

Vergleich muttersprachlicher und nicht-muttersprachlicher Vokale
mit Deutsch als Zielsprache:
Ein statistischer Vergleich in Corpora

von

Anja Hofmann

Philosophische Dissertation
angenommen von der Neuphilologischen Fakultät
der Universität Tübingen

am 11.1.2011

Tübingen

2011

Gedruckt mit Genehmigung der Neuphilologischen Fakultät der Universität Tübingen

Hauptberichterstatter: Prof. Hubert Truckenbrodt, Ph.D.

Mitberichterstatter: PD Dr. Ingo Hertrich

Dekan Prof. Dr. Jürgen Leonhardt

Danksagung

Hiermit möchte ich mich bei allen bedanken, die mich bei der wissenschaftlichen Arbeit und beim Schreiben meiner Dissertation unterstützt haben.

Prof. Hubert Truckenbrodt danke ich für die ausführliche und freundliche Betreuung meiner Dissertation, sowohl persönlich, per Telefon und online.

Ohne seine Bereitschaft, meine Arbeit zu betreuen und sich auf neue Programmiersprachen und statistische Verfahren einzulassen, würde diese Dissertation nicht existieren.

Prof. Ingo Hertrich hat sich als Zweitgutachter Zeit genommen, um mit mir statistische Verfahren und die Vorgehensweise im allgemeinen zu besprechen.

Prof. Ulrike Gut an der Albert-Ludwigs-Universität in Freiburg stellte mir ihr gesamtes Corpus unentgeltlich zur Verfügung und damit die Grundlage für meine phonetischen Untersuchungen.

Prof. Georg Carle, bei dem ich zuerst am Institut für Informatik angestellt war, danke ich für seinen Rat in technischen und organisatorischen Dingen und das sehr angenehme Arbeitsklima.

Peter Rempis und die gesamte UB Tübingen ermöglichten mir mit dem virtheo-Projekt, praktische Erfahrungen zu sammeln.

Meine Schwester Manuela Reiser geb. Hofmann brachte mich zuerst dazu, in Tübingen zu promovieren und übernahm zuletzt das Korrekturlesen meiner Arbeit.

Meinen Eltern, Kurt und Ruth Hofmann, bin ich für ihr Vertrauen in meine Fähigkeiten, ihre Ermutigung und ihre Geduld sehr dankbar.

Zusammenfassung

In dieser Arbeit wird verglichen inwieweit sich deutsche muttersprachliche Vokale und die Aussprache dieser Vokale durch Deutsch lernende Studenten unterscheiden. Dabei wird auf phonetische Messdaten wie die Formantwerte und der Vokaldauer zurückgegriffen.

Zuerst beschreibe ich die Ergebnisse neurolinguistischer und psycholinguistischer Studien zur Sprachverarbeitung im Gehirn und zum Zweitspracherwerb bei Kindern, Jugendlichen und Erwachsenen.

Da die Messungen auf gelesenen, vollständigen Texten durchgeführt werden und die Streuung der Daten für Zweitsprachenlerner größer ausfällt als für Muttersprachler, werden als nächstes die Vorbedingungen für den Einsatz parametrischer Verfahren erklärt wie z.B. des t-Tests oder der Varianzanalyse. Zusätzlich werden statistische Methoden beschrieben, die geringere Anforderungen an die Verteilung der Daten stellen und für den Einsatz auf Corpora entwickelt wurden.

Dann werden deutsche muttersprachliche Aufnahmen untersucht, deren Messergebnisse als Vorlage für den Vergleich mit nicht-muttersprachlichen Aufnahmen dienen. Die analysierten Unterschiede beim Lesen von Kurzgeschichten und Dialogen dienen als Begründung für die Materialauswahl, d.h. die Beschränkung auf gelesene Texte im Gegensatz zu Nacherzählungen, freien Gesprächen oder Wortlisten.

Als nächstes führe ich Untersuchungen zur Aussprache britischer, italienischer und chinesischer Muttersprachler durch. Die Auswahl der Sprachgruppen soll verschiedene Sprachfamilien mit einbeziehen, außerdem achte ich auf ausreichend große und ausgewogene Gruppen an Versuchspersonen.

Zuerst werden, soweit sinnvoll, in beiden Sprachen vorhandene Vokale verglichen. Danach beschäftige ich mich mit dem Gegensatz zwischen gespannten und ungespannten Vokalen, der im Italienischen und im Chinesischen nicht vorhanden ist, und als letztes mit unbekanntem Lauten, wobei sich vor allem Umlaute anbieten.

Im Chinesischen werden zusätzlich die Auswirkungen eines im Deutschen nicht vorhandenen Vokals besprochen.

Summary

This thesis uses phonetic measurements such as formant values or vowel duration to determine the extent to which native German speakers' vowel pronunciation differs from the pronunciation of students learning German.

First of all, the results of neurolinguistic and psycholinguistic studies concerning mental processing of speech and second language acquisition in children, teenagers and adults are described.

As my studies are based on complete texts being read aloud and as second language learners' data exhibits a greater dispersion than native speakers' measurements, the necessary requirements to conduct parametric tests are introduced, e.g. the t-test and the analysis of variance. Additionally, statistic methods with lower prerequisites are presented that were developed for use in corpora.

Then native German recordings are analyzed. The results serve as a comparison to values gained from non-native data.

Furthermore, recordings of short stories and transcribed dialogues are compared. The outcomes justify my choice of audio material in which I limit my studies to read texts and exclude retellings, free conversation and word lists.

Subsequently, the pronunciation of native speakers of British English, Italian and Mandarin (Standard Chinese) is examined. The languages have been selected to include a range of language families. Moreover, it was important to choose sufficiently large and balanced groups of speakers.

Firstly, vowels that exist in German and the test persons' native language are compared. Then the contrast between tense and lax vowels which does not exist in Italian or Chinese is examined. The final tests feature vowels that do not exist in the test groups' native languages especially umlaut vowels.

Concerning Chinese the effects of a native vowel that is not part of the German vowel inventory are presented.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Ausgangspunkt	1
1.1.1	Begründung der Fragestellung	1
1.1.2	These	2
1.2	Unterschiede zwischen Erstsprache und Zweitsprache	2
1.2.1	Verarbeitung von Sprache im Gehirn	2
1.2.2	Erwerb der Erstsprache	9
1.2.3	Rolle des Alters beim Erwerb der Zweitsprache	10
1.2.4	Schlussfolgerung	16
1.3	Zweitspracherwerb	17
1.3.1	Transfer und interlanguage	17
1.3.2	Perzeption	19
1.3.3	Artikulation	21
1.4	Elektronischer Vergleich von Sprachsignalen	23
1.4.1	Untersuchungsgegenstand	23
1.4.2	Einzelne Vokale	24
1.4.3	Formanten als wahrgenommene phonetische Eigenschaft	24
1.4.4	Vokale im Kontext einer Silbe oder intonatorischen Phrase	25
2	Vorgehensweise	27
2.1	Verwendetes Datenmaterial	27
2.1.1	Kiel Corpus of Read Speech	27
2.1.2	Learning Prosody in a Foreign Language (LeaP)	27
2.1.3	Eigene Annotationen	28
2.2	Abfolge der Hypothesen	29
2.3	Verteilungseigenschaften	30
2.3.1	Erstellung der Testmengen	30
2.3.2	Normalverteilung	30
2.3.3	Schiefe und Wölbung	31
2.3.4	Varianzhomogenität	33
2.3.5	Vorbedingungen für den t-Test	33
2.3.6	Vorbedingungen für die Varianzanalyse	34
2.3.7	Vorbedingungen der Kovarianzanalyse	36
2.4	Nichtparametrische Tests	38

2.4.1	Wilcoxon-Rangsummentest	38
2.4.2	H-Test nach Kruskal-Wallis	38
2.5	Darstellung der Unterschiede	38
2.5.1	Dunnett-Tukey-Kramer-Verfahren	39
2.5.2	Regularized Discriminant Analysis	39
2.5.3	Klassifikationsbäume	40
2.6	Ausschluss von Zufallsfaktoren	40
2.6.1	Gemischte Modelle	40
2.6.2	Erstellung eines linearen Modells	41
2.6.3	Bewertung der Modelle	42
2.7	Suche nach zusammenhängenden Daten	43
2.7.1	Modellbasiertes Clustering	43
2.8	Zusammenfassung	45
3	Statistische Eigenschaften der muttersprachlichen Aussprache	46
3.1	Einführung	46
3.2	Das deutsche Lautsystem	46
3.2.1	Vorgeschichte	46
3.2.2	Betonte und unbetonte Silben	47
3.2.3	Geschichte der Umlaute	49
3.2.4	Entwicklung von Varietäten	50
3.2.5	Die deutsche Standardaussprache heute	52
3.2.6	Vokaldauer der untersuchten Vokale	57
3.3	Kurzgeschichten und gelesene Dialoge	59
3.3.1	Problem	59
3.3.2	/i:/	60
3.3.3	/o:/	64
3.3.4	Zusammenfassung	65
4	Deutsche und britische Sprecher	66
4.1	Einleitung	66
4.1.1	Sprachgeschichte	66
4.1.2	Vokalraum	67
4.1.3	Vergleich der Durchschnittswerte	69
4.1.4	Vokalräume einzelner Sprecher	77
4.2	Gemeinsame Vokale	80
4.2.1	Verteilungsfreie Tests auf allen Testmengen	81
4.2.2	Testmengen für /ɪ/, männliche Sprecher	81
4.2.3	Testmengen für /ɪ/, weibliche Sprecher	82
4.2.4	Testmengen für /ʊ/, männliche Sprecher	86
4.2.5	Testmengen für /ʊ/, weibliche Sprecher	89

4.2.6	Zusammenfassung	90
4.3	Gespannte und ungespannte Vokale	92
4.3.1	Unterscheidung zwischen /e:/ und /ɛ/	92
4.3.2	Vergleich der Aussprache von /e:/ und /ɛ/ bei deutschen und britischen Sprechern	96
4.4	Unbekannte Vokale	98
4.4.1	Untersuchungsgegenstand	98
4.4.2	Aussprache von /ø:/	98
4.4.3	Aussprache von /y:/	101
4.4.4	Schlußfolgerung	102
5	Deutsche und italienische Sprecher	103
5.1	Einführung	103
5.1.1	Sprachgeschichte	103
5.1.2	Der italienische Vokalraum heute	105
5.1.3	Direkter Vergleich der Vokalräume	106
5.1.4	Vokaldauer	112
5.2	Gespannte und ungespannte Vokale	116
5.2.1	Unterscheidung zwischen /i:/ und /ɪ/	116
5.2.2	Vergleich von /ɪ/	120
5.3	Unbekannte Vokale	125
5.3.1	Der Vokal /ø:/	125
5.3.2	Statistische Tests	125
5.3.3	Diskriminanzanalyse	125
5.3.4	Vergleich nach Sprachkenntnissen	127
5.4	Zusammenfassung	130
6	Deutsche und chinesische Sprecher	132
6.1	Einführung	132
6.1.1	Die chinesischen Vokale	132
6.1.2	Betonte Silben	135
6.1.3	Durchschnittswerte für deutsche und chinesische Vokale	137
6.1.4	Eigene Messungen in chinesischen Aufnahmen	141
6.1.5	Physiologischer Vergleich	142
6.2	Direkter Vergleich der Werte aus dem LeaP-Corpus	144
6.2.1	Männliche Sprecher	144
6.2.2	Weibliche Sprecher	148
6.3	Vergleich der Vokaldauer	151
6.3.1	Männliche Sprecher	151
6.3.2	Weibliche Sprecher	152
6.4	Einfluss muttersprachlicher Vokale	154

6.4.1	/a/ und /ɤ/	154
6.5	Gespannte und ungespannte Vokale	161
6.5.1	Unterschied zwischen /i:/ und /ɪ/	161
6.5.2	Vergleich der Werte für /i:/, männliche Sprecher	165
6.5.3	Vergleich der Werte für /i:/, weibliche Sprecher	167
6.5.4	Vergleich der Werte für /ɪ/, männliche Sprecher	168
6.5.5	Vergleich der Werte für /ɪ/, weibliche Sprecher	169
6.5.6	Unterschied zwischen /u:/ und /ʊ/	170
6.5.7	Testmengen für /ɛ/ , weibliche Sprecher	173
6.6	Im Chinesischen unbekannte Vokale	175
6.6.1	Umsetzung von /ø:/ durch chinesische Sprecher	175
6.7	Schlussfolgerungen	178
7	Diskussion	179
7.1	Sprachspezifische Resultate	179
7.1.1	Deutsche Muttersprachler	179
7.1.2	Britische Sprecher	179
7.1.3	Italienische Sprecher	180
7.1.4	Chinesische Sprecher	180
7.2	Bewertung der Ausspracheunterschiede	181
7.3	Vergleich weiblicher Sprachaufnahmen	181
7.4	Parametrische statistische Verfahren und Data Mining	182
7.4.1	Gemischte lineare und generalisierte lineare Modelle	182
7.4.2	Bayesianische Statistik	182
7.4.3	Klassifikationsbäume und Clustering	182
7.5	Ausblick	183
A	Anhang	184
	Literatur	184

Tabellenverzeichnis

1.1	Verhältnis zwischen Wahrnehmung und Produktion in Erst- und Zweitsprache	18
1.2	Untersuchte Studentengruppen	20
1.3	Bewertung der gesprochenen Vokale	22
2.1	Übersicht über die Anwendung von Transformationsfunktionen	32
2.2	Auswahl des Tests für die Varianzhomogenität	32
3.1	Übersicht zu den oralen Monophthongen der deutschen Sprache	52
3.2	Durchschnittliche Vokaldauer für betonte und unbetonte Vokale	59
3.3	Vergleich zwischen /i:/ in gelesenen Kurzgeschichten und Dialogen für männliche Sprecher	61
3.4	Durchschnittliche Messwerte für /i:/ bei männlichen deutschen Sprechern in Dialogen und Kurzgeschichten	62
3.5	Vergleich zwischen /i:/ in gelesenen Kurzgeschichten und Dialogen für weibliche Sprecher	63
3.6	Durchschnittliche Messwerte für /i:/ bei weiblichen deutschen Sprechern in Dialogen und Kurzgeschichten	63
3.7	Vergleich zwischen /o:/ in gelesenen Kurzgeschichten und Dialogen für männliche Sprecher	64
3.8	Durchschnittliche Messwerte für /o:/ bei männlichen deutschen Sprechern in Dialogen und Kurzgeschichten	64
4.1	Ergebnisse des Wilcoxon-Rangsummentests für alle im Deutschen und Englischen vorhandenen untersuchten Vokale (p-Werte)	81
4.2	Ergebnisse eines gemischten Modells für /ɪ/ bei deutschen und britischen männlichen Sprechern	82
4.3	Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und britischen männlichen Sprechern	82
4.4	Eigenschaften des ermittelten gemischten Modells für /ɪ/ bei weiblichen Sprechern	84
4.5	Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und britischen weiblichen Sprechern	84
4.6	Ergebnisse für das gemischte Modell für /ʊ/ bei männlichen Sprechern	87
4.7	Durchschnittswerte des gemischten Modells für /ʊ/ bei deutschen und britischen männlichen Sprechern	88

4.8	Eigenschaften des gemischten Modells für /ʊ/ bei weiblichen Sprecherinnen . . .	90
4.9	Durchschnittswerte des gemischten Modells für /ʊ/ bei deutschen und britischen weiblichen Sprechern	90
4.10	Wilcoxon-Rangsummentest zur Unterscheidung zwischen /e:/ und /ɛ/ bei deutschen und britischen, männlichen und weiblichen Sprechern	93
4.11	Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei deutschen männlichen Sprechern	93
4.12	Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für deutsche männliche Sprecher	93
4.13	Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei deutschen weiblichen Sprechern	94
4.14	Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für deutsche weibliche Sprecher	94
4.15	Gemischtes Modell für die Unterscheidung zwischen /e:/ und /ɛ/ bei männlichen britischen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	95
4.16	Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für britische männliche Sprecher	95
4.17	Gemischtes Modell für die Unterscheidung zwischen /e:/ und /ɛ/ bei weiblichen britischen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	95
4.18	Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für britische weibliche Sprecher	96
4.19	Wilcoxon-Rangsummentests zur Aussprache von /e:/ und /ɛ/ bei deutschen und britischen Sprechern	97
4.20	Wilcoxon-Rangtest zu /ø:/ bei deutschen und britischen Sprechern	98
4.21	Durchschnittliche Werte für die Aussprache von /ø:/ durch deutsche und britische männliche Sprecher	99
4.22	Durchschnittliche Werte für die Aussprache von /ø:/ durch deutsche und britische weibliche Sprecher	99
4.23	Wilcoxon-Rangtest zu /y:/ bei deutschen und britischen Sprechern	101
5.1	Lateinische und italienische Monophthonge	104
5.2	Vokalviereck des heutigen Italienisch	105
5.3	Bedingungen für die Vokallänge im Italienischen	112
5.4	Gemischtes Modell für /i:/ und /ɪ/ bei deutschen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	116
5.5	Gemischtes Modell für /i:/ und /ɪ/ bei italienischen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	117
5.6	Durchschnittliche Messwerte für /i:/ und /ɪ/ bei deutschen und italienischen männlichen Sprechern	117
5.7	Gemischtes Modell für /i:/ und /ɪ/ bei deutschen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	118

5.8	Gemischtes Modell für /i:/ und /ɪ/ bei italienischen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	119
5.9	Durchschnittliche Messwerte für /i:/ und /ɪ/ bei deutschen und italienischen weiblichen Sprechern	119
5.10	Gemischtes Modell für /ɪ/ bei deutschen und italienischen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	121
5.11	Gemischtes Modell für /ɪ/ bei deutschen und italienischen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	123
5.12	Ergebnisse des Wilcoxon-Rangsummentests für /ø:/ bei deutschen und italienischen Sprechern	125
6.1	Struktur einer chinesischen Silbe	134
6.2	Zuordnung Pinyin – IPA	134
6.3	Vergleich der Vokaldauer bei deutschen und chinesischen männlichen Sprechern	151
6.4	Vergleich der Vokaldauer bei deutschen und chinesischen weiblichen Sprechern	152
6.5	Ergebnisse eines gemischten Modells für a-Cluster bei männlichen Sprechern .	159
6.6	Durchschnittswerte für das gemischte Modell der a-Cluster bei männlichen Sprechern	159
6.7	Ergebnisse eines gemischten Modells für a-Cluster bei weiblichen Sprechern . .	160
6.8	Durchschnittswerte für das gemischte Modell der a-Cluster bei weiblichen Sprechern	160
6.9	Ergebnisse des Wilcoxon-Rangsummentests für den Unterschied zwischen /i:/ und /ɪ/ bei deutschen männlichen und weiblichen Sprechern	162
6.10	Ergebnisse des Wilcoxon-Rangsummentests für den Unterschied zwischen /i:/ und /ɪ/ bei chinesischen männlichen und weiblichen Sprechern	162
6.11	Ergebnisse für das gemischte Modell für /i:/ bei männlichen deutschen und chinesischen Sprechern	166
6.12	Durchschnittswerte des gemischten Modells für /i:/ bei deutschen und chinesischen männlichen Sprechern	166
6.13	Ergebnisse für das gemischte Modell für /i:/ bei weiblichen deutschen und chinesischen Sprechern	167
6.14	Durchschnittswerte des gemischten Modells für /i:/ bei deutschen und chinesischen weiblichen Sprechern	167
6.15	Ergebnisse für ein gemischtes Modell für /ɪ/ bei männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	168
6.16	Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und chinesischen männlichen Sprechern	168
6.17	Ergebnisse für ein gemischtes Modell für /ɪ/ bei weiblichen Sprechern	169
6.18	Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und chinesischen weiblichen Sprechern	169
6.19	Unterschiede zwischen /u:/ und /ʊ/ für deutsche Sprecher	170

6.20	Unterschiede zwischen /u:/ und /ʊ/ für chinesische Sprecher	170
6.21	Mittelwerte für /u:/ und /ʊ/ bei männlichen und weiblichen chinesischen Sprechern	172
6.22	Ergebnisse für ein gemischtes Modell für /ɛ/ bei weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor	174
6.23	Durchschnittswerte des gemischten Modells für /ɛ/ bei deutschen und chinesischen weiblichen Sprechern	174
6.24	Ergebnisse des Wilcoxon-Rangsummentests für /ø:/ bei männlichen und weiblichen Sprechern	175
6.25	Durchschnittliche Messwerte für /ø:/ bei deutschen und chinesischen männlichen Sprechern	176
6.26	Durchschnittliche Messwerte für /ø:/ bei deutschen und chinesischen weiblichen Sprechern	177

Abbildungsverzeichnis

1.1	Ablauf des Hörvorgangs vom Ohr bis zum Hörzentrum	4
1.2	Für die Sprachverarbeitung zuständige Gehirnbereiche	5
2.1	Zusammenhang von Schiefe, Wölbung und Datenverteilung	31
2.2	Mögliche Clustertypen in modellbasierten Clustern	44
2.3	Auswahl statistischer Verfahren	45
3.1	Deutsche Vokalwerte bei männlichen Sprechern in einzelnen Worten	53
3.2	F_1 - und F_2 -Werte von männlichen Sprechern in gelesenen Texten und spontanen Unterhaltungen	54
3.3	F_1 - und F_2 -Werte von weiblichen Sprechern in gelesenen Texten und spontanen Unterhaltungen	55
3.4	Formantwerte aller deutschen Vokale (männliche Sprecher)	56
3.5	Formantwerte aller deutschen Vokale (weibliche Sprecher)	56
3.6	Vokaldauern aller untersuchten deutschen Vokale	57
3.7	Vokaldauer nach Textsorten für männliche Sprecher im Kiel Corpus	60
3.8	Vokaldauer nach Textsorten für weibliche Sprecher im Kiel Corpus	61
4.1	Englische Monophthonge	68
4.2	Übersicht zur Zungenmuskulatur	70
4.3	Durchschnittswerte für F_1 und F_2 bei von britischen männlichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen	71
4.4	Durchschnittswerte für F_1 und F_2 bei von britischen weiblichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen	71
4.5	Durchschnittswerte für F_1 und F_3 bei von britischen männlichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen	72
4.6	Durchschnittswerte für F_1 und F_3 bei von britischen weiblichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen	72
4.7	Durchschnittswerte für F_1 und F_2 bei deutschen und britischen männlichen Sprechern im LeaP-Corpus	74
4.8	Durchschnittswerte für F_1 und F_3 bei deutschen und britischen männlichen Sprechern im LeaP-Corpus	74
4.9	Durchschnittswerte für F_1 und F_2 bei deutschen weiblichen Sprechern und einer britischen weiblichen Sprecherin im LeaP-Corpus	75

4.10	Durchschnittswerte für F_1 und F_3 bei deutschen weiblichen Sprechern und einer britischen weiblichen Sprecherin im LeaP-Corpus	75
4.11	Vokale des zweisprachigen männlichen Sprechers ai	77
4.12	Vokale des zweisprachigen männlichen Sprechers cc	78
4.13	Verteilung der Formantwerte von F_1 und F_2 für /ɪ/ bei weiblichen Sprechern .	83
4.14	Klassifikationsbaum für die Formantwerte von /ɪ/ bei weiblichen deutschen und britischen Sprechern	85
4.15	Verteilung der Werte für /ʊ/ bei männlichen Sprechern	86
4.16	Verteilung der Formantwerte für /ʊ/ bei weiblichen Sprecherinnen	89
4.17	Verteilung der Werte von F_1 und F_2 bei /ø:/ für britische männliche Sprecher	100
4.18	Verteilung der Werte von F_1 und F_2 bei /ø:/ für britische weibliche Sprecher .	100
4.19	Klassifikationsbaum für die Formantwerte von /y:/ bei männlichen deutschen und britischen Sprechern	101
5.1	Werte für F_1 und F_2 bei männlichen deutschen und italienischen Sprechern . .	108
5.2	Werte für F_1 und F_2 bei weiblichen deutschen und italienischen Sprechern . .	108
5.3	Werte für F_1 und F_3 bei männlichen deutschen und italienischen Sprechern . .	109
5.4	Werte für F_1 und F_3 bei weiblichen deutschen und italienischen Sprechern . .	109
5.5	Werte für /ɪ/ und /i:/ in betonten und unbetonten Silben bei italienischen weiblichen Sprecherinnen	120
5.6	Aussprache von /ɪ/ durch deutsche und italienische männliche Sprecher	122
5.7	Aussprache von /ɪ/ durch deutsche und italienische weibliche Sprecher	124
5.8	Ergebnisse der Diskriminanzanalyse für /ø:/ bei männlichen deutschen und italienischen Sprechern	126
5.9	Ergebnisse der Diskriminanzanalyse für /ø:/ bei weiblichen deutschen und italienischen Sprechern	126
5.10	Vergleich der ersten drei Formanten für /ø:/ nach Sprachkenntnissen bei männlichen Sprechern	128
5.11	Vergleich der ersten drei Formanten für /ø:/ nach Sprachkenntnissen bei weiblichen Sprechern	129
5.12	Vergleich der durchschnittlichen Abweichungen für die Vokaldauer bei /ø:/ innerhalb und zwischen den Sprechergruppen bei weiblichen Sprechern	130
6.1	Verhältnis von lexikalischen zu neutralen Tönen im Chinesischen	135
6.2	Durchschnittliche Werte für F_1 und F_2 bei männlichen Sprechern	139
6.3	Durchschnittliche Werte für F_1 und F_3 bei männlichen Sprechern	139
6.4	Durchschnittliche Werte für F_1 und F_2 bei weiblichen Sprechern	140
6.5	Durchschnittliche Werte für F_1 und F_3 bei weiblichen Sprechern	140
6.6	Formantwerte einer chinesischen Muttersprachlerin nach Vokal	142
6.7	Durchschnittliche Werte für F_1 und F_2 bei deutsch sprechenden chinesischen Studenten	144

6.8	Formantwerte für Vokale von chinesischen Muttersprachlern, die im Deutschen und im Chinesischen auftreten	146
6.9	Formantwerte für Vokale von chinesischen Muttersprachlern, die nur im Deutschen und nicht im Chinesischen auftreten	147
6.10	Gesamte Werte für F_1 und F_2 bei Deutsch sprechenden chinesischen Studentinnen	148
6.11	Formantwerte für Vokale von chinesischen Muttersprachlerinnen, die im Deutschen und im Chinesischen auftreten	149
6.12	Formantwerte für Vokale von chinesischen Muttersprachlerinnen, die nur im Deutschen und nicht im Chinesischen auftreten	150
6.13	Automatische Klassifizierung von /a:/ und /a/ für chinesische männliche Sprecher	155
6.14	Automatische Klassifizierung von /a:/ und /a/ für chinesische weibliche Sprecherinnen	156
6.15	Messwerte weiblicher chinesischer Sprecherinnen für a-Cluster sortiert nach einzelnen Sprechern	156
6.16	F_1 -Werte für /a:/ und /a/ bei chinesischen Sprecherinnen nach dem Artikulationsort des vorherigen Konsonanten sortiert	157
6.17	F_2 -Werte für /a:/ und /a/ bei chinesischen Sprecherinnen nach dem Artikulationsort des vorherigen Konsonanten sortiert	158
6.18	Verteilung von /i:/ und /ɪ/ für deutsche männliche Sprecher	163
6.19	Verteilung von /i:/ und /ɪ/ für deutsche weibliche Sprecher	163
6.20	Verteilung von /i:/ und /ɪ/ für chinesische männliche Sprecher	164
6.21	Verteilung von /i:/ und /ɪ/ für chinesische weibliche Sprecher	164
6.22	Klassifikationsbaum für /u:/ und /ʊ/ für männliche chinesische Sprecher	171
6.23	Klassifikationsbaum für /u:/ und /ʊ/ für weibliche chinesische Sprecher	171
6.24	Verteilung der Formantwerte bei /ɛ/ für weibliche chinesische Sprecherinnen	173
6.25	Verteilung der Werte von F_1 und F_2 für /ø:/ bei männlichen Sprechern	176
6.26	Verteilung der Werte von F_1 und F_2 für /ø:/ bei weiblichen Sprecherinnen	177

1 Einleitung

1.1 Ausgangspunkt

1.1.1 Begründung der Fragestellung

Diese Arbeit soll dazu beitragen, durch eine genauere statistische Beschreibung der Ausspracheunterschiede zwischen deutschen Muttersprachlern und Sprechern, die Deutsch als Zweitsprache lernen, die Grundlage für ein verbessertes computerbasiertes Training der deutschen Aussprache zu legen. Wozu ist ein spezifisches Aussprachetraining überhaupt nötig?

Im englischen Sprachraum gibt es den Vorschlag stattdessen Muttersprachler darin zu schulen, „Englisch als internationale Sprache“ zu verstehen, s. (Jenkins, 2001, S. 227).

(Reinke, 2007) beschreibt die Reaktion einiger Germanistikstudenten (Deutsch als Fremdsprache) auf Aussprachefehler in einer Sprachaufnahme einer amerikanischen Studentin. Zuerst lehnen die Studenten Aussprachekorrekturen ab, da die Sprecherin trotzdem gut zu verstehen sei und der Akzent sich „richtig niedlich“ anhöre. Erst als die Dozentin nachfragt, ob die Sprecherin wirklich niedlich wirken wollte, werden die Studenten nachdenklich.

Durch dieses Beispiel wird deutlich, dass sowohl muttersprachliche dialektale als auch nicht-muttersprachliche Akzente auf bestimmte Art und Weise wahrgenommen werden können.

Dies trifft auf den Ausdruck regionaler Verbundenheit bei Dialektsprechern zu, genauso jedoch auf nicht-muttersprachliche Sprachvarietäten. So wird das „Lan-Deutsch“, das mit türkischen Jugendlichen in Deutschland verbunden wird und so auch z.B. in Comedy-Auftritten Verwendung findet, mit einer sozial benachteiligten Herkunft und einem zur Schau gestellten Machismo gleichgesetzt, s. (Byrd, 2009).

Das Ziel bei der Ausspracheschulung ist zuerst die gegenseitige Verständlichkeit, sowohl zwischen Muttersprachlern und Sprachlernenden als auch zwischen Zweitsprachlern mit unterschiedlichen Muttersprachen, z.B. Chinesisch oder Russisch.

Gleichzeitig sollte sie dem Lernenden helfen, ein Bewusstsein für die Wirkung seines muttersprachlichen Akzents sowie die eines u.U. erworbenen zweitsprachlichen Dialekts zu entwickeln.

Um Unterschiede in der Aussprache von Muttersprachlern und Nichtmuttersprachlern zu beschreiben, brauche ich Regeln, mit denen ich messbare Parameter bewerten kann.

Um solche Regeln auf der segmentalen Ebene formulieren zu können, greife ich hier auf die Vokalformanten als deutlich messbare Merkmale des Sprachsignals zurück.

1.1.2 These

Die akustischen Eigenschaften von muttersprachlichen und nicht-muttersprachlichen Vokalen in zusammenhängenden Äußerungen können mit statistischen Mitteln unterschieden werden, aus denen sich dann phonetische Regularitäten ableiten lassen.

1.2 Unterschiede zwischen Erstsprache und Zweitsprache

In diesem Unterkapitel möchte ich auf Erkenntnisse aus Neurolinguistik und Lernpsychologie eingehen, wie Sprachen erlernt und angewendet werden. Diese Informationen dienen auch dazu, den Unterschied zwischen Erstsprache bzw. -sprachen und Zweitsprache genauer zu beschreiben. Dabei lassen sich mit Hilfe der Neurolinguistik die zeitlichen Fenster für den Erst- und Zweitspracherwerb und die Unterschiede bei der Verarbeitung von Erst- und Zweitsprache im Gehirn genauer charakterisieren. Die Lernpsychologie dagegen beschreibt, mit welchen Schritten Fremdsprachenlerner eine neue Sprache erwerben und welche Strategien dabei mehr oder weniger geeignet sind.

1.2.1 Verarbeitung von Sprache im Gehirn

1.2.1.1 Untersuchungsmethoden

EEG Beim Elektroenzephalogramm (EEG) werden Elektroden an der Kopfhaut angebracht, um elektrische Impulse in darunter liegenden Hirnarealen zu messen. Da der Impuls eines einzelnen Neurons sehr schwach ist und das Signal durch mehrere Schichten hindurch gemessen wird (d.h. Gehirnflüssigkeit, Hirnhaut, Schädel, Haut) können nur Signale erfasst werden, die über Tausende von gleichzeitig aktivierten Neuronen übermittelt werden, s. (Bear, Connors & Paradiso, 2007, S. 586-588).

MEG Das Magnetenzephalogramm (MEG) misst durch die elektrische Ladung der Neuronen erzeugte Magnetfelder. Dabei ist der Aufwand zur Abschirmung des untersuchten Körperteils von anderen Magnetfeldern größer (z.B. Stromleitungen, Aufzüge oder das natürliche Magnetfeld der Erde), dafür können jedoch Vorgänge verfolgt werden, die nicht direkt auf der Oberfläche des Gehirns stattfinden, s. (Bear et al., 2007, S. 586-588).

PET Die *position emission tomography* (PET) wird durchgeführt, indem mit radioaktivem Fluor oder Sauerstoff versehener Traubenzucker in die Blutbahn injiziert wird. Danach wird die Verteilung und Aufnahme dieses Traubenzuckers im Gehirn gemessen, was Rückschlüsse auf die Gehirnaktivität zulässt, s. (Bear et al., 2007, S. 178f).

fMRI *functional magnetic resonance imaging* (fMRI) beruht auf dem Prinzip, dass sauerstoffreiches Hämoglobin andere magnetische Eigenschaften besitzt als sauerstoffarmes. Wird ein bestimmter Bereich des Gehirns stärker aktiviert, so nehmen seine Durchblutung und die Menge des abgegebenen Sauerstoffs zu, s. (Bear et al., 2007, S. 178).

NIRS Mit Hilfe der *near-infrared spectroscopy* (NIRS) werden Veränderungen im Hämoglobin- und im Sauerstoffgehalt des Blutes und damit in der Aktivität verschiedener Bereiche der Hirnrinde gemessen. Wird ein bestimmter Bereich des Gehirns aktiviert, erhöht sich die Menge des sauerstoffreichen Hämoglobin in den Blutgefäßen, s. auch (Hull, Bortfeld & Koons, 2009).

1.2.1.2 Perzeption

Der erste Teil des Weges vom Schall zur gehörten Sprache verläuft folgendermaßen:

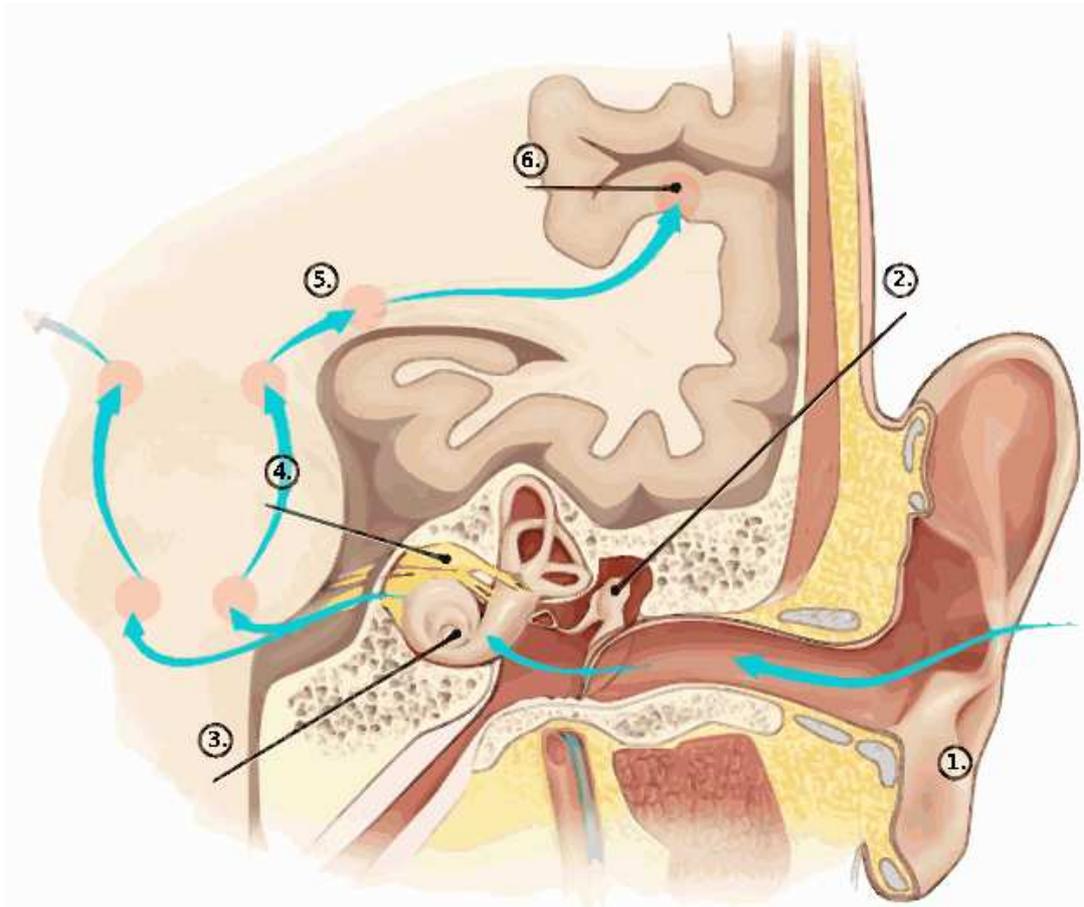
1. Der Schall erreicht das Ohr und über den Gehörgang das Trommelfell.
2. Die Erschütterungen des Trommelfells werden von den Gehörknöchelchen an eine zweite Membran vor der Schnecke weitergegeben.
3. Auf der Innenfläche der mit Flüssigkeit gefüllten Schnecke reagieren die Haarzellen in verschiedenen Bereichen auf unterschiedliche Frequenzen und geben die Vibrationen an Nervenzellen weiter.
4. Die Signale werden von der Schnecke über Nervenbahnen an das Stammhirn weitergeleitet.
5. Hier teilt sich die Hörbahn auf jeder Seite in zwei Bahnen, von denen eine zur jeweiligen Gegenseite kreuzt, während die andere auf der gleichen Seite bleibt. Im Mittelhirn kreuzen sich die Hörbahnen noch einmal, da hier das Richtungshören stattfindet.
6. Die Hörzentren, die auf beiden Seiten des Gehirns in den Heschl-Querwindungen liegen empfangen die Signale.

s. auch (Schindelmeiser, 2005, S. 197–220) sowie (Bear et al., 2007, S. 344–368)

Im Gegensatz zu Geräuschen, z.B. Tierstimmen, Maschinen etc. ist hier die Verarbeitung der Sprachsignale noch nicht zu Ende.

(Eulitz, Diesch, Pantev, Hampson & Elbert, 1995) analysiert die Messergebnisse eines EEG und eines MEG bei elf Versuchspersonen, denen synthetisierte Vokale und Töne präsentiert werden. Dabei werden für die synthetisierten Vokale die ersten drei Formanten

Abbildung 1.1: Ablauf des Hörvorgangs vom Ohr bis zum Hörzentrum



(Deretsky, 2006)

der Vokale /a/, /æ/, /i/, /u/ und /œ/ eingesetzt, der synthetisierte Ton hat eine zentrale Frequenz von 1000 Hertz.

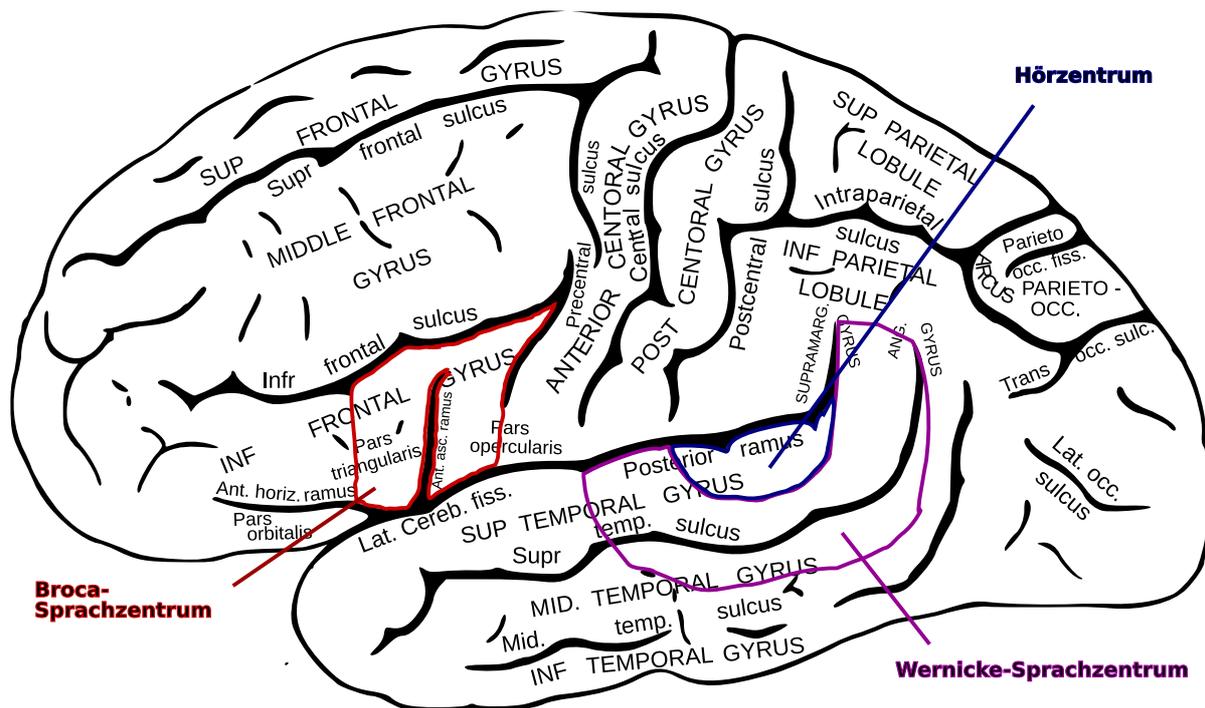
Als Reaktion auf auditive Stimuli werden dabei ein ausgeprägtes negatives Potential nach 100 ms (N100) und ein positives Signal nach ca. 150-200 ms gemessen. Dauert das gehörte Signal länger als 150-200 ms, so verschieben sich die elektrischen Impulse und elektromagnetischen Ladungen.

In diesen Versuchen fällt die Verschiebung für Vokale stärker aus als für synthetisierte Töne und der Unterschied zwischen der Verschiebung bei Vokalen und bei Tönen ist auf der linken Seite des Kortex größer als auf der rechten. Für das negative Potential nach 100 ms lassen sich keine eindeutigen Ergebnisse feststellen.

(Gootjes, Raij, Salmelin & Hari, 1999) vergleicht die elektromagnetischen Ladungen beim Hören von fünf gesprochenen finnischen Vokalen, am Klavier gespielten Noten und synthetisierten Tönen.

Hier sind die Unterschiede bei dem gemessenen negativen Potential eindeutig: die Reaktionen auf die Vokale in der linken Hirnhälfte fallen signifikant stärker aus als die Reaktionen auf Klaviernoten oder synthetische Töne.

Abbildung 1.2: Für die Sprachverarbeitung zuständige Gehirnbereiche



(Gray, 1918, Fig. 726), farbige Hervorhebungen von der Verfasserin

(Binder et al., 1997) setzt fMRI (s. 1.2.1.1 ein, um Unterschiede zwischen der Verarbeitung von Tönen und einzelnen gesprochenen Worten, in diesem Fall Tiernamen, festzustellen.

In einem ersten Durchgang hören die Versuchspersonen, 15 Männer und 15 Frauen, die Töne bzw. Sprachaufnahmen an, ohne darauf reagieren zu müssen. Im nächsten Durchgang sollen sie unter folgenden Bedingungen einen Knopf betätigen: bei zwei gleichen Tönen (750 Hz) oder bei dem Namen eines Nutztieres, das in den USA beheimatet ist. Beim Anhören der Töne wird vor allem das Hörzentrum innerhalb der linken Gehirnhälfte aktiviert sowie die hintere Hälfte des Gyrus temporalis medius (mittlere Schläfenwindung) innerhalb der rechten Gehirnhälfte und auf beiden Seiten der Gyrus supramarginalis, ein Teil des Lobus parietalis (Scheitellappen), der vordere Bereich des motorischen Kortex, der vordere Gyrus cinguli und der vordere Bereich des insulären Kortex (Inselregion). Beim Erkennen der Tiernamen werden folgende Bereiche aktiviert:

- der Gyrus temporalis superior (obere Schläfenwindung) auf beiden Seiten, dessen linke Seite das Wernicke-Areal enthält,
- ein großer Teil des Gyrus temporalis medius auf der linken Seite,
- der gesamte Gyrus frontalis inferior (untere Stirnwindung), der auch das Broca-Areal enthält,
- der vorderer und hintere Abschnitt des Gyrus frontalis medius (mittlere Stirnwindung),

- vor allem in der linken Hirnhälfte der Gyrus angularis, ein Bestandteil des Lobus parietalis,
- der vordere und hintere cinguläre Kortex,
- der innere, mittlere Bereich des Precuneus und der Isthmus gyri cinguli,
- sowie der rechte hintere Bereich des Cerebellum (Kleinhirn).

Die bei der Verarbeitung von Tönen und Worten aktivierten Hirnbereiche unterscheiden sich also deutlich ¹. Die Autoren erklären die Aktivität außerhalb des Wernicke-Areals damit, dass Schäden am Gyrus temporalis superior nicht zum totalen Ausfall des Sprachverständnisses führen, sondern zu Worttaubheit. Dagegen lösen Schäden im Bereich des Gyrus temporalis medius und Gyrus temporalis inferior (untere Schläfenwindung) multimodale Wortverständnisprobleme aus, die jeweils auf bestimmte Wortklassen beschränkt sind. Ansonsten ist der Gyrus supramarginalis bei Tönen mehr beansprucht, der mit für das akustische Kurzzeitgedächtnis verantwortlich ist. Bei den Tiernamen ist ein Bereich in der Nähe des Corpus callosum (Balken) stärker aktiviert, der mit dem Wortgedächtnis in Verbindung gebracht wird.

(Mäkelä, Alku, May, Mäkinen & Tiitinen, 2005) vergleicht mit Hilfe der Magnetenzephalographie (MEG) (s. 1.2.1.1) die Reaktionen des Gehirns bei Testpersonen, denen synthetisierte isolierte Vokale und Diphthonge präsentiert werden, während sie lesen oder einen Stummfilm betrachten.

Außer einer Reaktion nach 100 ms (N100m) wird bei Diphthongen, die in diesem Fall als Formantübergänge synthetisiert sind, während der nächsten 200-500 ms eine stärkere Aktivität des Kortex gemessen, die mit elektrischen Impulsen bei anderen Versuchen zum Hörverständnis, nämlich P200 (phonologische Verarbeitung) und N400 (Verarbeitung von Worten), in Verbindung gebracht wird. Die Reaktionen fallen für /a/ und /au/ stärker aus als für /u/ und /a/ und treten bei /a/ und /au/ 10 ms früher ein.

Dies könnte durch die Häufigkeit der Vokale im Finnischen bedingt sein.

(Schofield et al., 2009) misst ebenfalls mit Hilfe eines Magnetenzephalogramms die Reaktionen seiner Versuchspersonen auf vier verschiedene, einzeln aufgenommene Worte bzw. synthetische Töne. Es handelt sich dabei um eine Aufnahme des Wortes *bart*, in die ein synthetisierter Vokalabschnitt eingefügt wird. Dabei werden in einem Fall die Formantfrequenzen überhaupt nicht, in einem Fall nur leicht und in zwei Fällen entscheidend zu *burt* oder *beat* geändert wurden. Die Töne bestehen aus Sinusschwingungen, deren zentrale Frequenz jeweils dem zweiten Formanten der Vokale entspricht.

Den Testpersonen wird wiederholt die Originalaufnahme des Wortes bzw. der entsprechende synthetisierte Ton vorgespielt, zwischen diesen Wiederholungen werden abgeänderte Aufnahmen eingefügt (20% der eingesetzten Stimuli). Währenddessen müssen die Versuchspersonen eine optische Aufgabe lösen.

¹Zur Übersetzung der englischen Termini in deutsche bzw. lateinische Fachbegriffe, s. auch (Karnath & Thier, 2006, S. 703-706)

Danach werden die Reaktionen auf die abweichenden Worte bzw. Töne mit Hilfe des *dynamic causal modelling* (DCM) statistisch ausgewertet, wobei sich in beiden Fällen als Teil des *mismatch negativity field* 150-250 ms nach dem Signal Gehirnaktivität im Lobus temporalis, für die Töne zusätzliche Aktivität im primären auditiven Kortex der rechten Hirnhälfte und für die Worte zusätzliche Aktivität im hinteren Gyrus temporalis superior der linken Gehirnhälfte nachweisen läßt.

(Friederici, 1995) stellt auf der Grundlage verschiedener Untersuchungen ein Modell der Sprachwahrnehmung in drei Stufen vor. In der ersten Phase werden vor allem im vorderen Bereich der linken Gehirnhälfte Worte und Phrasen erkannt. Die zweite Phase besteht aus einer lexikalisch-semanticen und syntaktischen Analyse, bei der z.B. wenn das Ende eines Satzes erkannt wird, nach 400 ms ein ausgeprägter negativer Impuls gemessen werden kann (N400). In der dritten Phase findet eine syntaktische und semantische Reanalyse statt, z.B. bei Sätzen, die mehrere Lesarten aufweisen, die einen positiven Impuls nach ca. 600 ms auslöst (P600).

Abschließend läßt sich sagen, dass die Verarbeitung von Vokalen in die erste Phase dieses Modells fällt und dass das menschliche Gehirn sehr genau zwischen Sprachlauten und anderen Klängen unterscheidet. Auch die Reaktionen auf synthetisierte Vokale und natürliche Sprachaufnahmen unterscheiden sich in ihrer Intensität. Außerdem können schon die Formantübergänge innerhalb eines Diphthongs Wortverarbeitungsprozesse auslösen.

1.2.1.3 Produktion

(Levelt, 1999) entwirft ein Modell der Sprachproduktion in mehreren Schritten

Begriffliche Vorbereitung Dieser Schritt umfasst die Suche nach einem Begriff als Bestandteil einer größeren Nachricht oder um ein einzelnes Objekt zu benennen. Hier wird auch entschieden, ob ein allgemeiner Oberbegriff oder eine spezifischere Bezeichnung verwendet werden soll.

Lexikalische Auswahl Hier wird das konkret gemeinte Wort aus einer Menge von verwandten Begriffen ausgewählt.

Morphologisch-Phonologische Umsetzung Die syntaktisch notwendigen morphologischen Bestandteile, z.B. die Endung eines Verbs, werden hinzugefügt. Danach erfolgt auch die Aufteilung in Silben und die Betonung der Worte wird festgelegt.

Phonetik und Artikulation Die Worte werden in artikulatorische Gesten umgesetzt.

(Dogil et al., 2002) erweitert dieses Modell anhand von Ergebnissen aus Experimenten und unterscheidet zwischen dem rhetorisch/semantisch/syntaktischen System, das die Auswahl der Begriffe und die grammatischen Bedingungen der Sprache umfasst und dem phonologisch/phonetischen System, das die grammatische Struktur mit Hilfe des mentalen Lexikons füllt und anhand der mit einzelnen Silben verbundenen phonetischen Gesten die endgültige Artikulation erzeugt.

In insgesamt vier Versuchen werden die Unterschiede zwischen gedachten und artikulierten Worten, einzelnen Silben und mehrsilbigen Worten, monotonen und betonten Silbenfolgen sowie Wortlisten und ganzen Sätzen untersucht.

Gedachte und artikuliertete Worte Werden im ersten Versuch die zu sprechenden Monatsnamen nur gedacht, so zeigt sich dabei Aktivität im linken motorischen Kortex und im supplementär-motorischen Kortex sowie auf der rechten Hirnhälfte im Cerebellum. Bei tatsächlich ausgesprochenen Monatsnamen werden sowohl das Cerebellum als auch der motorische Kortex in beiden Hirnhälften angesprochen, zusätzlich wird der rechte Bereich des vorderen insulären Kortex aktiviert. Dieser spielt nur bei artikulierten, nicht bei gedachten Worten eine Rolle.

Mehrsilbige Worte und Silbenfolgen Der zweite Versuch zeigt beim Nachsprechen mehrsilbiger sinnvoller Worte oder reiner Silbenfolgen nur Aktivität in einem zentralen Bereich des motorischen Kortex der linken Hirnhälfte an. Werden dagegen nur einzelne Silben vorgegeben, so zeigt sich in beiden Hirnhälften im Bereich des motorischen Kortex eine Aktivierung. Die Hirnaktivität ist bei der komplexeren, aber im Deutschen häufigeren Silbe *stra* geringer als bei der einfacheren, aber seltener vorkommenden Silbe *ta*.

Intonation Das dritte Experiment besteht aus Silbenfolgen wie *dadadada*, die mit monotoner Stimme oder einer vorgegebenen Betonung gesprochen werden. Dabei werden entweder bestimmte betonte Silben vorgegeben, eine insgesamt steigende oder fallende Intonation oder eine fröhlicher bzw. trauriger Tonfall.

Das Aussprechen monotoner Silben aktiviert Bereiche im motorischen Kortex sowie im oberen seitlichen Bereich des Cerebellum. Bei betonten Silben wird zusätzlich der linke Gyrus temporalis superior eingesetzt und es wird eine insgesamt steigende oder fallende Intonation oder ein entsprechender Tonfall vorgegeben, so zeigt sich auch auf der rechten Hirnhälfte im Gyrus temporalis superior eine Reaktion.

Die Erzeugung der Intonation wird beim Menschen nur durch das Großhirn gesteuert, weder das limbische System noch das Mittelhirn waren daran beteiligt.

Wortlisten und vollständige Sätze Der vierte Versuch vergleicht die Gehirnaktivität beim Lesen von unterschiedlich angeordneten Wortlisten mit der beim Lesen syntaktisch unterschiedlich angeordneter, ansonsten identischer Sätze.

Dabei wird beim Lesen ganzer Sätze im Gegensatz zur Wortliste innerhalb der linken Gehirnhälfte im Lobus frontalis (Stirnappen) ein Bereich angesprochen, der sich bis zum Broca-Areal erstreckt und im Lobus temporalis ein Bereich, der teilweise im

Wernicke-Areal liegt. Außerdem spielen noch der vorherige cinguläre Kortex und das Cerebellum eine Rolle.

1.2.2 Erwerb der Erstsprache

1.2.2.1 Perzeption – Abgrenzung von anderen Sprachen

Die Wahrnehmung sprachlicher Merkmale lässt sich schon bei Neugeborenen feststellen. (Kujala et al., 2004) stellt anhand von MEG-Veränderungen bei Neugeborenen fest, dass diese ein /i:/ oder /a:/ sowie ein /a:/ mit gleichbleibender und /a:/ mit steigender Intonation voneinander unterscheiden können.

(Bosch & Sebastián-Gallés, 2003) untersucht die Veränderung der Wahrnehmung von phonetischen Kontrasten, die sich im ersten Lebensjahr abspielt und bei der zuerst für Vokale (5.-7. Monat) und später auch für Konsonanten (ab. dem 8. Monat) nur noch die Unterschiede registriert werden, die Teil der in der Umgebung des Säuglings gesprochenen, zuerst erworbenen Sprache sind.

Dazu vergleicht sie Kinder, die in einer spanischen einsprachigen, katalanischen einsprachigen oder spanisch-katalanisch zweisprachigen Umgebung aufwachsen. Dabei untersucht sie die Unterscheidung zwischen /e:/ und /ɛ/, die im Katalanischen als separate Phoneeme wahrgenommen werden, im Spanischen jedoch nicht. Die vier Monate alten Säuglinge können noch alle den Kontrast zwischen /e:/ und /ɛ/ wahrnehmen, nach acht Monaten ist dies nur noch bei der katalanischen einsprachigen Gruppe der Fall und im Alter von einem Jahr unterscheiden sowohl die katalanischen einsprachigen als auch die spanisch-katalanisch zweisprachigen Kinder zwischen /e:/ und /ɛ/, die spanischen einsprachigen jedoch nicht.

(Minagawa-Kawai, Mori, Naoi & Kojima, 2007) stellt mit Hilfe der near-infrared spectroscopy (s. 1.2.1.1) fest, dass japanische Säuglinge schon ab dem Alter von sechs Monaten Unterschiede in der Vokaldauer zwischen /a:/ und /a/ wahrnehmen. Die Reaktion des Gehirns findet jedoch zuerst in beiden Hirnhälften statt und erst ab dem Alter von 13 Monaten vor allem in der linken Gehirnhälfte wie bei Erwachsenen.

1.2.2.2 Produktion

(Wode, 1993, S. 185) fasst den Erwerb der muttersprachlichen Sprachproduktion in fünf Abschnitten zusammen:

1. Im ersten Lebensjahr findet vor allem eine Entwicklung der Sprachwahrnehmung statt, außerdem vorsprachliche Äußerungen wie Lallen oder Brabbeln.
2. In den darauf folgenden neun Monaten lernt das Kind die ersten 50 Worte, ein grundlegendes Lautinventar. Es kann zwischen unterschiedlich gesprochenen Wörtern unterscheiden, jedoch nicht zwischen einzelnen Lauten wie Erwachsene.

3. Mit ca. vier Jahren hat das Kind das Lautinventar seiner Muttersprache erworben. Außerdem erkennt es einfache Morpheme, phonologische Prozesse und einzelner Laute.
4. Ab sieben Jahren werden komplexe Morpheme und Wörter erkannt.
5. Zwischen sieben bis zwölf Jahren werden morphologische Alternationen gelernt, dazu gehören z.B. Bedeutungsunterschiede durch unterschiedliche Wortbetonung.

1.2.3 Rolle des Alters beim Erwerb der Zweitsprache

1.2.3.1 Critical Period Hypothesis

(Long, 1993) fasst Untersuchungen zur Critical Period Hypothesis zusammen, nach der die Fähigkeit, eine Sprache zu erlernen, bei Kindern optimal ausgeprägt ist, während der Pubertät nachlässt und bei Erwachsenen stark variiert. Er unterscheidet dabei drei verschiedene Altersstufen, in denen Personen mit dem Erlernen einer zweiten Sprache beginnen.

0 – 6 Jahre Die Lernenden können in diesem Alter bei ausreichendem Kontakt mit der Zweitsprache muttersprachliches Niveau erreichen.

6 – 15 Jahre Der Umfang der zweitsprachlichen Kenntnisse als Erwachsener ist hier desto größer, je früher die Kinder mit dem Erlernen der Zweitsprache beginnen. Die Wahrscheinlichkeit, dass eine muttersprachliche Aussprache im segmentalen und suprasegmentalen Bereich erreicht wird, nimmt rapide ab.

älter als 15 Jahre Hier liegt die erreichte Nähe zur muttersprachlichen Kompetenz unter derjenigen jüngerer Zweitsprachenlerner, wobei neben dem Alter viele weitere Faktoren das Lernergebnis beeinflussen.

(Johnson & Newport, 1989) vergleicht das Verständnis für grammatikalisch korrekte englische Sätze bei koreanischen und chinesischen Muttersprachlern, von denen die eine Hälfte der Versuchspersonen schon als Kinder und Jugendliche in die USA einwanderte, die andere Hälfte erst als Erwachsene. Den Testpersonen werden Aufnahmen verschiedener Sätze vorgespielt, sowohl grammatisch korrekte als auch in einem Punkt fehlerhafte Äußerungen, die sie dann als richtig oder falsch beurteilen müssen.

Dabei unterscheidet sich die Leistung der Personen, die zwischen drei und sieben Jahren in die USA kamen, nicht signifikant von derjenigen der Muttersprachler. Bei den später eingewanderten Gruppen zeigen sich jedoch signifikante Unterschiede. Diese Unterschiede bestehen sowohl im Vergleich zu den Muttersprachlern als auch im Vergleich der Gruppen untereinander. Von den als 8-10-Jährige angekommenen über die mit 11 bis 15 Jahren angekommenen Teilnehmer, bis zu denen, die erst mit 17 Jahren oder älter, also als Erwachsene, in den USA ankamen, nimmt die Anzahl der im Test erzielten Punkte ab und

die Streuung der Ergebnisse zu. Bis zu einem Ankunftsalter von 17 Jahren besteht eine direkte Korrelation zwischen Testergebnis und Ankunftsalter. Bei Erwachsenen zeigt sich diese Korrelation nicht mehr, dafür treten individuelle Unterschiede in den Vordergrund. Aufgrund dieser Ergebnisse und anderer Untersuchungen zum Erstspracherwerb, z.B. bei Gehörlosen, die teilweise erst als Schulkinder oder Jugendliche die Gebärdensprache erlernen, kommen die Autoren zu dem Schluss, dass sich der Zeitpunkt des Spracherwerbs deutlich darauf auswirkt, in welchem Umfang die Sprache erworben wird und dass sich erste Unterschiede schon vor der Pubertät zeigen.

1.2.3.2 Untersuchungen zu verschiedenen Altersstufen vor der Pubertät

(Díaz, Baus, Escera, Costa & Sebastián-Gallés, 2008) untersucht die mit einem EEG gemessenen *event-related potentials* (ERP) bei spanischen Muttersprachlern, die spätestens mit vier Jahren anfangen, Katalanisch zu lernen. Alle Versuchspersonen haben sehr gute Sprachkenntnisse in beiden Sprachen und wenden diese im täglichen Leben an. Dabei besteht die Hälfte der Versuchspersonen aus Sprechern, die sehr gute Ergebnisse in einem Perzeptionstest zu den katalanischen Vokalen /e/ und /ɛ/ erzielen und dabei keine Unterschiede zu katalanischen Muttersprachlern aufweisen. Die andere Hälfte besteht aus Sprechern, deren Testergebnisse im Perzeptionstest von denen katalanischer Muttersprachler abweichen.

Die *mismatch negativity* (MMN) tritt auf, sobald ein akustischer Impuls von den vorher gehörten akustischen Impulsen abweicht. Die Reaktion auf diesen Unterschied wird als starker negativer Ausschlag im *event related potential* sichtbar. Diese Reaktion tritt auch dann auf, wenn die Versuchsperson nicht aufmerksam zuhört und wird ausgelöst, wenn die Frequenz des derzeitigen Impulses von den vorherigen Impulsen abweicht, seine Intensität, seine Dauer oder die Richtung, aus der der Impuls wahrgenommen wird. Dies gilt für sprachliche und nichtsprachliche Klänge. Es ergibt sich keine MMN als Reaktion auf den ersten Impuls einer Reihe von Klängen auf oder auf Klänge, zwischen denen sehr lange Zeitintervalle liegen. Das Ausmaß der MMN nimmt zu, wenn Versuchspersonen vorher aktiv lernen Klänge zu unterscheiden. Der Frequenzabstand, ab dem zwischen Klängen unterschieden wird, ist bei sehr guten Versuchspersonen viermal geringer als bei schlechten, s. (Näätänen & Alho, 1995).

Anhand dieses Signals prüft (Díaz et al., 2008) die Reaktionen auf zwei Vokalkontraste: /e/ und /o/, die in beiden Sprachen vorhanden sind und /o/ und /ø/, die in beiden Sprachen unbekannt sind. Zur Kontrolle werden auch die Reaktionen auf synthetisierte Töne aufgezeichnet, die sich in Dauer, Frequenzstruktur und Reihenfolge unterscheiden. Die Teilnehmer sehen während des Versuchs einen Stummfilm und sollen die vorgespielten Aufnahmen nicht bewusst bewerten.

Die Reaktionen beider Gruppen auf die synthetisierten Töne unterscheiden sich nicht signifikant, diejenigen auf die Vokale fallen bei den Sprechern mit sehr guten Ergebnissen im Perzeptionstest im Frontallappen stärker aus als bei den mittelmäßigen Sprechern.

Aus dieser Untersuchung ergibt sich die interessante Schlussfolgerung, dass selbst bei Sprechern die eine Sprache sehr früh als Zweitsprache erlernen und fließend sprechen, Unterschiede in der Sprachwahrnehmung auftreten können. Leider werden in der Studie keine detaillierteren Angaben zur Art und Weise oder dem Umfang des Zweitspracherwerbs erhoben, z.B. ob die Versuchspersonen Katalanisch durch ein zweisprachiges Elternteil oder erst im Kindergarten erlernten.

(Dehaene et al., 1997) vergleicht die Verarbeitung von Erst- und Zweitsprache im Gehirn bei französischen Muttersprachlern, die ab dem Alter von sieben Jahren Englisch in der Schule gelernt hatten und durchschnittliche Sprachkenntnisse aufwiesen mit Hilfe des fMRI (s. 1.2.1.1).

Dabei hören die Testpersonen verschiedene Geschichten in ihrer Erst- und ihrer Zweitsprache. Während die muttersprachliche französische Aufnahme abgespielt wird, lässt sich Aktivität im linken Lobus temporalis und darin vor allem im linken Gyrus temporalis medius und superior, im linken Sulcus temporalis superior und im linken Gyrus angularis feststellen.

Wird eine zweitsprachliche Aufnahme präsentiert, variieren die aktivierten Areale deutlich zwischen den Testpersonen. Bei sechs Testpersonen werden der linke Lobus temporalis und dort der Gyrus temporalis medius und superior aktiviert, wobei die Gehirnaktivität schwächer und weiter gestreut ausfällt als in der Muttersprache. Gleichzeitig werden im rechten Lobus temporalis der Gyrus temporalis superior und der Sulcus temporalis superior aktiviert. Zwei Testpersonen weisen nur eine Gehirnaktivität im rechten Lobus temporalis auf, d.h. die Zweitsprache wird hauptsächlich in der rechten Gehirnhälfte verarbeitet.

Die Autoren sind sich nicht sicher, ob die Unterschiede bei der Zweitsprachwahrnehmung sich auf spezifische Lehr- und Lernmethoden oder auf individuelle Merkmale der Funktionsweise des Gehirns zurückführen lassen.

Leider fehlen hier weitere Angaben dazu, wie intensiv die Zweitsprache erlernt wurde (nur als fremdsprachlicher Unterricht oder auch in anderen Schulfächern und im Alltag) und wie oft sie von den Erwachsenen angewendet wird. Der Leser sieht jedoch deutlich, dass die Abläufe beim Verstehen der Zweitsprache sich hier grundsätzlich von denen beim Verstehen der Muttersprache abheben, obwohl das Lernen noch vor der Pubertät beginnt.

(Nenonen, Shestakova, Huotilainen & Näätänen, 2005) beschäftigt sich speziell damit, wie russische Schüler den im Finnischen bedeutungstragenden Kontrast im Hinblick auf die Vokallänge erwerben. Die Versuchsgruppe besteht aus elf Jugendlichen, die zwischen dem Alter von fünf und neun Jahren nach Finnland kamen und finnische Schulen besuchen. Einige von ihnen betrachten sich als komplett zweisprachig. In der Kontrollgruppe nehmen dreizehn einsprachige finnische Jugendliche an dem Versuch teil.

Dabei wird das Ausmaß der *mismatch negativity* 1.3.2.1 gemessen, während den Testpersonen Silben von [ka:] nach [kɑ] und von [kæ:] nach [kæ] vorgespielt werden, wobei /ɑ/ in beiden Sprachen, /æ/ jedoch nur im Finnischen ein Phonem darstellt.

Dabei ergibt sich für die Silben, die auf /æ/ enden, in beiden Gruppen eine Reaktion auf die Änderung der Vokallänge, für die Silben auf /ɑ/ jedoch nur in der finnischen Gruppe. Dies könnte dadurch bedingt sein, dass /ɑ/ als im Russischen bekannter Vokal in das muttersprachliche Vokalsystem eingeordnet wurde, während /æ:/ und /æ/ als „neue“ Vokale erlernt wurden.

In diesem Fall ist den Versuchspersonen das Erlernen neuer Phoneme vor der Pubertät gelungen, sie können jedoch nicht die in der Zweitsprache vorhandenen zusätzlichen phonetischen Kontraste, wie die Vokaldauer, auf schon aus der Muttersprache bekannte Vokale anwenden.

(Klein, Zatorre, Milner, Meyer & Evans, 1994) setzt die PET (s. 1.2.1.1) ein, um Unterschiede bei der Durchblutung des Gehirns festzustellen, während die Versuchspersonen muttersprachliche oder zweitsprachliche Begriffe nachsprechen. Bei den Teilnehmern der Studie handelt es sich um zwölf junge Erwachsene, die Englisch als Muttersprache erlernt hatten, Französisch jedoch erst, als sie älter als 5 Jahre waren. Alle Sprecher sprachen die Worte zu 99% (Englisch) bzw. 96% (Französisch) korrekt nach.

Während des Nachsprechens der französischen Worte, wurde jedoch eine zusätzliche Aktivität im linken Putamen gemessen. Schädigungen dieses Gehirnareals, z.B. durch Schlaganfälle, äußern sich im „foreign accent syndrome“. Dabei weisen die Patienten artikulatorische Probleme auf, die an einen fremdsprachlichen Akzent erinnern. Deshalb gehen die Autoren davon aus, dass die Aktivierung des linken Putamen auf einen erhöhten artikulatorischen Aufwand hinweist.

Während ich für Sprecher, die vor dem Alter von sechs Jahren ihre Zweitsprache erlernen, in der Fachliteratur keine Hinweise auf Unterschiede bei der Sprachproduktion finden kann, so zeigt sich hier, trotz des Erwerbs der zweiten Sprache noch vor der Pubertät, zumindest auf neurologischer Ebene ein zusätzlicher Arbeitsaufwand des Gehirns beim Sprechvorgang.

1.2.3.3 Zweitspracherwerb nach der Pubertät

(Frenck-Mestre, Anton, Roth, Vaid & Viallet, 2005) vergleicht die Gehirnaktivität bei Sprechern, die von Geburt an zweisprachig (Englisch und Französisch) aufgewachsen waren und von Sprechern, die Französisch erst mit 12 Jahren oder später gelernt hatten.

Die Versuchspersonen lesen englische oder französische Worte oder bedeutungslose Zeichenketten, die auf einem Bildschirm angezeigt werden. Die währenddessen gemessenen fMRI-Daten werden dann verglichen, wobei in beiden Gruppen in beiden Gehirnhälften folgende Bereiche aktiviert werden: motorischer Kortex, Gyrus temporalis medius und Gyrus temporalis superior, Cerebellum und supplementär-motorischer Kortex. Der einzige sichtbare, jedoch nicht signifikante Unterschied zwischen beiden Gruppen liegt in einer stärkeren Aktivierung des Putamens bei den Sprechern, die später Französisch gelernt hatten.

Da die Versuchspersonen nur Worte ablesen müssen, ohne diese zu bewerten oder die

Informationen weiter zu verarbeiten, ist es möglich, dass bei anderen Aufgaben größere Unterschiede in der Gehirnaktivität sichtbar werden.

(Videsott et al., 2010) untersucht die Vorgänge im Gehirn bei der Sprachproduktion multilingualer Sprecher, die mit Ladinisch als Muttersprache, Italienisch als Zweitsprache ab dem Alter von 5 Jahren und Englisch als vierter Fremdsprache (nach Deutsch) mit 14 Jahren aufwuchsen.

