Technische Universität Berlin
Fakultät I

Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

# Untersuchung des Zusammenhangs physikalischer Maße zur Bestimmung der *mixing time* in Raumimpulsantworten und ihrem perzeptiven Äquivalent

Magisterarbeit

vorgelegt von

Linda Kosanke

[redacted]

01.11.2010

Erstgutachter: Prof. Dr. Stefan Weinzierl
Zweitgutachter: Alexander Lindau M.A.

# Eidesstattliche Erklärung

Die selbstständige Anfertigung dieser Arbeit versichere ich an Eides statt.

Berlin, den 01. November 2010

---

Linda Kosanke

# Zusammenfassung

Die dynamische Binauralsynthese erfordert große Datensätze binauraler Raumimpulsantworten. Der Nachhall der Impulsantworten verhält sich nach einer bestimmten Zeit stochastisch, die zeitliche Feinstruktur kann zunehmend schlechter erkannt werden. Nach Ablauf einer gewissen Zeit kann der Nachhall unabhängig vom Ort gewählt werden. Damit ist eine Reduktion sowohl des Datensatzes als auch des Echtzeitrechenaufwandes möglich, ohne einen Qualitätsverlust hinnehmen zu müssen. Ziel der Arbeit war es, den Zeitpunkt zu bestimmen, ab dem eine Manipulation des Nachhalls nicht mehr wahrnehmbar ist. Dazu wurden in neun Räumen mit systematisch variierten mittleren Absorptionsgraden und Volumina binaurale Impulsantworten gemessen. In einem 3-AFC-Schwellwertverfahren wurde untersucht, ab welchem Zeitpunkt die Überblendung des frühen Anteils einer binauralen Raumimpulsantwort in einen diffusen Nachhall einer anderen binauralen Raumimpulsantwort mit unterschiedlicher Einfallsrichtung nicht mehr wahrnehmbar ist. Die Ergebnisse des Schwellwertversuches wurden deskriptiv und inferenzstatistisch ausgewertet. Zusammenhänge mit modellbasierten und empirischen Prädiktoren wurden mittels Regressionsrechnung untersucht.

# Inhaltsverzeichnis


**1 Einleitung** **1**

**2 Stand der Forschung** **6**

2.1 Zur Diffusität in Schallfeldern: Theoretische Grundlagen . . . . . 6

2.2 Modellbasierte Prädiktoren von Diffusität . . . . . 10

2.3 Empirische Prädiktoren von Diffusität . . . . . 14

2.3.1 Abel und Huang, 2006 . . . . . 15

2.3.2 Stewart und Sandler, 2007 . . . . . 17

2.3.3 Defrance u. a., 2009 . . . . . 19

2.3.4 Hidaka u. a., 2007 . . . . . 22

2.4 Perzeptive Evaluation . . . . . 24

**3 Methodik** **27**

3.1 Parameter der Raumauswahl . . . . . 27

3.2 Messdatenerhebung . . . . . 32

3.3 Ergebnisse aus dem Umgang mit empirischen Prädiktoren . . . . . 36

3.3.1 Abel und Huang, 2006 . . . . . 36

3.3.2 Stewart und Sandler, 2007 . . . . . 39

3.3.3 Defrance u. a., 2009 . . . . . 42

3.3.4 Hidaka u. a., 2007 . . . . . 43

3.4 Hörversuch . . . . . 45

3.4.1 Vorversuche . . . . . 45

3.4.2 Technischer Aufbau . . . . . 49

3.4.3 Versuchsdesign . . . . . 50

**4 Hörversuchsergebnisse** **55**

4.1 Stichprobenumfang und -beschreibung . . . . . 55

4.2 Deskriptive Auswertung . . . . . 57
4.3 Inferenzstatistische Auswertung . . . . . 62
4.4 Fehlerquellen . . . . . 63

**5 Regressionsanalyse 65**
5.1 Regression modellbasierter Prädiktoren . . . . . 68
5.1.1 Mittelwerte . . . . . 68
5.1.2 95%-Perzentile . . . . . 72
5.2 Regression empirischer Prädiktoren . . . . . 74
5.2.1 Mittelwerte . . . . . 75
5.2.2 95%-Perzentile . . . . . 79
5.3 Zusammenfassung . . . . . 82

**6 Zusammenfassung und Ausblick 85**

**Literatur 88**

**Abbildungsverzeichnis 93**

**Tabellenverzeichnis 95**

**A Quellcodes 96**
A.1 modellbasierte Prädiktoren . . . . . 96
A.2 empirische Prädiktoren . . . . . 97

**B Raumdaten 102**

**C Statistische Daten 103**
C.1 Deskriptive und inferenzstatistische Auswertung . . . . . 103
C.2 Regressionsanalyse . . . . . 105
C.2.1 Regressionsmodelle . . . . . 105
C.2.2 Signifikanz der Anstiege der Regressionsgeraden . . . . . 119

**D Fragebogen und Versuchsanleitung 120**

# Häufig verwendete Abkürzungen und Variablen

| | |
|---|---|
| 3-AFC | 3-Alternative-Forced-Choice |
| $A$ | äquivalente Absorptionsfläche |
| $\alpha$ | Absorptionsgrad |
| ANOVA | Varianzanalyse (analysis of variance) |
| BRIR | binaurale Raumimpulsantwort (binaural room impulse response) |
| $c$ | Schallgeschwindigkeit |
| $E$ | Effektgröße |
| $E(t, \omega)$ | zeit- und frequenzabhängige Energie |
| $f$ | Frequenz |
| FFT | schnelle Fouriertransformation (fast fourier transformation) |
| $H(\omega)$ | Übertragungsfunktion |
| HRIR | kopfbezogene Impulsantwort (head-related impulse response) |
| HRFT | Außenohrübertragungsfunktion (head-related transfer function) |
| IR, $h(t)$ | Impulsantwort (impulse response) |
| KI | Konfidenzintervall |
| $l_m$ | mittlere freie Weglänge |
| $\frac{dN}{dt}$ | Reflexionsdichte |
| $R^2$ | Varianzaufklärung |
| $S$ | Oberfläche |
| $t$ | Zeit |
| $T$ | Nachhallzeit |
| $t_m$ | physikalische mixing time |
| $t_{mp}$ | perzeptive mixing time |
| $t_{mp50\%}$ | geschätzte perzeptive mixing time auf Grundlage von Mittelwerten |
| $t_{mp95\%}$ | geschätzte perzeptive mixing time auf Grundlage von 95 %-Perzentilen |
| $V$ | Volumen |
| VP | Versuchsperson |

# 1 Einleitung

Zur Simulation akustischer Umgebungen existieren mehrere Verfahren, mit denen Schallfelder entweder in einem gegebenen Raumvolumen oder direkt am Ohr des Hörers realisiert werden. Im Rahmen dieser Arbeit wird die dynamische Binauralsynthese zur dreidimensionalen Wiedergabe von Schallquellen betrachtet. Das Schallfeld wird in diesem Fall kopfbezogen, d. h. direkt am Ohr des Hörers, simuliert.
Bei der Binauralsynthese wird die Tatsache genutzt, dass die Schallübertragung in Räumen zwischen Sender und Empfänger grundsätzlich ein lineares zeitinvariantes System bildet. Ein System dieser Art lässt sich vollständig durch seine Impulsantwort (IR, impulse response) charakterisieren, die die Reaktion des Systems auf einen unendlich kurzen Dirac-Stoß darstellt. Der Ausgang $y(t)$ eines solchen Systems lässt sich durch die Faltung des Eingangssignals $x(t)$ mit der Impulsantwort $h(t)$ berechnen:

$$y(t) = \int_{-\infty}^{\infty} x(\tau) \cdot h(t-\tau) d\tau = x(t) * h(t) \tag{1.1}$$

Die Faltung im Zeitbereich entspricht einer Multiplikation im Frequenzbereich. Die Übertragungsfunktion $H(\omega)$[1], die das System im Frequenzbereich beschreibt, wird mit dem Spektrum des Eingangssignals $X(\omega)$ multipliziert, um das Spektrum des Ausgangssignals zu erhalten:

$$Y(\omega) = X(\omega) \cdot H(\omega) \tag{1.2}$$

Zur Auralisation eines solchen Systems kann demnach ein beliebiges Eingangssignal mit der Impulsantwort gefaltet werden. Mit Hilfe dieses Prinzips kann die Schallausbreitung

[1] Die Übertragungsfunktion ist nichts anderes als die fouriertransformierte Impulsantwort.

zwischen Quelle und Empfänger in beliebigen Räumen mit einem beliebigen Quellsignal simuliert werden.

Die Binauralsynthese setzt ein Impulsantwortpaar voraus, das aus einer linken und einer rechten Impulsantwort besteht. Diese Impulsantworten enthalten alle Informationen, die durch Abschattung, Beugung, Reflexionen und Resonanzen an Torso, Kopf, Schultern und Ohrmuscheln abgeleitet werden können (Weinzierl, 2008, S. 586).
Die Impulsantworten werden mit Hilfe von Mikrofonen, die sich am Eingang des Ohrkanals befinden, gemessen. Der Ohrkanal ist meist Teil eines Kunstkopfes, der dem Kopf eines Menschen nachempfunden ist. Die Mikrofone können aber auch im menschlichen Gehörgang eingesetzt werden. Die Messungen finden im reflexionsarmen Raum statt, so dass nur die Wirkung von Außenohr, Kopf und, wenn vorhanden, Torso charakterisiert wird. Sie werden deshalb als kopfbezogene Impulsantwort (HRIR, head related impulse response) bezeichnet. Im Frequenzbereich werden sie Außenohrübertragungsfunktion (HRTF, head related transfer function) genannt.
Werden die Impulsantworten in einem akustisch wirksamen Raum gemessen, nennt man sie kopfbezogene bzw. binaurale Raumimpulsantworten (BRIR, binaural room impulse response). Sie enthalten nicht nur Informationen, die das Außenohr und den Kopf betreffen, sondern auch über den Raum, in dem sie gemessen wurden, sowie den Standort der Quelle und des Kunstkopfes. Im Gegensatz zu den kopfbezogenen Impulsantworten sind in den binauralen Impulsantworten die spezifischen Reflexionsmuster der Quell-Empfänger-Konstellation verankert. Die Lokalisation der Quelle ist aufgrund der Pegel- und Laufzeitunterschiede der linken und rechten binauralen Raumimpulsantwort möglich.
Die Binauralsynthese kann auch auf modellgenerierten Impulsantworten basieren, die in einem raumakustischen Computermodell berechnet wurden. Dazu wird jeder Schallstrahl, der beim Empfänger eintrifft, mit der der Einfallsrichtung entsprechenden HRTF gewichtet (Slavik und Weinzierl, 2008, S. 679).

Zur Auralisation wird die linke und rechte BRIR mit einem beliebigen nachhallfreien Quellsignal gefaltet und es entstehen zwei Ausgangssignale, die das Quellsignal in dem Raum, in dem die Aufnahme der BRIR stattgefunden hat, simulieren. Die Wiedergabe der binauralen Ausgangssignale erfolgt über Kopfhörer. Die Schallquelle wird im Gegensatz zur herkömmlichen Kopfhörerwiedergabe nicht im Kopf lokalisiert, sondern kann aufgrund der interauralen Unterschiede der BRIRs außerhalb des Kopfes lokali-

siert werden (Slavik und Weinzierl, 2008, S. 674-675).
Die Lokalisation einer Quelle ist jedoch erst eindeutig möglich, wenn der Kopf Peilbewegungen ausführen kann, um die Vorne-Hinten-Vertauschung zu umgehen. Durch die dynamische Binauralsynthese kann dieses Problem behoben werden. Die BRIRs müssen für unterschiedliche Kopfpositionen des Kunstkopfes aufgenommen werden, was je nach Schrittweite der BRIR-Messung und dem Bereich der möglichen Kopfbewegungen zu einem großen Datensatz führt. In der Wiedergabesituation wird über einen Positionssensor jeweils das BRIR-Paar mit dem Quellsignal zeitvariant und in Echtzeit gefaltet, welches der aktuellen Kopfposition des Hörers entspricht. Die Faltung wird im Frequenzbereich durchgeführt.

Das Prinzip der dynamischen Binauralsynthese, bestehend aus der Aufnahme der binauralen Impulsantworten in einem realen Raum mit einem Kunstkopftorsosimulator und die Wiedergabe der Simulation eines Quellsignals im Aufnahmeraum mittels dynamischer Binauralsynthese, verdeutlicht Abbildung 1.1.

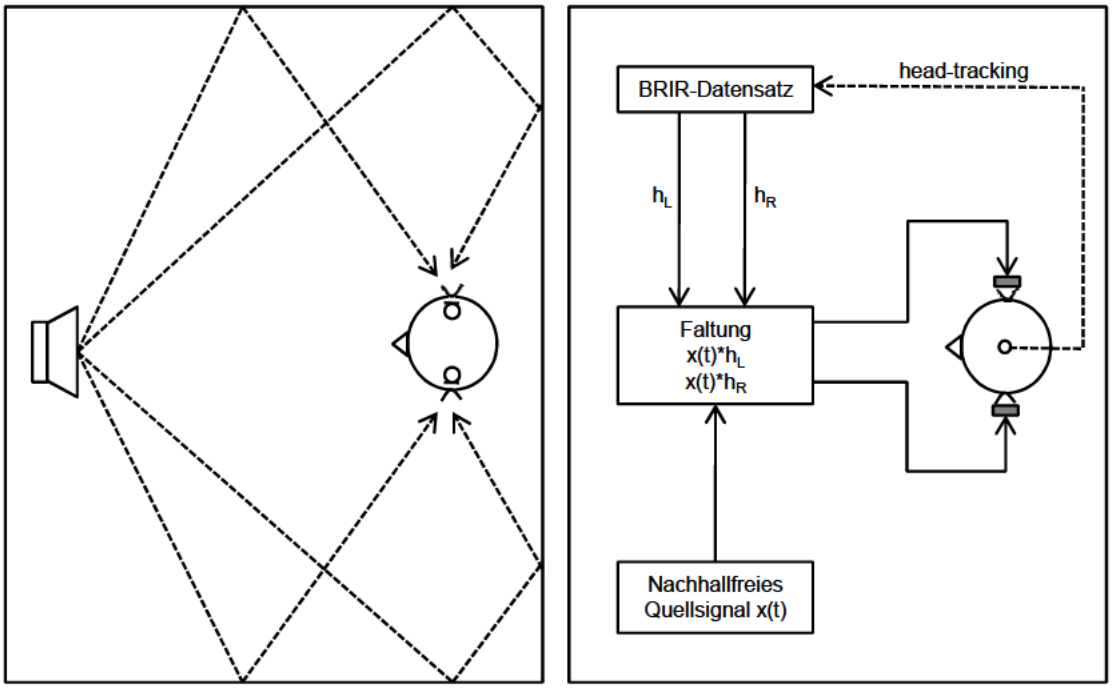



**Abbildung 1.1:** Prinzip der dynamischen Binauralsynthese
links: Aufnahme des BRIR-Datensatzes in einem realen Raum
rechts: Wiedergabe der Simulation eines beliebigen Quellsignals im Aufnahmeraum mittels dynamischer Binauralsynthese

Die für die Plausibilität der Simulation wichtige verzögerungsfreie Nachführung der Impulsantwort erfordert große Datenmengen im Zwischenspeicher des Faltungsrechners, da mehrere binaurale Impulsantworten um die aktuelle Kopfposition bereitgehalten werden. Es werden deshalb Möglichkeiten zur Datenreduktion und Verschnellerung der Echtzeitfaltung gesucht.

In dieser Arbeit wird ein Ansatz zur Datenreduktion verfolgt, der auf die Genauigkeit der Simulation, die für ein perzeptiv exaktes Hörereignis benötigt wird, abzielt.
Die modellhafte Annahme hierzu ist, dass sich Impulsantworten aus Direktschall, frühen Reflexionen und stochastischem Nachhall zusammensetzen. Die Intensität der Reflexionen und deren Abstand untereinander werden durch die Raumgeometrie und die Eigenschaften der reflektierenden Oberflächen, die Position von Quelle und Empfänger und deren Richtcharakteristik bestimmt (Meesawat und Hammershøi, 2003). Der Teil der Impulsantwort, in dem der Direktschall und die frühen Reflexionen enthalten sind, gibt Aufschluss über die Einfallsrichtung des Schalls und über die Entfernung zwischen Quelle und Empfänger. Im stochastischen Nachhall einer Impulsantwort sind nur noch Informationen über die Größe des Raumes enthalten, die Lokalisation der Quelle ist anhand dieses Teils der Impulsantwort nicht möglich (Reilly und McGrath, 1995; Stewart und Sandler, 2007). Die grobe spektrale und zeitliche Struktur des Nachhalls ist ausschlaggebend für den Höreindruck. Feinstrukturen des diffusen Nachhalls sind aufgrund der ansteigenden Reflexionsdichte des abklingenden Schallfeldes nicht unterscheidbar (Kuttruff, 1993). Der Übergangszeitpunkt zwischen den frühen Reflexionen und dem stochastischen Nachhall ist als die physikalische mixing time bekannt.
Der stochastische Nachhall hat die Eigenschaft, dass er keine Information über die Einfallsrichtung des Schalls enthält und somit für jede Position im Raum identisch ist. Der perzeptiv motivierte Ansatz dieser Arbeit ist, dass der individuelle Nachhall einer Impulsantwort durch einen konstanten Nachhall einer anderen beliebigen Kopfposition ersetzt werden kann, und so der Rechenaufwand bei der schnellen Echtzeitfaltung erheblich reduziert werden kann, ohne dass ein Qualitätsverlust hingenommen werden muss. Der Zeitpunkt, ab dem eine perzeptive Unterscheidung zwischen Originalnachhall und Nachhall einer anderen Position nicht mehr möglich ist, ist Untersuchungsgegenstand dieser Arbeit. Im Folgenden wird dieser Zeitpunkt als perzeptive mixing time bezeichnet.

Schon in frühen Veröffentlichungen über die dynamische Auralisation wurde vorge-

schlagen, die Faltung der BRIRs mit dem Quellsignal in zwei Abschnitte aufzuteilen. Die Anzahl der exakt zu simulierenden frühen Reflexionen soll möglichst reduziert werden, um einen geringeren Rechenaufwand zu erreichen. Der frühe Anteil der binauralen Impulsantworten wird deshalb zeitvariant entsprechend der Kopfposition mit dem Quellsignal gefaltet. Die Simulation des stochastischen Nachhalls erfolgt durch eine statische Faltung eines frei wählbaren Nachhalls einer repräsentativen Impulsantwort mit dem Quellsignal. Beide Faltungsprozesse müssen dann angemessen addiert werden (Reilly und McGrath, 1995). In der Auralisation mittels dynamischer Binauralsynthese wird also nur der frühe Anteil der Impulsantworten entsprechend der Kopfposition nachgeführt, der späte Anteil der Impulsantwort ist konstant.

Der Zeitpunkt der Trennung der Faltung wurde bisher anhand verschiedener Prädiktoren geschätzt. Eine Vorhersage des Übergangszeitpunktes auf Grundlage von perzeptiven Untersuchungen war bisher nicht möglich.
Die Kenntnis der mixing time ist nicht nur für die Auralisation in der dynamischen Binauralsynthese notwendig. Der Übergangszeitpunkt in den diffusen Nachhall einer Impulsantwort ist z. B. auch bei Hallgeräten zur Rechenzeitverkürzung relevant (Stewart und Sandler, 2007; Abel und Huang, 2006).

Ziel dieser Arbeit ist es, den Zusammenhang der perzeptiven mixing time mit den physikalischen Eigenschaften von Räumen sowie geeignete Prädiktoren zu finden.
Dazu werden in neun Räumen mit systematisch variierten mittleren Absorptionsgraden und Volumina binaurale Impulsantworten gemessen. In einem Hörversuch wird untersucht, ab welchem Zeitpunkt die Überblendung eines Direktschalls in einen diffusen Nachhall einer anderen Impulsantwort nicht mehr wahrnehmbar ist. Der Hörversuch wird mit dynamischer Binauralsynthese durchgeführt.
Mit einer Varianzanalyse wird geprüft, welche Effekte sich für die perzeptive mixing time bezüglich Volumen und mittleren Absorptionsgrad ergeben und ob Interaktionen auftreten.
Eine Regressionsanalyse soll zeigen, inwiefern Prädiktoren verschiedenster Art, basierend auf den physikalischen Eigenschaften eines Raumes oder gemessenen Raumimpulsantworten, zur Vorhersage der perzeptiven mixing time geeignet sind.

# 2 Stand der Forschung

In diesem Kapitel wird die Idee der mixing time erläutert sowie modellbasierte und empirische Prädiktoren vorgestellt.

## 2.1 Zur Diffusität in Schallfeldern: Theoretische Grundlagen

Die Diffusität eines Schallfeldes kann im Frequenz- und Zeitbereich betrachtet werden. Das stochastische Modell im Frequenzbereich wurde von Schroeder (1954) eingeführt. Die Dichte der Eigenfrequenzen in einem Raum ist bei hohen Frequenzen sehr viel größer als bei tiefen Frequenzen. Gleichung (2.1) zur Berechnung der Eigenfrequenzdichte pro Hertz mit dem Volumen $V$ in $m^3$, der Frequenz $f$ in Hz und der Schallgeschwindigkeit $c$ in m/s veranschaulicht den quadratischen Zusammenhang. Diese Gleichung gilt nur für hinreichend große Frequenzen in beliebigen Raumformen, für tiefe Frequenzen ist sie nur in rechteckförmigen Räumen gültig (Kuttruff, 2004, S. 165).

$$\frac{dN}{df} = 4\pi V \frac{f^2}{c^3} \tag{2.1}$$

Die Moden überlagern sich bei hohen Frequenzen zu einem diffusen Schallfeld, in dem einzelne Komponenten nicht mehr zuordenbar sind, bei tiefen Frequenzen hingegen spricht man dagegen von modalem Verhalten. Die Grenze, ab der die Theorie des diffusen Schallfeldes im Frequenzbereich gültig ist, ist bekannt als die Schroederfrequenz und berechnet sich wie folgt (Schroeder, 1996):[1]

[1] In der ersten Veröffentlichung von Schroeder zum genannten Thema wurde ein konservativerer Vorfaktor angenommen ($f_{Schroeder} = 4000\sqrt{\frac{T}{V}}$) (Schroeder, 1954)

$$f_{Schroeder} = 2000\sqrt{\frac{T}{V}} \tag{2.2}$$

mit der Nachhallzeit $T$ in s und dem Volumen $V$ in $m^3$. Diese Formel wurde ausgehend von der Halbwertsbreite einer Mode und dem mittleren Abstand benachbarter Eigenfrequenzen hergeleitet. Sie besagt, dass oberhalb der Schroederfrequenz innerhalb der Halbwertsbreite einer Mode mindestens drei Moden liegen.
Wenn die Moden statistisch unabhängig voneinander sind, folgt aus dem zentralen Grenzwerttheorem, dass der Real- und Imaginärteil des komplexen Frequenzganges Re{H(f)} und Im{H(f)} im Diffusfeld normalverteilt sind (Davenport und Root, 1958, S. 158-165). Daraus wiederum folgt, dass der Betragsfrequenzgang |H(f)| rayleighverteilt ist (Hirano, 2005; Jot u. a., 1997). Diese statistischen Eigenschaften sind unabhängig von der Messposition, vorausgesetzt, dass der Empfänger den Bereich des diffusen Schallfeldes nicht verlässt, d. h. solange er sich außerhalb des Hallradius befindet bzw. der Direktschall im Vergleich zum reflektierten Schall vernachlässigt werden kann (Jot u. a., 1997, S. 3).

Die Betrachtung der Diffusität im Zeitbereich basiert auf der modellhaften Annahme, dass sich Raumimpulsantworten aus Direktschall, frühen Reflexionen und spätem Nachhall zusammensetzen (siehe Abbildung 2.1, oben). Der Direktschall ist der Schall, der ohne Umwege von der Quelle zum Empfänger gelangt. Die nachfolgenden schwächeren Einzelreflexionen entstehen durch Schallrückwürfe der Seitenwände, der Decke und des Bodens. Sie sind in der Impulsantwort teilweise noch deutlich erkennbar und variieren je nach Quell- und Empfängerposition innerhalb des Raumes. Nach einer bestimmten Zeit werden die Reflexionen so dicht, dass keine Einzelreflexionen mehr erkennbar sind. Dieser Teil der Impulsantwort verhält sich diffus, d. h. es existiert keine Vorzugsrichtung des Schalleinfalls und der Schall kommt gleichwahrscheinlich aus allen Richtungen. Zudem besitzt das Schallfeld an jedem Punkt im Raum die gleiche Energiedichte (Schroeder, 1959). Das bedeutet, dass der Nachhall einer Impulsantwort im Gegensatz zu den frühen Reflexionen überall im Raum identisch ist. Der Zeitpunkt des Übergangs in den stochastischen, späten Nachhall wird auch als *physikalische mixing time* $t_m$ bezeichnet (Polack, 1992). Nach der mixing time ist der Energieausklang unabhängig von der Richtcharakteristik und von der Quell- und Empfängerposition, d. h. der Energieausklang ist im gesamten Raum gleichverteilt (Polack, 1992, S. 264).

Polack erklärt mixing unter Zuhilfenahme der Billiardtheorie. Demnach ist mixing erreicht, wenn sich eine Gruppe gleichzeitig ausgesendeter Teilchen, die anfangs ähnlichen Pfaden folgen, nach einer sehr langen Zeit (streng genommen unendlich lang) vollständig aufspaltet und im gesamten Raum verteilt. Die Zustände des Schallfeldes zu den beiden Beobachtungszeitpunkten sind statistisch unabhängig voneinander. Anders ausgedrückt, ist ein Raum 'gemischt', wenn kein Zusammenhang mehr zum Ausgangszustand des Schallfeldes erkennbar ist und die Position und Richtung der Schallteilchen unbekannt sind (Polack, 1993).
Eine Voraussetzung für mixing ist Ergodizität (Polack, 1992). Zur Definition von Ergodizität wird die Ausbreitung eines Teilchens entlang eines Pfades betrachtet: Ein Teilchen verbringt die gleiche Zeit an jeder Position im Raum, ohne dass eine Richtung bevorzugt wird (Joyce, 1975). Mathematisch bedeutet das, dass es für die Mittelung von Schallfeldgrößen keine Rolle spielt, ob sie für ein Teilchen über der Zeit oder für viele Teilchen über dem Ort durchgeführt wird.[2]
Das Eintreten von Ergodizität ist von der Raumform und den Reflexionseigenschaften der Oberflächen abhängig. Beispiele für nicht-ergodische Räume sind z. B. rechteckige, nicht-streuende Räume, weil die Ausbreitungsrichtung der Teilchen deterministisch bleibt. Ein weiteres Beispiel bilden kugelförmige, nicht-streuende Räume, da die Teilchen wegen der Bündelung an einem Punkt nicht alle Orte im Raum erreichen können (Polack, 1992).

Der Gültigkeitsbereich dieser beiden stochastischen Modelle nach Schroeder und Polack kann in der Zeit-Frequenz-Ebene zusammengefasst werden (Jot u. a., 1997). Die Abbildung 2.1 zeigt, in welchem Bereich sich eine Raumimpulsantwort diffus verhält. Diffusität bzw. mixing kann im tiefen Frequenzbereich und im Bereich der frühen Reflexionen nicht erreicht werden. Sie wird durch die Schroederfrequenz und die mixing time begrenzt.

In realen Konzertsälen ist der Übergang zwischen frühen und späten Reflexionen fließend (Reichardt und Lehmann, 1978). Wenn eine Schallwelle auf eine Begrenzungsfläche trifft, wird sie reflektiert. Je nach Oberflächeneigenschaften ist die Reflexion spiegelnd oder teilweise bzw. vollständig diffus. Aus der Art und Anzahl der Reflexionen ergibt sich, wie schnell ein Schallfeld in einen diffusen Zustand wechselt.

[2] „time average ensemble average“ (Kuttruff, 1979, S. 104), vgl. auch Nikias und Petropulu (1993, S. 19)

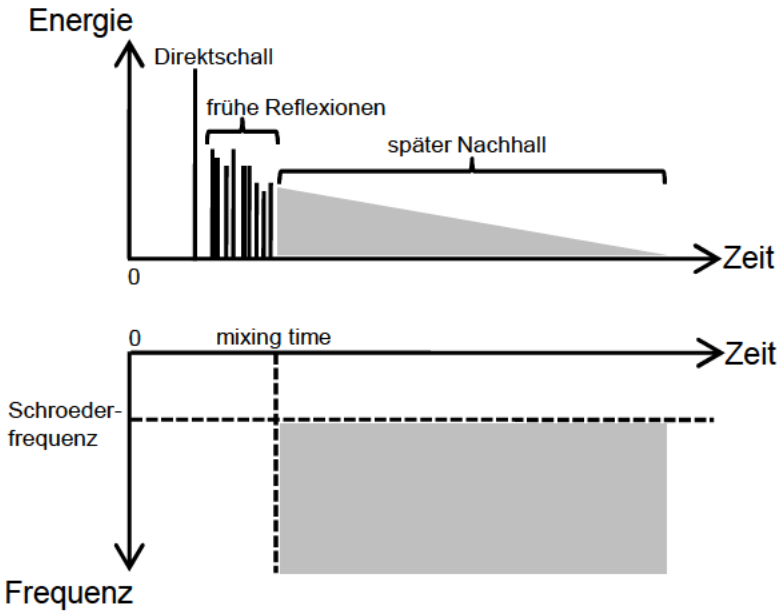



**Abbildung 2.1:** Gültigkeitsbereich des stochastischen Modells einer Impulsantwort
oben: schematische Darstellung einer Impulsantwort, diffuser Bereich grau gekennzeichnet
unten: Diffusitätsbereich in der Zeit-Frequenz-Ebene (grau)
Abbildung angelehnt an Jot u. a. (1997, Abb. 2 im Anhang)

In einem ideal diffus reflektierenden Raum verteilt sich die Schallenergie kontinuierlich mit der Zeit im gesamten Raumvolumen.
Die mittlere Reflexionsdichte steigt quadratisch mit der Zeit. Im Gegensatz dazu fällt die Energie aller Reflexionen wegen der größer werdenden Distanzen, die die Strahlen zurücklegen müssen, der Luftabsorption und der Absorption der Oberflächen exponentiell ab (Kuttruff, 1979, S. 87-88). Der Zeitpunkt, zu dem ein Schallfeld diffuses Verhalten annimmt, tritt in großen Räumen später ein, da die Dauer zwischen den Reflexionen aufgrund der großen freien Weglängen höher ist.
Diffusität wird in realen Räumen durch die diffus reflektierenden Oberflächen erreicht. Eine weitere Möglichkeit ist, die mittlere freie Weglänge zu verkürzen und schallharte Hindernisse frei im Raum mit unterschiedlicher Ausrichtung zu positionieren (Kuttruff, 2000, S. 120).

In einem realen Raum kann vollständige Diffusität nicht erreicht werden (Kuttruff, 2000). Besonders in Räumen, in denen der reflektierte Schall in bestimmten Regionen konzentriert wird, wird die Diffusität behindert. Beispiele dafür sind Räume mit sehr ungleichmäßig verteilten Absorptionsflächen, in denen bestimmte Pfade ausgelöscht werden. Das kann z. B. in Räumen mit einer hoch absorbierenden Publikumsfläche, aber nur sehr wenig absorbierenden Raumbegrenzungsflächen der Fall sein. Auch kann in Räumen mit einer sehr regelmäßigen Form und wenig diffus reflektierenden Begrenzungsflächen, in gekoppelten Räumen oder in sehr gedämpften Räumen nicht von einem perfekt diffusen Verhalten im Ausklang ausgegangen werden.

Polack betont jedoch, dass auch in Räumen, in denen aufgrund ihrer Form theoretisch kein diffuses Schallfeld eintreten kann, die unvermeidbaren Unebenheiten einer Wand ausreichen, dass sich das Schallfeld stochastisch verhält, wenn auch erst nach einer unverhältnismäßig langen Zeit (Polack, 1992, S. 263).

Ein Schallfeld verhält sich diffus, wenn die Energiedichte überall im Raum gleichverteilt ist und keine Information über die Einfallsrichtung des Schalls vorhanden ist, d. h. der Nachhallschall an jeder Position im Raum identisch ist. Aufgrund der hohen Reflexionsdichte sind Feinstrukturen nicht mehr wahrnehmbar. Perzeptiv betrachtet, bedeutet das, dass der stochastische Nachhall einer Impulsantwort ohne hörbare Konsequenzen durch den Nachhall einer anderen beliebigen Sender-/Empfängerkonstellation ausgetauscht werden kann. Idealerweise entspricht dieser Zeitpunkt der physikalischen mixing time, aber die Relation zwischen physikalischer und *perzeptiver mixing time* $t_{mp}$ ist nicht trivial ableitbar. Ein komplexes Zusammenspiel auditiver Verdeckungseffekte bestimmt, ab wann eine Nachhallmanipulation nicht mehr hörbar ist. Die Hörbarkeit jeder einzelnen Reflexion wird durch pegel-, richtungs- und zeitabhängige Maskierung und das Gesetz der ersten Wellenfront beeinflusst (Kuttruff und Mommertz, 2004). Zudem ist die Unterscheidbarkeit einzelner Reflexionen abhängig vom dargebotenen Audioinhalt (Olive und Toole, 1989).

## 2.2 Modellbasierte Prädiktoren von Diffusität

Die hier als modellbasiert verstandenen Prädiktoren umfassen sowohl Einzahlkennwerte als auch Maße, die auf den physikalischen Eigenschaften eines Raumes beruhen. Ihnen

gemeinsam ist, dass sie auf zumeist statistisch motivierten modellhaften Annahmen über den Aufbau eines stochastischen Raumschallfelds beruhen.

Die festen Zeiten konnten in Veröffentlichungen gefunden werden, deren Untersuchungsgegenstand meist nicht die mixing time war, die aber für Hörversuche mit synthetischen Schallfeldern trotzdem eine Startzeit für den späten Nachhall einer Impulsantwort angenommen haben. Begault (1992) nahm 50 ms an, Bradley und Soulodre (1995) und Reilly u. a. (1995) 80 ms und in der Untersuchung von Reichardt u. a. (1974) über die Grenze zwischen „brauchbarer und unbrauchbarer“ Durchsichtigkeit wurden die Nachhalleinsatzzeiten zwischen 50 und 168 ms gewählt. Gleichzeitig wurde jedoch betont, dass in realen Konzertsälen frühestens nach 120 ms von einem diffusen Schallfeld ausgegangen werden kann. Einige Jahre später veröffentlichten Reichardt und Lehmann (1978) einen Artikel über den Raumeindruck, in dem 150-200 ms als Übergangszeit in einen statistischen Nachhall genannt werden. Dieser Wert wurde aus Reflektogrammen von mittelgroßen Konzertsälen abgelesen.
Junius (1959) stellte anhand von zwei Beispielen (Musikaufführungsstätte und Theatersaal) fest, dass der Schall nach 50-100 ms gleichmäßig aus allen Richtungen eintrifft. Nach Kuttruff (1991) ist das Schallfeld eines Raumes „of some size“ (gewisser Größe) nach 100-150 ms diffus, der Übergangszeitpunkt zwischen frühen und späten Anteilen der Impulsantwort sollte demnach in diesem Bereich liegen.
Untersuchungen der Integrationsgrenzen zur Berechnung des interauralen Kreuzkorrelationskoeffizienten (IACC, interaural cross-correlation coefficient) von Hidaka u. a. (1995) zeigten, dass der Zeitpunkt, der frühe von späten Reflexionen trennt, zwischen 50 und 200 ms liegt.

Zudem existieren Maße, die auf den physikalischen Eigenschaften eines Raumes basieren. Dazu gehören die Reflexionsdichte, die mittlere freie Weglänge und die Nachhallzeit.
Das Spiegelschallquellenmodell liefert die Gleichung zur Bestimmung der Reflexionsdichte $\frac{dN}{dt}$ in Rechteckräumen[3] (Cremer, 1961):

$$\frac{dN}{dt} = 4\pi \frac{c^3 t^2}{V} \tag{2.3}$$

[3]Es konnte gezeigt werden, dass diese Gleichung auch für andere Raumformen Gültigkeit hat (Kuttruff, 1979)

mit der Schallgeschwindigkeit $c$ in m/s, der Zeit $t$ in s und dem Raumvolumen $V$ in $m^3$. Die Prädiktoren zur Schätzung der mixing time basierend auf der Reflexionsdichte sind perzeptiv motiviert und zielen auf das zeitliche Auflösungsvermögen des Gehörs ab. Schreiber (1960) fand heraus, dass abhängig von der Bandbreite ca. 2000-2500 Reflexionen pro Sekunde als gleichförmiges Rauschen wahrgenommen werden, wobei bei tieffrequentem Rauschen weniger Reflexionen pro Sekunde notwendig waren. Auch Cremer und Müller (1978) gehen davon aus, dass das Ohr ab einer Reflexionsdichte von 2000 $s^{-1}$ keine Feinstrukturen mehr erkennen kann. Stellt man Gleichung (2.3) nach der Zeit um (siehe Gleichung (2.4)) und setzt diese Reflexionsdichte ein, so ergibt sich Gleichung (2.5) als Berechnungsvorschrift für die mixing time.

$$t = \sqrt{\frac{dN}{dt} \cdot \frac{V}{4\pi c^3}} \quad [s] \tag{2.4}$$

$$t_{mp} = 2\sqrt{V} \quad [ms] \tag{2.5}$$

Zahlreiche weitere Autoren geben Vorschläge über die notwendige Reflexionsdichte für ein stochastisches Schallfeld. Schroeder (1962) fand heraus, dass das Gehör 1000 Reflexionen pro Sekunde nicht von einem höheren Wert unterscheiden kann. Nach Rubak und Johansen (1999) wird eine Reflexionsdichte von 4000 $s^{-1}$ benötigt, um einen perzeptiv hochwertigen Nachhall zu generieren. Griesinger (1989) schlägt 10000 Reflexionen pro Sekunde oder mehr vor, um eine gute perzeptive Qualität des künstlich erzeugten Nachhalls zu erzielen. Untersuchungen von Huang und Abel (2007) ergaben, dass 20000 Reflexionen pro Sekunde oder mehr notwendig sind, damit eine Unterscheidung von gaußverteiltem Rauschen nicht möglich ist.

Schmidt und Ahnert (1973) fanden in ihrer Untersuchung über den Raumeindruck heraus, dass bei 5 Reflexionen innerhalb eines Zeitraumes von 20 ms von zeitlicher Diffusität des Nachhallschalls ausgegangen werden kann. Das entspricht einer Reflexionsdichte von 250 $s^{-1}$, was eingesetzt in Gleichung (2.4) die Beziehung

$$t_{mp} = \sqrt{\frac{V}{2}} \quad [ms] \tag{2.6}$$

liefert.

