Einfluss der Nachhallzeit von tiefen Frequenzen auf die Sprachqualität

Diplomarbeit

durchgeführt von

Philipp Reents

am

Institut für Breitbandkommunikation der Technischen Universität Graz

Leiter: Univ.-Prof. Dipl.-Ing. Dr. techn. Gernot Kubin

in Kooperation mit

Müller BBM GmbH, Planegg

Betreuer: Dr.-Ing. Eckard Mommertz, Müller-BBM GmbH, Planegg

Begutachter: Ao. Univ.-Prof. Dr. Gerhard Graber

Graz, im April 2007

Zusammenfassung

Zur Verbesserung der Kommunikation in Klassen- und Seminarräumen ist ein besonders hohes Maß an Sprachverständlichkeit erstrebenswert. Einen wesentlichen Einfluss darauf erhält man über das Anpassen der Nachhallzeiten in den einzelnen Frequenzbändern. Während unter Fachleuten Einigkeit herrscht, bezüglich optimaler Nachhallzeiten für den mittleren bis hohen Frequenzbereich, werden für tiefe Frequenzen kontroverse Meinungen diskutiert. Im Rahmen dieser Arbeit wird der Einfluss der Nachhallzeit von Frequenzen unterhalb von 300 Hz auf die Sprachverständlichkeit untersucht. Zu diesem Zweck werden geeignete Impulsantworten durch Messung und Simulation erzeugt und auf Sprachsignale angewandt. Über Hörversuche in Form von Reimtests und Paarvergleichen werden die Signale beurteilt und bilden die Grundlage für Analysen hinsichtlich der Sprachqualität.

Abstract

Especially in classrooms, a high degree of speech intelligibility is of capital interest. Therefore, optimising the reverberation time in different frequency bands is a suitable method. Considering the reverberation time of lower frequency bands, different opinions lead to an ongoing discussion about an optimum time region. Goal of this study is to examine the influence of reverberation times of frequencies below 300 Hz on speech intelligibility. Therefore speech material is prepared and tested via comparison in pairs and a speech intelligibility test is being performed. The results are analyzed regarding the speech quality.

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen	2
2.1 Audiologie	2
2.1.1 Sprachspektrum	2
2.1.2 Hörspektrum	4
2.1.3 Verdeckungseffekte	5
2.2 Raumakustik	7
2.2.1 Wellentheorie	8
2.2.2 Geometrische Raumakustik	
2.2.3 Statistische Raumakustik	15
2.2.4 Messmethoden der Sprachverständlichkeit	18
3. Generierung des Audiomaterials	24
3.1 Simulation	24
3.1.1 Binaurale Impulsantworten	
3.1.2 Simulierte Raummoden	27
3.2 Messungen	
3.2.1 Messraum	
3.2.2 Planung und Bau von Plattenschwingern	
3.2.3 Messaufbau und Messgeräte	
3.2.4 Messung der Impulsantworten	
3.2.5 Erzeugung des Audiomaterials für den Hörtest	
4. Darbietung und Auswertung von Hörmaterial	
4.1 Reimtest	
4.1.1 Auswahl des Testverfahrens	
4.1.2 Testablauf	
4.1.3 Auswertung mittels Regressionsanalyse	
4.2 Paarvergleiche	40
4.2.1 Versuchsablauf	
4.2.2 Auswertung	
4.2.2.1 BTL-Modell	
4.2.2.2 Urteilskonsistenz	44

5. Ergebnisse des Reimtests	
5.1 Ergebnisse der Testreihen A und B	
5.2 Vergleich von Spektren der Testwörter	
5.3 STI-Werte des Reimtests	
6. Ergebnisse des Paarvergleichs	
6.1 Ergebnisse mit simulierten Impulsantworten	53
6.2 Ergebnisse mit simulierten Raummoden	
6.3 Ergebnisse mit realen Impulsantworten	60
7. Zusammenfassung und Ausblick	
7.1 Hörversuche mit simuliertem Klassenraum	
7.2 Hörversuche mit realen Raumaufzeichnungen	63
7.3 Folgerungen	
8. Literaturverzeichnis	

9. Anhang

1. Einleitung

Bei der Planung von Seminar- und Klassenräumen wird zunehmend berücksichtigt, ein möglichst hohes Maß an Sprachverständlichkeit in den Räumen zu erzielen. Wie durch zahlreiche Studien belegt ist, beeinflussen die raumakustischen Verhältnisse maßgeblich das Aufnahme- und Konzentrationsvermögen der anwesenden Personen, siehe z.B. Bradley[1], Sutherland, Lubman[2], oder Schick, Klatte, Meis [3]. Zur Bestimmung der Sprachverständlichkeit dienen Hörtests, sowie messtechnische Verfahren, die über eine Messung der Nachhallzeit, eine Aussage zur Sprachverständlichkeit ermöglichen. Die Kenntnisse über den Einfluss der Nachhallzeit eines Raumes auf die Sprachverständlichkeit dienen auch als Grundlage für Regelwerke wie die DIN 18041 [4], in der Vorgaben für die einzuhaltenden Nachhallzeiten zu finden sind, siehe Abbildung 1.1.



Abbildung 1.1: Toleranzbereich der empfohlenen Nachhallzeit T, normiert auf die vom Volumen abhängige Nachhallzeit T_{soll} in Abhängigkeit von der Frequenz für Sprache, aus Ruhe [5].

Während unter Fachleuten Einigkeit bezüglich der optimalen Nachhallzeit im mittleren, bis hohen Frequenzbereich herrscht, gibt es kontroverse Meinungen zu dem Bereich unterhalb von 300 Hz. Aus diesem Grund, wurde bei der Firma Müller-BBM, in Kooperation mit der TU Graz eine Arbeit durchgeführt, die den Einfluss der Nachhallzeit tiefer Frequenzen auf die Sprachverständlichkeit untersucht. Zu diesem Zweck wurden Sprachsignale erzeugt, deren Nachhallzeiten sich nur im tieffrequenten Bereich voneinander unterscheiden. Mittels Paarvergleichen und einem Sprachverständlichkeitstest wurden Daten erhoben, die als Grundlage für weitere Diskussionen dienen sollen.

2. Grundlagen

Im folgenden Kapitel werden jene Grundlagen erläutert, die zum Verständnis der Arbeit beitragen. In Abschnitt 2.1, Audiologie, werden die Frequenzbereiche dargelegt, die von der Stimme erzeugt, bzw. vom Ohr wahrgenommen werden können, und weiters wird aus dem Gebiet der Psychoakustik der Verdeckungseffekt vorgestellt. Abschnitt 2.2 behandelt die Grundlagen der Raumakustik und gibt einen Überblick über die verwendeten Messverfahren zur Klassifizierung von Räumen.

2.1 Audiologie

2.1.1 Sprachspektrum

Der menschliche Stimmapparat ist in der Lage, Frequenzen in einem Bereich zwischen 60 Hz und 8 kHz zu erzeugen. Für Sprache wird dabei nahezu der gesamte Frequenzbereich benutzt, der nach Fasold und Veres [6] in einen Bereich für Stimmklang und Stimmartikulation unterteilt werden kann, siehe Abbildung 2.1. Die erzeugten Schalldruckpegel liegen in einem Bereich von 20 dB bei leisem Flüstern, über 60 dB bei normaler Unterhaltung bis zu 110 dB bei Schreien. Aus der Abbildung wird die Lage der Grundtonbereiche und der Formantgebiete innerhalb der Hörfläche deutlich. Das "Flüstern" ist hierbei nicht berücksichtigt, da in diesem Fall keine tieffrequenten Anteile erzeugt werden und eine Verschiebung des Sprachspektrums zu hohen Frequenzen hin erfolgt. Für die Darstellung wurden Sprachsignale von mittlerer bis großer Lautstärke ausgewertet. Der Grundtonbereich beinhaltet die tiefsten Frequenzen, gefolgt von den Vokalen bis etwa 2 kHz. Die stimmhaften Konsonanten schließen den Bereich für Stimmklang ab, darüber befinden sich Frequenzanteile stimmloser Konsonanten, die der Stimmartikulation dienen.



Abbildung 2.1: Lage der Grundtonbereiche, und der Formantgebiete in der Hörfläche aus Fasold, Veres [6].

Die Buchstaben a bis g kennzeichnen dabei folgende Bereiche: (a) Hörschwelle, (b) Schmerzgrenze, (c) Grundtonhöhe Männer, (d) Grundtonhöhe Frauen, (e) Vokale, (f) stimmhafte Konsonanten und (g) stimmlose Konsonanten.

Der Grundtonbereich liegt bei Frauen etwa um eine Oktave höher als bei Männern, ansonsten besteht für den verwendeten Frequenzbereich kein Unterschied. Auch hinsichtlich der mittleren Schalldruckpegel (gemessen direkt vor dem Mund) bestehen für Vokale, sowie der stimmhaften und stimmlosen Konsonanten keine Unterschiede. In Abb. 2.2 sind die Beiträge der einzelnen Frequenzbereiche zur Gesamtlautstärke aufgetragen; a für männliche, b für weibliche Stimme. Den Bezugswert bildet der Schalldruckpegel bei 1000 Hz. Große Bedeutung kommt den mittleren bis tiefen Bereichen zu; hohe Frequenzen (> 4000 Hz) tragen nur mit geringerem Pegel zur Gesamtlautstärke des Sprachsignals bei.



Abbildung 2.2: Typischer Frequenzverlauf (a für männliche, b für weibliche Stimme) des Schalldruckpegels von Sprache, normiert auf den Schalldruckpegel bei 1 kHz, aus Fasold, Veres [6].

2.1.2 Hörspektrum

Der Bereich, in dem Schallereignisse wahrgenommen werden, ist abhängig von der Frequenz und vom Schalldruckpegel. Frequenzen tiefer als 16 Hz bezeichnet man als Infraschall, größer als 20000 Hz als Ultraschall. Die Wahrnehmung des Schalldrucks ist nach unten durch die Hörschwelle, nach oben durch die Schwelle der Schmerzempfindung beschränkt. Zwischen diesen Grenzen befindet sich der Hörbereich (siehe Abb.2.3).



Abbildung 2.3: Hörfeld mit Sprach- und Musikbereich, aus Dickreiter [7].

In einem Frequenzbereich zwischen 2 kHz und 5 kHz ist das menschliche Gehör besonders empfindlich, was auf den Einfluss des Gehörgangs zurückgeführt werden kann. Mit einer

Länge von etwa 3 cm hat er einen vergleichbaren Effekt, wie ein $\lambda/4$ – Resonator, der auf eine Frequenz von ca. 3000 Hz abgestimmt ist.

2.1.3 Verdeckungseffekte

Die Wahrnehmung von Schallereignissen beruht nicht allein auf den oben erwähnten Faktoren. Abgesehen vom geistigen und körperlichen Zustand des Hörers, wurde auf dem Gebiet der Psychoakustik der Einfluss von sog. Verdeckungseffekten untersucht. Diese können sowohl im Zeit-, als auch im Frequenzbereich stattfinden. D.h., das Störsignale, die vor, während, oder nach dem zu erkennenden Signal ertönen, einen Verdeckungseffekt zur Folge haben. Gleiches gilt für zeitgleich abgespielte Signale, deren Frequenzspektrum sich ganz oder teilweise überschneidet.

In Abbildung 2.4 wird der Verdeckungseffekt anhand Weißen Rauschens dargestellt. Dieses Störsignal ist einerseits tiefpassgefiltert (a), als auch hochpassgefiltert (b), für drei verschiedene, so genannte Dichtepegel l_R angegeben.



Abbildung 2.4:Mithörschwellen L_T von Teiltönen verdeckt durch Tiefpaßrauschen (a), bzw. Hochpaßrauschen (b) der Grenzfrequenz von 1 kHz für drei verschiedene Dichtepegel l_R in Abhängigkeit von der Testtonfrequenz (--- = Ruhehörschwelle), aus Zwicker [8].

Der Dichtepegel lässt sich über die Formel L = l_R + 10 log Δf [dB] in den Schalldruckpegel L umrechnen. Bei einem Dichtepegel von l_R = 40 dB und einer Bandbreite von Δf = 1000 Hz ergibt sich ein Schallpegel von L = 70 dB. Ein reiner Sinuston mit der Frequenz f = 500 Hz muss einen Schallpegel von mindestens 57 dB haben, um neben dem Störsignal (Maskierer) wahrgenommen zu werden. Ein Ton mit einer Frequenz von 1,5 kHz ist erst ab einem Pegel von 40 dB wahrnehmbar, obwohl das Rauschsignal in diesem Frequenzbereich von dem Tiefpass unterdrückt wird. Im Zeitbereich wird zwischen Vor-, Simultan- und Nachverdeckung unterschieden, wie aus Abbildung 2.5 hervorgeht. Hier wurde die Mithörschwelle mittels eines Schallimpulses bestimmt, der von einem Maskierer der Länge $\Delta t = 200$ ms verdeckt wurde. Ein Beispiel für zeitliche Verdeckung findet man in unserer Sprache. Da Vokale lauter sind als Konsonanten, werden Verschlusslaute von nachfolgenden Vokalen oft verdeckt.



Abbildung 2.5: Schema für die Zeiteffekte bei der Verdeckung: Aufgetragen wird der Pegel (über Ruhehörschwelle) SL_T des Testschalles in Abhängigkeit von der Zeit Δt nach dem Einschalten des Maskierers bzw. der Zeit t_v nach dem Abschalten, aus Zwicker [8].

2.2 Raumakustik

Die Wahrnehmung von Schallereignissen in einem geschlossenen Raum, hängt maßgeblich von dessen individueller Klangcharakteristik ab. Je nach Funktion, eignen sich bestimmte Räume eher für Musik- oder Sprachdarbietungen und müssen dafür unterschiedliche akustische Kriterien aufweisen.

Die Raumakustik beschäftigt sich mit der Beschreibung der akustischen Verhältnisse in geschlossenen Räumen. Bei der Lösung von Problemstellungen oder zur Festlegung von Kennwerten bedient man sich in Abhängigkeit des zu betrachtenden Frequenz- oder Zeitbereiches unterschiedlicher mathematischer Ansätze.

Die Wellentheorie (Kapitel 2.2.1) eignet sich für die Berechnung von Schallfeldern in geometrisch einfachen Räumen und im Besonderen auch für tiefe Frequenzen. Die Lösungen der Wellengleichungen bilden dabei die Grundlage für entsprechende Berechnungen.

Bei der geometrischen Raumakustik (Kapitel 2.2.2) verfolgt man die Schallausbreitung anhand von Schallstrahlen. Diese werden an Begrenzungsflächen den Gesetzen der Optik folgend absorbiert und/oder reflektiert. In Simulationsprogrammen wie dem später erwähnten ,Catt-Acoustic' wird diese Methode angewandt, um ein getreues Abbild des Direktschalls und der ersten Reflektionen berechnen zu können.

Mit fortschreitender Zeit bildet sich in geschlossenen Räumen ein diffuses Schallfeld aus. Anstelle einzelner Schallstrahlen betrachtet man nun die Energieverteilung im Raum. Sie wird bestimmt durch die Schallquellen und die Absorptionsfähigkeit der im Raum verwendeten Materialien. Die Beschreibung der zeitlichen Energieverteilung findet Anwendung in der statistischen Raumakustik (Kapitel 2.2.3) und ist für Berechnungen mittlerer bis hoher Frequenzen stationärer Signale geeignet, siehe auch Graber [9].

