Technische Universität Berlin
Fakultät I

Fachgebiet Audiokommunikation

Untersuchungen zur Sprachverständlichkeit in Bezug auf sprecherspezifische Eigenschaften und semantischer Vorhersagbarkeit

Masterarbeit

vorgelegt von

Dina Reusch
[redacted]

Erstgutachter: Prof. Dr. Stefan Weinzierl
Zweitgutachter: Dr. Alexander Raake

20. Oktober 2009

Die selbstständige Anfertigung dieser Arbeit versichere ich an Eidesstatt.

Berlin, den 20. Oktober 2009, _____

Dina Reusch

Mein Dank gilt an dieser Stelle Herrn Dr. A. Raake der Telekom Laboratories Berlin für die Betreuung und den daraus resultierenden zahlreichen Diskussionen, Ratschläge und fachliche Hilfestellungen.

Außerdem bedanke ich mich bei Herrn Prof. Dr. S. Weinzierl vom Fachgebiet Audiokommunikation der Technischen Universität Berlin für die Betreuung und Unterstützung während dieser Masterarbeit.

Für die grammatikalische und syntaktische Durchsicht bedanke ich mich herzlich bei meinem Freund Steffen Sauer. Ebenso möchte ich mich sehr bei meiner Familie für deren Unterstützungen in jeglicher Hinsicht während der Dauer der Masterarbeit und des Studiums bedanken.

Inhaltsverzeichnis

1 Einleitung und Fragestellung **1**
1.1 Motivation . . . 1
1.2 Stand der Forschung . . . 2
1.3 Ziel der Arbeit . . . 4

2 Grundlagen **6**
2.1 Auditory Scene Analysis (Sprachverständlichkeit und Cocktailparty-Effekt) . . . 6
2.2 Sprechereigenschaften . . . 8
2.2.1 Spracherzeugung . . . 8
2.2.2 Sprecherunterschiede . . . 9
2.3 Auditory Streaming . . . 11
2.4 Maskierung . . . 13
2.5 Top-Dow und Bottom-Up Prozesse . . . 14
2.6 Kontext . . . 16

3 Vorarbeit und Methode **17**
3.1 Methode der Sprachverständlichkeitsmessung . . . 17
3.1.1 Testmaterial . . . 17
3.1.2 Testmethode . . . 20
3.2 Sprachmaterial . . . 23
3.2.1 Wortmaterial . . . 23
3.2.2 Generierung des Wortkorpus . . . 24
3.2.3 Generierung der Testlisten . . . 25
3.2.4 Sprachaufnahmen der generierten Sätze und Normierung . 26
3.3 Testaufbau und allgemeiner Ablauf . . . 30

4 Praktischer Teil **33**
4.1 Test zur Ermittlung des SRT_mask . . . 33
4.1.1 Testprinzip . . . 33
4.1.2 Testablauf . . . 35
4.1.3 Ergebnisse . . . 37
4.2 Test zur Ermittlung der Sprachverständlichkeit mit vorgegebenen Pegeln . . . 43
4.2.1 Testprinzip . . . 43
4.2.2 Testablauf . . . 45
4.2.3 Ergebnisse . . . 46
4.3 Test zur Ermittlung des Einflusses der sinnhaften Verknüpfung . . 52

4.3.1 Testprinzip 52
4.3.2 Testablauf 55
4.3.3 Ergebnisse 57

5 Diskussion **60**
5.1 Einfluss der Maskierung 60
5.1.1 Informational Masking 63
5.1.2 Energetic Masking 65
5.2 Einfluss durch geänderte Sprechereigenschaften 67
5.2.1 Unterschiede zwischen den Testpersonen 69
5.3 Einfluss durch Kontext 71
5.3.1 Syntax 71
5.3.2 Semantik 75

6 Zusammenfassung und Ausblick **78**

Literaturverzeichnis **80**

A Anhang **83**
A.1 Wortliste aussortierter Wörter 83
A.2 Generierte und im Test verwendete Sätze (SUS) 83
A.3 Programmcode 89
A.4 Information zur Vorbereitung der Testpersonen 90
A.4.1 Anleitung: erster Test zur SRT-Ermittlung 90
A.4.2 Anleitung: zweiter Test zur Ermittlung der Sprachverständlichkeit mit festen Pegeln 94
A.4.3 Anleitung: dritter Test zusätzlich mit sinnbehafteten Satzpaarkombinationen 98

Abbildungsverzeichnis

2.1 Kehlkopf eines Menschen 9
2.2 Spektrogramm für Vokale 'i','u','a' 10

3.1 Beispiele für die gewählten 4 Satzstrukturen des SUS-Textmaterials nach Benoit et al. 1996 für die deutsche Sprache angepasst 19
3.2 Phonemverteilung für alle Schlüsselwörter gemittelt (rot durchgezogene Linie) und für die deutsche Sprache nach Helmut Meier (blaue Balken) 20
3.3 Aufnahmesituation im RAR TU Berlin 28

4.1 gemittelte SRT-Werte über alle Testpersonen für alle Maskierertypen 38
4.2 gemittelte Sprachverständlichkeit über alle Testpersonen für jeden Maskierertyp 38
4.3 gemittelte Sprachverständlichkeit für alle Testpersonen für Maskierertyp gleiche Stimme und um 3 Halbtöne erhöht 39
4.4 adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und den dafür errechneten Pegel (rechte y-Achse in grün) für den Ströspreacher der gleichen Stimme 40
4.5 adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und des dafür errechneten Pegels (rechte y-Achse in grün) für den Störsprecher der gleichen Stimme um 3 Halbtöne erhöht 40
4.6 adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und dem dafür errechneten Pegel (rechte y-Achse in grün) für den Maskierertyp gleiche Stimme der 3.Testperson . . 41
4.7 Differenz der Start- und Endwerte der berechneten Pegel 42
4.8 Minimierung der Abweichung zu 50 % -Sprachverständlichkeit mit den 5 letzten Werten pro Testperson pro Störsprecher der gleichen Stimme und um 3 Halbtöne erhöht 43
4.9 Sprachverständlichkeit in% für den Maskierertyp Rauschen 47
4.10 Sprachverständlichkeit in% für den Maskierertyp weibliche Stimme 47
4.11 Sprachverständlichkeit in % für den Maskierertyp um 6 Halbtöne erhöhte Stimme 48
4.12 Sprachverständlichkeit in% für den Maskierertyp um 3 Halbtöne erhöhte Stimme 48
4.13 Sprachverständlichkeit in% für den Maskierertyp gleiche Stimme . 49

4.14 gemittelte Sprachverständlichkeit in% über alle Testpersonen und Pegel für die einzelnen Maskierertypen entsprechend der Darbietungsreihenfolge 51
4.15 Verläufe der Sprachverständlichkeit über die Pegeländerung für die Störsprecher der gleichen Stimme (schwarze Sterne) und um 3 Halbtöne erhöht (blaue Dreiecke) aller Testpersonen (einfach numeriert) 52
4.16 Sprachverständlichkeit in % für beide Satzpaarformen (sinnhaft verknüpft und nicht sinnhaft verknüpft) für den Maskierertyp gleiche Stimme 58
4.17 Sprachverständlichkeit in% für beide Satzpaarformen (sinhaft verknüpft und nicht sinnhaft verknüpft) für den Maskierertyp um 3 Halbtöne erhöhte Stimme 58
4.18 Sprachverständlichkeit in% für den Maskierertyp Rauschen 59
4.19 gemittelte Sprachverständlichkeit in% über alle Testpersonen und Pegel für die Darbietungsreihenfolge aufgetragen 59

5.1 Analyse der Antworten für IM als Funktion der Pegeldifferenz für gleiche Stimme als Maskierer 65
5.2 Auftretenshäufigkeit der Maskiererfehler als Funktion der Pegeldifferenz für gleiche Stimme als Maskierer 65
5.3 Analyse der Eingaben für EM als Funktion der Pegeldifferenz zum SRT_mask für Rauschen und gleiche Stimme als Maskierer 66
5.4 SRT_mask für alle Maskierertypen 68
5.5 gemittelte Auftretenshäufigkeit der Fehler durch IM für alle Störsprecher 69
5.6 gemittelte Auftretenshäufigkeit der Wörter des Ziel- und Störsprechers gleicher Stimme 71
5.7 Abweichung der gemittelten Sprachverständlichkeit zwischen den einzelnen Maskierertypen für alle Testpersonen 72
5.8 Beispielsätze für alle 4 verwendete Satzstrukturen 73
5.9 gemittelte Sprachverständlichkeit je Satzstruktur für jede Position der Schlüsselwörter im Satz 74
5.10 gemittelte Auftretenshäufigkeit der Wörter des Maskierertyps gleiche Stimme für die Positionen im Satz 76

Tabellenverzeichnis

2.1 Vokal-Formant-Zentren 11

4.1 zwei 5x5-Matrizen zur Gleichverteilung der 5 Maskierertypen mit einfacherer Benennung: M1, M2, M3, M4, M5 35
4.2 Übersicht der ermittelten SRT_mask -Werte für alle 5 Maskierertypen 37
4.3 Übersicht der neu ermittelten SRT_mask -Werte für alle Maskierertypen 43
4.4 Zuordnung der Maskierertypen zur Gleichverteilung 44
4.5 3 mögliche Kombinationsformen der Eingabe mit „A“ als Schlüsselwort des Zielsprechersatzes und „B“ als Wort des Störsprechersatzes . . 53
4.6 Reihenfolge der Darbietung der nterschiedlichen Konditionen: Maskierertypen und Satzpaarkombinationen als Beispiel für die erste Testperson 56

5.1 Definition der Kombinationsmöglichkeiten der Eingabeform mit X = fehlerhaftes Wort, M = Maskiererwort, Z = Zielsprecherwort . 62
5.2 erweiterte Definition der Kombinationsmöglichkeiten der Eingabeform mit X = fehlerhaftes Wort, M = Maskiererwort, Z = Zielsprecherwort 63

1 Einleitung und Fragestellung

1.1 Motivation

Das Hauptziel von Sprache ist die Kommunikation. Der Sender von Sprachnachrichten will sich verständlich machen und der Empfänger der Nachricht versucht diese zu erkennen und zu verstehen. Dieser Ablauf basiert auf Spracherzeugungsmechanismen beim Sprecher, Übertragung im Medium und einem Sprachübertragungsprozess im Ohr bis hin zum Hirn des Empfängers. Diese Kette von Ereignissen, vom Konzept einer Nachricht im Verstand des Sprechers bis zur Ankunft der Nachricht im Verstand des Empfängers nennt sich Sprachkette [22].

Bei jedem dieser Kettenglieder können Fehler auftreten, die die Nachricht am Ende der Kette verfälschen könnten. Es ist aber dem Menschen möglich sein Kommunikationsverhalten der Umgebung und dem Gesprächspartner anzupassen und somit Fehler zu minimieren. Würde es beispielsweise durch zu leises oder undeutliches Reden oder durch mangelnde Kenntnis der Sprache des Gesprächpartners zu einem Informationsverlust kommen, dann könnte der Empfänger die Nachricht durch Nachfragen wiederherstellen. Würde die Kommunikation z.B. in einer Umgebung mit einer Geräuschkulisse stattfinden, würde der Sender die Stimme heben und die Silben anders betonen, um besser verstanden zu werden [22].

Das menschliche Gehör kann zudem in Anwesenheit mehrerer unabhängiger Schallquellen durch Lenkung der Aufmerksamkeit auf den Sprecher, der verstanden werden soll (Zielsprecher) seine Schallanteile aus dem Geräuschgemisch extrahieren und verstehen. Dieser Effekt ist besonders ausgeprägt wenn die Störgeräusche Sprache oder sprachähnlich sind, so dass der Zielsprecher rausgehört und verstanden werden kann,

siehe z.B. Hawley et al. 2004 [17]. Dieser Effekt ist eine Fähigkeit des Gehörs, die es dem Menschen ermöglicht in einer geräuschbehafteten Umgebung (bspw. Cocktailparty) ein Gespräch zu führen und wird als Cocktailpartyeffekt bezeichnet [4].

Zum Cocktailparty-Phänomen wurden zahlreiche Untersuchungen vorgenommen, wodurch Modelle erstellt werden konnten, die die Abhängigkeit der Sprachverständlichkeit von den spektralen Eigenschaften der Stimme (Langzeitsprektrum) und dem Amplitudenspektrum eines stationären, in seiner Energie gleichbleibenden Störgeräuschs beschreiben. Der Einfluss der Faktoren wie Art, Anzahl und Positionierung der Quellen wurden bereits untersucht. So sind Modelle vorhanden, die die Auswirkung der spektralen Unterschiede einer Stimme bezogen auf ihr Langzeitspektrum beschreiben, wie z.B. der Speech Intelligibility Index (SII) und der Articulation Index (AI). Ebenso wurde der Einfluss der Modulation der Sprache durch Formung im Vokaltrakt die zu zeitlichen Sprachschwankungen, der den Sprachfluss und damit u.a. die Pausen zwischen den Wörtern beschreibt, untersucht und modelliert. Ein Modell für Sprachverständlichkeit bei zeitlich instationären Störgeräuschen wurde entwickelt [25]. Es sind Modelle erstellt worden, die die Auswirkung der räumlichen Positionierung für ein oder mehrere Störquellen und für Unterschiede bezüglich zugewandtem Ohr oder binauralem Hören beschreiben [4]. Auch der Einflussfaktor des Raumschalls konnte modelliert werden und wird durch den Speech Transmission Index (STI) beschrieben [18]. Für den Aspekt der Sprecherähnlichkeit und der selektiven Aufmerksamkeit des Hörers hingegen sind noch keine Modelle vorhanden [4].

1.2 Stand der Forschung

Erst vor kurzer Zeit wurden die beiden Faktoren Sprecherähnlichkeit und selektive Aufmerksamkeit systematisch analysiert. Es konnte hierfür gezeigt werden, dass die selektive Aufmerksamkeit des Empfängers durch räumliche Positionierung und Lautstärkeunterschiede beeinflusst wird. Abhängig von der Positionierung der Geräuschquellen und der Pegelunterschiede ist es dem Hörer möglich seine Aufmerksamkeit unterschiedlich stark auf den Zielsprecher zu lenken und damit die Informationen des

gewünschten Senders zu erhalten [19].

Es hat sich gezeigt, dass im Falle unterschiedlicher Sprechereigenschaften des Zielsprechers und eines oder mehrerer Störsprecher unterschiedliche Ergebnisse in Bezug auf die Sprachverständlichkeit entstehen. Bei einer hohen Sprecherähnlichkeit zwischen Ziel- und Störsprecher nimmt die Verständlichkeit im Gegensatz zu deutlich unterschiedlichen Stimmen ab [12]. Die Signale des Zielsprechers und des Maskierers sind dann zwar hörbar, aber es ist dem Hörenden nicht möglich diese beiden Signale immer zu trennen und/oder zu unterscheiden, da die selektive Aufmerksamkeit sich nicht immer auf den Zielsprecher lenken lässt und ein Informationsverlust verursacht werden kann. Dieser Effekt wird als Informational Masking (IM) bezeichnet. Die Aufmerksamkeit die dem Zielsprecher gilt, kann also durch die sprecherspezifischen Eigenschaften des Störsprechers beeinflusst werden [13].

Bei der Untersuchung des Einfluss der Sprecherähnlichkeit auf das Informational Masking hat sich gezeigt, dass sich sowohl die Grundfrequenz als auch die Stimmschwankung des Störsprechers auf die Sprachverständlichkeit auswirken. Je unterschiedlicher die Grundfrequenz oder die Stimmschwankungen zwischen Ziel- und Störsprecher sind, desto deutlicher können die beiden Quellen unterschieden werden und desto seltener tritt Informational Masking auf. Vor allem die Kombination aus beiden Faktoren, oder die Stimme eines Sprechers eines anderen Geschlechts ermöglichen eine differenziertere Wahrnehmung des Störsprechers mit zusätzlichen Unterscheidungskriterien wie Länge des Vokaltrakts oder Sprechweise (Geschwindigkeit, Betonung, etc.) [12].

Durch diese Form der Maskierung der Informationen können beim Empfänger Lücken in der Verständlichkeit entstehen. Allerdings muss korrekte Sprachverständlichkeit nicht immer aus dem vollständigen Verstehen aller Wörter resultieren. Fehlende Sprachfragmente können oft aus dem Kontext, d.h. den phonologischen und lexikalischen Gegebenheiten sowie syntaktischen und semantischen Regeln erschlossen werden. Der Empfänger könnte somit aufgrund des Kontexts auf eine bestimmte Anzahl der maskierten Wörter des Senders verzichten und würde trotzdem die Botschaft des Senders

verstehen [6].

Abhängig der Faktoren, der räumlichen Positionierung, der Lautstärkeunterschiede und der Sprecherähnlichkeit zwischen Ziel- und Störsprecher wird die Aufmerksamkeit des Empfängers unterschiedlich stark abgelenkt. Wie sehr sich der Einfluss des Kontexts dabei auswirkt, ist noch nicht vollständig geklärt. Zwar kann bereits abhängig von der Lautstärke basierend auf dem Einfluss des Kontexts die Sprachverständlichkeit berechnet werden [5], aber in welchen Relationen die aufgezählten Faktoren (Positionierung der Quellen, Lautstärke und Sprechereigenschaften) und der Einfluss des Kontexts zueinander stehen und wann IM dabei am stärksten auftritt, ist noch nicht untersucht worden.

Zur Bestimmung menschlicher Sprachwahrnehmung wird die Abhängigkeit der Verständlichkeit vom relativen Signalpegel zwischen einem Störsignal und dem gewünschten Sprachsignal ermittelt. Bei Messungen mit z.B. Sprache als Maskierer wird vom Target-Maskier-Ratio (TMR), dem Pegelverhältnis zwischen Ziel- und Störsprecher ausgegangen (in dB). Dabei wird die Sprachverständlichkeit als der Anteil oder Prozentsatz der korrekt wiedergegebenen Testsegmente gemessen. Die gewählten Testsegmente (Silben, Wörter, Sätze) werden bei Messungen mit unterschiedlichen Pegeln dargeboten. Aus den resultierenden Messdaten wird dann der Pegel bei 50 % der richtig erkannten Segmente errechnet. Dieser Schwellwert der Sprachverständlichkeit (Speech Reception Threshold, SRT) dient als eine der möglichen Zielgrößen zur Sprachverständlichkeitsmessung [2].

1.3 Ziel der Arbeit

Mit Hilfe dieser adaptiven Messmethode nach Kollmeier et al. 2002 [2] kann Sprachverständlichkeit in Abhängigkeit von dem Lautstärkepegel von Ziel- und Störsprecher ermittelt werden. Dieses Verfahren soll im Rahmen des in dieser Arbeit geplanten Versuchs eingesetzt werden. Als Testmaterial werden sowohl für den Ziel- als auch für den Störsprecher Sätze nach Benoit et al. 1996 [1] verwendet, die kontextfrei sind,

wodurch davon ausgegangen werden kann, dass zum Verstehen nicht von der Semantik auf andere Bestandteile des Satzes geschlossen werden kann. Mit dieser Methode soll für verschiedene Sprechereigenschaften des Maskierers die Sprachverständlichkeit ermittelt werden.

Um gezielt den Einfluss der Semantik zu untersuchen, werden die normierten kontextfreien Sätze paarweise zusammengefügt, so dass bestimmte Wörter in Kombination beider Sätze sinnhaft sind. Diese Satzkombinationen sollen im finalen Versuch jeweils von Ziel- und Störsprecher mit unterschiedlich ähnlichen Stimmeneigenschaften dargeboten werden. Damit soll ermittelt werden wie stark der Einfluss des semantischen Kontexts gegenüber den Sprechereigenschaften des Maskierers sein kann.

Ziel dieser Arbeit wird es daher sein die Auswirkungen sprecherspezifischer Eigenschaften auf den Effekt des Informational Maskings zu untersuchen bei spezieller Berücksichtigung semantischer Vorhersagbarkeit des verwendeten Sprachmaterials.

2 Grundlagen

2.1 Auditory Scene Analysis (Sprachverständlichkeit und Cocktailparty-Effekt)

Sprachverständlichkeit ist gegeben durch die Fähigkeit des menschlichen Gehörs Klang, der in Form von Schallwellen auf das Ohr auftrifft, in wahrnehmbare bedeutungsvolle Elemente umzusetzen. Dieser Prozess, der es dem menschlichen Gehör ermöglicht die auditive Umgebung zu analysieren und zu verstehen, wird als Auditory Scene Analysis bezeichnet [3].

Ein ankommendes akustisches Signal kann als Ganzheitliches wahrgenommen werden, wie beispielsweise ein Akkord in der Musik, oder getrennt als individuelle Komponenten zugehörig zu verschiedenen Geräuschquellen. Durch Verknüpfung dieser einzelnen Komponenten über die Zeit hinweg (auditory streaming) wird es möglich Phoneme, Wörter, Sätze als Ganzes ausgehend von einem Sprecher wahrzunehmen [3].

Sprachwahrnehmung basiert auf verschiedenen Stadien der Analyse: der auditorischen, phonetsichen, phonologischen, lexikalischen, semantischen und syntaktischen. Diese können als ein serieller Prozess aufgefasst werden, in dem das Sprachsignal stufenweise in eine weitere verfeinerte Repräsentation transformiert wird und dieses dann als sprachliche Nachricht endet [22]. Nach der primären auditorischen Analyse auf dem Weg vom Trommelfell bis zum Gehörnerv, wird das ankommende Signal in Muster unterscheidbarer auditorischer Eigenschaften codiert. Auf Basis dieser Eigenschaften kann Periodizität von Sprachenergie und Frequenzen erkannt werden. Dies ermöglicht die Mustererkennung von Lauten. Aufbauend darauf wird eine Erkennung linguisti-

scher Einheiten auf der Ebene von Phonemen, der kleinsten Einheit der Sprache, und weiteren größeren Einheiten, wie Silben oder Wörter bis hin zur empfangenen Nachricht durchgeführt [22].

In alltäglichen Situationen treten durch mehrere Geräuschquellen, die aus unterschiedlichen Richtungen Schall abstrahlen, komplexe Signale auf, mit denen das Gehör konfrontiert wird. Es ist dem Gehör trotzdem möglich aus diesem Geräuschgemisch eine bestimmte Quelle zu extrahieren, deren Signale über die Zeit hinweg zu verknüpfen und zu verstehen. Dieses Phänomen wird als Cocktailparty-Effekt bezeichnet. Es beschreibt die Fähigkeit durch gerichtete Aufmerksamkeit einen einzelnen Sprecher aus einem Geräuschgemisch von Konversationen und Hintergrundgeräuschen rauszuhören [4].

Das Cocktailparty-Phänomen wurde bereits 1954 von Cherry [11] zitiert und wurde in den letzten 55 Jahren umfassend erforscht. Dennoch gibt es viele offene Fragen, da viele Aspekte Einfluss auf diesen Effekt zeigen. Die Sprachverständlichkeit des Sprechers, der verstanden werden soll (Zielsprecher), hängt zum einen von Typ und Anzahl der Störgeräusche ab, es kann sich hierbei um eine Geräuschkulisse durch stationäre oder instationäre Hintergrundgeräusche wie z.B. mehrere Sprecher oder um einzelne Störsprecher handeln. Dabei ist die akustische Umgebung in der sich die Geräuschquellen befinden von Bedeutung, da der Nachhall in einem Raum Einfluss auf die Sprachverständlichkeit haben kann. Einfluss hat zum anderen auch die räumliche Positionierung der störenden Geräuschquellen auf die Sprachverständlichkeit, da diese mit zunehmender Entfernung von dem Hörer unterschiedlich wahrgenommen werden. Hinzu kommen Unterschiede des binauralen gegenüber einem rein monauralen Hörens, die durch Ausnutzung des Vergleichs der beiden Ohrsignale zustande kommen [4]. Ebenso wurde gezeigt, dass Änderungen im Frequenz-Spektrum und der zeitlichen Modulation in den Stimmen des Ziel- und Störsprechers (Sprechereigenschaften) Unterschiede in der Sprachverständlichkeit erzeugen [4] [12].

2.2 Sprechereigenschaften

Durch unterschiedliche Stimmeigenschaften der Sprecher ist es möglich das Gesagte den einzelnen Sprechern zu zuordnen. Diese Sprechereigenschaften werden durch das sprecherspezifische Frequenzspektrum der vom Sprecher geformten Laute und durch die spezifischen zeitlichen Änderungen definiert. Beispiele sind eine sprecherspezifische Vokalartikulation, Konsonantenartikulation, Prosodie und Grundfrequenz [22].

2.2.1 Spracherzeugung

Stimmerzeugung wird initiiert, wenn die Stimmlippen des Kehlkopfs locker aneinander liegen und somit die Glottis verschließen, siehe Abbildung 2.1. Durch die vom Zwerchfell und den Lungen produzierten Phonations-Luftstrom werden die Stimmlippen zunächst auseinandergepresst, so dass die Luft in den darüber liegenden Vokaltrakt strömt. Aufgrund der Bernoulli-Kraft werden die Stimmlippen wieder zusammengepresst und die Glottis schließt sich. Dadurch wird Luft mit jeder Öffnung und Schließung stoßweise in den Artikulationsraum entlassen, wodurch periodische Schwingungen (das Anregungssignal der Sprache) entstehen. Die Frequenz dieser Schwingungen (etwa im Bereich 70-1000 Hz hängt u.a. von der Länge, sowie der während des Sprechens ausgeübten Spannung der Stimmlippen ab [15].

Die Frequenz der Schwingungen der Stimmlippen wird als Grundfrequenz (F0) bezeichnet. In der Realität sind periodische Schwingungen immer mit einem gewissen Anteil an Nebenwellen oder Oberwellen behaftet. Oberwellen sind höhere beteiligte Schwingungen, deren Frequenz ein Vielfaches der Grundfrequenz betragen. Die Grundfrequenz ist somit die tiefste Frequenz in einer harmonischen Serie und ist die kleinste Frequenzkomponente eines Signals [15].

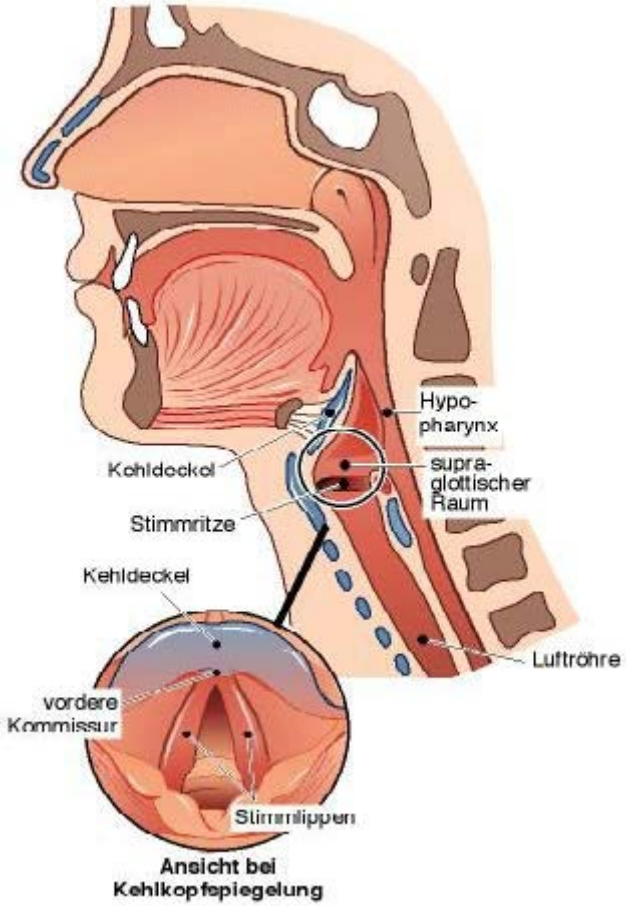


Abbildung 2.1: Kehlkopf eines Menschen

2.2.2 Sprecherunterschiede

Durch Resonanzeigenschaften, die von spezifischer Form des Vokaltrakts zu gegebenem Zeitpunkt abhängen, werden bestimmte Frequenzbereiche im Verhältnis zu anderen Frequenzbereichen verstärkt. Die Maxima der Energie in bestimmten Frequenzbereichen werden als Formanten bezeichnet. Die Lage und Ausprägung der Formanten formen maßgeblich die Klangfarbe einer Stimme. Durch sie lassen sich Stimmen voneinander unterscheiden. Bei der menschlichen Sprache charakterisiert die Lage der ersten beiden Formanten bestimmte Laute. Allerdings haben nur Formanten, die im Bereich oder oberhalb der Grundfrequenz eines Tons liegen, Auswirkungen auf die Klangfarbe des Tons, siehe als Beispiel in einem Spektrogramm die Vokale 'i', 'u', 'a' in Abbildung 2.2 gebildet durch die Formanten F1 und F2.

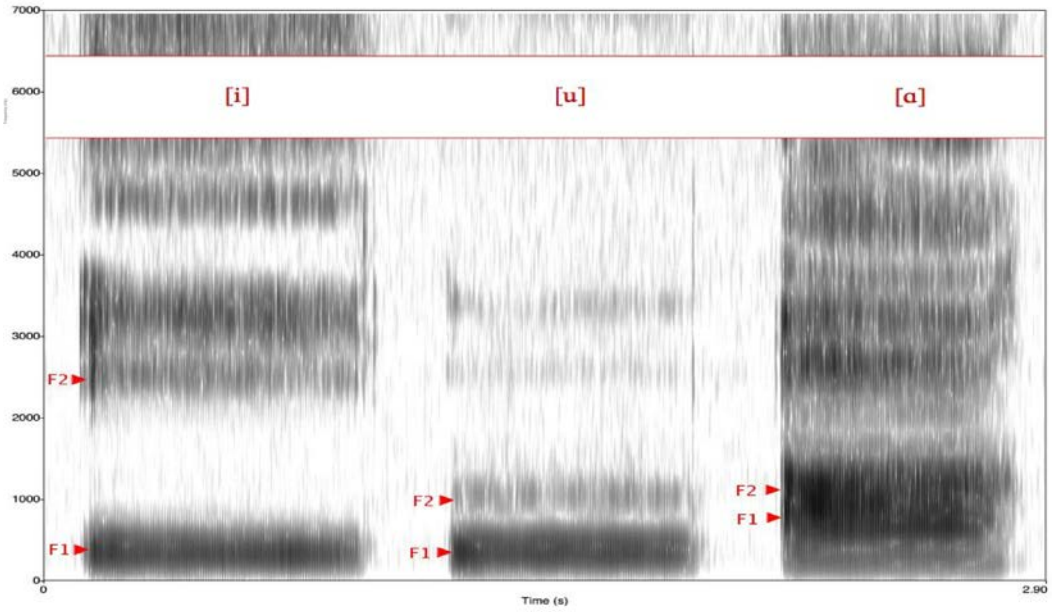


Abbildung 2.2: Spektrogramm für Vokale 'i','u','a'

Dabei ist die Frequenzlage der Vokaltrakt-Resonanzen für die Grundfrequenz F0: 50-500 Hz, und liegt beim Mann etwa bei 130 Hz und bei der Frau bei etwa 240 Hz. Für den ersten Formanten liegen sie bei F1: 200-1200 Hz für den zweiten Formanten bei F2: 0.7-3.3 kHz. Ihre Lage im Frequenzbereich charakterisiert den gesprochenen Vokal, siehe Tabelle 2.1. Neben der stimmhaften Anregung des gesprochenen Vokals, sind stimmlose Anregungen wie Reibe- oder Friktionsgeräusche [s],[f],[c],[S],[x],[z] (IPA Schreibweise, Internationales Phonetisches Alphabet) und Plosivanregungen (Verschlusslaute) [p],[t],[k],[b],[d],[c] aufgrund spezifischer Variationen, Kennzeichen für einen bestimmten Sprecher. Die Frequenzlage des dritten Formanten liegt bei F3: 1.5-3.7 kHz. F3 hat Auswirkungen auf die Klangfarbe einer Stimme und variiert je nach Sprecher [15].

deutscher Vokal	IPA	Formant 1	Formant 2
U	u	320 Hz	800 Hz
O	o	500 Hz	1000 Hz
å	?	700 Hz	1150 Hz
A	a	1000 Hz	1400 Hz
ö	ø	500 Hz	1500 Hz
ü	y	320 Hz	1650 Hz
ä	?	700 Hz	1800 Hz
E	e	500 Hz	2300 Hz
I	i	320 Hz	3200 Hz

Tabelle 2.1: Vokal-Formant-Zentren

Durch die zeitliche Änderung der Grundfrequenz lässt sich der Tonhöhenverlauf (Intonation) eines Wortes, Satzes oder Sprechakts beschreiben. Die Intonation und die zeitliche Strukturierung mittels Sprechtempo, -rythmus und -pausen bilden die Prosodie eines Sprechers. Dadurch wird unter anderem die Sprachmelodie des Sprechers bestimmt [31].

2.3 Auditory Streaming

Das auditory streaming führt Gruppierungen auf Basis der Frequenz und der Zeit durch. Je kleiner die Zeitabschnitte und je unterschiedlicher die Frequenzen, desto stärker ist der Effekt der Gruppierung. Durch spektrale und zeitliche Änderungen können Klangereignisse verschiedener Quellen getrennt und auch über eine zeitliche Dauer hinweg zugeordnet werden [3]. Darauf basierend können Sprecher bezüglich ihrer Sprechereigenschaften entsprechend deutlich voneinander getrennt wahrgenommen werden. Ein Sprecher allein in einem Raum kann i.A. deutlich verstanden werden. Anhand seiner Sprechereigenschaften kann derselbe Sprecher in einem Raum mit mehreren Sprechern, wie es beispielsweise bei einer Cocktailparty vorkommt, meist ebenfalls verstanden werden. Seine Stimme kann aus der Überlagerung mehrerer Geräusche aus dem Frequenzgemisch rausgefiltert werden. Wie deutlich die Stimme des Zielsprechers dann wahrgenommen und verstanden wird, hängt von den bereits genannten Faktoren, wie der Anzahl der Störsprecher, der Raumakustik usw. ab [3].