Oft genannt wird zudem noch die folgende Gleichung, die unter anderem aus Beobachtungen an mittelgroßen Konzertsälen abgeleitet wurde (Reichardt und Lehmann, 1978):

$$t_{mp} = \sqrt{V} \quad [ms] \tag{2.7}$$

Die erforderliche Reflexionsdichte, um mit dieser mixing time Gleichung (2.3) zu erfüllen, entspricht etwa 400 $s^{-1}$. Auch Jot u. a. (1997) und Polack (1988) nennen Gleichung (2.7) als gute Schätzung des Übergangszeitpunktes zwischen frühen Reflexionen und spätem Nachhall. Grundlage hierfür ist wiederum eine perzeptive Betrachtungsweise: Die Überlagerung von 10 zufälligen Rückwürfen ergibt eine näherungsweise gaußförmige Verteilung. Die mixing time ist erreicht, wenn diese 10 Reflexionen in einem bestimmten Zeitintervall auftreten. Als Zeitintervall dient hier die charakteristische Zeitauflösung des Gehörs, die auf 24 ms geschätzt wird (Polack, 1988, S. 48).

Rubak und Johansen (1999) führten einen neuen Ansatz zur Bestimmung der mixing time ein, der auf der mittleren freien Weglänge basiert. Es wird angenommen, dass sich das Schallfeld diffus verhält, sobald jedes Schallteilchen im Mittel einige Male reflektiert wurde. Die mixing time ist nach Rubak und Johansen (1999) die Laufzeit, die ein Teilchen benötigt, um die vierfache mittlere freie Weglänge zurückzulegen bzw. die Zeit, nach der vier Wandreflexionen aufgetreten sind:

$$t_{mp} = 4\frac{l_m}{c} \tag{2.8}$$

mit der Schallgeschwindigkeit $c$ und der mittleren freien Weglänge $l_m$, die sich aus dem Volumen $V$ und der Raumoberfläche $S$ zusammensetzt:

$$l_m = \frac{4V}{S} \tag{2.9}$$

Einen weiteren Ansatz zur Bestimmung der mixing time lieferten Hidaka u. a. (2007). Insgesamt mehr als 50 Räume unterschiedlichen Typs (Sinfonie, Kammermusiksäle, Opernsäle) und Größe wurden mittels einem von Hidaka u. a. (2007) entwickelten Algorithmus zur Bestimmung der physikalischen mixing time (siehe Abschnitt 2.3.4) aus Impulsantworten untersucht. Durch eine lineare Regression über die Ergebnisse dieser Studie zeigte sich, dass die Nachhallzeit als Prädiktor der mixing time des 500-Hz-

Oktavbandes genutzt werden kann (Varianzaufklärung: ca. 57 %).

$$t_{m500Hz} = 0.08 \cdot T_{500Hz} \quad [ms] \tag{2.10}$$

Die vorgestellten modellbasierten Prädiktoren zur Bestimmung der mixing time können auf die Reflexionsdichte bzw. die Quadratwurzel des Volumens $\sqrt{V}$, die mittlere freie Weglänge $l_m$ oder die Nachhallzeit $T$ zurückgeführt werden. Für die spätere Evaluation der perzeptiven mixing time können demnach drei unabhängige Beziehungen herausgefiltert werden. Die Reflexionsdichtebeziehungen (2.6) und (2.7) werden mit

$$t_{mp1} = k_{refl} \cdot \sqrt{V} \tag{2.11}$$

zusammengefasst, die Weglängebeziehung (2.8) lässt sich zu

$$t_{mp2} = k_{path} \cdot \frac{V}{S} \tag{2.12}$$

verallgemeinern und die Abschätzung der mixing time basierend auf der Nachhallzeit kann als

$$t_{mp3} = k_{reverb} \cdot T_f \tag{2.13}$$

geschrieben werden.

## 2.3 Empirische Prädiktoren von Diffusität

In der Literatur sind einige Methoden zu finden, die die physikalische mixing time auf Basis gemessener Raumimpulsantworten berechnen. Im folgenden Abschnitt werden vier dieser Algorithmen erläutert, deren perzeptive Evaluation im Abschnitt 5.2 vorgestellt wird.

### 2.3.1 Abel und Huang, 2006

Abel und Huang entwickelten eine Methode zur Bestimmung der mixing time in Raumimpulsantworten, die auf der Annahme beruht, dass die Amplituden des Schalldrucks bei vollständiger Diffusität des Raumes normalverteilt sind. Das ist der Fall, wenn die Reflexionen so dicht sind, dass sie sich gegenseitig überlappen und Einzelreflexionen nicht mehr erkennbar sind (Jot u. a., 1997, S. 5). Dazu wird ein Echodichteprofil erstellt, das den Anteil der Schalldruckamplituden der Impulsantwort innerhalb eines bestimmten Fensters darstellt, die außerhalb der Standardabweichung liegen. Es wird mit folgender Formel erzeugt:

$$\eta(t) = \frac{1}{erfc(1/\sqrt{2})} \cdot \frac{1}{2\delta} \sum_{\tau=t-\delta}^{t+\delta-1} \Theta(|h(\tau)| > \sigma) \tag{2.14}$$

mit

$$\sigma = \sqrt{\frac{1}{2\delta} \sum_{\tau=t-\delta}^{t+\delta-1} h^2(\tau)} \tag{2.15}$$

Die Standardabweichung $\sigma$ der Schalldruckamplituden innerhalb eines bestimmten Zeitfensters wird berechnet. Um zu testen, inwiefern die Verteilung der Schalldruckamplituden einer Normalverteilung entspricht, wird die Anzahl der Amplituden, die außerhalb der Standardabweichung liegen, mit der Anzahl aller Amplituden innerhalb des Fensters ins Verhältnis gesetzt. Anschließend wird durch den Anteil der Samples dividiert, der bei einer Normalverteilung erwartungsgemäß außerhalb der Standardabweichung liegt ($erfc(1\sqrt{2})$=0,3173). Dies entspricht einer Normierung des Echodichteprofils auf den Wert 1 bei Normalverteilung der Amplituden.
In Gleichung (2.14) entspricht $\delta$ der halben Fensterlänge, bei $\Theta(x)$ handelt es sich um die Sprungfunktion, die bei x-Werten kleiner 0 den Wert 0, bei x-Werten größer 0 den Wert 1 annimmt. In Gleichung (2.14) bedeutet das, dass nur die Amplituden der Impulsantwort gezählt werden, die größer als die Standardabweichung innerhalb des Fensters sind. Das Echodichteprofil entsteht, indem das Fenster sampleweise über die Impulsantwort 'geschoben' wird.

Im frühen Anteil einer Impulsantwort treten nur wenige Reflexionen auf, sie liegen weit auseinander und der Schalldruckpegel der Reflexionen ist sehr unterschiedlich. Die Standardabweichung der Amplituden ist dementsprechend sehr groß, es liegen nur wenige Samples außerhalb der Standardabweichung. Die Echodichte ist sehr gering. Mit der Zeit werden die Reflexionen immer dichter, das Echodichteprofil steigt monoton an. Wenn das Schallfeld vollständig diffus wird, verweilt das Echodichteprofil ungefähr bei 1. Die Amplituden sind dann normalverteilt. Der Zeitpunkt, an dem das Echodichteprofil das erste Mal den Wert 1 erreicht, ist der Beginn des stochastischen Nachhalls. Dieses Kriterium wird im Folgenden als das erste Kriterium von Abel und Huang (2006) bezeichnet.
Im Ausklang der Impulsantworten wird nach einer gewissen Zeit das Grundrauschen dominant, was dazu führt, dass die Echodichte wegen des stochastischen Verhaltens des Rauschens gegen 1 laufen muss. Ein gewisser Signal/Rausch-Abstand in der Impulsantwort muss dementsprechend auch noch nach dem Bereich der erwarteten mixing time eingehalten werden.
Durch eine Gewichtung der Samples und Vergrößerung des Fensters kann das Echodichteprofil geglättet werden, der grobe Verlauf bleibt jedoch ähnlich. In Abbildung 2.2 ist ein typischer Verlauf der Echodichte und die dazugehörige Impulsantwort dargestellt.

Die Autoren schlagen ein weiteres Kriterium zur Bestimmung der mixing time vor, um die Schwankungen des Echodichteprofils, die aufgrund der endlichen Fensterlänge auftreten, möglichst wenig in die Bestimmung der mixing time einfließen zu lassen: Der Beginn des späten Nachhalls setzt ein, wenn das Echodichteprofil das erste Mal größer als 1-$\sigma_l$ wird ($\sigma_l$ ist die Standardabweichung des späten Echodichteprofils). Dieses Kriterium wird im Folgenden als zweites Kriterium nach Abel und Huang (2006) bezeichnet.

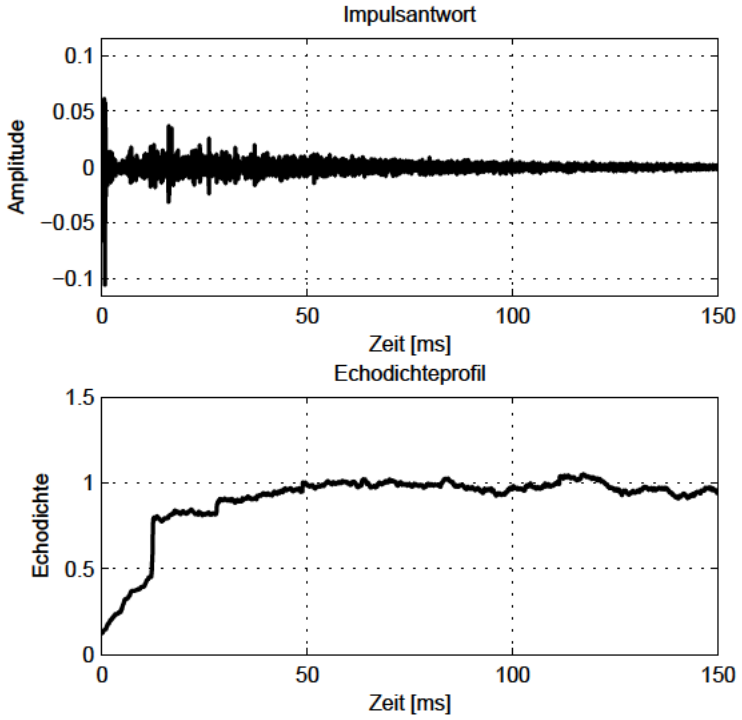



**Abbildung 2.2:** Impulsantwort und Echodichteprofil nach Abel und Huang (2006), exemplarisch für einen kleinen Seminarraum

### 2.3.2 Stewart und Sandler, 2007

Der Methode zur Bestimmung der physikalischen mixing time von Stewart und Sandler (2007) liegt wie auch bei Abel und Huang (2006) die Annahme zugrunde, dass die Amplituden im Zeitbereich eines diffusen Signals normalverteilt sind. Der Algorithmus ist den Autoren zufolge besser geeignet als die Methode zur Echodichtebestimmung, da bei ihrer Methode der Übergang zwischen frühen und späten Anteilen der Impulsantwort weniger graduell verliefe und folglich eindeutiger abgelesen werden kann.

Ihr Verfahren beruht auf der Kurtosis, die innerhalb eines gleitenden Fensters gemessen und über der Zeit aufgetragen wird. Die Kurtosis ist ein Maß für die Steilheit einer Verteilung (Zenga, 2005). Bei der Kurtosis handelt es sich um einen Kumulanten 4. Ordnung, d. h. sie misst direkt den Grad der „Nicht-Gaußheit“ von Zeitverläufen. Die normalisierte Kurtosis wird mit

$$\gamma_4 = \frac{E(x-\mu)^4}{\sigma^4} - 3 \tag{2.16}$$

berechnet, wobei $E()$ der Erwartungswert, $\mu$ der Mittelwert und $\sigma$ die Standardabweichung ist. Eine Normalverteilung ist gegeben, wenn die normalisierte Kurtosis wie alle Kumulanten der Ordnung größer 2 den Wert 0 annimmt (Nikias und Petropulu, 1993).

Abbildung 2.3 zeigt einen Kurtosisverlauf beispielhaft für einen kleinen Seminarraum. Im Bereich der frühen Reflexionen ist ein steiler Abfall erkennbar. Danach ist die Kurtosis nahezu 0. Das Kriterium für den Übergangszeitpunkt zwischen frühen Reflexionen und stochastischem Nachhall wird von den Autoren nicht eindeutig benannt. Anhand der Diskussion ihrer Kurtosisplots ist jedoch zu vermuten, dass die Autoren den stochastischen Nachhall als erreicht betrachten, wenn die Kurtosis ungefähr 0 wird. Dies stimmt mit der Theorie überein, nach der eine Normalverteilung gegeben ist, wenn die normalisierte Kurtosis dem Wert 0 entspricht (Zenga, 2005).

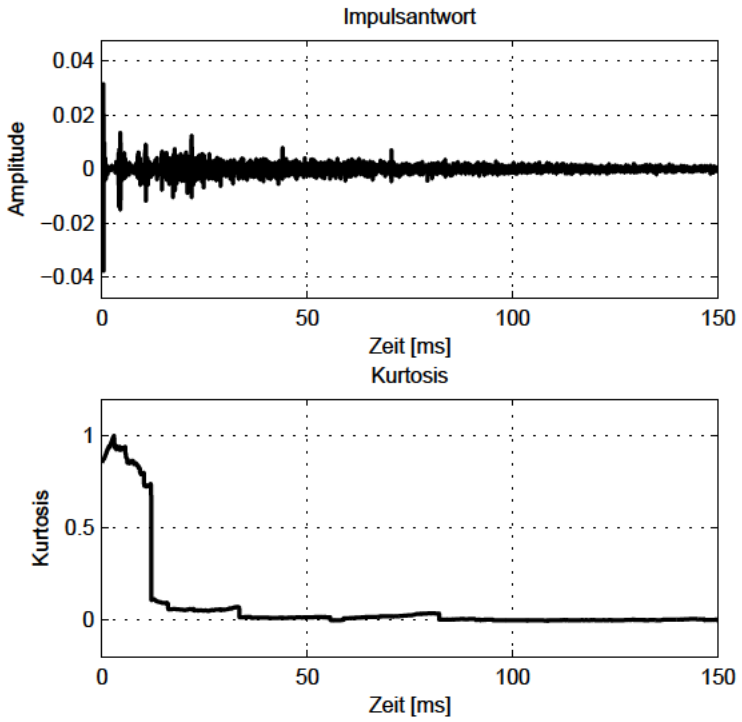



**Abbildung 2.3:** Impulsantwort und Kurtosisverlauf nach Stewart und Sandler (2007), exemplarisch für einen kleinen Seminarraum

### 2.3.3 Defrance u. a., 2009

Das Grundprinzip der Methode von Defrance u. a. (2009) ist, dass die kumulierte Anzahl der Reflexionen, die in einem Raum nach Anregung mit einem Impuls beobachtet werden kann, eine kubische Funktion über der Zeit bildet. Durch Integration von Gleichung (2.3) über der Zeit, kann dieser Zusammenhang folgendermaßen beschrieben werden:

$$N = \frac{4\pi}{3} \cdot \frac{c^3}{V} \cdot t^3 \tag{2.17}$$

Die mixing time wird nun zu dem Zeitpunkt angenommen, ab dem die Reflexionen so dicht sind, dass sie sich überlappen und nicht mehr voneinander unterscheidbar sind. Die Reflexionsdichte steigt nicht mehr quadratisch, sondern bleibt konstant. Die kumulierte Anzahl der Reflexionen (CNA, cumulated number of arrivals) steigt dann nicht mehr kubisch, sondern linear an. Abbildung 2.4 zeigt den typischen Verlauf einer CNA-Kurve sowie die kubische und lineare Anpassungsfunktion an den jeweiligen Bereich:

- $t < t_m$ kubischer Verlauf der CNA-Kurve
- $t > t_m$ linearer Verlauf der CNA-Kurve

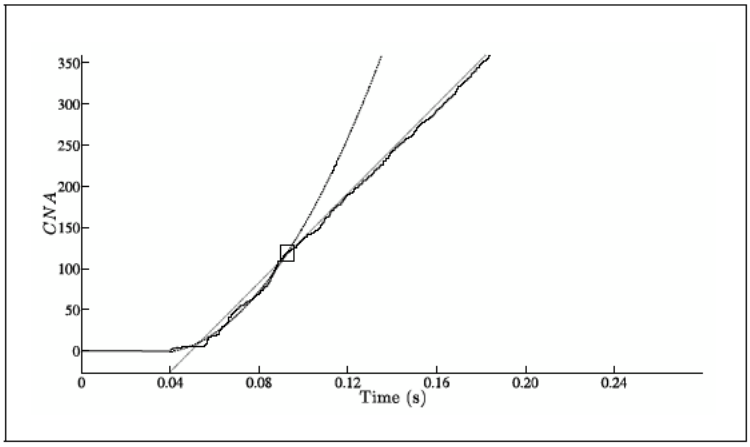



**Abbildung 2.4:** CNA (dicke Linie) mit kubischem ($t < t_m$) und linearem Fit ($t > t_m$)
□ $t_m$
(Quelle: Defrance u. a. (2009, S. 1076, Abb. 11))

Die Detektion dieser Reflexionen führen Defrance u. a. (2009) mit Hilfe der Matching Pursuit-Methode durch. Mit diesem Algorithmus können beliebige Signale zerlegt werden. Hierzu werden die Reflexionen in einer Impulsantwort als Reproduktionen des Direktschalls aufgefasst. Die Detektion der Reflexionen („arrivals“) kann erfolgen, indem der Direktschall mit der Impulsantwort korreliert wird. Matching Pursuit verfolgt genau dieses Prinzip: das sogenannte „mother atom“, in diesem Fall der Direktschall, wird schrittweise mit der Impulsantwort korreliert. Der Teil der Impulsantwort, der dem Direktschall am ähnlichsten ist, wird als „arrival“ detektiert, Amplitude und Zeitpunkt werden gespeichert. Die Zerlegung wird mit der verbleibenden Impulsantwort („residual“) fortgeführt, der sich durch die Subtraktion des mit der Amplitude des „arrivals“ gewichteten Direktschalls von der Impulsantwort ergibt. Die Zerlegung wird solange fortgesetzt, bis das Stopkriterium (SRR, „signal/residual ratio“) erreicht ist. Dazu muss das Energieverhältnis des rekonstruierten Signals und der verbleibenden Impulsantwort („residual“) einen bestimmten Wert unterschreiten. Defrance u. a. (2009) empfehlen ein SRR von 5 dB. Das Ergebnis dieser Art der Zerlegung ist ein Koeffizientenvektor mit den aus den Korrelationen abgeleiteten Gewichten und den Zeitpunkten der Reflexionen der Impulsantwort.

Matching Pursuit hat die Eigenschaft, dass die Wahrscheinlichkeit, „arrivals“ zu detektieren, vom Energieverlauf des Signals über der Zeit abhängt. Da die Energie einer Impulsantwort mit der Zeit sinkt, werden „arrivals“ eher am Anfang der Impulsantwort detektiert. Um trotzdem eine passende Zerlegung der Impulsantwort zu erhalten, muss der Energieausklang kompensiert werden (Defrance u. a., 2009). Damit entspricht die Impulsantwort einem stationären Signal und die Wahrscheinlichkeit, einen „arrival“ im Nachhallschall zu detektieren, ist genauso hoch wie bei den frühen Reflexionen.

Die Energiedichte in einem Raum klingt nach dem folgenden Gesetz ab (Kuttruff, 2004, S. 264):

$$w(t) = w_0 e^{\frac{c \cdot S}{4V} t \ln(1-\alpha)} \tag{2.18}$$

wobei $c$ die Schallgeschwindigkeit ist, $S$ die Raumoberfläche, $V$ das Volumen, $t$ die Zeit und $\alpha$ der Absorptionsgrad. Für den Absorptionsgrad setzt man den mittleren Absorptionsgrad $\alpha_{mittel}$ ein (dazu siehe auch Gleichung (3.1) im Abschnitt 3.1), da die „Absorptionsgrade der Raumbegrenzungsfläche i.Allg. nicht einheitlich sein wer-

den [...]“ (Kuttruff, 2004, S. 264). Es gilt $w(t) \sim h^2(t)$, aus diesem Grund muss die Multiplikation der Impulsantwort mit der Wurzel der Inversen der Energiedichte bzw. des Energieausklanges durchgeführt werden. Die Energiekompensation muss vor der Zerlegung der Impulsantwort erfolgen.

Die mixing time aus den CNAs abzulesen ist sehr zeitaufwendig, weshalb Defrance u. a. (2009) eine Methode zur automatischen Detektion vorschlagen. Demnach liegt die mixing time in einem Bereich der Impulsantwort, in dem die Dichte der Reflexionen sehr groß ist. Wenn der Abstand zwischen zwei „arrivals“ geringer ist als die Dauer des Direktschalls, kann Matching Pursuit nur einen „arrival“ detektieren. Die Dauer des Direktschalls bezeichnen Defrance u. a. (2009) als äquivalente Dauer des Impulses, angelehnt an die äquivalente statistische Bandbreite, die definiert ist als die „bandwidth of a hypothetical rectangular filter which would pass a signal with the same mean-square statistical error as the actual filter when the input is white noise“ (Stanley und Peterson, 1979, S. 1633). Wenn man das Spektrum durch das Zeitsignal, dem Direktschall $h_{dir}$, ersetzt, erhält man folgende Formel zur Berechnung der äquivalenten Dauer $\tilde{d}$ zwischen den Impulsen:

$$\tilde{d} = \frac{\left(\int_0^\infty |h_{dir}(t)|^2 dt\right)^2}{\int_0^\infty |h_{dir}(t)|^4 dt} \tag{2.19}$$

Als Kriterium zur Bestimmung der mixing time schlagen Defrance u. a. (2009) schließlich den Zeitpunkt vor, zu dem die Entfernung zweier aufeinander folgender „arrivals“ kleiner oder gleich der äquivalenten Dauer $\tilde{d}$ ist ($\Delta t \leq \tilde{d}$).

Dieser automatisch bestimmte Zeitpunkt entspricht nicht zwingend dem Knickpunkt der CNAs zwischen kubischem und linearen Verlauf, wie die praktischen Betrachtungen in Abschnitt 3.3.3 zeigen. Die automatisch bestimmte mixing time ist stark abhängig vom Stopkriterium, was auch von den Autoren gezeigt wurde. Mit ansteigendem SRR wird die detektierte mixing time geringer (Defrance u. a., 2009, S. 1077, Abb. 15).
Im Artikel (Defrance u. a., 2009, S. 1078, Abb. 17) sind die mixing times, die durch die Autoren selbst bestimmt wurden, in Abhängigkeit der Entfernung zwischen Quelle und Empfänger dargestellt. Es treten sehr geringe mixing times auf, wenn die Laufzeit der Impulsantworten (wie es in dieser Arbeit getan wurde) vor der Anwendung von Mat-

ching Pursuit entfernt wurden. Plausiblere mixing times finden Defrance u. a. (2009), wenn die Laufzeit der Impulsantworten nicht entfernt wird, was allerdings darauf hindeutet, dass die mixing time nach dem genannten Kriterium hauptsächlich durch die Laufzeit zwischen Quelle und Empfänger bestimmt wird. Auch der Entfernungszusammenhang zur mixing time, den Defrance u. a. (2009) fanden, ist durch die nicht entfernten Laufzeiten der Impulsantworten vor Anwendung des Matching Pursuit-Algorithmus zu erklären.

### 2.3.4 Hidaka u. a., 2007

Die Methode von Hidaka u. a. (2007) zur Bestimmung des Übergangszeitpunktes zwischen frühen und späten Anteilen einer Impulsantwort macht sich zu Nutze, dass die Korrelation zwischen beiden Anteilen sehr gering ist. Grob zusammengefasst, wird hier die Energie des Direkt- und Anfangsschalls mit der der nachfolgenden Schallanteile korreliert. Der Übergangszeitpunkt zwischen frühen und späten Anteilen der Impulsantwort ist erreicht, wenn die Korrelation einen speziellen geringen Wert annimmt.

Im Unterschied zu den bisher vorgestellten Methoden wird die mixing time im Frequenzbereich berechnet. Dazu wird die zeit- und frequenzabhängige Energie der Impulsantwort mit

$$E(t,\omega) = \left| \int_t^\infty h(\tau) e^{j\omega\tau} \right|^2 \tag{2.20}$$

bestimmt. Die Gleichung gibt die verbleibende Energie der Impulsantwort nach der Zeit $t$ an. Dabei ist $t = 0$ der Zeitpunkt, zu dem der Direktschall eintrifft. Eine Darstellung dieses Verlaufs ergibt das typische „Wasserfall“-Diagramm der frequenzabhängig abklingenden Nachhallenergie (siehe Abbildung 2.5).

Die Mittelung des Integrals für $t = 0$ über einen Frequenzbereich entspricht der Rückwärtsintegration der quadrierten Impulsantwort nach Schroeder (1965):

$$\langle E(t,\omega) \rangle = \frac{1}{\Delta\omega} \int_t^\infty h^2(\tau) d\tau \tag{2.21}$$

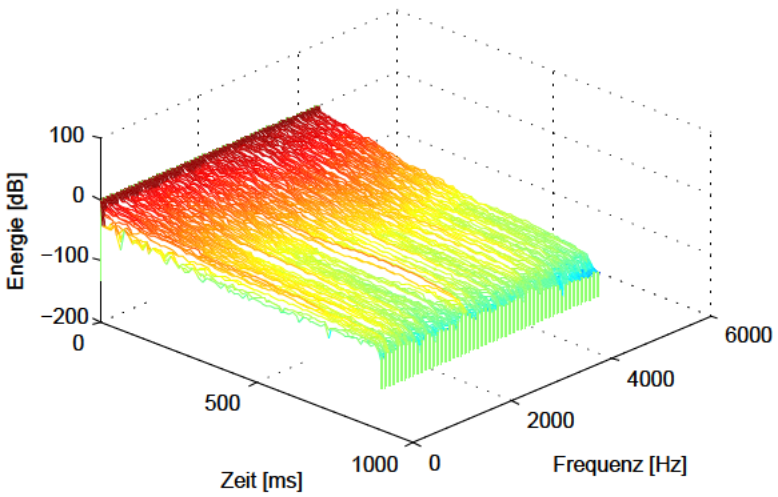



**Abbildung 2.5:** „Wasserfall"-Diagramm der frequenzabhängig abklingenden Energie einer Raumimpulsantwort

Eine Herleitung findet sich bei Hidaka u. a. (2007). $E(0, \Delta\omega)$ enthält den Direktschall und alle nachfolgenden Reflexionen, es handelt sich dabei um die (quadrierte) Übertragungsfunktion. Mit voranschreitender Zeit geht das gesamte Schallfeld in ein Schallfeld über, das nur von der Energie des Nachhallschalls dominiert wird. Um diesen Übergangszeitpunkt zu finden, nutzt Hidaka den Korrelationskoeffizienten nach Pearson, mit dem die Ähnlichkeit zwischen zwei Zuständen (in diesem Fall Zeitpunkten) untersucht werden kann. Es wird jeweils der für den frühen Zustand summierte frequenzabhängige Energieverlauf $E(0 \to \infty, \Delta\omega)$ mit dem über den gesamten folgenden Zeitraum summierten Verlauf $E(t \to \infty, \Delta\omega)$ für ein spezielles Frequenzband korreliert. Durch kontinuierliches Verschieben des Übergangszeitpunktes $t$ ergibt sich eine zeitabhängige Funktion des Korrelationsgrades $r(t)$, die die Ähnlichkeit zwischen der Energie des Ausgangszustandes mit der Energie ab irgendeinem späteren Zeitpunkt darstellt (siehe Gleichung (2.22)).

$$r(t) = \frac{\langle (E(0,\omega) - \mu(0))(E(t,\omega) - \mu(t)) \rangle}{\sqrt{\langle (E(0,\omega) - \mu(0))^2 \rangle \langle (E(t,\omega) - \mu(t))^2 \rangle}} \tag{2.22}$$

mit

$$\mu(t) = \langle E(t,\omega) \rangle \tag{2.23}$$

Der Übergangszeitpunkt in den stochastischen Nachhall ist nach Hidaka u. a. (2007) erreicht, wenn $r(t)$ einen speziellen geringen Wert erreicht, namentlich $r = e^{-1} = 0,367$. Dieser Wert wird gemäß Hidaka u. a. (2007) in der Akustik oder anderen Fachgebieten oft angewandt, um in stochastischen Prozessen geringere von höheren Korrelationen zu unterscheiden. Dazu werden im Artikel von Hidaka u. a. (2007) einige Quellen zitiert. Der Übergangszeitpunkt zwischen frühen und späten Anteilen der Impulsantwort zeigt demnach an, wann das Schallfeld nahezu stochastisch unabhängig vom Ausgangszustand $t = 0$ wird. Abbildung 2.6 zeigt beispielhaft für einen mittelgroßen Hörsaal den Verlauf der Korrelationsfunktion für die Oktave bei 500 Hz. Aus Gründen der Übersichtlichkeit werden die Verläufe der restlichen Oktaven zwischen 125 Hz und 16 kHz nicht mit dargestellt, der grundsätzliche Verlauf ist jedoch ähnlich.

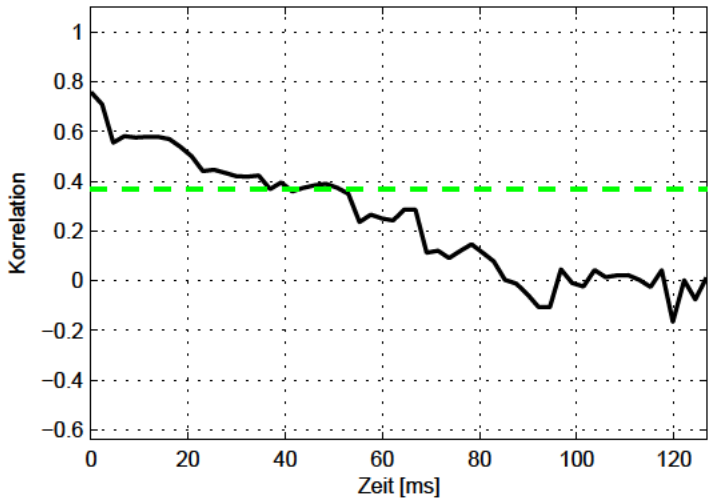



**Abbildung 2.6:** Korrelationsverlauf nach Hidaka u. a. (2007) (berechnet für die 500 Hz-Oktave) und Grenze zu geringer Korrelation $r = e^{-1}$ (grüne gestrichelte Linie), exemplarisch für einen mittelgroßen Hörsaal

## 2.4 Perzeptive Evaluation

Es existieren bisher nur wenige Untersuchungen zur perzeptiven mixing time. Meesawat und Hammershøi (2003) und Lindau u. a. (2007) führten Hörversuche zur Bestimmung der mixing time in einem kleinen und einem großen Hörsaal durch. Eine Untersuchung der perzeptiven mixing time für mehrere systematisch variierte Räume fehlte bisher.

Zudem wurden die genannten Hörversuche zwar mit gemessenen binauralen Raumimpulsantworten, jedoch nur mit statischer Auralisation durchgeführt.

Die Untersuchung von Meesawat und Hammershøi (2003) in einem kleinen Hörsaal ($V$=185 m$^3$, $T_{30,1kHz}$=0,72 s) zeigte, dass der stochastische Nachhall verschiedener Positionen und Richtungen eines Raumes tendenziell ausgetauscht werden kann. Der Übergangszeitpunkt zwischen frühen Reflexionen und stochastischem Nachhall wurde untersucht, indem verschiedene Kombinationen von frühen Reflexionen und Nachhallschwänzen mit dem 3-Alternative-Forced-Choice-Verfahren (3-AFC) getestet wurden. Dabei wurden die Kanäle links und rechts vertauscht und unterschiedliche Quell-/Empfänger-Konstellationen erzeugt, indem die Richtung des Kopfes oder die Position der Quelle und des Empfängers oder beides variiert wurde. Der Abstand zwischen Quelle und Empfänger wurde jedoch konstant gehalten. Keine der getesteten Varianten stellte sich als besonders kritisch heraus. Die perzeptive mixing time in diesem speziellen Raum ergab sich gemäß Meesawat und Hammershøi (2003) zu 40 ms. Dieser Wert erscheint jedoch fragwürdig, da die Mittelwerte der korrekten Antworten auch bei 80 ms noch deutlich über der Ratewahrscheinlichkeit lagen (siehe Meesawat und Hammershøi (2003, Abb. 7)). Zudem war der Stichprobenumfang sehr gering (4 Versuchspersonen), was eine ebenfalls geringe Teststärke nach sich zieht.

Da in größeren Räumen aufgrund der größeren mittleren freien Weglängen höhere mixing times erwartet werden können, führten Lindau u. a. (2007) einen Hörversuch zur Bestimmung des Schwellwertes der perzeptiven mixing time in einem großen Hörsaal ($V$=10000 m$^3$, $T_{30,1kHz}$=2 s) durch. Der Nachhallschall wurde einer anderen Kopfposition entnommen oder aus einer BRIR-Messung, bei der sich die Quelle erhöht hinten rechts befand (worst case-Situation mit unterschiedlicher Quell-Empfänger-Entfernung *und* -Richtung). Außerdem wurden die Audiobeispiele variiert: Es fanden Sprache und ein Schlagzeugsample Verwendung.
Die perzeptive mixing time für die Testbedingung, in der nur die Blickrichtung des Kopfes geändert wurde, stimmt in etwa mit den Ergebnissen von Meesawat und Hammershøi (2003) überein. Etwas höhere perzeptive mixing times wurden mit dem Schlagzeugsample erreicht. Zudem konnte festgestellt werden, dass die Versuchspersonen bei der Variante, für die der Nachhallschall von der Quelle hinten rechts entnommen wurde, empfindlicher waren. Einige Versuchspersonen konnten Unterschiede noch bis zu einer perzeptiven mixing time von 140 ms sicher erkennen. Die Autoren nennen das

modale Verhalten des Raumes als mögliche Ursache.
Der Hörversuch von Lindau u. a. (2007) wurde wiederum mit dem 3-AFC-Testverfahren durchgeführt, allerdings in einem adaptiven Schwellwertversuch. Die Anzahl der Versuchspersonen lag mit 23 höher als in der Untersuchung von Meesawat und Hammershøi (2003).

# 3 Methodik

Auf den folgenden Seiten wird die Methodik dargelegt, mit der eine geeignete Formel zur Vorhersage des perzeptiven Übergangszeitpunktes zwischen frühen Reflexionen und spätem Nachhall gefunden werden soll. Es wird die Auswahl der Räume, der Ablauf der binauralen Messungen sowie die perzeptive Evaluation erläutert.
Zudem werden Erfahrungen, die sich aus der praktischen Umsetzung der in Abschnitt 2.3 vorgestellten empirischen Prädiktoren ergaben, dargestellt.

## 3.1 Parameter der Raumauswahl

Hauptziel dieser Arbeit ist es, geeignete Prädiktoren der perzeptiven mixing time zu finden. Dazu wird der Einfluss des Raumvolumens und der Nachhallzeit bzw. der relativen Bedämpfung auf die perzeptive mixing time systematisch untersucht. Volumen und Nachhallzeit lassen sich nicht unabhängig voneinander variieren, so dass stattdessen versucht wurde, die Räume entsprechend des vom Volumen unabhängigen mittleren Absorptionsgrades zu variieren. Der mittlere Absorptionsgrad berechnet sich wie folgt (Kuttruff, 2004, S. 264):

$$\alpha_{mittel} = \frac{A}{S} \tag{3.1}$$

Die äquivalente Absorptionsfläche $A = \sum_{i=1}^{n} \alpha_i \cdot S_i$ wird dabei über die gesamte Raumoberfläche S gemittelt. Ohne Kenntnis der Absorptionsgrade der einzelnen Oberflächen kann die äquivalente Absorptionsfläche durch Umstellen der Gleichung zur Berechnung der Nachhallzeit nach Sabine berechnet werden (Cremer und Müller, 1978, S. 160):

$$A = 0,163 \cdot \frac{V}{T} \tag{3.2}$$

Die Faktoren Volumen und mittlerer Absorptionsgrad wurden jeweils dreifach gestuft, so dass insgesamt neun Räume Teil der Untersuchung waren. Die Auswahl der Räume ist in Tabelle 3.1 dargestellt. Die vollständige Variation der Räume ermöglicht eine unabhängige Untersuchung beider Einflussgrößen $V$ und $\alpha_{mittel}$ mittels einer Varianzanalyse.

| | $\alpha_{mittel}$ groß | $\alpha_{mittel}$ mittel | $\alpha_{mittel}$ gering | Ø $V$ |
|---|---|---|---|---|
| $V$ gering | **Raum 1 (EN324)**<br>V 216 m$^3$<br>$\alpha$ 0,36<br>$T_{30}$ 0,39 s | **Raum 2 (EN111)**<br>V 224 m$^3$<br>$\alpha$ 0,26<br>$T_{30}$ 0,62 s | **Raum 3 (EN190)**<br>V 182 m$^3$<br>$\alpha$ 0,17<br>$T_{30}$ 0,79 s | 207 m$^3$ |
| $V$ mittel | **Raum 4 (H0104)**<br>V 3300 m$^3$<br>$\alpha$ 0,28<br>$T_{30}$ 1,15 s | **Raum 5 (HE101)**<br>V 5179 m$^3$<br>$\alpha$ 0,23<br>$T_{30}$ 1,67 s | **Raum 6 (Teldex)**<br>V 3647 m$^3$<br>$\alpha$ 0,20<br>$T_{30}$ 1,83 s | 4042 m$^3$ |
| $V$ groß | **Raum 7 (UdK)**<br>V 8298 m$^3$<br>$\alpha$ 0,33<br>$T_{30}$ 1,52 s | **Raum 8 (H0105)**<br>V 8500 m$^3$<br>$\alpha$ 0,23<br>$T_{30}$ 2,08 s | **Raum 9 (JC-Kirche)**<br>V 7417 m$^3$<br>$\alpha$ 0,23<br>$T_{30}$ 2,36 s | 8072 m$^3$ |
| Ø $\alpha_{mittel}$ / Ø $T_{30}$ | 0,32 / 1,02 s | 0,24 / 1,46 s | 0,20 / 1,66 s | |

**Tabelle 3.1:** Volumen, mittlerer Absorptionsgrad und Nachhallzeit der ausgewählten Räume

Bei der Auswahl der Räume wurde darauf geachtet, dass der Abstand zwischen den Nachhallzeiten mindestens 10 % beträgt, damit ein Unterschied gerade wahrnehmbar ist. Die Unterschiedsschwelle der Nachhallzeit ist von ihrer Länge abhängig. Sehr kurze Nachhallzeiten erfordern eine größere prozentuale Änderung als lange Nachhallzeiten (Cremer und Müller, 1978, S. 406). Der Wert von 10 % ist nur bei den kleinen Räumen erforderlich, da diese eine geringe Nachhallzeit haben, bei den größeren Räumen würden 4 % Unterschied zwischen den Nachhallzeiten zur Unterscheidbarkeit ausreichen.

