



## *Aufgabenstellung*



BRANDENBURGISCHE TECHNISCHE UNIVERSITÄT COTTBUS  
FAKULTÄT 3 FÜR MASCHINENBAU, ELEKTROTECHNIK  
UND WIRTSCHAFTSINGENIEURWESEN  
LEHRSTUHL KOMMUNIKATIONSTECHNIK



# Diplomarbeit

eingereicht von:

Tobias Holfeld

Matrikel-Nr.: 2407254

Studiengang: Elektrotechnik

Thema:

## **Untersuchung von realistischen akustischen Umgebungen und ihren Einfluss auf internet-basierte Sprachgütemessungen**

Betreuer: Prof. Dr.-Ing. Klaus-Rüdiger Fellbaum

Dipl.-Ing. Christian Richter

Abgabe: Cottbus, September 2011

## **Kurzfassung**

In der heutigen Zeit tritt das Thema der internet-basierten Sprachgütemessung immer mehr in den Vordergrund. Sprache ist die bedeutendste Form der zwischenmenschlichen Kommunikation, und die Hauptaufgabe der Sprachübertragung liegt in der Informationsübermittlung. Daher ist die Sprachverständlichkeit die bedeutendste Größe zur Bestimmung der Qualität einer Sprachprobe.

Zur Ermittlung müssen auditive Testmethoden herangezogen werden, da zurzeit nur anhand von persönlichen Einschätzungen eine hinreichend aussagefähige Möglichkeit besteht die Güte einer Sprachprobe zu bewerten. Diese Tests sind jedoch sehr umfangreich und nehmen viel Zeit in Anspruch.

Die Sprachgütemessungen über das Medium Internet ist hierbei eher eine Randerscheinung in der Entwicklung von Sprachsynthesystemen. Die für diese Arbeit verwendete internet-basierte Bestimmung der Sprachgüte ist ein Versuch, diese Testverfahren zu optimieren.

Die vorliegende Arbeit soll anhand des Störeinflusses von realistischen akustischen Umgebungen und ihren Einfluss auf internet-basierte Sprachgütemessungen untersuchen, ob Sprachgütemessungen über das Medium Internet zu einem brauchbaren Ergebnis kommen.

Das Ergebnis zeigt, dass verschiedene akustische Bedingungen unterschiedliche Einflüsse auf die Messungen haben. Die Wertungen der Probanden zeigen dennoch, dass eine internet-basierte Sprachgütemessung unter normalen Umgebungsgeräuschen brauchbare Resultate liefern kann.

Diese Ergebnisse sind jedoch mit Vorsicht zu betrachten. An Resultate eines Studiotests kommen die Verfahren über das Internet heutzutage noch nicht heran. Sie können durch das vereinfachte Verfahren, ohne definierte Bedingungen, nur eine Tendenz geben.

# Inhaltsverzeichnis

Kurzfassung .....	V
Abkürzungsverzeichnis.....	VIII
Abbildungsverzeichnis.....	IX
Tabellenverzeichnis.....	XI
1. Einleitung .....	1
2. Definition und Einteilung von Sprachgüte-Messverfahren.....	3
2.1 Sprachgütemessung: Definition und Grundlagen .....	3
2.2 Unterscheidung der Verfahren .....	4
3. Auditive Testverfahren.....	6
3.1 Mean Opinion Score (MOS)-Test .....	8
3.2 Paarvergleichstest.....	9
3.3 Benotungstest .....	10
4. Internet-basierte Sprachgütemessung.....	12
5. Erstellung der akustischen Umgebungen .....	15
5.1 Herangehensweise.....	15
5.1.1 Räumliches Hören .....	15
5.1.2 Schall in Räumen.....	16
5.1.3 Aufnahmeverfahren .....	17
5.1.4 Aufnahmegerät - Kunstkopf .....	22
5.2 Umgebungen.....	26
5.2.1 Szenario 1 - „Zimmer an Straße“ .....	27
5.2.2 Szenario 2 - „Zimmer - nebenan Küchengeräusche und TV“ .....	29
5.2.3 Szenario 3 - „Raum 321 mit Baustellenlärm“.....	31
5.2.4 Szenario 4 - „Raum 119 PC Pool mit mehreren Computerlüftern“.....	33
5.2.5 Szenario 5 – „Modulated Noise Reference Unit (MNRU)“ .....	34
5.2.6 Szenario 6 - „Modulated Noise Reference Unit (MNRU) mit natürlicher Sprache“ ..	37

5.2.7	Szenario 7 - „Raum 321 Original“ .....	37
6.	Testumgebung und Randbedingungen.....	39
6.1	Hörtestumgebung.....	39
6.2	Randbedingungen des Tests.....	40
6.2.1	Versuchsteilnehmer .....	40
6.2.2	Räumliche und technische Ausstattung .....	41
6.2.3	Versuchsablauf.....	42
6.2.4	Störeffekte .....	42
7.	Ablauf und Durchführung des Tests .....	44
8.	Auswertung der Ergebnisse .....	48
8.1	Allgemeine Ergebnisse.....	48
8.2	Statistische Betrachtungen.....	54
8.2.1	Mittelwert, Varianz und Standardabweichung .....	54
8.2.2	Konfidenzintervall und Signifikanz .....	55
8.3	Mean Opinion Score für alle Szenarien .....	56
8.4	Auswertung der Szenarien .....	60
8.4.1	Szenario 1 – „Zimmer an Straße“ .....	60
8.4.2	Szenario 2 – „Zimmer nebenan Küchengeräusche und TV“ .....	64
8.4.3	Szenario 3 – „Raum 321 mit Baustellenlärm“ .....	69
8.4.4	Szenario 4 – „Raum 119 PC Pool mit mehreren Computerlüftern“ .....	73
8.4.5	Szenario 5 – „weißes Rauschen (MNRU) .....	77
8.4.6	Szenario 6 – „weißes Rauschen (MNRU) mit natürlicher Sprache“ .....	81
8.4.7	Szenario 7 – „Raum 321 Original“ .....	86
8.5	Diskussion der Ergebnisse .....	88
9.	Zusammenfassung und Ausblick.....	90
	Literaturverzeichnis .....	93
	Anhang.....	96
	Eidesstattliche Erklärung und Einverständnis zur Veröffentlichung.....	99

# Abkürzungsverzeichnis

ACR	Absolute Category Rating
CCITT	Comite´Consultatif International Telegraphique et Telephonique
CCR	Comparison Category Rating
dB	Dezibel
DCR	Degration Category Rating
HTML	Hyper Text Markup Language
Hz	Hertz (Einheit der Frequenz $1\text{Hz} = \text{s}^{-1}$ )
ITU	International Telecommunication Union
kHz	Kilo Hertz
LG	Lehrgebäude
MNRU	Modulated Noise Reference Unit
MOS	Mean Opinion Score
s	Sekunde
SNR	Signal-Störabstand

## Abbildungsverzeichnis

Abbildung 1: Aspekte der Sprachqualität [Lüd01] .....	4
Abbildung 2: Darstellung auditiver Bewertungsverfahren angelehnt an [Fel00] und [ESS01].	6
Abbildung 3: Übersicht Verständlichkeitstest [Fel00] .....	7
Abbildung 4: Richtungsbestimmende Bänder [Dic03] .....	15
Abbildung 5: AB-Mikrofonierung [Gör94] .....	18
Abbildung 6: MS-Mikrofonverfahren mit M als Kugel links und M als Niere rechts [Gör94] ...	19
Abbildung 7: XY-Mikrofonierung [Gör94] .....	20
Abbildung 8: Kunstkopf KU 100 der Firma Neumann mit Innenohr-Mikrofonen [Neu11] .....	21
Abbildung 9: Entzerrung einer Kunstkopfaufnahme für die korrekte Wiedergabe [Hea11] ....	22
Abbildung 10: Kunstkopf HMS II.3 der Firma Head acoustics .....	24
Abbildung 11: Mobiles Aufnahmegerät H2 der Firma Zoom [Zoo11] .....	25
Abbildung 12: Schallpegelmesser der Firma Brüel & Kjaer .....	25
Abbildung 13: Szenario 1 - "Zimmer an Straße" .....	27
Abbildung 14: Spektrogramm Szenario 1 .....	28
Abbildung 15: Szenario 2 - "Zimmer nebenan Küchengeräusche und TV" .....	29
Abbildung 16: Spektrogramm Szenario 2 .....	30
Abbildung 17: Szenario 3 - "Raum 321 mit Baustellenlärm" .....	31
Abbildung 18: Spektrogramm Szenario 3 .....	32
Abbildung 19: Szenario 4 - "Raum 119 PC Pool mit mehreren Computerlüftern" .....	33
Abbildung 20: Spektrogramm Szenario 4 .....	34
Abbildung 21: Vereinfachte Regelung der MNRU für digitale Breitbandversion [ITU810] .....	35
Abbildung 22: Spektrogramm Szenario 5 .....	36
Abbildung 23: Testplatz für den Versuch .....	41
Abbildung 24: Platz des Probanden in der schalldichten Kabine .....	42
Abbildung 25: Lautstärkeregelung der Soundkarte .....	44
Abbildung 26: Skala für die Sprachverständlichkeit .....	46
Abbildung 27: Skala für den Einfluss der Umgebung .....	46
Abbildung 28: Bildschirmoberfläche des Versuches .....	47
Abbildung 29: Altersverteilung .....	48
Abbildung 30: Schalldruckpegel am Kopfhörer .....	53
Abbildung 31: MOS-Werte für die Verständlichkeit der Sprache .....	57
Abbildung 32: MOS-Wert für den Einfluss der Umgebung .....	58
Abbildung 33: MOS-Wert über beide Beurteilungskriterien .....	59
Abbildung 34: Auswertung Szenario 1 – Sprachverständlichkeit .....	60
Abbildung 35: Auswertung Szenario 1 – Einfluss der Umgebung .....	61

## Abbildungsverzeichnis

Abbildung 36: Szenario 1 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	62
Abbildung 37: Szenario 1 Sprache und 4-facher Störpegel .....	62
Abbildung 38: Szenario 1 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	63
Abbildung 39: Szenario 1 Sprache und 8-facher Störpegel .....	63
Abbildung 40: Auswertung Szenario 2 – Sprachverständlichkeit .....	64
Abbildung 41: Auswertung Szenario 2 - Einfluss der Umgebung .....	65
Abbildung 42: Szenario 2 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	66
Abbildung 43: Szenario 2 Sprache und 4-facher Störpegel .....	67
Abbildung 44: Szenario 2 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	67
Abbildung 45: Szenario 2 Sprache und 8-facher Störpegel .....	68
Abbildung 46: Auswertung Szenario 3 – Sprachverständlichkeit .....	69
Abbildung 47: Auswertung Szenario 3 - Einfluss der Umgebung .....	70
Abbildung 48: Szenario 3 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	70
Abbildung 49: Szenario 3 Sprache und 4-facher Störpegel .....	71
Abbildung 50: Szenario 3 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	71
Abbildung 51: Szenario 3 Sprache und 8-facher Störpegel .....	72
Abbildung 52: Auswertung Szenario 4 – Sprachverständlichkeit .....	73
Abbildung 53: Auswertung Szenario 4 - Einfluss der Umgebung .....	74
Abbildung 54: Szenario 4 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	74
Abbildung 55: Szenario 4 Sprache und 4-facher Störpegel .....	75
Abbildung 56: Szenario 4 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	75
Abbildung 57: Szenario 4 Sprache und 8-facher Störpegel .....	76
Abbildung 58: Auswertung Szenario 5 – Sprachverständlichkeit .....	77
Abbildung 59: Auswertung Szenario 5 - Einfluss der Umgebung .....	78
Abbildung 60: Szenario 5 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	79
Abbildung 61: Szenario 5 Sprache und 4-facher Störpegel .....	79
Abbildung 62: Szenario 5 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	80
Abbildung 63: Szenario 5 Sprache und 8-facher Störpegel .....	80
Abbildung 64: Auswertung Szenario 6 – Sprachverständlichkeit .....	81
Abbildung 65: Auswertung Szenario 6 - Einfluss der Umgebung .....	83
Abbildung 66: Szenario 6 - 4-facher Störpegel (links) und Sprachbeispiel (rechts) .....	84
Abbildung 67: Szenario 6 Sprache und 4-facher Störpegel .....	84
Abbildung 68: Szenario 6 - 8-facher Störpegel (links) und Sprachbeispiel (rechts) .....	85
Abbildung 69: Szenario 6 Sprache und 8-facher Störpegel .....	85
Abbildung 70: Auswertung Szenario 7 – Sprachverständlichkeit .....	86
Abbildung 71: Auswertung Szenario 7 - Einfluss der Umgebung .....	87

## Tabellenverzeichnis

Tabelle 1: Vergleich Sprachverständlichkeit Szenario 5 und 6.....	82
Tabelle 2: Vergleich Einfluss der Umgebung Szenario 5 und 6 .....	83
Tabelle 3: Zeit des Versuchsbeginns mit Störpegel .....	88

# 1. Einleitung

Sprache ist die bedeutendste Form der zwischenmenschlichen Kommunikation. Die wichtigste Aufgabe der Sprachübertragung liegt hierbei in der Informationsübermittlung. Daher ist die bedeutendste Größe zur Bestimmung der Qualität einer Sprachprobe die Sprachverständlichkeit. Ohne Verständlichkeit kann keine Information übermittelt werden und das ist der Hauptzweck der Sprache. Nur anhand technischer Parameter ist die Sprachqualität oder auch Sprachgüte jedoch nicht zu bestimmen. Zur Ermittlung müssen auditive Testmethoden herangezogen werden. Diese Testverfahren sind zurzeit die einzige Möglichkeit, die Güte einer Sprachprobe zu bewerten, denn für eine Testperson ist die wichtigste Entscheidungsgrundlage der persönliche Höreindruck. Die Tests sind jedoch sehr umfangreich, nehmen viel Zeit in Anspruch, und gehören zu den schwierigsten und messtechnisch aufwendigsten Aufgaben.

Zur genauen Analyse ist eine Vielzahl von Testpersonen notwendig, die unter Studiobedingungen Sprachproben anhören und beurteilen. Am Lehrstuhl Kommunikationstechnik der Brandenburgischen Technischen Universität wurden bereits Untersuchungen hinsichtlich der technischen Machbarkeit solcher Tests über das Internet durchgeführt. Dies würde den Arbeits- und Zeitaufwand der Testteilnehmer erheblich verringern, und es könnten in kurzer Zeit viele Personen akquiriert werden.

Die Umsetzung von Sprachgütemessungen über das Internet wäre besonders für Entwickler von Sprachsynthesystemen interessant, da sie ihre Systeme schnell und effizient mit einer großen Anzahl von Probanden testen könnten.

Das größte Problem der internet-basierten Sprachgütemessung besteht in den akustischen Randbedingungen an verschiedenen Testorten. Jeder Testteilnehmer hat andere Gegebenheiten und Störeinflüsse.

In der vorliegenden Arbeit sollen typische Umgebungsszenarien und ihre Auswirkung auf die Ergebnisse näher untersucht werden. Dafür werden mit einer größeren Anzahl von Probanden Sprachgütemessungen in realistischen, virtuellen akustischen Umgebungen durchgeführt.

Im zweiten Kapitel werden die Sprachgüte-Messverfahren näher betrachtet. Dabei wird auf die Grundlagen der Sprachgütemessung eingegangen und es werden verschiedene Tests zur Bestimmung der Sprachgüte erläutert. Im folgenden Kapitel wird auf die auditiven Testverfahren eingegangen, wobei der hier verwendete MOS-Test beschrieben wird. Im vierten Kapitel wird die, der Arbeit zugrunde liegende, internet-basierte Sprachgütemessung dargestellt. Das fünfte Kapitel beschreibt die Erstellung der akustischen Umgebungen. Hierbei werden verschiedene Themen, wie das räumliche Hören, Schall in Räumen und verschiedene Aufnahmetechniken beschrieben. Weiterhin werden die sieben erzeugten Szenarien näher betrachtet und auf ihre Gegebenheiten analysiert. Im folgenden Kapitel werden die Testumgebung und ihre Randbedingungen betrachtet. Hierbei wird auf die Versuchsteilnehmer, die räumliche und technische Ausstattung, der Versuchsablauf und eventuell eintretende Effekte eingegangen. Im siebten Abschnitt werden der Versuch und seine Durchführung detailliert dargestellt. Das achte Kapitel beinhaltet die Auswertung und Interpretation der Ergebnisse. Dabei wird auch auf die verschiedenen Randbedingungen und Effekte eingegangen, die auf die Testteilnehmer wirken. Im letzten Kapitel werden eine Zusammenfassung der Ergebnisse und ein Ausblick für spätere Arbeiten gegeben.

# 2. Definition und Einteilung von Sprachgüte-Messverfahren

## 2.1 Sprachgütemessung: Definition und Grundlagen

Sprachgütemessungen werden durchgeführt, um Sprachqualität zu beurteilen. „Die Bestimmung der Sprachgüte gehört zu den schwierigsten messtechnischen Aufgaben“ [Fell84]. Sie ist mit apparativen Messmethoden kaum zu erfassen. Somit kann eine Beurteilung der Sprachgüte nur anhand auditiver Testverfahren stattfinden. Lange Zeit stand für die Sprachgüte das Kriterium der Sprachverständlichkeit an erster Stelle, denn ohne Verständlichkeit kann keine Information übermittelt werden und das ist der Hauptzweck von Sprache. „Eine hohe Sprachverständlichkeit ist jedoch keineswegs gleichbedeutend mit einer hohen Sprachgüte“ [Fell84]. Dennoch ist die Verständlichkeit eines der wichtigsten Kriterien für den Informationsaustausch. „Es verwundert daher nicht, dass für die Sprachgütebestimmung in erster Linie Verfahren der Sprachverständlichkeitsmessung entwickelt worden sind“ [Fell84].

Die angewendeten Verfahren sind heutzutage überwiegend auditiver Art. Das heißt, dass alle Messergebnisse anhand von Probandenbewertungen aufgenommen werden müssen.

Die Messung der Sprachgüte kann bzw. muss anhand weiterer Merkmale charakterisiert werden, da die Sprachverständlichkeit zu grob für die alleinige Charakterisierung ist. Laut Lüdtkke [Lüd01] ist Sprache auch dann noch gut verständlich, wenn die Gesamtqualität schon als mangelhaft empfunden wird. Berger [Ber98] beschreibt die Sprachqualität wie folgt:

*„Sprachqualität ist die Gesamtheit aller möglicherweise maßgeblichen Eigenschaften des in vielen Kategorien (z.B. Verständlichkeit, Natürlichkeit, Erkennung von Sprechereigenschaften) beschreibbaren Sprechschalls.“*

Weitere erfassbare Eigenschaften der Sprachgüte sind zum Beispiel Prosodie, Angenehmheit, Rauheit, Schärfe, Höranstrengung und Verzerrung [Ton02]. Abbildung 1 gibt einen Überblick über die Kriterien zur Bestimmung der Sprachgüte.

## 2. Definition und Einteilung von Sprachgüte-Messverfahren



Abbildung 1: Aspekte der Sprachqualität [Lüd01]

Da die Sprachgüte in erster Linie nur auditiv beurteilt werden kann, haben Erwartungshaltung, Erfahrung, Hörvermögen und das persönliche Befinden des Probanden einen starken Einfluss auf dessen Urteil. Durch die vielen verschiedenen Einflüsse entsteht ein relativ großer Versuchsaufwand bei dem die Ergebnisse meist großen Schwankungen unterworfen sind. Um dem entgegenzuwirken wird eine Vielzahl an Testpersonen benötigt, um die Abweichungen so gering wie möglich zu halten.

### 2.2 Unterscheidung der Verfahren

Wie bereits im Punkt 2.1 angedeutet, finden zwei verschiedene Arten der Qualitätsbestimmung Anwendung. Zum einen auditive und zum anderen instrumentelle Testverfahren.

Auf die instrumentellen Verfahren soll an dieser Stelle nur kurz eingegangen werden, da sie nicht relevant für die vorliegende Arbeit sind. Diese Testverfahren erlauben eine ausschließlich objektive Messung. Hierbei wird die Sprachqualität anhand verschiedener physikalischer, messtechnisch wahrnehmbarer Größen bestimmt. Das

## **2. Definition und Einteilung von Sprachgüte-Messverfahren**

Problem hierbei liegt im Finden der qualitätsrelevanten Merkmale der Sprache, die messtechnisch erfassbar sind. Die in Abbildung 1 aufgeführten Aspekte sind hierbei nur sehr schwer objektiv auswertbar. Doch nur durch diese Parameter kann ein aussagekräftiges Ergebnis geliefert werden.

Ziel der instrumentellen Messverfahren soll es sein, die aufwendigen auditiven Tests abzulösen. Es wird versucht, das Erfassen und Beurteilen der Probanden nachzubilden. Erst wenn dies gelungen ist, können die objektiven Testverfahren mit ihren Vorteilen angewandt werden und sichere Ergebnisse liefern. Ein großer Vorteil der objektiven Methode gegenüber den auditiven Versuchen ist der geringe Versuchsaufwand. Desweiteren kann zu jeder Zeit eine Qualitätseinschätzung von Sprachproben stattfinden.

Auf die auditiven Testverfahren soll im nächsten Abschnitt näher eingegangen werden.

### 3. Auditive Testverfahren

Durch die Komplexität der Sprachgütemessungen sind verschiedene Verfahren zur auditiven Bestimmung von Sprachproben entstanden. Diese spiegeln teilweise nur einige Gesichtspunkte der Sprachqualität wieder. „Ein wichtiger Aspekt bei der Auswahl eines geeigneten subjektiven Testverfahrens ist der Aufwand, der weniger durch technische Bedingungen (Versuchsraum und Audiotechnik), sondern durch die Beteiligung vieler repräsentativer Testpersonen bestimmt wird“ [Irm98]. Der wesentlichste Aspekt der Sprachgüte ist die Verständlichkeit. Somit wurden überwiegend Testverfahren zur Sprachverständlichkeit für die Sprachgütemessung entwickelt.

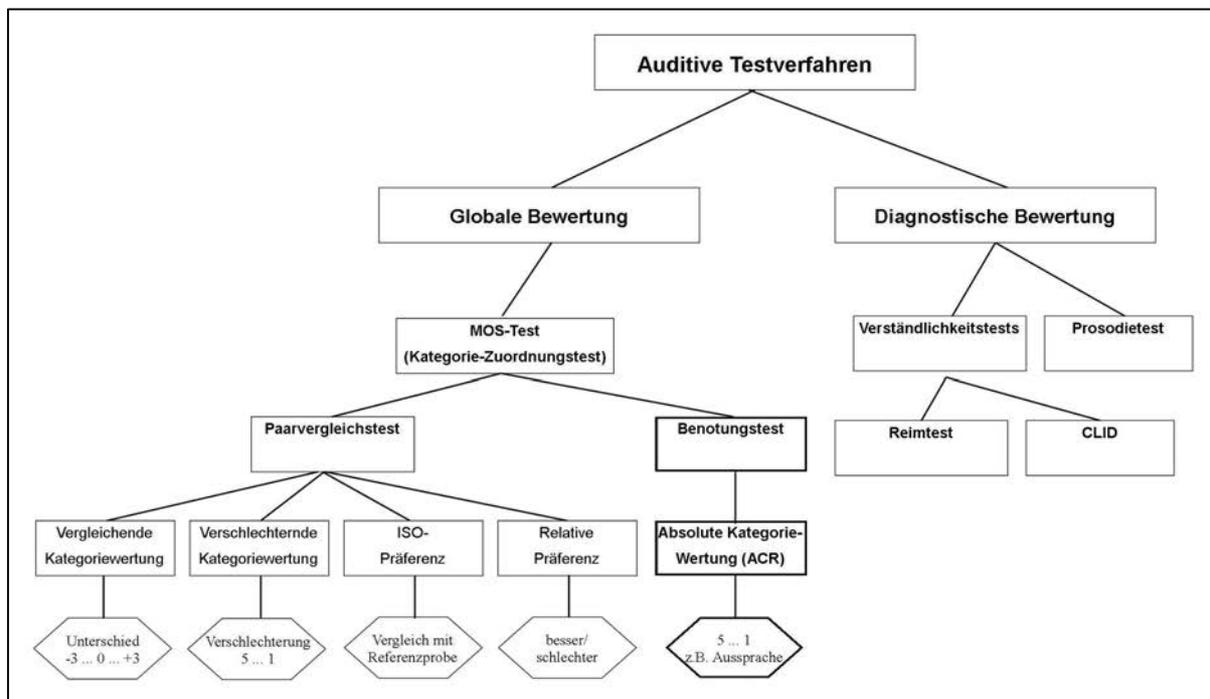


Abbildung 2: Darstellung auditiver Bewertungsverfahren angelehnt an [Fel00] und [ESS01]

In Abbildung 2 ist eine Darstellung der wichtigsten anwendbaren auditiven Testverfahren zu sehen. Es ist anzumerken, dass bei der Durchführung von diesen Untersuchungen das Bewertungsziel beachtet werden muss.

Die Verfahren werden in *globale* und *diagnostische* Bewertung untergliedert. Hierbei ist jedoch zu beachten, dass eine exakte Trennung zwischen beiden

### 3. Auditive Testverfahren

Bewertungsmöglichkeiten nicht immer möglich ist, da viele Verfahren Ergebnisse für beide Varianten liefern.

Bei den *globalen* Testverfahren werden alle abgegebenen Beurteilungen für jedes Kriterium gemittelt und ergeben so einen mittleren Meinungswert der Sprachproben. Dieser Mittelwert wird als Mean Opinion Score (MOS) bezeichnet [ESS01].

Zur Gewinnung *diagnostischer* Ergebnisse von Sprachproben müssen andere Verfahren als die MOS-Tests herangezogen werden. Hierbei kommen Verständlichkeitstests, wie der Reimtest nach Sotscheck und der Cluster-Identifizierungstest (CLID-Test) nach Jekosch, zur Anwendung. Desweiteren kann anhand eines Prosodietests der Prosodieverlauf einer Probe untersucht werden. Auf die diagnostischen Verfahren soll hier nicht weiter eingegangen werden. Detailliertere Aussagen sind der Arbeit von Lüdtkke [Lüd01] und dem Paper der 12. Konferenz der elektronischen Sprachverarbeitung 2001 in Bonn [ESS01] zu entnehmen.

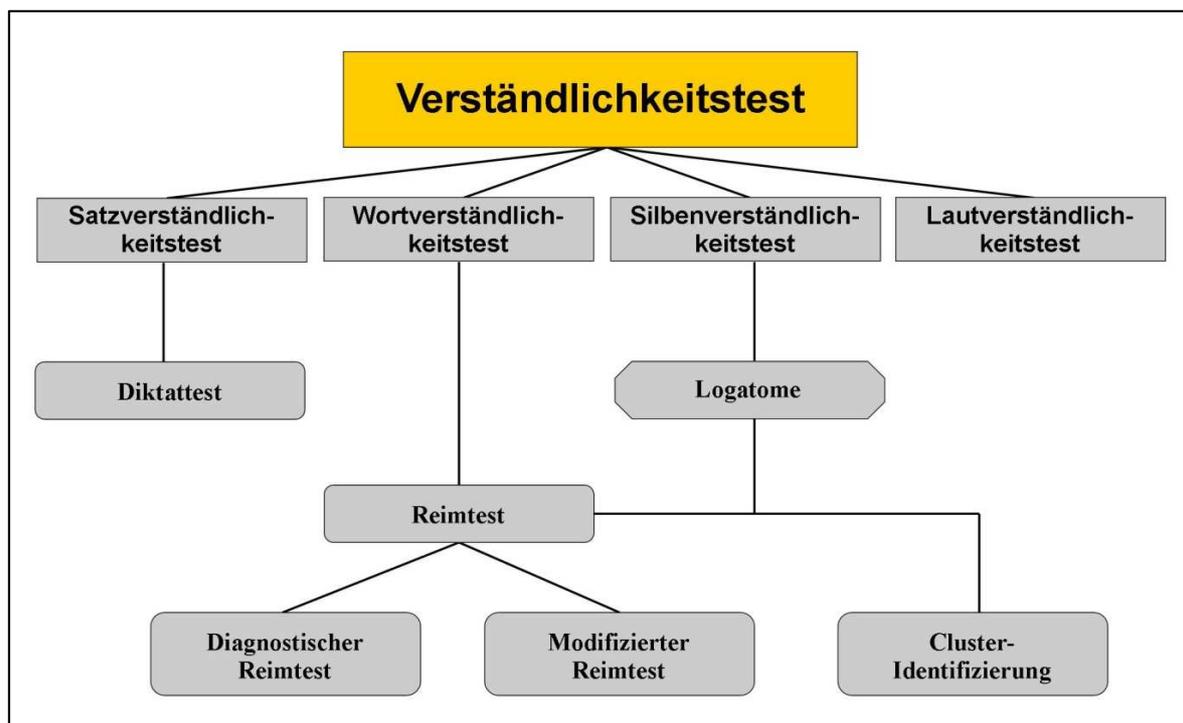


Abbildung 3: Übersicht Verständlichkeitstest [Fel00]

Abbildung 3 zeigt verschiedene Testmethoden zur Überprüfung der Verständlichkeit von Sprache. Der Verständlichkeitstest lässt sich in vier verschiedene Testverfahren aufteilen. Bei der Satzverständlichkeit wird durch einen Diktattest die Verständlichkeit der Sprache überprüft. Jedoch besteht hier die Möglichkeit, dass das Ergebnis

### 3. Auditive Testverfahren

verfälscht wird, da Wörter aus dem Kontext ergänzt werden können. Ein weiteres Verfahren ist der Wortverständlichkeitstest. Hier wird dem Probanden ein Wort akustisch wiedergegeben und gleichzeitig bekommt er per Monitor optisch ähnlich klingende Worte dargeboten, aus denen er das gehörte Wort herauszufinden hat. Diese Form bezeichnet man als *Reimtest*. Der Silbenverständlichkeitstest besteht aus sinnleeren Silben (Logatome). Logatome sollten nach der Empfehlung des CCITT (Comité Consultatif International Telegraphique et Telephonique) die Form konsonantischer Anlaut – Vokal – konsonantischer Auslaut besitzen. Nachteilig ist hierbei, dass Logatome wenig Ähnlichkeit mit dem natürlichen Sprachfluss haben. Die Lautverständlichkeit wird getestet, indem einzelne Laute in Reimtests wiedergegeben werden. Hier wird, wie im Silbenverständlichkeitstest, eine hohe Aufmerksamkeit von den Probanden gefordert.

#### 3.1 Mean Opinion Score (MOS)-Test

Der Mean Opinion Score Test ist ein Benotungstest bzw. ein Verfahren der Kategoriezuweisung. Beim Benotungstest bewerten die Probanden Sprachproben in verschiedenen Kategorien anhand von Noten. Heutzutage wird meist die fünfstufige Bewertung genommen. „Fünf“ steht für eine „ausgezeichnete“ Wiedergabe und „eins“ für „schlecht“. Dazwischen kann mit „gut“, „ziemlich gut / ordentlich“ und „mäßig / dürftig“ geurteilt werden.

Beim Opiniontest sollte die Sprachprobe die Mindestlänge eines Satzes aufweisen, damit die Satzmelodie als wichtige Komponente der Sprachgüte miterfasst werden kann. Anhand der abgegebenen Ergebnisse wird für jedes Kriterium ein Mittelwert gebildet, welcher einen durchschnittlichen Meinungswert der Sprachproben ergibt. Dieser Mittelwert wird als Mean-Opinion-Score (MOS) bezeichnet. Es wird davon ausgegangen, dass dieser Wert die Güte des jeweiligen Kriteriums widerspiegelt. Anhand der errechneten MOS-Werte wird eine qualitative Rangfolge der Sprachproben festgelegt. Trotz der bedingten Reproduzierbarkeit und der Ungenauigkeit der Opiniontests werden diese hauptsächlich angewandt, da sie mit relativ geringem Zeit- und Geräteaufwand durchführbar sind. Im Vergleich mit anderen Messverfahren treten bei hochwertiger Sprachübertragung keine Unterschiede in den Ergebnissen auf [Fel84].

Der Zuordnungstest lässt sich beispielsweise in den Paarvergleichstest und den Benotungstest aufteilen. Der Paarvergleichstest wird im Kapitel 3.2 und der Benotungstest im Kapitel 3.3 vorgestellt. Eine Alternative zu dieser absoluten Beurteilung zur globalen Bewertung kann der DCR-Test, der CCR-Test, der Präferenztest oder der ISO-Präferenztests sein. Im Folgenden wird kurz auf diese Verfahren eingegangen. Zur detaillierteren Veranschaulichung wird auf das Paper der 12. Konferenz der elektronischen Sprachsignalverarbeitung verwiesen [ESS01].

#### 3.2 Paarvergleichstest

Beim Paarvergleichstest (Präferenztest) werden zwei Sprachproben anhand einzelner Merkmale miteinander verglichen und auf ihre Sprachgüte untersucht. Die Hauptformen des Präferenztests sind die vergleichende Kategoriewertung (engl.: Comparison Category Rating; CCR) und die verschlechternde Kategoriewertung (engl.: Degration Category Rating; DCR).

Bei der vergleichenden Kategoriewertung werden zwei Sprachproben anhand ihrer Qualitätsunterschiede beurteilt. Dies geschieht mit Hilfe einer sieben stufigen Skala von „-3“ („viel schlechter“) bis „+3“ („viel besser“). Beim Test mit der verschlechternden Kategoriewertung wird anhand einer Referenzprobe die Verschlechterung gewertet. Hierzu dient eine fünfstufige Skala von „5“ (sehr belästigend) bis „1“ (unhörbar) [Fel03]. Weiterhin gibt es den ISO-Präferenztest und den relativen Präferenztest. Beim ISO-Testverfahren wird mit einer bestimmten Referenzprobe verglichen. Der Signalstörabstand (SNR) dieser Probe ist veränderbar. Nun wird das SNR solange verändert bis die Sprachprobe und die Referenzprobe als gleich bewertet werden. Bei der relativen Präferenz wird dem Probanden ein Sprachsignal vorgegeben, welches er mit einer zweiten Sprachprobe vergleichen soll. Er muss entscheiden, ob die zweite Probe „besser“, „gleich“ oder „schlechter“ ist [Fel84]. Es ist anzumerken, dass die Paarvergleichs- bzw. Präferenztests nicht zwangsläufig benotet werden müssen. Bei der relativen Präferenz wird beispielsweise nur entschieden, ob die zu bewertende Probe „besser“ oder „schlechter“ als eine Vergleichsprobe ist.

### 3.3 Benotungstest

Der Benotungstest, auch als Absolute Kategoriewertung (engl.: Absolute Category Rating; ACR) bezeichnet, wird von der ITU (engl.: International Telecommunication Union) als beste und effizienteste Testmethode der Kategorie-Zuordnungstests empfohlen. Laut Lüdtker sind die vergleichsweise einfache Versuchsstruktur und die gute Abbildung der Realität ausschlaggebend [Lüd01]. Durch dieses leicht zu automatisierende Testverfahren können mehrere Einzelwertungen zum Oberbegriff Sprachqualität abgefragt werden. „Durch die Möglichkeit der Bewertung von verschiedenen Kriterien kann die Qualität der getesteten Systeme recht differenziert beurteilt werden und Hinweise bieten, bei welchem Kriterium ein System besonders schlecht oder gut ist“ [Lüd01]. Nachteilig ist hierbei allerdings, dass die Notengebung am Ende der Skala („sehr gut“ bzw. „sehr schlecht“) zu grob für eine genauere Urteilsabgabe sein kann. Aufgrund des geringen Sprachprobenumfangs kann der ACR-Test im Vergleich zu anderen Verfahren als effizient eingestuft werden, da nicht für jedes Kriterium ein neues Referenzsignal angeboten werden muss. Somit werden wesentlich weniger Sprachproben für eine Sprachgütemessung benötigt und der zeitliche Aufwand wird stark verkürzt.

Im Gegensatz zum Paarvergleichstest steht der Versuchsperson hier nur die zu bewertende Sprachprobe zur Verfügung. Diese wird anhand von kurzen und sinnvollen Sätzen dargeboten. Hierbei sollte auf mehrere Aspekte geachtet werden. Damit jeder Satz nur einmal zur Verfügung steht, müssen genügend Testsätze vorhanden sein. Dadurch soll einer Gewöhnung an das Testmaterial entgegen gewirkt werden. Sie sollten einen einheitlichen Sprachcharakter aufweisen und in ihrer Phonemhäufigkeitsverteilung so gut wie möglich der deutschen Sprache angepasst sein. Um eine sichere Beurteilung der Proben zu gewährleisten, dürfen die Sätze nicht zu kurz sein [Lüd01].

Bei der Bewertung ist die Testperson auf sich allein gestellt, da sie nicht den Vergleich zu einer Referenz hat. Somit trifft sie ihre Entscheidung nach allgemein empfundenen Qualitätsaussagen. In einem Opinientest, beschrieben in [Kla97], wurden den Probanden beispielsweise acht Kriterien dargeboten. Sie sollten die Gesamtqualität, die Höranstrengung, die Schwierigkeit des Verstehens, die Deutlichkeit der Aussprache, die Aussprachefehler, die Betonungsfehler, die

### 3. Auditive Testverfahren

Benotungsfehler, die Sprechgeschwindigkeit und die Annehmlichkeit der Sprache beurteilen.

