

Studiengang: Computerlinguistik

Prüfer: Prof. Dr. phil. habil. Grzegorz Dogil

Betreuer: Dr. techn. Dipl.-Ing. Wolfgang Wokurek

begonnen am: 8. Juni 2004

beendet am: 26. August 2004

Studienarbeit Nr. 26

Untersuchung der Sprecherindividualität höherer Formanten

Gerhard Kremer



Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung
Azenbergstraße 12
D-70174 Stuttgart

Erklärung

Hiermit versichere ich, diese Arbeit
selbständig verfasst und nur die
angegebenen Quellen benutzt zu haben.

(Gerhard Kremer)

Inhaltsverzeichnis

1. Einführung	1
2. Datengrundlage	5
3. Auswertungsdaten	7
3.1. Wählen der Sprachlaute	7
3.2. Erstellen der Formantfrequenzdaten	8
3.3. Prüfen der erstellten Daten	12
4. Statistische Auswertung	19
4.1. Filtern der Daten	19
4.2. Vergleich der Formantfrequenzverteilungen	20
4.3. Prüfen der Sprecher auf Inkonsistenz	22
5. Zusammenfassung und Schlussbemerkung	26
A. Tabellen und Abbildungen	28
A.1. Inhalt der beigelegten CD-ROM	28
A.2. IPA-SAMPA-Umschrift	30
A.3. Vokalviereck	31
A.4. Datensatzmenge	32
A.5. T-Tests zur Inkonsistenz der Sprecher	36
Literatur	37

Abstract

Stating that higher-frequency formants are specific of a speaker seems to be reasonable. Yet I couldn't find studies providing experimental evidence. In order to see if there were any individual differences in the values from formants F4–F7 I examined data of the 16 speakers from the German “Kiel Corpus of Spontaneous Speech”. Using tools from Entropic, I created the necessary formant data on the basis of the LPC analysis method. A Perl script then extracted the relevant data of the vowels [a], [i], [u], and the nasals [m], [n], and [ŋ]. I used the t-test and ANOVA from the statistics tool R to analyse the data. The results showed highly significant differences in the distributions of all the seven formants, when they were tested on the factors speaker, sound, and the combination of both.

Kurzfassung

Die Behauptung, höhere Formanten seien sprecherspezifisch, scheint angemessen zu sein. Dennoch konnte ich keine Studien dazu finden, die dies belegen. In der vorliegenden Arbeit untersuchte ich die Individualität der Formanten F4–F7 der 16 Sprecher aus dem deutschen „Kiel-Korpus der spontanen gesprochenen Sprache“. Auf der Grundlage einer LPC-Analyse ließ ich daraus mit Hilfe von Programmwerkzeugen von Entropic die Formantfrequenzdaten aller Sprachsignaldateien erzeugen. Mit einem Perl-Skript extrahierte ich die Daten der Vokale [a], [i], [u] und der Nasale [m], [n] und [ŋ]. Ich wertete die Daten mit den statistischen Verfahren T-Test und ANOVA aus, wofür ich das Statistik-Programm R benutzte. Die Verteilungen der Formantfrequenzdaten unterschieden sich hochsignifikant bei allen sieben Formanten in Abhängigkeit der Faktoren Sprecher, Sprachlaut und der Sprecher-Sprachlaut-Kombination.

1. Einführung

Das Erkennen von Sprachschall ist ein Ziel im Fachgebiet akustische Phonetik. Nachdem festgestellt ist, dass ein akustisches Signal Sprachschall enthält, folgen weitere Untersuchungen. Das Sprachsignal wird herausgefiltert, segmentiert, und die Segmente werden klassifiziert. Um das Sprachsignal in einzelne Laute zu segmentieren, mag die visuelle Darstellung des Signals in einem Amplituden-Zeit-Diagramm der komplexen Schwingung noch ausreichen. Um den Segmenten jedoch die richtigen Laute zuzuordnen ist es nötig, die Spektrogramm-Darstellung zu wählen. Die Begründung dafür beruht auf den folgenden Überlegungen.¹

Gemäß dem Quelle-Filter-Modell der Sprachproduktion wird der an der Stimmquelle erzeugte Schall im Rachenraum gefiltert. Der erzeugte Schall besteht zunächst aus mehreren Schwingungsanteilen: der Grundfrequenz und ihren Harmonischen, den ganzzahligen Vielfachen der Grundfrequenz. Dabei besitzt die Grundfrequenz die größte Amplitude. Die Teilschwingungen mit höheren Frequenzen haben kontinuierlich kleiner werdende Amplituden. Die Amplitude einer Schwingung hängt mit der Schallintensität zusammen. Im Rachenraum werden die Amplitudenwerte der einzelnen Teilschwingungen durch dessen Resonanzeigenschaften verändert. Im Bereich der Resonanzfrequenzen werden sie verstärkt, ansonsten abgeschwächt, d. h. der Betrag ihrer Amplituden wird vergrößert bzw. verkleinert. Dadurch entstehen im Schall mehrere prominente Frequenzanteile, die sich durch eine höhere Schallintensität als andere Frequenzanteile im näheren Bereich auszeichnen. Mit der Veränderung der Form des Rachenraums mit Hilfe von Zunge, Kiefer und Lippen ändern sich dessen Resonanzeigenschaften und die damit produzierbare Lautqualität.² Deshalb sind Sprachlaute im Schall unter anderem durch ihre Resonanzspitzen, d. h. die lokal prominenten Frequenzanteile, unterscheidbar und klassifizierbar.

Im Spektrum – einem Amplituden-Frequenz-Diagramm – sind die Teilschwingungen eines Signals durch senkrechte Striche dargestellt. Die Position eines Strichs auf der Frequenzachse bestimmt den Frequenzwert einer Teilschwingung. Die Strichlänge gibt ihren Amplitudenwert an, der an der Amplitudenachse abgelesen werden kann. Diese Darstellung wird durch die sog. Fourier-Transformation erzeugt. Die Grundlage zu dieser Berechnungsmethode ist die Annahme, dass sich jedes komplexe, quasi-periodische Signal durch eine Überlagerung mehrerer Sinusschwingungen unterschiedli-

¹Vgl. CLARK & YALLOP (1995, Kap. 7) für eine detailliertere Einführung.

²Eine visuelle Verdeutlichung dieser Tatsache ist das Vokalviereck, das in Anhang A.3 abgebildet ist.

cher Amplitude und Frequenz beschreiben lassen kann. Um Lautänderungen im Zeitverlauf betrachten zu können, müssen mehrere kurz aufeinanderfolgende Spektren betrachtet werden.

Eine geeignete und gebräuchliche Darstellung dafür ist das Spektrogramm – ein Frequenz-Zeit-Diagramm, in dem der Grad der Prominenz der einzelnen Frequenzanteile eines Signals als dritte Dimension durch unterschiedliche Einfärbungen verdeutlicht wird. Bei Spektrogrammen wird unterschieden zwischen Schmalband- und Breitband-Spektrogrammen. Ein Schmalband-Spektrogramm wird erzeugt, indem die einzelnen Frequenzspektren aus längeren Signalabschnitten als beim Breitband-Spektrogramm berechnet werden. Die Filterfrequenz gibt an, wie oft Spektren des Signals erzeugt werden. Sie bestimmt die Länge der einzelnen Analysefenster, aus denen jeweils Spektren erzeugt werden. Schmalband-Spektrogramme werden mit einem schmalbandigen Analysefilter erzeugt (üblicherweise 50 Hz, Länge des Analysefensters: 20 ms), Breitband-Spektrogramme mit einem breitbandigen (üblicherweise 300 Hz, Länge des Analysefensters: 3,3 ms). Dadurch ist die Frequenzauflösung bei Breitband-Spektrogrammen schlechter als bei Schmalband-Spektrogrammen, die Zeitauflösung ist jedoch besser. Ein Vorteil von Breitband-Spektrogrammen besteht darin, dass Lautgrenzen zeitgenauer bestimmt werden können. Breitband-Spektrogramme werden außerdem zur Formantanalyse benutzt, da die Formanten einfacher abgelesen werden können: Nur ganze Frequenzbereiche höherer Schallintensität sind sichtbar – in Schmalband-Spektrogrammen dagegen alle einzelnen Frequenzanteile.

Die Resonanzfrequenzen des Rachenraums werden als Formanten bezeichnet (vgl. STEVENS, 1999). Da an diesen Stellen im Frequenzbereich die Schwingungen des Sprachsignals am Meisten verstärkt wurden, kann in einem Breitband-Spektrogramm auf Grund der Frequenzbereiche mit der lokal höchsten Schallintensität auf die Lage der Formanten geschlossen und ihr Zeitverlauf verfolgt werden. Die Lageänderung der Formanten im Zeitverlauf hilft bei der Lautunterscheidung, ihre Lage im Frequenzbereich bei der Bestimmung der Lautqualitäten. Durch Formänderungen des Racheninnenraums ändern sich die Frequenzwerte der Formanten, weshalb sich die prominenten Frequenzen auf der Frequenzachse im Spektrogramm verschieben. Die Lage der Formanten lässt begrenzt Rückschlüsse auf die Konstellation des Rachenraums zu. Zwischen dem Öffnungsgrad des Kiefers und dem ersten Formanten (F1)³, zwischen der Zungenposition und dem zweiten Formanten (F2) sowie zwischen der Lippenrundung und dem dritten Formanten (F3) wurde eine Korrelation festgestellt. Dadurch kann von ei-

³Die Nummerierung beginnt beim Formanten mit dem kleinsten Frequenzwert.

nem Formantmuster aus den ersten drei Formanten auf die Kiefer-Zunge-Lippen-Konstellation und die produzierte Vokalqualität geschlossen werden (vgl. STEVENS, 1999).

Die Form und die Größe des Rachenraums bestimmen die Lage der Formanten; sie können vom Sprecher durch Zungen-, Kiefer- und Lippenstellung beeinflusst werden. Da der restliche Teil des Rachenraums beim Sprechen wenig verändert werden kann und sich die Rachenraumformen zweier beliebiger Sprecher voneinander unterscheiden, ist die Frage zu stellen, ob sprecherindividuelle Unterschiede beim Untersuchen von Formantdaten festgestellt werden können.

Ich fand dazu in MALMBERG (1963, S. 18) den Hinweis:

The formants situated in the upper part (Formants 3, 4, etc.) [of the spectrogram] are characteristic of the speaker but have no linguistic value properly speaking. (After Potter, Kopp, Green.)

Hier wird ausgesagt, dass die Formanten ab F3 im Gegensatz zu F1 und F2, die als lautunterscheidend behandelt werden, eher sprecherindividuell sind. Dem Literaturverweis folgend fand ich in POTTER ET AL. (1947) jedoch keine entsprechende Erwähnung.

In MACHELETT & TILLMANN (1996, Kap. 2.1) erscheint der Hinweis:

Während erster, zweiter und dritter Formant vokalspezifisch sind, d. h. relativ sprecherunabhängig immer annähernd gleiche Frequenzwerte annehmen, sind die Frequenzwerte ab dem vierten Formanten überwiegend für Klangfarbe und Charakteristik der Sprecherstimme verantwortlich. Sie dienen in erster Linie der Identifikation eines Sprechers und nicht eines Vokals.