In drei Testläufen, je einmal für Ladinisch, Italienisch und Englisch, benennen die Teilnehmer als Bilder angezeigte Begriffe. dabei werden gleichzeitig fMRI-Daten erhoben und die gesprochenen Worte aufgezeichnet. Die Aufgabe wird sehr gut für Ladinisch und Italienisch gelöst, bei den englischen Begriffen sind durchschnittlich 76% der Bezeichnungen korrekt.

Bei der Benennung von Bildern werden für alle Sprachen folgende Bereiche aktiviert:

- der Lobus occipitalis (Hinterhauptlappen) und der Gyrus temporalis inferior zur Verarbeitung visueller Informationen und zum Erkennen von Objekten;
- der Hippocampus zum Abrufen von Informationen aus dem Gedächtnis;
- der Gyrus frontalis inferior und die Insula zur Planung der Artikulation und andere frontale Bereiche sowie der Gyrus cingulis anterior zur Steuerung des Sprachvorgangs;
- der Gyrus precentralis, der supplementär-motorische Kortex und das Cerebellum zur Sprachmotorik.

Ansonsten unterscheidet sich die Gehirnaktivität für sehr gute und mittelmäßige Sprachkenntnisse.

Bei Ladinisch, der Muttersprache der Versuchspersonen, wird anders als bei der englischen Testreihe der präfrontale Kortex in der rechten Gehirnhälfte stärker aktiviert sowie der Gyrus frontalis superior und der Gyrus frontalis medius.

Für Italienisch als fließend gesprochene Zweitsprache lässt sich ebenfalls eine verstärkte Aktivität im präfrontalen Kortex und im Gyrus frontalis medius feststellen, dazu innerhalb der rechten Gehirnhälfte in der Insula.

Dagegen werden für Englisch als vierte Fremdsprache auf der linken Hirnhälfte der Gyrus frontalis inferior und im Cerebellum eine verstärkte Aktivität festgestellt.

Die Genauigkeit beim Benennen der Bilder ist für die englischen Begriffe signifikant korreliert mit der Aktivität im rechten präfrontalen Kortex. Bei den italienischen Begriffen war dieser Zusammenhang schwächer und bei den ladinischen Begriffen nicht vorhanden. Die größere Aktivität im Cerebellum beim Aussprechen der englischen Begriffe deutet, wie in anderen Untersuchungen, auf höhere Anforderungen bei der phonetischen Aktivität hin. Der linke Gyrus frontalis inferior, der generell für die Steuerung von Sprachvorgängen zuständig ist, wird bei später erlernten Sprachen stärker aktiviert als bei früher erworbenen, bei denen der Ablauf der Sprachvorgänge optimiert ist.

1.2.3.4 Weitere Faktoren

Die zweitsprachlichen Fähigkeiten bei erwachsenen Sprechern variieren stark, je nach individuellen Fähigkeiten, was auch die folgende Studie belegt.

(Reiterer, Hemmelmann, Rappelsberger & Berger, 2005) vergleicht die Ergebnisse von EEG-Messungen deutscher Muttersprachler, die Englisch als Zweitsprache entweder sehr gut oder nur mittelmäßig erworben hatten. Alle Versuchspersonen hatten im Alter von neun Jahren begonnen Englisch zu lernen; die Gruppe mit sehr guten Sprachkenntnissen besteht aus Anglistik- und anderen Sprachstudenten, deren Englischkenntnisse von einem Dozenten als sehr gut bewertet werden und die Englisch auch in ihrem privaten Umfeld verwenden. Die Gruppe mit mittelmäßigen Kenntnissen umfasst Studenten aus naturwissenschaftlichen Fächern, Mathematik, Psychologie und Betriebswirtschaft, die ihre Sprachkenntnisse nach der Schulzeit nicht weiter ausbauten und sich selbst als nicht sehr talentiert einstufen.

Den Versuchspersonen werden britische, amerikanische und deutschsprachige Nachrichtensendungen vorgeführt und zwar entweder als normale Fernsehsendung mit Bild und Ton, als Radioprogramm ohne Bild oder als Fernsehsendung ohne Ton. Zur Kontrolle werden bei den Fernsehsendungen zwischendurch blaue, bzw. verrauschte Testbilder eingefügt, um den Unterschied zwischen Sprachverständnis und grundlegender Gehirnaktivität festzustellen. Danach füllen die Versuchspersonen einen Fragebogen aus, der sowohl Fragen zum Verständnis des Inhalts als auch zur Aufmerksamkeit, zur Verständlichkeit des Inhalts und dem Interesse an den behandelten Themen enthält.

Die durch das EEG gemessenen Gehirnströme werden über eine Fourier-Transformation in sechs Frequenzbereiche aufgeschlüsselt. Diese liegen für δ zwischen 0 und 3,5 Hertz, für θ zwischen 4 und 7,5 Hertz, für $\alpha 1$ zwischen 8 und 10 Hertz, für $\alpha 2$ zwischen 10,5 und 12,5 Hertz, für $\beta 1$ zwischen 13 und 18 Hertz und für $\beta 2$ zwischen 18,5 und 31 Hertz.

Bei der Auswertung des Fragebogens versteht die Gruppe mit den mittelmäßigen Englischkenntnissen ca. 50% der englischsprachigen Sendungen, die Gruppe mit den sehr guten Kenntnissen 95%. Bei den deutschsprachigen Sendungen liegt das Verständnis der Gruppe mit dem mittelmäßigen Englisch leicht unter dem der sehr guten Gruppe. Ansonsten zeigen sich keine signifikanten Zusammenhänge.

Die Frequenzen im δ -Bereich weisen keine signifikanten Unterschiede auf. Bei den Frequenzen im θ -Bereich zeigt sich nur bei den Fernsehsendungen mit Ton und den Radiosendungen eine stärkere Hirnaktivität in der linken Gehirnhälfte.

Die größten Unterschiede zwischen beiden Gruppen zeigen sich im $\alpha 1$ -Bereich, da bei der mittelmäßigen Gruppe sowohl für die Muttersprache als auch für die Zweitsprache größere Hirnareale aktiviert werden als bei der sehr guten Gruppe. Die Aktivität im $\alpha 2$ -Bereich ist bei der mittelmäßigen Gruppe in der linken Gehirnhälfte stärker ausgeprägt, auch bei den Fernsehsendungen mit abgeschaltetem Ton.

Für die mittelmäßige Gruppe ergibt sich im $\beta 1$ -Bereich eine stärkere Aktivität im linken Lobus temporalis, bei der sehr guten Gruppe war allgemein eine stärkere Gehirnaktivität

festzustellen. Im β 2-Bereich unterscheiden sich die Ergebnisse signifikant für beide Gruppen und alle Darstellungsformen (visuell und akustisch, akustisch, visuell).

Was die Darstellung betraf, so ist im α 1-Bereich die Aktivität bei den Radiosendungen im Lobus frontalis größer und bei den stumm geschalteten Fernsehsendungen im Lobus occipitalis. Der β 2-Bereich wird bei den Fernsehsendungen mit Ton in beiden Gruppen stärker aktiviert.

Die geringere Aktivität im α 1-Bereich bei den sehr guten Sprechern wird mit Bezug auf andere Studien mit einer besseren Leistung durch eine effizientere Informationsverarbeitung erklärt. Im α 2-Bereich, der in anderen Studien mit dem semantischen Gedächtnis in Verbindung gebracht wird, war die Gehirnaktivität bei den sehr guten Studenten ebenfalls geringer als bei der Vergleichsgruppe.

Das bessere Abschneiden der Gruppe mit den sehr guten Englischkenntnissen beim Verständnis deutschsprachiger Nachrichtensendungen wird durch eine insgesamt effizientere und effektive Verarbeitung von Sprache im Gehirn erklärt.

1.2.4 Schlussfolgerung

Bei der Einteilung von Sprechern in Muttersprachler und Zweitsprachler kann man von drei Gruppen ausgehen, was den Beginn des Spracherwerbs betrifft.

Eine oder zwei Muttersprachen können in den ersten drei Lebensjahren erworben werden, wobei die Sprachwahrnehmung im ersten Jahr und die Grundlagen der Sprachproduktion bis ungefähr zum Alter von vier Jahren erworben werden.

Wird die Zweitsprache in diesem Alter erworben, fällt die Sprachwahrnehmung bei Perzeptionstests und neurologischen Messungen teilweise unterschiedlich aus, für weitere Unterschiede finde ich jedoch keine Belege.

Vor der Pubertät kann eine Zweitsprache unter geeigneten Umständen, d.h. bei ausreichender sprachlicher Begabung und ausreichendem Kontakt mit der Zielsprache, auf einem muttersprachlichen Niveau erworben werden, ansonsten können die Ergebnisse jedoch abweichen.

Hier zeigen sich deutlichere Abweichungen bei der Wahrnehmung der Zweitsprache, nicht nur bei einzelnen Phonemen, sondern auch insgesamt bei phonetischen Kategorien. Außerdem werden beim Sprechen der Zweitsprache zusätzliche Bereiche des Gehirns in Anspruch genommen.

Nach der Pubertät unterscheidet sich der Zweitspracherwerb in der Sprachwahrnehmung und -produktion definitiv von der Muttersprache. Die Vorgänge im Gehirn müssen stärker bewusst gesteuert werden und die erworbene Kompetenz unterscheidet sich stark je nach Sprecher, erreicht jedoch nur in Ausnahmefällen das erstsprachliche Niveau.

Insgesamt könnten die Erkenntnisse aus der Neurolinguistik dazu beitragen, die *Critical Period Hypothesis* zu verfeinern. Außerdem ist eine genauere Differenzierung zwischen Zweitsprachwahrnehmung und -produktion möglich. Der steigende Einfluss von Umwelt-

faktoren wie des sozialen Umfelds oder der Gestaltung des Schulunterrichts müsste jedoch noch stärker untersucht werden.

1.3 Zweitspracherwerb

1.3.1 Transfer und interlanguage

Es gibt verschiedene wissenschaftliche Ansätze, um die Unterschiede zwischen der zuerst erworbenen Sprache (L1) und später erlernten Sprachen (L2) zu beschreiben.

(Major, 2001, S. 13) beschreibt einen allgemeinen fremdsprachlichen Akzent als Kombination aus nichtmuttersprachlichen Segmenten, Silben und Intonation. Bei der Beschreibung der nicht-muttersprachlichen Aussprache und des Transfers aus der Muttersprache unterscheidet er folgende Vorgänge:

Ersetzung von Phonemen ein nicht-muttersprachlicher Laut wird durch den nächsten muttersprachlichen Laut ersetzt

phonologische Prozesse z.B. Auslautverhärtung im Deutschen

fehlende Differenzierung die gelernte Sprache besitzt einen Kontrast, den die Muttersprache nicht kennt

übertriebene Differenzierung zwei Laute, die in der Muttersprache einen Kontrast bilden, sind in der gelernten Sprache nur Allophone, werden jedoch weiterhin wie ein Kontrastpaar behandelt

Uminterpretieren von Unterschieden z.B. deutsche Sprecher, die den Kontrast zwischen /i:/ und /ɪ/ im Englischen v.a. als Unterschied der Vokallänge interpretieren und nicht der Vokalqualität (s. S. 32).

Zusätzlich zu diesen Transferprozessen wird die Aussprache noch durch Entwicklungsprozesse beeinflusst, die der Entwicklung der Muttersprache bei einem Kind entsprechen. Die Transferprozesse nehmen mit der Zeit ab, wogegen die Entwicklungsprozesse anfangs vereinzelt auftreten, sich dann steigern und bei fortgeschrittenen Sprechern wieder verringern. (s. S. 48).

Im Unterschied zur Sprachentwicklung beim Erlernen der Muttersprache, bei dem die Zielvokale dem Kind klar sind, jedoch nicht in allen Fällen korrekt produziert werden, kann beim Erlernen einer Fremdsprache der Zielvokal dem Lernenden unklar sein, weil er z.B. nicht korrekt wahrgenommen wurde (s. S. 54).

Das hier entworfene Spracherwerbsmodell bezieht sich jedoch vor allem auf phonotaktische Regeln. Für den Erwerb einzelner Lautsegmente existieren noch weitere Beschreibungen.

(Bohn, 1993) beschreibt drei Verläufe des Vokalerwerbs beim Zweitspracherwerb:

Tabelle 1.1: Verhältnis zwischen Wahrnehmung und Produktion in Erst- und Zweitsprache

Wahrnehmung	Produktion	
	korrekt	nicht korrekt
korrekt	L_1, L_2	L_1, L_2
nicht korrekt	L_2	L_2

(nach (Major, 2001, S. 54))

1. Eine frühe Annäherung an die Standardwerte der Fremdsprache, jedoch keine oder nur unbedeutende Fortschritte.
Dies ist z.B. bei den Vokalen /i/ und /ɪ/ der Fall, die von Englisch lernenden Deutschen korrekt wahrgenommen, jedoch mit den höheren, deutschen Formantwerten ausgesprochen werden.
2. Eine frühe Annäherung an die Standardwerte der Fremdsprache, später jedoch eine Abweichung von diesen Werten in Richtung der Muttersprache.
Dies geschieht z.B. beim Vokal /ɛ/, dessen kürzere Vokaldauer und höhere F_1 -Werte bei fortgeschrittenen Lernenden stärker den deutschen muttersprachlichen Standardwerten entsprechen. Diese Abweichung im Detail entspricht jedoch einer verbesserten allgemeinen Aussprache, da der Vokal /æ/ als eigener Vokal wahrgenommen und produziert wird und im Gegensatz zu weniger erfahrenen Lernenden vom Vokal /ɛ/ unterscheidbar ist.
3. Eine frühe Abweichung von den Standardwerten der Muttersprache und der Fremdsprache und eine darauffolgende Annäherung an die Standardwerte der Fremdsprache.

Die bisher erwähnten Bücher und Aufsätze beziehen sich auf den Verlauf und mögliche Ursachen für eine abweichende Aussprache. Wie lässt sich jedoch die Aussprache eines Fremdsprachenlernenden beurteilen?

(Munro, Derwing & Morton, 2006) untersucht, wie Hörer, die Kantonesisch, Mandarin, Japanisch oder Englisch als Muttersprache sprechen, englischsprachige Äußerungen von kantonesischen, japanischen, polnischen oder spanischen Muttersprachlern beurteilen. Dabei wird zwischen folgenden Begriffen unterschieden:

intelligibility die tatsächlich verstandenen Bestandteile der Äußerung,

comprehensibility die Einschätzung der Verständlichkeit durch den Hörer und

accentedness die Abweichung der Aussprache einer Äußerung von der erwarteten Aussprache (S. 112).

Zu diesen Eigenschaften tragen „stimulus properties“ (SP) bei, d.h. linguistische Merkmale der Äußerung und „listener factors“ (LF), d.h. die Erfahrung des Hörers mit Sprache

und Akzenten. Daher untersucht der Autor, ob die linguistischen Merkmale oder die Erfahrung des Hörers ausschlaggebend für die objektive und subjektive Verständlichkeit der gesprochenen Äußerungen sind. Zählen vor allem die linguistischen Merkmale, müssten die Bewertungen von Hörern mit unterschiedlichen Muttersprachen übereinstimmen. Ist die Erfahrung des Hörers ausschlaggebend, müssten die Bewertungen sich stark unterscheiden (S. 114).

Bei der Auswertung der Ergebnisse waren für die objektive Verständlichkeit vor allem die linguistischen Merkmale maßgeblich. Die Korrelation der Genauigkeit mit der die Sprecher verstanden wurden zwischen den Sprachgruppen lag zwischen 0.600 und 0.893, mit einer Ausnahme bei 0.47 (S. 119-120). Der einzige signifikante Unterschied im Bezug auf die Sprecher fand sich bei den japanischen Muttersprachlern, die japanische Englisch-sprechende genauer als englische Muttersprachler transkribieren konnten (S. 121).

Bei der subjektiven Beurteilung der Verständlichkeit beurteilten die kantonesischen Muttersprachler die kantonesischen Sprecher als am verständlichsten, die japanischen Muttersprachler die japanischen Sprecher. Die Hörer, die Mandarin als Muttersprache hatten, empfanden die kantonesischen und japanischen Muttersprachler als verständlicher als die spanischen Sprecher. Die englischen Muttersprachler empfanden alle vier Sprachgruppen als gleich verständlich (S. 123).

1.3.2 Perzeption

In den folgenden Abschnitten möchte ich zuerst die Wahrnehmung nicht-muttersprachlicher Vokale beschreiben, welche die Grundlage für die Aussprache bildet. Diese lässt sich sowohl durch verschiedene neurologische Verfahren beobachten, wie z.B. EEG oder die Messung der Durchblutung verschiedener kortikaler Areale, als auch durch „klassische“ Perzeptionstests.

1.3.2.1 Einzelne Vokale

(Winkler, 1999) misst die Reaktion von finnischen Muttersprachlern sowie ungarischen Muttersprachlern, die entweder kein oder fließend Finnisch sprechen, auf die Laute /æ/ und /e/, die im Finnischen einen phonetischen Kontrast darstellen, im Ungarischen jedoch nicht. Dabei wird die *mismatch negativity* (MMN) gemessen, eine Reaktion des Gehirns auf Abweichungen in einer Reihe akustischer Stimuli. Bei den finnischen Muttersprachlern und den fließend Finnisch sprechenden ungarischen Muttersprachlern zeigte sich zwischen 130 und 190 ms ein deutlicher Impuls, nachdem in einer Reihe von /e/-Lauten ein /æ/ präsentiert worden war. Bei den ungarischen Muttersprachlern ohne Finnischkenntnisse zeigte sich keine Reaktion (S. 640).

(York, 2008) vergleicht die Wahrnehmung deutscher gerundeter Vokale durch verschiedene Gruppen von Germanistikstudenten (S. 35f):

Tabelle 1.2: Untersuchte Studentengruppen

Studienzeitpunkt	Aufenthalt in Deutschland
1. Semester	kein Aufenthalt
3. Semester	kein Aufenthalt
mindestens 6. Semester	weniger als vier Monate
mindestens 6. Semester	mindestens 16 Monate

Er geht davon aus, dass die Vokale /u/, /ʊ/, /o:/, /ɔ/ in mehr als der Hälfte der Fälle korrekt identifiziert werden, die Vokale /y:/, /ʏ/, /ø:/, /œ/ dagegen in weniger als 50% der Fälle.

Die Testpersonen müssen Silben unterscheiden. Diese enthalten die untersuchten Vokale und bestehen teilweise aus deutschen Worten, teilweise aus erfundenen Worten.

Bei der Auswertung des Perzeptionstests bestand zwischen den Studenten im 1. Semester und den Studenten mit mehr als 16 Monaten Auslandserfahrung ein signifikanter Unterschied. Insgesamt wurden nur die Laute /o:/ und /ɔ/ in mehr als 50% der Fälle korrekt identifiziert. Der Vokal /ʊ/ wurde mit 25,23% am schlechtesten erkannt. Für die Vokale /u:/, /ʏ/, /ø:/ und /œ/ lag die Erkennungsgenauigkeit bei allen Gruppen zwischen 25,3% und 30%. Bei den Vokalen /u:/, /o:/, /ɔ/, /ʏ/ und /œ/ war die Zuordnung der Vokale in den fortgeschrittenen Gruppen deutlich besser als bei den Anfängern (S. 49-52).

1.3.2.2 Vokallänge

Da im Deutschen der Kontrast zwischen langen, gespannten und kurzen, ungespannten Vokalen eine wichtige Rolle spielt, suche ich nach Beispielen für den Erwerb dieses Kontrasts in anderen Sprachen.

(McAllister, 1999) vergleicht die Wahrnehmung und Produktion schwedischer langer und kurzer Vokale durch englische, estnische und spanische Muttersprachler.

Im spanischen Vokalsystem spielt die Vokallänge keine Rolle, im englischen zusammen mit unterschiedlichen spektralen Eigenschaften und das estnische Vokalsystem besitzt sehr komplexe Kontraste zwischen Vokallängen. Bei schwedischen Vokalen spielen sowohl die Vokallänge als auch die spektralen Eigenschaften eine Rolle. Bei hohen und tiefen Vokalen ist das Spektrum wichtiger, bei mittleren Vokalen die Vokaldauer (S. 751).

Die Ergebnisse der estnischen Versuchspersonen lagen sehr nahe bei den Werten der schwedischen Kontrollgruppe. Die von den britischen Sprechern aufgenommenen Vokale wiesen ebenfalls einen deutlichen Unterschied zwischen langen und kurzen Vokalen auf, allerdings fielen die langen Vokale signifikant kürzer aus als die der schwedischen Sprecher (S. 753). Die Ergebnisse der spanischen Sprecher wichen sowohl bei der Produktion als auch bei der Identifikation der Vokale deutlich von den schwedischen Werten ab. Die Vokallänge unterschied sich vor allem bei mittleren, jedoch teilweise auch bei hohen und tiefen Vokalen nicht eindeutig voneinander (S. 754).

(Minegawa-Kawai, Mori & Sato, 2005) untersucht die Wahrnehmung japanischer langer und kurzer Vokale durch japanische und koreanische Muttersprachler. Im Japanischen existieren Mora als phonetische Einheit, die entweder aus einem Konsonant mit Vokal, einem Vokal oder /n/ bestehen. Indem man an einen Konsonant mit Vokal den entsprechenden Vokal anhängt, wird dieser zu einem verlängerten Doppelvokal. Zuerst findet ein Perzeptionsexperiment mit japanischen Sprechern, koreanischen Sprechern mit sehr guten Japanischkenntnissen und koreanischen Sprechern mit geringen Japanischkenntnissen statt. Dabei müssen die Versuchspersonen zwischen [mama], [ma:ma:] und [mama:] unterscheiden. Bei dieser Untersuchung zeigen sich signifikante Unterschiede zwischen den Sprechern mit geringen und mit sehr guten Japanischkenntnissen, jedoch nicht zwischen den Sprechern mit sehr guten Japanischkenntnissen und den Muttersprachlern (S. 1377f).

In einem zweiten Test werden den Testpersonen verschiedene Testworte von /mama/ bis /mama:/ vorgespielt, wobei sich die letzte Silbe bei jedem Wort verlängert. Bei diesem Test werden mit Hilfe der near-infrared spectroscopy (s. 1.2.1.1) Veränderungen in der Durchblutung verschiedener Gehirnbereiche gemessen, sobald die Testpersonen nicht mehr /mama/ sondern /mama:/ hören. Dabei zeigt sich eine verstärkte Durchblutung des Wernicke-Areals auf der linken Hirnhälfte nur bei den japanischen Muttersprachlern (S. 1379f), nicht bei den koreanischen Muttersprachlern.

1.3.3 Artikulation

Für Deutschlernende stehen zwei Punkte bei der Artikulation im Vordergrund: einerseits die Umlaute /ɛ:,ø:,œ,y,ʏ/ und ihre spektralen Eigenschaften und andererseits der Gegensatz zwischen gespannten und ungespannten Vokalen, der in vielen Sprachen der Welt nicht vorhanden ist.

(Flege & Hillenbrand, 1984) vergleicht die Aussprache der französischen Silben *tous* /tu/ und *tu* /ty/ durch amerikanisch-englische und französische Muttersprachler. Dabei existiert der Laut /y/ nur im Französischen, nicht im amerikanischen Englisch; der Laut /u/ dagegen existiert in beiden Sprachen, wird jedoch im Englischen mit einem niedrigeren F_2 -Wert ausgesprochen (S. 709).

In dieser Untersuchung werden französische Muttersprachler mit amerikanischen Muttersprachlern mit weniger Spracherfahrung (sechs Monate in Frankreich) und mehr Spracherfahrung (mehrere Jahre in Frankreich) verglichen. Dabei wurden die Silben als Bestandteile von Sätzen aufgenommen, extrahiert und französischen Muttersprachlerinnen zur Beurteilung vorgespielt. Die von unerfahrenen und erfahreneren Fremdsprachenlernern produzierten /ty/-Silben wurden zu 60% bzw. 75% als korrekt bewertet, bei den /tu/-Silben unterschieden sich die Bewertungen für die unerfahrene und die erfahrene Gruppe stärker, mit 45% bzw. 70% (S. 711f).

Beim Vergleich der akustischen Eigenschaften unterschieden sich die Formantwerte für /y/ nicht signifikant zwischen den französischen und amerikanischen Muttersprachlern.

Die F_2 -Werte für /u/ waren bei den französischen Sprechern niedriger als bei den erfahrenen amerikanischen Sprechern und bei diesen wiederum niedriger als bei den weniger erfahrenen Sprechern und unterschieden sich für alle drei Gruppen signifikant (S. 716). Insgesamt wird /y/ als neue phonetische Kategorie genauer erworben als /u/, das mit einer existierenden phonetischen Kategorie gleichgesetzt wird (S. 717).

(Flege, Bohn & Jang, 1997) untersucht wie Testpersonen, die Deutsch, Spanisch, Mandarin und Koreanisch als Muttersprache sprechen, die Vokale /i/ und /ɪ/ sowie /æ/ und /ɛ/ im Englischen aussprechen bzw. wahrnehmen. Dabei wird zwischen weniger erfahrenen und erfahrenen Teilnehmern unterschieden. Erstere haben weniger als ein Jahr in den USA gelebt, letztere sieben Jahre oder länger (S. 445).

Zuerst wird die Aussprache der Vokale /i/, /ɪ/, /æ/ und /ɛ/ innerhalb von einsilbigen Worten bei allen Testpersonen aufgenommen, wobei dann amerikanische Muttersprachler die Aufnahmen bewerten. Dann müssen die Versuchspersonen synthetisierte Vokale unterscheiden, bei denen sich F_1 und F_2 in jeweils elf Schritten ändern. Dies sind entweder /i/ und /ɪ/ oder /æ/ und /ɛ/ (S. 446).

Vergleicht man innerhalb der Sprachgruppen die Ergebnisse für weniger erfahrene und erfahrene Sprecher, so wirkt sich die Erfahrung nur bei /ɪ/ signifikant aus. Die Sprechergruppen unterscheiden sich hochsignifikant, mit $p = 0.01$, was /i/, /ɪ/ und /ɛ/ betraf und signifikant, mit $p = 0.05$ für /æ/ (S. 448).

In den Aufnahmen von deutschen Muttersprachlern waren /i/ und /ɪ/ deutlich von-

Tabelle 1.3: Bewertung der gesprochenen Vokale

Muttersprache	Erfahrung	Vokal (Genauigkeit in %)			
		/i/	/ɪ/	/ɛ/	/æ/
Englisch	Muttersprachler	100	100	99	100
Deutsch	erfahren	100	100	77	66
	unerfahren	100	99	88	53
Spanisch	erfahren	57	61	99	73
	unerfahren	69	51	91	70
Mandarin	erfahren	84	90	63	77
	unerfahren	80	83	60	58
Koreanisch	erfahren	60	92	81	43
	unerfahren	75	61	81	18

(s. S. 448)

einander zu unterscheiden, während bei den spanischen, chinesischen und koreanischen Sprechern einige /i/-Aufnahmen von den Bewertern als /ɪ/ eingestuft wurden und einige /ɪ/-Aufnahmen als /i/ (S. 450).

In dieser Aufstellung wird deutlich sichtbar, dass der Laut /æ/ weder im Deutschen noch

im Koreanischen existiert, da die Aufnahmen, die diesen Laut enthielten, bei deutschen und koreanischen Sprechern wesentlich seltener als korrekt bewertet wurden.

Bei der Untersuchung der aufgenommenen Silben am Rechner unterschieden sich die Vokallängen von /i/ und /ɪ/ sowie /ɛ/ und /æ/ bei den Muttersprachlern und bei den deutschen Sprechern signifikant, bei den anderen Sprechergruppen jedoch nicht (S. 452). Die Frequenzangaben lassen sich leider nicht direkt mit den Ergebnissen anderer Arbeiten vergleichen, da die Frequenzen von Hertz nach Bark transformiert wurden, und dann ein Wert für die Vokalhöhe (B1-B0) und ein Wert für die horizontale Vokalposition (B2-B1) errechnet wurden.

Für die so gewonnenen Parameter wird der Abstand zu den entsprechenden Werten bei den englischen Muttersprachlern anhand des euklidischen Abstands berechnet. Die Abstände zwischen erfahrenen Nutzern und Muttersprachlern sowie weniger erfahrenen Nutzern und Muttersprachlern unterscheiden sich, jedoch nur für /ɪ/ signifikant mit $p = 0.026$ (S. 454).

Bei den deutschen Sprechern weisen die Vokale folgende Unterschiede auf:

/i/ – /ɪ/ werden von weniger erfahrenen und erfahrenen Sprechern unterschiedlich ausgesprochen;

/ɛ/ – /æ/ weisen bei erfahrenen Sprechern deutlich unterschiedliche Werte für Vokalhöhe und horizontale Vokalposition auf, bei weniger erfahrenen Sprechern jedoch nicht.

Bei den chinesischen Sprechern sind folgende Unterschiede festzustellen (S. 454-455):

/i/ – /ɪ/ besitzen bei erfahrenen Sprechern deutlich unterschiedliche Werte für Vokalhöhe und horizontale Vokalposition, bei weniger erfahrenen Sprechern dagegen nicht;

/ɛ/ – /æ/ unterscheiden sich nur bei erfahrenen Sprechern im Hinblick auf die Vokalhöhe; die horizontale Vokalposition weist weder bei erfahrenen noch bei weniger erfahrenen Sprechern Unterschiede für die untersuchten Vokale auf.

1.4 Elektronischer Vergleich von Sprachsignalen

1.4.1 Untersuchungsgegenstand

Welche Aussprachemerkmale lassen sich überhaupt am PC untersuchen?

Ganze Worte oder Sätze werden mit einem vorgegebenen Sprachmodell verglichen. Das Ergebnis besteht jedoch nur aus den erkannten Worten und der Wahrscheinlichkeit der Übereinstimmung, wobei keine genaueren Angaben möglich sind.

Intonation Der Verlauf der Intonation kann bei Vokalen und stimmhaften Konsonanten angezeigt und zum Einüben von Intonationsverläufen eingesetzt werden.

Konsonanten lassen sich nur indirekt untersuchen, indem z.B. das Frequenzspektrum des Frikativrauschens, die *voice onset time*, also der Zeitpunkt, an dem nach einem Verschluss die Stimmhaftigkeit einsetzt, oder die Formantverläufe am Übergang zu einem Vokal gemessen werden.

Vokale sind akustisch sehr genau untersucht. Hier lassen sich sowohl die Grundfrequenz als auch die Formanten, deren Bandbreite und Intensität messen.

1.4.2 Einzelne Vokale

1.4.2.1 Was sind Formanten und wie werden sie gemessen?

Nach dem Quelle-Filter-Modell, wird im Kehlkopf die Stimme mit der Grundfrequenz und den Harmonischen erzeugt. Der Mund- und Rachenraum wirkt als Filter für dieses Signal, in dem einige Frequenzbereiche gedämpft und andere verstärkt werden. Die verstärkten Frequenzbereiche werden als Formanten bezeichnet, s. (Harrington, 1999, Kap. 3).

Da die Harmonischen als Vielfache der Grundfrequenz und die verstärkten Frequenzbereiche der Formanten nicht immer exakt zusammenfallen sind die Spitzen im Spektrum nicht unbedingt mit den Formantfrequenzen im Zentrum des verstärkten Bereichs identisch. Deshalb wird *linear predictive coding* (LPC) verwendet, um eine Funktion zu berechnen, die die Veränderungen des Spektrums im Mund- und Rachenraum darstellt. Die mit Hilfe dieser Funktion berechneten Höhepunkte bilden die Grundlage für die Bestimmung der Formanten, s. Kap. 8.

1.4.3 Formanten als wahrgenommene phonetische Eigenschaft

Die maschinelle Messbarkeit von Formanten ist für sich noch kein Beweis, dass diese so vom menschlichen Gehirn wahrgenommen werden.

(Kent & Forner, 1979) dokumentiert einen Versuch, in dem Testpersonen synthetisierte Vokale imitieren sollen. Danach wurden für die ersten drei Vokalformanten die Abstände zwischen Werten der synthetisierten und der tatsächlich gesprochenen Vokale gemessen. Dabei werden fünf englische Vokale – /i/, /u/, /ɑ/, /æ/ und /ε/ – und fünf zufällig synthetisierte Vokale verwendet. Zu den Versuchspersonen gehören erwachsene Männer und Frauen, 13-jährige männliche und weibliche Jugendliche und 4-jährige Kinder.

Die Stimuli werden auf der Basis eines Signals erstellt, bei dem die Grundfrequenz von 140 auf 105 Hz abfällt, entsprechen also von männlichen Sprechern geäußerten Vokalen. Daher ist die Übereinstimmung zwischen den synthetischen Vokalen und den Aufnahmen männlicher Erwachsener auch am größten (s. S.210). Bei den fünf englischen Vokalen beträgt die Abweichung bei erwachsenen männlichen Sprechern für F_1 bis zu 60 Hz, für F_2 und F_3 bis zu 100 Hz; bei vierjährigen Kindern das Doppelte (S. 216).

1.4.4 Vokale im Kontext einer Silbe oder intonatorischen Phrase

(Gottfried & Strange, 1980) vergleicht, wie gut Versuchspersonen Vokale identifizieren, die isoliert oder in einer Silbe präsentiert werden. Die Vokale /i,ɪ,e,æ,ɑ,ʌ,u,u/ werden isoliert und umgeben von labialen und velaren Plosiven aufgenommen, wie /bVb/, /pVp/, /kVk/, /kV/, /Vk/, /gVg/, /gV/ und /Vg/. Die Versuchspersonen müssen für jede Aufnahme eine von mehreren Alternativen auf einem Fragebogen ankreuzen (s. S. 1627).

Bei ungespannten Vokalen treten häufiger Fehler auf als bei gespannten. Ansonsten fallen die Ergebnisse für die von labialen Plosiven umgebenen Vokale besser aus als für isolierte Vokale. Bei den velaren Plosiven sind die Ergebnisse für Silben, die einen stimmlosen Konsonant enthielten, ebenfalls besser als bei isolierten Vokalen (s. S. 1629). Bei den Silben /gVg/ und /Vg/ dagegen fallen die Ergebnisse für vordere Vokale schlechter aus als in allen anderen Silben und bei isolierten Vokalen. Bei /gV/ fallen die Ergebnisse nicht besser als in isolierten Vokalen. Der Verfasser vermutet, dass das Fehlen eines phonetischen Kontrasts für /g/ am Wortende eine Rolle spielen könnte, da im Englischen z.B. [bɪg] existiert, jedoch nicht [big], "league"[lɪg] jedoch nicht [lɪg]. [eg] und [ɛk] sind Aussprachevarianten von "egg", die kein Minimalpaar bilden (s. S. 1632).

Da sowohl die Sprecher als auch die Versuchspersonen amerikanisches Englisch als Muttersprache haben, wäre in diesem Fall die Identifikation der Vokale auch durch phonotaktische Regeln beeinflusst.

(Hillenbrand, 2001) untersucht die automatische Klassifikation von isoliert und in CVC-Silben gemessenen Vokalformanten. Dabei werden die Vokale /i,ɪ,ɛ,æ,ɑ,u,u,ʌ/ isoliert, nach /h,b,p,d,t,g,k/ und vor /b,p,d,t,g,k/ gemessen (s. S. 748).

Die Messergebnisse zeigen einen starken Unterschied der Vokaldauer für Silben, die mit stimmhaften bzw. stimmlosen Konsonanten beginnen. Vokale zwischen einem stimmhaften und einem stimmlosen Konsonanten besitzen eine signifikant längere Dauer als Vokale zwischen zwei stimmlosen Konsonanten. Das gleiche gilt für Vokale zwischen zwei stimmhaften Konsonanten, bei denen die Vokaldauer im Vergleich zu den anderen Gruppen am längsten ausfällt und sich signifikant von der Vokaldauer zwischen einem stimmlosen und einem stimmhaften Konsonant unterscheidet (s. S. 752).

Für /u/ liegt F_2 zwischen zwei alveolaren Konsonanten ca. 500 Hz (Männer) bzw. 600 Hz (Frauen) höher als in einem isoliert gemessenen Vokal, die Vokale /u,ʌ,ɑ/ weisen zwischen /d,t/ ebenfalls deutlich höhere F_2 -Werte auf als in isolierten Vokalen.

Zwischen zwei velaren Konsonanten liegt F_2 für /u,u,ʌ,ɑ/ durchgehend 98 Hz (Männer) bzw. 117 Hz (Frauen) höher als in isolierten Vokalen. Die Werte für F_2 in vorderen Vokale /i,ɪ,ɛ,æ/ nehmen zwischen zwei labialen Konsonanten im Vergleich zu isolierten Vokalen für alle Sprecher um 85-100 Hz ab.

Gruppiert man die Vokale nur nach dem Artikulationsort des vorhergehenden oder des nachfolgenden Konsonanten, ohne dass beide Konsonanten identisch sein müssen, so wirkt sich der vorhergehende Konsonant stärker auf die Formantwerte aus als der nachfolgende (s. S. 754).

Die ungespannten vorderen und mittleren Vokale /ɪ, ε, æ/ besitzen zwischen stimmhaften Konsonanten deutlich niedrigere F_1 -Werte als in isolierten Vokalen.

Wird eine Varianzanalyse auf die zwischen Plosiven liegenden Vokale angewendet, so erhält man für F_1 sowohl für den Vokal selbst und für den Artikulationsort und die Stimmhaftigkeit des vorhergehenden Konsonanten signifikante Ergebnisse als auch für den Artikulationsort und die Stimmhaftigkeit des folgenden Konsonanten. Für F_2 erweisen sich der Vokal selbst, der Artikulationsort des vorhergehenden und des nachfolgenden Konsonanten und die Stimmhaftigkeit des vorhergehenden Konsonanten als signifikant (s. S. 755).

Bei F_3 waren die signifikanten Unterschiede auf die Vokale selbst und auf bestimmte vorhergehende Artikulationsorte und Vokale beschränkt. So fielen F_3 -Werte nach velaren Konsonanten signifikant niedriger aus. /i/ nahm nach labialen Konsonanten niedrigere Werte an, nach velaren Konsonanten höhere (s. S. 756).

2 Vorgehensweise

2.1 Verwendetes Datenmaterial

2.1.1 Kiel Corpus of Read Speech

Um mir eine Übersicht über den deutschen Vokalraum und relevante Faktoren zu verschaffen, arbeite ich zuerst mit dem Kiel Corpus of Read Speech. Das Corpus verteilt sich auf zwei Unterverzeichnisse (PH90 und PH92), die jeweils gelesene Kurzgeschichten bzw. Dialoge enthalten. Die Aufnahmen im Verzeichnis PH90 stammen von 50 erwachsenen Sprechern. In der Dokumentation lassen sich nähere Angaben zu Geschlecht, Alter und Herkunftsgegend finden. Das Verzeichnis PH92 enthält Aufnahmen von fünf Sprechern, von denen zwei, *KKO* und *RTD* auch im Verzeichnis PH90 als *K61* und *K62* vertreten sind.

Die Annotation der Sprachdateien, die ich mit Hilfe des EMU linguistic toolkit in ein für Praat geeignetes Format konvertiere, besteht aus vier Ebenen:

Phonetic Hier finde ich die einzeln segmentierten und transkribierten tatsächlich gesprochenen Vokale. Zur Transkription wird die deutsche SAMPA-Version verwendet.

LexAcc Auf dieser Ebene sind Vokale mit primärem oder sekundärem Wortakzent markiert.

Word In dieser Zeile steht der vorgegebene Text mit Groß- und Kleinschreibung. Die Umlaute sind jedoch als *ae*, *oe*, *ue* und *ß* als *ss* wiedergegeben.

Func Diese Ebene kennzeichnet Funktionswörter.

2.1.2 Learning Prosody in a Foreign Language (LeaP)

Das LeaP-Corpus (s. auch (Gut & Bayerl, 2004)) besteht aus deutschen und englischen Sprachaufnahmen von Zweitsprachlern und jeweils einer muttersprachlichen Kontrollgruppe. Die Aufnahmen lassen sich in Wortlisten, gelesene Texte, Nacherzählungen und freie Interviews aufteilen.

Für meine Vergleiche wähle ich gelesene Texte aus der deutschen und der britischen Kontrollgruppe und bei den Deutsch lernenden Studenten aus der britischen, italienischen und chinesischen Gruppe aus. Die Sprachgruppen suche ich so aus, dass jede Gruppe mindestens vier Sprecher umfasst und das Verhältnis zwischen männlichen und weiblichen Teilnehmern ausgeglichen ist.

Das LeaP-Corpus bietet für die britischen Sprecher bis zu acht Annotationsebenen, welche die Grundfrequenz, die Phrasengrenzen, die Silbe, die Intonation, den Vokal, das Wort, das Lemma und die Ergebnisse eines automatischen Part-of-Speech Taggings beinhalten. Die deutschen Aufnahmen können bis zu sechs Annotationsebenen besitzen, von denen jedoch nur vier durchgehend vorhanden sind.

phrase Diese Ebene gibt die Phrasengrenzen an.

syll Die gesprochenen, phonetisch transkribierten Silben.

vowel Diese Ebene enthält konsonantische bzw. vokalische Abschnitte. Dabei werden jedoch mehrere aufeinanderfolgende Konsonanten bzw. Vokale jeweils zu einem Abschnitt zusammengefasst.

words Hier stehen die klein geschriebenen Worte des vorgegebenen Textes. Die für das deutsche spezifischen Sonderzeichen sind wie im Kiel Corpus umgesetzt.

2.1.3 Eigene Annotationen

2.1.3.1 Transkription

Da das Kiel Corpus of Read Speech mit der deutschen SAMPA-Version und das LeaP-Corpus mit X-SAMPA (s. auch (Wells, 1995)) zwei verschiedene Transkriptionsstandards verwenden und das Symbol } sich unter R weder direkt verarbeiten noch auskommentieren lässt, konvertiere ich die phonemischen Annotationen und die phonetische Transkription in den Wortlisten mit Hilfe von Perl-Skripten nach Unicode.

Die Verarbeitung von Unicode-Zeichen ist mit R generell möglich, unter Windows XP lassen sich Unicode-Zeichen in R jedoch nur an der Konsole ausgeben, wenn ich R über Python mit dem Paket rpy aufrufe. Da rpy eigene Datentypen verwendet (Vector, Array), die sich von den üblichen Datentypen in Python unterscheiden, ist dieser Weg jedoch etwas mühsam und ich führe meine statistischen Untersuchungen mit R unter Linux durch. Während die Ausgabe von Unicode-Zeichen auf der Konsole und am Bildschirm unter R perfekt funktioniert, akzeptieren die Plot-Befehle zur Erstellung von Postscript- oder PDF-Grafiken wegen der verwendeten Schriftarten nur eine Untermenge an Zeichen. Deshalb gebe ich die Plotgrafiken als SVG-Dateien aus und konvertiere sie mit Inkscape nach Postscript, so dass ich sie in meine LaTeX-Dokumente einbauen kann.

2.1.3.2 Vorgegebene Vokale

Um die Abweichungen der muttersprachlichen Versuchspersonen und der deutsch lernenden Studenten korrekt beschreiben zu können, bin ich auf die in der Standardaussprache vorgegebenen Vokale angewiesen. Die in den vorliegenden Corpora vorhandene Transkription der tatsächlich gesprochenen Phoneme stimmt jedoch nicht immer mit den Versionen in Aussprachewörterbüchern überein.

Zuerst versuche ich für deutsche Vokale eine automatische Zuordnung von tatsächlich gesprochenen Vokalen zu der in einer separaten Wortliste angegebenen kanonischen Aussprache zu implementieren. Dies ist für die hohen Vokale /i:/, /ɪ/, /u:/ und /ʊ/ sowie für die gespannten mittleren Vokale /e:/ und /o:/ einfacher als für niedrige und ungespannte mittlere Vokale, bei denen die Verwechslungsrate größer ausfällt, obwohl ich die Schwa-Laute wegen ihrer großen Streuung von vorne herein ausgeschlossen habe.

Bei Nicht-Muttersprachlern ist die automatische Zuordnung von gesprochenen zu vorgegebenen Silben bzw. Vokalen jedoch nicht möglich, da hier nicht nur die Vokale, sondern auch die umgebenden Konsonanten teilweise deutlich von der Standardaussprache abweichen.

Deshalb exportiere ich die von mir erfassten Daten in ein Spreadsheet, sortiere sie dort nach vorgegebenem Wort und tatsächlich gesprochener Silbe und annotiere sie manuell. Danach importiere ich das Ergebnis als CSV-Datei in R.

2.1.3.3 Wortbetonung

Im Kiel Corpus ist der primäre bzw. sekundäre Wortakzent für betonte Vokale angegeben, in den deutschen Aufnahmen des LeaP-Corpus ist dies jedoch nur in einzelnen Aufnahmen der Fall. Im Sinne einer eindeutigen Auswertung erstelle ich für alle Daten zwei Spalten, *stress2* in der ich nur zwischen betonten und unbetonten Vokalen unterscheide und *stress3* in der die primäre bzw. sekundäre Wortbetonung angegeben wird.

2.1.3.4 Linguistische Kategorien

Um koartikulatorische Effekte besser bewerten zu können, führe ich jeweils vier Kategorien für die direkt vor und direkt nach dem Vokal gesprochenen Konsonanten ein:

Artikulationsart Diese richtet sich nach den akustischen Eigenschaften des Sprachsignals und beinhaltet Plosive, Frikative, Nasale, Liquide und Pausen.

Passiver Artikulator Hier wird die Stelle angegeben, an der ein Vokal artikuliert wird, also bilabial, labiodental, dental, alveolar, palatal, velar, uvular oder glottal.

Aktiver Artikulator Diese Kategorie verwende ich, um zwischen verschiedenen Artikulationsorganen wie Lippen (labial), Zungenspitze (coronal), Zungenrücken (dorsal) und Kehlkopf (glottal) zu unterscheiden.

Stimmhaftigkeit Hier gebe ich die Stimmhaftigkeit des vorhergehenden bzw. nachfolgenden Konsonanten an.

2.2 Abfolge der Hypothesen

Zuerst werden, wo möglich, Unterschiede bei der Aussprache von Vokalen untersucht, die in beiden Sprachen vorhanden sind. Dann wird untersucht, ob und inwieweit nicht-

muttersprachliche Lernende den im Deutschen vorhandenen Gegensatz zwischen gespannten und ungespannten Vokalen erlernt haben. Danach findet ein Vergleich der Aussprache von Vokalen statt, die in der Muttersprache der Deutschlernenden nicht vorhanden sind. Als letztes werden die relevante Teilbereiche der Vokalräume der Muttersprachler und deutsch lernenden Studenten auf charakteristische Cluster untersucht.

2.3 Verteilungseigenschaften

Sowohl der t-Test als auch die Varianzanalyse als übliche statistische Verfahren, um die Mittelwerte von Datenmengen zu vergleichen, sind parametrische Verfahren, d.h. sie stellen Anforderungen an die Verteilung der Daten. Sind diese erfüllt, so ist die Aussagekraft größer als bei nichtparametrischen Verfahren, die keine Vorbedingungen stellen.

In diesem Unterkapitel beschreibe ich zuerst die notwendigen Mengeneigenschaften und danach das Ausmaß, in dem die entsprechenden Bedingungen für die einzelnen Testverfahren erfüllt sein müssen.

2.3.1 Erstellung der Testmengen

Da verteilungsabhängige Verfahren wie der t-Test und die Varianzanalyse gleich große Vergleichsmengen verlangen, werden nach dem Zufallsprinzip aus der Menge der muttersprachlichen Testpersonen so viele Messwerte ausgewählt, wie Messwerte in der fremdsprachliche Vergleichsgruppe vorhanden sind.

2.3.2 Normalverteilung

Um zwei oder mehrere Testmengen miteinander zu vergleichen, wird zuerst geprüft, ob beide Testmengen normalverteilt sind. Dabei werden die Testmengen mit Hilfe des Kolmogorov-Smirnov-Verfahrens mit idealisierten, normalverteilten Kurven verglichen.

In diesem Fall könnte man auch den Shapiro-Wilks-Test einsetzen, der in R jedoch nur auf Mengen mit 3 bis 5000 Elementen angewendet werden kann.

Nicht-normalverteilte Messwerte bei Vokalformanten haben folgende Ursachen:

akustisch Das Rauschen vorhergehender oder nachfolgender Frikative sowie die Dämpfung von Formantbereichen nach oder vor Nasallauten kann zu Messfehlern führen. Dieses Problem habe ich insoweit behoben, dass ich nur in der Mitte extrahierter Audioabschnitte Messungen vorgenommen habe.

persönlich Da der Mund- und Rachenraum bei jedem Menschen unterschiedliche Dimensionen annimmt, können vor allem bei kleinen Sprechergruppen Abweichungen in der Artikulation eines Sprechers die Kurve verzerren. Deshalb setze ich in meinen Untersuchungen gemischte Modelle ein, bei denen ich den einzelnen Sprecher als Zufallsfaktor mit einbeziehen kann.

koartikulatorisch Folgt auf ein untersuchtes /i:/ in der nächsten Silbe ein /a/, fallen die Formantwerte zentraler aus als für /i:/ in anderen Silben.

Diphthonge und direkt aufeinanderfolgende Vokale habe ich in dieser Untersuchung von vorne herein ausgeschlossen.

vowel undershoot Ein besonderer Fall der Koartikulation ist auch der *vowel undershoot*, bei dem z.B. ein /a/ zwischen zwei alveolaren Konsonanten, z.B. in [tat], vor allem bei erhöhtem Sprechtempo nicht exakt artikuliert wird.

Fremdspracherwerb Bei Sprechern, die Deutsch erst als Erwachsene erlernen, gehen die Formantwerte noch weiter auseinander als bei Muttersprachlern, vor allem, wenn es sich um einen in ihrer Muttersprache unbekanntem Laut wie /y/ handelt, der u.U. als /i:/ oder /u:/ artikuliert wird.

Sind die Daten nicht normalverteilt, so gibt es noch zwei Möglichkeiten, trotzdem eine Varianzanalyse durchzuführen:

1. Überprüfung, ob die anderen zwei Kriterien für die Varianzanalyse, nämlich ähnliche Varianzen und gleiche Gruppengrößen erfüllt sind. Ist nur eine Bedingung verletzt, liefert die Varianzanalyse noch annehmbare Ergebnisse.
2. Die Anpassung der Daten durch eine Transformationsfunktion, z.B. den Logarithmus oder die Quadratwurzel der Werte.

2.3.3 Schiefe und Wölbung

2.3.3.1 Definition

Um herauszufinden,

- mit welcher Funktion man die Varianzhomogenität prüfen kann
- bzw. welche Transformationsfunktion sich bei nicht-normalverteilten Daten anwenden lässt,

verwendet man die Schiefe und die Wölbung der Datenverteilung.

Abbildung 2.1: Zusammenhang von Schiefe, Wölbung und Datenverteilung



2.3.3.2 Transformationsfunktionen

Um aus nicht-normalverteilten Datenmengen vergleichbare normalverteilte Daten zu gewinnen, bieten sich folgende Transformationsfunktionen an, s. (Leech & Onwuegbuzie, 2002, S. 3):

Tabelle 2.1: Übersicht über die Anwendung von Transformationsfunktionen

Abweichung	Transformationsfunktion
stark nach links (positive Schiefe)	logarithmische Transformation
schwach nach links	Quadratwurzel-Transformation
schwach nach rechts (negative Schiefe)	Quadratwurzel der Differenz zwischen dem größten Wert + 1 und dem jeweiligen Wert $\sqrt{(\max(x) + 1 - x)}$
stark nach rechts	Logarithmus der Differenz zwischen dem größten Wert + 1 und dem jeweiligen Wert $\log(\max(x) + 1 - x)$
J-förmige Verteilung	Invertierung $\frac{1}{x}$

Allerdings müssen die verglichenen Datenmengen in die gleiche Richtung von der Normalverteilung abweichen, damit diese Funktionen angewendet werden können.

Außerdem werden auch die transformierten Daten noch einmal auf Normalverteilung und Varianzhomogenität geprüft.

2.3.3.3 Prüfung der Varianzhomogenität

Die Auswahl des geeigneten Verfahrens verläuft nach folgenden Kriterien, s. (Sachs & Hedderich, 2006, S. 421), (Leech & Onwuegbuzie, 2002, S. 10):

Tabelle 2.2: Auswahl des Tests für die Varianzhomogenität

Verteilung	Maßzahl	Größe	
		<10	>= 10
leicht schief	$Schiefte \neq 0$	Cochran	
normalverteilt	$Schiefte \approx 0$ $Wölbung \approx 0$	Hartey, Cochran	Bartlett (mind. 5 Elemente)
flacher als normal	$Wölbung < 0$	Levene	
steiler als normal	$Wölbung > 0$	Cochran	Levene

2.3.4 Varianzhomogenität

Da meine Datenmengen teilweise leicht asymmetrisch ausfallen, verwende ich den **Cochran-Test**, um festzustellen, ob die Varianzen beider Testmengen gleich sind. Dieser Test verlangt gleich große Datenmengen, die ich durch zufällige Entnahme aus den zu untersuchenden Gruppen erhalte.

2.3.5 Vorbedingungen für den t-Test

(Ramsey, 1980) erwähnt voneinander unabhängige Werte aus normalverteilten Wertemengen, deren Varianzen übereinstimmen, als Voraussetzung für eine optimale Aussagekraft des t-Tests. Sind diese Bedingungen nicht erfüllt, so liegt τ , die Wahrscheinlichkeit eines Fehlers erster Ordnung, über α , der Signifikanzgrenze. Damit nimmt die Wahrscheinlichkeit zu, dass ein nicht existierender Unterschied festgestellt wird. Dies ist der Fall, wenn das Verhältnis der Gruppengröße $NR = N_1/N_2$ und das Verhältnis der Varianzen zueinander $\sigma = \sigma_1/\sigma_2$ deutlich von 1 abweichen und die verglichenen Mengen jeweils weniger als 16 Elemente umfassen.

(Gayen, 1949) beschreibt die Auswirkungen auf den t-Test, wenn Testwerte aus nicht-normalverteilten Mengen ausgewählt werden. Er greift dabei vor allem auf die Schiefe und die Wölbung der untersuchten Testmengen zurück.

Dabei sind die Auswirkungen einer flacheren Wölbung geringer als die einer zu steilen Wölbung und verringern sich, je größer die Testmengen ausfallen, vor allem bei mehr als 20 Werten pro Menge.

Die Auswirkungen einer extremen Schiefe werden als gewichtiger eingeschätzt, wobei der Wert für die Schiefe auch quadriert werden kann, wenn man nicht zwischen linkssteilen Kurven (positive Schiefe) und rechtssteilen Kurven (negative Schiefe) unterscheiden will. Bei extrem schiefen Kurven können signifikante Unterschiede übersehen werden, die sich vor allem auf eine Seite der Verteilung auswirken. Dies bedeutet, dass man z.B. bei einer quadrierten Schiefe von 4 und einer Wölbung von 6, die Signifikanzgrenze auf $\alpha = 10,37\%$ korrigieren muss.

(Srivastava, 1958) untersucht die Auswirkungen der Verletzung von Vorbedingungen beim t-Test auf Fehler zweiter Ordnung, d.h. dass tatsächlich existierende Unterschiede statistisch nicht erfasst werden, was die Aussagekraft des Tests reduzieren würde.

Dabei verringert sich die Aussagekraft bei einer schiefen Verteilung (± 0.5) stärker als bei einer zu steilen (> 2.4). Ist die Verteilung zu flach, sind die Auswirkungen der Schiefe noch stärker, ist sie zu steil, wirken sich Schiefe und Steilheit gleichermaßen aus.

(B.L.Welch, 1937) beschreibt, welche Auswirkungen unterschiedliche Varianzen für den Vergleich zweier normalverteilter Testmengen haben könnten. Er geht davon aus, dass unterschiedliche Varianzen dazu führen, dass signifikante Unterschiede zwischen Mittelwerten nicht erkannt werden.

Solange gleich große Mengen verglichen werden, halten sich die Auswertungen unterschied-

licher Varianzen in Grenzen, so dass die tatsächliche Irrtumswahrscheinlichkeit auch bei sehr unterschiedlichen Varianzen höchstens bei $\alpha = 0,065$ liegt.

Unterscheiden sich die Mengen in ihrer Größe, so wird der Unterschied zwischen den Mittelwerten der beiden Mengen unterschätzt, wenn $\sigma_1 < \sigma_2$ und überschätzt, wenn $\sigma_1 > \sigma_2$.

Zusammenfassend läßt sich sagen, dass der t-Test für gleich große Mengen deutlich zuverlässiger funktioniert, als bei deutlich unterschiedlichen. Dies spielt vor allem bei kleineren Datenmengen mit jeweils weniger als 20 Werten eine Rolle.

Extrem asymmetrisch verteilte Datenmengen, die sich oft auch in ihrer Streuung und in ihrer Varianz unterscheiden, können ebenfalls dazu beitragen, dass signifikante Beispiele übersehen werden.

Diese Überlegungen führen dazu, dass ich in meiner Arbeit bei t-Tests gewöhnlich auf nach dem Zufallsprinzip ermittelte gleich große Untergruppen meiner Messwerte zurückgreife, anstatt die z.B. durch eine unterschiedliche Anzahl an Sprechern oder Häufung von Testworten ungleich großen Mengen direkt zu vergleichen.

Da bei Nicht-Muttersprachlern z.B. unbekannte Laute durch unterschiedliche aus der Muttersprache bekannte Laute ersetzt und neue phonetische Kategorien wie z.B. die Vokallänge als phonemische Kategorie teilweise korrekt umgesetzt werden, teilweise jedoch nicht, ist grundsätzlich zu überlegen, ob man lineare Modelle durch Modelle auf der Grundlage anderer Verteilungen wie z.B. der Gamma- oder Weibullverteilung ersetzt, die asymmetrisch verteilte Daten genauer modellieren könnten.

2.3.6 Vorbedingungen für die Varianzanalyse

Für (Eisenhart, 1947) erfüllt die Varianzanalyse zwei Aufgaben: den Vergleich der Mittelwerte für mehrere Untergruppen einer Datenmenge und die Feststellung des Gewichtes von (auch zufälligen) Faktoren über das Ausmaß der Varianz.

Er gibt folgende Voraussetzungen für die Varianzanalyse an, um die Mittelwerte der Gruppen effektiv zu vergleichen:

1. Die Werte einer Gruppe müssen zufällig um ihren Mittelwert verteilt sein, es dürfen also keine Abhängigkeiten der Werte untereinander bestehen.

2. Additivität:

$$m_{ij} = m_{..} + (m_{i.} - m_{..}) + (m_{.j} - m_{..})$$

d.h. der Wert in der Zeile i und Spalte j sollte dem Durchschnitt aller Werte entsprechen, plus der Abweichung des Durchschnitts der Zeile i vom Durchschnitt aller Werte plus der Abweichung des Durchschnitts der Spalte j vom Durchschnitt aller Werte.

Lassen sich die Werte nicht so addieren, so bestehen Wechselwirkungen zwischen den Faktoren, die die Zeilen und die Spalten bestimmen.

3. Die verschiedenen Gruppen besitzen gleiche Varianzen.

4. Die untersuchten Daten sind normalverteilt.

Ist nur die erste Bedingung erfüllt, während die zweite und die dritte nicht erfüllt werden, lässt sich keine effektive Varianzanalyse durchführen. Sind die Mengen normalverteilt, wird die erste Bedingung redundant.

(Cochran, 1947) schließt aus Ergebnissen von Tests mit nicht-normalverteilten Zufallszahlen oder experimentellen Daten, dass in diesem Fall einem Signifikanzniveau von 5% eine tatsächliche Irrtumswahrscheinlichkeit zwischen 4 und 7% entspricht.

Den Anwendern statistischer Tests empfiehlt er vor allem auf zwei Punkte zu achten:

- Die Abwesenheit grober Fehler, z.B. Messfehler, die durch andere Werte ersetzt werden sollten.
- Unterschiedlich große Abweichungen in den verglichenen Gruppen, die in diesem Fall in Untergruppen mit homogenen Abweichungen zerteilt werden sollten.

(Hack, 1958) vergleicht Messergebnisse für Tomatenwurzeln in verschiedenen Feldern. Er kommt zu dem Schluss, dass sich die linksschiefen von der Normalverteilung abweichenden Messergebnisse des vierten Feldes noch mit der Varianzanalyse vergleichen lassen. Die stärker abweichenden Messergebnisse des siebten Feldes, in dem sehr kurze Wurzeln am häufigsten auftreten, führen zu einer wesentlich geringeren Standardabweichung, bei der mögliche signifikante Unterschiede übersehen werden können.

(Gayen, 1950) stellt die tatsächliche Signifikanzgrenze für Datenmengen fest, deren Schiefe und Wölbung deutlich von der Normalverteilung abweichen. Er verwendet dazu ein Integral über die Dichtefunktion der Verteilung.

Für alle flacher gewölbten Verteilungskurven liegt diese Grenze unter $\alpha = 5\%$ für eine entsprechende normalverteilte Datenmenge und sinkt bis zu 0,89% ab, wenn die Daten zusätzlich deutlich schief ausfallen.

Für steilere Verteilungskurven dagegen kann diese Grenze bis 11,47% ansteigen, wobei sie für symmetrische Datenmengen höher liegt als für deutlich schiefe.

(Rogan & Keselman, 1977) untersucht die Auswirkungen unterschiedlicher Varianzen bei gleich großen Testmengen. Sie geht davon aus, dass bei Mengen mit unterschiedlichen Varianzen der Abstand zwischen den Varianzen der jeweiligen Gruppe und der Varianz aller Messwerte unterschiedliche Werte annimmt. Dadurch wird das Signifikanzniveau niedriger angesetzt, als es notwendig wäre, z.B. auf 5%, statt 12%, in einem Fall mit sieben Untergruppen, nur drei Elementen pro Gruppe und Varianzen, die sich um das Zehnfache unterscheiden.

Die Autorin führt Tests mit unterschiedlich großen Mengen an Zufallszahlen durch, wobei bei größeren Testmengen die Auswirkung unterschiedlicher Varianzen abnimmt und die tatsächliche Wahrscheinlichkeit von Fehlern erster Ordnung nur 2-4% über der Signifikanzgrenze liegt.

(Kastenbaum, Hoel & Bowman, 1970) gibt die standardisierte Maximalabweichung

$$\tau = \frac{\mu_{max} - \mu_{min}}{\sigma}$$

für drei bis sechs Gruppen, verschiedene Gruppengrößen, Signifikanzgrenzen und eine unterschiedliche Aussagekraft der Vergleiche an.

Dabei nimmt τ ab, sobald die Anzahl der Gruppen steigt, und der notwendige Mindestabstand liegt für eine Aussagekraft von 70% etwa halb so hoch wie bei einer Aussagekraft von 95%.

Um Gruppen mit einem Abstand von 1, einer Signifikanzgrenze von 5% und einer Aussagekraft von 70% vergleichen zu können, benötigt man bei drei Gruppen eine Gruppengröße von 17 Elementen, bei sechs Gruppen eine Gruppengröße von 27.

Da es sich bei gemischten Modellen um lineare Modelle mit zusätzlichen Zufallsfaktoren handelt, können diese Modelle mit Hilfe einer Varianzanalyse verglichen werden.

Nach meiner Erfahrung lassen sich Abweichungen von der Normalverteilung durch entsprechende Zufallsfaktoren in linearen Modellen, z.B. einzelne Sprecher oder einen unterschiedlichen Sprachkenntnisstand ausgleichen. Sind die Abweichungen jedoch zu stark, erhalte ich auch bei gemischten Modellen keine sinnvollen Ergebnisse.

2.3.7 Vorbedingungen der Kovarianzanalyse

Die Kovarianzanalyse wird hier behandelt, da die dazu eingesetzten linearen Modelle mit mehreren festen Faktoren die Grundlage für gemischte Modelle bilden können.

(Cochran, 1957) gibt für die Kovarianzanalyse folgende Formel an:

$$y_{ij} = \mu + \tau_i + \rho_j + \beta(x_{ij} - x_{..}) + e_{ij}$$

Dabei repräsentiert $\beta(x_{ij} - x_{..})$ die Kovariate mit ihrem Regressionskoeffizienten, μ den bereinigten Mittelwert, τ_i die Auswirkung der i -ten Behandlung, ρ_j das Gewicht des j -ten Blocks und e_{ij} die Residuen.

Er empfiehlt den Einsatz der Kovarianzanalyse, um

- die Genauigkeit bei Experimenten zu erhöhen, bei denen Stichproben zufällig ausgewählt werden;
- eine Verzerrung durch zusätzliche Faktoren zu eliminieren;
- Einblicke in die Eigenschaften von Faktoren zu erhalten;
- Regressionskurven für mehrfache Klassifikationen zu erstellen;
- Daten mit einzelnen fehlenden Werten zu analysieren.

Dabei sollten jedoch auch folgende Vorbedingungen beachtet werden:

- Die Komponenten, die die Behandlung, den jeweiligen Block und die Regressionskurve darstellen, müssen im Modell eine Summe bilden.
- Die Residuen e_{ij} müssen unabhängig voneinander und normal verteilt sein, mit einem Durchschnitt von 0 und der gleichen Varianz.

(Greenberg, 1953) beschreibt die Bedingungen für den Einsatz von Kovariablen in statistischen Untersuchungen, am Beispiel der gemessenen Körpergröße von Kindern aus einer städtischen und einer ländlichen Grundschule.

Er geht von folgenden Voraussetzungen aus:

- einer linearen Regression zwischen abhängiger Variable und kontinuierlicher Kovariable, wobei es sich hier um die Körpergröße von Schülern zweier Schulen handelt und deren Alter in Monaten als Kovariable eingesetzt wird.
- einer ähnlichen Steigung bei den Regressionsverläufen. Würden sich die Veränderungen der Körpergröße an beiden Schulen mit zunehmendem Alter deutlich unterscheiden, z.B. durch ein langsames Wachstum, wäre das Alter als Kovariable ungeeignet.
- homogene Varianzen. Die Streuung der gemessenen Körpergröße sollte bei 10- oder 11-jährigen Schülern nicht wesentlich höher sein als bei 9-jährigen.
- überlappende Wertebereiche für das Alter als Kovariable.
- Die Anzahl der Kovariablen muss unter der Anzahl der Freiheitsgrade liegen.
- ausgeglichene Gruppen, die durch mehrstufiges Sampling erstellt werden können, in dem z.B. in beiden Schulen die Anzahl der männlichen gemessenen Schüler gleich groß ausfällt.

Diese Bedingungen lassen sich so auch auf gemischte Modelle mit mehreren festen Faktoren und zusätzlichen Zufallsfaktoren anwenden. Dabei müsste man dann z.B. bei muttersprachlichen und nicht-muttersprachlichen Studenten jeweils dieselbe Anzahl an betonten und unbetonten Silben auswählen, um die Silbenbetonung als Kovariable einsetzen zu können.

(Glass, Peckham & Sanders, 1972) betrachtet vor allem zwei Verletzungen von Vorbedingungen bei der Kovarianzanalyse als problematisch:

voneinander abhängige Werte In diesem Fall zeigen sich starke Auswirkungen auf die Signifikanzgrenze und die Aussagekraft bei gleich und ungleich großen Mengen.

heterogene Varianzen Bei ungleich großen Mengen haben diese ernsthafte Auswirkungen auf die Signifikanzgrenze.

Dies beschränkt den Einsatz der Kovarianzanalyse auf hierarchisch gesampelte Datenmengen, bei denen nicht nur die Teilmengen für den verglichenen Hauptfaktor, sondern auch für alle relevanten Kovariaten jeweils gleich groß ausfallen.

2.4 Nichtparametrische Tests

2.4.1 Wilcoxon-Rangsummentest

Werden nur **zwei** Gruppen miteinander verglichen, wird zuerst der verteilungsunabhängige Wilcoxon-Rangsummentest auf die untersuchten Datenmengen angewendet. Sind die Gruppen beide normalverteilt und besitzen ähnliche Varianzen, wird zusätzlich ein t-Test durchgeführt. Der Wilcoxon-Rangsummentest besitzt bei normalverteilten Mengen eine Aussagekraft von 95% des t-Tests. Bei nicht-normalverteilten, schiefen Mengen ist die Aussagekraft jedoch 3-4-mal höher.

Dabei werden jeweils **zweiseitige** Tests durchgeführt, da vorher nicht klar ist, ob sich die Datenmengen in genau eine Richtung unterscheiden, oder ob eine Menge breiter gestreut ist als die andere.

Die Tests beziehen sich auf **unabhängige** Stichproben, da die Testpersonen verschiedene Texte lesen und nur Abschnitte, die keine Diphthonge enthalten, mit einbezogen werden können. Deshalb können keine Wertepaare verglichen werden.

Bei abhängigen Stichproben, d.h. wenn die Stichproben beider Gruppen paarweise miteinander verglichen werden, hätte ich den t-Test auch auf Datenmengen durchführen können, die leicht von der Normalverteilung abweichen, solange die Differenzen der Wertepaare normal verteilt gewesen wären, s. (R.H.Baayen, 2008, S. 83f).

2.4.2 H-Test nach Kruskal-Wallis

Vergleicht man **mehr als zwei** Datengruppen, z.B. bei Vokalen zwischen Konsonanten an verschiedenen Artikulationsorten, so werden nicht-normalverteilte Datenmengen mit unterschiedlichen Varianzen mit dem **H-Test nach Kruskal-Wallis** verglichen. Dieser Test besitzt eine Aussagekraft von 95% der Varianzanalyse (s. (Sachs & Hedderich, 2006, S. 442)).

Außerdem existieren noch zwei Varianten, die für meine Untersuchung hier interessante Optionen darstellen: den Vergleich aller Gruppen mit einer Kontrollgruppe und multiple Vergleiche unabhängiger Stichproben miteinander.

2.5 Darstellung der Unterschiede

Erhalte ich z.B. bei einem Wilcoxon-Rangsummentest einen p-Wert, der unter 0.05 liegt, so weiß ich, dass sich die verglichenen Werte signifikant unterscheiden. Damit ist jedoch noch nicht klar, auf welche Art und Weise und aus welchem Grund sich die Messwerte unterscheiden. Diese Informationen benötige ich jedoch, um Vermutungen zur Ursache bestimmter Unterschiede anstellen zu können.

Die hier vorgestellten Verfahren stellen niedrigere Anforderungen an die Testmengen als

die vorher besprochenen parametrischen Verfahren, liefern jedoch im Hinblick auf feste Faktoren eindeutige Ergebnisse.

2.5.1 Dunnett-Tukey-Kramer-Verfahren

Das in (Dunnett, 1980) beschriebene DTK-Verfahren (Dunnett-Tukey-Kramer) dient dazu, Unterschiede zwischen den Mittelwerten für zwei oder mehr Gruppen festzustellen, die sich stark in ihrem Umfang unterscheiden, jedoch ähnliche Varianzen besitzen.

Der Verfasser verwendet für gleich große Gruppen folgende Formel:

$$\bar{y}_i - \bar{y}_j \pm SR_{\alpha,k,v} * \frac{s}{\sqrt{n}}$$

wobei \bar{y}_i den Mittelwert aller Werte beinhaltet, \bar{y}_j den Mittelwert der jeweiligen Gruppe j , $SR_{\alpha,k,v}$ eine angenommene Normalverteilung für die Signifikanzgrenze α , die Gruppenanzahl k und die Freiheitsgrade v , s die Standardabweichung und n die Anzahl der untersuchten Werte.

