Technische Universität Berlin

Fakultät I Institut für Sprache und Kommunikation Fachgebiet Audiokommunikation

> Einsteinufer 17c 10587 Berlin http://www.akt.tu-berlin.de



Masterarbeit für die Prüfung zum Master of Science im Studiengang Audiokommunikation und -technologie an der Technischen Universität Berlin, Fakultät I - Geistes und Bildungswissenschaften

Entwicklung eines Verfahrens zur Erfassung der raumakustischen Hörexpertise

Markus Martin von Berg

16.03.2020

Erstgutachter Prof. Dr. Stefan Weinzierl

Zweitgutachter Prof. Dr. Jochen Steffens

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.
Berlin, den 16.03.2020
(Unterschrift)

Abstract

Listening tests related to room acoustics are often conducted with either people with a professional or academic background in the subject, musiscians, or people with general interest in acoustics. Those people are supposed to have a certain expertise which makes them preferrable assessors for the tests. However, there does not seem to be a clear definition of this expertise and it is questionable if knowledge on room acoustics does improve the performance in a sensory experiment. An exploratory study was conducted to measure the listening expertise in room acoustics, which is defined as the ability to precisely perceive the sound of a room and hence draw conclusions about this room's properties. The test consisted of several listening tests, questions on experience and knowledge related to room acoustics and three components of the Gold-MSI to determine the musical sophistication. The test was presented to a listener panel of both experts and laymen with no specific knowledge about room acoustics in Germany and Great Britain. The results showed three latent factors representing practical experience, theretical knowledge and analytic listerning behaviour and another overall factor combining those three. Since the latter had a weak but significant impact on the correct conclusion about a room's size based on its sound this factor is considered to be the ability defined as listing experience in room acoustics. Yet, the measurement of this listening experience seemed to be not that precise yet and deserves further refinement in the future.

Zusammenfassung

Bei der Durchführung von Hörversuchen im raumakustischen Kontext wird oft angegeben, dass Probanden ausgewählt wurden, die sich beruflich und akademisch mit dem Thema auseinandersetzen, Musiker sind oder sich schlicht dafür interessieren. Diesen Personen wird dabei eine gewisse Expertise zugeschrieben, die sie zu geeigneten Versuchsteilnehmern macht. Die Definition dieser Expertise ist nicht nur uneinheitlich, es stellt sich auch die Frage, inwiefern Fachkenntnisse für die Teilnahme an sensorischen Experminenten qualifizieren. In einer explorativen Studie wurde ein Verfahren entwickelt, um die raumakustsiche Hörexpertise einer Person zu erfassen. Raumakustische Hörexpertise meint dabei die Fähigkeit Raumklang detailliert wahrzunehmen und auf Basis dieser Eindrücke Aussagen über die Beschaffenheit dieses Raumes treffen zu können. Dazu wurden verschiedene Hörtests, Fragen zur Fachlichen Expertise und Erfahrung mit dem Thema Raumakustik und drei Faktoren des Gold-MSI zur Bestimmung der Musikalität im weitesten Sinne in einem Versuch zusammengestellt. Getestet wurde mit einer Stichprobe aus sowohl Laien, als auch Sachverständigen an unterschiedlichen Orten in Deutschland und Großbritannien. Die Auswertung ergab drei latente Faktoren, die die Kompetenzen praktsiche Routine, Fachwissen und kritisches Hören widerspiegeln und einen weiteren, diesen dreien übergeordneten Faktor. Es konnte ein schwacher Einfluss der Ausprägung dieses Faktors auf die korrekte Zuordnung einer klanglichen zu einer physikalischen Eigenschaft eines Raumes festgestellt werden, weshalb dieser als Ausdruck der vermuteten raumakustischen Hörexpertise gewertet wird. Die Erfassung dieser Expertise war noch etwas unpräzise und es wird weitere Forschung zur genaueren Bestimmung empfohlen.

Danksagung

Zunächt möchte ich Jochen Steffens, Stefan Weinzierl und Daniel Müllensiefen für die Betreuung dieser Arbeit und umfassende Beratung und Unterstützung danken. Damit einher geht der Dank an Marc Voigt, Fabian Brinkmann, Henrik von Coler und alle Mitarbeiter des elektronischen Studios der TU Berlin, Siegbert Versümer, Jörg Becker-Schweizer und den Mitarbeitern des ISAVE der HS Düsseldorf, Bruno von Berg, Olaf Koch, Simon Zander und Matthias Zons vom AEG Kaarst und Nicolas Ruth und Shadan Salami vom Goldsmiths, University of London, die die technische Ausrüstung sowie die zeitlichen und räumlichen Gegebenheiten für die Versuchsdurchführung an vier verschiedenen Orten bereitgestellt haben. Danken möchte ich außerdem Florian Schlüter für die Unterstützung im Umgang mit HTML und Javascript und Jonathan Krebs für die Beratung an den verschiedensten Stellen.

Vor allem aber gilt mein Dank den fünf Vortestern und 100 Probanden, die an diesem Versuch teilgenommen und reges Interesse an dieser Arbeit gezeigt haben, in den meisten Fällen ohne selbst irgendeinen Nutzen davon zu haben.

Inhaltsverzeichnis

Ι.	Einie	• • •	1
	1.1.	Motivation	1
	1.2.	Aufbau und Struktur dieser Arbeit	2
2.		or outdoing - or annual agent	4
	2.1.		4
		2.1.1. Psychoakustische Grundlagen	4
		2.1.2. Physikalische Parameter der Raumakustik	5
		2.1.3. Raumklang als Störschall	7
			7
			8
	2.2.	Raumakustische Hörexpertise	
		Hörversuche	
	2.0.	2.3.1. Zielsetzungen von Hörversuchen	
		2.3.2. Versuchsverfahren	
		2.3.3. Eignung einer Person für die Teilnahme an Hörversuchen	
	2.4.	Probabilistische Testtheorie	
	2.4.		
		2.4.2. Vorteile der probabilisitschen Testtheorie	
		2.4.3. Kalibrierung der Test-Items	
	0.5	2.4.4. Explorative Faktoranalyse und probabilistische Testtheorie	
	2.5.	Virtuelle Akustik	
		2.5.1. Raumsimulation	
		2.5.2. Binauralsynthese	2
2	Mat	hodik 2	2
J.		Hörtests	
	J.1.		
		• •	
	0.0	3.1.2. Definition der Schwierigkeitsgrade	
	3.2.	Teststimuli	
		3.2.1. Aufnahmen	
		3.2.2. Raumsimulationen	
	3.3.	Korrelate	
		3.3.1. Fachliche Expertise und Erfahrung	
		3.3.2. Gold-MSI	
	3.4.	Versuchsaufbau	4
		3.4.1. Struktur des Versuchs	4
		3.4.2. Implementierung in LimeSurvey	5
		3.4.3. Kopfhörerwiedergabe	8
		3.4.4. Einmessung der Abhörpegel	0
	3.5.	Versuchsdurchführung	2
		3.5.1. Technische Universität Berlin	2
		3.5.2. Hochschule Düsseldorf	
		3.5.3. Albert-Einstein-Gymnasium Kaarst 5	

		3.5.4.	Goldsmiths, University of London	54
4.	Vers	uchserg	gebnisse und Auswertung 5	6
	4.1.	Korrek	turen und Anpassungen	66
	4.2.	Stichpi	robe	59
		4.2.1.	Erfahrung mit Raumakustik	59
		4.2.2.	Gold-MSI	59
	4.3.	Kalibri	erung und Berechnung der Fähigkeitsgrade	31
		4.3.1.	Vorauswertung der Antworten	31
		4.3.2.	Kalibirierung der Items	35
		4.3.3.	Fähigkeitsausprägung	37
	4.4.	Faktor	analyse aller Testkomponenten	70
		4.4.1.	Explorative Faktoranaylse	71
		4.4.2.	Prädiktion der Größeninterpretation	73
		4.4.3.	Inhaltliche Auswertung	74
5.	Disk	ussion	7	' 6
	5.1.	Testver	rfahren	76
	5.2.	Faktor	analyse und Multiple Regression	78
	5.3.		· · ·	79
6.	Fazi	t und A	usblick 8	31
Lit	teratu	ırverzei	chnis 8	32
Δr	nhang		g	88
Λ.	A.			38
	11.	A.1.		38
		A.2.	0	38
		A.3.	8	38
		A.4.	0)1
		A.5.)3
	В.		8)6
	D. С.		0)7
	\sim .	1100000		•

Abbildungsverzeichnis

2.1.	Psychometrische Funktion
	IRT-Modellparameter
3.1.	Versuchsinterface: raumakustische Wahrnehmungsfähigkeiten
3.2.	Versuchsinterface: Wiedererkennen des Raumes
3.3.	Versuchsinterface: Raumgröße vs. Nachhallzeit
3.4.	Nachhallspektren der ersten beiden Hörtests
3.5.	Spektren der Ausgangsstimuli
3.6.	Initiale Räume
3.7.	Initiale Nachhallzeiten
3.8.	Approximationsunschärfen
3.9.	Lautheiten der Stimuli
3.10.	Frequenzgänge der Kopfhörer und Entzerrungsfilter
3.11.	$Messaufbauten \dots \dots$
3.12.	Versuchsaubau TU Berlin
3.13.	Versuchsaubau HD Düsseldorf
3.14.	Versuchsaubau AEG Kaarst
3.15.	Versuchsaubau Goldsmiths
4.1.	Impulsantworten und Nachhallzeiten: Problem im ersten Hörtest
4.2.	Impulsantworten und Nachhallzeiten: Problem im zweiten Hörtest 57
4.3.	Antworten: zweiter Hörtest, variable Nachhallklangfarbe vor Problemlösung \dots 58
4.4.	Erfahrung und Gold-MSI
4.5.	Erfahrung und Gold-MSI in Abhähngigkeit der Institution 61
4.6.	Verteilung der Antworten im ersten Hörtest
4.7.	Erfolgsquoten: Wiedererkennen des Raumes
4.8.	Erfolgsquoten: Raumvolumen vs. Nachhallzeit
4.9.	Erfolgsquoten: Raumakustisches Wissen
4.10.	Ermittelte Itemparameter
4 11	Fähigkeitsgrade 70

1. Einleitung

1.1. Motivation

Schon vor zwei Jahrtausenden formulierte der römische Architekt Vitruv Regeln, um die Akustik von Amphitheatern zu verbessern (vgl. Weinzierl, 2008, S. 185). Damit wurde noch ein eher spezifischer "Raumtyp' behandelt, während heutzutage die Wirkung von Akustik in einem breiten Spektrum von räumlichen Umgebungen, Anwendungszwecken und Wahrnehmungsaspekten untersucht wird. Wie schon bei Vitruv beschäftigt sich die Forschung einerseits mit der Frage, welche akustischen Eigenschaften für die Präferenz von Konzertsälen ausschlaggebend sind (vgl. Lokki et al., 2012, S. 3148). Andrerseits wird auch die akustische Wahrnehmung von Raumeigenschaften (vgl. Hameed et al., 2004, S. 1), die Bestimmung der eigenen Position im Raum (vgl. Shinn-Cunningham, 2003, S. 1f) oder etwa der Umgang der neuronalen auditorischen Verarbeitung mit räumlichen Umgebungen als akustische Modulationssysteme behandelt (vgl. Devore et al., 2009, S. 123), (vgl. Slama und Delgutte, 2015, S. 4452) (vgl. Lavandier und Culling, 2008, S. 2237). Im Bereich der virtuellen Akustik steht hingegen die Authenzität der Simulationen räumlicher Umgebungen im Vordergrund (vgl. Brinkmann et al., 2019, S. 2746), (vgl. Cabrera et al., 2006, S. 417). Wie aber kann die Wirkung einer akustischen Eigenschaft eines Raumes auf die vielfältigen Aspekte der menschlichen Wahrnehmung erfasst werden?

In der DIN EN ISO 3382-1 (2009) wird eine Reihe von physikalischen, raumakustischen Parametern angegeben, die sich an perzeptiven Dimensionen der menschlichen Wahrnehmung orientieren (siehe Abschnitt 2.1.2 und 2.1.5). Allerdings können diese Parameter nicht alle Aspekte der menschlichen Wahrnehmung abbilden (vgl. Lokki et al., 2012, S. 3148) und generell dienen diese nur als Prädiktoren für perzeptive Eindrücke (vgl. Weinzierl und Vorländer, 2015, S. 47). Hörversuche sind also unerlässlich, um die auditorische Wahrnehmung vollständig erfassen und untersuchen zu können (vgl. Ellermeier et al., 2008, S. 7). Dabei spielt oftmals die Erfahrung oder Expertise der Versuchspersonen eine wichtige Rolle (vgl. Wickelmaier und Choisel, 2005, S. 1). Nach Scriven (2005) sind Experten in der Lage, die Ausprägung und Eigenschaften sensorischer Empfindungen objektiv zu beurteilen, während Laien "nur"eine bewertete Einschätzung abgeben können, also ob die Empfindung unangenehm oder angenehm ist (vgl. S. 528). Experten beurteilen Reize demnach einheitlicher und liefern damit eher reliable Ergebnisse (Wickelmaier und Choisel, 2005, vgl. [S. 1]. Und tatsächlich scheint es offensichtlich, dass Raumakustikexperten auf Basis ihres Wissen perzeptive Dimensionen wie z.B. Halligkeit semantisch besser verstehen und daher ihrem Höreindruck besser zuordnen können (vgl. Weinzierl und Vorländer, 2015, S. 42). Damit ist jedoch noch keine Aussage über die Leistungsfähgikeit des Gehörs und des auditorischen Systems der jeweiligen Person getroffen, was gerade bei psychoakustischen Experimenten doch ein relevanter Einflussfaktor zu sein scheint (vgl. Bech, 1992, S. 591).

Was macht also einen Expertenhörer aus? Zacharov und Lorho (2006) stellten fest, dass es im Bereich der auditorischen Forschung in der Tat keine einheitliche Verwendung von Begriffen wie "Experte' und "Laie' gibt (S. 2f), während etwa in der Lebensmittelindustrie mit der DIN EN ISO 8586 (2014) ein internationaler Standard für die Qualifikation von Versuchspersonen existiert. Um ein paar Beispiele zu nennen, Fleischmann et al. (2012) bezeichnen beim Vergleich von Kopfhörer- und Lautsprecherwiedergabekonfigurationen Probanden mit "Hintergrund als Toningenieure" (S. 5) als die einzigen Experten, die diese Systeme effizient vergleichen könnten, ohne eine Begründung dafür oder eine nährere Erläuterung zu geben. Lokki et al. (2012) ver-

weisen bei der Untersuchung des Zusammenhang zwischen raumakustischen Eigenschaften und der Präferenz von Konzertsälen darauf, dass es "obligatorisch" (S. 3151) sei, Versuchsteilnehmer umsichtig auswählen um aussagekräftige Ergebnisse zu garantieren. Ihrer Erfahrung nach seien regelmäßige Konzertgänger und aktive Hörer klassischer Musik "motivierte und gute Kandidaten" (Lokki et al., 2012, S. 3151), auch wenn die Kandidaten vor der Teilnahme audiometrische Tests absolvieren mussten. Wickelmaier und Choisel (2005) hingegen stellten fest, dass die Selbstauskunft der Probanden über musikalisches Interesse und Erfahrung kein geeigneter Prädiktor für die Ergebnisse nachfolgender Tests der Diskriminationsfähigkeit von Stimuli mit variabler stereofoner Breite war (vgl. Wickelmaier und Choisel, 2005, S. 6). Nach Zacharov und Lorho (2006, S. 16) lassen sich auch Verfahren und Konzepte aus der DIN EN ISO 8586 (2014) auf den Kontext des räumlichen Hörens übertragen, eine konkrete Anwendung für diesen Kontext steht jedoch noch aus. Auch zehn Jahre nach deren Publikation forderte Neidhardt (2016) eine einfache Methode, um die Qualifikation und Fähigkeit von Probanden einzuschätzen (S. 9).

Diese Arbeit zielt darauf ab, diese Lücke zu füllen uned ein passendes Verfahren zu entwickeln, um den Expertisegrad einer Person im Hören von raumakustischen Informationen, im Folgenden als raumakustische Hörexpertise bezeichnet, zu erfassen. Was genau diese Hörexpertise ausmacht und welche kognitiven Fähigkeiten damit einhergehen, wird dazu auf Basis bisheriger Forschung zur raumakustischen Wahrnehmung erarbeitet und in Abschnitt 2.2 definiert. Über einen Hörversuch sollen Aspekte der perzeptiven und kognitiven Wahrnehmung abgefragt werden, parallel zur fachlichen Bildung und Erfahrung der Testperson und ihrem Verhältnis zur Musik, um diese 'klassischen'Expertisekriterien der gemessenen Erkennungsleistung des auditorischen Systems gegenüber zu stellen. Die Analyse der statistischen Zusammenhänge zwischen perzeptiven und kognitiven Fähigkeiten und fachlicher Bildung und Musikalität soll zeigen, ob tatsächlich ein gemeinsamer, latenter Faktor die Performanz in diesen Teilgebieten beeinflusst. Das wäre im Sinne dieser Arbeit als Maß der raumakustischen Hörexpertise zu verstehen und kann Einblicke geben, ob Erfahrung und Interesse an Musik oder Akustik sowie Tontechnik tatsächlich als Prädiktoren für die Leistung in den Hörtests angesehen werden können. Durch diese Analyse soll auch erkannt werden, welche der untersuchten Eigenschaften redundante Ergebnisse liefern, um den Test zu verkürzen und im Idelafall im Sinne Neidhardts die Grundlage für einen schnellen einfachen Test zu schaffen, der reliable Aussagen über die individuelle Hörexpertise einer Person liefert.

Da bei dieser Untersuchung methodisch nach der probabilistischen Testtheorie vorgegangen wird, müssen in diesem Versuch zunächst die Testgrößen der einzelnen Hörtests kalibiert werden, um herauszufinden, in welchen quantitativen Dimensionen sich die abgefragten Erkennungsleistungen bewegen. Sollte sich eine messbare Hörexpertise abzeichnen, kann mit Hilfe dieser kalibrierten Hörtests und nach der Untersuchung möglicher redundanter Parameter im Versuch ein ggf. gekürztes, adaptives Verfahren konzipiert werden.

1.2. Aufbau und Struktur dieser Arbeit

Zunächst muss der Begriff der raumakustischen Hörexpertise als theoretisches Konstrukt definiert werden muss. Daher werden im folgenden Kapitel ein paar Grundlagen aus der Psychoakustik und der physikalischen Schallausbreitung im Raum zusammengefasst. Darauf basierend werden Aspekte des räumlichen Hörens beleuchtet, die von einer Hörexpertise beeinflusst werden könnten, um anschließend das Konstrukt der raumakustischen Hörexpertise zu formulieren. Da die Idee raumakustischer Hörexpertise der Frage nach der Eignung einer Person für die Teilnahme an Hörversuchen entspringt, werden auch hier relevante Grundlagen zu Versuchsverfahren und Anforderungen an Probanden diskutiert. Als theoretische Basis für die Entwicklung des Hörversuchs wird dann am Ende des Theorieteils die probabilistische Testtheorie als Paradig-

ma des Versuchs und Raumsimulation und Binauralsynthese als Themengebiete der virtuellen Akustik betrachtet, die für die Umsetzung der Hörtests ausschlaggebend sind.

Anschließend wird im Methodenteil die Konzeption und Umsetzung des Hörversuchs zur Kalibrierung der Testverfahren beschrieben. Nach der Entwicklung der Hörtest wird die Erzeugung der Stimuli diskutiert, gefolgt von den weiteren Teilfragen des Versuchs, der Implementierung und dem Versuchsaufbau. Der Versuch wurde an verschiedenen Institutionen durchgeführt, weshalb auch die Erläuterungen der Versuchsdurchführung entsprechend aufgeteilt sind.

In der Beschreibung der Versuchsergebnisse werden zunächst die Stichproben vorgestellt und einige Korrekturen an den Stimuli und Hörtests diskutiert, deren Notwendigkeit erst festgestellt wurde, nach dem die Ergebnisse der ersten Teilnehmer vorlagen. Unter Berücksichtigung dieser Korrekturen folgt die Auseinandersetzung mit den Versuchsergebnissen und die Analyse latenter Faktoren. Nach einer Diskussion der Ergebnisse und der entwickelten Verfahren schließt die Arbeit mit einem Fazit und dem Ausblick für weitere Forschung.

2. Theoretische Grundlagen

2.1. Raumakustisches Hören

Um ein Konstrukt raumakustischer Hörexpertise entwickeln zu können, werden in den folgenden Abschnitten Faktoren der Raumakustik und Hörwahrnehmung betrachtet, anhand derer eine solche Expertise beobachtet werden könnte. Diese Überlegungen schließen auch eine Identifikation von Aspekten des raumakustischen Hörens ein, die im Zusammenhang mit Hörexpertise keine Rolle spielen.

2.1.1. Psychoakustische Grundlagen

Ebenso wie die Wahrnehmung von Musik oder Sprache unterliegt auch die Wahrnehmung von Raumklang den Übertragungseigenschaften und -Mechanismen des menschlichen Gehörs, von denen einige kurz erläutert werden sollen. Diese Eigenschaften des Gehörs sind teilweise weniger für die raumakustische Hören als solches relevant, sondern mehr für die Durchführung und Auswertung von Hörversuchen, deren Methodik in Abschnitt 2.3 betrachtet wird.

Lautheitswahrnehmung

Diese Frequenzselektivität des Gehörs (vgl. Weinzierl, 2008, S. 55) führt dazu, dass Töne verschiedener Frequenz unterschiedliche subjektive Lautstärkeeindrücke, gemeinhin als Lautheit bezeichnet (vgl. Weinzierl, 2008, S. 59), erzeugen. Diese Differenzen der Lautheit bei zwei Tönen gleicher Amplitude nehmen mit ansteigendem Schalldruck ab, sodass bei hohen Schalldruckpegeln, der Lautstärkeeindruck über das hörbare Frequenzspektrum ähnlicher wird (vgl. Weinzierl, 2008, S. 55). Dieser Effekt wird durch Frequenzberwertungen bei der Berechnung von Schalldruckpegeln abzubilden versucht, die den Schalldruck in den verschiedenen Frequenzbändern unterschiedlich stark gewichten. Üblicherweise werden die A-, B- und C-Bewerung verwendet, wobei die A-Bewertung an für niedrige und die C-Bewertung für hohe Schalldruckpegel konzipiert sind, weshalb die Gewichtung tiefer Frequenzen von A nach C zunimmt (vgl. Weinzierl, 2008, S. 56). Auch Unterscheidungsschwellen liegen bei höheren und tieferen Frequenzen höher als im Bereich der maximalen Sensibilität (vgl. Fastl und Zwicker, 2007, S. 183).

Wahrnehmungsschwellen und Verdeckung

Da das Gehör letzten Endes auf mechanische Schwingungen von Membranen im Außen- und Innenohr und auf die der Gehörknöchelchen im Mittelohr reagiert, die alle eine gewisse Massenträgheit aufweisen, sind alle Aspekte der Hörwahrnehmung in ihrer Sensibilität begrenzt. Schalldrücke unterhalb einer bestimmten Hörschwelle können nicht wahrgenommen werden. Die Frequenzabhängigkeit ergibt sich aus der oben beschrieben frequenzselektiven Sensibilität des Gehörs (vgl. Weinzierl, 2008, S. 53f). Es gibt jedoch auch eine situationsabhängige Mithörschwelle, die beschreibt, ab wann ein Schallignal, das gleichzeitig mit anderen Geräuschen präsentiert wird, parallel zu diesen gehört werden kann. Ein Schallsignal kann ein anderes verdecken, d.h. ,übertonen', wenn es z. B. im gleichen oder einem benachbarten Frequenzbereich höhere Intensitäten aufweist. Dabei verdecken tiefe Frequenzen höhere eher als anders herum (vgl. Fastl und Zwicker, 2007, S. 64f).

Lokalisation von Schallquellen

Die Lokalisation der Einfallsrichtung eines Schallereignisses erfolgt anhand mehrerer Mechanismen. Über interaurale Laufzeit- und Pegeldifferenzen wird ermittelt, ob die Schallquelle sich rechts oder links von der Person befindet (vgl. Weinzierl, 2008, S. 96ff). Die Lokalisation auf der vertikalen Achse erfolgt mit Hilfe des Außenohres, das einfallenden Schall in Abhängigkeit von dessen Einfallsrichtung in einer spezifischen Weise filtert, die vom auditorischen System als Richtungsinformation erkannt wird (vgl. Weinzierl, 2008, S. 95). Die auditive Wahrnehmung von Entfernung funktioniert je nach Entfernung nach unterschiedlichen Prinzipien. Bei sehr kleinen Entfernungen bis 25 cm führt auch die Entfernung zu merklichen Änderungen in der Ohrübertraungsfunktion, die als entfernungsbedingt erkannt werden. Bei Entfernung bis 15 m bewertet das Gehör das Verhältnis vom Direktschall der Quelle zum Diffusschall der Raumreflexionen und, im Falle bekannter Schallereignisse und -Intensitäten, den Pegelabfall durch den Übertragungsweg (vgl. Weinzierl, 2008, S. 98). Ab einer Entfernung von mehr als 15 m kann nach Weinzierl (2008, S. 99) das auditorische System sensorisch keine Differenzen mehr erkennen, bei bekannten Schallsignalen jedoch durch Wissen und visuelle Informationen größere Entfernungen ebenfalls unterscheiden.

Ermüdung

Weniger für das raumakustische Hören selbst, als eher für die Konzeption und Durchführung von Hörversuchen ist das Phänomen der Hörermüdung relevant. Die Hörermüdung beschreibt das Verhalten des Gehörs, dass ein akustischer Stimulus mit hohem Schalldruckpegel mit der Zeit als leiser empfunden wird (vgl. Florentine et al., 2011, S. 93). Zusätzlich zu dieser Erschöpfung des Hörorgans kann konzentriertes Hören auch auf mentaler Ebene zu Ermüdungserscheinungen führen, deren Erforschung sich bisher allerdings vor allem auf Menschen mit Hörschädigungen, bei denen diese Ermüdung häufiger auftritt, konzentriert und durch uneinheitliche Methodik und Konstrukte zur Erfassung von Höranstrengung erschwert wird (vgl. McGarrigle et al., 2014, S. 2).

2.1.2. Physikalische Parameter der Raumakustik

Jegliche Wahrnehmung von Raumakustik basiert letzten Endes auf physikalischen Prozessen, die bei der Schallausbreitung im Raum entstehen. Durch die im vorigen Abschnitt beschriebenen Begrenzungen des Hörmvermögens, fallen diese Prozesse bei der Wahrnehmung mehr oder weniger ins Gewicht. Die Wirkung der physikalischen Raumakustik auf die menschliche Wahrnehmung werden in den Abschnitten 2.1.3, 2.1.5 thematisiert, hier wird vorab die physikalische Schallausbreitung im Raum als solche betrachtet.

Nachhallzeit und Anfangsnachhallzeit

Der älteste und wohl elementarste Parameter der Raumakustik beschreibt lediglich die Dauer, bis nach Ende der Schallabstrahlung durch eine Quelle der Schallpegel im Raum um 60 dB abgefallen ist (vgl. Weinzierl, 2008, S. 188). Die Nachhallzeit kann mit relativ wenig Aufwand gemessen und einfach berechnet werden und es gibt zudem zahlreiche Referenzmessungen bekannter Konzertsäle (vgl. Fuchs, 2018, S. 32).

Nach Carl. F. Eyring berechnet sich die Nachhallzeit T_{60} als

$$T_{60} = 0,163 \cdot \frac{V}{-\ln(1-\alpha)S + 4mV}$$
 (2.1)

wobei V und S Raumvolumen bzw. -Oberfläche, α den mittleren Absorptionsgrad der Wandoberflächen und m die Energiedämpfungskonstante der Luft bezeichnen (vgl. Weinzierl, 2008, S. 190).

Dabei ist die Nachhallzeit ebenso wie die Absorptionsgrade der Oberflächen frequenzabhängig. Auch die Luft als Übertragungsmedium wirkt absorbierend, was jedoch nur bei großen Raumvolumina und sehr hohen Frequenzen merkliche Auswirkungen hat (vgl. Möser, 2005, S. 193f). Für geringe Absorptionsgrade kann die einfachere Formel

$$T_{60} = 0,163 \cdot \frac{V}{\alpha \cdot S} \tag{2.2}$$

angewendet werden, die auf Wallace C. Sabine, den Entdecker der Nachhallzeit, zurückgeht (vgl. Möser, 2005, S. 193). Die Messung der Nachhallzeit ist in der DIN EN ISO 3382-1 beschrieben, in der DIN 18041 finden sich empfohlene Nachhallzeiten in Abhängigkeit des Anwendungszwecks und der Raumgröße. Die Nachhallzeit bildet die Basis der meisten den nachfolgenden physikalischen Raumakustikparameter.

Einer der aus der Nachhallzeit abgeleiteten Parameter ist die frühe Abklingzeit oder Anfangsnachhallzeit (auch early decay time bzw. EDT), die den Zeitraum bis zum Abfall des Schalldruckpegels umd 10 dB beschreibt (vgl. Weinzierl, 2008, S. 188). Die Anfangsnachhallzeit gilt als aussagekräftigerer in Bezug auf die subjektive Einschätzung der Nachhalldauer (vgl. Weinzierl, 2008, S. 188).

Stärkemaß

Durch die Reflexionen an den Wänden werden in Räumen Energieanteile der Schallerzeugung an einen Empfänger geleitet, die diesen im Freifeld nicht erreichen würden, weil sie von der Quelle eigentlich nicht in dessen Richtung abgestrahlt werden. Diese Schallanteile treffen wegen des Reflexionsweges zwar mit geringer zeitlicher Verzögerung am Empfänger ein, bewirken aber dennoch ab einer gewissen Signallänge effektiv eine Verstärkung des von der Quelle abgestrahlten Schallsignals. Die Verstärkung wird quantitativ mit dem Stärkemaß erfasst.

Dazu wird der aufsummierte Schalldruckpegel, der bei einer kugelförmig abstrahlenden Quelle im Raum am Empfänger gemessen wird, ins Verhältnis gesetzt zu dem aufsummierten Schalldruckpegel, den die gleiche Quelle in 10 m Entfernung im Freifeld erzeugt (vgl. Weinzierl, 2008, S. 201). Die Berechnung ist nachDIN EN ISO 3382-1 definiert als

$$G = 10 \cdot \lg \int_0^\infty \frac{p^2(t)}{p_{10}^2(t)} dt$$
 (2.3)

Dabei entspricht p(t) dem gemessenen Schallsignal im Raum und $p_{10}(t)$ dem Referenzsignal der gleichen Quelle im Freifeld in 10 m Entfernung.

Klarheitsmaß und Deutlichkeitsmaß

Das Klarheits- und Deutlichkeitsmaß Aufschluss über die zeitliche Balance einer Raumpimpuslantwort. Die Parameter repräsentieren das energetische Verhältnis zwischen dem frühen und dem späten Teil der Impulsantwort p(t) (vgl. DIN EN ISO 3382-1, 2009, S. 19):

$$C_{t_e} = 10 \cdot \lg \frac{\int_0^{t_e} p^2(t) dt}{\int_{t_e}^{\infty} p^2(t) dt}$$

$$\tag{2.4}$$

Als Grenzwert zwischen früher und später Impulsantwort t_e wird entweder 50 ms für das Deutlichkeitsmaß oder 80 ms für das Klarheitsmaß gewählt: Das erste ist an der Verständlichkeit von Sprache, letzteres an der von Musik orientiert (vgl. DIN EN ISO 3382-1, 2009, S. 19). Den beiden Parametern liegt die Annahme zugrunde, dass frühe Reflexionen durch verstärkende Wirkung den Verständlichkeit eines Schallsignals für den Empfänger fördern, während spätere Reflexionen diese eher beeinträchtigen (vgl. Weinzierl, 2008, S. 192). Die Rolle von Raumreflexionen als Störsignal wird in Abschnitt 2.1.3 näher betrachtet.

Seitlicher Schalleinfall, IACC und Schwerpunktzeit

Nicht nur der Anteil an abgestrahlter Schallenergie, der bei der Schallausbreitung im Raum an den Wänden reflektiert wird, kann sehr unterschiedlich ausfallen, sondern auch die Verteilung dieser Reflexionen im Raum. Das Gefühl, vom Hall eines Raumes umhüllt zu werden, scheint bei der Bewertung von Raumakustik eine wesentliche Rolle zu spielen (vgl. Lokki et al., 2012, S. 3160). Daher beschreiben Messgrößen für den Seitenschallgrad den Anteil an seitlich einfallender Schallenergie (vgl. DIN EN ISO 3382-1, 2009, S. 20f) und der interaurale Kreuzkorrelationskoeffizient (IACC) das Verhältnis der am linken und am rechten Ohr eintreffenden Impulsantworten (vgl. Weinzierl, 2008, S. 199).

Ein weiteres Maß für den Eindruck von Räumlichkeit ist die Schwerpunktzeit, die als das erste Moment der quadrierten Impulsantwort definiert ist (vgl. Weinzierl, 2008, S. 197):

2.1.3. Raumklang als Störschall

Da ein Schallsignal bei der Übertragung durch einen Raum mit zeitverzögerten Reflexionen überlagert und dadurch in seiner zetilichen und spektralen Form verändert wird, stellt der Raumklang technisch gesehen ein Störsignal dar (vgl. Shinn-Cunningham, 2003, S. 2). Die in Abschnitt 2.1.1 behandelten Lokalisation durch interaurale Zeitdifferenzen wird beispielsweise durch Raumklang beeinträchtigt (vgl. Darwin, 2008, S. 5), aber auch die Sprachverständlichkeit kann durch Nachhall beeinträchtigt werden (vgl. Lavandier und Culling, 2008, S. 2237).

Generell scheint die menschliche Hörwahrnehmung jedoch nicht nur robust gegen die Verzerrungen eines Signals durch Raumklang zu sein, sondern von diesen sogar profitieren zu können (vgl. Shinn-Cunningham, 2003, S. 1). Devore et al. (2009) konnten bei Messungen an Katzen feststellen, dass viele Neuronen in den Colliculi inferiores, dem auditorischen Mittelhirn, eine erhöhte Aktivität in den ersten Millisekunden nach Einsetzen der Anregung durch einen Stimulus aufweisen (vgl. S. 123). In diesem Zeitfenster ist der Direktschall noch relativ unverfälscht durch zeitverzögerte Reflexionen, die zusätzliche, widersprüchliche Richtungsinformationen enthalten, sodass auch zumindest in moderat halligen Umgebungen Schallquellen geortet werden können (vgl. Devore et al., 2009, S. 123). Die Tatsache, dass diese Mechanismen nur bei einigen Neuronen beobachtet wurden (vgl. Devore et al., 2009, S. 125), macht es im Zusammenhang mit raumakustischer Hörexpertise denkbar, dass es Personen gibt, deren auditorische Wahrnehmung durch die Eigenschaften der Neuronen im Colliculus inferiores robuster gegen die Beeinträchtigung der direktionalen und zeitlichen Informationen eines Schallsignals durch Raumklang ist, als bei anderen. Da in dieser Arbeit die Fähigkeit, Raumklang zu interpretieren im Vordergrund stehen soll und nicht, einen Direktschall trotz eines vorhandenen Raumklangs korrekt zu interpretieren, wird die Frage nach einer individuellen Resistenz gegen die Verfälschung eines Schallsignals durch Reflexionen nicht weiter untersucht, wohl aber die Fähigkeit, Raumklang und Direktschall auf kognitiver Ebene voneinander zu trennen (siehe Abschnitt 3.1.1).

2.1.4. Lokalisation im Raum

Lokalisation im Raum kann zum einen die Ortung von Schallquellen relativ zur eigenen Position in einer halligen Umgebung bedeuten. Dass diese Lokalisation prinzipiell durch Raumreflexionen beeinflusst wird, wurde bereits im vorigen Abschnitt ebenfalls erläutert. Wie ein empfangenes Schallsignal durch Raumreflexionen verändert wird, hängt aber nicht nur davon ab, ob sich Quelle und Empfänger in einem mehr oder weniger halligen Raum befinden, sondern auch von der konkreten Postion der beiden in diesem Raum. Lokalisation im Raum kann damit auch die Bestimmung der eigenen Position und der Position der Schallquelle im Raum, also relativ zu den Grenzflächen desselben meinen. Befinden sich Quelle oder Empfänger in der Nähe einer oder mehrerer Wände, kommt es zu Kammfilterverzerrungen, bei der in Abhängigkeit des genauen Wandabstands einige Frequenzen konstruktiv und andere destruktiv mit Reflexionen überlagert

werden, wodurch die ersten verstärkt und letztere gedämpft werden (vgl. Shinn-Cunningham, 2003, S. 2).

Wenn das Reflexionsmuster, das ein Empfänger als Raumklang wahrnimmt, positionsabhängig ist, stellt sich die Frage, ob ein Hörer daraus Informationen über seine eigene Position oder die der Schallquelle gewinnen kann und ob sich hier individuelle Expertisegrade feststellen lassen. Untersuchungen haben gezeigt, dass klangliche Veränderungen durch verschiedene Quell- und Empfängerpositionen im Raum zwar wahrgenommen werden, der Mensch jedoch kaum dazu in der Lage ist, aus diesen Unterschieden seine eigene Position oder die der Schallquelle im Raum abzuleiten (vgl. Hameed et al., 2004, S.2), (vgl. Shinn-Cunningham, 2003, S.5). Sie werden eher als Veränderungen der Halligkeit, Klangfarbe oder subjektiven Quellausdehnung interpretiert (vgl. Neidhardt, 2016, S.9). Lediglich die Einschätzung der Entfernung zwischen Quelle und Empfänger scheint nach Shinn-Cunningham (2003, S.3) in halligen Umgebungen einfacher zu sein, als unter nachhallfreien Bedingungen. Das ist jedoch weniger überraschend, da das Verhältnis von Direkt- zu Diffusschall schon in Abschnitt 2.1.1 als Faktor für die Bestimmung der Entfernung einer Schallquelle festgestellt wurde.

Auch wenn nach Shinn-Cunningham (2003, S. 4) starke individuelle Differenzen in der Unterscheidung verschiedener Empfängerpositionen feststellte, die einer persönlichen Hörexpertise der Versuchspersonen geschuldet sein könnten, wird auch die Sicherheit in der Bestimmung von Quell- und Empfängerpositionen im Raum in dieser Untersuchung als Faktor einer raumakustischen Hörexpertise nicht weiter untersucht. Denn, wie oben beschrieben scheint diese Fähigkeit trotz der genannten individuellen Differenzen beim Menschen sehr schwach ausgeprägt, sodass auch mit Blick auf Dauer und Umfang des späteren Hörversuchs anderen Aspekten der Interpretation eines wahrgenommenen Raumklangs der Vorzug gegeben wurde.

2.1.5. Auditorische Parameter des Raumklangs

Die Klangmodulation, die durch einen Raum hervorgerufen wird, kann sich nicht nur auf physikalischer, sondern auch auf perzeptiver Ebene sehr unterschiedlich gestalten und hängt von zahlreichen Eigenschaften des Raumes ab. Die Nachhallzeit z. B. wird nach Gl. 2.1 und Gl. 2.2 sowohl vom Raumvolumen, als auch von den Absorptionsgraden der Oberflächen bestimmt. Anders als in den vorigen Abschnitten stellt sich jetzt die Frage, welche Eigenschaften eines Raumklangs einem Hörer Aufschluss über die Beschaffenheit des Raumes selbst geben und nicht etwa über die eigene Position in demselben. Da der Klang eines Raumes nicht als eigenständiges Geräusch, sondern als eine Modulation eines Ausgangsklangs erfahren wird, muss für die Anaylse dieser Modulation dieselbe vom Ausgangsschall separiert werden. Der Hörer muss zuordnen, welche gehörte Klangeigenschaft die Schallquelle von sich aus verursacht und welche vom Raum als Übertragungssystem herrühren (vgl. Weinzierl und Vorländer, 2015, S. 42).

Es wurden schon einige Parameter vorgestellt, mit denen die physikalische Akustik eines Raumes gemessen und beschrieben werden kann. Diese Parameter sind in der Regel an perzeptiven Qualitäten wie Klangfarbe, Sprachverständlichkeit und Umhüllung orientiert (siehe Abschnitt 2.1.2). Es konnte jedoch vielfach gezeigt werden, dass die proklamierten Zusammenhänge zwischen physikalischen und perzeptiven Parametern, die dem Beispiel von Maempel (2012) folgend als akustische bzw. auditorische Parameter unterschieden werden, nicht immer zutreffend sind (vgl. von Dopr Schuitman et al., 2013, S. 1572). Die Folge sind zahlreiche Untersuchungen darüber, welche Qualitäten die Wahrnehmung eines Raumklangs formen und inwiefern diese durch akustische Eigenschaften des Raumes prädiziert werden können. Bisher konnte dabei kein valides Messinstrument für den gesamtheitlichen, klanglichen Eindruck eines Raumes gefunden werden konnte (vgl. Weinzierl und Vorländer, 2015, S. 44). Stattdessen gibt es eine Vielzahl von Erkenntnissen über einzelne Teilaspekte der Wahrnehmung und Beziehungen zwischen akustischen und auditorischen Parametern. Zur Definition einer möglichen raumakustischen Hörexpertise als eine grundlegende Fähigkeit, sich einen dedizierten Eindruck vom Klang eines Raumes und seiner

Bedeutung zu verschaffen, soll versucht werden, einzelne Befunde und Erkenntnisse zu einem Gesamtbild zusammenzusetzen.

Einerseits beschäftigt sich die bisherige Forschung mit der Herausarbeitung von Faktoren, die konkret bei der Beurteilung und Wahrnehmung von Raumklang eine Rolle spielen - mit anderen Worten, auf welche klanglichen Eigenschaften der Hörer konkret achtet. Nachdem Anfang des zwanzigsten Jahrhunderts im wesentlichen optimale Dimensionen und spektrale Verteilungen der Nachhallzeit gesucht wurden (vgl. Weinzierl und Vorländer, 2015, S. 42), folgten verschiedene Inventare und Kataloge aus beispielsweise vierzehn oder achtzehn definierten Attributen (vgl. Weinzierl und Vorländer, 2015, S. 43). Spätere Untersuchungen ermittelten mit Faktoranalysen oder dem Abgleich verschiedener individueller Vokabulare verschiedene Konstellationen zugrundliegender Faktoren bzw. Attributgruppen. Lehmann und Wilkens (1980) identifizierte die drei Beurteilungskriterien "Stärke und Ausdehnung der Schallquelle", "Empfindung der Deutlichkeit des Gesamtklanges", sowie "Beurteilung des Gesamtklanges bezüglich der Klangfarbe". Mit Hilfe individuell erhobener Vokabularkataloge definierten Lokki et al. (2011) neun Attributgruppen, wobei fast jeder Proband Begriffe mit Bezug zu Halligkeit und Klarheit verwendete. Die Arbeit von Lokki et al. (2012) ergab die drei Faktoren Lautheit, Umhüllung und Halligkeit als Kriterien für die Präferenz von Konzertsälen. In einer aktuelleren Studie definierten Weinzierl et al. (2018) den RAQI'-Katalog (Room Acoustic Quality Inventory), bestehend aus vier bis neun Attributen. Dazu wurde von einer Expertengruppe ein Inventar aus 50 Aspekten erstellt und aus diesem mittels Faktoranalyse die latenten Kriterien "Qualität", "Stärke", "Halligkeit", "Brillianz"und in den längeren Konfigurationen des Katalogs zusätzlich "Unregelmäßiges Abklingen", "Coloration", "Klarheit", "Lebhaftigkeit"und "Intimität"(Weinzierl et al., 2018, S. 1253) bestimmt. Es wird deutlich, dass, auch wenn jede Studie im einzelnen eine andere Konstellation auditorischer Raumakustikparameter ergibt, einige Qualitäten, inbesondere mit Bezug zur Verstärkungswirkung des Raumes, des Eindrucks der Umhüllung und der Halligkeit in allen Fällen relevant erschienen. Hinzu kommt auch die Deutlichkeit bzw. Klarheit, die nach Fuchs (2018) wesentlich mit der spektralen Verteilung, also der Klangfarbe der Nachhallzeit zusammenhängt (vgl. S. 43f).