Diese Aussage wird dort weder nachgewiesen, noch wird sie durch Literaturangaben belegt. Ich fand bisher keine weiteren Angaben oder Studien dazu, wie höhere Formanten mit Sprecherindividualität zusammenhängen.

In der vorliegenden Arbeit untersuche ich, ob es Anzeichen für die mögliche Sprecherindividualität höherer Formanten gibt. Dazu betrachte ich die Frequenzwerte der Formanten F4–F7 verschiedener Sprecher und vergleiche diese Datenmengen miteinander. Von den Einflussgrößen, die auf die Lage der Formanten einwirken (Stimmung, Gesundheitszustand, etc.) beziehe ich die beiden Größen ‚produzierte Lautqualität‘ und ‚lautproduzierendes Individuum‘ in die Untersuchung mit ein. Ich benutze als Grundlage das Datenmaterial aus dem deutschen „Kiel-Korpus der spontanen gesprochenen Sprache“, das in Kapitel 2 genauer beschrieben ist. Aus dem Kiel-Korpus

extrahiere ich die Formantfrequenzen der Vokale [a], [i], [u] und der Nasale [m], [n], [ŋ]. Zu diesem Zweck schrieb ich ein Perl-Skript ([PERL FOUNDATION, 2004](#)). Die genaue Vorgehensweise bei der Extraktion der Daten bzw. die Funktion des dazu nötigen Perl-Skripts erläutere ich in Kapitel 3. Ich beschreibe dort außerdem, wie sichergestellt wird, dass die Daten korrekt extrahiert werden. Anschließend beschreibe ich in Kapitel 4 die Auswertung der Formantdaten mit statistischen Methoden und das Ergebnis. Im letzten Kapitel fasse ich die Untersuchungsergebnisse zusammen und beschreibe Möglichkeiten zu weiterführenden Untersuchungen.

Die Sprachsignaldateien und Transkriptionsdateien des Kiel-Korpus, die extrahierten Formantdaten, die benutzten Skripte und R-Befehle sind dieser Arbeit auf einer CD-ROM beigelegt; ein Inhaltsverzeichnis mit Beschreibungen befindet sich in Anhang [A.1](#).

2. Datengrundlage

Für die vorliegende Untersuchung benutzte ich Sprachsignaldateien aus dem deutschen „Kiel-Korpus der spontanen gesprochenen Sprache“ (The Kiel Corpus of Spontaneous Speech Volume II, 1995), das am Institut für Phonetik und Digitale Sprachverarbeitung (IPDS) der Christian-Albrechts-Universität Kiel erstellt wurde.⁴ Neben den Aufnahmen enthält das aufbereitete Korpus die dazugehörigen Transkriptionsdateien, die nützlich sind, da sie zusätzlich zu den suprasegmentalen die segmentalen phonetischen Annotationen enthalten, die Transkription der segmentierten Äußerung. So können die Daten einzelner Sprachlaute gezielt untersucht werden.

Das Kiel-Korpus besteht aus 862 Sprachsignaldateien von insgesamt 16 Sprechern. Die Sprachaufnahmen entstanden, indem je zwei Sprechern in insgesamt acht Sitzungen jeweils die Aufgabe gestellt wurde, eine Terminvereinbarung anhand ausgehändigter Kalenderblätter miteinander zu verhandeln und diese Dialoge aufgezeichnet wurden. Die Sprecher konnten sich während des Dialogs nicht sehen. Es wurde sichergestellt, dass nicht beide Sprecher gleichzeitig redeten. Dazu musste jeder Sprecher während des Sprechens einen Knopf drücken, um von seinem Gesprächspartner über Kopfhörer gehört zu werden. Gleichzeitig wurde beiden angezeigt, wer gerade den Knopf gedrückt hielt. War einer der beiden Knöpfe gedrückt, blockierte das die Funktion des anderen Knopfs. Jedes erfolgreiche Knopfdrücken verursachte die Aufnahme und Speicherung des Sprachsignals in einer neuen Datei. Die Sprachsignale wurden digital mit einer Abtastrate von 16 kHz und einer Quantisierungsgenauigkeit von 16 Bit aufgenommen.

Danach wurden die Transkriptionen der Aufnahmen automatisch erzeugt und anschließend manuell korrigiert, bevor die Transkription jeder Sprachsignaldatei jeweils in einer gleichnamigen Datei mit anderer Dateiendung gespeichert wurde. Die so entstandenen Transkriptionsdateien enthalten die orthografische Umschrift der Äußerung und die daraus automatisch erzeugte kanonische Transkription, die auf einem modifizierten SAMPA-System⁵ basiert. Zusätzlich enthalten sie die manuell korrigierte Transkription der segmentierten Äußerung, die für die vorliegende Untersuchung benötigt wird, und die Annotationen prosodischer Merkmale. Jedem Segment können mehrere Etikettierungen (Label) zugeordnet sein, die unterschiedliche Merkmale des Segments beschreiben und zeilenweise voneinander getrennt sind. Zu jedem Label ist eine Sample-Nummer angegeben, die die entsprechende Seg-

⁴Kurzbeschreibung des Kiel-Korpus: URL <http://www.ipds.uni-kiel.de/publikationen/kcssp.de.html> (Zugriffdatum: Juli 2004)

⁵Die IPA-SAMPA-Umschrift-Tabelle zum Kiel-Korpus befindet sich in Anhang A.2.

mentierungsstelle bezeichnet, bis zu der das annotierte Label gültig ist. Die Sample-Nummer gibt die Anzahl der Abtastungen an, die bei der Digitalisierung eines Signals bis zu dieser Stelle vorgenommen wurden. Wenn die bei der digitalen Aufnahme benutzte Abtastrate bekannt ist, kann aus einer Sample-Nummer der entsprechende Aufnahmezeitpunkt errechnet werden.

Das Format der aller Label, die in den Transkriptionsdateien vorkommen, ist in KOHLER ET AL. (1995) ausführlich beschrieben. Im Folgenden stelle ich die für die Datenextraktion bei der vorliegenden Untersuchung wichtigen Annotationskonventionen für Sprachlaut-Label vor:

1. Morphologie

- a) Wortbeginn: ##
- b) wortinterner Kompositumsbeginn: \$#
- c) ansonsten (wortintern): \$

2. Betonung

- a) Hauptbetonung: '
- b) Nebenbetonung: "

3. Laut bzw. Korrektur der kanonischen Transkription

- a) korrekter Laut, wie in kanonischer Transkription
- b) Lautkorrektur: *falscher Laut - korrekter Laut*
- c) Lauteinfügung: - *Laut*
- d) Lautlöschung: *Laut -*

4. Funktionswortende: +

Ein Labeleintrag für einen Sprachlaut besteht aus einer Kombination dieser Annotationen in der angegebenen Reihenfolge. Optional sind alle bis auf die morphologische Annotation und einem darauffolgenden Laut. Detailliertere Informationen zum Sprachkorpus sind in KOHLER ET AL. (1995) aufgeführt.

3. Auswertungsdaten

In diesem Kapitel wird die Erzeugung der Formantdaten für die darauffolgende Auswertung beschrieben. Auf der Grundlage der Dateien des Kiel-Korpus werden die Frequenzwerte der ersten sieben Formanten⁶ ausgewählter Sprachlaute erzeugt. Nach der Datenermittlung wird geprüft, ob die korrekten Daten erzeugt wurden.

3.1. Wählen der Sprachlaute

Die Formantfrequenzdaten sollen für die Vokale [a], [i], [u], die nasalierten Vokale [ã], [ɛ̃], [ɔ̃], [œ̃], und die nasalen Konsonanten [m], [n], [ɲ] ermittelt werden. Die Vokale [a], [i], [u] gehören zu den Extrempunkten des Vokalvierecks (abgebildet in Anhang A.3), das die Kiefer-Zunge-Lippen-Konstellation jedes Vokals angibt. Die vier Extrempunkte des Vokalvierecks werden oft zu akustisch-phonetischen Untersuchungszwecken benutzt. Die Laute [ɑ] bzw. [ɒ] gehören zwar zu den Extrempunkten, können aber nicht in die Untersuchung einbezogen werden, denn sie kommen im deutschen Lautinventar nicht vor. Vokale haben typischerweise eine höhere Schallintensität als Konsonanten – der Signal-Rausch-Abstand ist größer. Dadurch sind bei Vokalen genauere Messungen als bei Konsonanten möglich.

Die Untersuchung der nasalen Laute ist mit Einschränkungen für die Untersuchung sprecherindividueller Merkmale empfehlenswert (vgl. [NOLAN, 1997](#), S. 750):

Nasal sounds have often attracted those looking for reliable cues to speakers' identity [...] because the shape and size of the nasal cavity is both highly variable between speakers and not volitionally alterable. Nasals might thus be the best candidates to provide acoustic manifestations of anatomical individuality. There are, however, two factors which complicate the situation: firstly, the resonance effects of the nasal cavity are not available in isolation, but are combined in a complex way with those of the rest of the vocal tract [...] and, secondly, nasal resonances are highly susceptible to changes in health [...].

⁶Beim Sprachsignal eines Erwachsenen ist maximal die Anzahl der Formanten ermittelbar, die man durch Halbieren der Abtastrate des Signals in kHz erhält; **formant** unterstützt bis zu sieben Formanten (vgl. [ENTROPIC RESEARCH LABORATORY, 1998](#)).

3.2. Erstellen der Formantfrequenzdaten

Zur automatischen Erstellung der Daten für meine Untersuchung schrieb ich das Skript `formanting.perl` in der Programmiersprache Perl ([PERL FOUNDATION, 2004](#)), das mehrere ESPS-Programme⁷ nutzt: `btopsp`, `formant` und `pplain`. Außerdem wird `s1h2phones` (geschrieben von Stefan Rapp) ausgeführt, ein GAWK-Skript ([FREE SOFTWARE FOUNDATION, 2003](#)), das die Transkriptionsdateien des Kiel-Korpus in das passende Format für `waves` umwandelt. Das grafische ESPS-Werkzeug `waves` ist nach der Beschaffung der Daten nützlich bei ihrer Überprüfung. Ich beschreibe im Folgenden die genaue Funktion von `formanting.perl`, das dazu dient, die Frequenzwerte der ersten sieben Formanten aus den Dateien des Kiel-Korpus zu ermitteln und sie in geeigneter Form zu speichern, so dass sie anschließend statistisch ausgewertet werden können.

Dateinamen und Pfade finden

Im ersten Schritt des Perl-Skripts wird in der gesamten Verzeichnisstruktur des Kiel-Korpus nach Sprachsignaldateien gesucht, die erkennbar sind an den Endungen `l16` bzw. `r16`. Die Verzeichnisstruktur ist zweistufig angelegt, um die Dateien zu gruppieren: Die Dialogsitzungen sind voneinander getrennt und in einzelne Aufgabensitzungen aufgespalten. Innerhalb jedes Aufgabensitzungsverzeichnisses sind alle Sprachsignaldateien durchnummeriert und ihr Name so gewählt, dass sie dem richtigen Verzeichnis zugeordnet werden können. Im gleichen Verzeichnis befinden sich auch die entsprechenden Transkriptionsdateien. Da jede Sprachsignaldatei bis auf die Endung den gleichen Namen hat wie die dazugehörige Transkriptionsdatei, wird ihr Dateiname ohne Endung in einer perl-internen Liste einschließlich der Pfadangabe gespeichert. Die entsprechenden Endungen, die den Typ der Datei auszeichnen, werden bei späteren Verarbeitungsschritten je nach dabei benötigtem Dateityp wieder an den Namen angehängt; neu erzeugte oder veränderte Dateien bekommen eine neue Endung, die den Dateityp bzw. den Verarbeitungsschritt eindeutig bestimmt. Damit ist stets eindeutig, welcher Dateityp vorliegt, auf welchem Verarbeitungsstand die Datei ist und zu welcher Äußerung die betrachtete Datei gehört. Alle bei der Verarbeitung entstehenden Dateien werden in einem gemeinsamen Arbeitsverzeichnis gespeichert, das im Perl-Skript festgelegt ist.