Zum Vergleich unterschiedlich großer Gruppen wird der letzte Term der Formel, den man auch $s * \sqrt{\frac{1}{n}}$ schreiben könnte, durch einen an die jeweilige Gruppengröße n_j angepassten Faktor ersetzt:

$$\bar{y}_i - \bar{y}_j \pm SR_{\alpha,k,v} * s * \sqrt{\frac{\frac{1}{n_i} + \frac{1}{n_j}}{2}}$$

2.5.2 Regularized Discriminant Analysis

Um herauszufinden, worin sich zwei Gruppen unterscheiden, z.B. bei der Streuung der Messdaten, wird bei normalverteilten Gruppen die Diskriminanzanalyse eingesetzt. Dabei wird hier die *regularized discriminant analysis* verwendet. Diese stellt im Gegensatz zur linearen Diskriminanzanalyse (LDA) oder zur quadratischen Diskriminanzanalyse (QDA), außer der Normalverteilung keine weiteren Anforderungen. Dies bedeutet, dass die Kovarianzen, also die Differenzen zwischen untersuchten Werten und dem Mittelwert einer Gruppe, nicht in allen Gruppen gleich sein müssen wie bei der LDA und dass anders als bei der QDA auch kleinere Datenmengen untersucht werden können (s. auch (Friedman, 1989)). Dies ermöglicht es mir, normalverteilte Mengen mit unterschiedlichen Varianzen zu vergleichen, bei denen z.B. die Werte einer Gruppe weiter gestreut sind als die der anderen.

Die Ergebnisse werden als partimat-Grafiken ausgegeben. Dabei werden die durch die Diskriminanzfunktionen definierten Bereiche als farbige Flächen dargestellt. Die mit Hilfe dieser Funktionen korrekt eingeordneten Werte werden durch schwarze Buchstaben, die falsch eingeordneten Werte durch rote Buchstaben dargestellt.

Lassen sich normalverteilte Formantmessungen klar verschiedenen Bereichen zuordnen,

so sind Diskriminanzfunktionen ein Mittel, um diese Bereiche deutlich darzustellen, z.B. unterschiedlich umgesetzte Vokale oder Formantwerte einzelner Sprecher.

2.5.3 Klassifikationsbäume

Bei grundsätzlich nicht-normalverteilten Gruppen wird ein binärer Klassifikationsbaum erstellt. Dabei bestehen die Blätter aus einzelnen Fällen und die Äste verzweigen sich, wobei für kontinuierliche Parameter zwei Gruppen mit $x_j < t$ und $x_j \geq t$ gebildet werden, mit x_j als untersuchtem Wert und t als Grenze. Um optimale Untergruppen zu erstellen wird der Gini-Index eingesetzt, der angibt, wieviele Fälle auf ein bestimmtes Merkmal kommen, s. auch (Ripley & Venables, 2002, S. 251ff), (Breiman, 2001) und (Segaran, 2007, S. 142ff). Mit Hilfe dieser Bäume können unabhängig von der Verteilung der Testmengen Regeln aufgestellt und auf ihre Genauigkeit überprüft werden, z.B. wenn es um unterschiedliche Untermengen bei der Umsetzung von gespannten und ungespannten Vokalen geht.

Allerdings habe ich in der praktischen Anwendung festgestellt, dass extreme Größenunterschiede zwischen Untermengen zu Verzerrungen führten. Versuche ich z.B. für alle mittleren Vokale einen Klassifikationsbaum zu erstellen, und belasse dabei die /ə/-Abschnitte in der untersuchten Menge, so dominieren sie die Ergebnisse und die Mehrzahl der anderen Vokale werden als /ə/ klassifiziert, was die Fehlerrate erhöht.

2.6 Ausschluss von Zufallsfaktoren

2.6.1 Gemischte Modelle

2.6.1.1 Korrelationstests

Um auszuschließen, dass die Unterschiede zwischen den Datenmengen z.B. auf individuelle Sprechercharakteristika oder auf zufällig gehäuft ausgewählte Worte zurückzuführen sind, wird v.a. bei normalverteilten Mengen versucht, mittels gemischter Modelle Resultate für männliche und weibliche Sprechergruppen zu erzielen, deren Abstand zu den tatsächlich gemessenen Daten möglichst gering ist. In diesen Modellen werden feste Faktoren und Zufallsfaktoren zu linearen Modellen kombiniert.

(Charles R. Henderson, 1982) befürwortet gemischte Modelle, die auch Zufallsfaktoren in Form unterschiedlicher Koeffizienten beinhalten. Er weist darauf hin, dass für feste Kovariablen innerhalb eines linearen Modells die selben Bedingungen gelten, wie für die Hauptfaktoren bei der Varianzanalyse, d.h. die gemessenen Abweichungen sollten normalverteilt sein und ähnliche Varianzen besitzen. Außerdem wird verlangt, dass die Kovariable vom Hauptfaktor linear unabhängig sein muss, d.h. es sollte möglichst keine Korrelation zwischen den festen Faktoren bestehen. Ungleich große Untermengen zeigen dieselben Auswirkungen wie bei der Varianzanalyse.

Deshalb sollte ich darauf achten, dass bei zufällig ausgewählten Wertemengen, nicht eine Ausprägung eines festen Faktors gehäuft in einer Gruppe auftritt, jedoch nicht in der anderen, d.h. keine Testmenge, in der die betonten Silben in der muttersprachlichen Gruppe überwiegen und in der nicht-muttersprachlichen selten sind.

Außerdem kann ich linguistische Merkmale, die stark ungleich verteilt sind, als feste Faktoren ausschließen. Dies gilt z.B. für einzelnen Konsonanten, von denen einige häufiger und andere selten vor bestimmten Vokalen auftreten.

(Robinson, 1991) erläutert den Einsatz von Zufallsfaktoren und die Entscheidung zwischen festen und Zufallsfaktoren. Er hält gemischte Modelle für ideal, um Abweichungen von einem grundsätzlich normal verteilten Modell zu errechnen oder um Unterschiede zwischen Blöcken zu neutralisieren. Er rät dazu, wenn die Auswirkungen bestimmter Kovariablen auf ein Ergebnis berechnet werden, diese Variablen als Zufallsfaktoren zu behandeln. Wenn es jedoch vor allem um die Beurteilung des Hauptfaktors geht, sollten sie als feste Faktoren eingesetzt werden.

Ist vor allem das Ausmaß der Auswirkungen verschiedener koartikulatorischer Bedingungen gefragt ist, wie des Artikulationsart des vorherigen Konsonanten oder der Stimmhaftigkeit des nachfolgenden Konsonanten, so können diese als Zufallsfaktoren behandelt werden könnten. Bin ich jedoch vor allem an den Ergebnissen eines Vergleichs für einen festen Faktor interessiert, sollte ich die Kovariablen als feste Faktoren einbinden, wenn ich z.B. betonte und unbetonte Vokale bei deutschen und italienischen Sprechern vergleiche.

2.6.2 Erstellung eines linearen Modells

In Fällen, in denen ich für die in der Mitte des Vokals gemessenen Formantwerte keine eindeutigen Ergebnisse erhalte, überprüfe ich in kritischen Fällen zusätzlich den ersten und den letzten gemessenen Formantwert innerhalb des extrahierten Abschnitts, die ich jeweils zu einem geordneten Tupel zusammenfasse.

Um geeignete Zufallsfaktoren für statistische Modelle zu finden, führe ich zuerst eine Reihe von Korrelationstests durch. Danach erstelle ich lineare Modelle für die in Frage kommenden Faktorenkombinationen und werte die Resultate mit Hilfe des MCMC-Samplings aus, mit dem ich auch die p-Werte berechne.

Wären meine Daten durchgehend normalverteilt mit homogenen Varianzen, so könnte ich die geeignetesten Faktorkombinationen für die linearen Modelle automatisch berechnen lassen, indem ich nach dem Modell mit dem niedrigsten AIC-Wert suche (s. u.). Ich ziehe hier die BIC-Werte vor, da sie nicht nur für lineare Modelle, sondern auch z.B. für die Genauigkeit modellbasierter Cluster verwendet werden und da übermäßig komplexe Modelle negativer bewertet werden als bei der Berechnung des AIC.

Sobald ich das geeigneteste Modell gefunden habe, gebe ich die t-Werte für die untersuchten Parameter, also die Formanten und die Vokaldauer, sowie die festen Faktoren an. Diese t-Werte bieten einen ungefähren Ausgangspunkt für die Beurteilung der Signifikanz. Bei einem festen Faktor entspricht ein t-Wert knapp über 2,0 einem p-Wert von 0,05, bei

mehreren festen Faktoren verringert sich diese Schwelle.

Zu diesem Zeitpunkt lässt sich außerdem feststellen, in welchem Maße der gewählte Zufallsfaktor die nicht durch feste Faktoren erklärbaren Schwankungen der gemessenen Daten abdeckt.

2.6.3 Bewertung der Modelle

Zur Bewertung der so gewonnenen Modelle werden verschiedene Kriterien herangezogen:

t-Wert Den *t-Wert* eines linearen Modells erhält man, wenn man den Koeffizienten durch den Standardfehler dividiert. Je kleiner der Standardfehler, desto geringer die Wahrscheinlichkeit, dass 0 innerhalb des Konfidenzintervalles liegt und dass die Werte des untersuchten Faktors sich nicht signifikant unterscheiden, s. (R.H.Baayen, 2008, S. 89f).

Signifikanzwert für das MCMC-Sampling Dabei wird ein ermitteltes gemischtes Modell mit einer großen Anzahl an zufällig erzeugten Werten (10000 pro Faktor) für die festen Faktoren und die Zufallsfaktoren aufgerufen. Der zurückgegebene p-Wert sagt aus, ob zwischen diesen Modellen ein signifikanter Unterschied besteht, s. (R.H.Baayen, 2008, S. 248).

Signifikanzwert Varianzanalyse Mehrere gemischte Modelle können mit Hilfe einer Varianzanalyse verglichen werden, um signifikante Unterschiede zwischen ihnen zu finden.

AIC (Akaike Information Criterion) Mit diesem Wert kann man angeben, wie gut das lineare Modell an die existierenden Daten angepasst ist:

$$AIC = -2 * \log(\text{likelihood}) + 2 * \text{Anzahl der Parameter}$$

Wobei bei der *likelihood* die Wahrscheinlichkeit $P(\text{Daten}|\hat{\beta})$ für den Schätzwert $\hat{\beta}$ des Koeffizienten β aus dem linearen Modells $Y = X\beta + \epsilon$ gesucht wird, bei dem der Abstand zwischen den Ergebnissen des Modells und den tatsächlich gemessenen Werten am geringsten ist, s. (Sachs & Hedderich, 2006, S. 575).

BIC (Bayesian Information Criterion) Wie beim AIC (s.o) gilt: Je kleiner der Betrag, desto besser die Anpassung des Modells an die vorhandenen Daten.

$$\begin{aligned} BIC &= -2 * \text{maximized log -likelihood} \\ &+ \ln(\text{Anzahl der untersuchten Werte}) \\ &* \text{Anzahl der Parameter} \end{aligned}$$

Der Unterschied zum AIC liegt darin, dass bei einer größeren Anzahl von untersuchten Werten kompaktere Modelle besser bewertet werden.

Es handelt sich beim BIC um eine Vereinfachung des Bayesschen Theorems, bei dem für das Auftreten einer Eigenschaft die Wahrscheinlichkeit berechnet wird, dass die entsprechende Stichprobe zu einer bestimmten Gruppe gehört, s. (Reinecke & Tarnai, 2008, S. 143). Dies wäre z.B. die Zuordnung eines Grundfrequenzwertes zu einer männlichen oder weiblichen Gruppe.

2.7 Suche nach zusammenhängenden Daten

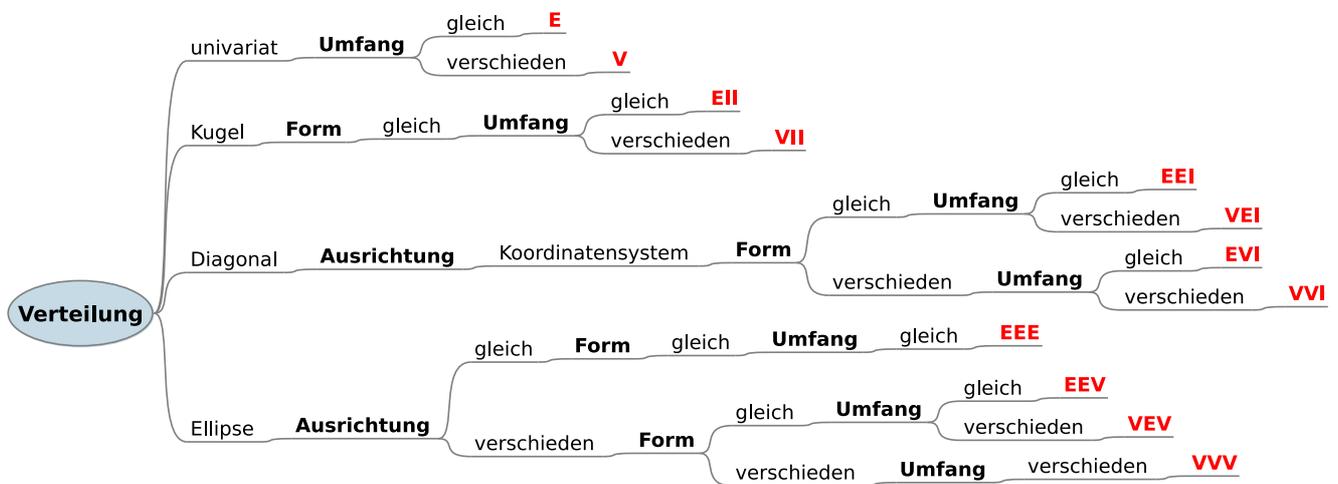
2.7.1 Modellbasiertes Clustering

2.7.1.1 Untersuchung

Um in muttersprachlichen und nicht-muttersprachlichen Datenmengen nach Zusammenhängen zu suchen, z.B. wenn ein deutscher muttersprachlicher Vokal zwei chinesischen Vokalen gegenübersteht, setze ich modellbasierte Clusteringverfahren ein. Der Vorteil dabei ist, dass ich bei diesen Verfahren nicht eine feste Anzahl an gewünschten Clustern oder eine bestehende Einteilung angeben muss, sondern nach zusammenhängenden Wertebereichen gesucht wird.

Dabei wird für eine unterschiedliche Anzahl von Clustern und für verschiedene Clusterformen geprüft, wie genau die so erzeugten Modelle an die tatsächlichen Daten angepasst sind.

Abbildung 2.2: Mögliche Clustertypen in modellbasierten Clustern



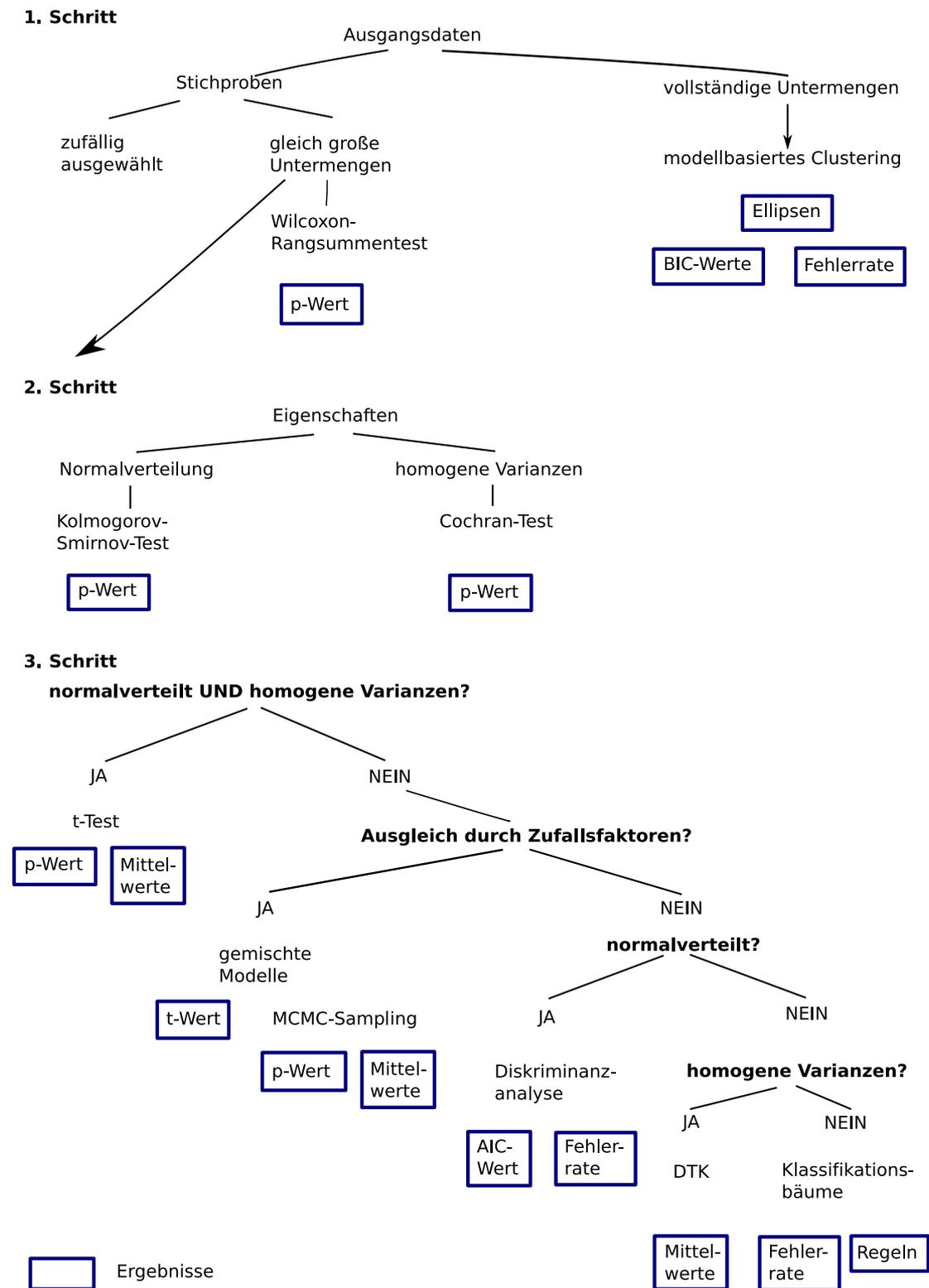
2.7.1.2 Bewertung

Vergleich mit vorgegebenen Kategorien Mit der Funktion `classError()` kann verglichen werden, inwieweit ermittelte Cluster mit vorgegebenen Kategorien übereinstimmen. Dazu müssen allerdings die Anzahl der Cluster und die Anzahl der Kategorien übereinstimmen.

2.8 Zusammenfassung

Um dem Leser einen Überblick darüber zu geben, wie ich die vorgestellten Verfahren anwende, stelle ich meine Entscheidungsfragen als Baumgrafik dar.

Abbildung 2.3: Auswahl statistischer Verfahren



3 Statistische Eigenschaften der muttersprachlichen Aussprache

3.1 Einführung

Für statistische Untersuchungen zur deutschen Aussprache stehen mir das Kiel Corpus of Read Speech zur Verfügung sowie die Aufnahmen der deutschsprachigen Kontrollgruppe des LeaP-Corpus.

Als nächstes untersuche ich, wie sich gespannte Vokale im Kiel Corpus in gelesenen Kurzgeschichten und Dialogen unterscheiden. Damit lässt sich begründen, warum ich mich im LeaP-Corpus nur auf direkt vorgelesene Geschichten beschränkt habe, ohne Nacherzählungen und freie Gespräche miteinzubeziehen.

Um die Entwicklung des deutschen Vokalsystems und seine speziellen Eigenschaften darzustellen und um später auch Vergleiche mit der Entwicklung anderer Sprachen ziehen zu können, gehe ich zuerst auf die sprachgeschichtliche Entwicklung des Deutschen ein.

Dabei gebe ich jedoch zu bedenken, dass es erst seit der Erfindung des Phonographen 1877 durch Edison möglich ist, Sprachschall direkt aufzuzeichnen und dass das internationale phonetische Alphabet (IPA) erst 1888 veröffentlicht wurde. Die Erstellung von Spektrogrammen ist seit 1943 möglich, (s. (Joos, 1948, S. 5)).

Deshalb sind die Forschungsergebnisse zur Phonetik vor dieser Zeit teilweise spekulativ und die entsprechende Literatur manchmal widersprüchlich. Unter diesen Umständen habe ich versucht, ein plausibles Bild aus den Einzelteilen zusammenzufügen.

3.2 Das deutsche Lautsystem

3.2.1 Vorgeschichte

Deutsch gehört zur Sprachfamilie der indoeuropäischen Sprachen, die unter anderem alle germanischen, romanischen und slawischen Sprachen sowie Griechisch, Altpersisch und Sanskrit umfasst (siehe auch (Bussmann, 2002, S. 299f)).

3.2.1.1 Lange und kurze Vokale

Durch Vergleiche mit dem auf dem Gebiet der heutigen Türkei entzifferten Hethitisch entsteht zuerst die Laryngal- und dann die Glottaltheorie. Gemäß der Laryngaltheorie

entwickelt sich der Gegensatz zwischen langen und kurzen Vokalen nach dem Verschwinden der glottalen Konsonanten h_1 , h_2 und h_3 (Bussmann, 2002, S. 390).

In der Glottaltheorie werden nicht mehr glottale Konsonanten, sondern glottalisierte Plosive $/p'/, /t'/, /k'/$ als die Ursache dieser Entwicklung betrachtet (Gamkrelidze, 2008, S. 141).

Ein Beispiel wäre $d^h\bar{e}$ (setzen, stellen, legen), vgl. im Hethitischen *da-a-i* (setzt, legt) bzw. im Altindischen *dád^hāti* (er setzt). Im Lateinischen werden daraus einerseits *abdere* (wegtun, verstecken) und *condere* (gründen, einlegen, bergen, bedecken). Im Althochdeutschen werden aus dieser Wurzel *tōm/tuom* (tun), *teta* (ich tat) und *gitān* (getan) (s. (Pokorny, 1959, S. 235)).

3.2.1.2 Entstehung der Wortbetonung

Die erste oder germanische Lautverschiebung, bei der sich die germanischen Sprachen von der indoeuropäischen Sprachgruppe trennen, führt unter anderem dazu, dass sich der Wortakzent auf die erste bedeutungstragende Silbe des Wortes verlegt und nicht z.B. wie im Lateinischen immer auf die vorletzte Silbe. Diese Veränderung wird auch daran sichtbar, dass $/e/$ und $/a/$ am Wortende in unbetonten Silben wegfallen. (s. (Lehmann, 1961, S. 69f)).

3.2.2 Betonte und unbetonte Silben

(Becker, 2008) erklärt die phonetischen Veränderungen der Vokale vom Althochdeutschen bis heute als Schwächungsprozesse unbetonter Silben bzw. Stärkungsprozesse betonter Silben.

Er zitiert dabei (Neppert, 1999, S.167–171), der folgende Merkmale als für den Wortakzent charakteristisch ansieht: eine Veränderung der Grundfrequenz, eine Steigerung der Intensität, eine verlängerte Dauer der betonten Vokale und ausgeprägtere, weiter von $/ə/$ entfernte Formantwerte. (Becker, 2008, S. 403) betrachtet jedoch die Vokaldauer und die Formantwerte als zuverlässigere Merkmale der Wortbetonung als die Grundfrequenz und die Intensität, da sie direkt mit der Vokalqualität verbunden sind.

3.2.2.1 Abschwächung unbetonter Nebensilben

Hier werden in (Becker, 2008, 14) z.B. der Übergang von *neman* (althochdeutsch) zu *nemen* (mittelhochdeutsch) und schließlich zu *nehmen* angeführt, bei dem der unbetonten Vokal heute nur noch als Schwa realisiert wird.

(Goettsch, 1908, S. 571) führt das schwache Verb *erblīhhan* (althochdeutsch) an, das im Mittelhochdeutschen zu *bleichēn* und heute zu *bleichen* wird.

(Moulton, 1961, S. 29) beschreibt, wie in offenen Silben am Wortende die unbetonten Vokale -i und -e zu -e, sowie -u und -o zu o zusammenfallen.

Innerhalb einer geschlossenen Silbe am Wortende entwickelt sich aus -e-, -a- und -o- sogar schon ein Schwa, wie z.B. in *offan*, das zu *offen* [-ən] oder *sibun*, das zu *siben* [-ən] wird.

(Szulc, 1987, S. 93) beschreibt das Zusammenfallen unbetonter Vorsilben, z.B. *gi-*, *ga-* und *ge-* zu *ge-* oder *fur-* und *fir-* zu *fer-*.

Beispiele für die Ersetzung kurzer Vokale in den Vorsilben durch e (heute /ə/) wären z.B. im Althochdeutschen *gastigan*, das im Mittelhochdeutschen zu *gestigen* wird, oder *gaholfan*, das im Mittelhochdeutschen *geholfen* gesprochen wird, s. (Weinhold, 1919, S. 9).

Unbetonte Vokale in Deklinations- und Konjugationsendungen werden zu /ə/ reduziert. So wird *gesti* im Mittelhochdeutschen zu *geste* (Gäste) und *zungû* zu *zungen* (Zungen), s. (Becker, 2008, S. 403f).

Im Mittelhochdeutschen verschwindet ein schwaches e im Wortstamm vor l oder r, z.B. in *werelt* (althochdeutsch), das zu *werlt* (Welt) wird oder in *dere*, das zu *der* verkürzt wird, s. (McCobb, 1915, S. 39).

Gleichzeitig entfallen auch andere unbetonte Mittelsilbenvokale, z.B. wird *meunisco* zu *mensche* (Mensch) oder *sâlida* zu *sælde*, s. (Weinhold, 1919, S. 17).

Während des Übergangs vom Althochdeutschen zum Mittelhochdeutschen entwickelt sich auch die Auslautverhärtung.

Dabei werden unter anderem die stimmhaften Plosive /b,d,g/ am Wortende und vor stimmlosen Konsonanten zu den stimmlosen Plosiven /p,t,k/, z.B. in *lob* (althochdeutsch), das im Mittelhochdeutschen *lop* gesprochen wird oder in *truobta* (trübte), das zu *truopta* wird, s. (Szulc, 1987, S. 120).

3.2.2.2 Stärkung betonter Silben

Während unbetonte Silben zu einem /ə/ reduziert werden oder wegfallen, verändern sich gleichzeitig zuerst die Struktur des Vokalraums und später die Silbenstruktur.

3.2.2.2.1 Lange Vokale (Vennemann, 1972) beschreibt die Veränderungen bei langen Vokalen im Rahmen der zweiten Lautverschiebung vom Germanischen zum Althochdeutschen. Dabei wird [ai] vor [r,h,w] zuerst zu /ē/ und dann zu /e:/ und [au] vor [h] und allen dentalen Lauten [t,d,s,z,n,r,l] zuerst zu /ō/ und dann zu /o:/.

Dabei handelt es sich jedoch nicht nur um zeitliche Veränderungen in der Aussprache, sondern auch um Strukturen innerhalb der Verbdeklinaton, z.B. bei *rītu* (ich reite) – *reit* (ich ritt) oder *friusu* (ich friere) *frōs* (ich fror), s. (Vennemann, 1972, S. 864).

Da /ē/ und /ō/ nahe bei /ē/ und /ō/ liegen, findet eine Diphthongisierung von /ē/ und /ō/ statt. Dabei wird /ē/ zuerst als /eə/ und dann als /iə/ gesprochen sowie /ō/ zuerst als /oə/ und dann als /uə/. Diese Verschiebung des Silbenkerns von /iə/ und /uə/ ermöglicht den Übergang von /ē/ zu /ē/ und von /ō/ zu /ō/ (s. (Vennemann, 1972, S. 871)).

Während des Übergangs vom Mittelhochdeutschen zum Neuhochdeutschen findet eine Monophthongierung der Diphthonge *ie* zu *i*, *uo* zu *u* und *üe* zu *ü* statt. So wird z.B. aus *guot* im Neuhochdeutschen *gut* und aus *Güete* *Güte*, s. (Weinhold, 1919, S. 13).

Andererseits fallen die Langvokale \hat{i} , \hat{u} und \hat{u} jeweils mit den Diphthongen ei , eu und au zusammen, *wiz* (weiß) und *weiz* ([ich] weiß) werden beide zu *weiß* (s. (Moulton, 1961, S 33)).

Im Gegensatz dazu bleibt im Jiddischen der Kontrast erhalten, so dass z.B. noch immer zwischen *vais* (weiß) und *ix vé's* (ich weiß) unterschieden wird (s. (Sapir, 1915, S. 234)).

3.2.2.2 Änderungen in der Silbenstruktur Was die Vokallänge betrifft, so werden im Neuhochdeutschen Kurzvokale in offener Silbe gedehnt, z.B. in *gebo-ten*, *Ha-se* oder *le-ben*. Dies ereignet sich zuerst im Niederdeutschen im 12. Jahrhundert, parallel zu Entwicklungen in anderen germanischen Sprachen, wie z.B. dem Dänischen. Später ändert sich auch die Aussprache des Mittel- bzw. im 14. Jahrhundert des Oberdeutschen. Vor Doppelkonsonanten findet keine Dehnung statt, deshalb wird z.B. das althochdeutsche *waskan* zu *wasschen* oder *hlahhan* zu *lachen*, s. (Szulc, 1987, S. 133f).

Andererseits werden einige mittelhochdeutsche Langvokale in geschlossener Silbe gekürzt, z.B. in *Jammer*, *Hochzeit*, *Licht*, *Mütter* oder *Futter*.

Durch diese Änderungen verschieben sich die Silbengrenzen, während das Phoneminventar gleich bleibt, s. (Moulton, 1961, S. 33f).

3.2.3 Geschichte der Umlaute

3.2.3.1 Einführung der Umlaute

Die ersten Umlaute entstehen während der zweiten Lautverschiebung vom Germanischen zum Althochdeutschen. Dabei findet ein kombinatorischer Lautwandel im Westgermanischen statt, die sogenannte Brechung, bei der sich das Althochdeutsche in der Entwicklung seines Vokalsystems z.B. vom Gotischen abspaltet.

Dabei wird vor /i,j,u/ und nasalen Konsonanten /e/ zu /ɪ/ und /i/ zu /ɛ/, s. (Twaddell, 1948, S. 148).

Beispiele wären z.B. das germanische *libān*, das im Althochdeutschen als *leben* gesprochen wird oder das germanische *nista*, das als *nest* in die althochdeutsche Sprache eingeht. Außerdem gibt es z. B. *fehu* im Germanischen, das zu *fihu* (Vieh) im Althochdeutschen wird, *seḅun*, das sich im Althochdeutschen zu *sibun* (sieben) entwickelt, s. (Twaddell, 1948, S. 145).

Laut (Szulc, 1987, S. 48) findet nach der Umwandlung von /e/ zu /ɪ/ vor hohen Vokalen und nasalen Konsonanten die Entstehung des Umlauts /æ/ statt, während sich die Umwandlung von /i/ zu /ɛ/ vor hohen Vokalen und nasalen Konsonanten erst später vollzieht.

(Moulton, 1961, S. 21) beschreibt, wie sich aus palatalisierten Varianten die Allophone /æ/ /œ/ und /y/ entwickeln, wenn nach /a/, /o/ oder /u/ in der nächsten Silbe ein i-Umlaut folgt, also /i/, /i:/ oder /j/, was sich im Althochdeutschen jedoch noch nicht in der Schreibung niederschlägt. Dies gilt für Worte wie z.B. *slagi* [slægi].

3.2.3.2 Verschiebungen im Vokalsystem

Im Übergang zum neuhochdeutschen Vokalsystem finden sowohl eine Senkung der hohen Vokale und der Diphthonge als auch eine Hebung der tiefen Vokale statt.

- Die Vokale werden z. B. bei *sunne* (Sonne) *wanen* (wohnen) oder *gewan* (gewohnt) gesenkt, s. (Weinhold, 1919, S. 22).
- Bei den Diphthongen wird in *klein* /aɪ/ statt /ei/ gesprochen. Die Senkung betrifft auch jeweils die ersten Vokale bei den Diphthongen in *ouge* (Auge) oder *erloben* (erlauben), s. (Weinhold, 1919, S. 27).
- Eine Hebung der Vokale findet bei \bar{a} und *a* statt, so z.B. in *māhe* (Mohn), *slâ* (Schlot), s. (Weinhold, 1919, S. 24).

3.2.3.3 Von /æ/ zu /ɛ/

Während des Übergangs vom Mittelhochdeutschen zum Neuhochdeutschen wird in der Mitteldeutschen Hebung /æ/ durch /ɛ/ ersetzt, z.B. in *geslähte*, das sich zu *Geschlecht* entwickelt (Moulton, 1961, S. 31f).

3.2.3.4 Rundung und Entrundung von Vokalen

Im 15. Jahrhundert wird zuerst im Oberdeutschen, dann auch in anderen Varietäten des Deutschen /ɛ/ zu /œ/ gerundet, z.B. in *schepfer* (Schöpfer), *zwelf* (zwölf) oder *leschen* (löschen). Für andere Vokale findet eine Entrundung statt, z.B. in *küssen* (Kissen), s. (Szulc, 1987, S. 128).

3.2.4 Entwicklung von Varietäten

3.2.4.1 Isoglossen

Nach der Gründung des fränkischen Reiches beginnt sich im Rahmen der zweiten Lautverschiebung eine regional stark gegliederte deutsche Volkssprache zu entwickeln. Dabei entsteht die Teilung in zwei bis heute deutlich getrennte Sprachbereiche (Isoglossen): das Niederdeutsche nördlich der „maken-machen“-Linie und das südlich davon gelegene Althochdeutsche. Das althochdeutsche Gebiet unterteilt sich wieder an der Speyrer Linie in das nördlich gesprochene Mitteldeutsch und das südliche Oberdeutsch, das z.B. das Alemannische und das Bayrische mit umfasst (s. (Szulc, 1987, S. 76)).

3.2.4.2 Entwicklung der Standardsprache

Ab Anfang des 14. Jahrhunderts werden kaiserliche Dokumente nicht mehr auf Latein, sondern auf Deutsch abgefasst. So entwickeln sich zuerst die Prager und die Wiener Kanzleisprache als überregionale, süddeutsch geprägte Schreibweisen. Im 16. Jahrhundert konkurrieren verschiedene „Druckersprachen“ miteinander, einerseits das „Gemeine Deutsch“

in katholischen Gegenden, andererseits die kursächsische Kanzleisprache, das „Meißnische Deutsch“, das von Luther für seine Bibelübersetzung verwendet wird. Das Meißnische Deutsch setzt sich durch, wird von Schriftstellern wie Herder, Schiller und Goethe verwendet und im 18. Jahrhundert vom Wiener Hof übernommen.

Zu dieser Zeit gilt noch die meißnisch-obersächsische Aussprache als ideal. Als nach der Einigung Deutschlands im 19. Jahrhundert Preußen politisch eine zentrale Rolle spielt und die Anzahl der niederdeutschen Sprecher zunimmt, entwickelt sich eine niederdeutsche Leseaussprache, die als lautrein betrachtet wird. Diese Aussprache etabliert sich im Laufe des 19. Jahrhunderts als Bühnenaussprache und bildet die Grundlage für Siebs' „Deutsche Bühnenaussprache“, die 1898 herausgegeben wird.

Das heutige Deutsch ist insofern ein Kompromiss, bei dem eine hochdeutsche Schreibung niederdeutsch ausgesprochen wird (s. (Szulc, 1987, S.142–144)).

3.2.5 Die deutsche Standardausprache heute

3.2.5.1 Phoneme des Deutschen

(Fiukowski, 2002, S. 10f) führt folgende Vokale auf: /i/, /ɪ/, /e/, /ɛ/, /ɛ:/, /ə/, /ɑ:/, /a/, /o/, /ɔ/, /u/, /ʊ/, /ø/, /œ/, /y/, /ʏ/ sowie die Diphthonge /æ/, /ao/ und /ɔø/.

Im Duden-Aussprachewörterbuch (Mangold & Dudenredaktion, 2005, S. 35ff) werden die im Deutschen gesprochenen Vokale in folgende Gruppen eingeteilt:

orale Monophthonge , zu denen in dieser Darstellung auch die Vokale der „Gymnasialausprache“ lateinischer und griechischer Fremdwörter gehören, also /i/, /e/, /y/, /u/ und /o/.

nasale Monophthonge , v.a. in Fremdwörtern, die aus dem Französischen übernommen wurden.

Diphthonge , d.h. /ai/, /au/ und /ɔy/.

englische Diphthonge , d.h. /ei/ und /ou/.

Die oralen Monophthonge werden folgendermaßen angeordnet:

Tabelle 3.1: Übersicht zu den oralen Monophthongen der deutschen Sprache

	vorn	zentral	hinten
geschlossen	/i:/, /y:/		/u:/
	/ɪ/, /ʏ/		/ʊ/
halbgeschlossen	/e:/, /ø:/		/o:/
	(/ɛ:/)	/ə/	
halboffen	/ɛ:/, /ɛ/, /œ/		ɔ
		/ɐ/	
offen		/a:/, /a/	

(s. (Mangold & Dudenredaktion, 2005, S. 37))

Vergleiche ich diese beiden Darstellungen, so treten sowohl bei den Diphthongen als auch bei einzelnen Monophthongen Unterschiede auf.

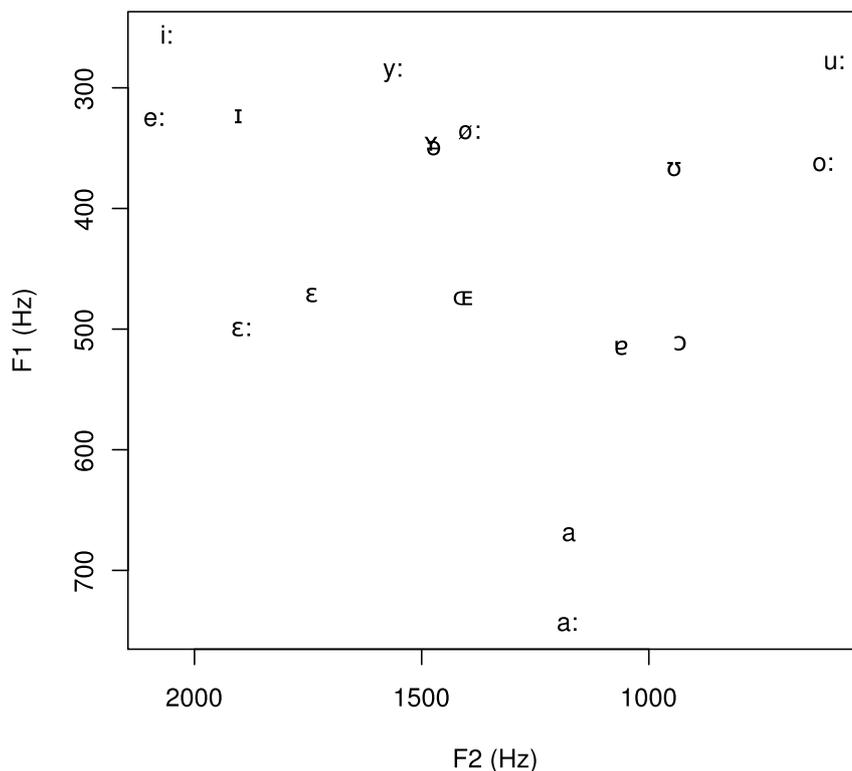
Die Vokale der Diphthonge liegen im Duden weiter auseinander als bei Fiukowski, wobei der zweite Vokal jeweils geschlossen und nicht halboffen transkribiert wird, also /au/ statt /ao/, /ai/ statt /æ/ und /ɔy/ statt /ɔø/.

Außerdem bietet der Duden zwei Varianten für /ɛ:/, halboffen oder zwischen halboffen und halbgeschlossen.

3.2.5.2 Referenzwerte für die Vokalformanten

3.2.5.2.1 Formantwerte für männliche Sprecher 1988 Suche ich nach Referenzwerten für deutsche Vokalformanten so findet man bei (Ramers, 1988, S. 189-192) Messergebnisse für vier männliche Sprecher, die Wortlisten lesen, deren Elemente sich nur im Hinblick den ersten Vokal unterscheiden, wie z.B. /bi:ten/, /bɛ:ten/, /ba:ten/ oder /buten/.

Abbildung 3.1: Deutsche Vokalwerte bei männlichen Sprechern in einzelnen Worten



(berechnete Durchschnittswerte nach (Ramers, 1988, S. 189-192))

Vergleiche ich die hier angegebenen Werte mit den im Kiel Corpus gemessenen Formanten so fällt eine noch größere Streuung der Formantwerte auf, die sich vor allem in höheren F_1 -Werten für /a/ und /a:/ sowie in niedrigeren F_2 -Werten für /u:/ und /o:/ zeigt.

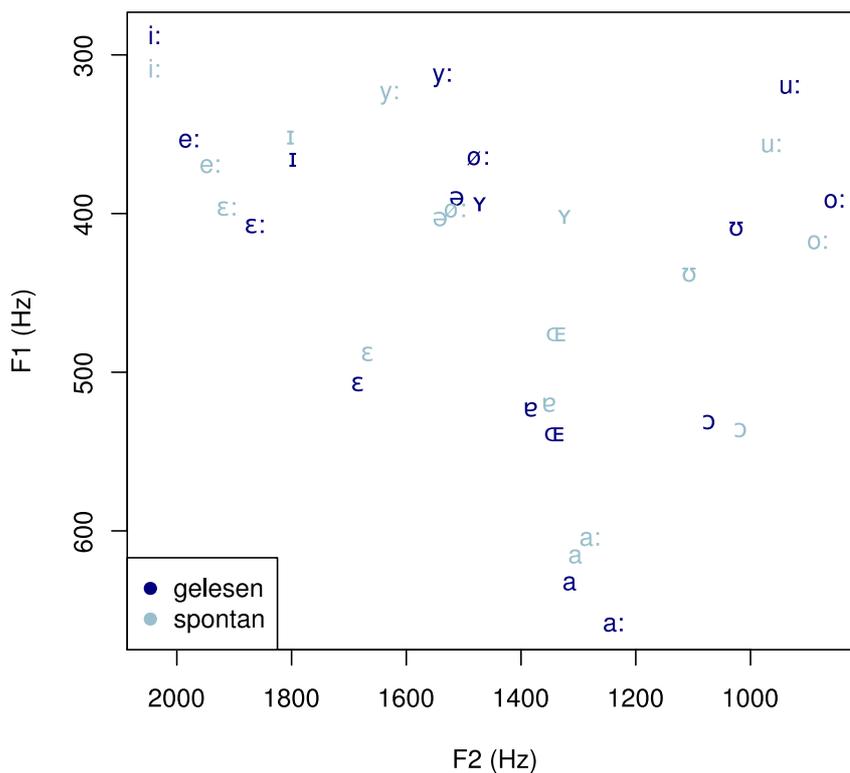
Dies spricht für eine noch ausgeprägtere Artikulation in Wortlisten als in gelesenen Texten, die sich hier in einer weiteren Öffnung des Kiefers und einer stärkeren Bewegung des Styloglossus äußern würde, s. S. 70.

Interessanterweise unterscheiden sich hier /a/ und /a:/ deutlich im Hinblick auf die F_1 -Werte, wobei sich für /a:/ durch die längere Sprechdauer der Kiefer weiter öffnen kann. Es existiert jedoch kein Unterschied im Hinblick auf die F_2 -Werte.

/ɛ:/ und /ɛ/, die z.B. von (Fiukowski, 2002, S. 365) als ein Vokal mit unterschiedlichen Längen betrachtet werden, z. B. in *Kälte* und *Kelte* weisen hier unterschiedliche F_2 -Werte und leicht verschiedene F_1 -Werte auf.

3.2.5.2.2 Medianwerte für männliche Sprecher im Kiel Corpus In der Dokumentation des Kiel Corpus ((Simpson, 1988)) finde ich Medianwerte, d.h. die häufigsten Werte in der Mitte einer geordneten Verteilung, für die gelesenen Texte und für weitere drei CDs mit Aufnahmen freier Unterhaltungen. Dies zeigt, dass der Verfasser nicht von normalverteilten Werten ausgeht, da sonst die Durchschnittswerte und die Standardabweichung ausreichend gewesen wären.

Abbildung 3.2: F_1 - und F_2 -Werte von männlichen Sprechern in gelesenen Texten und spontanen Unterhaltungen



(nach (Simpson, 1988, S. 216f))

Was hier sofort sichtbar wird, ist die größere Streuung der Formantwerte in gelesenen Texten im Vergleich zu spontan gesprochenen Sprache.

Für F_1 ergeben sich zwei Gruppen, die mittleren und niedrigen Vokalen $/\varepsilon/$, $/\text{œ}/$, $/a/$, $/a:/$ und $/\text{ɔ}/$, die sich durch höhere F_1 -Werte in gelesenen Vokalen auszeichnen und andere mittleren und hohe Vokalen, bei denen die F_1 -Werte für gelesene Vokale niedriger ausfallen.

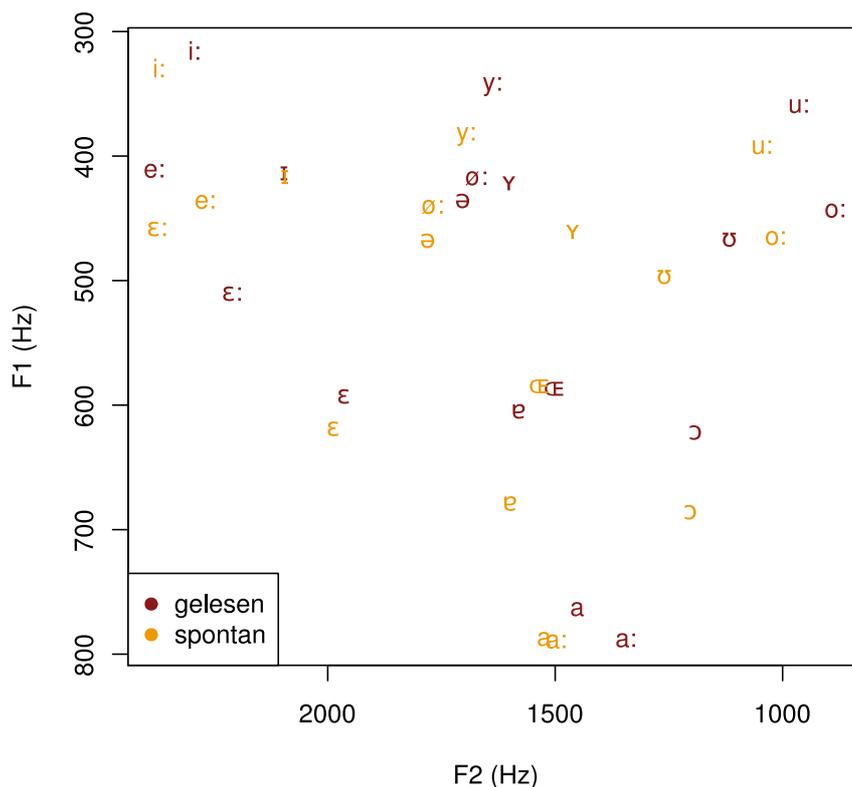
Was F_2 betrifft, ließe sich eine Grenze zwischen $/\text{œ}/$ und $/a/$ ziehen, so dass die Vokale links davon sich durch höhere F_2 -Werte in gelesenen Abschnitten auszeichnen, die Vokale rechts davon durch niedrigere F_2 -Werte. Eine Ausnahme bildet $/\text{ɔ}/$, dies könnte jedoch zur Abgrenzung von $/u/$ dienen, das einen ähnlichen F_2 -Wert besitzt.

In den gelesenen Aufnahmen sind $/a/$ und $/a:/$ auch im Hinblick auf ihre Vokalqualität deutlich als zwei getrennte Phoneme erkennbar, während ihre Werte in den spontansprachlichen Messdaten zusammenfallen. Dies könnte ein Hinweis darauf sein, dass diese Vokale

beim Vorlesen hyperartikulierte werden.

/ɛ:/ und /ɛ/ sind noch deutlicher voneinander getrennt als in den Messungen in (Ramers, 1988, S. 189f). Dafür nähert sich /ɛ:/ in frei gesprochenen Texten dem /e:/ an. Diese Tendenz ist in den Messungen von 1988 nicht sichtbar.

Abbildung 3.3: F_1 - und F_2 -Werte von weiblichen Sprechern in gelesenen Texten und spontanen Unterhaltungen



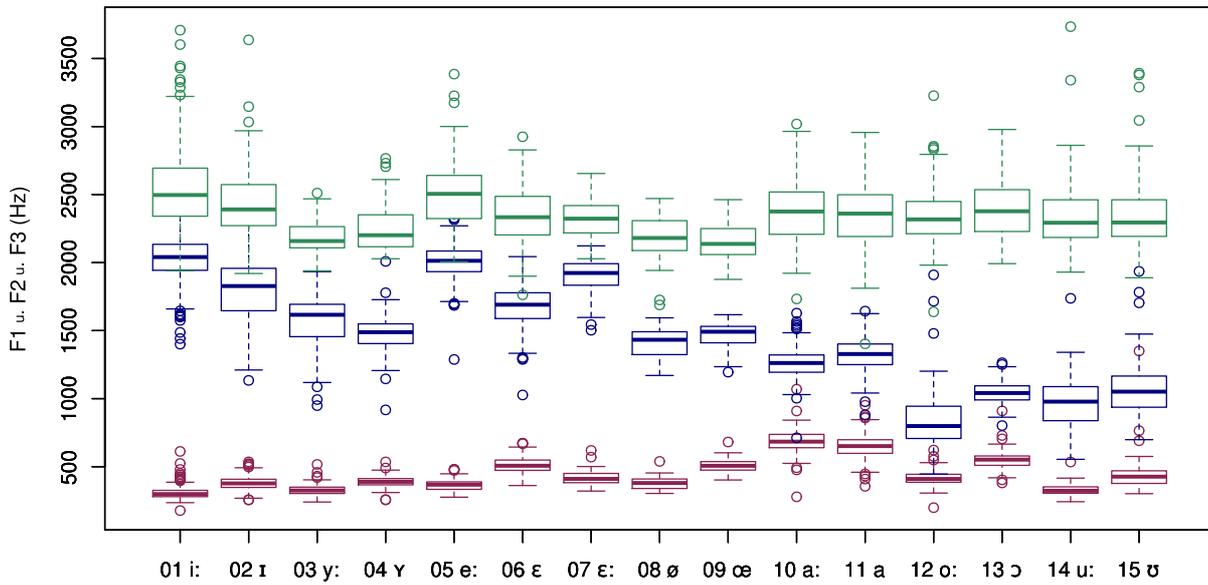
(nach (Simpson, 1988, S. 214f))

3.2.5.2.3 Medianwerte für männliche Sprecher im Kiel Corpus Wie zu erwarten fallen die maximalen F_1 - und die minimalen F_2 -Werte hier für Frauen höher aus als bei männlichen Sprechern, was mit der Länge des weiblichen Vokaltrakts zusammenhängt. Ansonsten fällt auf, dass /a:/ und /a/ hier in freien Gesprächen zusammenfallen, während der Abstand dieser Phoneme zu /ɐ/ und /ɔ/ größer ausfällt als in der männlichen Gruppe. /e:/ und /ɛ:/ sind weiter voneinander entfernt als bei den männlichen Sprechern und bleiben auch in spontanen Äußerungen unterscheidbar.

3.2.5.3 Eigene Messungen

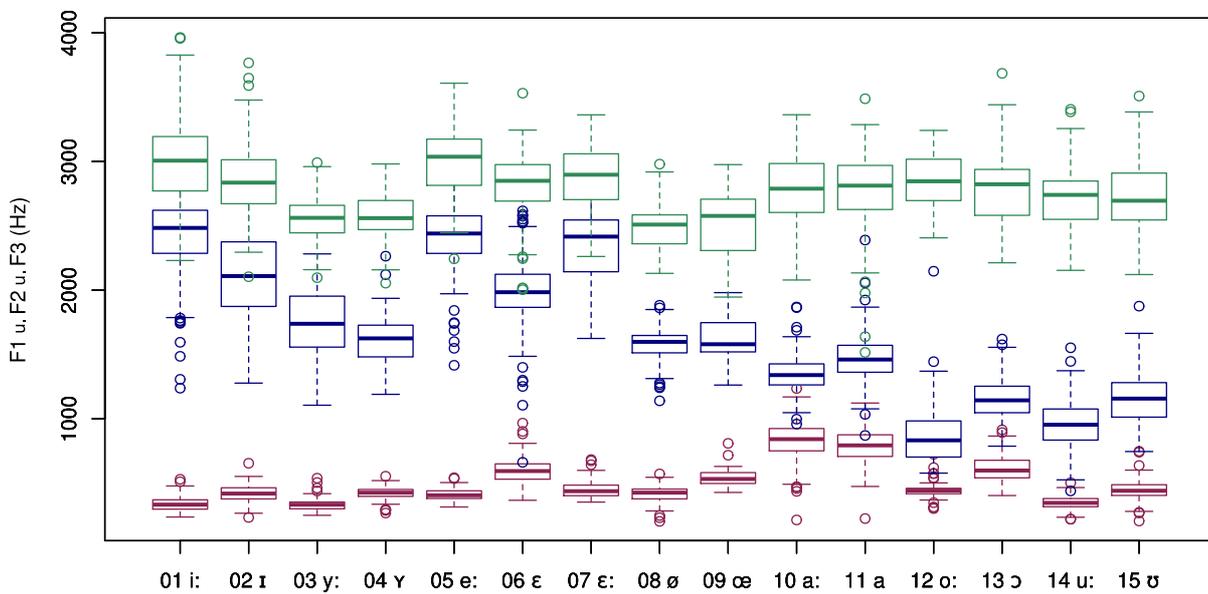
Die Boxplots auf der nächsten Seite zeigen die Messergebnisse für die ersten drei Formantwerte bei deutschen Muttersprachlern, getrennt nach Geschlecht. Dabei erstreckt sich der Bereich des Rechtecks von der ersten bis zur dritten Quartile der Messwerte, d.h. er umfasst die mittleren 50% der Werte, die der Größe nach sortiert wurden. Der breite Balken in der Mitte des Rechtecks zeigt den Mittelwert an und die Linien umfassen Wer-

Abbildung 3.4: Formantwerte aller deutschen Vokale (männliche Sprecher)



te, die nicht weiter von den Quartilen entfernt sind als das Eineinhalbfache des Abstands zwischen den Quartilen. Ausreißer außerhalb dieses Bereichs werden durch kleine Kreise dargestellt.

Abbildung 3.5: Formantwerte aller deutschen Vokale (weibliche Sprecher)



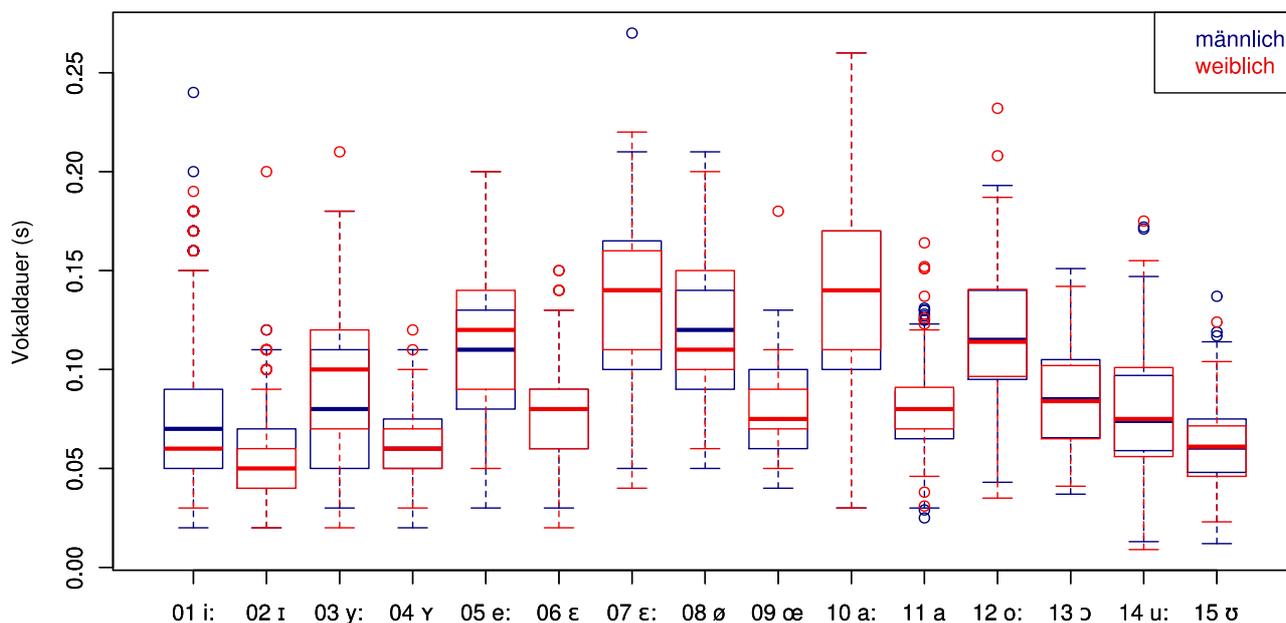
Sieht man sich die Formantmessungen der Vokale genauer an, so werden Eigenschaften deutlich, die in allen späteren Untersuchungen dieser Arbeit relevant bleiben. Zumindest bei den vorderen Vokalen /i:/ und /ɪ/, /y:/ und /ʏ/ und /e:/ und /ɛ/ nehmen die gespannten Vokale extremere, d.h. für F_1 geringere, für F_2 und F_3 höhere Werte an als die entspannten ungespannten Vokale. Ebenso besitzen die ungerundeten Vokale /i:/, /ɪ/, /e:/ und /ɛ/ höhere Werte für F_2 und F_3 als die gerundeten Vokale /y:/, /ʏ/, /ø:/ und /œ/.

Die Abstände zwischen den Quartilen und dem Mittelwert sind nicht bei allen Vokalen gleich, wobei bei den Vokalen /y:/, /ʏ/, /ø:/ und /œ/ die im Vergleich zu anderen Vokalen geringe Anzahl an untersuchten Abschnitten (<100) die Verteilung verzerrt.

Bei den weiblichen Sprechern fallen die im Vergleich zu den männlichen Sprechern breiteren Wertebereiche für F_2 und F_3 sowie die grössere Menge an Ausreißern für F_2 auf.

3.2.6 Vokaldauer der untersuchten Vokale

Abbildung 3.6: Vokaldauern aller untersuchten deutschen Vokale



Die gemessene Vokaldauer wird analog zu den vorher dargestellten Formantwerten nach männlichen und weiblichen Sprechern getrennt dargestellt. Dabei überschneiden sich die Werte in vielen Fällen - die kürzeren Vokallängen bei Umlauten sind auf sprecherspezifische Eigenheiten zurückzuführen. Bei anderen Vokalen mit einem größeren Umfang an Stichproben ist die Verteilung der Vokaldauer für beide Geschlechter gleich.

Nach einer Anfrage von Ingo Hertrich und Fabian Tomaschek, ob die Vokaldauern für /a:/ und /a/ im Kiel Corpus und in der deutschen Kontrollgruppe tatsächlich übereinstimmen, vergleiche ich die aus beiden Corpora verwendeten Worten.

Dabei fällt auf, dass die in (Simpson, 1988, S. 86) angegebenen Medianwerte für die Vo-

kaldauer der einzelnen Konsonanten sich je nach Sprechsituation und nach Geschlecht merklich unterscheiden. Für gelesene Worte liegt das Verhältnis der Vokaldauern von /a:/ und /a/ in der männlichen Sprechergruppe bei 1.72, in der weiblichen Gruppe bei 1.68. (Mooshammer, Fuchs & Fischer, 1999) vergleicht gespannte und ungespannte Vokale in betonter und unbetonter Position und greift dabei sowohl auf akustische Daten als auch auf kinematische Messungen durch Sensoren auf der Zungenspitze und am Kiefer zurück, die an insgesamt vier deutschen Muttersprachlern, zwei Männern und zwei Frauen durchgeführt werden. Sowohl die akustische Vokaldauer als auch der Zeitraum zwischen der Öffnung und dem Verschluß des Kiefers sind für betonte gespannte Vokale mit durchschnittlich 119 ms deutlich länger als für betonte ungespannte Vokale mit 49 ms. Zwischen betonten ungespannten Vokalen einerseits und unbetonten gespannten und ungespannten Vokalen andererseits besteht jedoch kein signifikanter Unterschied, was die Vokaldauer betrifft. Die Dauer von /a:/ und /a/ wies in unbetonten Silben keine Unterschiede auf.

(Jessen, 1993) beschreibt die Unterschiede in Formantwerten und Vokaldauer für Vokale in betonten Silben, Vokale vor einer betonten Silbe und Vokale, die in der zweiten Silbe vor einer betonten Silbe auftauchen, also z.B. in *Schieferei* oder *Hackerei*.

Dabei ergeben sich für die betonten Silben bei allen untersuchten gespannten und ungespannten Vokalen signifikante Unterschiede im Bezug auf die Vokaldauer. Werden hingegen Vokale vor einer betonten Silbe verglichen, weist die Vokaldauer bei gespannten und ungespannten Vokalen keine signifikanten Unterschiede auf.

Im Gegensatz dazu weist bei einem Vergleich zwischen betonten Silben und Vokalen, die zwei Silben vor einer betonten Silbe auftauchen, die Vokaldauer für gespannte und ungespannte Vokale in jedem Fall einen signifikanten Unterschied auf. Dies könnte jedoch auch durch einen sekundären Wortakzent in der zweiten Silbe vor der betonten Silbe bedingt sein.

Die Vokale /a:/ und /a/ unterscheiden sich sowohl in betonten Silben als auch in Vokalen, die zwei Silben vor einer betonten Silbe liegen, deutlich, was ihre Vokaldauer betrifft. Im Gegensatz zu anderen Vokalen ist hier die Entfernung zwischen den Formantwerten von /a:/ und /a/ jedoch wesentlich geringer.

Hier stelle ich folgende Abweichungen zwischen Corpora bzw. Sprechergruppen fest, die sich jedoch erklären lassen:

/e:/ Hier werden aus dem LeaP-Corpus nur fünf Worte verwendet, aus dem Kiel Corpus dagegen 56 (Männer) bzw. 53 (Frauen). Außerdem fallen einige Abschnitte des Sprechers *co* sehr kurz aus (<50 ms).

/ø:/ Hier habe ich aus dem LeaP-Corpus wieder nur zwei Worte, *böser* und *Löwe*, aus dem Kiel Corpus dagegen 29.

/a:/ und /a/ In diesem Fall sind es wieder nur neun Worte mit /a:/ und 14 Worte mit /a/ aus dem LeaP-Corpus, die überwiegend aus der Mitte der ersten Silbe eines zweisilbigen Wortes stammen, während ich aus dem Kiel Corpus 85 Worte für /a:/

Tabelle 3.2: Durchschnittliche Vokaldauer für betonte und unbetonte Vokale

Vokal	Vokaldauer (in ms)							
	Kiel Corpus of Read Speech				LeaP-Corpus			
	betont		unbetont		betont		unbetont	
	männlich	weiblich	männlich	weiblich	männlich	weiblich	männlich	weiblich
/i:/	84,59	86,75	62,83	61,48	80,00	90,00	61,69	60,43
/ɪ/	57,52	54,46	55,28	52,77	47,95	66,40	50,00	48,57
/y:/	86,44	97,64	80,56	96,15	58,00	70,00	53,33	–
/ʏ/	63,19	59,85	56,67	45,00	76,67	66,67	–	–
/e:/	106,72	117,95	113,00	111,25	75,00	83,75	–	–
/ɛ/	80,28	82,69	67,67	66,24	60,00	66,67	60,00	66,00
/ɛ:/	122,88	126,17	157,50	134,85	150,00	150,00	–	–
/ø/	111,72	115,40	111,86	112,67	149,23	135,00	–	–
/œ/	99,00	77,20	72,19	73,64	65,00	80,00	–	–
/a:/	153,13	153,58	126,83	127,26	93,16	99,63	130,00	135,00
/a/	78,35	82,82	78,55	81,71	64,33	81,83	39,25	73,25
/o:/	124,48	127,55	78,89	75,13	144,00	126,00	113,67	94,00
/ɔ/	87,70	89,18	88,47	80,33	64,65	70,89	–	–
/u:/	77,99	78,51	106,00	91,33	71,82	93,14	24,00	23,00
/ʊ/	62,67	63,43	61,89	57,20	63,29	67,33	46,40	59,40

und 135 Worte für /a/ zur Verfügung habe. Die Abschnitte aus dem Kiel Corpus stammen auch aus drei- oder viersilbigen Worten oder stehen am Wortanfang.

/o:/ Hier stehen die Abschnitte aus dem LeaP-Corpus wieder in vier zweisilbigen Worten in der Mitte der ersten Silbe, während die Abschnitte aus dem Kiel Corpus aus 47 bzw. 44 Worten am Wortanfang oder in der Mitte von zwei-, drei- oder viersilbigen Worten stehen.

/u:/ Die extrem kurzen Werte für die unbetonten Silben lassen sich auf Abweichungen bei der Sprecherin *dy* zurückführen, die diesen Vokal in weniger als 30 Milisekunden aussprach.

3.3 Kurzgeschichten und gelesene Dialoge

3.3.1 Problem

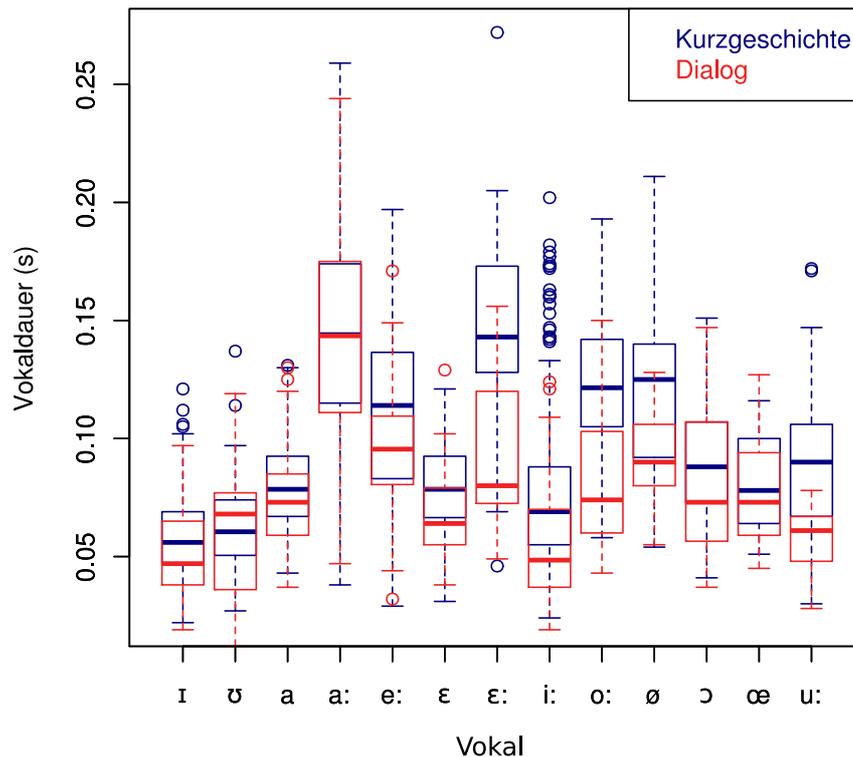
Um ein umfassendes Modell der deutschen Aussprache zu erhalten, analysiere ich zuerst die Formantwerte der oralen Monophthonge im Kiel Corpus of Read Speech.

Dabei fällt Folgendes auf:

Die Aufnahmen im Verzeichnis PH90 bestehen aus gelesenen Kurzgeschichten, im Verzeichnis PH92 finden sich dagegen gelesene Dialoge, z.B. Anfragen zu Zugverbindungen.

Bei den Formantwerten treten teilweise Unterschiede zwischen beiden Verzeichnissen auf. Noch deutlicher sind diese Unterschiede bei der Vokaldauer.

Abbildung 3.7: Vokaldauer nach Textsorten für männliche Sprecher im Kiel Corpus



Da sich mit Hilfe der Formantmessungen in der Vokalmitte keine eindeutigen Ergebnisse ermitteln lassen, greife ich auf Daten, die jeweils den ersten und den letzten Messwert eines extrahierten Abschnitts enthalten zurück.

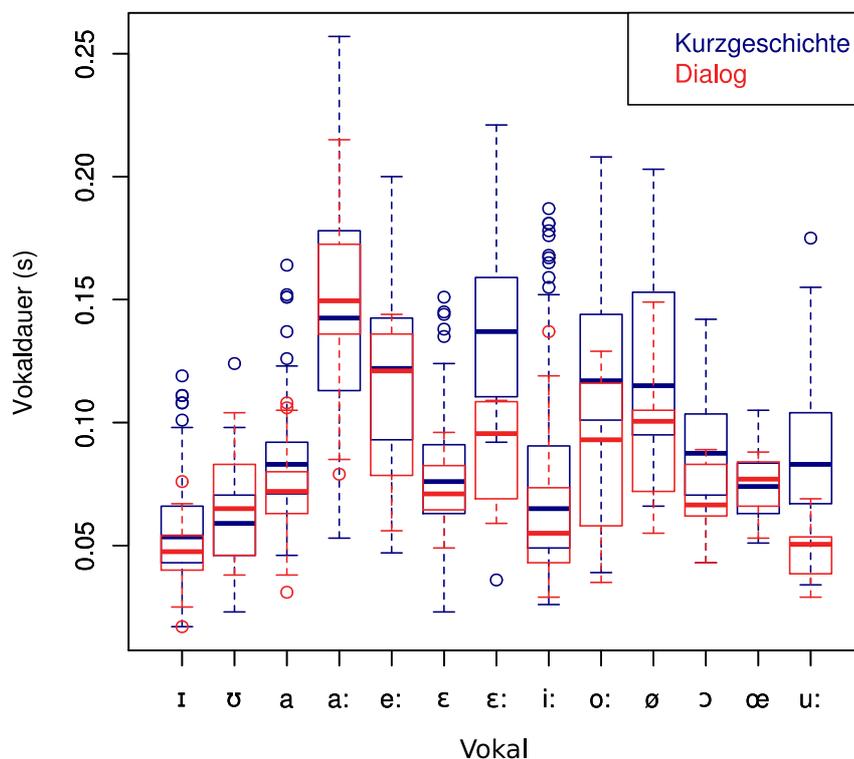
Da dies nur bei Abschnitten möglich ist, deren Dauer den Umfang eines untersuchten Fensters übersteigt, vergleiche ich hier nur gespannte Vokale, bei denen dies in der Regel der Fall ist.

Zeigen sich so schon innerhalb des Kiel Corpus of Read Speech messbare Unterschiede zwischen gelesenen Kurzgeschichten und Dialogen, so kann man damit die Beschränkung auf direkt gelesene Kurzgeschichten bei der Untersuchung der Vokale des LeaP-Corpus begründen.

3.3.2 /i:/

Da die Artikulationsorgane für /i:/ als vorderen, hohen Vokal vor allem bei der isolierten Artikulation eine Extremposition einnehmen, ist dieser Vokal für koartikulatorische Effekte besonders anfällig, was in den folgenden Untersuchungen deutlich sichtbar wird.

