(1), S. 276–280.

Ueno, K. und Tachibana, H. (2003). „Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system". In: *Acoustical science and technology* 24(3), S. 130–138.

Ueno, K. und Tachibana, H. (2005). „Cognitive modeling of musician's perception in concert halls". In: *Acoustical science and technology* 26(2), S. 156–161.

Ueno, K. und Tachibana, H. (2010). „A consideration on acoustic properties on concert-hall stages on concert-hall stages". In: *Proc. of the ISRA 2010.* Melbourne.

van den Braak, E. W. M., Hak, C. C. J. M. und van Luxemburg, L. C. J. (2009). „New (stage) parameter for conductor's acoustics? - part 2: further investigation on LQ_{7-40}". In: *Proc of inter-noise.* Ottawa.

van den Braak, E. W. M. und van Luxemburg, L. C. J. (2008). „New (stage) parameter for conductor's acoustics?“. In: *Proc. of Aoustics '08 Paris*. Paris.

Weir, J. P. (2005). „Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM“. In: *Journal of Strength and Conditioning Research* 19(1), S. 231–240.

Wilkens, H. (1977). „Mehrdimensionale Beschreibung subjektiver Beurteilungen der Akustik von Konzertsälen“. In: *Acustica* 38, S. 10–23.

Wirtz, M. und Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreabilität*. Hogrefe. Göttingen.

Datenblätter und Handbücher

Acoustic Design Ahnert. EASE Version 4.3. User's Guide & Tutorial (2009).
Online unter: http://www.renkus-heinz.com/support/software-support/ease_support/tutorials/tutorials_download/EASE4.3Tutorial/EASE_4.3_Tutorial.pdf.

DPA Microphones. d:vote™ 4099. User's manual.
Online unter: http://www.dpamicrophones.com/~/media/PDF/Download/Users%20Manuals/4099.pdf.

IBM. SPSS Advanced Statistics 21 (2012).
Online unter: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/de/client/Manuals/IBM_SPSS_Advanced_Statistics.pdf.

Reflexionsarmer Raum, Institut für Strömungsmechanik und Technische Akustik.
Spezifikationen online unter: http://www.akustik.tu-berlin.de/fileadmin/fg23/Sonstiges/Technische_Einrichtungen/Anechoic_Room.pdf.

Sony. Kondensatormikrofon C48 (1980). Bedienungsanleitung.
Online unter: http://www.sunsetwind.com/Sony_C48/C48_Manual.pdf.

Verwendete Software

EASE 4.3. ADA Acoustic Design Ahnert / AFMG Technologies.

fWonder. Fachgebiet Audiokommunikation. TU Berlin.

G*Power 3.1.5.
Online unter: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3

Matlab R2009b. The MathWorks.

Matlab-Skript-Paket: BRIR-Export. Fachgebiet Audiokommunikation. TU Berlin. Siehe Röhrbein (2011).

Matlab-Skript: anova_rm.m. Salarian, A. (2008).
Online unter: http://www.mathworks.com/matlabcentral/fileexchange/22088

Matlab-Skript: ICC.m, Salarian, A. (2008).
Online unter: http://www.mathworks.com/matlabcentral/fileexchange/22099

Matlab-Skript: kmo.m. Trujillo-Ortiz, A., Hernandez-Walls, R. , Castro-Perez, A., Barba-Rojo, K. und Otero-Limon, A. (2006).
Online unter: http://www.mathworks.com/matlabcentral/fileexchange/12736

Matlab-Skript: matlab2tikz.m. Schlömer, N. (2012).
http://www.mathworks.com/matlabcentral/fileexchange/22022

SPSS Statistics 21. IBM.

Alle im Literaturverzechnis angegebenen Internetquellen wurden am 19. Juni 2013 auf ihre Richtigkeit überprüft.

A. Versuchsanleitung

Lieber Versuchsteilnehmer,
vielen Dank, dass Sie an dieser Studie teilnehmen.

Sie werden nacheinander in virtuellen Räumen spielen, die Sie jeweils in Bezug auf die unten gelisteten Eigenschaften bewerten sollen. Bitte lesen Sie sich die Beschreibungen sorgfältig durch und fragen Sie bei Unklarheiten.

Gegensatzpaar		**Beschreibung**
tragfähig	wenig tragfähig	Trägt der Raum das Spiel oder nicht?
ich kann mich gut hören	ich kann mich nicht gut hören	Können Sie Ihr Instrument gut hören?
trocken	hallig	Hat der Raum einen langen oder kurzen Nachhall?
Raum spricht früh an	Raum spricht spät an	Kommt der Raumklang gleich oder verzögert?
groß	klein	Handelt es sich um einen großen oder kleinen Raum?
gute Akustik	schlechte Akustik	Gefällt die Akustik oder nicht?
hoher Raum	niedriger Raum	Wirkt der Raum niedrig oder hoch?
breiter Raum	schmaler Raum	Wirkt der Raum breit oder schmal?
dunkel	hell	Wie ist die Klangfarbe des Raums, des Nachhalls?
viel Raum	wenig Raum	Wie ist die Balance zwischen Instrument und Raumklang?
unterstützend	nicht unterstützend	Hilft der Raum beim Spielen?
laut	leise	Nehmen Sie Ihr Instrument laut oder leise wahr?
diffus	klar	Ist der Raumklang durchmischt oder klar?
dünn	voll	Hat der Raum einen dünnen oder vollen Klang?
weich	hart	Wirkt der Raum weich oder hart?
rund	metallisch	Hat der Raum einen dumpfen oder eher einen metallischen Klang?
Nachhall klingt weich aus	Nachhall bricht ab	Wie klingt der Nachhall aus?
ausgewogen	unausgewogen	Bietet der Raum ein in sich stimmiges Gesamtbild?
Klangfarbe ist gleichmäßig	Klangfarbe fluktuiert	Bleibt die Klangfarbe im Nachklang gleichmäßig oder schwankt sie?
Nachklang wird schärfer	Nachklang wird dumpfer	Wie ändert der Nachklang im Verlauf seine Klangfarbe?

Abbildung A.1.: Erklärung der 20 zu bewertenden Gegensatzpaare, welche den Probanden vorgelegt wurde

gelisteten Eigenschaften bewerten sollen. Bitte lesen Sie sich die Beschreibungen sorgfältig durch und fragen Sie bei Unklarheiten.

Gegensatzpaar		Beschreibung
tragfähig	wenig tragfähig	Trägt der Raum das Spiel oder nicht?
ich kann mich gut hören	ich kann mich nicht gut hören	Können Sie Ihr Instrument gut hören?
trocken	hallig	Hat der Raum einen langen oder kurzen Nachhall?
Raum spricht früh an	Raum spricht spät an	Kommt der Raumklang gleich oder verzögert?
groß	klein	Handelt es sich um einen großen oder kleinen Raum?
gute Akustik	schlechte Akustik	Gefällt die Akustik oder nicht?
hoher Raum	niedriger Raum	Wirkt der Raum niedrig oder hoch?
breiter Raum	schmaler Raum	Wirkt der Raum breit oder schmal?
dunkel	hell	Wie ist die Klangfarbe des Raums, des Nachhalls?
viel Raum	wenig Raum	Wie ist die Balance zwischen Instrument und Raumklang?
unterstützend	nicht unterstützend	Hilft der Raum beim Spielen?
laut	leise	Nehmen Sie Ihr Instrument laut oder leise wahr?
diffus	klar	Ist der Raumklang durchmischt oder klar?
dünn	voll	Hat der Raum einen dünnen oder vollen Klang?
weich	hart	Wirkt der Raum weich oder hart?
rund	metallisch	Hat der Raum einen dumpfen oder eher einen metallischen Klang?
Nachhall klingt weich aus	Nachhall bricht ab	Wie klingt der Nachhall aus?
ausgewogen	unausgewogen	Bietet der Raum ein in sich stimmiges Gesamtbild?
Klangfarbe ist gleichmäßig	Klangfarbe fluktuiert	Bleibt die Klangfarbe im Nachklang gleichmäßig oder schwankt sie?
Nachklang wird schärfer	Nachklang wird dumpfer	Wie ändert der Nachklang im Verlauf seine Klangfarbe?