Die Bewertung sollte laut ITU anhand einer fünfstufigen Skala mit den Noten „5“ für die beste Kategorie bis „1“ für die Schlechteste erfolgen. Die genaue Bezeichnung entspricht „5“ – Ausgezeichnet, „4“ – Gut, „3“ – Ordentlich, „2“ – Dürftig und „1“ – Schlecht [ITU800]. Die fünfstufige Skala soll Versuchspersonen ohne Erfahrungen die Bewertung erleichtern. Bei einer größeren Stufenzahl könnten sie in Entscheidungsschwierigkeiten geraten. Vorteil bei einer größeren Anzahl an Stufen wäre eine verfeinerte Einteilung der Bewertungsskala. Somit wäre eine differenziertere Urteilsabgabe der Testpersonen möglich. Am Lehrstuhl für Kommunikationstechnik an der Brandenburgischen Technischen Universität Cottbus wurde ein Test mit einer zehnstufigen Skala durchgeführt. Dieser Versuch wurde von den Probanden gut angenommen.

Eine Aufteilung der Skala in einen geradzahligen Bereich wäre vorzuziehen, da die Probanden somit angehalten sind, sich im mittleren Bereich differenzierter mit der Bewertung auseinanderzusetzen. Bei ungeradzahligen Skalen, wie in der ITU800 beschrieben, sind die Probanden gewillt, zum Mittelwert zu tendieren, wenn das Ergebnis nicht eindeutig scheint. Somit ist der Einsatz einer gradzahligen Skala zweckmäßiger [Lüd01].

Weiterhin sollte die Skala linear aufgeteilt sein, da die Probanden sich somit den Bewertungsbereich leichter einteilen und die empfundenen Qualitätsstufen leichter einordnen können. Die Auswertung erleichtert sich somit ebenfalls, da hier eine einfache Mittelwertbildung ausreicht. Es sollte trotzdem darauf geachtet werden, dass die Aufteilung der Beschreibung linear zur Notenskala gewählt wird, da die Qualitätsunterschiede zwischen zwei Beschreibungen größer oder kleiner als zwischen zwei anderen sein könnten.

Laut Lüdtker sind große Vorteile des ACR-Tests die gute Abbildung der Realität und die vergleichsweise einfache Versuchsstruktur [Lüd01]. Somit lassen sich der Test und die anschließende Auswertung sehr leicht automatisieren. Im Allgemeinen hat der Test eine relativ kurze Versuchsdauer und es können in kurzer Zeit viele Kriterien der Sprachqualität bewertet werden.

### 4. Internet-basierte Sprachgütemessung

Am Lehrstuhl Kommunikationstechnik der Brandenburgischen Technischen Universität in Cottbus wurden bereits einige Testverfahren zur Sprachgütemessung mit unterschiedlichen Techniken entwickelt und durchgeführt. Hierbei ist die Nutzung des Internets besonders zu erwähnen.

Ein großer Vorteil ist dabei, dass kein fester Ort für den Hörtest mehr von Nöten ist, da die Probanden den Test in ihrer häuslichen Umgebung oder am Arbeitsplatz durchführen können. Weiterhin kann positiv erwähnt werden, dass mit dem internet-basierten Verfahren wesentlich mehr Testperson in deutlich kürzerer Zeit erreicht werden können. Nachteilig ist jedoch, dass bei jeder Testperson andere Umgebungsbedingungen herrschen und die technische Ausstattung abweicht [Ric07].

Eines der ersten vom Lehrstuhl durchgeführten Testverfahren für das Internet war ein Bewertungstest. Der Test konnte auf statischen HTML-Seiten durchgeführt werden. Die wiederzugebenen Sprachproben wurden mit Hilfe des als Standard auf dem System festgelegten Players abgespielt. Hierbei sollte neben den Sprachproben die Akzeptanz der Benutzeroberfläche evaluiert werden. Im Anschluss an den Test hatten die Probanden die Möglichkeit, ihre Eindrücke auf vorgefertigten Fragebögen per Email abzugeben [Tei05]. Die Untersuchungen wurden mit ca. 50 Personen aus dem universitären Bereich und sonstigen Privatpersonen unterschiedlichen Bildungsgrades durchgeführt. Ein Vorteil der HTML-Version ist das nur geringe Datenaufkommen.

Während der Weiterentwicklung der internet-basierten Testverfahren wurde die Benutzeroberfläche für einen Hörtest geschaffen. Dieser realisierte die Speicherung der Daten mittels Cookies. Diese erlauben einem JavaScript-Programm, Daten auf der Festplatte des Anwenders zu speichern. Das Speichern geschieht hierbei in kontrollierter Form, so dass ein „Ausspionieren“ der Daten ebenso wenig möglich ist wie das Platzieren von Viren oder ähnlichen ausführbaren Codes.

Folgend wurde eine Benutzeroberfläche mit Hilfe von JavaScript-Funktionen entwickelt, wobei die Testparameter und Sprachproben aus JavaScript-Variablen geladen wurden. Eine weitere Veränderung war die Möglichkeit, die Farbe der

#### 4. Internet-basierte Sprachgütemessung

Oberfläche nach den individuellen Bedürfnissen anzupassen. Die Wiedergabe erfolgte, wie bereits bei der ersten Version beschrieben, mit Hilfe des gesetzten Standard-Players. Die abgegebenen Bewertungen wurden im Anschluss zur weiteren Verwendung per Email verschickt. Jedoch ist zu erwähnen, dass die Speicherung der Cookies auf eine Größenordnung von 4 KByte beschränkt und die Auswertung weiterhin manuell vorzunehmen war. Ein weiterer Nachteil dieses Verfahrens ist, dass Benutzer in ihrem Browser die Verwendung von JavaScript deaktiviert haben können und somit der Test nicht auf diesen Browsern ausgeführt werden kann.

In der Veröffentlichung von Herrn Richter [Ric07] ist weiter zu lesen, dass eine zweite Testoberfläche nach dem Client-Server-Prinzip um eine MySQL-Datenbank erweitert wurde. Mit Hilfe der Datenbank können die für den Test notwendigen Daten gespeichert werden und mittels PHP erfolgt aus diesen Daten die Generation dynamischer Webseiten. Somit können die vom Benutzer abgegebenen Ergebnisse direkt in die MySQL-Datenbank eingetragen werden. Durch die Datenbanktechnologie für internet-basierte Testverfahren können große Datenmengen gespeichert und zentral erfasst werden. Hierdurch entfällt die manuelle Auswertung, da es möglich ist, auf alle abgegebenen Bewertungen eine sofortige Analyse, von Grafiken oder statistische Methoden, anzuwenden. Die Wiedergabe erfolgte ebenfalls mit dem Standard-Player des Systems. Nachteilig an dem Verfahren mit Datenbanken sind eine deutliche Erhöhung des Datenverkehrs und die ständige Internetverfügbarkeit.

Gleichzeitig wurde ein Paarvergleichstest unter Verwendung der Datenbank-Technologie erstellt. Er wurde in die gleiche Datenbank wie der Bewertungstest integriert. Bei diesem Test werden dem Benutzer Sprachprobenpaare dargeboten, und er muss entscheiden, welche für das gewählte Kriterium besser oder schlechter ist.

Durch die Wiedergabe der Sprachproben mit Hilfe des Standard-Players traten oft Probleme auf. Ein Fall ist beispielsweise die jeweils unterschiedliche grafische Oberfläche bei verschiedenen Playern. Ein weiterer Nachteil ist die Wiedergabe der Proben als Livestream. Hierbei werden die Daten während der Wiedergabe vom Server geladen. Hierdurch kann es zu nicht korrekt abgespielten Audiodateien kommen, wenn ein Kommunikationsproblem mit dem Server auftritt. Dies kann sich unter Umständen in einer verfälschenden Bewertung des Benutzers niederschlagen.

#### 4. Internet-basierte Sprachgütemessung

Laut Richter treten ebenfalls Probleme auf, wenn eine Sprachprobe in einem nicht unterstützten Format auf dem Server vorliegt. In diesem Fall erhält man keine Rückmeldung, und ein unterbrechungsfreier vollständiger Test ist nicht mehr möglich [Ric07].

Später wurde am Lehrstuhl ein MOS-Test auf der Basis der beschriebenen Technologien umgesetzt und mit einem unter Studiobedingungen durchgeführten Test verglichen. Hierbei wurden Sprachsyntheseverfahren analysiert. Das Studioergebnis konnte nicht exakt reproduziert werden, da die Versuchsbedingungen der Probanden nicht identisch waren (unterschiedliche Hardware- und Umgebungsbedingungen). Dennoch ist zu erwähnen, dass die Teilnehmer, die den Versuch mit Kopfhörern durchgeführt hatten, annähernd an das Studioergebnis herangekommen sind [Cas06]. Aufgrund der vielen Störeinflüsse und unterschiedlichen Randbedingungen kann man schlussfolgern, dass bei internet-basierten Testverfahren immer eine große Anzahl von Versuchspersonen benötigt wird, damit durch eine Vielzahl von Bewertungen Störeinflüsse durch Mittelung der Ergebnisse minimiert werden.

# 5. Erstellung der akustischen Umgebungen

## 5.1 Herangehensweise

In den folgenden Abschnitten soll die Herangehensweise zur Erstellung der akustischen Umgebungen näher erläutert werden. Hierfür wird auf das räumliche Hören, Schall in Räumen, verschiedene Aufnahmeverfahren und das Aufnahmegerät eingegangen.

### 5.1.1 Räumliches Hören

Das räumliche Hören ist bei dieser Arbeit hervorzuheben, da nur so ein möglichst gutes und für den Probanden reproduzierbares virtuelles Szenario entstehen kann. Durch die räumliche Wahrnehmung des Menschen ist es möglich mit den Augen und den Ohren Informationen über die räumlichen Gegebenheiten aufzunehmen. Sehr genau ist das Richtungshören bei Menschen in Blickrichtung. Hierbei erreichen die Schallwellen beide Ohren zur gleichen Zeit und mit gleicher Intensität. Durch das Hören mit zwei Ohren sind bereits sehr kleine Richtungsunterschiede wahrnehmbar [Mey79]. Bei seitlich eintreffenden Schallwellen können Richtungsinformationen durch Phasen-, Pegel- und Klangunterschiede erkannt werden. Dies beruht auf der menschlichen Physiologie, denn trifft eine Schallwelle seitlich auf den Kopf, so ist der Weg zum abgewandten Ohr länger und es tritt ein Laufzeitunterschied auf. Des Weiteren treten durch Beugung an den Ohrmuscheln Frequenzänderungen auf. Durch charakteristische Stärkung und Schwächung einzelner Frequenzbereiche können die verschiedenen Richtungen unterschieden werden [Mey79] [Gör07]. In Abbildung 4 sind diese sogenannten *richtungsbestimmenden Bänder* veranschaulicht [Dic03].

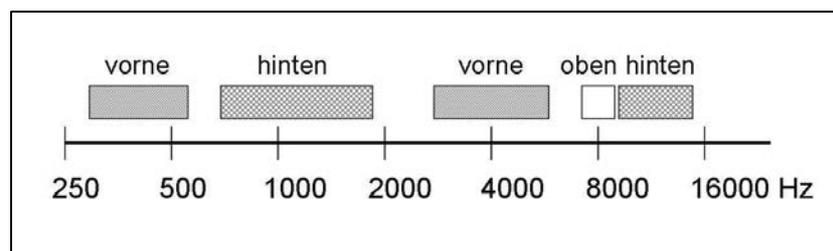


Abbildung 4: Richtungsbestimmende Bänder [Dic03]

## 5. Erstellung der akustischen Umgebungen

Ein weiterer wichtiger Punkt des räumlichen Hörens ist das Entfernungshören. In geschlossenen Räumen kann anhand der Nachhallstruktur ein Rückschluss auf die Entfernung des Schallsignals getroffen werden. Hierzu werden die zeitlichen Abstände der aus allen Raumrichtungen kommenden Wand- und Deckenreflektionen zum Direktschall ausgewertet [Gör07]. Im Freien spielt bei größeren Entfernungen die Dämpfung höherer Frequenzen durch Luftabsorption eine wichtige Rolle. Kürzere Distanzen von etwa 3 bis 15 Metern können durch Verringerung des Schallpegels wahrgenommen werden [Dic03].

### 5.1.2 Schall in Räumen

In geschlossenen Räumen treten Reflektionen, Beugung und Brechung des Schalls auf. Bei der Reflektion wird eine Schallwelle an einer Wand reflektiert. Hierbei ist der Einfallswinkel gleich dem Ausfallswinkel. Beim natürlichen Hören wird meist zuerst die Originalschallquelle gehört und anschließend die im Raum auftretenden Reflektionen. Der so entstehende Nachhall hilft dem Menschen bei der Einschätzung von Raumgrößen.

Wenn Schallwellen auf ein relativ großes Hindernis treffen, kann man eine Schattenwirkung beobachten. Der Frequenzbereich von 100 Hz bis 10 kHz entspricht einem Wellenlängenbereich von etwa 3 m bis 3 cm. In diesem Größenbereich liegen die meisten Gegenstände im Haushalt, und somit ist die Ausbreitung von Schallwellen im Wesentlichen von der Beugung bestimmt [Vei05].

Bei der Brechung treten Schallwellen von einem Medium in ein anderes ein. In diesem Fall werden die Wellen nicht nur reflektiert, sondern breiten sich im zweiten Medium mit einer Richtungsänderung aus.

Die genannten Merkmale des Schalls helfen dem Menschen sich in seiner Umgebung zu orientieren und diese wahrzunehmen. Zu den akustischen Eigenschaften eines Raumes zählen in erster Linie die Schallreflektionen an den Begrenzungsflächen. Nach einer akustischen Erregung verschwindet das Schallfeld nicht sofort; es klingt nach einer exponentiellen Zeitfunktion ab. Dieser Nachhall entsteht durch wiederholte Reflektionen an den Raumbegrenzungen. Bei jeder

## 5. Erstellung der akustischen Umgebungen

Reflektion findet eine Absorption der Schallenergie statt [Vei05]. Die Nachhallzeit eines Raumes kann man anhand der Gleichung 5.1 bestimmen.

$$T = 0,16 \frac{V}{\alpha * S} = 0,16 \frac{V}{\sum_n a_n * S_n} \quad (5.1)$$

V ist hierbei das Raumvolumen in m<sup>3</sup>, S<sub>n</sub> die Teilflächen in m<sup>2</sup> und α<sub>n</sub> ist der Schallabsorptionsgrad der Teilflächen. An einer schallharten Wand ist der Absorptionsgrad α = 0 und beispielsweise bei einem offenen Fenster gleich 1.

### 5.1.3 Aufnahmeverfahren

Zur realistischen Erzeugung einer für den Menschen gewohnten akustischen Umgebung muss dem räumlichen Hören nachgekommen werden. Somit kommen für diese Arbeit nur Aufnahmetechniken in Frage, die auf der Stereophonie beruhen. „Stereophonie („körperlicher Klang“) bezeichnet jedes Übertragungsverfahren mit mindestens zwei Kanälen im Gegensatz zur Monophonie mit nur einem Kanal“ [Gör07].

Es gibt drei Arten der Stereophonie. Die Hauptformen sind hierbei zum einen die raumbezogene und zum anderen die kopfbezogene Stereophonie. Das raumbezogene Verfahren ist dem Wiedergaberaum fest zugeordnet. Somit wird vom Hörer eine Ausrichtung auf die Lautsprecherbasis verlangt. Beim kopfbezogenen Verfahren wird das Klangbild von einem Kunstkopf aufgenommen und über Kopfhörer wiedergegeben. Für beide Verfahren gilt, dass die Stereosignale durch mindestens zwei Mikrofone erfasst werden [Dic03].

Bei Mischung beider Verfahren spricht man von der Äquivalenzstereomikrofonie. Anhand der Kombination von Lokalisationsschärfe und hoher Räumlichkeit, kann ein sehr natürliches Abbild der Schallquelle reproduziert und ein noch besseres Klangbild erzeugt werden.

Zur akustischen Stereobildgewinnung im raumbezogenen Verfahren finden zwei Aufnahmetechniken Anwendung: zum einen die Pegel- oder Intensitätsstereophonie und zum anderen die Laufzeitstereophonie.

## 5. Erstellung der akustischen Umgebungen

Auf das Laufzeitverfahren soll hier nur kurz eingegangen werden, da die Intensitätsstereofonie für diese Arbeit besser geeignet ist. Das Verfahren beruht auf Laufzeitdifferenzen zwischen zwei Mikrofonen. Signale ohne Laufzeit-unterschiede werden hierbei in der Mitte abgebildet. Die Unterschiede in der Laufzeit sind auf unterschiedliche Entfernungen der Schallquelle zu den Aufnahmemikrofonen zurückzuführen. In Abbildung 5 ist die Laufzeitmikrofonierung nach dem *AB-Mikrofonverfahren* bildlich dargestellt. Demnach erreichen Schallquellen von rechts das Mikrofon auf der rechten Seite zuerst und anschließend das linke Mikrofon. Hierdurch tritt eine Laufzeitverschiebung auf.

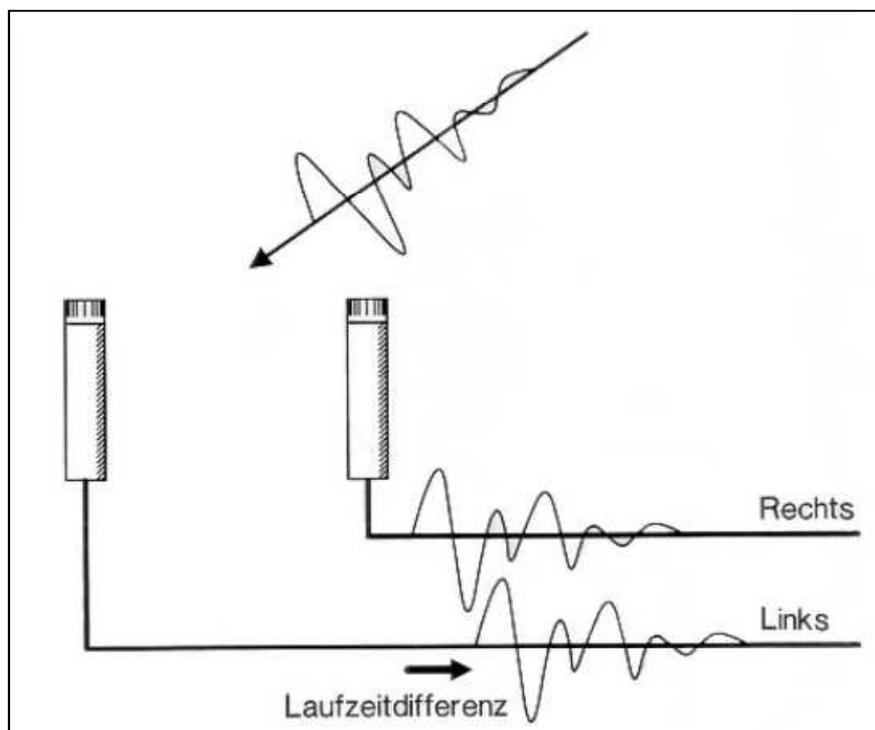


Abbildung 5: AB-Mikrofonierung [Gör94]

Beim Laufzeitmikrofonverfahren vermittelt das Stereoklangbild einen guten Raumeindruck. Jedoch sind die Ortbarkeit und Richtungsauflösung nicht so gut wie bei der Intensitätsstereofonie [Dic03].

Beim Intensitätsverfahren unterscheiden sich die Signale nur im Pegel und nicht in der Phasenlage. Somit ist ein Signal aus der linken Richtung aufgenommen auf dem rechten Kanal leiser als auf dem Linken. Dieses Verfahren beruht auf der Hörpsychologie. Wie bereits unter Punkt 5.1.1 beschrieben wurde, werden bei seitlich eintreffenden Schallsignalen die Pegelunterschiede zwischen rechtem und linkem Ohr ausgewertet [Gör07]. Laut Görne liefert die Intensitätsstereofonie

## 5. Erstellung der akustischen Umgebungen

Aufnahmen mit sehr präziser Richtungsabbildung der Schallquellen. Dieses Verfahren lässt sich mit Hilfe des *MS-* und *XY-Mikrofonverfahrens* oder des *Einzelmikrofonverfahrens* erzeugen. Beide Verfahren werden im Verlauf vorgestellt.

Beide Mikrofonverfahren sind theoretisch gleichwertig, jedoch ergeben sich in der praktischen Anwendung durch die Frequenzgänge der Richtcharakteristiken der Mikrofone sowie durch unterschiedliche Möglichkeiten der Fernbedienbarkeit erhebliche Unterschiede. Die MS- und XY-Mikrofonverfahren weisen keinerlei Laufzeitunterschiede auf. Sie zeigen nur Pegelunterschiede auf, die abhängig von der Einfallsrichtung des Schalls, von den eingestellten Richtcharakteristiken und den Vorverstärkern sind. Beim XY-Verfahren werden direkt die Signale für den linken Kanal (X) und für den rechten Kanal (Y) geliefert. Das MS-Verfahren ist etwas komplizierter, da hier erst durch Summen- bzw. Differenzbildung die Richtungsinformationen ausgewertet werden können. Beim Einzelmikrofonverfahren (Polymikrofonie) werden Mikrofone direkt an der Schallquelle aufgestellt, so dass möglichst wenig Übersprechen von anderen Quellen sowie wenig Diffusschall aufgenommen wird [Dic03].

Das MS-Mikrofonverfahren ist ein Hauptmikrofonverfahren in reiner Intensitätsstereofonie und arbeitet mit einem Koinzidenzmikrofon oder entsprechend angeordneten Einzelmikrofonen. M steht hierbei für das Signal aus der Mitte und S für die Signale von den Seiten. Die seitlich eintreffenden Signale werden von einem Mikrofon mit Achterrichtcharakteristik aufgenommen. Die Charakteristik für das Mikrofon des Mittensignals ist hierbei nicht festgelegt. In Abbildung 6 ist das MS-Verfahren mit einer Kugel- und einer Nierencharakteristik als M-Mikrofon zu sehen.

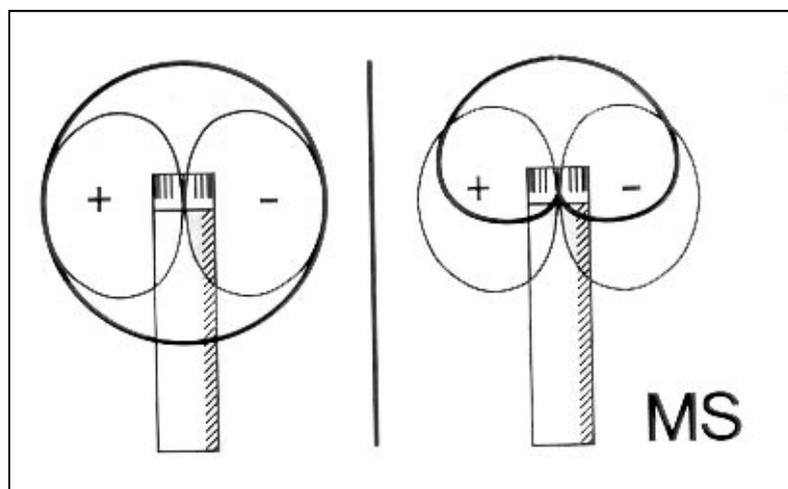


Abbildung 6: MS-Mikrofonverfahren mit M als Kugel links und M als Niere rechts [Gör94]

## 5. Erstellung der akustischen Umgebungen

Das XY-Mikrofonverfahren ist wie das MS-Verfahren ein Hauptmikrofonverfahren bei dem ein Koinzidenzmikrofon oder zwei entsprechend angeordnete Einzelmikrofone Verwendung finden.

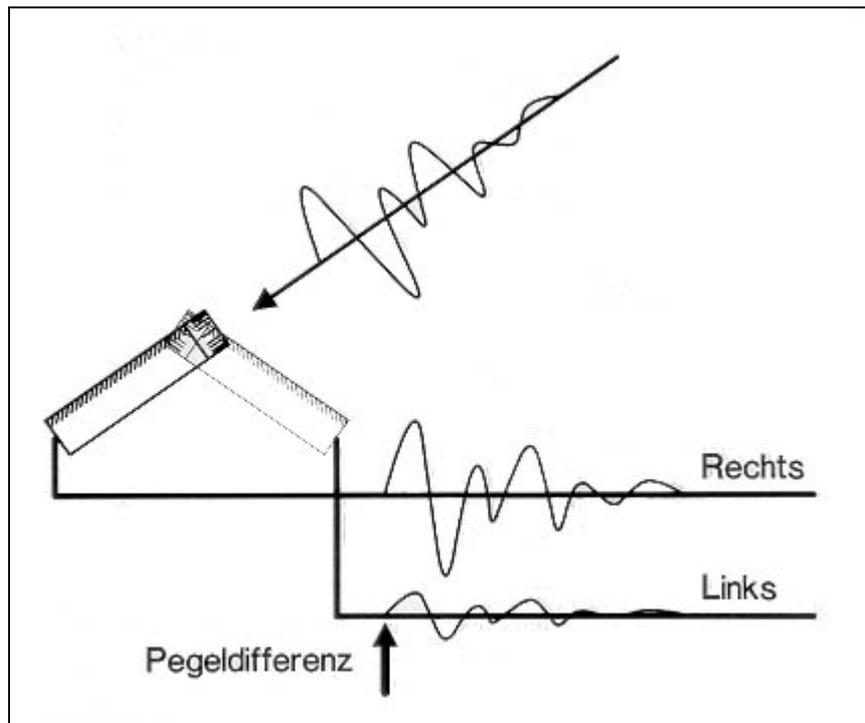


Abbildung 7: XY-Mikrofonierung [Gör94]

Das nach links zeigende Mikrofon gibt den linken Kanal wieder und das nach rechts zeigende den rechten Kanal. Wie in Abbildung 7 zu erkennen ist, werden durch Drehung der Mikrofonkapseln im gleichen Öffnungswinkel nach außen, die notwendigen Pegeldifferenzen durch gleiche Richtcharakteristiken, der Mikrofonkapseln, erreicht. Für Nierencharakteristik ist ein Winkel von  $\pm 135^\circ$ , für Superniere  $\pm 115^\circ$  und für Hyperniere  $\pm 105^\circ$  notwendig. Somit wird für die XY-Mikrofonierung immer ein besonders großer Aufnahmebereich erzeugt. In der Praxis sind meist zwei Mikrofone mit Nierencharakteristik im  $90^\circ$  Winkel übereinander angeordnet. Somit entsteht ein Aufnahmebereich von etwa  $135^\circ$ .

Der somit erzeugten Aufnahme fehlt jedoch die Verzerrung des Schallfeldes durch den Kopf, die für die Lokalisierung in der Medianebene benötigt wird. Schlussfolgernd eignet sich eine einfache Stereoaufnahme nicht für die vorliegende Arbeit, da nur eine ungefähre Ortbarkeit und Positionsbestimmung in der horizontalen Ebene möglich ist.

## 5. Erstellung der akustischen Umgebungen

Die kopfbezogene Stereophonie (Kunstkopfstereophonie) verbindet sowohl Laufzeit- als auch Intensitätsverfahren in der horizontalen Wiedergaberichtung. Sie ist eine sehr einfache und gleichzeitig naheliegende Möglichkeit, ein stereofones Klangbild hörgetreu zu übertragen [Vei05]. Die vertikale Ebene wird, wie beim natürlichen räumlichen Hören anhand von Frequenzgangunterschieden erzeugt. Das Aufnahmesystem ist eine Nachbildung eines menschlichen Kopfes mit Mikrofonen anstelle der Ohren, welche den Klang unmittelbar vor den Ohren des Hörers reproduzieren. Für die Kunstkopfstereophonie ist zur Wiedergabe ein Kopfhörer erforderlich. Die Aufnahme kommt hierbei sehr nah an das Originalklangbild heran, da die Schallquellen in allen Richtungen sehr gut lokalisierbar sind und eine gute Entfernungswahrnehmung sowie ein sehr natürlicher Raumeindruck entstehen. Abbildung 8 zeigt den Kunstkopf der Firma Neumann KU 100 mit Innenohr-mikrofonen.



Abbildung 8: Kunstkopf KU 100 der Firma Neumann mit Innenohr-Mikrofonen [Neu11]

Die Hauptanwendung der Kunstkopfstereophonie sind raumakustische Untersuchungen, wodurch sich dieses Verfahren am besten für die vorliegende Arbeit eignet. Ebenfalls sehr gut geeignet ist dieses Verfahren aufgrund seines akustisch-realistischen Klangbildes für Hörspielproduktionen [Dic03].

## 5. Erstellung der akustischen Umgebungen

### 5.1.4 Aufnahmegerät - Kunstkopf

Wie bereits in Punkt 5.1.3 erwähnt eignet sich die kopfbezogene Stereophonie am besten für die durchzuführenden Untersuchungen, da die Wiedergabe einer einfachen Mikrofonaufnahme für die Beurteilung eines Geräusches nicht dem hohen Höranspruch des Versuches genügen würde. Nur durch die Kunstkopftechnik ist man in der Lage, eine Umgebung so aufzunehmen und wiederzugeben, wie es ein Hörer selbst im entsprechenden Schallfeld erleben würde, da der Kunstkopf das Schallfeld ähnlich wie eine Person verzerrt. Diese Verzerrungen treten, wie bereits erwähnt, als richtungsabhängige und richtungsunabhängige Komponenten auf. Die unabhängigen Anteile durchlaufen den Weg zum Trommelfell direkt durch die Ohrmuschelhöhle und den Ohrkanal. Auf die richtungsabhängigen Anteile der Schallwellen haben die äußeren Körperpartien, wie Schulter, Kopf und Ohr Einfluss. An diesen Stellen werden die Wellen reflektiert und gelangen erst dann in die Gehörmuschel und den Gehörgang. Hierbei verändern sich die Klangfarbe und der Schallpegel, wodurch eine Ortbarkeit möglich wird.

Um diesen anatomischen Gegebenheiten auch bei der Wiedergabe Sorge zu tragen, muss die Aufnahme vor der Darbietung entsprechend entzerrt werden. Dies ist nötig, da die Kopfhörermembranen nicht direkt an den Trommelfellen der Testperson platziert sind und die Schallwellen noch die Ohrmuschelhöhle und den Ohrkanal durchlaufen müssen. Diese Eigenschaften lassen sich für die Wiedergabe herausrechnen, sodass die Kunstkopfaufnahme gehör richtig angehört werden kann (Abbildung 9).

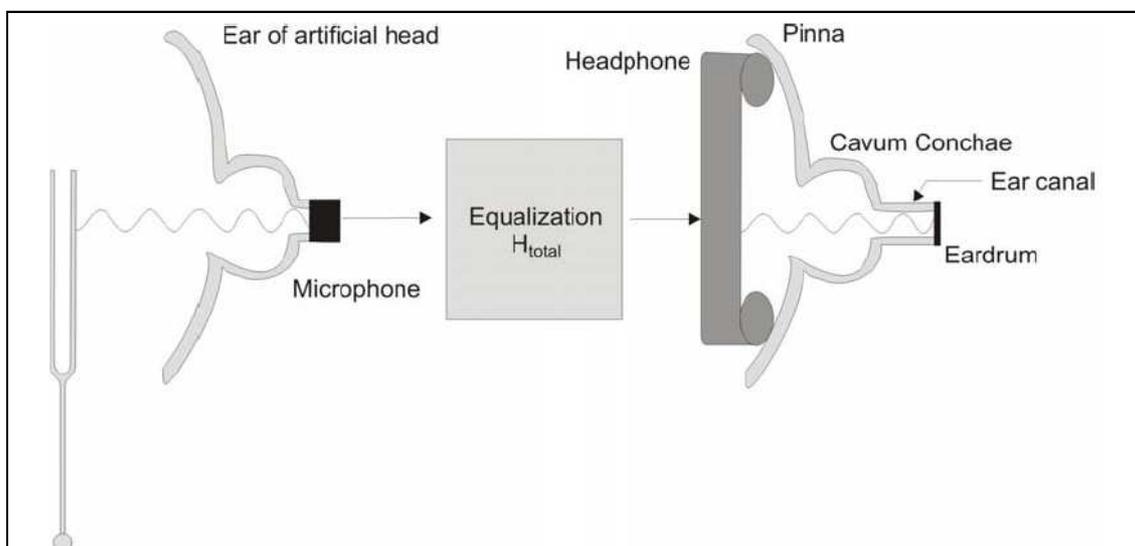


Abbildung 9: Entzerrung einer Kunstkopfaufnahme für die korrekte Wiedergabe [Hea11]

## 5. Erstellung der akustischen Umgebungen

*Head acoustics* bietet für die „Equalization“ drei Aufnahme-Entzerrungen an. Die „Freifeld“ (FF)-, die „Diffusfeld“ (DF)- und die „independent of direction“ (ID)-Entzerrung. Beim Freifeld und beim Diffusfeld handelt es sich um Schallfelder mit exakt festgelegten Rahmenbedingungen, welche in der Realität allerdings nur selten vorliegen. Bei diesen Entzerrungen werden die richtungsabhängigen Anteile aus dem Signal gefiltert, um die Aufnahme beispielsweise mit einer Mikrofonaufnahme vergleichbar zu machen. Die ID-Entzerrung filtert durch mathematische Berechnungen nur die richtungsunabhängigen Anteile aus dem Kunstkopfsignal heraus [Hea11].

Auf Grund der technischen Voraussetzungen an der Brandenburgischen Technischen Universität kann jedoch die beschriebene ID-Filterung nicht für den Versuch angewandt werden. Das Einbinden des entsprechenden Wiedergabegerätes zur Filterung erzeugte starke akustische Störungen des Signals, die für den Versuch nicht hinnehmbar waren.

Laut der Application Note von *Head acoustics* muss besondere Aufmerksamkeit bei der Versuchsdurchführung und Auswertung auf die verschiedenen Einflüsse, die der Proband aufnimmt, gelegt werden. Hierzu zählen die Umgebung des Versuchs, die Erwartungshaltung und die Wiedergabereihenfolge der Hörbeispiele. Die Testperson würde sofort feststellen, wenn sich die raumakustischen Eigenschaften des Aufnahmeortes und des Wiedergabeortes unterscheiden. Beim ersten Anhören einer Kunstkopfaufnahme wird die ungeübte Versuchsperson verunsichert, da sie eine Diskrepanz zwischen dem aktuellen und dem gehörten Raum wahrnimmt. Im vorliegenden Fall sitzt die Versuchsperson in einer schalldichten Kabine und hört verschiedene Raumszenarien, die nicht zu der Kabine passen. Dies kann zu einer Fehlbewertung der Umgebung führen. Diesen Einfluss könnte man verhindern, indem man den Versuch am Aufnahmeort durchführt oder eine geeignete Trainingsphase vor dem Test ausführt [Hea08].

Verschiedene Veröffentlichungen zeigten, dass die Erwartungshaltung an die Umgebung die Bewertung zu einem großen Teil beeinflusst. So führte der Vergleich von Bewertungen an Originalorten und in Abhörräumen zu deutlichen Unterschieden. Die richtige Umgebung sorgt für eine angemessene Erwartungshaltung beim Hören und somit für eine realitätsnahe Beurteilung. Um hier Abhilfe zu schaffen, könnte durch eine geeignete Visualisierung dem Probanden der Originalort dargeboten

## 5. Erstellung der akustischen Umgebungen

werden [Hea08]. In der vorliegenden Arbeit ist dies nur schwer möglich, da hierfür ein zweiter Monitor in der schalldichten Kabine installiert werden müsste und dieser wahrscheinlich eher vom Versuchsablauf ablenken würde.

Ein weiterer Einflusspunkt kann das Empfinden einer realitätsfremden Situation sein. Im Alltag wird ein Geräusch meist nur nebenbei wahrgenommen, und man hat nicht die Möglichkeit, sich ausschließlich dem Geräuschszenario zu widmen. Im vorliegenden Versuch, in dem sich die Versuchsperson hauptsächlich auf das Szenario konzentrieren soll, wird das Geräusch vermutlich anders bewertet als in einer realen Situation, wenn es nur unterschwellig erlebt und durch Gewohnheit vom Gehirn ausgeblendet wird [Hea08]. Diese Punkte zeigen, dass genaue Aufmerksamkeit bei der Auswertung sehr wichtig sein wird.

Zur Aufnahme der Beispielszenarien wurde der Kunstkopf HMS II.3 der Firma Head acoustics benutzt, welcher in Abbildung 10 zu sehen ist.



Abbildung 10: Kunstkopf HMS II.3 der Firma Head acoustics

Wie bereits beschrieben, eignet sich dieser Kunstkopf besonders durch seine anatomischen Gegebenheiten, die ebenfalls die Schulterpartien nachbilden. Dies ist nicht bei allen Kunstkopfherstellern berücksichtigt worden, wie bereits das Beispiel von Neumann zeigt. Mit dem Kunstkopf II.3 können sowohl ohrnahe Schallquellen, wie Kopfhörer, Handapparate, Headsets oder Gehörschütze sowie ohrferne Schallquellen, wie zum Beispiel Freisprecheinrichtungen oder im vorliegenden

## 5. Erstellung der akustischen Umgebungen

Beispiel akustische Umgebungen getestet und gemessen werden. Der Übertragungsbereich beträgt 3 Hz – 20 kHz mit einer Dynamik >110 dB.

Der Kunstkopf wurde für das jeweilige Szenario in den Raum gebracht und an der Stelle installiert, an der auch der Proband sitzen würde. Somit konnten sehr authentische akustische Umgebungsbedingungen für den Test geschaffen werden.

