

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

Direktor: Prof. Dr. med. Rudolf Hagen

**Entwurf und Evaluation
eines Prosodieverständnistests
in deutscher Pseudosprache**

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

Josefine Hirschfeld

aus Leipzig

Würzburg, April 2011

Referent: Prof. Dr. med. Joachim Müller
Korreferent: Prof. Dr. med. dent. Bernd Klaiber
Dekan: Prof. Dr. med. Matthias Frosch

Tag der mündlichen Prüfung: 28. Oktober 2011

Die Promovendin ist Zahnärztin

Gewidmet meinen Eltern, die mir mein Studium ermöglicht haben.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Deutsche Pseudosprache	2
1.2	Prosodie und emotionale Kommunikation	3
1.2.1	Grundlagen der Prosodie	3
1.2.2	Bedeutung der Prosodie im Alltag	4
1.2.3	Grundlagen der Sprachrezeption	7
1.3	Emotionen	9
1.3.1	Emotionstheorien	10
2	Methoden und Versuchsplanung	13
2.1	Erzeugung deutscher Pseudosprache	13
2.1.1	Eigenschaften der Originaltexte	14
2.1.2	Software „hmmtext“	16
2.2	Auswahl der Sätze und Erstellung einer Satzliste	29
2.3	Auswahl der Emotionen	30
2.4	Durchführung der Aufsprache	32
2.4.1	Räumlicher und technischer Aufbau	33
2.5	Schneiden der Sätze und Skalierung des SPL	34
2.6	Auswahl der gesprochenen Sätze und Nachaufsprache	36
2.7	Pilottest	37
2.8	Erstellung der Abspiellisten und Audiodateien	39

2.9	Probandenauswahl	41
2.10	Prosodietest	42
2.10.1	Aufstellung und Gerätschaften	42
2.10.2	Tonaudiogramm	42
2.10.3	Durchführung des Prosodietests	43
2.10.4	Erfassung der Evaluationsdaten	44
3	Ergebnisse	45
3.1	Pilottest: Reduktion auf 10 Emotionen	45
3.2	Offener Test	47
3.3	Geschlossener Test	51
4	Diskussion	63
4.1	Erfolge der Untersuchungen	63
4.2	Fehlerquellen	63
4.2.1	Einfluss der Versuchsplanung	64
4.2.2	Einfluss des Sprechers	65
4.2.3	Technische Einflüsse	66
4.2.4	Einfluss der Probandenauswahl	68
4.3	Urteilsverhalten der Probanden	70
5	Zusammenfassung	72
	Anhang	75
A	Statistische Eigenschaften deutscher Schriftsprache	76
A.1	Folgehäufigkeiten auf c, q, r und w	76
A.2	Ergebnisse bei steigender Synthesetiefe	78
B	Satzliste der Sätze in deutscher Pseudosprache	81

C	Probandeninstruktion	83
D	Messergebnisse der Tonaudiometrie	86
E	Zehn meistverwendete Probandenantworten im offenen Test	87
F	Einzelresultate der Probanden	92
G	Resultate aufgeschlüsselt nach Sätzen	96
	Literaturverzeichnis	102
	Danksagung	
	Lebenslauf	

1 Einleitung

Wenn wir mit unserem Gegenüber kommunizieren oder einfach bei einem Gespräch zuhören, haben wir ein Gefühl für die Stimmung, die währenddessen herrscht. Wir wissen oder ahnen zumindest, mit welcher inneren Einstellung, mit welchen Hintergedanken, mit welchen Emotionen unser Gesprächspartner zu uns spricht, selbst wenn uns die Mimik und Gestik verborgen bleiben, wie z. B. am Telefon. Doch woher kommen diese Informationen, wenn sie nicht gerade in den gesprochenen Worten enthalten sind? Es ist das nonverbale Beiwerk aus Melodie, Rhythmus und Klangfarbe, Prosodie genannt, das uns hilft mehr zu verstehen, als den bloßen Satzinhalt. Die Sprachmelodie und Stimmintonation können verschiedene psychische oder physische Zustände, wie Furcht, Schmerz, Angst, Überraschung, Ironie, Gereiztheit, Gleichgültigkeit, Ärger, Zärtlichkeit, Drohung usw. ausdrücken. Dabei ist jedoch die spezifische Art der Stimmführung interkulturell verschieden und durch die Muttersprache sowie auch dialektal geprägt (Isacenko und Schädlich, 1964). In dieser Dissertation wird die Frage erörtert, inwiefern hörgesunde Menschen, repräsentiert durch eine Gruppe von Probanden, in der Lage sind, einen wichtigen nonverbalen Aspekt der Sprache, nämlich die Prosodie, zu differenzieren und dadurch weitere Schlüsse auf die emotionale Bedeutung des Gesagten daraus zu ziehen. Ziel dieser Untersuchung ist es, zukünftig Hörtests bezüglich dieses Aspekts der gesprochenen Sprache an Cochleaimplantat-Trägern durchzuführen. So kann ermittelt werden, wo technische und rehabilitative Schwachstellen der Implantate selbst und der Nachsorge liegen und verbessert werden können. Diese Dissertation knüpft thema-

1 Einleitung

tisch an die Magisterarbeit von Bauer sowie an die Dissertation von Domes an, die zusammen mit Herrn Dipl.-Ing. Stefan Brill bereits 2005 und 2009 auf diesem Gebiet mehrere Versuchsreihen unternommen haben. Im Unterschied zur Magisterarbeit wurde in der vorliegenden Dissertation nicht mit deutscher Sprache, sondern mit deutscher Pseudosprache gearbeitet, um mögliche Einflüsse des Satzinhaltes auf das Urteilsverhalten des Probanden auszuschließen (s. Abschnitt 1.1, Deutsche Pseudosprache, S. 2). Es geht des Weiteren nicht mehr nur um die bei Bauer (2005) untersuchte Diskriminierungsfähigkeit zwischen Frage, Antwort und Aussage in der Sprachmelodie, sondern vielmehr um die diffizilere Diskriminierung von unterschiedlichen Tonhöhenverläufen und weiteren Parametern akustischer Signale. Geeignete Träger prosodischer Merkmale sind Emotionen, wie beispielsweise Fröhlichkeit, Trauer, Überraschung oder Zorn, die sich in der Stimme des Sprechers widerspiegeln (s. Abschnitt 1.3.1, Emotionstheorien, S. 10 f.).

1.1 Deutsche Pseudosprache

Deutsche Pseudosprache stellt in diesem Zusammenhang eine automatisiert erzeugte, künstliche Sprache in schriftlicher Form dar. Diese Sprache klingt für den Hörer zwar nach deutscher Sprache, bleibt jedoch ohne Sinngehalt und enthält auch kaum tatsächlich existierende Wörter. Die klangliche Ähnlichkeit der Pseudosprache zur deutschen Sprache kommt dadurch zustande, dass sie sich bezüglich der Art und Häufigkeit der Aneinanderreihung von Buchstaben und Silben an der Originalsprache orientiert (s. Abschnitt 2.1.2.3, Eigenschaften der synthetisierten Pseudotexte, S. 24 ff.). Die Semantik einer angewandten Sprache nimmt einen unmittelbaren Einfluss auf die subjektive Bewertung des Gesagten durch den Hörer (Schmidt-Atzert, 1996). Um dies im Hinblick auf die Interpretation des emotionalen Gehaltes zu vermeiden, kam die deutsche Pseudosprache zur Anwendung.

1.2 Prosodie und emotionale Kommunikation

Nach Rickheit und Strohner ist die menschliche Sprache ein kompliziertes Informationssystem, das der Kommunikation und Situationsbewältigung dient. Zwischen den Zeilen zu lesen ist ein wichtiger Bestandteil der sprachlichen Verständigung, denn ohne diese Fähigkeit blieben dem Hörer unterschwellige Botschaften verborgen, die zur beidseitigen Verständigung eine tragende Rolle spielen. *Das Griechische „prosodía“, aus dem sich das Wort Prosodie ableitet, bedeutet soviel wie „das zur Musik gesungene Lied“ oder „gesungene Begleitung“, was nahelegt, dass Prosodie die musikalische Begleitung zu den Worten selbst ist* (Fox, 2000, S. 1).

1.2.1 Grundlagen der Prosodie

Die Prosodie muss somit als ein Teilgebiet der Phonologie betrachtet werden, das ein eigenständiges Signalisierungssystem innerhalb der Sprache darstellt. Soll jedoch die Prosodie analysiert werden, setzt dies die Unabhängigkeit der Prosodie von anderen linguistischen Systemen voraus (Selting, 1995). Die prosodischen Eigenschaften eines lautsprachlichen Ereignisses sind nach Nöth (1991) Tonhöhe, Lautheit, zeitliche Strukturierung, Sprechtempo, Stimmlage, Stimmqualität, Klangfarbe, Rhythmus und auch das Fehlen eines sprachlichen Ereignisses, die Pause. Sprachliche Einheiten umfassen mehr als einen Laut, weshalb die prosodischen Eigenschaften auch als suprasegmentale Merkmale bezeichnet werden. Bei diesen sprachlichen Einheiten kann es sich um Silben, Wörter, Phrasen, Sätze oder Redebeiträge handeln. Aus diesem weiten Feld der prosodischen Merkmale ist nach Auer und Couper-Kuhlen (1994) und Traunmüller (2004) bisher vor allem die Intonation, also die Tonhöhenbewegung bzw. deren Korrelate, Mittelwert und Standardabweichung von der Grundfrequenz F_0 , untersucht worden. Insgesamt lassen sich prosodische Ereignisse auf drei wichtige Parameter zurückführen: Den Tonhöhenver-

1 Einleitung

lauf, die Lautheit und die Dauersteuerung, wobei unter Letzterer die Steuerung der Lautlänge sowie der Sprechpausenlänge zu verstehen ist (Pfister und Kaufmann, 2008; Rickheit et al., 2003). Nach Rickheit et al. (2003) sind diese drei Punkte dem akustischen Aspekt der Prosodie zuzuordnen. Neben diesem besitzt sie aber auch phonologische Eigenschaften, die Sprachmerkmale wie Satzakzent, Akzenttyp und Phrasierung umfassen. Vergleichbar der Musik ist somit auch das metrische und intonatorische Schema eines Satzes bedeutsam und kann in der gesprochenen Sprache Bedeutung über das Wort hinaus tragen. Bereits ein einfaches Beispiel macht dies deutlich: Derselbe Wortstrang, in unterschiedlicher Prosodie ausgesprochen, kann zu einer Frage werden, zu einem Ausruf oder zu einer Aussage (Speer, 1993), vgl. (Bauer, 2005).

1.2.2 Bedeutung der Prosodie im Alltag

Nach Scherer (1982) laufen die Prozesse der Stimmproduktion in ihren physiologischen Strukturen mit einer sehr hohen Komplexität ab, sodass ein Einfluss emotionaler Erregung auf die menschliche Vokalisation kaum vermeidbar ist. Scherer untersuchte im Jahr 1982 die Genauigkeit der Erkennung von Emotionen in Stimm- und Sprachproben. Er erhielt eine durchschnittliche Genauigkeit von 60 %, wobei die Probanden meist zwischen vier und zehn Emotionen auswählen konnten. Diese Treffsicherheit geht über bloße zufällig richtige Beurteilungen weit hinaus. Negative Emotionen wurden in seiner Studie durchgängig leichter erkannt: Ärger am besten, gefolgt von Traurigkeit, Gleichgültigkeit und Freude. Er beobachtete auch, dass in Untersuchungen, in denen eine gefilterte oder fremde Sprache verwendet wurde, der Genauigkeitsgrad zwar etwas reduziert war, die Beurteiler die dargestellten Emotionen aber noch immer mit mehr als Zufallswahrscheinlichkeit erkennen konnten. Williams und Stevens führten im Jahr 1982 ähnliche Versuche durch. Bei ihnen wur-

1.2 Prosodie und emotionale Kommunikation

de die Traurigkeit am besten erkannt, gefolgt von Ärger, Neutral, Freude und Furcht, wobei letztere mit einer Genauigkeit von 27% richtig erkannt wurde. Auch Stirling und Wales (1996) fanden in ihrem Experiment, in dem Personen auf einer Skala die Normalität von intonatorischen Mustern bestimmen sollten, heraus, dass Personen generell empfindlich auf prosodische Muster reagieren und mit einer prototypischen Vorstellung von möglichen prosodischen Mustern arbeiten. Da Banse und Scherer (1996) der Meinung sind, dass es dabei wichtige interindividuelle Unterschiede in der emotionalen Ausdrucksweise gibt, haben sie die emotional geprägten Sätze von möglichst vielen Sprechern festgehalten. Die vokalen Hinweise führten in ihrer Untersuchung dazu, dass Hörer die entsprechenden Emotionen zuverlässig erkannten. Die Autoren merken allerdings auch an, dass aufgrund soziokultureller Zensur der Sprecher bis zu einem gewissen Punkt seinen emotionalen Ausdruck kontrollieren und regulieren kann. Arndt und Janney (1991) bestätigen dies, indem sie unterstreichen, dass die in einer bestimmten Kultur aufwachsenden Menschen Techniken erlernen, um ihre Gefühle und Einstellungen verschiedenen Dingen gegenüber auszudrücken. Dies geschieht durch Erfahrungen mit anderen Menschen in dieser Kultur und lässt sie zu „normalen“ Mitgliedern derselben werden. Sie stellten im Jahr 1991 folgende These auf: *Emotionale Kommunikation ist psychobiologisch erworben und strukturiert, spontan bzw. nicht beabsichtigt expressiv, ist eine automatische, nicht symbolische Manifestation einer inneren affektiven Situation.* Im Hinblick auf das Studium der internationalen Literatur zu diesem Thema gilt es nach Arndt und Janney jedoch zu beachten, dass im englischsprachigen Raum unterschieden wird zwischen „emotional communication“ und „emotive communication“. Letztere bezeichnet die Art der Kommunikation, die durch das soziale Umfeld erlernt wurde. Hier kommt die bewusste, strategische Manipulation affektiver Signale zum Einsatz, um das Verhalten Anderer zu beeinflussen. Riggio erkannte im Jahr 1992, dass Menschen, die

1 Einleitung

gute soziale Interakteure sind, ihr nonverbales Verhalten kontrollieren und sich in verschiedene zwischenmenschliche Situationen einpassen können. Gute Fähigkeiten in nonverbaler Kommunikation sind nach seiner Ansicht eine Schlüsselkomponente für Erfolg im sozialen Leben und mit psychischem Wohlbefinden assoziiert. Sie sind wichtige Determinanten der Qualität und Quantität zwischenmenschlicher Beziehungen. Frick (1985) geht davon aus, dass durch eine persönliche Vertrautheit die Wahrscheinlichkeit des Erkennens emotionaler Ausdrücke erhöht wird, da bis zu einem gewissen Grad jeder Sprecher seine eigenen Muster der emotionalen Kommunikation aufweist. Außerdem nimmt er an, dass diese Kommunikation zusätzlich verbessert wird, wenn die Interaktionspartner sich gegenseitig mögen. Wichtig für die Durchführung unserer Arbeit ist auch die Annahme, dass verschiedene Individuen ähnliche prosodische Sprachmuster bzw. akustische Standardkonfigurationen verwenden, um eine Emotion auszudrücken. Die Versuche von Cosmides (1983) unterstützen diese Hypothese. Der Autor Traunmüller hat im Jahr 2004 für einige Emotionen exemplarisch zusammengetragen, wie diese sich in der deutschen Sprache prosodisch äußern:

1. **Zorn:** Hohe Stimmstärke, besonders an betonten Stellen. Hohe Dynamik der Artikulation und der Intonation, oft überdeutliche Aussprache, oft hohe Sprechgeschwindigkeit,
2. **Furcht:** Hohe Tonlage, oft hohe Sprechgeschwindigkeit, niedrige Dynamik der Artikulation, Reduktion und Assimilation von Segmenten, bei starker Erregung hohe Stimmstärke und Jitter [Jitter: Minimale Schwankung der Grundfrequenz F_0 einer Stimme (Orlikoff und Baken, 1989)],
3. **Betrübnis:** Niedrige Dynamik der Artikulation und der Intonation, Reduktion und Assimilation von Segmenten, niedrige Sprechgeschwindigkeit, niedrige Stimmstärke,

4. **Fröhlichkeit:** Hohe Dynamik der Artikulation und der Intonation, bei Lächeln erhöhte Resonanzlage, weiche Stimme, bei Euphorie auch hohe Stimmstärke und veränderliche Sprechgeschwindigkeit,
5. **Verachtung:** Niedrige Ton- und Resonanzlage, bei Erregung hohe Stimmstärke, eher hohe Dynamik und niedrige Sprechgeschwindigkeit.

Aus dieser Analyse ergibt sich bereits eine Grundlage für die spätere Aufsprache der emotionalen Sätze, an der sich der Sprecher orientieren kann, um die Emotionen möglichst sprachlich authentisch darzustellen.

1.2.3 Grundlagen der Sprachrezeption

Die Klänge der Sprache werden nach Speer (1993) wahrgenommen als ein hierarchisch organisiertes Muster aus Phonemen, Silben, Wörtern und Phrasen mit jeweils eigenen Betonungsmustern. Jeder gesprochene Satz hat schon dadurch eine intrinsische prosodische Struktur, deren Art und Bedeutung im Folgenden kurz besprochen werden sollen. Sprachrezeption bedeutet ganz wesentlich, die aus Sprachlautsequenzen entschlüsselten Informationen in das bereits im Bewusstsein Vorhandene zu integrieren. Der Empfänger verarbeitet den ihn erreichenden Informationsinput immer gleichzeitig und in Zusammenhang mit anderen situationsspezifischen Informationen. Hörer können das sprecherseitig Geäußerte in der Regel nur dann angemessen verstehen, wenn sie weitere, nichtsprachliche Informationen erhalten. Neben den kommunikativen Voraussetzungen gibt es nach Rickheit und Strohner (1993a) aber auch wichtige personenbedingte Einflussgrößen der Sprachrezeption, wie z. B. Wissen, Einstellung und Interessen des Hörers. Der Autor schreibt auch, dass durch die Charakteristika der Kommunikationssituation eine Zielorientierung des sprachlichen und nichtsprachlichen Verhaltens erzeugt wird. Zunächst einmal kann man nach Helfrich (1985) bei der Beschreibung einer sprachlichen Äußerung

1 Einleitung

sechs Ebenen abgrenzen: Eine akustische, eine phonetische, eine phonologische, eine morphologische, eine syntaktische und eine semantische Ebene. Dabei umfasst die phonetische Ebene die einen Laut definierenden Merkmale, die phonologische Ebene beschreibt den Laut als bedeutungsunterscheidende Einheit. Folgende Ebenen und Teilprozesse des Verstehens sprachlicher Äußerungen können laut Strube und Herrmann (2006) unterschieden werden:

1. Wahrnehmung von Lauten und Lautklassen (Phonemen),
2. Wahrnehmung von Wörtern und zusammenhängenden, grammatisch geordneten Wortfolgen,
3. Erkennen von kognitiven Inhalten (Begriffe und ihre Relationen),
4. Erkennen, was der Sprecher beabsichtigt (Sprecherintention),
5. Bewertung des Gesagten.

Strube und Herrmann schreiben, dass diese Teilprozesse sich zeitlich überlappen oder simultan ablaufen. Sie können sich auch gegenseitig korrigieren: Wenn der Hörer beispielsweise zunächst ein Wort wahrgenommen hat, das im Kontext des Satzes betrachtet keinen Sinn ergibt, muss es revidiert werden. Im Bezug auf die Betonung unterscheidet Nöth (1991) zwischen Wort-, Phrasen- und Satzakzent, wobei jeder Satz nur einen Satzakzent, aber mehrere Phrasenakzente enthält. Beim Wortakzent ist es eine Silbe, welche die Betonung trägt. Ein Beispiel soll dies näher erläutern:

„Mein Bruder wohnt in München.“

Der Satzakzent liegt hier auf „München“, ein Phrasenakzent auf „Bruder“ und Wortakzente auf „**Bruder**“ und „**München**“. Als Akzentuierungsmittel kommen die Modulation der Tonhöhe, die zeitliche Strukturierung, die Lautheit und die Klangfarbe in Betracht. Eine höhere Tonfrequenz wird meist mit Betonung assoziiert und betonte Silben bzw. Vokale sind tendenziell länger als unbetonte.

Diese Parameter treten in den häufigsten Fällen kombiniert auf, damit der vom Sprecher intendierte Eindruck entstehen kann.

1.3 Emotionen

Da Emotionen ein zentrales Thema dieser Dissertation sind, bedarf es einer Erläuterung des Begriffes „Emotion“. Emotionen sind Gefühle oder Gemütsbewegungen, die physiologische Veränderungen und Verhaltensreaktionen sowie das subjektive Erleben des Gefühls beinhalten (Pschyrembel, 2010). Auf eine Situation, die persönlich als bedeutsam wahrgenommen wurde, ergibt sich eine Antwort bestehend aus Verhaltensreaktionen, kognitiven Prozessen und psychologischer Erregung (Zimbardo und Gerrig, 2004). Diese Prozesse können bewusst, z. B. beim Gefühlserleben oder unbewusst, v. a. bei der unterbewussten Interpretation einer Situation oder als physiologische Reaktion, wie beispielsweise als Erhöhung des Blutdrucks, wahrgenommen werden. Nach Darwin (1998) treten Emotionen beim Menschen sowie bei höheren Tieren auf und werden immer akustisch, mimisch und gestisch ausgedrückt. Emotionen sind bereits seit der Antike ein weites Feld der Forschung unterschiedlichster Fachrichtungen und die Diskussion um eine einheitliche Definition dieses Begriffs dauert bis zum heutigen Tage an (Flick, 2009). Während einige Autoren die Begriffe „Emotion“, „Gefühl“, „Empfindung“ und „Stimmung“ voneinander abgrenzen und jeden für sich definieren, widmen sich andere den Emotionstheorien, die in ihrer Mannigfaltigkeit weiter kategorisiert werden. Meyer et al. stellten im Jahr 2001 fest, dass es kaum ein Lehrbuch oder zusammenfassendes Buchkapitel zur Emotionspsychologie gibt, in dem nicht einleitend über das Fehlen einer allgemein anerkannten Definition von „Emotion“ Klage geführt wird. Die Frage, worin diejenigen Merkmale bestehen, die Emotionen als solche kennzeichnen und wie sie von anderen intuitiv verwandten Phänomenen, wie z. B. Stimmungen, Bedürfnissen, Gedanken, Sinnes-

1 Einleitung

empfindungen zu unterscheiden sind, ist selbst eine zentrale Forschungsfrage der Emotionspsychologie.

1.3.1 Emotionstheorien

Emotionstheorien dienen der Definition des Begriffs und der Klassifikation derjenigen Phänomene, die diesem Begriff zugeordnet werden, aber auch der Erklärung des Zustandekommens von Emotionen und deren Einfluss auf Lebewesen. Bereits im Altertum beschäftigten Epikur und Aristoteles sich mit dem Thema der Gemütsbewegungen, später trugen auch Descartes, Kant, Nietzsche und Freud dazu bei. Die Emotionstheorien lassen sich aus den Blickwinkeln der Evolutions- und Entwicklungspsychologie, der behavioristischen oder auch Verhaltenspsychologie betrachten. Aber auch lernpsychologische, kognitive und mentalistische Ansätze tragen zur Klassifikation der Emotionstheorien bei. Ebenfalls zu beachten ist der Gesichtspunkt der psychophysiologischen Vorgänge, die für das Auftreten von Emotionen verantwortlich sind bzw. diese bedingen. Um dem Leser dieser Dissertation einen Einblick in die Emotionspsychologie verschaffen, werden zunächst vier erfolgreiche klassische Emotionstheorien aufgeführt, die in Abhandlungen über Emotionstheorien und deren Mannigfaltigkeit häufig vorgestellt werden (Plutchik, 2003):

1. Theorie nach Charles Darwin: Emotionen haben adaptive Funktionen und sind universell (*Darwinistischer Ansatz*),
2. Theorie nach William James: Emotionen sind Begleiterscheinungen körperlicher Vorgänge (*James-Theorie*),
3. Theorie nach Magda B. Arnold: Emotionen basieren auf Bewertungen (*Kognitiver Ansatz*),
4. Theorie nach James Averill: Emotionen sind soziale Konstrukte und sie dienen sozialen Zwecken (*Sozialkonstruktivistischer Ansatz*).

Vor allem die evolutions- und entwicklungspsychologischen Ansätze berufen sich immer wieder auf sogenannte Basisemotionen, die nach diesen Theorien angeboren und auch bei Tieren nachweisbar sind. Ulich und Mayring (2003) listen unterschiedliche Konzepte von Basisemotionen auf, von denen exemplarisch einige hier aufgeführt werden:

1. Basisemotionen nach Magda B. Arnold: Ärger, Aversion, Mut, Niedergeschlagenheit, Begierde, Verzweiflung, Furcht, Hass, Hoffnung, Liebe, Trauer,
2. Basisemotionen nach Nico H. Frijda: Begierde, Glück, Interesse, Überraschung, Verwunderung, Leid,
3. Basisemotionen nach Carrol Izard: Ärger, Verachtung, Ekel, Belastung, Furcht, Schuld, Interesse, Freude, Scham, Überraschung,
4. Basisemotionen nach William James: Furcht, Trauer, Liebe, Wut,
5. Basisemotionen nach William McDougall: Ärger, Ekel, Freude, Furcht, Abhängigkeit, Spannung, Verwunderung.

Der Autor Raymond Cattell fügte im Rahmen weiterführender Untersuchungen u. a. noch die Emotion „Stolz“ hinzu. Nach Cattell ist jede Emotion aber auch mit einem bestimmten Ziel verknüpft, z. B. „Furcht“ mit Flucht, „Ärger“ mit Kampfeslust, „Verzweiflung“ mit einem Appell nach außen und „Stolz“ mit Selbstbehauptung (Ewen, 1998). Rost versuchte im Jahr 2001, die Begriffe Empfindungen und Emotionen voneinander abzugrenzen und kommt zu folgendem Schluss: *Entweder beziehen wir alle Empfindungen in den Kreis der Emotionen mit ein, oder wir haben nur zwei bipolare Emotionspaare: Freude - Unlust und Angst - Aggressivität. Was machen wir? [...] Ich denke, wir überlassen es einer künftig sich entwickelnden Sprachregelung in der Psychologie, was noch oder wieder „Emotion“ genannt werden soll* (Rost, 2001, S. 68).

1 Einleitung

Im folgenden Kapitel 2 wird dargelegt, welche praktischen Schritte gewählt wurden, um einen Prosodietest zu kreieren, der die bisher beschriebenen Ziele und theoretischen Grundlagen umsetzt.

2 Methoden und Versuchsplanung

Ziel aller Vorbereitungen war es, Audiodateien in Form einzeln abhörbarer, emotional unterlegter Sätze in deutscher Pseudosprache zu erzeugen, die später über einen Lautsprecher den hörgesunden Probanden dargeboten werden sollten. Der erste Schritt auf diesem Weg war das manuelle Zusammentragen von digital vorliegendem Textmaterial, aus dem später die Pseudosprache erzeugt werden sollte. Zwar standen bei der Erarbeitung einer Pseudosprache lautsprachliche Aspekte im Vordergrund, jedoch war es nicht möglich, die großen benötigten Textmengen aus rein lautschriftlichen Quellen zu beschaffen. Daher fiel die Entscheidung auf Schriftsprache, welche später von einem Sprecher mündlich dargestellt wurde. Alle weiteren Schritte konnten automatisiert durchgeführt werden, was die eigentliche Satzerzeugung reproduzierbar macht. Für die Sprachaufnahmen waren die am besten geeigneten Räume, Gerätschaften und deren Aufstellung bereits durch Voruntersuchungen im Rahmen anderer wissenschaftlicher Arbeiten bekannt (Mahfoud, 2009). Schließlich wurde die eigentliche Aufnahme der Sätze durchgeführt. Diese wurden hinsichtlich verschiedener Aspekte überprüft und analysiert bevor sie den Probanden zum Zweck der Evaluation dargeboten wurden.

2.1 Erzeugung deutscher Pseudosprache

Im folgenden Unterkapitel werden die einzelnen Schritte dargelegt, die zur Schaffung einer deutsch klingenden Pseudosprache notwendig waren.

2.1.1 Eigenschaften der Originaltexte

Zunächst musste Textmaterial beschafft werden, aus dem das Programm „hmmtext“ im Folgenden die Pseudosprache erzeugen konnte (s. Abschnitt 2.1.2, Software „hmmtext“, S. 16 ff.). Die Texte wurden ausschließlich aus dem Internet (Internet-Textquellen, 2007) bezogen und nach folgenden Kriterien ausgewählt:

- deutsche Texte,
- möglichst wenige Fremd- und fremdsprachliche Wörter im Text,
- Texte aus unterschiedlichen Epochen: 18., 19. und 20. Jahrhundert, darin jeweils ein Drittel
 - Gebrauchsliteratur (Reden, Briefe, journalistische Texte),
 - Poetische Literatur (Prosa, dramatische Texte),
 - Erörternde Literatur (Essays, Lexikontexte, Erörterungen, wissenschaftliche Darlegungen etc.).

Auf diese Weise wurde ein möglichst großer Bereich der angewandten deutschen Sprache abgedeckt, um später eine Pseudosprache zu erhalten, welche die deutsche Sprache authentisch repräsentiert. Bevor jedoch die Software „hmmtext“ zur Anwendung kam, war noch ein weiterer Schritt notwendig: Mit Ausnahme von Leerzeichen mussten alle nicht-alphabetischen Zeichen (. ! , - “ (), Zahlen etc.) und Befehle, wie z. B. Zeilenumbrüche, aus den Texten eliminiert werden und große Buchstaben mussten durch kleine ersetzt werden. Der Grund hierfür ist, dass für die weitere Bearbeitung nur noch die Buchstabenabfolgen und Wortlängen, markiert durch die Leerzeichen, relevant sind. Diese Bereinigung wurde bewerkstelligt durch die Software „Bk Replace Em 2.0“. Sie steht im Internet zum kostenlosen Download zur Verfügung (Ecobyte, 2010). Bei der Arbeit mit dem Programm ist vorab im Einzelnen anzugeben, welche Zeichen und Befehle aus dem zu bearbeitenden Text eliminiert

bzw. durch andere Zeichen ersetzt werden sollen. Dies ist auch für größere Datenmengen, in diesem Fall 1,63 MB, möglich. Nun lag der Text in folgender Form vor (Ausschnitt aus Franz Kafkas „Brief an den Vater“ (Spiegel Online, 2010):

liebster vater du hast mich letzthin einmal gefragt warum ich behaupte
ich hätte furcht vor dir ich wußte dir wie gewöhnlich nichts zu antworten
zum teil eben aus der furcht die ich vor dir habe zum teil deshalb weil zur
begründung dieser furcht zu viele einzelheiten gehören als daß ich sie im
reden halbwegs zu sammenhalten könnte und wenn ich hier versuche dir
schriftlich zu antworten so wird es doch nur sehr unvollständig sein

In den so bearbeiteten Originaltexten befanden sich zunächst noch immer einige Sonderzeichen, die entfernt werden mussten. Eine automatisierte Bereinigung auf Basis der erzeugten Rohtexte war nicht möglich, da durch die Elimination von Trennzeichen auch viele neu zusammengesetzte Wörter darin waren. Daher wurden alle Originaldateien noch einmal mit dem Textprogramm von „OpenOffice“ geöffnet, als kodierter Text im ISO-8859-15-Format abgespeichert und dann einer automatisierten Sonderzeichen- und Zeilenumbruchsentsorgung unterworfen. Danach waren nur noch vier manuelle Korrekturen notwendig, die aus Fehlern in den Originaltexten resultierten. In diesen überarbeiteten Dateien kamen noch mehrere, jedoch nur zehn unterschiedliche Buchstaben mit Akzenten vor. Sie entstammten Lehn- bzw. Fremdwörtern oder Eigennamen. Da Importe fremdsprachiger Worte, die keine Buchstabenakzente enthalten, auf diese Weise nicht entdeckt werden konnten, waren von ihnen ebenfalls noch einige enthalten. Sie repräsentieren jedoch den natürlichen Anteil an Lehn- und Fremdwörtern im Deutschen und wurden deshalb in den Originaltexten belassen. Eine saubere Grenze zwischen Fremdsprache und Deutsch wäre darüber hinaus nicht definierbar gewesen, da viele fremdsprachliche Wörter inzwischen zum Alltagsgebrauch gehören.

2.1.2 Software „hmmtext“

Das Programm wurde konzipiert und geschrieben von Dipl.-Ing. Stefan Brill im Betriebssystem „Linux“. Es kann auch im Betriebssystem „Windows“ angewendet werden, wo es auf der MS-DOS-Ebene bedient wird. Es ist in der Lage, große Mengen an Rohtextmaterial zu analysieren und daraus die Pseudosprache zu synthetisieren. Die Funktionsweise des Programms „hmmtext“ basiert auf sogenannten „Hidden-Markov-Modellen“ (HMM). Um solch ein mathematisches Modell aufzubauen, wird die Erzeugung von Sprache als stochastischer Prozess angesehen, d. h. es wird davon ausgegangen, dass ein Spracherzeugungsprozess entsprechend einer bestimmten Wahrscheinlichkeit bestimmte Laute erzeugt. Dabei erhalten bestimmte Lautübergänge höhere Wahrscheinlichkeiten als andere, beispielsweise folgt in der deutschen Sprache „r“ häufiger auf „e“ als „f“ auf „n“. Mit diesem Modell lassen sich Sprachsignale als Zufallsereignisse beschreiben, wobei für die einzelnen Wörter, Silben oder Buchstaben, unter Berücksichtigung aller möglichen Zustandsfolgen, die Wahrscheinlichkeit ihres Auftretens berechnet werden kann. Ziel ist es, aus der Sequenz der Ausgabesymbole (Beobachtungen) auf die Sequenz der verborgenen (hidden) Zustände zu schließen. Die Markov-Kette liefert eine Modellvorstellung einer Korrelation zwischen Symbolen einer Sequenz und lässt sich verstehen als Serie aus Übergangswahrscheinlichkeiten verketteter Zustände bzw. von einem Symbol zum nächsten. In diesem Fall handelt es sich bei den Symbolen um Buchstaben und bei der Sequenz um den Rohtext. In einem zweistufigen stochastischen Zufallsprozess werden die statistischen Eigenschaften bestimmter Daten modelliert. Dabei ist der erste ein diskreter Prozess, d. h. es werden alle Ausgangssymbole (Buchstaben) aus einem begrenzten Vorrat gewählt. Der zweite steuert die Ausgabe des Markov-Modells (Euler, 2006; Fink, 2003; Hütt und Dehnert, 2006; Wendemuth, 2004). Das Programm „hmmtext“ modelliert demzufolge mathematisch eine alltägliche Beobach-

tung, in diesem Fall die deutsche Sprache, indem es diese analysiert und die beobachteten Zustände auf den synthetisierten Text überträgt.

