

**Experimentelle Untersuchungen zum
computergestützten Erlernen der Prosodie des
Deutschen für italienischsprachige Lernende**

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie an der
LUDWIG-MAXIMILIANS-UNIVERSITÄT
MÜNCHEN

im Cotutelle-Verfahren mit der Universität

UNIVERSITÀ DEGLI STUDI DI SASSARI

vorgelegt von

Maria Paola BISSIRI

Betreuer an der Ludwig-Maximilians-Universität München:
Prof. Dr. Hans G. Tillmann

Betreuerin an der Università degli Studi di Sassari:
Prof. Dr. Livia Tonelli

Gutachter der Ludwig-Maximilians-Universität München:
PD Dr. phil. habil. Hartmut R. Pfitzinger

Gutachter der Università degli Studi di Sassari:
Prof. Carlo Schirru

Tag der mündlichen Prüfung: 18.03.2008 (in Sassari)

UNIVERSITÀ DEGLI STUDI DI SASSARI
Dottorato di ricerca in TEORIE E PRATICHE DELLA
INTERPRETAZIONE E DELLA TRADUZIONE
LINGUISTICA, LETTERARIA E CULTURALE, XX ciclo
(A.A. 2004/2005 – A.A. 2006/2007)

in cotutela con l'Università:

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

**Experimentelle Untersuchungen zum
computergestützten Erlernen der Prosodie des
Deutschen für italienischsprachige Lernende**

Tesi di dottorato di
Maria Paola BISSIRI

Coordinatore del dottorato:
Prof. Dr. Baingio PINNA

Tutori di tesi:
Prof. Dr. Livia TONELLI

Prof. Dr. Hans G. TILLMANN

Bissiri, Maria Paola (2008): Experimentelle Untersuchungen zum computergestützten Erlernen der Prosodie des Deutschen für italienischsprachige Lernende. Dissertation, LMU München: Fakultät für Sprach- und Literaturwissenschaften

ISBN 978-3-00-040107-7

© Maria Paola Bissiri 2008

Riassunto

Titolo: Analisi sperimentali sull'apprendimento mediato dal computer della prosodia del tedesco per discenti italofoeni

Sono stati testati sperimentalmente nuovi metodi per insegnare mediante il computer la prosodia del tedesco ad apprendenti italofoeni. La ricerca si concentra in particolare sull'accento lessicale di parole morfologicamente complesse (composti nominali e verbi prefissati), spesso accentati erroneamente dai parlanti italofoeni.

Tramite tecniche di sintesi del parlato, sono stati corretti i parametri prosodici della frequenza fondamentale e della velocità di elocuzione locale nelle produzioni errate nella seconda lingua (L2). La correzione di tali parametri ha influenzato positivamente il giudizio di parlanti di madrelingua tedesca sulla correttezza dell'accento lessicale. Successivamente, produzioni errate in L2 sono state corrette tramite la manipolazione della frequenza fondamentale, della velocità di elocuzione locale e dell'intensità. Il feedback costituito da tali stimoli risintetizzati oppure da stimoli dall'accento enfaticizzato è risultato significativamente più efficace rispetto al feedback tradizionale, dato dalla voce registrata di una parlante di madrelingua tedesca. Infine, i risultati di un test sulla percezione dell'accento indicano che la difficoltà dei parlanti italofoeni nell'apprendere l'accento lessicale in tedesco può avere cause percettive.

Inizialmente è stato raccolto un corpus di parlato del testo standard tedesco "Die Buttergeschichte", letto da dodici apprendenti italofoeni del tedesco, parlanti della varietà regionale sarda dell'italiano, e da una parlante di madrelingua tedesca. Il corpus è stato ampliato successivamente da registrazioni di ulteriori parlanti italofoeni di ambo i sessi e di varia provenienza regionale effettuate per i singoli esperimenti. Dalla segmentazione ed etichettatura dei dati è emersa una grande varietà di errori segmentali, come l'epentesi dello schwa, la riduzione di cluster consonantici e la sostituzione delle vocali anteriori arrotondate

con le corrispondenti posteriori. Inoltre, la quasi totalità delle parlanti italofone sbagliava sistematicamente l'accento delle parole morfologicamente complesse, caratterizzate dalla presenza di un accento principale e di un accento secondario: pronunciavano l'accento principale sulla penultima sillaba o sul secondo componente della parola, invece che sul primo, invertendo pertanto la posizione di accento principale e accento secondario.

Per correggere le produzioni dei parlanti italofofoni, sono stati copiati sui loro enunciati i parametri prosodici degli enunciati corrispondenti della parlante germanofona. I parametri modificati sono: a) l'intonazione (ovvero la frequenza fondamentale f_0), b) la velocità di elocuzione locale e c) l'intensità.

Nel primo esperimento, è stata corretta tramite la sintesi esclusivamente la velocità di elocuzione locale in undici frasi pronunciate da otto parlanti italofone, parlanti della varietà regionale sarda dell'italiano. Per allineare i segnali delle parlanti italofone al segnale della parlante germanofona è stata utilizzata la segmentazione dei fonemi e delle sillabe. In un test percettivo, 16 parlanti germanofoni hanno espresso un giudizio sul miglioramento del ritmo degli stimoli corretti rispetto agli stimoli originali. I risultati del test non hanno riportato un miglioramento significativo del ritmo nelle frasi risintetizzate. Tuttavia una seppur lieve tendenza in questo senso è riscontrabile nei dati.

Il secondo esperimento aveva lo scopo di verificare l'influenza della correzione dei singoli parametri prosodici sull'accento lessicale di produzioni in L2. A tal fine sono stati manipolati dieci enunciati (sei frasi e quattro parole singole) pronunciati da otto parlanti italofone, parlanti della varietà regionale sarda dell'italiano. Di tali enunciati sono state risintetizzate tre versioni, correggendo la frequenza fondamentale, la velocità di elocuzione locale oppure entrambi i parametri insieme. La sintesi è stata effettuata in questo caso senza l'aiuto della segmentazione, ma allineando i segnali tramite un algoritmo di Dynamic Time Warping. In un test percettivo svolto da 31 parlanti tedeschi

nativi, la manipolazione della frequenza fondamentale è risultata significativamente più efficace rispetto alla manipolazione della velocità di elocuzione locale nel correggere l'accento lessicale.

Nel terzo esperimento, stimoli corretti anche rispetto all'intensità, oltre che alla frequenza fondamentale e alla velocità di elocuzione locale, sono stati impiegati in un training di pronuncia. L'ipotesi principale era che il feedback costituito da stimoli risintetizzati, dati dagli enunciati dell'apprendente corretti tramite la sintesi, fosse più efficace rispetto al feedback costituito da stimoli naturali, ovvero dalle registrazioni della voce della parlante nativa tedesca. È stato inoltre testato il feedback dato da enunciati in cui la posizione dell'accento lessicale delle parole morfologicamente complesse era enfatizzata, in modo che gli apprendenti potessero identificarla più facilmente. A questo scopo è stato chiesto alla parlante di madrelingua tedesca di rileggere gli enunciati una seconda volta, ponendo maggior enfasi sull'accento delle parole. Dodici parlanti italofoeni provenienti da varie regioni italiane hanno partecipato al training. I partecipanti sono stati suddivisi in due gruppi. Un gruppo ha ricevuto come feedback gli stimoli naturali (ovvero prodotti dalla parlante di madrelingua tedesca), mentre l'altro gruppo ha ricevuto come feedback gli stimoli risintetizzati derivanti dalla manipolazione della propria voce. Il feedback è stato fornito prima senza e poi con enfasi sulla posizione dell'accento.

Nel quarto esperimento sono stati valutati in un test percettivo i risultati del training di pronuncia. Gli enunciati pronunciati dai partecipanti prima e dopo il training sono stati presentati in ordine casuale a 37 parlanti germanofoni, che avevano il compito di giudicare la correttezza dell'accento lessicale delle parole morfologicamente complesse. I risultati dell'analisi statistica dei dati hanno evidenziato una differenza significativa tra il gruppo sottoposto a training con gli stimoli naturali e il gruppo sottoposto a training con gli stimoli risintetizzati: gli stimoli risintetizzati sono risultati più efficaci degli stimoli naturali. Inoltre vi sono evidenze che gli stimoli naturali con enfasi facilitino

il riconoscimento della corretta posizione dell'accento più degli stimoli naturali senza enfasi. Tuttavia diversi parlanti italofoeni hanno imitato la versione dall'accento enfaticizzato nelle loro produzioni, per cui tali stimoli sono stati giudicati come corretti dai parlanti germanofoni, ma è stato loro assegnato un punteggio più basso di quello che avrebbero ottenuto se non fossero stati enfaticizzati.

Il quinto esperimento aveva lo scopo di verificare se la difficoltà dei parlanti italofoeni nell'imparare l'accento lessicale di parole tedesche morfologicamente complesse sia di origine percettiva. A tal fine 23 parlanti italofoeni hanno svolto un test in cui dovevano memorizzare delle serie di due o quattro parole costituite da parole tedesche o pseudotedesche. Nel test veniva inizialmente presentata una coppia minima, in cui le due parole si differenziavano per via di un fonema. Tale coppia costituiva la linea di base ed era la stessa per tutti i partecipanti. Successivamente venivano presentate ad ogni partecipante altre due coppie di parole, scelte a caso tra 20 coppie disponibili: a) una in cui le parole si differenziavano esclusivamente per via dell'accento e b) una in cui le parole si differenziavano sia per l'accento che per la quantità vocalica nella penultima sillaba. Per un gruppo di partecipanti il contrasto accentuale era dato da parole contenenti una vocale lunga e per l'altro gruppo da parole contenenti una vocale breve nella penultima sillaba. Lo stesso test è stato svolto da sei parlanti germanofoni come gruppo di controllo. I risultati del test hanno riportato un numero di errori significativamente maggiore da parte dei parlanti italofoeni nel contrasto accentuale rispetto al contrasto segmentale e al contrasto accento + quantità. I partecipanti germanofoni hanno svolto il test con tutti e tre i contrasti senza difficoltà, ad eccezione del contrasto accentuale con parole contenenti una vocale lunga nella penultima sillaba, in cui la percentuale di errore dei partecipanti germanofoni era paragonabile a quella dei partecipanti italofoeni. Tuttavia, dato il numero ridotto di partecipanti germanofoni, questo risultato potrebbe essere casuale.

I partecipanti italofofoni erano in grado di distinguere l'accento nel test di identificazione preliminare, ma non nel test sopra descritto, che impegnava la memoria a breve termine e presentava variabilità fonetica negli stimoli. Le difficoltà non consistono pertanto nel percepire le differenze acustiche tra gli stimoli, ma sono probabilmente legate alla rappresentazione delle parole nel lessico mentale.

In conclusione, abbiamo effettuato delle registrazioni per replicare il quinto esperimento con non-parole. Infatti è necessario effettuare un confronto tra i parlanti italofofoni e germanofofoni sulla base di parole non tedesche, in modo da non avvantaggiare uno dei due gruppi rispetto all'altro. A tale scopo sono stati registrati due corpus di non-parole, costituite da fonemi appartenenti all'inventario fonemico sia dell'italiano che del tedesco e strutturate secondo regole fonotattiche comuni. Il primo corpus è stato letto da una parlante di madrelingua giapponese e il secondo da una parlante cantonese. In entrambi i casi si trattava di fonetiste esperte, con nessuna conoscenza dell'italiano e conoscenze molto ridotte del tedesco. Sono previste ulteriori registrazioni da parte di un parlante di sesso maschile e l'utilizzo di tali stimoli in un test percettivo simile a quello presentato nel quinto esperimento.

I risultati principali della tesi possono essere riassunti come segue. Abbiamo riscontrato errori sistematici nell'accento lessicale di parole tedesche morfologicamente complesse da parte della quasi totalità dei parlanti italofofoni. Nella maggior parte dei casi si trattava di parlanti del tedesco di livello avanzato, residenti in Germania da diversi anni, che non si erano mai resi conto dell'errore. Tale fatto sembra indicare un'origine percettiva dell'errore, forse dovuta al fatto che i parlanti italofofoni ricollegano la posizione dell'accento alla presenza di vocali lunghe e non necessariamente ad una frequenza fondamentale più alta. Difatti la correzione tramite la sintesi della frequenza fondamentale di parole accentate erroneamente da parlanti italofofoni è risultata più efficace rispetto alla correzione della durata dei fonemi, per cui le produzioni in L2 si differenzia-

vano dalle produzioni corrette più per il primo parametro che per il secondo. Inoltre, nel training di pronuncia, l'ascolto della pronuncia corretta da parte di una parlante di madrelingua tedesca in molti casi non era sufficiente perché i parlanti italofoeni riconoscessero la posizione corretta dell'accento. Era invece necessario fornire agli apprendenti ulteriori indicazioni sotto forma di stimoli dall'accento enfaticizzato oppure di stimoli risintetizzati nella loro voce e pertanto dalla variabilità fonetica ridotta. La sintesi del parlato può essere realizzata senza previa segmentazione del segnale, permettendo così la futura implementazione di questa tecnica in programmi di Computer Assisted Language Learning. Infine, i risultati di un test percettivo sull'accento lessicale, svolto sia da parlanti italofoeni che germanofoeni, indicano che gli errori di accentazione da parte dei parlanti italofoeni possono avere delle cause percettive. Probabilmente infatti, parole tedesche che si differenziano nella posizione dell'accento principale e secondario sono rappresentate dai parlanti italofoeni come omofoni nel lessico mentale.

A zia Grazia

Vorwort

Diese Arbeit entstand im Rahmen einer gemeinsamen Dissertationsbetreuung zwischen der Ludwig-Maximilians-Universität München und der Università degli Studi di Sassari in Italien.

Mein allererster Dank gilt meinen Doktoreltern Prof. Dr. Livia Tonelli und Prof. Dr. Hans G. Tillmann, ohne die diese Arbeit nie zustande gekommen wäre. Ich bin sehr dankbar für ihre Betreuung und Unterstützung und dafür, dass sie mich stets motiviert haben.

Ein besonderer Dank auch an meinen großen Doktorbruder Dr. habil. Hartmut R. Pfitzinger, der mich betreut und mir die Möglichkeit gegeben hat, mit seiner Software und seinem technischen Equipment zu arbeiten. Ganz herzlich möchte ich mich für seine Vorschläge zur Durchführung der Experimente bedanken. Seine Hilfe war unentbehrlich.

Herzlichen Dank an Prof. Dr. Jonathan Harrington, Leiter des Instituts für Phonetik und Sprachverarbeitung der Universität München, für seine bereitwillige Unterstützung meiner Arbeit.

Vielen Dank an Klaus Jänsch für seine Ratschläge und Hilfe bei der Programmierung in Java und für seine technische Unterstützung bei den Aufnahmen im Studio. Ich bedanke mich bei meinem Bruder Dr. Pier Giovanni Bissiri für seine wertvollen Hinweise zur statistischen Datenanalyse. Vielen herzlichen Dank an Georg Happe für die stilistischen Verbesserungen der Form dieser Arbeit.

Ich bedanke mich bei meinen Sprechern Juliana Ammer, Angela Baumann und Ulrich Reubold. Danke an die Studentinnen der Universität Sassari und an alle Probanden in Italien und in Deutschland, die an den Aufnahmen und den Tests teilgenommen haben. Danke an alle Mitarbeiter und Studenten des Instituts für Phonetik und Sprachverarbeitung der Universität München für die Teilnahme an die Perzeptionstests und für die freundliche Unterstützung. Ein spezieller Dank an die Kollegen vom Raum 249, Uwe Reichel, Dr. Katalin Mády und Jan Gorisch, für die nette und inspirierende Arbeitsatmosphäre. Ich bedanke mich insbesondere bei Uwe für seine Teilnahme an den Aufnahmen und für seine ständige Hilfsbereitschaft.

Schließlich möchte ich mich ganz herzlich bei meiner Familie dafür bedanken, dass sie mich stets motiviert und unterstützt haben. An alle herzlichen Dank.

Inhaltsverzeichnis

1	Einleitung	21
1.1	Zielsetzung	21
1.2	Methode	22
1.3	Struktur der Arbeit	24
	Erster Teil: Stand der Forschung und theoretische Grundlagen	25
2	Prosodie, Wortbetonung und Zweitspracherwerb	27
2.1	Die Pluralität der Prosodien	27
2.2	Definition der prosodischen Parameter	30
2.2.1	Lokale Sprechgeschwindigkeit	31
2.2.2	Intonation	31
2.2.3	Intensität	32
2.3	Wortbetonung	32
2.3.1	Akustische Korrelate der Betonung im Deutschen: Der Fall morphologisch komplexer Wörter	33
2.3.2	Interferenzquellen beim Erlernen der deutschen Wortbetonung durch Italiener	35

2.3.3	Wahrnehmung von L2-Kontrasten: eher ein phonologisches als ein akustisches Problem	38
2.3.4	Eine Methode zur Untersuchung von „stress deafness“	40
2.4	Prosodie und Zweitspracherwerb	41
3	Aussprachefehler italienischsprachiger Deutschlerner in der Forschungsliteratur	45
3.1	Segmentale Fehler	46
3.2	Prosodische Fehler	52
4	Computergestütztes Lernen der Aussprache	57
4.1	Computer Assisted Language Learning zum Training der Aussprache	57
4.1.1	Vorzüge von CALL-Systemen zum Erlernen der Aussprache	58
4.1.2	Automatische Spracherkennung in CALL	61
4.1.3	Input und Feedback in CALL	63
4.2	Aussprachetraining der Prosodie	66
4.2.1	Visuelles und auditives Feedback zum computergestützten Erlernen der Prosodie: Ein historischer Überblick	66
	 Zweiter Teil: Experimentelles Vorgehen	 71
5	Datenerhebung	73
5.1	Textmaterial	73
5.2	Sprecher	73
5.3	Aufnahme	75
5.4	Datenbearbeitung	75
5.4.1	Segmentation und Etikettierung	76

6 Aussprachefehler italienischsprachiger Probanden	79
6.1 Segmentale Fehler	79
6.2 Wortbetonungsfehler	83
7 Resynthese	87
7.1 Resynthese der lokalen Sprechgeschwindigkeit	87
7.1.1 Resynthese der perzipierten lokalen Sprechgeschwindigkeit PLSR	88
7.1.2 Resynthese der Dauer der Phone	91
7.2 Resynthese der Intonation	92
7.3 Resynthese der Intensität	93
7.4 Algorithmus zur Übertragung der Prosodie eines L1-Sprechers auf die Äußerung eines L2-Sprechers	93
8 Experimente zum computergestützten Lernen der Wortbetonung im Deutschen bzw. deren Wahrnehmung durch italienische Sprecher	97
8.1 Experiment 1: Einfluss der Korrektur der perzipierten lokalen Sprechgeschwindigkeit auf die Wahrnehmung des Sprechrhythmus	99
8.1.1 Methode	99
8.1.1.1 Stimuli	99
8.1.1.2 Testteilnehmer	102
8.1.1.3 Perzeptionstest	102
8.1.2 Ergebnisse und Diskussion	104
8.1.3 Betrachtungen nach dem Experiment 1	107
8.2 Experiment 2: Einfluss von Intonation und lokaler Sprechgeschwindigkeit auf die Betonung morphologisch komplexer Wörter im Deutschen durch L2-Sprecher	110
8.2.1 Methode	111
8.2.1.1 Stimuli	112

8.2.1.2	Testteilnehmer	114
8.2.1.3	Perzeptionstest zur Betonung morphologisch komplexer Wörter	114
8.2.2	Ergebnisse der statistischen Datenanalyse	115
8.2.3	Betrachtungen nach dem Experiment 2	118
8.3	Experiment 3: Training der Wortbetonung durch Resynthese und emphatische Betonung für italienischsprachige Lernende . .	120
8.3.1	Methode	121
8.3.1.1	Aufnahmen vor dem Training	122
8.3.1.2	Vorbereitung der Stimuli für das Aussprache- training	123
8.3.2	Aussprachetraining	124
8.3.3	Betrachtungen nach dem Experiment 3	127
8.4	Experiment 4: Perzeptionstest zur Evaluierung der Trainingsdaten	129
8.4.1	Methode	129
8.4.1.1	Stimuli	129
8.4.1.2	Testteilnehmer	129
8.4.2	Ergebnisse des Perzeptionstests und Diskussion	131
8.4.3	Betrachtungen nach dem Experiment 4	137
8.5	Experiment 5: Perzeption der Betonung morphologisch komple- xer Wörter im Deutschen durch italienische Sprecher	139
8.5.1	Methode	141
8.5.1.1	Auswahl der Stimuli	142
8.5.1.1.1	Aufnahmeversuche mit deutschen Mi- nimalpaaren	142
8.5.1.1.2	Gewählte Stimuli	144
8.5.1.2	Akustische Untersuchungen der Dauer der Sti- muli	148
8.5.1.3	Testteilnehmer	156

8.5.1.4	Perzeptionstest	157
8.5.1.5	Ergebnisse und Diskussion	161
8.5.1.6	Betrachtungen nach dem Experiment 5	165
8.6	Voruntersuchungen zu einem Test zur Betonungswahrnehmung von Nichtwörtern durch Deutsche und Italiener	166
8.6.1	Methode	167
8.6.1.1	Aufnahme einer japanischen Sprecherin	168
8.6.1.2	Aufnahme einer kantonesischen Sprecherin	170
8.6.2	Planung zur weiteren Durchführung des Experiments 6	172
9	Schlussbetrachtungen	173
9.1	Zusammenfassung der Ergebnisse	173
9.2	Ausblick	176
Anhang		179
A	Textmaterial	181
A.1	Die Buttergeschichte	181
A.2	Sätze aus der Buttergeschichte	182
B	Resynthese	185
B.1	Programm zum Kopieren der Prosodie eines L1-Sprechers auf einen L2-Sprecher	185
C	Experiment 1	189
C.1	Stimuli im Perzeptionstest	189
C.2	Instruktionen für die Probanden	190
D	Experiment 2	193
D.1	Wörter aus der Buttergeschichte	193

D.1.1	Morphologisch komplexe Wörter	193
D.1.2	Komplette Wortliste	193
D.2	Stimuli im Perzeptionstest	195
E	Experiment 5	197
E.1	Deutsche Minimalpaare	197
E.2	Sprecheranweisungen für die Aufnahme der Stimuli für das Experiment 5	197
E.3	Perzeptionstest	198
E.3.1	Anweisungen für die Probanden	198
F	Voruntersuchungen zum geplanten sechsten Experiment	201
F.1	Nichtwörter für die Aufnahme der japanischen Sprecherin	201
F.2	Nichtwörter für die Aufnahme der kantonesischen Sprecherin . .	202
	Literaturverzeichnis	203

1 Einleitung

1.1 Zielsetzung

Ziel dieser Arbeit ist die Erforschung geeigneter didaktischer Methoden, um mit Hilfe des Computers italienischsprachigen Lernenden die Prosodie der deutschen Sprache möglichst effektiv vermitteln zu können.¹

Unsere Untersuchung konzentriert sich hauptsächlich auf die Wortbetonung von morphologisch komplexen Wörtern des Deutschen wie Komposita (z.B. *Lebensmittel*, *Menschenmenge*) und trennbare Verben (z.B. *vorlassen*, *aufschließen*),² die auf ihrem Erstglied betont werden müssen, aber von italienischsprachigen Deutschlernenden häufig fehlerhaft auf ihrem Zweitglied betont werden.

In dieser Arbeit möchten wir die Effektivität zweier möglicher Feedbacks untersuchen, die ein Computerprogramm geben könnte, um italienischen Lernenden die Wortbetonung zu vermitteln:

1. resynthetisierte Stimuli in der Stimme der Lernenden, deren prosodische Parameter (F0, Dauer der Segmente und Intensität) an die entsprechen-

¹Auf die Forschungsziele von Computer Assisted Language Learning (CALL), das sich auf die Verwendung des Computers als Hilfsmittel beim Sprachenlernen bezieht, wird in Kap. 4 eingegangen.

²Die DUDEN Grammatik bezeichnet trennbare Verben als *Partikelverben* und definiert sie als „komplexe Verben mit einem morphologisch und syntaktisch trennbaren Erstglied“ (DUDEN 2006 [36, S.705]). Das Erstglied solcher Verben wird in der Literatur unterschiedlich bezeichnet, u.a. *trennbares Präfix*, *Halbpräfix* oder *Verbpartikel* (für Referenzen s. DUDEN 2006 [36, S.705]).

den Sprachsignale einer deutschen Muttersprachlerin angepasst worden sind, und somit in der richtigen Zielform präsentiert werden können.

2. Stimuli, die überbetont wurden, so dass die Lernenden die betonte Silbe besser erkennen können.

Zur Ermittlung der Effektivität der obengenannten Feedbacks werden diese in einer Trainingsmaßnahme miteinander sowie mit einem „traditionellen“ Feedback, normal betonte Stimuli in der aufgenommenen Stimme eines Muttersprachlers, und mit überbetonten resynthetisierten Stimuli verglichen.

Darüber hinaus möchten wir untersuchen, ob die Schwierigkeiten italienischsprachiger Lernender beim Erlernen der korrekten Betonung morphologisch komplexer Wörter im Deutschen perzeptive Ursachen haben. Wenn das Erstglied eines solchen Wortes betont wird, kann im Deutschen die betonte Silbe des Zweitglieds ihre Quantität bewahren (s. Kohler 1995 [74, S.188]). Dies würde Italienern den Eindruck vermitteln, die zweite Komponente sei betont, da im Italienischen die Dauer das Hauptkorrelat der Betonung ist, und eine nichtbetonte Pänultima normalerweise gekürzt wird (s. Bertinetto 1980 [9] und 1981 [10]).

1.2 Methode

Zunächst wurden die prosodischen Parameter der auf Deutsch gesprochenen Äußerungen der italienischsprachigen Lernenden mittels Resynthese korrigiert. Dafür wurde der deutsche Standardtext „Die Buttergeschichte“ (s. Kohler 2006 [75] und Anhang A.1) von italienischen Sprecherinnen und einer deutschen Muttersprachlerin vorgelesen und aufgenommen. Dann wurden die prosodischen Parameter *Intonation*, *lokale Sprechgeschwindigkeit* und *Intensität* aus den Sprachsignalen der deutschen Muttersprachlerin auf die Signale der italie-

nischen Deutschlernerinnen kopiert.

Im Folgenden untersuchten wir mittels zweier Perzeptionstests, welchen Einfluss die synthetische Korrektur der prosodischen Parameter auf die Beurteilung der Aussprache durch deutsche Muttersprachler hatte. Zu diesem Zweck wurde zunächst untersucht, wie die Resynthese der perzipierten lokalen Sprechgeschwindigkeit (s. Pfitzinger 2001 [116]) als einziger Parameter die Beurteilung des Sprechrhythmus beeinflusste. Nachfolgend haben wir Stimuli erzeugt, die hinsichtlich der Grundfrequenz und hinsichtlich der Dauer der Segmente korrigiert wurden, um den Einfluss der einzelnen Parameter auf die Betonung morphologisch komplexer Wörter zu testen.

Die mittels dieser Methode resynthetisierten Stimuli verwendeten wir in einem Aussprachetraining zur Wortbetonung für italienischsprachige Deutschlernende. Dafür korrigierten wir mittels Resynthese nicht nur die akustischen Parameter Grundfrequenz und Segmentdauer, sondern auch die Intensität in den Äußerungen von männlichen und weiblichen italienischen Sprechern. Um auch die Wirkung der emphatischen Betonung als Feedback zu testen, wurden Stimuli kreiert, in denen die prosodischen Parameter überbetonter Wörter auf die Sprachsignale der italienischen Sprecher kopiert wurden. Im Training wurden einer Gruppe italienischer Sprecher resynthetisierte Stimuli und einer anderen Gruppe natürliche Stimuli als Feedback angeboten. Beide Gruppen erhielten in einer ersten Phase des Trainings normal betonte Stimuli und in einer zweiten Phase überbetonte Stimuli.

Die Betonung der Wörter in den Äußerungen der italienischen Sprecher vor und nach dem Training wurde dann in einem Perzeptionstest von deutschen Muttersprachlern und Muttersprachlerinnen evaluiert.

Abschließend untersuchten wir in einem Perzeptionstest, ob die falsche Wortbetonung bei italienischsprachigen Deutschlernenden perzeptive Ursachen hat und nicht etwa durch die Übertragung des im Italienischen gängigen Beto-

nungsmusters, nach dem die Pänultima zu betonen ist, auf das Deutsche verursacht wird.

1.3 Struktur der Arbeit

Im ersten Teil dieser Arbeit wird eine knappe Zusammenfassung der einschlägigen Literatur gegeben. Zunächst werden der Begriff Prosodie und die in unserer Untersuchung resynthetisierten prosodischen Merkmale in unserem Sinne definiert. Dieser Teil schließt mit einem Überblick der Literatur zum computergestützten Lernen der Prosodie.

Im zweiten Teil der Arbeit werden die von uns durchgeführten Experimente vorgestellt. Zunächst beschreiben wir die Datenerhebung, die Segmentation und Etikettierung und dann die Analyse der Fehler in den Äußerungen der von uns aufgenommenen italienischsprachigen Probandinnen. Anschließend wird die Resyntheseprozedur erklärt. Es folgt die Beschreibung von fünf Experimenten und der Voruntersuchungen zu einem sechsten. Das erste Experiment befasst sich mit dem Einfluss der Korrektur der perzipierten lokalen Sprechgeschwindigkeit auf die Perzeption des Sprechrhythmus, und das zweite mit dem Einfluss der lokalen Sprechgeschwindigkeit und der Intonation auf die Wortbetonung. Das dritte Experiment besteht aus einem Training der Wortbetonung für italienischsprachige Deutschlernende und das vierte aus einem Perzeptionstest zur Evaluierung der Trainingsdaten. Im fünften Experiment untersuchen wir die Perzeption der Betonung von morphologisch komplexen Wörtern im Deutschen durch italienische Sprecher. Zum Abschluss des Experimentalteils werden Voruntersuchungen zu einem weiteren geplanten Experiment zur Betonungswahrnehmung der Italiener vorgestellt. Im letzten Kapitel werden die Ergebnisse zusammengefasst und ein kurzer Ausblick auf mögliche künftige Untersuchungen gegeben.

Erster Teil:
Stand der Forschung und
theoretische Grundlagen

2 Prosodie, Wortbetonung und Zweitspracherwerb

2.1 Die Pluralität der Prosodien

Als Prosodie bezeichnen wir alle phonemübergreifenden Eigenschaften der gesprochenen Sprache, die sich als Zeitfunktion ändern. Der Begriff der Prosodie wird häufig als Synonym für Intonation verwendet (oder umgekehrt). Firth 1948 [43] vertritt hingegen eine breitere Auffassung von Prosodie, die unserem Verständnis näher kommt. Er stellt die Prosodien als parallel zur phonematischen Struktur der Sprache verlaufende Spuren dar (s. Beispieldarstellung in Firth 1948 [43, S.151-152]) und befürwortet eine Analyse, die die Interaktionen zwischen segmentalen und suprasegmentalen Eigenschaften der gesprochenen Sprache betrachtet: „The phonological structure of the sentence and the words which comprise it are to be expressed as a plurality of systems of interrelated phonematic and prosodic categories.“ (Firth 1948 [43, S.151]). Laut Firth gehören zu den Prosodien auch Silbenstruktur [43, S.129 und 133], Vokalharmonie [43, S.146], durch Junktur der Silben bedingte Phänomene wie Geminatio (*junctional gemination*) [43, S.130], Aspiration und Glottalverschluss in mehreren Sprachen¹ [43, S.132] und Schwa-Elision [43, S.147].

¹Den Glottalverschluss bezeichnet Firth als *junction prosody* des Deutschen (s. Firth 1948 [43, S.143-144]).

„Den dynamischen Ablauf von lautsprachlichen Äußerungen“ nennt Tillmann in Anlehnung an Firth „ihre *Prosodie*“ (s. Tillmann 1980 [139, S.109]). Tillmann formulierte in seiner Dissertation seine Theorie des Ausprägungskodes (s. Tillmann 1964 [137]), die er acht Jahre später weiterentwickelte (s. Tillmann 1972 [138]). Nach der Theorie des silbischen Ausprägungskodes — auch Reduktionskode genannt (Tillmann 1972 [138, S.261]) — hat jeder Silbenkomplex „eine Skala von Ausprägungsstufen in der lautlich-segmentellen Struktur“, und „der Ausprägungsgrad variiert innerhalb einer Äußerung, von Silbe zu Silbe“ (Tillmann 1972 [138, S.263]). Aus diesem Grund kann man von einem Ausprägungsverlauf — oder Reduktionsverlauf — sprechen, den Tillmann als *silbisches Ausprägungsprofil* definiert (Tillmann 1972 [138, S.263]). Die Theorie des Ausprägungskodes ist relevant für die Phonetik des Zweitspracherwerbs:

Der kompetente Sprecher beherrscht den Ausprägungskode seiner Sprache. Der fremde Akzent des Ausländers verrät sich nicht zuletzt am falschen Ausprägungskode, z.B. durch fehlende oder falsche Reduktionen. Die Phonetik einer Sprache sollte darum die systematisierbaren Eigenschaften des Ausprägungskodes berücksichtigen. (Tillmann 1972 [138, S.264]).

Tillmann bekräftigt die Abhängigkeit der Ausprägung der Laute von Silbenkomplexität und -dauer und dadurch von der Redegeschwindigkeit, gemessen in Silben pro Zeiteinheit (Tillmann 1972 [138, S.263]). Zudem ist laut Tillmann „der reguläre Verlauf der Ausprägungsgrade mit dem Intonationsverlauf koordiniert“ (Tillmann 1972 [138, S.264]). Insbesondere mit dieser Theorie über die Zusammenhänge zwischen Intonationsverlauf, Sprechgeschwindigkeit und Ausprägungsprofil leistete Tillmann einen bemerkenswerten Beitrag zum damals aktuellen Forschungsstand. Umso mehr verwundert, dass seine These bis heute noch immer nicht durch umfangreichere prosodische Untersuchungen verifiziert worden ist.

Auch in seinem späteren Werk über die Phonetik deutet Tillmann eine Interaktion zwischen unterschiedlichen Prosodien an (s. Tillmann 1980 [139]). Er führte 1980 [139, S.39f, S.108ff] die Begriffe *A-*, *B-* und *C-Prosodie* ein. Die Art der Prosodie wird durch die Geschwindigkeit bestimmt, in der sich Klangeigenschaften wie Lautstärke, Tonhöhe bzw. Klangfarbe ändern. Tillmann erklärt diese Konzepte am Beispiel einer Klangeigenschaft eines Tones, deren Kontur eine sinusförmige Oszillation beschreibt (s. Tillmann 1980 [139, S.39]). Im *Fall A* handelt es sich um eine sich langsam ändernde Kontur. In Tillmanns Beispiel wird die Klangeigenschaft mit einer Geschwindigkeit von 0.5 Hz moduliert (s. Tillmann 1980 [139, S.39: Abb.5]). Der Klang wird als sich langsam kontinuierlich verändernd wahrgenommen. Im *Fall B* ändert sich die Kontur schneller, zum Beispiel mit einer Modulationsfrequenz von 2 Hz in Tillmanns bereits erwähntem Beispiel. So wird der Klang als eine Folge von rhythmischen Schlägen wahrgenommen. Bei einer noch höheren Modulationsfrequenz, wie 10 Hz, wird ein Vibrant wahrgenommen, dessen Schläge nicht mehr zählbar sind. Dies ist der *Fall C*. Die *A-Prosodie* betrifft also die sich langsam ändernden Merkmale der Sprachsignale, wie die Intonation einer Äußerung, die mehrere Silben und Wörter umfasst. Die *B-Prosodie* betrifft Änderungen auf Silbenebene, wie die Wortbetonung. Nicht nur die Vibranten, sondern alle Konsonantenbewegungen bestimmen die *C-Prosodie*.

Die Auffassung mehrerer lokal variierender und zugleich miteinander interagierender Prosodien wird auch von Pfitzinger vertreten. Vom Standpunkt der Sprachsynthese aus betrachtet, kann die Manipulation eines einzigen prosodischen Merkmals — ohne entsprechende Anpassung anderer akustischer Merkmale — zu resynthetisierten Stimuli mit widersprüchlicher Prosodie führen (s. Pfitzinger 2006 [117]). Pfitzinger nennt fünf Dimensionen der Prosodie: Intensität, Intonation, Sprechgeschwindigkeit, Stimmqualität und Reduktionsgrad. Die ersten drei Dimensionen sind traditionell bekannt. Die vierte, die Stimmqualität, wurde von Campbell & Mokhtari 2002 [11] vorgeschlagen. Pfit-

zinger hat die fünfte Dimension hinzugefügt, die des Reduktionsgrades, die mit dem Begriff des *Reduktionskodes* von Tillmann (s. oben in diesem Kapitel und Tillmann 1972 [138]) in Verbindung steht. Laut Pfitzinger hat jede der fünf Prosodien sowohl suprasegmentale als auch segmentale Auswirkungen auf die gesprochene Sprache (s. Pfitzinger 2006 [117, S.8]). Dies steht im Einklang mit der breiten Auffassung der Prosodie von Firth und mit der Bezeichnung der Konsonantenbewegungen als C-Prosodie von Tillmann. Zudem präzisiert Pfitzinger seine Aussage, indem er über Komponenten der Prosodie mit niedriger und mit hoher Frequenz spricht: „In this view, even the articulatory movements, and thus every detail that constitutes speech, could become prosodies.“ (Pfitzinger 2006 [117, S.8])

In seinem Aufsatz beschreibt Pfitzinger den Stand der Forschung zu Analyse, Manipulation und Synthese der fünf prosodischen Parameter. Einige der fünf Parameter sind mit heutigen Methoden schwierig zu messen oder zu synthetisieren, insbesondere Stimmqualität und Reduktionsgrad erweisen sich als problematisch (s. Pfitzinger 2006 [117, S.7]). Aus diesem Grund haben wir in dieser Arbeit nur die übrigen prosodischen Parameter in den Sprachsignalen der italienischsprachigen Deutschlernenden modifiziert: die lokale Sprechgeschwindigkeit, die Intonation und die Intensität (s. Kap. 7 zur Resynthese der Parameter).