Wahrnehmung ist ein Prozess des Aufbaus von Beschreibungen. Die Aufgabe der Wahrnehmung ist es aus den sensorischen Eingangsdaten eine nutzbare Repräsentation der Realität abzuleiten. Ein wichtiger Teil des Aufbaus einer Repräsentation ist es zu entscheiden, welcher Teil des sensorischen Stimulus dasselbe wahrgenommene Ereignis beschreibt. Also die Zuordnung der erhaltenen Informationen zum Verursacher, bereits bekannt als Auditory Scene Analysis [3]. Die Aneinanderreihung von mehreren Geräuschen die einer Ursache zuzuschreiben sind, ist eine Form der Ereignisrepräsentation und wird als das auditory streaming bezeichnet. Die Geräusche werden in unterschiedlich wahrnehmbare Repräsentationen der Geräuschquellen umgesetzt, so dass ihre wahrgenommenen auditiven Eigenschaften miteinander verbunden werden. Die Sinnesinformationen werden nach zugehörigen Eigenschaften zur Beschreibung eines auditiven Ereignisses gruppiert und zusammengeführt [3].

Sprache ist ein Breitbandsignal, mit einem Frequenzspektrum das sich als Funktion der Zeit verändert. Die wahrgenommenen Frequenzen bilden zusammen eine Art neurales Spektrogramm. Auf Basis des spektralen Musters und deren zeitlichen Änderungen werden bestimmte Gruppierungen durchgeführt [3]. Wenn alternierend hohe und tiefe Frequenzen in schneller zeitlicher Abfolge nacheinander zu hören sind, werden diese Töne als 2 verschiedene Klangereignisse, die parallel zueinander auftreten, wahrgenommen. Je höher der Frequenzunterschied der Töne ist, desto stärker werden sie jeweils als einzelne Töne gruppiert [3]. Dieser neuronale Mechanismus ist zeitlich abhängig und wird bei der Gruppierung nach Frequenzunterschieden mit ansteigender Zeit zwischen den Tönen weniger effektiv. Je schneller die unterschiedlich hohen Töne aufeinanderfolgen, desto eher werden diese in der Wahrnehmung zu ein und demselben Ereignis zugeordnet. Diese Trennung auf Basis von Frequenz und Zeit wird als streaming-Effekt bezeichnet. Dabei werden bestimmte Regeln beachtet, was als wahrgenommenes Ereignis selektiert und was verworfen wird [3].

Aus der Beziehung der Signaleigenschaften zueinander kann auf die Zugehörigkeit zum Ereignis geschlossen werden. Dabei kann ein Ton prinzipiell unterschiedlichen Ereignissen angehören, wird aber nur einem Ereignis zur gleichen Zeit zugeordnet und interpre-

tiert. Damit kommt es zu keiner Mehrfachzuweisung zum selben Zeitpunkt [3]. Treten mehrere akustische Ereignisse zur gleichen Zeit auf, werden diese i.A. zu verschiedenen Bereichen im Spektrum zugeordnet und als alleinstehende Eigenschaften für die jeweiligen Klangereignisse behandelt. Somit ist es möglich aus diesem Geräuschgemisch nicht nur Töne, sondern auch Spektren verschiedener Ereignisse voneinander zu trennen. Würden beispielsweise 2 Sprecher zur gleichen Zeit 2 unterschiedliche Vokale sagen, würden sich aufgrund der Klangfarbe der Sprecher und der gerade geäußerten Vokale die Spektren unterscheiden. Die verschiedenen Spektren treten zur gleichen Zeit auf und können bezüglich ihrer Quellen unterschieden werden. Dieser Prozess wird als spektrale Integration bezeichnet [3]. Bestimmte Änderungen im Spektrum und in der Geschwindigkeit der wahrgenommenen Ereignisse hingegen führen zur sequentiellen Integration. Dadurch können Ereignisse miteinander verbunden werden, die zu unterschiedlichen Zeitpunkten auftreten aber zur selben Geräuschquelle gehören. Somit ist es zum Beispiel möglich der Melodie eines Musikstücks zu folgen [3].

2.4 Maskierung

Treten akustische Elemente des Zielsprechers in den spektralen Regionen auf, in denen sie zeitlich und im Frequenzbereich mit einem stärkeren, maskierenden Signal z.B. von einem Störsprecher überlappen, werden sie von diesem Reiz eliminiert. Sie sind daher nicht in der Lage vom Hörer wahrgenommen zu werden. Diese spektrale Überlappung wird als Energetic Masking (EM) bezeichnet [3]. Diese energetische Maskierung ist definiert als Maskierung entstanden als Überlappung der Reizmuster des Ziel- und Störsprechers in der Cochlea oder dem Hörnerv. Sie ist eine Maskierung an der Peripherie des Hörsystems. Diese periphere Maskierung der lokalen spektrotemporalen Information, wie Frequenz, Räumlichkeit und Lautstärke beeinflusst die Kurzzeit-Segmentierung, der Verknüpfung von lokalen Zeit-Frequenz Komponenten [3].

Bei der Wahrnehmung eines Zielsprechers bei Störgeräuschen kann neben der Maskierung der Signalenergie eine andere, nicht energetische Art der Maskierung auftreten. Diese wird als Informational Masking (IM) bezeichnet und ist keine im eigentlichen

Sinne physische Maskierungsform. Das IM wird durch Fehler der gerichteten Aufmerksamkeit auf den Zielsprecher oder durch Fehler bei der Segregation der akustischen Quellen verursacht. Das Nutzsignal kann dann nicht von den Störsignalen getrennt und/oder zugeordnet werden [13].

Die Auswahl und die korrekte Zuordnung der Segmente des Zielsprechers durch auditory streaming erfolgen aufgrund der Gewissheit der Sprechereigenschaften des Zielsprechers. Die gelenkte Aufmerksamkeit auf die Selektion speziell dieser Segmente kann durch hohe Sprecherähnlichkeit des Störsprechers oder durch lokale spektrotemporale Informationen beeinflusst werden. Dies kann zu Fehlern der selektiven Aufmerksamkeit führen [27] [12] [8]. Bei dem Einfluss der Sprecherähnlichkeit zwischen Ziel- und Störsprecher in Bezug auf Wahrnehmung oder linguistischen Merkmalen auf die gerichtete Aufmerksamkeit [13], ist es entscheidend wie gut die lokalen Zeit-Frequenz Segmente des Zielsprechers aus dem Geräuschgemisch zur gleichen Zeit extrahiert werden können und wie stark die Verbindung dieser Segmente über längere Zeit hinweg erreicht werden kann. Die zeitliche Verknüpfung der Segmente wie die Verbindung von Silben zu Wörtern, wird durch die Kontinuität des Signals über die Zeit hinweg ermöglicht [19]. Diese Kontinuität kann beispielsweise durch die spezifische Lautstärke, Tonhöhe, Klangfarbe, Grundfrequenz oder den Abstand der Wörter eines Sprechers, eben den Sprechereigenschaften gegeben sein [3].

2.5 Top-Dow und Bottom-Up Prozesse

Top-Down und Bottom-Up sind Strategien der auditiven Informationsverarbeitung und der Wissensabwicklung. Die Informationsverarbeitung der beiden parallel arbeitenden Prozesse basiert auf Klangfarbe, Frequenz und Zeit. Wobei sie durch die verschieden langen Zeitabschnitte in denen die Informationen verarbeitet werden, zu unterscheiden sind [3].

Der Top-Down-Prozess ist für Informationen zuständig, die über einen längeren Zeitabschnitt andauern, wie zum Beispiel ein Wort, ein Satz, ein Text. Der Bottom-Up-

Prozess hingegen behandelt Informationen eher lokal. Damit ist gemeint, dass diesem Prozess die Kurzzeitsegmentierung, die Partitionierung und Aufteilung der ankommenden Signale zugeschrieben wird. Zur Aufschlüsselung der Informationen benutzt der Prozess akustische Eigenschaften, wie Frequenzumgebung, spektrale Ähnlichkeit, Korrelationen der Änderungen der akustischen Gegebenheiten [3]. Der Bottom-Up-Prozess kann als ein datengesteuertes System angenommen werden, dass Informationen, angefangen bei der akustischen Wellenform zu verständlichen hörbaren Gebilden zusammenfügt [14].

Aus der komplexen Form der sensorischen Aufnahme, müssen die einzelnen Eigenschaften auf ihre Effekte zurückgeführt werden. Dafür müssen die Effekte und ihre Eigenschaften bereits bekannt sein. Es muss daher eine mentale Beschreibung dieser Effekte geben, die es ermöglicht diese Geräusche zuzuordnen und sie damit zu unterscheiden [3]. Bei dem Top-Down-Prozess wird angenommen, dass aus gespeicherten Schemata, wie Wortgruppen, Frequenzkomponenten entnommen werden und diese mit den akustisch wahrgenommenen Mustern (Frequenzen) abgeglichen und nach Übereinstimmungen durchsucht werden. Der Top-Down-Prozess ist daher ein schemabasierter Prozess, wodurch Daten ausgewählt werden, die nach bestimmten Kriterien passen, basierend auf einem abstrakten internen Modell zur Vorhersage und einem bereits bestehendem Wissen. Es ist ein vorhersagegesteuertes System, das Beweise sammelt in Form von Mustern (aus dem Bottom-Up-Prozess), um bisherige Annahmen über das Wahrgenommene anzupassen oder zu ändern [14]. Diese gezielte Abtastung nach Musterübereinstimmung erfordert eine gelenkte Aufmerksamkeit auf die entsprechende Annahme. Dieser Prozess ist daher für die selektive Aufmerksamkeit verantwortlich und kann durch Training verstärkt werden (Trainingseffekt) [3].

In dem schemabasiereten Top-Down-Prozess wird durch diese abgespeicherte Form von Erfahrung, Mustererkennung ermöglicht. Jedes Schema bindet Informationen über eine spezielle Regularität in der Umgebung ein. Das Schema kann unterschiedlich komplex sein, wie für ein Phonem a', für ein Wort Apfel', für eine grammatikalische Struktur oder einer Satzart. Ein Schema wird aufgerufen sobald in den sensorischen Eingangs-

daten, Muster erkannt werden, die in dem Schema enthalten sind. Das Schema bleibt aktiviert, da es viele Muster enthält, die zeitlich nacheinander folgen, wie beispielsweise das A' in Apfel'. Ein Schema kann auch von anderen Schemen aktiviert werden, die miteinander zusammenhängen. Sie agieren nicht parallel, sondern bilden einen Aufbau einer Beschreibung der wahrgenommenen Umgebung. Dadurch ist es möglich ganze Sätze wahrzunehmen und zu verstehen [3].

2.6 Kontext

Ein Satz beinhaltet kontextabhängige Informationen. Kontext bezeichnet in der Sprachwissenschaft, insbesondere in der Kommunikationstheorie, alle Elemente einer Kommunikationssituation, die das Verständnis einer Äußerung bestimmen.

Es wird unterschieden zwischen allgemeinem Kontext, persönlichem und sozialem, sowie sprachlichem Kontext. Als allgemeiner Kontext der Sprechsituation wird dabei der Ort, die Zeit und der Handlungszusammenhang der Äußerung bezeichnet. Als persönlicher und sozialer Kontext gilt die "Beziehung zwischen Sprecher und Hörer, ihren Einstellungen, ihren Interessen und ihrem Wissen bzw. ihren wechselseitigen Wissensannahmen". Als sprachlicher Kontext wird der Zusammenhang der die Ausdrücke grammatisch und semantisch verknüpft, bewertet [10]. Dazu gehören syntaktische Informationen, die auf der Satzstruktur basieren und semantische Information, die die Bedeutung der Wörter beschreiben. Ebenso werden Koartikulationseffekte dazugezählt, die Informationen beinhalten, die beim Übergang zweier Laute, durch die Positionen zu anderen Lauten entstehen. Kontext ist ebenfalls gegeben durch a priori Wissen des Satzthemas oder durch die Art wie die Sätze dargeboten werden [6].

Kontext hängt weniger von der Menge der enthaltenen Informationen ab. Durch Kontext können z. B. verlorengegangene Elemente, wie Wörter oder Phoneme, aufgrund von phonlogischen und lexikalischen Festlegungen, semantischen und syntaktischen Regeln, die die Reihenfolge und das Auftreten beschreiben, oft ersetzt werden [5].

3 Vorarbeit und Methode

3.1 Methode der Sprachverständlichkeitsmessung

3.1.1 Testmaterial

Sprache ist ein Produkt kommunikativer Aktion und ein Objekt seines Gebrauchs. Zur konkreten Messung des Objekts Sprache sollte die Situation der Konversation, oder die Umgebung in der Sprache als Funktion entsteht, in Form von entsprechenden Teststimuli berücksichtigt werden. Zur Ermittlung des kommunikativen Prozesses in typischen Situationen müssen diese Sprachstimuli als Messobjekt passend und repräsentativ sein. Es ist daher wichtig die möglichen Anwendungsbereiche zu kennen und zu benennen, wodurch eine nähere Analyse der Sprache durch qualitativer Sprachmessung möglich wird. Daher sollte im Vorfeld der Sprachqualitätstests auf die Wahl der Wörter die verstanden werden sollen, des Sprachmaterials und der Testpersonen geachtet werden. Zudem ist das Frageformat und die Art der Repräsentation der Stimuli zu berücksichtigen [20].

Die Sprachqualitätsbewertung ist eine Routineprozedur, die eine oder mehrere empirisch beschränkende qualitative Eigenschaften von wahrgenommener Sprache prüft, mit dem Ziel eine quantitative Aussage über diese Eigenschaften zu treffen. Durch Sprachqualitätstests können Messgrößen des Messobjektes Sprache bestimmt werden. Messung ist ein Werkzeug um Information eines auditiven und qualitativen Ereignisses in eine quantitative Form zu bringen, die die Basis für Datenanalyse und Interpretation ist. Sie kann nicht einheitlich angewandt werden, da sie sich nach der zu definierenden charakteristischen Messgröße richtet. Die charakteristischen Messgrößen sind abgeleitet von einer formalen Berücksichtigung der Wahrnehmung, dem Verstehen und der

Mitteilung, zudem vom Kontrollieren und wenn nötig Lenken der Aufmerksamkeit der Testperson. Die auditiven Messungen ermöglichen somit eine Beurteilung und Skalierung der wahrgenommenen Qualität des Sprachereignisses. Das Ziel von beiden, Messung und Skalierung ist es den Wert einer Messgröße in Zahlen zu repräsentieren. Die Skalierung bemisst dabei Menge und Grad. In der Sprachqualitätsmessung sind Skalen numerische Beschreibungen, die die gemessenen Ergebnisse der sprachbezogenen Beobachtung belegen und mitteilen [20].

Hierbei gibt es im Bereich der Sprachqualitätsmessung eine Anzahl von Standardtests und Prozeduren die verschiedene Aspekte der wahrgenommenen Sprache skalieren. Diese Tests können benutzt werden, um Sprachverständlichkeit, Prosodie, Sprach- und Höreinschränkungen zu testen. Bei der Sprachverständlichkeit können einzelne Sprachkomponenten (Phoneme, Silben, Wörter) getestet werden. Für Satztests in denen Wörter erkannt werden sollen, wurden Sätze verwendet, bei denen Wörter des Ziel- und des Störsprechers (im Falle von zwei Quellen) äquivalent im Sinnzusammenhang des Ziel- oder Störsatzes sind. So wurden beispielsweise häufig Sätze gemäß des Korpus für das sogenannte Coordinate Response Measure (CRM) aus der Militärforschung verwendet. Hierbei handelte es sich um Sätze der Form "Ready, (Name), go to (Farbe) (Nummer) now", wobei Name, Farbe und Nummer aus einer festen Menge von möglichen Alternativen ausgewählt wurden. Die Rahmensätze lieferten keinen semantsichen Hinweis darauf, welches Wort eingesetzt worden ist, so dass alle Optionen aus Sicht einer Versuchsperson gleich wahrscheinlich waren [20].

Einer weiterer Sprachverständlichkeitstests ist der Test mit sogenannten Semantically Unpredictable Sentences (SUS) [1] zur Analyse der verstandenen Wörter in einem Satzkontext mit semantisch unvorhersagbaren Sätzen [20]. Bei diesen Sätzen besteht zwischen den Wörtern kein semantischer Zusammenhang, sie sind zwar entsprechend ihrer vom Syntax vorgegebener Wortart, wie Adjektiv, Nomen usw. eingesetzt, aber im Gegensatz zum Korpus des CRM ist die Wortart wie z.B. Name und Farbe nicht genau bekannt. Es muss daher jedes Wort verstanden und sich gemerkt werden und kann nicht aus dem Kontext erschlossen werden. Lediglich der syntaktische und der

Koartikulations-Kontext bleiben erhalten, vgl. Kapitel 2.6.

Im Rahmen dieser Arbeit wurden Sätze nach 4 der 5 vorhandenen Satzstrukturen entsprechend der SUS-Satzart von Benoit et al. 1996 [1] für das Deutsche erstellt und verwendet, siehe Abbildung 3.1. Diese Sätze sind untereinander ähnlich und einfach strukturiert und daher vergleichbar, ebenso sind die zugrunde liegenden Ansätze für verschiedene Sprachen anwendbar. Als zu erkennende Segmente, die zur Messung und Auswertung dienen sollen, werden bestimmte Wörter (Schlüsselwörter) verschiedener Wortarten (Adjektiv, Nomen, (in-)transitive Verben), verwendet.

Nr.	Frage-Wort	Adj.	Trans. Verb	Nomen	Trans. Verb	Intr. Verb	Prep.	Adj.	Nomen	Rel. Pron.	Intr. Verb	Punktierung
1				Der Schrott		brennt	nach dem	wahren	Ei			.
2		Das faule		Lob	liebt				den Spott			.
4	Wo		trinkt	der Knick				den scharfen	Hund			?
5				Die Wut	lernt				die Wahl	die	lacht	.

Abbildung 3.1: Beispiele für die gewählten 4 Satzstrukturen des SUS-Textmaterials nach Benoit et al. 1996 für die deutsche Sprache angepasst

Zur Minimierung des Trainingseffekts wird eine hohe Anzahl an verschiedenen Sätzen benötigt, um zu vermeiden, dass die gleichen Schlüsselwörter bzw. Sätze mehrfach einer Testperson vorgespielt werden. Wie in Kapitel 2.5 beschrieben, kann durch häufige Darbietung bestimmter auditiver Stimuli die Sprachverständlichkeit trainiert werden. Der Einsatz von Sätzen mit einer höheren Anzahl an Wörtern, die den typischen Konversationsthemen entsprechen würden, könnte den Trainingseffekt zwar reduziern, würde die Vergleichbarkeit der Sätze allerdings mindern [29]. Würde die Anzahl der Wörter in den Sätzen stark eingegrenzt werden, würde dies zu einer vergleichbaren Komplexität der Sätze führen, den Trainingseffekt hingegen verstärken.

Einen Kompromiss zwischen Trainingseffekt und homogener Satzkomplexität bildet die Testmethode mit Sätzen der SUS-Satzart mit entsprechend erhöhter Anzahl an Schlüsselwörtern und angeglichener mittlerer Worthäufigkeit pro Testliste [1]. Eine

Testliste besteht aus einer bestimmten Anzahl der generierten Sätze, in dieser Arbeit liegt die Anzahl bei 12 Sätzen, und beinhalten demnach eine gleich hohe Anzahl an Schlüsselwörtern. Ebenso sind gleich viele Sätze (3 Sätze) jeweils einer Satzstruktur (4 Satzstrukturen) in jeder Testliste enthalten. Die Testlisten sind ausgewogen im Hinblick auf die Phoneme, die entsprechend der mittleren Phonemverteilung der deutschen Sprache, siehe Abbildung 3.2 vertreten sind. Es sollte daher möglich sein die Testlisten trotz unterschiedlich phonetischer, syntaktischer und semantischer Komplexität untereinander vergleichen zu können [1].

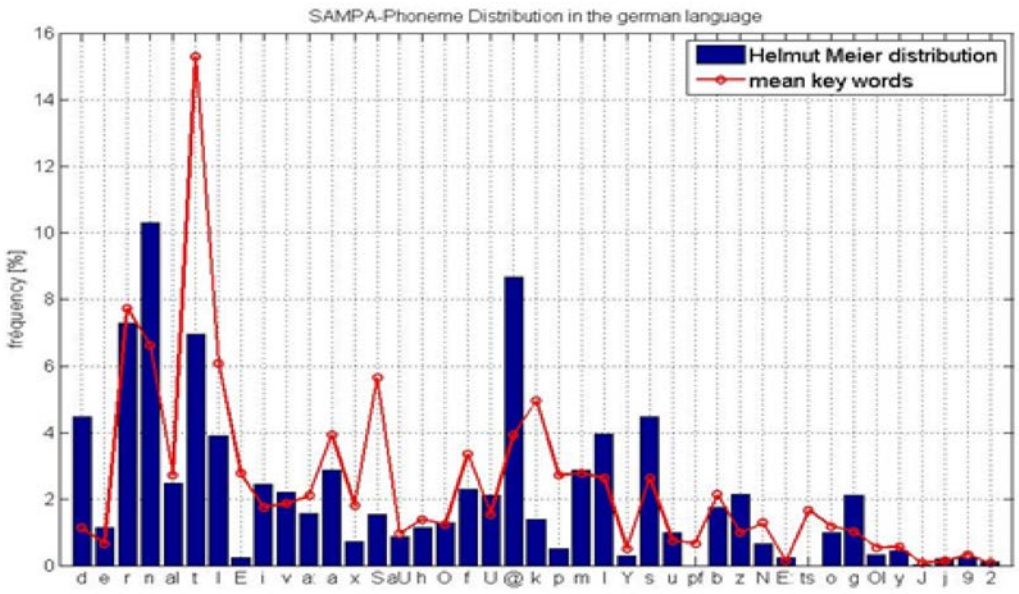


Abbildung 3.2: Phonemverteilung für alle Schlüsselwörter gemittelt (rot durchgezogene Linie) und für die deutsche Sprache nach Helmut Meier (blaue Balken)

3.1.2 Testmethode

Allgemein sind verschiedene Testarten für Sprachqualitätsbewertungen möglich, der Silbentest, Wörtertest (Freiburger Wörtertest, Döring Test) und Satztest (Oldenburger, Göttinger Satztest). Für die Sprachkommunikation ist ein Satztest repräsentativer als das Vorspielen eines Sprachsegments. Werden Sätze als Testmaterial verwendet ist die Messung des Schwellwerts, das Pegelverhältnis zwischen Zielsprecher und Rauschen (Signal-to-Noise Ratio, SNR), bzw. Störsprecher (Target-to-Masker Ratio, TMR) bei

dem 50 % Sprachverständlichkeit erreicht wird, präziser [21]. Dieser Schwellwert von 50 % -Sprachverständlichkeit (Speech Reception Threshold, SRT) dient als Parameter zur Sprachverständlichkeitsmessung [23]. Die sensible Messung basiert auf der schrittweisen psychometrischen Funktion von Sprachidentifikation bei Störgeräuschen. Die Sprachverständlichkeit als Funktion des Signal-zu-Rausch-Verhältnisses (SNR) wird als Diskriminationsfunktion bezeichnet [29], dessen Steigung der Literatur entnommen, zwischen 10 und 20 % pro dB liegt [2]. Es konnte bereits für phonemisch ausbalancierte Testlisten gezeigt werden, dass es keinen Unterschied der SRT-Werte zwischen ihnen gab, was die hohe Vergleichbarkeit zwischen den Testlisten bestätigte [23].

Der SRT wird für gewöhnlich in einem adaptiven Prozess für jeweils eine Testliste mit einer bestimmten Anzahl an Sätzen ermittelt. Allerdings wurde bislang bei der Wahl des Testmaterials mit Sätzen keine Kontextberücksichtigung vorgenommen. Bei Raake et al. 2006 [24] wurde entsprechend generiertes Sprachmaterial mit französischen Sätzen nach der Testmethode mit Sätzen der SUS-Satzart für einen SRT-Test entwickelt. Die vorliegende Arbeit basiert auf dieser SRT-Testmethode. Die Testreihe, die zur Ermittlung eines SRT-Wertes für verschiedene Störsprecher mit dem beschriebenen Satzmaterial für das Deutsche durchgeführt worden ist, bestand aus 3 Teilen.

Der erste Teil des Tests diente dazu den SRT-Wert für verschiedende Maskierer (SRT_mask) mit dem generierten Sprachmaterial bestehend aus Sätzen der SUS-Satzart durch die beschriebene adaptive Methode der Pegelanpassung zu ermitteln. Die so ermittelten Pegel des SRT_mask wurden im zweiten Test als fixe Werte der jeweiligen Maskierer dargeboten. Zudem wurden weitere feste SNR-Pegel, bzw. TMR-Pegel im Bereich von -6 und +6 dB um den jeweiligen SRT_mask -Wert getestet. Damit sollte die Sprachverständlichkeit in Abhängigkeit der ansteigenden Pegelunterschiede und die Auswirkung der Ähnlichkeit zwischen Ziel- und Störsprecher für dieses spezielle Sprachmaterial ermittelt werden. Im dritten Test wurde mit geänderten Satzpaarkombinationen bei festen Pegeln die Sprachverständlichkeit erneut getestet. Die Satzkombinationen wurden so erstellt, dass zwei aufeinander folgende Wörter beider Sätze sinnhaft verknüpft wurden. Damit sollte die Auswirkung des semantischen Kontexts

zwischen 2 Wörtern in Bezug auf die Sprachverständlichkeit untersucht werden und in inwiefern die semantische Verknüpfung dazu führte, dass die Wörter, die diese bilden anstatt der Wörter des Zielsprechersatzes wiedergegeben werden. Darüber hinaus sollte untersucht werden ob und wie stark sich dies von Sprechereigenschaften und dem Lautstärkeverhältnis zwischen Ziel- und Störsprecher abhängig zeigt.

Für die Tests wurden die generierten Sätze nach der SUS-Satzart, sowohl für den Ziel- als auch für den Störsprecher eingesetzt. Die Sätze wurden von dem Ziel- und Störsprecher gleichzeitig als ein Satzpaar gesprochen, vorgespielt. Als Zielsprecher wurde eine männliche Stimme benutzt. Es wurde davon ausgegangen, dass bei normal gesprochener Sprache eine bessere auditive Mustererkennung erfolgt und die natürlich gesprochene Sprache leichter erkannt wird, als bei normierter Sprache ohne Sprachmelodie [3]. Die Stimmen der Störsprecher sind die Stimme des Zielsprechers mittels der Software „PRAAT“ um unterschiedlich viele Halbtöne erhöht sowie eine Stimme einer weiblichen Sprecherin, so dass folgende Maskierertypen eingesetzt worden sind: gleiche Stimme, gleiche Stimme um 3 Halbtöne erhöht, gleiche Stimme um 6 Halbtöne erhöht, eine weibliche Stimme und zudem Rauschen. Es wurde bei den männlichen Stimmen als Störsprecher lediglich die Grundfrequenz geändert, bei der weiblichen Stimme hingegen sind alle Sprechereigenschaften, wie Grundfrequenz, Prosodie, Klangfarbe usw. an sich schon unterschiedlich und blieben unverändert. Das Rauschen ist ein stationäres, in seiner Energie relativ konstantes Rauschen und bestand aus derselben Grundfrequenz des Zielsprechers und enthielt keine Pausen. Es war zu berücksichtigen, dass der Zielsprecher bei entsprechenden Ziel- und Maskierer-Pegeln i.A. nicht energetisch maskiert wurde, also zu hören war. Zudem musste berücksichtigt werden, dass sich die Stimmen genau überlagern und nicht auf Grund von unterschiedlich lang gesprochenen Wörtern Pausen entstanden, in denen es möglich sein könnte den Zielsprecher herauszuhören, was als dip-listening Effekt bezeichnet wird [12].

3.2 Sprachmaterial

3.2.1 Wortmaterial

Es wurde mit Hilfe von WebCelex, einer online Datenbank für Wörter unterschiedlicher Sprachen, ein Wortkorpus für die deutsche Sprache erstellt. Unter dem Link: `http://celex.mpi.nl` kann auf das Programm zugegriffen werden.

WebCelex ist ein webbasiertes Interface mit einer Datenbank, die eine umfangreiche Wortbibliothek verschiedener Sprachen (Deutsch, Englisch, Niederländisch) enthält. Es ermöglicht durch vorgegebene Optionen bezüglich der Häufigkeit, Morphologie, Orthographie und Phonetik eine entsprechende Vorauswahl für die gewünschten Wörter zu treffen. Für die deutsche Sprache benutzt WebCelex den „Mannheim“ Korpus, entnommen 1984 dem Institut für Deutsche Sprache in Mannheim, welcher aus einer Anzahl von 6 Mio Wörtern, sowohl aus geschriebenen als auch aus gesprochenen Quellen besteht.

Als Grundlage für die Sätze der SUS-Satzart wurde Wortmaterial bestehend aus in der kanonischen Form einsilbigen Wörtern gewählt. Entsprechend der SUS-Satzart waren die Verben in der 3.Person Singular im Präsens, indikativ, aktiv in WebCelex ausgewählt worden [1]: Nach der Pfadwahl „Create Lexicon“ und „German Wordforms“ konnten die folgenden Eingaben zur Generierung dieser benötigten Wortformen getätigt werden. Die Häufigkeit, wie oft ein Wort in der Wortmenge von WebCelex vorkommt, konnte in der Kategorie „Frequency“ durch die Auswahl „Mann“ mitangegeben werden. Mit der gewählten Option „PhonSAM“ der Kategorie „Phonology“ wurde die phonetische Schreibweise jedes Wortes in dem phonetischen Alphabet SAMPA (Speech Assessment Methods Phonetic Alphabet) zusätzlich dargestellt. Durch die Auswahl Morphology „Sin3“ und „Pres“ sind Verben in der 3.Person Singular Präsens ausgewählt worden. Unter der Kategorie „lemmas“ wurde durch „LemmaSylCnt“ die Option angewählt die Anzahl der Silben der zu generierenden Wörter zu bestimmen. Und mit der Auswahl von „LemmaClass“ wurde die Wortart jedes Wortes mitangegeben, z.B.: für ein Verb ”V”. Im darauffolgenden Eingabefenster wurden durch die

logische Verknüpfung der Formel 3.1 die gewählten Angaben der zu generierenden Wortliste festgelegt.

$$(Mann >= 1)(Sin3 == 1)(Pres == 1)(LemmaSylCnt ==' 1') \tag{3.1}$$

Zur Ausgabe der Wortliste wurden für den Retrieve Mode „WWW“ gewählt und „word Include“, „counter Include“ und „query Include“ bestätigt. Die somit erzeugten ca. 300.000 Wörter wurden in eine txt-Datei kopiert, um sie in MATLAB einlesen zu können.

3.2.2 Generierung des Wortkorpus

Die generierte Wortliste wurde zunächst um einige Wörter reduziert, die besonders markant waren und sich somit leichter ins Gedächtnis einprägen konnten, siehe Anhang A.1. Hierzu gehörten Wörter, die politisch interpretiert werden konnten, die trotz relativ hoher Häufigkeit in der geschriebenen Mannheimverteilung im Gesprochenem ungeläufig waren, die als Schimpfwörter verwendet werden konnten, oder Symbolik aufwiesen. Ebenso wurden Wörter die aus Produkt- Firmen- und Personennamen bestanden aus dem Wortschatz entfernt, was durch eine Wortartangabe seitens WebCelex ermöglicht worden war. Sowohl Homographen als auch als Homophone, die mehrfach für ein Wort auftraten, wurden dem Wortschatz entnommen, um eine Dopplung in den zu generierenden Sätzen zu vermeiden. Dabei wurde jeweils das Wort mit der höheren Auftretenshäufigkeit beibehalten.

Durch die Beschreibung der Wortkategorien, die für jedes Wort durch WebCelex vorhanden waren, konnte nach der Reduzierung der Anzahl der Wörter, die Aufteilung der Wörter in Wortgruppen von Verben, Nomen, Adjektive, Adverbien, Präpositionen, Pronomen, Zahlen, Konjunktionen vorgenommen werden. Die Verben wurden manuell in Transitive und Intransitive aufgeteilt, da diese Untergliederung mit WebCelex nicht möglich und keine entsprechende Verbenliste in digitaler Form vorhanden war. Die Schlüsselwörter, die im Sprachverständlichkeitstest von den Testpersonen erkannt wer-

den sollten, sind Adjektive, Nomen, transitive und intransitive Verben [1]. Die Wörter dieser Gruppen wurden nach aufsteigender Häufigkeit sortiert, so dass entsprechend der Kategorie, alle Wörter die nicht häufiger als 3 mal vorkamen, entfernt werden konnten.

Bei einer zugelassenen Mehrfachnennung eines Wortes von 3 und einer Anzahl der Schlüsselwörter von 4 in einem Satz, lassen sich wie nach dem Vorbild der französischen Testmethode [24] 288 Sätze (24 Testlisten * 12 Sätze) bilden. Hierfür wurden 48 intransitive, 72 transitive Verben, 192 Nomen und 72 Adjektive benötigt. Zur Vorbereitung der Satzbildung wurden für die Adjektive und Nomen entsprechend des Kasus jede Form gebildet und abgespeichert. Somit konnte leichter zur Bildung der Sätze, entsprechend ihrer Satzstruktur auf Wörter im jeweiligen Kasus zugegriffen werden.

3.2.3 Generierung der Testlisten

Nach Benoit et al. 1996 [1] wurden die Sätze aus 4 unterschiedlichen syntaktischen Satztypen der Struktur 1,2,4,5 aus Abbildung 3.1 zu entnehmen, generiert. In den Sätzen wurde der direkte Artikel verwendet und Verben ausschließlich in der 3.Person Singular, Präsens, indikativ, aktiv. Durch die Satzbildung nach der jeweiligen Satzstruktur wichen allerdings die Adjektive von der kanonischen Form ab, wodurch sie nicht mehr einsilbig waren. Zudem war die Stellung der Adjektive zu dem Nomen im Deutschen anders als im Französischen. Das Adjektiv steht hier vor dem Nomen, wodurch sich die Abfolge der Schlüsselwörter änderte. Das hatte zudem zur Folge, dass der Artikel des Nomens nicht mehr bei seinem Nomen, sondern vor dem Adjektiv stand. Zudem musste bei Fragesätzen im Gegensatz zum Französischen das Verb vor dem Nomen gestellt werden.