Die folgenden Arbeitsschritte sind für jede der gefundenen Sprachsignaldateien gleich.

⁷ESPS: „Entropic Signal Processing System“ ([ENTROPIC RESEARCH LABORATORY, 1998](#)).

Signaldatei umwandeln

Mit `btosps` wird aus der ursprünglichen eine für `waves` lesbare Sprachsignaldatei erzeugt. Die Angabe der Abtastrate in der umgewandelten Datei ist auf den Wert 8 kHz voreingestellt. Über eine `btosps`-Programmooption änderte ich diesen Wert auf 16 kHz, da das die tatsächlich benutzte Abtastfrequenz bei der Aufnahme der Sprachsignale ist. Mit dem voreingestellten Wert für die Abtastfrequenz würden die Sprachsignale bezüglich der Sample-Zeitpunkt-Zuordnung falsch interpretiert. Die konvertierte Datei wird mit neuer Endung im Arbeitsverzeichnis des Perl-Skripts gespeichert.

Formantwerte ermitteln

Mit `formant` werden aus der konvertierten Sprachsignaldatei unter anderem die Frequenzwerte der ersten sieben Formanten ermittelt. Mehrere Ausgabedateien werden erzeugt. Da für die Untersuchung nur die Formantfrequenzdaten benötigt werden, reicht die Betrachtung der beiden Ausgabedateien mit den Endungen `fb` und `fb.sig`. Die Datei mit der Endung `fb.sig` hat das für `waves` geeignete Format, um die dort enthaltenen Formantwerte in der Spektrogrammansicht der zugehörigen Sprachsignaldatei anzeigen zu lassen. Diese Möglichkeit ist nützlich zur Überprüfung der Werte und wird in Kapitel 3.3 erläutert. Aus der Datei mit der Endung `fb` können die kodierten Formantdaten extrahiert werden, was unten beschrieben wird. Das Verzeichnis, in dem die beiden Dateien gespeichert werden sollen, kann als Option beim Programmaufruf angegeben werden. Ich legte hier das Arbeitsverzeichnis des Perl-Skripts fest. Beim Aufruf von `formant` können außerdem noch weitere voreingestellte Parameter verändert werden.

Folgende, bei der Formantanalyse interessierenden Parameter ließ ich unverändert:

- Präemphase-Konstante: 0,7,
- Fenstertyp: \cos^4 ,
- Fensterlänge, über die jeweils analysiert wird: 49 ms,
- Schrittweite, mit der das Analysefenster verschoben wird: 10 ms,
- LPC-Analyse-Methode: Autokorrelation.

Folgende Parameter änderte ich auf die angegebenen Werte:

- Anzahl zu ermittelnder Formanten: 7 (F1–F7),

- Ordnungszahl⁸ der LPC-Analyse: 18,
- Abtastfrequenz der zu analysierenden Datei: 16 kHz,
- Nominalfrequenz⁹ des ersten Formanten: 570 Hz.

Bei der Verarbeitung durch `formant` wird das Sprachsignal zuerst hochpassgefiltert; der „Cut-off“ erfolgt bei ca. 80 Hz. Nach der Präemphasierung folgt die LPC-Analyse (LPC: „linear predictive coding“), aus der die Formantfrequenzen für jedes betrachtete Analysefenster geschätzt werden. Die Formantverläufe, betrachtet über alle Analysefenster, werden mit einem modifizierten Viterbi-Algorithmus optimiert. Das Ziel ist deren größtmögliche Kontinuität.

Sprecherkürzel ermitteln

Aus der Transkriptionsdatei wird mit Hilfe eines regulären Ausdrucks das Sprecherkürzel ermittelt. Das Sprecherkürzel ist eine Kombination aus drei Buchstaben und soll zusammen mit den Ergebnisdaten abgespeichert werden, um eine korrekte Zuordnung der Formantdaten zu den Sprechern zu ermöglichen. Einfachere Methoden, die beispielsweise nicht erfordern, jede Transkriptionsdatei zu öffnen und darin nach dem Sprecherkürzel zu suchen, sind ausgeschlossen. Der Grund ist, dass von zwei Sprechern bei einem Dialog nicht vorhersehbar abwechselnd Sprachsignaldateien aufgenommen wurden. Die Aufnahme einer neuen Sprachsignaldatei wurde nicht durch den Sprecherwechsel ausgelöst, sondern dadurch, welcher Sprecher zuerst auf den dafür vorgesehenen Knopf drückte. Dadurch sind aufeinanderfolgende Nummerierungen der Sprachsignaldateien keine Hinweis darauf, dass diese von verschiedenen Sprechern stammen.

Transkriptionsdatei umwandeln

Mit `s1h2phones` wird die Transkriptionsdatei in eine Label-Datei umgewandelt. Diese Label-Datei kann bei der Durchsicht der Daten in `waves` benutzt werden. Sie enthält alle Label für die einzelnen Segmente. Die dazugehörigen Sample-Nummern der Transkriptionsdatei sind in die entsprechenden Zeitpunkte umgerechnet.

⁸Das Minimum der Ordnungszahl hängt von der Formantanzahl ab: $\text{Ordnungszahl-Minimum} = (\text{Formantanzahl} \cdot 2) + 4$ (vgl. [ENTROPIC RESEARCH LABORATORY, 1998](#)).

⁹Eine Begründung für die Änderung dieses Parameters folgt in Kapitel 3.3.

Relevante Formantfrequenzwerte extrahieren

Aus der Label-Datei werden vom Perl-Skript zunächst die Start- und Endzeiten der zu untersuchenden Laute extrahiert. Ein regulärer Ausdruck wird benutzt, um die Segmentlabel nach den gewünschten Lauten – unter Beachtung der Annotationskonventionen (s. Kapitel 2) – zu durchsuchen. Es reicht nicht aus, nach Vorkommen der Laute in SAMPA-Umschrift zu suchen, ohne auf den Rest des Label-Eintrags zu achten. Andere, suprasegmentale Merkmale sind mit denselben Zeichen annotiert, die auch in der SAMPA-Umschrift benutzt werden. Außerdem wurden Label während der Korpus-Erstellung so korrigiert, dass die automatisch erzeugte, falsche Transkription im Label-Eintrag erhalten blieb. Ein Label-Eintrag kann deshalb einen falschen Laut enthalten.

Anhand der relevanten Zeitabschnitte werden im nächsten Schritt die ihnen entsprechenden Formantfrequenzwerte extrahiert. Die Formantfrequenzwerte werden von `formant` für einzelne Signalabschnitte¹⁰ geschätzt und ohne Zeitangabe ausgegeben. Stattdessen wird jeder Analyse der jeweils sieben Formantwerte von F1–F7, einem Datensatz, eine fortlaufende Datensatznummer zugeteilt. Deshalb müssen zunächst die Zeitabschnitte, deren Formantfrequenzwerte extrahiert werden sollen, in die entsprechenden Datensatznummern umgerechnet werden. Danach werden mit `pplain` die Formantfrequenzwerte der ausgewählten Zeitabschnitte ausgegeben, die auf zwei Dezimalstellen hinter dem Komma gerundet sind. Von `pplain` werden die Formantfrequenzwerte von F1 auf drei Dezimalstellen gerundet.

Daten speichern

Die ermittelten Daten werden in der Datei `formantresults` gespeichert. In jede Zeile wird dabei ein Datensatz Formantfrequenzwerte der Formanten F1–F7 geschrieben. Zusammen mit jedem Datensatz werden die dazugehörigen Angaben Sprecherkürzel, der Sprachlaut in SAMPA-Umschrift und der Dateiname der Sprachsignaldatei (ohne Dateiendung) in derselben Zeile festgehalten. Zusätzlich wird dem Datensatz eine fortlaufende Nummer, die Lautnummer, zugeordnet und in der gleichen Zeile ausgegeben. Jeder Datensatzwert ist durch ein Leerzeichen von den anderen getrennt. Die Datensätze jeder Sprachsignaldatei werden dabei alphabetisch nach Lauten sortiert. In

¹⁰Diese Signalabschnitte entstehen durch Anwendung eines Analysefensters der Länge 49ms, das mit einer Schrittweite von 10ms über die Sprachsignaldatei geschoben wird. N.B.: Diese Fensterlänge darf nicht mit der Fensterlänge zur Erstellung von Breitband-Spektrogrammen in Zusammenhang gebracht werden.

der ersten Zeile von `formantresults` sind die Bezeichnungen der Datensatzwerte angegeben, so dass die Reihenfolge der Werte feststeht.

Das Sprecherkürzel und die Lautangabe sind für die Auswertung nötig, da die Daten möglicherweise unter dem Einfluß dieser beiden Größen stehen. Die Lautnummer hat den Zweck, die einzeln produzierten Laute voneinander unterscheiden zu können. Deshalb wird sie bei jeder neuen Datensatzreihe eines Lauts erhöht. Sie ist über die gesamte Formantdatenmenge aller Sprachsignaldateien fortlaufend und beispielsweise notwendig, um die Anzahl der untersuchten Laute zu bestimmen.

Dateien löschen

Die bei der Datenextraktion erzeugten Dateien im Arbeitsverzeichnis des Perl-Skripts werden wieder gelöscht, da sie keinen weiteren Wert für die Untersuchung haben. Ausgenommen ist die Datei `formantresults`, in der nach der Verarbeitung aller Sprachsignaldateien die gesamten Daten für die Auswertung gespeichert sind.

3.3. Prüfen der erstellten Daten

Vor der Datenauswertung prüfte ich zunächst in Testdurchläufen die Zuverlässigkeit der Formantdatenerzeugung und die Funktion des Perl-Skripts `formanting.perl`, dabei besonders die korrekte Datenextraktion. Notwendige Änderungen am Perl-Skript, die zum Verständnis der Untersuchung wichtig sind, werden erwähnt. Für die Datenprüfung änderte ich vorübergehend das Perl-Skript: Keine der erzeugten Dateien wurde gelöscht. Jeder erledigte Arbeitsschritt wurde in einer Datei protokolliert, in die ebenfalls alle Fehlermeldungen umgeleitet wurden. Außerdem ist in dieser Datei kommentiert, wann jeweils eine neue Sprachsignaldatei bearbeitet wurde. So sind dort ausgegebene Fehler leicht zu lokalisieren.