Um die oben diskutierten auditorischen Parameter nicht nur qualitativ zu erfassen, sondern auch die quantitative Ausprägung verstehen zu können, liegt es nahe, den Zusammenhang solcher explorativ ermittelter Kriterien mit den akustischen Parametern der DIN EN ISO 3382-1 (2009)zu untersuchen, die ja genau diese Kriterien abbilden sollen (vgl. Lokki et al., 2012, S. 3156). Nach Soulodre und Bradley (1995) besteht ein signifikanter Zusammenhang zwischen dem Stärkemaß und dem Lautheitseindruck, der durch eine A-Bewertung des Stärkemaßes noch erhöht wird und darüberhinaus eine Korrelation des Klarheitsmaßes und der Schwerpunktzeit mit der subjektiv empfundenen Klarheit (vgl. S. 299f). Lokki et al. (2012) stellten zwar ebenfalls eine hohe Korrelation der perzeptiven Parameter "Schallquellennähe"und "Bassigkeit"mit dem Stärkemaß in den unteren und mittleren Oktavbändern fest, allerdings auch merkliche Abweichungen des Klarheitsmaßes von dem perzeptiven Parameter "Definition"(vgl. S. 3159). de Vries et al. (2001) zeigte, dass der IACC, der als Maß für die subjektive Schallquellenausdehnung gilt (vgl. S. 947), bei leichten Drehungen des Mess-Kunstkopfes starke Fluktuation aufweist, während die wahgenommene Quellausdehnung jedoch gleich bleibt (vgl. S. 953).

Laut Skålevik (2010) können neben der Halligkeit auch die Lautheit, Umhüllung und Klarheit eines Raumklangs durch die Nachhallzeit prädiziert werden (vgl. S4), während Soulodre und Bradley (1995) in Übereinstimmung mit DIN EN ISO 3382-1 (2009) und Weinzierl (2008, S. 188) die Anfangsnachhallzeit als Prädiktor für den Eindruck von Räumlichkeit bestätigte (S. 300). Auch für die Empfindung von Raumgröße wurde in einer Studie von Hameed et al. (2004) die

Nachhallzeit als essentieller Parameter identifiziert, während die zeitliche Struktur der frühen Reflexionen eine untergeordnete Rolle zu spielen schien (vgl. S. 5). In diesem Zusammenhang ist jedoch anzumerken, dass nach Shinn-Cunningham (2003) kleine hallige Räume leicht von großen trockenen unterschieden werden können (vgl. S. 4), obwohl das Abklingverhalten dieser Räume sich vor allem durch die frühen Reflexionen unterscheidet, die im Fall des kleineren Raumes früher und mit höherer Amplitude am Hörer eintreffen (vgl. Hameed et al., 2004, S. 2).

Die Nachhallzeit ist also nicht nur der älteste raumakustische Parameter, sondern offenkundig auch sehr relevant für die perzeptive Wirkung eines Raumes. Unter diesem Gesichtspunkt liegt die Auseinandersetzung mit der Sensibilität des Gehörs für die Nachhallzeit nahe. Laut DIN EN ISO 3382-1 (2009) liegt die Wahrnehmungschwelle für Änderungen der Anfangsnachhallzeit bei 5 % (vgl. S. 16). Verschiedene nachfolgende Studien ergaben jedoch andere Werte. Die ermittelten Wahrnehmungsschwellen reichen von 3 % bis 39 %, wobei teils mit unterschiedlichen Stimuli und Ausgangsnachhallzeiten gearbeitet wurde (vgl. Belvins et al., 2013, S. 2). In einer jüngeren Untersuchung mit 1 kHz-Oktavbandrauschen und Ausgangsnachhallzeiten zwischen einer und drei Sekunden maßen Belvins et al. (2013) eine mittlere Wahrnehmungsschwelle von 24,5 % (vgl. S. 5). Dabei stellten sie außerdem fest, dass, wenn man sich im Hörtest von größeren Nachhallzeiten der Referenz annähert, die Wahrnehmungsschwelle höher liegt, als wenn der Vergleich mit niedrigeren Nachhallzeiten beginnt (vgl. Belvins et al., 2013, S. 5). Die Wahrnehmungsschwelle für Nachhallzeiten ist nach übereinstimmenden Ergebnissen darüberhinaus unabäagig vom Stimulus, der im Testverfahren zur Bestimmung der Schwelle verwendet wird (vgl. Belvins et al., 2013, S. 5).

2.2. Raumakustische Hörexpertise

Im vorigen Abschnitt wurden wesentliche Aspekte diskutiert, wie die Akustik eines Raumes die menschliche Wahrnehmung eines Schallsignals beeinflusst, das in diesem Raum präsentiertt wird. Daraus soll nun eine Definition raumakustischer Hörexpertise als individuelle Fähigkeit, diese Veränderung des Quellsignals durch den Raum zu verarbeiten, abgeleitet werden. Zunächst wird ein Schallsignal durch die Überlagerung mit zeitverzögerten Reflexionen in seiner zeitlichen und spektralen Form verfälscht. Es gibt offenbar Kompensationsmechanismen dieser Verzerrungen und für die Entfernungseinschätzung ist das Verhältnis des Direktschalls zu dem Reflexionsgemisch als Diffusschall sogar förderlich. Da eine Störresistenz gegen die Veränderungen eines Klangs durch Raumreflexionen jedoch weniger für eine Auseinandersetzung mit den akustischen Eigenschaften des Raumes zu tun hat, wird dieser Faktor hier nicht als Indikator raumakustischer Hörexpertise in Erwägung gezogen.

Außerdem ändert sich der Raumklang in Abhängigkeit der Positionen von Quelle und Empfänger im Raum. Wie in Abschnitt 2.1.4 gezeigt, kann der Mensch diese klanglichen Veränderungen zwar wahrnehmen, daraus jedoch kaum Informationen über Quell- und Empfängerpositionen ableiten. Dementsprechend sind auch bei individuellen Differenzen insgesamt nur geringe Erkennungsleistungen zu erwarten, weshalb auch die Sicherheit der eigenen Lokalisation im Raum als Komponente raumakustischer Hörexpertise an dieser Stelle vernachlässigt wird.

So bleibt zuletzt noch die Komponente, dem Raumklang Informationen über den Raum selbst zu entnehmen. Hier zeigt sich auf Basis bisheriger Forschung eine Vielzahl von Kriterien und subjektiven Parametern, die verschiedene Facetten eines Raumeindrucks formen und in unterschiedlichen Verhältnissen zu tatsächlichen, physikalischen Eigenschaften des Raumes stehen. Und genau das soll raumakustische Hörexpertise im Kontext dieser Untersuchung bedeuten: die Fähigkeit klangliche Eigenschaften eines Raumes zu erfassen und daraus Rückschlüsse uber die akustische, also letztendlich die materielle und geometrische Beschaffenheit dieses Raumes zu

ziehen. Diesem Modell zufolge kann ein Hörer mit hoher raumakustischer Hörexpertise die klangliche Aspekte eines Raumes oder die Differenzen zweier Räume hören und deren physikalische Ursache verstehen und benennen.

2.3. Hörversuche

Hörversuche sind nicht nur für die Untersuchung der im vorigen Abschnitt definierten Hörexpertise unerlässlich, sondern auch der wesentliche Bereich, für den diese Expertise von Interesse ist. Daher sollen hier grundlegende Paradigmen von Hörversuchen erläutert werden und die Basis für die methodische Entwicklung des hier präsentierten Hörversuchs in Abschnitt 3.1 bilden.

2.3.1. Zielsetzungen von Hörversuchen

Als psychologisches Messverfahren bieten Hörversuche die Möglichkeit, die menschliche Wahrnehmung direkt und zu erfassen (vgl. Weinzierl und Maempel, 2012, S. 316). So kann sich die Zielsetzung eines Hörversuchs an verschiedenen Erkenntnissen über unterschiedliche Wahrnehmungsaspekte orientieren. Nach Scriven (2005) kann die Wahrnehmung eines Stimulus' (bezogen auf den Geschmack von Lebensmitteln) in eine primäre und eine sekundäre Reaktion aufgeteilt werden (S. 528). In der primären Reaktion wird der Stimulus registriert und dessen Ausprägung gemessen und anschließend in der sekundären Reaktion subjektiv bewertet. Überträgt man dieses Modell in den auditiven Kontext, ist der Stimulus nicht notwendigerweise ein gesamter Klang, sondern eine einzelner Klangeindruck wie etwa Lautheit oder Halligkeit, dessen Ausprägung in der primären Reaktion bestimmt wird. Und tatächlich wird die Unterscheidung zwischen der bloßen Wahrnehmung eines klanglichen Merkmals und dessen Bewertung offenkundig auch bei Untersuchungen der raumakustischen Hörwahrnehmung vorgenommen, wenn etwa Weinzierl et al. (2018) mit dem RAQI ein Messinstrument zur (neutralen) auditorischen Wahrnehmung von Räumen vorgeschlagen (S. 1245), während Lokki et al. (2012) untersuchen, welche auditorischen Parameter für die Präferenz eines Konzertsaals ausschlaggebend sind (S. 3148).

In beiden Fällen ist es jedoch das Ziel, Parameter zu identifizieren, die im jeweiligen Kontext perzeptiv relevant sind. Bereits in Abschnitt 2.1.5 wurden derartige Studien Untersuchungen gegenübergestellt, die das wahrnehmbare quantitative Spektrum bereits definierter auditorischer oder auch akustischer Parameter untersuchen. In diesem Fall ändert sich auch die Zielsetzung der jeweils durchgeführten Hörversuche. Im ersten Fall, der Definition von Parametern, werden Dimensionen der Wahrnehmung oder auch Präferenz von Raumakustik gesucht, im zweiten Fall werden diese Dimensionen quantifiziert.

Wird der Wahrnehmungsschwelle der Nachhallzeit untersucht, ist der Unterschungsgegenstand rein perzeptiver Natur, d. h. es wird untersucht ob die Differenz zwischen zwei Abklingzeiten gehört werden kann. Dem gegenüber stehen Hörversuche, die nicht die Wahrnehmung eines Unterschiedes, sondern auch die kognitive Auswertung desselben abfragen. Dabei wird untersucht, ob auf Basis klanglicher Eigenschaften z. B. im Falle von Hameed et al. (2004) die Raumgröße oder im Fall von Neidhardt (2016) oder Shinn-Cunningham (2003) die Position von Quelle und Empfänger im Raum erkannt werden soll. Auch hier stellt sich aber nicht nur die Frage, ob Versuchspersonen in der Lage sind, diese Informationen zu gewinnen oder nicht, sondern auch, welche perzeptiven Eigenschaften des Raumklangs für diese abstraktere Interpretation ausschlaggebend sind. So untersuchten Hameed et al. (2004) nicht nur, wie zuverlässig die Raumgröße differenziert werden kann, sondern modulierten die Nachhallzeit und das Verhältnis von Diffusund Direktschall getrennt voneinander, um herauszufinden, ob eine der beiden akustischen Eigenschaften die Einschätzung der Raumgröße stärker beeinflusst.

Auch in dieser Arbeit geht es darum, die perzeptive Sensiblität der Probanden zu testen, und

parallel herauszufinden, welche raumakustischen Klangmerkmale bei der auditorischen Verarbeitung und der sachlichen Interpretation von Raumklang relevant sind.

2.3.2. Versuchsverfahren

Nicht nur beim Untersuchungsgegenstand, sondern auch beim methodischen Vorgehen eines Hörversehens gibt es ein breites Spektrum an Verfahren (vgl. Ellermeier et al., 2008, S. 10). Diese Verfahren basieren auf einigen zentralen Paradigmen, die hier vorgestellt werden, sollen, da sie auch die Grundlage der Hörtests bilden, die in dieser Studie entwickelt und verwendet wurden.

Objektivität, Reliabilität und Validität

Zunächst unterliegen Hörversuche den klassischen Gütekriterien Objektivität, Reliabilität und Validität (vgl. Weinzierl und Maempel, 2012, S. 315).

Objektivität bezeichnet die Unabhängigkeit der Untersuchungsergebnisse von den Ansichten und Wertungen des Untersuchenden. Da jede Erkenntnis über den Forscher als Subjekt festgestellt wird, ist eine absolut subjektunabhängige Wahrheit nur näherungsweise feststellbar (vgl. Schurz, 2011, S. 27). Ein Kriterium hierfür ist Intersubjektivität, die voraussetzt, dass sich jede Person mit dem notwendigen Hintergrundwissen von der Wahrheit einer aus erhobenen Daten formulierten Aussage überzeugen kann (vgl. Schurz, 2011, S. 27). Intersubjektivität kann durch eine standardisierte und vollständig dokumentierte Testverfahren gewährleistet werden (vgl. Weinzierl und Maempel, 2012, S. 315).

Reliabilität bezeichnet die Größe des unvermeidlichen Messfehlers eines Hörversuchs (vgl. Weinzierl und Maempel, 2012, S. 315) und kann nicht nur auf das Verfahren, sondern auch auf Testpersonen bezogen werden. Diese Subjektreliabilität wird im nachfolgenden Abschnitt 2.3.3 näher betrachtet, da sie in engem Bezug zum Konzept des Expertenprobanden steht (vgl. Wickelmaier und Choisel, 2005, S. 1).

Beim dritten der klassischen Testkriterien muss zwischen interner und externer Validität unterschieden werden. Interne Validität bezeichnet das Maß der Kontrolle möglicher Störeinflüsse im Hörversuch (vgl. Ellermeier et al., 2008, S. 18), durch die sichergestellt werden soll, dass die Varianz der Messergebnisse tatsächlich durch die vom Versuchsleiter gewollte Manipulation der Versuchsparameter verursacht wird, und nicht durch anderweitige, unbeabsichtigte Effekte bedingt ist (vgl. Weinzierl und Maempel, 2012, S. 315).

Die externe Validität hingegen beschreibt, inwiefern, die Ergebnisse auf den Alltag der Probanden übertragen werden können (vgl. Ellermeier et al., 2008, S. 18). Ein Hörversuch zur Einschätzung der Raumgröße durch akustische Stimuli allein kann durch strenge Kontrolle des Verfahrens eine hohe interne Validität besitzen, die externe Validität wird jedoch dadurch vermindert, dass der Mensch Räume tatsächlich sowohl auditorisch, als auch visuell wahrnimmt, was das Einschätzungsvermögen der Raumgröße nachweislich verbessert (vgl. Maempel und Jentsch, 2013, S. 2f). Tendenziell besitzen Versuche mit hoher interner Validität durch die strenge Kontrolle der Testbedingungen eine geringe externe Validität (vgl. Ellermeier et al., 2008, S. 18).

Kontrolle über das Entscheidungskriterium (Forced-choice)

Eine Problematik von Hörversuchen und anderen psychologischen Messungen ist, dass die menschliche Wahrnehmung anhand von Urteilen gemessen muss, die eine Versuchsperson im Hörtest fällt, z.B. dass ein Stimulus oder Unterschied gehört wird oder nicht (vgl. Ellermeier et al., 2008, S. 22). Ein Beispiel für derartige Schwellwertmessungen ist das Konstanzverfahren, bei dem einer Versuchsperson verschiedene Reizintensitäten wiederholt präsentiert werden, die teils

über und teils unter dem faktischen Schwellwert liegen (vgl. Ellermeier et al., 2008, S. 27f). Reiz meint in diesem Fall sowohl einen Klang, als auch den Unterschied zwischen zwei Klängen, der gehört wird, oder nicht. Die Rate an erkannten Reizen liefert dann eine in Abbildung 2.1 dargestellte psychometrische Funktion (in diesem Fall für die Messung der Absoluthörschwelle eines Stimulus'). Die Warhnehmungsschwelle markiert der Bereich, in dem in 50 % der Fälle der Reiz wahrgenommen wird. Bei kleineren Reizstärken nähert sich die Funktion asymptotisch einer Erkennungsrate von 0 %, bei größeren reizstärken der Rate von 100 % an. Die Entscheidung für

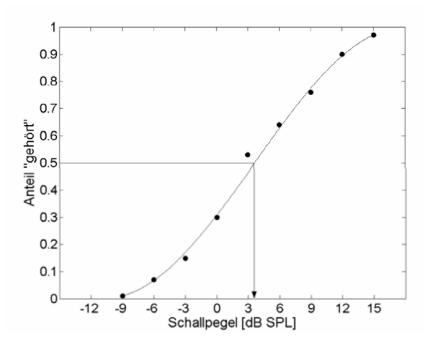


Abbildung 2.1.: Psychometrische Funktion zur Bestimmung der Absoluthörschwelle eines Stimulus. Der mit dem Pfeil markierten Punkt einer Erkennungsrate von $50\,\%$ markiert die Wahrnehmungsschwelle (Ellermeier et al., 2008, S. 27)

ein bestimmtes Urteil kann allerdings unter Umständen durch andere Faktoren als die Wahrnehmung beeinflusst werden, z. B. durch eine Antwortneigung, einen Stimulus erst dann als gehört anzugeben, wenn die Versuchsperson sich dabei absolut sicher ist und nicht bereits bei einem eher unsicheren Eindruck, ein Geräusch wahrgenommen zu haben (vgl. Ellermeier et al., 2008, S. 24). Eine Möglichkeit der Kontrolle über das Entscheidungskriterium bieten Forced-choice-Verfahren, bei denen nicht nur ein Reiz präsentiert und der Versuchsperson die Entscheidung überlassen wird, ob sie diesen gehört hat oder nicht, sondern zusätzlich die Situation abgefragt wird, in der der Reiz nicht vorhanden ist. Aufgabe der Versuchsperson ist es, korrekt zuzuordnen in welchem Fall ein Reiz gegeben ist und in welchem nicht (vgl. Ellermeier et al., 2008, S. 28). Gemessen wird alo effektiv, ob die Versuchsperson eine Situation (Reiz gegeben oder nicht) konsistent bewertet, oder rät, da sie nicht erkennen kann, ob ein Reiz vorhanden ist (vgl. Ellermeier et al., 2008, S. 28).

Ratewahrscheinlichkeit

Da bei den Forced-choice-Verfahren identifiziert werden soll, wann eine Testperson rät, anstatt auf Basis ihrer Wahrnehmung zu antworten, ist es notwendig, sich mit der Wahrscheinlichkeit auseinanderzusetzen, durch Raten die Testaufgabe korrekt zu lösen. Die Wahrscheinlichkeit $P(X \leq x)$, dass die Zufallsvariable X (in diesem Fall die Anzahl richtig geratener Antworten) bei n Testwiederholungen und einer Ratewahrscheinlichkeit p maximal den Wert x erreicht,

berechnet sich nach Papula (2009, S. 413) als:

$$P(X \le x) = \sum_{k \le x} \binom{n}{k} p^k \cdot (1 - p)^{n - k}$$
 (2.5)

Mit Hilfe dieser Gleichung kann ermittelt werden, wie viele Wiederholungen eines Forced-choice-Tests notwendig sind, damit die Wahrscheinlichkeit, durch Raten x richtige Antworten zu geben, unterhalb des üblichen Signifikanzniveaus liegt, das üblicherweise von 0.05 bzw. 5% beträgt (vgl. Ellermeier et al., 2008, S. 15).

Randomisierung

Randomisierung ist ein wichtiges Element bei der Planung von Hörversuchen (vgl. Ellermeier et al., 2008, S. 10). Wenn die Differenz zwischen zwei zu unterscheidenden Stimuli kontinuierlich verringert wird, besteht die Gefahr der Antizipation, in der eine Erwartungshaltung die Entscheidung einer Versuchsperson unabhängig vom tatsächlichen sensorischen Eindruck beeinflusst (vgl. Ellermeier et al., 2008, S. 26f) Ebenso kann durch Trainingseffekte eine Serie von Testwiederholungen mit mehreren Schwierigkeitsgraden zu unterschiedlichen Ergebnissen führen, je nach dem ob mit dem leichtesten oder schwersten Schwierigkeitsgrad begonnen und dieser kontinuierlich erhöht bzw. verringert wird (vgl. Shinn-Cunningham, 2003, S. 3). Dieser Effekt kann durch Randomisierung der Schwierigkeitsgrade begrenzt werden. Auch beim den oben diskutierten Forced-choice-Verfahren sollte die Reihenfolge von Beobachtungsintervallen, in denen der Reiz präsent ist oder nicht, zufällig generiert werden (vgl. Ellermeier et al., 2008, S. 30).

Adaptive Verfahren

Adaptive Tests bieten eine Möglichkeit, beispielsweise Schwellenmessverfahren deutlich zu verkürzen, ohne dass die Ergebnisse an Reliabilität einbüßen (Harrison und Müllensiefen, 2018, vgl. [S. 3). Bei adaptiven Testverfahren werden die Testparameter in Abhängigkeit der vorherigen Antworten der Versuchsperson variiert: falsche Antworten führen zu einer Vergrößerung der Differenzen zwischen den zu vergleichenden Beobachtungsintervallen, (meist zwei aufeinanderfolgende) richtige Antworten hingegen bedingen eine Verringerung der Differenz (vgl. Ellermeier et al., 2008, S. 31). In der Regel präsentiert man zu Beginn des Verfahrens deutlich wahrnehmbare Unterschiede, die dann sukzessiv verkleinert werden, bis sie den Bereich der Wahrnehmungsschwelle erreicht haben, in der der Verlauf der Parametervariation zwischen Vergrößerung und Verkleinerung schwankt (vgl. Ellermeier et al., 2008, S. 32). Somit erreichen die Testparameter schnell die Dimensionen, die in Abhängigkeit der individuellen Fähigkeiten der Probanden am aussagekräftigsten sind (vgl Harrison und Müllensiefen, 2018, S. 3). Dazu muss natürlich bereits bekannt sein, in welchen Größenordnungen die Wahrnehmungsschwellen prinzipiell liegen.

2.3.3. Eignung einer Person für die Teilnahme an Hörversuchen

Dass bei der Durchführung von Hörversuchen die Entscheidung zwischen Experten und Laien bisher eher inkonsistent ist, wurde in Abschnitt 1.1 bereits anhand einiger Beispiele gezeigt. Dennoch gibt es Konzepte und Überlegungen zur Eignung einer Person für die Teilnahme an Hörversuchen oder anderen, thematisch ähnlich gelagerten Studien. Diese beleuchten auch Aspekte bei der Durchführung von Hörversuchen mit Probanden unterschiedlicher Fachbildung, die auch für die Versuchsverfahren dieser Untersuchung relevant sind.

Kategorien von Prüfpersonen

Zacharov und Lorho (2006) empfehlen die Kategorisierung von Testsubjekten in der Audiodomäne nach Vorbild der DIN EN ISO 8586, in sechs Typen von Prüfpersonen: Laien (naïve assessor), eingeführte Prüfpersonen (initiated assessor), ausgewählte Prüfpersonen (selected assessor), Sachverständige (expert), Sensoriker (expert assessor) und spezialisierte Sensoriker (specialised expert assessor) (vgl. S. 4, für deutsche Übersetzungen siehe DIN EN ISO 5492 (2009, S. 6-9)). Definitionen dieser Kategorien sind in Tabelle 2.1 gegeben. Anhand der englischen Namen der Kategorien wird deutlich, dass dabei ein Unterschied besteht zwischen einem expert, also einer fachlich gebildeten Person, und einem expert assessor, einer Person mit "nachgewiesener sensorischer Empfindlichkeit und umfassender Schulung und Erfahrung hinsichtlich der sensorischen Prüfung "(DIN EN ISO 5492, 2009, S. 9). Diese Differenzierung, auf die auch Zacharov und Lorho (2006) explizit hinweisen (vgl. S. 4), bestätigt eine der Annahmen dieser Untersuchung, dass nämlich fachliche Qualifikation auf dem jeweiligen Themengebiet und die sensorische Leistungsfähigkeit zwei verschiedene Eigenschaften sind, weshalb zumindest anfechtbar ist, Raumakustikexperten automatisch als qualifizierte Hörversuchsteilnehmer einzustufen.

Kategorie	Definition
Laie	Eine Person ohne bestimmte
	Qualifikation
Eingeführte Prüfperson	Eine Person, die bereits an sensorischen Tests teilge-
	nommen hat
Ausgewählte Prüfperson	Eine Person, die wegen persönlicher Eignung für die
	Teilnahme an sensorischen Tests ausgewählt wurde
Sachverständiger	Eine Person, die durch Wissen oder Erfahrung eine
	kompetente Meinung zu einem Thema geben kann.
Sensoriker	Eine ausgewählte Prüfperson mit hoher Sensibilität
	und Erfahrung in der jeweiligen Wahrnehmungs-
	domäne, deren sensorische Bewertungen verschiedener
	Produkte konsistente und reproduzierbare Ergebnisse
	liefert.
Spezialisierter Sensoriker	Sensoriker mit zusätzlicher Expertise und Erfahrung
	zu einem Produkt, Prozess oder Marketing, der Pro-
	dukte sensorisch analysieren und Effekte durch Va-
	riation bestimmter Produkteigenschaften einschätzen
	kann.

Tabelle 2.1.: Kategorisierung von Prüfpersonen in Anlehnung an DIN EN ISO 8586. Gekürzte Fassung von Zacharov und Lorho (2006, S. 4)

Eine Person mit hoher raumakustischer Hörexpertise nach der Definition in Abschnitt 2.2 fiele in die Kategorie eines spezialisierten Sensorikers, da sie nicht nur eine ausgeprägte Sensibilität für die Wahrnehmung raumakustischer Klangeffekte hat, sondern auch die Wirkung der Eigenschaften des Raumes (gleichzusetzen mit dem 'Produkt'in Tabelle 2.1) auf diese Klangeffekte bewerten kann. Da in dieser Untersuchung sowohl perzeptive und kognitive Fähigkeiten als auch fachliches Wissen und Erfahrung abgefragt werden sollen, könnte zudem Eignung der Personen als Sensoriker bzw. Sachverständige bestimmt werden, die Grundlage der Eignung als spezialisierter Sensoriker bilden. Jeweils 'nur'als Sachverständiger oder Sensoriker tauglich zu sein, schränkt zwar die raumakustische Hörexpertise im engsten Sinne ein, könnte eine Person jedoch

für bestimmte Studien qualifizieren in denen eben diese beiden Typen von Prüfpersonen gefordert sind.

Trainingseffekte und Vorerfahrung

Der Zusammenhang zwischen Erfahrung und Performanz in sensorischen Hörversuchen nicht ganz unerforscht. Die von Scriven (2005) ebenfalls im Kontext der Lebensmittelindustrie diskutierte Eigenschaft von expert assessors, bei sensorischen Experimenten einheitliche und konsistente Ergebnisse zu liefern (vgl. S. 528f), wurde von Bech (1992) im Bereich der Hörversuche betrachtet. Konkret wurde untersucht, inwiefern die Fähigkeit, bei Hörtests zu Audiowiedergabesystemen reproduzierbare Ergebnisse zu produzieren, mit der absoluten Hörschwelle und der bisherigen Erfahrung mit Hörversuchen zusammenhängt (vgl. Bech, 1992, S. 590). Dabei wurde gezeigt, dass keine Korrelation zwischen der Hörschwelle und der Konsistenz des Antwortverhaltens besteht, wenn die Hörschwellen der Teilnehmer zwischen 250 Hz und 8 kHz eine Spanne von 15 dB nicht überschreiten. (vgl. Bech, 1992, S. 604). Allerdings konnten deutliche Trainingeffekte beobachtet werden. Demnach machen Probanden sich bei der erstmaligen Durchführung eines Tests zunächst mit dem Testverfahren und den Stimuli vertraut, wodurch sie sich noch nicht vollständig auf die abgefragten Testparameter konzentrieren. In nachfolgenden Wiederholungen hingegen sind experimentelle Umgebung und Verfahren bekannt, sodass die Probanden lernen, sich tatsächlich auf z.B. die Klangfarbe der Lautsprecher als singulären Faktor zu fokussieren und diese zu bewerten (vgl. Bech, 1992, S. 599).

Neben diesem Effekt des Trainings innerhalb eines Experiments sei noch angemerkt, dass erstens Vorerfahrung mit einem Antwortverhalten korreliert, das konsistenter aber nicht notwendiger qualitativ besser ist (vgl. Bech, 1992, S.). Die Antworten erfahrener Testsubjekte sind nicht zwingend richtiger, sondern gleichbleibender, als die von unerfahrenen. Bech (1992) versteht dabei Vorerfahrung nicht nur als Erfahrung mit Hörversuchen sondern auch generelle Vertrautheit mit dem Thema Klangfarbe durch kritisches Musikhören, Musizieren oder Konzertbesuchen (vgl. S. 593).

Der Trainingsprozess, sich zu Beginn eines Experiments zunächst verstärkt mit den Stimuli und Testbedingungen eines Hörversuchs zu beschätfigen, anstatt sich vollends auf die gestellte Aufgabe zu konzentrieren, ist von der untersuchten Wahrnehmungskomponente unabhängig und daher auch bei Hörversuchen über raumakustische Klangeigeschaften zu erwarten. Shinn-Cunningham (2003) beobachtete auch bei Experimenten zur Bestimmung der eigenen Position im Raum, bei der gezielt raumakustische Auswirkungen auf Stimuli untersucht wurden, Trainingseffekte, bei denen eine Versuchsperson sich nicht nur an die experimentelle Umgebung, sondern auch den Raum, innerhalb dessen verschiedene Positionen verglichen werden, gewöhnt (S. 3). Wenn auch die Fähigkeit der Lokalisation im Raum generell schwach ausgeprägt ist, verbessert sich die Einschätzung der Elevation einer Schallquelle nur in einer räumlichen Umgebung. In einem Vergleichsexperiment unter nachhallfreien Bedinugngen, blieb dieser Effekt aus (vgl. Shinn-Cunningham, 2003, S. 3).

Es gibt also einerseits den Befund, dass sich innerhalb eines Hörversuchs Probanden durch Gewöhnung an Versuchsumgebung und -Verfahren zunehmend auf die abgefragten Klangparameter allein konzentrieren, andererseits generellere Vorerfahrung mit konzentriertem Hören oder den abgefragten Hörsituationen ein konsistenteres Antwortverhalten hervorruft und drittens das auditorische System durch eine Art "textitroom learning"(Shinn-Cunningham, 2003, S. 4) die akustischen Einflüsse eines Raum zunehmend besser einschätzen kann. Vorerfahrung ist ohnehin ein Faktor, der in der Ausgangsfrage dieser Arbeit eine zentrale Rolle spielt. Der Trainingsaspekt wird in dieser explorativen Studie zunächst vernachlässigt, ist jedoch zweifelsohne eine Komponente, die auf Dauer nicht außen vor gelassen werden sollte.

Linguistische Aspekte

Ein weiterer Unterschied zwischen Laien und fachlichen Experten ergibt sich aus der Fachterminologie. Die Tatsache, dass fachkundige Personen ein intutives Verständnis der Konzepte besitzen, die hinter raumakustischen Klangparametern stehen, könnte die auditorische Bewertung derselben erleichtern (vgl. Weinzierl und Vorländer, 2015, S. 42). Um die perzeptiven Fähigkeiten von fachlichen Laien und Experten unterscheiden und vergleichen zu können, muss in den hier erstellten Hörtests darauf geachtet werden, solche sprachlichen Barrieren so gut es geht zu vermeiden. Begriffe wie 'hallig'oder 'dumpf' werden als inituitiv genug eingeschätzt, dass auch Laien sie verstehen könne. Bei abstrakteren Konzepten wie dem Klarheitsmaß können hingegen Verständnisschweirigkeiten als Störgröße die Versuchsergebnisse beeinträchtigen (vgl. Martelotta, 2010, S. 659).

2.4. Probabilistische Testtheorie

Die Probabilistische Testtheorie, oder Item-Response-Theorie, ist eine Alternative zur klassischen Testtheorie, auf der z.B. das in Abschnitt 2.3.2 gezeigte Konstanzverfahren basiert. Probabilistische Testtheorie eignet sich besonders für die Messung von persönlichen Kompetenzen (vgl. Hartig und Frey, 2013, S. 48) und wird deshalb auch für die hier durchgeführten Hörtests als geeignete Theorie angesehen. Daher wird diese Testtheorie in diesem Abschnitt kurz vorgestellt und einige Vorteile gegenüber der klassischen Testtheorie dargelegt. Bei der Entwicklung von Verfahren nach dieser Theorie muss zunächst die Metrik der Item-Schwierigkeitsgrade kalibriert werden (vgl. Baker, 2001, S. 130). Die Methodik dieser Kalibrierung wird im Folgenden ebenfalls näher erläutert..

2.4.1. Grundlagen der probabilistischen Testtheorie

In Abschnitt 2.3.2 wurde bereits diskutiert, dass bei einem Hörversuch das Urteil, dass eine Person in einer Testsituation fällt, Aufschluss über eine diesem Urteil zugrunde liegende Wahrnehmung geben soll. In der klassischen Testtheorie wird dieses Urteil als direktes, Maß für die Wahrnehmungsfähigkeit eingestuft. Dieses Maß unterliegt einem gewissen Messfehler (vgl. Kaspar, 2009, S. 105), der von verschiedenen zuälligen, situativen Einflüssen herrühren kann (Hambleton und van der Linden, Wim J., 1982, vgl. [S. 374) und durch eine hohe Reliabilität des Verfahrens möglichst gering halten werden soll (vgl. Weinzierl und Maempel, 2012, S. 315). Dazu wird unter anderem die ebenfalls in Abschnitt 2.3.2 behandelte Ratewahrscheinlichkeit durch Testwiederholungen begrenzt, um anhand des Ergebnisses des Gesamttests die Fähigkeit einer Person zu messen (vgl. Kaspar, 2009, S. 105).

In der probabilistischen Testtheorie wird auf Basis der Antwort (response) auf ein einzelnes Test-Item (eine Durchführung mit spezifischer Testparamterkonfiguration) ein Modell gebildet, um diese Fähigkeit zu schätzen (vgl. Hambleton und van der Linden, Wim J., 1982, S. 375). Die Wahrscheinlichkeit, ein Testitem J mit X zu beantworten, wird meist über ein logistisches Modell mit vier Parametern a_J, b_J, c_J, d_J ausgedrückt (Harrison und Müllensiefen, 2018, S. 3):

$$P(X_j = 1 | \theta, a_J, b_J, c_J, d_J) = c_J + (d_J - c_J) \cdot \frac{e^{a_J(\theta - b_J)}}{1 + e^{a_J(\theta - b_J)}}$$
(2.6)

Die Größe θ bezeichnet die Fähigkeit der Versuchsperson, a_J die Itemdiskrimination, b_J die Itemschwierigkeit, c_J die in Gleichung 2.5 als p enthaltene Ratewahrscheinlichkeit und d_J die Wahrscheinlichkeit momentaner Unaufmerksamkeit. Die Itemschwierigkeit beschreibt, welches Maß an Fähigkeit für die korrekte Beantwortung des Items erforderlich ist (vgl. Harrison und Müllensiefen, 2018, S. 3). Der exakte Wert der Schwierigkeit orientiert sich an der Fähigkeitsausprägung, bei der die Wahrscheinlichkeit, korrekt zu antworten, im Falle einer Ratewahrschein-

lichkeit von 0 bei 0,5 bzw. 50% liegt (vgl. Bejar, 1990, S. 239). Im Falle einer Unterscheidung zweier Stimuli als Test entspricht dieser Punkt der Wahrnehmungsschwelle einer Person mit dieser Fähigkeitsausprägung. Durch eine vorhandene Ratewahrscheinlichkeit von beispielsweise 50% wird dieser Punkt entsprechend von 50% auf 75% verschoben. Abbildung 2.2 zeigt die Auswirkungen einer Ratewahrscheinlichkeit auf die spezifische Lösungswahrscheinlichkeit der Itemschwierigkeit).

Im Umkehrschluss kann bei bekannter Itemschwierigkeit die individuelle Fähigkeit quantitativ erfasst werden (vgl. Baker, 2001, S. 130). In den Modellen der probabilistischen Testtheorie werden also die Fähigkeit der Versuchsperson und die Schwierigkeit des Test-Items auf der gleichen Skala erfasst (vgl. Hartig und Frey, 2013, S. 48). Diese Skala ist so angelegt, dass die Verteilung der vorkommenden Fähigkeitsgrade einen Mittelwert von 0 und eine Standardabweichung von 1 hat (vgl. Harrison und Müllensiefen, 2018, S. 3).

Die Itemdiskrimination beschreibt die Trennschärfe des Items. Bei einem sehr diskriminativen Item ist diese Lösungswahrscheinlichkeit bei dem einen Fähigkeitswert sehr gering und bei einem nur geringfügig größeren Wert wesentlich höher. So können auch benachbarte Fähigkeitsgrade klar differenziert werden (vgl. Kaspar, 2009, S. 108).

Der Unaufmerksamkeitsparameter beschreibt die Resistenz der Beantwortung des Test-Items gegen Ablenkungen der Versuchspersonen, durch die diese mit einer gewissen Wahrscheinlichkeit falsche Antworten geben, auch wenn sie eigentlich die Fähigkeiten besitzen, um sicher korrekt zu antworten (vgl. Harrison und Müllensiefen, 2018, S. 3).

Wie auch beim Konstanzverfahren wird eine psychometrische Funktion angenommen, nach der

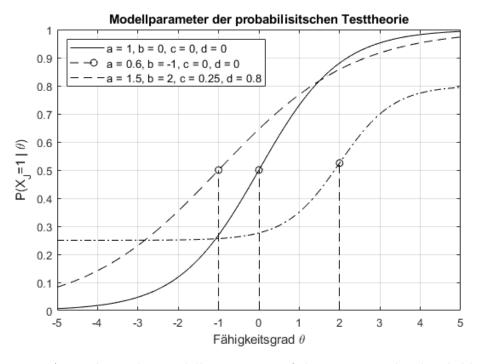


Abbildung 2.2.: Auswirkung der Modellparameter auf die Lösungswahrscheinlichkeit von drei Test-Items in Anhängigkeit der Fähigkeit der Probanden. Dabei repräsentiert a die Itemdiskrimination, b die Itemschwierigkeit, c die Ratewahrscheinlichkeit und d den Unaufmerksamkeitsparameter. Die Itemschwierigkeit ist durch Kreise markiert und wird auf derselben Skala wie der Fähigkeitsgrad abgebildet.

die Lösungwahrscheinlichkeit in Abhängigkeit der Fähigkeitsausprägung der Versuchsperson eine Sigmoidfunktion abbildet, deren Wendepunkt durch die Itemschwierigkeit charakterisiert wird. Die Steigung im Wendepunkt wird durch die Itemdiskrimination beschrieben, die Ratewahr-

scheinlichkeit beschreibt, welchem Wert die Funktion sich nach unten asymptotisch annähert (vgl. Kaspar, 2009, S. 107ff). Der Unaufmerksamkeitsparameter beeinflusst nach Gleichung 2.6 den oberen Wert der Annäherung. Abbildung 2.2 stellt die Auswirkungen der Modellparameter auf die psychometrische Funktion graphisch dar.

2.4.2. Vorteile der probabilisitschen Testtheorie

Die probabilistische Testtheorie hat gegenüber der klassischen Testtheorie einige Vorteile, sowohl im testtheoretischen, als auch im Sinne einer effizienten Versuchsgestaltung. Zum einen lassen sich Itemschwierigkeit und Fähigkeitsgrad der Testperson im Modell separat schätzen. Damit lässt sich sowohl die Fähigkeit einer Person unabhängig vom Testverfahren beschreiben (vgl. Hambleton und van der Linden, Wim J., 1982, S. 375), als auch die Itemschwierigkeit unabhängig von den verfügbaren Probanden, sodass Item-Batterien mit universell einsetzbaren Schwierigkeitsgraden erstellt werden können. Dies bietet die Grundlage für adaptive Testverfahren (vgl. Hambleton und van der Linden, Wim J., 1982, S. 376). Zudem muss nicht jede Testperson alle Items beantworten, um die Ergebnisse verschiedener Probanden vergleichen zu können (vgl. Harrison und Müllensiefen, 2018, S. 3). Es sei allerdings angemerkt, dass die Metrik der Fähigkeitsausprägung, anhand arbiträrer Regeln durch die initial verwendete Item-Batterie und Stichprobe definiert sind (vgl. Baker, 2001, S. 132).

Auch sind wiederholte Beantwortungen der Test-Items nicht notwendig, da die Fähigkeit der Person nicht, wie in der klassischen Testtheorie, von dem Ergebnis des Gesamttests ausgedrückt wird, in dem der Einfluss der Ratewahrscheinlichkeit deshalb durch Testwiederholungen kompensiert werden muss. Ratewahrscheinlichkeit und Fähigkeit wirken als unabhängige Modellparameter auf das stochastische Testergebnis eines Items ein (vgl. Hambleton und van der Linden, Wim J., 1982, S. 376). Damit kann nicht nur die Durchführungsdauer erheblich verkürzt werden, sondern auch das Problem umgangen werden, dass exakte Wiederholungen desselben Tests sich real nicht umsetzen lassen, da sich während der mehrfachen Durchführung situative Einflüsse auf die Versuchsperson ändern, indem sie z. B. auf neue Details in den Stimuli achten, unaufmerksam werden, die Motivation nachlässt, Müdigkeit einsetzt etc. (vgl. Hambleton und van der Linden, Wim J., 1982, S. 374).

Während bei der klassischen Testtheorie das Konzept der idealen Wiederholung nur begrenzt haltbar ist, ist bei der probabilistischen Testtheorie die Annahme stochastischer Unabhängigkeit nicht ganz unproblematisch, die besagt, dass die Beantwortung eines Testitems absolut unabhängig von der Beantwortung anderer Items des gleichen Tests ist. Gerade bei mehreren Durchführungen des selben Tests mit variablen Itemschwierigkeiten sind lokale Abhängigkeiten zwischen den einzelnen Tests möglich (vgl. Hartig und Frey, 2013, S. 49). Aus pragmatischer Sicht stellt gerade in der Phase der Kalibrierung der Itemschwierigkeiten, die im folgenden Abschnitt diskutiert wird, der Bedarf an sehr großen Stichproben ebenfalls ein Problem dar, da bei eindimensionalen Modellen etwa 100 Antworten pro Item, bei komplexeren Modellen auch sechsstellige Teilnehmerzahlen erforderlich sein können (vgl. Hartig und Frey, 2013, S. 50).

2.4.3. Kalibrierung der Test-Items

Die Schätzung der Schwierigkeiten und weiteren Parameter der Test-Items stellt einen initialen Schritt bei der Entwicklung eines Testverfahrens nach der probabilistischen Testtheorie dar (vgl. Harrison und Müllensiefen, 2018, S. 3). Daher müssen auch die in dieser Arbeit entwickelten Hörtests als Kalibrierungsexperimente entwickelt werden, da es nicht möglich ist, die Itemparameter a priori zu bestimmen (vgl. Baker, 2001, S. 130), bzw. hypothetische Prädiktoren für die Itemschwierigkeit gefunden und während der Kalibrierung auf ihre Eignung überprüft werden müssen (vgl. Harrison und Müllensiefen, 2018, S. 5). Dazu empfiehlt es sich, die strukturellen

Parameter der Test-Items in radicals und incidentals aufzuteilen, d. h. in Faktoren, bei deren Variation Auswirkungen auf die Itemschwierigkeit erwartet wird und solche, bei denen das nicht der Fall ist (vgl. Harrison und Müllensiefen, 2018, S.5). Die genaue Festlegung der Prädiktoren für die Schwierigkeitsgrade der hier entwickelten Hörtests folgt in Abschnitt 3.1.2.

In der probabilistischen Testtheorie ist der Einsatz von Rasch-Modellen beliebt (vgl. Baker, 2001, S. 136), da nur ein itemspezifischer Parameter variiert wird und dieses Verfahren für eine geringe Anzahl an Items und Probanden geeignet ist (vgl. Baker, 2001, S. 132). Üblicherweise wird nämlich eine konstante Itemdiskrimination für alle Items angenommen (vgl. Harrison und Müllensiefen, 2018, S.6). Bei der Kalibrierung werden alle Items einer entwickelten Testbatterie allen Probanden einer Stichprobe vorgelegt (vgl. Baker, 2001, S. 130). Für die anschließende Schätzung der Modellparameter ist das Birnbaum-Paradigma verbreitet (oder auch joint maximum likelihood estimation procedure), bei dem in einem zweistufigen Verfahren erst die Itemschwierigkeit und dann die individuelle Fähigkeit bestimmt wird (vgl. Baker und Kim, 2017, S. 106). Um die Itemschwierigkeit zu bestimmen werden dabei zu nächst Werte für die Fähigkeitsgrade der Probanden geschätzt und als tatsächliche Werte angenommen, um für jedes Item die Schwierigkeit zu schätzen (vgl. Baker, 2001, S. 131). Initial wird die Fähigkeit einfach über die Anzahl an richtiger Antworten bestimmt (vgl. Baker, 2001, S. 133). Dann werden wiederum die geschätzten Itemschwierigekeiten als die tatsächlichen angenommen, um daraus die Fähigkeitsgrade neu zu schätzen und der Prozess so lange wiederholt, bis Verbesserung der Anpassung der modellierten Parameter an den Datensatz ein jeweils definiertes Konvergenzkriterium erfüllt (vgl. Baker und Kim, 2017, S. 106).