2.2 Definition der prosodischen Parameter

Im Folgenden geben wir eine kurze Definition der von uns in dieser Arbeit manipulierten prosodischen Parameter: lokale Sprechgeschwindigkeit, Intonation und Intensität.

2.2.1 Lokale Sprechgeschwindigkeit

Mit lokaler Sprechgeschwindigkeit (vgl. Pfitzinger 2001 [116, S.139]) wird die Sprechgeschwindigkeit bezeichnet, die sich bereits von Silbe zu Silbe, und sogar auch innerhalb von Silben verändern kann.

Pfitzinger entwickelte ein Modell zur Vorhersage der perzipierten lokalen Sprechgeschwindigkeit (PLSR: *perceptual local speech rate*) im Deutschen auf Basis von Silben- und Phonrate (s. Pfitzinger 2001 [116] und 1999 [114]). In Fenstern von 625 ms wird die Silben- und Phonrate anhand der manuell durchgeführten Segmentation berechnet. Die Größe des Analysefensters wurde mittels mehrerer Perzeptionsexperimente bestimmt. Das Modell zur Berechnung der PLSR ist folgendes (s. Pfitzinger 2001 [116, S.208, Abb.8.3]):

$$\text{PLSR-Modell: } 7.834 * \text{Silbenrate} + 4.060 * \text{Phonrate} - 1.41$$

Durch dieses Modell ist es möglich, zu jeder segmentierten Äußerung eine Sprechgeschwindigkeitskontur mit Maxima und Minima zu berechnen (s. Abb. 7.1 auf S.90), die wie eine Intonationskontur den prosodischen Verlauf darstellt.

In unserem Experiment 1 (s. Kap. 8.1) haben wir die Resynthese der lokalen Sprechgeschwindigkeit durch die Manipulation der gemessenen PLSR-Kontur durchgeführt. Ab Experiment 2 haben wir stattdessen die Dauer der Phone mittels eines *dynamic-time-warping*-Algorithmus (DTW) ohne Handsegmentation resynthetisiert (s. Kap. 7.1 zur Resynthese der lokalen Sprechgeschwindigkeit).

2.2.2 Intonation

Wir bezeichnen mit Intonation den Grundfrequenzverlauf im Sprachsignal, genauer: das akustische Korrelat der quasiperiodischen Schwingungen der

Stimmlippen (s. Lehiste 1970 [82, S.60ff]).

2.2.3 Intensität

Die Intensität ist ein akustisches Korrelat der Lautstärke. Die Schallintensität wird als die Schalleistung definiert, die durch eine Schallwelle auf eine gegenüber der Schallrichtung befindliche senkrechte Fläche von einem Quadratzentimeter übertragen wird (s. Lehiste 1970 [82, S.111]).

Zur Intensitätsanalyse haben wir die Einhüllende der Amplitude (die maximale Amplitude als Zeitfunktion) berechnet.

2.3 Wortbetonung

Mit dem Begriff *Betonung* bezeichnen wir in dieser Arbeit die Hervorhebung von Silben auf Wortebene — auf Englisch *stress* — im Gegensatz zu *Akzent*, die Hervorhebung von Wörtern auf Äußerungsebene — auf Englisch *accent* (s. Harrington & Cassidy 1999 [61, S.111f]).

Die Betonung ist ein komplexes Phänomen und hat mehrere akustische Korrelate wie Phondauer, Grundfrequenz, Intensität, Vokalqualität und spektrale Balance (s. Sluijter & van Heuven 1996 [135]). Die Relevanz der einzelnen Korrelate ist von Sprache zu Sprache oft unterschiedlich.

Die traditionelle Sichtweise, dass in akzentzählenden Sprachen die F₀-Änderung das Hauptkorrelat der Betonung sei (s. Fry 1955 [46] und 1958 [47] und Dauer 1983 [22]), rührt laut Sluijter & van Heuven 1996 [135] daher, dass Wörter in Fokusposition untersucht wurden. Laut Sluijter & van Heuven ist die F₀-Änderung hingegen eher akustisches Korrelat des Fokus als der Betonung.

2.3.1 Akustische Korrelate der Betonung im Deutschen: Der Fall morphologisch komplexer Wörter

Die bisher durchgeführten Untersuchungen über die akustischen Korrelate der Wortbetonung im Deutschen variieren hinsichtlich Methode, Stimulusmaterial und Ergebnissen. Den meisten Studien zufolge ist die Dauer das Hauptkorrelat der Betonung im Deutschen (s. Jessen, Marasek, Schneider & Clahßen 1995 [67], Rapp 1995 [122] und Dogil & Williams 1999 [33]).² Doch eigentlich testeten die erwähnten Untersuchungen Nichtwörter oder Simplizia und keine morphologisch komplexen Wörter, die sich hinsichtlich der primär- und sekundär betonten Silbe unterscheiden. Dogil & Williams 1999 [33] zitieren Goldbeck & Sendlmeier 1988 [52], die die Korrelate der Betonung von zwei Präfixverben untersuchten, als eine Studie, die demonstriert, dass Dauer das Hauptkorrelat der Betonung im Deutschen ist. Jedoch sind die Ergebnisse bei Goldbeck & Sendlmeier nicht so eindeutig. Aufgezeichnet wurden die Betonungsminimalpaare der beiden zweisilbigen Wörter *Tenor* und *modern* und der Präfixverben *umfahren* und *übersetzen*, die von zwei Sprechern jeweils vier mal ausgesprochen wurden. Die Autoren maßen F0 und Silbendauer der in deklarativen Sätzen eingebetteten Präfixverben *umfahren* und *übersetzen* und fanden eine höhere F0 in den betonten Silben, jedoch keine Unterschiede in der Dauer. Bei den in Interrogativsätzen eingebetteten Präfixverben gab es ebenfalls keine Dauerunterschiede zwischen betonten und unbetonten Silben, dafür aber einen unterschiedlichen Grundfrequenzverlauf. Bei den zweisilbigen Wörtern *tenor* und *modern* konnte hingegen eine längere Dauer der betonten Silben sowohl in den Deklarativ- als auch in den Interrogativsätzen nachgewiesen werden. Mittels der akustischen Messungen konnte also ein Dauerun-

²Jessen, Marasek, Schneider & Clahßen 1995 [67] untersuchten isoliert stehende Wörter. Rapp 1995 [122] und Dogil & Williams 1999 [33] untersuchten in Sätzen eingebettete Wörter sowohl in Fokus- als auch in Nichtfokusposition. Sie fanden heraus, dass — unabhängig vom Fokus — die Dauer das Hauptkorrelat der Betonung der Wörter war.

terschied zwischen betonten und unbetonten Silben bei den Simplicia, jedoch nicht bei den Präfixverben, festgestellt werden.

Da sich die Korrelation zwischen F0 und Betonung als so eindeutiger Befund der akustischen Messungen der Deklarativsätze erwiesen hatte, entschieden sich die Autoren, nur die interrogativen Sätzen für einen Perzeptionstest zu verwenden (Goldbeck & Sendlmeier 1988 [52, S.313]). Resynthetisierte Stimuli wurden geschaffen, indem die akustischen Parameter jedes Wortes auf das andere Wort des Minimalpaars kopiert wurden. Da die betonten und die unbetonten Silben in *umfahren* und *übersetzen* die gleiche Dauer hatten, und sich daher nichts verändert hätte, wenn die Dauer der Silben eines Wortes auf das andere Wort kopiert worden wäre, wurde die Dauer der Silben halbiert bzw. verdoppelt. Aus diesem Grund und angesichts der Beschränkungen des Stimulusmaterials auf Interrogativsätze zum einen, sowie des Wortinventars auf nur zwei Verben zum anderen, kann man aus den Ergebnissen dieses Experiments nicht verallgemeinernd schließen, dass die Dauer das perzeptive Hauptkorrelat der Betonung für die Präfixverben sei.

Für die Dreisilber [*umfahren*] liefern die Daten ebenfalls Evidenz für die perzeptive Relevanz des Faktors Dauer; hierbei ist jedoch zu berücksichtigen, daß die akustische Analyse für die Drei- und Viersilber ursprünglich keine Dauerunterschiede der in Opposition stehenden Silben aufwies. Daher sollten die Faktoren F0-Anstiegszeitpunkt und Intensitätsmaximum bei der Interpretation dieser Stimuli in stärkerem Maße als relevante Parameter für die Wahrnehmung der Akzentposition betrachtet werden. (Goldbeck & Sendlmeier 1988 [52, S.320])

Auch Isačenko & Schädlich 1966 [65] manipulierten das Minimalpaar *über'setzen* 'übersetzen'. Zunächst wurden die zwei Wörter durch Konstanthaltung ihrer Grundfrequenz monotonisiert. 82 bzw. 96 Prozent der Versuchsp-

sonen in einem Perzeptionstest waren trotzdem in der Lage, die betonte Silbe zu erkennen. Den Autoren zufolge war der Grund dafür eine Hervorhebung der Intensität, die in den Stimuli noch vorhanden war. Dann manipulierten sie den F0-Verlauf jedes Wortes, schafften damit zwei F0-Stufen und vergaben die höhere Stufe an die Silbe *ü* des Wortes *über'setzen* und an die Silbe *set* in *'übersetzen*. Für 85 Prozent der Versuchspersonen war in dem manipulierten Stimulus aus *über'setzen* die erste Silbe betont, und für 98 Prozent war in dem manipulierten Stimulus aus *'übersetzen* die dritte Silbe betont (s. Isačenko & Schädlich 1966 [65, S.22]). Offensichtlich war die alleinige Manipulation der Grundfrequenz ausreichend, um das Betonungsmuster umzukehren.

Nach unserer Meinung sollten künftige Studien über die akustischen Korrelate der Betonung im Deutschen Wörter mit unterschiedlichen morphologischen Strukturen differenziert betrachten. nicht reduziert werden und sich dadurch von den Silben mit Hauptbetonung möglicherweise hinsichtlich der Grundfrequenz unterscheiden. Die Betonung eines von mehreren Morphemen ist womöglich ein ähnliches Phänomen wie die Hervorhebung eines Wortes in einem Satz, die durch einen Grundfrequenzgipfel ausgedrückt wird.

2.3.2 Interferenzquellen beim Erlernen der deutschen Wortbetonung durch Italiener

Sowohl im Italienischen als auch im Deutschen ist — anders als zum Beispiel im Französischen — die Wortbetonung phonologisch kontrastiv. Deswegen könnte verwundern, dass italienische Sprecher Schwierigkeiten beim Lernen der Betonung im Deutschen zeigen. Die italienische Sprache weist mehrere Betonungsminimalpaare auf.³ Da jedoch die Darstellung der Betonung sprach-

³Beispiele von Betonungsminimalpaaren im Italienischen aus Malagoli 1946 [87, S.23-30]: *'abitino* (konjugierte Form von *abitare* „wohnen“) vs. *abi'tino* (Diminutiv von *'abito* „Kleid“), *'compito* („Aufgabe“) vs. *com'pito* („höflich“), *'impari* („ungleich“) vs. *im'pari*

spezifisch ist, können Schwierigkeiten bei der Betonungswahrnehmung in der L2 auftreten, wenn etwa in der L2 die prosodischen Parameter durch die Betonung unterschiedlich ausgeprägt werden. Die Dauer spielt im Italienischen als akustisches und perzeptives Korrelat der Betonung offensichtlich eine größere Rolle als im Deutschen. Tatsächlich ist im Italienischen die Dauer das akustische Hauptkorrelat der Betonung, selbst wenn die untersuchten Wörter im Fokus stehen. Zudem ist für Italiener die Dauer das Hauptkorrelat bei der Betonungswahrnehmung (s. Bertinetto 1980 [9]). Da das Italienische keine Quantitätsopposition der Vokale kennt, könnten Langvokale von italienischen Sprechern fälschlicherweise als betont wahrgenommen werden.

Dieser Unterschied kann am Beispiel des dreisilbigen italienischen Wortes *subito* und des unfest zusammengesetzten deutschen Verbs *abfahren* aufgezeigt werden. Ein italienischer Sprecher kann die erste (*'subito*) oder die zweite Silbe (*su'bito*) dieses Wortes betonen. Wenn die erste Silbe (*'su-*) betont wird, wird sie normalerweise länger, während die zweite Silbe (*-bi-*) kürzer werden kann.

Wenn deutsche Muttersprachler dagegen die erste Silbe in einem deutschen dreisilbigen morphologisch komplexen Wort betonen, kann diese Silbe eine höhere Grundfrequenz haben, muss aber nicht notwendigerweise lang sein. Die zweite Silbe kann eine niedrigere Grundfrequenz haben, aber auch lang bleiben. Im Verb *abfahren* [*'apfa:rən*] ist der Vokal im Präfix kurz, während der in der zweiten Silbe lang auszusprechen ist. Trotzdem ist der erste Vokal betont. Insbesondere in solchen Wörtern kann die Betonungsstelle von italienischsprachigen Deutschlernenden missverstanden werden.

In einem morphologisch komplexen Wort, in dem das Erstglied die Hauptbetonung trägt, bewahrt laut Kohler die betonte Silbe des Zweitglieds ihre Quantität:

(konjugierte Form von *imparare* „lernen“, *'leggere* („lesen“) vs. *leg'gere* (fem. pl. vom Adjektiv *leggero* „leicht“) und *'tendine* („Sehne“) vs. *ten'dine* („Gardinen“).

Im zusammengesetzten Wort bleiben entweder die einzelnen Wortakzente erhalten, oder sie ordnen sich als Sekundärakzente einem einzigen von diesen unter, der damit zum *Kompositionsakzent* wird und die prominenteste Silbe hervorruft. Die Prominenzabstufungen der übrigen Silben ergeben sich vor allem aus der Bewahrung der Vokalqualitäten und -quantitäten in den sekundären Wortakzentsilben. (Kohler 1995 [74, S.188])

Möglicherweise wird in diesem Fall die Betonung durch eine höhere Grundfrequenz ausgedrückt, während italienischsprachige Lernende die Hauptbetonung auf dem Zweitglied orten könnten. Der ebenfalls denkbare Fall, dass die betonte Silbe des Zweitglieds ihre Betonung ganz verliert (z.B. durch „stress clash“) wird von Kohler hier nicht berücksichtigt. Jedoch wäre es selbst in diesem Fall möglich, dass Italiener die Betonung auf einer unbetonten Pänultima, die nicht signifikant kürzer als die betonte Silbe ist, lokalisieren würden.

Zudem zeigten Untersuchungen, dass Sprecher einer Sprache mit einer bevorzugten Wortbetonungsstelle die Tendenz haben, die Betonung an dieser Stelle wahrzunehmen (s. Berinstein 1979 [6]). Berinstein untersuchte 1979 [6], ob K'ekchi⁴, Englisch- und Spanischsprecher eine bestimmte Betonungsstelle bevorzugen. Im Englischen liegt die Betonung häufiger auf der ersten Silbe und im Spanischen, wie im Italienischen, auf der Pänultima. Berinstein fand heraus, dass, wenn dieselbe Silbe /bi/ dreimal hintereinander wiederholt wurde, Englischsprachige erwartungsgemäß dazu tendierten, die erste und die K'ekchi Sprecher die letzte Silbe als betont wahrzunehmen. Demgegenüber bevorzugten Spanischsprechende keine bestimmte Betonungsstelle. Die Autorin behauptet, dies sei nicht überraschend, da es im Spanischen mehrere Betonungsminimalpaare gibt, denn „stress is in part morphologically determined“ (s. Berinstein 1979 [6, S.22f]). Dasselbe gilt für das Italienische, und deswegen scheint die

⁴K'ekchi ist eine Maya-Sprache aus Guatemala mit fester Betonung auf der letzten Silbe.

Bevorzugung einer bestimmten Betonungsstelle bei Italienern weniger wahrscheinlich. Uns sind keine spezifischen Untersuchungen zu diesem Phänomen bei Italienischsprachigen bekannt, so dass dieser Aspekt ein künftiges Forschungsthema werden könnte.

2.3.3 Wahrnehmung von L2-Kontrasten: eher ein phonologisches als ein akustisches Problem

Laut Pallier, Colomé & Sebastián-Gallés 2001 [111] ordnen Lerner, selbst wenn sie in der Lage sind, die akustischen Unterschiede zwischen L2-Kontrasten wahrzunehmen, diese der selben Kategorie im mentalen Lexikon zu. Die Autoren führten ein Priming-Experiment⁵ mit bilingualen Sprechern des Spanischen und des Katalanischen durch. Die Teilnehmer sprachen ihre L2 — Spanisch bzw. Katalanisch — fließend, denn sie hatten sie in der Kindheit spätestens ab dem sechsten Lebensjahr gelernt. In einem Worterkennungstest wurden den Probanden Minimalpaare präsentiert, die sich hinsichtlich eines katalanischen Phonemkontrasts unterschieden. Die spanischen und nicht die katalanischen Muttersprachlern hatten nach dem zweiten Element eines Minimalpaars kürzere Reaktionszeiten, das heißt, dass sie die Wörter des Minimalpaars als Homophone kategorisierten. Aus den Studienergebnissen schließen die Autoren: „although acoustic information may be available to the speech perception system, it does not percolate into the lexicon“ (Pallier, Colomé & Sebastián-Gallés 2001 [111, S.448]).

Das Phänomen gilt sowohl für segmentale als auch für suprasegmentale

⁵Unter *priming* verstehen wir in diesem Zusammenhang „Voraktivierung“. Durch die Untersuchung von Priming-Effekten mittels Worterkennungsexperimenten kann man herausfinden, ob Probanden bestimmte Wörter als Homophone kategorisieren. Die Probanden erkennen ein Wort schneller, wenn sie dasselbe Wort bzw. ein Wort, das für sie ein Homophon ist, vorher im Test schon gehört haben (s. Cutler & van Donselaar 2001 [21, S.178] und Pallier, Sebastián-Gallés & Colomé 1999 [110]).

Kontraste. Der Umstand, dass die Betonung im Französischen anders als im Spanischen kontrastiv ist, veranlasste Dupoux, Pallier, Sebastián-Gallés & Mehler 1997 [37] zu untersuchen, ob französische Sprecher im Vergleich zu spanischen Sprechern Probleme bei der Unterscheidung von Betonungsmustern haben. Die Ergebnisse ihrer Untersuchung zeigten, dass französische Sprecher Betonungsminimalpaare identifizieren konnten, wenn sie in einem AX-Diskriminationstest mit Stimuli derselben Sprecherin konfrontiert wurden. Probleme hatten französische Sprecher dagegen mit einem ABX-Diskriminationstest, in dem die Stimuli von unterschiedlichen Sprechern ausgesprochenen wurden. Offensichtlich bezog der ABX-Test mehr phonetische Variabilität ein und beanspruchte das Kurzzeitgedächtnis stärker.

This suggests two things. First, unlike what happens with the perception of consonants for which within-category discrimination is very difficult, French listeners can use the acoustic stress cues in order to perform standard discrimination tasks flawlessly. Second, French listeners are nevertheless „deaf“ to stress contrasts at a more abstract processing level, which is revealed only with tasks that are more demanding as far as memory and perceptual resources are concerned. (Dupoux, Peperkamp & Sebastián-Gallés 2001 [38, S.1607])

Die Schwierigkeit italienischsprachiger Deutschlernender, die korrekte Betonung von morphologisch komplexen Wörtern im Deutschen zu lernen, könnte also eher ein phonologisches als ein akustisches Problem sein. Für ein Experiment zur Wahrnehmung der Wortbetonung morphologisch komplexer Wörter durch italienische Sprecher wäre daher notwendig, eine anspruchsvollere Aufgabe anstelle eines einfachen Diskriminationstests zu verwenden.

2.3.4 Eine Methode zur Untersuchung von „stress deafness“

Dupoux, Peperkamp & Sebastián-Gallés entwickelten 2001 [38] eine Methode, mit der sie die von ihnen so genannte „*stress deafness*“ bei französischen Hörern untersuchten.⁶

Bereits 1997 hatten Dupoux et al. [37] herausgefunden, dass Franzosen Schwierigkeiten haben, zwischen Betonungsmustern zu unterscheiden (s. dazu Kap. 2.3.3). Die Autoren betrachteten aber ihren ABX-Diskriminationstest als nicht robust genug, da die Perzeptionsergebnisse eine zu hohe individuelle Variabilität zwischen den Versuchspersonen zeigten, und die Leistung der französischen Hörer signifikant besser als der Zufall war (s. Dupoux, Peperkamp & Sebastián-Gallés 2001 [38, S.1607]).

Dupoux, Peperkamp & Sebastián-Gallés 2001 [38] entwickelten daher eine robustere Methode, die aus einem Kurzgedächtnistest mit Sequenzen von zwei bis sechs Nichtwörtern besteht. Dafür variierten sie in fünf Experimenten die Kurzzeitgedächtnisbelastung und die phonetische Variabilität der Stimuli, um ihren Einfluss auf die Fähigkeit französischer Hörer, die Betonung zu diskriminieren, zu testen. Mit ihrer Methode, die auf ihren Experimenten 3 und 4 basiert, werden die Versuchspersonen gezwungen, die Stimuli phonologisch zu kategorisieren, da ihnen die Möglichkeit nicht gegeben wird, akustische Strategien oder echoisches Gedächtnis in ihrer Aufgabe zu verwenden. Als Baseline wurde derselbe Test mit einem Phonemminimalpaar durchgeführt, und die Kontrollgruppe bestand aus spanischen Hörern.

In einer ersten Trainingsphase lernten die Versuchspersonen bei Dupoux, Peperkamp & Sebastián-Gallés 2001 [38], zwei unterschiedlich betonte Nichtwörter wie *mípa* vs. *mipá* mit den Tasten 1 und 2 des Computers zu assozi-

⁶Es bietet sich an, „*stress deafness*“ als „Betonungstaubheit“ zu übersetzen.

ieren. In der zweiten Phase, der eigentlichen Testphase, hörten sie Sequenzen der Nichtwörter von zwei bis sechs Elementen und mussten beim Tippen der entsprechenden Tasten die Sequenz in der richtigen Reihenfolge reproduzieren. Um phonetische Variabilität zu gewährleisten, wurde mittels Resynthese die Grundfrequenz der Stimuli variiert, und im Experiment 4 wurden dazu die Stimuli von einem weiblichen und von einem männlichen Sprecher ausgesprochen. Ihre Experimente 3 und 4 zeigten keine überlappende Verteilung der Einzelergebnisse der zwei Grundgesamtheiten der französischen und der spanischen Hörer. Im Test mit dem Betonungsminimalpaar war die Leistung der Franzosen nicht besser als der Zufall in den Sequenzen von fünf und sechs Elementen des Experiments 4 (mit den von dem männlichen und dem weiblichen Sprecher ausgesprochenen Stimuli), und sie war fast signifikant in den 6er Sequenzen des Experiments 3 (mit den Stimuli ausschließlich vom weiblichen Sprecher). Es konnte also nachgewiesen werden, dass es „stress deafness“ bei französischen Hörern gibt.

2.4 Prosodie und Zweitspracherwerb

L2-Lerner — wie man weiß — übertragen nicht nur die segmentalen sondern auch die prosodischen Eigenschaften ihrer Muttersprache auf die L2 (Munro 1995 [103], Tajima, Port & Dalby 1997 [136], Missaglia 1999 [95]). Die daraus resultierenden prosodischen Fehler lassen nicht nur den muttersprachlichen Akzent der L2-Lerner erkennen, sondern sie verringern auch die Verständlichkeit der produzierten Äußerungen (Tajima, Port & Dalby 1997 [136]). Die Relevanz der Prosodie bei der Perzeption des fremden Akzents durch Muttersprachler wurde bereits häufig festgestellt.⁷ Nach Anderson-Hsieh, Johnson & Koehler 1992 [4] spielt die Prosodie verglichen mit den segmentalen Aspekten

⁷Anderson-Hsieh, Johnson & Koehler 1992 [4], Anderson-Hsieh & Koehler 1988 [5], James 1976 [66] und Johannson 1978 [69].

und der Silbenstruktur die größte Rolle in der Beurteilung des Akzents von L2-Sprechern durch Muttersprachler. Für ihre Studie wurden 60 Aufnahmen, die ursprünglich für den *SPEAK Test*⁸ verwendet wurden, untersucht. Die Aufnahmen bestanden aus einem kurzen englischen Text (s. Anderson-Hsieh, Johnson & Koehler 1992 [4, S.554]), der von 60 L2-Sprechern des Englischen unterschiedlicher Herkunft gelesen worden war.⁹ Die Gesamtaussprache des Textes wurde von drei Englischlehrern bewertet, und die Prosodieanalyse wurde ebenso wie die Analyse der segmentalen Fehler von drei Phonetikern durchgeführt. Die Ergebnisse der Untersuchung zeigten, dass die Aussprachebewertungen am stärksten mit den Prosodiebewertungen korrelierten. Mit Blick auf künftige Forschungen formulierten die Autoren die Forderung, durch Resynthese die prosodischen Merkmale getrennt und gezielt zu manipulieren, um so den Einfluss jedes einzelnen Merkmals auf die Beurteilung der Aussprache untersuchen zu können (vgl. Anderson-Hsieh, Johnson & Koehler 1992 [4, S.549]). Diese Anforderung versuchen wir in unserem Experiment 2 zu realisieren (s. Kap. 8.2).

Neue Erkenntnisse über den Einfluss einzelner prosodischer Merkmale auf die Perzeption der Betonung sollten helfen, Trainingsmaßnahmen insbesondere in Form von computergestütztem Lernen gezielt auf die Bedürfnisse der L2-Lerner auszurichten. Solche Trainingmaßnahmen sollten schon bei Anfängern eingesetzt werden, da es wünschenswert ist, dass Sprachenlerner die Prosodie der L2 von Anfang an lernen (Hirschfeld & Trouvain 2007 [64]). Tatsächlich beeinflusst ein prosodiezentriertes Training auch die segmentalen Aspekte der Aussprache positiv (vgl. Missaglia 1999 [95]). Missaglia entwickelte die von

⁸Der *SPEAK Test* war ein üblicher Test, um die Englischkenntnisse fremdsprachlicher Assistenten an Universitäten der Vereinigten Staaten zu testen (s. Anderson-Hsieh, Johnson & Koehler 1992 [4, S.530 und 550], Clark & Swinton 1979 [18] und 1980 [19]).

⁹Die Muttersprachen der Sprecher waren Arabisch, Armenisch, Chinesisch, Farsi, Deutsch, Griechisch, Indonesisch, Koreanisch, Serbo-Kroatisch, Spanisch und unterschiedliche Sprachen des indischen Subkontinents.

ihr so genannte *Kontrastive Prosodiemethode*, die der Bilingualität der Lernenden besondere Relevanz beimisst und die L1 im Prozess des Erlernens der L2 mit einbezieht. Ziel der Methode ist es, die „prosodische Aufmerksamkeit“ der Lernenden zu aktivieren, das heißt, „die Lerner auf spezifische prosodische Eigenheiten der Mutter- und der Fremdsprache aufmerksam zu machen“ (Missaglia 1999 [96, S.148] und 2001 [97, S.79]). Nach Missaglias Methode üben die Lernenden selbstständig in der Gruppe mit deutschen Texten, die sie ins Italienische übersetzen. Sie versuchen, den Satzakzent zunächst in dem italienischen Satz zu erkennen und ihn zu übertreiben, so dass alle anderen Wörter im Satz deakzentuiert werden. Danach versuchen sie, dasselbe mit dem deutschen Text zu tun. Laut Missaglia verschwindet damit ein Großteil der für Italiener typischen Aussprachefehler in den Sprechakten der Lernenden (s. Kap. 3), denn durch Übertreibung des Satzakzents bleibt für die nicht betonten Wörter wenig Energie, z.B. für volle Vokale in unbetonter Stellung, stimmhafte Laute in Silbenkoda und Epenthese nach wortfinalen Konsonanten (Missaglia 2001 [97, S.86]). Die Lernergruppe übernimmt eine Kontrollfunktion und erteilt der übenden Person Feedback über ihre eigenen Realisierungen. Die Methode wurde an 20 italienischen Deutschstudenten der Katholischen Universität Mailand getestet. Die Ergebnisse der Untersuchungen der Vor- und Nachtestmaterialien zeigen, dass sich die Gesamtaussprache der nach der *Kontrastiven Prosodiemethode* trainierten Gruppe stärker als die der segmentorientiert trainierten Gruppe verbessert hatte. Darüber hinaus machten die Probanden der Prosodiegruppe weniger segmentale Fehler (s. Missaglia 1999 [96] und Missaglia 2007 [98, S.248ff]). Die nach dem Prosodietraining übrig gebliebenen segmentalen Fehler sollten durch ein gezieltes Hörtraining, das die Aufmerksamkeit der Lernenden auf die Merkmale der Einzellaute dirigieren soll, korrigiert werden (Missaglia 2001 [97, S.88]).

Gleich ob im Präsenzunterricht oder auch im Training mit einem Computerprogramm: Lernende sollten in jedem Fall ein geeignetes Feedback bekommen,

mit dem sie die eigenen Fehler korrigieren und ihre Aussprache verbessern können, denn sie sind normalerweise nicht in der Lage, ihre falsche Aussprache ohne externe Hinweise als solche zu erkennen. *Recasts* — die Reformulierungen unkorrekter Äußerungen der Lernenden durch den Lehrer (Nicholas, Lightbown & Spada 2001 [108, S.720]) — hat sich als eine effektive Art des Feedbacks für phonologische Fehler erwiesen (s. Lyster 1998 [85]). Recasts, die mittels Resynthese der Äußerungen der Lernenden geschaffen worden sind, sollten ihnen helfen, zwischen korrekter und unkorrekter Aussprache zu unterscheiden.¹⁰ Zudem könnte die Korrektur in dem Feedback hervorgehoben werden (s. Chaudron 1977 [15, „repetition with change and emphasis“]), um den Lernernden dabei zu helfen, die eigenen Fehler zu identifizieren.

¹⁰Nagano & Ozawa 1990 [105], Germain-Rutherford & Martin [49], Hirose, Gendrin & Minematsu [62], Hirose 2004 [63] und Martin 2004 [89]).

3 Aussprachefehler italienischsprachiger Deutschlernender in der Forschungsliteratur

L2-Sprecher filtern die Laute der Zweitsprache durch ihr „muttersprachliches Sieb“. Diese Hypothese formulierte bereits Trubetzkoy 1939 [144]. Italienischsprachige Deutschlernende nehmen also die deutsche Sprache mit Hilfe der phonologischen Kategorien, die durch ihre Muttersprache bedingt sind, wahr. Aus diesem Grund perzipieren sie möglicherweise einige relevante phonologische Eigenschaften der Sprache nicht, oder, wenn sie es tun, betrachten sie diese als redundant. Aufgrund der Interferenzen, die die Muttersprache beim Lernen der Zweitsprache verursachen kann, ist es sinnvoll, die phonologischen und phonetischen Eigenschaften der Erst- und Zweitsprache zu vergleichen, um mögliche Fehlerquellen zu identifizieren.

Sowohl kontrastive phonologische bzw. phonetische Untersuchungen von Deutsch und Italienisch für den Fremdsprachunterricht als auch empirische Untersuchungen zu den Aussprache Fehlern italienischsprachiger Deutschlernender sind selten in der wissenschaftlichen Literatur. Zu den wenigen Ausnahmen zählen die von Zuanelli Sonino 1976 [156], Mioni 1976 [92], Figge &

De Matteis 1982 [42], De Dominicis 1997 [24] und Grassegger 1986 [56]. Die meisten Werke beruhen auf Beschreibung und Vergleich des deutschen und des italienischen Phoneminventars. Ein Vergleich der prosodischen Eigenschaften der beiden Sprachen ist meistens nicht vorhanden (Zuanelli Sonino 1976 [156], Mioni 1976 [92]) oder, wenn überhaupt, sehr kurz gefasst (wie in De Dominicis 1997 [24, S.245-247]).

Noch seltener als kontrastive Untersuchungen sind empirische Untersuchungen zum Deutschen von italienischsprachigen Deutschlernenden. Missaglia untersuchte in ihrer Dissertation 1999 [96] sowohl segmentale als suprasegmentale Aspekte der deutschen Aussprache von italienischen Muttersprachlern.

3.1 Segmentale Fehler

Die segmentalen Fehler italienischsprachiger Deutschlerner hängen sowohl vom unterschiedlichen Phoneminventar und der Phonotaktik im Deutschen und Italienischen, als auch vom unterschiedlichen Rhythmus ab.

Das Vokalsystem des Italienischen ist im Vergleich zum Deutschen sehr einfach. Es besteht aus den geschlossenen /i/ und /u/, halb geschlossenen /e/ und /o/ und halb offenen /ɛ/ und /ɔ/ Vokalen und aus einem zentralisierten tiefen Vokal /a/ in betonter Stellung (s. Abb. 3.1). In unbetonten Silben weist das Inventar lediglich fünf Vokale auf, da nicht zwischen halb geschlossenen und halb offenen Vokalen /e/ vs. /ɛ/ und /o/ vs. /ɔ/ unterschieden wird. Die Wissenschaftler sind sich über die Qualität der mittleren Vokale in unbetonter Stellung nicht einig. Mioni listet drei Meinungen auf, denen zufolge es entweder: a) nur eine halb geschlossene Realisierung als [e] und [o], b) eine mittlere Realisierung zwischen [e] [o] und [ɛ] [ɔ], und c) eine Abhängigkeit der Realisierung von der folgenden Silbe, also Vokalharmonie, gebe (s. Mioni 1993 [93, S.122]). Jedoch existiert die Opposition zwischen /e/ und /ɛ/, /o/

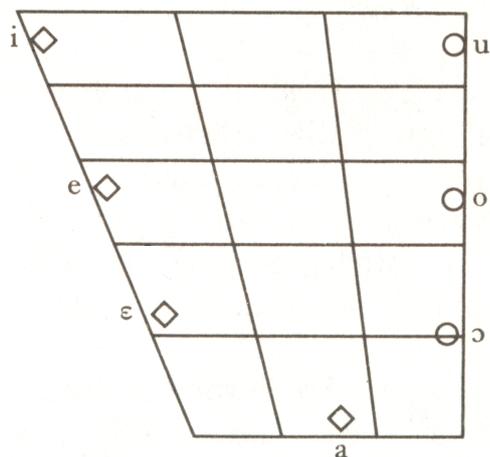


Abbildung 3.1: Vokale des Italienischen in betonter Stellung (aus. Canepari 1979 [12, S.33])

und /ɔ/ für die Hälfte der Regionalvarietäten des Italienischen nicht einmal in betonter Stellung (s. Mioni 1993 [93, p. 122]). Da das deutsche Vokalsystem viel komplexer ist und aus 15 Vokalen in betonter Silbe und dazu Schwa in unbetonter Silbe besteht (s. Kohler 1995 [74, S.169]), könnte es für Italiener problematisch sein, die Diskrimination und Produktion der deutschen Vokale zu lernen. Missaglia untersuchte die Aussprache von zehn italienischsprachigen Deutschlernerinnen, die den Standardtext „Nordwind und Sonne“ lasen. Sie fand heraus, dass die Sprecherinnen in der Regel ungespannte Vokale sowohl in betonter als auch in unbetonter Position gespannt aussprachen (s. Missaglia 1999 [96, S.132]). Zudem tendierten die Sprecherinnen dazu, vordere gerundete Vokale als hintere gerundete Vokale auszusprechen (s. Missaglia 1999 [96, S.133]).

Die Phonotaktik des Deutschen erlaubt Konsonantenkluster, die im Italienischen nicht möglich sind, und die daher mittels Schwa-Epenthese oder Auslassung von Konsonantsegmenten von den italienischen Sprechern erfahrungsge-

mäß vereinfacht werden. Zudem dürfen die Endsilben im Italienischen nur eine CV-Struktur haben, und daher sprechen italienische Sprecher nach Endsilben im Deutschen, die mit einem Konsonant enden, oft einen Schwa aus.