Eine Untersuchung des Übergangszeitpunktes zwischen frühen und späten Anteilen der Impulsantwort in mehr als 50 Opern- und Sinfoniesälen von Hidaka u. a. (2007) zeigte, dass die physikalische mixing time in rechteckförmigen Räumen maximal ist. Die regelmäßige Form und lange freie Weglängen lassen nach der Theorie von diffusen Schallfeldern in Räumen ebenfalls schuhkartonartige Räume am kritischsten erscheinen (vgl. Abschnitt 2.1). Die Raumauswahl beschränkt sich daher auf im Wesentlichen rechteckförmige Räume. Gekoppelte Räume, wie sie z. B. in Theatern vorkommen, wurden vermieden. Die ausgewählten Räume gehören meist zur TU Berlin, da diese am einfachsten (und kostengünstigsten) gemietet werden können. Die drei kleinen Räume sind das elektronische Studio der TU Berlin (Raum 1) und zwei kleine Seminarräume: EN111 und EN190 (Raum 2 und 3). Bei den Räumen mit mittlerem Volumen handelt es sich um die beiden Hörsäle H0104 (Raum 4) und HE101 (Raum 5) sowie den Aufnahmesaal der Teldex-Studios in Berlin (Raum 6). Die drei Räume mit großem Volumen sind der Konzertsaal der Universität der Künste (UdK) in Berlin (Raum 7), das Audimax (H0105) der TU Berlin (Raum 8) sowie die Jesus-Christus-Kirche in Berlin (Raum 9).

**Raum 1** Im elektronischen Studio des Fachgebiet Audiokommunikation (EN324) ist der Fußboden mit Filz ausgestattet, an zwei Wänden befinden sich Diffusoren, die beiden anderen Wände sind Fensterflächen, wovon eine von Vorhängen verdeckt ist. Die Decke besteht aus unregelmäßig angeordneten Lochplatten.

**Raum 2** Der Seminarraum EN111 verfügt über schallharte Wände und Fußboden. Die Decke ist teilweise mit absorbierendem Material verkleidet. Während der Messung standen Tische, Stühle und Schränke im Raum. Eine Seitenwand ist nahezu vollständig verglast.

**Raum 3** Im Seminarraum EN190 sind die Wände und der Fußboden wie auch in Raum 2 schallhart beschaffen, eine Seitenwand ist nahezu vollständig verglast. Der Raum ist mit einer Akustikdecke versehen und war zum Zeitpunkt der Messung möbliert.

**Raum 4** Der Hörsaal H0104 ist mit schallhartem Fußbodenbelag und gepolsterten Holzstühlen im Publikumsbereich ausgestattet. Die Wände sind im unteren Bereich absorbierend, darüber größtenteils reflektierend beschaffen (Behrens, 2007). Im Hörsaal befindet sich das Wellenfeldsynthesesystem der TU Berlin.

**Raum 5** HE101 ist ebenfalls ein Hörsaal der TU Berlin. Die Wände sind mit Holzpaneelen verkleidet. Der Fußbodenbelag ist schallhart. Im Zuhörerbereich befindet sich Holzgestühl. Die Rückwand hinter der Bühne ist mit Diffusoren ausgestattet.

**Raum 6** Der Fußboden im Aufnahmesaal der Teldex-Studios ist mit Parkett belegt. Die Wände sind nach Augenschein schallhart mit diffundierenden Elementen ausgestattet. Zum Zeitpunkt der Messung befanden sich gepolsterte Stühle im Raum.

**Raum 7** Der Konzertsaal der UdK Berlin verfügt über gepolsterte Stühle. Die Wände sind teilweise absorbierend und teilweise schallhart gestaltet. Die Decke ist mit Holz verkleidet.

**Raum 8** Im Audimax der TU Berlin (H0105) befinden sich an den Seitenwänden Ziegel mit integrierten Helmholtzresonatoren. Die Rückwand ist holzvertäfelt. Die Decke ist weithin reflektierend ausgekleidet. Der Fußboden ist größtenteils mit schallharten Materialien belegt. Die Publikumsfläche besteht aus Holzstühlen mit gepolsterter Rückenlehne.

**Raum 9** Die Jesus-Christus-Kirche in Berlin-Dahlem verfügt über einen schallharten Fußboden. Die Seitenwände sind ebenfalls schallhart bzw. mit Fensterflächen belegt. Bei der Decke handelt es sich um eine Holzdecke. Im Publikumsbereich befinden sich Holzstühle.

Zur Verdeutlichung der Raumauswahl zeigt Abbildung 3.1 die maßstabsgetreuen Grundrisse der untersuchten Räume.

In Abbildung 3.2 ist die zweidimensionale Faktorstufung über dem Raumvolumen und dem mittleren Absorptionsgrad aufgetragen. Auf der x-Achse ist der zunehmende mittlere Absorptionsgrad bzw. die sinkende Nachhallstärke aufgetragen, auf der y-Achse das zunehmende Volumen. Die Stufen der Räume sind relativ gut zu erkennen. Jeweils drei Räume bilden eine Volumenstufe, die in drei Absorptionsgradstufen unterteilt ist. Die Abstufungen innerhalb des 3x3-Designs sind für eine Auswahl realer Räume recht regelmäßig. Eine Ausnahme bildet die Jesus-Christus-Kirche (Raum 9), deren mittlerer Absorptionsgrad etwas zu hoch ist und ungefähr dem von Raum H0105 (Raum 8) entspricht. Die Absorptionsgradstufe zwischen diesen Räumen ist sehr gering. Nichtsdestotrotz hat die Jesus-Christus-Kirche (Raum 9) die längste Nachhallzeit innerhalb des Untersuchungsschemas.

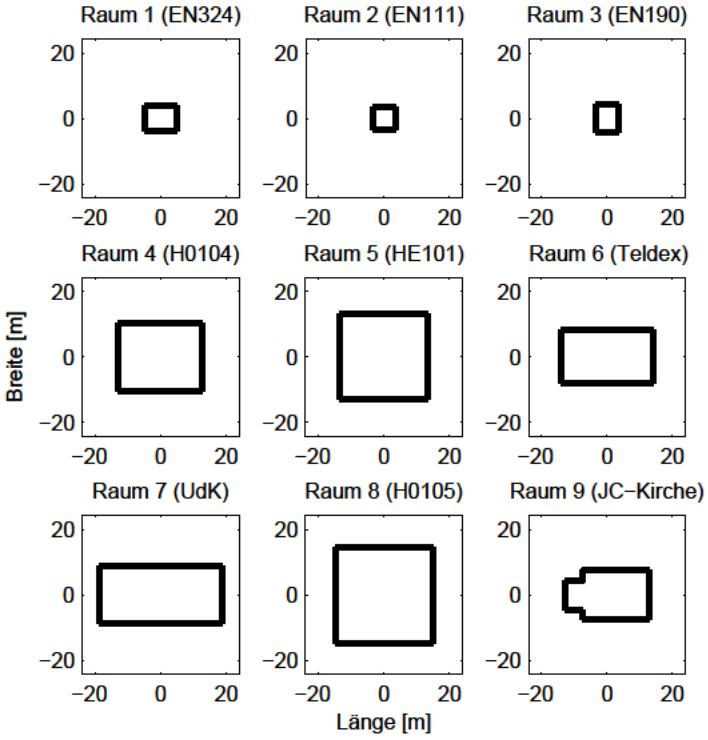



**Abbildung 3.1:** Maßstabsgetreue Grundrisse der neun Räume

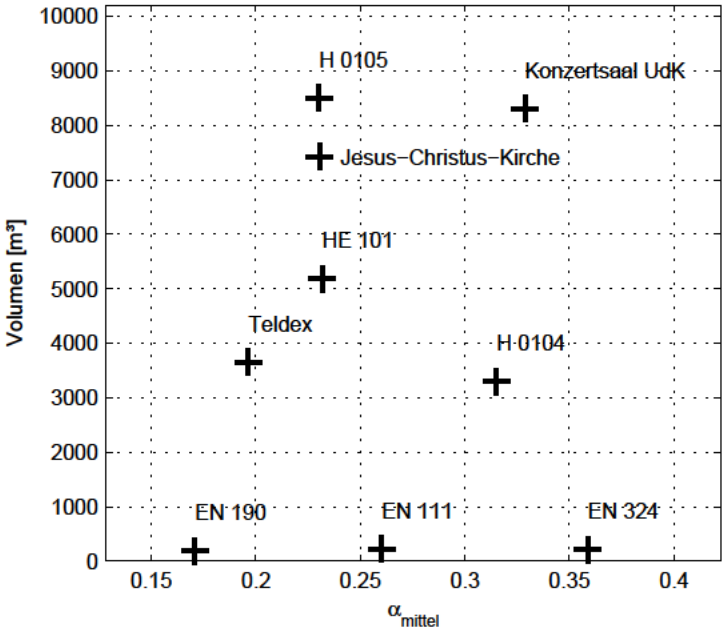



**Abbildung 3.2:** Stimuluskontinuum

## 3.2 Messdatenerhebung

Zur Ermöglichung des interaktiven Hörversuchs mussten in allen neun Räumen binaurale Raumimpulsantworten gemessen werden. Zudem musste die Nachhallzeit und das Volumen bestimmt werden, um die Eignung des Raumes für das Untersuchungsschema sicherzustellen. Die Messungen waren in zwei Abschnitte unterteilt: erst wurde die Nachhallzeit und der Hallradius bestimmt, dann folgte die Messung der binauralen Raumimpulsantworten. Als Messsignal dienten jeweils Sweeps.
Für beide Messungen wurde als Quelle ein frequenzgangkompensierter 3-Wege-Dodekaeder verwendet, der an der RWTH Aachen entwickelt wurde.[1]
Der Lautsprecher wurde auf einen linearen Frequenzgang und Kugelcharakteristik zwischen 40 Hz und 17 kHz mit einer Genauigkeit von ±3 dB eingestellt, um einen möglichst großen Frequenzbereich der BRIRs zu erhalten. Dieser ist notwendig für die spätere Auralisation. Der Lautsprecher wurde auf der Bühne bzw. dem vorderen Drittel entlang der Längsachse des Raumes positioniert (siehe Abbildung 3.3).

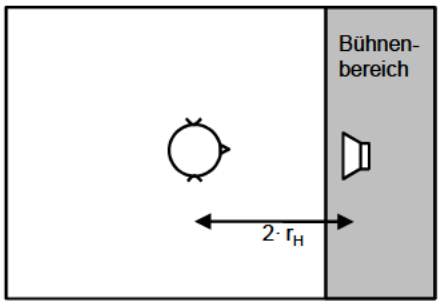



**Abbildung 3.3:** Messaufbau: Platzierung von FABIAN und Lautsprecher innerhalb der Räume

Die Nachhallzeit wurde durch die Messung von Raumimpulsantworten mit einem Messmikrofon (Behringer ECM 8000) an drei Mikrofonpositionen und anschließender Schroederintegration (Schroeder, 1965) mit der Messsoftware Monkey Forest® bestimmt. Das Mikrofon wurde im Diffusfeld aufgestellt, jeweils im linken, rechten und mittleren Raumbereich. Die Position des Lautsprechers wurde nicht verändert. Die Nachhallzeit wurde über die drei Messpositionen und die Oktaven zwischen 125 Hz und 4 kHz gemittelt. Die Ergebnisse der Nachhallzeitmessungen sind in Abbildung 3.4 und Tabelle B.2 zu finden. Alle Nachhallzeiten wurden im unbesetzten Zustand gemessen.

[1] Das Vorgängermodell des verwendeten Lautsprechers wurde von Behler (2000) beschrieben.

Der Hallradius wurde über die bekannte Formel $r_h = 0,057 \cdot \sqrt{\frac{V}{T}}$ bestimmt. Zur Bestimmung der Raumabmessungen wurde ein Laserentfernungsmessgerät verwendet.

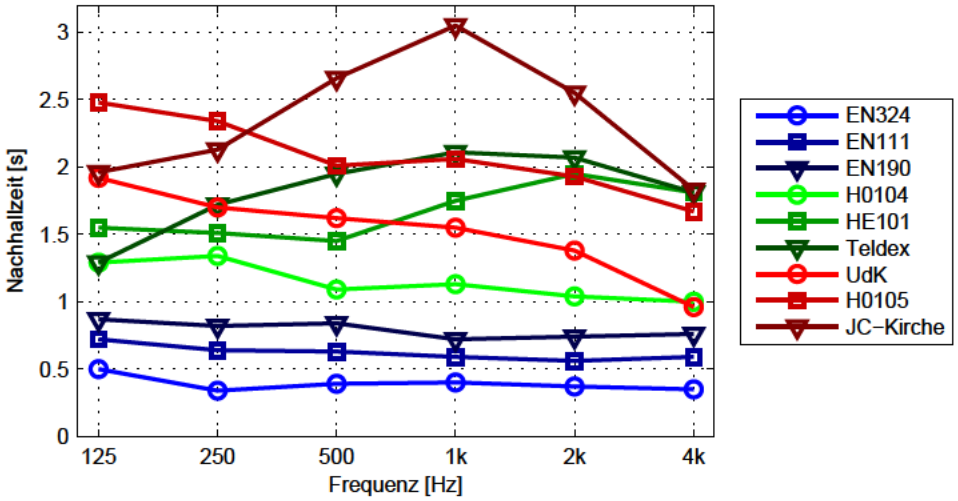



**Abbildung 3.4:** Gemessene Nachhallzeiten der neun untersuchten Räume. Die verschiedenen Farben kennzeichnen die drei Volumenstufen, die Farbabstufungen bzw. Symbole kennzeichnen die Absorptionsgradstufen innerhalb einer Volumenstufe.

Im zweiten Abschnitt wurden die binauralen Raumimpulsantworten mit dem Kunstkopftorsosimulator FABIAN gemessen. Der Kunstkopf und Torso sind den Abmessungen eines menschlichen Kopfes bzw. Oberkörpers nachempfunden. FABIAN verfügt über Mikrofone am Eingang der Ohrkanäle und ein bewegliches Halsgelenk, das eine automatische Bestimmung der BRIRs für verschiedene Kopfpositionen ermöglicht (Lindau, 2006).
Der Kunstkopftorsosimulator wurde auf einem repräsentativen Hörerplatz ungefähr mittig im Raum, aber etwas außerhalb der Raumachse im doppelten Hallradius zum Lautsprecher positioniert. In diesem Abstand kann von einem stochastischen Schallfeld ausgegangen werden (Kuttruff, 1991). FABIAN wurde frontal zum Lautsprecher ausgerichtet. Die BRIRs wurden mit einer Auflösung von 1° zwischen ±80° für horizontale Kopfpositionen gemessen. Das entspricht dem maximalen Bewegungsradius des Kopfes, den ein Mensch unter normalen Abhörbedingungen ausnutzen kann.

Den Messaufbau verdeutlicht die Abbildung 3.5. Der Messlaptop war für die Erzeugung der Sweeps, die Berechnung und Speicherung der Impulsantworten und die automati-

sche Steuerung des Halsgelenkes des Kunstkopftorsosimulators verantwortlich. Eine genaue Aufstellung der Geräte und deren technische Spezifikationen sind bei Lindau (2006, S. 216-217) verzeichnet. Mit Hilfe des Equalizers wurde die Frequenzgangkompensation des Lautsprechers vorgenommen.

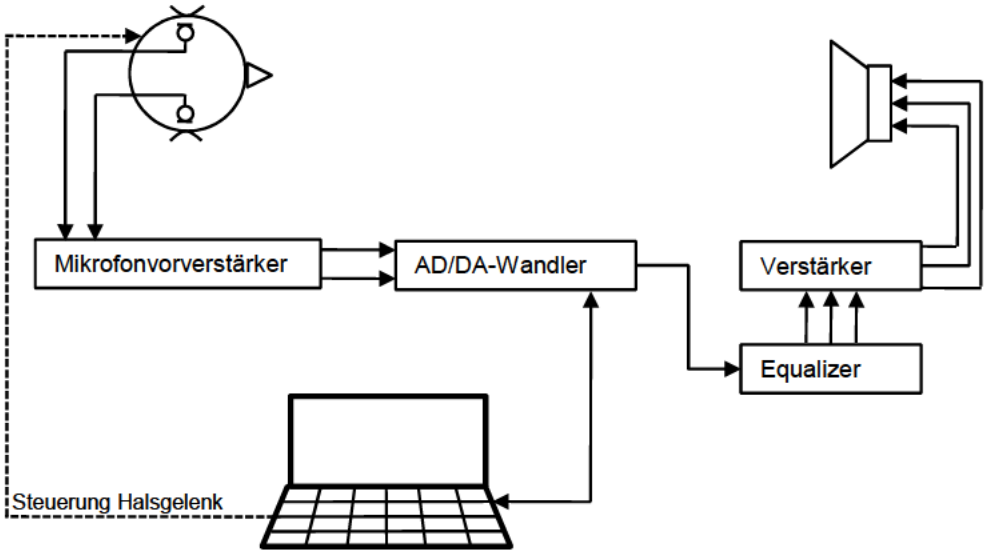
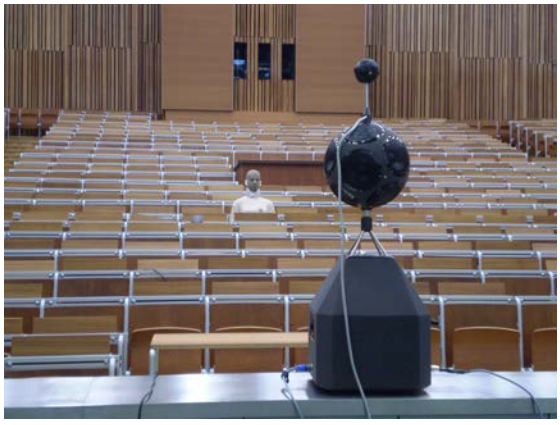



**Abbildung 3.5:** Messaufbau: verwendete Geräte

**Abbildung 3.6:** Kunstkopftorsosimulator und Lautsprecher im Detail

**(a)** Nachhallzeitmessung im Hörsaal H0104 (Raum 4)

**(b)** Nachhallzeitmessung im elektronischen Studio (Raum 1)



**(c)** BRIR-Messung im Hörsaal HE101 (Raum 5)

**(d)** BRIR-Messung in der Jesus-Christus-Kirche (Raum 9)

**Abbildung 3.7:** Beispiele einiger Messaufbauten

## 3.3 Ergebnisse aus dem Umgang mit empirischen Prädiktoren

Alle in Abschnitt 2.3 vorgestellten empirischen Methoden zur Bestimmung der mixing time wurden in Matlab® implementiert. Die Ergebnisse der Berechnungen wurden für die spätere Regressionsanalyse genutzt, die die Eignung der verschiedenen Prädiktoren zur Vorhersage der perzeptiven mixing time untersuchen sollte.
Bei der praktischen Umsetzung der Methoden stellten sich einige Schwierigkeiten heraus, die im folgenden Abschnitt neben der Vorgehensweise bei der Implementierung erläutert werden.
Vor der Anwendung der verschiedenen Algorithmen wurden die Impulsantworten laufzeitbefreit, damit die mixing times untereinander vergleichbar und unabhängig von der Entfernung zwischen Quelle und Empfänger sind. Alle Algorithmen werden schrittweise im Zeitbereich auf die Impulsantworten angewendet. Um Rechenzeit zu sparen, wurden die Impulsantworten am Ende gekürzt, so dass deren berücksichtigte Dauer noch etwa der doppelten erwarteten mixing time entsprach.

### 3.3.1 Abel und Huang, 2006

Für die Berechnung der mixing time in Matlab® mit Hilfe der Echodichte nach Abel und Huang (2006) wurde ein Rechteckfenster verwendet. Der Empfehlung der Autoren zufolge wurde eine Fensterlänge von 1024 Samples (23 ms) gewählt. Die Länge garantiert, dass genügend Reflexionen im Fenster enthalten sind und die Standardabweichung nicht zu groß ist. Das Fenster lag symmetrisch um den Zeitpunkt, für den die Echodichte berechnet wurde. Das resultiert in einer steigenden bzw. fallenden Fensterlänge zu Beginn und am Ende der Impulsantwort. Daraus folgt, dass die Fensterlänge ab 23 ms und bis 23 ms vor Ende der Impulsantwort konstant ist.

Es zeigte sich, dass die mixing times nach Abel und Huang (2006) (Kriterium 1) innerhalb eines Raumes stark variieren. Zum Vergleich dienten drei Impulsantworten (IR), die an verschiedenen Positionen im Raum gemessen wurden und eine binaurale Raumimpulsantwort (BRIR), die mit einem Kunstkopftorsosimulator mit Blickrichtung nach vorn gemessen wurde (Beschreibung der Messungen siehe Abschnitt 3.2). Die Abtast-

frequenz der Impulsantworten betrug 44100 Hz.

In Abbildung 3.8 sind die mixing times nach Abel und Huang (2006) von fünf Impulsantworten im Diffusfeld von neun verschiedenen Räumen dargestellt (zur Beschreibung der Räume siehe Abschnitt 3.1). Es ist kein Zusammenhang zwischen der Höhe der mixing times und der Art der Impulsantwort bzw. der Messposition erkennbar, d. h. keine der Impulsantworten lieferte über alle Räume hinweg besonders hohe oder besonders niedrige mixing times.

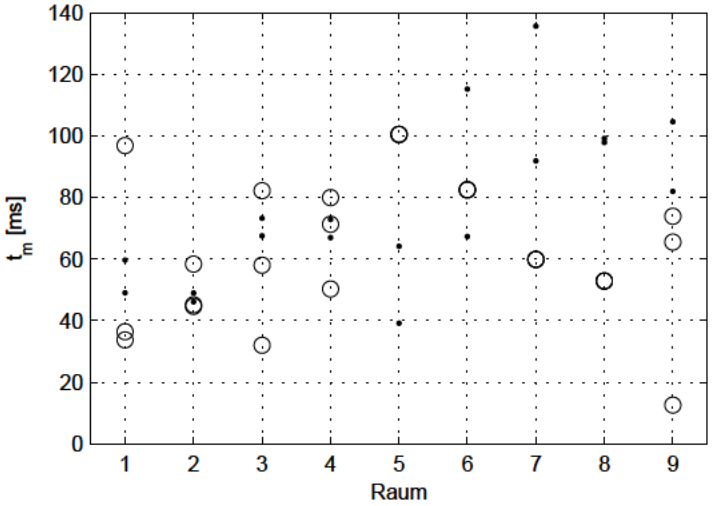



**Abbildung 3.8:** Ortsvarianz der $t_m$-Werte nach Abel und Huang (2006)
• BRIR (links und rechts)
○ IR (3 Messpositionen)

Die Analyse verschiedener Echodichteprofile zeigte, dass nicht alle Impulsantworten den theoretisch erwarteten kontinuierlich zunehmenden Verlauf der Echodichte, wie er in Abschnitt 2.3.1 beschrieben wurde, hervorrufen. Einige Beispiele für untypische Verläufe sind in Abbildung 3.9 dargestellt. Der Verlauf des Echodichteprofils steigt teilweise schon sehr früh an und fällt dann wieder ab. Zudem sind auch beim Vergleich der Echodichteprofile verschiedener Messpositionen eines Raumes Abweichungen der Verläufe untereinander erkennbar.

Der charakteristische Verlauf des Echodichteprofils ließ vermuten, dass eine visuelle Bestimmung der mixing time eventuell konsistentere mixing times innerhalb eines Raumes liefern würde.

Die visuelle Ablesung der mixing times erfolgte anhand eines neuen Kriteriums, das nach folgendem Muster bestimmt wurde: Nach dem nahezu senkrechten Anstieg wur-

den die folgenden Maxima des Echodichteprofils interpoliert (im Zweifel die beiden größten Maxima, auch wenn dazwischen noch andere Peaks lagen). Der Schnittpunkt der Interpolationsgeraden mit $y = 1$ entsprach der visuell bestimmten mixing time. Die visuelle Bestimmung der mixing time wurde nur für die BRIRs durchgeführt, da nur diese Impulsantworten für die spätere perzeptive Evaluation benutzt wurden.

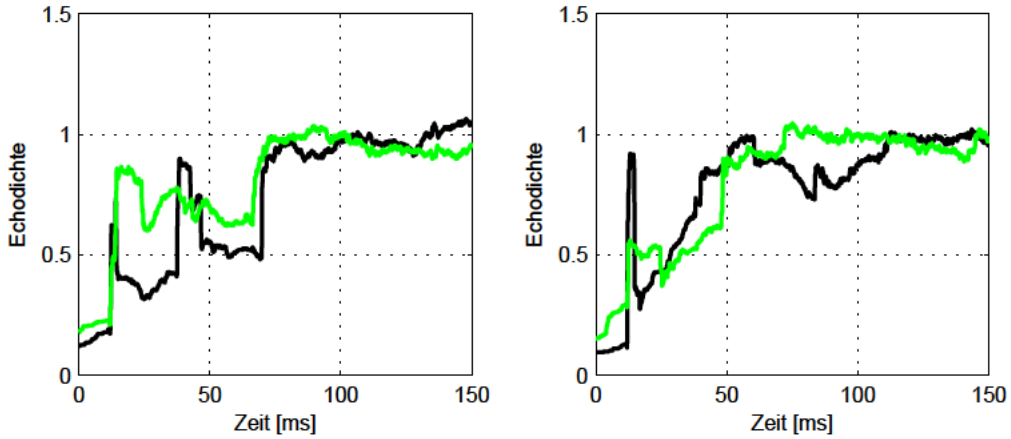



**Abbildung 3.9:** Beispielhafte Echodichteprofile. Alle Berechnungen wurden mit einer Fensterlänge von 1024 Samples durchgeführt.
links: untypische Verläufe
rechts: unterschiedliche Verläufe bei verschiedenen Messpositionen

Es konnte festgestellt werden, dass die visuelle Bestimmung der mixing time aus den Echodichteprofilen weniger variante Ergebnisse liefert (siehe Abbildung 3.10). Im Abschnitt 5 wird jedoch gezeigt, dass die visuell bestimmten physikalischen mixing times weniger zur Prädiktion der perzeptiven mixing time geeignet sind, als die automatisch bestimmten mixing times.

Für das zweite Kriterium nach Abel und Huang (2006), wonach das Echodichteprofil 1-$\sigma_{late}$ überschreiten soll, wurde festgelegt, dass die letzten 20 % des Echodichteprofils als 'spät' aufgefasst werden.

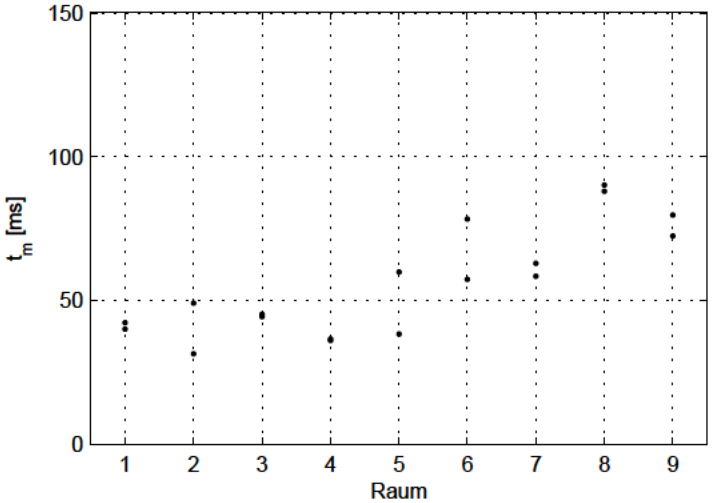



**Abbildung 3.10:** Ortsvarianz der visuell bestimmten $t_m$-Werte aus den Echodichteprofilen der BRIRs (links und rechts) nach Abel und Huang (2006)

### 3.3.2 Stewart und Sandler, 2007

Die Berechnungen der physikalischen mixing time nach Stewart und Sandler (2007) in Matlab® wurden wie auch bei der Echodichtebestimmung nach Abel und Huang (2006) mit einem 1024 Samples (23 ms) dauernden Rechteckfenster durchgeführt, dessen Länge zu Beginn ansteigt und am Ende der Impulsantwort kürzer wird. Die Plots der Kurtosis wurden zur besseren Darstellung normiert. Die physikalische mixing time wurde auf den Zeitpunkt festgelegt, zu dem die Kurtosis das erste Mal kleiner als 0 wurde.

Die genauere Untersuchung der für diese Arbeit gemessenen Impulsantworten (siehe Abschnitt 3.2) mit dem Algorithmus von Stewart und Sandler (2007) zeigte, dass sich auch mit dieser Methode nicht immer eindeutige Ergebnisse ergeben. Die Kurtosis folgt nicht immer dem theoretisch erwarteten Verlauf, der im Abschnitt 2.3.2 erläutert wurde. Die automatische Detektion liefert wie das Kriterium von Abel und Huang (2006) mixing time-Werte, die innerhalb eines Raumes stark variieren (siehe Abbildung 3.11). Beispielhaft sind in Abbildung 3.12 einige Verläufe der Kurtosis dargestellt.

Bei vielen Impulsantworten ist der Kurtosisverlauf schon sehr früh einmal kleiner 0 und steigt dann wieder an. Die automatische Detektion findet den ersten Wert, der kleiner 0 ist als mixing time. Es konnte keine über alle Impulsantworten konsistente Gemeinsamkeit der Kurtosisverläufe nach dem frühen Abfall festgestellt werden. Viele

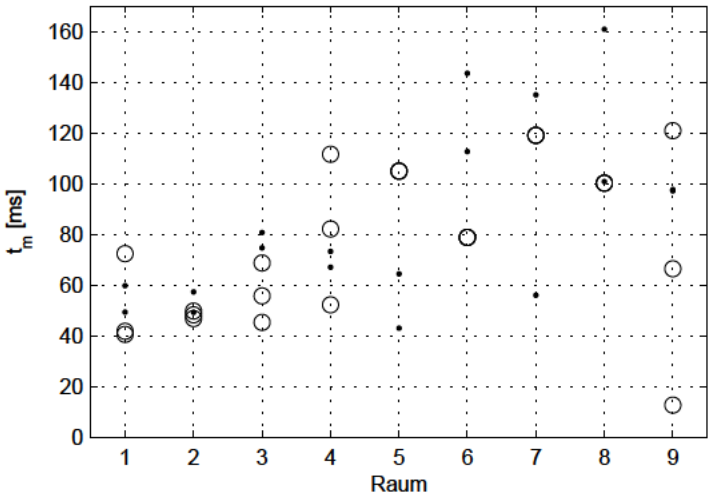



**Abbildung 3.11:** Ortsvarianz der $t_m$-Werte nach Stewart und Sandler (2007)
• BRIR (links und rechts)
○ IR (3 Messpositionen)

der Kurtosisverläufe springen auf einen relativ konstanten Wert. Da dies aber nicht bei allen untersuchten Impulsantworten der Fall war, konnte kein eindeutiges neues automatisches Kriterium festgelegt werden.

Deshalb wurde wiederum ein Kriterium festgelegt, mit dem die mixing time visuell bestimmt werden konnte. Nach dem nahezu senkrecht fallenden Kurvenverlauf wurden die nachfolgenden Minima interpoliert (im Zweifel die letzten beiden Minima vor dem waagerechten Verlauf der Kurtosis). Der Schnittpunkt der Interpolationsgeraden mit $y = 0$ entsprach der abgelesenen mixing time. Falls der Verlauf nach dem stark fallenden Kurvenabschnitt nur waagerecht verlief (ohne Extrempunkte), dann lag die visuell bestimmte mixing time im Knickpunkt zwischen steilem Abfall und waagerechtem Verlauf der Kurtosis.

Die mixing times wurden nur für die linken und rechten BRIRs der Vorne-Richtung abgelesen, da nur diese Datensätze in der perzeptiven Evaluation Verwendung fanden. Die Ortsvarianz war auch hier geringer als mit der automatischen Methode. Wie in Abschnitt 5.2 gezeigt wird, sind die visuell bestimmten mixing times trotzdem nicht besser zur Prädiktion der perzeptiven mixing time geeignet als die automatisch detektierten mixing times.

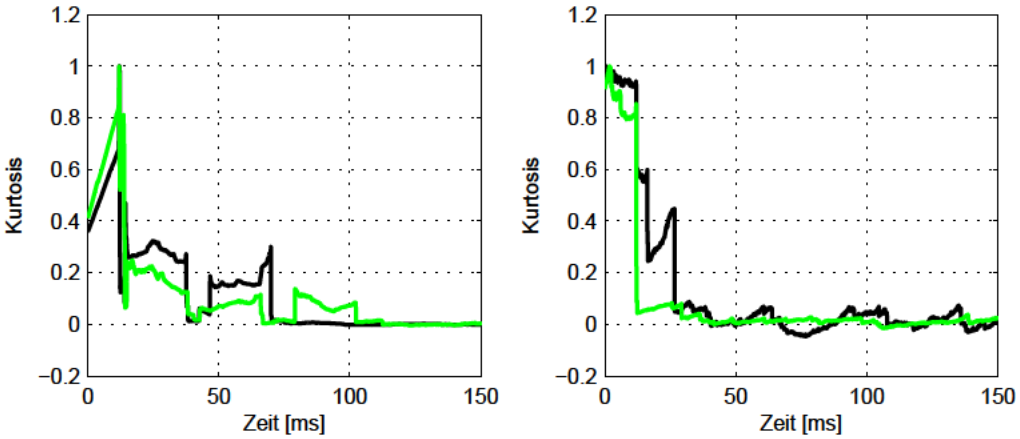



**Abbildung 3.12:** Beispielhafte Kurtosisverläufe. Alle Berechnungen wurden mit einer Fensterlänge von 1024 Samples durchgeführt.
links: untypische Verläufe
rechts: unterschiedliche Verläufe bei verschiedenen Messpositionen

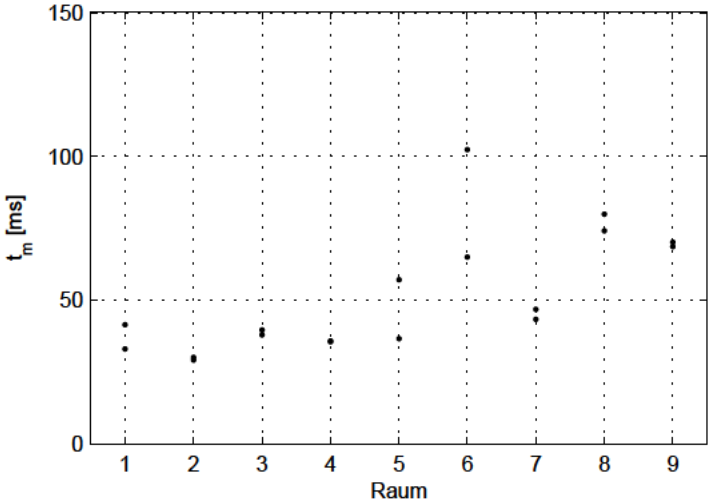



**Abbildung 3.13:** Ortsvarianz der visuell bestimmten $t_m$-Werte aus den Kurtosisverläufen der BRIRs (links und rechts) nach Stewart und Sandler (2007)

### 3.3.3 Defrance u. a., 2009

Der Matlab®-Code zur Zerlegung der Impulsantwort mit Hilfe von Matching Pursuit wurde freundlicherweise von Guillaume Defrance zur Verfügung gestellt. Die Implementierung der Energiekompensation sowie der Berechnung der äquivalenten Dauer des Direktschalls erfolgte durch Alexander Lindau.

Die berechneten mixing times nach Defrance u. a. (2009) sind im Gegensatz zu den bisher betrachteten Methoden nicht nur abhängig vom Ort, sondern zudem auch bis auf eine Ausnahme sehr gering. Abweichend zum Vorgehen von Defrance u. a. (2009) wurden die Impulsantworten laufzeitbefreit.
Abbildung 3.14 zeigt die berechneten Werte (ein Ausreißer in Raum 1 wurde nicht mit dargestellt (324 ms)).

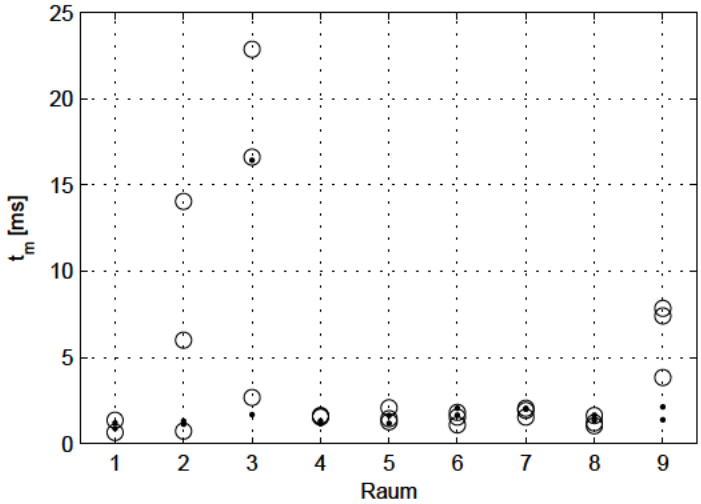



**Abbildung 3.14:** $t_m$-Werte nach Defrance u. a. (2009), 1 Ausreißer in Raum 1 nicht dargestellt
- • BRIR (links und rechts)
- ○ IR (3 Messpositionen)

Die Werte liegen größtenteils unter 5 ms, was vermuten lässt, dass die automatische Detektion nicht den Übergang in den stochastischen Nachhall findet. Das könnte daran liegen, dass das Stopkriterium für den Matching Pursuit-Algorithmus (SRR = 5 dB) zwar von den Autoren vorgeschlagen wurde, jedoch nur auf Basis eines Raumes ermittelt wurde. Die mixing time ist stark vom Stopkriterium abhängig, was auch von den Autoren gezeigt wurde. Die mixing times, die Defrance in seinem Artikel detek-

tiert, sind ebenfalls sehr gering, wenn die Laufzeit vor der Anwendung der Matching Pursuit-Methode entfernt wurde (siehe auch Abschnitt 2.3.3).

Um für den späteren Vergleich mit perzeptiven mixing times plausiblere Werte zu erhalten, wurden die mixing times durch das visuelle Kriterium, das von Defrance beschrieben wurde, anhand der CNAs abgelesen. Das führte zwar grundsätzlich zu höheren mixing times, allerdings war der Übergangszeitpunkt zwischen einem kubischen und einem linearen Verlauf teilweise sehr schwer zu detektieren (siehe Abbildung 3.15). Da zudem der Aufwand sehr hoch ist und die abgelesenen mixing times nicht reliabel[2] sind, wurde für die weitere Betrachtung von dieser Methode abgesehen.