Die Aufnahme erfolgte mit dem in Abbildung 11 gezeigtem mobilen Aufnahmegerät H2 der Firma Zoom.



Abbildung 11: Mobiles Aufnahmegerät H2 der Firma Zoom [Zoo11]

Dieses Gerät bietet die Möglichkeit, das zweikanalige Ausgangssignal des Kunstkopfes per Line-Eingang auf einer SD-Karte aufzunehmen und abzuspeichern. Die Szenarien werden mit einer Quantisierung von 24 bit und einer Abtastrate von 96 kHz erstellt.

Bei jeder Aufnahme des Kunstkopfes wird mit dem Schallpegelmesser 2237 Controller der Firma Brüel & Kjaer der entsprechende Schallpegel der Umgebung gemessen. Das Gerät ist in Abbildung 12 zu sehen.



Abbildung 12: Schallpegelmesser der Firma Brüel & Kjaer

## 5. Erstellung der akustischen Umgebungen

Vor der Messung wird der Messbereich eingestellt. Dieser lässt sich für verschiedene Situationen zwischen 30 bis 100 dB, 50 bis 120 dB und 70 bis 140 dB anpassen. Durch diese Anpassung kann man eine genauere Messung vornehmen und trägt Sorge dafür, dass die Messung nicht übersteuert. Zur Aufnahme der Szenarien wird der Wertebereich von 30 dB bis 100 dB gewählt, da die Umgebungen teilweise sehr ruhig sind und keine höhere Lautstärke als 50 dB erreichen werden. Die Messung wird in einem Zeitintervall von 10 Sekunden aufgenommen und anschließend der Schallpegel ausgegeben. Dieser Wert steht für den Pegel, den der Proband in der Umgebung wahrnehmen würde.

Bei verschiedenen Messungen ist immer darauf achtzugeben, dass im gleichen Abstand gemessen wird, da der Schallpegel stark von der Entfernung der Schallquelle abhängig ist. Im vorliegenden Fall wurde am Kunstkopf gemessen, um den tatsächlichen Pegel an den Ohren des Probanden zu erhalten.

### 5.2 Umgebungen

Jede Umgebung wird mit einem Foto veranschaulicht, auf dem der Kunstkopf an seinem Testplatz zu sehen ist. Diese Position würde im Versuch der Position des Probanden entsprechen. Desweiteren werden die Umgebungsbedingungen zur besseren Vorstellung beschrieben. Um einen guten akustischen Eindruck des Raumes zu bekommen, wird jeweils die Nachhallzeit der einzelnen Umgebungen anhand der Gleichung 5.1 berechnet und angegeben. Zur Vereinfachung werden die Schallabsorptionsgrade der einzelnen Materialien der Testräume für eine Frequenz von 1 kHz betrachtet und etwaige Möblierung vernachlässigt, da diese in den vorliegenden Fällen keinen großen Einfluss auf die Dämpfung der Nachhallzeit aufweisen würde.

Anhand eines Spektrogrammes wird jeweils die akustische Umgebung veranschaulicht. Ein Spektrogramm stellt das Leistungsspektrum eines Signals zeitlich dar. Die Frequenz ist logarithmisch auf der Ordinatennachse von 40 Hz bis 22,1 kHz aufgetragen. Die Signale werden im Bereich zwischen -90 und 0 dB durch die Farben Blau, Gelb und Rot kenntlich gemacht. Durch diese Kennzeichnung wird jeweils die Stärke der entsprechenden Frequenz dargestellt. Bereiche mit nur geringen Anteilen werden blau abgebildet. Frequenzanteile im mittigen Pegelbereich

## 5. Erstellung der akustischen Umgebungen

werden gelb und Frequenzanteile mit dem größten Pegel - 0dB - rot angezeigt. Durch diese Farbzuoordnung kann bei der Analyse genau auf störende Bereiche eingegangen werden.

Die im Folgenden zu sehenden Spektrogramme wurden bei der Originallautstärke der einzelnen Szenarien erstellt.

### 5.2.1 Szenario 1 - „Zimmer an Straße“



Abbildung 13: Szenario 1 - "Zimmer an Straße"

Die erste Umgebung demonstriert ein Zimmer an einer befahrenen Straße. Damit die Verkehrsräusche besser wahrgenommen werden können, erfolgt die Aufnahme des Kunstkopfes bei geöffnetem Fenster. Das Fenster befindet sich, wie in Abbildung 13 zu erkennen ist, hinter dem Probanden. Die Schallwellen erreichen den Kunstkopf mit einem Schallpegel von etwa 45 dB. Durch das Eintreffen der Schallwellen von hinten, wird das Verkehrsräusch nur sehr leise und indirekt

## 5. Erstellung der akustischen Umgebungen

wahrgenommen. Die Testumgebung hat eine Fläche von etwa 15 m<sup>2</sup> und eine Höhe von 2,90 m. Von den 45,9 m<sup>2</sup> Wandfläche sind 2,4 m<sup>2</sup> Fensterfläche vorhanden. Das Zimmer ist mit einem dämpfenden Teppichbelag ausgestattet, und die Wände sind mit Raufaser versehen. Die Nachhallzeit dieses Raumes beträgt bei 1 kHz etwa 0,38 Sekunden. Demzufolge ist in diesem Raum nur sehr gering mit störenden Reflektionen zu rechnen.

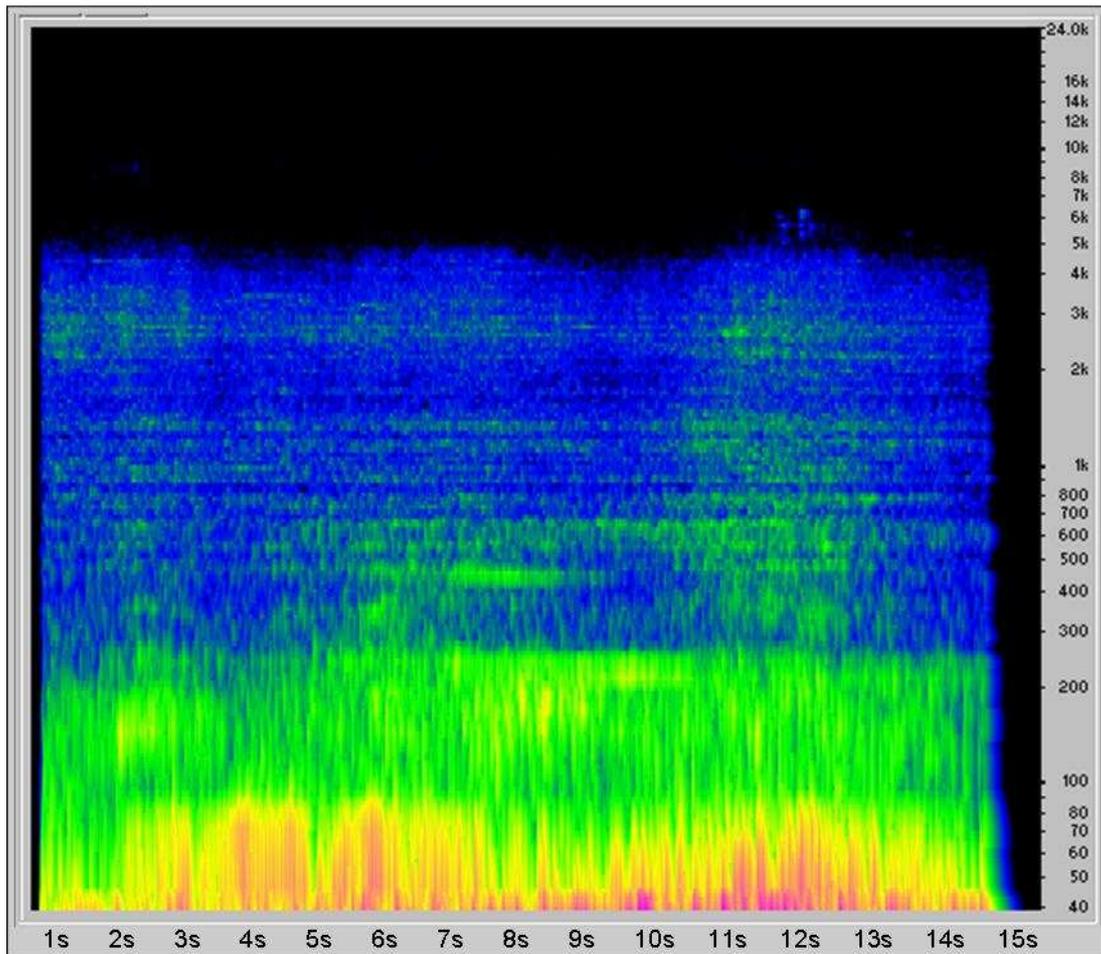


Abbildung 14: Spektrogramm Szenario 1

Das Spektrogramm in Abbildung 14 zeigt die Umgebungsgeräusche des Szenario 1. Der Verkehrsschall ist weitestgehend nur im Frequenzbereich zwischen 20 und 250 Hz zu finden. Die ansteigenden tieffrequenten Anteile im Bereich von 20 bis 80 Hz treten jeweils beim Vorbeifahren eines Fahrzeuges auf. Dies ist bei 3 bis 6 und 11 bis 13 Sekunden wahrnehmbar. Ab 250 Hz ist das Eigenrauschen des Raumes erkennbar.

## 5. Erstellung der akustischen Umgebungen

### 5.2.2 Szenario 2 - „Zimmer - nebenan Küchengeräusche und TV“



Abbildung 15: Szenario 2 - "Zimmer nebenan Küchengeräusche und TV"

In der zweiten Umgebung wurde ein Abhörort demonstriert, der ebenfalls einer Wohnumgebung entspricht. Hier treten im angrenzenden Zimmer (der Küche) Fernseh- sowie Küchengeräusche auf. Die Schallwellen erreichen den Kunstkopf mit einem Schallpegel von etwa 43 dB von hinten. Somit wird das Umgebungsgeräusch nur sehr leise und indirekt wahrgenommen, wie es bereits in der Testumgebung 1 der Fall war. Das Testzimmer hat eine Fläche von etwa 15 m<sup>2</sup> und eine Höhe von 2,90 m. Von den 47,6 m<sup>2</sup> Wandfläche sind 2,4 m<sup>2</sup> Fensterfläche vorhanden. Das Zimmer ist mit einem Teppichbelag ausgestattet und die Wände sind mit Raufaser tapeziert. Die Nachhallzeit dieses Raumes beträgt bei 1 kHz etwa 0,38 Sekunden. Demzufolge ist dieser Raum sehr ruhig, und es treten nur sehr gering störende Reflektionen auf.

## 5. Erstellung der akustischen Umgebungen

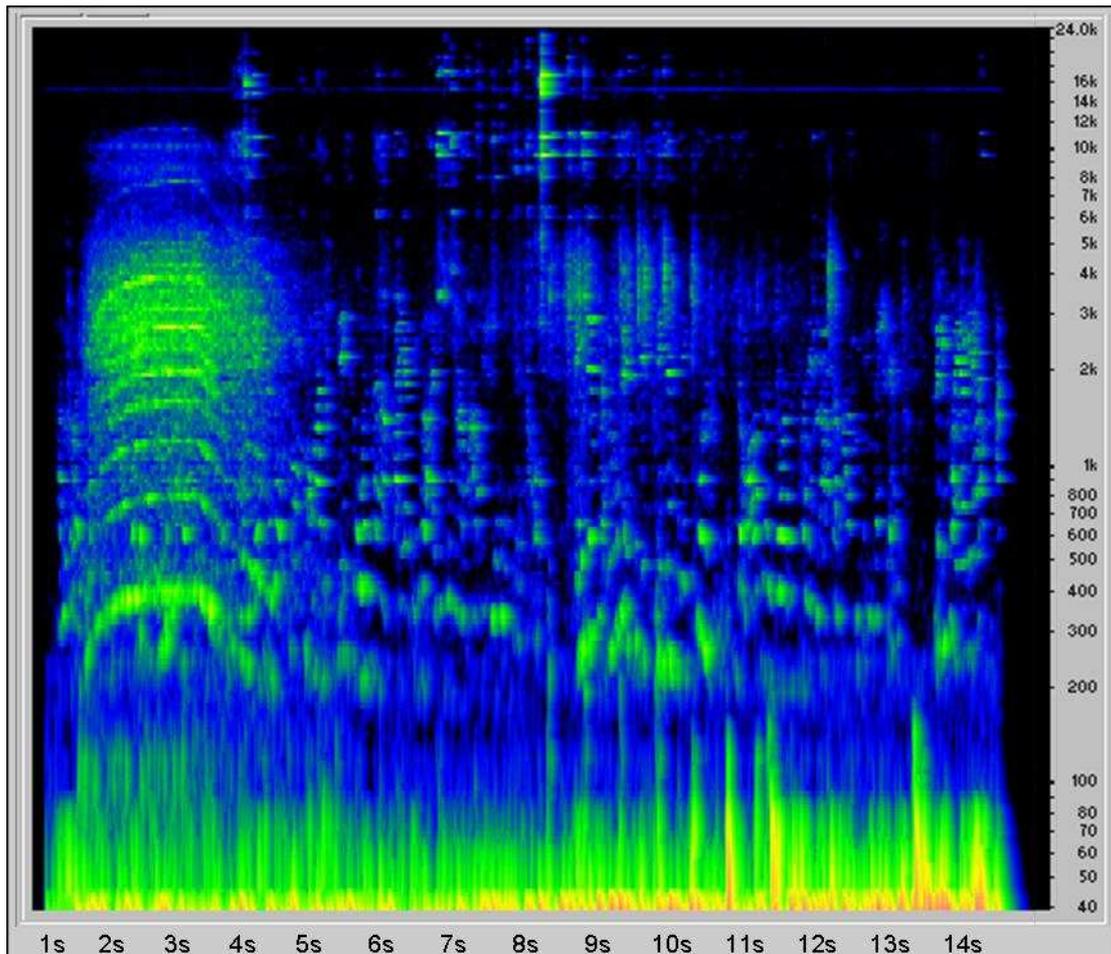


Abbildung 16: Spektrogramm Szenario 2

In Abbildung 16 ist das Spektrogramm der Umgebung 2 dargestellt. Es zeigt die Frequenzen eines laufenden Fernsehers, Geklapper eines Schlüssels und das Bedienen einer Brot Schneidemaschine. Es ist sehr gut zu erkennen, dass die Frequenzen im Bereich von 20 bis etwa 50 Hz mehr hervortreten, als alle anderen. Im Bereich von 50 bis 100 Hz werden die Geräusche mit einem geringen Pegel relativ konstant abgebildet. Ein Frequenzloch herrscht etwa zwischen 100 und 200 Hz. Nur sehr gering treten Geräuschanteile im Frequenzbereich zwischen 200 Hz und 7 kHz auf. Dies ist vor allem die Brot Schneidemaschine, die zu Beginn für 5 Sekunden betätigt wird. Bei 8,5 Sekunden kann man das hochfrequente Klappern des Schlüssels sehr gut erkennen.

## 5. Erstellung der akustischen Umgebungen

### 5.2.3 Szenario 3 - „Raum 321 mit Baustellenlärm“



Abbildung 17: Szenario 3 - " Raum 321 mit Baustellenlärm"

Aufgrund des aktuellen Baugeschehens auf dem Campus der Brandenburgischen Technischen Universität wurde das dritte Umgebungsszenario entworfen. Der Proband befindet sich hier virtuell im Lehrgebäude 3A im Raum 321. In diesem Raum befindet sich auch die schalldichte Kabine, in der der Versuchsablauf durchgeführt wird. Dieses Szenario wurde aufgenommen, da sich jeder Student mit dieser Situation identifizieren kann, da er dieses Szenario jeden Tag erlebt. Der Raum hat eine Grundfläche von etwa 47 m<sup>2</sup> und eine Deckenhöhe von 2,75 m. Zur akustischen Dämpfung des Raumes sind vier Schallabsorber mit den Maßen 1 x 2 m an der Decke angebracht. Der Fußboden ist mit Linoleum ausgelegt. Die Fensterfläche des Testraumes beträgt etwa 11,6 m<sup>2</sup>. Der Baulärm konnte mit 41,8 dB erfasst werden; dies entspricht etwa einer leisen bis normalen Unterhaltung. Die Nachhallzeit des Raumes beträgt ungefähr eine Sekunde.

## 5. Erstellung der akustischen Umgebungen

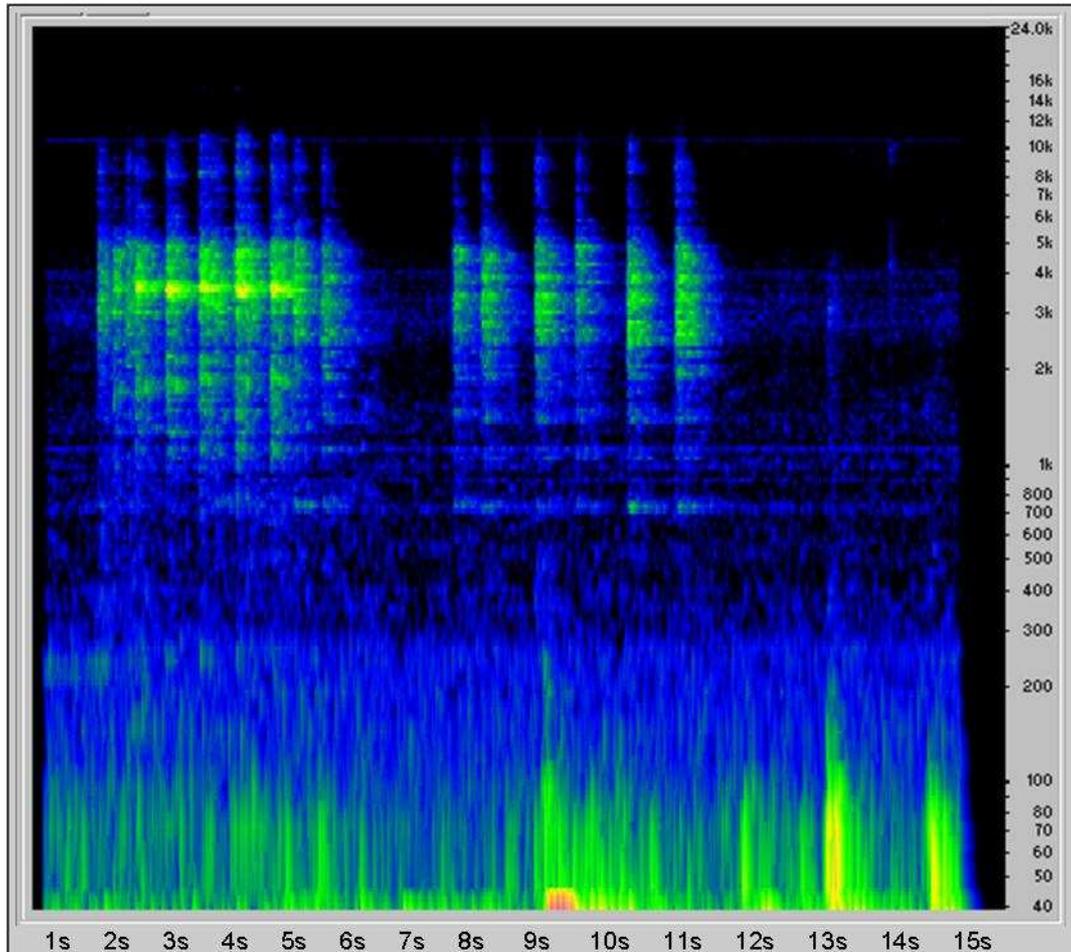


Abbildung 18: Spektrogramm Szenario 3

In Abbildung 18 ist das Spektrogramm von Szenario 3 zu sehen. Von 20 bis 300 Hz kann man eine Art Umgebungsrauschen wahrnehmen. Im Bereich von 1 bis etwa 7 kHz ist das Hämmern der Bauarbeiten sehr gut zu erkennen. Das Schlagen mit dem Hammer auf das Baumaterial kann im Bereich von 2 bis 6 Sekunden und von 8 bis 12 Sekunden gut wahrgenommen werden. Kennzeichnend wird es in den gelben Bereichen. Diese Bereiche könnten später im Test besonders störend wirken.

## 5. Erstellung der akustischen Umgebungen

### 5.2.4 Szenario 4 - „Raum 119 PC Pool mit mehreren Computerlüftern“



Abbildung 19: Szenario 4 - " Raum 119 PC Pool mit mehreren Computerlüftern "

Die vierte Umgebung zeigt einen PC-Pool an der Universität. Die Aufnahmen für dieses Szenario wurden im Lehrgebäude 3B im Raum 119 durchgeführt. Hierzu wurden mehrere Computer vor und hinter dem Kunstkopf in Betrieb genommen, um eine laufende Pool-Umgebung zu simulieren. Die Lüftergeräusche wurden mit einem Schallpegel von 34,8 dB erfasst. Dies ist relativ leise und nur wenig wahrnehmbar. Die Maße des Raumes betragen 10,2 m in der Tiefe und 8,4 m in der Breite. An der hinteren Wand sind mit Gipskarton Ecken mit den Maßen 0,8 x 0,8 m montiert. Somit ergibt sich eine Grundfläche von etwa 84,4 m<sup>2</sup>. Die Raumhöhe beträgt 2,9 m. Im Raum befinden sich zwei große Fensterflächen; eine mit 3,9 m<sup>2</sup> und eine mit 9,3 m<sup>2</sup>. Der komplette Fußboden ist mit einem Teppichbelag ausgelegt, wodurch eine gute Dämpfung des Nachhalls erreicht wird. Die Nachhallzeit beträgt etwa 0,65 Sekunden.

## 5. Erstellung der akustischen Umgebungen

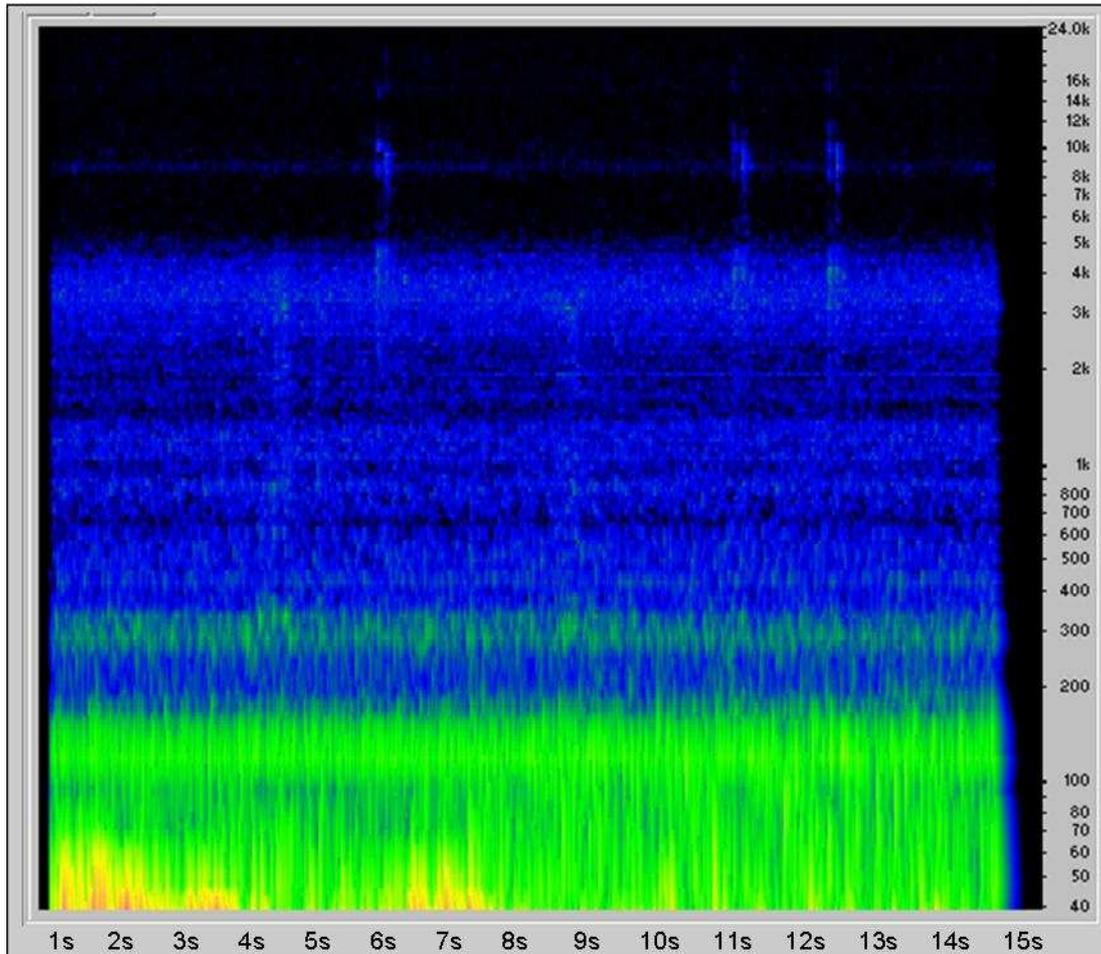


Abbildung 20: Spektrogramm Szenario 4

Abbildung 20 zeigt das Spektrogramm von Szenario 4. Die hier laufenden Lüfter der einzelnen Computer im PC Pool erzeugen ein tief- bis mittelfrequentes Rauschen im Bereich von 20 Hz bis 5 kHz. Besonders intensiv ist das Rauschen im Bereich von 20 bis etwa 180 Hz wahrzunehmen. Der Störpegel ist im kompletten Aufnahmebereich konstant.

### 5.2.5 Szenario 5 – „Modulated Noise Reference Unit (MNRU)“

Im Szenario 5 soll anhand der Modulated Noise Reference Unit (MNRU) der Störeinfluss über den kompletten Frequenzbereich von 20 Hz bis 20 kHz getestet werden. Durch die MNRU kann die Versuchsperson durch ein Rauschsignal auf den ganzen Qualitätsbereich stimuliert.

## 5. Erstellung der akustischen Umgebungen

Die Grundlagen für die Modulated Noise Reference Unit (MNRU) sind in der Norm ITU-T P.810 zu finden [ITU810]. Die ITU-T (Telecommunication Standardization Sector) ist ein Organ der internationalen Fernmeldeunion (ITU). Sie ist zuständig für die Untersuchung von technischen, betrieblichen und tariflichen Fragen und der Abgabe von Empfehlungen zur Standardisierung der weltweiten Telekommunikation.

Die Empfehlung P.810 beschreibt die MNRU. Sie ist ein Mittel zur kontrollierten Beeinflussung von Sprachsignalen. Sie wurde ursprünglich entwickelt, um subjektive Verzerrungen zu produzieren, die in ähnlicher Form bei logarithmisch-komprimierten PCM-Systemen auftreten. Ursprünglich wurde die MNRU unter Verwendung analoger Schaltungstechnik erzeugt. Heutzutage wird sie meist mit digitalen Schaltungen erzeugt oder auf dem Computer simuliert.

In Abbildung 21 ist eine vereinfachte Darstellung für die digitale Breitbandversion der MNRU zu sehen.

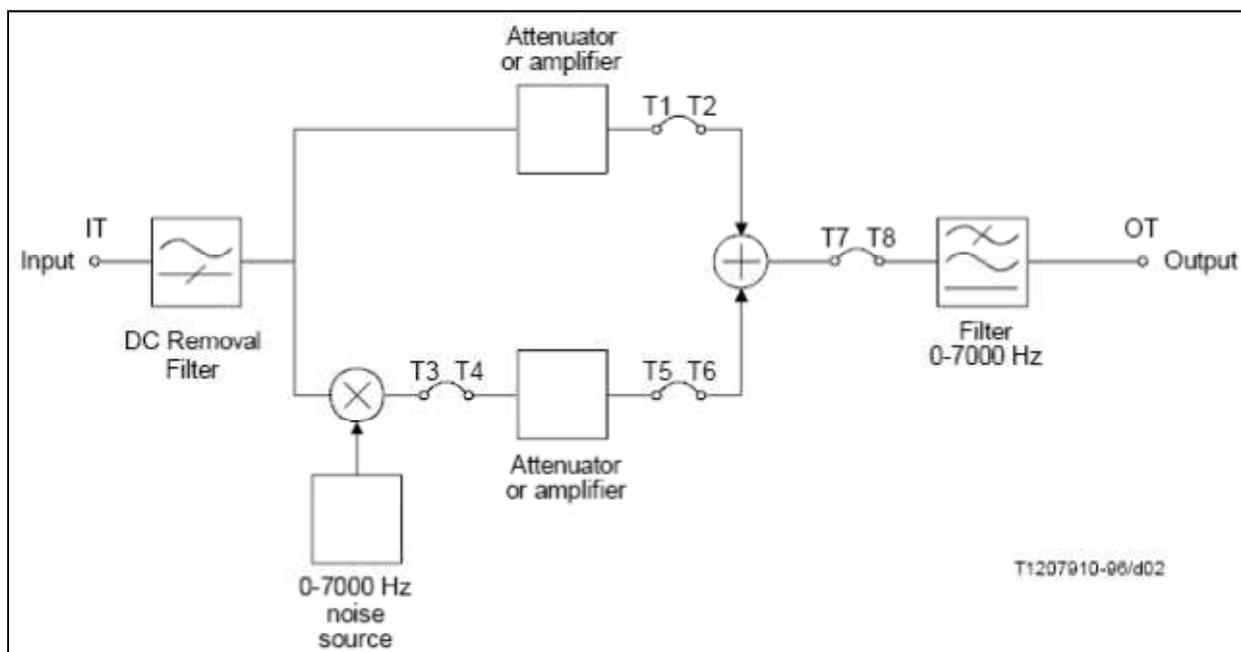


Abbildung 21: Vereinfachte Regelung der MNRU für digitale Breitbandversion [ITU810]

Das Eingangssignal wird auf zwei Wege aufgeteilt. Einerseits in einen unverfälschten Signalweg und andererseits in einen Geräuschweg, bei dem das Sprachsignal mit einem Gauß'schen Rauschen multipliziert wird. Der Ausgang des Multiplikators besteht aus dem Rauschen des modulierten Sprachsignals und wird dann zum Original-Sprachsignal addiert. Auf diese Weise wird das gestörte Signal erzeugt.

## 5. Erstellung der akustischen Umgebungen

Durch den Abschwächer und den Schalter im Signal- und Rauschweg wird eine unabhängige Einstellung des Sprach- und Rauschsignalanteils am Ausgang erreicht. Normalerweise ist das System so kalibriert, dass die Einstellungen des Dämpfers (in dB) im Rauschweg das Verhältnis der momentanen Sprachleistung zur Rauschleistung widerspiegeln, wenn beide am Ausgang des Bandpasses gemessen werden. Insbesondere wenn beide Dämpfer auf 0 dB gesetzt sind, beträgt der Geräuschpegel gemessen am Anschluss OT mit separatem ohmschen Abschluss am Anschluss T1 und T2 (nicht verbunden) den gleichen Pegel, wie der Sprachpegel gemessen am Anschluss OT mit separatem ohmschen Abschluss an den Anschlüssen T5 und T6 (nicht verbunden). Zur Überprüfung kann kontrolliert werden ob das Verhältnis von Sprach-plus-Rausch-modulierter Leistung am Ausgang im Vergleich zur Leistung der Eingangssprache am Eingang 3 dB beträgt.

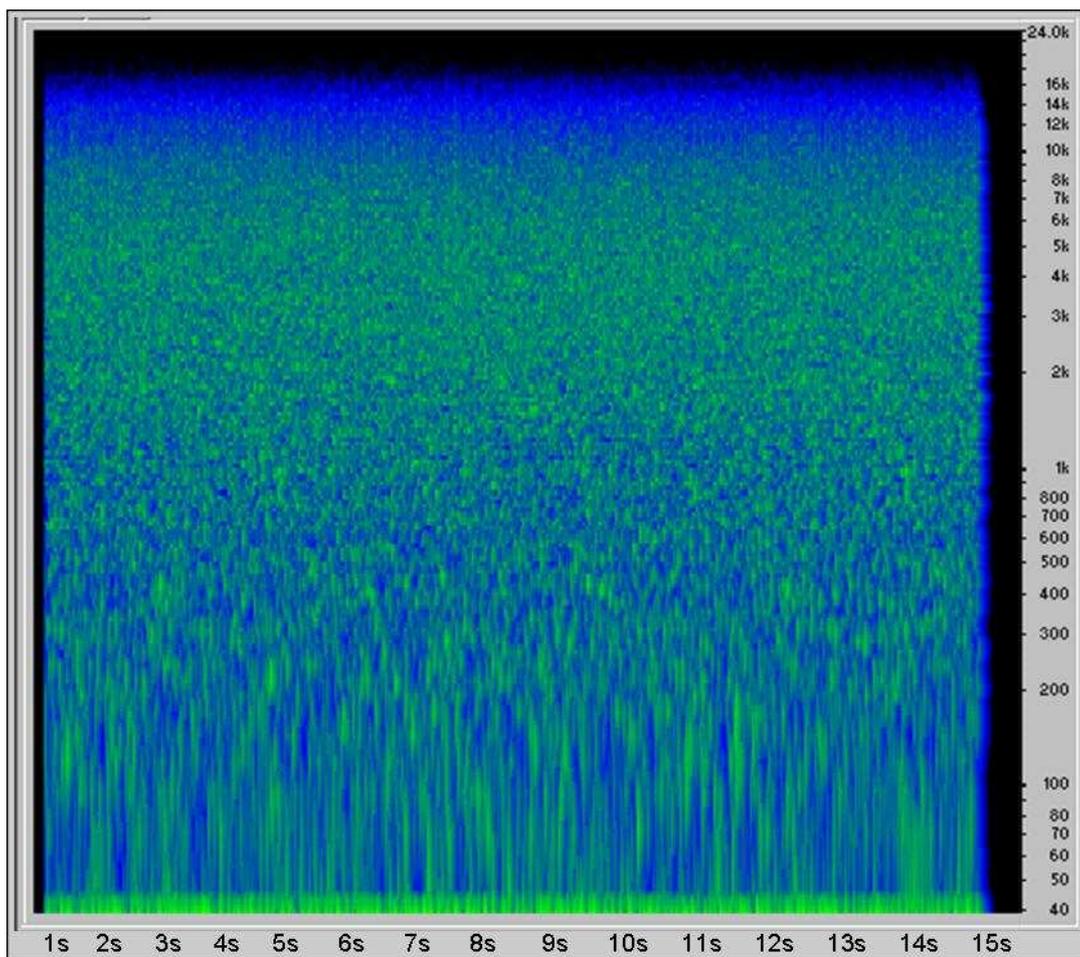


Abbildung 22: Spektrogramm Szenario 5

## 5. Erstellung der akustischen Umgebungen

Das Spektrogramm eines MNRU beziehungsweise eines weißen Rauschens ist in Abbildung 22 zu sehen. Das Signal erstreckt sich weitestgehend über den kompletten Frequenzbereich mit gleichbleibender Amplitude. Je nach Zumischung zum Nutzsignal wird die Sprachverständlichkeit leiden bzw. der Einfluss auf die Umgebung groß sein.

### 5.2.6 Szenario 6 - „Modulated Noise Reference Unit (MNRU) mit natürlicher Sprache“

Wie bereits im Szenario 5 wird hier ein weißes Rauschen zum Nutzsignal zugemischt. Bei diesem Beispiel besteht das Nutzsignal aus natürlicher Sprache eines anderen Sprechers. Hierdurch soll geprüft werden, ob gleiche Rückschlüsse, wie bei hochwertiger professioneller Sprache gezogen werden können. Die unterschiedliche Aufnahmequalität sollte bei der späteren Auswertung berücksichtigt werden, da die Sprachaufnahmen des Profisprechers eine wesentlich höhere Aufnahmequalität aufweisen.

### 5.2.7 Szenario 7 - „Raum 321 Original“

Szenario 7 findet in derselben Umgebung wie bereits Szenario 3 statt. Der Proband sitzt am originalen Aufnahmeplatz. Hierdurch sollen Rückschlüsse zwischen originaler und virtueller Umgebung gezogen werden. Besonderes Augenmerk muss bei der Auswertung auf die wechselnden Umgebungsbedingungen gelegt werden. Der Test erfolgt, wie die Aufnahme, mit angeklapptem Fenster. Jedoch ist der Baulärm zu jeder Versuchszeit unterschiedlich. Um später eine genauere Aussage über die Störgeräusche vornehmen zu können, wird bei jedem Test die Lautstärke mit dem bereits in Punkt 5.1.4 vorgestellten Schallpegelmessgerät der Firma Brüel & Kjaer dokumentiert. Der Störpegel des Szenarios wird eine Minute lang aufgenommen. Somit kann ein Durchschnittswert während der Versuchszeit ermittelt werden. Bei der späteren Auswertung kann mit Hilfe der Pegel auf einzelne Gruppen näher eingegangen und es können einzelne Effekte besser erläutert werden. Weiterhin stellt dieser Testplatz einen Originalplatz für internet-basierte Untersuchungen dar, da die Probanden mit relativ einfacher Technik den Versuch

## 5. Erstellung der akustischen Umgebungen

durchführen, wie es in der heimischen Umgebung wäre. Als Kopfhörer kommt hier der PC 131 der Firma Sennheiser zur Anwendung. Dies ist ein einfacher Stereokopfhörer mit Headsetfunktion. Sein Übertragungsbereich beträgt 30 Hz bis 18 kHz. Der Kopfhörer ist dynamisch sowie offen und lässt somit sehr gut Störgeräusche an die Ohren des Probanden herandrängen. Somit wird eine nahezu identische akustische Umgebung, wie im Szenario 3 geschaffen, welches in der schalldichten Kabine abgespielt wird. Die Bedienung des Versuchsablaufes ist identisch mit dem Ablauf in der Kabine. Die Testperson sitzt vor einem Monitor mit dem entsprechenden Testablauf und hat eine Maus zur Steuerung des Versuches. Probanden mit sehr empfindlichem Gehör werden den Netzwerkschitch auf dem Fensterbrett als störend empfinden, da er einen sehr hochfrequenten Pfeifton von sich gibt. Somit kann für diese Testpersonen die Umgebung auch dann sehr störend wirken, wenn kein Baulärm vorhanden ist. Mit Hilfe der Verdunkelung und der Raumbelichtung wurde für jeden Testteilnehmer die gleiche visuelle Versuchsumgebung geschaffen. Der Grundstörerschallpegel des Raumes beträgt bei geschlossenen Fenstern und laufenden Computern bereits 39 dB.

