

Aus der Klinik und Poliklinik für Hals-, Nasen- und Ohren-  
krankheiten, plastische und ästhetische Operationen  
der Universität Würzburg

Direktor: Prof. Dr. med. Rudolf Hagen

**Evaluation des Einsilber-Sprachmaterials M-2007  
und Entwurf einer Methodik für die  
Zusammenstellung gleichwertiger Listen**

Inaugural-Dissertation  
zur Erlangung der Doktorwürde der  
Medizinischen Fakultät  
der  
Julius-Maximilians-Universität Würzburg  
vorgelegt von

Juliane Friederike Qualen

aus Kiel

Würzburg, Dezember 2010



Referent: Prof. Dr. med. Joachim Müller  
Korreferentin: Prof. Dr. rer. nat. Kathleen Wermke  
Dekan: Prof. Dr. med. Matthias Frosch

Tag der mündlichen Prüfung:

Die Promovendin ist Zahnärztin



Für meine Eltern  
Diese Arbeit ist ihnen in Liebe und Dankbarkeit gewidmet



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Entwicklung, Nutzen und Anwendung der Sprachaudiometrie . . . . .	1
1.2	Der Freiburger Einsilberverständnistest . . . . .	7
1.2.1	Voraussetzungen, Durchführung und Merkmale . . . . .	7
1.2.2	Einflussgrößen des Freiburger Sprachverständnistests . . . . .	9
1.2.3	Kritikpunkte am Freiburger Sprachverständnistest . . . . .	10
1.2.4	Methodik zur Erstellung der Listen . . . . .	14
1.3	Evaluation von Sprachtests: Durchführung und Merkmale . . . . .	16
1.4	Anforderungen an den neuen Einsilber-Sprachverständnistest . . . . .	17
<b>2</b>	<b>Versuchsplanung und Methodik</b>	<b>19</b>
2.1	Aufsprache der fehlenden Einsilber . . . . .	20
2.1.1	Räumlichkeit, Aufbau und verwendete Gerätschaften . . . . .	20
2.1.2	Durchführung der Aufsprache . . . . .	21
2.1.3	Schneiden der Wörter, Abgleichen des SPL und anschließende Qualitätskontrollen . . . . .	23
2.2	Versuchsvorbereitungen für die Evaluation des neuen Sprachmaterials . . . . .	25
2.2.1	Belegen der Einsilber mit einem SNR von -8 dB . . . . .	26
2.2.2	Erstellung der „silence-padded“ Versionen der WAV-Dateien . . . . .	28
2.2.3	Listenerstellung: Mechanismus der Verwürfelung . . . . .	29
2.3	Evaluation des Sprachmaterials M-2007 . . . . .	31
2.3.1	Auswahl der Probanden . . . . .	31
2.3.2	Räumlichkeiten und verwendete Gerätschaften . . . . .	32
2.3.3	Durchführung der Evaluation . . . . .	33
2.4	Methodik der statistischen Auswertung . . . . .	34
2.5	Zusammenstellung der Listen . . . . .	37
2.5.1	Vorbereitung . . . . .	37
2.5.2	Vorgehensweise . . . . .	39

<b>3</b>	<b>Ergebnisse</b>	<b>49</b>
3.1	Auswertung der Versuche . . . . .	49
3.1.1	Einschätzung des Schwierigkeitsgrades aller einsilbigen Testwörter . . . . .	49
3.1.2	Einzelauswertung der Probanden / Probandenleistung . . . . .	50
3.1.3	Einflussfaktoren der Versuchsergebnisse . . . . .	52
3.1.4	Vergleich der Versuchsergebnisse dieser Studie mit den Versuchsergebnissen der Pilotstudie . . . . .	72
3.1.5	Anzahl der verfügbaren Wörter . . . . .	76
3.2	Verteilung der Wörter in Listen . . . . .	79
<b>4</b>	<b>Diskussion</b>	<b>85</b>
4.1	Stellungnahme zu den Kritikpunkten . . . . .	85
4.2	Erweiterungs- und Verbesserungsvorschläge für die Zukunft . . . . .	91
4.2.1	Entwicklung des Tests . . . . .	91
4.2.2	Aufsprache . . . . .	93
4.2.3	Vorbereitung des Testmaterials . . . . .	93
4.2.4	Bildung der Listen . . . . .	94
4.2.5	Das Prüfsystem . . . . .	96
4.2.6	Auswertung . . . . .	97
<b>5</b>	<b>Ausblick</b>	<b>99</b>
<b>6</b>	<b>Zusammenfassung</b>	<b>101</b>
<b>A</b>	<b>Verzeichnis verwendeter Abkürzungen</b>	<b>103</b>
<b>B</b>	<b>Probandenaufklärung</b>	<b>113</b>
B.1	Informationen zu den Versuchsdurchführungen . . . . .	113
B.2	Einverständniserklärung . . . . .	114
<b>C</b>	<b>Listen</b>	<b>115</b>
<b>D</b>	<b>Unbekannte Wörter</b>	<b>121</b>
<b>E</b>	<b>Danksagung</b>	<b>127</b>
	<b>Lebenslauf</b>	<b>129</b>



# Abbildungsverzeichnis

2.1	Tonschwellenaudiogramme der 20 Probanden . . . . .	31
3.1	Histogramm der Wörterzahl je nach richtigen Antworten . . . . .	50
3.2	Boxplot der Abhängigkeit korrekter Antworten von dem Geschlecht . . . . .	53
3.3	Linienplot der Korrektheit im Zeitverlauf des Versuches . . . . .	55
3.4	Linienplot der Korrektheit im Zeitverlauf der Blöcke . . . . .	56
3.5	Linienplot der Korrektheit im Zeitverlauf der Blöcke, ohne Einbeziehung des 1.und 2. Blocks . . . . .	56
3.6	Scatterplot der Abhängigkeit der Korrektheit der Antworten von der Bekanntheit der Wörter . . . . .	58
3.7	Scatterplot der Abhängigkeit der Korrektheit der Antworten von der Bekanntheit der Bedeutungen . . . . .	59
3.8	Histogramm der Antworten des gesamten Sprachmaterials und der vollständig bekannten Wörter . . . . .	60
3.9	Boxplot der Abhängigkeit der Korrektheit der Antworten von der Wortstruktur . . . . .	63
3.10	Boxplot der Abhängigkeit der Verständlichkeit der Wörter von den in ihnen enthaltenen Vokalen . . . . .	67
3.11	Boxplot der Abhängigkeit der Verständlichkeit von der Länge des im Wort enthaltenen Vokals . . . . .	69
3.12	Boxplot der Abhängigkeit der Verständlichkeit von der Wortherkunft . . . . .	71
3.13	Vergleich der Versuchsergebnisse der Pilotstudie mit der Evaluation . . . . .	73
3.14	Histogramm der richtigen Antworten aus der Schnittmenge der 360 Wörter . . . . .	75
3.15	Linienplot der verfügbaren Wörter je nach den korrekten Antworten . . . . .	77
3.16	Histogramm der absoluten Häufigkeiten der Bekanntheit der Wörter . . . . .	78
3.17	Histogramm der absoluten Häufigkeiten der Bekanntheit der Bedeutungen . . . . .	79

3.18	Histogramm der richtigen Antworten aller Wörter, der bekannten Wörter und der Wörter, die in den Freiburger Testverfahren zur Anwendung kommen . . . . .	80
3.19	Linienplot der Anzahl der verfügbaren Wörter je nach deren Bekanntheitsgrad . . . . .	81
3.20	Scatterplots der vier Optimierungsmethoden . . . . .	82

## Tabellenverzeichnis

2.1	Permutation der Testlisten . . . . .	30
3.1	Leistung und Wortkenntnis der Probanden . . . . .	51
3.2	Tabelle über den Zusammenhang der Bekanntheit der abgehörten Wörter und der Korrektheit der Antworten (p-Werte des $\chi^2$ -Tests) . . .	61
3.3	Übersichtstabelle der besten zwei Listenzuordnungen . . . . .	83
4.1	Korrektheitsgrade der Wörter gleicher Aussprache . . . . .	88
C.1	Übersichtstabelle der entstandenen Listen Teil 1 . . . . .	116
C.2	Übersichtstabelle der entstandenen Listen Teil 2 . . . . .	117
C.3	Übersichtstabelle der entstandenen Listen Teil 3 . . . . .	118
C.4	Übersichtstabelle der entstandenen Listen Teil 4 . . . . .	119
C.5	Übersichtstabelle der entstandenen Listen Teil 5 . . . . .	120

# 1. Einleitung

## 1.1. Die geschichtliche Entwicklung, Nutzen und Anwendung der Sprachaudiometrie

„Der Mensch zeichnet sich unter allen anderen Lebewesen der Gattung Mammalia dadurch aus, dass bei ihm die Möglichkeit zur Sprachkommunikation mit einem oder mehreren Partnern evolutionär entwickelt ist“ (Keidel, 1980, s. Seite 172). Das dazu benötigte auditorische System, welches zusammen mit dem visuellen System zu den Fernsinnen des Menschen gehört, ist für ihn von elementarer Bedeutung. Personen, die eine Minderung der Hörleistung erfahren mussten, können in ihrem Alltag deutlichen Einschränkungen unterworfen sein. Ohne eine normale Hörfähigkeit ist das Verstehen von Sprache, und damit die zwischenmenschliche Kommunikation, für die Betroffenen entweder deutlich erschwert oder sogar völlig unmöglich. Eine Vielzahl genetisch, traumatisch oder tumorös bedingter Minderungen der Hörleistung sind bekannt. Durch audiometrische Reihenuntersuchungen an Millionen schwerhöriger Menschen im Ausland wurde errechnet, dass mindestens 1 %, wahrscheinlich aber 2-3 % der Bevölkerung, Hörstörungen derartigen Umfangs aufweisen, dass ernste Schwierigkeiten bei der Verständigung auftreten (Hahlbrock, 1970, s. Seite 147).

Zur genauen quantitativen und qualitativen Untersuchung der Hörstörungen werden präzise diagnostische Verfahren benötigt. Deren Ergebnisse sollten dabei replizierbar sein, damit es möglich ist, Vergleiche zwischen verschiedenen Versuchsergebnissen desselben Patienten und zwischen Versuchsergebnissen verschiedener Patienten anzustellen. Nur durch diagnostische Verfahren, die genaue Angaben über das Ausmaß der Hörminderung geben und eventuell auch Rückschlüsse auf den Entstehungsort zulassen, kann die Wahl adäquater und an die Bedürfnisse des Patienten angepasster Therapien ermöglicht werden. Dabei kann man tonschwellenaudiometrische- und sprachaudiometrische Tests unterscheiden.

Die Entwicklung der Testverfahren zu der Bestimmung von Hörverlusten begann am Anfang des 19. Jahrhunderts, denn schon 1801 hatte Grappengiesser versucht,

## 1. Einleitung

Schwerhörigkeit oder Taubheit zu heilen, indem er die Ohren tauber Kinder mit Gleichstrom reizte. Er konnte sogar von gewissen Erfolgen berichten (Feldmann, 2004, s. Seite 735 f.). Aufgrund dieser Berichte erschien es notwendig, das Hörvermögen zu messen, um das Ausmaß der Erkrankung erfassen und etwaige Therapieerfolge belegen zu können.

Den Grundstein für die Tonschwellenaudiometrie hatte Berliner 1878 gelegt, indem er aus dem Phonographen das Grammophon entwickelte (Hahlbrock, 1970, s. Seite 1). Allerdings ist die Tonschwellenaudiometrie allein für eine umfangreiche Diagnostik noch nicht ausreichend, da es nur in Einzelfällen möglich ist, von dem Tonaudiogramm Rückschlüsse auf das Sprachverständnis eines Patienten zu ziehen, was darauf beruht, dass das Audiogramm eine Schwellenwertmessung für einzelne Töne ist, während es sich bei der Sprache um einen akustischen Komplex handelt (Koch und Weiland, 1950, s. Seite 424). Sprache eignet sich außerdem besser für die Hörgeräteanpassung und für die Prognose und Beurteilung des Erfolges bei Ohroperationen (Roeser, 1963, s. Seite 851). Smoorenburg und Bosman drückten 1992 die Notwendigkeit für die Durchführung der sprachaudiometrischen Diagnostik folgendermaßen aus: „...tone audiometry does not provide us with the direct insight into the handicap that is experienced by the individual with hearing loss in perceiving speech in everyday life“ (Smoorenburg und Bosman, 1992, s. Seite 86).

Die Methode Grappengiessers wurde drei Jahre später von Pflingsten übernommen, aber mit dem gravierenden Unterschied, dass Pflingsten zum ersten Mal Sprache einsetzte, um verschiedene Grade des Hörverlustes unterscheiden zu können (Bosman, 1992, s. Seite 15). Durch Schmalz wurde es ab 1846 üblich, die Hörweite für Sprache als Maß einer Hörminderung zu verwenden, das heißt, dass die Abstufung der Lautheit der Sprache durch die Entfernung definiert wurde (Feldmann, 2004, s. Seite 737 f.). Dabei war ein Ablesen der Wörter in Flüstersprache noch nicht üblich. Auch gab es noch keine Empfehlungen, welche Wörter zu verwenden waren.

Somit waren die Sprachtests noch großen Schwankungen unterworfen, je nachdem, welcher Prüfer den Hörtest mit welchem Vokabular durchführte. Diese Problematik erkannte auch Kramer. Er stellte 1861 fest: „...dass es niemandem möglich ist, seine Stimme in stets gleicher Stärke und Modulation zu erhalten. [...] Dieselbe (Tonsprache) [...] könne nicht als Maßstab für den Grad der Schwerhörigkeit benutzt werden“ (Feldmann, 2004, s. Seite 737). Lucae forderte aus diesem Grund, die Hörprüfung nur noch in Flüstersprache durchzuführen, da man erkannt hatte, dass die Lautheitsunterschiede zwischen Vokalen und Konsonanten wesentlich geringer ausfielen, wenn man die Hörweitenprüfung in Flüstersprache, also tonlos vornahm.

## 1.1. Entwicklung, Nutzen und Anwendung der Sprachaudiometrie

Helmholtz hatte 1863 mit dem Werk „Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik“ neue Erkenntnisse publiziert, die die Grundlagen der Gehörprüfung mit Sprache entscheidend veränderten. Er hatte nachgewiesen, dass Vokale aus einfachen Tönen zusammengesetzt sind, und es war ihm gelungen, Sprachlaute mit einem Sortiment von elektrisch betriebenen Stimmgabeln zu synthetisieren. Außerdem hatte er wichtige physiologische Grundlagen erkannt, indem er davon ausging, dass die einzelnen Töne in der Cochlea durch Resonanz auf der Basilarmembran getrennt und aufgenommen werden (von Helmholtz, 1863, s. Seite 197 ff.). Zwar hat sich die Annahme des Resonanzmechanismus nicht bestätigt, wohl aber die Erkenntnis, dass jeder Frequenz ein Abschnitt auf der Basilarmembran zugeordnet ist. Somit hatte sich in der Praxis eine Hörweitenprüfung in Flüster- und Umgangssprache mit zweisilbigen Zahlwörtern etabliert.

In den 40er Jahren waren die USA die Vorreiter in der Sprachaudiometrie (Hahlbrock, 1970, s. Seite 69). Dort war zwischen 1947 und 1948 in psychoakustischen Laboratorien der Harvard Universität unter anderem der sogenannte Spondee-Sprachtest von Hudgins, Hawkins, Karlin und Stevens entwickelt worden, der aus Wörtern mit zwei gleich stark betonten Silben, wie zum Beispiel *baseball*, besteht.

Auch in Europa gab es daraufhin bald erste repräsentative Sprachgehörprüfungen. Amersbach und Meister stellten 1950 in Düsseldorf einen „isophonen Wortkatalog“ zusammen, der durch eine spektrographische Analyse der Wörter begründet war (Amersbach und Meister, 1950, s. Seite 355 ff.). Schubert kombinierte für seinen Sprachtest in Bonn jeweils drei Wörter mit zehn bis dreizehn Lauten, einerseits in einem sinnvollen, und andererseits in einem sinnlosen Zusammenhang, wie zum Beispiel: *derselbe fort- fort selbe der*. Diese drei Wörter mussten gemeinsam nachgesprochen werden. Damit sollte Zeit eingespart werden (Schubert, 1950a, s. Seite 99 ff. und Schubert, 1950b, s. Seite 328 ff.). Aber weder der Sprachtest von Amersbach und Meister, noch der von Schubert hielten der klinischen Erprobung stand, was nach Weiland durch die mangelhafte Beachtung der psychologischen Gesichtspunkte begründet werden muss (Weiland, 1954, s. Seite 7). Siehe dazu Abschnitt 1.2.2.

Der anerkannteste Einsilber-Sprachtest wurde 1953 von Karl Heinz Hahlbrock veröffentlicht. Er hatte erkannt, dass unter den seinerzeit üblichen Prüfbedingungen zu viele Fehlerquellen die Resultate der Gehörprüfungen nahezu wertlos machten. Weiland hatte 1954 die unzulänglichen Prüfbedingungen wie folgt zusammengefasst:

## 1. Einleitung

Zum einen sei die Flüstersprache und das Sprechen in Umgangslautstärke nicht einheitlich, denn: „... bei der Prüfung mit Flüster- und Umgangssprache bleibt es dem jeweiligen Untersucher völlig überlassen, welche Flüster- und Umgangslautstärke er wählt, wie deutlich er artikuliert und vor allem, mit welchen Prüfwörtern er arbeitet“ (Weiland, 1954, s. Seite 2). Allgemein anerkannte Wörterlisten gab es zu diesem Zeitpunkt noch nicht. Außerdem sei es bei weitem zu ungenau, das Hörvermögen der Patienten in Meterabständen zu messen. Und schlussendlich würde der unterschiedlichen Raumakustik verschiedener Prüfungsräume als möglicher Fehlerquelle keine Beachtung geschenkt, was sich für eine Vereinheitlichung der Prüfsituation unbedingt ändern müsse (Weiland, 1954, s. Seite2).

Nachdem Hahlbrock nach dem amerikanischen Vorbild des Spondee-Tests zunächst den sogenannten „Freiburger Hörverlusttest“ mit Wortlisten aus Zweisilbern und Zahlwörtern zusammengestellt hatte, widmete er sich schließlich einsilbigen Hauptwörtern um einen geeigneten „Sprachverständnistest“ zu entwickeln (Feldmann, 2004, s. Seite 739). Der Unterschied zwischen den Testverfahren liegt darin, dass mit der Messung des Hörverlustes für Sprache diejenige Lautstärke bestimmt wird, bei der der Patient den Speech Reception Threshold (SRT), also ein Wortverständnis von 50 % erreicht. Die Differenz zu der bei Normalhörenden erforderlichen Intensität ergibt den Hörverlust für Sprache in der benutzten Lautstärkeeinheit. Für diesen Test wählte Hahlbrock zweisilbige, gut verständliche und bekannte Hauptwörter, die aus einsilbigen Begriffen zusammengesetzt sind, und Zahlenlisten. In der täglichen Sprachaudiometrie haben sich die Zahlenlisten durchgesetzt.

Ein Sprachverständnistest misst im Gegensatz dazu die Güte des Unterscheidungsvermögens und zeichnet sich durch die Schwierigkeit aus, die er dem einhundertprozentig richtigen Verstehen aller Wörter entgegensetzt (Weiland, 1954, s. Seite 15).

Am Anfang stand die Überlegung, welches Testmaterial für die Entwicklung eines Sprachverständnistests geeignet wäre. Es mussten Test-Items gewählt werden, die einerseits möglichst schwierig sind und andererseits die Redundanz auf ein Minimum reduzieren. Falconnet stellte 1950 für die französische Sprache eine Verständlichkeitsskala auf, nach der ganze Sätze am einfachsten zu verstehen sind. In der Reihenfolge ansteigender Schwierigkeit folgten erst Zahlen, dann Wochentage, Monate, Vornamen, gebräuchliche Wörter, ungebräuchliche Wörter und schließlich Logatome. Logatome sind sinnleere, einsilbige aus Sprachlauten zusammengesetzte Kunstwörter. Auch wurde festgestellt, dass die Verständlichkeit von Wörtern deutlich

## 1.1. Entwicklung, Nutzen und Anwendung der Sprachaudiometrie

mit ihrer Lautzahl zunimmt (Egan, 1948, s. Seite 955 ff.). Dabei ist besonders die Zunahme eines Wortes um einen weiteren Vokal, das dann natürlich zu einem zweisilbigen Wort wird, mit einem erheblichen Gewinn der Verständlichkeit verbunden. Nach diesen Kriterien stehen zwei Möglichkeiten für die Auswahl des Testmaterials zur Verfügung. Es könnten Logatome oder Einsilber zur Anwendung kommen. Logatome bieten zwar den Vorteil nicht vorhandener Redundanz, was ein Auswendiglernen solcher Listen unmöglich macht, sie weisen aber so viele Nachteile auf, dass sich Hahlbrock und Weiland gegen eine Prüfung mit Logatomen entschieden haben. Zum Beispiel sei die Hörprüfung mit Logatomen „unnötig schwer und zu künstlich“ (Caussé und Falconnet, 1947, s. Seite 436). Außerdem würden alle Menschen dazu neigen, dem Gehörten einen Sinn geben zu wollen: „Niemand hört, als was er weiß“ (von Goethe, 1833, s. Seite 156 ff.). Schlussendlich sei die Prüfung mit Logatomen für den Patienten sehr viel unangenehmer als mit Einsilbern, da er dabei sehr schnell ermüde und schließlich sogar weghöre (Hahlbrock, 1970, s. Seiten 82 f.).

Aus diesem Grund wählte Hahlbrock für die Entwicklung des Sprachverständnistests einsilbige Hauptwörter, die in jeder Sprache als die am schwersten verständlichen Wörter gelten (Weiland, 1954, s. Seite 15). Die relative Häufigkeit der Wörter entnahm er einer vorangegangenen Arbeit von Kaeding (1897), die auf der Grundlage von 11 Millionen Wörtern oder 20 Millionen Silben berechnet worden war. Unter den 13 215 häufigsten Wörtern waren nur 1194 einsilbige Hauptwörter vertreten. Hahlbrock wählte aus dieser Gruppe 800 Wörter aus, die er in 40 Gruppen mit jeweils 20 Wörtern verteilte, mit dem Ziel, die Gruppen phonetisch auszugleichen. Allerdings stellte sich heraus, dass in einigen dieser Gruppen trotzdem viele sehr seltene und unbekannte Wörter enthalten waren. Aus diesem Grund wurde die Anzahl der Gruppen auf 20 reduziert. Die Länge der Gruppen blieb dabei gleich (Feldmann, 2004, s. Seite 739).

Für die Aufsprache und Reproduktion der Einsilber diente ein im Eigenbau konstruiertes Tonbandgerät. Aufgesprochen wurden die Wörter von einem Rundfunksprecher, der in zahlreichen Vorversuchen die am besten reproduzierbaren Ergebnisse erzielt hatte.

Schon 1952 wurde dieser sogenannte Freiburger Sprachtest von der Arbeitsgemeinschaft deutscher Audiologen (ADA) für die klinische Sprachaudiometrie als Standardverfahren empfohlen. Trotz einiger Kritikpunkte am Freiburger Sprachtest wird er bis heute im klinischen Alltag standardmäßig verwendet (Dieroff und Meissner, 1994, s. Seite 60).

## 1. Einleitung

Inzwischen gibt es eine Vielzahl von Variationen und Überarbeitungen des Freiburger Sprachtests: So fand 1968 eine Neuaufnahme des Einsilber- und des Zahlentests in den Studios des Norddeutschen Rundfunks Hamburg statt (Plath, 1971, s. Seite 283), bei der auf eine flüssigere Aussprache geachtet werden sollte (Doerr, 1994, s. Seite 493). Und 1976 wurde der Test von Keller und Moser erneut überarbeitet und aufgenommen (Baumann, 2006, s. Seite 4). Auch wurde die phonetische Ausgewogenheit von Keller 1977 überarbeitet. Außerdem hat man einen Stör-schall entwickelt, der durch 32fach zeitverdeckte Überlagerung der Einsilber die gleiche spektrale Verteilung wie das Sprachmaterial aufweist (Döring und Hamacher, 1992, s. Seite 149). Ferner wurde von Schubert der sogenannte „Verhallte Freiburger Sprachtest“ eingeführt, der unter anderem bei der apparativen Rehabilitation von Hörgeschädigten und bei der Erkennung von Akustikustumoren von Bedeutung ist (Dieroff und Meissner, 1985, s. Seite 466). Vor allem eine schnelle und zuverlässige Früherkennung dieser Erkrankung ist mit dem verhallten Freiburger Sprachtest möglich (Dieroff und Mangoldt, 1989, s. Seite 376).

Die vorgelegte Arbeit beruht auf der von Mahfoud 2009 publizierte Dissertation zur Entwicklung eines neuen Sprachtests, der in seiner Form dem Freiburger Sprachverständnis test ähneln, aber die bestehenden Mängel beheben sollte. Siehe hierzu Abschnitt 1.2.3. Mahfoud hatte die einsilbigen Hauptwörter der deutschen Sprache aus dem Duden Version 3.0 extrahiert und einen Großteil der Wörter sowohl mit einer weiblichen Sprecherin als auch mit einem männlichen Sprecher aufgenommen. Die Schnittmenge der Wörter, die auch in den Aufnahmen des Freiburger Sprachtests von 1968 und 1976 enthalten waren, hatte sie in einer Pilotstudie mit vier Probanden evaluiert (Mahfoud, 2009, s. Seite 26 ff.). In der vorgelegten Dissertation soll die Dissertation Mahfouds fortgeführt und das Gerüst für einen Sprachtest zur Verfügung gestellt werden, der den modernen Anforderungen genügt.

Die Sprachaudiometrie ist inzwischen zu einem unverzichtbaren Bestandteil der audiologischen Diagnostik geworden, die zusammen mit der Tonaudiometrie am Anfang einer jeden Untersuchung stehen sollte. Dabei wird sie nicht nur als einschätzende Maßnahme der Schwerhörigkeit eines neuen Patienten, sondern auch für die Verlaufskontrolle, für das Feststellen von Hörentwicklungen nach chirurgischen Eingriffen und für die Anpassung von Hörgeräten verwendet. In der Forschung ist die Sprachaudiometrie zum Nachweis technischer Fortschritte im Bereich von Hörgeräten, Cochlea- und Hirnstammimplantaten notwendig.



## 1.2. Der Freiburger Einsilberversständnistest

### 1.2.1. Voraussetzungen, Durchführung und Merkmale

Der Freiburger Sprachverständlichkeitstest ist bei Kindern, wie auch bei Erwachsenen, seit Jahrzehnten der meistgebrauchteste Sprachtest im gesamten deutschsprachigen Raum (Kompis et al., 2006, s. Seite 445). Die Durchführung des Freiburger Sprachverständnistests ist nach DIN 45 621-1 standardisiert (Niemeyer, 1967, s. Seite 335 und Kiessling, 2006, s. Seite 12 ff.), was u. a. ein Grund für die starke Akzeptanz dieses Tests im klinischen Alltag ist (Sukowski et al., 2009, s. Seite 240).

In der Sprachaudiometrie befindet sich der Patient zusammen mit dem Hörprüfer in einer Audiometrikammer, die nach ISO 8253-2 genormt ist (Mahfoud, 2009, s. Seite 4). Diese schallgedämpften Räume dienen zum einen dem Ausschluss ablenkendem, hauptsächlich niederfrequentem Störschalls, der vor allem im unteren Frequenzbereich verdeckend wirkt, und zum anderen einer Vereinheitlichung der Prüfatmosphäre, die schon 1954 von Weiland gefordert worden war. Siehe dazu Abschnitt 1.1.

Nach den bestehenden Richtlinien muss die Größe des Raumes mindestens 2×3 m betragen. Die Wand-, Boden- und Deckenverkleidung muss reflektionsarm sein. Der Störschall darf bei der Routine-Audiometrie maximal 40 dB betragen. Erreicht wird dies durch Doppelwände und Doppeltüren. Außerdem darf der Patient von seinem Sitzplatz keine Einsicht in das Audiogramm haben und die Lautsprecher für die Freifeldmessung müssen einen Meter von seinen Ohren entfernt sein. Eine gute Klangwirkung wird durch eine Schallabsorption der Ausstattung des Prüfraumes erreicht. Die Schallabsorption lässt sich durch Ausstattung des Hörprüfraumes mit einer porösen Wandverkleidung, welche eine große Absorption für tiefe Frequenzen aufweist, verwirklichen (Hahlbrock, 1970, s. Seite 45).

Die Sprachgehörprüfung kann seitengetrent über Kopf- oder Einsteckkopfhörer oder im freien Schallfeld vorgenommen werden (Kompis, 2004, s. Seite 96). Bei der Freifeldmessung werden dem Patienten aus dem Lautsprecher vor ihm zunächst die Zahlwörter und anschließend die einsilbigen Hauptwörter jeweils einer Liste vorgespielt. Hat der Patient eine stark unterschiedliche Schwerhörigkeit auf beiden Ohren, so muss zur differenzierten Prüfung das Ohr mit der besseren Innenohrleistung veräubt werden (Feldmann, 1979, s. Seite 11.51).

Der Freiburger Sprachverständlichkeitstest beinhaltet zwei Testabschnitte und besteht somit aus 10 Gruppen mit jeweils 10 zweistelligen Zahlen und 20 Gruppen mit jeweils 20 einsilbigen Wörtern (Lehnhardt und Laszig, 2000, s. Seite 175).

## 1. Einleitung

Der Zahlentest dient der groben Orientierung über den vorliegenden Hörverlust und zur Prüfung sehr schwerhöriger Patienten, also der Bestimmung des „Hörverlusts für Sprache“. Um ihn zu vereinheitlichen, kommen die Zahlen eins bis zwölf in dem Test nicht vor, da sie einsilbig sind. Die einzelnen Gruppen bestehen vielmehr aus jeweils zwei zweisilbigen und acht viersilbigen Zahlwörtern und sollen phonetisch annähernd ausgeglichen sein. Unter den insgesamt 100 Zahlwörtern sind 14 doppelt vertreten, aber so in den Gruppen verteilt, dass ein Wiedererkennen dieser Wörter bei mehrfacher Prüfung möglichst ausgeschlossen sein soll (Hahlbrock, 1970, s. Seite 86). Wie unter Punkt 1.1 bereits beschrieben, wird bei diesem Abschnitt des Freiburger Sprachtests der SRT ermittelt. Er entspricht derjenigen Intensität, bei dem ein Wortverständnis von 50 % erreicht wird. Dieser außerordentlich wichtige Wert entspricht der Lautstärke, bei dem ein Schwerhöriger dem Sinn fortlaufender Sprache gerade noch folgen kann (Weiland, 1954, s. Seite 6). Zahlen sind zur alleinigen Prüfung des Gehörs mit Sprache jedoch nicht geeignet, unter anderem deswegen, weil von den 53 bekannten deutschen Sprachlauten zur Bildung der Zahlwörter nur 24 benutzt werden (Hahlbrock, 1970, s. Seite 86). Außerdem weisen Zahlwörter eine sehr hohe Redundanz auf.

Die Güte des Unterscheidungsvermögens wird mit einsilbigen Hauptwörtern, genauer gesagt, mit einsilbigen Substantiven geprüft. Dabei wird auf sinnlose Wörter als Testmaterial verzichtet. Auch Ortsnamen, Vornamen, Monatsnamen, Farben, Himmelsrichtungen, *Ja* und *Nein* kommen in dem Freiburger Sprachverständnistest nicht vor, um zufällige „Begünstigungen“ zu vermeiden (Weiland, 1954, s. Seite 15). Die Testwörter werden den Patienten mit einer Schallleitungsschwerhörigkeit mit 20 dB über dem SRT, und den schallempfindungsschwerhörigen Patienten mit 40 dB über dem SRT, angeboten (Feldmann, 1979, s. Seite 11.50). Der Freiburger Sprachtest ist ein offenes Testverfahren. Das heißt, der Patient muss die Wörter, die er verstanden hat, wiederholen und kann nicht aus einer Gruppe ihm dargebotener Wörter eines auswählen, wie es etwa bei dem Einsilber-Reimtest nach Wallenberg und Kollmeier der Fall ist (Sukowski et al., 2009, s. Seite 244). Gemessen wird jedes nicht verstandene Wort, das bei einer Gesamtheit von 20 Wörtern, einem Sprachverlust von jeweils 5 % entspricht. Die Ergebnisse beider Testabschnitte werden in das Sprachaudiogrammformular, das nach der DIN-Vornorm 45624 standardisiert ist, eingetragen (Feldmann, 1979, s. Seite 11.51). Aus ihm können dann die Werte für den Hörverlust für Zahlen, für den Diskriminationsverlust und die optimale Intensität abgelesen werden.

Die Bedeutung dieses Prüfverfahrens lag in den neu entstandenen Möglichkeiten, verfeinerte Diagnosen zu stellen, die praktischen Erfolge gehörverbessernder Operationen zu prüfen, Hörgeräte viel präziser anpassen zu können und in der Möglichkeit, gutachterliche Beurteilungen von Gehörschäden stellen zu können (Plath et al., 1973, s. Seite 457). Bis Hahlbrock 1953 seine Sprachgehörprüfung vorgestellt hatte, gab es diese Vereinheitlichungen noch nicht. Viele Fehlerquellen, wie das Vortragen des Sprachmaterials in Flüstersprache oder in der Lautstärke normaler Umgangssprache, unterschiedliche Artikulation der Prüfer, die Durchführung der Gehörprüfungen in nicht einheitlichen Räumlichkeiten und das Fehlen eines einheitlichen Wortkataloges wurden dadurch ausgeschlossen.

### 1.2.2. Einflussgrößen des Freiburger Sprachverständnistests

Hahlbrock legte als Kriterien für die Aufstellung seines Sprachverständnistests fest, dass alle verwendeten Wörter sinnhafte Wörter und keine Logatome sein sollten. Außerdem sollten die verwendeten Wörter allgemein verständlich sein und eine hohe Gebräuchlichkeit aufweisen. Das Testmaterial als Ganzes sollte eine hohe Homogenität haben und keine zu starke Vereinfachung enthalten. Dabei durften keine zwei aufeinanderfolgenden Wörter innerhalb einer Gruppe eine Sinnverwandtschaft aufweisen. Schlussendlich wurde festgelegt, dass die Gruppen relativ kurz sein sollten, damit die Durchführung des Tests nicht zu viel Zeit in Anspruch nähme (Hahlbrock, 1960a, s. Seite 102).

Außer dem Aufbau des Tests, der Zusammenstellung der Wörter und der technischen Durchführung, gibt es eine Reihe von Einflussgrößen, die nicht mit der Auswahl des Tests aufgehoben werden können. Diese Einflussgrößen kann man als psychologische Faktoren zusammenfassen. Die persönlichen Faktoren sind nie völlig auszuschalten. Gemeint sind unter anderem die unterschiedliche Bildung und die damit verbundene Größe des persönlichen Wortschatzes, das soziale Niveau, die Intelligenz und phonetisch-physiologische Gesichtspunkte. Dadurch kann der Vergleich der Ergebnisse erheblich erschwert werden (Weiland, 1954, s. Seite 8). Ebenso kann die Kenntnis der verwendeten Sprache die Ergebnisse stark beeinflussen. Außerdem spielen auch die individuellen Schwankungen von Interesse, Aufmerksamkeit, Gewissenhaftigkeit, Konzentration, augenblicklicher Stimmungslage, Temperament und Ermüdbarkeit eine Rolle (Hahlbrock, 1970, s. Seite 81). Teilweise können diese unterschiedlichen Einflussgrößen ausgeglichen werden, indem zu schwierige, veral-

## 1. Einleitung

tete und unbekannte Wörter aus dem Sprachmaterial eliminiert werden. Auch sollte die Abfolge der Wörter nicht zu schnell hintereinander erfolgen, um die Reaktionsfähigkeit des Patienten nicht zu überfordern. Letztendlich wirken sich die persönlichen Einflussfaktoren aber nicht nur im Sprachtest, sondern auch im täglichen Leben aus, weswegen sie die Vergleichbarkeit nicht allzu sehr einschränken (Weiland, 1954, s. Seite 7).

### 1.2.3. Kritikpunkte am Freiburger Sprachverständnistest

Im klinischen Alltag sind schon seit langem einige Schwachstellen des Freiburger Sprachtests aufgefallen, was weitergehende Untersuchungen notwendig erscheinen ließ. Dadurch wurde der Freiburger Sprachtest Gegenstand vieler wissenschaftlicher Untersuchungen. In den Publikationen hierüber sind die nachfolgenden Kritikpunkte aufgezeigt worden.

Obwohl Hahlbrock bei der Entwicklung des Sprachverständnistests und der Zusammenstellung der Listen genauestens auf eine phonetische Ausgeglichenheit und eine Ausgeglichenheit des Schwierigkeitsgrades geachtet hatte, ist im klinischen Alltag deutlich geworden, dass die Listen von den Patienten als unterschiedlich schwer empfunden wurden. Daraufhin wurde die Schwierigkeit der Listen und ihre phonetische Ausgeglichenheit sowohl mit normalhörenden als auch mit schwerhörigen Patienten unter anderem von Bangert (1980), Alich (1985) und von Wedel (1986) untersucht. Dabei hatte von Wedel festgestellt, „... dass bezweifelt werden müsse, ob von einem phonetisch ausbalancierten Test gesprochen werden könne. [...] Die Interpretation und Bewertung der Sprachaudiogramme könne bei Verwendung des Freiburger Sprachtests in seiner jetzigen Form (Erstaufsprache) Fehlbeurteilungen der Hörstörungen zur Folge haben. Dies wäre für die Begutachtung, für die Hörgeräteindikation, für die Hörgeräteanpassung und für die Kontrolle einer Hörgeräteversorgung sicherlich nicht zu verantworten“ (von Wedel, 1986, s. Seite 72). Von Wedels Untersuchungen zeigten, dass die Listen 1, 8, 14, 15, 19 und 20 deutlich von dem mittleren Diskriminationswert abweichen. In einer vorangegangenen Studie hatte er schon 1984 festgestellt, dass die Konsonant-Vokal-Struktur innerhalb der einzelnen Gruppen des Freiburger Sprachverständnistests nicht übereinstimmen (von Wedel, 1984, s. Seite 106). Auch Bangert hatte Untersuchungen mit dieser Zielsetzung durchgeführt und festgestellt, „... dass es Gruppen gibt, die bei dieser Stichprobe (der Probanden) besonders gut verstanden werden (Gruppe 1, 15 und 20),

aber auch solche (Gruppe 5, 9, 12, 13 und 14), die den Versuchspersonen große Schwierigkeiten bereiten. [...] In jedem Fall würde eine Fehleinschätzung bei der Verwendung einer solchen Gruppe die Hörgeräteanpassung stark beeinträchtigen“ (Bangert, 1980, s. Seite 168). Grund für die unterschiedliche Schwierigkeit der Listen sei nach Bangert ihre phonetische Unausgeglichenheit, die sich in der Häufung schwierig zu verstehender Cluster widerspiegele. So sei in den schwierigen Listen zum Beispiel eine Häufung von plosiven Initialphonemen mit nachfolgendem „f“ und ein vermehrtes Vorkommen von Affrikaten aufgetreten (Bangert, 1980, s. Seite 170). Affrikate sind die Folge eines Frikkativs auf ein Plosiv. Frikkative wiederum sind Konsonanten, die durch eine Verengung des Artikulationstraktes entstehen, wodurch der Luftstrom eingeengt wird. Beispiele sind das „f“ und das „s“. In vorangegangenen Untersuchungen hatte Bangert 1975 festgestellt, dass die Verständlichkeit einiger dieser Cluster nicht sehr gut ist (Bangert, 1980, s. Seite 170). Auch in anderen Untersuchungen war gezeigt worden, „... dass bei zweiformantigen Stimuli die Anwesenheit eines ersten Formanten eine deutliche Reduktion der Diskriminationsschärfe für die Entdeckung einer Transition im zweiten Formanten ausmacht. Dieser Maskierungseffekt [...] ist für Hörgestörte größer, als für Normalhörer“ (Sendlmeier, 1992, s. Seite 53). Auf ein ausgeglichenes Auftreten von Plosiven war bei Erstellung der Listen allerdings nicht geachtet worden. Auch Kiessling kritisierte, dass zwar die Phonemverteilung deutscher Sprache entspräche, die Phonemübergänge bei der Konzeption der Testlisten aber nicht berücksichtigt worden wären. Eine andere Erklärung für die ungleiche Schwierigkeit der Listen sehen Kiessling und Alich in dem unterschiedlichen Bekanntheitsgrad einiger Einsilber. „... Da die Verständlichkeit auch mit dem Bekanntheitsgrad korreliere, wiesen die einzelnen Testlisten (Gruppen) entgegen der ursprünglichen Absicht unterschiedliche Verständlichkeiten auf“ (Kiessling, 2000, s. Seite 633). Aber der Bekanntheitsgrad der Wörter würde nicht mit der Häufigkeit, mit der sie in gesprochener oder geschriebener Sprache vorkommen, korrelieren. Vielmehr wiesen Wörter, die in der Kinder- und Schulzeit erworben werden und eine hohe emotionale Bedeutung hätten, einen hohen Bekanntheitsgrad auf (Alich, 1985, s. Seite 2 f.). Außerdem gibt Alich zu bedenken, dass die Bekanntheit einiger Wörter sehr großen regionalen Schwankungen unterliegt. Aus der unterschiedlichen Schwierigkeit der Testlisten resultiert, dass eine Vergleichbarkeit der Untersuchungsergebnisse mit den verschiedenen Listen nicht mehr möglich ist. „... Grundsätzlich gelte, dass Sprachaudiogramme hinsichtlich der Feststellung des Diskriminationsverlustes nur dann miteinander vergleichbar seien, wenn mit der gleichen Gruppe

## 1. Einleitung

gemessen werde“ (Alich, 1985, s. Seite 2). Darüber hinaus wurde die universelle Verwendung des Tests für alle Arten der Schwerhörigkeit kritisiert. So sei „... bei Patienten mit einer Hochtonschwerhörigkeit die sprachaudiometrische Beurteilung nach dem Freiburger Zahlen- und Einsilbertest [...] unzulänglich, da aus dem Diskriminationsvermögen für einsilbige Wörter das tatsächliche Ausmaß der sprachlichen Kommunikationsbehinderung des Schwerhörigen im Alltag nicht zu bestimmen sei“ (Schultz-Coulon, 1974, s. Seite 735).

Anlass zu weiterer Kritik gab die übertriebene Artikulation mit übermäßiger Melodiebewegung der Wörter (Plath et al., 1973, s. Seite 458) und die Befürchtung, dass einige Testwörter wie zum Beispiel *Aas*, *Sau*, *Krebs*, etc. psychologische Hemmnisse hervorrufen könnten, und daher vorzugsweise aus dem Test eliminiert werden sollten. Außerdem seien „... die Testlisten mit nur jeweils 20 Wörtern zu klein, da dadurch die Streuung der Messergebnisse zu groß werde“ (Kiessling, 2000, s. Seite 633). Auch Alich kritisiert die Kürze der Testlisten. Denn bei häufiger Anwendung würde ein Lerneffekt einsetzen, der zu besseren Ergebnissen führe, als eigentlich zu erwarten wäre.

Andere Kritikpunkte beschäftigen sich nicht mit dem Aufbau und der Struktur des Testmaterials, sondern mit der Technik, der Durchführung und der Qualität der Aussprache. Kiessling kritisiert zum Beispiel, dass es kein Ankündigungssignal vor den Testwörtern und keinen geeigneten Störschall für die Prüfung in geräuschvollen Situationen gäbe (J. Kiessling und Wagner, 1994, s. Seite 8). Außerdem wären Sätze nach seiner Überzeugung für die Bestimmung der Sprachverständlichkeit im Störschall geeigneter (Kiessling, 2000, s. Seite 633 f.). Eine weitere Fehlerquelle sehen sowohl Alich als auch Kiessling in dem offenen Antwortsystem des Freiburger Sprachtests. Bei einigen Probanden sei es sehr schwierig, ihre Antwort korrekt zu verstehen. Gerade wenn die Patienten einen sehr starken regionalen Dialekt hätten oder wenn es sich um schwerhörige Patienten handle, die häufig eine verwaschene Aussprache hätten, sei es für den Prüfer oftmals schwierig zu verstehen, ob die Antwort zum Beispiel *Pflug*, *Flug*, oder *Fluch* lautete (Alich, 1985, s. Seite 4). Nach Kiessling hat das offene Antwortsystem den Nachteil, dass dadurch ein automatisierbarer, also computergesteuerter Testablauf nicht möglich ist.

Die Kritik anderer Autoren zielt in eine andere Richtung und betrifft technische Mängel. Obwohl die Tonträger ebenso wie die Wiedergabegeräte Normvorschriften unterliegen, hätten entsprechende Untersuchungen gezeigt, dass die verwendeten Tonträger häufig keine vergleichbaren Qualitäten aufweisen. Eine Vielzahl der Tonträger sei nicht zeitgemäß. So variere die Qualität vom einfachen Kassettenrecor-

der bis zum Tonbandgerät mit HIFI-Norm (von Wedel, 1986, s. Seite 72). Daraus resultieren natürlich qualitätsbedingte Unterschiede, die sich zum Beispiel auf Abnutzungserscheinungen der Tonträger oder Verschmutzungen der Tonköpfe zurückführen lassen. Dies habe weitreichende Konsequenzen. So hätten seine Untersuchungen mit extrem unterschiedlichen Wiedergabegeräten (Sprachaudiometer mit einfachem Kassettenrecorder, Sprachaudiometer mit Tonbandgerät mit HIFI-Norm) und entsprechend unterschiedlichen Tonträgern ohne Berücksichtigung der Gruppenauswahl im Freiburger Sprachtest MdE-Unterschiede bis zu 15% erbracht (von Wedel, 1986, s. Seite 72). Auch Alich (1985, s. Seite 1) kritisiert die Verwendung unterschiedlicher und häufig veralteter Audiometrietypen und gibt außerdem noch die unregelmäßigen und unterschiedlich häufig durchgeführten Wartungen der Audiometriegeräte als mögliche Fehlerquelle an. Kritisiert wurde auch die Qualität der Aufnahme. Nach Keller 1985 sind die Aufnahmen technisch und akustisch nicht einwandfrei. Dass weder Hahlbrocks Originalaufsprache noch die Neuaufsprache von 1968 heutigen technischen Anforderungen entspricht, liegt auf der Hand. Auch Kiessling (2000, s. Seite 633) kritisiert die technisch veraltete Aufsprache und die unnatürliche Pegelnivellierung.