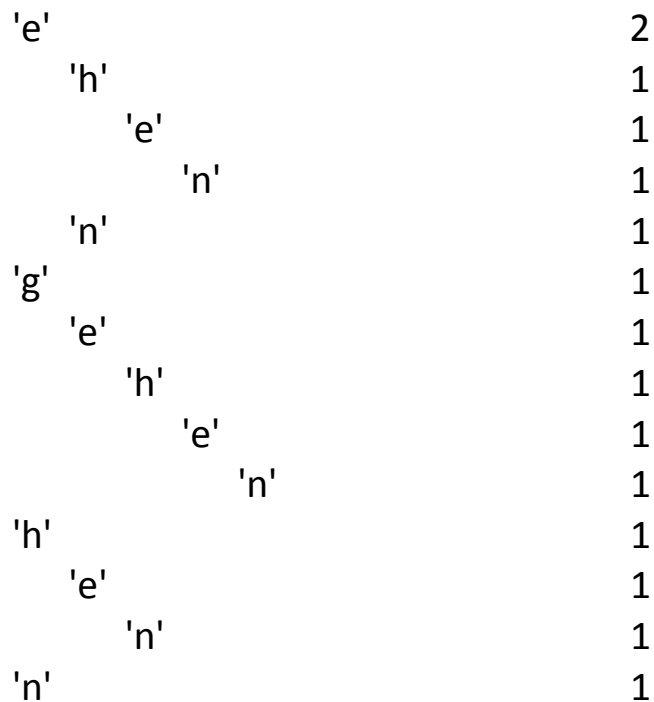


Abbildung 2.1: Baumstruktur der Analyse des Wortes „gehen“

Dabei wird im ersten Zwischenschritt der Text nach allen vorkommenden Buchstabenkombinationen in ihrer natürlichen Abfolge aufgegliedert und in einer Baumstruktur dargestellt. In dieser ist zu erkennen, welche Buchstabenkombination wie oft im Text vorkommt. Vor Beginn dieser Textanalyse wird durch einen Befehl festgelegt, bis zu welcher Tiefe die Buchstabenabfolgen der einzelnen Worte analysiert werden sollen. Ein Beispiel soll dies verdeutlichen: Bei einer Eingabe der Tiefe 5 als Befehl wird

2 Methoden und Versuchsplanung

die Buchstabenabfolge jedes Wortes und der aufeinanderfolgenden Worte inklusive der Leerzeichen bis zum fünften Buchstaben ausgezählt und wiedergegeben, angefangen von Zweier-Kombinationen bis hin zu fünf aufeinanderfolgenden Buchstaben. Die Analyse des Wortes „gehen“ mit der Tiefe 5 führt zu dem in Abb. 2.1 gezeigten Ergebnis. Die Baumstruktur ist von links oben nach rechts unten zu lesen, die Zahlen geben an, wie oft ein Buchstabe oder eine Buchstabenkombination in dem analysierten Wort auftritt. Das Wort „gehen“ wurde hier bis hin zu fünf aufeinanderfolgenden Buchstaben analysiert. Der Buchstabe „e“ kommt in dem Wort zweimal vor, die Kombinationen „eh“, „ehe“ und „ehen“ nur einmal usw. Alle in dem Wort „gehen“ vorkommenden Einzelbuchstaben sind in der ersten Spalte in alphabetischer Reihenfolge aufgelistet. In der zweiten Spalte steht der Buchstabe, der demjenigen aus der ersten bzw. im Wort direkt folgt, in der dritten Spalte der dritte folgende Buchstabe usw. Daher verweist das einzeln stehende „n“ in der zweiten Spalte auf die Kombination „en“ in „gehen“, es gehört noch zu dem Hauptast des „e“. So kann das Programm durch die Analyse von Häufigkeitsverteilungen feststellen, welche Kombinationen in der geschriebenen deutschen Sprache auftreten.

2.1.2.1 Eigenschaften der analysierten Originaltexte

Bei obigem Beispiel handelt es sich der besseren Übersichtlichkeit halber nur um ein einzelnes Wort. Im Rahmen dieser Arbeit jedoch hat „hmmtext“ auf diese Weise 55 Texte von zum Teil mehr als 36.000 Zeichen eingelesen. Wie bereits im (s. Abschnitt 2.1.2, Software „hmmtext“, S. 16 f.) erläutert, wird von dem Programm dokumentiert, wie in der deutschen Sprache Buchstaben miteinander kombiniert sind. Diese Muster werden bei der Erzeugung der Pseudosprache beibehalten und nach dem Zufallsprinzip verteilt, so dass besonders häufige Kombinationsmuster der deutschen Sprache auch in der Pseudosprache wieder vermehrt auftauchen können. Dabei ist

es möglich, dass zufällig Wörter generiert werden, die im Deutschen tatsächlich existieren. Das Programm arbeitet mit sogenannten Rohtextdateien, einem Textformat, welches im Betriebssystem Windows die Endung „.txt“ trägt. In diesem Format sind alphabetische und nicht-alphabetische Zeichen enthalten, Schriftgröße und Schriftart sind einheitlich. Daher wurden alle Textdateien, die für „hmmtext“ genutzt werden sollten, ins Format „.txt“ überführt.

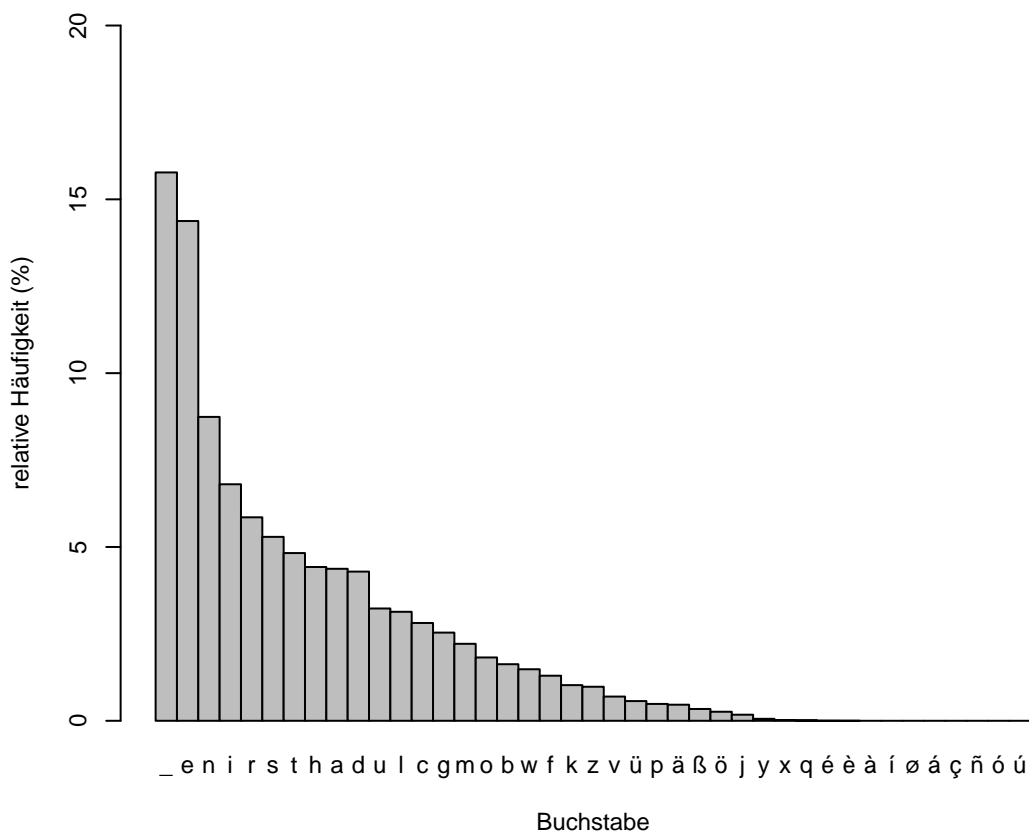


Abbildung 2.2: Relative Häufigkeiten aller vorkommenden Buchstaben

Zunächst wurde ein Analysedurchlauf der Originaltexte bis zur Abhängigkeitstiefe 15 durchgeführt, was zu einer „hmmtext“-Datei von ca. 333 MB führte, bei einer Gesamtgröße der Originaltexte von ca. 2 MB. Wie sich herausstellte, war die Analy-

2 Methoden und Versuchsplanung

se bis zur Tiefe 15 das Maximum des Erreichbaren und Sinnvollen, da eine größere Tiefe im Verhältnis zu deutlich größeren Dateien führte und keine zusätzlich brauchbaren Informationen lieferte. Aus der „hmmtext“-Datei können Informationen über die Originaltexte extrahiert werden, die die relativen Häufigkeiten aller vorkommenden Buchstaben bei einer Analysetiefe von 1 betreffen. Dies entspricht einer Abhängigkeit nullter Ordnung (s. Abb. 2.2). Der Unterstrich steht für das Leerzeichen, das wie ein beliebiger Buchstabe behandelt wird. E, N und I allein machen bereits ca. 30% der Texte aus. Rechnet man das Leerzeichen nicht mit, dann erhöht sich ihr Anteil.

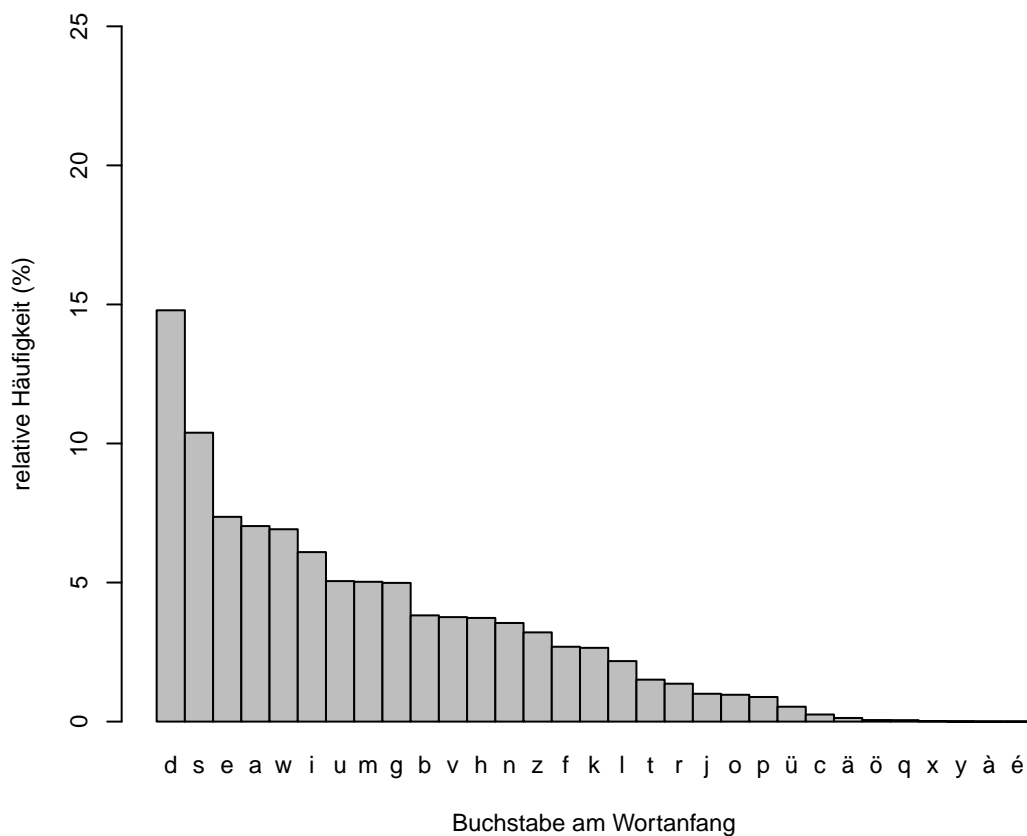


Abbildung 2.3: Buchstabenhäufigkeit am Wortanfang

2.1 Erzeugung deutscher Pseudosprache

Im nächsten Schritt wurden die Folgehäufigkeiten erster Ordnung, also die Analysetiefe 2, betrachtet. Die Graphiken A.1, A.2, A.3 und A.4 im Anhang zeigen die relativen Folgehäufigkeiten auf einige ausgewählte Buchstaben. Extreme Beispiele sind das R, nach dem fast jeder andere Buchstabe vorkommen kann oder aber das C und das Q, nach denen fast ausschließlich H und K, bzw. U im Falle des Q, folgen. Beim Q gab es nur ein Vorkommen ohne darauffolgendes U, dies war die Nennung einer „FAQ“ in einer wissenschaftlichen Abhandlung. Die relativen Häufigkeiten von Buchstaben am Wortanfang und Wortende finden sich in den Graphiken 2.3 und 2.4.

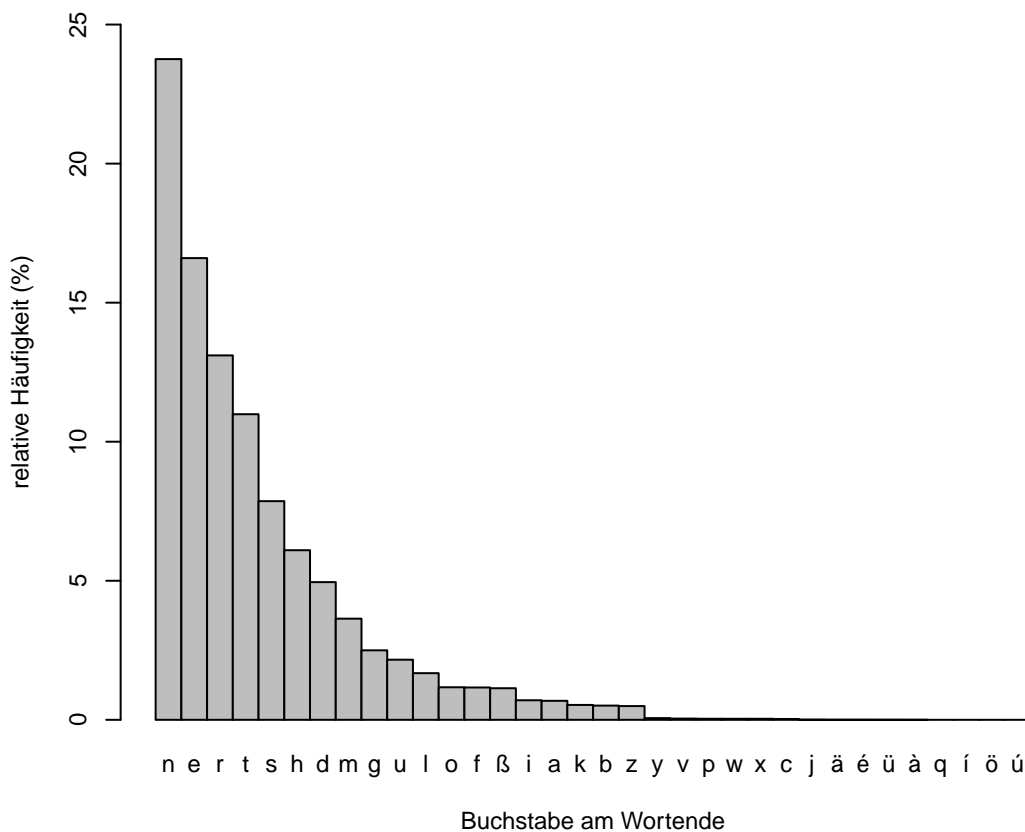


Abbildung 2.4: Buchstabenhäufigkeit am Wortende

2 Methoden und Versuchsplanung

D, S, E, A, W und I stehen bei über 50 % aller Wörter am Anfang, während N, E und R allein bereits über 50 % aller Wörter beenden. Die Graphik 2.5 zeigt die relativen

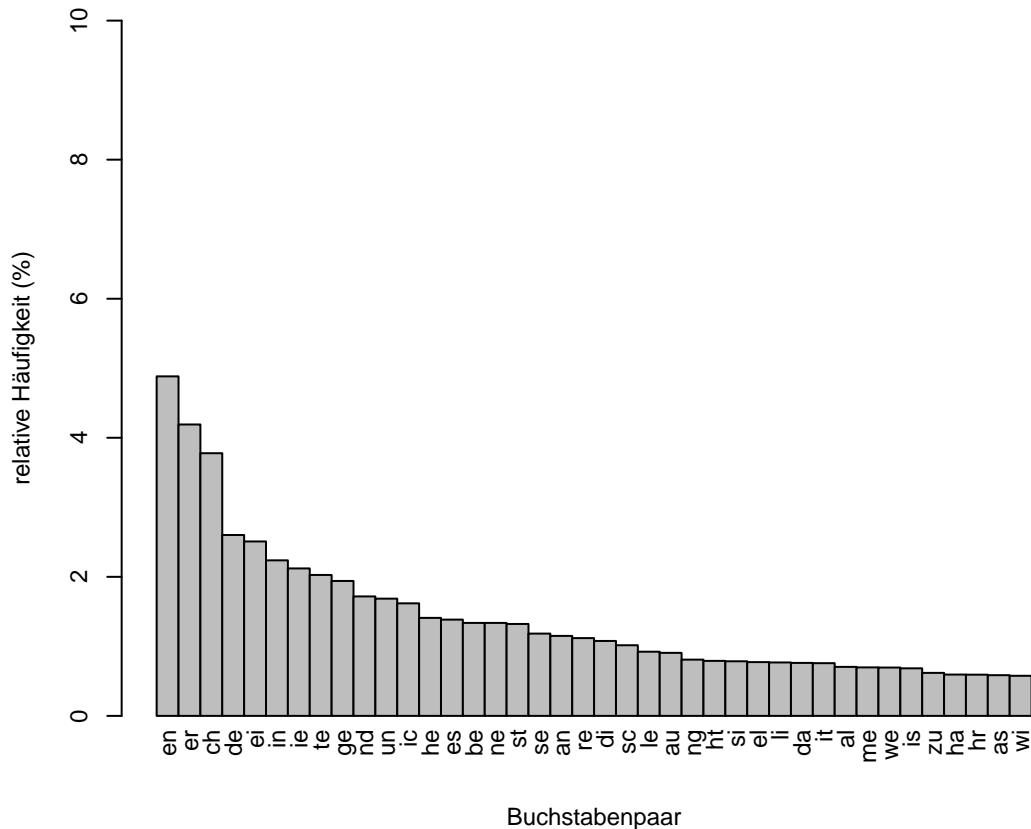


Abbildung 2.5: Häufigkeit der Buchstabenpaare im Wort

Häufigkeiten der 40 häufigsten Buchstabenpaare innerhalb von Wörtern. Im Folgenden wurde das Leerzeichen willkürlich herausgegriffen und als Wortgrenze definiert, so wie es im Alltagsgebrauch üblich ist. Zuerst wurde eine Liste aller vorkommenden Wörter des gesamten Originaltextes erstellt. Es gibt darin 36.906 verschiedene Wörter, wovon natürlich viele nur Flexionen und eventuell unterschiedliche Schreibweisen desselben Wortes sind. Eine Häufigkeitsstatistik der vorkommenden Wortlängen zeigt die Graphik 2.6. Dreibuchstabile Wörter kommen bei weitem am häufigsten vor

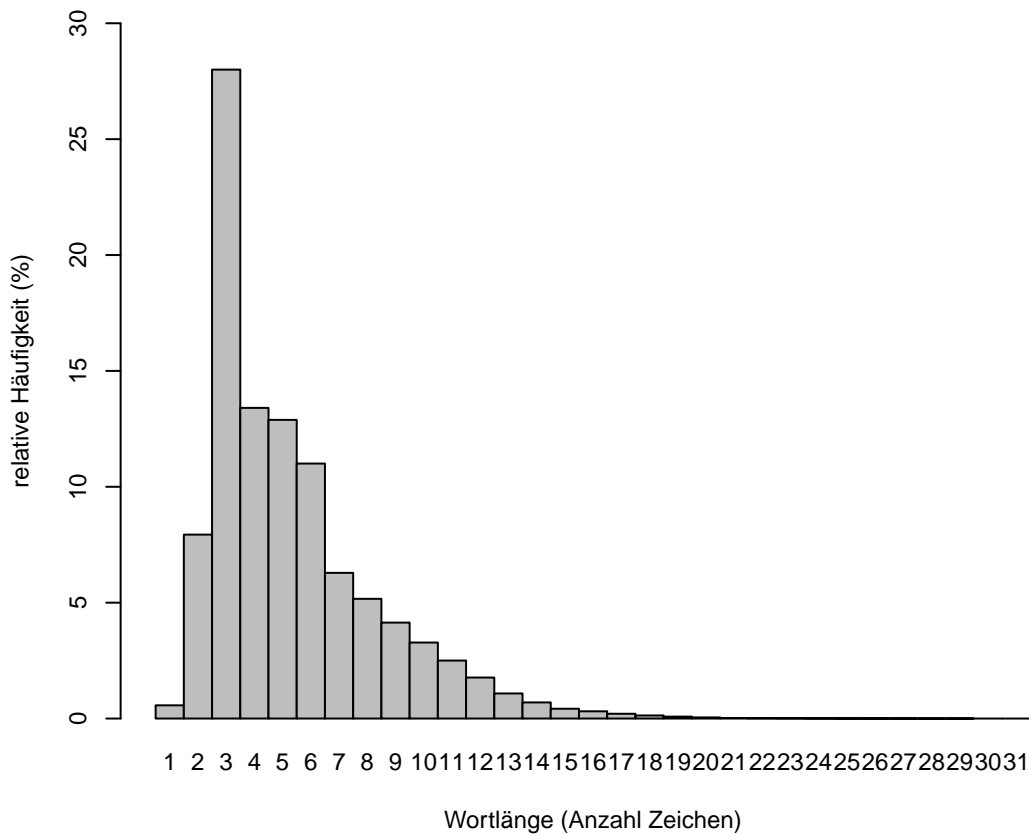


Abbildung 2.6: Wortlängenhäufigkeit

und stellen über 25 % des Originaltextes. Mehrfach auftretende gleiche Wörter wurden entsprechend mehrfach gezählt. Das längste vorkommende Wort in diesem Fall war „Tatverdächtigenbelastungszahlen“ mit 31 Buchstaben.

2.1.2.2 Vorbemerkung zur Pseudotexterzeugung

Nachdem das Programm die Analyseschritte durchgeführt hatte, wurden darauffolgend die Texte in deutscher Pseudosprache synthetisiert, basierend auf den gewonnenen Daten aus den Originaltexten. Bei der Synthese des Pseudotextes trat jedoch ein unvermeidbares Phänomen auf: es wurden nach einer gewissen Anzahl an

2 Methoden und Versuchsplanung

zufallserzeugten Zeichen nur noch repetitive Sequenzen erzeugt. Die Wahrscheinlichkeit, mit der dieser Effekt auftrat, nahm mit zunehmender Analyse-, bzw. Synthesetiefe zu. Offensichtlich kommen in dem Baum der Aufeinanderfolge-Wahrscheinlichkeiten Zweige vor, aus denen es keinen Ausweg mehr gibt, wenn der Syntheseprozess einmal dort hineingeraten ist. Es sind aber auch repetitive Zwischensequenzen mitten im Text möglich, aus denen ein Ausweg existiert, der irgendwann durch das Programm gefunden wird. Auf die Bedeutung dieses Phänomens für diese Arbeit wird im Abschnitt 2.1.2.3 noch einmal eingegangen.

2.1.2.3 Eigenschaften der synthetisierten Pseudotexte

Bei einer Betrachtung der Anzahl aller vorkommenden Zeichen finden sich 41 verschiedene auf der Synthesetiefe 1, was einer Abhängigkeit nullter Ordnung ohne Vorgängerabhängigkeit entspricht. Bei der Synthese ist daher 41 die Anzahl der möglichen wählbaren Zufallsbuchstaben, wenn noch kein vorangehender Buchstabe festliegt. Durch einen weiteren Syntheseschritt kann ausgezählt werden, wie viele zufallswählbare Buchstaben für jeden der vorangegangenen Buchstaben existieren, wobei es z. B. für das R viele und für das Q wenige Wahlmöglichkeiten gibt. Aus dem Mittelwert aller Wahlmöglichkeiten für die 41 Buchstaben des ersten Schrittes ergibt sich die durchschnittliche Größe der Wahlmöglichkeiten zur Pseudotextsynthese nachdem ein vorangegangener Buchstabe feststeht (Abhängigkeit erster Ordnung). Dies lässt sich entsprechend fortsetzen, so dass die durchschnittliche Größe der Wahlmöglichkeiten für verschiedene Analyse- bzw. Synthesetiefen berechnet und graphisch dargestellt werden kann (s. Abb. 2.7). Die mittlere Verzweigungshäufigkeit nimmt mit zunehmender Synthesetiefe rapide ab. So liegen schon ab der Tiefe 7 durchschnittlich nur noch 1,8 verschiedene Buchstaben zur (Zufalls-) Auswahl vor. Dabei liegt ein Zusammenhang mit der gewählten Größe der Textbasis vor.

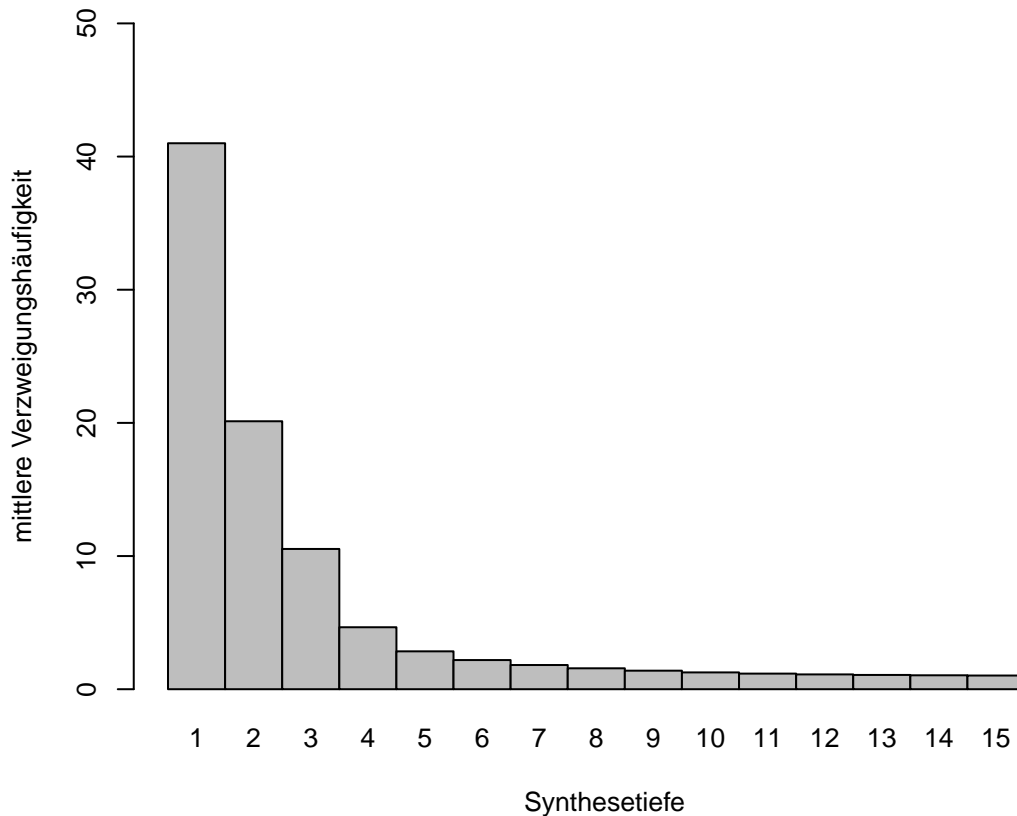


Abbildung 2.7: Mittlere Verzweigungshäufigkeit (Anzahl der Buchstaben) in Abhängigkeit von der Synthesetiefe

Mit Hilfe der einmal erzeugten „hmmtext“-Datei wurden für jede Synthesetiefe von 1 bis 15 jeweils 20 Pseudotext-Dateien der Größe 1 Mio. Zeichen erzeugt, insgesamt waren dies 300 Dateien. Aus diesen wurde wieder jeweils eine Liste aller vorkommenden verschiedenen Pseudowörter erzeugt. So konnte ausgezählt werden, wieviele verschiedene Wörter in Abhängigkeit von der gewählten Synthesetiefe entstanden. Die Graphik 2.8 zeigt den Zusammenhang, wobei jeder einzelne Boxplot die Werte der 20 entsprechenden Pseudotext-Dateien enthält.

Bei den ersten vier Synthesetiefen ist die Streuung der jeweiligen 20 Dateien bzw. Wortlisten so klein, dass der Boxplot sich auf einen einzelnen Querbalken reduziert.

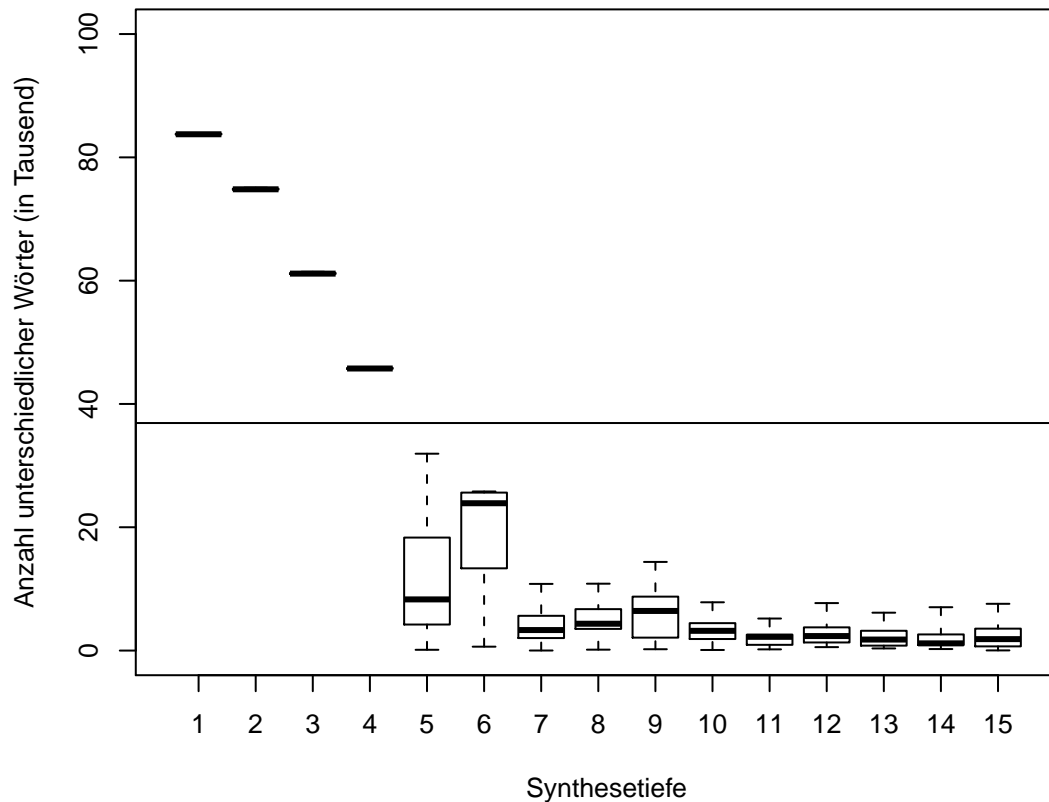


Abbildung 2.8: Anzahl unterschiedlicher Wörter in Abhängigkeit von der Synthesetiefe

Die Linie über die ganze Breite der Graphik gibt die Lage der Zahl 36.906 an, also die Anzahl der vorkommenden Wörter im Gesamt-Originaltext. Die ersten vier Synthesetiefen liegen deutlich darüber, was bedeutet, dass mindestens entsprechend viele neue Wörter erzeugt worden sind. Zu den hohen Synthesetiefen hin - sehr deutlich ab 11 - bleibt die Anzahl der Wörter im Pseudotext weit unter der Anzahl der Wörter im Originaltext. Die großen Streuungen ab der Synthesetiefe 5 sind vermutlich auf Dateien zurückzuführen, die schon früh in eine repetitive Sequenz geraten und ab da nur noch eine geringe Anzahl unterschiedlicher Wörter aufweisen. Wie groß der Anteil der Wörter ist, die schon im Originaltext vorkommen, zeigt die Graphik 2.9. Bis

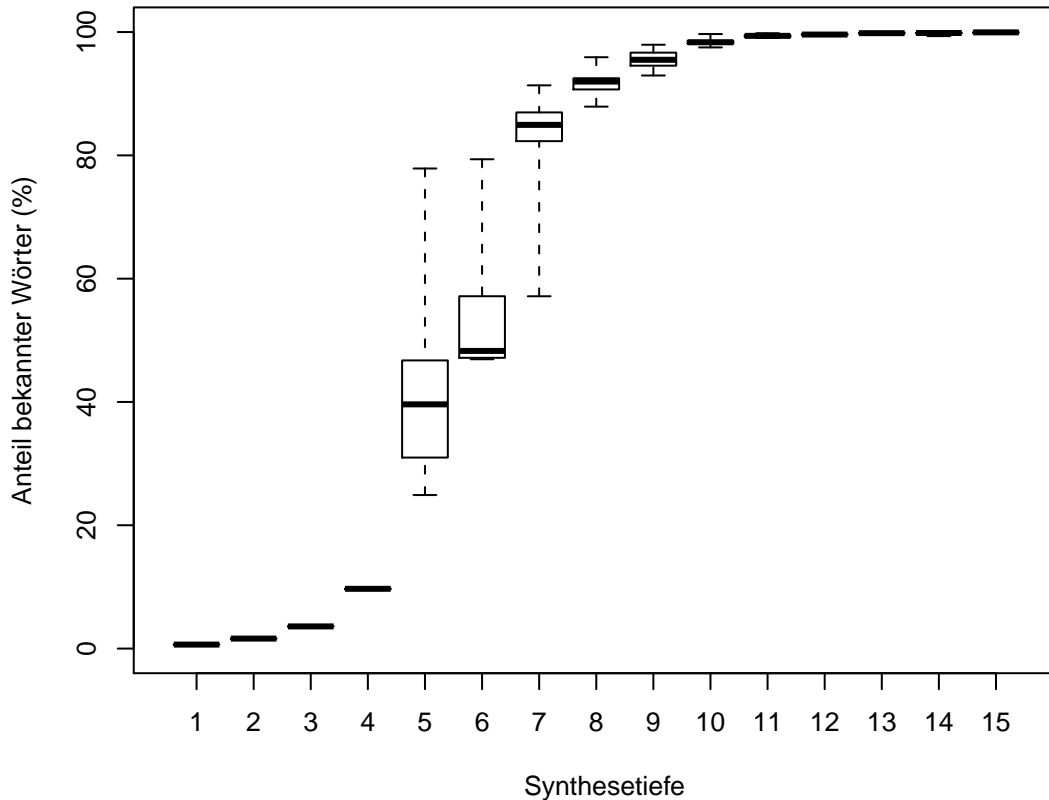


Abbildung 2.9: Anteil bekannter Wörter in Abhängigkeit von der Synthesetiefe

zur Tiefe 4 bleibt er deutlich unter 20 %, d.h. über 80 % der erzeugten Wörter sind Neuschöpfungen. Ab der Tiefe 8 ist es umgekehrt, hier sind über 80 % der Wörter bereits im Originaltext enthalten und bei weniger als 20 % handelt es sich um neuartige Wörter. Auch hier sind wieder in jedem Einzel-Boxplot die 20 Werte der jeweiligen Pseudotext-Dateien enthalten. Die Streuungen sind für die Tiefen 1 - 4 und 10 - 15 sehr klein, so dass die Boxplots fast nur aus einem Querbalken bestehen. Auch hier spielen die repetitiven Sequenzen eventuell eine manipulative Rolle.