Aus diesen drei Teilbereichen werden im Folgenden einzelne Aspekte näher beschrieben, die zur Erklärung der in der Arbeit verwendeten Begriffe dienen.

2.2.1 Wellentheorie der Raumakustik

Die Wellengleichungen lassen sich aus physikalischen Grundgesetzen herleiten und bieten die Möglichkeit einer exakten Beschreibung des Schallfeldes. Die Definition der Randbedingungen gestaltet sich aber oftmals so schwierig, dass die Wellengleichungen nur für die Lösung von geometrisch "einfachen" Räumen zur Anwendung kommen. Auf eine detaillierte Herleitung wird hier verzichtet, siehe dazu Kuttruff [10] oder Graber [9]. Ausgehend von den Wellengleichungen wird der Fall der Totalreflexion und die Bildung von Eigenmoden erläutert.

Durch die Wellengleichungen lassen sich Kenngrößen wie Schalldruck p und Schallschnelle \vec{v} beschreiben als:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \cdot \Delta p \qquad (2.1) \quad \text{Wellengleichung für den Druck}$$
$$\frac{\partial^2 \vec{v}}{\partial t^2} = c^2 \cdot \Delta \vec{v} \qquad (2.2) \quad \text{Wellengleichung für die Schallschnelle}$$

Sie stehen über die Eulerschen Bewegungsgleichungen in folgendem Zusammenhang:

- grad
$$p = \rho \cdot \frac{\partial \vec{v}}{\partial t}$$
 (2.3)

Wie aus der Formel 2.3 ersichtlich wird, ist die maximale Änderung des Drucks betragsmäßig proportional zur Beschleunigung aber entgegengesetzt gerichtet. Der Parameter ρ beschreibt in dieser Gleichung die Dichte des Mediums.

2.2.1.1 Totalreflexion

Der Fall einer Totalreflexion kann, wie in Abbildung 2.6 dargestellt, als Überlagerung aus einer hinlaufenden und einer reflektierten Schallwelle betrachtet werden. Ausgehend von einfachen Randbedingungen (ebene Welle, harmonische Schwingung, senkrechter Schalleinfall) wird nur eine Ausbreitungsrichtung (x-Richtung) beschrieben, siehe Graber [9].



Abbildung 2.6: Totalreflexion, aus Graber [9]

Für die Betrachtung in x-Richtung lässt sich der Schalldruck nach Gleichung (2.1) schreiben als:

$$\frac{\partial^2 p(x,t)}{\partial t^2} = c^2 \cdot \frac{\partial^2 p(x,t)}{\partial x^2} \qquad (2.4)$$

Gleichung (2.3) lässt sich vereinfacht darstellen als:

$$-\frac{\partial p}{\partial x} = \rho \cdot \frac{\partial v_x}{\partial t} \qquad \text{und für ebene Wellen in der Form} \qquad \underline{v}_x = \frac{\underline{p}}{\rho \cdot c} \qquad (2.5)$$

Verwendet man als Ansatz für den Schalldruckverlauf p eine komplexe harmonische Schwingung der Form $\underline{\hat{A}} \cdot e^{j\omega t}$ folgt daraus:

$$\underline{p} = \underline{p}_{h} + \underline{p}_{r} = (\underline{\hat{p}}_{h} \cdot e^{-jkx} + \underline{\hat{p}}_{r} \cdot e^{jkx}) \cdot e^{j\omega t} \quad (2.6)$$

Der Unterstrich <u>p</u> kennzeichnet die komplexe Darstellung des Schalldruckes. Die Wellenzahl k ist dabei definiert als:

$$k = \frac{\omega}{c} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda} \quad (2.7)$$

Aus der Überlagerung der hin- und rücklaufenden Welle ergibt sich für die Schallschnelle \underline{v}_x mit Gleichung (2.5) :

$$\underline{\underline{v}}_{x} = \underline{\underline{v}}_{xh} + \underline{\underline{v}}_{xr} = \frac{1}{\rho \cdot c} (\underline{\underline{p}}_{h} - \underline{\underline{p}}_{r}) = \frac{1}{\rho \cdot c} (\underline{\hat{p}}_{h} \cdot e^{-jkx} - \underline{\hat{p}}_{r} \cdot e^{jkx}) \cdot e^{j\omega t}$$
(2.8)

Die Randbedingung bei schallharter Wand lautet:

$$\underline{v}_x(x=0) = 0$$

d.h. die Schallschnelle am Reflexionsort ist Null. Damit ergibt sich aus Gleichung (2.8) für die Scheitelwerte von Schalldruck und Schallschnelle:

$$\underline{\hat{p}}_{h} = \underline{\hat{p}}_{r}$$
 und $\underline{\hat{v}}_{xh} = -\underline{\hat{v}}_{xr}$.

Das bedeutet, dass bei Totalreflexion an einer schallharten Wand der Schalldruck pgleichphasig reflektiert wird und am Reflexionsort die Resultierende maximal ist, während die Schallschnelle \vec{v} mit einem Phasensprung von 180° reflektiert wird und die Resultierende am Reflexionsort immer null ergibt. Für den Druck- und den Schnelleverlauf bilden sich stehende Wellen, da die Nulldurchgänge und Scheitelwerte ortsfest sind.

2.2.1.2 Eigenmoden eines Raumes

Stehende Wellen bilden sich aus, wenn der Schallstrahl zwischen zwei reflektierenden Flächen quasi in sich selbst zurückläuft. Die tiefste Resonanz tritt dann auf, wenn der Abstand der beiden Flächen genau eine halbe Wellenlänge beträgt. Höhere Resonanzen bilden sich bei ganzzahligen Vielfachen der tiefsten Resonanzfrequenz aus. Es gilt folgender Zusammenhang:

$$l = n \frac{\lambda_r}{2} \rightarrow \lambda_r = \frac{2l}{n};$$
 $f_r = \frac{c}{\lambda_r} = \frac{nc}{2l}$ mit $n \in \mathbb{N}$

Dabei ist l eine Kantenlänge des Raumes, f_r die Resonanzfrequenz, λ_r ihre Wellenlänge, c die Schallgeschwindigkeit und n eine ganze Zahl mit Werten größer, oder gleich Null. Für den Schalldruck ergibt sich dann aus Gl. (2.6) und den Randbedingungen der Totalreflexion:

$$\underline{p} = \underline{\hat{p}}_{h} (e^{-jkx} + e^{jkx}) \cdot e^{j\omega t} = 2\hat{p}_{h} \cdot \cos(kx) \cdot e^{j\omega t}$$
(2.9)

Den Cosinus-Term lässt sich nach Gleichung (2.7) auch schreiben als:

$$\cos(k_r x) = \cos\left(\frac{2\pi}{\lambda_r}x\right) = \cos\left(\frac{n\pi}{l}x\right)$$

10

Erweitert man die Gl. (2.9) auf drei Dimensionen, so folgt für die Schalldruckverteilung:

$$\underline{p} = 8\underline{\hat{p}}_{h} \cdot \cos\left(\frac{n_{x} \cdot \pi \cdot x}{l_{x}}\right) \cdot \cos\left(\frac{n_{y} \cdot \pi \cdot y}{l_{y}}\right) \cdot \cos\left(\frac{n_{z} \cdot \pi \cdot z}{l_{z}}\right) \cdot e^{j\omega t}$$
(2.10)

Jedem Wertetripel n_x , n_y , n_z entspricht eine Eigenschwingung des Raumes, deren räumliche Schalldruckverteilung sich mit obiger Gleichung berechnen lässt. Die Ordnungszahlen n_x , n_y und n_z geben die Zahl der Knotenebenen an, die man beim Durchqueren des Raumes parallel zur entsprechenden Koordinatenachse passiert, Graber [9].

Je nach Ausbreitungsrichtung unterscheidet man zwischen axialen, tangentialen und schrägen Moden. Besondere Bedeutung kommt den axialen Moden zu, deren Ausbreitungsvektor parallel zu einer Raumkante liegt (Abb.2.7). Sie bewirken die größten Resonanzüberhöhungen und können in kleinen Räumen zu deutlich ausgeprägten, diskreten Raummoden führen.



Abbildung 2.7: Eigenmoden, aus Everest A. [11]

2.2.2 Geometrische Raumakustik

Bei der geometrischen Raumakustik wird die Schallausbreitung anhand von Schallstrahlen verfolgt. Dabei ist zu beachten, dass die betrachteten Bereiche (Entfernungen, Abmessungen reflektierender Flächen usw.) groß im Vergleich zu den vorkommenden Wellenlängen sind.

2.2.2.1 Spiegelquellenverfahren

In Abbildung 2.8 sind einige Strahlenwege eingezeichnet, auf denen der Schall von der Quelle A zum Punkt P gelangt. Der Direktschall erreicht den Empfänger dabei auf dem kürzesten Weg. Alle nachfolgenden Beiträge gelangen über Reflexionen an Decke und Rückwand zum Beobachtungspunkt.



Abbildung 2.8: Längsschnitt eines Auditoriums mit einigen Schallstrahlen und Spiegelschallquellen. A=Schallquelle, A_n=Spiegelquellen n-ter Ordnung, aus Kuttruff [10]

Die mit A_n bezeichneten Quellen sind Spiegelbilder der Originalschallquelle. Der Index n gibt dabei die Ordnung der Spiegelquelle an. Vom Beobachtungspunkt P aus gesehen, scheinen die reflektierten Strahlen von diesen virtuellen Schallquellen herzukommen. Eine Abschwächung des Signals aufgrund unvollkommener Reflexion kann durch Multiplikation des Signals mit einem Faktor (1- α) berücksichtigt werden. α gibt dabei den Absorptionsgrad der Fläche an. Auf einem leistungsstarken Rechner kann dieser Prozess auf alle Raumwände ausgedehnt und mit vielen Schallstrahlen wiederholt werden und bildet so die Grundlage zur Berechnung des Schallfeldes in Simulationsprogrammen, siehe Kuttruff [10].

2.2.2.2 Impulsantwort

Die zeitliche Abfolge der Rückwürfe kann auch als Summe dargestellt werden. Das von der Schallquelle erzeugte Signal s(t), wobei sich die Funktion s(t) auf den Schalldruck bezieht, sowie die Beiträge der Spiegelschallquellen lassen sich zusammenfassen als:

$$s'(t) = \sum a_n s(t - t_n)$$
 (2.11)

Die Zeiten t_n sind die Verzögerungen des Rückwurfs gegenüber dem Direktschall. Bei jeder Wandreflexion wird das Signal nicht nur abgeschwächt zurückgeworfen, sondern auch klanglich verfärbt, da der Reflexionsfaktor einer Wand frequenzabhängig ist. Ist das von der Schallquelle ausgesandte Signal ein sehr kurzer Impuls, dargestellt durch die Diracsche Deltafunktion, dann stellt die Gleichung (2.12) die Impulsantwort des Raumes dar:

$$g(t) = \sum_{n} a_n \delta(t - t_n)$$
 (2.12)

In Abbildung 2.9, die eine solche Impulsantwort zeigt, kennzeichnet der erste Strich den Direktschall, jeder weitere Strich stellt einen Rückwurf dar. Die zeitliche Dichte der Rückwürfe nimmt quadratisch mit der Zeit zu.



Abbildung 2.9: schematische Impulsantwort eines Raumes, aus Kuttruff [10]

Das allmähliche Abklingen der Schallenergie bezeichnet man als Nachhall und wird im nächsten Abschnitt genauer behandelt. Eine in einem realen Raum gemessene Impulsantwort ist in Abbildung 2.10 dargestellt. Ihre kompliziertere Struktur ist auf die Signalverzerrung bei der Reflexion zurückzuführen, sowie darauf, dass reale Wände die auftreffende Schallenergie nicht ausschließlich nach dem Reflexionsgesetz zurückwerfen, sondern dass ein Teil davon gestreut wird, Kuttruff [10].



Abbildung 2.10: exemplarische Impulsantwort eines Raumes, aus Müller, Möser [12]

2.2.3 Statistische Raumakustik

Betrachtet man eine Schallquelle als Energiequelle und stellt die Frage nach der sich einstellenden Energiedichte, so bewegt man sich auf dem Gebiet der statistischen Raumakustik. Dabei wird auf die Verfolgung einzelner Schallstrahlen verzichtet und ein diffuses Schallfeld vorausgesetzt. Durch die Bildung einer Energiebilanz

zeitliche Änderung des Energieinhalts = zugeführte Energie – absorbierte Energie gewinnt man weitere Erkenntnisse über das Schallfeld in einem Raum. Die relevanten Größen sind hierbei die Energiedichte w, das Raumvolumen V, die Oberfläche S und der Schallabsorptionsgrad α.

2.2.3.1 Hallradius und Nachhall

Der Energieinhalt ist definiert als die Energiedichte w multipliziert mit dem Raumvolumen V, die sekundlich zugeführte Energie wird durch die Schallquellenleistung P gekennzeichnet. Die Oberfläche S eines Raumes kann unterteilt werden in die Teilflächen S_i mit den jeweils einheitlichen Absorptionsgraden α_i . Jede dieser Teilflächen absorbiert pro Sekunde eine Energie von α_i S_icw/4. Durch Einführen der "äquivalenten Absorptionsfläche"

$$A = \sum_{i} \alpha_{i} S_{i} \tag{2.13}$$

geht die Energiebilanz über in

$$V\frac{dw}{dt} = P(t) - \frac{c}{4}Aw \qquad (2.14)$$

und damit in eine Differentialgleichung erster Ordnung für die Energiedichte. Zwei Sonderfälle sollen im Folgenden näher betrachtet werden.

Bei konstanter Leistung P und Energiedichte w (stationärer Fall) ergibt die Energiebilanz:

$$w = \frac{4P}{cA} \tag{2.15}$$

In der Nähe der Schallquelle überwiegt die von ihr direkt, also ohne Mitwirkung des Raumes erzeugte Energiedichte. Bei gleichmäßiger Abstrahlung des Schalls in alle Richtungen ist sie:

$$w_d = \frac{P}{4\pi c r^2} \tag{2.16}$$

wobei r der Abstand von der Schallquelle ist. Die Entfernung, in der beide Energiedichten gleich groß sind, wird als Hallradius r_h bezeichnet (siehe Abbildung 2.11). Er ergibt sich aus den Gleichungen (2.15) und (2.16) zu:

$$r_h = \sqrt{\frac{A}{16\pi}} \tag{2.17}$$

Die gesamte Energiedichte $w + w_d$ lässt sich nun schreiben als:

$$w_{ges} = \frac{P}{4\pi c} \left(\frac{1}{r^2} + \frac{1}{r_h^2} \right)$$
 (2.18)

Der erste Term stellt das sog. Direktfeld dar, der zweite das Hallfeld, oder auch "Diffusfeld", Kuttruff [10].