Alternativen bietet der Einsatz generalisierter linearer gemischter Modelle (GLMMs). Bei der Umsetzung von Rasch-Modellen werden die Itemparameter als feste und die persönlichen Fähigkeiten als zufällige Effekte behandelt (vgl. de Boeck et al., 2011, S.8), da diese die Einflüsse der verschiedenen Einheiten abbilden, zu denen sich die Einzelbeobachtungen gruppieren. In diesem Fall sind die Einheiten die Personen, zu denen jeweils ein Antwortsatz aller Items gehört (vgl. Rijmen et al., 2003, S. 185). Item-Response-Modelle können über GLMMs beschrieben werden, wenn Ratewahrscheinlichkeit und Unaufmerksamkeitsparameter ungleich 0 und die Itemdiskrimination für alle Items gleich sind (vgl. Harrison und Müllensiefen, 2018, S.6).

2.4.4. Explorative Faktoranalyse und probabilistische Testtheorie

Die explorative Faktoranalyse ist der Berechnung von Modellen nach der probabilistischen Testtheorie nicht unähnlich. Bei der Faktoranalyse wird die Varianz einer metrischen Variablen, in
diesem Fall eines Items, in eine durch die latenten Faktoren bedingte Varianz und eine Einzelrestvarianz zerlegt (vgl. Backhaus et al., 2016, S. 392). Dementsprechend repräsentieren die
latenten Faktoren die Fähigkeitsausprägung und die Einzelrestvarianz den Effekt der Itemparameter. Die genaue Funktionsweise der Faktoranalyse wird hier nicht vertieft, ein Unterschied
liegt vor allem darin, dass anders als bei der logistischen Regression der probabilistischen Testtheorie metrische statt dichotomer responses betrachtet werden. Bei der Faktoranalyse werden
hoch korrelierte Items mit unterschiedlicher Gewichtung zu einem Faktor zusammengefasst (vgl.
Backhaus et al., 2016, S. 366). Dazu ist es erforderlich, dass entsprechende Korrelationen gegeben sind (vgl. Backhaus et al., 2016, S. 375). Das Kaiser-Meyer-Olkin-Kriterium zur Beruteilung
der Korrelationsverteilungen gilt als das derzeit beste, um die Eignung eines Datensatzes von
Itemantworten für eine Faktoranalyse zu bestimmen (vgl. Backhaus et al., 2016, S. 379).

Außerdem stellt sich die Frage, wie viele Faktoren sinnvollerweise aus den Antwortsätzen mehrerer Items berechnet werden sollten (vgl. Backhaus et al., 2016, S. 370). Neben Überlegungen der Plausibilität gibt es einige deskriptive Methoden, um die geeignete Anzahl an zu extrahierenden Faktoren zu berechnen. Nach dem Kaiserkriterium sollten alle Faktoren mit einem Eigenwert

über eins zugelassen werden (vgl. Fabrigar et al., 1999, S. 278). Beim Scree-Test werden zur Bestimmung der Faktorenanzahl die Eigenwerte für alle möglichen Faktoren auf substantielle Verminderungen des Eigenwerts von einem Faktor zum nächsten untersucht (vgl. Backhaus et al., 2016, S. 397). Darüberhinaus kann die Verbesserung der Modellanpassung an die Daten mit zunehmender Faktoranzahl iterativ getestet werden, Fabrigar et al. (1999) empfehlen hier den Effektivwert des Approximationsfehlers (Root Mean Error of Approximation, RMSEA) als Gütekriterium, der weniger von der Stichprobengröße abhängt als der Chi-Quadrat-Test, der bestimmt, ob das Faktorenmodell als korrektes Modell mit Blick auf die gegebenen Variblenausprägung zugelassen werden kann (vgl. S. 280). Für den RMSEA bedeuten Werte unterhalb von 0,05 eine gute und Werte bis 0,08 eine akzeptable Modellanpassung, beim Chi-Quadrat-Test ist ein nicht signifikantes Ergebnis das Gütekriterium für eine gute Anpassung (vgl. Fabrigar et al., 1999, S. 280). Ebenfalls weniger abhängig von der Stichporbengröße ist der Tucker-Lewis-Index, der zudem auch auf eine Überanpassung des durch die Faktoren spezifierten Modells an den konkreten Datensatz hinweisen kann (vgl. Hu und Bentler, 1998, S. 429). Für diesen Index werden Werte oberhalb von 0,9 als Indiz guter Modellanpassung angesehen (vgl. Marsh et al., 1988, S. 6).

Bei der Analyse selbst stellt sich die Frage nach Schätzungmethode für die Faktorladungen. Fabrigar et al. (1999) empfehlen die Maximum-Likelihood-Methode (vgl. S. 283), die auch hier in allen Faktoranalysen verwendet wurde. Darüberhinaus sollte festgelegt werden, ob bei der Extraktion der Faktoren Korrelationen derselben erlaubt sein sollen oder nicht. Unkorrelierte Faktoren sind zwar einfacher zu interpretieren, da sie voneinander unabhängige latente Merkmale bedeuten, nach Fabrigar et al. (1999) gibt es jedoch im Bereich kognitiver Kompetenzen Grund zur Annahme, dass korrelierte Faktoren die tatsächlichen Beziehungen latenter Merkmale besser darstellen (vgl. S. 282.). Zudem lassen sich unkorrelierte Faktoren auch dann extrahieren, wenn Korrelation prinzipiell erlaubt ist. Für die dafür erforderliche Rotation der Faktorachsen stehen verschiedene Methoden zur Verfügung. Hier wird die "Promax"-Methode benutzt, die nach Rennie (1997) die oblique Rotationstechnik ist, die am ehesten reproduzierbare Ergebnisse produziert (vgl. S. 15).

Wird die Interpretation der Faktoren durch Korrelationen erschwert, kann zudem eine weitere Faktoranalyse derselben druchgeführt werden, um zugrundeliegende Faktoren höherer Ordnung zu identifizieren, die zugleich einen tieferen Einblick in die untersuchten latenten Strukturen vermitteln können (vgl. Wolff und Preising, 2005, S. 49). Hierbei ist die Schmid-Leimann-Transformation (Schmid und Leiman, 1957) ein hilfreiches Werkzeug, da diese sowohl Faktoren höherer als auch niedriger Ordnung durch ihren Einfluss auf die beobachteten Items darstellt (vgl. Wolff und Preising, 2005, S. 49f). Das bedeutet, dass die Faktoren, wenn auch hierarchisch organisiert, unabhängig voneinander ausgedrückt werden, indem die Faktorladungen der Faktoren aus der ersten Faktoranalyse in einen Anteil dieses jeweiligen Faktors selbst und einen des Faktors höherer Ordnung aufgeteilt werden. Ein Beispiel für die Anwendung der Schmid-Leimann-Transformation auf Faktoren unterschiedlicher Ordnung ist die Erstellung des Gold-MSI von Müllensiefen et al. (2014), der in Abschnitt 3.3.2 beschrieben wird (vgl. S. 5).

2.5. Virtuelle Akustik

Bevor im nächsten Kapitel die Entwicklung und Umsetzung des Hörversuchs erörtert wird, werden hier zwei Teilgebiete der virtuellen Akustik behandelt, mit Hilfe derer die Stimuli erstellt wurden: Raumsimulation und Binauralsynthese. Dieser Realitätsgrad von Raumsimulationen unterliegt gewissen Einschränkungen (vgl. Brinkmann et al., 2019, S. 2746). Trotzdem werden Simualtion der Arbeit mit realen Räumen vorgezogen, da für die Hörversuche eine Vielzahl verschiedener Räume und Nachhallcharakteristika benötigt wird, was in realen Räumen so nicht

umsetzbar wäre. Für die Durchführung Kopfhörer und Binauralsynthese zu verwenden, bietet den Vorteil, dass das Wiedergabesystem wesentlich kompakter und weitestgehend frei von Reflexionen durch den umgebenden Raum ist (vgl. Fleischmann et al., 2012, S.2). Für diesen Hörversuch, der an mehreren Orten mit Räumen unterschiedlicher Größe und Nachhallcharakteristik durchgeführt wurde, bedeutet dies, sowohl ähnlichere Testbedingungen, als auch weniger materiellen Aufwand gegenüber einer Stimuluswiedergabe über dreidimensionale Lautsprechersysteme.

2.5.1. Raumsimulation

Raumsimulationen basieren überwiegend auf geometrischer Akustik (vgl. Schröder und Vorländer, 2011, S. 1541), und modellieren das Schallfeld durch Spiegelschallquellen- und Ray-tracing-Algorithmen (vgl. Brinkmann et al., 2019, S. 2748), d. h. zusätzliche Schallquellen für jede Wandreflexion bzw. jedes Schallpartikel, deren Ausbreitungsweg im Raum verfolgt wird (vgl. Weinzierl, 2008, S 244f.). Dabei werden frequenzabhängige Absorption und diffuse Streuung, Richtcharakteristiken von Sender und Empfänger und Luftabsoprtion berücksichtigt (vgl. Brinkmann et al., 2019, S. 2748), andere Phänomene der tatsächlichen Schallausbreitung wie Interferenz jedoch nicht (vgl. Weinzierl, 2008, S. 242). In einem Vergleich von fünf Simulationsystemen von Brinkmann et al. (2019) konnte keine simulierte Auralisation eine physikalische Messung so exakt nachbilden, dass reale und simulierte Impulsantwort nicht mehr voneinander zu unterscheiden gewesen wären. Die Simulationen wurden jedoch überwiegend als plausibel bewertet (vgl. S. 2754).

2.5.2. Binaural synthese

Ziel der Binauralsynthese ist es, die richtungsspezifischen Informationen, die einem Schallsignal vom Außenohr aufgeprägt werden (vgl. Abschnitt 2.1.1), zu reproduzieren, um den gleichen räumlichen Eindruck zu erzeugen (vgl. Møller et al., 1995, S. 203). Solche binauralen Aufnahmen können mit einem Kunstkopf mit Mikrofonen in den künstlichen Gehörgängen durchgeführt werden (vgl. Paquier und Koehl, 2010, S.3). Um in der Wiedergabe die richtungsabhängige Verzerrung des Schalls bei der Kopfhörerwiedergabe korrekt abzubilden, müssen die Übertragungseigenschaften des an das Ohr angekoppelten Kopfhörers (auch headphone transfer function, HPTF) kompensiert werden (vgl. Paquier und Koehl, 2010, S. 1). Die dazu angewendeten Entzerrungsfilter können und sollten jedoch nicht die Übertragung im gesamten hörbaren Frequenzbereich kompensieren bzw. linearisieren. Einerseits begrenzen die Möglichkeiten der elektroakustischen Wandler, tiefe Frequenzen wiederzugeben, die Bandbreite, über die entzerrt werden sollte. Andrerseits sind Außenohrübertraungsfunktionen oberhalb von 12 kHz so individuell, dass ab dieser Frequenz nur grob entzerrt werden muss (vgl. Fleischmann et al., 2012, S. 5). Zudem führen Variationen in der Positionierung eines Kopfhörers auf den Ohren zu deutlich wahrnehmbaren klanglichen Unterschieden (vgl. Paquier und Koehl, 2010, S. 7), sodass einmal entworfene Entzerrungsfilter in Abhängigkeit der konkreten Position des Kopfhörers nicht immer gleichermaßen akkurat sind. Die HPTF-Messungen, auf denen die Entzerrungsfilter basieren, sind bei ohrumschließenden Kopfhörern zuverlässiger, als bei ohraufsitzenden Modellen (vgl. Paquier und Koehl, 2010, S. 3).

3. Methodik

In diesem Kapitel wird der Versuch, bzw. die Testverfahren beschrieben, die zur Erfassung der raumakustischen Hörexpertise entwickelt und durchgeführt wurden. Es wurden Hörtests zur direkten Messung der perzeptiven und kognitiven Fähigkeiten auf der einen Seite und Selbstauskünfte und Fachfragen auf der anderen verwendet, um das Verhältnis zwischen Sachverständigen, Sensorikern und spezialisierten Sensorikern nach den in Abschnitt 2.3.3 vorgestellten Kategorien untersuchen zu können. Der Begriff Hörtest wird hier dem des Hörversuchs vorgezogen, da es sich bei den Tests um Bestandteile eines einzigen Versuchs handelt, der von jedem Probanden einmalig und in voller Länge durchgeführt wurde. Da der Versuch unter anderem mit Schülern durchgeführt wurde, durfte die Bearbeitungsdauer maximal 45 Minuten betragen (vgl. Moosbrugger und Kelava, 2012, S. 35). Es wird angenommen, dass einerseits Laien ohne Erfahrung mit Hörversuchen für die Durchführung länger brauchen, um sich mit den Verfahren und dem generellen Konzept vertraut zu machen, einen Stimulus bezüglich des enthaltenen Raumklangs zu bewerten. Andererseits wird es als wahrscheinlich angesehen, dass einige Personen mit hoher fachlicher Expertise sich viel Zeit nehmen, um die Raumklänge kritisch zu analysieren. Dementsprechend wird von sehr unterschiedlichen Testdauern ausgegangen und die angestrebte durchschnittliche Dauer auf 35-40 Minuten verkürzt.

3.1. Hörtests

Die Wahrnehmung eines gehörten Raumklangs wird in dieser Arbeit als dreistufiger Vorgang modelliert. Zunächst wird der Schall vom Gehör wahrgenommen und an das auditorische System weitergeleitet. Nach Weinzierl und Vorländer (2015) wird dann auf kognitiver Ebene der Schall, der ursprünglich von der Quelle abgestrahlt wurde, von den Modulationen, die er durch die Übertragung durch einen Raum erfahren hat, separiert (vgl. S. 42). Zuletzt wird der Raumklang analysiert,um daraus Rückschüsse über die Beschaffenheit des Raumes zu ziehen. Vor allem im letzten Schritt käme die raumakustsiche Hörexpertise der Person zum Tragen. Es ist offensichtlich, dass die Performanz in den ersten beiden Stufen jeweils Voraussetzung für die in den nächsten sind. Somit muss, um die klanglichen Merkmale von Raum und Schallquelle zu trennen, ein Stimulus detailliert und umfassend wahrgenommen werden. Zudem können nur durch eine präzise und "korrekte' Abstraktion des Raumklangs differenzierte Aussagen über den Raum getroffen werden. Korrekte Abstraktion meint in diesem Fall, keine räumlichen Klangmerkmale der Schallquelle zuzuschreiben und andersherum. Um diesen dreistufigen Verarbeitungsprozess möglichst nachzubilden und die individuelle Leistung in allen diesen Etappen getrennt erfassen zu können, werden drei Hörtests durchgeführt. Die Ergebnisse sollen dann drei Fähigkeitsgrade liefern, deren Beziehung untereinander gemeinsam mit den weiteren Faktoren, die im Versuch erhoben werden, ausgewertet werden kann.

3.1.1. Testparadigmen

Um Messfehler durch nicht kontrollierte Entscheidungskriterien zu vermeiden, sind alle Hörtests als Forced-Choice-Verfahren konzipiert (siehe Abschnitt 2.3.2). In jedem Hörtest soll eine andere Komponente der auditorischen Verarbeitung geprüft werden. Dementsprechend ändern sich auch die methodischen Testparadigmen. Darüber hinaus flossen auch pragmatische Überlegungen in die Gestaltung der Testverfahren mit ein. Wie im Abschnitt 2.4.2 erklärt, folgen die Hörtests

den Paradigmen der probabilistischen Testtheorie und müssen in diesem ersten Versuch kalibriert werden. Das bedeutet, dass jeder Proband alle Items beantworten muss, was bei drei Hörtests und zusätzlichen Fragen zu Erfahrung, Musikalität und Fachwissen die Dauer der Versuchsdurchführung zu einem kritischen Faktor macht. Einerseits stellt die Hörermüdung auf perzeptiver Ebene ein Problem dar (siehe Abschnitt 2.1.1). Andererseits bedingen die vielen Testdurchläufe mehrerer Items die Gefahr von zunehmender Demotivation (vgl. Hambleton und van der Linden, Wim J., 1982, S. 347). Das kann ein gewissenhaftes Antwortverhalten generell einschränken und durch zunehmende Erschöpfung die Reihenfolge der Itembeantwortung zu einem Einflussfaktor machen, der das Prinzip unabhängiger Items des Birnbaum-Paradigmas verletzt (vgl. Baker, 2001, S. 131). Um die Unabhängigkeit der Item-Beantwortung von einander zu gewährleisten, erhält die Versuchsperson zudem keine Rückmeldung, ob eine gegebene Antwort korrekt ist. Zuletzt wird das Risiko der Demotivation bei einigen Probanden unabhängig von der Testdauer dadurch erhöht, dass eine Reihe von Fachfragen gestellt wird, die Probanden mit geringer oder keiner Fachexpertise nur eingeschränkt beantworten können.

Um solcher Frustration entgegenzuwirken wurde versucht, die Verfahren möglichst unterschiedlich zu gestalten. Das diente zudem dem Zweck, die Ratewahrscheinlichkeit, die auch in der probabilistischen Testtheorie die Interpretation der Ergebnisse beeinflusst (siehe Abschnitt 2.4.1 und Abbildung 2.2), möglichst gering zu halten und auch die Anzahl an Items zu minimieren. Die Ratewahrscheinlichkeit bestimmt sich jeweils über den Kehrwert der vorhandenen Antwortoptionen (vgl. Harrison und Müllensiefen, 2018, S. 5).

Da auch Laien den Versuch problemlos absolvieren können sollen, galt es außerdem, sprachliche Barrieren zu beachten und zu vermeiden. Gerade in den Hörtests, die perzeptive und kognitive Fähigkeit unabhängig vom Fachwissen erfassen sollen, galt es, ein möglichst allgemeinverständliches Vokabular zu verwenden (siehe Abschnitt 2.3.3).

Test 1: Raumakustische Wahrnehmungsfähigkeiten

Der erste Hörtest sollte die rein perzeptive Sensibilität für raumakustische Klangparameter messen. Wenn zwei identische Quellsignale verglichen werden und nur der simulierte Raum sich ändert, ist die Änderung des zunächst abstrakten Raumklangs gleichzeitig die einzige tatsächlich vorhandene und wahrnehmbare Änderung zwischen Stimuli. Durch einen derartigen Vergleich kann die Wahrnehmungsschwelle für bestimmte Formen klanglicher Unterschiede erfasst werden (just noticable difference, JND).

Das entwickelte Verfahren basiert auf der Idee eines Schiebereglers, auf dem eine Versuchsperson ein klangliches Merkmal eines Stimulus variieren und so einstellen kann, dass dieser veränderbare Stimulus sich für die Person genauso anhört wie eine gegebene Referenz. Die Abweichung der Antwort der Versuchsperson von der eigentlich korrekten Einstellung des Reglers zeigt, welchen Unterschied die Person nicht mehr wahrnimmt. Die Umgebung, in der der Versuch implementiert wurde, bietet jedoch keine Möglichkeit, Schieberegler einzubinden. Stattdessen wurde eine diskrete 21-stufige Skala eingesetzt. Jede Stufe der Skala entspricht einem Stimulus, der so moduliert ist, dass die Ausprägung eines raumakustischen Parameters von links nach rechts gleichmäßig zunimmt. Eine dieser Stufen repräsentiert den exakt gleichen Stimulus wie die Referenz.

Abbildung 3.1 zeigt die Oberfläche des Hörtests, beispielhaft für die Variation der Nachhallzeit. Die detaillierte Beschreibung der Implementierung des Versuchs folgt in Abschnitt 3.4.2. Die roten Schaltflächen repräsentieren die Stimuli, deren Wiedergabe durch Klicken der Schaltflächen gestartet und gestoppt werden kann. Unterhalb der Schaltflächen kann die jeweilige Stufe der Skala ausgewählt werden (in Abbildung 3.1 die Stufe 12). In keinem Test war zu Beginn eine Antwort vorausgewählt, sodass die Versuchsperson eine Option auswählen muss, um fortzufahren. So wird versucht, das Forced-Choice-Paradigma einzuhalten.

Referenz

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

weniger Hall

Wählen Sie auf der untenstehenden Skala den Klang aus, in dem der Raum die gleiche Nachhallzeit wie im Referenzklang hat.

Abbildung 3.1.: Versuchsinterface des ersten Hörtests. Klicken der roten Schaltflächen startet und stoppt die Wiedergabe der Stimuli. Durch Klicken auf den jeweiligen Kreis unter den Nummern-Schaltflächen kann die Stufe ausgewählt werden (hier Stufe 12). In diesem Beispiel wird die Nachhallzeit variiert.

Ein Vorteil dieses Verfahrens ist die eher geringe Ratewahrscheinlichkeit von 4,5 %. Um den Probanden das Konzept der linear angeordneten Skala verständlich zu machen, sind die Eckpunkte der Skala mit als allgemein verständlich angenommenen Beschriftungen versehen (in Abbildung 3.1 "weniger Hall"und "mehr Hall"), die den Verlauf des variierten raumakustischen Klangmerkmals beschreiben sollen. Vor der eigentlichen Durchführung wird den Versuchspersonen zudem eine Beispielskala mit einer ausführlichen Beschreibung gezeigt, damit sie sich mit dem Prinzip vertraut machen können. Dieses Beispiel sollte bereits eine minimale Gewöhnung an das Verfahren gewährleisten, da eine initiale Auseinandersetzung mit Stimuli und Testprozedur die Ergebnisse der ersten Testwiederholungen beeinflussen kann (vgl. Bech, 1992, S. 599).

Test 2: Wiedererkennen des Raumes

Im zweiten Hörtest sollte die Fähigkeit getestet werden, Räume als abstrakte Klangmodulationssysteme zu erfassen. Hier war der sprachliche Aspekt eine besondere Herausforderung, da diese kognitive Fähigkeit auch bei Probanden ohne raumakustische Fachkenntnisse und das entsprechende Vokabular untersucht werden sollte. Diese haben zwar oft ein intuitives Verständnis der klanglichen Eigenschaften dieses abstrahierten Übertragungssystems haben, können dieses aber vermutlich nicht differenziert oder gar nicht in Worte fassen (vgl. Weinzierl und Maempel, 2012, S. 316).

Deswegen ist der Test so angelegt, dass ein Raum von einem anderen unterschieden werden soll, ohne dass konkrete Angaben gemacht werden müssen, in welcher Hinsicht sich die Räume unterscheiden. Dazu werden in einem Forced-Choice-Verfahren drei Beobachtungsintervalle, d. h. drei Stimuli, präsentiert, zwei mit dem gleichen Raumklang und der dritte mit einem anderen. Die Versuchsperson muss nun lediglich angeben, in welchen Stimuli der Raum identisch sind. Damit bei dieser Aufgabe tatsächlich das kognitiv erfasste Konzept des Raumes abgefragt wird, werden für die Stimuluserzeugung jeweils unterschiedliche Ausgangssignale verwendet. Dieser Hörtest unterscheidet sich also vom ersten, da nicht die reine Diskriminationsähigkeit sensorischer Eindrücke abgefragt wird, denn alle Stimuli sind eindeutig verschieden. Um die Ergebnisse der ersten beiden Hörtests möglichst direkt vergleichen zu können, wurden in beiden Tests die selben raumakustischen Parameter anhand der gleichen Räume variiert (siehe Abschnitt 3.2.2).

Abbildung 3.2 zeigt die Benutzeroberfläche des Hörtests. Die Stimuli waren jeweils so angeordnet, dass sich die Räume von A und B unterschieden und der von X entweder dem von A oder B entsprach. In allen Items wurden die selben drei Quellsignale verwendet und die Zuweisung der Signale zu den verschiedenen Räumen und den Pseudonymen A, B, oder X für jedes Item zufällig neu angeordnet.

Hören Sie sich die Klänge A, B und X an. A und B stammen aus unterschiedlichen Räumen. Aus welchem Raum stammt X?



Abbildung 3.2.: Versuchsinterface des zweiten Hörtests. Klicken der roten Schaltflächen startet und stoppt die Wiedergabe der Stimuli. Die Räume der Stimuli A und B sind jeweils verschieden, der von X entspricht entweder A oder B. Durch Klicken auf den jeweiligen Kreis wird ausgewählt, welche Räume gleich sind (hier ist B ausgewählt).

Die Ratewahrscheinlichkeit liegt in diesem Verfahren mit 50 % sehr hoch. Theoretisch wäre es denkbar, mehr als zwei Räume zum Vergleich mit dem von Stimulus X anzubieten. Gerade für Laien, die einen evtl. nur sehr diffusen Raumklangeindruck von einem Gesamtklang auf einen anderen übertrage müssen, wird jedoch die Unterscheidung von nur zwei Räumen bereits als ausreichende Herausforderung angesehen.

Um die Testdauer anderweitig zu verkürzen wurde der Test so implementiert, dass in diesem Verfahren jeder Stimulus nur einmal angehört werden kann. Diese Methode wurde auch von Hameed et al. (2004) angewendet (vgl. S. 4), und macht den Test sehr anfällig für Momente der Unaufmerksamkeit da sich die Person z. B. beim Hören von X nicht mehr gut genug an A erinnern kann. Das wurde zugunsten der Dauer jedoch in Kauf genommen und zudem auch in diesem Test den Versuchspersonen zu Beginn ein Beispiel präsentiert, in dem zwei Räume mit erkennbar unterschiedlichen Nachhallzeiten verwendet wurden. Im Beispiel konnten die Stimuli beliebig oft angehört werden und es wurde auch explizit darauf hingewiesen, dass Raum A und X eine wesentlich höhere Nachhallzeit als B haben. Das sollte gerade Laien ein Gefühl dafür geben, wie ein Raumklang trotz wechselndem Quellsignal wiedererkannt werden kann. Vor dem eigentlichen Test folgte dann noch ein expliziter Hinweis, das nun keine Mehrfachwiedergabe mehr möglich ist. Die Einführungsbeispiele und Erklärungen zu den ersten beiden Hörtests befinden sich im Anhang.

Test 3: Raumvolumen vs. Nachhallzeit

Im letzten Hörtest sollten die Probanden nicht nur die Übertragungseigenschaften eines Raumes abstrahieren, sondern auch Aussagen über die Beschaffenheit desselben treffen. Dabei sollte ebenfalls vermieden werden, dass sehr spezifische Kenntnisse nötig sind um z.B. eine diskrete Raumresonanz bei einer bestimmten Frequenz erkennen zu können. Außerdem sollten im Sinne der probabilistischen Testtheorie eine eindeutig richtige und falsche Antworten zur Auswahl stehen. Aussagen wie etwa "Raum A hat eine starke Absorption bei hohen Frequenzen"lassen viel Interpretationsspielraum, welche Frequenzen als 'hoch'gelten und bedingen, dass man sowohl die Begriffe Absorption und Frequenz versteht. Damit würde die Notwendigkeit fachlicher Expertise zur Beantwortung der Tests nicht untersucht, sondern vorausgesetzt. Obwohl also Aussagen über einen Raum anhand seiner Akustik getroffen werden sollen, sollte keine akustisches Fachterminologie erforderlich sein.

Die anderen beiden Hörtests wurden zunächst unabhängig von den raumakustischen Parametern entwickelt, deren Einfluss schlussendlich untersucht wird. In diesem Fall ist das Verfahren direkt an einen spezifischen raumakustischen Sachverhalt gekoppelt und basiert auf der Feststellung von Shinn-Cunningham (2003), dass man problemlos kleine hallige von größen 'trockenen'Räumen unterscheiden kann (vgl. S. 4). Diese räumlichen Gegebenheiten implizieren, dass die Nachhallzeiten, anders als die Raumgrößen, nicht sehr unterschiedlich sind.

Das Raumvolumen beeinflusst den Wandabstand und damit vor allem die frühen Reflexionen, für deren Struktur der Mensch nach Hameed et al. (2004) eher unempfindlich ist (vgl. S. 2). In kleinen Räumen steigt durch kürzere Reflexionswege auch der Diffusschallpegel und die Gesamtlautheit eines Stimulus. Nach Hameed et al. (2004) konnten das einige Probanden korrekt dem Raum zuordnen, andere fassten dies als eine lautere Schallquelle auf (vgl. S. 3). Diese Diskrepanz deutet auf einen Einfluss individueller Hörexpertise hin.

Daraus wurde das Testparadigma dieses Hörtests abgeleitet dass Räume mit gleicher Nachhallzeit und variablem Volumen korrekt voneinander unterschieden werden müssen. Anders als im vorigen Test musste nicht festgestellt werden, welche beiden Räume sich klanglich unterschieden, sondern welcher größer ist. Dazu muss die Versuchsperson im Test den Raumklang in einem Stimulus erkennen, verstehen und damit den Raum physikalisch beurteilen. Der Begriff Nachhallzeit wird als intuitiv verständlich angenommen, sodass auch Laien die Aufgabe prinzipiell verstehen und beantworten können sollten.

Um die Ratewahrscheinlichkeit gegenüber z.B. dem obigen ABX-Test zu verringern und Abwechslung zu schaffen, wurden drei Räume präsentiert, die der Größe nach sortiert werden sollen. Den Probanden wird dazu die in Abbildung 3.3 dargestellte Matrix präsentiert, in der sie jedem Raum eine relative Größe zuweisen können.

	kleinster Raum	zweigrößter Raum	größter Raum
Raumklang anhören		•	
Raumklang anhören	•		
Raumklang anhören			•

Abbildung 3.3.: Versuchsinterface des dritten Hörtests. Klicken der roten Schaltflächen startet und stoppt die Wiedergabe der Stimuli. Durch Klicken auf die jeweiligen Kreise kann jeder Raum einer Größe zugeordnet werden.

Auch wenn nur drei Stimuli verwendet werden, gibt es die doppelte Anzahl an möglichen Kombinationen, wenn nie zwei Räume die gleiche Größe haben. Damit liegt die Ratewahrscheinlichkeit bei 16,6 %. Ausgerechnet drei Räume zu verwenden geht auf die Annahme zurück, dass relativ schnell zwischen zwei Räumen ein Größenverhältnis hergestellt werden kann und dann nur noch für einen weiteren Raum bestimmt werden muss, ob dieser eben größer als beide, kleiner die beiden oder größer als der eine und kleiner als der andere ist. Bei mehr Räumen werden deren Beziehungen untereinander komplizierter und die Bearbeitungsdauer des Test ggf. länger.

Da das Konzept der Sortierungsmatrix als intuitiv eingestuft wird und in diesem Test die Mehrfachwiedergabe der Stimuli wieder möglich ist (das wird für den Vergleich der drei Klänge untereinander als definitiv notwendig angesehen), wurde auf ein Einführungsbeispiel verzichtet.

Raumakustische Parameter und Gruppenaufteilung

Auch wenn die Anzahl an Items pro Test erst im nächsten Abschnitt besprochen wird, soll hier vorweggenommen werden, dass es nicht möglich war, eine Person die ersten beiden Hörtests für mehr als einen raumakustischen Parameter durchführen zu lassen, ohne die angestrebte Testzeit zu überschreiten. In Abschnitt 2.1.5 wurde bereits dargelegt, dass, auch wenn die Ergebnisse bisheriger Forschung sehr breit gestreut sind, die Wahrnehmung von Raumakustik von einigen zentralen Faktoren abhängt, die als variable Klangparameter in den ersten beiden Tests in Frage kommen. Da nur ein Parameter pro Person getestet werden konnte, wurde der Test in zwei Gruppen aufgeteilt, die jeweils mit einer anderen Größe getestet wurden.

In Abschnitt 2.1.5 wurden die Verstärkungswirkung, Halligkeit, Umhüllung, sowie die spektrale Balance der Nachhallzeit und die daraus resultierende Deutlichkeit und Nachhallzeit als wesentliche Faktoren der Wahrnehmung herausgearbeitet. Die mit dem Stärkemaß korrelierte Verstärkungswirkung (vgl. Soulodre und Bradley, 1995, S. 299), die Deutlichkeitsmaße und der Eindruck von Halligkeit hängen von der Nachhallzeit ab. Bleibt die von der Schallquelle abgestrahlte Energie über einen längeren Zeitraum als Schallenergie in Form von Reflexionen im Raum erhalten, steigt der Diffusschallpegel und das Stärkemaß nach Gleichung 2.3, in der der Diffusschall zusätzliche Energie in der Impulsantwort bedeutet. Dass ein langer Nachhall den Eindruck von Halligkeit verstärkt ist offenkundig. Bei der Berechnung des Deutlichkeits- bzw. Klarheitsmaßes steigt im Fall einer langsam abfallenden Nachhallkurve die Energie im späteren Teil der Impulsantwort und das Deutlichkeits- bzw. Klarheitsmaß sinkt. Die Auswirkungen auf die tatsächliche Verständlichkeit von Sprache und Musik hängt dann nach Fuchs (2018) von der spektralen Verteilung der Nachhallzeit ab (vgl. S. 42f).

Aus diesem Grund wurden die Dauer der Nachhallzeit und ihre Frequenzabhängigkeit, die auch als Klangfarbe des Nachhalls verstanden werden kann, als Parameter für die Test gewählt. Ein Vorteil dieser Größen ist neben ihrem Bezug zu wesentlichen Kriterien des auditorischen Raumeindrucks, dass die Begriffe und Konzepte auch für Laien einigermaßen zugänglich sind. Sowohl eine hohe oder niedrige Nachhallzeit als auch eine 'helle'der 'dumpfe'Klangfarbe werden als allgemeinverständliche semantische Differentiale angesehen. Zudem wird den Versuchspersonen im ersten Versuch ein Beispiel einer Skala gezeigt wird, auf dem das jeweilige Merkmal sich linear von einem 'Extremwert'zum anderen bewegt.

Dabei nicht abgebildet wird der Eindruck von Umhüllung, für dessen Variation eine Manipulation der Streuungsgrade der Wandoberflächen, die die Verteilung der Schallenergie im Raum verändern, eine wirksame Methode erscheint. Ein dritter Parameter hätte jedoch eine dritte Testgruppe mit sich gebracht, wodurch die, für belastbare Ergebnisse erforderliche, Probandenzahl zusätzlich ansteigt (siehe Abschnitt 2.4.2). Es wurde ausgerechnet die Umhüllung als Wahrnehmungsdimension außen vor gelassen wird, da das Konzept dieses Raumeindrucks für Laien weniger intuitiv verständlich erscheint als die Nachhallzeit und die Klangfarbe des Nachhalls.

Dementsprechend ist der Versuch in zwei Gruppen aufgeteilt. Bei Gruppe A wird in den ersten beiden Hörtests die Nachhallzeit der Räume variiert, bei Gruppe B die spektrale Balance derselben. Dass bei den Gruppen in beiden Tests mit dem jeweils gleichen Parameter getestet wird, soll den direkten Vergleich der Testergebnisse einer Person ermöglichen. Alle weiteren Testbestandteile, der dritte Hörtests und die nachfolgend diskutiere Abfragung von Musikalität und Fachexpertise und -Erfahrung, ist bei beiden Gruppen identisch.

3.1.2. Definition der Schwierigkeitsgrade

Die Prädiktion der Itemschwierigkeiten der Hörtests basiert auf einer einfachen Hypothese: je geringer der Unterschied zwischen zwei raumakustischen Parametern, desto schwieriger die Unterscheidung der jeweiligen Räume. Demnach ist die maximale Schwierigkeit durch die Wahrnehmungsschwelle begrenzt, ab der Unterschiede vom Gehör nicht mehr zuverlässig erfasst werden können. Im ersten Hörtest soll diese Wahrnehmungsschwelle jedoch erst festgestellt werden. Das bedeutet für die späteren Hörtests, dass für maximale Schwierigkeit arbiträre Unterschiede festgelegt werden müssen, die der Erwartung nach den tatsächlichen Wahrnehmungsschwellen entsprechen oder geringer sind.

Raumakustische Wahrnehmungsfähigkeiten

Im ersten Hörtest soll die Wahrnehmungsschwelle gemessen werden. Das ist im Kontext der probabilistischen Testtheorie wichtig, da die Beantwortungen der Items keine dichotomen Ergebnisse liefert. Es wird nicht nur gemessen, ob eine Referenz identifiziert wurde oder nicht, sondern wie groß die Abweichung ist. Damit ist auch die Bestimmung von verschiedenen Itemschwierigkeiten weniger problematisch, dass niedrigere oder höhere Wahrnehmungsschwellen nicht anhand von unterschiedlicher Beantwortung verschiedener Schwierigkeitsgrade bestimmt, sondern direkt metrisch erfasst werden können. Stattdessen sollten verschiedene Test-Items die gleiche latente Komponente abfragen. Für jede Person wird der individuelle Wert des latenten Faktors als Kennwert für die jeweilige Sensibilität für Unterschiede verwendet. Dieser Faktor wird auf Basis von Korrelationen berechnet und enthält keine Information über die absoluten Wahrnehmungsschwellen. Da für diese Arbeit vor allem von Interesse ist, wie das Verhältnis der gemessenen Sensbilität zu den Ergebnissen der anderen Tests ist, ist dieser Umstand weniger problematisch.

Das die bisher gemessenen Werte für die Wahrnehmungsschwelle von Nachhallzeiten breit gestreut sind, wurde in Abschnitt 2.1.5 gezeigt. Einheitlich ist jedoch - der DIN EN ISO 3382-1 (2009) folgend - die Messung und Angabe der Schwelle in Prozent, die auch hier übernommen wird. Aus den Ergebnissen von Belvins et al. (2013) geht ferner hervor, dass die Schwelle von der absoluten Dimension der Nachhallzeit zumindest im untersuchten Bereich von Referenznachhallzeiten zwischen einer und drei Sekunden unabhängig scheint (vgl. S. 4f). Um Itemunabhängigkeit zu gewährleisten, wurde bei jedem Item der Referenzstimulus geändert. Parallel wurde jeweils die Auflösung der Skala verändert, um auch eine leichte Varianz in der Itemschwierigkeit zu erzeugen. Die Annahme hierbei ist, dass eine Beantwortung schwieriger wird, je mehr Antwortoptionen es gibt, bei denen die Unterschiede zur Referenz im unmittelbaren Bereich der Wahrnehmungsschwelle oder darunter liegen.

Bleibt die Skala gleich und die Referenz ändert sich, variiert auch die Auflösung. Angenommen es wird für eine bestimmte Referenznachhallzeit eine Skala erstellt, auf der die Nachhallzeit mit jeder Stufe um $5\,\%$ (relativ zur Referenz) steigt bzw. sinkt. Wird nun in einem zweiten Item die Nachhallzeit der Referenz halbiert, steigt die Schrittweite auf der Skala auf $10\,\%$ der neuen Referenznachhallzeit.

Abbildung 3.4 zeigt links die sukzessiv um 5 % variierten Nachhallzeiten der Räume die in diesem Hörtest verwendet werden. Diese Modulationen sind bezogen auf einen 'initalen'Raum (gestrichelte Linie in der Abbildung), dessen Nachhallcharakteristik in Abschnitt 3.2.2 erklärt wird. Tabelle 3.1 zeigt die Variation der Referenzen und die daraus resultierenden Schrittweiten der Skala im ersten Hörtest. Die Referenz ist dabei bezogen auf die initiale Nachhallzeit des Musterraums. Eine Referenz von 100 % hat exakt diese Nachhallcharakteristik, im Falle einer 50 %-Modulation ist die Nachhallzeit ist halb so groß etc. Die Schrittweite von 5 % orientiert sich an der Wahrnehmungsschwelle der DIN EN ISO 3382-1 (2009). Da diese von Belvins et al. (2013)

und anderen als zu niedrig angesehen wurde (vgl. S. 1 und S. 4) und zudem bei zwei Items mit höheren Referenznachhallzeiten die Schrittweite damit noch darunter liegt, wurde dieser Wert als hinreichend geringe Auflösung eingestuft. Da die Auflösung bei zwei Items höher ist, kann gleichzeitig die Itemschwierigkeit merklich variiert werden, sollte die Wahrnehmungsschwelle einer Person tatsächlich bei dem in der Norm spezifizierten Wert liegen.

Der Raum, in dem die Simulationen durchgeführt werden, hat dabei in allen Fällen die gleiche Größe, es ändert sich nur die Absorption der Oberflächen. Wie in Abschnitt 4.1 erläutert wird, wurde während der Druchführung des Versuchs die Referenz des dritten Nachhallzeit-Items von $100\,\%$ auf $110\,\%$ geändert, weswegen in Tabelle 3.1 bei Item 3 zwei Werte angegeben sind.

Item-Nr.	Parameter	Referenz (rel.)	Schrittweite
1	Nachhallzeit (A)	50%	10,00%
2		75%	6,67%
3		100% bzw. $110%$	5,00% bzw. $4,54%$
4		125%	4,00%
5		150%	3,33%
1	Nachhallklangfarbe (B)	125%	4 %
2		150%	3,33%
3		175%	$2,\!86\%$
4		200%	2,50%
5		225%	$2{,}22\%$

Tabelle 3.1.: Dimensionierungen der Referenzen und die resultierende Schrittweite der Skala im ersten Hörtest. Die Referenz ist auf die Abweichung des jeweiligen Parameters von einem initialen Referenzraum bezogen. Die dritte Referenz der Nachhallzeitgruppe musste während der Durchführung von 100 % auf 110 % geändert werden.

Im Fall der Nachhallklangfarbe repräsentiert 100 % eine frequenzunabhängige Nachhallzeit, wobei ein Abfall der Nachhallzeit bei sehr hohen Frequenzen durch Luftabsorption unvermeidlich ist (vgl. Möser, 2005, S. 193). Die Manipulation der Klangfarbe wurde so umgesetzt, dass eine Modulation von 125 % eine Anhebung der Nachhallzeit im Oktavband von 32 Hz bedeutet und die Nachhallzeiten in den darüberliegenden Oktavbändern derart linear interpoliert werden, dass die Nachhallzeit bei 1 kHz in allen Räumen gleich ist und zu darüberliegenden Frequenzen weiter abfällt. Durch den exponentiell wachsenden Frequenzumfang der Oktavbändern bedeutet das einen exponentiellen Abfall der Nachhallzeit über die Frequenz. Die final daraus resultierenden Nachhallkurven sind rechts in Abbildung 3.4 dargestellt, die 100 %-Referenz ist auch hier als gestrichelte Linie markiert.

Wie auch bei der Nachhallzeit beträgt die Schrittweite der Klangfarbenmodulation (und die der Skala im Versuch) ausgehend von der 100 %-Referenz 5 %. Dieser Wert ist im Grunde arbiträr, da die Nachhallklangfarbe durch Standardparameter zur Beschreibung von Raumakustik kaum abgebildet wird (vgl. Weinzierl et al., 2018, S. 1255) und dementsprechend keine Referenzwerte bekannt sind. Da aber ebenso wie im Test der anderen Versuchsgruppe Nachhallzeiten variiert werden - wenn auch je nach Frequenz auf unterschiedliche Art - wurde die gleiche Schrittweite wie im Nachhallzeit-Test gewählt.

In Abbildung 3.4 weichen die Nachhallkurven in den oberen Oktavbändern von dem beschriebenen Verlauf ab. Das liegt daran, dass dort die Ergebnisse der Simulationen dargestellt sind, die die vorgegebenen Nachhallkurven nur approximmieren, wobei gerade bei hohen Frequenzen keine beliebigen Nachhallzeiten möglich sind (siehe Abschnitt 3.2.2). Ein Nachhallspektrum, das zu

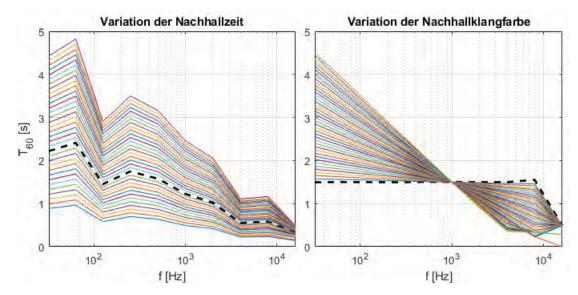


Abbildung 3.4.: Nachhallspektren der Räume, die in den ersten beiden Hörtests die Itemschwierigkeit variieren sollen. Links steigt die Nachhallzeit sukzessive an. Der spektrale Verlauf (siehe Abschnitt3.2.2) bleibt konstant. Rechts nimmt der Nachhall zu tiefen Frequenzen und fällt zu hohen hin ab, wobei die Nachhallzeit bei 1 kHz stets gleich bleibt. Da die Simulationsresultate abgebildet sind, weichen die Spektren zu sehr hohen Frequenzen und starker Coloration von dem linearen Verlauf ab.

tiefen Frequenzen abfällt und zu hohen ansteigt, erscheint eher unrealistisch. Hierfür wäre eine breitbandige Absorption erforderlich, die zu hohen Frequenzen abfällt, was sich in einem realen Raum mit üblichen porösen Breitbandabsorbern nicht umsetzen lässt, deren Absorption zu hohen Frequenzen tendenziell zunimmt (vgl. Weinzierl, 2008, S. 234f). Aus diesem Grund wurde die frequenzunabhängige Nachhallzeit in keinem Item als Referenz verwendet. Die Skalen waren in allen Items so angelegt, dass sie auf beiden Seiten der Referenz so viele Stufen enthielten, dass der Unterschied zwischen den Extremwerten der Skala und der Referenz mindestens 20 % betrug, was in etwa dem Mittelwert verschiedener Messungen der Nachhallzeits-Wahrnehmungsschwelle entspricht (vgl. Belvins et al., 2013, S. 2). So sollte der Versuchsperson die Möglichkeit gegeben werden, die Referenz von beiden Seiten einzugrenzen. Da nun bei der Variation der Nachhallklangfarbe die frequenzunabhängige Nachhallzeit (100 %-Modulation) einen äußersten Punkt der Skala darstellte, konnte diese nicht auch als Referenz eingesetzt werden.