Die deutschen Konsonanten bereiten italienischen Deutschlernenden auch manche Schwierigkeiten. Italiener neigen dazu, Doppelkonsonanten im Deutschen zu realisieren. Zusätzlich wiesen die Produktionen der Italiener oft mangelnde Auslautverhärtung auf (s. Missaglia 1999 [96, S.133]). Der im italienischen Phoneminventar ebenfalls unbekannt Glottalfrikativ fehlt zudem oft in der Aussprache italienischer Deutschlernender. Möglicherweise wird er als redundant empfunden, obwohl er in der Orthographie gekennzeichnet wird. Jedoch tritt auch häufig der umgekehrte Fehler auf: Die Aspiration von Vokalen, denen im Schrift ein <h> folgt, und die nicht aspiriert werden sollten. Der Fehler wird möglicherweise durch *Übergeneralisierung* verursacht:

Italienische Lerner wissen, daß sie im Deutschen aufgrund der fehlenden Aspiration von Vokalen die im Schriftbild einem <h> folgen, häufig Fehler machen, und sie neigen dazu, Vokale selbst dann zu aspirieren, wenn sie unaspiriert sind. (s. Missaglia 1999 [96, S.134])

Das Digramm <ng> wird zudem von Italienern oft als Aufeinanderfolge von velarem Nasal und stimmhaftem velaren Plosiv [ŋg] statt als velarer Nasal [ŋ] ausgesprochen (Missaglia 1999 [96, S.133]). Laut Missaglia ist die Orthographie die Ursache für viele Fehler. Die Italiener lernen zunächst das Deutsche zu schreiben und erst später auszusprechen, daher gewöhnen sie sich daran, die Graphem-Phonem Korrespondenz des Italienischen auf das Deutsche anzuwenden (Missaglia 1999 [96, S.134]).

Kontrastive phonologische Untersuchungen basieren auf der These, dass nach einem Vergleich des italienischen mit dem deutschen Phonemsystem die fehlerhaften Realisierungen italienischsprachiger Deutschlernender vorhersagbar wären. Jedoch lassen sich solche Vorhersagen durch empirische Erfahrung

bzw. Untersuchungen bisweilen widerlegen. Als Beispiel gilt die von Zuanelli Sonino 1976 [156] vorhergesagte Substitution der vorderen gerundeten Vokale im Deutschen durch italienische Sprecher. Es stellt sich die Frage, ob Italiener die deutschen vorderen gerundeten Vokale als vordere ungerundete oder als hintere gerundete interpretieren, das heißt, ob bei ihrer Diskriminierung die Artikulationsstelle oder die Lippenrundung relevanter ist. Laut Zuanelli Sonino 1976 [156, S.71ff] ist die Lippenrundung kein distinktives Merkmal im Italienischen, und daher sollten die Italiener nur die Artikulationsstelle des Vokals als relevant — und folgerichtig auch das deutsche /y:/ als italienisches /i/ — interpretieren. Jedoch belegt die Erfahrung mit italienischen Deutschlernenden, dass sie die vorderen gerundeten Vokale als hintere gerundete interpretieren und aussprechen (s. Missaglia 1999 [96]). Zu Zuanelli Soninos Vorhersage bemerkt Missaglia: „Dabei zeigt sich, daß nach theoretisch ableitbaren Erwartungen erstellte Hypothesen, die sich auf die starke Version der Kontrastiven Analyse-Hypothese stützen [...] nicht immer korrekt sind.“ (Missaglia 1999 [96, S.79]) Missaglia begründet die fehlerhafte Interpretation der vorderen gerundeten Vokale im Deutschen durch Italiener mit der Orthographie. /y:/ und /ʏ/ werden als <ü> und /ø/ und /œ/ als <ö> geschrieben, weswegen italienische Deutschlernende sie als *u-* bzw. *o-Laute* kategorisieren. Möglicherweise könnte aber die Lippenrundung eine wichtigere Rolle als die Artikulationsstelle in der Diskrimination der Vokale durch italienische Sprecher spielen. Die Vokale /i/, /e/ und /ɛ/ unterscheiden sich von /u/, /o/ und /ɔ/ im Italienischen sowohl hinsichtlich der Artikulationsstelle als auch hinsichtlich der Lippenrundung. Allein durch Betrachtung des Vokalsystems kann man nicht erkennen, welches der beiden Merkmale für die Kategorisierung der Laute ausschlaggebend ist. Die perzeptive Relevanz der Merkmale sollte durch empirische Untersuchungen erforscht werden.

Über die Unterschiede zwischen italienischem und deutschem Vokalismus existieren bereits phonologisch ausgerichtete Untersuchungen (Zuanelli Sonino

1976 [156], Mioni 1976 [92], Canepari 1979 [12] und De Dominicis 1997 [24]). Jedoch sind in der Literatur phonetisch ausgerichtete Untersuchungen ausgesprochen selten. Dazu zählen die Untersuchungen von Missaglia 1999 [96] und von Gut 2003 [57].

Studien belegen, dass sich die Laute bilingualer L2-Sprecher von denen monolingualer L1-Sprecher unterscheiden (s. MacKay, Flege, Piske & Schirru 2001 [86] und Flege, Schirru & MacKay 2003 [44]).¹ Missaglia untersuchte die Vokalproduktion deutscher Muttersprachler und italienischsprachiger Deutschlerner. Sie fand heraus, dass sich italienische Vokale — von italienischen Sprechern gesprochen — über das gesamte Vokaldreieck verteilen, ähnlich wie deutsche Vokale, die von deutschen Muttersprachlern gesprochen wurden (s. Missaglia 1999 [96, S.111ff]). Das heißt, dass Italiener auch reduzierte Vokale produzieren können, was Albano Leoni, Cutugno & Savy bereits 1995 [1] in ihrer Studie zu italienischer *connected speech* am Beispiel eines Korpus aus Fernsehrichtensendungen zeigten. Sie sind also an große Artikulationsfreiheit gewöhnt, was die Produktion der Vokale betrifft, denn im Italienischen kann dasselbe Vokalphonem durch Vokale unterschiedlicher Qualität dargestellt werden. Dies ist jedoch im Deutschen nicht möglich, denn bereits kleinere spektrale Differenzen um ca. 100-200 Hz können den Unterschied zwischen zwei Vokalen charakterisieren (Missaglia 1999 [96, S.120]). Im Unterschied zu den italienischen Vokalen konzentrieren sich die Streuungsbereiche der deutschen Vokale der italienischen Muttersprachler auf die fünf Vokale /i/, /e/, /a/, /o/ und /u/. Die Realisierung der deutschen Vokale wird also nicht durch die phonetische Realisierung der italienischen Vokale gefiltert, sondern vielmehr durch das Phoneminventar

¹Laut einer mit italienischen Sprechern des Englischen von Flege, Schirru & MacKay 2003 [44] durchgeführten Studie unterscheiden sich die Vokale der bilingualen L2-Sprecher von denen der monolingualen L1-Sprecher aufgrund von Assimilation- bzw. Dissimilationsphänomenen. Solche Phänomene beruhen auf einer neuen Strukturierung phonetischer Kategorien, die vom Kontakt des phonetischen Subsystems der L1 mit dem der L2 verursacht werden (s. MacKay, Flege, Piske & Schirru 2001 [86] und Flege, Schirru & MacKay 2003 [44]).

des Italienischen (Missaglia 1999 [96, S.116]). Bereits Trubetzkoy behauptete, dass der „fremde Akzent“ nicht durch mangelnde Artikulationsfähigkeit sondern durch fehlerhafte, im Phoneminventar der Muttersprache wurzelnde Beurteilung der Laute entstehen würde. Trubetzkoy schreibt:

der sogenannte „fremde Akzent“ [hängt] gar nicht davon [ab], daß der betreffende Fremde irgendeinen Laut nicht aussprechen kann, sondern vielmehr davon, daß er diesen Laut nicht richtig beurteilt. Und diese falsche Beurteilung des Lautes einer fremden Sprache ist durch den Unterschied zwischen der phonologischen Struktur der fremden Sprache und der Muttersprache des Redenden bedingt (1939 [144, S.49]).

Nach Kuhls *Native Language Magnet Theory* wird die Wahrnehmung und Produktion der Laute in der L2 durch phonetische Prototypen bedingt, die durch die phonologischen Kategorien der L1 bestimmt werden. Diese Prototypen bewirken, dass benachbarte Laute auf die selbe Kategorie der Prototypen zurückgeführt werden (s. Kuhl 1991 [77], 1993 [78] und Kuhl & Iverson 1995 [79]).

Missaglia 1999 [96] untersuchte die Wahrnehmung deutscher Vokale durch italienische Muttersprachler. Dafür replizierte sie ein Experiment, das Sendmeier 1981 [134] durchgeführt hatte, um die Wahrnehmung deutscher Vokale durch deutsche Muttersprachler zu untersuchen. Missaglia fand heraus, dass sich die Italiener für die Diskriminierung deutscher gespannter vs. ungespannter Vokale eher an der Vokalquantität als an der -qualität orientieren (Missaglia 1999 [96, S.122ff]). Weder die Opposition gespannt vs. ungespannt noch die Vokalquantität hat im Italienischen phonematischen Status, jedoch sind betonte Vokale durch längere Dauer charakterisiert (Bertinetto 1981 [10]), und dies kann das Ergebnis des Experiments erklären.

3.2 Prosodische Fehler

Bislang existieren noch keine umfassenden kontrastiven Vergleiche der Prosodie des Deutschen mit der Prosodie des Italienischen. Die vorliegenden Untersuchungen betreffen stattdessen das Deutsche und weitere Fremdsprachen (s. Dellatre 1965 [25], Gibbon 1980 [50] und 1984 [51], Wunderli 1981 [154], Scuffil 1982 [133], Jin 1990 [68]). In den kontrastiven Untersuchungen zwischen Deutsch und Italienisch spielt die Prosodie im Vergleich zu den segmentalen Aspekten eine sekundäre Rolle und wird meistens nur kurz behandelt (Canepari 1979 [12] und De Dominicis 1999 [24]). Canepari bietet in seiner Einführung in die Phonetik (Canepari 1979 [12]) eine Darstellung des Phonemsystems des Deutschen, der Realisierungen der Phoneme im Kontext und eine knappe Beschreibung der Betonung von Komposita und der Intonation (Canepari 1979 [12, S.266-267]). De Dominicis 1999 widmet wenige Zeilen der Betonung, liefert eine kurze Darstellung der Silbenstruktur im Deutschen, und bietet beziehungsweise auf Canepari 1979 [12, S.266-267] einen kurzen Vergleich zwischen deutscher und italienischer Intonation. [24, S.245-247]).

Tonellis Untersuchung, 1984 [142] entstanden und 2002 [143] überarbeitet, besteht aus einem kontrastiven Vergleich des Bozner Deutschen mit dem Bozner Italienischen. Bezugnehmend auf die schwache Form der Isochronie-Hypothese von Bertinetto 1977 [8] widmet Tonelli 2002 [143] ein Kapitel der Untersuchung des Rhythmus der beiden Varietäten und kommt zu dem Schluss, dass es sich dabei um zwei unterschiedliche Rhythmustypen handelt. Schwer zu bestimmende Silbengrenzen, Vokalreduktion und Tilgung der Silbenkerne in unbetonten Silben im Bozner Deutschen deuten auf einen akzentzählenden Rhythmus hin. Im Bozner Italienischen lassen hingegen leicht auffindbare Silbengrenzen, Tensing-Prozesse der unbetonten Vokale und eine tendenzielle Erhaltung der Silbenstruktur in unbetonter Stellung auf einen silbenzählenden Rhythmustyp schließen.

In silbenzählenden Sprachen wie dem Italienischen wird die Betonung hauptsächlich mittels Daueränderung markiert (s. Rigault 1962 [123], Delattre 1966 [26], Wunderli 1981 [154], Wenk & Wioland 1982 [151], Wenk 1983 [150] zum Französischen und Bertinetto 1976 [7] und 1981 [10], Farnetani & Kori 1982 [40] und 1983 [41], Marotta 1985 [88], Vayra, Avesani & Fowler 1987 [147] zum Italienischen). Möglicherweise spielt im Deutschen die Intonation als Korrelat der Betonung eine wichtigere Rolle. Italiener können hingegen Langvokale fälschlicherweise als betont wahrnehmen. (s. Kap. 2.3.1 dieser Arbeit).

Zusätzlich kann es — bedingt durch die im Italienischen und Deutschen üblichen Unterschiede hinsichtlich der Alignierung des auf der betonten Silbe liegenden F0-Gipfels mit den Lauten — zu Missverständnissen der Identifizierung der Betonungsstelle kommen. Da im Italienischen ein früherer F0-Gipfel üblich ist, könnten Italiener einen späteren F0-Gipfel — kurz vor dem Ende der betonten Silbe — im Italienischen deutscher Lernender als fehlerhafte Betonung auf der folgenden Silbe interpretieren (Ladd 1996 [80, S.128f] und Mennen 2007 [91, S.58f]). Gleichfalls könnten auch italienische Lerner des Deutschen wegen eines späteren F0-Gipfels im Deutschen die Pänultima statt der Antepänultima als betont wahrnehmen.²

Im Italienischen kann gewöhnlich der Fokus im Satz durch eine Änderung der Wortstellung — vor allem durch Verschiebung des zu akzentuierenden Inhaltswortes ans Satzende — erreicht werden. Hingegen erfolgt im Deutschen häufig die Hervorhebung durch prosodische Mittel. Dies kann zur Deakzentuierung des letzten lexikalischen Elements im Satz führen, was für Italiener schwierig zu lernen sein kann, denn sie tendieren dazu, das letzte Element des Synthagmas zu akzentuieren — selbst wenn, wie in Nebensätzen, das Verb am Satzende steht — (s. Missaglia 1999 [96, S.91-95] und Kap. 5.2.2 „Deaccenting and *semantic weight*“ in Ladd 1996 [80, S.174ff]). Zudem sind Aussprachefehler der

²Zur Alignierung des F0-Gipfels mit dem Beginn des betonten Vokals und deren semantischen Funktion im Deutschen s. Kohler 1991 [73].

Italiener in der fünften Prosodiedimension, der Reduktion (s. Pfitzinger 2006 [117] und Kap. 2.1), zu erwarten. Das Italienische kennt anders als das Deutsche und als andere akzentzählenden Sprachen keine ausgeprägte Reduktion der unbetonten Silben (s. Rossi 1998 [127]). Wenn wir dies zur unterschiedlichen Fokusrealisierung und zur unterschiedlichen Relevanz der prosodischen Parameter in der Darstellung der Betonung in den beiden Sprachen addieren, finden wir eine große Vielfalt an Interferenzquellen, die die Wahrnehmung der phonologischen Eigenschaften des Deutschen für Italiener erschweren können.

Gut untersuchte 2003 [57] die Realisierung von Vokalreduktion durch drei deutsche Muttersprachler und 18 fremdsprachliche Deutschlerner, darunter sechs Italiener. Der untersuchte Korpus bestand sowohl aus gelesener Sprache als auch aus Spontansprache (Wiedergabe des gelesenen Textes). Die Italiener elidierten die Vokale in Endsilben mit der Struktur C+<-en> bzw. C+<-em> seltener als die Deutschen und produzierten einige davon als Vollvokale. Darüber hinaus wurden die zu reduzierenden Silben von den Italienern in der gelesenen Sprache nicht genug reduziert: Bei ihnen war das Verhältnis der Dauer von nicht-reduzierten und der unmittelbar folgenden reduzierten Silben signifikant niedriger als bei den Deutschen. Bei Spontansprache waren die Unterschiede nicht signifikant. In einer Untersuchung 2006 [58] maß Gut den ersten und den zweiten Formanten von Vokalen in Silben mit C+<-en>-Struktur, die von sechs Italienern und drei Deutschen gesprochen wurden und fand heraus, dass die Italiener die Vokale signifikant gespannter sprachen als die Deutschen (s. auch Gut 2007 [59, S.160ff]).

Anhand ihrer empirischen Untersuchungen belegt auch Missaglia die Tendenz der Italiener, unbetonte Reduktionsvokale an den Wort- bzw. Morphemgrenzen nicht zu elidieren, sondern sie als Vollvokale zu realisieren. Italiener elidieren hingegen oftmals fälschlicherweise die auslautenden Konsonanten und realisieren dadurch die im Italienischen häufigste Silbenstruktur CV auch im Deutschen (s. Missaglia 1999 [96, S.131]). Darüber hinaus unterscheiden Ita-

liener nicht zwischen betonten und unbetonten Satzelementen und reduzieren letztere nicht (Missaglia 1999 [96, S.134]). Über Wortbetonungsfehlern berichten Missaglia 1999 [95] und Delmonte 2002 [30].

Missaglia fand heraus, dass sowohl Anfänger als auch fortgeschrittene Deutschlernende — wenngleich in unterschiedlichen Maßen — von den selben Arten von prosodischen Fehlern betroffen sind (s. Missaglia 1997 [94] und Missaglia 2007 [98, S.238f]). Laut Missaglia sind viele Schwierigkeiten in der Aussprache einzelner Segmente nicht auf eine unzulängliche Realisierung einzelner Phoneme, sondern auf mangelnde Kompetenz im suprasegmentalen Bereich zurückzuführen (Missaglia 1997 [94]). Die Aussprachefehler italienischsprachiger Deutschlernender beruhen hauptsächlich auf dem Kontakt zwischen einer silbenzählenden — dem Italienischen — und einer akzentzählenden — dem Deutschen — Sprache (Missaglia 1999 [96]).

4 Computergestütztes Lernen der Aussprache

In diesem Kapitel wird ein Überblick über die bestehenden Computer Assisted Language Learning Systeme (CALL) zur Vermittlung der Aussprache gegeben. Der Schwerpunkt hierbei liegt auf Systemen, mit denen prosodische Aspekte der gesprochenen Sprache, wie Intonation, Reduktionsphänomene, Wortbetonung, Satzakzent und Rhythmus der L2 trainiert werden können. Zunächst wird eine Definition für CALL gegeben. Unter anderem soll die Frage beantwortet werden, welche Vorzüge CALL-Systeme zum Training der Aussprache bieten, und wie sie eine optimale Lernumgebung zum Trainieren von Kommunikationsfähigkeiten schaffen können. Auch die Nachteile der Verwendung automatischer Spracherkennung in CALL und unterschiedliche Möglichkeiten der Darstellung von Stimuli — als Input und Feedback — für den Lernenden in CALL werden dargestellt.

4.1 Computer Assisted Language Learning zum Training der Aussprache

Computer Assisted Language Learning bezieht sich auf die Verwendung des Computers als Hilfsmittel beim Sprachenlernen und im Sprachunterricht.

Hierfür wurden unterschiedliche Bezeichnungen vorgeschlagen, wie Computer-Based Education (CBE) und Computer-Managed Learning (CML) (vgl. Levy 1997 [83]). Der Begriff CALL setzt im Gegensatz zu CBE und CML einen breiteren Lernkontext voraus, in dem der Computer nicht als ausschließliches Lernmittel, sondern als Ergänzung zum Präsenzunterricht eingesetzt wird.

Im Rahmen von CALL kann der Computer als Hilfsmittel zum Erlernen der Aussprache eingesetzt werden. Die meisten bestehenden CALL-Programme für Aussprache haben als Ziel das Erlernen der segmentalen Aspekte der gesprochenen Sprache, aber einige trainieren auch die Prosodie der L2. Die Vorzüge von CALL-Systemen werden im nächsten Abschnitt beschrieben.

4.1.1 Vorzüge von CALL-Systemen zum Erlernen der Aussprache

CALL-Systeme können im Zusammenhang mit Präsenzunterricht eine optimale Lernumgebung schaffen und die Effektivität der gesamten Lernmaßnahme verbessern. Hardison fand heraus, dass „computer-assisted prosody training“ des Französischen die Prosodie der Produktionen von Lernerinnen amerikanisch-englischer Muttersprache signifikant verbessern konnte (Hardison 2004 [60]). Die Forscherin gelangte zu nachstehender Schlussfolgerung: „frequent input, use of contextualized vocabulary with application to daily life, a range of syntactic and prosodic structures, practice opportunities, and auditory and visual feedback contribute to learning“ (Hardison 2004 [60, S.49]). Mit einem CALL-System stünden dem Lerner in der Tat stets ausreichende Inputmaterialien und Übungsmöglichkeiten zur Verfügung.

Zweitspracherwerbsforschungen haben gezeigt, wie verständlicher Input (s. dazu Kap. 4.1.3) und die eigene Sprachproduktion die sprachliche Entwicklung des Lerners fördern (Pallotti 2000 [112], Jones 1997 [70], Derwing, Munro

& Wiebe 1998 [32], Flege 1999 [45]). Im Präsenztraining hat der Lerner selten Gelegenheit zu sprechen, weil er sich den Lehrer mit anderen Lernenden teilen muss. Mit CALL könnten Lerner hingegen jederzeit üben. Ein weiterer Vorteil eines CALL-Programms ist, dass der Lerner die Möglichkeit hat, verschiedene Sprecher zu hören. Dies ist einer der Grundsätze, die laut Eskenazi 1999 [39] zu einem erfolgreichen Aussprachetraining beitragen. Sowohl im Präsenztraining als auch im One-to-one-Training wird der Lerner nur mit einem Sprachmuster — dem des Lehrers — konfrontiert, der häufig und insbesondere in der Schule kein L1-Sprecher ist. Mit einem entsprechend entwickelten CALL-Programm hätte der Lerner nicht nur die Möglichkeit, mehrere Sprecher zu hören, sondern könnte sich auch an verschiedene Aussprachevarianten der Zielsprache gewöhnen. Hardison schreibt, dass Feedback idealerweise sowohl akustisch als auch optisch erfolgen soll (Hardison 2004 [60, S.49]). Ein Vorteil von CALL besteht darin, dass es mehrere Medien zur Darstellung des Feedbacks nutzen kann. Das System könnte sprachlich erklären, was in der Äußerung des Lerners nicht korrekt ist, eine schriftliche Meldung oder eine graphische Darstellung der Intonation des Lerners einblenden und sie mit der graphischen Darstellung der Intonation eines L1-Sprechers vergleichbar machen.

Damit der Vorteil des multimedialen Feedbacks genutzt werden kann, muss ein CALL-Programm in der Lage sein, Feedback korrekt und verständlich darzustellen (Delmonte 2000 [29, S.146], Eskenazi 1999 [39, S.63], Neri et al. 2002 [106, S.452ff]).

Ein weiterer Vorzug von CALL besteht darin, dass auch Lerner, die Angst haben, vor der Klasse zu reden und deswegen die Übungsmöglichkeiten während des Unterrichts nicht nutzen, mit einem Sprachlernprogramm ausreichend üben können (Eskenazi 1999 [39, S.64]). Angst, vor der Klasse in einer wenig vertrauten Sprache zu reden, ist ein Faktor, der die Leistungen des Lerners in der L2 beeinträchtigt. Während dieser Faktor seine Wirkungen auf Schreib- und Leseleistungen weniger zeigt, ist er entscheidend in der sprachlichen Kom-

munikation. Für einige Lerner kann es angenehmer sein, mit einem CALL-Programm statt im Präsenzunterricht zu üben.

Darüber hinaus erklärt Hardison einige positive Aspekte, die CALL für die Benutzer haben kann:

The computer program provides a visual display in real-time and the opportunity to overlay the NS [native speaker Anm. d. V.] version on the learner's — a salient form of feedback that drew many positive comments from participants. [...] For some learners, technology holds greater interest, which influences motivation. Some enjoy a greater comfort level in working with a computer program than in face-to-face interaction where other personality factors are involved. (Hardison 2004 [60, S.48])

Dabei erklärt Hardison, dass mit den oben genannten Aspekten nicht die höhere Effektivität von Sprachlernprogrammen im Vergleich zum Präsenzunterricht gemeint ist. Beide Methoden bringen Vorteile für die Lerner, die am Besten genutzt werden können, wenn sie miteinander kombiniert werden.

Im Hinblick auf eine Kombination von CALL und Präsenzunterricht, sollen CALL-Programme für Lehrer einfach zu bedienen sein und keine besonderen Kenntnisse über Phonetik und Sprachsignalverarbeitung voraussetzen. Leider verfügen die meisten Fremdsprachenlehrer nicht über solche Kenntnisse, daher soll ein CALL-Programm die meisten Prozesse, wie die Segmentierung der Sprachsignale, automatisch durchführen können.

Die Anwendung eines CALL-Programms hätte außerdem die Vorteile, einen Lehrer mit der Phonetik vertrauter zu machen, und ihn erfahren zu lassen, wie sinnvoll Präsenzunterricht damit ergänzt werden kann. Lehrer haben leider häufig Vorurteile gegen den Einsatz des Computers im Unterricht. So schließen manche Lehrer wegen falscher Bewertungen der Aussprache der Lerner

durch einige Programme darauf, dass Computer grundsätzlich für das Sprachenlernen nicht geeignet sind. Andere befürchten, Computer könnten eines Tages den Lehrer ersetzen. Diese beiden gegensätzlichen Behauptungen sind unbegründet. Tatsächlich kann ein sinnvoll pädagogisch geplantes CALL eine große Hilfe für den Lehrer im Unterricht sein.

Damit die Vorzüge von CALL genutzt werden können, soll CALL-Systemen eine didaktische und kognitive Theorie zu Grunde liegen (Roche 2003 [125] und Plass 1998 [120]). Leider ist dies jedoch häufig nicht der Fall, insbesondere bei kommerziellen CALL-Programmen. Die Hersteller werben mit den technischen Besonderheiten ihrer Programme, ohne ihre didaktische Effektivität gemessen zu haben (Neri, Cucchiaroni, Strik & Boves 2002 [106], Roche 2000 [124], Roche 2003 [125] und Murray & Barnes 1998 [104]). Vorrangiges Ziel bei der Entwicklung eines CALL-Programms sollte die optimale Nutzung der vorhandenen Technik zur Unterstützung der Didaktik sein.

Die Sprachtechnologie hat in den letzten Jahrzehnten erhebliche Fortschritte gemacht. Heutzutage ist es möglich, auf dem eigenen PC Sprachsignale zu analysieren und zu synthetisieren. Solche Technologien können und sollen für die Sprachdidaktik angewendet werden, jedoch soll die Didaktik vor der Technologie kommen und nicht umgekehrt. Des Weiteren stellt sich die Frage, inwieweit die betreffende Technologie zum Erreichen des didaktischen Ziels geeignet ist, zum Beispiel bei der im Folgenden beschriebenen Anwendung der automatischen Spracherkennung in CALL.

4.1.2 Automatische Spracherkennung in CALL

Verschiedene CALL-Programme bieten eine Evaluierung der Aussprache des Lernalters mittels automatischer Spracherkennung (ASR). Das ursprüngliche Ziel der ASR besteht jedoch darin, anhand eines gegebenen Sprachsignals die ent-

sprechende Wortkette zu erkennen. Meistens basieren ASR-Systeme auf Hidden Markov Modellen (HMM), statistischen Modellen, die für jede in Frage kommende Wortkette die Wahrscheinlichkeit berechnen, dass sie das Sprachsignal produziert hat (s. Jurafsky & Martin 2000 [71, S.241ff]). Die Wortkette mit der höchsten Wahrscheinlichkeit wird schließlich erkannt. Die Frage ist, ob ASR-Technologie auch für die Sprachdidaktik geeignet ist. Für Prosodietraining ist sie auf jeden Fall ungeeignet, weil sie gerade die Prosodiemerkmale des Sprachsignals nicht berücksichtigt:

Current ASR systems are insensitive to fundamental frequency, to amplitude and to details of vocal cord activity. They are also tolerant to differences in the durations and amplitude of speech sounds since they contribute little to determine the phonetic identity of speech sounds which is the primary source of information to recover their lexical correspondence. (Delmonte 2000 [29, S.146])

Für die automatische Spracherkennung sind also nur die Informationen von Bedeutung, mittels derer Phoneme — und dadurch die Wörter — erkannt werden können. Weitere Eigenschaften, die für die Prosodie relevant sind, werden vernachlässigt.

Die Anwendung der ASR für segmentales Feedback ist jedoch ebenso problematisch, weil es nicht nur darum geht, Wortketten zu erkennen — wofür ASR entwickelt wurde — sondern die Aussprache des Lerners zu bewerten. Die ASR-Technologie wird in verschiedenen Aussprachetrainingsystemen mit segmentalem Feedback angewendet:

In all systems based on HMMs (Kawai and Hirose, 1997 [72]; Ronen et al., 1997 [126]), student's speech is segmented and then matched against native acoustic models. The comparison is done using HMM loglikelihoods, phone durations, HMM phone posteri-

or probabilities, and a set of score is thus obtained. They should represent the degree of match between non-native speech and native models. [...] there are typically two databases, one for native and another for non-native speech which are needed to model the behaviour of HMMs. (Delmonte 2000 [29, S.148])

Ein großes Problem der Aussprachetrainingssysteme, die auf HMM basieren, ist die mangelnde Erkennungsgenauigkeit aufgrund der statistischen Modellierung. Die daraus resultierenden Fehler können laut Delmonte 2000 [29, S.149] für Diktiersysteme unerheblich, jedoch für CALL-Systeme unannehmbar sein. Tatsächlich können sie bewirken, dass die gute Aussprache des Lerners irrtümlicherweise vom System abgelehnt (*false rejection*) und die falsche Aussprache als korrekt anerkannt (*false acceptance*) wird. Wenn dies geschieht, wird der Lerner nicht mehr glauben, dass das System seine Aussprache bewerten kann, und er wird das Interesse am Üben mit der Sprachlernsoftware verlieren (Delmonte 2000 [29, S.147]).

Idealerweise sollte ein CALL-Programm nur Technologie verwenden, die die in einer Lernsituation notwendige hohe Präzision und Zuverlässigkeit anbieten kann.

4.1.3 Input und Feedback in CALL

Mit dem Begriff *Input* ist die Zielsprache, der der Lerner ausgesetzt wird, und im Fall von CALL die Stimuli, die der Benutzer vom System erhält, gemeint. (s. das Zweitspracherwerbmodell von Gass 1997 [48] und deren vereinfachte Version von Chapelle 1998 [14, S.23]). Entsprechend ist der *Output* das, was der Lerner selber in der Zielsprache produziert, und was seine sprachliche Entwicklung auf zwei Weisen fördert: erstens zwingt er den Lerner, auf die Form der Sprache zu achten, um sie zu produzieren, zweitens entlockt er dem Ge-

sprächspartner bzw. dem CALL-Programm weiteren Input — das *Feedback*, das Hinweise zu Fehlern des Lernalers beinhalten kann, und ihn dazu bringt, weiteren Output zu produzieren. Laut der *interactionist theory* der *Second Language Acquisition*-Forschung (SLA) ist diese Interaktion die Basis für den Zweitspracherwerb:

The SLA model of interest hypothesizes that large language input acts as the potential starting point for acquiring aspects of the L2. The model expands considerably on Krashen's idea that a lot of comprehensible input is what is needed for SLA. It attempts to articulate what makes input comprehensible and how it is processed to influence the development of the learner's linguistic knowledge. (Chapelle 1998 [14, S.23])

Nach unserer Auffassung besteht *comprehensible input* (verständlicher Input) zum Erlernen der Aussprache aus Stimuli, in denen segmentelle bzw. prosodische Eigenschaften der gesprochenen Sprache, auf die der Lerner seine Aufmerksamkeit richten soll, hervorgehoben wurden. Die Hervorhebung kann zum Beispiel aus einer visuellen Darstellung der Intonation, der Dauer der Sprachsegmente (Delmonte 2000 [29]) oder — falls die Aufmerksamkeit auf die Betonung gerichtet werden soll — aus überbetonten Stimuli bzw. aus mittels Resynthese korrigierten Stimuli in der Stimme des Lernalers bestehen (s. Kap. 7 und 8.3). Idealerweise sollte ein CALL-Programm nicht nur die Stimme des Lehrers, sondern auch die Stimme anderer Muttersprachler als Muster verwenden. Eine große Vielfalt an Mustern soll die Induktion allgemeiner phonetischer Kategorien ermöglichen (Logan, Lively & Pisoni 1991 [84]).

Der Input kann durch eine multimediale Darstellung verständlicher gemacht werden. Laut Massaro 1987 [90] und Jones 1997 [70] fördert die visuelle in Zusammenhang mit der auditiven Information die Wahrnehmung der gesprochenen Sprache. Empirische Untersuchungen zeigen, dass visuelles Feedback

im Zusammenhang mit auditivem Feedback effektiver als das alleinige auditive Feedback ist (s. Hardison 2004 [60] und De Bot 1983 [23]). De Bot 1983 [23] untersuchte in einem Experiment mit holländischen Lernenden des Englischen die Effektivität von aus graphischer Darstellung der Intonation bestehendem Feedback. Sie fand heraus, dass — unabhängig von der Übungsdauer — akustisches Feedback mit Darstellung des Intonationsverlaufs effektiver als alleiniges akustisches Feedback war. Eine weitere vielversprechende Möglichkeit ist die Verwendung von Videos mit sprechenden Köpfen als virtuellen Sprachtutoren (Granström 2004 [53], Granström & House 2006 [54] und 2007 [55]). Die prosodischen Eigenschaften der gesprochenen Sprache können durch die Gestik hervorgehoben werden, zudem kann ein Video darstellen, wie die Laute artikuliert werden.

Der Lerner soll nicht nur Input bekommen, sondern auch viel Output produzieren (Eskenazi 1999 [39]). Ein CALL-System sollte dem Lerner die Möglichkeit geben, zu sprechen und seine Aussprache aufzunehmen. Das Problem ist dabei, in welcher Art und Weise und mit welchem Inhalt dem Lerner ein Feedback gegeben werden soll.

Das Feedback soll zuverlässig und informationsreich sein. Eine automatische Aussprachebewertung (Witt & Young 2000 [153] und Neumeyer et al. 2000 [107]) – wie bei manchen kommerziellen Softwareangeboten – wäre kein gutes Feedback, nicht nur wegen möglicher *false rejection* und *false acceptance* (s. Kap. 4.1.2), sondern auch weil dadurch der Lerner keine Informationen bekommt, um seine Aussprache zu verbessern. Im Klassenunterricht ist *recast* ein effektives Feedback bei phonologischen Fehlern (s. Lyster 1998 [85] und Kap. 2.4). Der Lerner ist tatsächlich häufig nicht in der Lage, die eigenen Fehler zu erkennen, weil er zum Beispiel bestimmte phonetische Kategorien in der Zielsprache nicht diskriminieren kann. In CALL kann die Hervorhebung der Korrekturen multimedial erfolgen: mittels Audio, Text oder Graphik.

Der Lerner soll außerdem das Feedback unmittelbar nach seiner Aussprache bekommen. Bei einem CALL-System werden die Äußerungen des Lerners aufgenommen und erst später von einem Lehrer evaluiert (Germain-Rutherford & Martin 2000 [49], Martin 2004 [89], WinPitch [152], s. dazu Kap. 4.2.1). Jedoch ist unmittelbares Feedback effektiver, weil es dem Lerner ermöglicht, seine Aussprache mit der Aussprache des Modells zu vergleichen (Nicholas, Lightbown & Spada 2001 [108]). Die Herausforderung besteht also darin, CALL-Systeme zu entwickeln, die das Feedback automatisch produzieren. Und dieses Feedback muss zuverlässig sein, um den didaktischen Zwecken zu genügen.

4.2 Aussprachetraining der Prosodie

Aufgrund ihrer Komplexität und ihrer schwierigen Erlernbarkeit wurde die Prosodie viele Jahre von der Sprachdidaktik nicht berücksichtigt. Neben den segmentalen sollten jedoch auch die suprasegmentalen Aspekte vom Anfang an erlernt werden (Eskenazi 1999 [39, S.64]).

4.2.1 Visuelles und auditives Feedback zum computergestützten Erlernen der Prosodie: Ein historischer Überblick

Eine gängige Methode, um Lernenden die Prosodie der L2 zu vermitteln, ist die visuelle Darstellung des Intonationsverlaufs. Der erste Versuch in diese Richtung wurde von Vardanian (1964 [146]) unternommen, der brasilianischen Lernenden die melodische Variation des Englischen beibringen wollte. Lane & Buiten (1969 [81]) entwickelten danach das System SAID (Speech Auto Instructional Device), das Intonation, Intensität und Dauer der Äußerung des Lerners visualisieren und ihre Abweichung vom Satz des Modells bezeichnen

konnte. Diese ersten Versuche waren nicht sonderlich vielversprechend. Vardanian behauptete sogar, dass die visuelle Darstellung der Intonation den L2-Sprecher vom Lernen ablenken würde.