Abbildung 2.11: Zur Definition des Hallradius r_h. (w_d, w: Energiedichte im Diffusfeld, bzw. im Direktschallfeld), aus Kuttruff [10]

Bei dem zweiten Sonderfall geht man davon aus, dass die Schallquelle bis zum Zeitpunkt t= 0 in Betrieb war und bis dahin eine gewisse Energiedichte w_0 aufgebaut hat. Bei Abschalten der Quelle (P=0) ergibt sich für die nunmehr homogene Differentialgleichung folgende Lösung:

$$w(t) = w_0 e^{-cA/4V}$$
 für $t \ge 0$ (2.19)

Diesen durch Gleichung (2.19) beschriebenen Abklingvorgang bezeichnet man als Nachhall des Raumes. Die Geschwindigkeit des Abklingens kennzeichnet man durch die sogenannte Nachhallzeit. Man versteht darunter die Zeit, in welcher die Energiedichte auf den millionsten Teil ihres Anfangswertes abfällt. Eingesetzt in die Gleichung (2.19)

$$10^{-6} = e^{-cA/4V}$$

ergibt sich durch Logarithmieren:

$$T = \frac{24 \cdot \ln 10}{c} \cdot \frac{V}{A} \tag{2.20}$$

und mit einsetzen des Wertes der Schallgeschwindigkeit c in Luft:

$$T = 0.163 \frac{V}{A} \tag{2.21}$$

Neben dem Reflexionsgesetz bildet diese Gleichung (2.21) die wichtigste Regel der Raumakustik und ist nach ihrem Entwickler W.C. Sabine benannt, siehe Kuttruff [10].

2.2.4 Messmethoden der Sprachverständlichkeit

In diesem Abschnitt werden neben der Nachhallzeit noch weitere Begriffe vorgestellt, die es ermöglichen, Räume nach einem akustischen Merkmal zu bewerten und zu vergleichen.

2.2.4.1 Nachhallzeiten T₂₀, T₃₀ und EDT

Messtechnisch erfolgt eine Berechnung der Nachhallzeit über die Raumimpulsantwort mittels Rückwärtsintegration nach Schröder [13]. In Abbildung 2.12 ist ein Schallpegelverlauf exemplarisch dargestellt. Die Parameter T_{20} und T_{30} geben dabei das Messintervall an, aus dem sie ermittelt werden, nämlich den Bereich zwischen -5 dB und -25 dB, bzw. -35 dB des Schallpegels, bezogen auf den stationären Pegel.



Abbildung 2.12: Schallpegelverlauf

Die Frühe Zerfallszeit, oder Early-Decay-Time (EDT) nach Jordan [14] misst nur den Abfall von 0 dB auf -10 dB, wodurch sich bei der Interpolation meist eine längere Nachhallzeit ergibt. Besonders in kleinen Räumen erhält man mit dieser Kennzahl oft eine bessere Übereinstimmung mit den realen Gegebenheiten, da der Einfluss von Raumeigenmoden in diesen deutlicher zur Geltung kommt.

2.2.4.2 Deutlichkeitsmaß für Sprache C₅₀

Der Begriff ,Deutlichkeit' oder ,Definition' wurde von R. Thiele [15] im Jahr 1953 vorgestellt. Zur Bestimmung der Deutlichkeit wird das Verhältnis der Schallintensität über zwei Zeitintervalle gebildet:

$$D = \int_{0}^{50ms} I(t)dt / \int_{0}^{\infty} I(t)dt \quad [\%]$$

Die Wahl der Zeitspanne von 50ms ergibt sich aus der Erkenntnis, dass alle Rückwürfe, die innerhalb der ersten 50ms nach Eintreffen des direkten Schalls beim Zuhörer ankommen, zur Erhöhung der Lautstärke beitragen, ohne die Richtungslokalisation der Schallquelle zu stören. Die ursprüngliche Berechnung ist in die heute verwendete Form der ISO 3382 [16] mit kleinen Änderungen übernommen worden.

$$C_{50} = 10 \cdot \log \left(\int_{0}^{50ms} p^2(t) dt \middle/ \int_{50ms}^{\infty} p^2(t) dt \right) \quad [dB]$$

Man unterscheidet dabei zwischen folgenden Bereichen:

 $\begin{array}{ll} C_{50} > \ 2 \ dB & \mbox{sehr gute Verständlichkeit} \\ 2 \ dB > C_{50} > -3 \ dB & \mbox{gute Verständlichkeit} \\ -3 \ dB > C_{50} > -8 \ dB & \mbox{noch ausreichende Verständlichkeit} \end{array}$

2.2.4.3. Schwerpunktzeit t_s

Bei der von Kürer [17] definierten Schwerpunktzeit handelt es sich ebenfalls um einen Parameter, der aus dem Verhältnis der Energie von frühen zu späten Reflexionen gebildet wird. Die Aufteilung in Zeitbereiche von nützlichen und schädlichen Rückwürfen wird dabei vermieden.

$$t_s = \int_0^\infty t p^2(t) dt \bigg/ \int_0^\infty p^2(t) dt$$

Niedrige Werte von t_s deuten dabei auf ein hohes Maß an Sprachverständlichkeit hin.

2.2.4.4 Articulation Loss of Consonants (AL_{cons})

Die Niederländer V.M.A. Peutz und W. Klein [18] stellten fest, dass bei der Artikulation von Sprache, speziell der Verlust von Konsonanten zu einer schlechteren Verständlichkeit führt. Der von Ihnen danach benannte Parameter AL_{cons} wird über einen Hörtest ermittelt. Der Vergleich der Ergebnisse mit einer Prozentskala ermöglicht eine Aussage über die Verständlichkeit. Dabei entspricht:

$Al_{cons} < 2\%$	ideale Verständlichkeit	(,sehr gut')
$Al_{cons} = 279$	% gute Verständlichkeit	(,gut')
$Al_{cons} > 7\%$	befriedigende Verständlichkeit	(,befriedigend')
$Al_{cons} > 20\%$	unbrauchbare Verständlichkeit	(,schlecht')

2.2.4.5 Bass – Ratio, RT

Die 'Wärme' eines Klanges wird von den tiefen Frequenzen darin beeinflusst. Das Bass – Verhältnis beschreibt diese Klangcharakteristik und bestimmt sich aus dem Verhältnis der Nachhallzeiten von tiefen zu mittleren Frequenzen.

$$BR = \frac{RT_{125} + RT_{250}}{RT_{500} + RT_{1000}}$$

RT steht hierbei für ,Reverberation Time' (Nachhallzeit), der Index gibt das jeweilige Oktavband an. Für Sprache sollte laut Beranek [19] ein Wert von BR = 1.0 nicht überschritten werden.

2.2.4.6 Echo Kriterium nach Dietsch und Kraak [20]

Ausgehend von der Raumimpulsantwort wird folgende Aufbaufunktion der Schwerpunktzeit t_s, siehe 2.2.4.3, gebildet:

$$t_{s}(\tau) = \int_{0}^{\tau} \left| p(t)^{n} \right| \cdot t dt / \int_{0}^{\tau} \left| p(t) \right|^{n} dt \quad [\text{ms}] \qquad \text{mit } n = 2/3 \text{ für Sprache.}$$

Danach wird der über diese Aufbaufunktion $t_S(\tau)$ gleitende Differenzenquotient

$$EK(\tau) = \frac{\Delta t_{s}(\tau)}{\Delta t_{E}} \qquad \text{mit } \Delta t_{E} = 9 \text{ ms für Sprache}$$

ermittelt und mit echokritischen Grenzwerten $EK_{gr N}$ verglichen. Werden diese Grenzwerte $EK_{gr N}$ vom Zeitverlauf von $EK(\tau)$ überschritten, so sind bei N% der Hörer Echostörungen zu erwarten. Diese Grenzwerte betragen: $EK_{gr 10} = 0,9$ und $EK_{gr 50} = 1,0$ für Sprache. Werden die Grenzwerte $EK_{gr N}$ mehrfach überschritten, so tritt dann ein Mehrfachecho auf, wenn zwischen zwei aufeinanderfolgenden Überschreitungen ein Mindestzeitabstand von $\Delta t_{E \min} = 50$ ms liegt. Erfolgen die Überschreitungen in regelmäßigen Abständen $\Delta t_E > \Delta t_{E \min}$, so tritt ein Flatterecho auf.

2.2.4.7 Articulation Index (AI)

Der Artikulationsindex (AI) geht zurück auf N.R. French und J.C. Steinberg [21], die ein Kriterium zur Bewertung von Sprachverständlichkeit entwickelten, welches sich in einem Wertebereich zwischen 0 und 1 bewegt. Dazu unterteilten Sie einen Frequenzbereich zwischen 250 Hz und 7000 Hz in 20 Bänder, so dass jedes Band maximal einen Anteil von 5% zum AI beisteuert. Gewichtet werden diese Bänder mit einem Faktor W, der sich aus den Lautstärkeverhältnissen von Störgeräuschen zu Sprachsignal zusammensetzt. In ANSI 3.5-1969 [22] wurde der AI als standardisierte Methode zur Ermittlung der Sprachverständlichkeit festgelegt.

2.2.4.8 Speech Intelligibility Index (SII)

In der überarbeiteten Version der Norm ANSI 3.5 [23] wurde ein neues Verfahren eingeführt, dass zur Ermittelung der Sprachverständlichkeit dienen soll. Der Speech Intelligibility Index (SII) beruht von der Methode her auf dem AI, verfügt aber über ein verfeinertes Berechnungsverfahren. Der SII setzt sich dabei aus dem Produkt der ,band – importance – function' und der ,band – audibility – function' zusammen. Erste gibt dabei an, wie stark ein Frequenzband zur Sprachverständlichkeit beiträgt, die zweite Funktion legt einen Wert fest, der den effektiv beisteuernden Teil des Dynamikbereiches von Sprache innerhalb eines Bandes bestimmt.

$$S = \sum_{i=1}^{n} I_i A_i$$

Die Werte des SII bewegen sich zwischen 0 und 1, wobei gute Sprachverständlichkeit ab Werten größer 0.75 und schlechte Sprachverständlichkeit für Werte unter 0.45 zu erwarten ist.

2.2.4.9 Speech Transmission Index (STI)

Der Speech Transmission Index (STI) ist der gebräuchlichste Kennwert, der zur Klassifizierung der Sprachverständlichkeit verwendet wird. Auch im Rahmen dieser Arbeit werden Aussagen mit dem ermittelten STI unterlegt. Aus diesem Grund wird das Zustandekommen dieses Kennwertes im Folgenden ausführlicher beschrieben.

Die Modulation – Transfer – Funktion (MTF)

Der Raum zwischen Sprecher und Hörer kann als ein System zur Schallübertragung angesehen werden, welches durch die MTF beschrieben wird. Das Sprachsignal kann durch die Einhüllende der Schallintensität beschrieben werden, die man sich wiederum aus sinusförmigen Komponenten zusammengesetzt vorstellen kann. Die Güte des Raumes im Hinblick auf die Sprachverständlichkeit wird dadurch bestimmt, wie getreu die Modulationen der Schallintensitäten, die vom Sprecher produziert werden, an der Hörerposition ankommen. Für ein Eingangssignal der Form \overline{I}_i (1 + cos $2\pi Ft$), mit F = Modulationsfrequenz ergibt sich ein Ausgangssignal \overline{I}_0 [1 + $m \cos 2\pi F(t - \tau)$], wobei m = Modulationsindex und τ der zeitliche Versatz ist. Die Funktion m(F) ist definiert als die Modulation – Transfer – Funktion, siehe Abbildung 2.13. Die MTF ist nur für den Bereich von F von Interesse, der relevant für Sprache ist, also für den Bereich, in dem rhythmische Schwankungen auftreten. Normalerweise liegt dieser Bereich zwischen 0.4 Hz und 20 Hz, siehe auch Houtgast [24].

MTF und Speech – Transmission – Index (STI)

Für die Umwandlung der Funktion m(F) in den STI, wird zunächst zu jedem Funktionswert das Signal – Rausch - Verhältnis (S/N) gebildet. Daraus wird ein mittleres Signal – Rausch – Verhältnis errechnet $(\overline{S/N})$ und der Dynamikumfang auf ±15 dB begrenzt. Nach entsprechender Normierung liegen die Werte des Speech – Transmission – Index in einem Bereich zwischen 0 und 1, wobei die Sprachverständlichkeit nach folgender Tabelle eingeordnet werden kann, aus Houtgast [24].

$$STI = 0.2 \dots 0.4$$
schlecht $STI = 0.4 \dots 0.6$ mäßig $STI = 0.6 \dots 0.8$ gut $STI = 0.8 \dots 1.0$ exzellent



Abbildung 2.13: Die Modulation-Transfer-Funktion (MTF): m(F) quantifiziert die Abnahme des Modulationsindex bei Übertragung eines Signals im Raum. Als Testsignal wird ein sprachsimulierendes Rauschen mit 100%-iger Intensitätsmodulation verwendet, mit der Modulationsfrequenz F als Parameter. Von der MTF wird der Speech Transmission Index abgeleitet als Maß für die Sprachverständlichkeit, Houtgast [24].

3. Generierung des Audiomaterials

Um den Einfluss der Nachhallzeit tiefer Frequenzen auf die Sprachverständlichkeit bestimmen zu können, werden Hörtests durchgeführt. Das dabei verwendete Sprachmaterial soll sich hinsichtlich der Nachhallzeiten ausschließlich im tieffrequenten Bereich unterscheiden. Zunächst werden dafür die entsprechenden Impulsantworten generiert und verwendet, um das Sprachmaterial für den Hörtest zu erhalten. Durch Faltung der Impulsantwort mit dem Sprachsignal wird diesem die Klangcharakteristik des jeweiligen Raumes beigefügt. Dieses Kapitel beschreibt die Erzeugung der Impulsantworten auf zwei verschiedene Arten, aus der Simulation und aus realen Messungen.

3.1 Simulation

Für die Erstellung des Computermodells eines Raumes wurde das Programm ,Catt-Acoustic' verwendet. Durch Eingabe von Koordinaten werden in diesem Programm Punkte und in weiterer Folge Flächen festgelegt. Den Flächen können beliebige Absorptions- und Diffusionskoeffizienten zugeordnet werden und ermöglichen so die Simulation beliebiger Raumgeometrien. Für die Ermittlung raumakustischer Kennwerte werden die Positionen von Schallempfängern und Schallquellen und deren Abstrahl- und Richtcharakteristik definiert. Unter Verwendung des Spiegelquellenverfahrens (Kapitel 2.2.2) und Methoden der statistischen Raumakustik (Kapitel 2.2.3) können verschiedene Kennwerte und auch die Impulsantwort des Raumes ermittelt werden.

3.1.1 Binaurale Impulsantworten

Als Simulationsobjekt wird ein Klassenzimmer verwendet, das vorher vermessen worden ist. Mit Hilfe der Software ,Catt-Acoustic' wird das Computermodell zunächst so angepasst, dass die simulierten Werte mit den gemessenen übereinstimmen. Anschließend werden die unterschiedlichen Nachhallzeiten berechnet und die Sprachverständlichkeit mittels STI ermittelt (siehe Kap. 2.2.4.9). Um dies zu erreichen, werden die Absorptionskoeffizienten α eines Materials variiert, welches als ringförmige Verkleidung an der Decke und flächig an der rückwärtigen Wand angebracht ist (siehe Abbildung 3.1). Ein α -Wert von 0 entspricht dabei einer Totalreflexion im entsprechenden Frequenzband ein Wert von 100 einer kompletten Absorption.