Tatsächlich wurden 33 bzw. 36 Stimuli generiert, in denen sich die Nachhallzeit bzw. deren Klangfarbe sukzessive erhöht bzw. moduliert wurde. Ursprünglich waren für beide Varianten 36 Stufen angesetzt - die Nachhallzeit wäre von 1,5 s von 25 % bis 200 % variiert worden. Allerdings führte die Raumsimulation nur bis zu einem unteren Limit von 40 % zu genügend exakten Abbildungen der jeweils angestrebten Nachhallzeiten. Die Nachhallzeiten betrugen also das 0,4- bis 2-fache der initialen Nachhallzeit und die Klangfarbenmodulation reichte nach dem zuvor beschriebenen System vom Faktor 1 (frequenzunabhängig) bis zum Faktor 2,75. Es wurden jedoch in jedem Item immer nur 21 Skalenstufen präsentiert, sodass die Skala jedesmal einen andere Teilmenge der vorhandenen Räume abbildete. Zudem wurden, sofern die Einhaltung von minimal 20 % Prozent Abweichung den Extremwerten der Skala von der Referenz das ermöglichte, von jedem Item zwei Skalenausschnitte angelegt, von denen zufällig einer ausgewählt wurde. So verschob sich in jedem Item nicht nur die Referenz, sondern auch die Skala in eine andere Richtung, um zu vermeiden, dass ein Test-Item auf Basis des vorangegangenen einfacher zu be-

antworten ist.

Aus Tabelle 3.1 geht hervor, dass dieser Test nur aus fünf Items besteht. Mit der Itemanzahl steigt zwar generell die Präzision und Reliabilität der Testergebnisse (vgl. Moosbrugger und Kelava, 2012, S. 34f), in Anbetracht der niedrigen Ratewahrscheinlichkeit und der Aussagekraft einer metrisch skalierten Antwortvariable wird diese Itemanzahl jedoch als hinnehmbar angesehen, um die Durchführungsdauer in einem akzeptablen Umfang zu halten.

Wiedererkennen des Raumes

Für den zweiten Hörtest wurden die gleichen Raumimpulsantworten verwendet wie im ersten. Wegen der hohen Ratewahrscheinlichkeit wurde frühzeitig eine Anzahl von 20 Items festgelegt. In einer vergleichbaren Entwicklung eines Hörversuchs nach der probabilistischen Testtheorie verwendeten Harrison und Müllensiefen (2018) in der Kalibrierungsphase 27 verschiedene Items (vgl. S. 8). In dem hier durchgeführten Test die Anzahl etwas niedriger anzusetzen, ist ein erneutes Zugeständnis an die Dauer des Tests. Darüberhinaus gaben die Testpersonen in Vortests an, dass bereits die Durchführung von 20 Items gegen Ende sehr monoton und anstrengend würde.

In diesem Hörtest wurden pro Item jeweils zwei Räume miteinander verglichen. Um linear ansteigende Itemschwierigkeiten zu erzeugen, wurde ein Raum als eine Referenz festgelegt, die in allen Items gleich blieb. Im anderen Raum wurde die Nachhalllzeit bzw. -Klangfarbe mit jedem Item sukzessive erhöht. Da sich in diesem Test mit jedem Item zufällig änderte, welche Quellsignale mit dem gleichen Raum verhallt wurden und ob die Referenz oder der variierende Raum zwei mal vorkam, wurden keine erheblichen Effekte durch eine Gewöhnung an den Referenzraum erwartet. Die Modulationsschrittweite zwischen den Items betrug 10 % relativ zur Referenz. Wenn auch die genauen Wahrnehmungsschwellen noch nicht bekannt waren, wurde diese minimale Differenz als nahe genug der Schwellen angenommen, um bei wechselnden Quellsignalen die Räume nicht mehr zuverlässig zuordnen zu können. Wenn in 20 Items die Abweichung von der Referenz jeweils um 10 % erhöht wird, ergeben sich Modulationsgrade von 1,1 bis 3.

Im Fall der Veränderung der Nachhallzeit ergab sich die 10 % Schrittweite daraus, dass der Raum mit der halben Nachhallzeit (50 %) des initial definierten Raums als gleichbleibende Referenz festgelegt wurde. Der Raum mit der dreifachen Nachhallzeit dieses Raumes repräsentierte auf der Skala des ersten Tests die 150 %-Modulation, sodass alle Items mit den gleichen Räumen wie im ersten Test konstruiert werden konnten. Unter der Prämisse, dass die initiale Nachhallzeit, die nach der DIN 18041 (2016) festgelegt wurde (siehe Abschnitt 3.2.2), realistisch ist, bietet sich die gegebene Auswahl auch deshalb an, da durch Verwendung kleinerer und größerer Nachhallzeiten die Abweichungen vom initialen Raum insgesamt begrenzt werden. Wenn nur Nachhallzeiten oberhalb der initalen Nachhallcharakteristik verwendet würden, käme es bei gleichem Volumen zu sehr hohen Nachhallzeiten, die zu eher unrealistischen Raumsimulationen führen könnten.

Ähnliches gilt auch für die Variante des Hörtests, in der Klangfarbe des Nachhalls variiert. Das eine 'hellere'Klangfarbe als eine frequenzunabhängige Nachhallzeit eher unrealistisch erscheint wurde bereits erläutert, allerdings scheint auch ein beliebig steiler Anstieg der Nachhallzeit zu tiefen Frequenzen in realen Räumen nicht umsetzbar. Daher wird in diesem Test die frequenzunabhängige Nachhallzeit als Referenz gewählt. So ist im extremsten Fall die Nachhallzeit bei 32 Hz dreimal so hoch wie die bei 1 kHz und fällt bei darüberliegenden Frequenzen entsprechend steil ab. Diese Nachhallcharakteristik ist bereits sehr tiefenlastig, für die methodische Konsistenz des Verfahrens jedoch notwendig.

Es wurde bereits beschrieben, dass in diesem Versuch mit jedem Item sich nicht nur die Akustik des einen Raums ändert, sondern randomisiert auch die Verteilung der drei verschiedenen

		1. Hörtest	2. Hörtest
Item-Nr.	T_{60}	T_{60}/T_{Ref100}	T_{60}/T_{Ref50}
Referenz	$0,750\mathrm{s}$	0,50	1,00
1	$0,\!825\mathrm{s}$	0,55	1,10
2	$0,900{\rm s}$	0,60	1,20
3	$0,975{ m s}$	0,65	1,30
4	$1,050{\rm s}$	0,70	1,40
5	$1,125{ m s}$	0,75	1,50
6	$1,200{ m s}$	0,80	1,60
7	$1,275{ m s}$	0,85	1,70
8	$1,\!350{ m s}$	0,90	1,80
9	$1,425{ m s}$	0,95	1,90
10	$1,500{ m s}$	1,00	2,00
11	$1,575{ m s}$	1,05	2,10
12	$1,\!650{ m s}$	1,10	2,20
13	$1,725{ m s}$	1,15	2,30
14	$1,800\mathrm{s}$	1,20	2,40
15	$1,875\mathrm{s}$	1,25	2,50
16	$1,950{ m s}$	1,30	2,60
17	$2,025\mathrm{s}$	1,35	2,70
18	$2,100\mathrm{s}$	1,40	2,80
19	$2,175\mathrm{s}$	1,45	2,90
20	$2,\!250\mathrm{s}$	1,50	3,00

Tabelle 3.2.: Anstieg der Nachhallzeit in den Räumen des zweiten Hörtests. In der zweiten Spalte sind die absoluten mittleren Nachhallzeiten aufgelistet (die tatsächlichen Nachhallkurven folgen in Abschnitt 3.2.2). die dritte Spalte zeigt das Verhältnis zur initalen Referenznachhallzeit von 1,5 s nach DIN 18041 (2016, S. 13), die im ersten Hörtest als 100-Referenz bezeichnet wird (T_{Ref100}). Die letzte Spalte zeigt das Verhältnis dieser Nachhallzeiten zur Referenznachhallzeit des zweiten Hörtests, T_{Ref50} , von 750 ms.

Ausgangsstimuli auf die jeweils zu vergleichenden Räume. Wenn die Ausgangsstimuli jedoch ähnliche Spektren und Dynamiken besitzen und sich die klanglichen Effekte durch den Raum etwa ähnlich auswirken, kann die genaue Zuordnung der Ausgangsstimuli zu den Räumen als incidental eingestuft werden und hat keinen Einfluss auf die Itemschwierigkeit.

Raumvolumen vs. Nachhallzeit

Auch wenn die auditorische Wahrnehmung von Raumgröße noch nicht vollständig erforscht ist (vgl. Hameed et al., 2004, S. 2), wurde auch im dritten Hörtest dieses Versuchs nach der Prämisse vorgegangen, dass große Differenzen zwischen den vergleichenden akustischen Parametern einfacher zu erkennen sind. Auf diesen Test bezogen sind im Fall einer niedrigen Itemschwierigkeit die Unterschiede der Volumina der Räume besonders groß. Um einen gleichmäßigen Anstieg der vermeintlichen Schwierigkeitsgrade zu schaffen, wurde der mittlere der drei Räume, die es zu vergleichen galt, zwischen den Items nicht verändert. Die Volumina des größeren bzw. kleineren Raums nahmen somit kontinuierlich zu- bzw. ab. Darüberhinaus entspricht das Verhältnis des kleineren Raumes zum mittleren dem Reziprok des größeren zum mittleren, um auch innerhalb eines Items die Herstellung der Beziehungen der Räume gleichermaßen schwer oder einfach zu machen.

Doch wieviel größer bzw. kleiner sollen die Räume von einem Item zu nächsten werden? In der Wahrnehmung von Reizen und Reizänderungen dominieren nach dem Weber-Fechnerschen Gesetz logarithmische Zusammenhänge. Sowohl die Empfindung von Lautstärke und Tonhöhe (vgl. Möser, 2005, S. 5) als auch die z. B. von (visueller) Helligkeit sind proportional zum Logarithmus des physikalischen Reizes (vgl. Schmidt, 2013, S. 26). Aus diesem Grund wurden auch die Änderungen der Raumgrößen logarithmisch dimensioniert. Der Test besteht aus zwölf Items (nach den ersten Druchführungen wurden noch zwei weitere hinzugefügt, die Gründe folgen in Abschnitt 4.1). Dieser Umfang wurde in Anbetracht der Ratewahrscheinlichkeit von 16,67 % Prozent als genügend eingestuft.

Da die Fähigkeiten einer breit gestreute Zielgruppe erfasst werden sollten, musste auch die Schwierigkeit einen großen Bereich umfassen (vgl. Moosbrugger und Kelava, 2012, S. 33). Deshalb wurden teils sehr große Volumendifferenzen präsentiert, um leicht zu beantwortende Items in den Test zu integrieren. Nachdem in einem informellen Hörtest die doppelte bzw. halbe Größe des mittleren Referenzraums als noch zu uneindeutig eingestuft wurde, wurde das Vierfache des Referenzraumvolumens für den größten Raum und ein Viertel dieses Volumens für den kleinsten Raum als maximale Differenzen festgelegt.

	kleiner	Referenz	größer
Item-Nr.	V/V_{Ref}	V/V_{Ref}	V/V_{Ref}
1	0,89	1,00	1,12
2	0,79	1,00	1,26
3	0,71	1,00	1,41
4	0,62	1,00	1,59
5	$0,\!56$	1,00	1,78
6	0,50	1,00	2,00
7	0,45	1,00	2,24
8	0,40	1,00	$2,\!52$
9	$0,\!35$	1,00	2,83
10	0,31	1,00	3,17
11	0,28	1,00	$3,\!56$
12	$0,\!25$	1,00	4,00

Tabelle 3.3.: Größenverhältnisse der Räume des dritten Hörtests. V_Ref bleibt in allen Items unverändert, sodass die Volumina der anderen beiden Räume über die zwölf Items logarithmisch sinken bzw. steigen.

Für den jeweils größeren Raum galt es also, den Größenanstieg bis zum vierfachen Volumen in zwölf logarithmisch äquidistante Stufen einzuteilen. In Anlehnung an die Aufteilung musikalischer Intervalle, die in der gleichstufigen Stimmung eine Verdopplung der Frequenz (Oktave) in zwölf Stufen unterteilt (vgl. Weinzierl, 2008, S. 124) wurde das Volumen für den größeren Raum $V_{groß}$ als

$$V_{groß} = V_{Ref} \cdot 2^{\frac{x}{6}} \tag{3.1}$$

und das Volumen für den kleineren Raum V_{klein} als

$$V_{klein} = V_{Ref} \cdot \frac{1}{2^{\frac{x}{6}}} \tag{3.2}$$

berechnet, wobei V_{Ref} dem konstanten Volumen des mittleren Raumes entspricht und x = 1, 2, ..., 12 dem jeweiligen Item. Tabelle 3.3 zeigt die resultierenden Größenverhältnisse für alle Items. Die angenommene Itemschwierigkeit fällt zu den höheren Itemnummern ab, sodass Item 12 als das leichteste und Item 1 als das schwerste angenommen wird.

3.2. Teststimuli

3.2.1. Aufnahmen

Aus Abschnitt 3.1.1 geht hervor, dass für den zweiten Hörtest drei verschiedene Quellsignale bzw. Ausgangsstimuli notwendig sind. Im ersten und dritten Hörtest wird in allen Stimuli das gleiche Quellsignal verwendet, damit die gehörten Differenzen ausschließlich vom wechselnden Raumklang herrühren.

Der Eindruck, den ein Hörer sich von einem Raumklang anhand eines in diesem Raum präsentierten Schallsignal macht, hängt stark von der Beschaffenheit dieses Schallsignals ab (vgl. Weinzierl und Vorländer, 2015, S. 43). Sprach- und Musikaufnahmen repräsentieren die Schallereignisse, mit denen der Mensch im Alltag konfrontiert wird und bei denen er es eher gewohnt ist, konzentriert hinzuhören. Synthetische Signale wie weißes Rauschen weisen Energieanteile in einem breiten Frequenzspektrum auf (vgl. Möser, 2005, S. 13) und werden auch zur Messung der Nachhallzeit nach DIN EN ISO 3382-1 (2009) verwendet (vgl. S. 11). Dennoch wurde sich bei der Konzeption dieses Versuchs gegen den Einsatz synthetischer Messsignale als Stimuli entschieden. Gerade im raumakustischen Kontext ist ein wesentlicher Vorteil von Sprach- und Musiksignalen, dass die Schallquellen den meisten Probanden bekannt sein dürften, d. h. sie haben eine Vorstellung des Klangs gängiger Instrumente wie einem Klavier oder dem Klang von Stimme ohne die Einflüsse eines umgebenden Raums. Daher dürfte die Trennung von Quellsignal und Raumklang leichter fallen, als wenn die Klangcharakteristik der Schallquelle (z. B. eines Rauschgenerators) eher unbekannt ist.

Nach Weinzierl et al. (2018) erzeugen Sprach- und Musiksignale einen ähnlichen Eindruck der räumlichen Umgebung (vgl. S. 1255). Es werden jedoch ausschließlich Musiksignale eingesetzt. Sprachsignale transportieren immer einen sehr konkreten semantischen Inhalt, der die Probanden leicht davon ablenken kann, sich auf die eigentlich relevanten Komponente - die raumakustischen Klangeinflüsse - zu konzentrieren. Außerdem müssen kurze Stimulussegmente in ständiger Wiederholung angehört werden. Bei drei Hörtests mit teils vielen Wiederholungen ist es unvermeidbar, dass die Stimuli dadurch irgendwann monoton und lästig klingen. Für Sprache ist das jedoch ein besonders unnatürlicher Anwendungsfall, in der Musik hingegen ist die Wiederholung ein gängiges, teils explizit gefordertes Stilmittel (vgl. Luko, 2008, S. 133).

Die Stimuli wurden daher mit vier Musiksignalen generiert. In allen fällen handelte es sich um monophone Aufnahmen unter annähernd anechoischen Bedingungen, die der Datenbank der Raumsimulationssoftware RAVEN (Schröder und Vorländer, 2011) entnommen sind. Im ersten und letzten Hörtest wurde eine kurze Cellosequenz eingesetzt, für den zweiten Tests drei Passagen aus einer Aufnahme eines Klavierstücks. Dass für das Wiedererkennen des Raumes drei Signale der gleichen Schallquelle eingesetzt werden, liegt daran, dass Instrumente unabhängig vom umgebenden Raum eigene Abklingzeiten aufweisen (vgl. Weinzierl und Vorländer, 2015, S. 42). Auch wenn die Signale sich im Zuge des Testverfahrens unterscheiden mussten, sollten die Abklingphasen der Instrumente in allen Fällen gleich sein, damit Probanden diese nicht in unterschiedlicher Weise den Räumen zuordnen. Sonst hätten z. B. Probanden mit profunden Wissen um den Klang dieser Instrumente im Test einen Vorteil gehabt.

Abbildung 3.5 zeigt die Betragsspektren der vier Ausgangsstimuli. Da keine entsprechenden Informationen über die Aufnahmetechnik bekannt sind, konnte der physikalische Schallpegel, der aufgenommen wurde, nicht rekonstruiert werden. Aus der Darstellung geht hervor, dass alle Signale ein breitbandiges Spektrum aufweisen, wenn auch die Klavierpassagen deutlich weniger Energie unterhalb von 100 Hz enthalten als das Cello. Da Cello- und Klavierstimuli jedoch nie im gleichen Test eingesetzt wurden, wurde dies als hinnehmbar angesehen.

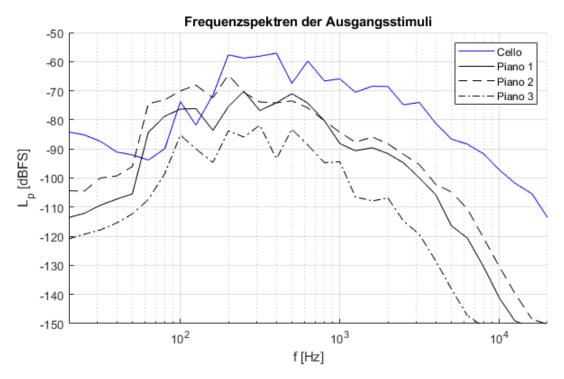


Abbildung 3.5.: Betragsspektren der Ausgangsstimuli in Terzbandauflösung. Die Terzbandpegel sind in auf die digitale Vollausteuerung bezogen, da keine Informationen zum physikalischen Aufnahmepegel vorliegen.

Bei den drei Ausgangsstimuli des zweiten Hörtests bildeten die Anforderungen einen Kompromiss aus gewollten Ähnlichkeiten und Unterschieden. Damit sich die Klangfarbe und Dauer des Nachhalls vergleichbar auf alle Quellsignale auswirken, sollte die spektrale Bandbreite und die Energie des Signals, das den Raum anregt, möglichst ähnlich sein. Gleichzeitig sollten auch deutliche Unterschiede zwischen den Klängen bestehen, damit der Raumklangeinfluss tatsächlich von merklich verschiedenen Quellsignalen abstrahiert werden muss.

Ein wenig kritisch ist die dritte Klavieraufnahme ("Piano 3"in Abbildung 3.5), die insgesamt niedrigere Terzbandpegel aufweist als die anderen beiden. Dies ist der Tatsache geschuldet, dass alle Klavierpassagen aus der Aufnahme eines Stückes von George Gershwin herausgeschnitten wurden. Wegen der eigenen Abklingzeit eines Klaviers gibt es nur wenige Stellen, an denen im Material geschnitten werden kann, ohne das ein unterbrochener Abklingvorgang hörbare Artefakte erzeugt. Gerade in Hörtests, in denen der Nachhall eines Instruments im Raum bewertet wird, rücken solche Artefakte am Stimulusende in den Fokus der Aufmerksamkeit und sollten daher vermieden werden, um die Authentizität des Wahrnehmungssituation nicht zu vermindern.

Aus diesem Grund wurde in Kauf genommen, dass es drei Passagen der gleichen Aufnahme des gleichen Instruments gibt, von denen eine einen insgesamt leiseren Schallpegel aufweist, als die anderen beiden. Auch wenn für alle Ausgangsstimuli in diesem Test aus einer Aufnahme stammen, konnte die zweite Anforderung, deutliche Unterschiede zwischen den Ausgangsklängen zu schaffen, erfüllt werden. Die erste Sequenz bildet eine kontinuierliche Folge kurzer Notenanschläge mit mittlerer und vereinzelt hoher Intensität, die zweite beginnt mit getragenem einzelnen Noten und geht dann in ein über in eine Folge zunehmend lauterer kurz andauernder Akkordklänge, während die dritte Passage ausschließlich aus leicht angeschlagenen, lang ausklingenden Akkorden und Einzeltönen besteht.

3.2.2. Raumsimulationen

Die Simulation der Raumimpulsantworten wurde mit dem C++-basierten Framework RA-VEN (Schröder und Vorländer, 2011) durchgeführt. Dieses Framework ist nicht nur kostenfrei für den akademischen Gebrauch verfügbar, sondern berücksichtigt auch Streuung und Beugung von Reflexionen und ermöglicht eine (hier nicht benutzte) Änderung der Quell- und Empfängerpositionen und der Raumgeometrie in Echtzeit (vgl. 1546). Teil von RAVEN ist eine Einbindung in die 3D-Modellierungs-Software SketchUp (Trimble Inc., 2000), sodass Auralisationen von in SketchUp erstellten Räumen erzeugt werden können (vgl. Pelzer et al., 2014, S. 113). Darüberhinaus können die mit dem SketchUp-Plugin generierten Raumkonfigurationen auch in der Software MATLAB (The MathWorks, 1994) bearbeitet und in dieser Umgebung Simulationen durchgeführt werden. Damit können erstellte Räume automatisiert variiert werden, was für die Erzeugung der Itemschwierigkeiten in diesen Hörtests sehr praktisch war.

Raummodelle

In Abschnitt 3.1.2 wurde erklärt, dass zur Variation der Itemschwierigkeit zahlreiche Raumkonfigurationen eingesetzt werden, die jedoch letztendlich auf zwei verschiedenen, 'initialen'Räumen basieren: ein Raum mit konstanter Größe und Geometrie, in dem in den ersten beiden Test jeweils die Nachhallzeit bzw. Nachhallklangfarbe durch Manipulation der Absorptionsgrade der Wandoberflächen variiert wurde und ein Raum, dessen Volumen im dritten Hörtest variiert wurde. Diese beiden initialen Räume wurden so erzeugt, dass eine Raumgeometrie und ein Zielnachhallzeitspektrum in Oktavbandauflösung vorgegeben wurde. RAVEN bietet die Möglichkeit, ein solches Nachhallspektrum für eine gegebene Raumgeometrie vorzugeben und iterativ die Raumparameter so manipulieren, dass das vorgewünschte Nachhallspektrum approximiert wird.

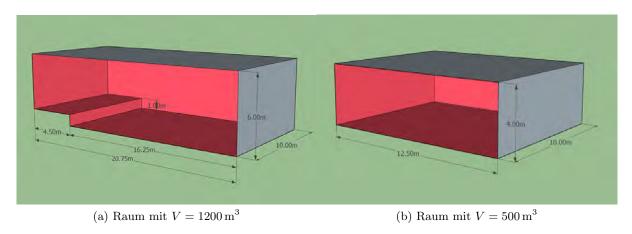


Abbildung 3.6.: SketchUp-Modelle der initialen Räume mit Abmessungen. Der erste Raum (a) wird in den ersten beiden Hörtests eingesetzt, Raum (b) im dritten Test.

Für die ersten beiden Hörtests wurde ein Volumen von 1200 m³ festgelegt. Vorbild des Raummodells, das in Abbildung 3.6a dargestellt ist, war ein großer Aufführungsraum mit Bühne. Die genauen Maße wurden letztendlich willkürlich festgelegt. Nach DIN 18041 (2016) sind für Musiknutzung eher Räume mit einem Volumen bis 1000³ typisch (vgl. S. 12f), Raumtyp und -Größe wurden jedoch unter der Annahme erstellt, dass gerade Laien die Auseinandersetzung mit Raumakustik mit dieser Art von Räumen assoziieren. Große Volumina bieten in diesem Zusammenhang generell den Vorteil, dass sie tendenziell große Nachhallzeiten verursachen, sodass auch Personen, die sich üblicherweise nicht mit Raumakustik beschäftigen, diesen Raumklang deutlich wahrnehmen. Im Laufe der Tests wurde diese Nachhallzeit freilich vielfältig variiert.

Dennoch sollten die initialen Räume einen Archetyp eines Raumes mit deutlich hörbaren Einflüssen auf die Schallübertragung darstellen. Auch wenn ein Raumvolumen von $1200\,\mathrm{m}^3$ nach der DIN 18041 (2016) für Musiknutzung wie gesagt untypisch ist, wird für diese Nutzungsart und Volumen eine Nachhallzeit von $T_{\mathrm{Soll}} = 1,5\,\mathrm{s}$ empfohlen.

Die spektrale Verteilung der Nachhallzeit ist im Fall der Hörtests, die mit der Nachhallklangfarbe als variablem Parameter testen, durch das Verfahren vorgegeben. Hier wird im Fall der frequenzunabhängigen Nachhallzeit eine konstante Abklingzeit von 1,5 s angestrebt (abgesehen von den obersten Frequenzbändern, in denen durch die bereits erwähnte Luftabsorption ein Abfall der Nachhallzeit gerade bei so großem Raumvolumen unvermeidlich ist).

Bei den Tests mit variabler Nachhallzeit ist der Frequenzgang derselben im Grunde irrelevant. Daher wurde die Nachhallzeit des initialen Raumes so festgelegt, dass zufällige Nachhallzeiten für jedes Oktavband generiert wurden, wobei der Wertebereich, an die, in der DIN 18041 (2016, S. 14) spezifizierten Toleranzbereiche angelehnt ist. Abbildung 3.7 zeigt die generierte Nachhallzeit und den Toleranzbereich. Da es sich dabei um eine Approximation zuvor zufällig generierter Zahlenwerte für die Nachhallzeit handelt, weicht die Simulationen stellenweise von den Vorgaben durch die Toleranzbereiche ab. Für letztere wurden die Grenzen aus der Norm so verändert, dass die Nachhallzeit in den oberen Frequenzbändern in jedem Fall abfällt, um zu vermeiden, dass zufällig ein Anstieg der Nachhallzeit in diesem Frequenzbereich entsteht, der in realen Räumen durch Luftabsorption so nicht auftreten würde. Parallel wurde die untere Toleranzgrenze bei 100 Hz etwas erhöht, unter der Annahme, dass sich realen Räumen die Nachhallzeit tendenziell relativ hoch ist. Außerdem wurden Toleranzbereiche für die Oktavbänder mit den Mittenfrequenzen 32 Hz und 16 kHz ergänzt. Für 16 kHz wurde ein natürlicher Abfall der Nachhallzeit angenähert. Für 32 Hz wurde die zunehmende Vergrößerung des unteren Toleranzbereichs in den darüberliegenden Oktavbändern fortgeführt und die obere Toleranz zu des darüberliegenden Oktavbands beibehalten.

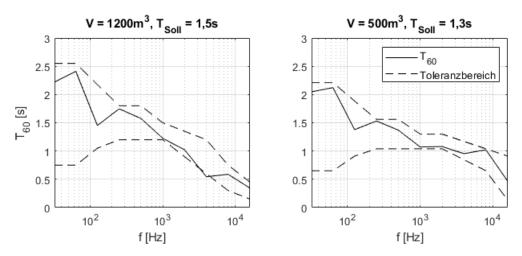


Abbildung 3.7.: Zufällig generierte initale Nachhallzeiten und Toleranzbereiche für zulässige Nachhallzeiten, basiert auf DIN 18041 (2016, S. 14) Die Nachhallzeiten sind bereits das Ergebnis der Raumsimulation, und überschreiten daher stellenweise den Tolernazbereich.

Für den letzten Hörtest wurde ein Quaderraum mit einem Ausgangsvolumen von $500\,\mathrm{m}^3$ festgelegt, dessen Raummodell und Abmessungen in Abbildung 3.6b dargestellt sind. Die Raumabmessungen durch die Variation des Volumens befinden sich im Anhang. Generell wurden jedoch alle Raumdimensionen gleichmäßig erhöht. Die Raumgröße des initialen Raums wurde für diesen Test etwas verringert, da sich daraus für die zu vergleichenden Räume Volumina zwischen $125\,\mathrm{m}^3$

und $2000\,\mathrm{m}^3$ ergeben. Damit sind auch mittlere Raumgrößen im Repertoire des Tests enthalten, mit denen die Probanden in ihrem Alltag eher in Kontakt kommen, als mit den Räumen, die sich im Falle eines Ausgangsvolumens von ebenfalls $1200\,\mathrm{m}^3$ ergäben (Modulationsbereich von $300\,\mathrm{m}^3$ bis $4800\,\mathrm{m}^3$).

Nach DIN 18041 (2016, S. 14) ergibt sich eine Soll-Nachhallzeit von 1,3 s. Der Frequenzgang der Nachhallzeit wurde analog zum vorigen Raum zufällig generiert und ist in Abbildung 3.7 dargestellt.

Quell- und Empfängerposition

Im ersten Raum wurde die Quelle mittig auf der Bühne platziert, der Empfänger frontal davor mit einer Entfernung von 10 m. Die Quelle befand sich 1 m Meter über der Bühnenhöhe von 1 m und der Empfänger in einer Höhe von 1,5 m. Die Höhen waren ein Kompromiss zwischen den in der (DIN EN ISO 3382-1, 2009, S. 10) vorgeschriebenen Messpositionen (1,5 für die Quelle und 1,2 m für den Empfänger), die von Weinzierl et al. (2018) auch für Auralisationen übernommen wurden (vgl. S. 1248) und der Bemühung, zu große Höhenunterschiede zwischen Quelle und Empfänger zu vermeiden. Im zweiten Raum betrug der Abstand 5 m, Quelle und Empfänger waren von beiden Seitenwänden gleich weit entfernt und so platziert, dass der Abstand der beiden zur jeweils rückwärtigen Wand identisch war. Sowohl Quelle als auch Empfänger befanden sich in diesem Raum 1,5 m über dem Boden.

Da im ersten Raum zwei verschiedene Instrumente als Quellen eingesetzt wurden und für ein Klavier in RAVEN keine Richtcharakteristik verfügbar ist, wurde für alle Räume und Instrumente als Empfänger eine omnidirektionale Richtcharakteristik festgelegt. Für das Cello hätte sich die korrekte Richtcharakteristik einstellen lassen. Da diese jedoch einen Einfluss auf die raumakustischen Parameter haben (vgl. Weinzierl und Vorländer, 2015, S. 45), wurde stattdessen für alle Situationen eine gleichermaßen unakkurate Kugelcharakteristik festgelegt, um zu vermeiden, dass die Simulationen in einigen Tests authentischer wirken als andere. Für den binauralen Empfänger wurde die Übertragungsfunktion des FABIAN-Kunstkopfes der TU Berlin (Brinkmann et al., 2017) für frontalen Schalleinfall (ohne Elevation der Schallquelle) festgelegt.

Approximation der Materialparameter

Die akkurate Simulation vorgegebener Nachhallzeiten war für die hier erstellten Hörtests von großer Bedeutung. Bei sukzessiver Modulation der Nachhallzeit, ihrer Klangfarbe und des Raumvolumens mussten die Simulationen so exakt sein, dass bei zwei zu vergleichenden Räumen die Differenzen der gezielt variierten Parameter stets größer waren als die Abweichungen zwischen approximierten und vorgegebenen Nachhallcharakteristiken, damit diese nicht die Itemschwierigkeit beeinträchtigten.

RAVEN bietet die Möglichkeit, Materialparameter (Absoroptions- und Streuungsgrad) eines Raummodells so zu approximieren, dass die Nachhallzeit eines Raumes sich einer definierten Vorgabe nähert. Diese Funktion wurde zur Simulation der Raumimpulsantworten iterativ eingesetzt. Als Startkonfiguration diente ein Material ohne Streuung und einem frequenzunabhängigem Absoprtionsgrad von 0,2. Die Iterationszahl wurde jeweils vorgegeben, zudem war die Iterationsfunktion so aufgebaut, dass die Schleife unterbrochen wurde, sobald die simulierte Nachhallzeit sich (in allen Oktavbändern) um weniger als 10 ms von der Vorgabe unterschied. Die Anzahl an Iterationsdurchläufen wurden, ebenso wie andere Simualtionsparameter durch Ausprobieren bestimmt. Das Ziel dabei war es, bei möglichst geringer Rechenzeit die vorgegebenen Nachhallzeiten in der Simulation möglichst exakt abzubilden.

Abbildung 3.8 zeigt die Abweichungen der Approximationen von den jeweiligen Soll-Nachhallzeiten für alle Räume. Bis 2 kHz weichen die Simulationsergebnisse kaum von den angestrebten Nachhallzeiten ab. In den darüberliegenden Oktavbändern kommt es gerade bei den verschiedenen

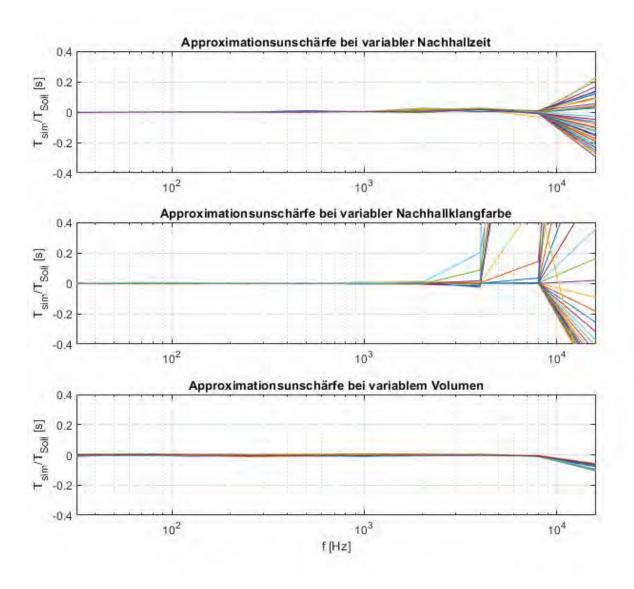


Abbildung 3.8.: Approximationsunschärfen der Nachhallzeiten aus Abbildung 3.4 und 3.7.

Klangfarben (mittlerer Graph) zu deutlichen Abweichungen, die in Abbildung 3.4 bereits deutlich zu sehen sind. Auch bei der Variation der Nachhallzeit weicht diese bei 16 kHz um bis zu ca. 220 ms vom Soll-Wert ab. Bei den variablen Volumina gibt es ebenfalls leichte Ungenauigkeiten. Dass RAVEN die Luftabsorption berücksichtigt (vgl. Brinkmann et al., 2019, S. 2748), könnte ein Grund sein, warum in den oberen Oktavbändern nicht beliebige Nachhallzeiten simuliert werden können. Die eingesetzten Insturmentalstücke weisen in den oberen Frequenzbändern ab 4kHz immer weniger Energie auf (siehe Abbildung 3.5). Zudem wird das Ohr für Frequenzen oberhalb von 4kHz zunehmend insensibler und hochfrequente Klangkomponenten werden tendenziell von den darunterliegenden verdeckt, die bei der Klangfarbenmodulation parallel zu den zunehmend Ungenauigkeiten immer dominanter werden (vgl. Abschnitt 2.1.1). Daher wurden die inakkuraten Simlationsergebnisse in den oberen drei Oktavbändern als hinnehmbar angesehen.

Lautheit

Die Verstärkungswirkung bzw. subjektiv empfundene Lautheit eines Raumes wurde in Abschnitt 2.1.5 als zentraler Aspekt der Wahrnehmung von Raumakustik identifiziert und ist zudem ei-

ner der auditorischen Parameter, für den es neben dem Stäremaß als physikalischen Prädiktor eine direkte Berechnung gibt, die in der DIN 45631 (1991) spezifiziert ist. Abbildung 3.9 zeigt, wie sich die Lautheit nach der DIN 45631 der Stimuli in Abhängigkeit der jeweils variierten Raumparameter ändert (man beachte, dass die X-Achse im Graph der variablen Volumina entsprechend der Modulationsschrittweite zwischen den Test-Items logarithmisch skaliert ist). Dargestellt werden die Lautheiten der Faltung der Instrumentalstimuli mit den jeweiligen Räumen (ohne Kopfhörerentzerrungen und Übertragungsfunktionen zu beachten). Da die Quellsignale jeweils identische Intensitäten haben, geben die angebenen Lautheiten direkt Aufschluss über die Verstärkungswirkungen der Räume.

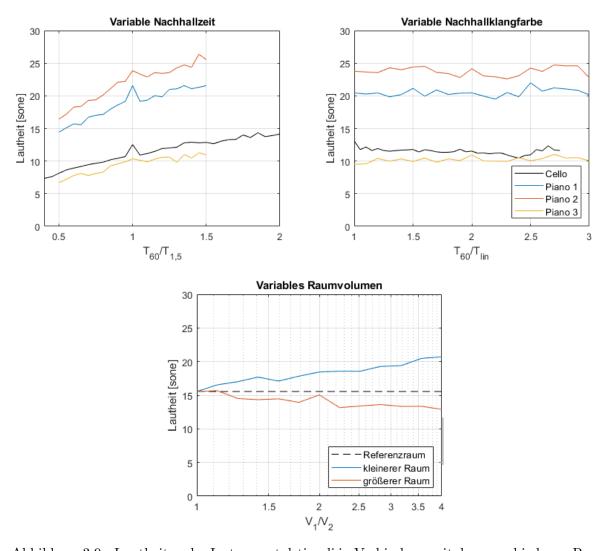


Abbildung 3.9.: Lautheiten der Instrumentalstimuli in Verbindung mit den verschiedenen Raumklängen, berechnet nach DIN 45631 (1991). Im Fall des variablen Volumens stellt die X-Achse jeweils das Verhältnis zum nächstkleineren Raum dar, also einerseits das der größeren zur Referenz und andrerseits das der Referenz zu den kleineren Räumen. Zusätzlich ist die Lautheit des Referenzraums eingezeichnet.

In den ersten beiden Hörtests wurde versucht, die Itemschwierigkeiten und das Verfahren für die beiden unterschiedlichen raumakustischen Testgrößen möglichst analog zu gestalten. Hier zeigt sich nun ein klarer Unterschied zwischen den beiden Varianten, da im Fall der variablen Nachhallzeit die Lautheit mit der Nachhallzeit durch eine zunehmende Überlagerung mit Refle-

xionen in allen Frequenzbändern ansteigt, während sie im Fall der variierten Nachhallklangfarbe etwa gleich bleibt. Das macht deutlich, dass die Lautheit auf der Frequenzbewertung des Gehörs (siehe Abschnitt 2.1.1) basiert (vgl. DIN 45631, 1991, S. 2f), sodass sich der Anstieg der Nachhallzeit bei tiefen Frequenzen ebenso wie der Abfall bei hohen weniger stark auf die Lautheit auswirken, als die gleichbleibende Nachhallzeit bei 1 kHz (siehe Abbildung 3.4). Damit ist die Lautheit - zumindest im Falle des ersten Hörtests mit nur einem Instrumentalstimulus - ein mögliches Diskriminationskritierium zwischen den einzelnen Stimuli, das bei der Testung mit variabler Nachhallzeit stärker ausgeprägt war als bei der Variation der Nachhallklangfarbe. Deshalb könnte die Unterscheidung einzelner Stimuli im letzteren Falle insgesamt schwerer sein.

Es wird außerdem deutlich, dass im dritten Hörtest mit den Volumendifferenzen der drei jeweils zu vergleichenden Räume auch die Lautheitsunterschiede zunehmend größer werden. Dabei steigen die Lautheiten der größeren Räume starker an, als die der kleineren abfallen. Außerdem ist die Lautheit im Item, bei dem der kleinere Raum halb so groß wie die Referenz ist, deutlich höher als bei den benachbarten Raumgrößen. Gleiches gilt für den initialen Referenzraum im Falle der variablen Nachhallzeit $(T_{60}/T_{Ref} = 1 \text{ im linken, oberen Graph})$. In diesem Zusammenhang darf nicht vergessen werden, dass sowohl die Vorgaben, als auch die Approximationen von frequenzhabhängigen Nachhallzeiten in RAVEN prinzipiell in Oktavbandauflösung durchgeführt werden. Innerhalb eines Oktavbandes wid dabei der Mittelwert mehrerer Nachhallzeiten gebildet. Dabei ist es durchaus möglich dass in den Simulationen die Nachhallzeit bei einzelnen Frequenzen deutlich über der Vorgabe für das jeweilige Oktavband liegen. Wenn nun eine solche überdurchschnittlich hohe Nachhallzeit bei einer Frequenz auftritt, in der auch das Quellsignal hohe Intensitäten aufweist, könnte eine wesentlich höhere Lautheit entstehen als bei Räumen, deren Nachhallzeit sich - in Oktavbandauflösung - nur geringfügig von der des ersten Raums unterscheidet. Da kurz vor der Versuchsdurchführung kurzfristig die Konfigurationen der Simulationen geändert und diese erneut zeitaufwändig berechnet werden mussten, sind diese lokalen Lautheitsanstiege erst während der ersten Durchführung aufgefallen, wie in Abschnitt 4.1 dargelegt wird.

3.3. Korrelate

In diesem Abschnitt werden die Teile des Versuchs erläutert, die zusätzlich zu den Hörtests die fachliche Kenntnis und Erfahrung der Versuchsperson sowie ihren generellen Bezug zu Musik abfragen sollten. Für letzteres wurden Komponenten der Gold-MSI von Müllensiefen et al. (2014) verwendet. Fachliche Expertise und Erfahrung wurden mit für diesen Versuch konstruierten Fragebögen erhoben.

3.3.1. Fachliche Expertise und Erfahrung

Erfahrung und fachliche Expertise wurden zunächst über Selbstauskünfte erhoben. Der Empfehlung von Weinzierl et al. (2018) folgend, wurde die Dauer der akademischen und beruflichen Auseinandersetzung mit Raumakustik ermittelt. Zusätzlich wurde auch die Beschäftigung mit musikalischen Aufnahmen im Rahmen von Freizeitaktivitäten abgefragt. Der konkrete Bezug zu Musik-Aufnahmen in der Fragestellung sollte vermeiden, dass in Zeiten in denen mit Mobiltelefonen jederzeit Aufnahmen erstellt werden können, auch die Erfahrung mit solchen Aufnahmen angegeben werden, wie etwa beim Versenden von Sprachnachrichten in Kurznachrichtendiensten. Es gab jeweils eine 5-stufige Antwortskala für den Grad an Zustimmung zu einer gegebenen Aussage.

Da die Beschäftigung mit Raumakustik im Rahmen der akademischen Laufbahn nur bedingt

Aufschluss über die tatsächlichen Kenntnisse auf diesem Fachgebiet gibt (es wurde nur die Dauer der Beschäftigung abgefragt), enthielt der Versuch zusätzlich acht konkrete Fragen zur Raumakustik, bei denen die Probanden eine von vier möglichen Antworten auswählen müssen. Die Fragen wurden in Zusammenarbeit mit einem Universitätsprofessor für Audiotechnik entworfen und sind ebenfalls im Sinne der probabilistischen Testtheorie aufgebaut, d. h. es wurde bei der Auswahl der Fragen versucht, unterschiedliche Schwierigkeitsgrade zu schaffen. Außerdem gibt es keine Möglichkeit, anzugeben, dass man die Antwort nicht kennt, sodass jede Frage eine dichotom verteilte Antwortvariable liefert.