Trotz dieser Vielzahl seit langem bekannter Mängel des Freiburger Sprachverständnistests und seiner Anwendung, wird er weiterhin standardmäßig im klinischen Alltag verwendet. Dies liegt unter anderem daran, dass der Freiburger Sprachverständnistest seinen Eingang in die ISO-Normierung gefunden hat, dass das Tabellenwerk zur Ermittlung der MdE und anderer Kenngrößen auf dem Freiburger Testmaterial beruht, und dass für die Anwendung anderer bestehender Testverfahren mit einem höheren Zeitaufwand gerechnet werden muss.

Ziel dieser Dissertation ist es, aufbauend auf der Dissertation von M. Mahfoud 2009, ein neues Testmaterial zur Verfügung zu stellen und zu evaluieren, das dem Freiburger Sprachverständnistest ähnelt, aber eine Vielzahl der hier genannten Mängel behebt. Dabei sollen neue Listen zusammengestellt werden, die nicht nur den selben Schwierigkeitsgrad aufweisen und phonetisch ausgeglichen sind, sondern auch auf eine Vielzahl von Einflussfaktoren Rücksicht nehmen, die bei der Zusammenstellung des Freiburger Einsilbertests nicht beachtet worden waren.



#### **1.2.4. Methodik zur Erstellung der Listen des Freiburger Sprachverständnistests und die Voraussetzungen für die Listenzusammenstellung des neuen Sprachmaterials M-2007**

Hahlbrock und Weiland wendeten für die Zusammenstellung ihrer Listen folgende Methode an: Wie in Abschnitt 1.1 bereits beschrieben, suchte Hahlbrock von den nach Kaeding (1897) 1194 häufigsten Einsilbern 800 einsilbige und allgemein verständliche Substantive aus, die er zunächst auf 40 Listen mit jeweils 20 Wörtern verteilte. Diese wurden später auf 20 Listen reduziert. Dabei waren in den einzelnen Gruppen jeweils 73 Sprachlaute vertreten, die sich aus 10 Wörtern mit 4 Sprachlauten, 7 Wörtern mit 3 Sprachlauten, 2 Wörtern mit 5 Sprachlauten und einem Wort mit zwei Sprachlauten zusammensetzten. Bei der Zusammenstellung der Listen sollte nicht nur auf eine phonetische Ausgeglichenheit zwischen den einzelnen Listen geachtet werden, sondern das Testmaterial sollte in seiner phonetischen Zusammensetzung fließender geschriebener oder gesprochener Sprache gleichen. Dieses Ziel ist bei einem Einsilber-Sprachtest nur annäherungsweise zu erreichen. Da es sich bei dem Testmaterial nur um eine einzige Wortgruppe handelt, kommen einige Sprachlaute wie das ablautende „e“, der sog. „Schwa“-Laut ə (Beispiel: *Sonne*) überhaupt nicht vor (Weiland, 1954, s. Seite 20), während andere Sprachlaute nur eine veränderte Häufigkeitsverteilung aufweisen. Beispiele für vermindert vorkommende Sprachlaute in Hahlbrocks Sprachverständnistest sind das *d*, das hauptsächlich in Artikeln zu finden ist, das *ch*, das dem Wort *Ich* seine Häufigkeit verdankt, und alle mittleren Vokalschattierungen, weil sie fast ausschließlich in mehrsilbigen Wörtern (Beispiel: *Aluminium*) zu finden sind. Insgesamt kommt es zu einer Verschiebung des Vokal- / Konsonantverhältnisses. Bei Hahlbrock und Weiland tritt eine deutliche Häufung von Konsonanten auf.

Nach der Einteilung der Wörter hinsichtlich ihrer phonetischen Zusammensetzung und ihrer Wortlänge, wurden sie anschließend nach der Stellung des Vokals innerhalb des Wortes sortiert. Die Listen wurden so zusammengesetzt, dass jede Liste ein Wort mit einem Vokal als Anlaut enthielt. Alle übrigen Wörter der Liste begannen mit einem Konsonanten. Außerdem wurde bei der Zusammenstellung der Listen strikt auf eine Ausgeglichenheit der Vokale geachtet. Es wurden folgende 6 Vokalsorten unterschieden: *a*, *e*, *i*, offenes *o*, geschlossenes *o*, *u* und die Diphthonge *ei* und *au*. Nicht als eigenständige Vokale galten *ö* und *ä*, die als *e* gewertet wurden, das *ü*, das



als *i* gewertet wurde, und der Diphthong *eu*, der als *ei* zu Buche schlug. Bei den Konsonanten zählten ausschließlich *r*, *l*, *ss*, *sch* und *h* als eigenständige Sprachlaute. Zu Gruppen zusammengefasst wurden jeweils *m*, *n*, *ng*; *f* und *ch* (pal); *k*, *p*, *t*, *ch* (gutt); *g*, *b* und *d*; *w* und das stimmhafte *s*. Der konsonantische Diphthong *ts* wurde als *ss* gewertet. Somit entstanden Listen, die jeweils 73 Sprachlaute enthalten und nach ihren Vokalhäufungen, Wortlängen und Anlauten sortiert sind (Hahlbrock, 1970, s. Seite 92 und Weiland, 1954, s. Seite 23).

Der Sprachverständnistest von Mahfoud (2009) bietet andere Voraussetzungen. Das Sprachmaterial beschränkt sich nicht auf die häufigsten Substantive, sondern erfasst alle Einsilber des digitalen Duden (2004) Version 3.0. Ausnahmen sind nach Mahfoud (2009) Einsilber,

- die nur in bestimmten Regionen Deutschlands bekannt sind, bzw. sich dort variieren,
- die in anderen Ländern gebräuchlich sind oder aus anderen Ländern importiert wurden und nur regional angewandt werden,
- die nur von bestimmten Gesellschaftsgruppen benutzt werden und dort eine Art Modesprache darstellen,
- die einer Fachsprache angehören,
- die durch Konjugation von Verben entstanden sind,
- die aufgrund des veränderten Sprachgebrauchs nicht mehr im Alltag benutzt werden,
- die Kurzformen oder Abkürzungen darstellen,
- die hauptsächlich in der Lyrik und Dichtung ihre Anwendung finden,
- die trotz unterschiedlicher Schreibweise und Bedeutung die gleiche Aussprache haben,
- die die gleiche Bedeutung, aber eine unterschiedliche Schreibweise haben.

Eine Aufgabe und ein Ziel dieser Dissertation ist es, das zur Verfügung stehende Sprachmaterial, dem Beispiel Hahlbrocks und Weilands folgend, auf möglichst viele Gruppen à 20 Wörter zu verteilen, wobei den oben aufgeführten Gesichtspunkten die gleiche Beachtung geschenkt werden muss. Zusätzlich soll auch auf die Einflussfaktoren Rücksicht genommen werden, die bei der Zusammenstellung des Freiburger Sprachtests keine Beachtung gefunden hatten, wie zum Beispiel der Bekanntheitsgrad der enthaltenen Wörter. Wenn möglich, soll weiterhin darauf geachtet werden, den Schwierigkeitsgrad des entstehenden Sprachverständnistests an den bestehenden Freiburger Sprachverständnistest anzugleichen.

### 1.3. Evaluation von Sprachtests: Durchführung und Merkmale

Aufgrund der Menge an Sprachtests, die in Deutschland seit dem zweiten Weltkrieg entwickelt wurden, gibt es eine Vielzahl von Studien, die sich mit dem Vergleich und der Auswertung verschiedener Sprachtests beschäftigen. So gibt es zum Beispiel von [Doerr \(1994\)](#) vergleichende sprachaudiometrische Untersuchungen mit Kompaktdiskette und Tonbandkassette. [Sukowski et al. \(2009\)](#) hatte sich mit der Vergleichbarkeit des Freiburger Sprachtests mit dem Göttinger-Satztest und dem Einsilber-Reimtest nach von Wallenberg und Kollmeier beschäftigt. [Kompis et al. \(2006\)](#) setzte sich mit der Vergleichbarkeit der Bezugskurven der Schweizer Version des Freiburger Zahlen- und Einsilbertests auseinander. Und natürlich haben auch Weiland und Hahlbrock den von ihnen entwickelten Sprachtest mit dem Ziel evaluiert, seine Brauchbarkeit zu erproben und zu beweisen. Zusätzlich zu diesen Untersuchungen, sind Evaluationen notwendig, um Normalkurven aufzustellen, die aussagen, wie viel Prozent eines bestimmten Testmaterials bei bestimmten Lautstärken verstanden wird.

Bei diesen vergleichenden Untersuchungen ist es von großer Wichtigkeit, standardisierte Meßmethoden zu verwenden. Der Versuch muss stets in denselben schalldichten oder schallreduzierten Räumlichkeiten mit der gleichen Versuchsanordnung durchgeführt werden. Außerdem sollten die Audiometer vor Versuchsbeginn geeicht sein. Auch die Anforderungen an das Probandenkollektiv müssen genau definiert sein. Für die Erstevaluation eines neuen Testmaterials bietet sich ein junges und normalhörendes Probandenkollektiv an. „Im Gegensatz zu normalen sprachaudiometrischen Untersuchungen sollen die Probanden nicht mit dem Testmaterial, sondern das Testmaterial durch die Probanden geprüft werden“ ([Hahlbrock, 1970](#), s. Seite 105). Ein wesentlicher Teil dieser Dissertation bestand darin, das gewonnene Sprachmaterial zu evaluieren, um die Schwierigkeit jedes einzelnen Einsilbers zu ermitteln.

## 1.4. Anforderungen an den neuen Einsilber-Sprachverständnistest

Aus den im Kapitel [Kritikpunkte am Freiburger Sprachverständnistest](#) genannten Mängeln ergeben sich eine Reihe von Anforderungen für die Konzeption des neuen Einsilber-Sprachverständnistests. Siehe hierzu Punkt [1.2.3](#).

Hudgins hatte zusätzlich folgende Anforderungen an einen Sprachverständnistest gestellt: Nach ihm ist es vor allem wichtig, dass es sich um bekannte Testwörter handelt, da das Gehör für Sprache getestet werden soll und nicht die Intelligenz oder das Vokabular des Patienten. Des Weiteren sollen die Test-Items phonetisch unähnlich sein, das heißt, innerhalb einer Liste sollen weder Wörter enthalten sein, die sich nur in einer Silbe unterscheiden, noch dürfen sich in einer Liste Wörter befinden, die sich reimen. Es würden den Patienten so nämlich nur Unterscheidungsfähigkeiten abverlangt, die so genau sind, dass sie die Effektivität des Tests nicht verbesserten. Darüber hinaus soll eine ausgeglichene Auswahl der Sprachlaute bestehen. Diese Anforderung ist jedoch nach Hudgins am ehesten zu vernachlässigen, da man das Verstehen von Sprache auch schon mit Zahlen testen könne, die nur einen Bruchteil aller vorkommenden Sprachlaute enthalten. Überdies ist auf eine Homogenität der Darbietung des Testmaterials zu achten ([Hudgins et al., 1947](#), s. Seite 58 ff.).

Ziel dieser Dissertation ist es, wie bereits unter Ziffer [1.1](#) dargestellt, aufbauend auf der Dissertation von [Mahfoud \(2009\)](#) einen Sprachverständnistest zu entwickeln und zu evaluieren, der einen Großteil der bekannten Mängel des Freiburger Sprachverständnistests behebt und die hier beschriebenen Anforderungen, so weit es sinnvoll ist, berücksichtigt. Große Beachtung sollte einer möglichst natürlichen und trotzdem deutlichen Aussprache der Einsilber geschenkt werden. Überdies wurde besonderes Augenmerk darauf gerichtet, den Schwierigkeitsgrad der einzelnen Listen möglichst konstant zu halten. Bei der Verteilung der Wörter soll, anders als bei dem Freiburger-Einsilbertest, besonderes Gewicht auf eine Ausgeglichenheit der Listen in Bezug auf die Anzahl der enthaltenen Lehnwörter, der verschiedenen Vokale und deren Länge und in Bezug auf ein gleichmäßiges Vorkommen der verschiedenen Wortstrukturen gelegt werden. Des Weiteren soll durch eine Erfassung des Bekanntheitsgrades darauf geachtet werden, unbekannte Wörter aus dem Testmaterial zu eliminieren, um die psychologischen Faktoren so weit wie möglich auszuschalten. Der neue Sprachtest soll den Anforderungen des klinischen Alltags gerecht werden und bei allen verwendeten Gruppen gleichbleibende und untereinander kompatible Ergebnisse liefern.

## 1. Einleitung

## 2. Versuchsplanung und Methodik

Diese Dissertation und die durchgeführten Versuche bauen, wie in Kapitel 1 bereits beschrieben, auf den Ergebnissen der von Mahfoud (2009) durchgeführten Arbeiten zur Erstellung eines neuen Einsilber-Sprachverständnistests auf. Die Kriterien zur Auswahl der Wörter wurden unter Abschnitt 1.2.4 beschrieben. Bevor mit der Evaluation des neuen Sprachtests begonnen werden konnte, mussten die Aufnahmen von ca. 400 Einsilbern wiederholt werden. Die Neuaufnahme der bis dahin fehlenden Wörter wurde mit einem männlichen Sprecher zur Vervollständigung des Sprachmaterials M-2007 durchgeführt. Dabei wurde dieselbe Versuchsanordnung verwendet, die Mahfoud für die Aufnahmen der Einsilber vorgeschlagen hatte. Siehe dazu Abschnitt 2.1.2 und Abschnitt 2.1.1. Ausgehend von den Überlegungen zu der Durchführung, den Merkmalen, den Voraussetzungen, sowie den Einflussgrößen des Freiburger Sprachverständnistests wurden Versuche konzipiert, mit dem Ziel, die durchschnittliche Schwierigkeit jedes einzelnen Einsilbers des neuen Sprachmaterials M-2007 zu evaluieren. Grundlage hierfür war wiederum der Versuchsaufbau von Mahfoud, der bei vorangegangenen vergleichenden Untersuchungen des neu aufgenommenen Sprachmaterials mit den Freiburger Aufnahmen von 1968 und 1976 verwendet worden war. Des Weiteren wurde ein Fragebogen erstellt, in dem die Probanden angeben mussten, welche der einsilbigen Wörter ihnen gänzlich unbekannt waren und welches Wort ihnen zwar bekannt, aber die Bedeutung unklar war. Dies war die Voraussetzung um evaluieren zu können, ob der Bekanntheitsgrad eines Wortes auch die Wahrscheinlichkeit beeinflusst, mit der es verstanden wird. Anschließend galt es, ausgehend von den Kritikpunkten am Freiburger Sprachverständnistest, eine neue Zusammenstellung der Wörter in Listen vorzunehmen, die die bekannten Schwachstellen seines Vorgängers vermeidet und die Anforderungen des klinischen Alltags berücksichtigt.

## 2.1. Aufsprache der fehlenden Einsilber

### 2.1.1. Räumlichkeit, Aufbau und verwendete Gerätschaften

Als Räumlichkeiten für die Aufsprache der verbleibenden 400 Einsilber wurde die Audiometrikabine B in der Audiometrieabteilung der Kopfklinik der Julius-Maximilians-Universität Würzburg ausgewählt. Diese Audiometrikammer besteht aus drei voneinander abgeschirmten Räumen, in deren mittlerer Abteilung sich die Eingangstür und die Steuerungskonsolen für die Gehörprüfungen befinden. In den äußeren beiden Abteilungen der Audiometrikammer sitzt im klinischen Alltag der Patient, während sich der Prüfer im mittleren Raum aufhält. In Vorversuchen durch Mahfoud 2009 war das Grundrauschen in den verschiedenen Audiometrikammern mit einem Schallpegelmesser 2260 INVESTIGATOR von Brüel & Kjær gemessen und verglichen worden. Dabei hatte sich ergeben, dass in der rechten Kammer der Audiometrikabine B in der Ecke diagonal gegenüber der Eingangstür das Grundrauschen mit 18 dB am geringsten war (Mahfoud, 2009, s. Seite 20). Dieser Standort bot sich für den Aufbau des Mikrofons an. Es wurde dort mit einem Abstand von 80 cm zum Raumwinkel aufgestellt. Um etwaige Übertragungen eines tieffrequenten Brummens von dem Boden auf das Mikrofon zu verhindern, wurden die Füße des Stativs mit Schaumstoff unterlegt. Die Höhe des Mikrofons wurde der Höhe des Mundes des Sprechers angeglichen. Für die Aufsprachen wurde ein Kondensationsmikrofon C-48 von Sony verwendet, das auf Kugelcharakteristik eingestellt worden war. Dabei ist die Aufzeichnung von Schallquellen in unterschiedlicher Raumrichtung mit annähernd gleicher Qualität möglich (Ederhof, 2004, s. Seite 88). Qualitätsminderungen durch geringfügige Positionsänderungen des Sprechers während der Aufnahme sollten dadurch möglichst ausgeschlossen werden. Um keine Dämpfungen zu erzeugen, wurde das Mikrofon auf 0 dB geschaltet und mit der nach unten gerichteten Membran an dem Stativ aufgehängt. Dies sollte ein Rauschen durch mögliche Wärmeströmungen verhindern. Außerdem wurde ein Ploppschutz angebracht, damit bei der Aussprache der Plosivlaute eine Übersteuerung verhindert werden konnte. Durch Vorversuche von Mahfoud war ein optimaler Mikrofon-Sprecherabstand von 12 cm ermittelt worden. Dieser Abstand wurde auch bei unseren Aufsprachen so genau wie möglich eingehalten. Mit einem Signalgenerator Typ 1049 von Brüel & Kjær wurden reine Sinustöne mit einer Frequenz von 1000 Hz für jeweils genau eine Sekunde generiert. Diese Sinustöne dienten später als Schnittmarkierungen. Neben dem Sprecher stand ein Passivlautsprecher, damit die generierten Sinustö-

ne über das Mikrofon mit aufgenommen werden konnten. Außerdem erhielt der Sprecher einen MP3-Player, der ein beispielhaftes Wort einer früheren Aufnahme enthielt. Dieses ermöglichte es ihm, seine Stimmlage an die vorherigen Aufsprachen anzupassen. Nachdem der Sprecher Platz genommen hatte, wurde auf Höhe seines Kopfes eine Abdämmung aus Schaumstoff und Pappe angebracht, um eine mögliche Echobildung zu verhindern. Die analogen Signale des Mikrophons wurden an den Mikrofonvorverstärker PHASE X24 von Terratec geleitet. Dort wurden sie digitalisiert und mittels FireWire an ein W1N Notebook von ASUS weitergeleitet. Dieses war mit Cubase LE(Steinberg) ausgestattet, mit denen die Audiosignale als WAV-Dateien abgespeichert werden konnten. Der Aufbau für die Aufnahmen der verbliebenen Einsilber glich in allen Punkten dem Aufbau, der von Mahfoud bei der Aufnahme der übrigen 1100 Einsilber verwendet worden war.

### **2.1.2. Durchführung der Aufsprache**

Für diese Dissertation wurden ausschließlich die männlichen Aufsprachen des von Mahfoud zur Verfügung gestellten Sprachmaterials vervollständigt. Zuvor waren alle Einsilber mit demselben Sprecher schon durch Mahfoud aufgenommen worden, konnten allerdings zum Teil, selbst bei wiederholt durchgeführten Aufsprachen, den anschließenden Qualitätskontrollen nicht standhalten. Von den 1554 einsilbigen Hauptwörtern mussten insgesamt 395 Wörter erneut aufgenommen werden. Dies hatte verschiedene Ursachen: Bei einigen Wörtern war bei dem Abhören eine nicht zufriedenstellende Artikulation aufgefallen oder sie wiesen trotz der oben beschriebenen Maßnahmen immer noch einen subjektiv empfundenen, zu starken Lautheitsunterschied zu den anderen Einsilbern auf. Auch ein zum Vorschein kommendes Hintergrundgeräusch oder ein bei der Artikulation aufgetretener Zungenschlag, der bei der Aufsprache nicht aufgefallen war, machte eine Neuaufsprache erforderlich. Das weibliche Sprachmaterial W-2007 wurde nicht vervollständigt.

Wie bei den vorangegangenen Aufnahmen durch Mahfoud, erfolgte die Aufsprache durch Herrn Dipl.-Ing. Stefan Brill, der auch der Betreuer dieser Dissertation ist. Es wurde darauf verzichtet, professionelle Sprecher zu engagieren, zum Teil aus finanziellen und zum Teil aus praktischen Gründen. So hätte der Sprecher mehrmals einbestellt werden müssen, wenn das Wort bei den anschließenden Qualitätskontrollen auch nach wiederholter Aufnahme immer noch mangelhaft erschien. Selbst dann wäre ein zufriedenstellendes Ergebnis nicht immer gewährleistet gewesen,

## 2. Versuchsplanung und Methodik

denn nicht nur die Güte des Sprechers, sondern auch die Aufnahmetechnik kann für die Mängel ursächlich sein. Mahfoud hatte sich in ihrer Dissertation gegen einen professionellen Sprecher entschieden, da für diesen Test eine möglichst natürliche und eher monotone Aussprache angestrebt werden sollte, während professionelle Sprecher dazu neigen würden, jedes Wort für sich interessant aussprechen zu wollen, was zu einer übertriebenen Artikulation führen könne (Mahfoud, 2009, s. Seite 21 ff.). Diesen Effekt hatte auch schon Hahlbrock 1951 bei der Entwicklung seines Sprachtests mit folgenden Worten beschrieben: „... diese Schauspieler (Bariton und Sopranistin) sprachen aber mit sehr viel Dramatik und Pathos, die Lautstärke und das Sprachtempo wechselten völlig regellos, so dass wir mit diesen Testbändern keine zufriedenstellenden Resultate bei Normalhörenden erzielten“ (Hahlbrock, 1970, s. Seite 98).

Als Orientierung während der Aussprache der verbliebenen Einsilber diente das Aussprachewörterbuch des Dudens (Duden, 2006a). Besondere Aufmerksamkeit wurde der Aussprache der An- und Auslaute gewidmet, deren genaue Wiedergabe besonders wichtig für das richtige Verständnis eines Wortes ist (Dekant, 1992, s. Seite 14). Auch andere Autoren betonen mit folgender Begründung die Wichtigkeit der Anlaute für das richtige Verständnis eines Wortes: „... der Beginn eines Wortes öffnet im Sinne einer Parallelverschiebung eine große Liste möglicher Bedeutungskandidaten, die alle mit einer ähnlichen akustischen Folge beginnen [...] bis am Ende das am besten passende Ergebnis übrigbleibt“ (Spreng, 1984, s. Seite 186). Als auffallend schwierig erwies sich die Aussprache der Plosivlaute und Affrikate, da die reine Aussprache häufig von Beigeräuschen des Artikulationsraumes überlagert wurde. Dabei ist eine genaue Aussprache der plosiven Initialphoneme mit den nachfolgenden Konsonanten (vor allem „l“) ebenfalls von großer Wichtigkeit, da die Verständlichkeit dieses Clusters ohnehin eingeschränkt ist (Bangert, 1980, s. Seite 170).

Bei aufeinander folgend gesprochenen Wörtern einer Wortliste besteht die Gefahr, dass der Sprecher das letzte Wort mit absinkender Intonation spricht. Aus diesem Grund sollten die Wörter so ausgesprochen werden, als wenn ihnen gedanklich ein Komma (anstatt eines Punktes) folgt. Außerdem musste darauf geachtet werden, die Wörter weder zu schnell, noch übertrieben langsam auszusprechen. Bei all diesen Bemühungen einer möglichst reinen und korrekten Aussprache kam es darauf an, trotzdem nicht die Natürlichkeit der Aussprache zu vernachlässigen. Der in Abschnitt 2.1.1 bereits erwähnte MP3-Player diente dazu, die Wörter in annähernd gleicher



Grundfrequenz aufnehmen zu können. So konnte der Sprecher seine Stimmlage den vorangegangenen Aufnahmen angleichen und die Frequenz über die Vielzahl der Aussprachetermine annähernd konstant halten. Vor Aufnahmebeginn wurde eine Liste aller aufzusprechenden Wörter erstellt, die dann während der Aufnahme als Strichliste diente. So konnte ohne größeren Aufwand festgehalten werden, wie oft der Sprecher jedes einzelne Wort artikulierte. Das hatte den Vorteil, dass der Sprecher sich ausschließlich auf die Aussprache der Wörter konzentrieren konnte und nicht durch das Abzählen einer vorgegebenen Anzahl von Wörtern abgelenkt wurde. Des Weiteren konnte er vor allem schwierige oder misslungene Wörter sofort häufiger aussprechen als es bei einfacheren vonnöten war. Damit vergrößerte sich die Wahrscheinlichkeit, dass in der Aufnahme mindestens ein geeignetes Wort zu finden war. Die Strichliste war auch notwendig, um den späteren Schneidevorgang automatisieren zu können, denn nur so konnten die geschnittenen Items auch richtig benannt werden. Als Mindestzahl wurden fünf Aussprachen für jeden einzigen Einsilber festgelegt. Aus ihnen sollte später die jeweils beste Version ausgewählt werden. Allerdings blieb es dem Sprecher überlassen, die Häufigkeit der Aussprache eines besonders schwierigen Wortes spontan zu erhöhen, wenn er den Eindruck hatte, bisher keine zufriedenstellende Version zustande gebracht zu haben.

Die Qualität der Aussprache lag bei 96000 Hz / 24 Bits. Vor dem ersten Wort, zwischen den einzelnen Wörtern und nach dem zuletzt aufgesprochenen Einsilber wurde jeweils ein Sinuston eingespielt. Besonderes Augenmerk wurde darauf gelegt, eine Überlappung des Sinustons und der Artikulation zu vermeiden. Es folgten immer etwa 36 verschiedene, jeweils mehrmals ausgesprochene Einsilber am Stück. Diese wurden auf der Liste und bei Cubase LE zusammenhängend in einer Spalte eingetragen. Die Zeit zwischen den Aufnahmen der einzelnen Spalten nutzte der Sprecher, um seine Stimme zu entspannen und die Aussprache der folgenden Wörter einzuüben. Auf diese Weise wurden die WAV-Dateien nicht zu groß. Das vereinfachte ihre Übertragung für das späteren Schneiden. Es waren insgesamt sieben Aussprachetermine nötig.

### **2.1.3. Schneiden der Wörter, Abgleichen des SPL und anschließende Qualitätskontrollen**

Der Schneidevorgang der WAV-Dateien wurde automatisiert durchgeführt. Zu diesem Zweck wurde ein Skript mit der Numerik-Software Matlab erstellt. Das Skript

## 2. Versuchsplanung und Methodik

erkannte die Sinustöne als Schnittstellen zwischen den Wörtern. Damit war für jedes Wort jeweils Anfang und Ende markiert. So konnte das Skript erkennen, an welchen Stellen der WAV-Dateien sich die Einsilber befanden. Der eigentliche Beginn und das Ende der Wörter wurde anhand bestimmter Schwellenwerte ermittelt. Von diesem Bereich ausgehend, befanden sich die Schnittstellen genau 0,5 Sekunden zeitversetzt in beide Richtungen. So war zu Beginn der Wort-Dateien nur das Eigenrauschen der Kammer zu vernehmen, das über das Mikrofon mit aufgenommen wurde, und das Eigenrauschen des Mikrophons selbst. Um nun ein Knackgeräusch zu verhindern, das durch abruptes Einschalten von absoluter Stille in das Rauschen auftreten könnte, wurde durch das Skript am Anfang und am Ende jeder Wort-Datei eine Sequenz von 0,5 Sekunden absoluter Stille angehängt. Damit ergab sich ein Zeitraum, in den Rampen integriert werden konnten. Diese Rampen bestanden aus einer von 0 bis 1 ansteigenden Sinuskurve zu Beginn der Wort-Datei und einer von 1 bis 0 abfallenden Sinuskurve am Ende der Datei. Sie waren erforderlich, um sich der tatsächlichen Lautstärke des Wortes anzunähern und diese anschließend wieder abzuschwächen. Dadurch konnten die Einsilber gewissermaßen ein- und ausgeblendet werden. Dies stellte eine fundierte Prävention gegen das Eintreten eines Knackgeräusches dar.

Darauf folgte die Benennung der neu entstandenen WAV-Dateien. Als Grundlage diente die Strichliste, auf der die genaue Anzahl der Aussprachen jedes Einsilbers festgehalten worden war. Die Anzahl war für die Verteilung der Wörter in der Reihenfolge der Aussprache, in eigenen Dateien, mit dem jeweils richtigen Namen maßgeblich.

Auch die Gleichsetzung des Sound Pressure Levels (SPL) wurde automatisiert durchgeführt, indem die Maximalwerte der SPL-Zeitverläufe in Dezibel aller aufgenommenen Wörter aneinander angeglichen wurden. Dies hatte zum Ziel, alle Einsilber in der gleichen Lautstärke verwenden zu können. Das Skript maß den Schallpegelverlauf aller 1554 Einsilber und glich die jeweiligen Maximalwerte aneinander an. Dabei durfte der höchste Ausschlag der WAV-Dateien bei einer Verstärkung des Signals die obere Grenze der Codierbarkeit nicht überschreiten, da dies zu einem „Clippen“ und somit zu spektralen, deutlich hörbaren Verzerrungen geführt hätte.

Als nächstes musste die Qualität aller aufgesprochenen Einsilber überprüft werden. Dazu wurden die Wörter in einer Reihenfolge, die weder der Reihenfolge der Aufnahme noch der alphabetischen Reihenfolge entsprach, abgehört. Dadurch sollten Klangabweichungen deutlicher auffallen als beim alphabetischen Abhören der

## 2.2. Versuchsvorbereitungen für die Evaluation des neuen Sprachmaterials

Wörter. Überprüft wurden die Artikulation und Betonung des Wortes selbst, die Natürlichkeit der Aussprache, Hintergrundgeräusche, Nebengeräusche des Artikulationsraumes, Frequenz der Aufsprachen, subjektiv empfundene Lautheit der Wörter und die allgemeine Qualität der Aufsprache. Wie auch schon in der Dissertation von Mahfoud beschrieben, wurden auch bei diesen Aufnahmen teilweise erhebliche Differenzen in der Lautheit und in der Grundfrequenz der Wörter festgestellt. Deshalb ist die Forderung nach einem Prüfmaterial, das in diesen Bezugspunkten exakt gleich ist, nicht erfüllbar. Diese Anforderungen mussten also etwas relativiert werden. Trotzdem wurden alle diejenigen Wörter erneut aufgesprochen, bei denen gravierende Schwankungen in Frequenz oder Lautheit auftraten, während geringe Schwankungen toleriert werden konnten. Siehe hierzu auch Kapitel 4. Die Ergebnisse der Qualitätskontrollen wurden anschließend noch einmal überprüft und die Wörter, die die Qualitätskontrollen überstanden hatten, in einer neuen Reihenfolge gegengehört. Die Aufnahmen wurden so oft wiederholt, bis die Wörter in allen Punkten unseren Qualitätsansprüchen genügten. Am Ende lagen alle 1554 Einsilber richtig benannt als einzelne WAV-Dateien in der angestrebten Qualität zur weiteren Verarbeitung vor.

## 2.2. Versuchsvorbereitungen für die Evaluation des neuen Sprachmaterials

Die Evaluation des männlichen Sprachmaterials hatte zum Ziel, die Schwierigkeit jedes einzelnen Einsilbers zu ermitteln und die Beziehung zwischen der Bekanntheit eines Wortes und dessen Verständlichkeit festzustellen. Außerdem sollten weitere Einflussfaktoren auf die Verständlichkeit der Einsilber herausgearbeitet werden. Dazu mussten zunächst folgende Parameter festgelegt werden:

- Anzahl der Probanden
- Art der Probanden (normalhörend vs. hörgeschädigt)
- Alter und Geschlecht der Probanden
- Umfang des Tests
- Stärke und Art des Rauschens bei der Testung an normalhörenden Probanden
- Jeweilige Reihenfolge der Einsilber für die einzelnen Probanden

Wir entschieden uns für Versuche mit einer homogenen Gruppe junger normalhörender Probanden. Die Geschlechtsverteilung sollte gleich sein. Aufgrund der Nor-

## 2. Versuchsplanung und Methodik

malhörigkeit mussten die Einsilber vor Versuchsbeginn mit einem definierten Rauschen belegt werden, um verwertbare Ergebnisse erhalten zu können. Auf das Rauschen wird in Abschnitt 2.2.1 näher eingegangen. Nach einer Zeitaufwandsabschätzung, die etwa dreieinhalb Stunden erwarten ließ, schien es möglich, jeden Probanden die Gesamtliste aller Wörter abhören zu lassen. Aus diesem Grund entschieden wir uns dafür, den Test mit 20 Probanden durchzuführen. Die Alternative, 40 Probanden jeweils nur die Hälfte der Einsilber abhören zu lassen, hätte durch individuelle Unterschiede der Probanden ggf. eine zusätzliche Varianz verursacht.

### 2.2.1. Belegen der Einsilber mit einem SNR von -8 dB

Um verwertbare Ergebnisse zu erhalten, sollte nun ein nichtsprachliches Störgeräusch auf die Sprachsignale gelegt werden, wodurch eine Verschlechterung der Verständlichkeit erreicht wird (Schlöndorff und Tegtmeier, 1973, s. Seite 31). Mahfoud hatte in einer vorangegangenen vergleichenden Untersuchung das neue Sprachmaterial M-2007 und W-2007 mit den Aufnahmen des Freiburger Sprachverständnistests von 1968 und 1976 verglichen. Dazu waren alle vier verschiedenen Aufnahmen gefünftelt worden und jedes Fünftel jeweils mit einem Rauschen so belegt worden, dass ein Signal to Noise Ratio (SNR) von 0 dB, -4 dB, -8 dB, -12 dB und -16 dB erreicht wurde. Der SNR gibt den Rauschabstand an, was bei -8 dB zum Beispiel bedeutet, dass das Rauschen um 8 dB lauter ist, als das Signal selbst. Mit -8,54 dB war dies in der Pilotstudie von Mahfoud für den männlichen Sprecher auch der Bereich, in dem die Probanden den SRT erreichten (Mahfoud, 2009, s. Seite 42). Wie in Abschnitt 1.1 bereits beschrieben, ist der SRT der Punkt auf der Regressionskurve, an dem genau die Hälfte des angebotenen Sprachmaterials verstanden wird, und an dem schwerhörige Patienten fortlaufender Sprache gerade noch folgen können. Er ist einfacher und genauer zu messen, als das 100-prozentige Verständnis der Wörter, das zum Beispiel bei einigen Patienten mit einer Innenohrschwerhörigkeit gar nicht mehr erreicht werden kann.

Für die Evaluation der Einsilber musste also die Stärke und die Art des Rauschens festgelegt werden. Wir entschieden uns für ein Rauschen von -8 dB. Auf die Gründe für diese geringfügige Vereinfachung der Prüfsituation wird unter Punkt 4.2.3 näher eingegangen. Für die Versuche sollte genau der SNR angestrebt werden, bei dem die Probanden den SRT erreichen. Ist das gewählte Rauschen zu laut, dann werden die meisten richtigen Wörter nur noch erraten. Ist es aber zu leise, wird alles

## 2.2. Versuchsvorbereitungen für die Evaluation des neuen Sprachmaterials

verstanden und die Versuche sind nicht mehr aussagekräftig. Die Verwendung eines Störrauschens ist notwendig, um von den normalhörenden Probanden verwertbare Ergebnisse zu erhalten. Aber auch bei schwerhörigen Patienten kann ein Störrauschen notwendig sein, um das Kommunikationshandicap wirklichkeitsnah zu bestimmen (Schultz-Coulon, 1974, s. Seite 735).

Für die Sprachtests suchten wir nach einem Störgeräusch, das sprachähnlich ist, definierbare akustische Eigenschaften besitzt, kalibrierbar und reproduzierbar ist, und das sich zur Anwendung bei Sprachtests eignet. Diese Anforderungen hatte auch Döring bei der Konstruktion des „Dreieinsilbertests“ mit Störschall gestellt (Döring und Hamacher, 1992, s. Seite 147). Grundsätzlich in Frage kamen unmodulierte Störgeräusche mit sprachähnlichem Frequenzspektrum, wie zum Beispiel Stimmengewirr, „Party“-Geräusche sowie „sprachsimulierendes“ Rauschen und Störgeräusche mit Modulation, die nach Brand besonders für die Diagnostik schwerhöriger Patienten geeignet sind, wie zum Beispiel das Fastl-Rauschen (Brand, 2007, s. Seite 1). Für die Evaluation des Sprachmaterials M-2007 war ein fluktuierendes Störrauschen tendenziell ungeeignet, da Normalhörende in den „Lücken“ des Störgeräusches einen Teil der Sprache noch relativ gut erkennen können, weshalb ihr Wortverständnis durch diese Art des Störrauschens nur in geringem Ausmaß beeinträchtigt wird (Kollmeier, 2009, s. Seite 90). Außerdem hängt die Störwirkung ganz wesentlich von der Pulsrate der Fluktuation ab (Schultz-Coulon, 1974, s. Seite 736).

Aus diesem Grund entschieden wir uns gegen das Fastl-Rauschen und für die Empfehlung G227 des Comité Consultatif International Télégraphique et Téléphonique, kurz CCITT-Rauschen nach Niemeyer. Das Spektrum des CCITT-Rauschens ist dem Spektrum der Sprache sehr ähnlich, weshalb es gut geeignet ist, das Sprachgewirr einer Cocktailparty zu simulieren (Fastl, 1987, s. Seite 7). Weiterhin ist von Bedeutung, dass die verdeckende Wirkung eines Störschalles umso größer ausfällt, je ähnlicher es dem Nutzschall ist. Dabei sind die spektralen und die zeitlichen Eigenschaften des Störschalls von Bedeutung (Döring und Hamacher, 1992, s. Seite 147). Zusätzlich handelt es sich bei dem CCITT-Rauschen um ein genormtes Rauschen, weshalb es große Verwendung in der heutigen Audiometrie findet. Das CCITT-Rauschen wird durch Filterung des weißen Rauschens erzeugt. Das weiße Rauschen an sich muss als eher umweltfremd betrachtet werden, da es hauptsächlich in hochfrequenten Bereichen verdeckend wirkt (Schultz-Coulon, 1974, s. Seite 736). Das CCITT-Rauschen dagegen enthält im hohen Frequenzbereich keine und im mittleren Frequenzbereich nur geringe Schallstärken, während die Hauptenergie

## 2. Versuchsplanung und Methodik

im Bereich tiefer Frequenzen liegt (Wömpner, 2003, s. Seite 8). Damit hat es, wie die meisten Umweltgeräusche, die Hauptenergieanteile im tieffrequenten Bereich unter 1 kHz (Schultz-Coulon, 1974, s. Seite 736).

Das Rauschen wurde so auf die Einsilber gelegt, dass es das Wortsignal jeweils mit einer Ein- und Ausblendrampe um 300 ms überragte. So wirkten der Anfang und der Abschluss der Wörter nicht zu abrupt.

### Erzeugung der „non-frozen“ Varianten der Einsilber

Um eine etwaige Fehlerquelle auszuschließen, entschlossen wir uns, das Rauschen weiter zu überarbeiten. In der ersten Version der Verrauschung war bei allen 20×1554 Einsilbern derselbe Abschnitt des CCITT-Rauschens verwendet worden. Obwohl es sich um ein unmoduliertes Rauschen handelt, mussten wir in Betracht ziehen, dass bestimmte Wortanteile eventuell eher verdeckt werden könnten, als andere. Wenn nun für alle Patienten immer die gleichen Wortanteile stärker und andere Wortanteile schwächer verdeckt sind, könnte dies eine Verständlichkeit oder Unverständlichkeit einiger Wörter vortäuschen, die sich bei nachfolgenden Untersuchungen mit schwerhörigen Patienten nicht bestätigen würden. Also beschlossen wir, die Einsilber für jeden Probanden zwar mit demselben Rauschtyp, aber mit verschiedenen Abschnitten aus einem langen Rauschsignal zu versehen. Das CCITT-Rauschen wurde daher in unterschiedliche Abschnitte unterteilt, von denen jeder Einsilber einen anderen erhielt. So hörte jeder der zwanzig Probanden zum Beispiel das Wort *Minsk* mit einem anderen Abschnitt des CCITT-Rauschens ab. Die Einsilber wurden also erneut mit einem Rauschen von -8 dB belegt. Dabei wurden dieselben Kriterien eingehalten, die in Abschnitt 2.2.1 beschrieben sind. Somit standen jetzt alle Einsilber in einer „non-frozen“ Variante zur Verfügung, denn sie waren mit einem definierten Rauschen, welches für jeden Probanden individuell und zufällig auf die Einsilber gelegt worden war, versehen.

### 2.2.2. Erstellung der „silence-padded“ Versionen der WAV-Dateien

Beim probeweisen Abspielen der WAV-Dateien in der für die Versuche vorgesehenen Audiometrikammer war bei einer Vielzahl der Wörter gegen Ende der WAV-Dateien ein Störsignal aufgefallen. Dies äußerte sich als ein Knacken in dem dem Wort nachfolgenden Rauschen. Obwohl dieses Störgeräusch sich nicht direkt im Wort befand, musste eine Lösung für die Entfernung des Störsignals gesucht werden. Diejenigen Einsilber, bei denen dieses Störgeräusch als besonders gravierend aufgefallen war,

wurden erneut mittels Sennheiser HD 570 Kopfhörern vom WIN Notebook abgehört und ihre Güte mittels Audacity, einem freien Audioeditor und Recorder, analysiert. Bei dem Abhören über die Kopfhörer von dem WIN Notebook traten die Störgeräusche jedoch nicht auf. Auch konnten über Cubase LE keine Nebengeräusche herausgefiltert werden. Daraus schlossen wir, dass die Ursache nicht in den WAV-Dateien selber, sondern in den technischen Bedingungen der Audiometrikammer zu finden war. Am wahrscheinlichsten schien uns die Abspiel-Software, evtl. in Kombination mit dem (veralteten) Betriebssystem, die Ursache zu sein. Da die technischen Geräte nicht ausgetauscht werden konnten, wurde das Problem umgangen, indem an sämtliche WAV-Dateien eine Sekunde absolute Stille angehängt wurde. Bei dem darauf folgenden Abhören der Wörter in der Audiometrikabine konnte in keinem dieser „silence-padded“ Versionen der WAV-Dateien noch ein Störgeräusch festgestellt werden.

### **2.2.3. Erstellung der 20 verschiedenen Listen für die jeweiligen Probanden: Mechanismus der Verwürfelung**

Um für jeden der 20 Probanden eine unterschiedliche Reihenfolge für das Abhören der Wörter festzulegen, wurden die 1554 Einsilber in eine neue, zufällig generierte Reihenfolge gebracht und anschließend in 10 Listen unterteilt. Sechs dieser neu entstandenen Listen enthielten 155 Wörter, während die übrigen vier Listen jeweils 156 Wörter enthielten. Der Inhalt dieser Listen änderte sich in der folgenden Verarbeitung nicht mehr. Die 10 Listen wurden für die 20 Probanden jeweils fortlaufend verwürfelt. So erhielt zum Beispiel Proband Nr.1 die Listen in der Reihenfolge 1, 2, 3, . . . , 10; während Proband Nr. 2 die Listen in der Reihenfolge 2, 3, 4, . . . , 10, 1 erhielt. Dieser Mechanismus wurde für alle Probanden weitergeführt, so dass jeweils die Probanden 1 & 11, 2 & 12, etc., die gleiche Reihenfolge der Listen hatten. Siehe Tabelle 2.1 auf Seite 30. Ferner mussten noch die Wörter innerhalb der Listen verwürfelt werden, damit keines der Wörter stets am Anfang oder am Ende einer Liste abgehört wurde. Durch mögliche Ermüdungseffekte hätte das die Ergebnisse eventuell verfälscht. Erreicht wurde diese Verwürfelung, indem die WAV-Dateien mittels einer vorangestellten dreistelligen Ziffer umbenannt, und anschließend in eine Zufallsreihenfolge gebracht wurden. So bildeten sie die Grundlage für die neu entstandene Reihenfolge der WAV-Dateien. Dadurch wurden die Einsilber für jeden Probanden in eine individuelle Reihenfolge gebracht.

Im Ergebnis befanden sich für die 20 Probanden jeweils zehn fertig zusammen-

## 2. Versuchsplanung und Methodik

Tabelle 2.1.: Permutation der Testlisten

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

gestellte Blöcke in einem Ordner, die direkt für die Hörversuche verwendet werden konnten. Die Gesamtliste, die die Ordner aller Probanden enthielt, hatte eine Größe von etwa 4,7 GB.

Für die spätere Kalibrierung der richtigen Lautheit wurde eine CD gebrannt, die die Listen der ersten drei Probanden enthielt. Damit war die Voraussetzung geschaffen, dass jeder Einsilber von den Probanden zu einem anderen Zeitpunkt innerhalb des Versuches abgehört werden konnte. Geht man davon aus, dass die Probanden am Anfang des Versuches eine Eingewöhnungsphase benötigen, bis sie sich an die Testsituation gewöhnt haben, und dass am Ende des Versuches eine Ermüdung eintritt, so ist diese potentielle Fehlerquelle durch die systematische Verwürfelung der WAV-Dateien ausgeschlossen worden.