Abbildung 3.1: Simulation des Klassenzimmers mittels Catt-Acoustic. Rote Bereiche kennzeichnen die Flächen, deren Absorptionskoeffizienten zur Erzeugung unterschiedlicher Nachhallzeiten variiert wurden.

Die mit Catt-Acoustic ermittelten Nachhallzeiten und die zugehörigen STI-, bzw. α-Werte sind in Tabelle 3.1 und 3.2 bzw. Abbildung 3.2 dargestellt. Es wurde darauf geachtet, Nachhallzeiten zu generieren, die auch im echten Klassenraum mit entsprechender Absorberanordnung realisiert werden können. Wie der STI zeigt, tritt eine deutliche Veränderung nur zwischen dem Ausgangszustand (Zustand 0) und dem Zustand 1 auf. Bei allen weiteren Raumzuständen bleibt der STI-Wert nahezu konstant. Eine Erklärung dafür findet sich in Dalenbäck [25]. Zur Berechnung des STI-Wertes wird der Signal-Rausch-Abstand (SNR) bestimmt. Dabei gehen die einzelnen Frequenzbänder, nach den Vorgaben von Steeneken, Houtgast [26] mit unterschiedlicher Gewichtung ein. Der Anteil der tiefen Frequenzbereiche hat auf das gesamte ausgewertete Spektrum nur einen kleinen Einfluss und führt daher nur zu geringfügigen Veränderungen des STI.

f in Hz	125	250	500	1000	2000	4000	STI
Zustand 0	2,21	1,82	1,49	1,28	1,18	1,08	0,53
Zustand 1	1,61	1,14	0,67	0,63	0,64	0,6	0,67
Zustand 2	1,62	0,73	0,67	0,64	0,63	0,61	0,69
Zustand 3	1,26	0,7	0,68	0,62	0,62	0,61	0,69
Zustand 4	0,91	0,71	0,65	0,63	0,63	0,6	0,7
Zustand 5	0,68	0,75	0,67	0,64	0,62	0,6	0,7

 Tabelle 3.1: aus der Computersimulation ermittelte

 Nachhallzeiten in Sekunden, sowie STI-Werte

f in Hz	125	250	500	1000	2000	4000
α-Werte Zstd. 0	6	6	3	4	3	0
α-Werte Zstd. 1	18	31	78	74	69	60
α-Werte Zstd. 2	18	81	78	74	69	60
α-Werte Zstd. 3	31	81	78	74	69	60
α-Werte Zstd. 4	56	81	78	74	69	60
α-Werte Zstd. 5	99	81	78	74	69	60

Tabelle 3.2: α-Werte der verschiedenen Raumzustände



Abbildung 3.2: graphische Darstellung der realisierten Nachhallzeiten verschiedener Zustände

Für den Hörtest werden drei Positionen innerhalb des Raumes ausgewählt (Position 1, 6 und 8; siehe Abbildung 3.3). Dadurch soll auch der Einfluss von Raummoden und des Verhältnisses von Direkt- zu Diffusschall berücksichtigt werden. Zu jeder Position werden sechs Impulsantworten der Zustände 0 bis 5 generiert.

Aus den erzeugten Raumimpulsantworten werden mit der Übertragungsfunktion eines Kunstkopfes die binauralen Raumimpulsantworten gebildet. Durch den speziellen Frequenzverlauf wird das Klangempfinden, dem eines echten Zuhörers nachgebildet. Nach der Faltung mit einem nachhallfreien Sprachsignal erhält man schließlich Audiodateien, welche die Klangcharakteristik des jeweiligen Raumzustandes wiedergeben.



Abbildung 3.3: Hörpositionen 01-09 und Quellposition A0 im Simulationsobjekt

3.1.2 Simulierte Raummoden

Für eine weitere Versuchsreihe wird Audiomaterial derart bearbeitet, dass der Hörer scheinbar im Schalldruckmaximum einer Raumeigenfrequenz sitzt (siehe Kap. 2.2.1.2). Dazu wird das Frequenzspektrum einer Impulsantwort analysiert, die aus der Messung eines Klassenzimmers des Bundesrealgymnasiums Kepler in Graz stammt. Mit Hilfe der Software ,Samplitude' wird eine Fast-Fourier-Transformation (FFT) vorgenommen, um das Zeitsignal in den Frequenzbereich transformiert darzustellen (Abbildung 3.4). In einem Bereich zwischen 80 Hz und 300 Hz werden nacheinander die Pegel innerhalb einer Terzbandbreite um jeweils 18 dB angehoben. Die dafür verwendete Impulsantwort, sowie die Bearbeitungen finden sich als Wave-Dateien in Anhang A. Anschließend wird mit dem Programm durch Faltung der veränderten Impulsantworten mit einem nachhallfrei aufgenommenen Sprachsignal, das Audiomaterial für einen Hörtest generiert. Abbildung. 3.4 zeigt das Spektrum einer Impulsantwort mit einer Pegelanhebung bei 258 Hz, Abb. 3.5 das Frequenzspektrum des damit gefalteten Sprachsignals.



Abbildung 3.4: Frequenzspektrum einer Impulsantwort, mit Pegelanhebung im Terzband um 258 Hz



Abbildung3.5: Frequenzspektrum des Sprachsignals nach Faltung mit obiger Impulsantwort

3.2 Messungen

Um eine weitere Testreihe durchzuführen, die auf realen Impulsantworten basiert, wird ein Raum der Firma Müller-BBM GmbH ausgewählt. Ziel der Messungen ist es, mehrere Impulsantworten zu erzeugen, deren Nachhallzeiten für Frequenzen oberhalb 200 Hz möglichst gleich und für darunter liegende Frequenzen verschieden sind.

3.2.1 Messraum

Der Raum entspricht seiner Größe nach dem eines kleinen Seminarraumes (siehe Abb. 3.6). Bei der ersten Besichtigung des Raumes wird ein "Dröhnen" wahrgenommen und auf den unbehandelten Zustand des Raumes zurückgeführt. Trotz Maßnahmen zur Bedämpfung bleibt das "Dröhnen" während der ersten Messungen erhalten. Deshalb wird versucht, durch die gezielte Anordnung von Plattenschwingern die Auswirkungen der Raummoden abzuschwächen, um eine bessere akustische Ausgangssituation zu erhalten.



Abbildung 3.6: Grundriß des Messraums mit Meß- und Quellpositionen (QP).

3.2.2 Planung und Bau von Plattenschwingern

Um die Plattenschwinger auf ihre Resonanzfrequenz abstimmen zu können, werden zunächst folgende Überlegungen angestellt. Nach Fasold und Veres [6] kann eine Grenzfrequenz f_{gr} bestimmt werden, ab welcher der Frequenzabstand der Raummoden klein genug ist, um die Gleichmäßigkeit des Schallfeldes nicht weiter zu stören. Sie berechnet sich aus folgender Gleichung:

$$f_{gr} \approx 2 \cdot 10^3 \sqrt{\frac{T}{V}}$$
 [Hz] (3.1)

dabei ist T die Nachhallzeit in Sekunden und V das Raumvolumen in m³. Für den Messraum mit den Abmessungen 8,3 x 4,1 x 3,1 m (Länge x Breite x Höhe) und einer angenommenen, maximalen Nachhallzeit von T = 1,2 Sekunden ergibt sich aus der Gleichung (3.1) eine Grenzfrequenz von f_{gr} = 205 Hz. Die hohe Grenzfrequenz ist bedingt durch das kleine Volumen. Eine weitere Abschätzung wird mit einer minimalen Nachhallzeit von 0,5 Sekunden vorgenommen und liefert eine Grenzfrequenz von f_{gr} = 133 Hz. Der gewählte Nachhallzeitbereich zwischen 1,2 und 0,5 Sekunden ist in etwa der, in dem sich die unterschiedlichen Impulsantworten später befinden sollen. Der kritische Bereich wird nochmals mit einer Software (Hunecke [27]) überprüft, die eine Auflistung der ersten Raummoden ermöglicht, siehe Abbildung 3.7.



Abbildung 3.7: Bestimmung der Raumeigenmoden, aus Hunecke [27]

Bei der Verteilung der Raummoden fällt auf, dass bis etwa 100 Hz Eigenmoden isoliert auftreten. Aus der Tabelle finden jene Frequenzen besondere Beachtung, bei denen mehrere Moden zusammenfallen. Das sind bei den gewählten Raummaßen die Frequenzen um 41, 59, 69 und 83 Hz, die auch mittels eines Signalgenerators im Messraum überprüft werden. Für die exakte Platzierung der Absorber werden aus den Graphiken die Orte der Schalldruckmaxima notiert und nach Gehör mit denen im Raum verglichen.

Da bei einem durchschnittlichen Frequenzspektrum von Sprache(vergl. Abb. 2.1) nur Frequenzen bis etwa 80 Hz hinsichtlich der Schalldruckpegel von Bedeutung sind, wird auf die gezielte Bedämpfung tieferer Frequenzen durch Plattenschwinger verzichtet. Es wird angenommen, dass der Bereich zwischen 80 Hz und 100 Hz für das ,Dröhnen' verantwortlich ist und in weiterer Folge werden zwei Plattenschwinger konstruiert, deren Resonanzfrequenz auf 84 Hz, bzw. 94 Hz abgestimmt ist. Der Vergleich der Messungen mit und ohne Plattenschwinger (siehe Abbildung 3.8) zeigt eine deutliche Abnahme der Nachhallzeit auch weit bis in höhere Frequenzbereiche hinein.





Abbildung 3.8: Vergleich der Nachhallzeiten mit und ohne Plattenschwingern

3.2.3 Messaufbau und Messgeräte

Um einen möglichst linearen Frequenzgang über einen Bereich von 40 Hz bis 10 kHz zu gewährleisten, werden für die Messungen zwei Lautsprecher verwendet. Ein Frequenzsweep von 20 Hz bis 400 Hz wird über den Lautsprecher des Typs MP Tube 30 der Marke MAC Audio, ein zweiter Sweep von 20 Hz bis 10 kHZ über den Lautsprecher des Typs K+H M51, der Firma Klein+Hummel wiedergegeben. Aus dem Frequenzgang des Klein+Hummel Lautsprechers (Abbildung 3.9) wird ersichtlich, dass dieser Frequenzen unterhalb von 100 Hz mit abgeschwächtem Pegel wiedergibt. Aus diesem Grund wurde der Subwoofer benutzt, der für einen Frequenzbereich zwischen 20 Hz und 200 Hz abgestimmt ist.



Abbildung 3.9:Frequenzgang des Lautsprechers Klein+Hummel, M 51 aus K+H [28]

Für das zeitgleiche Abspielen und Aufzeichnen wird die Software "Monkey-Forest' und eine Soundkarte des Typs Hammerfall von RME verwendet. Die Signale aus beiden Lautsprechern werden mit einem Kunstkopf (MK1) der Firma Cortex nacheinander aufgenommen, zur Soundkarte übertragen und anschließend mit "Monkey-Forest' zwischen 100 Hz und 200 Hz zu einer Messkurve zusammengefügt.

Die Abbildungen 3.10 bis 3.12 zeigen den Messaufbau und die verwendeten Absorptionsmaterialien. Wie auf den Photos zu erkennen ist, handelt es sich bei den Materialien um Schaumstoffplatten mit einer Stärke von 3,5, bzw. 8 cm und Glasfasermatten mit einer Stärke von 10 cm. Die zumeist in den Ecken platzierten Rollwägen bestehen ebenfalls aus Glasfasermatten derselben Stärke, sind jedoch rundum geschlossen. In der Abbildung 3.12 ist der maximal bedämpfte Raumzustand dargestellt, wobei folgende Abkürzungen verwendet wurden: QP = Quellpunkt, HP = Hörposition, RW = Rollwagen, SS = Schaumstoff, MF = Mineralfaser und PS = Plattenschwinger. Durch unterschiedliche Anordnung dieser Absorber können nun jene Raumzustände realisiert werden, welche die eingangs erwähnten Kriterien erfüllen (siehe Kapitel 3.2, S.28).



Abbildung 3.10, 3.11: Anordnung von Absorbern und Messaufbau im Prüfraum



Abbildung 3.12: Skizze des Messraums mit Absorberanordnung bei maximaler Bedämpfung. QP=Quellposition, HP=Hörposition, S =Schaumstoff, PS=Plattenschwinger, MF=Mineralfaser, RW=Rollwagen

3.2.4 Messung der Impulsantworten

Mit einem Kugelmikrophon (Brüel & Kjaer) werden je sieben Messpunkte bei zwei verschiedenen Quellpositionen ausgewertet (siehe Abb.3.6). Die Ergebnisse dieser Messungen sind für verschiedene Raumzustände in den Abbildungen 3.13- 3.16 als senkrechter Balken dargestellt. Sie zeigen die Schwankungsbreite, sowie den Mittelwert innerhalb eines Frequenzbandes an. Auf der Grundlage dieser Ergebnisse wird nun ein Messpunkt ausgewählt (Messpunkt Nr. 4, Abb. 3.6), an dem die Impulsantworten mittels Kunstkopf (Cortex) gemessen werden. Diese Messergebnisse sind ebenfalls in den Abbildungen 3.13- 3.16 als rote Linie dargestellt.

Nachhallzeit Zustand 1 2,2 2,0 **Nachhallzeit in Sekunden** 1,8 1,6 1,4 1,2 1,0 0,8 0,6 0,4 0,2 0,0 40 100 160 250 400 630 1000 1600 2500 4000 6300 63 Frequenz in Hz

Abbildung 3.13: Nachhallzeitverlauf des Raumzustands 1, Balken = Schwankungsbreite aller Messpunkte, rot = Messkurve Punkt Nr. 4



Nachhallzeit Zustand 2

Abbildung 3.14: Nachhallzeitverlauf des Raumzustands 2, Balken = Schwankungsbreite aller Messpunkte, rot = Messkurve Punkt Nr. 4

Nachhallzeit Zustand 3



Abbildung 3.15: Nachhallzeitverlauf des Raumzustands 3, Balken = Schwankungsbreite aller Messpunkte, rot = Messkurve Punkt Nr. 4

Nachhallzeit Zustand 4



Abbildung 3.16: Nachhallzeitverlauf des Raumzustands 4, Balken = Schwankungsbreite aller Messpunkte, rot = Messkurve Punkt Nr. 4

Bei der Auskleidung des Raumes mit Absorbermaterial wurde besonders darauf geachtet, die Nachhallzeit der höheren Frequenzen möglichst nicht zu verändern, damit bei den Hörtests der Einfluss der tieffrequenten Bereiche isoliert betrachtet werden kann. Wie aus den Abbildungen hervorgeht, variieren die Frequenzen oberhalb von 1 kHz in einem Bereich von einer Zehntelsekunde. Auffällig ist hingegen die große Schwankungsbreite im tieffrequenten Bereich. Die Ergebnisse hängen hier stark von der Messposition ab und unterscheiden sich teilweise um mehr als eine Sekunde. Für einen direkten Vergleich der Impulsantworten untereinander sind ihre Nachhallzeiten im tieffrequenten Bereich (bis 800 Hz) in Abbildung 3.17 zusammen dargestellt.



Nachhallzeiten verschiedener Raumzustände

Abbildung 3.17: Vergleich der Nachhallzeiten für Frequenzen bis 800 Hz bei Messpunkt Nr. 4 aus Abbildung 3.6.