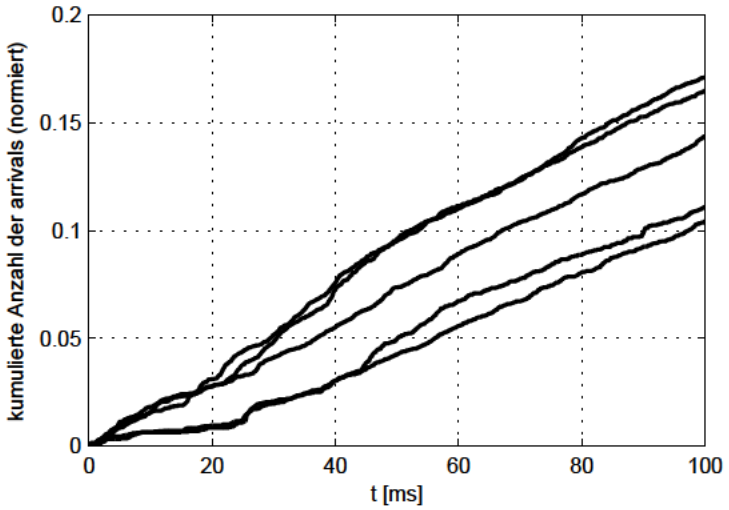



**Abbildung 3.15:** Normierte CNAs (5 Messpositionen in Raum 5)

### 3.3.4 Hidaka u. a., 2007

In Matlab® wurde die zeit- und frequenzabhängige Energie (Gleichung (2.20)) implementiert, indem die schnelle Fouriertransformation (FFT) auf entsprechende Zeitausschnitte der Impulsantwort angewendet und anschließend der Betrag quadriert wurde. Um die Rechenzeit der FFT zu verkürzen, wurden an die Impulsantwort Nullen angehängt, um eine Länge, die einer Potenz von 2 entspricht, zu erhalten.
Der Algorithmus von Hidaka liefert frequenzabhängige mixing times, weil die Korrelation der Energie für einzelne Oktaven durchgeführt wurde. Hidakas Berechnungen der

[2] Die Werte sind stark von der ablesenden Person abhängig und auch durch mehrmaliges Ablesen einer Person nicht eindeutig wiederholbar.

mixing time wurden für die Oktaven von 125 Hz bis 4 kHz durchgeführt. Da der Rechenaufwand nicht erheblich größer wurde, enthält die Berechnung der mixing time nach Hidaka für diese Arbeit alle Oktaven von 125 Hz bis 16 kHz. Für die spätere perzeptive Evaluation sollte möglichst der ganze Hörbereich abgedeckt werden. Die Energie der gesamten Impulsantwort wurde mit Hilfe der Matlabfunktion *corr()* mit der Energie der Impulsantwort, beginnend zu späteren Zeitpunkten, korreliert. Die Zeitauflösung wurde, um Rechenzeit zu sparen auf 100 Samples gesetzt, was 2,3 ms entspricht. Die mixing time ist der Zeitpunkt, zu dem die Korrelationsfunktion das erste Mal den Wert 0,367 unterschreitet bzw. gleich ist.

Wie auch bei den anderen bisher vorgestellten Algorithmen unterscheiden sich die berechneten mixing times nach der Methode von Hidaka u. a. (2007) abhängig von der Messposition und der Art der Impulsantwort untereinander (siehe Abbildung 3.16).

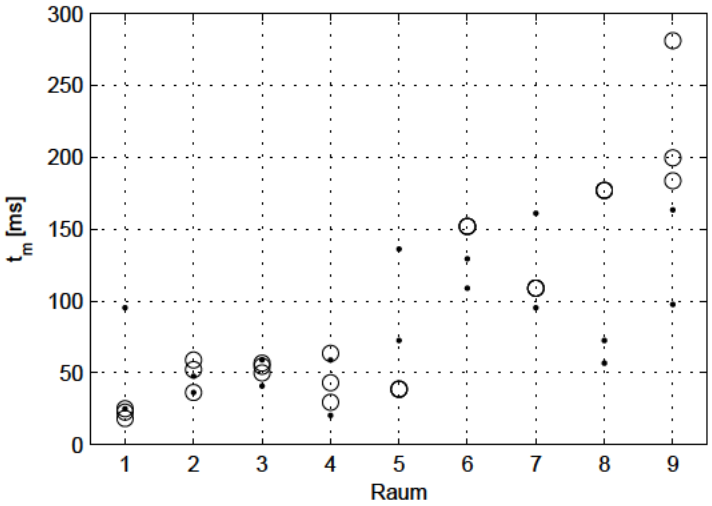



**Abbildung 3.16:** Ortsvarianz der $t_m$-Werte der 500 Hz-Oktave nach Hidaka u. a. (2007)
- • BRIR (links und rechts)
- ○ IR (3 Messpositionen)

Über alle Räume konstante Zusammenhänge zwischen den Werten und den berechneten Oktaven konnten nicht gefunden werden, ebenso nicht zwischen der Messposition bzw. der Art der Impulsantwort und den mixing times. Wie auch bei den anderen Methoden entsprechen die Verläufe der Korrelationsfunktion nur selten dem theoretischen Verlauf, der im Artikel von Hidaka vorgestellt wird. Die Korrelation fällt nicht stetig, sondern springt teilweise wieder über den kritischen Wert von 0,367 (siehe Abbildung 3.17).

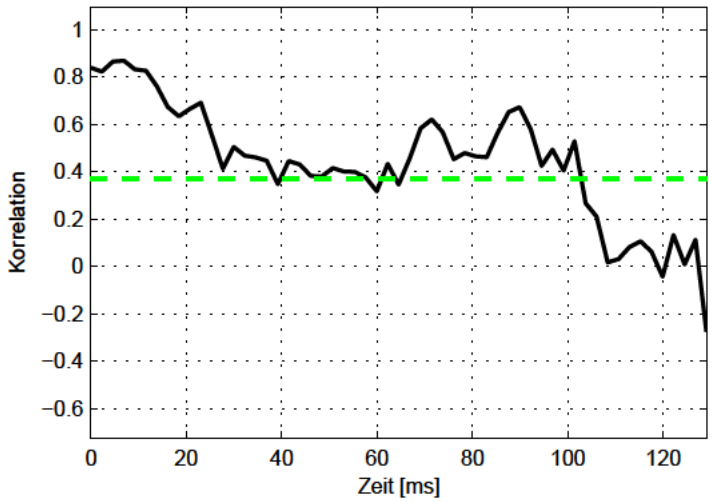



**Abbildung 3.17:** Beispiel für einen untypischen Verlauf der Korrelationsfunktion nach Hidaka u. a. (2007) und Grenze zu geringer Korrelation $r = e^{-1}$ (grüne gestrichelte Linie), berechnet für die 500 Hz-Oktave

## 3.4 Hörversuch

Um einen geeigneten Prädiktor für die perzeptive mixing time zu finden, musste ein Hörversuch durchgeführt werden. Hauptsächlicher Anwendungsbereich der Ergebnisse am Fachgebiet Audiokommunikation wird in der dynamischen Binauralsynthese liegen. Daher wurde im Hörversuch die dynamische Binauraltechnik genutzt. Es handelt sich um einen Schwellwertversuch, bei dem der Überblendzeitpunkt zwischen frühen und späten Anteilen einer Impulsantwort variiert wird. Der Versuchsaufbau wird im folgenden Abschnitt dargelegt.

### 3.4.1 Vorversuche

Wie in Abschnitt 1 erläutert, wird bei der dynamischen Binauralsynthese nur der erste Teil der Impulsantworten entsprechend der aktuellen Kopfposition des Hörers nachgeführt, der späte Nachhall wird aus einer statischen Kopfposition entnommen. Die Faltung muss nur für den frühen Anteil der Impulsantworten ständig aktualisiert werden. Im Hörversuch konnte diese Strategie nur in der Simulation des Testschallfeldes

verfolgt werden, in der der Übergangszeitpunkt in den späten Nachhall variiert wurde, mit dem Ziel, durch den Vergleich mit einer Referenzsituation die perzeptive mixing time der jeweiligen Räume zu finden. Zur Simulation der Referenzsituation musste die BRIR in vollständiger Länge in Echtzeit mit dem nachhallfreien Quellsignal gefaltet und der aktuellen Kopfposition nachgeführt werden. Demzufolge waren die Anforderungen an den Rechner sehr hoch.

In Vorversuchen wurde überprüft, welche perzeptive Genauigkeit für den Hörversuch notwendig ist und welche Beschränkungen sich technisch ergeben. Aus beiden Vorgaben musste ein Kompromiss gefunden werden.
Es wurde getestet, zu welchen Überblendzeitpunkten der Übergang der dynamischen BRIR-Nachführung in eine statische von den Vorversuchspersonen nicht mehr wahrgenommen werden konnte. Als Vorversuchspersonen dienten vier Institutsangehörige. Besonders die untere Grenze der Überblendung stellte sich als kritisch heraus, da die Probanden im Vorversuch nur geringe Schwellwerte erreichten. Technische Beschränkungen mussten in diesem Bereich besonders untersucht werden.
Die Anzahl und Lage der Stufen und die Auflösung, d. h. die Schrittweite zwischen den Stufen, mussten unter Beachtung der daraus resultierenden Genauigkeit des Schwellwertes und der vermuteten Lage des Schwellwertes festgelegt werden. Insgesamt standen 16 mögliche Stufen zur Verfügung.[3]
Um die Unabhängigkeit der Übertragungseigenschaften der Simulation von der Variation der unabhängigen Variablen *mixing time* sicherzustellen, wurden Validierungsmessungen durchgeführt. Es stellte sich heraus, dass bei sehr frühen Überblendzeiten Frequenzgangsinvarianzen auftraten. Aus diesem Grund wurden die frühestmöglichen artefaktfreien Überblendzeitpunkte für jeden Raum einzeln festgelegt. Die Artefakte traten vor allem bei den kleinen Räumen auf, weshalb eine unerwartete Wechselbeziehung der blockbasierten Überblendung in den Nachhall und der kurzen Ausklingdauer der BRIRs vermutet wird.

Die unterste Stufe des Hörversuchs konnte aus den genannten technischen Gründen nicht tiefer als 8,7 ms gelegt werden. Da in den mittleren und großen Räumen von einer höheren perzeptiven mixing time als in den kleinen ausgegangen werden konnte, wurden die Schrittweiten größer und die erste Stufe etwas höher gewählt als in den

[3] In Raum 1 konnten die untersten beiden Stufen aufgrund von Artefakten nicht im Hörversuch dargeboten werden. Das verringerte die Anzahl der Stufen auf 14.

kleinen Räumen (siehe Tabelle 3.2). Die Überblendung der dynamisch nachgeführten Impulsantworten in eine konstante Impulsantwort konnte in den kleinen Räumen mit Schrittweiten von 5,8 ms (256 Samples) realisiert werden. Die mittleren und großen Räume wurden mit Schrittweiten von 11,6 ms (512 Samples) simuliert. So konnte sichergestellt werden, dass im Hörversuch auch in den mittleren und großen Räumen genügend hohe Schwellwerte dargeboten werden können. Die maximalen Überblendzeitpunkte sind in Tabelle 3.2 zu finden.

| Raum | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | EN324 | EN111 | EN190 | H0104 | HE101 | Teldex | UdK | H0105 | JC |
| $t_{min}$ [ms] | 20,3 | 8,7 | 8,7 | 29,0 | 29,0 | 29,0 | 29,0 | 29,0 | 29,0 |
| $t_{max}$ [ms] | 95,8 | 95,8 | 95,8 | 197,3 | 197,3 | 197,3 | 197,3 | 197,3 | 197,3 |

**Tabelle 3.2:** unterste und oberste realisierbare Stufen der Überblendung im Hörversuch

Zudem wurde in einer weiteren Validierungsmessung messtechnisch überprüft, ob sich das Referenzschallfeld (Faltung mit kompletter BRIR) und das Testschallfeld (Faltung mit nach unterschiedlichen Zeiten in konstanten Nachhall übergeblendeten BRIRs) unabhängig von der Überblendung verhalten. Dazu wurde ein Testdatensatz aus einer BRIR erzeugt und die Übertragungsfunktionen für die Referenzsimulation sowie die Simulation mit Überblendung in eine konstante BRIR mit Hilfe der Messsoftware Monkey Forest® bestimmt. Bei der Benutzung nur einer BRIR, die in beiden Simulationen identisch ist, müssten die Amplitudenfrequenzgänge (und demzufolge auch die Impulsantworten) in beiden Zuständen übereinstimmen. Entgegen der Erwartung waren die Frequenzgänge der vollständig gerenderten BRIR und der nach dem Überblendzeitpunkt in sich selbst übergeblendeten Variante nicht identisch. Vor allem im Bassbereich unterschieden sich die Frequenzgänge. Die Ursache für dieses Verhalten konnte bisher nicht ermittelt werden. Um diese hörbaren Frequenzgangsvariationen im Hörversuch auszuschließen, wurden während des Hörversuchs alle BRIRs großzügig hochpassgefiltert. Die Übertragungsfunktion des Hochpassfilters ist in Abbildung 3.18 dargestellt. Es handelt sich um zwei kaskadierte Butterworthfilter 2. Ordnung mit einer Grenzfrequenz von 183 Hz. Durch das Filter wurden die Anteile unterhalb der Schroederfrequenz (bei dieser Raumauswahl war $f_{Schroeder} < 132$ Hz)[4] nicht dargeboten. Dies hat zusätzlich den Vorteil, dass Unterschiede zwischen den Simulationen, die auf dem

[4]Die Schroederfrequenz war in Raum 3 am höchsten.

modalen Verhalten des Raumes beruhen, demnach nicht auftreten sollten.

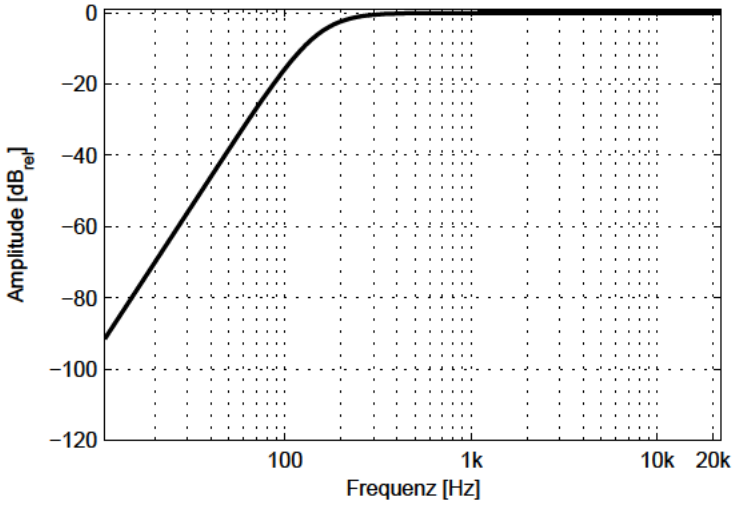



**Abbildung 3.18:** Verwendetes Filter im Hörversuch

In AB-Vergleichen zweier BRIRs aus unterschiedlichen Kopfpositionen wurde festgestellt, dass das Grundrauschen unterschiedlich spektral verfärbt war. Dies würde die Ergebnisse des Hörversuchs beeinflussen, da nicht nur der Zeitpunkt der Überblendung sondern auch die spektrale Zusammensetzung verändert worden wäre. Aus diesem Grund wurden die BRIRs auf 2/3 ihrer Gesamtlänge gekürzt. Die Dauer der BRIRs lag zwischen 14000 und 100000 Samples (0,32 s bzw. 2,26 s). Für die binauralen Simulationen jedes Raumes konnte ein Signal-Rauschabstand von mindestens 60 dB erreicht werden. Der Datensatz der binauralen Raumimpulsantworten wurde normalisiert, um Lautheitsunterschiede zwischen den simulierten Räumen zu minimieren.

In den Vorversuchen stellte sich heraus, dass der Übergangszeitpunkt der Überblendung in Raum 6 (Teldex) bzw. Raum 2 (EN111) für die meisten Versuchspersonen des Vorversuchs leichter zu detektieren war, als in den anderen Räumen (siehe Abbildung 3.19). Bei Raum 6 hatten die Versuchspersonen eher das „Gefühl“, einen Unterschied gehört zu haben. Aus diesem Grund wurde Raum 6 für alle Probanden des Hauptversuchs als Raum für die Trainingssimulation festgelegt. Das Training diente dazu, die Versuchspersonen mit ihrer Aufgabe vertraut zu machen.

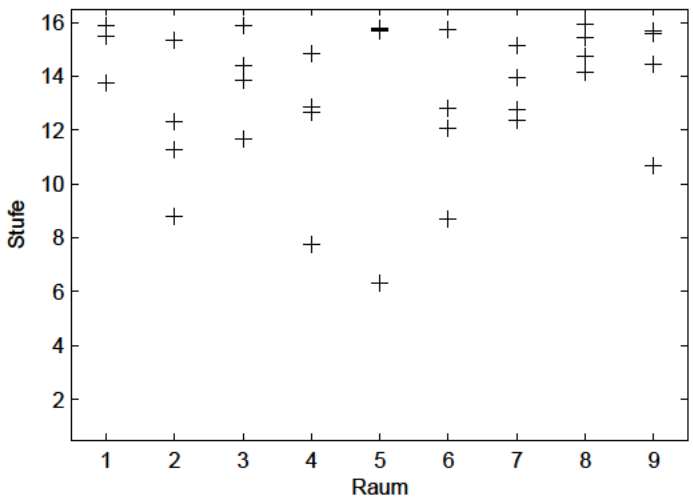



**Abbildung 3.19:** Erreichte Schwellwerte in den Vorversuchen. Eine hohe Stufe kennzeichnet einen geringen Schwellwert.

### 3.4.2 Technischer Aufbau

Der Hörversuch fand im kleinen Studio EN325 des Fachgebiet Audiokommunikation statt. Er wurde mit der am Fachgebiet entwickelten Hörversuchsumgebung WhisPER durchgeführt (Ciba u. a., 2009). Die Software sendet Steuerdaten via OSC-Protokoll an den Faltungsrechner zur Auswahl der Stimuli bzw. des Überblendzeitpunktes in den späten Nachhall. Außerdem werden die Ergebnisse des Hörversuchs gespeichert. WhisPER stellt eine grafische Benutzeroberfläche zur Verfügung, die eine einfache Bedienbarkeit durch die Versuchspersonen gewährleistet.
Die Auswahl der passenden BRIRs zur aktuellen Kopfposition wurde durch den Headtracker gesteuert. Dieser war fest auf dem Kopfhörer montiert und sendete permanent aktuelle Positionsdaten an einen Empfänger. Die schnelle Faltung der jeweils zur Kopfposition passenden BRIR mit dem nachhallfreien Stimuli erfolgte in Echtzeit durch die Software fWonder. Die Überblendung der BRIRs in den diffusen Nachhall erfolgte mit einem linearen Filter, dessen Länge der Schrittweite zwischen den Überblendzeitpunkten entsprach, und wurde ebenfalls mit Hilfe von fWonder durchgeführt.
Im Hörversuch kamen elektrostatische Kopfhörer (STAX SR-2050II) zum Einsatz, die frequenzkompensiert wurden. Dabei wurde ein Filter mit Hochpassregulierung verwendet, das auf Messungen mit dem Kunstkopftorsosimulator FABIAN basiert (Schärer und Lindau, 2009).

Den Aufbau des Hörversuchs verdeutlicht Abbildung 3.20.

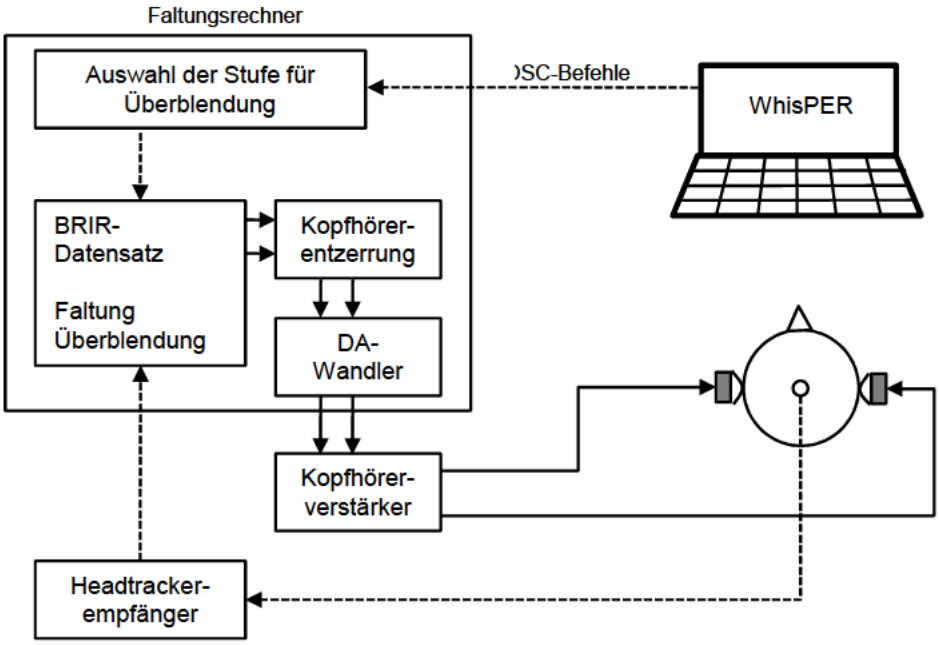



**Abbildung 3.20:** Versuchsaufbau
Durchgezogene Linien kennzeichnen Audiodaten, gestrichelte Linien kennzeichnen Steuerdaten.

### 3.4.3 Versuchsdesign

Zur Evaluation der perzeptiven mixing time wurde die 3-AFC-Methode verwendet. Die Versuchspersonen musste dabei anhand von drei Simulationen entscheiden, welche Simulation sich von den anderen beiden unterscheidet. Eine Simulation war dabei manipuliert, bei den anderen beiden handelte es sich um Referenzsimulationen. Wenn die Versuchspersonen keinen Unterschied zwischen den Alternativen wahrgenommen haben, sind sie gezwungen zu raten.
Der Vorteil dieses Verfahrens ist, dass es objektiv richtige und falsche Antworten gibt. Dadurch wird der Einfluss der Neigung der Versuchspersonen zu einem bestimmten Kriterium bzw. einer bestimmten Antwort (ja/nein) minimiert. Die Ratewahrscheinlichkeit liegt bei 33 %, d. h. die Wahrscheinlichkeit, dass die Versuchsperson den Unterschied

bei einer richtigen Antwort tatsächlich wahrgenommen hat, ist relativ hoch.
Die Benutzeroberfläche des 3-AFC-Testdesigns in WhisPER zeigt Abbildung 3.21.

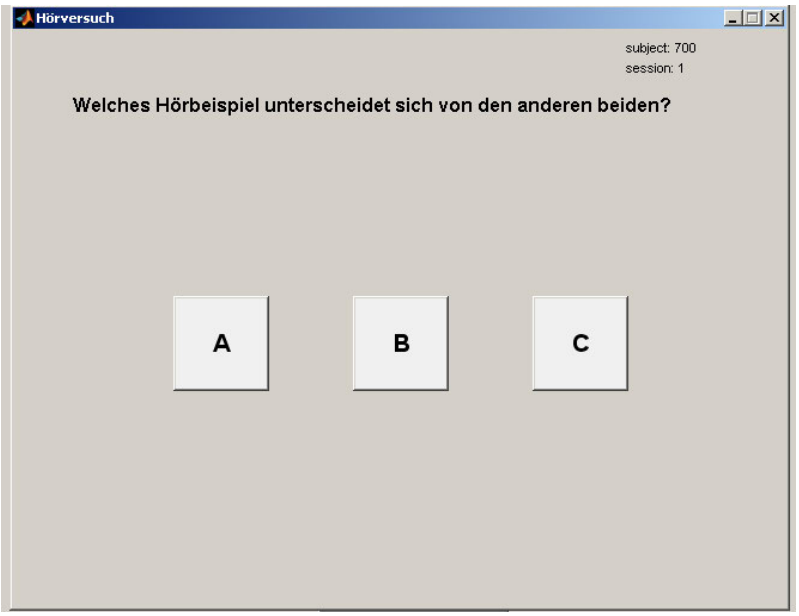



**Abbildung 3.21:** Benutzeroberfläche des 3-AFC-Testdesigns in WhisPER

Die Versuchspersonen mussten eine manipulierte dynamische binaurale Simulation von der Originalsimulation unterscheiden. In der Originalsimulation wurden die kompletten BRIRs entsprechend der aktuellen Kopfposition des Hörers nachgeführt. Bei der manipulierten Simulation entsprach nur der frühe Teil der BRIRs der aktuellen Kopfposition. Der späte Nachhall wurde immer aus der BRIR, die mit Blickrichtung nach vorn gemessen wurde, entnommen (siehe Abbildung 3.22). Dieses Vorgehen entspricht dem praxisnahen Fall, bei dem während der Simulation der konstante späte Anteil der Impulsantworten aus einer spezifischen BRIR des vorhandenen Datensatzes entnommen wird.
Gemäß Meesawat und Hammershøi (2003) gibt es keinen signifikanten Unterschied zwischen verschiedenen Quell-/Empfängerrelationen bezüglich der perzeptiven mixing time. Lindau u. a. (2007) fanden zumindest eine Abhängigkeit der perzeptiven mixing time von der Quellposition (siehe Abschnitt 2.4).
In Vorversuchen, die im Rahmen dieser Arbeit durchgeführt wurden, konnte dies bestätigt werden: Simulationen mit Nachhallschwänzen, die aus Messungen mit einer anderen

Quellposition entnommen wurden[5], lieferten hohe perzeptive mixing times. Besonders im tiefen Frequenzbereich unterschieden sich die späten Anteile der Impulsantworten verschiedener Quellpositionen, was auf das modale Verhalten der Räume zurückzuführen ist. Wie in Abschnitt 2.1 erläutert, ist mixing nicht mit dem modalen Verhalten des Schallfeldes vereinbar.
Die vorliegende perzeptive Evaluation der mixing time untersucht nicht den worst case, sondern ist an der praktischen Anwendung der Ergebnisse in der dynamischen Binauralsynthese orientiert. Aus diesem Grund wird bei der manipulierten Simulation der Nachhall aus der BRIR einer anderen Kopfrichtung entnommen und nicht aus einer anderen Quellposition.

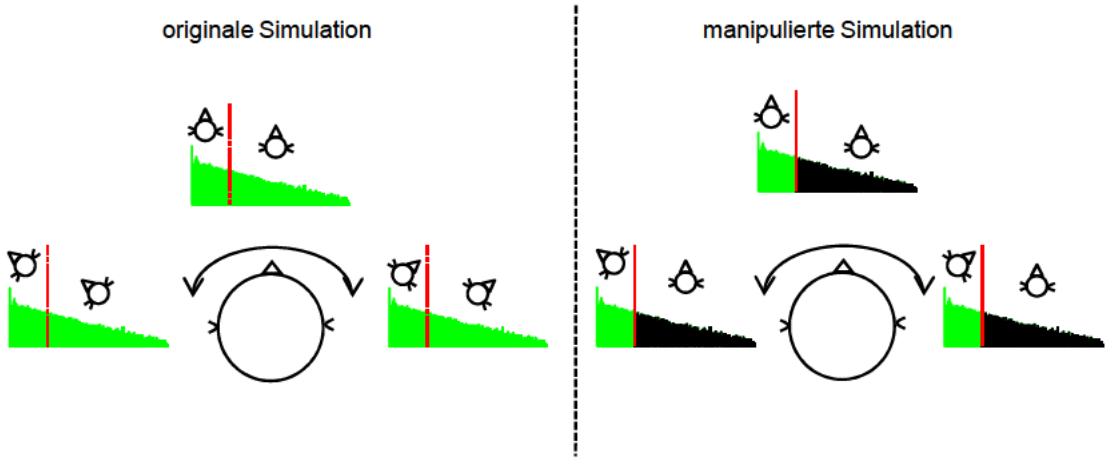



**Abbildung 3.22:** Im adaptiven Hörversuch zu unterscheidende Situationen. Der Überblendzeitpunkt zwischen frühen und späten Anteilen der BRIRs konnte adaptiv verändert werden.

Die Reizstärke bzw. der Überblendzeitpunkt in den späten Nachhall wurde adaptiv an die Erkennungsleistung der Versuchsperson angepasst. Die Adaption wurde durch ein parametrisches Verfahren mit Bayes-Schätzung realisiert. Dabei werden Informationen über die Wahrscheinlichkeitsverteilung der möglichen Schwellwerte zur Platzierung der ersten Stimuli genutzt, was verhindern soll, dass die ersten Trials eines Runs zu lange in Bereichen verweilen, die von der Schwelle weit entfernt sind. In dieser Arbeit wurde als a-priori-Wahrscheinlichkeitsdichtefunktion eine breite Normalverteilung mit

[5] Die Quellposition befand sich hinten rechts, der Abstand zum Kunstkopftorsosimulator entsprach dem vierfachen Hallradius. Es wurden Richtung und Entfernung der Quelle zum Empfänger verändert.

dem Mittelwert in der Mitte des Stimulusranges angenommen. Diese etwas konservative Schätzung stützt sich auf Erfahrungswerte.
Bei Bayes-Methoden wird der nachfolgende Stimulus auf Basis des besten aktuellen Schätzwertes der Schwelle festgelegt. Es wird eine Wahrscheinlichkeitsverteilung für die möglichen Schwellwerte gesucht. In WhisPER wird der Mittelwert der a-posteriori-Verteilung zur Auswahl des nächsten Stimulus und zur abschließenden Schätzung der Schwelle genutzt. WhisPER legt die logistische Funktion als psychometrisches Modell zugrunde. Die vorgenommenen Einstellungen in WhisPER entsprechen nahezu dem ZEST-Verfahren (Treutwein, 1995). Eine Abweichung liegt in der psychometrischen Funktion, da die Autoren des ZEST-Verfahrens die Weibull-Funktion benutzen. Zudem entspricht die a-priori-Wahrscheinlichkeitsdichtefunktion, die in WhisPER genutzt wurde, nicht der Anforderung, dass sie aus einer repräsentativen Stichprobe gewonnen wurde. Der Vorteil des ZEST-Verfahrens ist, dass es frei von Bias[6] ist und schon nach 20 Trials konvergiert (Otto, 2008, S. 63).
Als Stimulus wurde das Schlagzeugsample verwendet, das sich in der Untersuchung von Lindau u. a. (2007) als besonders kritisch erwies. Das nachhallfreie Hörbeispiel dauerte 1,4 s.

Die Probanden wurden, nachdem sie eine Anleitung zum Versuch gelesen haben, in einem Training an die Aufgabe herangeführt. Das Training wurde mit der Simulation von Raum 6 durchgeführt (siehe auch Abschnitt 3.4.1). Im Unterschied zum richtigen Hörversuch wurde den Versuchspersonen mitgeteilt, bei welcher der drei Darbietungen es sich um die manipulierte Simulation handelte. Um größtmögliche Unterschiede zwischen der manipulierten und der Originalsimulation zu erreichen, wurden die Versuchspersonen angewiesen, den Kopf möglichst weit zu drehen. Die endgültige Wahl der idealen Kopfposition zur Unterscheidung der Simulationen wurde den Probanden überlassen. Die Versuchspersonen durften die Lautstärke selbst einstellen.
Die Simulationen der Räume wurden in zufälliger Reihenfolge dargeboten. Jede Versuchsperson hörte die Simulation aller neun Räume. Ein Run entsprach einem Raum. Die Simulationen mussten zwischen den Runs geladen werden, was den Versuchspersonen eine kurze Pause verschaffte. Zusammen mit der definierten Länge von 20 Trials je Run führte das zu einer Testdauer von ca. 60 Minuten pro Person.
Nach dem Hörversuch wurden die Versuchspersonen gebeten, einen kurzen Fragebogen

[6] Bias ist der systematische Fehler. Er ist definiert als Differenz zwischen Schätzwert und wahrem Schwellwert.

zu Alter, Geschlecht und Hörversuchserfahrung auszufüllen und eventuell Unterscheidungsmerkmale zwischen manipulierter und Originalsimulation zu nennen.

# 4 Hörversuchsergebnisse

## 4.1 Stichprobenumfang und -beschreibung

Um den Zusammenhang der perzeptiven mixing time mit den Faktoren Volumen und mittlerem Absorptionsgrad zu finden, wurde eine Varianzanalyse (ANOVA, analysis of variance) mit vollständiger Messwiederholung durchgeführt. Wie in Abschnitt 3.1 beschrieben, sind die beiden Faktoren jeweils 3-fach gestuft. Der Vorteil des Testdesigns mit vollständiger Messwiederholung ist, dass sich der erforderliche Stichprobenumfang erheblich reduzieren lässt, wenn die gleiche Stichprobe mehrmals untersucht wird. Wie im vorherigen Abschnitt erläutert, bedeutet das, dass alle Versuchspersonen unter allen Testbedingungen getestet wurden, d. h. die Simulationen aller neun Räume unterscheiden mussten.
Der optimale Stichprobenumfang einer 2-faktoriellen univariaten[1] ANOVA mit vollständiger Messwiederholung sollte vom erwarteten Interaktionseffekt abhängig gemacht werden (Bortz und Döring, 2006, S. 631). Die Anzahl der Freiheitsgrade $df$ der Interaktion berechnet sich aus der Anzahl der Stufen innerhalb eines Faktors $p$ und $q$ mit:

$$df = (p - 1) \cdot (q - 1) \tag{4.1}$$

Die Anzahl der Stufen $p$ des Faktors Volumen und $q$ des Faktors mittlerer Absorptionsgrad beträgt jeweils 3, die Interaktion 1. Ordnung hat demnach vier Freiheitsgrade. Für einfaktorielle Varianzanalysen mit Messwiederholung und vier Freiheitsgraden werden nach Bortz und Döring (2006, Tabelle 9.9 auf Seite 631) 20 Versuchspersonen benötigt. Bedingung dafür ist, dass die durchschnittliche Korrelation $\bar{\rho}$ zwischen den Messreihen

[1] Univariate Methoden untersuchen nur eine abhängige Variable, in diesem Fall die perzeptive mixing time.

0,5 beträgt und mittlere Effekte ($E$=0,25) auf dem Signifikanzniveau $\alpha$=0,05 abgesichert werden sollen. Die Teststärke, also die Wahrscheinlichkeit, dass eine Alternativhypothese angenommen wird, wenn diese wahr ist, wird allgemein bei 0,8 angenommen. Der optimale Stichprobenumfang für eine Zelle eines mehrfaktoriellen Versuchsplanes berechnet sich wie folgt (Bortz und Döring, 2006, S. 631):

$$n_{Zelle} = \frac{(n-1)(df+1)}{N_{Zellen}} + 1 \tag{4.2}$$

Mit einer Versuchspersonenanzahl von $n$=20 bei einfaktoriellen Versuchsplänen, der Anzahl der Zellen $N_{Zellen}$=3·3=9 und den oben genannten Bedingungen ergibt sich der optimale Stichprobenumfang von $n_{opt}$=12 bei mehrfaktoriellen Versuchsplänen. Aufgrund der sehr unterschiedlichen Ergebnisse aus den Vorversuchen kann vermutet werden, dass die Korrelation zwischen den Messreihen nicht 0,5 ist, wie bisher angenommen, sondern geringer. Deshalb wurde die empirisch ermittelte durchschnittliche Korrelation aus den Vorversuchen in die Bestimmung des Stichprobenumfangs einbezogen. Dazu wird die korrigierte Effektgröße mit

$$E_{korrigiert} = \frac{E}{\sqrt{1-\bar{\rho}}} \tag{4.3}$$

bestimmt (Bortz und Döring, 2006, S. 632). Für einen mittleren Effekt $E$=0,25 und die durchschnittliche Korrelation zwischen den Messreihen $\bar{\rho}$=0,15 (ermittelt aus den Vorversuchen) ergibt sich eine korrigierte Effektgröße von 0,27. Mit Hilfe des frei erhältlichen Programms G*POWER[2] kann nun der optimale Stichprobenumfang mit beliebiger Effektgröße berechnet werden (Bortz und Döring, 2006, S. 627). Für eine ANOVA ohne Messwiederholung ergibt sich eine optimale Stichprobengröße von $n_{opt}$=168. Die Messwiederholung erlaubt es, diese Anzahl durch die Anzahl der Zellen zu teilen. Folglich ist die erforderliche Versuchspersonenanzahl

$$n_{opt} = 19$$

für eine ANOVA mit Messwiederholung mit einer Teststärke von 0,8 und der durchschnittlichen Korrelation von $\bar{\rho}$=0,15. Ein Effekt von $E$=0,27 kann auf dem Signifi-

[2] `http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/`

kanzniveau $\alpha$=0,05 mit diesem Stichprobenumfang abgesichert werden.

Die dynamische Binauralsynthese soll Schallfelder in Zukunft nicht nur plausibel simulieren können, sondern von einem realen Schallfeld im direkten Vergleich nicht unterscheidbar sein. Um diesem ambitionierten Anspruch gerecht zu werden, sollten Hörversuche zur dynamischen Binauralsynthese mit geschulten Hörern durchgeführt werden. Unter diesem Aspekt wurden die Versuchspersonen ausgewählt. Die meisten Probanden waren Studenten und Mitarbeiter des Fachgebiets Audiokommunikation, da sie als geübte Hörer eingestuft werden können. Insgesamt nahmen 24 Versuchspersonen an der perzeptiven Evaluation der Wahrnehmbarkeit der Diffusfeldgrenze teil. Die Probanden waren zwischen 20 und 38 Jahren alt, der Altersdurchschnitt lag bei 28,3 Jahren. Bis auf drei Versuchspersonen gaben alle an, über eine musikalische Ausbildung zu verfügen. Alle Versuchspersonen haben schon vorher an Hörversuchen teilgenommen. Es kann demnach von Probanden mit einer angemessenen Hörerfahrung ausgegangen werden.

## 4.2 Deskriptive Auswertung

In der Darstellung der Ergebnisse aller 24 Versuchspersonen ist erkennbar, dass sich die perzeptiven mixing times über einen sehr großen Bereich erstrecken (siehe Abbildung 4.1). In den großen Räumen 7 bis 9 unterscheiden sich die mixing times der einzelnen Versuchpersonen innerhalb eines Raumes sogar um bis zu 168 ms (Raum 7). Viele Probanden erreichten mixing time-Werte, die sich nahe der untersten Stufe befinden.

Auch die genauere Analyse der Hörversuchsverläufe zeigte, dass die Aufgabe, die originale Simulation von der manipulierten Simulation zu unterscheiden, für viele der Versuchspersonen sehr schwierig war. Einige Probanden sind dauerhaft an der untersten realisierbaren Stufe angeschlagen. In Abbildung 4.2(a) und (b) sind zwei kritische Hörversuchsverläufe dargestellt. Versuchsdurchläufe, in denen bei der untersten Stufe öfter falsch als richtig geantwortet wurde, mussten aus der weiteren Untersuchung ausgeschlossen werden (z. B. der Verlauf in Abbildung 4.2(a)). Bei diesen Versuchspersonen muss davon ausgegangen werden, dass die wahre Schwelle unter der untersten dargebotenen Stufe liegt.