Abbildung A.1.: (Fortsetzung) Erklärung der 20 zu bewertenden Gegensatzpaare, welche den Probanden vorgelegt wurde

B. Quellcode

Berechnung raumakustischer Parameter – AcousticMeasuresFromIR.m

```
%% calculation of T30, EDT, C50, C80, Tc, LQ 7-40, SP1, SP2, SP_late etc. with Impulse response

clear all; close all; clc;
%% wavread of IR
filename = '10Theater_4mio_1000_S1';
[h fs] = wavread(['Data\',filename,'.wav']);
h = h(:,2); %we need only one channel

% order of octave filters
order = 4;
% factor to multiply with group delay
fac = 0.5;
doplots = 1;

%% wavread of IR with 10 m distance between source and receiver in free field for G Measure
h10mff = wavread('Data\10mff.wav');
h10mff = h10mff.*0.1;% because of normalization of wav-Export in EASE, -20d
h10mff = h10mff(:,2); %we need only one channel
%% Find beginning of IR
thr = 30; % threshold of beginning of impulse response in dB
hmax = max(abs(h)); % find maximum of impulse response
thr_val = hmax*10^(-thr/20); % absolute value of chosen threshold

% find index of maximum
index_max = find(abs(h)==max(abs(h)),1,'first');
% find index of threshold
%[~,index_thr] = min(abs(abs(h(1:index_max)) - thr_val));
index_thr = find(abs(h(1:index_max)) > thr_val,1,'first');

h = h(index_thr:end); % impulse response is now beginning at threshold
h2 = h.^2; % squared impulse response
N = length(h); % length of IR in samples
h_dur = N/fs; % duration of IR in seconds
t = ((0:(N-1))/fs)'; % timescale in seconds

%% EDC unfiltered
a = fliplr(cumtrapz(fliplr(h2')));
b = cumtrapz(h2);
edc_lin = (a./b');
N_edc = length(edc_lin);
% logarithmic EDC
EDC_unf = 10*log10(a./b(N));
```

```
%% Octave Filters and Filtering for calculation of T30 and EDT
% 125 Hz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F125 = design(fdesign.octave(1,'Class 2','N,F0',order,125,fs));
h_125 = filter(F125,h);
[gd125,~]=grpdelay(F125);
fdelay(1) = max(gd125);
% 250 Hz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F250 = design(fdesign.octave(1,'Class 2','N,F0',order,250,fs));
h_250 = filter(F250,h);
[gd250,~]=grpdelay(F250);
fdelay(2) = max(gd250);
% 500 Hz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F500 = design(fdesign.octave(1,'Class 2','N,F0',order,500,fs));
h_500 = filter(F500,h);
[gd500,~]=grpdelay(F500);
fdelay(3) = max(gd500);
% 1 kHz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F1k = design(fdesign.octave(1,'Class 2','N,F0',order,1000,fs));
h_1k = filter(F1k,h);
[gd1k,~]=grpdelay(F1k);
fdelay(4) = max(gd1k);
% 2 kHz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F2k = design(fdesign.octave(1,'Class 2','N,F0',order,2000,fs));
h_2k = filter(F2k,h);
[gd2k,~]=grpdelay(F2k);
fdelay(5) = max(gd2k);
% 4 kHz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F4k = design(fdesign.octave(1,'Class 2','N,F0',order,4000,fs));
h_4k = filter(F4k,h);
[gd4k,~]=grpdelay(F4k);
fdelay(6) = max(gd4k);
% 8 kHz %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
F8k = design(fdesign.octave(1,'Class 2','N,F0',order,8000,fs));
h_8k = filter(F8k,h);
[gd8k,~]=grpdelay(F8k);
fdelay(7) = max(gd8k);

fdelay = round(fdelay*fac);

% filtered and squared IR's
h_oct = [h_125,h_250,h_500,h_1k,h_2k,h_4k,h_8k];
h2_oct = h_oct.^2;

% integration limits in samples
t7  = round(0.007*fs); % 7ms
t10 = round(0.01*fs); % 10ms
t20 = round(0.02*fs); % 20ms
t50  = round(0.05*fs); % 50ms
t80  = round(0.08*fs); % 80ms
t100  = round(0.1*fs); % 100ms
t1000  = round(1*fs); % 1000ms
t40  = round(0.04*fs); % 40ms

%% energy relation measures, method 1
% Filtering time slots seperately
% First index is the beginning of broadband IR according to DIN EN ISO
% 3382-1:2009
```

```
G_1(1,1) = 10*log10(trapz((filter(F125,h)).^2)/trapz((filter(F125,h10mff)).^2));
G_1(1,2) = 10*log10(trapz((filter(F250,h)).^2)/trapz((filter(F250,h10mff)).^2));
G_1(1,3) = 10*log10(trapz((filter(F500,h)).^2)/trapz((filter(F500,h10mff)).^2));
G_1(1,4) = 10*log10(trapz((filter(F1k,h)).^2)/trapz((filter(F1k,h10mff)).^2));
G_1(1,5) = 10*log10(trapz((filter(F2k,h)).^2)/trapz((filter(F2k,h10mff)).^2));
G_1(1,6) = 10*log10(trapz((filter(F4k,h)).^2)/trapz((filter(F4k,h10mff)).^2));
G_1(1,7) = 10*log10(trapz((filter(F8k,h)).^2)/trapz((filter(F8k,h10mff)).^2));
G_1(1,8) = (G_1(1,3)+G_1(1,4))/2;

Ge_1(1,1) = 10*log10(trapz((filter(F125,h(1:t80))).^2)/trapz((filter(F125,h10mff)).^2));
Ge_1(1,2) = 10*log10(trapz((filter(F250,h(1:t80))).^2)/trapz((filter(F250,h10mff)).^2));
Ge_1(1,3) = 10*log10(trapz((filter(F500,h(1:t80))).^2)/trapz((filter(F500,h10mff)).^2));
Ge_1(1,4) = 10*log10(trapz((filter(F1k,h(1:t80))).^2)/trapz((filter(F1k,h10mff)).^2));
Ge_1(1,5) = 10*log10(trapz((filter(F2k,h(1:t80))).^2)/trapz((filter(F2k,h10mff)).^2));
Ge_1(1,6) = 10*log10(trapz((filter(F4k,h(1:t80))).^2)/trapz((filter(F4k,h10mff)).^2));
Ge_1(1,7) = 10*log10(trapz((filter(F8k,h(1:t80))).^2)/trapz((filter(F8k,h10mff)).^2));
Ge_1(1,8) = (Ge_1(1,3)+Ge_1(1,4))/2;

Gl_1(1,1) = 10*log10(trapz((filter(F125,h(t80:end))).^2)/trapz((filter(F125,h10mff)).^2));
Gl_1(1,2) = 10*log10(trapz((filter(F250,h(t80:end))).^2)/trapz((filter(F250,h10mff)).^2));
Gl_1(1,3) = 10*log10(trapz((filter(F500,h(t80:end))).^2)/trapz((filter(F500,h10mff)).^2));
Gl_1(1,4) = 10*log10(trapz((filter(F1k,h(t80:end))).^2)/trapz((filter(F1k,h10mff)).^2));
Gl_1(1,5) = 10*log10(trapz((filter(F2k,h(t80:end))).^2)/trapz((filter(F2k,h10mff)).^2));
Gl_1(1,6) = 10*log10(trapz((filter(F4k,h(t80:end))).^2)/trapz((filter(F4k,h10mff)).^2));
Gl_1(1,7) = 10*log10(trapz((filter(F8k,h(t80:end))).^2)/trapz((filter(F8k,h10mff)).^2));
Gl_1(1,8) = (Gl_1(1,3)+Gl_1(1,4))/2;

C50_1(1,1) = 10*log10(trapz((filter(F125,h(1:t50))).^2)/trapz((filter(F125,h(t50:end))).^2));
C50_1(1,2) = 10*log10(trapz((filter(F250,h(1:t50))).^2)/trapz((filter(F250,h(t50:end))).^2));
C50_1(1,3) = 10*log10(trapz((filter(F500,h(1:t50))).^2)/trapz((filter(F500,h(t50:end))).^2));
C50_1(1,4) = 10*log10(trapz((filter(F1k,h(1:t50))).^2)/trapz((filter(F1k,h(t50:end))).^2));
C50_1(1,5) = 10*log10(trapz((filter(F2k,h(1:t50))).^2)/trapz((filter(F2k,h(t50:end))).^2));
C50_1(1,6) = 10*log10(trapz((filter(F4k,h(1:t50))).^2)/trapz((filter(F4k,h(t50:end))).^2));
C50_1(1,7) = 10*log10(trapz((filter(F8k,h(1:t50))).^2)/trapz((filter(F8k,h(t50:end))).^2));
C50_1(1,8) = (C50_1(1,3)+C50_1(1,4))/2;
```