Erst in den 70er und 80er Jahren wurde die Methode der visuellen Darstellung erfolgreich eingesetzt (James 1976 [66], De Bot 1983 [23] und Cranen et al. 1984 [20]). Seitdem wurden Displays des Intonationsverlaufs häufig in CALL-Systemen implementiert (vgl. u.a. Chun 1989 [16] und 1998 [17], Anderson-Hsieh 1992 [2] und 1994 [3], Delmonte, Petrea & Bacalu 1997 [31], Delmonte 1997 [28] und 2000 [29], Kommissarchik & Kommissarchik 2000 [76] und Hardison 2004 [60]). Das *Prosodic Module* der Software SLIM, Software Linguistico Interattivo Multimediale (interaktive multimediale Sprachlernsoftware, Delmonte 1995 [27] und 2000 [29]) bietet als Feedback auch einen stilisierten Intonationsverlauf dar. Die Software, die zur Verbesserung der Perception und Produktion der Prosodie des Englischen bei italienischen Muttersprachlern entwickelt wurde, enthält zudem Übungen zu Wortbetonung, Satzakkzent, Silben- und Wortdauer. Das *Prosodic Module* ist — dem Prinzip des Lernens-im-Kontext folgend — zusammen mit weiteren Modulen wie *Grammar*, *Dialogue Understanding* and *Production Activities* integriert. Eines der drei Übungsfenster — benannt „Enunciato e parole fonologiche“ (Äußerung und phonologische Wörter) — möchte die Benutzer mit den Reduktionsphänomenen der gesprochenen Sprache in fließender Rede vertraut machen. Das Fenster zeigt zwei Varianten derselben Phrase: die *lazy* und die *fluent* Variante. In der *lazy* Variante werden keine phonologischen Regeln angewendet. Die *fluent* Variante zeigt hingegen die phonologischen Änderungen, die in fließender Rede üblich sind. Im zentralen Teil des Fensters wird die Aussprache des Benutzers an der Aussprache des Muttersprachlers ausgerichtet. Die Dauer jeder einzelnen Silbe wird durch die Länge entsprechender Felder sowohl in der Aussprache des Musters als auch in der Aussprache des Lerners angezeigt. Wenn der Benutzer die falsche Silbe betont, leuchtet das entsprechende

Feld rot. Gleichzeitig wird im Musterwort das Feld der korrekt betonten Silbe grün. Diese Methode der visuellen Darstellung hat im Test mit Studenten der Universität Venedig zu positiven Ergebnissen geführt (Delmonte 2000 [29]).

In den letzten Jahren wurden erste Versuche mit einer Methode der „auditiven“ statt visuellen Darstellung der Korrektur der Aussprache des Lehrers durchgeführt. Die Idee ist, die Fehler in der Aussprache des Lerners durch Resynthese zu korrigieren, so dass er die richtige Aussprache in seiner eigenen Stimme hört. Damit könnte der Lerner den Unterschied zwischen korrekter und falscher Aussprache besser wahrnehmen, als wenn er nur die Aussprache des im System aufgenommenen Muttersprachlers hört.

Die Effektivität dieser Methode wurde 1990 von Nagano & Ozawa bewiesen (s. Nagano & Ozawa 1990 [105]). Sie wird in dem System WinPitch LTL (Germain-Rutherford & Martin 2000 [49], Martin 2004 [89], WinPitch [152]) angewendet. Mit WinPitch LTL kann der Lerner die eigene Stimme als Oszillogramm, Intensitäts- und Tonhöhenkurve visualisieren. WinPitch LTL ist ein Autorensystem für den Lehrer. Der Lehrer kann damit die eigene Stimme als Lernmaterial aufnehmen und die aufgenommenen Äußerungen des Lerners manuell segmentieren und nach Korrektur z.B. der Dauer der Segmente und der Tonhöhe resynthetisieren. So kann der Lerner bei seinem erneuten Log-in ins System seine eigene Stimme mit der korrigierten Aussprache hören. Die Vorteile dabei sind, dass der Lehrer manuell Kommentare und Hinweise für den Lerner ins Programm einfügen kann, und dass er die Ergebnisse der Resynthese kontrollieren kann, bevor der Lerner sie bekommt. Die Nachteile von WinPitch LTL sind, dass es voraussetzt, dass der Lehrer mit der Phonetik vertraut ist, und vor allem dass der Lerner das Feedback nicht unmittelbar bekommt, sondern erst viel später, nachdem der Lehrer seine Äußerungen bearbeitet hat. Wie bereits erwähnt (s. Kap. 4.1.3) sollte jedoch das Feedback unmittelbar erfolgen, wenn es effektiv sein soll. Ein unmittelbares Feedback würde man nur erreichen, wenn das System in der Lage wäre, die Sprachsignale automatisch

zu segmentieren oder auf die Segmentation zu verzichten.

Keikichi Hirose von der Graduate School of Frontier Sciences der Universität Tokyo hat ein Programm entwickelt, das Nicht-Japanischsprechenden beibringt, japanische Wörter mit dem richtigen lexikalischen Akzent auszusprechen (Hirose, Gendrin & Minematsu 2003 [62], Hirose 2004 [63]). Hirose entwickelte eine Methode, um lexikalische Akzente — als Gaußskurven modelliert — automatisch zu erkennen. Damit konnte das System die lexikalischen Akzente in der Aussprache des Lerners als richtig oder falsch erkennen. Im letzteren Fall konnte die Aussprache des Lerners mittels der Methode TD-PSOLA (time-domain pitch synchronous overlap add) resynthetisiert werden, so dass sie korrekt wird.

Der jetzige Stand der Forschung ermöglicht die Entwicklung solcher Verfahren, die die Stimme des Sprechers in real-time korrigieren. Diese Methode, die bereits didaktische Vorteile aufweist, soll unbedingt für künftiges CALL angewendet werden.

Zentrales Thema des nun folgenden experimentellen Teils dieser Arbeit ist die Verwendung resynthetisierter und mit korrigierten akustischen Parametern versehenen Stimuli in der Stimme des Lerners als Feedback für das Training der Prosodie des Deutschen. Die experimentellen Untersuchungen wurden folgendermaßen durchgeführt. Zunächst wurde aus der Analyse eines Sprachkorpus italienischer Lernender des Deutschen (Kap. 5) ein systematischer prosodischer Fehler — die Wortbetonung morphologisch komplexer Wörter — identifiziert (Kap. 6). Die nachfolgenden Untersuchungen basieren dann auf der Behebung und auf der Erforschung der Ursachen dieser Fehler. Im Folgenden wird eine Methode entwickelt, um die Resynthese der prosodischen Parameter in den Äußerungen der Lernenden automatisch durchzuführen (Kap. 7). Die nach dieser Methode resynthetisierten Stimuli werden zunächst in Perzeptionsexperimenten von deutschen Muttersprachlern bewertet (Kap. 8.1 und

8.2). Dies dient zur Überprüfung der Korrektheit der Prosodie der resynthetisierten Äußerungen und der unterschiedlichen Relevanz der Korrektur einzelner prosodischer Merkmale. Die resynthetisierten Stimuli werden in einem Aussprachetraining verwendet (Kap. 8.3), und ihre Effektivität im Vergleich mit weiteren Arten von Feedback wird in einem Perzeptionstest mit deutschen Muttersprachlern bewertet (Kap. 8.4). Schließlich wird mittels Perzeptionstests untersucht, ob die systematischen Fehler in der Wortbetonung morphologisch komplexer Wörter perzeptive Ursachen hat (Kap. 8.5 und 8.6).

Zweiter Teil: Experimentelles Vorgehen

5 Datenerhebung

Für die in dieser Arbeit dargestellten Untersuchungen wurde ein Korpus von Äußerungen, die von italienischen Sprechern und einer deutschen Muttersprachlerin produziert wurden, verwendet.

5.1 Textmaterial

Das Textmaterial bestand aus dem deutschen Standardtext „Die Buttergeschichte“ (s. Anhang A.1). Der Text ist bereits für die Datenbank PhonDat aufgenommen und transkribiert worden (s. Pompino-Marschall 1992 [121]). Dies war einer der Gründe für seine Wahl. Außerdem enthält der Text vielfältige prosodische Muster, und er ist für L2-Sprecher einfach verständlich.

Für eine erste Analyse wurde der Text den Sprechern als Ganzes präsentiert. Für die nachfolgenden Experimente wurde der Text in einzelne Sätze geteilt und dann vorgelesen. Zusätzlich wurde eine Liste mit aus dem Text stammenden Wörtern aufgenommen (s. Experimente in Kap. 8.2, 8.3 und 8.4).

5.2 Sprecher

Die deutsche Muttersprachlerin war 32 Jahre alt, ausgebildete Phonetikerin aus München und sprach ohne bayrischen Akzent. Die von ihr ausgesprochenen

Äußerungen dienten als Referenz für eine korrekte deutsche Aussprache.

Als Probandinnen wurden neun Studentinnen deutscher Linguistik an der Universität Sassari in Italien aufgenommen. Die Daten von drei von ihnen konnten nicht verwendet werden. Eine dieser drei Probandinnen las den Text, offensichtlich ohne ihn zu verstehen, wobei das Textverständnis eine Voraussetzung für die Aufnahme war. Die Intonation dieser Probandin war nur von der Interpunktion bestimmt. Die Aussprache der zweiten war zu stark von segmentalen Fehlern geprägt, die die spätere Manipulation der akustischen Parameter problematisch gemacht hätten. Die dritte Probandin sprach fast akzentfrei Deutsch und unterschied sich daher zu sehr von den anderen Sprecherinnen. Aus diesem Grund konnten nur die Daten der übrigen sechs Sprecherinnen verwendet werden. Sie waren zwischen 20 und 25 Jahre alt und hatten zwischen einem und sieben Jahren lang Deutsch gelernt.

Zusätzlich wurden sechs in München wohnhafte italienische Sprecherinnen im Alter zwischen 24 und 42 Jahren aufgenommen. Die Dauer ihres Aufenthalts in Deutschland war unterschiedlich und lag zwischen zwei Monaten und zwölf Jahren. Die Probandin, die seit weniger als einem Jahr in Deutschland lebte, hatte jedoch bereits zwei Jahre Deutschunterricht.

Alle Probandinnen waren Sprecherinnen der sardischen Varietät des Italienischen. In München wurden ausschließlich sardische Sprecherinnen für die Aufnahme gesucht, um die Stichprobe homogen zu halten. Die Intonation in den regionalen Varietäten des Italienischen ist unterschiedlich, und dies hätte zu unterschiedlichen Transferphänomenen in die L2 führen können.¹

Für die in 8.3 und 8.4 beschriebenen Experimente wurden später weitere italienischsprachige Probanden beiderlei Geschlechts und unterschiedlicher re-

¹Zu einer kontrastiven Analyse der Prosodie des Sardischen und des Italienischen s. Schirru 1992 [132], zur Intonation der sardischen Varietät des Italienischen Schirru 1981 [131] und Sardelli & Marotta [130] zur Relevanz prosodischer Parameter für die Erkennung regionaler Varietäten des Italienischen.

gionaler Herkunft aufgenommen. Die in diesem Kapitel beschriebene Analyse betrifft nur die Sprachsignale der sechs Sprecherinnen aus Sassari und der sechs aus München.

5.3 Aufnahme

Vor der Aufnahme wurde den Sprecherinnen eine Version des Textes mit Fußnoten ausgehändigt, die Übersetzungen von Wörtern und Sätzen enthielt. Der Aufnahme der Studentinnen aus Sassari ging zusätzlich eine Textanalyse voran, die das Textverständnis während des Lesens garantieren sollte.

Die Studentinnen aus Sassari wurden in einem professionellen Aufnahmestudio in Sassari, und die Sprecherinnen aus München in dem reflexionsarmen Raum des Instituts für Phonetik und Sprachverarbeitung der Universität München aufgenommen. In beiden Fällen wurde die Aufnahme mit einem Mikrofon Neumann TLM 103 bei 96 KHz Abtastrate und 24 Bit Auflösung durchgeführt.

Die Aufnahmen sollten möglichst reflexionsarm sein, denn Nachhall in den Daten hätte die Resynthese beeinträchtigt. Um die Reflexion der Aufnahmekabine des Studios in Sassari zu dämpfen, wurden absorbierende Materialien an die Wände gehängt und auf den Boden gelegt. Die Aufnahmequalität war schließlich mit der im reflexionsarmen Raum in München vergleichbar, und die Daten konnten problemlos resynthetisiert werden.

5.4 Datenbearbeitung

Die Sprachsignale der zwölf italienischsprachigen Probandinnen und die der deutschen Muttersprachlerin wurden manuell segmentiert und etikettiert. Damit konnten die segmentalen Fehler – Einfügungen, Ersetzungen und Aus-

lassungen – identifiziert werden. Im Folgenden wird die Segmentations- und Etikettierungsmethode der Daten beschrieben.

5.4.1 Segmentation und Etikettierung

Die Sprachsignale wurden mit dem Tool ProFIS (s. Pfitzinger 2007 [119]) segmentiert. Die Phone wurden manuell, die Silben hingegen automatisch anhand der Sonoritätsskala segmentiert.²

Die Segmentation und Etikettierung basieren auf dem Verfahren, welches für die Datenbank PhonDat verwendet wurde (s. Pompino-Marschall 1992 [121]). Nach dieser Prozedur wird für die Etikettierung nicht das IPA- sondern das SAMPA-Alphabet (Speech Assessment Methods Phonetic Alphabet, s. SAMPA 2005 [128] und Wells 1997 [149]) für eine einfachere maschinelle Verarbeitung der Transkription verwendet.

Zu den Phonsymbolen des PhonDat Zeicheninventars (s. Pompino-Marschall 1992 [121, S.116f] und Pfitzinger 2002 [115, S.370]) wurde auch das SAMPA-Symbol [J] für den palatalen Nasal [ɲ] verwendet (s. SAMPA Italian 1996 [129]), der zum Phoneminventar des Italienischen gehört. Der Grund dafür war, dass einige italienische Probandinnen die Grapheme <gn> wie im Italienischen als palatalen Nasal statt als zwei Laute — stimmhafter velarer Plosiv und alveolarer Nasal — aussprachen (z.B. <resigniert> als [r ε z ɪ ɲ i:ɐ t]).

Eine Besonderheit der Segmentation sind die Diphthonge des vokalisiertes /r/ ([ʀ] in IPA-, [6] in SAMPA-Notation), die in einem Segment zusammengefasst wurden. Damit werden unsichere Segmentgrenzen zwischen [ʀ] und dem vorherigen Vokal vermieden (s. van Dommelen 1992 [145, S.201]).

Die Etikettierung wurde auf Basis der für die Datenbank PhonDat kodierten

²Die Silbenkerne wurden anhand der manuellen Phonsegmentierung mittels eines Perl-Skripts von Uwe Reichel ermittelt.

kanonischen Form durchgeführt (s. Anhang A.2 für die kanonische Transkription der Sätze der „Buttergeschichte“). Es wurden die Einfügungen und die Ersetzungen gegenüber der kanonischen Form notiert, zum Beispiel [-@], um die Schwa-Epenthese zu markieren, oder [x-k] (beide Beispiele in SAMPA-Notation), um die Ersetzung des velaren Frikativs mit dem homorganischen Plosiv zu markieren.

Anders als in der Datenbank PhonDat wurden die Auslassungen nicht etikettiert. Sie wurden später beim Vergleich zwischen der durchgeführten Transkription und der kanonischen Form automatisch berechnet. Zudem wurden keine Funktionswörter, Satz-, Wort- und Morphemgrenzen gekennzeichnet und keine weitere Diakritika verwendet, die für unseren Zweck nicht notwendig waren.

Die Pausen wurden ebenfalls segmentiert und etikettiert.

Im folgenden Kap. 6 werden die segmentalen und Betonungsfehler der italienischen Sprecherinnen dargestellt.

6 Aussprachefehler italienischsprachiger Probanden

Anhand der Segmentierung und Etikettierung der Daten zeigte sich eine große Vielfalt an segmentalen Fehler. Zudem betonten die italienischen Sprecherinnen die morphologisch komplexen Wörter, die auf der Antepänultima betont werden sollten, systematisch fehlerhaft auf der Pänultima bzw. auf ihrem Zweitglied.

6.1 Segmentale Fehler

Wenn die aus Ersetzungen und Einfügungen resultierenden Labels mitberechnet werden, steigt die Anzahl der zur Transkription der Sprachdaten der zwölf italienischen Sprecherinnen verwendeten Labels auf fast 300 unterschiedliche, wobei die Anzahl der Phoneme in der kanonischen Transkription lediglich 41 beträgt.

Da, um die Resynthese zu ermöglichen, ursprünglich die Phone der Äußerungen der italienischen Probandinnen den entsprechenden Phonen der deutschen Muttersprachlerin anhand der Segmentierung hätten angeglichen werden sollen, sollten im Idealfall die Äußerungen der italienischen Probandinnen die sel-

be Anzahl an Segmenten wie die Äußerungen der deutschen Muttersprachlerin aufweisen. Aus diesem Grund wurden Schwa-Einfügungen der italienischen Probandinnen, wenn der Schwa nicht besonders ausgeprägt war, zusammen mit dem vorangehenden Konsonanten in einem Etikett zusammengefasst (zum Beispiel: Markierung der Ersetzung des alveolaren Plosivs durch Plosiv und Schwa [t-t@], statt Markierung von Plosiv und Schwa-Einfügung [t -@]).¹ Das erhöhte die Etikettenanzahl um ein zusätzliches Dutzend. Auch wenn dies berücksichtigt wird, bleibt die Gesamtzahl an Etiketten beachtlich.

Der häufigste Fehler überhaupt ist die Schwa-Epenthese mit 273 Vorkommnissen. Zu den häufigsten Fehler zählen ferner die Ersetzung von glottalem Plosiv durch den glottalen Frikativ [Q-h] und umgekehrt [h-Q] (32 bzw. 29 Vorkommnisse), sowie die Realisierung von velarem Nasal und nachfolgendem velaren stimmhaften Plosiv anstelle des bloßen velaren Nasals [N-Ng] (31 Vorkommnisse). Häufig waren ebenfalls die Ersetzungen von vorderen gerundeten Vokalen durch die entsprechenden hinteren Vokale (zum Beispiel [y:-u:]) oder umgekehrt und die Ersetzungen von Frikativen (wie [C-S] oder [C-x]). Die Fehler werden möglicherweise von der unterschiedlichen Phonotaktik und von der Tatsache, dass viele Phoneme des Deutschen — wie Vokale oder Frikative — im Phoneminventar des Italienischen nicht existieren, verursacht.

Tabelle 6.1 zeigt die Abweichungen gegenüber der kanonischen Form — Ersetzungen, Einfügungen und Auslassungen — für die zwölf italienischen Sprecherinnen und die deutsche Muttersprachlerin. Bemerkenswert ist, dass die meisten Schwa-Einfügungen wegen der oben beschriebenen Segmentationsstrategie zu den Ersetzungen zählen. Die Abweichungen gegenüber der kanonischen Form von der deutschen Muttersprachlerin sind in diesem Fall keine Fehler, sondern reguläre Reduktionsphänomene oder unterschiedliche /r/-Realisierungen.

¹Alle Beispiele dieses Kapitels werden in SAMPA-Transkription gegeben.

Sprecherin	Tot. Abw.	Ersetz.	Einfüg.	Auslass.	Tot. Labels
deutsche Spr.	72	35	1	36	990
Italienische Sprecherinnen aus München					
Spr. 1	194	153	7	34	999
Spr. 2	151	119	0	32	993
Spr. 3	162	129	0	33	993
Spr. 4	155	112	0	43	982
Spr. 5	169	128	8	33	1000
Spr. 6	99	55	0	44	981
Italienische Studentinnen der Universität Sassari					
Spr. 2	147	101	3	43	985
Spr. 3	186	122	34	30	1029
Spr. 4	167	129	10	28	1007
Spr. 7	162	107	7	48	984
Spr. 8	173	100	19	54	990
Spr. 9	97	74	4	19	1010

Tabelle 6.1: Abweichungen gegenüber der kanonischen Transkription in Form von Ersetzungen, Einfügungen und Auslassungen für die deutsche Muttersprachlerin, die sechs italienischsprachigen Studentinnen aus Sassari und die sechs italienischen Sprecherinnen aus München.

Die italienischen Sprecherinnen wendeten vielfältige Strategien an, um gegen die Ausspracheschwierigkeiten anzukommen. Die einzelnen Fehler, die nur einmal im gesamten Korpus der zwölf italienischen Sprecherinnen realisiert wurden, belaufen sich auf ca. einhundert verschiedene.

Abb. 6.1 zeigt das Sprachsignal von „auf acht kam“ der Äußerung „Je näher der Zeiger auf acht kam, desto unruhiger wurden die Leute.“, die von Sprecherin 7 der Studentinnen der Universität Sassari ausgesprochen wurde. Die dunklen Markierungen sind die Phonemgrenzen, während die hellen Markierungen die Silbenkerne zeigen. Das Signal enthält zwei häufige segmentale Fehler: Ersetzung des glottalen Plosiv durch den glottalen Frikativ [Q-h] zu Beginn der

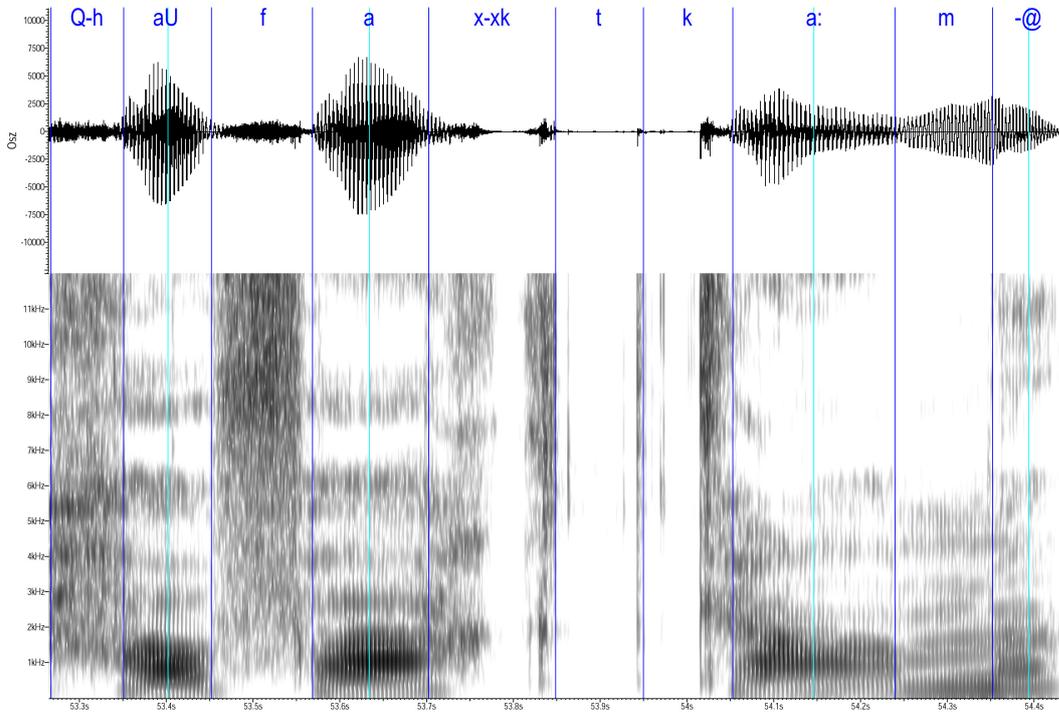


Abbildung 6.1: Beispiel von segmentalen Fehlern: „auf acht kam“ der Äußerung „Je näher der Zeiger auf acht kam, desto unruhiger wurden die Leute.“ von Sprecherin Nr. 7 aus Sassari.

Äußerung und Schwa-Epenthese [-@] am Wortende. In der Mitte des Signals befindet sich hingegen einer der hundert Einzelfehler, der im Korpus nur einmal und nur von dieser einzigen Sprecherin realisiert wurde. Es handelt sich um die Ersetzung des stimmlosen velaren Frikativs, den es im Phoneminventar des Italienischen nicht gibt, durch einen undefinierten Laut, als [xk] etikettiert. In der Produktion dieses Lautes folgt der Frikation ein im Sonagramm sichtbarer kompletter Verschluss.

6.2 Wortbetonungsfehler

Die Daten zeigten bei fast allen Probanden systematische Fehler der Betonung von morphologisch komplexen Wörtern, wie Komposita und Präfixverben. Die einzige Ausnahme war die italienische Sprecherin Nummer 6 aus München, die auch weniger segmentale Fehler als die anderen Probandinnen gemacht hatte. Ansonsten betonten alle anderen Probandinnen die Pänultima oder das Zweitglied des komplexen Wortes statt die erste Silbe. Tabelle 6.2 zeigt die falsch betonten morphologisch komplexen Wörter, die auf der ersten Silbe betont werden sollten, und die Fehlerhäufigkeit. Die Hälfte der Wörter wurde von mehr als der Hälfte der Sprecherinnen falsch betont. Aufgrund der massiven Häufigkeit dieser Fehler widmen sich die Untersuchungen der Experimente 2, 3 und 4 (s. Kap. 8.2, 8.3 und 8.4) der Korrektur der Wortbetonung mittels Resynthese und einem Training der Wortbetonung.

Wort	Fehler	Wort	Fehler
1. durchlassen	9	9. Lebensmittel	4
2. unruhiger	9	10. anzustellen	3
3. vorlassen	8	11. Buttergeschichte	3
4. aufschließen	7	12. meinetwegen	3
5. Menschenmenge	7	13. anschließen	2
6. ausverkauft	6	14. ebenfalls	2
7. eingetroffen	6	15. Ladentür	2
8. Schimpfwörter	6	16. rücksichtslos	1

Tabelle 6.2: Nach der Häufigkeit geordnete, fehlerhaft auf der Pänultima bzw. auf dem Zweitglied betonte morphologisch komplexe Wörter, die auf der Antepänultima betont werden sollten. Da jedes Wort im Text nur ein Mal vorkommt, sind pro Wort maximal 12 Fehler möglich.

Da die Betonung im Italienischen am häufigsten auf die Pänultima fällt, könnten die Sprecherinnen diese Eigenschaft auf ihre Betonung der deutschen Wörter übertragen haben. Im Italienischen ist aber die Betonung der dritt-

auf Antepänultima statt Pän. bzw. Ultima		auf Pänultima statt Antepänultima		auf Pänultima statt Ultima	
Wort	Fehler	Wort	Fehler	Wort	Fehler
1. Drängelei	3	6. wartenden	3	9. Berlin	4
2. erhalten	2	7. beachtliche	2	10. bereits	1
3. genügend	1	8. forderten	1	11. empört	1
4. Polizei	1				
5. vorhanden	1				

Tabelle 6.3: Weitere fehlerhaft betonte Wörter: a) auf der Antepänultima statt auf der Pänultima, b) auf der Pänultima statt auf der Antepänultima und c) auf der Pänultima statt auf der letzten Silbe.

letzten, viertletzten und sogar fünftletzten Silbe möglich. Aus diesem Grund könnten die Fehler eher von einer unterschiedlichen Interpretation der akustischen Parameter der Dauer und der Grundfrequenz, die die Wortbetonung bestimmen, stammen. Wie bereits oben erwähnt (s. Kap. 2.3.2 auf S.37), kann im Deutschen die betonte Silbe des Zweitglieds eines morphologisch komplexen Wortes, in dem das Erstglied die Hauptbetonung trägt, ihre Quantität bewahren (s. Kohler 1995 [74, S.188]), während im Italienischen die Quantität nicht phonologisch distinktiv und die Dauer das Hauptkorrelat der Betonung ist (s. Bertinetto 1980 [9] und 1981 [10]). Aus diesem Grund könnten italienischsprachige Deutschlernende vermuten, das Zweit- statt des Erstglieds eines morphologisch komplexen Wortes sei betont. Um zu überprüfen, ob die Fehler perzeptive Ursachen haben, wurde das Experiment 5 durchgeführt (s. Kap. 8.5). Die Hypothese wird außerdem durch die Tatsache gestützt, dass auch Sprecherinnen, die seit mehreren Jahren in München lebten, die Wortbetonungsfehler systematisch machten und dies nie bemerkten.

Tabelle 6.3 stellt weitere weniger häufige Betonungsfehler dar. Diese Fehler wurden hauptsächlich von den Studentinnen aus Sassari gemacht und basieren möglicherweise zum Teil auf mangelnden lexikalischen Kenntnissen.

Möglicherweise handelt es sich bei morphologisch komplexen Wörtern im Deutschen um besondere Fälle, in denen die Grundfrequenz als Korrelat der Betonung eine größere Rolle als bei anderen Wörtern spielt (s. Kap. 2.3.1). Dies würde erklären, warum italienischsprachige Deutschlernende Schwierigkeiten haben, die korrekte Betonung morphologisch komplexer Wörter im Deutschen zu lernen. Da, wie bereits oben erwähnt, die Betonung im Italienischen hauptsächlich von der Dauer abhängt, könnte es für italienische Sprecher schwierig sein, die Betonung auf Tonhöheunterschiede zurückzuführen.

7 Resynthese

Die prosodischen Parameter lokale Sprechgeschwindigkeit, Intonation und Intensität der Äußerungen der deutschen Muttersprachlerin wurden auf die entsprechenden Äußerungen der italienischen Sprecher übertragen. Die durch die Korrektur eines oder mehrerer Parameter resynthetisierten Stimuli wurden in den Experimenten 1 und 2 sowie 3 und 4 verwendet. Während sie in den ersten beiden Experimenten dazu dienten, den Einfluss der Korrektur der Parameter auf die Wahrnehmung des Sprechrhythmus und der Wortbetonung durch deutsche Muttersprachler zu untersuchen (s. Kap. 8.1 und 8.2), wurden sie in den Experimenten 3 und 4 im Rahmen eines Aussprachetrainings eingesetzt (s. Kap. 8.3 und 8.4).

Die Resynthese wurde mit dem Tool ProFIS (PROmpts For Information Systems) durchgeführt (s. Pfitzinger 2007 [119]).

7.1 Resynthese der lokalen Sprechgeschwindigkeit

Für die Resynthese der lokalen Sprechgeschwindigkeit wurden zwei Methoden angewendet: a) die Anpassung der perzipierten lokalen Sprechgeschwindigkeit (PLSR) nach dem Modell von Pfitzinger 2001 [116] (s. Kap. 2.2.1), und b) die Angleichung der Dauer der Phone.

7.1.1 Resynthese der perzipierten lokalen Sprechgeschwindigkeit PLSR

In dem ersten Experiment wurde lediglich die perzipierte lokale Sprechgeschwindigkeit (PLSR) der Stimuli manipuliert (s. Kap. 8.1). Die Alignierung der Signale der italienischen Sprecherinnen und der deutschen Muttersprachlerin sollte mittels Segmentierung erfolgen. Aus diesem Grund wurde zunächst versucht, die Signale der italienischen Sprecherinnen so zu segmentieren, dass die Anzahl ihrer Phone der des Signals der deutschen Muttersprachlerin entsprach (s. dazu Kap. 6.1). Die Äußerungen waren jedoch dermaßen unterschiedlich, dass eine andere Lösung für die Alignierung gesucht werden mußte.

Schließlich wurde mittels eines Dynamic Time Warping-Algorithmus (s. Vintsyuk 1968 [148]) die Segmentierung der deutschen Muttersprachlerin dem Signal jeder italienischen Sprecherin angepasst. Auf diese Weise entstand eine eindeutige Entsprechung zwischen den Segmenten der Äußerungen der deutschen Muttersprachlerin und denen jeder italienischen Sprecherin. Die gesamte vom Algorithmus erzeugte Segmentierung wurde dann manuell überprüft und korrigiert.

Die Übertragung der Segmentierung war jedoch problematisch, denn die Segmente des Signals der deutschen Muttersprachlerin und die der Signale der italienischen Sprecherinnen waren meistens verschieden. Beispielsweise waren in den Signalen der italienischen Sprecherinnen viele Einfügungen (vor allem Schwa-Einfügungen) und Auslassungen (wie Reduktionen von Konsonantenklustern). In diesen Fällen wurde trotzdem darauf geachtet, dass es Eineindeutigkeit zwischen den Segmenten in der Äußerung der deutschen Muttersprachlerin und denen in den entsprechenden Äußerungen der italienischen Sprecherinnen gab. Deswegen wurden Einfügungen seitens der italienischen Sprecherinnen mit dem vorherigen Phon in einem Segment zusammengefasst (zum Beispiel wurden [t] und [ə] in [t] zusammengefasst), und es wurden zur

Not einige überkurze Segmente markiert, obwohl die entsprechenden Phone von den italienischen Sprecherinnen ausgelassen worden waren (zum Beispiel wurden Glottalverschlüsse [Q] der deutschen Sprecherin für die ersten zwei bis drei Perioden des Vokals der italienischen Sprecherin markiert, selbst wenn es im Signal der italienischen Sprecherin weder Glottalverschluss noch Laryngalisierung gab).

Anhand der Segmentierung wurden Silben- und Phonrate der Äußerung der italienischsprachigen Probandinnen und der deutschen Muttersprachlerin ermittelt und dadurch die perzipierte lokale Sprechgeschwindigkeit — PLSR nach dem Modell von Pfitzinger 2001 [116] (s. dazu oben 2.2.1) — der Äußerung berechnet.

Die Dauer der Phone in den Äußerungen jeder italienischen Sprecherin wurde modifiziert, so dass ihre PLSR-Konturen denen der entsprechenden Äußerungen der deutschen Muttersprachlerin entsprachen.

Die lokalen Veränderungen der Sprechgeschwindigkeit wurden mit Hilfe eines PSOLA-ähnlichen Verfahrens durchgeführt (PSOLA: *pitch synchronous overlap and add*, s. Moulines & Charpentier 1990 [102]).

Abb. 7.1 zeigt die Anpassung der Segmentierung der deutschen Muttersprachlerin an das Signal einer italienischen Sprecherin. Die Pause vor dem Wort [d o x] der italienischen Sprecherin, die im Signal der deutschen Muttersprachlerin nicht vorhanden ist, wurde in dem Segment [d] mit dem stimmhaften alveolaren Plosiv zusammengefasst. Die italienischen Probandinnen sprachen in den meisten Fällen, wie in diesem Beispiel, langsamer als die deutsche Muttersprachlerin. Zudem ist die relative Geschwindigkeit der Einzelwörter bei den italienischen Probandinnen und bei der deutschen Muttersprachlerin unterschiedlich. In der Tat sprach die deutsche Muttersprachlerin die Funktionswörter schneller aus. Die PLSR-Kurve unseres Beispiels zeigt zwei Peaks bei den Wörtern [d i:] und [n u n] und, in der Phase des *pre-final lengthening*,

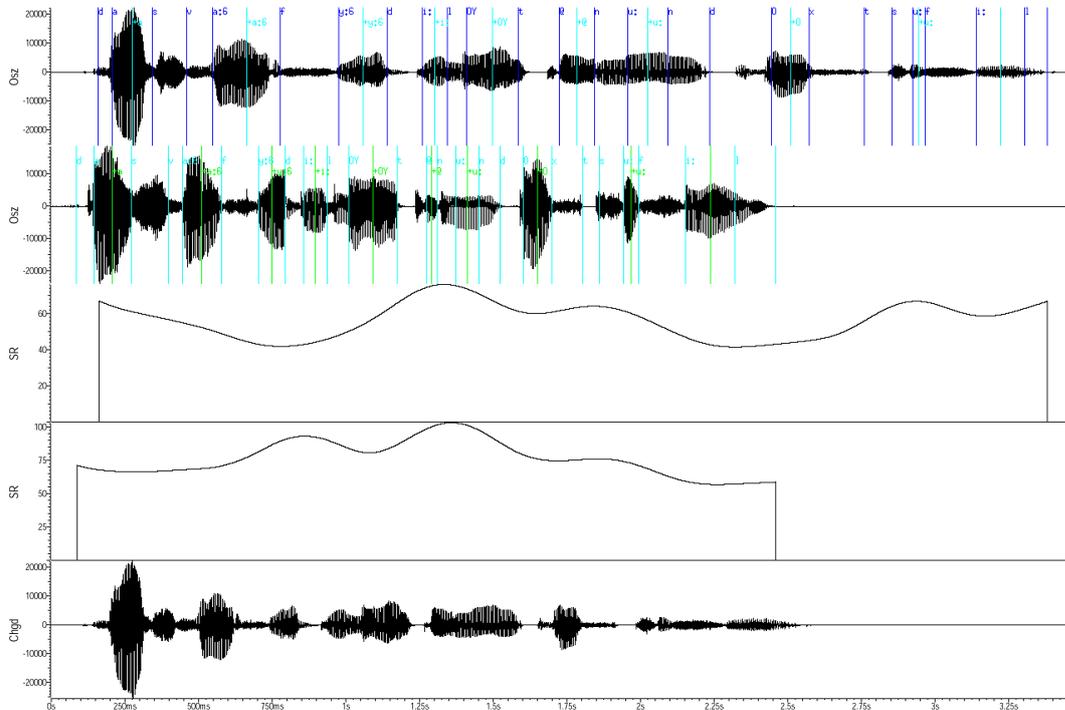


Abbildung 7.1: Äußerung: „*Das war für die Leute nun doch zu viel!*“. Von oben nach unten: *Osz* = Signal der italienischen Sprecherin Nr. 9 aus Sassari, *Osz* = Signal der deutschen Muttersprachlerin, *SR* = PLSR-Kontur der italienischen Sprecherin, *SR* = PLSR-Kontur der deutschen Muttersprachlerin, *Chgd* = Resynthetisiertes Signal der italienischen Sprecherin mit der PLSR-Kontur der deutschen Muttersprachlerin.

einen kleineren Peak beim Wort [t s u:].¹ Hingegen verlangsamt sich die Sprache der italienischen Sprecherin kurz nach [n u n] und beschleunigt sich am Ende der Äußerung wieder. Anders als die deutsche Muttersprachlerin unterschieden die italienischen Sprecherinnen meistens nicht zwischen bedeutungstragenden Wörtern und Funktionswörtern und sprachen die letzteren in einem ähnlich

¹In der Transkription sind die in einem Segment zusammengefassten Symbole zusammengeschrieben. Die zu unterschiedlichen Segmente gehörenden Symbole sind hingegen durch ein Leerzeichen getrennt.

langsamen Tempo aus.