Abbildung 3.8: Vokaldauer nach Textsorten für weibliche Sprecher im Kiel Corpus



Innerhalb dieses Abschnitts untersuche ich für die männlichen Sprecher jeweils 58 Abschnitte pro Textsorte, für die weiblichen Sprecher jeweils 43 Abschnitte.

3.3.2.1 /i:/, männliche Sprecher

3.3.2.1.1 Eigenschaften des gemischten Modells Die AIC-Werte fallen hier mit bis zu 3000 Punkten sehr hoch aus, so dass die Anpassung an die gemessenen Daten nur ungefähr ausfällt. Der Anteil der Abweichungen individueller Sprecher ist für F_3 mit 39,8% am höchsten und für F_1 mit 16,2% am niedrigsten.

Dies bedeutet, dass andere Faktoren, z.B. die Silbenbetonung oder koartikulatorische Aspekte sich ebenfalls stark auf die Formantwerte auswirken.

Tabelle 3.3: Vergleich zwischen /i:/ in gelesenen Kurzgeschichten und Dialogen für männliche Sprecher

Parameter	t-Wert	pMCMC	p-Wert
F_1	-2.14	0.0362	0.0337
F_2	-0.19	0.8566	0.8459
F_3	-0.073	0.9354	0.9420
Vokaldauer	1.297	0.1310	0.196

Vergleicht man die Formantwerte und die Vokaldauer für /i:/ in gelesenen Kurzgeschichten und Dialogen insgesamt, so zeigt sich nur für F_1 ein signifikanter Unterschied. Dies ist

durch die Abhängigkeit des ersten Formanten von der vertikalen vorderen Zungenposition begründet, bei der sich der Genioglossus stark nach vorn und oben streckt. Diese Bewegung wird in schneller gesprochenen Silben möglicherweise weniger intensiv ausgeführt. Dies bestätigt sich, wenn ich die durch das gemischte Modell errechneten Durchschnittswerte vergleiche, bei denen das Ergebnis für F_1 in Dialogen über dem Ergebnis in Kurzgeschichten liegt.

Tabelle 3.4: Durchschnittliche Messwerte für /i:/ bei männlichen deutschen Sprechern in Dialogen und Kurzgeschichten

Parameter	Kurzgeschichte	Dialog
F_1 (Hz)	306	324
F_2 (Hz)	2025	2010
F_3 (Hz)	2529	2557
Vokaldauer (ms)	69,68	54,80

Die durchschnittliche Länge der Vokale unterscheidet sich zwar nicht signifikant, fällt in den Dialogen doch tendenziell kürzer aus.

3.3.2.2 /i:/, weibliche Sprecher

3.3.2.2.1 Eigenschaften des gemischten Modells Betrachte ich die Eigenschaften des resultierenden gemischten Modells, so fällt auf, dass trotz der bei weiblichen Sprechern üblichen größeren Streuung der Messwerte niedrigere BIC-Werte erzielt werden, als bei dem gemischten Modell für /i:/ von männlichen Sprechern. Der Unterschied zwischen AIC- und BIC-Werten ist ebenfalls leicht geringer als bei männlichen Sprechern. Dazu könnte das wesentlich größere Gewicht individueller Abweichungen beitragen, die für F_1 und F_2 jeweils ca 51% und für F_3 immerhin noch 41,6% betragen.

Tabelle 3.5: Vergleich zwischen /i:/ in gelesenen Kurzgeschichten und Dialogen für weibliche Sprecher

Parameter	t-Wert	pMCMC	p-Wert
F_1	-0.198	0.7712	0.8431
F_2	1.505	0.0302	0.1344
F_3	1.049	0.1544	0.2960
Vokaldauer	1.038	0.0828	0.3007

Für weibliche Sprecher ergeben sich für /i:/ keine signifikanten Unterschiede mehr, was vor allem durch den starken Einfluss individueller Sprechereigenschaften bedingt ist.

Tabelle 3.6: Durchschnittliche Messwerte für /i:/ bei weiblichen deutschen Sprechern in Dialogen und Kurzgeschichten

Parameter	Kurzgeschichte	Dialog
F_1 (Hz)	329	333
F_2 (Hz)	2479	2270
F_3 (Hz)	3024	2850
Vokaldauer (ms)	72,27	58,69

Suche ich nach Gründen, warum sich F_1 für /i:/ bei männlichen Sprechern signifikant unterscheidet, bei weiblichen Sprechern jedoch nicht, so muss ich zuerst feststellen, dass die weiblichen Formantmesswerte breiter gestreut sind, was auch innerhalb des gemischten Modells mit dem einzelnen Sprecher als Zufalsfaktor nicht vollständig ausgeglichen werden kann.

3.3.3 /o:/

3.3.3.1 /o:/, männliche Sprecher

Für /o:/ untersuche ich nur die Aufnahmen der männlichen Sprecher mit 15 Abschnitten pro Textsorte, da für die weiblichen Sprecher zu wenige Abschnitte verfügbar waren.

3.3.3.1.1 Eigenschaften des gemischten Modells Das gemischte Modell ist hier sehr viel besser an die Daten angepasst als für /i:/, und der Abstand zwischen AIC- und BIC-Werten fällt mit weniger als 10 Punkten sehr gering aus. Dies bedeutet, dass es sich um ein sehr genau an die bestehenden Daten angelehntes, nicht übermäßig komplexes Modell handelt.

Der einzelne Sprecher als Zufallsfaktor beeinflusst F_1 am stärksten. Dort erklärt er 49,4% der Abweichungen. Die Auswirkungen sind für die Vokaldauer mit 43,9% der Abweichungen am geringsten.

Tabelle 3.7: Vergleich zwischen /o:/ in gelesenen Kurzgeschichten und Dialogen für männliche Sprecher

Parameter	t-Wert	pMCMC	p-Wert
F_1	-0.454	0.6032	0.6532
F_2	-0.547	0.5290	0.5881
F_3	0.442	0.7102	0.6618
Vokaldauer	3.659	0.0020	0.0009

Die durchschnittlichen Werte für die Vokaldauer fallen in Dialogen deutlich niedriger aus, als in Kurzgeschichten, während die Formantwerte keine großen Unterschiede aufweisen.

Tabelle 3.8: Durchschnittliche Messwerte für /o:/ bei männlichen deutschen Sprechern in Dialogen und Kurzgeschichten

Parameter	Kurzgeschichte	Dialog
F_1 (Hz)	410	418
F_2 (Hz)	889	927
F_3 (Hz)	2327	2322
Vokaldauer (ms)	125,78	69,78

3.3.4 Zusammenfassung

Für männliche Sprecher unterscheiden sich folgende Messwerte signifikant in gelesenen Kurzgeschichten und Dialogen:

F_1 für /i:/

Vokaldauer für /o:/

Im Gegensatz zu t-Tests oder Rangsummentests erhielt ich für gemischte Modelle in denen der einzelne Sprecher als Zufallsfaktor eingesetzt wurde, sonst keine signifikanten Unterschiede mehr.

Die größere Streuung der Werte bei weiblichen Sprechern konnte nicht vollständig durch die einzelnen Sprecherinnen als Zufallsfaktoren aufgefangen werden.

Andere mögliche Zufallsfaktoren wie z.B. die Grundfrequenz oder die Intensität kommen nicht in allen Fällen in Frage, da die Anzahl der Ausprägungen der Zufallsfaktoren geringer sein muss als die Anzahl der untersuchten Abschnitte.

Die Silbenbetonung spielt zwar eine wichtige Rolle, da in den Testmengen jeweils etwa gleich viel betonte und unbetonte Samples pro Verzeichnis enthalten waren, trägt hier sie hier nicht zu Verzerrungen bei.

Obwohl in keinem Fall signifikante Unterschiede zwischen Formantwerten am Silbenanfang und -ende festgestellt werden konnten, ist die Streuung geringer als bei meinen vorherigen Messungen, bei denen ich nur Formantwerte aus der Silbenmitte untersucht habe.

4 Deutsche und britische Sprecher

4.1 Einleitung

Deutsch und Englisch besitzen als Angehörige der germanischen Sprachfamilie beide ein sehr umfangreiches und ähnliches Vokalsystem. Betrachtet man die jeweiligen Vokale und ihre Umsetzung jedoch näher, werden Unterschiede sichtbar.

4.1.1 Sprachgeschichte

4.1.1.1 Altenglisch

Beim Übergang vom Germanischen zum Altenglischen verändern sich mehrere phonetische Merkmale (s. (Gelderen, 2006, S.51-55))

Stimmhaftigkeit Stimmlose labiodentale, dental-alveolare und alveolare Frikative zwischen zwei Vokalen werden stimmhaft, z.B. wird /f/ zu /v/, /s/ zu /z/ und /θ/ zu /ð/.

Palatalisierung Velare Plosive werden zu palatalen bzw. post-alveolaren Frikativen, so wird /sk/ zu /ʃ/, /k/ zu /tʃ/ und /g/ zu /j/.

breaking Die vorderen Vokale i, e und æ werden vor [l,r] und einem weiteren Konsonanten bzw. vor /h/ zu *io/eo*, *eo* und *ea* diphthongisiert.

i-mutation (i-Umlaut) Hierbei werden hintere Vokale vor einer Wortendung mit /i/ durch vordere Vokale ersetzt, z.B. /u/ durch /y/. Da die Wortendung später entfällt, dient dieser Vokalkontrast zur Unterscheidung von Singular und Plural sowie von intransitiven und transitiven Verben.

Der altenglische Vokalraum umfasst die Vokale /a:/, /æ/, /ɛ:/, /ɛ/, /ɪ/, /i/, /ɒ/, /o:/, /ʊ/, /u:/, /y/ und /y:/ sowie die Diphthonge /æə/, /ɛə/ /eə/ und /e:ə/, s. (McIntyre, 1975, S. 38).

4.1.1.2 Mittelenglisch

(Moessner, 2003, S. 75 ff) führt folgende Veränderungen vom Altenglischen zum Mittelenglischen an:

Kürzung von Vokalen Im Altenglischen besteht die Vokalqualität unabhängig von der Silbenstruktur, so dass lange und kurze Vokale in offenen und geschlossenen Silben

stehen. Im Mittelenglischen dagegen werden lange Vokale verkürzt, wenn sie vor einem Konsonantencluster stehen, z.B. wird *sōfte* zu *softə* (heute: soft) und *brēost* zu *brest* (heute: breast).

Längung von Vokalen Die altenglischen kurzen, betonten Vokale /e/, /a/ und /o/ werden im Mittelenglischen zu /ɛ:/, /a:/ und /ɔ:/ gelängt. Dieser Vorgang umfasst nur zweisilbige Worte mit offener erster Silbe, z.B. wird /e-tan/ zu /ɛ:-tə/ (heute: eat), /nama/ zu /na:-mə/ (heute: name) und /talʊ/ zu /ta:-lə/ (heute: tale).

Monophthongierung Die altenglischen Diphthonge *ēo*, *eo* werden zu /e:/ und /e/. *ēa* und *ea* fallen zu *æ* zusammen.

Änderung der Vokalqualität Im 11. und 12. Jahrhundert wird /æ/ durch /a/ ersetzt, z.B. in *sæt*, das zu *sat* wird. Gleichzeitig wird /a:/ zu /ɔ:/, z.B. wird /sta:n/ zu /stɔ:n/.

Das Verschwinden der Flexionsformen während des Übergangs zum Mittelenglischen lässt sich unter anderem darauf zurückführen, dass unbetonte Silben wie z.B. die Endungen in *faru* (Reise, Sg.) und *fara* (Reisen, Pl.) zuerst als /ə/ gesprochen werden. Dabei ist jedoch der grammatische Unterschied nicht mehr hörbar, so dass der unbetonte Vokal später entfällt, s. auch (McIntyre, 1975, S. 51).

4.1.1.3 Die Great Vowel Shift (GVS)

Während der Great Vowel Shift, die zwischen 1400 und 1700 stattfindet, nehmen niedrige und halb-hohe lange Vokale höhere Positionen ein, während hohe lange Vokale diphthongisiert werden. Dabei lassen sich zwei Reihen unterscheiden:

vordere Reihe /a:/ ⇒ /ɛ:/ ⇒ /e:/ ⇒ /i:/ ⇒ /ɪ/

hintere Reihe /ɔ:/ ⇒ /o:/ ⇒ /u:/ ⇒ /ʊ/

Die heutige Schreibweise englischer Worte spiegelt den Stand der Sprache vor dem Abschluss der Great Vowel Shift wider, s. (McIntyre, 1975, S. 57-59).

Die Veränderung begann zuerst bei den hohen und niedrigen Vokalen, so wurde z.B. /i/ schon um 1500 als /ej/, /e/ als /i/ und /o/ als /u/ gesprochen, während die Verschiebungen bei den halbhohen Vokalen später begannen, z.B. wird der Vokal in *great* zuerst als /ɛ/ gesprochen und erst um 1600 als /ej/. Der Vokal in *boat* wird noch im 16. Jahrhundert als /ɔ/ und erst ab 1700 als /ow/ gesprochen, s. (Gelderen, 2006).

4.1.2 Vokalraum

(Jones, Roach, Hartmann & Setter, 2003, S. ix) bemerkt, dass /æ/ zwar zu den kurzen Vokalen gehöre, vor stimmhaften Plosiven und Nasalen wie /b,d,g,ʒ,m,n/ jedoch deutlich länger ausgesprochen werde als vor anderen Konsonanten. Im britischen Englisch

Abbildung 4.1: Englische Monophthonge

	vorn	zentral	hinten
geschlossen	/i:/, /i/		/u:/, /u/
halbgeschlossen	/ɪ/	/ʊ/ /ə/, /ɜ:/	/ɔ:/
halboffen	/e/	/ʌ/	
offen	/æ/		/ɒ/ /ɑ:/

nach (Dretzke, 1998, S. 28)

werde dieser Vokal zunehmend offener ausgesprochen, so dass man ihn eines Tages als /a/ transkribieren könne. Im amerikanischen Englisch dagegen werde /æ/ zunehmend geschlossener gesprochen.

(Ladefoged, 2006, S. 96ff) beschreibt den Unterschied zwischen gespannten und ungespannten Vokalen. Er erwähnt zuerst einige Gegensatzpaare, bei denen der gespannte Vokal in Worten mit einem „stillen e“, der ungespannte Vokal in Worten ohne „stilles e“ steht, z.B. *mete* – *met* oder *cute* – *cut*. Generell stellt er fest, dass in geschlossenen Silben alle, in offenen Silben nur bestimmte Vokale auftreten können. So treten in betonten offenen Silben die Vokale /ɪ/, /ɛ/, /æ/, /ʊ/ und /ʌ/ nicht in Erscheinung. Die ungespannten Vokale fallen in der Regel kürzer, niedriger und zentraler aus als ihre gespannten Gegenstücke. Er fügt hinzu, dass im britischen Englisch /ɒ/ als zusätzlicher ungespannter Vokal existiere. Dieser Vokal bildet ein Gegensatzpaar mit /ɑ/, z.B. für *card* [kɑ:d] und *cod* [kɒd].

Insgesamt stellt er für englische Vokale folgende Regeln auf:

- Sind alle anderen Bedingungen gleich, so ist die Vokaldauer bei einer offenen Silbe am längsten, kürzer in einer Silbe, die mit einem stimmhaften Vokal endet und am kürzesten in einer Silbe, an deren Ende ein stimmloser Vokal steht.
- Vokale in betonten Silben sind länger als Vokale in unbetonten Silben.
- Sind alle anderen Bedingungen gleich, so ist die Vokaldauer in einsilbigen Worten am größten, kürzer in zweisilbigen Worten und am kürzesten in Worten mit mehr als zwei Silben.
- Vokale werden in Silben nasalisiert, die durch einen nasalen Konsonant abgeschlossen werden.

4.1.3 Vergleich der Durchschnittswerte

4.1.3.1 Vergleich von Referenzwerten

(Deterding, 1997) misst Formantwerte bei britischen Muttersprachlern. Berechnet man aus seinen Messwerten pro Sprecher Durchschnittswerte für Männer und Frauen, sehen die Ergebnisse folgendermaßen aus:

- Die F_2 -Werte für die gespannten hohen Vokale /i:/ und /u:/ fallen deutlich höher aus als bei deutschen Muttersprachlern.
- Alle F_3 -Werte fallen ebenfalls höher aus als bei deutschen Muttersprachlern.
- Die britischen ungespannten Vokale nehmen eine zentralere Position ein als ihre deutschen Äquivalente.
- Der Vokal /æ/ nimmt bei den britischen Muttersprachlern für alle Formantwerte eine Extremposition ein, die von keinem deutschen Monophthong erreicht wird.

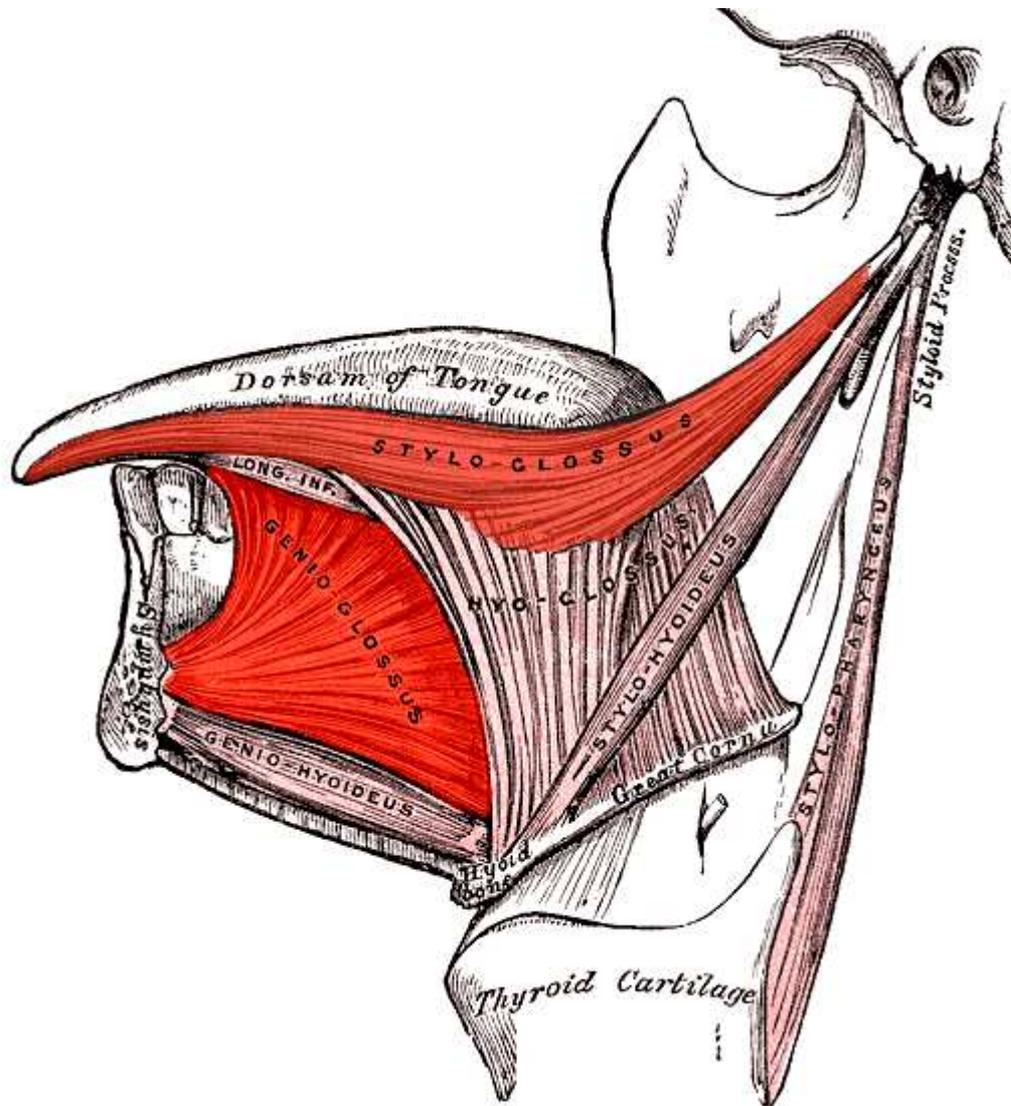
(Harshman, Ladefoged & Goldstein, 1977) erstellen auf der Grundlage von Röntgenbildern zehn englischer Vokale mit Hilfe der Principal Component Analysis ein statistisches Modell um Vokale nach ihren artikulatorischen Eigenschaften zu kategorisieren. Dabei werden fünf Versuchspersonen dabei aufgenommen, wie sie einsilbige Worte mit der Struktur /hVd/, z.B. *heed*, *had* oder *hood* aussprechen. Bei der Auswertung ergeben sich zwei aussagekräftige Faktoren:

Front-Raising Dazu gehören eine Vorwärtsbewegung der Zungenwurzel und eine Anhebung der Zungenspitze. Diese Bewegungen entsprechen etwa dem Einsatz des Genioglossus.

Back-Raising Dieser Faktor umfasst eine Bewegung der Zunge nach oben und nach hinten und entspricht ungefähr dem Anziehen des Styloglossus.

Insgesamt deckt das Modell 96% der gemessenen Unterschiede ab.

Abbildung 4.2: Übersicht zur Zungenmuskulatur



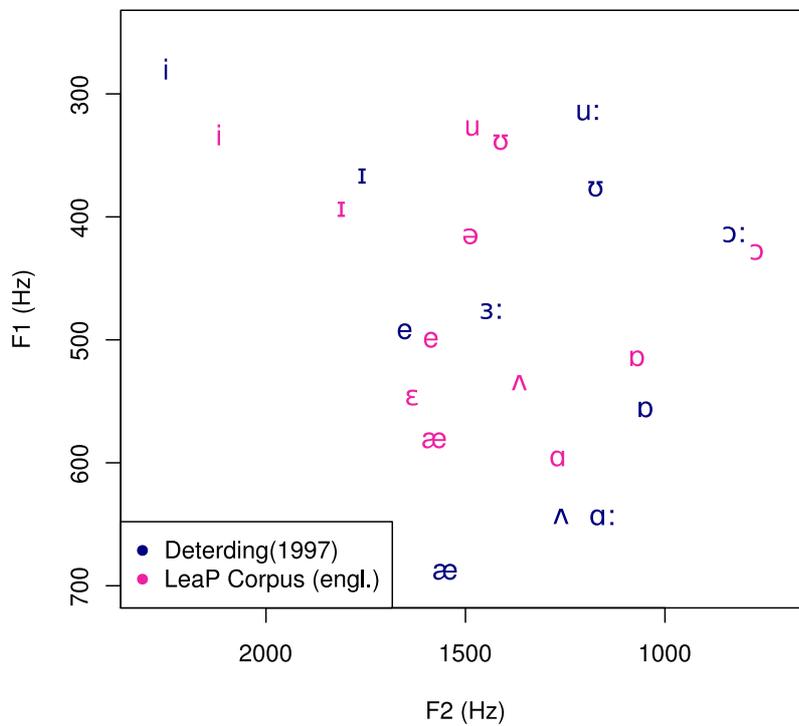
(Gray, 1918, Fig. 177), farbige Hervorhebungen durch die Verfasserin

Wendet man diese Erkenntnisse auf die Messergebnisse hier an, so werden die britischen Vokale insgesamt mit einer weiter nach vorne bewegten Zungenwurzel artikuliert. Da in (Deterding, 1997) jedoch die Formanten in einzelnen einsilbigen Worten gemessen wurden, im Kiel Corpus jedoch in zusammenhängenden gelesenen Texten, sind bei den britischen Messungen die koartikulatorischen Effekte geringer.

4.1.3.2 Gegenüberstellung von britischen muttersprachlichen Vokalen in einzelnen Worten und gelesenen Texten

Um für diese Untersuchungen besser vergleichbare Werte zu finden, führe ich zusätzlich Messungen an muttersprachlichen Aufnahmen der vier britischen Sprecher innerhalb des LeaP-Corpus durch.

Abbildung 4.3: Durchschnittswerte für F_1 und F_2 bei von britischen männlichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen



(Messwerte nach (Deterding, 1997) und eigenen Messungen im englischen Teil des LeaP-Corpus)

Abbildung 4.4: Durchschnittswerte für F_1 und F_2 bei von britischen weiblichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen

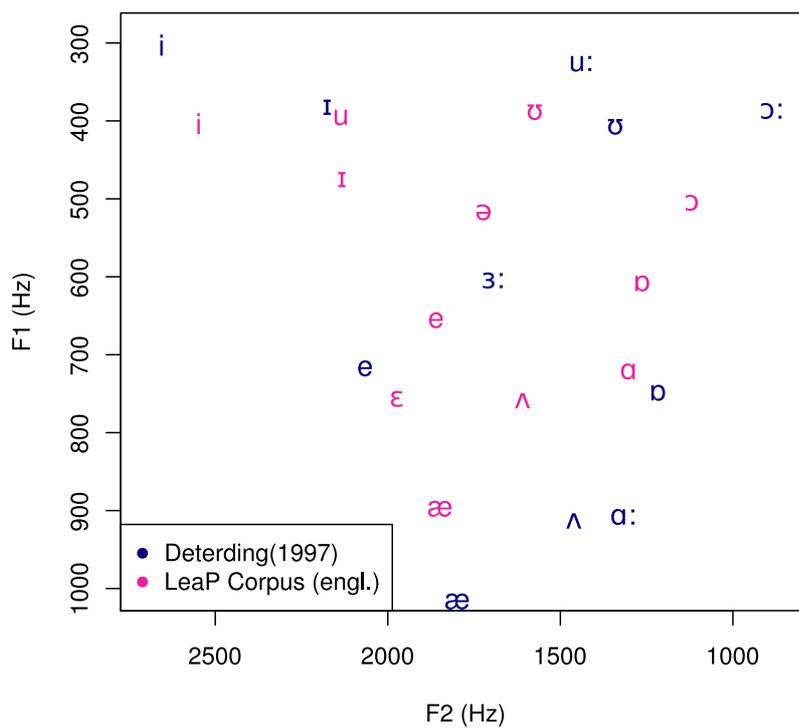
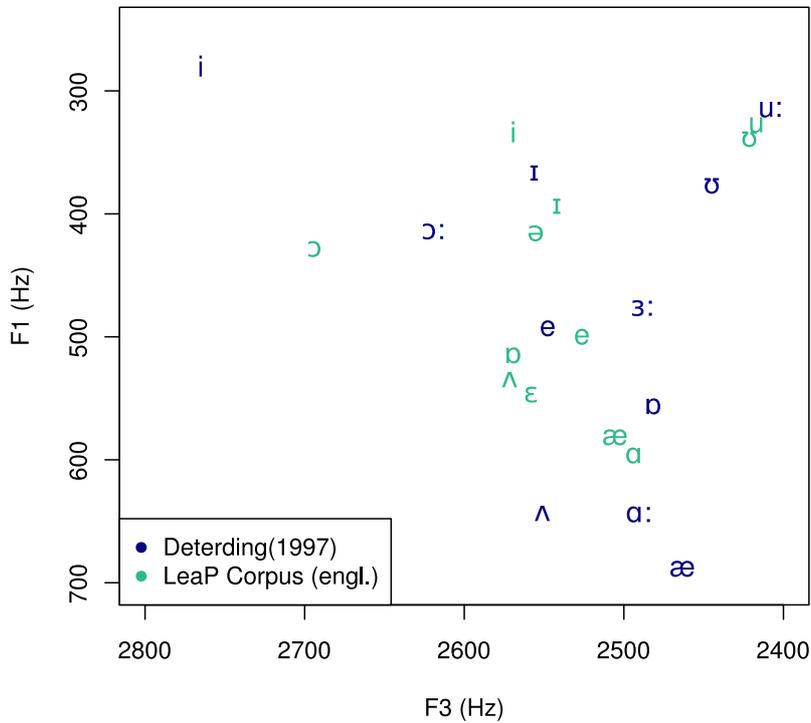
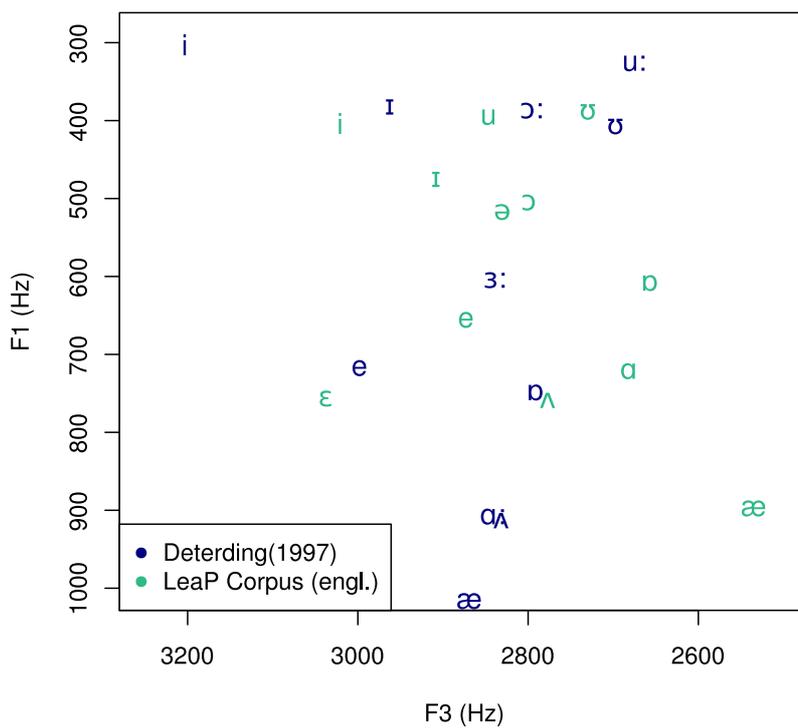


Abbildung 4.5: Durchschnittswerte für F_1 und F_3 bei von britischen männlichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen



(Messwerte nach (Deterding, 1997) und eigenen Messungen im englischen Teil des LeaP-Corpus)

Abbildung 4.6: Durchschnittswerte für F_1 und F_3 bei von britischen weiblichen Sprechern einzeln gesprochenen oder im Kontext gelesenen Vokalen



Dabei fällt die vor allem die zentralere Position der gemessenen Vokale auf. Sowohl für F_1 als auch vor allem für F_2 nehmen die innerhalb eines gelesenen Textes gemessenen Vokale zentralere Positionen ein, als die Vokale in (Deterding, 1997), die innerhalb einzelner Worte gemessen wurden.

Da bei den britischen Sprechern im englischen Teil des LeaP-Corpus Aufnahmen von drei männlichen Sprechern, jedoch nur einer weiblichen Sprecherin vorhanden sind, kann man die relativ vordere Position des /u/ bei der weiblichen Sprecherin als Verzerrung aufgrund individueller Ausspracheigenschaften betrachten. Ansonsten bestätigen sich bei den weiblichen Messwerten die gleichen Tendenzen wie bei den männlichen Daten.

Interessanterweise liegen die Formantwerte für F_1 und F_2 bei /ʌ/ und /ɑ/ bei den im Kontext gelesenen Vokalen für alle Sprecher weiter auseinander, als bei den einzelnen Vokalen. Da /ʌ/ und /ɑ/ Minimalpaare bilden, z.B. in *hut* [hʌt] und *hart* [hɑ:t] (männlicher Hirsch), werden sie in gelesenen Texten, in denen beide Vokale vorkommen, offensichtlich deutlicher artikuliert, um Verwechslungen zu vermeiden.

Die F_3 -Werte von /ʌ/ und /ɑ/ liegen bei den gelesenen Vokalen ebenfalls stärker auseinander, dafür liegen die F_3 -Werte von /ɑ/ und /ɒ/ sehr nahe beieinander. /ɑ/ und /ɒ/ unterscheiden sich nur in ihren F_2 -Werten deutlich voneinander.

Für /æ/, /ʌ/, /ɑ/ und /ɒ/ gilt, dass die F_3 -Werte innerhalb gelesener Texte höher liegen, als bei einzeln artikulierten Vokalen. Dies wird durch eine schwächere Anspannung der Lippen hervorgerufen.

Da sowohl die Werte von (Deterding, 1997) mit fünf männlichen und fünf weiblichen Sprechern als auch die Werte aus dem Kiel Corpus of Read Speech mit 22 männlichen und 20 weiblichen Sprechern auf einer relativ großen Datenbasis beruhen, werde ich sie nicht verwerfen, sondern die Ergebnisse meiner eigenen Messungen zusätzlich untersuchen.

Abbildung 4.7: Durchschnittswerte für F_1 und F_2 bei deutschen und britischen männlichen Sprechern im LeaP-Corpus

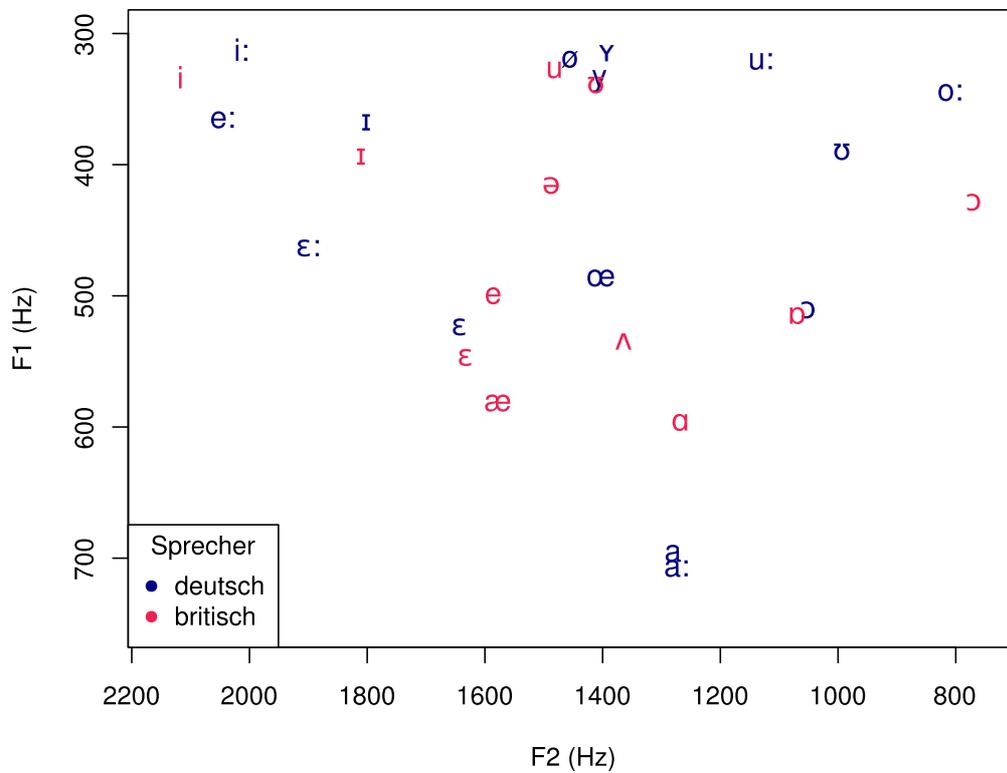


Abbildung 4.8: Durchschnittswerte für F_1 und F_3 bei deutschen und britischen männlichen Sprechern im LeaP-Corpus

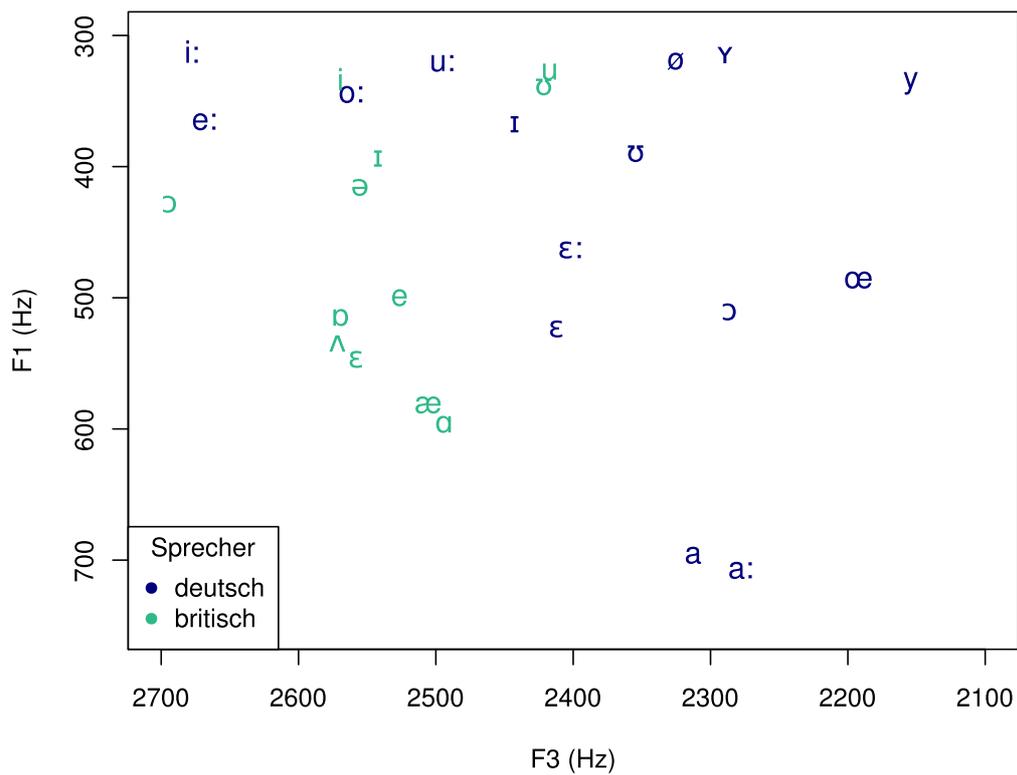


Abbildung 4.9: Durchschnittswerte für F_1 und F_2 bei deutschen weiblichen Sprechern und einer britischen weiblichen Sprecherin im LeaP-Corpus

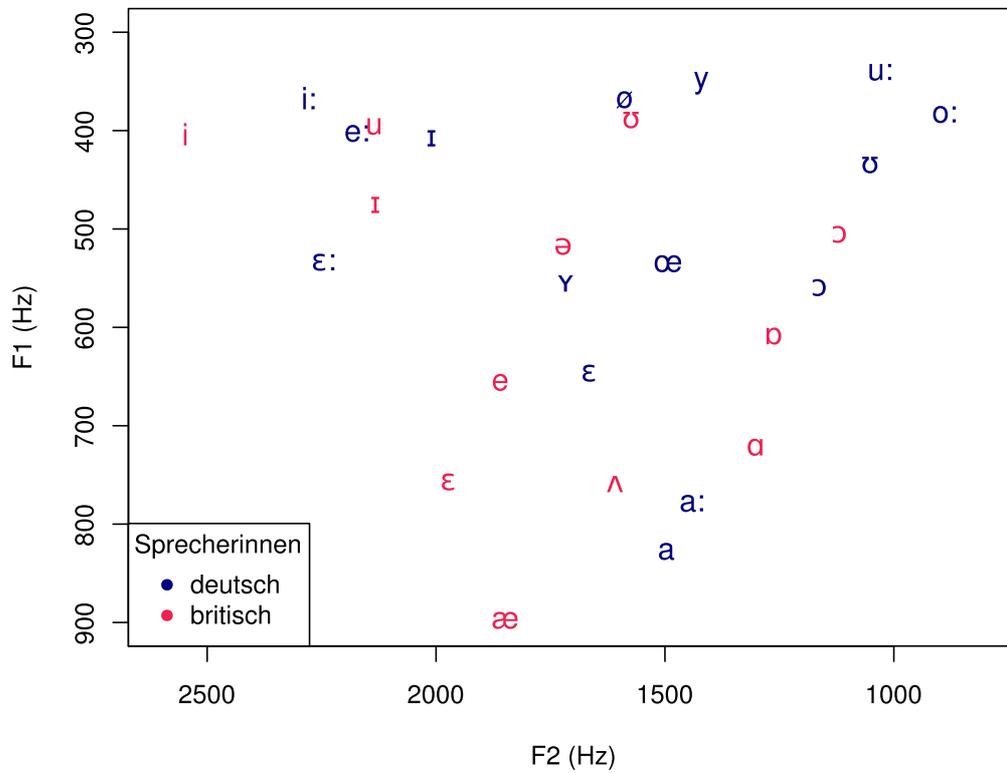
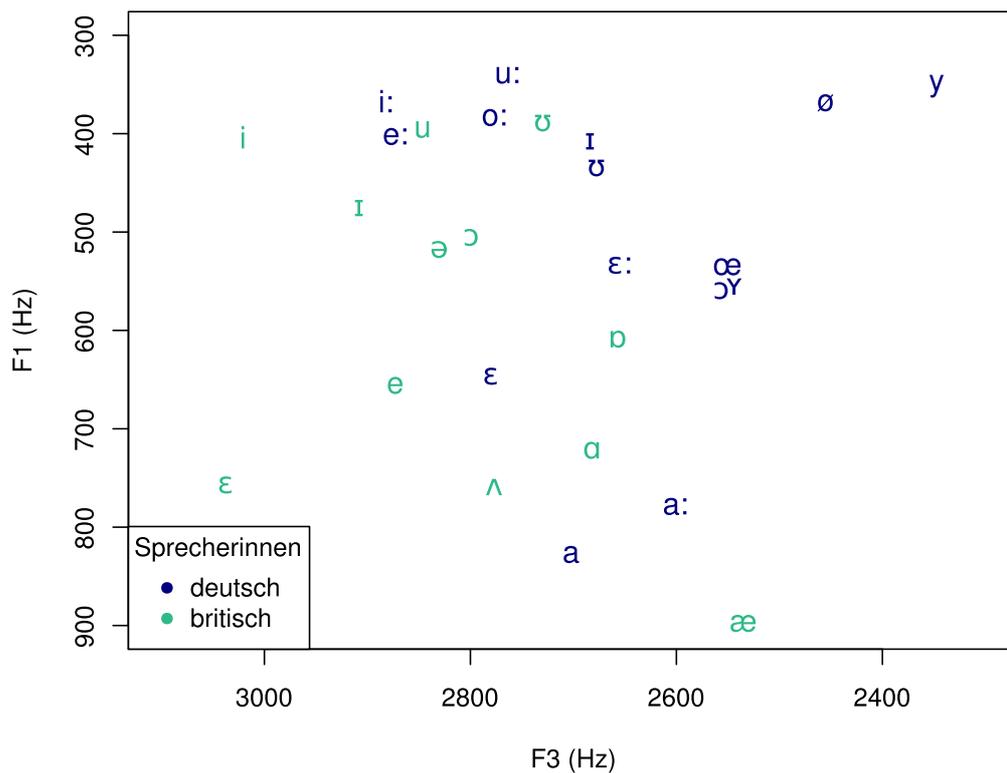


Abbildung 4.10: Durchschnittswerte für F_1 und F_3 bei deutschen weiblichen Sprechern und einer britischen weiblichen Sprecherin im LeaP-Corpus



/e:/ und */ɛ/* besitzen bei britischen Sprechern höhere F_1 -Werte und werden so deutlicher von */ə/* unterschieden.

Das britische */ɔ/* liegt mit seinem F_1 -Wert zwischen dem höheren Wert des vergleichbaren deutschen Vokals und dem niedrigeren F_1 -Wert, den das deutsche */o:/* aufweist. Dafür steht das britische */ɒ/* dem deutschen */ɔ/* relativ nahe.

Die F_1 -Werte der britischen Laute */ɑ/* und */ɒ/* liegen weit unter den deutschen F_1 -Werten für */a:/* und */a/*. Dafür liegen die F_2 - und F_3 -Werte von */ɑ/* und */ɒ/* deutlich über denen der deutschen Vokale.

Für */i/* und */ʊ/* sind die F_2 -Werte britischer Sprecher deutlich höher, wobei F_1 und F_2 für */ʊ/* fast wie ein deutsches */ø:/* ausfallen, so dass der Unterschied zwischen */ʊ/* und */ø:/* sich nur im größeren Abstand für F_3 zeigt. Dies bedeutet, dass sich das britische */ʊ/* und das deutsche */ø:/* vor allem im Hinblick auf die Lippenrundung unterscheiden.

Der hohe F_2 -Wert für */ʊ/* bei der weiblichen britischen Sprecherin, der schon beim Vergleich der britischen Formantwerte auffällt, kann als individuelle Eigenheit betrachtet werden.

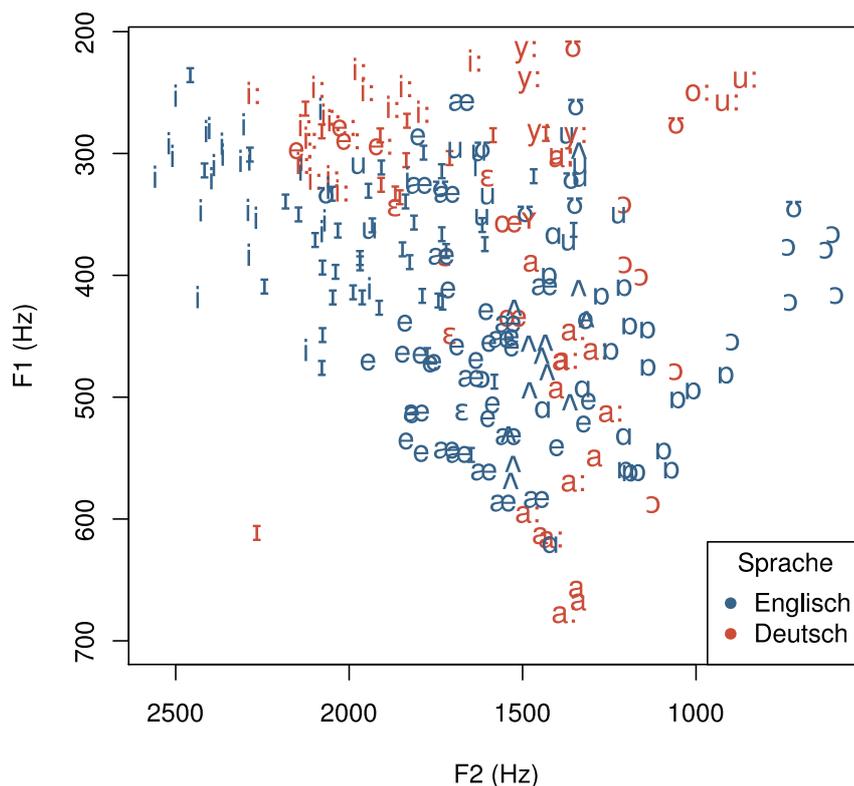
Die F_3 -Werte für */u/,/ʊ/* und */ɔ/* liegen bei den britischen Sprechern höher als bei den deutschen Sprechern. Dies spricht für eine schwächere Rundung aller hinteren Vokale.

4.1.4 Vokalräume einzelner Sprecher

4.1.4.1 Der männliche Sprecher ai

Für zwei britische Sprecher liegen im LeaP-Corpus muttersprachliche englische Aufnahmen und gleichzeitig zweitsprachliche deutsche Aufnahmen vor.

Abbildung 4.11: Vokale des zweisprachigen männlichen Sprechers ai



Die hohen Vokale /i:/, /ɪ/, /u:/ und /ʊ/ weisen hier für den gelesenen fremdsprachlichen deutschen Text niedrigere F_1 -Werte auf, als für den muttersprachlichen englischen Text. Die F_1 -Werte für /y:/ sind sehr niedrig, die F_2 -Werte dieses Vokals liegen zwischen /i:/ und /ʊ/. Damit entsprechen sie einem muttersprachlichen deutschen /y:/, was zeigt, dass der im Englischen unbekannt Laut erlernt wurde.

Die F_1 -Werte für /a:/ und /ʌ/ dagegen, weisen eine breitere Streuung auf und liegen teilweise deutlich tiefer, als die britischen Vokale /ɑ/ und /ɒ/. Ein Teil der F_1 - und F_2 -Werte für den deutschen Vokal /a:/ und den britischen Vokal /ʌ/ überlappt sich. Dies ist insbesondere für niedrigere F_1 - und weniger breit gestreute F_2 -Werte der Fall, könnte also darauf hinweisen, dass diese Vokale in unbetonten Silben ähnlich artikuliert werden.

Die Bandbreite der Formantwerte ist hier insgesamt bei deutschen Vokalen für F_1 größer, für F_2 geringer.

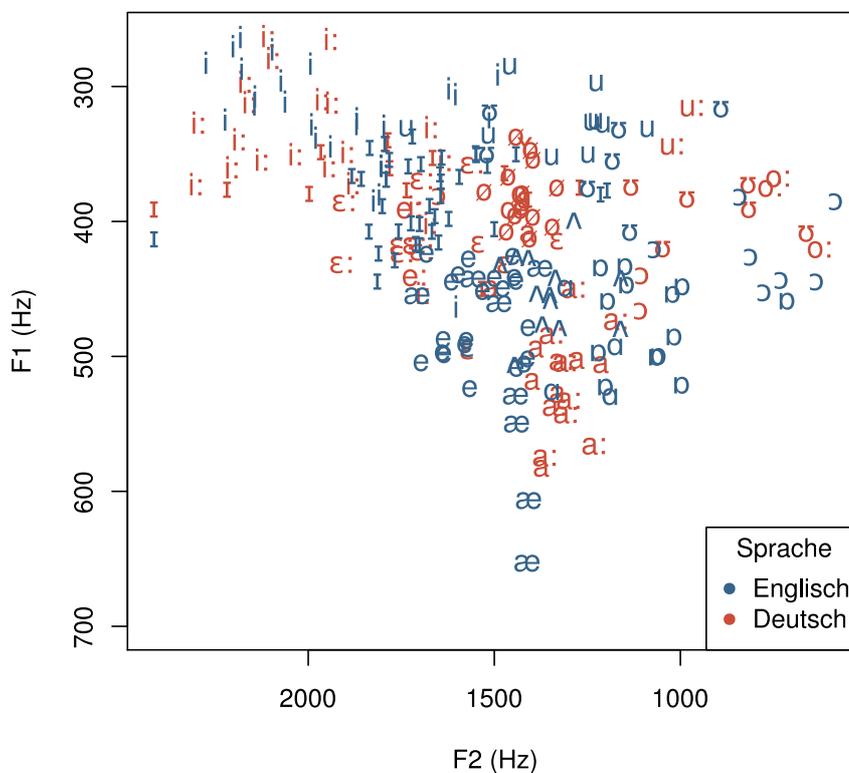
4.1.4.2 Der männliche Sprecher cc

Bei dem Sprecher *cc* überlappen sich britische und deutsche Vokale stark, im Gegensatz zum Sprecher *ai*.

Die F_1 -Werte für die britischen /i:/-Abschnitte liegen niedriger, als bei den deutschen Abschnitten.

Allerdings sind wie bei *ai* auch die F_2 -Werte für /ɪ/ niedriger und für /ɔ/ höher als bei den britischen Vokalen, was bedeutet, dass diese Vokale in gelesenen deutschen Texten zentraler ausgesprochen werden.

Abbildung 4.12: Vokale des zweisprachigen männlichen Sprechers cc



Der Vokal /u:/ wird hier wie bei dem Sprecher *ai* in den deutschen gelesenen Texten mit einem niedrigeren Wert für F_1 artikuliert. Dies könnte dadurch bedingt sein, dass /u/ und /ʊ/ im Deutschen auch durch eine längere bzw. kürzere Vokaldauer abgegrenzt werden. So würde /u:/ als gelängter Vokal vor allem in betonten Silben extremer artikuliert werden.

Der deutsche Vokal /a:/ besitzt für höhere F_1 -Werte überlappende Bereiche mit dem englischen /æ/, d.h. in betonten Silben, die extremer artikuliert werden, könnte *cc* /a:/ ähnlich wie ein englisches /æ/ aussprechen. Für niedrigere F_1 -Werte fällt /a:/ wie bei *ai* mit /ʌ/ zusammen.

In den hier untersuchten Werten sind leider keine Messungen für /y:/, jedoch einige für /ø:/ vorhanden. Für diesen Vokal werden ein höherer F_1 -Wert und zentralere F_2 -Werten gemessen, d.h. er wird eher wie ein deutsches /ø/ ausgesprochen.

Die interessantesten Ergebnisse erhalte ich hier für die Vokale /o:/ und /ɔ/. Der im Englischen unbekannt Vokal /o:/ wird wie ein englisches /ɔ/ gesprochen. Dafür fällt das deutsche /ɔ/ mit /ɒ/ zusammen. Dies könnte bedeuten, dass der Sprecher den Kontrast zwischen /o:/ und /ɔ/ hört, sie allerdings durch britische muttersprachliche Vokale ersetzt.

4.2 Gemeinsame Vokale

Untersuchungsgegenstand 1 (Aussprache gemeinsamer Vokale) *In diesem Unterkapitel wird untersucht, ob die in beiden Sprachen vorhandenen Vokale /ɪ/ und /ʊ/ von deutschen und britischen Versuchspersonen gleich oder unterschiedlich artikuliert werden.*

4.2.0.3 Begründung

In (Flege & Hillenbrand, 1984) wird der Erwerb der französischen Vokale /u/ und /y/ durch amerikanische Sprecher beschrieben. Das /y/ als neuer Laut wurde nach einigen Jahren gelernt, die Aussprache des /u/, die sich leicht unterschied, jedoch beibehalten. Deshalb würde ich hier gerne überprüfen, ob dies auch für fortgeschrittene britische Deutschlernende zutrifft.

(Disner, 1983) vergleicht die Vokalproduktion bei zweisprachigen Sprechern für verschiedene germanische Sprachen, unter anderem Englisch und Deutsch. Es handelt sich dabei um Personen, die beide Sprachen fließend beherrschen, allerdings haben einige davon ihre Englischkenntnisse erst während der Pubertät bzw. als Erwachsene erworben, was sich auch in den Bewertungen ihrer Aussprache innerhalb der Untersuchung niederschlägt.

Allgemein zeigt sich in ihren Messungen vor allem bei kurzen Vokalen ein deutlicher Unterschied zwischen isoliert und innerhalb eines konsonantischen Kontexts aufgenommenen Vokalen. Letztere verschieben sich in Richtung einer neutraleren, weniger peripheren Position. Diese Unterschiede werden v.a. bei hohen bzw. vorderen gerundeten Vokalen signifikant, bei /ɑ/, /o/, /e/ und /ɛ/ sind die Unterschiede nicht signifikant (s. S. 31).

Vergleicht man Werte aus vorherigen Untersuchungen für gemeinsame Monophthonge im Englischen und Deutschen für /i/, /u/ und /ɛ/, so ist /u/ im Englischen weniger gerundet bzw. weiter vorne, da es eine signifikant höhere F_2 als das deutsche /u/ besitzt. Der Vokal /ɛ/ ist im Englischen signifikant höher als im Deutschen, für /i/ ergeben sich keine signifikanten Unterschiede (s. S. 34).

Die Formantwerte von englisch-deutschen zweisprachigen Sprechern fallen für alle deutschen Vokale extremer aus als für die englischen. Das /u/ als hinterer Vokal wird bei deutschen Sprechern stärker gerundet (s. S.90).

Bei bilingualen amerikanischen Sprechern ist /ɛ/ im Englischen nicht niedriger als das deutsche /ɛ/, bei britischen fällt es niedriger aus. Für /i/ fallen die Werte bei deutschen Vokalen für F_2 etwas niedriger, für F_3 wesentlich höher aus als bei englischen Vokalen (s. S.91).

Das vermutete Prinzip der Streuung der Vokale (vowel dispersion) trifft nicht zu: manchmal liegen Formantwerte für verschiedene Vokale eng zusammen, während andere Bereiche des Vokalraums kaum belegt sind. Und selbst bei Sprachen, die die gleiche Anzahl an Vokalen besitzen, wie Norwegisch und Schwedisch, sind deren Formantwerte nicht auf die gleiche Art und Weise verteilt. Niederländisch besitzt niedrigere und keine höheren Werte bei F_2 und F_3 für /i/ als Englisch, obwohl im Niederländischen dem /i/ ein gerundetes

/y/ gegenübersteht (s. S.115).

Die Unterschiede zwischen ungerundeten und gerundeten Vokalen im Deutschen sind extremer als im Englischen (v.a.) bei F_2 und F_3 (s. S. 121).

4.2.1 Verteilungsfreie Tests auf allen Testmengen

Tabelle 4.1: Ergebnisse des Wilcoxon-Rangsummentests für alle im Deutschen und Englischen vorhandenen untersuchten Vokale (p-Werte)

Untersuchte Daten	F_1	F_2	F_3	Vokaldauer
/i:/, männliche Sprecher	0.9924	0.0245	0.7257	0.0502
/i:/, weibliche Sprecher	3.764 * 10⁻⁵	0.0060	0.9829	0.0201
/ɪ/, männliche Sprecher	0.1714	0.8661	0.1381	0.4004
/ɪ/, weibliche Sprecher	6.417 * 10⁻⁵	0.1305	0.7054	0.8284
/u:/, männliche Sprecher	0.3527	0.8534	0.6842	0.2260
/u:/, weibliche Sprecher	0.0931	0.0649	0.2403	0.6623
/ʊ/, männliche Sprecher	0.07802	0.4095	0.01449	0.5251
/ʊ/, weibliche Sprecher	0.0973	0.0070	0.6200	0.7914

Die signifikanten Unterschiede für /i:/ verschwinden, sobald gemischte Modelle mit dem einzelnen Sprecher als Zufallsfaktor eingesetzt werden und lassen sich daher durch sprecherspezifische Unterschiede erklären.

Für /u:/ zeigen sich auch in einem gemischten Modell mit dem einzelnen Sprecher als Zufallsfaktor keine signifikanten Unterschiede.

4.2.2 Testmengen für /ɪ/, männliche Sprecher

Hier werden jeweils 45 Stichproben für jede Muttersprache verglichen, wobei aus den britischen Messdaten 45 von 46 Abschnitten ausgewählt werden. Im Hinblick auf die Testmenge für F_1 , sind die Werte für britische Muttersprachler nicht normalverteilt. Außerdem unterscheiden sich die Varianzen beider Gruppen.

Die Testmengen für F_2 und F_3 sind normalverteilt und besitzen gleiche Varianzen.

4.2.2.1 Eigenschaften des gemischten Modells

Die Anpassung des Modells an die bestehenden Daten fällt hier mit einem BIC-Wert von über 1000 Punkten bei den ersten drei Formanten relativ ungenau aus.

Der einzelne Sprecher als Zufallsfaktor wirkt sich mit 46,8% der Abweichungen am stärksten auf F_3 aus. Die Auswirkungen auf F_1 sind mit 21,7% am geringsten.

Tabelle 4.2: Ergebnisse eines gemischten Modells für /ɪ/ bei deutschen und britischen männlichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	0.106	0.9176	0.9157
F_2	-0.22	0.8238	0.8277
F_3	-0.82	0.3118	0.4118
Vokaldauer	-0.625	0.4814	0.5335

Tabelle 4.3: Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und britischen männlichen Sprechern

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	368	366
F_2 (Hz)	1799	1815
F_3 (Hz)	2440	2531
Vokaldauer (ms)	48,25	52,12

Wie im gemischten Modell für /i:/ zeigen sich für /ɪ/ bei den männlichen Sprechergruppen insgesamt keine signifikanten Unterschiede.

Die durchschnittlichen Werte für F_1 und F_2 in diesem gemischten Modell bei deutschen und britischen Sprechern stehen noch näher als die unter 4.7 errechneten Durchschnittswerte.

Nur für die Vokaldauer höre ich bei der Resynthese der durchschnittlichen Werte einen Unterschied.

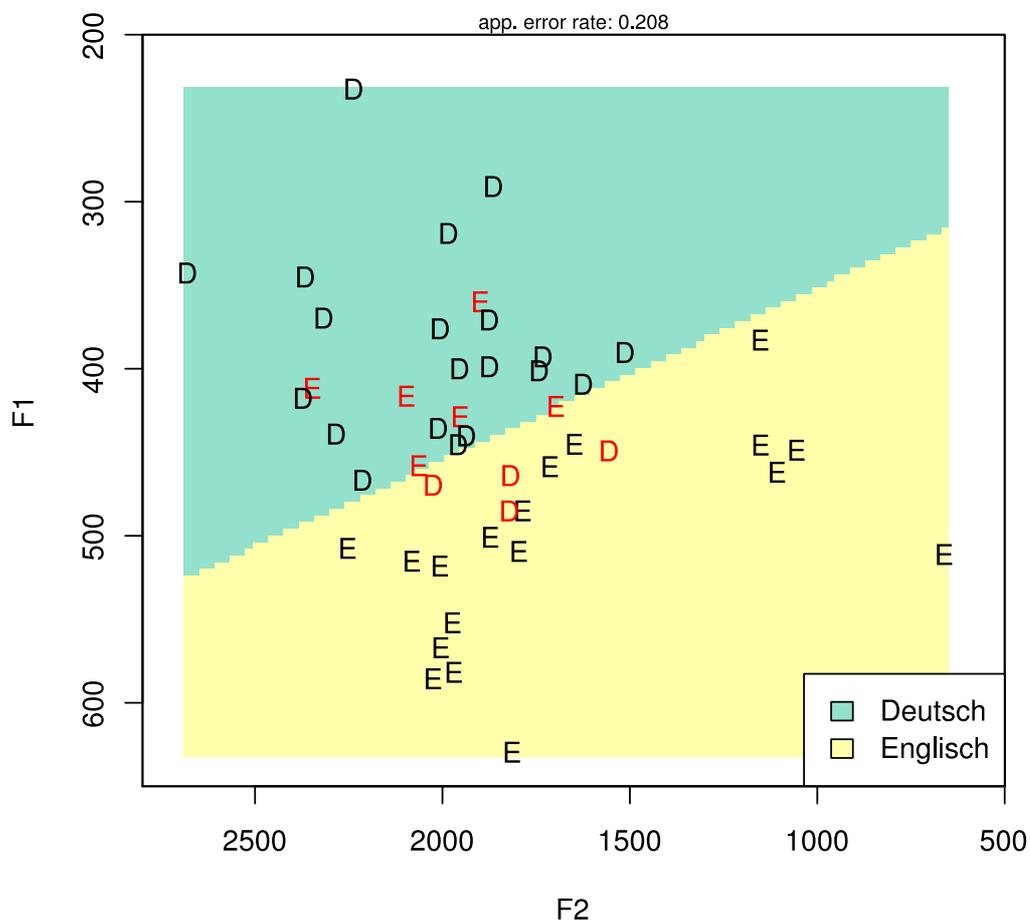
Da die Vokaldauern für /i:/ bei 64,53 ms für deutsche Sprecher und bei 72,07 ms für britische männliche Sprecher liegen, könnten die längeren Werte für die Vokaldauer bei /i:/ und /ɪ/ in diesem Fall durch ein aufmerksameres und deutlicher artikuliertes Lesen der nicht-muttersprachlichen Gruppe bedingt sein.

4.2.3 Testmengen für /ɪ/, weibliche Sprecher

Für diese Untersuchung vergleiche ich 24 Abschnitte pro Sprechergruppe, wobei ich aus der deutschen Gruppe 24 von 32 Stichproben auswähle. Die Testmengen für /ɪ/ bei weiblichen deutschen und britischen Sprecherinnen sind alle normalverteilt, die Varianzen sind jedoch für die Vokaldauer nicht homogen.

4.2.3.1 Diskriminanzanalyse

Um zu sehen, wo die im Rangsummentest festgestellten Unterschiede für F_1 genau liegen, führe ich eine Diskriminanzanalyse (rda) durch:

Abbildung 4.13: Verteilung der Formantwerte von F_1 und F_2 für /ɪ/ bei weiblichen Sprechern

Hier sind die Formantwerte bei den Nichtmuttersprachlern für F_1 tendenziell höher und für F_2 tendenziell niedriger. Damit entsprechen die Ergebnisse für F_1 denen für britische und deutsche /ɪ/-Laute im vorherigen Kapitel.

Die Fehlerrate von 20,8% ist durch einen mittleren gemeinsamen Bereich von ca. 360-480 Hz für F_1 und 1600-2400 Hz. für F_2 bedingt, in dem sich die Messwerte für Muttersprachler und britische Deutschler überlagern. Außerhalb dieses Bereichs finden sich peripherer artikulierte deutsche und zentraler artikulierte britische Abschnitte für /ɪ/.

Da sich in jeder Untermenge 19 betonte und 5 unbetonte Abschnitte finden, kann ich eine Verzerrung durch unterschiedlich viele betonte Silben ausschließen.

4.2.3.2 Eigenschaften des gemischten Modells

Die BIC-Werte liegen nur knapp über den AIC-Werten, was bedeutet, dass das gemischte Modell mit dem einzelnen Sprecher als Zufallsfaktor nicht zu komplex ausfällt. Insgesamt liegen die AIC-Werte bei ca 55% der AIC-Werte des Modells für die männlichen Sprecher, d.h. das gemischte Modell ist deutlich besser an die gemessenen Daten angepasst.

Im Gegensatz zu den männlichen Sprechern wirkt sich hier der einzelne Sprecher als Zufallsfaktor auf F_1 mit 40,8% am stärksten und auf F_3 17,8% am geringsten aus.

Tabelle 4.4: Eigenschaften des ermittelten gemischten Modells für /ɪ/ bei weiblichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	-2.091	0.0468	0.0421
F_2	1.595	0.1430	0.1175
F_3	0.51	0.6264	0.6132
Vokaldauer	0.351	0.6856	0.7273

Im Gegensatz zum vorher durchgeführten t-Test zeigen sich hier für F_2 keine signifikanten Unterschiede für deutsche und britische weibliche Sprecher mehr. Da die ermittelten Durchschnittswerte für den t-Test und das gemischte Modell sich kaum unterscheiden, ist die im gemischten Modell geringere Streuung der Messwerte für das Verschwinden des im t-Test festgestellten Unterschiedes verantwortlich.

Der ermittelte Werteintervall für F_1 umfaßt bei deutschen Sprecherinnen 102 Hz, von 350 Hz bis 452 Hz, bei britischen Sprecherinnen fällt er mit 122 Hz, von 422 bis 544 Hz, umfangreicher aus.

Das gleiche gilt für F_2 , wo der Intervall für deutsche Sprecherinnen zwischen 1754 Hz und 2245 Hz liegt und für britische Sprecherinnen zwischen 1478 Hz und 2022 Hz.

Damit kann ich davon ausgehen, dass /ɪ/ von deutschen Sprecherinnen extremer artikuliert wird als von britischen, möglicherweise um es von /y/ abzugrenzen.

Das von britischen Sprecherinnen produzierte /ɪ/ klingt bei der Resynthese wie ein verkürztes deutsches /ɛ:/, was zumindest für F_1 und F_2 mit dem direkten Vergleich der muttersprachlichen Werte übereinstimmt, s. S. 75.

Tabelle 4.5: Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und britischen weiblichen Sprechern

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	402	484
F_2 (Hz)	1988	1752
F_3 (Hz)	2679	2625
Vokaldauer (ms)	64,30	60,00

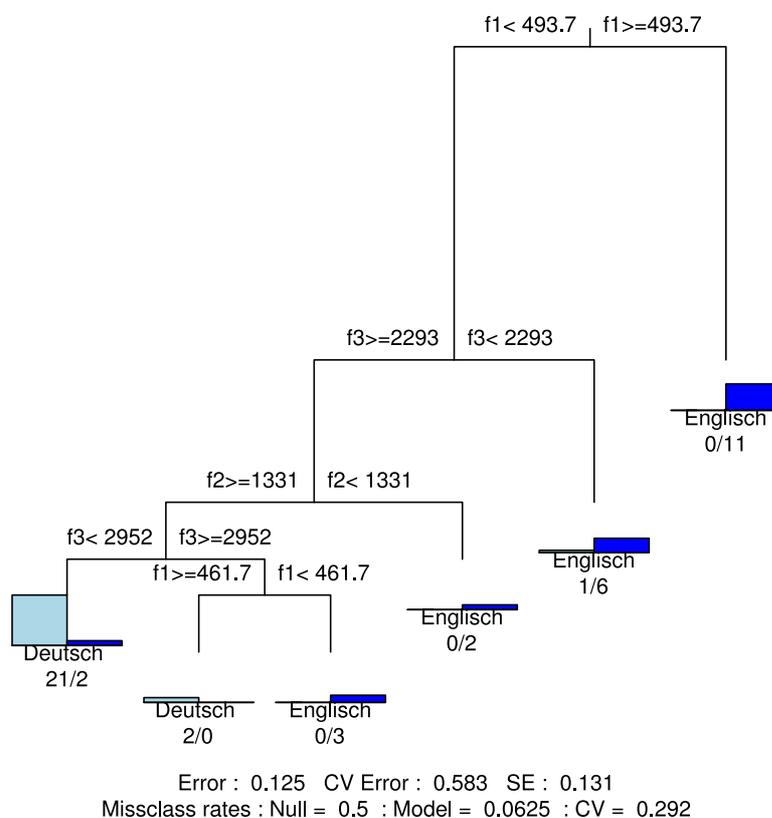
4.2.3.3 Klassifikationsbäume

Um doch noch eine eindeutige Antwort zu erhalten, wende ich Klassifikationsbäume auf die Formantwerte der Sprecher an.

Bei den männlichen Sprechern erhalte ich für eine maximale Baumtiefe von 10 Ästen eine relativ ungenaue Klassifikation mit einer Fehlerrate von 33%, bei der nur einige Vokalabschnitte britischer Sprecher durch höhere F_3 - und niedrigere F_1 -Werte auffallen.

Für die weiblichen Versuchspersonen dagegen erhalte ich eine relativ eindeutige Klassifikation mit einer Fehlerrate von 12,5% bei einer maximalen Baumtiefe von 7 Ästen. Hier

Abbildung 4.14: Klassifikationsbaum für die Formantwerte von /ɪ/ bei weiblichen deutschen und britischen Sprechern



befinden sich die meisten deutschen F_1 -Werte innerhalb eines engeren Bereichs von 461 bis 493 Hertz, während die F_1 -Werte der britischen Sprecher entweder in knapp der Hälfte der Fälle höher liegen oder in einem Achtel der Fälle niedriger. Bei den F_3 -Werten liegen die deutschen Werte wieder ebenfalls innerhalb eines Bereichs von 2293 bis 2952 Hertz während 37,5% der britischen Werte außerhalb dieses Bereichs liegen.

Die teilweise niedrigeren F_2 -Werte stimmen mit dem vorher durch eine Diskriminanzanalyse ermittelten Modell überein.

Die Vokaldauer taucht in den errechneten Klassifikationsbäumen nicht auf, obwohl sie als möglicher Parameter mit angegeben wurde.

Insgesamt unterscheiden deutsche und britische Sprecher deutlich zwischen den in beiden Sprachen vorhandenen Vokalen /i:/ und /ɪ/. Mögliche Unterschiede, die nicht von

den individuellen Sprechereigenschaften der jeweiligen Versuchsgruppen abhängig sind, zeigen sich höchstens im spektralen Bereich, wo die Anzahl möglicher Formantwerte bei deutschen Sprechern durch die zusätzlich vorhandenen gerundeten Vokale /y:/ und /ɤ/ eingeschränkt ist.