Ein Ziel des Hörversuchs war es, die Haupteffekte und Interaktion 1. Ordnung mittels Varianzanalyse mit vollständiger Messwiederholung zu untersuchen. Um Probleme mit

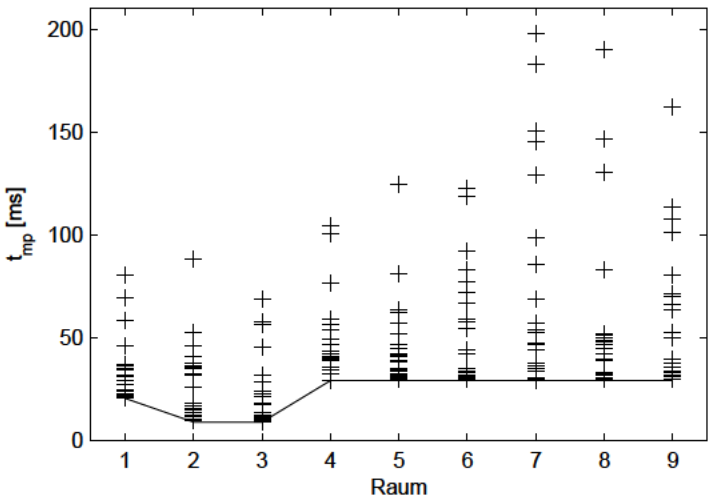
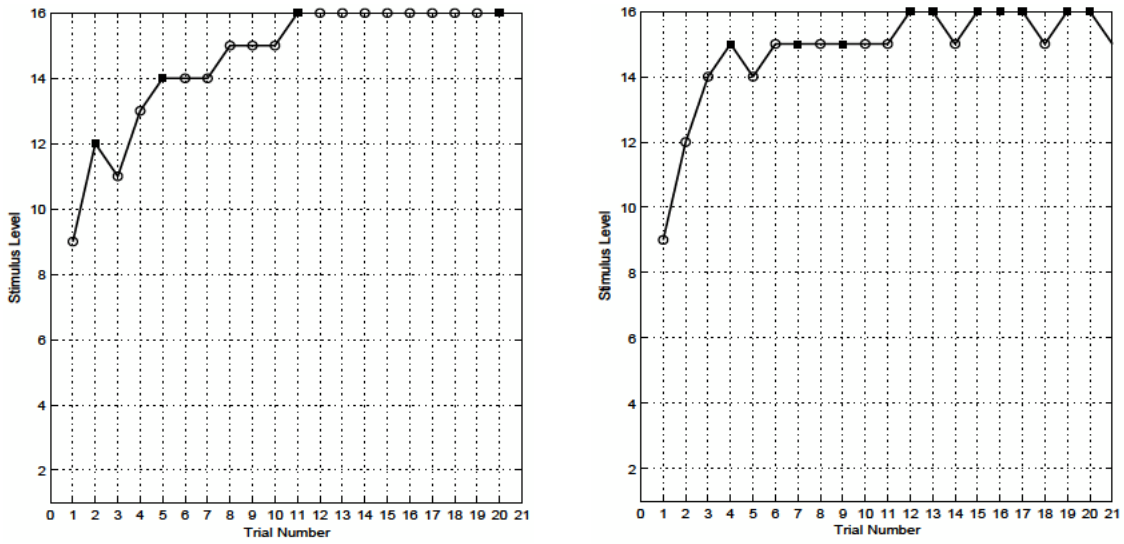



**Abbildung 4.1:** Hörversuchsergebnisse aller 24 Versuchspersonen. Die durchgezogene Linie zeigt die unterste Stufe des Überblendzeitpunktes zum Nachhall der Impulsantwort mit Blickrichtung nach vorn, die im Hörversuch realisierbar war.

fehlenden Daten zu umgehen, müssen die Datensätze vollständig sein (Bortz und Döring, 2006, S. 549), folglich mussten alle Versuchspersonen ausgeschlossen werden, die bei mindestens einem Raum an der untersten Schwelle mehr als die Hälfte der Simulationen nicht unterscheiden konnten.
Von den 24 Versuchspersonen konnten 10 Expertenhörer herausgefiltert werden, die unter jeder Testbedingung ein valides Ergebnis geliefert haben. Diese 10 Versuchspersonen wurden in die weitere Untersuchung mit einbezogen.

Die reduzierte Stichprobengröße ließ es nur zu, einen eher großen Effekt ($E$=0,34) zu testen.[3] Die tatsächlich getestete Effektgröße wurde wiederum mit Hilfe von G*POWER bestimmt. Dazu wurde erst die durchschnittliche Korrelation zwischen den Messreihen des Hauptversuchs bestimmt ($\bar{\rho}$=0,17) und anschließend die von G*POWER berechnete Effektgröße ($E$=0,37) mit dem Korrekturfaktor für Effektgrößen bei Varianzanalysen mit Messwiederholung $\sqrt{1-\bar{\rho}}$ multipliziert.
Zur Feststellung der internen Konsistenz der Messdaten wurde Cronbach's $\alpha$ mit SPSS® berechnet. Mit Hilfe von Cronbach's $\alpha$ als ein Reliabilitätsmaß kann eine Aussage darüber getroffen werden, ob das Ergebnis des Hörversuchs generalisierbar ist (Bortz, 2005, S. 559). Die Berechnung erfolgt mit dieser Formel:

[3] Bei Bortz und Döring (2006, S. 606) wird ein Effekt von $E$ 0,25 als mittlerer, ein Effekt von $E$ 0,40 als großer Effekt klassifiziert.




**(a)** ungültig **(b)** gültig

**Abbildung 4.2:** Beispielhaft zwei kritische Verläufe der Schwellenmessung
■ richtige Antwort
○ falsche Antwort

$$\alpha = \frac{p}{p-1} \cdot \left(1 - \frac{\sum_i s_i^2}{s_{tot}^2}\right) \tag{4.4}$$

wobei $p$ hierbei die Anzahl der Variablen, also der Versuchspersonen, $s_i^2$ die Varianz der Versuchspersonen und $s_{tot}^2$ die Varianz der Summe ist. Cronbach's $\alpha$ lag bei 0,635. Dieser Wert ist als eher gering einzustufen. Der Grund dafür ist, dass eine einzelne Versuchsperson sehr viel höhere Schwellwerte erreichte als die anderen. Da diese Versuchsperson aber sehr empfindlich war und der Hörversuch mit dem Ziel durchgeführt wurde, dass auch Expertenhörer keine Unterschiede zwischen den Simulationen wahrnehmen können, wurde sie trotzdem in die weitere Betrachtung eingeschlossen.

Anhand der Boxplots der zehn besten Versuchspersonen (siehe Abbildung 4.3) ist erkennbar, dass die mittleren 50 % der $t_{mp}$-Werte einer breiten Streuung unterliegen (besonders in Raum 7) und dass die Schwellwerte einiger Versuchspersonen sehr von den mittleren 50 % der gemessenen Schwellwerte abweichen. Auch kann anhand der Lage des Medians innerhalb des Interquartilbereiches sowie der Lage der maximalen bzw. minimalen $t_{mp}$ vermutet werden, dass die gemessenen Schwellwerte je Raum nicht immer normalverteilt sind.

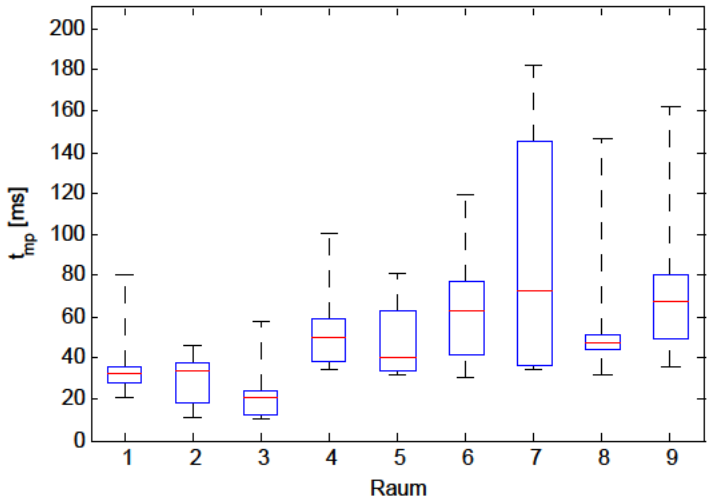



**Abbildung 4.3:** Schwellwerte der 10 besten Versuchspersonen als Boxplots mit Median, Interquartilbereich und maximal bzw. minimal erreichten Schwellwerten

Voraussetzung für die Rechnung mit Mittelwerten und Konfidenzintervallen ist die Normalverteilung der Hörversuchsergebnisse. Um diese zu untersuchen, wurde ein Lilliefors-Test durchgeführt. Das Resultat war, dass die Daten der Räume 1, 3, 7 und 8 nicht normalverteilt sind. Da jedoch die Mehrheit der simulierten Räume normalverteilte $t_{mp}$-Werte hatte, wurden trotzdem die Mittelwerte und Konfidenzintervalle bestimmt. Die Konfidenzintervalle kennzeichnen denjenigen Wertebereich, in dem sich „95 % aller Populationsparameter befinden, die den empirisch ermittelten Stichprobenkennwert erzeugt haben können“ (Bortz, 2005, S. 102).
In Abbildung 4.4 sind die Mittelwerte und 95 %-Konfidenzintervalle, geordnet nach den beiden Testbedingungen Volumen und mittlerem Absorptionsgrad, dargestellt. Die unterschiedlichen Graustufen kennzeichnen die Volumenabstufungen, die abfallenden Pfeile die Abstufungen des mittleren Absorptionsgrades. Die Werte der perzeptiven mixing time steigen mit wachsendem Volumen an. An den großen Konfidenzintervallen bei den Räumen 7 bis 9 ist zu erkennen, dass sich auch die Unsicherheit der Versuchspersonen mit dem Volumen erhöht.

Die Mittelwerte der perzeptiven Evaluation der mixing time je dargebotenem Raum befinden sich in Tabelle 4.1.

Nach dem Hörversuch wurden die Versuchspersonen gebeten, die Merkmale, die zur Unterscheidbarkeit der Simulationen beigetragen haben, anzugeben. Die genannten

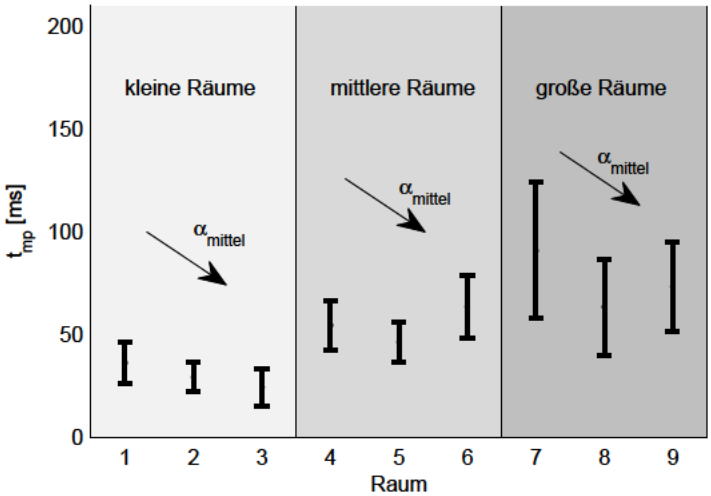



**Abbildung 4.4:** Mittelwerte der perzeptiven mixing time und die dazugehörigen 95 %-Konfidenzintervalle in Abhängigkeit der Volumen- und Absorptionsgradstufen

| Raum | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| | EN324 | EN111 | EN190 | H0104 | HE101 | Teldex | UdK | H0105 | JC |
| $t_{mp}$ [ms] | 36,11 | 29,19 | 24,19 | 54,47 | 46,24 | 63,54 | 90,77 | 63,31 | 73,30 |

**Tabelle 4.1:** Mittelwerte der perzeptiven mixing times je Raum

Unterscheidungsmerkmale lassen sich grob zu

- Richtung des Nachhalls
- Klangfarbe
- Artefakte

zusammenfassen. Es wurde angemerkt, dass die Cues bei jeder Simulation bzw. sogar Schwellwertstufe wechselten. Die meisten Versuchspersonen drehten ihren Kopf zur Seite. Bei der manipulierten Simulation kamen Seitenwandreflexionen dementsprechend auf dem Ohr an, das der Quelle abgewandt war. Einige Räume des Untersuchungsschemas haben sehr ausgeprägte Seitenreflexionen, weshalb die Unterscheidung zwischen manipulierter und originaler Simulation einfacher möglich war.

## 4.3 Inferenzstatistische Auswertung

Mit Hilfe der zweifaktoriellen Varianzanalyse wird untersucht, ob eine abhängige Variable (die perzeptive mixing time) von zwei unabhängigen Variablen (Volumen und mittlerer Absorptionsgrad) beeinflusst wird. Die beiden unabhängigen Variablen (Faktoren) sind im Untersuchungsschema dieser Arbeit jeweils 3-fach gestuft. Die Voraussetzung für eine Varianzanalyse mit Messwiederholung ist, dass die Varianzen[4] der Differenzen der einzelnen Messungen von jeweils zwei Faktorstufen homogen sein müssen (Bortz, 2005, S. 354). Diese Zirkularitätsannahme wird in SPSS® mit dem Test von Mauchly überprüft. Für den Fall, dass der Test nicht signifikant ist, ist die Voraussetzung für eine ANOVA mit Messwiederholung gegeben. In dieser Untersuchung trifft dies zu.

Die Untersuchung der Haupteffekte bezüglich der perzeptiven mixing time ergab, dass nur der Effekt des Volumens hoch signifikant ist ($p<0{,}01$). Die perzeptive mixing time ist demnach vom Raumvolumen abhängig. Den positiven Trend kann man auch in Abbildung 4.4 erkennen. Eine Aussage über die Signifikanz des Trends oder einen linearen bzw. quadratischen Zusammenhang zur perzeptiven mixing time lässt sich nicht treffen, da die Faktoren Volumen und mittlerer Absorptionsgrad nur ordinalskaliert sind. Voraussetzung für Trendtests ist jedoch die Intervallskalierung der Faktorstufen.

Beachtet werden muss, dass ein ansteigendes Volumen mit einer ansteigenden Nachhallzeit konfundiert ist. Trotzdem konnte ein Effekt des mittleren Absorptionsgrades nicht gefunden werden ($p=0{,}178$). Die relative Halligkeit unabhängig vom Volumen hat folglich keinen signifikanten Einfluss auf die perzeptive mixing time.

Mit Hilfe einer ANOVA kann zusätzlich zu den Haupteffekten auch die Kombinationswirkung der beiden Faktoren Volumen und mittlerer Absorptionsgrad untersucht werden. Eine Interaktion liegt vor, wenn die Haupteffekte nicht nur additiv, sondern in anderer Weise zusammenwirken. Die Interaktion von Volumen und mittlerem Absorptionsgrad ist nicht interpretierbar ($p=0{,}229$).
Den Interaktionsdiagrammen (Abbildung 4.5) kann jedoch ebenfalls ein Trend ansteigender perzeptiver mixing time bei ansteigendem Volumen entnommen werden. Ein Trend des mittleren Absorptionsgrades bezüglich der perzeptiven mixing time ist nicht einfach abzuleiten, da sich die drei Verläufe im Interaktionsdiagramm nicht einheitlich

[4] Die Varianz kennzeichnet die Unterschiedlichkeit der einzelnen Werte einer Verteilung (Bortz, 2005, S. 798).

verhalten.

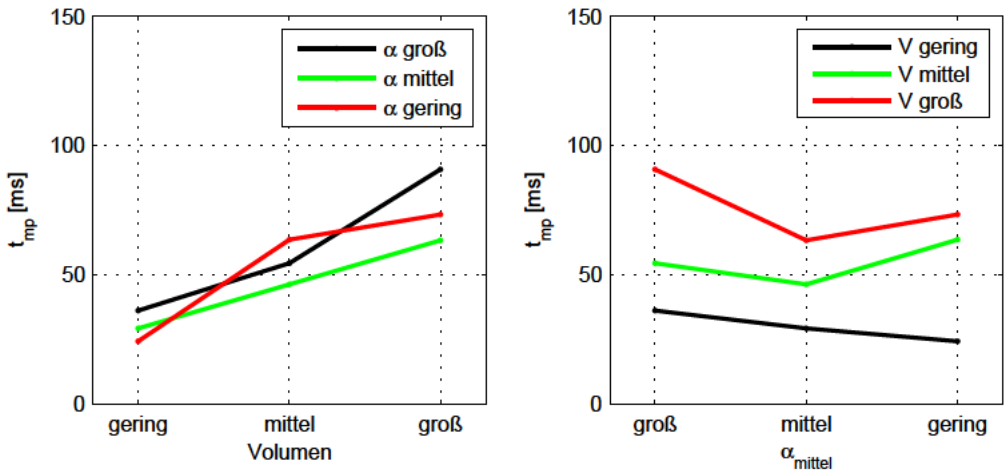



**Abbildung 4.5:** Interaktionsdiagramme

## 4.4 Fehlerquellen

Aufgrund des Messwiederholungsdesigns ist die Dauer des Hörversuches kritisch zu betrachten. Vor allem die Effekte Ermüdung und Motivationsverlust sind in diesem Zusammenhang zu nennen. Um den Einfluss dieser Effekte auf das Ergebnis der perzeptiven Evaluation gering zu halten, wurden die Simulationen der Räume in zufälliger Reihenfolge dargeboten. Jede Versuchsperson hat eine andere Reihenfolge der Simulationen gehört. So wurde vermieden, dass immer die selbe Raumsimulation zum Ende bzw. Beginn des Hörversuchs dargeboten wurde.

Das Verständnis der Aufgabe im Hörversuch sollte durch das Training gewährleistet werden. Das Training endete wie auch ein Run im Hörversuch nach 20 Trials. Ein gezieltes Training mit verschiedenen Raumsimulationen zur Demonstration der möglichen Unterscheidungsmerkmale hätte möglicherweise dazu geführt, dass mehr als zehn Versuchspersonen die Einstufung als Expertenhörer erreicht hätten. Dies war jedoch aus technischen Gründen nicht möglich, da das Laden einer Raumsimulation bis zu zwei Minuten in Anspruch nahm und so die Dauer des Hörversuchs erheblich erhöht hätte.

Die realisierbaren Schwellwertstufen waren in einigen Räumen relativ hoch. Durch nied-

rigere unterste Stufen hätten eventuell die Ergebnisse von mehr Versuchspersonen in die Auswertung einfließen können, da es einfacher möglich gewesen wäre, valide Schwellwerte[5] zu erreichen.
Der Vorteil der etwas hoch liegenden Stufen ist, dass die Versuchspersonen, die trotzdem einen validen Schwellwert lieferten, als gute Expertenhörer eingestuft werden können. Die Ergebnisse des Hörversuchs beruhen demnach auf *wahren* Experten.

[5]Wie in Abschnitt 4.2 erläutert, konnten nur die Versuchspersonen in die Auswertung einbezogen werden, die unter keiner Schwellwertbedingung dauerhaft an der untersten Stufe angeschlagen sind und dort weniger als die Hälfte richtig erkannt haben.

# 5 Regressionsanalyse

Die Regressionsanalyse dient dazu, eine schwer zu erfassende Variable durch eine einfach messbare Variable vorherzusagen. Im Rahmen dieser Arbeit soll die Kriteriumsvariable *perzeptive mixing time* $t_{mp}$ vorhergesagt werden. Als Prädiktorvariablen dienen hier die modellbasierten und empirischen Prädiktoren von Diffusität, die in den Abschnitten 2.2 und 2.3 vorgestellt wurden. Mit Hilfe dieser schneller und einfacher zu bestimmenden Prädiktoren kann eine Vorhersage getroffen werden, ab wann der Nachhall einer Impulsantwort durch den Nachhall einer anderen Impulsantwort ausgetauscht werden kann. In diesem Kapitel wird die Eignung der einzelnen Prädiktoren zur Diffusitätswahrnehmung erläutert.

Zunächst wurden nur lineare Zusammenhänge zwischen den Prädiktoren und der Kriteriumsvariablen getestet, weil andere Zusammenhänge in der Natur seltener vorkommen. Während der Erstellung der Regressionsmodelle konnte anhand der Lage der empirisch ermittelten perzeptiven mixing time (siehe Abschnitt 4) bezüglich der physikalischen Prädiktoren festgestellt werden, dass eine Regression mit nichtlinearen Zusammenhängen nicht sinnvoll erscheint.

Mit einer linearen Regression wird die kleinste quadratische Abweichung der $t_{mp}$-Werte aus dem Hörversuch zu den $t_m$-Werten der in den Abschnitten 2.2 und 2.3 diskutierten Prädiktoren gesucht. Es ergibt sich eine Regressionsgerade der Form

$$y = b \cdot x + a \tag{5.1}$$

mit dem Anstieg $b$ und dem Schnittpunkt $a$ mit der y-Achse. Das Kriterium der kleinsten Quadrate lässt sich zu

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = min \tag{5.2}$$

zusammenfassen. Bei $\hat{y}_i$ handelt es sich um den durch die Regressionsgerade geschätzten y-Wert. Setzt man Gleichung (5.1) in (5.2) ein, so erhält man

$$\sum (y_i - \hat{y}_i)^2 = \sum [y_i - (b \cdot x_i + a)]^2 \tag{5.3}$$

Der Anstieg $b$ und Schnittpunkt mit der y-Achse $a$ lassen sich berechnen, indem die Gleichung (5.3) partiell nach $a$ und $b$ differenziert wird, jeweils die 1. Ableitung 0 gesetzt und nach $a$ und $b$ aufgelöst wird:

$$a = \bar{y} - b \cdot \bar{x} \tag{5.4}$$

$$b = \frac{n \cdot \sum\limits_{i=1}^{n} x_i \cdot y_i - \sum\limits_{i=1}^{n} x_i \cdot \sum\limits_{i=1}^{n} y_i}{n \cdot \sum\limits_{i=1}^{n} x_i^2 - (\sum\limits_{i=1}^{n} x_i)^2} \tag{5.5}$$

In Gleichung (5.4) kennzeichnet $\bar{x}$ den Mittelwert der Prädiktoren und $\bar{y}$ den Mittelwert der durch die Prädiktoren hervorgerufenen physikalischen mixing time $t_m$.
Eine ausführliche Herleitung der Regressionskoeffizienten findet sich bei Bortz (2005, S. 185-186).

Für den Fall, dass der Anstieg $b$ der Regressionsgeraden den Wert 1 und der konstante Term $a$ den Wert 0 haben, wären physikalische und perzeptive mixing time identisch. Der konstante Term $a$ müsste den Wert 0 annehmen, wenn eine physikalische mixing time von 0 ms eine perzeptive mixing time von 0 ms vorhersagen soll.

Die Berechnung der Regressionskoeffizienten wurde in Matlab® mit der Funktion *fit()* durchgeführt.
Zudem wurden die Konfidenzintervalle der linearen Modelle bestimmt, um eine Aussage über die Genauigkeit der Vorhersage treffen zu können. Innerhalb des Konfidenzintervalls einer Regressionsvorhersage befindet sich der wahre Erwartungswert des vorher-

gesagten y-Wertes (Bortz, 2005, S. 194). Breite Konfidenzintervalle zeigen an, dass ein Regressionsmodell weniger zur Prädiktion der $t_{mp}$ geeignet ist.
Mit Hilfe der Konfidenzintervalle kann entschieden werden, ob der Regressionskoeffizient signifikant ist: Liegt der Wert 0 außerhalb des Konfidenzintervalls, ist der Regressionskoeffizient signifikant. Zur Bestimmung der Signifikanz der Regression wurde die Nullhypothese getestet, welche besagt, dass kein linearer Zusammenhang zwischen abhängiger und unabhängiger Variable existiert.
Ein weiterer Kennwert zur Charakterisierung des Regressionsmodells ist die Varianzaufklärung. Sie gibt an, welcher Anteil der Varianz der abhängigen Variablen (in diesem Fall $t_{mp}$) auf die unabhängige Variable (die Prädiktoren) zurückzuführen ist (Bortz, 2005, S. 280).
Die Regressionsanalyse wurde auf Grundlage von neun Räumen durchgeführt. Die Anzahl ist für eine Regressionsanalyse etwas gering, was auch an den großen Konfidenzintervallen der Modelle erkennbar ist. Aus praktischen Gründen konnte die Anzahl der Räume jedoch nicht erhöht werden, da der Messaufwand und die Dauer des Hörversuchs sonst zu hoch gewesen wären.
Die Vorhersage einer Kriteriumsvariablen anhand einer Regressionsgleichung wird umso unsicherer, je mehr der zur Vorhersage verwendete x-Wert vom Mittelwert der in der Stichprobe der Prädiktoren enthaltenen x-Werte abweicht.
Dieses Problem und die geringe Anzahl von Räumen wurde durch die systematische und breite Raumauswahl versucht auszugleichen. Die Raumauswahl bildet eine gute Teilmenge der Realität.

Die Regressionsanalyse wurde für die modellbasierten und empirischen Prädiktoren vorgenommen. So kann die Vorhersage der perzeptiven mixing time entweder anhand von physikalischen Eigenschaften des Raumes oder gemessenen bzw. berechneten Impulsantworten getroffen werden.
Die Regressionen wurden für die mittlere perzeptive mixing time aus dem Hörversuch und für die 95 %-Perzentile der Werte der perzeptiven mixing time berechnet. Die 95 %-Perzentile kennzeichnen die unteren 95 % einer Verteilung (Bortz, 2005, S. 794). Mit der Regression über die 95 %-Perzentile der empirisch ermittelten Schwellwerte erhält man einen konservativeren Wert für $t_{mp}$. Die daraus resultierenden Regressionsformeln garantieren eine nahezu unhörbar exakte Simulation. Das bedeutet, dass 95 % der Expertenhörer, die die Schwellwerte dieser Untersuchung erzeugt haben, zwischen

Original- und manipulierter Simulation keinen Unterschied hören würden.
Zur eindeutigeren Kennzeichnung der vorhergesagten perzeptiven mixing times werden im Folgenden die vorhergesagten perzeptiven mixing times anhand der Mittelwerte des Hörversuchs mit $t_{mp50\%}$ gekennzeichnet. Diejenigen perzeptiven mixing times, die anhand der 95 %-Perzentile vorhergesagt wurden, werden mit $t_{mp95\%}$ bezeichnet.

## 5.1 Regression modellbasierter Prädiktoren

Wie in Abschnitt 2.2 erläutert, können die meisten modellbasierten Prädiktoren der mixing time aus der Literatur a) auf die Quadratwurzel des Volumens $\sqrt{V}$, b) auf die mittlere freie Weglänge, die sich proportional zum Verhältnis von Volumen und Raumoberfläche $\frac{V}{S}$ verhält und c) auf die Nachhallzeit $T$ zurückgeführt werden.
Zusätzlich zu diesen Prädiktoren wurden das Volumen $V$, die Raumoberfläche $S$ und der mittlere Absorptionsgrad $\alpha_{mittel}$ getestet. Der mittlere Absorptionsgrad wurde wie in Abschnitt 3.1 über die Sabine'sche Nachhallzeitgleichung berechnet, die Raumoberfläche wurde aus den drei Hauptdimensionen eines Raumes bestimmt.

Zusätzlich zu den einfachen linearen Regressionen wurden auch multiple Regressionen gerechnet. Je nach Verfahren zum Einschluss von Prädiktoren (alle gleichzeitig, schrittweise und rückwärts) lieferte SPSS® Regressionsmodelle mit einem oder drei Prädiktoren. Die Güte der Regressionsgleichung wurde durch die korrigierte Varianzaufklärung beurteilt (Janssen und Laatz, 2007, S. 427-428). Die multiplen Regressionen mussten jedoch alle verworfen werden, weil die zusätzlichen Koeffizienten nicht signifikant waren, Kollinearitätsprobleme (hohe Interkorrelation) auftraten und die Konfidenzintervalle den Wert 0 einschlossen (Janssen und Laatz, 2007, S. 424-436, 445-448).

### 5.1.1 Mittelwerte

Die Regressionsanalyse der Mittelwerte der perzeptiven mixing times, die im Hörversuch bestimmt wurden, lieferte das Verhältnis von Volumen und Raumoberfläche $\frac{V}{S}$ als besten Prädiktor zur Vorhersage der perzeptiven mixing time. Bei dem Verhältnis $\frac{V}{S}$ handelt es sich um den Kern der Formel zur Bestimmung der mittleren freien Weglänge $l_m$. Die Varianzaufklärung dieses Prädiktors erreichte $R^2$=81,5 %, was einer

Korrelation von $r$=0,9 entspricht.

Die Quadratwurzel des Raumvolumens $\sqrt{V}$ erwies sich als zweitbester Prädiktor der perzeptiven mixing time. Die Varianzaufklärung lag bei $R^2$=78,6 %.
Darauf folgte das Volumen $V$, welches eine Varianzaufklärung von $R^2$=77,4 % erreichte. Eine etwas geringere Varianzaufklärung konnte das Regressionsmodell der Raumoberfläche $S$ erreichen. Diese lag bei 73,3 %.
Die Nachhallzeit erwies sich als eher ungeeignet zur Prädiktion der perzeptiven mixing time $t_{mp50\%}$, die Varianzaufklärung war nur $R^2$=53,4 %. Dieser Wert kann nahezu vollständig dem Fakt zugeschrieben werden, dass die Nachhallzeit mit dem Volumen konfundiert ist, welches als Prädiktor eher geeignet ist. Das bestätigt auch das Regressionsmodell des mittleren Absorptionsgrades $\alpha_{mittel}$, dessen Varianzaufklärung nur 4,3 % erlangt. Ein linearer Zusammenhang zwischen $\alpha_{mittel}$ und $t_{mp50\%}$ besteht demnach nicht.
Auch in der Untersuchung von Hidaka u. a. (2007) wurde festgestellt, dass die physikalische mixing time eher von der Raumform abhängig ist als von der Nachhallzeit.

Die Anstiege der Regressionsmodelle waren alle positiv und signifikant. Eine Ausnahme bildet nur der mittlere Absorptionsgrad. Die Regressionsmodelle basierend auf den Mittelwerten der perzeptiven mixing time und den modellbasierten Prädiktoren sind in Abbildung 5.1 dargestellt. Außerdem zeigt die Abbildung die Mittelwerte und 95 %-Konfidenzintervalle (KIs) der perzeptiven mixing time sowie die 95 %-Konfidenzintervalle der Regressionen.

Die Regressionsgleichung für den besten Prädiktor der perzeptiven mixing time basierend auf Mittelwerten $t_{mp50\%}$ lautet:

$$t_{mp50\%} = 20 \cdot \frac{V}{S} + 12 \qquad [ms] \tag{5.6}$$

Dabei muss beachtet werden, dass die Raumoberfläche $S$ aus den drei Hauptabmessungen Länge, Breite und Höhe eines idealen Rechteckraumes bestimmt wurde. Zusätzliche Oberflächen wie Mobiliar, Emporen usw. wurden nicht berücksichtigt.
Die Regressionsgleichung des zweitbesten Prädiktors ist:

$$t_{mp50\%} = 0,58 \cdot \sqrt{V} + 21 \qquad [ms] \tag{5.7}$$

Basierend auf den Regressionsgleichungen der mittleren freien Weglänge und des Kerns der Reflexionsdichte $\sqrt{V}$ können aus den modellbasierten Prädiktoren die gerade hörbaren Größen aus den empirisch ermittelten perzeptiven mixing times abgeleitet werden. Durch Einsetzen der Gleichung (5.6) in Gleichung (5.8) zur Bestimmung der Anzahl der aufgetretenen Reflexionen bis zur Zeit $t$ (Kuttruff, 2000, S. 122) mit:

$$n_{Reflexionen} = \frac{c \cdot t}{l_m} \tag{5.8}$$

erhält man

$$n_{Reflexionen} = \frac{c \cdot S}{4V} \cdot \frac{t_{mp50\%}}{10^3} \tag{5.9}$$

Diese Gleichung gibt die Anzahl der Reflexionen an, die aufgetreten sein müssen, damit ein Schallfeld als diffus wahrgenommen wird. In den Gleichungen (5.8) und (5.9) bezeichnet $c$ die Schallgeschwindigkeit in m/s, $t$ die Zeit in s, $S$ die Raumoberfläche in $m^2$, $V$ das Raumvolumen in $m^3$ und $t_{mp50\%}$ die perzeptive mixing time in ms.
Für die Auswahl der Räume dieser Arbeit tritt die perzeptive Diffusität des Schallfeldes nach 2-3 Reflexionen ein.[1]
Die erforderliche Reflexionsdichte für ein als diffus wahrgenommenes Schallfeld kann berechnet werden, indem die Ergebnisse der perzeptiven mixing time aus der Regressionsgleichung (5.7) in Gleichung (2.3) eingesetzt werden. Bei der vorliegenden Raumauswahl lagen die Reflexionsdichten zwischen 322 $s^{-1}$ (Raum 8) und 2254 $s^{-1}$ (Raum 3). Um eine volumenunabhängige untere Grenze für die Reflexionsdichte zu erhalten, muss der konstante Term in Gleichung (5.7) vernachlässigt werden. In Gleichung (2.3) eingesetzt, erhält man eine erforderliche Reflexionsdichte von 171 $s^{-1}$. Dieser Wert gilt nur für sehr große Räume, weil in diesem Fall der konstante Term in der Regressionsgleichung keinen großen Einfluss auf die vorhergesagte perzeptive mixing time hat.
Es muss jedoch beachtet werden, dass es sich hierbei nicht um gemessene Größen handelt, sondern um Werte, die sich aus den Regressionsgleichungen und den modellbasierten Prädiktoren ergeben.

[1] Tendenziell reichen in den großen Räumen eher 2 Reflexionen aus, in den kleinen Räumen sind 3 Reflexionen erforderlich, um die perzeptive Diffusität eines Schallfeldes zu erreichen.

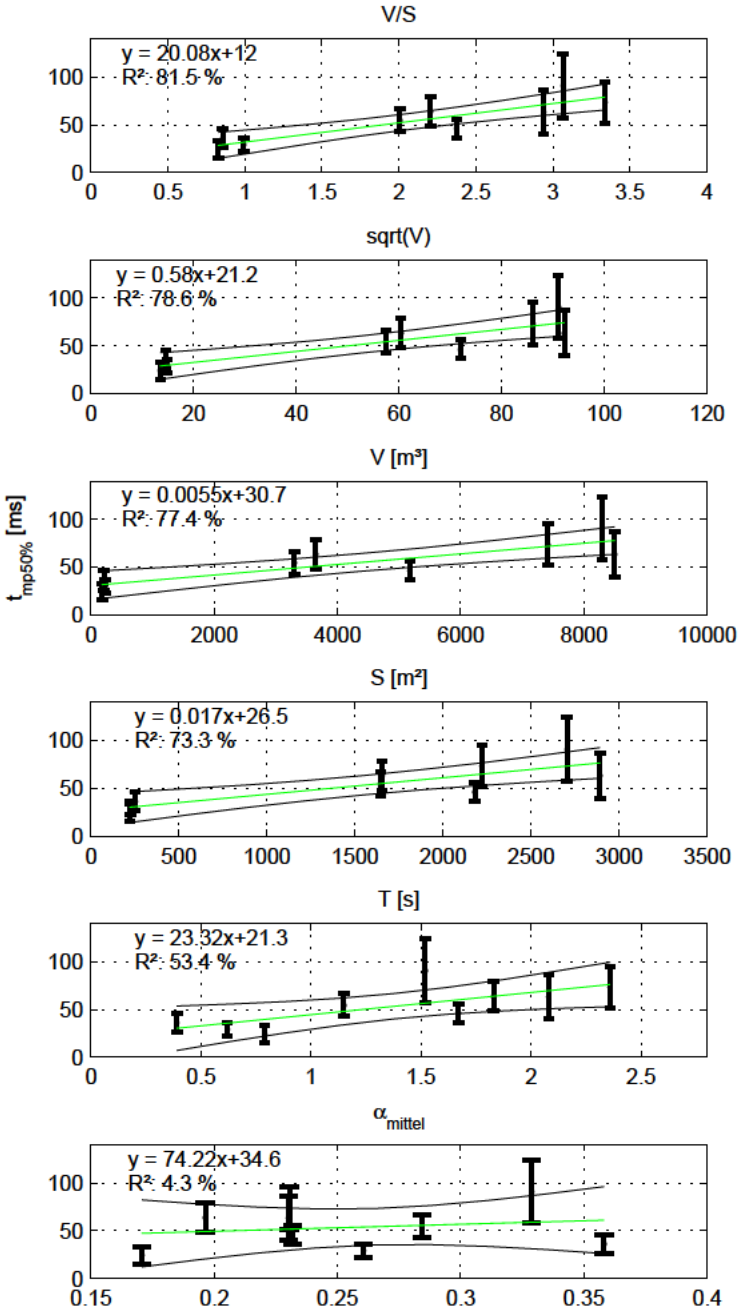



**Abbildung 5.1:** Mittelwerte der perzeptiven mixing times (inklusive 95 %-KIs) aufgetragen über den modellbasierten Prädiktoren. Lineare Regressionsmodelle und dazugehörige 95 %-Konfidenzintervalle.

### 5.1.2 95%-Perzentile

Aus der Regressionsanalyse basierend auf den 95 %-Perzentilen der empirisch ermittelten mixing time-Werte ergibt sich eine konservativere Schätzung der perzeptiven mixing time. Die Verwendung der $t_{mp95\%}$-Werte garantiert eine nahezu unhörbar exakte Simulation. Die Modelle der Mittelwertprädiktion und der 95 %-Prädiktion unterscheiden sich aufgrund der unterschiedlichen Abweichungen der perzeptiven mixing time innerhalb eines Raumes voneinander. Die Reihenfolge der Modelle, belegt durch die Varianzaufklärung, ändert sich im Vergleich zur Mittelwertprädiktion nur bei den Prädiktoren, die auch bei der vorhergehenden Betrachtung nah beieinander liegende Varinzaufklärungen hatten. Die weniger geeigneten Prädiktoren wie Nachhallzeit und mittlerer Absorptionsgrad liegen im Ranking der Prädiktoren wiederum an letzter Stelle.
Die Anstiege der Regressionsgeraden waren wie auch bei der Mittelwertprädiktion alle positiv und signifikant. Eine Ausnahme stellt der mittlere Absorptionsgrad dar, dessen Regressionskoeffizient $b$ nicht signifikant war. Das bedeutet, dass der Anstieg ebenso auch negativ sein könnte.

Die beste Varianzaufklärung konnte durch das Volumen $V$ mit $R^2$=78,7 % erreicht werden. Die beiden nachfolgenden Prädiktoren $\frac{V}{S}$ und $\sqrt{V}$ erreichten beide Varianzaufklärungen über 70 %. Insgesamt sind die Varianzaufklärungen bei der Regressionanalyse basierend auf den 95 %-Perzentilen etwas geringer als bei der Mittelwertprädiktion.

In Abbildung 5.2 sind die linearen Regressionsgleichungen mit den dazugehörigen 95 %-Konfidenzintervallen sowie die 95 %-Perzentile der perzeptiven mixing times über den modellbasierten Prädiktoren dargestellt.

Die Regressionsgleichung des besten Prädiktors Volumen der perzeptiven mixing time $t_{mp95\%}$ ist:

$$t_{mp95\%} = 0,0117 \cdot V + 50.1 \qquad [ms] \tag{5.10}$$

Die Regressionsgleichungen und Varianzaufklärungen der übrigen Prädiktoren können Abbildung 5.2 entnommen werden.