So wie mit dem Originaltext wurde auch mit den Pseudotexten eine Statistik der vorkommenden Wortlängen erstellt, für jede Synthesetiefe separat. Diese befindet

2 Methoden und Versuchsplanung

sich im Anhang (s. Abschnitt A.2, Ergebnisse bei steigender Synthesetiefe, S. 78 f.). In der Graphik 2.10 werden die Wortlängen der jeweils längsten vorkommenden Wörter in den Pseudotext-Dateien gezeigt. In dieser Graphik finden sich insgesamt 300

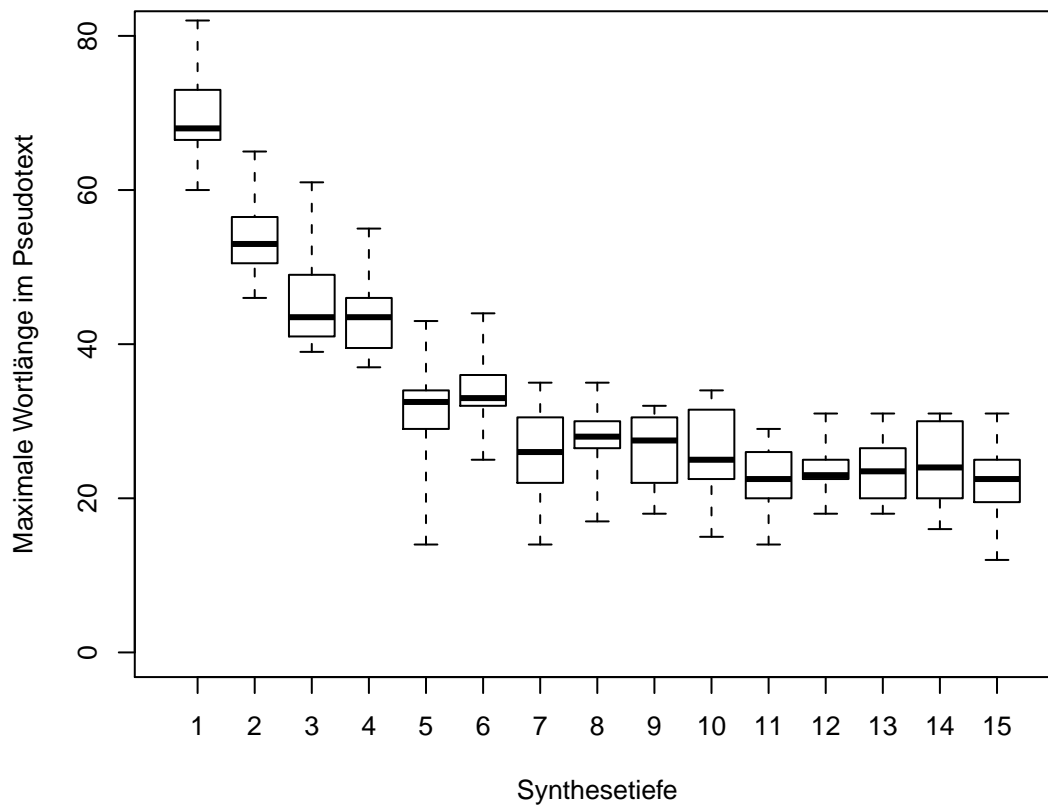


Abbildung 2.10: Wortlängenmaxima (Anzahl der Buchstaben) in Abhängigkeit von der Synthesetiefe

Wortlängenmaxima, da auch hier für jede der 15 Synthesetiefen 20 Dateien in einen Boxplot gelegt wurden. In jeder dieser aus 1 Mio. Zeichen bestehenden 20 Dateien wurde jeweils nach dem längsten vorkommenden Wort gesucht. Diese Statistik ist vom geringen Effekt der repetitiven Sequenzen kaum beeinflusst, da nur das jeweils längste Wort jeder Pseudotextdatei eingebracht wird. Der Einfluss könnte höchstens dadurch zustande kommen, daß bei frühem Verfallen in repetitive Sequenzen weni-

2.2 Auswahl der Sätze und Erstellung einer Satzliste

ger Gelegenheit besteht, längere Wörter zu generieren. Die längsten Wörter sind bei den Synthesetiefen 1 bis 4 unnatürlich lang. Ab der Tiefe 5, spätestens 7, befinden sich die maximalen Wortlängen in Orientierung an der deutschen Sprache aber im Normalbereich. Das längste Wort im Originaltext hat 31 Zeichen, was von beinahe allen Synthesetiefen ab 5 sowohl über- als auch unterboten wird. Nachfolgend befindet sich beispielhaft ein Ausschnitt eines Pseudotextes der Analyse- und Synthesetiefe 4:

durchzeich schlich ware südhaftlicht leit dassene näches festimmenschen
gesen gen wette im her s davon wird es wenkenn der selbst die nun
destammer ihrer kinde größer die mitt jeher tugehenarrom und gütigeri
lach wem kumme [...] einerverst mir verschliegt sal passen auch die
größere erröte diesen von tod so phob singen frieferteil es öffnet eine
aufgefindlichte bösen unsetzbar immer müde taus vorgenz du des sein
höhe gali abhänder ichtung grobleiteten

Die Tiefe 5 für den Analyse- und Syntheseschritt hat sich als die sinnvollste erwiesen, da auf diese Weise die höchste Anzahl an direkt aufeinanderfolgenden Pseudowörtern erzeugt werden konnte und außerdem die repetitiven Sequenzen vermindert auftraten.

2.2 Auswahl der Sätze und Erstellung einer Satzliste

Aus den synthetisierten Pseudotexten wurden nun Sequenzen von direkt aufeinanderfolgenden Wörtern ausgewählt, die möglichst ausschließlich aus Pseudowörtern bestehen sollten. Diese Sequenzen dienten als Sätze, wobei sie so auszuwählen waren, dass die Silbenlänge entweder 10 oder 15 betrug, um Sätze weitestgehend gleicher Länge zu erhalten. Aus der Vorauswahl von 30 zehnsilbigen und 30 fünf-

2 Methoden und Versuchsplanung

zehnsilbigen Sätzen wurden diejenigen ausgewählt, die am besten sprechbar waren, d. h. die das geringste Potential für Sprechfehler während der Aufsprachen boten. Auf diese Weise entstanden 22 Zehnsilber und 22 Fünfehsilber, wobei zunächst jeweils zwei Sätze zur Reserve dienen sollten für den Fall, dass später noch einige verworfen werden würden. Für diese 44 Sätze wurde eine Liste erstellt, die dem Sprecher als Vorlage diente. In jedem Satz wurde nun für jedes Wort die jeweils zu betonende Silbe nach dem Kriterium der Akzentsprache ausgewählt und markiert, um die Aussprache zu normieren und reproduzierbar zu gestalten. Die Betonung einzelner Wörter innerhalb eines Satzes durfte je nach intendierter Emotion variieren. Als Beispiel soll der Satz 25 der Satzliste dienen: **besicht** für den kub**ik**kilometer ent**fer**nunft**prinzip**.

2.3 Auswahl der Emotionen

Bei der Auswahl der emotionalen Stimmungen, in denen die Sätze gesprochen werden sollten, kam es zum einen darauf an, den Gemütszustand mit der Stimme überzeugend darstellen zu können. Zum anderen wurde ein möglichst gleichmäßiges Verhältnis zwischen positiv und negativ belegten Emotionen angestrebt. Es kamen zunächst 18 Emotionen für eine sprachliche Umsetzung in Frage (s. Tab. 2.1).

Positiv	Negativ	Weitere
Begeisterung	Aggression	Ekel
Erleichterung	Besorgnis	Gleichgültigkeit
Freude	Furcht	Mitleid
Hoffnung	Panik	Schüchternheit
Interesse	Trauer	Überraschung
Stolz	Zorn	Verachtung

Tabelle 2.1: Ausgewählte Emotionen für die sprachliche Umsetzung

Für jede Emotion wurden nun Beispielsätze entworfen, die dem Sprecher beispielhaft für emotionale Alltagssituationen als Sprechübungen dienen sollten.

1. **Freude:** Ich habe es geschafft!
2. **Erleichterung:** Endlich ist es geschafft.
3. **Begeisterung:** Das ist einfach unglaublich!
4. **Stolz:** Dort drüben steht mein neues Auto!
5. **Hoffnung:** Hoffentlich schaffe ich es diesmal.
6. **Interesse:** Das müssen Sie mir unbedingt noch einmal erläutern!
7. **Zorn:** Du hast mein neues Auto zerstört!
8. **Aggression:** Jetzt habe ich aber langsam genug!
9. **Furcht:** Ich habe dort hinten ein Geräusch gehört!
10. **Panik:** Lauft um Euer Leben!
11. **Trauer:** Ich habe es wieder nicht geschafft.
12. **Besorgnis:** Sie hätte schon vor zwei Stunden hier sein sollen.
13. **Verachtung:** Du hast doch noch nicht mal einen Schulabschluss!
14. **Ekel:** Das da kann ich einfach nicht essen!
15. **Überraschung:** Das Fenster stand ja die ganze Zeit offen!
16. **Gleichgültigkeit:** Ob Du das schaffst oder nicht, ist mir völlig egal.
17. **Schüchternheit:** Ich traue mich nicht, allein hinzugehen.
18. **Mitleid:** Beim nächsten Mal schaffst Du das schon.

2 Methoden und Versuchsplanung

Interesse, Hoffnung und Besorgnis erwiesen sich frühzeitig als ungeeignet, nachdem sich in den Sprechübungen herausgestellt hatte, dass diese Emotionen bezüglich ihrer Eindeutigkeit nicht befriedigend dargestellt werden konnten. Die verbleibenden 15 Emotionen wurden in Form der Pseudosätze im nächsten Schritt aufgesprochen.

2.4 Durchführung der Aufsprache

Die Aufsprachen erfolgten ausschließlich durch Herrn Dipl.-Ing. Stefan Brill, alle Emotionen wurden nur mit einer männlichen Stimme dargestellt. Es wurden keine professionellen Sprecher oder Schauspieler engagiert, um alle Aufsprachen sowie später notwendige Korrekturaufsprachen zeitlich und praktisch flexibel gestalten zu können. Des Weiteren verfügt Herr Dipl.-Ing. Stefan Brill über eine langjährige positive Erfahrung als Sprecher für wissenschaftliche Arbeiten, die er in der Vergangenheit betreute. Die Liste mit den 44 Sätzen diente ihm als Vorlage, von der er die Sätze jeweils vorlas. Jeder Satz wurde fünf bis zehn Mal gesprochen und aufgezeichnet. So ergab sich später die Möglichkeit, den besten Satz auszuwählen, bzw. nicht korrekt gesprochene Sätze oder solche mit Störgeräuschen verwerfen zu können. Trotz aller Vorkehrungen können Störgeräusche auftreten, wenn der Sprecher z. B. versehentlich das Mikrofon berührt, der Sprecherabstand zum Mikrofon temporär zu gering ist oder mit höherem Schalldruckpegel gesprochen wird, was zu einem Clipping in der betreffenden Sounddatei führt. Der Begriff „Clipping“ beschreibt eine Übersteuerung des Eingangssignals, wenn dessen Maximalwert die Grenze des zulässigen Pegels, der in der Software festgelegt wurde, überschreitet (Lerch und Weinzierl, 2008; Mahfoud, 2009).

Vor Beginn der Aufnahmen fanden einige Testdurchläufe statt, die Hinweise auf Mängel der sprachlichen Umsetzung oder des Aufnahmeverfahrens und Gelegen-

heit zur Korrektur dieser Defizite gaben. Zur Erleichterung des späteren automatischen Schneidevorgangs wurde während der Aufnahmen die Summe der Wiederholungen pro Satz und Emotion mit Hilfe einer Strichliste protokolliert. Dadurch konnte der Sprecher diese bei Bedarf spontan erhöhen und musste sich nicht darauf konzentrieren, eine vorgegebene Anzahl einzuhalten. Jeder gesprochene Satz wurde durchschnittlich sieben bis acht Mal aufgezeichnet, die Aufnahmequalität lag bei 96000Hz/24Bits.

2.4.1 Räumlicher und technischer Aufbau

Die Aufnahmen der Pseudosätze für den Prosodieverständnistest wurden in Kabine B der Audiometrieabteilung der Würzburger Kopfklinik durchgeführt, die normalerweise für Audiometriemessungen benutzt wird. Sie besteht aus drei schallisolierten Räumen, die durch ebenfalls schalldichte Türen und Sichtfenster miteinander in Verbindung stehen. In der mittleren Kammer befinden sich die Steuerungskonsolen für die Tonaudiometrie, rechts und links sind die Räume, in denen sich im alltäglichen Klinikbetrieb die Patienten für die Hörtests befinden. Während der Sprachaufnahmen befand sich der Sprecher (Dipl.-Ing. Stefan Brill) in der rechten Kammer an einem festgelegten Platz in der hinteren linken Raumecke vor dem Mikrofon und führte die Aufsprachen durch. Diese wurden analog den Ergebnissen ausgeführt, die sich aus den Voruntersuchungen im Rahmen der Dissertation von Mahfoud (2009) ergeben hatten. Die Aufsprachen wurden ebenfalls mit dem Kondensatormikrofon „Sony C-48“ aufgenommen, das auf Kugelcharakteristik eingestellt war. Dies erwies sich als günstigste Aufnahmeeinstellung in der Magisterarbeit von Bauer (2005). Auch im Aufbau dieser Arbeit wurden die analogen Signale des Mikrofons an die „Phase X24“-Soundkarte von „Terratec“, die einen hochwertigen Mikrofonvorverstärker besitzt, weitergegeben. Diese Signale wurden digitalisiert und über einen Firewire-

2 Methoden und Versuchsplanung

Anschluss an einen PC weitergegeben. Zum Einsatz kam dabei ein „Asus W1N“-Notebook, das über die Software „Cubase LE (Steinberg)“ verfügt. Mit diesem Programm können die digitalen Signale als Wav-Dateien abgespeichert und in Form von Zeitverlaufsdarstellungen sichtbar gemacht werden. Mit dem Signalgenerator „Typ 1049“ von „Brüel & Kjær“ wurde ein Sinuston mit der Frequenz von 1000 Hz generiert, der genau 1 s andauerte. Er wurde zu Beginn des zuerst aufgesprochenen Satzes sowie im Folgenden an jedem Satzende aktiviert. Über einen Passivlautsprecher, der in einem Abstand von etwa einem Meter zum Mikrofon aufgestellt war, wurde der Sinuston abgespielt und über das Mikrofon aufgenommen. Dieses Signal diente später als Schneidemarkierung für die Sätze. Ein Aktivlautsprecher wäre möglicherweise durch seine Eigengeräusche in den späteren Aufnahmen nachweisbar gewesen (Mahfoud, 2009), (s. Abschnitt 2.5, Schneiden der Sätze und Skalierung des SPL, S. 34). Den Vorverstärker, das Notebook und den Signalgenerator bediente die Doktorandin (Josefine Hirschfeld), die ihren Arbeitsplatz für die Sprachaufnahmen im mittleren Raum hatte. So wurde vermieden, dass Nebengeräusche, wie z. B. Lüftergeräusche der Geräte, vom Mikrofon mit aufgenommen wurden. Das Mikrofongabel und das Verbindungskabel zwischen Signalgenerator und Passivlautsprecher wurden unter der schalldichten Tür hindurchgeführt. Die Aufstellungen für die Sprachaufnahmen orientierten sich exakt an denjenigen aus der Dissertation von Mahfoud (2009) und sind in der Skizze 2.11 dargestellt.

2.5 Schneiden der Sätze und Skalierung des SPL

Beim Schneiden der Sätze dienten die eingespielten Sinustöne als Orientierung für Anfang und Ende des Satzes. Das automatisierte Schneiden erfolgte unter der Numerik-Software „Matlab“ mit Hilfe eines Schnittskriptes, das im Editor „Emacs“ erstellt wurde. Bei der Arbeit mit diesen beiden Programmen kam die freie „GNU/Linux“-

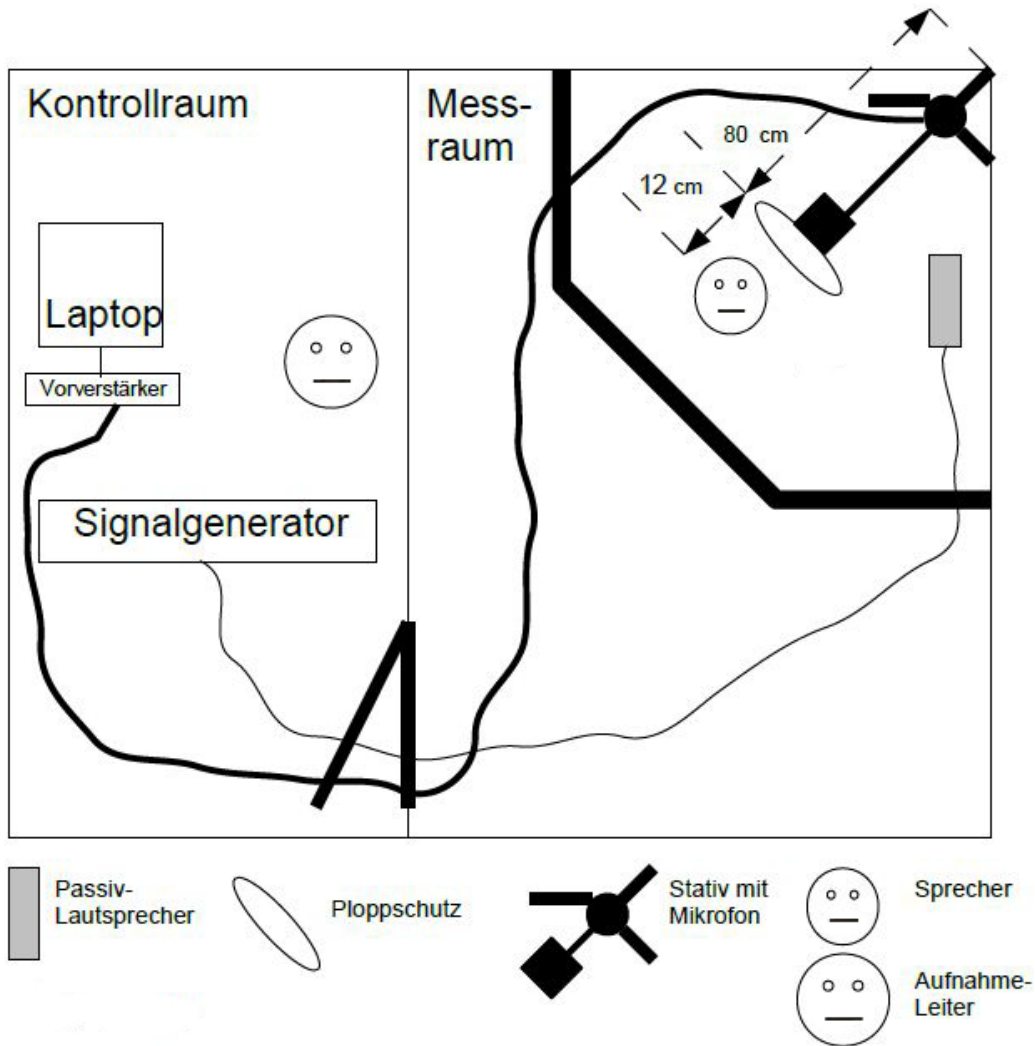


Abbildung 2.11: Skizze der Aufstellungen in der Kabine für die Sprachaufnahmen nach Mahfoud (2009)

Distribution „Knoppix“ zum Einsatz. Das automatisierte Schneiden anhand der Sinustöne erfolgte genau wie bereits in der Dissertation von Mahfoud (2009) beschrieben. Dabei diente jeder Sinuston als Schnittmarke. Auf diese Weise wurde jeder Satz zwischen zwei Sinustönen herausgeschnitten und in einer eigenen Datei mit entsprechender automatischer Benennung abgespeichert. Im letzten Schritt wurden die Schalldruckpegel (Sound Pressure Level = SPL) aller zu beurteilenden Sätze

2 Methoden und Versuchsplanung

automatisiert gleichgesetzt, d. h. die Maximalwerte der SPL-Zeitverläufe in dB wurden einander angeglichen. Durch diese Maßnahme konnte erreicht werden, dass die Sätze sich in ihrer maximalen Lautstärke nicht unterscheiden (Mahfoud, 2009), vgl. (s. Abschnitt 4.2.3, Technische Einflüsse, S. 66). Diese Lautstärke-Kalibrierung wurde ebenfalls mit Hilfe von „Matlab“ durchgeführt. Sämtliche Sätze wurden auf -14,3 dB nivelliert. Durch diesen Wert konnte für alle aufgezeichneten Daten ein einheitlicher maximaler SPL im Kurvenverlauf erzeugt werden unter der Voraussetzung, dass keine der Dateien aufgrund dieser Nivellierung den festgelegten Schwellenwert mit ihren Pegelspitzen überschritt.

2.6 Auswahl der gesprochenen Sätze und Nachaufsprache

Es lagen nun pro Emotion etwa 350 Sätze vor, je zur Hälfte Zehnsilber und Fünfzehnsilber. Diese wurden einzeln abgehört und der jeweils beste Satz wurde selektiert. Dabei waren die Kriterien die gelungene sprachliche Umsetzung der intendierten Emotion, es durften keine Aussprache- und Betonungsfehler vorliegen und es sollten keine externen oder internen Störgeräusche wahrnehmbar sein. Insgesamt wurden 660 Sätze aus rund 5.300 Sätzen manuell ausgewählt, dies sind 44 Sätze für jede der 15 Emotionen. Dabei stellte sich heraus, dass vier Einzelsätze aufgrund falscher Betonung verworfen werden mussten. Des Weiteren war die gesamte Satzreihe der Emotion „Gleichgültigkeit“ nach Einschätzung der Durchführenden unbefriedigend emotional dargestellt. Daher fanden Nachaufsprachen der betroffenen Sätze statt. Die wiederholten Aufnahmen sowie die anschließenden Bearbeitungsschritte fanden nach der gleichen Vorgehensweise statt wie in den Hauptaufsprachen.

2.7 Pilottest

Der Pilottest diente als Vorstudie, in der getestet werden sollte, ob der gesamte Ablauf reibungslos gelang, die Pilotprobanden die Aufgabenstellung verstanden und mit welchem Zeiteinsatz zu rechnen war. Des Weiteren konnte vorab ermittelt werden, ob am geplanten Studiendesign noch Änderungen vorzunehmen waren und welche sich als die am besten erkannten Emotionen herausstellten im Hinblick auf eine weitergehende Selektion. Die Pilotgruppe bestand aus 15 zufällig zusammengestellten wissenschaftlichen und ärztlichen Mitarbeitern der HNO-Klinik Würzburg zwischen 28 und 68 Jahren. Eine gleichmäßige Geschlechterverteilung lag nicht vor. Es nahmen elf männliche und vier weibliche Personen teil. Auch wurden sie eingangs keinem Audiometrie-Hörtest unterzogen. Alle Probanden der Pilotgruppe befanden sich sitzend und gleichmäßig verteilt gemeinsam in einem Raum und wurden dort mündlich und schriftlich über Ablauf und Ziel der Untersuchung aufgeklärt. Beispielhaft wurden ihnen sechs Pseudosätze über einen Lautsprecher präsentiert. Im Anschluss wurde die Gruppe aufgefordert, 30 Pseudosätze (pro Emotion zwei Sätze) offen zu evaluieren, indem jeder Teilnehmer für sich handschriftlich auf einem vorgefertigten Bogen mit möglichst wenigen Worten die intendierte Emotion benennen sollte. Dabei wurde jeder Satz einmal vorgespielt, danach wurden jeweils etwa 20 Sekunden zur Bewertung eingeräumt, bevor der nächste Satz abgespielt wurde. Die Probanden benötigten insgesamt etwa 15 Minuten für die reine Testphase. Die Sätze wurden in jeweils einer von insgesamt 15 möglichen Emotionen präsentiert, wobei die Probanden über die Auswahl an Emotionen nicht informiert wurden, um den Test offen ansetzen zu können. Die intendierten Emotionen waren: Aggression, Begeisterung, Ekel, Erleichterung, Freude, Furcht, Gleichgültigkeit, Mitleid, Panik, Schüchternheit, Stolz, Verachtung, Trauer, Überraschung, Zorn. Die 30 Sätze waren zuvor willkürlich ausgewählt worden unter dem Gesichtspunkt des nach Einschätzung der Autorin

2 Methoden und Versuchsplanung

repräsentativsten Satzes für die jeweilige Emotion. Anhand der Ergebnisse des Pilottests (s. Abschnitt 3.1, Pilottest: Reduktion auf 10 Emotionen, S. 45 f.) wurden nun diejenigen Emotionen verworfen, die keine eindeutigen Antworten seitens der Pilotgruppe erbracht hatten. Dabei wurde auf sinngemäße Übereinstimmung mit der intendierten Emotion geachtet sowie auf wortwörtliche Deckungsgleichheit innerhalb der Pilotgruppe. Als Grenze diente hier eine Kongruenz von weniger als drei. Solche Begriffe, die sinngemäß der intendierten Emotion entsprachen, jedoch von den Probanden mehrheitlich andersartig bezeichnet wurden, waren gute Hinweise auf die Notwendigkeit der Überarbeitung der vorgegebenen Emotionsbegriffe. Daher wurde wie folgt vorgegangen: Fünf Emotionen wurden gestrichen, diese waren: Begeisterung, Erleichterung, Furcht, Mitleid und Schüchternheit. Somit war die Gesamtanzahl der Sätze von 660 auf 440 reduziert. Die Bezeichnungen der verbleibenden zehn Emotionen wurden um je einen weiteren Ausdruck ergänzt und in Adjektive überführt, sodass pro Emotion ein Wortpaar entstand. Bei diesem neu hinzugekommenen Ausdruck handelte es sich um die jeweils am häufigsten genannte Angabe, die gleichzeitig sinngemäß übereinstimmend mit dem ursprünglich vorgegebenen Emotionsbegriff war.

Eine Ausnahme bildet hier die Ergänzung von „angeekelt“ um „weinerlich“. Diese beiden Gemütszustände entsprechen nicht direkt einander, jedoch war „weinerlich“ anstelle von „angeekelt“ so vielfach genannt worden, dass diese Interpretation nicht unberücksichtigt bleiben konnte, sondern im Hinblick auf die Versuche mit der Hauptgruppe aufgegriffen wurde. In der zweiten Spalte befindet sich zusätzlich eine festgelegte einbuchstabige Abkürzung für jede Emotion, mit der im späteren Verlauf der Versuchsdurchführung und -auswertung übersichtlich gearbeitet werden konnte.

Emotion	Abkürzung
aggressiv / genervt	A
überrascht / verwundert	B
angeekelt / weinerlich	E
freudig / fröhlich	F
gleichgültig / gelangweilt	G
aufgeregt / panisch	P
stolz / prahlerisch	Z
traurig / betrübt	T
verächtlich / herablassend	V
zornig / verärgert	Z

Tabelle 2.2: Emotionswortpaare und ihre Abkürzungen

2.8 Erstellung der Abspiellisten und Audiodateien

Folgende Fragestellungen ergaben sich für den Versuchsaufbau:

1. Verteilung der Emotionen auf die Sätze,
2. Aufteilung der Sätze auf die Probanden,
3. Reihenfolge der Sätze und
4. Anzahl der Probanden.

Die Grundidee zur Erstellung der Abspielliste für die Probanden war, dass jeder Proband alle 440 Sätze anhören sollte, jedoch jeweils in einer anderen Reihenfolge. Damit sollte vermieden werden, dass das Endergebnis durch Faktoren wie Eingewöhnen beim Hören, Ermüdung, Lerneffekte usw. beeinflusst wird. Von der Annahme ausgehend, dass nicht alle Aufsprachen gleichwertig sind, war es auch aus diesem Grund sinnvoll, nicht jedem Probanden die gleiche Abhörreihenfolge zuzuteilen, sondern die Aufnahmen so zu vermischen, dass die 440 Sätze gleichmäßig auf die Probanden verteilt und somit die Probanden gleichmäßig belastet wurden. Zu diesem Zweck erstellte Herr Dipl.-Ing. Stefan Brill mithilfe der Linux-

2 Methoden und Versuchsplanung

Shell „bash“ ein Programmskript, das ein Verwürfelungsschema für die 440 Satz-Emotions-Kombinationen erzeugte. Das Schema kam folgendermaßen zustande: Es wurden 20 Spalten einer Rechtecktafel erstellt, wobei jeweils 10 Spalten die 10 Emotionen enthielten, der erste Block die 10-silbigen Sätze, der zweite Block die 15-silbigen. Es gab 22 Zeilen für die je 22 Pseudosätze jeder Silbenzahl. Die Spalten wurden jeweils in sich mit einem Zufalls-Offset so durchrotiert, dass jeder Pseudosatz nur einmal pro Zeile vorkam. Im nächsten Schritt wurden die Spalten dieses Rechtecks in sich unverändert belassen, aber die Zeilen zufallsverwürfelt. Dann wurden die 22 Zeilen dieser Tabelle von oben nach unten der Reihe nach entnommen, in sich zufallsverwürfelt und durchnummerierend aneinandergehängt. Das Resultat war eine Reihenfolge von 440 Sätzen, bestehend aus 22 20er-Blöcken (den ehemaligen Zeilen), innerhalb derer jede Emotion zweimal auftrat. Außerdem lagen zehn 10- und zehn 15-silbige Sätze vor und kein Pseudosatz wiederholte sich. Die ganze Prozedur wurde so oft wiederholt, bis ein Schema gefunden wurde, in dem keine Pseudosatz-Doubletten und genau 22 EmotionsDoubletten vorkamen ($22 = 5\%$ von 440). Unter Pseudosatz-Doubletten ist die Aufeinanderfolge zweier identischer Pseudosätze unterschiedlicher Emotion zu verstehen, wohingegen mit Emotions-Doubletten die Aufeinanderfolge zweier identischer Emotion in unterschiedlichen Pseudosätzen gemeint ist. Nach der achten Durchrotation dieses Schemas lag wieder dieselbe Reihenfolge vor wie zu Beginn. Es hatten sich also acht verschiedene Aufteilungsmuster ergeben, sodass die Probandenzahl ein Vielfaches dieser Zahl erreichen musste. Die Entscheidung fiel auf 32 Probanden, die in acht Gruppen à vier Personen (je zwei männliche und zwei weibliche) aufgeteilt, jeweils dieselbe Reihenfolge im Versuch hörten. Die Verteilung der Probanden auf die Gruppen stellte sich folgendermaßen dar:

Gruppe	01	02	03	04	05	06	07	08
Probandin	01	02	03	04	05	06	07	08
Probandin	09	10	11	12	13	14	15	16
Proband	17	18	19	20	21	22	23	24
Proband	25	26	27	28	29	30	31	32

Tabelle 2.3: Aufteilung der Probanden auf die acht Gruppen

Die bereits vorliegenden wav-Dateien wurden durch Umbenennen in acht verschiedene Abspiellisten kopiert und gemeinsam mit einem CCITT-Rauschen, das ebenfalls auf -14,3 dB (s. Abschnitt 2.5, Schneiden der Sätze und Skalierung des SPL, S. 34) skaliert war, auf zwei CDs gebrannt. Bei diesem Rauschen handelt es sich um ein gemäß CCITT Rec. G.227 genormtes Störgeräusch, das die spektralen Eigenschaften gesprochener Sprache nachbildet (International Telecommunication Union, 1988; Stavrianou, 2010).

2.9 Probandenauswahl

Die Probanden, mit denen die Versuche durchgeführt wurden, gliederten sich in zwei Gruppen. Mit der ersten wurden Vorversuche innerhalb der Pilotgruppe durchgeführt (s. Abschnitt 2.7, Pilottest, S. 37). Mit der zweiten, die durch die Hauptprobandengruppe repräsentiert wurde, fand der eigentliche Hauptversuch in einzelnen Sitzungen statt. Diese Hauptgruppe bestand aus 32 Probanden. Davon waren 16 Probanden Frauen, die andere Hälfte waren Männer, jeweils zwischen 20 und 30 Jahren mit Abitur, akademischer Ausbildung und audiometrisch überprüfter Normalhörigkeit.

2.10 Prosodietest

Alle 32 Probanden wurden auf freiwilliger Basis innerhalb eines Zeitraums von sechs Wochen einzeln getestet. Der Test bestand aus drei Teilen: Dem Tonaudiogramm, dem offenen Testteil und dem geschlossenen Test.

2.10.1 Aufstellung und Gerätschaften

Die Hörtests wurden in der Kabine C der Audiometrie in der HNO-Universitätsklinik Würzburg durchgeführt und es kamen folgende Geräte zum Einsatz (vgl. Mahfoud, 2009):

- IBM-kompatibler PC mit Soundkarte TYPHOON 3 D Acoustic Five Audio Card,
- WESTRA Audiometer CAD-03/1,
- WESTRA Audiometriebox LAB-1001.

Das Audiometer wurde auf Freifeldmessung eingestellt, um eine möglichst authentische Konstellation einer Kommunikationssituation zu schaffen. Der Schalldruck, mit dem die Wörter vorgespielt werden sollten, wurde auf 65 dB festgelegt, da dies der Lautstärke von Sprache in einem geschlossenen Raum entspricht. In 1 m Entfernung wurde dafür vorbereitend auf die spätere Position des Probanden ein Schalldruckpegelmessgerät von „Brüel & Kjær“ aufgestellt, mit dem der Schalldruckpegel (SPL) des über den Freifeldlautsprecher abgespielten CCITT-Rauschens gemessen wurde. Das Audiometer musste nominell auf 75 dB eingestellt werden, damit am Ort des Hörers 65 dB vorlagen. Die Einregelung fand an der Steuerungskonsole des Audiometers statt, am PC befanden sich alle Lautstärkeregelungen in maximaler Position.

2.10.2 Tonaudiogramm

Von jedem Probanden der Hauptgruppe wurde zu Beginn ein Tonaudiogramm erstellt. Hörverluste von 20 dB im Frequenzbereich von 250 Hz bis 4.000 Hz und Hör-

verluste von 40 dB in den Bereichen 125 - 250 Hz bzw. 4.000 Hz - 8.000 Hz durften dabei nicht überschritten werden, um eine Eignung des Probanden für den Prosodietest zu bestätigen. Diese Anforderung wurde von allen Probanden erfüllt, die Ergebnisse der Tonaudiometrie befinden sich im Anhang (s. Kapitel D, Messergebnisse der Tonaudiometrie, S. 86).

2.10.3 Durchführung des Prosodietests

Nachdem jeder Proband über den Ablauf und die Dauer des Test schriftlich und mündlich informiert und instruiert worden war, wurde er zunächst aufgefordert, 110 Sätze mit eigenen Worten zu beurteilen. Dabei kam es darauf an, den vermuteten emotionalen Gemütszustand des Sprechers zu benennen und dabei so wenige Worte wie möglich zu verwenden. Im zweiten Teil hörte sich jeder Proband nacheinander 330 Sätze an, wobei er diesmal die Emotion des Sprechers mit Hilfe einer Vorgebenliste auszuwählen hatte. Dabei waren die zehn möglichen Antworten schriftlich aufgelistet, jede Emotion beschrieben durch ein Wortpaar (s. Tab. 2.2). Es war jedem Probanden freigestellt, wann er eine Pause machen wollte und wie lange diese dauern sollte. Damit sollte vermieden werden, dass Probanden überbeansprucht oder aus der Konzentration gebracht wurden. Die vorgeschlagenen Pausen in der schriftlichen Instruktion waren nicht bindend. Die Sätze wurden manuell und pro Satz nur einmal abgespielt. Das Tempo wurde bestimmt durch die Antwortgeschwindigkeit der Probanden, die zum Nachdenken über Richtigkeit ihrer Antwort zwischen 0 Sekunden und 02 Minuten 55 Sekunden benötigten. Auch diese Bedenkzeit nach Anhörung eines Satzes war jedem Probanden vorbehalten. Weder Kommentare noch Korrekturen oder nonverbale Signale wurden den Probanden bezüglich ihrer Antworten übermittelt.