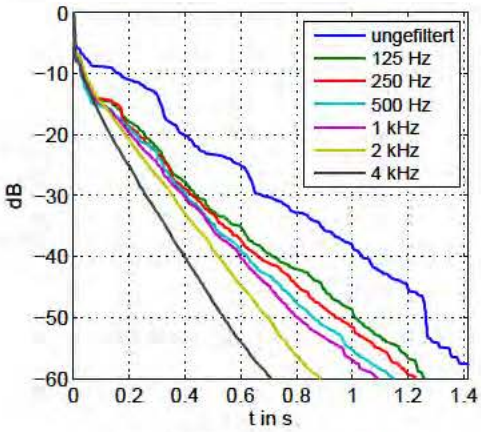
```
C80_1(1,1) = 10*log10(trapz((filter(F125,h(1:t80))).^2)/trapz((filter(F125,h(t80:end))).^2));
C80_1(1,2) = 10*log10(trapz((filter(F250,h(1:t80))).^2)/trapz((filter(F250,h(t80:end))).^2));
C80_1(1,3) = 10*log10(trapz((filter(F500,h(1:t80))).^2)/trapz((filter(F500,h(t80:end))).^2));
C80_1(1,4) = 10*log10(trapz((filter(F1k,h(1:t80))).^2)/trapz((filter(F1k,h(t80:end))).^2));
C80_1(1,5) = 10*log10(trapz((filter(F2k,h(1:t80))).^2)/trapz((filter(F2k,h(t80:end))).^2));
C80_1(1,6) = 10*log10(trapz((filter(F4k,h(1:t80))).^2)/trapz((filter(F4k,h(t80:end))).^2));
C80_1(1,7) = 10*log10(trapz((filter(F8k,h(1:t80))).^2)/trapz((filter(F8k,h(t80:end))).^2));
C80_1(1,8) = (C80_1(1,3)+C80_1(1,4))/2;

Tc_1(1,1) = 1000*(sum(t(1:t1000).*((filter(F125,h(1:t1000))).^2)) / sum((filter(F125,h(1:t1000))).^2)); % in ms
Tc_1(1,2) = 1000*(sum(t(1:t1000).*((filter(F250,h(1:t1000))).^2)) / sum((filter(F250,h(1:t1000))).^2)); % in ms
Tc_1(1,3) = 1000*(sum(t(1:t1000).*((filter(F500,h(1:t1000))).^2)) / sum((filter(F500,h(1:t1000))).^2)); % in ms
Tc_1(1,4) = 1000*(sum(t(1:t1000).*((filter(F1k,h(1:t1000))).^2)) / sum((filter(F1k,h(1:t1000))).^2)); % in ms
Tc_1(1,5) = 1000*(sum(t(1:t1000).*((filter(F2k,h(1:t1000))).^2)) / sum((filter(F2k,h(1:t1000))).^2)); % in ms
Tc_1(1,6) = 1000*(sum(t(1:t1000).*((filter(F4k,h(1:t1000))).^2)) / sum((filter(F4k,h(1:t1000))).^2)); % in ms
Tc_1(1,7) = 1000*(sum(t(1:t1000).*((filter(F8k,h(1:t1000))).^2)) / sum((filter(F8k,h(1:t1000))).^2)); % in ms
Tc_1(1,8) = (Tc_1(1,3)+Tc_1(1,4))/2;

% ST1 = STearly
ST1_1(1,1) = 10*log10(trapz((filter(F125,h((t20:t100)))).^2)/trapz((filter(F125,h(1:t10))).^2));
ST1_1(1,2) = 10*log10(trapz((filter(F250,h((t20:t100)))).^2)/trapz((filter(F250,h(1:t10))).^2));
ST1_1(1,3) = 10*log10(trapz((filter(F500,h((t20:t100)))).^2)/trapz((filter(F500,h(1:t10))).^2));
ST1_1(1,4) = 10*log10(trapz((filter(F1k,h((t20:t100)))).^2)/trapz((filter(F1k,h(1:t10))).^2));
ST1_1(1,5) = 10*log10(trapz((filter(F2k,h((t20:t100)))).^2)/trapz((filter(F2k,h(1:t10))).^2));
ST1_1(1,6) = 10*log10(trapz((filter(F4k,h((t20:t100)))).^2)/trapz((filter(F4k,h(1:t10))).^2));
ST1_1(1,7) = 10*log10(trapz((filter(F8k,h((t20:t100)))).^2)/trapz((filter(F8k,h(1:t10))).^2));
ST1_1(1,8) = (ST1_1(1,2)+ST1_1(1,3)+ST1_1(1,4)+ST1_1(1,5))/4;% average of octave bands 250Hz - 2kHz

% STtotal
STtot_1(1,1) = 10*log10(trapz((filter(F125,h(t20:t1000))).^2)/trapz((filter(F125,h(1:t10))).^2));
STtot_1(1,2) = 10*log10(trapz((filter(F250,h(t20:t1000))).^2)/trapz((filter(F250,h(1:t10))).^2));
STtot_1(1,3) = 10*log10(trapz((filter(F500,h(t20:t1000))).^2)/trapz((filter(F500,h(1:t10))).^2));
```

```
STtot_1(1,4) = 10*log10(trapz((filter(F1k,h(t20:t1000))).^2)/trapz((filter(F1k,h(1:t10))).^2));
STtot_1(1,5) = 10*log10(trapz((filter(F2k,h(t20:t1000))).^2)/trapz((filter(F2k,h(1:t10))).^2));
STtot_1(1,6) = 10*log10(trapz((filter(F4k,h(t20:t1000))).^2)/trapz((filter(F4k,h(1:t10))).^2));
STtot_1(1,7) = 10*log10(trapz((filter(F8k,h(t20:t1000))).^2)/trapz((filter(F8k,h(1:t10))).^2));
STtot_1(1,8) = (STtot_1(1,2)+STtot_1(1,3)+STtot_1(1,4)+STtot_1(1,5))/4; % average of octave bands 250Hz - 2kHz

STlate_1(1,1)= 10*log10(trapz((filter(F125,h(t100:t1000))).^2)/trapz((filter(F125,h(1:t10))).^2));
STlate_1(1,2)= 10*log10(trapz((filter(F250,h(t100:t1000))).^2)/trapz((filter(F250,h(1:t10))).^2));
STlate_1(1,3)= 10*log10(trapz((filter(F500,h(t100:t1000))).^2)/trapz((filter(F500,h(1:t10))).^2));
STlate_1(1,4)= 10*log10(trapz((filter(F1k,h(t100:t1000))).^2)/trapz((filter(F1k,h(1:t10))).^2));
STlate_1(1,5)= 10*log10(trapz((filter(F2k,h(t100:t1000))).^2)/trapz((filter(F2k,h(1:t10))).^2));
STlate_1(1,6)= 10*log10(trapz((filter(F4k,h(t100:t1000))).^2)/trapz((filter(F4k,h(1:t10))).^2));
STlate_1(1,7)= 10*log10(trapz((filter(F8k,h(t100:t1000))).^2)/trapz((filter(F8k,h(1:t10))).^2));
STlate_1(1,8) = (STlate_1(1,2)+STlate_1(1,3)+STlate_1(1,4)+STlate_1(1,5))/4; % average of octave bands 250Hz - 2kHz

LQ_1(1,1) = 10*log10(trapz((filter(F125,h(t7:t40))).^2)/trapz((filter(F125,h(t40:t1000))).^2));
LQ_1(1,2) = 10*log10(trapz((filter(F250,h(t7:t40))).^2)/trapz((filter(F250,h(t40:t1000))).^2));
LQ_1(1,3) = 10*log10(trapz((filter(F500,h(t7:t40))).^2)/trapz((filter(F500,h(t40:t1000))).^2));
LQ_1(1,4) = 10*log10(trapz((filter(F1k,h(t7:t40))).^2)/trapz((filter(F1k,h(t40:t1000))).^2));
LQ_1(1,5) = 10*log10(trapz((filter(F2k,h(t7:t40))).^2)/trapz((filter(F2k,h(t40:t1000))).^2));
LQ_1(1,6) = 10*log10(trapz((filter(F4k,h(t7:t40))).^2)/trapz((filter(F4k,h(t40:t1000))).^2));
LQ_1(1,7) = 10*log10(trapz((filter(F8k,h(t7:t40))).^2)/trapz((filter(F8k,h(t40:t1000))).^2));
LQ_1(1,8) = (LQ_1(1,3)+LQ_1(1,4)+LQ_1(1,5))/3; % %average of octave bands 500Hz - 2kHz

for i = 1:7
    %% Reverberation Time
    a_oct = fliplr(cumtrapz(fliplr(h2_oct(1:N,i)')));
    b_oct = cumtrapz(h2_oct(1:N,i));
    %edc_lin = (a_oct./b_oct');
    N_edc = length(edc_lin);
    % logarithmic EDC
    EDC(1:N_edc,i) = 10*log10(a_oct./b_oct(N));
    % EDC has to be normalized
    EDC_reg1(i) = find (EDC(:,i) <= -5, 1, 'first');
    EDC_reg2(i) = find (EDC(:,i) <= -35, 1, 'first');

    % DIN EN ISO 3382-1:2009, Seite 13:
    % Innerhalb des Auswertungsbereichs ist für die Kurve rechnerisch
```

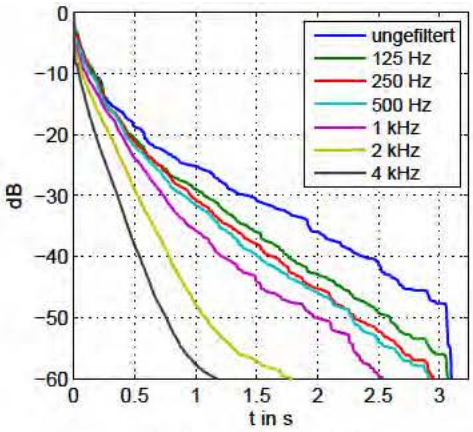
```
    % eine Gerade nach der Anpassung mit Hilfe des Verfahrens der
    % kleinsten Fehlerquadrate zu erstellen ...
    EDC_reg12{i} = EDC(EDC_reg1(1,i):EDC_reg2(1,i),i);
    x{i}=(1:length([EDC_reg12{i}]))';
    p(:,i) = polyfit([x{i}],[EDC_reg12{i}],1); % linear least square fitting

    y{i}=p(1,i)*[x{i}]+p(2,i);
    y0(1,i)=[y{i}(1,1)]-p(1,i)*EDC_reg1(i);

    % intersection of polyfit line with -60dB
    x_rt(i) = (-60-y0(1,i))/p(1,i);
    T30(i)=x_rt(i)/fs;   % Reverberation time in s

    %% EDT
    % Die Neigung der Abklingkurve sollte aus linearen
    % Regressions-Geraden bester Anpassung aus den anfänglichen 10 dB
    % (zwischen 0 dB und -10 dB) des Abklingvorgangs ermittelt werden.
    % Die Abklingzeiten sollten aus der Steigung innerhalb der Zeit, die
    % für einen 60 dB-Abklingvorgang erforderlich ist, ermittelt werden.
    EDCedt_reg1(i) = find (EDC(:,i) <= 0, 1, 'first');
    EDCedt_reg2(i) = find (EDC(:,i) <= -10, 1, 'first');%

    EDCedt_reg12{i} = EDC(EDCedt_reg1(1,i):EDCedt_reg2(1,i),i);
    xedt{i}=(1:length([EDCedt_reg12{i}]))';
    pedt(:,i) = polyfit([xedt{i}],[EDCedt_reg12{i}],1); % linear least square
        fitting

    yedt{i}=pedt(1,i)*[xedt{i}]+pedt(2,i);
    y0edt(1,i)=[yedt{i}(1,1)]-pedt(1,i)*EDCedt_reg1(i);

    % intersection of polyfit line with -60dB
    xedt_edt(i) = (-60-y0edt(1,i))/pedt(1,i);

    EDT(i)= xedt_edt(i)/fs;   % Early Decay Time in s
end
```