7.1.2 Resynthese der Dauer der Phone

Nach der Resynthese der perzipierten lokalen Sprechgeschwindigkeit (s. oben 7.1.1) bekamen die Äußerungen der italienischen Sprecherin die PLSR-Kontur der Äußerungen der deutschen Muttersprachlerin. Wobei dies nicht bedeutet, dass die einzelnen Segmente gleich lang wurden. Für die weiteren Experimente 2, 3 und 4 (s. Kap. 8.2, 8.3 und 8.4) war die Synchronie der Phone der zu manipulierenden Signale mit den Phonen der entsprechenden Signale der deutschen Muttersprachlerin notwendig, denn nur so konnte die Intonation der deutschen Muttersprachlerin auf die anderen Signale übertragen werden.

Aus diesem Grund wurde die Resynthese für die weiteren Experimente nicht mittels Segmentierung und Übertragung der PLSR-Kurve durchgeführt, sondern nach zeitlicher Alignierung der Laute im Sprachsignal jeder italienischen Sprecherin mit den Lauten im Sprachsignal der deutschen Muttersprachlerin durch einen Dynamic Time Warping (DTW) Algorithmus.

Die Resynthese der lokalen Sprechgeschwindigkeit war also möglich, ohne dass eine manuelle Segmentation der Sprachdaten durchgeführt werden musste. Der DTW-Algorithmus funktionierte mit Einzelwörtern und kurzen Sätzen in der Regel fehlerfrei. Nur bei längeren Sätzen war häufig die Angleichung der Pausen des Sprachsignals der italienischen Sprecherin an die des entsprechenden Signals der deutschen Muttersprachlerin erforderlich. Das war vor allem notwendig, wenn die Sprachsignale der italienischen Sprecherin und der Muttersprachlerin von erheblich unterschiedlicher Dauer waren, zum Beispiel wenn die italienische Sprecherin besonders langsam sprach.

Die Dauer der Phone des Sprachsignals jeder italienischen Sprecherin wurde auch in diesem Fall mittels eines PSOLA-Algorithmus modifiziert, so dass

sie der Dauer der Phone des Sprachsignals der deutschen Muttersprachlerin entsprach, und beide Signale dadurch synchron wurden.

7.2 Resynthese der Intonation

Die Resynthese der Intonation wurde mit einem insel-gesteuerten AMDF-basierten (Average Differential Magnitude Function) Algorithmus durchgeführt. Damit die Stimmlage der korrigierten Version der italienischen Sprecherin zur Originalversion passte, wurde der aus der Äußerung der deutschen Muttersprachlerin extrahierte Grundfrequenzverlauf durch seinen eigenen Mittelwert dividiert und anschließend mit dem Mittelwert des Grundfrequenzverlaufs der Äußerung der italienischen Sprecherin multipliziert, bevor er auf ihr Signal übertragen wurde. Mit dieser Methode und durch die Reduktion der Grundfrequenzbandbreite in der Detektion konnten wir den F0-Verlauf der deutschen Sprecherin auch auf die Signale des männlichen italienischen Sprechers übertragen.

Für Experiment 2 (s. Kap. 8.2) sollten unter anderem Stimuli erzeugt werden, in denen ausschließlich die Intonation korrigiert werden sollte. Jedoch war eine direkte Übertragung der Intonation nicht ohne Weiteres möglich, weil die Originalsignale der deutschen Muttersprachlerin und der italienischen Sprecherin nicht synchron waren. In diesem Fall wurde zuerst die Dauer der Phone und dann die Intonation korrigiert. Anschließend wurde die Dauer der Phone wieder in den ursprünglichen Zustand gebracht.

7.3 Resynthese der Intensität

Für die Erzeugung der resynthetisierten Stimuli für das Aussprachetraining war zusätzlich zur Resynthese der Dauer und der Intonation auch die Resynthese der Intensität notwendig, denn die Intensitätsunterschiede waren in den überbetonten Stimuli wahrnehmbar. Zudem sollte ein Intonationsgipfel der betonten Silbe idealerweise immer auch einen Intensitätsgipfel mit sich bringen.

Die Einhüllende der Amplitude der deutschen Muttersprachlerin wurde auf das Signal des italienischen Sprechers, in dem die Amplitude normalisiert worden war, übertragen.

7.4 Algorithmus zur Übertragung der Prosodie eines L1-Sprechers auf die Äußerung eines L2-Sprechers

Abb. 7.2 auf der nächsten Seite zeigt das Blockdiagramm des Algorithmus, der die prosodischen Parameter der Dauer der Phone, der Intonation und der Intensität von einem L1-Sprecher auf einen L2-Sprecher überträgt.

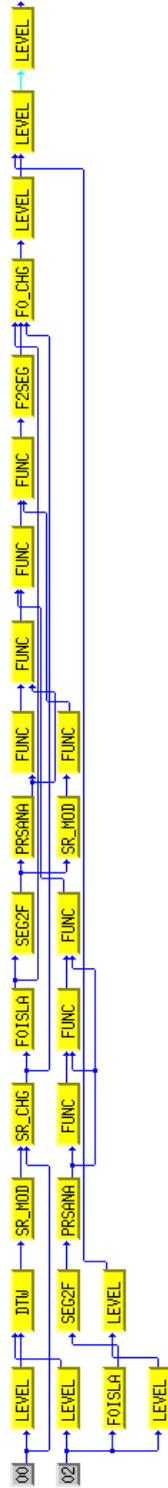


Abbildung 7.2: Blockdiagramm des Algorithmus zur Übertragung der prosodischen Parameter eines L1-Sprechers auf einem L2-Sprecher: 00 = das Signal des L2-Sprechers, 02 = das Signal des L1-Sprechers. Die Verbindungspfeile zwischen den Modulen stellen die Zwischenergebnisse dar.

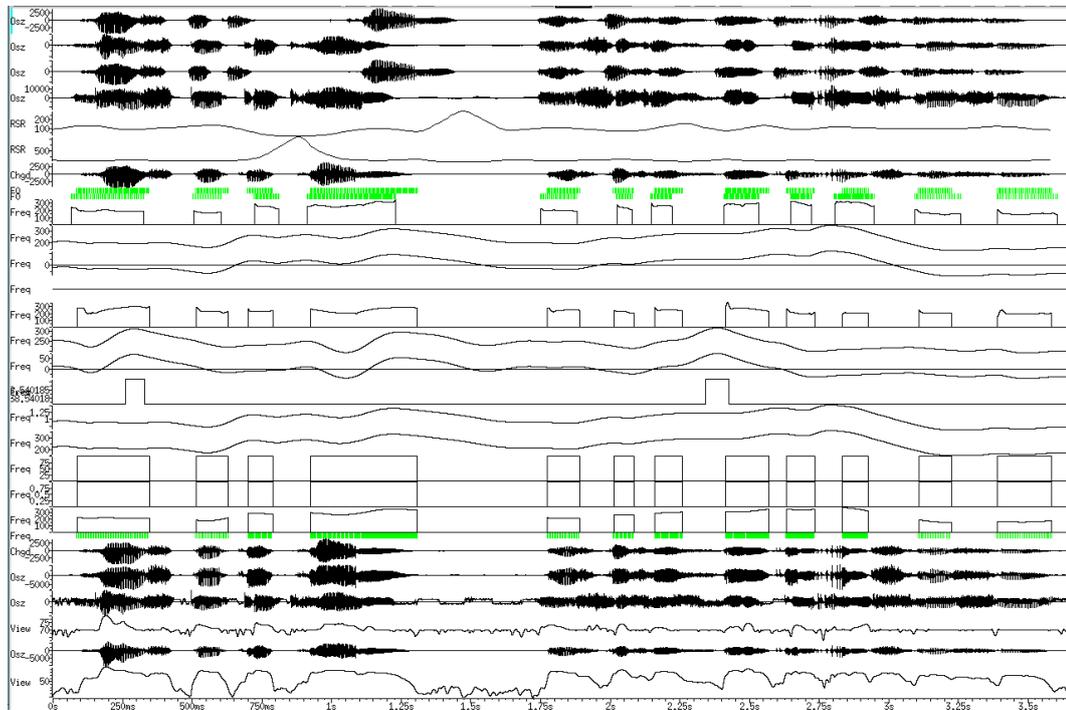


Abbildung 7.3: Äußerung: „*Wer später kam, musste sich hinten anschließen.*“ Erstes Signal von oben = Signal der italienischen Sprecherin Nr. 24 aus München, zweites Signal von oben = Signal der deutschen Muttersprachlerin und vorletztes Signal unten = Endergebnis der Resynthese.

Jedes Programmmodul (die Kästchen in Abb. 7.2) erzeugt ein Zwischenergebnis, und zwar ein Signal, das vom nächsten Modul bearbeitet wird. Diese Zwischenergebnisse gemeinsam mit dem Gesamtergebnis werden in Abb. 7.3 dargestellt. Die vorletzte Spur unten zeigt das Endergebnis der Resynthese: Das Signal der italienischen Sprecherin nach Korrektur der Dauer der Phone, der Intonation und der Intensität. Das Zögern der italienischen Sprecherin zwischen den Wörtern „*Wer später*“ und „*kam*“ wird vom Algorithmus problemlos behoben (vgl. die erste mit der vorletzten Spur in Abb. 7.3).

Der komplette Algorithmus mit den Zwischenschritten wird als Unix-Shell-

Programm in Anhang B.1 dargestellt.

Die sprecherspezifischen Eigenschaften bleiben nach der Korrektur der prosodischen Parameter erhalten, und das ist gewünscht, denn die damit verbundene Reduktion der phonetischen Variabilität vereinfacht den Vergleich zwischen der unkorrekten Betonung im Originalsignal und der korrekten Betonung im resynthetisierten Signal. Die Vokal- und Stimmqualität bleiben ebenfalls erhalten. In der Tat wäre eine Übertragung des Reduktionsverhaltens der unbetonten Silben wünschenswert, denn die Vokalqualität ist akustisches Korrelat der Betonung im Deutschen (s. Dogil & Williams 1999 [33], Mooshammer & Harrington 2005 [99]), aber der jetzige Stand der Forschung ermöglicht sie noch nicht.

8 Experimente zum computergestützten Lernen der Wortbetonung im Deutschen bzw. deren Wahrnehmung durch italienische Sprecher

Aus unserem Hauptziel, neue Methoden zu entwickeln, mit denen italienischsprachige Deutschlernende die Prosodie des Deutschen mit Hilfe des Computers erlernen können, resultierten insgesamt fünf Experimente und eine Voruntersuchung zu einem sechsten, die neben dem Hauptthema weitere Forschungsfelder eröffnet haben.

Die beiden ersten Perzeptionsexperimente testen, welchen Einfluss die Korrektur der prosodischen Parameter der Äußerungen der italienischen Sprecher auf die Wahrnehmung von deutschen Muttersprachlern hat. Insbesondere interessierte uns, a) welche Wirkung die Korrektur einzelner akustischer Parameter im Vergleich zu den anderen hat, und b) ob diese Wirkung signifikante Verbesserungen der Bewertung der Wortbetonung verursacht, so dass derart

gemeinte resynthetisierte Stimuli in einem Aussprachetraining verwendet werden können. Im ersten Experiment werden resynthetisierte Stimuli mit lediglich korrigierter lokaler Sprechgeschwindigkeit verwendet. 16 deutsche Muttersprachler sollen die Verbesserung des Sprechrhythmus von 88 resynthetisierten Äußerungen im Vergleich zu ihren entsprechenden Originaläußerungen bewerten. Im zweiten Experiment werden die Dauer der Phone und die Intonation manipuliert, und es wird in einem Perzeptionstest mit 31 deutschen Muttersprachlern, die insgesamt je 256 Stimuli bewerten, getestet, welcher Parameter die Betonungswahrnehmung effektiver beeinflusst.

Das dritte und das vierte Experiment bestehen aus einem Aussprachetraining zur Wortbetonung für italienische Sprecher und aus einer Evaluierung der Trainingsmethoden durch einen Perzeptionstest. Es werden zwei Gruppen aus je sechs italienischen Sprechern trainiert, die eine Gruppe mit natürlichen, die andere mit resynthetisierten Stimuli. Zudem wird das Training sowohl mit normal betonten als auch mit überbetonten Stimuli durchgeführt. In dem Perzeptionstest bewerten 37 deutsche Muttersprachler die Wortbetonung in den von den italienischen Sprechern vor und nach dem Training ausgesprochenen Äußerungen (insgesamt 255 Stimuli).

Mit dem fünften Experiment wollten wir testen, ob die systematischen Wortbetonungsfehler der Italiener perzeptive Ursachen haben. Eine solche Erkenntnis wäre hilfreich, um Trainingsmaßnahmen zielgerichtet vorzubereiten. Im fünften Experiment müssen 23 italienischsprachige Probanden und eine Kontrollgruppe von 6 deutschsprachigen Probanden lernen, drei Wortpaare aus deutschen bzw. pseudo-deutschen Wörtern zu unterscheiden: a) ein Phonemminimalpaar, b) ein Betonungsminimalpaar, und c) ein Paar, das sich sowohl hinsichtlich der Betonung als auch hinsichtlich der Vokalquantität in der Pänultima unterscheidet. Im Test kommen für das Phonemminimalpaar 12 Stimuli, für die 10 Betonungs- und 10 Betonungs-Quantitätswortpaare 180 Stimuli zum Einsatz. In der Voruntersuchung zu dem geplanten sechsten Experiment wer-

den Aufnahmen von Nichtwörtern durchgeführt, um das fünfte Experiment auf gleicher Basis für italienische und deutsche Versuchspersonen zu wiederholen.

8.1 Experiment 1: Einfluss der Korrektur der perzipierten lokalen Sprechgeschwindigkeit auf die Wahrnehmung des Sprechrhythmus

Das Ziel des ersten Experiments war es zu testen, ob die Korrektur der perzipierten lokalen Sprechgeschwindigkeit den Sprechrhythmus von Äußerungen italienischer Sprecher verbessern kann. In der Tat sprachen die italienischen Sprecherinnen Funktionswörter häufig genauso langsam wie Inhaltswörter (s. dazu Kap. 7.1), während die deutsche Muttersprachlerin erstere, da sie keine bedeutungstragenden Elemente sind, deutlich schneller aussprach.

8.1.1 Methode

Für das Experiment wurden die Äußerungen der italienischen Sprecherinnen in der lokalen Sprechgeschwindigkeit den entsprechenden Äußerungen der deutschen Muttersprachlerin angeglichen.

8.1.1.1 Stimuli

Der Perzeptionstest wäre zu umfangreich geworden, wenn wir sämtliche Sätze der Buttergeschichte in den Versionen aller zwölf italienischen Sprecherinnen als Stimuli benutzt hätten, so dass wir eine Auswahl treffen mussten. Da die segmentalen Fehler das Ergebnis der Resynthese verschlechtert hätten, wurde beschlossen, lediglich die Äußerungen zu benutzen, die die wenigsten segmen-

talen Fehler aufwiesen.

Die Fehler, die die größten Probleme bei der Resynthese verursacht hätten, wären die Einfügungen und die Auslassungen gewesen, da sie eine Vermehrung bzw. Verminderung der Anzahl der Phone gegenüber der Kanonischen Form und wahrscheinlich auch gegenüber der Transkription der Äußerungen der deutschen Muttersprachlerin verursachen. Mit der Übertragung der Segmentierung der Äußerungen der deutschen Muttersprachlerin auf die entsprechenden Äußerungen der italienischen Sprecherinnen hatten wir die eindeutige Korrespondenz zwischen den Segmenten in den Äußerungen erreicht (zur verwendeten Segmentierung für die Resynthese der PLSR s. Kap. 7.1.1), jedoch verbargen sich bei Einfügungen mehr als ein Phon auf demselben Segment, bzw. wurde bei einer Auslassung dasselbe Phon auf zwei Segmente verteilt, was zur Berechnung einer weniger wahrheitsgetreuen PLSR-Kontur geführt hätte. Aus diesem Grund sollten vor allem Einfügungen und Auslassungen vermieden werden. Die Ersetzungen hätten nicht die Berechnung der PLSR-Kontur beeinflusst, und zudem sollten einige Ersetzungen gegenüber der kanonischen Transkription, wie die Ersetzung von [6] (vokalisiertem /r/) durch [r], nicht als gravierende Fehler, sondern als mögliche Aussprachevariationen betrachtet werden.

Nach diesen Überlegungen wurde für jeden Satz und für jede italienische Sprecherin eine Punktzahl aus der gewichteten Summe der Fehler berechnet, die helfen sollte, die Sätze anhand der Schwere der Fehler zu klassifizieren, und bei der die Ersetzungen mit einem niedrigeren Koeffizienten multipliziert werden:

$$\text{Punktzahl} = \text{Einfügungen} + \text{Auslassungen} + \\ (\text{Ersetzungen} - \text{/r/ Ersetz.}) * 0.5 + \text{/r/ Ersetzungen} * 0.3$$

Der höchsten Punktzahl wurde dann der Wert 1 zugewiesen, und die anderen Werte wurden entsprechend skaliert, so dass eine Punktzahl zwischen 0 und 1

erreicht werden konnte. Für den Perzeptionstest wurden dann die Äußerungen mit der niedrigsten Punktzahl ausgewählt, die von den Sprecherinnen mit insgesamt der niedrigsten Punktzahl ausgesprochen worden waren. Tab. 8.1 stellt die Rangliste der Sätze anhand der Punktzahl dar. Die ersten elf Sätze wurden für den Perzeptionstest ausgewählt (s. Satzliste in Anhang C.1).

	Sätze	Punktzahl	Fehl.	Ers.	Ers-/r/	Einf.	Ausl.	Phoneme
1.	s00	0.02	6	4	1	0	2	167
2.	s12	0.13	36	32	18	3	1	327
3.	s06	0.23	49	33	23	4	12	421
4.	s07	0.31	70	51	31	4	15	574
5.	s15	0.32	62	37	27	0	25	560
6.	s16	0.36	82	63	42	2	17	765
7.	s02	0.50	103	73	62	4	26	719
8.	s03	0.55	106	62	46	5	39	954
9.	s01	0.56	117	80	57	10	27	815
10.	s04	0.60	127	87	58	2	38	1108
11.	s14	0.60	132	95	64	1	36	719
12.	s08	0.67	141	100	83	16	25	901
13.	s10	0.73	172	138	97	8	26	1022
14.	s13	0.76	158	110	85	8	40	839
15.	s05	0.77	158	103	71	6	49	919
16.	s11	0.95	202	144	109	13	45	982
17.	s09	1.00	213	152	113	7	54	1149

Tabelle 8.1: Rangliste der Sätze der Buttergeschichte gemäß der Punktzahl, die anhand der Abweichungen gegenüber der kanonischen Transkription berechnet wurde. Die kanonische Transkription aller Sätze ist in Anhang A.2 aufgelistet. *Fehl.* = gesamte Fehler, *Ers.* = gesamte Ersetzungen, *Ers-/r/* = gesamte Ersetzungen - Ersetzungen von [6] durch [r], *Einf.* = Einfügungen, *Ausl.* = Auslassungen, *Phoneme* = gesamte Anzahl der Phoneme.

Es wurden die Sätze folgender Sprecherinnen genommen: Nr. 2, 7 und 9 der Studentinnen aus Sassari und 2, 3, 4, 5 und 6 der italienischen Sprecherinnen aus München. Bei diesen handelte es sich um die Sprecherinnen, die die

niedrigsten Punktzahlen nach der oben erklärten Berechnung erzielt hatten. Die einzige Ausnahme war die Sprecherin 7 aus Sassari mit der Punktzahl 0.85. Nach dem obengenannten Kriterium hätten wir eigentlich die Sprecherin 4 aus Sassari nehmen sollen (Punktzahl 0.83). Jedoch wurde uns während der Segmentation klar, dass die Fehler der Probandin 4 gravierender als die der Probandin 7 waren. Aus diesem Grund entschieden wir uns, die Sätze der nächsten Probandin aus Sassari, der Nr. 7, für den Perzeptionstest zu benutzen.

Die elf Sätze der acht Sprecherinnen wurden resynthetisiert, indem die perzipierte lokale Sprechgeschwindigkeit (PLSR) der entsprechenden Äußerungen der deutschen Muttersprachlerin darauf kopiert wurde (zur Resynthese der PLSR s. Kap. 7.1.1).

8.1.1.2 Testteilnehmer

Die Testteilnehmer waren 16 deutsche Muttersprachler, alle Mitarbeiter oder Studenten des Instituts für Phonetik der Universität München. Sie waren zwischen 23 und 49 Jahre alt und hatten keine bis sehr geringe Italienischkenntnisse. Die Probanden erhielten eine kleine Aufwandsentschädigung.

8.1.1.3 Perzeptionstest

Der Perzeptionstest wurde mit dem Tool CoDIT (s. Pfitzinger 2007 [118]) implementiert. Auf der Testoberfläche waren die Stimuli in Spalten angeordnet, in der selben Reihenfolge, in der sie in der „Buttergeschichte“ vorkamen. Jede Spalte bestand aus zwei Knöpfen mit der selben Zahlen- und Buchstabenkombination (nach dem Muster A1, a1). Hinter den gelben Knöpfen verbargen sich die Originalversionen der Sätze der italienischen Sprecherinnen und hinter den entsprechenden grauen Knöpfen die zugehörigen resynthetisierten Stimuli nach

Korrektur der perzipierten lokalen Sprechgeschwindigkeit. Die Testteilnehmer sollten die zwei Versionen miteinander vergleichen und die grauen Knöpfe entsprechend ihrer Beurteilung des Sprechrhythmus in den oberen bzw. unteren Bereich des Fensters an die nach ihrer Meinung zutreffende Stelle verschieben. Folgende sechs Bereiche standen den Teilnehmern innerhalb der Bewertungsskala zur Verfügung: 1. große Verbesserung, 2. mittlere Verbesserung, 3. leichte Verbesserung, 4. leichte Verschlechterung, 5. mittlere Verschlechterung und 6. große Verschlechterung (s. Abb. 8.1).

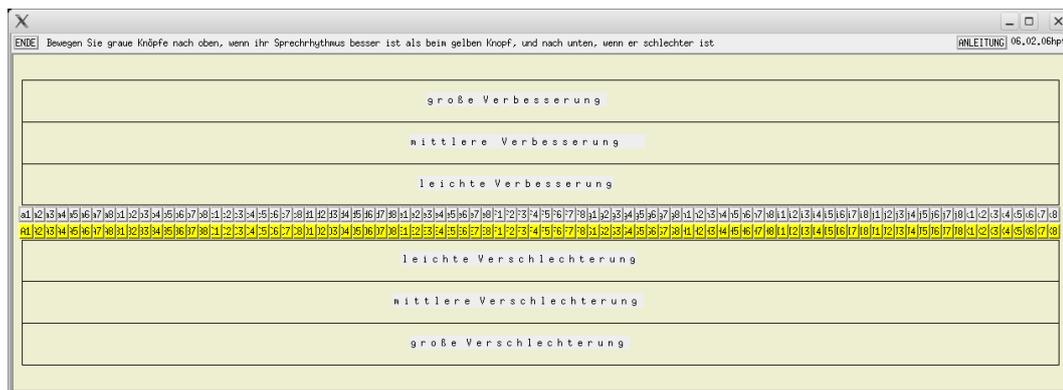


Abbildung 8.1: Benutzeroberfläche.

Je größer den Teilnehmern die Verbesserung des Sprechrhythmus im Vergleich zur Originalversion erschien, desto weiter oben — auch innerhalb eines Bereichs — sollten sie den Knopf positionieren. So konnten sie auch sehr kleine Unterschiede ausdrücken. Entsprechend sollten sie verfahren, falls sie eine Verschlechterung des Sprechrhythmus wahrnahmen: Je größer die Verschlechterung, desto weiter unten sollten sie den Knopf positionieren. Wenn sie den Sprechrhythmus in beiden Versionen gleich gut empfanden, dann sollten sie den grauen Knopf nicht bewegen und ihn im ursprünglichen Bereich lassen. Die Anweisungen für die Testteilnehmer befinden sich in Anhang C.2.

Der Test dauerte ca. eine halbe Stunde.

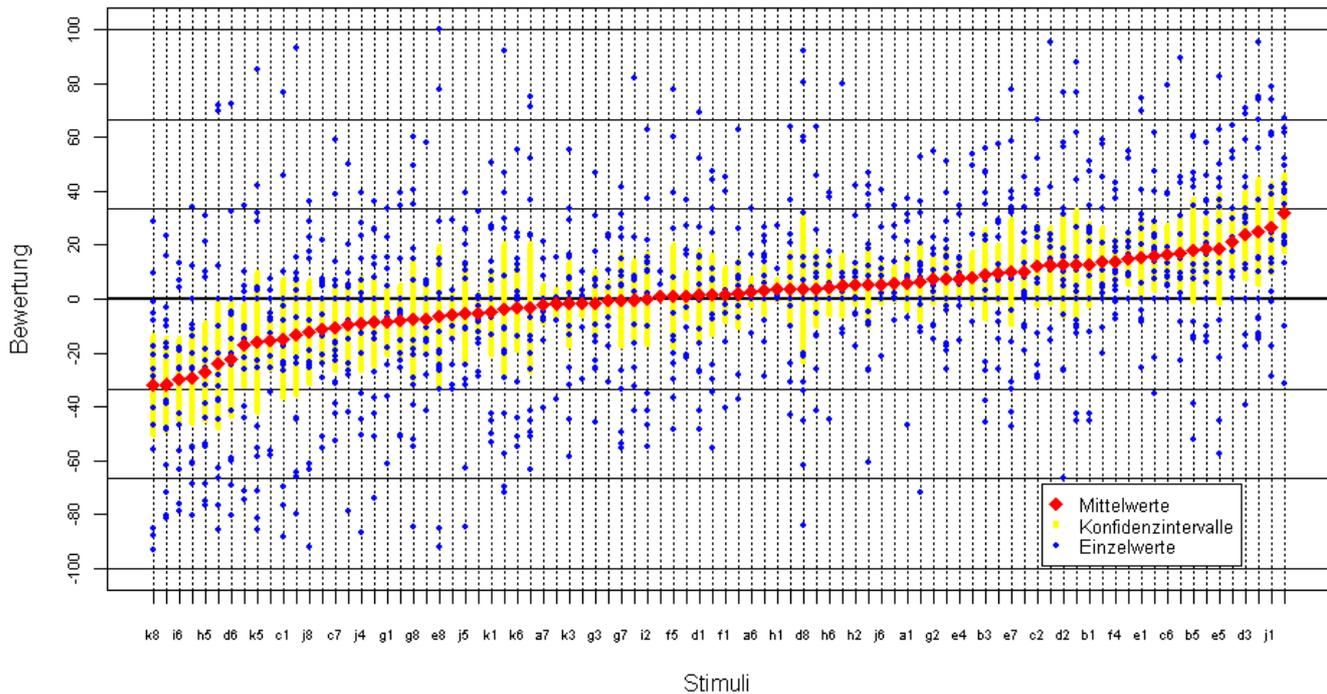


Abbildung 8.2: Die Bewertungsergebnisse der 16 Probanden zu 88 Stimuli, die nach Mittelwerten sortiert wurden. Die vertikalen Balken stellen die Konfidenzintervalle und die kleinen Punkte die einzelnen Bewertungen dar.

8.1.2 Ergebnisse und Diskussion

Die Bewertungsergebnisse wurden in eine Skala von -100 bis +100 konvertiert, in der -100 die schlechtestmögliche Verschlechterung, +100 die bestmögliche Verbesserung des Sprechrhythmus und 0 gleichbleibenden Sprechrhythmus bedeutet.

Abb. 8.2 stellt die Rohdaten des Experiments dar. Die Darstellung entspricht der Benutzeroberfläche des Tests (vgl. Abb. 8.1). Die horizontalen Li-

nien trennen die sechs Bewertungsbereiche. Der mittlere Streifen der Benutzeroberfläche, in dem die Knöpfe vor dem Test positioniert waren, wird von der Null-Linie dargestellt, auf der sich die als gleichbleibend bewerteten Stimuli befinden. Auf der x-Achse sind die 88 Stimuli anhand des Mittelwerts angeordnet. Die Konfidenzintervalle überschneiden in den meisten Fällen die Null-Linie, und dies bedeutet, dass die Resynthese der PLSR meistens keine signifikante Verbesserung des Sprechrhythmus brachte. Die Anzahl der Stimuli, die eine leichte Verbesserung aufwiesen (auf der rechten Seite der Abbildung dargestellt), überwog dabei gegenüber der Anzahl der Stimuli, die sich leicht verschlechtert hatten (linke Seite der Abbildung). Die durchschnittliche Bewertung aller Stimuli beträgt 0.96.

Damit die wiederholten Messungen nicht berücksichtigt werden müssen, wurden die Daten anhand der Stimuli gruppiert, indem der Mittelwert der 16 Bewertungen für jeden Stimulus berechnet wurde. Da die Daten normalverteilt waren, wurden sie mit dem t-Test für eine Stichprobe auf Signifikanz überprüft. Wie erwartet ist der Mittelwert nicht signifikant unterschiedlich von 0 ($p = 0.2510$). Die Bewertungsergebnisse für die einzelnen Sätze und Sprecherinnen wurden ebenfalls auf Signifikanz überprüft. Nur die Bewertungsergebnisse des Satzes s06 („*Wer später kam, musste sich hinten anschließen.*“) waren signifikant höher als 0 (Mittelwert = 7.47, $p = 0.0453$, t-Test für eine Stichprobe). Es handelt sich um einen der Sätze mit den wenigsten segmentalen Fehlern (Punktzahl = 0.23 in Tabelle 8.1). Bei den Sprecherinnen bringt die Manipulation der PLSR der Sprachsignale einer einzigen Sprecherin, der Nr. 9 aus Sassari, eine signifikante Verbesserung des Sprechrhythmus (Mittelwert = 7.96, $p = 0.0047$, t-Test für eine Stichprobe). Es handelt sich hier um die Sprecherin mit der niedrigsten Punktzahl bei der Fehlerberechnung (0.45).

Der t-Test wurde bei den Bewertungsergebnissen aller 88 Stimuli durchgeführt. 23 Stimuli wiesen eine signifikante Verbesserung und zehn eine signifikante Verschlechterung des Sprechrhythmus nach der Resynthese auf. Daher

gibt es eine, wenn auch nur minimale, Tendenz ins Positive bei den Urteilen der PLSR-manipulierten Stimuli der deutschen Muttersprachlerin.

Abb. 8.3 stellt den Stimulus *d3* von Sprecherin Nr. 9 aus Sassari dar, bei dem eine signifikante Verbesserung des Sprechrhythmus nachgewiesen wurde (Mittelwert = 24.13, $p = 0.0035$, t-Test für eine Stichprobe).

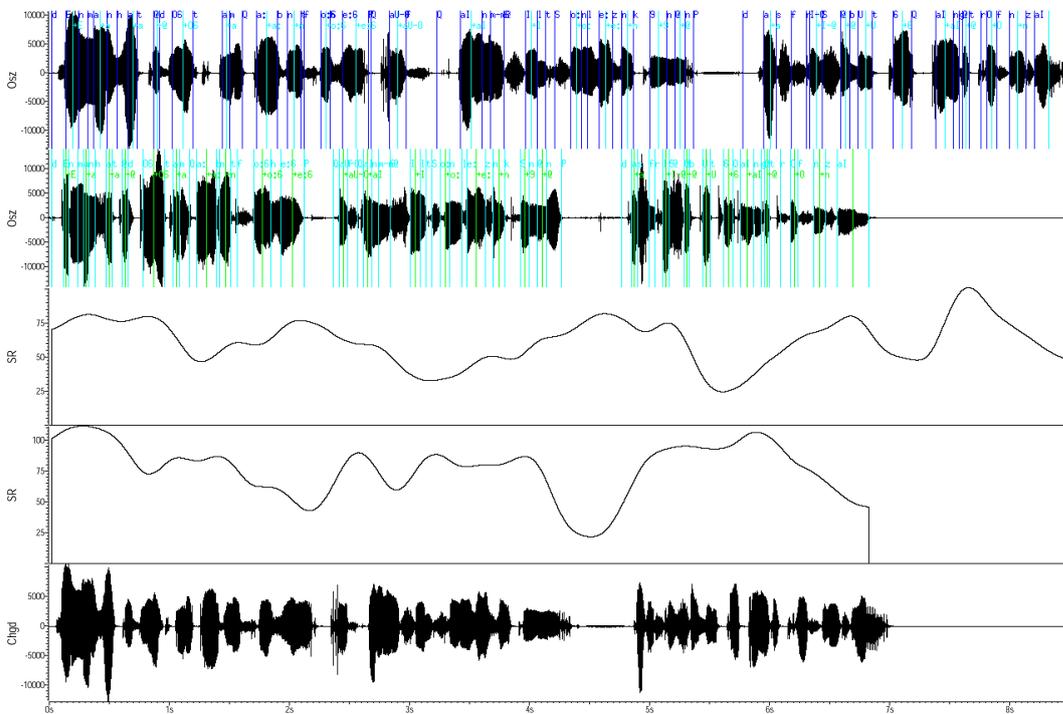


Abbildung 8.3: Stimulus *d3*: „denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter eingetroffen sei.“ von Sprecherin Nr. 9 aus Sassari. Von oben nach unten: *Osz* = Originalsignal der italienischen Sprecherin, *Osz* = Signal der deutschen Muttersprachlerin, *SR* = PLSR-Kontur der italienischen Sprecherin, *SR* = PLSR-Kontur der deutschen Muttersprachlerin, *Chgd* = PLSR-korrigiertes Signal der italienischen Sprecherin.

Die italienische Sprecherin produzierte Pausen bzw. Verzögerungen vor „am

Abend“, zwischen „*auf*“ und „*einem Schild*“, und vor „*eingetroffen*“, die im Oszillogramm und in den Vertiefungen des entsprechenden PLSR-Kontur sichtbar sind. Die Behebung dieser Pausen durch die Resynthese hat der Äußerung einen fließenden Rhythmus verliehen und den Urteil der Testteilnehmer positiv beeinflusst.

Abb. 8.4 zeigt hingegen den Stimulus *k8*, dessen Sprechrhythmus sich im Vergleich zum Originalsignal verschlechtert hatte (Mittelwert = 24.13, $p = 0.0035$). In diesem Fall produzierte die Sprecherin eine lange Atempause zwischen „*können*“ und „*meinetwegen*“. An dieser Stelle zeigt die PLSR-Kontur der italienischen Sprecherin ein tiefes Tal, das in der Kontur der deutschen Muttersprachlerin nicht vorhanden ist. Die Komprimierung des Atemgeräusches und die Beschleunigung des Wortes „*meinetwegen*“ führten möglicherweise zu den negativen Bewertungen des Sprechrhythmus. Stimulus *d8* (s. Abb. 8.5) zeigte hingegen ein sehr breites Konfidenzintervall (-23.47, 30.97), und dies bedeutet, dass er mal als Verbesserung und mal als Verschlechterung des Sprechrhythmus bewertet wurde, ohne eine Tendenz in eine der beiden Richtungen. Das Bewertungsergebnis war in der Tat nicht signifikant unterschiedlich als 0 (Mittelwert = 3.75, $p = 0.773$). Das Originalsignal ist voller Pausen, die den Sprachfluss ständig unterbrechen, was auch anhand der Spitzen und Tälern in der PLSR-Kontur zu sehen ist. Einerseits haben die Testteilnehmer möglicherweise den Sprechrhythmus der Originaläußerung als besonders mangelhaft empfunden, andererseits klang die Resynthese auch wegen der Komprimierung der häufigen Pausen besonders unnatürlich. Einige Testteilnehmer haben wahrscheinlich die Unnatürlichkeit der Resynthese der gestammelten Originaläußerung bevorzugt, andere umgekehrt, was zu diesem gespaltenen Ergebnis führte.