2.10.4 Erfassung der Evaluationsdaten

Für die Erfassung der Probandenantworten lag eine „OpenOffice-Calc“-Tabelle vor. Diese enthielt für jeden einzelnen Teilnehmer eine Auflistung aller Sätze, die diesem vorgespielt werden sollten, sodass die Promovendin stets im Blick hatte, ob der Proband gerade den Satz hörte, den er laut Abspielliste auch hören sollte. Außerdem eingetragen waren die laufende Numerierung sowie die dem Satz zugeordnete Nummer (1 - 44) und die Anzahl der Silben dieses Satzes. Hinter jedem Satz konnte in eine freie Spalte eingetragen werden, welche Emotion der Proband angegeben hatte. Im offenen Teil des Tests wurde die Antwort in Worten genau protokolliert, im geschlossenen Teil reichte die einbuchstabige Abkürzung aus, die vorher festgelegt worden war (s. Tab. 2.2). Außerdem wurden in dem Dokument vor jeder Sitzung Name und Geburtstag des jeweiligen Probanden vermerkt.

3 Ergebnisse

In der Durchführung der Versuche mit den 32 Hauptprobanden sind zahlreiche Daten gesammelt und einer statistischen Auswertung zugeführt worden. Um sie für den Leser dieser Arbeit komplett zugänglich zu machen, werden sie im Anhang (s. Kapitel 5, Anhang, S. 75 ff.) in Graphiken dargestellt. Die zehn Emotionen wurden in jeweils 44 Sätzen präsentiert, so dass jeder Proband 440 Sätze hörte. Insgesamt kamen 14080 Urteile zustande, davon 3.520 in offener und 10.560 in geschlossener Fragestellung. Jede Satz-Emotions-Kombination wurde acht Mal offen und 24 Mal geschlossen beurteilt. Für den geschlossenen Teil gilt demzufolge, dass pro Emotion 1056 Urteile vorliegen. Dies ergibt 240 Urteile pro Satz. Im offenen Teil liegen dementsprechend 352 Urteile pro Emotion und 80 pro Satz vor.

3.1 Pilottest: Reduktion auf 10 Emotionen

Die Auswertung der Ergebnisse des Pilottests beschränkte sich auf die Auszählung und Zusammenfassung aller Antworten der Pilotprobanden ohne Berücksichtigung anderer Parameter. Diese Auswertung erfolgte mit Hilfe einer „OpenOffice-Calc“-Tabelle, in der für jede der 15 dargebotenen Emotionen alle abgegebenen Antworten miteinander verglichen wurden, um sie daraufhin ihrer absteigenden Häufigkeit nach zu sortieren. Da im Pilottest 30 Sätze vorgespielt wurden und 15 Probanden teilnahmen, kam jede Emotion zweimal vor und wurde dementsprechend 30 Mal beurteilt. Das Ergebnis wurde verwendet, um diejenigen Emotionen zu identifizieren, die im

3 Ergebnisse

Hinblick auf den Test mit der Hauptgruppe verworfen werden können (s. Abschnitt 2.7, Pilottest, S. 37) und auch um Hinweise auf bestimmte Begriffe zu bekommen, die häufig von den Pilotprobanden zur Beschreibung der gehörten Emotion gewählt wurden. Solche Begriffe kamen in Frage, später als Vorgaben für den geschlossenen Test bei der Hauptgruppe übernommen zu werden. Die intendierten Emotionen, die von den Probanden als solche erkannt wurden, waren:

- Zorn,
- Trauer,
- Freude,
- Verachtung,
- Aggression,
- Gleichgültigkeit,
- Panik.

Emotionen, für deren Erkennen durch die Probanden kein Hinweis erbracht werden konnte, waren:

- Erleichterung,
- Ekel,
- Begeisterung,
- Furcht,
- Mitleid,
- Stolz,
- Schüchternheit,
- Überraschung.

Die drei Emotionen „Ekel“, „Stolz“ und „Überraschung“ wurden dabei jedoch mit weitgehender Übereinstimmung innerhalb der Probandengruppe andersartig beschrieben. Ein Großteil der Probanden hielt die Gemütsverfassung „angeekelt“ für eine

„weinerliche“, „Stolz“ wurde als „ärgerliche Erregung“ und „Nachdrücklichkeit“ interpretiert und die „Überraschung“ als „Neugierde“ oder „Unsicherheit“. Da der Pilottest u. a. darauf abzielte, mangelhaft dargestellte bzw. schlecht zu erkennende Emotionen zu identifizieren und ggf. auszusondern, fiel die Entscheidung auf das Verwerfen der fünf Emotionen „Erleichterung“, „Begeisterung“, „Furcht“, „Mitleid“ und „Schüchternheit“. In den Probandenantworten auf diese Emotionen kam weder ein Konsens zustande, noch wurden die intendierten Gemütsbewegungen richtig erkannt. In der Tabelle 3.1 wird zusammengefasst, welche die am häufigsten genannten Begriffe für jede der zehn selektierten Emotion waren.

3.2 Offener Test

Die Ergebnisse des offenen Testteils, die sich aus den freien Antworten ergeben, sind vermutlich stark probandenabhängig. Im geschlossenen Teil lag eine Einschränkung durch vorgegebene Antwortoptionen vor und somit eine Möglichkeit des Ratens (Trefferwahrscheinlichkeit = 1/10). Daher konnte die wahre Leistungsfähigkeit in der

Dargebotene Emotion	Häufigste Probandenantworten
A Aggression	genervt, aufgebracht, ruhig
B Überraschung	neugierig, unsicher, verwundert
E Ekel	traurig, lustig, weinerlich, angeekelt
F Freude	fröhlich, erfreut
G Gleichgültigkeit	gelangweilt, traurig
P Panik	aufgeregt, erregt
S Stolz	erregt, ärgerlich, nachdrücklich
T Trauer	traurig, depressiv
V Verachtung	herablassend, überheblich
Z Zorn	ärgerlich, verärgert

Tabelle 3.1: Häufigste Antworten im Pilottest

3 Ergebnisse

Erkennung der Emotionen nicht sicher beurteilt werden. Mit Hilfe des offenen Teils lässt sich ein Vergleich ziehen, denn er zeigt, ob es zwischen den Beurteilungen der beiden Testteile eine Deckungsgleichheit gibt. Ebenfalls werden Abweichungen zwischen den intendierten Emotionen und den Interpretationen durch die Probanden im offenen Teil deutlich. Jedem Probanden war es freigestellt, das Gehörte in einem oder mehreren Worten zu beschreiben. So kam eine Antwortliste mit insgesamt 3.520 Einzelurteilen zustande (352 pro Emotion), wobei jedes Einzelurteil aus einem oder mehreren Begriffen bestehen konnte. In den Antworten der Probanden tauchen allerdings auch Benennungen auf, die weniger einem Gemütszustand als vielmehr der Beschreibung eines physischen Zustands gleichen oder rhetorische Besonderheiten des Sprechers aufgreifen. Einige Antworten bestanden dabei nicht aus einem oder mehreren Einzelbegriffen, sondern aus beschreibenden kurzen Sätzen, wie z. B. „fühlt sich überlegen“ oder „positive Nachricht überbringend“. Solcherlei Antworten mussten in einen auswertbaren Begriff überführt werden. Dies geschah mit Hilfe des Online-Synonym-Wörterbuches der Universität Leipzig (Universität Leipzig, 2010). Die einzelnen Schritte der Evaluation wurden in einer „OpenOffice-Calc“-Tabelle umgesetzt. Zunächst wurden diejenigen Antworten als korrekt eingestuft, die wortwörtlich mit der vorgegebenen Emotion übereinstimmten. Wenn dies nicht der Fall war, fand ein Abgleich mit dem Synonymwörterbuch statt.

Auf der ersten Ebene wurden die Antworten als richtig gewertet, deren direktes Synonym ein Wort jenes Wortpaares war, das als Vorgabe für die jeweilige Emotion galt. Bei mehr als einer Antwort pro Urteil wurde, falls vorhanden, derjenige Begriff ausgewertet, der mit der intendierten Emotion übereinstimmte. Dieses Vorgehen findet Begründung in der Annahme, dass ein Proband unter Beweis stellt, die Emotion erkannt zu haben, sobald er den korrekten Begriff nennt, auch wenn dieser begleitet ist von anderen, nicht zutreffenden Bezeichnungen. Dabei ist die Wahrscheinlichkeit

für den Probanden, durch Raten auf die richtige Antwort zu kommen, verschwindend gering. Römer und Matzke (2005) zufolge umfasst der Begriffe-Pool eines Durchschnittssprechers der deutschen Sprache 6000 - 10.000 Wörter. Von den Teilnehmern des Haupttests wurden bis auf insgesamt drei Ausnahmen niemals mehr als drei Begriffe pro Stimulus angegeben.

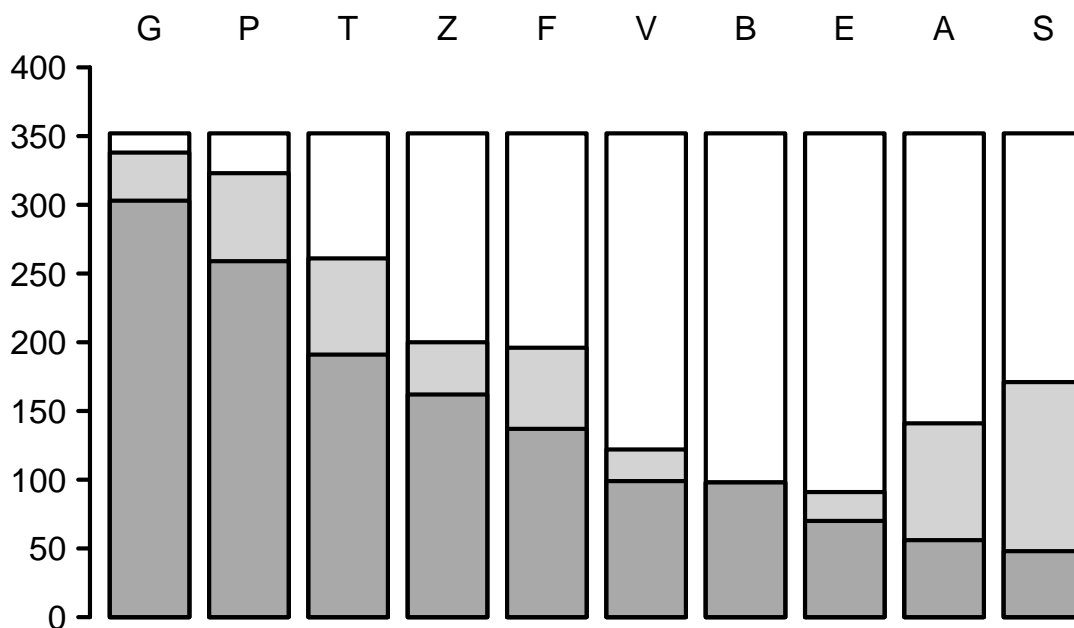


Abbildung 3.1: Erkennungsleistung aller Probanden im offenen Test. Übereinstimmung wortwörtlich oder via direktes Synonym (dunkelgrau), Übereinstimmung via „Brückensynonym“ (hellgrau), Keine Übereinstimmung (weiß).

Auf der zweiten Ebene gab es Wortnennungen der Probanden, die nicht direkt, sondern über ein anderes Synonym dem vorgegebenen Emotionsbegriff zugeführt werden konnten. Beispielsweise ist die Beschreibung „missgelaunt“ weder ein Synonym von „zornig“, noch von „verärgert“. Das Wort „ärgerlich“ hingegen ist ein Synonym von beiden, so dass es als überbrückendes Synonym gelten kann. Diese Art der Zuordnung war für eine beträchtliche Anzahl an Probandenurteilen möglich. Auf diese Weise konnten Probandenantworten, die sinngemäß mit den intendierten Emotionen übereinstimmten, mit berücksichtigt werden. Alle Antworten, die auch über ein

3 Ergebnisse

solches „Brückensynonym“ nicht übereinstimmend zugeordnet werden konnten, wurden als falsche Antwort gewertet. Im zweiten Schritt der Auswertung wurde gezählt, wie oft die intendierten Emotionsbegriffe wortwörtlich genannt wurden. Ebenfalls ausgezählt wurden die am häufigsten genannten Begriffe für jede Emotion, diese wurden auf sinngemäße Übereinstimmung mit dem vorgegebenen Begriff geprüft. Aus dieser Auswertung ergeben sich Hinweise auf mögliche Mängel bei der Wahl der Emotionsbezeichnung oder dass es nicht ausreichend gelungen ist, die intendierte Emotion überzeugend sprachlich darzustellen (vgl. Abschnitt 4.2.1 (Einfluss der Versuchsplanung, S. 64)). Ein eindrucksvolles Beispiel für solch einen Fall liefert die Gemütsregung „weinerlich/angeekelt“. Hier interpretiert die Mehrheit der Probanden den Stimulus in einer gänzlich abweichenden Richtung, der Sprecher wird hier am häufigsten als „belustigt“ oder „amüsiert“ bezeichnet. Hier lässt sich eine deutliche Abweichung von den Ergebnissen des Pilottests erkennen (s. Abschnitt 3.1, Pilottest: Reduktion auf 10 Emotionen, S. 45 f.), jedoch eine Korrelation mit den Resultaten aus dem geschlossenen Test (s. Abschnitt 3.3, Geschlossener Test, S. 51 f.). Die Graphiken zu den meistgenannten Begriffen befinden sich im Anhang (s. Abb. E.1 bis E.10). Die Abbildung 3.1 zeigt in Zahlen, welche Emotionen am besten erkannt wurden. Dabei kennzeichnet der dunkelgraue Abschnitt die wortwörtlichen Übereinstimmungen bzw. direkten Synonyme, der hellgraue diejenigen Antworten, die via Synonym der vorgegebenen Emotion zuzuordnen sind und der weiße Balkenteil die inkongruenten Antworten. Diese Graphik lässt sich mit dem globalen Blasendiagramm in der Abbildung 3.2 aus dem geschlossenen Test vergleichen. Dabei ergibt sich eine weitreichende Deckungsgleichheit zwischen den Erkennungsleistungen bezüglich der dargebotenen Emotionen. Dieses Ergebnis lässt den Rückschluss zu, dass die Trefferquote im geschlossenen Test kein Zufallsereignis ist und nicht durch ausschließliches Raten (Trefferwahrscheinlichkeit = 1:10) zustande kommt.

3.3 Geschlossener Test

Die Auswertung der Evaluationsergebnisse aus dem geschlossenen Testteil erfolgte mit dem Statistikprogramm *R* (Hellbrück, 2009). Durch die gezielte Analyse der fünf Einflussfaktoren

- Proband (01 - 32),
- Probandengeschlecht (m/w),
- Satz (01 - 44),
- Anzahl der Silben (10/15) und
- Emotion (A, B, E, F, G, P, S, T, V, Z)

konnten folgende Zusammenhänge und Abhängigkeiten untersucht werden:

- Erkennen der Emotionen abhängig vom Probanden,
- Erkennen der Emotionen abhängig vom Geschlecht der Probanden,
- Erkennen der Emotionen abhängig vom Satz,
- Erkennen der Emotionen abhängig von der Anzahl der Silben eines Satzes,
- Erkennen der Emotionen abhängig von der jeweils dargebotenen Emotion.

Bei der Erarbeitung einer Statistik mit den angeführten fünf Einflussfaktoren ist zu beachten, dass sie nicht alle demselben Skalentyp zuzuordnen sind. Die Probanden tragen das dichotome und nominalskalierte Merkmal „Geschlecht“. Die Sätze enthalten sowohl dichotome Merkmale (10 oder 15 Silben) als auch polytome Merkmale (Emotion A, B, E, F, G, P, S, T, V oder Z), die dem jeweiligen Satz zugeordnet sind. Die Anzahl der Silben gehört ebenfalls der nominalen Skala an, denn sie finden in diesem Kontext keine ordinale oder kardinale Verwendung dahingehend, dass beispielsweise eine höhere Silbenanzahl als höherwertiges Merkmal einzustufen ist. Auch die Emotionen selbst sind nominalskaliert (Bortz und Döring, 2002).

Zur Signifikanzprüfung, in diesem Fall bei dichotomer Datenausprägung (Emotion erkannt/nicht erkannt), kam der Chi-Quadrat-Test zum Einsatz. Das resultierende be-

3 Ergebnisse

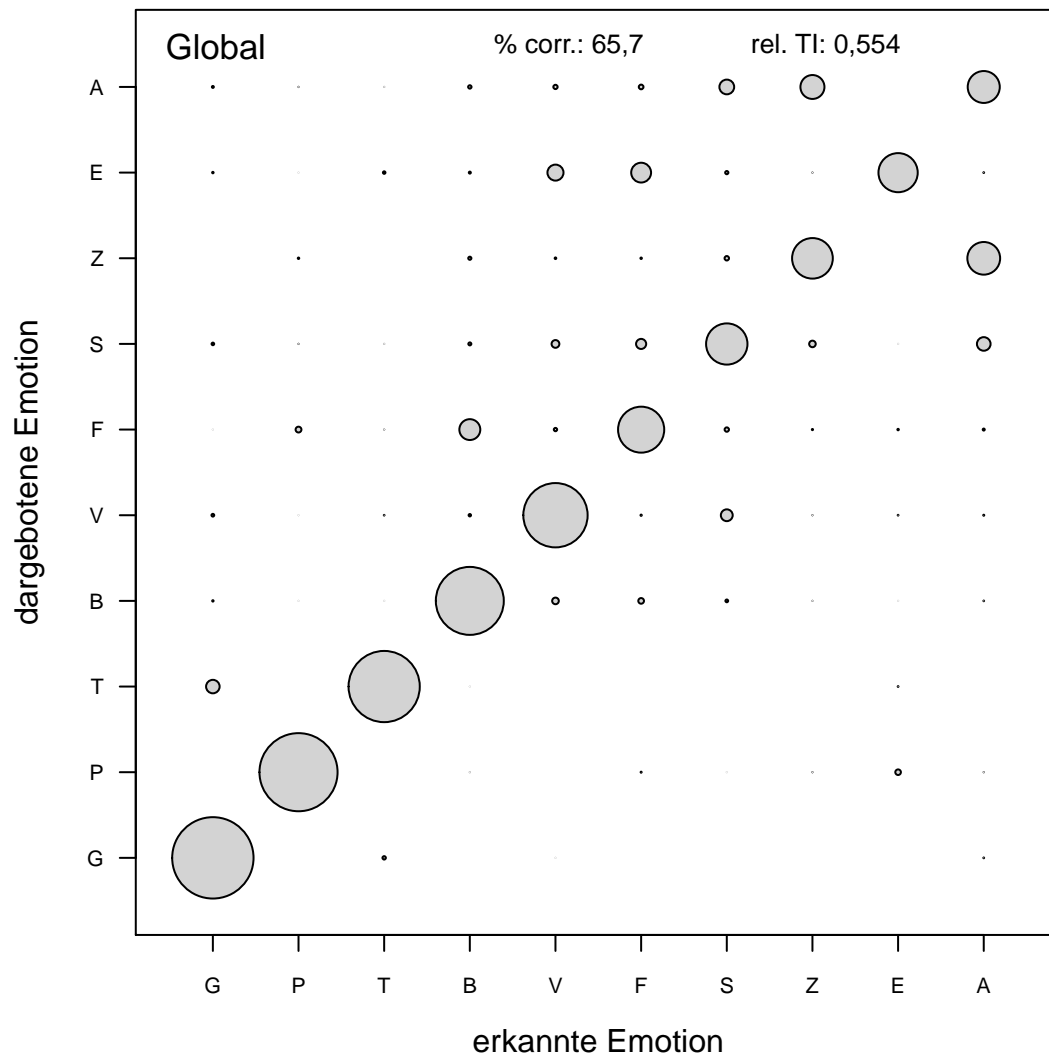


Abbildung 3.2: Erkennungsleistung und Vertauschungen aller Emotionen durch die Probanden

rechnete Signifikanzniveau trifft eine Aussage darüber, wie hoch die Wahrscheinlichkeit ist, dass die Unterschiede der Antworthäufigkeiten nur zufällig zustande kommen (Irrtumswahrscheinlichkeit). Je geringer das Signifikanzniveau, desto wahrscheinlicher ist tatsächlich vorhandene Erkennungsleistung (Kuckartz et al., 2010). Die Abbildung 3.2 zeigt die Absolutwerte der Häufigkeiten der richtigen Antworten aller Probanden. Die Emotionen sind nach der Häufigkeit der richtigen Antworten angeordnet,

der Durchmesser einer Blase gibt Aufschluss über die Anzahl der korrekten Beurteilungen im Einzelnen.

Da die 440 Sätze insgesamt verwürfelt wurden, lagen im Ergebnis zwar Listen vor, in denen jede Satz-Emotions-Kombination nur einmal vorkam und jeder Proband jede Emotion und jeden Satz gleich oft hörte, dies galt jedoch nur für die Betrachtung über alle 440 Sätze. Beachtet man die Tatsache, dass der Test in einen offenen (110 Sätze) und einen geschlossenen Teil (330 Sätze) gegliedert ist, dann trifft diese gleichmäßige Verteilung für jeden einzelnen Abschnitt nicht mehr zu, was sich insbesondere in der Auswertung des geschlossenen Teils auswirkte. Bei der Auszählung ergab sich, dass jede Emotion entweder 8, 9, 10 oder 11 Mal gehört wurde, je nachdem, wie häufig sie schon im offenen Teil evaluiert worden war. Im Kapitel 4 (Diskussion, S. 63), Abschnitt 4.2.1 (Einfluss der Versuchsplanung, S. 64) wird dies noch einmal erläutert. Der beschriebenen Diskrepanz wurde mit prozentualer Gewichtung der richtigen und falschen Antworten begegnet, so dass sich die oben genannte Ungleichverteilung zwischen den Probandengruppen nicht auf die Blasendurchmesser auswirkte. Schon auf den ersten Blick ist erkennbar, dass es Emotionen gibt, die insgesamt besser erkannt wurden als andere. Dies sind in erster Linie G, P und T, sie wurden dementsprechend selten mit anderen verwechselt. Emotionen, die nie miteinander vertauscht wurden, waren:

- G und P,
- G und Z,
- T und Z,
- T und P.

Häufig verwechselt wurden:

- A mit S,
- A mit Z,

3 Ergebnisse

- E mit V,
- E mit F,
- Z mit A,
- S mit A,
- F mit B,
- V mit S,
- T mit G.

In der Graphik 3.2 fällt auf, dass es zwei Emotionen gibt, die konsequent von allen Probanden miteinander verwechselt wurden: Z und A. Dies ließ sich bereits während der Probandenversuche dadurch erkennen, dass die Probanden ihre Unsicherheit bezüglich der Unterscheidung dieser beiden Emotionen äußerten und wird in der graphischen Darstellung deutlich sichtbar. An dieser Stelle bietet sich ein Vergleich mit den Ergebnissen der Untersuchungen von Scherer (1982) an. Die drei bei ihm am besten erkannten Emotionen waren nach absteigender Häufigkeit: Ärger, Traurigkeit und Gleichgültigkeit. Bei einer gedankenexperimentellen Zusammenführung der Emotionen Z und A zu einer Emotion „Ärger“, zeigen sich in dieser Arbeit ganz ähnliche Ergebnisse.

Außerdem lässt sich feststellen, dass bei P eine Verwechslung mit E verhältnismäßig häufiger vorkam als mit anderen Emotionen. In nachträglichen Gesprächen zwischen Probanden und der Promovendin stellte sich heraus, dass die intendierte weinerlich-angeekelte Stimmlage des Sprechers für ein glucksendes Lachen gehalten wurde. In diesen Fällen war die Zuordnung konsequent und in der logischen Folge wurde die Emotion F als eine andere angenommen, oftmals war dies B, möglicherweise aufgrund der hier ebenfalls erhöhten Stimmfrequenz (vgl. Abschnitt 3.2 (Offener Test, S. 47 f.)).

In der Darstellung 3.3 sind die Transinformationsindizes ersichtlich. Der Transinformationsindex, der gleichbedeutend ist mit dem Begriff der „relativen Transinfor-

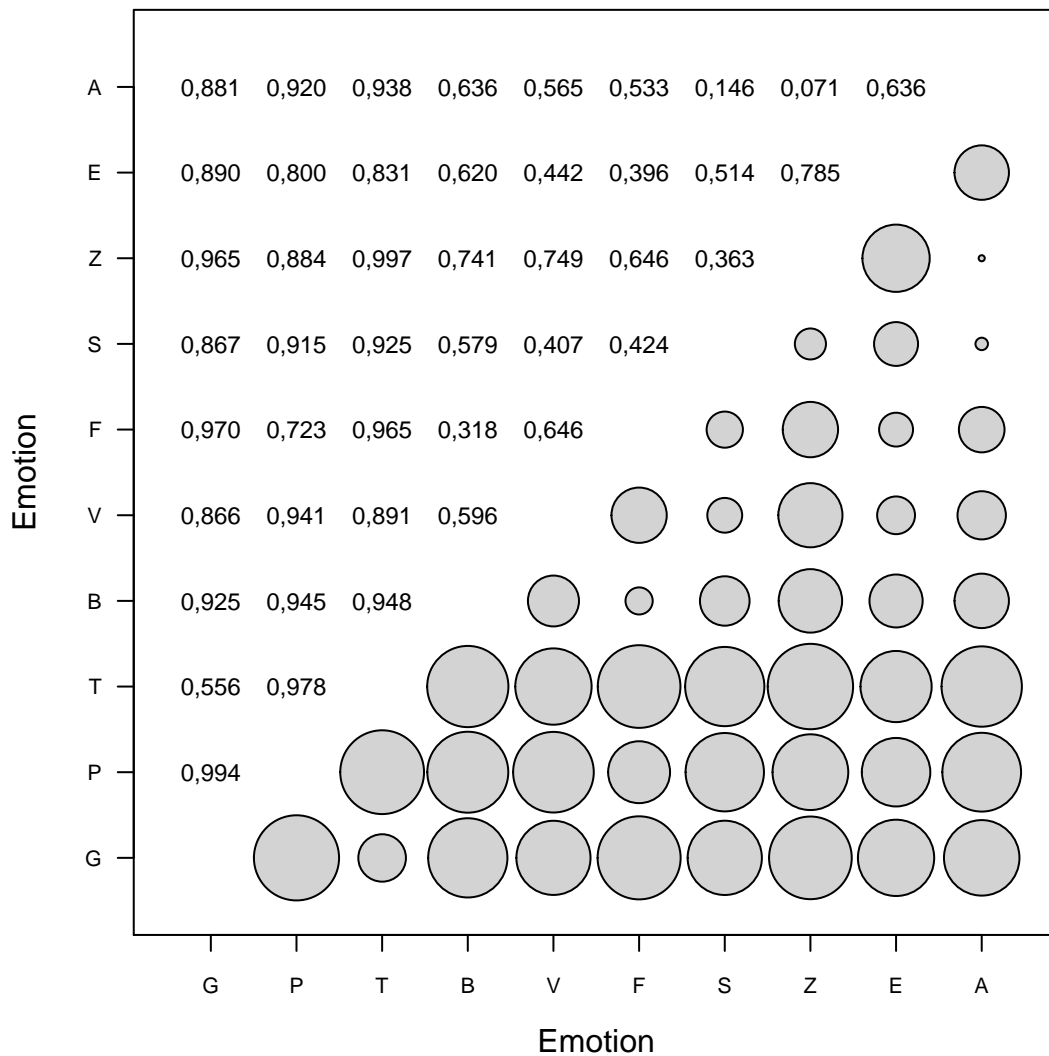


Abbildung 3.3: Transinformationswert der Erkennungsleistung und Vertauschungen aller Emotionen durch die Probanden

mation“ aus der Nachrichtentechnik, stellt dar, wie groß der Anteil einer gesendeten Information ist, der beim Empfänger ankommt. Ein Beispiel soll dies verdeutlichen:

Eine gedachte Quelle sendet ausschließlich die Nachrichten A und B zu gleichen Teilen. Kommt beim Empfänger jede Nachricht korrekt an, so liegt der TI-Wert bei 1, der Prozent-korrekt-Wert bei 100. Wird jede Nachricht A jedoch als B und jede Nachricht B als A empfangen, dann liegt der TI-Wert ebenfalls bei 1, denn es hat auch hier

3 Ergebnisse

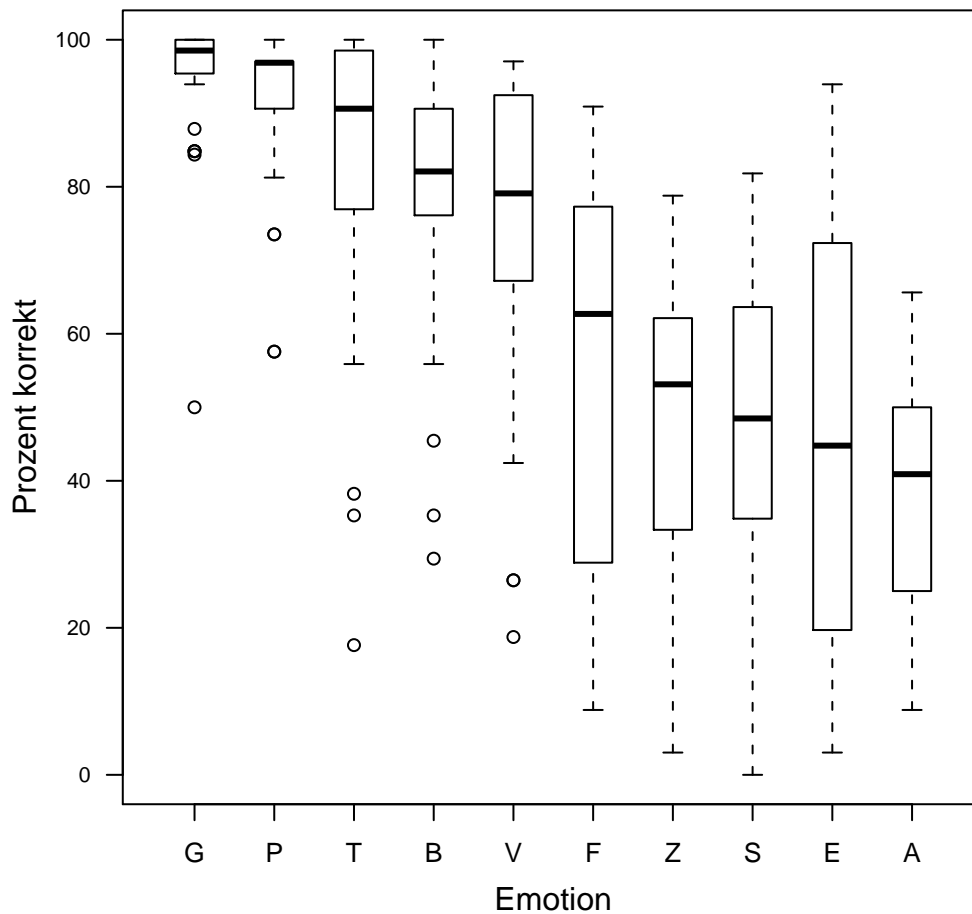


Abbildung 3.4: Erkennungsleistung der Probanden in Abhängigkeit von der angebotenen Emotion (in % korrekt)

eine vollständige Transmission stattgefunden. Durch eine einfache Umdefinition der Nachrichten wäre eine Rekonstruktion möglich. Der Prozent-korrekt-Wert liegt hier jedoch bei 0. Werden A und B hingegen nur zufällig zur Hälfte richtig oder falsch empfangen, so liegt der TI-Wert bei 0, denn es ist keinerlei Information des Eingangs zuverlässig auf den Ausgang übertragen worden. Der Prozent-korrekt-Wert läge in diesem Fall bei 50 (Wendler et al., 2005).

Die Graphik ist als spiegelbildliche Darstellung zu verstehen und zeigt die relative Transinformation als Maß für die Stärke der Störung des Übertragungskanal, immer zwei Emotionen betreffend. Sie bietet in Zahlen Aufschluss darüber, wie jeweils zwei

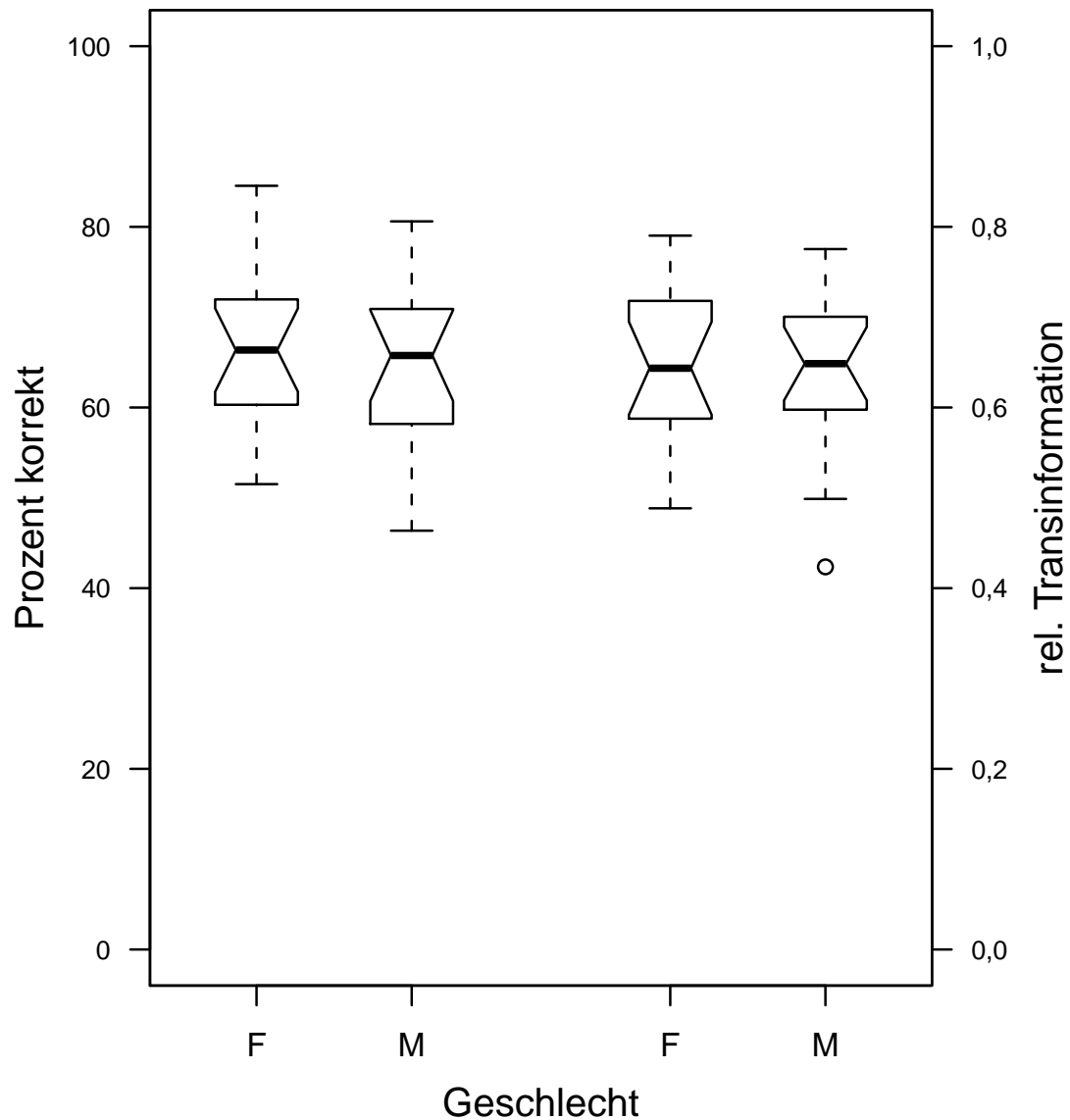


Abbildung 3.5: Erkennungsleistung der Probanden in Abhängigkeit vom Geschlecht

Emotionen miteinander verwechselt wurden bzw. wie häufig eine Emotion richtig erkannt wurde.

Die Abbildung 3.4 stellt die Streuungen in der Emotionserkennung noch einmal in Boxplots dar. Wie das globale Blasendiagramm (s. Abb. 3.2) bereits zeigte, sind die Streuungen bei den Emotionen A, E, F, S und Z besonders hoch.