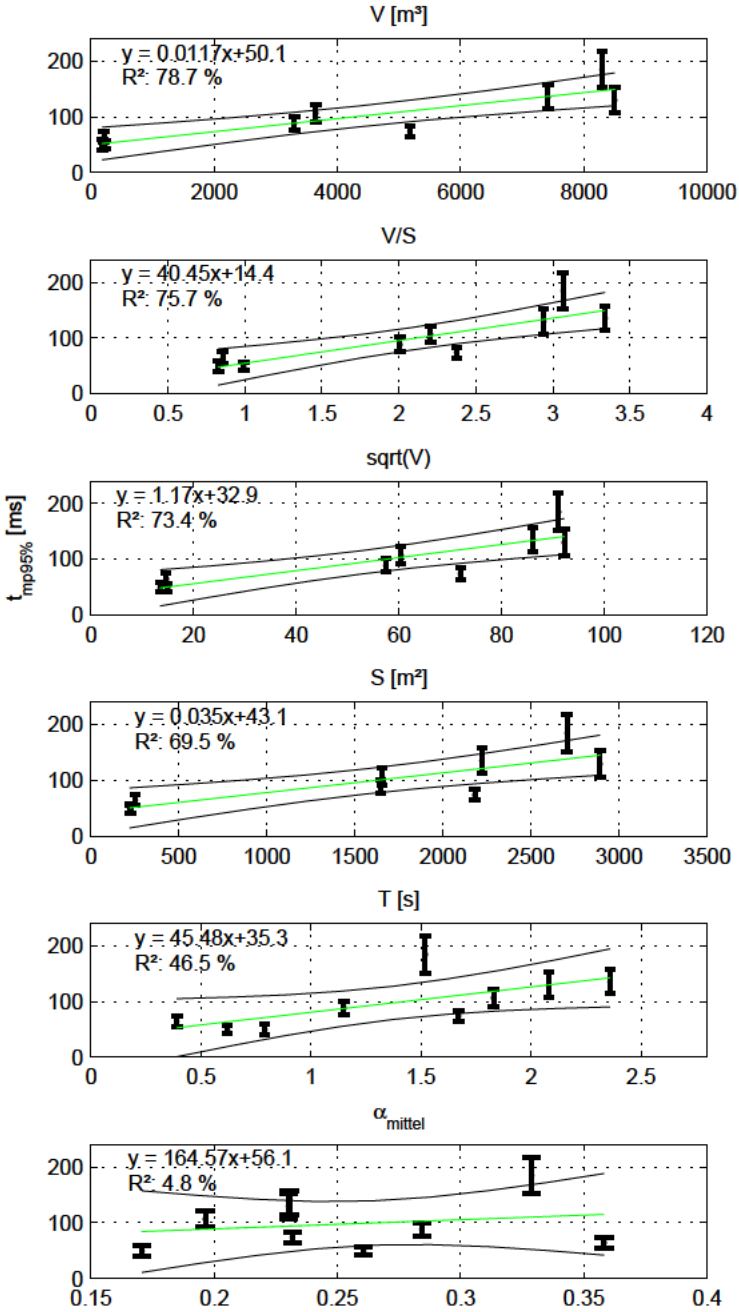



**Abbildung 5.2:** 95 %-Perzentile der perzeptiven mixing times (inklusive 95 %-KIs) aufgetragen über den modellbasierten Prädiktoren. Lineare Regressionsmodelle und dazugehörige 95 %-Konfidenzintervalle.

## 5.2 Regression empirischer Prädiktoren

Im Abschnitt 2.3 wurden vier empirische Prädiktoren der mixing time vorgestellt. Die praktische Umsetzung zeigte, dass die Prädiktoren abhängig vom Messort und der Art der Impulsantworten (einfache Impulsantwort oder binaurale Raumimpulsantwort) unterschiedliche mixing times liefern. Es konnte jedoch nicht festgestellt werden, dass eine bestimmte Art bzw. Messposition immer besonders hohe oder besonders niedrige mixing times hervorruft (siehe Abschnitt 3.3). Aus diesem Grund wurden zur Bestimmung der physikalischen mixing times die linke und rechte BRIR genutzt. Das hat auch einen praktischen Nutzen, da BRIRs im Gegensatz zu den Nachhallzeitmessungen zu jedem Datensatz gehören.

Die Regressionsanalyse wurde für jeden Prädiktor mit der mixing time der linken und rechten BRIR und dem Mittelwert aus beiden mixing times berechnet.
Um die Hörversuchssituation nachzuempfinden, auf der die perzeptiven mixing times basieren, wurde auf die Impulsantworten das Hochpassfilter, das im Hörversuch verwendet wurde, angewendet (siehe Abschnitt 3.4.1). Die Impulsantworten mit dem vollen gemessenen Frequenzbereich kamen ebenfalls zum Einsatz.

Insgesamt wurden mit Hilfe jedes empirischen Prädiktors sechs physikalische mixing times berechnet:

- linke BRIR (fullrange)
- rechte BRIR (fullrange)
- Mittelwert der mixing times basierend auf der linken und rechten BRIR (fullrange)
- linke BRIR (hochpassgefiltert)
- rechte BRIR (hochpassgefiltert)
- Mittelwert der mixing times basierend auf der linken und rechten BRIR (hochpassgefiltert)

In Abschnitt 3.3 konnte gezeigt werden, dass die Methoden von Abel und Huang (2006) sowie Stewart und Sandler (2007) weniger variante mixing times liefern, wenn sie visuell bestimmt werden. Die Eignung der abgelesenen mixing times zur Vorhersage der

perzeptiven mixing time wurde zusätzlich untersucht.

Die Methode zur Bestimmung der mixing time nach Hidaka u. a. (2007) wurde für alle Oktavbänder zwischen 125 Hz und 16 kHz bezüglich ihrer Fähigkeit, die perzeptive mixing time vorherzusagen, geprüft. Auf die Filterung der BRIRs wie im Hörversuch wurde verzichtet, da die mixing times für Frequenzbereiche errechnet wurden.

Alle impulsantwortbasierten Prädiktoren wurden bezüglich ihrer Vorhersagegenauigkeit auf Grundlage der Mittelwerte und der 95 %-Perzentile der empirisch bestimmten perzeptiven mixing times getestet.

### 5.2.1 Mittelwerte

Um einen Überblick über die Vielzahl der Regressionsmodelle zu gewinnen, wurden erst die Ergebnisse jedes einzelnen Prädiktors, die auf den verschiedenen BRIRs beruhen, betrachtet. Die Abbildungen befinden sich aufgrund ihrer großen Anzahl im Anhang.
Die Analyse der Varianzaufklärungen durch die Regressionsmodelle basierend auf den Mittelwerten der perzeptiven mixing times zeigte, dass die Regressionsmodelle der visuell bestimmten mixing times etwa 10-15 % schlechtere Varianzaufklärungen erreichten als die der automatischen Detektion. Aus diesem Grund und wegen des großen Aufwandes dieser Methode zur Bestimmung der mixing times kann die visuelle Detektion nicht empfohlen werden.
Zudem konnte festgestellt werden, dass die Regressionsmodelle der gefilterten BRIRs keine besseren Ergebnisse lieferten, weshalb diese Methode ebenfalls nicht praktikabel ist.
Die Ortsvarianz der mixing times spiegelte sich auch in den Regressionsmodellen wider. Einige Regressionsmodelle erreichten hohe Varianzaufklärungen, wenn eine bestimmte BRIR (links bzw. rechts) zur Prädiktion verwendet wurde. Jedoch trat dies zufällig für das linke bzw. rechte Ohr auf, eine Systematik konnte nicht festgestellt werden.

Diese Erkenntnisse führten dazu, dass im Folgenden für das Ranking der Vorhersagegenauigkeit der empirischen Prädiktoren nur die Mittelwerte der mixing times, die mit den linken und rechten ungefilterten BRIRs berechnet wurden, analysiert werden.

Die Anstiege der Regressionsmodelle waren nahezu alle positiv und signifikant. Davon ausgenommen ist die Methode von Stewart und Sandler (2007), deren Regressionsko-

effizient $b$ nicht signifikant war.
Die Betrachtung der Varianzaufklärungen zeigte, dass die Methode von Defrance u. a. (2009) am besten zur Vorhersage der perzeptiven mixing time $t_{mp50\%}$ geeignet ist. Dies trifft jedoch nur zu, wenn die mixing time aus Raum 3 nicht in die Regression einbezogen wird, wobei es sich dabei um einen Ausreißer handelt: Im Vergleich zu den anderen Werten, die der Algorithmus von Defrance u. a. (2009) liefert, ist die mixing time von Raum 3 ein offensichtlich falscher Wert. Das ist jedoch problematisch, wenn man anhand einer mixing time nach Defrances Methode die perzeptive mixing time bestimmen möchte. Die Methode von Defrance ist nicht reliabel, da man nie wissen kann, ob der Algorithmus einen Ausreißer liefert oder nicht. Zudem ist unklar, wie die sehr geringen mixing time-Werte, die der Algorithmus detektiert hat, interpretiert werden können. Anhand der visuell bestimmten mixing times aus den CNAs kann geschlossen werden, dass die automatische Detektion nicht den gleichen Punkt, d. h. den Übergang des CNA-Verlaufes von kubisch in linear detektiert (siehe Abschnitt 3.3.3).

Die beste Vorhersagegenauigkeit, ohne Vernachlässigung von Ausreißerwerten, konnte durch den Algorithmus von Abel und Huang (2006) mit dem ersten Kriterium zum Eintreten eines diffusen Schallfeldes erreicht werden (das Echodichteprofil nimmt das erste Mal den Wert 1 an). Die Varianzaufklärung lag bei $R^2$=74,7 %. Das erste Kriterium von Abel und Huang (2006) wird für die impulsantwortbasierte Prädiktion auf Grundlage von Mittelwerten empfohlen. Die dazugehörige Regressionsgleichung lautet wie folgt:

$$t_{mp50\%} = 0,8 \cdot t_{mix_{Abel_I}} - 8 \qquad [ms] \tag{5.11}$$

Regressionsmodelle, die im Ranking folgen, konnten nur mäßige Varianzaufklärungen erreichen. Die Vorhersage der $t_{mp50\%}$ anhand der Methode von Hidaka u. a. (2007) hatte eine Varianzaufklärung von 57,3 % bzw. 56,5 % ($t_m$ berechnet für 1 kHz und 500 Hz). Das zweite Kriterium von Abel und Huang (2006), das als mixing time den Punkt, wenn das Echodichteprofil das erste Mal größer als $1 - \sigma_{late}$ ist, definiert (mit $\sigma_{late}$ als die Standardabweichung des späten Echodichteprofils), erreichte im Vergleich zum ersten Kriterium nur eine mäßige Varianzaufklärung ($R^2$=50,7 %).
Mit der Methode von Stewart und Sandler (2007) ergab sich nur eine Varianzaufklärung von 37,6 %. Abgesehen davon war der Anstieg der Regressionsgeraden nicht signifikant,

weshalb dieses Modell zur Prädiktion der perzeptiven mixing time nicht geeignet ist. In Abbildung 5.3 sind die Regressionsmodelle der sechs besten Mittelwert-Prädiktoren dargestellt.

In Abschnitt 2.1 wurde die Annahme getroffen, dass die physikalische mixing time einen oberen Grenzwert für die perzeptive mixing time bildet. Diese Voraussetzung wurde getestet, indem geprüft wurde, ob die vorhergesagte $t_{mp50\%}$ kleiner bzw. gleich der physikalischen mixing time $t_m$ ist:

$$t_m \geq t_{mp} \tag{5.12}$$

Wenn man $t_{mp}$ in Gleichung (5.12) durch die allgemeine Gleichung der linearen Regression (Gleichung (5.1)) ersetzt, erhält man

$$t_m(1-b) \geq a \tag{5.13}$$

Die Lösung dieser Ungleichung erfordert eine Fallunterscheidung.

$$t_m \geq \frac{a}{1-b} \qquad , b < 1 \tag{5.14}$$

$$t_m \leq \frac{a}{1-b} \qquad , b > 1 \tag{5.15}$$

Gleichung (5.15) findet nur bei dem auf Defrance u. a. (2009) basierenden Regressionsmodell Anwendung. Die Anstiege aller anderen Regressionsmodelle sind kleiner 1. Zudem gibt es nur positive Zusammenhänge, ein negativer Anstieg tritt bei keinem der Regressionsmodelle auf.
Die Anwendung der beiden Ungleichungen auf die jeweiligen Regressionsmodelle zeigte, dass nur die Regressionsgleichung, die auf dem Prädiktor von Abel und Huang (2006) (Kriterium 1) basiert, immer perzeptive mixing times liefert, die kleiner als die physikalische mixing time sind. Da das erste Echodichtekriterium von Abel und Huang (2006) bisher der beste empirische Prädiktor und der einzige ist, der Gleichung (5.12) erfüllt, wird davon ausgegangen, dass dieser Prädiktor eine gute Abschätzung der perzeptiven mixing time liefert.

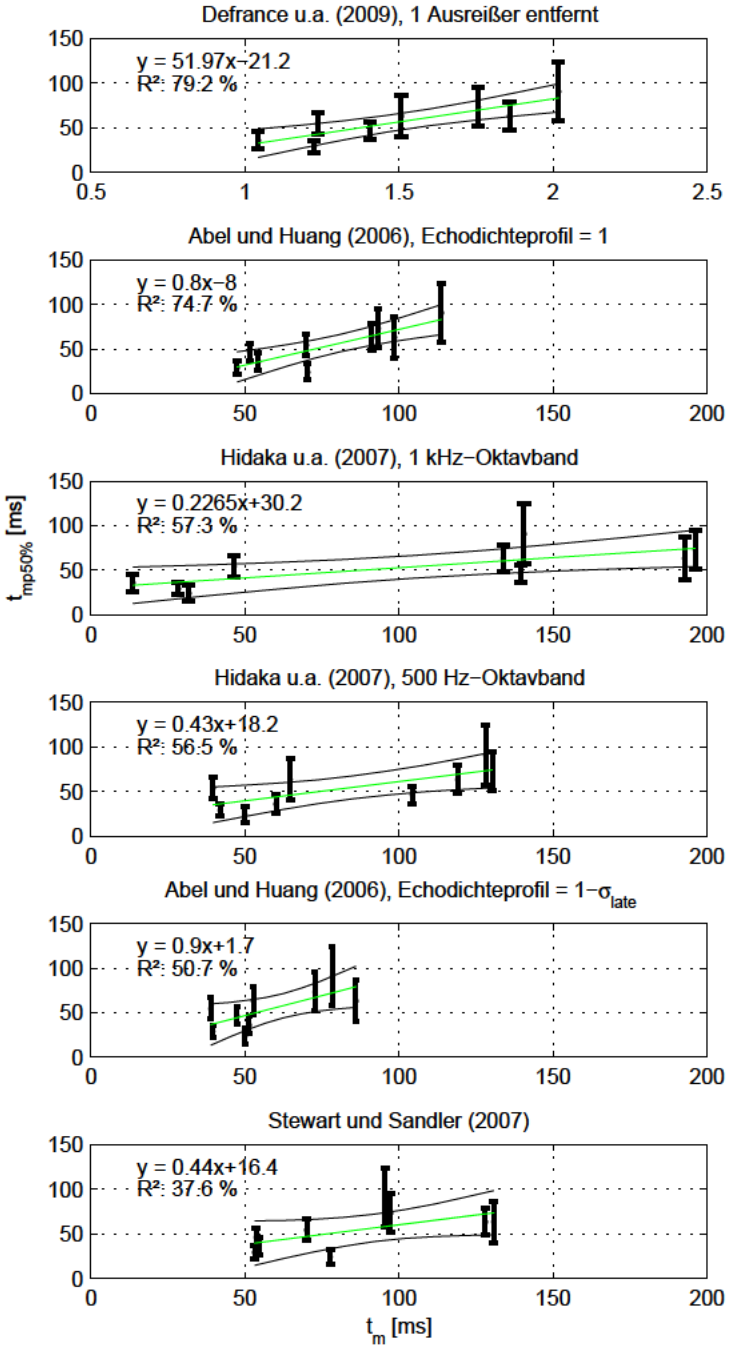



**Abbildung 5.3:** Mittelwerte der perzeptiven mixing times (inklusive 95 %-KIs) aufgetragen über den impulsantwortbasierten Prädiktoren. Lineare Regressionsmodelle und dazugehörige 95 %-Konfidenzintervalle.

### 5.2.2 95%-Perzentile

Aus den im vorigen Abschnitt 5.2.1 genannten Gründen, werden dem abschließenden Ranking der Prädiktoren auf Basis der 95 %-Perzentile der empirisch ermittelten perzeptiven mixing times die Mittelwerte der physikalischen mixing times der linken und rechten BRIR zugrunde gelegt. Wie auch bei der Prädiktion auf Basis von Mittelwerten waren die Regressionsmodelle der gefilterten Impulsantworten und visuell bestimmten mixing times nicht besser als die fullrange-Impulsantworten und automatisch detektierten mixing times.

Die Anstiege der Regressionsgeraden waren alle signifikant und positiv bis auf die Methode der normalisierten Kurtosis von Stewart und Sandler (2007), deren Anstieg wie auch bei der Mittelwertprädiktion nicht signifikant war.

Die beste Vorhersagestärke erreichte wiederum die Methode von Abel und Huang (2006) (Kriterium 1) mit einer Varianzaufklärung von $R^2$=83,7 %. Die beste Regressionsgleichung zur Prädiktion der perzeptiven mixing time ist demnach:

$$t_{mp95\%} = 1,77 \cdot t_{mix_{Abel_I}} - 38 \tag{5.16}$$

Mit dem Algorithmus von Defrance u. a. (2009) konnte wieder eine hohe Varianzaufklärung erreicht werden, jedoch nur, wenn Raum 3 nicht in die Regressionsanalyse einbezogen wurde. Das macht diesen Prädiktor untauglich zur Vorhersage der perzeptiven mixing time.
Das zweite Kriterium von Abel und Huang (2006) lieferte im Fall der 95 %-Prädiktion bessere Regressionsmodelle als Hidaka u. a. (2007), jedoch war das erste Kriterium von Abel und Huang (2006) erneut geeigneter. Die Varianzaufklärung des zweiten Abel-Kriteriums lag bei $R^2$=66,7 %. Mit der Methode von Hidaka u. a. (2007) ergaben sich nur mittelmäßige Varianzaufklärungen ($R^2$=55 % und $R^2$=49,2 %), jeweils wieder für die 1 kHz- und 500 Hz-Oktave. Die anderen Oktaven, für die die Regressionsanalyse auf Basis von Hidakas Methode durchgeführt wurde, hatten wesentlich geringere Varianzaufklärungen (siehe Abbildung C.9 im Anhang).
Ungeeignet zur Prädiktion der perzeptiven mixing time ist, wie oben schon erwähnt, die Methode von Stewart und Sandler (2007). Abgesehen davon, dass der Anstieg der Regressionsgeraden nicht signifikant war, lag die Varianzaufklärung dieses Regressions-

modells nur bei 40,3 %. Ein weiterer Hinweis, dass die Vorhersagequalität der Methode von Stewart und Sandler (2007) gering ist, sind die breiten Konfidenzintervalle des Regressionsmodells.

Abbildung 5.4 zeigt die sechs besten Regressionsmodelle, die auf Basis der 95 %-Perzentile des Hörversuchs erzeugt wurden. Die Regressionsmodelle aller weiteren untersuchten Prädiktoren befinden sich im Anhang C.2.1.

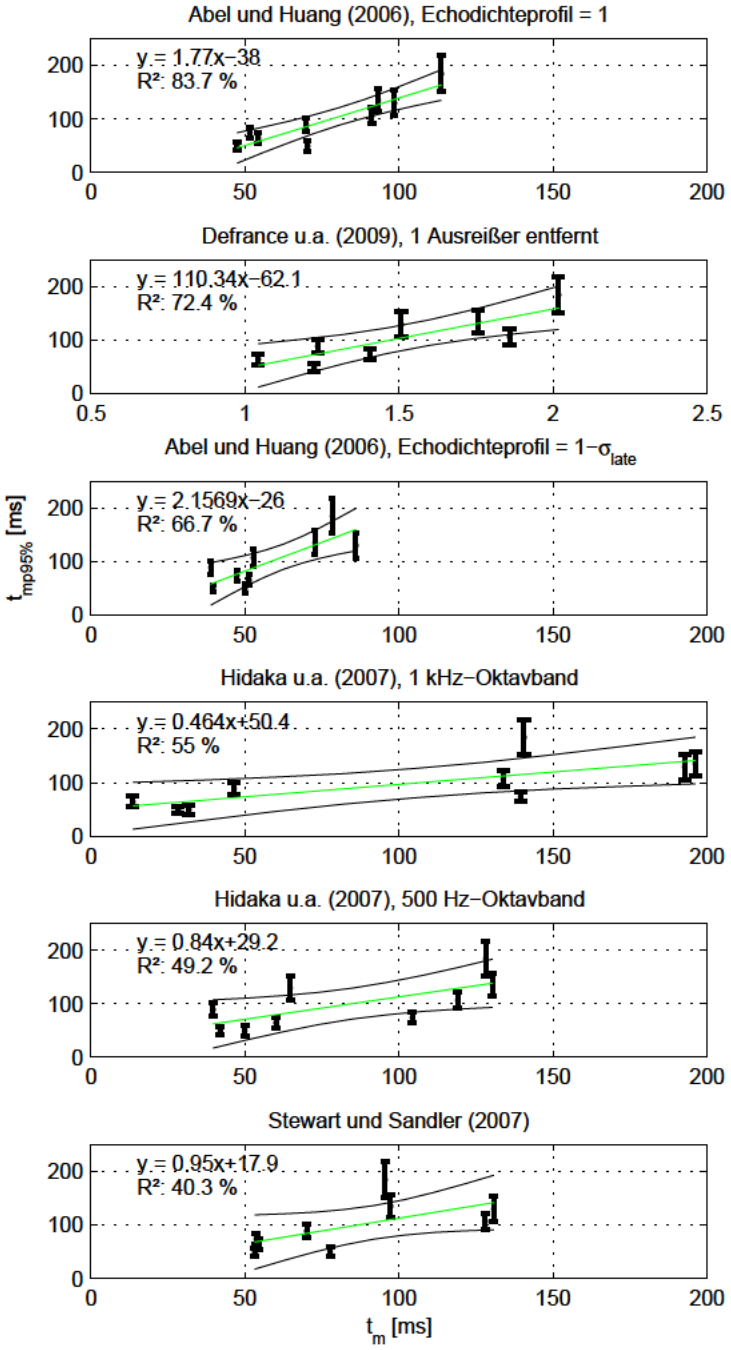



**Abbildung 5.4:** 95 %-Perzentile der perzeptiven mixing times (inklusive 95 %-KIs) aufgetragen über den impulsantwortbasierten Prädiktoren. Lineare Regressionsmodelle und dazugehörige 95 %-Konfidenzintervalle.

## 5.3 Zusammenfassung

Die Regressionsanalyse verschiedener Prädiktoren zeigte, dass für beide Möglichkeiten – modell- und impulsantwortbasiert – gute Prädiktoren existieren. Die Vorhersage der perzeptiven mixing time ist folglich für die modellbasierte Auralisation und die Auralisation basierend auf gemessenen Impulsantworten möglich.
Regressionsmodelle, deren Grundlage die Mittelwerte der empirisch ermittelten mixing times sind, erreichten insgesamt betrachtet mit den modellbasierten Prädiktoren etwas bessere Varianzaufklärungen. Bei den Regressionsmodellen, die auf Basis der 95 %-Perzentile der empirisch ermittelten Schwellwerte bestimmt wurden, war es umgekehrt: In diesem Fall lieferten die empirischen Prädiktoren bessere Varianzaufklärungen. Insgesamt betrachtet, erreichten die modellbasierten Prädiktoren die besten Varianzaufklärungen. Vier der sechs getesteten Prädiktoren lagen im guten bis sehr guten Bereich.[2] Von den empirischen Prädiktoren erreichte bei der Mittelwert- sowie auch bei der 95 %-Prädiktion nur ein Prädiktor eine gute bzw. sehr gute Varianzaufklärung.[3]
Der Grund für das eher schlechte Abschneiden der empirischen Prädiktoren wird in den im Abschnitt 3.3 erläuterten Problemen der Konsistenz der mixing time-Werte vermutet.

Bester modellbasierter Prädiktor der Mittelwerte der perzeptiven mixing time war $\frac{V}{S}$, der Kern der Formel zur Bestimmung der mittleren freien Weglänge $l_m$. Die Raumoberfläche $S$ wurde aus den drei Hauptdimensionen des Raumes berechnet.
Die Regression auf Basis der 95 %-Perzentile der empirisch ermittelten mixing times dient dazu, einen konservativeren Wert für die perzeptive mixing time zu erhalten, bei dem davon ausgegangen werden kann, dass die darauf basierende Simulation exakt ist. In diesem Fall war der beste modellbasierte Prädiktor das Volumen.

Der beste empirische, d. h. impulsantwortbasierte, Prädiktor war in beiden Betrachtungen das erste Kriterium von Abel und Huang (2006). Die Varianzaufklärung bei der 95 %-Prädiktion erreichte den höchsten Wert aller betrachteten Prädiktoren. Die Ergebnisse der empirischen Prädiktoren beruhen auf den gemittelten $t_m$ aus den linken und rechten Impulsantworten.

[2] Als sehr gut werden Varianzaufklärungen mit $R^2$ 80-100 % aufgefasst. Gute Varianzaufklärungen liegen zwischen 70 % und 80 %.
[3] Die Regressionsergebnisse von Defrance u. a. (2009) sind davon ausgeschlossen.

Die Methoden von Defrance u. a. (2009) und Stewart und Sandler (2007) sind nicht zur Prädiktion der $t_{mp}$ geeignet: Der Algorithmus von Defrance liefert nicht-reliable Werte und das Regressionsmodell zur Methode von Stewart ist nicht signifikant.
Außerdem kann das zweite Kriterium von Abel und Huang (2006) nicht zur Prädiktion empfohlen werden, da die physikalischen mixing times stark von der gewählten Länge der Impulsantwort abhängen.

Die Ergebnisse der Regressionsanalyse der modellbasierten Prädiktoren und der sechs besten empirischen Prädiktoren, basierend auf den Mittelwerten der berechneten mixing times der linken und rechten BRIR, sind in Tabelle 5.1 zusammengefasst. Ein komplettes Ranking aller untersuchten Prädiktoren mit gefilterten und ungefilterten Impulsantworten, einzelner Untersuchung der linken und rechten BRIRs und deren Mittelwerten sowie der visuell bestimmten mixing times befindet sich im Anhang C.2.1.

| | Modellbasierte Prädiktoren | | Empirische Prädiktoren | |
|---|---|---|---|---|
| # | $t_{mp50\%}$ | $t_{mp95\%}$ | $t_{mp50\%}$ | $t_{mp95\%}$ |
| 1 | $\frac{V}{S}$<br>$R^2$ 81,5 % | $V$<br>$R^2$ 78,7 % | Abel I<br>$R^2$ 74,7 % | Abel I<br>$R^2$ 83,7 % |
| 2 | $\sqrt{V}$<br>$R^2$ 78,6 % | $\frac{V}{S}$<br>$R^2$ 75,7 % | Hidaka 1 kHz<br>$R^2$ 57,3 % | Abel II<br>$R^2$ 66,7 % |
| 3 | $V$<br>$R^2$ 77,4 % | $\sqrt{V}$<br>$R^2$ 73,4 % | Hidaka 500 Hz<br>$R^2$ 56,5 % | Hidaka 1 kHz<br>$R^2$ 55 % |
| 4 | $S$<br>$R^2$ 73,3 % | $S$<br>$R^2$ 69,5 % | Abel II<br>$R^2$ 50,7 % | Hidaka 500 Hz<br>$R^2$ 49,2 % |
| 5 | $T_{30_{125Hz-4kHz}}$<br>$R^2$ 53,4 % | $T_{30_{125Hz-4kHz}}$<br>$R^2$ 46,5 % | Stewart<br>$R^2$ 37,6 % | Stewart<br>$R^2$ 40,3 % |
| 6 | $\alpha_{mittel}$<br>$R^2$ 4,3 % | $\alpha_{mittel}$<br>$R^2$ 4,8 % | Defrance<br>$R^2$ 79,2 %* | Defrance<br>$R^2$ 72,4 %* |

**Tabelle 5.1:** Ranking der besten Prädiktoren der perzeptiven mixing time
grün $100\% \leq R^2 \leq 80\%$, gelb $80\% \leq R^2 \leq 60\%$, orange $60\% \leq R^2 \leq 40\%$, rot $R^2 < 40\%$
(*ausgeschlossen)

In Abschnitt 4.2 wurde darauf hingewiesen, dass die Schwellwerte einer einzelnen Ver-

suchsperson in einigen Räumen als Ausreißer zu betrachten sind. Wenn die Ergebnisse dieser Versuchsperson von der Regressionsanalyse ausgeschlossen wurden, waren die Varianzaufklärungen etwas geringer. Die Reihenfolge der Prädiktoren änderte sich dadurch jedoch nicht.

Die Ergebnisse der Regressionsanalyse wurden in Matlab® implementiert. Damit ist eine schnelle Berechnung der perzeptiven mixing time mit unterschiedlichen Eingabeparametern möglich (siehe beiliegende CD).

# 6 Zusammenfassung und Ausblick

Für die Auralisation von virtuellen akustischen Umgebungen mittels dynamischer Binauralsynthese werden große Datenmengen in Echtzeit verarbeitet, was mit einem großen Rechenaufwand einhergeht.
Deshalb werden perzeptive Reduktionsmöglichkeiten gesucht, ohne dass die Auralisation an Qualität verliert. Der Ansatz zur Reduktion des Rechenaufwands bei der Auralisation bestand in dieser Arbeit darin, den individuellen Nachhall einer Impulsantwort durch einen für den gesamten Raum repräsentativen Nachhall zu ersetzen. Untersuchungsgegenstand war die perzeptive mixing time, d. h. der Zeitpunkt, ab dem der Nachhall einer Impulsantwort ohne hörbare Konsequenzen durch den Nachhall einer beliebigen Quell-/Empfängerkonstellation ausgetauscht werden kann. Zudem wurde der Zusammenhang der perzeptiven mixing time und Parametern des Raumes untersucht.

Dazu wurden binaurale Raumimpulsantworten im Diffusfeld von unterschiedlichen Räumen mit einem Kunstkopftorsosimulator gemessen. Die Auswahl der Räume beschränkte sich auf rechteckförmige Räume, bei denen wegen der großen mittleren freien Weglängen von einer spät eintreffenden Diffusität ausgegangen wird.
Die Räume wurden so ausgewählt, dass sich ein 3x3-Design ergab, in dem das Volumen und der mittlere Absorptionsgrad bzw. die relative Bedämpfung jeweils dreifach gestuft wurden. Das Untersuchungsschema umfasste demzufolge neun Räume.

Um den perzeptiven Übergangszeitpunkt in den diffusen Nachhall zu untersuchen, wurde ein Hörversuch mit einem 3-AFC-Testdesign durchgeführt. Der Hörversuch wurde mit dynamischer Binauralsynthese realisiert. Dabei mussten zwei Simulationen voneinander unterschieden werden: Bei der Referenzsimulation wurden die BRIRs entsprechend der Kopfposition in vollständiger Länge ausgetauscht, bei der manipulierten Simulation wurde nur der frühe Anteil der Impulsantwort entsprechend der Kopfposition nachgeführt, der späte Anteil der Impulsantwort wurde einer statischen Kopfposition

mit Blickrichtung nach vorn entnommen. Der Überblendzeitpunkt der dynamischen Simulation in eine statische wurde adaptiv an die Erkennungsleistung der Versuchspersonen angepasst.
Die Unterscheidung der beiden Simulationen stellte sich als schwierig heraus. Der Hörversuch lieferte relativ geringe Schwellwerte.
Eine Varianzanalyse konnte einen Volumeneffekt belegen. Es konnte gezeigt werden, dass die perzeptive mixing time in Rechteckräumen mit dem Raumvolumen ansteigt. Ein Effekt des mittleren Absorptionsgrades konnte nicht nachgewiesen werden. Ebenso ist eine Interaktion der beiden untersuchten Faktoren nicht interpretierbar.

Mit einer Regressionsanalyse sollte verdeutlicht werden, inwiefern sich verschiedene Prädiktoren der physikalischen mixing time zur Schätzung der perzeptiven mixing time eignen.
Als Prädiktoren dienten modellbasierte, d. h. auf den physikalischen Eigenschaften eines Raumes beruhende Maße und empirische, d. h. auf Impulsantworten basierende Maße. Die Regressionsanalyse wurde jeweils basierend auf den Mittelwerten und den 95 %-Perzentilen der empirisch ermittelten Schwellwerte durchgeführt. Die Prädiktion auf Basis der 95 %-Perzentile liefert einen konservativen Wert für den Überblendzeitpunkt und garantiert eine nahezu unhörbar exakte Simulation.

Sowohl unter den modellbasierten als auch den empirischen Prädiktoren existieren reliable Prädiktoren der perzeptiven mixing time.
Es zeigte sich, dass die mittlere freie Weglänge als modellbasierter Prädiktor, basierend auf dem Volumen und den drei Hauptdimensionen eines Raumes, zur Prädiktion anhand von Mittelwerten am besten geeignet ist. Als bester modellbasierter Prädiktor für eine konservative perzeptive mixing time stellte sich das Raumvolumen heraus.
Der geeignetste empirische Prädiktor war ein Verfahren zur Erstellung von Echodichteprofilen nach Abel und Huang (2006). Die Schätzung der perzeptiven mixing time erfolgte auf Basis von Mittelwerten der physikalischen mixing times der linken und rechten BRIRs.

Kürzlich erschienene Artikel beschreiben neue Methoden zur Bestimmung der physikalischen mixing time, die ebenfalls auf der Auswertung von Impulsantworten beruhen.
Ein interessanter Ansatz nimmt die Schroederfrequenz zum Vorbild zur Bestimmung des Übergangs in einen diffusen Nachhall, indem eine bestimmte Anzahl der Reflexionen innerhalb einer festgelegten Pulsdauer, die an die Halbwertsbreite einer Mode

angelehnt ist, betrachtet wird. Durch die Bestimmung der mittleren Dauer der Reflexionen wird die Dämpfung des Raumes in dieser neuen Methode im Gegensatz zu den bisher untersuchten empirischen Methoden berücksichtigt (Jeong u. a., 2010).
Defrance und Polack (2010) entwickelten einen weiteren Algorithmus zur Bestimmung der mixing time, dessen Grundlage die Fouriertransformation von Impulsantwortausschnitten mit ansteigender Fensterlänge bildet. Die Autoren räumen ein, dass die Ergebnisse, die der Algorithmus liefert, stark von den Parametern der Berechnung abhängen. Aus diesem Grund sowie aufgrund der Ähnlichkeit der Ergebnisse zu denen der Matching-Pursuit-Methode (Defrance u. a., 2009), die nach Aussage der Autoren besteht, ist die Eignung dieser Methode zur Bestimmung der mixing time fraglich.
Nichtsdestotrotz könnten diese beiden Prädiktoren ebenfalls einer Regressionsanalyse unterzogen werden, um ihre Eignung als Prädiktor der perzeptiven mixing time zu testen.

Für die Zukunft stellt sich die Frage, ob die Ergebnisse auch auf andere Kunstkopftorsosimulatoren übertragbar sind, oder ob individuelle Eigenschaften wie Ohrabstand und Kopf-/bzw. Außenohrgeometrie Einfluss auf die Wahrnehmbarkeit der Diffusfeldgrenze haben.
Zudem sind die Ergebnisse dieser Arbeit nicht dahingehend verallgemeinerbar, dass jeder beliebige Nachhall einer beliebigen Quell-/Empfängerkonstellation zur Simulation des statischen Anteils während der dynamischen Binauralsynthese genutzt werden kann. Die perzeptiven mixing times wären vermutlich höher, wenn der späte Nachhall aus einer anderen Quellposition entnommen worden wäre (siehe auch die Ergebnisse von Lindau u. a. (2007)).

Die Ergebnisse dieser Arbeit können in Zukunft in weiteren Hörversuchen zur dynamischen Binauralsynthese genutzt werden. Zudem können Schallfelder realer und modellbasierter Räume auf Basis der gefundenen Regressionsgleichungen plausibel und mit einem angemessenen Rechenaufwand simuliert werden.

# Literatur


Abel, Jonathan S. und Patty Huang (2006). "A Simple, Robust Measure of Reverberation Echo Density". In: *Proc. of the 121st AES Convention.* San Francisco.

Begault, Durand R. (1992). "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems". In: *J. Audio Eng. Soc.* 40.11, S. 895–904.

Behler, Gottfried K. (2000). "Ein neuer Messlautsprecher für die Raumakustik". In: *DAGA*. Oldenburg.

Behrens, Tobias (2007). "Raumakustische Konzeption von Wiedergaberäumen für Wellenfeldsynthese am Beispiel eines Hörsaals der TU Berlin". In: *DAGA*. Stuttgart.

Bortz, Jürgen (2005). *Statistik für Human- und Sozialwissenschaftler.* 6. Aufl. Heidelberg: Springer.

Bortz, Jürgen und Nicola Döring (2006). *Forschungsmethoden und Evaluation.* 4. Aufl. Heidelberg: Springer.

Bradley, John S. und Gilbert A. Soulodre (1995). "The influence of late arriving energy on spatial impression". In: *J. Acoust. Soc. Am* 97.4, S. 2263–2271.

Ciba, Simon (2009). *WhisPER v1.0 - User documentation.*

Ciba, Simon, André Wlodarski und Hans-Joachim Maempel (2009). "WhisPER - A new tool for performing listening tests". In: *Proc. of the 126th AES Convention.* preprint 7749. München.

Cremer, Lothar (1961). *Die wissenschaftlichen Grundlagen der Raumakustik (Band II). Statistische Raumakustik.* Stuttgart: S. Hirzel Verlag.

Cremer, Lothar und Helmut Müller (1978). *Die wissenschaftlichen Grundlagen der Raumakustik (Band I). Geometrische Raumakustik. Statistische Raumakustik. Psychologische Raumakustik.* 2. Aufl. Stuttgart: S. Hirzel Verlag.

Davenport, Wilbur B. und William L. Root (1958). *An Introduction to the Theory of Random Signals and Noise.* New York Toronto London: McGraw-Hill Book Company, Inc.

Defrance, G. und J.-.D Polack (2010). "Estimating the mixing time of concert halls using the eXtensible Fourier Transform". In: *Applied Acoustics* 71.9, S. 777–792.

Defrance, G., L. Daudet und J.-D. Polack (2009). "Using Matching Pursuit for Estimating Mixing Time Within Room Impulse Responses". In: *Acta Acustica united with Acustica* 95, S. 1071–1081.

Griesinger, David (1989). "Practical Processors and Programs for Digital Reverberation". In: *Proc. of the 7th International AES Conference: Audio in Digital Times.*

Hidaka, Takayuki, Leo L. Beranek und Toshiyuki Okano (1995). "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls". In: *J. Acoust. Soc. Am* 98.2, S. 988–1007.

Hidaka, Takayuki, Yoshinari Yamada und Takehiko Nakagawa (2007). "A new definition of boundary point between early reflections and late reverberation in room impulse responses". In: *J. Acoust. Soc. Am* 122.1, S. 326–332.

Hirano, Katuomi (2005). "Rayleigh Distribution". In: *Encyclopedia of statistical sciences.* Hrsg. von Samuel Kotz u. a. Hoboken, New Jersey: John Wiley & Sons, S. 6986–6988.

Huang, Patty und Jonathan S. Abel (2007). "Aspects of Reverberation Echo Density". In: *Proc. of the 123rd AES Convention.* preprint 7163. New York.

Janssen, Jürgen und Wilfried Laatz (2007). *Statistische Datenanalyse mit SPSS für Windows.* 6. Aufl. Berlin Heidelberg: Springer.

Jeong, Cheol-Ho, Jonas Brunskog und Finn Jacobsen (2010). "Room acoustic transition time based on reflection overlap". In: *J. Acoust. Soc. Am.* 127.5, S. 2733–2736.