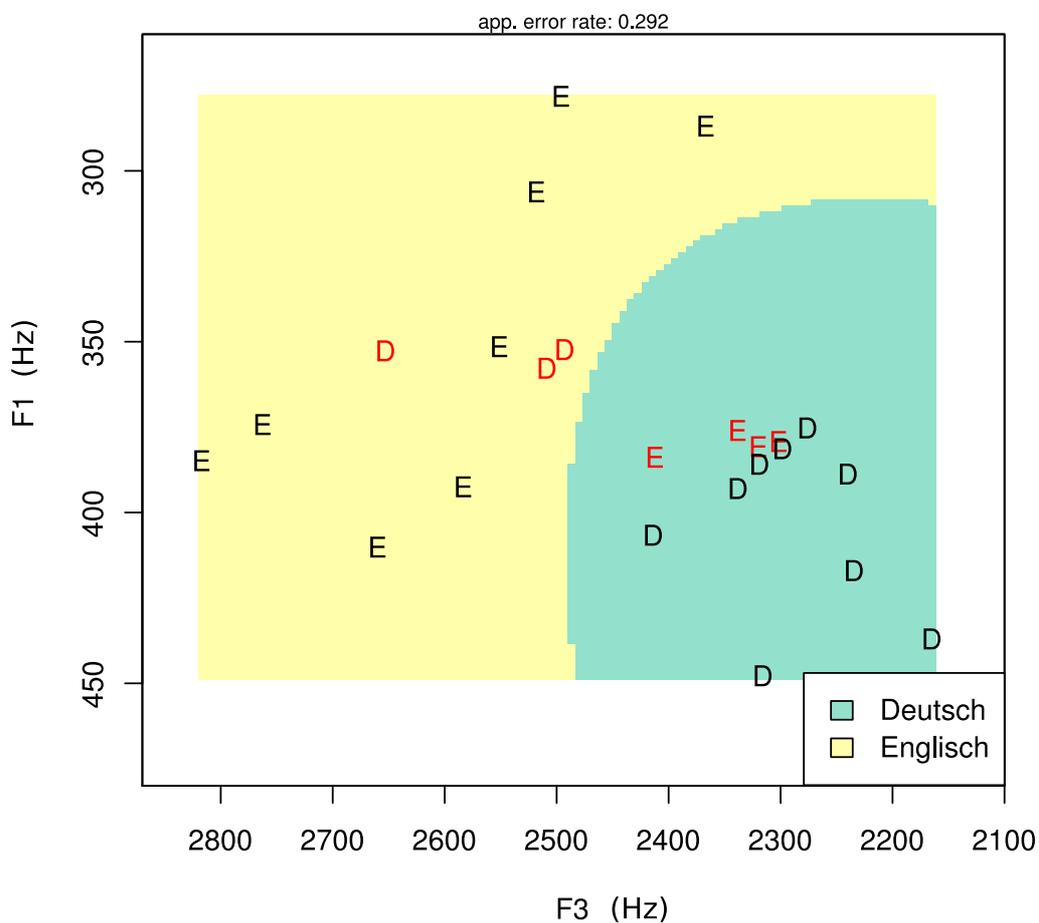
4.2.4 Testmengen für /u/, männliche Sprecher

Die Testmengen sind alle normalverteilt und besitzen ähnliche Varianzen. Für jede Sprechergruppe werden 12 Abschnitte untersucht. Dabei werden bei den britischen Muttersprachlern 12 von 15 Stichproben zufällig ausgewählt.

4.2.4.1 Diskriminanzanalyse für /u/ bei deutschen und britischen männlichen Sprechern

Wende ich eine Diskriminanzanalyse (rda) auf die Testmenge an, sind die Formantwerte für F_1 und F_3 weiter gestreut als bei Muttersprachlern. F_1 tendiert dabei zu tieferen, F_3 zu höheren Werten.

Abbildung 4.15: Verteilung der Werte für /u/ bei männlichen Sprechern



Die hohe Fehlerrate von 29,2% ist durch einige britische Werte im „deutschen“ Bereich und durch einige deutsche Abschnitte mit einem höheren F_3 -Wert bedingt.

Da es sich in diesem Fall um einzelne Werte von insgesamt acht Sprechern handelt, kann der Einfluß individueller Sprechercharakteristika ausgeschlossen werden.

Berechne ich die Durchschnittswerte für F_3 in betonten und unbetonten Silben für deutsche (7 betonte und 5 unbetonte Silben) und britische Sprecher (8 betonte und 4 unbetonte Silben), so betragen diese in betonten Vokalen 2364 Hz für die deutschen und 2575 Hz für die britischen Sprecher. In unbetonten Vokalen sind es 2340 für die deutschen und 2379 für die britischen Versuchspersonen.

4.2.4.2 Eigenschaften des optimalen gemischten Modells für /u/ bei männlichen Sprechern

Die niedrigen BIC-Werte zwischen 238 und 299 Punkten und der sehr geringe Abstand zwischen BIC- und AIC-Werten sprechen für eine sehr gute Anpassung der Daten an ein sehr kompaktes Modell.

Hier fällt auf, dass der einzelnen Sprecher als Zufallsfaktor für F_2 und die Vokaldauer kein nennenswertes Gewicht besitzt. Die Auswirkungen des Zufallsfaktors sind bei F_3 mit 45,5% am stärksten sichtbar.

Tabelle 4.6: Ergebnisse für das gemischte Modell für /u/ bei männlichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	1.56	0.0748	0.133
F_2	1.106	0.3172	0.2805
F_3	-1.52	0.0630	0.1417
Vokaldauer	0.80	0.4968	0.4322

Vergleicht man die Ergebnisse des gemischten Modells mit dem vorhergehenden t-Test, so liegen jetzt sowohl die Unterschiede für F_1 als auch für F_3 über der Signifikanzgrenze, was eine direkte Auswirkung des Zufallsfaktors sein dürfte, der für beide Parameter einen sichtbaren Anteil der Abweichungen abdeckt. Damit sind die vorher gemessenen Unterschiede zwischen den Formantwerten teilweise durch individuelle Sprechercharakteristika bedingt.

Tabelle 4.7: Durchschnittswerte des gemischten Modells für /ʊ/ bei deutschen und britischen männlichen Sprechern

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	386	350
F_2 (Hz)	993	928
F_3 (Hz)	2369	2513
Vokaldauer (ms)	56,25	51,08

Resynthetisiere ich die Durchschnittswerte, so erhalte ich für die britischen Sprecher einen Vokal, der eher wie ein deutsches /u:/ klingt. Dies stimmt mit dem allgemeinen Vergleich der Formantfrequenzen überein, in dem die F_1 -Werte für das britische /ʊ/ und das deutsche /u:/ etwa auf derselben Ebene lagen.

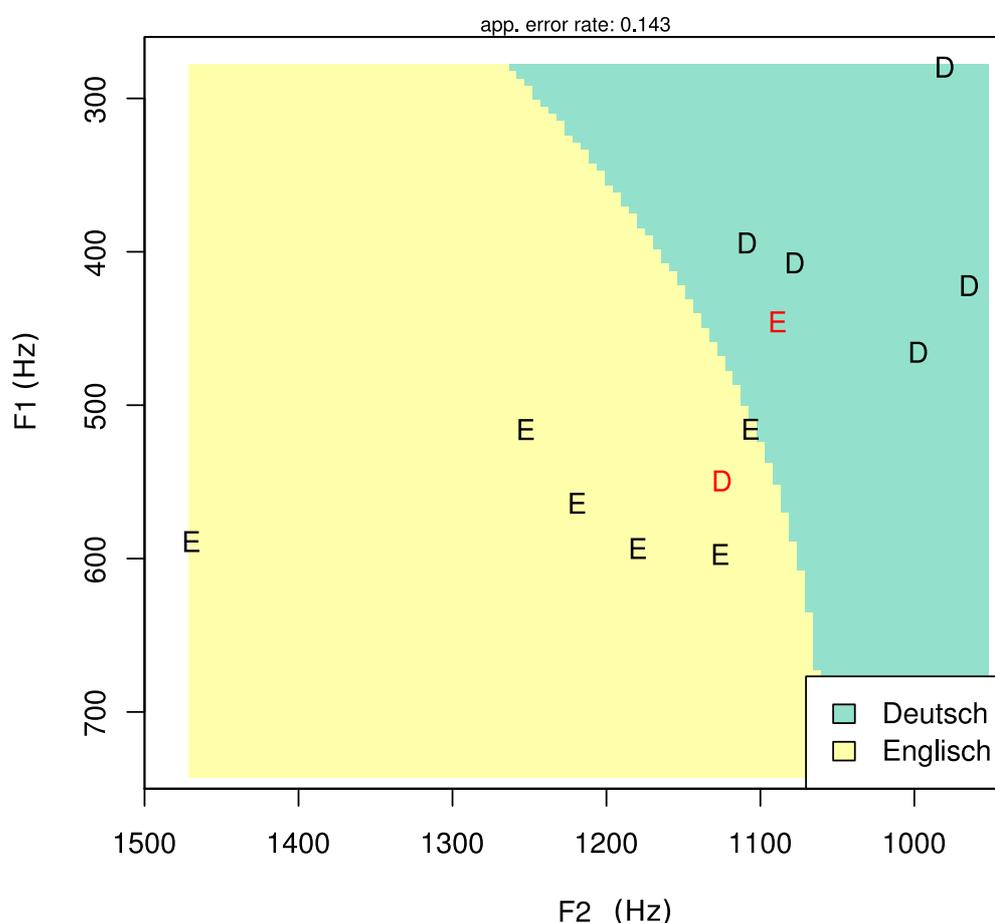
4.2.5 Testmengen für /u/, weibliche Sprecher

4.2.5.1 Diskriminanzanalyse

Für diese Untersuchung habe ich 7 Abschnitte pro Sprechergruppe zur Verfügung. Aus der deutschen Gruppe wähle ich dabei 7 von 14 Stichproben aus. In diesem Fall sind die Testmengen alle normalverteilt, die Werte für F_1 besitzen jedoch bei deutschen und britischen Sprecherinnen unterschiedliche Varianzen.

Da die Werte für F_1 und F_2 beide normalverteilt sind, bei den Werten für F_1 jedoch die Varianzen nicht übereinstimmen, führe ich eine Diskriminanzanalyse durch (rda). Die

Abbildung 4.16: Verteilung der Formantwerte für /u/ bei weiblichen Sprecherinnen



geringere Fehlerrate von 14,3% spricht für eindeutigere Unterschiede, kann jedoch auch durch die kleineren verglichenen Datenmengen bedingt sein.

/u/ wird hier von britischen Sprecherinnen zentraler artikuliert als von deutschen, wobei die extremeren deutschen Formantwerte der Abgrenzung zu /ø:/ und /y:/ dienen könnten.

4.2.5.2 Eigenschaften des gemischten Modells

Die AIC- und BIC-Werte sind hier noch geringer als in der männlichen Versuchsgruppe, genau wie der Abstand zwischen den Kennzahlen, was angesichts der üblicherweise größeren Streuung weiblicher Messwerte beachtenswert ist.

Interessanterweise hat der einzelne Zufallsfaktor hier keinen Einfluss auf F_1 und die Vokaldauer, während er sich auf F_2 mit 66,1% und F_3 mit 60,9% umso stärker auswirkt.

Tabelle 4.8: Eigenschaften des gemischten Modells für /u/ bei weiblichen Sprecherinnen

Parameter	t-Wert	pMCMC	p-Wert
F_1	-1.366	0.2382	0.1971
F_2	-2.211	0.0156	0.0472
F_3	-0.081	0.6606	0.9368
Vokaldauer	-1.608	0.1968	0.1338

Da der signifikante Unterschied für F_2 trotz eines starken Zufallsfaktors zustande kommt, kann man davon ausgehen, dass er tatsächlich durch sprachliche Unterschiede begründet ist.

Tabelle 4.9: Durchschnittswerte des gemischten Modells für /u/ bei deutschen und britischen weiblichen Sprechern

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	466	547
F_2 (Hz)	1032	1301
F_3 (Hz)	2770	2794
Vokaldauer (ms)	62,29	73,43

Vom Höreindruck her liegen die durchschnittlichen Werte des britischen /u/ hier nahe an einem deutschen /ø:/ mit einer etwas kürzeren Vokaldauer, was durch den direkten Vergleich von deutschen und britischen muttersprachlichen Formantwerten bestätigt wird, s. S. 75.

Der ermittelte Wertebereich reicht bei deutschen Sprecherinnen für F_2 von 925 Hz bis 1133 Hz und bei britischen Sprecherinnen von 1121 Hz bis 1380 Hz.

Die Vokaldauer fällt für /u:/ bei deutschen Sprecherinnen mit 75 ms und bei britischen Sprecherinnen mit 89,6 ms jeweils deutlich länger aus.

4.2.6 Zusammenfassung

Die Vokale /i:/, /I/, /u:/ und /U/ werden von deutschen und britischen Sprechern sehr ähnlich ausgesprochen. Für die Vokale /I/ und /U/, die häufiger sowohl in betonten als

auch in unbetonten Silben auftauchen, können Ausreißer v.a. bei weiblichen Sprecherinnen die ermittelten Durchschnittswerte stark beeinflussen.

4.3 Gespannte und ungespannte Vokale

Untersuchungsgegenstand 2 (Gespannte und ungespannte Vokale) *Am Beispiel der Vokale /e:/ und /ɛ/ wird untersucht, ob die Unterscheidung zwischen gespannten und ungespannten Vokalen in beiden Sprachen gleich umgesetzt wird.*

(Stevens, 1998, S. 294) beschreibt die artikulatorischen Eigenschaften gespannter und ungespannter Vokale. Dabei nehmen die Artikulationsorgane bei gespannten Vokalen eine extremere Position ein. Durch die nach vorne geschobene Zungenwurzel weitet sich der Rachenraum und der Zungenrücken wird angehoben.

Bei ungespannten Vokalen nehmen die Artikulationsorgane zentrale, weniger extreme Positionen ein. Da der Zungenrücken gesenkt bleibt, ist der Rachen weniger geweitet und der Mundraum größer.

Ansonsten unterscheiden sich gespannte und ungespannte Vokale in folgenden Punkten:

Vokaldauer Gespannte Vokale zeichnen sich in vielen Sprachen durch eine längere Vokaldauer aus als ungespannte Vokale.

Diphthongisierung Der Formantverlauf gespannter Vokale weist Übergänge zu einer extremeren artikulatorischen Position auf.

Stellung des Vokaltrakts Die Amplituden der spektralen Peaks oberhalb von F_1 sind bei ungespannten Vokalen höher. Wenn bei gespannten Vokalen der Vokaltrakt verengt wird, kommt es zu einer längeren Öffnungsphase des Kehlkopfs. Diese zeigt sich in einer geringeren Amplitude bei höheren Frequenzen und einer größeren Bandbreite für F_1 .

Die Unterschiede bei den Amplituden des Spektrums für gespannte und ungespannte Vokale können jedoch durch die Kehlkopfstellung ausgeglichen werden. Eine behauchte oder gedehnte Kehlkopfstellung führt in diesem Fall zu geringeren Amplituden in den höheren Frequenzbereichen, während eine gepresste oder zusammengezogene Kehlkopfstellung die Amplituden verstärkt.

4.3.1 Unterscheidung zwischen /e:/ und /ɛ/

4.3.1.1 Statistischer Vergleich

Dieser Vergleich wird mit jeweils 14 Abschnitten für /e:/ und /ɛ/, davon 14 aus 15 für /ɛ/, bei den deutschen männlichen Sprechern durchgeführt und mit 11 Abschnitten, davon 11 aus 16 für /ɛ/ bei den britischen männlichen Sprechern.

In den weiblichen Sprechergruppen setze ich bei den deutschen Sprecherinnen 8 und bei den britischen Sprecherinnen 6 Stichproben pro Gruppe ein. Dafür wähle ich in der deutschen Gruppe 8 aus 11 Stichproben für /ɛ/ aus.

Die Testmengen für /e:/ und /ɛ/ sind bei britischen und deutschen, männlichen und weiblichen Sprechern normalverteilt. Allerdings stimmen die Varianzen in der deutschen

Gruppe und bei britischen männlichen Sprechern bei F_1 und für die deutschen weiblichen Sprecherinnen bei F_3 nicht überein.

Tabelle 4.10: Wilcoxon-Rangsummentest zur Unterscheidung zwischen /e:/ und /ɛ/ bei deutschen und britischen, männlichen und weiblichen Sprechern

Parameter	Deutsche Muttersprachler		Britische Muttersprachler	
	Männer	Frauen	Männer	Frauen
F_1	0.0000009	0.00016	0.02331	0.24026
F_2	0.00003	0.01476	0.21695	0.13203
F_3	0.02412	0.19487	0.60632	0.48489
Vokaldauer	0.08890	0.63601	0.13709	0.69913

4.3.1.2 Gemischtes Modell für /e:/ und /ɛ/ bei deutschen männlichen Sprechern

Die Unterschiede bei den AIC- und BIC-Werten fallen mit 309 bis 382 Punkten relativ gering aus, wobei hier das geringe Gewicht des einzelnen Sprechers als Zufallsfaktor für F_3 mit 0% und für die Vokaldauer mit nur 18,3% auffallen.

Tabelle 4.11: Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei deutschen männlichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	6.766	0.0001	0
F_2	-5.985	0.0001	0
F_3	-2.44	0.0238	0.0217
Vokaldauer	-2.11	0.0476	0.0447

Die Ergebnisse für die deutschen Sprecher lassen einen deutlichen signifikanten Unterschied für alle gemessenen Werte erkennen, wobei vor allem für F_3 der Einfluss sonstiger koartikulatorischer Faktoren noch überprüft werden könnte (s.u).

Tabelle 4.12: Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für deutsche männliche Sprecher

Parameter	/e:/	/ɛ/
F_1 (Hz)	366	521
F_2 (Hz)	2027	1641
F_3 (Hz)	2668	2424
Vokaldauer (ms)	75,04	57,68

Die hier errechneten Durchschnittswerte fallen vor allem für F_2 und F_3 etwas extremer aus als die Medianwerte für gelesene Texte im KielCorpus. Dazu könnte beigetragen ha-

ben, dass beide Datenmengen überwiegend betonte Silben umfassen, für /e:/ sind es 14 betonte, für /ɛ/ 11 betonte und 3 unbetonte Silben.

4.3.1.3 Gemischtes Modell für /e:/ und /ɛ/ bei deutschen weiblichen Sprechern

Die Anpassung des gemischten Modells fällt deutlich besser aus als in der männlichen Gruppe und dies trotz der fehlenden Auswirkungen des Zufallsfaktors auf F_2 , F_3 und die Vokaldauer.

Tabelle 4.13: Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei deutschen weiblichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	5.361	0.0006	0.0002
F_2	-2.879	0.0140	0.0139
F_3	-0.39	0.7148	0.706
Vokaldauer	-1.678	0.1290	0.1191

Wie bei den männlichen Sprechern zeigen sich hier signifikante Unterschiede für F_1 und F_2 . Das Fehlen von Unterschieden für F_3 legt jedoch nahe, dass zusätzliche koartikulatorische Faktoren existieren und für die Vokaldauer könnten auch noch die Auswirkungen der Unterschiede zwischen betonten und unbetonten Silben eine Rolle spielen.

Tabelle 4.14: Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für deutsche weibliche Sprecher

Parameter	/e:/	/ɛ/
F_1 (Hz)	414	687
F_2 (Hz)	2170	1727
F_3 (Hz)	2872	2822
Vokaldauer (ms)	83,75	68,33

Die Werte für F_2 und F_3 fallen hier leicht niedriger aus als die Medianwerte im Kiel Corpus, dies könnte jedoch auch ein Ergebnis der kleinen Datenmenge sein.

4.3.1.4 Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei britischen männlichen Sprechern

Sobald ich ein gemischtes Modell mit dem einzelnen Sprecher als Zufallsfaktor erstelle, unterscheiden sich die Werte bei F_1 für /e:/ und /ɛ/ in beiden Gruppen nicht mehr signifikant.

Für die Erstellung komplexerer gemischter Modelle war der Umfang der Testmengen mit jeweils 11 Abschnitten bei den britischen männlichen und jeweils 6 Abschnitten bei den britischen weiblichen Sprechern zu gering.

Tabelle 4.15: Gemischtes Modell für die Unterscheidung zwischen /e:/ und /ɛ/ bei männlichen britischen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	1.382	0.1714	0.1823
F_2	-1.756	0.1082	0.0944
F_3	0.053	0.8840	0.9586
Vokaldauer	-1.748	0.0810	0.0957

Der in beiden Sprachen existierende Vokal /ɛ/ dessen Werte für F_1 und F_2 sich jeweils relativ nahe liegen könnte zum Fehlen signifikanter Unterschiede beigetragen haben.

Tabelle 4.16: Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für britische männliche Sprecher

Parameter	/e:/	/ɛ/
F_1 (Hz)	379	422
F_2 (Hz)	1980	1755
F_3 (Hz)	2548	2552
Vokaldauer (ms)	81,82	67,27

Bei der Resynthese ergibt sich hier kein großer Unterschied zwischen deutschem und britischem /e:/. Vergleicht man die Durchschnittswerte deutscher und britischer männlicher Sprecher direkt miteinander, so fallen die zentraleren Formantwerte und die längere Vokaldauer bei den britischen Teilnehmern auf.

4.3.1.5 Eigenschaften des gemischten Modells für /e:/ und /ɛ/ bei britischen weiblichen Sprechern

Tabelle 4.17: Gemischtes Modell für die Unterscheidung zwischen /e:/ und /ɛ/ bei weiblichen britischen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

	AIC	t-Wert	pMCMC	p-Wert
F_1	126.4	2.114	0.0794	0.0606
F_2	154.4	-1.562	0.154	0.1492
F_3	160.4	1.329	0.191	0.2132
Vokaldauer	-27.97	-0.217	0.8426	0.8327

Die Unterschiede für F_2 und die Vokaldauer bei männlichen Sprechern und vor allem für F_1 bei weiblichen Sprechern liegen geringfügig über der Signifikanzgrenze, sonst lassen sich keine relevanten Unterschiede feststellen.

Dies legt die Vermutung nahe, dass die britischen Sprecher Schwierigkeiten damit haben könnten, gleichzeitig den im Englischen unbekanntem Vokal /e:/ umzusetzen und auf den Gegensatz zwischen /e:/ und /ɛ/ zu achten.

Tabelle 4.18: Durchschnittswerte für /e:/ und /ɛ/ in gemischten Modellen für britische weibliche Sprecher

Parameter	/e:/	/ɛ/
F_1 (Hz)	542	621
F_2 (Hz)	2090	1803
F_3 (Hz)	2615	2981
Vokaldauer (ms)	92,11	86,73

In diesem Fall ergeben sich für die britischen Sprecherinnen wie in der männlichen Gruppe wieder zentralere Formantwerte und eine längere Vokaldauer. Die Unterschiede sind für /e:/ deutlich, für /ɛ/ jedoch gering.

4.3.2 Vergleich der Aussprache von /e:/ und /ɛ/ bei deutschen und britischen Sprechern

Hier untersuche ich für /e:/ jeweils 11 Stichproben aus der männlichen und 6 Stichproben aus der weiblichen Gruppe, für /ɛ/ sind es 15 Stichproben für die männlichen und 7 für die weiblichen Teilnehmer. Dabei wähle ich für /e:/ 11 von 14 Abschnitten aus der deutschen männlichen Gruppe sowie 6 von 8 aus der deutschen weiblichen Gruppe aus; für /ɛ/ sind es 15 von 16 aus der britischen männlichen Gruppe und 6 von 11 aus der deutschen weiblichen Gruppe.

Die Testmengen für /e:/ sind bei Männern und Frauen normalverteilt, bei den Männern unterscheiden sich die Varianzen für F_1 jedoch signifikant. Für /ɛ/ sind alle Testmengen normalverteilt und besitzen ähnliche Varianzen.

Tabelle 4.19: Wilcoxon-Rangsummentests zur Aussprache von /e:/ und /ɛ/ bei deutschen und britischen Sprechern

Parameter	/e:/		/ɛ/	
	Männer	Frauen	Männer	Frauen
F_1	0.60632	0.00216	0.00022	0.62004
F_2	0.84700	0.69913	0.74370	0.80478
F_3	0.51903	0.39394	0.09753	1
Vokaldauer	0.64548	0.58875	0.80314	0.1792339

Die Unterschiede für F_1 von /e:/ bei weiblichen Sprechern könnten sich als zutreffend erweisen, die der männlichen Sprecher eher ein Produkt individueller Sprecherabweichungen sein.

4.4 Unbekannte Vokale

4.4.1 Untersuchungsgegenstand

Untersuchungsgegenstand 3 (Der unbekannt Vokal /ø:/) *In diesem Unterkapitel untersuche ich, ob der im britischen Englisch nicht vorhandene Vokal /ø:/ von deutschen und britischen Sprechern gleich oder unterschiedlich ausgesprochen wird.*

4.4.2 Aussprache von /ø:/

4.4.2.1 Statistischer Vergleich

Für diese Vergleiche habe ich für britische und deutsche, männliche und weibliche Sprecher jeweils 13 Messwerte zur Verfügung. Dabei wähle ich aus der britischen männlichen Gruppe 13 von 21 Abschnitten aus und aus der deutschen weiblichen 13 von 26.

Für /ø:/ sind alle Testmengen normalverteilt, bei den weiblichen Sprecherinnen unterscheiden sich jedoch die Varianzen für F_1 .

Tabelle 4.20: Wilcoxon-Rangtest zu /ø:/ bei deutschen und britischen Sprechern

Parameter	Männer	Frauen
F_1	0.000004	0.00004
F_2	0.02560	0.01017
F_3	0.50491	0.31073
Vokaldauer	0.01108	0.05764

Die Unterschiede für F_1 und F_2 sind hier bei beiden Geschlechtern eindeutig, der Wert für die Vokaldauer liegt bei den weiblichen Sprecherinnen nur knapp über der Signifikanzgrenze, so dass auch dort von einem tatsächlichen Unterschied ausgegangen werden kann.

Leider existieren für /œ/ zu wenige Abschnitte, um zu vergleichen ob es eine Beziehung zwischen der Vokaldauer und der Unterscheidung von /ø:/ und /œ/ gibt.

4.4.2.2 Diskriminanzanalyse

Um die Unterschiede bei der Aussprache des Vokals /ø:/ deutlicher darzustellen, führe ich eine Diskriminanzanalyse (rda) durch. Dabei zeigen sich niedrigere F_1 -Werte und eine größere Streuung der F_2 -Werte in beiden Gruppen.

Zusätzlich berechne ich die durchschnittlichen Formantwerte und Vokallängen für jede Sprechergruppe.

Tabelle 4.21: Durchschnittliche Werte für die Aussprache von /ø:/ durch deutsche und britische männliche Sprecher

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	320	375
F_2 (Hz)	1454	1399
F_3 (Hz)	2325	2293
Vokaldauer (ms)	149,23	113,08

Die für die britischen männlichen Sprecher gemessenen Formantwerte tendieren stark in Richtung /ə/, obwohl es sich ausschließlich um betonte Vokale handelt.

Tabelle 4.22: Durchschnittliche Werte für die Aussprache von /ø:/ durch deutsche und britische weibliche Sprecher

Parameter	deutsche Sprecher	britische Sprecher
F_1 (Hz)	371	508
F_2 (Hz)	1590	1653
F_3 (Hz)	2479	2544
Vokaldauer (ms)	134,62	113,08

In diesem Fall tendiert die Aussprache der britischen Sprecherinnen in Richtung /ɛ/.

Abbildung 4.17: Verteilung der Werte von F_1 und F_2 bei / \emptyset :/ für britische männliche Sprecher

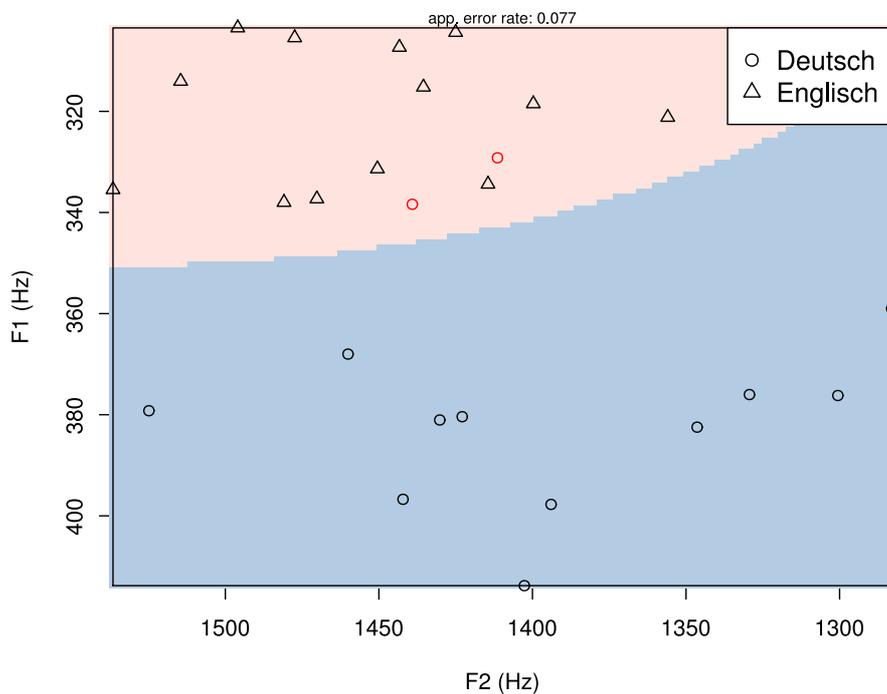
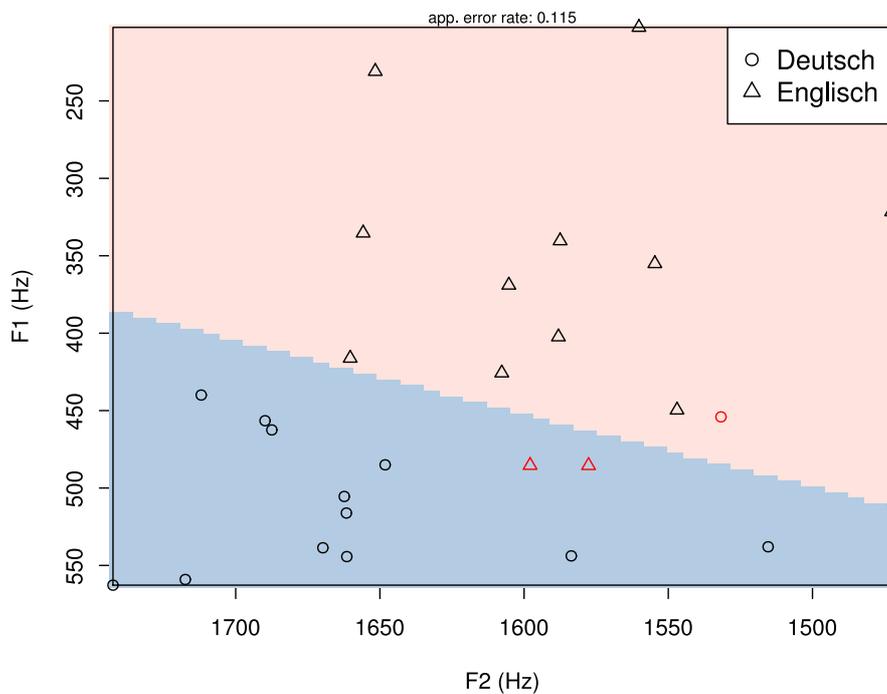


Abbildung 4.18: Verteilung der Werte von F_1 und F_2 bei / \emptyset :/ für britische weibliche Sprecher



4.4.3 Aussprache von /y:/

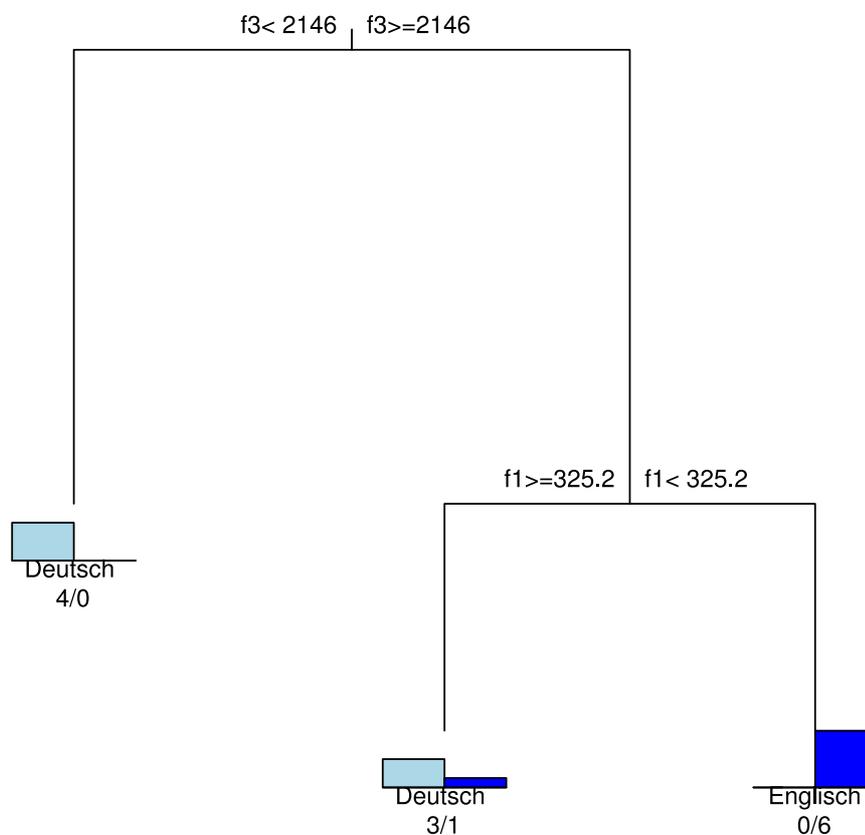
In diesem Fall können nur die männlichen Sprecher mit jeweils sieben Samples verglichen werden, da bei den weiblichen Sprechern nur zwei Samples zur Verfügung stehen.

Tabelle 4.23: Wilcoxon-Rangtest zu /y:/ bei deutschen und britischen Sprechern

Parameter	Männer
F_1	0.2086
F_2	0.7104
F_3	0.02622
Vokaldauer	1

Aufgrund der geringen Datenmenge erstelle ich kein lineares Modell, sondern einen Klassifikationsbaum, der ein eindeutiges Ergebnis liefert. Durch die hier deutlich sichtba-

Abbildung 4.19: Klassifikationsbaum für die Formantwerte von /y:/ bei männlichen deutschen und britischen Sprechern



Error : 0.143 CV Error : 0.857 SE : 0.265
 Missclass rates : Null = 0.5 : Model = 0.0714 : CV = 0.429

ren höheren F_3 - und niedrigeren F_1 -Werte bei den britischen männlichen Sprechern lässt sich zeigen, dass der Vokal /y:/ dort eher als /i:/ ausgesprochen wird.

4.4.4 Schlußfolgerung

Deutliche Unterschiede zeigen sich in dieser Gruppe vor allem bei den unbekanntem Vokalen, die teilweise in Annäherung an bekannte Vokale ausgesprochen, teilweise durch muttersprachliche Laute ersetzt werden.

Die bekannten Phoneme /i:/ und /u:/ werden wie im Deutschen umgesetzt. Der in beiden Sprachen vorhandene und sehr ähnliche Vokal /ɛ/ trägt zu einem ausreichenden Kontrast zum deutschen Vokal /e:/ bei, dessen muttersprachliche Formantwerte sich in der Nähe britischer /ɪ/-Laute befinden.

Für den Vokal /ʊ/ dagegen wirkt sich die andere Umsetzung des britischen /ʊ/ im Vergleich zum deutschen /ʊ/ aus, das mit einer weiter nach hinten bewegten Zunge artikuliert wird, um es von /ø:/ abzugrenzen.

5 Deutsche und italienische Sprecher

5.1 Einführung

In diesem Kapitel beschreibe ich zuerst die diachrone Entwicklung des Italienischen nach der Entstehung des Lateinischen als indogermanischer Sprache. Nach einer Beschreibung des italienischen Vokalraums heute und einem ersten visuellen Vergleich zwischen dem Vokalinventar italienischer Deutschlernender und deutscher Muttersprachler, stelle ich die Ergebnisse meiner statistischen Vergleiche für zwei relevante Unterschiede vor, die Aussprache von Vokalen vor Doppelkonsonanten und die Aussprache des im Italienischen unbekanntem Vokals /ø:/.

5.1.1 Sprachgeschichte

(Genot, 1984, S. 10ff) beschreibt die Entwicklung des Italienischen aus dem Lateinischen, bei der sich zuerst das allgemein gesprochene Vulgärlatein immer stärker von der klassischen literarischen Sprache entfernt. Aus dieser Form des Lateinischen entwickelt sich dann das Proto-Romanische und Dialekte wie das Toskanische oder das Sizilianische. Das Toskanische in der in Florenz gesprochenen und von Dante Aligheri verwendeten Form bildete die Grundlage des klassischen Italienisch.

Dieses bildet wiederum die Grundlage für das offizielle Italienisch und die gehobene Standardsprache. Die auf der Grundlage verschiedener Dialekte entstandene nationale Umgangssprache gehört heute zur gelockerten Standardsprache.

Das Ende des Vulgärlateinischen und die Entstehung der romanischen Sprachen findet zwischen dem 8. und 10. Jahrhundert nach Christus statt. Dabei wird 813 auf dem Konzil von Tours entschieden, dass Predigten in der „rustica romana lingua“ oder auf Deutsch gehalten werden sollen. 960 sind in der Charta des Klosters Monte Cassino Schwurformeln für Analphabeten in der Umgangssprache festgehalten (s. auch (Genot, 1984, S. 24).

Im Laufe der Entwicklung vom Vulgärlateinischen zum Italienischen ändert sich die Art der Betonung, kurze und lange Vokale fallen zusammen, das Vokalsystem wird neu strukturiert und Konsonanten werden stärker von ihrem artikulatorischen Umfeld beeinflusst. Das lateinische Vokalsystem besteht aus den fünf Vokalen *i*, *e*, *a*, *o* und *u* in jeweils zwei Längen und den häufig auftretenden Diphthongen *ae*, *au* und *oe*, sowie den selteneren Diphthongen *eu* und *ie* (s. (Genot, 1984)).

(Pertsch, 1999, S. 1310) beschreibt die Silbenbetonung im Lateinischen, bei der grundsätzlich drei Möglichkeiten existieren:

Die vorletzte Silbe wird betont, wenn sie einen langen Vokal enthält, z.B. in *abdōmen* (Unterleib), oder ein kurzer Vokal vor zwei oder mehr Konsonanten steht, wie in *alsus* (kühl) solange diese nicht aus einem Plosiv und einem Liquid bestehen (muta cum liquida).

Die letzte Silbe wird betont, wenn Partikel wie *-que*, *-ve* oder *-ne* an ein Wort angehängt werden, z.B. in *atque* (und).

Die drittletzte Silbe wird in allen anderen Fällen betont, wie in *fenestra* (Fenster).

Während es im Lateinischen jeweils betonte und unbetonte lange und kurze Vokale gibt, enthalten im Italienischen betonte Silben lange und unbetonte Silben kurze Vokale.

Tabelle 5.1: Lateinische und italienische Monophthonge

Latein	\bar{i}	$i \bar{e}$	e	$\bar{a} a$	o	$\bar{o} u$	\bar{u}
Italienisch	/i/	/e/	/ɛ/	a	/ɔ/	/o/	/u/

(nach (Genot, 1984, S. 38))

Die Veränderungen vom Lateinischen zum Italienischen lassen sich in folgenden Schritten beschreiben:

Lange Vokale im Lateinischen behalten im Italienischen ihre Vokalqualität.

i und u werden in betonten Silben geöffnet, außer vor /n/, /l/ und /nj/. So wird aus *silva selva* (Wald), während das erste *u* in *puntu* (von *punctum*) auch in *punto* (Punkt) erhalten bleibt.

e und o werden in offener Silbe innerhalb eines zweisilbigen Wortes diphthongisiert, wie z.B. in *pede*, das zu *piede* und *foco*, das zu *fuoco* wird. Stehen *e* und *o* jedoch nach einem Konsonanten, dem ein /r/ folgt, fallen die Diphthonge wieder zu Monophthongen zusammen, wie in *breve* → *brieve* → *breve* oder in *tropa* → *truova* → *trova*.

In dreisilbigen Worten und zweisilbigen Worten mit geschlossener Silbe bleiben *e* und *o* erhalten, z.B. in *opera*, *dente* oder *porta* sowie vor palatalisierten Konsonanten, z.B. bei *venio* → *vegno* oder *folia* → *foglia*.

y kommt in aus dem Griechischen übernommenen Worten vor und wird entweder wie *u* zu *o* z.B. bei *byrsa* → *bursa* → *borsa* (Tasche, Beutel) oder zu *i* wie in *crypta* → *cripta* oder *lynce* → *lince* (Luchs).

Unbetonte Vokale führen am Wortende zu einer Palatalisierung des vorherigen Konsonanten, z.B. in *fortis* → *fortia* → *forza* (Stärke) oder *prandium* → *prandiu* → *pranzo* (Mittagessen).

Vor einem betonten Vokal am Wortanfang oder nach einem betonten Vokal in der Wortmitte entfallen unbetonte Vokale wie z.B. in *bonitate* → *bontate*, *bont* (Güte),

cerebellu → *cervello* (Gehirn), sowie in *calidu* → *caldo* (warm), *solidus* (Goldmünze) → *solidu* → *soldo* (Geld) oder *lepore* → *lepre* (Hase).

Initiale Vokale entfallen, wo sie fälschlich *de*, *ad* und anderen Präpositionen zugeordnet werden, z.B. bei *abatissa* → *badessa*, *instrumento* → *strumento*, *elemosina* → *lemosina*.

Diphthonge *ae* wird zu *e* bzw. *ie*, wie in *maesto* → *mesto* und *caelu* → *cielo*.

au wird zu *o* z.B. in *causa* → *cosa* *thesauru* → *tesoro*.

Lehnwörter aus anderen Sprachen werden der italienischen Silbenstruktur durch die Einfügung von Vokalen angepasst, wie z.B. *blank* → *bianco* oder *blond* → *biondo* die während der Völkerwanderung übernommen werden, sowie *Landsknecht* → *lanzicheneco* und *bringdir's* → *brindisi* (Trinkspruch) im 16. Jahrhundert.

(s. (Genot, 1984))

5.1.2 Der italienische Vokalraum heute

(Zamboni, 1992, S. 48) gibt für das Italienische ein Vokalviereck mit folgenden Vokalen an:

Tabelle 5.2: Vokalviereck des heutigen Italienisch

/i/ 'vino	/u/ fu'tu:ro
/e/ 'vetta	/o/ 'rotto
/ɛ/	/σ/
/ɛ/ 'setta	/ɔ/ 'oro
/a/ pa'ta:ta	

Die Vokale /ɛ/ und /σ/ stehen dabei für dialektale Varietäten, in denen /e/ und /ɛ/ zu /ɛ/ zusammenfallen, sowie /o/ und /ɔ/ zu /σ/.

(Caldognetto, Zmarich & Cosi, 1998) untersucht die Bedeutung der visuellen Information beim Sprechen, also der Position der Oberlippe, der Unterlippe und des Kiefers für italienische Konsonanten und Vokale.

Folgende Parameter werden durch Sensoren gemessen:

lip height Abstand zwischen Ober- und Unterlippe

lip width Abstand der Sensoren am Ende der Lippen

protrusion Abstand jeweils zwischen Ober-/Unterlippe und dem Ohrläppchen

Die Position des Kiefers selbst wird jedoch nicht in die Untersuchung mit aufgenommen. An der Untersuchung nehmen sechs Versuchspersonen teil, davon zwei Männer und vier

Frauen.

Die Messungen zu italienischen Vokalen werden durchgeführt, indem jeweils fünfmal die sieben betonten Vokale, /a/, /e/, /ɛ/, /i/, /o/, /ɔ/ und /u/ und die fünf unbetonten Vokale /a/, /e/, /i/, /o/ und /u/ untersucht werden, wobei die betonten Vokale in der ersten Silbe eines zweisilbigen Wortes und die unbetonten in der ersten Silbe eines dreisilbigen Wortes innerhalb eines vorgegebenen Satzes gesprochen werden.

Für betonte Vokale werden dabei verschiedene charakteristische Merkmale deutlich. Im Hinblick auf den Abstand der Lippen (lip height) ergeben sich drei Gruppen:

/i,e,o,u/ geschlossene und geschlossene mittlere Vokale

/a,ɔ/ offene zentrale, mittlere bzw. hintere Vokale

/ɛ/ offener mittlerer vorderer Vokal

Bei der Weitung der Lippen (lip width) lassen sich /i,e,ɛ,a/ als ungerundete Vokale und /u,o,ɔ/ als gerundete Vokale deutlich voneinander unterscheiden.

Was die Hervorstülpung der Lippen (lip protrusion) betrifft, so kann man die betonten italienischen Vokale in vier Gruppen einteilen:

/a,ɛ/ offene mittlere vordere bzw zentrale Vokale bei denen die Lippen stark zurückgezogen werden.

/i,e/ geschlossene hohe bzw. mittlere vordere Vokale, mit leicht zurückgezogenen Lippen.

/ɔ/ ein offener mittlerer hinter Vokal mit leicht gerundeten Lippen.

/u,o/ geschlossene hohe oder mittlere hintere Vokale, bei denen die Lippen stark gerundet werden.

Die Ergebnisse für unbetonte Vokale sind weniger differenziert als bei betonten Vokalen. So fällt der Lippenabstand bei /a/ stärker aus als für /i,e,o,u/, die Weite der Lippen bei /i,e,a/ stärker als für /u,o/. Für die Hervorstülpung der Lippen lassen sich vier Gruppen wie bei den betonten Vokalen beschreiben : /a/, /i,e/, /o/ und /u/.

Insgesamt besteht für die Vokale eine Korrelation zwischen der Weite (lip width) und der Hervorstülpung der Lippen (lip protrusion).

Dies bedeutet für den Sprachvergleich eine möglicherweise extremere Artikulation der Vokale, auch wenn sie nicht isoliert ausgesprochen werden. Da nur hintere Vokale gerundet ausgesprochen werden, werden die vorderen gerundeten Vokale im Deutschen, also /ø:/ und /y:/ neu erworben.

5.1.3 Direkter Vergleich der Vokalräume

Die Messwerte von (Cosi, Ferrero & Vaggis, 1995) basieren auf Aufnahmen von 20 Männern und 20 Frauen aus Rom und der Toskana, die zu einem Corpus gehören, das

für das automatische Training von Spracherkennungsmodellen entwickelt wurde. Dabei wurden Äußerungen mit der Struktur $CV/ta/ /e/ /t/CV$ verwendet. Es handelt sich dabei also nicht um isolierte Vokale, jedoch auch nicht um solche, die in kontinuierlichem gelesenen Text untersucht wurden.

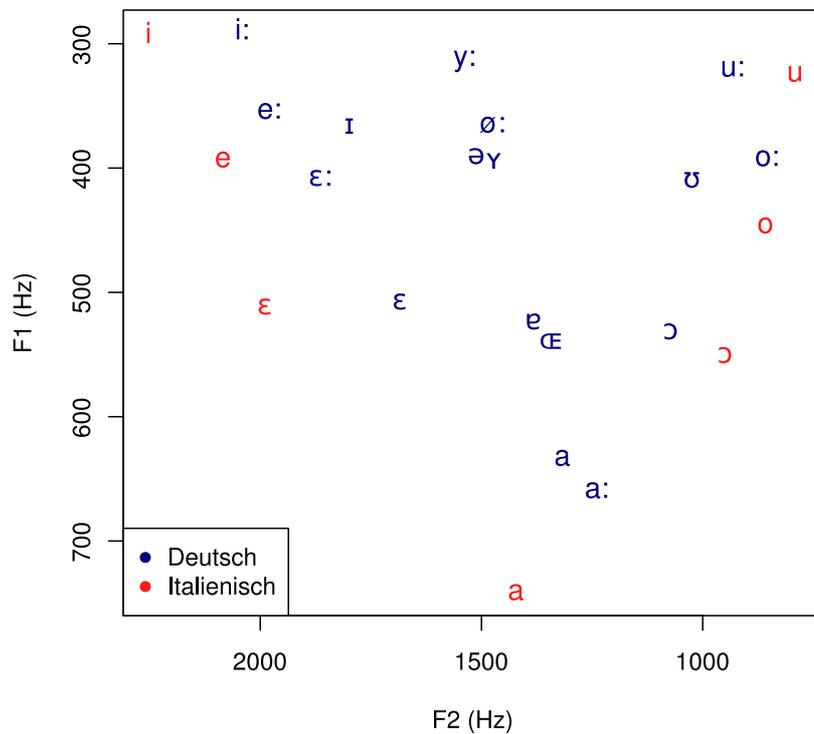
Vergleicht man die Messwerte des Kiel Corpus mit denen aus der Dokumentation von (Cosi et al., 1995), so nehmen letztere extremere Positionen ein, bei denen die F_1 -Werte von $/e/$ genau zwischen dem $/e:/$ und dem $/\varepsilon:/$ deutscher Muttersprachler liegen. Die F_2 -Werte für $/e/$ sind höher als bei $/e:/$ und $/\varepsilon:/$.

Die F_1 -Werte für $/a/$ sind bei italienischen Sprechern höher als bei deutschen. Alle anderen Vokale weisen geringere Unterschiede auf.

Alle Vokale, die von männlichen italienischen Sprechern aufgenommen wurden, besitzen höhere F_3 -Werte als bei den deutschen Sprechern, bei den weiblichen italienischen Sprechern ist dies nur für die vorderen Vokale $/i/$, $/e/$ und $/\varepsilon/$ der Fall.

Vergleiche ich die italienischen Referenzwerte in (Cosi et al., 1995) mit denen in (Ramers, 1988, S. 189-192) (s. auch 53), so dass die Messungen in beiden Fällen aus vergleichbaren Wortlisten stammen, so besitzt $/i/$ im Italienischen immer noch einen niedrigeren F_1 -Wert und einem höheren Wert für F_2 d.h. es wird extremer artikuliert. $/e/$ besitzt einen niedrigeren F_1 -Wert. Da es im Italienischen kein ungespanntes Gegenstück zu $/i/$ und kein Schwa gibt, könnte die deutlichere Artikulation dieser Laute vor allem durch Unterschiede in der Vokalqualität für betonte und unbetonte Silben verursacht sein, s. 114.

Die Formantwerte von F_1 für das italienische $/a/$ und das deutsche hyperartikulierte $/a:/$ entsprechen einander, während die Werte für F_2 für das italienische $/a/$ etwas höher und damit zentraler angesiedelt sind als für $/a:/$ im Deutschen. Die Formantwerte für $/\text{ɔ}/$ sind in beiden Sprachen identisch. Für $/u/$ und $/o/$ fallen die italienischen F_1 - und F_2 -Werte jeweils etwas höher und damit etwas zentraler aus als die deutschen Messwerte.

Abbildung 5.1: Werte für F_1 und F_2 bei männlichen deutschen und italienischen Sprechern

(Messwerte nach (Simpson, 1988) und (Cosi et al., 1995))

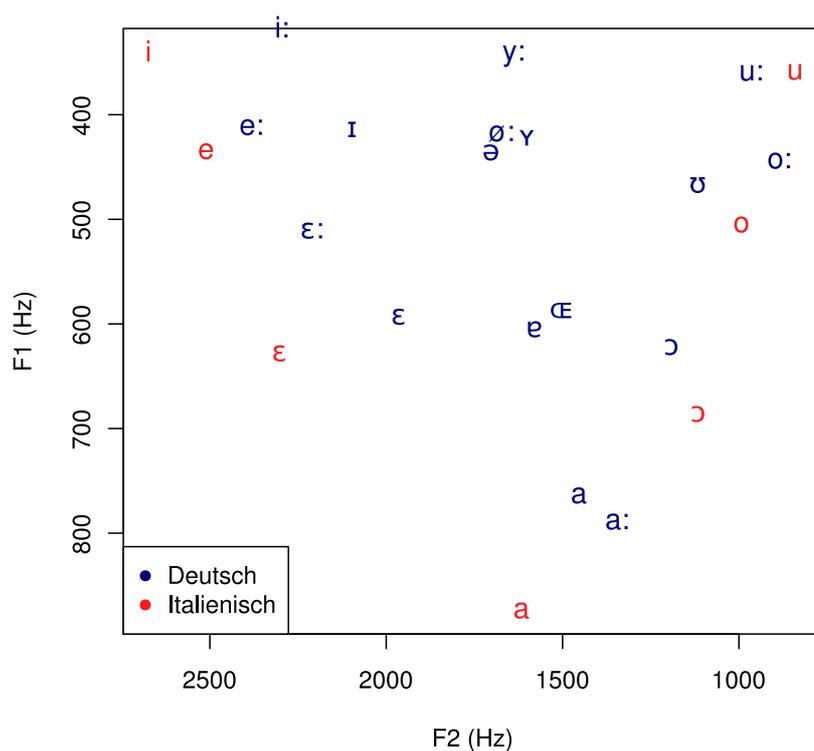
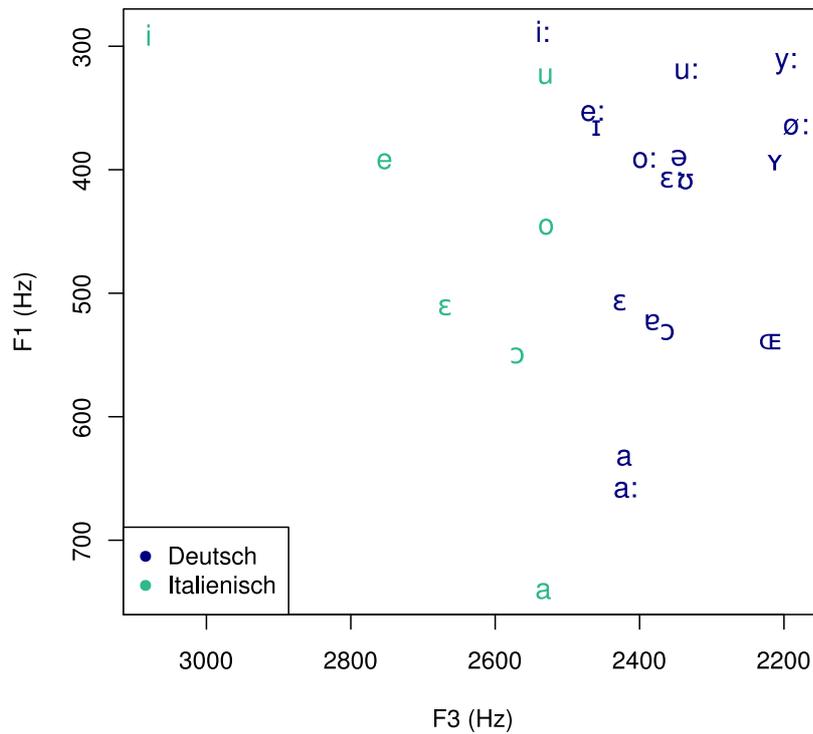
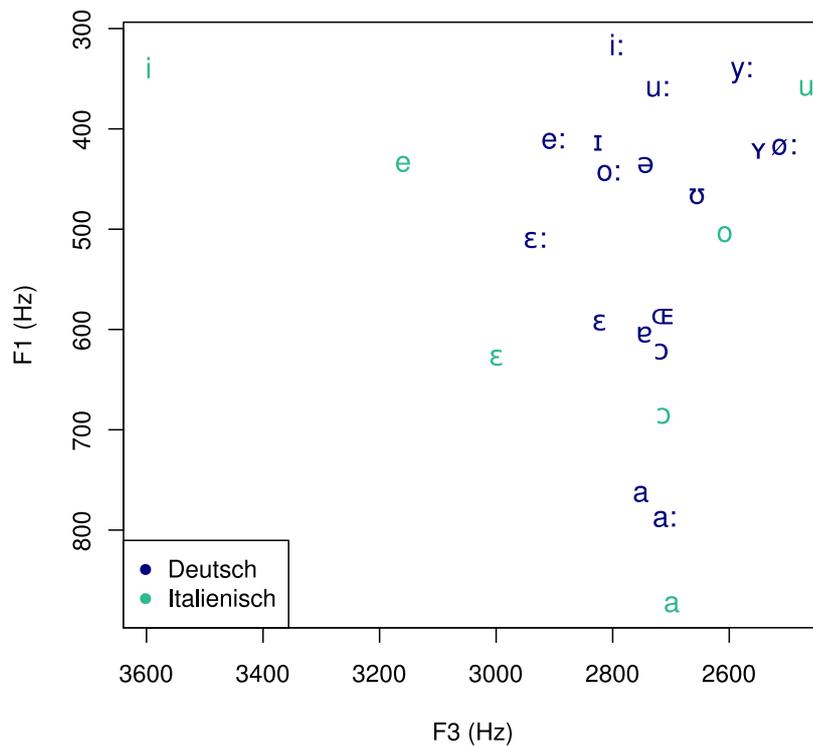
Abbildung 5.2: Werte für F_1 und F_2 bei weiblichen deutschen und italienischen Sprechern

Abbildung 5.3: Werte für F_1 und F_3 bei männlichen deutschen und italienischen Sprechern



(Messwerte nach (Simpson, 1988) und (Cosi et al., 1995))

Abbildung 5.4: Werte für F_1 und F_3 bei weiblichen deutschen und italienischen Sprechern



(Savy & Cutugno, 1997) untersuchen zwei Arten von Vokalreduktion, die Änderung der Vokalqualität durch Unterartikulation und die Vokalreduktion bei unbetonten im Vergleich zu betonten Vokalen.

Dazu verwenden sie zwei Corpora:

PC Die mit PC (*parlato connesso*) beschriebenen Aufnahmen stammen aus der Database on Italian Vowel Acoustics (DIVA), die Ausschnitte aus Nachrichtensendungen von 20 Sprecherinnen enthält, von denen jeweils fünf aus den Regionen Lombardei, Toskana, Latium und Kampanien stammen. Aus diesen Aufnahmen werden pro Person 160 Vokalabschnitte ausgewählt, davon 70 betonte, 50 unbetonte und 40 wortfinale.

Psp Die unter Psp (*parlato spontaneo*) gruppierten Aufnahmen bestehen aus frei geführten Gesprächen mit zwei Sprecherinnen aus der Region Kampanien. Aus diesen Aufnahmen werden wieder 160 Vokalabschnitte pro Person ausgewählt, nach denselben Kriterien wie bei den Nachrichtensendungen.

Da im Italienischen kein Schwa-Laut existiert, wird hier der *centroid*, als Referenzwert verwendet, d.h. der Mittelpunkt aller F_1 - und F_2 -Werte. Allerdings wird jeweils ein Mittelwert für die betonten und für die unbetonten Vokale berechnet. Der Abstand dieser Werte dient als Gradmesser für die Zentralisierung der unbetonten Vokale. Die im Zuge der Untersuchung errechneten Mittelwerte stimmen jedoch mit den Schwa-Werten in anderen Sprachen überein.

Die Aufnahmen aus den Nachrichtensendungen weisen in allen untersuchten Regionen bei unbetonten Vokalen deutlich zentralere F_1 - und F_2 -Werte auf als bei betonten Vokalen. In den aufgenommenen Gesprächen fällt die Vokalreduktion für /i/, /o/ und /u/ ähnlich aus. /e/ und /a/ werden in unbetonten und wortfinalen Vokalen stärker zentralisiert und der Abstand zwischen finalen und unbetonten Vokalen für /e/ fällt sehr gering aus.

Für /ε/ und /ɔ/ werden in allen Aufnahmen nur betonte Vokalabschnitte untersucht.

Insgesamt lassen sich die Vokalabschnitte aus den Nachrichtensendungen und den frei geführten Gesprächen in folgende Gruppen einteilen:

betonte Vokale Während sich die Ellipsen für die F_1 - und F_2 -Werte aus den Aufnahmen der Nachrichtensprecher nur peripher überlagern, überschneiden sich die Ellipsen für /i/ und /e/ in den Aufnahmen aus freien Gesprächen deutlich. /e/ wird dort zentraler artikuliert, als in den Nachrichtensendungen.

unbetonte Vokale Hier fällt auf, dass die Messwerte für F_2 bei den Vokalen /i/, /e/, und /a/ in den freien Gesprächen niedriger ausfallen, als in den Nachrichtensendungen. Für /a/, /e/, und /o/ und ergeben sich zusätzlich niedrigere F_1 -Werte. Alle Vokale außer /a/ besitzen in den frei aufgenommenen Gesprächen F_2 -Werte, die stärker streuen, als die F_1 -Werte.

wortfinale Vokale Hier nehmen die Formantwerte der Vokale /a/, /e/, und /o/ für die freien Gespräche zentralere Werte an, die sich stark überlappen. Nur die Werte für /i/ bleiben deutlich unterscheidbar.

Die von mir untersuchten Sprachaufnahmen enthalten bewusst gelesene Texte, d.h. die Vokale entsprechen eher der Aussprache in einer Nachrichtensendung. Da ich Abschnitte mit mehreren aufeinander folgenden Vokalen ausgeschlossen habe, fällt die Anzahl der wortfinalen Vokale niedriger aus. Dagegen macht sich der Unterschied zwischen betonten und unbetonten Silben deutlich bemerkbar.

Vergleiche ich diese Ergebnisse mit den Messungen in gelesenen Texten und spontanen Gesprächen im Kiel Corpus (s.(Simpson, 1988, S. 214-217)), so liegt der häufigste F_1 -Wert für /i:/ in freien Gesprächen niedriger, ist jedoch trotzdem noch weit von den zentraleren Formantwerten für /e:/ entfernt. Leider wird bei den Referenzwerten nicht zwischen betonten und unbetonten Silben unterschieden, sondern nur ein Wertebereich von der ersten Quartile (bei 25% der gemessenen sortierten Werte) über den Median (häufigster Wert in der Mitte) zur dritten Quartile (bei 75% der gemessenen sortierten Werte) definiert. Dabei überlappen sich die F_1 -Werte für /i:/ und /e:/ nicht, die für F_2 nur teilweise.

5.1.4 Vokaldauer

5.1.4.1 Lange und kurze Vokale und Konsonanten

Die Vokaldauer spielt im Italienischen keine bedeutungsunterscheidende Rolle, trägt jedoch zum Wort- und Satzakkzent bei. Das Verhältnis von Vokallänge und Silbenbetonung ist allerdings komplexer als im Deutschen oder Englischen und kann durch verschiedene phonologische und morphologische Faktoren beeinflusst werden, so dass Italienisch in der Regel wie andere romanische Sprachen als silbenzählende Sprache eingeordnet wird.

Tabelle 5.3: Bedingungen für die Vokallänge im Italienischen

Betonung	Silbentyp	
	CV	VC
	(nicht am Wortende)	
betont	gelängt	kurz
unbetont	kurz	kurz

Beispiele wären *capa* [ˈka:pa] (Kopf) und *cappa* [ˈkappa] (Umhang, Mantel), aber *sera-ta* [seˈra:ta] (Abend) und *serrata* [serˈrata] (Aussperrung) (s. (*Pons Kompaktwörterbuch Italienisch*, 2009)). Dabei nimmt die Dauer des Rauschens bei Frikativen, der charakteristischen Formantstruktur bei Nasalen oder der Lösung des Verschlusses bei Plosiven gegenüber „einfachen“ Konsonanten deutlich zu.

Die Längung der Konsonanten ist

möglich wenn es sich um die Laute /p,b,t,d,k,g,tʃ,dʒ,f,v,s,m,n,l,r/ handelt.

immer üblich, wenn /ts, dz, ʃ, ɲ, λ/ am Ende einer Silbe stehen.

nie vorhanden, wenn /z,w,j/ am Silbenende auftreten.

(s (Zamboni, 1992, S133)).

(Youngblood, 2007, S. 7-12) untersucht die Rolle der Vokaldauer im Standarditalienischen und zwei norditalienischen Dialekten. Dabei treten lange Vokale im Standarditalienischen nur in betonten Silben auf, die offen sind und nicht am Wortende stehen, während Vokale in unbetonten Silben kurz ausgesprochen werden. Minimalpaare basieren eher auf gelängten Konsonanten als auf langen Vokalen, wie in *fato* [fa:to] (Schicksal) und *fatto* [fatto] (Tatsache).

Um das Verhältnis zwischen Vokallänge und Silbenbetonung zu beschreiben, werden italienischer Silben in Form von Moren interpretiert. Betonte lange Vokale bestehen aus zwei Moren, kurze nur eine, gelängte Konsonanten besitzen ihre eigene Mora im Hintergrund, tragen jedoch zum Abschluss der vorherigen Silbe bei. Kurzen Vokalen in betonten Silben am Wortende wird ebenfalls eine zusätzliche Mora im Hintergrund zugeordnet, die jedoch nur bei *raddoppiamento sintattico* über Wortgrenzen hinweg in Erscheinung tritt.

raddoppiamento sintattico ist die Längung eines Konsonanten innerhalb eines Satzes, also nicht als lexikalische Eigenschaft des Wortes, die unter folgenden Umständen stattfindet:

- Der erste Konsonant eines Wortes, das auf einen betonten finalen Vokal folgt, wird verdoppelt, z.B. in *città triste* (traurige Stadt):

[tʃit'ta] + ['triste] → [tʃit'tat'triste]

- Auf ein Wort, das mit einer geschlossenen betonten Silbe endet, folgt ein Wort, das mit einem Vokal beginnt. In diesem Fall wird der Konsonant am Wortende verdoppelt, wie in *tram elettrico* (elektrische Straßenbahn)

['tram] + [e'lettriko] → ['tramme'lettriko]

- Zuletzt gibt es noch die morphologische Verdoppelung eines Konsonanten am Wortanfang nach bestimmten Schlüsselwörtern, z.B. "o" (oder) in *Maria o Carla* (Maria oder Carla)

[ma'ria] + [o] + ['karla] → [ma'ria ok'karla]

(s. (Youngblood, 2007, S. 6f))

(Celata & Kaeppli, 2004) untersucht die Vokallänge vor einfachen Konsonanten, Doppelkonsonanten und heterogenen Konsonantclustern. Sie verwendet dabei Aufnahmen von männlichen und weiblichen Studenten aus Brescia, der Toskana und aus Palermo und erfasst jeweils die Dauer des Konsonanten und des vorhergehenden Vokals.

Dabei ergibt sich eine deutliche Verkürzung des vorherigen betonten Vokals vor Obstruenten und eine schwächere vor Frikativen. Vor /n:/ und /l:/ werden die Vokale jedoch nicht einheitlich ausgesprochen: bei Sprechern aus Palermo findet keine Kürzung der Vokallänge statt, bei Teilnehmern aus der Toskana ist sie tendenziell signifikant und bei Studenten aus Brescia eindeutig signifikant.

Bei einem direkten Vergleich von Worten, die im Deutschen mit einem Doppelkonsonanten geschrieben werden, davon ein Drittel mit *nn* und *ll*, weisen die italienischen Sprecher deutlich längere Vokaldauern auf, als die deutschen. Da ich jedoch keine Informationen über die Heimatregion der italienischen Sprecher besitze, kann ich nicht eindeutig feststellen, ob die größere Vokallänge dialektal oder durch ein langsames Sprechtempo bedingt ist.

5.1.4.2 Unterschiede zwischen betonten und unbetonten Silben

(Santen & Imperio, 1999) untersucht die Auswirkungen der Position einer betonten Silbe innerhalb eines Wortes auf die Vokaldauer im Italienischen. Dabei stellt er fest, dass nicht alle Vokale innerhalb eines Wortes länger sind als diejenigen am Wortende, sondern nur die Vokale in der vorletzten Silbe. Die gemessene Vokallänge am Wortende hängt stark von den Messkriterien ab. Da auf betonte Vokale am Wortende ein glottaler Verschluss folgt, ergibt sich ein charakteristischer knarzender oder behauchter Bereich nach dem

eigentlichen Vokal. Dieser Verschlusslaut könnte dazu dienen, das Gewicht der Silbe zu erhöhen. Wird dieser stark glottalisierte Bereich mit einbezogen, so sind die Vokaldauer der letzten und der vorletzten Silbe beinahe gleich groß.

In dreisilbigen Worten ist die vorletzte Silbe nicht nur länger als die letzte, sondern auch als die erste Silbe, was betonte italienische Silben von betonten Silben in anderen Sprachen unterscheidet, unter anderem Deutsch und Mandarin.

(D’Imperio & Rosenthal, 1999) führen im erster Teil Versuche mit Sprechern aus Neapel, einem Sprecher aus Rom, einem Sprecher aus Sardinien, sowie zusätzlich mit einem kalabrischen und einem sizilianischen Sprecher durch.

Dabei wird Vokaldauer bei offenen Silben in 2-/3- und 4-silbigen Worten verglichen, also z.B. *'dite* (sagt), *'ditele* (sagt sie), *'ditemelo* (sagt es mir), sowie bei geschlossenen Silben in 2-silbigen Worten *ditte* (jeweils in den selben Satz eingebettet).

Hierbei ergeben sich signifikante Unterschiede für die Vokaldauer zwischen 2- und 3-silbigen, bzw. 2- und 4-silbigen Worten, jedoch nicht zwischen 3- und 4-silbigen Worten. Im zweiten Teil wird mit Hilfe der *optimality theory* nachgewiesen, dass diese Gesetzmäßigkeiten für prosodische Worte sind, d.h. für Worte und ihre Suffixe bzw. für durch *raddoppiamento sintattico* miteinander verbundene Silben gültig.

Um einen adäquaten Vergleich zwischen der Vokaldauer italienischer bzw. deutscher Vokale durchzuführen, müsste man die Testmengen jeweils nach Silbentyp (offen oder geschlossen), sowie nach der Silbenposition innerhalb des Wortes bzw. der Phrase zusammenstellen.

(Vayra, 1999) analysiert die Vokaldauer und die Werte für F_1 in betonten offenen und geschlossenen Silben bei drei Versuchspersonen. Dabei verwendet er Pseudoworte wie [ˈtata], [ˈtapa] oder [ˈtama] in offenen Silben, [ˈtatta], [ˈtappa], [ˈtamma] in durch Doppelkonsonanten abgeschlossenen Silben und [ˈtapta], [ˈtapla] oder [ˈtampa] in geschlossenen Silben mit Konsonantclustern.

Dabei wirkt sich der Gegensatz zwischen offenen und geschlossenen Silben am stärksten aus ($p = 0.0001$). Die Vokaldauer vor den stimmlosen Konsonanten /p/ und /t/ ist signifikant kürzer als vor /m/ ($p = 0.019$). Vor den Doppelkonsonanten /mm/ und /pp/ fällt die Vokaldauer etwas länger aus als vor den Konsonantenclustern /mp/ und /pt/, jedoch erreicht der Unterschied nicht die Signifikanzgrenze.

Die Werte für F_1 unterscheiden sich nicht für offene und geschlossene Silben.

Betrachtet man zum Vergleich die Vokaldauer in unbetonten und betonten Silben ergibt sich ein Unterschied von 1:2,5 (unbetont: 70,75 ms; betont: 178,53 ms).

Für F_1 ergibt sich ein signifikanter Unterschied zwischen betonten und unbetonten Silben, wobei in betonten Silben bis zu 1000 Hz, in unbetonten Silben nur bis zu 700 Hz erreicht werden, was bedeutet, dass in betonten Silben der Kiefer wesentlich stärker gesenkt wird.