3 Ergebnisse

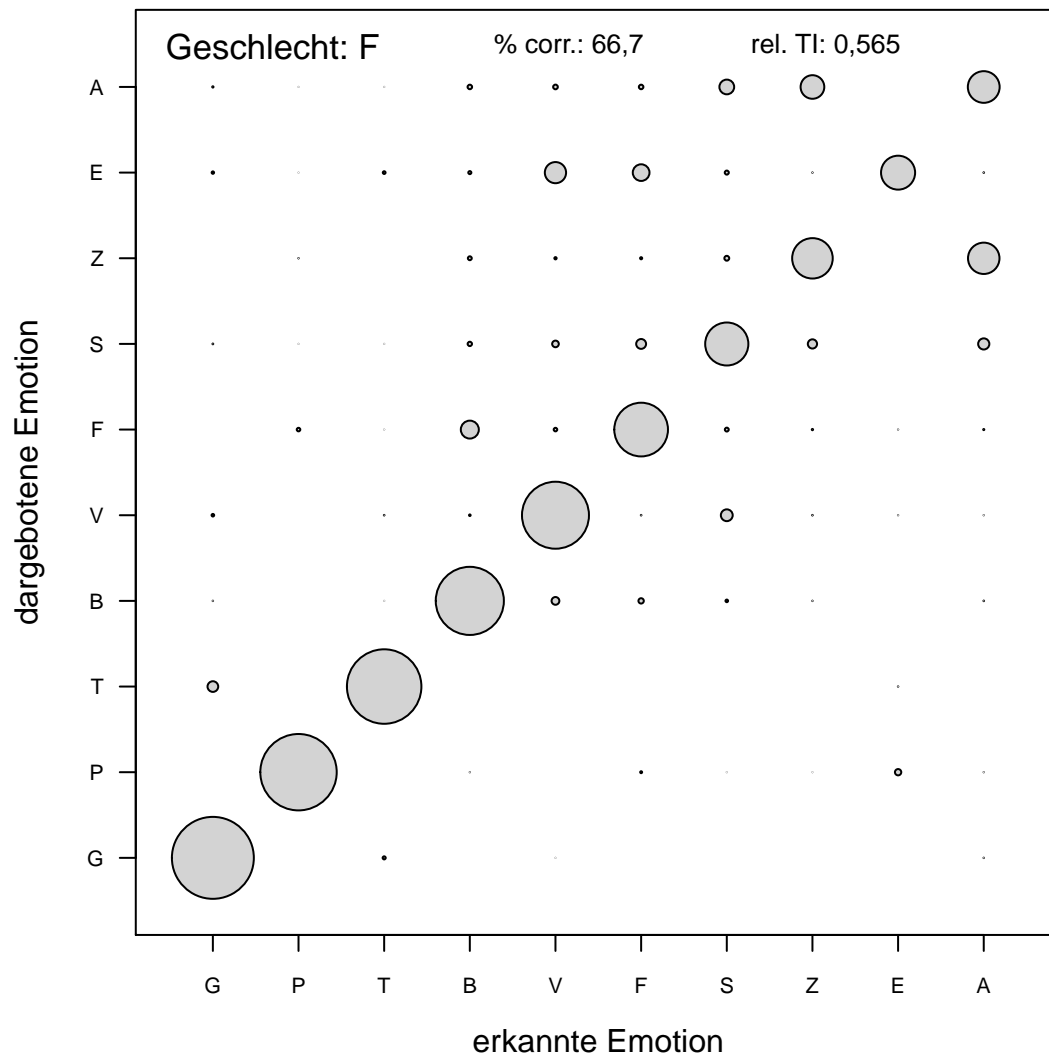


Abbildung 3.6: Erkennungsleistung der Probanden weiblichen Geschlechts

Ohne Einfluss auf die Emotionserkennung ist dagegen das Geschlecht der Probanden. In der Darstellung 3.5 treffen die Boxplots u.a. eine Aussage darüber, wie groß die Unterschiede in der Streuung der Emotionserkennung zwischen männlichen und weiblichen Probanden sind. Diese sind auf der linken Seite dargestellt als prozentuale Anteile aller richtig erkannten Emotionen und auf der rechten Seite als TI-Werte. Zur Signifikanzprüfung kam der Wilcoxon-Rangsummentest zum Einsatz. Der p-Wert der TI-Werte lag bei 0,9556, derjenige der Prozent-korrekt-Werte bei 0,534.

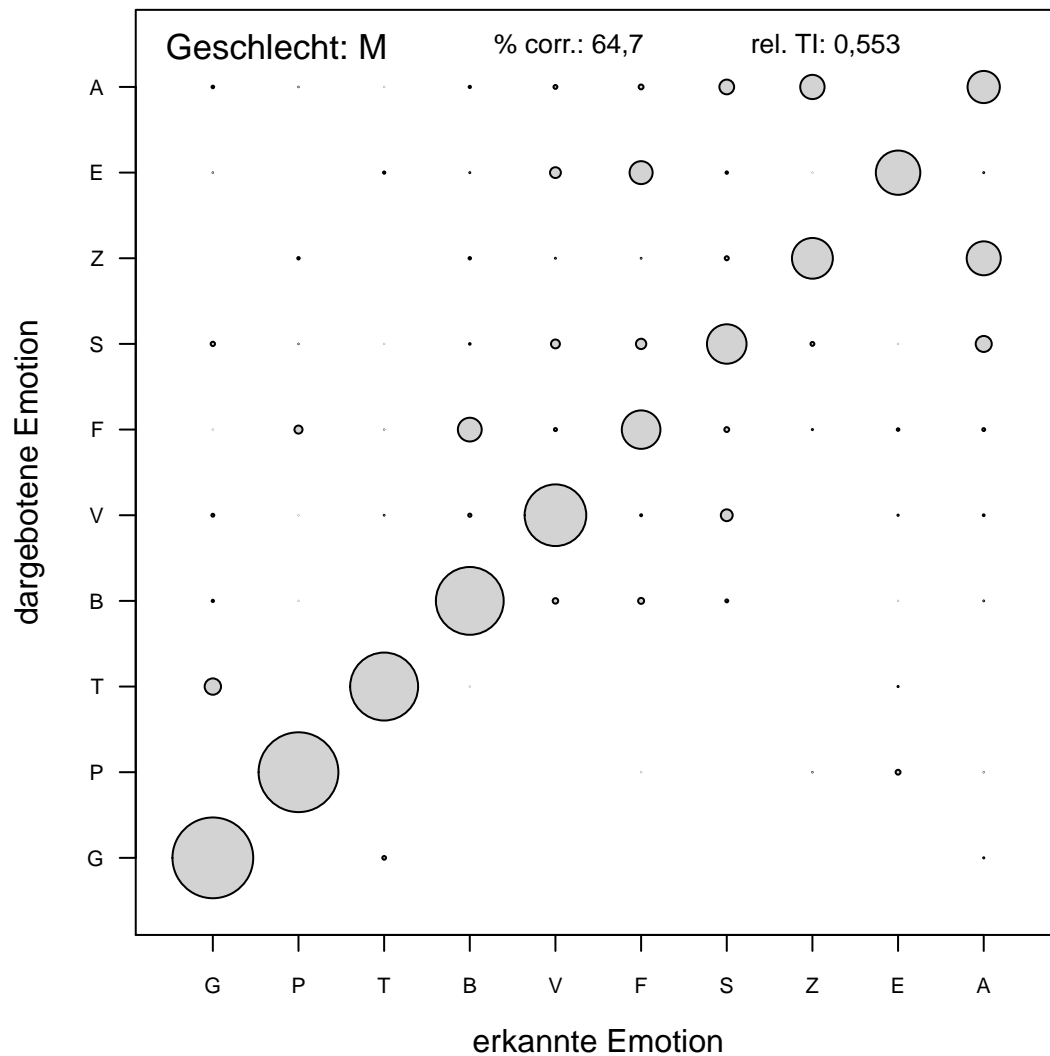


Abbildung 3.7: Erkennungsleistung der Probanden männlichen Geschlechts

Die Nullhypothese, dass es keinen Unterschied zwischen den Geschlechtern bezüglich der Emotionserkennung gibt, konnte somit nicht verworfen werden. Der Vergleich ist im Einzelnen in den Graphiken 3.6 und 3.7 zusammengestellt, hier beträgt der Unterschied zwischen den Prozent-korrekt-Werten zwei Prozentpunkte zugunsten der Probandinnen. Das Emotionserkennen der beiden Geschlechter weist kaum Unterschiede auf. Im Vergleich wurde E von den weiblichen Teilnehmern häufiger für V und S für Z gehalten, während die männlichen Probanden E öfter für F hielten, F für

3 Ergebnisse

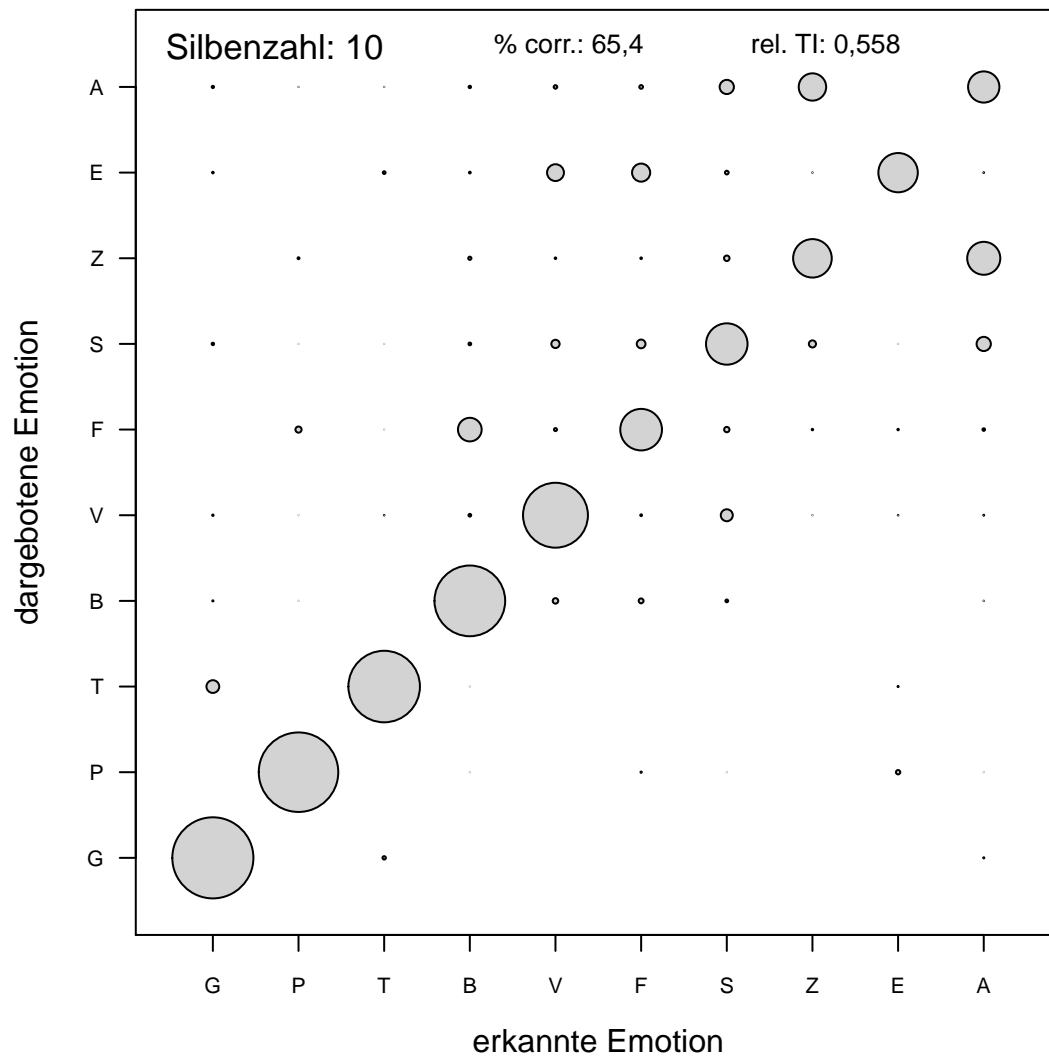


Abbildung 3.8: Erkennungsleistung der Probanden bei 10-silbigen Sätzen

B und T für G.

Die Anzahl der Silben übt ebenfalls keinen Einfluss auf das Erkennen der Emotionen aus, der Chi-Quadrat-Test ergab ein Signifikanzniveau von $p=0,594$. Graphisch dargestellt ist dies in den Abbildungen 3.8 und 3.9. Die Aufschlüsselung der Erkennungsleistungen für jeden einzelnen Satz befindet sich im Anhang in den Graphiken G.1 bis G.6. Auch hier zeigt das Ergebnis des Chi-Quadrat-Tests von $p=0,206$, dass die Emotionserkennung satzunabhängig zu betrachten ist. Der nicht-

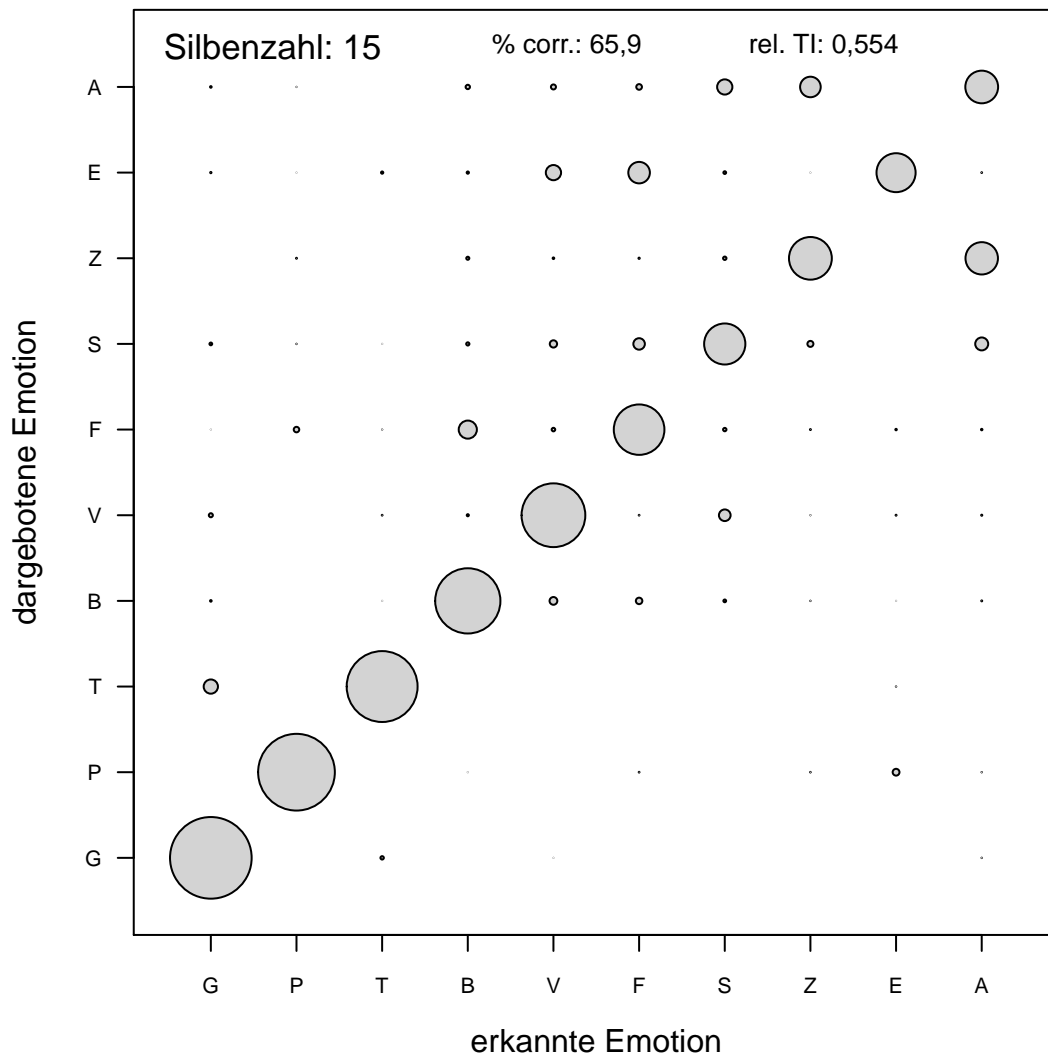


Abbildung 3.9: Erkennungsleistung der Probanden bei 15-silbigen Sätzen

parametrische Kruskal-Wallis-Test wurde zusätzlich zur Signifikanzprüfung der Prozent-korrekt-Werte und der TI-Werte der Sätze herangezogen und ergab $p=0,140$ bzw. $p=0,494$. Bezüglich der Anzahl der Silben und der einzelnen Sätze wäre ein Einfluss durch die Länge des Satzes möglich gewesen, weil der Sprecher in 15-silbigen Sätzen zeitlich betrachtet mehr Gelegenheit hatte, die Emotion möglichst glaubhaft darzustellen als in den kürzeren Zehnsilbern. Den Graphiken ist zusätzlich zu entnehmen, dass in einem Erkennungsbereich von 57,1 % (Satz 44) bis maxi-

3 Ergebnisse

mal 73,3 % (Satz 35) beide Sätze zu den 15-Silbern gehören. Weiterhin kann noch betrachtet werden, ob die Probanden in ihrem relativen Erkennen oder Nichterkennen bzw. Verwechseln der Emotionen untereinander übereinstimmten. Die Blasen-diagramme jedes einzelnen Probanden befinden sich im Anhang (s. Abb. F.1 bis F.4). Sie zeigen, dass die Stimuli zu maximal 84,5 % richtig erkannt wurden bei einem TI-Wert von 0,79 (Probandin 4), die niedrigste Erkennungsquote liegt bei 46,4 % und einem entsprechenden TI-Wert von 0,424 (Proband 29).

4 Diskussion

4.1 Erfolge der Untersuchungen

Die Versuche zur Erkennbarkeit der Prosodie bzw. Emotionen in deutscher Pseudosprache haben gezeigt, dass normalhörende Personen in der Lage sind, schauspielerisch dargestellte Emotionen in Form von inhaltsleeren Sätzen richtig zu erkennen. Trotz der nachstehend aufgeführten potentiellen Fehlerquellen hatten die Probanden im Gesamtergebnis eine Trefferquote, die über bloße Zufallserfolge weit hinausgeht. Insofern ist dieser Satzverständnistest geeignet, nach geringfügiger Weiterentwicklung in der klinischen Forschung angewendet zu werden. Diese Untersuchungen knüpfen thematisch an die Prosodietests der Masterarbeit und der Dissertation von Bauer und Domes an. Bauer forderte bereits im Jahr 2005 für nachfolgende Arbeiten zum einen die Erhöhung der Probandenzahl und die Integration von Pseudowörtern, zum anderen die Einbeziehung weiterer Aspekte der Prosodie. Dies ist in der vorliegenden Arbeit erfolgreich umgesetzt worden.

4.2 Fehlerquellen

Dennoch war die Versuchsplanung und Auswertung aufgrund bisher unerprobter Methoden komplex, daher ist die Zielsetzung weiterer Untersuchungen das Aufgreifen sowie der Ausschluss von Kritikpunkten.

4.2.1 Einfluss der Versuchsplanung

Sowohl die Auswahl der Emotionen selbst, als auch deren sprachliche Umsetzung oblag vollständig den Durchführenden. Dies bedeutet, dass die aufgezeichneten Sätze und Emotionen subjektiv geprägt sind. Es gibt viele Möglichkeiten, eine Emotion prosodisch darzustellen und auch das Gelingen dieser Umsetzung ist abhängig von der Ausführung durch den Sprecher. In den Testergebnissen und im Dialog mit den Probanden zeigte sich, dass die Auswahl der vorgespielten Emotionen optimiert werden kann. Die Wortliste, welche den Probanden im zweiten Versuchsteil vorlag, enthielt für jede Emotion zwei Begriffe, die einander ergänzen sollten (s. Tab. 2.10). Dies hatte den Vorteil, dass den Probanden die Zuordnung des Stimulus zu einer der vorgegebenen Emotionen erleichtert wurde und die Antwortmöglichkeiten für die Begriffe limitiert waren. Diese zweifache Benennung sollte nach Meinung der Autorin jedoch zukünftig durch eine einfache ersetzt werden. Gründe hierfür sind die übersichtlichere Gestaltung der Emotionsliste und die exaktere Festlegung auf eine Emotionsbenennung. Dies ist bei der Weiterentwicklung des Prosodietests möglich, da mit der Auswertung dieser Arbeit die benötigten Informationen über häufig genannte Begriffe und das Erkennen bzw. Nichterkennen prosodischer Strukturen vorliegen. Weiterhin zu erwähnen ist die im Abschnitt 3.3 (Geschlossener Test, S. 51) erläuterte Art der Zufallsverwürfelung und Satzaufteilung. Hier kann Abhilfe geschaffen werden, indem der offene und der geschlossene Abschnitt separat verwürfelt werden. Auf diese Weise kann gewährleistet werden, dass in jedem der beiden Abschnitte jede Emotion gleich häufig vorkommt. Ein möglicher Einfluss auf die Emotionserkennung bestand in der ungleichen Chancenverteilung zwischen den acht Probandengruppen. Sie hatten jeweils mit unterschiedlicher Häufigkeit Gelegenheit, die Sätze anzuhören und zu beurteilen. Dieser Einfluss kann jedoch ausgeschlossen werden, da die Gesamtanzahl sowohl der Probanden als auch der Satz-Emotions-

Kombinationen in jeder Gruppe ausreichend hoch war. Außerdem lassen sich beim Vergleich der Ergebnisse, die aus den Blasendiagrammen und ihren Zahlenwerten im Anhang (s. Abb. F.1 bis F.4) hervorgehen, keine signifikanten Unterschiede zwischen den acht Gruppen feststellen.

4.2.2 Einfluss des Sprechers

Die sprachliche Gestaltung der Pseudosätze war für den Sprecher eine neuartige Aufgabe. Dem Sprecher war es zwar möglich, für die prosodische Umsetzung von Emotionen in der Sprache Hinweise und Anregungen aus den visuellen und auditiven Medien sowie aus der eigenen Erfahrung zu sammeln und einfließen zu lassen. Trotzdem ist eine auf schauspielerische Weise dargestellte Emotion nur fiktiv und nicht tatsächlich vorhanden. Alle physiologischen Abläufe, die sich aus einer empfundenen Emotion ergeben und sich auf den Stimm-, Sprech- und Atemapparat auswirken, können nur imitiert werden (Stemmler, 2009). Dabei sind die Grenzen durch die Fähigkeiten des Schauspielers gesetzt. Hierbei spielt es jedoch keine tragende Rolle, ob er eine professionelle Ausbildung für solche Aufgaben erhalten hat, da auch Sprecher und Schauspieler, die in den audiovisuellen Medien auftreten, eine solche Ausbildung nicht immer vorweisen können (z. B. Kinder). Im Falle dieser Arbeit verfügte der Sprecher (Dipl.-Ing. Stefan Brill) nicht über professionelle schauspielerische Fähigkeiten, dafür jedoch über Erfahrungen in der erfolgreichen Aufsprache prosodischer und intonatorischer Variationen von Sätzen aus der Magisterarbeit von Bauer (2005). Trotzdem sollte als Verbesserungsmöglichkeit der Einsatz professioneller Schauspieler erwogen werden, um einen Einfluss auf das Ergebnis verifizieren bzw. ausschließen zu können. Während der Aufsprachen erwies es sich bei einigen Emotionen als Herausforderung, die zuvor eingeübte Intonation während der gesamten Aufnahme beizubehalten. Um diesem Anspruch adäquat zu begegnen war

4 Diskussion

es notwendig, einige Aufnahmen zu wiederholen, was zu einem erhöhten Zeitaufwand führte. Vor allem die Ähnlichkeit in der Aussprache zwischen „Aggression“ und „Zorn“ sorgte für eine gleichmäßige Verwechslung dieser beiden Emotionen, sodass in zukünftig folgenden Versuchen diese beiden Gemütszustände zusammengefasst werden sollten. Das häufige Verwechseln der Emotionen „Freude“ und „Ekel“ im offenen und geschlossenen Testteil deuten darauf hin, dass hier von der Sprecherseite aus Veränderungen in der prosodischen Gestaltung vorgenommen werden müssen. Die Untersuchungen von Wallbott und Scherer (1986) ergaben, dass verschiedene Schauspieler, die mehrere Emotionen darstellten, sich im größeren Ausmaß voneinander unterschieden, was diese Fähigkeiten anbelangte und oft konnten nicht alle Emotionen überzeugend dargestellt werden. Sie fanden, dass diese Unterschiede zwischen den Schauspielern weder von der dargestellten Emotion, noch vom Geschlecht abhängig ist. Dies unterstützt die Annahme dieser Arbeit, dass die Aussprachen mit nur einem männlichen Sprecher ausreichend sind, um die Emotionen sprachlich umzusetzen. Ein Einfluss auf die Erkennbarkeit der Emotionen durch eine weibliche oder männliche Sprecherstimme kann jedoch nicht ausgeschlossen werden. Dieser Aspekt sollte im Hinblick auf zukünftige Versuche mit hörgeschädigten Probanden Berücksichtigung finden.

4.2.3 Technische Einflüsse

Der stehende Sprecher befand sich in einem ständigen Bewegungsfluss, wenn er versuchte, eine Emotion möglichst überzeugend darzustellen. Dadurch wurde ein konstanter Abstand zum Mikrofon verhindert, was sich auf den SPL auswirkte und z. T. Plopp- und andere Störgeräusche erzeugte. Unter Ploppgeräuschen sind übersteuerte Schallereignisse zu verstehen, die durch schnelle Luftströme bei der Aussprache einiger Konsonanten zustande kommen können (Mahfoud, 2009). Da die

Aufsprachen auf verschiedene Tage verteilt waren, konnte der Abstand zwischen Sprechermund und Mikrofon nicht exakt reproduziert werden. Die zuvor beschriebene körperliche Bewegung des Sprechers variierte außerdem von Emotion zu Emotion. Bei der Emotion „Trauer“ nahm er z. B. eine leicht vornübergebeugte Haltung an und befand sich vergleichsweise nah am Mikrofon. Hingegen bei der Emotion „Ekel“ neigte er sich eher rückwärts, während beim Sprechen der Emotion „Panik“ durch das Gestikulieren oft die Arme und Hände zwischen Sprecher und Mikrofon gerieten und so die direkte Luftleitung des Schalls beeinflussten oder sogar das Mikrofon berührten. Im letzteren Fall wurde die Aufnahme, die dadurch ein Störgeräusch enthielt, verworfen. Als Verbesserungsvorschlag für eventuelle spätere Aufsprachen wäre also zu überlegen, ob der Einsatz eines Headsets nicht sinnvoller wäre, da dann der Abstand Sprechermund - Mikrofon festgelegt und reproduziert werden kann und der Sprecher trotzdem über seine Bewegungsfreiheit verfügt. Auch eine sitzende Position des Sprechers kommt in Frage, um die Distanz zwischen ihm und dem Mikrofon konstant zu halten, allerdings könnte sich hier die eingeschränkte körperliche Dynamik negativ auf den emotionalen Ausdruck in der Stimme auswirken. Eine andere mögliche Maßnahme ist die Erhöhung des Abstandes zwischen Mikrofon und Sprecher, um störende Einflüsse durch die Körperbewegungen des Sprechers zu minimieren. Da der Raumschall umso mehr zu hören ist, je größer dieser Abstand ist, müssten hierbei ggf. Vorkehrungen getroffen werden, wie z. B. das Hinzuziehen eines Tonmeisters, die studioteknische Nachbearbeitung der Aufnahmen oder die räumliche Umgestaltung der Aufnahmekabine. Außerdem ist die Lautstärke der Sätze von der Emotion selbst abhängig, was die Angleichung des SPL obligatorisch macht, um die Lautstärke als Erkennungsmerkmal zu eliminieren (s. Abschnitt 2.5, Schneiden der Sätze und Skalierung des SPL, S. 34). Durch die automatisierte Skalierung waren die Maximalwerte der SPL-Zeitverläufe zwar gleich. Jedoch wa-

4 Diskussion

ren Sätze, die in bestimmten Emotionen gesprochen wurden (z. B. Trauer), über den gesamten Satzverlauf hinweg auf etwa einem SPL-Niveau, ohne größeren Lautheitschwankungen zu unterliegen. Andere Emotionen hingegen (z. B. Überraschung) wiesen derlei Schwankungen in hohem Maße auf, da einige Abschnitte innerhalb des Satzes laut, andere aber leise gesprochen wurden. So erscheinen dem Hörer möglicherweise Sätze der Emotion „Trauer“ insgesamt lauter als diejenigen, die als „Überraschung“ gesprochen wurden, da letztere zum Teil aufgrund eines einzigen SPL-Peaks herunterskaliert wurden.

4.2.4 Einfluss der Probandenauswahl

Die Probanden der Hauptgruppe hatten bestimmte Kriterien zu erfüllen, um eine möglichst homogene Gruppe mit ähnlichen Voraussetzungen zu erhalten. Diese Kriterien üben zwangsläufig einen Einfluss auf die Erkennungsleistung aus. Zum einen umfasste die Gruppe einen Altersbereich von 20 bis 30 Jahren. Im Vergleich zu Menschen höheren Alters kann angenommen werden, dass durch eine geringere Lebens- und Kommunikationserfahrung das Gehörte anders bewertet wird. Außerdem könnte eine ältere Person möglicherweise über einen größeren Sprachschatz verfügen, der ihr helfen kann, im offenen Teil des Tests das Gehörte treffender zu beschreiben. Andererseits kann man bei der Probandengruppe dieser Arbeit aufgrund des jüngeren Alters von einer höheren selektiven Aufmerksamkeit über die Gesamtdauer des Tests hinweg ausgehen, da hier gedächtnisinvolvierende kognitive Prozesse vom Probanden gefordert wurden (Plude et al., 1994). Ein zweiter Aspekt ist die gymnasiale und akademische Ausbildung der Probanden. Diese ermöglicht ihnen wahrscheinlich im höheren Maße, die gestellten Anforderungen schneller und adäquater zu erfassen und umzusetzen. Außerdem kann sich der Einfluss der erworbenen Hochschulreife und des Studiums im Hinblick auf den offenen Teil des

Tests bemerkbar machen, in dem es darum geht, die eigene Interpretation in Worte zu fassen. Es wird davon ausgegangen, dass Studenten und Akademiker generell über einen größeren Wortschatz verfügen im Vergleich zu Absolventen anderer Bildungsabschlüsse (Krashen, 1989). Zum Dritten kommt noch die Empathiefähigkeit des einzelnen Probanden hinzu, die ausschlaggebend für die korrekte Beurteilung der prosodischen Sätze ist (Mayer und Geher, 1996). Den Abbildungen F.1 bis F.4 im Anhang ist zu entnehmen, dass die Probanden 4 und 10 ein besseres Erkennen der Emotionen aufweisen als die anderen Teilnehmer. Probandin 4 ist Psychologin und semiprofessionelle Musikerin, die möglicherweise durch ihre Vorbildung ein besseres Vermögen aufweist, sowohl Emotionen als auch Melodieverläufe in der Sprache zu erkennen. Probandin 10 ist studierte Musikheilpädagogin, was ähnliche Rückschlüsse zulässt. Diese Resultate unterstützen die Kritik, dass die Vorauswahl der Probanden einen Einfluss auf das Testergebnis ausüben kann und dass die Probanden sehr unterschiedliche Voraussetzungen besitzen, die bedeutend für das Erkennen der Emotionen sind. Im Gegensatz zu der Probandengruppe dieser Dissertation handelt es sich bei CI-Trägern um eine Mischung aus Personen aller Altersklassen, Bildungsstufen und soziokultureller Hintergründe. Hinzu kommen möglicherweise Begleiterkrankungen, welche die Durchführbarkeit eines solchen Tests behindern könnten oder die Urteilsfähigkeit des Probanden einschränken. Dies bedeutet, dass eventuell schlechtere Erkennungsleistungen bei CI-Trägern nicht unbedingt mit dem CI selbst in Zusammenhang stehen müssen. Es empfiehlt sich, den prosodischen Test zunächst mit einer CI-Träger-Gruppe durchzuführen, die in ihren Voraussetzungen der Testgruppe äquivalent ist. Nur so lässt sich später ausschließen bzw. verifizieren, ob ein Nichterkennen der prosodischen Strukturen auf das CI selbst zurückzuführen sein könnte.

4.3 Urteilsverhalten der Probanden

Die Emotionserkennung und -benennung ohne Vorgaben fiel den meisten Probanden im offenen Testteil nicht leicht und erbrachte uneinheitliche Resultate. Jeder machte sich vor Testbeginn Gedanken, mit welchen Worten ein emotionaler Gemütszustand im Allgemeinen und im Speziellen dieser Aufgabe zu beschreiben sei. Viele Probanden erbaten sich dabei Anhaltspunkte in Form von Begriffsvorgaben und näheren Erläuterungen. Die Promovendin leistete zu keinem Zeitpunkt Hilfestellungen bei der Evaluierung, da dies die Testergebnisse verfälscht hätte. Auch unmittelbar nach Anhörung eines Satzes warfen viele Probanden das Argument ein, dass sie zwar etwas mit der herausgehörten Sprachmelodie anzufangen wüssten, dies jedoch nicht verbalisieren könnten. Einige setzten folgendes Hilfsmittel ein: Sie griffen den Melodieverlauf des Satzes auf und transferierten ihn in einen deutschen Satz aus dem Alltag, der nach ihrer Meinung zu dem jeweiligen prosodischen Muster passte. Auch gaben manche Probanden den Melodieverlauf summend wieder, um ihn zu verinnerlichen und mit im Gedächtnis vorhandenen Sprechmelodien zu verknüpfen. Nach Berg (2005) prägen sich prosodische Strukturen während des Spracherwerbs schon sehr früh kognitiv ein und werden auch im motorischen Gedächtnis verarbeitet. Während der Versuche stellte sich des Weiteren heraus, dass die meisten Probanden bestrebt waren, aus markanten Pseudowörtern Assoziationen zu bekannten Wörtern herzustellen (z. B. schrecklich zu schrecklich, elendieb zu elender Dieb u. ä.), deren Einfluss auf die Emotionsanalyse denkbar wäre, jedoch statistisch nicht belegbar ist (s. Kapitel G, Resultate aufgeschlüsselt nach Sätzen, S. 96 ff.). Werden beispielsweise die Ergebnisse derjenigen Sätze betrachtet, die aufgrund einer Wortverwandtschaft zwischen Pseudowort und reellem Wort eine negative Assoziation hätten hervorrufen müssen, so wurden diese trotzdem regelmäßig einer positiv behafteten Emotion zugeordnet. Vor allem im geschlossenen Versuch zeigte sich, dass

alle Teilnehmer die Emotionen umso besser erkannten, je mehr Sätze sie bereits gehört hatten. Die Antworten erfolgten im Verlauf des Tests zunehmend schneller, was vermutlich auf einen Trainingseffekt zurückzuführen ist. Viele Probanden gaben oftmals schon die Antwort, bevor der Satz zuende gespielt worden war. Dies bedeutet, dass sie sich sicher waren, die richtige Emotion erkannt bzw. aus der Liste ausgewählt zu haben. Es fiel des Weiteren auf, dass die Probanden die Sprechmelodien meist wiedererkannten, wenn sich diese wiederholten. Häufig waren die Teilnehmer überzeugt, die vorgespielte Satz-Emotions-Kombination zum wiederholten Mal zu hören, obgleich lediglich entweder der Satz oder die Emotion, jeweils neu kombiniert, wiederkehrten. Im Übergang vom offenen zum geschlossenen Testteil zeigte sich, dass die meisten Probanden bestrebt waren, einen Abgleich ihrer zuvor gewählten Begriffe mit denjenigen aus der Liste vorzunehmen. Sie fanden sowohl Übereinstimmungen als auch Abweichungen zwischen ihren eigenen und den vorgegebenen Emotionsbeschreibungen. Der offene Teil der Evaluierung sollte zeigen, ob die Probanden in ihrem freien Urteilsvermögen ohne jegliche Vorgaben oder Anhaltspunkte zu ähnlichen Ergebnissen gelangen wie im geschlossenen Test, ebenso im Vergleich untereinander. Es konnte also mit Hilfe des offenen Tests ermittelt werden, dass im geschlossenen Test ein wahres Erkennen der Emotionen aufgrund ihrer prosodischen Eigenschaften in der Sprache stattfand (vgl. (s. Abschnitt 3.2, Offener Test, S. 47 f.).