C. Eigenschaften der auralisierten Raummodelle

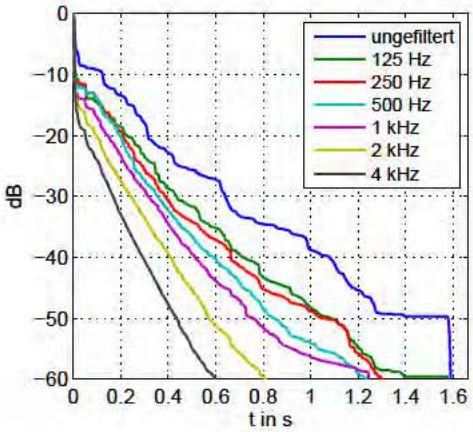
C.1. Schroeder-Plots



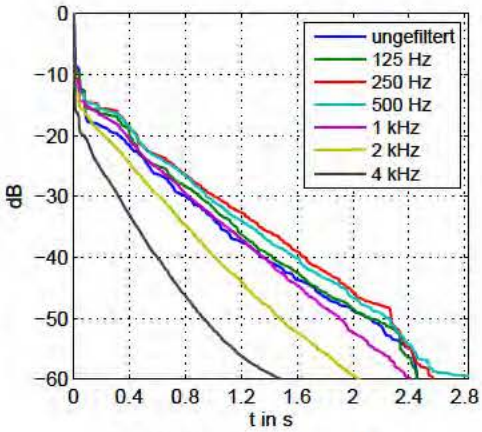
(a) R1: Kammersaal1



(b) R2: Kammersaal2

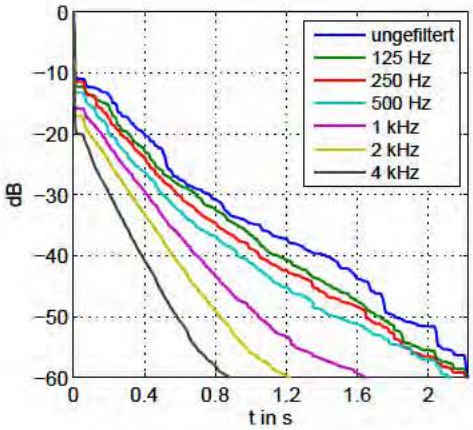


(c) R3: Oper

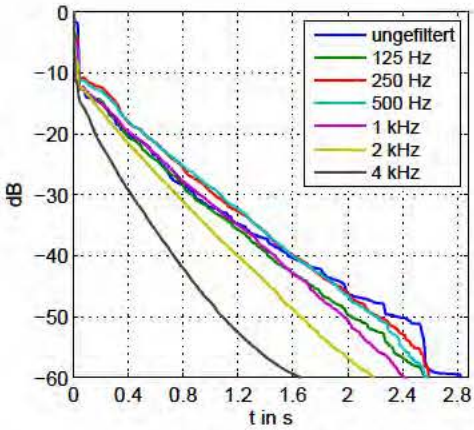


(d) R4: Konzertsaal1

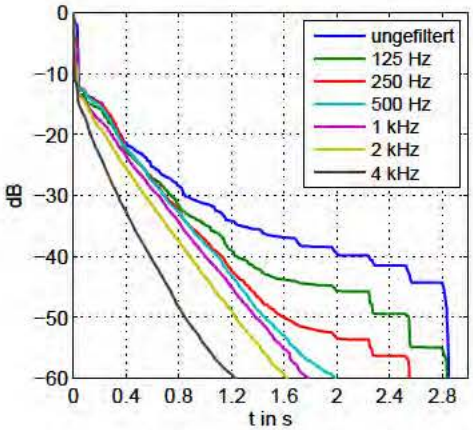
Abbildung C.2.: Schroeder-Plots der zur Berechnung der raumakustischen Parameter verwendeten Impulsantworten, Entfernung von Quelle zu Empfänger: 1 m, die simulierte Quelle hatte eine omnidirektionale Richtcharakteristik.



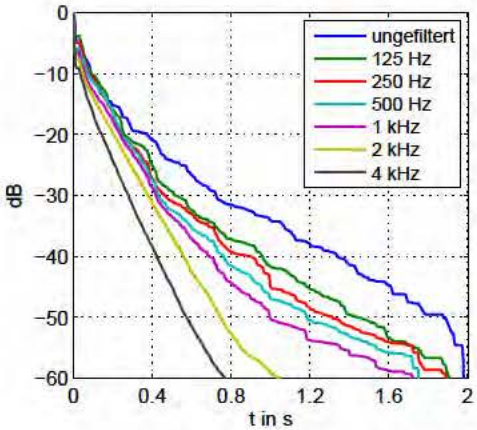
(e) R5: Konzertsaal2



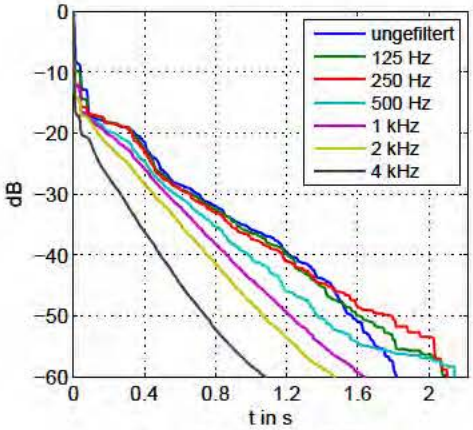
(f) R6: Kirche



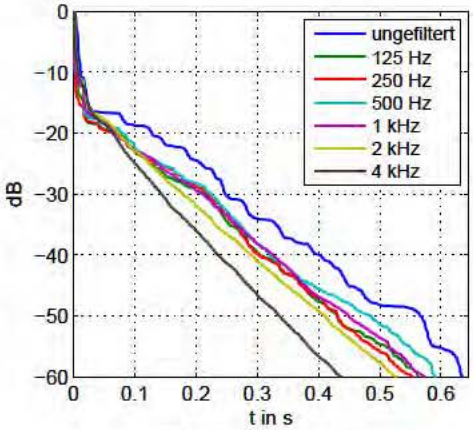
(g) R7: KircheShort



(h) R8: Kammersaal2Short



(i) R9: Kammersaal1Short



(j) R10: Theater

Abbildung C.2.: (Fortsetzung:) Schroeder-Plots der zur Berechnung der raumakustischen Parameter verwendeten Impulsantworten

C.2. Raumakustische Parameter

Tabelle C.1.: $T_{30\,1\mathrm{m}}$ in s

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	1,57	1,10	1,03	1,01	0,96	0,82	0,62	1,34	1,02	0,89
R2	3,66	2,59	2,31	2,16	1,70	1,21	0,84	3,13	2,24	1,46
R3	1,48	1,13	0,98	0,88	0,84	0,67	0,49	1,31	0,93	0,76
R4	2,23	2,37	2,71	2,57	2,25	1,74	1,02	2,30	2,64	2,00
R5	1,91	1,80	1,56	1,39	1,19	1,00	0,69	1,86	1,48	1,10
R6	2,40	2,29	2,56	2,62	2,41	1,93	1,16	2,35	2,59	2,17
R7	2,49	1,99	1,74	1,71	1,66	1,43	0,91	2,24	1,73	1,55
R8	1,92	1,24	1,17	1,03	0,98	0,88	0,65	1,58	1,10	0,93
R9	1,90	1,86	1,83	1,71	1,51	1,33	0,84	1,88	1,77	1,42
R10	0,69	0,60	0,64	0,61	0,58	0,47	0,37	0,65	0,63	0,53