3.2.5 Erzeugung des Audiomaterials für den Hörtest

Aufgrund von Störgeräuschen, die in dem Ergebnis aus der Faltung der Impulsantworten mit dem Sprachmaterial auftreten, wird das Audiomaterial mit einer anderen Methode hergestellt. Anstelle des 'Frequenzsweeps', der für die Erzeugung der Impulsantwort benutzt wird, wird das Sprachsignal direkt in den Raum gespielt und über Kunstkopf aufgenommen. Die Anordnung der Absorber ist identisch zu der, bei den Impulsantwortmessungen. Statt der zwei ursprünglich verwendeten Lautsprecher, wird nun auf einen Dodekaeder-Lautsprecher zurückgegriffen, der die anschließende Überlagerung von zwei Audiosignalen hinfällig macht. Das Sprachmaterial wird zuvor mit dem invertierten Frequenzgang des Lautsprechers multipliziert, um eine nahezu lineare Übertragung des Signals, zwischen 50 Hz und 10 kHz zu erhalten. Die Abbildung 3.18 zeigt die Frequenzspektren des Originalsignals (grün) und des bearbeiteten Signals (blau).



Abbildung 3.18: Frequenzspektrum des Audiomaterials, Original (grün) und bearbeitet (blau)

Der an der Hörposition (Messpunkt Nr.4) verwendete Kunstkopf vom Typ HMS III, der Firma Head Acoustics ermöglicht verschiedene Aufnahmeeinstellungen und eine anschließende Nachbearbeitung, näheres dazu siehe Head Acoustics [29]. Die Abbildung 3.19 zeigt den neuen Messaufbau.



Abbildung 3.19: Messaufbau für direkte Aufnahme des Sprachsignals, links Dodekaeder-Lautsprecher, rechts Kunstkopf HMS III

4. Darbietung und Auswertung von Hörmaterial

In diesem Kapitel werden die beiden Testverfahren und die angewandten Auswerteverfahren vorgestellt. Zur Messung der Sprachverständlichkeit wird der Reimtest ausgewählt, dessen Ergebnisse mittels Regressionsanalyse (Abschnitt 4.1.3) ausgewertet werden. Die Daten der Paarvergleiche werden mittels BTL-Skalierung (Abschnitt 4.2.2) veranschaulicht. Dieses Verfahren ermöglicht es, die gewonnenen Daten auf einer Differenzenskala darzustellen und sie so bei verschiedenen Hörpositionen miteinander vergleichen zu können.

4.1 Reimtest

4.1.1 Auswahl des Testverfahrens

Zur Bestimmung eines geeigneten Sprachtests wurden folgende Aspekte berücksichtigt. Der Test muss für Normalhörende geeignet sein. Die Dauer eines Testlaufes soll einen zeitlichen Rahmen von zehn Minuten nicht überschreiten und die verwendeten Testwörter dürfen nicht aus dem Kontext zu erschließen sein.

Bei dem sog. 'Reimtest' nach v. Wallenberg & Kollmeier (1989), gesprochen von Sotschek sind diese Kriterien erfüllt. Er dient zur Messung der Sprachverständlichkeit (in Prozent) und kann in Ruhe oder im Störgeräusch durchgeführt werden. Der Test wird dabei als Reimtest in einer geschlossenen Form durchgeführt, d. h. mit einer Anzeige von fünf sich reimenden (bzw. ähnlichen) Antwortalternativen nach jeder Darbietung eines Wortes. Die Antworten werden durch die Testperson selbst gegeben (nach der Einweisung in der Regel ohne weitere Betreuung durch eine Aufsichtsperson), siehe auch Hörtech [30]. Ein Auszug aus dem Reimtest befindet sich in Form von Wave-Dateien in Anhang A.

4.1.2 Testablauf

Den Versuchspersonen (Vpn) wurden die zu erkennenden Worte, über einen Kopfhörer des Typs Beyerdynamic HD 8 vorgespielt. Der Frequenzgang des Kopfhörers wurde bei der Erstellung der Impulsantworten in dem Programm ,Catt-Acoustic' berücksichtigt. Vor jedem Testwort ertönt zunächst der Satz: "Bitte markieren Sie das Wort…", um den Raum anzuregen und die Vpn vorzubereiten. Das Audiomaterial umfasst 100 Worte und ist in zehn Durchgänge zu je zehn Wörtern unterteilt. Nach jedem Durchgang findet ein Wechsel der Impulsantwort statt. Die Vpn wurden instruiert, das von Ihnen gehörte, einsilbige Wort auf der Auswahlliste zu markieren. Aus der Fehlerquote wurde anschließend ermittelt, inwieweit die Sprachverständlichkeit mit der verwendeten Impulsantwort korreliert.

Die ersten Vorversuche zeigen, dass das Sprachmaterial zusammen mit einem Störgeräusch abgespielt werden muss, da sonst die Fehlerquote, unabhängig von der gewählten Impulsantwort gegen Null geht. Aus der Arbeit von Rau [31] geht hervor, dass der Einfluss von tieffrequenten Störgeräuschen (< 250 Hz) auf die Sprachverständlichkeit vernachlässigbar ist. Deshalb handelt es sich bei dem hier verwendeten Störgeräusch um sprachsimulierendes Rauschen, das aus sämtlichen Zielwörtern des Reimtests durch zeitversetztes Überlagern erzeugt wird.

Es werden zwei Testreihen (A und B) mit jeweils zehn Personen durchgeführt, wobei sich diese innerhalb der Testreihen teilweise überschneiden. Da aber zwischen den beiden Zeitpunkten der Durchführung ein großer Abstand besteht, sind die Antworten, der an beiden Testreihen teilnehmenden Vpn, als voneinander unabhängig zu betrachten. Bei der Testreihe A wird nur das Sprachmaterial mit den Impulsantworten gefaltet, bei der Testreihe B auch das Störsignal. Die Reihenfolge der dargebotenen Impulsantworten wird innerhalb einer Testreihe variiert, um eine Fehlerhäufung bei einem Zustand, auf Grund von Gewöhnungs- oder Ermüdungseffekten ausschließen zu können. In Tabelle 4.1 ist die Anordnung der Impulsantworten für die Vpn aufgeführt.

Durchgang	Vp 1/6	Vp 2/7	Vp 3/8	Vp 4/9	Vp 5/10
1	Zstd. 0	Zstd. 1	Zstd. 2	Zstd. 3	Zstd. 4
2	Zstd. 1	Zstd. 2	Zstd. 3	Zstd. 4	Zstd. 0
3	Zstd. 2	Zstd. 3	Zstd. 4	Zstd. 0	Zstd. 1
4	Zstd. 3	Zstd. 4	Zstd. 0	Zstd. 1	Zstd. 2
5	Zstd. 4	Zstd. 0	Zstd. 1	Zstd. 2	Zstd. 3
6	Zstd. 0	Zstd. 1	Zstd. 2	Zstd. 3	Zstd. 4
7	Zstd. 1	Zstd. 2	Zstd. 3	Zstd. 4	Zstd. 0
8	Zstd. 2	Zstd. 3	Zstd. 4	Zstd. 0	Zstd. 1
9	Zstd. 3	Zstd. 4	Zstd. 0	Zstd. 1	Zstd. 2
10	Zstd. 4	Zstd. 0	Zstd. 1	Zstd. 2	Zstd. 3

Tabelle 4.1: Reihenfolge der Impulsantworten Zstd. 0 bis 4 aus der Simulation

Durch diese Anordnung wird jede Impulsantwort zwei Mal innerhalb eines Tests bewertet. Der Einfluss von einzelnen Wörtern auf die Messergebnisse, die aufgrund ihrer Endung besonders schlecht, oder gut verstanden werden, soll dadurch ebenfalls eingeschränkt werden.

Die Pegeldifferenz zwischen Sprach- und Störsignal beträgt bei Testreihe A 8dB und bei Testreihe B 4dB. Das Störsignal wurde jeweils so laut eingestellt, dass ein richtiges Erkennen der Worte möglich, aber deutlich erschwert wurde und sich dadurch die Fehlerquote erhöhte.

4.1.3 Auswertung mittels Regressionsanalyse

Die Regressionsanalyse ist ein statistisches Verfahren zur Analyse von Daten und geht von der Aufgabenstellung aus, 'einseitige' statistische Abhängigkeiten (d.h. statistische Ursache-Wirkung-Beziehungen) durch so genannte 'Regressionsfunktionen' zu beschreiben. Vermutet man einen zumindest näherungsweise linearen Zusammenhang zwischen zwei Merkmalen in einer Grundgesamtheit, so kann mittels linearer Regressionsrechnung dieser Zusammenhang näher spezifiziert und untersucht werden. Dazu werden die beiden interessierenden Merkmale X und Y an n Objekten aus der Grundgesamtheit beobachtet. Die 2n Realisationen $x_1,..., x_n$ und $y_1,..., y_n$ werden dann zur Untersuchung der linearen Beziehung verwendet. Man geht davon aus, dass gilt:

$$y_i = \alpha + \beta x_i + e_i$$
 für i = 1,..., n

wobei α das Absolutglied und β den Steigungsparameter der linearen Beziehung darstellen und e₁,..., e_n zufällige Fehler sind, aus Hartung [32]. Nähere Ausführungen zu Methoden der Punktschätzung und Tests für die Parameter α und β und die Analyse der zufälligen Fehler e_i sind der entsprechenden Literatur zu entnehmen, z.B. Hartung [32] oder Cochran [35].

4.2 Paarvergleiche

Paarvergleiche dienen zur Ermittlung einer Rangfolge von Merkmalen. Durch den direkten Vergleich von jeweils zwei Merkmalen wird ein jedes mit jedem anderen verglichen und als wichtiger, oder weniger wichtig beurteilt. Durch die Summation der einzelnen Gewichtungen ergibt sich eine Gesamtrangfolge aller Merkmale.

4.2.1 Versuchsablauf

Im Rahmen dieser Arbeit wurden drei Testreihen im einseitigen Paarvergleich durchgeführt. Einseitig bedeutet, dass ein Paar nur einmal abgefragt wird, unabhängig von der dargebotenen Reihenfolge. Die Paarvergleiche wurden mit Hilfe einer Software (PAK) von Müller-BBM durchgeführt. Sie ermöglicht das Einlesen des Audiomaterials und erzeugt aus den Beispielen eine Sequenz für den einseitigen Paarvergleich. Die von den Vpn zu beantwortenden Fragen, sowie die Antwortmöglichkeiten können vorformuliert werden und erscheinen automatisch nach dem Abspielen eines Paares auf dem Bildschirm, siehe Abbildung 4.1. Die Antworten werden anschließend tabellarisch gespeichert.

Abmelden Frage 1 von 1 Bei welchem Beispiel ist die Sprachverständlichkeit besser ? 1 2	PAK-Jury Pre 1 Image: Constant State System Antworten Ansicht Einstellungen Hilfe Name Testperson (beschäftigt)	 ✓ Bildanzeige ✓ Textanzeige Sequenz SeqNr.: 1 ✓ B Wiederholen Weiter
	Abmelden Frage 1 von 1 Bei welchem Beispie besser ? 1 2	el ist die Sprachverständlichkeit

Abbildung 4.1: Grafische Oberfläche bei der Durchführung des Paarvergleiches mit PAK

Für die Wiedergabe der simulierten Raummoden (Kap. 3.1.2) und der Kunstkopfaufnahmen (Kap. 3.2.5) wurde ein System von Head Acoustics verwendet, bestehend aus Equalizer (PEQ IV), Verstärker (PVA IV.2) und elektrostatischem Kopfhörer. Dadurch kann der Einfluss des Gehörgangs des Kunstkopfes während des Abspielens herausgerechnet werden, näheres dazu siehe Head Acoustics [29].

Bei den Vorversuchen der Kunstkopfaufnahmen zeigte sich, dass hohe Frequenzanteile überbetont waren und ein Vergleich, der auf unterschiedlichen Nachhallzeiten im tieffrequenten Bereich beruht, dadurch erschwert würde. Aus diesem Grund wurden die Sprachsignale nochmals gefiltert und der Frequenzbereich zwischen 10 kHz und 17 kHz im Pegel verringert, siehe Abbildung 4.2.



Abbildung 4.2: Frequenzanteile Sprachsignal original grün, gefiltert blau

Bei dem Sprachsignal handelte es sich um einen Textausschnitt aus einem Mathematikvortrag (Anhang A). Für alle zu vergleichenden Beispiele wurde derselbe Satz verwendet. Es sollten Unterschiede bezüglich der Sprachverständlichkeit der einzelnen Aufnahmen beurteilt werden. Auf die Frage: "Bei welchem Beispiel ist die Sprachverständlichkeit besser?" musste zwischen den Antworten: "Beispiel 1" oder "Beispiel 2" gewählt werden.

Bei dem Paarvergleich mit künstlich erzeugten Impulsantworten wurde ebenfalls das oben genannte Sprachmaterial verwendet. Die Wiedergabe erfolgte über die interne Soundkarte des Computers und einen Kopfhörer des Typs DT 990 von Beyerdynamic. Für diesen Kopfhörertyp fand die Entzerrung des Audiosignals im Simulationsprogramm ,Catt-Acoustic' statt.

4.2.2 Auswertung

Für die Auswertung der Daten aus den Paarvergleichen wurde ein Verfahren benutzt, das unter dem Namen ,BTL-Modell' (Bradley-Terry-Luce) bekannt ist. Der folgende Abschnitt erläutert diese Methode, die speziell für die Auswertung von Paarvergleichen entwickelt worden ist.

4.2.2.1 Das Bradley-Terry-Luce Modell (BTL – Modell)

Bei dem hier vorgestellten Modell geht man davon aus, dass es eine absolute Dominanz-Relation der Reize gibt, die unabhängig vom Beobachter existiert. Die Datenerhebung erfolgt über den vollständigen Paarvergleich: Es werden jedem Beurteiler alle Reizpaare zur Beurteilung vorgelegt und er muss angeben, ob ein Reiz S_i dominanter ist, als ein Reiz S_j . Die Relation wird durch $S_i > S_j$ ausgedrückt. Bei N Befragten kann man eine sog. Dominanzmatrix aufstellen, in der die Häufigkeiten aller Fälle gespeichert wird, in denen S_i dominanter als S_j gewählt wurde. Durch Prozentuieren (teilen aller Häufigkeiten durch N) erhält man die zur Dominanzmatrix zugehörige Prozentmatrix. Ziel dieser Skalierungs-Methode ist es, den Reizen Skalenwerte so zuzuordnen, dass die sich aus den Skalenwerten ergebene Prozentmatrix so gut wie möglich mit der empirisch erhobenen Prozentmatrix übereinstimmt, siehe Gediga [33].

Wahlaxiom von Luce

Folgendes wird angenommen: Aus einer gegebenen Reizmenge $A=\{S_1, S_2, ..., S_n\}$ besteht für jeden Reiz eine feste Wahrscheinlichkeit $p(S_i)$ bei einer Präferenzwahl gewählt zu werden. Die Wahrscheinlichkeit, die Alternative S_i aus einer beliebigen Menge von T Alternativen auszuwählen, lässt sich berechnen durch:

$$p(S_i | T) = \frac{p(S_i)}{p(T)}$$
(4.5)

Hierbei bezeichnet:

 $p(S_i)$: Die Wahrscheinlichkeit der Wahl von S_i .