# 6. Testumgebung und Randbedingungen

## 6.1 Hörtestumgebung

In erster Linie ist bei der Hörtestumgebung darauf zu achten, dass die Umgebungsbedingungen während des gesamten Tests unverändert bleiben. Hierzu zählen die Temperatur, die Akustik, die Beleuchtung und die Einrichtung. Die Lautstärke der Sprachproben sollte ebenfalls konstant gehalten werden. Während des gesamten Testablaufes dürfen die Versuchsanordnung des Testes und der Ablauf nicht verändert werden. Somit wird der Versuch mit anderen, unter gleichen Bedingungen durchgeführten Tests vergleichbar und eine Repräsentativität erreicht. Bereits bei kleinen Abweichungen der genannten Randbedingungen können Unterschiede im Abstimmungsverhalten der Testpersonen auftreten. Desweiteren wird in der ITU-Norm festgelegt, dass der Test in einer schalldichten Kabine mit einem Volumen von nicht mehr als 30 m<sup>3</sup> durchgeführt werden soll. Störgeräusche müssen so gering wie möglich gehalten werden.

Weiterhin gibt es einige Faktoren, die den gesamten Versuchsablauf über konstant gehalten werden müssen, denn vergleichbare Ergebnisse können nur erreicht werden, wenn unter gleichen Bedingungen getestet wird. Beispielsweise genügt schon eine veränderte Lichtsituation, um die Bewertung zu beeinflussen. Bei einer zu grellen Beleuchtung könnten Probanden beispielsweise gereizt reagieren oder abgelenkt werden. Desweiteren sollte der Hörtest mit Kopfhörern anstelle von Lautsprechern durchgeführt werden, da so das Schallereignis direkt an die Ohren gelangt und störende Reflektionen oder Beugungen des Schalls an Gegenständen im Raum keine Auswirkungen haben. Schlussfolgernd aus diesen Betrachtungen müssen für die Durchführung des Versuchs eine schalldichte Kabine und Kopfhörer benutzt werden. Um keine Veränderungen am Sprachmaterial hervorzurufen, muss die hardwaremäßige Ausstattung ebenfalls unveränderlich bleiben. Bereits eine Änderung der Soundkarte kann die Hörproben negativ beeinflussen, da jede Hardwarekomponente andere signalcharakteristische Eigenschaften besitzt.

Ein weiterer wichtiger und nicht zu vernachlässigender Aspekt ist der Sprecher des Testmaterials. Er muss versuchen, das gesamte Material mit dem gleichen Stimmcharakter wiederzugeben. Somit müssen beispielsweise die Intonation, die Ge-

## 6. Testumgebung und Randbedingungen

schwindigkeit bzw. der Rhythmus des Gesprochenen zwingend konstant gehalten werden [Cas06]. Weitere Randbedingungen folgen in Kapitel 6.2.

### 6.2 Randbedingungen des Tests

Die folgenden vier Absätze sollen Randbedingungen des Tests aufzeigen. In erster Linie zählen hierzu die Versuchsteilnehmer, die verschiedene Voraussetzungen haben sollten. Im Anschluss wird etwas zur technischen und räumlichen Ausstattung, sowie zu auftretenden Effekten gesagt.

#### 6.2.1 Versuchsteilnehmer

Um einzelne Tendenzen der Versuchspersonen ausgleichen zu können, wird eine große Anzahl an Probanden benötigt. Die Testpersonen sollten normalhörend sein und in ihrem Alter, Geschlecht und Beruf einem repräsentativen Durchschnitt entsprechen. Sie sollten keine Spezialisten für die Sprachqualitätsbewertung oder Sprachsynthese sein. Weiterhin sollten die Probanden in den letzten sechs Monaten nicht an einem subjektiven Test und mindestens seit einem Jahr nicht an einem Einschätzungstest teilgenommen haben. Ebenfalls dürfen sie die im Test verwendeten Sätze noch nie gehört haben [ITU800]. Es ist darauf zu achten, dass Stress für die Probanden vermieden wird, da dies einen großen Einfluss auf die Bewertung haben kann. Weitere Einflüsse auf die Ergebnisse können nachlassende Aufmerksamkeit oder Müdigkeit sein. Somit sollte der Test so gestaltet werden, dass die Versuchsteilnehmer nicht überfordert werden und die Dauer auf ein erträgliches Maß gehalten wird. Weiterhin können eventuelle Hörprobleme der Probanden anhand von audiometrischen Untersuchungen im Vorfeld des Tests erkannt und gegebenenfalls entgegengewirkt werden bzw. diese vom Test ausgeschlossen werden. Eine Möglichkeit ist die Hörschwellenermittlung. Besonders wichtig ist es darauf zu achten, dass die Versuchsteilnehmer mit der entsprechenden Motivation am Test teilnehmen, denn nur so ist ein realistisches Ergebnis zu erwarten. Zur Motivation kann eine einführende Erläuterung über Sinn und Zweck des Versuches beitragen.

## 6. Testumgebung und Randbedingungen

### 6.2.2 Räumliche und technische Ausstattung

Die Versuche fanden in einer schalldichten Kabine statt. Die Maße dieser Kabine betragen 2,14 m \* 2,22 m \* 2,13 m (B\*T\*H). Dies ergibt ein Volumen von 10,1 m<sup>3</sup>, welches entsprechend der ITU800 unter dem Maximalwert von 30 m<sup>3</sup> liegt. Am Testplatz steht als Steuerungscomputer ein Pentium(R) 4 CPU mit 1,90 GHz zur Verfügung. Dieser hat zur D/A-Umsetzung die Soundkarte DIGI96/8 PAD der Firma RME mit einer Auflösung von 24 bit integriert. Der Signal-Störgeräusch-Abstand der Karte beträgt 112 dBA. In Abbildung 23 ist der Testplatz zu sehen, von dem aus der Versuchsablauf gesteuert wird. Durch den Zweibildschirmmodus der Grafikkarte wird dem Probanden in der Kabine ausschließlich die PowerPoint-Präsentation gezeigt, und der Versuchsleiter hat einen kompletten Überblick über den Versuchsablauf und dessen Steuerung.



Abbildung 23: Testplatz für den Versuch

Abbildung 24 zeigt den Platz des Probanden in der schalldichten Kabine. Hier steht der Testperson der geschlossene, dynamische Kopfhörer HDA 200 der Firma Sennheiser zur Verfügung. Der Hörer ist speziell für den Einsatz an Hochton-Audiometern entwickelt worden und besitzt einen Frequenzgang von 20 Hz bis 20 kHz mit einer Nennbelastbarkeit von 500 mW. Durch seine geschlossene Bauweise weist er eine sehr hohe Außengeräuschdämmung auf. Zur optischen Darstellung des

## 6. Testumgebung und Randbedingungen

Versuches befinden sich ein Monitor und eine Maus zur Kontrolle des Ablaufes in der Kabine.



Abbildung 24: Platz des Probanden in der schalldichten Kabine

### 6.2.3 Versuchsablauf

Im Vorfeld ist eine genaue Planung des Versuchsablaufs nötig, da sich die gesamte Struktur auf die Beurteilung auswirken kann. Somit kann sich die Länge des Tests negativ auf die Aufmerksamkeit der Testpersonen auswirken. Die Ergebnisse könnten durch Verlust der nötigen Konzentration beeinflusst werden. Laut Casper [Cas06] wurde bei der Befragung im Anschluss seines Internettests deutlich, dass für die Probanden eine Testdauer von 25-30 Minuten erträglich ist. Somit wurde alles, was in diesem Zeitrahmen lag, als angenehm und akzeptabel bewertet. Weiterhin sollte man den Probanden an keiner Stelle im Test ein Zeitlimit setzen, da die Versuchspersonen nicht unter Druck geraten dürfen, sowie um Stresssituationen und damit eine eventuelle Verfälschung des Tests auszuschließen [Cas06].

### 6.2.4 Störeffekte

Bei der Bewertung der Probanden werden verschiedene Effekte auftreten. In Erscheinung treten voraussichtlich der Lombard-Effekt, der Cocktailparty-Effekt und

## 6. Testumgebung und Randbedingungen

die Verdeckung. Auf diese Effekte sollte bei der Auswertung später genaues Augenmerk gelegt werden und eine ausführliche Diskussion erfolgen.

Vom „Cocktailparty-Effekt“ spricht man, wenn bei mehreren unterschiedlichen Signalen das eigentlich wichtige Signal (Nutzsignal, hier die Sprache) aus mehreren Schallquellen herausgefiltert wird. Somit wird nur der Sprecher wahrgenommen und das Störsignal unterdrückt. Diesen Effekt bezeichnet man auch als selektives oder intelligentes Hören. Der Cocktailparty-Effekt beruht auf der Fähigkeit des Richtungshörens. Der Mensch ist in der Lage, das Geräusch, welches er hören möchte, um einige dB anzuheben, so dass es aus dem Störgeräusch herausgefiltert werden kann und lauter wahrgenommen wird [Deg11].

Vom „Lombard-Effekt“ spricht man, wenn die Kommunikation bei verschiedenen Umweltsituationen entsprechend den Umgebungsbedingungen angepasst wird. Dies tritt zum Beispiel in Stresssituationen durch Änderung der Wortdauer, der Aussprachefrequenz und zahlreicher weiterer Parameter ein. Besonders wird dies in Notsituationen deutlich, da der Mensch hier instinktiv zu lauterem Sprechen tendiert um gehört zu werden. Stress kann aber bereits bei erhöhtem Umgebungsgeräusch oder Lärm auftreten. Laut Lombard nimmt der Sprechpegel um jeweils ein halbes dB zu, wenn sich das Störgeräusch um ein dB erhöht. Dies hat zur Folge, dass sich Gesprächspartner ihrer Umweltsituation anpassen. Diese Anpassung hat Einfluss auf den Stimmcharakter, denn eine laute, erhobene Stimme wirkt beispielsweise weniger sympathisch und eher gestresst. Hierdurch kann die Motivation sinken und der Zuhörer kann in Stress verfallen, wodurch eine Steigerung der Fehlerrate möglich ist [AEU11].

Von „Verdeckung“ spricht man, wenn dem Gehör zwei Schallereignisse gleichzeitig dargeboten werden und das leisere durch das lautere Ereignis durch gewisse Bedingungen verdeckt oder maskiert wird. Ein wichtiger Faktor ist hierbei die Mithörschwelle. Sie bezeichnet denjenigen Schalldruck des verdeckten Schalls, bei dem dieser gerade nichtmehr hörbar ist. Die Mithörschwelle ist abhängig vom Schalldruckpegel des verdeckenden Signals und den Frequenzspektren des verdeckten und des verdeckenden Schalls. Hierbei werden typischerweise bei kleiner Bandbreite des verdeckenden Schalls tieffrequente Signale weniger verdeckt als hochfrequente. Die Mithörschwelle sinkt bei wachsendem Frequenzabstand vom verdeckten Frequenzbereich auf die absolute Hörschwelle ab [Ler09].

### 7. Ablauf und Durchführung des Tests

Zu Beginn werden die Probanden in den Test eingewiesen. Damit jeder Versuchsteilnehmer die gleichen Voraussetzungen besitzt, erfolgt die Einweisung in schriftlicher Form anhand eines Informationsblattes. Im Anschluss können die Testpersonen Fragen an den Versuchsleiter stellen.

Anschließend wird anhand einer vereinfachten Gehörschwellenermittlung die optimale Testlautstärke eingestellt, sodass jeder Proband den Versuch gleich laut empfindet. Die Hörschwelle ist eine sogenannte absolute Schwelle. Unterhalb dieser Grenze werden Schalle nicht gehört. Durch Motivation, Lärm oder Alter kann sich die Hörschwelle verändern [Deg11]. Die Anpassung an die Hörschwelle hat auf die Lautstärkeempfindung der Szenarien keine Auswirkung, da eine schlechter bzw. besser hörende Person die Umgebungsgeräusche ebenfalls schlechter bzw. besser wahrnehmen würde. Zur Anpassung an das Gehör wird der Testperson ein Sinus-Ton mit einer Frequenz von 1 kHz und einer Lautstärke von -60 dB dargeboten. Dieser Wert spiegelt die Hörschwelle wider. In Abbildung 25 ist die Lautstärkeregelung der Soundkarte zu sehen.

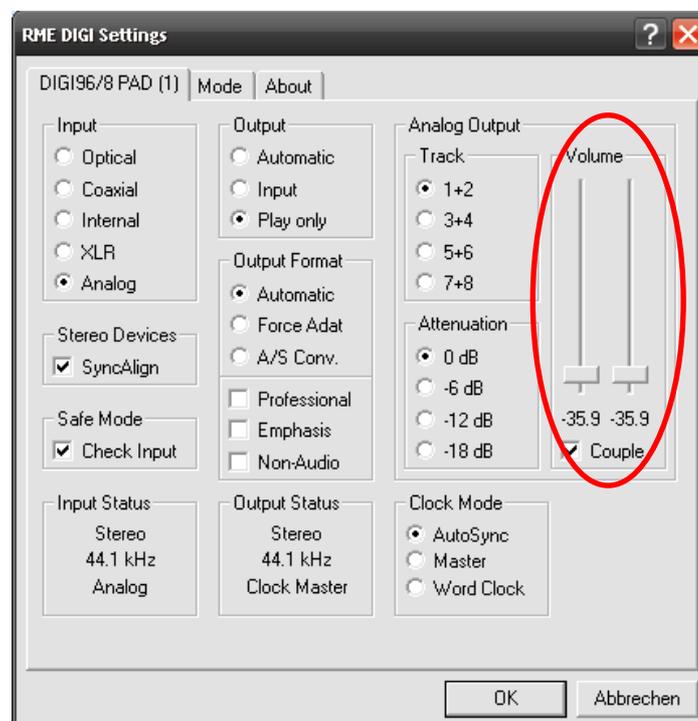


Abbildung 25: Lautstärkeregelung der Soundkarte

## 7. Ablauf und Durchführung des Tests

Mit Hilfe der markierten Regler kann der Proband nun die Lautstärke soweit verringern, bis er den Ton gerade noch wahrnimmt. Um dies zu gewährleisten, soll der Testton erst soweit herunter geregelt werden, bis er nicht mehr gehört wird und anschließend soll die Lautstärke wieder erhöht werden, bis der Sinus-Ton das erste Mal wieder wahrgenommen wird.

Auf eine Testphase wird verzichtet, da die Probanden rein intuitiv bewerten sollen. Somit kann jeder Teilnehmer individuell entscheiden, was ihn bezüglich der Sprachqualität stört oder er noch für ertragbar empfindet.

Im Anschluss beginnt der eigentliche Test. Den Versuchspersonen werden nun nacheinander Hörproben, eingebettet in einer PowerPoint Präsentation, dargeboten. Die Proben sind nach bestimmten Kriterien zu beurteilen. Um Beurteilungstendenzen auf alle Sprachproben gleichmäßig zu verteilen, werden die Proben in zufälliger Reihenfolge wiedergegeben. Somit werden diese Tendenzen, die durch Gewöhnung oder durch Lerneffekte entstehen können, vermieden.

Die Sprachproben besitzen eine Länge von etwa 10 bis 13 Sekunden. Dies entspricht im Schnitt etwa zwei gesprochenen Sätzen. Zwischen den Testsätzen ist jeweils eine gut wahrnehmbare Pause vorhanden. Die Testsätze bestehen aus alltäglichen Wörtern und enthalten keine unverständlichen Fremdwörter. Dies würde die Verständlichkeit zusätzlich erschweren. Die jeweiligen Umgebungssignale wurden ins Verhältnis zum Sprachsignal gesetzt und mit entsprechendem Pegel abgesenkt.

Um den Störeinfluss der einzelnen Szenarien zu testen, wurden fünf unterschiedliche Störschallpegel je Szenario erzeugt. Die Pegelvariation betrug jeweils 6 dB. Das entspricht einer Verdopplung des Schalldruckes. Die Pegeländerungen für die Szenarien betragen -6 dB, 0 dB, 6 dB, 12 dB und 18 dB. Hierbei entspricht -6dB dem halben Schalldruckes, 0 dB dem Original, 6 dB dem doppelten Schalldruck, 12 dB dem vierfachen und 18 dB dem achtfachen Schalldruckpegel. Eine noch extremere Lautstärkeerhöhung wäre nicht sinnvoll gewesen, da derartige Geräuschpegel in der natürlichen Hörumgebung nicht vorkommen. Es zeigte sich, dass von den Probanden eine Vervierfachung bereits als unangenehm empfunden wurde.

Es waren insgesamt sieben Szenarien - wie in den Punkten 5.2.1 bis 5.2.7 bereits vorgestellt - mit Sprachproben anhand der jeweils aufgezeigten 5-stufigen Skala zu

## 7. Ablauf und Durchführung des Tests

bewerten. Die ersten sechs Szenarien werden in der schalldichten Kabine und das siebte Szenario außerhalb der Kabine von den Probanden bewertet. Für jede Umgebung waren fünf Hörbeispiele mit den entsprechenden Lautstärkeänderungen zu beurteilen. Diese wurden in zwei Durchgängen jeweils einmal dargeboten. Im ersten Durchgang war die Sprachverständlichkeit zu bewerten und im zweiten der Einfluss der Umgebung auf die Sprachprobe. Die Reihenfolge und die Zuordnung der Sprachbeispiele waren in jedem Szenario und Durchgang unterschiedlich und zufällig gewählt. In Abbildung 26 und 27 sind die Bewertungsskalen zu sehen, anhand derer die Probanden ihr Urteil abzugeben hatten. Die Bewertungsskala für die Sprachverständlichkeit, welche durch „sehr gut“, „gut“, „mittelmäßig“, „schlecht“ und „nicht verständlich“ abgestuft ist, kann man in Abbildung 26 sehen.

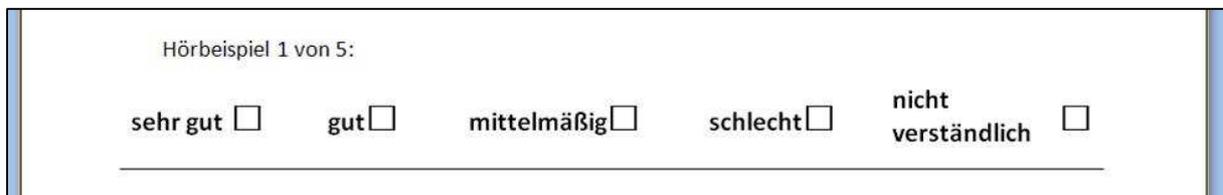


Abbildung 26: Skala für die Sprachverständlichkeit

Abbildung 27 zeigt die fünfstufige Skala für den Einfluss der Umgebung, welche mit „nicht hörbar“, „nicht störend“, „leicht störend“, „störend“ und „extrem störend“ gewichtet ist.



Abbildung 27: Skala für den Einfluss der Umgebung

Auf Grund des internationalen Standards wird durch die Bewertung „5“ das beste und durch „1“ das schlechteste Ergebnis angegeben. Hierdurch wird der Test eindeutig für weitere Untersuchungen angepasst. Die in Deutschland typische umgedrehte Skala würde hier nur irritieren. Eine weitere positive Eigenschaft der internationalen Norm ist in der späteren Auswertung ersichtlich. Denn hierdurch stehen die besseren Ergebnisse in den Grafiken auch an höherer Position, als die Schlechteren.

## 7. Ablauf und Durchführung des Tests

Im Anschluss an jedes Hörbeispiel wird der Proband aufgefordert, seine Bewertung abzugeben und die Beurteilung mit dem Betätigen der „Weiter“-Taste zu bestätigen. Nach dem Drücken der Taste wird das nächste Beispiel dargeboten.

Die Visualisierung des Bildschirms gibt dem Probanden stets Auskunft über den aktuellen Stand des Versuches. Auf dem Bildschirm ist der Name des zu bewertenden Szenarios oben zu sehen. Darunter wird auf der linken Seite der Durchgang und daneben die jeweilige Hörprobe angezeigt. Mittig auf dem Monitor ist ein Foto des Aufnahmeortes mit dem Kunstkopf zu sehen. Somit kann sich der Proband sich ein besseres Bild von der zu bewertenden Umgebung machen. Unter dem Foto wird der Gesamtfortschritt kenntlich gemacht. In Abbildung 28 ist die Bildschirmoberfläche zu sehen.



Abbildung 28: Bildschirmoberfläche des Versuches

## 8. Auswertung der Ergebnisse

### 8.1 Allgemeine Ergebnisse

Der Versuch lief vom 04. Juli 2011 bis zum 03. August 2011. In dieser Zeit nahmen insgesamt 40 Personen an dem Versuch teil. Abbildung 29 zeigt die Altersverteilung der Testteilnehmer.

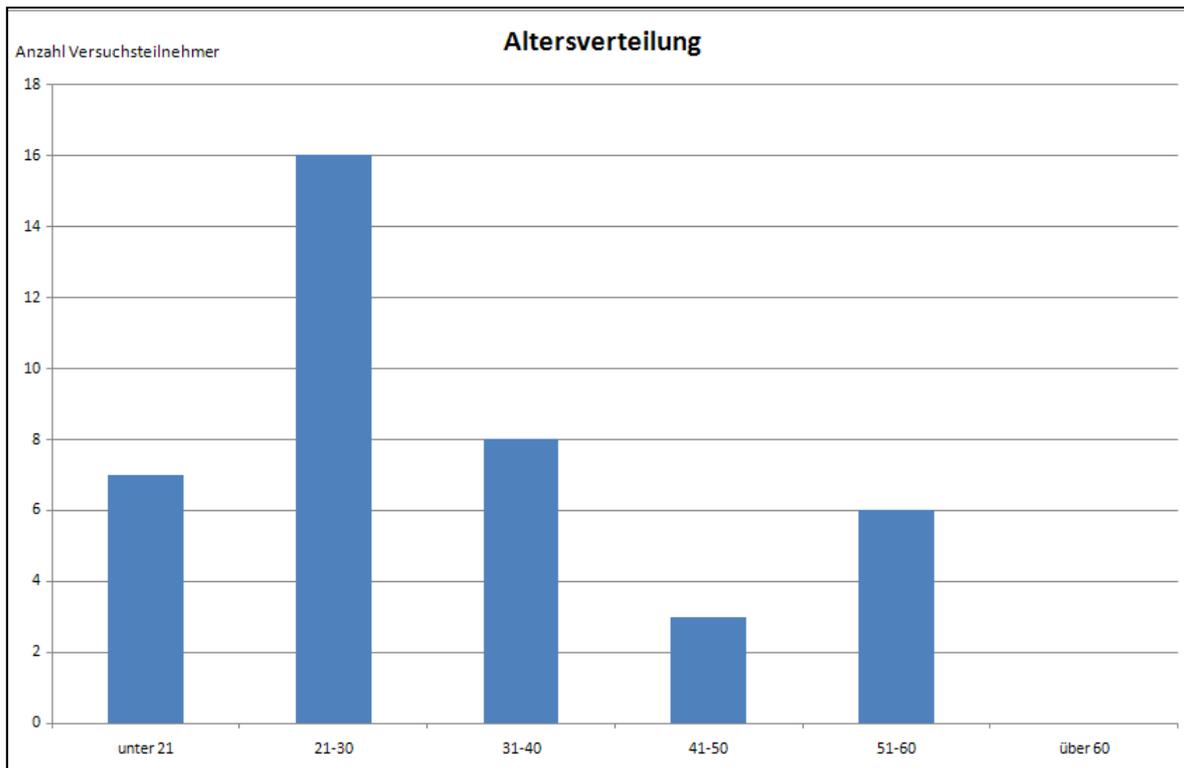


Abbildung 29: Altersverteilung

In der Grafik ist ersichtlich, dass 40% der Probanden im Alter zwischen 21 und 30 Jahren waren. Dies ist auf die vorrangige Akquise von Studenten und Schülern zurückzuführen. Die vier anderen Altersgruppen waren weitestgehend gleich stark. Allein die Gruppe „41-50“ war durch 3 Teilnehmer nur sehr gering vertreten.

Die 40 Probanden teilten sich in etwa ein Drittel weibliche und zwei Drittel männliche Teilnehmer auf. Diese Akquise scheint bei vielen Studien üblich zu sein. Die Frage, warum mehr Männer sich für derartige Tests bereit erklären, kann hier jedoch nicht geklärt werden.

Im Schnitt benötigten die Probanden ca. 25 Minuten für die Versuchsdurchführung. Dies ist dem geregelten Ablauf geschuldet. Hierdurch kann nur bei der Bewertung in

## 8. Auswertung der Ergebnisse

den zeitlichen Ablauf durch Klicken der „Weiter“-Taste auf dem Bildschirm eingegriffen werden.

Nach Beurteilung der Probanden war die Entscheidung für das Ergebnis bereits nach den ersten Sekunden des Hörbeispiels gut zu treffen, sodass die Bewertung ohne große zeitliche Einbuße erfolgte. Die Testdauer wurde von den meisten Probanden als sehr angenehm empfunden.

Im Anschluss an den Test hatten die Probanden die Möglichkeit, ein Feedback zu dem Versuch zu geben, und in einem kurzen Gespräch konnten etwaige Schwierigkeiten oder Fragen geklärt werden. Folgende Kommentare sind bei den Feedbacks abgegeben worden (entnommen aus den Bewertungsbögen):

- Bemerkungen zur Sprachgütemessung allgemein:
  - Ungewohnt, da sich trotz stärkerem Störpegel die Stimme nicht ändert.
  - Weißes Rauschen ist unangenehmer als die anderen Geräusche.
  - Sehr interessant, Gegensätze könnten noch extremer sein.
  - Lärm ist kein Problem durch Gewöhnung des Alltags.
  - Gleiche Sprachprobe → Ergebnis bei Verständlichkeit verfälscht.
  - Unterschiedliche Sprecherqualitäten (Aufnahme).
  
- Technische Bemerkungen:
  - Kalibrierung der Hörschwelle zu ungenau.
  - Störpegel stark abhängig von eingestellter Hörschwelle (Fehler?!).
  - Kopfhörer drücken zu doll, unangenehm für Brillenträger.
  - Lüfter in 321 stören sehr stark (mehrere Computer).
  - Starkes Flackern des Bildschirms bei Szenario 7.
  - Switch hat ein sehr störendes hochfrequentes Fiepen.

## 8. Auswertung der Ergebnisse

- Bemerkungen zum Testverlauf:
  - Trainingsphase wäre gut gewesen, um besser beurteilen zu können, was genau „gut“ und „schlecht“ ist.
  - Gut bewertbar.
  - Bessere Trennung zwischen dem ersten und zweiten Durchgang → was wird gerade bewertet.
  - „Weiter“-Taste gut, da individuelle Übergangszeit.
  - „Weiter“ lieber per Tastatur.
  - Pausenzeiten zu lang → Mausproblem (Maus verschwindet immer wieder von der Bildschirmoberfläche).
  - Maus weg.
  - Positiv, dass Störsituation bildlich dargestellt wurde.
  - Schnellere Bewertung ermöglichen (Maus).
  - Entscheidung kann oft schon früher getroffen werden.
  - Bildbeschreibung → Kunstkopf am jeweiligen Aufnahmeort → war nicht klar.
  - Sehr realistisch klingende Szenarien.

Anhand dieser Feedbacks lassen sich Rückschlüsse für spätere Sprachgütemessungen ziehen. Viele Probanden bemerkten, dass die akustischen Umgebungen sehr real wirkten, so dass sie teilweise sogar kurz die Hörmuschel vom Kopfhörer von den Ohren nahmen um zu überprüfen, ob die Geräusche aus dem Raum oder dem Hörer kommen. Diese Aussagen zeigen, dass die erzeugten Umgebungen sehr realitätsnah waren und der Proband sich ohne große Probleme in das Szenario versetzen konnte. Einige der Probanden empfanden das weiße Rauschen als sehr unangenehm. Es hatte wesentlich mehr Einfluss auf die Sprachproben als alle anderen Szenarien. Im Gespräch wurde den Probanden erläutert, wie dieses „Phänomen“ zustande kam. Das weiße Rauschen erstreckt sich über den kompletten

## 8. Auswertung der Ergebnisse

Frequenzbereich mit gleicher Amplitude und verdeckt somit mehr Informationen als die nur punktuell in gewissen Frequenzbereichen auftretenden Störungen der anderen Szenarien. Da einige Testteilnehmer täglich beruflich mit Baulärm oder in der heimischen Umgebung mit Straßenlärm leben müssen, gaben sie an, dass sie sich bereits an die Störungen gewöhnt hätten und diese sehr gut ausblenden könnten.

Für spätere Versuche in dieser Testumgebung sollten die technischen Bemerkungen berücksichtigt werden. Die Kalibrierung der Hörschwelle ist zu ungenau und sollte eventuell für jedes Ohr extra erfolgen. Die Einstellsoftware der Soundkarte DIGI96/8 PAD der Firma RME, mit der die Hörschwellenermittlung durchgeführt wurde, ist zu ungenau für dieses Verfahren, da sie nur sprunghaft die Werte ändert. Dies wurde später auch an den Werten der Hörschwellen sichtbar. Somit hatten viele Probanden einen gleichen Wert, obwohl er differenzierter hätte sein können. Weiterhin wird der Störpegel mit zu hoher Hörschwelle auch unangenehm. Im Umkehrschluss wird bei einer zu gering eingestellten Hörschwelle die Sprache zu leise. Da diese Hörschwellenermittlung jedoch rein subjektiv ist und nur durch den Probanden durchgeführt werden konnte, können hier keine relevanten Fehler aufgezeigt werden.

Brillenträger gaben an, dass der Kopfhörer zu stark an den Kopf drückt und dies nach einiger Zeit sehr unangenehm wird. Dieses Problem kann jedoch nicht geändert werden, da dieser geschlossene Kopfhörer für den Versuch unverzichtbar ist. In Ergänzung zur schalldichten Kabine ist so gesichert, dass keine Nebengeräusche den Probanden beeinflussen oder ablenken; die Testpersonen bekamen nur das zu hören, was sie hören sollten. Da sich die meisten Personen beim Szenario 7 auf den Baulärm konzentriert haben und sie nur mit diesem gerechnet hatten, empfanden sie die Computerlüfter teilweise stärker störend als den Baustellenlärm. Die Mehrheit der Probanden führte den Versuch ohne störende Bauarbeiten durch. Bei weiteren Experimenten im Raum 321 sollte auf geeignete Testtechnik geachtet werden. Der Monitor am Testplatz für das Szenario 7 wies ein sehr starkes Flackern auf. Die Probanden, denen dies auffiel, bemerkten, dass ein durchgängiger Test an diesem Platz sehr stark die Konzentration beeinflusst und sehr unangenehm ist.

Weiterhin gab es kurze Bemerkungen zum Testverlauf, auf die hier eingegangen werden soll. Viele Probanden haben ohne Probleme intuitiv ihre Bewertung abgegeben. Jedoch gab es auch einige Testpersonen, die gerne im Vorfeld eine

## 8. Auswertung der Ergebnisse

Trainingsphase gehabt hätten. In dieser Phase hätte eine bessere Zuordnung für die Probanden ersichtlich werden können, in dem eine nicht störende und eine stark störende Umgebung zum Sprachsignal zugemischt worden wäre. Durch die Testbewertung in Papierform hatten die Testteilnehmer jedoch die Möglichkeit ihre Angaben während des Versuches zu korrigieren und die Bewertung anzupassen. Von dieser Möglichkeit machten viele der Probanden Gebrauch.

Ein Teilnehmer empfand die Trennung zwischen dem ersten und dem zweiten Durchgang als ungeeignet, da er nicht genau wusste, wann er was beurteilen musste. Jedoch bemerkte der Proband, dass dies ab dem zweiten Szenario keine großen Umstände mehr machte.

Das größte Problem hatten die Probanden mit der Weiterschaltung im Anschluss an ihre Bewertung. Auf Grund der PowerPoint Präsentation verschwand der Mauszeiger bei jeder neuen Folie. Somit mussten die Probanden vor ihrer Bewertung immer erst die Maus bewegen, bis der Zeiger wieder auftauchte. Dies empfanden viele als störend und zeitaufwendig. Eine Alternative wäre die Weiterschaltung per Tastatur, jedoch hätte die Testperson dann auch die Möglichkeit, sich frei durch den Versuch zu bewegen oder die Hörbeispiele mehrmals anzuhören – und dies sollte vermieden werden. Dennoch wurde die Mausvariante positiv beurteilt, da somit eine individuelle Beurteilungszeit bestand und keine Stresssituation entstehen konnte. Als großen Vorteil empfanden die meisten Probanden die bildliche Veranschaulichung der Aufnahmeumgebung, da sie sich so einen besseren Eindruck von dem Raum machen konnten, in dem sie sich virtuell befanden. Jedoch wäre eine detailgenauere Beschreibung des Bildes hilfreich gewesen, da sich herausstellte, dass eine große Anzahl der Teilnehmer nicht wussten, warum der Kunstkopf auf dem Bild zu sehen war. Viele Probanden bemerkten im Gespräch, dass eine frühe Urteilsabgabe wünschenswert wäre, da der generelle Eindruck bereits nach den ersten Sekunden des Hörbeispiels getroffen werden konnte. Sehr positiv ist zu bemerken, dass die Probanden die erzeugten akustischen Umgebungen als sehr realistisch empfanden. Hier gab es nie den Eindruck, dass die Szenarien unnormale Klänge.

Ein Proband bemerkte, dass eine Dopplung der Sprachproben beim professionellen Sprecher und dem natürlichen Sprecher in Szenario 6 auftrat. Diese Gewöhnung an die Sprachproben sollte eigentlich nicht stattfinden, da hierdurch die Urteilsabgabe verfälscht würde. Jedoch konnte dies in dem vorliegenden Beispiel etwas

## 8. Auswertung der Ergebnisse

vernachlässigt werden, da mit dem Szenario 6 ausschließlich ein Vergleich zwischen den Sprechern gezogen werden sollte.

Die Hörschwellenermittlung wurde jeweils mit Hilfe des Versuchsleiters durchgeführt. Somit konnte bei etwaiger falscher Einstellung des Probanden auf einen erneuten Versuch der Ermittlung hingewiesen werden. Die eingestellten Werte unterlagen teilweise großen Schwankungen, jedoch konnte trotzdem von einer Normalhörigkeit aller Probanden ausgegangen werden. Der eingestellte Wertebereich auf der Soundkartensoftware betrug -31,0 bis -44,2 dB. Als Mittelwert ergab sich bei allen Probanden ein Wert von -37,2 dB. In der folgenden Abbildung wird die entsprechende Zuordnung zwischen Soundkartenlautstärke und dem Schalldruck am Kopfhörer verdeutlicht. Zur Ermittlung der Werte wurden für jeden eingestellten Wert an der Soundkartensoftware mehrere Messungen mit dem Schallpegelmesser von Brüel & Kjaer an der Ohrmuschel des Kopfhörers aufgenommen und gemittelt. Die zusätzlich eingefügte Trendlinie verdeutlicht den entsprechenden Verlauf.