## 2.3. Evaluation des Sprachmaterials M-2007

### 2.3.1. Auswahl der Probanden

Die Kriterien für die Auswahl der Probanden wurde unter Punkt 2.2 erläutert. Es handelte sich um 20 deutschsprachige Probanden, deren Normalhörigkeit vor Versuchsbeginn mit einem Tonschwellenaudiogramm überprüft wurde. Vorher war festgelegt worden, dass die Probanden im Hauptsprachfeld zwischen 250 Hz und zwei KHz einen maximalen Verlust von 20 dB, und in den Nebenfrequenzen einen maximalen Verlust von 30 dB aufweisen durften. Alle Probanden, die diese Anforderungen nicht erfüllten, wurden von der Teilnahme an dem Versuch ausgeschlossen. Die durchschnittliche Hörfähigkeit der Probanden lag aber weit über diesen festgelegten Minimalwerten. In Abbildung 2.1 können die übereinandergelegten Tonschwellenaudiogramme der Probanden eingesehen werden. Insgesamt nahmen zehn weibli-

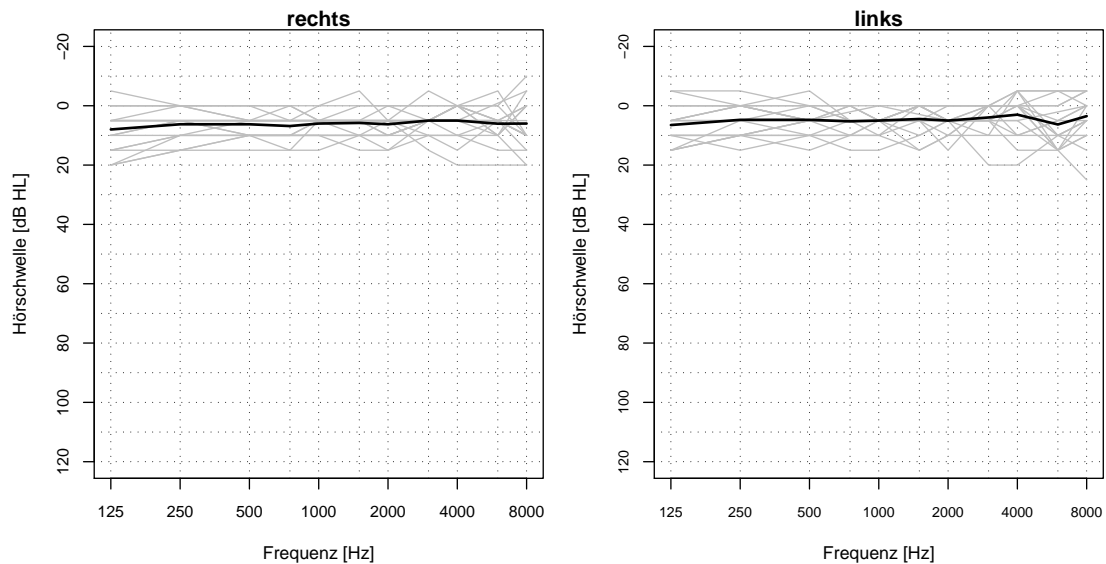


Abbildung 2.1.: Tonschwellenaudiogramme der 20 Probanden

che und zehn männliche Probanden im Alter zwischen 21 und 31 Jahren an dem Versuch teil. Das Durchschnittsalter der weiblichen Probanden lag bei 25,4 und das der männlichen Probanden bei 26,1 Jahren. Es ergab sich ein allgemeiner Altersdurchschnitt von 25,75 Jahren. Des Weiteren wurde auch die Herkunft der Probanden mit aufgenommen. Von den 20 Probanden kamen 6 aus Norddeutschland (vorwiegend Schleswig-Holstein), während die restlichen 14 Probanden in Süddeutschland (hauptsächlich Bayern) aufgewachsen waren. Wie in Abschnitt 2.2.1 beschrieben,

## 2. Versuchsplanung und Methodik

hatte keiner der Probanden schon Erfahrungen mit Einsilber-Sprachtests gesammelt und auch beruflich bestanden bei keinem der Probanden Berührungspunkte zu der Entwicklung oder Durchführung von Sprachgehörprüfungen.

### 2.3.2. Räumlichkeiten und verwendete Gerätschaften

Die Versuche wurden in der Kabine C der Audiometrieabteilung der Kopfklinik in Würzburg durchgeführt. Diese Audiometriekammer ist anders aufgebaut als die, in der die Aufnahmen der Einsilber durchgeführt worden waren. Sie besteht aus einem Raum ohne Unterteilungen, in dessen Mitte sich die Steuerungskonsolen befinden. Der Prüfer nahm vor den Steuerkonsolen Platz und saß dem Probanden gegenüber. Dies hatte den Vorteil, dass eine möglichst normale Konversationssituation entstand. Dieser Aufbau war 2007 auch von Mahfoud für die Durchführung der Pilotstudie verwendet worden. In allen vier Raumrichtungen aus Sicht des Patienten befindet sich jeweils ein Lautsprecher, der genau einen Meter von seinen Ohren entfernt ist, wenn er die ihm vorgegebene Position einhält. Die Audiometriekammer ist mit einem Westra-Audiometer CAD-03/1 und einer Westra Audiometriebox LAB-1001 ausgestattet. Für die Kalibrierung des Audiometers wurde der Schallpegelmesser 2260 INVESTIGATOR von Brüel & Kjær verwendet. Die Sprachgehörprüfung wurde im freien Sprachfeld binaural und nicht über Kopfhörer durchgeführt. Die Sprachsignale kamen aus dem Lautsprecher, der sich vor den Patienten befand. Abgespielt wurden die WAV-Dateien über einen Winamp-Player von einem IBM-kompatiblen Computer, der mit einer TYPHON 3D Acoustic Five Audio Card (Soundkarte) ausgestattet ist. Für die Dokumentation der Versuche wurde ein Medion-Notebook von 2007 verwendet.

Vor Versuchsbeginn musste das Sprachaudiometer kalibriert werden. Dies wurde mit Hilfe der fertig verrauschten Einsilber der ersten drei Probanden vorgenommen. Diese Einsilber waren zu diesem Zweck auf Compact Disc gebrannt worden. Um das Sprachsignal bei 65 dB anzubieten, was natürlicher Sprachlautstärke entspricht, musste das verrauschte Signal aufgrund des SNR von -8 dB auf 73 dB eingestellt werden. Bei einer Einstellung des Audiometers auf Freifeldmessung und nominell 70 dB wurde im freien Schallfeld ein Schalldruckpegel von 72 bis 73,5 dB gemessen. Damit lagen die vorgespielten Signale im Bereich des natürlichen Sprachfeldes und wiesen äußerst geringe Schwankungen auf. Bei der Messung war darauf geachtet worden, den Schallpegelmesser in der späteren Position des Kopfes der Probanden aufzustellen.

### 2.3.3. Durchführung der Evaluation

Zunächst wurde eine Instruktion des Versuches mit Einverständniserklärung für die Probanden erstellt. Ein Muster der Einverständniserklärung befindet sich im Anhang in Abschnitt B. Nun konnte mit den Versuchen zur Evaluation der Einsilber begonnen werden. Vor dem eigentlichen Versuchsbeginn wurden die Probanden noch einmal mündlich und schriftlich über den Ablauf des Versuches und seine Ziele aufgeklärt. Wie in Abschnitt 2.3.1 bereits beschrieben, mussten zur Verifizierung der Normalhörigkeit zunächst Tonschwellen-Audiogramme durchgeführt werden. Dabei wurde allerdings nur die Luftleitung getestet. Da ohnehin nur normalhörende Probanden an den Versuchen teilnehmen durften und die Luftleitung bei allen zugelassenen Probanden nicht beeinträchtigt war, war eine Testung der Knochenleitung nicht notwendig.

Bevor mit der Evaluation der Einsilber begonnen wurde, bekamen die Probanden zur Eingewöhnung und Verdeutlichung der Versuchssituation einen sehr kleinen Teil einer anderen Sprachgehörprüfung vorgespielt. Dabei handelte es sich um einen veräuschten Satztest, der von I. Angermüller 2007 entwickelt worden war. Dieser Test besteht aus einer Neuaufsprache des HSM-Satztests in Verbindung mit zusätzlichen Sätzen. Auch dieser Sprachtest ist von Herrn Dipl.-Ing. Stefan Brill aufgesprochen worden. Da es sich um einen Satztest handelt, bei dem eine wesentlich größere Redundanz als bei Einsilbertests möglich ist, betrug der SNR für die Evaluation hier -10 dB, obwohl in der Pilotstudie der SRT bei einem SNR von -9.91 dB ermittelt worden war. Von diesem Satztest wurden allen Probanden dieselben 20 Sätze des Blockes 0 von Proband Nr.1 vorgespielt. Ob die Probanden die Übungssätze richtig verstanden hatten oder nicht, wurde ihnen nicht mitgeteilt.

Nun wurden für jeden Probanden die zehn Blöcke nacheinander abgespielt. Die Länge der Pausen zwischen den einzelnen Blöcken konnte von den Probanden frei bestimmt werden und unterlag keiner Limitierung. Die Einsilber wurden manuell über den Computer der Audiometriekammer abgespielt. Die Antwortgeschwindigkeit der Patienten bestimmte dabei den Abspielrhythmus, der sehr großen interindividuellen Schwankungen unterlag. Die Probanden waren vor Versuchsbeginn aufgefordert worden, nur das zu wiederholen, was sie auch tatsächlich verstanden hatten und nicht über die Richtigkeit ihrer Antwort nachzudenken. Wie bei den Übungssätzen erhielten sie keine Rückmeldung über die Richtigkeit ihrer Antworten.

Die korrekten oder falschen Antworten der Probanden wurden parallel von der Versuchsleiterin in eine vorbereitete OpenOffice™-Tabelle eingetippt, in der sie dem

## 2. Versuchsplanung und Methodik

tatsächlich abgespielten Wort gegenüberstanden. Aus der Tabelle ergab sich ferner, welche Probanden welchen Geschlechts getestet wurden und um welche Liste, welchen Block und welche Wortnummer es sich handelte.

Nach Abschluss des Versuches wurde schlussendlich der Fragebogen zu der Bekanntheit der Wörter an die Probanden ausgeteilt. In diesem sollten sie angeben, welche der in dem Versuch enthaltenen Einsilber ihnen gänzlich unbekannt waren und bei welchen Einsilbern ihnen zwar das Wort, aber nicht dessen Bedeutung bekannt war.

### 2.4. Methodik der statistischen Auswertung

Für die Auswertung der Ergebnisse, aus den Versuchen mit den 20 Probanden, wurde ein Skript mit dem Statistikprogramm „R“ (R Development CoreTeam, 2010) erstellt. Die während der Versuche gewonnenen Daten wurden als Rohdaten in das Programm eingelesen. Der Umfang dieser Datenmenge war beträchtlich. Zusammengefasst ergaben sich für alle 20 Probanden 31080 Einzelwerte. Außer den direkten Versuchsergebnissen galt es, auch die Auswertung des Fragebogens und die beeinflussenden Faktoren in der Statistik zu berücksichtigen. Dabei sollten die Einflussfaktoren möglichst vollständig erfasst und bei der Verteilung der Wörter in Prüflisten berücksichtigt werden. Folgende Fragestellungen wurden bei der Auswertung untersucht:

- Einschätzung des Schwierigkeitsgrades aller einsilbigen Testwörter
- Einzelauswertungen der Probanden / Leistung der Probanden
- Einfluss folgender Faktoren auf die Versuchsergebnisse: Geschlecht der Probanden, Zeitposition der abgehörten Wörter innerhalb des Tests, Bekanntheit der abgehörten Wörter, Wortstruktur, der in dem jeweiligen Einsilber enthaltene Vokal oder Diphthong, die Länge dieses Vokals beziehungsweise Vokal-komplexes und der Wortstamm
- Absolute Anzahl der bekannten Wörter und der bekannten Bedeutungen
- Anzahl der Wörter in Bezug auf die absolute Häufigkeit richtiger Antworten
- Anzahl der Wörter in Bezug auf die absolute Häufigkeit richtiger Antworten, bezogen nur auf die Schnittmenge der bekannten Wörter/ Bedeutungen

- Anzahl der Wörter in Bezug auf die absolute Häufigkeit richtiger Antworten, aber nur aus der Schnittmenge mit den in der Pilotstudie ebenfalls getesteten 360 Wörtern
- Vergleich mit den Versuchsergebnissen der Pilotstudie von Mahfoud

Zunächst galt es, das statistische Skalenniveau festzulegen. Liegen, wie bei diesen Rohdaten, zwei Merkmalsausprägungen vor, so können diese nur gleich oder ungleich sein und werden dann in der Nominalskala angegeben. Bei solchen Daten werden üblicherweise Häufigkeitsauswertungen vorgenommen. Wenn sie der Größe nach geordnet werden können, gehören sie der Rang- oder Ordinalskala an. Die Intervall- oder Verhältnisskala verwendet man, wenn der Abstand zwischen beiden Werten gemessen werden kann (Sachs und Hedderich, 2006, s. Seite 15 f.).

Das Skalenniveau der erhobenen Daten ist metrisch nicht erfassbar und besitzt keine ordinale, sondern nur eine kategoriale Rangfolge und gehört somit der nominalen Skala an. Dieses niedrige Skalenniveau erlaubt, wie oben bereits beschrieben, nur eine Aussage über die Gleichheit oder Ungleichheit zweier Werte. Dies trifft sowohl auf die Daten der Probanden und auf ihre Angaben im Fragebogen als auch auf die Versuchsergebnisse zu. Bei den untersuchten Daten handelt es sich, außer bei den zum Vergleich herangezogenen Werten von Mahfoud 2009, um dichotome Variablen (Wort verstanden / Wort nicht verstanden, bzw. Wort oder Bedeutung bekannt / Wort oder Bedeutung nicht bekannt).

Für die itemspezifische Auswertung der Daten wurde, um etwaige Signifikanz-Zusammenhänge zu ermitteln und um Häufigkeitsverteilungen zu überprüfen, der Chi-Quadrat ( $\chi^2$ ) - Test angewendet. Bei diesem Test wird eine Nullhypothese angenommen und überprüft, die besagt, dass die beiden untersuchten Merkmale voneinander unabhängig sind (Voß, 2004, s. Seite 446 ff.). Dabei können die betrachteten statistischen Merkmale beliebig skaliert sein. Die beobachteten Häufigkeiten der betrachteten Merkmale und ihrer Kategorien werden in eine Kreuztabelle eingetragen. Daraufhin wird die Häufigkeit bestimmt, die zu erwarten gewesen wäre, wenn die betrachteten Merkmale völlig unabhängig voneinander wären. Als nächstes wird für jedes der Felder die Differenz aus beobachteten und erwarteten Häufigkeiten gebildet, quadriert und durch die erwartete Häufigkeit geteilt. Der  $\chi^2$ -Wert wird so durch die Summe der entsprechenden Werte für alle Zeilen gebildet. Zusätzlich muss der Freiheitsgrad beachtet werden. Das Ergebnis lässt eine Aussage darüber zu, ob die untersuchten Parameter einen signifikanten Einfluss aufeinander haben oder nicht. Je kleiner der errechnete Wert ist, desto stärker ist die reale Abhängigkeit der untersuchten Parameter. Ab einem p-Wert von 0,05 wird ein Zusammenhang als si-

## 2. Versuchsplanung und Methodik

gnifikant und ab einem p-Wert von 0,01 als hochsignifikant angesehen. Angewendet wurde der  $\chi^2$ -Test zum Beispiel, um die Abhängigkeit der Bekanntheit der getesteten Wörter und der Wahrscheinlichkeit einer richtigen Antwort zu untersuchen.

Probandenspezifische Zusammenhänge wurden mit Hilfe des Kruskal-Wallis-Tests untersucht. Es handelt sich dabei um einen parameterfreien statistischen Test, der eine Erweiterung des U-Tests darstellt, mit dem untersucht werden kann, ob sich verschiedene Stichproben (d. h. ein mehrstufiger Einflussfaktor) auf den Erwartungswert auswirken. Die hierfür angenommene Nullhypothese besagt, dass mehrere Stichproben aus der gleichen Verteilung stammen (Voß, 2004, s. Seite 484 ff.). Der Test wird ausgeführt, indem die Daten aus allen Stichproben zusammengefasst und in eine Rangfolge gebracht werden. Dann werden den Daten Rangplätze zugeordnet und für jede Stichprobe die Rangsumme berechnet. Der Kruskal-Wallis-Test wurde zum Beispiel verwendet, um den Einfluss der verschiedenen Wortstrukturen auf das richtige Verständnis eines Wortes zu untersuchen.

Auch der Wilcoxon-Mann-Whitney-Test kam bei der statistischen Auswertung der Daten zum Einsatz. Er dient der Überprüfung von Hypothesen zur Lage statistischer Verteilungen (Voß, 2004, s. Seite 480 ff.), zum Beispiel zur Untersuchung eines möglichen Einflusses des Geschlechts der Probanden auf die Korrektheit der Antworten. Anders als beim Kruskal-Wallis-Test können mit dem Wilcoxon-Test nur zwei Stichproben verglichen werden. All diese statistischen Untersuchungsverfahren wurden herangezogen, um etwaige Fehlerquellen bei der Zusammenstellung der Listen zu vermeiden. Denn eine korrekte Zusammenstellung der Wörter in Listen ist nur unter Beachtung aller beeinflussenden Faktoren möglich.

Für die deskriptive Statistik der beobachteten Werte kamen hauptsächlich „Box-and Whisker-Plots“ zum Einsatz. Ein Boxplot gibt die Verteilung der Werte an. Die Abszisse dient der Unterteilung nach möglichen Untergruppen. Die gemessenen Punkte werden zu den Werten auf der Ordinate in einem Koordinatensystem aufgetragen (Bortz, 2005, s. Seite 40). Der Mittlere Querstrich gibt den Medianwert an, also den Wert, von dem alle übrigen in der Weise abweichen, dass die Summe der Absolutbeträge der Abweichungen ein Minimum ergibt. Dies ist derjenige Wert, der die Häufigkeitsverteilung halbiert (Bortz, 2005, s. Seite 36). Die Box zeigt, in welchem Bereich 50 % der Daten liegen. Die beiden Querstriche, die die Box begrenzen, geben die Grenzen des ersten und des dritten Quartils an. Die äußeren beiden Querstriche, auch Whisker genannt, stellen den Maximal- und den Minimalwert dar. Die Box inklusive Whisker macht deutlich, in welchem Bereich der Großteil der Daten liegt. Ausnahmen sind nur die sogenannten Ausreißer.

## 2.5. Zusammenstellung der Listen

### 2.5.1. Vorbereitung

Zunächst musste definiert werden, auf welche Einflussfaktoren bei der Zusammenstellung der Wörter in Listen Rücksicht genommen werden sollte und wie umfangreich die Listen werden sollten. Aus Konstanzgründen entschieden wir uns, Hahlbrocks Beispiel zu folgen und eine Wörteranzahl von 20 Einsilbern pro Liste anzustreben, auch wenn von einigen Kritikern des Freiburger Sprachverständnistests längere Listen gefordert worden waren. Siehe dazu Abschnitt 1.2.3. Auch die vorgesehene Anzahl der Listen sollte in etwa festgelegt werden. Dies war nur eingeschränkt möglich, da mit dem zur Verfügung stehenden Sprachmaterial und unter Berücksichtigung der beschriebenen Einflussfaktoren so viele Listen wie möglich entstehen sollten. Ausschlaggebend für die Zusammenstellung war vor allem die bei dem Versuch ermittelte Schwierigkeit der Einsilber. Auf folgende Faktoren sollte aber außerdem noch Rücksicht genommen werden:

- Bekanntheit der Wörter und die Bekanntheit der Bedeutungen
- Wortstruktur der Einsilber
- der in den Einsilbern enthaltene Vokal, Umlaut oder Diphthong
- Länge des gesprochenen Vokals oder Vokalkomplexes
- Herkunft des Wortstammes: Lehnwörter oder deutschstämmige Wörter
- Zugehörigkeit der Wörter zu den Sprachverständnistests von 1968 und 1976

Nachdem die gesammelten Daten als Rohdaten in das Statistikprogramm eingelesen und strukturiert worden waren, wurde eine Haupttabelle „e“ angelegt. Diese Tabelle enthielt nach der Strukturierung der benötigten Daten alle Informationen über die Einsilber und war der Ausgangspunkt für die nachfolgenden Untersuchungen. Sie bestand aus 20×1554, also 31080 Zeilen (je Zeile eine Stimulus-Antwort-Kombination), und enthielt die Informationen über die Nummer und das Geschlecht des jeweiligen Probanden, die Nummer, die Liste und den Block des jeweiligen Wortes, d. h., die Zeitposition des Wortes innerhalb des Versuches. Die übrigen benötigten Informationen über die Bekanntheit des Wortes, beziehungsweise die Bekanntheit der Bedeutung des jeweiligen Probanden, die Wortstruktur der Wörter, den in den Wörtern enthaltenen Vokal und dessen Länge, sowie die Zugehörigkeit des Wortstammes mussten schrittweise an die bestehende Tabelle „e“ angefügt werden. Die dazu nötigen Informationen wurden teilweise aus der Organisation der Daten und teilweise aus dem ausgeteilten Fragebogen entnommen.

## 2. Versuchsplanung und Methodik

Als erstes erfolgte das Einlesen der Information über die Bekanntheit der Wörter. Die Fragebögen dazu hatten die Patienten nach Versuchsende ausgefüllt. Wie in Abschnitt 2.3.3 bereits beschrieben, sollten sie unterscheiden, ob ihnen das Wort gänzlich unbekannt war, oder ob sie es zwar schon einmal gehört hatten, aber seine Bedeutung nicht kannten. Nachdem die Fragebögen eingelesen worden waren, musste entschieden werden, welche Wörter in die Listen aufgenommen werden sollten und welche nicht. Ausschlaggebend hierfür war, welche Auswirkungen das Nichtkennen eines Wortes auf dessen richtiges Wortverständnis hätte (s. Abschnitt 3.1.3, [Bekanntheitsgrad des abgehörten Wortes](#), S. 57). Nach der Analyse der Daten entschieden wir uns dafür, das Nichtkennen eines Wortes als absolutes Ausschlusskriterium zu betrachten. Das galt nicht nur, wenn das Wort selbst, sondern auch seine Bedeutung unbekannt war. Wir trafen diese Entscheidung, da mit diesem Test das Gehör der Probanden und nicht ihre Allgemeinbildung getestet werden sollte. Außerdem hatten die Kritiker des Freiburger Einsilberhörtestes bemängelt, dass in den Listen immer noch einige recht unbekannte Wörter vertreten seien. Siehe dazu Abschnitt 1.2.3. Ein Auftreten unbekannter Einsilber sollte durch diese Maßnahme möglichst vermieden werden.

Die übrigen Informationen konnten durch Organisation der Rohdaten direkt gewonnen werden. Hierzu erfolgte als nächstes die Ergänzung der Tabelle um eine Spalte, die die Wortstruktur jedes Einsilbers angab. Dazu wurden zunächst die Vokale, beziehungsweise Vokalkomplexe sowie die Konsonanten, beziehungsweise Konsonantenkomplexe und deren Zusammenstellung in vier möglichen Wortstrukturen (VV, VC, CV, CVC) als reguläre Ausdrücke definiert. Die Zuordnung jedes Einsilbers zu einer der vier Wortstrukturen konnte so automatisiert durchgeführt werden. Einige Ausnahmewörter mussten gesondert zugeordnet werden. Dies betraf vor allem die Wörter, die keinen Vokal, sondern stattdessen ein y an der betreffenden Stelle enthalten. Eine genaue Beschreibung der sich ergebenden Wortstrukturen und der korrekten Verteilung befindet sich in Abschnitt 3.1.3 auf Seite 61.

Darauf folgte die Erweiterung der Tabelle um eine Spalte, die den in den Wörtern enthaltenen Vokal wiedergab. Auch hier konnten durch eine Definition regulärer Ausdrücke für die Vokale die meisten Wörter automatisiert zugeordnet werden. Eine Ausnahme waren wiederum die Wörter, die ein y enthielten, ferner die Lehnwörter und diejenigen Wörter, deren Aussprache sich von Ihrer Schreibweise unterscheidet, wie es zum Beispiel bei *Soest* der Fall ist. Siehe hierzu Abschnitt 3.1.3 auf Seite 64.



Die Einteilung der Wörter nach der Länge der in den Wörtern enthaltenen Vokale wurde ebenfalls so weit wie möglich automatisiert durchgeführt. Dazu wurden reguläre Ausdrücke für lange Wörter definiert. In diese Kategorie fielen zum Beispiel alle Einsilber, die einen Doppelvokal enthalten, alle Wörter, die einen Vokal mit einem anschließenden *h* oder *β* enthalten oder alle Einsilber, die ein *ie* enthalten. Auch für die Worte, die einen kurzen Vokal enthalten, wurde eine Reihe von regulären Ausdrücken definiert, von denen die meisten Konsonantenkombinationen wie *bt*, *kt* oder *lp* sich an einen Vokal anschlossen. Alle Wörter, die durch diese Definitionen nicht erfasst wurden, wurden einzeln in die entsprechenden Listen einsortiert. Auch die Ergebnisse der automatisierten Zuordnung wurden noch einmal manuell überprüft. Siehe hierzu auch Abschnitt 3.1.3 auf Seite 68.

Die Information über die Zugehörigkeit des Wortstammes konnte nicht durch die Organisation der Daten gewonnen werden. Die benötigten Daten wurden dem Herkunftswörterbuch des Dudens entnommen (Duden, 2006b). Siehe auch Abschnitt 3.1.3 auf Seite 70. Nach Aufnahme aller erforderlichen Informationen in die Tabelle „e“ wurden diese in eine neue Tabelle „et“ umgerechnet. Hierbei handelte es sich um eine einsilberspezifische Worttabelle, die nur 1554 Zeilen hatte, d. h. eine Zeile pro Einsilber. Zusätzlich enthielt sie Informationen darüber, welche Wörter schon in den Aufnahmen des Freiburger Einsilberverständnistests von 1968 und 1976 enthalten gewesen waren und welche Wörter aus dem Sprachmaterial M-2007 in der Pilotstudie von Mahfoud getestet worden waren. Die hierfür benötigten Informationen wurden entsprechenden Listen entnommen, die Mahfoud für ihre Arbeit zusammengestellt hatte. Die so entstandene Tabelle war die Ausgangstabelle für die Verteilung der Einsilber in die entsprechenden Listen.

### 2.5.2. Vorgehensweise

Für die Verteilung der Wörter in Listen gab es zwei grundsätzlich verschiedene Vorgehensweisen, die beide durchgeführt und anschließend gegeneinander abgewogen wurden.

Zum einen bestand die Möglichkeit, die Wörter des zur Verfügung stehenden Sprachmaterials mit Hilfe einer Zufallsverteilung in möglichst viele Listen à 20 Wörter aufzuteilen und sie in mehreren nachgeschalteten Arbeitsschritten immer weiter zu optimieren, bis sie in Bezug auf ihren Schwierigkeitsgrad und die Einflussfaktoren möglichst ausgewogen waren. Diese flexible Zuordnung setzt voraus, dass Schwankungen in gewissem Maß akzeptiert werden müssen. Bei Verwendung aller 1554

## 2. Versuchsplanung und Methodik

Einsilber und einer Listengröße von 20 Testwörtern würde so eine maximale Anzahl von 77 Gruppen resultieren. Wird nur die Untermenge der vollständig bekannten Wörter betrachtet, existieren noch 1286 Einsilber, die in maximal 64 Listen aufgeteilt werden können.

Als Alternative bot sich eine fixe Zuordnung der Wörter in Bezug auf ihren Schwierigkeitsgrad an. Hierbei mussten die Wörter eines Korrektheitsgrades vor Verteilungsbeginn eliminiert werden, da es insgesamt 21 mögliche Korrektheitsgrade von 0 bis 20 in 0,5er Schritten gab. Im Fall dieser Herangehensweise entschieden wir uns, die Wörter, die von allen Probanden verstanden worden waren, aus dem Test zu entfernen. Die Pilotstudie von Mahfoud hatte ergeben, dass das Testmaterial M-2007 wesentlich besser zu verstehen ist, als die heute noch gebräuchliche Version des Freiburger Einsilbertests von 1968 (Mahfoud, 2009, s. Seite 42 f.). Dies war der Grund, die leichten Wörter aus dem Testmaterial zu entfernen, da ansonsten die Diskrepanz des Schwierigkeitsgrades zwischen den beiden Einsilbertests noch größer geworden wäre. Natürlich wäre es ebenfalls möglich gewesen, einen anderen Schwierigkeitsgrad zu eliminieren, was zu einer größeren Anzahl von Testlisten mit einem leichteren Schwierigkeitsgrad geführt hätte. Zum Beispiel hätte die Eliminierung derjenigen Wörter, die von 16 Probanden korrekt verstanden worden waren, das Zustandekommen von 45 Listen ermöglicht. Bei dieser fixen Verteilung wurden die Listen vorerst so gestaltet, dass jede Liste jeweils ein Wort eines jeden Korrektheitsgrades erhielt mit Ausnahme den vollständig verstandenen Wörter. Auf die anderen Einflussfaktoren wurde bei diesem Schritt noch keine Rücksicht genommen. Damit ergab sich für die resultierenden Listen ein Schwierigkeitsgrad von 0,475 und eine maximale Listenanzahl von 42. Somit war dieser Schwierigkeitsgrad noch weit entfernt von dem Schwierigkeitsgrad des heute gebräuchlichen Freiburger Einsilbertests von 1968, der bei 0,24 liegt. Der Zufallsfaktor bei der Verteilung der Wörter blieb damit gewahrt, allerdings unter der strikten Vorgabe der Einhaltung eines identischen Korrektheitsgrades. Und ebenso wie bei der zuerst beschriebenen Alternative waren auch hier mehrere nachgeschaltete Optimierungsvorgänge notwendig.

Beide Möglichkeiten hatten Vor- und Nachteile. Mit der ersten Möglichkeit war eine relativ große Listenanzahl erreichbar, allerdings mit dem Resultat größerer Varianz der Schwierigkeitsgrade der Listen. Bei der zweiten Möglichkeit ergab sich eine starke Einschränkung der Gruppenanzahl, da die maximale Gruppenanzahl von der Wortanzahl des minimal verstandenen Korrektheitsgrades festgelegt wird. Außerdem waren die resultierenden Listen noch viel leichter als die des Freiburger Einsilber-

tests, und in ihrem Schwierigkeitsgrad nur noch in eine Richtung veränderbar. Dies würde eine deutliche Einschränkung der Einsetzbarkeit des neuen Sprachmaterials bedeuten und ihn unter anderem für die Erstellung von Gutachten auf der Grundlage der Tabellen zur Ermittlung der MdE ausschließen. Siehe dazu Abschnitt 4.1 und 4.2.4.

Aufgrund der geringen Anzahl und dem niedrigen resultierenden Schwierigkeitsgrad bei der fixen Zuordnung der Listen entschieden wir uns für die flexible Zuordnung.

Zu diesem Zweck wurde in dem Statistikprogramm „R“ eine Funktion (improve-list) erstellt, die durch Veränderung der Parameter in der Lage sein sollte, verschiedene Varianten der flexiblen Zuordnung vorzunehmen. Innerhalb dieser Funktion wurde eine komplexe Datenstruktur „lt“ erstellt, die alle resultierenden Parameter einer Listenzuordnung enthielt, nämlich

- die Information, welches Wort in welche Liste einsortiert wurde
- welche Worte aus der Liste „et“ überhaupt zugeordnet worden waren
- die Ausgewogenheit der Listen in Bezug auf die verschiedenen Einflussfaktoren
- den durchschnittlichen Korrektheitsgrad jeder Liste
- den durchschnittlichen Bekanntheitsgrad der Wörter einer jeden Liste
- die Anzahl der enthaltenen Lehnwörter, Vokale, Wortstrukturen und kurz gesprochenen Wörter
- die Anzahl der Wörter, die auch schon in den Sprachverständnistests von 1969, 1976 und in der Pilotstudie von Mahfoud enthalten gewesen waren

Der Funktion improve-list wurden die Tabellen „et“ und „lt“ übergeben, wobei „lt“ durch die Funktion manipuliert werden durfte, während „et“ konstant bleiben musste.

Bevor die Funktion mit der Verteilung begann, wurden sogenannte „subsets“ definiert, mit denen die Funktion arbeiten sollte. Dies konnte zum Beispiel eine Auswahl der Wörter sein, die allen Probanden vollständig bekannt waren. Durch diese Limitierung der Wörter nahm zwar die Anzahl der erreichbaren Listen weiter ab, aber dies musste akzeptiert werden, um Listen zu erhalten, die nur das Gehör und nicht die Allgemeinbildung der Patienten testen. Somit wurde nicht die gesamte Tabelle „et“ an die Funktion übergeben, sondern nur der „subset“ dieser Tabelle, der die vollständig bekannten Wörter enthielt. Die bei der Zufallsverteilung nicht ausgewählten Wörter dieser Untermenge sollten später für die Optimierung der Listen zur Verfügung stehen, indem aus ihnen Wort-Alternativen herausgegriffen werden konnten.

## 2. Versuchsplanung und Methodik

Außer den Tabellen „lt“ und „et“ wurde der Funktion das jeweils aktuelle Zielkriterium übergeben. Für den ersten Schritt war dies der Korrektheitsgrad der Listen. In allen darauf folgenden Schritten konnten sogenannte „heilige“ Kriterien übergeben werden, die in den nachfolgenden Optimierungsvorgängen nicht mehr verändert werden durften. Dies waren alle diejenigen Einflussfaktoren, die schon optimiert worden waren, was ab dem zweiten Durchlauf zum Beispiel den Korrektheitsgrad betraf. Als „freie Kriterien“ galten all jene, die nicht gesondert definiert wurden und somit zur möglichst genauen Einhaltung des Zielkriteriums manipuliert werden konnten. Dazu sollten Wörter aus dem Pool der noch nicht benutzten Einsilber und Wörter aus den Listen unter Berücksichtigung der „heiligen“ Kriterien gegeneinander austauschbar sein.

Außerdem wurde die Möglichkeit in das Programm integriert für die Einflussfaktoren einen Idealwert oder eine Ober- und Untergrenze der tolerierbaren Schwankungsbreite zwischen den Listen anzugeben.

Innerhalb der Funktion (improve-list) wurde „lt“ wieder vollständig zusammengefügt und enthielt dann alle wichtigen Informationen die Listen und die Einsilber betreffend, die bei der Verteilung beachtet werden sollten. Außerdem enthielt sie noch das jeweilige Zielkriterium, die „heiligen“ Kriterien, die aus „et“ ausgesuchten Wörter, beziehungsweise ihre Nummern, und den durchschnittlichen Korrektheitsgrad der Listen.

Um den Erfolg der Optimierung überprüfen zu können, wurde nach den jeweiligen Vertauschungsvorgängen ein Globalwert errechnet. Dieser für die Erfolgskontrolle zuständige Wert entstand, indem jeder Einflussfaktor mit einem Gewichtungsfaktor multipliziert wurde. Der Korrektheitsgrad erhielt den Gewichtungsfaktor 0,5, während der Bekanntheitsgrad, die Verteilung der Lehnwörter, die Wortstruktur, die Verteilung der Vokale und deren Länge einen Gewichtungsfaktor von 0,1 erhielt.

Auf das gleichmäßige Vorkommen von Wörtern, die auch in den alten Aufsprachen des Freiburger Einsilbertests von 1968 und 1976 sowie in der Pilotstudie enthalten gewesen waren, wurde bei der Errechnung des Globalkriteriums nicht geachtet.

### **Paarweise Verähnlichung zweier Listen**

Daraufhin wurde eine zweite Funktion erstellt, die „similize two lists“ genannt wurde. Diese Funktion sollte den nachgeschalteten eigentlichen Optimierungsvorgang der

Listen durchführen. Die verschiedenen Kriterien sollten dabei nacheinander betrachtet werden. Wiederum war der Schwierigkeitsgrad der Listen das erste betrachtete Kriterium.

Um das Ziel der Verähnlichung der Listen zu erreichen, sollten zunächst diejenigen zwei Listen aus der Gesamtmenge herausgegriffen werden, die am stärksten von dem durchschnittlichen Schwierigkeitsgrad aller Listen abweichen. Nun ergaben sich wiederum zwei mögliche Alternativen fortzufahren. Zum einen konnten die Listen so lange durch den gegenseitigen Austausch von Wörtern optimiert werden, bis zwischen ihnen keine weitere Verbesserung mehr möglich war, bevor die erneute Berechnung der extremsten Listen durchgeführt wurde. Zum anderen war es möglich, nach dem Austausch des ersten Einsilberpaares eine erneute Berechnung der extremsten Listen vorzunehmen und nun diese Listen miteinander zu verähnlichen. Sollten dies nun immer noch die zuerst betrachteten Listen sein, so sollte ein erneuter Austausch zwischen diesen Listen stattfinden.

Wir entschieden uns für die erste Möglichkeit mit der Vorgabe, dass vor jedem Vertauschungsvorgang berechnet werden sollte, ob der Austausch der ausgewählten Wörter auch wirklich zu einer Verähnlichung beider Listen in Richtung des Gesamtdurchschnitts stattgefunden hatte. Wir hatten in vorgeschalteten Testdurchläufen feststellen müssen, dass es ohne diese Limitierung auch zu einer Überkompensation der Listen mit resultierender Verschlechterung kommen konnte. Diese Möglichkeit konnte nun ausgeschaltet werden, indem ein neuer Parameter „similize success“ gleich *true* oder *false* vor jedem Vertauschungsvorgang angab.

Nun fand also der Ausgleich der beiden extremsten Listen in Bezug auf das Zielkriterium so lange statt, wie der Parameter „similize success“ gleich *true* angab. Als Tauschpartner wurden bei einer besonders schweren Liste die schlecht verstandenen Wörter und bei einer besonders leichten Liste die besonders gut zu verstehenden Einsilber unter der Berücksichtigung der „heiligen“ Kriterien herausgesucht. Falls für die Vertauschung mehrere Wörter zur Verfügung standen, wurden diese zufallsverwürgelt und aus jeder Liste eines zufällig herausgegriffen. So stand jeweils ein Wortpaar für den Austauschvorgang zur Verfügung. Dies wurde zwischen den beiden betrachteten Listen so lange wiederholt, bis keine geeigneten Tauschpartner mehr gefunden wurden. Danach wurden die beiden neuen extremsten Listen aneinander angeglichen. Sollte zwischen zwei Listen „similize success“ gleich *false* sein, bevor ein Austausch stattgefunden hatte, z. B. weil keine geeigneten Tauschpartner gefunden wurden oder weil die beiden Listen schon miteinander ausgetauscht wor-

## 2. Versuchsplanung und Methodik

den waren, wurde die Funktion „similize two lists“ an der Stelle abgebrochen und es wurden die nächsten beiden Listen zur Verähnlichung herausgesucht.

Nun musste noch eine Verwaltung für die Reihenfolge der Vertauschungsvorgänge geschaffen werden. Wie oben bereits beschrieben, sollte der Vorgang mit den beiden extremsten Listen in Bezug auf das jeweilige Zielkriterium begonnen werden. Hatte zwischen diesen Listen allerdings schon ein Austausch stattgefunden und waren sie trotzdem noch die extremsten Listen in Bezug auf das Zielkriterium, so durfte zwischen ihnen kein erneuter Verähnlichungsversuch ohne dazwischengeschaltetem Austausch mit einer anderen Liste stattfinden. Um hierfür eine geeignete Struktur zu entwickeln, wurden zwei Matrizes erstellt. In der ersten Matrix wurden die Differenzen der Zielkriterien eingetragen. Das sich ergebende Raster hatte die Differenzen spiegelverkehrt mit umgekehrtem Vorzeichen. Die untere Hälfte wurde auf Null gesetzt. In dieser Matrix konnten nun durch das Herausgreifen der größten Werte die geeignetsten Tauschpartner ermittelt werden. Dabei wurden die Absolutwerte der Differenzen betrachtet. Kamen mehrere gleich große Differenzen vor, wurde eine zufällig herausgegriffen.

In der zweiten Matrix gleicher Größe wurde eingetragen, welche der Listen schon miteinander vertauscht worden waren. Hierzu wurden zunächst alle Felder gleich *false* gesetzt. Das Listenpaar mit der größten Differenz konnte nur herausgegriffen werden, wenn das entsprechende Feld in der zweiten Matrix auf *false* stand. Fand zwischen den herausgegriffenen Listen ein erfolgreicher Austauschvorgang statt, wurde das entsprechende Feld in der zweiten Matrix auf *true* gesetzt. Wurde allerdings in einem der darauf folgenden Schritte eine der beiden Listen durch einen anderen Austauschvorgang wieder verändert, so wurde die gesamte Zeile, beziehungsweise die gesamte Spalte, wieder auf *false* gesetzt, außer in dem Schnittpunkt mit dem neuen Tauschpartner, da ja nun wieder neue potentielle Tauschpartner zwischen diesen Listen zur Verfügung standen. Auch die Werte in der ersten Matrix, die die Differenzen der Listen angab, mussten nach jedem Austauschvorgang in der entsprechenden Zeile und in der entsprechenden Spalte neu berechnet werden. Dies war natürlich nur bei einem erfolgreichen Austauschvorgang nötig.

Die Listen konnten nun so lange aneinander angeglichen werden, bis alle Listenkombinationen entweder einen erfolgreichen oder, mangels passender Tauschpartner, einen erfolglosen Austauschvorgang durchlaufen hatten.

### **Verbesserung der Listen mithilfe der Reserve der nicht vergebenen Wörter**

Eine weitere Möglichkeit zur Verbesserung der Listen war dabei noch nicht beachtet worden. Dies war der Pool der Wörter, die bei der Zufallsverteilung nicht vergeben worden waren.

Um diese Gruppe von Wörtern in den Verbesserungsvorgang mit einzubeziehen, wurde eine zweite Variante der „similize“-Funktion dem Verbesserungsdurchlauf nachgeschaltet. Diese Funktion sollte sich aber jeweils mit nur einer Liste und dem Pool der Wörter befassen. Dabei sollte keine Annäherung zweier Listen, sondern eine Annäherung einer Liste an einen Idealwert auf Kosten des Pools stattfinden. Als Idealwert stand zum einen der durchschnittliche Schwierigkeitsgrad aller Listen und zum anderen der durchschnittliche Schwierigkeitsgrad aller Listen exklusive der gerade betrachteten Liste zur Verfügung. Wir entschieden uns, letztere Variante als Idealwert zu verwenden, da es sonst zu Veränderungen des Idealwertes durch die Veränderung der betrachteten Liste hätte kommen können. Es wurde jeweils die Liste, die in ihrem Schwierigkeitsgrad am stärksten von dem Mittelwert aller übrigen Listen abwich, an die neue Funktion „similize one list towards value“ übergeben. In der Funktion wurde der Pool nach geeigneten Wörtern für die Verbesserung der Liste durchsucht. Standen diese zur Verfügung, konnte die Liste nach geeigneten Wörtern für die Abgabe an den Pool durchsucht werden. Die geringere Anzahl der dabei gefundenen Wörter bestimmte die Anzahl der austauschbaren Wörter. Standen zum Beispiel drei Wörter für die Aufnahme aus dem Pool, aber nur zwei Wörter für eine sinnvolle Abgabe aus der betrachteten Liste zur Verfügung, so kam es nur zum Austausch von zwei Einsilberpaaren. Welche zwei der drei zur Verfügung stehenden Wörter dabei aus dem Pool aufgenommen wurden, wurde dabei wiederum durch eine Zufallsverwürfelung bestimmt. Enthielt der Pool aber keine oder nicht ausreichend Wörter der gewünschten Schwierigkeitsgrade, so wurde dadurch die Anzahl der austauschbaren Wörter limitiert.

Um nun eine geeignete Struktur für die Festlegung der Reihenfolge der Angleichung der Listen an einen Idealwert zu schaffen, wurde dasselbe Prinzip wie in der „similize two lists“-Funktion angewendet. Nur mussten in diesem Fall keine Matrizes sondern zwei Vektoren erstellt werden. Der erste Vektor enthielt die Differenzen der jeweiligen Liste zu dem durchschnittlichen Schwierigkeitsgrad aller übrigen Listen und der zweite Vektor enthielt die Information, ob diese Liste in dem vorherigen Schritt bereits einen Verbesserungsversuch durch den Pool erfahren hatte. Auch hier wurde der zweite Vektor zunächst überall auf *false* gesetzt. Kam es zwischen dem

## 2. Versuchsplanung und Methodik

Pool und einer Liste zu einem Wort austausch, so wurde das entsprechende Feld in dem zweiten Vektor auf *true* gesetzt. Wie bei der anderen Verwaltungsstruktur wurden bei einem erfolgreichen Austausch alle übrigen Felder wieder gleich *false* gesetzt. So wurde die Reihenfolge, in der die Listen an die Funktion übergeben wurden, und das Abbruchkriterium wie in der ersten Funktion festgelegt.

### **Angleichungskriterium: numerischer Wert vs. Verteilung**

Mit den Funktionen „similize two lists“ und „similize one list towards value“ war eine erste funktionierende Implementierung der zwei wesentlichen Werkzeuge für die Verähnlichung der Listen geschaffen. „Similize two lists“ war von vorneherein auf ein Angleichen der Histogramme ausgelegt worden. Bei „similize one list towards value“ bestand allerdings die Problematik unterschiedlicher Skalierung der verschiedenen Einflussfaktoren.