 $p(S_i|T)$: Die Wahrscheinlichkeit der Wahl von S_i , wenn die Teilmenge T zur Auswahl steht. p(T): Die Wahrscheinlichkeit, dass die Wahl auf ein Element von T fällt.

Herleitung der Skalenwerte aus dem Wahlaxiom

 $p(S_i > S_j)$ ist die Wahrscheinlichkeit, dass beim Paarvergleich von S_i und S_j der Reiz S_i bevorzugt wird. Nach dem Wahlaxiom aus Gleichung (4.5) gilt mit $T = \{S_i, S_j\}$

$$p(S_i \succ S_j) = p(S_i \mid T = \{S_i, S_j\}) = \frac{p(S_i)}{p(\{S_i, S_j\})}$$
(4.6)

$$p(S_j \succ S_i) = p(S_j \mid T = \{S_i, S_j\}) = \frac{p(S_j)}{p(\{S_i, S_j\})}$$
(4.7)

mit $p(S_j \succ S_i) = 1 - p(S_i \succ S_j)$ ergibt sich das Risiko R_{ij} zu

$$R_{ij} = \frac{p(S_i \succ S_j)}{1 - p(S_i \succ S_j)} = \frac{p(S_i)}{p(S_j)} = \frac{u_i}{u_j}$$
(4.8)

D.h., dass das Risiko zu jeder Wahlwahrscheinlichkeit $p(S_i > S_j)$ abhängig vom Quotienten zweier zugeordneter Zahlen u_i , u_j ist. Die Wahrscheinlichkeiten $p(S_i > S_j)$, $p(S_j > S_i)$ lassen sich aus dem Paarvergleichsexperiment empirisch bestimmen. Mit den Skalenwerten u_i , u_j , lässt sich eine Verhältnisskala bilden. Durch Logarithmieren $\sigma_i = log (u_i)^2$ wird die grafische Darstellung der Skalenwerte auf einer Dimension als Differenzskala ermöglicht.

BTL-Modell und logistische Regression

Das BTL-Modell beschreibt die Wahrscheinlichkeit, dass ein Reiz S_i einem anderen Reiz S_j bevorzugt wird, durch das Modell

$$p(S_i > S_j) = \frac{u_i}{u_i + u_j} = \frac{\exp(\sigma_i)}{\exp(\sigma_i) + \exp(\sigma_j)}$$
$$p(S_i > S_j) = \frac{\exp(\sigma_i - \sigma_j)}{1 + \exp(\sigma_i - \sigma_j)}$$
(4.9)

Die Gleichung (4.9) wird auch als Grundgleichung des BTL-Modells bezeichnet. Aufgabe für die Parameterschätzung ist es, aus den empirisch gefundenen Wahlwahrscheinlichkeiten die Skalenwerte σ_i auszurechnen, so dass die rückgerechneten Wahlwahrscheinlichkeiten aus Gleichung (4.9) möglichst gut mit den empirischen Werten übereinstimmen. Nachdem in allen Gleichungen immer Differenzen von Skalenwerten eingehen, kann man einen Skalenwert beliebig festlegen.

Logarithmiert man die Gleichung (4.9) lässt sie sich auch als logit-Modell schreiben:

$$logit (p(S_i > S_j)) = \sigma_i - \sigma_j$$
.

Weitere Erläuterungen zum logit-Modell, siehe Hartung [32]. Zusammenfassend kann man jede Differenz $\sigma_i - \sigma_j$ als Differenz von Differenzen schreiben: $\sigma_i - \sigma_j = (\sigma_i - \sigma_k) - (\sigma_j - \sigma_k)$. Setzt man einen Reiz als Anker und vergleicht die anderen Reize mit diesem, entsteht eine Datenmatrix mit der sich eine logistische Regression durchführen lässt, siehe Gediga [33].

4.2.2.2 Urteilskonsistenz

Paarvergleichsurteile können bezüglich ihrer Konsistenz, d.h. Widerspruchsfreiheit überprüft werden. Bevorzugt eine Person bei einem Paarvergleich ein Objekt A vor B (A > B) und B vor C (B > C), so gilt bei konsistentem Präferenzurteil auch A > C. Wenn die Präferenzen hingegen intransitiv und damit die Vergleichsurteile inkonsistent sind, ergeben sich sog. ,Zirkularitäten', bzw. für den Fall mit drei Objekten ,zirkuläre Triaden'(siehe Tabelle 4.2). Die Anzahl der zirkulären Triaden aus einer Menge mit n Objekten stellt ein Maß für die Inkonsistenz der Paarvergleichsurteile dar, siehe auch Bortz, Lienert, Boehnke [34]. Überschreitet ihre relative Häufigkeit einen bestimmten Wert, werden die Urteile der Vp für die weitere Auswertung ausgeschlossen. Anhand des folgenden Beispiels werden die bei der Berechnung verwendeten Formeln erläutert:

n = 6	А	В	С	D	Е	F	D_i
А		-	-	-	-	-	0
В	+		-	-	-	-	1
С	+	+		+	-	+	4
D	+	+	-		-	-	2
Е	+	+	+	+		-	4
F	+	+	-	+	+		4

Tabelle 4.2: Paarvergleichsurteile, bei (+) wird Zeilenobjekt vor Spaltenobjekt bevorzugt

Die Tabelle 4.2 zeigt die Beurteilungen einer Versuchsperson zu einem Hörtests mit 15 Paarvergleichsurteilen. Ein Plus wurde signiert, wenn ein Zeilenobjekt einem Objekt aus der Spalte vorgezogen wurde. Die Zeilensummen der Plussignaturen sind mit D_i bezeichnet und in der letzten Spalte angegeben.

Nach Gediga [33] lässt sich die Zahl der zirkulären Triaden angeben durch

$$d = \frac{n(n-1)(2n-1)}{12} - \frac{1}{2} \sum_{i=1}^{n} D_i^2 .$$
(4.1)

Die Anzahl der maximal möglichen zirkulären Triaden berechnet sich aus folgenden Gleichungen:

für n ungeradzahlig:
$$d_{\text{max}} = \frac{n(n^2 - 1)}{24}$$
 (4.2)

für n geradzahlig:
$$d_{\text{max}} = \frac{n(n^2 - 4)}{24}$$
 (4.3)

Der Konsistenzkoeffizient errechnet sich aus der relativen Häufigkeit der zirkulären Triaden gemäß:

$$T = 1 - \frac{d}{d_{\max}} \quad . \tag{4.4}$$

Für das Beispiel ergeben sich aus den Gleichungen 4.1 – 4.4 folgende Werte:

$$d = \frac{6(6-1)(2 \cdot 6 - 1)}{12} - \frac{1}{2} \left(0^2 + 1^2 + 4^2 + 2^2 + 4^2 + 4^2 \right) = 1$$
$$d_{\max} = \frac{6(6^2 - 4)}{24} = 8$$
$$T = 1 - \frac{1}{8} = 0,875$$

Ist $T \ge 0.85$ spricht man von konsistenter Beurteilung, was in diesem Fall bedeutet, dass das Beispiel das Konsistenzkriterium erfüllt und für die Analyse verwendet werden kann.

5. Ergebnisse des Reimtests

Nach Erläuterung des ausgewählten Testverfahrens und dessen Durchführung, werden in diesem Kapitel die Ergebnisse des Reimtests präsentiert.

5.1 Ergebnisse der Testreihen A und B

Tabelle 5.1 und 5.2 zeigen die Ergebnisse aus dem Reimtest. In der zweiten Zeile ist der Erkennungs-Score dargestellt, der die Anzahl der richtig erkannten Wörter einer Vp angibt (maximal 100). Die weiteren Zeilen zeigen die prozentuale Verteilung des jeweiligen Erkennungsscores auf die Raumzustände 0 bis 4. In der letzten Spalte ist der Mittelwert der entsprechenden Zeile angegeben.

	1	1	1	1	1	1	1	1	1	1	1
Testreihe A	VP 1	VP 2	VP 3	VP 4	VP 5	VP 6	VP 7	VP 8	VP 9	VP 10	Mittelwert
Erkennungs –Score	75	76	71	70	63	80	77	64	72	69	72
Zstd. 0	0,20	0,21	0,21	0,19	0,17	0,21	0,23	0,19	0,19	0,19	0,20
Zstd. 1	0,21	0,24	0,15	0,20	0,17	0,20	0,22	0,25	0,21	0,22	0,21
Zstd. 2	0,17	0,16	0,23	0,20	0,19	0,16	0,19	0,16	0,24	0,17	0,19
Zstd. 3	0,20	0,22	0,17	0,20	0,25	0,21	0,17	0,25	0,21	0,20	0,21
Zstd. 4	0,21	0,17	0,24	0,21	0,21	0,21	0,18	0,16	0,15	0,22	0,20

 Tabelle 5.1: Erkennungs-Score der Versuchspersonen (Vp) und Zusammensetzung des Erkennungsscores nach Raumzustand (=erreichter Score bei Zustand / 100), Testreihe A.

	1	1	1	1	1	1		1	1	1	1
Testreihe B	VP 1	VP 2	VP 3	VP 4	VP 5	VP 6	VP 7	VP 8	VP 9	VP 10	Mittelwert
Erkennungs –Score	85	88	86	78	80	77	83	78	85	90	83
Zstd. 0	0,18	0,23	0,21	0,22	0,19	0,17	0,24	0,17	0,21	0,19	0,20
Zstd. 1	0,19	0,22	0,2	0,17	0,21	0,14	0,19	0,26	0,21	0,21	0,20
Zstd. 2	0,20	0,18	0,21	0,22	0,19	0,22	0,19	0,22	0,20	0,18	0,20
Zstd. 3	0,21	0,19	0,20	0,19	0,24	0,22	0,22	0,19	0,19	0,21	0,21
Zstd. 4	0,22	0,18	0,23	0,21	0,18	0,25	0,16	0,17	0,19	0,21	0,20

 Tabelle 5.2: Erkennungs-Score der Versuchspersonen (Vp) und Zusammensetzung des Erkennungsscores nach Raumzustand (=erreichter Score bei Zustand / 100), Testreihe B.

In den Abbildungen 5.1 und 5.2 sind die Werte der Tabellen 5.1 und 5.2 grafisch dargestellt. Der Mittelwert des Erkennungsscores liegt für beide Testreihen A und B bei allen Raumzuständen in einem Bereich zwischen 19 und 21%.





Abbildung5.1: Erkennungsscore Testreihe A



Bei der statistischen Auswertung des Reimtests wurden folgende Zusammenhänge analysiert. Zum einen die Korrelation zwischen dem Erkennungsscore und der Nachhallzeit des 125 Hz-, sowie des 250 Hz-Bandes. Außerdem fand eine Auswertung der Korrelation zwischen dem Erkennungsscore und der Durchgangszahl, sowie deren Quadrat statt. Ein Durchgang besteht dabei aus zehn Testwörtern und markiert den Wechsel der Impulsantwort. Die Durchgangszahl gibt an, in welchem Abschnitt des Hörtests sich eine Versuchsperson befindet (siehe Tabelle 4.1, S. 38).

Das Ergebnis der linearen Regression für den Erkennungs-Score, in Abhängigkeit von der Nachhallzeit des 125 Hz Bandes und des 250 Hz Bandes adjustiert danach, um welchen Durchgang es sich handelt, zeigt für die Variable Durchgang eine hochsignifikante positive Assoziation (Koeff.=1,146; p<0,0004) und für die Variable Durchgang zum Quadrat eine ebenfalls hochsignifikante, nun aber negative Assoziation (Koeff.=-0,074; p=0,001) (siehe Tabelle 5.3 - 5.5).

Der Koeffizient gibt dabei an, wie sich der Erkennungs-Score verändert, wenn sich die Bezugsvariable um eine Einheit (Nachhallzeit in Sekunden, oder Durchgangszahl) erhöht.

Betrachtet man nun die Koeffizienten aus Tabelle 5.3 der beiden Frequenzbänder (Koeff.= -0,531 für 125 Hz-Band und Koeff.= 0,568 für 250 Hz-Band) so bedeutet dies, dass sich der Erkennungsscore um eine Einheit erhöht, wenn die Nachhallzeit im 125 Hz-Band um 0,53 Sekunden abnimmt, bzw. um 0,57 Sekunden im 250 Hz-Band zunimmt.

Der p-Wert drückt die Wahrscheinlichkeit aus, dass ein solcher oder extremerer Wert unter der Annahme, dass der wahre Wert gleich Null ist, beobachtet wird. Das Signifikanzniveau gibt an, bis zu welchem p-Wert die Nullhypothese (H0) verworfen wird, z.B.:

H0: Durchgangszahl hat keinen Einfluss auf Erkennungsscore \rightarrow Koeff. = 0.

H1: Durchgangszahl hat einen Einfluss (Richtung egal) \rightarrow Koeff. $\neq 0$.

Koeffizient beobachtet: 1,146

p-Wert: 0,0004

Der beobachtete Koeffizient von 1,146 für die Variable Durchgang ist nur mit einer Wahrscheinlichkeit von 4% durch Zufall zu erklären. Dieser Wert liegt unter dem gewählten Signifikanzniveau von 5%, daher wird die Nullhypothese, dass die Durchgangszahl keinen Einfluss auf den Erkennungsscore hat, auf diesem Niveau verworfen. Man kann sagen, dass der beobachtete Koeffizient systematisch, bzw. signifikant von Null verschieden ist.

Das Konfidenzintervall gibt den Wertebereich an, in welchem bei einer Wiederholung des Versuchs, der Koeffizient in 95% aller Fälle liegen wird. Schließt das Konfidenzintervall die Null mit ein, dann bedeutet das, dass der Koeffizient nicht signifikant (p>0,05) von Null verschieden ist.

Für die Veränderung der Nachhallzeit in den beiden berücksichtigten Frequenzbändern lässt sich in diesem Modell keine signifikante Assoziation (p=0,44 und p=0,40) mit dem von den Vpn erreichten Erkennungs-Score beobachten.

			95% Ko	nfidenzintervall
	Koeffizient	p-Wert	untere	obere Grenze
Fq.band 125 Hz	-0,531	0,466	-1,972	0,909
Fq.band 250 Hz	0,568	0,435	-0,871	2,006
Durchgang	0,329	0,000	0,219	0,439
Konstante	5,590	0,000	4,218	6,962

Tabelle 5.3: Ergebnis der mulitvariablen linearen Regression für Erkennungsscore in Abhängigkeit von Nachhallzeit auf zwei Frequenzbändern adjustiert nach Durchgang, Testreihe A.

			95% Konfidenzinterval			
	Koeffizient	p-Wert	untere	obere Grenze		
Fq.band 125 Hz	-0,531	0,440	-1,891	0,828		
Fq.band 250 Hz	0,568	0,408	-0,789	1,924		
Durchgang	1,146	0,000	0,683	1,609		
Durchgang ²	-0,074	0,001	-0,115	-0,332		
Konstante	3,956	0,000	2,379	5,534		

Tabelle 5.4: Ergebnis der mulitvariablen linearen Regression für Erkennungsscore in Abhängigkeit von Nachhallzeit auf zwei Frequenzbändern adjustiert nach Durchgang², Testreihe A.