Zur Satzgenerierung durch das Programm „MATLAB“ wurde für jede Wortkategorie eine Matrix zur Index-Verteilung erstellt. Jedem Wort wurden anfangs eine Zahl und ein Index der entsprechenden Wortgruppe zugeordnet. Dadurch sollte eine Mehrfachnennung eines Wortes in einer Testliste vermieden werden. Aus den 4 Matrizen wurden dann jeweils die 4 Schlüsselwörter pro Satz entnommen. Es wurden für jede

Testliste 48 (4 Schlüsselwörter * 12 Sätze) Schlüsselwörter, mit der zugrunde liegenden gemittelten Phonem-Verteilung ausgewählt, siehe Abbildung 3.2. Danach wurden aus diesem Wortpool die Sätze nach den 4 Satzstrukturen mit Hilfe der vorbereiteten und abgespeicherten Formen der einzelnen Schlüsselwörter für die entsprechenden grammatikalischen Fälle, gebildet. Die noch fehlenden Wörter (Präpositionen, Konjunktionen) zur Komplettierung des Satzes wurden randomisiert aus den bereits vorsortierten Wortgruppen ausgewählt.

Für alle Schlüsselwörter einer Testliste wurde die Summe ihrer Worthäufigkeiten nach der Wortart berechnet. Weichen die Werte von denen anderer bereits ermittelter Testlisten um mehr als 20 % ab, wurde die Liste erneut generiert. Bei 1000 Wiederholungen wurde die Konstellation mit der besten Chi-Quadrat-Statistik ausgewählt. Es wurde also geprüft, ob die Wortauswahl für jede Liste ähnlich gut war, wie der Mittelwert der Häufigkeit der zugrunde liegenden Wortliste. Die Wörter wurden mittels einer Minimierung des Abstands zwischen der mittleren Phonemverteilung des Deustchen und der Phonemverteilung im Korpus vorgenommen. Ein Chi-Quadrat Test ergab dabei aber keine signifikante Abweichung von einer zufälligen Auswahl.

Nach der automatisierten Testlistengenerierung wurden alle Sätze manuell auf eine eventuelle Sinnhaftigkeit geprüft. Falls ein Satz Sinn ergab, wurden die Wörter derselben Wortgruppe jeweils innerhalb einer Testliste ausgetauscht.

3.2.4 Sprachaufnahmen der generierten Sätze und Normierung

Die 288 generierten Sätze nach der SUS-Satzart wurden mit 8 professionellen Sprechern aufgenommen. Sie wurden durch eine Kollegin der T-Systems telefonisch interviewt, ausgewählt und für ihre Teilnahme bezahlt. Die Teilnehmergruppe bestand aus 4 weiblichen und 4 männlichen Teilnehmern, wobei innerhalb der geschlechtlichen Verteilung je zwei Teilnehmer hohe bzw. tiefe Stimmlagen aufwiesen. Ihnen wurde vorab das verwendete Sprachmaterial zugesandt, so dass sie sich auf die Sprachaufnahmen vorbereiten konnten. Bei dem Termin selbst wurden die Sprecher bezüglich Ablauf und

Sprechweise eingewiesen. Ihnen wurde mitgeteilt, die Sätze so natürlich wie möglich zu formulieren, um eine normale Sprechsituation zu simulieren. Wegen unterschiedlicher Interpretationsmöglichkeiten wurde jeder Satz zweimal gesprochen. Die Natürlichkeit des Redeflusses konnte so gewahrt bleiben.

Nach der Hälfte der Anzahl der zu sprechenden Sätze, für die im Schnitt ca. 30 Minuten gebraucht wurden, ist eine Pause von ca. 15 Minuten eingelegt worden. Zur Minimierung des Listeneffekts, der sich durch die mit der Zeit einstellende Änderung der Geschwindigkeit und Lautstärke des Sprechers ergibt, wurden die Sätze nicht der Listennummer nach, sondern in randomisierter Reihenfolge dargeboten und aufgenommen.

Für die Aufnahmen stand der reflektionsarme Raum (RAR) der Technischen Universität Berlin zur Verfügung. In dem Raum war ein Stahlnetz gespannt, der als Fußboden diente. Zusätzlich waren Gitterroste vorhanden, die weitestgehend entfernt werden konnten, aber aus Sicherheitsgründen zum Teil noch vorhanden sein mussten. Diese Konstruktion führte zu einem minimalen höher frequentem Nachhall, so dass die Sprecherposition möglichst nah an den absorbierenden Wänden gewählt worden ist, siehe Abbildung 3.3.

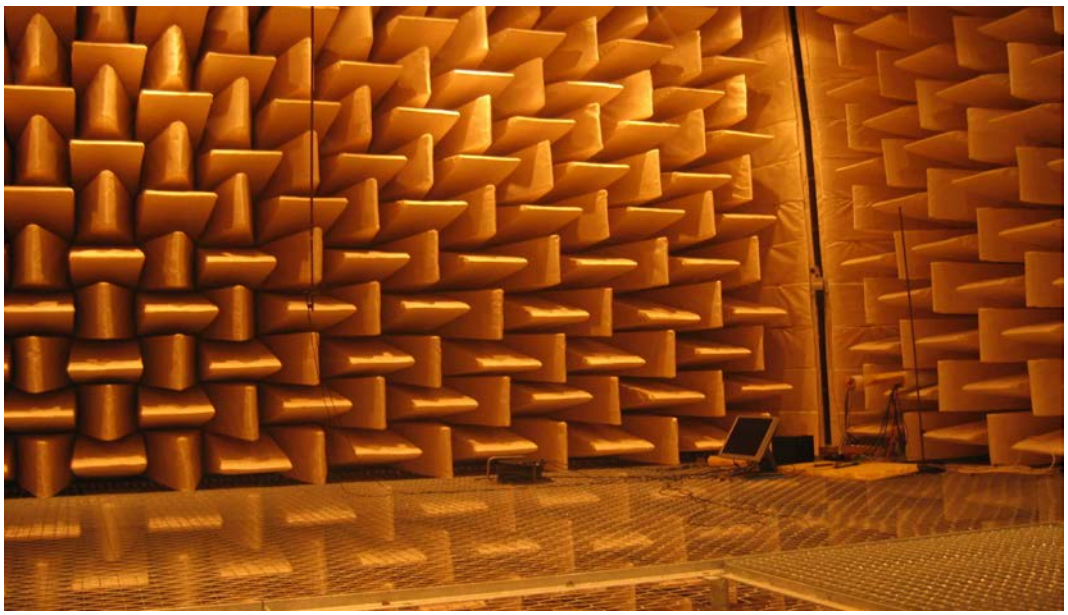


Abbildung 3.3: Aufnahmesituation im RAR TU Berlin

Die Sätze wurden dem Sprecher auf einem Monitor und nicht auf Papier dargeboten, so dass Knistern oder Rascheln vermieden werden konnte. Die Aufnahme- und Abhörmöglichkeit war in dem Flur vor dem RAR aufgebaut. Über einen Monitor und einen Lautsprecher konnte Kontakt zu dem Sprecher in dem RAR gehalten werden. Während der Aufnahmen wurde der Lautsprecher allerdings ausgestellt.

Nach einem digitalen Mikrofonverstärker „D19“ der Firma Studer wurden die Sprachsignale (20 Bit) per Soundkarte „Digi96“ von der Firma RME auf einen Computer mit Hilfe des Programms „Adobe Audition“ (hier Version 2.0) im Format "48kHz, 32bitäufgezeichnet. Der Pegel wurde am Mikrofonverstärker, in Adobe Audition und mit Hilfe des Softwaretools „DigiCheck“ von RME, das zusätzlich eine frequenzbezogene Pegelbewertung ermöglicht, geprüft. Der Aufnahmepegel am Mikrofonverstärker wurde so eingestellt, dass er im Schnitt zwischen 18 und 12 dBFS lag. Durch den Einsatz des digitalen Mikrofonverstärkers konnte der Störpegel auf 80 dBFS abgesenkt werden, was bei einem durchschnittlichen Signalpegel von 32 dBFS zu einem Signal-Rauschabstand von 48 dB führte. Für die Mikrofonierung wurde mit Ausnahme des nachträglich eingezogenen Deckengestänges ausschließlich Technik und Zubehör der Firma Schoeps verwendet. Die Mikrofone wurden per Gestänge von der Decke gehängt. Zusätzlich

zu dem Hauptmikrofon mit Kugelcharakteristik (MK2 + CMC6 + KC5), wurden die Aufnahmen mit einem Nierenmikrofon (MK4) durchgeführt.

Das Rohmaterial bestehend aus 30 Stunden Aufnahme, wurde auf einen genormten Vor- und Nachlauf eingestellt. Pegel- und Laufzeitanpassungen der aufgenommenen unterschiedlichen stereofonen Signale wurden durchgeführt. Das Sprachmaterial, das am geeignetsten erschien, wurde jeweils für einen männlichen und einer weiblichen Sprecherin ausgewählt. Diese Sätze wurden auf ihren Effektivwert normiert, um sie für die folgenden Sprachverständlichkeitstests verwenden zu können. Hierfür wurden mit Hilfe eines Voice Activity Detector (VAD) die geräuschlosen Pausen, definiert bei -70 dBFS entfernt. Für die extrahierten Sätze wurde der Effektivwert aus den Sample S mit Formel 3.2 berechnet, wobei N für die Anzahl der Sample steht.

$$S_eff = \frac{\sqrt{\sum S(i)^2}}{\sqrt{N}} \tag{3.2}$$

Zur Normierung des digitalen Sprachmaterials S_norm auf -26 dB wurde die Formel 3.3 angewandt. Aus dem normierten Sprachmaterial wurde das stationäre Rauschen, ebenfalls normiert auf -26 dB für die Testdurchführung erzeugt. Zur Generierung des Rauschens wurden die 288 Sätze des ausgewählten Sprechers jeweils mit zufällig ausgewähltem Startpunkt und Cosinus-Fensterung hintereinandergelegt bis für jeden Satz 20 Sekunden Signaldatei entstanden. Diese 288 Dateien wurden nun 20 mal überlagert, um ein stationäres, sprachsimulierendes Rauschen zu erhalten und ohne die unbesprochenen Start- und Endlängen, zu einem pausenlosen Sprachstück aneinandergehängt. Davon wurden Abschnitte von 20 Sekunden durch Fensterung randomisiert entnommen. Durch diese Methode konnten die Enden des Abschnitts sanft ausgeblendet werden, so dass ein abrupter Übergang und damit ein „Klick“-Geräusch verhindert werden konnte. Das Sprachmaterial aller entnommenen 20 s -Abschnitte wurde aufsummiert und gemittelt. Nach 10-maliger Durchführung dieser Prozedur und Mittelung der Abschnitte, entstand so ein Sample von 20 Sekunden Rauschen mit dem Langzeitspektrum des Sprechers, dem sogenannten sprachgeformten Rauschen.

$$S_norm = S \cdot 10^{\frac{-26\ \mathrm{dB} - S_eff}{20\ \mathrm{dB}}} \tag{3.3}$$

Für die Aufbauten und die Bearbeitung des Sprachmaterials wurden aus einem parallel laufenden Projekt heraus 2 Kollegen der T-Systems beauftragt; die Erstellung des Rauschens wurde von einem Kollegen der Telekom Laboratories vorgenommen.

3.3 Testaufbau und allgemeiner Ablauf

Die Testpersonen wurden vor dem Sprachverständlichkeitstest auf ihr Gehör mit einer 10-minütigen Audiometrie geprüft. Hierfür wurden 11 Töne von 125 Hz bis 8 kHz in Oktav-Bändern schrittweise für 30 Sekunden für jeweils ein Ohr vorgespielt. Bei mindestens zweifacher Überschreitung der Grenze von 15 dB HL wurde die betreffende Testperson von der Studie ausgeschlossen. Die Testpersonen wurden für ihren Aufwand entschädigt.

Der Sprachverständlichkeitstest selbst wurde mit Kopfhörern der Sorte „ProNew“ der Firma Stax und dessen darauf abgestimmter SRM-Vorverstärker durchgeführt, der zuvor auf den Pegel von 70 dB kalibriert worden ist. Mit der Software „Adobe Audition“ (Versio 3.0) wurden die Hörproben über die externe Soundkarte „Multiface 2“ des „Hammerfall DSP Systems“ der Firma RME verstärkt. Über eine PCMCIA-Karte Typ II und einem FireWire-Kabel wurde der LapTop, auf dem das Testprogramm lief mit der Karte verbunden. Über eine Patchbay wurden die Kopfhörer in der schallisolierenden Sprachkabine, in der die Testpersonen den Test durchführen sollten mit dem Multiface angeschlossen. Tastatur und Mouse sowie der Ausgabebildschirm, die sich in der Sprachkabine befanden, wurden ebenfalls über diese Patchbay mit dem Laptop verbunden.

Die vorgespielten Hörproben sollten von den Testpersonen vollständig angehört und dann eingegeben werden. Bei Wiedergabe der erkannten Schlüsselwörter musste die Reihenfolge beachtet werden [24]. Die Eingabe erfolgte über die Tastatur, wodurch

sie automatisch auf Homophone und Fehler geprüft werden konnte. Dadurch war es möglich für den ersten Test der adaptiven Pegelanpassung, aus der Fehleranzahl den Pegel für die darauffolgende Hörprobe sofort zu berechnen. Mit dieser vollautomatisierten Testdurchführung konnten die eingegebenen Wörter schnell und korrekt auf Homophone und Rechtschreibfehler überprüft werden. Das automatisierte Testprogramm war ausgehend von [24] für das Deutsche in einer eigenen Vorarbeit zu dieser Masterarbeit angepasst worden.

Rechtschreibfehler wurden mit dem Befehl „grep“ in dem Programm „SCLite“ als eine Unix-Anwendung mit Hilfe des Portier-Programm „Cygwin“ für eine Unix-Umgebung, geprüft, siehe Anhang A.3. Sämtliche im Test verwendeten Sätze waren in einer Liste mit ihren Schlüsselwörtern in extra Spalten gespeichert. Aus dieser Liste wurden die entsprechenden Schlüsselwörter ausgelesen und als txt-Datei (refD) sowie die eingegebenen Wörter der Testperson (hypD) gespeichert. Diese konnten so dem Programm SCLite übergeben werden. Das Ergebnis der Rechtschreibüberprüfung wurde in einer txt-Datei abgespeichert und konnte wieder in „MATLAB“ eingelesen werden.

Für die Homophonüberprüfung wurde zuvor die phonemische Beschreibung der Schlüsselwörter mit der phonemischen Beschreibung aller deutschen Wörter aus WebCelex auf Übereinstimmungen verglichen. Trat der Fall der Übereinstimmung auf und waren die dazugehörigen Wörter unterschiedlich, wurde das entsprechende phonemisch gleich klingende Wort aus der deutschen Wortschatzliste bei dem jeweiligen Schlüsselwort in die beschriebenen Liste eingetragen. Neben dem ursprünglichen Schlüsselwort stand nun auch das gleich klingende, semantisch unterschiedliche Wort, wie z.B. für das eigentliche Schlüsselwort „Meer“ das Homophon „mehr“ eingetragen worden ist. Die Eingaben der Testpersonen wurden dann neben der Schlüsselwörter des Satzes auch auf diese Homophone hin überprüft, so dass im Falle eines korrekten Verstehens, aber einer falschen Interpretation, dies nicht als Fehler gewertet wurde.

Fehler wurden der Testperson in Form einer einmaligen unkonkreten Fehlermeldung mitgeteilt. Somit musste der eingegebene Satz im Ganzen von der Testperson geprüft

werden, ohne dass suggestive Hinweise auf die Art oder die Position der Fehler gegeben wurden. Unterscheidungen zwischen Groß- und Kleinschreibung wurde nicht gemacht, so dass dies nicht als Fehler gewertet wurde. Wegen bestimmter Satzstrukturen, steht der Artikel nicht immer bei seinem Nomen, so wurde nur das Nomen ohne Artikel als korrekt verstanden gewertet.

4 Praktischer Teil

4.1 Test zur Ermittlung des SRT_mask

4.1.1 Testprinzip

In dem ersten Teil der praktischen Durchführung wurde der SRT-Wert für jeden einzelnen Maskierertyp SRT_mask ermittelt. Hierfür wurde vorab der Pegel des Zielsprechers von der Testperson selbst nach der individuellen Einschätzung von 50 % des Verstandenen eingestellt. Es wurde dadurch lediglich die Lautstärke des Zielsprechers verändert, die Lautstärke des Rauschens bzw. Störsprechers blieb fix. Dabei wurde beginnend mit dem SNR-, bzw. TMR-Wert von -30 dB ein Satzpaar mit dem jeweiligen Maskierertyp akustisch dargeboten. Wenn die Testperson eine bestimmte Taste bestätigte, wurde der Zielsprecher dann um 3 dB erhöht, erneut vorgespielt. Die Testperson minderte dadurch solange den SNR-, TMR-Wert bis sie das Gefühl hatte 2 der 4 Schlüsselwörter des lauter gewordenen Zielsprechers rauszuhören. Der so ermittelte Pegel wurde als Startwert für die adaptive Prozedur der SRT-Ermittlung benutzt.

Hierfür wurde das erste Schlüsselwort des Zielsprechersatzes in geschriebener Form vom Programm vorgegeben, um der Testperson erkenntlich zu machen, welcher der beiden vorgespielten Sprecher der Zielsprecher war. Die restlichen 3 Schlüsselwörter sollten von der Testperson verstanden und eingegeben werden. Die Anzahl der korrekt eingegebenen Wörter wurde vom Programm gezählt und darauf basierend wurde der neue Pegel des Zielsprechersatzes der darauffolgenden Hörprobe automatisch ermittelt. Der neue Pegel wurde nach der Formel 4.1 adaptiv berechnet. Es wurden allerdings pro Pegeländerung 2 aufeinanderfolgende Sätze verwendet, so dass maximal 6 Fehler erreicht werden konnten. Der Pegel wurde daher erst nach jedem zweiten dargebotenem

Satzpaar um den berechneten dB-Wert geändert. Durch den verhältnismäßig großen Fehlerbereich vgl. Benoit et al. 1996 mit 4 Fehlern [1], ließ sich der zu berechnende Pegel feiner abstimmen.

$$Pegel = Pegel + ((Fehleranzahl - 3) \cdot 2) \tag{4.1}$$

Zum testen jedes Maskierertyps wurden 2 Testlisten mit je 12 Satzpaaren verwendet. Von den damit zur Verfügung stehenden 24 Satzpaaren wurde das erste Satzpaar zur Justierung des Pegels des Zielsprechers eingesetzt. Und erst bei jedem zweiten Satzpaar wurde der Pegel erneut berechnet, so dass 11 SNR-, TMR-Werte für jeden Maskierertyp pro Testperson ermittelt werden konnten. Zur Berechnung des SRT_mask -Werts wurde dann die jeweils letzten 8 der 11 Ergebnisse gemittelt. Es wurde davon ausgegangen, dass für ein Einpendelverhalten ca. 3 bzw. hier 6 Satzpaare benötigt werden [24].

Die Sätze des Zielsprechers wurden entsprechend der Reihenfolge der Testlisten phonemisch aubalanciert zur SRT-Ermittlung dargeboten. Die Sätze des Störsprechers entsprechen keiner Testlistenreihenfolge und wurden entsprechend der jeweiligen Satzstruktur passend zu den Sätzen des Zielprechers ausgewählt. Die Störsprechersätze durften allerdings nicht wesentlich länger als die Zielsprechersätze sein, um zu gewährleisten, dass sich die Wörter beider Sätze weitestgehend zeitlich überlagern. Es wurde daher ein zeitlicher Rahmen von 0 bis 0.2 Sekunden festgelegt in dem die Zeitdifferenz der Ziel- und Störsprechersätze schwanken durfte. Zudem wurden 2 verschiedene Störsprechersätze jeweils einem Zielsprechersatz zugeordnet. Jeweils einer Hälfte der Testpersonen wurde die eine bzw. die andere Kombination aus Ziel- und Störsprechersatz dargeboten. Dadurch sollten eventuell auftretende Effekte, die durch ungünstige Kombination (sinnhafte Verbindung) zweier Sätze entstehen konnten, minimiert werden.

Die so erstellten Testlisten mit ihren Satzpaarkombinationen wurden in unterschiedlicher Reihenfolge für alle beschriebenen Maskierertypen den Testpersonen dargeboten.

Dadurch war das Sprachmaterial zum einen nicht an den Maskierertyp und zum anderen nicht an die Darbietungsreihenfolge gebunden. Damit sollte vermieden werden, dass durch Effekte abhängig der Position im Test oder durch zufällig auffällige Satzkombinationen auf die Resultate der Maskierertypen geschlossen werden konnte.

Die Reihenfolge der Darbietung der 5 bereits beschriebenen Maskierertypen (Rauschen, gleiche männliche Stimme, um 3 bzw. 6 Halbtöne erhöhte Stimme und weibliche Stimme) war über alle Testpersonen gleich verteilt, d.h. dass die Positionen der Maskierertypen im Test für jede Testperson unterschiedlich, aber insgesamt gleich oft von jeweils einem Maskierertyp vertreten worden war. Würde beispielsweise ein bestimmter Maskierertyp bei allen Testpersonen stets am Ende des Tests auftreten, könnte es sein, dass durch die zunehmende trainierte Aufnahmefähigkeit der Testpersonen über den Testverlauf hinweg für diesen Maskierertyp bessere Resultate erzielt werden können. Es wurde daher für die 5 Maskierertypen manuell eine 5x5 Matrix erstellt, die die gleichmäßige Verteilung der Maskierertypen jeder Position im gesamten Testablauf zuordnet. Allerdings hätten somit nur 5 Testpersonen eingesetzt werden können. Es wurden somit 2 der 5x5 Matrizen erstellt, die aneinandergefügt worden sind, siehe Tabelle 4.1. Dadurch konnte eine Testpersonenanzahl von wenigstens 10 ermöglicht werden.

Testperson Nr.	1	2	3	4	5		6	7	8	9	10
1. Position im Test	M1	M2	M3	M4	M5		M1	M2	M3	M4	M5
2. Position im Test	M2	M3	M4	M5	M1		M2	M3	M4	M5	M1
3. Position im Test	M3	M4	M5	M1	M2		M3	M4	M5	M1	M2
4. Position im Test	M4	M5	M1	M2	M3		M4	M5	M1	M2	M3
5. Position im Test	M5	M1	M2	M3	M4		M5	M1	M2	M3	M4

Tabelle 4.1: zwei 5x5-Matrizen zur Gleichverteilung der 5 Maskierertypen mit einfacherer Benennung: M1, M2, M3, M4, M5

4.1.2 Testablauf

Wie beschrieben wurden pro Maskierertyp 24 Satzpaare benötigt, so dass für 5 Maskierertypen insgesamt 216 Sätze pro Testperson verwendet wurden. Es verblieben damit 72 Sätze, bzw. 36 Satzpaare der insgesamt 288 vorhandenen Sätze des Sprachmateri-

als. Diese restlichen Sätze wurden für die Trainingsphase verwendet. Dadurch sollten die Testpersonen vorab auditiv auf die Höraufgaben vorbereitet werden, woruch die gemessenen Werte im Verlauf des Tests relativ konstant bleiben und wenige Schwankungen aufweisen sollten, vgl. Kapitel **??**.

Zu Anfang des Tests wurden 2 Trainingsphasen von 9 Sätzen und 30 Satzpaaren dargeboten. In der ersten Trainingsphase wurden den Testpersonen 9 Sätze des Zielsprechers ohne Degration vorgespielt, die eingegeben werden sollten. Mit dieser relativ einfachen Aufgabe sollte die Testperson mit der Eingabemethodik, der Art des Satzmaterials und der Stimme des Zielsprechers vertraut gemacht werden. Der Pegel des Sprechers blieb hierbei konstant. In der zweiten Trainingsphase wurden alle Maskierertypen dargeboten und den Testpersonen in folgender Reihenfolge vertraut gemacht: Rauschen, gleiche Stimme, um 3 Halbtöne erhöhte Stimme, um 6 Halbtöne erhöhte Stimme und weibliche Stimme. Der Zielsprecher wurde in 6 Satzpaaren für jeweils einen Maskierertyp für 6 unterschiedliche Pegel nach Formel 4.2 vorgespielt. Die Satzpaare der Trainingsphasen waren für alle Testpersonen gleich, um die Wahrscheinlichkeit zu erhöhen, dass sie alle gleich stark trainiert worden waren.

$$TMR = SRT_mask + [9, 3, 0, -3, -6, -9\,]dB \tag{4.2}$$

Die Testpersonen wurden mit dem Testablauf in Form einer Beschreibung, die unter anderem Programmbilder enthielt vertraut gemacht, siehe Anhang A.4.1. Darüber hinaus wurden den Testpersonen vor jedem Abschnitt im Test mitgeteilt, um welchen Maskierertyp es sich im Folgenden handelte. Dadurch sollten die Testpersonen auf ihre Hörproben vorbereitet werden. Für die Dauer der Bearbeitung einer Hörprobe (Hören, Eingeben, Kontrollieren) wurde von 30 Sekunden ausgegangen, so dass sich die voraussichtliche Gesamtdauer des Tests auf eine Stunde und 15 Minuten belaufen sollte. Vor jedem ersten Testdurchlauf jeder Testperson wurde mittels der Audiometrie das Gehör der Testpersonen getestet, so dass sich die Testdauer um zusätzliche 10 Minuten erhöhte. Zudem wurde eine feste Pause von 5 Minuten nach der Trainingsphase und weiteren 48 Hördarbietungen eingeführt.

4.1.3 Ergebnisse

Der Test lieferte den SRT_mask -Wert für alle getesteten Maskierertypen über alle Testpersonen gemittelt, deren Werte aus Tabelle 4.2 zu entnehmen sind. Jeder aufgetragene Wertepunkt besteht aus 80 Ergebnissen.

Rauschen	Gl. Stimme	3 Halbtöne erhöht	6 Halbtöne erhöht	Weibl. Stimme
-1.75 dB SNR	-1.6 dB TMR	-3.75 dB TMR	-9.7 dB TMR	-14 dB TMR

Tabelle 4.2: Übersicht der ermittelten SRT_mask -Werte für alle 5 Maskierertypen

Diese Ergebnisse wurden in Abbildung 4.1 für alle Maskierertypen dargestellt, wodurch sich deutlicher zeigt, dass sich mit zunehmender Ähnlichkeit zwischen Ziel- und Störsprecher der SRT_mask -Wert verringert. Damit ist ein Unterschied des SRT_mask -Werts für den Maskierertyp weibliche Stimme und gleiche Stimme von 12.4 dB TMR aufzuweisen. Eine Erhöhung der Grundfrequenz der gleichen Stimme auf um 3 Halbtöne bewirkt hier eine Minderung des SRT_mask -Werts um das Doppelte auf -2,15 dB TMR. Bei einer weiteren Erhöhung um 3 Halbtöne (insgesamt 6 Halbtöne) verringert sich der SRT_mask -Wert um das Vierfache und liegt bei -8,1 dB TMR.

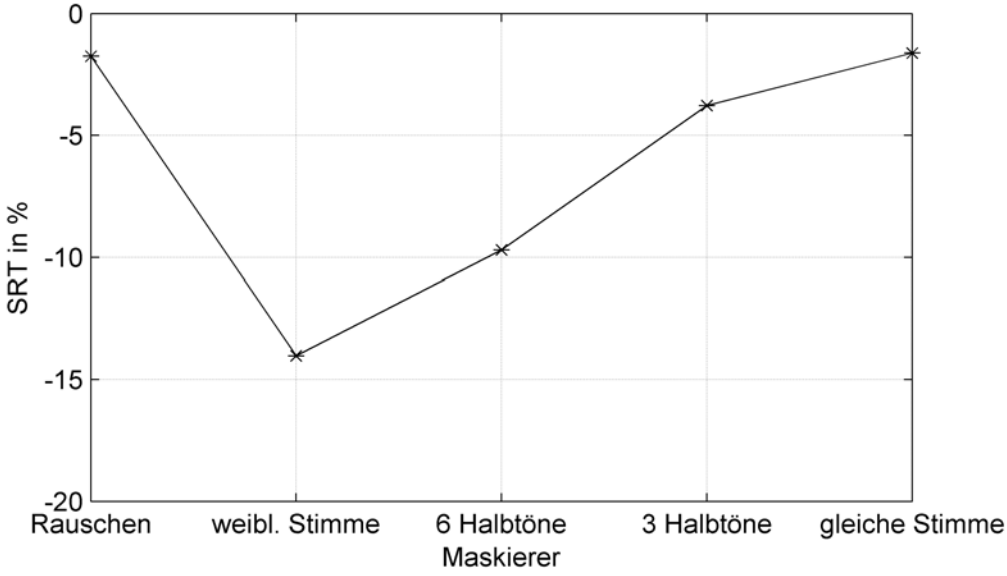


Abbildung 4.1: gemittelte SRT-Werte über alle Testpersonen für alle Maskierertypen

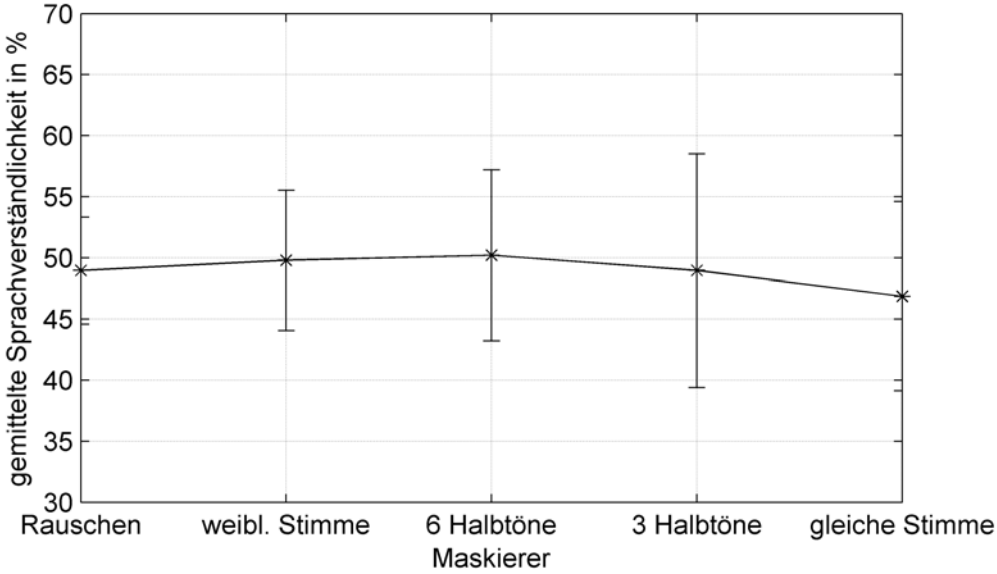


Abbildung 4.2: gemittelte Sprachverständlichkeit über alle Testpersonen für jeden Maskierertyp

Bei Betrachtung der über alle Testpersonen gemittelten Sprachverständlichkeit für die bewerteten TMR-Pegel zur SRT-Berechnung, siehe Abbildung 4.2 fällt auf, dass die Werte der Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme, entgegen des angestrebten 50 % -Schwellwerts um 1 % bis 4 % niedriger sind. Die adaptive

Methode sollte den TMR-Wert bei einer Sprachverständlichkeit von 50 % ermitteln. Kann diese nicht erreicht werden, kann der dafür ermittelte TMR-Wert nicht alls Schwellwert (SRT) angenommen werden.

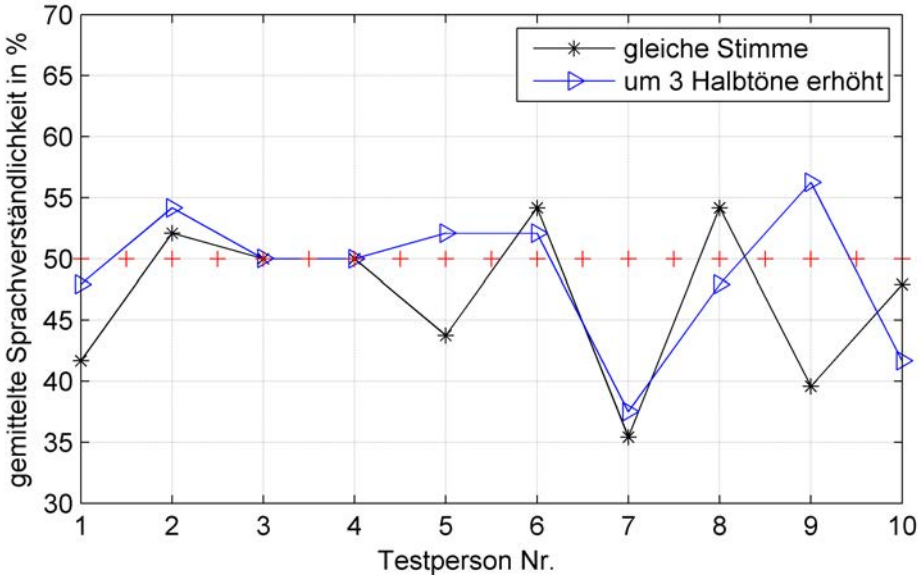


Abbildung 4.3: gemittelte Sprachverständlichkeit für alle Testpersonen für Maskierertyp gleiche Stimme und um 3 Halbtöne erhöht

Bei genauerer Untersuchung der Sprachverständlichkeit für die Ergebnisse der einzelnen Testpersonen für die beiden Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme zeigt sich, wie der Abbildung 4.3 zu entnehmen ist, dass die Testpersonen unterschiedlich stark von dem 50 % -Schwellwert abweichen. Vor allem die Testpersonen 1, 5, 7 und 9 weisen mit zwischen 46 % bis 35 % der Sprachverständlichkeit höhere Abweichungen bei dem Maskierertyp gleiche Stimme auf. Der Abbildung 4.4 kann die adaptive Pegelanpassung abhängig der Fehleranzahl für die aufgezählten Testpersonen entnommen werden. Aufgrund der sich auf die Fehleranzahl anpassenden Pegelwerte sollte sich die Sprachverständlichkeit mit jeder weiteren Berechnung der 50 7% nähern. Allerdings zeigen alle Verläufe stark schwankende Werte für die Sprachverständlichkeit auf und es ist kein Einpendelverhalten zu erkennen. Der Endwert weicht deutlich von dem Startwert ab.