Der Schwierigkeitsgrad der Listen gehört im Gegensatz zu allen anderen beeinflussenden Faktoren der Intervallskala an, während diese nominal skaliert sind. Für die Erweiterung des Programms auf die übrigen Einflussfaktoren ergab sich dadurch ein deutliches Problem, denn es war nur bei Betrachtung des Schwierigkeitsgrades möglich, einen Mittelwert zu errechnen und diesen als ausschlaggebenden Faktor für die Verähnlichung heranzuziehen. Um die Funktion „similize one list towards value“ auch auf die anderen Kriterien anwenden zu können, mussten die jeweiligen Listen nicht über den errechneten Schwierigkeitsdurchschnitt der übrigen Listen angeglichen werden, sondern über ein „normiertes“ Histogramm, das die Häufigkeiten der jeweiligen Korrektheitsgrade aller Listen außer der betrachteten Liste enthielt, geteilt durch die Anzahl aller Listen abzüglich einer.

Die Ergebnisse waren die jeweiligen Zielwerte der Wortanzahlen für die Liste. Diese Anzahl wurde mit dem tatsächlichen Zustand in der betrachteten Liste verglichen. Wenn die tatsächliche Anzahl um 0,5 nach unten oder nach oben ab, konnte ein Wort aus dem Pool der nicht vergebenen Wörter aufgenommen, bzw. an diesen abgegeben werden. Enthielt der Pool keine oder nicht ausreichend Wörter der gewünschten Schwierigkeitsgrade, so wurde dadurch die Auswahl der austauschbaren Wörter limitiert. Durch diesen Mechanismus wurde das Histogramm der betrachteten Liste an das angestrebte Idealhistogramm angeglichen.



### Globale Methoden für die Verähnlichung der Listen

Mit der Erweiterung der „similize one list towards value“-Funktion auf Histogramm-Angleichung standen beide Werkzeuge zur Verähnlichung der Listen zur Verfügung. Allerdings musste die ideale Reihenfolge für das Anwenden der einzelnen Programmschnitte ermittelt werden. Hierfür wurden vier verschiedene Methoden implementiert:

1. Verähnlichung aller Listenpaare gegeneinander, danach Verähnlichung aller Listen gegenüber dem Pool
2. Verähnlichung aller Listen gegenüber dem Pool, danach Verähnlichung aller Listenpaare gegeneinander
3. Repetitives Vorgehen wie unter 1., limitiert auf max. 100 Durchläufe, ggf. früherer Abbruch, wenn keine Verbesserung mehr möglich war
4. Repetitives Vorgehen wie unter 2., limitiert auf max. 100 Durchläufe, ggf. früherer Abbruch, wenn keine Verbesserung mehr möglich war

Um zu eruieren, welche der vier verschiedenen Methoden das beste Ergebnis erzielt, wurde jede von ihnen in das Auswertungs-Skript integriert und die resultierenden Listen miteinander verglichen. Zu diesem Zweck wurde das Skript „auswertung.R“ vierfach kopiert und als „auswertung-methode1.R“, „auswertung-methode2.R“, etc. gespeichert. Jede dieser Kopien wurde nun so verändert, dass die entsprechende Methodik damit durchgeführt werden konnte.

Zusätzlich wurde ein weiteres Skript „auswertung-methoden-rahmen.R“ erstellt. Dieses Skript verwaltete die vier Skripte, die die unterschiedlichen Methoden ausführten, und sorgte dafür, dass die vier Methoden jeweils einhundertmal wiederholt und die wichtigsten Ergebnisse in einer Tabelle gespeichert wurden. Diese Informationen bestanden für jeden der 100×4 Durchläufe aus der Anzahl der resultierenden Listen, dem Mittelwert, der jeweilige Standardabweichung, dem Medianwert, dem IQR und dem MAD-Wert der jeweiligen Korrektheitsgrade der resultierenden Listen. Außerdem sorgte das Verwaltungsskript dafür, dass jeweils die gesamte Struktur derjenigen Verteilungsversion mit der größten Listenanzahl und derjenigen Verteilungsversion mit dem geringsten MAD-Wert gespeichert wurde.

Dazu wurde zunächst der Parameter des bisher besten MAD-Wertes auf 1 und der Parameter der bisher größten Listenanzahl auf 0 gesetzt. Dadurch wurde sichergestellt, dass die Struktur der ersten Verteilungsversion gespeichert wurde. War der MAD-Wert eines darauf folgenden Durchgangs geringer, so wurde dieser durch die

## *2. Versuchsplanung und Methodik*

neue Version ersetzt. Dasselbe galt für die Anzahl der resultierenden Listen, so dass jeweils die Strukturen von zwei Verteilungsversionen aufgehoben wurden. Bei einem gleichen MAD-Wert der neuesten Version mit der im Speicher enthaltenen Version wurde diejenige Version aufgehoben, die die größere Listenanzahl aufwies. Parallel galt, dass bei gleicher Listenanzahl diejenige gespeichert wurde, die den geringeren MAD-Wert aufwies. Die Ergebnisse dieses Vergleiches der vier verschiedenen Verähnlichungsmethoden können unter Punkt [3.2](#) eingesehen werden.

Damit war die grundsätzliche Methodik für die Zusammenstellung der Listen geschaffen. Das bestehende Programm muss auf die übrigen zu beachtenden Faktoren für das Entstehen von ausgewogenen Listen erweitert werden, was jedoch den Rahmen dieser Arbeit sprengen würde. Aber die notwendige Methodik kann durch das Betrachten der Häufigkeitsverteilungen des Korrektheitsgrades als Histogramme auch auf die anderen beeinflussenden Faktoren übertragen werden, da diese sich ebenfalls als Häufigkeitsverteilungen betrachten lassen. Um diesen Test in der klinischen Praxis nutzbar zu machen, ist eine Erweiterung des Programms in der Zukunft unumgänglich.

# 3. Ergebnisse

## 3.1. Auswertung der Versuche

### 3.1.1. Einschätzung des Schwierigkeitsgrades aller einsilbigen Testwörter

Das primäre Ziel der Versuche zur Evaluation der aufgenommenen Wörter war, die Schwierigkeit jedes einzelnen Einsilbers abschätzen zu können, was für die Verteilung der Wörter in Listen mit einem ausgeglichenen Schwierigkeitsgrad maßgeblich war. Dazu wurden die Antworten der zwanzig Probanden für die Einschätzung eines jeden Wortes zusammengefasst und mit der Anzahl der Probanden verrechnet. Somit erhielt jeder der 1554 Einsilber einen Prozentwert in Fünfprozent-Schritten zwischen 0 und 100 % bedingt durch die Anzahl der Probanden. Es hat sich gezeigt, dass die Probanden durchschnittlich etwa 43,11 % der abgehörten Worte bei einem SNR von -8 dB korrekt verstanden haben. Dieser Wert lag etwas unter dem Wortverständnis, das nach der Pilotstudie zu erwarten gewesen wäre. Dort hatten die Probanden ein fünfzigprozentiges Wortverständnis bei einem SNR von -8,54 dB erreicht. Siehe hierzu auch Abschnitt [2.2.1](#). Auf die möglichen Gründe für diese Differenz wird in Abschnitt [4.2.3](#) näher eingegangen. [Abbildung 3.1](#) zeigt ein Histogramm mit dem die absoluten Häufigkeiten richtiger Antworten dargestellt sind. Ein Histogramm dient der graphischen Veranschaulichung einer Häufigkeitsverteilung von diskontinuierlichen Variablen ([Bortz, 2005](#), s. Seite 31). Auf der Abszisse ist die jeweilige Anzahl der Probanden eingetragen, während auf der Ordinate die absolute Häufigkeit richtiger Antworten verzeichnet ist. Die flächenproportionale Darstellung der vorliegenden Häufigkeiten zeigt, wie viele Wörter zum Beispiel von einem, von acht oder von neunzehn Probanden richtig verstanden wurden. Daraus lässt sich erkennen, dass es weit mehr Wörter gibt, die gar nicht oder nur von wenigen Probanden verstanden wurden, als solche, die von allen oder den meisten Probanden verstanden wurden. So liegt die Zahl der von einem Probanden verstandenen Wörter bei ca. 107, wäh-

### 3. Ergebnisse

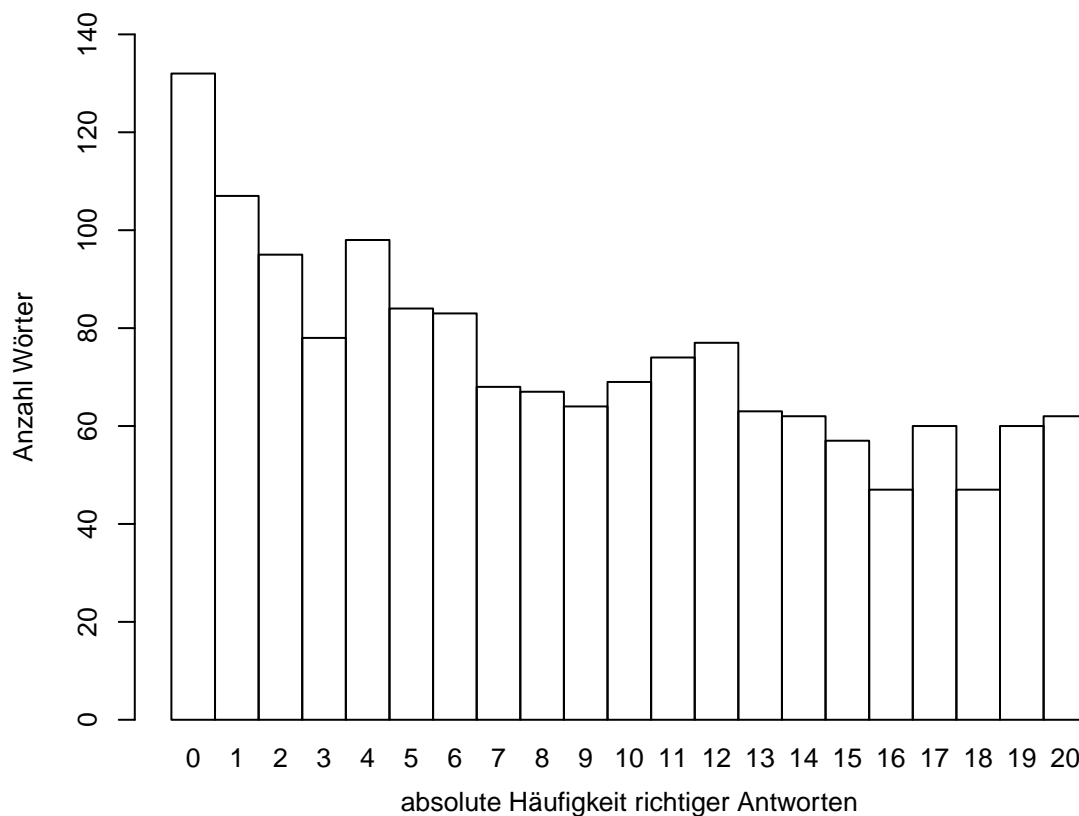


Abbildung 3.1.: Histogramm der Wörterzahl je nach richtigen Antworten

rend die Zahl der von 15 und 17 Probanden richtig verstandenen Wörter bei nur ca. 47 Einsilbern liegt. Natürlich traten auch eine Reihe von Wörtern auf, die entweder von keinem Probanden erkannt wurden, und solche, die von allen Probanden richtig erkannt wurden, wie in Abbildung 3.1 deutlich zu sehen ist. Wie mit diesen Wörtern weiter verfahren wird, wird in Abschnitt 3.1.5 auf Seite 76 noch näher besprochen.

#### 3.1.2. Einzelauswertung der Probanden / Probandenleistung

Um einen etwas genaueren Überblick über die Leistungsschwankungen der Probanden zu erhalten, wurde eine Tabelle erstellt, in der die durchschnittliche Leistung aller Probanden eingetragen wurde. Diese Tabelle gibt die Spannweite der Leistungen der zwanzig Probanden wieder. Insgesamt liegen die Ergebnisse der Probanden recht nah beieinander. Die Probandin mit der geringsten Erfolgsrate hat 34,62 % der angebotenen Wörter korrekt verstanden und der beste Proband hat in 51,09 % der Fälle die korrekte Antwort genannt. Die Extremwerte liegen also 16,47 Prozentpunk-

Tabelle 3.1.: Leistung und Wortkenntnis der Probanden

Proband	Geschlecht	% korrekt	% korrekt (360)	unbek. Wörter	unbek. Bedeutg.
1	F	46,07	50,83	44	60
2	F	34,62	35,83	43	89
3	F	36,55	39,44	41	83
4	F	48,33	49,44	45	61
5	F	41,76	45,28	79	96
6	F	44,79	48,33	13	20
7	F	46,78	50,83	51	60
8	F	43,76	43,89	38	67
9	F	41,76	45,28	3	68
10	F	43,37	48,33	38	62
11	M	45,82	52,22	52	79
12	M	51,09	55,28	69	95
13	M	41,96	46,94	47	74
14	M	39,45	43,61	52	79
15	M	43,37	50,00	29	54
16	M	46,01	50,56	37	37
17	M	41,12	48,33	81	83
18	M	38,22	41,67	32	35
19	M	39,06	43,61	75	113
20	M	48,20	53,89	48	72

te auseinander. Insgesamt gibt es aber nur einen Probanden der etwas weniger als 35% korrekt verstanden hat und einen Probanden der über 50% der angebotenen Wörter korrekt verstanden hat. Abgesehen von diesen beiden Eckwerten liegen die übrigen Ergebnisse sehr viel enger beieinander. Die Werte der meisten Probanden bewegen sich zwischen 36 und 46% und differieren somit nur um 10 Prozentpunkte. Außer der Nummer der Probanden, ihrem Geschlecht und dem Prozentwert der korrekten Antworten über alle getesteten Einsilber gibt die Tabelle noch den Prozentwert der Korrektheit der Schnittmenge der Wörter an, die auch in beiden Aufnahmen der Freiburger Einsilbertests und in der Pilotstudie vorkamen. Außerdem kann der Tabelle entnommen werden, wie viele Wörter beziehungsweise Bedeutungen dem jewei-

### 3. Ergebnisse

ligen Probanden nicht bekannt waren. Auf diese zusätzlichen Informationen wird in den Abschnitten 3.1.3 und 3.1.4 noch näher eingegangen.

#### 3.1.3. Einflussfaktoren der Versuchsergebnisse

Um etwaige Fehlerquellen auszuschalten, sollten diejenigen Faktoren untersucht werden, die die Versuchsergebnisse beeinflussen. Dazu wurden einerseits nicht beeinflussbare Faktoren, wie das Geschlecht der Probanden, und andererseits prüfsystembedingte äußere Faktoren, wie die Zeitposition der abgehörten Wörter innerhalb des Tests, berücksichtigt. Aber auch die Worteigenschaften selbst, wie die Wortstruktur, der in dem jeweiligen Wort enthaltene Vokal, dessen Länge oder die Bekanntheit, beziehungsweise Unbekanntheit der Wörter, waren für diese Untersuchung relevant. Dabei war es auch für die Verteilung der Wörter in Listen ausschlaggebend, welche Worteigenschaften einen Einfluss auf die Versuchsergebnisse hatten.

##### Geschlecht

Als erstes wurde der Einfluss des Geschlechts der Probanden auf die Testergebnisse untersucht. Zu diesem Zweck wurde ein Boxplot erstellt. Siehe Abbildung 3.2. Es wurde das Geschlecht auf der Abszisse gegen die Leistung der Probanden, also der Prozentsatz korrekt verstandener Wörter, auf der Ordinate aufgetragen. Die beiden Ergebnisgruppen wurden als Boxplots in dieser Grafik eingetragen. Bei dieser graphischen Darstellung lässt sich erkennen, dass das Geschlecht keinen signifikanten Einfluss auf die Leistung der Probanden zu haben scheint. Die Medianwerte liegen sehr nah beieinander. Sie werden bei einer geraden Probandenanzahl bestimmt, indem die Messwerte der Probanden der Größe nach geordnet und das arithmetische Mittel zwischen dem größten der zu den unteren 50 % gehörenden Werte und dem darauf folgenden Wert errechnet werden (Bortz, 2005, s. Seite 36 f.). Der Medianwert der weiblichen Probanden liegt bei etwa 43,5 % und der der männlichen Probanden bei etwa 42,7 %. Die oberen Quartile beider Gruppen liegen bei ca. 46 % während das untere Quartil der weiblichen Probanden bei ca. 42 % und das der männlichen Probanden bei etwa 39,5 % liegt. Damit haben die männlichen und die weiblichen Probanden einen sehr ähnlichen Interquartilsabstand. Außerdem ist zu bemerken, dass die Streuung der weiblichen Probanden eine geringfügig rechtsschiefe Verteilung aufweist. Die Extremwerte liegen bei den weiblichen Probanden bei 48,33 % und 36,55 %, während die Extremwerte der männlichen Probanden bei 51,09 % und

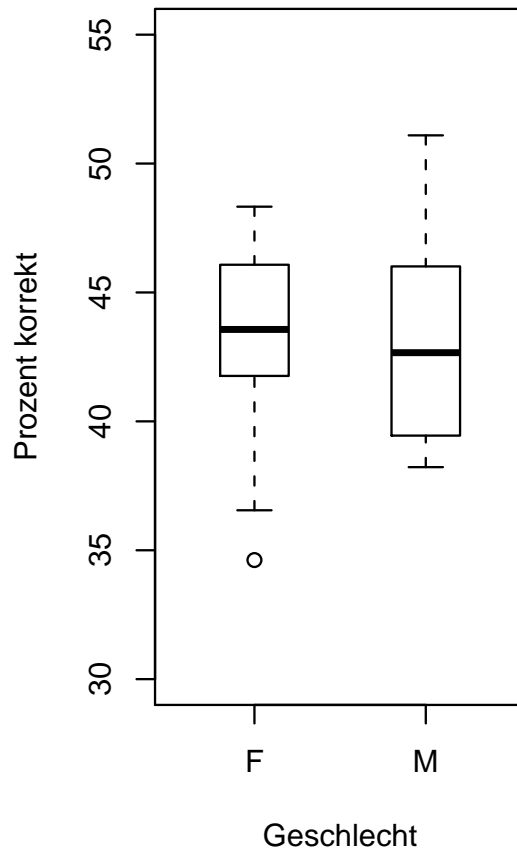


Abbildung 3.2.: Boxplot der Abhängigkeit korrekter Antworten von dem Geschlecht

38,22% liegen. Allerdings ist bei den weiblichen Probanden ein ausreißender Wert auffällig. Dieser durch Probandin Nr. 2 entstandene Wert liegt bei 34,62%. Nach der Definition von John W. Tukey handelt es sich bei diesem Wert aber nur um einen „milden“ Ausreißer, da er sich zwischen  $1,5 \times \text{IQR}$  und  $3 \times \text{IQR}$  befindet. Um einen möglichen Signifikanzzusammenhang genauer zu untersuchen, wurde der Wilcoxon-Test auf diese Daten angewendet. Es ergab sich ein p-Wert von 1,000000. Auch der  $\chi^2$ -Test wurde auf diese Daten angewendet. Es wurde ein p-Wert von 0,252042 errechnet. Damit war bestätigt, dass das Geschlecht die Leistung der Probanden nicht signifikant beeinflusst.

### **Zeitposition des abgehörten Wortes innerhalb des Testverlaufs**

Ein weiterer möglicher Einflussfaktor der Versuchsergebnisse war die Zeitposition des jeweiligen Wortes innerhalb des Versuches. Jeder der 20 Probanden musste al-

### 3. Ergebnisse

le 1554 einsilbigen Hauptwörter abhören. Die Wörter wurden, wie in Abschnitt 2.2.3 bereits beschrieben, in 10 Listen aufgeteilt, von denen vier Listen 156 Wörter und die übrigen sechs Listen 155 Wörter enthielten. Zwischen den jeweiligen Listen wurden Pausen eingehalten, deren Länge von den Probanden selbst bestimmt werden konnte. Während des Abhörens der einzelnen Listen war eine Unterbrechung allerdings nicht möglich. Durch diesen Versuchsaufbau und die unterschiedlichen Antwortgeschwindigkeiten der Probanden schwankte die Dauer des Versuches zwischen drei und viereinhalb Stunden pro Proband. Wir nahmen an, dass die Probanden zu Beginn des Versuches eine kurze Eingewöhnungsphase brauchen würden, bis sie ihr Leistungsmaximum erreichten, und dass bei einer derartig langen Versuchsdauer gegen Ende eine Ermüdungsphase zu erwarten sei, die die Leistung der Probanden sehr wahrscheinlich herabsetzen würde.

Um diesen Zusammenhang zu untersuchen, wurden zwei Liniendiagramme erstellt, auf denen die prozentuale Korrektheit der Antworten gegen die Zeitposition aufgetragen wurde. Abbildung 3.3 zeigt die Korrektheit der Antworten im Zeitverlauf des gesamten Versuches, wobei die Leistung aller 20 Probanden über den Zeitverlauf gemittelt wurde. Die zehn Abschnitte, die durch Pausen voneinander getrennt waren, sind durch die dünnen schwarzen senkrechten Linien dargestellt. Jede der grauen senkrechten Linien symbolisiert die Zeitposition eines einsilbigen Testwortes. Dabei ist zu bemerken, dass diese Linien keinen bestimmten Einsilber symbolisieren, sondern nur die Zeitposition, die bei den 20 Probanden jeweils von einem anderen Einsilber besetzt war.

Über die histogrammartige Darstellung der Einzelwerte sind in orange und in rot zwei geglättete Linien eingezeichnet (Lowess-Glättung). Die weniger geglättete orangefarbene Linie (Glättungsspanne ca. 50) lässt noch die Leistungsschwankungen der Probanden innerhalb der Blöcke erkennen, während die stark geglättete rote Linie (Glättungsspanne ca. 1000) den Leistungsverlauf über den gesamten Versuch wiedergibt. Wie sich in dieser Abbildung deutlich erkennen lässt, trifft unsere Annahme einer Ermüdungsphase nicht zu. Zwar benötigen die Probanden eine Eingewöhnungsphase, um ihr Leistungsmaximum zu erreichen, aber die Leistung sinkt zum Ende des Versuches nicht wieder ab, sondern steigert sich immer weiter. Innerhalb der ersten vier Blöcke kommt es zu einem relativ starken Anstieg der Leistung, dessen Steigung zwar etwas nachlässt, sich aber trotzdem bis zum Ende des Versuches fortsetzt. Abbildung 3.4 zeigt dagegen die Leistung der Probanden im Zeit-



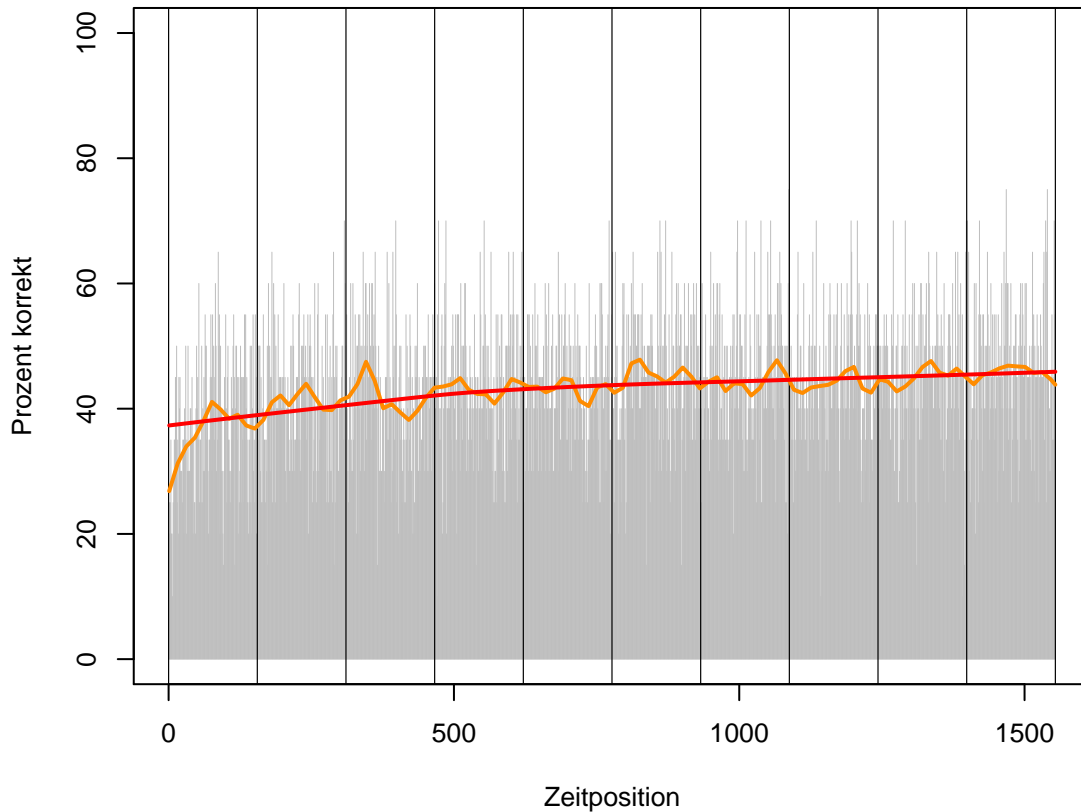


Abbildung 3.3.: Linienplot der Korrektheit im Zeitverlauf des Versuches

verlauf eines Blockes, wobei die Daten aller Blöcke und aller Probanden gemittelt wurden. Wie in [Abbildung 3.3](#) stellen auch hier die senkrechten grauen Linien die Zeitpositionen der Einsilber innerhalb der Blöcke dar. Überraschenderweise scheint eine kurze Eingewöhnungsphase am Anfang eines jeden Blockes notwendig zu sein, bevor die Probanden ihr Leistungsmaximum erreichen. Die schlechtere Leistung zu Beginn der einzelnen Blöcke ist nicht durch die relativ intensive Lernphase in den ersten zwei Blöcken bedingt. Dies wurde dadurch ausgeschlossen, dass diese zwei Blöcke in einer weiteren Graphik (s. [Abb. 3.5](#)) nicht mit einbezogen wurden und nur die Leistung der Probanden im Zeitverlauf der Blöcke 3 bis 10 aufgetragen wurde. Der Verlauf erscheint zwar etwas flacher, aber auch hier erreichen die Probanden ihr Leistungsmaximum erst nach etwa 50 Wörtern. Somit scheint der Gewöhnungseffekt eine wichtige Rolle zu spielen, der sich wahrscheinlich auch bei kurzen Gehörprüfungen, bei denen eine Liste nur etwa 20 Wörter enthält, auf die Ergebnisse auswirkt.

### 3. Ergebnisse

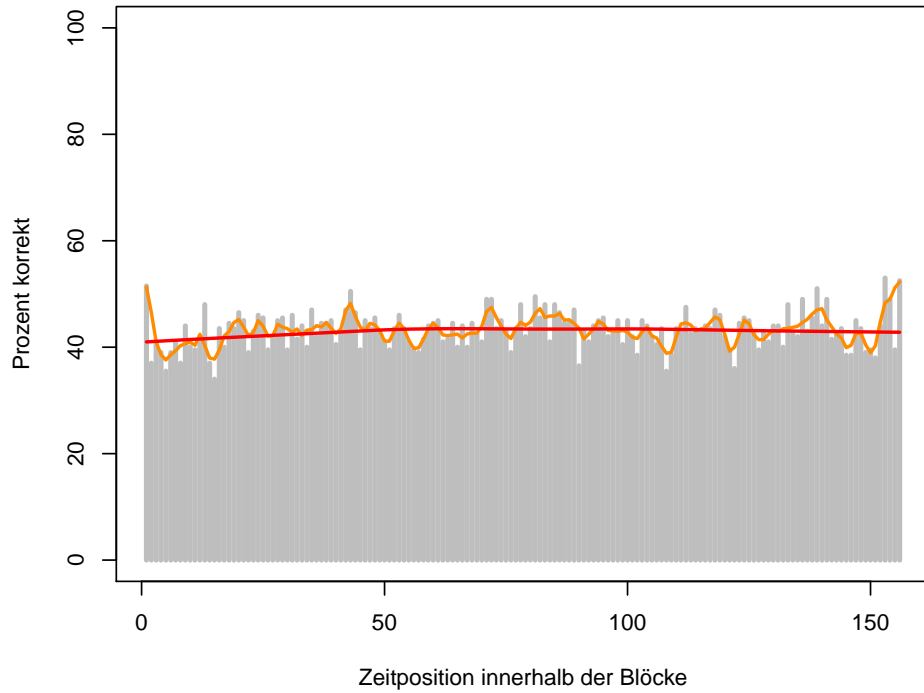


Abbildung 3.4.: Linienplot der Korrektheit im Zeitverlauf der Blöcke

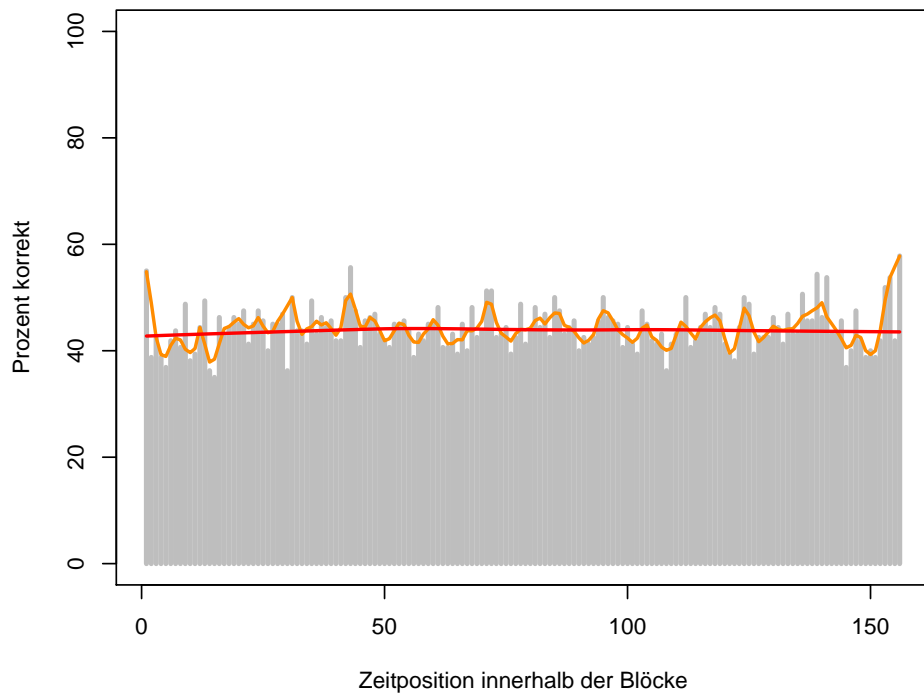


Abbildung 3.5.: Linienplot der Korrektheit im Zeitverlauf der Blöcke, ohne Einbeziehung des 1. und 2. Blocks

#### **Bekanntheitsgrad des abgehörten Wortes**

Die Information über die Bekanntheit der einzelnen Wörter wurde von den Probanden auf dem nach Versuchsende ausgeteilten Fragebogen angegeben. Dabei sollten die Probanden unterscheiden, ob ihnen das Wort gänzlich unbekannt war oder ob ihnen zwar das Wort bekannt, aber seine Bedeutung unbekannt war, was zum Beispiel bei den Wörtern *Schwyz*, *Gracht* oder *Fjaell* häufiger der Fall war. Als gänzlich unbekannt wurden zum Beispiel die Wörter *Kwass* und *Lodsch* häufig angegeben. Eine Zusammenfassung der Antworten dieses Fragebogens befindet sich im Anhang im Abschnitt **D** ab Seite 121. Insgesamt waren 223 Wörter und 286 Bedeutungen den 20 Probanden unbekannt.

Abbildung 3.6 zeigt einen Scatterplot, bzw. ein Streudiagramm, bei dem die relative Häufigkeit richtiger Antworten gegen die Bekanntheit der Wörter aufgetragen wurde. Je heller die eingetragenen Punkte sind, desto weniger Wörter liegen in dem jeweiligen Bereich. Schon in dieser Abbildung wird deutlich, dass diejenigen Wörter, die nur einen geringen Bekanntheitsgrad aufweisen, auch eine schlechte Verständlichkeit haben. Abbildung 3.7 zeigt einen ähnlichen Scatterplot mit dem Unterschied, dass hier die Unbekanntheit nicht nur das Wort selbst, sondern auch seine Bedeutung umfasst. Das sich ergebende Bild ist trotzdem ähnlich. In beiden Abbildungen ist zu erkennen, dass gänzlich unbekannte Wörter auch fast nie richtig erkannt werden, während Clusterbildungen in den Bereichen guter Bekanntheit der Wörter auftreten. Auch in Abbildung 3.8 wird dieser Zusammenhang deutlich. In diesem Histogramm wurde die absolute Häufigkeit richtiger Antworten gegen die Anzahl der Wörter aufgetragen. Die weißen Säulen umfassen dabei das gesamte Sprachmaterial, während die grauen Säulen die Reduktion des Sprachmaterials auf die vollständig bekannten Wörter (als Wort und in der Bedeutung) darstellen. Es wird deutlich, dass die Anzahl der zur Verfügung stehenden Wörter gegenüber dem gesamten Sprachmaterial abnimmt. Allerdings ist diese Abnahme in den Bereichen der schlecht verstandenen Wörter besonders ausgeprägt. Die Häufung der schlecht oder kaum verstandenen Wörter scheint also in einer prozentual größeren Unbekanntheit dieser Wörter begründet zu sein.

Um diesen Zusammenhang noch etwas genauer zu untersuchen, wurde ein  $\chi^2$ -Test durchgeführt. Dafür wurden die Ergebnisse aller Probanden gemittelt. Bei der Untersuchung der Bekanntheit des Wortes an sich und der Korrektheit der dazugehörigen Antworten ergab sich ein p-Wert von  $< 1,0 \times 10^{-6}$ . Dieser Wert zeigt deutlich, dass die Unbekanntheit eines Wortes signifikant mit der steigenden Wahrrscheinlich-

### 3. Ergebnisse

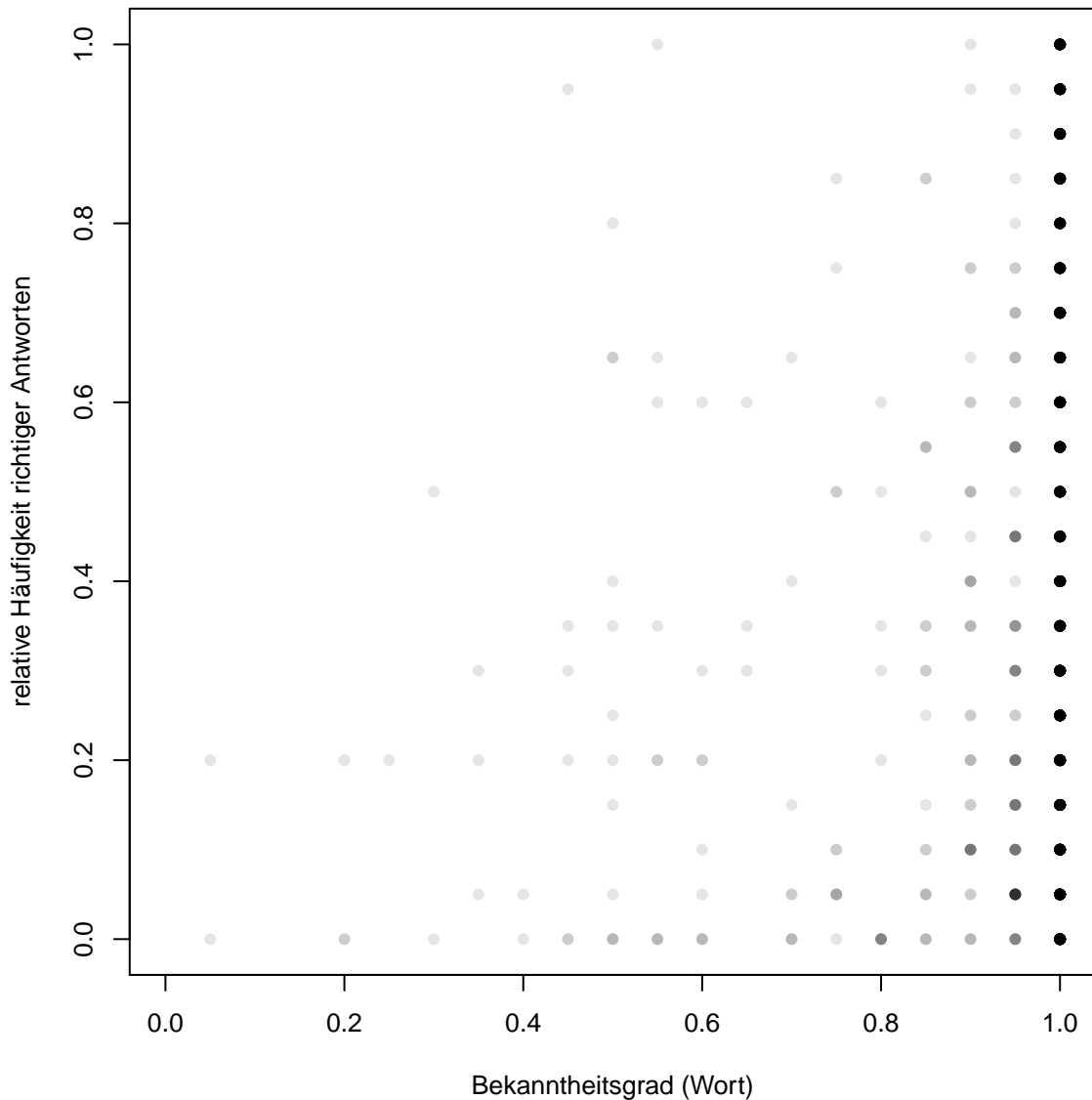


Abbildung 3.6.: Scatterplot der Abhängigkeit der Korrektheit der Antworten von der Bekanntheit der Wörter

keit einer falschen Antwort korreliert. Auch die Unbekanntheit der Bedeutung eines Wortes hat eine signifikante Auswirkung auf die Korrektheit der Antwort. Denn auch hier hat eine Unbekanntheit schlechtere Ergebnisse zur Folge. Der p-Wert liegt bei  $< 1,0 \times 10^{-6}$ . Um einen genaueren Einblick zu erlangen, wurden die p-Werte auch probandenspezifisch errechnet. Die Ergebnisse sind in Tabelle 3.2 zusammengefasst. Die zweite Spalte gibt den p-Wert für die Bekanntheit des Wortes und die dritte Spalte den p-Wert für die Bekanntheit der Bedeutungen an. Hier zeigt sich ein ähnliches Bild. Bei fast allen Probanden hat die Unbekanntheit des Wortes selbst oder seiner

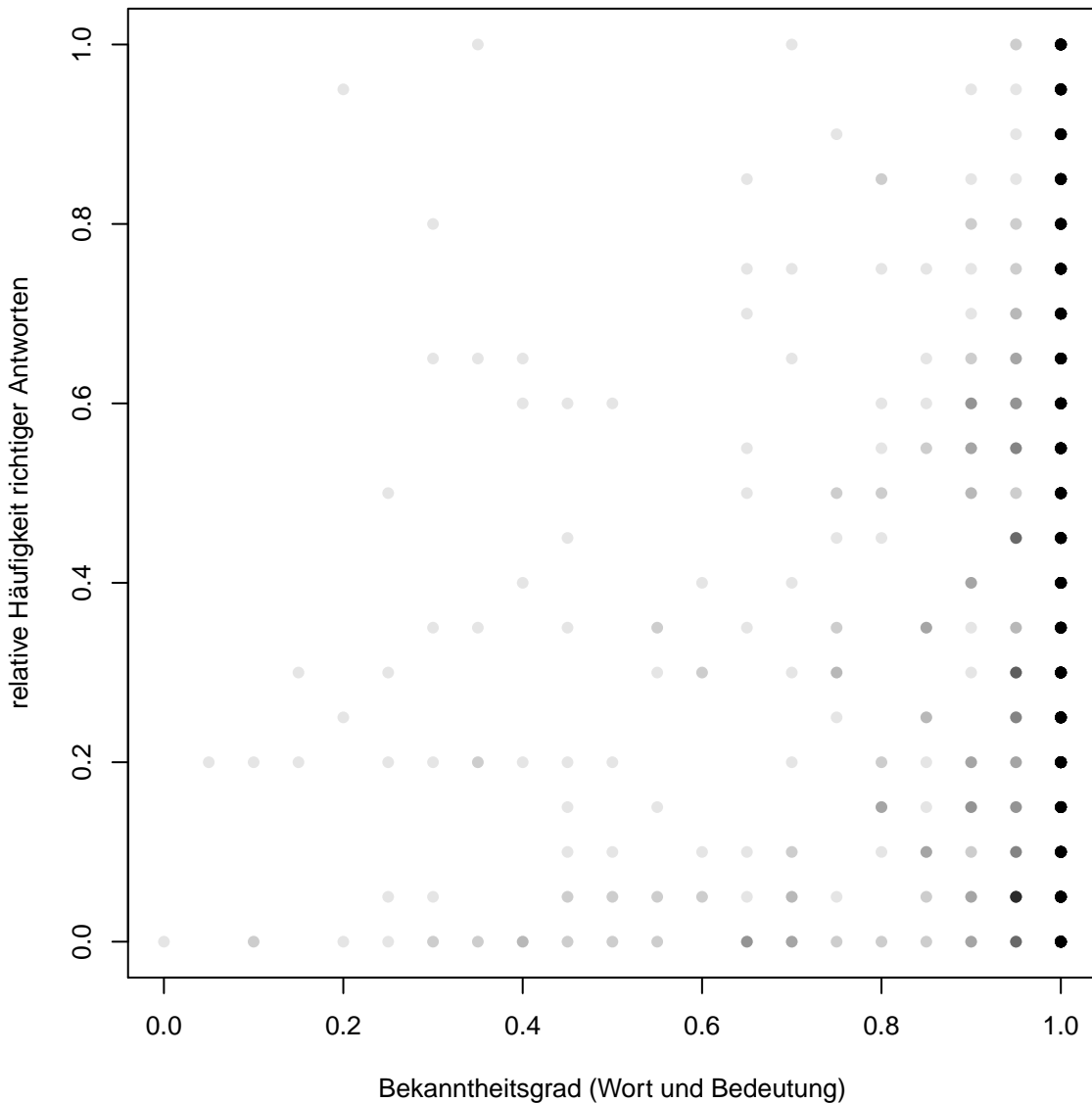


Abbildung 3.7.: Scatterplot der Abhängigkeit der Korrektheit der Antworten von der Bekanntheit der Bedeutungen

Bedeutung ein signifikant schlechteres Versuchsergebnis zur Folge. Allerdings gibt es einige Ausnahmen. Bei Probandin Nummer 6 ergaben sich die p-Werte von ca. 0,19 für die Bekanntheit des Wortes und 0,12 für die Bekanntheit der Bedeutung. Diese p-Werte geben an, dass bei dieser Probandin keine signifikanten Zusammenhänge zwischen der Bekanntheit eines Wortes und der Korrektheit der Antworten bestehen. Bei Probandin Nummer 9 ist der p-Wert für die Bekanntheit des Wortes nicht signifikant. Er liegt bei 0,38. Der p-Wert für die Bekanntheit der Bedeutung zeigt dagegen ein signifikantes Verhältnis an. Bei den männlichen Probanden gibt es

### 3. Ergebnisse

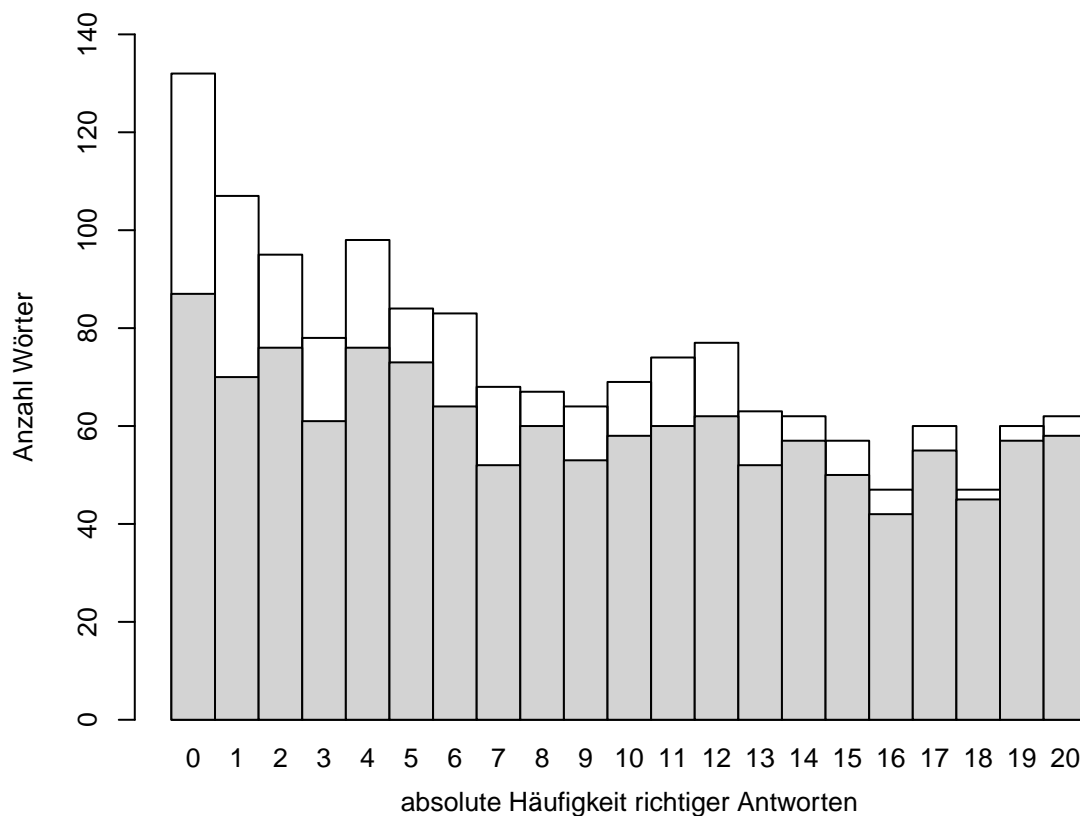


Abbildung 3.8.: Histogramm der Antworten des gesamten Sprachmaterials und der vollständig bekannten Wörter

ebenfalls Beispiele für diesen Zusammenhang. Bei Proband Nummer 16 betragen beide p-Werte 0,24. Damit besteht auch bei ihm kein signifikanter Zusammenhang zwischen der Bekanntheit der Wörter und der Korrektheit der Antworten. Auch die Probanden Nr. 15 und 18 weisen bei einem von beiden Werten keinen signifikanten Zusammenhang auf. Bei Proband Nr. 15 ist es mit einem p-Wert von 0,053 der Wert für die Bekanntheit der Bedeutungen und bei Proband Nr. 18 mit einem p-Wert von etwa 0,082 der Wert für die Bekanntheit der Wörter.