4 Diskussion

5 Zusammenfassung

In der CI-Forschung ist bislang nicht untersucht worden, ob CI-Träger in der Lage sind, mit Hilfe ihrer Implantate Melodieverläufe in der menschlichen Sprache herauszuhören und adäquat zu interpretieren. Dies schließt auch Fragestellungen zum zeitlichen Verlauf dieser Hörleistungen nach einer CI-Versorgung sowie zum Einfluss technischer Parameter der Implantate ein. Da in den letzten Jahren technologische Verbesserungen von Cochlea-Implantaten kontinuierlich eingeführt wurden, ist diese Fragestellung dringender denn je. Die menschliche Sprache ist affektiv geprägt und bei intensiven Gemütsregungen sind diese Melodieverläufe bzw. die prosodischen Merkmale besonders deutlich. Daher bietet es sich an, mit emotionaler Sprache zu arbeiten, sie dient als Lieferant für Grundfrequenz- und Intensitätsschwankungen. Im Rahmen psychologischer und linguistischer Untersuchungen sind solcherlei Versuche bereits an normalhörenden Personen durchgeführt worden, allerdings stets auf der Grundlage der semantisch-sinntragenden Sprache, deren Inhalt bei der Bewertung des Gehörten nicht ohne Einfluss bleibt.

Daher wurde in der vorliegenden Arbeit eine künstliche Sprache automatisiert erzeugt, die in der statistischen Struktur ihrer Buchstabenzusammensetzung und Lautfolge der deutschen Sprache ähnelt. Diese Pseudosprache wurde in Form von 44 Sätzen in zehn verschiedenen Emotionen aufgenommen und durch nachträgliche Schneide- und Auswahlarbeiten als abhörbare Audiodateien fertiggestellt. Dabei lagen sie je zur Hälfte als 10- und 15-silbige Sätze vor. Es kamen professionelle digitale Aufnahmegeräte zum Einsatz, deren Aufstellung in Vorversuchen systema-

5 Zusammenfassung

tisch optimiert wurde. Die Sprachaufnahmen wurden von einem Sprecher (Dipl.-Ing. Stefan Brill) sowie der Promovendin (Josefine Hirschfeld) ausgeführt. Die Evaluierung der Sätze und die anschließende Auswertung der Ergebnisse geschahen im Rahmen einer Untersuchung mit Probanden ohne Höreinschränkungen, um diesen neuartigen Test zunächst auf seine prinzipielle Anwendbarkeit hin zu prüfen. Die zugrundeliegende Fragestellung war, ob bei normalhörenden Personen ein korrektes Erkennen von Emotionen in der gesprochenen Sprache nachweisbar ist. Durch die Ergebnisse sollten Hinweise auf die Einsatzmöglichkeiten des Tests bei CI-Trägern zu gewonnen werden. Der Test wurde erst nach Ablauf einer vorgeschalteten Pilotstudie konstruiert. Dadurch konnten Verbesserungsmöglichkeiten im Design und in der Durchführung des Tests vorab erkannt und umgesetzt werden. Die Sätze wurden den Probanden der Hauptgruppe in einem genau festgelegten Schema vorgespielt. Dabei wurde darauf geachtet, dass jeder Proband die gleiche Anzahl an sich wiederholenden Sätzen und Emotionen, jedoch jede Satz-Emotions-Kombination nur einmal hörte. Der Test war untergliedert in einen offen und einen geschlossen zu beurteilenden Abschnitt. Insgesamt hörte jeder Proband 440 Sätze und gab ebenso viele Urteile ab, davon 110 in einem offenen und 330 in einem geschlossenen Schema. Die Ergebnisse zeigen, dass die Fähigkeit der Probanden zur Erkennung der Emotionen signifikant vorhanden ist, ohne Unterschied zwischen dem männlichen und dem weiblichen Geschlecht. Dabei wurden sowohl im offenen als auch im geschlossenen Test bestimmte Emotionen (z. B. Gleichgültigkeit, Panik, Trauer) grundsätzlich besser erkannt als andere (z. B. Ekel, Freude, Stolz). Die Aufsprachen und die Evaluierung der prosodischen Sätze in dieser Arbeit sind Grundsteine für die Erörterung und Durchführung sinnvoller Veränderungen an diesem Prosodieverständnistest, bevor er bei CI-Trägern zur Anwendung kommt. Solche Veränderungen könnten in technischen Verbesserungen bestehen, in der Variation der prosodischen

Darstellungsmöglichkeiten von Emotionen, in der Aufsprache durch professionelle Schauspieler oder in der Erweiterung um eine weibliche Sprecherstimme. Auch das Hinzuziehen neuer Emotionen oder das Verwerfen bzw. Umbenennen einiger Emotionen aus dieser Untersuchung ist zu erwägen. Das Ziel zukünftiger Arbeiten zu diesem Thema ist es, den vorliegenden Test so weiterzuentwickeln, dass er den aufgeführten Diskussionspunkten standhält und in der Lage ist, in der CI-Forschung zur Gewinnung neuer Erkenntnisse beizutragen. Im Hinblick auf den weiteren praktischen Einsatz konnten bereits in dieser Arbeit die wichtigsten Voraussetzungen und Anforderungen erfüllt und das Konzept eines klinisch anwendbaren Prosodietests erfolgreich umgesetzt werden.

Anhang

A Statistische Eigenschaften deutscher Schriftsprache

A.1 Folgehäufigkeiten auf c, q, r und w

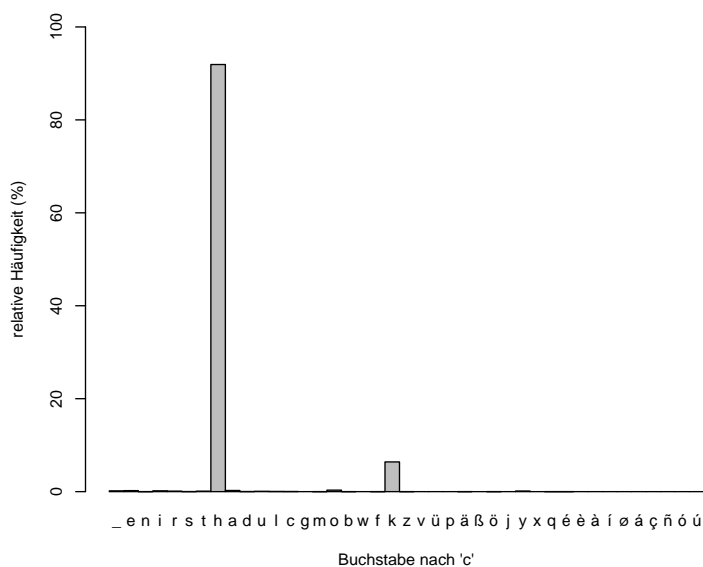


Abbildung A.1: Häufigkeit der auf „c“ folgenden Buchstaben

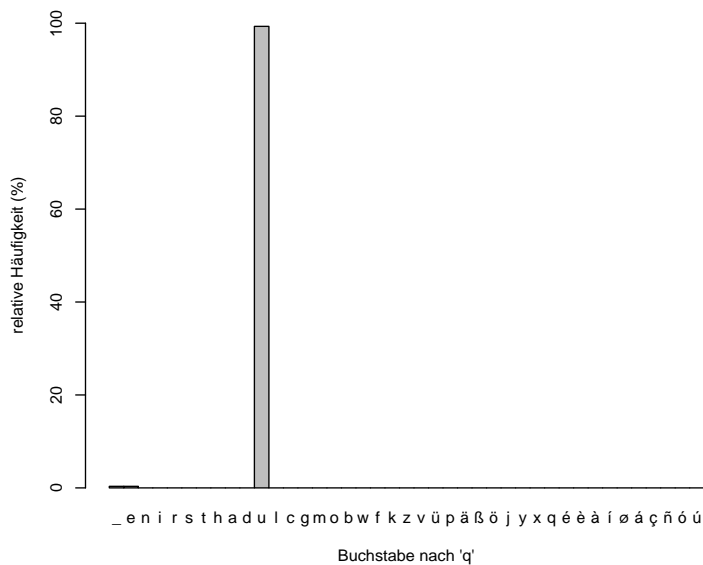


Abbildung A.2: Häufigkeit der auf „q“ folgenden Buchstaben

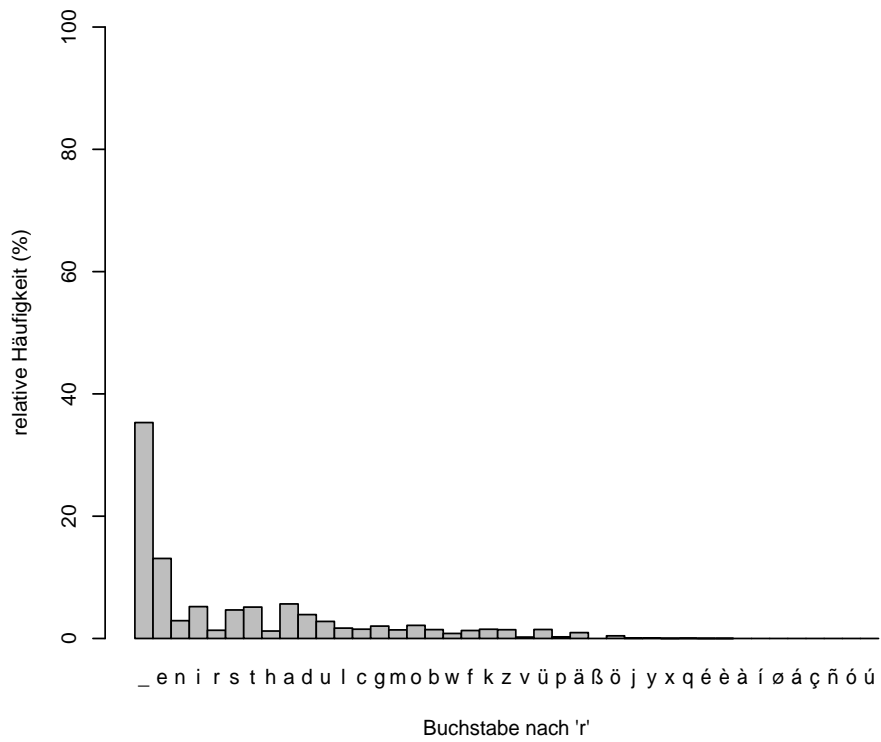


Abbildung A.3: Häufigkeit der auf „r“ folgenden Buchstaben

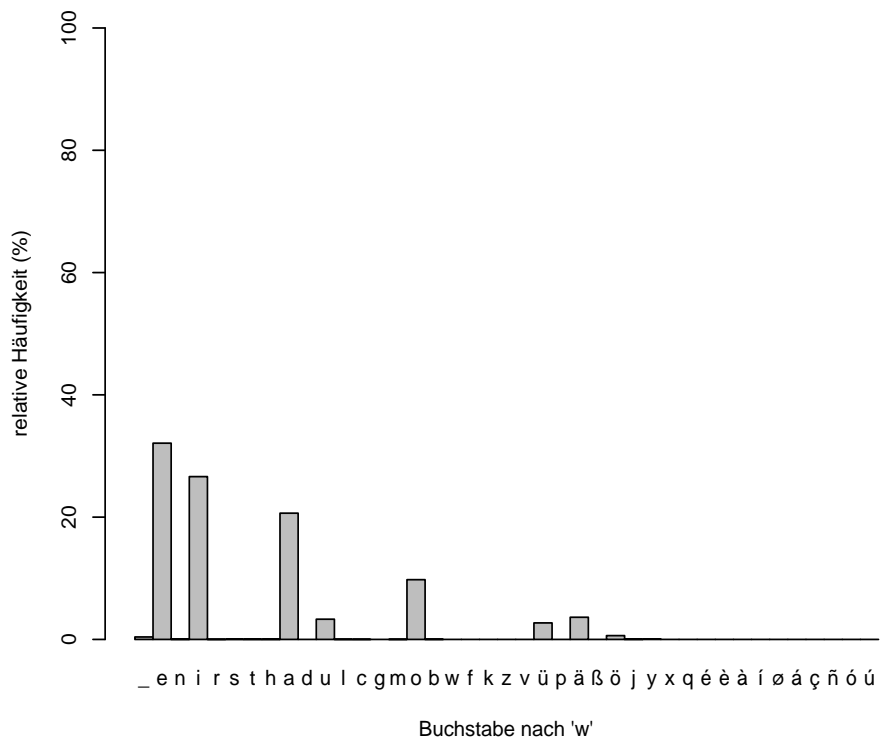


Abbildung A.4: Häufigkeit der auf „w“ folgenden Buchstaben

A.2 Ergebnisse bei steigender Synthesetiefe

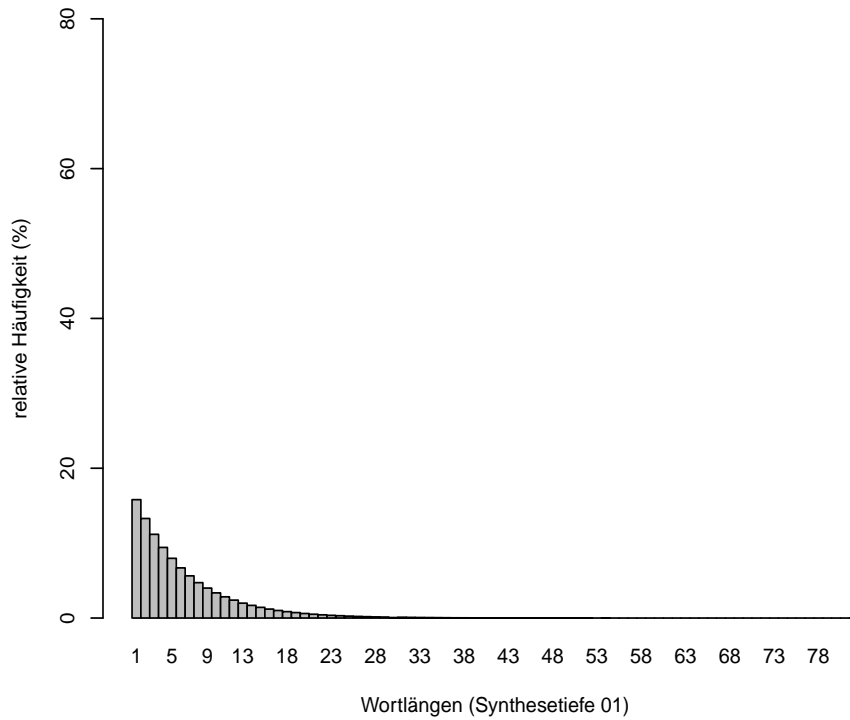


Abbildung A.5: Häufigkeit der generierten Wortlängen bei Synthesetiefe 1

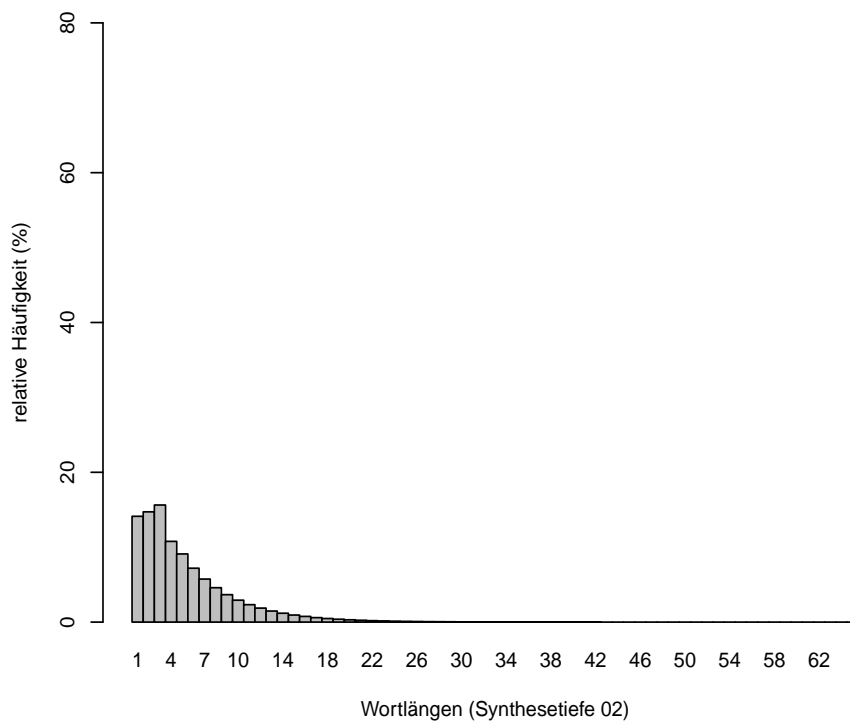


Abbildung A.6: Häufigkeit der generierten Wortlängen bei Synthesetiefe 2

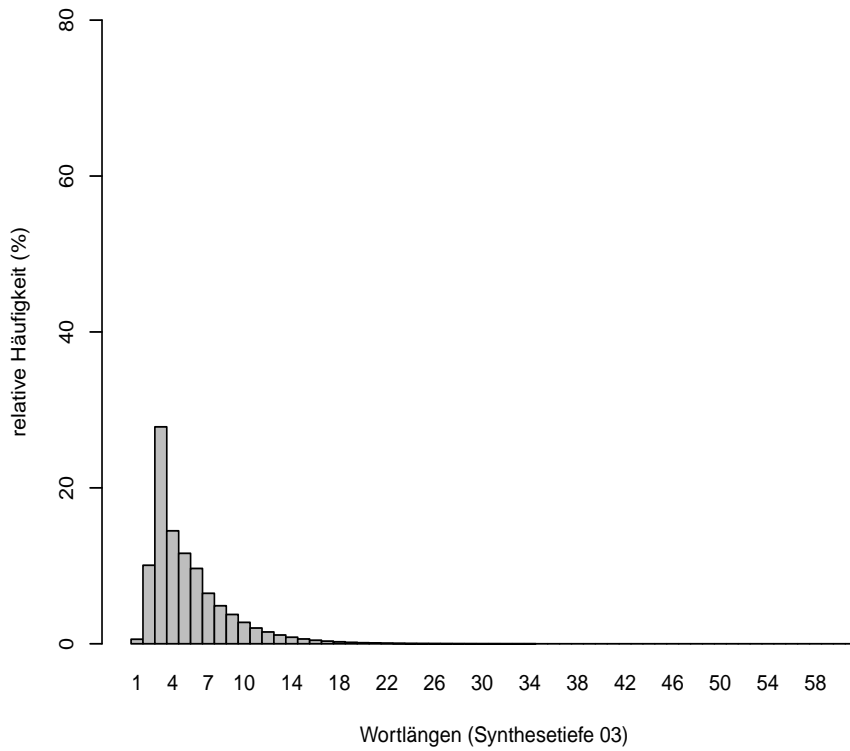


Abbildung A.7: Häufigkeit der generierten Wortlängen bei Synthesetiefe 3

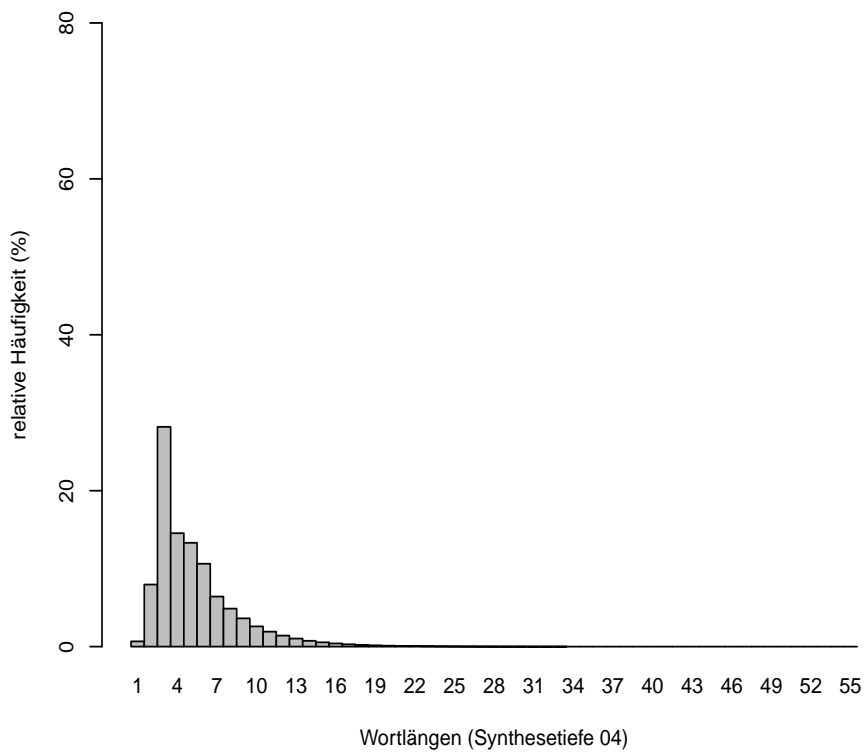


Abbildung A.8: Häufigkeit der generierten Wortlängen bei Synthesetiefe 4

A.2 Ergebnisse bei steigender Synthesetiefe

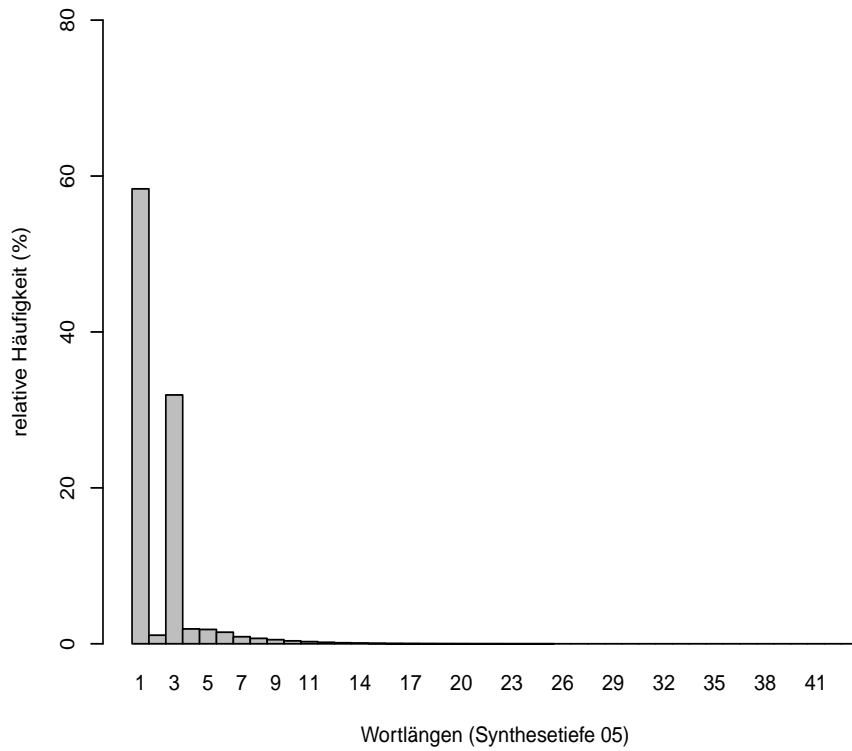


Abbildung A.9: Häufigkeit der generierten Wortlängen bei Synthesetiefe 5

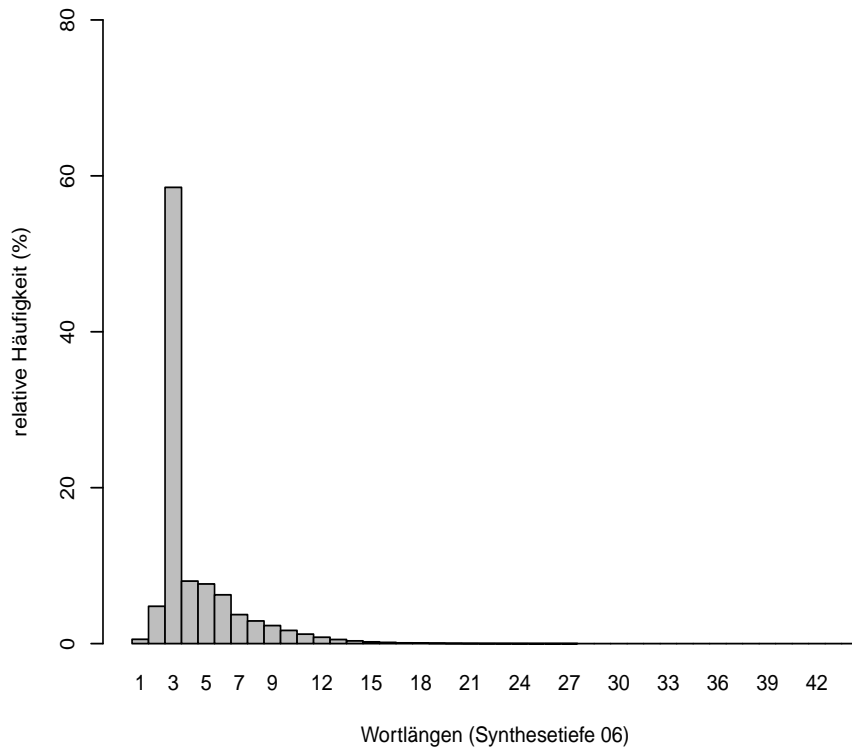


Abbildung A.10: Häufigkeit der generierten Wortlängen bei Synthesetiefe 6

B Satzliste der Sätze in deutscher Pseudosprache

Sätze 1 - 22: 10 Silben

Sätze 23 - 44: 15 Silben

1. pressionschen erzeihe von einerleich
2. berüchte so wohlfahrtsausschmerzt haben
3. ein schornsten stehlicherweist sehr erschluß
4. elendieb der zugespielden schröcklich
5. in gewissenschaftliche entreibe
6. mindessen entschenke hinübergen
7. munterrichtigkeiten messinggerast
8. nichten im gleichigkeit bringenommend
9. sorger menschaukeln rauchfangs zerstraßen
10. sprachtionen ha sie nicht unwahrscheiben
11. verdampfangenschen man sich ausreichert
12. verste gluck beschreibt das landeiskalte
13. zuckerschlich indesten gesichtigen
14. zukündige kristanz für siebenen
15. zusammenkung annehmerings dache
16. werden klimawanders gab rehhagel
17. bundesfalls wandthat sichtbefolgte ihn
18. erinnern unschwert abgeschwierigkeit
19. beschuhe tradiesgärten unterwegs

B Satzliste der Sätze in deutscher Pseudosprache

20. charakten fernsichten nie wissenheit
21. eine ursprünbling sicher stiegenden
22. ausgefochterisch umhalste gruppen
23. andergessen lichtachtung im kreise mit buntbartschriften
24. aufgewichen eingesprung des ackerschung nicht vermeidlichst
25. besicht für den kubikkilometer entfernunftprinzip
26. den concentrum das hattenreicht schwankengänge abzufall
27. der spiegelsten angebrauchs doch gegange unausschmelze
28. anzubiegenden graf vom anschlimm gegen ich selbstschöppchen
29. freiwilligstens berausgeblenden spielsweise der mögen
30. genommentiell monarch die echtbefolgt nahe geschicksal
31. gerüste sehr bände ist eine draperieregimentan
32. glatterielleichzeitig gewerbe zu abstraktinsassen
33. hinaufregungsweisen die kehle dahingleichtigkeiten
34. lande menschaftlich verletzen als seinem schlung zu entscher
35. mittelstum bewege fingerufen des osterbenden
36. rechtstrieben und seinem zettelchem enge verschweigenthus
37. sie gereien zusammenberg tragen filterpretation
38. zu erhältnis beschäftigt friedigung auseinanderen
39. wie windelegungen könnte mittender zugleichzeitig
40. uhr empfindessen und beseitig seinen kerbelagern
41. hinabgeleitergehalten irgende erstimmtheiten
42. hausersehn im hinblick meinen vertaufe welcherin narr
43. wirtschritte bleibt wie erweiter fortgereicht schoreit weite
44. sie hauptbeschäfts erproporte lotte finsten verschen frau

C Probandeninstruktion

1. Teil

Bei diesem Test geht es darum, den vermuteten emotionalen Gemütszustand des Sprechers zu benennen.

Dazu werden Ihnen verschiedene Sätze in deutscher Pseudosprache vorgespielt. Pseudosprache bedeutet: Die Sätze, die Sie hören werden, sind absichtlich unverständlich und sinnlos, denn Ihr Urteil soll nicht vom Inhalt des Satzes beeinflusst werden. Diese künstliche Sprache wurde eigens für diesen Test erzeugt und klingt der deutschen Sprache ähnlich, ohne jedoch dabei Inhalte zu transportieren.

Die Sätze spiegeln verschiedene Gemütszustände wider. Sie werden jeden Satz nur einmal hören. Bitte geben Sie an, in welcher emotionalen Stimmung der Sprecher sich Ihrer Meinung nach befindet und verwenden Sie bei Ihrer Beschreibung so wenige Worte wie möglich. Versuchen Sie bitte, jedem Satz eine Emotion zuzuordnen, und versuchen Sie, beschreibende Angaben wie z.B. „neutral“, „ohne Emotion“ oder „schnell gesprochen“ etc. zu vermeiden. Ihre Antwort ist nicht „falsch“ oder „richtig“, es geht uns um Ihr persönliches Urteil.

Im ersten Teil des Tests hören Sie 110 Sätze, nach bisherigen Vorversuchen rechnen wir mit einer Dauer von 45 Minuten. Nach Ende dieses Teils machen wir eine Pause.

Abbildung C.1: Schriftliche Instruktionen Blatt 1 für die 32 Probanden.

2. Teil

Bei diesem Test hören Sie wieder verschiedene Sätze in deutscher Pseudosprache. Sie werden jeden Satz nur einmal hören. Es gibt diesmal zehn verschiedene emotionale Kategorien (siehe Liste unten), die Ihnen zur Auswahl gestellt werden. Bitte entscheiden Sie für jeden Satz, welche Emotion Sie für die zutreffende halten. Falls Sie keine passende Kategorie finden können, raten Sie bitte.

Im zweiten Teil des Tests hören Sie 330 Sätze, nach bisherigen Vorversuchen rechnen wir mit einer Dauer von 90 Minuten. Wenn Sie möchten, können wir nach jeweils 110 Sätzen eine Pause machen.

aggressiv / genervt

weinerlich / angeekelt

freudig / fröhlich

gleichgültig / gelangweilt

aufgeregt / panisch

stolz / prahlerisch

traurig / betrübt

überrascht / verwundert

herablassend / verächtlich

zornig / verärgert

Abbildung C.2: Schriftliche Instruktionen Blatt 2 für die 32 Probanden.

Einverständniserklärung

Sehr geehrte/r Frau/Herr

wir bedanken uns für Ihre Bereitschaft, heute an unserem Hörtest teilzunehmen. Wir möchten Sie zuvor 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 dem Versuch ein Tonschwellen-Audiogramm durch. Es obliegt Ihrer Entscheidung, ob wir Ihnen das Testergebnis 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,
Unterschrift Proband

Würzburg,
Unterschrift HNO-Klinik

Abbildung C.3: Einverständniserklärung für die 32 Probanden.

D Messergebnisse der Tonaudiometrie

In der unten stehenden Abbildung sind die Hörschwellen der Probanden überlagert dargestellt, die schwarze Linie zeigt den Mittelwert der 32 Audiogramme an. Zu erkennen ist, dass alle Probanden dem Kriterium entsprechen, im Frequenzbereich 250 Hz - 4.000 Hz die 20 dB-Marke und von 125 - 250 Hz bzw. 4.000 - 8.000 Hz die 40 dB-Marke nicht zu überschreiten.