8.1.3 Betrachtungen nach dem Experiment 1

Die Manipulation der PLSR als alleinigem Parameter führte nicht zu einer signifikanten Verbesserung des Sprechrhythmus. Die Einfügungen und Aus-

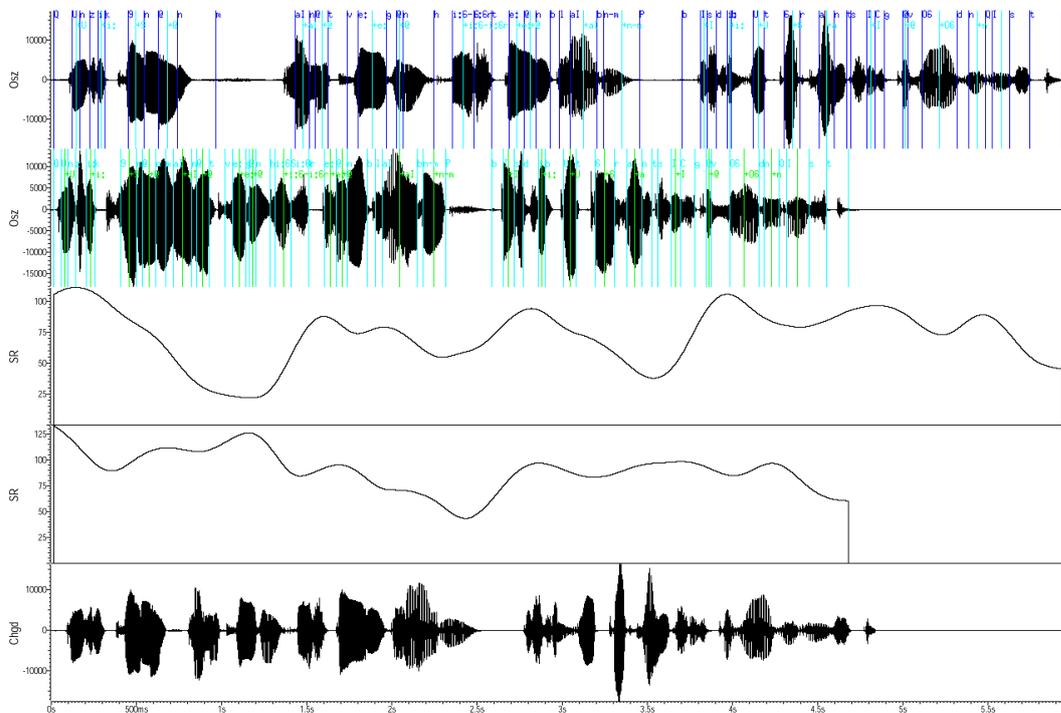


Abbildung 8.4: Stimulus *k8*: „und Sie können meinetwegen hier stehen bleiben, bis die Butter ranzig geworden ist.“ von Sprecherin Nr. 6 aus München. Zur Legende s. Beschreibung von Abb. 8.3.

lassungen in den Sprachdaten der italienischen Sprecherinnen und vor allem die Beschleunigung langer Pausen können eine Verschlechterung der Qualität der PLSR-korrigierten Stimuli bewirkt haben. Die Sprachsignale des einzigen Satzes bzw. der einzigen Sprecherin, deren Sprechrhythmus sich durch Resynthese signifikant verbesserte, waren diejenigen Signale, die eine niedrige Anzahl segmentaler Fehler aufwiesen. Zudem darf man keine großen Verbesserungen erwarten, wenn nur ein Parameter korrigiert wird. Im Gegenteil, da die neu manipulierte PLSR-Kontur nicht in Einklang mit dem originalen Intonationsverlauf und der originalen Intensität steht, wurden die daraus resultierenden Sprachsignale möglicherweise oft als unnatürlich empfunden, und das könnte

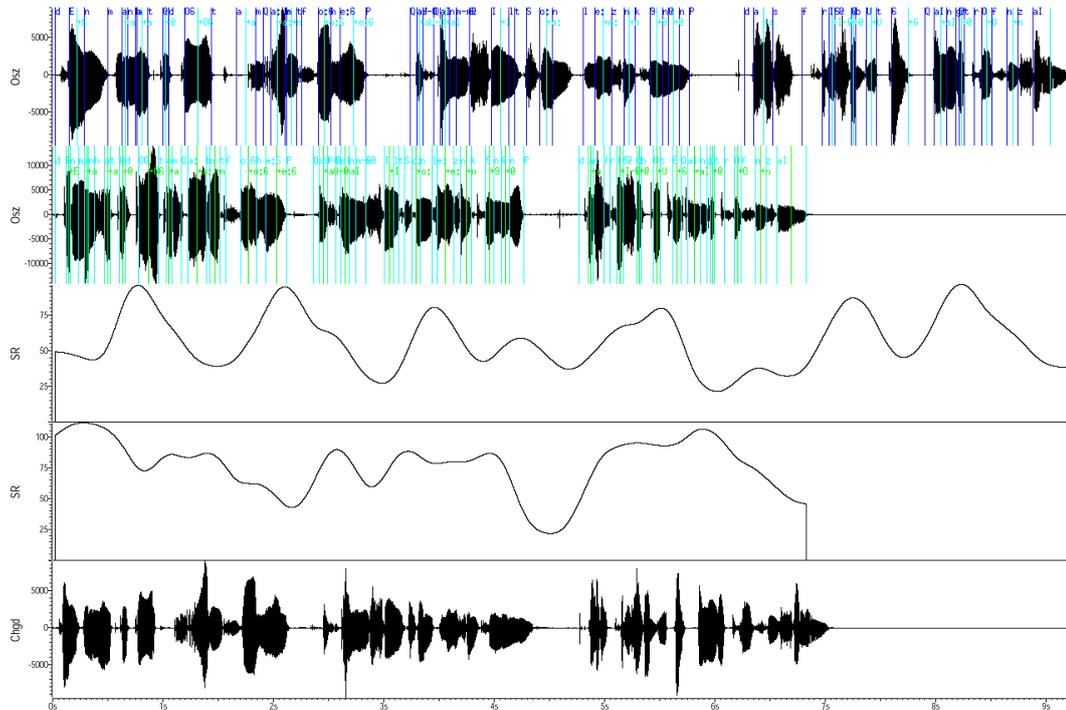


Abbildung 8.5: Stimulus d8: „denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter eingetroffen sei.“ von Sprecherin Nr. 6 aus München. Zur Legende s. Beschreibung von Abb. 8.3.

das Urteil der deutschen Testteilnehmer beeinflusst haben. Außerdem bewirkt die Resynthese der PLSR keine exakte Angleichung der Dauer der Phone in den Äußerungen der italienischen Sprecherinnen an die Phone der Äußerungen der deutschen Muttersprachlerin, die vielleicht zu einer größeren Verbesserung des Sprechrhythmus geführt hätten.

Aus diesen Gründen — und auch wegen der Notwendigkeit einer exakten Korrespondenz zwischen der Dauer der Phone in den Quell- und Zieläußerungen bei der zusätzlichen Resynthese der Intonation und der Intensität — wurde für die folgenden Experimente keine Resynthese der PLSR durchgeführt. Stattdessen wurde die Dauer der Phone der Äußerungen der italienischen Sprecher

an die der entsprechenden Äußerungen der deutschen Muttersprachlerin angeleglichen (zur Resynthesemethode s. Kap. 7.1.2).

8.2 Experiment 2: Einfluss von Intonation und lokaler Sprechgeschwindigkeit auf die Betonung morphologisch komplexer Wörter im Deutschen durch L2-Sprecher

In diesem Experiment wird der Einfluss einzelner prosodischer Parameter — der Intonation und der lokalen Sprechgeschwindigkeit — auf die Betonung morphologisch komplexer Wörter im Deutschen wie Komposita und Präfixverben, die von italienischsprachigen Deutschlernenden ausgesprochen werden, untersucht. Ziel ist es, zu ermitteln, welches dieser zwei Merkmale mit der Beurteilung der Betonung der Wörter durch Muttersprachler — und daher auch mit der Perzeption der falschen Betonung — am stärksten korreliert.

Wie bereits oben erwähnt (s. 2.4 auf S. 41) beeinflusst laut Anderson-Hsieh, Johnson & Koehler 1992 [4] die Prosodie die Beurteilung des Akzents von L2-Sprechern durch Muttersprachler in stärkerem Maße als die segmentalen Aspekte und die Silbenstruktur. Unser Experiment versucht, die Forderung der Autoren umzusetzen, den Einfluss einzelner prosodischer Merkmale auf die Beurteilung der Aussprache zu untersuchen (s. Anderson-Hsieh, Johnson & Koehler 1992 [4, S.549]).

8.2.1 Methode

Für die hier präsentierte Studie wurden 16 morphologisch komplexe Wörter aus dem Text „Die Buttergeschichte“ — isoliert und in Sätzen eingebettet — von einer Muttersprachlerin und anschließend von acht italienischsprachigen Deutschlernerinnen vorgelesen. Es handelte sich ausschließlich um Wörter, die auf der ersten Silbe betont werden müssen.

Die prosodischen Merkmale der Sätze und Wörter wurden dann mittels Resynthese den entsprechenden Äußerungen der Muttersprachlerin angeglichen. Daraus ergaben sich drei weitere korrigierte Versionen jedes Satzes bzw. Wortes durch Korrektur folgender Merkmale:

1. Intonation,
2. Lokale Sprechgeschwindigkeit,
3. Intonation und lokale Sprechgeschwindigkeit.

Um die Wirkung der Korrektur von Grundfrequenz und lokaler Sprechgeschwindigkeit zu überprüfen, bewerteten anschließend 31 deutsche Muttersprachler in einem Perzeptionstest die Korrektheit der Betonung der Wörter in der Originalversion und in den drei korrigierten Versionen.

Mittels statistischer Methoden wurde untersucht,

1. ob die Korrektur der oben genannten prosodischen Merkmale in der Aussprache der Deutschlernerinnen die Bewertung der Betonung signifikant beeinflusst,
2. ob die Korrektur der Intonation oder die der lokalen Sprechgeschwindigkeit die größere Wirkung zeigt, und

3. ob es einen multiplikativen Effekt gibt, d.h. ob die Wirkung der Korrektur beider Merkmale größer als die Summe der Wirkungen der einzelnen Merkmale ist.

8.2.1.1 Stimuli

Für dieses Experiment wurde eine zweite Aufnahme notwendig, da sowohl Sätze als auch Einzelwörter aus der „Buttergeschichte“ als Stimuli verwendet werden sollten.

Die Aufnahme der Sprachdaten wurde ebenfalls in dem reflexionsarmen Raum des Instituts für Phonetik der Ludwig-Maximilians-Universität München mit dem Mikrofon Neumann TLM 103, bei 48 KHz Abtastrate und 16 Bit Auflösung, durchgeführt.

Die deutsche Muttersprachlerin, die bereits den gesamten Text der „Buttergeschichte“ aufgenommen hatte (s. Kap. 5.2), nahm auch das neue Korpus auf. Es wurden außerdem acht italienische Sprecherinnen aufgenommen. Drei von ihnen waren die sardischen Sprecherinnen 1, 4 und 5, die bereits an der vorherigen Aufnahme teilgenommen hatten. Die anderen fünf waren unterschiedlicher regionaler Herkunft. Die italienischen Sprecherinnen waren zwischen 20 und 52 Jahre alt und wohnten in München. Die Dauer ihres Aufenthalts in Deutschland war unterschiedlich und lag zwischen einem Monat und 15 Jahren. Die zwei Probandinnen, die seit weniger als einem Jahr in Deutschland lebten, hatten jedoch bereits vier bzw. sechs Jahre Deutschunterricht. Alle Probandinnen erhielten eine Aufwandsentschädigung.

Die Probandinnen lasen zunächst 16 Sätze aus der Buttergeschichte. 14 davon enthielten die insgesamt 16 auf der ersten Silbe zu betonenden morphologisch komplexen Wörter. Zusätzlich lasen sie die 16 isolierten Wörter (s. Anhang D.1.1). Um zu verhindern, dass die Probandinnen dasselbe Betonungs-

muster von einem Wort auf das andere übertragen, lasen sie die morphologisch komplexen Wörter in einer Liste zusammen mit weiteren 16 auf unterschiedlichen Silben betonten Wörtern, die ebenfalls aus der Buttergeschichte stammten. Die Reihenfolge der Wörter in der Liste entsprach der Reihenfolge im Originaltext (s. Wortliste im Anhang D.1.2).

Jeder von den Deutschlernerinnen vorgelesene Satz und jedes isoliert ausgesprochene Wort wurde dann von einem deutschen Muttersprachler und von der Autorin bezüglich der Korrektheit der Betonung überprüft. Alle Sätze und Wörter, die in dieser Hinsicht als fehlerhaft beurteilt wurden, wurden resynthetisiert. Ihre lokale Sprechgeschwindigkeit und Intonation wurde an die des entsprechenden Satzes bzw. Wortes der Muttersprachlerin angeglichen. Für die Resynthese der lokalen Sprechgeschwindigkeit wurde die Dauer der Phone der Signale jeder italienischen Sprecherin an die der entsprechenden Signale der deutschen Muttersprachlerin angeglichen (zur verwendeten Resynthesemethode der lokalen Sprechgeschwindigkeit und der Intonation s. Kap. 7.1.2 bzw. 7.2).

Auf die Manipulation eines dritten Parameters — der Intensität — wurde verzichtet, da mehreren Studien zufolge (s. Lehiste 1970 [82]) die Perzeption der Betonung stärker vom Grundfrequenzverlauf und von der Dauer der Silben als von der Intensität abhängt. Die Berücksichtigung der Intensität hätte die Anzahl der Stimuli im Perzeptionstest unnötig verdoppelt (von 2^2 auf 2^3 Versionen pro Äußerung). Zur Bewertung standen also vier Versionen derselben Äußerung einer jeden italienischen Sprecherin:

1. original
2. mit korrigierter lokaler Sprechgeschwindigkeit
3. mit korrigierter Intonation
4. mit korrigierter Sprechgeschwindigkeit und Intonation.

8.2.1.2 Testteilnehmer

Die Teilnehmer an dem Perzeptionstest waren 31 deutsche Muttersprachler zwischen 20 und 49 Jahren, in den meisten Fällen mit keinen oder minimalen Italienischkenntnissen. Sie waren alle Mitarbeiter oder Studenten am Institut für Phonetik der Ludwig-Maximilians-Universität München. Ein Teil der Probanden hatte bereits an dem Perzeptionstest für Experiment 1 teilgenommen (s. Kap. 8.1.1.2). Für die Teilnahme am Test, der zirka 40 bis 50 Minuten dauerte, wurden die Probanden entschädigt.

8.2.1.3 Perzeptionstest zur Betonung morphologisch komplexer Wörter

Da zu viele Stimuli zur Verfügung standen, und der Test für die Teilnehmer deshalb zu umfangreich geworden wäre, wurde beschlossen, die Anzahl der Stimuli auf 256 zu begrenzen. Um diejenigen 64 originalen Äußerungen zu wählen, aus denen die 256 Stimuli generiert werden sollten, wurden nur die am häufigsten falsch betonten Äußerungen gewählt. Es handelte sich um zehn Äußerungen, darunter sechs Sätze und vier Wörter (s. Anhang D.2). Das hatte zum einen den Vorteil, dass für jede Äußerung Exemplare mehrerer Sprecher beurteilt werden konnten, zum anderen, dass wir uns so auf die häufigsten — und somit wahrscheinlich auch auf die relevantesten — Fehler konzentrieren konnten. Wie beim Experiment 1 wurde der Test ebenfalls mit dem Tool Co-DIT entwickelt (s. Pfitzinger 2007 [118]). Auf der Testoberfläche waren die Stimuli in Spalten angeordnet (s. Abb. 8.6). Hinter den Knöpfen einer jeden Spalte mit derselben Zahlen- und Buchstabenkombination (nach dem Muster a1, A1, 1a, 1A) verbargen sich die vier Versionen jeweils derselben Äußerung derselben italienischen Sprecherin. Die vier Versionen einer Äußerung wurden auf die vier Knöpfe jeder Spalte zufällig verteilt. Die Teilnehmer sollten die vier Versionen miteinander vergleichen und die Knöpfe entsprechend ihrer Beurtei-

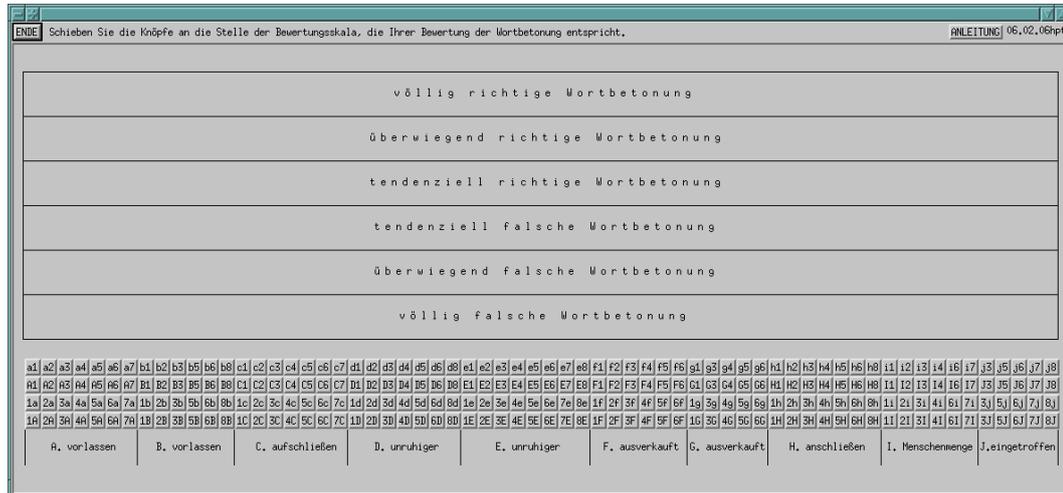


Abbildung 8.6: Benutzeroberfläche des Perzeptionstests

lung der Betonung an die ihrer Meinung zutreffende Stelle im oberen Bereich des Fensters verschieben. Folgende sechs Bereiche standen den Teilnehmern innerhalb der Bewertungsskala zur Verfügung: 1. völlig richtige Wortbetonung, 2. überwiegend richtige Wortbetonung, 3. tendenziell richtige Wortbetonung, 4. tendenziell falsche Wortbetonung, 5. überwiegend falsche Wortbetonung und 6. völlig falsche Wortbetonung.

Je korrekter den Teilnehmern die Betonung eines Kompositums erschien, desto weiter oben — auch innerhalb eines Bereichs — sollten sie den Knopf positionieren. So konnten sie auch sehr kleine Unterschiede in der Betonung ausdrücken.

8.2.2 Ergebnisse der statistischen Datenanalyse

Die aus dem Perzeptionstest resultierenden Daten wurden statistisch analysiert, um zu überprüfen, ob die Bewertungsergebnisse der vier Versionen der Äußerungen der italienischen Sprecherinnen signifikant unterschiedlich sind.

Die Bewertungsergebnisse wurden in eine Skala von 0 bis 100 konvertiert, in der 0 die schlechtestmögliche und 100 die bestmögliche Note bedeutet.

Die Betrachtung der Mittelwerte der Beurteilungen der gesamten Stimuli in den vier Versionen zeigt das höchste Bewertungsergebnis für die Stimuli mit ausschließlich resynthetisierter F0 und für diejenigen mit beiden resynthetisierten Merkmalen. Sie zeigt auch höhere Bewertungsergebnisse für die Versionen mit resynthetisierter lokaler Sprechgeschwindigkeit im Vergleich zur Originalversion (s. Tab. 8.2). Zu denselben Erkenntnissen kommt man auch beim Betrachten der Mediane und der Quartile im Boxplot (s. Abb. 8.7).

	Original	Sprechgesch.	Intonation	Sprechgeschw. + Inton.
Mittelwert	40.08	44.24	58.26	62.32

Tabelle 8.2: Mittelwerte der Bewertungsergebnisse der vier Versionen der Stimuli.

Um zu überprüfen, ob die Bewertungsergebnisse der Originalversionen und der drei resynthetisierten Versionen voneinander signifikant abweichen, wurden der Friedman-Test und der Wilcoxon-Test durchgeführt.

Der Friedman-Test ist ein nicht-parametrischer Test für den verteilungsunabhängigen Vergleich mehrerer abhängiger Stichproben. Der Test zeigte einen höchst signifikanten p-Wert ($p < 0.001$).

Der Wilcoxon-Test für Paardifferenzen mit multiplen Vergleichen wurde durchgeführt, um jede einzelne Version jeweils mit allen anderen zu vergleichen. Für alle sechs Vergleiche ermittelte der Test höchste Signifikanz ($p < 0.001$).

Um die Interaktion der beiden Parameter Intonation und Sprechgeschwindigkeit zu überprüfen, wurde eine Varianzanalyse (ANOVA) durchgeführt, da keiner der gängigen nicht-parametrischen Tests die Wechselwirkungen berücksichtigt. Zudem ist ANOVA ein robuster Test, der auch bei nicht-

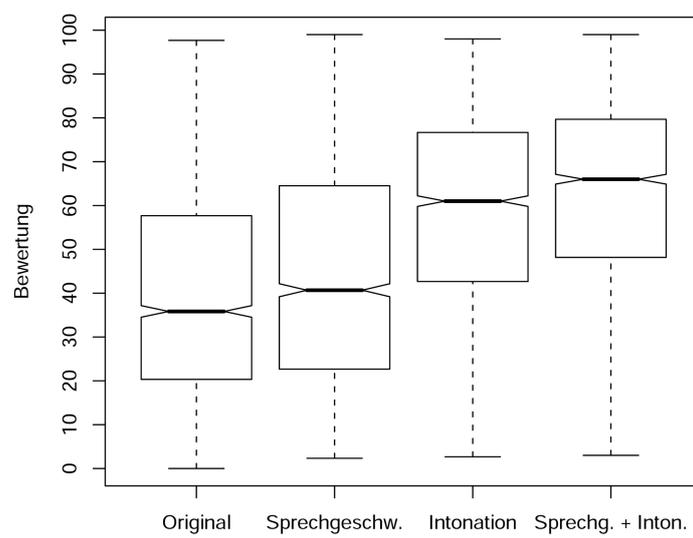


Abbildung 8.7: Bewertungsergebnisse der vier Versionen der Stimuli.

normalverteilten Daten noch zu brauchbaren Ergebnissen führen kann. Hinzu kommt, dass wenn mehr Daten zur Verfügung gestanden hätten, diese wahrscheinlich normalverteilt gewesen wären.

Die mit ANOVA ermittelte Interaktion ergab einen nicht-signifikanten p-Wert von 0.92. Das heißt, dass die Effekte der Resynthese der Intonation und der lokalen Sprechgeschwindigkeit nur additiv sind. Bei der Modifikation der Parameter lieferte die ANOVA höchste Signifikanz sowohl für die lokale Sprechgeschwindigkeit als auch für die Intonation ($p < 0.001$).

Die ANOVA-Koeffizienten von 4.16 für die Resynthese der lokalen Sprechgeschwindigkeit und von 18.18 für die Resynthese der Intonation, zeigen einen höheren Effekt der Intonation auf die Beurteilung der Testteilnehmer.

8.2.3 Betrachtungen nach dem Experiment 2

Die Korrektur der Intonation und der lokalen Sprechgeschwindigkeit durch Resynthese hat die Bewertung der Betonung durch deutsche Muttersprachler signifikant beeinflusst, wobei sich die Korrektur der Intonation gegenüber der der lokalen Sprechgeschwindigkeit als effektiver erwies. Untersuchungen belegen, dass die Betonung im Deutschen hauptsächlich mit der Dauer statt mit der Grundfrequenz korreliert (s. Dogil und Williams 1999 [33] und Jessen u.a. 1995 [67]). In den meisten Fällen waren die im Test zu beurteilenden Sätze in Fokusposition, und dies könnte den größeren Einfluss der Grundfrequenz erklären. Jedoch berichteten mehrere Testteilnehmer, dass sie die Versionen eines Stimulus paarweise gleich empfanden, dass also zwei Versionen die falsche und zwei die richtige Betonung hatten. Tatsächlich stellte sich während der Resynthese heraus, dass bei vielen Stimuli die Korrektur der lokalen Sprechgeschwindigkeit anscheinend keine Verbesserung der Betonung brachte, während erst die Korrektur der Intonation effektiv zu sein schien. In diesen Fällen unterschiede-

den sich also die Dauerverhältnisse der Laute in den von den Italienerinnen ausgesprochenen morphologisch komplexen Wörtern nicht erheblich von den Dauerverhältnissen in denselben von der Muttersprachlerin ausgesprochenen Wörtern. Erst die Modifikation der F0 konnte die Betonung korrigieren. Das entspricht der bereits in Kap. 2.3.2 erläuterten Vermutung, dass bei morphologisch komplexen Wörtern im Deutschen die Grundfrequenz als Korrelat der Betonung eine größere Rolle als bei anderen Wörtern spielt, und könnte also erklären, warum italienischsprachige Deutschlernende Schwierigkeiten haben, die Betonung solcher Wörter zu lernen.

Die in der Sprechgeschwindigkeit modifizierten Sätze können Artefakte enthalten, vor allem, wenn die Sätze lang sind, und der Unterschied zwischen der Dauer der Äußerung der italienischen Sprecherin und der Dauer der Äußerung der deutschen Muttersprachlerin erheblich ist. Die Artefakte im Sprachsignal könnten die Wirkung der Korrektur der lokalen Sprechgeschwindigkeit beeinträchtigt haben, falls sie die Testteilnehmer dazu gebracht haben, die Betonung schlechter zu bewerten. Dies relativiert sich jedoch, wenn man bedenkt, dass die auf der Subgruppe der Wörter berechneten ANOVA-Koeffizienten ebenfalls eine höhere Wirkung der Intonation zeigten (lokale Sprechgeschwindigkeit 2,31, Intonation 18,57), obwohl die Wörter selten Artefakte enthielten.

Ein Ziel des hier präsentierten Experiments war, zu überprüfen, ob sich die synthetische Korrektur der prosodischen Merkmale effektiv auf die Betonung auswirkt, so dass die resynthetisierten Stimuli — mit der zusätzlichen Korrektur der Intensität und nach Überprüfung ihrer Qualität — in einem Computer-Sprachlernprogramm als Feedback eingesetzt werden könnten. Für solche künftige Anwendungen ist es wichtig, dass die Resynthese mit Verzicht auf die manuelle Segmentation durchgeführt werden kann. Dadurch könnten die Lernenden das Feedback in der eigenen Stimme unmittelbar, nachdem sie eine Äußerung falsch ausgesprochen haben, erhalten und hätten so die Möglichkeit, die korrekte und die falsche Aussprache miteinander zu vergleichen.

Mit der zusätzlichen Korrektur der Intensität und nach Überprüfung der Qualität der Stimuli konnten Äußerungen der Lernenden in einem Aussprachetraining getestet werden. Im folgenden Kap. 8.3 geht es deshalb um ein Training der Betonung morphologisch komplexer Wörter im Deutschen für italienische Sprecher.

8.3 Experiment 3: Training der Wortbetonung durch Resynthese und emphatische Betonung für italienischsprachige Lernende

In diesem Experiment testen wir eine neue Methode, mit der italienischsprachigen Lernenden die Betonung morphologisch komplexer Wörter vermittelt werden kann. Nach dieser Methode erhalten die Lernenden ein Feedback in Form ihrer eigenen Äußerungen, deren prosodische Parameter allerdings mittels Resynthese korrigiert und/oder überbetont wurden, so dass die Lernenden die betonte Silbe einfacher identifizieren können.

In dieser Studie lasen zwölf italienischsprachige Deutschlernende isolierte oder in Sätzen eingebettete morphologisch komplexe Wörter im Deutschen. Die Personen wurde in zwei Gruppen aufgeteilt, die jeweils zwei Arten von Feedback erhielten. Die eine Gruppe erhielt:

1. ihre eigene Äußerungen nach Korrektur der F0, Dauer der Segmente und Intensität
2. eine zweite korrigierte Version, deren prosodische Parameter an überbetonte Stimuli der deutschen Muttersprachlerin angepasst worden waren

Die zweite Gruppe erhielt:

1. aufgezeichnete Äußerungen der deutschen Muttersprachlerin (das traditionelle Feedback)
2. überbetonte Äußerungen der deutschen Muttersprachlerin

Die erste zu verifizierende Hypothese lautet, dass Feedback in der eigenen Stimme den Lernenden hilft, den Unterschied zwischen korrekter und inkorrekt Aussprache wahrzunehmen (s. Tillmann & Pfitzinger 2004 [140, 141]). Die zweite Hypothese lautet, dass die Überbetonung der zu betonenden Silbe den Lernenden hilft, die eigenen Fehler besser zu identifizieren.

8.3.1 Methode

Das Textmaterial bestand aus Sätzen und Einzelwörtern aus der „Buttergeschichte“, gelesen von zwölf italienischsprachigen Deutschlernenden, zehn weiblichen und zwei männlichen. Dasselbe Material wurde von der deutschen Muttersprachlerin zweimal gelesen. Zunächst las die deutsche Muttersprachlerin die Äußerungen normal, dann betonte sie die morphologisch komplexen Wörter mit didaktischer Emphase.

Die italienischen Sprecher wurden nach dem Zufallsprinzip einer Test- und einer Kontrollgruppe zugeteilt. Die Korrektheit der Wortbetonung der von den italienischen Lernenden ausgesprochenen Wörter wurde von der Autorin und von einem deutschen Muttersprachler überprüft. Die von den Sprechern der Testgruppe ausgesprochenen Äußerungen, die als nicht korrekt betont beurteilte Wörter enthielten, wurden mittels Resynthese prosodisch korrigiert. Die Resynthese wurde ausgeführt, indem folgende prosodische Parameter von den Äußerungen der deutschen Muttersprachlerin auf sie übertragen wurden: lokale Sprechgeschwindigkeit, Grundfrequenz und Intensität. Es wurden zwei Versionen der resynthetisierten Stimuli geschaffen. Für die erste Version wurden die prosodischen Parameter der normal betonten Äußerung der Muttersprachle-

rin verwendet. Für die zweite wurden die Parameter der Äußerungen mit den überbetonten Wörtern verwendet.

Während der Trainingsphase hörten die italienischen Sprecher zwei Versionen jeder Äußerung: zunächst die mit der falschen Betonung, die sie während der ersten Aufnahmesession selbst gesprochen hatten, danach dieselbe Äußerung mit korrigierter Prosodie. Die Sprecher der Testgruppe hörten die resynthetisierten Stimuli in der eigenen Stimme, während die Sprecher der Kontrollgruppe die deutsche Muttersprachlerin hörten. Nachdem sie dieses Feedback gehört hatten, sprachen die Probanden die Äußerung ein zweites Mal, und diese wurde aufgenommen. Sie wiederholten dann die Übung. Dieses Mal bekamen sie als Feedback die Äußerungen mit den überbetonten Wörtern. Somit standen nach dem Training also nun zwei nachgesprochene Versionen jeder Äußerung für jeden italienischen Sprecher zu Verfügung: die eine nach Feedback aus normal betonten Stimuli und die andere nach Feedback aus überbetonten Stimuli. Die Korrektheit der Betonung in den Äußerungen vor und nach dem Training sollte von deutschen Muttersprachlern in einem Perzeptionstest bewertet werden (s. Experiment 4 in Kap. 8.4).

8.3.1.1 Aufnahmen vor dem Training

Während der in Kap. 8.2.1.1 beschriebenen Aufnahmesession wurden die 16 Sätze und 32 Wörter aus der „Buttergeschichte“ aufgenommen, die sowohl für das Experiment 2 als auch für dieses Experiment verwendet wurden. Zusätzlich zu den italienischen Sprecherinnen, deren Sprachdaten für das Experiment 2 verwendet wurde, wurden nun noch vier weitere italienische Sprecher, zwei weibliche und zwei männliche, unterschiedlicher regionaler Herkunft aufgenommen. Insgesamt handelt es sich also nun um zehn Frauen und zwei Männer im Alter zwischen 20 und 52 Jahren, alle wohnhaft in München. Ihre Aufenthaltsdauer in Deutschland lag zwischen 2 Monaten und 23 Jahren. Acht von ihnen

waren fortgeschrittene Lerner und vier, darunter auch die zwei männlichen Sprecher, waren Anfänger.

Die deutsche Muttersprachlerin las die Äußerungen ein zweites Mal, wobei sie die morphologisch komplexen Wörtern in den Sätzen und alle Wörter auf der Liste überbetonte. Es wurde ihr empfohlen, sich vorzustellen, dass die Lernenden die Wörter falsch betont hätten, und sie sollte ihnen klar machen, welche Silbe zu betonen war.

Die italienischen Sprecher wurden per Zufall einer Test- und einer Kontrollgruppe zugewiesen. Um zu vermeiden, dass alle vier Anfänger, inklusive die beiden männlichen Sprecher, derselben Gruppe zugeteilt wurden, wiesen wir per Zufall je einen männlichen und einen weiblichen Anfänger der Testgruppe und die übrigen beiden der Kontrollgruppe zu.

8.3.1.2 Vorbereitung der Stimuli für das Aussprachetraining

Die Äußerungen der Sprecher der Testgruppe, deren Wörter als nicht korrekt betont beurteilt worden waren, wurden resynthetisiert, um den Fehler zu korrigieren. Intonation, lokale Sprechgeschwindigkeit und Intensität der Äußerungen der deutschen Muttersprachlerin wurden auf die entsprechenden Äußerungen jedes italienischen Sprechers kopiert.

Die Resynthese wurde wie in Kap. 7 erläutert durchgeführt, wobei die Dauer der Phone der Äußerungen der italienischen Probanden an die Dauer der Phone der entsprechenden Äußerungen der deutschen Muttersprachlerin angeglichen wurde (s. dazu Kap. 7.1.2). Die Resynthese der lokalen Sprechgeschwindigkeit war daher ohne Handsegmentierung der Daten möglich.

8.3.2 Aussprachetraining

Zirka eine Woche nach der Aufnahme fand das Aussprachetraining für die italienischen Sprecher statt. Es war sehr unwahrscheinlich, dass sich ihre Aussprache morphologisch komplexer Wörter in der Zwischenzeit verbessert hätte. Denn erstens hatten die Probanden keinerlei Information über Themen und Inhalte des Trainings bekommen, und zweitens wohnten die meisten von ihnen bereits seit mehreren Jahren in Deutschland und hatten den Fehler nie bemerkt. Bemerkenswert ist auch, dass die italienischen Sprecherinnen 1, 4 und 5 aus München, die auch an der allerersten Aufnahme teilgenommen hatten (s. Kap. 5) genau dieselben Wortbetonungsfehler in der zweiten Aufnahme machten. Dies zeigt, wie sehr sich diese Fehler in ihrer Aussprache verfestigt hatten.

Das Aussprachetraining umfasste für jeden Teilnehmer diejenigen seiner Äußerungen, die falsch betonte morphologisch komplexe Wörter enthielten, weshalb das Trainingsmaterial für jeden Teilnehmer unterschiedlich war.

Den meisten Sprechern bereiteten die Sätze erwartungsgemäß größere Probleme als die Wörter. In manchen Fällen war die falsche Betonung nur in den Sätzen zu finden und nicht in den entsprechenden isolierten Wörtern. In solchen Fällen wurden die Wörter sowohl als Einzelwort als auch im Satz ins Training aufgenommen, auch wenn das isolierte Wort ursprünglich korrekt ausgesprochen worden war. Damit wollten wir die Aufmerksamkeit des Sprechers auf das Einzelwort lenken und den Lerneffekt maximieren. Wenn die italienischen Sprecher weitere Wörter außer den ausgewählten 16 morphologisch komplexen Wörtern falsch betont hatten, wurden diese auch ins Trainingsmaterial eingefügt.

Auch das Aussprachetraining wurde im reflexionsarmen Raum des Instituts für Phonetik in München durchgeführt. Vor dem Teilnehmer wurde ein Lautsprecher platziert, aus dem er die vorbereiteten Stimuli hörte (s. Abb. 8.8). Die

Teilnehmer bekamen eine Liste mit den Wörtern und Sätzen, in denen sie Betonungsfehler gemacht hatten. Da das Trainingsmaterial für jeden Teilnehmer unterschiedlich war, war die Dauer des Trainings auch für jeden unterschiedlich, aber sie dauerte nie länger als eine Stunde.

Das Training bestand aus zwei Teilen. Im ersten Teil bekamen die Teilnehmer der Testgruppe folgende Instruktionen. Ihnen wurde gesagt, dass sie zwei Versionen jedes Wortes und jedes Satzes hören würden, zum einen ihre Originaläußerung und zum anderen ihre gemäß folgender prosodischen Merkmale korrigierte Äußerung: Intonation, Dauer der Vokale und Konsonanten (Sprechgeschwindigkeit) und Intensität. Sie würden jede Version zwei Mal hören. Danach sollten die Teilnehmer die Äußerungen nochmals je zweimal aussprechen, wobei sie die korrekte Aussprache nachahmen sollten. Die neue Aussprache wurde dann aufgenommen. Nur die erste Äußerung wurde danach in einem Perzeptionstest evaluiert. Die zweite wurde nur für den Fall, dass die erste unbrauchbar wäre, aufgenommen. Dies passierte nur ein Mal, als eine Sprecherin der Testgruppe zu früh zu sprechen begann, und das erste Wort nicht komplett aufgenommen wurde.

Während sie die Wörter aussprachen, sollten die Teilnehmer darauf achten, die oben erwähnten prosodischen Eigenschaften korrekt zu reproduzieren. In der Aussprache eines Satzes, der einem Wort folgte, sollten sie insbesondere auf das betreffende, gerade einzeln trainierte Wort achten.

Nachdem die Teilnehmer der Testgruppe den ersten Teil des Trainings beendet hatten, wurden ihnen die Instruktionen für den zweiten Teil gegeben: Die Teilnehmer würden die zwei Versionen, die sie vorher schon gehört hatten, noch einmal hören. Zusätzlich würden sie eine dritte Version hören. Diese bestand aus einer resynthetisierten Äußerung, in der das morphologisch komplexe Wort überbetont war, so dass die Teilnehmer die Betonungsstelle einfacher erkennen konnten. Nachdem die Teilnehmer die drei Versionen gehört hatten, sollten sie



Abbildung 8.8: Probandin während des Aussprachetrainings.

die Äußerung nochmals aussprechen.