Obwohl es sich um einen Einsilberhörttest handelt, scheinen die meisten Probanden sich nicht ausschließlich auf ihr Gehör, sondern auch auf ihr Sprachgedächtnis zu verlassen. Diesen Zusammenhang hatte auch schon Goethe erkannt: „niemand hört, als was er weiß“ (von Goethe, 1833, s. Seite 156 ff.). Es sieht also so aus, als ob die Redundanz selbst bei Einsilberhörttests nicht gänzlich auszuschalten wäre.

Die oben beschriebenen Beispiele scheinen aber zu zeigen, dass es auch einige Ausnahmen von dieser Regel gibt. Es muss also davon ausgegangen werden, dass nur Wörter, die allen Prüflingen bekannt sind, die gleichen Grundvoraussetzungen für

Tabelle 3.2.: Tabelle über den Zusammenhang der Bekanntheit der abgehörten Wörter und der Korrektheit der Antworten (p-Werte des  $\chi^2$ -Tests)

Proband	p (Wort bekannt)	p (Bedeutung bekannt)
1	0,000304	0,000063
2	0,016342	0,017966
3	0,001823	0,000029
4	0,028244	0,018915
5	0,000112	0,000004
6	0,193347	0,117623
7	0,000421	0,000343
8	0,000799	0,000504
9	0,377634	0,012804
10	0,000940	0,000057
11	0,000050	0,000005
12	0,001676	0,000055
13	0,000758	0,000023
14	$< 1,0 \times 10^{-6}$	$< 1,0 \times 10^{-6}$
15	0,002248	0,053077
16	0,239428	0,239428
17	0,000013	0,000019
18	0,081979	0,038621
19	0,004216	0,000190
20	0,004695	0,001429

einen aussagekräftigen Einsilber-Hörtest bieten. Wie mit den unbekanntem Wörtern weiter verfahren wird, wird in Abschnitt 2.5 besprochen.

### Wortstruktur des abgehörten Wortes

Als nächstes wurde untersucht, ob die Struktur der abgehörten Wörter signifikante Auswirkungen auf das Wortverständnis hat. Dabei wurden, anders als bei Hahlbrock und Weiland, zunächst nur vier grundlegende Wortstrukturen voneinander unterschieden:

- Wörter, die nur aus Vokalen bestehen, wurden durch VV gekennzeichnet

### 3. Ergebnisse

- Wörter, die mit einem Konsonanten oder einem Konsonantenkomplex beginnen und mit einem Vokal oder Vokalkomplex enden, wurden durch CV gekennzeichnet
- Wörter, die mit einem Vokal oder einem Vokalkomplex beginnen und mit einem Konsonanten oder Konsonantenkomplex enden, wurden durch VC gekennzeichnet
- Wörter, die mit einem Konsonanten oder Konsonantenkomplex beginnen, mit einem Vokal oder Vokalkomplex fortfahren und wieder mit einem Konsonanten oder Konsonantenkomplex enden, wurden durch CVC gekennzeichnet

Damit ist die Struktur der Wörter in unserem Einsilberverständnis etwas anders definiert worden, als es bei Weiland 1954 der Fall gewesen war. Weiland hatte jeden Buchstaben einzeln angegeben und somit auch gleich die Länge der Wörter definiert. Das ergab weit mehr mögliche Wortstrukturen, wie zum Beispiel VV, VK, KV, VKK, KKV, KVK, VKKK, usw. (Anders als in der vorgelegten Dissertation nutzten Hahlbrock und Weiland das „K“ als Abkürzung für einzelne Konsonanten). Insgesamt ergaben sich durch diese Definition 15 mögliche Wortstrukturen (Weiland, 1954, s. Seite 20 und 21). In der vorliegenden Arbeit wird dagegen die Information über die Länge der Wörter in einem zusätzlichen Schritt untersucht. Siehe dazu Abschnitt 3.1.3 auf Seite 68.

Die in die vier verschiedenen Strukturen eingeteilten einsilbigen Hauptwörter traten in sehr unterschiedlicher Menge auf. So gab es zum Beispiel nur zwei Wörter, die in die Kategorie VV fielen. Dies waren *Au* und *Ei*. Die große Mehrzahl der aufgenommenen Wörter fielen in die Kategorie CVC. Insgesamt 1391 der 1554 Einsilber gehörten zu dieser Gruppe, während nur 80 Wörter der Kategorie VC und 81 Wörter der Kategorie CV angehörten. Um nun zu überprüfen, ob die Struktur der Wörter einen Einfluss auf die Verständlichkeit hat, wurde in Grafik 3.9 die Korrektheit der Antworten gegen die vier Wortstrukturen aufgetragen. Die Verteilungen der Ergebnisse wurden in vier Boxplots in diese Grafik eingetragen. Oberhalb der Boxen ist die jeweilige Anzahl der in die betreffende Kategorie fallenden Wörter angegeben. Die Verteilungen lassen eine erste Einschätzung zu. Der Medianwert des Boxplots der Wortstruktur CVC liegt mit ca. 45 % weit oberhalb der Medianwerte der übrigen Wortstrukturen. Die Medianwerte der Strukturen CV und VC liegen mit ca. 30,5 % und 29 % recht nah beieinander. Die Wortstruktur VV kann wegen der äußerst geringen Anzahl der betreffenden Wörter nur für eine sehr grobe Einschätzung in diese Betrachtung mit einbezogen werden. In den meisten Fällen wurden die Wörter, die



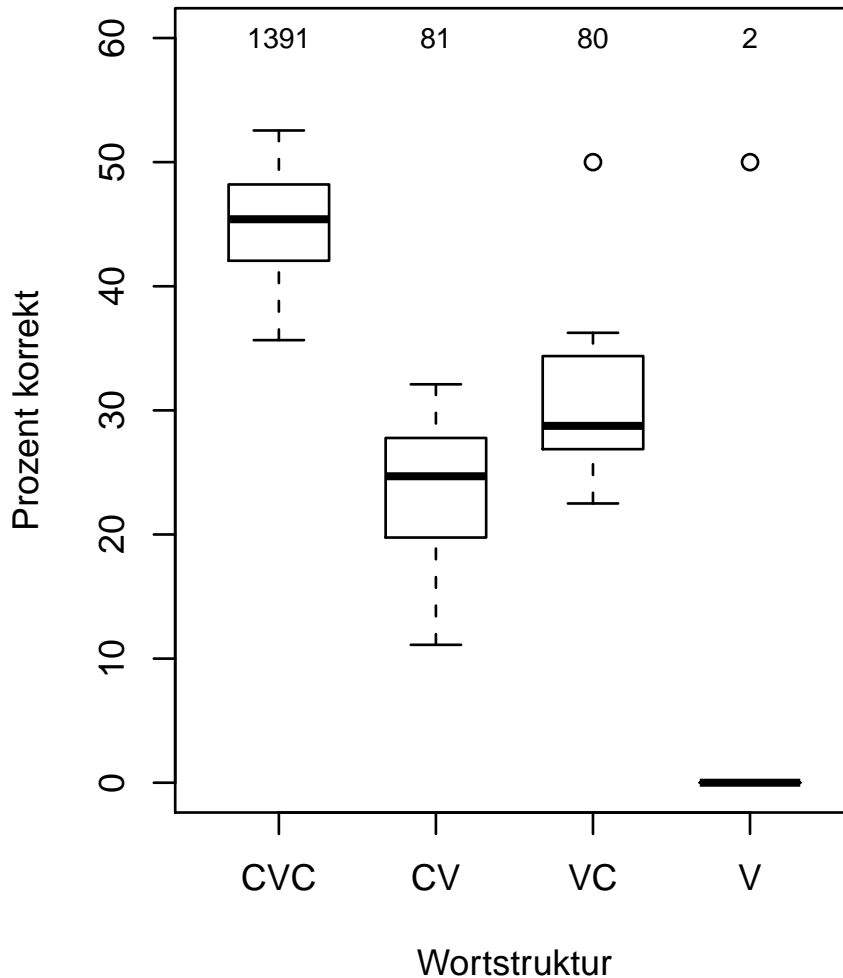


Abbildung 3.9.: Boxplot der Abhängigkeit der Korrektheit der Antworten von der Wortstruktur

dieser Struktur entsprechen, von den Probanden nicht korrekt verstanden. Nach dieser Grafik scheint es also der Fall zu sein, dass die Struktur der betreffenden Wörter sehr wohl einen Einfluss auf ihre Verständlichkeit hat. Um diesen Zusammenhang genauer zu untersuchen, wurde die prozentuale Korrektheit der Wörter der vier Wortstrukturen errechnet (Mittelwert). Dabei ergaben sich recht unterschiedliche Werte. Wie auch schon in Abbildung 3.9 zu erkennen ist, wurden Wörter der Struktur CVC am besten verstanden. Insgesamt wurde in 45,01 % der Fälle korrekt geantwortet. Die Wörter der Struktur CV wurden in 23,70 % und die Wörter der Struktur VC in 30,56 % der Fälle richtig erkannt. Wörter der Kategorie VV wurden nur in 2,50 % richtig verstanden. Dies bestätigt, dass es einen signifikanten Zusammenhang der

### 3. Ergebnisse

Struktur der Wörter und ihrer Verständlichkeit zu geben scheint. Um diese Hypothese zu belegen, wurde auf diese Daten für die probandenspezifische Auswertung der Kruskal-Wallis-Test und für die itemspezifische Auswertung der  $\chi^2$ -Test angewendet. Der Kruskal-Wallis-Test ergab ebenso wie der  $\chi^2$ -Test einen p-Wert von  $< 1,0 \times 10^{-6}$ , was den angenommenen Signifikanzzusammenhang zwischen den Wortstrukturen und der Verständlichkeit der Wörter belegt. Es muss also bei der Verteilung der Wörter in Listen auch auf die Strukturen der Wörter geachtet werden

#### **In den Einsilbern enthaltener Vokal**

Vokale werden weit besser verstanden als Konsonanten, was vor allem dadurch begründet ist, dass diese im Sprachfluss mit einer höheren Intensität gesprochen werden als Konsonanten (Wiehe, 2000, s. Seite 81 f. und Lazarus et al., 2007, s. Seite 47). Die Verständlichkeit von Vokalen in sinnlosen Silben wurde von Miller 1951 getestet. Es ergab sich ein 70-90-prozentiges Verständnis schon bei einem Pegel von 25 dB. Konsonanten wurden dagegen erst ab einem Schallpegel von 30-60 dB verstanden (Lazarus et al., 2007, s. Kap. 2 auf Seite 26). Allerdings wird nicht jeder Vokal gleich gut verstanden. Vokale bestehen, als die einfachsten Phoneme, aus einem Grundton und zahlreichen, im Mittel etwa drei Formanten, deren Frequenzen stets höher liegen als der Grundton (Keidel, 1980, s. Seite 174). Die Unterschiede in der Verständlichkeit werden durch die Formanten geprägt. Formanten sind die charakteristischen Frequenzen oder Frequenzbereiche eines Vokals und des dazugehörigen Sprechers. Mit zwei bis drei Formanten im Vokaldreieck lassen sich alle Vokale eines Lautsystems voneinander unterscheiden. Besonders wichtig für die Unterscheidung der Vokale sind dabei die Formanten F1 und F2 (Kollmeier, 2009, s. Seite 81 f.).

Der nächste Einflussfaktor der untersucht wurde, war also der in den Wörtern enthaltene Vokal. Es kann in einsilbigen Wörtern nur ein Vokal oder ein Vokalkomplex vorkommen, da ein Wort andernfalls zu einem zweisilbigen Wort würde. Um die Wörter automatisiert nach dem Vokal zu kategorisieren, wurden zunächst reguläre Ausdrücke definiert. Wörter, die dadurch nicht korrekt erfasst werden konnten, wurden durch Verwendung von Ausnahmelisten zugeordnet. Ausnahmelisten wurden also bei allen Fällen verwendet, in denen sich die Aussprache nicht eindeutig aus der Schreibweise ableiten lässt.

Die Diphthonge *ai*, *au*, *oi* und *äi* galten ebenso als Vokale beziehungsweise Vokalkomplexe, wie die Hauptvokale *a*, *e*, *i*, *o*, *u* und die Umlaute *ä*, *ö*, *ü* und konnten über reguläre Ausdrücke zugeordnet werden. Diphthonge mit unterschiedlicher Schreib-

weise aber gleicher Aussprache fielen in dieselbe Kategorie. Zum Beispiel kann der Diphthong (in Lautschrift)  $\text{ai}$  mit den Schreibweisen *ai*, *ay*, *ei* und *ey* dargestellt werden. Der Diphthong  $\text{oy}$  kann als *oi*, *eu* und *oy* geschrieben werden. Für die anderen beiden Diphthonge waren solche Ausnahmeregelungen nicht notwendig.

Einen Sonderfall bei dieser Definition stellt das *y* dar, das zwar ein Konsonant ist, aber in einigen Wörtern die Rolle eines Vokals übernimmt. Ein Beispiel für ein solches Wort ist *Typ*. Problematisch ist auch, dass das *y* nicht in jedem Wort wie ein *ü* ausgesprochen wird, sondern in einigen Wörtern wie ein *i*, zum Beispiel in *Schwyz*. In Kombination mit einem anderen Vokal kann es zu einem Diphthong werden, zum Beispiel  $\text{ai}$  in *Frey*. Hinzu kommen noch Lehnwörter wie *Boy*, deren Aussprache des *y* wiederum anders ist.

Zusätzlich wurden noch zwei Ausnahmelisten für die französischen Lehnwörter *Fond* und *Teint* definiert. Diese wurden mit dem Namen *on* beziehungsweise *en* belegt.

Über die Lehnwörter hinaus traf dies auch auf die Wörter zu, deren Vokal ein kurz gesprochenes *e* ist. Nach dem Aussprachewörterbuch des Dudens wird ein kurzes *e* in der überwiegenden Zahl der Fälle wie ein *ä* ausgesprochen (Duden, 2006a, s. Kapitel G auf Seite 64 ff.). Diese Fälle wurden ausnahmslos in die entsprechende Sonderliste übernommen.

Nachdem alle Wörter in die entsprechenden Vokalkategorien einsortiert worden waren, wurde untersucht, ob der in dem Wort enthaltene Vokal einen Einfluss auf die Verständlichkeit des Wortes hat. Zu diesem Zweck wurde für die probandenspezifische Auswertung der Kruskal-Wallis-Test auf diese Daten angewendet. Zunächst wurden dabei alle Vokale und alle Wortstrukturen berücksichtigt. Es ergab sich ein *p*-Wert von  $< 1,0 \times 10^{-6}$  und damit ein signifikanter Zusammenhang zwischen dem Vokal und der Verständlichkeit der Wörter. Zusätzlich zu dieser Betrachtungsweise sollte der Kruskal-Wallis-Test auf eine Untergruppe der oben getesteten Wörter angewendet werden. Denn die Anzahl der Wörter mit einem Hauptvokal unterschied sich erheblich von denen mit einem Diphthong. So kamen Wörter, die den Vokal *a* enthielten, zum Beispiel 378 mal vor, während Wörter, die den Diphthong *oi* enthielten, nur 20 mal vorkamen. Als zusätzlicher Kritikpunkt an der obigen Vorgehensweise könnte eine mögliche Überlagerung des Einflusses der Wortstruktur und der enthaltenen Vokale auf die Verständlichkeit der Wörter genannt werden. Und zwar könnte der signifikante Zusammenhang dadurch entstehen, dass Wörter mit allen Wortstrukturen für die Berechnung mit herangezogen worden waren. Dabei wäre

### 3. Ergebnisse

denkbar, dass die beiden Einflussfaktoren der Vokale und der Wortstrukturen sich überlagern und der signifikante Zusammenhang zwischen dem Vokal und der Verständlichkeit nur vorgetäuscht wird. Um diese etwaige Fehlerquelle auszuschließen, wurde der Kruskal-Wallis-Test noch einmal nur auf die Wörter angewendet, die in die Kategorie CVC fielen, was bei der Mehrzahl der Wörter der Fall war. Es ergab sich ein p-Wert von  $< 1,0 \times 10^{-6}$ . Dies wurde unter dem Ausschluss der sehr seltenen Diphthonge *ui*, *en* und *on* aus der Berechnung wiederholt. Der entsprechende p-Wert lag bei  $< 1,0 \times 10^{-6}$ . Dieser Wert belegt einen signifikanten Zusammenhang zwischen dem in dem Wort enthaltenen Vokal und seiner Verständlichkeit.

Offen ist durch diese Tatsache allerdings noch, welche Vokale besonders gut und welche besonders schlecht verstanden werden. Ebenfalls klärungsbedürftig ist, ob die Signifikanz durch einen einzelnen Extremwert hervorgerufen wird oder ob alle Werte betroffen sind. Um dies zu überprüfen, wurde der Kruskal-Wallis-Test noch auf verschiedene Untermengen der Einsilber der Kategorie CVC angewendet. Zunächst wurden nur diejenigen Wörter, die einen der fünf Hauptvokale oder einen Umlaut enthielten, in die Berechnung einbezogen. Der errechnete p-Wert ergab  $< 1,0 \times 10^{-6}$ . Anschließend wurden nur Einsilber, die einen Hauptvokal enthielten, untersucht. Dabei ergab sich ein p-Wert von  $< 1,0 \times 10^{-6}$ . Des Weiteren erfolgte eine Berechnung derjenigen Wörter, die einen Umlaut enthielten. Der errechnete p-Wert lag bei  $< 1,0 \times 10^{-6}$ . Daraufhin wurde auf diejenigen Wörter, die einen Diphthong enthielten, der Kruskal-Wallis-Test angewendet. Auch hier lag der p-Wert bei  $< 1,0 \times 10^{-6}$ . Auch bei der Untersuchung dieses Einflussfaktors wurde wiederum der  $\chi^2$ -Test auf diese Daten angewendet. Zum Einen kam er unter Einbeziehung aller Vokale und zum Anderen unter Nichtbeachtung der seltenen Diphthonge *oi*, *en* und *on* zum Einsatz. In beiden Fällen wurde ein p-Wert von  $< 1,0 \times 10^{-6}$  errechnet.

Bei all diesen Berechnungen ergaben sich also p-Werte, die einen signifikanten Zusammenhang zwischen den einbezogenen Vokalen, beziehungsweise Vokalkomplexen, und der Verständlichkeit der Wörter belegen. Um nun deutlich zu machen, bei welchen Vokalen es sich um gut verständliche und bei welchen es sich um schlecht verständliche Vokale handelt, wurde ein Boxplot mit diesen Daten erstellt. Auf der Abszisse sind die elf möglichen Kategorien, also die Vokale, Umlaute und Diphthonge, eingezeichnet und auf der Ordinate ist die Verständlichkeit der Wörter in Prozent aufgetragen. Das sich ergebende Bild ist anders als erwartet. Wie oben bereits beschrieben, ist die Verständlichkeit der Vokale ganz wesentlich von den Formanten abhängig. Sie lassen erkennen, an welcher Position welche Teiltöne verstärkt sind und

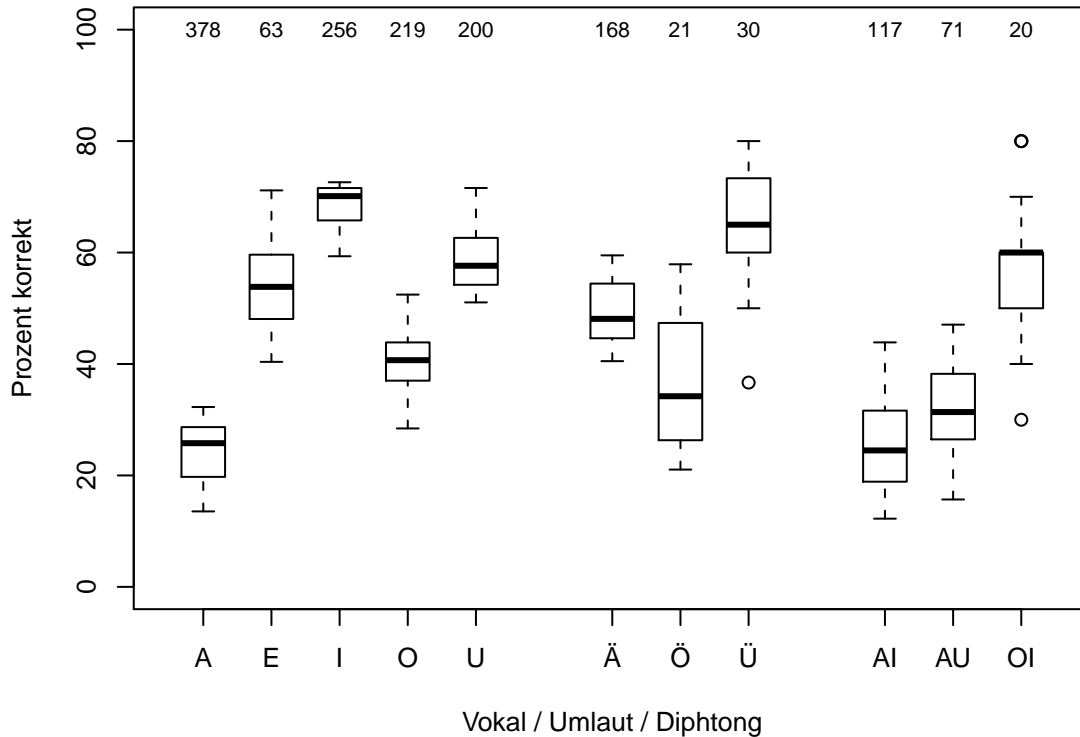


Abbildung 3.10.: Boxplot der Abhängigkeit der Verständlichkeit der Wörter von den in ihnen enthaltenen Vokalen

charakterisieren damit den gesprochenen Vokal. Die beiden tiefrequenten Formanten F1 und F2 sind wichtig für die Verständlichkeit der Vokale (Willinger, 2005, s. Kap. 6.1.1 auf Seite 21 f.), während Formant F3 und F4 eher den Charakter einer Stimme ausmachen. Der Vokal *u* hat zum Beispiel einen Formantenbereich von 200-400 Hz (Variation durch verschiedene Sprecher), das *o* hat einen typischen Formantenbereich von 400-600 Hz. Das *a* einen Bereich von 800-1200 Hz, das *e* wieder nur von 400-600 Hz und das *i* nur noch von 200-400 Hz. Dabei sind bei den letzten beiden Vokalen die Werte sehr stark davon abhängig, ob der Vokal von einem männlichen oder von einem weiblichen Sprecher ausgesprochen wird. Bei weiblichen Sprechern ergeben sich sehr viel höhere Werte. Die wichtigsten zur Spracherkennung notwendigen Frequenzen liegen zwischen 300 und 3000 Hz (von Eiff et al., 1985, s. Seite 95 ff.). Auf einer Formantkarte (siehe z. B. Vary et al., 1998, S. 47) wird erkennbar, dass das *a* weit abseits der anderen Vokale liegt, was die Verwechslungsgefahr reduziert. Man könnte daraufhin schlussfolgern, dass das *a* im Gegensatz zu dem *u* besonders gut verstanden wird. Die Graphik zu der Verständlichkeit der Einsilber belegt aber genau das Gegenteil. Wörter, die den Vokal *a* enthalten, wurden besonders

### 3. Ergebnisse

schlecht verstanden, während Einsilber mit dem Vokal *u* besonders gut verstanden wurden. Die Medianwerte liegen bei 27 und bei 59% der Wortverständlichkeit. Mit am besten wurden die Wörter verstanden, die ein *i* enthalten. Hier liegt der Medianwert sogar bei 70%. Auch Wörter mit einem *e* wurden sehr gut verstanden. Das *o* liegt in einem mittleren Bereich. Bei den Umlauten bietet sich ein ähnliches Bild. Hier weist das *ü* mit einem Medianwert von 67% die beste Verständlichkeit auf. Die Verständlichkeit von *ö* und dem *ü* liegen etwa 20 bis 30 Prozentpunkte darunter. Die Verständlichkeit der Diphthonge ist insgesamt etwas schlechter. Aber auch hier ist überraschend, dass die Verständlichkeit des ungebräuchlichsten Diphthongs *oi* am besten war. Über den Boxen ist die jeweilige Anzahl der Wörter, die in diese Kategorie fallen, in die Grafik eingetragen. Auf die möglichen Gründe für diese überraschenden Ergebnisse wird in Abschnitt 4.1 auf Seite 85 näher eingegangen.

#### **Länge des enthaltenen Vokals**

Um einen möglichen Einfluss der Länge des in dem Wort enthaltenen Vokals und dessen Verständlichkeit zu untersuchen, wurde zunächst ähnlich wie bei der Untersuchung des Zusammenhangs zwischen dem enthaltenen Vokal und der Verständlichkeit der betreffenden Wörter vorgegangen. Die Information über die Länge wurde durch Definition verschiedener regulärer Ausdrücke festgehalten. Die regulären Ausdrücke für lang gesprochene Vokale waren zum Beispiel Doppelvokale, Vokale auf die ein *h* folgte und ein *i* in Kombination mit einem *e*. Reguläre Ausdrücke für kurz gesprochene Vokale waren zum Beispiel Vokale auf die ein *ck* folgte oder ein Vokal mit einem darauf folgenden Konsonantenkomplex, der mit einem *r* beginnt. Insgesamt wurde eine Vielzahl möglicher Vokal-Konsonantenkombinationen in diese Listen aufgenommen, um die Zuordnung der Wörter nach der Länge der gesprochenen Vokale möglichst automatisiert durchführen zu können. Das Ergebnis dieser Zuordnung wurde mehrmals überprüft. Wie bei der Zuordnung der Vokale wurden auch hier Ausnahmelisten erstellt, die vor allem bei der Zuordnung der Lehnwörter gebraucht wurden.

Insgesamt beinhalteten 1123 der 1554 Einsilber kurz gesprochene Vokale, während bei den übrigen 431 Wörtern die Vokale lang ausgesprochen wurden. Für die deskriptive Auswertung dieser Daten wurde nun ein Boxplot der richtigen Antworten je nach der Dauer des gesprochenen Vokals erstellt. Schon auf dieser Abbildung liegen die Verteilungen der Werte recht nah beieinander. Damit scheint die Länge des Vokals die Verständlichkeit der Wörter nicht sehr stark zu beeinflussen. Die Median-

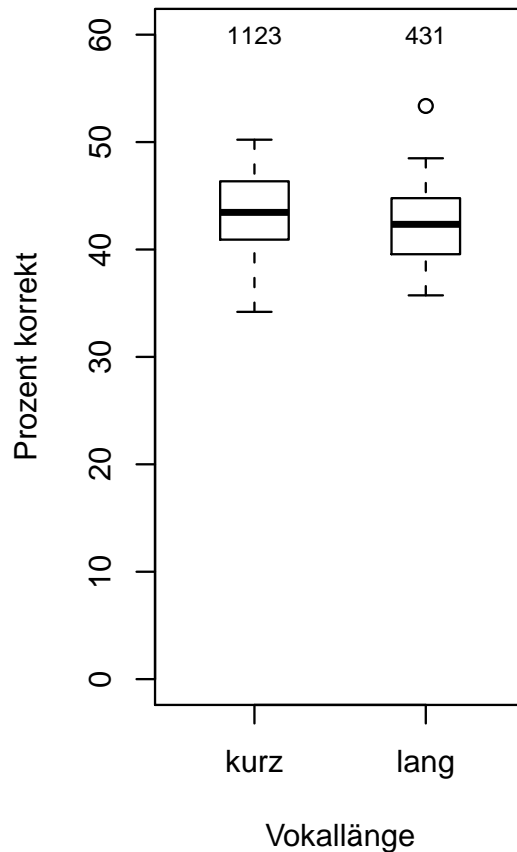


Abbildung 3.11.: Boxplot der Abhängigkeit der Verständlichkeit von der Länge des im Wort enthaltenen Vokals

werte liegen bei ca. 47 und 43 %, was diesen Eindruck noch verstärkt. Prozentual betrachtet, wurden 43,32 % der kurzen und 42,55 % der Wörter mit langen Vokalen korrekt verstanden. Um die Annahme des fehlenden Signifikanzzusammenhanges zu untermauern, wurde der Wilcoxon-Test auf diese Daten angewendet. Es ergab sich ein p-Wert von 0,430433. Dies belegt, dass die Länge des Vokals die Verständlichkeit der Wörter nicht signifikant beeinflusst. Um die Daten nicht nur probandenspezifisch, sondern auch itemspezifisch auszuwerten, wurde zusätzlich noch der  $\chi^2$ -Test auf diese Daten angewendet. Es ergab sich ein p-Wert von 0,227777. Damit ergibt sich auch bei dieser Auswertungsart kein signifikanter Zusammenhang zwischen der Länge des enthaltenen Vokals eines Einsilbers und seiner Verständlichkeit. Schon in früheren Versuchen ist die Beziehung der Länge der Wörter zu ihrer Verständlichkeit untersucht worden. Allerdings lagen dabei die Konsonanten im Blickfeld der Untersuchung. Dabei hatte sich ergeben, dass „... die Versuche zur Verbesserung der Konsonantenverständlichkeit bei Hörgestörten mit Hilfe von Lautdehnungen kei-

### 3. Ergebnisse

ne eindeutigen Ergebnisse liefern. In einigen Versuchen zeigte sich sogar, dass eine Kombination aus Lautdehnung und Verstärkung des Sprachpegels zu einer höheren Konsonantenverwechslung führt, als bei der Verstärkung des Sprachpegels allein“ (Sendlmeier, 1992, s. Seite 52). Es zeigt sich also, dass bei der Verteilung der Wörter in Listen auf die Länge des gesprochenen Vokals keine Rücksicht genommen werden muss.

#### **Lehnwörter**

Ein Lehnwort ist ein aus einer anderen Sprache stammendes Wort, das in Flexion, Gebrauch und Schreibweise gut an die Nehmersprache, also in diesem Fall die deutsche Sprache, angepasst ist. Fremdwörter sind ebenfalls Wörter, die aus einer anderen Sprache übernommen wurden, aber im Gegensatz zu den Lehnwörtern kaum an die Empfängersprache angepasst sind (Saglam, 2003, s. Seite 87). Allerdings ist eine eindeutige Unterscheidung zwischen Fremdwörtern und Lehnwörtern in vielen Fällen schwierig und nicht eindeutig, weshalb hier durchgehend der Ausdruck Lehnwort verwendet wird.

Zu überprüfen war, ob die in diesem Test enthaltenen Lehnwörter einen Unterschied in ihrer Verständlichkeit zu den deutschstämmigen Wörtern aufweisen. Hierfür mussten alle 1554 Wörter nach Lehnwörtern durchsucht werden. Diese Aufgabe gestaltete sich schwieriger als zunächst erwartet, da viele der betreffenden Wörter schon vor sehr langer Zeit in die deutsche Sprache integriert wurden und inzwischen so gebräuchlich sind, dass sie auf Anhieb kaum noch als Lehnwörter erkannt werden. Aus diesem Grund mussten alle Wörter mit Hilfe des Herkunftswörterbuches des Dudens auf ihren Wortstamm untersucht werden (Duden, 2006b). Wir entschieden uns für eine strenge Wortauswahl. Außer den deutschsprachigen Wörtern wurden nur Wörter, die aus dem Lateinischen stammen, nicht den Lehnwörtern zugeordnet. Außerdem ließen wir einige Ausnahmen bei Wörtern zu, die schon seit Anfang des 18. Jahrhunderts in der deutschen Sprache gebräuchlich sind, wie zum Beispiel *Keks* und *Koks*. Des Weiteren gab es eine Reihe von Wörtern, die sowohl im Deutschen als auch im Englischen vorkommen, wie zum Beispiel *link*, was im Deutschen soviel wie „gemein“ heißt und im Englischen „Verbindung“ bedeutet. Das Wort *Troll* hat ebenfalls zwei Wortstämme. Außer dem besser bekannten skandinavischen Wortstamm ist es auch aus dem Mittelhochdeutschen bekannt. All die Wörter, die sowohl einen deutschen, als auch einen anderen Sprachstamm aufweisen, wurden in die Liste der Lehnwörter nicht mit aufgenommen. Von allen im Sprachmaterial



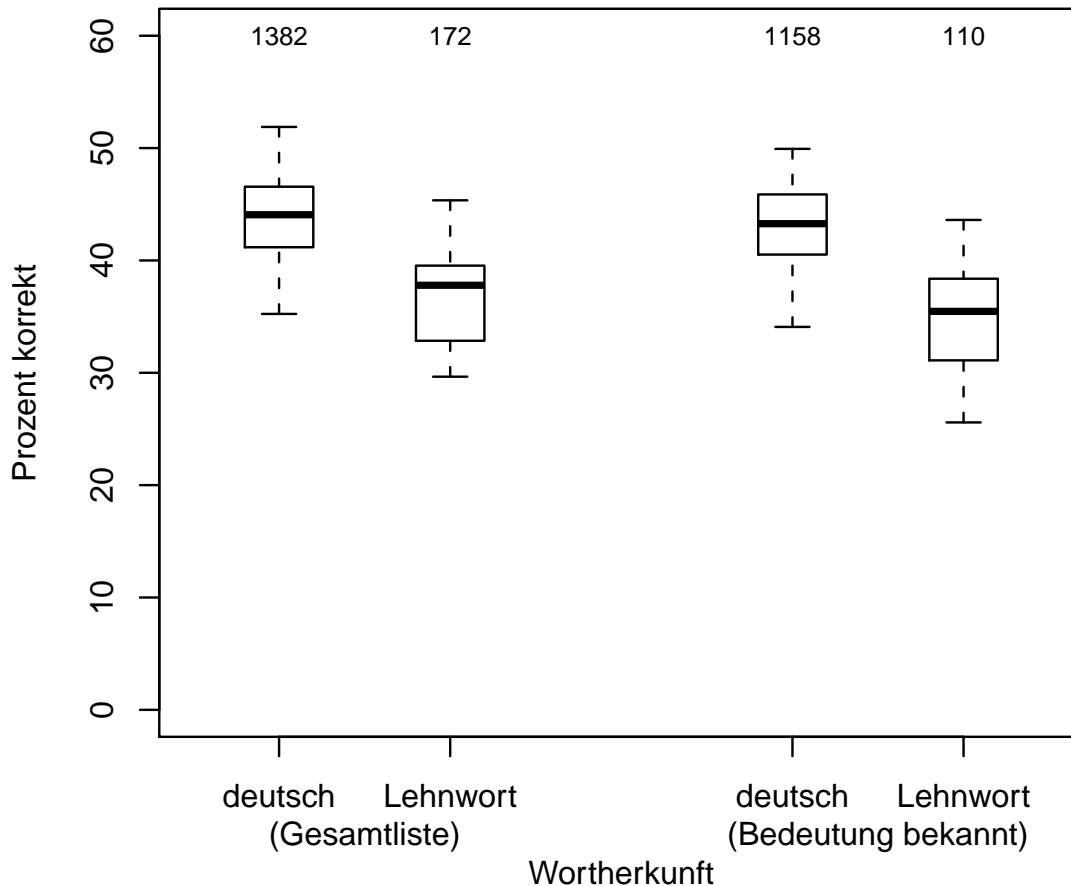


Abbildung 3.12.: Boxplot der Abhängigkeit der Verständlichkeit von der Wortherkunft

M-2007 enthaltenen Wörtern hatten 1382 einen deutschen Wortstamm, während die übrigen 172 Einsilber den Wortstamm einer anderen Sprache aufwiesen. Somit enthielt die Liste der Lehnwörter rund 10 % aller Einsilber.

Sie wurde für die weitere Verarbeitung in das Statistikprogramm „R“ eingelesen und als zusätzliche Einflussfaktorspalte an die Tabellen „e“ und „et“ angehängt. Nun konnte auf diese Daten ein  $\chi^2$ -Test angewendet werden. Es ergab sich ein p-Wert von  $< 1,0 \times 10^{-6}$  und damit ein hochsignifikanter Zusammenhang zwischen der Sprachzugehörigkeit des Wortstammes und seiner Verständlichkeit. Um die Daten auch probandenspezifisch auszuwerten, wurde zusätzlich der Wilcoxon-Test angewendet. Auch bei dieser Auswertungsart ergab sich mit einem p-Wert von 0,000010 ein signifikanter Zusammenhang. Insgesamt wurden 43,87 % der deutschstämmigen Wörter, aber nur 36,98 % der Lehnwörter korrekt verstanden. Fraglich war aber, ob die schlechtere Verständlichkeit der Lehnwörter auf ihre größere Unbekanntheit zurückzuführen war. Aus diesem Grund wurde der  $\chi^2$ -Test noch einmal auf dieje-

### 3. Ergebnisse

nigen Lehnwörter angewendet, die alle Probanden als bekannt angegeben hatten. Der errechnete p-Wert lag bei  $< 1,0 \times 10^{-6}$ . Auch der Wilcoxon-Test wurde auf diese Wörter erneut angewendet. Der errechnete p-Wert lag bei 0,000004. Abbildung 3.12 zeigt einen Boxplot, der die prozentuale Korrektheit der deutschstämmigen Wörter und die der Lehnwörter gegenüberstellt. Es ist deutlich zu erkennen, dass die Lehnwörter deutlich schlechter verstanden wurden. Die Medianwerte liegen bei 37 bzw. 35 Prozent für die Lehnwörter der Gesamtliste und für die Auswahl der bekannten Wörter und bei 44 bzw. 43 Prozent für die deutschsprachigen Prüfwörter beider Kategorien. Man kann also den Effekt der schlechteren Verständlichkeit der Lehnwörter nicht ihrer geringeren Bekanntheit zuschreiben. Daher muss auch auf ein gleichmäßiges Vorkommen der Lehnwörter bei der Verteilung der Einsilber in Listen Rücksicht genommen werden.

#### 3.1.4. Vergleich der Versuchsergebnisse dieser Studie mit den Versuchsergebnissen der Pilotstudie

Ein Teil der Einsilber des neuen Sprachmaterials war auch schon von Mahfoud 2007 getestet worden. Dies war im Rahmen einer Pilotstudie mit anderer Zielsetzung als in der vorliegenden Arbeit geschehen. Mahfoud hatte die Verständlichkeit des männlichen und des weiblichen Sprachmaterials mit den Aufnahmen des Freiburger Einsilbertests von 1968 und 1976 bei verschiedenen SNRs verglichen. Dabei wurde nur die Schnittmenge der in allen vier Aufnahmen vorkommenden Wörter betrachtet. Beide Aufnahmen des Freiburger Sprachverständnistests bestehen aus 400 einsilbigen Hauptwörtern, von denen allerdings nur 380 übereinstimmen. Zusätzlich kommt in beiden dieser Aufsprachen das Wort *Rad* und das Wort *Rat* vor, die zwar unterschiedliche Bedeutungen, aber die gleiche Aussprache aufweisen und somit in einem Hörtest nicht unterschieden werden können. Außerdem kommt in beiden Versionen des Freiburger Sprachverständnistests ein Wort doppelt vor. In der Aufnahme von 1968 ist dies das Wort *Schrift* und in der Aufnahme von 1976 der Einsilber *Stück*. Dass heißt, dass die Schnittmenge der Wörter der drei Sprachtests insgesamt 378 Wörter umfasst, die in der Pilotstudie von Mahfoud an 4 Probanden getestet worden waren. Dabei hörte nicht jeder Proband 1512 Wörter ab, sondern nur 360, zuzüglich einer Eingewöhnungsphase von 18 Wörtern. Dies kam zustande, indem jeweils 24 Wörter einer Aufsprache mit einem SNR von 0 dB, -4 dB, -8 dB, -12 dB oder -16 dB belegt wurden. So kamen Blöcke von 120 (5×24) Wörtern zusammen, von denen ins-

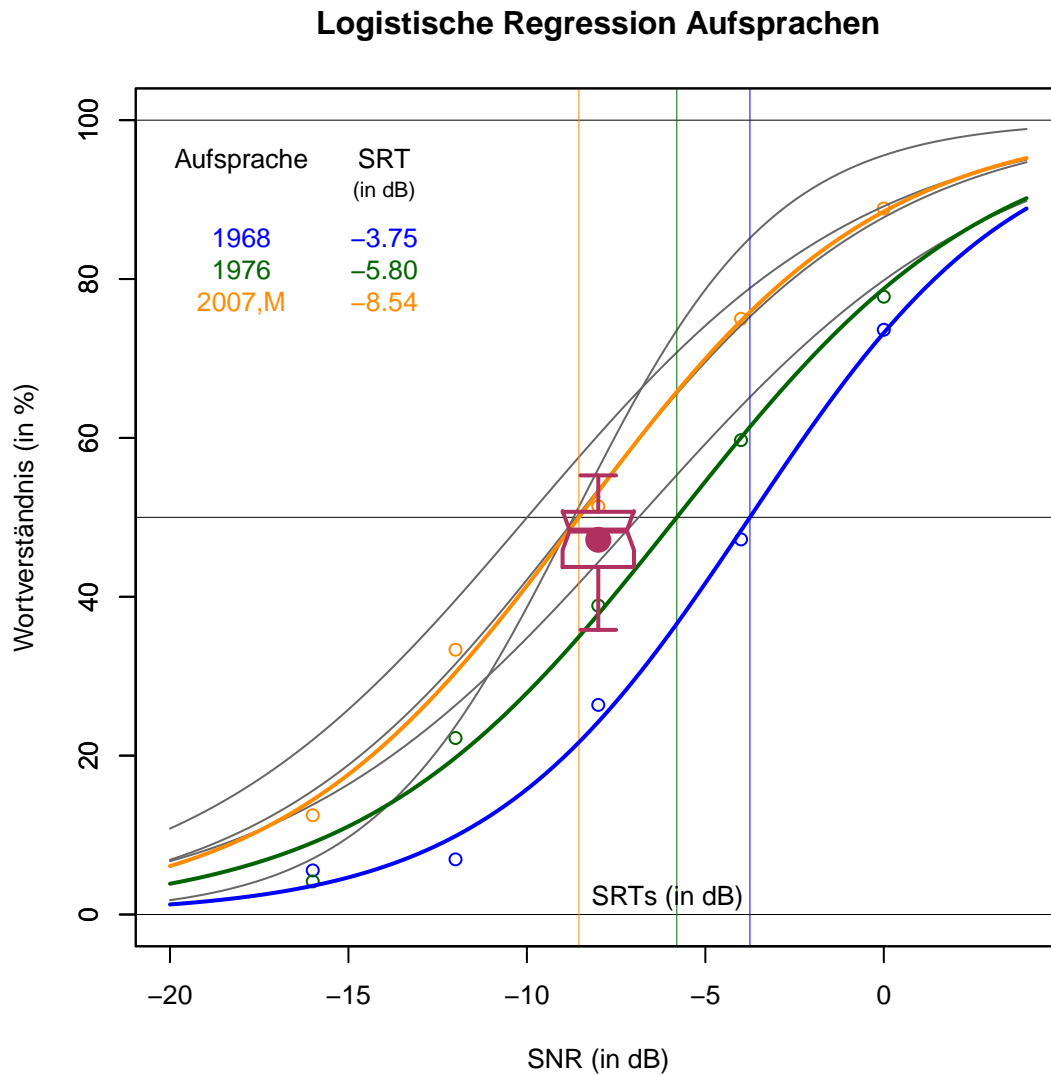


Abbildung 3.13.: Vergleich der Versuchsergebnisse der Pilotstudie mit der Evaluation

gesamt drei an jedem Probanden getestet wurden (Mahfoud, 2009, s. Seite 31 ff.). Da in der vorliegenden Arbeit alle Wörter bei einem SNR von -8 dB getestet worden sind, konnte der Vergleich der Ergebnisse der Pilotstudie und der vorliegenden Evaluation des Sprachmaterials M-2007 nur in dem Bereich dieses Rauschabstandes stattfinden. Auch wurde nur die Schnittmenge der 360 Wörter für diesen Vergleich herangezogen. Hinzu kommt, dass diese Arbeit nur die Evaluation der männlichen Stimme umfasst. Die Ergebnisse der weiblichen Sprecherin der Pilotstudie wurden daher aus dem Vergleich entfernt. Abbildung 3.13 zeigt eine Graphik des Vergleiches

### 3. Ergebnisse

der Versuchsergebnisse. Auf der Abszisse ist der Rauschabstand von -20 bis 0 dB in 5 dB-Schritten und auf der Ordinate das prozentuale Wortverständnis eingetragen. Die Versuchsergebnisse der Pilotstudie sind als logistische Regressionskurven eingezeichnet. Die blaue Linie gibt die Verständlichkeit des Freiburger Einsilberverständnistests von 1968 und die grüne Linie die von 1976 an. Schon bei diesen beiden Aufsprachen fällt der sehr unterschiedliche SRT auf. Bei der 68er Aufsprache liegt er bei nur -3,75 dB und bei der 76er Aufsprache schon bei -5,8 dB. Das neue Sprachmaterial M-2007 ist noch wesentlich besser zu verstehen. Der SRT liegt bei -8,54 dB. Die gelbe Linie stellt die gemittelten Versuchsergebnisse der vier Probanden in der Pilotstudie dar. Die dünnen grauen Linien zeigen die Einzelergebnisse der Probanden. Auch diese fallen sowohl in dem SRT als auch in der Steigung der logistischen Regression recht unterschiedlich aus. Die Ergebnisse aller 20 Probanden über die 360 Wörter aus der vorliegenden Studie sind in dem hellroten Boxplot in die Graphik eingetragen. Die Box gibt die Verteilung der Werte an. Zusätzlich zum Medianwert ist auch der Mittelwert durch den Punkt in der Box eingetragen. Es lässt sich erkennen, dass die Probanden, die an dieser Studie teilgenommen haben, durchschnittlich etwas schlechter abgeschnitten haben als die Probanden der Pilotstudie. Das mittlere Wortverständnis liegt bei 44,65 %, während die Probanden der Pilotstudie ein mittleres Wortverständnis von ca. 51 % erreicht haben. Auch bei der Betrachtung der Streubreiten bestätigt sich dieses Ergebnis. Die Streubreite bei den Probanden der Pilotstudie liegt in etwa zwischen 41 und 59 %, während die Streubreite der Ergebnisse der Probanden dieser Studie zwischen 35,8 und 55,3 % liegt.