Das vollständige "Raumakustik-Quiz'ist im Anhang aufgeführt. Eine allgemeingültige Itemschwierigkeit festzulegen ist insofern schwierig, als dass letztendlich einzelne Sachverhalte des gesamten Themenfeldes der Raumakustik mit teilweise sehr ähnlichen Antworten abgefragt wurden. Es bedeutet aber nicht zwingend eine geringe Fachexpertise, wenn eine Person eine bestimmte, in dem Moment notwendige Information nicht auswendig wusste. Deswegen waren die niedrigen Schwierigkeitsgrade so angelegt, dass Zusammenhänge abgefragt werden, die auch mit Prinzipien aus der allgemeinen Physik beantwortet werden können, wie die Energieerhaltung oder der Einfluss der Temperatur auf die Schallgeschwindigkeit (vgl. Möser, 2005, S. 22). Beispielhaft ist hier die Frage zu nennen, wie sich die Nachhallzeit im Raum ändert, wenn ein Fenster geöffnet wird. Dabei wird davon ausgegangen, dass keine Fachkenntnisse notwendig sind, um sich zu erschließen, dass dabei Reflexionsfläche verschwindet und Schallenergie dem Raum entweichen kann, wodurch sich die Nachhallzeit verringert.

Zudem wurde versucht, möglichst allgemein verständliche Sprache zu verwenden, auch wenn eher spezifische Begriffe wie Absorber oder stehende Wellen in ein paar Fragen unvermeidbar waren. Fragen, bei denen eine 'mittlere'Schwierigkeit vermutet wurde, fragten akustische Grundkenntnisse ab, etwa wie der Schalldruckpegelabfall mit wachsender Entfernung unter Freifeldbedingungen und mit einer Kugelschallquelle sinkt. Fragen mit höherer Schwierigkeit behandelten dann konkrete Zusammenhänge wie die Ausbreitung von stehenden Wellen oder die Wirkungsweise von porösen Absorbern.

3.3.2. Gold-MSI

Als letzte Komponente wurden drei Faktoren des Gold-MSI (Goldsmiths Musical Sohpistication Index) abgefragt. Der Gold-MSI ist ein Messinstrument zur Erfassung der musical sophistication bzw. "musikalischen Erfahrenheit"(Schaal, 2014, S. 426), die vielfältige Aspekte von musikalischer Expertise zusammenfasst, die über die Musiklität im Sinne der Beherrschung eines Instruments oder musiktheoretischen Kenntnissen hinausgehen (vgl. Müllensiefen et al., 2014, S. 1). Das gesamte Frageninventar umasst 38 Teilfragen, die mit einem siebenstufigen Grad an Zustimmung beantwortet werden.

Aus den Angaben können ein Gesamt-Index für musikalische Erfahrenheit und die fünf Faktoren "Aktiver Umgang mit Musik", "Musikalische Wahrnehmungsfähigkeiten", "Musikalische Ausbildung", "Gesangsfähigkeiten" und "Emotionen" bestimmt werden (Schaal, 2014, S. 429). Die ersten drei Faktoren wurden auch in diesem Versuch als Kovariate zur raumakustischen Hörexpertise erhoben.

Aktiver Umgang mit Musik meint im Kontext des Gold-MSI die aktive Beschäftigung mit Musik im Alltag, d. h. wie viel Musik täglich gehört wird und wie aktiv z. B. eine Auseinandersetzung mit neuer Musik stattfindet (vgl. Müllensiefen et al., 2014, S. 5). Für die Bestimmung der musikalischen Wahrnehmungsfähigkeiten werden Selbstauskünfte über kognitive Fähigkeiten im Zusammenhang mit Musik gegeben, z. B. Gesang bezüglich der tonalen und rhythmischen Korrektheit zu bewerten (vgl. Müllensiefen et al., 2014, S. 5). In einer Folgestudie konnte eine verhältnismäßig hohe Korrelation dieser Selbstauskunft mit in Laborversuchen gemessenen Hörfähigkeiten nach-

gewiesen werden (vgl. Müllensiefen et al., 2014, S. 9). Der letzte hier berücksichtigte Faktor, "Musikalische Ausbildung", wird aus Angaben zu Intrumentalunterricht und -Praxis und der Bildung in Musiktheorie bestimmt (vgl. Müllensiefen et al., 2014, S. 5). Die einzelnen Teilfragen, aus denen die Faktoren berechnet werden, sind den öffentlich zugänglichen Zusatzunterlagen von Müllensiefen et al. (2014) entnommen¹. Für die deutsche Version dieses Versuchs wurde die validierte Übersetzung des Gold-MSI von Schaal (2014) verwendet.

3.4. Versuchsaufbau

Der Versuch wurde an vier verschiedenen Orten durchgeführt: der Technischen Universtiät Berlin (TU), der Hochschule Düsseldorf (HS), dem Goldsmiths, Universtiy of London und dem Albert-Einstein-Gymnasium (AEG) in Kaarst nahe Neuss. Da die Items der Hörtests und des Raumakustik-Quiz' kalibriert werden mussten, war eine große Stichprobe erforderlich, wie in Abschnitt 2.4.2 bereits erläutert wurde. Die Durchführung auf mehrere, geografisch entfernte Orte zu verteilen, erhöht nicht nur die Repräsentanz der Stichprobe (vgl. Scriven, 2005, S. 528), sondern auch die Anzahl möglicher Kandidaten, die sich für den Versuch interessieren könnten. Wie in Abschnitt 3.5 gezeigt wird, konnte durch die konkrete Auswahl der Institutionen zudem ein breites Spektrum professioneller Hintergründe erreicht werden. Die Durchführung an mehreren Orten machte einen mobilen, schlichten Versuchsaufbau erforderlich, der trotz der unterschiedlichen räumlichen Gegebenheiten überall aufgebaut werden konnte. Daher wurde auch eine Wiedergabe der Audiostimuli über Kopfhörer vorgezogen (siehe Abschnitt 2.5.2). Zudem war die Implementierung des Versuchs im Online-Umfragesystem LimeSurvey in diesem Zusammenhang sehr nützlich, da für diese Testform keine besondere Software beötigt wird.

3.4.1. Struktur des Versuchs

Der vollständige, im Zuge dieser Arbeit durchgeführte Versuch umfasste die drei Hörtests, Selbstauskünfte zur fachlichen Erfahrung und Expertise mit Raumakustik, acht konkrete Fachfragen und die drei Faktoren des Gold-MSI, die sich aus 25 Teilfragen zusammensetzen. Zusätzlich wurden zu Beginn des Versuchs Alter, Geschlecht und das Vorhandensein einer Hörschädigung abgefragt. Soziodemographische Aspekte wurden nicht abgefragt, da relevante Aspekte der Bildung bereits durch die Selbstauskünfte abgefragt werden.

In Abschnitt 3.1.2 wurde bereits ausführlich die Problematik der Durchführungsdauer angesichts der zahlreichen Teilverfahren behandelt und eine maximale Dauer von 35-40 Minuten festgelegt. Durch die Reihenfolge der Teilverfahren sollte diese Zeitspanne zusätzlich weniger belastend gestaltet werden. Nach 15 Minuten Daueraufmerksamkeit kann es zu Leistungseinbußen der Versuchsperson kommen (vgl. Schmitz, 2004, S. 7). Die Hörtest wurden als Situationen eingestuft, die eine derartige Daueraufmerksamkeit erfordern. Aus diesem Grund und um der in Abschnitt 2.1.1 genannten Hörermüdung vorzubeugen, wurden nicht nur die Hörtests jeweils so kurz wie möglich gehalten, sondern auch Hörtests, Selbstauskünfte und Fachfragen abwechselnd präsentiert. Dieser Wechsel wurde den Probanden zu Beginn des Versuchs erklärt und außerdem während der Durchführung immer wieder darauf hingewiesen, welcher Test als nächstes folgt.

Zu Beginn des Tests wurde die Institution abgefragt, an der der Hörtest durchgeführt wird. Dies war für eine adäquate Kopfhörerwiedergabe erforderlich und wird im folgenden Abschnitt behandelt. Anschließend wurden die allgemeinen Informationen, Alter, Geschlecht und das Vorhandensein einer Hörschädigung abgefragt, gefolgt von der bisherigen Erfahrung mit Raumakustik. Als nächstes der erste Hörtest, bei dem auch unerfahrenen Probanden die Skala des

¹digital verfügbar unter doi:10.1371/journal.pone.0089642

einführenden Beispiels nutzen konnten, um sich ein Bild von den klanglichen Auswirkungen einer ansteigenden Nachhallzeit oder veränderlichen Nachhallklangfarbe zu machen. Nachdem auf den ersten Hörtest zunächst die Gold-MSI-Fragen zu den musikalischen Wahrnehmungsfähigkeiten folgten, bildete der zweite Hörtest die Mitte des Versuchs, der wegen der hohen Item-Anzahl als der eintönigste und ermüdenste angesehen wurde.

Nach dem zweiten Hörtest folgten weitere Fragen des Gold-MSI zum aktiven Umgang mit Musik. Die anschließenden Fachfragen zur Raumakustik erforderten wieder etwas mehr Konzentration. Danach kam der letzte Hörtest, der wegen der etwas abwechslungsreicheren Sortieraufgabe als geeignet für diesen späten Teil des Versuchs angesehen wurde. Den Schluss bildeten dann die Fragen des Gold-MSIs zur musikalischen Ausbildung. So wurde auch vermieden, dass die Versuchsperson zu viele Selbstauskunftfragen hintereinander beantworten musste.

3.4.2. Implementierung in LimeSurvey

Der Versuch wurde als Online-Fragebogen in LimeSurvey implementiert. Gerade die Randomisierung der Hörtests war in dieser Umgebung eher umständlich. Der Vorteil von LimeSurvey war jedoch, dass keine besondere Software erforderlich war, um den Fragebogen an einem Computer durchzuführen. Die Umfragen werden über einen Link in einem Internetbrowser aufgerufen und durchgeführt. Sowohl Umfragen als auch die Ergebnisse werden auf einem Server gespeichert und sind online überall zugängig. Damit ließ sich der Versuch problemlos an mehreren Orten bzw. Rechnern simultan durchführen.

LimeSurvey bietet die Möglichkeit, Designvorlagen zu verwenden und eigene Gestaltungskonfigurationen in CSS und Funktionalitäten in Javascript hinzuzufügen (vgl. LimeSurvey GmbH, 11.02.2020). Der Javascript-Code und die Definition der CSS-Klassen zur Implememtierung der im folgenden beschriebenen Funktionalitäten befinden sich im digitalen Anhang. Ferner gibt es in LimeSurvey 28 verschiedene Fragetypen (Mehrfachauswahl, Texteingabem etc.), die zur Erstellung von Umfragen erstellt werden können. Neben simplen Textanzeigen und der Zahleneingabe des Alters wurden zwei dieser Typen verwendet. Einfachauswahl-Fragen bieten zu einer Frage eine Reihe von Antwortoptionen, von denen nur eine ausgewählt werden kann (LimeSurvey GmbH, 03.04.2019, vgl.). Neben den Fachfragen wurde auch der zweite Hörtest (siehe Abbildung 3.2) als dieser Fragetyp implementiert. Bei sogenannten "Matrix-Fragen"wird eine Reihe von Unterfragen angeboten, für die jeweils die gleichen Antwortoptionen zur Wahl stehen (vgl. LimeSurvey GmbH, 03.04.2019). Die Implementierung des dritten Hörtests ist ein typisches Beispiel einer solchen Matrix-Frage (siehe Abbildung 3.3). Auch die Teilfragen der Gold-MSI-Faktoren und auch der erste Hörtest waren Matrix-Fragen, bei letzterem gab es jedoch immer nur eine "Teilfrage" (siehe Abbildung 3.1).

Audiowiedergabe

Da die Lautheit der Stimuli ein wichtiger Einflussfaktor auf die Wahrnehmung von Raumakustik ist (siehe Abschnitt 2.1.5), der sich wähnred der Hörtests merklich änderte (siehe Abschnitt 3.2.2), musste die Wiedergabelautstärke im gesamten Versuch annähnernd konstant sein. Das ließ sich bei der Kopfhörerwiedergabe nicht vollständig realisieren, jedoch sollte wenigstens vor der D/A-Wandlung der Audiosignale im gesamten Versuch der gleiche Signalpegel garantiert werden. Dazu ist es notwendig, dass die Versuchsperson keine Kontrolle über die Lautstärke hat.

Aus diesem Grund wurden den Versuchspersonen nur die roten Schaltflächen präsentiert, die in den Abbildung 3.1, 3.2 und 3.3 bereits gezeigt wurden. Über diese konnte Wiedergabe abwechselnd gestartet und gestoppt werden. Nach Stoppen und erneutem Start begann die Stimuluswiedergabe von vorn. Um ein unmittelbares Wechseln zwischen zwei Audiowiedergaben

für den direkten Vergleich zu ermöglichen, stoppt der Start der Wiedergabe eines Stimulus alle weiteren aktiven Wiedergaben.

Diese Mechanismen wurde über Javascript-Funktionen realisiert, in denen die Interaktion zwischen den Schaltfächen und nicht sichtbare HTML-Audioplayern (jeweils einer für jeden Stimulus) umgesetzt wurde. Im zweiten Hörversuch sollte jeder Stimulus nur einmal angehört werden können. Dazu wurde eine separate Funktion implementiert, die jede Schaltfäche nach einmaliger Betätigung deaktivierte Die Deaktivierung wurde der Versuchsperson durch eine Farbänderung der Schalftläche angezeigt, die über eine CSS-Klasse definiert wurde.

Kopfhörerauswahl

Der Versuch wurde mit verschiedenen Kopfhörern durchgeführt, für die die Stimuli jeweils individuell entzerrt werden mussten. Es gab für jeden Stimulus separate Audiodateien für jede Entzerrung, von denen immer die richtige ausgewählt werden musste.

Da der Versuch parallel mit verschiedenen Kopfhörern durchgeführt wurde, musste dynamisch eine der verschiedenen Entzerrungen, gewählt werden können. Aus diesem Grund wurde zu Beginn des Versuchs gefragt, an welcher Institution der Versuch durchgeführt wird. Je nach dem, welche Antwort dabei gegeben wurde, wurde in einer Variable gespeichert, für welchen Kopfhörer die Stimuli jeweils geladen werden sollen. Diese Variable wurde lokal im Browser gespeichert (vgl. Mozilla, 2005), da in LimeSurvey mit dem Aufruf einer neuen Frage jeweils eine neue Website geladen wird.

In den Audioplayern, die über die Schaltflächen angesteuert wurden, waren als Quelldateipfade tatsächlich nicht die vollständigen Namen von Audiodateien enthalten. Die Dateien der Stimuli folgen einer festen Nomenklatur, wobei Hörtest und Modulationsgrad des Raumes definiert wurden und zudem jeweils eine Kennung für den Ausgangsstimulus und die Kopfhörerentzerrung spezifiziert waren. Die Kennung des Kopfhörers war immer am Ende des Dateinamens. Wurde nun eine Schaltfläche zur Audiowiedergabe erstmalig gestartet, wurde vor der eigentlichen Wiedergabe, die Variable mit der Kopfhörerkennung geladen und diese Kennung als Endung an den Quellpfad der Audiodatei angefügt, sodass der korrekt entzerrte Stimulus für die Wiedergabe ausgewählt wurde.

Randomisierung

Randomisierung ist ein wichtiger Faktor bei allen Hörversuchsverfahren (vgl. Ellermeier et al., 2008, S. 10). In den hier erstellten Hörtests wurde jeweils auf mehreren Ebenen randomisiert, und auch die Fachfragen wurden in zufälliger Reihenfolge präsentiert, damit konstante Effekte durch eine gleichbleibende Fragenabfolge nicht die Unabhängigkeit der Items beeinträchtigte (siehe Abschnitt 2.4.1).

In LimeSurvey sind die erstellten Fragen in Fragegruppen zusammengefasst (vgl. LimeSurvey GmbH, 18.03.2019). Sowohl einzelne Fragen, als auch ganze Fragengruppen lassen sich Randomisierungsgruppen zuweisen. Ist in der Umfrage ein Element (Frage oder Fragengruppe) einer solchen Randomisierungsgruppe an der Reihe, wird stattdessen irgendein Element aus dieser Gruppe ausgewählt. Folgt das nächste Element aus dieser Gruppe, wird der Vorgang wiederholt, wobei bereits präsentierte Elemente nicht noch einmal wiederholt werden.

Gibt es in einer Randomisierungsgruppe genau so viele Elemente, wie Aufrufe dieser Elemente in der Umfrage, führt das schlicht zu einer zufälligen Anordnung aller Elemente einer Randomisierungsgruppe. Handelt es sich bei den Elementen um Fragen, kann auch die Anzahl verfügbarer Fragen größer sein, als die Anzahl an Fragen, die tatsächlich in der Umfrage gestellt werden. Dazu werden alle verfügbaren Fragen in zwei Fragegruppen aufgeteilt. In einer aktiv geschalteten Fragengruppe sind so viele Fragen aus der Randomisierungsgruppe enthalten, wie tatsächlich angezeigt werden sollen. In einer zweiten Fragengruppe, die deaktiviert ist, also theoretisch gar

nicht angezeigt wird, sind die übrigen Fragen enthalten. Ist nun in der aktiven Fragengruppe eine Frage an der Reihe, wird aus beiden Fragengruppen zufällig eine Frage ausgewählt, da alle der gleichen Randomisierungsgruppe zugeordnet sind.

Innerhalb einer Frage kann nur die Reihenfolge, in der die Antworten einer Liste oder die Teilfragen einer Matrix präsentiert werden, randomisiert werden. Für weitere Randomisierungen, z. B. welche Audiodateien von den Schalfächen aufgerufen werden müssen eigene Javascript-Funktionen angelegt werden, deren Konfiguration jedoch nicht aufgezeichnet wird. Im zweiten Hörtest musste randomisiert werden, ob im zweiten Hörtest A und X oder B und X den gleichen Raum beinhalteten und ob der variable oder der gleichbleibende Raum derjenige sein sollte, der doppelt vorkam. Um die jeweilige Konfigurationen aufzeichnen zu können, mussten alle diese Möglichkeiten als separate Fragen angelegt und zufällig eine davon ausgewählt werden. Aus diesem Grund gab es für jedes Item dieses Hörtests vier verschiedene Versionen (zwei Optionen, welche Buchstaben den gleichen Raum abbilden und zwei Optionen, welcher Raum zweimal vorkommt), die alle einer spezifischen Randomisierungsgruppe für jedes Item zugeordnet waren. Diese Versionen unterscheiden sich lediglich durch die hinterlegten Namen der abzuspielenden Audiodateien, bei denen die Raumklänge fest den Schaltflächen A, B und X zugeordnet sind. Nach dem beschriebenen System gab es für jedes Item eine aktive Fragegruppe, in der jeweils eine dieser vier Versionen enthalten war und alle anderen Versionen aller Items wurden in einer deaktivierten Fragegruppe, deponiert'. Die 20 aktiven Fragegruppen für alle Items waren ihrerseits einer weiteren Randomisierungsgruppe zugeordnet, sodass die Items zufällig angeordnet wurden und für jedes Item immer eine von vier möglichen Konfigurationen ausgewählt wurde.

Zusätzlich sollte randomisiert werden, welche der drei Klaviersquenzen welchem Raum zugeordnet wird. Diese Zuordnung wurde in Abschnitt 3.1.1 als incidental charakterisiert, die nicht
aufgezeichnet werden musste und deshalb über eine Javascript-Funktion umgesetzt wurde. Es
wurde bereits erklärt, dass in den HTML-Playern initial unvollständige Dateinamen hinterlegt
waren, um dynamisch die Dateien mit der richtigen Kopfhörerentzerrung auszuwählen. Nach
dem gleichen Verfahren wurden im zweiten Hörtest die Klaviersequenzen ausgewählt. Bei den
hinterlegten Dateinamen in den entsprechenden Fragen fehlte nicht nur die Kopfhörerkennung
am Ende der Datei, sondern auch die Nummer der Klaviersequenz in der Stimuluskennung. Mit
jedem Aufruf einer Frage wurden nun in einem Array die Zahlen 1 bis 3 in zufälliger Reihenfolge angelegt. Wurde die erste Audiowiedergabe über eine Schaltfläche gestartet, wurde nicht
nur der Kopfhörername angehängt, sondern auch die erste Nummer in dem genannten Array,
wodurch der Stimulus mit der entsprechenden Klaviersequenz geladen wurde. Mit der zweiten
Wiedergabe wurde die zweite Nummer angefügt, mit der dritten die dritte. Da jede Schaltfälche
in diesem Test nur einmal bedient werden konnte, war sichergestellt, dass jeder Stimulus über
eine andere Schalfläche wiedergegeben wurde.

Wie beim zweiten Hörtest wurden im ersten für jedes Item zwei Versionen mit zwei unterschiedlichen Skalenausschnitten angelegt, von denen immer eine ausgewählt wurde. Auch hier gab es für jedes Item eine aktive Fragegruppe und eine deaktivierte mit den alternativen Skalenausschnitten. Die aktiven Fragegruppen bildeten wiederrum eine Randomisierungsgruppe, um die Item-Reihenfolge zu randomisieren. Bei sehr kleinen und sehr großen Referenzen konnten nicht immer zwei Versionen der Skala erstellt werden, da diese so nah an den Maximal- oder Minimalwerten lagen, dass nur eine Anordnung den festgelegten Mindestabstand der Referenz zum Skalenende ermöglichte. Die Konfigurationen der Skalen befinden sich im Anhang.

Im letzten Hörtest wurde die Option genutzt, die Teilfragen in einer Matrix-Frage, die in diesem Fall die Schaltflächen für die einzelnen Stimuli repäsentierten, in eine zufällige Reihenfolge zu

bringen. Alle Fragen gehörten derselben Randomisierungsgruppe an, sodass auch deren Reihenfolge zufällig festgelegt wurde.

Die gesamte LimeSurvey-Umfrage ist außerdem mit umfassenden Instruktionen zum Gesamtversuch und den Teilverfahren ausgestattet, um eine Durchführung ohne Unterstützung durch einen Versuchsleiter während des Versuchs zu ermöglichen. In einigen Fällen war eine Betreuung paralleler Durchführungen durch einen Versuchsleiter nur eingeschränkt oder mit Hilfe von Drittpersonen möglich.

3.4.3. Kopfhörerwiedergabe

Der Versuch konnte nicht an jeder Institution mit der gleichen Technik durchgeführt werden. Sowohl Audiointerfaces als auch Kopfhörer waren verschieden. Tabelle 3.4 zeigt eine Auflistung der Wiedergabetechnik, die jeweils verwendet wurde. Besonders bei der Durchführung am Albert-Einstein-Gymnasium kamen viele verschiedene Gerätschaften zum Einsatz, da die parallele Durchführung an mehreren Arbeitsplätzen erforderlich war. Die Kopfhörerauswahl umfasste elektrostatische und elektrodyanmische sowie offene und geschlossene Modelle (siehe Tabelle 3.4). Bei allem Modellen handelt es sich um ohrumschließende Kopfhörer.

Institution	Audiointerface	Kopfhörer
TU Berlin	Focusrite Scarlett 6i6	Sennheiser HD650
		(elektrodyn., offen)
HS Düsseldorf	HEAD acoustics PEQ V	STAX SR-303
		(elektrostat., offen)
		mit
		HEAD acoustics HPS IV
Goldsmiths	Focusrite Scarlett 6i6	AKG K702
		(elektordyn., offen)
AEG Kaarst	RME Babyface Pro	AKG K702
	Behringer U-Control UCA222	(elektordyn., offen)
	Steinberg UR22	Beyerdynamic DT770 Pro
		(eletkrodyn., geschlossen)

Tabelle 3.4.: Kopfhörer und Interfaces, die an den verschiedenen Orten für die Durchführung eingesetzt wurden. Da es sich bei den STAX SR-303 um elektrostatische Kopfhörer handelt, ist ein separater Vorverstärker notwendig.

Da es kaum einheitliche Richtlinien für die Frequenzgänge von Kopfhörern gibt (vgl. Fleischmann et al., 2012, S. 1), bestehen deutliche Differenzen in den Übertragungsverhalten der Kopfhörer. Messungen der Übertragungsfunktionen der Modelle Sennheiser HD650 und Beyerdynamic DT770 Pro sind in der FABIAN-Datenbank von Brinkmann et al. (2017) enthalten und in Abbildung 3.10 dargestellt. Es ergeben sich deutliche Abweichungen, sowohl zwischen den beiden Kopfhörern, als auch zwischen dem linken und rechten Kanal desselben Modells. Hinzu kommen individuelle Asymmetrien zwischen Pegeln der beiden Stereokanäle bei der Kopfhörerwiedergabe (vgl. Merimaa et al., 2011, S. 9) und die Tatsache, dass in diesem Versuch auch die Abhörpegel der verschiedenen Versuchsaufbauten unterschiedlich eingemessen werden mussten (siehe Abschnitt 3.4.4). Trotz der Kopfhörerentzerrung, die für eine authentische Binarualsynthese erforderlich ist (siehe Abschnitt 2.5.2), konnten deshalb klangliche Unterschiede in der Wiedergabe an den verschiednen Durchführungsorten nicht vermieden werden.

Für den Sennheiser HD650 und den Beyerdynamic DT770 Pro befinden sich auch Entzerrungs-

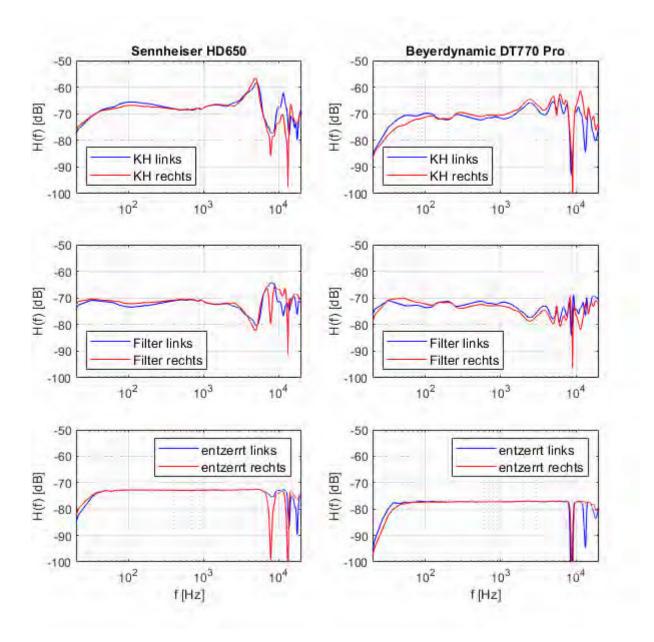


Abbildung 3.10.: Frequenzgangmessungen (KH), Entzerrungsfilter (Filter)und linearisierte Frequenzgänge (entzerrt) von zwei der verwendeten Kopfhörer aus der FABIAN-Datenbank (Brinkmann et al., 2017).

filter in der FABIAN-Datenbank, in der auch ein Entzerrungsilter für den Kopfhörer AKG K701 enthalten ist. Die technische Spezifikationen des Herstellers für dieses Modell (vgl. Harman International Industries, 2020a) sind mit denen des hier verwendeten Modells AKG K702 identisch (vgl. Harman International Industries, 2020b), weshalb das Filter auch zur Entzerrung der Stimuli für den K702 verwendet wurde. Als elektrostatische Kopfhörer benötigen die STAX SR-303 eine Versorgungsspannung der elektroakustischen Wandler durch einen zusätzlichen Vorverstärker (vgl. Dickreiter et al., 2009, S. 179). Der hier benutzte Vorverstärker HPS IV von HEAD acoutstics stellt neben der genannten Polarisationsspannung auch ein internes Entzerrungsfilter bereit, sodass keine Entzerrung bei der Erzeugung der Stimuli erforderlich war. Die mittleren und unteren Graphen in Abbildung3.10 zeigen zwei der verwendeten Entzerrungsfilter und die linearisierten Kopfhörerfrequenzgänge. Es ist zu beachten, dass diese Filter und Entzer-

rungsergebnisse auf Messungen beruhen, die zwar mit den gleichen Kopfhörermodellen, wie in diesem Versuch, jedoch nicht mit den identischen Exemplaren durchgefürt wurden. Daher kann es zu Abweichungen im Übertragungsverhalten der genutzten Modelle kommen. Wegen der individuellen Differenzen beim Kopfhörergebrauch und der Tatsache, dass mit einem Kunstkopf und nicht mit den Ohren der Probanden gemessen wurde, waren ohnehin keine idealen Entzerrungen möglich, die darüberhinaus generell nicht für beliebige Frequenzen sinnvoll realisiert werden können (siehe Abschnitt 2.5.2).

3.4.4. Einmessung der Abhörpegel

Da die Ausgangsstimuli nur als digitale Signale vorlagen und keine Dokumentation der tatsächlichen Schallleistungspegel der Quellen verfügbar war, musste die Abhörlaustärke letzten Endes willkürlich eingestellt werden. Um eine möglichst realitätsnahe Repräsentation der jeweils simulierten Aufnahmesituationen, wurde dem Beispiel von Grigoriev et al. (2016) folgend von drei fachkompetenten und routinierten Personen (zwei Professoren für Audiotechnik bzw. Musikpsychologie und einem Studenten der Audiokommunikation) zunächst für den kompletten Versuchsaufbau an der Technische Universität Berlin ein Abhörpegel festgelegt, der als realistisch und im Hinblick auf die Durchführungsdauer als nicht unangenehm laut angesehen wurde. Die Einstellung der Lautstärke wurde für die Stimuli mit Cello- und die mit Klavierstücken separat vorgenommen. Daraus ergab sich für die Cellostimuli ein gewünschter Abhörpegel, der 2,8 dB unter dem der Klavierstimuli lag. Diese Pegeldifferenz wurde bei der Stimuluserzeugung berücksichtigt, ebenso wie eine Anpassung der Lautstärken bei den jeweiligen Kopfhörerwiedergaben und -Entzerrungen (wie in den unteren beiden Graphen in Abbildung 3.10 zu erkennen, führt die Entzerrung der Kopfhörerübertragungsfunktionen zu unterschiedlichen Dämpfungsfaktoren). So konnte einmalig ein Abhörpegelfür den gesamten Versuch festgelegt werden.

Nachdem die Wiedergabelautstärke für den ersten Versuchsaufbau festgelegt war, wurde diese mit einem digital vollausgesteuerten, unentzerrten Sinuston mit einer Frequenz von 1 kHz und einer Dauer von 10 s gemessen. Der so gemessene Signalpegel galt als Referenz für die Einmessung aller weiteren Versuchsaufbauten.

Die Wiedergabesysteme konnten nicht alle mit der gleichen Messtechnik eingemessen werden. Eine genaue Aufstellung der technischen Ausrüstung für die Messungen der verschiedenen Versuchsaufbauten befindet sich in Tabelle 3.5. Die Messung des Referenzsignalpegels an der Technischen Universtität Berlin wurde mit einem künstlichen Ohr durchgeführt. Abbildung 3.11 zeigt oben links das künstliche Ohr und den Messverstärker und unten links die Messung des Referenzpegels des Abhörpegels mit dem Sennheiser HD650 als Kopfhörer und dem Focusrite Scarlett 6i6 als Interface.

Für die Messung wurde jeweils ein Kopfhörertreiber über dem künstlichen Ohr fixiert (siehe Abbidlung 3.11 links unten). Für jeden Treiber wurde fünfmal gemessen und nach jeder Messung der Kopfhörer von künstlichen Ohr entfernt und wieder neu angebracht, um durch eine anschließende Mittelung der Messergebnisse Schwankungen im Frequenzgang durch leichte Unterschiede in der Positionierung des Kopfhörers zu kompensieren (vgl. Paquier und Koehl, 2010, S. 7). Die Messungen wurden in einem Seminarraum der Akustik-Prüfstelle der Technischen Universität Berlin durchgeführt. Es ergab sich ein Signalpegel von 89,4 dBSPL auf beiden Kanälen als Referenzpegel für die Justierung der Abhörlaustärke der übrigen Veruschsaufbauten.

Die Arbeitsplätze für die Durchführung an der Hochschule Düsseldorf wurden ebenso wie die für die Durchführung am Gymnasium im nahegelegenen Kaarst mit einem Kunftkopf von HEAD acoustics im Hörversuchslabor der Hochschule duchgeführt. Für die Messungen am Goldsmiths in London stand lediglich ein Hand-Schallpegelmessgerät zur Verfügung. Die verwendete Messtechnik unterscheidet sich deutlich, sodass es gut möglich ist, dass die so unterschiedlich ein-



Abbildung 3.11.: Künstliches Ohr und Messverstärker (oben links) und Messaufbau mit dem künstlichen Ohr an der Technischen Unversität Berlin (unten links) und mit einem Kunstkopf an der Hochschule Düsseldorf (rechts).

gemessenen Wiedergabelautstärken faktisch voneinander abweichen. Die Stimuli weisen jedoch selbst starke Schwankungen in den Schallpegeln und Lautheiten auf (siehe Abschnitt 3.2.2). Darüberhinaus entstehen beim Kopfhörergebrauch ohnehin durch leichte Unterschiede in der Positionierung (vgl. Paquier und Koehl, 2010, S. 7) und individuelle interaurale Asymmetrien in der Übertragung (vgl. Merimaa et al., 2011, S. 9) von Versuchsperson zu Versuchsperson leichte Veränderungen im Klangbild. Deshalb wurden geringe Abweichungen in den eingestellten Abhörpegeln als hinnehmbar angesehen.

Da die Lautstärke jeweils an den Audiointerfaces eingestellt wurde, wurden die Probanden vor

Institution	Messverstärker	Mikrofon
TU Berlin	Brüel & Kjaer Type 2609	Brüel & Kjaer Type 4144
		(Mikrofonkapsel)
		Brüel & Kjaer Type 4152
		(künstliches Ohr)
HS Düsseldorf	HEAD acoustics BEQ II	HEAD acoustics HMS IV
AEG Kaarst		(Kunstkopf)
		HEAD acoustics HTB IV
		(Torso-Box)
Goldsmiths	Hand-Schallpegelmesser	in Schallpegelmesser eingebaut

Tabelle 3.5.: Messtechnik, mit denen die verschiedenen Versuchsaufbauten eingemessen wurden. Die Empfindlichkeit des Messaufbaus der TU Berlin wurde mit einem Kalibrator (Cirrus CR: 511F) gemessen, der Kunstkopf der HS Düsseldorf ist fest kalbriert, ebenso wie der Schallpegelmesser am Goldsmiths.

und während der Versuchsdurchführung mehrfach darauf hingewiesen, das sie die Abhörlautstärke nicht verstellen durften.

3.5. Versuchsdurchführung

Wie in den vorigen Abschnitten beschrieben, waren für den Versuchsaufbau lediglich ein Computer mit Internetzugang, ein Audiointerface und ein Kopfhörer erforderlich, für den ein Entzerrungfilter verfügbar war. Außerdem sollte eine ansatzweise einheitliche Wiedergabelauststärke gegeben sein. Da überwiegend offene Kopfhörer verwendet wurden, war es auch wichtig, eine räumliche Umgebung ohne störende Umgebungsgeräusche bereitzustellen, die die Probanden ablenken oder Klangkomponenten der Stimuli verdecken könnten.

Da pro Proband aus Zeitgründen in den ersten beiden Hörtests immer nur entweder mit variabler Nachhallzeit oder variabler Nachhallklangfarbe gestestet werden konnte, gab es zwei identisch aufgebaute Fragebögen, die sich nur durch die Stimuli in diesen beiden Hörversuchen unterschieden. Zudem wurden die Erklärungstexte und die Beispielskala im ersten Hörtest jeweils an den gestesteten Parameter angepasst (siehe Absdchnitt 3.1.2). Probanden der Gruppe A wurden mit variabler Nachhallzeit, die der Gruppe B mit variabler Nachhallklangfarbe getestet.

3.5.1. Technische Universität Berlin

Die erste Durchführung an der Technischen Universtiät Berlin fand in einem kleinen, akustisch optimierten Hörversuchslabor an einem Arbeitsplatz statt (siehe Abbildung 3.12). Über Mailinglisten wurde der Versuch bei Studenten und Mitarbeitern am Fachgebiet Audiokommunikation der Universität sowie allgemein an diesem Themengebiet interessierten Personen beworben. Audiokommunikationsstudenten konnten sich die Durchführung für eine Prüfungsleistung an-



Abbildung 3.12.: Versuchsaufbau im Hörversuchslabor der TU Berlin

rechnen lassen, außerdem wurde damit geworben, die persönlichen Ergebnisse unmittelbar nach dem Versuch einsehen zu können. Im Zuge dieser informellen Gespräche wurde zudem versucht, die Strategien, nach denen die Probanden insbesondere den dritten Hörtest beantworteten, zu erfragen und nachzuvollziehen. Die Durchführung in Berlin wurde über einen Zeitraum von drei Wochen (abzüglich der Wochenenden) und zwei Tagen angeboten. Die Probanden wurden abwechselnd der Gruppe A oder B zugewiesen.

3.5.2. Hochschule Düsseldorf

Am Institue for Sound and Vibration Engineering der Hochschule Düsseldorf war die parallele Durchführung an zwei identisch eingerichteten Arbeitsplätzen möglich. Wie in Abbildung 3.13 zu erkennen, saßen sich bei gleichzeitiger Durchführung beide Probanden in einem Hörversuchslabor gegenüber. Das Labor ist zur Optimierung der Akustik als Raum-in-Raum-Konstruktion angelegt (vgl. Hochschule Düsseldorf, 17.12.2019). Über Mailinglisten wurden alle Studenten und Mitarbeiter der Hochschule zur Teilnahme eingeladen. Auch hier wurde den Probanden angeboten, direkt nach der Durchführung die Ergebnisse einzusehen, wobei ebenfalls nach Möglichkeit die Vorgehensweisen bei der Durchführung angesprochen wurden. In Düsseldorf dauerte die Durchführung sieben Werktage. Die Probanden wurden abwechselnd der Gruppe A oder B zugewiesen. Vor der Teilnahme mussten die Versuchspersonen zudem im Rahmen der Datenschutzrichtlinien der Hochschule Düsseldorf die Zustimmung zur Erhebung persönlicher Daten geben. Diese Datenschutzerklärung befindet sich im digitalen Anhang dieser Arbeit.



Abbildung 3.13.: Versuchsaufbau im Hörversuchslabor der HS Düsseldorf

3.5.3. Albert-Einstein-Gymnasium Kaarst

Die Durchführung am Gymnasium in Kaarst stellte technisch die größten Anforderungen, da die zeitliche Planung an den 90-Minuten-Rhythmus (doppelte Unterrichtsstunden) angepasst werden

musste. An einem Tag wurde der Versuch mit zwei Kursen durchgeführt. In einem Laborraum ohne besondere akustische Optimierungen außer einer für Klassenräume typsiche Akustik-Decke wurden neun Arbeitsplätze eingerichtet. Wie in Abbildung 3.14 dargestellt, saßen die meisten Probanden an Einzeltischen mit einem gewissen Abstand zueinander. Drei weitere Arbeitsplätze befanden sich an der Fensterfront des Raums - auch hier wurde auf einen Abstand von mindestens zwei Metern zwischen den Plätzen geachtet. Wie aus Tabelle 3.4 hervorgeht, mussten für die Ausstattung der vielen Arbeitsplätze zwei verschiedene Kopfhörermodelle und drei verschiedene Audiointerfaces verwendet werden.

Für die Teilnahme ausgesucht wurden zwei Physikkurse der elften (17 Personen, ca. 16-17 Jahre) bzw. zwölften Jahrgangsstufe (25 Personen, ca. 17-18 Jahre). Der erste Kurs wurde in zwei Gruppen aufgeteilt, die der Reihe nach den Versuch durchführten. Beim zweiten Kurs gab es Schwierigkeiten mit dem LimeSurvey-Server, sodass nur wenige Probanden den Versuch tatsächlich abschließen konnten. Neben den Schülern nahmen auch zwei Lehrer am Versuch teil. In jeder Durchführungsrunde wurde abwechselnd auf fünf Rechnern der Versuch der Gruppe A oder B eingerichtet und in der nächsten Runde entsprechend viermal A und fünfmal B. Wegen der Problematiken beim zweiten Kurs kam es zu einer unregelmäßigen Verteilung der Versuche A und B auf die Probanden. Die Eltern der Schüler wurden vorab über die Versuchsdurchführung informiert und die Teilnahme den Schülern ausdrücklich freigestellt (das Informationsschreiben befindet sich im digitalen Anhang).



Abbildung 3.14.: Versuchsaufbau im Physiklabor des AEG Kaarst

3.5.4. Goldsmiths, Univeristy of London

Am Goldsmiths, University of London wurde die Versuchsdruchführung von einer Drittperson betreut und fand ebenfalls in einem Hörversuchslabor statt (siehe Abbildung 3.5.4). Daher war keine direkte Einsicht der Ergebnisse mit den Probanden möglich. Eingeladen wurden Studenten

der Universität. Studenten der Musikpsychologie konnten sich die Dauer der Teilnahme für eine Prüfungsleistung anrechnen lassen. Die Durchführung dauerte drei Wochen, abzüglich der Wochenenden. Im Rahmen der Ethikrichtlinien der Universität mussten die Probanden in London zudem einer speziellen Einverständniserklärung zur Teilnahme zustimmen und ihre Volljährigkeit bestätigen, die in der LimeSurvey-Umfrage implementiert und nur Probanden am Goldsmiths präsentiert wurde (die Einverständniserklärung befindet sich als Teil der englischen Version des Versuchs im Anhang). Die Versuchsdurchführung wurde von der Ethikkomission des Goldsmiths, University of London bewilligt.



Abbildung 3.15.: Versuchsaufbau im Hörversuchslabor des Goldsmiths, University of London

4. Versuchsergebnisse und Auswertung

4.1. Korrekturen und Anpassungen

Während der ersten Durchführungsperiode an der Technischen Universität, zeigte sich, dass an drei Stellen Korrekturen erforderlich waren. Da während der laufenden Durchführung keine zeitaufwändigen Wiederholungen von Raumsimulationen möglich waren, wurden die Probleme durch minimale Änderungen zu beheben versucht. Dabei waren immer nur einzelne Items des jeweiligen Hörtests betroffen, sodass in der Auswertung bei diesen Items weniger Antworten vorlagen.

Raumakustische Wahrnehmungsfähigkeiten - variable Nachhallzeit

In den Antworten des ersten Hörtests mit variabler Nachhallzeit häuften sich bei der $100\,\%$ -Referenz zu Beginn des Versuchs auffallend große Abweichungen. In Abschnitt 3.2.2 wurde bereits die ungweöhnlich hohe Lautheit dieses Stimulus festgestellt.

Abbidlung 4.1 zeigt die binauralen Impulsantworten (in Terzbanddarstellung) und Nachhallzei-

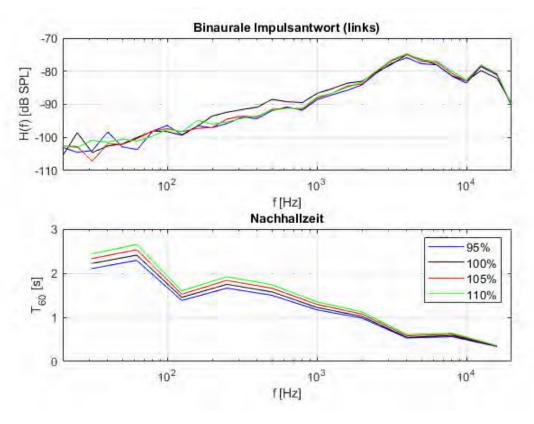


Abbildung 4.1.: Impulsantworten (linker Kanal) und Nachhallzeiten der 100 %-Referenz und der "benachbarten'Räume im ersten Hörtest, variable Nachhallzeit.

ten (in Oktavbanddarstellung) des $100\,\%$ -Raumes und der Räume mit den nächst kleineren- und größeren Nachhallzeiten. Die binauralen Impulsatworten enthalten neben der Übertragungsfunktion

des Raumes auch die Außenohrübertragungsfunktion des Kunstkopfes. Deren Effekt ist jedoch bei allen dargestellten Impulsantworten gleich. Auch wenn die oktavbandgemittelte Nachhallzeit dem vorgegebenen Verlauf folgt, kommt es der 100 %-Referenz zwischen 125 Hz und 2 kHz zu einer erhöhten Verstärkung durch prägnante Reflexionen, weshalb viele Probanden diesen Stimulus auf der Skala im Bereich wesentlich höherer Nachhallzeiten vermuteten. Eine erneute Simulation dieser Raumimpulsantwort lieferte das gleiche Resultat. Aus diesem Grund wurde für alle weiteren Durchführungen stattdessen eine Nachhallzeit von 110 % der initialen Nachhallkurve als Referenz verwendet. Die resultierende Skalenauflösung ist ebenfalls in Tabelle 3.1 dargestellt.