Tabelle C.2.: $EDT_{1\mathrm{m}}$ in s

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	1,75	0,30	0,35	0,32	0,37	0,47	0,34	1,03	0,34	0,42
R2	1,17	1,21	1,11	1,19	0,96	0,75	0,54	1,19	1,15	0,86
R3	1,38	0,04	0,04	0,04	0,04	0,05	0,06	0,71	0,04	0,05
R4	0,28	0,34	0,03	0,01	0,05	0,06	0,07	0,31	0,02	0,06
R5	0,06	0,04	0,03	0,03	0,04	0,06	0,06	0,05	0,03	0,05
R6	0,39	0,47	0,44	0,55	0,84	0,45	0,07	0,43	0,50	0,65
R7	0,38	0,43	0,63	0,59	0,82	0,37	0,07	0,41	0,61	0,60
R8	0,70	0,61	0,65	0,77	0,69	0,62	0,41	0,66	0,71	0,66
R9	0,29	0,31	0,02	0,01	0,03	0,07	0,06	0,30	0,02	0,05
R10	0,06	0,03	0,02	0,01	0,03	0,07	0,08	0,05	0,02	0,05

Tabelle C.3.: $G_{1\mathrm{m}}$ in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	17,05	22,31	20,7	20,99	21,25	21,64	22,53	19,68	20,85	21,45
R2	18,96	22,36	21,61	21,24	21,35	21,68	21,99	20,66	21,43	21,52
R3	18,43	21,4	20,28	20,48	20,51	20,73	21,56	19,92	20,38	20,62
R4	19,55	19,89	19,62	19,68	20,13	20,46	21,37	19,72	19,65	20,3
R5	19,51	20,25	19,75	19,94	20,15	20,48	20,87	19,88	19,85	20,32
R6	17,26	23,4	21,72	21,56	22,33	21,43	21,61	20,33	21,64	21,88
R7	17,61	23,71	21,45	21,1	22,18	21,61	21,69	20,66	21,28	21,9
R8	19,23	21,65	20,94	20,67	20,9	21,16	21,78	20,44	20,81	21,03
R9	18,94	20	19,44	19,56	20,12	20,61	21,28	19,47	19,5	20,37
R10	18,37	19,89	19,25	19,74	20,13	20,72	21,29	19,13	19,5	20,43

Tabelle C.4.: $C_{50\,1\mathrm{m}}$ in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	7,32	11,93	12,28	12,99	11,48	10,53	11,98	9,63	12,64	11,01
R2	3,75	3,53	3,61	4,75	6,52	7,95	10,29	3,64	4,18	7,24
R3	8,13	12,74	12,13	12,46	14,68	16,91	19,93	10,44	12,30	15,80
R4	12,26	12,46	13,01	12,95	14,15	15,41	19,61	12,36	12,98	14,78
R5	10,65	11,96	11,16	13,04	15,75	17,11	20,02	11,31	12,10	16,43
R6	12,50	12,30	10,43	10,92	12,02	12,13	14,79	12,40	10,68	12,08
R7	12,44	13,27	12,25	12,41	13,57	13,49	15,37	12,86	12,33	13,53
R8	6,24	5,77	6,74	7,35	8,72	9,79	11,58	6,01	7,05	9,26
R9	12,61	14,27	15,58	16,52	16,52	16,99	20,26	13,44	16,05	16,76
R10	16,54	19,35	19,46	18,54	18,13	17,80	18,51	17,95	19,00	17,97

Tabelle C.5.: $C_{80\,1\mathrm{m}}$ in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	8,16	14,09	13,77	14,91	14,06	13,30	15,07	11,13	14,34	13,68
R2	6,18	6,00	6,32	7,15	8,68	10,21	12,82	6,09	6,74	9,45
R3	8,46	13,83	12,84	12,86	15,59	18,64	22,63	11,15	12,85	17,12
R4	17,06	15,26	14,18	14,29	14,82	16,36	20,25	16,16	14,24	15,59
R5	12,04	13,28	13,38	15,00	17,37	19,43	22,08	12,66	14,19	18,40
R6	13,61	12,76	10,69	11,08	12,51	12,73	15,83	13,19	10,89	12,62
R7	13,16	13,76	12,69	12,63	14,17	14,31	16,76	13,46	12,66	14,24
R8	9,25	8,88	9,81	10,10	11,62	12,49	14,78	9,07	9,96	12,06
R9	16,64	16,75	16,52	17,19	16,97	17,43	20,75	16,70	16,86	17,20
R10	18,40	20,52	20,92	19,93	21,03	20,83	22,46	19,46	20,43	20,93

Tabelle C.6.: $T_{C\,1\mathrm{m}}$ in ms

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	41,98	16,95	14,90	12,15	12,63	14,07	10,41	29,47	13,53	13,35
R2	55,31	55,07	50,27	42,59	32,04	22,40	12,83	55,19	46,43	27,22
R3	30,48	12,68	11,05	9,59	5,89	3,95	2,79	21,58	10,32	4,92
R4	17,57	17,24	16,78	15,04	12,53	7,62	3,71	17,41	15,91	10,08
R5	22,79	15,41	13,00	8,65	5,12	3,70	2,09	19,10	10,83	4,41
R6	36,13	27,94	33,87	28,30	24,33	16,62	7,50	32,04	31,09	20,48
R7	35,21	23,99	23,55	19,19	17,82	11,75	6,55	29,60	21,37	14,79
R8	34,37	34,68	27,93	23,39	18,34	14,53	9,69	34,53	25,66	16,44
R9	17,66	13,36	9,91	6,83	6,97	5,52	3,12	15,51	8,37	6,25
R10	8,92	4,89	3,29	3,67	3,23	3,78	3,45	6,91	3,48	3,51

Tabelle C.7.: ST_{early} in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{250\text{-}2\mathrm{k}}$
R1	−7,56	−7,34	−8,01	−8,02	−7,68	−6,70	−7,97	−7,76
R2	−3,17	−2,06	−3,31	−4,91	−5,87	−6,99	−8,76	−4,04
R3	−15,08	−14,52	−14,99	−16,87	−16,57	−16,40	−18,09	−15,74
R4	−8,34	−10,19	−12,66	−15,32	−12,54	−15,10	−17,90	−12,68
R5	−15,93	−17,52	−15,12	−17,01	−19,69	−19,82	−22,12	−17,34
R6	2,98	−2,04	−3,31	−5,69	−3,56	−8,67	−12,56	−3,65
R7	2,69	−2,57	−3,56	−6,10	−3,84	−9,12	−12,27	−4,02
R8	−4,20	−2,57	−4,35	−6,08	−6,54	−7,68	−9,30	−4,89
R9	−8,16	−10,18	−13,52	−16,68	−13,52	−16,41	−18,74	−13,48
R10	−18,46	−19,66	−19,42	−18,35	−16,78	−14,59	−13,87	−18,55

Tabelle C.8.: ST_{total} in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{250\text{-}2\mathrm{k}}$
R1	−4,46	−6,39	−6,88	−7,15	−6,80	−5,94	−7,39	−6,81
R2	−1,01	−0,21	−1,25	−2,74	−3,95	−5,39	−7,61	−2,04
R3	−7,13	−11,30	−11,12	−11,90	−13,61	−14,79	−17,09	−11,98
R4	−7,74	−8,99	−10,38	−11,85	−10,55	−12,80	−16,12	−10,44
R5	−10,27	−11,94	−11,16	−13,03	−15,71	−17,00	−19,79	−12,96
R6	3,24	−1,55	−2,38	−4,46	−2,91	−7,22	−11,08	−2,83
R7	2,99	−2,16	−2,94	−5,18	−3,39	−8,00	−11,13	−3,42
R8	−2,75	−1,45	−3,08	−4,66	−5,39	−6,57	−8,48	−3,65
R9	−7,51	−9,29	−11,81	−14,12	−12,05	−14,15	−16,92	−11,82
R10	−15,25	−17,95	−17,78	−16,81	−15,77	−13,93	−13,52	−17,08

Tabelle C.9.: ST_{late} in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{250\text{-}2k}$
R1	−7,40	−13,46	−13,30	−14,53	−14,17	−13,87	−16,41	−13,87
R2	−5,09	−4,84	−5,49	−6,81	−8,44	−10,51	−13,95	−6,40
R3	−7,95	−14,17	−13,67	−13,67	−16,70	−19,85	−23,88	−14,55
R4	−16,75	−15,17	−14,30	−14,44	−14,90	−16,67	−20,87	−14,70
R5	−11,65	−13,34	−13,40	−15,25	−17,94	−20,20	−23,61	−14,98
R6	−9,10	−11,43	−9,55	−10,56	−11,44	−12,69	−16,48	−10,75
R7	−8,81	−12,70	−11,73	−12,39	−13,40	−14,45	−17,50	−12,56
R8	−8,22	−7,91	−9,06	−10,21	−11,73	−13,04	−16,12	−9,73
R9	−16,19	−16,63	−16,65	−17,61	−17,53	−18,11	−21,58	−17,11
R10	−18,09	−23,28	−22,95	−22,19	−22,59	−22,53	−24,62	−22,75

Tabelle C.10.: LQ_{7-40} in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{500\text{-}2k}$
R1	−3,03	1,08	0,96	2,65	0,74	0,06	0,77	1,45
R2	−4,51	−5,05	−6,40	−6,08	−3,66	−3,23	−1,32	−5,38
R3	−1,93	−4,95	−9,66	−6,94	−4,17	1,38	5,13	−6,92
R4	−2,70	−4,28	−2,47	−4,35	1,69	0,77	7,32	−1,71
R5	−18,07	−17,95	−11,11	−10,52	−4,76	0,45	3,85	−8,80
R6	10,36	8,08	5,39	4,10	6,79	2,81	2,60	5,43
R7	10,13	8,70	7,02	5,20	8,00	4,19	3,33	6,74
R8	−2,97	−3,47	−4,09	−4,06	−1,75	−1,60	−0,49	−3,30
R9	−0,74	−2,31	−0,57	−1,55	2,34	1,85	4,59	0,07
R10	4,05	3,04	1,80	6,55	5,61	5,76	7,43	4,65