(Landi & Savy, 1996) vergleicht Sprachaufnahmen von fünf männlichen Nachrichtensprechern aus der Region Latium, die 800 Silben umfassen, davon 350 betonte, 250 unbetonte und 200 am Wortende. Für jeden untersuchten Abschnitt werden die Dauer des

gesamten Wortes, der Silbe und des Vokals innerhalb der Silbe erfasst, sowie die Silbenanzahl und die Position der Silbe.

Werden die Aufnahmen im Hinblick auf die Sprechgeschwindigkeit, die in Silben pro Sekunde gemessen wird, in homogene Klassen eingeteilt, so ist der Unterschied der Vokaldauer zwischen offenen und geschlossenen betonten Silben nicht signifikant. Die Vokaldauer in unbetonten Silben nimmt für offene und geschlossene Silben beinahe die gleichen Werte an.

Zusammenfassend läßt sich sagen, dass die Vokaldauer im Italienischen keine phonemische Rolle spielt, wie im Deutschen, wo sie z.B. das einzige Unterscheidungsmerkmal zwischen /a:/ und /a/ in fließend gesprochenen Äußerungen darstellt, s. S.54. Dafür spielt sie eine wichtigere Rolle bei der Abgrenzung von betonten und unbetonten Silben, die im Deutschen teilweise mit einer anderen Vokalqualität realisiert werden, z.B. als /ə/.

5.2 Gespannte und ungespannte Vokale

Untersuchungsgegenstand 4 (Umsetzung von gespannten und ungespannten Vokalen)

Im Gegensatz zu den mittleren Vokalen /e/ und /ɛ/, die in betonten offenen Silben gelängt werden, existieren im Italienischen bei den hohen Vokalen nur /i/ und /u/, jedoch nicht den Kontrast zu /ɪ/ und /ʊ/ als ungespannte Phoneme. Deshalb prüfe ich in diesem Kapitel, ob italienische Muttersprachler den Gegensatz zwischen /i:/ und /ɪ/ und im Deutschen auf dieselbe Art und Weise umsetzen wie deutsche Muttersprachler.

Zuerst wollte ich hier auch noch den Gegensatz zwischen /u:/ und /ʊ/ miteinbeziehen, kann dort jedoch bei den deutschen Muttersprachlern nur für F_1 signifikante Unterschiede feststellen, die auch von den italienischen Muttersprachlern korrekt umgesetzt werden.

5.2.1 Unterscheidung zwischen /i:/ und /ɪ/

5.2.1.1 Gemischte Modelle für /i:/ und /ɪ/ bei männlichen Sprechern

Um festzustellen, wie sich /i:/ und /ɪ/ bei deutschen und italienischen Sprechern unterscheiden, erstelle ich gemischte Modelle mit dem einzelnen Sprecher als Zufallsfaktor und dem vorgegebenen Vokal als festem Faktor.

Für die vier deutschen männlichen Sprecher vergleiche ich dabei je 45 Messwerte für /i:/ und /ɪ/, wobei ich für /i:/ 45 von 70 Werten zufällig ausgewählt habe.

Tabelle 5.4: Gemischtes Modell für /i:/ und /ɪ/ bei deutschen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	7.127	0.0001	0
F_2	-7.15	0.0001	0
F_3	-5.24	0.0001	0
Vokaldauer	-3.013	0.0048	0.0034

Die Anpassung des Modells an die gemessenen Daten bleibt hier mit BIC-Werten zwischen 900 und 1200 Punkten für die gemischten Modelle der Formantwerte sehr ungenau, der einzelne Sprecher als Zufallsfaktor wirkt sich am stärksten auf F_2 mit 43,7% aus, am schwächsten auf die Vokaldauer mit 24,6%.

In diesem Fall ist der Unterschied für alle Messwerte signifikant. Leider ist die Anzahl der betonten und unbetonten Abschnitte schon in den Ausgangsdaten ungleich, so dass ich keine Kovarianzanalyse mit der Silbenbetonung als zusätzlichem festem Faktor durchführen.

Gruppieren Sie die Messwerte nach Vokal und Silbenbetonung, so weisen diese für die ersten drei Formanten ähnliche Varianzen auf. Deshalb verwende ich den DTK-Test, um festzustellen, wie weit die Werte der betonten und unbetonten Vokale jeweils voneinander

entfernt sind. Für alle Formanten war die Distanz zwischen betontem /i:/ und betontem /ɪ/ am größten, mit 75 Hz für F_1 , 362 Hz für F_2 und 395 Hz für F_3 . Der Abstand zwischen den unbetonten Vokalen fiel dagegen am geringsten aus, mit 28 Hz für F_1 , 84 Hz für F_2 und 71 Hz für F_3 . Damit zeigen sich für die betonten Vokale stärkere Unterschiede als für unbetonte, diese bleiben jedoch in beiden Fällen wahrnehmbar.

Für die drei italienischen männlichen Sprecher vergleiche ich je 57 Messwerte für /i:/ und /ɪ/, wobei ich für /i:/ 57 von 84 Werten zufällig ausgewählt habe.

Tabelle 5.5: Gemischtes Modell für /i:/ und /ɪ/ bei italienischen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	-0.651	0.5222	0.5161
F_2	0.80	0.4384	0.4231
F_3	-0.44	0.6674	0.6597
Vokaldauer	1.559	0.1154	0.1217

Die Anpassung des gemischten Modells an die gemessenen Daten ist mit BIC-Werten zwischen 1200 und 1642 Punkten für die Formantwerte noch ungenauer, als bei den deutschen männlichen Sprechern. Der einzelne Sprecher als Zufallsfaktor wirkt sich vor allem auf F_1 mit 47,6% aus, am geringsten auf die Vokaldauer mit 15,4%. In diesem Fall zeigen sich keine signifikanten Unterschiede zwischen den Messwerten für /i:/ und /ɪ/.

Tabelle 5.6: Durchschnittliche Messwerte für /i:/ und /ɪ/ bei deutschen und italienischen männlichen Sprechern

Parameter	Mittelwerte (deutsch)		Mittelwerte (italienisch)	
	/i:/	/ɪ/	/i:/	/ɪ/
F_1 (Hz)	310	368	316	322
F_2 (Hz)	2016	1799	2038	1989
F_3 (Hz)	2685	2440	2658	2678
Vokaldauer (ms)	65,93	48,17	73,38	64,23

In diesem Fall liegen die Formantwerte für /i:/ und /ɪ/ bei den italienischen Sprechern sehr viel näher beieinander.

Um keine möglichen Unterschiede zwischen betonten und unbetonten Vokalen zu übersehen, gruppiere ich die Testmengen wieder nach Vokal und Betonung und führe auf den Daten für die ersten drei Formanten, die ähnliche Varianzen besitzen, jeweils den DTK-Test durch.

In diesem Fall zeigt sich, dass die Abstände zwischen betonten und unbetonten Abschnitten eines Vokals teilweise größer ausfallen, als wenn jeweils betonte oder unbetonte Ab-

schnitte zweier Vokale verglichen werden. So beträgt der Abstand zwischen unbetontem /ɪ/ und betontem /ɪ/ für F_1 22 Hz, für F_2 223 Hz und für F_3 150 Hz. Der Abstand zwischen betontem /i:/ und betontem /ɪ/ beträgt für F_1 21 Hz, für F_2 113 Hz und für F_3 6 Hz.

Dies zeigt, dass die italienischen Sprecher wie in ihrer Muttersprache betonte und unbetonte Vokale unterschiedlich artikulieren, hier jedoch nicht zwischen /i:/ und /ɪ/ unterscheiden.

5.2.1.2 Gemischte Modelle für /i:/ und /ɪ/ bei weiblichen Sprechern

Für die drei weiblichen deutschen Sprecher vergleiche ich jeweils 32 Abschnitte für /i:/ und /ɪ/, wobei ich für /i:/ 32 von 59 Abschnitten auswähle.

Tabelle 5.7: Gemischtes Modell für /i:/ und /ɪ/ bei deutschen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	3.79	0.0001	0.0003
F_2	-3.722	0.0004	0.0004
F_3	-2.93	0.0034	0.0048
Vokaldauer	-0.549	0.5724	0.5852

Die Anpassung der gemischten Modelle an die Formantwerte ist mit BIC-Werten zwischen 689 und 910 Punkten wesentlich besser als bei den männlichen deutschen Sprechern. Dies könnte durch das größere Gewicht des Zufallsfaktors bedingt sein, der für F_1 49% beträgt.

Bilde ich Untergruppen für betonte und unbetonte Vokale, so besitzen diese für F_1 , F_2 und die Vokaldauer ähnliche Varianzen. Die Ergebnisse des DTK-Tests fallen für diese Werte ähnlich aus wie bei den deutschen männlichen Sprechern, da sich zumindest für die Formantwerte die betonten Vokale am stärksten und die unbetonten Vokale am wenigsten unterscheiden. So betragen die Differenzen für betontes /i:/ und /ɪ/ bei F_1 113 Hz, bei F_3 596 Hz, für unbetontes /i:/ und /ɪ/ jedoch nur 20 Hz bei F_1 und 85 Hz bei F_3 . Die Vokaldauer hängt für /i:/ wesentlich mehr von der Silbenbetonung ab, hier liegt zwischen unbetonten und betonten Abschnitten ein Abstand von 44,87 ms. Der Unterschied zwischen betontem /i:/ und betontem /ɪ/ beträgt 36 ms und zwischen unbetontem /i:/ und /ɪ/ nur 9 ms.

Für die drei italienischen Sprecherinnen vergleiche ich jeweils 61 Werte für /i:/ und /ɪ/, wobei ich für /i:/ 61 von 88 Werten auswähle.

Für die ersten drei Formantwerte liegen die BIC-Werte zwischen 1430 und 1923 Punkten, also wesentlich schlechter als in den deutschen Gruppen oder der italienischen männlichen Gruppe. Der Zufallsfaktor spielt hier eine geringere Rolle als in den anderen Modellen, er deckt für F_1 28,1% der ermittelten Abweichungen ab, für F_3 nur 21,8%.

Tabelle 5.8: Gemischtes Modell für /i:/ und /ɪ/ bei italienischen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	-1.555	0.1270	0.1225
F_2	0.35	0.7340	0.7266
F_3	1.56	0.1246	0.1222
Vokaldauer	0.712	0.4750	0.4777

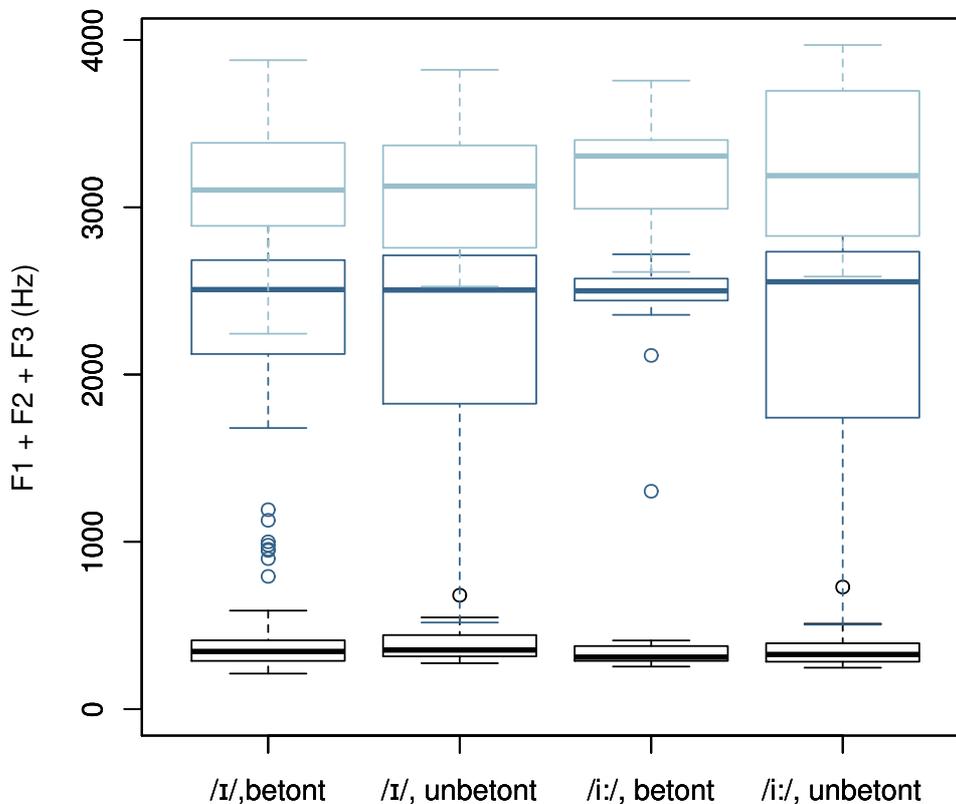
Tabelle 5.9: Durchschnittliche Messwerte für /i:/ und /ɪ/ bei deutschen und italienischen weiblichen Sprechern

Parameter	Mittelwerte (deutsch)		Mittelwerte (italienisch)	
	/i:/	/ɪ/	/i:/	/ɪ/
F_1 (Hz)	363	410	336	360
F_2 (Hz)	2261	2005	2236	2195
F_3 (Hz)	2904	2683	3199	3090
Vokaldauer (ms)	66,38	62,37	79,74	75,12

Wie bei den männlichen Sprechern liegen die Messwerte für /i:/ und /ɪ/ in der italienischen Gruppe wesentlich näher beieinander als in der deutschen.

Wieder prüfe ich die Varianzhomogenität, in diesem Fall ist jedoch die Streuung der Werte für /ɪ/, vor allem bei unbetonten Silben zu groß.

Abbildung 5.5: Werte für /ɪ/ und /i:/ in betonten und unbetonten Silben bei italienischen weiblichen Sprecherinnen



Was deutlicher als bei den männlichen Sprechern auffällt, ist die Nähe zwischen betontem /i:/ und betontem /ɪ/ sowie unbetontem /i:/ und unbetontem /ɪ/. Diese lässt den Schluss zu, dass die Sprecherinnen /i:/ und /ɪ/ als betonte und unbetonte Version von /i/ betrachten, analog zu der unterschiedlichen Aussprache betonter und unbetonter Vokale im Italienischen.

5.2.2 Vergleich von /ɪ/

Da der Vokal /ɪ/ im italienischen Vokalsystem nicht existiert und von italienischen Deutschlernenden auch nicht von /i:/ unterschieden wird, möchte ich die Aussprache dieses Vokals noch einmal im Detail untersuchen, um Gründe für die Abweichungen zu finden.

5.2.2.1 Gemischtes Modell für /ɪ/ bei männlichen Sprechern

In diesem Fall vergleiche ich jeweils 45 Werte von vier deutschen und drei italienischen Sprechern, wobei ich bei den italienischen Sprechern 45 von 57 zufällig auswähle. Die Anzahl der betonten Silben überwiegt deutlich, so sind bei den deutschen Sprechern 39, bei den italienischen 30 betont.

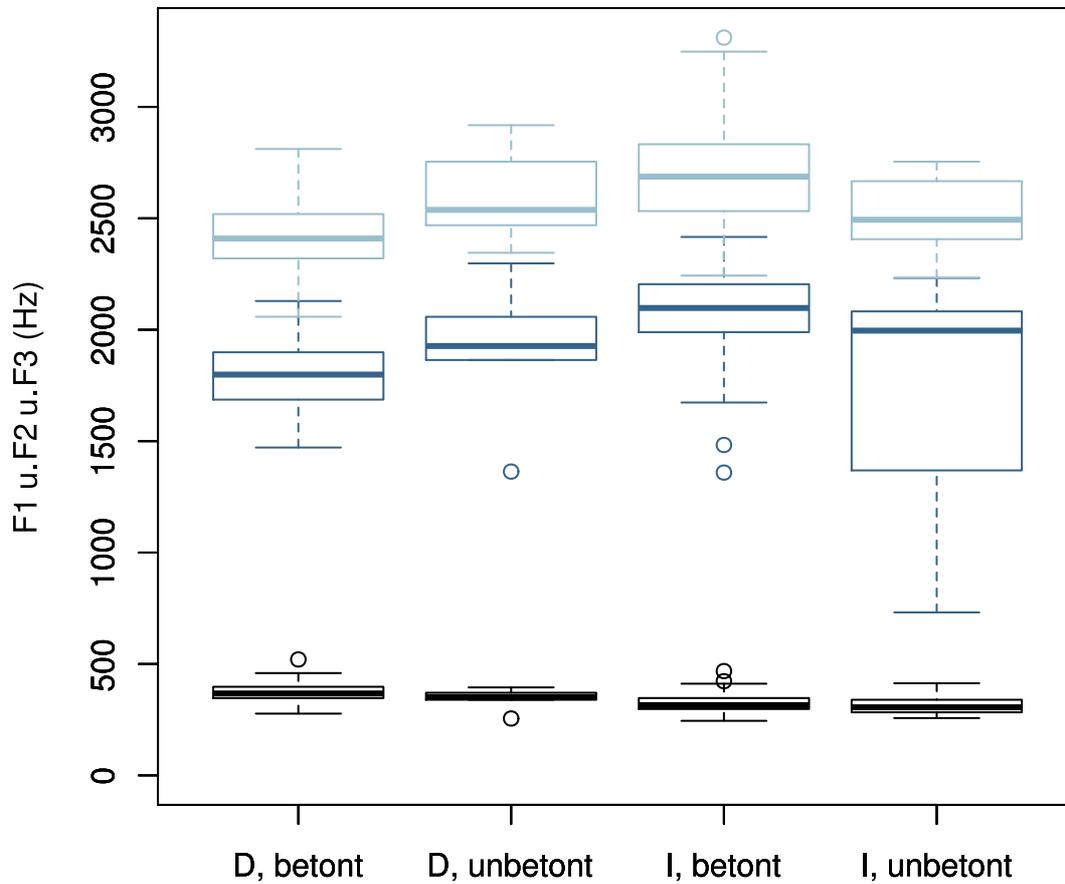
Tabelle 5.10: Gemischtes Modell für /ɪ/ bei deutschen und italienischen männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	-2.236	0.0264	0.0279
F_2	1.601	0.1182	0.1129
F_3	2.18	0.0328	0.0319
Vokaldauer	3.073	0.0156	0.0028

Die BIC-Werte der gemischten Modelle bewegen sich im gleichen Bereich wie die der gemischten Modelle für /i:/ und /ɪ/ bei deutschen männlichen Sprechern. Das Gewicht des einzelnen Sprechers als Zufallsfaktor liegt für F_1 und F_3 bei 39,6% bzw. 39,5%, für die Vokaldauer bei 23,1%.

Da sich die Varianzen zu stark unterscheiden, wenn ich Untergruppen für betonte und unbetonte Vokale erstelle, füge ich hier eine Plotgrafik für die entsprechenden Gruppen ein, die ich auf der Grundlage der tatsächlichen Messwerte erstelle.

Abbildung 5.6: Aussprache von /ɪ/ durch deutsche und italienische männliche Sprecher



In diesem Fall wird deutlich, dass die betonten Abschnitte trotz der größeren Anzahl an Messwerten eine geringere Streuung aufweisen und sich die betonten Abschnitte für deutsche und italienische Sprecher deutlich unterscheiden.

Die Formantwerte für /ɪ/ in betonten Silben bei italienischen männlichen Versuchspersonen liegen deutlich näher an den Werten für /i:/ in betonten Silben in der deutschen männlichen Kontrollgruppe als an den Werten für /ɪ/ in der deutschen Gruppe.

Was die Vokaldauer betrifft, so liegen die Werte im gemischten Modell bei 48,24 ms für deutsche und bei 62,60 ms für italienische Sprecher. Damit befinden sich die italienischen Sprecher wieder deutlich näher an der Vokaldauer für /i:/ bei deutschen Sprechern als an der Vokaldauer für /ɪ/.

5.2.2.2 Gemischtes Modell für /ɪ/ bei weiblichen Sprechern

Hier untersuche ich jeweils 32 Abschnitte von drei deutschen und drei italienischen weiblichen Sprechern. Dabei wurden aus der italienischen Gruppe nach dem Zufallsprinzip 32 von 61 Werten ausgewählt.

Tabelle 5.11: Gemischtes Modell für /ɪ/ bei deutschen und italienischen weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

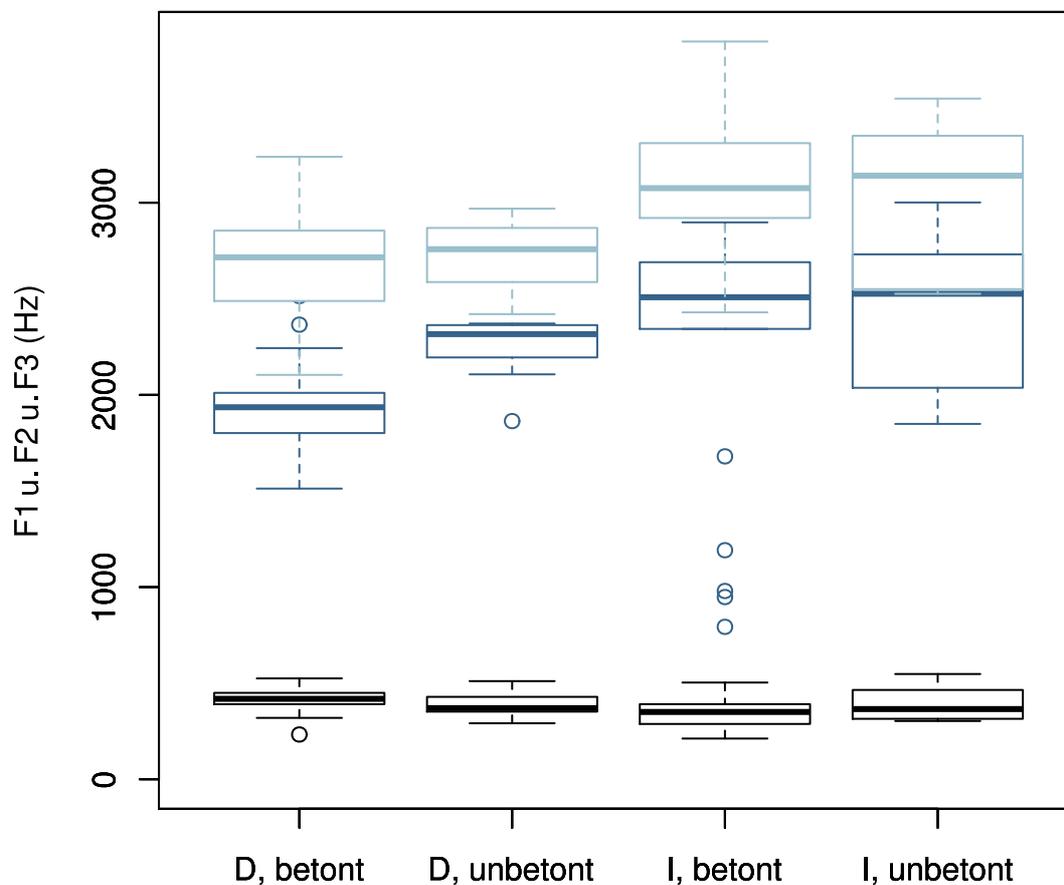
Parameter	t-Wert	pMCMC	p-Wert
F_1	-1.651	0.1046	0.1038
F_2	2.089	0.0822	0.0408
F_3	5.24	0.0118	0
Vokaldauer	0.96	0.3342	0.3407

Vergleicht man das gemischte Modell hier mit dem für die männlichen Sprecher, so fällt auf, dass die Werte für F_1 und die Vokaldauer sich nicht signifikant unterscheiden. Dagegen wird hier für F_2 ein signifikanter Unterschied sichtbar.

Die BIC-Werte liegen hier zwischen 728 und 961 Punkten, also niedriger als in der männlichen Gruppe. Der Anteil der Abweichungen, die durch den Zufallsfaktor erklärt werden, ist hier ebenfalls niedriger und liegt für F_1 bei 30,47%, für F_3 sogar bei 0.

Sowohl bei deutschen als auch bei italienischen Sprecherinnen überwiegt die Anzahl der betonten Vokale mit 25 bei der deutschen und 22 bei der italienischen Gruppe. Allerdings fallen gerade für F_2 die Varianzen nicht einheitlich aus, was auf eine starke Streuung der Werte für /ɪ/ in unbetonten Silben bei den italienischen Sprecherinnen zurückzuführen ist.

Abbildung 5.7: Aussprache von /i/ durch deutsche und italienische weibliche Sprecher



Die Plotgrafik beruht wie bei den männlichen Sprechern auf den tatsächlichen Messwerten, nicht den angepassten Werten des gemischten Modells, um Verzerrungen zu vermeiden.

Wie bei den männlichen Sprechern liegen die Formantwerte für betontes /i/ bei italienischen Sprecherinnen deutlich näher an den Werten für /i:/ bei deutschen Muttersprachlern.

Die im gemischten Modell errechnete Vokaldauer von 72,28 ms für italienische Sprecherinnen liegt deutlich über den 62,37 ms für deutsche Sprecherinnen. Errechnet man die durchschnittlichen Vokaldauern für betonte und unbetonte Silben, so erhält man bei betonten Silben für die italienischen Versuchspersonen 79,55 ms, für deutsche 66,40 ms. Die Vokaldauer in unbetonten Silben liegt bei 59 ms für die italienischen Teilnehmerinnen und bei 48,57 ms für die deutschen.

5.3 Unbekannte Vokale

5.3.1 Der Vokal /ø:/

Untersuchungsgegenstand 5 (Aussprache des Vokals /ø:/) *In diesem Unterkapitel wird untersucht, inwieweit die Aussprache des im Italienischen unbekanntes Vokals /ø:/ bei deutschen und italienischen Sprechern übereinstimmt.*

5.3.2 Statistische Tests

Die Testmengen sind alle normalverteilt, für F_1 und F_2 besitzen sowohl die Werte der männlichen als auch die der weiblichen Sprecher unterschiedliche Varianzen.

Tabelle 5.12: Ergebnisse des Wilcoxon-Rangsummentests für /ø:/ bei deutschen und italienischen Sprechern

Parameter	p-Wert	
	Männer	Frauen
F_1	0.0000937	0.02529
F_2	0.801	0.00024
F_3	0.03383	0.00052
Vokaldauer	0.1236	0.5337

5.3.3 Diskriminanzanalyse

Da sich in gemischten Modellen keine signifikanten Unterschiede zeigen, mir die Abwesenheit jeglicher statistischer Unterschiede für die Messwerte jedoch wenig plausibel erscheint, führe ich zusätzlich eine Diskriminanzanalyse (rda) durch. Dabei zeigen sich bei den männlichen Sprechern eindeutige Unterschiede, während sich die weiblichen Messwerte teilweise überlappen. In beiden Fällen sind die höheren F_3 -Werte bei den italienischen Muttersprachlern deutlich sichtbar, was für eine geringere Lippenrundung sprechen könnte.

Abbildung 5.8: Ergebnisse der Diskriminanzanalyse für /ø:/ bei männlichen deutschen und italienischen Sprechern

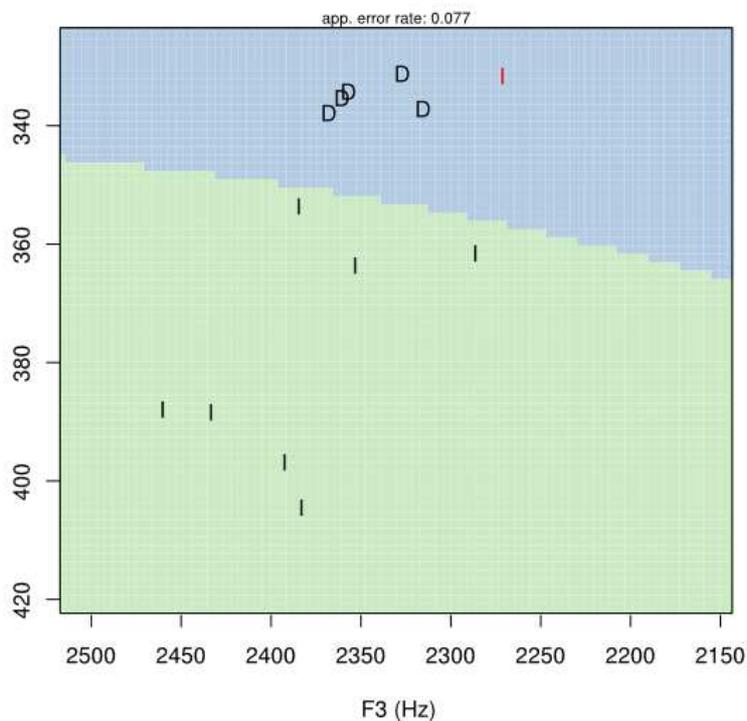
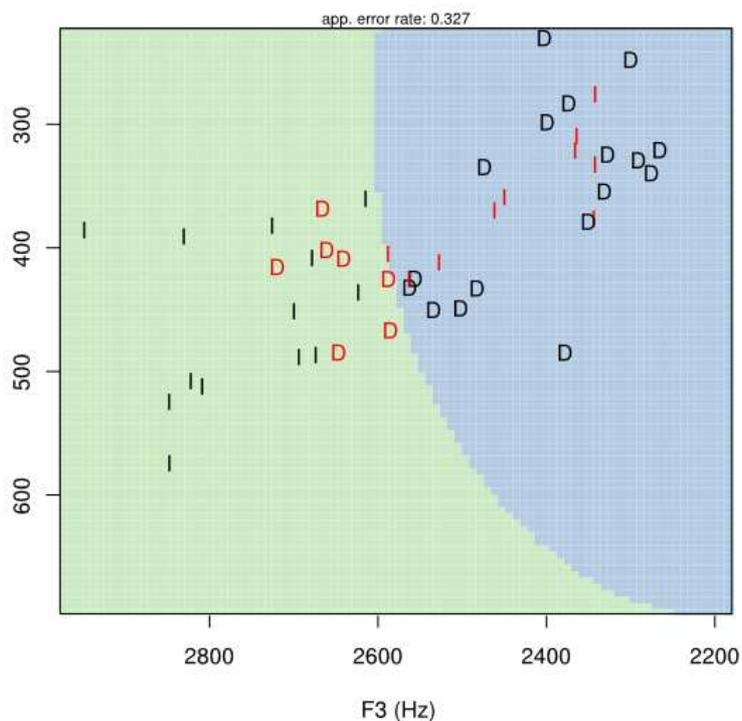


Abbildung 5.9: Ergebnisse der Diskriminanzanalyse für /ø:/ bei weiblichen deutschen und italienischen Sprechern



5.3.4 Vergleich nach Sprachkenntnissen

Um noch genauer herauszufinden, warum sich /ø:/ bei deutschen und italienischen Sprechern unterscheidet, obwohl diese Unterschiede im gemischten Modell verloren gehen, führe mich mittels des von Dunnett für ungleich große Gruppen modifizierten Tukey-Kramer-Tests, den ich auch schon im vorigen Unterkapitel auf /a/ vor Doppelkonsonanten angewandt habe, einen Vergleich zwischen den Gruppen mit unterschiedlichen Sprachkenntnissen durch.

Die Gruppe der italienischen Versuchspersonen besteht dabei aus

- zwei männlichen Studenten, *ar* und *dg*, die jeweils vor und nach dem Besuch eines Sprachkurses aufgenommen wurden,
- einer weiblichen Studentin *di*, die nur vor ihrem Sprachkurs aufgezeichnet wurde,
- zwei weiblichen Studentinnen, *am* und *al*, die sich in Deutschland aufhielten, jedoch keinen Sprachkurs besuchten, wobei *am* zu Beginn und am Ende ihres Aufenthalts an der Studie teilnahm, *al* jedoch nur vor ihrer Rückkehr nach Italien.
- Zusätzlich wurde noch der männliche Teilnehmer *de* mit einbezogen, der nicht studierte und zu dem ich keine weiteren Angaben besitze.

Leider reicht dabei die Anzahl der Abschnitte nicht aus, um gezielt einzelne Sprecher oder die Unterschiede zwischen der ersten und der zweiten Aufnahme für einige Sprecher zu untersuchen.

5.3.4.1 Männliche Sprecher

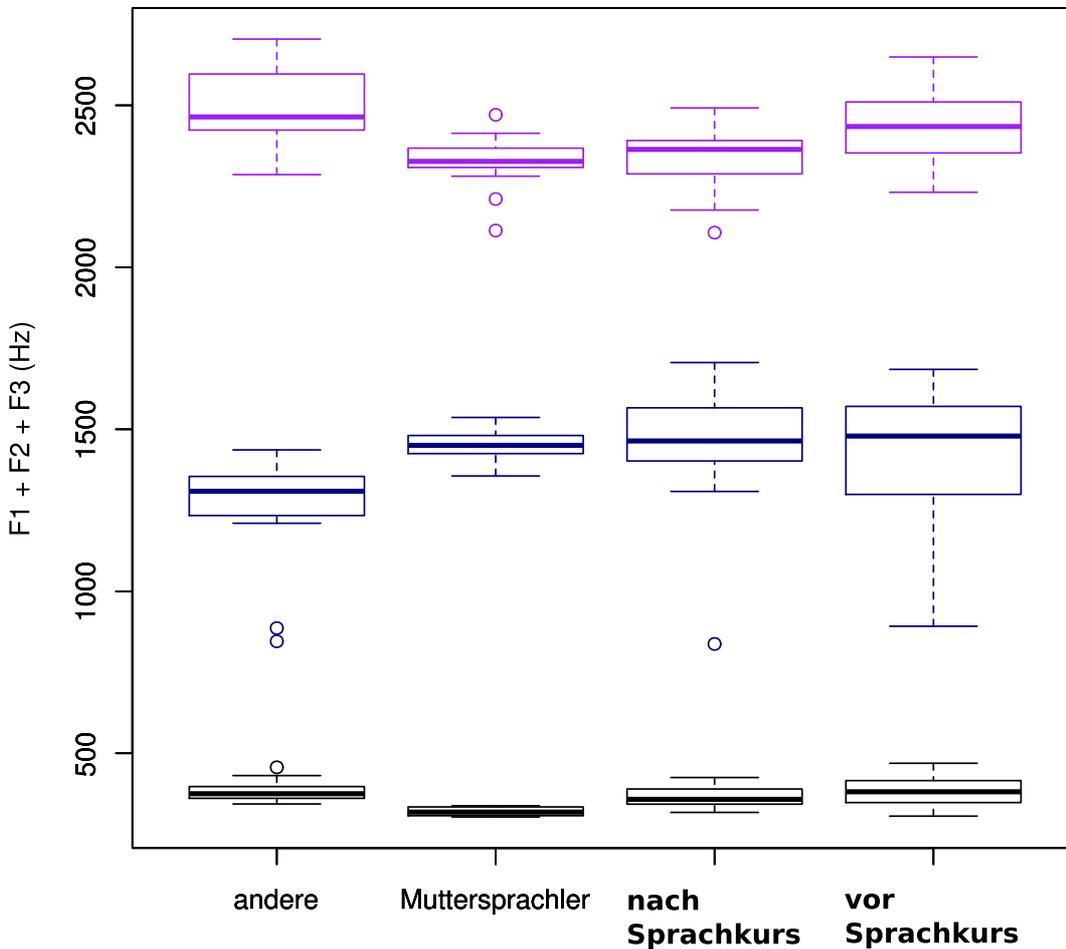
Dabei ist für F_2 und F_3 bei männlichen Sprechern der Abstand zwischen den Sprechern nach dem Sprachkurs und den Muttersprachlern geringer als der Abstand zwischen den Sprechern vor dem Sprachkurs und nach dem Sprachkurs.

Die Streuung der Formantwerte nimmt von den Versuchspersonen vor dem Besuch eines Sprachkurses über diejenigen nach einem Sprachkurs bis zu den Muttersprachlern hin deutlich ab, vor allem für F_1 . Bei F_1 und F_2 fällt auf, dass die Formantwerte von den Muttersprachlern weniger zentral artikuliert werden. Die F_3 -Werte liegen bei den deutschen deutlich niedriger als bei den italienischen Sprechern. Die Vokaldauer ist bei dem italienischen Sprecher am kürzesten, der keinen Sprachkurs besucht, nach dem Besuch eines Sprachkurses länger als vorher und bei den Muttersprachlern am längsten.

Die zentraleren Formantwerte und die kürzere Vokaldauer tragen vor allem bei den italienischen Sprechern vor dem Sprachkurs zu einem Klang bei, der zwischen /ø:/ und /ε/ liegt.

Da es sich bei den männlichen Studenten vor und nach dem Sprachkurs um dieselben Personen handelt, kann man davon ausgehen, dass sich ihre Aussprache in der zweiten Aufnahme nach dem Kurs verbessert hat.

Abbildung 5.10: Vergleich der ersten drei Formanten für /ø:/ nach Sprachkenntnissen bei männlichen Sprechern



5.3.4.2 Weibliche Sprecher

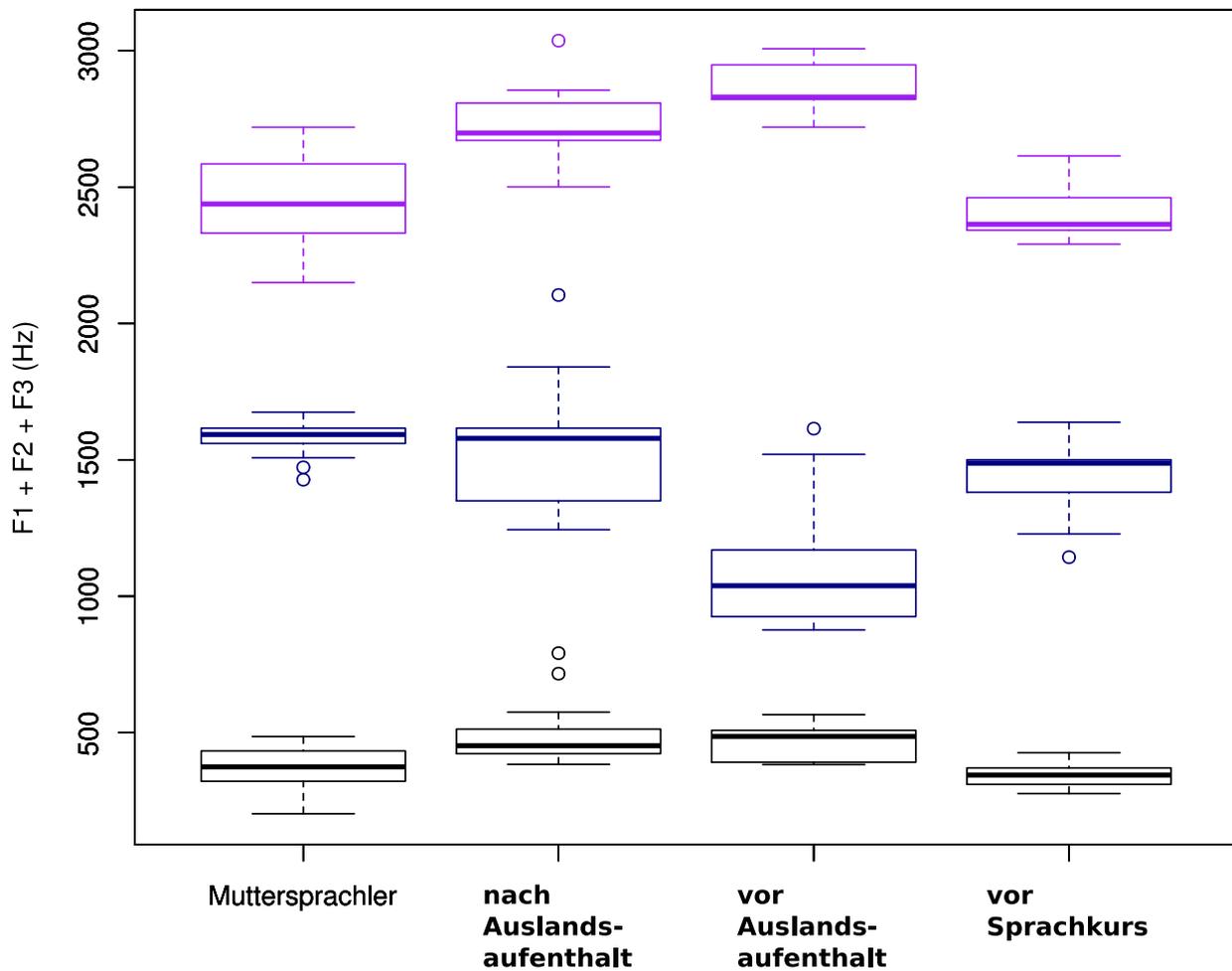
Bei den weiblichen Sprecherinnen ist für F_2 und F_3 ebenfalls der Abstand zwischen den Sprecherinnen nach einem Aufenthalt in Deutschland und den Muttersprachlerinnen geringer als der Abstand zwischen den Sprecherinnen vor und nach einem Aufenthalt in Deutschland.

Die Sprecherinnen, die vor einem Aufenthalt in Deutschland ohne Sprachkurs aufgenommen wurden, artikulieren das /ø:/ so, dass es vom Klang zu einem /ɔ/ tendiert, wozu der niedrige F_2 -Wert entscheidend beiträgt.

Bei den niedrigeren Werten für F_1 , F_2 und F_3 der Sprecherin, die vor dem Besuch eines Sprachkurses aufgenommen wurde, könnte es sich auch um individuelle Eigenheiten handeln, da hier für eine weibliche Sprecherin relativ niedrige- F_0 und F_4 -Werte vorhanden sind. Diese Sprecherin artikuliert das /ø:/ mit einem F_3 -Wert, die nahe an der muttersprachlichen Aussprache liegt, jedoch mit einem niedrigeren Wert für F_1 , so dass die Aussprache nach einem hyperkorrekten /y/ klingt.

Die Vokaldauer ist bei den Sprecherinnen, die sich in Deutschland aufhalten, ohne einen

Abbildung 5.11: Vergleich der ersten drei Formanten für /ø:/ nach Sprachkenntnissen bei weiblichen Sprechern

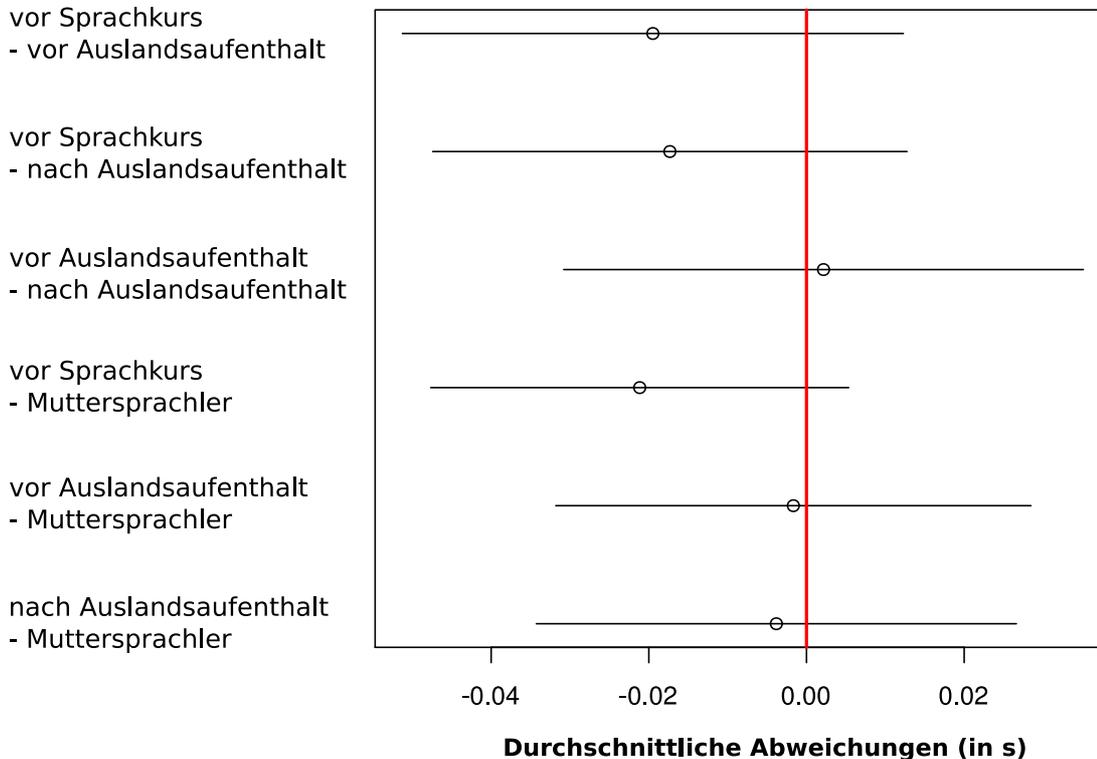


Sprachkurs zu besuchen, länger als bei der Sprecherin, die vor einem Sprachkurs aufgenommen wurde. Zwischen den Aufnahmen vor und nach dem Aufenthalt in Deutschland besteht jedoch kein großer Unterschied und der Abstand zur muttersprachlichen Vokallänge bleibt gleich.

Abbildung 5.12: Vergleich der durchschnittlichen Abweichungen für die Vokaldauer bei /ø:/ innerhalb und zwischen den Sprechergruppen bei weiblichen Sprechern

Konfidenzintervall (95%) für die Unterschiede bei der Vokaldauer

Vergleich der Mittelwerte



5.4 Zusammenfassung

Das Italienische ist, was seinen Vokalraum betrifft, stärker dialektal geprägt als das Deutsche, so dass z.B. der Vokalkontrast zwischen /e:/ und /ɛ/ bei einigen Sprechern vorhanden sein kann, bei anderen nicht.

Untersuche ich den Gegensatz zwischen /i:/ und /ɪ/, so unterscheiden italienische Sprecher deutlich zwischen /ɪ/ in betonten und unbetonten Positionen, wie in ihrer eigenen Sprache üblich, artikulieren das betonte /ɪ/ jedoch wie /i:/.

Außerdem variieren beim Erlernen unbekannter Vokale die Strategien der einzelnen Sprecher stark, wie man hier bei der Aussprache des /ø:/ als /ɛ/ in der männlichen und als /ɔ/ in der weiblichen Gruppe beobachten kann.

Trotzdem lässt sich die Vokalqualität eines neuen Vokals noch leichter erlernen, als die grundsätzliche Unterscheidung zwischen unterschiedlichen Vokallängen, wenn diese wie hier im Italienischen nicht Teil des muttersprachlichen Vokalsystems ist.

Um ausführlichere Untersuchungen zu den unterschiedlichen Vokallängen, die sich auch auf die Umsetzung der Silbenbetonung auswirken, durchführen zu können, wären weitere Vergleiche mit muttersprachlichen italienischen und deutschen Aufnahmen notwendig. Beim Vergleich unbetonter Silben wäre zusätzlich noch die Rolle des Schwa-Vokals im Deutschen relevant, der so im Italienischen nicht existiert, den ich jedoch aufgrund seiner

extremen Streuung in deutschen muttersprachlichen Aufnahmen noch nicht berücksichtigt habe.

6 Deutsche und chinesische Sprecher

6.1 Einführung

6.1.1 Die chinesischen Vokale

In diesem Kapitel möchte ich die Aussprache von Personen vergleichen, die Deutsch oder Mandarin (Hochchinesisch) als Muttersprache sprechen. Dabei führe ich zuerst eine visuelle Gegenüberstellung der Formantwerte durch und vergleiche die Vokaldauern aller Vokale. Danach beschäftige ich mich mit den Auswirkungen des chinesischen Vokals /y/ auf die Aussprache der deutschen Vokale /a:/ und /a/. Als nächstes untersuche ich, inwieweit die chinesischen Sprecher den im Deutschen vorhandenen Kontrast zwischen gespannten und ungespannten Vokalen erlernt haben und als letztes beschäftige ich mich mit deutschen Vokalen, die im Chinesischen unbekannt sind.

Eine der ersten Aufgaben beim Vergleich der von deutschen und chinesischen Sprechern produzierten Vokale ist es, eine passende Basis für einen Vergleich zu finden.

Dabei geht es im weiteren Sinne um die Definition dessen, was mit Chinesisch gemeint ist und im engeren Sinne um die passende Transkription der Laute.

Einerseits umfasst die chinesische Sprachfamilie mindestens sechs Sprachen, darunter Mandarin, auch Putonghua, offizielle Amtssprache in der Volksrepublik China und in Taiwan, zusätzlich Kantonesisch in Südchina und Hongkong, Min auf Taiwan und andere. Im chinesischen Sprachgebrauch werden diese Sprachen als Dialekte bezeichnet, da sie alle dieselben Schriftzeichen verwenden. Aus westlicher phonetischer Sicht weisen sie jedoch starke phonetische Unterschiede auf, die z.B. denen zwischen europäischen Sprachen entsprechen. So besitzt Mandarin vier charakteristische Tonverläufe, Kantonesisch dagegen neun, s. auch (Bussmann, 2002, S. 136).

(Liu, Shuang, Qin, Zhang & Yan, 2007) vergleicht die Formantwerte von Mandarin sprechenden Personen in verschiedenen Regionen Chinas und stellt signifikante Unterschiede für F_2 bei /o/ und /u/ sowie für F_1 und F_3 bei /e/ und F_2 und F_3 bei /i/ fest.

Betrachtet man das Diagramm für F_1 und F_2 , so wird /u/ von nordchinesischen Sprechern in Beijing und Henan mit einer besonders tiefen F_1 artikuliert, /o/ dagegen von Sprechern in Shanghai und Guangdong (Kanton). Die F_1 -Werte für /e/ sind bei Sprechern aus Guangdong (Kanton) am niedrigsten und die höheren F_2 -Werte für /i/ in Beijing heben sich deutlich von den anderen Messungen ab. Ob die sichtbaren Abweichungen für /o/ und /e/ mit dem in Guangdong gesprochenen Kantonesisch begründet werden können, wird leider nicht erläutert.

Da ich jedoch keine defilierten Informationen zur Herkunft und zu den Sprachkenntnissen der chinesischen Teilnehmer abgesehen von Mandarin besitze, kann ich diese Differenzen nicht weiter analysieren.

6.1.1.1 Entstehung der Tonverläufe in betonten Silben

(Hombert, Ohala & Ewan, 1979) erklärt die Entstehung von Tonsprachen einerseits dadurch, dass z.B. der Kontrast zwischen stimmhaften und stimmlosen oder stimmlos aspirierten Konsonanten vor einem Vokal verschwindet und bei stimmhaften Konsonanten durch einen tieferen Ton ersetzt wird, bei stimmlosen Konsonanten durch einen höheren Ton. Dabei verweist er unter anderem auf eigene Messungen, bei denen sich die Grundfrequenz für /i/ nach stimmhaften Plosiven als niedriger erwies, die nach stimmlosen Plosiven als höher. Diesen Unterschied erklärt er durch die vertikale Position des Kehlkopfs, der am Ende des Verschlusses bei einem stimmlosen Plosiv eine höhere Position einnimmt, als bei einem stimmhaften Plosiv.

Andererseits kann für ihn auch ein auf einen Vokale folgender glottaler Verschluss Ausgangspunkt für eine tonale Kontur sein, z.B. im Vietnamesischen und im Mittelchinesischen, wo ein auf einen Vokal folgender glottaler Verschluss durch einen steigenden Ton ersetzt wurde. In diesen Sprachen entwickelte sich außerdem aus einem auf einen Vokal folgenden /h/ ein fallender Ton. Der Autor sieht sich durch Messungen bei arabischen Sprechern bestätigt, bei denen die Grundfrequenz innerhalb eines Vokals vor /ʔ/ anstieg und vor /h/ abfiel.

Damit dienen im letzteren Fall glottale Konsonanten, also ein glottaler Verschluss oder /h/, als Ausgangspunkt für Gegensätze, die in betonten Silben eine Rolle spielen. Im Deutschen und im Englischen unterscheidet man in betonten Silben zwischen gespannten und ungespannten Vokalen, während im Chinesischen verschiedene Tonverläufe verwendet werden. Unbetonte chinesische Silben besitzen dagegen oft einen neutralen Ton, der dem Verlauf betonter Silben angepasst wird.

6.1.1.2 Silbenstruktur

Wie man in (Duanmu, 2009, S. 80) sieht, bilden in China selbst nicht Phoneme die Grundlage der Beschreibung der Aussprache, sondern ganze Silben. Diese besitzen wie in anderen ostasiatischen Sprachen eine klar definierte Struktur:

Initial, Medial und Coda sind optional, es gibt z.B. Silben, die nur aus einem Vokal bestehen, z.B. [ʎe] (Motte).

6.1.1.3 Der chinesische Vokalraum

Die chinesische Pinyin-Umschrift, die auch in China selbst z.B. bei der Bedienung von Computern verwendet wird, enthält vier Diakritika zur Darstellung von Tonhöhen, und einige im Deutschen nicht vorhandene Konsonanten. Bei den Vokalen sieht man jedoch,

Tabelle 6.1: Struktur einer chinesischen Silbe

Initial	Medial	Reim	
		Vokal	Coda
/l/, /m/, /n/; stimmlose aspirierte Plosive, Frikative und Affrikate	die Gleitlaute /w/, /j/, /ɥ/	ein Vokal, /ɤ/ oder ein anderer Diphthong	/j/, /w/, /r/ oder ein Nasal, d.h. /n/ bzw. /ŋ/

dass einige Laute, die in westlichen Sprachen als Phoneme auftreten, dort Allophone eines Vokals sind (s. auch (Cheng, 1973)).

Tabelle 6.2: Zuordnung Pinyin – IPA

Pinyin	Vokal (IPA)
i	/i/, /ɪ/ nach /s,ts,ts ^h /, /ɿ/ nach /ʃ,tʃ,tʃ ^h /
yu/ü	/y/
u	/u/
e	/ɛ/, /ə/, /ɤ/
o	/ɔ/
a	/a/

Auf der Basis ausgewerteter moderner Reimlexika kommt (Duanmu, 2000, S. 37ff) jedoch zu einer kompakteren Darstellung, die die drei hohen Vokale /i/, /y/ und /u/, einen mittleren Vokal /e/ und einen niedrigen Vokal /a/ umfasst.

In diesem Modell umfassen der mittlere Vokal /e/ und der niedrige Vokal /a/ jeweils eine Reihe von Allophonen.

6.1.1.3.1 Allophone des mittleren Vokals /e/

o steht in offenen Silben, nach Labialen, z.B. in ɿpo (Abhang).

E liegt akustisch zwischen /e/ und /ɛ/ und wird in offenen Silben nach Palatalen gesprochen, d.h. nach /j/, /ɥ/, /j/ und /ɥ/.

ɤ wird in offenen Silben gesprochen, wenn kein Labial oder Palatal vorausgeht.

e steht vor i.

ə wird vor [-u,-n,ŋ] ausgesprochen.

(s. (Duanmu, 2000, S.38))

6.1.1.3.2 Allophone des niedrigen Vokals /a/

ʌ steht in offenen Silben.

ɑ wird in geschlossenen Silben vor [u] und [ŋ] gesprochen.

a steht in geschlossenen Silben nach einem Konsonant der kein Palatal ist und vor /n/ oder /i/.

æ steht in geschlossenen Silben zwischen /j/ oder [C^j] und /n/.

e steht in geschlossenen Silben nach /ŋ/ und [C^ŋ] und vor /n/.

(s. (Duanmu, 2000, S.40))

Der Vokal /ɛ/ wird hier als Diphthong betrachtet, genau wie /ai/, /au/, /əi/ und /əu/, die als Kombinationen des niedrigen oder mittleren Vokals mit einem hohen Vokal betrachtet werden.

Die in anderen Arbeiten als Allophone von /i/ behandelten Vokale /ɪ/ und ɹ/ werden hier als syllabische Konsonanten [z] und [r] betrachtet, was ich aufgrund der Formantmessungen in anderen Untersuchungen nicht nachvollziehen kann.

6.1.2 Betonte Silben

(Huang, 1992, S. 371) beschreibt den neutralen Ton einer unbetonten Silbe, der sich im Intonationsverlauf an die vorherige Silbe anbläst. Sie betrachtet den neutralen Ton einer

Abbildung 6.1: Verhältnis von lexikalischen zu neutralen Tönen im Chinesischen

Ton	Verlauf	nachfolgender neutraler Ton
1. Ton	⌈	∨
2. Ton	↗	∨
3. Ton	↘	↘
4. Ton	∨	∨

unbetonten Silbe als „fünften“ Ton, da sich in Komposita die Bedeutung zwischen zwei betonten oder einer betonten und einer unbetonten Silbe unterscheiden kann. Sie erwähnt dabei Beispiele wie [ʈtɛ^hi ʋtsi] (Ehefrau) und [ʈtɛ^hi ʌtsi] (Frau und Kind) oder [ʈɛyoŋ ʋti] (jüngerer Bruder) und [ʈɛyoŋ ʋti] (Brüder).

(Surendran & Levow, 2004) vergleicht den Informationsgehalt der lexikalischen Töne im Chinesischen und des Wort- bzw. Satzakkzents im Deutschen, Englischen und Niederländischen mit dem Informationsgehalt einzelner Phoneme. Zuerst wird dabei die Entropie $H(L)$ für L berechnet. Dabei steht L entweder für ein Wort, das als eine Folge von Silben verarbeitet wird oder für Silbe, die aus einer Folge von Phonemen besteht.

$$H(L) = - \sum_x p(x) \log_2 p(x)$$

Mit dieser Formel wird die Entropie der im Korpus enthaltenen Worte bzw. Silben $H(L_u)$ ermittelt. Als nächstes wird die Entropie von Silben und Worten berechnet, bei denen z.B. einzelne Vokale nicht mehr unterschieden, sondern als eine Gruppe, $H(L_u^{-vowel})$, behandelt werden. Der Informationsgehalt der Vokale ergibt sich dann aus folgender Formel:

$$FL_u(vowel) = \frac{H(L_u) - H(L_u^{-vowel})}{H(L_u)}$$

Bei den Vokalen ist der Informationsgehalt innerhalb eines Wortes im Deutschen am größten (0,042) und im Chinesischen am geringsten (0,022). Der Unterschied zwischen Konsonanten und Vokalen hat in niederländischen Worten und Silben den größten Informationsgehalt (0,634 bzw. 0,512) und in chinesischen Worten und Silben den geringsten (0,453 bzw. 0,231). Das Gewicht des lexikalischen Tons in chinesischen Silben (0,108) ist wesentlich stärker als das Gewicht der Silbenbetonung im Deutschen (0,034) oder im Niederländischen (0,026).

Vergleicht man den Informationsgehalt der lexikalischen Töne im Chinesischen miteinander, so hat der Kontrast zwischen hohem \uparrow und fallendem \searrow den stärksten Informationsgehalt (2,9), der Kontrast zwischen dem dritten Ton \sphericalangle und dem entsprechenden neutralen Ton \downarrow den geringsten (0,02).

(Gendrot & Adda-Decker, 2007) untersuchen die Formanten und die Vokaldauer in acht Sprachen, nämlich im Arabischen, Englischen, Französischen, Deutschen, Italienischen, Chinesischen, Portugiesischen und Spanischen. Dabei zeichnet sich für ihn in allen untersuchten Sprachen eine Tendenz zur Vereinfachung der Formantstruktur bei Vokalen mit kurzer Dauer ab.

Die Formantwerte werden zuerst von Hertz in Bark umgerechnet. Danach wird die Differenz zwischen F_1 und F_0 als charakteristischer Wert für den Öffnungsgrad des Kiefers verwendet, die Differenz zwischen F_3 und F_2 als charakteristischer Wert für die horizontale Zungenposition. Für jede Sprache wird der Abstand zwischen /i/, /a/ und /u/ anhand dieser Maßzahlen errechnet, um die Ausdehnung des Vokalraums zu beschreiben. Als nächstes werden die Vokale nach ihrer Dauer in kurze ($< 60\text{ms}$), mittlere ($60\text{ms} < d < 80\text{ms}$) und lange Abschnitte ($> 80\text{ms}$) eingeteilt. Als letztes wird pro Gruppe wieder der Abstand zwischen /i/, /a/ und /u/ errechnet, der sich je nach Vokaldauer für die europäischen Sprachen deutlich unterscheidet, für Chinesisch und Arabisch jedoch nicht.

(Gu, Mori & Kasuya, 2003) vergleicht Aufnahmen der fünf Silben $[\uparrow da]$, $[\uparrow d\epsilon]$, $[\uparrow di]$, $[\uparrow do]$ und $[\uparrow du]$ in denen diese Silben entweder den Satzakkzent tragen oder innerhalb des Satzes nicht betont werden. Bei der Analyse der Ergebnisse liegt F_1 für /a/ in Silben mit Satzakkzent wesentlich höher als in unbetonten Silben, bei / ϵ / und /o/ ist der Unterschied weniger ausgeprägt und bei /i/ und /u/ nicht vorhanden. F_2 liegt für /i/ in betonten Silben wesentlich höher und für /u/ wesentlich niedriger.

(Chen, 2002) untersucht die phonetischen Eigenschaften einsilbiger Worte, die innerhalb eines Satzes den Satzakkzent tragen oder unbetont bleiben. Sie stellt einen signifikanten

Unterschied für die Vokallänge bei betonten und unbetonten Worten fest.

In den letzten beiden Arbeiten kann man, im Gegensatz zur der vorher zitierten Untersuchung einen deutlichen Unterschied zwischen betonten und unbetonten Silben ausmachen. Dies könnte jedoch dadurch bedingt sein, dass der Wortakzent, der in indogermanischen Sprachen unter anderem durch die Vokalqualität umgesetzt wird, im Chinesischen schon durch die Töne abgedeckt wird. Der Unterschied zwischen zwei verschiedenen Lesarten eines Satzes, in denen das gleiche Wort einmal den Satzakzent trägt und einmal nicht, kann jedoch nicht wie im Deutschen durch die Satzintonation signalisiert werden. Deshalb werden v.a. periphere Vokale wie /i/, /a/ oder /u/ in diesem Kontext möglicherweise deutlicher artikuliert.

6.1.3 Durchschnittswerte für deutsche und chinesische Vokale

6.1.3.1 Vergleich des Vokalraums

Die durchschnittlichen Werte für deutsche Muttersprachler sind aus (Simpson, 1988, S. 214 u. 216), entnommen, die für chinesische Muttersprachler aus (Meng, Chen & Li, 2006). Die deutschen Vokale stammen dabei aus Formantmessungen in entsprechenden Abschnitten gelesener Texte, die nachträglich noch durch eine gewichtete Bewertung möglicher Messwerte korrigiert wurde.

Die chinesischen Aufnahmen bestehen aus einsilbigen Worten, die von angehenden Peking-Radiosprechern aufgenommen wurden. Die chinesischen Silben werden alle im 1. Ton gesprochen, d.h. mit gleichbleibend hoher F_0 : ɿyi (Kleidung), ɿwu (Rabe, Krähe, schwarz), ɿyü (auf Umwegen, indirekt), ɿbo (Schwingung, Welle), ɿa (ach, autsch), ɿzi (Mittel, Kapital), ɿzhi (wissen) und ɿei (hey). Eine Ausnahme bilden [ɿɛ] (Gans, Gänse) und [dɿ], das als Attributpartikel in einem neutralen Ton ausgesprochen wird.

Die Durchschnittswerte für einzelne Vokale liegen bei chinesischen Sprechern weiter auseinander, obwohl das chinesische Vokalsystem insgesamt weniger Vokale enthält.

Im grafischen Vergleich fallen vor allem die Allophone von /i/, /ɪ/ und /ɨ/ auf, deren Formantwerte zentraler ausfallen, als die der deutschen Phoneme /i:/ und /ɪ/.

Bei den Allophonen von /i/, /ɪ/ und /ɨ/, liegen die Werte für F_2 und F_3 bei /ɪ/ weiter auseinander als bei /ɨ/, wobei /ɪ/ einen niedrigeren F_2 -Wert, jedoch einen höheren F_3 -Wert besitzt als /ɨ/.

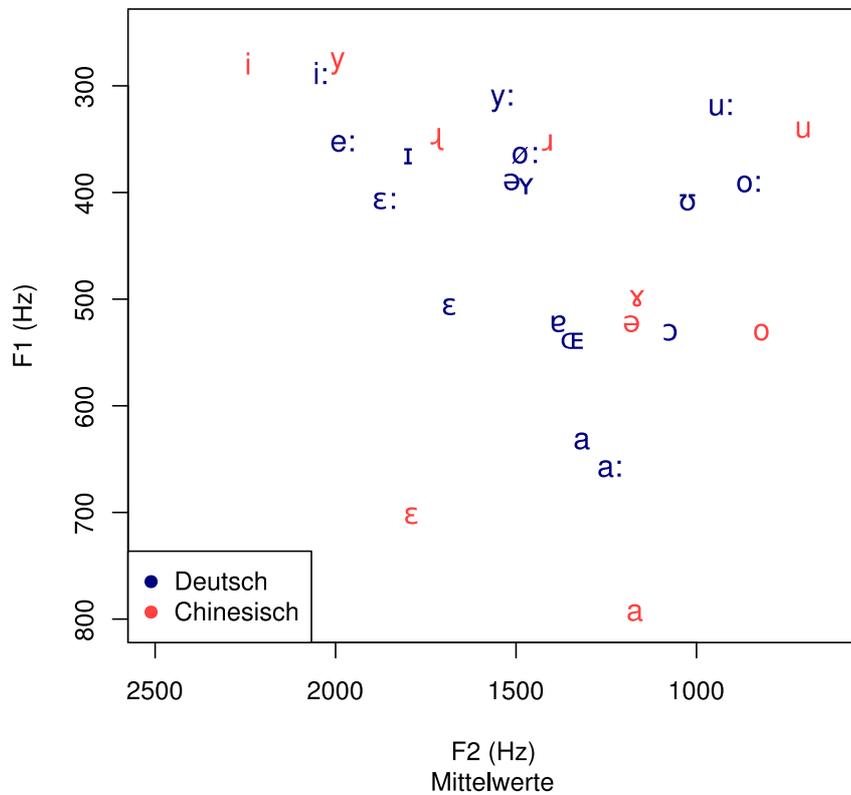
Vergleiche ich die männlichen Formantwerte mit denen der Wortlisten in (Ramers, 1988, S. 189-192), so werden /i/ /ɛ/, /o/ und /u/ zentraler artikuliert als im Deutschen, /a/ dagegen extremer.

Die zentralere Artikulation der hohen und mittleren Vokale entspricht dem geringeren Informationsgehalt der Vokalqualität. /a/ könnte von chinesischen Sprechern als isoliert ausgesprochener Vokal stärker betont worden sein.

6.1.3.2 Die Allophone /i/, /ɪ/ und /ɨ/

(Man, 2004, S. 3) führt Perzeptionstests mit synthetisierten Versionen von /ɪ/ bzw. /ɨ/ durch. Nach einem ersten Versuchsdurchgang mit den synthetisierten Silben [sɪ] und [ʃɨ] ersetzt er die vorhergehenden Frikative /s/ oder /ʃ/ durch Rauschen, da die Hörer sonst dazu tendieren unabhängig von den Werten für F_1 und F_2 die auf /s/ folgenden Vokale als /ɪ/ und die auf /ʃ/ folgenden Vokale als /ɨ/ einordnen. Was die F_3 -Werte betrifft, so werden synthetisierte Stimuli mit Werten von 1700 – 2300 Hertz eher als /sɨ/ eingeordnet, Stimuli mit Werten über 2300 Hertz eher als /sɪ/

Abbildung 6.2: Durchschnittliche Werte für F_1 und F_2 bei männlichen Sprechern



(jeweils nach (Simpson, 1988, S. 214 u. 216) für die deutschen und (Meng et al., 2006) für die chinesischen Formantwerte)

Abbildung 6.3: Durchschnittliche Werte für F_1 und F_3 bei männlichen Sprechern

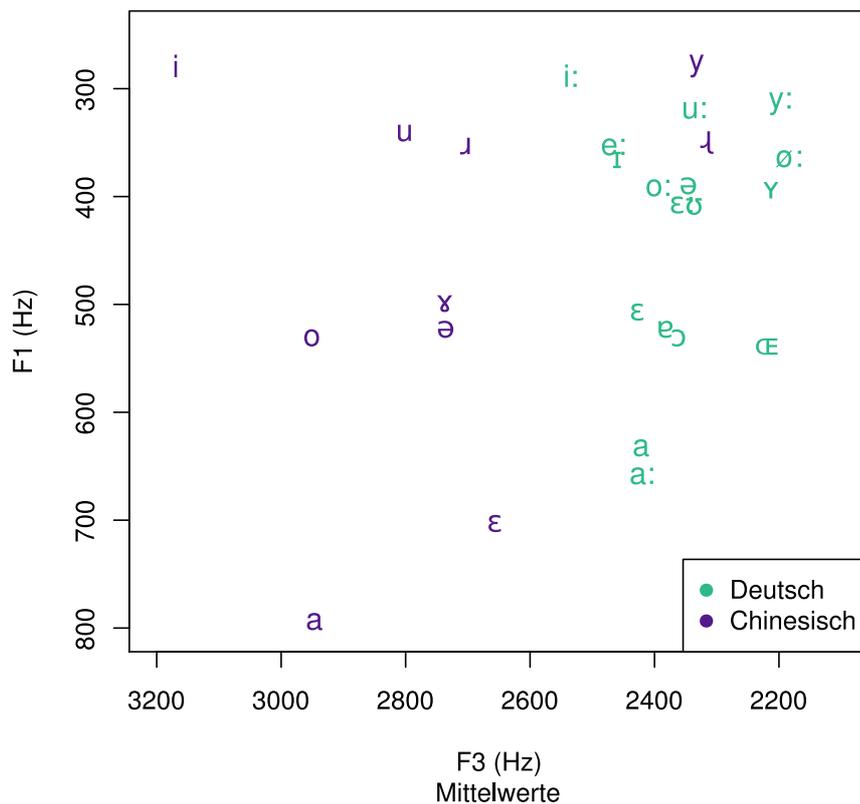
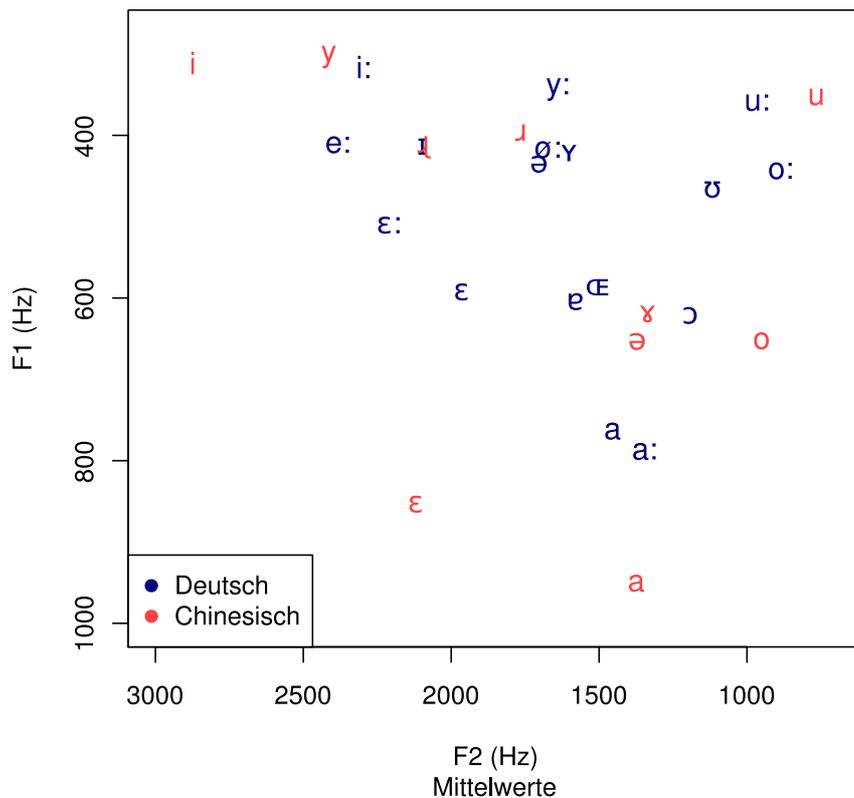
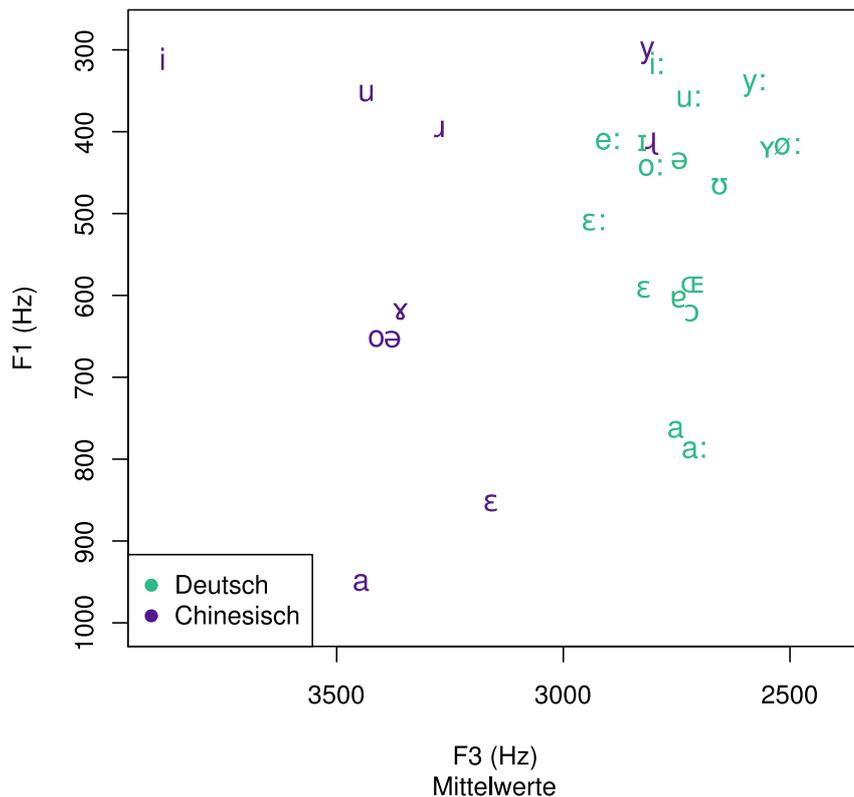


Abbildung 6.4: Durchschnittliche Werte für F_1 und F_2 bei weiblichen Sprechern



(jeweils nach (Simpson, 1988, S. 214 u. 216) für die deutschen und (Meng et al., 2006) für die chinesischen Formantwerte)

Abbildung 6.5: Durchschnittliche Werte für F_1 und F_3 bei weiblichen Sprechern



(Howie, 1976) mißt bei Studien mit Aufnahmen eines chinesischen Muttersprachlers, die als Grundlage für Perzeptionstests mit fünf anderen chinesischen Muttersprachlern dienen, ebenfalls eher hohe F_3 -Werte.