			95% Konfidenzintervall		
	Koeffizient	p-Wert	untere	obere Grenze	
Fq.band 125 Hz	0,080	0,900	-1,173	1,333	
Fq.band 250 Hz	-0,381	0,547	-1,632	0,870	
Durchgang	0,822	0,000	0,396	1,249	
Durchgang ²	-0,055	0,005	-0,093	-0,017	
Konstante	6,159	0,000	4,704	7,613	

Tabelle 5.5: Ergebnis der mulitvariablen linearen Regression für Erkennungsscore in Abhängigkeit vo
Nachhallzeit auf zwei Frequenzbändern adjustiert nach Durchgang ² , Testreihe B.

Wie in Abschnitt 4.1.2 bereits erläutert, gliedert sich eine Testreihe in zehn Durchgänge zu je zehn Testwörtern, wobei mit jedem Durchgang ein Wechsel der Impulsantwort einhergeht. Der Erkennungs-Score gibt die Anzahl der richtig erkannten Wörter innerhalb eines Durchgangs an, und kann somit Werte zwischen 0 und 10 annehmen.

In den nachstehenden Abbildungen 5.3 und 5.4 ist der Zusammenhang von erreichtem Erkennungsscore (Ordinate) und Durchgang (Abszisse) dargestellt. Die blauen, offenen Kreise stehen für den Erkennungs-Score der Vpn. Zusätzlich zeigen die roten offenen Kreise, die auf der Grundlage des multivariablen linearen Regressionsmodells vorhergesagten Werte. Es zeigt sich ein Maximum beim erreichten Score im achten Durchgang mit nachfolgendem leichten Rückgang.



Abbildung 5.3: Erkennungs-Score über Anzahl der Durchgänge der Testreihe A.



Abbildung 5.4: Erkennungs-Score über Anzahl der Durchgänge der Testreihe B.

Bei der Testreihe B fällt der Erkennungs-Score um etwa zehn Punkte höher aus, als bei Testreihe A. Jene Vpn, die an beiden Testreihen teilnehmen, beschreiben Testreihe B als ,leichter', oder ,besser zu verstehen'. Das lässt sich einerseits auf die Faltungsoperation mit der Impulsantwort zurückführen, die bei dieser Testreihe auf Sprach- und Rauschsignal angewendet wurde, andererseits auf das Lautstärkeverhältnis von Sprach- zu Rauschsignal, das bei den Testreihen A und B unterschiedlich festgelegt wurde, siehe dazu Kapitel 4, S. 38.

Der in Abbildung 5.3 und 5.4 gezeigte Verlauf bestätigt die Vermutung eines Lerneffektes bei den Vpn. Der Koeffizient (Tabelle 5.4) gibt an, dass eine Steigerung des Erkennungs-Scores von 1,146 Punkten pro Durchgang stattfindet. Ebenso lässt sich ein, wenngleich auch schwacher Ermüdungseffekt ab dem achten Durchgang feststellen.

Bezüglich der Nachhallzeit kann keine Korrelation mit dem Erkennungs-Score abgeleitet werden, was durch den p-Wert und das Konfidenzintervall in Tabelle 5.4 und 5.5 belegt wird. Die Möglichkeit, dass der Reimtest unter den hier gewählten Bedingungen nicht sensibel genug ist, um eine Korrelation zu messen, ist nicht auszuschließen.

5.2 Vergleich von Spektren der Testwörter

Bei der Auswertung der Testergebnisse wurde folgendes festgestellt: Einzelne Testwörter wurden, unabhängig von der verwendeten Impulsantwort, von allen Vpn falsch, bzw. richtig erkannt. Um Aufschluss über die Ursache zu erhalten, wurden die Spektren der Wörter untersucht. Bei den falsch verstandenen Wörtern ("wann' und "mull') fiel auf, dass sie auf die Konsonanten "n' und "l' enden, die der Gruppe der Sonoranten zugeordnet werden. Sonoranten sind immer stimmhaft und zeichnen sich dadurch aus, dass bei ihrer Bildung kein Geräusch entsteht. Die Spektren der Wörter "wann" und "mull" in Abbildung 5.5 erreichen oberhalb von 4 kHz nur geringe Pegel und liegen unterhalb des Rauschpegels. Testwörter und Rauschsignal stammen dabei aus der Simulation des Ausgangszustands (Zstd. 0).



Abbildung 5.5: Frequenzspektren des Rauschsignals und der Worte 'mull' und , 'wann'.

Die Wörter "buff" und "saat" wurden durchweg richtig verstanden. Die Endung dieser Wörter bilden Konsonanten aus der Gruppe der Obstruenten oder auch Geräuschlauten. Ihre Frequenzspektren sind in Abbildung 5.6 zusammen mit dem Rauschsignal dargestellt. Ihre Pegel überschreiten im hochfrequenten Bereich den des Rauschsignals.



Abbildung 5.6: Frequenzspektren des Rauschsignals und der Worte 'saat' und 'buff'.

Da beim Reimtest anhand einer vorliegenden Auswahl an Testwörtern entschieden wird, trägt das Verstehen des Schlusskonsonanten wesentlich zum richtigen Erkennen bei. Bleibt der Schluss eines Wortes unverständlich, kann das gehörte Wort nicht korrekt zugeordnet werden, was auch den beobachteten Effekt erklärt.

5.3 STI-Werte des Reimtests

Um das Ergebnis des Reimtests zu überprüfen, wurde der STI (siehe Kap. 2.2.4.9) nochmals mit dem Programm Catt-Acoustic ermittelt. Als Sprachsignal wurde diesmal das Material aus dem Reimtest verwendet. Mit Hilfe eines Schallpegelmessers (B&K 2260) wurden die Terzpegel gemessen und als Quellsignal in die Software übernommen. Das Störsignal wurde für die verschiedenen Raumsimulationen (Zstd.0 – Zstd.4) ebenfalls in Terzpegeln ermittelt und unter dem Menupunkt ,background noise' miteinbezogen. Die Signale wurden, den Vorgaben von Houtgast [24] entsprechend, verwendet. Die Ergebnisse sind der Tabelle 5.6 zu entnehmen. Der STI wurde für die Position Nr.6 (siehe Abb. 3.3, Kap. 3) berechnet, was auch der Hörposition beim Reimtest entsprach. Bei dieser Berechnung lässt sich ebenfalls keine deutliche Verbesserung des STI mit Verringerung der Nachhallzeit erkennen. Die Verbesserung des STI-Wertes von Zustand 0 auf Zustand 1 wird durch den Hörtest nicht bestätigt.

	STI ohne Rauschen	STI mit Rauschen		
Zustand 0	0,530	0,280		
Zustand 1	0,667	0,342		
Zustand 2	0,686	0,355		
Zustand 3	0,694	0,350		
Zustand 4	0,706	0,363		

Tabelle 5.6: STI aus Catt-Acoustic ohne/mit Rauschen

6. Ergebnisse des Paarvergleichs

Die Skalierung nach dem BTL – Modell (Kap. 4.2.2.1) ermöglicht es, Reize, die in Paarvergleichen miteinander verglichen werden, auf einer Differenzenskala anzuordnen. Der Abstand der Reize zueinander (die Differenz) sagt etwas darüber aus, wie ähnlich, bzw. unterschiedlich die Reize von den Versuchspersonen bewertet wurden. Die Ergebnisse aus den Vergleichen mit simulierten und realen Impulsantworten, sowie mit simulierten Raummoden, sind auf diese Weise skaliert und in den folgenden Abschnitten dargestellt.

6.1 Ergebnisse mit simulierten Impulsantworten

Die aus der Simulation (Kap. 3.1) gewonnenen Impulsantworten werden in den folgenden Abbildungen 6.1 bis 6.3, mit S0 bis S5 bezeichnet. Dabei steht S0 für die Impulsantwort des Ausgangszustands (Zustand 0), S5 für den Zustand 5. Aus dem Simulationsmodell wurden drei Positionen ausgewählt (Pos.1, 6 und 8, siehe Abbildung 3.3) und die Ergebnisse dargestellt. Bei dieser Art von Skala können die Differenzen der Skalenwerte über die Skalen hinweg miteinander verglichen werden, und erleichtern dadurch die Beurteilung des Raumes.



Abbildung 6.1: Paarvergleichsurteile simulierter Impulsantworten für Position 1

Aus der Abbildung 6.1 geht hervor, dass der Ausgangszustand (Zustand 0, bzw. S0) hinsichtlich der Sprachverständlichkeit, als schlechtester bewertet wurde. Der Abstand auf der Skala zum nächsten Reiz (S1) zeigt, dass dieser deutlich von dem Reiz S0 unterschieden werden konnte. Ordnen alle Vpn die Reize S0 bis S5 in der gleichen Reihenfolge an, so ergibt sich zwischen ihnen ein idealer Abstand von 1. Betrachtet man die Reize S4 und S3, so fällt auf, dass S3 etwas besser beurteilt wurde, als S4. Der geringe Abstand deutet aber darauf hin, dass die Vpn kaum in der Lage waren, einen Unterschied zwischen diesen Reizen wahrzunehmen. Die Graphik belegt weiters, dass die Sprachverständlichkeit bei der kürzesten Nachhallzeit des Raumes (S5), für die Hörposition 1, als am besten bewertet wurde.

Beim Paarvergleich an Position 6 zeichnet sich ein anderes Bild. Von den Vpn werden drei Bereiche klar unterschieden, wobei S0 wieder als der unverständlichste Reiz bewertet wird. S1, S2 und S3 werden besser eingestuft, liegen aber recht dicht zusammen und waren daher beim Paarvergleich einander ähnlich. Für die Vpn noch schlechter zu unterscheiden waren jedoch die Reize S4 und S5. Im direkten Vergleich mit einem Reiz aus dem Mittelfeld (S1 bis S3) wurden sie zwar klar besser bewertet, konnten aber untereinander kaum unterschieden werden. Zu dieser Bewertung trug die Tatsache bei, dass die Abnahme des Direktschallanteils gegenüber der Position 1 auf den Aufnahmen deutlich zu hören war.



Abbildung 6.2: Paarvergleichsurteile simulierter Impulsantworten für Position 6

Vergleicht man die Abbildung 6.1 mit der Abb. 6.2, so lassen sich folgende Schlüsse ziehen. Die Tatsache, dass S5 in Abb. 6.1 einen Wert von 3,5 und in Abb. 6.2 über 4 erreicht, lässt sich dahingehend interpretieren, dass für die Position 1 die Sprachverständlichkeit bei verschiedenen Raumzuständen insgesamt ähnlicher beurteilt worden ist, als für die Position 6 (vergl. Abb. 3.3, Kap. 3.1.1, S. 26). Eine Erklärung dafür liefert die Tatsache, dass der Einfluss des Direktschallanteils an Position 1 größer ist als der des diffusen Anteils, da man sich noch innerhalb des Hallradius r_H befindet. Mit zunehmendem Abstand von der Schallquelle spielt die Nachhallzeit des Raumes eine immer bedeutendere Rolle. Es wäre jedoch falsch, aufgrund der Absolutwerte von S5 zu behaupten, dass die Sprachverständlichkeit für diesen Reiz in Abb. 6.2 besser beurteilt worden ist als in Abb. 6.1. Einen Vergleich von Differenzen über die Abbildungen hinweg anzustellen, ist allerdings zulässig. Betrachtet man das Paar S3, S2, so ist in Abb. 6.1 eine große positive Differenz zu verzeichnen, in Abb. 6.2 eine kleine Differenz und in Abb. 6.3 zeigt sich eine klar negative Differenz der Skalenwerte. Der Wechsel in der Beurteilung der Reize S2, S3 steht also in unmittelbarem Zusammenhang mit der Hörposition.

BTL-Skala Position 8





Für Hörposition 8 ergibt sich wieder eine ähnliche Beurteilung der Zustände S4 und S5, sowie 1 und 3 (siehe Abbildung 6.3). Bei diesen Hörbeispielen tritt eine Verstärkung der tiefen Frequenzen aufgrund der rückwärtigen Sitzposition hervor. Bemerkenswert ist, dass trotz längerer Nachhallzeiten, der Zustand S2 gegenüber S3 bevorzugt wird.

Zusammenfassend kann gesagt werden, dass eine Bevorzugung der Sprachsignale stattfindet, die im tieffrequenten Bereich über kürzere Nachhallzeiten verfügen. In allen Paarvergleichen wird das Beispiel S0 am schlechtesten bewertet, die Raumzustände S4, bzw. S5 liefern die beste Sprachverständlichkeit. Zwischen diesen beiden Zuständen können die Vpn jedoch nicht genau unterscheiden, wie der geringe Abstand zwischen den Reizen belegt.

Anders als die ermittelten STI-Werte aus Tabelle 6.1 vermuten lassen, werden die Impulsantworten von den Vpn in eine einheitliche Reihenfolge gebracht. Für die Position 1, ergibt sich nahezu eine Reihenfolge entsprechend der Nachhallzeitdauer in Tab. 6.1 mit der Einschränkung, dass Zustand 3 und 4 (S3, S4) als sehr ähnlich beurteilt werden.

f in Hz	125	250	500	1000	2000	4000	STI
Zustand 0	2,21	1,82	1,49	1,28	1,18	1,08	0,53
Zustand 1	1,61	1,14	0,67	0,63	0,64	0,60	0,67
Zustand 2	1,62	0,73	0,67	0,64	0,63	0,61	0,69
Zustand 3	1,26	0,70	0,68	0,62	0,62	0,61	0,69
Zustand 4	0,91	0,71	0,65	0,63	0,63	0,60	0,70
Zustand 5	0,68	0,75	0,67	0,64	0,62	0,60	0,70

Tabelle 6.1: aus der Computersimulation ermittelte Nachhallzeiten in Sekunden, sowie STI-Werte

Die Bewertung mittels STI, der für die Zustände 1 bis 5 einen Wert von ungefähr 0,7 annimmt, scheint daher nicht geeignet, um unterschiedliche Nachhallzeiten im tieffrequenten Bereich angemessen zu berücksichtigen. Der Vergleich der Abbildungen 6.1 bis 6.3 zeigt, dass kurze Nachhallzeiten für die Verständlichkeit des Sprachsignals umso wichtiger sind, je geringer der Anteil des Direktschalls ist.

6.2 Ergebnisse mit simulierten Raummoden

Aus den Paarvergleichen mit simulierten Impulsantworten (Kap. 6.1) geht ein deutlicher Einfluss der Hörposition auf die Präferenzurteile hervor. In der Versuchsreihe mit simulierten Raummoden (Kap. 3.2.1) zeigt sich, dass die Terzbänder bezüglich ihrer störenden Wirkung auf die Sprachverständlichkeit klassifiziert werden können. Die Abbildungen 6.4-6.6 zeigen die Ergebnisse der Paarvergleiche, bei unterschiedlich gewähltem Konsistenzkoeffizient T (siehe Kap. 4.2.2.2). Dieser Wert dient als Kriterium bei der Überprüfung der Antworten der Vpn und beeinflusst, wie viele Vpn bei der Auswertung berücksichtigt werden. In Abbildung 6.4, mit der strengsten Testgröße T > 0.85, sind die Antworten von sechs Vpn, bei einer Gesamtzahl von 19 Vpn dargestellt.



Abbildung 6.4: Vergleich simulierter Raummoden von 86Hz bis 258Hz, für T > 0,85.