In Abbildung 3.14 wird ein anderer Zusammenhang deutlich. Hier wurde die absolute Häufigkeit richtiger Antworten gegen die Anzahl der Wörter aufgetragen. Allerdings ist nur die Schnittmenge der 360 Wörter in diese Betrachtung mit einbezogen worden. Es fällt sofort auf, dass die Verteilung der Antworten sich völlig von den Ergebnissen des gesamten Sprachmaterials unterscheidet. Anders als in Abbildung 3.8 ist keine Häufung nicht oder kaum verstandener Wörter zu beobachten. Dies könnte durch die Auswahl der getesteten Einsilber begründet sein. Die Wörter, die für den Freiburger Sprachverständnistest ausgewählt worden waren, galten als sehr gebräuchliche und einfache Wörter. Hahlbrock hatte diese Wörter ausgesucht, damit das Ergebnis des Hörtests möglichst wenig durch einen möglicherweise geringen Wortschatz der Probanden beeinflusst werden würde. Um zu untersuchen, ob diese Wörter auch heute noch als besonders gebräuchlich betrachtet werden, wurde der Prozentsatz der unbekanntem Wörter, beziehungsweise der unbekanntem

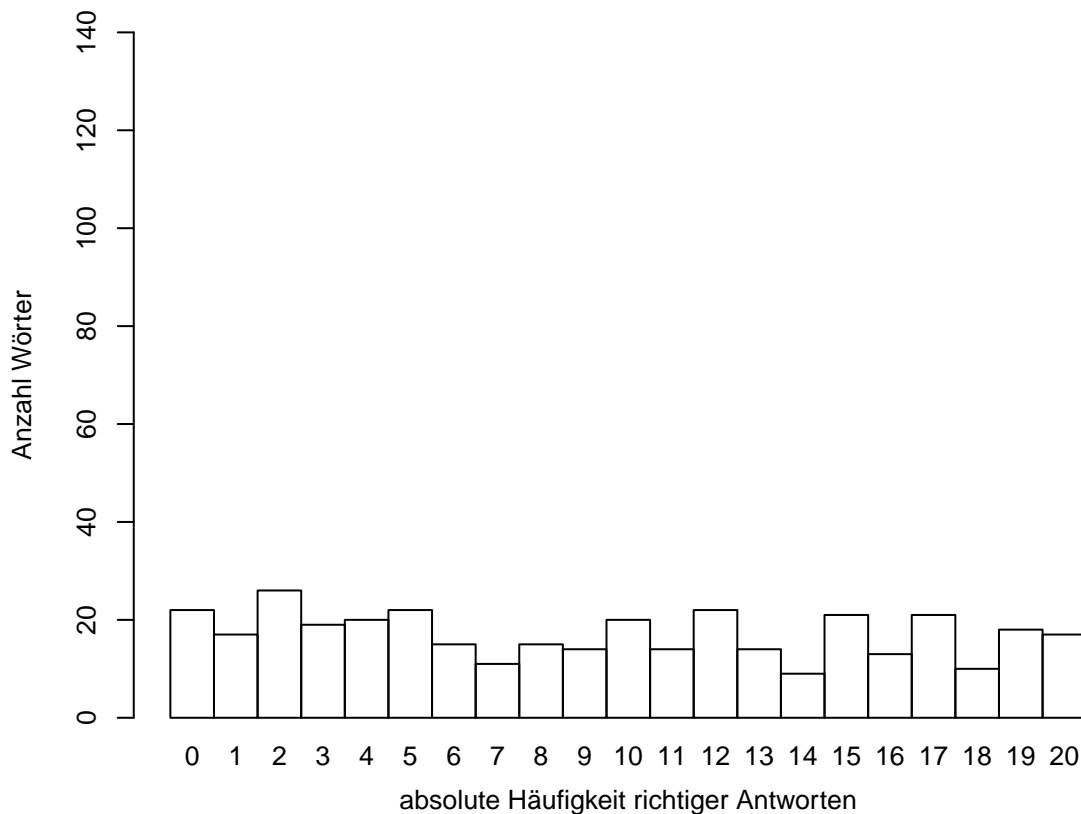


Abbildung 3.14.: Histogramm der richtigen Antworten aus der Schnittmenge der 360 Wörter

Bedeutungen der Schnittmenge der 360 Wörter, dem aller Wörter und dem der Gesamtzahl der Wörter abzüglich der 360 Wörter errechnet. Bei der Gruppe der 360 Wörter waren 0,153 % gänzlich unbekannt und 0,222 % zumindest in der Bedeutung unbekannt. Im Gegensatz dazu lagen die entsprechenden Prozentwerte der Gesamtmenge der Einsilber (abzüglich der 360 Wörter) bei 3,794 % für gänzlich unbekannte Wörter und bei 5,741 % für in der Bedeutung unbekannte Wörter. Die Prozentwerte für unbekannte Wörter des gesamten Sprachmaterials lagen wiederum bei 2,950 % und bei 4,463 % für unbekannte Bedeutung der Wörter. Anschließend wurde mit Hilfe des  $\chi^2$ -Tests untersucht, ob die Bekanntheit der Wörter, beziehungsweise die Bekanntheit der Bedeutungen der in der Schnittmenge der 360 enthaltenen Wörter sich von den im übrigen Sprachmaterial enthaltenen Wörtern signifikant unterscheidet. Die p-Werte für die Wörter und die Bedeutungen lagen beide bei  $< 1,0 \times 10^{-6}$  und bestätigen damit den signifikanten Zusammenhang. Da die Ergebnisse dieser Evaluation gezeigt haben, dass eine Unbekanntheit der Wörter mit einer geringeren Verständlichkeit korreliert, kann man schlussfolgern, dass die guten Er-

### 3. Ergebnisse

gebnisse der Schnittmenge durch die höhere Gebräuchlichkeit der getesteten Wörter begründet ist. Siehe hierzu auch Abschnitt 3.1.3 auf Seite 57. Auf die möglichen Ursachen für die unterschiedlichen Ergebnisse der Probanden der Pilotstudie und der Probanden dieser Evaluation wird in Abschnitt 4.2.3 auf Seite 93 eingegangen. Durch die Einschränkung auf gebräuchliche Wörter und die Auswahl der Probanden sind die Ergebnisse der Pilotstudie nicht für die Verteilung der Wörter in Listen anzuwenden.

#### 3.1.5. Anzahl der verfügbaren Wörter

Wie in Abschnitt 2.5 bereits beschrieben, wurde bei der Verteilung der Einsilber in Listen eine große Gruppenanzahl angestrebt. Um dies verwirklichen zu können, war ein möglichst großes zur Verfügung stehendes Sprachmaterial notwendig. Allerdings sollte aus Gründen der Homogenität des entstehenden Sprachtests auf verschiedene Wörter verzichtet werden. Bei der Zusammenstellung der Listen sollte auf einen gleichen Korrektheitsgrad sowie auf einen ausgeglichenen Bekanntheitsgrad der Wörter geachtet werden. Abbildung 3.15 zeigt in einem Linienplot die Anzahl der zur Verfügung stehenden Einsilber, wenn die Auswahl ausschließlich anhand des Korrektheitsgrades getroffen würde. Wie in dieser Grafik deutlich wird, stehen 1554 Wörter für die Verteilung zur Verfügung, wenn die Wörter keines Korrektheitsgrades bei der Verteilung ausgeschlossen werden. Wird die Auswahl allerdings auf die ersten zwanzig Korrektheitsgrade beschränkt, wie es bei der fixen Verteilung vorgesehen war, so bleiben noch 1360 Einsilber. Somit kommt es mit Reduktion der akzeptierten Korrektheitsgrade zur linearen Abnahme der Wortanzahl. Wenn für die Zusammenstellung der Listen sogar nur Wörter verwendet werden sollten, die von der Hälfte der Probanden korrekt verstanden wurden, dann stünden nur noch 69 Einsilber für die Verteilung zur Verfügung.

Außer auf den Korrektheitsgrad musste aber vor allem auf die unbekanntes Wörter Rücksicht genommen werden. Die folgenden beiden Abbildungen zeigen die absoluten Häufigkeiten der Wörter verschiedener Bekanntheitsgrade. Abbildung 3.16 bezieht sich dabei auf bekannte Wörter, während Abbildung 3.17 sich auf darüber hinaus in ihrer Bedeutung bekannte Wörter bezieht. Erwartungsgemäß ist der Großteil der Wörter vollständig bekannt. Aber auch die Gruppe der Wörter, die nur eine der Versuchspersonen als nicht bekannt angegeben hat, umfasst noch fast 100 Einsilber. Die Anzahl der für mehrere Probanden unbekanntes Wörter nimmt immer

weiter ab. Schlussendlich gibt es kein Wort, das allen Probanden unbekannt war.

Abbildung 3.18 zeigt ein Histogramm, das die Anzahl der Wörter der absoluten Anzahl richtiger Antworten gegenüberstellt. Dabei umfassen die weißen Säulen das gesamte zur Verfügung stehende Sprachmaterial, während die hellgrauen Balken nur die Auswahl der vollständig bekannten Wörter darstellt. Die dunkelgrauen Säulen zeigen die zur Verfügung stehende Wortanzahl bei Verwendung der 360 Wörter, die auch in den älteren Sprachtests verwendet worden waren. Die Anzahl der zur Verfügung stehenden Wörter wird durch die Reduktion auf die vollständig bekannten Wörter stark eingeschränkt, allerdings hauptsächlich im Bereich der gar nicht oder nur kaum verstandenen Wörter, während bei den gut verstandenen Einsilbern nur eine äußerst geringe Reduktion der Wortanzahl stattfindet. Es wird aber deutlich, dass im Vergleich der beiden Aufnahmen des Freiburger Einsilbertests eine wesentlich größere Auswahl an Wörtern zur Verfügung steht. Einen etwas genaueren Überblick über diesen Sachverhalt bietet Abbildung 3.19. Der Linienplot zeichnet die Anzahl der zur Verfügung stehenden Wörter gegen die Bereiche absoluter Häufigkeiten be-

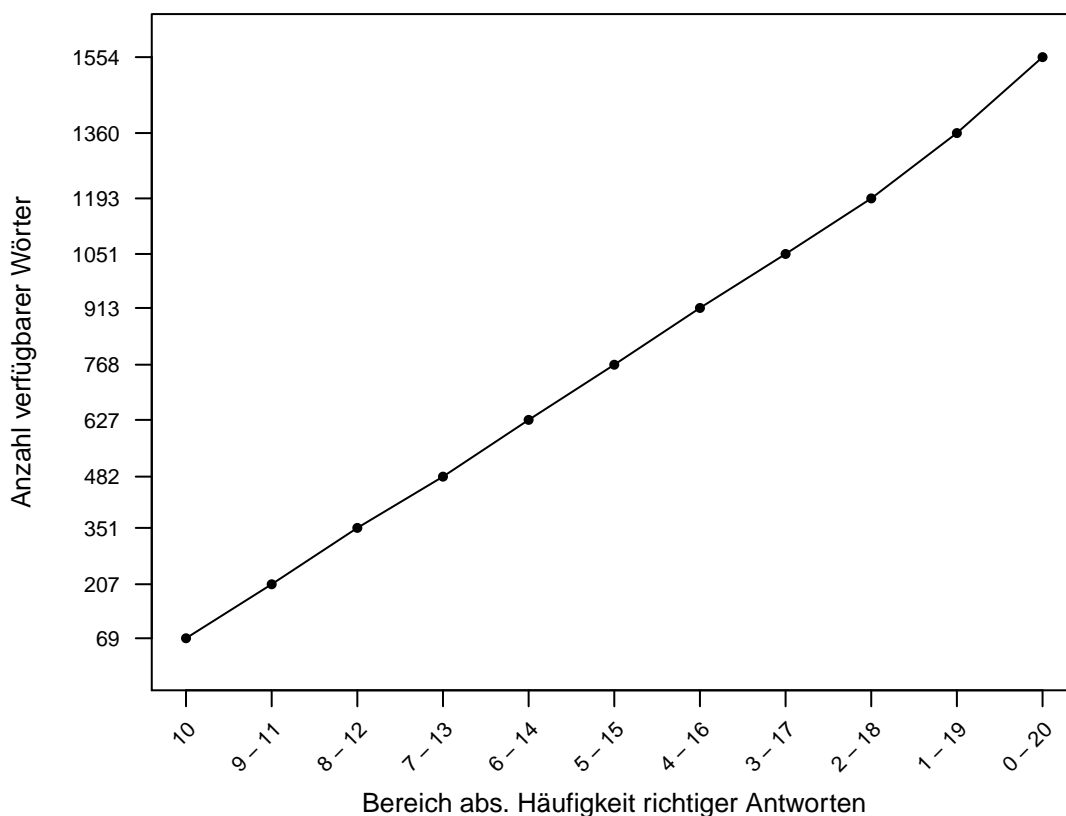


Abbildung 3.15.: Linienplot der verfügbaren Wörter je nach den korrekten Antworten

### 3. Ergebnisse

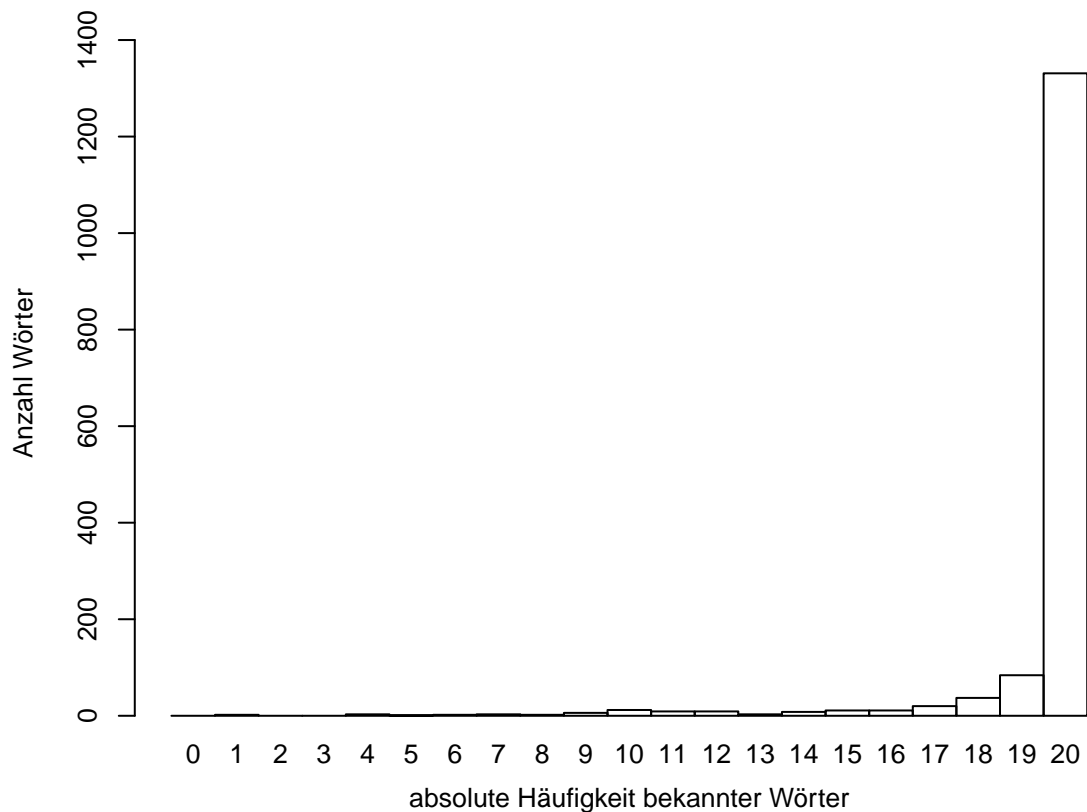


Abbildung 3.16.: Histogramm der absoluten Häufigkeiten der Bekanntheit der Wörter

kannter Wörter, beziehungsweise bekannter Bedeutungen, mithin die Bekanntheitsgrade der Wörter auf. Die schwarze Linie gibt die Anzahl der zur Verfügung stehenden Wörter in Bezug auf die Bekanntheit des Wortes und die graue Linie in Bezug auf die Bekanntheit auch der Bedeutung an. Wie bereits beschrieben, stehen im Bereich hoher Bekanntheit 1554 Wörter zur Verfügung. Die beiden Linien verlaufen hier noch parallel. Bis zur Abnahme des Bekanntheitsgrades auf etwa 50 % bleibt die Anzahl der zur Verfügung stehenden Wörter noch weitgehend konstant. Verwendet man nur allen Probanden bekannte Wörter, d. h. Wörter unbekannter Bedeutung werden toleriert, so bleiben 1331 Einsilber. Akzeptiert man allerdings nur Wörter, die allen Probanden vollständig bekannt sind, stehen noch 1268 Wörter für die Verwendung in dem Sprachtest zur Verfügung. Es zeigt sich also, dass die Gesamtheit der Wörter, aus denen für die Erstellung der Listen geschöpft werden kann, relativ großen Schwankungen unterliegt, je nachdem, welche Auswahl getroffen wird.



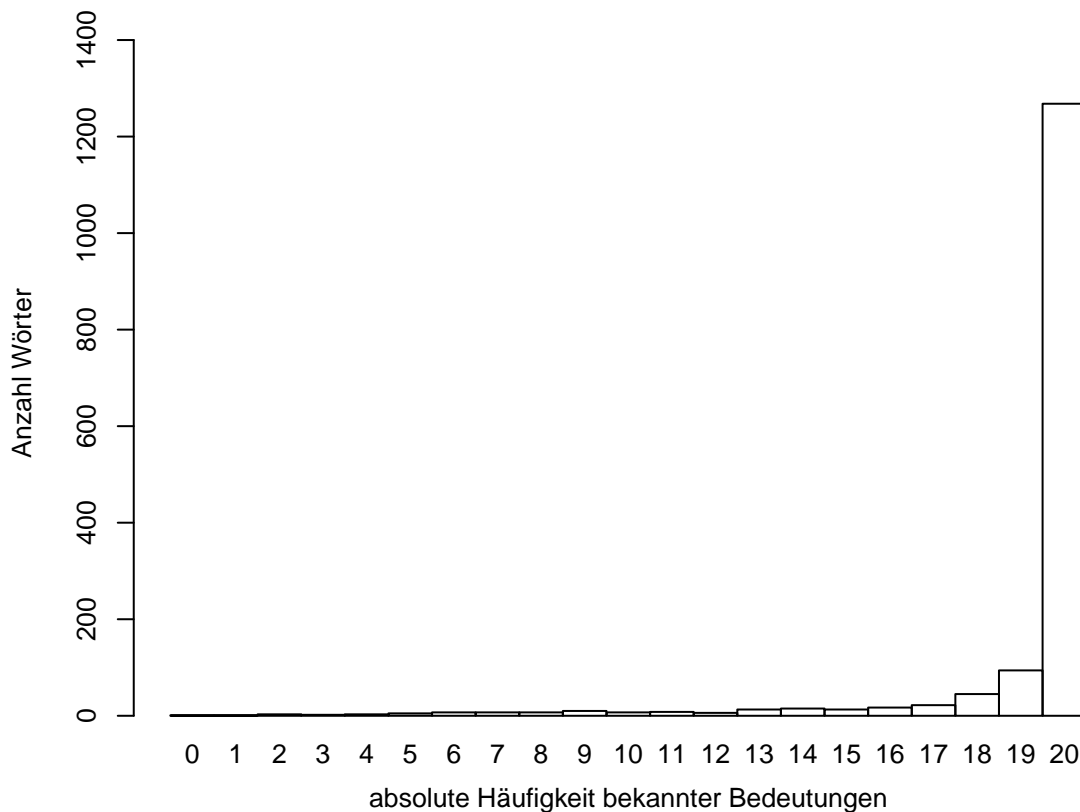


Abbildung 3.17.: Histogramm der absoluten Häufigkeiten der Bekanntheit der Bedeutungen

## 3.2. Verteilung der Wörter in Listen

In Abschnitt 2.5 ist die Methodik für die Zusammenstellung der Listen dargelegt. Wie dort beschrieben, wurde die Menge der Wörter, die allen Probanden vollständig bekannt waren, zunächst zufallsverteilt, um anschließend in nachgeschalteten Optimierungsvorgängen in Bezug auf den Korrektheitsgrad aneinander angeglichen zu werden. Die vier hierfür entwickelten Methoden sind in Abschnitt 2.5.2 ([Globale Methoden für die Verähnlichung der Listen](#), S. 47) dargestellt.

Um zu evaluieren, welche der vier beschriebenen Methoden für die Verähnlichung der Listen am geeignetsten ist, wurde jede der Methoden einhundertmal durchgeführt. Die Ergebnisse sind in vier Streudiagrammen dargestellt (siehe Abb. 3.20), die die Anzahl der Listen gegen den MAD-Wert auftragen. Der MAD-Wert ist eine Maßzahl zur Beschreibung der Streubreite bei Häufigkeitsverteilungen. Er misst die Streuung vom Zentrum des Medians, wobei die Absolutwerte betrachtet werden, da

### 3. Ergebnisse

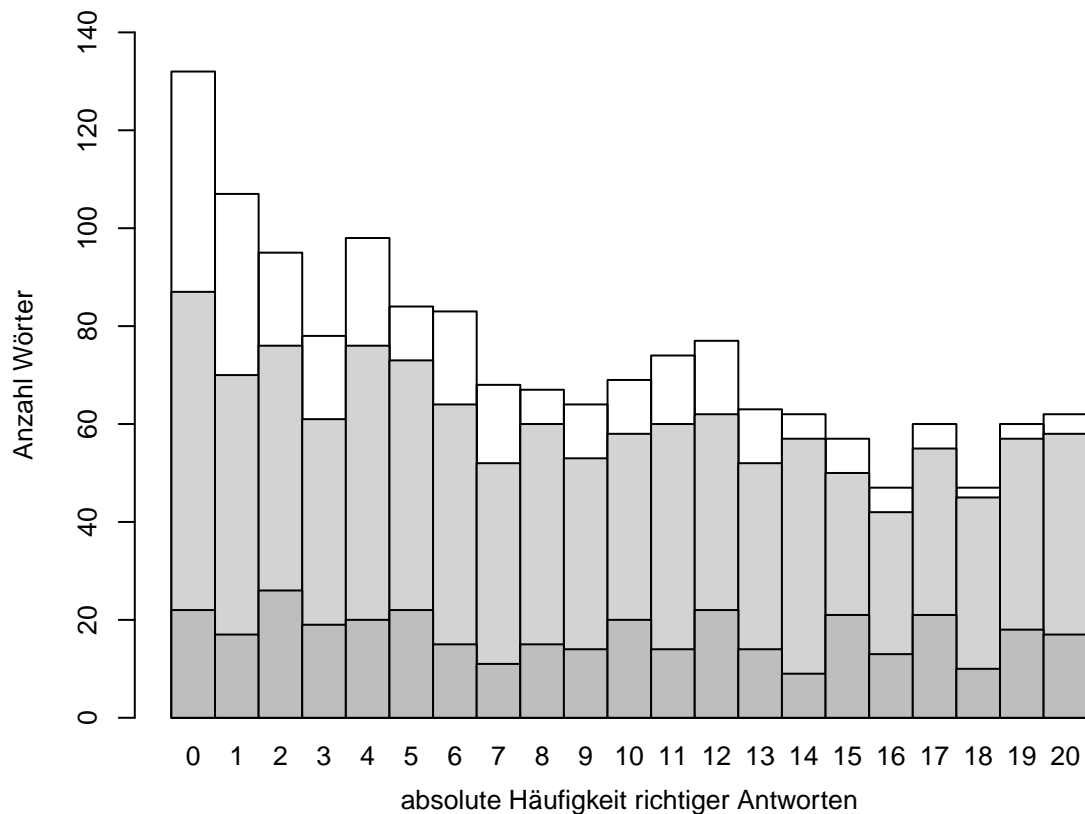


Abbildung 3.18.: Histogramm der richtigen Antworten aller Wörter, der bekannten Wörter und der Wörter, die in den Freiburger Testverfahren zur Anwendung kommen

sich positive und negative Abweichungen vom Zentrum nicht aufheben dürfen. So errechnet sich der MAD aus der Summe der absoluten Fehler geteilt durch die Anzahl der Fehler. Im Gegensatz zum IQR gehen in den MAD alle Werte der Verteilung ein (Kohn, 2000, s. Seite 73).

Die Intensität der betrachteten Punkte gibt in Abbildung 3.20 die Häufigkeit der in diesem Bereich liegenden Ergebniswerte an. Bei der Zufallsverteilung der Wörter schwankte die Anzahl zwischen 56 und 61 entstandenen Listen, wobei 61 Listen nur in zwei Fällen zustande gekommen sind, was bei 400 Verteilungsversuchen einem Prozentwert von 0,005 % entspricht. Der durch die Optimierungsvorgänge erreichte MAD-Wert hat mit Werten zwischen 0,004 und 0,035 eine große Schwankungsbreite. Bei dem Vergleich der vier verschiedenen Optimierungsmethoden fallen deutliche Unterschiede ins Auge.

Das Streudiagramm der ersten Methode, das durch die blaue Farbe dargestellt ist,

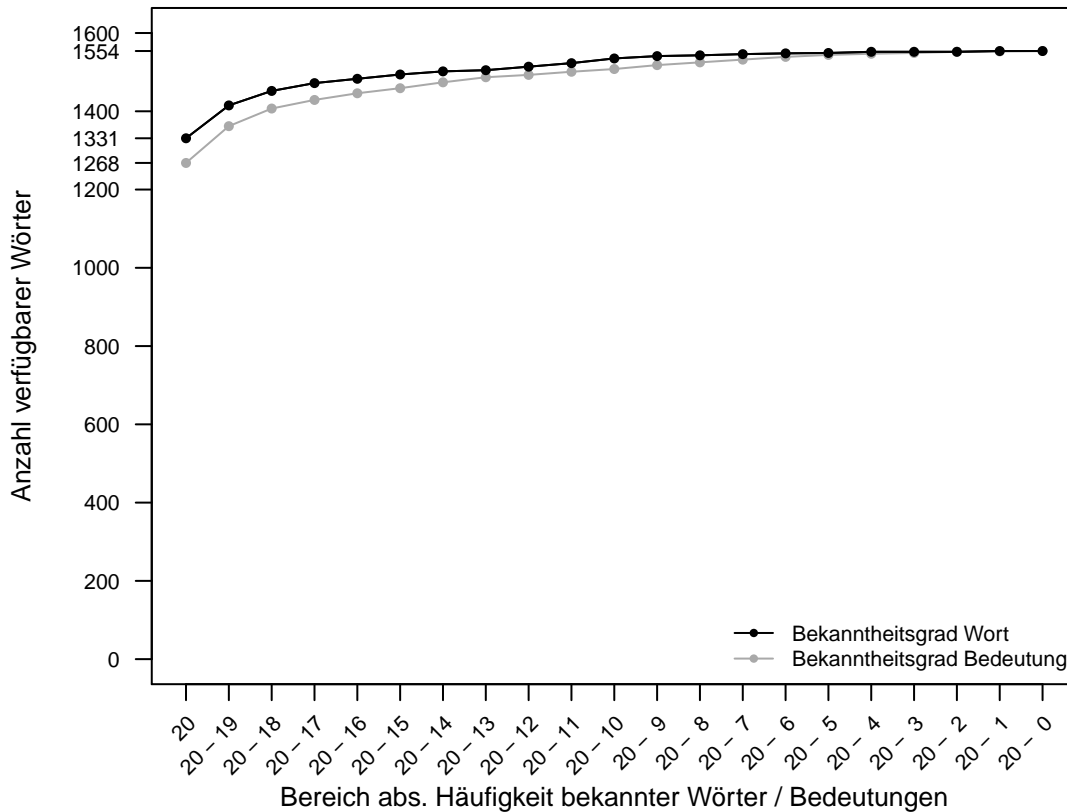


Abbildung 3.19.: Linienplot der Anzahl der verfügbaren Wörter je nach deren Bekanntheitsgrad

zeigt die größte Schwankungsbreite in Bezug auf den MAD-Wert. Bei dieser Verteilungsmethode ist in der Mehrzahl der Fälle nur eine Verkleinerung des MAD-Wertes in einem mittleren Bereich zwischen 0,01 und 0,025 gelungen. Bei Werten bis zu 0,035 zeigt diese Methode auch die Ergebnisse mit den größten MAD-Werten insgesamt.

Die zweite Methode ist in rot dargestellt und zeigt ein ganz anderes Bild. Sie scheint für die Optimierung der Listen wesentlich besser geeignet zu sein. Die erreichten MAD-Werte liegen im Ganzen wesentlich niedriger als bei der ersten Methode. Außerdem ist dies auch diejenige Methode, die die Verteilungsversion mit dem niedrigsten MAD-Wert erreicht hat. Dieser Wert lag bei 0,00371 und ist in der Grafik durch den schwarz umrandete Punkt in dem rechten oberen Streudiagramm eingezeichnet. Im Bereich höherer Listenzahlen sind die erreichten MAD-Werte nicht mehr entsprechend gering. Es scheint ein Zusammenhang zwischen dem MAD-Wert und der Anzahl der Listen zu bestehen. Die niedrigsten MAD-Werte sind im Bereich der

### 3. Ergebnisse

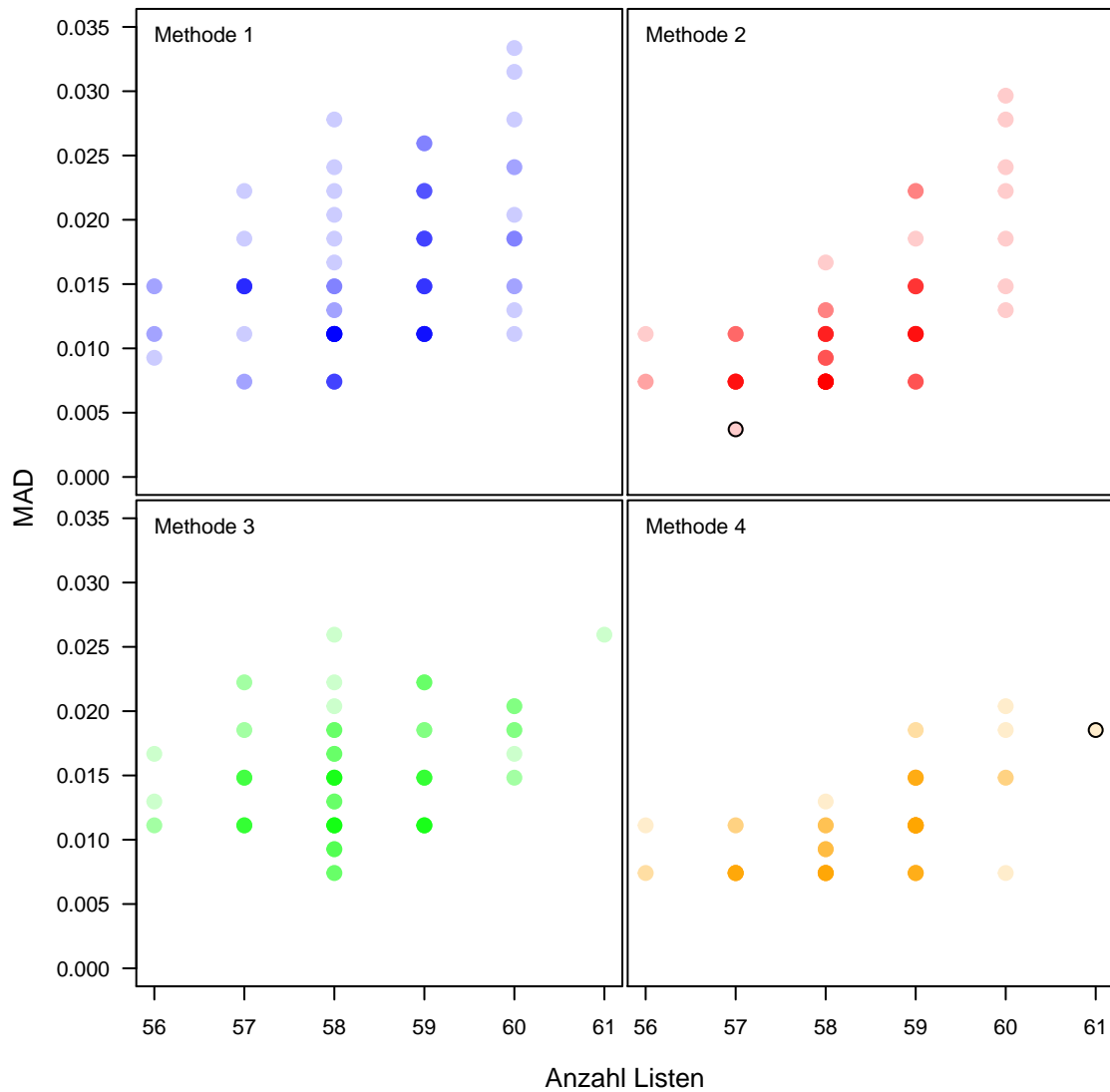


Abbildung 3.20.: Scatterplots der vier Optimierungsmethoden

geringeren Listenanzahlen zu finden.

Die dritte Methode zeigt wiederum ein Bild, das der ersten Methode ähnelt, jedoch etwas geringere Schwankungsbreiten in Bezug auf den MAD-Wert zeigt. Insgesamt fällt aber auch in diesem grün dargestellten Streudiagramm auf, dass trotz der einhundertmal wiederholten Optimierungsschleife nur wenige Verteilungsversionen in einem MAD-Wert unter 0,01 resultierten.

Die vierte orange dargestellte Methode ähnelt den Ergebnissen der zweiten Methode. Auch hier wurden sehr gute MAD-Werte erzielt. Obwohl die Version mit dem absolut geringsten MAD-Wert aus Methode 2 hervorgegangen ist, zeigt diese Metho-

Tabelle 3.3.: Übersichtstabelle der besten zwei Listenzuordnungen

	min. MAD	max. n
Eigenschaft	166	357
Anzahl Listen	57	61
MAD	0.00371	0.01853
Arithm. Mittel	0.48697	0.47389
Standard- Abweichung	0.00872	0.02161
Median	0.48500	0.47250
IQR	0.00500	0.02750

de die durchschnittlich besten Ergebnisse. Die Version mit dem größten MAD-Wert liegt bei nur 0,02 und der Großteil liegt noch wesentlich darunter. Auch die Version mit der größten Listenanzahl und dem dabei geringsten MAD-Wert ist durch diese Methode entstanden. In Methode 3 war ebenfalls eine Verteilungsversion mit 61 Listen entstanden, allerdings liegt hier der MAD-Wert über 0,25. Die Verteilungsversion mit der größten Listenanzahl und dem dabei geringsten MAD-Wert ist im rechten unteren Streudiagramm ebenfalls durch eine schwarze Umrandung dargestellt.

Für die Optimierung der durch eine Zufallsverteilung entstandenen Listen sind also die Methoden 2 und 4 am besten geeignet. Es können mithin wesentlich bessere Ergebnisse erzielt werden, wenn einer Optimierung durch den Pool eine Optimierung der Listen untereinander nachgeschaltet wird. Eine einhundertfache Wiederholung dieser Optimierungsreihenfolge hat nur noch eine relativ geringe Verbesserung der Ergebnisse zur Folge.

Weiterhin auffällig ist die Lage der Ergebnisversionen bezüglich der MAD-Werte. Durch den Austausch von Wörtern streuen die resultierenden Ergebnisse nicht, sondern zeigen eine deutliche Rasterung.

Wie in Abschnitt 2.5 bereits beschrieben, wurden die beiden besten Versionen bezüglich der Anzahl der Listen und dem MAD-Wert aufbewahrt. Die 166. Verteilungsversion zeigte dabei den geringsten MAD-Wert und die 357. Verteilungsversion zeigte die höchste Listenanzahl bei einem niedrigeren MAD-Wert. Tabelle 3.3 stellt die Eigenschaften dieser beiden Versionen einander gegenüber. Betrachtet wurden dabei, außer der Anzahl der entstandenen Listen und dem MAD-Wert, das arithmetische Mittel, die Standardabweichung, der Interquartilsabstand und der Median. Diese Werte beziehen sich jeweils auf den Korrektheitsgrad der entstandenen

### 3. Ergebnisse

Listen. Das arithmetische Mittel ist ein Lokalisationsparameter zur Kennzeichnung der mittleren Lage von Häufigkeitsverteilungen metrischer Merkmale auf der Merkmalsachse (Vogel, 1997, s. Seite 32). Dieses erhält man, indem alle beobachteten Werte aufsummiert werden und diese Summe durch die Anzahl der Beobachtungen dividiert wird (Fahrmeir et al., 2003, s. Seite 53). Diese Werte liegen mit 0,49 und 0,47 sehr nahe beieinander. Im durchschnittlichen Schwierigkeitsgrad unterscheiden sich diese Listen also nur äußerst gering. Die Standardabweichung ist die bekannteste Maßzahl für die Streuung einer Verteilung. Sie misst die Streuung der Daten um ihr Mittel (Fahrmeir et al., 2003, s. Seite 69). Verteilungsversion 166 hat eine wesentlich geringere Standardabweichung als Zuordnung 357 und hat damit eine wesentlich geringere Streubreite um den Mittelwert. Die Berechnung des Medians wurde bereits unter Punkt 2.4 besprochen. Wie das arithmetische Mittel, liegen auch die Median-Werte mit 0,48 und 0,47 sehr nah beieinander. Der IQR ist ebenfalls ein Streuungsmaß für komparative statistische Merkmale. In ihn gehen nur die beiden Werte des ersten und des dritten Quartils ein. Der IQR der 166. Verteilungsversion liegt mit 0,005 wesentlich unter dem IQR der 357. Verteilungsversion, der bei 0,03 liegt. Die Listen der 166. Optimierungsversion zeigen also eine deutlich geringere Streubreite. Obwohl bei dieser Version eine um vier Listen geringere Anzahl zustande gekommen ist als bei der 357. Version, ist aus Gründen der Qualität diese Listenzusammenstellung zu bevorzugen. Das Ergebnis dieser Verteilungsversion kann im Anhang unter Punkt C auf Seite 115 eingesehen werden. Die Wörter einer Spalte ergeben dabei jeweils eine Liste.

## 4. Diskussion

### 4.1. Stellungnahme zu den Kritikpunkten des Freiburger Einsilbertests mit Bezug auf das neue Sprachmaterial M-2007

Wie in Abschnitt 1.2.3 beschrieben wurde, sind seit 1953 eine Vielzahl von Arbeiten erschienen, die den Freiburger Sprachverständnistest kritisieren und Verbesserungsvorschläge vorstellen. Wir haben versucht, auf die wichtigsten und sinnvollsten Vorschläge während der Erstellung des neuen Sprachtests einzugehen und mögliche Fehlerquellen zu vermeiden.

Eine Vielzahl der Kritikpunkte beschäftigt sich mit der unzulänglichen Aufnahmequalität, die sich in einem Rauschen und Knacken der Tonaufnahmen und in einer Unausgeglichenheit der Lautstärke und des Schallpegels der aufgenommenen Wörter manifestiert (Kießling, 2000, s. Seite 633 und Keller, 1977, s. Seite 292 ff.). Auch eine unnatürliche Artikulation der Einsilber ist wiederholt kritisiert worden (Kießling, 2000, s. Seite 633, Doerr, 1994, s. Seite 493. und Plath et al., 1973, s. Seite 458).

Viele dieser im Freiburger Einsilbertest bestehenden Mängel sind durch die Arbeit von Mahfoud behoben worden. So hatte sie durch wiederholte Vorversuche die idealen Aufnahmebedingungen herausgearbeitet, die auch in unseren Aufnahmesitzungen eingehalten wurden. Besondere Aufmerksamkeit hatte sie auf die Aussprache der Testwörter gelegt. Dabei folgte sie nicht dem Vorschlag Kießlings, der eine möglichst langsame Aussprache gefordert hatte, sondern strebte eine natürliche Aussprache der Wörter an (Mahfoud, 2009, s. Seite 21 ff.). Gestützt wird diese Vorgehensweise durch eine Untersuchung von Smoorenburg und Bosman, die die Verstehbarkeit von einsilbigen und zweisilbigen Wörtern in Abhängigkeit von der Artikulation getestet hatten. Dabei hatte sich ergeben, dass eine überdeutliche Aussprache zwar bei normalhörenden Probanden bessere Ergebnisse geliefert hatte, nicht jedoch bei Patienten mit Presbyakusis, bei denen eine normale Artikulation, vor allem bei den

#### 4. Diskussion

Vokalen, dieselben Ergebnisse geliefert hatte, wie die überdeutliche Artikulation der Wörter (Smooenburg und Bosman, 1992, s. Seite 90). Aus diesem Grund scheint es sinnvoll, dem Vorschlag Mahfouds zu folgen, damit die Übertragung der Ergebnisse von den normalhörenden Probanden auf schwerhörige Patienten so wenig wie möglich verfälscht wird. Also wurde auf die Einhaltung dieser Forderung auch bei unseren Aufnahmesitzungen größter Wert gelegt. Die Durchführung dieser Vorgaben gestaltete sich jedoch während der Aufnahmesitzungen zunehmend als schwierig. Wie bereits beschrieben, sollten die Wörter so ausgesprochen werden, als folge ihnen kein Punkt, sondern ein Komma. Dies hatte zum Ziel, die Stimme der Sprecher so konstant wie möglich zu halten und ein Absacken am Ende des Wortes möglichst zu verhindern. Beim Abhören der Wörter ergab sich aber häufig genau das entgegengesetzte Problem. Bei vielen Wörtern stieg die Stimme der Sprecher am Ende des Wortes an, was eine Neuaufsprache der betreffenden Einsilber erforderlich machte. Ein weiteres Problem stellte die reine Artikulation des zu sprechenden Wortes ohne das Beimischen von Nebengeräuschen des Artikulationsraumes dar. Viele Versionen der aufgesprochenen Wörter mussten wegen eines deutlich hörbaren Zungenschlages oder sonstiger Hintergrundgeräusche eliminiert werden. Dieses Problem wurde durch das mehrfache Aufsprechen jedes Einsilbers und durch die wiederholt stattfindenden Aufnahmesitzungen gelöst. So blieb am Ende der Aufnahmen mindestens eine Version guter Qualität jedes Einsilbers übrig.

Auch auf die Einhaltung der gleichen Lautstärke konnte schon während der Aufnahmesitzungen durch das Anbieten eines Referenz-Sprachsignals geachtet werden. Eine weitere Vereinheitlichung des Prüfmaterials wurde erreicht, indem die Maxima der SPL-Zeitverläufe aller Einsilber aneinander angeglichen wurden. Allerdings kann die Forderung nach einem Hörtest, in dem alle enthaltenen Wörter die exakt gleiche Lautstärke und den gleichen Schallpegel aufweisen, nicht erfüllt werden. Denn trotz der beschriebenen Maßnahmen zur Vereinheitlichung des Sprachmaterials in dieser Hinsicht wiesen die aufgenommenen Wörter noch erkennbare Lautheitsunterschiede auf. Dies könnte zum Einen in der Auswahl der Sprecher begründet sein. Denn obwohl die Sprecher große Erfahrung im Bereich der Hörteste hatten, sind ihre Stimmen nicht professionell ausgebildet worden. Zu den Gründen der Auswahl der Sprecher siehe Abschnitt 2.1.2. Zum Anderen könnte ein möglicher Grund für dieses Phänomen in der Gleichrechnung der Lautheit aller Wörter begründet sein. Als Ausgangspunkt für die Gleichrechnung der Lautheit wurden die SPL-Zeitverläufe aller Einsilber gemessen und das jeweilige Maximum bestimmt. Anhand dieser Ma-



xima wurden die Wörter aneinander angeglichen. Es wäre nun also möglich, dass diejenigen Wörter, die durch einen bestimmten lautereren Vokal ein höheres Maximum haben, im Ganzen etwas zu leise berechnet wurden. So könnten Wörter mit einem sehr hohen Maximum im Verhältnis leiser gerechnet werden, als Wörter mit einem geringeren SPL-Maximum. Dies würde auch erklären, warum Wörter, die einen sehr lauten Vokal enthalten (zum Beispiel ein *a*) im Vergleich viel schlechter verstanden wurden, als Wörter die einen leisen Vokal enthalten (zum Beispiel ein *i* oder ein *ue*). Siehe hierzu auch Abschnitt 3.1.3 auf Seite 64. Somit kommen trotz Gleichrechnung über die SPL-Zeitverläufe deutlich hörbare Lautstärkeunterschiede zustande. Dieser problematische Zustand muss vorerst akzeptiert werden, da ein genaueres Verfahren zur Angleichung der subjektiv wahrgenommenen Lautstärke sich noch nicht allgemein durchsetzen konnte. Möglicherweise sollte bei einer zukünftigen Version auf Lautheitsabgleich durch professionelle Toningenieure zurückgegriffen werden. Ein geeignetes automatisiertes Verfahren zur Gleichsetzung der Lautstärke bleibt dennoch erstrebenswert.