Das Training der Teilnehmer der Kontrollgruppe war genauso strukturiert. Der einzige Unterschied war, dass sie die korrekten Äußerungen in der Version der deutschen Muttersprachlerin anstelle des resynthetisierten Stimulus in ihrer eigenen Stimme erhielten. Im ersten Teil des Trainings wurde ihnen zudem gesagt, dass sie auf die oben beschriebenen prosodischen Eigenschaften achten sollten.

Die resynthetisierten Stimuli in der eigenen Stimme wirkten motivierend auf die Lernenden. Die Teilnehmer der Testgruppe zeigten mehr Interesse für das Training und waren neugierig und überrascht, die korrekte Aussprache in ihrer Stimme zu hören. Am Ende des Trainings sagten einige, sie würden sich freuen, in einer Woche wiederzukommen, um weitere Übungen dieser Art zu machen. Viele Teilnehmer der Kontrollgruppe waren hingegen vom Training enttäuscht und zeigten nicht dasselbe Interesse wie die Testgruppe.

8.3.3 Betrachtungen nach dem Experiment 3

In unserem Aussprachetraining wurden resynthetisierte Stimuli als Feedback verwendet, die ohne vorherige Handsegmentation realisiert werden konnten. Für eine künftige Anwendung von resynthetisierten Stimuli in computergestützten Programmen (CALL) ist dies ein sehr wichtiger Aspekt. Dadurch bekämen die Lernenden das Feedback in der eigenen Stimme, unmittelbar nachdem sie die Äußerung ausgesprochen haben, und sie wären dann in der Lage, die korrekte und die inkorrekte Aussprache direkt miteinander zu vergleichen. Für dieses Experiment wurden den Probanden die prosodisch korrigierten Äußerungen aus zweierlei Gründen nicht unmittelbar angeboten. Erstens enthielt das Textmaterial einige sehr lange Sätze, was eine Vorbereitung des Signals des italienischen Sprechers erforderte, um seine Sprechpausen denen

des Signals der deutschen Muttersprachlerin anzupassen. Das passierte insbesondere, wenn zwei Signale eine ganz unterschiedliche Dauerstruktur hatten, zum Beispiel wenn der italienische Sprecher sehr langsam sprach und zauderte. Ohnehin funktionierte der DTW-Algorithmus gut bei Einzelwörtern und kurzen Sätzen. Der zweite Grund war, dass wir die Qualität jedes resynthetisierten Stimulus, der als Feedback benutzt werden sollte, kontrollieren wollten. Es war wichtig, dass die Lernenden keine Äußerungen schlechter Qualität als Feedback erhielten. Zum Beispiel wurden einige Äußerungen verworfen, weil *creaky voice* oder eine starke Nasalierung auf der zu betonenden Silbe im Signal der italienischen Sprecher eine erfolgreiche Korrektur mittels Resynthese verhinderte. Eine hochqualitative Resynthese mancher Äußerungen war nicht möglich, weil die Signale der deutschen und der italienischen Sprecher hinsichtlich Dauer und Stimmqualität zu unterschiedlich waren. Wir sollten auch erwähnen, dass die italienischen Sprecher noch nie einen deutschen Muttersprachler beim Lesen dieses Textes gehört hatten. Dies war eine viel schwierigere Situation als das übliche Training mit einem CALL-Programm. Normalerweise sollten die Lernenden eine Äußerung wiederholen, nachdem sie deren Aussprache von einem Muttersprachler gehört haben. Wenn die Sprecher in diesem Experiment die Äußerungen nach vorherigem Anhören der Aussprache durch einen deutschen Muttersprachler ausgesprochen hätten, wäre die Manipulation ihrer Signale wahrscheinlich einfacher gewesen. Die Implementierung dieser von uns getesteten Technik in ein CALL-Programm könnte dies künftig ermöglichen.

Darüber hinaus wurde die Qualität der resynthetisierten Stimuli von den Trainingsteilnehmern insofern gut angenommen, als die Stimuli einen motivierenden Effekt auf sie ausübten.

8.4 Experiment 4: Perzeptionstest zur Evaluierung der Trainingsdaten

Um die Korrektheit der Betonung der morphologisch komplexen Wörter vor und nach dem Training zu evaluieren, wurde ein Perzeptionstest mit deutschen Muttersprachlern durchgeführt. Das Ziel des Perzeptionstests war es, zu verifizieren, a) ob Feedback in der eigenen Stimme effektiver als Feedback aus der aufgenommenen Stimme der Muttersprachlerin ist, b) ob Feedback aus überbetonten Stimuli effektiver als Feedback aus normal betonten Stimuli ist, und c) ob Feedback aus resynthetisierten überbetonten Stimuli am effektivsten ist.

8.4.1 Methode

8.4.1.1 Stimuli

Im Perzeptionstest wurden nur die ursprünglich falsch betonten Äußerungen und die entsprechenden nach dem Training aufgenommenen Versionen, präsentiert. Jedoch waren es — wie bereits für die Experimente 1 und 2 geschehen — zu viele Äußerungen. Daher reduzierten wir die Anzahl der Stimuli auf 255. Diese bestanden aus 85 vor dem Training mit jeweils zwei dazugehörigen nach dem Training aufgenommenen Äußerungen. Ausgewählt wurden die im Training verwendeten Äußerungen, die bei der Aufnahme vor dem Training am häufigsten falsch betont worden waren. Es handelte sich um acht isoliert ausgesprochene und sechs in Sätze eingebettete morphologisch komplexe Wörter.

8.4.1.2 Testteilnehmer

Der Perzeptionstest wurde von 37 deutschen Muttersprachlern im Alter zwischen 21 und 53 Jahren durchgeführt. 25 von ihnen waren Mitarbeiter oder Stu-

zenten des Instituts für Phonetik der Ludwig-Maximilians-Universität München, und mehrere von ihnen hatten bereits an den Perzeptionstests für die Experimente 1 oder 2 teilgenommen.

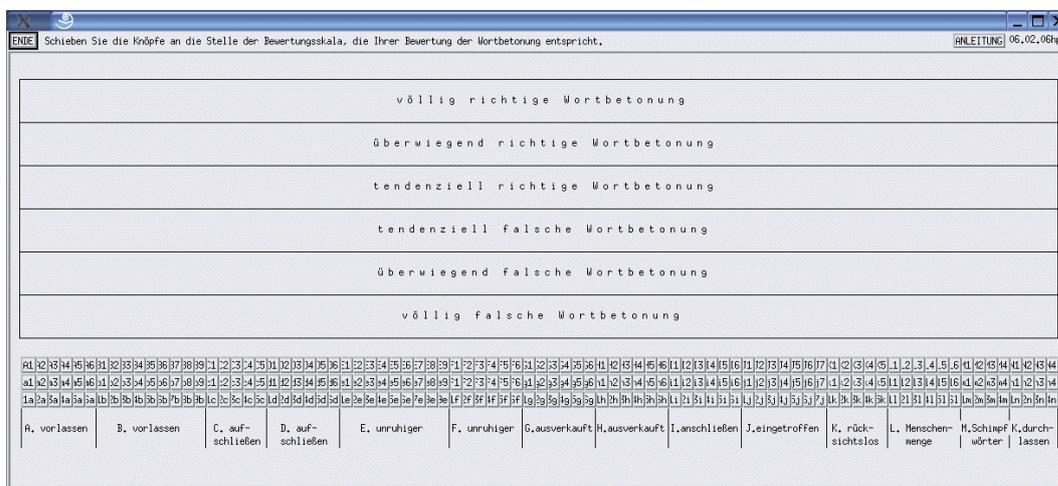


Abbildung 8.9: Benutzeroberfläche des Perzeptionstests.

Der Test wurde mit dem Tool CoDIT (s. Pfitzinger 2007 [118]) analog zu den Experimenten 1 und 2 implementiert, allerdings gab es in diesem Test drei Stimuli pro Spalte (s. Abb. 8.9). Hinter den drei Knöpfen jeder Spalte mit der selben Buchstaben-Zahlen-Kombination verbargen sich die vor dem Training und die zwei entsprechenden nach dem Training aufgenommenen Äußerungen desselben Sprechers. Auch in diesem Test war sowohl die Zuordnung der drei Versionen in jeder Spalte als auch die Reihenfolge der Sprecher für jede Äußerung zufällig. Die Durchführung des Tests durch die Probanden geschah ebenfalls nach den selben Regeln wie der Test im Experiment 2 (s. Kap. 8.2.1.3). Abb. 8.10 zeigt die Testoberfläche, nachdem ein Teilnehmer den Test beendet hatte.

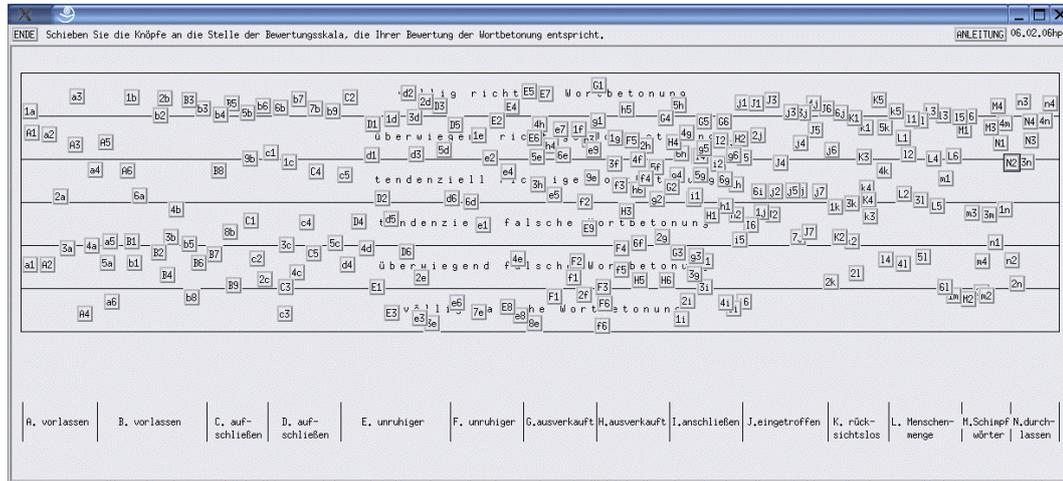


Abbildung 8.10: Test von Teilnehmer al.

8.4.2 Ergebnisse des Perzeptionstests und Diskussion

Die Bewertungsergebnisse der deutschen Muttersprachler wurden, wie bereits für Experiment 2 entschieden, auf eine Skala mit Werten von 0 (schlechtmögliche Bewertung) bis 100 (bestmögliche Bewertung) übertragen.

Die Tabelle 8.3 zeigt die Mittelwerte der drei Versionen (vor dem Training, nach dem Training mit normal betonten Stimuli und nach dem Training mit überbetonten Stimuli). Die Versionen nach dem Training, die nach Feedback aus resynthetisierten oder überbetonten Stimuli ausgesprochen wurden, erhielten die höchsten Bewertungsergebnisse.

Abb. 8.11 stellt die durch Feedback aus normal betonten und aus überbetonten Stimuli verursachte Verbesserung jeweils für die mit natürlichen und für die mit resynthetisierten Stimuli trainierte Gruppe dar. Die Werte auf der y-Achse errechnen sich durch Subtraktion der Bewertungsergebnisse der Aussprache vor dem Training von denen der entsprechenden nach dem Training ausgesprochenen Stimuli. Das Feedback aus überbetonten Stimuli bewirkte in

Versuchspersonen trainiert durch		
	natürliche Stimuli	resynthetisierte Stimuli
vor dem Training	37.35	38.13
normal betonte Stimuli	55.21	61.36
überbetonte Stimuli	62.70	63.80

Tabelle 8.3: Mittelwerte der Bewertungsergebnisse der drei Versionen der Äußerungen von Trainingsteilnehmern trainiert durch natürliche und durch resynthetisierte Stimuli.

beiden Teilnehmergruppen eine größere Verbesserung als das Feedback aus normal betonten Stimuli.

Da die Daten nicht normalverteilt waren, wurden für die statistische Analyse nichtparametrische Signifikanztests verwendet. Um zu testen, ob die Bewertungsergebnisse der Äußerungen vor dem Training und der zwei Versionen nach dem Training signifikant unterschiedlich waren, wurde der Wilcoxon Test angewendet.

Der Wilcoxon Rank Sum Test für unabhängige Stichproben zeigt keinen signifikanten Unterschied zwischen den Bewertungsergebnissen der vor dem Training ausgesprochenen Äußerungen der Test- und der Kontrollgruppe ($p = 0.1099$). Das deutsche Sprachniveau beider Gruppen kann als äquivalent betrachtet werden. Stattdessen zeigte der Vergleich zwischen den Stimuli nach dem Training mit normal betonten Stimuli, dass das Feedback aus resynthetisierten Stimuli in der Stimme des Lerners effektiver als das Feedback mit natürlichen Stimuli (Wilcoxon rank sum test, $p < 0.001$) war. Um den Vergleich durchzuführen, wurden die Verbesserungswerte wie in Abb. 8.11 verwendet.

Die Anwendung des Wilcoxon Signed Rank Tests für abhängige Stichproben auf die Bewertungen der Äußerungen der mit natürlichen Stimuli trainierten Gruppe zeigt einen signifikanten Unterschied zwischen den Bewertungsergebnissen der Äußerungen vor und nach dem Training ($p < 0.001$). Die nach Feed-

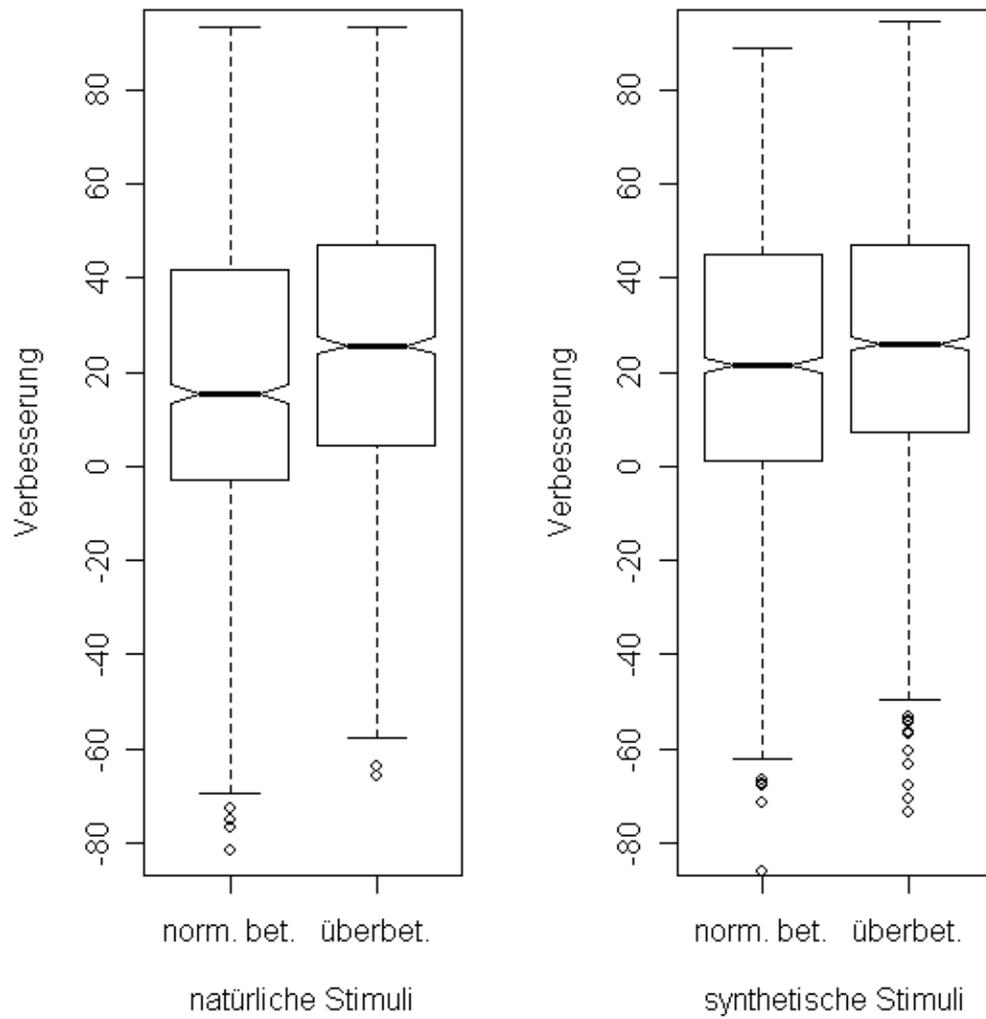


Abbildung 8.11: Verbesserung nach Feedback aus normal betonten Stimuli bzw. nach Feedback aus überbetonten Stimuli für die mit natürlichen Stimuli trainierte Gruppe und die mit resynthetisierten Stimuli trainierte Gruppe.

back aus natürlichen Stimuli gesprochenen Äußerungen waren nach Feedback aus überbetonten Stimuli signifikant besser als nach Feedback mit normal betonten Stimuli ($p < 0.001$).

Wir verglichen auch die drei Versionen der mit resynthetisierten Stimuli trainierten Gruppe. Beide Versionen, sowohl die nach Feedback mit normal betonten Stimuli als auch die nach Feedback mit überbetonten Stimuli, bekamen signifikant bessere Bewertungsergebnisse als die vor dem Training aufgenommene Version ($p < 0.001$). Zwischen den beiden Versionen nach dem Training war der Unterschied fast signifikant ($p = 0.052$). Wir erwarteten, dass Feedback aus resynthetisierten überbetonten Stimuli effektiver als Feedback aus resynthetisierten normal betonten Stimuli gewesen wäre, und wahrscheinlich hätten wir mit mehr Daten eine höhere Signifikanz nachweisen können. Jedoch imitierten viele italienische Sprecher die überbetonte Version. Aus diesem Grund bewerteten die deutschen Muttersprachler diese Version als korrekt, aber möglicherweise bewerteten sie sie niedriger, als wenn das Wort nicht überbetont gewesen wäre. Zwölf der 37 deutschen Perzeptionstestteilnehmer wurden gefragt, ob sich im Test überbetonte Stimuli befanden, und wie sie solche Stimuli bewertet hatten. Alle zwölf erklärten, dass sie überbetonte Stimuli gehört hatten. Acht von zwölf Testteilnehmern sagten, sie bewerteten solche Stimuli als korrekt, weil die Sprecher in diesen Stimuli die richtige Silbe betont hatten. Trotzdem bewerteten sie Stimuli in solchen Fällen schlechter, als sie es bei normal betonten Stimuli getan hätten.

Abb. 8.12 stellt die Mediane der drei Versionen für die 52 Äußerungen der mit resynthetisierten Stimuli trainierten italienischen Sprecher dar. Die Mediane sind auf der Benutzeroberfläche des Perzeptionstests dargestellt, und sie sind vom niedrigsten bis zum höchsten Wert gemäß dem Feedback mit normal betonten Stimuli angeordnet. Die Mediane der Version nach dem Training mit überbetonten Stimuli sind auf dem oberen Teil der Grafik platziert, und dies zeigt, dass sie von den deutschen Muttersprachlern als korrekt betont bewer-

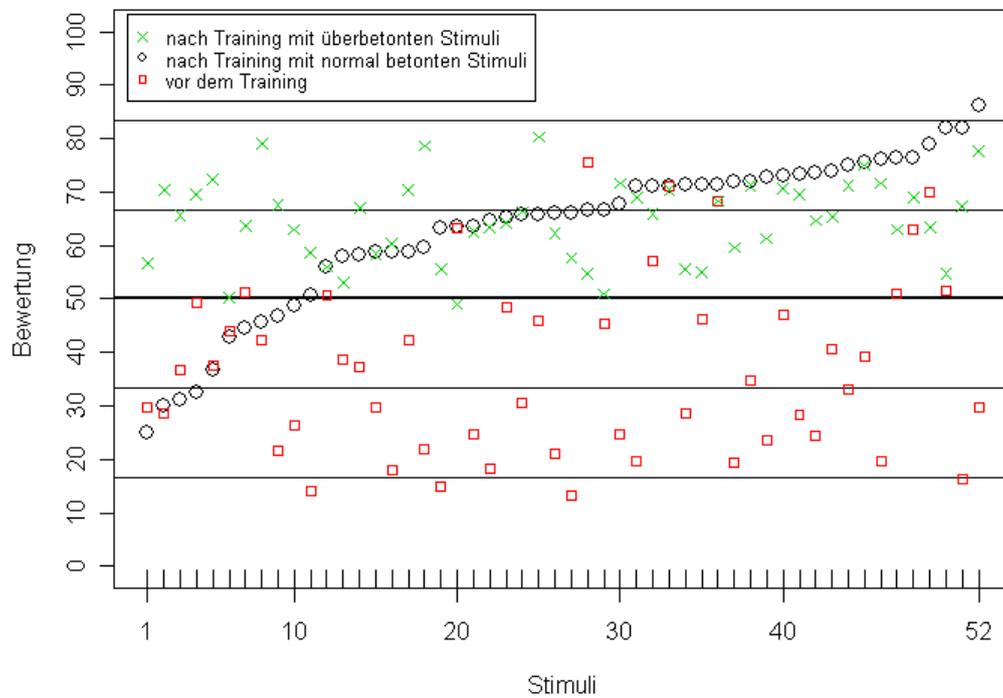


Abbildung 8.12: Mediane der drei Versionen für die mit resynthetisierten Stimuli trainierten italienischen Sprecher. *Rechtecke* = vor dem Training aufgenommene Version. *Kreise* = Version nach dem Training mit normal betonten Stimuli. *Kreuze* = Version nach dem Training mit überbetonten Stimuli.

tet wurden. Jedoch wurden sie meistens schlechter bewertet als die Stimuli nach dem Feedback aus normal betonten Stimuli, möglicherweise weil sie von den Testteilnehmern wegen der Überbetonung als schlechter betont „bestraft“ wurden. Ganz links in Abb. 8.12 befinden sich die zehn der 52 Stimuli nach Feedback mit normaler Betonung, die als inkorrekt bewertet wurden, und die deshalb unter der Mittellinie landeten. Die entsprechenden nach Feedback mit Überbetonung ausgesprochenen Stimuli wurden als korrekter beurteilt, und sind daher in der oberen Hälfte platziert. Im Fall dieser Stimuli waren die italienischen Sprecher möglicherweise nicht in der Lage, die Betonung auf der korrekten Silbe auszusprechen, nachdem sie das Feedback aus normal betonten Stimuli gehört hatten, und die Überbetonung war für sie notwendig, um die korrekte Betonungsstelle zu erkennen und reproduzieren.

Der Vergleich zwischen resynthetisierten und natürlichen überbetonten Stimuli zeigte, dass überbetontes Feedback in der Stimme des Lernalters nicht effektiver als das überbetonte Feedback in der Stimme der deutschen Muttersprachlerin war (Wilcoxon Rank sum test for unpaired samples, $p = 0.5956$).

Wir hatten erwartet, dass Feedback aus überbetonten Stimuli in der Stimme des italienischen Sprechers effektiver als Feedback aus überbetonten Stimuli in der Stimme der deutschen Muttersprachlerin gewesen wäre. Ein Grund dafür, dass es nicht so war, könnte sein, dass der Vergleich zwischen der inkorrekten und der korrekten Version in seiner eigenen Stimme für den Lerner ausreichend war, um den Unterschied wahrzunehmen, und dass möglicherweise die Überbetonung keine weitere Verbesserung brachte. Die Ergebnisse könnten auch vom Testdesign und von der Tatsache, dass die Testteilnehmer die überbetonten Versionen „bestraften“, beeinflusst worden sein. Die Testteilnehmer sollten drei Stimuli miteinander vergleichen: die vor dem Training ausgesprochene Version und die beiden nach dem Training ausgesprochenen Versionen. Wenn die Stimuli von der mit resynthetisiertem Feedback trainierten Gruppe stammten, verglichen die Testteilnehmer möglicherweise häufiger zwei auf

der korrekten Silbe betonte Stimuli miteinander. Wenn einer der Stimuli überbetont war, wurde er dann in der Bewertung „bestraft“. Die Stimuli der mit natürlichem Feedback trainierten Gruppe enthielten häufiger nur eine, auf der korrekten Silbe betonte Version, und zwar die nach dem überbetonten Stimulus ausgesprochene Äußerung. War diese Äußerung überbetont, wurde sie möglicherweise von den Testteilnehmern weniger „bestraft“, da sie die einzige richtige war.

8.4.3 Betrachtungen nach dem Experiment 4

Um eine Fremdsprache korrekt zu lernen, sollten Lernende in die Lage versetzt werden, den Unterschied zwischen richtiger und falscher Aussprache zu begreifen. Dies betrifft italienische Sprecher, die möglicherweise nicht in der Lage sind, die korrekte Betonung morphologisch komplexer Wörter im Deutschen nachzuvollziehen, ohne dass sie entsprechende Hinweise dafür bekommen. Unserer Hypothese zufolge wären in diesem Fall sowohl resynthetisierte Stimuli in der Stimme des Lerners als auch überbetonte Stimuli geeignete Arten des Feedbacks.

Das Hauptergebnis dieses Experiments ist, dass resynthetisierte und/oder überbetonte Stimuli eine effektivere Art des Feedbacks für Aussprachetraining sind als natürliche normal betonte Stimuli.

Da die Resynthese der lokalen Sprechgeschwindigkeit auch ohne Handsegmentierung durchgeführt werden kann, könnte eine solche Technik verwendet werden, um effektivere CALL-Programme zu schaffen. Die Ergebnisse belegten zudem, dass resynthetisierte Äußerungen mit normaler Betonung effektiver als normal betonte natürliche Äußerungen sind. Jedoch haben sich resynthetisierte überbetonte Stimuli als gleich effektiv wie natürliche überbetonte Stimuli erwiesen.

Mögliche Ursachen für dieses Ergebnis können folgende sein: a) die Kombination aus Überbetonung und resynthetisierten Stimuli in der Stimme des Lerner bewirkte keine weitere Verbesserung, da das auf eine einzige der beiden Methoden basierende Feedback für den Lerner ausreichend sein könnte, um den Unterschied zwischen der korrekten und der unkorrekten Aussprache zu erkennen, und b) die Perzeptionstestteilnehmer haben möglicherweise die überbetonten Äußerungen der mit resynthetisierten Stimuli trainierten Gruppe häufiger als die der anderen Gruppe bestraft (s. dazu Kap. 8.4.2).

Der Unterschied zwischen Feedback aus überbetonten und aus normal betonten resynthetisierten Stimuli war fast signifikant. Mit mehr Daten hätten wir möglicherweise eine höhere Signifikanz erreicht. Nichtsdestoweniger könnte die Tatsache, dass überbetonte Stimuli als weniger korrekt bewertet wurden, ein Grund für dieses Ergebnis sein.

Schließlich war ein deutliches Ergebnis, dass natürliche überbetonte Stimuli effektiver als natürliche normal betonte Stimuli waren.

Wir hatten nicht genügend italienische Teilnehmer, um vier Gruppen zu bilden, von denen jeweils eine Gruppe jeweils eine Art von Feedback erhalten hätte: 1. Feedback aus natürlichen normal betonten Stimuli, 2. Feedback aus resynthetisierten normal betonten Stimuli, 3. Feedback aus natürlichen überbetonten Stimuli, und 4. Feedback aus resynthetisierten überbetonten Stimuli. Auf diese Weise hätten die Perzeptionstestteilnehmer nur zwei statt drei Stimuli miteinander vergleichen müssen, und die nach dem Training ausgesprochenen Äußerungen hätten sie ausschließlich mit der ursprünglichen, vor dem Training ausgesprochenen Äußerung vergleichen müssen.

Mit vier Gruppen von Trainingsteilnehmern hätten wir zudem Übertragungseffekte zwischen den zwei Trainingsteilen vermieden. Jedoch bekamen die italienischen Lernenden im zweiten Teil des Trainings durch die überbetonten Stimuli eine neue Information, die sie im ersten Teil nicht bekommen

hatten. Das mehrmals wiederholte Hören des normal betonten Stimulus hätte nicht dieselbe Wirkung gehabt, denn die Lernenden wären nicht in die Lage versetzt worden, die korrekte Betonungsstelle wahrzunehmen.

8.5 Experiment 5: Perzeption der Betonung morphologisch komplexer Wörter im Deutschen durch italienische Sprecher

Unsere bisherigen Ergebnisse deuten darauf hin, dass die Schwierigkeiten italienischer Sprecher, die Betonung morphologisch komplexer Wörter im Deutschen zu lernen, perzeptive Ursachen haben. Die Beobachtung, dass italienischsprachige Personen, die seit zehn oder mehr Jahren in Deutschland leben, diesen systematischen Fehler nie bemerkten, stützt die Hypothese. In diesem Fall handelt es sich um eine Ausspracheeigenschaft, die für italienischsprachige Lernende eigentlich nicht schwierig zu reproduzieren ist, vorausgesetzt, sie nehmen sie wahr. Zudem zeigen unsere Experimente zum Training der Wortbetonung (s. Kap. 8.3 und Kap. 8.4), dass überbetonte und resynthetisierte Stimuli in der Stimme der Lernenden ein effektiveres Feedback als normal betonte bzw. natürliche Stimuli sind. Die Überbetonung ist hilfreich, da italienische Lernende einen Hinweis brauchen, um die betonte Silbe zu identifizieren. Zudem hilft ihnen die Reduktion der phonetischen Variabilität in den resynthetisierten Stimuli dabei, die unkorrekte von der korrekten Betonung zu unterscheiden. Die Ergebnisse des Perzeptionstests im Experiment 2 (s. Kap. 8.2) zeigten, dass die Resynthese der Grundfrequenz einen stärkeren Einfluss auf die Korrektur der Betonung hat als die Resynthese der lokalen Sprechgeschwindigkeit, und dass daher die von den Italienern falsch betonten Wörter hauptsächlich aufgrund ihrer Grundfrequenz von der korrekten Betonung abwichen. Da sich italienisch-

sprachige Lernende möglicherweise auf die Dauer der Silben verlassen, um die Betonung zu erkennen, könnten wir annehmen, dass sie Schwierigkeiten haben, die Wortbetonung mit höherer Grundfrequenz in Zusammenhang zu bringen, insbesondere wenn der Vokal in der unbetonten Silbe lang ist.

In diesem Experiment möchten wir testen, ob die Schwierigkeit italienischsprachiger Lernender, die korrekte Betonung morphologisch komplexer Wörter im Deutschen zu lernen, von ihrer sprachspezifischen phonologischen Darstellung der Betonung abhängt, die sie zum Beispiel dazu bringt, Minimalpaare wie *'umfahren* und *umf'ahren* dem selben Betonungsmuster zuzuordnen. Italienische Deutschlernende könnten in der Lage sein, eine Aufgabe wie einen einfachen Diskriminationstest von deutschen Betonungsminimalpaaren zu lösen, da die Betonung durch verschiedene prosodische Parameter ausgedrückt wird, die sie akustisch wahrnehmen können. Jedoch könnte ein solcher Test nicht zeigen, wie diese Art von Wörtern auf einem abstrakteren Niveau wahrgenommen wird. Daher verwendeten wir die Methode von Dupoux, Peperkamp & Sebastián-Gallés 2001 [38] (s. dazu Kap. 2.3.4), die phonetische Variabilität, Einsatz des Kurzzeitgedächtnisses und Zeitbeschränkung in die Aufgabe miteinbezieht. Die Methode wurde von den Autoren verwendet, um die Betonungswahrnehmung von französischen und später auch von finnischen, ungarischen und polnischen Sprechern zu untersuchen (Peperkamp & Dupoux 2002 [113]). In unserem Fall testen wir Sprecher einer Sprache, Italienisch, in der die Betonung stattdessen kontrastiv ist und in der zahlreiche Betonungsminimalpaare existieren. In der Tat könnte die unterschiedliche Relevanz der einzelnen akustischen Korrelate der Betonung in den einzelnen Sprachen der Grund dafür sein, dass Sprecher einer betonungskontrastiven Sprache Schwierigkeiten haben, die Betonung einer anderen betonungskontrastiven Sprache wahrzunehmen.

8.5.1 Methode

Die von uns verwendete Methode basiert — von einigen Abweichungen abgesehen — auf Experiment 4 von Dupoux et al. 2001 [38]. Während jedoch Dupoux et al. die Grundfrequenz der Stimuli mittels Resynthese variierten, verwendeten wir ausschließlich natürliche Stimuli. Wir wollten möglichst keine Artefakte durch unnatürliche Grundfrequenzkonturen in die Stimuli einbauen. Die notwendige phonetische Variabilität wurde durch die Verwendung von Stimuli eines männlichen und eines weiblichen Sprechers gewährleistet. Die italienischen Versuchspersonen wurden mit drei Wortpaaren getestet: a) einem Phonemkontrast, b) einem deutschen Betonungskontrast, und c) einem Betonungs-Quantitätskontrast, der einem Betonungskontrast im Italienischen ähnlich sein soll, in dem die Betonung durch längere Vokaldauer „unterstützt“ wird. Der Phonemkontrast war für jeden Teilnehmer derselbe. Die Stimuli für die anderen beiden wurden aus zehn unterschiedlichen Wortpaaren per Zufall ausgewählt. Derselbe Perzeptionstest wurde ebenfalls von deutschen Muttersprachlern als Kontrollgruppe durchgeführt.

Nach unserer Hypothese sollten die italienischen Probanden mehr Schwierigkeiten mit dem Betonungskontrast, den sie möglicherweise nicht ohne Weiteres einer phonologischen Kategorie zuordnen können, haben als mit dem Phonem- und mit dem Betonungs-Quantitätskontrast. Die deutschen Probanden sollten bei allen drei Kontrasten ähnlich gut abschneiden.

Da wir deutsche Wörter und Pseudowörter als Stimuli verwendeten, hatten die deutschen gegenüber den italienischen Versuchspersonen einen Vorteil. Jedoch wollten wir zunächst den Test mit deutschen Wörtern durchführen, um zu sehen, ob der Effekt überhaupt auftreten würde. Wir planen einen entsprechenden Test mit für beide Sprechergruppen unbekanntem Nichtwörtern. Dafür wurden bereits Aufnahmen durchgeführt (s. Kap. 8.6).

8.5.1.1 Auswahl der Stimuli

Es wurden mehrere vorbereitende Aufnahmen durchgeführt, um die Stimuli für den Perzeptionstest zu finden. Zunächst hatten wir vor, bestehende deutsche Minimalpaare zu verwenden. Jedoch erwiesen sich diese Wörter für unseren Zweck als ungeeignet, da keines ein /a/ sowohl in der ersten als auch in der zweiten Silbe enthielt. Wie wir feststellten, könnte die unterschiedliche Vokalqualität in betonten und unbetonten Silben den Testteilnehmern einen zusätzlichen Hinweis liefern. Im Folgenden werden die ersten mit Minimalpaaren unternommenen Aufnahmeversuche und die definitive Auswahl der Stimuli beschrieben.

8.5.1.1.1 Aufnahmeversuche mit deutschen Minimalpaaren Die deutschen Wörter, die als Stimuli in dem Perzeptionstest dienen sollten, wurden per Zufall aus der CELEX Datenbank ausgewählt (s. CELEX 2004 [13]), die 51728 deutsche Lemmata enthält. Es wurden 170 Betonungsminimalpaare in der Datenbank gefunden. Aus den 99 dreisilbigen Minimalpaaren, die darin enthalten waren, wurden zunächst zehn nach dem Zufallsprinzip ausgewählt. Es handelte sich um neun Minimalpaare aus Präfixverben und das Minimalpaar '*Misstrauen* vs. *miss'trauen*, und sie wurden von einem männlichen und einer weiblichen Sprecher vorgelesen. Dazu mussten die Sprecher – ausgebildete Phonetiker – die auf der ersten Silbe betonten Wörter nochmals mit extra gekürzter Pänultima vorlesen. Dafür wurden die Wörter in phonetischer Transkription präsentiert. Jedoch war es für die Sprecher schwierig, die Stimuli wie gewünscht zu produzieren, da es sich in den meisten Fällen darum handelte, einen Kurzvokal oder einen Diphthong überkurz auszusprechen. Aus diesem Grund wurde beschlossen, nur Stimuli zu verwenden, in denen die Längung oder Kürzung der Pänultima mittels Orthographie darstellbar wäre.

Es wurden daher weitere acht Minimalpaare aus der Datenbank CELEX

zufällig ausgewählt: vier, in denen die Pänultima lang war, und vier, in denen die Pänultima kurz war. Die Minimalpaare, in denen die Verkürzung bzw. Verlängerung der Pänultima mittels Orthographie nicht darstellbar war, wurden aussortiert und durch weitere ersetzt.¹ Es wurden nur Minimalpaare mit dem Vokal /a/ oder /a:/ in der Pänultima gesucht. Wenn man Wörter mit weiteren Vokalen genommen hätte, wäre der Unterschied zwischen den Kurz- und Langvokalen nicht nur quantitativer, sondern auch qualitativer Art (gespannt vs. ungespannt) gewesen (z.B. /i:/ vs. /ɪ/ und /u:/ vs. /ʊ/). Das Minimalpaar 'umschlagen vs. um'schlagen wurde durch das in CELEX nicht enthaltene Minimalpaar 'umfahren vs. um'fahren ersetzt, um eine größere Vielfalt im Korpus zu gewährleisten, da das Minimalpaar aus *durchschlagen* ebenfalls per Zufall gewählt worden war. Die somit gewählten Minimalpaare sind im Anhang E.1 aufgelistet.