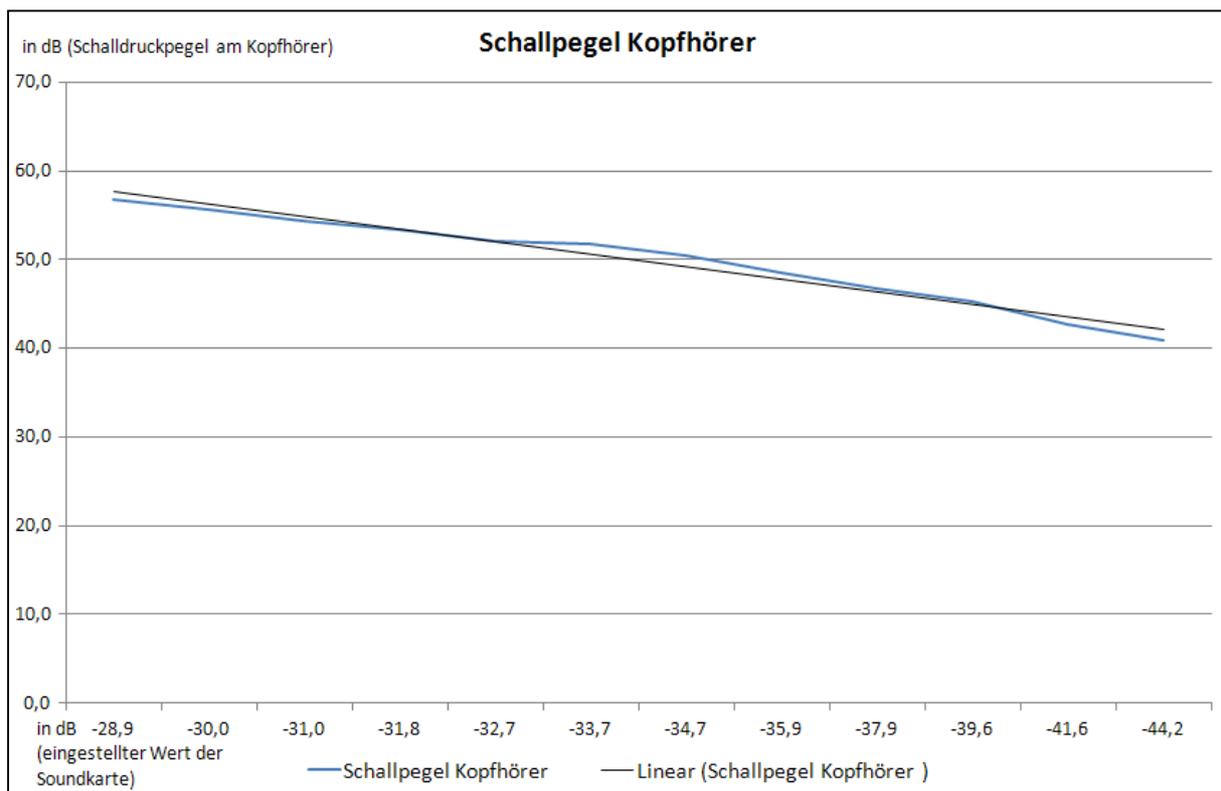


Abbildung 30: Schalldruckpegel am Kopfhörer

Der standardmäßig eingestellte Wert von etwa -37,9 dB entspricht 46,6 dB beim Anhören der Sprachproben unter dem Kopfhörer. Dies ist ungefähr der Wert, den

## 8. Auswertung der Ergebnisse

auch eine Unterhaltung im normalen Umfeld entsprechen würde. Die Extremfälle waren -31,0 dB und -44,2 dB. Diese Werte entsprechen 56,8 dB und 40,8 dB direkt am Kopfhörer gemessen. Der Höchstwert würde schon einer lauten Unterhaltung entsprechen. Dies zeigt, dass ein bis zu 4-facher Schalldruckunterschied bei den Probanden für den Versuch vorlag. Dennoch konnten bei den Bewertungen keine gravierenden Unterschiede festgestellt werden. Die Annahme, dass besonders ältere Versuchsteilnehmer eher höhere Lautstärken eingestellt hätten, traf ebenfalls nicht zu. Die Hörschwellenermittlung streute quer durch die Altersgruppen. Ob dies an wirklichen Hördefiziten oder an einer falschen bzw. ungenauen Einstellung lag, konnte, wie bereits erwähnt, nicht überprüft werden.

### 8.2 Statistische Betrachtungen

#### 8.2.1 Mittelwert, Varianz und Standardabweichung

Für den vorliegenden Versuch ist das wichtigste Vergleichsmittel der Mittelwert oder auch MOS-Wert. Dieser wird nach folgender Formel berechnet:

$$MOS = \mu = \frac{1}{n} \sum_{i=1}^n x_i \quad [8.1]$$

*n...Stichprobenumfang; x...abgegebene Noten*

Zur Berechnung ob und wie stark die abgegebenen Beurteilungen der Probanden um den Mittelwert variieren, werden die Varianz und die Standardabweichung zur Berechnung herangezogen. Die Varianz  $\sigma^2$  berechnet sich wie folgt:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - MOS)^2 \quad [8.2]$$

*n...Stichprobenumfang; x...abgegebene Noten*

## 8. Auswertung der Ergebnisse

Die Standardabweichung  $S$  berechnet sich aus der Wurzel der Varianz:

$$s = \pm\sqrt{\sigma^2} \quad [8.3]$$

Die Standardabweichung gibt den Abstand vom Mittelwert nach oben und unten an und welche Werte sich im Bereich von ca. 2/3 aller Einzelwerte befinden. Bei einem Mittelwert von 4 und einer Standardabweichung von 0,5, würden etwa 66% aller Einzelwerte, statistisch gesehen, zwischen 3,5 und 4,5 liegen.

Zur Berechnung des Gesamtergebnisses  $\mu$  über alle Kriterien wird folgende Formel angewandt:

$$MOS = \mu = \frac{1}{n \cdot k} \sum_{i=1}^n \sum_{j=1}^k x_{ij} \quad [8.4]$$

*n...Stichprobenumfang; k...Anzahl der Kriterien; x...abgegebene Noten*

Die Varianz  $\sigma^2$  für das Gesamtergebnis berechnet sich durch:

$$\sigma^2 = \frac{1}{n \cdot k - 1} \sum_{i=1}^n \sum_{j=1}^k (x_{ij} - MOS)^2 \quad [8.5]$$

*n...Stichprobenumfang; k...Anzahl der Kriterien; x...abgegebene Noten*

### 8.2.2 Konfidenzintervall und Signifikanz

Als Hauptkriterium für die Auswertung wird für jedes Szenario der Mittelwert und das dazugehörige Konfidenzintervall grafisch dargestellt. Anhand des Konfidenzintervalles kann man einen empirisch bestimmten Stichprobenwert mit einer bestimmten Wahrscheinlichkeit abschätzen [Bor93]. Als Wahrscheinlichkeit wird in der Auswertung 95% festgelegt. Somit können die Ergebnisse der Probanden besser

## 8. Auswertung der Ergebnisse

beurteilt werden. Wünschenswert wäre ein schmales Intervall, welches zeigt, dass die Probanden ein ähnliches Ergebnis wahrgenommen haben. In der Statistik werden Unterschiede als signifikant (bedeutsam) bezeichnet, wenn diese durch Zufall nur mit einer bestimmten geringen Wahrscheinlichkeit zustande gekommen sind. Die Größe des Konfidenzintervalles wird für große Stichproben mittels der Normalverteilung und für kleine Stichproben mit der t-Verteilung berechnet. Das Konfidenzintervall wird nach folgender Formel berechnet:

$$\bar{x} \pm q_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \quad [8.6]$$

$\bar{x}$  ... Mittelwert;  $q_{1-\frac{\alpha}{2}}$  ... 0,975-Quantil der Standardnormalverteilung;  $s$  ... Standardabweichung;  
 $n$  ... Stichprobenumfang

Der Wert  $\alpha$  steht für die Irrtumswahrscheinlichkeit (standardmäßig  $\alpha = 0,05$ ). Je geringer dieser Wert ist, desto höher ist die Informationsqualität. Das erhaltene Konfidenzniveau beträgt 95%, das am Anfang und am Ende der Normalverteilung je 2,5% aller ermittelten Werte „abschneidet“. Dieses sogenannte Quantil der zu überprüfenden Signifikanz muss immer vorher festgelegt werden.

### 8.3 Mean Opinion Score für alle Szenarien

Wie bereits im Kapitel 3.1 beschrieben, wird für jedes Kriterium nun ein Mittelwert gebildet, welcher einen durchschnittlichen Meinungswert aufweist. Der sogenannte Mean-Opinion-Score wird anhand der Formeln 8.1 für die beiden Kriterien „Verständlichkeit der Sprache“ und „Einfluss der Umgebung“ berechnet.

In den folgenden Diagrammen soll die Güte der Verständlichkeit der Sprache und die Güte des Störeinflusses der Umgebung dargestellt und eine Rangfolge des jeweiligen Kriteriums aufgezeigt werden.

## 8. Auswertung der Ergebnisse

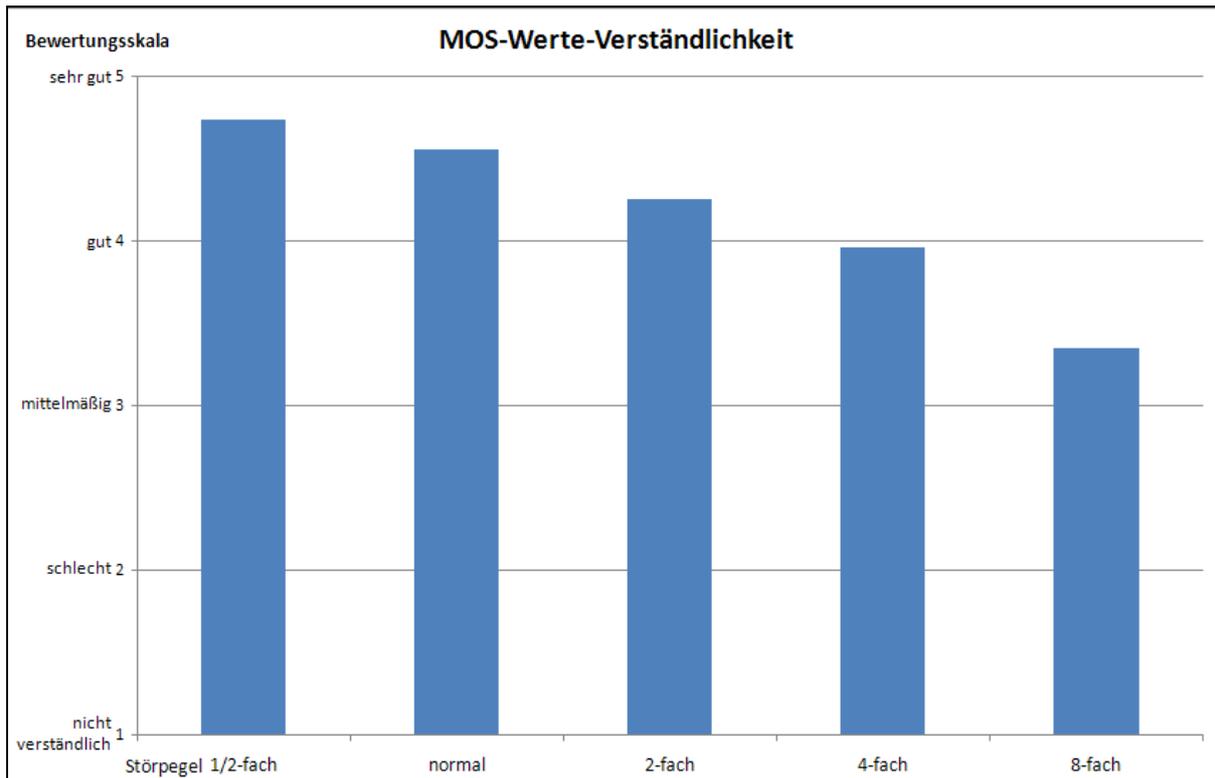


Abbildung 31: MOS-Werte für die Verständlichkeit der Sprache

Für jedes Szenario wurde im ersten Durchgang die Verständlichkeit der Sprache untersucht. Abbildung 31 zeigt die Mittelung der MOS-Werte über alle Szenarien. Die Verständlichkeit bzw. die Sprachqualität kann anhand des MOS-Tests bewertet werden, da eine absolute Bewertung in den Bewertungsbögen gegeben war. Es ist klar zu erkennen, dass nur mit dem halben Schallpegel des Störeinflusses eine bis zu sehr gute Sprachverständlichkeit gegeben ist. Der MOS-Wert entspricht hier 4,74. Beim normalen Störeinfluss der verschiedenen Szenarien liegt das Ergebnis bei einem Wert von 4,56. Bei einer Verdopplung oder sogar Vervierfachung des originalen Störpegels sinkt der MOS-Wert nur gering ab, bleibt aber immer noch mit 4,26 und 3,96 im Bereich von „gut“. Nur bei einer 8-fachen Erhöhung des Störschallpegels tritt eine deutliche Änderung des Ergebnisses ein. Die Sprachverständlichkeit ist hier nur noch mit einem Wert von 3,35 im Bereich von „mittelmäßig“ anzusehen. Bis auf den letzten Wert fällt die Kurve der MOS-Werte relativ linear ab. Dies zeigt, dass eine kontinuierliche Änderung des Störpegels auch zu einer kontinuierlichen Änderung der abgegebenen Bewertung der Probanden führt.

## 8. Auswertung der Ergebnisse

In Abbildung 32 ist die Zusammenfassung aller MOS-Werte für den zweiten Durchgang dargestellt. Hier wurde der Einfluss der Umgebung von den Probanden bewertet.

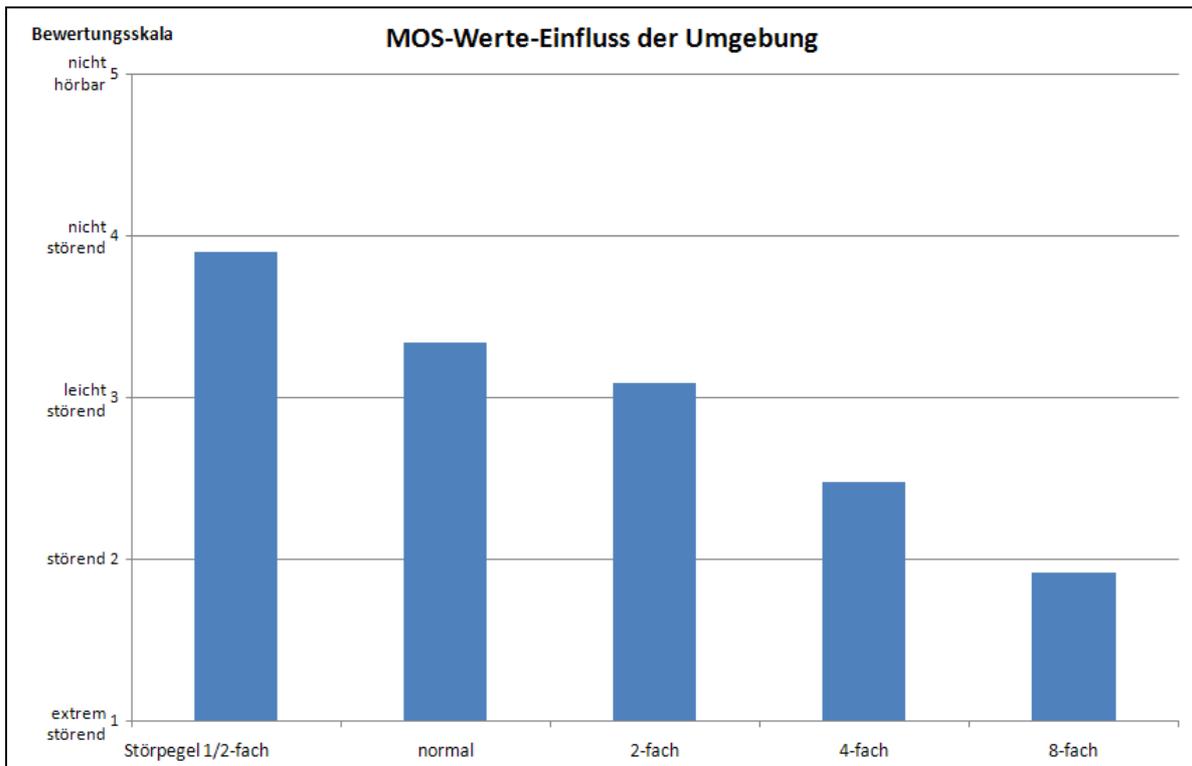


Abbildung 32: MOS-Wert für den Einfluss der Umgebung

Im Vergleich zum ersten Kriterium, der Sprachverständlichkeit, sinken beim Einfluss der Umgebung die MOS-Werte stärker ab. Beim halben Störpegel der Umgebung kann nur noch ein Wert von 3,90 erreicht werden. Der normale Störeinfluss und das Doppelte von ihm, liegen im leicht störenden Bereich. Beide Werte sind mit 3,34 und 3,09 relativ ähnlich. Hingegen weisen die 4-fache und die 8-fache Verstärkung des Originalsignals des Szenarios einen „störenden“ Mittelwert auf. Die Vervielfachung liegt mit ca. 2,5 grenzwärtig zum nur „leichten Stören“ jedoch ist die Verachtfachung eindeutig für die Probanden störend gewesen und wirkt sich wie bereits in Abbildung 31 gezeigt auf die Sprachverständlichkeit bzw. die Sprachqualität aus.

In der nächsten Grafik in Abbildung 33 soll das Gesamtergebnis über die zwei untersuchten Kriterien Sprachverständlichkeit und Einfluss der Umgebung auf die Sprache betrachtet werden. Zur Berechnung wurde Formel 8.4 herangezogen.

## 8. Auswertung der Ergebnisse

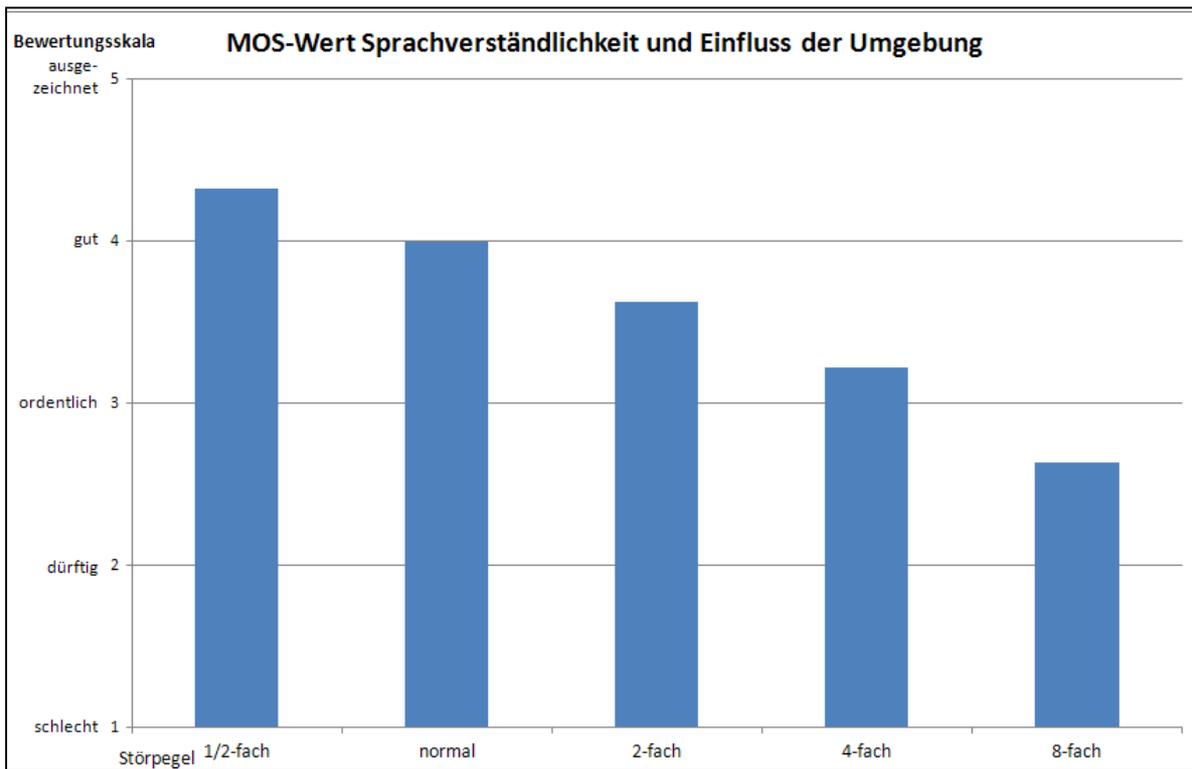


Abbildung 33: MOS-Wert über beide Beurteilungskriterien

Das Gesamtergebnis über beide Kriterien wird mit der absoluten Kategoriebewertung „ausgezeichnet“, „gut“, „ordentlich“, „dürftig“ und „schlecht“ dargestellt. Es ist klar ersichtlich, dass die Grafik zu den bereits gezeigten Abbildungen 31 und 32 nicht stark abweicht. Der lineare Abfall bei höheren Störpegeln wird hier noch etwas deutlicher. Es ist jedoch zu bemerken, dass selbst die Verachtfachung des Schalldruckpegels der Szenarien nur einen MOS-Wert von 2,64 hervorruft und dies immer noch zu einem „ordentlichen“ Ergebnis gezählt werden kann. Bei der Betrachtung der ersten drei Balken wird klar, dass alltägliche Störsituationen keinerlei Einfluss auf Sprachgütemessungen haben dürften. Diese Erkenntnis soll im Weiteren für jedes Szenario separat ausgewertet werden, um den Einfluss der verschiedenen Störgeräusche auf Sprache zu detaillieren. Denn nicht jedes Störgeräusch wird das gleiche Gewicht auf die Urteilsabgabe haben und somit die Bewertung für eine internet-basierte Sprachgütemessung beeinflussen.

### 8.4 Auswertung der Szenarien

Im Folgenden werden alle Szenarien genauer betrachtet. Hierbei werden die Extrembeispiele analysiert, welche Auskunft geben, ob noch eine Verständlichkeit vorliegt bzw. die Störung zu groß ist. Zeigen beide Störpegel ein positives Ergebnis, kann von einem ebenfalls positiven Ergebnis der niedrigeren Pegel ausgegangen werden.

#### 8.4.1 Szenario 1 – „Zimmer an Straße“

In diesem Szenario sollte ein Zimmer an einer befahrenen Straße getestet werden. Die Auswertung sollte zeigen, inwiefern Verkehrsgeräusche Auswirkungen auf die Verständlichkeit von Sprachproben haben.

Im ersten Durchgang wurde die Sprachverständlichkeit von den Probanden bewertet. Abbildung 34 zeigt das nach den Störpegellautstärken geordnete Ergebnis.

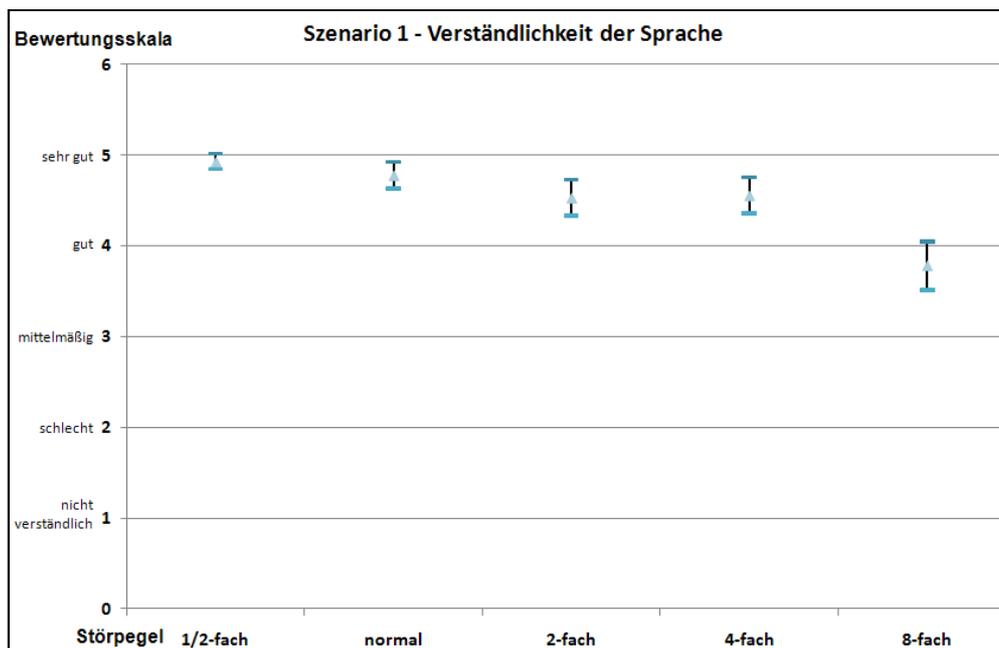


Abbildung 34: Auswertung Szenario 1 – Sprachverständlichkeit

Der Mittelwert (MOS-Wert) ist durch die hellblauen Dreiecke verdeutlicht. Die Konfidenzintervalle werden anhand der senkrechten Linien dargestellt und mit waagerechten Strichen begrenzt. Es ist auf den ersten Blick zu erkennen, dass die Sprachproben immer „sehr gut“ bis „gut“ bewertet wurden. Die Konfidenzintervalle weisen auch keine großen Differenzen unter den Bewertungen auf. Bei nur halbem

## 8. Auswertung der Ergebnisse

Störeinfluss beurteilten die Testteilnehmer mehrheitlich die Probe als sehr gut verständlich. Die größten Schwankungen gab es erst bei der 8-fachen Lautstärkeerhöhung des Störschallpegels. Jedoch ist auch hier zu bemerken, dass das Ergebnis trotzdem eindeutig für „gut“ befunden wurde. Scheinbar machte es auch keinen großen Unterschied, ob der doppelte oder der 4-fache Störpegel in diesem Szenario vorlag. Beide Fälle wurden im Mittel gleich bewertet. Allein die Schwankung des Ergebnisses ist beim doppelten Pegel der Umgebung etwas größer. Die genauen Werte können im Anhang eingesehen werden.

In der folgenden Abbildung wird der 2. Durchgang grafisch dargestellt. Hier wurde der Störeinfluss der Umgebung bewertet.

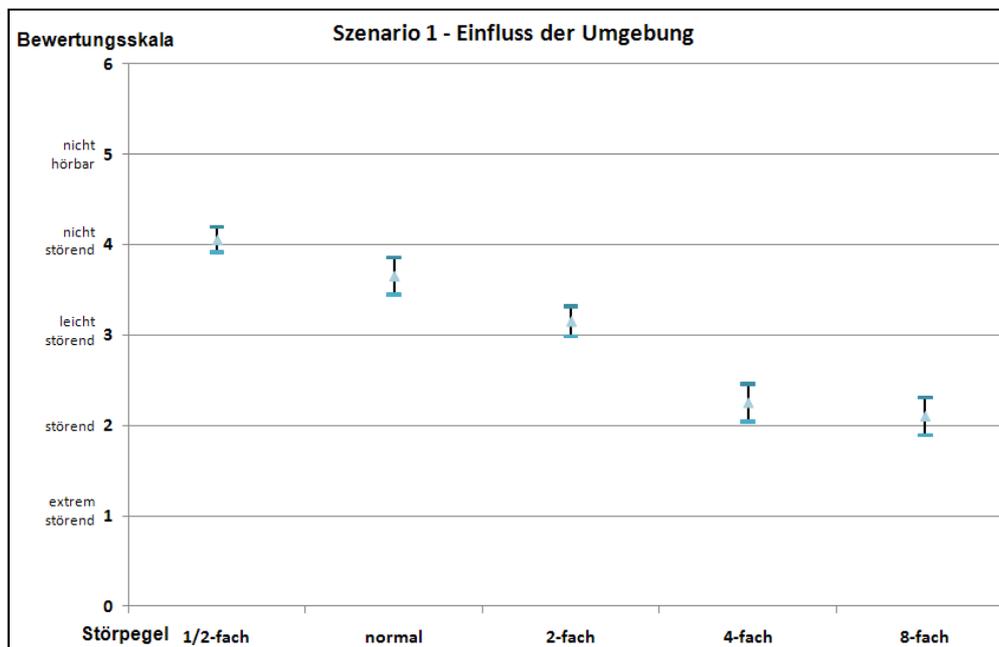


Abbildung 35: Auswertung Szenario 1 – Einfluss der Umgebung

Im Gegensatz zum ersten Durchgang sind hier stärkere Schwankungen in den Ergebnissen zu erkennen. Jedoch wird auch ersichtlich, dass die Bewertungen der Probanden relativ einheitlich waren. Anhand von Spektrogrammen sollen die beiden am meisten störenden Hörbeispiele mit dem 4- und 8-fachen Störpegel detaillierter untersucht werden.

Abbildung 36 zeigt im rechten Spektrogramm die Sprachprobe addiert mit dem 4-fachen Störschallpegel, welche den Probanden dargeboten wurde. Im Hörbeispiel ist folgender Satz zu hören: „Das Identifizieren von sprachlichen Einheiten, zum Beispiel von Sprachlauten, Silben oder ganzen Wörtern, aber auch von prosodischen

## 8. Auswertung der Ergebnisse

Elementen, durch automatische Klassifikation von Sprachsignalen.“ Man kann gut erkennen, dass die markante Stimme des Sprechers große Anteile im Frequenzbereich zwischen 60 und 300 Hz aufweist, wie es typisch für tiefe männliche Stimmen ist. Dieser Bereich ist allerdings nur wenig im Umgebungsspektrum wahrzunehmen. Der Effekt der akustischen Verdeckung könnte allerdings im hohen bis mittleren Frequenzbereich auftreten, denn wie die Spektrogramme sehr gut zeigen, haben der Störpegel und die Sprache im Bereich von 400 Hz bis 3 kHz etwa die gleichen Pegel. Hierdurch könnten wahrscheinlich Verständlichkeitsprobleme bei den Probanden aufgetreten sein.

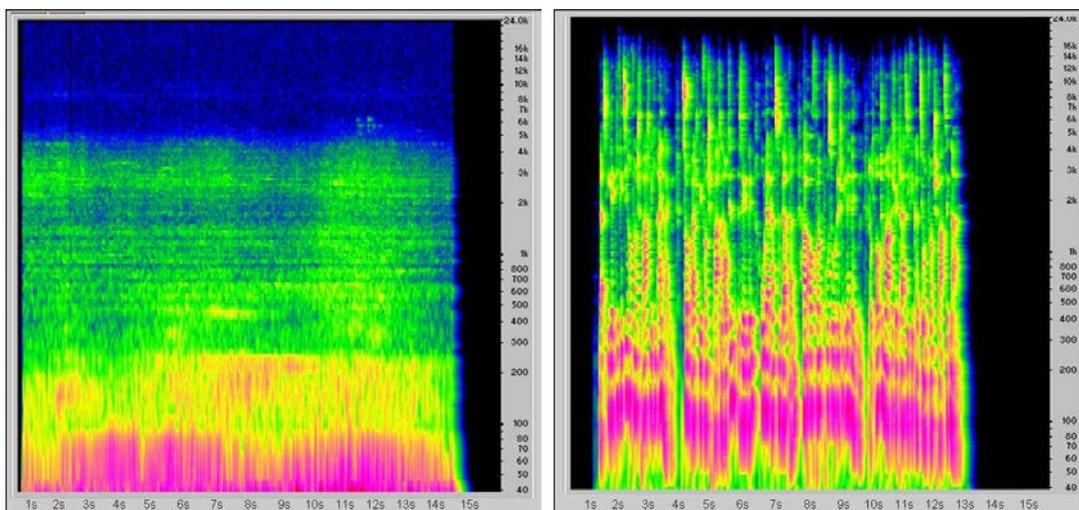


Abbildung 36: Szenario 1 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Der für die Sprachverständlichkeit sehr wichtige Bereich um 4-8 kHz ist jedoch nicht von der Verdeckung betroffen. Hier weist eindeutig das Sprachbeispiel den höheren Schallpegel auf und müsste so noch zum Verstehen beigetragen haben, wie man am Diagramm für die Sprachverständlichkeit gut erkennen kann.

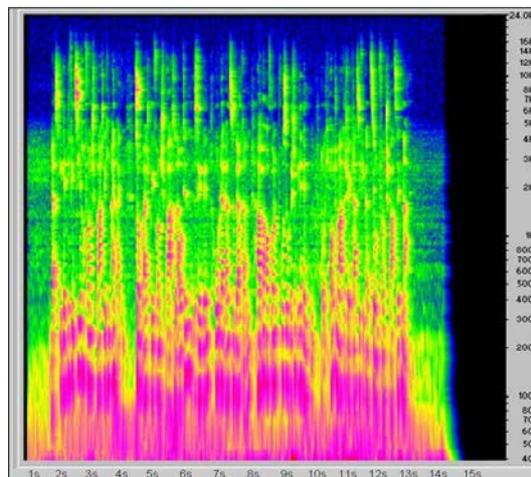


Abbildung 37: Szenario 1 Sprache und 4-facher Störpegel

## 8. Auswertung der Ergebnisse

Nach der Addition der beiden Spektrogramme aus Abbildung 36 lässt sich das Sprachsignal immer noch gut erkennen. Dies bestärkt trotz des hohen Störpegels ebenfalls die Bewertung aus Abbildung 34.

Laut der Bewertung der Testteilnehmer wirkte sich der 8-fache Störpegel des Originalszenarios stärker, als der 4-fache Pegel aus. Das soll an den nächsten Spektrogrammen dargestellt werden.

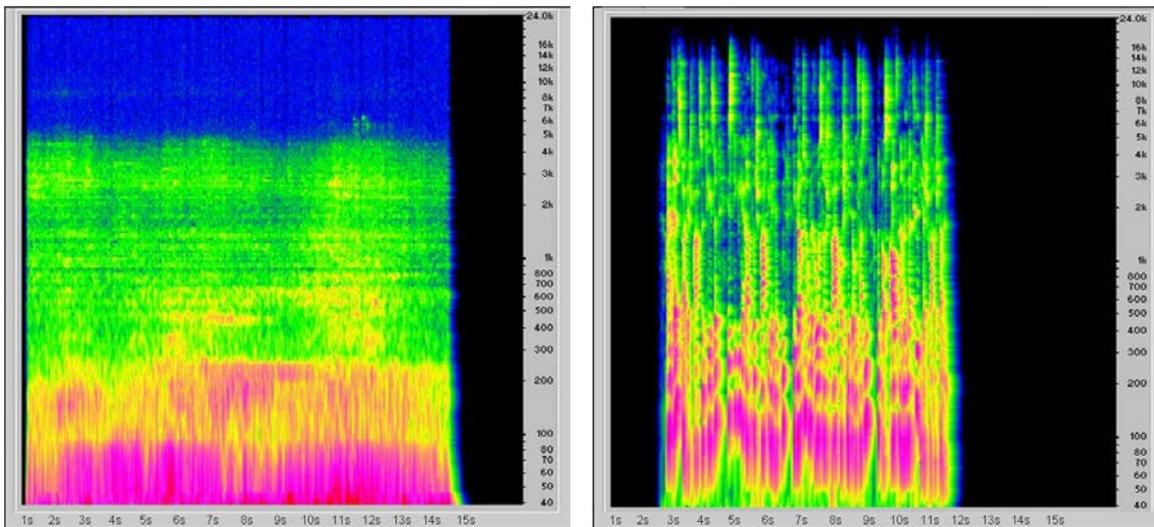


Abbildung 38: Szenario 1 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

In Abbildung 38 wird sichtbar, dass ein höherer Störpegel vorliegt. Alle Farbbereiche treten intensiver auf, was für einen höheren Pegel in den einzelnen Frequenzbereichen spricht.

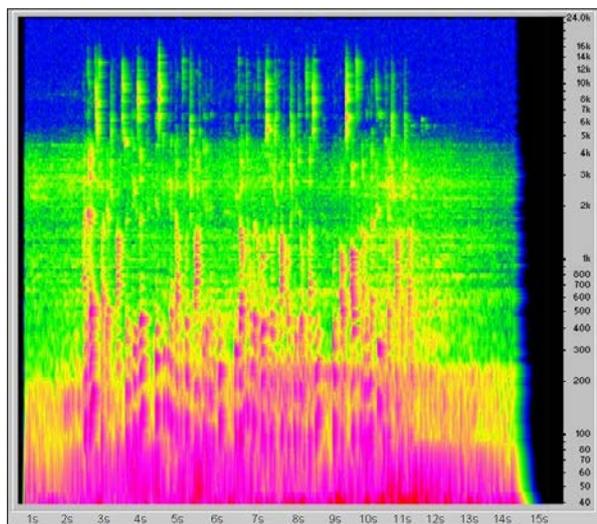


Abbildung 39: Szenario 1 Sprache und 8-facher Störpegel

## 8. Auswertung der Ergebnisse

Der Testsatz „Ein Grenzfall der Sprachsynthese, ist die Wiedergabe von Sprachelementen, die Wörter oder größere Sprachausschnitte umfassen und aus Aufnahmen natürlicher Sprache stammen“, ist vom Spektrogramm her ähnlich dem des vorangegangenen Testsatzes. Somit kommen hier gleiche Frequenzanteile zur Geltung.

Die Addition des Testsatzes und des Umgebungsgeräusches ist in Abbildung 39 bildlich dargestellt. Trotz des hohen Störsignals kann das Sprachmaterial gut erkannt werden. Dies erklärt die Bewertung der Testpersonen, dass das Umgebungsszenario störend ist, aber trotzdem eine gute Sprachverständlichkeit vorliegt.

Dieser Extremfall mit der 4- oder auch 8-fachen Lautstärke des Störeinflusses wird praktisch in der Realität nicht auftreten, und man kann sagen, dass normaler Verkehrslärm eine internet-basierte Sprachgütemessung nicht negativ beeinflussen würde.

### 8.4.2 Szenario 2 – „Zimmer nebenan Küchengeräusche und TV“

In der zweiten Umgebung wurde ein Abhörort demonstriert, der wie bereits Szenario 1 einer Wohnumgebung entsprach. Im angrenzenden Zimmer (der Küche) traten Fernseh- sowie Küchengeräusche auf.