Formantdatenextraktion prüfen

Um die automatische Extraktion der Formantwerte zu prüfen, wählte ich als Stichprobe von jedem Sprecher jeweils die erste Sprachsignaldatei der Dialogsitzung. Aus diesen Dateien werden mit `formant` die Formantdaten ermittelt. Nachdem mit `waves` aus jeder Sprachsignaldatei ihr Breitband-Spektrogramm erzeugt wurde, ließ ich die ermittelten Formantdaten in demselben Schaubild anzeigen (s. Abbildung 1). Dabei werden die ermittelten Formantdaten auf dem Bildschirm als farbige Kästchen angezeigt, deren Farben die Formanten unterscheiden.

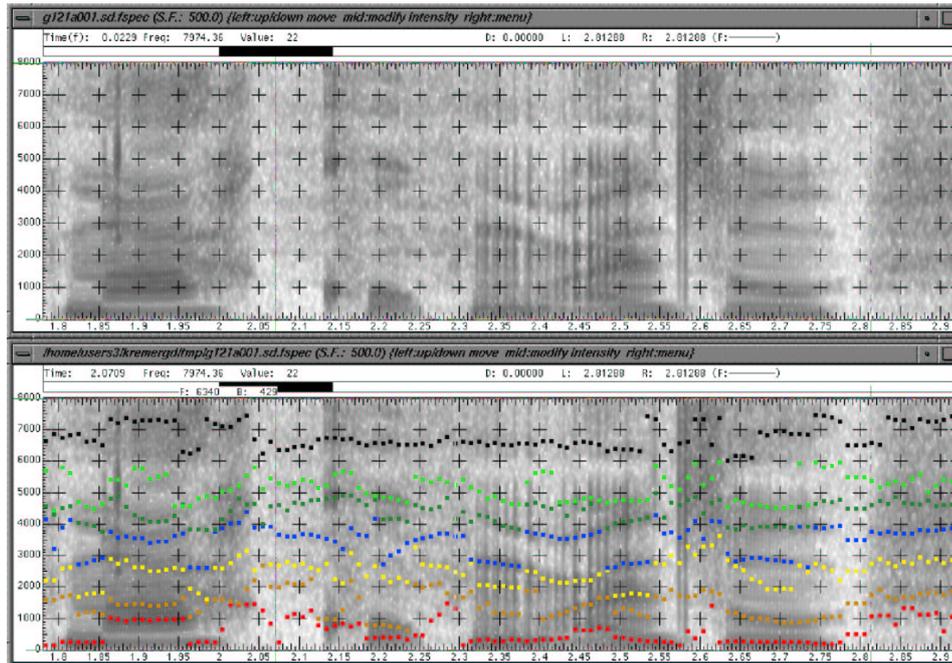


Abbildung 1: Identische Spektrogramme, ohne (oben) und mit (unten) überlagerten Formantwerten.

Dies ermöglichte, die automatisch erzeugten Formantfrequenzen mit den aus dem Spektrogramm visuell schätzbaren zu vergleichen und ihre Korrektheit zu prüfen. Außerdem konnte ich die Ausgabe der Formantfrequenzwerte von `pplain` überprüfen, indem ich sie mit den angezeigten Werten im Spektrogramm verglich. Um gezielt die Formantdaten der zu untersuchenden Laute zu betrachten, kann in `waves` zusätzlich die dazugehörige Label-Datei (Dateiendung: `lab`) geöffnet werden, so dass die Label ebenfalls angezeigt werden.

In den Spektrogrammen, die ich prüfte, waren bei manchen Formantverläufen der von `formant` erzeugten Daten Sprünge erkennbar. Bei manchen Sprüngen erschienen die Formantverläufe nach oben bzw. nach unten verschoben. Manche Formanten wurden in Frequenzbereichen angezeigt, in denen aus dem Spektrogramm visuell keine geschätzt werden konnten. An manchen Stellen im Spektrogramm konnte die genaue Lage der Formanten nicht geschätzt werden, da dort kein lokal begrenztes Amplituden-Maximum sichtbar war. Dort konnten die ermittelten Formantdaten nicht mit den aus

dem Spektrogramm visuell schätzbaren Formantfrequenzwerten verglichen werden.

Die genannten Auffälligkeiten bespreche ich beispielhaft für die anderen überprüften Sprachsignaldateien an Hand von Abbildung 2 auf der nächsten Seite. Dort sind drei Fenster zu sehen. Das oberste Fenster enthält das Breitband-Spektrogramm eines Sprachsignalausschnitts mit den automatisch erzeugten Formantdaten. Darunter, im mittleren Fenster, sind die dazugehörigen Label entsprechend der Segmentierung des Spektrogramms angezeigt. Jedes Label befindet sich an der rechten Markierung des zu bezeichnenden Segments.

Am Ende des Segments mit dem Label »\$' a:« sind bei einigen Formantverläufen Sprünge zu sehen, die aussehen, als wären sie dadurch entstanden, dass an diesen Stellen fälschlicherweise die nächsthöhere Formantnummer für einen Formanten vergeben wurde. Andernfalls wäre der Formantverlauf fließend. Im gleichen Segment kann der siebte Formant aus dem Spektrogramm visuell nicht genau bestimmt werden, weshalb die Genauigkeit der automatisch erzeugten Formantfrequenzdaten hier nicht überprüft werden kann. Am Anfang des Segments mit dem Label »\$n« sind keine Formanten visuell aus dem Spektrogramm schätzbar.

Formantfrequenzextraktion verbessern

Um die Erzeugung der Formantfrequenzdaten zu verbessern, änderte ich den Parameter bei `formant`, der den Nominalfrequenzwert des Formanten F1 festlegt. Dieser Parameter ändert gleichzeitig auch die Nominalwerte der anderen Formanten und ihre Bereichsgrenzen, innerhalb derer sie verlaufen dürfen (vgl. [ENTROPIC RESEARCH LABORATORY, 1998](#)). Bei einem höheren F1-Nominalwert sollte deshalb die falsche Vergabe von nächsthöheren Formantnummern verhindert werden, indem diese Formantnummern auf höhere Frequenzbereiche beschränkt werden.

Beim Experimentieren mit verschiedenen Werten stellte ich fest, dass es nicht möglich ist, den optimalen Parameterwert zu finden. Für jeden der dabei benutzten Werte 600 Hz, 700 Hz und 800 Hz treten in manchen Dateien Formantverlaufssprünge auf. Bei der Überprüfung, ob möglicherweise Formanten der weiblichen Stimmen bedeutend schwieriger oder einfacher erkannt werden als die der männlichen Stimmen, stellte ich keinen Unterschied fest. Bei unterschiedlichen Parameterwerten sind die Formantdaten unterschiedlich zuverlässig, unabhängig vom Genus des Sprechers.

Ich entschied mich für den Parameterwert 570 Hz, bei dem die zugelassenen Bereiche der einzelnen Formantverläufe gleichmäßig auf fast den ge-

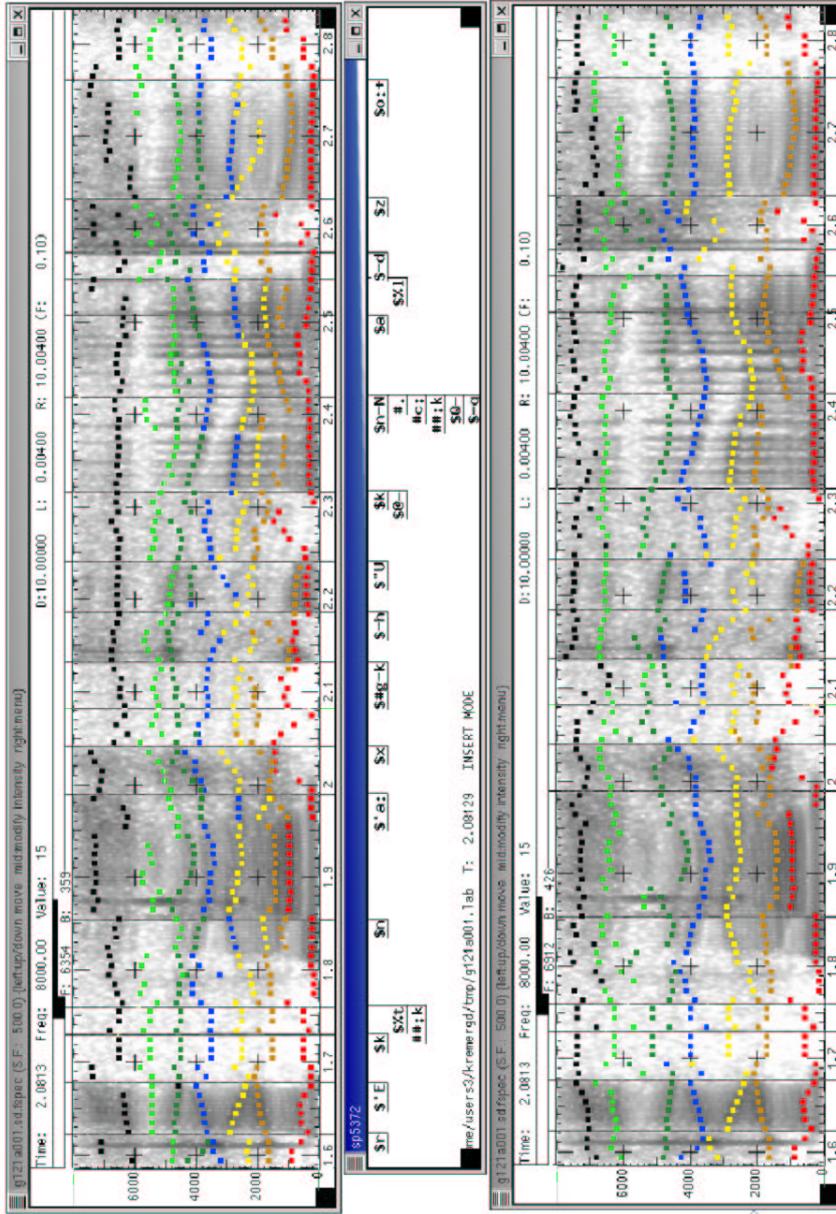


Abbildung 2: Breitband-Spektrogramme, erzeugt unter Angabe unterschiedlicher Werte für den Parameter bei `formant`, der die F1-Nominalfrequenz festlegt: 500 Hz (oben) und 570 Hz (unten).

samten Frequenzbereich von 8 kHz verteilt sind. Die Abstände zwischen den Nominalwerten aufeinanderfolgender Formanten entsprechen dem doppelten Nominalwertbetrag des ersten Formanten, 1040 Hz. Ausgehend vom Nominalwert des jeweiligen Formanten haben die Formantverläufe einen zugelassenen Frequenzbereich von 570 Hz darüber und 570 Hz darunter. Bei sieben Formanten erstreckt sich ihr gesamter Verlaufsgebiet über 7980 Hz.