Jot, Jean-Marc, Laurent Cerveau und Olivier Warusfel (1997). "Analysis and synthesis of room reverberberation based on a statistical time-frequency model". In: *Proc. of the 103rd AES Convention.* New York.

Joyce, W.B. (1975). "Sabine's reverberation time and ergodic auditoriums". In: *J. Acoust. Soc. Am.* 58.3, S. 643–655.

Junius, W. (1959). "Raumakustische Untersuchungen mit neueren Messverfahren in der Liederhalle Stuttgart". In: *Acustica* 9.4, S. 289–303.

Kuttruff, H. und E. Mommertz (2004). "Raumakustik". In: *Taschenbuch der Technischen Akustik.* Hrsg. von Gerhard Müller und Michael Möser. 3. Aufl. Berlin Heidelberg: Springer. Kap. 11.

Kuttruff, Heinrich (1979). *Room Acoustics.* 2. Aufl. London: Applied Science Publisher Ltd.

Kuttruff, Heinrich (1991). "On the Audibility of Phase Distortion in Rooms and its Significance for Sound Reproduction and Digital Simulation in Room Acoustics". In: *Acustica* 74, S. 3–7.

– (1993). "Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results". In: *J. Audio Eng. Soc.* 41.11, S. 876–880.

– (2000). *Room Acoustics.* 4. Aufl. London: Spon Press.

– (2004). *Akustik. Eine Einführung.* Stuttgart: S. Hirzel Verlag.

Lindau, Alexander (2006). "Ein Instrument zur softwaregestützten Messung binauraler Impulsantworten in mehreren Freiheitsgraden". Magisterarbeit. Technische Universität Berlin.

Lindau, Alexander, Torben Hohn und Stefan Weinzierl (2007). "Binaural resynthesis for comparative studies of acoustical environments". In: *Proc. of the 112th AES Convention.* Wien.

Meesawat, Kittiphong und Dorte Hammershøi (2003). "The time when the reverberation tail in a binaural room impulse response begins". In: *Proc. of the 115th AES Convention.* preprint 5859. New York.

Nikias, Chrysostomos L. und Athina P. Petropulu (1993). *Higher-order spectra analysis: a nonlinear signal processing framework.* Englewood Cliffs: Prentice Hall.

Olive, Sean E. und Floyd E. Toole (1989). "The Detection of Reflections in Typical Rooms". In: *J. Audio Eng. Soc.* 37.7/8, S. 539–553.

Otto, Stefanie (2008). "Vergleichende Simulation adaptiver, psychometrischer Verfahren zur Schätzung von Wahrnehmungsschwellen". Magisterarbeit. Technische Universität Berlin.

Polack, Jean-Dominique (1988). "La transmission de l'énergie sonore dans les salles". Dissertation. Université du Maine.

– (1992). "Modifying Chambers to play Billiards: the Foundations of Reverberation Theory". In: *Acustica* 76, S. 257–272.

– (1993). "Playing Billiards in the Concert Hall: The Mathematical Foundations of Geometrical Room Acoustics". In: *Applied Acoustics* 38, S. 235–244.

Reichardt, W. und U. Lehmann (1978). "Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit, Erläuterungen des Raumeindrucksmaßes R". In: *Acustica* 40.5, S. 277–290.

Reichardt, W., O. Abdel Alim und W. Schmidt (1974). "Abhängigkeit der Grenzen zwischen brauchbarer und unbrauchbarer Durchsichtigkeit von der Art des Musikmo-

tives, der Nachhallzeit und der Nachhalleinsatzzeit". In: *Applied Acoustics* 7, S. 243–264.

Reilly, Andrew und David McGrath (1995). "Real-time auralization with head-tracking". In: *Proc. of the 5th Australian Regional AES Convention.* Sydney.

Reilly, Andrew, David McGrath und Bengt-Inge Dalenbäck (1995). "Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers". In: *Proc. of the 99th AES Convention.* New York.

Rubak, Per und Lars G. Johansen (1999). "Artificial Reverberation based on a Pseudo-random Impulse Response II". In: *Proc. of the 106th AES Convention.* preprint 4900. München.

Schmidt, Wolfgang und Wolfgang Ahnert (1973). "Einfluss der Richtungs- und Zeitdiffusität auf den Raumeindruck". In: *Wiss. Z. Techn. Univers. Dresden* 22.2, S. 313–317.

Schreiber, L. (1960). "Was empfinden wir als gleichförmiges Rauschen?" In: *Acustica* 14.12, S. 399–403.

Schärer, Zora und Alexander Lindau (2009). "Evaluation of Equalization Methods for Binaural Signals". In: *Proc. of the 126th AES Convention.* München.

Schroeder, Manfred R. (1954). "Die statistischen Parameter der Frequenzkurve von großen Räumen". In: *Acustica* 4, S. 594–600.

– (1959). "Measurement of Sound Diffusion in Reverberation Chambers". In: *J. Acoust. Soc. Am.* 31.11, S. 1407–1417.

– (1962). "Natural Sounding Artificial Reverberation". In: *J. Audio Eng. Soc.* 10.3, S. 219–223.

– (1965). "New Method of Measuring Reverberation Time". In: *Journal of the Acoustical Society of America* 37.3, S. 409–412.

– (1996). "The 'Schroeder frequency' revisited". In: *J. Acoust. Soc. Am.* 99.5, S. 3240–3241.

Slavik, Karl M. und Stefan Weinzierl (2008). "Wiedergabeverfahren". In: *Handbuch der Audiotechnik.* Hrsg. von Stefan Weinzierl. Berlin Heidelberg: Springer. Kap. 11.

Stanley, William D. und Steven J. Peterson (1979). "Equivalent Statistical Bandwidths of Conventional Low-Pass Filters". In: *IEEE Transactions on Communications* 27.10, S. 1633–1634.

Stewart, Rebecca und Mark Sandler (2007). "Statistical Measures Of Early Reflections Of Room Impulse Responses". In: *10th Int. Conference on Digital Audio Effects*. Bordeaux.

Treutwein, Bernhard (1995). "Adaptive Psychophysical Procedures". In: *Vision Research* 35.17, S. 2503–2522.

Weinzierl, Stefan (2008). "Aufnahmeverfahren". In: *Handbuch der Audiotechnik*. Hrsg. von Stefan Weinzierl. Berlin Heidelberg: Springer. Kap. 10.

Zenga, Michele (2005). "Kurtosis". In: *Encyclopedia of statistical sciences*. Hrsg. von Samuel Kotz u. a. Hoboken, New Jersey: John Wiley & Sons, S. 3902–3915.

# Abbildungsverzeichnis


1.1 Prinzip der dynamischen Binauralsynthese . . . . . 3

2.1 Gültigkeitsbereich des stochastischen Modells einer Impulsantwort . . . 9
2.2 Impulsantwort und Echodichteprofil (Abel und Huang, 2006) . . . . . 17
2.3 Impulsantwort und Kurtosisverlauf (Stewart und Sandler, 2007) . . . . 18
2.4 CNA mit kubischem und linearem Fit (Defrance u. a., 2009) . . . . . 19
2.5 „Wasserfall“-Diagramm der abklingenden Energie einer IR . . . . . 23
2.6 Korrelationsverlauf nach Hidaka u. a. (2007) (500 Hz) . . . . . 24

3.1 Maßstabsgetreue Grundrisse der neun Räume . . . . . 31
3.2 Stimuluskontinuum . . . . . 31
3.3 Messaufbau: Platzierung von FABIAN und Lautsprecher . . . . . 32
3.4 Gemessene Nachhallzeiten der neun untersuchten Räume . . . . . 33
3.5 Messaufbau: verwendete Geräte . . . . . 34
3.6 Kunstkopftorsosimulator und Lautsprecher im Detail . . . . . 35
3.7 Beispiele einiger Messaufbauten . . . . . 35
3.8 Ortsvarianz der $t_m$-Werte nach Abel und Huang (2006) . . . . . 37
3.9 Beispielhafte Echodichteprofile . . . . . 38
3.10 Ortsvarianz der $t_m$-Werte (visuell) nach Abel und Huang (2006) . . . . 39
3.11 Ortsvarianz der $t_m$-Werte nach Stewart und Sandler (2007) . . . . . 40
3.12 Beispielhafte Kurtosisverläufe . . . . . 41
3.13 Ortsvarianz der $t_m$-Werte (visuell) nach Stewart und Sandler (2007) . . 41
3.14 $t_m$-Werte nach Defrance u. a. (2009) . . . . . 42
3.15 Normierte CNAs (5 Messpositionen in Raum 5) . . . . . 43
3.16 Ortsvarianz der $t_m$-Werte der 500 Hz-Oktave nach Hidaka u. a. (2007) . 44
3.17 Beispielhafter Korrelationsverlauf . . . . . 45
3.18 Verwendetes Filter im Hörversuch . . . . . 48
3.19 Ergebnisse des Vorversuchs . . . . . 49
3.20 Versuchsaufbau . . . . . 50
3.21 Benutzeroberfläche des 3-AFC-Testdesigns in WhisPER . . . . . 51
3.22 Im adaptiven Hörversuch zu unterscheidende Situationen . . . . . 52

4.1 Hörversuchsergebnisse aller 24 Versuchspersonen . . . . . 58
4.2 Beispielhaft zwei kritische Verläufe der Schwellenmessung . . . . . 59
4.3 Schwellwerte der 10 besten VPn als Boxplots . . . . . 60

4.4 Mittelwerte und 95 %-Konfidenzintervalle der perzeptiven mixing time 61
4.5 Interaktionsdiagramme . . . . . 63

5.1 Lineare Regressionsmodelle (modellbasierte Prädiktoren und Mittelwerte der $t_{mp}$) . . . . . 71
5.2 Lineare Regressionsmodelle (modellbasierte Prädiktoren und 95 %-Perzentile der $t_{mp}$) . . . . . 73
5.3 Lineare Regressionsmodelle (impulsantwortbasierte Prädiktoren und Mittelwerte der $t_{mp}$) . . . . . 78
5.4 Lineare Regressionsmodelle (impulsantwortbasierte Prädiktoren und 95 %-Perzentile der $t_{mp}$) . . . . . 81

C.1 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und Mittelwerte der $t_{mp}$, linke BRIR) . . . . . 105
C.2 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und Mittelwerte der $t_{mp}$, rechte BRIR) . . . . . 106
C.3 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und Mittelwerte der $t_{mp}$, Mittelwert (linke/rechte BRIR)) . . . . . 107
C.4 Lineare Regressionsmodelle (Prädiktor Defrance u. a. (2009) und Mittelwerte der $t_{mp}$) . . . . . 107
C.5 Lineare Regressionsmodelle (Prädiktor Abel und Huang (2006) und Mittelwerte der $t_{mp}$) . . . . . 108
C.6 Lineare Regressionsmodelle (Prädiktor Stewart und Sandler (2007) und Mittelwerte der $t_{mp}$) . . . . . 109
C.7 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und 95 %-Perzentile der $t_{mp}$, linke BRIR) . . . . . 110
C.8 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und 95 %-Perzentile der $t_{mp}$, rechte BRIR) . . . . . 111
C.9 Lineare Regressionsmodelle (Prädiktor Hidaka u. a. (2007) und 95 %-Perzentile der $t_{mp}$, Mittelwert (linke/rechte BRIR)) . . . . . 112
C.10 Lineare Regressionsmodelle (Prädiktor Defrance u. a. (2009) und 95 %-Perzentile der $t_{mp}$) . . . . . 112
C.11 Lineare Regressionsmodelle (Prädiktor Abel und Huang (2006) und 95 %-Perzentile der $t_{mp}$) . . . . . 113
C.12 Lineare Regressionsmodelle (Prädiktor Stewart und Sandler (2007) und 95 %-Perzentile der $t_{mp}$) . . . . . 114

# Tabellenverzeichnis

3.1 Raumauswahl . . . . . 28
3.2 unterste und oberste realisierbare Stufen der Überblendung im Hörversuch 47

4.1 Mittelwerte der perzeptiven mixing times je Raum . . . . . 61

5.1 Ranking der besten Prädiktoren der perzeptiven mixing time . . . . . 83

B.1 Raumabmessungen, Oberflächen und Hallradien der untersuchten Räume 102
B.2 Nachhallzeiten der untersuchten Räume . . . . . 102

C.1 Deskriptive Auswertung . . . . . 103
C.2 Test von Mauchly . . . . . 103
C.3 Zweifaktorielle ANOVA . . . . . 104
C.4 Ranking der Varianzaufklärung der Mittelwertprädiktion . . . . . 116
C.5 Ranking der Varianzaufklärung der 95 %-Prädiktion . . . . . 118
C.6 Signifikanz der Anstiege der Regressionsgeraden (modellbasierte Prädiktoren) . . . . . 119
C.7 Signifikanz der Anstiege der Regressionsgeraden (impulsantwortbasierte Prädiktoren) . . . . . 119

# A Quellcodes

Es werden hier die wichtigsten Quellcodes zur Berechnung der physikalischen mixing time durch impulsantwortbasierte und modellbasierte Methoden gezeigt, wobei der Vorspann teilweise etwas gekürzt wurde. Nicht eigenentwickelte Quellcodes, Quellcodes mit untergeordneter Bedeutung, Quellcodes zur Prädiktion der perzeptiven mixing time und die m-files zur statistischen Auswertung des Hörversuches inklusive Regressionsanalysen befinden sich auf der beiliegenden CD.

## A.1 modellbasierte Prädiktoren

```
function [t_rubak_4000,t_rubak_mfw,t_cremer_2000,t_hidaka_RT,t_schmidt_250,...
...t_reichardt_sqrtV_1000] = geom(V,S,T)
%
% computes the starting time for stochastic reverberation in an impulse
% response based on geometric and physical properties of a room
%
% [t_rubak_4000, t_rubak_mfw, t_cremer_2000, t_hidaka_RT,t_schmidt_250,...
% ...t_reichardt_sqrtV_1000] = geom(V,S,T)
%
% V - volume
% S - surface area
% T - reverberation time
%
% (C) L. Kosanke, TU Berlin, audio communication group, 2009
%-------------------------------------------------------------------------%

c           = 340;          %speed of sound
dN_dt_r     = 4000;         %number of reflections per second, that are
                            %necessary for stochastic reverberation
                            %(Rubak1999)
dN_dt_c     = 2000;         %number of reflections per second, that are
                            %necessary for stochastic reverberation
                            %(Cremer1961)
l_m         = 4*V/S;        %mean free path
dN_dt_s     = 250;          %number of reflections per second, that are not
                            %distinguishable (getrennt wahrnehmbar)
                            %(zeitliche Diffusität)(Schmidt1973)
```

```
%starting time for stochastic reverberation in ms (reflections > 4000/s)
t_rubak_4000    = sqrt(dN_dt_r*V/(4*pi*c^3))*1000;

%starting time for stochastic reverberation in ms (mean free path)
t_rubak_mfw    = (4*l_m/c)*1000;

%starting time for stochastic reverberation in ms (reflections > 2000/s)
t_cremer_2000    = sqrt(dN_dt_c*V/(4*pi*c^3))*1000;

%transition time in ms for 500-Hz-octave band from regression equation
t_hidaka_RT   = 0.08*T*1000;

%starting time for stochastic reverberation in ms (reflections > 250/s)
t_schmidt_250    = sqrt(dN_dt_s*V/(4*pi*c^3))*1000;

%starting time for stochastic reverberation in ms (corresponds to
%reflections > 1000/s)
t_reichardt_sqrtV_1000  = sqrt(V);
```

## A.2 empirische Prädiktoren

### Abel und Stewart

```
function [t_abel, t_stewart, t_abelII, echo_dens, kurt, d, e] = stat(RIR,N,fs)
%berechnet den Zeitpunkt für den Einsatz des stochastischen Nachhalls
%basierend auf statistischen Verteilungen (Zeitbereich)
%
%t_abel     - mixing time nach Abel2006     (echo density = 1)
%t_stewart  - mixing time nach Stewart2007  (kurtosis = 0)
%t_abelII   - mixing time nach Abel2006     (echo density = 1-std_late)
%
%[t_abel,t_stewart,echo_dens, kurt]     = stat(RIR,N,fs1)
%RIR                     - monaurale Raumimpulsantwort (einkanalig)
%fs1                     - Abtastrate der wav-Datei
%N                       - Fensterlaenge
%
%dependencies:
%getOnsetPosition()
%plot_abel_stewart()
%
% (C) L. Kosanke, TU Berlin, audio communication group, 2009
%-------------------------------------------------------------------------%

len = length(RIR);

%Laufzeit entfernen
t_start = getOnsetPosition(RIR,-35,10);
```

```
RIR = RIR(t_start:length(RIR));

for n = 1:length(RIR)

    %Fenster am Anfang (steigende Fensterlaenge)
    if n <= N/2+1
        %Standardabweichung fuer Abel2006
        s(n)        = std(RIR(1:n+N/2-1));
        %Anzahl der tips außerhalb der Standardabweichung
        anz_s(n)    = sum(abs(RIR(1:n+N/2-1))>s(n));
        %Abel2006 echo density
        echo_dens(n)= anz_s(n)/N;
        %Stewart2007 kurtosis
        kurt(n)     = kurtosis(RIR(1:n+N/2-1))-3;
    end

    %Fenster in der Mitte (konstante Fensterlaenge)
    if n > N/2+1 & n <= length(RIR)-N/2+1
        s(n)        = std(RIR(n-N/2:n+N/2-1));
        anz_s(n)    = sum(abs(RIR(n-N/2:n+N/2-1))>s(n));
        echo_dens(n)= anz_s(n)/N;                           %Abel2006
        kurt(n)     = kurtosis(RIR(n-N/2:n+N/2-1))-3;       %Stewart2007
    end

    %Fenster am Ende (sinkende Fensterlaenge)
    if n > length(RIR)-N/2+1
        s(n)        = std(RIR(n-N/2:length(RIR)));
        anz_s(n)    = sum(abs(RIR(n-N/2:length(RIR)))>s(n));
        echo_dens(n)= anz_s(n)/N;
        kurt(n)     = kurtosis(RIR(n-N/2:length(RIR)))-3;
    end

end

%normieren
echo_dens       = echo_dens./erfc(1/sqrt(2));                   %Abel2006
echo_dens       = echo_dens';
kurt            = kurt./max(kurt);                              %Stewart2007
kurt            = kurt';

%Uebergangszeitpunkt Abel
d           = min(find(echo_dens>1));
t_abel      = d/fs*1000;

%Uebergangszeitpunkt Stewart
e           = min(find(kurt<0));
t_stewart   = e/fs*1000;

%Uebergangszeitpunkt Abel II
STD_late    = std(echo_dens(round(length(echo_dens)*0.8):length(echo_dens)-2000));
```

```
g           = min(find(echo_dens>(1-STD_late)));
t_abelII    = g/fs*1000;

figure
plot_abel_stewart(echo_dens, d, kurt, e, RIR, N, fs, t_abel, t_stewart)
```

## Hidaka

```
function [t_hidaka_all, r_all,t,sample_hidaka_all] = hidaka(RIR,fs,dsample)
% computes the transition time by correlating the energy of the direct sound
% and the energy of the sound at any time
%
% correlation between both should be r=1/e=0.367 for a stochastic process
% (e.g. diffuse sound field)
%
% [t_hidaka_all, r_all,t,sample_hidaka_all] = hidaka(RIR,fs,dsample)
%
% RIR     - room impulse response, one channel, shortened
% fs      - sample rate
% dsample - steps in samples
%
% t_hidaka_all contains the mixing time for octaves (125 - 16000 Hz) (this order):
% t_hidaka(1) - mixing time for 125 Hz-octave
% t_hidaka(2) - mixing time for 250 Hz-octave
% t_hidaka(3:8) - mixing time for the following octaves (500 Hz - 16 kHz)
%
%dependencies:
%getOnsetPosition.m
%plot_hidaka.m
%
% (C) L. Kosanke, TU Berlin, audio communication group, 2009
% ---------------------------------------------------------------------------%

t_start     = getOnsetPosition(RIR,-35,10);
RIR         = RIR(t_start:length(RIR));
len         = length(RIR);
N           = nextpow2(len);

fm          = [125 250 500 1000 2000 4000 8000 16000]';    %octaves

fu          = fm./sqrt(2);   %lower frequency boundary
fo          = 2.*fu;         %upper frequency boundary

%time-frequency-distribution of acoustic energy
energy1 = abs(fft([RIR(1:len);zeros(2^N-len,1)])).^2;

for m = 1:length(fm)
      for n = 1:floor(len/dsample)
        RIR_mod     = [RIR((n)*dsample+1:len);zeros(2^N-len+(n)*dsample,1)];
```

```
        energy      = (abs(fft(RIR_mod))).^2;
        %Pearson's correlation coefficient (correlation of energy of direct sound
        %and energy at any time)
        %rows: r for different times
        %columns: r for different frequencies

       %correlation for octaves:
       r_all(n,m) = corr(energy(ceil(length(energy).*fu(m)/fs):...
       ...ceil(length(energy).*fo(m)/fs)),...
         ...energy1(ceil(length(energy1).*fu(m)/fs):...
         ...ceil(length(energy1).*fo(m)/fs)));
      end
end

%convert samples of vectors to time
for k = 1:floor(len/dsample)
    samples(k)  = (k-1)*len/floor(len/dsample);
end

t               = samples./fs*1000;

for i = 1:size(r_all,2)
    sample_hidaka_all(i)    = min(find(r_all(:,i)<1/exp(1)));
    t_hidaka_all(i)         = samples(sample_hidaka_all(i))/fs*1000;
end

figure
plot_hidaka(fs,r_all,sample_hidaka_all,t_hidaka_all,dsample)
```

## Hauptskript

```
%Skript zur Berechnung der objektiven Maße zur Bestimmung des Zeitpunkts
%der vollständigen Diffusität einer RIR

% (C) L. Kosanke, TU Berlin, audio communication group, 2009
%-------------------------------------------------------------------------%

clear all
close all
clc

%_____________________________________________
%Geometrische Berechnung der mixing time
%Raummaße hier eintragen:
l   = 9.51;     %Laenge in m
b   = 7.06;     %Breite in m
h   = 3.15;     %Hoehe in m
T   = 0.62;     %Nachhallzeit in s
%V   = 3647;    %Raumvolumen
```

```
V   = l*b*h;                      %Raumvolumen
S   = 2*l*b + 2*b*h + 2*l*h;      %Raumoberflaeche

display('Geometrische Berechnung der mixing time')
[t_rubak_4000, t_rubak_mfw, t_cremer_2000, t_hidaka_RT,t_schmidt_250,...
...t_reichardt_sqrtV_1000] = geom(V,S,T)

%_____________________________________________
%zur statistischen Berechnung der mixing time
N       = 1024;  %Fensterlaenge (Potenzen von 2) in samples für Abel/Stewart
dsample = 100;   %Schrittweite in samples (für Hidaka)

display('statistische Berechnung der mixing time')

file       = input('Bezeichnung der wav-Datei (Impulsantwort) ','s');
[RIR,fs]   = wavread(file);
%RIR kann hier begrenzt werden:
RIR        = RIR(1:40000,:);

%Abfrage Kanalnummer
num_chan    = input('Welcher Kanal soll für die Berechnung benutzt werden? [1/2/...] ');
if size(RIR,2) < num_chan
     display('Kanal existiert nicht, Kanal 1 wird benutzt')
     num_chan = 1;
end
RIR         = RIR(:,num_chan);

%statistische Berechnung der mixing time
[t_abel, t_stewart, t_abelII, echo_dens, kurt, d, e]    = stat(RIR,N,fs);

%Korrelation der Energie
[t_hidaka_all, r_all,t,sample_hidaka_all]               = hidaka(RIR,fs,dsample);
```

# B Raumdaten

| Raum | Länge [m] | Breite [m] | Höhe [m] | Oberfläche [m$^2$] | Hallradius [m] |
|---|---|---|---|---|---|
| (1) EN324 | 9,51 | 7,60 | 3,15 | 252 | 1,34 |
| (2) EN111 | 7,06 | 7,06 | 4,49 | 225 | 1,08 |
| (3) EN190 | 7,10 | 8,55 | 3,16 | 220 | 0,87 |
| (4) H0104 | 25,65 | 20,90 | 6,15 | 1645 | 3,06 |
| (5) HE101 | 26,88 | 26,00 | 7,41 | 2181 | 3,17 |
| (6) Teldex | 28,31 | 16,20 | 8,28 | 1654 | 2,55 |
| (7) UdK | 37,70 | 17,40 | 12,65 | 2706 | 4,21 |
| (8) H0105 | 29,60 | 29,40 | 9,76 | 2892 | 3,64 |
| (9) JC-Kirche | | | | 2222[1] | 3,19 |

**Tabelle B.1:** Raumabmessungen, Oberflächen und Hallradien der untersuchten Räume

| Raum | 125 Hz | 250 Hz | 500 Hz | 1000 Hz | 2000 Hz | 4000 Hz |
|---|---|---|---|---|---|---|
| (1) EN324 | 0,5 s | 0,34 s | 0,39 s | 0,40 s | 0,37 s | 0,35 s |
| (2) EN111 | 0,72 s | 0,64 s | 0,63 s | 0,59 s | 0,56 s | 0,59 s |
| (3) EN190 | 0,87 s | 0,82 s | 0,84 s | 0,72 s | 0,74 s | 0,76 s |
| (4) H0104 | 1,29 s | 1,34 s | 1,09 s | 1,13 s | 1,04 s | 1,00 s |
| (5) HE101 | 1,55 s | 1,51 s | 1,45 s | 1,75 s | 1,95 s | 1,81 s |
| (6) Teldex | 1,29 s | 1,72 s | 1,95 s | 2,11 s | 2,07 s | 1,81 s |
| (7) UdK | 1,92 s | 1,70 s | 1,62 s | 1,55 s | 1,38 s | 0,96 s |
| (8) H0105 | 2,48 s | 2,34 s | 2,01 s | 2,06 s | 1,93 s | 1,67 s |
| (9) JC-Kirche | 1,96 s | 2,13 s | 2,66 s | 3,05 s | 2,55 s | 1,83 s |

**Tabelle B.2:** Nachhallzeiten der untersuchten Räume

[1] aus Teilvolumina bestimmt

# C Statistische Daten

## C.1 Deskriptive und inferenzstatistische Auswertung

**Descriptive Statistics**

| | Mean | Std. Deviation | N |
|---|---|---|---|
| Raum1 | ,03610980 | ,017199215 | 10 |
| Raum2 | ,02919200 | ,012099382 | 10 |
| Raum3 | ,02418737 | ,015431489 | 10 |
| Raum4 | ,05447240 | ,020773671 | 10 |
| Raum5 | ,04624160 | ,016818262 | 10 |
| Raum6 | ,06353690 | ,026314549 | 10 |
| Raum7 | ,09077400 | ,057151056 | 10 |
| Raum8 | ,06330520 | ,040324692 | 10 |
| Raum9 | ,07329770 | ,037708304 | 10 |

**Tabelle C.1:** Deskriptive Auswertung

**Mauchly's Test of Sphericity**

Measure: MEASURE_1

| | | | | | Epsilon | | |
|---|---|---|---|---|---|---|---|
| Within Subjects Effect | Mauchly's W | Approx. Chi-Square | df | Sig. | Greenhouse-Geisser | Huynh-Feldt | Lower-bound |
| volume | ,617 | 3,867 | 2 | ,145 | ,723 | ,825 | ,500 |
| absorb | ,909 | ,762 | 2 | ,683 | ,917 | 1,000 | ,500 |
| volume * absorb | ,380 | 7,171 | 9 | ,629 | ,674 | ,991 | ,250 |

**Tabelle C.2:** Test von Mauchly

**Tests of Within-Subjects Effects**

Measure: MEASURE_1

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power(a) |
|---|---|---|---|---|---|---|---|---|---|
| volume | Sphericity Assumed | ,032 | 2 | ,016 | 12,174 | ,000 | ,575 | 24,348 | ,987 |
| | Greenhouse-Geisser | ,032 | 1,446 | ,022 | 12,174 | ,002 | ,575 | 17,601 | ,952 |
| | Huynh-Feldt | ,032 | 1,649 | ,019 | 12,174 | ,001 | ,575 | 20,077 | ,970 |
| | Lower-bound | ,032 | 1,000 | ,032 | 12,174 | ,007 | ,575 | 12,174 | ,873 |
| Error(volume) | Sphericity Assumed | ,023 | 18 | ,001 | | | | | |
| | Greenhouse-Geisser | ,023 | 13,012 | ,002 | | | | | |
| | Huynh-Feldt | ,023 | 14,843 | ,002 | | | | | |
| | Lower-bound | ,023 | 9,000 | ,003 | | | | | |
| absorb | Sphericity Assumed | ,003 | 2 | ,002 | 1,905 | ,178 | ,175 | 3,810 | ,343 |
| | Greenhouse-Geisser | ,003 | 1,834 | ,002 | 1,905 | ,182 | ,175 | 3,493 | ,326 |
| | Huynh-Feldt | ,003 | 2,000 | ,002 | 1,905 | ,178 | ,175 | 3,810 | ,343 |
| | Lower-bound | ,003 | 1,000 | ,003 | 1,905 | ,201 | ,175 | 1,905 | ,235 |
| Error(absorb) | Sphericity Assumed | ,014 | 18 | ,001 | | | | | |
| | Greenhouse-Geisser | ,014 | 16,502 | ,001 | | | | | |
| | Huynh-Feldt | ,014 | 18,000 | ,001 | | | | | |
| | Lower-bound | ,014 | 9,000 | ,002 | | | | | |
| volume * absorb | Sphericity Assumed | ,003 | 4 | ,001 | 1,480 | ,229 | ,141 | 5,921 | ,412 |
| | Greenhouse-Geisser | ,003 | 2,698 | ,001 | 1,480 | ,246 | ,141 | 3,993 | ,325 |
| | Huynh-Feldt | ,003 | 3,963 | ,001 | 1,480 | ,229 | ,141 | 5,865 | ,409 |
| | Lower-bound | ,003 | 1,000 | ,003 | 1,480 | ,255 | ,141 | 1,480 | ,193 |
| Error(volume*absorb) | Sphericity Assumed | ,019 | 36 | ,001 | | | | | |
| | Greenhouse-Geisser | ,019 | 24,278 | ,001 | | | | | |
| | Huynh-Feldt | ,019 | 35,665 | ,001 | | | | | |
| | Lower-bound | ,019 | 9,000 | ,002 | | | | | |

a Computed using alpha = ,05

**Tabelle C.3:** Zweifaktorielle ANOVA

# C.2 Regressionsanalyse

## C.2.1 Regressionsmodelle

### Mittelwertprädiktion

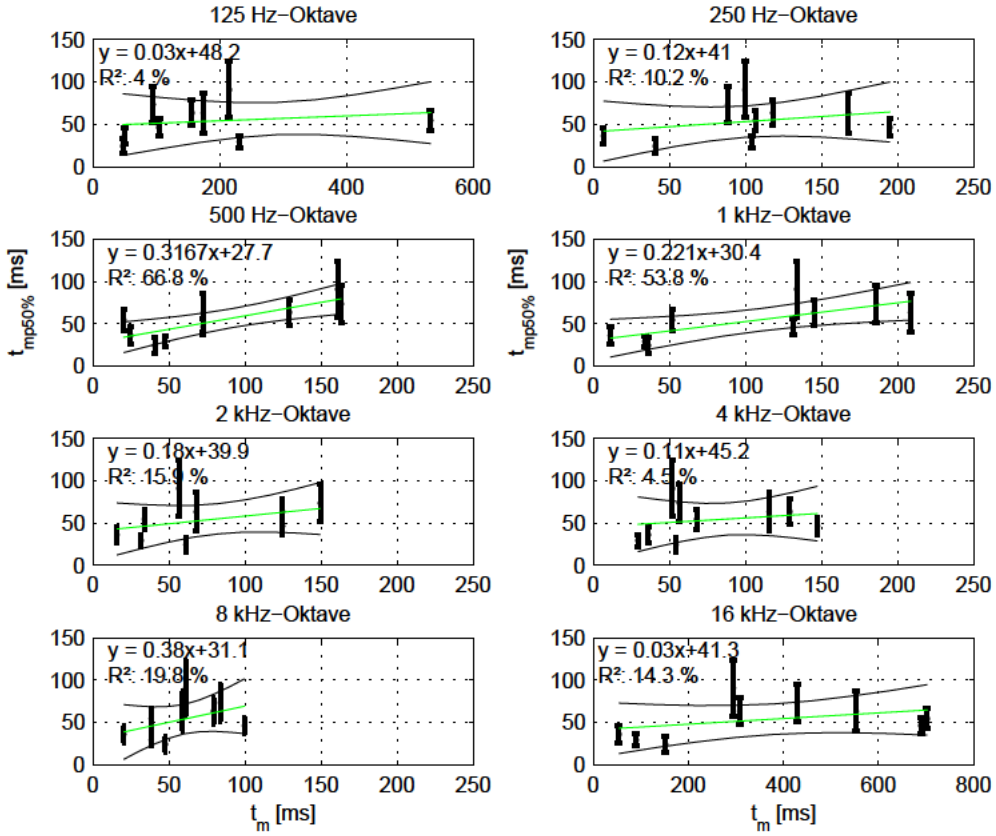



**Abbildung C.1:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times der linken BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

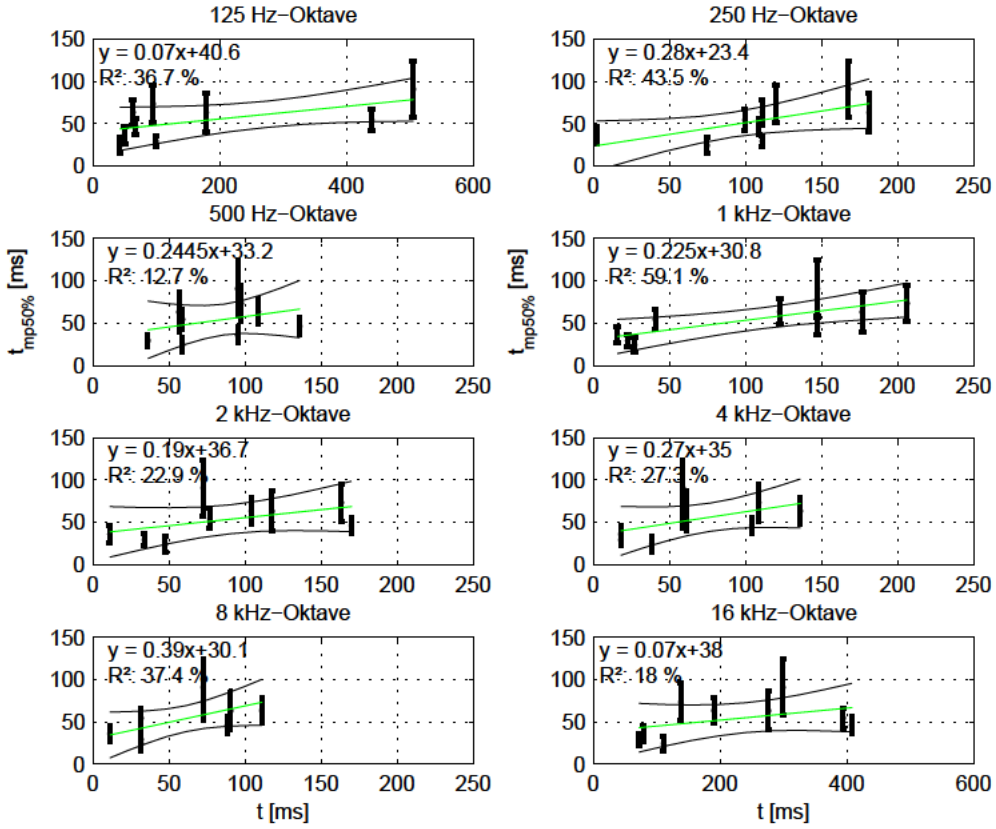



**Abbildung C.2:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times der rechten BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

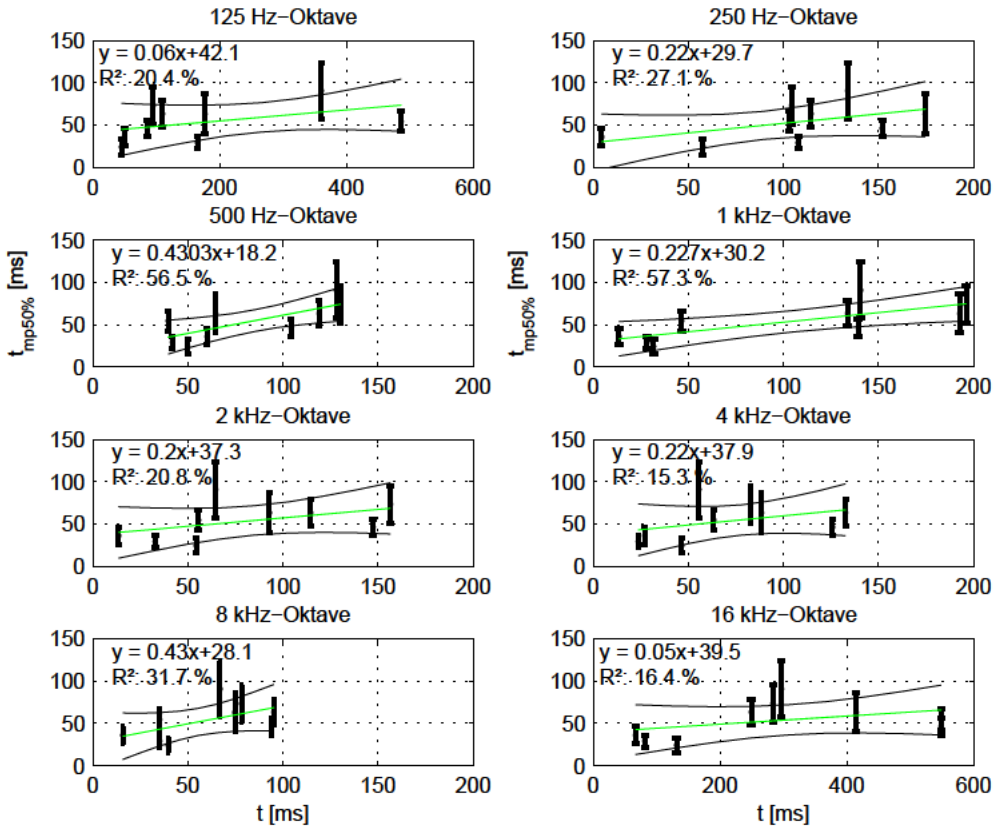



**Abbildung C.3:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den Mittelwerten der mixing times der linken und rechten BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