Tabelle C.11.: $G_{e\,1\mathrm{m}}$ in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1k}$	$\bar{x}_{2k,4k}$
R1	17,7	22,14	20,52	20,85	21,8	21,44	22,4	19,92	20,69	21,92
R2	20,24	21,38	20,70	20,48	20,8	21,28	21,77	20,81	20,59	21,53
R3	18,66	21,22	20,06	20,26	20,39	20,67	21,54	19,94	20,16	21,11
R4	19,99	19,74	19,46	19,52	19,99	20,36	21,33	19,87	19,49	21,33
R5	19,6	20,04	19,55	19,81	20,07	20,43	20,84	19,82	19,68	20,64
R6	21,78	23,16	21,36	21,24	22,09	21,21	21,5	22,47	21,3	21,36
R7	21,96	23,52	21,21	20,87	22,01	21,45	21,6	22,74	21,04	21,53
R8	20,29	21,11	20,5	20,27	20,61	20,92	21,64	20,7	20,38	21,28
R9	19,42	19,89	19,35	19,48	20,03	20,54	21,24	19,66	19,41	20,89
R10	18,45	19,85	19,21	19,7	20,1	20,69	21,27	19,15	19,46	20,98

Tabelle C.12.: $G_{l\,1\mathrm{m}}$ in dB

f in Hz	125	250	500	1k	2k	4k	8k	$\bar{x}_{125,250}$	$\bar{x}_{500,1\mathrm{k}}$	$\bar{x}_{2\mathrm{k},4\mathrm{k}}$
R1	9,54	8,05	6,76	5,94	7,01	8,14	7,33	8,8	6,35	7,74
R2	14,05	15,39	14,38	13,33	12,12	11,07	8,95	14,72	13,85	10,01
R3	10,2	7,39	7,22	7,4	4,8	2,03	−1,09	8,8	7,31	0,47
R4	2,93	4,48	5,27	5,23	5,17	4	1,07	3,71	5,25	2,54
R5	7,56	6,76	6,17	4,8	2,7	1	−1,24	7,16	5,49	−0,12
R6	8,17	10,4	10,67	10,16	9,58	8,47	5,66	9,29	10,41	7,07
R7	8,81	9,76	8,53	8,24	7,84	7,14	4,83	9,29	8,38	5,99
R8	11,04	12,23	10,69	10,17	9,00	8,43	6,86	11,64	10,43	7,65
R9	2,78	3,15	2,83	2,29	3,07	3,1	0,49	2,97	2,56	1,8
R10	0,06	−0,67	−1,71	−0,23	−0,94	−0,15	−1,19	−0,31	−0,97	−0,67

Tabelle C.13.: Verhältnismaße aus den angegebenen Parametern

	$G_{w\,1\mathrm{m}}$	$BR_{T\,1\mathrm{m}}$	$BR_{EDT\,1\mathrm{m}}$	$BR_{G\,1\mathrm{m}}$	$TR_{T\,1\mathrm{m}}$	$TR_{EDT\,1\mathrm{m}}$	$TR_{G\,1\mathrm{m}}$	$EDTF_{1\mathrm{m}}$
R1	22,49	1,31	3,06	0,94	0,87	1,25	1,03	0,94
R2	24,81	1,40	1,03	0,96	0,65	0,74	1,00	1,08
R3	22,54	1,40	17,75	0,98	0,81	1,13	1,01	1,00
R4	21,96	0,87	15,50	1,00	0,76	2,75	1,03	6,17
R5	22,20	1,26	1,67	1,00	0,74	1,67	1,02	1,00
R6	24,63	0,91	0,87	0,94	0,84	1,30	1,01	0,65
R7	24,86	1,30	0,66	0,97	0,90	0,98	1,03	0,75
R8	23,70	1,44	0,92	0,98	0,85	0,92	1,01	0,86
R9	21,61	1,06	20,00	1,00	0,80	3,33	1,04	8,25
R10	20,89	1,03	3,00	0,98	0,84	3,33	1,05	1,25

D. Daten der statistischen Auswertung

D.1. Varianzanalysen mit Messwiederholung

D.1.1. MANOVA mit z-transformierten Daten; I1–I20

Multivariate[a,b]

Innersubjekteffekt		Wert	F	Hypothese df	Fehler df	Sig.	Partielles Eta-Quadrat	Nichtzentralitäts-Parameter	Beobachtete Schärfe[d]
Raum	Pillai-Spur	2,063	1,576	180,000	954,000	,000	,229	283,663	1,000
	Wilks-Lambda	,043	2,136	180,000	826,725	,000	,295	346,552	1,000
	Hotelling-Spur	6,438	3,442	180,000	866,000	,000	,417	619,496	1,000
	Größte charakteristische Wurzel nach Roy	4,885	25,891[c]	20,000	106,000	,000	,830	517,830	1,000

a. Design: Konstanter Term
Innersubjektdesign: Raum

b. Die Tests basieren auf den gemittelten Variablen.

c. Die Statistik ist eine Obergrenze auf F, die eine Untergrenze auf dem Signifikanzniveau ergibt.

d. Unter Verwendung von Alpha = ,05 berechnet

D.2. ANOVAs mit z-transformierten Daten; I1-I20

Mauchly-Test auf Sphärizität[a]

Innersubjekteffekt	Maß	Mauchly-W	Approximiertes Chi-Quadrat	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Raum	I1	,016	40,606	44	,699	,560	,958	,111
	I2	,009	45,895	44	,484	,560	,957	,111
	I3	,010	44,935	44	,523	,516	,837	,111
	I4	,011	44,082	44	,558	,568	,981	,111
	I5	,001	67,161	44	,025	,463	,708	,111
	I6	,013	42,412	44	,627	,636	1,000	,111
	I7	,005	51,436	44	,279	,523	,857	,111
	I8	,011	44,273	44	,550	,621	1,000	,111
	I9	,017	39,766	44	,730	,577	1,000	,111
	I10	,001	65,773	44	,033	,571	,988	,111
	I11	,008	47,185	44	,432	,559	,954	,111
	I12	,009	46,308	44	,467	,590	1,000	,111
	I13	,005	52,747	44	,239	,531	,878	,111
	I14	,011	44,353	44	,547	,582	1,000	,111
	I15	,002	59,366	44	,096	,592	1,000	,111
	I16	,003	55,774	44	,162	,567	,977	,111
	I17	,001	63,890	44	,046	,513	,830	,111
	I18	,041	31,232	44	,950	,674	1,000	,111
	I19	,092	23,386	44	,997	,707	1,000	,111
	I20	,006	49,858	44	,332	,624	1,000	,111

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term
Innersubjektdesign: Raum