In Abbildung 2 auf der vorherigen Seite können Formantfrequenzdaten verglichen werden, die mit unterschiedlichen Werten für den Parameter erstellt wurden, der den F1-Nominalfrequenzwert festlegt. Im oberen Fenster wurden die Formanten mit dem voreingestellten Parameterwert 500 Hz erstellt, im unteren Fenster durch Angabe des Werts 570 Hz. Im oberen Fenster scheinen stellenweise Formantdaten zwischen zwei Formanten ermittelt worden zu sein, obwohl im Spektrogramm dort keine zusätzlichen lokalen Amplituden-Maxima erkennbar sind. Im unteren Fenster sind die ermittelten Formantwerte passender zu den lokalen Amplituden-Maxima. Die Änderung des voreingestellten Parameterwerts ergab zusätzliche Sprünge von Formantverläufen an Stellen von Sprachsignalen, an denen unter Angabe des voreingestellten Parameterwerts keine aufgetreten waren (z. B. beim siebten Formanten bei »\$n«). In beiden Fenstern sind Sprünge in den Formantverläufen an den Stellen der zu untersuchenden Vokale zu erkennen. Die Formantverläufe im unteren Fenster sind im Vergleich zu denen im oberen Fenster nicht so sehr aneinander gedrängt, sondern besser auf dem gesamten Frequenzbereich verteilt. Da im unteren Fenster die angezeigten Formantdaten besser zu dem Spektrogramm passen und keine Formantverläufe auftreten, die aussehen, als wären die Formantnummern verschoben worden, bedeutet die Änderung der Nominalfrequenz von F1 auf den Wert 570 Hz eine Verbesserung der Formantdatenextraktion.

Eine einmalige Änderung des Parameterwerts für die Präemphase-Konstante ergab keine sichtbaren Änderungen der Formantverläufe. Eine direkte Abhängigkeit der Formantverläufe von der Präemphase-Konstante ist nicht denkbar, weshalb ich damit nicht weiter experimentierte.

Nominalfrequenzwerte ausschließen

Als Formantfrequenzdaten der wenigen ersten Millisekunden jeder Sprachsignaldatei werden von `formant` die Nominalfrequenzwerte der Formanten ausgegeben. Die Ausgabe dieser Werte führe ich auf die anfängliche Stille im Sprachsignal zurück, bei der scheinbar Formantfrequenzwerte ermittelt werden, obwohl aus dem Spektrogramm aufgrund der Stille visuell keine Formanten schätzbar sind. Ich fand in der Dokumentation zu `formant` keine

Erläuterung dazu. Es scheint, als würden durchgängig für jedes betrachtete Fenster des Sprachsignals Formantfrequenzwerte, und, falls diese nicht schätzbar sind, die Nominalfrequenzwerte der Formanten ausgegeben. Um diese falschen Werte aus der Untersuchungsdatenmenge auszuschließen, änderte ich das Perl-Skript so, dass diejenigen Daten nicht ausgegeben werden, bei denen für alle Formanten eines Datensatzes ihre Nominalfrequenzwerte ermittelt werden. Um zu verhindern, dass in den Analysedaten eine Lücke in der Zählung der extrahierten Laute entsteht, wenn für ein Label keine Werte vom Perl-Skript ausgegeben werden, ist dieser Fall berücksichtigt, indem dann die Lautummer nicht weitergezählt wird.

Fehlerprotokoll-Datei prüfen

In der Fehlerprotokoll-Datei taucht häufig die Fehlermeldung „Exceeded maximum trial count in `_lbpoly`“ auf, die von `formant` erzeugt wird. Sie wird bei der Bearbeitung von 511 Sprachsignaldateien (59,28 % aller bearbeiteten Sprachsignaldateien) ausgegeben, bei jeder dieser Sprachsignaldateien zwischen ein- und zehnmal. Insgesamt kam sie 1117-mal vor. Diese Fehlermeldung sei lediglich eine Angabe zur Überschreitung von Kontinuitätsgrenzen und ignorierbar, da dadurch die Erzeugung der Formantwerte nicht beeinflusst werde (aus einem persönlichen Gespräch zwischen PD Dr. phil. Bernd Möbius, IMS, Universität Stuttgart und David Talkin, Mitentwickler von ES-PS, AT&T Bell Laboratories, stattgefunden 1994).

Label-Extraktion prüfen

Die Funktion von `s1h2phones` prüfte ich stichprobenhaft durch Anhören mehrerer Segmente auf die korrekte Setzung der Sprachlaut-Label und ihrer Start- und Endmarkierungen in der Spektrogrammansicht.

Um zu prüfen, ob die korrekten Label extrahiert wurden, schrieb ich ein Perl-Skript. Es hat die Funktion, alle Zeilen einer Transkriptionsdatei auszugeben, die einen der gewählten Laute als SAMPA-Umschrift enthält. Diese Zeilen werden außerdem von demselben regulären Ausdruck wie in `formanting.perl` auf Kandidaten für die Extraktion überprüft. Das Ergebnis wird jeweils zusammen mit der entsprechenden Zeile ausgegeben. Ich überzeugte mich in Stichproben von der Korrektheit des regulären Ausdrucks für die Label-Extraktion.

Durch einen Vergleich der von `formanting.perl` ausgegebenen Formantfrequenzwerte mit den zugehörigen Werten zu einem Label überzeugte ich

mich in der Spektrogrammansicht von der Korrektheit der Umrechnung von Sample-Nummern in Zeitpunkte.

Datensatzmenge prüfen

In der Tabelle in Anhang [A.4](#) ist die Menge der gesamten Datensätze für die Auswertung, gruppiert nach Sprechern und Sprachlauten, aufgeführt. In der Datensatzmenge kommen keine nasalierten Vokale vor. Ich stellte fest, dass diese Laute im Kiel-Korpus nicht annotiert wurden. Nasalierung wurde dort mit separaten Labeln als suprasegmentale phonetische Kennzeichnung annotiert.

Bei jedem Sprecher wurden durchgehend deutlich mehr Daten bei den Lauten [a], [m] und [n] extrahiert als bei [i], [u] und [ŋ]. Bei der Durchsicht einer kleinen Stichprobe der Transkriptionsdateien stellte ich fest, dass das orthografische »i« sehr oft als [ɪ] im Kiel-Korpus transkribiert ist, das orthografische »u« als [ʊ]. Eine ähnliche Erklärung für [ŋ] fand ich nicht. Dieser Laut kam scheinbar nicht so oft vor wie die anderen extrahierten Laute.

4. Statistische Auswertung

Die extrahierten Analysedaten sind in einer Text-Datei im ASCII-Format in tabellarischer Anordnung gespeichert. Sie kann vom Statistikprogramm R (R DEVELOPMENT CORE TEAM, 2004) geladen und als sog. Data-Frame, bestehend aus mehreren namentragenden Spaltenvektoren, interpretiert werden. Die Namensgebung der Vektoren ist durch die erste Zeile der Datei `formantresults` vorgegeben, die folgende Bezeichnungen enthält: Sprecher, Lautnummer, Sprachlaut, F1, F2, F3, F4, F5, F6, F7 und Datei. Über diese Bezeichnungen kann in R auf die Daten der entsprechenden Spalten zugegriffen werden. Das Statistikprogramm R enthält bereits implementierte statistische Tests, die durch einfache Funktionsaufrufe an den Daten durchgeführt werden können. Zusätzlich sind eigene Funktionen in R implementierbar und grafische Darstellungen der Daten erzeugbar.

Die entscheidende Frage für die vorliegende Untersuchung ist, ob sich die Verteilungen der ermittelten Formantfrequenzdaten abhängig vom Faktor Sprecher signifikant unterscheiden und dadurch Sprecherindividualität andeuten. Ich nutze den T-Test (ein statistischer Mittelwertsvergleich) und die multivariate Varianzanalyse (ANOVA: „analysis of variances“), um die Verteilungen unterschiedlich gruppierter Formantfrequenzdaten zu untersuchen. Die Eingaben in R sowie Perl-Skripte zur weiteren Verarbeitung der Auswertungsergebnisse und einige Dateien mit Zwischenergebnissen sind auf der beigelegten CD-ROM in der Datei `R.Befehle` gespeichert.

4.1. Filtern der Daten

Die extrahierten Formantdaten enthalten Werte, die im Spektrogramm als Sprünge sichtbar sind und nicht ausgewertet werden sollen. Sie entstanden möglicherweise durch eine fehlerhafte Schätzung und würden die Auswertung verfälschen. Um diese Werte von der Untersuchung auszuschließen, ist eine Vorverarbeitung der Daten nötig. Davon ausgehend, dass die auszu-schließenden Werte aufgrund der Sprünge im Formantverlauf zu den Extremwerten des jeweiligen Sprachlauts eines Sprechers gehören, benutze ich dazu eine von mir erstellte Funktion für R. Sie ermittelt bei jedem Sprecher bei jeweils jedem Formanten die Frequenzwerte, die unter dem 10%-Quantilwert oder über dem 90%-Quantilwert der Werte aller produzierten Einzellaute des jeweiligen Sprachlauts beim jeweiligen Sprechers liegen. Diese Werte werden durch den Eintrag »NA« ersetzt. Der Eintrag »NA« zeigt R einen fehlenden Wert an und ist empfehlenswerter als beispielsweise der Eintrag des Werts »0«, da er in weitere Berechnungen nicht versehentlich

mit einfließen kann. Bei dieser Vorbehandlung der Daten entfallen bei jedem Formanten ungefähr 20 % seiner Datenmenge. Die analysierbare Datenmenge ist kleiner, denn die folgenden statistischen Analysen beruhen auf der Betrachtung vollständiger Datensätze. Ein Datensatz besteht aus den sieben Formantwerten der Formanten F1–F7, die aus der Betrachtung eines einzelnen Analysefensters ermittelt wurden. Die ausgeschlossenen Werte eines Formanten kommen nicht zwingend im gleichen Datensatz vor wie die der anderen Formanten. Deshalb verringert sich die Datenmenge nochmals, wenn nur die Datensätze betrachtet werden, die keine »NA«-Einträge enthalten. Bei der Untersuchung der Formanten F4–F7 können nach der beschriebenen Vorverarbeitung ungefähr 48 % (50429 von 105610 Datensätzen) der extrahierten Datenmenge analysiert werden. Die entstehende, gefilterte Tabelle (mit allen Datensätzen, einschließlich der »NA«-Einträge) ist als Data-Frame im R-Format gespeichert.

4.2. Vergleich der Formantfrequenzverteilungen

Ich prüfte die Formantfrequenzdaten mit der multivariaten ANOVA-Analyse auf Unterschiede in den Verteilungen. Dabei ließ ich die Antwortmatrix, bestehend aus den vollständigen Datensätzen der Formanten F4–F7, analysieren. Als Abhängigkeiten für die Antwortmatrix gab ich bei ANOVA die Faktoren Sprecher und Sprachlaut an. Der Sprachlaut-Faktor hat den Zweck festzustellen, ob mögliche Unterschiede in den Verteilungen der Formantfrequenzwerte auch zwischen den verschiedenen Sprachlauten bestehen oder ob die möglichen Unterschiede zwischen Sprechern lautunabhängig sind.

Das Ergebnis ist bei allen Formanten bei den Faktoren Sprecher, Sprachlaut und der Kombination aus Sprecher und Sprachlaut ein p-Wert, der nicht genau angegeben wird, da er kleiner ist als $2,2 \cdot 10^{-16}$. Der p-Wert gibt die Wahrscheinlichkeit an, einen Unterschied in den Verteilungen geschätzt zu haben, obwohl tatsächlich keiner besteht. Je kleiner der p-Wert ist, desto signifikanter ist seine Aussage. Die üblichen Signifikanzniveaus an Hand derer entschieden wird, ob ein Ergebnis aussagekräftig genug ist bzw. verworfen werden sollte, werden bei den p-Werten 0,1, 0,05, 0,01 oder 0,001 festgelegt. Das Ergebnis der Analyse liegt im hochsignifikanten Bereich und lässt schließen, dass die Frequenzwerte der höheren Formanten F4–F7 der untersuchten Sprechergruppe unterschiedlich sind und abhängig davon, welcher Sprecher sprach und welcher Sprachlaut produziert wurde. Da die Faktoren Sprecher und Sprachlaut auch einzeln dieselben kleinen p-Werte liefern, schließe ich, dass zwischen den Sprachlauten ohne Sprecherberücksichtigung Unterschie-

Tabelle 1: Anzahl analysierbarer Datensätze der gefilterten Daten (Gesamtzahl der Ursprungsdatensätze: 105610).