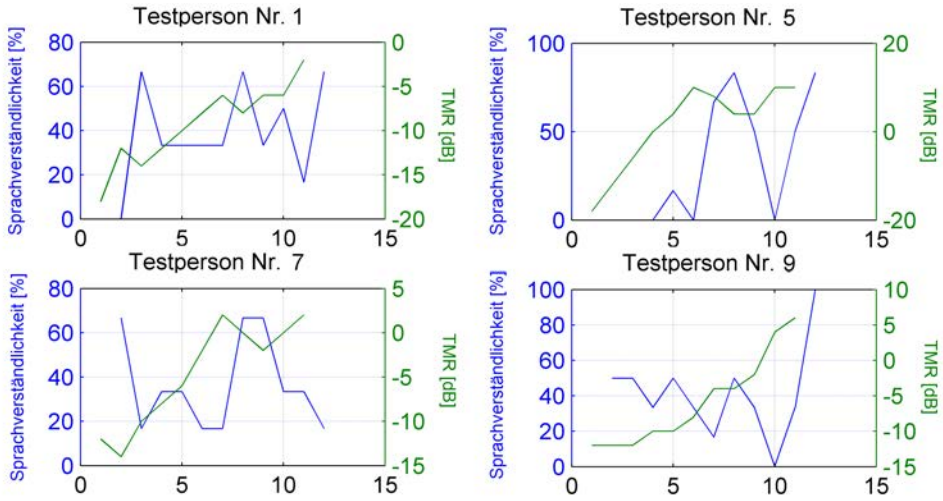


Abbildung 4.4: adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und den dafür errechneten Pegel (rechte y-Achse in grün) für den Ströspreher der gleichen Stimme

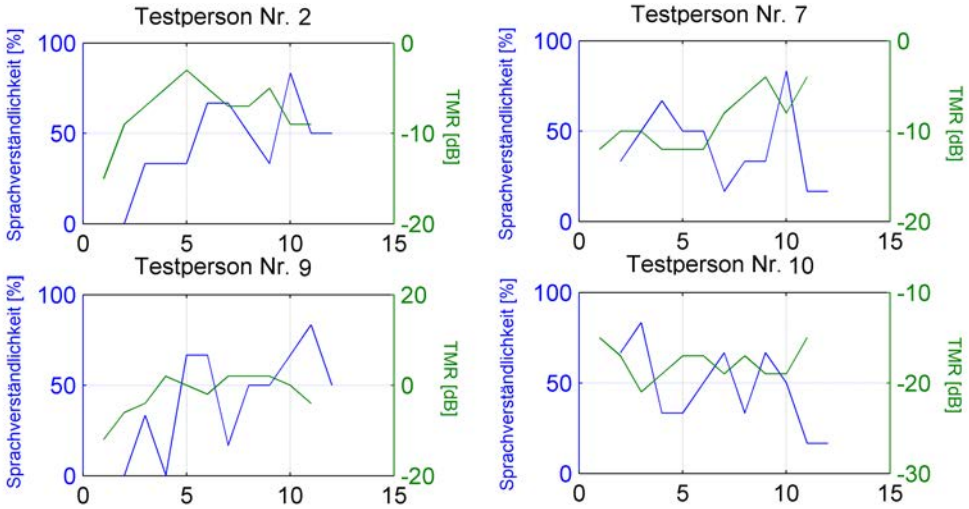


Abbildung 4.5: adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und des dafür errechneten Pegels (rechte y-Achse in grün) für den Störsprecher der gleichen Stimme um 3 Halbtöne erhöht

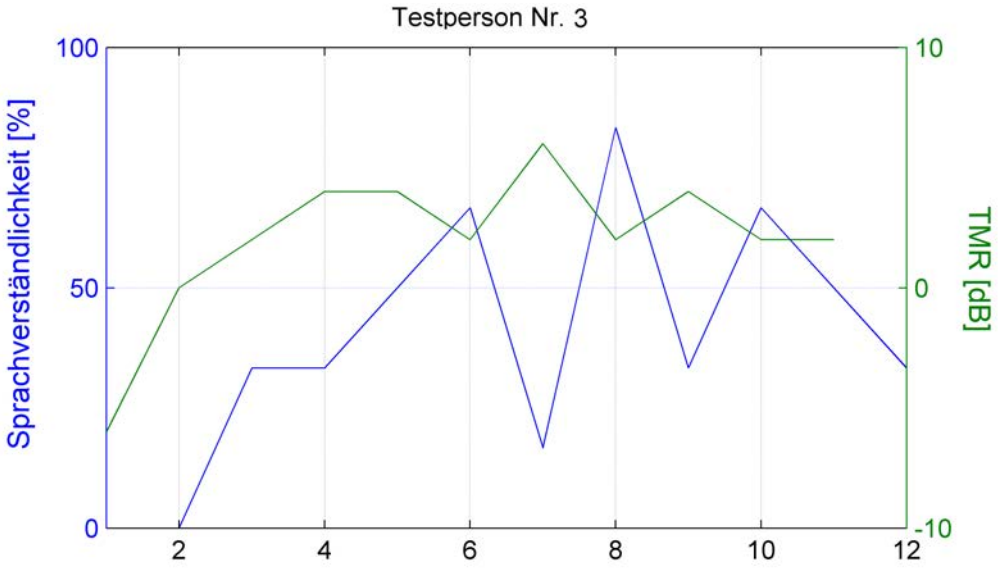


Abbildung 4.6: adaptive Pegelanpassung der Sprachverständlichkeit in% (linke y-Achse in blau) und dem dafür errechneten Pegel (rechte y-Achse in grün) für den Maskierertyp gleiche Stimme der 3.Testperson

Ebenso verhält es sich für den Maskierertyp um 3 Halbtöne erhöhte Stimme, siehe Abbildung 4.3. Hier lässt sich erkennen, dass die Testpersonen 2, 7, 9 und 10 die meiste Abweichungen von den 50 % Sprachverständlichkeit mit 4 bis 12 % aufweisen. Diese zeigen ebenso, wie sich der Abbildung 4.5 entnehmen lässt, einen unregelmäßigen und schwankenden Verlauf der Sprachverständlichkeit. Dieses gilt besonders für die unregelmäßigen Verläufe der Testpersonen 7 und 9, die keine Annäherung an die 50 % -Sprachverständlichkeit aufweisen. Zum Vergleich ist ein typisches Einpendelverhalten in Abbildung 4.6 aufgetragen.

Es zeigt sich, dass bei den aufgeführten Testpersonen 1, 5, 7 und 9 aus Abbildung 4.4 zwischen Start- und Endwert der berechneten Pegel, Unterschiede für 10 bis 15 dB TMR für den Maskierertyp gleiche Stimme bestehen. Für den Maskierertyp um 3 Halbtöne erhöhte Stimme, zeigen die Testpersonen 7 und 9 ebenfalls die höchsten Abweichungen mit 6 bis 8 dB TMR, siehe Abbildung 4.7. Um diese hohen Unterschiede der berechneten Pegel zu minimieren und damit der hohen Schwankung der Sprachverständlichkeit Rechnung zu tragen, wurden nicht die 8 letzten Werte der 11 gesamten Ergebnisse pro Testperson zur Berechnung des SRT-Wertes verwendet, sondern nur die

letzten 5.

Mit der neuen Berechnung der gemittelten Sprachverständlichkeit pro Testperson, kann eine größere Annäherung an den Schwellwert der Sprachverständlichkeit erreicht werden, siehe Abbildung 4.8. Es konnte somit für die beiden Maskierertypen eine Minimierung der Abweichung von den 50 % -Sprachverständlichkeit um knapp 4 % zu den verhergehenden Ergebnissen erreicht werden (durchgehende Kreuze). Die daraus resultierende Neuberechnung der SRT_mask-Werte führte zu leicht veränderten Ergebnissen als die anfangs Berechneten, siehe Tabelle 4.3 und wurden im darauffolgenden Test eingesetzt.

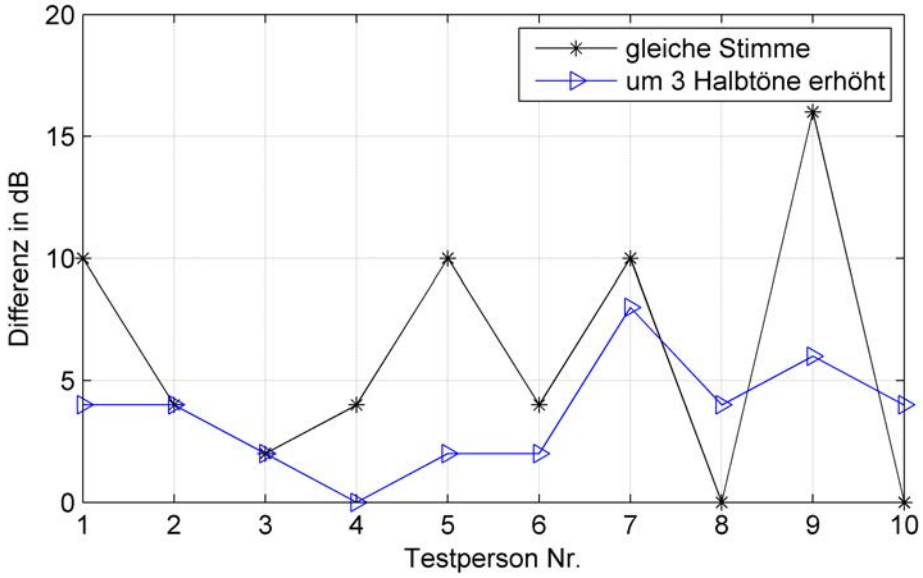


Abbildung 4.7: Differenz der Start- und Endwerte der berechneten Pegel

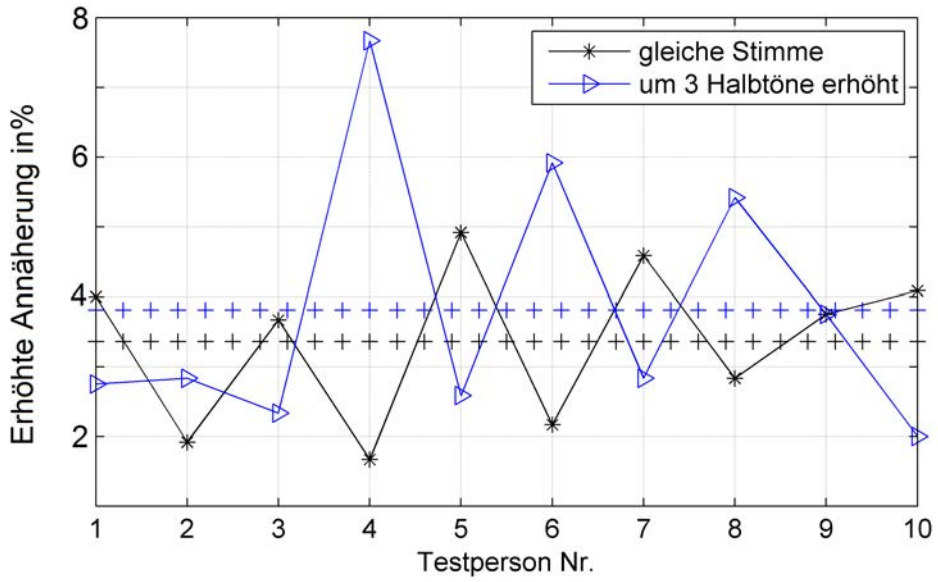


Abbildung 4.8: Minimierung der Abweichung zu 50 % -Sprachverständlichkeit mit den 5 letzten Werten pro Testperson pro Störsprecher der gleichen Stimme und um 3 Halbtöne erhöht

Rauschen	Gl. Stimme	3 Halbtöne erhöht	6 Halbtöne erhöht	Weibl. Stimme
-2 dB SNR	-1.6 dB TMR	-3.8 dB TMR	-9.6 dB TMR	-13.8 dB TMR

Tabelle 4.3: Übersicht der neu ermittelten SRT_mask -Werte für alle Maskierertypen

4.2 Test zur Ermittlung der Sprachverständlichkeit mit vorgegebenen Pegeln

4.2.1 Testprinzip

Nachdem im vorangegangenen Test für das verwendete Testmaterial der SRT_mask -Wert für eine Sprachverständlichkeit von 50 % ermittelt worden war, wurde in dem folgenden Test das Verhalten der Sprachverständlichkeit in Abhängigkeit der SNR-, bzw. TMR-Werte um den SRT_mask-Wert des jeweiligen Maskiertyps gemessen. Hierbei wurde nur der Lautstärkepegel des Zielsprechers verändert, der Pegel des Störsprechers blieb unverändert. Dafür wurden zur feineren Abmessung Schrittgrößen von 2 dB gewählt. Mit bis zu 12 dB Pegeländerung sollte ein deutlicher Verlauf der Sprachverständlichkeit in Abhängigkeit der Pegelunterschiede aufgezeigt werden können. Es

wurden für jeden einzelnen Maskiertyp die Sprachverständlichkeit mit Pegelunterschieden zwischen Zielsprecher und Maskierer nach Formel 4.3 gemessen. Die Pegel wurden im Test randomisiert angeordnet dargeboten, allerdings durften identische Pegel nicht direkt aufeinander folgen, um eine kurzzeitige Anpassung des Gehörs an den Pegel und damit einen Trainingseffekt zu vermeiden. Den Pegeln wurden somit unterschiedliche Satzpaare für jede Testperson zugeordnet.

$$TMR = SRT_mask + [-6, -4, -2, 0, 2, 4, 6\,]dB \tag{4.3}$$

Es wurden alle 5 Maskierertypen beibehalten, so dass die Sprachverständlichkeit der Maskierertypen Rauschen, gleiche Stimme, um 3 Halbtöne erhöhte Stimme, um 6 Halbtöne erhöhte Stimme und weibliche Stimme ermittelt wurde. Die Anordnung der Maskierertypen wurde über den gesamten Testverlauf über alle Testpersonen, ebenfalls gleich verteilt dargeboten. Hierfür wurde mittels eines Algorithmus die Verteilung mit Hilfe einer 4x4 Matrix nach Wagenaar 1969 [28] erstellt. Der Algorithmus erzeugt Matrizen mit gleicher Spalten- und Zeilenanzahl für eine gerade Anzahl. Bei 5 Maskierertypen konnte dieser Algorithmus nicht direkt angewandt werden. Es wurden daher drei 4x4 Matrizen für 4 Maskierertypen erzeugt und der letzte verbleibende Maskierertyp, hier Rauschen, alternierend eingefügt, siehe Tabelle 4.4. Die mit „X“ gekennzeichneten Felder beschreiben die Maskierertypen, die nach dem Algorithmus verteilt worden sind. Die Nummer hinter dem „X“ beschreibt die Verteilung der unterschiedlichen 4 Maskierertypen. Die „0“ steht für den einen Maskierertyp, der nachträglich eingefügt worden ist. Auf die Art konnten 12 Testpersonen für diesen Versuch eingesetzt werden.

Testperson Nr.	1	2	3	4	5	6	7	8	9	10	11	12
1. Position im Test	0	X2	X3	X4	X1	0	X3	X4	X1	X2	0	X4
2. Position im Test	X1	0	X4	X1	X2	X2	0	X1	X2	X3	X2	0
3. Position im Test	X2	X3	0	X2	X3	X3	X4	0	X3	X4	X4	X1
4. Position im Test	X3	X4	X1	0	X4	X4	X1	X2	0	X1	X1	X2
5. Position im Test	X4	X1	X2	X3	0	X1	X2	X3	X4	0	X2	X3

Tabelle 4.4: Zuordnung der Maskierertypen zur Gleichverteilung

Die Sätze des Zielsprechers wurden zufällig angeordnet dargeboten, sie entsprachen keiner Testlistenreihenfolge mehr. Allerdings gehörten die maskierenden Sätze des Störsprechers weiterhin derselben Satzstruktur an, ebenso wurden die Störsprechersätze, die in dem Bereich von 0 bis 0.2 Sekunden länger als die Zielsprechersätze waren, gewählt. Auch wurden 2 verschiedene Störsprechersätze jeweils einem Zielsprechersatz zugeordnet. Jeweils einer Hälfte der Testpersonen wurde die eine bzw. die andere Kombination aus Ziel- und Störsprechersatz dargeboten.

Wie im vorangegangenen Test wurde das erste Schlüsselwort des Satzes des Zielsprechers vorgegeben. Wegen der abwechselnden Darbietung der Pegel eines jeden Satzes, wurde die Fehleranzahl hier für jeden Satz einzeln ermittelt, so dass 3 Schlüsselwörter zur Berechnungsgrundlage dienten.

4.2.2 Testablauf

Aus Tabelle 4.4 geht hervor, dass aufgrund der gleich verteilten Anordnung der 5 Maskierertypen 12 Testpersonen erforderlich waren. Bei allen Testpersonen sollte pro Maskierertyp jeder Pegel jeweils 3 mal dargeboten werden. Damit wurden 105 Ziel- und 84 Störsprechersätze für jede Testperson benötigt. Von den 288 Sätzen verblieben damit 99 Sätze für die Trainingsphase.

Alle Satzpaare eines Maskierertyps wurden wie im vorangegangenem Test zusammenhängend nacheinander getestet. Vor den 21 Satzpaaren pro Maskierertyp wurde jeweils eine Trainingsphase bestehend aus 8 Satzpaaren durchgeführt. Damit bestand der Test aus 5 Teilabschnitten (bei 5 Maskierertypen) zusammengesetzt aus jeweils einer Trainings- und einer Testphase. Die Trainingsphase trainierte jeden SNR-, bzw. TMR-Wert, der in der Testphase getestet wurde, jeweils einmal nach Formel 4.3. Zusätzlich wurde der Zielsprecher als erstes ohne Maskierer dargeboten, um den Hörer an die Stimme zu gewöhnen. Die Sätze aller Trainingsphasen pro Testperson waren unterschiedlich, damit der Trainingseffekt für das Satzmaterial minimiert werden konnte. Allerdings waren die Satzpaare der Trainingsphasen für alle Testpersonen gleich, um

die Wahrscheinlichkeit zu erhöhen, dass sie gleich stark trainiert worden sind.

Die Testpersonen wurden mit dem Testablauf in Form einer Beschreibung, die unter anderem Programmbilder enthielt vertraut gemacht, siehe Anhang **??**. Darüber hinaus wurde ihnen vor jedem Abschnitt im Test mitgeteilt, um welchen Maskierertyp es sich im Folgenden handeln würde. Dadurch sollten die Testpersonen auf ihre Hörproben vorbereitet werden. Für die Dauer der Bearbeitung einer Hörprobe (Hören, Eingeben, Kontrollieren) wurde von 30 Sekunden ausgegangen, so dass sich die Gesamtdauer des Tests auf eine Stunde und 12 Minuten belaufen sollte. Vor jedem Testdurchlauf wurde mittels einer Audiometrie das Gehör der Testpersonen getestet, so dass sich die Testdauer um zusätzliche 10 Minuten erhöhte. Zudem wurde eine Pause in jedem Testdurchlauf nach dem ersten der fünf Abschnitte (29 Hörproben) und nach weiteren 2 Abschnitten (58 Hörproben) von 5 Minuten eingeführt, um den Grad der Konzentration aufrechtzuerhalten.

4.2.3 Ergebnisse

Auf Basis der Berechnung für die Fehleranzahl von 3 Schlüsselwörtern konnten die folgenden gemittelten Verläufe für die Sprachverständlichkeit der einzelnen Maskierertypen über die Pegeländerung ermittelt werden. Jeder dargestellte Wert besteht aus 36 Ergebnissen. Es zeigte sich hierbei, dass für alle Maskierertypen die Konfidenzintervalle hoch waren und das Signifikanzniveau von 5 % überschritten wurde. Somit können keine genauen Aussagen über die Ergebnisse getroffen werden, lediglich über ihre Verläufe.

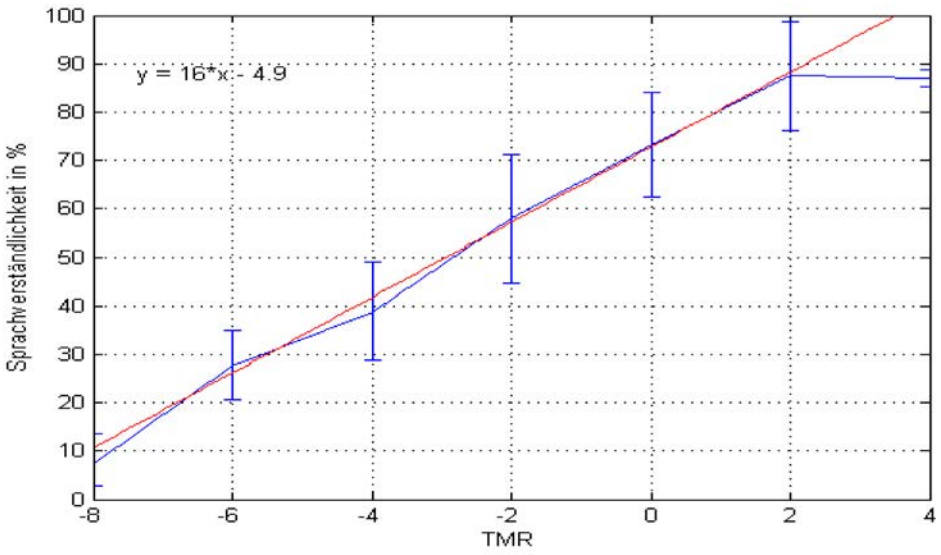


Abbildung 4.9: Sprachverständlichkeit in% für den Maskierertyp Rauschen

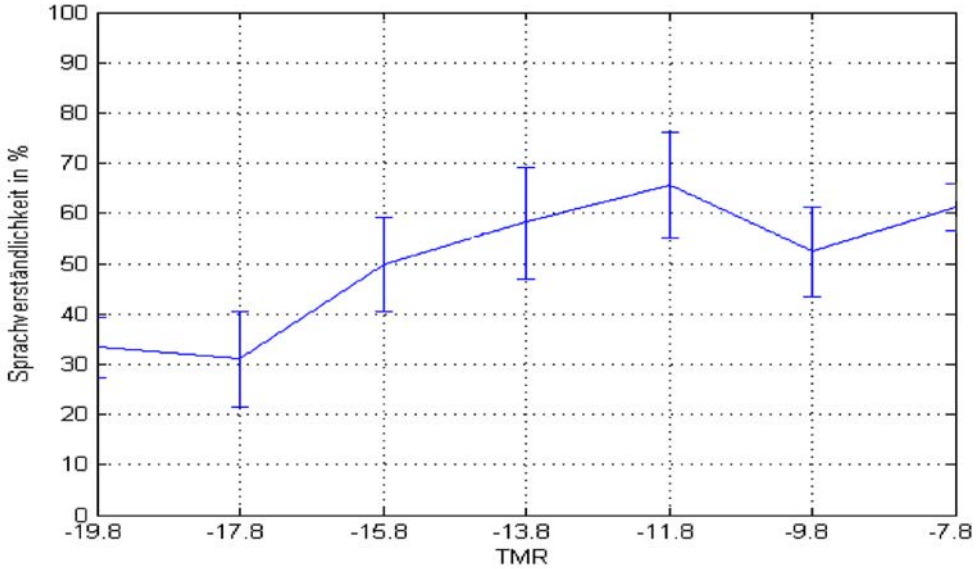


Abbildung 4.10: Sprachverständlichkeit in% für den Maskierertyp weibliche Stimme

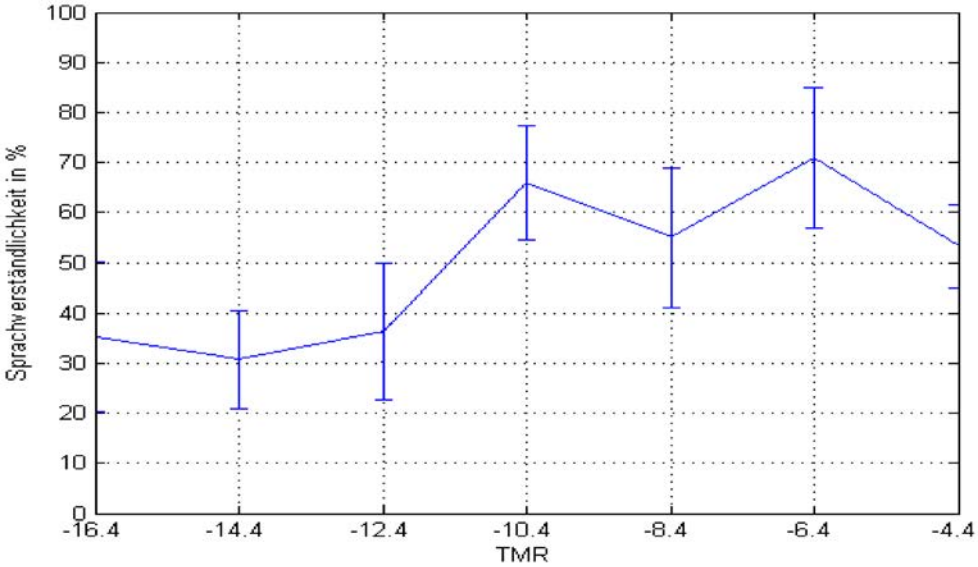


Abbildung 4.11: Sprachverständlichkeit in % für den Maskierertyp um 6 Halbtöne erhöhte Stimme

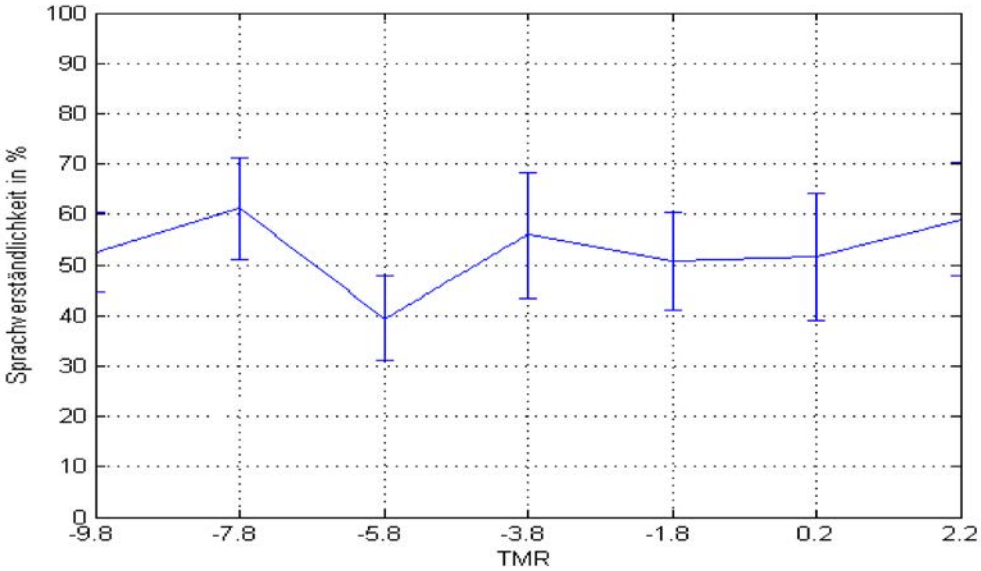


Abbildung 4.12: Sprachverständlichkeit in% für den Maskierertyp um 3 Halbtöne erhöhte Stimme

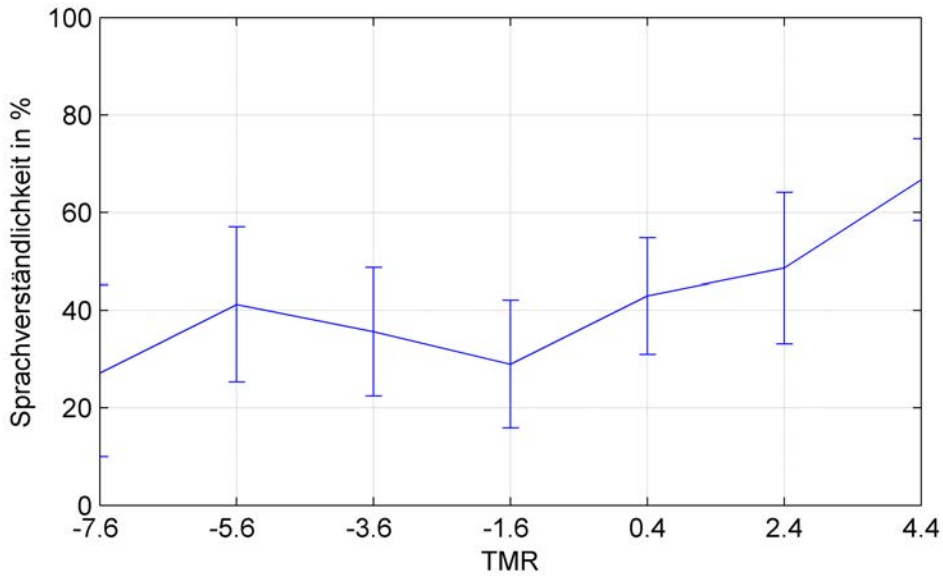


Abbildung 4.13: Sprachverständlichkeit in% für den Maskierertyp gleiche Stimme

Der Schwellwert der Sprachverständlichkeit bei 50 % für Rauschen als Maskierer liegt mit -3 dB SNR etwas unter dem ermittelten SRT_mask -Wert. Der hier aufgeführte Wert entspricht eher dem in der Literatur beschriebenem SRT_mask -Wert bei ca. -4 dB TMR für die französische Sprache [24]. Die Steigung der Diskriminierungsfunktion liegt bei 14 %, was ebenfalls durch die Literatur [24] mit 15 % bestätigt wurde. Die Testmethode scheint somit vergleichbar zu sein und die Ergbenisse zuverlässig.

Die Ergebnisse der Sprachverständlichkeit für den Maskierertyp weibliche Stimme, Abbildung 4.10 zeigen ein Maximum der Sprachverständlichkeit mit 65 % bei -11.8 dB TMR. Das Minimum der Sprachverständlichkeit mit 31 % liegt bei -17.8 dB TMR. Die 50 %-Sprachverständlichkeit liegt auch hier etwas unter dem ermittelten SRT_mask -Wert bei -15.8 dB TMR. Hier hätten offenbar noch niedrigere TMR-Werte getestet werden können.

Ein geringerer TMR-Wert bei 50 %-Sprachverständlichkeit als berechnet, ist für den Maskierertyp um 6 Halbtöne erhöhte Stimme zu erkennen, siehe Abbildung 4.11. Mit ansteigendem TMR kommt es zu steigender Sprachverständlichkeit.

Insgesamt liegt die Sprachverständlichkeit für die getesteten Pegel bei dem Maskierertyp um 3 Halbtöne erhöhte Stimme über 50 %, siehe Abbildung 4.12. Eine auffälige aber nicht signifikante Abweichung der Sprachverständlichkeit liegt bei -5.8 dB TMR.

Bei dem Maskierertyp gleiche Stimme 4.13 ist erst ab -1.6 dB TMR eine relativ konstante Steigung der Sprachverständlichkeit zu erkennen. Bis -1.6 dB TMR liegt die Sprachverständlichkeit unter 50 %. Der Einbruch der Sprachverständlichkeit auf 32 % bei dem SRT_mask -Wert (-1.6 dB TMR) ist mit 18 % Unterschied zum erwateten Sprachverständlichkeit von 50 % auffällig.

Aus Abbildung 4.14 lassen sich die einzelnen Verläufe der gemittelten Sprachverständlichkeit über alle Testpersonen und Pegel für die einzelnen Maskierertypen für die Darbietungsreihenfolge im Testablauf (28 Werte pro dargestelltem Punkt) entnehmen.

Die erste Pause im Test fand gleich nach dem ersten Abschnitt statt. Dadurch könnte sich die Sprachverständlichkeit für den Störsprecher gleicher Stimme um knapp 10 % (Position 2) verbessert haben. Die Werte der anderen Maskierertypen zeigen wenig Änderung. Insgesamt verhalten sich die Werte der Sprachverständlichkeit bis zur zweiten Pause konstant. Danach (Position 4) verbesserte sich die Sprachverständlichkeit für den Maskierertyp um 6 Halbtöne erhöhte Stimme um 15 %. Die Werte für den Maskierertyp weibliche Stimme hingegen sanken um 10 %. Die Sprachverständlichkeit der Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme, sank am Ende der Testdurchläufe auf 40 % ab, die Werte aller anderen Maskierertypen hingegen lagen konstant bei knapp 60 %.

Es ist zu erkennen, dass die Werte für den Maskierertyp um 6 Halbtöne erhöhte Stimme, weibliche Stimme und Rauschen weniger Schwankung aufzeigen als die Werte der Maskierertypen gleichen Stimme und um 3 Halbtöne erhöhte Stimme. Diese zeigen allerdings auch einen Anstieg der Sprachverständlichkeit mit der Darbietungsabfolge zwischen der ersten und vorletzten Position um 24 % des Maskierertyps gleiche Stimme und 24 % des Maskierertyps um 3 Halbtöne erhöhte Stimme.