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

Tests auf Univariate

Quelle	Maß		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.	Partielles Eta-Quadrat	Nichtzentralitäts-Parameter	Beobachtete Schärfe[a]
Raum	I1	Sphäriztät angenommen	62,526	9	6,947	12,806	,000	,496	115,251	1,000
		Greenhouse-Geisser	62,526	5,042	12,400	12,806	,000	,496	64,573	1,000
		Huynh-Feldt	62,526	8,620	7,253	12,806	,000	,496	110,387	1,000
		Untergrenze	62,526	1,000	62,526	12,806	,003	,496	12,806	,911
	I2	Sphäriztät angenommen	20,000	9	2,222	2,453	,014	,159	22,076	,912
		Greenhouse-Geisser	20,000	5,041	3,967	2,453	,042	,159	12,366	,741
		Huynh-Feldt	20,000	8,617	2,321	2,453	,015	,159	21,137	,902
		Untergrenze	20,000	1,000	20,000	2,453	,141	,159	2,453	,306
	I3	Sphäriztät angenommen	92,882	9	10,320	36,459	,000	,737	328,129	1,000
		Greenhouse-Geisser	92,882	4,642	20,008	36,459	,000	,737	169,251	1,000
		Huynh-Feldt	92,882	7,537	12,324	36,459	,000	,737	274,787	1,000
		Untergrenze	92,882	1,000	92,882	36,459	,000	,737	36,459	1,000
	I4	Sphäriztät angenommen	5,699	9	,633	,616	,782	,045	5,543	,291
		Greenhouse-Geisser	5,699	5,114	1,114	,616	,692	,045	3,150	,214
		Huynh-Feldt	5,699	8,625	,646	,616	,778	,045	5,435	,287
		Untergrenze	5,699	1,000	5,699	,616	,447	,045	,616	,113
	I5	Sphäriztät angenommen	88,514	9	9,835	30,697	,000	,702	276,269	1,000
		Greenhouse-Geisser	88,514	4,164	21,255	30,697	,000	,702	127,833	1,000
		Huynh-Feldt	88,514	6,372	13,891	30,697	,000	,702	195,602	1,000
		Untergrenze	88,514	1,000	88,514	30,697	,000	,702	30,697	,999
	I6	Sphäriztät angenommen	10,935	9	1,215	1,235	,280	,087	11,119	,583
		Greenhouse-Geisser	10,935	5,726	1,910	1,235	,299	,087	7,074	,446
		Huynh-Feldt	10,935	9,000	1,215	1,235	,280	,087	11,119	,583
		Untergrenze	10,935	1,000	10,935	1,235	,286	,087	1,235	,178
	I7	Sphäriztät angenommen	86,690	9	9,632	28,669	,000	,688	258,021	1,000
		Greenhouse-Geisser	86,690	4,710	18,407	28,669	,000	,688	135,017	1,000
		Huynh-Feldt	86,690	7,712	11,241	28,669	,000	,688	221,086	1,000
		Untergrenze	86,690	1,000	86,690	28,669	,000	,688	28,669	,999
	I8	Sphäriztät angenommen	49,822	9	5,536	8,502	,000	,395	76,520	1,000
		Greenhouse-Geisser	49,822	5,592	8,909	8,502	,000	,395	47,545	1,000
		Huynh-Feldt	49,822	9,000	5,536	8,502	,000	,395	76,520	1,000
		Untergrenze	49,822	1,000	49,822	8,502	,012	,395	8,502	,769
	I9	Sphäriztät angenommen	10,578	9	1,175	1,191	,307	,084	10,723	,564
		Greenhouse-Geisser	10,578	5,191	2,038	1,191	,323	,084	6,185	,406
		Huynh-Feldt	10,578	9,000	1,175	1,191	,307	,084	10,723	,564
		Untergrenze	10,578	1,000	10,578	1,191	,295	,084	1,191	,173
	I10	Sphäriztät angenommen	74,758	9	8,306	18,966	,000	,593	170,695	1,000
		Greenhouse-Geisser	74,758	5,136	14,555	18,966	,000	,593	97,417	1,000
		Huynh-Feldt	74,758	8,890	8,409	18,966	,000	,593	168,612	1,000
		Untergrenze	74,758	1,000	74,758	18,966	,001	,593	18,966	,980
	I11	Sphäriztät angenommen	22,944	9	2,549	2,894	,004	,182	26,048	,956
		Greenhouse-Geisser	22,944	5,031	4,560	2,894	,020	,182	14,562	,818
		Huynh-Feldt	22,944	8,589	2,671	2,894	,005	,182	24,858	,949
		Untergrenze	22,944	1,000	22,944	2,894	,113	,182	2,894	,351
	I12	Sphäriztät angenommen	39,364	9	4,374	5,907	,000	,312	53,160	1,000
		Greenhouse-Geisser	39,364	5,310	7,413	5,907	,000	,312	31,365	,993
		Huynh-Feldt	39,364	9,000	4,374	5,907	,000	,312	53,160	1,000
		Untergrenze	39,364	1,000	39,364	5,907	,030	,312	5,907	,614
	I13	Sphäriztät angenommen	38,296	9	4,255	5,676	,000	,304	51,088	1,000
		Greenhouse-Geisser	38,296	4,783	8,007	5,676	,000	,304	27,151	,986
		Huynh-Feldt	38,296	7,906	4,844	5,676	,000	,304	44,879	,999
		Untergrenze	38,296	1,000	38,296	5,676	,033	,304	5,676	,597
	I14	Sphäriztät angenommen	38,517	9	4,280	5,724	,000	,306	51,512	1,000
		Greenhouse-Geisser	38,517	5,242	7,348	5,724	,000	,306	30,001	,991
		Huynh-Feldt	38,517	9,000	4,280	5,724	,000	,306	51,512	1,000
		Untergrenze	38,517	1,000	38,517	5,724	,033	,306	5,724	,600
	I15	Sphäriztät angenommen	20,943	9	2,327	2,591	,009	,166	23,323	,929
		Greenhouse-Geisser	20,943	5,331	3,928	2,591	,030	,166	13,816	,786
		Huynh-Feldt	20,943	9,000	2,327	2,591	,009	,166	23,323	,929
		Untergrenze	20,943	1,000	20,943	2,591	,131	,166	2,591	,320
	I16	Sphäriztät angenommen	9,695	9	1,077	1,084	,380	,077	9,753	,515
		Greenhouse-Geisser	9,695	5,103	1,900	1,084	,378	,077	5,530	,366
		Huynh-Feldt	9,695	8,795	1,102	1,084	,380	,077	9,530	,508
		Untergrenze	9,695	1,000	9,695	1,084	,317	,077	1,084	,162
	I17	Sphäriztät angenommen	18,622	9	2,069	2,255	,023	,148	20,291	,882
		Greenhouse-Geisser	18,622	4,617	4,033	2,255	,065	,148	10,410	,668
		Huynh-Feldt	18,622	7,472	2,492	2,255	,033	,148	16,847	,829
		Untergrenze	18,622	1,000	18,622	2,255	,157	,148	2,255	,285

	I18	Sphäriztät angenommen	15,730	9	1,748	1,855	,066	,125	16,691	,797
		Greenhouse-Geisser	15,730	6,064	2,594	1,855	,098	,125	11,245	,662
		Huynh-Feldt	15,730	9,000	1,748	1,855	,066	,125	16,691	,797
		Untergrenze	15,730	1,000	15,730	1,855	,196	,125	1,855	,243
	I19	Sphäriztät angenommen	23,533	9	2,615	2,986	,003	,187	26,870	,962
		Greenhouse-Geisser	23,533	6,360	3,700	2,986	,010	,187	18,987	,896
		Huynh-Feldt	23,533	9,000	2,615	2,986	,003	,187	26,870	,962
		Untergrenze	23,533	1,000	23,533	2,986	,108	,187	2,986	,360
	I20	Sphäriztät angenommen	11,964	9	1,329	1,364	,212	,095	12,275	,635
		Greenhouse-Geisser	11,964	5,617	2,130	1,364	,243	,095	7,661	,484
		Huynh-Feldt	11,964	9,000	1,329	1,364	,212	,095	12,275	,635
		Untergrenze	11,964	1,000	11,964	1,364	,264	,095	1,364	,191
Fehler(Raum)	I1	Sphäriztät angenommen	63,474	117	,543					
		Greenhouse-Geisser	63,474	65,552	,968					
		Huynh-Feldt	63,474	112,061	,566					
		Untergrenze	63,474	13,000	4,883					
	I2	Sphäriztät angenommen	106,000	117	,906					
		Greenhouse-Geisser	106,000	65,539	1,617					
		Huynh-Feldt	106,000	112,024	,946					
		Untergrenze	106,000	13,000	8,154					
	I3	Sphäriztät angenommen	33,118	117	,283					
		Greenhouse-Geisser	33,118	60,349	,549					
		Huynh-Feldt	33,118	97,980	,338					
		Untergrenze	33,118	13,000	2,548					
	I4	Sphäriztät angenommen	120,301	117	1,028					
		Greenhouse-Geisser	120,301	66,482	1,810					
		Huynh-Feldt	120,301	114,729	1,049					
		Untergrenze	120,301	13,000	9,254					
	I5	Sphäriztät angenommen	37,486	117	,320					
		Greenhouse-Geisser	37,486	54,137	,692					
		Huynh-Feldt	37,486	82,838	,453					
		Untergrenze	37,486	13,000	2,884					
	I6	Sphäriztät angenommen	115,065	117	,983					
		Greenhouse-Geisser	115,065	74,435	1,546					
		Huynh-Feldt	115,065	117,000	,983					
		Untergrenze	115,065	13,000	8,851					
	I7	Sphäriztät angenommen	39,310	117	,336					
		Greenhouse-Geisser	39,310	61,224	,642					
		Huynh-Feldt	39,310	100,252	,392					
		Untergrenze	39,310	13,000	3,024					
	I8	Sphäriztät angenommen	76,178	117	,651					
		Greenhouse-Geisser	76,178	72,696	1,048					
		Huynh-Feldt	76,178	117,000	,651					
		Untergrenze	76,178	13,000	5,860					
	I9	Sphäriztät angenommen	115,422	117	,987					
		Greenhouse-Geisser	115,422	67,483	1,710					
		Huynh-Feldt	115,422	117,000	,987					
		Untergrenze	115,422	13,000	8,879					
	I10	Sphäriztät angenommen	51,242	117	,438					
		Greenhouse-Geisser	51,242	66,773	,767					
		Huynh-Feldt	51,242	115,573	,443					
		Untergrenze	51,242	13,000	3,942					
	I11	Sphäriztät angenommen	103,056	117	,881					
		Greenhouse-Geisser	103,056	65,409	1,576					
		Huynh-Feldt	103,056	111,654	,923					
		Untergrenze	103,056	13,000	7,927					
	I12	Sphäriztät angenommen	86,636	117	,740					
		Greenhouse-Geisser	86,636	69,031	1,255					
		Huynh-Feldt	86,636	117,000	,740					
		Untergrenze	86,636	13,000	6,664					
	I13	Sphäriztät angenommen	87,704	117	,750					
		Greenhouse-Geisser	87,704	62,180	1,410					
		Huynh-Feldt	87,704	102,779	,853					
		Untergrenze	87,704	13,000	6,746					
	I14	Sphäriztät angenommen	87,483	117	,748					
		Greenhouse-Geisser	87,483	68,142	1,284					
		Huynh-Feldt	87,483	117,000	,748					
		Untergrenze	87,483	13,000	6,729					
	I15	Sphäriztät angenommen	105,057	117	,898					
		Greenhouse-Geisser	105,057	69,307	1,516					
		Huynh-Feldt	105,057	117,000	,898					
		Untergrenze	105,057	13,000	8,081					