Die 110%-Referenz wurde als Ersatz gewählt, da sich die Nachhallzeiten zur alten Referenzen nur geringfügig unterscheiden und von den gleichmäßigen Abständen der verschiedenen Item-Referenzen kaum abgewichen wurde. Außerdem gab es auf der Skala zu beiden Seiten dieses Stimulus mindestens eine Stufe, die sich wie erwartet von der Referenz unterschied. Wäre der 105%-Raum als neue Referenz gewählt worden, läge direkt daneben die problematische 100%-Referenz und die Bestimmung der exakten Position der Referenz wäre daher einfacher. Durch diese Umstellung mussten für die ersten 19 Probanden die Antworten des dritten Items dieses Hörtests verworfen werden.

Wiedererkennen des Raumes - variable Nachhallklangfarbe

Beim zweiten Hörtest mit variabler Nachhallklangfarbe wurde nach 19 Teilnahmen ebenfalls eine Korrektur vorgenommen. Hier zeigte sich ein ähnlicher Effekt bei der frequenzunabhähngigen Nachhallzeit, da bei der korrespondierenden binauralen Impulsantwort eine ungewöhnliche Anhebung zwischen 200 kHz und 2 kHz beobachtet wurde, die in der oktavbandgemittelten Nachhallzeit nicht erkannt werden konnte (siehe Abbildung 4.2).

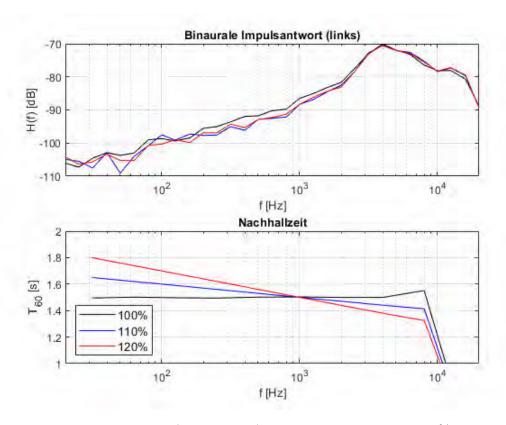


Abbildung 4.2.: Impulsantworten (linker Kanal) und Nachhallzeiten der 100 %-REferenz und der "benachbarten'Räume im zweiten Hörtest, variable Nachhallklangfarbe.

Das führte dazu, dass Items mit geringer Koloration der Nachhallzeit, die eigentlich nur einen geringen klanglichen Unterschied erzeugen sollten, sicherer von der Referenz unterschieden werden konnten als Räume, bei denen eine stärker Modulation der Klangfarbe deutlichere Differenzen bewirken sollte. Somit nahm die Anzahl richtiger Antworten nur bis zu einem gewissen Grad mit den Unterschieden zwischen den Nachhallzeiten ab, da bei den geringsten Modulationsstufen die Referenz unbeabsichtigt basslastiger klang und so leicht von den anderen Räumen unterschieden werden konnte. Abbildung 4.3 zeigt die Häufigkeiten richtiger Antworten im Test für die ersten 19 Teilnehmer, die zunächst wie erwartet abnehmen und dann zu vermeintlich geringeren Differenzen wieder ansteigen.

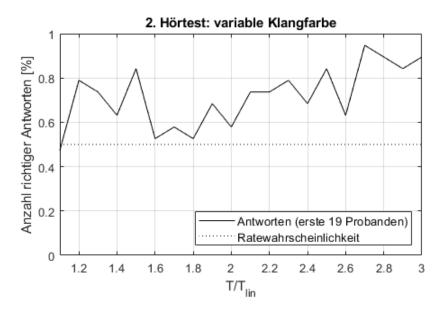


Abbildung 4.3.: Antwortverhalten der Versuchspersonen im zweiten Hörtest bei variabler Klangfarbe, bevor der Test korrigiert wurde. Anstatt mit der Annäherung der Nachhallzeiten abzufallen, steigt die Anzahl richtiger Antworten durch unexakte Simulationsergebnisse beim Referenzraum ab einem Modulationsfaktor von 1,5 wieder an.

Um dieses Problem schnell zu lösen, wurden für die nachfolgenden Durchführungen die Stimuli mit dem "nächsten"Raum mit einer Modulation von 1,1 als neue Referenz eingesetzt. Dazu war lediglich erforderlich, eine zusätzliche Modulationsstufe hinzuzufügen, da das vormals zwanzigste Item zum neunzehnten wurde usw. Die Schrittweite der Modulation, bezogen auf die neue Referenz, änderte sich nur minimal von 10,% auf 9,09 %, weshalb in der Auswertung die Testergebnisse mit beiden Referenzen zusammengefasst wurden. Bei den ersten 19 Teilnahmen wurden jedoch die 5 Items mit den geringsten Modulationsdifferenzen ausgeschlossen.

Raumvolumen vs. Nachhallzeit

Im dritten Hörtest zeichnete sich schnell ab, dass auch bei den geringsten Volumendifferenzen die klanglichen Auswirkungen von einigen Probanden offenkundig erkannt und bezogen und die klanglichen Differenzen in die richtige Reihenfolge gebracht werden konnten. Damit erreichte die maximale, angenommene Itemschwierigkeit nicht wie geplant den Bereich der Wahrnehmungsschwelle (siehe Abschnitt 3.1.2). Aus diesem Grund wurde dieser Hörtest nach Ende der ersten Durchführung in Berlin um zwei weitere Items auf eine Gesamtzahl von 14 ergänzt: eines in dem die Volumendifferenzen - nach wie vor auf einer logarithmischen Skala - genau zwischen denen des vormals ersten und zweiten Items lagen und eines, in dem die Volumenverhältnisse

des vormals ersten Items noch einmal halbiert wurden.

Konkret wurden die Gleichungen 3.1 und 3.2 so umgestaltet, dass für x nicht wie bisher die Zahlen zwischen von 1 bis 12, sondern zusätzlich 0,5 und 1,5 eingesetzt wurden. Da hierfür zwei Fragen in LimeSurvey ergänzt werden mussten, wofür ein Abschalten der Umfragen erfoderlich ist, wurde mit der Korrektur bis nach Ende der Durchführung in Berlin gewartet, zumal sich auch erst dann der eindeutige abzeichnete, dass die bisherigen Items nicht schwierig genug waren.

In den nachfolgenden Durchführungen schien bei den neuen minimalen Differenzen keine zuverlässigen Unterscheidungen mehr möglich.

4.2. Stichprobe

Insgesamt nahmen 100 Personen am Versuch teil: 44 Probanden in Berlin, 29 an der Hochschule Düsseldorf, 22 am AEG in Kaarst und 5 am Goldsmiths, Univeristy of London. Die Teilnehmer waren zwischen 16 und 59 Jahre alt (Mittelwert: 25,80, Standardabweichung: 7,92). Als Geschlecht gaben 71 Personen männlich, 28 weiblich und eine Person divers an. Die durchschnittliche Durchführungsdauer betrug 32 Minuten (Standardabweichung: 7,78), sodass die Maximaldauer von 40 Minuten meist eingehalten wurde. Die längste Durchführungsdauer lag bei 61 Minuten, über die kürzeste lässt sich keine präzise Aussage treffen, dadurch die technischen Probleme bei der Durchführung am AEG Kaarst einige Probanden den Versuch abbrechen und neu beginnen mussten. In der zweiten Durchführung ging dann die Beantwortung bekannter Fragen tendenziell schneller, sodass die aufgezeichnete Durchführungsdauer verzerrt wurde. 53 Probanden absolvierten den A-Versuch mit variabler Nachhallzeit und 47 den B-Versuch mit variabler Nachhallklangfarbe.

4.2.1. Erfahrung mit Raumakustik

Die Erhebung der Erfahrung mit Raumakustik und der drei Faktoren des Gold-MSI sind nicht nur für die Auswertung der Hörtests relevant, sie bieten auch eine Möglichkeit, die Heterogenität der Probanden bezüglich ihres fachlichen Hintergrundes und Interesses abzuschätzen.

Abgefragt wurde die Häufigkeit der Beschäftigung mit Raumakustik im Rahmen akademischer, künstlerischer und musikalischer Tätigkeiten sowie beruflicher oder privater Musikaufnahmen. Abbildung 4.4 zeigt links die Häufigkeiten der fünf möglichen Antworten für die drei genannten Komponenten. Aus der Grafik geht hervor, dass über die Hälfte der Probanden mit dem Thema auf akademischer Ebene noch keinen Kontakt hatte. Etwa 45 % hatten sich nach eigenen Angaben mit Raumakustik noch nie im Rahmen von Aufnahmetätigkeiten und fast 40 % im Zusammenhang mit musikalischen Aktivitäten auseinandergesetzt.

Dass etwa die Hälfte der Versuchspersonen damit als Laien eingestuft werden können, spricht für die Respräsentativität der Stichprobe. Tatsächlich ist diese Menge, bezogen auf die Gesamtbevölkerung vermutlich noch zu niedrig. Für die Auswertung möglicher Korrelationen mit den Ergebnissen der Hörtests ist es jedoch von Vorteil, alle Expertisegrade in der Stichprobe mit einer gewissen Prozentzahl vertreten zu finden. Hier ist anzumerken, dass in allen Kategorien nur selten angegeben wurde 'sehr oft'mit Raumakustik in Kontakt zu kommen. Das trägt ebenfalls zur Respräsentativität für eine Gesellschaft bei, in der es tendenziell nur wenige Experten auf einem Fachgebiet gibt.

4.2.2. Gold-MSI

Auch außerhalb des raumakustischen Kontexts ist der Bezug zu Musik in der Stichprobe breit gefächert. Abbildung 4.4 zeigt rechts die aus den einzelnen Items berechneten Ausprägungen

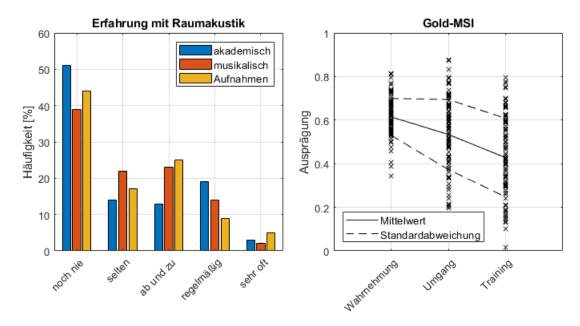


Abbildung 4.4.: Antworthäufigkeiten zur Erfahrung mit Raumakustik (links) und Gold-MSI-Faktorwerte (rechts). Die Gold-MSI-Werte sind normiert auf den Bereich zwischen dem maximal (1) und dem minimal möglichen, gewichteten Wert (0).

der drei Gold-MSI-Faktoren "Musikalische Wahrnehmungsfähigkeiten", "Aktiver Umgang mit Musik "und "Musikalische Ausbildung". Anders als die Datennormen, die von Müllensiefen et al. (2014) bestimmt wurden, berücksichtigt die hier durchgeführte Normierung die faktorspezifische Gewichtung der einzelnen Items (vgl. S. 9). Der Maximalwert der Skalierung entspricht also der gewichteten Summe aller Items mit der höchsten möglichen Antwort, der Wert 0 entsprechend der geringsten. Auch die Gold-MSI-Faktoren können als Indikator für die Diversität der Versuchspersonen ausgewertet werden.

Der Gold-MSI basiert auf Selbsteinschätzung und interessanterweise werden die musikalischen Wahrnehmungsfähigkeiten (siehe Abschnitt 3.3.2) im Mittel am höchsten eingestuft. Die anderen beiden Faktoren sind relativ gleichmäßig über den möglichen Bereich verteilt, wobei der aktive Umgang mit Musik außerhalb vom Musizieren ausgeprägter scheint.

Die breite Streuung der Erfahrung mit Raumakustik und der Ausprägung der Gold-MSI-Faktoren deutet auf eine heterogene Stichprobe hin und bietet somit eine günstige Ausgangssituation, um mögliche Zusammenhänge dieser Eigenschaften mit den Ergebnissen der Hörtests zu untersuchen.

Vergleich der Institutionen

Zusätzlich wurde betrachtet, wie die Stichproben in Abhängigkeit der jeweiligen Institution klassifziert werden können. Abbildung 4.5 zeigt die mittleren Erfahrungs- und Gold-MSI-Faktorwerte der verschiedenen Durchführungsorte.

Wenig überraschend fallen die Angaben zur Erfahrung mit Raumakustik gerade bei der Schule (AEG) deutlich niedriger aus. Aber auch bei den Hochschulen gibt es merkliche Unterschiede. Bei den Gold-MSI-Faktoren ist der erste Faktor über alle Institutionen eher konstant, deutliche Schwankungen gibt es beim Faktor "aktiver Umgang mit Musik". Durch die Heterogenität der Teilstichproben untereinander ist es auch schwerlich möglich, Auswirkungen der verschiedenen Wiedergabesysteme und räumlichen Situationen auf die Ergebnisse der Hörtests zu prüfen, da diese Unterschiede ebenso von unterschiedlicher Erfahrung und Musikalität herrühren können.

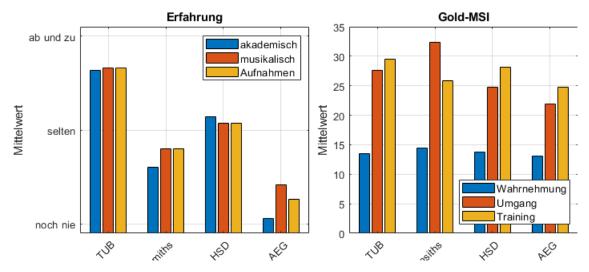


Abbildung 4.5.: Mittlere Angaben für Erfahrungswerte und Gold-MSI-Faktoren in Abhähngigkeit der Institutionen.

4.3. Kalibrierung und Berechnung der Fähigkeitsgrade

Für die Auswertung wurden zunächst die prädizierten Itemschwierigkeiten überprüft. Das Vorgehen ist an der Arbeit von Harrison und Müllensiefen (2018) orientiert, bei der ebenfalls ein nach probabilistischer Testtheorie entworfener Hörversuch entwickelt, kalibriert und ausgewertet wurde (vgl. S. 1).

4.3.1. Vorauswertung der Antworten

Schon während der Versuchsdurchführung wurden die Erfolgsquoten für jedes Item kontrolliert, um festzustellen, ob es Auffälligkeiten bei der Beantwortung einzelner Items gab, oder ob wie im Fall des dritten Hörtests die Items nicht fordernd genug waren. Vor der Schätzung der Itemparameter wurden die finalen Anzahlen richtiger Antworten für jedes Item erneut betrachtet, um Items mit Erfolgsquoten unterhalb der Ratewahrscheinlichkeit zu erkennen und bei der Auswertung nicht zu berücksichtigen (vgl. Harrison und Müllensiefen, 2018, S. 7).

Der erste Hörtest unterscheidet sich von den übrigen, da keine binäre Antwort, sondern eine metrische Abweichung von der korrekten Antwort. Statt Anzahlen richtiger Antworten sind daher in Abbildung 4.6 die Streuung der Antworten um die tatsächliche Referenz (obere Graphen) und die Häufigkeiten der verschiedenen Antworten (untere Graphen) dargestellt.

Die gewählten Antworten sind in prozentuale Abweichungen der Nachhallzeit bzw. Klangfarbe vom Referenzraum umgerechnet. Mittelwerte und Standardabweichungen sind eingezeichnet. Aus diesen geht, ebenso wie aus den unteren Abbildungen, hervor, dass die Abweichungen bei jedem Item näherungsweise normalverteilt sind.

Bei der Auswertung der Streuung der Antworten ist zu beachten, dass die (relative) Auflösung der Skala sich jeweils mit der Referenz ändert, sodass eine Abweichung auf der Skala um eine Stufe bei der 50 %-Referenz eine wesentlich größere prozentuale Abweichung von der Zielnachhallzeit bedeutet als bei der 150 %-Referenz. Dementsprechend verteilen sich die Antworten bei den niedrigeren Referenzen auf weniger Stufen als bei den hohen. Dass bei der, erst im Verlauf des Versuchs gewählten 110 %-Referenz die Abweichungen kaum zu geringeren Nachhallzeiten streuen, ist der ungewöhnlichen Nachhallcharakteristik bei 100 % geschuldet.

Sonst bleibt der Mittelwert bei allen Items etwa gleich und die Standardabweichung wird zu Referenzen mit höheren Nachhallzeiten etwas geringer. Diese Entwicklung lässt sich auch im Fall

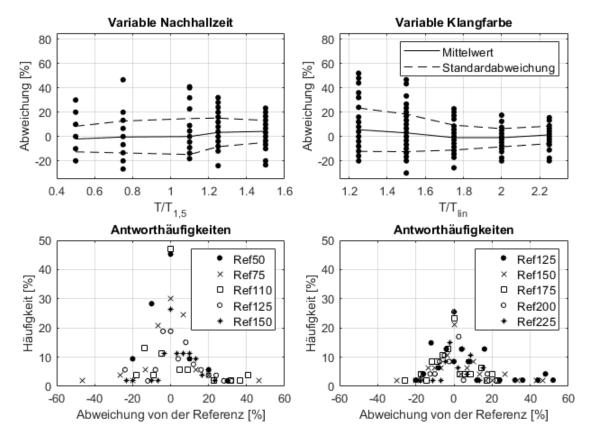


Abbildung 4.6.: Bandbreite (oben) und Häufigkeiten (unten) der angegeben Abweichungen von der jeweiligen Referenz. Die Schrittweite in Abhängigkeit von der Refernz beeinflusst, welche prozentuale Nachhallzeit- bzw. Klangfarbendifferenz die Abweichung um eine Stufe bedeutet. Mittelwerte und Standardabweichungen sind eingezeichnet.

der variablen Nachhallzeit beobachten und dürfte in beiden Fällen teils der zunehmend feiner aufgelösten Skala geschuldet sein, obschon bei der Variation der Klangfarbe auch darüberhinaus der Test bei basslastigeren Referenzen einfacher zu sein schien.

Beim zweiten Hörtest wurden vorab die Erfolgsquoten ausgewertet und auch untersucht, ob es einen Einfluss auf die Anzahl an korrekten Antworten gab, wenn im Test jeweils zwei Stimuli aus dem gleichbleibenden Raum oder dem mit variabler Nachhallzeit oder -Klangfarbe stammen. Die Frage war also, ob der Test leichter ist, wenn der Raum mit dem stärker ausgeprägten Merkmal wiedererkannt werden musste. Dazu wurde mit einem Student's t-Test geprüft, ob die Verteilung der Räume (zwei mal Referenz oder zwei mal der variable Raum) einen Einfluss auf die Anzahl richtiger Beantwortungen des Items hat. Sowohl bei der Testung mit variabler Nachhallzeit, als auch bei der Testung mit variabler Klangfarbe konnte kein signifikanter Einfluss festgestellt werden (p=0,389 im Fall der variablen Nachhallzeit, p=0,118 im Fall der variablen Nachhallklangfarbe).

Abbildung 4.7 zeigt die Anzahlen richtiger Antworten pro Item, die sich über die Variation der Nachhallzeit und Nachhallklangfarbe verteilen. Zudem ist jeweils die Ratewahrscheinlichkeit von $50\,\%$ markiert. Es zeigt sich, dass bei variabler Nachhallzeit ein Item und bei variabler Klangfarbe drei Items unterhalb dieser Schwelle lagen und zudem jeweils ein Items mit einer Quote von gerundet $51\,\%$ so nah daran lag, dass sie ebenfalls nicht in den Kalibrierungsprozess

einbezogen wurden. Darüberhinaus deutet sich in beiden Fällen der erwartete Trend ab, dass die Anzahl richtiger Antworten mit der Annäherung der zu vergleichenden Raumklänge sinkt. Bei variabler Nachhallzeit beginnt der Abfall erst unterhalb von Differenzfaktoren von 2,5, wobei das Item mit doppelter Nachhallzeit - hier handelt es sich um den gleichen Raum wie bei der $100\,\%$ Referenz im ersten Hörtest - dieses mal mit einer ungewöhnlich hohen Erfolgsrate auffällt. Dieses Item wurde wegen der problematischen Raumsimulation ebenfalls nicht in die Auswertung mit aufgenommen.

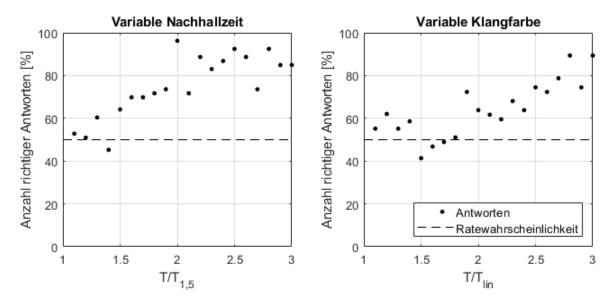


Abbildung 4.7.: Relative Anzahlen richtiger Antworten für im zweiten Hörtest. Die Ratewahrscheinlichkeit ist als gestrichelte Linie eingezeichnet.

Für die Auswertung des dritten Hörtests muss beachtet werden, dass Beantwortung des Tests ßen in zwei Stufen stattfand. Zunächst musste die Versuchsperson erkennen, dass die drei Stimuli sich bezüglich klanglicher Merkmale derart unterscheiden, dass diese Merkmale bei allen dreien unterschiedlich stark ausgeprägt ist. In informellen Gesprächen mit den Probanden nach der Durchführung bestätigte sich vielfach, was auch in Abbildung 3.9 bereits nahegelegt wird: dass die Lautheit eine wesentliche Einflussgröße darstellte. Die Versuchsperson sortierte die Stimuli also zunächst z. B. nach der Lautheit. In einem zweiten Schritt musste dann entschieden werden, wie diese Reihenfolge die Größenverhältnisse der Räume abbildet. Hier gibt es zwei Möglichkeiten: Raumgröße und Lautheit sind entweder proportional oder antiproportional. Ein Item wird erst dann richtig beantwortet, wenn auch die richtige Zuordnung stattfindet, nämlich, dass die Lautheit mit steigender Raumgröße abnimmt (vgl. Hameed et al., 2004, S. 5).

Die in Abschnitt 3.1.2 definierte Itemschwierigkeit hängt jedoch "nur'davon ab, wie nah die hörbaren Differenzen der Stimuli der Unterschiedschwelle kommen und hat damit vermutlich auf diesen zweiten Schritt der Beantwortung keinen Effekt. Im Umkehrschluss scheint es widersinnig, bei der Schätzung der Itemschwierigkeiten die Tatsache, dass ein Item falsch beantwortet wurde, weil Versuchspersonen die korrekt gehörten und sortierten Merkmalsdifferenzen nur falsch interpretiert haben, als Indiz für eine hohe Itemschwierigkeit anzusehen. Daher erfolgte auch die Auswertung in zwei Stufen.

Für die Berechnung der Itemschwierigkeit und der korrespondierenden Fähigkeitsausprägung wurden sowohl korrekt angegebene Reihenfolgen miteinbezogen als auch solche, bei denen der kleinste und größte Raum vertauscht wurden, da hier offenkundig die Unterschiede korrekt erkannt, aber falsch zugeordnet wurden. Zudem wird für jeden Probanden anhand des Verhältnisses

korrekter zu umgedrehten Reihenfolgen (beschränkt auf eine Vertauschung von groß und klein) eine dichotome Variable erstellt, die bei einer Mehrzahl an korrekten Antworten 1 und bei einer Überzahl an verdrehten Reihenfolgen 0 annimmt. Diese Variable bildet ab, ob die Person den Zusammenhang zwischen Raumgröße und klanglichen Eigenschaften verstanden hat oder nicht. Um Antworten auszuschließen, bei denen generell kein System der Sortierung erkennbar ist, wurde eine Mindestanzahl von vier korrekten Antworten als Bedingung für die Wahl des Variablenwerts festgelegt. 48 Probanden konnten demnach den Zusammenhang zwischen Klangveränderung und Raumvolumen korrekt herstellen.

Einige wenige Probanden waren der Auffassung, dass nicht bei jedem Item tatsächlich drei unterschiedliche Stimuli präsentiert wurden und ordneten deshalb alle oder zwei Klänge der selben Raumgröße zu. Dieses Antwortverhalten verletzt im Grunde das Testparadigma, dass die Räume in eine zusammenhängende Reihenfolge gebracht werden müssen, bestätigt nur, was ohnehin bei den schwersten Items zu erreichen versucht wurde, nämlich, dass die Differenzen im Bereich der Wahrnehmungsschwellen liegen. Da solche Antworten zudem in beiden erläuterten Schritten der Auswertung ebenso als falsch gewertet werden, wie eine durch Raten gewählte unkorrekte Reihenfolge verschiedener Raumgrößen, wurden diese Antworten nicht in besonderer Weise berücksichtigt.

Abbildung 4.8 zeigt sowohl die relativen Anteile korrekter Antworten (schwarz), als auch den Anteil verdrehter Reihenfolgen (rot) und die Summer beider Antwortkategorien (blau). Auch wenn zwei als Antworten 'korrekt'gelten, bleibt die Ratewahrscheinlichkeit unverändert, da je nach Deutungsansatz der Versuchsperson nach wie vor immer nur eine Kombination korrekt ist. So ergab sich, dass bei keinem Item die Anzahl korrekter Sortierungen (unabhängig vom Bezug zur Raumgröße) unterhalb dieser Wahrscheinlichkeit lag und alle Items im Kalbirierungsprozess berücksichtigt werden konnten.

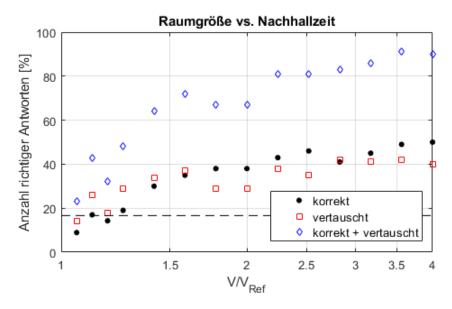


Abbildung 4.8.: Relative Anzahlen richtiger Antworten für im dritten Hörtest. Die Ratewahrscheinlichkeit ist als gestrichelte Linie eingezeichnet. Angegeben werden sowohl die Anzahlen korrekter Reihenfolgen (schwarze Punkte), exakt vertauschter Reihenfolgen (rote Quadrate), sowie die Summer beider Kategorien (blaue Rauten).

In Abschnitt 3.3.1 wurde bereits angeführt, dass die Festlegung der Itemschwierigkeit über den Inhalt einer Fachfrage nicht ganz indiskutabel ist. Deswegen war hier eine Auswertung der Er-

folgsquoten besonders wichtig, um festzustellen, ob die Ergebnisse überhaupt für eine Schätzung der Itemschwierigkeit in Frage kamen.

Abbildung 4.9 zeigt die Anzahl richtiger Antworten für die acht Fragen. Tatsächlich kommt der Verlauf der Erfolgsquoten einem linearen Abfall relativ nah und bewegt sich zwischen etwa 90% und 35%, etwas oberhalb der Ratewahrscheinlichkeit. Daher musste auch hier kein Item verworfen werden. Eine Frage wurde wesentlich öfter richtig beantwortet als die anderen. Hier handelt es sich um die schon erwähnte Frage zum Effekt eines geöffneten Fensters auf die Nachhallzeit, den sich offenbar auch viele Probanden ohne spezifische Kenntnisse über Akustik erschließen konnten.

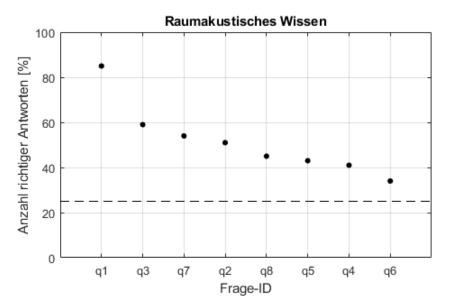


Abbildung 4.9.: Relative Anzahlen richtiger Antworten bei den fachlichen Fragen zur Raumakustik. Die Ratewahrscheinlichkeit ist als gestrichelte Linie eingezeichnet. Die konkreten Fragen und Antworten finden sich im Anhang.

4.3.2. Kalibirierung der Items

Da im ersten Hörtest metrische Antworten vorliegen, wurden die Fähigkeitsgrade über eine Faktoranalyse bestimmt, die ebenso die Ausprägung einer latenten Variable misst, wie ein Modell der probabilistischen Testtheorie (vgl. Fabrigar et al., 1999, S. 275). Aus diesem Grund werden die folgenden Modellanalysen nur für die beiden anderen Hörtests und die Fachfragen zur Raumakustik betrachtet und die Faktoranalyse des ersten Hörtests im Anschluss separat behandelt.

Die Kalibrierung der Items erfolgte in zwei Schritten. Zunächst wurde jeweils ein dreiparametrisches Rasch-Modell berechnet (siehe Abschnitt 2.4.3), um die Hypothesen zu prüfen, dass erstens die Itemschwierigkeit wie erwartet durch die jeweilige Variation der Nachhallcharakteristik bzw. die Art der Frage zur Raumakustik prädiziert wird. Zweitens wurde getestet, ob die Antworten auf die Items als voneinander unabhähngig betrachtet werden können. Für diese Modellberechnung muss eine definierte Anzahl an Items einer definierten Anzahl an Probanden vorgelegt werden (vgl. Baker, 2001, S. 130). Da dafür vollständige Datensätze jeder Person erforderlich sind, wurden diese Analysen nicht immer mit allen Probanden bzw. Items durchgeführt. Konkret wurden beim zweiten Hörtest mit variabler Klangfarbe nur die vollständigen Datensätze beachtet, also nur 29 von 47 Probanden, und beim dritten Hörtest zwar die Antworten aller Personen, nicht aber die nachträglich hinzugefügten Items.

In einem zweiten Schritt wurden generalisierte lineare gemischte Modelle berechnet, aus denen dann die mit dem gesamten Datensatz die finalen Itemparameter und Fähigkeitsgrade berechnet wurden, die allerdings nicht die Möglichkeit bieten, die genannten Gütekriterien zu prüfen (vgl. Harrison und Müllensiefen, 2018, S.6). Die Analysen wurden in R (The R Foundation, 1993) durchgeführt und sind stark an der Kalibrierung des CA-BAT-Tests von Harrison und Müllensiefen (2018) und ihrem öffentlich zu Verfügung gestellten R-Code orientiert, der im Paper verlinkt ist (vgl. S. 17).

Prädiktion der Itemschwierigkeit und Itemunabhängigkeit

Im zweiten und dritten Hörtest wurde die Prädiktion der Itemschwierigkeiten durch die jeweils variierten raumakustischen Parameter mit dem logistischen Modell aus Gleichung 2.6 geprüft. Das Modell wurde in R als dreiparametrisches Modell nach Birnbaum mit dem 1tm-Package (Rizopoulos, 2006) nach der Maximum-Likelihood-Methode geschätzt. Dabei handelt es sich um ein Raschmodell, bei dem die Itemdiskrimination als konstant angenommen wird (siehe Abschnitt 2.4.1).

Dass die Anzahlen richtiger Antworten mit der Ausprägung eines Unterschieds zweier Raumklänge anstiegen, deutet sich bereits in den Abbildungen 4.7 und 4.8 an. Um dies zu verifizieren, wurde für jedes Item die Modellgüte geprüft, indem die tatsächliche lokale Antwortverteilung mit Monte-Carlo-Antwortschätzungen auf Basis der berechneten Modellparameter verglichen wurde (vgl. Harrison und Müllensiefen, 2018, S.7).

Bei einer Monte-Carlo-Simulation wird eine vermutete Verteilung einer Prädiktorvariable genutzt, um zufällig Werte aus dieser Verteilung generieren und das resultierende Ergebnis der abhängigen Variablen durch das spezifizierte statistische Modell zu simulieren (vgl. Raychaudhuri, 2008, S. 91). Wird dabei die Verteilung der tatsächlich vorliegenden Daten angenähert, kann eine Testwiederholung simuliert und auf merkliche Unterschiede in den Ergebnissen getestet werden (vgl. ?, S. ,92). Die Implementierung dieser und der beiden nachfolgenden Analysen wurde im genannten Code von Harrison und Müllensiefen (2018) in R realisiert und für diese Auswertung übernommen. Bei nur zwei Items im zweiten Hörtest mit variabler Nachhallzeit und einem Item im dritten Hörtest kam es zu signifikanten Abweichungen zwischen Schätzung und tatsächlichen Daten (eine vollständige Aufstellung der Teststatistik befindet sich im Anhang, als Signifikanzniveau galt 0.05).

Zusätzlich wurde die Eindimensionalität der Items untersucht, die ausdrückt, dass die Items tatsächlich ein latentes Merkmal (in dem Fall eine Fähigkeit) messen und nicht mehrere (vgl. Moosbrugger und Kelava, 2012, S. 229f). Dazu wurde die Modifizerte Parallelanaylse von Drasgow und Lissak (1983) verwendet. Dabei wird die tetrachorische Korrelationsmatrix der Antworten mit der eines simulierten Datensatzes, bei dem durch die Art der Simulation tatsächliche Eindimensionalität garantiert wird, verglichen (vgl. Drasgow und Lissak, 1983, S. 365f). Weder für die beiden Versionen des zweiten Hörtests, noch für den dritten konnten signifikante Abweichungen festgestellt werden (p=0,874 für variable Nachhallzeit, p=0,892 für variable Nachhallklangfarbe, p=0,840 für variables Raumvolumen).

Ebenfalls dem Beispiel von Harrison und Müllensiefen (2018) folgend wurde in einem letzten Schritt die Unabhängigkeit der Items getestet. Dabei werden die Zweier- und Dreierkombinationen aus den einzelnen Items gebildet und die beobachteten Wahrscheinlichkeiten, jeweils alle Items eines Pärchens oder Trios korrekt zu beantworten, mit den Wahrscheinlichkeiten verglichen, die sich Antworten ergeben, die mit den berechneten Modellparametern simuliert wurden (vgl. Harrison und Müllensiefen, 2018, S. 7). Aus der Differenz de Wahrscheinlichkeiten wird nach Bartholomew (1998) ein Residuum gebildet, das einen Wert von 4 nicht überschreiten sollte (vgl. S. 7). Die Menge an Zweier- und Dreierkombinationen, die dieses Kriterium nicht erfüllen,

wird als ein Indiz auf die Unabhängigkeit der Items betrachtet.

Im zweiten Hörtest mit variabler Nachhallzeit (Laborgruppe A) wurde bei $0\,\%$ der Zweier- und nur $0.98\,\%$ der Dreierkombinationen ein Residuum oberhalb der genannten Grenze festgestellt, und auch beim dritten Test nur bei $1.52\,\%$ bzw. $2.56\,\%$, sodass Itemunabhängigkeit angenommen werden kann. Beim zweiten Hörtest der Gruppe B liegen die Werte mit $13.05\,\%$ bzw. $16.25\,\%$ deutlich höher, allerdings darf hier nicht vergessen werden, dass durch die nachträgliche Korrektur nur 29 Antworten ausgewertet werden konnten. Da der zweite Hörtest mit variabler Klangfarbe jedoch bei den anderen beiden Gütetests ebenso gute Ergebnisse erzielte wie die anderen, wurde er trotzdem in die weitere Auswertung miteinbezogen.

Ähnlich fielen die genannten Test bei den Fachfragen zur Raumakustik aus. Auch hier konnte die Annahme der Itemunanhängigkeit voneinander bestätigt werden, da nur 0.89% bzw. 6.70% Zweier- bzw. Dreierkombinationen an Items das Baertholomew-Kriterium nicht erfüllten. Die Schätzung der Modellgüte der einzelnen Items lieferte ebenfalls in keinem Fall signifikante Abweichungen zu den Ergebnissen der Monte-Carlo-Simulation. Die Eindimensionalität konnte zwar rechnerisch bestätitgt werden (p=0.874), ist inhaltlich jedoch fraglich, wie in der Diskussion näher erläutert wird.

4.3.3. Fähigkeitsausprägung

Faktoranalyse des ersten Hörtests

Um einen latenten Faktor zu messen, der den Grad an Abweichung der einzelnen Antworten von der Referenz beschreibt, wurden zunächst nur die absoluten Distanzen dieser Abweichungen betrachtet und nicht, ob die gewählte Skalenstufe zu hoch oder zu niedrig war. So werden große und kleine Abweichungen in beide Richtungen gleichermaßen behandelt. Das Kaiser-Meyer-Olkin-Kriterium ergab für alle Items einen Wert über 0,6, der die Eignung zur Faktoranalyse bestätigt (vgl. Backhaus et al., 2016, S. 379). Einzig das Item der 110 %-Referenz im Test mit variabler Nachhallzeit, das von 19 Probanden weniger beantwortet wurde als die anderen, liegt mit 0,57 knapp unter der Schwelle, wurde aber dennoch in die Analyse miteinbezogen. Da ein siguläres latente Fähigkeit gemessen werden soll, macht die Extraktion von mehr als einem Faktor keinen Sinn. In Anbetracht der geringen Stichprobe für die zwei verschiedenen Faktoranalysen (Laborgruppe A und B) ist eine Schätzung der Modellgüte generell problematisch.

Dennoch wurde die Modellanpassung nach den Kriterien aus Abschnitt 2.4.4 getestet. Bei der Testung mit variabler Nachhallzeit erfüllte die Extraktion eines Faktors das Eigenwert-Kriterium, bestand den Scree-Test und auch den Chi-Quadrat-Test (p < 0,84). Der RMSEA ergab 0 und ein Tucker-Lewis-Index von mehr als 1 (TLI = 1,353) wies auf eine Überanpassung des Modells hin (vgl. Brant et al., 2011, S. 11). Diese wurde jedoch der geringen Datenmenge, insbesondere durch die fehlenden Antworten beim dritten Item, angerechnet und daher in Kauf genommen, zumal bereits die minimale Faktoranzahl berechnet wurde. Beim Test mit variabler Nachhallzeit implizierten Kaiser-Kriterium und Scree-Test zwar zwei Faktoren, allerdings ergab bereits ein Faktor eine perfekte Anpassung nach dem Tucker-Lewis-Index (TLI = 1). Auch hier lag also bereits eine Überanpassung vor, dementsprechend bestätigte auch der Chi-Quadrat-Test eine gute Modellanpassung (p = 0,88). Die detaillierten Ergebnisse aller Faktoranalysen befinden sich im Anhang.

Die ermittelten Fähigkeitsgrade enthielten keine Information zum absoluten Wert der Wahrnehmungsschwelle. Die Mittelung der absoluten Distanzen über alle Versuchspersonen und Items ergab ein mittleren Schwellwert von 12,7% (Standardabweichung: 13,7,%) für die Nachhallzeit und einen Wert von 8,5% (Standardabweichung: 11,62%) für die Klangfarbe. Für letzteren sind keine Vergleichswerte bekannt, die mittlere Wahrnehmungsschwelle für Unterschiede in der Nachhallzeit liegt im erwarteten Bereich von 10% (siehe Abschnitt 3.1.2), auch wenn die Stan-

dardabweichungen für große individuelle Unterschiede sprechen.

Generalisierte lineare gemischte Modelle

Nachdem festgestellt werden konnte, dass die Itemschwierigkeit adäquat durch die raumakustischen Parameter modelliert werden konnte und das wesentliche Paradigma der Itemunabhängigkeit bei den anderen beiden Hörtests und den Fachfragen eingehalten wurde, wurden für alle diese Tests generalisierte lineare gemischte Modelle berechnet, um die finalen Parameter aller Items und die individuellen Merkmalsausprägungen zu extrahieren. Zur Berechnung der Modelle in R wurde das 1me4-Package von de Boeck et al. (2011) verwendet.

Die Effektstärke der modellierten Zusammenhänge wurde mit der von Nagakawa und Schielzeth (2013) vorgestellten Anpassung des R^2 für generalisierte lineare gemischte Modelle geprüft. Bei dem Verfahren werden zwei Maße berechnet. $R^2_{\rm GLMM}(m)$ (marginal) schätzt die durch die festen Effekte erklärte Varianz. $R^2_{\rm GLMM}(c)$ (conditional) bestimmt die durch das gesamte Modell erklärte Varianz, bestehend aus festen und zufälligen Effekten (vgl. Nagakawa und Schielzeth, 2013, S. 137). Dabei wird jeweils eine "distributionsspezifische Vairanz"berücksichtigt (Nagakawa und Schielzeth, 2013, S. 137). Für diese wurde nach Browne et al. (2005) eine Varianz von $\pi^2/3$ gewählt, die auf der Annahme einer kontinuierlichen, latenten Variable basiert, die sich in den binären Antworten der Versuchspersonen äußert (vgl. S. 604).

Wie bereits in Abschnitt 2.4.3 beschrieben, werden im Modell die Itemparameter als feste und die individuellen Fähigkeiten als zufällige Effekte abgebildet.

Als fester Effekt wurde den Fachfragen dazu auf Basis der Gesamtanzahl richtiger Antworten ein Pseudo-Rang von 1 bis 8 zugeordnet, wobei 1 die geringste und 8 die höchste Erfolgsquote repräsentiert. Dieser Rang ergab einen signifikanten Effekt bei der Modellbildung (p < 0,00), der jedoch nur eher geringen Teil der beobachteten Varianz erklärte ($R_{\rm GLMM}^2(m) = 0,17$). Das gesamte Modell erklärte hingegen schätzungsweise $R_{\rm GLMM}^2(c) = 60\,\%$ der Gesamtvarianz.

Beim zweiten Hörtests, getestet mit variabler Nachhallzeit, ergab diese einen signifkanten Effekt (p=0,00), der 21 % der Gesamtvarianz erklärte. Das gesamte Modell lieferte ein $R_{\rm GLMM}^2$ von 0,47. Auch beim Test mit variabler Klangfarbe war der Effekt derselben signifikant (p<0,00). Der $R_{\rm GLMM}^2(m)$ -Index lag mit 0,28 etwas höher als bei der variablen Nachhallzeit, das $R_{\rm GLMM}^2(c)$ des Gesamtmodells mit 0,37 jedoch niedriger.

Im dritten Hörtest belief sich die geschätzte Varianzerklärung des signifikanten Effekts der Volumenänderung (p < 0,00) auf $R^2_{\rm GLMM}(m) = 0,28$ und die des gesamten Modells auf $R^2_{\rm GLMM}(c) = 0,59$. Tabelle 4.1 zeigt eine vollständige Aufstellung der Teststatistiken der Modelle. Neben den R^2 -Indizes sind das Akaike-Informationskriterium (AIC), ein Maß für den Messfehler einer Modellschätzung und das Bayesische Informationskriterium angegeben, dass dem AIC ähnelt, jedoch die Anzahl vieler Variablen stärker bei der Berechnung des Testfehlers berücksichtigt (vgl. James et al., 2017, S. 213) . Die Tabelle zeigt außerdem die Devianz als Maß für die Abweichung der tatsächlichen Messwerte von der Modellschätzung (vgl. Mayer und Butler, 1993, S. 24).

Test	AIC	BIC	Devianz	$R^2_{\text{GLMM}}(m)$	$R_{\rm GLMM}^2(c)$
2. Hörtest: Nachhallzeit	908,3	922,7	902,3	0,21	0,47
2. Hörtest: Nachhallklangfarbe	878,2	892,0	872,2	$0,\!28$	0,37
3. Hörtest (Sortierung)	1334,9	1350,4	1328,9	$0,\!28$	0,59
Quiz	986,0	1000,1	980,0	$0,\!17$	0,60

Tabelle 4.1.: Modellparameter der generalisierten linearen gemischten Modelle. Neben dem AIC

Zur Bestimmung der Itemparameter wurden nach Harrison und Müllensiefen (2018) die Itemdiskrimination und Itemschwierigkeiten aus der Standardabweichung der zufälligen Effekte und der Koeffizienten der festen Effekte berechnet (vgl. S. 6). Die Ratewahrscheinlichkeit war bekannt, als Unaufmersamkeitparameter wurde auch hier der Wert 1 angenommen (kein Effekt durch Unaufmerksamkeit).

Abbildung 4.10 zeigt die so berechneten Itemdiskriminationen und -Schwierigkeiten. Da im Modell ein konstanter Effekt der raumakustischen Größen auf die Schwierigkeit angenommen wird, stimmen die Abstände der Itemschwierigkeiten mit den Modulationsgraden der raumakustischen Parameter überein. Die Unterschiede zwischen den jeweiligen Räumen in den Items nehmen von linsk nach rechts ab. Auch die Itemdiskrimination wurde initial als konstant angenommen und ist daher innerhalb der jeweiligen Tests konstant. Dennoch lässt sich ablesen, welche vermuteten Schwierigkeitsbereiche abgedeckt werden, bzw. wie stark die Schwierigkeit zwischen den Items steigt.