(Zee & Lee, 2001) analysiert die ersten drei Formantwerte in einsilbigen Worten, die von zehn männlichen und zehn weiblichen Sprechern des Hochchinesischen aus Beijing gelesen werden.

Seine Messergebnisse für F_3 liegen bei männlichen Sprechern zwischen 3385 Hz für /i/ und 1962 Hz für /ɿ/ und bei weiblichen Sprechern zwischen 3847 Hz für /i/ und 2466 Hz für /ɿ/. Beim Vergleich von offenen Silben und Silben die auf Nasale enden, ergeben sich für F_3 keine signifikanten Unterschiede.

Vergleicht man die Messwerte für /ɿ/ und für /ɿ/ mit den deutschen Vokalwerten aus (Ramers, 1988, S. 189-192), so ist /ɿ/ dem deutschen /ɪ/ am nächsten, weist jedoch einen niedrigeren F_2 -Wert auf. Die Position von /ɿ/ entspricht der des Vokals /ø:/ im deutschen Vokalraum, was auch mit meinem Höreindruck bei der Resynthese dieses Vokals übereinstimmt.

6.1.4 Eigene Messungen in chinesischen Aufnahmen

Da die ungewöhnlich hohen F_3 -Werte in unterschiedlichen Arbeiten auftauchen und daher nicht durch individuelle Messfehler bedingt sind, führe ich selbst Messungen an muttersprachlichen chinesischen Silben durch.

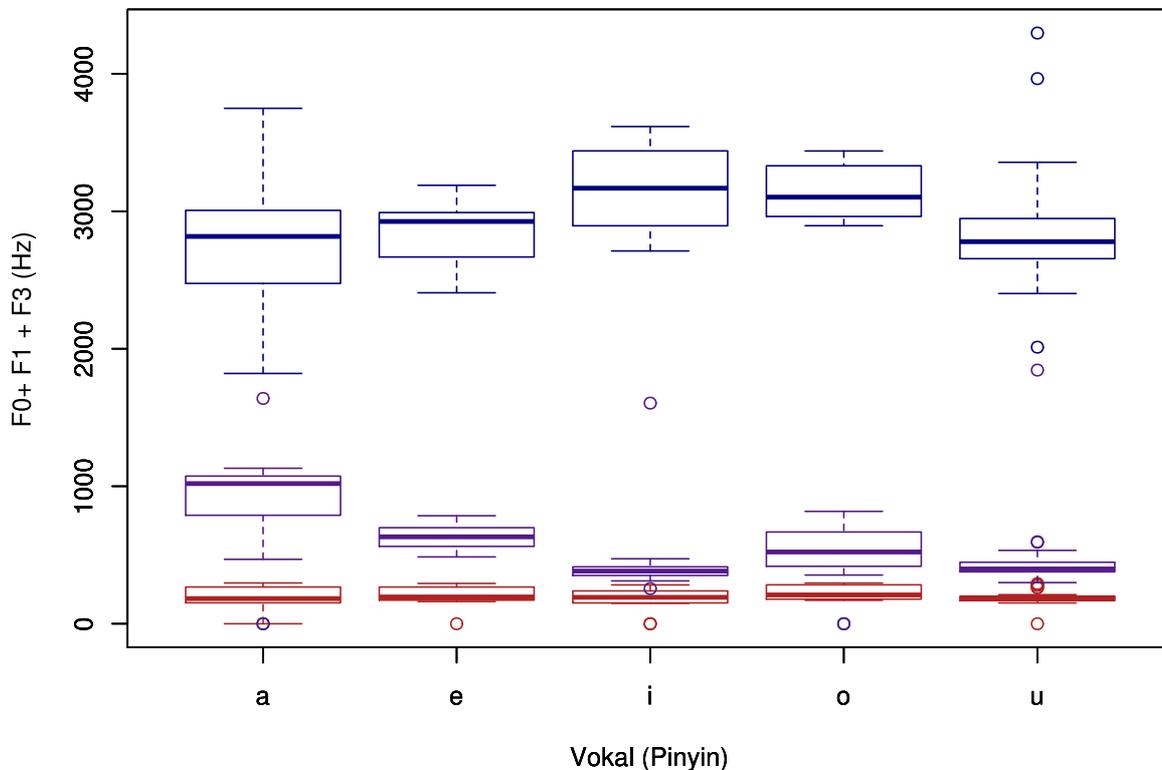
Dabei habe ich glücklicherweise vom Betreiber der Website <http://www.chineselessons.com>, Michael Thigpen, die Erlaubnis erhalten, seine Sprachaufnahmen chinesischer Silben zu verwenden. (s. (Thigpen, 2010)).

Diese Aufnahmen umfassen 1189 Silben, die von einer chinesischen Muttersprachlerin aufgenommen wurden und decken damit das gesamte Silbeninventar ab. Damit ich eindeutige Ergebnisse erhalte, wähle ich aus den Audiodateien 89 Aufnahmen aus, bei denen Monophthonge in offenen Silben auf stimmlose unaspirierte Plosive, Laterale oder Glides folgen oder isoliert aufgenommen sind. Würde ich die im Chinesischen zahlreichen Frikative oder Nasale bzw. nasalierte Vokale mit einbeziehen, würde dies die Messergebnisse u.U. verzerren. Viele der hohen Formantwerte treten entweder für den Vokal /i/ auf oder bei Vokalen, die im ersten Ton, d.h. mit einer höheren Grundfrequenz gesprochen wurden. Da die Aufnahmen von einer weiblichen Sprecherin stammen, möchte ich Verwechslungen mit der Grundfrequenz, durch die die restlichen Formanten nicht korrekt gemessen werden könnten, ausschließen. Dabei wird sichtbar, dass die Werte für F_3 um ca. 3000 Hertz liegen, was phonetisch einem Gesangsformanten entspricht.¹

Laut (Sundberg, 1977) entsteht der Gesangsformant bei allen gesungenen Vokalen im Bereich von 2500 bis 3000 Hertz zwischen F_3 und F_4 , indem sich der Kehlkopf senkt, wodurch der Rachenraum erweitert wird. Für ihn ändert sich damit bei gesungenen Vokalen die Vokalqualität, so dass z.B. /i/ als /y/ gesungen wird.

¹Hinweis von Ingo Hertrich

Abbildung 6.6: Formantwerte einer chinesischen Muttersprachlerin nach Vokal



Bei (Sundberg, 1972) liegen die Werte für F_2 und F_3 bei den vorderen gesungenen Vokalen /i/, /y/ und /e/ niedriger als bei den entsprechenden gesprochenen Vokalen während gleichzeitig zwischen F_3 und F_4 der Gesangsformant bei ca. 3000 Hz erscheint.

Wenn es sich bei den hohen gemessenen F_3 -Werten um einen Gesangsformanten mit einer entsprechenden Stimmqualität handeln würde, müsste bei den Formantmessungen ein niedrigerer F_3 -Wert ausgelassen oder von dem entsprechenden Gesangsformanten überdeckt worden sein. Um dies zu prüfen, resynthetisiere ich die Messergebnisse mit Praat und kann danach ausschließen, dass ein F_3 -Wert bei den Messungen ausgelassen wurde, da der Klang der resynthetisierten chinesischen Vokale dem der untersuchten Aufnahmeabschnitte entspricht.

6.1.5 Physiologischer Vergleich

(Jackson, 1988, S. 91ff) vergleicht anhand von Röntgenaufnahmen den chinesischen Vokalraum mit dem französischen und dem Vokalraum für Akan, auch Twi, eine in Ghana gesprochene Tonsprache. Außerdem greift er auf die Ergebnisse einer Studie mit Röntgenaufnahmen amerikanischer Sprecher zurück.

Er setzt ein PARAFAC-Modell ein, das auf der principal component analysis (PCA) basiert, um aus den Stellungen der Artikulationsorgane in den Röntgenaufnahmen Gemeinsamkeiten abzuleiten. Anstelle der üblichen front-back und high-low Kategorien findet er für Chinesisch zwei Faktoren.

In dem englischen Modell in (Harshman et al., 1977) , das er mit seinen Ergebnissen vergleicht, bestehen die zwei Faktoren aus:

Front-Raising der Grad der Hebung der Zungenspitze nach vorne von /i/ (am stärksten) bis /o/ (kaum vorhanden).

Back-Raising der Grad der Hebung des Zungenrückens nach hinten von /u/ bis /ɑ/.

In seinem chinesischen Modell erhält er ebenfalls zwei Faktoren, wobei die Anhebung des Kiefers und damit auch der Unterlippe eine größere Rolle spielt als im Englischen. Die Lippenrundung wird in diesem Modell nicht als Faktor berücksichtigt, dies wäre jedoch möglich, wenn nicht nur Phoneme, sondern auch Allophone mit einbezogen würden.

Wird ein sprachübergreifendes statistisches Modell für alle drei Sprachen erstellt, so ergeben sich fünf Faktoren unter denen der dritte, der die Lippenrundung von einem gerundeten /o/ zu einem /i/ und der vierte, der die Bewegung des Zungenrückens und die Kehlkopfsenkung von /i/ nach /o/ beschreibt, sich für Akan und Chinesisch als relevant herausstellen, s. (Jackson, 1988, S 110)).

Da das Absinken der F_3 sowohl durch die Rundung der Lippen als auch durch das Absinken des Kehlkopfs hervorgerufen werden kann, gehe ich grundsätzlich davon aus, dass die höheren F_3 -Werte, insbesondere im ersten Ton \uparrow , durch eine höhere Position des Kehlkopfs hervorgerufen werden.

6.2 Direkter Vergleich der Werte aus dem LeaP-Corpus

In diesem und den folgenden Unterkapiteln verwende ich noch eigene Messungen im LeaP-Corpus, die sowohl von chinesischen Studenten in Deutschland als auch von einer deutschen Kontrollgruppe stammen.

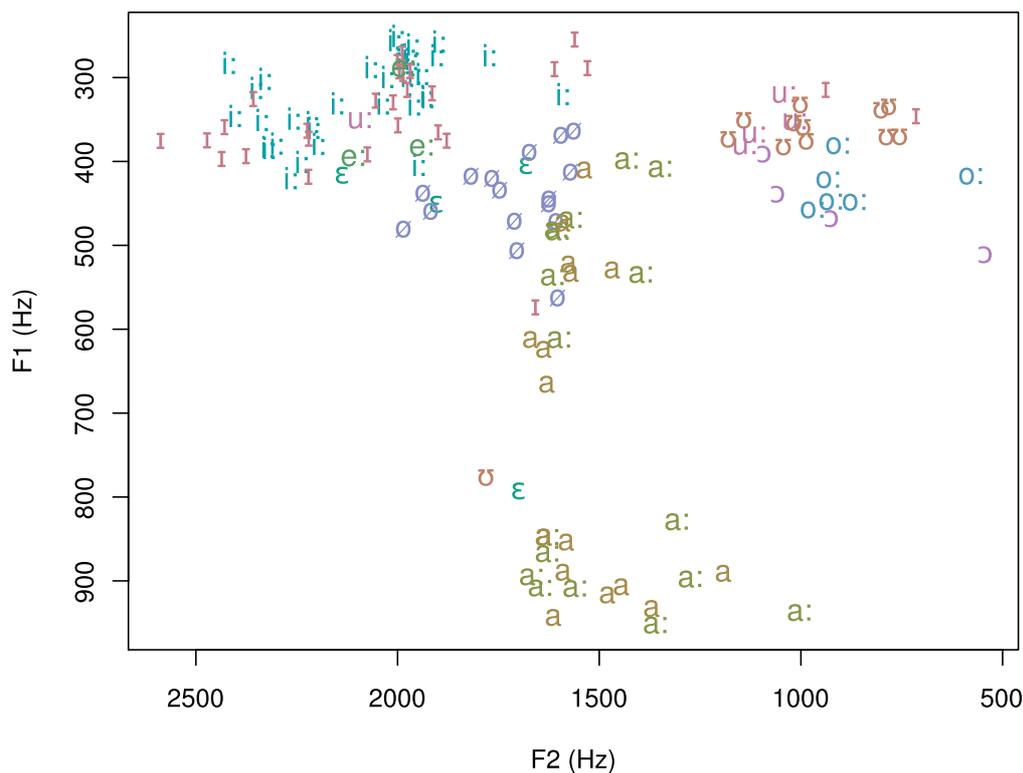
Leider besitze ich keine genaueren Details zur Herkunft der chinesischen Sprecher, z.B. ob sie aus Taiwan oder einer der Provinzen der VR China stammen. Es handelt sich um zwei Männer und vier Frauen, wobei alle Teilnehmer vor dem Besuch eines Sprachkurses aufgenommen wurden sowie eine Teilnehmerin nach dem Kurs.

Dieses Unterkapitel enthält einen ersten rein visuellen Vergleich der Formantwerte mittels verschiedener grafischer Darstellungen. Die aus diesen Darstellungen gezogenen Schlussfolgerungen können als Grundlage für die statistischen Untersuchungen in den nächsten Kapiteln dienen.

6.2.1 Männliche Sprecher

6.2.1.1 Vokalraum

Abbildung 6.7: Durchschnittliche Werte für F_1 und F_2 bei deutsch sprechenden chinesischen Studenten



Diese Grafik enthält die Messwerte aller Monophthonge aus den deutschen Äußerungen chinesischer Studenten.

Dabei fallen die stark überlappenden Bereiche bei den hohen, vorderen Vokalen auf, während ein Teil der /a/-Werte einen deutlich abgegrenzten Bereich bildet. Die restlichen /a/-Werte bilden zusammen mit den und /a:/ und einem Teil der Abschnitte für /ø:/ einen Bereich, der durch den muttersprachlichen Vokal /ɤ/ bedingt ist. Dieser Vokal liegt vom Klang her zwischen /ø:/ und /a/.

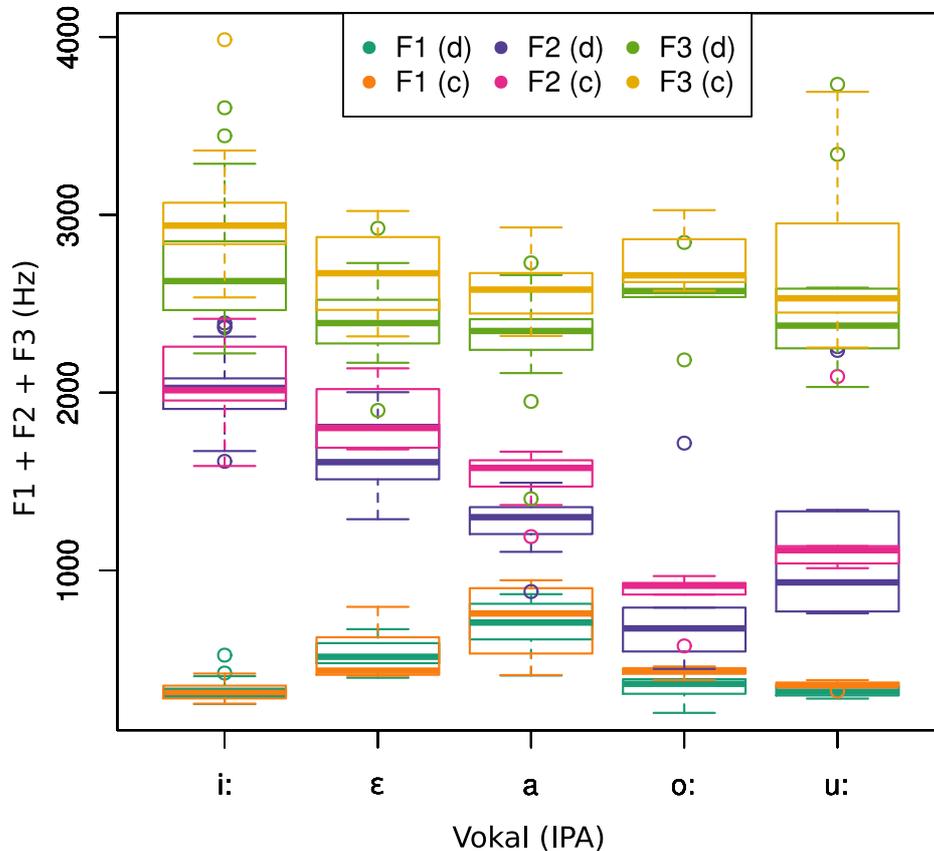
Die restlichen Abschnitte für /ø:/ und ein Teil der Abschnitte für /ɛ/ stellen einen weiteren Bereich dar, was ebenfalls durch die chinesische Vokalstruktur beeinflusst sein könnte, in der /ɤ/ und /ɛ/ Allophone bilden.

Die starke Streuung der Werte für /i/ und /ɪ/ läßt sich dadurch erklären, dass der vordere Vokal *i* im Chinesischen mittlere Allophone besitzt.

Bei den hinteren Vokalen bilden die Werte für /o/ und /ɔ/ sowie für /u/ und /ʊ/ Gruppen, was dadurch bedingt sein könnte, dass im Chinesischen nur die Vokale /o/ und /u/, jedoch nicht /ɔ/ und /ʊ/ vorhanden sind, und so der Gegensatz zwischen gespannten und ungespannten hinteren Vokalen fehlt.

6.2.1.2 Gemeinsame Vokale

Abbildung 6.8: Formantwerte für Vokale von chinesischen Muttersprachlern, die im Deutschen und im Chinesischen auftreten



Anmerkung: /y/ wäre auch ein gemeinsamer Vokal, leider existieren für diesen Vokal keine geeigneten Samples von chinesischen Sprechern.

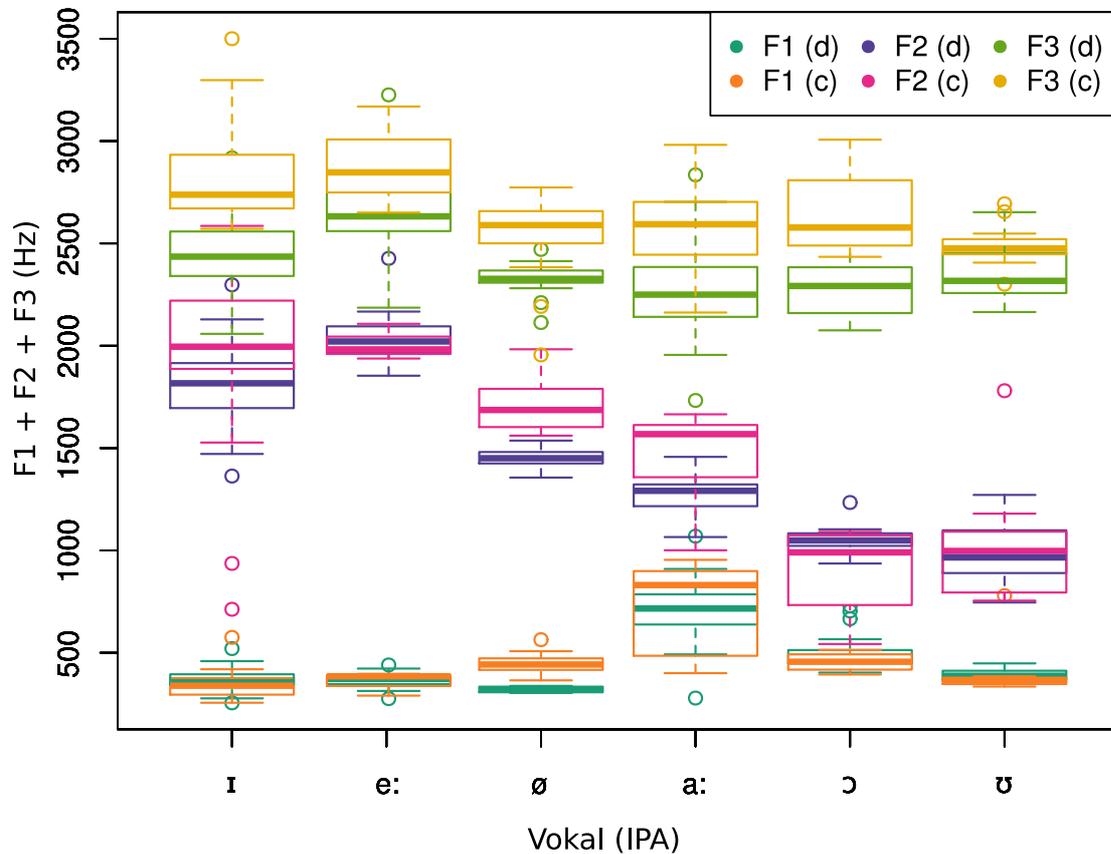
Der Wertebereich ist für die nicht-hohen Vokale /ε/ und /a/ bei den chinesischen Sprechern größer, wobei sich bei /a/ auch der koartikulatorische Aufwand durch die Einbettung zwischen im Chinesischen nicht vorhandene Konsonantencluster bemerkbar macht, z.B. bei den Wort *Spatz*, das mehrmals in den gelesenen Texten vorkam.

Die F_2 -Werte liegen bei allen nicht-hohen Vokalen, also /ε/, /a/, /o:/ und /u:/ über den deutschen Werten. Der Umfang der Formantwerte ist bei /i:/ und /ε/ geringer, bei /a/ gleich groß und bei /o/ und /u/ größer. Dabei werden bei den hinteren Vokalen ebenfalls die Auswirkungen unbekannter Konsonantencluster, die z.B. bilabial oder alveolar artikuliert werden, wie in *Pfote* oder *herunterzufallen* [tsu:f] sichtbar.

Die F_1 -Werte bei /o:/ und /u:/ liegen bei den chinesischen Muttersprachlern höher als bei den deutschen Sprechern und damit näher an /ɔ/ und /ʊ/. Da im Chinesischen die Laute /ɔ/ und /ʊ/ fehlen, müssen chinesische Sprecher darauf achten, zwischen diesen Lauten und den bekannten Vokalen /o/ und /u/ deutlich zu trennen.

6.2.1.3 Nur im Deutschen vorhandene Vokale

Abbildung 6.9: Formantwerte für Vokale von chinesischen Muttersprachlern, die nur im Deutschen und nicht im Chinesischen auftreten



Die F_1 -Werte bei /ø:/ sind bei chinesischen Sprechern höher als bei deutschen. Der Wertebereich bei /a:/ ist deutlich weiter. Bei /ø:/ und /a:/ liegen die F_2 -Werte höher. Die F_2 -Werte liegen für /e:/ und /ɔ/ niedriger, wobei der Wertebereich für /ɔ/ einen größeren Umfang besitzt.

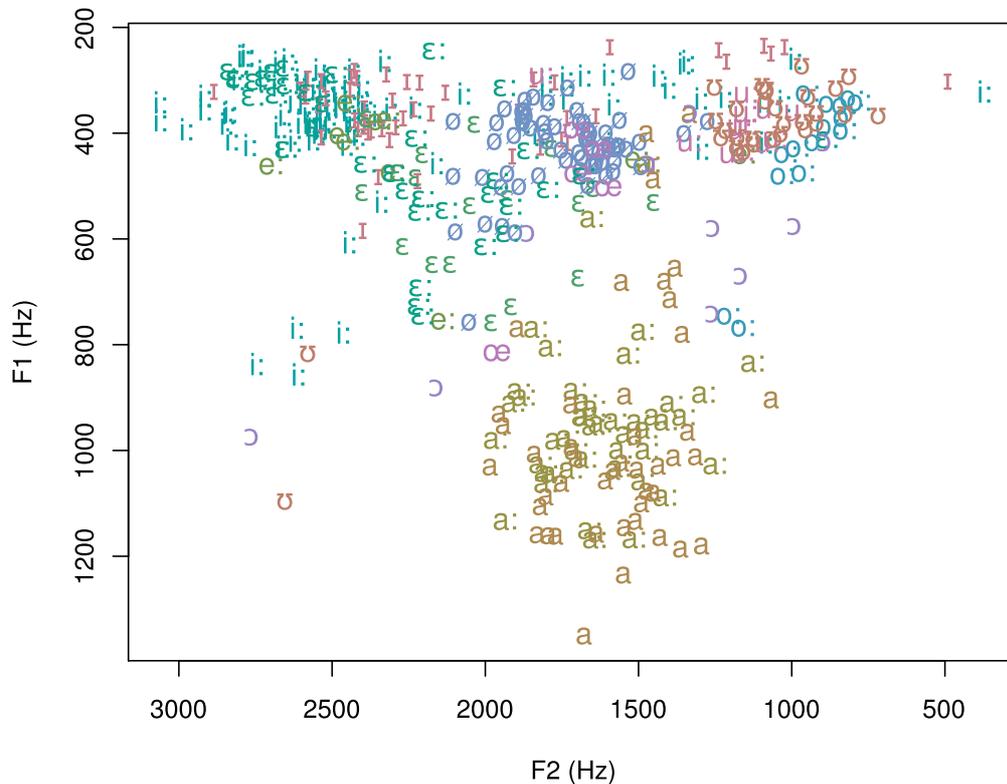
/i/ und /u/ werden extremer artikuliert, als im Deutschen, da F_1 und F_2 hier bei chinesischen Sprechern niedriger liegen.

Für /e:/ liegen wenige Messwerte vor, die vorliegenden F_2 -Werte liegen jedoch näher bei den F_2 -Werten für /ɛ/ als bei den deutschen Sprechern.

6.2.2 Weibliche Sprecher

6.2.2.1 Vokalraum

Abbildung 6.10: Gesamte Werte für F_1 und F_2 bei Deutsch sprechenden chinesischen Studentinnen



Da bei den Sprachaufnahmen vier weibliche chinesische Sprecherinnen, jedoch nur zwei männliche chinesische Sprecher vorhanden sind, werden hier durch die größere Anzahl der Abschnitte einige Merkmale deutlicher.

So besitzt hier nicht nur /i:/, sondern auch /ɛ:/ sehr stark gestreute Formantwerte, wobei die Formantwerte für /e:/ und /ɛ:/ innerhalb des Wertebereichs von /ɛ:/ liegen und stärker begrenzt sind.

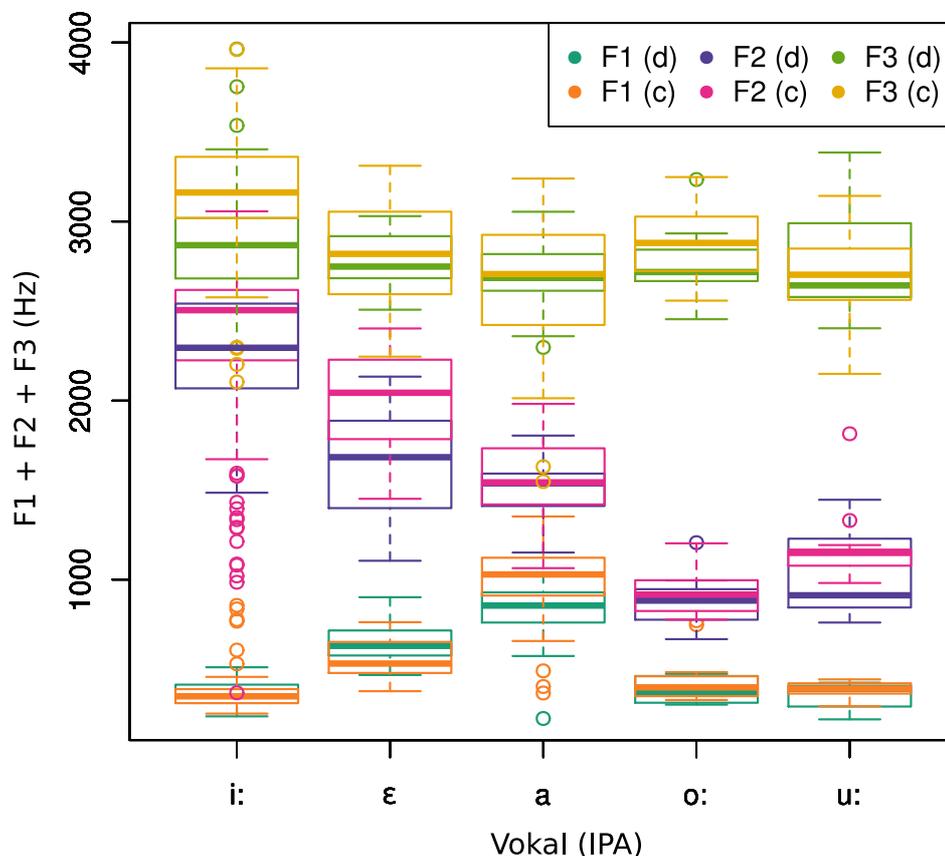
Die Formantbereiche für /ø:/ und /œ:/ überlappen sich ebenfalls, wobei die F_1 -Werte für /œ:/ höher liegen als für /ø:/.

Die Werte für /o:/ bilden zwei Gruppen, eine Gruppe mit niedrigeren F_1 - und F_2 -Werten, die stark mit /u/ überlappt und einzelne Abschnitte mit höheren F_1 - und F_2 -Werten, die eher dem muttersprachlichen /o/ entsprechen.

Die Werte für /a/ und /a:/ sind stark gestreut, liegen jedoch in den meisten Fällen weiter von /ø:/ entfernt als bei den männlichen Sprechern.

6.2.2.2 Gemeinsame Vokale

Abbildung 6.11: Formantwerte für Vokale von chinesischen Muttersprachlerinnen, die im Deutschen und im Chinesischen auftreten



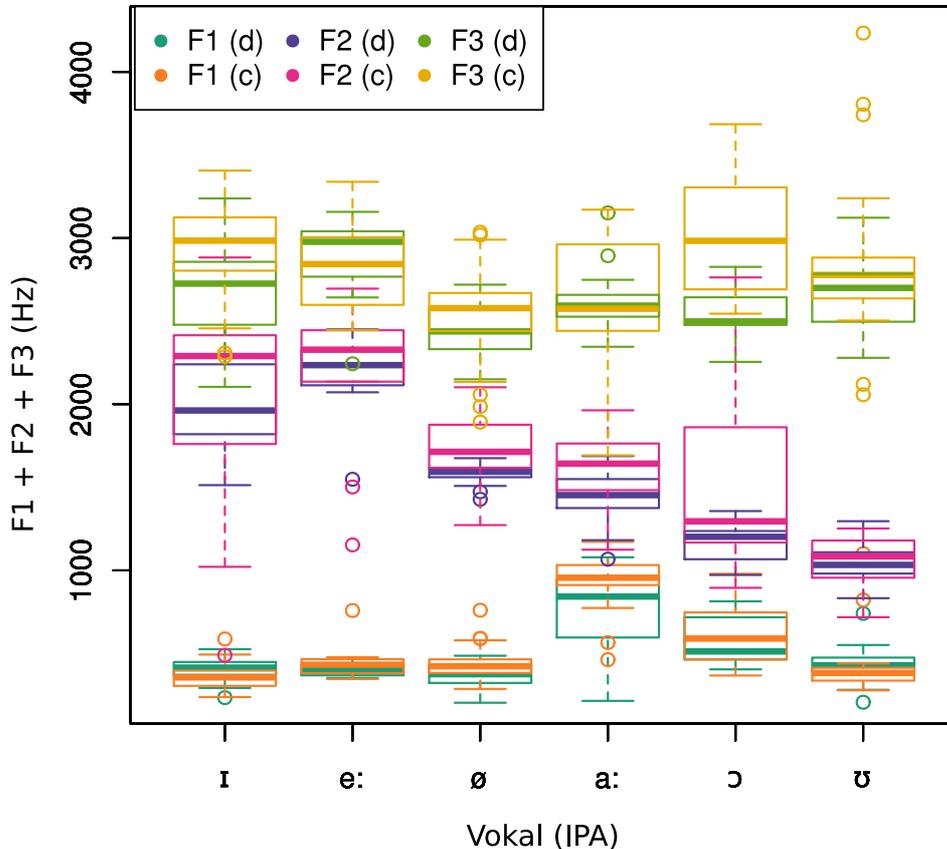
Die Formantwerte für F_1 und F_2 liegen bei $/\epsilon/$ für chinesische Sprecherinnen weiter auseinander als bei deutschen, was bedeutet, dass der Vokal peripherer artikuliert wird. Betrachtet man $/i:/$, sind die Werte für F_1 und F_2 bei chinesischen Sprecherinnen tendenziell weiter voneinander entfernt, es existieren jedoch mehr Ausreißer, bei denen die Formantwerte enger zusammenliegen.

Die Werte für F_1 bei $/a/$ fallen bei chinesischen Sprecherinnen höher aus als bei deutschen, der Umfang der Formantwerte unterscheidet sich jedoch nicht so stark wie bei den männlichen Sprechern.

Was am stärksten auffällt, sind die Werte für F_3 , die sich hier für chinesische und deutsche Sprecherinnen nicht so deutlich unterscheiden, wie bei den männlichen Sprechern. Dies ist natürlich auch durch die stärkere Streuung der Formantwerte bei weiblichen Sprecherinnen bedingt. Trotzdem wird durch die näher beieinander liegenden Mittelwerte deutlich, dass sich die Aussprache der weiblichen chinesischen Sprecherinnen näher am deutschen muttersprachlichen Modell befindet.

6.2.2.3 Nur im Deutschen vorhandene Vokale

Abbildung 6.12: Formantwerte für Vokale von chinesischen Muttersprachlerinnen, die nur im Deutschen und nicht im Chinesischen auftreten



Die Werte für F_1 sind für /a:/ und /ɔ/ bei chinesischen Sprecherinnen höher als bei deutschen Muttersprachlerinnen, für /i/ und /u/ niedriger. Für /ɔ/ sind die Wertebereiche in allen drei Formanten bei chinesischen Sprecherinnen weiter. Allerdings liegen bei den chinesischen Versuchspersonen viele überprüfte Abschnitte vor einem Nasal, was zu Verzerrungen im Frequenzspektrum führen könnte.

Für /i/, /ø:/ und /u/ sind die F_2 -Werte bei chinesischen Sprecherinnen sowohl höher als auch weiter gestreut, was bei den hohen Vokalen /i/ und /u/ auf eine peripherere Artikulation hinweisen könnte.

Für F_3 sind die Formantwerte bei chinesischen Versuchspersonen nicht generell höher als bei deutschen Sprechern, bei /e:/ liegen sie sogar niedriger. Die Durchschnittswerte für F_3 bei /i:/, /ø:/, /ɔ/ und /u/ liegen bei den chinesischen Sprecherinnen deutlich höher, wobei sich hier eine fehlende Kehlkopfsenkung und Lippenrundung zeigt.

6.3 Vergleich der Vokaldauer

Um festzustellen, ob die Vokaldauer bei den chinesischen Sprechern allgemein länger ist, als bei den deutschen Sprechern, vergleiche ich alle Vokale, die ausreichend große Wertemengen vorweisen.

Da ich im LeaP-Corpus Vokalabschnitte mit mehreren Vokalen ausschließen musste, sind hier Abschnitte in der Wortmitte gehäuft vertreten, jedoch nur in Worten, die nicht auf /ɐ/ enden (s. auch 3.2.6). Dabei wende ich den verteilungsunabhängigen Wilcoxon-Rangsummentest auf gleich große Testmengen an und führe zuerst zweiseitige Tests aus, bei denen keine Annahmen darüber vorliegen, wie sich die Testmengen unterscheiden. Als nächstes führe ich noch einseitige Tests aus, bei denen ich prüfe, ob die Vokaldauern der chinesischen Sprecher grundsätzlich länger sind als die der deutschen.

6.3.1 Männliche Sprecher

Tabelle 6.3: Vergleich der Vokaldauer bei deutschen und chinesischen männlichen Sprechern

Vokal	p-Wert				Durchschnitt (ms)	Durchschnitt (ms)
	(zweiseitig)		(einseitig)		(deutsch)	(chinesisch)
	original	korrigiert	original	korrigiert		
/i/	0.00746	0.03732	0.00373	0.01866	75,31	97,25
/ɪ/	0.00506	0.03034	0.00253	0.01517	50,82	70,11
/ø:/	0.64367	0.64367	0.69636	0.93308	149,8	148
/a:/	0.14302	0.42906	0.93308	0.93308	89,71	79,65
/a/	0.00012	0.00083	0.00006	0.00042	60,50	111,80
/u/	0.46334	0.64367	0.23167	0.69502	44,8	57,6
/ʊ/	0.14078	0.42906	0.07039	0.28156	56,25	68,50

Die p-Werte wurden mit Hilfe der Simes-Hochberg-Prozedur korrigiert, um die Wahrscheinlichkeit falsch positiver Ergebnisse bei mehrfachen Tests zu verringern. Auf der Grundlage dieser Vergleiche kann ich nicht von einer generell längeren Vokaldauer bei chinesischen Sprechern ausgehen.

Die peripheren Vokale /i:/ und /ɪ/, die auch bei Formantmessungen eine große Streuung an Werten aufweisen, besitzen sowohl bei männlichen als auch bei weiblichen Sprechern signifikante verschiedene Werte.

Da /a:/ von chinesischen Sprechern kürzer und /a/ länger artikuliert wird als bei deutschen Sprechern, kann man davon ausgehen, dass die chinesischen Sprecher bei der Vokalquantität nicht zwischen /a:/ und /a/ unterscheiden.

/i:/ und /ɪ/ werden länger artikuliert als bei deutschen Sprechern, was durch chinesische

Lautkombinationen wie [ji] am Wortanfang bedingt sein könnte.

Die Vokallängen von /u:/ und /ʊ/ nähern sich dagegen an, was darauf hinweist, dass chinesische Sprecher diese Vokale als einen Laut betrachten.

6.3.2 Weibliche Sprecher

Tabelle 6.4: Vergleich der Vokaldauer bei deutschen und chinesischen weiblichen Sprechern

Vokal	p-Wert				Durchschnitt (ms)	
	(zweiseitig)		(einseitig)		(deutsch)	(chinesisch)
	(original)	(korrigiert)	(original)	(korrigiert)	größer	
/i/	0.00016	0.00179	0.00008	0.00090	69,57	103,7
/ɪ/	0.21142	1	0.10571	0.73995	63,06	67,44
/e/	0.14090	0.98627	0.07045	0.56359	82,25	102
/ɛ/	0.02091	0.19442	0.01045	0.09721	73,85	103,3
/ø:/	0.02160	0.19442	0.01080	0.09721	134,5	158,5
/æ/	1	1	0.51152	0.95197	148,8	150,9
/a:/	0.09920	0.79360	0.95197	0.95197	101,1	83,24
/a/	0.00004	0.00044	0.00002	0.00022	79,19	119,30
/o/	0.37057	1	0.18528	0.92642	121,4	128,6
/ɔ/	0.25075	1	0.12538	0.75226	70,89	78,22
/u/	0.42654	1	0.81151	0.95197	77,56	65,67
/ʊ/	0.66204	1	0.33102	0.95197	64,50	68,29

Der Vokal /ø:/ fällt in diesem Vergleich extrem lang aus. Im Gegensatz zu den männlichen Sprechern ist dieser Vokal bei den chinesischen Sprecherinnen signifikant länger als bei deutschen Sprecherinnen.

Der Vokale /i:/ besitzt eine signifikant längere Vokallänge, /ɪ/ wird jedoch, anders als bei den männlichen Sprechern bei deutschen und chinesischen Sprecherinnen gleich lang ausgesprochen.

Der Vokal /ɛ/ wird von chinesischen Sprecherinnen länger artikuliert, so dass seine Vokallänge sich /e:/ annähert. Wie bei den Männern fällt /a:/ bei chinesischen Teilnehmerinnen kürzer und /a/ länger als bei deutschen Sprechern. Analog zu den a-Lauten wird /u:/ bei chinesischen Sprecherinnen ebenfalls kürzer und /ʊ/ länger artikuliert.

Die Vokale /e/, /æ/, /o/ und /ɔ/ werden länger artikuliert, der Unterschied ist jedoch nicht signifikant. Hier könnte eine sorgfältigere Artikulation beim Lesen eines fremdsprachlichen Textes eine Rolle spielen. Da es sich bei /e/, /æ/ und /o/ laut (Duanmu, 2000, S 38) im Chinesischen um Allophone handelt, ist es vorstellbar, dass die Studenten die

deutschen Vokale mit ihrer größeren phonetischen Bedeutung benutzt deutlicher artikulieren.

6.4 Einfluss muttersprachlicher Vokale

6.4.1 /a/ und /ɤ/

6.4.1.1 Hypothese

Untersuchungsgegenstand 6 (Einfluss des muttersprachlichen Vokals /ɤ/)

In diesem Unterkapitel soll untersucht werden, ob die Formantwerte für /a:/ und /a/ bei deutschen und chinesischen Sprechern jeweils eine Gruppe bilden oder sich bei chinesischen Sprechern, anders als bei deutschen Muttersprachlern, unter dem Einfluss des Vokals /ɤ/ zwei separate Gruppen ergeben.

6.4.1.2 Begründung

Betrachtet man die Gesamtansicht des Vokalraums, so werden v.a. bei den chinesischen männlichen Sprechern zwei Gruppen von /a/-Lauten sichtbar, von denen eine stark mit den /ø:/-Lauten vermischt ist.

(Zhang, 1996) beschreibt den chinesischen Vokal /ɤ/ als ungerundeten, mittel-hohen, hinteren Vokal mit konsonantischen Eigenschaften. Sie führt an, dass die F_1 -Übergänge innerhalb dieses Vokals mit den Übergängen bei /jɛ/, /yɛ/ und /wo/ verglichen werden können und dass während der Artikulation des Vokals die Zunge gesenkt wird.

Außerdem unterscheiden sich die ersten 25% der untersuchten Vokalabschnitte deutlich von der zweiten Hälfte des Vokals. Im ersten Teil ist die Intensität von F_2 , F_3 und F_4 geringer und die Formantwerte sind für F_1 höher und für F_2 niedriger als in der zweiten Hälfte. Daher beschreibt sie das erste Viertel des Vokalabschnitts als pharyngealen Approximanten, der in einen velaren Approximanten übergeht. Die restlichen drei Viertel bilden den eigentlichen Vokal /ɤ/.

In Versuchen stellt sie fest, dass Versuchspersonen vor vokalischen Silben, die mit /a/ oder /o/ beginnen, eher einen glottalen Verschlusslaut einfügen, als vor solchen mit /ɤ/.

6.4.1.3 Vergleich

Zuerst erstelle ich für männliche und weibliche Sprecher Untermengen, die alle Abschnitte von /a:/ und /a/ enthalten.

Für alle Testmengen führe ich ein modellbasiertes Clustering auf der Basis der Werte für F_1 , F_2 und die Vokaldauer durch und gebe die Ergebnisse als Grafik für F_2 und F_1 aus. Die Werte für F_3 waren zu stark gestreut, so dass sich zu viele Untercluster bildeten, wenn man sie mit einbezog.

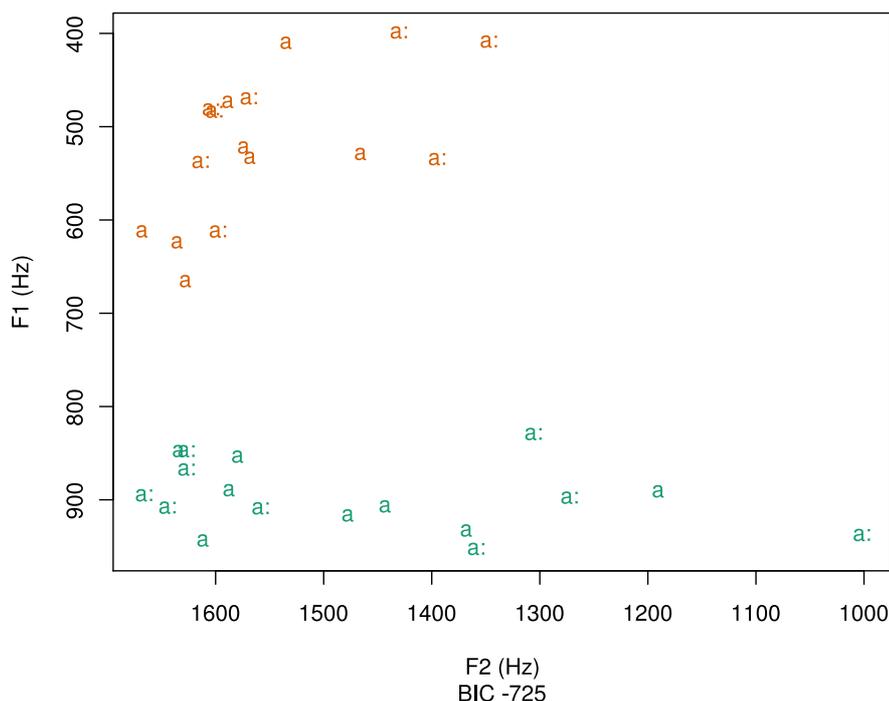
Die Formantwerte von deutschen männlichen Sprechern bilden nur ein Cluster für /a:/ und /a/, obwohl im Kiel-Corpus die durchschnittlichen Werte für /a:/ und /a/ separat angegeben werden.

Für deutsche weibliche Sprecher erhalte ich mehrere Cluster, dies ist jedoch auf sprecher-

spezifische Unterschiede zurückzuführen, unter anderem besonders niedrige Werte von einer Sprecherin.

Die Werte für F_1 und F_2 bei den chinesischen männlichen Sprechern lassen sich dagegen deutlich in zwei Gruppen einteilen:

Abbildung 6.13: Automatische Klassifizierung von /a:/ und /a/ für chinesische männliche Sprecher



In der weiblichen chinesischen Sprechergruppe ist der Unterschied nicht so deutlich, aber immer noch sichtbar.

Als nächstes stellt sich die Frage, wie ich sichergehen kann, dass diese Gruppen nicht durch sprecherspezifische Unterschiede oder verschiedene phonetische Umgebungen entstanden sind.

Bei den chinesischen männlichen Sprechern stammen die Werte der Gruppe mit den niedrigeren F_1 -Frequenzen alle von einem Sprecher (cz) und die Werte der anderen Gruppe mit den höheren F_1 -Frequenzen alle ebenfalls von einem einzigen Sprecher (dl). Die Übereinstimmung der Cluster mit den einzelnen Sprechern beträgt hier 100%.

Untersuche ich dagegen die Werte der vier weiblichen Sprecherinnen, sind Messergebnisse von drei von vier Sprecherinnen in der Gruppe mit den niedrigeren F_1 -Werten enthalten, so dass man eine Sprecherabhängigkeit der Cluster ausschließen kann. Die Übereinstimmung der Cluster mit den einzelnen Sprechern beträgt hier nur 37,8%.

Abbildung 6.14: Automatische Klassifizierung von /a:/ und /a/ für chinesische weibliche Sprecherinnen

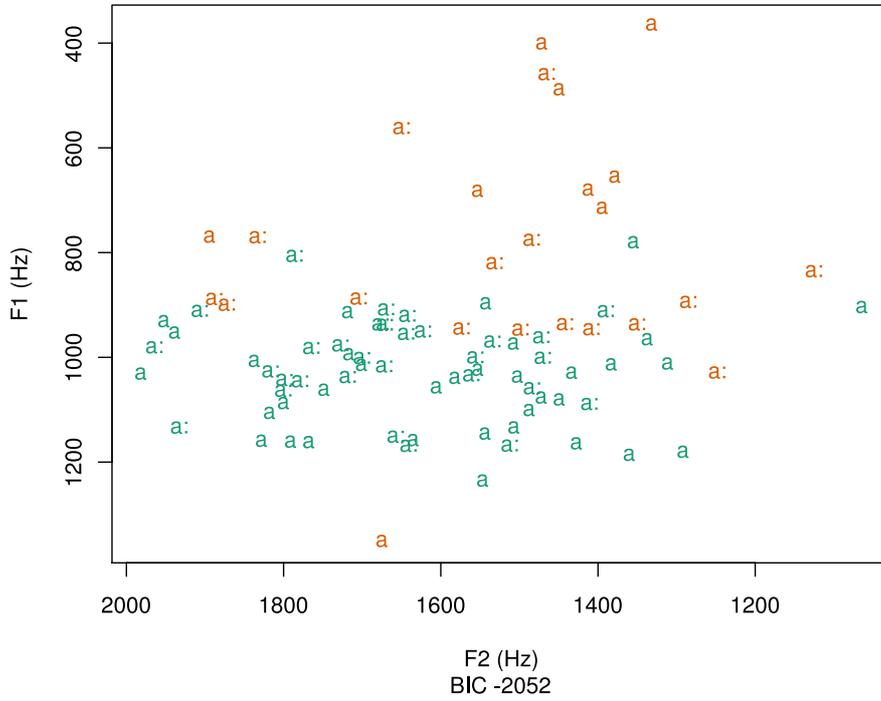
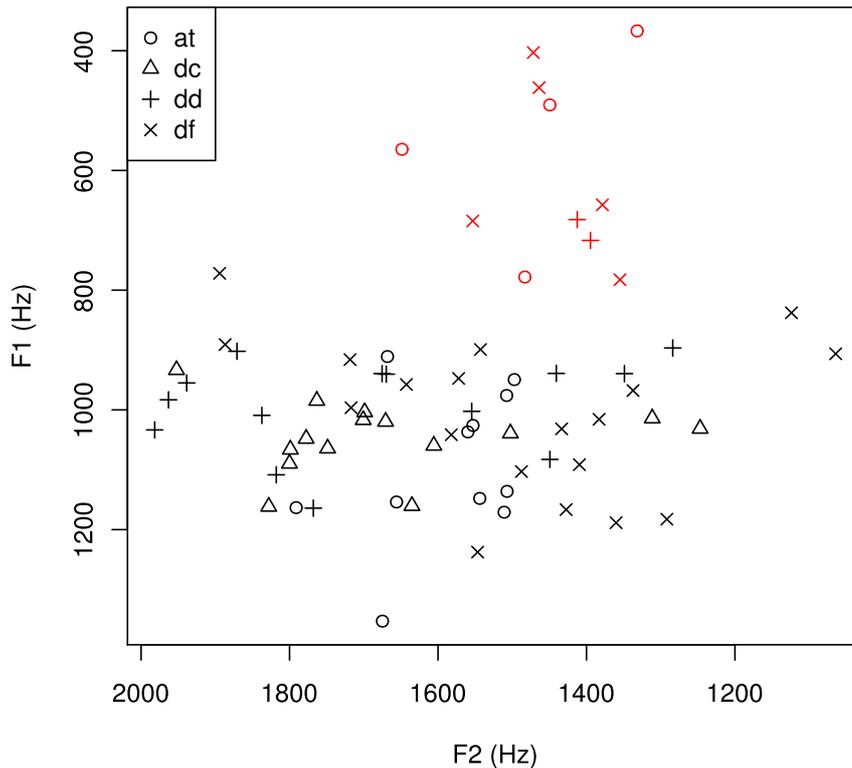


Abbildung 6.15: Messwerte weiblicher chinesischer Sprecherinnen für a-Cluster sortiert nach einzelnen Sprechern



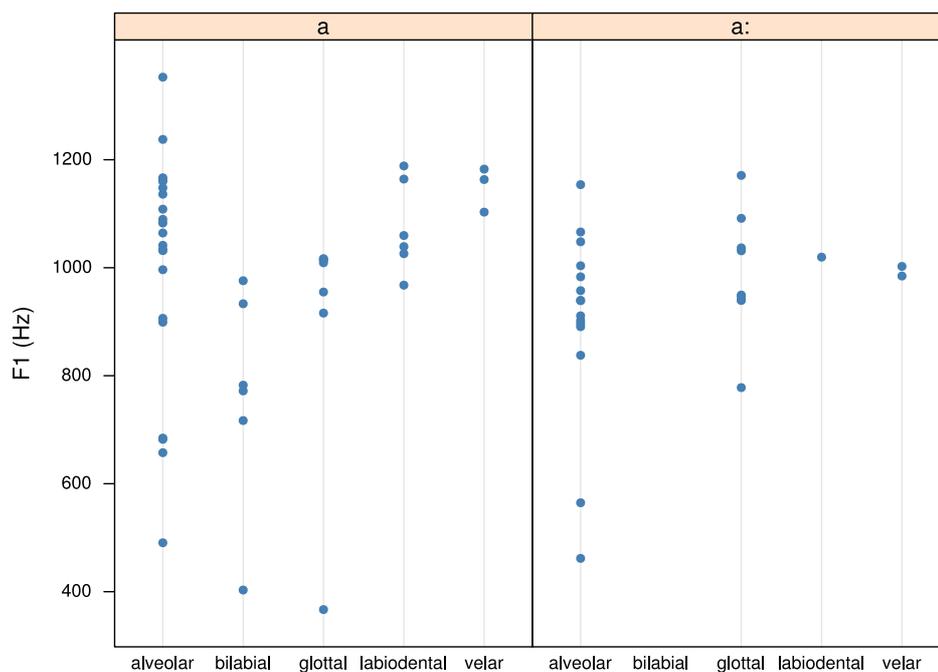
6.4.1.4 Kontexteffekte

Der Artikulationsort des vorherigen bzw. folgenden Konsonanten kann sich auf die Bandbreite der Formantwerte auswirken. So gibt z.B. (Recasens, 1985) für katalanische Vokale unterschiedlich große Formantverteilungen an, z.B. eine große Spannweite von Werten für F_2 vor velaren und labialen Konsonanten und eine geringere Spannweite für F_2 vor palatalen Konsonanten und /w/. Die Vokale vor dentalen und alveolaren Konsonanten liegen dazwischen.

Um den Einfluss der vorherigen und nachfolgenden Konsonanten richtig einzuschätzen, lasse ich die Formanten nach vorherigen bzw. nachfolgenden Artikulationsorten anzeigen.

Da bei den zwei männlichen Sprechern die Auswirkungen der individuellen Unterschiede überwiegen, stelle ich hier nur die Ergebnisse für die weiblichen Sprecher dar.

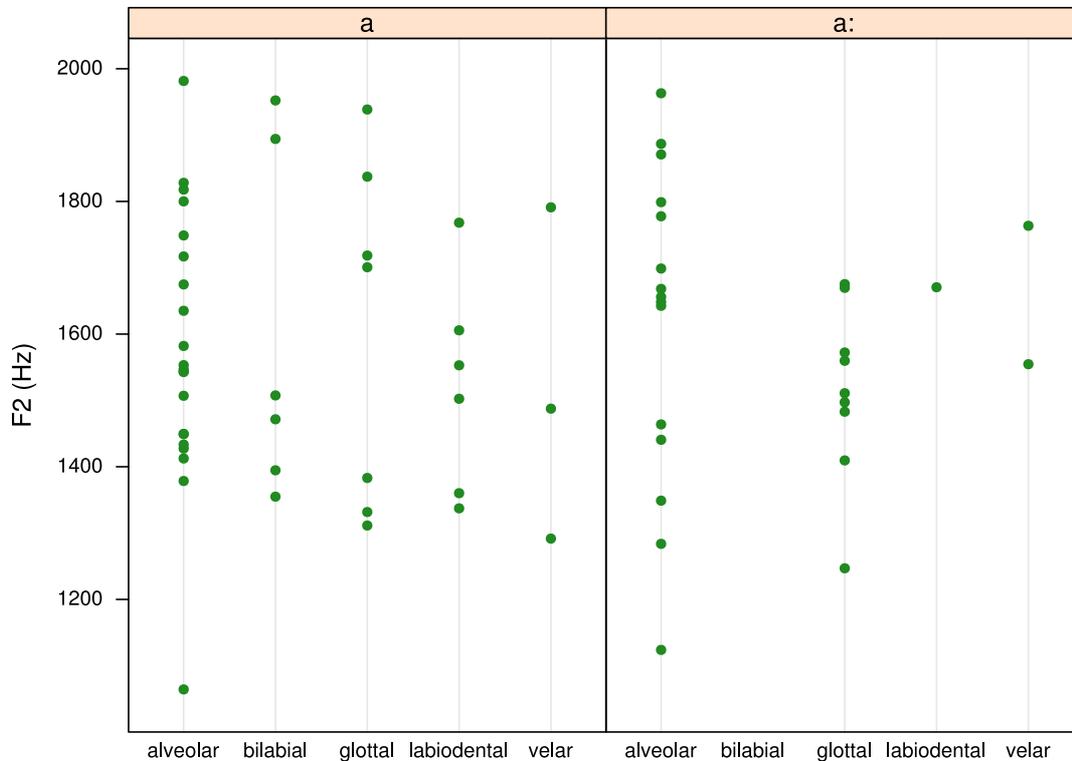
Abbildung 6.16: F_1 -Werte für /a:/ und /a/ bei chinesischen Sprecherinnen nach dem Artikulationsort des vorherigen Konsonanten sortiert



Hier wird deutlich die größere Spannweite der Formantwerte nach alveolaren Konsonanten deutlich, die durch die weitere Zungenbewegung aus der Position oben hinter dem Zahndamm nach unten bedingt ist.

Dies wird auch dadurch bestätigt, dass /a/ mit seiner kürzeren Vokaldauer eine größere Streuung der Werte aufweist als /a:/.

Abbildung 6.17: F_2 -Werte für /a:/ und /a/ bei chinesischen Sprecherinnen nach dem Artikulationsort des vorherigen Konsonanten sortiert



Beim zweiten Vokalformanten ist die Spannweite der Werte ebenfalls nach alveolaren Konsonanten am größten. Dies könnte dadurch bedingt sein, dass sich die Zunge von alveolaren Konsonanten wie /d/ oder /t/ zu /a/ nicht nur nach unten sondern auch nach hinten bewegt.

Da hier deutliche Unterschiede in der Streuung bestehen, jedoch nicht unbedingt in der Formanthöhe, gehe ich davon aus, dass die in den Clustern sichtbaren Untergruppen tatsächlich existieren und nicht nur ein Zufallsprodukt kontextueller Einflüsse sind.

6.4.1.5 Gemischte Modelle für männliche Sprecher

Die errechneten Cluster fallen mit 19 Werten im ersten und 17 Werten im zweiten Cluster ähnlich groß aus, so dass sich eine Auswahl von Werten nach dem Zufallsprinzip erübrigt.

6.4.1.5.1 Eigenschaften des gemischten Modells Die BIC-Werte fallen hier mit 378 bis 436 Punkten für die ersten drei Formanten höher aus als für das gemischte Modell bei den weiblichen Sprechern. Der einzelne Zufallsfaktor erklärt für die ersten drei Formanten jeweils 28,4% der gemessenen Abweichungen.

Tabelle 6.5: Ergebnisse eines gemischten Modells für a-Cluster bei männlichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	8.792	0.0302	0
F_2	-0.70	0.6032	0.4889
F_3	-1.94	0.3082	0.0618
Vokaldauer	-0.387	0.7458	0.7015

In diesem Fall unterscheiden sich die Werte für F_1 signifikant und die Werte für F_3 liegen knapp über der Signifikanzgrenze.

Tabelle 6.6: Durchschnittswerte für das gemischte Modell der a-Cluster bei männlichen Sprechern

Parameter	Cluster 1 /ɤ/	Cluster 2 /a/
F_1 (Hz)	518	897
F_2 (Hz)	1538	1467
F_3 (Hz)	2677	2471
Vokaldauer (ms)	99,53	91,00

Vergleiche ich die Durchschnittswerte, so fallen vor allem der niedrige F_1 -Wert des ersten Clusters und der niedrigere F_3 -Wert des zweiten Clusters auf. Resynthetisiert man einen Vokal mit den Werten des ersten Clusters, so liegt dieser vom Höreindruck näher bei /ə/ und von den Frequenzen her zwischen /ɛ/ und /ə/. Der Klang des zweiten Clusters entspricht einem deutschen muttersprachlichen /a:/.

6.4.1.6 Gemischte Modelle für weibliche Sprecher

6.4.1.6.1 Eigenschaften des gemischten Modells Die BIC-Werte des gemischten Modells fallen hier mit 272 bis 318 Punkten niedriger aus als in der männlichen Gruppe. Der einzelne Sprecher als Zufallsfaktor besitzt für F_1 , F_2 und die Vokaldauer kein Gewicht, für F_3 erklärt er 25,9% der Abweichungen.

Tabelle 6.7: Ergebnisse eines gemischten Modells für a-Cluster bei weiblichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	6.863	0.0001	0
F_2	1.32	0.2144	0.2011
F_3	0.204	0.8396	0.8401
Vokaldauer	1.228	0.2468	0.2338

Wie bei den männlichen Sprechern unterscheidet sich F_1 signifikant, der Wert für F_3 dagegen liegt weit über der Signifikanzgrenze.

Tabelle 6.8: Durchschnittswerte für das gemischte Modell der a-Cluster bei weiblichen Sprechern

Parameter	Cluster 1 /ɤ/	Cluster 2 /a/
F_1 (Hz)	599	1018
F_2 (Hz)	1449	1529
F_3 (Hz)	2544	2582
Vokaldauer (ms)	82,09	96,82

In diesem Fall tendiert der Vokal des ersten Clusters vom Klang her wieder gegen /ə/, der des zweiten Clusters zu einem /a/, das heller ausfällt, als der entsprechende deutsche Vokal.

Im Gegensatz zu den männlichen Sprechern zeichnet sich der erste Cluster zusätzlich durch eine hörbar kürzere Vokaldauer aus.

Der Vergleich der Cluster in der weiblichen Sprechergruppe, die aus vier Sprecherinnen besteht, von denen drei in beiden Clustern mit Abschnitten vertreten sind, belegt, dass die Unterschiede für F_1 mehr als nur sprecherspezifische Abweichungen darstellen. In beiden Clustern treten ähnliche dem Vokal vorausgehenden bzw. nachfolgenden Konsonanten auf. Da der erste Cluster vier betonte und sieben unbetonte, der zweite sieben betonte und vier unbetonte Werte enthält, errechne ich jeweils die Durchschnittswerte. Diese unterscheiden sich mit 544 Hz für betonte Vokale und 630 Hz für unbetonte im ersten Cluster sowie 1053 Hz für betonte Vokale und 958 Hz für unbetonte im zweiten Cluster trotzdem noch deutlich voneinander.

6.5 Gespannte und ungespannte Vokale

Untersuchungsgegenstand 7 (Gespannte und ungespannte Vokale) *Da die Aussprache der in beiden Sprachen bekannten Vokale /i/, /u/, /ε/ und /o/ eng mit der Aussprache der nur im Deutschen vorhandenen Vokale /ɪ/, /ʊ/, /e/ und /ɔ/ verbunden ist, werden die Fragen, inwieweit sich die Aussprache bereits bekannter Vokale unterscheidet und inwieweit gespannte und ungespannte Vokale getrennt werden hier in einem Unterkapitel bearbeitet.*

Insgesamt untersuche ich, wie sehr gespannte und ungespannte Vokale von deutschen und chinesischen Sprechern unterschieden werden.

6.5.0.7 Begründung

Laut (Lin, 2007, S. 252) fallen die Laute /i/ und /ɪ/ bei chinesischen Sprechern zu /i/, die Laute /u/ und /ʊ/ zu /u/ zusammen. (Liao, 2006, S. vii) stellt fest, dass seine Testpersonen, die Mandarin als Erst- und Englisch als Zweitsprache sprechen, nicht zwischen /u/ und /ʊ/ bzw. /o/ und /ɔ/ unterscheiden. (Chen, 2006) vergleicht die Aussprache der Gegensatzpaare /i/ und /ɪ/, /æ/ und /ε/ sowie /u/ und /ʊ/ im amerikanischen Englisch durch chinesische Muttersprachler.

Dabei unterscheidet sich die Vokaldauer für /æ/ und /ε/ sowie für /u/ und /ʊ/ bei chinesischen Muttersprachlern nicht wesentlich – für /i/ und /ɪ/ können dagegen deutliche Unterschiede, ähnlich wie bei amerikanischen Muttersprachlern festgestellt werden. Der Abstand zwischen den F_1 - und F_2 -Werten ist bei allen drei Vokalpaaren signifikant geringer als bei amerikanischen Muttersprachlern. Dabei liegen die Formantwerte für /u/ und /ʊ/ am nächsten.

Werden amerikanischen Muttersprachlern aus den Sprachaufnahmen isolierte Vokalabschnitte vorgespielt, so unterscheiden sich für die Zuhörer 52% der /i/ und /ɪ/-Paare, jedoch nur 32% der /u/ und /ʊ/-Paare. Bei von amerikanischen Muttersprachlern produzierten Vokalen werden 94% bzw. 85% der Vokalpaare als unterschiedlich beurteilt.

(Gut, 2006) untersucht Unterschiede in der Aussprache deutscher betonter und unbetonter Silben bei Nicht-Muttersprachlern des Deutschen, in diesem Fall britische, italienische und chinesische Muttersprachler. Signifikante Abweichungen in der Aussprache unbetonter Vokale zeigten sich bei den chinesischen Sprechern vor allem bei der Vokaldauer, bei der Nachsilbe *-en* für F_1 und F_2 , und bei der Vorsilbe *ver-* für F_1 .

6.5.1 Unterschied zwischen /i:/ und /ɪ/

6.5.1.1 Vorgehensweise

Zuerst werden alle Messwerte für /i:/ und /ɪ/ gesammelt und nach Geschlecht und Muttersprache getrennt. Dann werden für /i:/ und /ɪ/ jeweils gleich große Testmengen erstellt. Zuerst wird die Verteilung dieser Mengen untersucht, dann werden sie statistisch verglichen.

Im Hinblick auf die deutschen Muttersprachler vergleiche ich 44 Werte pro Vokal in der männlichen und 31 Werte in der weiblichen Kontrollgruppe. Diese werden zufällig aus den 45 Werten für /ɪ/ und 70 für /i:/ bei männlichen Sprechern sowie den 32 Werten für /ɪ/ und 59 Werten für /i:/ bei weiblichen Sprechern ausgewählt.

Wie erwartet unterscheiden sich die Messwerte für /i:/ und /ɪ/ bei deutschen männlichen und weiblichen Sprechern signifikant:

Tabelle 6.9: Ergebnisse des Wilcoxon-Rangsummentests für den Unterschied zwischen /i:/ und /ɪ/ bei deutschen männlichen und weiblichen Sprechern

Parameter	p-Wert (Männer)	p-Wert (Frauen)
F_1	$1.17219 * 10^{-7}$	0.03866
F_2	$2.59188 * 10^{-8}$	0.00481
F_3	$3.89481 * 10^{-6}$	0.02744
Vokaldauer	0.01162	0.835

Die chinesischen Sprechergruppen umfassen 28 Werte pro Vokal für die männlichen Sprecher, die aus 37 Werten für /i:/ und 28 für /ɪ/ ausgewählt werden. Die Testmengen für die weiblichen Sprecher bestehen aus 47 Werten pro Vokal, ausgewählt aus 91 Werten für /i:/ und 51 Werten für /ɪ/.

Bei chinesischen Sprechern erhalte ich folgende Ergebnisse:

Tabelle 6.10: Ergebnisse des Wilcoxon-Rangsummentests für den Unterschied zwischen /i:/ und /ɪ/ bei chinesischen männlichen und weiblichen Sprechern

Parameter	p-Wert (Männer)	p-Wert (Frauen)
F_1	0.2782	0.6831
F_2	0.597	0.00352
F_3	0.01413	0.00526
Vokaldauer	0.00559	0.00010

Da die Daten für chinesische Sprecherinnen für F_1 und F_2 nicht normalverteilt sind und für F_1 und F_3 keine gleichen Varianzen besitzen, wende ich auf jede der vier untersuchten Mengen Klassifikationsbäume an und gebe das Ergebnis als partimat-Grafik aus.