Formanten	Filterebene	
	Sprachlaut	Einzellaut
F1–F3	58425 ($\approx 55\%$)	44839 ($\approx 42\%$)
F4–F7	50429 ($\approx 48\%$)	34579 ($\approx 33\%$)

de in den Verteilungen der Daten der höheren Formanten bestehen, genauso wie zwischen den Sprechern lautunabhängig.

Da die Unterschiede der Verteilungen dadurch begründet sein könnten, dass die Daten nicht korrekt bzw. streng genug gefiltert wurden, filterte ich die Ursprungsdaten nach einer anderen Methode. Diesmal ließ ich die beiden Quantilwerte (10% und 90%) jeweils zu den einzeln produzierten Lauten errechnen und bei jedem Einzellaut die Einträge ersetzen, die außerhalb seiner Quantilgrenzen liegen. Die Anzahl der analysierbaren Datensätze ist bei der Filterung auf Einzellautebene geringer als bei der Filterung auf Sprachlautebene: ungefähr 33% der Ursprungsdatenmenge (34579 von 105610 Datensätzen). Das Ergebnis der multivariaten ANOVA ist dasselbe: hochsignifikante p-Werte (kleiner als $2,2 \cdot 10^{-16}$) bei allen drei Faktoren.

Die Analyse der Verteilungen bei den ungefilterten Daten zeigt dasselbe Ergebnis: Auch hier ist der p-Wert bei allen Formanten für alle Faktoren kleiner als $2,2 \cdot 10^{-16}$.

Aus diesen drei Analyseergebnissen kann geschlossen werden, dass die höheren Formanten F4–F7 der betrachteten Daten sprecherabhängig sind. Dabei sind die sprecherindividuellen Unterschiede mit und ohne Sprachlautabhängigkeit feststellbar. Die Analyse ist robust gegen die einzelnen falsch ermittelten Formantfrequenzwerte.

Tiefe Formanten

Bei der Analyse der Formantfrequenzdaten von F1–F3 mit ANOVA ergeben sich dieselben Ergebnisse. Bei allen betrachteten Formanten ist der p-Wert für alle drei Faktoren kleiner als $2,2 \cdot 10^{-16}$, unabhängig davon, ob die Analyse auf Grundlage einer der gefilterten Datenmengen oder auf der nicht gefilterten Datenmenge durchgeführt wurde. Die Anzahl der jeweils analysierbaren Datensätze ist in Tabelle 1 aufgeführt.

Getrennte Lautgruppen

Bei der getrennten Untersuchung der beiden Lautgruppen Nasale und Vokale mit ANOVA werden nichtsignifikante p-Werte für den Sprachlaut-Faktor errechnet. In der Gruppe der Nasale bei F7 (0,7808 bei der ungefilterten Datenmenge; 0,09485 wenn gefiltert auf Einzellautebene; 0,06003 wenn gefiltert auf Sprachlautebene) und in der Gruppe der Vokale bei F5 (0,4618 bei der ungefilterten Datenmenge; 0,4838 wenn gefiltert auf Einzellautebene). Bei der auf Sprachlautebene gefilterten Datenmenge ergibt ANOVA in der Gruppe der Vokale bei F5 einen – von den anderen hochsignifikanten Ergebnissen abweichenden – p-Wert von $6,17 \cdot 10^{-5}$. In der Gruppe der Vokale wird bei F4 bei der auf Einzellautebene gefilterten Daten der p-Wert $1,707 \cdot 10^{-15}$ errechnet. Bei den Formanten F1–F3 sind in beiden Sprachlautgruppen Vokale und Nasale alle p-Werte hochsignifikant.

Histogramme

Zusätzlich zur rechnerischen Analyse können die Verteilungen der Formantfrequenzdaten in der grafischen Darstellung – Histogramme mit relativen Häufigkeitsangaben – untersucht werden. Abbildung 3 auf der nächsten Seite zeigt die Histogramme aller 16 Sprecher des Formanten F7 aus den Daten des Sprachlauts [a]. Zusätzlich ist oben links das Histogramm von F7 aus den Daten aller Sprecher des Sprachlauts [a] abgebildet. Dabei wurden die Histogramme auf der Grundlage der auf Sprachlautebene gefilterten Datenmenge erstellt. Aus den Abbildungen der Histogramme ist ersichtlich, weshalb die Verteilungen der Formantfrequenzdaten so unterschiedlich sind: Ihre jeweiligen Mittelwerte liegen entweder an unterschiedlichen Punkten auf der Frequenzskala oder ihre Streuungen sind verschieden groß. Außerdem haben manche Histogramme scheinbar zwei Häufungspunkte. In Abbildung 3 gibt es einzelne Verteilungen, die einander sehr ähnlich sind (z. B. die der Sprecher KAE und KAP, REK und NAR). Die Abbildung ist repräsentativ für die Unterschiede der Histogramme bei anderen Formanten und Sprachlauten und der lautunabhängigen Histogramme. Der Vergleich der Histogramme bestätigt das Ergebnis aus der multivariaten ANOVA-Analyse.

4.3. Prüfen der Sprecher auf Inkonsistenz

Als weiteren Aspekt prüfte ich die ungefilterte Datenmenge auf Inkonsistenz der Formantfrequenzdaten innerhalb der Sprecher. Dazu halbierte ich bei jedem Formanten die Datenmenge jeweils eines Sprachlauts eines Sprechers und analysierte die Verteilungen der beiden Datenhälften mit dem T-Test.

Formant F7 – Sprachlaut a

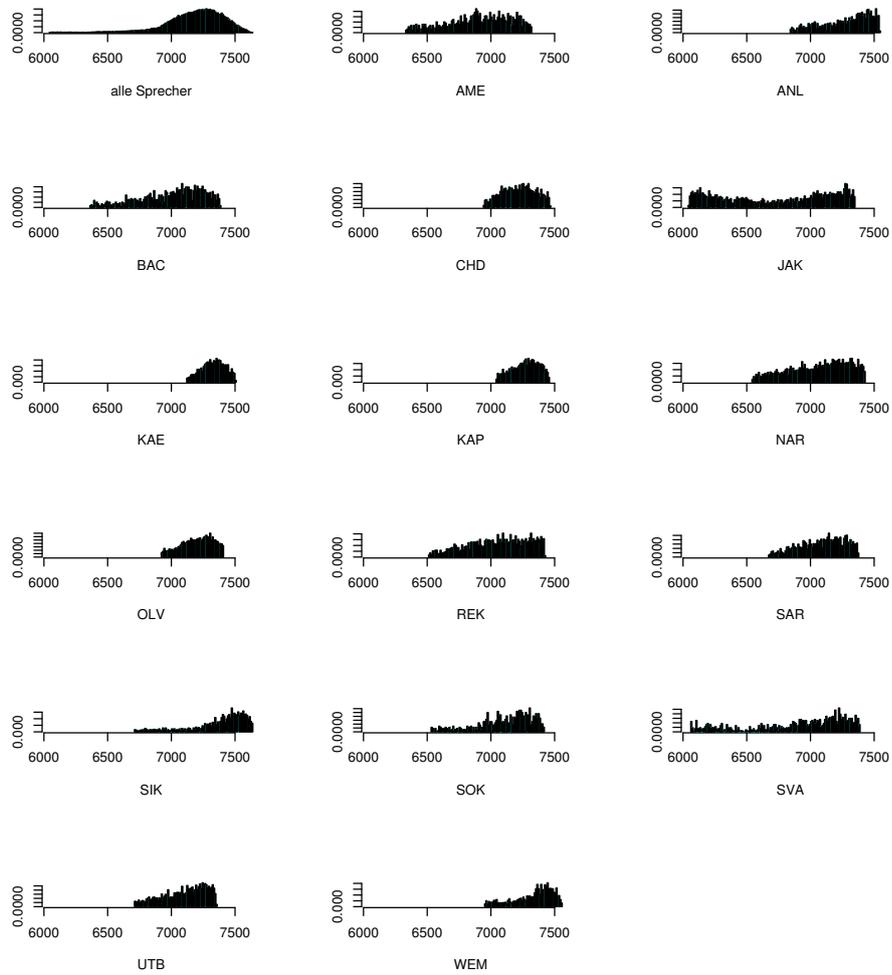


Abbildung 3: Histogramme relativer Häufigkeit der Formantfrequenzen von F7 bei Laut [a] (oben links: Gesamthistogramm aus den Frequenzdaten des Lauts [a] aller Sprecher).

Tabelle 2: Anzahl signifikanter p-Werte ($p < 0,01$) der T-Test-Analysen, kumuliert über die höheren Formanten F4–F7.

Analysemethode	Filterebene		ungefiltert	alle T-Tests
	Sprachlaut	Einzellaut		
lautunabhängig	29	33	36	64
lautabhängig	118	121	128	384

Ich führte den T-Test zusätzlich lautunabhängig durch, d. h. ich verglich jeweils die gesamten Datenhälften eines Formanten des jeweils untersuchten Sprechers. Die p-Werte der T-Tests ließ ich zusammen mit den dazugehörigen Daten Formant, Sprecher und gegebenenfalls Sprachlaut ausgeben. Da ich die T-Tests lautabhängig sowie lautunabhängig an jeweils allen drei Datenmengen (ungefilterte und zwei unterschiedlich gefilterte) durchführte, speicherte ich die Ergebnisse in sechs Dateien, um sie getrennt voneinander untersuchen zu können. Mit einem selbsterstellten Perl-Skript extrahierte ich aus diesen Dateien jeweils die Zeilen mit den p-Werten, die kleiner als 0,01 waren. Diese p-Werte sind signifikant für den Unterschied in den betrachteten Verteilungen und zeigen so die Inkonsistenz des Sprechers. Tabelle 2 zeigt die Anzahl der T-Tests mit signifikanten p-Werten, kumuliert über alle höheren Formanten F4–F7. Eine Tabelle der einzelnen signifikanten p-Werte für alle sieben Formanten ist in Anhang A.5 aufgeführt.

Tabelle 2 verdeutlicht, dass die Sprechergruppe bei allen drei Datenmengen in ungefähr 50 % der lautunabhängigen Analysen inkonsistent war, bei den lautabhängigen Analysen in ungefähr 30 % der Fälle. Dabei variieren die relativen Häufigkeiten der inkonsistenten Fälle einzelner Formanten bei der lautunabhängigen Analyse stärker und sind in fast allen Fällen höher als bei der lautabhängigen (s. Tabellen in Anhang A.5). Die Verringerung des Prozentsatzes an inkonsistenten Fällen bei lautabhängigen im Vergleich zu lautunabhängigen Analysen deutet an, dass auch die Werte höherer Formanten lautabhängig sind.