Einer der Hauptkritikpunkte, der sich in den verschiedenen Arbeiten vielfach wiederholt, ist die Tatsache, dass die Listen des Freiburger Einsilbertests nicht gleich schwer seien und die Ergebnisse verschiedener Listen somit nicht miteinander verglichen werden könnten, z. B. (Bangert, 1980, s. Seite 168 und von Wedel, 1984, s. Seite 106 ff.). Bei der Zusammenstellung des neuen Sprachtests hatte die Ausgewogenheit der Listen in Bezug auf den Schwierigkeitsgrad die höchste Priorität. Allerdings beruhen die Daten zur Evaluation des Schwierigkeitsgrades der einzelnen Einsilber auf den Versuchsergebnissen von 20 Probanden. Eine geringfügige Änderung der Datenlage würde sich also schon durch den Austausch eines besonders gut verstehenden durch einen eher schlecht hörenden Probanden ergeben. Aus diesem Grund wurden diejenigen Wörter, die zwar eine unterschiedliche Schreibweise, aber eine gleiche Aussprache aufweisen und bei der Evaluation des Sprachmaterials noch in dem Test enthalten gewesen waren, genauer untersucht. Bei einer ausreichend hohen Probandenanzahl müssten die betroffenen Wörter jeweils den gleichen Korrektheitsgrad aufweisen. Dies war bei genauerer Betrachtung jedoch nicht der Fall. Die Wörter *in* und *Inn* wiesen noch eine der geringsten Schwankungen im Korrektheitsgrad auf. Als Einsilber mit unterschiedlicher Schreibweise, aber gleicher Aussprache sind in dem Sprachmaterial außerdem noch die Einsilber *Jazz* und *Jazz\_e*; *Heer*, *hehr* und *her*; *Steg* und *Steak\_dsch*; *Mob* und *Mopp*; *Ried* und *Riet* sowie *Spray\_de* und *Spree* enthalten. Die Unterstriche mit den Endungen *e*,

#### 4. Diskussion

Tabelle 4.1.: Korrektheitsgrade der Wörter gleicher Aussprache

Wort	Lautschrift	% korrekt
Heer	he:ɾ	10
hehr		5
her		10
in	in	50
Inn		70
Jazz	dʒɛs	75
Jazz_e		50
Mob	mɒp	15
Mopp		0
Ried	ri:t	75
Riet		75
Spray_de	ʃpre:	70
Spree		85
Steak_dsch	ʃte:k	80
Steg		95

*dsch* und *de* geben dabei die jeweilige landestypische Aussprache des Wortes an, wobei *e* zum Beispiel für die englische Aussprache steht. Die Schwankungen im Korrektheitsgrad sind bei diesen Wörtern teilweise nicht unerheblich. Tabelle 4.1 stellt die Wörter unterschiedlicher Schreibweise, aber gleicher Aussprache, dem jeweiligen prozentualen Korrektheitsgrad gegenüber. In der zweiten Spalte ist noch einmal durch die Lautschrift die Aussprache der Wörter in der Tabelle der Wörter vertreten. Die mit diesem Versuchsaufbau ermittelten Korrektheitsgrade sind also nicht eindeutig reproduzierbar, was einerseits an einem unterschiedlich ausgeprägten Kaschieren des Sprachsignals durch das überlagerte Rauschen liegen könnte. Andererseits könnte die Anzahl der Probanden auch zu gering sein um reproduzierbare Ergebnisse zu erhalten, d. h., es könnte an rein statistischen Gründen liegen. Auch ist zu bedenken, dass die Evaluation mit normalhörenden Probanden im Störrauschen durchgeführt wurde. Die Frage, inwieweit sich die Ergebnisse auf schwerhörige Probanden übertragen lassen, ist nicht genau zu beantworten. Sicher ist jedoch, dass sich dieses Problem gleichermaßen bei der Entwicklung jedes neuen Hörtests ergibt. Die Antwort wird nur eine Evaluation der Listen mit schwerhörigen Probanden und eine langjährige klinische Erfahrung mit diesem Test liefern können.

Auch die Auswahl der Einsilber ist vielfach kritisiert worden. Schon Mahfoud hatte veraltete Wörter, Wörter, die nur in bestimmten Dialekten oder Regionen Deutschlands gebräuchlich sind und Wörter, die bestimmten Fachgebieten angehören, nicht mit in den Hörtest aufgenommen (Mahfoud, 2009, s. Kap.2.1 auf Seite 13 ff.). Trotzdem waren durch diese Maßnahmen noch recht ungebräuchliche Wörter in dem Sprachmaterial enthalten. Als Beispiel könnten *Kwass*, *Pfriem* oder *Flom* genannt werden. Dieses Problem wurde gelöst, indem diejenigen Wörter aus dem Hörtest entfernt wurden, die den Probanden dieser Dissertation unbekannt waren. Damit ist dieser Test gegenüber dem Freiburger Einsilbertest aussagekräftiger, denn bei dem Freiburger Einsilbertest war die Bekanntheit der Wörter bei der Zusammenstellung der Listen nicht formal überprüft worden. Allerdings muss auch hier beachtet werden, dass es sich bei den Probanden hauptsächlich um Studenten, Ingenieure und Zahnärzte handelte, deren Allgemeinbildung und Wortschatz wohl etwas über dem Durchschnitt liegt. Das Sprachmaterial eines Hörtests, das allen Patienten vollständig bekannt ist, dürfte nur aus Wörtern wie *Tisch*, *Sport*, *Baum*, etc. bestehen und wäre dann für eine umfassende Prüfung mit möglichst vielen Listen unzureichend. Das Sprachmaterial ist nicht nur besser an den heute gebräuchlichen Wortschatz angepasst, weil Wörter entfernt wurden, sondern auch, weil Wörter und Anglizismen neu aufgenommen wurden, die aus dem modernen Wortschatz nicht mehr wegzudenken sind. Damit wurde die Phonemverteilung der Listen an den in der deutschen Sprache vorkommenden Sprachgebrauch angepasst. Auch entsteht dadurch ein wesentlicher größerer Pool von Wörtern, mit denen das Sprachverständnis getestet werden kann. Hahlbrock hatte 800 Einsilber zur Verfügung, während wir auf 1554 Wörter zurückgreifen konnten.

Auch bei der Zusammenstellung der Wörter in Listen waren noch weitergehende Forderungen gestellt worden. Weiland (1954, s. Seite 16) hatte zum Beispiel verlangt, dass innerhalb der Listen keine Wörter aufeinander folgen sollten die einen Sinnzusammenhang ergeben. Obwohl Mahfoud (2009, s. Seite 51) diese Forderung als nicht sinnvoll betrachtet hatte, sind wir der Meinung, dass bei den fertigen Listen durchaus darauf geachtet werden sollte, dies zu vermeiden, da zum Beispiel ein *Hans* mit einem nachfolgenden *Wurst* die Patienten belustigen und so von den nachfolgenden Wörtern ablenken könnte. Oder dass bei den Wörtern *Brot* und einem darauf folgenden *Hel*, statt dessen *Mehl* geantwortet wird, weil der Patient gedanklich noch dem Wort *Brot* nachhängt. Es dürfen also nicht nur keine Wörter aufeinander folgen, die einen Sinnzusammenhang haben, sondern es dürfen auch solche nicht hintereinander stehen, die einen sehr ähnlichen Klang haben und sich reimen. Diese

#### 4. Diskussion

Forderungen sind durch Umstellung der Wortreihenfolge innerhalb der Listen leicht zu erfüllen, ohne dass eine Änderung der Listenzusammenstellung vorgenommen werden muss.

Auf verschiedene andere Forderungen der Kritiker wurde jedoch nicht eingegangen. So hatten Alich und Kiessling längere Testlisten gefordert. Alich hatte dies mit der Verbesserung der Hörergebnisse der Patienten bei wiederholter Anwendung der Testlisten begründet. Auf diese Forderung sind wir aus zwei Gründen nicht eingegangen. Zum Einen sollte der entstehende Sprachtest zumindest in der äußeren Form dem Freiburger Sprachtest ähneln. Zum Anderen ist zu bedenken, dass längere Listen im klinischen Alltag zu einem größeren Zeitaufwand bei der Durchführung des Test führen würden. Dies hätte mit großer Wahrscheinlichkeit erheblichen Einfluss auf die Akzeptanz des Tests. Außerdem kann dem Effekt der Verbesserung der Hörergebnisse der Patienten bei wiederholter Anwendung des Tests auch durch eine größere Anzahl der Testlisten entgegengewirkt werden. Auf das Einfügen eines Ankündigungssignals vor jedem Einsilber wurde aus denselben Gründen verzichtet.

Auch auf Kiesslings Forderung nach einem geschlossenen Antwortsystem für diesen Test wurde nicht eingegangen. Bei Hörtests mit geschlossenen Antwortsystemen können die Probanden das abgespielte Wort mit einer Reihe in geschriebener Form angebotener Wörter vergleichen und von diesen dasjenige auswählen, das dem Gehörten am nächsten kommt. Der Grund für die Entscheidung gegen ein geschlossenes Antwortsystem war wiederum, dass die Durchführung des Tests dem des Freiburger Tests ähneln sollte.

Aber auch die Problematik der verschiedenen Vorkenntnisse der Patienten ist noch nicht entgültig gelöst. So wäre dieser Test mit geschlossenem Antwortsystem nicht mehr für Analphabeten oder Legastheniker anwendbar. Auch für Patienten mit einer fortgeschrittenen Sehminderung wäre so ein Test nicht mehr durchführbar. Die Forderung nach einem Einsilbertest, der dem Freiburger Sprachtest ähnelt, aber ein geschlossenes Antwortsystem aufweist und somit computergesteuert durchführbar ist, ist durchaus sehr sinnvoll. Er müsste allerdings in offener und in geschlossener Form anwendbar sein, um die Patienten, die aus verschiedenen Gründen nicht lesen oder schreiben können, ebenfalls testen zu können. Eine Erweiterung dieses Tests um diese Möglichkeit würde allerdings die Vergleichbarkeit der Ergebnisse unmöglich machen.

Kiesslings Forderung, Wörter, die psychologische Hemmnisse hervorrufen könnten, aus dem Sprachtest zu eliminieren, wurde ebenfalls nicht entsprochen, was

## 4.2. Erweiterungs- und Verbesserungsvorschläge für die Zukunft

hauptsächlich an den Schwierigkeiten lag, diese Wörter zu definieren. Ein Wort, das Kiessling als Beispiel anführt, ist das Wort *Krebs*. Aber außer für die bekannte Krankheit steht dieses Wort auch für das Lebewesen, das wohl nur in den seltensten Fällen ein psychologisches Hemmnis auslösen wird. An welche der beiden Bedeutungen beim Hören dieses Wortes gedacht wird, wird von Patient zu Patient unterschiedlich sein. Andere Einsilber, die Kiessling nicht mit aufgeführt hatte, könnten durchaus eine negative psychologische Reaktion hervorrufen, wie zum Beispiel die Wörter *Bruch* oder *fett*, bei entsprechender Vorerfahrung beziehungsweise Konstitution eines Patienten. Aus diesen Gründen wurden negative Wörter nicht aus dem Pool aller zur Verfügung stehenden Einsilber eliminiert.

Testlisten zu erstellen, die in all diesen Punkten völlig homogen sind und keine Schwankungen in Bezug auf die verschiedenen Einflussfaktoren zulassen, erscheint illusorisch. Einen Test, der diesen Anforderungen gerecht wird, kann es unter der Vorgabe, viele Listen zu erstellen, nicht geben. Auch die Schwankungen, die durch den Einsatz verschiedener Probanden entstehen, können nicht ausgeschaltet werden. Eine gewisse Inkonsistenz muss durch die unterschiedliche Bildung mit den dadurch verbundenen Unterschieden im Wortschatz, die unterschiedliche Bereitschaft der Patienten, an dem Test teilzunehmen, die augenblickliche Stimmungslage der Probanden, ihre Konzentrationsfähigkeit etc. akzeptiert werden.

## 4.2. Erweiterungs- und Verbesserungsvorschläge für die Zukunft

### 4.2.1. Entwicklung des Tests

Was während der Entwicklung des Tests wiederholt Schwierigkeiten bereitet hatte, war die Inkonsistenz der Sprache bezüglich der Aussprache. Die deutsche Sprache ist plurizentrisch. Dies bedeutet, dass die Sprache über mehrere Zentren verfügt und es keine einheitliche Aussprache für den gesamten deutschen Sprachraum gibt. Es gibt also keine Standardaussprache aller Wörter, die als ideal betrachtet werden kann. Denn von vielen Wörtern existieren unterschiedliche Varianten, die als gleichwertig betrachtet werden müssen (Ammon et al., 2004, s. Seite XXXI). Diese verschiedenen Abweichungen des Standard-Deutschen üben in den jeweiligen Regionen des deutschsprachigen Gebiets eine Vorbildfunktion aus. Daher ist es nicht möglich, eine dieser verschiedenen Aussprachen als die „richtige“ und die anderen

#### 4. Diskussion

als dialektgefärbte Abweichungen zu betrachten (Muhr, 2000). Aus diesem Grund haben wir uns bei einigen der Wörter entschieden, mehrere Varianten aufzusprechen und in den Test mit aufzunehmen. Dabei folgten wir den Vorschlägen des Aussprachewörterbuchs des Duden (2006a).

Eine andere, wesentlich größere Schwierigkeit, entstand durch die häufige Differenz zwischen der Schreibweise und der Aussprache. Viele der notwendigen Kategorisierungen, wie zum Beispiel die Einteilung der Wörter nach den Wortstrukturen und nach den in den Wörtern enthaltenen Vokalen, könnten so nicht vollständig automatisiert durchgeführt werden. Ein in Abschnitt 3.1.3 schon erwähntes Beispiel betrifft den Buchstaben *y*, der zwar offiziell ein Konsonant ist, aber häufig die Funktion eines Vokals übernimmt. Ein anderes Beispiel für diese Schwierigkeit ist die Unterscheidung zwischen einem *e* und einem *ä*. In vielen Wörtern ist die Aussprache dieser beiden Buchstaben identisch. Das /ɛ:/, wie in *spät*, stellt in dem System der Vokalpaare betonter Silben eine Ausnahme dar, da er der einzige lange offene Vokal im Deutschen ist. In zahlreichen Varianten der Standardsprache fällt dieses Phonem jedoch mit /e:/ zusammen, was dazu führt, dass die Vokale in *spät* und in *Lese* gleich ausgesprochen werden. Daher ist der Status dieses Lauts als eigenständiges Phonem des Deutschen umstritten (Bunk, 2009, s. Seite 7). Das Duden-Aussprachewörterbuch weist seit der 4. Auflage (2000) auf die Aussprachevariante hin, das /ɛ:/ lang als /e:/ auszusprechen.

Ein weiteres Problem ist das Vorkommen von Wörtern in diesem Sprachtest, die zwar unterschiedlich geschrieben werden und auch eine unterschiedliche Bedeutung haben, aber gleich ausgesprochen werden. Auch in Hahlbrocks Sprachtest kamen solche Wörter vor, zum Beispiel das Wort *Rad*, wie auch das Wort *Rat*. Auch bei der wiederholten Kontrolle des Sprachmaterials des neuen Tests fielen einige Wörter gleicher Aussprache auf. Beispiele finden sich in Abschnitt 4.1 auf Seite 87.

All diese hier geschilderten Schwierigkeiten könnten durch die Änderung der Darstellung der Wörter in Lautschrift umgangen werden. Wörter gleicher Aussprache würden direkt zusammenfallen. Die Einteilung der Einsilber in die verschiedenen Kategorien könnte automatisiert in Bezug auf die Aussprache und nicht auf die Schreibweise stattfinden. Auch in Bezug auf die Aufsprachesitzungen hätte dies den erheblichen Vorteil gehabt, dass dem Sprecher jedes aufzusprechende Wort direkt in Lautschrift vorgelegen hätte und er nicht bei Unklarheiten auf das Aussprachewörterbuch des Dudens hätte zurückgreifen müssen, was eine nicht unerhebliche Zeitersparnis bedeutet hätte. Ich schlage daher vor, bei einer künftigen Verfeinerung der Testver-

fahren die geschriebenen Wörter in Lautschrift mit Hilfe eines dazu zu entwickelnden Programms umzuwandeln. Durch diese Maßnahme könnte auch eine erleichterte Erweiterung des Tests durch eine Analyse von Phonemverwechslungen stattfinden. So könnten individuelle Schwachstellen der Patienten aufgedeckt werden, ohne hierfür ein spezielles Testverfahren zu entwickeln.

### 4.2.2. Aufsprache

Eine Vielzahl der aufgesprochenen Wörter konnte den anschließenden Qualitätskontrollen nicht standhalten. Eine mögliche Ursache war die Auswahl von nicht-professionell ausgebildeten Sprechern. Schon [Mahfoud \(2009, s. Kap. 4.2.1 auf Seite 49\)](#) hatte in ihrer Arbeit zum Ausdruck gebracht, dass professionellen Sprechern in der Zukunft der Vorzug zu geben sei. Dieser Forderung schließe ich mich an, obwohl die Entwicklung von Sprachtests dann mit einem erheblich größeren Kostenaufwand verbunden ist. Das Konstanthalten der Stimme in Bezug auf Lautstärke und Schallpegel sowie die Einhaltung der Genauigkeit der Aussprache ohne das Beimischen von störenden Geräuschen des Artikulationsraumes ist von nicht ausgebildeten Sprechern nur sehr schwer zu realisieren. So musste die Aussprache einiger besonders schwieriger Wörter diverse Male wiederholt werden, wodurch insgesamt über 15 Aufsprachesitzungen zustande kamen. Auch die Durchführung der Aufsprache in einem professionellen Tonstudio wäre im Zweifelsfall zu bevorzugen. Durch die fest-installierten Geräte und einen erfahrenen Tontechniker könnte der Aufnahmeprozess deutlich erleichtert werden. Der komplizierte Aufbau der Gerätschaften und die Schalldämmung der Rückwand durch Schaumstoff wäre dadurch unnötig gewesen. Auch das vereinzelt aufgetretene Knacken in den WAV-Dateien hätte so eventuell sofort vermieden werden können.

### 4.2.3. Vorbereitung des Testmaterials

Um brauchbare Ergebnisse der Evaluation des Testmaterials zu erhalten, musste das Sprachmaterial durch ein hinzugefügtes Rauschen künstlich erschwert werden. Die Stärke dieses Rauschabstandes wurde der vorangegangenen Pilotstudie von Mahfoud entnommen, die einen SRT von -8,54 dB ermittelt hatte.

Für die Evaluation der Einsilber musste also die Stärke und die Art des Rauschens festgelegt werden. Wir entschieden uns für ein Rauschen von -8 dB. Es gab drei Gründe für diese geringgradige Vereinfachung der Prüfungssituation. Erstens sollten

#### 4. Diskussion

die Probanden dieser Dissertation nicht nur 400 Einsilber, sondern alle 1554 Einsilber abhören. Mit der Vereinfachung sollte also die Ermüdung der Probanden in geringem Maße ausgeglichen werden. Zweitens waren die Probanden in der Pilotstudie von Mahfoud mit dem Prüfmaterial vertraut, da es sich teilweise um Mitarbeiter der Klinik handelte und auch ein Mitarbeiter von MED-EL vertreten war (Mahfoud, 2009, s. Seite 42).

Die Probanden, die dagegen an dieser Studie teilnahmen, hatten keine Erfahrung mit dem Prüfmaterial des Freiburger Sprachverständnistests und nahmen zum ersten Mal an einem Einsilber-Sprachtest teil. Wir gingen bei der Wahl des SNR davon aus, dass Probanden, die mit dem Prüfmaterial vertraut sind, eventuell etwas bessere Ergebnisse erzielen, als ungeübte Probanden. Drittens war in der Pilotstudie nur ein Teil der 400 Wörter verwendet worden, die Hahlbrock in seinem Sprachtest als besonders gebräuchlich beschrieben hatte. In dem von uns getesteten Sprachmaterial sind allein durch die Menge der Wörter viele recht ungebräuchliche Wörter enthalten, die eventuell etwas schwieriger zu verstehen sind. Auf diese Vermutung wird in Kapitel 3.1.3 näher eingegangen. Aus diesen Gründen wurde für die Evaluation ein geringeres Stör rauschen ausgewählt, als nach den vergleichenden Untersuchungen in der Pilotstudie von Mahfoud vermeintlich nahe gelegen hätte. Probanden der vorliegenden Arbeit erreichten zwar nicht genau das angestrebte mittlere Sprachverständnis von 50 %, jedoch liegt es mit ca. 45 % sehr gut im empfindlichen Bereich der Messung.

#### 4.2.4. Bildung der Listen

Bei der Bildung der Listen sollte Rücksicht auf eine Vielzahl von Einflussfaktoren genommen werden. Damit sollte sichergestellt werden, dass die entstehenden Listen nicht nur in Bezug auf den Schwierigkeitsgrad, sondern auch in Bezug auf den Bekanntheitsgrad, die Wortstruktur der Einsilber, die Vokale und ihre Dauer, die Anzahl der Lehnwörter und das Vorkommen der Einsilber in den beiden Aufnahmen des Freiburger Sprachtests von 1968 und 1976 ausgeglichen sind. Dadurch war die Anzahl der zu berücksichtigenden Faktoren bei diesem Sprachtest deutlich größer, als es bei dem Freiburger Sprachtest der Fall gewesen war. Vergleiche hierzu Hahlbrock (1970, s. Seite 81 ff.): „Zusammenstellung der Freiburger Wörterlisten“ und Abschnitt 1.2.4. Andererseits sind einige Punkte bei der Aufstellung der Listen nicht beachtet worden.



#### 4.2. Erweiterungs- und Verbesserungsvorschläge für die Zukunft

Keine Rücksicht wurde auf die Verteilung der Konsonanten und auf das ausgeglichene Vorkommen von Affrikaten in den Testlisten genommen. Auch die genaue Lautanzahl wurde nicht berücksichtigt, sondern nur die der Phonemkomplexe. Für Patienten, die nur unter einer Schalleitungsschwerhörigkeit leiden, ist eine Rücksichtnahme auf diese Faktoren nicht notwendig, da die Schwierigkeit der Listen abhängig von diesen Faktoren schon getestet worden ist und eine Änderung bestimmter Konsonantenhäufigkeiten keine Änderung der gesamten Liste in Bezug auf die Schwierigkeit verursachen würde. Es muss aber beachtet werden, dass das Konsonantenverständnis durch eine bestehende Hochtonschwerhörigkeit wesentlich stärker beeinträchtigt wird, als das Vokalverständnis, was daran liegt, dass 60 % der Konsonantenfrequenzen über 1000 Hz auftreten. „Es sind gerade die hohen Frequenzanteile, die wesentlich zur Konsonantenunterscheidung beitragen“ (Betzel, 2004, s. Seite 1). Nun ist es aber möglich, dass bei einigen Patienten mit einer Hochtonschwerhörigkeit das Verständnis bestimmter Konsonanten stärker beeinträchtigt wird, als das anderer. So haben *f*-, *s*- und *sch*- Laute besonders hohe Frequenzanteile, weshalb sie von diesen Patienten etwas schlechter verstanden und häufiger verwechselt werden, als die übrigen Konsonanten (Boenninghaus und Roeser, 1961, s. Seite 448 ff.). Der Schwierigkeitsgrad der Listen kann also für Patienten, die an einer Hochtonschwerhörigkeit leiden, variieren, wenn die Verteilung oben genannter Laute ungleichmäßig ist. Um diese eventuelle Ungleichheit der Listen auszuschalten, sollten die Listen an einem schwerhörigen Probandenkollektiv, bei dem mindestens eine Gruppe an einer Hochtonschwerhörigkeit leidet, evaluiert werden. Sollte sich dabei eine Ungleichheit der Listen in Bezug auf die Schwierigkeit bei diesen Patienten ergeben, so muss die Methode der Listenzusammenstellung um die hier dargelegten Punkte erweitert werden.

Was die Phonemanzahl betrifft, so hatte schon Weiland (1954, s. Seite 19) dargelegt, dass die Phonemverteilung eines Einsilbertests nie mit derjenigen flüssiger deutscher Sprache übereinstimmen kann, da einige Phoneme der deutschen Sprache in einsilbigen Wörtern überhaupt nicht vorkommen, während andere in verminderter oder vermehrter Anzahl vertreten sind. Einen Einsilberhörtest, der die gleiche Phonemverteilung aufweist, die in der gesprochenen deutschen Sprache vorkommt, kann es also nicht geben.

##### 4.2.5. Das Prüfsystem

Eine weitere Fehlerquelle, die bei der Durchführung der Evaluation bei einigen Probanden zu erheblichen Schwierigkeiten geführt hatte, war durch die Art des Prüfsystems begründet. Und zwar stellte es sich als ausgesprochen schwierig heraus, die Antworten einiger Probanden zu verstehen. Je unsicherer die Probanden bezüglich der Korrektheit ihrer Antwort waren, desto leiser und ungenauer formulierten sie diese. Auch die Antwortgeschwindigkeit verringerte sich. Wiederholtes Nachfragen der Prüferin führte häufig zu einer weiteren Verunsicherung der Probanden, mit der Folge, dass Antworten von der Prüferin eventuell falsch aufgenommen worden sein könnten. Diese Gefahr besteht bei der Durchführung des Tests vor allem dann, wenn Patienten mit einem ausgeprägten regionalen Dialekt sprechen.

Diese Fehlerquelle, die durch den Prüfer selbst entsteht, kann dadurch reduziert werden, dass seine Normalhörigkeit einer ständigen Kontrolle unterliegt. Dies ist aber nur eine unzureichende Maßnahme, um eine mögliche Inkorrektheit der notierten Antworten auszuschalten. Eine effektivere Maßnahme wäre das vollständige Ausschalten dieser Fehlerquelle zum Beispiel durch das Aufschreiben der Antworten anstelle der üblichen mündlichen Übertragung des Gehörten. Hierzu könnten vorgedruckte Formulare oder ein PC verwendet werden. Dies hätte noch den zusätzlichen Vorteil, dass die Anwesenheit eines Prüfers nicht mehr erforderlich wäre. Problematisch wären allerdings Eingabefehler oder eventuelle Rechtschreibschwächen einiger Patienten. Es gibt verschiedene Möglichkeiten dieses Problem zu umgehen. Zum einen könnten die Antworten kurz mit einem Audiometrieprüfer besprochen werden, der überprüft, ob die Wörter tatsächlich falsch verstanden wurden, oder ob sie einfach nicht korrekt aufgeschrieben wurden.

Eine andere und weitaus elegantere Möglichkeit wäre die Umwandlung des Geschriebenen in Lautschrift. Dies setzt allerdings voraus, das gesamte Sprachmaterial in Lautschrift aufzubereiten. Die Umwandlung könnte synchron durch ein dafür konstruiertes Programm erfolgen. Auch die Erweiterung des Tests durch ein geschlossenes Antwortsystem würde das vorliegende Problem beheben. So könnte auch hier eine computergesteuerte Durchführung ermöglicht werden. Dillier beschrieb die Vorteile dieser Vorgehensweise auf dem 12. Kolloquium audiologisch tätiger Physiker und Ingenieure wie folgt: „Mit Hilfe einer Informations-Transmissions Analyse können die Wortverwechslungen automatisch ausgewertet werden. Submatrizen würden eine Fehleraufschlüsselung nach phonetischen und artikulatorischen Merkmalen ermöglichen. Die computerisierte Analyse erlaubt eine gezielte Auswahl von Rehabili-

tationsmaterial und Signalverarbeitungsmöglichkeiten. Verschiedene Hörgeräte und Parametervariationen können gezielt und wiederholt mit dem gleichen Testmaterial (jeweils neu randomisiert) evaluiert werden“ (Dillier, 1985, s. Seite 180). Eine Erweiterung des Tests um eine der hier dargelegten Möglichkeiten sollte in der Zukunft in Betracht gezogen werden.

### 4.2.6. Auswertung

„Die wichtigste Grundlage für die Hörverlustbestimmung und damit für die MdE-Einschätzung ist der sprachaudiometrische Befund unter Berücksichtigung des gewichteten Gesamtwortverstehens“ (Brusis, 1999a, s. Seite 142 und Brusis, 1999b, s. Seite 211).

Hierfür standen bisher Tabellen von Boenninghaus und Roeser zur Bestimmung des prozentualen Hörverlusts zur Verfügung (Boenninghaus und Roeser, 1973, s. Seite 153). Aus diesen Tabellen ließ sich der Hörverlust in Prozentzahlen ablesen und durch getrenntes Aufsuchen des prozentualen Hörverlustwertes für jedes Ohr in der Tabelle die von der Deutschen Gesellschaft für Hals-, Nasen-, Ohrenärzte festgesetzte prozentuale Minderung der Erwerbstätigkeit ablesen (Brusis und Feldmann, 1973, s. Seite 161). Feldmann konstruierte 1995 eine neue Tabelle in Verbindung zu denen von Boenninghaus und Roeser mit angehobenen Eckwerten für die MdE-Ermittlung bei der ein- und der beidseitigen Taubheit bei 20 bis 80 % (Brusis, 1999b, s. Seite 213). Diese MdE-Tabellen sind seither mit einigen Abänderungen noch allgemein in Gebrauch, obwohl seither einige Beanstandungen aufgetaucht sind. „Seit einer Reihe von Jahren gibt es ein wachsendes Unbehagen bei vielen Gutachtern, weil sie den Eindruck haben, dass die strikte Anwendung der Tabelle nach Boenninghaus und Roeser zur Bestimmung des prozentualen Hörverlustes und, davon ausgehend, die Bemessung der MdE nach der oben genannten Tabelle dem tatsächlichen Hörverlust nicht immer gerecht wird. Dieser Meinung sind auch einige Geschädigte und daher rührt eine Flut von sozialgerichtlichen Verfahren mit dem Ziel, die MdE durch Gerichtsentscheid festsetzen, d. h. höher einstufen zu lassen“ (Feldmann, 1988, s. Seite 320). Allerdings stellen die im Königssteiner Merkblatt enthaltenen Tabellen und Übersichten zur Einschätzung der MdE allgemeine Anhaltspunkte dar und eröffnen dem Gutachter einen Beurteilungsspielraum für die Bewertung des Einzelfalles (Brusis, 1999a, s. Seite 140).

Ein Ergebnis der moderneren Aufnahmebedingungen und der Auswahl der Wörter bei der Entwicklung des neuen Sprachtests ist die deutlich bessere Verstehbarkeit

#### 4. Diskussion

des neuen Sprachmaterials M-2007 im Vergleich zu den Aufnahmen des Freiburger Sprachtests von 1968 und 1976. Dies hatte sich auch in der Pilotstudie von Mahfoud abgezeichnet. Bei der älteren Aufnahme des Freiburger Sprachverständnistests, die auch heute noch standardmäßig verwendet wird, lag der SNR bei Erreichen des SRT bei -3,75 dB, während er bei dem neuen Sprachverständnistest M-2007 bei -8,54 dB lag (Mahfoud, 2009, s. Seite 42 ff.). Dies hat die Konsequenz, dass der Schwierigkeitsgrad der Listen bei der fixen Zuordnung bei 0,475 (47,5% der in den Listen enthaltener Wörter wurde von den Probanden korrekt verstanden) liegt, womit die Listen deutlich einfacher zu verstehen sind, als die des Freiburger Einsilbertests.

Damit können die Ergebnisse der beiden Testverfahren nicht mehr miteinander verglichen und aufeinander bezogen werden. Auch die Auswertungstabellen der Ergebnisse des Freiburger Tests zur Ermittlung der MdE können nicht ohne Weiteres auf den neuen Test angewendet werden. Somit kommt der Test in seiner jetzigen Form für die Erstellung von Gutachten nicht in Frage. Die Möglichkeit mit dem zur Verfügung stehenden Sprachmaterial schwierigere Listen zusammenzustellen, musste aufgrund der Vorgabe, eine möglichst große Listenanzahl zu erreichen, verworfen werden. Aus diesem Grund müssen in der Zukunft neue Verfahren zur Auswertung gefunden werden oder es müssen Verfahren zur Angleichung der Ergebnisse entwickelt werden. Eine Möglichkeit den Test zu erschweren, wäre, die Wörter mit einer geringeren Intensität anzubieten oder sie in Verbindung mit einem definierten Störgeräusch abzuspielen. Der dazu nötigen SNR, beziehungsweise die Stärke der Intensitätsminderung, könnte durch vergleichende Untersuchungen ermittelt werden. Eine andere Möglichkeit wäre die Umrechnung der Ergebnisse, so dass weiterhin die alten Formulare für die Ermittlung der MdE verwendet werden könnten. Schlussendlich wäre auch die Erarbeitung neuer Tabellen für die Erstellung der Gutachten möglich. Es ist aber festzuhalten, dass der Test auch ohne diese Erweiterungen sehr wohl für Forschungszwecke verwendet werden kann. Durch die Methode der Zusammenstellung konnte ein sehr homogener Schwierigkeitsgrad der Listen erreicht werden.

## 5. Ausblick

Durch die Evaluation des Sprachmaterials M-2007 durch eine homogene Gruppe 20 normalhörender junger Erwachsener konnte ein Einsilbertest konstruiert werden, der aus Listen sehr gleichmäßiger Schwierigkeit besteht. Diese Listenzusammenstellung ist jedoch noch nicht mit hörgeschädigten Probanden, der eigentlichen Zielgruppe, evaluiert worden. Dies wäre eine erforderliche Aufgabe zur Überführung des Tests in die Forschungs- bzw. die klinische Anwendung.

In diesem Zusammenhang ist weiterhin ungeklärt, ob die bezüglich der Phonemverteilung, Wortstruktur und Vokallänge nur zufällig zusammengestellten Listen bei anderen Probanden oder Patienten ebenfalls gleiche Schwierigkeitsgrade aufweisen. Es wäre zu überprüfen, ob diese erweiterten Ähnlichkeitskriterien verbessert werden können, ohne die nominell gleichen Listen wieder zu verschlechtern. Sollte eine solche Verbesserung möglich sein, dann wäre diese einer Evaluierung durch Probanden voranzustellen.

Um das Sprachmaterial des neuen Einsilber-Sprachtests zu vervollständigen, wäre eine vollständige Aufsprache einer weiblichen Stimme wünschenswert. Auch diese würde eine Evaluierung erfordern, was den Gesamtaufwand dupliziert. Fraglich wäre, ob eine beiden Aufsprachen gemeinsame Listenzusammenstellung möglich ist, die dennoch gleiche Listenschwierigkeiten besitzt.

Die neue, technisch bessere Aufsprache M-2007 besitzt eine deutlich höhere durchschnittliche Verständlichkeit, als die Aufsprache nach DIN 45 621-1. Würde diese Aufsprache für die Erstellung von MdE-Gutachten verwendet, so hätte dies eine sachlich falsche, für den Patienten nachteilige Einstufung zur Folge. Daher darf das Sprachmaterial M-2007 nur nach Beaufschlagung mit Korrekturfaktoren für diesen Zweck verwendet werden, wobei die Festlegung einer solchen rechnerischen Korrektur eine eigene aufwendige Evaluierung erfordert.

Unterschiedlich geschriebene Wörter gleicher Aussprache sollten entfernt werden. Durch eine anschließende neue Listenzusammenstellung mit erhöhtem Rechenzeitaufwand könnten möglicherweise sogar Verteilungsversionen mit 62 Listen erreicht

## 5. Ausblick

werden. Eine Modifikation der Methodik der Listenzusammenstellung könnte darin bestehen, nach der initialen Zufallszuordnung nur Listenanzahlen von mindestens 60 zur Optimierung zuzulassen, und diese dann der 2. und der 4. Optimierungsmethode zuzuführen.

## 6. Zusammenfassung

Der im klinischen Alltag gebräuchlichste Sprachverständnistest ist der von Hahlbrock 1953 entwickelte Freiburger Sprachtest. Dieser Sprachtest steht allerdings wegen einer Vielzahl bestehender Mängel in der Kritik. Um einen neuen Sprachtest zur Verfügung stellen zu können, der diesen Kritikpunkten standhält, hatte Mahfoud (2009) einen Großteil der in Deutschland gebräuchlichen Einsilber mit moderneren digitalen Geräten mit Hilfe eines Sprechers (Dipl.-Ing. S. Brill) aufgenommen. In dieser Arbeit wurden die 378 Wörter, die die Schnittmenge der beiden Aufnahmen des Freiburger Einsilberverständnistests von 1968 und 1976 bilden, in drei Aufsprachevarianten an vier Probanden evaluiert. Dabei hatten sich Sprachverständlichkeitsschwellen (SRT) von -3,75 dB (1968), -5,8 dB (1976) und -8,54 dB (M-2007) ergeben. Es hatte sich also gezeigt, dass das neu aufgenommene Sprachmaterial wesentlich besser verstanden wird, als die Aufnahmen von 1968 und 1976. Mahfoud hatte eine Basis für die Konstruktion eines neuen Sprachtests geschaffen. Diese wurde ausgebaut, indem die Aufnahmen der ausgewählten Wörter vervollständigt, die Qualität des gesamten Materials überprüft und gegebenenfalls ersetzt wurde.

Ziel der vorliegenden Arbeit war die Erzeugung eines Einsilber-Sprachtests mit ausgeglichener durchschnittlicher Verständlichkeit der Wortlisten. Die Evaluation erfolgte mit einer homogenen Gruppe 20 normalhörender, junger Erwachsener. Um in einen empfindlichen Messbereich zu gelangen, wurden die WAV-Dateien nach Lautheitsabgleich aller Sprachsignale anhand des Maximums des SPL-Zeitverlaufs, mit einem CCITT-Rauschen von -8 dB belegt. Jeder der Probanden hörte alle 1554 Wörter in jeweils unterschiedlicher Reihenfolge ab und gab später Auskunft, ob er die Wörter und ihre Bedeutung kannte.

Die Ergebnisse wurden mit Hilfe eines Auswertungsprogramms, das in der Statistik-Programmiersprache „R“ erstellt wurde, ausgewertet. So konnte nicht nur der Schwierigkeitsgrad aller 1554 Einsilber ermittelt werden, sondern auch der Bekanntheitsgrad jedes einzelnen Wortes. Ebenso konnten die Wortstrukturen, die in den Einsilbern enthaltenen Vokale und deren Länge, die Anzahl der Lehnwörter und die Infor-

## 6. Zusammenfassung

mation, ob die betreffenden Wörter auch in den jeweiligen Aufnahmen des Freiburger Sprachverständnistests enthalten gewesen waren, mit einbezogen werden.

Diese Informationen wurden für die Verteilung der Wörter in Listen verwendet. Das wichtigste Kriterium war dabei der Schwierigkeitsgrad der Wörter. Mit Hilfe der so erhobenen Daten konnte automatisiert eine zufällige initiale Verteilung der Wörter in möglichst viele Listen und eine anschließende Harmonisierung der Listen in Bezug auf den Schwierigkeitsgrad vorgenommen werden. Es wurden vier verschiedene Methoden der Zuordnung implementiert, die sich als unterschiedlich leistungsfähig erwiesen. Das bezüglich der *Homogenität* der Listen beste Resultat bestand aus 57 Listen mit einem MAD-Wert der Schwierigkeitsgrade von 0,00371 und einem Median der Verständlichkeit von 48,50 %. Das bezüglich der *Anzahl* günstigste Resultat bestand aus 61 Listen mit einem MAD-Wert von 0,01853 und einem Median von 47,25 %.

Das Geschlecht der Probanden hatte keine Auswirkungen auf das Sprachverständnis, ebenso konnte keine Abhängigkeit von der Vokallänge und von der Zeitposition im Testverlauf nachgewiesen werden. Die Eigenschaften Wortstruktur, Zentralvokal, Sprachherkunft und Bekanntheitsgrad erwiesen sich jedoch als statistisch signifikante Einflussfaktoren.

Damit ist die Grundlage eines neuen Sprachtests entstanden, der für Forschungszwecke mit Normalhörenden geeignet ist. Für die Verwendung mit hörgeschädigten Probanden ist jedoch die weitere Evaluierung erforderlich.



# A. Verzeichnis verwendeter Abkürzungen

ADA	Arbeitsgemeinschaft deutscher Audiologen
CCITT	Comité Consultatif International Téléphonique et Télégraphique
dB	Dezibel
HSM-Test	Hochmair-Schulz-Moser-Test
IQR	Interquartilsabstand (inter quartile range)
MAD	Mittlere absolute Abweichung vom Median (median absolute deviation)
MdE	Minderung der Erwerbstätigkeit
SD	Standardabweichung (standard deviation)
SNR	Signal-Rausch-Abstand (signal to noise ratio)
SPL	Schalldruckpegel (sound pressure level)
SRT	Sprachverständnisschwelle (speech reception threshold)
WAV	Waveform Audio File
$\chi^2$	Chi-Quadrat (Teststatistik)

## *A. Verzeichnis verwendeter Abkürzungen*

# Literaturverzeichnis

- G. Alich. Anmerkung zum Freiburger Sprachverständnistest. *Sprache, Stimme, Gehör*, 9:1–6, 1985.
- K. Amersbach und F.J. Meister. Ein Wortkatalog für isophone Sprachgehörprüfung. *HNO*, 157:352–358, 1950.
- U. Ammon, H. Bickel, J. Ebner, R. Esterhammer, M. Gasser, L. Hofer, Kellermeier-Rehbein, H. Löffler, D. Mangott, H. Moser, R. Schläpfer, M. Schloßmacher, R. Schmidlin, R. Vallaster, G. Unter Mitarb. v. Kyvelos, R. Nyffenegger und T. Oehler. *Variantenwörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz und Deutschland, sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. De Gruyter, 2004.
- H. Bangert. Probleme der Ermittlung des Diskriminationsverlustes nach dem Freiburger Sprachtest. *Audiologische Akustik*, 19:166 – 170, 1980.
- R. Baumann. *Der Würzburger Kindersprachtest: Entwicklung, Standardisierung und dessen Erprobung bei normalhörenden und cochleaimplantierten Kindern*. Dissertation, Bayerische Julius-Maximilians-Universität zu Würzburg, 2006.
- S. Betzel. *Der HSM-Sprachtest mit Hochpassfilter und Rauschen bei Patienten im Alter von 50 bis 60 Jahren*. Dissertation, Klinik und Poliklinik für Hals-, Nasen- und Ohrenkranke der Bayerischen Julius-Maximilians-Universität Würzburg, 2004.
- H.-G. Boenninghaus und D. Roeser. Sprachaudiometrische Untersuchungen bei Zusatz von Schmalbandgeräuschen. *Archiv für Ohren-, Nasen- und Kehlkopf-Heilkunde*, 177:448–457, 1961.
- H.-G. Boenninghaus und D. Roeser. Neue Tabellen zur Bestimmung des prozentualen Hörverlustes für das Sprachgehör. *Z. Laryng. Rhinol*, 52:153–161, 1973.

- J. Bortz. *Statistik für Human- und Sprachwissenschaftler*. Nr. 6. Springer-Verlag, 2005.
- A. Bosman. *Review of Speech Audiometric Tests, in Moderne Verfahren der Sprachaudiometrie*. Median-Verlag, 1992.
- T. Brand. Sprachaudiometrie in der Forschung. In *8th EFAS Congress/ 10th Congress of the German Society of Audiology*, 2007.
- T. Brusis. Begutachtung der Lärmschwerhörigkeit nach dem neuen Königssteiner Merkblatt, Teil 1: Audiometrie und Diagnose. *HNO*, 47:140–153, 1999a. Redaktion H. Schultz-Coulon, Neuss.
- T. Brusis. Begutachtung der Lärmschwerhörigkeit nach dem neuen Königssteiner Merkblatt, Teil 2: Begutachtung. *HNO*, 47:211–219, 1999b. Redaktion: H. Schultz-Coulon, Neuss.
- T. Brusis und H. Feldmann. Die neuen Tabellen des prozentualen Hörverlustes von Boenninghaus und Roeser in der Erprobung. *Z. Laryng. Rhinol*, 52:161–164, 1973.
- J.S. Bunk. *Phonetik aktuell*. Hueber Verlag Hans Huber, 2009.
- R. Caussé und P. Falconnet. Occupational Aviation Deafness. *Ann. d'Oto-laryng.*, 64:436, 1947.
- A. Dekant. *Vergleich der Aufsprache des Freiburger Sprachverständnistests des Norddeutschen Rundfunks, Sprecher Claus Wunderlich, von 1969 und der Aufsprache des sogenannten „verbesserten Freiburgertests Schweizeraufsprache“ unter Anleitung von Tonmeister Neukomm, Basel, Sprecher Peter Richner, von 1987*. Dissertation, Bayerische Julius-Maximilians-Universität, Würzburg, 1992.
- H. G. Dieroff und W. Mangoldt. Experiences with the distorted Freiburg Speech Test—diagnosis and rehabilitation of hearing disorders. *Laryngorhinootologie*, 68 (7):372–378, Jul 1989.
- H. G. Dieroff und W. Meissner. Distorted Freiburg Speech Test. *Laryngol Rhinol Otol (Stuttg)*, 64(9):466–469, Sep 1985.