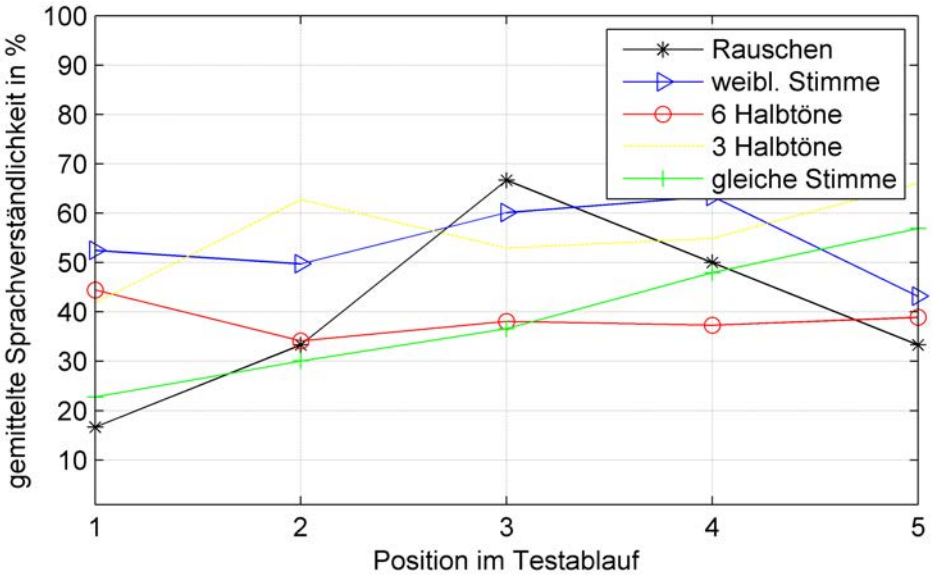


Abbildung 4.14: gemittelte Sprachverständlichkeit in% über alle Testpersonen und Pegel für die einzelnen Maskierertypen entsprechend der Darbietungsreihenfolge

Die Verläufe der Sprachverständlichkeit (mit 3 Werten pro TMR-Wert) der einzelnen Testpersonen über die Pegeländerung für die Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme können der Abbildung 4.15 entnommen werden. Ein regelmäßiger Verlauf, der einen Anstieg mit der Pegelerhöhung des Zielsprechers aufweist, zeigen die Werte der Testpersonen 3, 5, 7, 8 und 10. Die Verläufe der anderen Testpersonen sind unregelmäßig und zeigen aus unerklärlichen Gründen keine eindeutigen Tendenzen mit dem Pegelanstieg. Der bereits beschriebene Einbruch der Sprachverständlichkeit bei dem Maskierertyp der gleichen Stimme zwischen -2 und 2 dB TMR (siehe Abbildung 4.13), kann aus den Ergebnissen der Testpersonen 1, 2, 4, 8, 10 und 11 ebenfalls abgelesen werden.

Die in diesem Test ermittelten Resultate zeigen offenbar Auswirkungen des Trainingseffekts bei den Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme, die 15 % Schwankung der Sprachverständlichkeit ausmachen. Zudem sind unregelmäßige Verläufe der Sprachverständlichkeit dieser Maskierertypen für 5 von 12 Testperso-

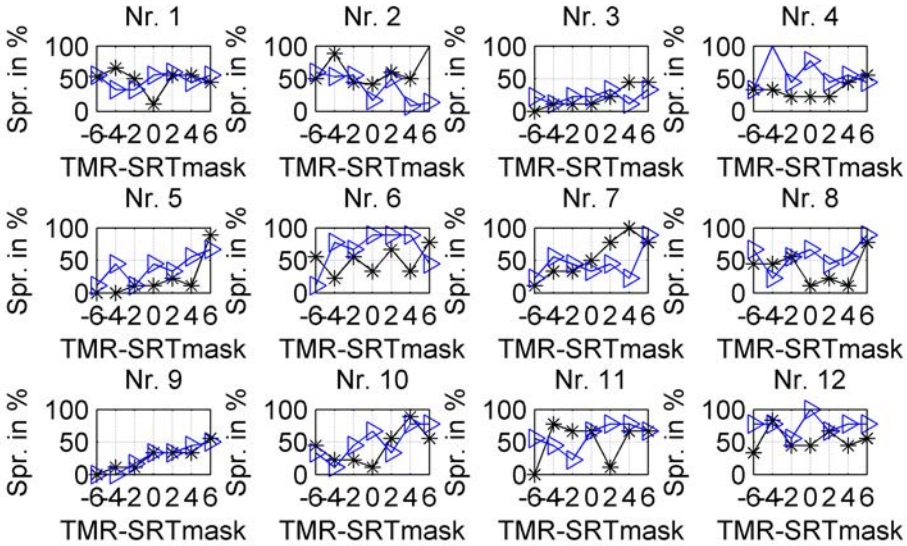


Abbildung 4.15: Verläufe der Sprachverständlichkeit über die Pegeländerung für die Störsprecher der gleichen Stimme (schwarze Sterne) und um 3 Halbtöne erhöht (blaue Dreiecke) aller Testpersonen (einfach numeriert)

nen aufgezeigt worden. Ebenso wurde ein abweichender Verlauf (Einbruch bei -1.6 dB TMR) der Sprachverständlichkeit des Maskierertyps gleiche Stimme gegenüber anderen Maskierertyp deutlich.

4.3 Test zur Ermittlung des Einflusses der sinnhaften Verknüpfung

4.3.1 Testprinzip

Das Testprinzip basiert weitestgehend auf dem Prinzip des vorangegangenen Tests 4.2. Die Satzpaare wurden zu feststehenden TMR-Werten dargeboten. Darüber hinaus wurden sie hinsichtlich der bekannten Kriterien (Satzlänge und Satzstruktur) ausgewählt. Im Unterschied zu den vorherigen Tests wurden Satzpaare in denen einzelne Schlüsselwörter zwischen Ziel- und Störsprechersatz sinnhaft bzw. hoch assoziativ miteinander verknüpft worden sind, dargeboten. Dabei wurden nicht die gleichzeitig auf-

tretenden Schlüsselwörter sinnhaft verknüpft, sondern stets das Schlüsselwort aus dem Zielsprechersatz mit dem zeitlich um eine Position weiter versetzten Schlüsselwort aus dem Störsprechersatz. Es können dafür 3 mögliche Kombinationsformen für ein Sinn ergebendes Wortpaar gebildet werden, siehe Tabelle 4.5.

Kombinationsform	Position 1	Position 2	Position 3
1	A	A	B
2	A	B	A
3	B	A	A

Tabelle 4.5: 3 mögliche Kombinationsformen der Eingabe mit „A“ als Schlüsselwort des Zielsprechersatzes und „B“ als Wort des Störsprechersatzes

Allerdings können aufgrund der Satzstruktur nicht alle Schlüsselwörter hoch assoziativ miteinander verknüpft werden. Aus Abbildung 3.1 in Kapitel 3 geht hervor, dass die Kombinationsmöglichkeit 2 nur für Satzstruktur 2 und 5 möglich ist. Nur die Kombinationsmöglichkeiten 1 und 3 sind für alle Satzstrukturen möglich. Um eine möglichst gleich hohe Anzahl dieser beiden Kombinationsformen zu erhalten bzw. testen zu können, sind daher allein die Kombinationsmöglichkeiten 1 und 3 gewählt worden.

Zur Erstellung möglicher Wortpaare wurden manuell alle im Satzmaterial vorhandenen Schlüsselwörter sinnhaft einander zugeordnet. Aufgrund der Kombinationsformen, die in allen Satzstrukturen ein Wortpaar aus Nomen und Adjektiv oder Nomen und Verb (sowohl Transitives als auch Intransitives) vorsieht, wurden allen Nomen, Sinn ergebende Adjektive und Verben zugeordnet. Mittels dem Programm „MATLAB“ konnten für die Nomen aller Sätze (Zielsprechersätze) entsprechende Sätze gleicher Satzstruktur, die die zugeordneten assoziativen Verknüpfungen nach den oben genannten Kombinationsformen 1 und 3 zeigten, gefunden werden. Diese Sätze wurden für jedes Nomen gespeichert. Bei der darauffolgenden manuell durchgeführten Kontrolle wurden diejenigen Kombinationen ausgewählt, die eine Übereinstimmung des Falls, der Konjugation, sowie eine besonders hohe Assoziation aufwiesen. Zielsprechersätze durften nur einmal Verwendung finden, um ein eindeutiges Ergebnis bezüglich der Verknüpfung zu

erzielen und um eventuelle Trainingseffekte zu vermeiden. Dabei hat sich während der manuell durchgeführten Auswahl der Satzpaare gezeigt, dass sich für die Satzstruktur 4 fast keine Verknüpfungsmöglichkeit ergaben. Für beide Kombinationsformen 1 und 3 konnten jeweils 32 Satzpaare erzeugt werden.

Diese im Verhältnis zu den vorherigen Tests geringere Anzahl an Testmaterial ließ daher nur eine geringere Anzahl an Testkonditionen zu, so dass lediglich 4 Pegelunterschiede bei nur 2 verschiedenen Störsprechern getestet werden konnten. Für jeden Pegelunterschied pro Maskierertyp einer Kombinationsform konnten somit pro Testperson 4 Werte erzielt werden. Es wurden dafür Schrittweiten der Pegelunterschiede von 4 dB festgelegt und wie aus Formel hervorgeht der Gesamtbereich der Pegelvariation von 12 dB beibehalten. Dabei sollten sowohl Messpunkte für besonders hohe und niedrige Pegel des Zielsprechers zur Untersuchung des energetic Maskings gegeben sein, sowie ausreichende Messwerte im Bereich der geringen Pegelunterschiede zwischen Ziel- und Störsprecher zur Untersuchung des Einflusses des Informational Maskings, siehe Formel 4.4.

$$TMR = SRT_mask + [-6, -2, 2, 6\,]dB \tag{4.4}$$

Als Maskierertypen wurden gleiche Stimme, sowie um 3 Halbtöne erhöhte Stimme getestet, da diese besonders hohe Schwankungen und Unsicherheiten der Testpersonen im letzten Test gezeigt hatten. Aus den insgesamt 64 Satzpaaren die sich bezüglich der sinnhaften Verknüpfung ergaben, wurden je 32 Satzpaare für die Maskierertypen, gleiche Stimme und um 3 Halbtöne erhöhte Stimme vorgesehen. Diese 64 Satzpaare, wurden zusäztlich durch 32 weiterere Hörproben mit dem Maskierertyp Rauschen zur Überprüfung der Testmethode (Vergleich der Steigung der Diskriminierungsfunktion) auf insgesamt 94 Hörproben ergänzt.

Innerhalb der für einen Maskierertyp vorgesehenen Satzmenge wurden jeweils die Hälfte der Satzpaare derart verändert, dass durch Austauschen der Störsprechersätze unter sich die sinnhafte Verknüpfung verloren ging. Jedem Zielsprechersatz wurde

ein weiterer Störsprechersatz, der keine sinnhafte Verknüpfung ergab zugeordnet, um die Sprachverständlichkeit der gewählten Zielsprechersätze im Vergleich mit diesen sinnfreien Satzkombinationen zu testen. Alle Testpersonen erhielten somit Zielsprechersätze mit unterschiedlichen Störsprechersätzen in Bezug auf die Verknüpfung. Dadurch konnte die Sprachverständlichkeit für jeden Zielsprechersatz jeweils mit und ohne sinnhafter Verknüpfung ermittelt werden. Es resultierten daher für jeden Maskierertyp jeweils 16 sinnhaft verknüpfte, sowie 16 nicht sinnhaft verknüpfte Satzpaare.

Jede der 4 Testkonditionen: Maskierertyp (gleiche Stimme, um 3 Halbtöne erhöhte Stimme), Satzpaarkombination (sinnhaft oder nicht sinnhaft verknüpfte Satzpaare) sollten bei jedem der 4 unterschiedlichen Lautstärkepegel aus Formel 4.4 jeweils 4 mal dargeboten werden. Die Reihenfolge der Hörproben entschied sich nach Maskierertyp und Form der Satzkombination. Die Pegelreihenfolge wurde nicht beachtet, allerdings durften identische Pegel nicht direkt aufeinander folgen, um eine kurzzeitige Anpassung des Gehörs an den Pegel und damit einen Trainingseffekt zu vermeiden.

4.3.2 Testablauf

Wie die Ergebnisse des vorangegangenen Tests zeigten, verbesserte sich die Sprachverständlichkeit für die Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme im Laufe des Tests. Um die Auswirkungen dieses Trainingseffekts und den Einfluss der Darbietungsreihenfolge auf die Resultate gering zu halten, wurden die 32 Satzpaare eines jeden Maskierertyps zu je 16 Hörproben aufgeteilt und verteilt im Testablauf dargeboten. Auf diese Weise konnte sowohl der Reihenfolgeunabhängigkeit als auch der Konzentration der Testpersonen auf den jeweiligen Maskierertyp Rechnung getragen werden. Jeder Maskiertyp wurde sowohl vor der Pause in der ersten Hälfte, als auch nach der Pause, in der zweiten Hälfte des Testes mit 16 Hörproben dargeboten. Mittels einer 4x4 Matrix [28] konnte die Darbietungsreihenfolge der beiden Maskierertypen gleiche Stimme und um 3 Halbtöne erhöhte Stimme gleich verteilt werden. Diese wurden insgesamt 4 mal angewandt um die Gleichverteilung für 16 Satzpaare einer Testkondition (Maskierertyp und Satzpaarkombination) zu ermöglichen, woraus

16 Testpersonen resultierten. Der Maskierertyp Rauschen wurde stets zu Beginn jeder Hälfte des Tests durch jeweils 16 Hörproben dargeboten, siehe Tabelle 4.6.

Zusätzlich wurde zu Beginn eines Testdurchlaufs eine längere Trainingsphase bestehend aus 23 Hörproben durchgeführt. Hierbei wurde zuerst die Zielsprecherstimme 7 mal ohne Degration vorgespielt. Danach wurden für jeden Maskierertyp (Rauschen, gleiche Stimme, um 3 Halbtöne erhöhte Stimme) jeweils 4 unterschiedliche SNR-, bzw. TMR-Werte nach Formel 4.4 dargeboten. Für alle Hördarbietungen musste die Testperson das Gehörte eingeben. Jede Trainingsphase beinhaltete die gleichen Sätze für alle Testpersonen, um die Wahrscheinlichkeit zu erhöhen, dass sie gleich stark trainiert wurden.

Position im Test	Maskierer	Satzpaarkombination
1.	Training	
2.	Rauschen	
3.	3 Halbtöne erhöhte Stimme	8 mit u. 8 ohne sinnhafter Verknüpfung
4.	gleiche Stimme	8 mit u. 8 ohne sinnhafter Verknüpfung
Pause		
5.	Rauschen	
6.	gleiche Stimme	8 mit u. 8 ohne sinnhafter Verknüpfung
7.	3 Halbtöne erhöhte Stimme	8 mit u. 8 ohne sinnhafter Verknüpfung

Tabelle 4.6: Reihenfolge der Darbietung der nterschiedlichen Konditionen: Maskierertypen und Satzpaarkombinationen als Beispiel für die erste Testperson

Für die Dauer der Bearbeitung einer Hörprobe (Hören, Eingeben, Kontrollieren) wurde von 30 Sekunden ausgegangen, so dass sich die voraussichtliche Gesamtdauer des Tests auf eine Stunde belaufen sollte. Vor jedem Testdurchlauf wurde mittels einer Audiometrie das Gehör der Testpersonen getestet, so dass sich die Testdauer um zusätzliche 10 Minuten erhöhte. Es wurde eine Pause in dem Testdurchlauf nach der Trainingsphase und weiteren 48 Hördarbietungen von 5 Minuten eingeführt. Die Testpersonen wurden mit dem Testablauf in Form einer Beschreibung, die unter anderem Programmbilder, siehe Anhang A.4.3 enthielt, vertraut gemacht. Darüber hinaus wurden den Testpersonen vor jedem Abschnitt im Test mitgeteilt um welchen Maskierertyp es sich im Folgenden handeln würde, um sie auf ihre Hörproben vorzubereiten.

4.3.3 Ergebnisse

Die dargestellten Verläufe (64 Ergebnisse pro dargestelltem Wertepunkt) in den Abbildungen 4.16 und 4.17 zeigen, dass die Werte der Sprachverständlichkeit als Funktion des TMR für Satzpaare mit und ohne sinnhafter Verknüpfung für beide Maskierertypen sich entgegen der Annahmen nur sehr gering unterscheiden. Die Konfidenzintervalle sind hoch und überschreiten das Signifikanzniveau von 5 %. Es sind daher keine signifikanten Unterschiede durch die sinnhafte Verknüpfung in Bezug auf die Sprachverständlichkeit zu erkennen.

Aufgrund der wenigen möglichen Messpunkte der Pegelunterschiede konnte der Bereich zwischen 0 und -4 dB TMR nicht getestet werden, so dass der Abfall der Sprachverständlichkeit wie der Abbildung 4.13 zu entnehmen hier nicht untersucht werden kann.

Die Werte der Sprachverständlichkeit bei dem Maskierertyp Rauschen zeigen in diesem Versuch zwar etwas geringere Konfidenzintervalle, allerdings befinden sie sich über dem Signifikanzniveau von 5 %. Die Steigung der Sprachverständlichkeit zwischen -4 und 0 dB TMR beträgt hier 35 %. Allerdings kann bei den zu wenigen Messpunkten der Steigungsansatz verschoben werden (Messung zwischen 0 und 4 dB TMR), wodurch sich ein anderer Wert ergeben würde.

Die gemittelte Sprachverständlichkeit über alle Testpersonen und über alle Pegel wurde für die Darbietungsreihenfolge dargestellt, siehe Abbildung 4.19. Die Werte schwanken zwischen 39 und 52 % bei dem Maskierertyp gleiche Stimme und zwischen 45 und 57 % bei dem Maskierertyp um 3 Halbtöne erhöhte Stimme. Die Schwankungen konnten hier für den Maskierertyp gleiche Stimme um 9 % zum vorangegangen Test verringert werden.

Die Ergebnisse dieses Tests zeigen keine deutlichen Unterschiede in der Sprachverständlichkeit

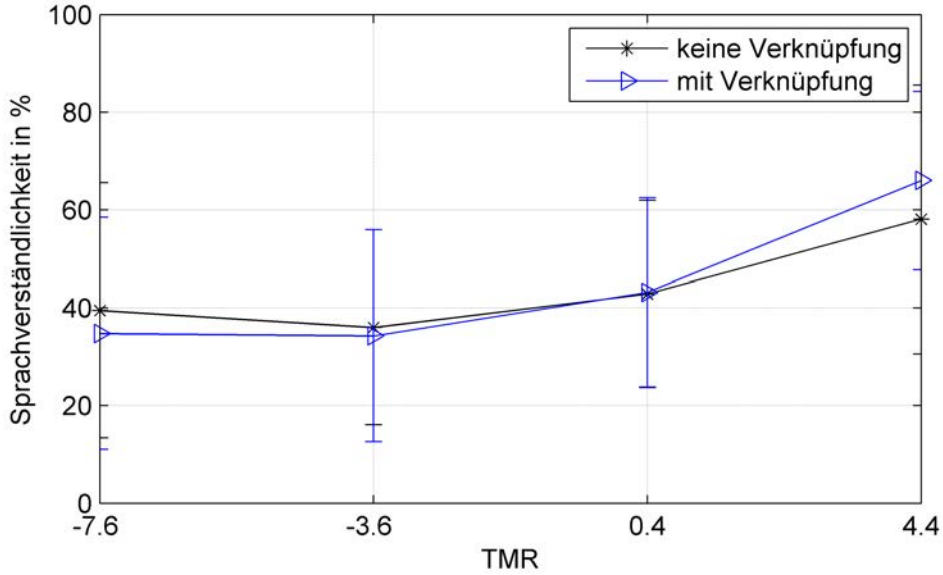


Abbildung 4.16: Sprachverständlichkeit in % für beide Satzpaarformen (sinnhaft verknüpft und nicht sinnhaft verknüpft) für den Maskierertyp gleiche Stimme

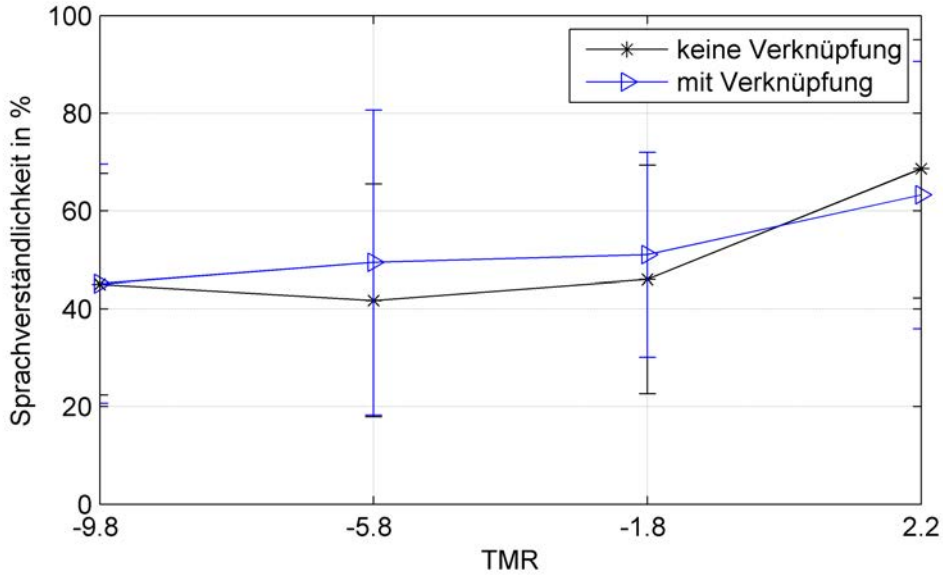


Abbildung 4.17: Sprachverständlichkeit in% für beide Satzpaarformen (sinhaft verknüpft und nicht sinnhaft verknüpft) für den Maskierertyp um 3 Halbtöne erhöhte Stimme

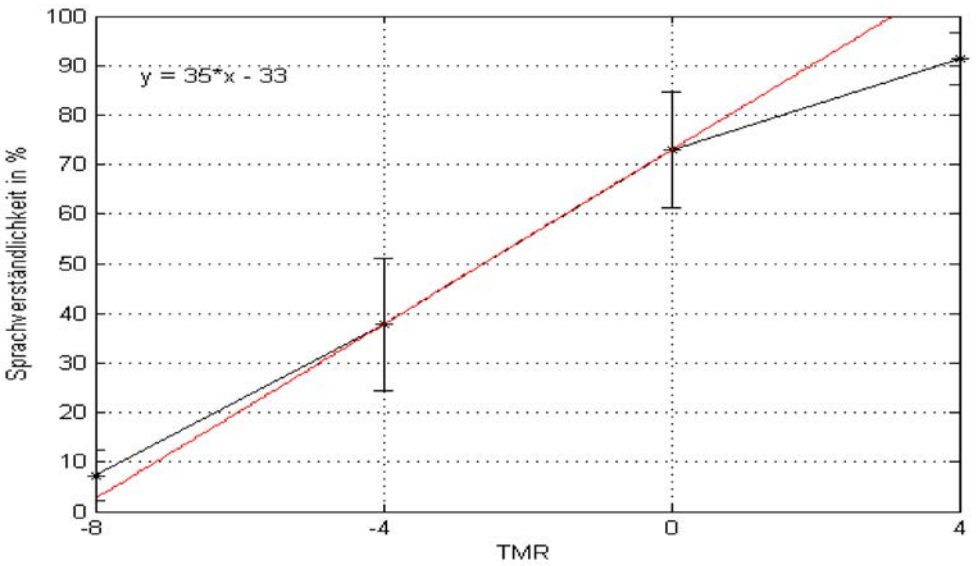


Abbildung 4.18: Sprachverständlichkeit in% für den Maskierertyp Rauschen

zwischen den Satzpaarformen mit und ohne sinnhafter Verknüpfung. Es verringerte sich die Schwankung über die Zeit in der Darbietungsreihenfolge für den Maskierertyp gleiche Stimme.

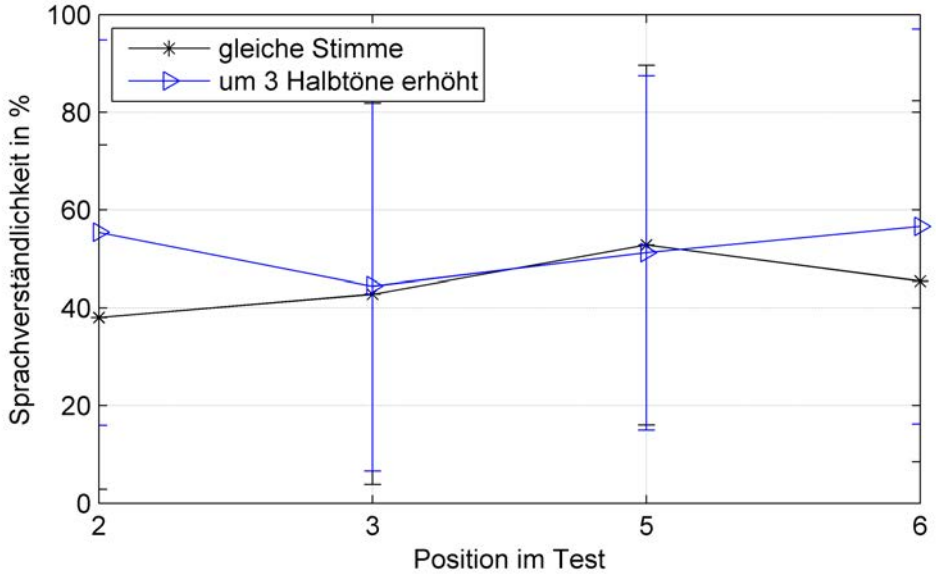


Abbildung 4.19: gemittelte Sprachverständlichkeit in% über alle Testpersonen und Pegel für die Darbietungsreihenfolge aufgetragen

5 Diskussion

5.1 Einfluss der Maskierung

Auffällig bei den ermittelten Ergebnissen bezüglich der Sprachverständlichkeit in Abhängigkeit der Pegelunterschiede, wie sie in Test 2 ermittelt worden waren, war der unregelmäßige Verlauf der Sprachverständlichkeit des Maskierertyps gleiche Stimme, siehe Abbildung 4.13. Bei ähnlicher Lautstärke beider Sprecher und dem gleichen Sprecher als Maskierer war die Sprecherähnlichkeit in diesem Test am höchsten. Die Sprachverständlichkeit wurde im Vergleich zu allen anderen Maskierertypen in diesem Bereich bei gleicher Stimme am meisten beeinträchtigt. Statt der erwarteten 50 % Sprachverständlichkeit bei dem SRT-Wert von -1.6 dB TMR wurde eine Sprachverständlichkeit von 28 % ermittelt.

Der Einbruch zwischen -3.6 und 0.4 dB TMR könnte der Bereich sein an dem die Unsicherheit bezüglich der Richtigkeit des Zielsprechers der Testpersonen am höchsten war [12]. Verändert sich die Lautstärke des Zielsprechers und die Pegelunterschiede werden ab 0.4 dB TMR bis 4.4 dB TMR erhöht, waren die Testpersonen in der Lage das Gesprochene leichter zu unterscheiden, wodurch die Sprachverständlichkeit anstieg. Sank der Pegel des Zielsprechers in dem Bereich zwischen -7.6 und -3.6 dB TMR konnte aufgrund des deutlicheren Pegelunterschieds der Zielsprecher wieder besser von dem Störsprecher unterschieden werden. Allerdings hatte dies aufgrund der energetischen Maskierung durch die sinkende Lautstärke des Zielsprechers nicht die gleiche starke Auswirkung auf die Sprachverständlichkeit wie im Bereich ab 0.4 dB TMR aufwärts. Dieser Verlauf und die Annahme wurde bereits durch Darwin et al. 2003 festgestellt [12].

Die Unsicherheit der Testpersonen über das Verstandene könnte zu einer Zuordnung des Gehörten zum Ziel- bzw. Störsprecher führen [13]. Die Wörter des Störsprechers könnten als Schlüsselwörter des Zielsprechers gedeutet worden sein. Bei Betrachtung der Abbildung 5.2 in der die Eingaben der Wörter des Störsprechers über die TMR-Werte aufgetragen worden sind, zeigt sich, dass für alle Pegelunterschiede Unsicherheit in Form von Wiedergabe der Wörter des Störsprechers auftrat. Die Auftretenshäufigkeit der Wörter des Strösprechers schwankt hier zwischen 13 % und 34 %. Mit einem Maximum von 34 % der Auftretenshäufigkeit bei -1.6 dB TMR ist die Anzahl der Maskiererwörter ungefähr gleich der Anzahl der Schlüsselwörter des Zielsprechers mit 29 %, vgl. Abbildung 4.13. In diesem Bereich der Pegelunterschiede zwischen Ziel- und Störsprecher konnte zwischen dem Ziel- und Störsprecher offenbar nicht mehr genau unterschieden werden, so dass hier die Wahrscheinlichkeit die Wörter des Zielsprechers verstanden zu haben, gleich war.

Zur genaueren Analyse des Umfangs und damit der Skalierbarkeit von Informational Masking wurden die Antwortmöglichkeiten der Testpersonen nach Shinn-Cunningham et al. 2008 [19] in Gruppen aufgeteilt. In dem dort beschriebenen Versuch wurde mit nur 2 Schlüsselwörtern pro Satz nach der CRM-Methode getestet. Bei 3 verschiedenen Eingabeformen, ob Wort des Zielsprechers (Z), des Maskierers (M), oder keine bzw. fehlerhafte Eingabe (X) würde sich demnach nach Formel 5.1 eine Anzahl von 9 Kombinationsmöglichkeiten ergeben.

$$Eingabeform^{Schlüsselwortanzahl} = Kombinationsmöglichkeit \quad (5.1)$$

Bei der Gruppierung der Eingabemöglichkeiten wurde die Position der Schlüsselwörter im Satz nicht beachtet. Es wurden insgesamt 6 Eingabekategorien definiert, die jeweils einer Form der Maskierung, Informational Masking oder Energetic Masking zugeordnet wurden, siehe hierzu Tabelle 5.1. Demnach wurde das Informational Masking durch Fehler der Eingabekategorie Maskiererfehler und Gemischte Fehler definiert, d.h. dass ausschließlich Wörter die von dem Ziel- (Z) und dem Störsprecher (M), oder nur von dem Störsprecher wiedergegeben wurden, gewertet wurden. Das Energetic Masking

Kategorienummer	Eingabekategorie	Eingabekombination	Maskierungsform
1	Maskierer Fehler	MM	IM
2	Gemischte Fehler	ZM, MZ	IM
3	1 Falscheingabe	ZX, XZ	EM
4	2 Falscheingaben	XX	EM
5	Kombinierte Fehler	MX, XM	EM, IM
6	Korrekt	ZZ	–

Tabelle 5.1: Definition der Kombinationsmöglichkeiten der Eingabeform mit X = fehlerhaftes Wort, M = Maskiererwort, Z = Zielsprecherwort

wurde definiert durch Eingabe der Schlüsselwörter des Zielsprechers (Z) und falscher, bzw. keiner Eingabe (X). Traten fehlerhafte Wörter (X) sowie Wörter des Störsprechers (M) zusammen auf, wurde angenommen, dass sie durch beide Maskierungsformen IM und EM hervorgerufen wurden [19].

Bei dem hier durchgeführten Versuch wurden allerdings 3 Schlüsselwörter pro Satz verwendet, wodurch nach der oben genannten Formel 5.1 27 Kombinationsmöglichkeiten auftreten konnten. Die Definitionen und Benennung dieser Kombinationsmöglichkeiten müsste um die dritte Position erweitert werden. Aufgrund der Nichtbeachtung der Position im Satz und der Anzahl der Eingabeformen pro Satz können die Kategorien 1 und 2, "Maskierer Fehleründ „Gemischter Fehler“ übernommen werden, siehe Tabelle 5.1. Dasselbe gilt für den Fall 3 und 4, der einmaligen Falscheingabe und der zweifachen Falscheingabe, siehe hierzu Tabelle 5.2. Aus der, für die um die dritte Position erweiterte Eingabekombination ersteltte Tabelle 5.2 läßt sich entnehmen, dass für beide Maskierungsformen IM und EM damit gleich viele Kombinationsmöglichkeiten der Eingabeformen vorhanden sind. Treten allerdings alle 3 Eingabeformen X,Z und M zusammen in Kombination auf, bildet dies eine neue, noch nicht definierte Kombinationsform und könnte sowohl dem Informational Masking als auch dem Energetic Masking zugeordnet werden. Für die weitere Diskussion werden aber nur die eindeutigen Fälle des IM und des EM, die Kombinationsformen 1-4 (Tabelle 5.2) betrachtet, um möglichst genaue Aussagen über das Verhalten der beiden Maskierungsformen treffen zu können, wodurch dieser Fall der Kombinationsform 6 außer Acht gelassen wird.

Nr.	Eingabekategorie	Eingabekombination	Maskierungsform
1	Maskierer Fehler	MMM	IM
2	Gemischte Fehler	ZMM, MMZ, MZM, ZZM, ZMZ, MZZ	IM
3	1-2 Falscheingaben	ZXX, XXZ, XZX, ZZX, ZXZ, XZZ	EM
4	3 Falscheingaben	XXX	EM
5	Kombinierte Fehler	MXX, XXM, XMX, XMM, MMX, MXM	EM, IM
6	Alle Eingabeformen	XMZ, ZMX, MZX, ZXM, XZM, MXZ	EM, IM
7	Korrekt	ZZZ	–

Tabelle 5.2: erweiterte Definition der Kombinationsmöglichkeiten der Eingabeform mit X = fehlerhaftes Wort, M = Maskiererwort, Z = Zielsprecherwort

5.1.1 Informational Masking

Es wurde die Auftretenshäufigkeit aller Fälle, die nach der in Tabelle 5.2 der Definition für IM entsprechen, gemittelt über alle Testpersonen für den Maskierertyp gleiche Stimme aufgetragen, siehe Abbildung 5.1. Zudem wurden zum Vergleich unabhängig der Kombinationskategorien alle Wörter, die von dem Störsprechersatz stammen und statt des Schlüsselwortes des Zielsprechers (Maskiererwort M) eingegeben worden sind, gemittelt aufgetragen. Es wurde somit die allgemeine genormte Häufigkeit der Maskiererwörter quantitativ gegen die Fälle in denen sie durch IM auftraten qualitativ, aufgetragen, siehe 5.2.