Ich ließ mir von den Fällen mit signifikanten p-Werten für jeden Formanten die Anzahl der gleichen Sprecher-Sprachlaut-Kombinationen von einem Perl-Skript ausgeben. Mir fielen beim Betrachten dieser Liste keine erwähnenswerten Regelmäßigkeiten auf, auf Grund derer bestimmte Formantdaten von der Untersuchung ausgeschlossen werden sollten – durchgehend inkonsistente Sprecherdaten wären wert gewesen, die Analyse ohne diese zu wiederholen. Bei Sprechern, bei denen die lautunabhängige Analyse bei al-

len höheren Formanten F4–F7 inkonsistente Verteilungen der Frequenzwerte zeigt, sind die Verteilungen bei der lautabhängigen Analyse weder bei allen Sprachlauten noch bei allen Formanten inkonsistent. Bei keinen bestimmten Sprachlauten, Sprechern oder Sprecher-Sprachlaut-Kombinationen ist eine herausragende Anzahl von Vorkommen zu finden, die auf Inkonsistenz in besonderer Abhängigkeit von einem der drei Faktoren schließen ließe. Alle Sprachlaute und alle Sprecher kommen in der Liste vor, d. h. bei keinem Sprachlaut oder Sprecher ist eine herausragende Konsistenz feststellbar.

5. Zusammenfassung und Schlussbemerkung

In der vorliegenden Arbeit untersuchte ich die Formantfrequenzdaten der Sprachlaute [a], [i], [u], [m], [n] und [ŋ], die ich auf der Grundlage des „Kiel-Korpus der spontanen gesprochenen Sprache“ ermittelt hatte, um herauszufinden, ob sprecherindividuelle Unterschiede feststellbar sind. Besonders interessierten mich dabei die Frequenzwerte der höheren Formanten F4–F7, da von ihnen behauptet wird, sie seien sprecherspezifisch. Ich betrachtete im Vergleich die Formanten F1–F3, denen eine Korrelation mit Artikulationswerkzeug-Konstellationen nachgewiesen wurde.

Bei der Untersuchung der Verteilungen mit ANOVA waren die p-Werte aller höheren Formanten für alle drei Faktoren Sprecher, Sprachlaut und der Kombination aus Sprecher und Sprachlaut hochsignifikant unterschiedlich. Dieses Ergebnis war unabhängig davon, ob die Auswertungsdaten auf Sprachlaut- bzw. Einzellautebene gefiltert waren oder ob die Ursprungsdatenmenge analysiert wurde. Die Analyse der Formanten F1–F3 zeigte dasselbe Ergebnis. Daraus folgt, dass die Formantfrequenzdaten aller untersuchten Formanten F1–F7 sprecherindividuell verteilt und nicht ausschließlich auf die höheren Formanten beschränkt sind. Dieses Ergebnis kann robust mit ANOVA aus den Daten ermittelt werden. Der Vergleich der Histogramme (die für jeweils einen Sprachlaut eines Sprechers erstellt wurden) aus den auf Sprachlautebene gefilterten Daten bestätigte das Ergebnis. In den Histogrammen war erkennbar, dass sich nicht alle Verteilungen völlig voneinander unterschieden. Bei der getrennten Analyse der Gruppen Nasale und Vokale lieferte ANOVA einzelne nichtsignifikante p-Werte, die nahe an der Signifikanzgrenze lagen: In der Gruppe der Nasale war bei Formant F7 und in der Gruppe der Vokale bei Formant F5 kein Unterschied in den Formantfrequenzverteilungen der einzelnen Sprachlaute festzustellen. Dieses Ergebnis muss so interpretiert werden, dass die Daten in diesen Fällen nicht ausreichend waren, um einen signifikanten Unterschied in den Verteilungen feststellen zu können. Insgesamt deuten die Ergebnisse zur Untersuchung der Formantfrequenzverteilungen darauf, dass die untersuchten tiefen und höheren Formanten in sprecherindividuellen Frequenzbereichen verlaufen und abhängig von der Lautqualität sind.

Mit einzelnen T-Tests prüfte ich bei jedem Sprecher die Inkonsistenz der Formantfrequenzdaten bei jeweils allen Sprachlauten. Die Verteilungen waren bei allen Formanten in ungefähr 30 % der Sprecher-Sprachlaut-Kombinationen signifikant voneinander verschieden. Bei der lautunabhängigen Analyse war bei allen Formanten in ungefähr 50 % der Fälle eine signifikante Inkonsistenz feststellbar. Die niedrigere Inkonsistenzrate bei der lautabhän-

gigen Analyse lässt den Schluss zu, dass alle Formanten lautabhängig sind und bestätigt darin die Untersuchung der Formantdatenverteilungen.

Die Untersuchung zur Sprecherindividualität höherer Formanten ist auf der Grundlage des gegebenen Korpus erweiterbar. Eine genauere Datenunterscheidung ist möglich durch die Trennung der Sprachlaute nach ihren Kontexten, um Einflüsse der Koartikulation auszuschließen. Im Korpus gibt es außerdem noch weitere, phonetische Annotationen, die ebenfalls dazu benutzt werden können, Sprachlaute genauer in Untergruppen zu unterteilen, z. B. Annotationen zur Stimmqualität. Bei den Nasalen können zusätzlich die Antiformanten in die Untersuchung mit einbezogen werden (die ich mit den ESPS-Werkzeugen nicht berechnen konnte, da ich keine entsprechende Funktion dafür fand). Da die Formantfrequenzen in dieser Untersuchung an Hand der LPC-Analyse ermittelt wurden, die nicht verlässlich ist (vgl. [CLARK & YALLOP, 1995](#), S. 261), sollte die statistische Auswertung bei Formantfrequenzdaten wiederholt werden, die auf eine andere Art ermittelt werden, um die Ergebnisse zu vergleichen. Frauenstimmen sollten getrennt untersucht werden, da dort der erste Formant leicht mit der Grundfrequenz oder einer Harmonischen verwechselt wird.

Die Ergebnisse der vorgestellten Auswertung haben sprecherindividuelle Unterschiede bei der Verteilung der Formantfrequenzen aller Formanten gezeigt. Es bleibt die Frage, ob dieses Ergebnis mit ausgefeilteren Auswertungsmethoden und anderen, möglicherweise umfangreicheren Daten in seiner Grundaussage verändert werden kann. Größere Frequenzdatenmengen werden entweder die Verteilungsformen besser herauskristallisieren oder so verändern, dass stochastisch produzierte Daten angenommen werden müssen.

A. Tabellen und Abbildungen

A.1. Inhalt der beigelegten CD-ROM

Tabelle 3: Inhalt der beigelegten CD-ROM und Beschreibung der einzelnen Dateien, Skripte und Verzeichnisse.

Datei/Skript/Verzeichnis	Beschreibung
kiel-cd/	Verzeichnis, das die Sprachsignal- und Transkriptionsdateien des Kiel-Korpus, gruppiert in Unterverzeichnissen, enthält
*.l16, *.r16	Sprachsignaldateien des Kiel-Korpus (Endungen: l16, r16)
*.s1h	Transkriptionsdateien des Kiel-Korpus (Endung: s1h)
btosps	ESPS-Programm zur Umwandlung der Sprachsignaldateien in Dateien mit ESPS-lesbarem Format
s1h2phones	GAWK-Skript zur Umwandlung der Transkriptionsdateien in Segmentlabel-Dateien mit für waves lesbarem Format
formant	ESPS-Programm zur Ermittlung von Formantfrequenzwerten aus einer Signaldatei
pplain	ESPS-Programm zur Anzeige der Daten aus einer mit formant erzeugten Datei
formanting.perl	Perl-Skript, das aus den Daten des Kiel-Korpus die nötigen Daten für die vorliegende Untersuchung erstellt
formantresults	Text-Datei im ASCII-Format, in der alle zu analysierenden Daten für die Untersuchung in geeigneter tabellarischer Form gespeichert sind
ERRORS	Text-Datei im ASCII-Format mit allen Fehlermeldungen, die während der Ausführung von formanting.perl entstanden sind
quant9010ges.R	Binär-Datei in komprimiertem R-Format, die die auf Sprachlautebene gefilterte Datentabelle der Datei formantresults enthält
quant9010einz.R	Binär-Datei in komprimiertem R-Format, die die auf Einzellautebene gefilterte Datentabelle der Datei formantresults enthält
R.Befehle	Kommentierte Befehle, die innerhalb des Statistikprogramms R benutzt wurden
Perl.Skripte	Text-Datei im ASCII-Format, die alle restlichen, kommentierten Perl-Skripte enthält, die für die vorliegende Untersuchung benutzt wurden

(Tabelle 3 wird fortgesetzt auf der nächsten Seite)

(Fortgesetzte Tabelle 3)

Datei/Verzeichnis	Beschreibung
inkonsistenz.*.R	Text-Dateien im ASCII-Format, die die p-Werte der T-Test-Analysen zur Sprecher-Inkonsistenz enthalten – im Dateinamen ist festgehalten, wenn die Analyse auf gefilterten Daten (auf Sprachlautebene: ges, auf Einzellautebene: einz) und wenn sie lautunabhängig durchgeführt wurde: lautunabh
studa.pdf, studa.ps	Schriftliche Ausarbeitung der Studienarbeit in den angegebenen Formaten
cd-inhalt.txt	Text-Datei im ASCII-Format, die den Inhalt der beigelegten CD-ROM beschreibt (entspricht Tabelle 3)

A.2. IPA-SAMPA-Umschrift

Tabelle 4: Die IPA-SAMPA-Tabelle der deutschen Sprachlaute, die bei der Annotation des Kiel-Korpus benutzt wurden.

	IPA	SAMPA		IPA	SAMPA
Vokale	a	a	Konsonanten	b	b
	e	e		d	d
	ɛ	E		f	f
	i	i		g	g
	ɪ	I		h	h
	o	o		j	j
	ɔ	O		k	k
	u	u		l	l
	ū	U		m	m
	y	y		n	n
	ʏ	Y		p	p
	ø	2		r	R
	œ	9		s	s
	ã	a~		t	t
	ẽ	E~		v	v
	õ	O~		z	z
	œ̃	9~		ç	C
	aɪ	aI		x	x
	aʊ	aU		ʃ	S
	ɔʏ	OY		ʒ	Z
	ə	@		ŋ	N
	ɐ	6		ʔ	Q
Längung	:	:			

A.3. Vokalviereck

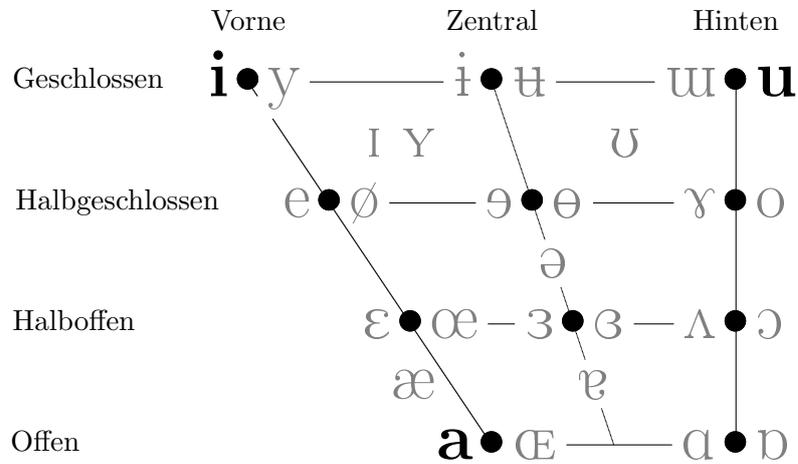


Abbildung 4: Vokalviereck (vgl. [INTERNATIONAL PHONETIC ASSOCIATION, 1999](#)). Jeder Vokal wird jeweils durch die entsprechende Kiefer-Zunge-Lippen-Konstellation produziert, wie sie hier angegeben ist. Die Horizontale bezeichnet die Lage der Zunge, die Vertikale den Öffnungsgrad des Kiefers (der mit der vertikalen Lage der Zunge korreliert). Bei Vokalpaaren wird der Vokal auf der rechten Seite mit gerundeter Lippenstellung produziert, der auf der linken mit ungerundeter. Die hervorgehobenen Vokale wurden für die vorliegende Untersuchung benutzt.