Für den betrachteten Frequenzbereich unterhalb 300 Hz lassen sich folgende Aussagen treffen. Bei der Ausbildung von Raummoden wird das Terzband um 216 Hz als besonders störend empfunden. Dieses Band liegt genau zwischen den Grundtonbereichen von Männern (130 Hz) und Frauen (240 Hz) (vergleiche Abb. 2.1). Ausgehend von diesem Frequenzband nimmt der Einfluss in beide Richtung ab, sodass die tiefste Modenfrequenz mit 86 Hz, als am wenigsten störend empfunden wird. Bei Aufweichen des Konsistenzkriteriums, (T > 0,75, bzw. T > 0,625) lässt sich der Trend erkennen, dass die Frequenzbereiche um 216 Hz immer ähnlicher beurteilt werden, siehe Abbildung 6.5, 6.6. Dies lässt sich durch die Aussagen von mehreren Vpn erklären, die angeben nach folgendem Muster geurteilt zu haben. Bis zu einer

gewissen Frequenz wird jenes Signal bevorzugt, das von der tieffrequenteren Raummode beeinflusst wird. Bei überschreiten dieser ,Trennfrequenz', fallen die Urteile zu Gunsten der Raummoden höherer Frequenz aus.



Abbildung 6.5: Vergleich simulierter Raummoden von 86Hz bis 258Hz, für T > 0,75.



Abbildung 6.6: Vergleich simulierter Raummoden von 86Hz bis 258Hz, für T > 0,625.

Vergleicht man die Aussagen der Vpn mit der Spektralanalyse des verwendeten Audiosignals in Abbildung 6.7, so erkennt man einen deutlichen Zusammenhang. Bei ca. 220 Hz befindet sich auch das Maximum des verwendeten Sprachmaterials, weshalb diese Raumeigenfrequenz besonders stark angeregt wird.



Abbildung 6.7: Spektralanalyse des Sprachsignals, welches beim Paarvergleich mit simulierten Raummoden verwendet wurde.

6.3 Ergebnisse mit realen Impulsantworten

Wie in Abschnitt 3.2 bereits beschrieben, wurden neben den simulierten auch reale Impulsantworten erzeugt, um diese näher zu untersuchen. Dieses Audiomaterial wurde verwendet, um einen weiteren Paarvergleich durchzuführen. Insgesamt wurden vier Sprachsignale miteinander verglichen. Ziel dieser Testreihe war es, einen Vergleich mit den Ergebnissen aus den Tests mit den simulierten Impulsantworten anzustellen. Die Messapparatur war dieselbe, die für den Paarvergleich mit simulierten Impulsantworten verwendet wurde (siehe Abschnitt 4.2). Insgesamt wurden 19 Vpn nach Ihren Präferenzen im Paarvergleich befragt. Beurteilt wurde dabei die Sprachverständlichkeit der Beispiele. Die Antworten wurden mittels BTL-Skalierung ausgewertet und in Abbildung 6.8 graphisch dargestellt. Die Werte IR1 bis IR4 entsprechen dabei den Raumzuständen 1 bis 4 in Abbildung 6.9.





Abbildung 6.8: Bewertung von vier real gemessenen Impulsantworten in BTL-Skalierung.

Die Grafik zeigt, dass zwischen den Beispielen nicht deutlich unterschieden werden konnte. Bei optimaler Differenzierbarkeit und eindeutiger Zuordnung durch die Vpn hätte sich eine Verteilung über den Bereich von 0 bis 3 mit einem Abstand von jeweils 1 zwischen den Impulsantworten ergeben. Die geringe Differenz der Messergebnisse zueinander, wirft die Frage auf, ob das gewählte Modell zur Auswertung dieser Daten geeignet ist. Zu diesem Zweck wird mittels des Q-Tests von Cochran [34] überprüft, ob sich die Vpn in ihrem Urteilsverhalten substantiell unterscheiden. Der Test überprüft für zwei oder mehrere Beurteiler, ob ihre Dominanzmatrizen besser durch das Zufallsmodell:

H₀: "Die Wahrscheinlichkeit einer Dominanz ist für alle Objektpaare gleich.", oder durch die Alternative

H₁: "Es gibt eine Übereinstimmung zwischen den Dominanzurteilen." Beschrieben werden können, näheres dazu siehe auch Hartung [32].

Der Test ergibt, dass keine Übereinstimmung zwischen den Dominanzurteilen der Vpn besteht. Das bedeutet, dass für diese Daten eine Auswertung der Daten mittels BTL-Skalierung keine brauchbaren Ergebnisse liefert. Eine Erklärung hiefür findet sich in Gediga [33], der darauf hinweist, dass sehr homogene oder heterogene Reize zu einem Perspektivenwechsel während der Beurteilung führen und somit eine Auswertung nach dem BTL-Modell ausschließen. Alle Vpn geben nach dem Hörtest an, keinen erkennbaren Unterschied bezüglich der Sprachverständlichkeit zwischen den einzelnen Sprachsignalen feststellen zu können.





Abbildung 6.9: Nachhallzeitvergleich der Raumzustände 1 bis 4 für Frequenzen bis 1kHz.

7. Zusammenfassung

Um Aussagen über den Einfluss der Nachhallzeit bei tiefen Frequenzen auf die Sprachverständlichkeit zu ermöglichen, wurden für exemplarische Räume und unterschiedliche akustische Raumzustände binaurale Impulsantworten simuliert, bzw. gemessen. Nach Faltung mit einem nachhallfreien Sprachsignal eines männlichen Sprechers und Darbietung über Kopfhörer wurde die Sprachverständlichkeit über Hörversuche im Paarvergleich beurteilt. Die Auswertung erfolgte nach dem sogenannten Bradley-Terry-Luce-Modell (BTL-Modell). Dies erlaubt es, die Ergebnisse auf einer Präferenzskala aufzutragen, sowie inkonsistente Antworten zu ermitteln und mit Hilfe des Kriteriums des Konsistenzkoeffizienten die Antworten einzelner Probanden ggf. auszuschließen.

7.1 Hörversuche mit simuliertem Klassenraum

Wie in Kapitel 3.1 beschrieben, wurde ein typischer Klassenraum modelliert und so angepasst, dass sich die in Abb. 7.1 gezeigten Sabine'schen Nachhallzeiten einstellten. Für verschiedene Raumzustände (Zstd. 0-5) wurden bei drei verschiedenen Empfängerpositionen (vorn - mitte – hinten) mit dem Simulationsprogramm CATT binaurale Raumimpulsantworten ermittelt. Die Grenzfrequenz des Raumes liegt bei Nachhallzeiten von weniger als 1 s unter 130 Hz, was die Anwendung eines geometrisch-statistischen Simulationsverfahrens rechtfertigt. Bei dem verwendeten Simulationsprogramm wird darüber hinaus der Frequenzbereich unterhalb der 125 Hz-Oktave extrapoliert, so dass auch hier Energieanteile vorhanden sind.



Abbildung 7.1: SabinescheNachhallzeiten des simulierten Klassenraums

Die Auswertung der Hörversuche zeigte (vgl. Abb. 7.2), dass der Raumzustand 0 (unbehandelt) von sämtlichen Probanden hinsichtlich der Sprachverständlichkeit als am ungünstigsten bewertet wurde. Auch die Raumzustände mit einem starken Anstieg der Nachhallzeit im mittelfrequenten Bereich, bzw. einem starken Anstieg in der 250 Hz-Oktave wurden ungünstiger bewertet, als dies bei den Raumzuständen 4 und 5 der Fall war. Die auf der BTL-Skala sehr nah beieinander liegenden Balken für die Zustände 4 und 5 zeigen jedoch, dass zwischen einem Abfall der Nachhallzeit (5) und einem leichten Anstieg der Nachhallzeit (4) zu tiefen Frequenzen kaum, bzw. keine Unterschiede hinsichtlich der Sprachverständlichkeit erkennbar waren. Diese Aussage gilt nicht nur für den hier betrachteten Platz, sondern auch für die untersuchten Plätze im hinteren und vorderen Raumbereich.



Abbildung 7.2: Exemplarisches Ergebnis der Hörversuche (simulierter Raum, Sitzplatz Mitte)

7.2 Hörversuche mit realen Raumaufzeichnungen

Um neben den Simulationen auch Ergebnisse mit gemessenen Schallfeldern und damit ggf. einer realistischeren Situation unterhalb der Grenzfrequenz des Raumes zu erhalten, wurden die Untersuchungen auf einen realen Raum ausgedehnt (siehe Kap. 3.2, S.28 ff). Hier wurden ähnlich wie bei Untersuchungen von Rau [31] durch Einbringen von Absorbern unterschiedliche akustische Raumzustände eingestellt. In Abbildung 7.3 sind die gemessenen Nachhallzeiten gezeigt.



Abbildung 7.3: Nachhallzeiten für unterschiedliche Raumzustände

Die Messungen erfolgten mit einem Kunstkopf und die Darbietung wie oben beschrieben über Kopfhörer. Die Ergebnisse der Hörversuchsreihe sind in Abb. 7.4 dargestellt. Es ist zu erkennen, dass die Beurteilungen sehr eng beieinander liegen. Auch war die Anzahl inkonsistenter Antworten sehr hoch. Dies bedeutet, dass im Hörversuch ein deutlicher Anstieg der Nachhallzeit unter 125 Hz respektive eine unterschiedliche Modenbedämpfung nicht zuzuordnen war. Auch umfangreiche Reimtests mit, bzw. ohne verdeckendem Rauschen führten zu diesem Ergebnis.



Abbildung 7.4: Ergebnis der Hörversuche (realer Raum)

7.3 Folgerungen

Aus den Untersuchungen lassen sich für Unterrichtsräume und verallgemeinernd auch für andere Sprachräume mit einem Volumen ab etwa 100 m³ die folgenden Schlüsse ziehen:

- Eine ausreichend niedrige Nachhallzeit in der 250 Hz-Oktave ist f
 ür eine gute Sprachverst
 ändlichkeit erwartungsgem
 äß sehr wichtig
- Die Nachhallzeit verliert unterhalb der 250 Hz-Oktave und insbesondere unterhalb der 125 Hz-Oktave in Bezug auf die Sprachverständlichkeit zunehmend an Bedeutung.
- Ein leichter Anstieg der Nachhallzeit zu tiefen Frequenzen, z.B. wie in der Literatur f
 ür klassische Musik gezeigt, ist akzeptabel und beeintr
 ächtigt nicht die raumakustische Qualit
 ät f
 ür die Sprachnutzung.

Die hier angestellten Betrachtungen gelten nicht für Räume mit deutlich kleinerem Raumvolumen. Hier spielt der modale Charakter des Schallfeldes im Grundtonbereich der Stimme eine wesentlich wichtigere Rolle.

8. Literaturverzeichnis

- [1] Bradley J.S.: *Speech intelligibility studies in classrooms*, JASA, Vol. 80, Nr. 3 (1986), pp. 846-854
- [2] Sutherland L.C, Lubman D.: *The Impact of Classroom Acoustics on Scholastic Achievement*, Proceedings of 17th International Congress on Acoustics, Rome (2001)
- [3] Schick A., Klatte M., Meis M.: *Die Lärmbelästigung von Lehrern und Schülern ein Forschungsstandsbericht*, Zeitschrift für Lärmbekämpfung, 46, Nr. 3 (1999), S. 77-87
- [4] DIN 18041: Hörsamkeit in kleinen bis mittelgroßen Räumen, (2004)
- [5] Ruhe C.: *Der Nachhall wird kürzer*, Sonderdruck aus Trockenbau Akustik, Sonderheft (2003)
- [6] Fasold W., Veres E.: *Schallschutz und Raumakustik in der Praxis*, 1. Auflage, Berlin: Verlag für Bauwesen (1998)
- [7] Dickreiter M.: *Handbuch der Tonstudiotechnik*, Band I. 6. Auflage. München: Saur Verlag KG (1997)
- [8] Zwicker, E.: *Psychoakustik*, Berlin, Heidelberg, Springer-Verlag (1982)
- [9] Graber G.: *Raumakustik*, Vorlesungsskript, Institut für Breitbandkommunikation, T.U. Graz (2004)
- [10] Kuttruff H.: Akustik, Stuttgart, S. Hirzel Verlag (2004)
- [11] Alton Everest, (Handbook for Sound Engineers, 1st Edition 1987) Acoustics of Small Rooms, Kapitel 3
- [12] Müller, Möser: *Taschenbuch der Technischen Akustik*, Springer Verlag, 3. Auflage, (2004)
- [13] Schroeder MR, J. Acoust. Soc. Amer., Vol. 37 (1965), S. 409 412 New method of measuring reverberation time
- [14] Jordan V.L., JASA Vol. 47 (1970), S. 408 412 Acoustical criteria for auditoriums and their relation to model techniques
- [15] Thiele R., Acustica Vol.3 (1953), S.291-302 Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen
- [16] ISO 3382, second edition (1997) Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters
- [17] Kürer R., Acustica Vol. 21 (1969), S. 370 372 Zur Gewinnung von Einzahlkriterien bei Impulsmessungen in der Raumakustik

- [18] Peutz V.M.A., Klein W., Acust. Soc. Netherlands, Vol. 28 (1973), pp. 11 18 Articultion Loss of Consonants influenced by Noise, Reverberation and Echo
- [19] Beranek L.L., Acoustical Society of America Publications (1996) *How they sound, Opera and Concert Halls*
- [20] Dietsch L., Kraak W., Acustica Vol.60 (1986), S. 205-216 Ein objektives Kriterium zur Erfassung von Echostörungen bei Musik- und Sprachdarbietungen
- [21] N.R. French, J.C. Steinberg, JASA Vol. 19 (1947), S. 90 119 Intelligibility of Speech Sounds Entwicklung vom AI
- [22] American National Standard, ANSI 3.5 (1969) Methods for the calculation of the articulation index
- [23] ANSI S 3.5 (1996) oder auch ANSI S 3.79 (1998)
- [24] Houtgast T., Steeneken H.J.M., Plomp R., Acustica Vol.46 (1980), S.60 72 Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I General Room Acoustics
- [25] Dalenbäck, B.I.: *CATT-Acoustic v8 User's Manual*, 1st edition. Gothenburg: Spectra D'Sign Kopiering AB (2002)
- [26] Steeneken H.J.M., Houtgast T., JASA, Vol. 67 (1980), S.318 A physical method for measuring speech-transmission quality
- [27] Hunecke: http://www.hunecke.de/german/rechenservice/raumeigenmoden.html
- [28] Klein + Hummel: <u>www.kleinhummel.de</u>
- [29] Head Acoustics: www.head-acoustics.de
- [30] Hörtech: http://www.hoertech.de
- [31] Rau, C.: *Einfluss tieffrequenter Störgeräusche auf die Sprachverständlichkeit*, Diplomarbeit, I.E.M., Graz (2004)
- [32] Hartung J.: *Statistik*, 10. Auflage, Oldenbourg Verlag, München (1995)
- [33] Gediga G.: Osnabrücker Schriften zur Psychologie: *Skalierung*, Bd.5, LIT Verlag (1998)
- [34] Boehnke K., Bortz J., Lienert G.: *Verteilungsfreie Methoden in der Biostatistik*, 2. Auflage Springer Verlag (2000), Kapitel 9, S. 488 502
- [35] Cochran, W.G.: *The comparison of percentages in matched samples*. Biometrika, Kap. 10, S. 417 451