Für das verwendete syntaktisch korrekte aber semantisch unzusammenhängende Sprachmaterial konnte somit ein Einfluss durch IM bei Sprache als Maskierer von maximal 30 % bei 0.4 und 2.4 dB TMR aufgezeigt werden (Abbildung 5.1). Zwischen -5.6 und 2.4 dB TMR liegt die Fehlerhäufigkeit versuracht durch IM bei 20 % und mehr. Außerhalb dieser TMR-Werte liegt der Einfluss bei 15 % und weniger. Die Ziel- und Störsprecher sind offenbar bei 4.4 dB TMR deutlicher zu unterscheiden und bei -7.6 dB TMR könnte sich verstärkt der energetische Einfluss bemerkbar gemacht haben. Der Verlauf verhält sich ähnlich zum zitierten Versuch nach Shinn-Cunningham 2008 [19] innerhalb des Bereichs dieser TMR-Werte.

Die Auftretenshäufigkeit der Maskiererwörter ist bei -1.6 dB TMR mit 34 % am höchsten bei verhältnismäßig geringerem Einluss durch IM, siehe Abbildung 5.2. Offenbar wur-

den bei -1.6 dB TMR zwar mehr Fehler pro Satz durch den Störsprecher erzeugt, aber für insgesamt weniger Sätze die durch IM betroffen waren als beispielsweise bei 2.4 dB TMR, vgl. Abbildung 5.1. Somit traten Maskiererfehler häufiger komplett im ganzen Satz und nicht einzelnd auf. Der Störsprecher wurde demnach öfters für den Zielsprecher gehalten. Die selektive Aufmerksamkeit und damit die korrekte Wahl des Zielsprechers, war bei diesem TMR-Wert eingeschränkter als bei den anderen. Allerdings spricht die konsequente Einhaltung der Wahl der Wörter des Störsprechers über den Satz hinweg, für eine bessere Segregation als bei 2.4 dB TMR.

Im Vergleich dazu zeigen die Ergebnisse bei 2.4 dB TMR (in Abbildung 5.2) hingegen, dass weniger Maskiererfehler auftreten, aber dafür ein stärkerer Einfluss von IM vorhanden war. Dieses könnte für eine schlechtere Segregation sprechen, da häufiger Sätze auftreten, in denen Wörter sowohl des Ziel- als auch des Störsprechers auftreten, aber für eine bessere Selektion, da oft die Wörter des Zielsprechers eingegeben wurden. Bei nur 24 % Auftretenshäufigkeit der Maskiererfehler aber mit 30 % Auftretenshäufigkeit des Fälle von IM hat bei diesem TMR-Wert relativ zu den anderen TMR-Werten Informational Masking den größten Einfluss. Dies könnte bedeuten, dass eher durch geringerer Segregation als durch Selektion der Einfluss des Informational Masking am höchsten ist.

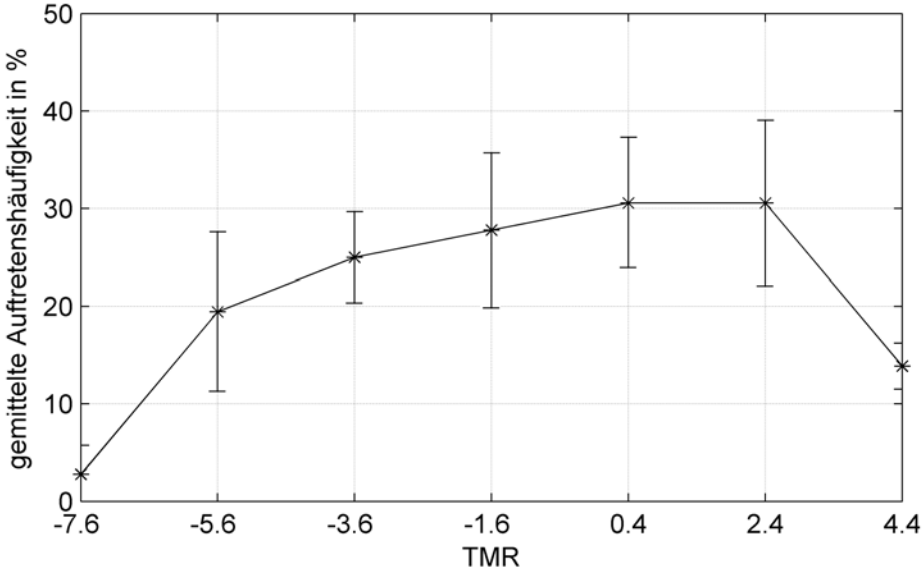


Abbildung 5.1: Analyse der Antworten für IM als Funktion der Pegeldifferenz für gleiche Stimme als Maskierer

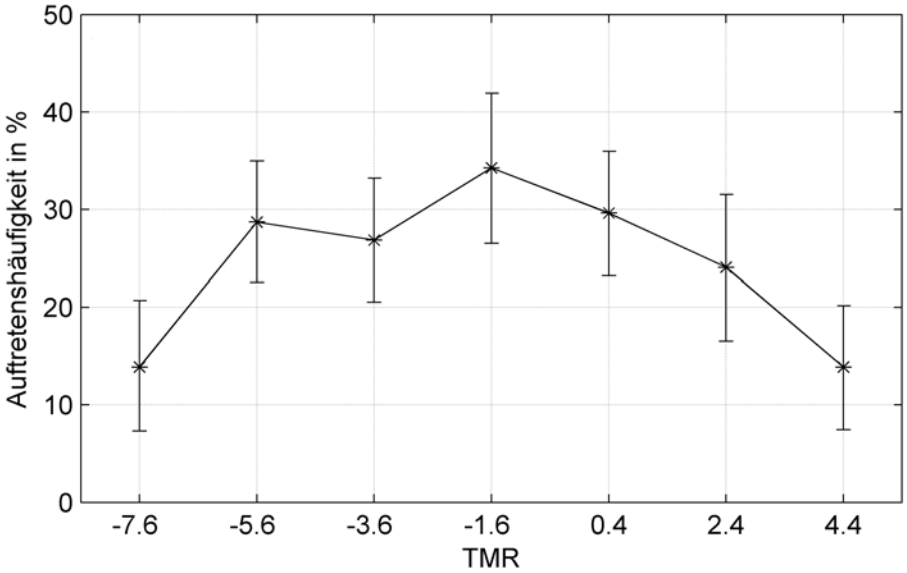


Abbildung 5.2: Auftretenshäufigkeit der Maskiererfehler als Funktion der Pegeldifferenz für gleiche Stimme als Maskierer

5.1.2 Energetic Masking

Bei Sprache als Maskierungsform hat das EM ab einem TMR-Wert von -9 dB TMR und größer im Vergleich zum IM eher geringere Auswirkungen [9]. Das Energetic Mas-

king ist ein Resultat der Maskierung an der Peripherie und hat Auswirkungen auf die Sprachverständlichkeit. Das stationäre Rauschen als Maskierer erzielt durch permanente energetische Überlagerung des Zielsprechers eine höhere energetische Maskierung als wenn die gleiche Stimme den Sprecher maskieren würde [8]. Durch die Möglichkeit in die Pausen bei Sprache als Maskierer den Zielsprecher rauszuhören (dip-listening Effekt), wird ein besseres SRT für Sprache als für Rauschen ermittelt [16] [7]. Wegen der ähnlichen TMR-Werte die sich um 0.4 dB TMR unterscheiden, bei denen die beiden Maskierertypen Rauschen und gleiche Stimme hier getestet worden sind, wurden beide Verläufe gemeinsam zum Vergleich des Einfluss der energetischen Maskierung in Abbildung 5.3 aufgetragen.

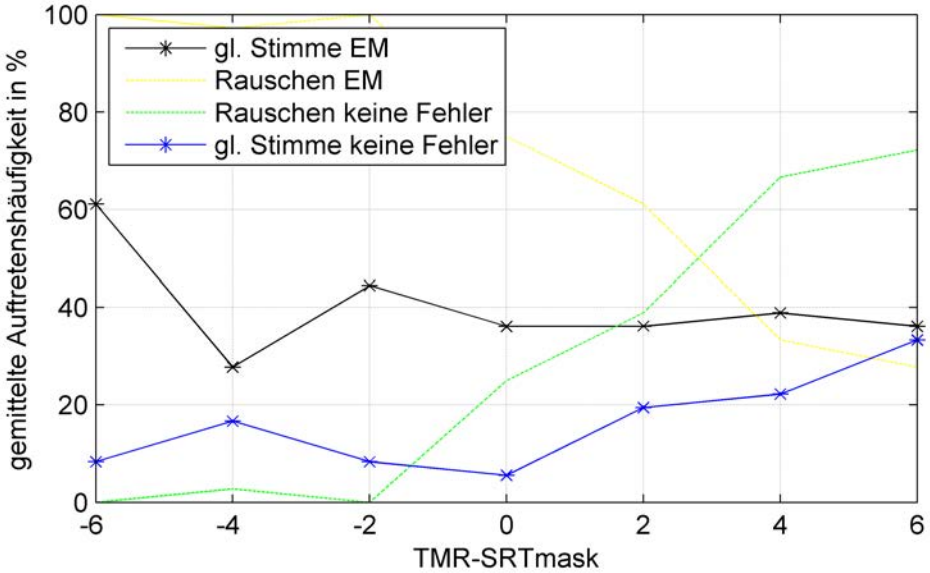


Abbildung 5.3: Analyse der Eingaben für EM als Funktion der Pegeldifferenz zum SRT_mask für Rauschen und gleiche Stimme als Maskierer

Für den Maskierertyp Rauschen konnte ein höherer energetischer Einfluss (gelb gestrichelt) bei niedrigeren SNR-Werten, zwischen SRT_mask + (-6 dB TMR) bis SRT_mask + 2 dB TMR als bei dem Maskierertyp gleiche Stimme (schwarze Sterne) wie oben angenommen, bestätigt werden. Hier liegt der Einfluss der Maskierung um bis zu maximal 40 % höher. Der Einfluss durch EM liegt beim Maskierertyp gleiche Stimme relativ konstant bei 40 % ab SRT_mask + (-2 dB TMR). Ab SRT_mask + (-2 dB TMR) ist ein

konstanter Anstieg der Anzahl der komplett korrekten Eingaben bei Rauschen (grün gestrichelt) als Maskiertyp zu erkennen. Erst ab SRT_mask + 0 dB TMR kann auch ein Anstieg der komplett richtig Eingaben bei dem Maskierertyp gleiche Stimme (blaue Sterne) gesehen werden.

Es hat sich gezeigt, dass der Anstieg der komplett richtigen Eingaben der beiden Maskierertypen, Rauschen und gleiche Stimme ab SRT_mask + 0 dB TMR unterschiedlich stark ansteigt. Bei gleicher Grundfrequenz, aber mit Modulation und Semantik der Wörter ist die Sprachverständlichkeit ab diesem Bereich für die gleiche Stimme als Maskierer geringer. Offenbar führt diese Information zu Irritation, wodurch mehr Fehler erzeugt worden sind als für Rauschen als Maskierer.

5.2 Einfluss durch geänderte Sprechereigenschaften

Wie in Test 4.1 gezeigt, können durch unterschiedliche Grundfrequenzen der Störsprecherstimmen bei der gleichen Sprachverständlichkeit (50 %) unterschiedliche TMR-Werte (SRT_mask) ermittelt werden, siehe Tabelle 4.3. Aus Abbildung 5.4 geht hervor, dass mit zunehmender Änderung der Grundfrequenz zwischen Ziel- und Störsprecher, der TMR-Wert für eine Sprachverständlichkeit von 50 % logarithmisch sinkt. Eine Erhöhung der gleichen Stimme um 6 Halbtöne des Störsprechers scheint derart unterscheidbar zu sein, dass sich die Werte der weiblichen Stimme, die komplett anders ist, trotzdem annähern.

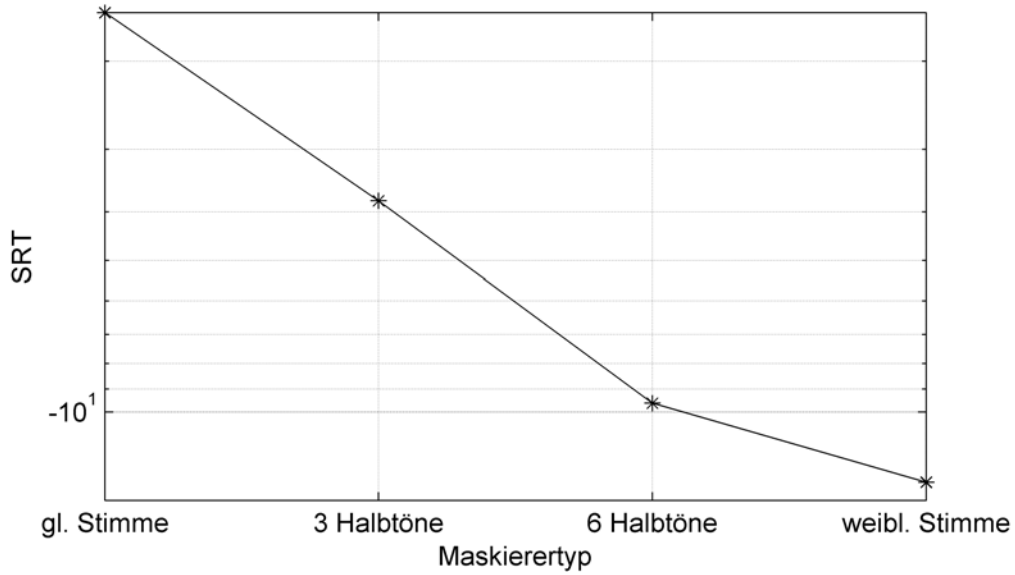


Abbildung 5.4: SRT_mask für alle Maskierertypen

Das IM wird durch die Sprecherähnlichkeit in Bezug auf hörbare und linguistische Eigenschaften beeinflusst. Wird die Ähnlichkeit des Störsprechers zum Zielsprecher erhöht, gleiches Geschlecht oder sogar gleicher Sprecher, würde die Anzahl der Maskiererfehler steigen [8]. Änderungen in der Grundfrequenz der Stimmen führen zu Unterschieden in dem wahrgenommenem auditiven Spektrum. Je ähnlicher die Spektren der beiden Stimmen des Ziel- und Störsprechers sind, desto ähnlicher sind die spektralen auditiven Muster, die vom Hörer wahrgenommen werden, siehe hierzu Kapitel 2.3. Die Segregation des Gehörten und die selektive Aufmerksamkeit zum Zielsprecher könnte demnach entsprechend schlechter durchgeführt werden [13]. Bei dem hier durchgeführten Versuch wurden als Störsprecher unter anderem die gleiche Stimme, die gleiche Stimme um 3 Halbtöne und um 6 Halbtöne erhöht, getestet. Die weibliche Stimme kann wegen beispielsweise anderer Artikulation und Klangfarbe nicht direkt in diesem Konetxt der Grundfrequenzerhöhung mit den Ergebnissen der anderen Störsprecherstimmen verglichen werden [12] [8], würde aber aufgrund der hohen Unterschiede zur Zielsprecherstimme die höchste Sprachverständlichkeit bei gleichen TMR-Werten erzielen [12].

Zur Veranschaulichung wurde, die in Abbildung 5.5 aufgetragene Häufigkeit der gege-

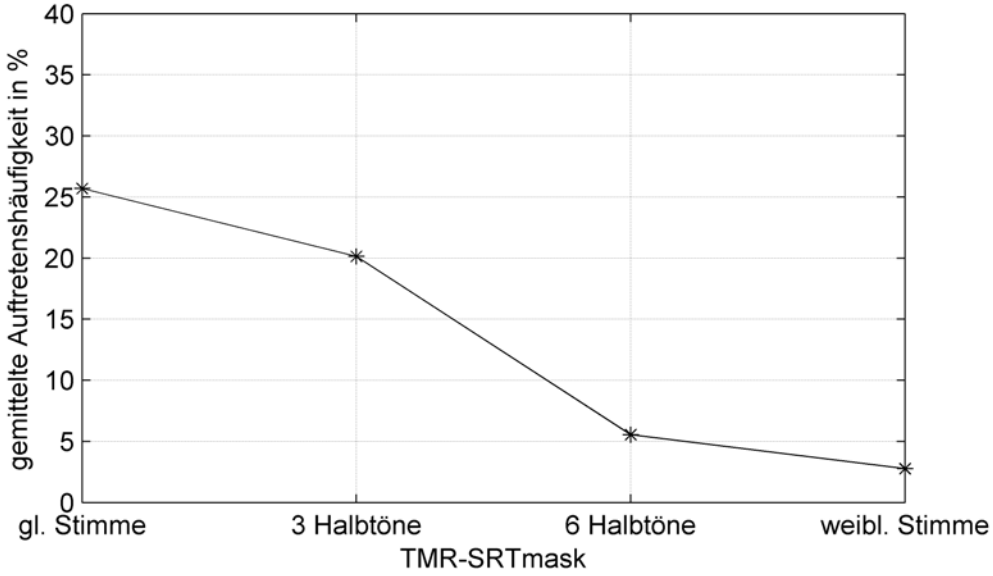


Abbildung 5.5: gemittelte Auftretenshäufigkeit der Fehler durch IM für alle Störsprecher

benen Antworten, die aus Informational Masking resultieren, gemittelt über alle Pegel und Testpersonen für jeden Störsprecher dargestellt. Wurde die Stimme um 3 Halbtöne erhöht war die Fehlerhäufigkeit verursacht durch IM bei 20 %. Für den Störsprecher um 6 Halbtöne erhöht, sinkt der Einfluss von IM von 26 % auf 6 % und nähert sich dem Wert der weiblichen Stimme als Störsprecher an. Der Einfluss von IM bei der weiblichen Stimme liegt trotz der hohen Stimmenunterschiede zwischen Stör- und Zielsprecher bei 3 %. Der Verlauf des Einfluss von IM abhängig der Grundfrequenzerhöhung des Störsprechers verhält sich nicht linear. Die oben getätigten Annahmen bezüglich der Unterscheidbarkeit zwischen den Störsprechern und des Einfluss von IM konnten bestätigt werden. Über die Genauigkeit der Werte läßt sich allerdings, wegen der hohen Schwankungen der Ergebnisse zwischen den Testpersonen, keine zuverlässige Aussage treffen.

5.2.1 Unterschiede zwischen den Testpersonen

Die Unterschiede zwischen den Störsprechern kann bei jeder Testperson, aufgrund ihrer physischen und neuronalen Konstitution unterschiedlich stark Einfluss nehmen

[27]. Die Ergebnisse der 3 durchgeführten Tests haben gezeigt, dass sich die Sprachverständlichkeit zwischen den Testpersonen stark unterschieden, siehe Abbildung **??** und Kapitel 4.2.3. Durch den Audiometrietest, der mit den Testpersonen wie in Kapitel 3.1 beschrieben, durchgeführt worden ist, kann davon ausgegangen werden, dass die Testpersonen, die an diesem Sprachverständlichkeitstest teilnahmen, in etwa gleich gut hörten. Als Faktor für die Unterschiede kann Unsicherheit in der Sprachverständlichkeit abhängig der Sprechereigenschaften angenommen werden [27]. Vor allem bei einer hohen Sprecherähnlichkeit steigt die Unsicherheit [27], was bei dem Maskierertyp gleiche Stimme in diesem Test am höchsten wäre. Zudem wurde gezeigt, dass erst ab 3 Halbtönen Grundfrequenzunterschied zwischen beiden Stimmen unterschieden werden kann [12].

Daher wird für diesen Maskierertyp die gemittelte Auftretenshäufigkeit der Wörter des Ziel- und des Störsprechers für alle Testpersonen einzeln untersucht, siehe Abbildung 5.6. Es zeigt sich, dass z.B. die Testpersonen 2,7,11 und 12 relativ viele Wörter des Zielsprechers wiedergaben, aber verhältnismäßig wenig Wörter des Störsprechers. Die Testpersonen 4,5,9 und 10 hingegen weisen eine eher geringe Sprachverständlichkeit bezüglich des Zielsprechers auf, aber gaben häufiger als die anderen die Wörter des Störsprechers wieder.

Aus den Ergebnissen aus Test 4.2.3 sind Unterschiede zwischen den einzelnen Testpersonen in Bezug auf die unterschiedlichen Störsprecher zu erkennen, vgl. zudem Abbildung 4.15. Die Resultate einiger Testpersonen zeigten deutliche Unterschiede in der Sprachverständlichkeit hier am Beispiel zwischen den beiden Störsprechern, gleiche Stimme und um 3 Halbtöne erhöhte Stimme. Bei näherer Betrachtung der gemittelten Sprachverständlichkeit für alle Maskierertypen, dargestellt für alle Testpersonen in Abbildung 5.7 mit 21 Ergebnissen pro Wertepunkt, zeigt sich, dass auch die anderen Maskierertypen unterschiedlich deutlich wahrgenommen wurden. Die Sprachverständlichkeit eines einzelnen Maskierers variiert zwischen den Testpersonen.

Zudem ist zu erkennen, dass die Relationen der Sprachverständlichkeit der unterschied-

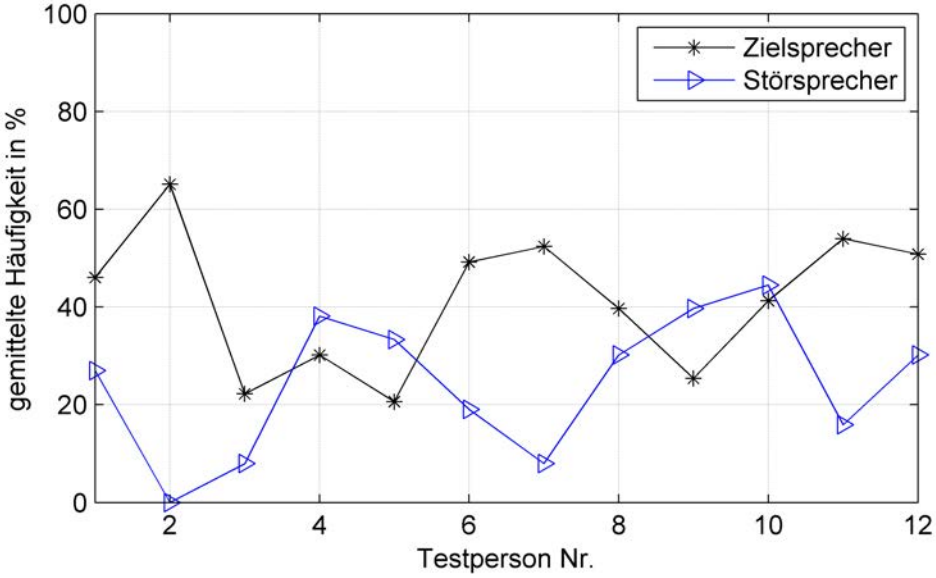


Abbildung 5.6: gemittelte Auftretenshäufigkeit der Wörter des Ziel- und Störsprechers gleicher Stimme

lichen Maskierertypen zueinander für jede Testperson unterschiedlich sind. Die gemittelten Sprachverständlichkeitswerte der einzelnen Maskierertypen liegen bei den Testpersonen 1, 2 und 3 beispielsweise nah beieinander und schwankten um 20 % der Sprachverständlichkeit. Für die Resultate der Testpersonen 5 und 9 variieren sie zwischen 20 und 60 % der Sprachverständlichkeit. Für 7 von 12 Testpersonen erzielt die gleiche Stimme als Maskierer die geringsten Sprachverständlichkeitswerte, im Vergleich zu den anderen Maskierertypen.

5.3 Einfluss durch Kontext

5.3.1 Syntax

Wie stark der verbleibende Einfluss des Kontexts durch Syntax und Koartikulation sein kann, soll hier näher betrachtet werden. Wie bereits beschrieben, wurde für den Maskierer- und den Zielsprecher der gleiche Sprecher verwendet, so dass kein großer Unterschied in der Betonung, Klangfarbe usw. vorlag. Ebenso wurden die Unterschiede der gesprochenen Geschwindigkeit in hohem Maße klein gehalten, indem die Satzpaa-

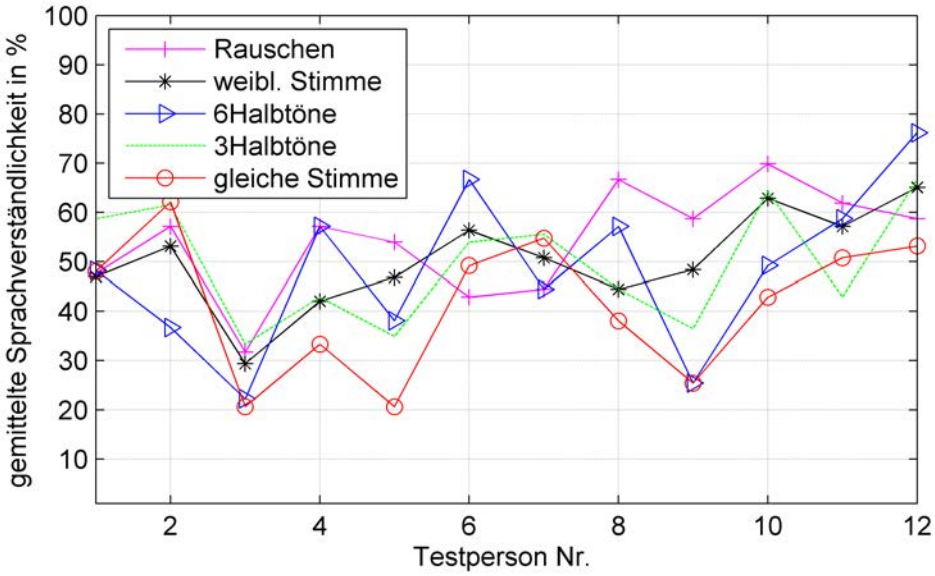


Abbildung 5.7: Abweichung der gemittelten Sprachverständlichkeit zwischen den einzelnen Maskierertypen für alle Testpersonen

re nach ähnlicher Länge kombiniert worden sind. Die Satzpaare entstammten jeweils der gleichen Satzstruktur. Es wurden für das Satzmaterial 4 unterschiedliche syntaktische Satzstrukturen [1] verwendet, siehe Kapitel 3.1. Sie beinhalteten alle gleich viele Schlüsselwörter.

Es könnte trotzdem sein, dass Wörter einiger Satzstrukturen durch die Syntax eine größere syntaktische Beziehung zu einander haben als in anderen Satzstrukturen. In der folgenden Abbildung 5.8 werden Beispiele für jede einzelne Satzstruktur aufgeführt. Diese sind mit einem Unterstrich, der die Austauschbarkeit der Wörter zwischen den beiden Sätzen darstellt, da sie keinen syntaktischen Bezug auf das Folgewort im Satz aufweisen, aufgeführt. Die farblich markierten Wörter zeigen den syntaktischen Zusammenhang.

Unterschiedliche Syntax erfordert unterschiedliche Betonung der Wörter, wodurch eine Art Melodie des Satzes erzeugt wird. So wird beispielsweise eine Frage anders ausgesprochen als eine Assage. Der Sprecher hat die Sätze in natürlicher Form vorgetragen und nicht in monotoner, unbetonter Form. Aufgrund der entsprechend unterschiedli-

Struktur 1:
Die Hand | fällt _ mit _ dem _ dünnen _ Kurs.
Die Schau | hofft _ unter _ der _ knappen _ Macht.

Struktur 2:
Der steile | Berg _ schließt _ den Spruch.
Der dichte | Zoll _ weckt _ das Tuch.

Struktur 4:
Wann dient | das Pfand _ der _ tiefen _ Schnur?
Wo trinkt | der Elch _ den _ großen _ Stein?

Struktur 5:
Das Spiel | senkt _ die Haut die _ schaut.
Der Plan | liebt _ den Trost der _ schimpft.

Abbildung 5.8: Beispielsätze für alle 4 verwendete Satzstrukturen

chen Betonung der Wörter kann es sein, dass einer Satzstruktur leichter gefolgt werden kann als einer anderen. Durch unterschiedlich stark ausgeprägter Stimmdynamik kann unterschiedlich hohe Sprachverständlichkeit erzeugt werden [12] [22].

Die Abbildung 5.9 zeigt die Sprachverständlichkeit für jede Satzstruktur für alle drei Positionen der Schlüsselwörter im Satz. Das erste Wort wurde im Test vorgegeben, so dass es nicht als Schlüsselwort gezählt wird, daher werden die Positionen der 3 verbleibenden Wörter ab 1 gezählt. Die Ergebnisse der Satzstruktur 4, des Fragesatzes weisen die höchte Sprachverständlichkeit auf. Es wird daher vermutet, dass aus dem syntaktischen Muster eines Satzes auf den Klang der Wörter im Verlauf des Satzes geschlossen werden kann. Somit haben die unterschiedlichen Satzstrukturen Einfluss auf die Sprachverständlichkeit und können Schwankungen von bis zu maximal 20 % für einzelne Wortpositionen hervorrufen. Mit insgesamt 9 Werten für jede Testkondition einer jeden Satzstruktur und dem Einflussfaktor des IM sind es relativ wenig Testdurchläufe pro Testperson, um zuverlässige Aussagen über die genaue Schwankung für diese Satzstrukturen treffen zu können.

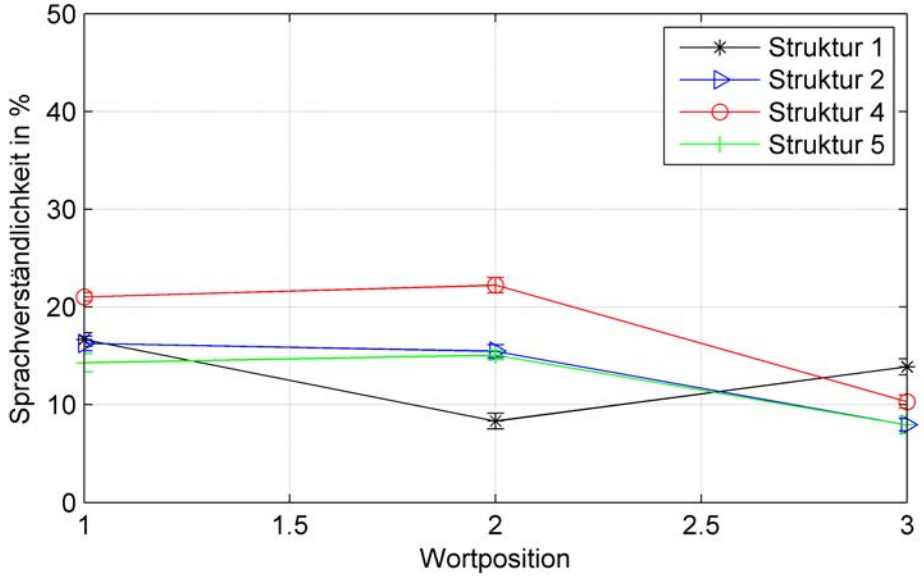


Abbildung 5.9: gemittelte Sprachverständlichkeit je Satzstruktur für jede Position der Schlüsselwörter im Satz

Bei genauerer Betrachtung der Strukturen zeigt sich auch, dass sie unterschiedlich viele Wörter enthalten. Nach Watson et al. 1987 wurde gezeigt, dass die Anzahl der Komponenten (Phoneme) einer auditiven Abfolge entscheidend für das Gedächtnis ist und nicht die Dauer dieser Abfolge [30]. Bei einer geringerern Anzahl der Phoneme einer Satzstruktur, würde somit die Wahrscheinlichkeit, dass alle Komponenten sich gemerkt worden sind, als für Satzstrukturen mit einer höheren Anzahl steigen. Dadurch könnte es sein, dass die Ergebnisse für die längeren Strukturen vergleichsweise schlechter ausfielen.

Einem gesprochenem Satz zu folgen ist das Ergebnis des temporal ausgedehnten Prozess der Aufmerksamkeit, auditives streaming, vgl. Kapitel 2.3. Wie bereits beschrieben kann Prososdie diesen Vorgang erleichtern. Durch die so erhaltenen neuronalen auditiven Muster, die übereinstimmen und aneinandergefügt werden, kann das Material einfacher gemerkt werden. Aufmerksamkeit und Gedächtnis sind eng mit dem Top-Down-Prozess verknüpft [3].

5.3.2 Semantik

Mit dem in Kapitel 4.3 beschriebenen Test wurde durch die sinnbehaftete Verknüpfung der Satzpaare des Ziel- und des Störsprechers zusätzlich Kontext konstruiert. Es wurden Sätze semantisch verknüpft, die aber getrennt wahrgenommen werden sollten. Dadurch könnte der schemabasierte Top-Down-Prozess das semantsiche Muster, der sinnvollen Folge von 2 Wörtern beider Sätze erkennen und zusammenfügen wollen. Allerdings steht die Verbindung dieser beiden Wörter im Konflikt mit dem Verknüpfen der akustischen Muster, basierend auf den Stimmeigenschaften (Grundfrequenz, Prosodie, Laustärke) des Zielsprechers.

Allerdings hat sich bereits gezeigt, dass bei dieser Form von Kontext keine Unterschiede in der Sprachverständlichkeit zwischen den unterschiedlichen Satzpaarkombinationen ermittelt worden sind, siehe 4.16. Wegen der beiden gewählten sinnbehafteten Verknüpfungsformen an der ersten und letzten Position im Satz, ist es daher von Bedeutung die Auftretenshäufigkeiten der Maskiererwörter (M) für diese Positionen darzustellen. Es hat sich zudem gezeigt, dass im Bereich zwischen -3.6 dB TMR und kleiner als 4.4 dB TMR die Unsicherheit für den Maskierertyp gleiche Stimme am meisten auftritt. Es wurden daher die Auftretenshäufigkeiten der Maskiererwörter für alle Positionen im Satz nur über die beiden TMR-Werte gemittelt aufgetragen, siehe 5.10.