A.4. Datensatzmenge

Tabelle 5: Anzahl der analysierbaren Datensätze der Formanten F4–F7 (detaillierte Erklärung in Kapitel 3).

Sprecher	Sprachlaut	Ursprungsdaten		gefiltert auf Sprachlautebene	
		Vorkommen	Datensätze	Vorkommen	Datensätze
AME	a	129	1112	109	542
	i	34	190	29	101
	u	19	108	18	55
	m	64	686	56	320
	n	129	889	112	452
	ŋ	9	73	8	34
	gesamt	384	3058	332	1504
ANL	a	223	1654	184	745
	i	43	270	34	118
	u	18	131	18	66
	m	133	1170	118	520
	n	232	1612	182	690
	ŋ	32	238	30	88
	gesamt	681	5075	566	2227
BAC	a	180	1462	148	650
	i	39	252	28	118
	u	17	98	14	52
	m	125	1000	112	503
	n	230	1295	193	629
	ŋ	34	320	30	144
	gesamt	625	4427	525	2096
CHD	a	196	1911	183	852
	i	45	302	39	133
	u	10	80	9	35
	m	96	966	94	474
	n	207	1760	186	825
	ŋ	41	351	36	172
	gesamt	595	5370	547	2491
JAK	a	387	3153	321	1512
	i	72	439	58	227
	u	28	194	22	89
	m	196	1631	167	827
	n	489	2870	416	1466

(Tabelle 5 wird fortgesetzt auf der nächsten Seite)

(Fortgesetzte Tabelle 5)

Sprecher	Sprachlaut	Ursprungsdaten		gefiltert auf Sprachlautebene	
		Vorkommen	Datensätze	Vorkommen	Datensätze
KAE	ɨ	64	318	47	164
	gesamt	1236	8605	1031	4285
	a	318	2741	269	1215
	i	51	382	43	167
	u	43	391	41	189
	m	145	946	127	508
	n	342	2435	302	1265
	ɨ	53	452	49	239
	gesamt	952	7347	831	3583
KAP	a	304	2709	251	1212
	i	65	412	48	189
	u	29	236	25	104
	m	165	1951	144	920
	n	353	2663	295	1152
	ɨ	40	379	37	177
	gesamt	956	8350	800	3754
NAR	a	383	3362	343	1648
	i	92	640	83	344
	u	35	286	34	125
	m	192	1458	178	808
	n	510	3806	468	1984
	ɨ	60	394	52	192
	gesamt	1272	9946	1158	5101
OLV	a	527	4206	443	1847
	i	186	1117	146	502
	u	57	462	49	213
	m	270	1980	239	1062
	n	701	4096	604	1942
	ɨ	72	436	63	237
	gesamt	1813	12297	1544	5803
REK	a	301	2125	249	993
	i	79	502	64	230
	u	39	307	31	137
	m	163	1441	140	683
	n	450	2583	393	1267
	ɨ	28	177	24	84
	gesamt	1060	7135	901	3394

(Tabelle 5 wird fortgesetzt auf der nächsten Seite)

(Fortgesetzte Tabelle 5)

Sprecher	Sprachlaut	Ursprungsdaten		gefiltert auf Sprachlautebene	
		Vorkommen	Datensätze	Vorkommen	Datensätze
SAR	a	284	2177	244	982
	i	35	225	29	105
	u	19	121	15	57
	m	132	1013	115	486
	n	240	1577	197	731
	ŋ	15	97	13	42
	gesamt		725	5210	613
SIK	a	201	1892	174	864
	i	39	286	33	123
	u	13	171	13	80
	m	136	1168	115	534
	n	332	2520	280	1182
	ŋ	30	249	26	115
	gesamt		751	6286	641
SOK	a	116	851	97	410
	i	29	210	23	96
	u	13	128	13	56
	m	69	396	58	198
	n	157	959	136	509
	ŋ	21	194	17	104
	gesamt		405	2738	344
SVA	a	143	1088	126	490
	i	30	201	24	100
	u	18	126	17	62
	m	81	671	73	321
	n	225	1588	207	785
	ŋ	19	131	16	54
	gesamt		516	3805	463
UTB	a	325	2696	290	1402
	i	116	795	93	377
	u	36	266	31	123
	m	250	1940	224	920
	n	516	3508	438	1715
	ŋ	60	441	50	226
	gesamt		1303	9646	1126
WEM	a	206	1829	176	856
	i	50	388	42	183

(Tabelle 5 wird fortgesetzt auf der nächsten Seite)

(Fortgesetzte Tabelle 5)

Sprecher	Sprachlaut	Ursprungsdaten		gefiltert auf Sprachlautebene	
		Vorkommen	Datensätze	Vorkommen	Datensätze
	u	18	183	16	84
	m	148	1332	131	641
	n	285	2201	259	1010
	ŋ	40	382	37	168
	gesamt	747	6315	661	2942
alle Sprecher	a	4223	34968	3607	16220
	i	1005	6611	816	3113
	u	412	3288	366	1527
	m	2365	19749	2091	9725
	n	5398	36362	4668	17604
	ŋ	618	4632	535	2240
	gesamt	14021	105610	12083	50429

A.5. T-Tests zur Inkonsistenz der Sprecher

Tabelle 6: Lautunabhängige Analyse der Sprecher-Inkonsistenz: Anzahl der T-Tests mit signifikanten p-Werten ($p < 0,01$); rechte Spalte: Anzahl aller durchgeführten T-Tests.

Formant	Filterebene		ungefiltert	alle T-Tests
	Sprachlaut	Einzellaut		
F1	9	6	4	16
F2	10	7	8	16
F3	9	6	6	16
F4	6	8	11	16
F5	9	10	10	16
F6	7	10	10	16
F7	6	5	5	16
gesamt	57	52	54	112

Tabelle 7: Lautabhängige Analyse der Sprecher-Inkonsistenz: Anzahl der T-Tests mit signifikanten p-Werten ($p < 0,01$); rechte Spalte: Anzahl aller durchgeführten T-Tests.

Formant	Filterebene		ungefiltert	alle T-Tests
	Sprachlaut	Einzellaut		
F1	34	29	29	96
F2	42	32	33	96
F3	37	37	31	96
F4	26	23	27	96
F5	29	36	32	96
F6	23	32	39	96
F7	37	30	31	96
gesamt	225	215	221	672

Literatur

- [CLARK & YALLOP 1995] CLARK, J. ; YALLOP, C.: *An Introduction to Phonetics and Phonology*. 2. Aufl. Oxford : Blackwell, 1995 (Zitiert auf den Seiten 1 und 27.)
- [ENTROPIC RESEARCH LABORATORY 1998] ENTROPIC RESEARCH LABORATORY: *ESPS/waves+: Manuals of Product Release 5.3*. Washington, DC: Entropic Research Laboratory (Veranst.), 1998. – Online-Dokumentation: URL <http://ciips.ee.uwa.edu.au/~roberto/research/speech/local/entropic/ESPSDoc/> (Zugriffsdatum: Juli 2004) (Zitiert auf den Seiten 7, 8, 10 und 14.)
- [FREE SOFTWARE FOUNDATION 2003] FREE SOFTWARE FOUNDATION: *GAWK – String Manipulation Language*. URL <http://www.gnu.org/software/gawk/gawk.html>. November 2003 (Zitiert auf Seite 8.)
- [INTERNATIONAL PHONETIC ASSOCIATION 1999] INTERNATIONAL PHONETIC ASSOCIATION (Hrsg.): *Handbook of the International Phonetic Association*. Cambridge, UK : Cambridge University Press, 1999 (Zitiert auf Seite 31.)
- [KOHLETER ET AL. 1995] KOHLER, K. ; PÄTZOLD, M. ; SIMPSON, A.: From Scenario to Segment. The Controlled Elicitation, Transcription, Segmentation and Labelling of Spontaneous Speech / Institut für Phonetik und digitale Sprachverarbeitung Universität Kiel. K. J. Kohler, 1995 (Arbeitsberichte des Instituts für Phonetik der Universität Kiel: AIPUK 29). – Forschungsbericht (Zitiert auf Seite 6.)
- [MACHELETT 1994] MACHELETT, K.: *Das Lesen von Sonagrammen in der Phonetik*. München, Institut für Phonetik und sprachliche Kommunikation, Magisterarbeit, 1994 (Zitiert auf Seite 37.)
- [MACHELETT & TILLMANN 1996] MACHELETT, K. ; TILLMANN, H.G.: *Das Lesen von Sonagrammen*. URL <http://www.phonetik.uni-muenchen.de/SGL/SGLHome.html> (Zugriffsdatum: Juli 2004). 1996. – basiert auf MACHELETT (1994) (Zitiert auf Seite 3.)
- [MALMBERG 1963] MALMBERG, B.: *Phonetics*. Dover Publications, 1963 (Zitiert auf Seite 3.)
- [NOLAN 1997] NOLAN, F.: Speaker Recognition and Forensic Phonetics. In: HARDCASTLE, William J. (Hrsg.) ; LAVER, J. (Hrsg.): *The Handbook*

of *Phonetic Sciences*. Oxford : Blackwell, 1997 (Blackwell Handbooks in Linguistics), Kap. 25, S. 744–767 (Zitiert auf Seite 7.)

[PERL FOUNDATION 2004] PERL FOUNDATION: *The Perl Programming Language*. URL <http://www.perl.org/>. Juli 2004 (Zitiert auf den Seiten 4 und 8.)

[POTTER ET AL. 1947] POTTER, R.K. ; KOPP, G.A. ; GREEN, H.C.: *Visible Speech*. New York : D. Van Nostrand, 1947 (Bell telephone laboratories series) (Zitiert auf Seite 3.)

[R DEVELOPMENT CORE TEAM 2004] R DEVELOPMENT CORE TEAM: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (Veranst.), Juni 2004. – URL <http://www.R-project.org> (Zitiert auf Seite 19.)

[STEVENS 1999] STEVENS, K.: Articulatory-Acoustic-Auditory Relationships. In: HARDCASTLE, William J. (Hrsg.) ; LAVER, J. (Hrsg.): *The Handbook of Phonetic Sciences*. Oxford : Blackwell, 1999, Kap. 15, S. 462–507 (Zitiert auf den Seiten 2 und 3.)