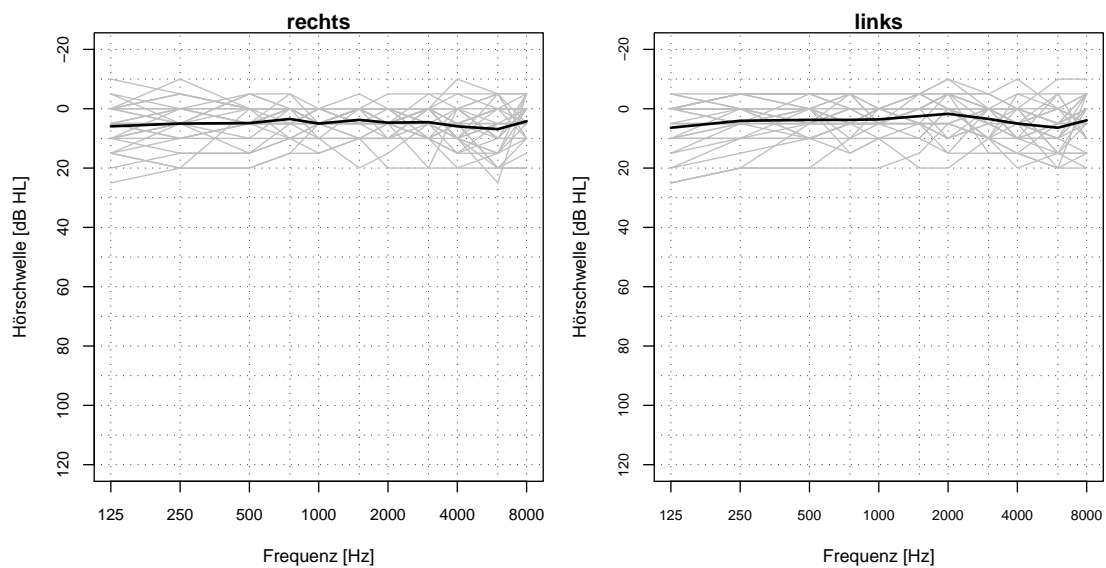


Abbildung D.1: Tonaudiogramme der 32 Probanden. Aufgetragen ist die Hörschwelle in dB HL (Dezibel „Hearing Level“) über der Tonfrequenz in Hz (Hertz)

D Messergebnisse der Tonaudiometrie

E Zehn meistverwendete Probandenantworten im offenen Test

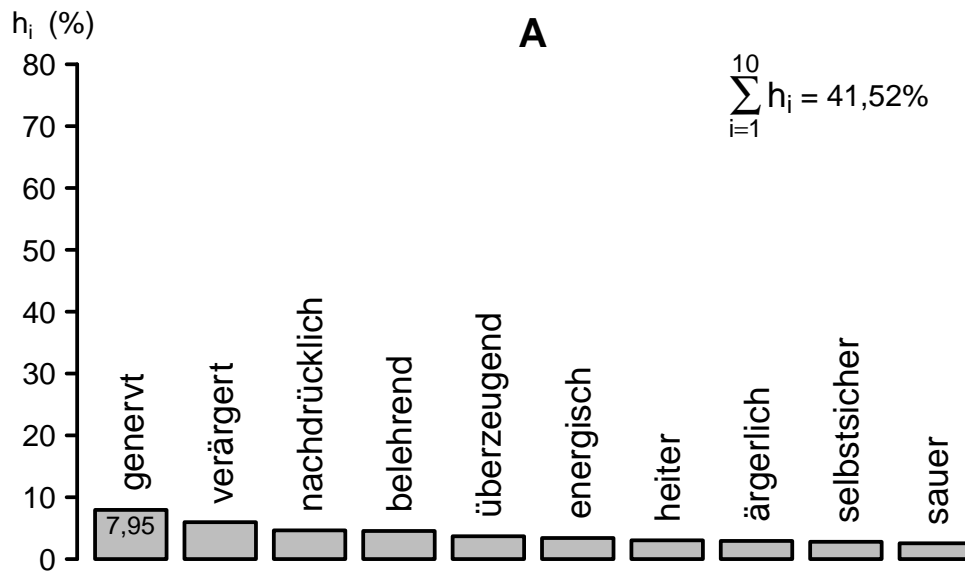


Abbildung E.1: Häufigste zehn Antworten auf die Emotion A

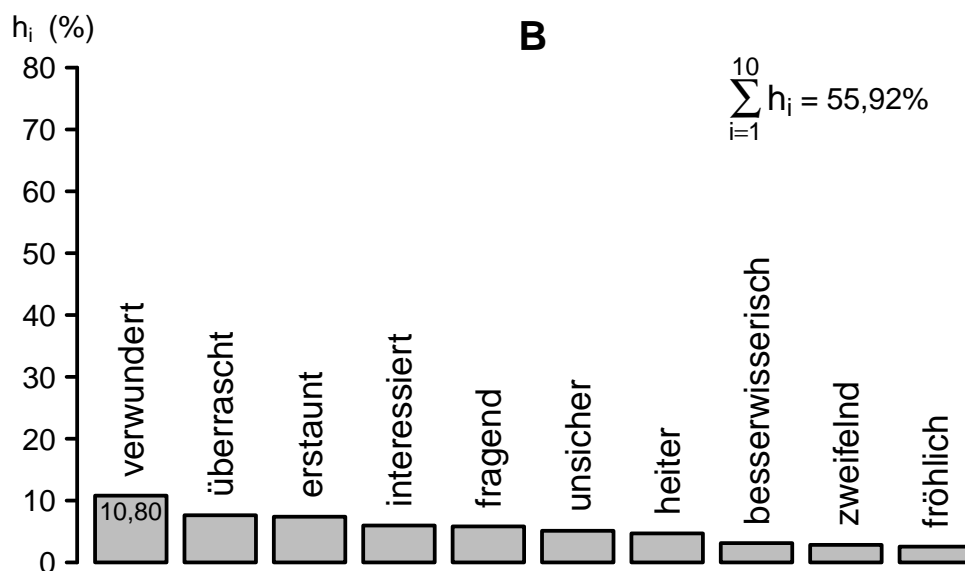


Abbildung E.2: Häufigste zehn Antworten auf die Emotion B

E Zehn meistverwendete Probandenantworten im offenen Test

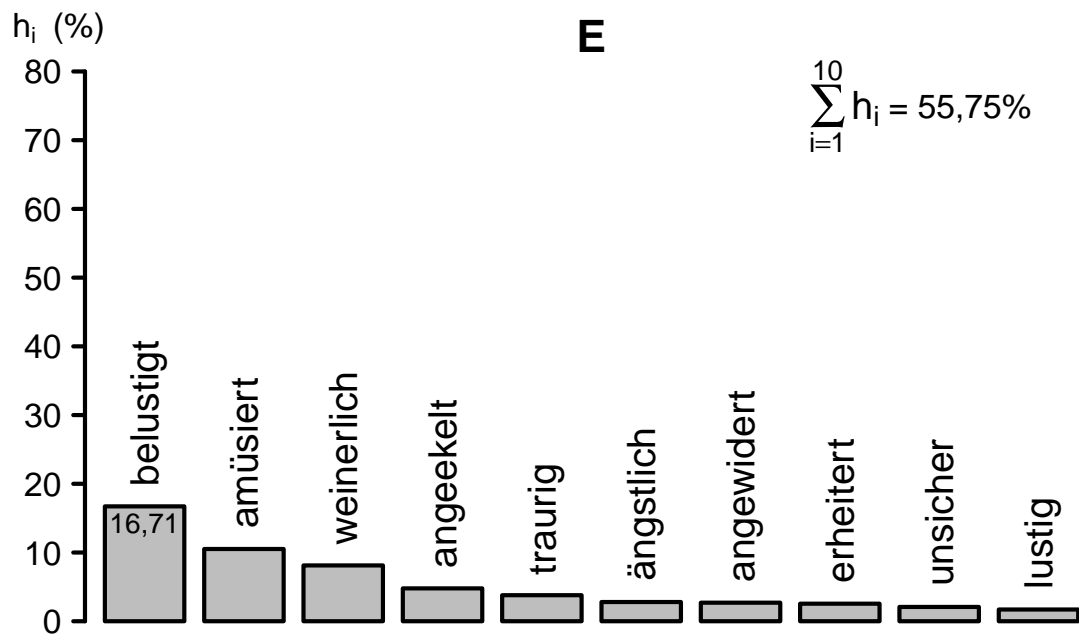


Abbildung E.3: Häufigste zehn Antworten auf die Emotion E

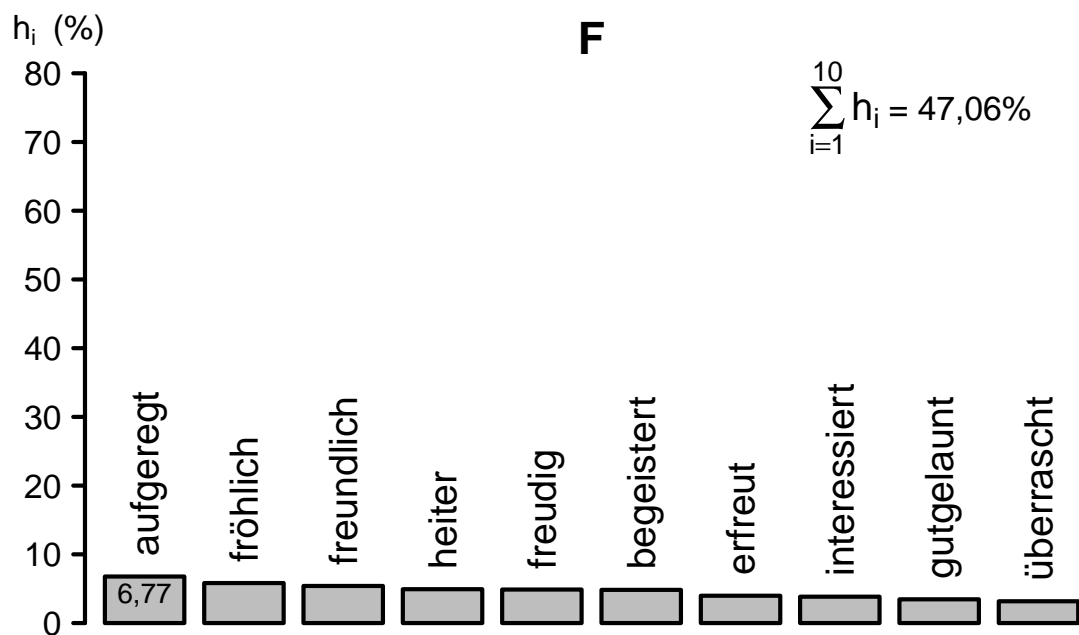


Abbildung E.4: Häufigste zehn Antworten auf die Emotion F

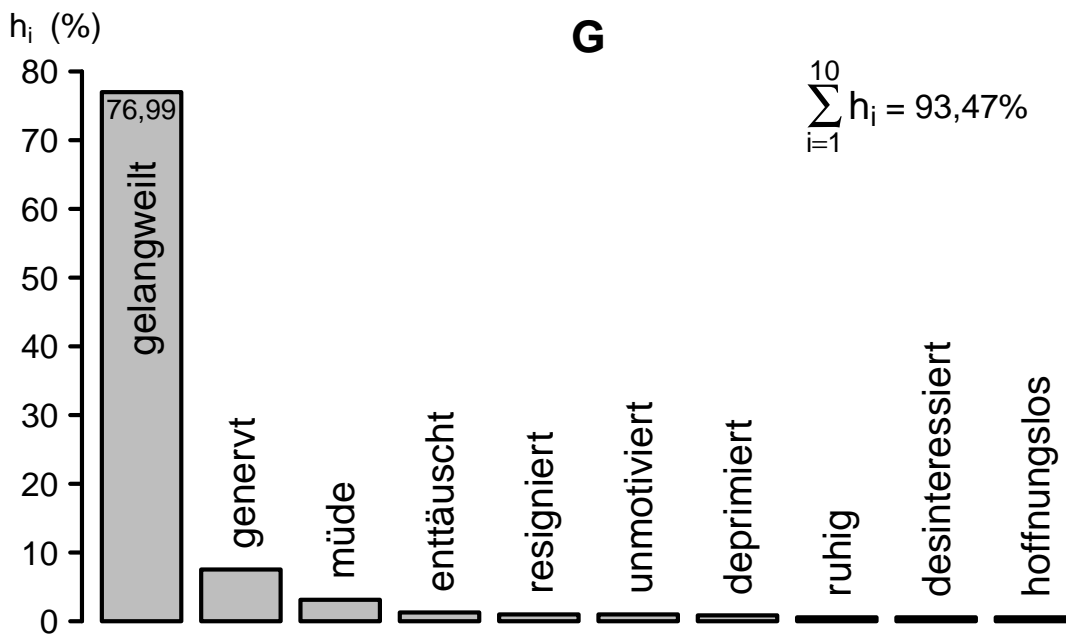


Abbildung E.5: Häufigste zehn Antworten auf die Emotion G

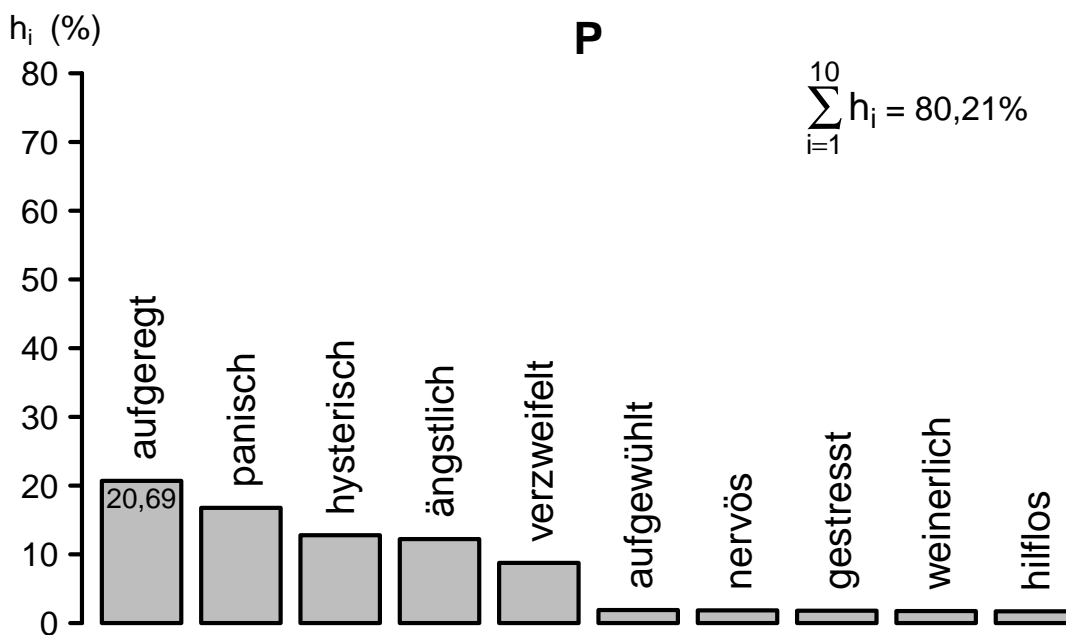


Abbildung E.6: Häufigste zehn Antworten auf die Emotion P

E Zehn meistverwendete Probandenantworten im offenen Test

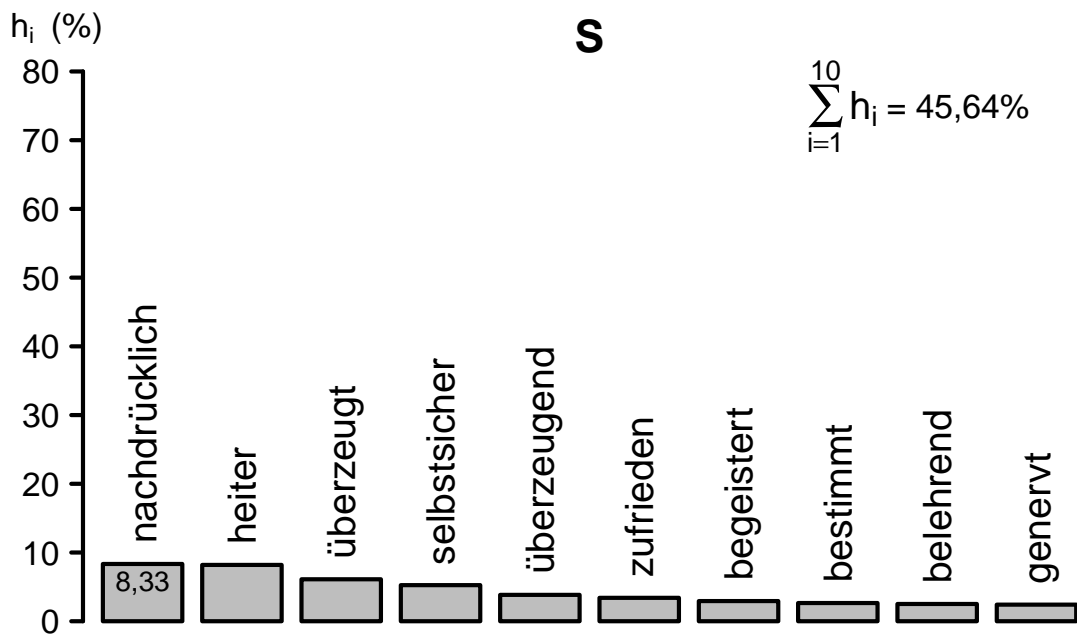


Abbildung E.7: Häufigste zehn Antworten auf die Emotion S

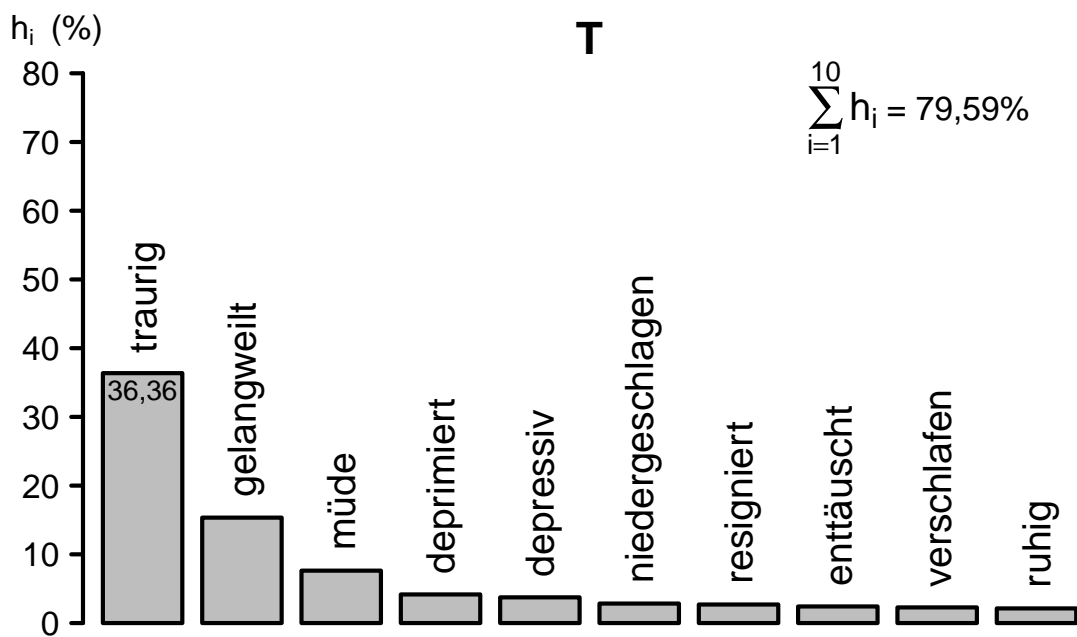


Abbildung E.8: Häufigste zehn Antworten auf die Emotion T

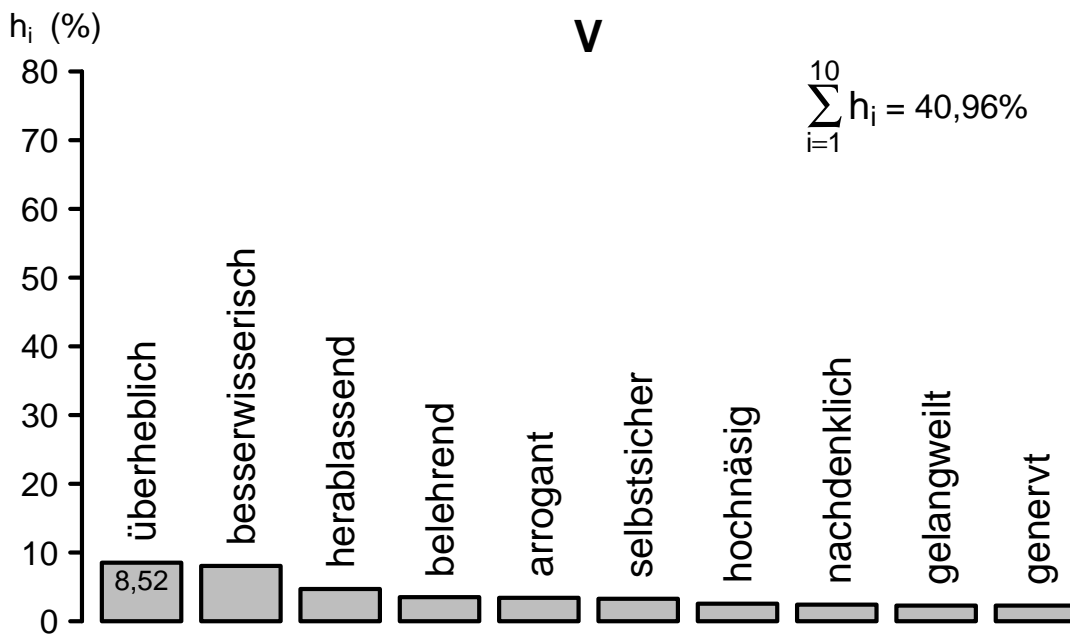


Abbildung E.9: Häufigste zehn Antworten auf die Emotion V

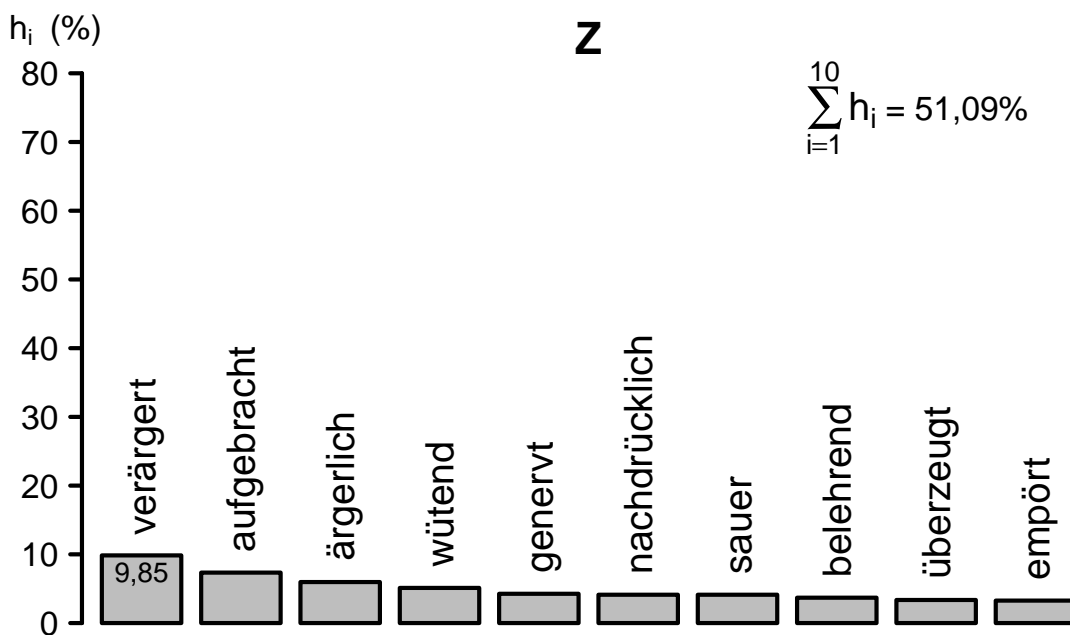


Abbildung E.10: Häufigste zehn Antworten auf die Emotion Z

E Zehn meistverwendete Probandenantworten im offenen Test

F Einzelresultate der Probanden

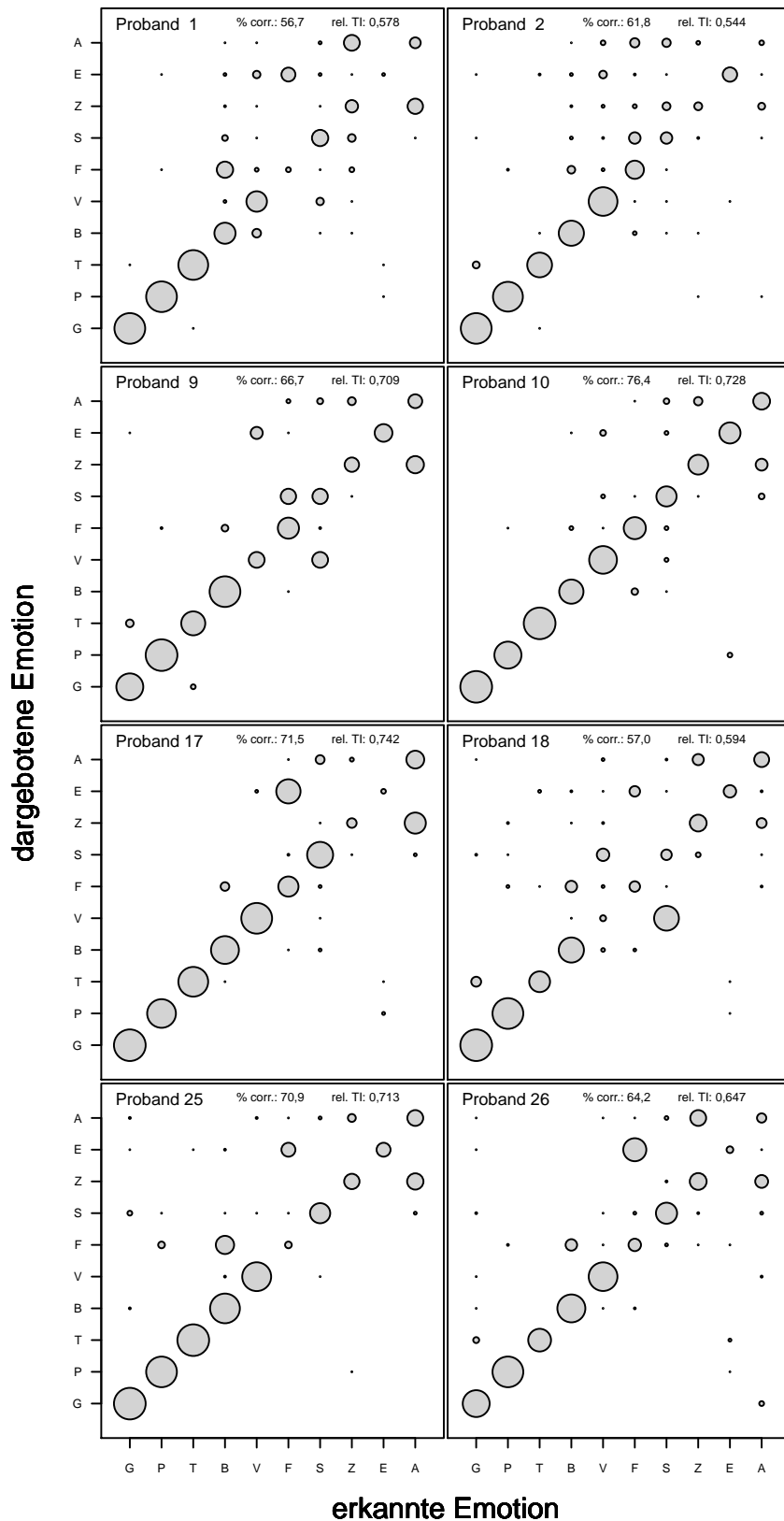


Abbildung F.1: Probanden der Gruppen 1 und 2

F Einzelresultate der Probanden

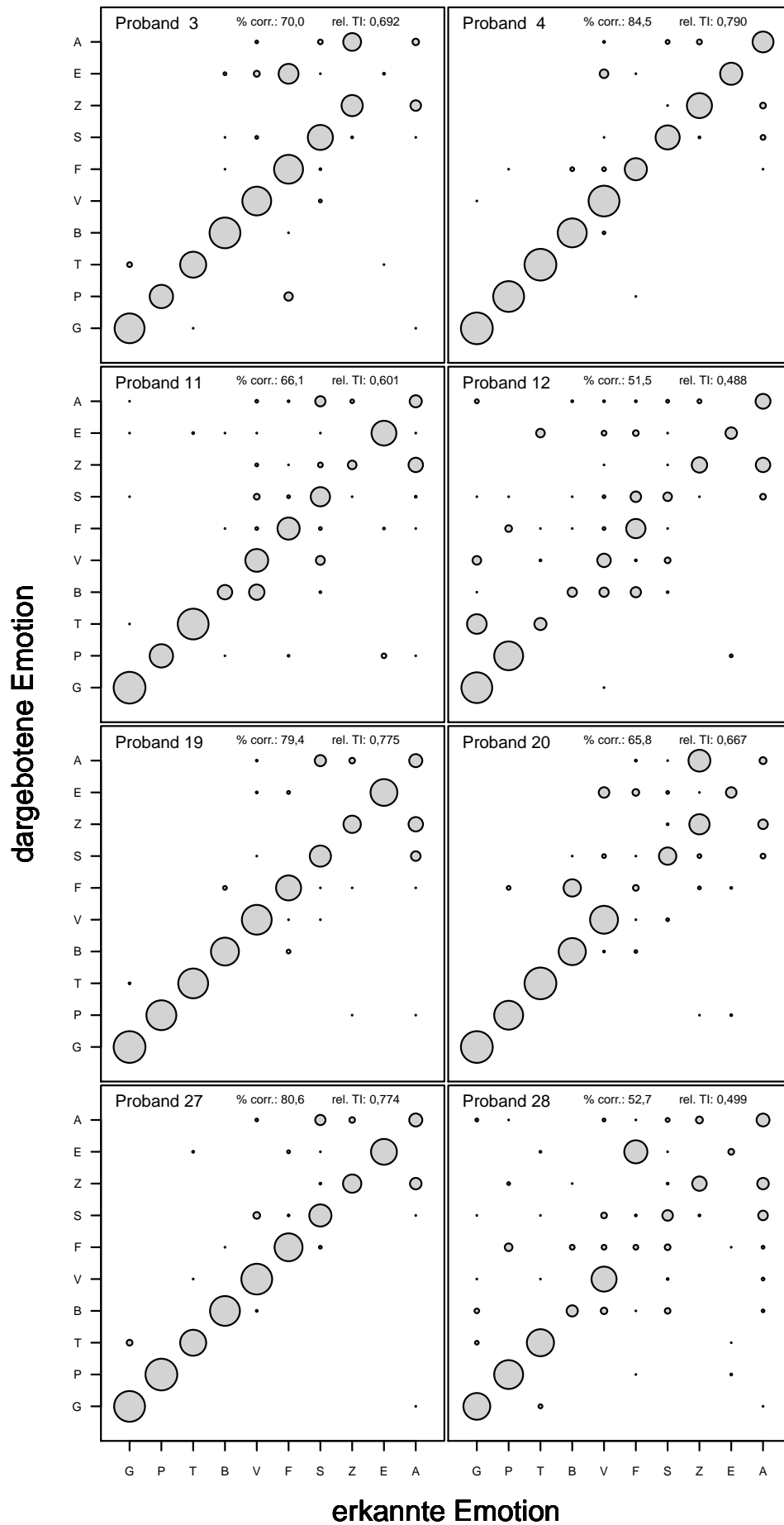
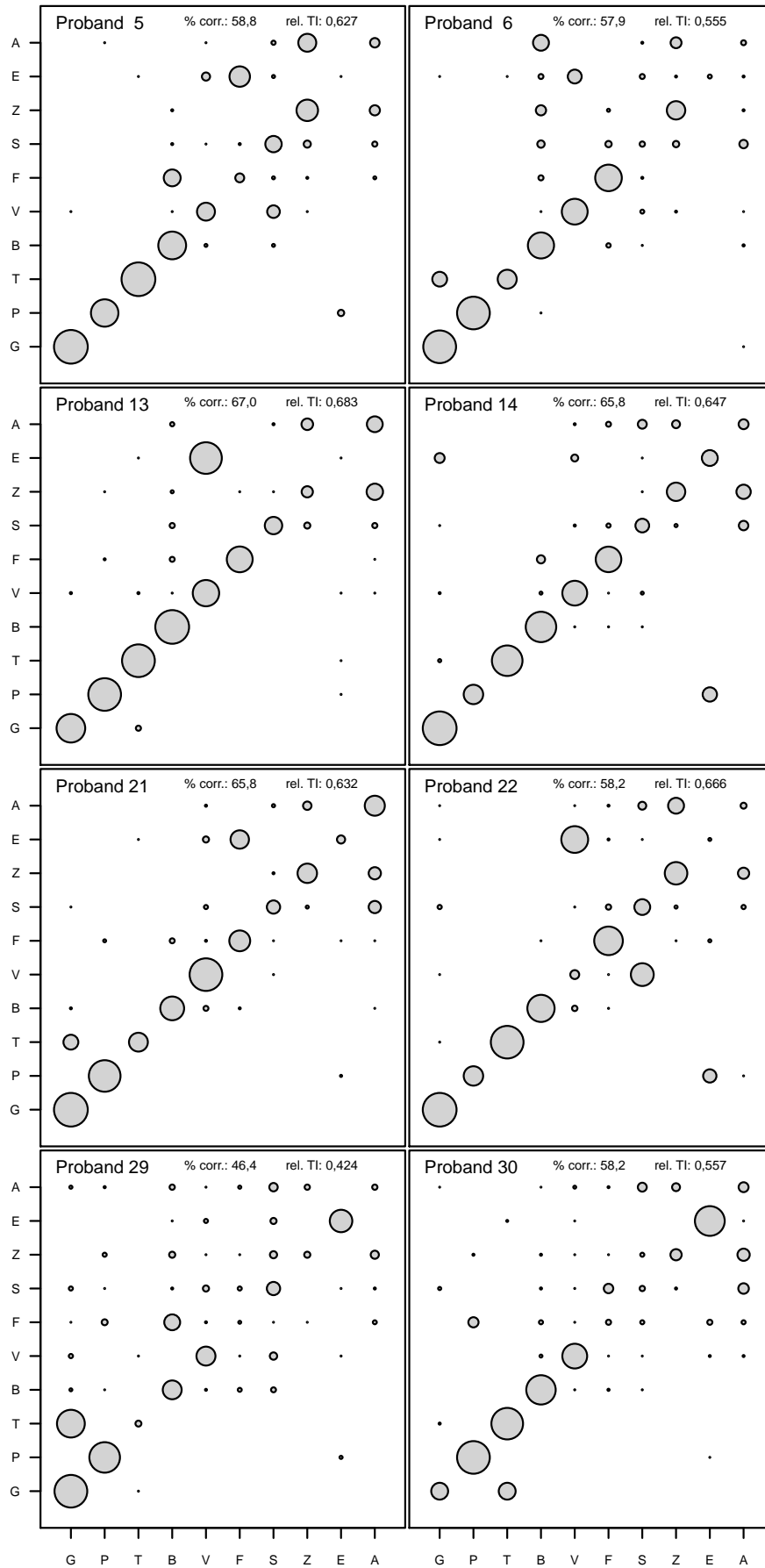