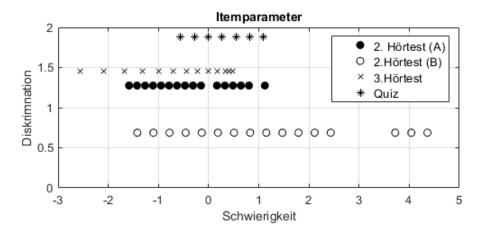


Abbildung 4.10.: Geschätzte Itemschwierigkeit und -Diskrimination der GLMMs. Die Items mit geringer Schwierigkeit entsprechen denen mit großen Differenzen in der Akustik der jeweiligen Räume.

Auf Basis dieser Itemparameter und der individuellen Antwortsätze der Probanden wurden die individuellen Fähigkeitsausprägungen mit Hilfe des catR-Packages in R (Magis und Raîche, 2011) geschätzt. Abbildung 4.11 zeigt die aus den Hörtests und Fachfragen ermittelten Fähigkeitsgrade. Da die Ergebnisse der Faktoranalysen für die ersten Hörtests andere Metriken besitzen als die der GLMMs, wurden alle Verteilungen z-standardisiert, also auf die Standardabweichung normiert. Im ersten Hörtest sind die Ergebnisse sehr ähnlich und stellen jeweils nach links verlagerte Normalverteilungen dar.

Die anderen Fähigkeitsgrade sind etwa gleichverteilt. Beim zweiten Hörtest sind die Messergebnisse bei variabler Nachhallklangfarbe etwas gleichmäßiger verteilt und nehmen zu hohen Fähigkeitsgraden relativ konstant ab. Die Ergebnisse der Fachfragen, konzentrieren sich hingegen stärker auf den Null- und die Extremwerte. Gerade beim höchsten Wert ist das insofern wenig verwunderlich, als dass ein Zehntel der Personen alle Fragen richtig beantworten konnte. Aus den Messungen geht hervor, dass auch in Bezug auf die perzeptiven und kognitiven Fähigkeiten die Stichprobe als relativ heterogen eingestuft werden kann.

Die Ergebnisse der Hörtest wurden auf Korrelationen untersucht, um festzustellen, ob sich hier nicht eine übergeordnete auditorische Kompetenz messen lässt. Die Korrelationsmatrizen für die Laborgruppe A und B sind in Tabelle 4.2 dargestellt. Es ist zu beachten, dass beim ersten Hörtest niedrige Werte ein gutes Ergebnis bedeuten, da die Abweichung von der korrekten Antwort gemessen wurde. Eine negative Korrelation mit einem der anderen Hörtests bedeutet dementsprechend ein gutes Ergebnis beiden Tests. Es konnten jedoch keine signifikanten Kor-

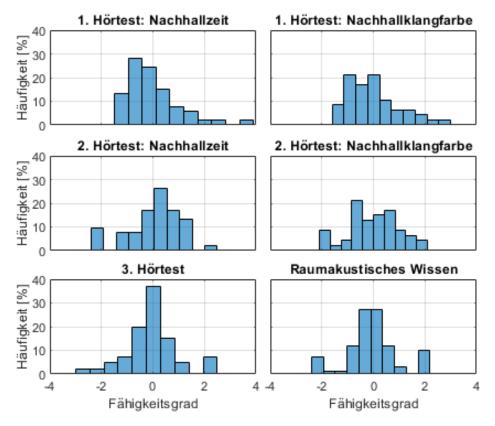


Abbildung 4.11.: Verteilung der ermittelten Fähigkeitsgrade. Da die Faktoranalyse eine andere Metrik lieferte als die GLMMs, sind die Darstellungen z-standardisiert.

relationen festgestellt werden. Der geringste p-Wert der positiven Korrelation zwischen dem ersten Hörtest mit variabler Nachhallzeit und dem dritten Test verfehlte mit p=0,06 knapp das definierte Signifikanzniveau von 0,05.

	Na	chhallz	eit	Nachhallklangfarbe			
Test	HT1	HT2	HT3	HT1	HT2	HT3	
HT1	1			1			
HT2	-0,18	1		-0,26	1		
HT3	-0,18	-0,13	1	0,28	0,05	1	
Signif	ikanztes	sts					
HT1	0,00			0,00			
HT2	0,19	0,00		0,08	0,00		
HT3	0,21	$0,\!37$	0,00	0,06	0,75	0,00	

Tabelle 4.2.: Korrelationsmatrix der drei Hörtests. HT1 bezeichnet den ersten, HT2 den zweiten Hörtest und HT3 den dritten Hörtest.

4.4. Faktoranalyse aller Testkomponenten

Die finale Auswertung des gesamten Versuchs erfolgte in zwei Schritten. Zunächst wurden durch eine weitere exploratorische Faktoranalyse latente Faktoren extrahiert, um Strukturen aufzudecken, die eine raumakustische Hörexpertise widerspiegeln können. In diese Faktoranalyse nicht

miteinbezogen wurde die Information, ob die Probanden im dritten Hörtest den Zusammenhang zwischen wahrgenommenen Klangunterschieden und Raumgröße korrekt hergestellt haben. Diese Fähigkeit, anhand eines Raumklangs den Raum charakterisieren zu können, stellt nach Abschnitt 2.1 den Kern raumakustischer Hörexpertise dar. Daher wird im Anschluss an die Faktoranalyse eine multiple Regression mit den extrahierten Faktoren durchgeführt, um zu testen, ob diese tatsächlich einen Einfluss darauf haben, wie der Zusammenhang zwischen Raumgröße und Raumklang beurteilt wird.

4.4.1. Explorative Faktoranaylse

Für die Faktoranalyse wurden in den ersten beiden Hörtests die Fähigkeitswerte der Testungen mit variabler Nachhallzeit und variabler Klangfarbe jeweils zu einer Variablen zusammengefasst. Das basiert auf der Überlegung, dass im ersten Hörtest in beiden Fällen eine physiologische Sensibilität und im zweiten Test jeweils die Fähigkeit gemessen wird, einen Raum als auditorisches Objekt zu abstrahieren. Ob diese Zusammenfassung bedenkenlos vorgenommen werden kann, muss in nachfolgenden Untersuchungen getestet werden.

Da jeweils relative Fähigkeitswerte gemessen und die einzelnen Variablen vor der Faktoranalyse standardisiert wurden, ist die Annahme, dass eine Person mit überdurchschnittlicher Merkmalsausprägung beim Test mit dem einen raumakustischen Parameter auch beim Test mit einem anderen einen solche höhere Ausprägung aufweisen würde. Vor der Faktoranalyse wurde die Kor-

	HT1	HT2	HT3	Quiz	M-W	M-U	M-T	E-A	E-M	E-R
HT1	1,00									
HT2	-0,20	1,00								
HT3	$0,\!12$	-0,06	1,00							
Quiz	-0,15	0,17	$0,\!12$	1,00						
M-W	-0,09	$0,\!32$	-0,06	$0,\!18$	1,00					
M-U	-0,06	0,14	-0,12	$0,\!13$	$0,\!34$	1,00				
M-T	-0,10	$0,\!25$	0,03	$0,\!26$	$0,\!16$	$0,\!19$	1,00			
E-A	-0,24	$0,\!30$	0,00	$0,\!67$	$0,\!22$	$0,\!22$	$0,\!36$	1,00		
E-M	-0,15	$0,\!31$	-0,11	$0,\!40$	$0,\!25$	$0,\!45$	$0,\!41$	$0,\!52$	1,00	
E-R	-0,20	$0,\!43$	-0,09	$0,\!33$	$0,\!21$	$0,\!42$	$0,\!28$	$0,\!32$	0,68	1,00
MSA	0,65	0,69	0,52	0,67	0,68	0,77	0,88	0,68	0,75	0,69

Tabelle 4.3.: Korrelationsmatrix der Variablen, die in die Faktoranalyse einbezogen wurden: die drei Hörtests (HT1-3), das Ergebnisse der Fachfragen (Quiz), die Gold-MSI-Faktoren "musikalische Wahrnehmungsfähigkeiten" (M-W), "aktiver Umgang mit Musik" (M-U), "musikalisches Training" (M-U) und die drei Erfahrungskategorien "akademisch" (E-A), "musikalisch" (E-M) und "im Rahmen von Aufnahmen" (E-A). Signifikante Korrelationen (p < 0,05) sind hervorgehoben und für jede Varialbe die MSA angegeben.

relationsmatrix und die individuelle Eignung der Variablen zur Faktoranalyse (mean-sampling-adequacy, MSA) betrachtet (siehe Tabelle 4.3), aus denen sich ein Kaiser-Meyer-Olkin-Kriterium von 0,71 berechnete (vgl. Backhaus et al., 2016, S. 378). Daraus ergab sich, dass fast alle Variablen und die Gesamtheit aller für eine Faktoranalyse geeignet sind (vgl. Backhaus et al., 2016, S. 378f). Einzige Ausnahme bildet das Ergebnis des dritten Hörtests, also die Fähigkeit, die Reihenfolge der Räume anhand ihres Klangs (unabhähgig von der Bedeutung des Klangs) in eine korrekte Reihenfolge zu bringen. Das Ergebnis des dritten Hörtests wurde daher aus der Faktorextraktion ausgeschlossen.

Wie in Abschnitt 2.4.4 bereits erwähnt, wurde den Empfehlungen von Fabrigar et al. (1999) folgend zur Faktorextraktion die Maximum-Likelihood-Methode und oblique Rotation verwendet, die eine Korrelation der Faktoren erlaubt (vgl. Backhaus et al., 2016, S. 400).

Zur Bestimmung der Faktoranzahl wurde eine Reihe von Kriterien getestet. Gemäß dem Kaiserkriterium gab es nur einen Faktor mit einem Eigenwert über eins (wenn auch der Eigenwert eines zweiten bei 0,96 lag). Der Scree-Test hingegen unterstützte eine 3-Faktor-Löung. Die Berechnung weiterer Kriterien mit Hilfe des psych-Packages (Revelle, 2019) in R ergab nur bis zu einer Anzahl von fünf Faktoren in allen Fällen Lösungen. Tabelle 4.4 zeigt den RMSEA, BIC (normal berechnet und an die Stichprobengröße angepasst), die Chi-Quadrat-Statistik mit einem p-Index und das very-simple-structure-Kriterium (VSS) für zwei Komplexitätsstufen, dass die Annäherung an ein einfaches Modell mit einfach zu interpretierenden Faktoren beschreibt (vgl. Revelle und Rocklin, 1979, S. 405f). Auch der Velicer's-minimum-average-partial-Test ist angegeben, der den Anteil an Varianz in der Korrelationsmatrix darstellt, der nicht durch die Faktoranalyse erklärt wurde (vgl. O'Connor, 2000, S. 396).

Faktoren	VSS1	VSS2	MAP	χ^2	p_{χ^2}	RMSEA	BIC	SABIC
1	0,67	0,00	0,043	73,00	0,00	0,131	-51,0	34,278
2	0,56	$0,\!54$	0,062	28,00	0,08	0,069	-59,4	0,657
3	0,55	$0,\!53$	0,087	15,00	0,23	0,051	-40,0	-2,123
4	0,57	$0,\!62$	0,137	5,70	0,46	0,000	-21,9	-2,972
5	0,50	0,65	0,215	1,40	$0,\!23$	0,063	-3,2	-0,023

Tabelle 4.4.: Verschiedene Kriterien zur Auswahl der besten Faktoranzahl: very-simple-strucutre für Komplexitäten 1 oder 2 (VSS1 bzw. VSS2), minimal average partial (MAP), Chi-Quadrat-Test (χ^2) mit Signifkanzlevel (p_{χ^2}), RMSEA, BIC und an die Stichprobengröße angepasstes BIS (SABIC).

Aus der Tabelle 4.4 wird ersichtlich, dass die verschiedenen Kriterien entweder einen (VSS1, MAP), zwei (BIC), vier (RMSEA, SABIC,, χ^2) oder fünf Faktoren (VSS2) favorisieren. Tests der Faktoranzahlen ergaben, dass das Modell mit vier Faktoren eine Überanpassung an die Daten lieferte (Tucker-Lewis-Index von 1,008) und daher aus dem Vergleich ausgeschlossen wurde. Somit wäre nach dem RMSEA und dem SABIC die Lösung mit drei Faktoren vorzuziehen. Der Chi-Quadrat Test ist sehr von der Stichprobengröße abhähngig und daher bei 100 Probanden nicht unbedingt aussagekräftig (vgl. Fabrigar et al., 1999, S. 280). Dass auch die anderen Kriterien so unterschiedliche Werte liefern, könnte auch ein Indiz für Präsenz von Faktoren höherer Ordnung sein (vgl. Müllensiefen et al., 2014, S. 4). Daher wurde nach Fabrigar et al. (1999, S. 280) das Modell mit dem geringsten RMSEA bevorzugt und es ergab sich unter Verwendung der Schmid-Leiman-Transfomration und 'oblimin'-Rotation die beste Modellanpassung bei drei Faktoren erster Ordnung und einem weiteren, übergeordneten Faktor.

Dieser erklärt 42% der durch die Faktoren ingesamt beschriebenen Varianz und kann damit als eine relevante Komponente angesehen werden (vgl. Wolff und Preising, 2005, S. 50). Wie in Tabelle 4.4 dargestellt, liegt der RMSEA des Modells mit 0,051 genau auf der Schwelle zur guten Modellanpassung und nach dem Scree-Test, SABIC und Chi-Quadrat-Test ergeben drei Faktoren erster Ordnung die beste Lösung. Zur Bestimmung der Reliabilität wurden McDonald's ω_t und Cronbach's α verwendet, die das Verhältnis der beobachteten zur tatsächlichen Varianz der gemessenen Kompetenzen eines Probanden beschreiben. Letzteres nimmt dabei eine direkte Messung dieser Kompetenzen durch die Items und ersteres lediglich eindimensionale Kompetenzen an (vgl. Zhang und Yuan, 2016, S. 388). Cronbach's α lag für das gesamte Mo-

dell bei 0,77 und McDonald's ω_t bei 0,85. Nach Revelle und Zinbarg (2009, S. 151) und Dunn et al. (2014, S. 409) ist gerade bei mehrdimensionalen Konstrukten McDonald's ω_t der bessere Reliabilitätsindex. Bei dem hier getesteten Modell liefert dieser einen akzeptablen Wert. Wie gut die Items den übergeordneten, latenten Faktor beschreiben, wird mittels ω_t gemessen, das in diesem Fall mit 0,56 deutlich niedriger lag (vgl. Revelle und Zinbarg, 2009, S. 152). Als letztes wurde in einer Faktoranalyse mit nur einem Faktor getestet, ob die Faktoranalyse auf zwei Ebenen überhaupt erfordlich ist. Hier ergab sich jedoch eine deutlich schlechtere Modellanpassung (RMSEA: 0,144).

Variable	G1	F1*	F2*	F3*	Kommunalität
1. Hörtest	-0,18	-0,11	-0,17	0,17	0,10
2. Hörtest	0,39	$0,\!22$	0,09	$0,\!19$	0,24
Quiz	0,46	$0,\!07$	$0,\!49$	0,01	0,46
Gold-MSI: Wahrnehmung	0,37	0,00	0,00	0,93	1,00
Gold-MSI: Umgang	0,39	$0,\!29$	-0,01	$0,\!21$	0,28
Gold-MSI: Training	0,35	0,18	0,19	0,03	0,19
Erfahrung: akademisch	0,60	-0,01	0,80	0,00	1,00
Erfahrung: musikalisch	0,65	$0,\!46$	$0,\!17$	$0,\!02$	0,66
Erfahrung: Aufnahmen	0,63	$0,\!62$	-0,06	0,02	0,79
Eigenwerte	1,99	0,78	0,99	0,97	
Anteil Gesamtvarianz					
Faktor (vor Transformation)		0,21	0,20	0,12	
kumuliert		$0,\!21$	$0,\!40$	$0,\!52$	
Anteil an gemeinsamer Varianz	0,42				
McDonald's ω_t	0,85	0,77	0,70	0,99	

Tabelle 4.5.: Ladungen des übergeordneten Faktors G1 und der drei Subfaktoren (F1*, F2*, F3*) nach der Schmid-Leiman-Transformation und Kommunalitäten der Variablen. Zur besseren Übersicht sind nur Faktorladungen über 0,2 hervorgehoben. Zudem sind die erklärte Varianz, die Eigenwerte und die individuellen ω_t -Werte der Faktoren angegeben.

Tabelle 4.5 zeigt die Faktorladungen und Eigenwerte der vier Faktoren nach der Schmid-Leiman-Transformation. Die Kommunalitäten messen den Anteil an Varianz eines Parameters, der durch die Faktoren erklärt wird (vgl. Backhaus et al., 2016, S. 411). Es ist auch hier zu beachten, dass beim ersten Hörtest kleine Werte eine geringe Wahrnehmungsschwelle für Unterschiede bedeuten. Ladungen über 0,2 sind hervorgehoben. Die Faktorladungen der drei Faktoren erster Ordnung vor der Schmid-Leiman-Transformation befinden sich im Anhang.

4.4.2. Prädiktion der Größeninterpretation

In einem letzten Schritt wurden die extrahierten Faktorwerte als Prädiktoren für die korrekte Zuordnung der Raumgrößen zur erkannten Reihenfolge der Stimuli im dritten Hörtest getestet. Auch die ebenfalls in diesem Test bestimmte Fähigkeit, die korrekte Reihenfolge unabhängig von der Interpretation derselben zu finden, wurde als Prädiktor in die Modellbildung miteinbezogen. Das Ergebnis der ersten Hörtests wurde bei der Berechnung der individuellen Faktorwerte ausgeschlossen, da es hier keine Faktorladung über 0,3 gab, was nach Howard (2016) ein unteres Limit für relevante Faktorladungen in der bisherigen Literatur darstellt (vgl. S. 55).

Zunächst wurden sechs Modelle verglichen: nur der übergeordnet Faktor, nur die drei anderen Faktoren, alle Faktoren, der übergeordnete Faktor und das Ergebnis der Sortierung, die drei

anderen Faktoren und das Sortierergebnis sowie zuletzt alle Faktoren und das Sortierergebnis zusammen. Ein Chi-Quadrat-Test ergab keine signifikanten Verbesserungen zwischen den Modellen.

Um unter allen weiteren Kombinationen der Prädiktoren das beste Modell zu finden, wurde dann nach dem Prinzip der Forward Stepwise-Selection ausgehend von einem Basismodell mit nur dem übergeordneten Basisfaktor begonnen und schrittweise die drei anderen Faktoren und das Ergebnis der Sortierung hinzugefügt (vgl. James et al., 2017, S. 207f). Parallel wurde nach der Backward Stepwise-Selection ausgehend vom Modell mit allen Parametern schrittweise jeweils einer entfernt, um das Modell mit dem besten Informationskriterium zu ermitteln (vgl. James et al., 2017, S. 209). Der Auswahlprozess wurde mit Hilfe der stepAIC-Funktion des MASS-Packages (Ripley et al., 2019) in R implementiert und orientiert sich an der Minimierung des Akaike-Informationskriteriums.

Parameter	B	σ	B/σ	$P(> B/\sigma)$
Konstantterm	-0,399	0,216	-1,851	0,064
G1	0,274	0,089	3,072	0,002

Tabelle 4.6.: Regressionskoeffizienten B mit Standardabweichung σ der multiplen logistischen Regression zur Prädiktion der korrekten Zuordnung der Raumgrößen im dritten Hörtest. Eine letzte Spalte zeigt das Ergebnis eines t-Tests, ob der Regressionskoeffizient unter Berücksichtigung der Standardabweichung einen signifikanten Effekt darstellt. Signifikante Werte sind hervorgehoben (Signifikanzniveau $\alpha = 0,05$).

Das beste Modell enthielt nur den übergeordneten Faktor als Prädiktor und ergab einen positiven signifikanten Effekt. Tabelle 4.6 zeigt den berechneten Regressionskoeffizienten. Zur zusätzlichen Testung des Modells wurde die Anpassung des R^2 -Index für GLMs nach Nagelkerke genutzt, der eine Varianzaufklärung von 13 % ergab (vgl. Nagakawa und Schielzeth, 2013, S. 135).

4.4.3. Inhaltliche Auswertung

Der erste Faktor zeigt vor allem bei den Parametern hohe Ladungen auf, die einen praktischen Umgang mit Raumakustik und Musik widerspiegeln. Hinzu kommt eine leicht erhöhte Faktorladung beim zweiten Hörtest, also der kognitiven Separation von Quelle und Raum. Dass diese durch praktische Routine trainiert wird, scheint plausibel, weswegen der Faktor als diese Eigenschaft, bzw. Erfahrung interpretiert wird.

Mit nur zwei hohen Faktorladungen bei den Fachfragen und der akademischen Erfahrung hängt der zweite Faktorhängt offenbar mit theoretischem Wissen über Raumakustik zusammen. Beim dritten Faktor gibt es eine besonders hohe Faktorladung beim Gold-MSI-Faktor 'musikalische Wahrnehmungsfähigkeiten'. Bei diesem bewertet die Person ihre eigene Fähigkeit, musikalische Darbietungen vergleichen und z.B. falsch gesungene Tonhöhen zu erkennen (vgl. Müllensiefen et al., 2014, S. 5).

Eine hohe Selbsteinschätzung hierbei korreliert nicht nur erwiesenermaßen mit audiometrischen Testergebnissen (siehe Abschnitt 3.3.2), sondern impliziert auch zu einem gewissen Grad, dass sie tendenziell auf derartiges achtet, wenn sie Musik hört und analytische Hörgewohnheiten hat. Der dritte Faktor kombiniert dies mit leicht erhöhten Faktorladungen beim Ergebnis des zweiten Hörtest, der ebenfalls eine kritische Auseinandersetzung mit gehörtem Klang erfordert und einer ähnlich hohen Verbindung mit dem Gold-MSI-Faktor 'aktiver Umgang mit Musik'. Aus diesem Grund wird dieser Faktor als Ausdruck eines kritischen Gehörs interpretiert, das zu einem gewissen Grad mit einem musikalischen Interesse einhergeht. Interessanterweise ist dieser Faktor der

einzige mit einer positiven Faktorladung beim ersten Hörtest, der eigentlich genau die perzeptive Sensibilität im raumakustischen Kontext messen soll. Durch die inverse Metrik dieses Faktors ergibt die Faktorladung jedoch, dass dieses kritische Gehör eher ein schlechteres Ergebnis bei diesem Test bedeuten würde. Allerdings muss beachtet werden, dass der erste Hörtest ohnehin durch die geringe Kommunalität von nur 0,1 vor allem durch andere Parameter beeinflusst zu werden scheint.

Der letzte Faktor weist bei fast allen Parametern hohe Ladungen auf und ist ja auch für fast die Hälfte der Varianz verantwortlich, die durch die Faktoren insgesamt erklärt wird. Dieser Faktor zeigt die höchsten Korrelationen mit allen Variablen, die die Erfahrung mit Raumakustik beschreiben. Die Faktorladung bei den Fachfragen liegt etwas niedriger, gefolgt von etwa gleichen Korrelationen beim zweiten Hörtest und den Gold-MSI-Faktoren.

Da dieser Faktor Erfahrung, Fachwissen und die Leistung im Hörtest in sich vereint, kann hierin durchaus die in Abschnitt 2.1 definierte raumakustische Hörexpertise vermutet werden. Bekräftigt wird das durch den signifikanten, wenn auch insgesamt geringen Einfluss auf die Interpretation im dritten Hörtest. Dem Modell nach trägt eine Kompetenz, bestehend aus Wissen, Erfahrung und kritsichem Gehör dazu bei, dass eine Person auf Basis klanglicher Differenzen zweier Räume physikalische Unterschiede zwischen diesen Räumen korrekt beurteilen kann. Diese Beobachtung entspricht der gegebenen Definition raumakustischer Hörexpertise und bekräftigt die Vermutung, dass ein solche Fähigkeit existiert, bzw. sich durch Erfahrung und Fachwissen ausbildet.

5. Diskussion

5.1. Testverfahren

Die drei Hörtests wurden in dieser Untersuchung entworfen und erstmalig angewendet. Die Auswertung der Item-Reponse-Theorie-Modelle ergab, dass die grundlegenden Annahmen zur Kontrolle der Itemschwierigkeit bestätigt wurden und geringe Unterschiede die Schwierigkeit erhöhen. Eine Gegenüberstellung der Itemschwierigkeiten aller Tests hat allerdings gezeigt, dass die Verteilung der Schwierigkeitsgrade und auch die Itemdiskrimination unterschiedlich ausfällt (siehe Abbildung 4.7 und 2.2). Die jeweilige Metrik ist im Kalibrierungsprozess an die Stichprobe gebunden (vgl. Baker, 2001, S. 131), sodass zumindest bei den beiden Varianten des ersten und zweiten Hörtests eine Vergleichbarkeit der beiden Teilstichproben nur angenommen werden kann. Dennoch wird offenkundig, dass etwa der dritte Hörtest insgesamt niedrigere Schwierigkeiten enthält als der zweite Hörtest mit variabler Nachhallklangfarbe.

Zur Kalibrierung der Itemparameter wurden 100 Probanden als notwendige Stichprobengröße festgehalten (siehe Abschnitt 2.4.3). Durch die Aufteilung in zwei Gruppen wurde in den ersten beiden Hörtests de facto nur etwa die Hälfte dieser Zahl erreicht. Dieser Umstand wurde zugunsten der zusätzlichen Information durch die Testung mit zwei raumakustischen Parametern in Kauf genommen, schwächt jedoch prinzipiell die Aussagekraft der ermittelten Itemparamter und Fähigkeitsgrade.

Unter der Annahme, dass in den Teilstichproben A und B die Kompetenzen in etwa gleich verteilt sind, ergeben sich unterschiedliche Schwierigkeitsbandbreiten für die Test-Items, obwohl diese nach möglichst identischen Schritten generiert wurden (siehe Abschnitt 3.1.2). Interessanterweise sind zudem im ersten Hörtest die Abweichungen vom korrekten Ergebnis und die über alle Probanden gemittelte Wahrnehmungsschwelle mit ca. 8% (Standardabweichung: 11,62%) bei der Testung mit der Nachhallklangfarbe etwas niedriger als beim Test mit der Nachhallzeit (Mittelwert: 12,7%, Standardabweichung: 13,7%). Dabei sind bei ersterem die Schwankungen in der Lautheit der Stimuli bzw. dem Stärkemaß, die als wichtiges Kriterium der Wahrnehmung von Raumklang gelten (siehe Abschnitt 2.1.5), kleiner.

Die Variation der Klangfarbe ergab also im ersten Test etwas bessere Ergebnisse, die für eine höhere Sensibilität sprechen, führte aber im zweiten Test zu höheren Schwierigkeitsgraden bzw. niedrigeren Erfolgsquoten. Das kann verschiedene Gründe haben. Einerseits besitzen die einzelnen Klavierpassagen, die als Ausgangssignale fungierten, unterschiedliche Frequenzgänge und Lautstärken (siehe Abbildung 3.5). Da die Sensibilität des menschlichen Gehörs für Schall in einzelnen Frequenzbändern von dessen Intentistät abhängt (vgl. Weinzierl, 2008, S. 55), kann die absolute Lautstärke der Stimuli beeinflussen, wie die Klangfarbe des Raumklangs wahrgenommen wird. Auch besitzen die drei Klavierauzüge "eigene"Klangfarbendifferenzen. Diese von denen des Raumes zu separieren, ist zwar der Kern dieses Tests, dennoch sollte in nachfolgenden Anwendungen eine größere Homogenität der Ausgangsstimuli, zumindest in Bezug auf die Lautstärke, angestrebt werden. Eine Möglichkeit wäre, in den beiden Referenzen des Tests (A und B) das gleiche Instrumentalstück zu nehmen und nur den Raum zu variieren und dann nur für den Vergleichsstimulus (X) ein anderes Ausgangssignal zu verwenden.

Ein weiterer kritischer Punkt war die Einschränkung der Wiedergabe in diesem Test, durch die jeder Stimulus nur einmal gehört werden konnte. Dieser Schritt war zwar notwendig, um die Testlänge kontrollieren zu können, macht die Durchführung jedoch sehr anfällig für Unauf-

merksamkeit. Dieser Parameter wurde zwar bei der Berechnung der IRT-Modelle mangels einer Methode, um einen geeigneten Wert festzulegen, vernachlässigt, wird aber vermutlich vielfach einen Einfluss auf die Ergebnisse gehabt haben. Einerseits berichteten viele Versuchspersonen, dass dadurch der Vergleich erschwert wurde und sie gelegentlich mit momentaner Unaufmerksamkeit zu kämpfen hatten. Andrerseits spricht in beiden Tests auch die relativ geringe Varianzaufklärung durch die berechneten GLMMs dafür, dass hier noch andere, situative Komponenten eine Rolle gespielt haben.

Dass es keine Korrelationen zwischen den Hörtests gab, unterstützt den Eindruck, dass jeweils unterschiedliche Aspekte der auditorischen Verarbeitung getestet werden konnten. Im zweiten Hörtest konnte also offenbar die kognitive Separierung von Schallquelle und Raum ansatzweise unabhängig von der physiologischen Sensiblität für Raumklang getestet werden, sodass dieser Test in Zukunft ausgebaut werden sollte. Hierbei könnten zum einen die Performanz derselben Person bei den beiden ausgesuchten raumakustischen Parametern (und noch weiteren) untersucht werden, und zum anderen versucht werden, die konkreten perzeptiven Eindrücke und Strategien, die bei der Unterscheidung genutzt werden, zu identifizieren.

Beim dritten Hörtest wurde festgestellt, dass der Test im Grunde zwei Ergebnisse liefert: die Fähigkeit einer Person, klangliche Differenzen zu erkennen und in eine Reihenfolge zu bringen und die korrekte Interpretation der gehörten Unterschiede. Das ist ein großer Vorteil dieser Sortieraufgabe gegenüber beispielsweise einem ABX-Test, in dem es "nur'richtige und falsche Antworten gibt. Auch hier gibt es zunächst nur eine richtige Reihenfolge der Stimuli. Die falschen Antworten können jedoch, da durch die Sortierung ein explizites Antwortmuster eingestellt werden muss, in zwei Kategorien aufgeteilt werden. Es also wird ein Anhaltspunkt geliefert, welchen Fehler die Person gemacht hat: ob sie die Reihenfolge der klanglichen Unterschiede nicht erkennen konnte, oder ob sie den Zusammenhang zur Raumgröße falsch verstanden hat.

Mit Blick auf die Testkonstruktion nach der probabilistischen Testtheorie muss festgehalten werden, dass nur im Bezug auf das erste Ergebnis graduelle Schwierigkeiten erzeugt wurden. Ob die Versuchsperson die Raumklänge gut unterscheiden kann oder nicht, ändert vermutlich nichts an ihrem Verständnis, wie z.B. die Lautheit eines Stimulus mit der Größe des Raumes zusammenhängt. Hier sind natürlich auch innerhalb einer gegebenen Aufgabe keine Schwierigkeitsvariationen möglich: entweder wird eine erhöhte Lautstärke mit einem kleineren oder einem größeren Raum in Verbindung gebracht.

Was sich hingegen ändern kann, sind die konkreten Eindrücke, die zur Interpretation ausgewertet werden, also ob auf die Lautstärke geachtet wird, oder auf etwas anderes. Die Nachbesprechungen mit den Probanden ergaben außerdem, dass gerade von Probanden mit geringer Fachkenntnis die Deutungshypothese während der Durchführung geändert wurde und zum Beispiel in den ersten drei Items kleine Räume mit lauten Klängen assoziiert wurden und in den restlichen Items größere. Die Probanden erklärten, dass die wiederholte Beantwortung der Items sie zunehmend dazu gebracht hätte, den Zusammenhang zwischen Klang und Raumgröße kritisch zu hinterfragen, weswegen sie aus Unsicherheit die Interpretation geändert hätten. Das legt die bereits in Abschnitt 2.3.3 erwähnten Trainingseffekte nahe.

Durch die Randomisierung der Items ließ sich die Reihenfolge, in der die Items beantwortet wurden nicht aufzeichnen. Somit konnten diese Änderungen in den Ergebnissen nicht nachvollzogen werden. In zukünftigen Anwendungen sollte das geändert werden, um zu beobachten, wie oft, nach welcher Itemanzahl und in welche Richtung sich die Deutungshypothese ändert. Außerdem wurde mit dem Vergleich der Raumgrößen bei gleicher Nachhallzeit eine relativ spezielle Situation abgefragt. Weitere Tests sollten gefunden werden, die auf ähnliche Weise eine bestimmte inhaltliche Interpretation gehörter Unterschiede durch die Versuchsperson erfordern.

Zuletzt wurden auch die Fachfragen nach der probabilistischen Testtheorie organisiert. Die Schaffung gradueller Schwierigkeitsgrade gelang erstaunlich gut und auch die Modelltests auf Itemunabhängigkeit und Eindimensionalität lieferten positive Ergebnisse.

Das ist inhaltlich jedoch fraglich, wenn man bedenkt, dass in diesem Fall keine eindeutige Fähigkeit wie in den Hörtests abgefragt wurde, sondern die richtige Antwort zu einem raumakustischen Sachverhalt. Die Fragen bedienen dabei einerseits verschiedene Unterthemen dieses Fachs, andererseits können auch verschiedene Kompetenzen der Person ausschlaggebend sein, etwa ob sie die Information auswendig kennt, oder sich die Antwort auf Basis ähnlicher Sachverhalte herleitet. Unabhängig von dem Ergebnis der Modifizierten Parallelanalyse kann daher nicht von Eindimensionalität ausgegangen werden. Eine solche singuläre Fähigkeit zu messen, war aber auch nie das Ziel des Tests. Vielmehr sollte eine Kontrollvariable zur Selbstauskunft geschaffen werden, die anfällig für Antwortneigungen und verschiedene Interpretationen der Antwortoptionen ist. Der gemeinsame Faktor und eine hohe Korrelation sprechen für eine gute Übereinstimmung zwischen Selbsteinschätzung und dem Ergebnis der Fachfragen.

Generell ist noch anzumerken, dass in den Hörtests teilweise eer unrealistische Nachhallcharakteristiken simuliert wurden. Dass zwei Räume mit Volumina von 250 m³ und 2000 ³ die gleiche Nachhallzeit aufweisen, ist eher unwahrscheinlich, ebenso wie die steilen Frequenzgänge der Nachhallzeiten durch die Klangfarbenvariation. Bei letzterem wäre eine Möglichkeit, statt einer perfekt frequenzunabhängigen Nachhallzeit eine spektral ausgewogene, aber dennoch variable Charakteristik als Referenz zu nehmen und auf Basisv dieser die Nachhallzeit bei höhen und tiefen Frequenzen systematisch anzuheben bzw. abzusenken.

Was die extremen Raumgrößen angeht, so hat sich gezeigt, dass die Itemschwierigkeiten bei diesen ohnehin sehr gering und die Erfolgsquoten relativ konstant hoch sind. Daher könnten in zukünfitgen Tests die Items mit sehr extremen Differenzen entfernt werden.

Zudem sollte im Vorfeld die Konfiguration der Raumsimulationen noch einmal genauer geprüft und optimiert werden, um Approximationsfehler und die erforderlichen Korrekturen der Tests zu vermeiden, die letzten Endes zu Datenverlust führen.

5.2. Faktoranalyse und Multiple Regression

Die Ergebnisse der Faktoranalyse und der anschließenden Prädiktion der Interpretation im dritten Hörtest sind prinzipiell im Einklang mit den theoretischen Vorüberlegungen dieser Arbeit. Aus dem vorigen Abschnitt geht jedoch hervor, dass gerade die Ergebnisse des zweiten Hörtests relativ vielen Störeinflüssen unterlagen und noch einmal präziser und differenzierter erfasst werden sollten. Auch die extrahierten Faktoren, insbesondere die drei Subfaktoren der Schmid-Leiman-Transformationen zeigen jeweils nur bei zwei bis vier Items nennenswerte Faktorladungen. Daher sollte in Zukunft die Reproduzierbarkeit dieses Faktormodells, im Idealfall mit mehr Items für die verschiedenen Faktoren, getestet werden.

Zudem ist die Relevanz des dritten Faktors (kritisches Gehör) fraglich. In den getesteten Regressionsmodellen konnten für diesen Faktor keine Koeffizienten berechnet werden, der im wesentlichen bloß einen Faktor des Gold-MSI widerspiegelt. Die beiden anderen Subfaktoren, praktische Routine und Fachexpertise, drücken Komponeten aus, die auch im übergeordneten Faktor hohe Faktorladungen aufweisen und sollten daher weiter präzisiert werden.

Vor allem die Erfahrung einer Person kann und sollte noch detaillierter bestimmt werden. In diesem Versuch waren hierzu drei Selbstauskünfte enthalten. Sowohl der erste Faktor, als auch die hohen Faktorladungen von allen drei Items beim übergeordneten Faktor sprechen jedoch für die überaus wichtige Bedeutung der Erfahrung für die Ausprägung der raumakustischen Hörexpertise. Erfahrung sollte deshalb so differenziert wie möglich erfasst werden. Die große

Bedeutung der Vorerfahrung stimmt mit den Beobachtugnen von Bech (1992) überein, die in Abschnitt 2.3.3 diskutiert wurden.

Es stellt sich außerdem die Frage, ob es noch weitere Subfaktoren gibt, die in diesem Versuch nicht gemessen werden konnten, da keine geeigneten Items enthalten waren. Diese These wird durch das relativ geringe McDonald's ω_t gestützt, das darauf hindeutet, dass dieser latente Faktor nicht besonders akkurat bestimmt werden konnte (vgl. Revelle und Zinbarg, 2009, S. 152).

Dem ersten Hörtest schien hier keine besondere Bedeutung zuzukommen, denn die Varianzaufklärung der Ergebnisse durch die Faktoren ist ausgesprochen niedrig. Das kann daran liegen,
dass der Test ein unzureichendes Messinstrument darstellt, oder aber ein Indiz dafür sein, dass
die physiologische Wahrnehmung, die hier gemessen werden soll, tatsächlich eine untergeordnete
Rolle spielt. Das scheint plausibel, da weder Wissen noch praktische Routine einen Einfluss auf
die Sensibilität des Ohrs als Sinnesorgan haben sollten. Auch die Fähigkeit, wahrgenommene
Zusammenhänge korrekt zu interpretieren, wird natürlich durch das Gehör eingeschränkt, kann
die Leistung desselben jedoch kaum verbessern.

In Abschnitt 2.3.3 wurden Kategorien von Prüfpersonen nach der DIN EN ISO 8586 (2014) vorgestellt und eine Person mit hoher raumakustischer Hörexpertise mit dem 'spezialisierten Sensoriker'verglichen. Auch wenn die Bestimmung der Hörexpertise offenbar noch Unzulänglichkeiten aufweist, konnte durch den übergeordneten Faktor der Analyse genau diese Kompetenz, einen sensorischen Eindruck fachlich korrekt zu bewerten, nachgewiesen werden.

Darüber hinaus findet sich im Faktor "Fachwissen" ein Äquivalent des Sachverständigen der Norm: eine ausgeprägte Fachkompetenz unabhängig von den sensorischen Fähigkeiten. Auf Basis dieser Beobachtung kann Fachwissen oder Interesse einer Person an dem Thema Raumakustik jedoch noch nicht als hinreichendes Eignungskriterium für die Teilnahme an Hörversuchen gelten, bei denen Raumklang gehört und verstanden werden muss. Die gemessene übergeordnete Kompetenz hingegen scheint hier einen zumindest geringfügigen Einfluss zu haben.

5.3. Vorschläge für ein adaptives Verfahren

Für die letzten beiden Hörtests könnten in Zukunft adaptive Verfahren durchgeführt werden. Der erste Test hat bei der späteren Extraktion der latenten Faktoren keine nennenswerte Rolle mehr gespielt und sollte daher entweder in seiner Methodik verbessert oder aus dem Test gestrichen werden. Beim zweiten Hörtest sollten für den Test mit variabler Nachhallklangfarbe andere Nachhallcharakteristiken simuliert werden.

Da bei adaptiven Verfahren die nachfolgenden Items von der Beantwortung der vorigen abhängen, stellt sich gerade beim dritten Test die Frage, ob schon eine korrekt eingestellte Reihenfolge der klanglichen Differenzen, oder nur eine korrekte Interpretation der Reihenfolge in diesem Zusammenhang als richtige Antwort gewertet werden sollte. Denkbar ist auch, beides separat abzufragen, also erst die Person aufzufordern, die Stimuli in eine Reihenfolge zu bringen und anschließend separat Raumgrößen zuzuordnen.

Sollten in den zweiten Hörtest noch weitere Modulationsparameter aufgenommen werden, wären hier erneut Kalibrierungsexperimente erforderlich. Ferner sollte, wie bereits erwähnt, die Auswahl der Instrumentalstücke überdacht werden. Auch für die bestehenden Parameter stellt sich außerdem die Frage, ob der Test nicht noch einmal mit Items wiederholt werden sollte, da derzeit pro Parameter nur etwa 50 Ergebnisse vorliegen. Dann könnte das adaptive Verfahren jeweils inital mit einem Item beginnen, bei dem die Erfolgsqouten noch im oberen Bereich liegen. Bei der variablen Nachhallzeit wäre hier etwa ein Modulationsgrad von 2,5 denkbar. Wie oft ein Schwierigkeitsgrad korrekt beantwortet werden sollte, um zum nächsthöheren zu gelangen,

kann in Zukunft erörtert werden. Wegen der zufälligen Anordnung der Instrumentalstimuli und Räume empfehlen sich jedoch zwei bis drei Durchgänge, da sich in jedem die Testsituation trotz konstanter Schwierigkeit ändert.

Für die praktische Umsetzung ist noch anzumerken, dass LimeSurvey als eher ungeeignet angesehen wird. Die Randomisierung von Items ist in dieser Umgebung sehr aufwändig. Die Implementierung adaptiver Versuchsverfahren ist hingegen kaum umsetzbar, da keine automatisierte Rückwärtsnavigation möglich ist, die bei adaptiven Verfahren jedoch essentiell ist, um im Falle einer falschen Antwort wieder zum Item mit einem niedrigeren Schwierigkeitsgrad "zurückzukehren".

6. Fazit und Ausblick

Zu Beginn dieser Arbeit wurde die Frage gestellt, ob sich raumakustische Hörexpertise als Fähigkeit messen lässt, Raumklang differenziert wahrzunehmen und die Ursachen bestimmter Klangeigenschaften zu verstehen. Dazu wurde ein Hörversuch durchgeführt, für den drei Testverfahren entwickelt und angewendet wurden. Diese Hörtests konnten offenbar die jeweils angestrebte Fähigkeit einer Person messen und lieferten im Fall des dritten Tests sogar zusätzliche Hinweise auf die Ursache unkorrekter Antworten. Es konnte eine latente Fähigkeit gemessen werden, die sich in ausgeprägtem Fachwissen, Erfahrung und der Fähigkeit, Klangeigenschaften von Schallquelle und Raum auseinanderzuhalten, äußert und offenbar dazu beiträgt, dass die Person dazu in der Lage ist, in einer gegebenen klanglichen Situation korrekte Aussagen über die räumlichen Gegebenheiten zu treffen.

Dieser latente Faktor unterteilt sich in die Komponeten Fachwissen und Erfahrung. Die beobachteten Zusammenhänge zwischen Wissen, Erfahrung und Interpretation von Raumklang sind naheliegend, hier ist es offenbar gelungen, die resultierende raumakustische Hörexpertise nicht nur festzustellen, sondern auch zu messen. In dieser Pilotstudie ging es vor allem darum, die Dimensionen dieser Expertise auszuloten, weswegen eine Vielzahl von Tests in einer entsprechend kurzen Fassung zu einem Versuch zusammengestellt wurde.

Es gibt Hinweise darauf, dass der Test in seiner jetzigen Form diese Expertise noch nicht sehr exakt misst, bzw. diverse Störeinflüsse die Ergebnisse beeinträchtigen. Weitere Forschung sollte daher der Optimierung dieses Tests gelten. Die beobachteten Faktoren und Konstrukte können differenzierter erfasst werden. Konkret sollte hierzu die Erfahrung einer Person genauer erfasst werden, die relevanten akustischen Parameter bei der kognitiven Abstraktion und Interpretation von Raumklang gezielt identifiziert werden und zuletzt Trainingseffekte durch wiederholte Durchführungen erforscht werden. Letztere könnten genutzt werden, um die raumakustische (oder auch generelle) Hörexpertise nicht nur zu erfassen, sondern explizit zu fördern und Personen so aktiv für Hörversuche zu qualifizieren. Zudem sollten weitere Testverfahren miteinbezogen werden, um die kognitive Analyse von Raumklang besser zu verstehen und weitere, mögliche Komponenten des Gesamtkonstrukts der raumakustischen Hörexpertise aufzudecken.