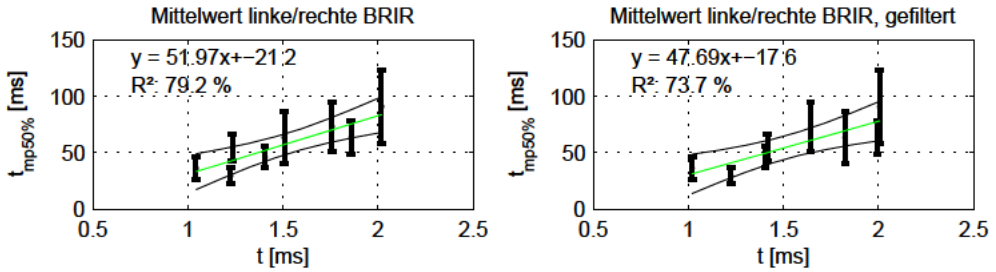
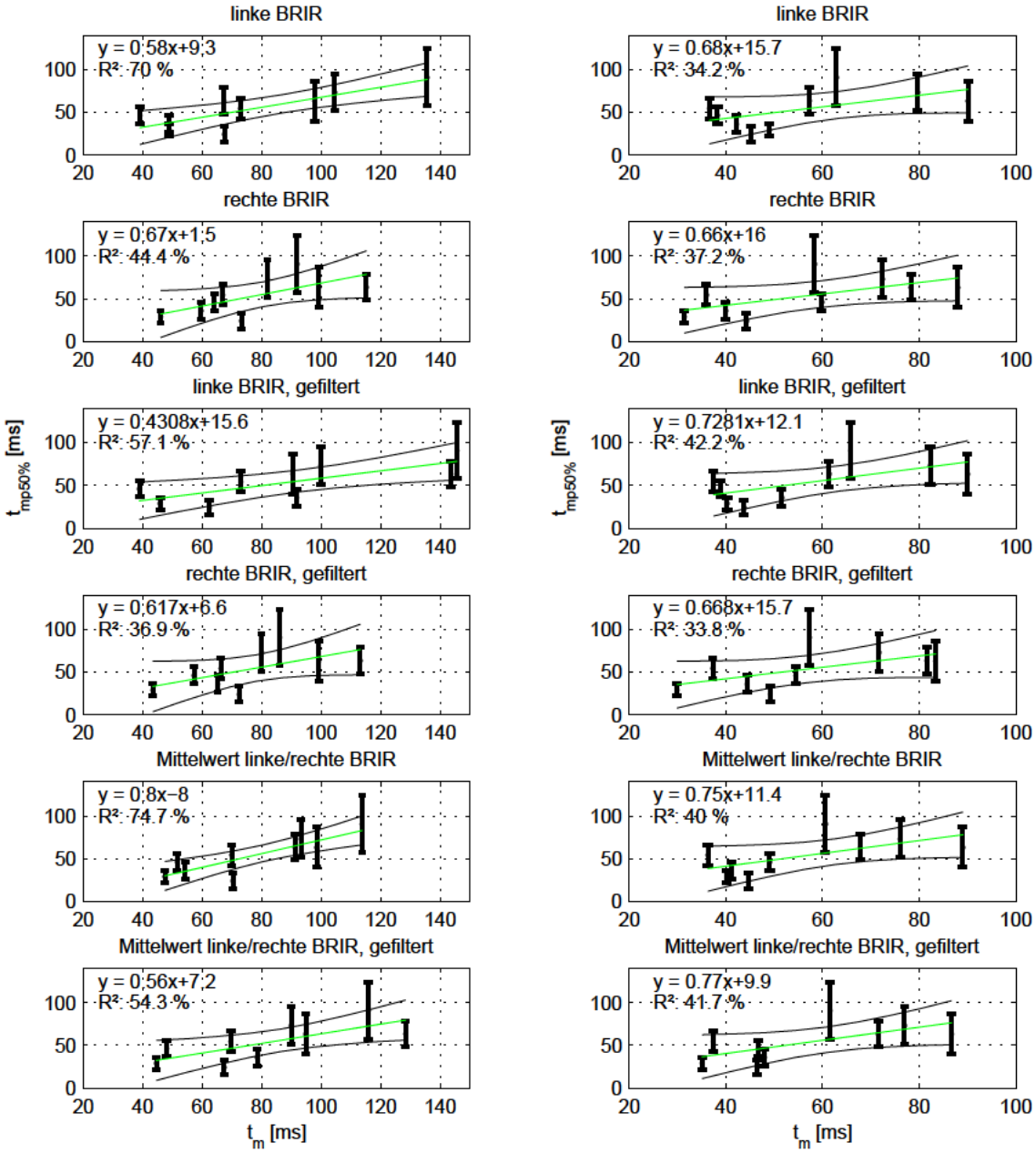


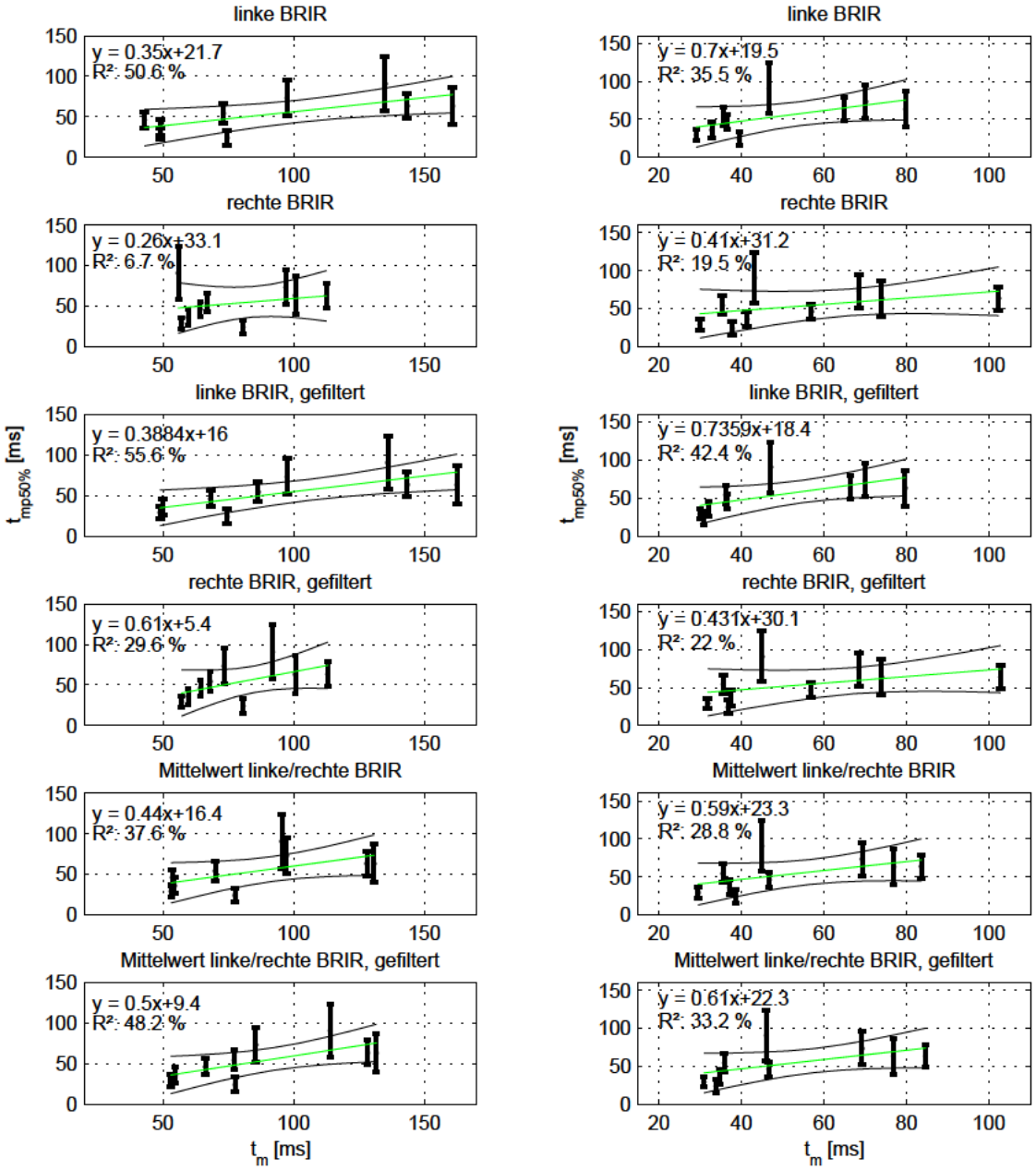

**Abbildung C.4:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Defrance u. a. (2009)




**(a)** Abel und Huang (2006) automatisch

**(b)** Abel und Huang (2006) visuell

**Abbildung C.5:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Abel und Huang (2006)




**(a)** Stewart und Sandler (2007) automatisch

**(b)** Stewart und Sandler (2007) visuell

**Abbildung C.6:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. Mittelwerte der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Stewart und Sandler (2007)

## 95 %-Prädiktion

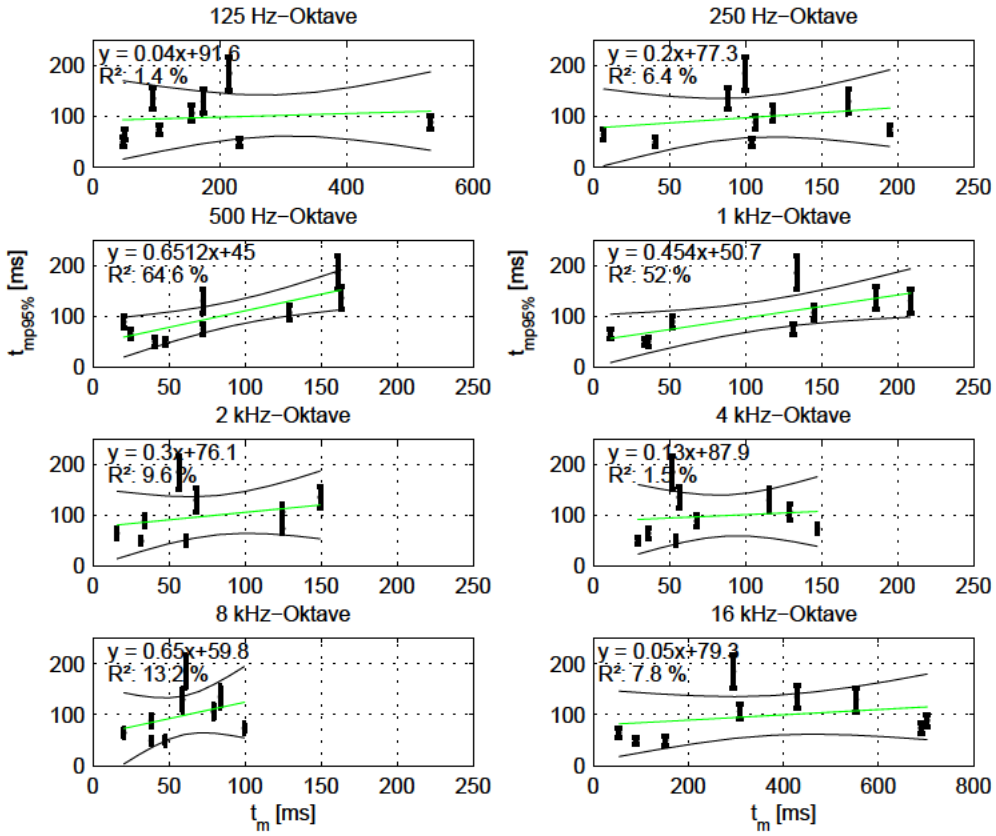



**Abbildung C.7:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times der rechten BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

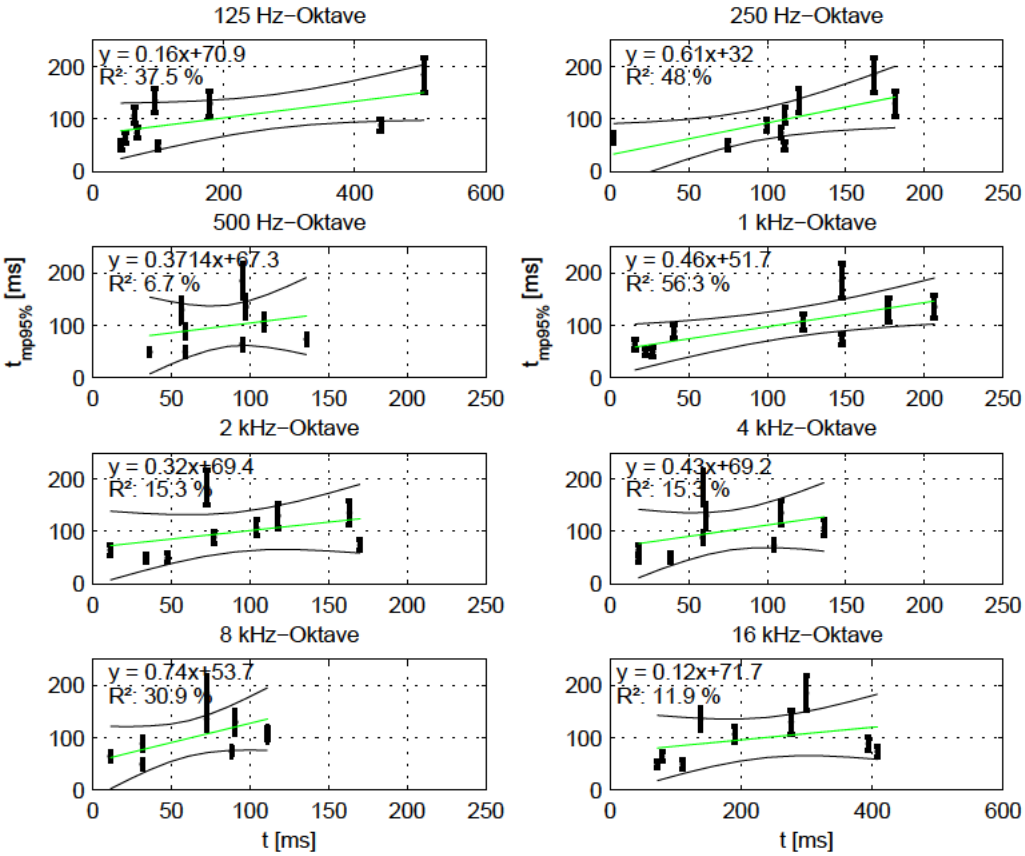



**Abbildung C.8:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times der rechten BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

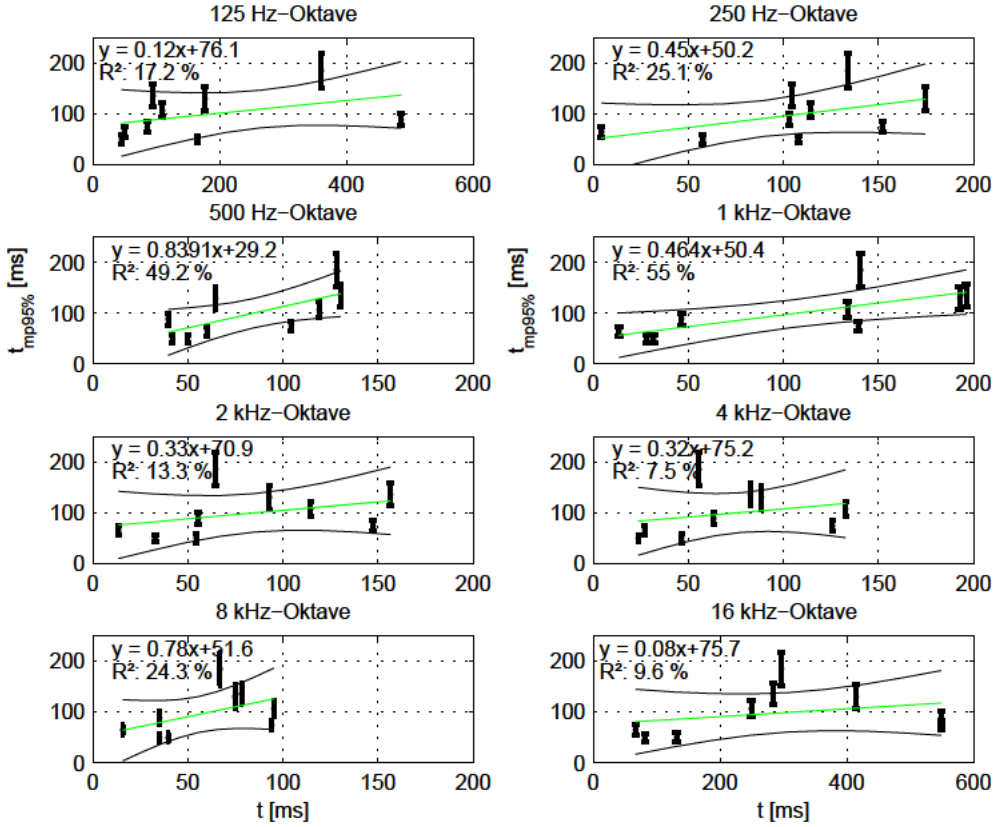



**Abbildung C.9:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den Mittelwerten der mixing times der linken und rechten BRIRs berechnet durch den Prädiktor von Hidaka u. a. (2007)

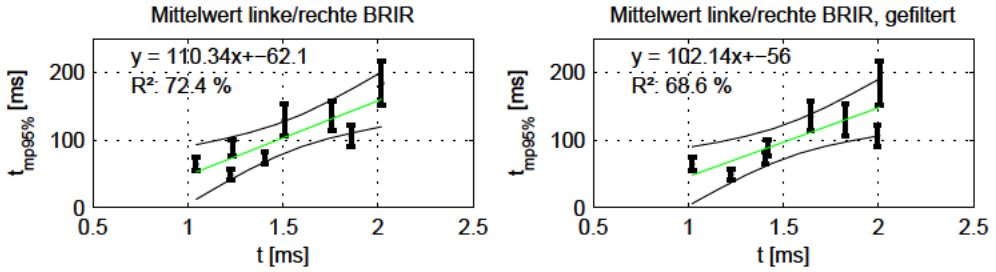
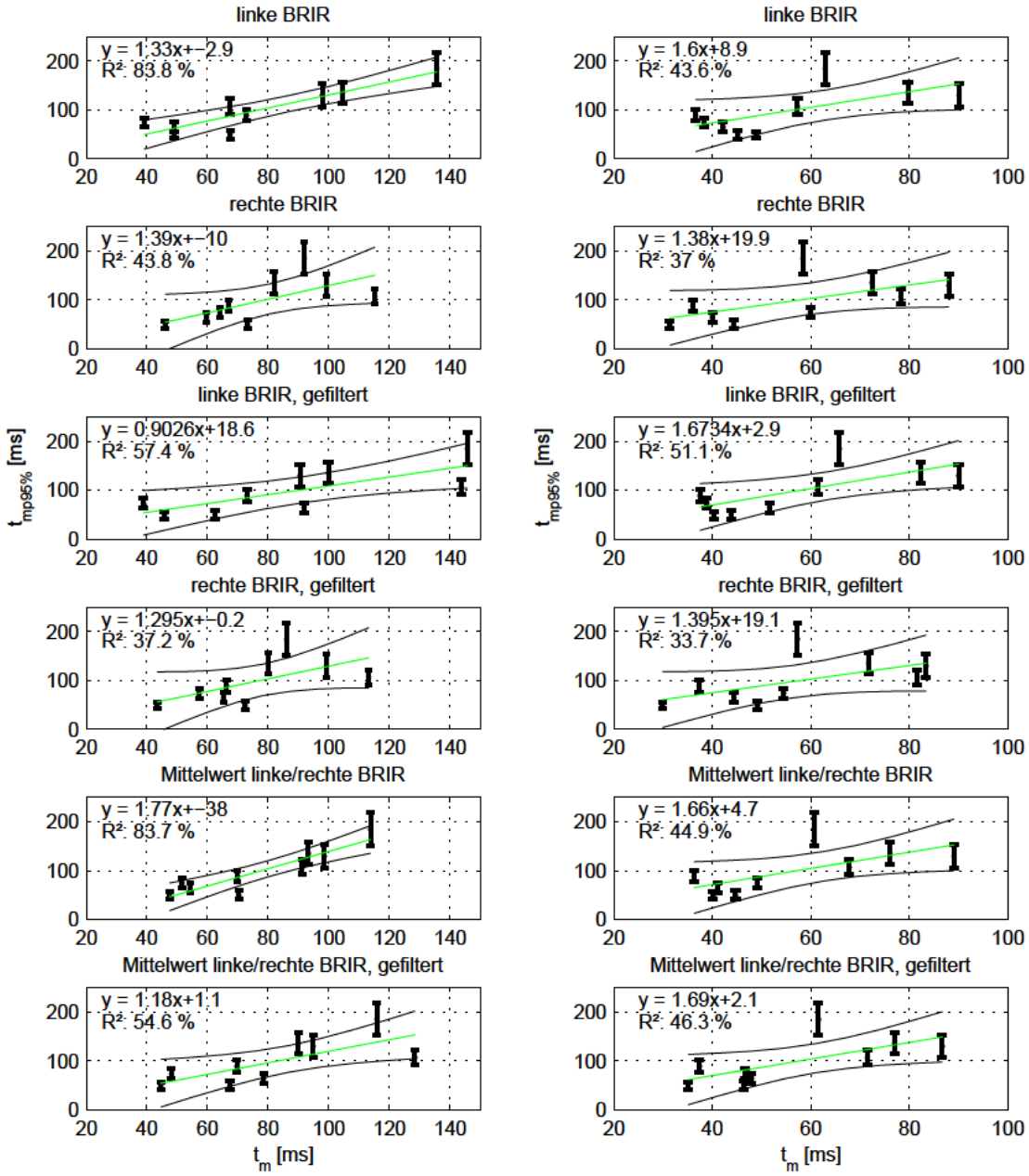


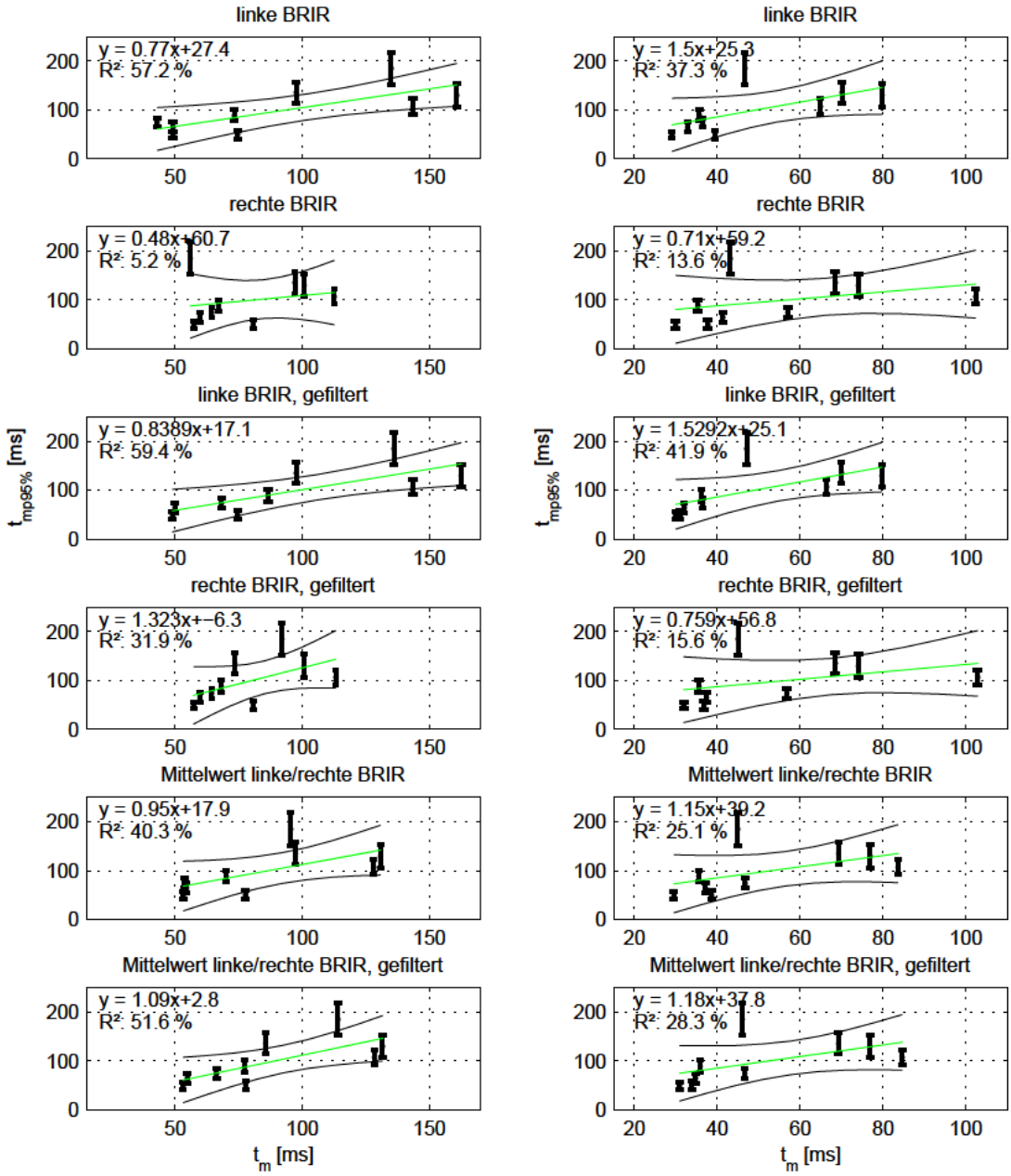

**Abbildung C.10:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Defrance u. a. (2009)




**(a)** Abel und Huang (2006) automatisch

**(b)** Abel und Huang (2006) visuell

**Abbildung C.11:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Abel und Huang (2006)




**(a)** Stewart und Sandler (2007) automatisch

**(b)** Stewart und Sandler (2007) visuell

**Abbildung C.12:** Lineare Regressionen und dazugehörige 95 %-Konfidenzintervalle. 95 %-Perzentile der perzeptiven mixing times und 95 %-Konfidenzintervalle dargestellt über den mixing times berechnet durch den Prädiktor von Stewart und Sandler (2007)

## Ranking der Ergebnisse der Mittelwertprädiktion

| # | Prädiktor | BRIR | Oktavband / HP-gefiltert | automatisch / visuell | $R^2$ |
|---|---|---|---|---|---|
| 1 | $\frac{V}{S}$ | | | | 81,5 % |
| 2 | Defrance | Mittelwert | | automatisch | 79,2 % |
| 3 | $\sqrt{V}$ | | | | 78,6 % |
| 4 | $V$ | | | | 77,4 % |
| 5 | Abel I | Mittelwert | | automatisch | 74,7 % |
| 6 | Defrance | Mittelwert | gefiltert | automatisch | 73,7 % |
| 7 | $S$ | | | | 73,3 % |
| 8 | Abel I | links | | automatisch | 70,0 % |
| 9 | Hidaka | links | 500 Hz | | 66,8 % |
| 10 | Hidaka | rechts | 1 kHz | | 59,1 % |
| 11 | Hidaka | Mittelwert | 1 kHz | | 57,3 % |
| 12 | Abel I | links | gefiltert | automatisch | 57,1 % |
| 13 | Hidaka | Mittelwert | 500 Hz | | 56,5 % |
| 14 | Stewart | links | gefiltert | automatisch | 55,6 % |
| 15 | Abel I | Mittelwert | gefiltert | automatisch | 54,3 % |
| 16 | Hidaka | links | 1 kHz | | 53,8 % |
| 17 | $T_{30_{125Hz-4kHz}}$ | | | | 53,4 % |
| 18 | Abel II | Mittelwert | | automatisch | 50,7 % |
| 19 | Stewart | links | | automatisch | 50,6 % |
| 20 | Stewart | Mittelwert | gefiltert | automatisch | 48,2 % |
| 21 | Abel I | rechts | | automatisch | 44,4 % |
| 22 | Hidaka | rechts | 250 Hz | | 43,5 % |
| 23 | Stewart | links | gefiltert | visuell | 42,4 % |
| 24 | Abel I | links | gefiltert | visuell | 42,2 % |
| 25 | Abel I | Mittelwert | gefiltert | visuell | 41,7 % |
| 26 | Abel I | Mittelwert | | visuell | 40,0 % |
| 27 | Defrance | Mittelwert | | automatisch | 39,4 % |
| 28 | Stewart | Mittelwert | | automatisch | 37,6 % |
| 29 | Hidaka | rechts | 4 kHz | | 37,4 % |

*Fortsetzung auf der nächsten Seite*

| # | Prädiktor | BRIR | Oktavband / HP-gefiltert | automatisch / visuell | $R^2$ |
|---|---|---|---|---|---|
| 30 | Abel I | rechts | | visuell | 37,2 % |
| 31 | Abel I | rechts | gefiltert | automatisch | 36,9 % |
| 32 | Hidaka | rechts | 125 Hz | | 36,7 % |
| 33 | Stewart | links | | visuell | 35,5 % |
| 34 | Abel I | links | | visuell | 34,2 % |
| 35 | Abel I | rechts | gefiltert | visuell | 33,8 % |
| 36 | Stewart | Mittelwert | gefiltert | visuell | 33,2 % |
| 37 | Hidaka | Mittelwert | 8 kHz | | 31,7 % |
| 38 | Stewart | rechts | gefiltert | automatisch | 29,6 % |
| 39 | Stewart | Mittelwert | | visuell | 28,8 % |
| 40 | Hidaka | rechts | 4 kHz | | 27,3 % |
| 41 | Hidaka | Mittelwert | 250 Hz | | 27,1 % |
| 42 | Hidaka | rechts | 2 kHz | | 22,9 % |
| 43 | Stewart | rechts | gefiltert | visuell | 22,0 % |
| 44 | Hidaka | Mittelwert | 2 kHz | | 20,8 % |
| 45 | Hidaka | Mittelwert | 125 Hz | | 20,4 % |
| 46 | Hidaka | links | 8 kHz | | 19,8 % |
| 47 | Stewart | rechts | | visuell | 19,5 % |
| 48 | Hidaka | rechts | 16 kHz | | 18,0 % |
| 49 | Hidaka | Mittelwert | 16 kHz | | 16,4 % |
| 50 | Hidaka | links | 2 kHz | | 15,9 % |
| 51 | Hidaka | Mittelwert | 4 kHz | | 15,3 % |
| 52 | Hidaka | links | 16 kHz | | 14,3 % |
| 53 | Hidaka | rechts | 500 Hz | | 12,7 % |
| 54 | Hidaka | links | 250 Hz | | 10,2 % |
| 55 | Stewart | rechts | | automatisch | 6,7 % |
| 56 | Hidaka | links | 4 kHz | | 4,5 % |
| 57 | $\alpha_{mittel}$ | | | | 4,3 % |
| 58 | Hidaka | links | 125 Hz | | 4,0 % |

**Tabelle C.4:** Ranking der Varianzaufklärung der Mittelwertprädiktion

## Ranking der Ergebnisse der 95 %-Prädiktion

| # | Prädiktor | BRIR | Oktavband / HP-gefiltert | automatisch / visuell | $R^2$ |
|---|---|---|---|---|---|
| 1 | Abel I | links | | automatisch | 83,8 % |
| 2 | Abel I | Mittelwert | | automatisch | 83,7 % |
| 3 | $V$ | | | | 78,7 % |
| 4 | $\frac{V}{S}$ | | | | 75,7 % |
| 5 | $\sqrt{V}$ | | | | 73,4 % |
| 6 | Defrance | Mittelwert | | automatisch | 72,4 % |
| 7 | Defrance | Mittelwert | gefiltert | automatisch | 68,6 % |
| 8 | Abel II | Mittelwert | | automatisch | 66,7 % |
| 9 | Hidaka | rechts | 1 kHz | | 65,3 % |
| 10 | Hidaka | links | 500 Hz | | 64,6 % |
| 11 | Stewart | links | gefiltert | automatisch | 59,4 % |
| 12 | Abel I | links | gefiltert | automatisch | 57,3 % |
| 13 | Stewart | links | | automatisch | 57,2 % |
| 14 | Hidaka | rechts | 1 kHz | | 56,3 % |
| 15 | Hidaka | Mittelwert | 1 kHz | | 55 % |
| 16 | Abel I | Mittelwert | gefiltert | automatisch | 54,8 % |
| 17 | Hidaka | links | 1 kHz | | 52 % |
| 18 | Stewart | Mittelwert | gefiltert | automatisch | 51,6 % |
| 19 | Abel I | links | gefiltert | visuell | 51,1 % |
| 20 | Hidaka | Mittelwert | 500 Hz | | 49,2 % |
| 21 | Hidaka | rechts | 250 Hz | | 48 % |
| 22 | $T_{30_{125Hz-4kHz}}$ | | | | 46,5 % |
| 23 | Abel I | Mittelwert | | visuell | 44,9 % |
| 24 | Abel I | links | | visuell | 43,6 % |
| 25 | Abel I | rechts | | automatisch | 43,8 % |
| 26 | Stewart | links | gefiltert | visuell | 41,9 % |
| 27 | Stewart | Mittelwert | | automatisch | 40,3 % |
| 28 | Stewart | links | | visuell | 37,3 % |
| 29 | Defrance | Mittelwert | | visuell | 37,3 % |

*Fortsetzung auf der nächsten Seite*

| # | Prädiktor | BRIR | Oktavband / HP-gefiltert | automatisch / visuell | $R^2$ |
|---|---|---|---|---|---|
| 30 | Abel I | rechts | gefiltert | automatisch | 37,2 % |
| 31 | Abel I | rechts | | visuell | 37 % |
| 32 | Hidaka | rechts | 125 Hz | | 35,7 % |
| 33 | Abel I | rechts | gefiltert | visuell | 33,7 % |
| 34 | Stewart | rechts | gefiltert | automatisch | 31,9 % |
| 35 | Hidaka | rechts | 8 kHz | | 30,9 % |
| 36 | Stewart | Mittelwert | gefiltert | visuell | 28,3 % |
| 37 | Stewart | Mittelwert | | visuell | 25,1 % |
| 38 | Hidaka | Mittelwert | 250 Hz | | 25,1 % |
| 39 | Hidaka | Mittelwert | 8 kHz | | 24,3 % |
| 40 | Hidaka | Mittelwert | 125 kHz | | 17,2 % |
| 41 | Abel | Mittelwert | gefiltert | visuell | 16,3 % |
| 42 | Stewart | rechts | gefiltert | visuell | 15,6 % |
| 43 | Hidaka | rechts | 2 kHz | | 15,3 % |
| 44 | Hidaka | rechts | 4 kHz | | 15,3 % |
| 45 | Stewart | rechts | | automatisch | 13,6 % |
| 46 | Hidaka | Mittelwert | 2 kHz | | 13,3 % |
| 47 | Hidaka | links | 8 kHz | | 13,2 % |
| 48 | Hidaka | rechts | 16 kHz | | 11,9 % |
| 49 | Hidaka | links | 2 kHz | | 9,6 % |
| 50 | Hidaka | Mittelwert | 16 kHz | | 9,6 % |
| 51 | Hidaka | links | 16 kHz | | 7,8 % |
| 52 | Hidaka | Mittelwert | 4 kHz | | 7,5 % |
| 53 | Hidaka | rechts | 500 Hz | | 6,7 % |
| 54 | Hidaka | links | 250 Hz | | 6,4 % |
| 55 | Stewart | rechts | | automatisch | 5,2 % |
| 56 | $\alpha_{mittel}$ | | | | 4,8 % |
| 57 | Hidaka | links | 4 kHz | | 1,5 % |
| 58 | Hidaka | links | 125 Hz | | 1,4 % |

**Tabelle C.5:** Ranking der Varianzaufklärung der 95 %-Prädiktion

### C.2.2 Signifikanz der Anstiege der Regressionsgeraden

| Prädiktor | $\frac{V}{S}$ | $\sqrt{V}$ | $V$ | $S$ | $T_{30_{125Hz-4kHz}}$ | $\alpha_{mittel}$ |
|---|---|---|---|---|---|---|
| 50 %-Prädiktion | 0,000859 | 0,0015 | 0,0018 | 0,0032 | 0,0253 | 0,5939 |
| 95 %-Prädiktion | 0,023 | 0,0032 | 0,0014 | 0,0052 | 0,0431 | 0,5712 |

**Tabelle C.6:** Signifikanz der Anstiege der Regressionsgeraden (modellbasierte Prädiktoren)

| Prädiktor | Abel I | Abel II | Hidaka (1 kHz) | Hidaka (500 Hz) | Stewart | Defrance |
|---|---|---|---|---|---|---|
| 50 %-Prädiktion | 0,0026 | 0,0313 | 0,0182 | 0,0195 | 0,079 | 0,0031 |
| 95 %-Prädiktion | 0,000543 | 0,0353 | 0,0222 | 0,0353 | 0,0661 | 0,0074 |

**Tabelle C.7:** Signifikanz der Anstiege der Regressionsgeraden (impulsantwortbasierte Prädiktoren)

# D Fragebogen und Versuchsanleitung

**Fragebogen zum Hörversuch**

**„Wahrnehmbarkeit der Diffusfeldgrenze (mixing time) in BRIRs“**

**Geschlecht**

☐ weiblich

☐ männlich

**Alter**

_____

**Musikalische Ausbildung**

☐ ja, _____ (_____ Jahre)

☐ nein

**Hörversuchserfahrung**

☐ ja

☐ nein

**Woran hast du den Unterschied zwischen Referenz und manipuliertem Signal erkannt?**

VP ___

Liebe Versuchsteilnehmer!

In diesem Hörversuch geht es um die Bestimmung des subjektiven Übergangszeitpunkts zwischen frühen Reflexionen und diffusem Nachhall in binauralen Raumimpulsantworten.
Dazu wird ein Schlagzeugsample in neun verschiedenen akustischen Umgebungen präsentiert.

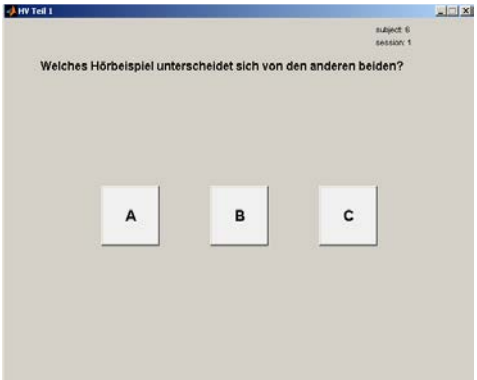
Das Hörbeispiel wird je 2x ohne Veränderung und 1x als manipulierte Variante dargeboten. Bei der manipulierten Variante wird der Nachhall entsprechend einer festen, frontalen Kopfausrichung generiert.
Die Simulation reagiert auf horizontale Kopfbewegungen (siehe Grafik).
Es ist also am einfachsten, einen Unterschied zwischen Original und manipuliertem Schlagzeugsample zu hören, wenn der Kopf möglichst weit nach links oder rechts gedreht wird.

***Training:***

Zu Beginn wird der Versuch für einen Beispielraum durchgeführt. Der Versuchsleiter gibt eine Rückmeldung, **welches der 3 Hörbeispiele sich von den anderen beiden unterschieden hat**.
Ziel des Trainings ist es, eine feste Kopfposition oder auch eine Rotationsbewegung zu finden, bei der du die Unterschiede möglichst leicht wahrnehmen kannst.

***Versuch:***

Es werden immer 3 Hörbeispiele nacheinander dargeboten, von denen nur eines manipuliert ist. Die beiden anderen sind Hörbeispiele, die aus der Originalimpulsantwort simuliert wurden.




Nachdem die Schaltfläche mit der Antwort angeklickt wurde, **folgt sofort das nächste Hörbeispiel**.
Dem Hörversuch liegt eine adaptive Prozedur zugrunde, die auf die Antwortleistung reagiert.
Nach jedem Raum gibt es eine kurze Pause.