Aus der Abbildung 5.10 läßt sich erkennen, dass bei der ersten Position im Satz die Auftretenshäufigkeit für Satzpaare die sinnbehaftet verknüpft worden sind um 14 % höher ist als für die unverknüpften Satzpaare. Demnach könnte der semantische Kontext die Testpersonen dazu verleitet haben die Wörter des Störsprechers einzugeben, anstatt die des Zielsprechers. Für die dritte Position hingegen, verhält sich die Auftretenshäufigkeit für Satzpaarformen (mit und ohne sinnbehafteter Verknüpfung) genau umgekehrt. Hier liegt die Auftretenshäufigkeit der Wörter des Störsprechers bei den Satzpaaren mit semantischer Verknüpfung um 10 % höher als mit. Der semantische Kontext hatte hier keine starken Auswirkungen. Allerdings muss beachtet werden, dass der Fokus bei der Auswahl der Satzpaare der sinnhaften Verknüpfung mehr auf der Verknüpfung lag. Die nicht verknüpften Satzpaare konnten, freier nach ihrer zeitlichen Dauer aus-

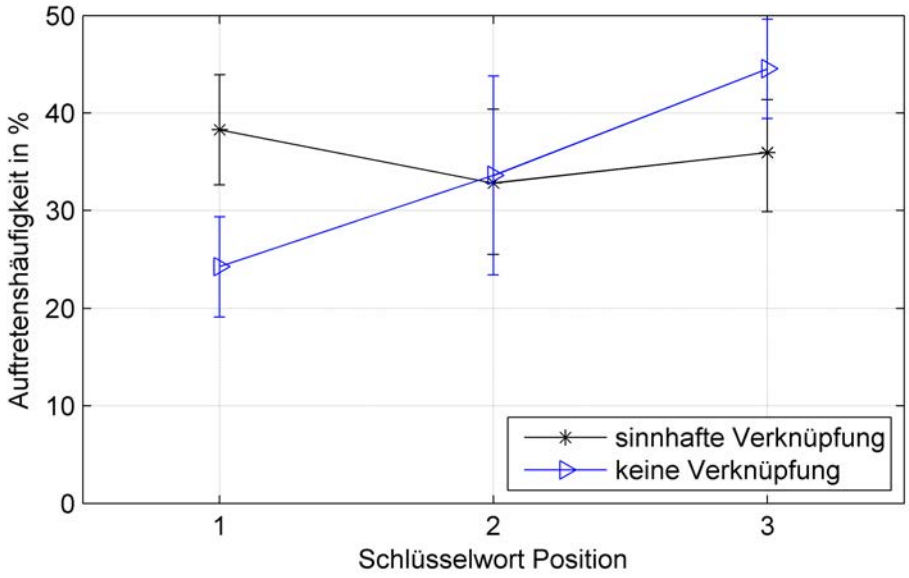


Abbildung 5.10: gemittelte Auftretenshäufigkeit der Wörter des Maskierertyps gleiche Stimme für die Positionen im Satz

gewählt werden, so dass diese öfters zeitlich weniger Längenunterschiede aufweisen. Dies könnte ein Grund für das unterschiedliche Verhalten der Sprachverständlichkeit bezüglich der beiden Positionen im Sat, an denen sinnhafte Verknüpfung vorhanden war, sein.

Dieses Ergebnis, das den Einfluss der semantischen Verküpfung zwischen 2 Wörtern zwischen Ziel- und Störsprecher trotz Konzentration auf den Zielsprecher zeigt, könnte die Annahme bestätigen, dass auf lexikalischer Ebene eine Sprachverarbeitung der Wörter des Störsprechers trotz fehlender Aufmerksamkeit auf ihn, erfolgt [26].

6 Zusammenfassung und Ausblick

Die Erhöhung der Grundfrequenz des Störsprechers zwischen 0 und 6 Halbtönen führten zu einer logarithmischen Erhöhung der Sprachverständlichkeit und Minderung des Einfluss von Informational Masking. Bei gleicher Stimme als Maskierer und bei -1.6 dB TMR konnten Ziel- und Störsprecher am wenigsten bis gar nicht auseinandergehalten werden. Informational Masking war hier am höchsten, offenbar als Folge von eher geringerer Segregationsfähigkeit zwischen Ziel- und Störsprecher und weniger durch Fehler der Selektion. Weitere Untersuchung zur genaueren Definition der Auftretensursache des IM sollten bezüglich des Top-Down-Prozesses durchgeführt werden, da dieser Prozess maßgeblich beteiligt für Informationsverarbeitung auf lexikalischer Ebene und die darauf basierenden Fehler ist.

Es wurde gezeigt, dass besonders bei geringen Pegelunterschieden zwischen Ziel- und Störsprecher von -1.6 dB TMR und -2.4 dB TMR und für hohe Sprecherähnlichkeiten (gleicher Sprecher als Maskierer mit gleicher Grundfrequenz) semantischer Kontext zwischen 2 Wörtern Auswirkungen auf die Sprachverständlichkeit hatte. Die semantische Mustererkennung des neuronalen Top-Down-Prozess beeinflusste hierbei die akustische Mustererkennung bezüglich der Stimmeigenschaften (Grundfrequenz, Prosodie, Laustärke) des Zielsprechers. Allerdings kann dieser Effekt durch zeitlich verschobene Positionierung der beiden Wörter stark gemindert werden.

Der Einfluss der Semantik auf die Sprachverständlichkeit bei hoher Sprecherähnlichkeit könnte im weiteren durch Tests mit Zielsprechersätzen des SUS-Prinzips und Störsprechersätzen die derselben Satzstruktur angehören, aber deren Wörter zueinander Sinn ergeben, untersucht werden. Der semantische Kontext zwischen den einzelenen

Wörtern wären im ganzen Satz des Störsprechers vorhanden. Inwieweit die lexikalische Sprachverarbeitung des Störsprechersatzes trotz Konzentration auf den Zielsprechersatz stattfindet und wie sehr der semantische Kontext zwischen den Wörtern Einfluss auf die Sprachverständlichkeit hätte, könnte somit in Abhängigkeit der Sprechereigenschaften und für das Informational Masking untersucht werden. Hierbei muss allerdings berücksichtigt werden, dass wie sich zeigte, unterschiedliche Syntax bei Sprache als Maskierer zu unterschiedlicher Sprachverständlichkeit führen kann.

Bei der Testdurchführung hat sich gezeigt, dass die Unsicherheit bei der Erkennung des Zielsprechers zwischen den Testpersonen unterschiedlich hoch war. Ebenso wurden die Unterschiede zwischen den verschiedenen Maskiererstimmen unterschiedlich stark wahrgenommen. Aus der stark variierenden Unsicherheit zwischen den Testpersonen resultierten hohe Schwankungen der Ergebnisse. Für weitere Test sollte daher mehr Testmaterial oder weniger aber dafür besonders ausgewählte Testpersonen eingesetzt werden, bei einer speziellen Berücksichtigung der Trainierbarkeit der Wahrnehmung über den Testverlauf.

Literaturverzeichnis

[1] Benoit, C., Grice, M., Hazan, V. (1996), „The SUS test: A method for assessment of text-to-speech sythesis intelligibility using Semantically Unpredictable Sentences.“ Speech Comm.18, 381-392

[2] Brand, T., Kollmeier, B. (2002), „Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysiscs and speech intelligbility tests.“ J. Acoust. Soc. Am. 111, 2801-2810

[3] Bregman, A. S. (1990), „Auditory scene analysis.“ MIT Press: Cambridge, MA

[4] Bronkhorst, A. (2000), „The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multi Talker Conditions.“ Acta Acustica utd w. Acustica 86, 117-128

[5] Bronkhorst, A. W., Bosman, A. J., Smoorenburg G. F. (1993), „A model for context effects in speech recognition.“ J. Acoust. Soc. Am. 93, 499-509

[6] Bronkhorst, A. W., Brand, T., Wagener, K. (2002), „Evaluation of context effects in sentence recognition.“ J.Acoust. Soc. Am. 93, 499-509

[7] Bronkhorst, A., Plomp, R. (1992), „A clinical test for the assessment of binaural speech perception in noise.“ Audiology 29, 275-285

[8] Brungart, D. S. (2001), „Informational and energetic masking effects in the perception of two simoultaneous talkers.“ J.Acoust. Soc. Am. 109, 1101-1109

[9] Brungart, D. S., Chang, P. S., Simpson, B. D., Wang, D. (2006), „Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation.“ J. Acoust. Soc. Am. 120(6), 4007-4018

[10] Bußmann, H. (2002), „Lexikon der Sprachwissenschaft.“ 3. Aufl. Kröner, Stuttgart

[11] Cherry, E. C. (1953), „Some experiements on the recognition of speech, with one and with two ears.“ J. Acoust. Soc. A. 25, 975-979

[12] Darwin, C. J., Brungart, D. S., Simpson, B. D. (2003), „Effects on fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers.“ J. Acoust. Soc. Am. 114, 2913-2921

[13] Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., Shinn-Cunningham, B. G. (2003), „Note on informational masking (L).“ J. Acoust. Soc. Am. 113, 2984-2987

[14] Ellis, D. P. W. (1996), „Prediction-driven computational auditory scene analysis.“ Ph.D. thesis, Dept. of Elec. Eng Comp. Sci., M.I.T.

[15] Euler, S. (2006), „Grundkurs Spracherkennung: Vom Sprachsignal zum Dialog-Grundlagen und Anwendungen verstehen- Mit praktischen Übungen.“ Friedr. Vieweg und Sohn Verlag — GWV Fachverlage, Wiesbaden

[16] Festen, J.M., Plomp, R. (1990), „Effects of fluctuating noise and interfering speech.“ J. Acoust. Soc. Am. 88(4), 1725-1736

[17] Hawley, M.L., Litovsky, R.Y., Culling, J.F. (2004), „The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer.“ J. Acoust Soc. Am. 115(2), 833-843

[18] Houtgast, T., Steeneken, H.J.M, (1973), „The modulation transfer function in room acoustics as a predictor of speech intelligibilty.“ Acustica 28, 66-73

[19] Ihlefeld, A., Shinn-Cunningham, B. (2008), „Spatial release from energetic and informational masking in a selective speech identification task.“ J. Acoust. Soc. Am. 123, 4369-4378

[20] Jekosch, U. (2005), „Voice and Speech Quality Perception: Assessment and Evaluation.“ Springer-Verlag, Heidelberg

[21] Kollmeier, B., Wesselkamp, M. (1997), „Development and evaluation of a German sentence test for objective and subjective speech intellgibility assessment.“ J. Acoust. Soc. Am. 102, 2412-2412

[22] O'Shaughnessy, D. (1987), „Speech Communication: Human and Machine.“ Reading, Massachusetts: Addison-Wesley Publishing Company

[23] Plomp, R., Mimpen, A.M. (1979), „Improving the Reliability of Testing the Speech Reception Threshold for Sentences.“ Audiology 18, 43-52

[24] Raake, A., Katz, B.F.G. (2006), „SUS-based Method for Speech Reception Threshold Measurements in French.“ Deutsche Telekom Laboratories, Berlin

[25] Rhebergen, K., Versfeld, N. (2005), „A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners.“ J. Acoust. Soc. Am. 117, 2181-92

[26] Rivenez, M., Darwin, C., Guillaume, A. (2004), „Perception of unattended speech.“ Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia

[27] Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, L. E. (2005), „Bottom-up and Top-down Influences on Spatial Unmasking.“ Acta Acustica utd w. Acustica 91, 1-13

[28] Wagenaar, W. A. (1969), „Note on the construction of diagram-balanced latin squares.“ Psychological Bulletin 72, 384-386

[29] Wagener, K., Kollmeier, B. (2004), „Göttinger und Oldenburger Satztest.“ Z. Audiol 43, 134-141

[30] Watson, C. S., (1987) „Uncertainty, informational maskingand the capacity of immediate auditory memory.-In: Auditory Processing of Complex Sounds.“ Erlbaum Hillsdale NJ 267-277

[31] Zwicker, E., Fastl, H. (1999), „Psychoacustics: Facts and Models.“ Springer-Verlag, Berlin Heidelberg

A Anhang

A.1 Wortliste aussortierter Wörter

Gott, Sack, Typ, Affe, Sex, Glied, brach, Boom, Tang, heil, Gas, braun, Schütz, rar, Ski, Tab, Wams, Matt, Fonds, prompt, Platt, Kuh, Schwein, Rechts, rechts, links, wehe, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, Grün, Gelb, Rot, Blau, Schwarz, Weiß, arg, krass, Kiel, Bund, West, Ost

A.2 Generierte und im Test verwendete Sätze (SUS)

```
'L01S01''Der Neid lernt die Wahl die lacht.'
'L01S02''Die Schau hofft unter der knappen Macht.'
'L01S03''Die Hand fällt mit dem dünnen Kurs.'
'L01S04''Der Flur pflegt in dem reifen Block.'
'L01S05''Wann dient das Pfand der tiefen Schnur?'
'L01S06''Das Spiel senkt die Haut die schaut.'
'L01S07''Wie teilt der Kreis das späte Bein?'
'L01S08''Der dichte Zoll weckt das Tuch.'
'L01S09''Der steile Berg schließt den Spruch.'
'L01S10''Der Pfau bricht den Marsch der singt.'
'L01S11''Der satte Schuss drückt den Tag.'
'L01S12''Wo trinkt der Elch den großen Stein?'
'L02S01''Der Plan liebt den Trost der schimpft.'
'L02S02''Wann spart die Uhr den scheuen Wolf?'
'L02S03''Die Bar spricht neben dem blinden Fleck.'
'L02S04''Der Trumpf springt bei dem lieben Sohn.'
'L02S05''Der März fällt in den dummen Sarg.'
'L02S06''Wann weiß der Kuss den schmalen Fall?'
'L02S07''Die Bucht drängt das Glas das singt.'
'L02S08''Der weiche Toast frisst die Naht.'
'L02S09''Der Zahn schiebt den Schlamm der denkt.'
'L02S10''Der kahle Draht übt den Bau.'
'L02S11''Der schlaue Kohl klappt das Blei.'
'L02S12''Wo fragt der Sinn den treuen Arzt?'
'L03S01''Der Plan schließt den Schrott der wirkt.'
'L03S02''Der Draht schreit auf dem lahmen Markt.'
'L03S03''Wann neigt der Wolf den schlappen Stall?'
'L03S04''Wo trinkt der Knick den scharfen Hund?'
```

'L03S05''Das Ei fasst den Trip der reist.'
'L03S06''Der hohle Schrank drückt den Sarg.'
'L03S07''Der flotte Chef trennt die Tat.'
'L03S08''Das Lied schwimmt bei dem flachen März.'
'L03S09''Das frische Bier beißt den Quatsch.'
'L03S10''Der Hass sitzt mit dem milden Akt.'
'L03S11''Wo droht die Bar dem glatten Arzt?'
'L03S12''Der Bau druckt den Schweiss der fehlt.'
'L04S01''Wie schützt der Fuchs das feine Geld?'
'L04S02''Das nette Team macht den Schmutz.'
'L04S03''Der Bach schmeckt in der stillen Flucht.'
'L04S04''Wo holt der Turm das platte Lob?'
'L04S05''Das Bier rührt mit der runden Kur.'
'L04S06''Der Akt würgt den Stall der glaubt.'
'L04S07''Das freie Lied prägt die Not.'
'L04S08''Der Sekt spürt den Chef der wirkt.'
'L04S09''Der nahe Hass liest den Schwung.'
'L04S10''Wo ruft das Gramm den neuen Schweiss?'
'L04S11''Das Hirn schaut bei dem leeren Kranz.'
'L04S12''Das Tor schreckt den Hund der stürzt.'
'L05S01''Die milde Jagd braucht den Bart.'
'L05S02''Wie senkt der Raum den flachen Witz?'
'L05S03''Der Test meint den Leib der sitzt.'
'L05S04''Der Marsch übt das Gold das fließt.'
'L05S05''Der Spruch bricht die Haut die schimpft.'
'L05S06''Wann füllt das Grab den schlichten Schnee?'
'L05S07''Der nahe Kampf prüft das Haus.'
'L05S08''Wie weiß die Macht den flotten Kurs?'
'L05S09''Das leere Rad dient dem Freund.'
'L05S10''Die Schau herrscht in der hübschen Schnur.'
'L05S11''Der Mensch blüht über dem neuen Mehl.'
'L05S12''Das Bein läuft bei dem kranken Brot.'
'L06S01''Der scharfe Hirsch hört den Arm.'
'L06S02''Der Sturz dreht das Bad das strömt.'
'L06S03''Das Vieh treibt den Stein der passt.'
'L06S04''Wann bringt der Knopf den warmen Kreis?'
'L06S05''Der Pfau klingt über dem glatten Kalk.'
'L06S06''Der Dank saugt neben dem lahmen Pfund.'
'L06S07''Wo wäscht das Glück das kurze Nest?'
'L06S08''Der Tag schmeißt das Spiel das geht.'
'L06S09''Wie stützt das Dorf den starken Flur?'
'L06S10''Das Kleid weilt mit dem nackten Pfand.'
'L06S11''Die kalte Bahn nutzt das Fleisch.'
'L06S12''Der süße Zoll plant den Stoß.'
'L07S01''Der Kern schaut vor dem schmalen Nest.'
'L07S02''Die kranke Front macht den Gast.'
'L07S03''Der hübsche Flug schreckt das Jahr.'

'L07S04''Wie würgt der Sinn den warmen Fleck?'
'L07S05''Der Spott stützt das Geld das tanzt.'
'L07S06''Der Golf greift den Stoß der stirbt.'
'L07S07''Wo rollt die Faust das kahle Ding?'
'L07S08''Der Schlamm lacht in dem blinden Prinz.'
'L07S09''Das Fleisch glaubt unter dem starken Mai.'
'L07S10''Der schlichte Turm wiegt das Dorf.'
'L07S11''Wo hört das Blei das süße Hemd?'
'L07S12''Das Hirn putzt die Flucht die stoppt.'
'L08S01''Die Hast bringt den Dom der liegt.'
'L08S02''Wann nutzt der Pfahl das heiße Pfund?'
'L08S03''Der Stamm läuft neben dem knappen Pfau.'
'L08S04''Wie wischt der Streich das feste Salz?'
'L08S05''Das lahme Kleid hilft dem Hang.'
'L08S06''Das Glück nennt den Biss der rührt.'
'L08S07''Wie stürzt das Hemd das stumme Ding?'
'L08S08''Die Front schwingt auf dem milden Brief.'
'L08S09''Der Tag fliegt mit dem glatten Spiel.'
'L08S10''Der dichte Schein treibt den Bruch.'
'L08S11''Der Flur hängt den Kern der geht.'
'L08S12''Das steile Bad kennt die Stadt.'
'L09S01''Wann hängt der Stock den fremden Wald?'
'L09S02''Der Gips lässt den Müll der geht.'
'L09S03''Das Salz spricht neben dem lauten Bruch.'
'L09S04''Der Wind läuft bei der stillen Hast.'
'L09S05''Der Pfahl trinkt den Schuh der strömt.'
'L09S06''Der neue Korb schließt den Dom.'
'L09S07''Wie zahlt die Stadt das schwache Zeug?'
'L09S08''Das Jahr senkt in dem schlanken Stamm.'
'L09S09''Die Faust nimmt den Streich der rennt.'
'L09S10''Wie neigt der Hals den nahen Start?'
'L09S11''Das schöne Kind plant den Park.'
'L09S12''Der liebe Mai wischt den Golf.'
'L10S01''Der Zahn löst den Trumpf der wirkt.'
'L10S02''Wo pflegt die Tat die stumme Hand?'
'L10S03''Wo merkt der Neid den dichten Leib?'
'L10S04''Die Jagd schießt unter dem breiten Block.'
'L10S05''Der flotte Trip sucht den Fall.'
'L10S06''Der scharfe Berg neigt das Grab.'
'L10S07''Der Schuss schwimmt neben dem reinen Test.'
'L10S08''Der Schrott brennt nach dem wahren Ei.'
'L10S09''Der Mensch streicht den Trost der klopft.'
'L10S10''Das knappe Rad beißt den Raum.'
'L10S11''Das Tuch gibt den Witz der fehlt.'
'L10S12''Wo löscht die Wahl den flachen Elch?'
'L11S01''Der hübsche Film dreht das Werk.'
'L11S02''Der Witz mischt das Maul das bleibt.'

'L11S03''Der schwache Raum lässt das Volk.'
'L11S04''Das Kreuz stützt den Fisch der schweigt.'
'L11S05''Der Bart schwingt bei der großen Jagd.'
'L11S06''Das tiefe Knie hört den Tank.'
'L11S07''Die Nacht rollt auf dem starken Schub.'
'L11S08''Wo packt der Schnee den kranken Scherz?'
'L11S09''Der Rat zielt mit dem schönen Tee.'
'L11S10''Wo fängt die Scheu den späten Freund?'
'L11S11''Der Leib riecht das Mehl das rührt.'
'L11S12''Wann zahlt das Erz den faulen Pass?'
'L12S01''Der Start macht das Paar das pflegt.'
'L12S02''Der Schein klopft auf dem weiten Müll.'
'L12S03''Der schnelle Film stellt den Arm.'
'L12S04''Der Ton herrscht vor dem alten Blatt.'
'L12S05''Der Sturz schluckt das Bild das schweigt.'
'L12S06''Die nette Scheu braucht den Hang.'
'L12S07''Wann wiegt das Fach den braven Zweig?'
'L12S08''Wo putzt der Busch den runden Kalk?'
'L12S09''Der platte Brief baut den Tee.'
'L12S10''Das Knie sieht den Trend der hofft.'
'L12S11''Wo nimmt der Biss den faulen Park?'
'L12S12''Der Dank bleibt neben dem klugen Stock.'
'L13S01''Der Zweig weilt bei dem lieben Tier.'
'L13S02''Der Saft liest das Schiff das packt.'
'L13S03''Der Wolf schießt mit dem reichen Lied.'
'L13S04''Die böse Wut sucht den Hof.'
'L13S05''Das laute Bier löst das Kinn.'
'L13S06''Wie löscht der Mond den fremden Akt?'
'L13S07''Das Paar druckt den Lohn der liegt.'
'L13S08''Die bunte Bar ruft den Hund.'
'L13S09''Wie droht der Busch den scheuen März?'
'L13S10''Die Milch zielt über den treuen Nerv.'
'L13S11''Das Blatt trennt den Mut der blüht.'
'L13S12''Wo spürt der Mann den armen Draht?'
'L14S01''Die Macht weint vor dem braven Bach.'
'L14S02''Der Trend schwebt über dem breiten Tuch.'
'L14S03''Der Ton stellt den Flug der passt.'
'L14S04''Der schöne Block nennt die Saat.'
'L14S05''Der Neid stirbt bei dem reinen Bein.'
'L14S06''Das faule Lob liebt den Spott.'
'L14S07''Wo fängt der Gast die kühne Haut?'
'L14S08''Wie sieht der Stress das schwache Tor?'
'L14S09''Der Kranz schmeißt den Trieb der prüft.'
'L14S10''Wann schiebt das Bild das alte Fach?'
'L14S11''Der Prinz wäscht die Schau die klappt.'
'L14S12''Der schlanke Stern fragt den Elch.'
'L15S01''Der Schuh füllt den Schmutz der tanzt.'

```
'L15S02''Die Saat fließt mit dem kalten Stern.'
'L15S03''Der frische Kranz fühlt das Team.'
'L15S04''Der Preis saugt neben dem stillen Holz.'
'L15S05''Die Not baut den Stamm der scheint.'
'L15S06''Der Bach lacht bei dem hohlen Gramm.'
'L15S07''Die nackte Schlacht bringt den Staat.'
'L15S08''Wie nutzt das Salz den kurzen Stress?'
'L15S09''Das Lob schluckt den Trieb der klingt.'
'L15S10''Wann braucht das Tor den freien Bruch?'
'L15S11''Wo kennt der Wind den schlappen Wald?'
'L15S12''Der feine Streich meint das Kind.'
'L16S01''Das Bild lernt den Lohn der steigt.'
'L16S02''Der Flug stoppt auf dem treuen Scherz.'
'L16S03''Das Fach mischt das Jahr das springt.'
'L16S04''Der freie Spott kennt den Start.'
'L16S05''Die frohe Faust schwingt den Hof.'
'L16S06''Wie teilt der Mond das schiefe Volk?'
'L16S07''Wo hilft der Rat dem leeren Park?'
'L16S08''Wann rollt der Müll den dicken Tank?'
'L16S09''Der Golf riecht den Stock der fliegt.'
'L16S10''Der Mai denkt neben dem bunten Prinz.'
'L16S11''Die Milch fällt bei dem feinen Gast.'
'L16S12''Der scheue Ton streicht den Trend.'
'L17S01''Die Wut lernt die Nacht die brennt.'
'L17S02''Der lange Nerv drängt den Stall.'
'L17S03''Wie weckt das Knie den kühnen Saft?'
'L17S04''Das Blatt steigt vor dem breiten Mut.'
'L17S05''Das krumme Erz gibt den Plan.'
'L17S06''Der Sarg frisst den Hass der saugt.'
'L17S07''Der Tee streicht das Werk das schießt.'
'L17S08''Das reine Paar teilt den Chef.'
'L17S09''Wo merkt der Pass den süßen Zweig?'
'L17S10''Die Scheu fließt auf dem warmen Busch.'
'L17S11''Wann spart der Film den schlichten Bau?'
'L17S12''Der Schweiß schmeckt auf dem wahren Arzt.'
'L18S01''Die hohle Hand fragt den Hirsch.'
'L18S02''Der Preis druckt das Pfund das passt.'
'L18S03''Die dumme Schnur löscht die Wahl.'
'L18S04''Der Spruch weilt bei der blinden Not.'
'L18S05''Wann putzt das Bad das laute Glück?'
'L18S06''Das fremde Gramm greift den Staat.'
'L18S07''Der Kurs wiegt das Vieh das herrscht.'
'L18S08''Das Kleid klopft neben dem bunten Schmutz.'
'L18S09''Wann spart der Berg das nackte Holz?'
'L18S10''Wie sucht die Bahn den schmalen Schuss?'
'L18S11''Der Marsch liebt den Knopf der stirbt.'
'L18S12''Die Schlacht stoppt auf dem schlappen Team.'
```

'L19S01''Der schlanke Schnee drückt den Freund.'
'L19S02''Wann wischt der Sohn die dünne Bucht?'
'L19S03''Das Haus schmeißt das Brot das fasst.'
'L19S04''Das Zeug hängt den Pfahl der schwebt.'
'L19S05''Der Knick steigt in den alten Dom.'
'L19S06''Der feste Kampf nennt die Hast.'
'L19S07''Das Mehl brennt den Kohl der weint.'
'L19S08''Die Naht schmeckt bei dem reifen Schrank.'
'L19S09''Wo plant der Korb das satte Gold?'
'L19S10''Die Stadt reist über dem braven Gips.'
'L19S11''Der stumme Markt beißt den Bart.'
'L19S12''Wie wäscht der Hals den heißen Quatsch?'
'L20S01''Die Kur rennt in dem festen Glas.'
'L20S02''Das steile Rad schluckt den Sekt.'
'L20S03''Das Blei prägt den Fleck der weint.'
'L20S04''Die Flucht strömt über den krummen Sinn.'
'L20S05''Der Turm schimpft nach dem schlauen Test.'
'L20S06''Wie ruft das Gold das weiche Brot?'
'L20S07''Der Mensch spürt den Toast der springt.'
'L20S08''Der lange Fuchs baut das Hirn.'
'L20S09''Wie übt der Schwung den kühnen Schlamm?'
'L20S10''Wo fühlt das Geld das dumme Grab?'
'L20S11''Das Haus glaubt dem Kampf der packt.'
'L20S12''Die heiße Uhr bricht den Kuss.'
'L21S01''Das Pfand dient in dem runden Kreis.'
'L21S02''Wie mischt das Schiff das reiche Kind?'
'L21S03''Der Fisch dient den Hirsch der scheint.'
'L21S04''Das Kinn schweigt über dem schiefen Knopf.'
'L21S05''Der platte Wald singt den Zoll.'
'L21S06''Der Gips weiß das Zeug das klingt.'
'L21S07''Der Mann holt das Maul das schwebt.'
'L21S08''Wann liest der Schub das reife Tier?'
'L21S09''Die Bahn zielt vor dem großen Wind.'
'L21S10''Der tiefe Hals riecht den Stein.'
'L21S11''Wann schützt das Vieh den armen Schuh?'
'L21S12''Der kluge Korb dreht das Kreuz.'
'L22S01''Wann greift der Mut den schnellen Biss?'
'L22S02''Das Werk fliegt mit dem netten Schein.'
'L22S03''Wann stürzt das Ding die dicke Nacht?'
'L22S04''Der Kern schreckt die Wut die hofft.'
'L22S05''Der böse Pass hilft dem Kalk.'
'L22S06''Das Dorf rennt vor dem weiten Stoß.'
'L22S07''Das wahre Nest würgt das Hemd.'
'L22S08''Das Erz treibt die Front die scheint.'
'L22S09''Das lange Fleisch fühlt den Sturz.'
'L22S10''Wann füllt der Dank den frohen Hang?'
'L22S11''Der Saft liegt unter dem krummen Brief.'

```
'L22S12''Der Arm meint den Nerv der schreit.'
'L23S01''Wo gibt die Milch das dünne Kreuz?'
'L23S02''Das späte Maul fängt das Schiff.'
'L23S03''Der schiefe Hof stellt den Schub.'
'L23S04''Das Kinn nimmt die Uhr die sitzt.'
'L23S05''Das satte Tier zahlt den Fisch.'
'L23S06''Der Tank blüht in dem frohen Mond.'
'L23S07''Der Sekt fehlt über dem armen Toast.'
'L23S08''Der Lohn denkt mit dem bösen Mann.'
'L23S09''Die Kur weckt den Schwung der prüft.'
'L23S10''Wo sieht der Fuchs den dicken Rat?'
'L23S11''Wie merkt der Kuss den reichen Scherz?'
'L23S12''Das Glas lässt das Volk das klappt.'
'L24S01''Die schlaue Naht löst den Staat.'
'L24S02''Der Fall tanzt vor dem schnellen Schrott.'
'L24S03''Die Tat spricht in dem kurzen Trumpf.'
'L24S04''Die frische Schlacht drängt den Kohl.'
'L24S05''Der Knick holt den Trost der schreit.'
'L24S06''Wie trennt der Preis den kahlen Sohn?'
'L24S07''Wo schützt die Bucht den klugen Zahn?'
'L24S08''Der Trip schwimmt mit dem kalten Holz.'
'L24S09''Der Trieb schiebt den Markt der bleibt.'
'L24S10''Der Quatsch frisst den Stern der reist.'
'L24S11''Die weiche Saat prägt das Ei.'
'L24S12''Wie droht der Stress dem weiten Schrank?'
```

A.3 Programmcode

Die Rechtschreibprüfung fand nach der Homophonüberprüfung unter MATLAB statt.
Ein Auszug aus dem Progagammcode veranschaulicht dies.

```
refStr(1,1)=lower(refStr(1,1)); %kleinschreibweise
cellrefStr=cellstr([refStr{1,1} ' (cmh_sa01) \r\n']);  %vorlage der schluesselwoert

hypStr(1,1)=lower(hypStr(1,1));
cellhypStr=cellstr([hypStr{1,1} ' (cmh_sa01) \r\n']); %eingabe testperson

fRefStr=CellArray2File(cellrefStr,'f','refD.txt'); %in txt.format zum auslesen in S
fHypStr=CellArray2File(cellhypStr,'f','hypD.txt');

cd('c:\cygwin\usr\local\bin');
!sclite -s -r {'D:\Intel_Test\Program\refD.txt'} trn -h {'D:\Intel_Test\Program\hyp
cd('D:\Intel_Test\Program');
```

A.4 Information zur Vorbereitung der Testpersonen

Die Anleitung mit der die Testperson auf den Sprachverständlichkeitstest vorbereitet wurden, sind im Folgenden für alle 3 Tests aufgeführt.

A.4.1 Anleitung: erster Test zur SRT-Ermittlung

Test zur Worterkennung

Vielen Dank für Ihre Teilnahme an diesem Experiment. Bitte nehmen Sie sich ein paar Minuten, um die folgenden Hinweise zu lesen. Bitte zögern Sie nicht nach dem Lesen Ihre Fragen zu stellen.

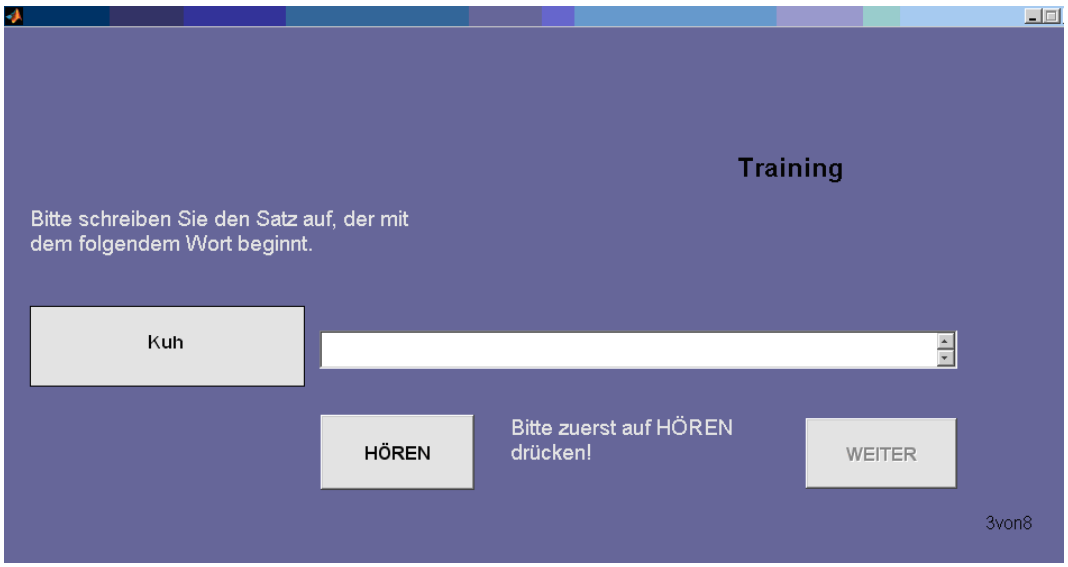
Sie nehmen heute an einem Test zur Worterkennung teil. Dieser Test dauert ca. 1h. Ihre Aufgabe ist es so viele Wörter wie möglich zu verstehen.