Die Ergebnisse der Hörtests bieten die Grundlage für adaptive Testverfahren und es wurden einige Vorschläge für deren Gestaltung gemacht. Die Implementierung würde den Rahmen dieser Arbeit überschreiten, zudem gibt es die genannten Indizien, dass die Testverfahren noch zusätzlich optimiert und erweitert werden können, bevor ein adaptiver Test konzipiert wird.

Literaturverzeichnis

- Backhaus, Klaus; Bernd Erichson; Wulff Pinke und Rolf Weiber (2016): Multivariate Analyse-methoden: Eine anwendungsorientierte Einführung. 14. Berlin: Springer-Verlag.
- Baker, Frank B. (2001): The basics of item response theory. 2. College Park, Maryland: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, Frank B. und Soeck-Ho Kim (2017): The Basics of Item Response Theory Using R. Statistics for Social and Behavioral Sciences. Cham (Schweiz): Springer.
- Bartholomew, David J. (1998): "Scaling unobservable construct in social science." In: *Applied Statistics*, 47(1) S. 1–13.
- Bech, Søren (1992): "Selection and Training of Subjects for Listening Tests on Sound Reproducing-Equipment." In: *Jornal of the Acoustic Society of America*, 40(7/8) S. 590–610.
- Bejar, Isaac I. (1990): "A Generative Analysis of a Three Dimensional Spatial Task." In: *Applied Psychological Measurement*, 14(3) S. 237–245.
- Belvins, Matthew G.; Adam T. Buck; Zhao Peng und Lily M. Wang (2013): "Quantifying the just noticeable difference of reverberation time with band-limited noise centered around 1000 Hz using a transformed up-down adaptive method." In: *Proceedings of the International Symposium on Room Acoustics*.
- Brant, Jeannine M. et al. (2011): "Symptom trajectories during chemotherapy in outpatients with lung cancer colorectal cancer, or lymphoma." In: *European Journal of Oncology Nursing*, 15(5).
- Brinkmann, Fabian et al. (2017): "The FABIAN head-related transfer-function data base."
- Brinkmann, Fabian et al. (2019): "A round robin on room acoustical simulation and auralization." In: *Jornal of the Acoustic Society of America*, 145(4).
- Browne, W. J.; S. V. Subramanian; K. Jones und H. Goldstein (2005): "Variance partitioning in multilevel lositic models that exhibit overdispersion." In: *Journal of the Royal Statistic Society*, 168(3) S. 599–613.
- Cabrera, Densil; Claudiu Pop und Jeong Daeup (2006): "Auditory Room Size Perception: A Comparison of Real versus Binarual Sound-Fields." In: *Proceedings of the 1st Australasian Acoustic Societies' Conference*. S. 417–422.
- Darwin, Christopher J. (2008): "Spatial Hearing and Perceiving Sources." In: Auditory Perception of Sound Sources. Springer-Verlag.
- de Boeck, Paul et al. (2011): "The Estimation of Item Response Models with the Imer Function from the Ime4 Package in R." In: *Journal of Statistical Software*, 39(12).
- de Vries, Diemer; Edo M. Huslebos und Jan Baan (2001): "Spatial fluctuations in measures for spaciousness." In: *Jornal of the Acoustic Society of America*, 110(2) S. 947–954.

- Devore, Sasha; Antje Ihlefeld; Kenneth Hancock; Barbara G. Shinn-Cunningham und Bertrand Delgutte (2009): "Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain." In: *Neuron*, 6(1) S. 123–134.
- Dickreiter, Michael; Volker Dittel; Wolgang Hoeg und Martin Wöhr (2009): Handbuch der Tonstudiottechnik, vol. 1. 7. Berlin: Walter de Gruyter.
- DIN 18041 (2016): "Hörsamkeit in Räumen Anforderungen, Empfehlungen und Hinweise für die Planung."
- DIN 45631 (1991): "Berechnug des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum."
- DIN EN ISO 3382-1 (2009): "Akustik Messung von Parametern der Raumakustik Teil 1: Aufführungsräume."
- DIN EN ISO 5492 (2009): "Sensorische Analyse Vokabular."
- DIN EN ISO 8586 (2014): "Sensorische Analyse- Allgemeiner Leitfaden für die Auswahl, Schulung und Überprüfung ausgewählter Prüfer und Sensoriker."
- Drasgow, Fritz und Robin I. Lissak (1983): "Modified Parallel Analysis: A Procedure for Examining the Latent Dimensionality f Dichotomously Scored Item Responses." In: *Journal of Applied Psychology*, 68(3) S. 363–373.
- Dunn, Thomas J.; Thom Baguley und Vivienne Brunsden (2014): "From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation." In: *British Journal of Psychology*, 105(1) S. 399–412.
- Ellermeier, Wolfgang; Jürgen Hellbrück; Armin Kohlrausch und Alfred Zeitler (2008): Kompendium zur Durchführung von Hörversuchen in Wissenschaft und industrieller Praxis. Berlin: Deutsche Gesellschaft für Akustik e.V.
- Fabrigar, Lenadre R.; Duane T. Wegener; Robert C. MacCallum und ErinJ. Strahan (1999): "Evaluating the Use of Exploratory Factor Analysis in Psychological Research." In: *Psychological Methods*, 4(3) S. 272–299.
- Fastl, Hugo und Eberhard Zwicker (2007): Psychoacoustics: Facts and Models. 3. Springer-Verlag Berlin-Heidelberg.
- Fleischmann, Felix; Andreas Silzle und Jan Plogsties (2012): "Identification and Evaluation of Target Curves for Headphones." In: Audio Engineering Society Convention Paper 8740.
- Florentine, Mary; Arthur N. Popper und Richard R. Fay (2011): Loudness, vol. 37 von Springer Handbook of Auditory Research. New York, NY: Springer Science+Business Media LLC.
- Fuchs, Helmut V. (2018): "Thesen zur Akustik anspruchsvoller Räume: Teil 1: Bass-Verhältnis und Hörsamkeit in größeren Räumlichkeiten." In: Akustik Journal, (2) S. 31–45.
- Grigoriev, Dimitry; David Ackermann; Sönke Pelzer und Stefan Weinzierl (2016): "Ein psychologisches Messinstrument für die akustische Wahrnehmung von Räumen für Musik und Sprache: Stimulus-Erzeugung." In: Deutsche Gesellschaft für Akustik e.V. (Hrg.) DAGA 2016.
- Hambleton, Ronald K. und van der Linden, Wim J. (1982): "Advances in Item Response Theory and Applications: An Introduction." In: *Applied Psychological Measurement*, 6(4) S. 373–378.

- Hameed, Sharaff; Pakarinen, Jyri Valde, Kari und Ville Pulkki (2004): "Psychoacoustic Cues in Room Size Perception." In: Audio Engineering Society Convention Paper 6084.
- Harman International Industries, Inc. (2020a): "K701 Reference class premium headphone." URL https://www.akg.com/Headphones/Professional%20Headphones/K701.html.
- Harman International Industries, Inc. (2020b): "K702 Reference Studio-Kopfhörer." URL https://de.akg.com/headphones/K702.html?cgid=headphones&dwvar_K702_color= Black-GLOBAL-Current.
- Harrison, Peter M. C. und Daniel Müllensiefen (2018): "Development and Validation of the Computerised Adaptive Beat Alignment Test (CA-BAT)." In: *Scientific Reports*, (8:12395).
- Hartig, Johannes und Andreas Frey (2013): "Sind Modelle der Item-Response-Theorie das "Mittel der Wahl" für die Modellierung von Kompetenzen?" In: Zeitschrift für Erziehungswissenschaft, 16(1) S. 47–51.
- Hochschule Düsseldorf (17.12.2019): "Labore." URL https://isave.hs-duesseldorf.de/profil/labore.
- Howard, Matt C. (2016): "A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices. What We Are Doing and How Can We Improve?" In: *International Journal of Human-Computer-Interaction*, 32(1) S. 51–62.
- Hu, Li-tze und Peter M. Bentler (1998): "Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification." In: *Psychological Methods*, 3(4) S. 424–453.
- James, Gareth; Daniela Witten; Trevor Hastie und Robert Tibshirani (2017): An Introduction to Statistical Learning with Applications in R. 8. Springer Science+Business Media.
- Kaspar, Roman (2009): Psychometrische Beurteilung verhaltensgestützer Schmerzassessments für Menschen mit Demenz: Potentiale von Item-Response-Theorie und Latent Variable Modellen am Beispiel der Verhaltensinventare CNPI und BESD. Dissertation, Ruprecht-Karls-Universität, Heidelberg.
- Lavandier, Mathieu und John F. Culling (2008): "Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer." In: *Jornal of the Acoustic Society of America*, 123(4) S. 2237–2248.
- Lehmann, P. und H. Wilkens (1980): "Zusammenhang subjektiver Beurteilung von Konzertsälen mit raumakustischen Kriterien." In: *Acta Acustica united with Acustica*, 45(4) S. 256–268.
- LimeSurvey GmbH (03.04.2019): "Fragetypen LimeSurvey Manual." URL https://manual. limesurvey.org/Question_types/de.
- LimeSurvey GmbH (11.02.2020): "New Template System in LS3.x." URL https://manual.limesurvey.org/New_Template_System_in_LS3.x.
- LimeSurvey GmbH (18.03.2019): "Umfragestruktur LimeSurvey Manual." URL https://manual.limesurvey.org/Survey_structure/de.
- Lokki, Tapio; Jukka Pätynen; Antti Kuusinen und Sakari Tervo (2012): "Disentangling preference ratings of concert hall acoustics using subjective sessory profiles." In: *Jornal of the Acoustic Society of America*, 132(5).

- Lokki, Tapio; Jukka Pätynen; Antti Kuusinen; Heikki Vertanen und Sakari Tervo (2011): "Concert hall assessment with individually elicited attributtes." In: *Jornal of the Acoustic Society of America*, 130(2).
- Luko, Alexis (2008): "Tinctoris on varietas." In: Early Music History, 27 S. 99–136.
- Maempel, Hans-Joachim (2012): "Experiments on audio-visual room perception: A methodological discussion." In: Deutsche Gesellschaft für Akustik e.V. (Hrg.) DAGA 2012. S. 783–784.
- Maempel, Hans-Joachim und Matthias Jentsch (2013): "Audio-visual interaction of size and distance perception in concert halls a preliminary study." In: *Proceedings of the International Symposium on Room Acoustics*.
- Magis, David und Gilles Raîche (2011): "catR: An R Pacakge for Computerized Adaptive Testing." In: Applied Psychological Measurement, 35(7) S. 576–577.
- Marsh, Herbert; Jon R. Balla und Roderick P. McDonald (1988): "Goodness-of-fit indexesin confirmatory facotr analysis: The effect of sample size." In: *Psycholigical Bulletin*, 103(3).
- Martelotta, Francesco (2010): "The just noticeable difference of center time and clarity index in large reverberant spaces." In: *Jornal of the Acoustic Society of America*, 128(2).
- Mayer, D. G. und D. G. Butler (1993): "Statistical validation." In: *Ecological Modeling*, 68(1) S. 21–32.
- McGarrigle, Ronan et al. (2014): "Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'." In: *International Journal of Audiology*, 53(7).
- Merimaa, Juha; V. Ralph Algazi und Richard O. Duda (2011): "Individual Perception of Headphone Reproduction Asymmetry." In: AES Convention Paper 8540.
- Møller, Henrik; Dorte Hammershøi; Clemen Boje Jensen und Michael Friis Sørensen (1995): "Transfer Characteristics of Headphones Measured on Human Ears." In: *Jornal of the Acoustic Society of America*, 43(4) S. 203–217.
- Moosbrugger, Helfried und Augustin Kelava (2012): Testtheorie und Fragebogenkonstruktion. 2. Berlin: Springer-Verlag GmbH.
- Möser, Michael (2005): Technische Akustik. 6. Berlin: Springer-Verlag Berlin Heidelberg.
- Mozilla (2005): "Window.localStorage Web API Referenz." URL https://developer.mozilla.org/de/docs/Web/API/Window/localStorage.
- Müllensiefen, Daniel; Bruno Gingras; Jason Musil und Lauren Stewart (2014): "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population." In: *PLoS ONE*, 9(2).
- Nagakawa, Shinichi und Holger Schielzeth (2013): "A general and simple method for obtaining \mathbb{R}^2 from generalized linear mixed-effects models." In: *Methods in Ecology and Evolution*, 4(1) S. 133–142.
- Neidhardt, Annika (2016): "Perception of the reverberation captured in a real room, depending on position and direction." In: *Proceedings of the 22nd Interational Congress on Acoustics*.

- O'Connor, Brian P (2000): "SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test." In: *Behavior Research Methods, Instruments, & Computers*, 32(3) S. 396–402.
- Papula, Lothar (2009): Mathematische Formelsammlung für Ingenieure und Naturwissenschaftler. 10. Wiesbaden: Vieweg +Teubner GWV Fachverlage GmbH.
- Paquier, Mathieu und Vincent Koehl (2010): "Audibility of headphone postioning variability." In: AES Convention Paper 8147. London, S. 1–9.
- Pelzer, Sönke; Lukas Aspöck; Dirk Schröder und Michael Vorländer (2014): "Integrating Real-Time Room Acoustics into a CAD Modeling Software to Enhance the Architectural Design Process." In: *Buildings*, 4(2) S. 113–138.
- Raychaudhuri, Samik (2008): "Introduction to Monte Carlo Simulation." In: S. J. Mason; R. R. Hill; L. Mönch; O. Rose; T. Jefferson und J. W. Fowler (Hrgs.) *Proceedings of the 2008 Winter Simulation Conference*.
- Rennie, Kimberly M. (1997): "Exploratory and Confirmatory Rotation Strategies in Exploratory Factor Analysis." In: *Proceedings of the Annual Meeting of the Southwest Educational Research Association*.
- Revelle, William (2019): Package ,psych'. URL https://personality-project.org/r/psych-manual.pdf.
- Revelle, William und Thomas Rocklin (1979): "Very simple structure: an alternative procedure for estimating the optimal number of interpretable factors." In: *Multivariate Behavourial Research*, 14(1) S. 403–414.
- Revelle, William und Richard E. Zinbarg (2009): "Coefficents alpha, beta, omega, and the glb: comments on Sijtsma." In: *Psychometrika*, 74(1) S. 145–154.
- Rijmen, Frank; Francis Tuerlinckx; Paul de Boeck und Peter Kuppens (2003): "A Nonlinear Mixed Model Framework for Item Response Theory." In: , 8(2) S. 185–205.
- Ripley, Brian et al. (2019): Package ,MASS'. URL http://www.stats.ox.ac.uk/pub/MASS4/.
- Rizopoulos, Dimitris (2006): "ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses." In: *Journal of Statistical Software*, 17(5).
- Schaal, Nora K. (2014): "Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe." In: *Musicae Scientiae*, 18(4) S. 423–447.
- Schmid, John und John M. Leiman (1957): "The development of hierarchical factor solutions." In: *Psychometrica*, 22(1) S. 53–61.
- Schmidt, Ulrich (2013): Professionelle Videotechnik: Grundlagen, Filmtechnik, Fernsehtechnik, Geräte und Studiotechnik in SD, HD, DI, 3D. 6. Berlin: Springer-Verlag.
- Schmitz, Katrin (2004): Standardisierung psychometrischer Diagnostik in der neurologischen Rehabilitation: Eine computergestützte neurosycholigische Testbatterie. Dissertaion, Universität Konstanz, Konstanz.
- Schröder, Dirk und Michael Vorländer (2011): "RAVEN: A real-time frameworkv for the auralization of interactive virtual environments." In: Forum Acusticum. S. 1541–1546.

- Schurz, Gerhard (2011): Einführung in die Wissenschaftstheorie. 3. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Scriven, Frances (2005): "Two types of sensory panels or are there more?" In: *Journal of Sensory Studies*, 20(1) S. 526–538.
- Shinn-Cunningham, Barbara G. (2003): "Acoustics and peception of sound in everyday environments." In: *Proceedings of the 3rd Int. Workshop on Spatial Media*.
- Skålevik, Magne (2010): "Reverberation Time The Mother of All Room Acoustic Parameters." In: Proceedings of 20th International Congress on Acoustics, ICA 2010.
- Slama, Michaël C. C. und Bertrand Delgutte (2015): "Neural Coding of Sound Envelope in Reverberant Environments." In: *The Journal of Neuroscience*, 35(10) S. 4452–4468.
- Soulodre, Gilbert A. und John S. Bradley (1995): "Subjective evaluation of new room acoustic measures." In: *Jornal of the Acoustic Society of America*, 98(1) S. 294–301.
- The MathWorks (1994): "MATLAB." URL https://www.mathworks.com/products/matlab. html.
- The R Foundation (1993): "R." URL https://www.r-project.org/.
- Trimble Inc. (2000): "SketchUp." URL https://www.sketchup.com/.
- von Dopr Schuitman, Jan; Diemer de Vries und Alexander Lindau (2013): "Derving contentspecific measures of room acoustic perception using a binaural, nonlinear auditory model." In: Jornal of the Acoustic Society of America, 133(3) S. 1572–1585.
- Weinzierl, Stefan (Hrg.) (2008): Handbuch der Audiokommunikation. 1. Springer-Verlag Berlin-Heidelberg.
- Weinzierl, Steffan; Steffen Lepa und David Ackermann (2018): "A measuring instrument for the auditory perception of rooms: The Room Acoustical Quality Inventory (RAQI)." In: *Jornal of the Acoustic Society of America*, 144(3) S. 1245–1257.
- Weinzierl, Stefan und Hans-Joachim Maempel (2012): "Sind Hörversuche subjektiv? Zur Objektivität akustischer Maße." In: Deutsche Gesellschaft für Akustik e.V. (Hrg.) DAGA 2012. S. 315–316.
- Weinzierl, Stefan und Michael Vorländer (2015): "Room Acoustical Parameters as Predictors of Room Acoustical Impression: What Do We Know and What Would We Like to Know?" In: *Acoustics Australia*, 43(1) S. 41–48.
- Wickelmaier, Florian und Sylvain Choisel (2005): "Selecting participants for listening tests of mulitchannel reproduced sound." In: AES Convention Paper.
- Wolff, Hans-Georg und Katja Preising (2005): "Exploring item and higher oder factor structure with the Schmid-Leiman solution: Syntax codes for SPSS and SAS." In: *Behaviour Research Methods*, 37(1) S. 48–58.
- Zacharov, Nick und Gaëtan Lorho (2006): "What are the requirements of a listening panel for evaluating spatial audio quality?-." In: *Proc. Int. Workshop on Spatial Audio and Sensory Evaluation Techniques*.
- Zhang, Zhiyong und Ke-Hai Yuan (2016): "Robust Coefficient Alpha and Omega and Conficende Intervals With Outlying Observations and Missing Data: Methods and Software." In: *Educational and Psychological Measurement*, 76(3) S. 387–411.

Anhang

A. Fragebogen

Hier sind alle Fragen und Textelemente aufgelistet, die für diesen Versuch erstellt und im Haupttext noch nicht aufgeführt wurden. Die Quellen der englischen und deutschen Versionen der Gold-MSI-Fragen sind im Abschnitt 3.3.2 referenziert. Eine Kopie der LimeSurvey-Datei des Fragebogens befindet sich in den digitalen Anlagen.

A.1. Allgemeine Informationen

A.2. Selbstauskunft: Erfahrung

		F	ragen							
akademisch	Ich hab	oe mich mit der Rau	makustik als akademis	sches Fach beschäftigt.						
	I have	I have dealt with room acoustics as part of my academic career.								
musikalisch	Ich hab	oe mich mit der Rau	makustik im Rahmen	einer künstlerischen/						
	musika	lischen Tätigkeit bes	schäftigt.							
	I have	I have dealt with room acoustics as part of my artistic/musical activites.								
Aufnahmen	Ich füh	Ich führe im beruflichen oder privaten Rahmen Musikaufnahmen durch,								
	oder ha	abe solche durchgefü	hrt.							
	I condu	uct or have conducte	d music recordings as	part of my professional						
	carrer	or free time.								
		Antv	vortskala							
noch nie	kaum	ab und zu	regelmäßig	sehr oft						
never	rarely	from time to time	on a regular basis	very often						

A.3. Fachfragen

q1: Sie befinden sich in einem halligen Raum und öffnen ein Fenster. Wie wirkt sich das auf die Nachhallzeit des Raumes aus?

You are in a reverberant room and open a window. How does this affect the reverberation time of the room?

- Die Nachhallzeit wird kleiner. The reverberation time decreases.
- Die Nachhallzeit wird größer.

 The reverberation time increases.
- Die Nachhallzeit bleibt gleich.

 The reverberation time does not change.
- Es gibt dann keinen Nachhall mehr. The reverberation disappears.

q2: Wie ändert sich der Schalldruckpegel eines Sprechers im Freien mit der Entfernung? How does the sound pressure level of a speaker in the open air change with increasing distance?

- Der Schalldruckpegel sinkt um 3 dB mit jeder Verdopplung des Abstands. Doubling the distance decreases the sound pressure level 3 dB.
- Der Schalldruckpegel sinkt um 6 dB mit jeder Verdopplung des Abstands. Doubling the distance decreases the sound pressure level 6 dB.
- Der Schalldruckpegel sinkt um 10 dB mit jeder Verdopplung des Abstands. Doubling the distance decreases the sound pressure level 10 dB.
- Der Schalldruckpegel sinkt um 12 dB mit jeder Verdopplung des Abstands. Doubling the distance decreases the sound pressure level 12 dB.

q3: Welche dieser Größen beeinflusst die Schallgeschwindigkeit? Which of these parameters has an influence on the sound propagation speed?

- Die Umgebungstemperatur.

 The temperature of the surrounding environment.
- Die Lautstärke der Schallquelle.

 The sound pressure level emitted by the sound source.
- Die Größe der Schallquelle. The size of the sound source.
- Der statische Luftdruck. The static air pressure.

q4: Wo tritt der höchste Schalldruckpegel für alle stehenden Wellen auf, die sich im Raum ausbreiten?

Where is the location of the maximum sound pressure for all standing waves that can appear in a certain room?

- In der Mitte des Raumes.

 At the center of the room.
- An den Wänden des Raumes. At the walls of the room.
- Eine halbe Wellenlänge von den Wänden entfernt. Half a wavelength away from the walls.
- Eine ganze Wellenlänge von den Wänden entfernt. A whole wavelength away from the walls.

q5: Die Nachhallzeit ist definiert als The reverberation time is defined as

- die wahrgenommene Dauer des Nachhalls im Raum. the perceived duration of the reverberation in a room.
- die Zeit, in der der Schalldruckpegel im Raum nach Abschalten einer Schallquelle um 60 dB abfällt

the time it takes until the sound pressure level has decreased by 60 dB after the sound source has been turned off.

- der Schalldruckpegel, um den der Schalldruck in einer Sekunde abfällt. the decrease of sound pressure level per second.
- die Zeit, bis der Schalldruckpegel im Raum nach Abschalten einer Schallquelle nicht mehr messbar ist.

the time until the sound pressure level cannot be measured anymore after the sound source has been turned off.

q6: Sie wollen mit einem Vorhang die Nachhallzeit im Raum bei einer bestimmten Frequenz senken. Wo ist die optimale Position für den Vorhang?

You want to use a curtain to decrease the reverberation time in a room at a certain frequency. Which is the optimal position for this curtain?

- Unmittelbar an der Wand. Closely in front of the walls.
- In einem Abstand zur Wand, welcher der halben Wellenlänge entspricht. Half a wavelength away from the walls.
- An einer Position, an der Schallschnelle und Schalldruck gleich groß sind. At a spot of equal sound pressure and sound particle velocity.
- In einem Abstand zur Wand, welcher einem Viertel der Wellenlänge entspricht. One quarter of the wavelength away from the walls.

q7: Ein Lochplattenabsorber hat seine maximale Absorptionwirkung bei einer bestimmmten Frequenz. Welches physikalische Phänomen liegt diesem Verhalten zugrunde? A perforated plate shows its maximum absorption efficiency at a certain frequency. Which physical phenomenon is the reason for that?

- Beugung. Diffraction.
- Streuung. Scattering.
- Resonanz.

 Resonance.
- Dissonanz. Dissonance.

q8: Der eben hörbare Schalldruck für einen Sinuston bei 1 kHz liegt f $\tilde{A}\frac{1}{4}$ r normal hörende Personen ungefähr bei

For a person with normal hearing, the just audible sound pressure is at approximately

- 20 Kilopascal. 20 kilopascal.
- 20 Pascal. 20 pascal.
- 20 Milipascal. 20 milipascal.
- 20 Mikropascal. 20 micropascal.

A.4. Textelemente - Deutsch

Willkommenstext

Herzlich Willkommen zu dieser Umfrage und vielen Dank, dass Sie sich die Zeit für die Teilnahme nehmen möchten! Ziel der Untersuchung ist es, die raumakustische Hörexpertise einer Person einschätzen zu können. Dazu werden sowohl Hörtests durchgeführt, als auch Fragen zur fachlichen Kenntnis und zur persönlichen musikalischen Erfahrung gestellt. Um einer Ermüdung Ihrer Ohren entgegenzuwirken, werden Höraufgaben und Selbstauskunft-Fragen abwechselnd präsentiert.

BITTE BEACHTEN SIE: Einige der Fragen und Hörtests behandeln spezielle Themen der Raumakustik. Es ist jedoch ausdrücklich **KEIN** Vorwissen oder fachliche Bildung in diesem Gebiet erforderlich, um an der Umfrage teilzunehmen! Es wird lediglich ein einwandfreies Gehör ohne bekannte Hörstörungen benötigt. Sollten Sie bei einer Frage keine Antwort wissen, wählen Sie bitte aus dem Bauch heraus eine der möglichen Antworten.

BITTE ÄNDERN SIE AUßERDEM NICHT DIE VOREINGESTELLTE ABHÖRLAUTSTÄRKE!

Dadurch würden Ergebnisse verfälscht.

Bei den Hörtests gibt es jeweils mehrere Durchläufe des gleichen Tests - auch wenn das Erscheinungsbild jeweils gleich bleibt, gibt es jeweils Veränderungen in den Klangbeispielen, sodass sich die Schwierigkeit mit jedem Durchlauf ändert.

Die erhobenen Daten werden ausschließlich für meine Masterarbeit am Fachgebiet Audiokommunikation der Technischen Universität Berlin verwendet.

Einführung: 1. Hörtest

Beim folgenden Hörtest wird Ihnen einerseits die hier gezeigte Skala von 21 Vergleichsklängen präsentiert. Durch Klicken der roten Schaltflächen lässt sich die Audiowiedergabe der jeweiligen Klänge starten und stoppen. Es handelt sich immer um das gleiche kurze Instrumentalstück, die Vergleichsklänge sind auf der Skala jedoch so angeordnet, dass ein bestimmtes raumakustisches Merkmal (z. B. die Dauer des Nachhalls) von links nach rechts zunimmt. Hören Sie sich exemplarisch ein paar Klänge auf dieser Skala an, um sich mit dem Prinzip vertraut zu machen Zusätzlich zu dieser Skala wird Ihnen im Versuch ein Referenzklang präsentiert, der über eine



zusätzliche Schaltfläche angehört werden kann. Diese Referenz entspricht immer genau einem Klang auf der Skala: Ihre Aufgabe ist es, anzugeben, welche Stufe auf der Skala das ist.

Dieser Test besteht aus 5 Durchläufen, es geht jedoch immer um das gleiche akustische Merkmal. Nach jedem Durchlauf wird der Wertebereich der Skala zufällig nach links oder rechts verschoben, sodass z.B. nicht mehr Nachhallzeiten von 1,5-2 Sekunden, sondern Nachhallzeiten von 1,7-2,2 Sekunden abgebildet werden.

Ändern Sie bitte nicht die voreingestellte Abhörlaustärke!

Einführung: 2. Hörtest

Im folgenden Hörtest werden Ihnen jeweils drei verschiedene Klänge präsentiert: A, B und X. A und B haben einen unterschiedlichen Raumklang, der von X entspricht entweder dem von A oder von B. Ihre Aufgabe ist es, zu erkennen, ob entweder A und X oder B und X aus dem gleichen Raum stammen. Hören Sie sich das Beispiel an: die Klänge A und X stammen aus einer wesentlich halligeren Umgebung als der Klang B. Wie bei den bisherigen Versuchen können Sie durch Klicken der jeweiligen Schalflächen die Wiedergabe starten und stoppen. Es gibt 20



Durchläufe bei diesem Versuch, in jedem Durchlauf werden die drei Instrumentalklänge und die Raumklänge zufällig neu zugeordnet.

BITTE BEACHTEN SIE: Anders als im Beispiel können Sie sich bei diesem Verfahren jeden Klang **NUR EINMAL** anhören. Seien Sie daher bitte von Anfang an aufmerksam!

Ändern Sie bitte nicht die voreingestellte Abhörlaustärke!

Einführung: Fachfragen

Vor dem letzten Hörtest folgen nun ein paar inhaltliche Fragen zur Raumakustik. Einige Aufgaben fragen spezifische, physikalische Sachverhalte ab. Sollten Sie daher die Antwort auf eine Frage nicht kennen, entscheiden Sie sich einfach aus dem Bauch heraus für eine der möglichen Antwortoptionen.

Einführung: 3. Hörtest

Im letzten Hörversuch ist es Ihre Aufgabe, drei Räume der Größe nach zu sortieren. Die Räume haben zwar die gleiche Nachhallzeit, unterscheiden sich aber im Raumvolumen. Äceber die Schaltflächen "Raumklang anhören"können Sie sich die jeweiligen Raumklänge an einem Beispiel anhören und vergleichen (hier ist wieder Mehrfachwiedergabe der Klänge möglich).

Von diesem Test gibt es 14 Wiederholungen, in denen die Räume jeweils neu angeordnet werden.

Ändern Sie bitte nicht die voreingestellte Abhörlaustärke!

A.5. Textelemente - Englisch

Willkommenstext

Welcome to this survey and thank you for taking the time to participate! The goal of this research is to learn about people's listening expertise in room acoustics. The survey consists of listening tests, as well as questions regarding your professional knowledge in room acoustics and your general musical sophistication. To prevent ear fatigue, the listening tests and self-report-questions appear in a mixed order.

PLEASE NOTE: some of the questions and tests treat specific aspects of room acoustics. Yet, there is **NO** professional or any other experience in this field needed to participate in this survey. All you need are healthy ears with no known hearing-disability. If you do not know the answer to a question, please pick spontaneously one of the possible options.

PLEASE DO NOT CHANGE THE PLAYBACK VOLUME WHILE TAKING THE SURVEY!

This would corrupt the results.

In the listening tests, there will always be several repititions of each test - although the visual appearance remains the same, the audio examples used change in a certain manner changing the difficulty of the taks in each repitition.

The data collected will be used for my master thesis at the Audiocommunication Group of the Technische Universität Berlin only.

Einführung: 1. Hörtest

In the following listening test you will be presented the scale of 21 sounds shown below. By clicking the red buttons you can start and stop the audio playback of each sound. It is always the same instrumental melody but the sounds are ordered by a certain room acoustic property (e.g. the reverberation time) increasing from left to right. Please listen to a few sounds on this scale to get familiar with the principle. In addition to this scale, there will always be a reference



sound which can be listened to by another red button labeled 'Reference'. This reference is always the same as one specific sound on the scale: your task is to find out which sound on the scale matches the reference.

There are five repititions of this test, all of them dealing with the same acoustic property. After each repitition, the range represented by the scale is randomly shifted, e.g. changing the range of reverberation times from 1.5 seconds to 2 seconds to a range of 1.7 seconds to 2.2 seconds.

Please do not change the playback volume!

Einführung: 2. Hörtest

In the following listening test you will listen to three different sounds: A, B and X. A and B have different reverberation, while the reverberation in X is either the same as in A or B. Your task is to recognize in which two sounds the same reverberation is being used: A and X or B and X. Listen to the example sounds given below: the sounds A and X represent a significantly more reverberant environemt as the sound B. As in the previous tests, you can start and stop the audio playback of each sound by clicking the according button. This test consists of



20 repititions. In each repitition, the three instrumental sounds and the different reverberations will be reordered randomly.

PLEASE NOTE: Unlike in the given example, in these tests you can listen to the sounds **ONLY ONCE**. Thus, please be attentive from the beginning!

Please do not change the playback volume!

Einführung: Fragen

Before the last listening test, you will now be presented a few questions about room acoustics. Some of these questions deal with specific physical facts. Thus, if you do not know the answer to a question, just pick spontaneously one of the possible answers.

Einführung: Fragen

In this final listening test you have to order three rooms by their size. All four rooms have the same reverberation time but different sizes. By clicking on the buttons "Listen to Room" you can listen to sound examples for each room and compare the rooms' acoustics (in this test, mulitple playback is possible).

There will be 14 repititions of this test and after each, the four rooms will be reordered randomly.

Please do not change the playback volume!

Einwillgungserklärung (nur in englischer Version)

Information sheet for the participants at Goldsmiths, University of London

You have been invited to take part in a research study at Goldsmiths University of London. Before you decide if you want to participate, please take some time to read the following information.

You must be at least 18 years old to participate in this study. Participation is completely voluntary. If you decide to take part, you will be asked to tick a box on the consent page and you will be free to withdraw at any time without giving any reason. All data is collected anonymously, and you will not be asked for any information that could be used to personally identify you. In addition, all information obtained throughout this study will be kept strictly confidential.

This study is investigating how well different people are able to distinguish between different rooms used for audio recordings. You will be presented with three different tasks. On the first two tasks you will be asked to listen to short music excerpts recorded in different rooms and to find the room that matches a target recording. The third task asks you to sort three short audio excerpts according to the size of the room they were recorded in. In addition, you will be asked to fill out a short questionnaire on your musical background as well as on your previous experience and expertise with room acoustics.

In total, the study should take no longer than 40 minutes and you are free to take breaks between any of the parts of the study. If, however, you do experience any psychological distress then please don't hesitate to alert the research assistant.

This study has been approved by Goldsmiths Ethics Committee.

If you have any questions or would like any further information about the study, you can get in touch with Prof Daniel Müllensiefen (d.mullensiefen@gold.ac.uk)

Thank you for your participation!

B. Raumkonfiguration

	Nach	nhallzei	it-Refe	renzen	[%]		Kla	ngfarbe	en-Refe	erenzen	[%]
$T_{60} [s]$	50	75	110	125	150	T_{60}/T_{lin}	125	150	175	200	225
0,600	80	53	36			100	80	67			
0,675	90	60	41			105	84	70			
0,750	100	67	45	40		110	88	73	63		
0,825	110	73	50	44		115	92	77	66		
0,900	120	80	55	48		120	96	80	69		
0,975	130	87	59	52		125	100	83	71		
1,050	140	93	64	56		130	104	87	74		
1,125	150	100	68	60		135	108	90	77		
1,200	160	107	73	64	53	140	112	93	80	70	
1,275	170	113	77	68	57	145	116	97	83	73	
1,350	180	120	82	72	60	150	120	100	86	75	
1,425	190	127	86	76	63	155	124	103	89	78	
1,500	200	133	91	80	67	160	128	107	91	80	
1,575	210	140	95	84	70	165	132	110	94	83	
1,650	220	147	100	88	73	170	136	113	97	85	76
1,725	230	153	105	92	77	175	140	117	100	88	78
1,800	240	160	109	96	80	180	144	120	103	90	80
1,875	250	167	114	100	83	185	148	123	106	93	82
1,950	260	173	118	104	87	190	152	127	109	95	84
2,025	270	180	123	108	90	195	158	130	111	98	87
2,100	280	187	127	112	93	200	160	133	114	100	89
2,175		193	132	116	97	205		137	117	103	91
2,250		200	136	120	100	210		140	120	105	93
2,325		207	141	124	103	215		143	123	108	96
2,400		213	145	128	107	220		147	126	110	98
2,475			150	132	110	225			129	113	100
2,550			155	136	113	230			131	115	102
2,625			159	140	117	235			134	118	104
2,700			164	144	120	240			137	120	107
2,775				148	123	245				123	109
2,850				152	127	250				125	111
2,925				156	130	255				128	113
3,000				160	133	260				130	116
						265					118
						270					120
						275					122
$\Delta T_{60} \ [\%]$	10,00	6,67	4,54	4,00	3,33		4,00	3,33	2,85	2,50	2,22

Tabelle 1.: Skalenstufen im ersten Hörtest. Die Linien markieren die Anfangs- und Endpunkte der Skalen. Die Spalten repräsentieren jeweils eine Referenz. Meist gab es zwei Skalen zur Auswahl, die sich teilweise überlagern. Die Schrittweiten jeder Skala sind dargestellt und die Skalenstufen in Abhängigkeit der jeweiligen Referenz angegeben.

Mod.	B x L x H [m]	$V [m^3]$	$B \times L \times H [m]$	$V [m^3]$			
	Referenz						
	10,00 x 12,50 x 4	500					
Item	kleiner		größer $[m^3]$				
0,5	$9,81 \times 12,26 \times 3,92$	472	10,19 x 12.74 x 4.08	530			
1	$9,62 \times 12,02 \times 3,84$	445	$10,39 \times 12,99 \times 4,15$	561			
1,5	$9,44 \times 11,80 \times 3,76$	420	$10,59 \times 13,24 \times 4,24$	595			
2	$9,26 \times 11,57 \times 3,70$	396	$10.80 \times 13.50 \times 4.32$	630			
3	$8,91 \times 11,14 \times 3,56$	353	11.22 x 14.03 x 4.49	707			
4	$8,57 \times 10,72 \times 3,43$	314	11.67 x 14.58 x 4.67	794			
5	$8,25 \times 10,31 \times 3,30$	281	12.12 x 15.15 x 4.85	890			
6	$7,94 \times 9,92 \times 3,17$	250	12.60 x 15.75 x 5.04	1000			
7	$7,64 \times 9,55 \times 3,05$	223	$13.09 \times 16.37 \times 5.24$	1122			
8	$7,35 \times 9,19 \times 2,94$	198	13.61 x 17.01 x 5.44	1260			
9	$7,07 \times 8,84 \times 2,83$	177	14.14 x 17.68 x 5.66	1414			
10	$6,80 \times 8,50 \times 2,72$	157	14.70 x 18.37 x 5.88	1587			
11	$6,55 \times 8,18 \times 2,62$	140	15.27 x 19.09 x 6.11	1782			
12	$6,30 \times 7,87 \times 2,52$	125	15.87 x 19.84 x 6.35	2000			

Tabelle 2.: Raumvolumina und Abmessungen (Breite, Länge, Höhe) im dritten Hörtest. Der Modulationsfaktor (Mod.) wird in die Gleichungen 3.1 und 3.2 eingesetzt

C. Auswertung

2. H	örtest (A)	2. H	örtest (1	B)	3	. Hörtes	t	Fa	chfrage	n
$T/T_{1,5}$	χ^2	p_{χ^2}	T/T_{lin}	χ^2	p_{χ^2}	Mod.	χ^2	p_{χ^2}	Rang	χ^2	p_{χ^2}
1,1	7,87	0,62	1,1	11,00	0,65	1	27,08	0,03	1	8,93	0,68
1,3	2,82	0,99	1,2	$5,\!51$	0,80	2	6,26	$0,\!82$	2	20,92	0,06
1,5	3,50	0,97	1,3	6,04	0,71	3	7,57	0,70	3	21,74	0,07
1,6	6,49	0,76	1,4	1,94	0,98	4	5,81	0,84	4	9,01	0,75
1,7	8,05	0,64	1,8	$7,\!67$	$0,\!82$	5	4,42	0,93	5	$4,\!64$	0,97
1,8	13,98	0,11	1,9	$15,\!27$	$0,\!10$	6	11,61	$0,\!29$	6	18,20	0,18
1,9	17,16	0,03	2,0	9,03	$0,\!32$	7	10,29	$0,\!27$	7	$12,\!47$	0,77
2,1	4,18	0,93	2,1	$15,\!67$	$0,\!55$	8	15,20	0,06	8	$33,\!58$	0,05
2,2	12,06	0,17	2,2	10,36	$0,\!33$	9	5,29	0,79			
2,3	7,61	$0,\!53$	2,3	$13,\!12$	$0,\!14$	10	2,83	0,97			
2,4	7,48	$0,\!49$	2,4	6,73	$0,\!60$	11	16,44	0,06			
2,5	3,57	0,90	2,5	$6,\!22$	$0,\!64$	12	10,42	$0,\!23$			
2,6	3,67	0,90	2,6	11,78	$0,\!15$						
2,7	17,94	0,03	2,7	10,14	$0,\!23$						
2,8	7,70	$0,\!44$	2,8	3,09	0,92						
2,9	10,03	$0,\!28$	2,9	6,10	0,71						
3,0	8,99	$0,\!41$	3,0	$13,\!31$	0,11						

Tabelle 3.: Test der Items durch den Vergleich mit einer Monte-Carlo-Simulation. Signifikante Abweichungen zeigen eine schlechte Anpassung. Der Modulationsfaktor (Mod.) im dritten Hörtest wird in die Gleichungen 3.1 und 3.2 eingesetzt.

	feste	e Effekte	е		Mo	dell				
Parameter	В	σ	B/σ	$P(>B/\sigma)$	AIC	BIC				
2. Hörtest: Nachhallzeit										
Konstantterm	-4,074	0,819	-4,969	> 0,000	908,3	922,7				
Nachhallzeit	2,032	0,353	5,753	> 0,000						
2. Hörtest: Nac	hhallklar	ngfarbe								
Konstantterm	-5,876	1,519	-3,868	> 0,000	878,2	892,0				
Klangfarbe	2,211	0,574	3,849	> 0,000						
3. Hörtest	•		,							
Konstantterm	-2,414	0,321	-7,522	> 0,000	1334,9	1350,4				
Volumen	1,530	0,139	10,988	> 0,000						
Fachfragen										
Konstantterm	-3,100	0,486	-6,376	> 0,000	986,0	1000,1				
Rang	0,519	0,080	6,511	> 0,000						

Tabelle 4.: Modellparamter der GLMMs. Dargestellt sind für die festen Effekte die Regressionskoeffizienten B mit Standardabweichung σ und das Ergebnis eines t-Tests, ob der Regressionskoeffizient unter Berücksichtigung der Standardabweichung einen signifikanten Effekt ergibt. Zudem sind AIC und BIC des jeweiligen Modells angegeben.

	Nachhal	lzeit		Nachhallklangfarbe					
			Faktorla	adungen					
Variable	MSA	F	Kom.	Variable	MSA	F	Kom.		
Ref050	0,64	0,50	0,25	Ref125	0,71	0,73	0,53		
Ref075	0,64	0,76	$0,\!57$	Ref150	0,68	0,53	$0,\!28$		
Ref110	$0,\!57$	0,09	0,01	Ref175	0,78	0,64	$0,\!41$		
Ref125	0,71	0,56	$0,\!32$	Ref200	0,72	0,69	$0,\!47$		
Ref150	0,68	0,41	$0,\!17$	Ref225	0,80	0,46	$0,\!21$		
Erklärte '	Varianz	0	,26	Erklärte '	Varianz	0	,41		
KMO		0	,66	KMO		0	,74		
]	Modellp	arameter					
		χ^2	p_{χ^2}			χ^2	p_{χ^2}		
Chi-Quad	lrat	2,04	0,84	Chi-Quad	lrat	6,28	0,28		
RMSEA		0		RMSEA		0,07			
TLI		1	,35	TLI		1			

Tabelle 5.: Ergebnisse der Faktoranalysen im ersten Hörtest. Angegeben sind die individuellen MSA und das KMO, die Faktorladungen (F) und Kommunalitäten (Kom.), die erklärte Varianz, der Chi-Qaudrat-Test, der RMSEA und der Tucker-Lewis-Index (TLI)

Variable	F1	F2	F3	Kom
1. Hörtest	-0,16	-0,24	0,14	0,10
2. Hörtest	0,39	0,21	$0,\!22$	0,76
Quiz	,22	$0,\!64$	0,09	$0,\!46$
Gold-MSI: Wahrnehmung	0,23	0,07	0,97	1,00
Gold-MSI: Umgang	0,47	$0,\!12$	$0,\!23$	0,28
Gold-MSI: Training	0,30	$0,\!30$	0,07	0,19
Erfahrung: akademisch	0,17	0,97	$0,\!12$	1,00
Erfahrung: musikalisch	0,70	$0,\!40$	0,07	$0,\!66$
Erfahrung: Aufnahme	0,87	0,18	0,00	0,79

Tabelle 6.: Faktorladungen der orthogonalen Faktoren vor der Anwendung der Schmid-Leiman-Transformation. Neben den Faktorladungen sind auch die Kommunalitäten (Kom.) angegeben.