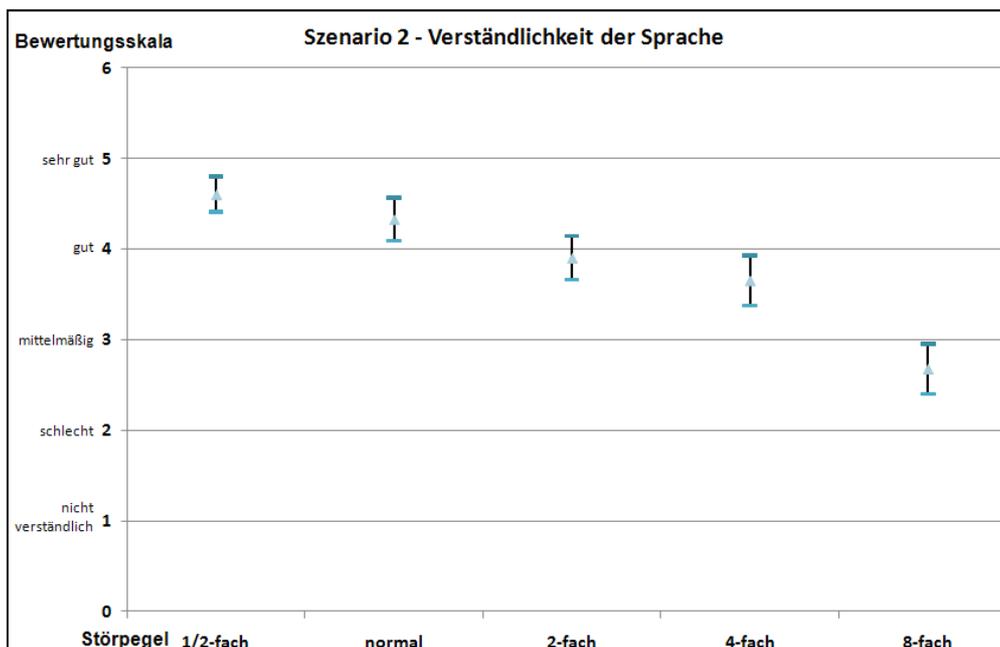


Abbildung 40: Auswertung Szenario 2 – Sprachverständlichkeit

## 8. Auswertung der Ergebnisse

Es wurde nun getestet, ob sich diese teilweise hochfrequenten Signalanteile und die Sprache des Fernsehsers störend auf die Sprachgütemessung auswirken würden.

Die Sprachverständlichkeit lies bei diesem Szenario etwas nach (Abbildung 40). Dies ist sicher den bereits erwähnten Umgebungsbedingungen geschuldet. Dennoch unterliegen die Einzelergebnisse wiederum keiner großen Streuung, wie man sehr gut am Konfidenzintervall erkennen kann. Bei der 8-fachen Erhöhung des Störpegels bewerteten die Testteilnehmer die Sprachverständlichkeit nur noch als „mittelmäßig“, was bereits auf eine große Störung hinweist. Dies wird im Folgenden an den Spektrogrammen verdeutlicht.

In Abbildung 41 ist der zweite Durchgang des Szenarios 2 zu sehen. Die Bewertung sinkt hier erneut ab, da die Probanden den Störeinfluss bewerten sollten und sich somit mehr auf die Umgebungsgeräusche konzentrierten.

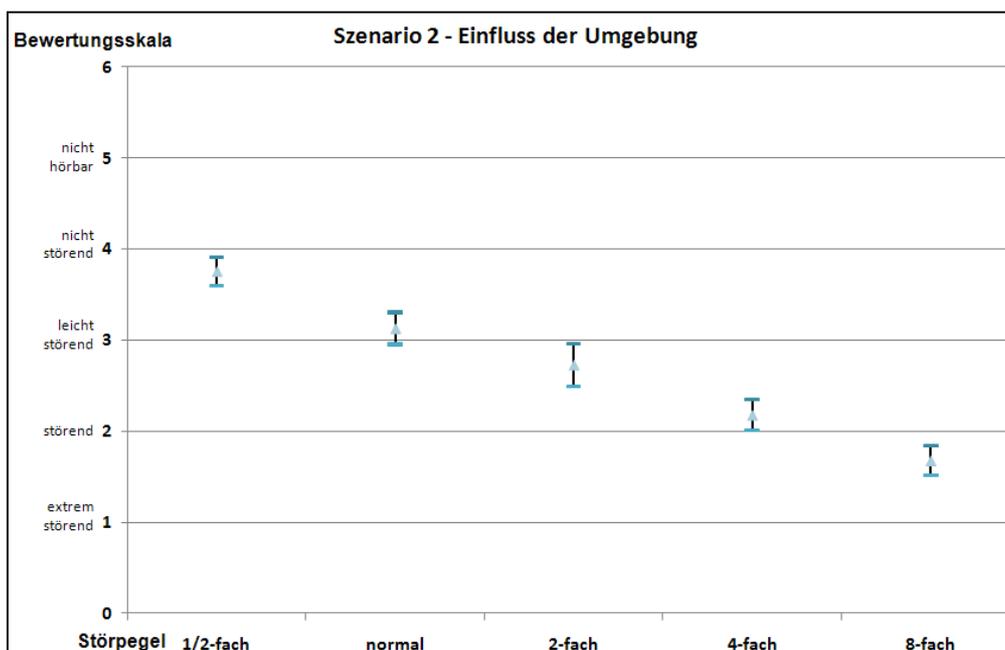


Abbildung 41: Auswertung Szenario 2 - Einfluss der Umgebung

Es ist gut zu erkennen, dass die Schwankungen in den Bewertungen der Probanden nicht so gravierend sind, wie es im ersten Durchgang bei der Sprachverständlichkeit zu sehen war. Es ist bei diesem Szenario sogar ein fast linearer Abfall pro Erhöhung des Störpegels um je 6 dB zu verzeichnen. Die nun folgende Auswertung anhand der Spektrogramme mit den beiden Extremfällen, der 4- und 8-fachen Lautstärke des

## 8. Auswertung der Ergebnisse

Störpegels soll die gravierenden Einflüsse auf die Sprachverständlichkeit verdeutlichen.

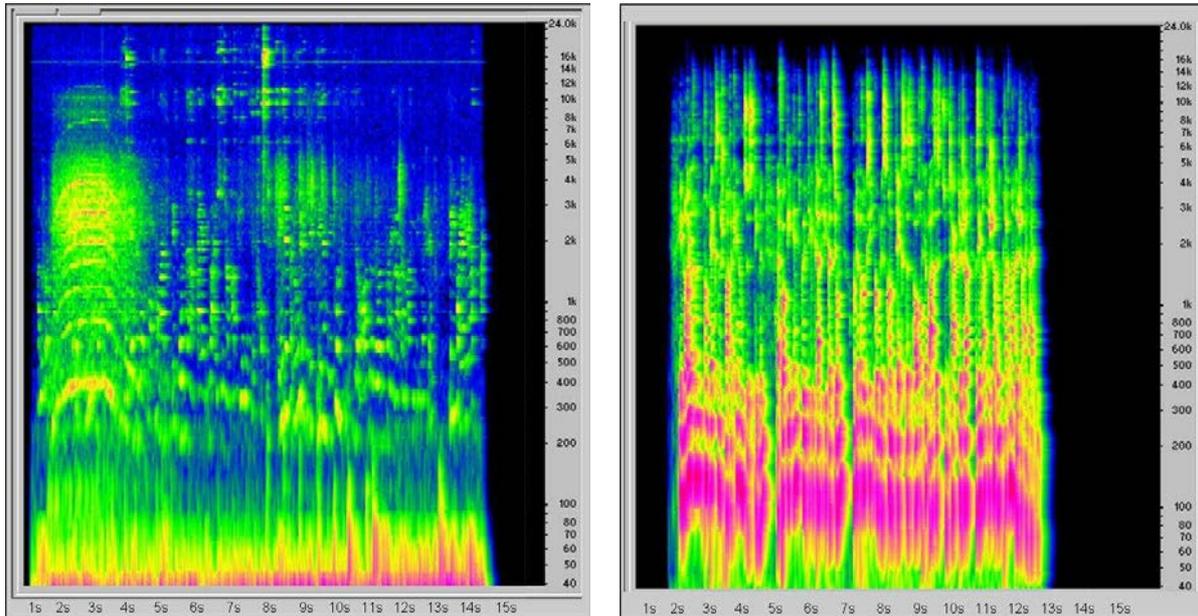


Abbildung 42: Szenario 2 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Im linken Spektrogramm in Abbildung 42 ist das Umgebungsgeräusch dargestellt. Dieses weist besonders in den ersten vier Sekunden durch die Brotschneidemaschine einen sehr hohen Störanteil über die wichtigen Frequenzbereiche der Sprache auf. Vor allem diese ersten Augenblicke des Hörbeispiels werden sehr störend auf die Probanden gewirkt haben, wie dies auch im Diagramm in Abbildung 40 deutlich wird. Im Vergleich des Störpegels und des Sprachbeispiels ist gut zu erkennen, dass der Testsatz: „In neueren komplexen Spracherkennungssystemen erfolgt die endgültige Klassifikation unter Berücksichtigung des Kontextes auch auf der Basis syntaktischer, semantischer und pragmatischer Einheiten.“, dennoch gut von den Probanden aus dem Störpegel herausgefiltert werden konnte. Dies wird in der Addition der beiden Signale in Abbildung 43 nochmals verdeutlicht. Somit wird das Ergebnis der Probanden, welche die Sprachverständlichkeit durchschnittlich mit 3,65 bewertet haben, bestärkt. Es ist somit immer noch eine „gute“ bis „mittelmäßige“ Verständlichkeit der Sprache gegeben.

## 8. Auswertung der Ergebnisse

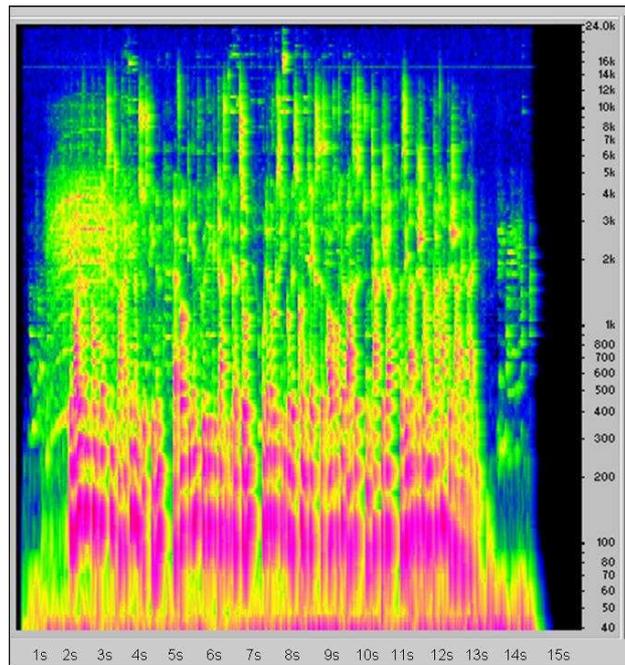


Abbildung 43: Szenario 2 Sprache und 4-facher Störpegel

Es wird auch gut sichtbar, dass der tieffrequente Bereich, welcher für das Volumen der Sprache steht, in diesem Szenario nicht stark durch die Umgebung gestört wird. Alle wichtigen Frequenzinformationen der Sprache bleiben weitestgehend beim 4-fachen Störpegel erhalten und können durch die Testteilnehmer noch gut ausgewertet werden.

Die folgenden zwei Grafiken zeigen das Spektrogramm der Umgebung mit der 8-fachen Erhöhung des normalen Störgeräusches des Szenarios und dem dazugehörigen Sprachbeispiel.

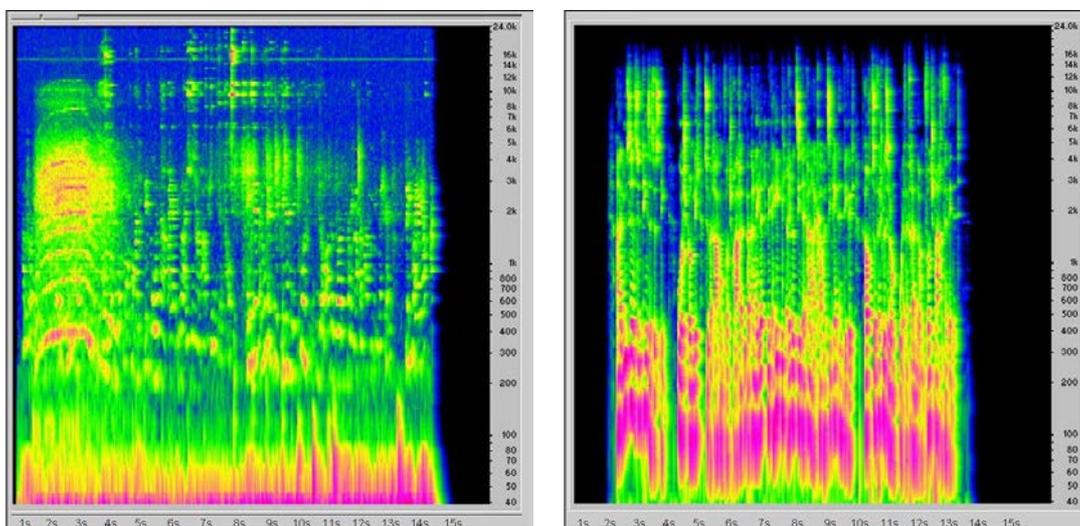


Abbildung 44: Szenario 2 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

## 8. Auswertung der Ergebnisse

Vergleicht man das linke Spektrogramm in Abbildung 42 mit dem in Abbildung 44, wird schnell deutlich, dass hier ein wesentlich höherer Störeinfluss der Umgebung eintreten wird. Dies wird auch durch die Bewertungen der Testteilnehmer verdeutlicht, welche die 8-fache Erhöhung als „störend“ empfanden. Durch diesen hohen Störpegel leidet die Sprachverständlichkeit ebenfalls enorm. Dies wird durch die Wertung der Probanden ebenfalls bestärkt. Sie beurteilten die Verständlichkeit im Durchschnitt nur noch mit 2,68. Dieser Wert entspricht nur noch einer mittelmäßigen Verständlichkeit.

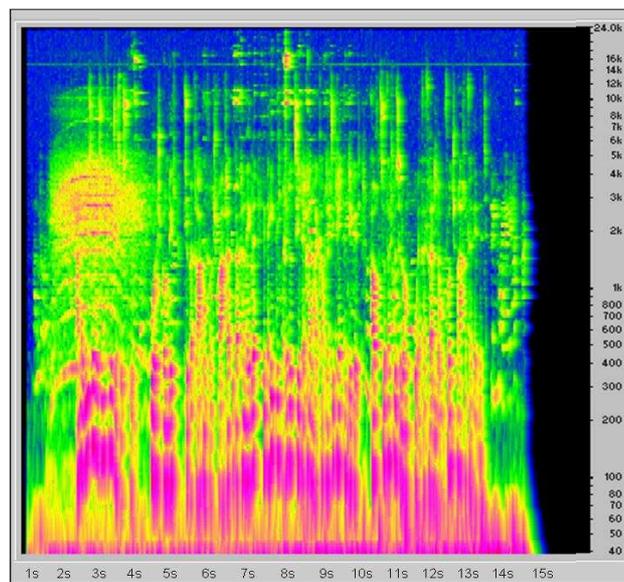


Abbildung 45: Szenario 2 Sprache und 8-facher Störpegel

Die Addition von Sprach- und Störsignal bestätigt die Bewertungen der Probanden ebenfalls. Es ist nur noch schwer möglich, detailliert das Sprachsignal vom Signal der Umgebung zu unterscheiden, da das Störsignal die Sprache annähernd verdeckt. Im hochfrequenten Bereichen ab 5 kHz kann das Sprachsignal noch identifiziert werden. Da dieser Frequenzbereich einen wichtigen Teil der Sprachverständlichkeit ausmacht, konnte wahrscheinlich dennoch ein akzeptables Ergebnis erzielt werden.

Es muss jedoch auch hier gesagt werden, dass kein Testteilnehmer unter diesen Extrembedingungen einen Hörtest durchführen würde und somit diese ausgeprägten Störpegel nicht auftreten werden. Somit wird sich der Störpegel immer in einem nicht bis leicht störenden Bereich bewegen. Damit ist eine Beurteilung für die Probanden immer noch sehr gut möglich.

### 8.4.3 Szenario 3 – „Raum 321 mit Baustellenlärm“

Im nächsten Szenario sollte untersucht werden, inwiefern sich Baulärm negativ auf Sprachgütemessungen auswirken würde. Diese Umgebung sollte möglichst nicht sehr störend für die Probanden sein, da sie täglich mit dieser Situation im Lehrgebäude 3A konfrontiert sind.

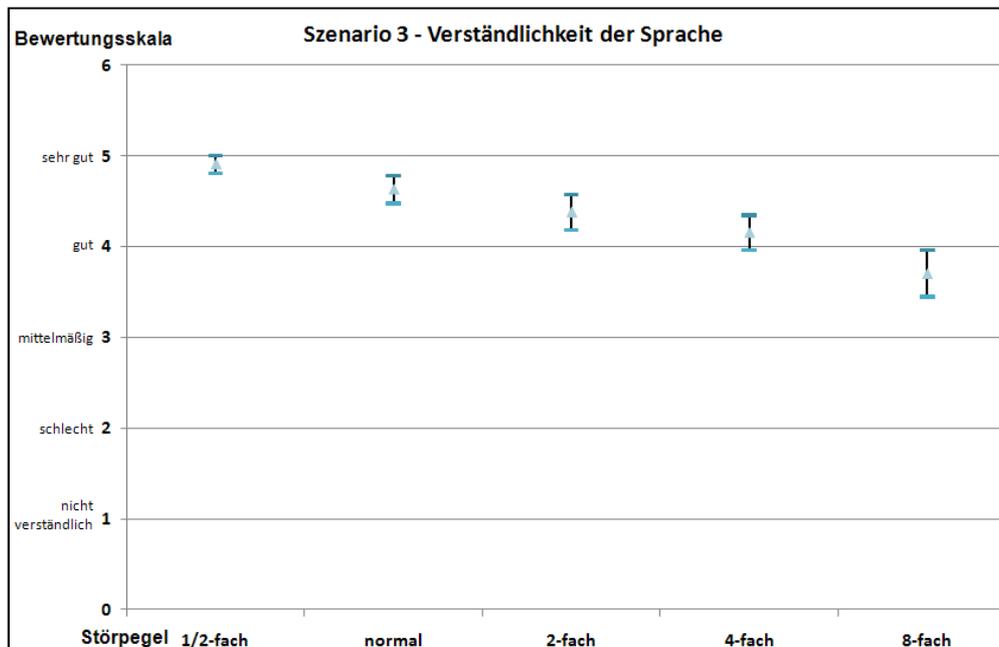


Abbildung 46: Auswertung Szenario 3 – Sprachverständlichkeit

Wie bereits bei den zuvor ausgewerteten Szenarien, sinkt die Sprachverständlichkeit relativ linear mit der Erhöhung des Störpegels ab. Ungeachtet des hohen Störpegels bei der Verachtfachung, beurteilten die Probanden die Verständlichkeit mit „gut“. Bis zum 4-fachen Störpegel tendierten die Ergebnisse sogar zu einer sehr guten Sprachverständlichkeit bzw. Sprachqualität.

Im zweiten Durchgang wurde dieses Ergebnis erneut etwas relativiert. In Abbildung 47 ist ersichtlich, dass das Baustellenszenario die Testteilnehmer bereits stark stört. Dies wird wiederum besonders bei der Vervielfachung und der Verachtfachung des störenden Signals deutlich. Wie im vorangegangenen Durchgang ist in der Grafik ein linearer Abfall zu sehen. Je stärker das Störsignal ist, desto mehr empfinden die Probanden die Umgebung als störend. Dies ist ein sehr gutes Ergebnis für die Untersuchung und zeigt auch, dass sich die Probanden gut in das Szenario hineinversetzen und real erleben konnten.

## 8. Auswertung der Ergebnisse

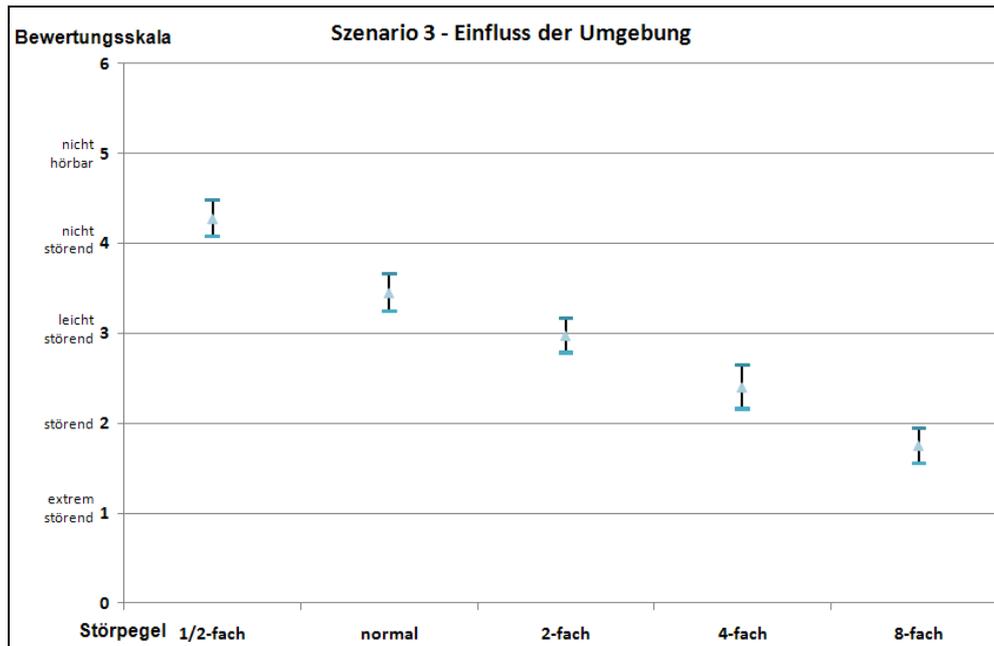


Abbildung 47: Auswertung Szenario 3 - Einfluss der Umgebung

Auch hier wird durch das Konfidenzintervall erneut sichtbar, dass die Streuung der Bewertungen nur in ganz geringem Maße variiert und wiederum von einem aussagekräftigen Ergebnis ausgegangen werden kann.

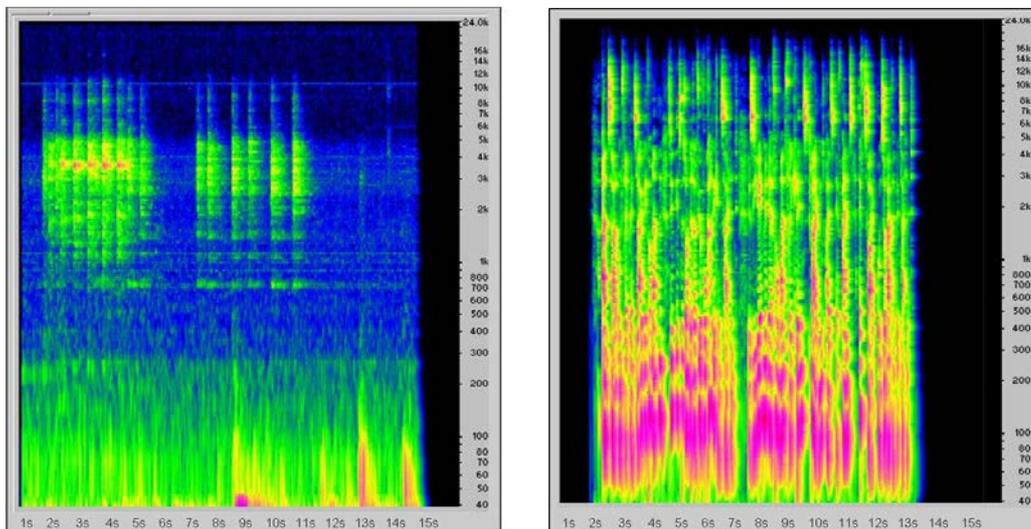


Abbildung 48: Szenario 3 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Das linke Spektrogramm für das Umgebungsszenario weist im Bereich von 20 Hz bis 300 Hz einen verstärkten Schallanteil auf. Dies dürfte allerdings das Sprachsignal nicht markant beeinflussen, wie im rechten Bild zu sehen ist. Die punktuellen Signalanteile im Bereich von 1 kHz bis 12 kHz sind ebenfalls nicht so stark und

## 8. Auswertung der Ergebnisse

durchgängig, dass sie die Sprachverständlichkeit einschränken dürften. Dies spiegelt sich auch in der Bewertung der Probanden wider, da sie alle Hörbeispiele mit „sehr gut“ und „gut“ beurteilt haben.

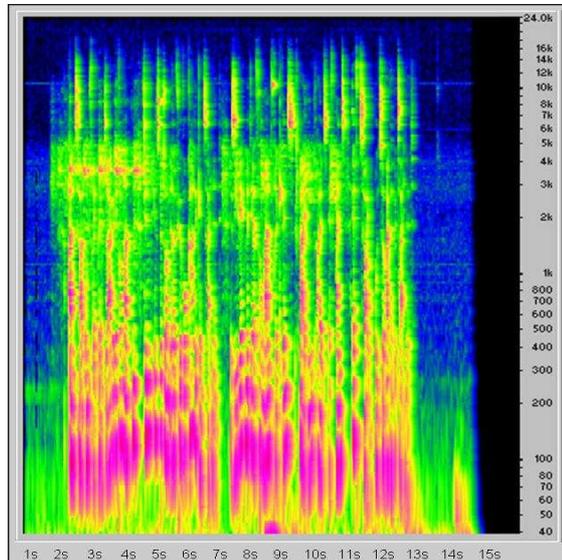


Abbildung 49: Szenario 3 Sprache und 4-facher Störpegel

In der Addition der beiden Signale ist das Sprachsignal noch gut zu erkennen. Dies bestärkt das Ergebnis der Probanden, welche den Störeinfluss der Umgebung mit „leicht störend“ bis „störend“ für die 4-fache Erhöhung des Störpegels beurteilten.

Im Folgenden soll der 8-fache Störeinfluss der Baustellenumgebung näher betrachtet werden.

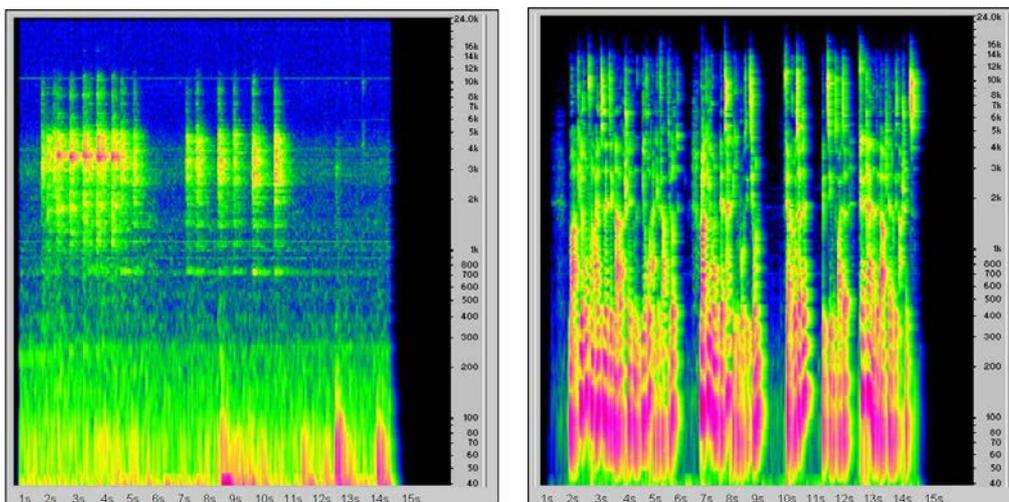


Abbildung 50: Szenario 3 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

## 8. Auswertung der Ergebnisse

Anhand der kräftigeren Farbtöne im Spektrogramm auf der linken Seite wird deutlich, dass das Störsignal stärker geworden ist und mehr Einfluss auf die Verständlichkeit der Sprache haben wird. Die ersten 11 Sekunden werden demzufolge einen negativen Einfluss auf die Beurteilung der Testteilnehmer haben. Diese Vermutung aus den Spektrogrammen spiegelt sich auch in der Wertung der Probanden wider. Somit bewerteten fast alle Teilnehmer die 8-fache Lautstärkeerhöhung als störend.

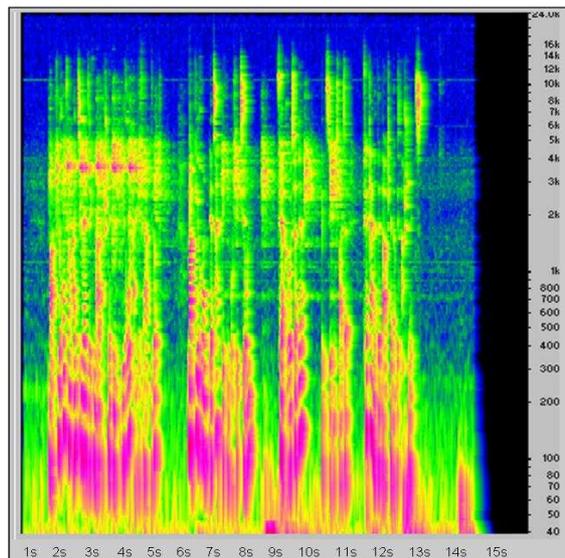


Abbildung 51: Szenario 3 Sprache und 8-facher Störpegel

Durch das Zusammenfügen beider Signale wird deutlich, dass die Sprachprobe in den ersten Sekunden sehr stark gestört wird und eine ausreichende Verständlichkeit des Gesagten nicht mehr gegeben ist. Unter diesen Bedingungen wäre eine Sprachgütemessung nicht mehr sinnvoll. Das Ergebnis dieses Szenarios zeigt, dass eine gute Verständlichkeit des gegebenen Testsatzes: „Markierung von einzelnen Signalabschnitten, die rot eingefärbt wird, Markierungen der zeitlichen Position, vergrößerte bzw. verkleinerte Darstellung des Abschnitts.“ nur noch bedingt möglich ist. Die besondere Schwierigkeit liegt hier sicher auch am Testmaterial, welches für die Probanden eher ungewohnt schien.

Zusammenfassend lässt sich dennoch auch für das 3. Szenario sagen, dass diese extremen Lautstärkebedingungen nicht bei einer derartigen Untersuchung gegeben sein werden. Bis zum doppelten Störpegel bewerteten die Probanden die Verständlichkeit mit „sehr gut“ bis „gut“ bewerten und empfanden nur eine leichte Störung durch die Umgebung.

### 8.4.4 Szenario 4 – „Raum 119 PC Pool mit mehreren Computerlüftern“

Das vierte Szenario sollte eine der typischen Umgebungen für internet-basierte Sprachgütemessungen widerspiegeln. Es wurde hier eine virtuelle Testumgebung durch den PC-Pool des Lehrstuhles erzeugt. Markant an einer Computerpoolumgebung sind die Lüftergeräusche der vorhandenen PC.

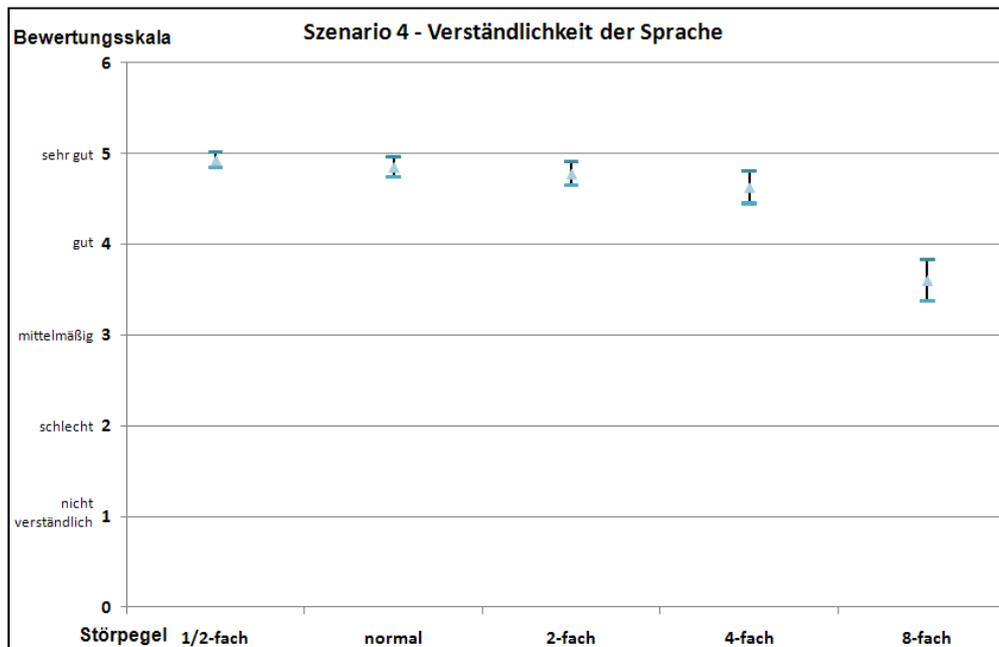


Abbildung 52: Auswertung Szenario 4 – Sprachverständlichkeit

In Abbildung 52 ist zu sehen, dass die Computerlüfter des PC-Pools keinen markanten Einfluss auf die reine Sprachverständlichkeit haben. Bis zur Verdopplung des Störgeräusches haben die Probanden die Verständlichkeit mit „sehr gut“ beurteilt. Beginnend ab einer Vervielfachung sinkt der MOS-Wert auf 4,63 ab. Dies würde aber immer noch für „sehr gut“ stehen. Erst ab dem 8-fachen Störpegel tritt für die Probanden eine Verschlechterung ein, welche aber immer noch als eine „mittelmäßige“ bis „gute“ Sprachverständlichkeit empfunden wurde.

Im zweiten Durchgang wurde bei der Bewertung des Störeinflusses der Umgebung ein ähnliches Ergebnis gewonnen. Hierbei beurteilten die Testpersonen die Hälfte des originalen Störsignals und das originale Lüfterrauschen als „nicht hörbar“ bzw. mit einer leichten Tendenz zu „nicht störend“. Bei einer Verdopplung oder Vervielfachung des Störeinflusses empfanden die Probanden das Rauschen immer noch im Durchschnitt als „nicht störend“. Erst bei einer 8-fachen Erhöhung des Rauschsignals wurde der Einfluss der Umgebung als störend beurteilt.

## 8. Auswertung der Ergebnisse

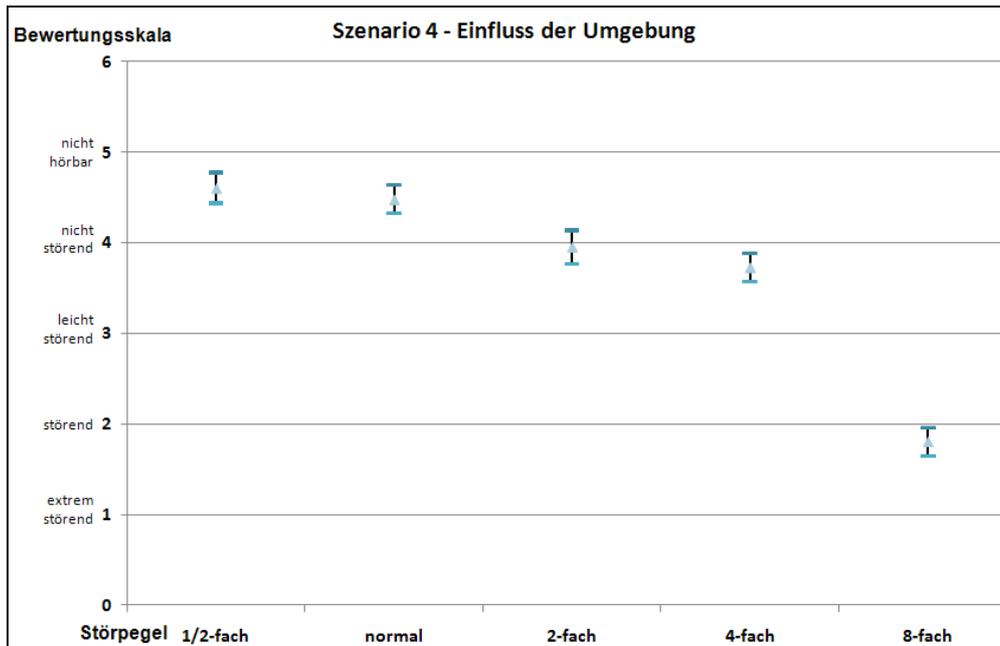


Abbildung 53: Auswertung Szenario 4 - Einfluss der Umgebung

Bei normalen Lüftergeräuschen kann also von einer geringen, aber nicht gravierenden Störung ausgegangen werden. Die beiden Extremfälle sollen im Folgenden näher analysiert werden.