Abbildung 6.18: Verteilung von /i:/ und /ɪ/ für deutsche männliche Sprecher

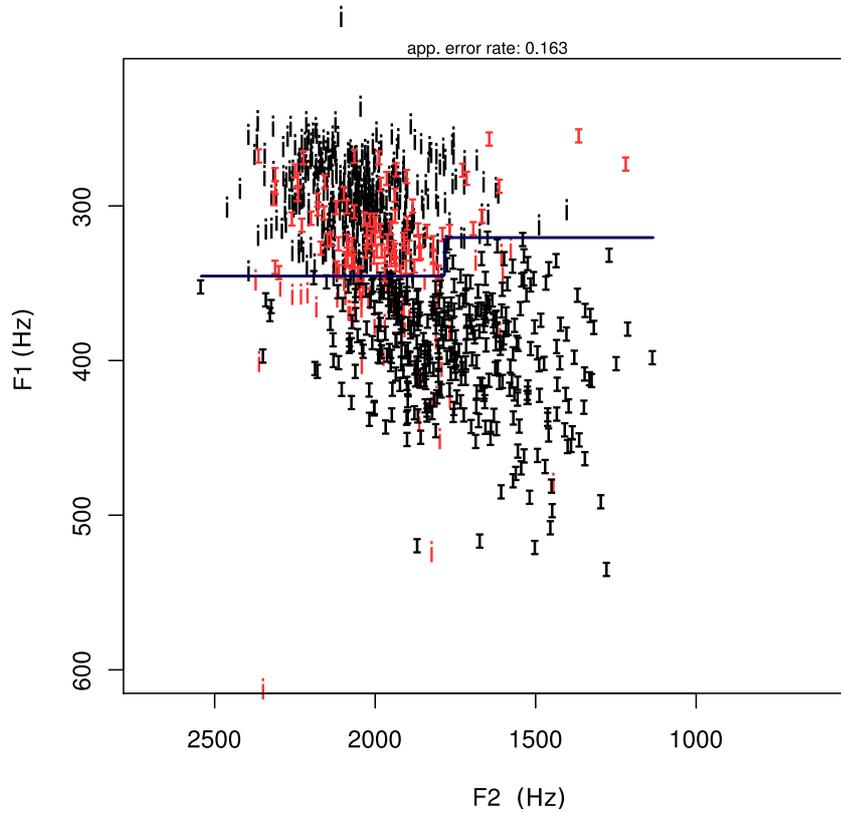


Abbildung 6.19: Verteilung von /i:/ und /ɪ/ für deutsche weibliche Sprecher

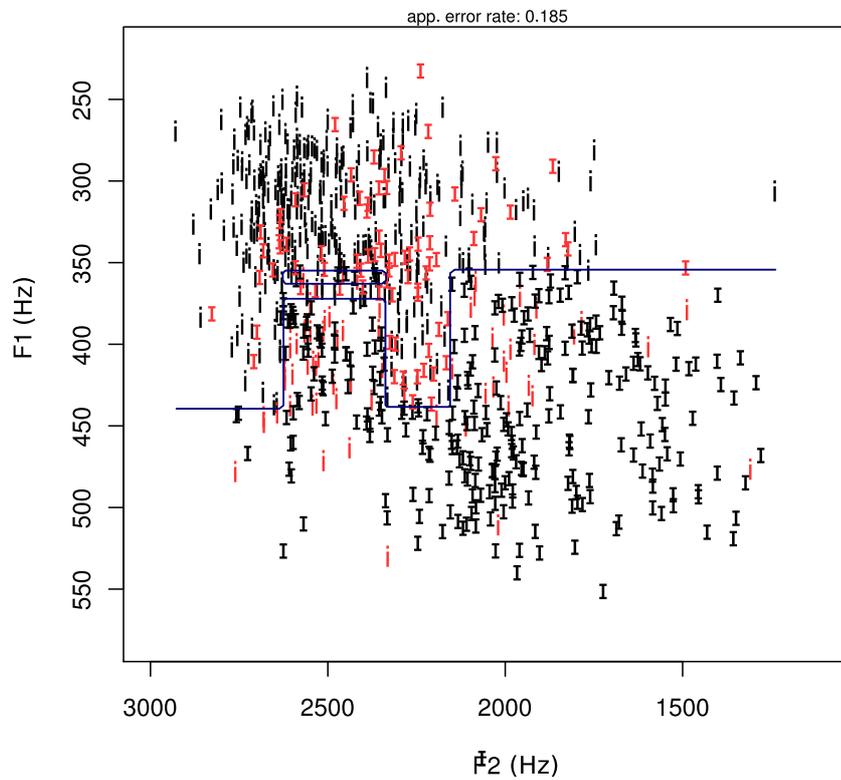


Abbildung 6.20: Verteilung von /i:/ und /ɪ/ für chinesische männliche Sprecher

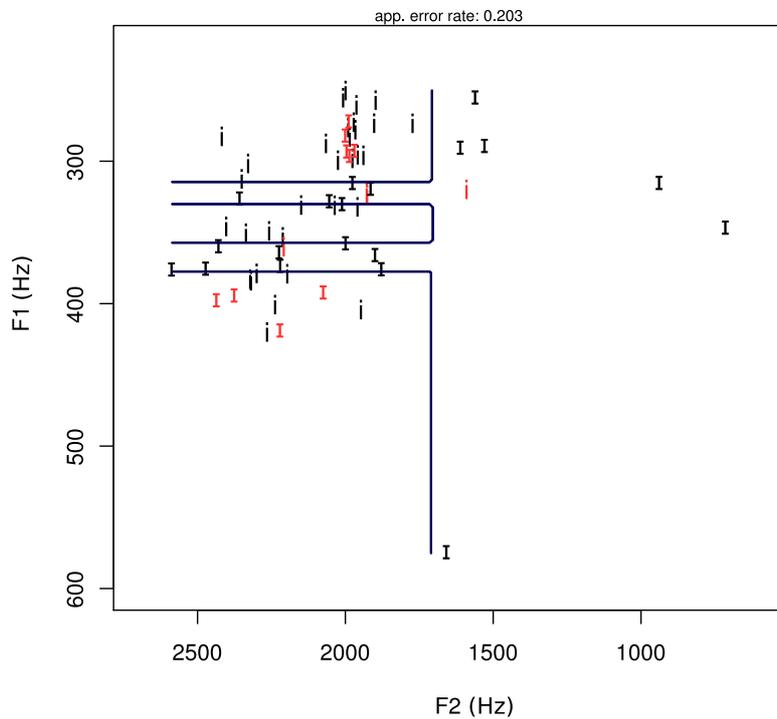
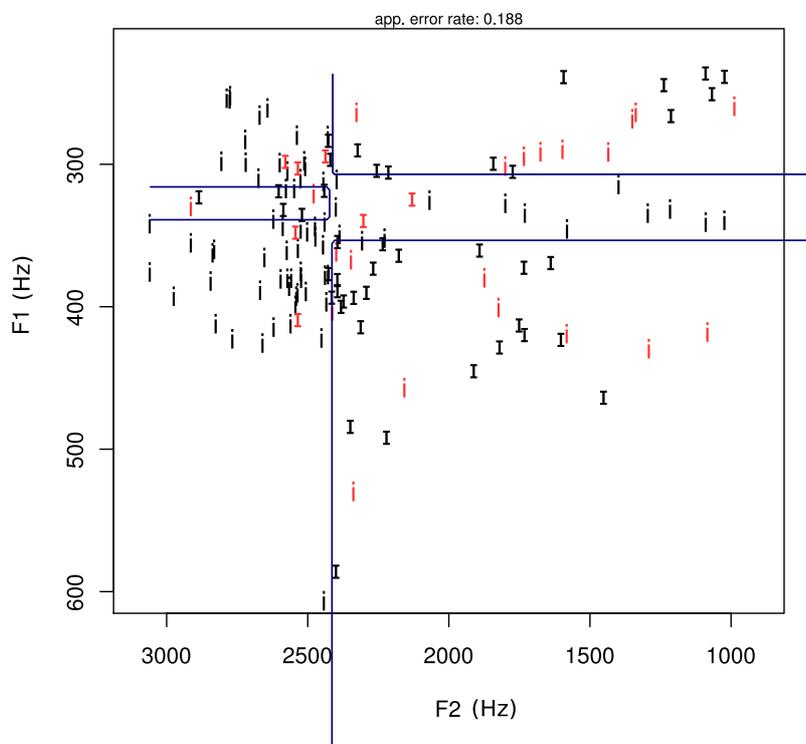


Abbildung 6.21: Verteilung von /i:/ und /ɪ/ für chinesische weibliche Sprecher



Die Formantwerte der deutschen männlichen Sprecher lassen sich bei einer Fehlerrate von 16,3% besser trennen als die der chinesischen männlichen Sprecher mit einer Fehlerrate von 20,3%. Da die Formantwerte der weiblichen Sprecherinnen in beiden Sprachen stark gestreut sind, liegen die Fehlerraten mit 18,5% bzw. 18,6% auf dem gleichen Niveau.

Bei den chinesischen männlichen Sprechern fallen die Formantwerte unter 1700 Hz sehr spärlich aus. Die Formantwerte der chinesischen Sprecherinnen für /i:/ verteilen sich unter anderem auf einen Block zwischen 300 und 350 Hz für F_1 und 1203 bis 1797 Hz für F_2 . Vergleiche ich die vorhergehenden Konsonanten miteinander, finde ich kein /ʃ/ oder /s/, das für eine Übertragung der Formantwerte der muttersprachlichen Vokale /ɪ/ oder /ɪ̥/ sprechen würde. Stattdessen tauchen in 50% der Fälle Pausen, Kehlkopfverschlüsse am Wortanfang oder der Glide /j/ auf, ansonsten die vorderen Plosive /d/ und /b/. Damit liegt eher der Schluss nahe, dass die Diphthongierung von /i/ zu /ji/ am Wortanfang im Chinesischen die Aussprache des deutschen /i:/ in derselben Position beeinflussen könnte.

Mit Ausnahme eines Bereichs, in dem F_1 zwischen 300 und 400 Hz liegt, gilt für chinesische Sprecher die Tendenz, /i:/ vor allem mit einer F_2 über 1700 Hz bei männlichen und über 2400 Hz bei weiblichen Sprechern zu artikulieren, /ɪ/ dagegen mit F_2 -Messwerten unter dieser Grenze. Damit würde /i:/ durch eine extremere Zungenbewegung stärker betont. Dies bestätigt sich jedoch nur für weibliche Sprecherinnen, bei männlichen Sprechern werden die hohen F_2 -Werte für /i:/ durch eine koartikulatorisch bedingte Streuung der unbetonten Abschnitte hervorgerufen.

6.5.2 Vergleich der Werte für /i:/, männliche Sprecher

Hier vergleiche ich jeweils 37 Abschnitte pro Sprachgruppe, wobei ich in der deutschen Gruppe 37 aus 70 vorhandenen Werten zufällig auswähle lasse.

Alle vier Datenmengen sind normalverteilt und besitzen ähnliche Varianzen.

6.5.2.1 Eigenschaften des gemischten Modells

In diesem Fall ist die Anpassung des gemischten Modells an die gemessenen Daten mit BIC-Werten zwischen 781 und 1019 Punkten sehr ungenau. Der einzelne Sprecher als Zufallsfaktor deckt für F_2 46,3% der Abweichungen ab und für F_3 32,9%, hat jedoch im Hinblick auf die Vokaldauer keine Aussagekraft.

Das Einbeziehen des einzelnen Sprechers als Zufallsfaktor hat hier die Ergebnisse im Vergleich zum Rangsummentest nicht verändert.

Die höheren F_3 -Werte in der chinesischen Gruppe könnten durch die hohen charakteristischen F_3 -Werte für hochchinesische Vokale bedingt sein.

Die für /i:/ sehr kurze Vokaldauer in der deutschen Gruppe wird auch durch die große Anzahl an unbetonten Silben ausgelöst. So enthalten die Messwerte der deutschen Gruppe sieben betonte und dreißig unbetonte Abschnitte, die der chinesischen Gruppe zehn betonte und dreißig unbetonte.

Tabelle 6.11: Ergebnisse für das gemischte Modell für /i:/ bei männlichen deutschen und chinesischen Sprechern

Faktor	t-Wert	pMCMC	p-Wert
F_1	-0.063	0.9504	0.9496
F_2	-0.833	0.3252	0.4078
F_3	-2.90	0.0224	0.005
Vokaldauer	-2.811	0.0522	0.0064

Tabelle 6.12: Durchschnittswerte des gemischten Modells für /i:/ bei deutschen und chinesischen männlichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	320	321
F_2 (Hz)	1996	2086
F_3 (Hz)	2653	2980
Vokaldauer (ms)	69,18	97,30

Errechne ich die durchschnittliche Vokallänge für betonte und unbetonte Vokale in beiden Gruppen, so wird der Unterschied zwischen deutschen und chinesischen Sprechern vor allem bei den unbetonten Vokalen noch deutlicher. Hier beträgt die durchschnittliche Vokaldauer der deutschen Sprecher 67 ms, die der chinesischen Sprecher 94,4 ms. Die betonten Vokale werden von den deutschen Muttersprachlern mit durchschnittlich 78,57 ms, von den chinesischen Deutschlernenden mit 105 ms artikuliert.

6.5.3 Vergleich der Werte für /i:/, weibliche Sprecher

Für den statistischen Vergleich setze ich jeweils 59 Abschnitte aus beiden Sprachgruppen ein, wobei ich in der chinesischen Gruppe 59 von 91 Abschnitten auswählen lasse.

Die hier verglichenen Messwerte sind für F_1 , F_2 und die Vokaldauer weder normalverteilt noch besitzen sie gleiche Varianzen.

6.5.3.1 Eigenschaften des gemischten Modells

Die Anpassung des gemischten Modells an die Messdaten fällt hier mit BIC-Werten von 1408 bis 1769 deutlich schlechter aus als in der männlichen Gruppe. Der Anteil des einzelnen Sprechers als Zufallsfaktor an den errechneten Abweichungen liegt zwischen 43,3% für F_1 und 26,9% für die Vokaldauer.

Tabelle 6.13: Ergebnisse für das gemischte Modell für /i:/ bei weiblichen deutschen und chinesischen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	-0.435	0.5932	0.6645
F_2	0.352	0.6660	0.7252
F_3	-1.77	0.0994	0.0799
Vokaldauer	-2.38	0.0430	0.0189

Im Vergleich zur männlichen Gruppe nimmt der Unterschied für F_3 keinen signifikanten Wert mehr an.

Tabelle 6.14: Durchschnittswerte des gemischten Modells für /i:/ bei deutschen und chinesischen weiblichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	368	391
F_2 (Hz)	2275	2191
F_3 (Hz)	2889	3113
Vokaldauer (ms)	66,96	95,38

Wie bei den männlichen Sprechern fällt die wesentlich kürzere Vokaldauer in der deutschen Gruppe auf. Beide Gruppen enthalten jeweils 13 betonte und 46 unbetonte Abschnitte. Errechne ich die Mittelwerte für die Vokaldauer in betonten und in unbetonten Silben, so erhalte ich für unbetonte Silben 60,43 ms bei deutschen und 87,83 bei chinesischen Sprecherinnen; für betonte Silben sind es 90 ms in der deutschen und 116,15 ms in der chinesischen Gruppe.

Der niedrigere F_2 -Wert in der chinesischen Gruppe macht sich bei der Resynthese der Formantwerte bemerkbar und trägt zu einem Klang zwischen /i:/ und /ɪ/ bei.

6.5.4 Vergleich der Werte für /ɪ/, männliche Sprecher

In diesem Fall vergleiche ich 28 Messwerte pro Sprachgruppe, davon 28 von 45 zufällig gezogene Werte in der deutschen Gruppe.

Die untersuchten Messwerte für /ɪ/ sind alle normalverteilt, die Wertemengen für F_2 und für die Vokaldauer besitzen jedoch bei männlichen Sprechern unterschiedliche Varianzen.

6.5.4.1 Eigenschaften des gemischten Modells für /ɪ/ bei männlichen Sprechern

Die BIC-Werte, die beschreiben, wie gut das gemischte Modell an die Messdaten angepasst ist, fallen hier mit 611 bis 800 Punkten niedriger aus, als im gemischten Modell der männlichen Gruppen für /i:/.

Der einzelne Sprecher als Zufallsfaktor deckt 32,7% der Abweichungen für F_1 ab, jedoch nur 19% für F_2 und die Vokaldauer.

Tabelle 6.15: Ergebnisse für ein gemischtes Modell für /ɪ/ bei männlichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	0.702	0.4432	0.4856
F_2	-1.717	0.1334	0.0917
F_3	-4.25	0.0034	0.0001
Vokaldauer	-2.714	0.0344	0.0089

Wie für /i:/ in der männlichen Gruppe lassen sich signifikante Unterschiede für F_3 und die Vokaldauer feststellen.

Tabelle 6.16: Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und chinesischen männlichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	365	346
F_2 (Hz)	1781	1965
F_3 (Hz)	2446	2826
Vokaldauer (ms)	50,81	71,43

Die chinesischen Formantwerte fallen hier hörbar extremer aus, wobei F_1 und F_2 sich den Werten für ein muttersprachliches deutsches /i:/ annähern, sich jedoch immer noch vom gesprochenen /i:/ unterscheiden.

6.5.5 Vergleich der Werte für /ɪ/, weibliche Sprecher

Für beide Sprechergruppen sind die Wertemengen für alle drei Formanten normalverteilt. Die Messwerte für die Vokaldauer sind nicht normalverteilt und sowohl die Werte für F_2 als auch die für die Vokaldauer besitzen unterschiedliche Varianzen.

6.5.5.1 Eigenschaften des gemischten Modells

Die BIC-Werte fallen hier mit 719 bis 943 Punkten höher aus als in dem Modell für /ɪ/ bei männlichen Sprecher.

Das Gewicht des einzelnen Sprechers beträgt für F_1 37,9%, für F_3 jedoch nur 17,4%. Damit wirkt sich der Zufallsfaktor weniger intensiv aus als für /i:/ bei weiblichen Sprechern.

Tabelle 6.17: Ergebnisse für ein gemischtes Modell für /ɪ/ bei weiblichen Sprechern

Parameter	t-Wert	pMCMC	p-Wert
F_1	1.963	0.0508	0.0542
F_2	-0.358	0.6404	0.7219
F_3	-3.85	0.0062	0.0003
Vokaldauer	-0.603	0.5020	0.5485

Die Vokaldauer ist im Gegensatz zu den männlichen Sprechern nicht mehr signifikant, wobei die größere Streuung der Werte durch eine größere Anzahl an Sprecherinnen eine Rolle gespielt haben könnte. Dafür liegen die Unterschiede für F_1 nur gering über der Signifikanzgrenze, was durch den im Vergleich von /i:/ und /ɪ/ schon sichtbaren Block mit Abschnitten nach /j/ und am Wortanfang bedingt sein könnte.

Tabelle 6.18: Durchschnittswerte des gemischten Modells für /ɪ/ bei deutschen und chinesischen weiblichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	410	345
F_2 (Hz)	2007	2059
F_3 (Hz)	2684	2979
Vokaldauer (ms)	62,33	68,53

Das /ɪ/ der chinesischen Sprecherinnen klingt wieder heller und befindet sich mit seinen Formantwerten genau in der Mitte zwischen dem muttersprachlichen /i:/ und /ɪ/, wird jedoch immer noch anders artikuliert als /i:/.

6.5.6 Unterschied zwischen /u:/ und /ʊ/

6.5.6.1 Statistischer Vergleich

Die jeweils 11 Messwerte für deutsche männliche Sprecher, davon 11 zufällig aus 14 ausgewählte Werte für /u:/, sind für F_1 und F_2 nicht normalverteilt und besitzen unterschiedliche Varianzen.

Die jeweils 8 Werte für deutsche weibliche Sprecher, davon 8 aus 14 für /ʊ/, besitzen für F_1 unterschiedliche Varianzen.

Die jeweils fünf Messwerte für chinesische männliche Sprecher, davon 5 aus 12 für /ʊ/, sind für F_1 nicht normalverteilt und besitzen für F_1 und F_3 unterschiedliche Varianzen.

Die jeweils 13 Werte für chinesische weibliche Sprecherinnen, davon 13 aus 30 für /ʊ/, sind für F_1 und F_2 nicht normalverteilt und besitzen unterschiedliche Varianzen.

Bei einer Prüfung mit dem Wilcoxon-Rangsummentest ergeben sich für deutsche Sprecher signifikante Unterschiede zwischen /u:/ und /ʊ/ bei F_1 , für chinesische Sprecher jedoch nicht.

Tabelle 6.19: Unterschiede zwischen /u:/ und /ʊ/ für deutsche Sprecher

Parameter	p-Wert (Männer)	p-Wert (Frauen)
F_1	$7.396 * 10^{-7}$	0.03998
F_2	0.7553	0.4363
F_3	0.8428	0.8633
Vokaldauer	0.4023	0.2694

Tabelle 6.20: Unterschiede zwischen /u:/ und /ʊ/ für chinesische Sprecher

Parameter	p-Wert (Männer)	p-Wert (Frauen)
F_1	0.5476	0.2930
F_2	0.4206	0.0686
F_3	0.6905	0.7388
Vokaldauer	0.3095	0.6813

6.5.6.2 Klassifikationsbäume

In diesem Fall weist der Klassifikationsbaum für männliche Sprecher eine deutliche Tendenz zu niedrigeren Werten für F_1 und höheren Werten für F_2 und F_3 bei /u:/ auf. Damit wird der gespannte Vokal /u:/ wie im Deutschen extremer artikuliert als der ungespannte Vokal /ʊ/.

Abbildung 6.22: Klassifikationsbaum für /u:/ und /ʊ/ für männliche chinesische Sprecher

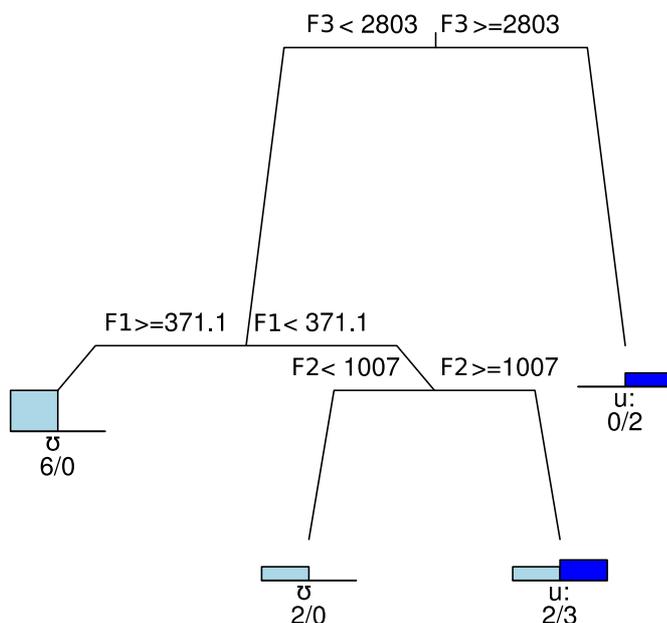
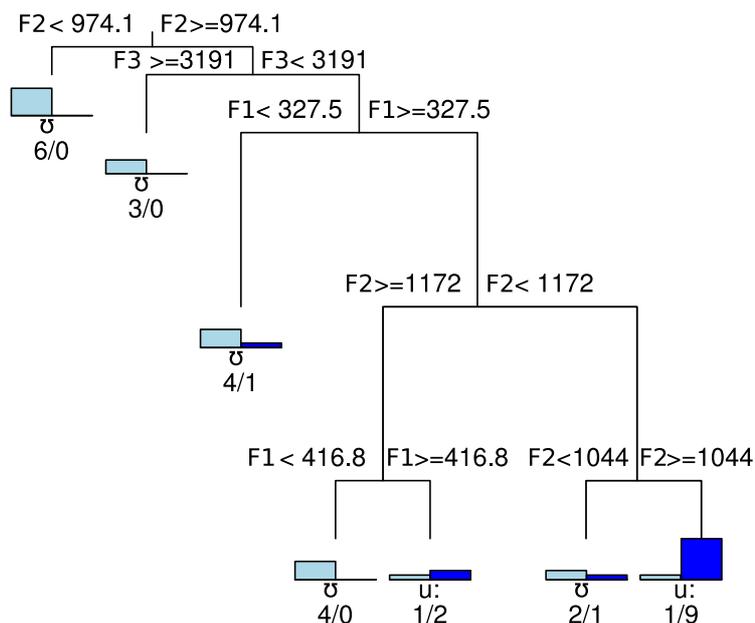


Abbildung 6.23: Klassifikationsbaum für /u:/ und /ʊ/ für weibliche chinesische Sprecher



Für die chinesischen männlichen Sprecher liegt der F_1 -Wert für /u:/ leicht über dem deutscher Muttersprachler, der F_2 -Wert dagegen sogar über dem für /ʊ/, so dass die Formantwerte dem Vokal /ɤ/ am nächsten kommen. Dass sich der Vokal in der Resynthese trotzdem eher zwischen /ʊ/ und /ɤ/ anhört, kann durch den extrem hohen F_3 -Wert erklärt werden.

Die kürzere Vokaldauer für /u:/ bei männlichen chinesischen Sprechern ist nicht durch die Silbenbetonung bedingt, da in dieser Testmenge vier von fünf Abschnitten betont sind. Die Wertemenge der weiblichen chinesischen Sprecher enthält jeweils acht betonte und fünf unbetonte Silben.

Die F_1 - und F_2 -Werte der männlichen chinesischen Studenten für /u:/ dagegen liegen nur

Tabelle 6.21: Mittelwerte für /u:/ und /ʊ/ bei männlichen und weiblichen chinesischen Sprechern

Parameter	männliche Sprecher		weibliche Sprecher	
	/u:/	/ʊ/	/u:/	/ʊ/
F_1 (Hz)	356	444	387	368
F_2 (Hz)	1279	1089	1189	1025
F_3 (Hz)	2776	2460	2697	2707
Vokaldauer (ms)	57,6	73,0	70,85	81,23

leicht höher als bei deutschen Sprechern und immer noch im Bereich des entsprechenden muttersprachlichen Vokals.

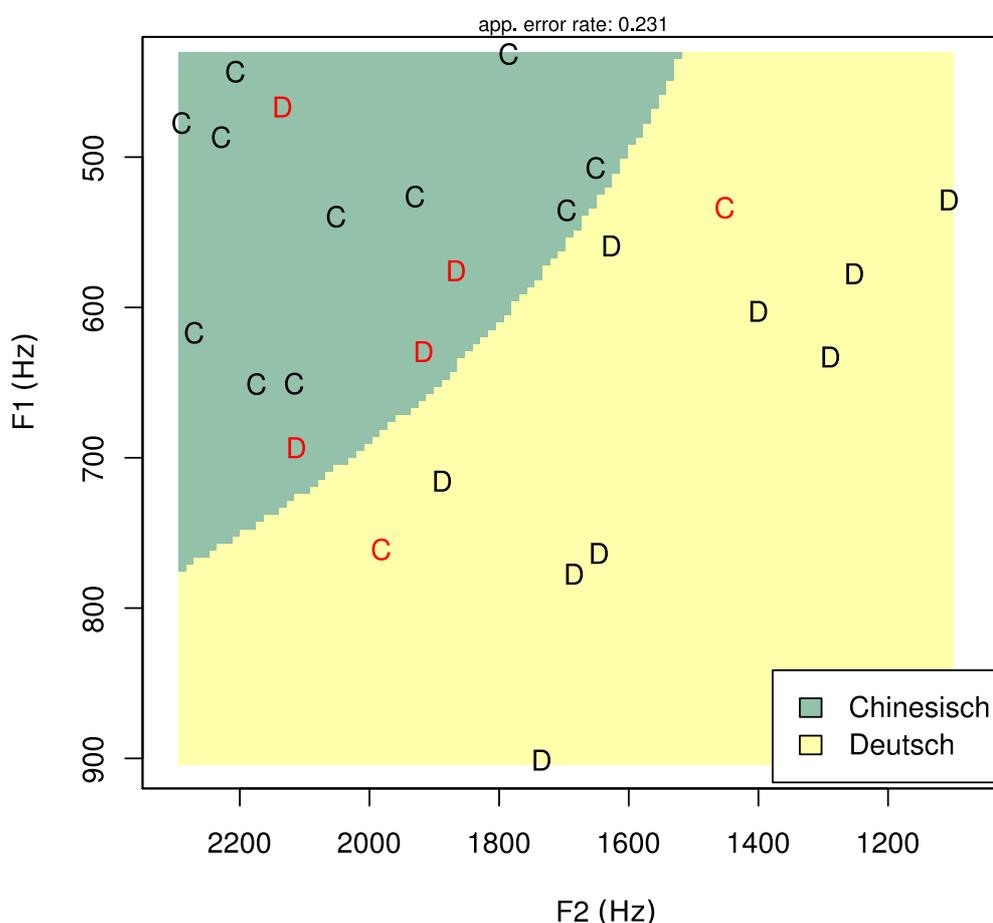
Das von weiblichen Sprecherinnen artikulierte /u:/ besitzt leicht höhere F_1 - und deutlich höhere F_2 -Frequenzen. Es liegt im Klang wie das der männlichen Sprecher zwischen /ʊ/ und /ɤ/. Der Vokal /ʊ/ hingegen wird mit seinem niedrigen F_1 -Wert zwischen einem deutschen /u:/ und einem deutschen /ʊ/ artikuliert.

6.5.7 Testmengen für / ϵ /, weibliche Sprecher

Da für die männlichen Sprecher jeweils nur vier Testwerte zur Verfügung standen, werden hier nur die Testmengen für die weiblichen Sprecher untersucht. Dazu wähle ich 11 Werte pro Sprache aus, von denen 11 Werte aus den 16 Werten der chinesischen Gruppe zufällig ausgewählt werden.

Um herauszufinden, worin sich F_1 und F_2 für deutsche und chinesische Sprecher unterscheiden, führe ich eine Diskriminanzanalyse durch.

Abbildung 6.24: Verteilung der Formantwerte bei / ϵ / für weibliche chinesische Sprecherinnen



Die Vokale chinesischer Sprecher weisen höhere Werte für F_2 und niedrigere Werte für F_1 auf.

Sehe ich mir die Worte an, aus denen die untersuchten Abschnitte entnommen sind, so kann ich keine auffällige Häufung bestimmter Artikulationsarten oder -orte entdecken.

6.5.7.1 Eigenschaften des gemischten Modells

In diesem Fall sind die BIC-Werte um nur 4 Punkte höher als die AIC-Werte, was für eine ausreichende Komplexität des Modells spricht. Die Unterschiede zwischen den BIC-Werten für die einzelnen Formantwerte von 271 bis 301 Punkten sind niedrig, was auf die

geringere Streuung bei kleineren Testwertemengen zurückzuführen sein könnte.

Für F_1 ist der Anteil der durch den Zufallsfaktor erklärten Daten gleich 0, für F_2 dagegen liegt er bei 50,5%.

Tabelle 6.22: Ergebnisse für ein gemischtes Modell für / ϵ / bei weiblichen Sprechern mit dem einzelnen Sprecher als Zufallsfaktor

Parameter	t-Wert	pMCMC	p-Wert
F_1	2.027	0.1142	0.0556
F_2	-1.91	0.0178	0.0699
F_3	-0.129	0.7418	0.8985
Vokaldauer	-2.783	0.0114	0.0111

In diesem gemischten Modell unterscheidet sich nur die Vokaldauer signifikant, obwohl die Ergebnisse für F_1 und F_2 knapp über der Signifikanzgrenze liegen.

Tabelle 6.23: Durchschnittswerte des gemischten Modells für / ϵ / bei deutschen und chinesischen weiblichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	651	556
F_2 (Hz)	1587	1978
F_3 (Hz)	2776	2798
Vokaldauer (ms)	66,02	95,42

Das von chinesischen Sprechern artikulierte / ϵ / hat höhere Formantwerte und tendiert vom resynthetisierten Klang her in Richtung /e:/, wobei F_1 zwischen den muttersprachlichen Werten für /e:/ und / ϵ / liegt, F_2 jedoch noch im Bereich des muttersprachlichen / ϵ /. Es unterscheidet sich allerdings deutlich vom artikulierten /e:/, dessen Formantwerte noch extremer liegen.

6.6 Im Chinesischen unbekannte Vokale

6.6.1 Umsetzung von /ø:/ durch chinesische Sprecher

Untersuchungsgegenstand 8 (Der Vokal /ø:/) *Der im Chinesischen unbekannte Laut /ø:/ wird von deutschen und chinesischen Sprechern gleich ausgesprochen.*

6.6.1.1 Statistischer Vergleich

Tabelle 6.24: Ergebnisse des Wilcoxon-Rangsummentests für /ø:/ bei männlichen und weiblichen Sprechern

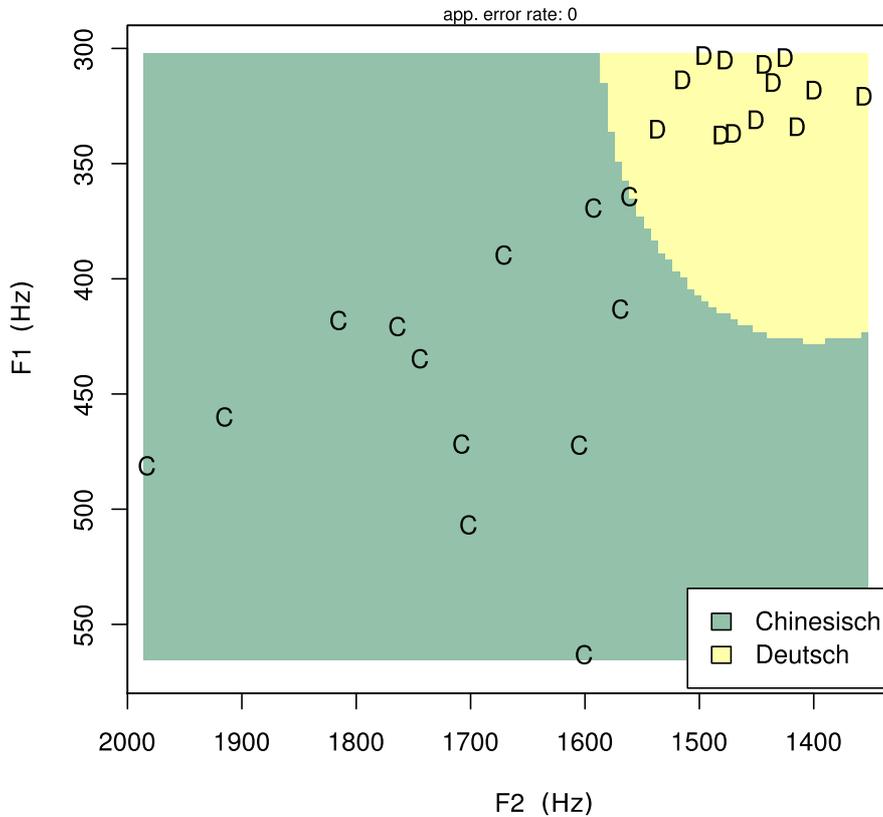
Parameter	p-Wert	
	Männer	Frauen
F_1	1.923 * 10⁻⁷	0.00433
F_2	1.923 * 10⁻⁷	3.133 * 10⁻⁵
F_3	1.346 * 10⁻⁶	0.02656
Vokaldauer	0.6437	0.02160

Dies könnte darin bedingt sein, dass in der überwiegenden Mehrheit der untersuchten Abschnitte das Wort *Löwe* gelesen wurde (92% bei männlichen Sprechern, 96% bei weiblichen Sprechern).

6.6.1.2 Diskriminanzanalyse

Da die Varianzen der Wertemengen sich für ein gemischtes Modell zu stark unterscheiden, führe ich hier eine Diskriminanzanalyse durch.

Abbildung 6.25: Verteilung der Werte von F_1 und F_2 für /ø:/ bei männlichen Sprechern

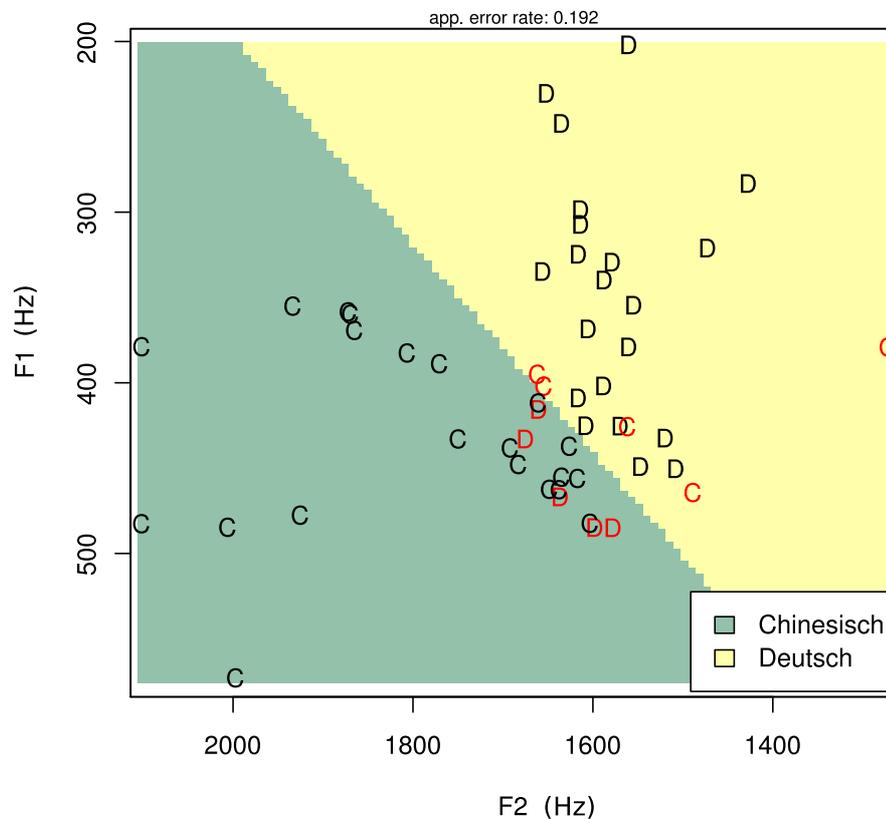


F_1 und F_2 sind bei chinesischen Sprechern jeweils höher und viel weiter gestreut.

Tabelle 6.25: Durchschnittliche Messwerte für /ø:/ bei deutschen und chinesischen männlichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	320	444
F_2 (Hz)	1454	1710
F_3 (Hz)	2325	2605
Vokaldauer (ms)	149,23	146,92

Errechne ich die Mittelwerte für die deutschen und die chinesischen Sprecher, so liegt der resynthetisierte Vokal für die chinesischen Werte vom Klang her näher bei dem Vokal /ε:/ deutscher Muttersprachler.

Abbildung 6.26: Verteilung der Werte von F_1 und F_2 für /ø:/ bei weiblichen Sprecherinnen

In diesem Fall ist der Abstand zwischen chinesischen und deutschen Formantwerten geringer und es existiert ein Bereich, in dem sich beide überlappen. Trotzdem bleibt eine deutliche Tendenz zu höheren F_1 - und F_2 -Werten in der chinesischen Gruppe sichtbar.

Tabelle 6.26: Durchschnittliche Messwerte für /ø:/ bei deutschen und chinesischen weiblichen Sprechern

Parameter	deutsche Sprecher	chinesische Sprecher
F_1 (Hz)	370	430
F_2 (Hz)	1586	1748
F_3 (Hz)	2454	2563
Vokaldauer (ms)	135,00	159,23

Die durchschnittlichen Formantwerte der chinesischen Gruppe liegen hier näher bei den Werten für /ə/ von deutschen Muttersprachlern. Die Vokaldauer fällt jedoch deutlich länger aus als bei den männlichen Sprechern oder in den deutschen Sprechergruppen.

6.7 Schlussfolgerungen

Für die chinesischen Sprecher läßt sich insgesamt sagen, dass sich die Werte für F_3 für alle Vokale von deutschen Messwerten unterscheiden. Die Vokaldauer unterscheidet sich bei Männern für die extremen Vokale /i:/, /ɪ/ und /a/, für die Frauen für die vorderen und mittleren Vokale /i:/, /ɛ/, /ø:/ und /a/.

Interessanterweise machen die Sprecher einen Unterschied zwischen dem gespannten Vokal /u:/ und dem ungespannten /ʊ/, auch wenn die konkrete Umsetzung noch anders klingt, als bei deutschen Sprechern.

Was in diesen Daten am meisten auffällt ist die breitere Streuung der nicht-muttersprachlichen Daten, die noch extremer ausfallen würde, wenn man all die Abschnitte mit einbezöge, bei denen muttersprachliche Monophthonge als Diphthonge realisiert wurden.

Zusätzlich fallen in vielen Fällen die Formantwerte der chinesischen Fremdsprachenlernenden extremer aus, vor allem bei F_2 höher, bei F_1 für /u/ auch niedriger.

Die Abweichungen für /ø:/ (und möglicherweise auch für andere im Chinesischen unbekannte Vokale) sind zu groß, um durch lineare Modelle beschrieben zu werden.

7 Diskussion

In diesem Kapitel möchte ich noch einmal die wichtigsten Punkte meiner Arbeit aufgreifen und erläutern.

7.1 Sprachspezifische Resultate

7.1.1 Deutsche Muttersprachler

Von den Ergebnissen im muttersprachlichen Kapitel halte ich folgende für besonders interessant:

/a:/ und /a/ Diese Vokale unterscheiden sich nur in einzeln gelesenen Worten eindeutig, in denen sie überdeutlich ausgesprochen werden, um Verwechslungen zu vermeiden. In gelesenen Texten geht dieser Unterschied zurück, in spontanen Gesprächen verschwindet er ganz.

/ɛ:/ und /ɛ/ Hier liegt im Gegensatz zu /a:/ und /a/ zusätzlich zur Vokalquantität ein deutlicher Unterschied in der Vokalqualität vor. Dieser verstärkt sich in spontanen sprachlichen Aufnahmen noch, in denen sich /ɛ:/ leicht /e:/ annähert, jedoch nicht mit ihm zusammenfällt.

/ɛ/ und /ə/ In den Referenzdaten handelt es sich um zwei deutlich voneinander getrennte Laute, da /ɛ/ im Vergleich zu /ə/ auch in betonten Silben auftreten kann, z.B. in *Bett*. Muttersprachler aus Sprachen, die kein /ə/ in ihrem Phoneminventar kennen, wie Italienisch bzw. bei denen /ə/ ein Allophon zu /ɛ/ darstellt, wie im Chinesischen, würde man deshalb am besten deutlich auf diesen Unterschied aufmerksam machen.

7.1.2 Britische Sprecher

Zur Aussprache von /a:/ und /a/ und der möglichen Substitution durch muttersprachliche Laute konnte ich hier keine Belege finden, dies könnte jedoch auch daran liegen, dass es sich bei den Versuchspersonen durchweg um als sehr gut eingestufte Sprecher handelt. Der Gegensatz zwischen /o:/ und /ɔ/ könnte dagegen auch durch die Auswirkungen zu kleine Datenmengen überlagert worden sein.

Das Merkmal, das britische und deutsche Sprecher am deutlichsten voneinander unterscheidet, ist die Aussprache gerundeter Vokale.

Dies trifft auf den in beiden Sprachen vorhandenen Vokal /u/ zu, der im britischen Vokalraum mit einem wesentlich zentraleren F_2 -Wert ausgesprochen wird, da er dort nicht von /ø:/ unterschieden werden muss.

Noch deutlicher wird es bei den gerundeten Umlauten /y:/ und /ø:/. Dabei wird /y:/ durch Werte ersetzt, die in Richtung eines extremer artikulierten /i:/ tendieren (s. S. 101), wozu beitragen könnte, dass /y:/ meistens in betonten Silben auftaucht.

/ø:/ dagegen wird oft durch zentralere Laute wie /ε/ oder sogar /ə/ ersetzt. Da die Formantwerte für F_1 und F_2 denen eines britischen /u/ entsprechen, könnte es für deutsch lernende Sprecher verwirrend sein, eine gewohnte Zungenposition von einem muttersprachlichen auf einen neu erlernten Laut zu übertragen.

7.1.3 Italienische Sprecher

Da die Vokallänge vor allem durch die Silbenbetonung, zusätzlich noch durch die Position der Silbe im Wort bestimmt wird und keine lexikalische Rolle spielt, ist es wichtig, italienische Lernende auf Gegensatzpaare wie *fahl* [fa:l] und *Fall* [fal] hinzuweisen, da diese sonst unter Umständen nicht wahrgenommen werden.

Im Gegensatz zu /u:/ und /ʊ/, die sich vor allem durch ihre F_1 -Werte und damit durch den Grad der Anhebung der Zunge nach hinten unterscheiden und von den italienischen Sprechern korrekt realisiert werden, werden /i:/ und /ɪ/ anders als bei deutschen Muttersprachlern umgesetzt.

Dabei entsprechen die Messwerte für /ɪ/ in betonten Silben denen für /i:/ in betonten Silben deutscher Muttersprachler, die Messwerte für /ɪ/ in unbetonten Silben dagegen nähern sich eher der deutschen Aussprache von /ɪ/ an. Da sich im Italienischen unbetonte Silben durch eine zentralere Artikulation von betonten Silben abheben, scheinen die italienischen Lernenden /i:/ und /ɪ/ als betonte bzw. unbetonte Allophone von /i/ zu behandeln.

Für Umlaute wie /ø:/ ist zu beachten, dass die Strategien zum Erwerb dieses neuen Vokals differieren können, so dass er bei einigen Sprechern mehr in Richtung eines /ε/ bei einem anderen eher in Richtung eines /ɔ/ umgesetzt wird.

7.1.4 Chinesische Sprecher

In dieser Gruppe fällt es mir schwer, generelle Schlussfolgerungen zu ziehen, da ich keine Informationen über die regionale Herkunft der Sprecher und damit über die eventuellen Auswirkungen einer Zweisprachigkeit (z.B. mit Kantonesisch oder Min) besitze.

Insgesamt wirkt sich die komplexe chinesische Vokalstruktur stärker auf die Aussprache der deutschen Vokale aus, als in europäischen Sprachen. Dies gilt vor allem in den Clustern für /a:/ und /a/, von denen eines dem chinesischen Vokal /ʏ/ entspricht.

Für /i:/ und /ɪ/ müsste man dagegen prüfen, inwieweit sich die Diphthongisierung von /i/ zu /ji/ am Wortanfang auf die Vokalqualität auswirkt.

Interessanterweise unterscheiden alle chinesischen Sprecher zwischen /u:/ und /ʊ/, wobei sich die F_1 - und F_2 -Werte für /ʊ/ zwischen den Werten für ein deutsches /u:/ und ein deutsches /ʊ/ befinden. Auch die Formantwerte für /ɛ/ würden sich auf einer Linie zwischen dem deutschen /e:/ und /ɛ/ abtragen lassen.

Dies bedeutet, dass chinesische Sprecher z.B. den Kontrast zwischen /u:/ und /ʊ/ hören, ihn jedoch weniger ausgeprägt artikulieren als deutsche Sprecher.

Für /ø:/ kann ich eine sehr starke Streuung beobachten, mit einer Tendenz, diesen Vokal wie das deutsche /ɛ/ auszusprechen. Während sich die F_3 -Werte mit einer anderen Kehlkopfstellung für alle Vokale begründen lassen, weist der höhere F_2 -Wert deutlich auf eine fehlende Lippenrundung hin. Dies würde auch die höheren F_2 -Werte für /o/ und /u/ bei männlichen und weiblichen Sprechern erklären.

Dabei wäre dieses Merkmal im Chinesischen nicht unbekannt, da auch der vordere gerundete Vokal /y/ zum Phoneminventar gehört. So muss mit /ø:/ nur ein neuer Laut, jedoch nicht ein neues phonetisches Merkmal erworben werden.

7.2 Bewertung der Ausspracheunterschiede

Im Rahmen meiner Untersuchungen stellt sich die Frage, ob neue Phoneme, z.B. Umlaute, oder neue linguistische Kategorien, z.B. eine kontrastive Vokallänge einfacher zu erwerben sind.

Der Gegensatz zwischen gespannten und ungespannten Vokalen existiert in den hier untersuchten Sprachen außer im Deutschen nur noch im Englischen, wo er jedoch eher mit der Vokalqualität, also den Formantfrequenzen, als mit der Vokalquantität verknüpft ist. Im Italienischen und im Chinesischen existiert kein Gespanntheitskontrast und die Vokallänge hat dort keine phonemische Bedeutung.

In meinen Untersuchungen unbekannter Vokale ist die Streuung der Werte deutlich größer, da diese auch von Sprechern mit der gleichen Muttersprache durch unterschiedliche bekannte Laute substituiert werden. Diese Unterschiede nehmen bei sehr guten Zweitsprachenlernern jedoch ab.

Die mangelnde Unterscheidung zwischen gespannten und ungespannten Vokalen lässt sich schwerer ermitteln, bleibt jedoch auch bei sehr guten Zweitsprachlern erhalten, sofern der Gegensatz gespannt – ungespannt nicht schon in der Muttersprache existiert.

7.3 Vergleich weiblicher Sprachaufnahmen

In meinen Vergleichen weiblicher Sprachaufnahmen zeichnen sich diese durch eine größere Variabilität aus, auch was die Unterschiede zwischen einzelnen Sprecherinnen betrifft. Leider kann ich dabei für gemessene Formantwerte keine eindeutige Korrelation mit F_0 feststellen, auch nicht, wenn ich auf Formantverläufe zurückgreife. Setze ich in gemischten Modellen jedoch den einzelnen Sprecher als Zufallsfaktor ein, so erhalte ich eine bessere

Anpassung des Modells an die gemessenen Daten und damit klarere Resultate als bei männlichen Sprechern.

7.4 Parametrische statistische Verfahren und Data Mining

7.4.1 Gemischte lineare und generalisierte lineare Modelle

Solange es um rein zufällige Faktoren, wie z.B einzelne Sprecher geht, können gemischte Modelle leichte Abweichungen von der Normalverteilung ausgleichen, wenn diese durch sprecherspezifische Eigenheiten bedingt sind.

Spiegeln die Abweichungen jedoch grundlegende Unterschiede wieder, z.B. durch ko-artikulatorische Einflüsse, so habe ich entweder die Wahl, sie als feste Faktoren in mein gemischtes Modell einzubauen und dabei dieselben Bedingungen einzuhalten wie bei der Kovarianzanalyse oder andere Verfahren mit einzubeziehen, die auch unterschiedliche Gruppengrößen und -verteilungen akzeptieren, wie den DTK-Test bei italienischen und die Klassifikationsbäume bei britischen, italienischen und chinesischen Sprechern.

7.4.2 Bayesianische Statistik

Die AIC- und BIC-Werte, die für gemischte Modelle, Ergebnisse der Diskriminanzanalyse und modellbasierte Cluster errechnet werden, sind eine interessante Möglichkeit, um zu vergleichen, inwieweit die Ergebnisse eines Tests mit den tatsächlich gemessenen Werten übereinstimmen und so das beste Modell auszuwählen.

Allerdings sind diese Werte stark vom Umfang der untersuchten Daten abhängig und ich konnte in der Fachliteratur noch keine Bewertung finden, welche Kennzahlen für welche Gruppengröße und -anzahl ideale Werte darstellen.

7.4.3 Klassifikationsbäume und Clustering

An den von mir verwendeten Klassifikationsbäumen und den im chinesischen Kapitel angewendeten modellbasierten Clustern schätze ich die Möglichkeit, übersichtliche grafische Darstellungen und bei einer entsprechend geringen Fehlerrate auch eindeutige linguistische Regeln erstellen zu können.

Ein Problem liegt darin, dass für die Fehlerrate anders als beim Signifikanzniveau parametrisch definierter Verfahren keine üblichen Grenzwerte existieren. Nach meiner Erfahrung sprechen Fehlerraten von unter 10% für ein sehr gutes, von unter 20% für ein gutes und von mehr als 40% für ein unbrauchbares Ergebnis.

7.5 Ausblick

Mittlerweile existieren viele Untersuchungen, die Sprachsignale nicht nur akustisch, sondern zusätzlich auch kinematisch mit Hilfe von Sensoren oder visuell mit Hilfe von gefilmten Sprechbewegungen analysieren.

Würde ich diese Untersuchungen noch einmal durchführen, würde ich stärker auf Formantverläufe und ihr Verhältnis zu Grundfrequenzverläufen eingehen. Außerdem würde ich nicht grundsätzlich von einer Normalverteilung ausgehen, sondern eher ein generalisiertes lineares Modell auf der Grundlage einer Gamma- oder Weibull-Verteilung einsetzen, um unterschiedliche artikulatorische Charakteristika direkt beschreiben zu können.

A Anhang

Literatur

- Bear, M. F., Conners, B. W. & Paradiso, M. (2007). *Neuroscience: Exploring the brain* (3 Aufl.). Baltimore, MD, USA: Lippincott Williams & Wilkins.
- Becker, T. (2008, Dezember). Akzent und Lautwandel der Vokale seit althochdeutscher Zeit. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 130(3), 401–419.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M. & Prieto, T. (1997, January). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience*, 17(1), 353–362.
- B.L.Welch. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350–360.
- Bohn, O.-S. (1993). Interactions between new and similar vowels in L2 development. In B. Kettermann & W. Wieden (Hrsg.), (S. 71–79). Tübingen: Gunter Narr Verlag.
- Bosch, L. & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46(2–3), 217–243.
- Breiman, L. (2001, August). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199–215.
- Bussmann, H. (2002). *Lexikon der Sprachwissenschaft* (3 Aufl.). Stuttgart: Alfred Kröner Verlag.
- Byrd, B. R. (2009). Media representations of Turkish-German and hip-hop language as a uniform ethnolect. In *Texas linguistic forum* (Bd. 53, S. 72–78).
- Caldognetto, E. M., Zmarich, C. & Cosi, P. (1998). Statistical definition of visual information for Italian vowels and consonants. In *Ausp'98*. Terrigal, Sydney, Australia.
- Celata, C. & Kaeppli, B. (2004). Affricazione e rafforzamento in italiano: alcuni dati sperimentali. In *il primo congresso nazionale dell' AISV (Associazione Italiana di Scienze della Voce*. Padova, Italia.
- Charles R. Henderson, J. (1982, September). Analysis of covariance in the mixed model: Higher-level, nonhomogeneous, and random regressions. *Biometrics*, 38(3), 623–640.
- Chen, Y. (2002). Accentual lengthening of monosyllabic constituents in Beijing Mandarin. *Speech Prosody*, 211–214.
- Chen, Y. (2006). Production of tense-lax contrast by Mandarin speakers of English. *Folia Phoniatica et Logopaedica*, 58, 240–249.

- Cheng, C.-C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague, The Netherlands: Mouton & Co., N.V.
- Cochran, W. G. (1947). Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 3(1), 22–38.
- Cochran, W. G. (1957, September). Analysis of covariance: Its nature and uses. *Biometrics*, 13(3), 261–281.
- Cosi, P., Ferrero, F. & Vagges, K. (1995). Rappresentazioni acustiche e uditive delle vocali italiane.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D. et al. (1997, November). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8, 3809–3815.
- Deretsky, Z. (2006). *Hearing mechanics*. National Science Foundation. (This work is in the public domain in the United States because it is a work of the United States Federal Government under the terms of Title 17, Chapter 1, Section 105 of the US Code.)
- Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, 27, 47–55.
- D’Imperio, M. & Rosenthal, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, 16(1), 1–28.
- Disner, S. F. (1983). Vowel quality: The relation between universal and language-specific factors. *Working Papers in Phonetics*(58).
- Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J. et al. (2002). The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, 15, 59–90.
- Dretzke, B. (1998). *Modern British and American English pronunciation*. Paderborn: Verlag Ferdinand Schöningh.
- Duanmu, S. (2000). *The phonology of Standard Chinese*. Walton Street, Oxford OX2 6DP, UK: Oxford University Press.
- Duanmu, S. (2009, March). Chinese syllable structure. In C. Ewen, B. Hume, M. van Oostendorp & K. Rice (Hrsg.), *Companion to phonology*. Wiley.
- Dunnett, C. (1980, December). Pairwise multiple comparisons in the homogeneous variance, unequal sample size case. *Journal of the American Statistical Association*, 75(372), 789–795.
- Díaz, B., Baus, C., Escera, C., Costa, A. & Sebastián-Gallés, N. (2008, October). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences of the USA*, 105(42), 16083–16088.
- Eisenhart, C. (1947, March). The assumptions underlying the analysis of variance. *Biometrics*, 3(1), 1–21.
- Eulitz, C., Diesch, E., Pantev, C., Hampson, S. & Elbert, T. (1995, March). Magnetic and

- electric brain activity evoked by the processing of tone and vowel stimuli. *Journal of Neuroscience*, 15(4), 2748–2755.
- Fiukowski, H. (2002). *Sprecherzieherisches Elementarbuch*. Tübingen: Max Niemeyer Verlag GmbH.
- Flege, J. E., Bohn, O.-S. & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, J. E. & Hillenbrand, J. (1984, September). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustic Society of America*, 76(3), 708–721.
- Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J. & Viallet, F. (2005, May). Articulation in early and late bilinguals' two languages: evidence from functional magnetic resonance imaging. *Neuroreport*, 16(7), 761–765.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, 50, 259–281.
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84, 165–175.
- Gamkrelidze, T. V. (2008). The indo-european glottalic theory: A new paradigm in indo-european comparative linguistics. *Bulletin of the Georgian Academy of Sciences*, 2(1), 140–143.
- Gayen, A. K. (1949, December). The distribution of 'student's t in random samples of any size drawn from non-normal universes. *Biometrika*, 36(3/4), 353–369.
- Gayen, A. K. (1950, December). The distribution of the variance ratio in random samples of any size drawn from non-normal universes. *Biometrika*, 37(3/4), 236–255.
- Gelderen, E. van. (2006). *A history of the English language*. Amsterdam, The Netherlands: John Benjamins B.V.
- Gendrot, C. & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. In S. U. Institute of Phonetics (Hrsg.), *ICPhS 2007 proceedings*.
- Genot, G. (1984). *Linguistique italienne*. Nanterre, France: Univ. Nanterre.
- Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3), 237–288.
- Goettsch, C. (1908, April). "Ablaut"-relations in the weak verb in Gothic, Old High German, and Middle High German. *Modern Philology*, 5(4), 569 – 616.
- Gootjes, L., Raij, T., Salmelin, R. & Hari, R. (1999). Left-hemisphere dominance for processing of vowels: a whole-scalp neuromagnetic study. *Neuroreport*, 10, 2987–2991.
- Gottfried, T. L. & Strange, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustic Society of America*, 68(6), 1626–1635.

- Gray, H. (1918). *Henry Gray's anatomy of the human body* (20 Aufl.). Philadelphia, USA: Lea & Febiger.
- Greenberg, B. G. (1953). The use of analysis of covariance and balancing in analytical surveys. *American Journal of Public Health*, 43, 692–699.
- Gu, Z., Mori, H. & Kasuya, H. (2003). Analysis of vowel formant frequency variations between focus and neutral speech in Mandarin Chinese. *Acoustic Science & Technology*, 24(4), 192–193.
- Gut, U. (2006). Unstressed vowels in non-native German. *Speech Prosody*.
- Gut, U. & Bayerl, P. S. (2004). Measuring the reliability of manual annotations of speech. In (S. 565-568). Nara, Japan.
- Hack, H. R. B. (1958, June). An empirical investigation into the distribution of the f-ratio in samples from two non-normal populations. *Biometrika*, 45(1/2), 260–265.
- Harrington, J. (1999). *Techniques in speech acoustics*. Dordrecht, NL: Kluwer Academic Publishers.
- Harshman, R., Ladefoged, P. & Goldstein, L. (1977, September). Factor analysis of tongue shapes. *Journal of the Acoustic Society of America*, 62(3), 693–707.
- Hillenbrand, J. M. (2001, February). Effect of consonant environment on vowel formant patterns. *Journal of the Acoustic Society of America*, 109(2), 748–763.
- Hombert, J.-M., Ohala, J. J. & Ewan, W. G. (1979, March). Phonetic explanations for the development of tones. *Language*, 55(1), 37–58.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. The Edinburgh Building, CB2 8RU, UK: Cambridge University Press.
- Huang, L. M. (1992). Remarks on the phonological structure of Mandarin Chinese. *Bul. Nat 1 Taiwan Nor. Univ.*, 37, 363–383.
- Hull, R., Bortfeld, H. & Koons, S. (2009). Near-infrared spectroscopy and cortical responses to speech production. *The Open Neuroimaging Journal*, 3, 26–30.
- Jackson, M. T. T. (1988). *Phonetic theory and cross-linguistic variation in vowel articulation*. Unveröffentlichte Dissertation, University of California Los Angeles (UCLA).
- Jenkins, J. (2001). *The phonology of English as an international language* (2 Aufl.). Walton Street, Oxford OX2 6DP, UK: Oxford University Press.
- Jessen, M. (1993). Stress conditions on the vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory*, 8, 1–27.
- Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Jones, D., Roach, P., Hartmann, J. & Setter, J. (2003). *English pronouncing dictionary* (16 Aufl.). The Edinburgh Building, CB2 8RU, UK: Cambridge University Press.
- Joos, M. (1948). Acoustic phonetics. *Language*, 24(2), 5–136.
- Karnath, H.-O. & Thier, P. (2006). *Neuropsychologie* (2 Aufl.). Heidelberg: Springer Medizin Verlag.

- Kastenbaum, M. A., Hoel, D. G. & Bowman, K. O. (1970). Sample-size requirements: one-way analysis of variance. *Biometrika*, 57(2), 421–430.
- Kent, R. D. & Forner, L. L. (1979). Developmental study of vowel formant frequency in an imitation task. *Journal of the Acoustic Society of America*, 65(1), 208–217.
- Klein, D., Zatorre, R. J., Milner, B., Meyer, E. & Evans, A. C. (1994). Left putaminal activation when speaking a second language: evidence from PET. *Neuroreport*, 5(17), 2295–2297.
- Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V. et al. (2004). Speech-sound discrimination in neonates as measured with meg. *Neuroreport*, 15(13), 2089–2092.
- Ladefoged, P. (2006). *A course in phonetics* (5 Aufl.). Boston, MA, USA: Thomson Wadsworth.
- Landi, R. & Savy, R. (1996). Durata vocalica, struttura sillabica e velocita' di eloquio nel parlato connesso. In (S. 65–70). Trento, Italia.
- Leech, N. & Onwuegbuzie, A. J. (2002). A call for greater use of nonparametric statistics. In *Annual meeting of the mid-south educational research association*. Chattanooga, Tennessee, USA.
- Lehmann, W. P. (1961). A definition of Proto-Germanic: A study in the chronological delimitation of languages. *Language*, 37(1), 67–74.
- Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232.
- Liao, J.-S. (2006). *English vowel production of Mandarin speakers*. Unveröffentlichte Dissertation, University of Texas at Austin.
- Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge University Press.
- Liu, K., Shuang, Z., Qin, Y., Zhang, J. & Yan, Y. (2007). Mandarin accent analysis based on formant frequencies. In *Ieee international conference on acoustics, speech and signal processing*. Honolulu, USA.
- Long, M. H. (1993). Second language acquisition as a function of age: research findings and methodological issues. In K. Hyltenstan & Åke Viberg (Hrsg.), *Progression & regression in language — sociocultural, neuropsychological & linguistic perspectives*. Cambridge University Press.
- Major, R. C. (2001). *Foreign accent: the ontology and phylogeny of second-language phonology*. USA: Lawrence Erlbaum Associates.
- Man, C. Y. (2004). *A perceptual analysis of the apical vowels in Beijing Mandarin*. Ma thesis, City University of Hong Kong.
- Mangold, P. D. M. & Dudenredaktion. (2005). *Duden. Das Aussprachewörterbuch* (6 Aufl.). Mannheim: Bibliographisches Institut & F.A. Brockhaus AG.
- McAllister, R. (1999). The acquisition of Swedish long vs. short vowel contrasts by native speakers of English, Spanish and Estonian. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, USA.

- McCobb, A. L. (1915, February). The loss of unaccented e in the 'transition period'. *Modern Language Notes*, 30(2), 39-41.
- McIntyre, D. (1975). *History of English: a resource book for students*. London, UK: Routledge.
- Meng, Z., Chen, Y. & Li, X. (2006). Statistical survey of monophthong formants in Mandarin for students being trained as broadcasters. *Acta Acoustica*, 25(2). (Communication Acoustics Laboratory, Communication University of China, Beijing P.R.C.)
- Minagawa-Kawai, Y., Mori, K., Naoi, N. & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience*, 27(2), 315-321.
- Minegawa-Kawai, Y., Mori, K. & Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of Cognitive Neuroscience*, 17(9), 1376-1385.
- Moessner, L. (2003). *Diachronic English linguistics: An introduction*. Tübingen: Gunter Narr Verlag.
- Mooshammer, C., Fuchs, S. & Fischer, D. (1999). Effects of stress and tenseness on the production of CVC syllables in German. *ICPhS99*.
- Moulton, W. G. (1961). Zur Geschichte des deutschen Vokalsystems. *Beiträge zur Geschichte der deutschen Sprache und Literatur*.
- Munro, M. J., Derwing, T. M. & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111-131.
- Mäkelä, A. M., Alku, P., May, P. J. C., Mäkinen, V. & Tiitinen, H. (2005, April). Left-hemispheric brain activity reflects formant transitions in speech sounds. *Neuroreport*, 16(6), 549-553.
- Nenonen, S., Shestakova, A., Huotilainen, M. & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language*, 92, 26-32.
- Neppert, J. M. H. (1999). *Elemente einer akustischen Phonetik* (4 Aufl.). Hamburg: Buske Verlag.
- Näätänen, R. & Alho, K. (1995). Mismatch negativity - a unique measure of sensory processing in audition. *Journal of Neuroscience*, 80, 317-337.
- Pertsch, D. E. (1999). *Langenscheidts Großes Schulwörterbuch* (13 Aufl.). Berlin und München: Langenscheidt KG.
- Pokorny, J. (1959). *Indogermanisches etymologisches Wörterbuch* (26 Aufl.). Bern: A. Francke AG Verlag.
- Pons Kompaktwörterbuch Italienisch*. (2009). Stuttgart.
- Ramers, K. H. (1988). *Vokalqualität und -quantität im Deutschen*. Tübingen: Max Niemeyer Verlag GmbH.
- Ramsey, P. H. (1980). Exact type 1 error rates for robustness of student's t test with unequal variances. *Journal of Educational and Behavioral Statistics*, 5(4), 337-349.

- Recasens, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech*, 28(2), 97-114.
- Reinecke, J. & Tarnai, C. (Hrsg.). (2008). *Klassifikationsanalysen in Theorie und Praxis*. Münster: Waxmann Verlag GmbH.
- Reinke, K. (2007). Zur emotionalen Wirkung phonetischer Mittel bei DaF-Lernenden. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 12(2).
- Reiterer, S., Hemmelmann, C., Rappelsberger, P. & Berger, M. L. (2005). Characteristic functional networks in high- versus low-proficiency second language speakers detected also during native language processing: An explorative EEG coherence study in 6 frequency bands. *Cognitive Brain Research*, 25, 566-578.
- R.H.Baayen. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. The Edinburgh Building, CB2 8RU, UK: Cambridge University Press.
- Ripley, B. D. & Venables, W. N. (2002). *Modern applied statistics with S* (4 Aufl.). New York, NY, USA: Springer-Verlag Inc.
- Robinson, G. K. (1991, February). That blup is a good thing: The estimation of random effects. *Statistical Science*, 6(1), 15-32.
- Rogan, J. C. & Keselman, H. J. (1977). Is the anova f-test robust to variance heterogeneity when sample sizes are equal?: An investigation via a coefficient of variation. *American Educational Research Journal*, 14(4), 493-498.
- Sachs, L. & Hedderich, J. (2006). *Angewandte Statistik: Methodensammlung mit R* (12 Aufl.). New York, NY, USA: Springer-Verlag Inc.
- Santen, J. van & Imperio, M. D. (1999). Positional effects on stressed vowel duration in standard Italian. In *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, USA.
- Sapir, E. (1915, October). Notes on Judeo-German phonology. *The Jewish Quarterly Review*, 6(2), 231-266.
- Savy, R. & Cutugno, F. (1997). Ipoarticolazione, riduzione vocalica, centralizzazione: come interagiscono nella variazione diafasica? In F. Cutugno (Hrsg.), *Fonetica e fonologia degli stili dell'Italiano parlato* (S. 177-194). Roma, Italia.
- Schindelmeiser, J. (2005). *Anatomie und Physiologie für Sprachtherapeuten* (1 Aufl.). München: Elsevier GmbH.
- Schofield, T. M., 1, P. I., Kiebel, S. J., Stephan, K. E., Kilner, J. M., Friston, K. J. et al. (2009). Changing meaning causes coupling changes within higher levels of the cortical hierarchy. *Proceedings of the National Academy of Sciences of the USA*, 106(29), 11765-11770.
- Segaran, T. (2007). *Programming collective intelligence: Building smart web 2.0 applications* (1 Aufl.). 1005 Gravenstein Highway North, Sebastopol, CA 95472, USA: O'Reilly Media, Inc.
- Simpson, A. P. (1988). Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. *Arbeitsberichte des Instituts für*

Phonetik und digitale Sprachverarbeitung(33), 233.

- Srivastava, A. B. L. (1958, December). Effect of non-normality on the power function of t-test. *Biometrika*, 45(3/4), 421–430.
- Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA, USA: MIT Press.
- Sundberg, J. (1972). An articulatory interpretation of the 'singing formant'. *Dept. for Speech, Music and Hearing Quarterly Progress and Status Report*, 13(1), 045–053.
- Sundberg, J. (1977). *The acoustics of the singing voice*.
- Surendran, D. & Levow, G.-A. (2004). The functional load of tone in Mandarin is as high as that of vowels. *Speech Prosody*, 99–102.
- Szulc, A. (1987). *Historische Phonologie des Deutschen*. Tübingen: Max Niemeyer Verlag GmbH.
- Thigpen, M. (2010). *chinese-lessons.com: Mandarin sounds*. (<http://www.chinese-lessons.com>)
- Twaddell, W. F. (1948). The prehistoric Germanic short syllabics. *Language*, 24(2), 139 – 151.
- Vayra, M. (1999). On the phonetic bases of vowel-consonant coordination in Italian: a study of stress and "compensatory shortening". In *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, USA.
- Vennemann, T. (1972, December). Phonetic detail in assimilation: Problems in Germanic phonology. *Language*, 48(4), 863–892.
- Videsott, G., Herrnberger, B., Hoenig, K., Schilly, E., Grothe, J., Wiater, W. et al. (2010). Speaking in multiple languages: Neural correlates of language proficiency in multilingual word production. *Brain and Language*, 1–10.
- Weinhold, K. (1919). *Kleine mittelhochdeutsche Grammatik* (5 Aufl.). Wien: Braumüller.
- Wells, J. C. (1995, April). *Computer-coding the IPA: a proposed extension of SAMPA*.
- Winkler, I. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, 36, 638–642.
- Wode, H. (1993). *Psycholinguistik: eine Einführung in die Lehr- und Lernbarkeit von Sprachen; Theorien, Methoden, Ergebnisse* (1 Aufl.). Ismaning: Max Hueber Verlag.
- XIVth International Congress of Phonetic Sciences*. (1999). San Francisco, USA.
- York, B. J. (2008). *The effects of experience on the perception of German rounded vowels by native speakers of American English*. MA thesis, Center for Language Studies, Brigham Young University, USA.
- Youngblood, J. L. (2007). *Vowel length in standard Italian and northern Italian dialects*. MA thesis, University of Texas at Austin.
- Zamboni, A. (1992). *Manuale di pronuncia italiana* (1 Aufl.). Bologna, Italia: Zanichelli editore S.p.A.
- Zee, E. & Lee, W.-S. (2001). An acoustic analysis of the vowels in Beijing Mandarin. In *Eurospeech*.
- Zhang, J. (1996). The consonantal properties of the unrounded high-mid back vowel /ɤ/

in Beijing Mandarin. In *131st meeting of the Acoustical Society of America (asa)*. Indianapolis, IN.