Die Sätze, die verstanden werden sollen, werden von einem männlichen Sprecher gesprochen. Diese Sätze sind auf Deutsch. Sie sind grammatikalisch korrekt, ergeben aber keinen Sinn. Jeder Satz besteht aus geläufigen, kurzen Wörtern. Ihre Aufgabe hierbei ist es die verstandenen Hauptwörter, von denen insgesamt immer 4 in jedem Satz vorhanden sind und Adjektive, Nomen und Verben sein können, einzugeben.

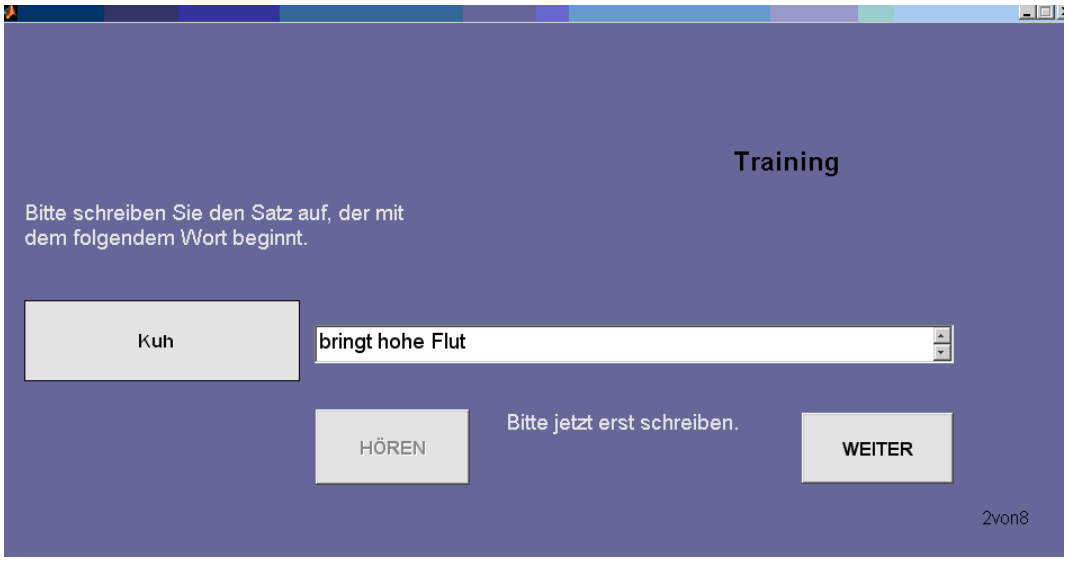
Die Lautstärke des Sprechers kann für jeden Satz unterschiedlich sein. Zudem wird der Sprecher von einem anderen Sprecher (männlich oder weiblich) oder von Rauschen überlagert. Die Lautstärke der störenden Sprecher/Rauschen bleibt allerdings konstant.

Bitte seien Sie daher vorbereitet, dass es nicht einfach sein wird die Sätze des Sprechers im Ganzen zu verstehen.

Jeder Satz wird Ihnen einzeln vorgespielt, durch Drücken der ¡Hören¿ Taste. Das erste Wort, eines jeden Satzes den sie verstehen sollen, wird Ihnen in dem Feld vor der Eingabezeile angezeigt.



Sie werden nach dem Abspielen des Satzes gebeten die verstandenen Wörter in die Eingabezeile einzutippen. Bitte hören Sie sich den Satz bis zu Ende an, bevor Sie Ihre Eingabe tätigen. Danach fahren Sie bitte fort mit ¡Weiter¿ und der nächste Satz ist bereit abgespielt zu werden.



Das erste vorgegebene Wort muss nicht mitangegeben werden. Es interessieren lediglich die weiteren 3 Hauptwörter (siehe Abbildung oben).

Bei Ihrer Eingabe bitten wir Sie Zeichensetzung (; ? !) und Artikel zu vermeiden, da sie Ihnen Zeit und Energie kosten. Es wird Wert auf die genannten Hauptwörter gelegt. Ebenso wird Groß- und Kleinschreibung nicht berücksichtigt.

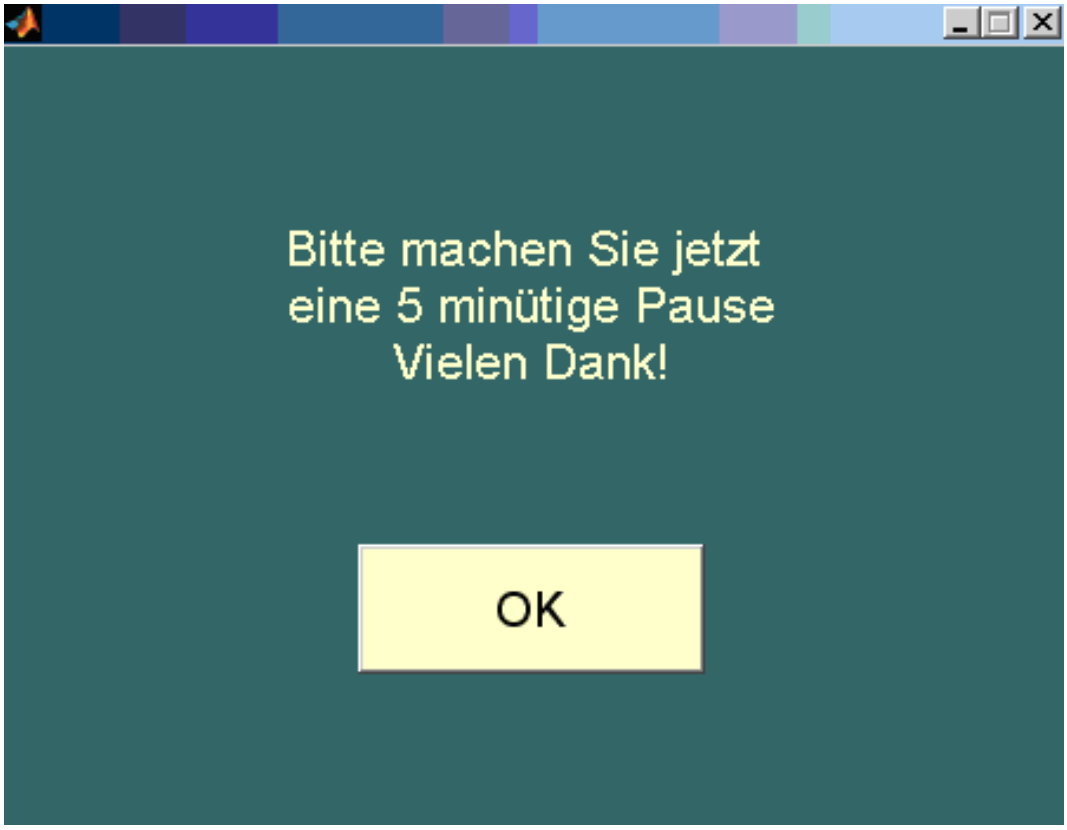
Es kann vorkommen, dass Sie nichts verstanden haben, bitten geben Sie in diesem Fall

ein x' ein.

Falls Ihnen bei der Eingabe ein oder mehrere Rechtschreibfehler unterlaufen sind, wird einmalig pro Satz eine Fehlermeldung ausgegeben. Sie werden dann gebeten die verstandenen Wörter erneut in korrekter Schreibweise einzugeben. Nach der Eingabe fahren Sie bitte fort mit ¡OK¿.

Der Testablauf besteht aus einer Trainingsphase und 5 Abschnitten. Am Anfang eines jeden Abschnitts wird die Lautstärke des Zielsprechers an Ihr Hörverständnis angepasst. Dafür wird Ihnen ein Satz vorgespielt, dessen Verständlichkeit sie nach jedem Abspielen erhöhen können, indem Sie den Knopf ¡Hören¿ betätigen. Wenn Sie das Gefühl haben 2 der 4 Hauptwörter zu verstehen, fahren Sie bitte fort im Programm mit ¡Weiter¿.

Nach der Trainingsphase und dem 2. Abschnitt wird eine Pause von 5 min eingelegt. Ihnen wird dies in Form des folgenden Fensters mitgeteilt. Wenn Sie bereit sind fortzufahren drücken Sie dann bitte auf ¡ OK¿. Sie können wenn Sie möchten in der Zeit das Zimmer verlassen.



Nach dem 5. Abschnitt ist der Test vorbei, was Ihnen dann mit dem entsprechendem Fenster mitgeteilt wird.

Vielen Dank für Ihre Mitarbeit und viel Erfolg.

A.4.2 Anleitung: zweiter Test zur Ermittlung der Sprachverständlichkeit mit festen Pegeln

Test zur Worterkennung

Vielen Dank für Ihre Teilnahme an diesem Experiment. Bitte nehmen Sie sich ein paar Minuten, um die folgenden Hinweise zu lesen. Bitte zögern Sie nicht nach dem Lesen Ihre Fragen zu stellen.

Sie nehmen heute an einem Test zur Worterkennung teil. Dieser Test dauert ca. 1h. Ihre Aufgabe ist es so viele Wörter wie möglich zu verstehen.

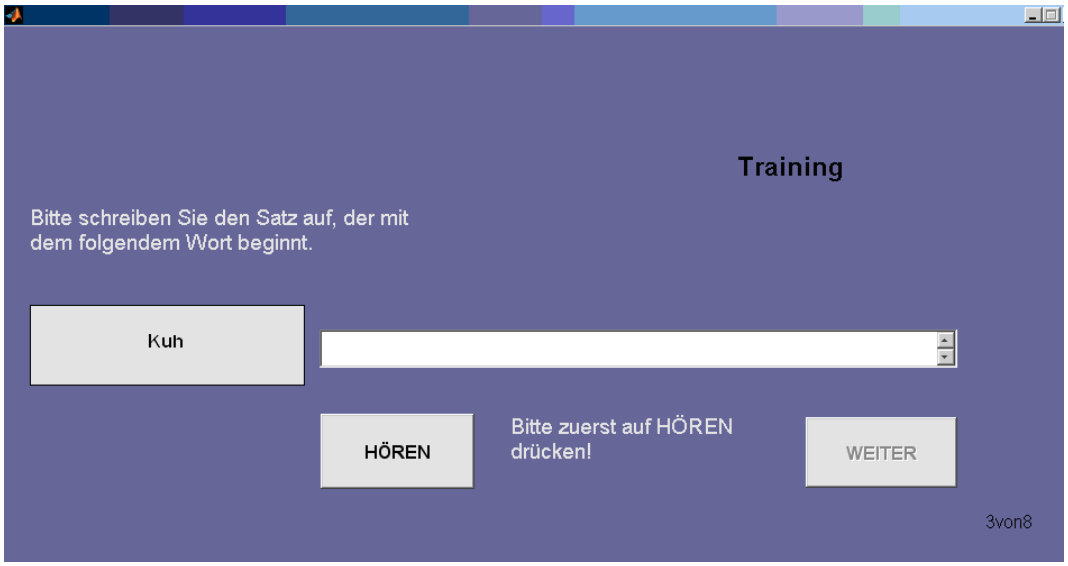
Die Sätze, die verstanden werden sollen, werden von einem männlichen Sprecher gesprochen. Diese Sätze sind auf Deutsch. Sie sind grammatikalisch korrekt, ergeben aber keinen Sinn. Jeder Satz besteht aus geläufigen, kurzen Wörtern. Ihre Aufgabe hierbei ist es die verstandenen Hauptwörter, von denen insgesamt immer 3 in jedem Satz vorhanden sind und Adjektive, Nomen und Verben sein können, einzugeben.

Die Lautstärke des Sprechers ist für jeden Satz unterschiedlich. Zudem wird der Sprecher von einem anderen Sprecher (männlich oder weiblich) oder von Rauschen überlagert.

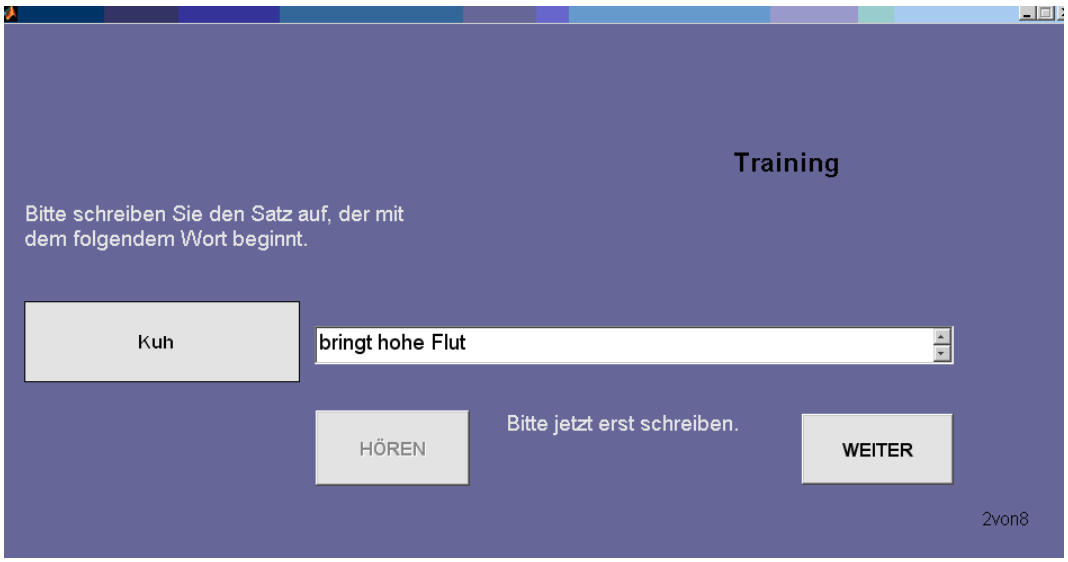
Die Lautstärke der störenden Sprecher/Rauschen bleibt allerdings konstant.

Bitte seien Sie daher vorbereitet, dass es nicht einfach sein wird die Sätze des Sprechers im Ganzen zu verstehen.

Jeder Satz wird Ihnen einzeln vorgespielt, durch Drücken der ¡Hören¿ Taste. Das erste Wort, eines jeden Satzes den sie verstehen sollen, wird Ihnen in dem Feld vor der Eingabezeile angezeigt.



Sie werden nach Abspielen des Satzes gebeten die verstandenen Wörter in die Eingabezeile einzutippen. Bitte hören Sie sich den Satz bis zu Ende an, bevor Sie Ihre Eingabe tätigen. Danach fahren Sie bitte fort mit ¡Weiter¿ und der nächste Satz ist bereit abgespielt zu werden.



Das erste vorgegebene Wort muss nicht mitangegeben werden. Es interessieren lediglich die weiteren 3 Hauptwörter (siehe Abbildung oben).

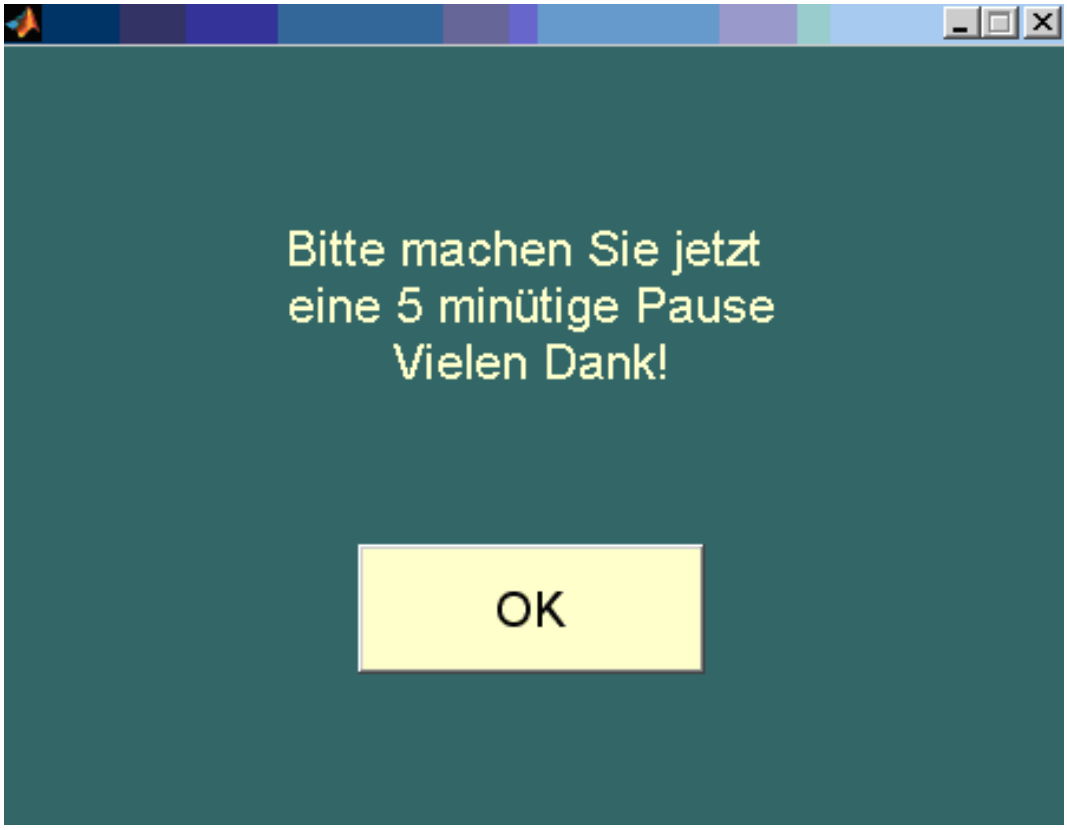
Bei Ihrer Eingabe bitten wir Sie Zeichensetzung (; ? !) und Artikel zu vermeiden, da sie Ihnen Zeit und Energie kosten. Es wird Wert auf die genannten Hauptwörter gelegt. Ebenso wird Groß- und Kleinschreibung nicht berücksichtigt.

Es kann vorkommen, dass Sie nichts verstanden haben, bitten geben Sie in diesem Fall ein x' ein.

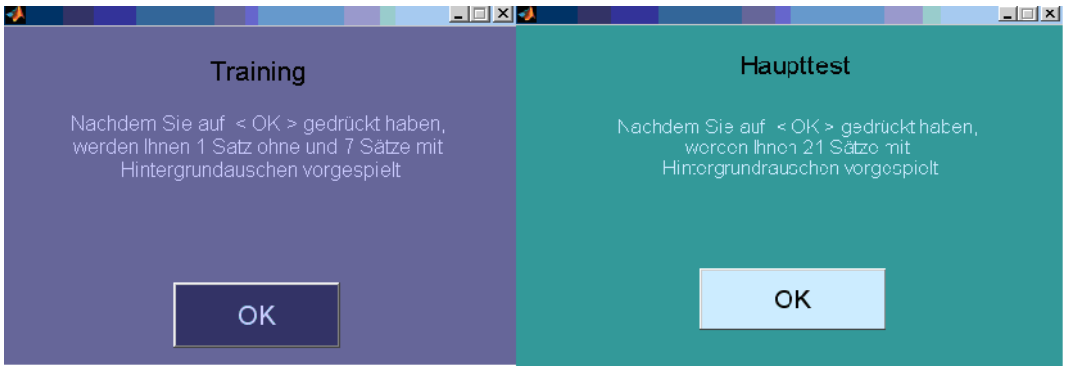
Falls Ihnen bei der Eingabe ein oder mehrere Rechtschreibfehler unterlaufen sind, wird einmalig pro Satz eine Fehlermeldung ausgegeben. Sie werden dann gebeten die verstandenen Wörter erneut in korrekter Schreibweise einzugeben. Nach der Eingabe fahren Sie bitte fort mit ¡OK¿.

Der Testablauf besteht aus 5 Abschnitten und 2 Pausen.

Nach dem 1. und 3. Abschnitt wird eine reguläre Pause von 5 min eingelegt. Ihnen wird dies in Form des folgenden Fensters mitgeteilt. Wenn Sie bereit sind fortzufahren drücken Sie dann bitte auf ¡ OK¿. Sie können wenn Sie möchten in der Zeit das Zimmer verlassen.

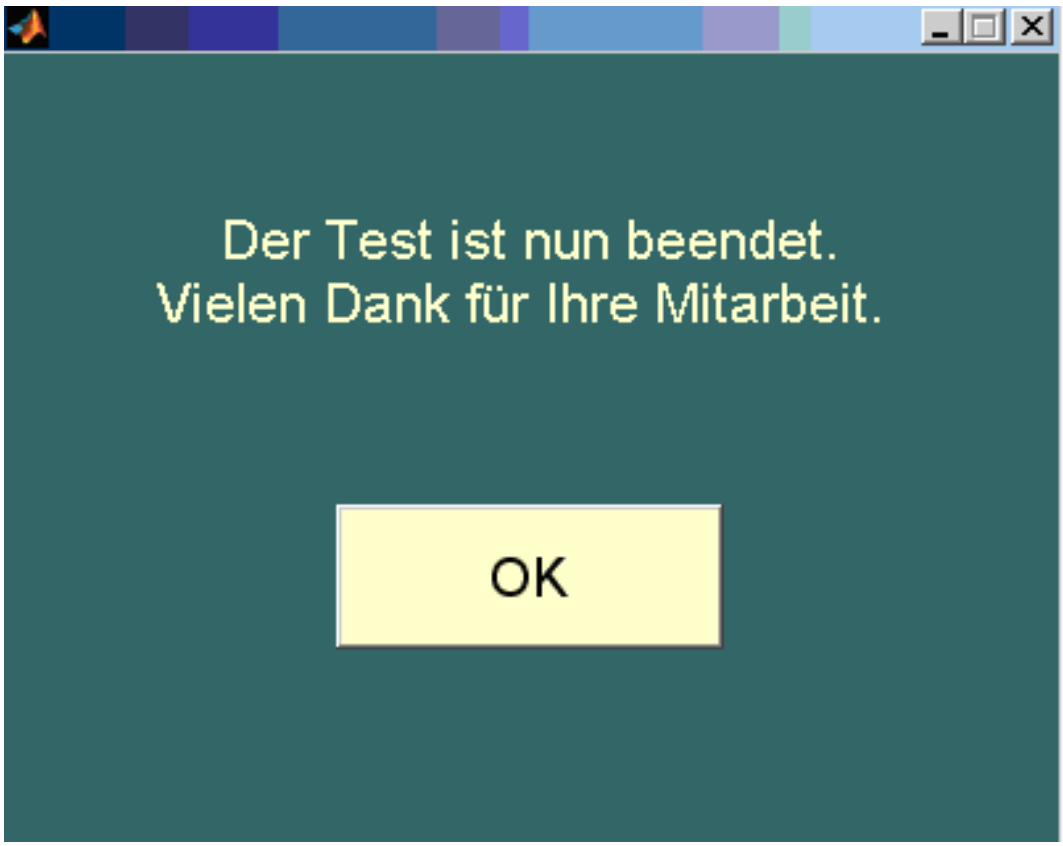


Jeder dieser 5 Abschnitte besteht aus einer Trainingsphase von jeweils 8 Sätzen und einer Testphase von jeweils 21 Sätzen.



Diese Fenster sollen Sie auf die darauffolgende Anzahl der Sätze und auf die Form des Störers vorbereiten. Danach folgt das bereits beschriebene Eingabefenster.

Nach dem 5. Abschnitt ist der Test vorbei, was Ihnen dann mit dem entsprechendem Fenster mitgeteilt wird.



Vielen Dank für Ihre Mitarbeit und viel Erfolg.

A.4.3 Anleitung: dritter Test zusätzlich mit sinnbehafteten Satzpaarkombinationen

Test zur Worterkennung

Vielen Dank für Ihre Teilnahme an diesem Experiment. Bitte nehmen Sie sich ein paar Minuten, um die folgenden Hinweise zu lesen. Bitte zögern Sie nicht nach dem Lesen Ihre Fragen zu stellen.

Sie nehmen heute an einem Test zur Worterkennung teil. Dieser Test dauert ca. 1h. Ihre Aufgabe ist es so viele Wörter wie möglich zu verstehen.

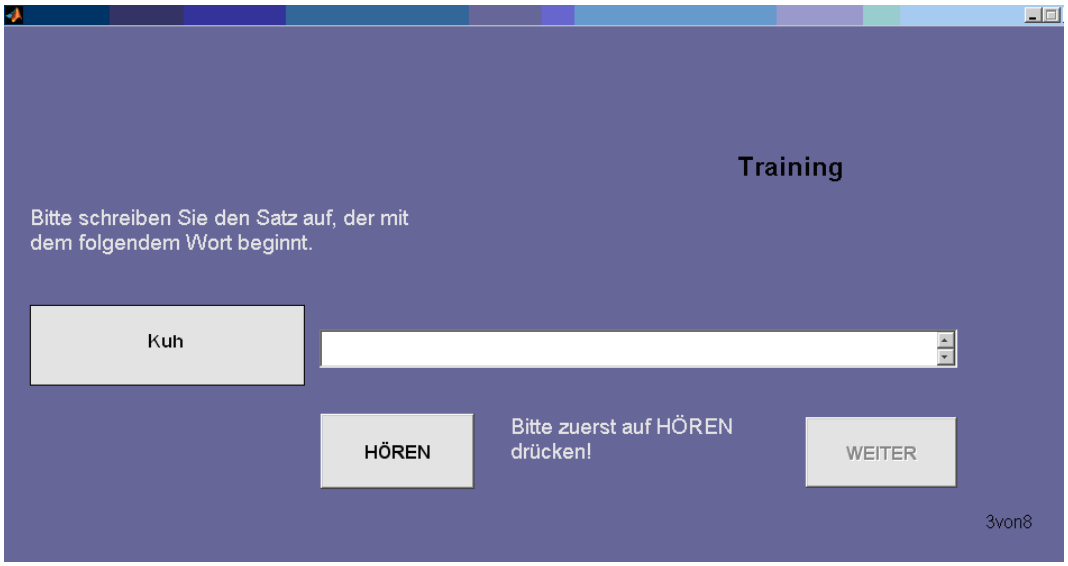
Die Sätze, die verstanden werden sollen, werden von einem männlichen Sprecher gesprochen. Diese Sätze sind auf Deutsch. Sie sind grammatikalisch korrekt, ergeben aber keinen Sinn. Jeder Satz besteht aus geläufigen, kurzen Wörtern. Ihre Aufgabe hierbei ist es die verstandenen Hauptwörter, von denen insgesamt immer 4 in jedem Satz vorhanden sind und Adjektive, Nomen und Verben sein können.

Die Lautstärke des Sprechers ist für jeden Satz unterschiedlich. Zudem wird der Sprecher von einem anderen Sprecher (männlich) oder von Rauschen überlagert. Die Lautstärke

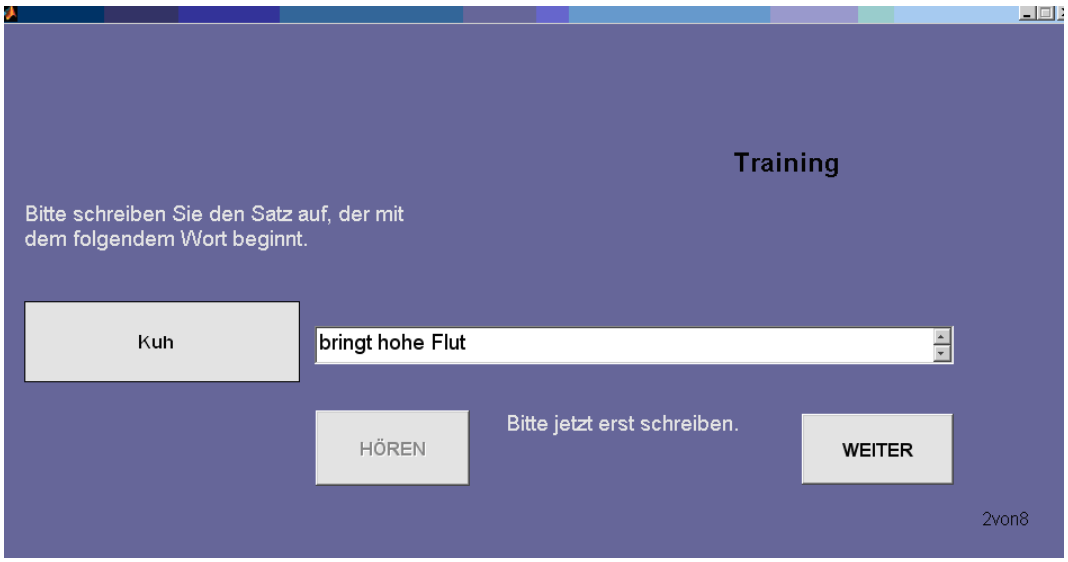
der störenden Sprecher/Rauschen bleibt allerdings konstant.

Bitte seien Sie daher vorbereitet, dass es nicht einfach sein wird die Sätze des Sprechers im Ganzen zu verstehen.

Jeder Satz wird Ihnen einzeln vorgespielt, durch Drücken der ¡Hören¿ Taste. Das erste Wort, eines jeden Satzes den sie verstehen sollen, wird Ihnen in dem Feld vor der Eingabezeile angezeigt.



Sie werden nach Abspielen des Satzes gebeten die verstandenen Wörter in die Eingabezeile einzutippen. Bitte hören Sie sich den Satz bis zu Ende an, bevor Sie Ihre Eingabe tätigen. Danach fahren Sie bitte fort mit ¡Weiter¿ und der nächste Satz ist bereit abgespielt zu werden.



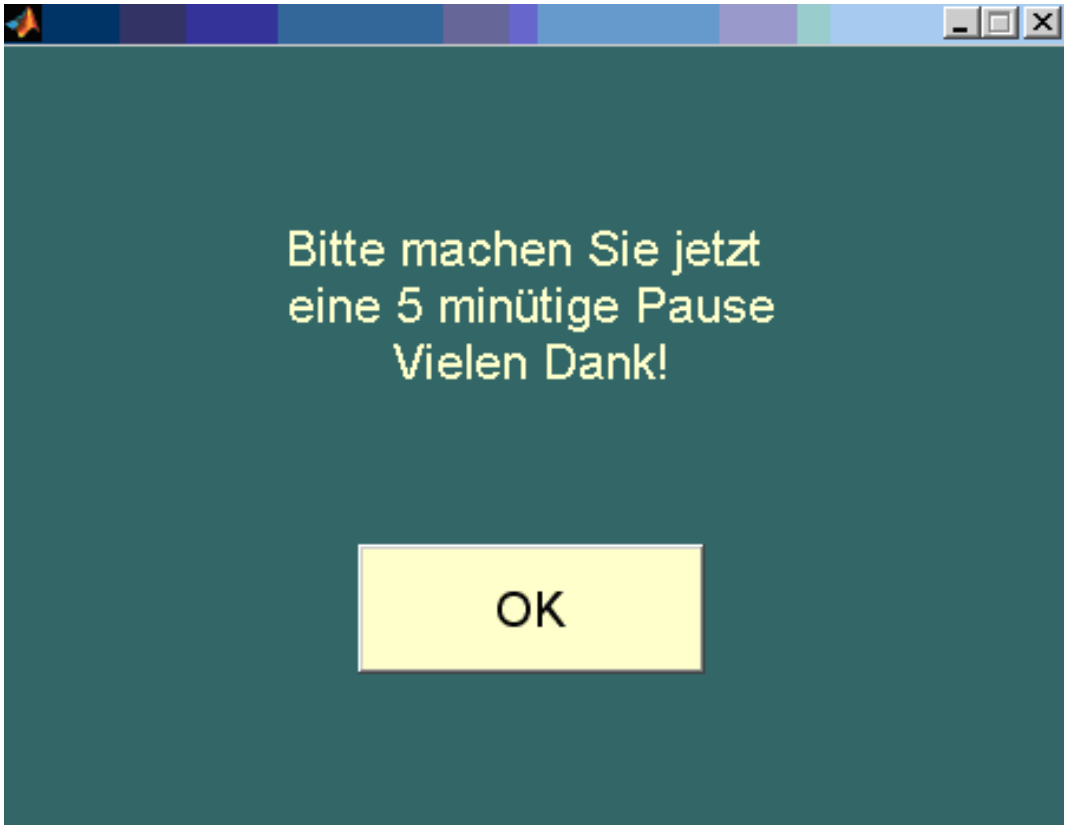
Das erste vorgegebene Wort muss nicht mitangegeben werden. Es interessieren lediglich die weiteren 3 Hauptwörter (siehe Abbildung oben).

Bei Ihrer Eingabe bitten wir Sie Zeichensetzung (; ? !) und Artikel zu vermeiden, da sie Ihnen Zeit und Energie kosten. Es wird Wert auf die genannten Hauptwörter gelegt. Ebenso wird Groß- und Kleinschreibung nicht berücksichtigt.

Es kann vorkommen, dass Sie nichts verstanden haben, bitten geben Sie in diesem Fall ein x' ein.

Falls Ihnen bei der Eingabe ein oder mehrere Rechtschreibfehler unterlaufen sind, wird einmalig pro Satz eine Fehlermeldung ausgegeben. Sie werden dann gebeten die verstandenen Wörter erneut in korrekter Schreibweise einzugeben. Nach der Eingabe fahren Sie bitte fort mit ¡OK¿.

Der Testablauf besteht aus einer Trainingsphase und 6 Abschnitten. Nach dem 3. Abschnitt wird eine Pause von 5 min eingelegt. Ihnen wird dies in Form des folgenden Fensters mitgeteilt. Wenn Sie bereit sind fortzufahren drücken Sie dann bitte auf ¡OK¿. Sie können wenn Sie möchten in der Zeit das Zimmer verlassen.



Diese Fenster sollen Sie auf die darauffolgende Anzahl der Sätze und auf die Form des Störers vorbereiten. Danach folgt das bereits beschriebene Eingabefenster.

Nach dem 6. Abschnitt ist der Test vorbei, was Ihnen dann mit dem entsprechendem Fenster mitgeteilt wird.

Vielen Dank für Ihre Mitarbeit und viel Erfolg.