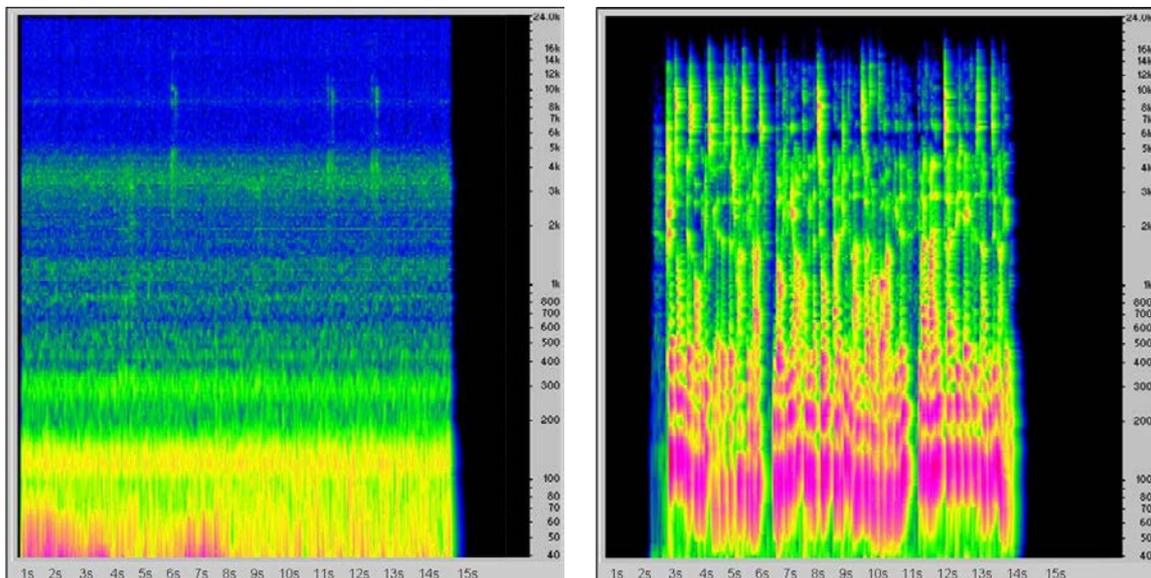


Abbildung 54: Szenario 4 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Auf den ersten Blick ist zu erkennen, dass sich die Lüftergeräusche verstärkt nur im Bereich zwischen 20 Hz und etwa 600 Hz bewegen. Diese Bereiche sind nicht ganz so gravierend und haben keinen großen Einfluss auf die Sprachverständlichkeit, wie in Abbildung 52 anhand des MOS-Wertes und des Konfidenzintervalles bereits

## 8. Auswertung der Ergebnisse

gezeigt wurde. Das etwas höherfrequente Signal zwischen 3 und 4 kHz dürfte ebenfalls keinen merklichen Einfluss auf das Sprachsignal haben, da dieser Frequenzbereich nur mit geringem Pegel auftritt.

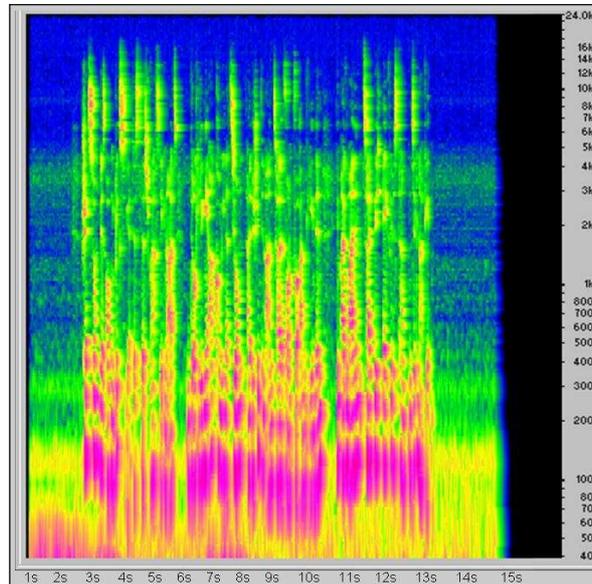


Abbildung 55: Szenario 4 Sprache und 4-facher Störpegel

Abbildung 55 macht deutlich, dass das Rauschen mit 4-facher Erhöhung keinen merklichen Einfluss auf das Sprachbeispiel hat. Man kann noch gut das rechte Spektrogramm aus Abbildung 54 in der Addition erkennen. Dies bestätigt auch die Bewertung der Probanden, welche den 4-fachen Störpegel eigentlich immer noch für nicht störend empfanden.

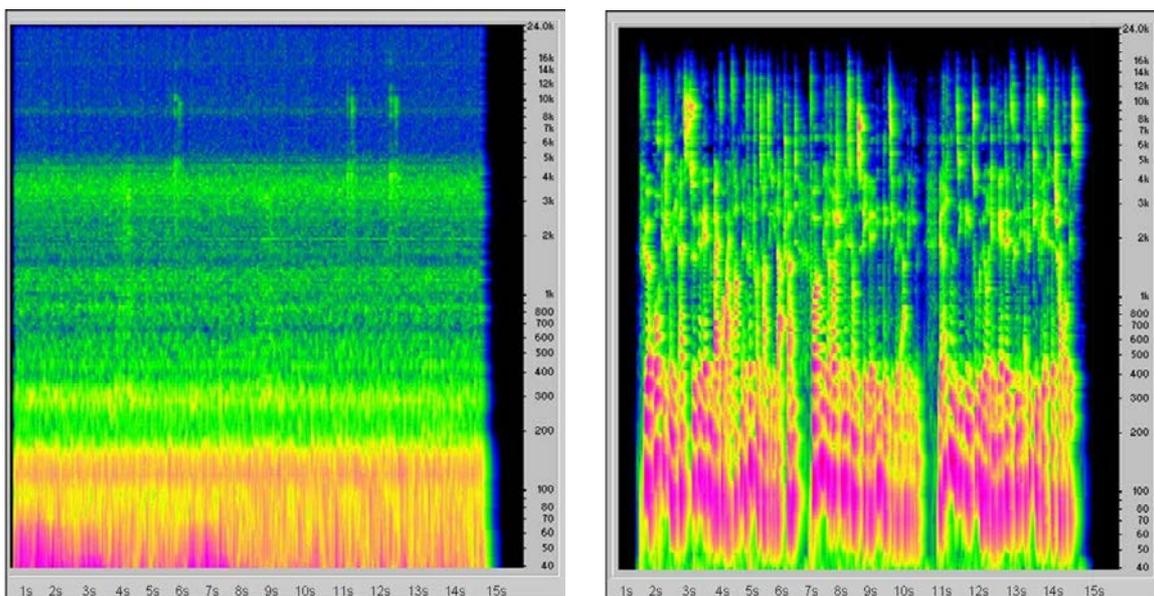


Abbildung 56: Szenario 4 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

## 8. Auswertung der Ergebnisse

Bei einer Verachtfachung des Störpegels sinkt die Verständlichkeit der Sprache ab, da sich nun das Störgeräusch intensiver über den Frequenzbereich legt. Dies ist auch in den Bewertungen der Probanden wahrnehmbar. Die Verständlichkeit wird nur noch für „gut“ befunden und der Einfluss der Umgebung als „störend“. Bei näherer Betrachtung der beiden Spektrogramme wird diese Beurteilung schnell kenntlich. Besonders verdeutlicht wird dies durch die Addition des Sprach- und des Störsignals. Das entsprechende Spektrogramm ist in Abbildung 57 zu sehen.

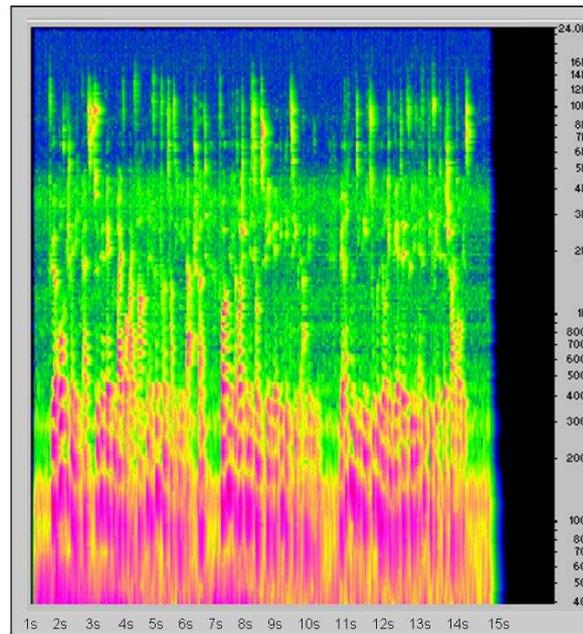


Abbildung 57: Szenario 4 Sprache und 8-facher Störpegel

Eine Reproduktion des reinen Sprachsignals fällt hier schwer. Dies zeigt, dass ein lautes Rauschen eine Sprachgütemessung stark negativ beeinflusst. Jedoch handelt es sich hierbei wieder um einen Extremfall. Die Analyse mit den MOS-Werten zeigt für die Sprachverständlichkeit und den Einfluss der Umgebung, dass sogar bei einer bis zur 4-fachen Erhöhung des Rauschens eines normalen Computerpools keine markanten Bewertungsschwierigkeiten auftreten würden. Somit kann man sagen, dass eine Sprachgütemessung unter diesen Bedingungen durchaus durchführbar wäre.

### 8.4.5 Szenario 5 – „weißes Rauschen (MNRU)“

Im Szenario 5 sollte anhand der Modulated Noise Reference Unit (MNRU) der Störeinfluss über den kompletten Frequenzbereich von 20 Hz bis 20 kHz getestet werden. Im Versuch wurde hierbei ein weißes Rauschen zum Sprachsignal addiert. Das Störsignal erstreckte sich hier weitestgehend über den kompletten Frequenzbereich mit gleichbleibender Amplitude. Demzufolge war zu erwarten, dass je nach Zumischung zum Nutzsignal die Sprachverständlichkeit leiden bzw. sich der störende Einfluss auf die Umgebung erhöhen würde. Hierdurch kann sehr gut ein allgemeiner Test durchgeführt werden, da der Störeinfluss nicht nur punktuell in gewissen Frequenzbereichen auftritt, sondern sich kontinuierlich über den gesamten Hörbereich erstreckt.

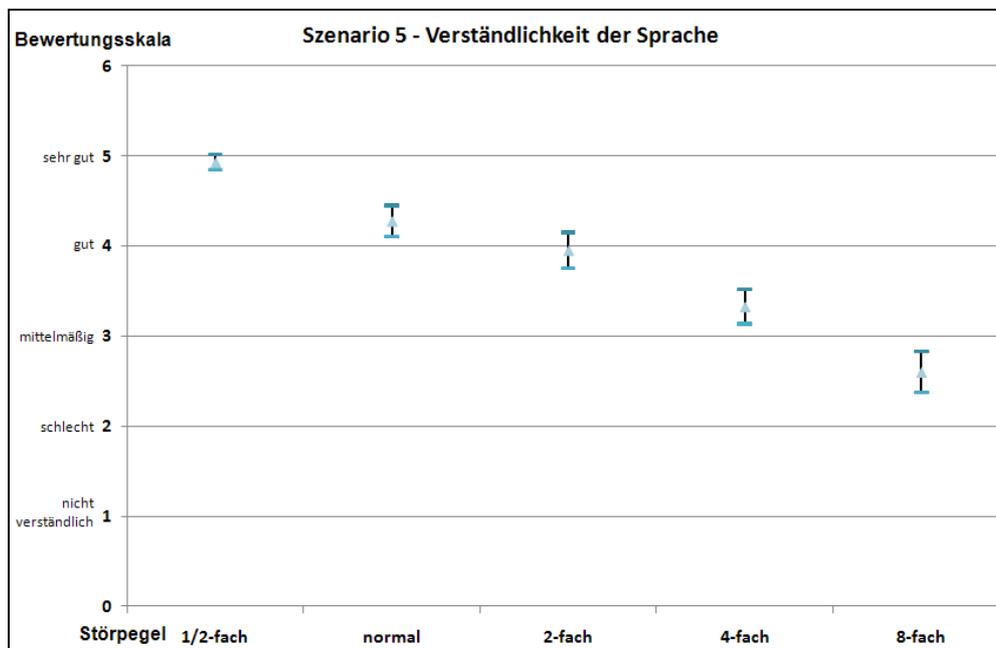


Abbildung 58: Auswertung Szenario 5 – Sprachverständlichkeit

In Abbildung 58 ist erneut zu beobachten, dass die Sprachverständlichkeit relativ linear mit der Erhöhung des Störpegels absinkt. Je größer das Störsignal, desto schlechter ist die Verständlichkeit der Sprache. Trotzdem ist auch hier zu bemerken, dass die Testsätze dennoch von allen Probanden im Durchschnitt mit „mittelmäßig“ als schlechtester Verständlichkeit bewertet wurden.

Wie bei den bereits ausgewerteten Szenarien fällt die Bewertung des Einflusses der Umgebung gegenüber der Sprachverständlichkeit deutlich ab. Dies lässt sich, wie bei

## 8. Auswertung der Ergebnisse

den anderen Umgebungen, damit erklären, dass die Testteilnehmer nun besonders auf die Störung geachtet haben. Trotz des hohen Störsignals über den gesamten Frequenzbereich bewegten sich die halbe und die originale Lautstärke des Rauschens im „nicht störend“ und nur „leicht störend“ Bereich. Eine Verdopplung des Pegels des weißen Rauschens rief aber ein Stören der Umgebung bei den Probanden hervor.

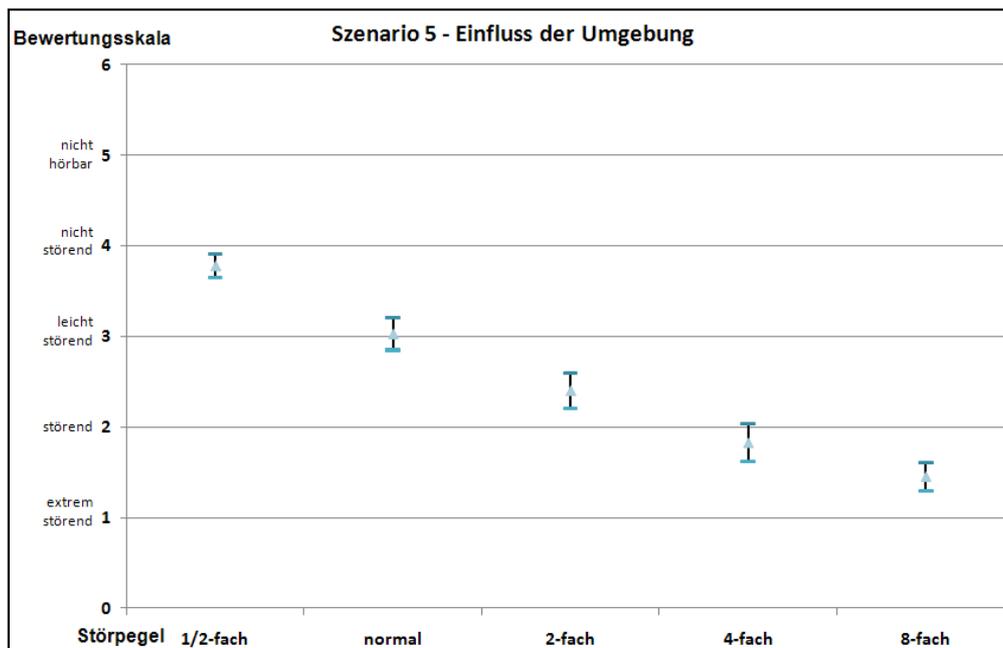


Abbildung 59: Auswertung Szenario 5 - Einfluss der Umgebung

Bei einer Vervielfachung oder sogar Verachtfachung empfanden die Probanden relativ eindeutig die Umgebung als „störend“ und der MOS-Wert von 1,45 entspricht sogar einer extremen Störung. Die Konfidenzintervalle der einzelnen Störpegel zeigen, dass die Testpersonen eine relativ einheitliche Wertung abgegeben und alle das weiße Rauschen für unangenehm und störend für die Verständlichkeit empfunden hatten.

Im Folgenden sollen wiederum die Spektrogramme für die beiden Extremfälle der Lautstärkeerhöhung untersucht werden. Abbildung 60 zeigt die 4-fache Erhöhung.

## 8. Auswertung der Ergebnisse

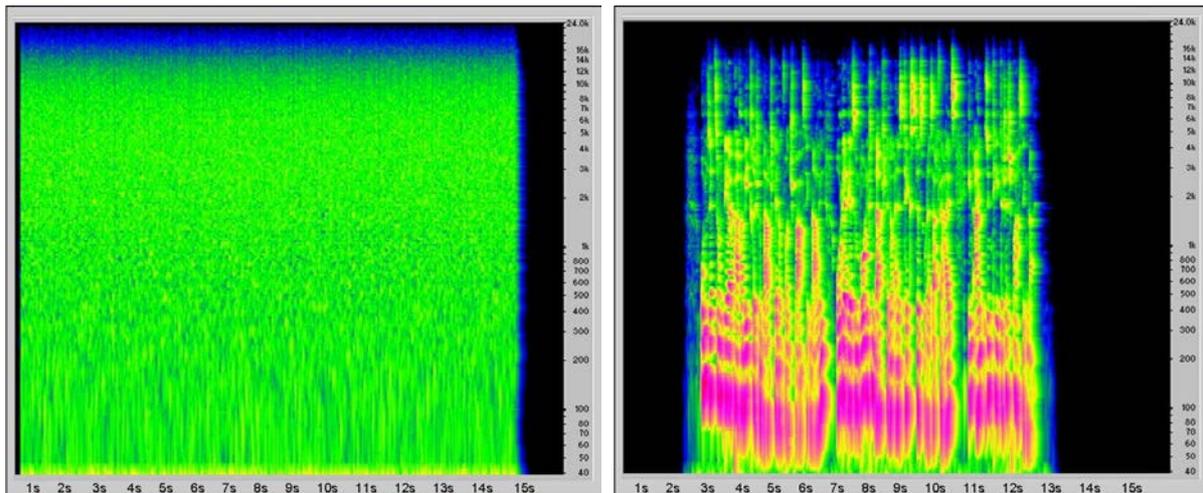


Abbildung 60: Szenario 5 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Bereits beim ersten Blick auf das Spektrogramm des Störpegels und des Sprachbeispiels fällt auf, dass eine Addition der beiden Signale gravierend in die Sprachverständlichkeit eingreifen musste, da das weiße Rauschen das Sprachsignal zum größten Teil verdecken würde.

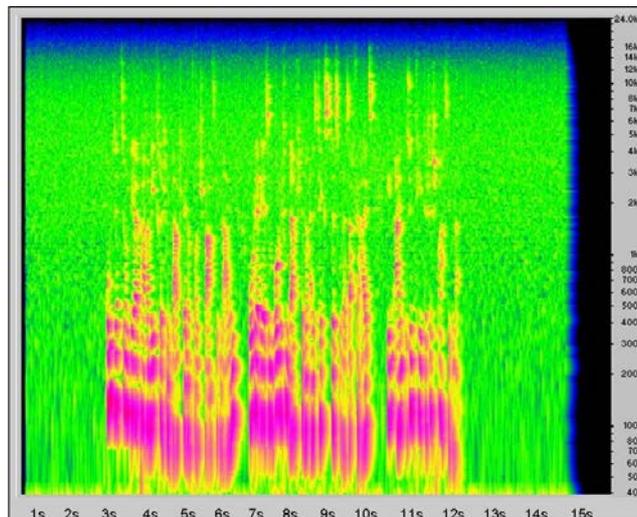


Abbildung 61: Szenario 5 Sprache und 4-facher Störpegel

Das Spektrogramm in Abbildung 61 zeigt diese Addition, und es wird klar, warum die Probanden den Einfluss der Umgebung bzw. die Sprachverständlichkeit bei einer Vervielfachung des Störpegels so negativ bewertet hatten. Nur die tiefen bis mittleren Frequenzbereiche der Stimme sind noch eindeutig zu erkennen. Die hohen Frequenzanteile des Sprachsignals kommen nur noch sehr gering zur Geltung und können nur schlecht im Spektrogramm erkannt werden. Dies erklärt die negative Empfindung der Probanden.

## 8. Auswertung der Ergebnisse

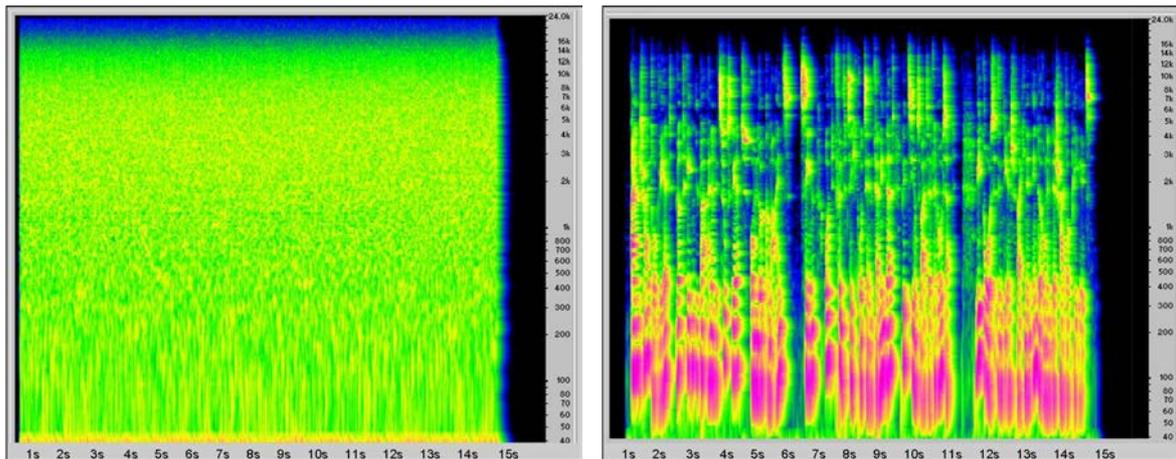


Abbildung 62: Szenario 5 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

Abbildung 62 zeigt links das Spektrogramm des 8-fachen Störpegels. Dem Betrachter fällt sofort anhand der kräftigeren Farben ins Auge, dass sich der Störpegel nochmals erhöht hat. Im Vergleich der beiden Bilder lässt sich bereits vor der Addition sagen, dass bis auf den tiefen Frequenzbereich nur noch wenig von dem Sprachsignal gut hörbar sein würde. In der folgenden Grafik wird die Addition näher betrachtet.

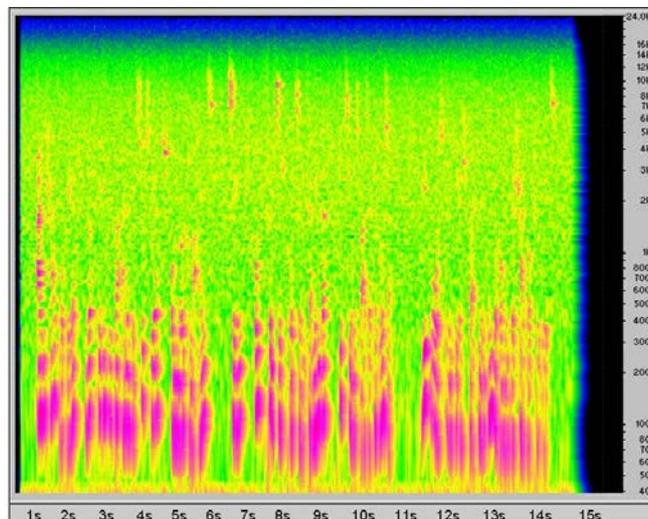


Abbildung 63: Szenario 5 Sprache und 8-facher Störpegel

Anhand dieses Spektrogrammes wird die Bewertung der Testteilnehmer nochmals verdeutlicht. Sie empfanden die Verachtfachung des weißen Rauschens größtenteils als extrem störend und somit sank die Sprachverständlichkeit ebenfalls auf „mittelmäßig“ bis „schlecht“. Unter diesen Bedingungen wäre eine Sprachgütemessung auf keinen Fall sinnvoll. Die Probanden hätten keine aussagefähigen

## 8. Auswertung der Ergebnisse

Ergebnisse liefern können, da die Umgebung eine genaue Bewertung zu stark beeinflussen würde.

Es ist jedoch auch hier wiederum zu bemerken, dass diese Extremfälle bei einer solchen Analyse nicht auftauchen werden und jeder Versuchsleiter versuchen würde dies zu unterbinden. In diesem Szenario wurde durch die Störung über den kompletten Frequenzbereich ein Ausnahmefall geschaffen, der in der Realität nicht bei den Messungen auftreten würde. Dennoch könnte bei einem normalen Pegel des weißen Rauschens auf einem Niveau von Umgebungsgeräuschen ein aussagekräftiges Urteil abgegeben werden, wie das relativ schmale Konfidenzintervall zeigt.

### 8.4.6 Szenario 6 – „weißes Rauschen (MNRU) mit natürlicher Sprache“

In Szenario 6 sollte anhand eines anderen Sprechers mit „alltäglicher/natürlicher“ Sprache die Situation eines Störsignals über den ganzen Frequenzbereich nochmal überprüft werden. Um zu gewährleisten, dass ein ähnliches Ergebnis erreicht wird, wurde ebenfalls ein männlicher Sprecher eingesetzt. Es ist vorab zu bemerken, dass das Ergebnis etwas schlechter ausfallen wird, da die Aufnahmequalität nicht der hochwertigen Sprache der anderen Szenarien entspricht.

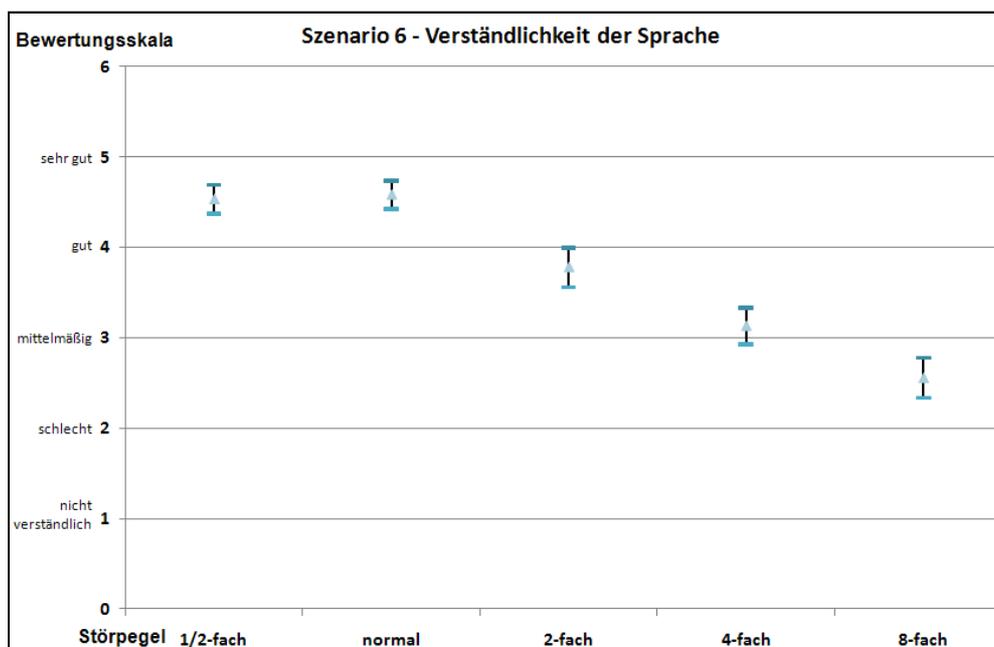


Abbildung 64: Auswertung Szenario 6 – Sprachverständlichkeit

## 8. Auswertung der Ergebnisse

In Vergleich zu Abbildung 58 ist, wie bereits erwartet, eine geringe Verschlechterung der Bewertung zu erkennen. Zum besseren Vergleich soll die folgende Tabelle der MOS-Werte dienen.

		Störeinfluss		Szenario 5	Szenario 6
Durchgang1 Verständlichkeit	1/2-fach	1		4,925	4,525
	normal	2		4,275	4,575
	2-fach	3		3,95	3,775
	4-fach	4		3,325	3,125
	8-fach	5		2,6	2,55

Tabelle 1: Vergleich Sprachverständlichkeit Szenario 5 und 6

Tabelle 1 zeigt, dass das 6. Szenario mit der natürlichen Sprache jeweils mit einer schlechteren Sprachverständlichkeit bewertet wurde. Lediglich beim Normalpegel des weißen Rauschens wurde bei der natürlichen Sprache die Verständlichkeit als etwas besser empfunden. Dies kann jedoch auch am entsprechenden Hörbeispiel und Schwierigkeitsgrad des zu hörenden Satzes gelegen haben. In Szenario 5 wurde der folgende Satz gesprochen. „Beginnen wir mit der Anregung, im Bild links dargestellt. Der aus der Lunge kommende Luftstrom ist für die Sprechmaschine gewissermaßen der Energielieferant.“ Im 6. Szenario hören die Probanden diesen Satz: „Dieses Filter, heute ausschließlich ein digitales Filter, wird durch die Filterkoeffizienten auf eine Resonanzfrequenz eingestellt, die der Resonanz des Mund- bzw. Nasenraumes entspricht.“ Man kann davon ausgehen, dass der Satz des natürlichen Sprechers schwieriger vom Sinn her zu verstehen war. Dies würde den Bewertungsunterschied erklären. Die anderen Hörbeispiele wiesen eine ähnliche Beurteilung auf. Somit sollte für die reine Sprachverständlichkeit der Sprecher keine große Rolle bei der Beurteilung mit einer Störung über den gesamten Frequenzbereich spielen.

## 8. Auswertung der Ergebnisse

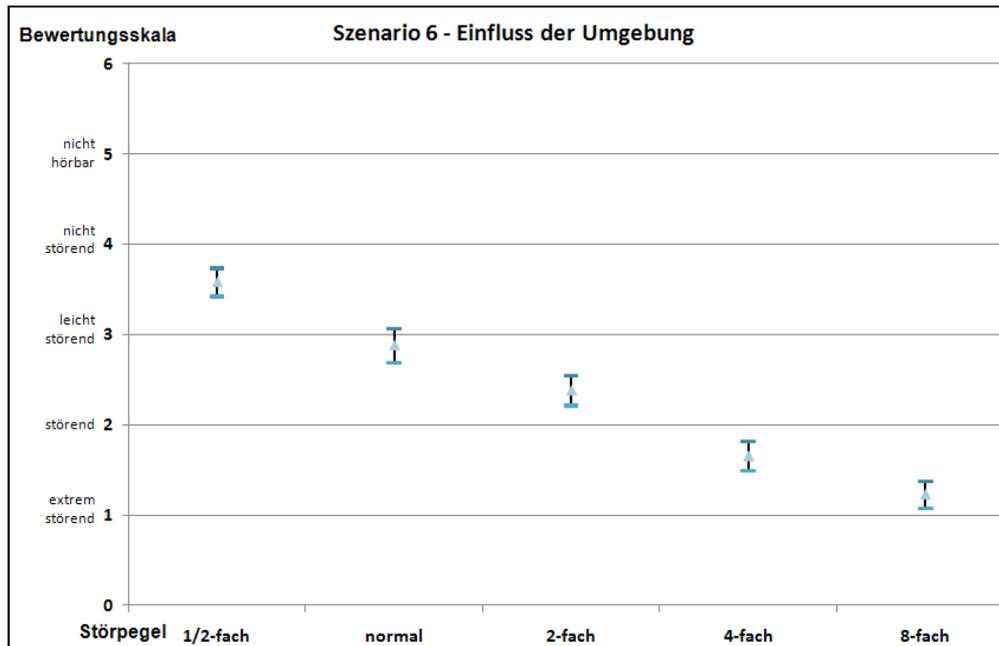


Abbildung 65: Auswertung Szenario 6 - Einfluss der Umgebung

In Abbildung 65 sind die Ergebnisse der MOS-Werte des 2. Durchgangs zusammengestellt. Wie bereits im 5. Szenario ist hier eine Linearität in der Grafik zu sehen. Je größer die Störung, desto störender der Einfluss der Umgebung auf das Sprachsignal. Tabelle 2 lässt erkennen, dass die Bewertung der Testteilnehmer lediglich um 0,2 Punkte im MOS-Wert abweicht.

		Störeinfluss	Szenario 5	Szenario 6
Durchgang2 Umgebung	1/2-fach	1	3,775	3,575
	normal	2	3,025	2,875
	2-fach	3	2,4	2,375
	4-fach	4	1,825	1,65
	8-fach	5	1,45	1,225

Tabelle 2: Vergleich Einfluss der Umgebung Szenario 5 und 6

Diese Abweichung lässt erneut die Schlussfolgerung zu, dass der Sprecher keinen großen Einfluss auf diesen Versuch hatte und das Ergebnis somit allgemein Gültigkeit haben dürfte.

Anhand der Spektrogramme werden im Folgenden die beiden extremen Störungen näher betrachtet.

## 8. Auswertung der Ergebnisse

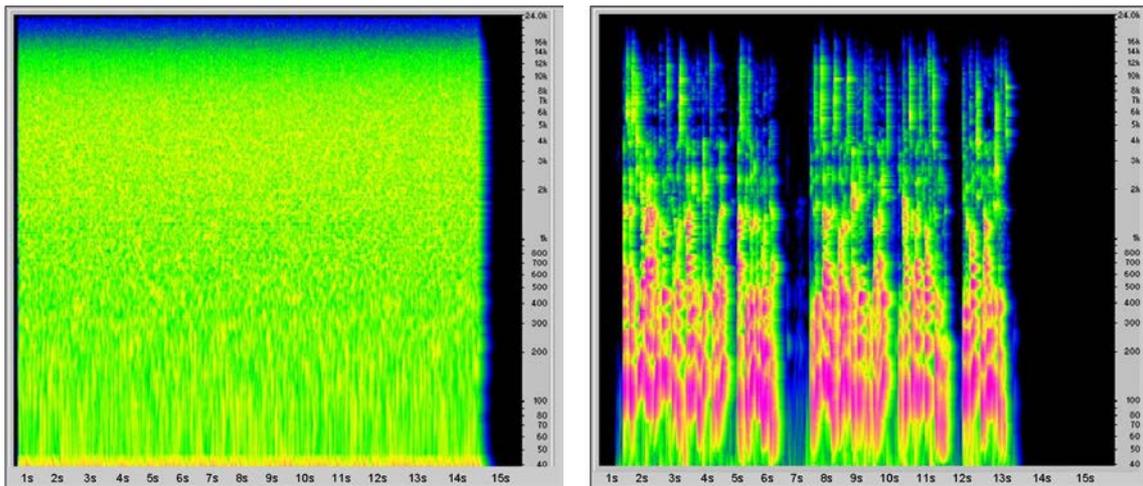


Abbildung 66: Szenario 6 - 4-facher Störpegel (links) und Sprachbeispiel (rechts)

Wie zuvor im 5. Szenario führte die Addition des Sprachsignals mit dem weißen Rauschen zu einer schlechten Verständlichkeit. Anhand der rechten Grafik lässt sich bei genauer Betrachtung erkennen, dass das Sprachsignal im Vergleich zum originalen Sprecher des Versuches nicht ganz so durchsetzungsfähig gegenüber dem Störsignal ist. Dies bestätigte auch die Wertung der Probanden.

Abbildung 67 zeigt die Addition der beiden Signale. Wie bereits im vorhergehenden Szenario lässt sich hier das Sprachsignal nicht mehr eindeutig reproduzieren. Allein der tief- bis mittelfrequente Bereich kann noch wahrgenommen werden.

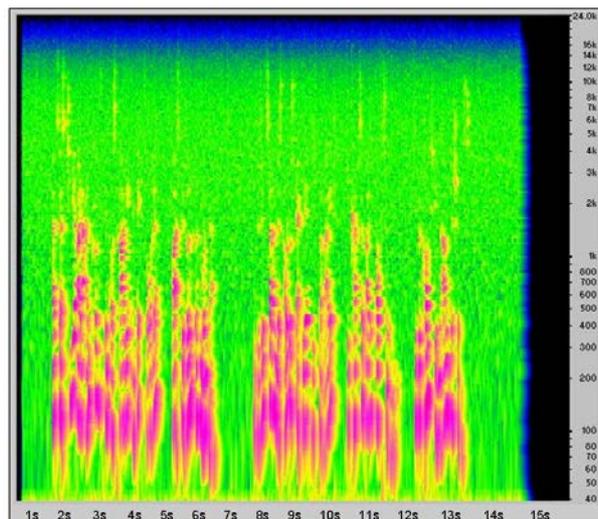


Abbildung 67: Szenario 6 Sprache und 4-facher Störpegel

Im Folgenden wird die 8-fache Erhöhung des Störsignals betrachtet. Die beiden Spektrogramme sollen veranschaulichen, was den Probanden akustisch dargeboten wurde.

## 8. Auswertung der Ergebnisse

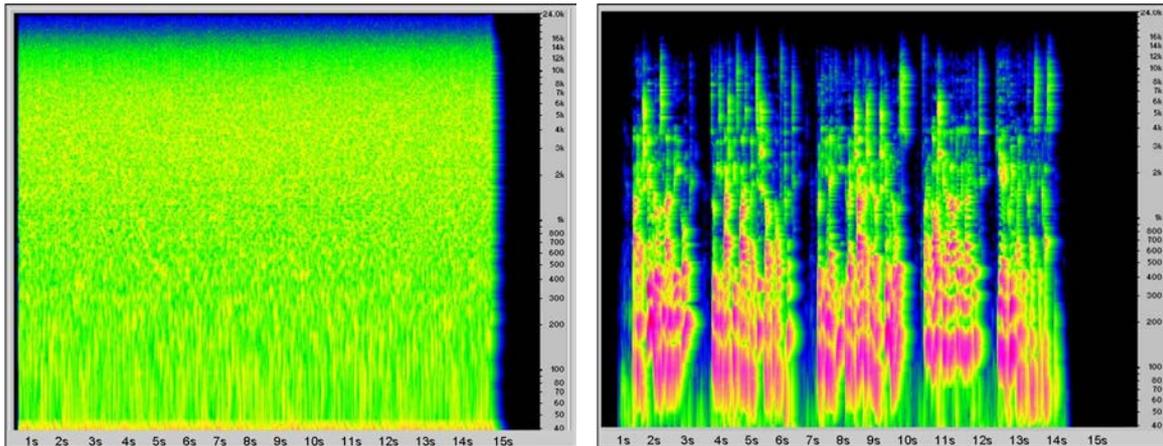


Abbildung 68: Szenario 6 - 8-facher Störpegel (links) und Sprachbeispiel (rechts)

Der Vergleich der beiden Spektrogramme lässt schon erkennen, dass das weiße Rauschen das Sprachbeispiel sehr stark verdecken würde. Ebenfalls ist wiederum bei genauer Betrachtung wahrzunehmen, dass das Sprachsignal nicht ganz so hochwertig aufgelöst ist, wie in den vorangegangenen Szenarien. Besonders gut sieht man dies im hochfrequenten Bereich, da hier die Amplituden nicht durchgängig die gleiche Intensität aufweisen und stark im Pegel schwanken. Dies deutet auf eine ungenaue Aufnahmetechnik ohne entsprechende Nachverarbeitung.