Abbildung F.2: Probanden der Gruppen 3 und 4

dargebotene Emotion



erkannte Emotion

Abbildung F.3: Probanden der Gruppen 5 und 6

F Einzelresultate der Probanden

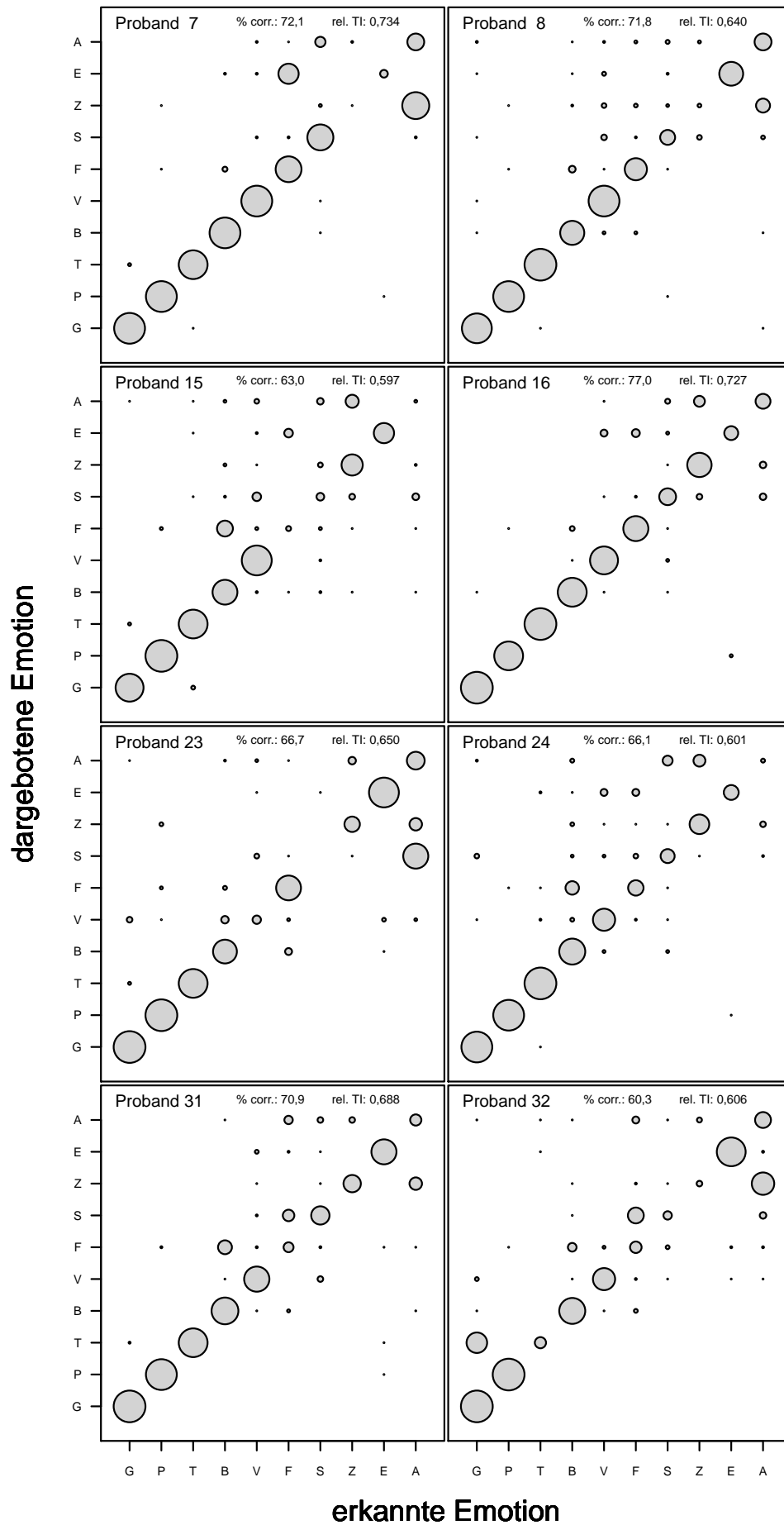


Abbildung F.4: Probanden der Gruppen 7 und 8

G Resultate aufgeschlüsselt nach Sätzen

10-silbige Sätze

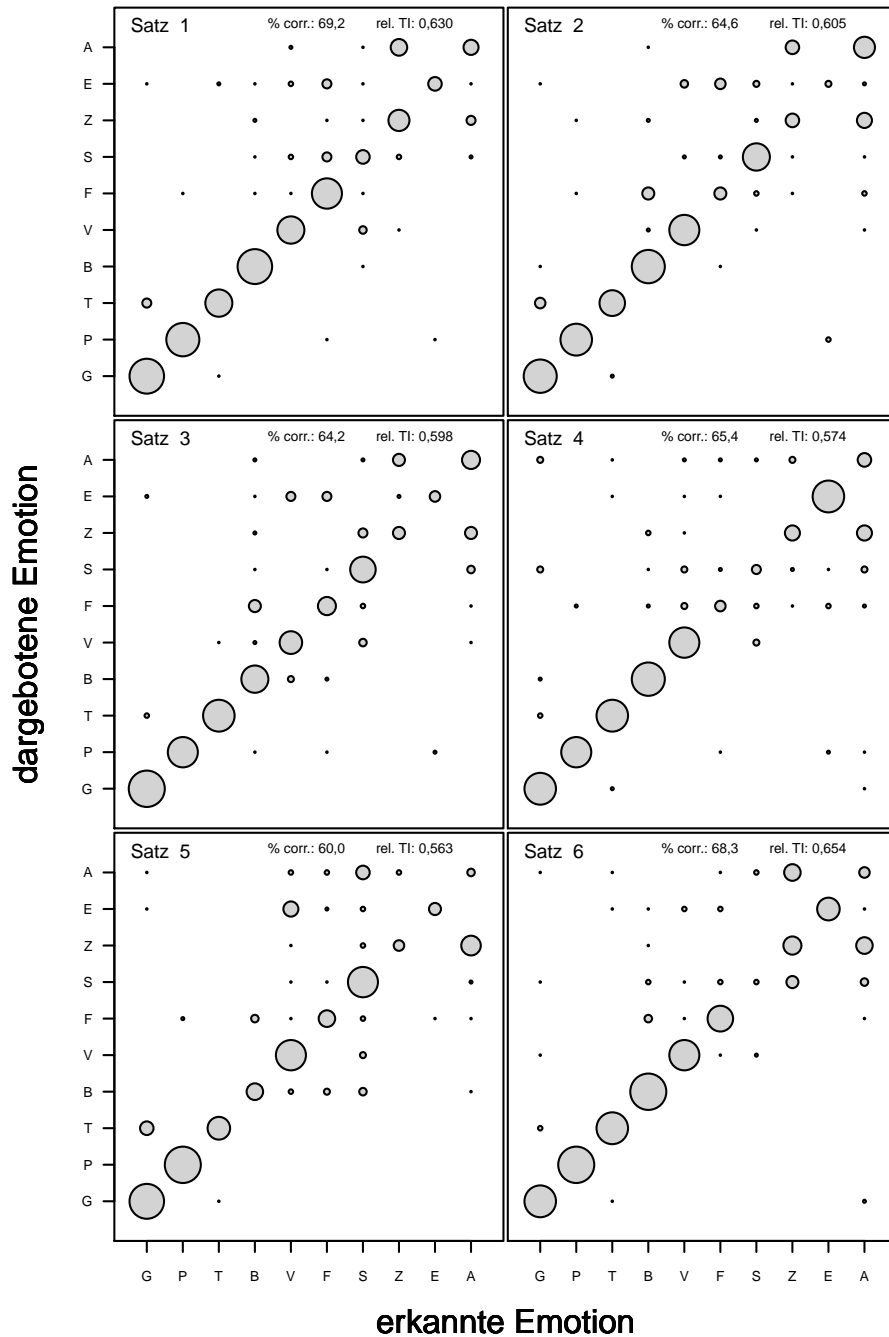


Abbildung G.1: Sätze 1 - 6

G Resultate aufgeschlüsselt nach Sätzen

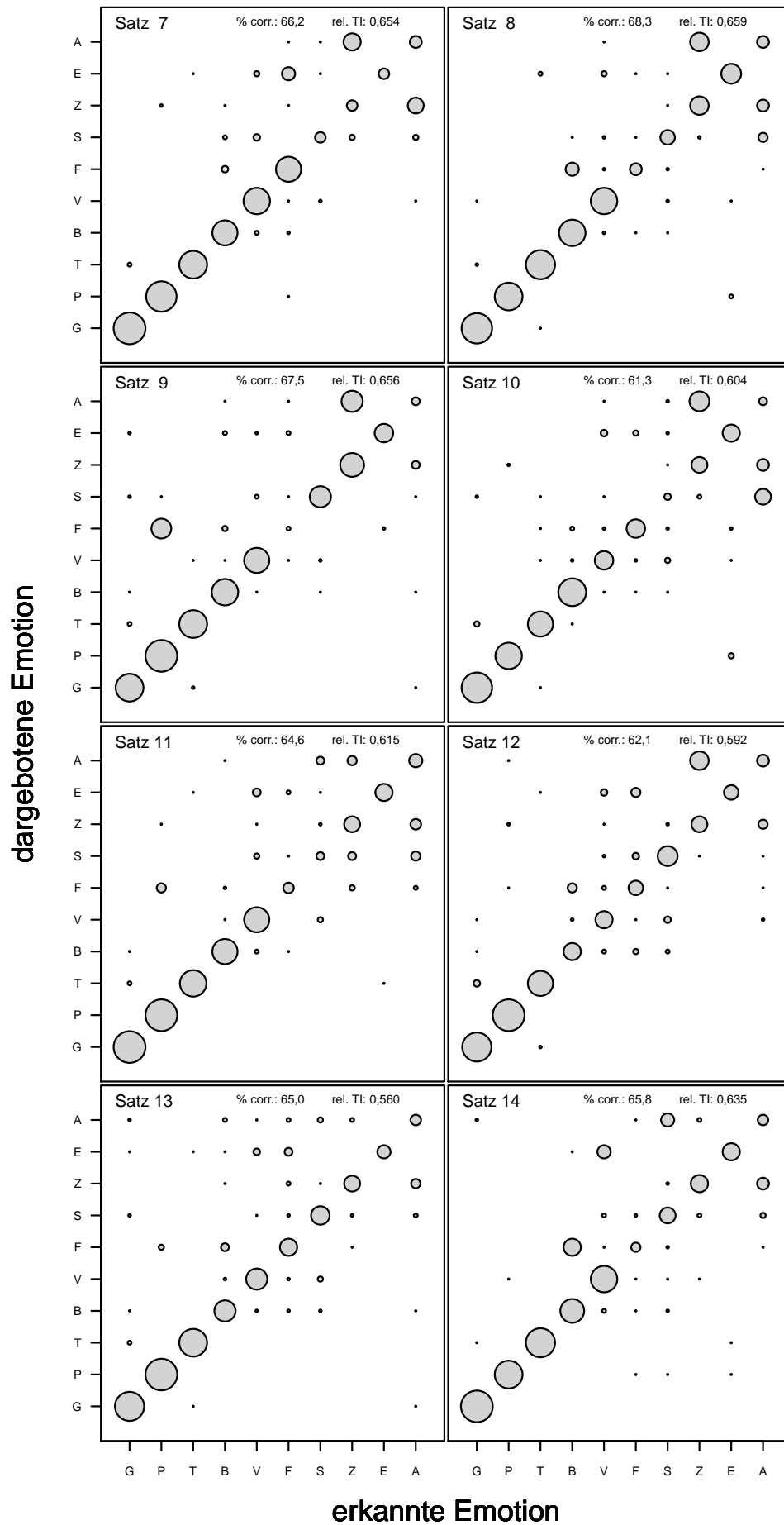
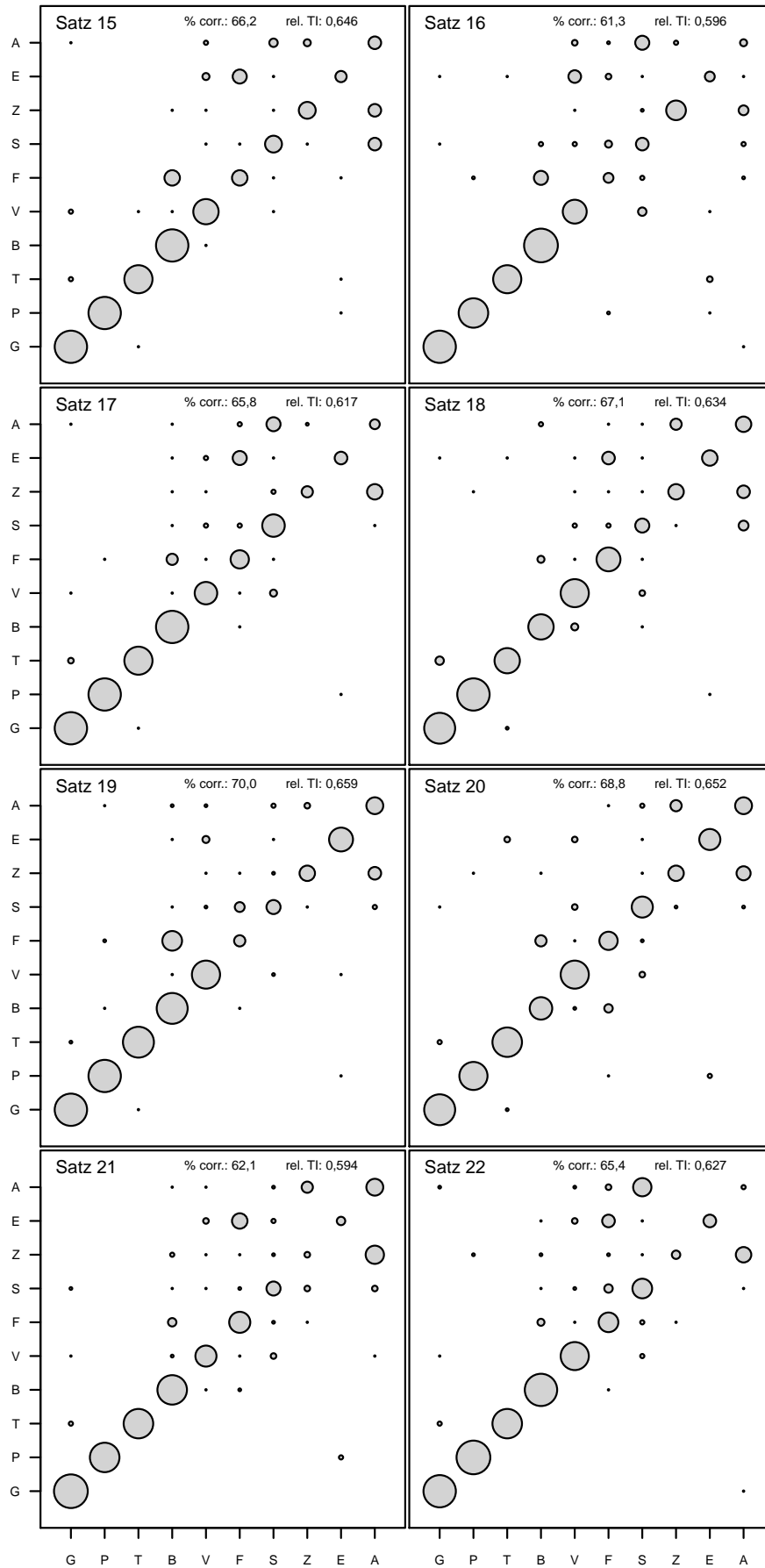


Abbildung G.2: Sätze 7 - 14

dargebotene Emotion



erkannte Emotion

Abbildung G.3: Sätze 15 - 22

15-silbige Sätze

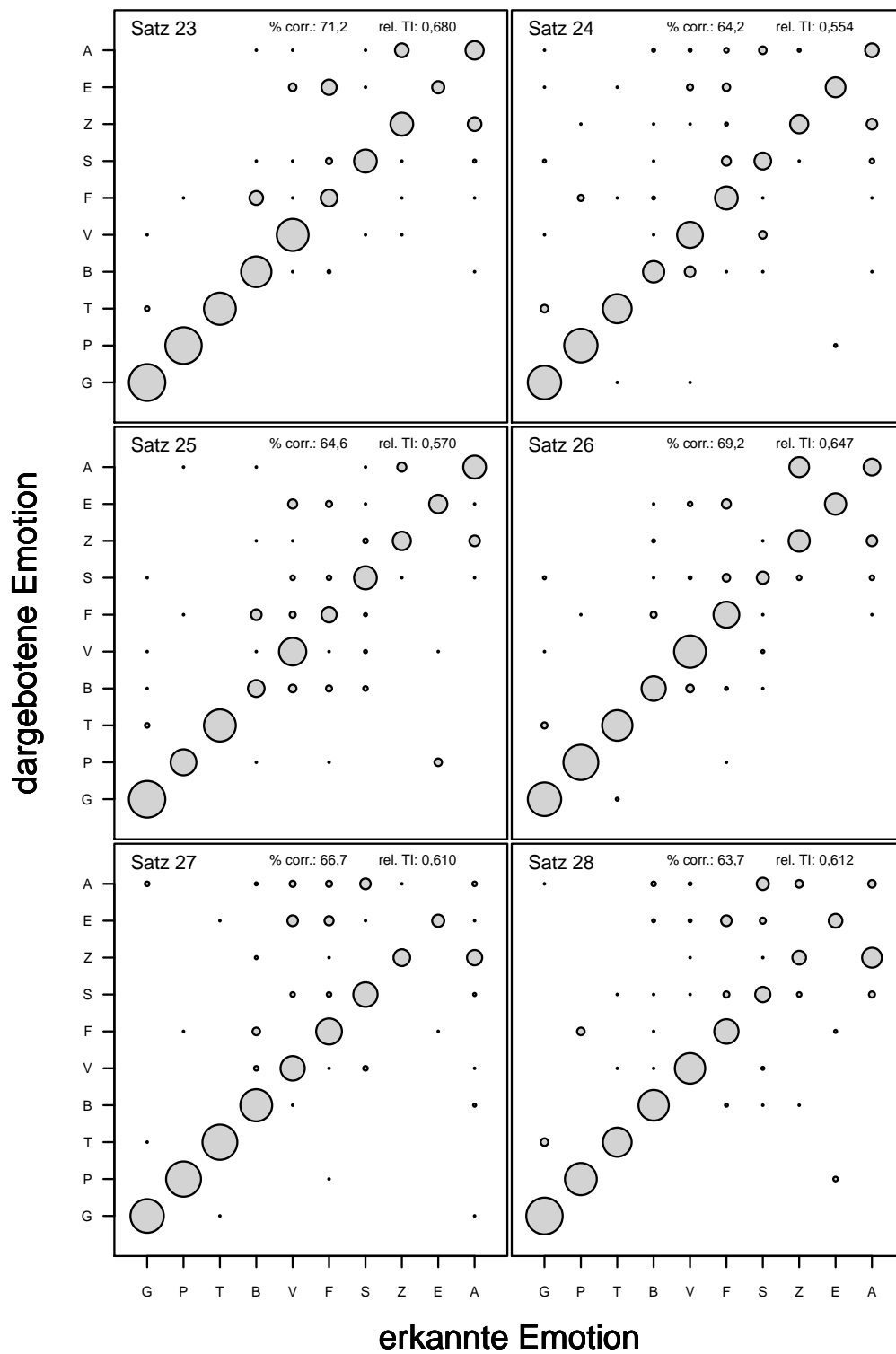
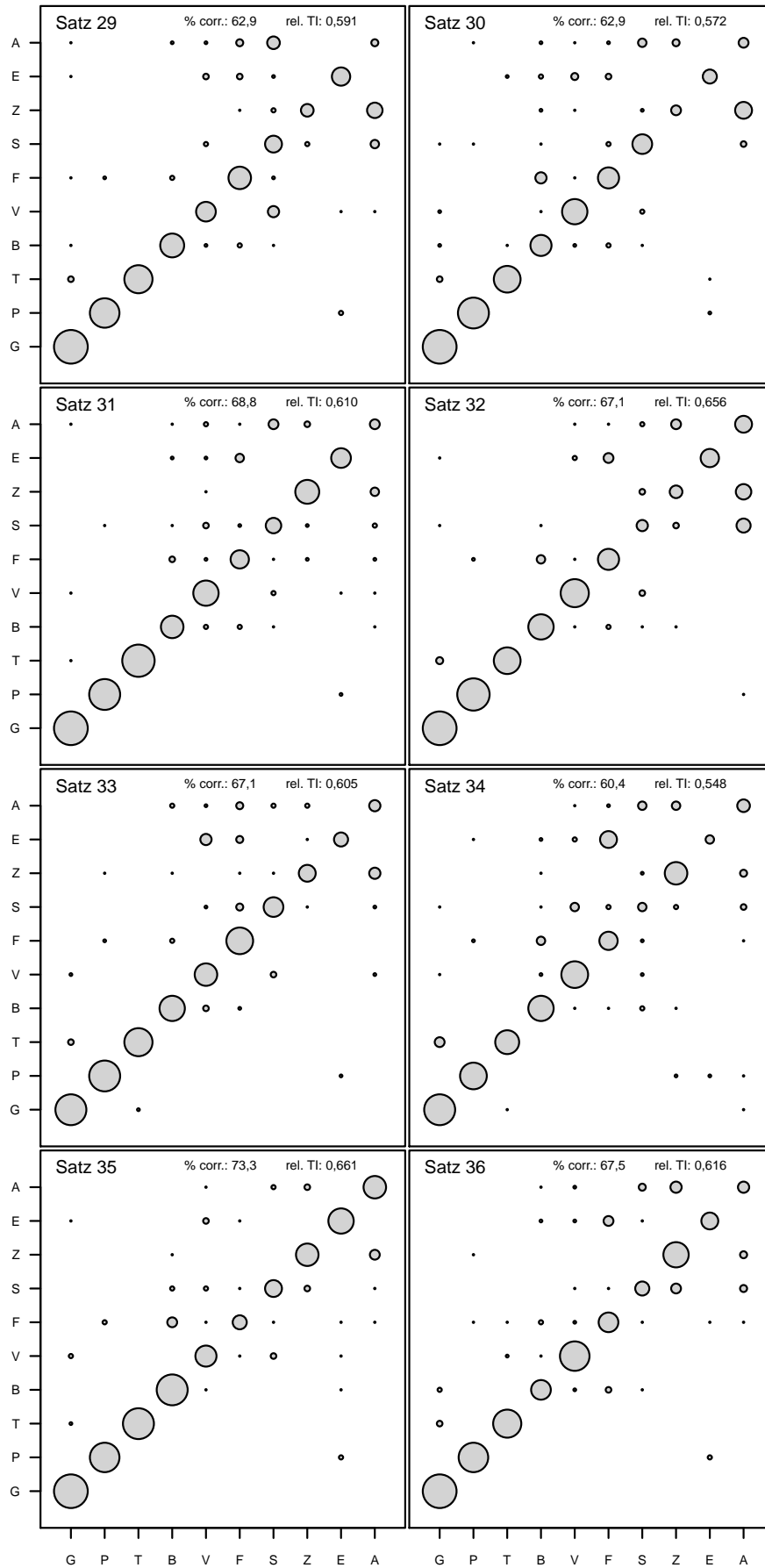


Abbildung G.4: Sätze 23 - 28

dargebotene Emotion



erkannte Emotion

Abbildung G.5: Sätze 29 - 36

G Resultate aufgeschlüsselt nach Sätzen

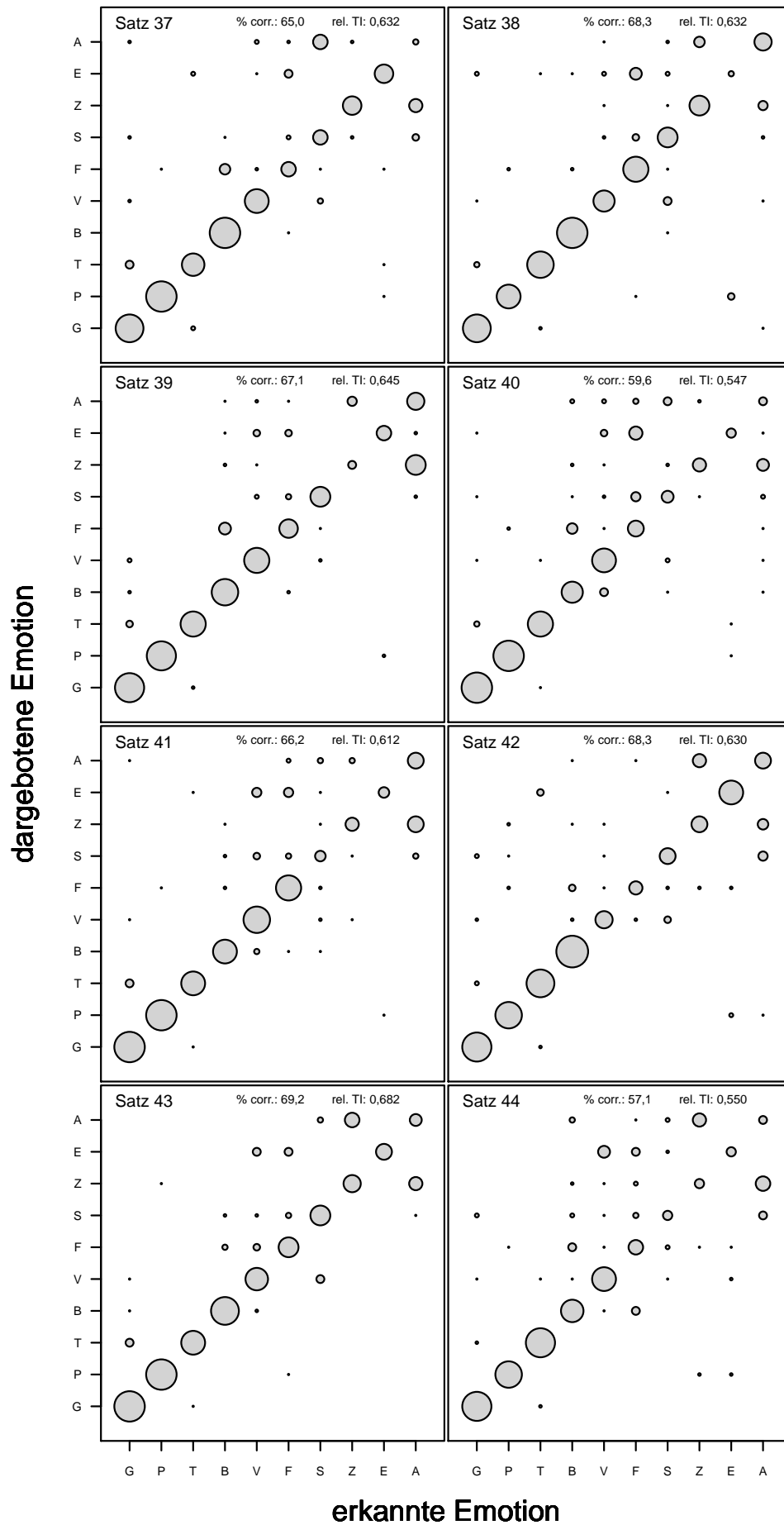


Abbildung G.6: Sätze 37 - 44

Literaturverzeichnis

Horst Arndt und Richard W. Janney. Verbal, prosodic, and kinesic emotive contrasts in speech. *Journal of Pragmatics*, 15:521 – 549, 1991.

Peter Auer und Elisabeth Couper-Kuhlen. Rhythmus und Tempo konversationeller Alltagssprache. *Zeitschrift für Literaturwissenschaft und Linguistik*, 24(96):78, 1994.

Rainer Banse und Klaus R. Scherer. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3):614 – 636, 1996.

Monika Bauer. Zum Erkennen von Stimmqualitäten und prosodischen Mustern bei cochleaimplantierten Patienten. Magisterarbeit, Julius-Maximilians-Universität Würzburg, 2005.

Klaus Berg. *Gedichte im Gedächtnis? Vom Verlust der Gedächtniskultur in und außerhalb der Schule*. Königshausen & Neumann, Würzburg, 2005.

Jürgen Bortz und Nicola Döring. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer Verlag, Heidelberg, 2002.

Leda Cosmides. Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception & Performance*, 9(6):864 – 881, 1983.

Charles Darwin. *The Expression Of The Emotions In Man And Animals*. Oxford University Press, New York, 1998.

Literaturverzeichnis

- Carmen Cordelia Domes. *Evaluierung und Weiterentwicklung eines Prosodie-verständnistests*. Dissertation, Julius-Maximilians-Universität Würzburg, 2009. URL http://www.opus-bayern.de/uni-wuerzburg/volltexte/2010/4901/pdf/Dissertation_Cordelia_Domes.pdf.
- Ecobyte. Bk ReplaceEm Version 2.0, 2010. URL www.ecobyte.com.
- Stephan Euler. *Grundkurs Spracherkennung: Vom Sprachsignal zum Dialog*. Vieweg +Teubner Verlag, Wiesbaden, 2006.
- Robert B. Ewen. *Personality, a topical approach: theories, research, major controversies, and emerging findings*. Lawrence Erlbaum Associates, Hove, 1998.
- Gernot A. Fink. *Mustererkennung mit Markov-Modellen: Theorie-Praxis-Anwendungsgebiete*. Vieweg +Teubner Verlag, Wiesbaden, 2003.
- Sabine Flick. *Emotionen in Geschlechterverhältnissen: Affektregulierung und Gefühlsinszenierung im historischen Wandel*. Transcript Verlag, Bielefeld, 2009.
- Anthony Fox. *Prosodic features and prosodic structure*. Oxford University Press, New York, 2000.
- Robert Frick. Emotion: The role of prosodic features. *Psychological Bulletin*, 97(3): 412 – 429, 1985.
- Hede Helfrich. *Satzmelodie und Sprachwahrnehmung*, Kapitel Die Grundfrequenz als Komponente der sprachlichen Äußerung, Seiten 8 – 15. De Gruyter Verlag, Berlin, New York, 1985.
- Reiner Hellbrück. *Angewandte Statistik mit R: Eine Einführung für Ökonomen und Sozialwissenschaftler*. Gabler Verlag, Wiesbaden, 2009.

- Marc-Thorsten Hütt und Manuel Dehnert. *Methoden der Bioinformatik: Eine Einführung*. Springer, 2006.
- International Telecommunication Union. *ITU-T Recommendation G.227; Conventional Telephone Signal*, Kapitel Fascicle III.2, 11. The Blue Book, 1988.
- Internet-Textquellen, 2007.
- Aleksandr V. Isacenko und Hans Joachim Schädlich. Untersuchungen über die deutsche Satzintonation. *Studia Linguistica*, 18(2):122 – 131, 1964.
- Stephen Krashen. We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4):440 – 464, 1989.
- Udo Kuckartz, Stefan Rädiker, Thomas Ebert und Julia Schehl. *Statistik - Eine verständliche Einführung*. VS Verlag für Sozialwissenschaften, Wiesbaden, 2010.
- Alexander Lerch und Stefan Weinzierl. *Handbuch der Audiotechnik*. Springer Verlag, Berlin, Heidelberg, 2008.
- Maria Mahfoud. *Neuaufsprache und Evaluation des Einsilber-Sprachverständnistests*. Dissertation, Julius-Maximilians-Universität Würzburg, 2009. URL http://www.opus-bayern.de/uni-wuerzburg/volltexte/2010/4854/pdf/Dissertation_Maria_Mahfoud.pdf.
- John D. Mayer und Glenn Geher. Emotional intelligence and the identification of emotion. *Intelligence*, 22(2):89–113, 1996.
- Wulf-Uwe Meyer, Rainer Reisenzein und Achim Schützwohl. *Einführung in die Emotionspsychologie, Bd.1, Die Emotionstheorien von Watson, James und Schachter*. Verlag Hans Huber, Bern, 2001.

Literaturverzeichnis

- Elmar Nöth. *Prosodische Information in der automatischen Spracherkennung*. Max Niemeyer Verlag, Tübingen, 1991.
- Robert F. Orlikoff und Ronald J. Baken. The effect of the heartbeat on vocal fundamental frequency perturbation. *Journal of Speech and Hearing Research*, 32(3): 576 – 582, 1989.
- Beat Pfister und Tobias Kaufmann. *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Springer Verlag, Berlin, Heidelberg, 2008.
- Dana J. Plude, Jim T. Enns und Darlene Brodeur. The development of selective attention: A life-span overview. *Acta Psychologica*, 86(2-3):227 – 272, 1994.
- Robert Plutchik. *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*. American Psychological Association (APA), Washington D.C., 2003.
- Pschyrembel. *Pschyrembel Klinisches Wörterbuch Online*, Kapitel Emotion. De Gruyter Verlag, Berlin, 2010. URL <http://www.wdg.pschyrembel.de>.
- Gert Rickheit und Hans Strohner. *Grundlagen der kognitiven Sprachverarbeitung*, Kapitel Pragmatische Rezeption, Seite 253. Francke Verlag, Tübingen, Basel, 1993a.
- Gert Rickheit und Hans Strohner. *Grundlagen der kognitiven Spachverarbeitung*, Kapitel Das Konzept der kognitiven Sprachverarbeitung, Seite 24. Francke Verlag, Tübingen, Basel, 1993b.
- Gert Rickheit, Theo Herrmann und Werner Deutsch. *Psycholinguistik: Ein internationales Handbuch*. De Gruyter Verlag, Berlin, New York, 2003.

- Ronald E. Riggio. Social interaction skills and nonverbal behavior. In Robert S. Feldman, Hrsg., *Applications of nonverbal behavioral theories and research*, Kapitel Social interaction skills and nonverbal behavior, Seiten 3 – 31. Lawrence Erlbaum Associates, Hove, 1992.
- Christine Römer und Brigitte Matzke. *Lexikologie des Deutschen: eine Einführung*. Gunter Narr Verlag, 2005.
- Wolfgang Rost. *Emotionen. Elixiere des Lebens*. Springer Verlag, Berlin, Heidelberg, 2001.
- Klaus R. Scherer. *Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens*, Kapitel Einführung, Seiten 284 – 303. Beltz Verlag, Weinheim, Basel, 1982.
- Lothar Schmidt-Atzert. *Lehrbuch der Emotionspsychologie*. Kohlhammer Verlag, Stuttgart, 1996.
- Margret Selting. *Prosodie im Gespräch*. Max Niemeyer Verlag, Tübingen, 1995.
- Shari R. Speer. Prosodic structure and sentence recognition. *Journal of memory and language*, 32:336 – 358, 1993.
- Spiegel Online. Projekt Gutenberg, 2010. URL <http://gutenberg.spiegel.de/>.
- Eleni Stavrianou. *Griechischer Satztest nach dem Vorbild des HSM-Tests*. Dissertation, Julius-Maximilians-Universität Würzburg, 2010.
- Gerhard Stemmler. *Psychologie der Emotion (Enzyklopädie der Psychologie : Themenbereich C : Ser. 4 ; Bd. 3)*, Kapitel Somatoviszzerale Aktivierung, Seiten 291 – 327. Hogrefe Verlag, Göttingen, 2009.

Literaturverzeichnis

- Lesley Stirling und Roger Wales. Prosody and parsing. In Paul Warren, Hrsg., *Prosody and parsing*, Kapitel Does prosody support or direct sentence processing?, Seiten 193 – 213. Psychology Press, Hove, 1996.
- Gerhard Strube und Theo Herrmann. *Lehrbuch Allgemeine Psychologie*, Kapitel Sprechen und Sprachverstehen, Seiten 285 – 316. Verlag Hans Huber, Bern, Stuttgart, Toronto, 2006.
- Hartmut Trau Müller. *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society / Ein Internationales Handbuch Zur Wissenschaft Von Sprache Und Gesellschaft*, Kapitel The Social Implications of Levels of Linguistic Analysis / Soziale Implikationen von Sprachanalyse-Ebenen, Seiten 653 – 666. De Gruyter Verlag, Berlin, New York, 2004.
- Dieter Ulich und Philipp Mayring. *Psychologie der Emotionen*. Kohlhammer Verlag, Stuttgart, 2003.
- Universität Leipzig. Wortschatz-Portal, 2010. URL <http://wortschatz.uni-leipzig.de>.
- Harald Wallbott und Klaus Scherer. Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*, 51(4):690 – 699, 1986.
- Andreas Wendemuth. *Grundlagen der stochastischen Sprachverarbeitung*. Oldenbourg Wissenschaftsverlag, München, Wien, 2004.
- Jürgen Wendler, Wolfram Seidner und Ulrich Eysholdt. *Lehrbuch der Phoniatrie und Pädaudiologie*. Thieme Verlag, Stuttgart, New York, 2005.
- Carl E. Williams und Kenneth N. Stevens. Akustische Korrelate diskreter Emotionen. In Klaus R. Scherer, Hrsg., *Vokale Kommunikation. Nonverbale Aspekte des*

Sprachverhaltens, Kapitel Akustische Korrelate diskreter Emotionen, Seiten 307 – 325. Beltz Verlag, Weinheim, Basel, 1982.

Philip G. Zimbardo und Richard J. Gerrig. *Psychologie. Eine Einführung*. Pearson Studium, München, Boston, 2004.

Danksagung

Herrn Prof. Dr. med. Rudolf Hagen, Direktor der Klinik und Poliklinik für Hals-, Nasen- und Ohrenkrankheiten, plastische und ästhetische Operationen der Universität Würzburg und Herrn Prof. Dr. med. Joachim Müller bin ich für die Überlassung des Themas dieser Arbeit und für die Übernahme des Referates zu besonderem Dank verpflichtet. Prof. Dr. med. dent. Bernd Klaiber, Direktor der Poliklinik für Zahnerhaltung und Parodontologie der Universitätszahnklinik Würzburg, danke ich für die Übernahme des Korreferats. In aller Form möchte ich mich bei Herrn Dipl.-Ing. Stefan Brill für die Vergabe des Dissertationsthemas und die Einarbeitung sowie die gute Betreuung während der gesamten Arbeit bedanken. Mit seiner Hilfe war es mir möglich, die technischen Herausforderungen bei der Vorbereitung, Durchführung und Bearbeitung der Aufsprachen zu bewältigen. Er investierte viel Zeit in die gemeinsamen Aufsprachen und unterstützte mich darin, den Versuchsaufbau für die Pilotstudie sowie die Haupttests zu konstruieren, die Ergebnisse auszuwerten und zu interpretieren. Vielen Dank für die außerordentliche Einsatzbereitschaft, die er mir entgegenbrachte, wann immer sie gefragt war. Herr Dipl.-Ing. Stefan Brill vermittelte mir außerdem weitreichende Kenntnisse über die kritische Herangehensweise an Problemstellungen und deren wissenschaftliche Lösung. Gleichzeitig kreierte er ein stets angenehmes und motivierendes Arbeitsklima. Des Weiteren möchte ich den stets hilfsbereiten Mitarbeitern der Abteilung für Cochlea-Implantate danken sowie insbesondere den Teilnehmern des Pilottests. Mein besonderer Dank gilt schließlich den 32 Probanden, die Zeit investierten, viel Ausdauer unter Beweis stellten und ihre Aufgabe sehr ernst nahmen.