I16	Sphäriziität angenommen	116,305	117	,994					
	Greenhouse-Geisser	116,305	66,344	1,753					
	Huynh-Feldt	116,305	114,330	1,017					
	Untergrenze	116,305	13,000	8,947					
I17	Sphäriziität angenommen	107,378	117	,918					
	Greenhouse-Geisser	107,378	60,022	1,789					
	Huynh-Feldt	107,378	97,139	1,105					
	Untergrenze	107,378	13,000	8,260					
I18	Sphäriziität angenommen	110,270	117	,942					
	Greenhouse-Geisser	110,270	78,828	1,399					
	Huynh-Feldt	110,270	117,000	,942					
	Untergrenze	110,270	13,000	8,482					
I19	Sphäriziität angenommen	102,467	117	,876					
	Greenhouse-Geisser	102,467	82,676	1,239					
	Huynh-Feldt	102,467	117,000	,876					
	Untergrenze	102,467	13,000	7,882					
I20	Sphäriziität angenommen	114,036	117	,975					
	Greenhouse-Geisser	114,036	73,021	1,562					
	Huynh-Feldt	114,036	117,000	,975					
	Untergrenze	114,036	13,000	8,772					

a. Unter Verwendung von Alpha = ,05 berechnet

Tabelle D.1.: ICC(3,1)

		Vertrauensintervall $\alpha = 0,05$		Hypothesen-Test ICC = 0			
	ICC(3,1)			F	df1	df2	p
I1	0,472	0,266	0,762	13,52	9	117	0,000
I2	0,164	0,047	0,456	3,75	9	117	0,000
I3	0,694	0,495	0,887	32,73	9	117	0,000
I4	−0,024	−0,052	0,084	0,67	9	117	0,732
I5	0,658	0,452	0,870	27,90	9	117	0,000
I6	0,016	−0,033	0,185	1,23	9	117	0,283
I7	0,611	0,400	0,846	23,00	9	117	0,000
I8	0,341	0,162	0,659	8,26	9	117	0,000
I9	0,011	−0,035	0,173	1,16	9	117	0,328
I10	0,565	0,352	0,820	19,15	9	117	0,000
I11	0,108	0,015	0,369	2,70	9	117	0,007
I12	0,266	0,110	0,583	6,06	9	117	0,000
I13	0,312	0,141	0,631	7,34	9	117	0,000
I14	0,261	0,106	0,578	5,94	9	117	0,000
I15	0,062	−0,010	0,284	1,93	9	117	0,054
I16	0,002	−0,040	0,150	1,02	9	117	0,427
I17	0,117	0,020	0,383	2,86	9	117	0,004
I18	0,068	−0,006	0,295	2,02	9	117	0,043
I19	0,129	0,027	0,403	3,08	9	117	0,002
I20	0,018	−0,032	0,188	1,25	9	117	0,272

Tabelle D.2.: ICC(2,1)

	ICC(2,1)	Vertrauensintervall $\alpha = 0,05$		Hypothesen-Test ICC = 0 F	df1	df2	p
I1	0,446	0,246	0,742	13,52	9	117	0,000
I2	0,134	0,037	0,398	3,75	9	117	0,000
I3	0,632	0,422	0,857	32,73	9	117	0,000
I4	−0,024	−0,052	0,083	0,67	9	117	0,732
I5	0,638	0,431	0,860	27,90	9	117	0,000
I6	0,017	−0,035	0,195	1,23	9	117	0,283
I7	0,573	0,362	0,825	23,00	9	117	0,000
I8	0,323	0,152	0,640	8,26	9	117	0,000
I9	0,010	−0,033	0,164	1,16	9	117	0,328
I10	0,515	0,307	0,790	19,15	9	117	0,000
I11	0,099	0,014	0,345	2,70	9	117	0,007
I12	0,220	0,087	0,522	6,06	9	117	0,000
I13	0,268	0,117	0,581	7,34	9	117	0,000
I14	0,223	0,088	0,526	5,94	9	117	0,000
I15	0,061	−0,009	0,280	1,93	9	117	0,054
I16	0,001	−0,036	0,138	1,02	9	117	0,427
I17	0,107	0,019	0,359	2,86	9	117	0,004
I18	0,069	−0,006	0,298	2,02	9	117	0,043
I19	0,126	0,026	0,395	3,08	9	117	0,002
I20	0,017	−0,031	0,184	1,25	9	117	0,272

Tabelle D.3.: Ergebnisse aus zweifaktoriellen ANOVAs je Item

	Gegensatzpaar	QS_e	df_e	MQ_e	QS_{rater}	df_{rater}	MQ_{rater}	F_{rater}	p_{rater}
I1	tragfähig – wenig tragfähig	347,57	117	2,97	81,64	13	6,28	2,11	0,018
I2	ich kann mich gut hören – ... nicht gut hören	455,63	117	3,89	186,69	13	14,36	3,69	0,000
I3	trocken – hallig	280,17	117	2,39	129,85	13	9,99	4,17	0,000
I4	Raum spricht früh an – Raum spricht spät an	615,33	117	5,26	78,44	13	6,03	1,15	0,327
I5	groß – klein	272,50	117	2,33	58,07	13	4,47	1,92	0,035
I6	gute Akustik – schlechte Akustik	720,83	117	6,16	31,03	13	2,39	0,39	0,972
I7	hoher Raum – niedriger Raum	282,17	117	2,41	84,81	13	6,52	2,70	0,002
I8	breiter Raum – schmaler Raum	398,13	117	3,40	81,88	13	6,30	1,85	0,043
I9	dunkel – hell	369,93	117	3,16	69,82	13	5,37	1,70	0,070
I10	viel Raum – wenig Raum	362,91	117	3,10	128,71	13	9,90	3,19	0,000
I11	unterstützend – nicht unterstützend	542,94	117	4,64	123,27	13	9,48	2,04	0,023
I12	laut – leise	275,08	117	2,35	116,78	13	8,98	3,82	0,000
I13	diffus – klar	534,01	117	4,56	198,45	13	15,27	3,34	0,000
I14	dünn – voll	330,04	117	2,82	121,48	13	9,34	3,31	0,000
I15	weich – hart	572,90	117	4,90	74,78	13	5,75	1,17	0,306
I16	rund – metallisch	509,98	117	4,36	111,67	13	8,59	1,97	0,029
I17	Nachhall klingt weich aus – ... bricht ab	658,29	117	5,63	151,45	13	11,65	2,07	0,021
I18	ausgewogen – unausgewogen	505,38	117	4,32	48,41	13	3,72	0,86	0,594
I19	Klangfarbe ist gleichmäßig – ... fluktuiert	651,49	117	5,57	95,43	13	7,34	1,32	0,212
I20	Nachklang wird schärfer – ... wird dumpfer	564,99	117	4,83	80,95	13	6,23	1,29	0,229

Tabelle D.4.: Ergebnisse der Tukey-Tests je Item nach Wirtz und Caspar (2002, S. 256). Der kritische F-Wert liegt bei bei 1,34 (α = 0,25; df_{nonadd} = 1; df_{bal} = 116) (Bortz und Schuster, 2010).

	Gegensatzpaar	QS_{bal}	QS_{nonadd}	QS_{res}	$F = \frac{MQ_{nonadd}}{MQ_{bal}}$
I1	tragfähig – wenig tragfähig	340,43	7,14	347,57	2,43
I2	ich kann mich gut hören – ... nicht gut hören	415,68	39,95	455,63	11,15
I3	trocken – hallig	279,93	0,24	280,17	0,1
I4	Raum spricht früh an – Raum spricht spät an	615,3	0,03	615,33	0,01
I5	groß – klein	268,87	3,63	272,5	1,56
I6	gute Akustik – schlechte Akustik	720,77	0,07	720,83	0,01
I7	hoher Raum – niedriger Raum	269,71	12,46	282,17	5,36
I8	breiter Raum – schmaler Raum	394,46	3,67	398,13	1,08
I9	dunkel – hell	365,86	4,07	369,93	1,29
I10	viel Raum – wenig Raum	346,04	16,87	362,91	5,65
I11	unterstützend – nicht unterstützend	538,75	4,19	542,94	0,9
I12	laut – leise	271,22	3,86	275,08	1,65
I13	diffus – klar	471,61	62,4	534,01	15,35
I14	dünn – voll	322,48	7,56	330,04	2,72
I15	weich – hart	560,14	12,76	572,9	2,64
I16	rund – metallisch	509,93	0,05	509,98	0,01
I17	Nachhall klingt weich aus – ... bricht ab	647,91	10,38	658,29	1,86
I18	ausgewogen – unausgewogen	497,15	8,23	505,38	1,92
I19	Klangfarbe ist gleichmäßig – ... fluktuiert	641,58	9,91	651,49	1,79
I20	Nachklang wird schärfer – ... wird dumpfer	564,89	0,09	564,99	0,02