- H. G. Dieroff und W. Meissner. Detection of inactivity of the auditory system in the beginning stage with the Freiburg masked speech test. *HNO*, 42(1):58–65, Jan 1994.
- N. Dillier. Automatisierung der Sprachaudiometrie durch eine computerisierte Messung der Phonem-Diskrimination, 12. Kolloquium audiologisch tätiger Physiker und Ingenieure. *Audiologische Akustik*, 24:178–186, 1985.
- L. Doerr. Comparative speech audiometry examinations with compact disk and cassette tape. *HNO*, 42(8):493–498, Aug 1994.
- W.H. Döring und V. Hamacher. *Neue Sprachverständlichkeitstests in der Klinik: Aachener Logatomtest und Dreieinsilbertest im Störschall, in Moderne Verfahren der Sprachaudiometrie*. Median-Verlag, 1992.
- Duden. *Deutsches Universalwörterbuch*. Bibliographisches Institut, Mannheim, 2004. Version 3.0.
- Duden. *Das Aussprachewörterbuch*, Band 6. Bibliographisches Institut Mannheim, 2006a.
- Duden. *Das Herkunftswörterbuch: Etymologie der deutschen Sprache. Die Geschichte der deutschen Wörter bis zur Gegenwart*. Bibliographisches Institut, Mannheim, 2006b.
- A. Ederhof. *Das Mikrophonbuch. Optimaler Einsatz im Studio und auf der Bühne*, Band 1. Carstensen Verlag, 2004.
- J.P. Egan. Articulation testing methods. *The Laryngoscope*, 58:955–991, 1948.
- L. Fahrmeir, R. Künstler, I. Pigeot und G. Tutz. *Statistik. Der Weg zur Datenanalyse*. Springer Verlag, 2003.
- H. Fastl. Ein Störgeräusch für die Sprachaudiometrie. *Audiol. Akustik*, 26:2–13, 1987.
- H. Feldmann. Audiometrie bei Erwachsenen. Kapitel 11, Seiten 1–63. Georg Thieme Verlag, 1979.
- H. Feldmann. Die Problematik der quantitativen Bewertung von Hörstörungen in der Begutachtung. *Laryngo-Rhino-Otologie*, 67:319–325, 1988.

- H. Feldmann. 200 years testing hearing disorders with speech, 50 years german speech audiometry – a review. *Laryngorhinootologie*, 83(11):735–742, Nov 2004. doi: 10.1055/s-2004-825717. URL <http://dx.doi.org/10.1055/s-2004-825717>.
- K. H. Hahlbrock. Kritische Betrachtungen und vergleichende Untersuchungen der Schubertschen und Freiburger Sprachteste. *Z Laryngol Rhinol Otol*, 39:100–115, 1960a.
- K. H. Hahlbrock. *Sprachaudiometrie - Grundlagen und praktische Anwendung einer Sprachaudiometrie für das deutsche Sprachgebiet*. Georg Thieme Verlag, Stuttgart, 1970.
- C. V. Hudgins, J. E. Hawkins, J. E. Karlin und S. Stevens. The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech. *Laryngoscope*, 57(1):57–89, Jan 1947.
- M. Schubert J. Kiessling und I. Wagner. Sprachverständlichkeitsmessungen an Normalhörenden und Schallempfindungsschwerhörigen - fünf Sprachtests im Vergleich. *Audiologische Akustik*, 1 und 2:6 – 15,18,19; 11 – 15, 1994.
- W. D. Keidel. Physiologisch-akustische Grundlagen des menschlichen Sprachverständnisses. *Audiologische Akustik*, 19:172–180, 1980. Vortrag, gehalten auf dem 24. Hörgeräte-Akustiker-Kongreß 1979 in Baden Baden.
- F. Keller. Verschiedene Aufsprachen des Sprachverständlichkeitstests nach DIN 45 621 ("Freiburger Test"). *Biomed. Techn.*, 22:292– 298, 1977.
- J. Kiessling. Moderne Verfahren der Sprachaudiometrie. *Laryngo-Rhino-Otologie*, 79:633 – 635, 2000.
- J. Kiessling. Fehlerquellen in der Audiometrie – Grundlagen und Abhilfe. *Praktische Arbeitsmedizin*, 6:12–16, 2006.
- J. Koch und E. Weiland. Investigations on the relation between the audiometric curve and the understanding of speech. *Z Laryngol Rhinol Otol*, 29(9):423–429, Sep 1950.
- W. Kohn. *Statistik, Datenanalyse und Wahrscheinlichkeitsrechnung*. Springer Verlag, 2000.

- B. Kollmeier. Medizinische Akustik und Audiologie: Psychoakustik und Sprachperzeption bei Normal- und Schwerhörigen. Technischer Bericht, XVII. Winterschule für Medizinische Physik: Deutsche Gesellschaft für Audiologie, Deutsche Gesellschaft für Medizinische Physik, Österreichische Gesellschaft für Medizinische Physik, Schweizerische Gesellschaft für Strahlenbiologie und Medizinische Physik, 2009.
- M. Kompis. *Audiologie*. Verlag Hans Huber, 2004.
- M. Kompis, M. Krebs und R. Häusler. Verification of normative values for the Swiss version of the Freiburg speech intelligibility test. *HNO*, 54(6):445–450, Jun 2006. doi: 10.1007/s00106-005-1337-8. URL <http://dx.doi.org/10.1007/s00106-005-1337-8>.
- H. Lazarus, C.A. Sust, R. Steckel, M. Kulka und P. Kurtz. *Akustische Grundlagen sprachlicher Kommunikation*. Springer Verlag, Berlin, 2007. 1. Auflage.
- E. Lehnhardt und R. Laszig. *Praxis der Sprachaudiometrie*. Georg Thieme Verlag, Stuttgart, 2000.
- M. Mahfoud. *Neuaufsprache und Evaluation des Einsilber- Sprachverständnistests*. Dissertation, Julians-Maximilians-Universität Würzburg, 2009. URL <http://www.opus-bayern.de/uni-wuerzburg/volltexte/2010/4854>.
- R. Muhr. Österreichisches Staatsdiplom. CD-ROM, 2000.
- W. Niemeyer. Audiologie und Phoniatrie. Sprachaudiometrie mit Sätzen: Grundlagen und Testmaterial einer Diagnostik des Sprachverständnisses. *HNO*, 15:335–43, 1967.
- P. Plath. New recording of the Freiburg speech discrimination test. *HNO*, 19(9): 282–284, Sep 1971.
- P. Plath, H. W. Stühlen, H. Graf und H. Pelzer. Investigations on the intelligibility of a new record of the German Freiburg speech discrimination test. *Z Laryngol Rhinol Otol*, 52(6):457–469, Jun 1973.
- R Development CoreTeam. *R: A Language and Environment for Statistical Computing*. R foundation for Statistical Computing, 2010. URL <http://www.R-project.org>. ISBN 3-900051-07-0.

- D. Roeser. Sprachgehör und Tonaudiometrie. *Z Laryngol Rhinol Otol*, 42:851–861, 1963.
- L. Sachs und J. Hedderich. *Angewandte Statistik, Methodensammlung mit R*. Springer Verlag, 2006. 12. vollständig neu bearbeitete Auflage.
- M. Y. Saglam. Eine lexikologische Wortschatzuntersuchung des einsprachigen türkischen Wörterbuches TÜRKÇE SÖZLÜK aus dem Jahre 1945. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 20:85–94, 2003.
- G. Schlöndorff und W. Tegtmeier. Untersuchung zur Beziehung zwischen Selektionsfähigkeit und dichotischem Hörvermögen. *Z. Laryng. Rhinol.*, Georg Thieme Verlag, Stuttgart, 52:31–33, 1973.
- K. Schubert. Vergleich alter und neuer Hörprüfmethoden. *Archiv für Ohren-, Nasen- und Kehlkopf-Heilkunde*, 157:99–116, 1950a.
- K. Schubert. Der klinische Wert verschiedener audiometrischer Prüfmethode. *Archiv für Ohren-, Nasen- und Kehlkopf-Heilkunde*, 157:328–340, 1950b.
- H.-J. Schultz-Coulon. Sprachaudiometrie mit Sätzen und Geräusch. *Laryngo-Rhinology*, 53:734–749, 1974.
- W.F. Sendlmeier. *Moderne Verfahren der Sprachaudiometrie, Tests zur Optimierung der Sprachverständlichkeit- phonetische Feinstrukturanalysen des Artikulationsortes bei Plosiven*. Median Verlag, 1992.
- G.F. Smoorenburg und A. Bosman. *Choosing Speech Materials to Assess Hearing Impairment in Moderne Verfahren der Sprachaudiometrie*. Median Verlag, 1992.
- M. Spreng. Risikofaktor Lärm - Physiologische Aspekte. *Therapiewoche*, 34:3765–3772, 1984.
- H. Sukowski, T. Brand, K. C. Wagener und B. Kollmeier. Comparison of different speech intelligibility tests in German language (Freiburg speech test vs. Göttingen sentence test and monosyllabic rhyme test). *HNO*, 57(3):239–250, Mar 2009. doi: 10.1007/s00106-008-1727-9. URL <http://dx.doi.org/10.1007/s00106-008-1727-9>.
- P. Vary, U. Heute und W. Hess. *Digitale Sprachsignalverarbeitung*. B.G. Teubner Verlag, Stuttgart, 1. Ausgabe, 1998.



- W. Voß. *Taschenbuch der Statistik*, Band 2. Auflage. Fachbuchverlag Leipzig, 2004.
- F. Vogel. *Beschreibende und schließende Statistik*. R. Oldenbourg Verlag München Wien, 1997.
- A.W. von Eiff, H. Ising, G. Jansen, R. Martin, B. Rohrmann, W. Schönpflug und M. Spreng: Interdisziplinärer Arbeitskreis für Lärmwirkungsfragen beim Umweltbundesamt in Berlin. Die Beeinträchtigung der Kommunikation durch Lärm. *Zeitschrift für Lärmbekämpfung*, 32:95–99, 1985.
- J.W. von Goethe. Hör-, Schreib- und Druckfehler. In *Goethe's nachgelassene Werke. Fünfter Band. II. Deutsche Literatur*, Seiten 156–162. Stuttgart und Tübingen, in der J. F. Cotta'schen Buchhandlung, 1833.
- H. von Helmholtz. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Verlag von Friedrich Vieweg und Sohn, 1863.
- H. von Wedel. Reichen die heutigen verfügbaren sprachaudiometrischen Verfahren zur Hörgeräteanpassung? *Audiologische Akustik*, 23:66 – 76, 102 – 119, 1984.
- H. von Wedel. Suitability of speech audiometry study procedures for current demands in clinical and general practice. *HNO*, 34(2):71–74, Feb 1986.
- E. Weiland. *Neue Wörtertестe für Sprachaudiometrie und über damit angestellte Versuche an 96 Normalhörenden*. Dissertation, Universität Freiburg i. Br., 1954.
- J. Wiehe. *Lärmbedingte Verdeckungseffekte- eine Unfallursache?* Technische Universität Bergakademie Freiberg, 2000.
- E. Willinger. *Wahrnehmung synthetischer Laute durch Cochlea-Implantat-Patienten*. Dissertation, Bayerische Julius- Maximilians- Universität zu Würzburg, 2005.
- K. Wömpner. *Sprachverstehen im Störlärm mit dem HSM-Satztest mit digitalen Hörgeräten*. Dissertation, Bayerische Julius-Maximilian-Universität zu Würzburg, 2003.



## **B. Probandenaufklärung**

### **B.1. Informationen zu den Versuchsdurchführungen**

Teil 1: Vor dem eigentlichen Versuchsbeginn werden wir einen Hörtest, ein sogenanntes Tonschwellen-Audiogramm, mit Ihnen durchführen. Dieser Hörtest dient dazu festzustellen, ob sie normalhörend sind, was eine Voraussetzung für die Teilnahme bei diesem Versuch ist. Ob Ihnen das Versuchsergebnis mitgeteilt werden soll oder nicht, obliegt ganz allein ihrer Entscheidung. [1ex] Teil 2: In diesem Test soll die Verstehbarkeit von einsilbigen Wörtern getestet werden. Damit Verstehbarkeitsunterschiede messbar werden, wurde die Schwierigkeit durch hinzugefügtes Störrauschen künstlich erhöht. Die Einsilber werden ihnen aus einem Lautsprecher direkt vor Ihnen vorgespielt. Sie werden jedes Wort nur einmal hören und sollen dann angeben, was sie verstanden haben, auch wenn es nur ein Teil des Wortes war. Die Reihenfolge der Einsilber ist willkürlich durchmischt und wird weder durch alphabetische, noch durch inhaltliche Aspekte bestimmt. Es werden ihnen zehn Blöcke mit jeweils etwa 160 Einsilbern vorgespielt, zwischen denen wir Pausen machen. Für den Versuch sind insgesamt etwa vier Stunden angesetzt. Bitte versuchen Sie beim Hören der Wörter aufmerksam zu bleiben. Wenn gewünscht, können wir jederzeit zwischen den Blöcken längere Pausen machen.

## **B.2. Einverständniserklärung**

Sehr geehrte/r Frau/Herr ..... Wir bedanken uns für ihre Bereitschaft, heute an unserem Hörtest teilzunehmen. Vorab möchten wir sie noch auf einige Punkte aufmerksam machen und sie um ihr schriftliches Einverständnis bitten.

- Die Tests sind freiwillig und können jederzeit von Ihnen abgebrochen werden.
- Im Falle eines Testabbruchs Ihrerseits werden die bis dahin aufgezeichneten Daten verworfen.
- Es handelt sich bei diesen Hörtests um rein wissenschaftliche Untersuchungen.
- Um sicherzustellen, dass sie normalhörend sind, was für diesen Versuch Voraussetzung ist, führen wir vor diesem
- Versuch ein Tonschwellen–Audiogramm durch. Es obliegt ihrer Entscheidung, ob wir ihnen das Ergebnis mitteilen oder nicht.
- Die erhobenen Daten werden ausschließlich zu wissenschaftlichen Zwecken und in anonymisierter Form verwendet.
- Die Hörtests finden in einem schallgedämpften Raum statt.
- Die Schallsignale werden von einem Lautsprecher aus abgegeben, der sich vor ihnen befindet.
- Die Schallsignale werden nur bis zu einem maximalen Pegel von 70 dB angeboten (normale Sprechlautstärke) und überschreiten nie die Dauer von 10 Sekunden. Eine Schädigung des Gehörs ist daher nicht möglich. Trotzdem können wir nicht mit Sicherheit ausschließen, dass sie nicht doch ein Schallsignal als unangenehm laut oder erschreckend empfinden könnten.

Ich habe die oben genannten Punkte gelesen und zur Kenntnis genommen und erkläre mich mit der Durchführung der Hörtests einverstanden.

Würzburg,..... Würzburg,.....  
Unterschrift Proband    Unterschrift HNO-Klinik

## C. Listen

Auf den folgenden Seiten (Tabellen [C.1](#) bis [C.5](#)) ist die 166. Listenzusammenstellung im Detail aufgeführt. Es handelt sich um diejenige Zusammenstellung, die die geringste Streuung der mittleren Listenschwierigkeitsgrade ergeben hat. Eine genauere Beschreibung befindet sich in Abschnitt [3.2](#) ([Verteilung der Wörter in Listen](#), S. [79](#)).

Tabella C.1.: Übersichtstabelle der entstandenen Listen Teil 1

Knauf	acht	stur	Foehn	lang	Sprung	Schlacht	Not	Tor	Durst	spitz	Talk
nein	per	Rand	stumm	Spiss	Stein	Brunch_e	Netz	Bart	Reh	Board_e	Lift
Moll	Kohl	Mund	Marsch	aus	grau	fies	bunt	Bann	Pep	Sand	Form
weg	Haft	greis	Fleisz	Mai	Darm	nun	blau	Reis	Pfund	Krach	Flaum
Wirt	Schwert	Trunk	als	Fakt	santf	frisch	bar	Busch	Kitz	Kult	Typ
Wort	Obst	Krampf	Frack	Klaus	dir	ins	Schreck	krumm	Tanz	Kai	Pelz
Zoll	Bauch	Heft	leer	Troll	Zeh	Klotz	roh	Fleck	Land	lieb	Trott
den	Papst	Hans	nebst	Krieg	Duft	Sack	Star_e	Graus	zur	Nil	Herbst
Fels	Ding	flach	Jazz_e	das	Schar	Kluft	Kraut	Ei	Schliff	Wand	Dorf
Schlauch	Jeep	trueb	oft	klug	halt	Bad	Kind	Stock	Forst	gut	Tisch
Markt	Brief	glatt	Ski	echt	Knecht	dreist	loll	Stand	Latz	neu	Gischt
Spross	wrack	weil	so	Fusz	Flausch	Schwamm	gruen	Spray_e	Horn	Moench	Hahn
Eis	Strick	Strand	kein	schon	Schein	Freak_e	Grill	Freak	rau	Ruhm	krass
Pflug	uns	Terz	Huhn	Hast	Strip	grosz	siebt	Matsch	Trend	sich	Rom
Span	hoch	Klecks	Wuchs	Show	Zimt	Schlimpf	Band_d	Brett	Kost	zwecks	Klamm
Gips	bleich	Jacht	Ziel	Halm	vom	stramm	Strauch	laengs	Haut	Luft	schlimm
doch	Mond	Spray_day	Wind	nachts	Putsch	Zwei	leicht	Berg	Schrank	Dach	Schatz
Zoo	Pfeil	sein	Raps	schick	Schluss	Chlor	Pech	hier	ihr	Wahl	Rahm
Reif	weisz	dich	Pasch	Tuer	Sturm	Frost	Schleim	Bucht	Meer	duerr	Huf
wen	Fax	Brauch	Ur	Schelm	Brand	Moos	Fall	Spiesz	Teint	Dreck	Lachs

C. Listen

Tabelle C.2.: Übersichtstabelle der entstandenen Listen Teil 2

zum	Laus	Schall	Wut	Steg	Dachs	Vier	schier	Streit	Tipp	Trab	faul
taub	fast	Flur	Bug	Norm	wahr	Schlumpf	tot	Haar	Plan	Bein	Prinz
nackt	Schrift	boes	Eins	gleich	Lob	Schwarm	Slip	sacht	Jeans	Stuhl	Koks
Hirsch	Tritt	Fell	wo	Scharm	Stoff	Satz	Zehn	Ist	Star_d	Glied	Hirt
Trakt	Squash	Kauf	Katz	Rutsch	Vers	Snob	dumm	Pirsch	Schicht	Pest	es
forsch	Sucht	Kran	Start	Skat	Zink	Sarg	Coach	Kies	Beet	Scheck	Helm
Bus	Schmalz	sehr	Rumpf	Zwoelf	Arzt	Park	Hund	Vieh	Toast_d	Speck	Tweed_e
manch	Brom	Au	wie	Horst	Fan	Ruf	Pracht	Hort	Drang	statt	Tal
Punsch	grob	Stress	Neun	schwer	Wolf	mir	Faust	Schwur	Streu	Boy	Dreh
Schlaf	kraus	rechts	feucht	Ulm	Stift	Schmutz	Wurst	frueh	Hecht	Deck	Knack
Genf	Witz	Wacht	man	Gaul	Team	fuers	Slum	Zins	Top	Haus	Flut
steif	Wels	Specht	Trotz	derb	Weib	Hut	Wurf	Schlund	um	Tag	Senf
Schrot	Nerv	Korn	Pflock	laengst	braun	frech	Frist	Mob	dort	fuer	wund
Weg	Funk_d	Moor	Baer	Scotch	nett	kalt	Watt	Furcht	duenn	keck	Tour
Nest	Fest	Keim	fett	Klatsch	Mais	der	Tropf	Zeit	Stuck	schlau	rein
Streif	Takt	schrill	Feld	Pflicht	Gras	Bau	Schmerz	hin	Leim	nur	neunt
Chat	Raub	Eich	breit	auf	Stahl	Funk_e	Hof	fix	Pferd	Zorn	Blatt
Zwang	Stoer	Wert	matt	Maerz	Saar	Clown	froh	Pub	Geist	Zwerg	Akt
Masz	heim	schnell	Staub	voll	link	List	Gans	Hengst	Trip	nichts	Knopf
Kot	Fracht	Puls	Weih	Pult	hoechst	Lot	stets	Putz	Zier	schwuel	Freund

Tabella C.3.: Übersichtstabelle der entstandenen Listen Teil 3

Bett	Strolch	Bord	Dill	ich	Smog	Kuh	Ruck	Bund	Rost	im	Malz
Schliff	Heer	Spot	Wien	blond	Lack	Knicks	des	Job	schraeg	ab	Kampf
flott	schlapp	just	Kork	Post	Hub	Zahn	dick	Snack	Mohn	Ort	Sechs
Pakt	flink	Tee	Steak_ds	wann	sproed	Zaun	Po	Ernst	Prunk	Gicht	Topf
Steak_e	Kranz	Stroh	Ball	ganz	heut	blass	Traum	Match	Schlag	Couch	euch
Pop	Burg	Los	schwarz	pink	Last	er	Held	drall	eins	zaeh	knapp
Art_e	Schmied	Brust	alt	Christ	Jazz	Turm	Hack	Graf	krank	Zaum	Angst
Dank	Sinn	Soll	Schaft	Fuchs	rank	Talg	Chor	Freak_d	Lahn	Toast_e	Laerm
See	an	Muell	Punkt	Amt	Bier	Rind	Bub	Boot	Grund	Spazz	Bob
Schweisz	Lauch	Schub	doof	bei	Spind	suesz	zu	Mast	Wurm	leck	kahl
dumppf	Keks	Mist	fromm	Schrott	Schwanz	Eid	Schuft	Brot	Hals	Ohm	Test
Pfau	Pimpf	Stil	Strang	jaeh	Stich	Lamm	Spur	durch	Schiff	Aal	Stretch
Schnitt	Stier	Gas	schmal	feig	Mut	Wohl	fair	Pacht	Teer	Blick	Schopf
Schnaps	Spalt	grell	Fund	Fach	und	Rum	Schluck	ihn	Pin	rot	Leib
Wall	Air	Pfalz	Fink	Knie	Wucht	Milch	frei	Laich	Sau	Werk	keusch
Null	Maus	tief	Frucht	noch	Sporn	Kahn	Holz	Nord	Zwist	Jet	Dom
Feind	zehnt	Gag	Schock	nass	Ritt	wild	Dienst	Rhoen	Trug	Shrimp_d	Nuss
Fahrt	Lust	Herr	mein	Quark	Stern	dies	herb	Kuss	Schwein	Schlitz	Lurch
Quatsch	Pfui	Kropf	weich	Platz	Tusch	Ost	Zweig	Pfiff	Spruch	Tat	Sekt
Psalm	zart	Schrein	Volk	Druck	Stosz	je	Gott	hell	Kamm	Tracht	Zweck

C. Listen



Tabelle C.4.: Übersichtstabelle der entstandenen Listen Teil 4

mehr	Stiepp	Geld	schief	Koeln	Treff	Wink	Punk	prompt	Klee	Rein	Zug
Bild	Lauf	Magd	fort	Fisch	Fluch	Drink_d	Floh	Hanf	Dampf	Draht	Storch
vor	Blitz	Pack	Film	Schnee	bis	am	wach	wer	Welt	Napf	Schott
Main	weh	Sieg	Flair	Deich	in	Macht	Tank	Griff	Zinn	harsch	Rausch
Schmelz	blink	Reich	Naht	hohl	Trieb	Heil	scharf	Schaf	deutsch	Schweif	Koelsch
Reck	Schrieb	Zopf	schmuck	mich	vorn	Schach	Sitz	jetzt	Streik	gern	nie
Kerl	Gleis	Salz	Grusz	Seil	Schritt	Part	pur	Spann	Swing	Schloss	Frau
Spuk	Schuh	Wicht	Nacht	Zunft	Laub	Loch	selbst	Block	klein	Recht	Jazz_d
Ohr	heisz	Schutt	Drill	stumpf	Pol	Klosz	schlecht	Schuss	gar	falsch	Zahl
wir	nach	Sport	links	Trog	Hass	Shorts	Text	Gel	Sohn	Axt	Bonn
Schlucht	Dieb	Farm	ja	Dolch	spaat	Twist	Schrei	Mal	Ochs	Qualm	Stab
satt	dann	Schacht	Crew	Pfand	rund	ob	Wachs	Golf	Preis	Strich	dritt
schlank	Boss	Ton	Rest	Kopf	Bach	nicht	schroff	Tausch	Trier	Puck	Kunst
Mord	Wok	Klub	Lehm	Uhr	Chef	blind	Max	Schnur	Drei	Mensch	Quell
Hirn	Bruch	Stadt	lahm	Riff	Wehr	Brut	Tran	sonst	Strumpf	du	Gast
Flop	naechst	Pfad	Fuerst	fit	Mops	viel	Stopp	Look	First	Alm	Strom
Teich	kuehl	Ass	Neid	mies	Schlips	Frust	Alp	Schmant	blank	Rast	Sicht
Korb	jung	Rad	Trost	von	Blues	Kalb	plus	West	Staat	Leid	sechst
Milz	meist	Harz	Glueck	Flosz	da	ein	Gier	schwach	Zank	Spree	Schirm
Kreis	Fang	Sumpf	Krebs	Stau	was	Herd	Beil	Wunsch	Tod	Pluesch	Schenk

Tabella C.5.: Übersichtstabelle der entstandenen Listen Teil 5

Riss	rasch	Hemd	Ring	Dritt	Cent	Pfahl	bald	Kur
blasz	ihm	Camp_e	Schaum	Wein	Fluss	Thron	Harn	Biest
Stamm	Prag	Wald	Haupt	Braut	Bahn	Schank	Track_e	Bluff
Biss	fern	Bock	mint	Sued	Groll	strikt	Klon	Schutz
Jahr	Arsch	Licht	Schweiz	Mehl	tells	Hang	Kap	kurz
Flucht	Gunst	Balz	mit	Schwund	Rap_d	Strahl	Stiel	Schlamm
Quarz	Schosz	Zwirn	Gruft	Steisz	Kleid	Zucht	Ross	frank
Sog	Speer	Board	erst	Schwenk	Krug	Gurt	Hieb	Gin
Torf	Schwan	Tau	seit	Tuch	rings	sie	Spatz	Dock
Trick	Mars	Fee	Zeus	Dorn	plump	warm	die	Ritz
Saat	wenn	Sud	Schild	Braeu	light	Koch	Gift	Oel
weit	Dunst	Letzt	Elf	straff	Sieb	gelb	Griesz	Schmaus
Hauch	Schwall	schlaff	Verb	Stolz	Flip	Ast	Fass	samt
Kuehn	Blech	Frosch	Eck	Soul_d	Werft	wuest	Aids	huebsch
Glanz	Gold	Docht	Fleisch	falls	Kniff	Scherz	Knick	laut
dicht	Schuld	Kiel	Knall	Lohn	Mett	wem	Trumpf	Filz
bloed	Zeit	Herz	Band_e	Maul	Kelch	fein	Tramp_e	Heck
Dings	Zeug	Quiz	Kinn	brav	Stueck	Heu	denn	Glut
Spot	Strausz	Kitsch	rar	still	Saft	Skript	Reim	Narr
schoen	schal	Kuer	Flug	Sturz	Tier	Rauch	Kreuz	kraft

## D. Unbekannte Wörter

Proband 1 gänzlich unbekannte Wörter	Kwass, Fjaell, Flom, Pfriem, Spelz, Faun, Triel, Sterz, Ster, Phot, Warp, Etsch, Straps, Twill, Schoerl, Ziest, Spelt, Kar, Flinz, Tand, Waadt, Percht, Floez, Kif, Aast, Bai, Gneis, Lotz, Lolch, Sass, Wams, Bilch, Pulp, Phlox, Zent, Trutz, Lein, Prau, Dost, Borg, Frigg, Seim, Zwilch, Tort
Wortlaut bekannt, aber Bedeutung unbekannt	Schwyz, Treck, Waag, Flor, Ra, Styx, Gams, Priel, Speik, Ried, Klips, Baal, Nuuk, Geest, Salm, Piz
Proband 2 gänzlich unbekannte Wörter	Kwass, Krux, Fjaell, Lodsch, Brunft, Hehl, Spelz, Aar, Rick, Warp, Soest, bruesk, Styx, Pud, zag, Flett, Brigg, Crew_e, Rap_e, Brunch_dsch, Spray_de, Kar, Hohn, Huld, Gams, Nerz, Olm, Pulk, Waadt, Percht, Noeck, Baal, Enns, Gneis, Nuuk, Wams, Plebs, flugs, patt, Phlox, Kauz, Balg, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Kiez, Schwyz, Spund, Gracht, hehr, Dung, Tramp_d, Tang, Dorsch, Kilt, Phot, bleu, Butt, jein, Etsch, seicht, siech, kess, Spleen_d, Frasz, Spelt, Sphinx, Speil, Krim, Tross, Flinz, Priel, Brunch_dtsch, elft, Tand, Pein, Lehn, Borschtsch, Gent, Fron, Steak_dsch, Strass, Prahm, Belt, Fond, Pan, Gnom, gen, gram, Geest, Jak
Proband 3 gänzlich unbekannte Wörter	Kwass, Sims, Schwyz, Erz, Fjaell, Flachs, Lodsch, Flom, Brunft, Triel, Aar, Flansch, Warp, Soest, Frett, Bries, Styx, Gilb, Schoerl, Flett, Porst, Olm, Geisz, Tand, Ried, Waadt, Borschtsch, Noeck, Gnu, Enns, Nuuk, Lolch, muerb, Wams, Plebs, Aas, Phlox, Tweed_d, Quent, Frigg, Zwilch

#### D. Unbekannte Wörter

Wortlaut bekannt, aber Bedeutung unbekannt	Krapp, Strunk, Spreu, Schur, Krux, Spund, Gracht, Pfriem, wirr, hehr, falb, Dung, Spelz, Fjord, Kilt, Brunst, Kulm, bruesk, Twill, Lump, Krill, kess, Hatz, Kar, Sund, Huld, Speil, Grind, Stint, Percht, Drusch, Gau, Gneis, Schlick, Prahm, Bilch, Holm, Trutz, Trupp, Saus, Dost, Piz
Proband 4 gänzlich un- bekannte Wörter	Kwass, Flom, Pfriem, falb, Rips, Triel, Sterz, Ster, Stunt, Phot, Etsch, Kulm, Styx, Pud, Twill, Schoerl, Flett, Porst, Spelt, Kar, Huld, Speil, Flinz, Waadt, Percht, Drusch, Kif, Baal, Enns, Bai, Lotz, Prahm, Mink, Lolch, Sass, Kral, Frey, Phlox, Salm, Prau, Dost, Frigg, Piz, Seim, Tort
Wortlaut bekannt, aber Bedeutung unbekannt	Kien, Sext, hehr, zag, Quast, Brigg, Ziest, Olm, Floez, Graz, Noeck, Bilch, Phon, Priem, Quent, Zwilch
Proband 5 gänzlich un- bekannte Wörter	Krapp, Kwass, Met, Schwyz, Spreu, Kien, Schur, Krux, Flachs, Mull, Flom, Pfriem, Brunft, falb, Dung, Waag, Spelz, Faun, Schmach, Phot, Flansch, Warp, Etsch, Kulm, Styx, Skalp, Straps, Stent, Gilb, Schoerl, Quast, Brigg, Ziest, Beiz, Hatz, Dur, Porst, Spelt, Hohn, Quart, Speil, Gams, Krim, Furt, Tand, Waadt, Percht, Borschtsch, Drusch, Gent, Fron, Pfuhl, Noeck, Baal, Gnu, Enns, Maat, Prahm, Mink, Nuuk, Wams, Tschad, Plebs, flugs, patt, Phlox, Pan, Worms, Trutz, gram, Jak, Priem, Prau, Scheit, Dost, Borg, Pomp, Lug, Piz
Wortlaut bekannt, aber Bedeutung unbekannt	Rips, kess, Sund, Tross, Vamp, Speik, Schalk, Gral, Lotz, Lolch, Sol, Frey, Lein, Salm, Frigg, Seim, Zwilch
Proband 6 gänzlich un- bekannte Wörter	Kwass, Schwyz, Flom, Phot, Styx, Schoerl, Kar, Flinz, Olm, Waadt, Nuuk, Lolch, Geest
Wortlaut bekannt, aber Bedeutung unbekannt	Krapp, Fjaell, Ska, Spelt, Tand, Frigg, Zwilch

Proband 7 gänzlich unbekannte Wörter	Krapp, Kwass, Strunk, Hel, Wust, Kien, Hall, Fjaell, Flom, Pfriem, hehr, falb, Spelz, Sterz, Ska, Schlepp, Ster, Ra, Phot, Butt, Trumm, Warp, Frett, Etsch, Pud, Twill, Schoerl, Flett, Fries, Hatz, Krem, Speil, Flinz, Speik, Stint, Waadt, Floez, Naab, Aast, Noeck, Baal, Lotz, Haff, Prahm, Mink, Nuuk, Phlox, Geest, Salm, Seim, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Maer, Krill, Porst, Sund, Probst, Gnu, Enns, Frey, Quent
Proband 8 gänzlich unbekannte Wörter	Kwass, Hel, Schwyz, Fjaell, Lodsch, Flom, falb, Spelz, Triel, Phot, Garn, Warp, Frett, Styx, Pud, Stent, Spelt, Sund, Speil, Flinz, Stint, Waadt, Drusch, Floez, Noeck, Baal, Gneis, Lotz, Prahm, Mink, Nuuk, Lolch, Bilch, Plebs, Phlox, Salm, Prau, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Krapp, Maas, firm, Fizz, hehr, Sterz, Aar, Butt, jein, Soest, Etsch, Kulm, Schoerl, Krill, Porst, Kar, Priel, Kif, Aast, Bai, Haff, Tschad, Frey, Trutz, Geest, Jak, Dost, Frigg, Piz
Proband 9 gänzlich unbekannte Wörter	Borschtsch, Phlox, Jak
Wortlaut bekannt, aber Bedeutung unbekannt	Krapp, Kwass, Kien, zwoelft, Krux, Lodsch, Flom, Fizz, Pfriem, hehr, Rips, Triel, Patsch, Sterz, Ster, Ra, Frett, Etsch, Kulm, Styx, Pud, Schoerl, Quast, Flett, Fries, Hatz, Spelt, Kar, Sund, Gams, Priel, Blend, elft, Stint, Tand, Ried, Waadt, Percht, Drusch, Floez, Naab, Kif, Aast, Noeck, Maat, Bai, Riet, stet, Lotz, Haff, Prahm, Nuuk, Bilch, Pulp, Frey, Trutz, Geest, Quent, viert_i, Dost, Borg, Lug, Frigg, Zwilch, Tort
Proband 10 gänzlich unbekannte Wörter	Krapp, Kwass, Schwyz, Spreu, Treck, Ulk, Krux, Fjaell, Gracht, Flom, Minsk, falb, Sterz, Rusz, Phot, Fug, Flansch, Warp, Etsch, Brigg, Ziest, Hatz, Kar, Tun, Tand, Floez, Gent, Aast, Lotz, stracks, Phlox, Trutz, Vogt, Piz, Seim, Zwilch, Tort, Teil

#### D. Unbekannte Wörter

Wortlaut bekannt, aber Bedeutung unbekannt	Maer, Sold, Dung, Lutz, Raft, Braus, Flor, Ra, Pik, Rick, Poch, Styx, Gilb, Twill, Schurz, Harm, Huld, Lehn, Waadt, Krepp, Quirl, Salm, Prau, Dost
Proband 11 gänzlich un- bekannte Wörter	Krapp, Kwass, Hel, Schwyz, Schur, Flom, Spelz, Triel, Raft, Ska, Phot, Flansch, Warp, Etsch, Styx, Twill, Schoerl, Flett, Fries, Porst, Spelt, Kar, Sund, Speil, Flinz, Stint, Tand, Ried, Waadt, Percht, Drusch, Floez, Noeck, Baal, Enns, Bai, Gneis, Lotz, Prahm, Mink, Nuuk, Lolch, Bilch, Phlox, Kir, Geest, Tuell, Priem, Prau, Dost, Frigg, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Bims, Fjaell, Spund, Fizz, Pfriem, Dung, Faun, Petz, Sterz, Aar, Frett, Kulm, Schurz, Brigg, Oehr, Bor, Huld, Priel, Rain, Speik, Kif, Probst, Haff, Tschad, patt, Quent, Piz
Proband 12 gänzlich un- bekannte Wörter	Molch, Krapp, Maer, Kwass, Spreu, Kien, Lodsch, Flom, Pfriem, falb, Waag, Spelz, Triel, Petz, Sterz, Flansch, Trumm, Frett, Etsch, Bries, Styx, Pud, Gilb, Twill, Schoerl, Krill, Flett, Brigg, Ziest, Porst, Spelt, Krem, Kar, Sund, Speil, Flinz, Rain, Speik, Tand, Waadt, Percht, Borschtsch, Drusch, Naab, Kif, Aast, Noeck, Maat, Bai, Lotz, Prahm, Nuuk, Lolch, Belt, Sass, Bilch, Plebs, Kral, Phlox, Zent, Trutz, Geest, Prau, Qent, Dost, Borg, Frigg, Piz, Tort
Wortlaut bekannt, aber Bedeutung unbekannt	Fjaell, hehr, Klan, Raft, Warp, Poch, quitt, Kulm, Hain, siech, Schurz, Quast, Fries, Falz, Furt, Stint, Ried, Floez, Fron, Riet, Greif, Frey, Tuell, Salm, Jak, Priem
Proband 13 gänzlich un- bekannte Wörter	Krapp, Kwass, Kien, Flom, falb, Boom_e, Waag, Spelz, Triel, Sterz, Flansch, Klipp, Rick, Pud, Gilb, Twill, Schoerl, Brigg, Speil, Flinz, Priel, Speik, Boom_d, Waadt, Percht, Borschtsch, Floez, Kif, Aast, Noeck, Baal, Bai, Fakts, Prahm, Mink, Nuuk, Bilch, Phlox, Geest, Prau, Ruch, Dost, Borg, Frigg, Piz, Seim, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Hel, Met, Fjaell, Lodsch, Rips, Faun, Tramp_d, Ster, Pferch, Phot, Kulm, Styx, Schurz, Quast, Porst, Grind, Stint, Schlich, Pfuhl, Lotz, Sass, Wams, Kral, Pulp, Frey, Quent, Tort

Proband 14 gänzlich unbekannte Wörter	Krapp, Maas, Maer, Kwass, Kien, Fjaell, Lodsch, Flom, falb, Spelz, Phot, Aar, Trumm, Etsch, Styx, Pud, Schoerl, Schurz, Hatz, Porst, Spelt, Krem, Speil, Flinz, Stint, feil, Tand, Waadt, Borschtsch, Floez, Aast, Noeck, Baal, Bai, Riet, Gneis, Haff, Prahm, Lolch, zig, Wams, Bilch, Plebs, Kral, Phlox, Stank, Zent, Trutz, Geest, Dost, Frigg, Seim
Wortlaut bekannt, aber Bedeutung unbekannt	Raft, Brunst, bleu, Fug, Flansch, Butt, Frett, Kulm, Twill, Flett, Ziest, Fries, Sund, Huld, Priel, Olm, Lehn, Gent, Fron, Lotz, Nuuk, Sass, Salm, Borg, Lug, Zwilch, Taft
Proband 15 gänzlich unbekannte Wörter	Kwass, Strunk, Hel, Pfriem, hehr, Spelz, Triel, Phot, bleu, Poch, Pud, Twill, Schoerl, Krill, Flett, Speil, Percht, Aast, Noeck, Bai, stet, Lotz, Prahm, Nuuk, Lolch, Bilch, Phlox, Geest, Seim
Wortlaut bekannt, aber Bedeutung unbekannt	Kien, Flom, falb, Rips, Tramp_d, Flor, Butt, Etsch, Bries, Gilb, Ziest, Stint, Drusch, Dutt, Kif, Probst, Fakts, lind, Mink, Sass, Plebs, Frey, Kir, Tuell, Frigg
Proband 16 gänzlich unbekannte Wörter	Krapp, Kwass, Strunk, Schwyz, Fizz, falb, Spelz, Ra, Phot, Flansch, Etsch, Pud, Schoerl, Brigg, Spelt, Flint, Flinz, Tand, Waadt, Percht, Borschtsch, Drusch, Naab, Kif, Noeck, Enns, Bai, Haff, Prahm, Mink, Nuuk, Bilch, Tschad, Phlox, Geest, Ren, Frigg
Proband 17 gänzlich unbekannte Wörter	Krapp, Kwass, firm, Kien, zwoelft, Trift, Krux, Iod, Flom, Pfriem, Waag, Spelz, Triel, Petz, Sterz, Ska, Flor, Ster, Phot, bleu, Aar, fehl, Fug, Klipp, Trumm, Poch, viert_ Frett, Bries, Kulm, bruesk, Pud, Stent, Gilb, Twill, Hab, Schoerl, Farn, Flett, Ziest, Harm, Spelt, Krem, Speil, Gams, Flinz, Rain, Tand, Waadt, Percht, Borschtsch, Floez, Schlich, Kif, Grog, Aast, Pfuhl, Schoss, Bai, Fakts, Riet, lind, Gneis, stet, Haff, Krepp, zig, Wams, Bilch, Phlox, Stank, Phon, Tuell, viert_i, Dost, Pomp, Lug, Piz, Seim, Zwilch, Tort
Wortlaut bekannt, aber Bedeutung unbekannt	Drink_e, Boom_e

#### D. Unbekannte Wörter

Proband 18 gänzlich unbekannte Wörter	Kwass, Flom, Phot, Bries, Pud, Schoerl, Ziest, Oehr, Krem, Flinz, Speik, Waadt, Drusch, Naab, Stick, Kif, Noeck, Baal, Fakts, Lotz, Prahm, Nuuk, Lolch, Frey, Phlox, Phon, Jak, Prau, Dost, Frigg, Piz, Zwilch
Wortlaut bekannt, aber Bedeutung unbekannt	Spelz, Gams, Ried
Proband 19 gänzlich unbekannte Wörter	Krapp, Kwass, Schwyz, Kien, Trift, Krux, Fjaell, Gracht, Flom, hehr, lasch, Dorsch, Flor, Ra, Pferch, Phot, bleu, Flansch, Rick, Soest, Frett, Bries, Loft, Styx, Pud, Stent, Twill, Schoerl, Krill, Flett, Brigg, Bor, Porst, Spelt, Sund, Mumps, Falz, Flinz, Olm, Speik, Waadt, Percht, Borschtsch, Drusch, Floez, Kif, Aast, Noeck, Strass, Maat, Bai, Fakts, Mink, Nuuk, Bilch, Tschad, Fond, Frey, Phlox, Pan, Zent, Kir, Gramm, Geest, Tuell, Scheit, Ruch, Quent, Dost, Borg, Frigg, Piz, Seim, Tort, Taft
Wortlaut bekannt, aber Bedeutung unbekannt	Maas, Maer, Hel, Lodsch, Pfriem, falb, Waag, Triel, Patsch, Geck, fehl, Bursch, Vlies, Kliff, Schurz, Quast, Schmiss, Spleen_de, Schwulst, Blend, Shrimp_e, Lehn, Fron, Pfuhl, Baal, Gneis, Kral, Sol, Quirl, Pulp, Phon, gen, Salm, Jak, Pfropf, Lid, Pomp
Proband 20 gänzlich unbekannte Wörter	Krapp, Kwass, Hel, Kien, Fjaell, Flom, Dung, Spelz, Triel, Sterz, Ra, Geck, Phot, Pud, zag, Twill, Schoerl, Flett, Porst, Spelt, Krem, Kar, Speil, Krim, Flinz, Olm, Speik, Stint, Tand, Percht, Kif, Noeck, Baal, Enns, Lotz, Haff, Prahm, Mink, Bilch, Kral, Phlox, Zent, Geest, Prau, Dost, Lug, Frigg, Seim
Wortlaut bekannt, aber Bedeutung unbekannt	Maer, Fizz, hehr, falb, Faun, Flor, Trumm, Schurz, Fries, Grind, Ried, Borschtsch, Drusch, Klips, lind, Bausch, Kord, Nuuk, Sass, Frey, Stank, Zwilch, Tort



## E. Danksagung

Herrn Prof. Dr. Rudolf Hagen, Direktor der Klinik und Poliklinik für Hals-, Nasen- und Ohrenkrankheiten, plastische und ästhetische Operationen der Universität Würzburg und Herrn Prof. Dr. Joachim Müller, Oberarzt an der HNO-Klinik der Universität Würzburg, bin ich für die Überlassung des Themas dieser Arbeit zu besonderem Dank verpflichtet. Auch für die Übernahme des Referates möchte ich mich bei Herrn Prof. Joachim Müller herzlich bedanken.

Frau Prof. Dr. rer. nat. Kathleen Wermke, Leiterin des Zentrums für vorsprachliche Entwicklung und Entwicklungsstörungen in Würzburg, danke ich für die Übernahme des Koreferats.

Zu ganz besonderem Dank bin ich Herrn Dipl.-Ing. Stefan Brill verpflichtet, der das Dissertationsthema vergeben und die Begleitung in allen Phasen der Arbeit übernommen hat. Besonders dankbar bin ich ihm für sein außergewöhnliches Engagement. Nur mit seiner Hilfe war es mir möglich, die technischen Schwierigkeiten bei der Vorbereitung der Messungen am Probanden zu bewältigen, die umfangreichen Ergebnisse auszuwerten und zu interpretieren, sowie das Programm zu entwickeln, das die Verteilung der Wörter in die die Testlisten vornahm. Allein durch seine Bereitschaft, so ein hohes Maß an Zeit an diversen Freitagabenden zu investieren, war es mir möglich, die Dissertation auch von Frankfurt aus abzuschließen.

Allen Mitarbeitern der Abteilung für Cochlea-Implantate sowie den Mitarbeitern der Audiometrieabteilung danke ich für die Unterstützung.

Besonders dankbar bin ich allen Probanden, die viele Stunden investierten und ein hohes Maß an Geduld während der Versuche bewiesen haben.

Zu guter Letzt möchte ich meiner Familie danken, die mir das Studium ermöglicht und mich zu der Arbeit an dieser Dissertation stets ermutigt hat. Besonders meinem Vater bin ich für die aktive Unterstützung und die konstruktive Kritik besonders dankbar.