Die Addition bestärkt die Bewertungen der Testpersonen, da das Sprachbeispiel kaum noch von dem weißen Rauschen trennbar ist. Dennoch können spektrale Einzelheiten des Sprachsignals erkannt werden.

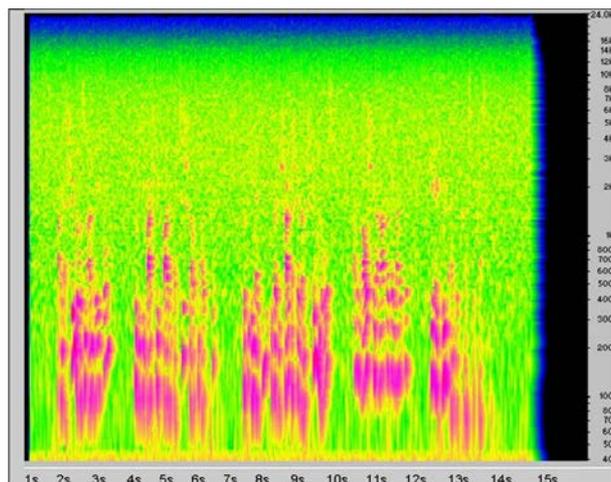


Abbildung 69: Szenario 6 Sprache und 8-facher Störpegel

Das Nutzsinal wird größtenteils vom Rauschen verdeckt und somit ist eine Sprachverständlichkeit nur noch bedingt gegeben und die Probanden empfanden das Rauschsignal als sehr störend.

## 8. Auswertung der Ergebnisse

Zusammenfassend muss dennoch gesagt werden, dass es sich hier um Extremfälle handelt, die so nicht auftreten werden. Man kann sagen, dass die Bewertung nur gering bei einem „natürlichen“ Sprecher abweichen würde und somit kann das Ergebnis von diesem Test auch allgemein für Sprachgütemessungen betrachtet werden.

### 8.4.7 Szenario 7 – „Raum 321 Original“

Szenario 7 fand real in derselben Umgebung wie das virtuelle Szenario 3 statt. Dabei sollten Rückschlüsse zwischen originaler und virtueller Umgebung gezogen werden. So sollte analysiert werden, ob eine gleiche Urteilsabgabe im virtuellen und realen Rahmen stattfinden würde. Im Nachhinein stellte sich dies jedoch als Problem heraus, da – wie bereits vor den Versuchen schon angemerkt wurde – jeder Proband unterschiedliche Störeinflüsse hatte. Durch die Aufnahme der aktuellen Störpegel konnte so für jeden Probanden individuell das Szenario analysiert werden.

Abbildung 70 zeigt das Ergebnis der MOS-Werte aller Probanden, die an dem Test teilnahmen. Anhand der vergrößerten Konfidenzintervalle lassen sich bereits die unterschiedlichen Gegebenheiten vermuten.

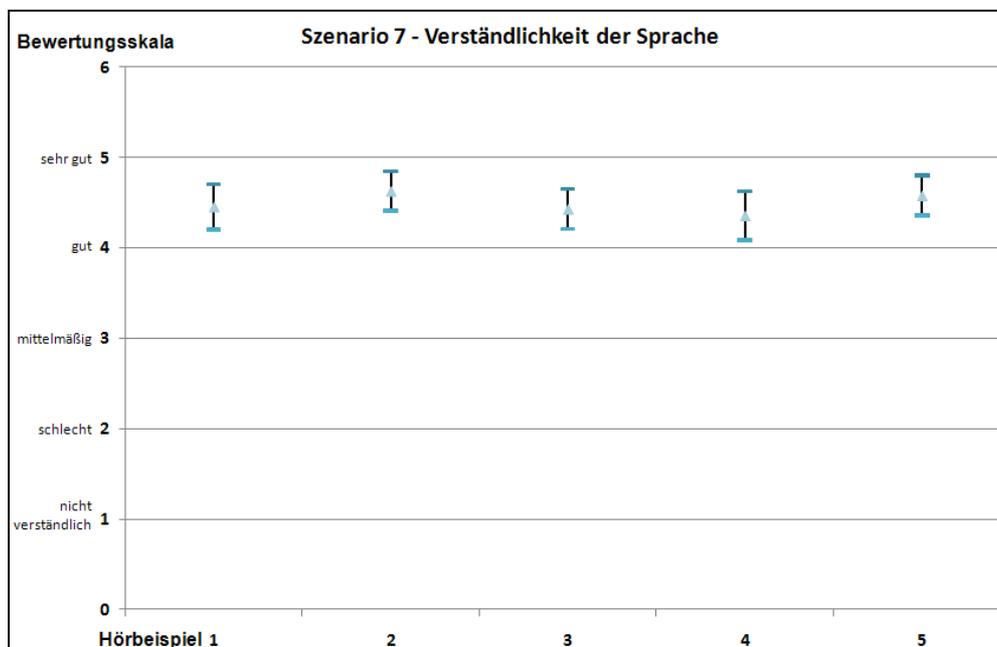


Abbildung 70: Auswertung Szenario 7 – Sprachverständlichkeit

## 8. Auswertung der Ergebnisse

Da die Probanden im 7. Szenario keine aktiven Lautstärkeänderungen den Probanden beeinflusst haben, hätten erwartungsgemäß alle fünf Hörbeispiele gleich bewertet werden müssen. Die Schwankungen in den MOS-Werten sind auf die unterschiedlichen Bedingungen je Beispiel zurückzuführen. Beispielsweise traten bei einem Testsatz mehr Störungen von außen auf und bei einem anderen weniger. Dies würde den nicht ganz konstanten Verlauf erklären. Dennoch bewegen sich die Werte im Mittel zwischen einer „guten“ und „sehr guten“ Sprachverständlichkeit. Dieses Ergebnis stimmt etwa mit dem Wert aus Szenario 3 bei originalem Störpegel überein und das sollte hierbei überprüft werden.

Wie bereits beim ersten Durchgang erwartet, sollten die Bewertungen der Testteilnehmer beim zweiten Durchlauf ebenfalls für alle Hörbeispiele gleich sein. Das Ergebnis hierfür ist in der folgenden Abbildung zu sehen. Im Mittel empfanden die Probanden das originale Szenario als „nicht störend“ bis „leicht störend“. Diese Wertung war jedoch abhängig von der Zeit der Versuchsdurchführung und ob auf der gegenüberliegenden Baustelle ein großer Störpegel produziert oder beispielsweise gerade Pause gemacht wurde. Die unterschiedlichen Gegebenheiten spiegeln sich in den größeren Konfidenzintervallen wider. Die Bewertung der einzelnen Probanden betrug je nach Situation „nicht störend“, „leicht störend“ und „störend“ – s. Anhang.

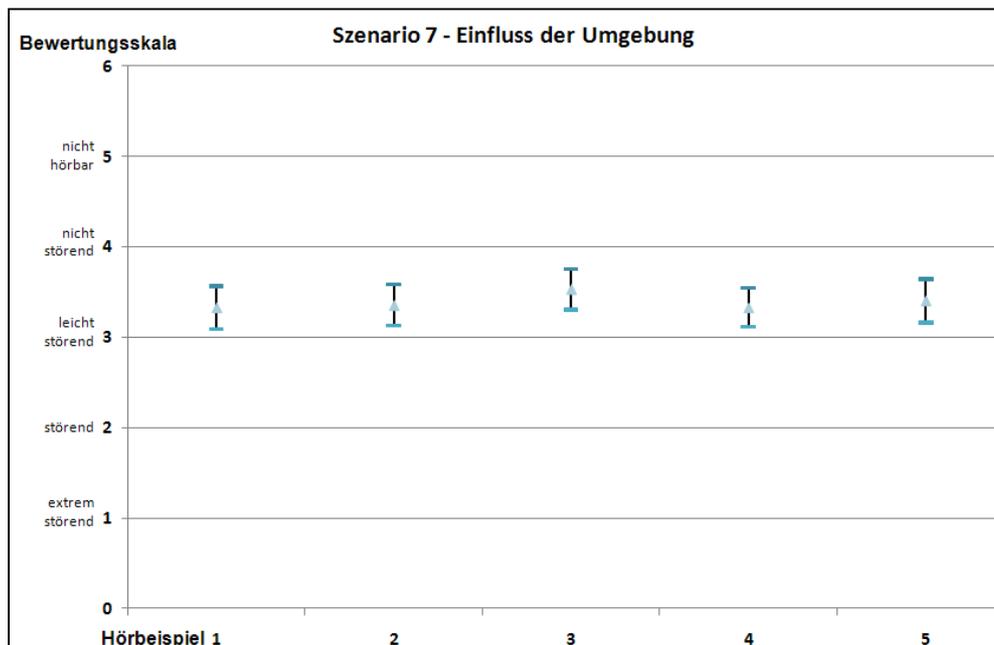


Abbildung 71: Auswertung Szenario 7 - Einfluss der Umgebung

## 8. Auswertung der Ergebnisse

In Tabelle 3 wird ein kurzer beispielhafter Überblick über die Versuchszeiten und die vorherrschenden akustischen Bedingungen am Versuchsort gezeigt. Hiermit sollen die unterschiedlichen Gegebenheiten für die einzelnen Probanden aufgezeigt werden. Der Schalldruckpegel wurde über eine Minute mit dem Messgerät dokumentiert und gemittelt. Somit tragen punktuelle Störpegel nur gering zum gesamten Schallpegel bei.

16:45 Uhr	19:00 Uhr	13:00 Uhr	16:00 Uhr	19:00 Uhr	19:30 Uhr	10:00 Uhr	10:45 Uhr	11:30 Uhr	13:00 Uhr	13:45 Uhr	15:15 Uhr	16:00 Uhr	10:00 Uhr	10:45 Uhr	11:30 Uhr	13:45 Uhr	17:00 Uhr	10:00 Uhr	10:45 Uhr	11:30 Uhr	12:15 Uhr	17:45 Uhr
44,8 dB	39,2 dB	43,7 dB	39,1 dB	40,6 dB	40,6 dB	40,6 dB	49,0 dB	44,5 dB	41,1 dB	44,3 dB	42,4 dB	50,6 dB	44,4 dB	44,4 dB	41,7 dB	42,2 dB	39,2 dB	46,2 dB	48,8 dB	50,4 dB	51,2 dB	39,6 dB

Tabelle 3: Zeit des Versuchsbeginns mit Störpegel

Anhand der Tabelle lässt sich schnell erkennen, dass der geringste Störschallpegel bei 39,2 dB und der größte bei 51,2 dB lag. Es ist nur offensichtlich, dass Probanden mit höherem Störsignal schlechter urteilten. Da dieser Umstand nicht diskutiert werden muss, soll hier auch nicht detailliert auf die unterschiedlichen Bedingungen eingegangen werden.

Zusammenfassend kann man trotzdem von einer guten Variante zur Überprüfung der virtuellen Umgebungen sprechen, da das Szenario trotz nicht konstanter Bedingungen zeigte, dass mit einem einfachen Kopfhörer und einem undefinierten Störpegel ein gutes Ergebnis erzielt werden kann und sich die Störungen der Umgebung in Grenzen halten.

### 8.5 Diskussion der Ergebnisse

Bei genauer Betrachtung der einzelnen Szenarien und des Gesamtergebnisses fällt die durchgängig weitestgehend einheitliche Bewertung der Probanden auf. Diese ist bei der Sprachverständlichkeit eindeutig auf die sehr hochwertigen Sprachbeispiele, deren exakte Aussprache, die hochwertige Aufnahmetechnik und Nachbearbeitung zurückzuführen. Die Störeinflüsse waren definiert mit dem halben, dem originalen, dem doppelten, dem 4-fachen und dem 8-fachen Schalldruckpegel. Anhand dieser definierten Pegeländerung konnte jeder Proband nach seinem Empfinden urteilen und eine Erhöhung um 6 dB ist deutlich merkbar. Desweiteren hatten die Probanden durch die Bewertungsbögen die Möglichkeit, bei einer für sie als falsch empfundenen

## 8. Auswertung der Ergebnisse

Bewertung, vorangegangene Wertungen anzupassen. Von dieser Möglichkeit machten einige Testteilnehmer Gebrauch. Hierauf sind sicher auch die meist linearen Verläufe der Grafiken zurückzuführen.

Die durchweg für „sehr gut“ bis „gut“ empfundene Sprachverständlichkeit ist vermutlich auf den „Cocktailparty-Effekt“ zurückzuführen. Denn es ist dem Menschen möglich bei mehreren unterschiedlichen Signalen das eigentlich wichtige Signal, hier die Sprache, herauszufiltern. Diesen Effekt nutzten die Testteilnehmer ganz automatisch, indem sie sich im ersten Durchgang nur auf die Sprache konzentrierten. Somit wird das Störgeräusch unbewusst abgesenkt, und die Sprache tritt hervor.

Trotzdem wirkten die Sprachbeispiele, nach Aussage einiger Probanden, teilweise etwas befremdlich, was man möglicherweise auf den „Lombard-Effekt“ zurückführen kann. Normalerweise würde ein Sprecher seine Stimme der Umgebung anpassen. Bei einer lauten und störenden Umgebung würde er seine Stimme erheben und sie würde kräftiger und lauter erscheinen. Laut Lombard nimmt im Normalfall der Sprechpegel um jeweils ein halbes Dezibel zu, wenn sich der Störpegel um ein Dezibel erhöht. Dieser Umstand war im vorliegenden Versuch nicht gegeben, da der Sprecher immer mit gleicher Lautstärke, Intonation und Aussprachefrequenz zu den Umgebungsgeräuschen addiert wurde.

Der Effekt der „Verdeckung“, bei dem lautere Schallanteile leisere verdecken oder maskieren, war besonders bei 4-facher und 8-facher Erhöhung des Störschallpegels zu beobachten. Dies konnte besonders gut an den einzelnen Spektrogrammen kenntlich gemacht werden und spiegelte sich in den Probandenbewertungen ebenfalls wider. Tieffrequente Signale sind hierbei weniger von der Verdeckung betroffen, was ebenfalls anhand der Spektrogramme gut dargestellt werden konnte. Jedoch ist trotzdem anzumerken, dass die Hauptinformation der Sprache nicht im tieffrequenten Bereich liegt und somit nicht ganz so wichtig für die Sprachverständlichkeit ist.

Man kann von einem allgemeingültigen Ergebnis ausgehen, da anhand von zwei verschiedenen Sprechern getestet wurde. Beide Sprachsignale wurden mit einem weißen Rauschen addiert. Diese Störung über den kompletten Frequenzbereich kann jegliches Szenario beinhalten. Somit kann allein durch das 5. und 6. Szenario von einer Allgemeingültigkeit der Ergebnisse gesprochen werden.

### 9. Zusammenfassung und Ausblick

Anhand eines Studiotests sollten realistische akustische Umgebungen und ihr Einfluss auf internet-basierte Sprachgütemessungen untersucht werden. Diese Untersuchungen sollten analysieren, inwiefern Umgebungen eine auditive Messung beeinflussen und ob trotzdem ein brauchbares Ergebnis entstehen kann.

Die vorliegende Arbeit bestätigte die Annahme, dass sich akustische Umgebungsgeräusche störend auf eine Sprachgütemessung auswirken können. Jedoch ist dies zu relativieren, denn die Untersuchungen zeigten ebenfalls, dass dieser Einfluss stark von der Intensität des Störpegels abhängig ist.

Anhand von typischen Umgebungen, in denen üblicherweise internet-basierte Sprachgütemessungen stattfinden, wurden 7 unterschiedliche Szenarien entworfen.

Ein typischer Fall war hierbei das Szenario 1, bei dem ein Zimmer an einer befahrenen Straße simuliert wurde. Als Ergebnis kann man sagen, dass sich eine normalbefahrene Straße nicht stark auf die Urteilsabgabe der Probanden auswirken wird.

Desweiteren wurde ein Zimmer mit angrenzenden Küchengeräuschen und einem Fernseher getestet. Trotz des erhöhten Störeinflusses konnte die Sprache gut vom Störgeräusch getrennt werden, und es entstand ein aussagekräftiges Ergebnis.

Der zurzeit herrschende Baustellenlärm auf dem Campus wurde im dritten Szenario überprüft. Die Probanden empfanden hier selbst bei stark erhöhtem Störpegel der Umgebung nur eine leichte Störung.

Ein weiterer typischer Ort für derartige Untersuchungen wurde im vierten Szenario anhand eines virtuellen PC-Pools erzeugt. Die hier auftretenden Lüftergeräusche zeigten, dass sich ein starkes, rauschförmiges Signal negativ auf die Sprachverständlichkeit auswirken kann.

Um den Einfluss einer Störung im kompletten Frequenzbereich zu untersuchen, wurde im Szenario 5 mit der MNRU gearbeitet. Dies ist, vereinfacht gesagt, ein weißes Rauschen, welches zum Nutzsignal addiert wird. Dieser Extremfall zeigt, dass hier durch Verdeckungseffekte eine Beeinträchtigung der Sprachverständlichkeit eintreten kann, - was auch zu erwarten war.

## 9. Zusammenfassung und Ausblick

Das sechste Szenario wurde ebenfalls mit einem weißen Rauschen gestört. Jedoch wurde hier der Sprecher gewechselt, um einen Eindruck zu bekommen, was passiert, wenn sich die Stimmcharakteristik und die Sprachaufnahmequalität ändern. Das Ergebnis sah hier ganz ähnlich wie beim vorhergehenden Szenario aus. Hierdurch wurde gezeigt, dass die Aussagen der anderen Szenarien allgemeingültig sind.

Im letzten Test wurden die Probanden an den originalen Aufnahmeplatz des Szenarios 3 gesetzt, um einen Vergleich zwischen virtueller und reeller Umgebung zu ziehen. Auf Grund der ständig wechselnden akustischen Umgebungsbedingungen konnte hier jedoch keine eindeutige Aussage getroffen werden.

Zusammenfassend lässt sich für alle Szenarien sagen, dass jeweils eine Linearität zwischen dem Störpegel und dem Einfluss der Umgebung vorlag. Je höher der Störpegel, desto störender empfanden die Probanden die Umgebungsgeräusche, und die Sprachverständlichkeit sank ab.

Im vorliegenden Versuch wurden den Probanden unter definierten Bedingungen jeweils der halbe, der originale, der doppelte, der 4-fache und der 8-fache Störpegel einer Umgebung dargeboten. Die Auswertung zeigte, dass die Vervierfachung und die Verachtfachung des Pegels die größten Störungen hervorriefen, und die Verständlichkeit teilweise nur noch bedingt gegeben war. Diese Extremfälle werden bei keiner Untersuchung auftreten, denn jeder Testteilnehmer sollte im Vorfeld einer internet-basierten Sprachgütemessung darüber informiert werden, dass externe Geräuschpegel so gering wie möglich zu halten sind.

Man kann sagen, dass internet-basierte Sprachgütemessungen durchaus ihre Daseinsberechtigung haben. Hierdurch können aussagefähige und vor allem schnelle Ergebnisse mit vielen Probanden in kurzer Zeit erzielt werden. Jedoch gilt auch zu sagen, dass eine Untersuchung unter Studiobedingungen immer bessere Ergebnisse liefern wird, da die Testpersonen keinerlei Störungen ausgesetzt sind und definierte Bedingungen herrschen. Diese sorgen auch dafür, dass der Test eindeutig vergleichbar wird. Bei einer internet-basierten Untersuchung gelten immer wieder andere Randbedingungen, sei es z.B. die Software oder die Hardware und vor allem eben die akustischen Umgebungsbedingungen.

Für spätere Tests mit realistischen akustischen Umgebungen sollte eine zusätzliche Filterung, z.B. die ID-Entzerrung, für die Kunstkopfaufnahmen ermöglicht werden.

## 9. Zusammenfassung und Ausblick

Denn nur durch diese Filterung wird der Höreindruck noch realistischer für die Testteilnehmer. Laut Aussagen der Probanden spielte sich die Umgebung wirklich um sie herum ab, aber vielleicht wäre so eine noch plastischere Wiedergabe möglich.

Bei der vorliegenden Arbeit wurde für die Versuchsdurchführung eine PowerPoint Präsentation genutzt. Im Testverlauf stellte sich heraus, dass dies auf Grund technischer Probleme keine optimale Lösung war. Zukünftige Tests sollten, wie die bereits durchgeführten Tests, wieder anhand einer extra für MOS-Tests entworfenen Software durchgeführt werden. Dies würde den Probanden den Versuch erleichtern, und die abgegebenen Bewertungen könnten sofort statistisch ausgewertet werden.

Weiterhin sollte die Anpassung an die Hörschwelle optimiert werden. Dies könnte beispielsweise durch eine automatisierte Software erfolgen. Eventuell könnte bei der Adaption eine separate Einstellung für das linke und rechte Ohr vorgenommen werden.

Um das Ergebnis noch weiter zu verallgemeinern, wäre es wünschenswert, einen derartigen Versuch mit diversen – auch weiblichen – Sprechern erneut durchzuführen. Hier könnte analysiert werden, wie weibliche Stimmen von entsprechenden Szenarien beeinflusst werden. Diese zusätzliche Untersuchung hätte sicherlich den Rahmen dieser Arbeit gesprengt und bietet somit die Möglichkeit für weitere Evaluationen.

Da es zu dem vorliegenden Versuch keine reale Überprüfung gibt, wäre eine weitere Arbeit zu diesem Thema wünschenswert. Hier sollte der Test tatsächlich über das Internet stattfinden und die Probanden vielleicht sogar gebeten werden, entsprechende originale Szenarien zu erzeugen. Da heutzutage fast jeder einen Laptop besitzt, wäre ein Wechsel in verschiedene Räume während des Versuchsablaufes gut vorstellbar. Das Ergebnis könnte Aufschluss über den hier analysierten Studioversuch geben. Somit könnte die Anwendbarkeit und Aussagekraft einer internet-basierten Sprachgütemessung bestärkt werden.

Ein Vergleich mit der Masterarbeit von Lutz Casper aus dem Jahr 2005 konnte auf Grund der unterschiedlichen Sprachproben in synthetischer Sprache (Stand 2001) und der hochwertigen Sprache eines ausgebildeten Nachrichtensprechers im vorliegenden Versuch nicht gezogen werden. Weitere Arbeiten in dem Themengebiet wären sehr wünschenswert.

## Literaturverzeichnis

- [AEU11] AEU Acoustics AG – Homepage, URL: <http://www.aeuacoustics.com>, Recherche: 12.07.2011
- [Ahn81] Ahnert, Wolfgang; Reichardt, Walter: Grundlagen der Beschallungstechnik, S. Hirzel Verlag, Stuttgart, 1981
- [Ber98] Berger, J.: Instrumentelle Verfahren zur Sprachqualitätseinschätzung. Arbeiten über digitale Signalverarbeitung, Shaker Verlag, 1998
- [Bor93] Dr. Jürgen Bortz: Statistik für Sozialwissenschaftler; Institut für Psychologie, TU Berlin, 1993
- [Cas06] Casper, Lutz: Internet-basierte Sprachgütemessungen, Masterarbeit, Cottbus, 2006
- [Deg11] Deutsche Gesellschaft für Akustik – Homepage, URL: <http://www.dega-akustik.de>, Recherche: 12.07.2011
- [Dic03] Dickreiter, Michael: Mikrofon- Aufnahmetechnik, 3. Auflage, S. Hirzel Verlag, 2003
- [ESS01] Elektronische Sprachsignalverarbeitung, 12. Konferenz, Bonn 2001
- [Fel84] Fellbaum, Klaus: Sprachverarbeitung und Sprachübertragung, Springer Verlag Berlin, 1984
- [Fel00] Fellbaum, Klaus: Elektronische Sprachsignalverarbeitung, Tagungsband der 11. Konferenz, w.e.b. Universitätsverlag & Buchhandel Eckhard Richter & Co OHG, Dresden, 2000
- [Fel03] Fellbaum, Klaus; Höpfner, Dirk; Richter, Christian: Auditive Verfahren zur Sprachqualitätsmessung – Übersicht, Gestaltung und Anwendungen, Festschrift, Cottbus, 2003
- [Gör94] Görne, Thomas: Mikrofone in Theorie und Praxis, Elektor-Verlag GmbH, Aachen, 1994

- [Gör07] Görne, Thomas: Mikrofone in Theorie und Praxis – 8. überarbeitete und erweiterte Auflage, Elektor-Verlag GmbH, Aachen, 2007
- [Hea08] Head acoustics – Homepage, URL: [http://www.head-acoustics.de/downloads/de/application\\_notes/Richtig\\_Messen\\_und\\_Wiedergeben\\_02\\_08.pdf](http://www.head-acoustics.de/downloads/de/application_notes/Richtig_Messen_und_Wiedergeben_02_08.pdf) , Recherche: 20.05.2011
- [Hea11] Head acoustics – Homepage, URL: [http://www.head-acoustics.de/downloads/de/application\\_notes/Entzerrung\\_01\\_11d.pdf](http://www.head-acoustics.de/downloads/de/application_notes/Entzerrung_01_11d.pdf) , Recherche: 20.05.2011
- [Irm98] Irmer, Ralf: Aufbau einer Softwareumgebung für Reimtests, Studienarbeit, TU Dresden, 1998
- [ITU800] ITU-T Recommendation P.800: Methods for subjective determination of transmission quality, International Telecommunication Union, 1996
- [ITU810] ITU-T Recommendation P.810: Modulated Noise Reference Unit (MNRU), International Telecommunication Union, 1996
- [Kla97] Klaus, H.; Fellbaum, K.; Sotscheck, J.: Auditive Bestimmung und Vergleich der Sprachqualität von Sprachsynthesystemen für die deutsche Sprache, Acustica 83, 1997
- [Ler09] Lerch, Reinhard; Sessler, Gerhard; Wolf, Dietrich: Technische Akustik, Grundlagen und Anwendungen, Springer Verlag, Berlin, 2009
- [Lüd01] Lüdtker, Tobias: Messplatz für auditive Beurteilungsverfahren von Sprachqualität. Studienarbeit, Cottbus, 2001
- [Neu11] Neumann-Homepage: URL: [http://www.neumann.com/?lang=de&id=current\\_microphones&cid=ku100\\_description](http://www.neumann.com/?lang=de&id=current_microphones&cid=ku100_description) , Recherche: 07.05.2011
- [Mey79] Meyer, Erwin; Neumann, Ernst-Georg: Physikalische und Technische Akustik, 3. Auflage, Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig, 1979

- [Ric07] Richter, Christian: Internetbasierte auditive Testverfahren. In: Hentschel, C. (Hrsg.): Sprachsignalverarbeitung – Analyse und Anwendungen. Studentexte zur Sprachkommunikation, Band 44, TUDpress Verlag, Dresden, 2007, S. 113-120
- [Tei05] Teich, Sven: Erstellung eines Audiotools zur Durchführung von internetbasierten MOS-Tests. Studienarbeit, Cottbus, 2005
- [Ton02] Tontch, Kilian: Instrumentelle und auditive Beurteilung der Qualität des Sprachsignals nach räumlicher Filterung des Sprechschalls in Personenkraftwagen, Dissertation, Shaker Verlag Aachen, 2002
- [Vei05] Veit, Ivar: Technische Akustik, 6. erweiterte Auflage, Vogel Buchverlag, 2005
- [Zoo11] Zoom-Homepage: URL: <http://www.zoom.co.jp/products/h2>, Recherche: 08.05.2011

## Anhang

Schallabsorptionsgrad Materialien:

Material der Raumbegrenzungsfläche	Oktavband-Mittenfrequenz $f_m$ [Hz]					
	125	250	500	1000	2000	4000
	<b>Schallabsorptionsgrad <math>\alpha</math></b>					
Absorber <sup>1)</sup>	0.30	0.69	1.01	0.81	0.66	0.62
Akustikplatten, Flumroc	0.26	0.88	0.99	0.91	1.04	1.17
Akustikplatten, Heraklit	0.13	0.11	0.22	0.54	0.85	0.71
Beton, Stuckgips, Naturstein	0.02	0.02	0.03	0.04	0.05	0.05
Bühnenöffnung	0.40	0.40	0.60	0.70	0.80	0.80
Dielen, Parkett, Holzboden (hohlliegend, auf Leisten)	0.10	0.08	0.06	0.05	0.05	0.05
Dielen, Parkett, Spanplatten (festaufliegend)	0.03	0.04	0.04	0.05	0.05	0.05
Fenster, Spiegel	0.12	0.10	0.05	0.04	0.02	0.02
Gestühl, gepolstert unbesetzt	0.45	0.60	0.70	0.80	0.80	0.80
Gestühl, Holz ungepolstert, unbesetzt	0.05	0.05	0.05	0.05	0.05	0.05
Kalkzementputz, Tapete, Gipskartonplatten	0.02	0.03	0.04	0.05	0.06	0.08
Linoleum, PVC-Belag, Gummi	0.02	0.03	0.03	0.04	0.04	0.05
Linoleum, PVC-Belag, Gummi auf Filzschicht	0.02	0.05	0.10	0.15	0.07	0.05
Marmor, Fliesen, Klinker	0.01	0.01	0.02	0.02	0.02	0.03
Mauerwerk, glatte Oberfläche	0.02	0.03	0.03	0.04	0.05	0.07
Mauerwerk, gerippt 10 mm	0.08	0.09	0.12	0.16	0.22	0.24
Metaldecke, GEMA	0.34	0.76	0.81	0.91	0.97	0.90
Pavatex	0.37	0.30	0.48	0.71	0.88	0.89
Publikum auf Holzgestühl	0.40	0.60	0.75	0.80	0.85	0.80
Publikum auf Polstergestühl	0.60	0.75	0.80	0.85	0.90	0.85
Schaumstofftapete, etwa 8 mm dick	0.03	0.10	0.25	0.40	0.50	0.60
Spannteppich (PVC-Folie auf 5 mm Filz)	0.02	0.09	0.20	0.15	0.07	0.05
Teppichboden bis etwa 5 mm Dicke	0.03	0.04	0.06	0.20	0.30	0.40
Teppichboden bis mehr als 5 mm Dicke	0.03	0.06	0.10	0.30	0.50	0.60
Tür (Holz, lackiert)	0.10	0.08	0.06	0.05	0.05	0.05
Vorhänge, Plüsch gerüscht	0.15	0.45	0.96	0.91	1.06	1.02

1) Gipskarton-Lochplatte (Lochanteil 19.6%), mit Faservlies hinterlegt, 10 cm vor massiver Wand, mit 3cm Mineralwolleinlage

Quelle:

[http://www.bph.hbt.arch.ethz.ch/Filep/Schall/Schall\\_Grundl/Kennwerte/Schallabsorptionsgrad.html](http://www.bph.hbt.arch.ethz.ch/Filep/Schall/Schall_Grundl/Kennwerte/Schallabsorptionsgrad.html)

## Auswertung – MOS-Wert, Standardabweichung, Konfidenzintervall

		Variation	Standard-		Konfidenz-			
		Störeinfluss	MOS-Wert	abweichung	intervall	Höchstwert	Tiefstwert	Mitte
Szenario 1	Durchgang1	1/2-fach	4,925	0,267	0,083	5,008	4,842	4,925
	Verständlichkeit	normal	4,775	0,480	0,149	4,924	4,626	4,775
		2-fach	4,525	0,640	0,198	4,723	4,327	4,525
		4-fach	4,55	0,639	0,198	4,748	4,352	4,550
		8-fach	3,775	0,862	0,267	4,042	3,508	3,775
	Durchgang2	1/2-fach	4,05	0,450	0,139	4,189	3,911	4,050
	Umgebung	normal	3,65	0,662	0,205	3,855	3,445	3,650
		2-fach	3,15	0,533	0,165	3,315	2,985	3,150
		4-fach	2,25	0,670	0,208	2,458	2,042	2,250
		8-fach	2,1	0,672	0,208	2,308	1,892	2,100
Szenario 2	Durchgang1	1/2-fach	4,6	0,632	0,196	4,796	4,404	4,600
	Verständlichkeit	normal	4,325	0,764	0,237	4,562	4,088	4,325
		2-fach	3,9	0,778	0,241	4,141	3,659	3,900
		4-fach	3,65	0,893	0,277	3,927	3,373	3,650
		8-fach	2,675	0,888	0,275	2,950	2,400	2,675
	Durchgang2	1/2-fach	3,75	0,494	0,153	3,903	3,597	3,750
	Umgebung	normal	3,125	0,563	0,175	3,300	2,950	3,125
		2-fach	2,725	0,751	0,233	2,958	2,492	2,725
		4-fach	2,175	0,549	0,170	2,345	2,005	2,175
		8-fach	1,675	0,526	0,163	1,838	1,512	1,675
Szenario 3	Durchgang1	1/2-fach	4,9	0,304	0,094	4,994	4,806	4,900
	Verständlichkeit	normal	4,625	0,490	0,152	4,777	4,473	4,625
		2-fach	4,375	0,628	0,195	4,570	4,180	4,375
		4-fach	4,15	0,622	0,193	4,343	3,957	4,150
		8-fach	3,7	0,823	0,255	3,955	3,445	3,700
	Durchgang2	1/2-fach	4,275	0,640	0,198	4,473	4,077	4,275
	Umgebung	normal	3,45	0,677	0,210	3,660	3,240	3,450
		2-fach	2,975	0,620	0,192	3,167	2,783	2,975
		4-fach	2,4	0,778	0,241	2,641	2,159	2,400
		8-fach	1,75	0,630	0,195	1,945	1,555	1,750
Szenario 4	Durchgang1	1/2-fach	4,925	0,267	0,083	5,008	4,842	4,925
	Verständlichkeit	normal	4,85	0,362	0,112	4,962	4,738	4,850
		2-fach	4,775	0,423	0,131	4,906	4,644	4,775
		4-fach	4,625	0,586	0,181	4,806	4,444	4,625
		8-fach	3,6	0,744	0,231	3,831	3,369	3,600
	Durchgang2	1/2-fach	4,6	0,545	0,169	4,769	4,431	4,600
	Umgebung	normal	4,475	0,506	0,157	4,632	4,318	4,475
		2-fach	3,95	0,597	0,185	4,135	3,765	3,950
		4-fach	3,725	0,506	0,157	3,882	3,568	3,725
		8-fach	1,8	0,516	0,160	1,960	1,640	1,800

		Variation	Standard-		Konfidenz-				
		Störeinfluss	MOS-Wert	abweichung	intervall	höchstwert	Tiefstwert	Mitte	
Szenario 5	Durchgang1	1/2-fach	4,925	0,267	0,083	5,008	4,842	4,925	
	Verständlichkeit	normal	4,275	0,554	0,172	4,447	4,103	4,275	
		2-fach	3,95	0,639	0,198	4,148	3,752	3,950	
		4-fach	3,325	0,616	0,191	3,516	3,134	3,325	
		8-fach	2,6	0,744	0,231	2,831	2,369	2,600	
		Durchgang2	1/2-fach	3,775	0,423	0,131	3,906	3,644	3,775
	Umgebung	normal	3,025	0,577	0,179	3,204	2,846	3,025	
		2-fach	2,4	0,632	0,196	2,596	2,204	2,400	
		4-fach	1,825	0,675	0,209	2,034	1,616	1,825	
		8-fach	1,45	0,504	0,156	1,606	1,294	1,450	
		Szenario 6	Durchgang1	1/2-fach	4,525	0,506	0,157	4,682	4,368
	Verständlichkeit		normal	4,575	0,501	0,155	4,730	4,420	4,575
2-fach			3,775	0,698	0,216	3,991	3,559	3,775	
4-fach			3,125	0,648	0,201	3,326	2,924	3,125	
8-fach			2,55	0,714	0,221	2,771	2,329	2,550	
Durchgang2			1/2-fach	3,575	0,501	0,155	3,730	3,420	3,575
Umgebung	normal		2,875	0,607	0,188	3,063	2,687	2,875	
	2-fach		2,375	0,540	0,167	2,542	2,208	2,375	
	4-fach		1,65	0,533	0,165	1,815	1,485	1,650	
	8-fach		1,225	0,480	0,149	1,374	1,076	1,225	
	Szenario 7		Durchgang1	normal	4,45	0,815	0,253	4,703	4,197
Verständlichkeit			normal	4,625	0,705	0,218	4,843	4,407	4,625
		normal	4,425	0,712	0,221	4,646	4,204	4,425	
		normal	4,35	0,864	0,268	4,618	4,082	4,350	
		normal	4,575	0,712	0,221	4,796	4,354	4,575	
		Durchgang2	normal	3,325	0,764	0,237	3,562	3,088	3,325
Umgebung		normal	3,35	0,736	0,228	3,578	3,122	3,350	
		normal	3,525	0,716	0,222	3,747	3,303	3,525	
		normal	3,325	0,694	0,215	3,540	3,110	3,325	
		normal	3,4	0,778	0,241	3,641	3,159	3,400	

# **Eidesstattliche Erklärung und Einverständnis zur Veröffentlichung**

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Mit meiner Unterschrift stimme ich einer Veröffentlichung der Arbeit oder Teilen aus der Arbeit zu.

Cottbus, September 2011

Tobias Holfeld