In einem informellen Test, sollte ein italienischsprachiger Proband einige der Stimuli mit Langvokal in der Pänultima hören und mit der Betonung transkribieren. Der Proband behauptete, dass auch die auf der ersten Silbe betonten Wörter auf der Pänultima betont wären. Jedoch war er offensichtlich in der Lage, die unterschiedlich betonten Wörter durch den Vokal im Präfix [ʊm-] zu unterscheiden. Wenn das Präfix unbetont war, transkribierte er den Vokal als <a>, da er die Zentralisierung des Vokals [ʊ] eher dem Vokal /a/ als dem /u/ des Italienischen zuordnen konnte. Die italienischen Sprecher hätten daher — trotz einer möglicherweise unkorrekten Wahrnehmung der Betonung — die Möglichkeit, die Wörter der Minimalpaare voneinander zu unterscheiden. Aus diesem Grund entschieden wir uns, für den Perzeptionstest nur Wörter zu benutzen, die sowohl in der zweiten als auch in der ersten Silbe den Vokal /a/ enthielten. Dies sollte vor extremeren Vokalqualitätsunterschieden zwischen

¹Zum Beispiel ist in *umblasen* die Kürzung des Vokals /a:/ durch das Graphem <ss> darstellbar. Daraus ergibt sich aber kein Minimalpaar, da das Phonem /z/ durch /s/ ersetzt wird: /'ʊmbla:zən/ und */'ʊmblasən/.

den unterschiedlich betonten Wörtern schützen.

8.5.1.1.2 Gewählte Stimuli Wie vermutet, ergab die Suche nach deutschen dreisilbigen Minimalpaaren mit /a/ sowohl in der ersten als auch in den zweiten Silbe in der Datenbank CELEX keine Ergebnisse. Deswegen suchten wir in der Datenbank dreisilbige Wörter mit einer Morphemgrenze zwischen der ersten und der zweiten Silbe, mit dem Vokal /a/ sowohl in der ersten als auch in der zweiten Silbe und Betonung auf der ersten Silbe. Die Suche in der Datenbank ergab 156 Wörter mit dem Langvokal /a:/ und 276 mit dem Kurzvokal /a/ in der Pänultima. Aus dieser Menge wurden fünf mit Langvokal und fünf mit Kurzvokal nach dem Zufallsprinzip ausgewählt.

Die Wörter wurden mit drei unterschiedlichen Aussprachen von den Sprechern vorgelesen.

1. Die erste Version wurde korrekt mit der Betonung auf der ersten Silbe vorgelesen.
2. Dann wurde das Wort mit der Betonung auf der Pänultima vorgelesen.
3. Die dritte Version war für die Wörter mit Lang- und Kurzvokal unterschiedlich:
 - a) In den Wörtern ursprünglich mit Langvokal in der Pänultima wurde dieser gekürzt und die erste Silbe betont (zum Beispiel *'Anlage statt 'Anlage).
 - b) In den Wörtern ursprünglich mit Kurzvokal in der Pänultima, wurde dieser verlängert und die Pänultima wurde betont (zum Beispiel *an'klahmern statt 'anklammern)

Aus den drei Versionen jedes Wortes konnten dann zwei Wortpaare gebaut werden: ein Betonungs- und ein Betonungs-Quantitätskontrast (s. Ta-

ble 8.4). Die Quantität sollte, wie bei den Minimalpaaren, mittels Orthographie ausgedrückt werden.

Stimuli für den Betonungs- bzw. Betonungs-Quantitätskontrast		
lang und unbetont	lang und betont	kurz und unbetont
1 'Stadtplanung	*Stadt'planung	*'Stadtplanung
2 'Anlage	*An'lage	*'Anlagge
3 'abstrahlen	*ab'strahlen	*'abstrallen
4 'abnabeln	*ab'nabeln	*'abnabbeln
5 'Astgabel	*Ast'gabel	*'Astgabbel
kurz und unbetont	kurz und betont	lang und betont
6 'absacken	*ab'sacken	*ab'saaken
7 'abflachen	*ab'flachen	*ab'flaachen
8 'anklammern	*an'klammern	*an'klahmern
9 'abzwacken	*ab'zwacken	*ab'zwaacken
10 'anstarren	*an'starren	*an'stahren
Stimuli für den Phonemkontrast		
alveolar lateral-approximant		bilabial nasal
11 *ab'lagern		*ab'magern

Tabelle 8.4: Stimuli für den Perzeptionstest, angeordnet entsprechend der Betonung und Quantität von /a/ in der Pänultima bzw. des vorherigen Konsonants.

Die Wörter des Phonemminimalpaars, das als Baseline dienen sollte, sollten ebenso wie die Wörter der Betonungsminimalpaare dreisilbig sein, ein /a/ sowohl in der ersten als auch in der zweiten Silbe haben und eine Morphemgrenze zwischen der ersten und zweiten Silbe aufweisen. Dazu sollten sie einen im Italienischen vorhandenen Betonungskontrast aufweisen und dem geläufigen italienischen Betonungsmuster mit betonter langer Pänultima entsprechen. Da in der Datenbank keine Entsprechungen zu diesen Kriterien zu finden waren, wurde stattdessen ein auf der ersten Silbe betontes Phonemminimalpaar gesucht, das dann die Sprecher auf der Pänultima betonen sollten. Nach dem

Zufallsprinzip wurde dann das Minimalpaar *ablagern* vs. *abmagern* gewählt.

Die Wörter wurden sowohl von einer Frau als auch von einem Mann, beide ausgebildete Phonetiker, vorgelesen. Sie waren deutsche Muttersprachler aus Bayern, und ihr Deutsch war nicht dialektal geprägt. Die Frau war 22 und der Mann 30 Jahre alt. Sie verfügten über keine bzw. sehr geringe Italienischkenntnisse.

Die Daten wurden unter denselben Bedingungen wie für die Experimente 2 und 3 aufgenommen (s. Kap. 8.2.1.1). Jedoch wurden diesmal die zu lesenden Wörter auf einem Bildschirm präsentiert (s. Abb. 8.13).

Die Stimuli wurden mit dem Software SpeechRecorder aufgenommen (s. Draxler & Jänsch 2004 [34] und Draxler & Jänsch 2007 [35]). Um sich an die Aussprache der Nichtwörter zu gewöhnen, lasen die Sprecher die gesamten Wörter zunächst einmal. Die Anweisungen und die zwei Übungsfensterprompts mit dem gesamten Textmaterial, die die Sprecher vor der Aufnahme erhielten, sind in Anhang E.2 dargestellt. Im Folgenden wurde jedes Wort drei mal hintereinander vorgelesen, um einen Kontrastakzent zu vermeiden (s. Abb. 8.14), und nur das dritte Wort wurde dann für die Untersuchung verwendet.

Das gesamte Wortmaterial wurde von jedem Sprecher fünf mal vorgelesen. Da wir für den Perzeptionstest sechs Tokens für jede Wortversion benötigten, wurden drei der fünf Realisierungen für jeden Sprecher gewählt. In einem informellen Test wurden die gewählten Stimuli einem deutschen Muttersprachler ohne phonetische Vorkenntnisse präsentiert. Er sollte die betonte Silbe und die Vokalquantität der Pänultima in den Stimuli erkennen, und er tat dies fehlerfrei.



Abbildung 8.13: Studioeinstellungen für die Aufnahme. Ein Pop-Filter wurde ebenfalls verwendet.

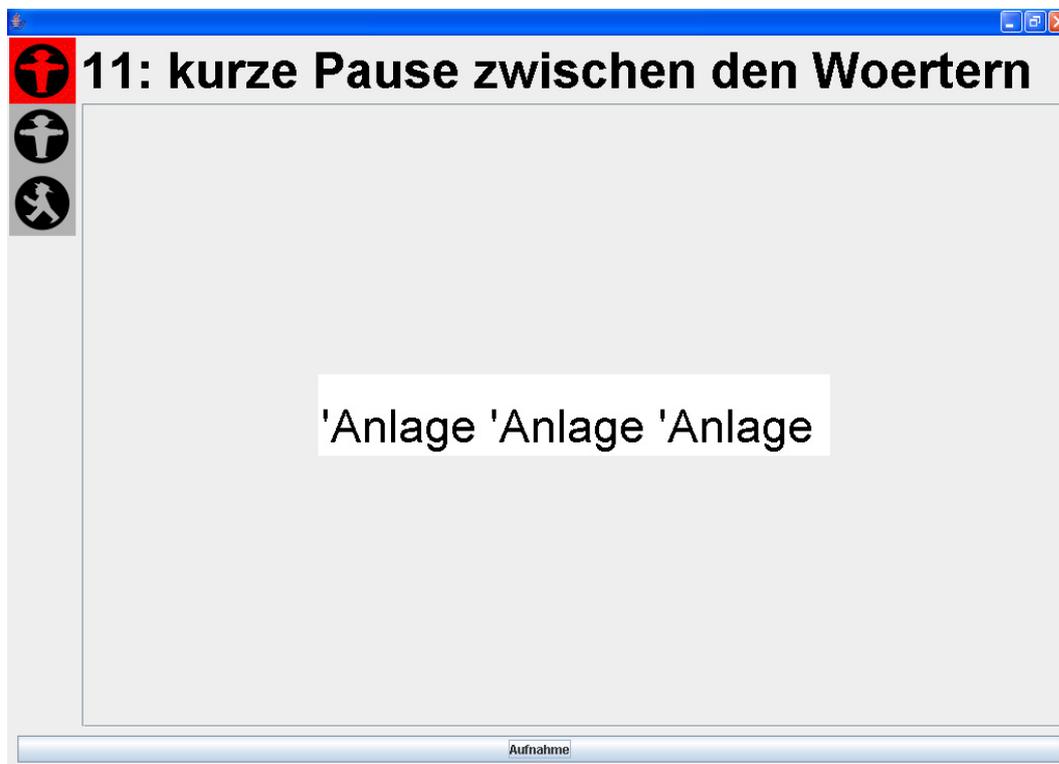


Abbildung 8.14: Beispiel eines Textprompts für die Aufnahme.

8.5.1.2 Akustische Untersuchungen der Dauer der Stimuli

Die Stimuli für den Perzeptionstest umfassten insgesamt 10 Wörter \times 3 Versionen \times 6 Wiederholungen = 180 Tokens für den Betonung- und 2 Wörtern \times 6 Wiederholungen = 12 Tokens für den Phonemkontrast. Alle Stimuli wurden manuell segmentiert, anschließend wurden akustische Messungen der Stimuli durchgeführt. Es wurde die Gesamtdauer der Stimuli, die Dauer der Silben und die des Vokals /a/ in der ersten und zweiten Silbe der Stimuli des Betonungskontrasts gemessen.

Die durchschnittliche Dauer der gesamten Stimuli beträgt 925.8 ms. Tab. 8.5 stellt die durchschnittliche Dauer der Stimuli für die drei Versionen der Wör-

Stimuli für den Betonungs- bzw. Betonungs-Quantitätskontrast		
lang und unbetont	lang und betont	kurz und unbetont
999.2	999.7	887.3
kurz und unbetont	kurz und betont	lang und betont
844.6	843.8	944.4
Stimuli für den Phonemkontrast		
alveolar lateral-approximant	bilabial nasal	
998.1	1033.8	

Tabelle 8.5: Mittelwerte der Dauer in Millisekunden der Stimuli für den Perzeptionstest, angeordnet entsprechend der Betonung und Quantität von /a/ in der Pänultima bzw. des vorherigen Konsonants.

ter des Betonungskontrasts im Original mit Lang- bzw. Kurzvokal und für die Wörter des Phonemkontrasts. Die Längung bzw. Kürzung des Vokals in den Stimuli des Betonungskontrasts hat die Gesamtdauer der Stimuli beeinflusst. Um zu überprüfen, ob die Dauer der unterschiedlichen Stimuliversionen signifikant voneinander abweichen, wurde ein paarweiser Vergleich durch den t-Test durchgeführt. In den aus dem Originalwort mit Langvokal stammenden Stimuli waren die Stimuli mit kurzer Pänultima signifikant kürzer als die mit langer betonter und unbetonter Pänultima (in beiden Fällen $p < 0.001$). In den aus dem Originalwort mit Kurzvokal stammenden Stimuli war der Dauerunterschied zwischen Stimuli mit unbetonter und betonter kurzer Pänultima und Stimuli mit langer Pänultima ebenfalls signifikant (in beiden Vergleichen $p < 0.001$).

Es stellte sich dann die Frage, ob die unterschiedliche Dauer der Stimuli die Ergebnisse des Perzeptionstests beeinflussen könnte, da die Testteilnehmer sie als zusätzlichen Hinweis verwenden könnten, um die Stimuli des Betonungs- und Quantitätskontrast voneinander zu unterscheiden. Zunächst wurde überlegt, ob deswegen die Dauer der Stimuli mittels Resynthese angeglichen werden

Stimuli im Original mit Langvokal			
	Erste Silbe	Zweite Silbe	Dritte Silbe
korrekt betont	322.8	323.5	352.8
auf 2. Silbe betont	276.0	365.1	358.6
mit Vokalkürzung	306.6	244.4	336.3
Stimuli im Original mit Kurzvokal			
	Erste Silbe	Zweite Silbe	Dritte Silbe
korrekt betont	251.8	269.6	323.2
auf 2. Silbe betont	218.1	272.4	353.3
auf 2. Silbe bet. mit Vokallängung	216.5	379.2	348.7

Tabelle 8.6: Mittelwerte der Silbendauer in Millisekunden für die unterschiedlichen Versionen der Stimuli des Betonungscontrasts: a) normal betonte, b) mit Betonungsverschiebung auf der Pänultima und c) mit Vokalkürzung bzw. mit Vokallängung und zusätzlicher Betonungsverschiebung.

sollte. Es wurde dann aber beschlossen, in diesem Test nur natürliche Stimuli zu verwenden. In künftigen Perzeptionstests könnte man die Versuchspersonen sowohl mit natürlichen als auch mit resynthetisierten Stimuli, in denen die Dauer normalisiert wurde, testen, um zu überprüfen, ob die Normalisierung der Dauer die Ergebnisse beeinflusst.

Abb. 8.15 und 8.16 stellen die Silbendauer der ersten bzw. der zweiten Silbe für die drei Versionen beider Wortgruppen dar.

Die Mittelwerte der Dauer der ersten, zweiten und dritten Silben werden in Tab. 8.6 gezeigt.

Abb. 8.17 und 8.18 stellen die Dauer des Vokals /a/ in der ersten bzw. in der zweiten Silbe für die drei Versionen beider Wortgruppen dar.

Die Mittelwerte der Dauer des Vokals /a/ in den ersten und zweiten Silben sind in Tab. 8.7 dargestellt.

Die Vokalquantitätsänderungen beeinflussten die Silben- und die Vokaldau-

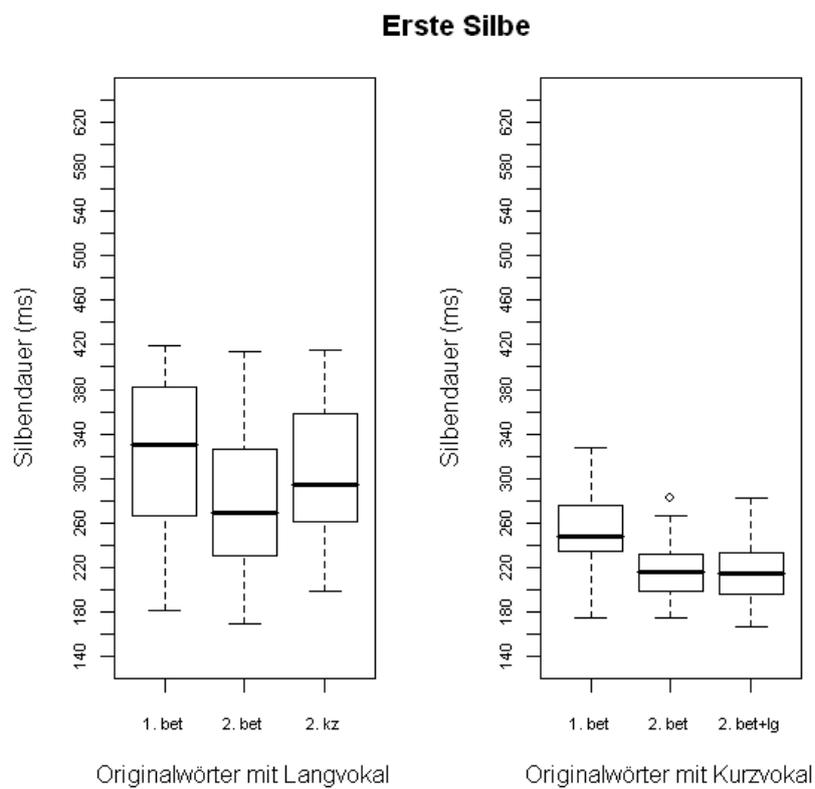


Abbildung 8.15: Dauer der ersten Silbe für die unterschiedlichen Versionen der Stimuli im Original mit Lang- bzw. Kurzvokal. 1. bet = auf der ersten Silbe betont, 2. bet = auf der zweiten Silbe betont, 2. kz = mit Kurzvokal in der zweiten Silbe, 2. bet+lg = auf der zweiten Silbe betont und mit Langvokal.

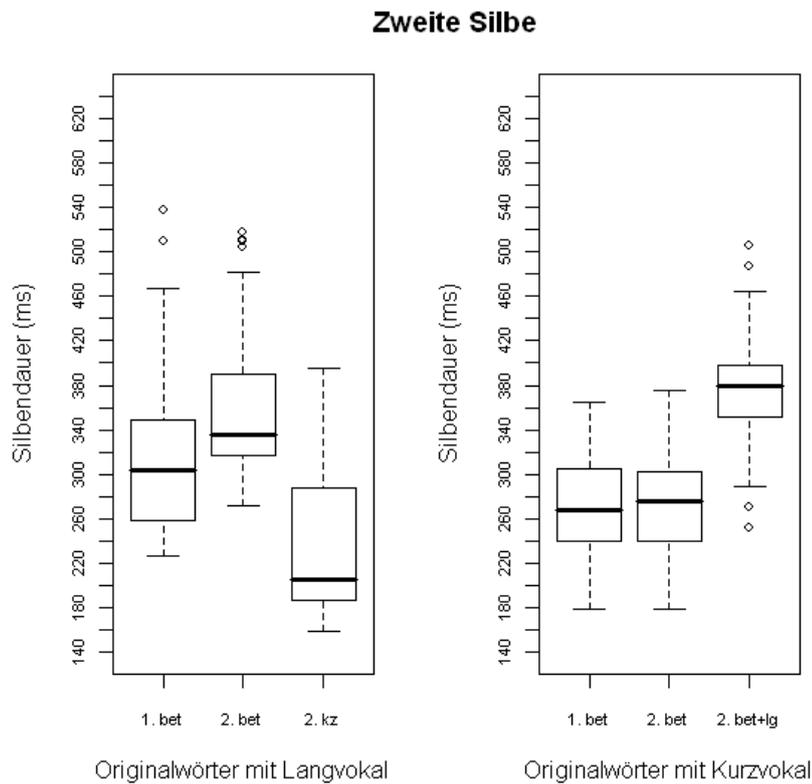


Abbildung 8.16: Dauer der zweiten Silbe für die unterschiedlichen Versionen der Stimuli im Original mit Lang- bzw. Kurzvokal (zur Legende s. Abb. 8.15).

er. Wir verglichen die gleich betonten Versionen mit Unterschieden in der Vokalquantität miteinander. Die vorletzten Silben in den mit Vokalkürzung ausgesprochenen Versionen sind durchschnittlich 79.1 ms kürzer als die entsprechenden Silben in der Originalversion mit langer Pänultima. Die vorletzten Silben der Versionen mit Vokallängung sind durchschnittlich 106.8 ms länger als die der Versionen mit Kurzvokal. Derselbe Vergleich für den Vokal /a/ in der zweiten Silbe zeigt, dass gekürzte Vokale durchschnittlich 81 ms kürzer und verlängerte Vokale 102.3 ms länger als ihre Gegenstücke sind. In allen

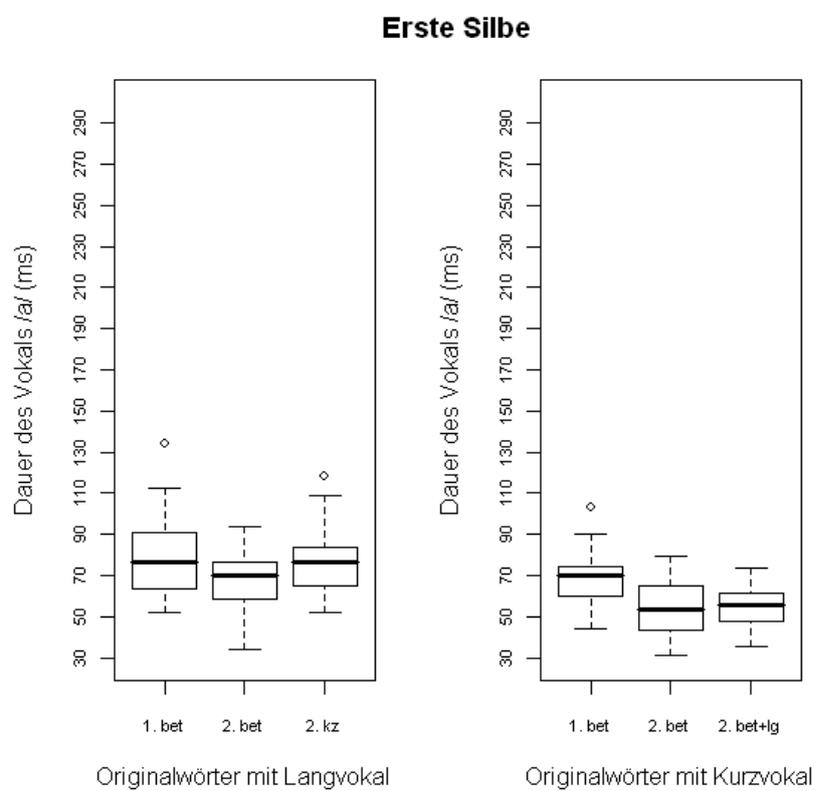


Abbildung 8.17: Dauer des Vokals /a/ in der ersten Silbe der unterschiedlichen Versionen der Stimuli im Original mit Lang- bzw. Kurzvokal. 1. bet = auf der ersten Silbe betont, 2. bet = auf der zweite Silbe betont, 2. kz = mit Kurzvokal in der zweiten Silbe, 2. bet+lg = auf der 2. Silbe betont und mit Langvokal.

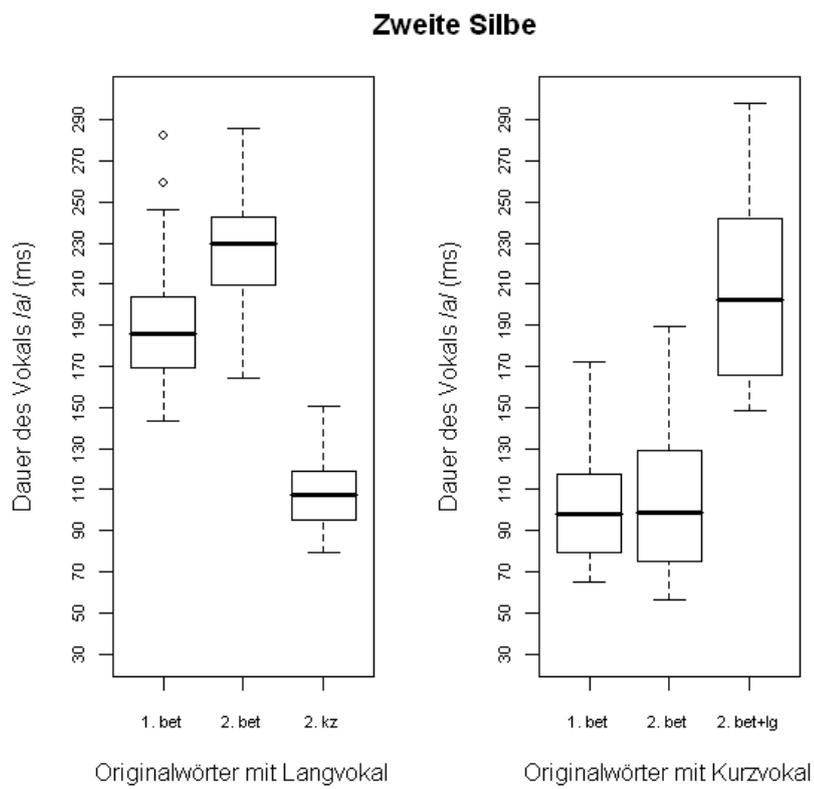


Abbildung 8.18: Dauer des Vokals /a/ in der zweiten Silbe der unterschiedlichen Versionen der Stimuli im Original mit Lang- bzw. Kurzvokal (zur Legende s. Abb. 8.15).

Stimuli im Original mit Langvokal		
	Erste Silbe	Zweite Silbe
korrekt betont	79.9	190.7
auf 2. Silbe bet.	69.6	228.2
mit Vokalkürzung	76.7	109.7
Stimuli im Original mit Kurzvokal		
	Erste Silbe	Zweite Silbe
korrekt betont	69.6	100.6
auf 2. Silbe bet.	54.2	103.6
auf 2. Silbe bet. mit Vokallängung	55.5	205.9

Tabelle 8.7: Mittelwerte der Dauer in Millisekunden des Vokals /a/ für die unterschiedlichen Versionen der Stimuli des Betonungskontrasts: a) normal betonte, b) mit Betonungsverschiebung auf der Pänultima und c) mit Vokalkürzung bzw. mit Vokallängung und zusätzlicher Betonungsverschiebung.

Vergleichen sind die Unterschiede hochsignifikant ($p < 0.001$, t-Test für gepaarte Stichproben).

Hingegen beeinflusste die Betonung nicht immer die Silben- und Vokaldauer. Wir verglichen die ersten und die zweiten Silben und Vokale der Wörter des Betonungskontrasts. Die ersten Silben in den Wörtern mit langer Pänultima sind durchschnittlich 46.8 ms kürzer, wenn sie unbetont sind. Die ersten Silben in den Wörtern mit kurzer Pänultima sind durchschnittlich 33.7 ms kürzer, wenn sie unbetont sind. Der Unterschied ist in beiden Fällen signifikant ($p = 0.002$ bzw. $p = 0.0419$). Die Dauer der zweiten Silben in den Wörtern mit langer Pänultima unterscheidet sich um 41.6 ms und in den Wörtern mit kurzer Pänultima nur um 2.8 ms. Diese Unterschiede sind nicht signifikant ($p = 0.088$ bzw. $p = 0.874$). Führt man denselben Vergleich mit den Vokalen durch, zeigt sich, dass die unbetonten Vokale in der ersten Silbe bei Wörtern mit Langvokal in der Pänultima durchschnittlich 10.3 ms kürzer, und bei Wörtern mit Kurzvokal in der Pänultima 15.4 ms kürzer sind als die betonten Vokale. Der

Unterschied ist fast signifikant im ersten Fall ($p = 0.052$) und hochsignifikant im zweiten Fall ($p < 0.001$). Die unbetonten Langvokale in der Pänultima sind durchschnittlich 37.5 ms kürzer als ihre betonten Gegenstücke. Die unbetonten Kurzvokale in der Pänultima sind durchschnittlich nur 3 ms kürzer. Im ersten Fall ist der Unterschied hochsignifikant ($p < 0.001$), während er im zweiten Fall nicht signifikant ist.

Da die Dauer der betonten und unbetonten Silben nicht immer signifikant unterschiedlich ist, könnten italienische Muttersprachler die Betonung der Stimuli falsch interpretieren, denn sie verlassen sich möglicherweise auf die Dauerunterschiede, um die Wortbetonung zu bestimmen.

8.5.1.3 Testteilnehmer

Die Versuchspersonen waren 23 italienische Muttersprachler im Alter zwischen 20 und 51 Jahren. 21 waren aus Sardinien und zwei aus der Provinz Mailand. Wir warben italienische Sprecher, die kein Deutsch sprachen, um zu vermeiden, dass unterschiedliche lexikalische Kenntnisse die Ergebnisse beeinflussten. Nur ein Teilnehmer berichtete über sehr geringe Deutschkenntnisse, erklärte aber zugleich, er habe die Wörter im Test noch nie vorher gehört.

Die Teilnehmer der Kontrollgruppe waren sechs deutsche Muttersprachler im Alter zwischen 14 und 49 Jahren. Fünf waren aus Wuppertal und einer aus München. Die Hälfte berichtete über keine und die anderen Hälfte über geringe Italienischkenntnisse. Eine auch noch so kleine Kontrollgruppe war notwendig, um zu überprüfen, ob die Aufgabe für Deutsche lösbar war.

Keine der deutschen und italienischen Versuchspersonen besaß eine phonetische Ausbildung.

8.5.1.4 Perzeptionstest

Der Perzeptionstest wurde in Java implementiert. Die Probanden wurden zunächst trainiert, zwei Wörter eines Wortpaares jeweils mit den Tasten *1* und *2* zu assoziieren. Im Folgenden wurden sie instruiert, sich Sequenzen aus zwei bzw. vier Wörtern einzuprägen und sie durch Drücken der entsprechenden Tasten in der korrekten Reihenfolge wiederzugeben. Jeder Teilnehmer wurde mit drei Wortpaaren getestet: einem Phonem-, einem Betonungs-Quantität- und einem Betonungs-Quantitätskontrast.

In dem Phonemkontrast wurde **ab'lagern* der Taste *1* und **ab'magern* der Taste *2* zugeordnet. In dem Betonungs- und Betonungs-Quantitätskontrast war das auf der ersten Silbe betonte Wort der Taste *1* und das auf der Pänultima betonte Wort der Taste *2* zugeordnet.

Alle Teilnehmer wurden zunächst mit dem Phonemkontrast und danach mit einem Betonungs- und einem Betonungs-Quantitätskontrast getestet. Die Reihenfolge dieser beiden letzten Kontraste war für jeden Teilnehmer zufällig. Die Wortpaare wurden für jeden Teilnehmer nach folgenden Kriterien gewählt: a) Die Kontraste sollten sich von unterschiedlichen Wörtern ableiten, damit der Teilnehmer nicht dasselbe Wort in beiden Kontrasten hörte, und b) die Wörter beider Kontraste sollten zur selben Wortgruppe gehören (Kurz- oder Langvokal in der Pänultima), damit die intra-subjektiven Ergebnisse in den beiden Kontrasten miteinander vergleichbar blieben.

Der Ablauf entsprach dem bei Dupoux et al. 2001 [38] und war derselbe für jeden Kontrast (s. Anhang E.3.1 für die Probandenanweisungen). Die Teilnehmer lernten in einer Trainingsphase, die zwei Wörter im Kontrast den Tasten *1* und *2* zuzuordnen. Zunächst hörten sie alle sechs Tokens eines jeden Wortes, die in zufälliger Reihenfolge präsentiert wurden. Danach konnten sie auf die Tasten drücken, sooft wie sie wollten, um jeweils einen der Tokens zu hören. Wenn sie sich sicher fühlten, machten sie einen Test, um zu überprüfen,

ob sie die Wörter den Tasten zuordnen konnten. Sie hörten einzelne Wörter und sollten die entsprechenden Tasten drücken. Wenn die Antwort richtig war, erschien 800 ms lang in grünen Buchstaben „OK!“ auf den Bildschirm, wenn nicht, erschien in roter Schrift „ERRORE!“, und der Stimulus wurde nochmals abgespielt, bis der Proband die richtige Antwort gab (s. die Benutzeroberfläche in Abb. 8.19). Um den Test zu bestehen, mussten die Probanden fünf korrekte Antworten in Folge geben.

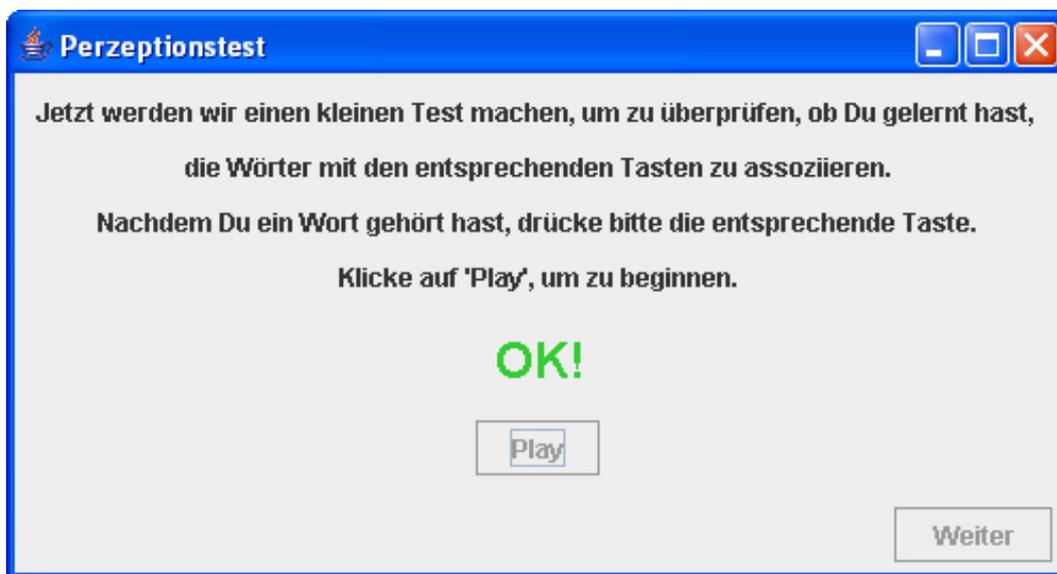


Abbildung 8.19: Benutzeroberfläche des Tests, der überprüft, ob der Proband die Wort-Taste Zuordnung gelernt hat.

In der folgenden Testphase sollten die Probanden acht Wortsequenzen aus zwei und acht aus vier Elementen hören und anschließend die Computertasten in der richtigen Reihenfolge der Wortsequenz entsprechend drücken. Zum Beispiel sollten sie für 'Anlage', 'Anlage', *An'lage, 'Anlage die Tasten 1121 drücken. Während dieser Phase bekamen die Probanden kein Feedback zu ihren Antworten. Abb. 8.20 stellt die Testoberfläche für den Test mit den Serien aus

zwei Wörtern dar.

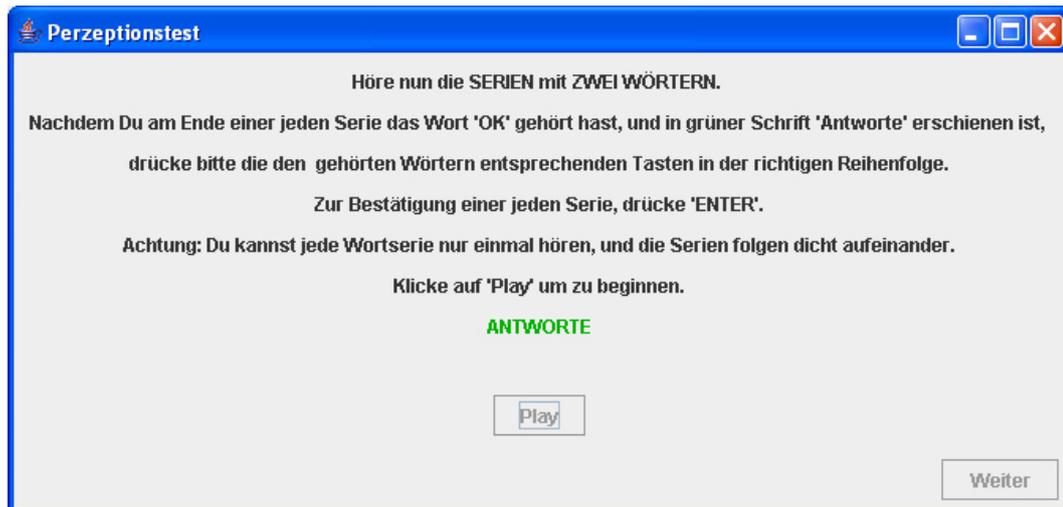


Abbildung 8.20: Benutzeroberfläche des Tests mit den Serien aus zwei Wörtern.

Anders als Dupoux et al. 2001 [38] präsentierten wir in unserem Test keine Serien aus sechs Wörtern. Serien aus sechs der dreisilbigen Wörter, die im Durchschnitt jeweils 925.8 ms lang waren, wären wahrscheinlich zu lang gewesen. In einem Pilottest mit einem italienischen und einem deutschen Probanden berichteten die Teilnehmer übereinstimmend, dass es für sie unmöglich war, sich die Serien aus sechs Wörtern jedweden Kontrasts zu merken.

Die vier möglichen Serien aus zwei Wörtern wurden zweimal im Test präsentiert. Von den 16 möglichen Serien aus vier Wörtern wurden die selben wie bei Dupoux et al. 2001 [38] ausgewählt. Es handelt sich um die Serien, die mehr Übergänge haben. Zum Beispiel hat die Serie 1121 zwei Übergänge: von 1 auf 2 und von 2 auf 1. Regelmäßige Muster wie 1212 wurden nicht genommen, da sie einfacher sind (s. Dupoux et al. 2001 [38, S.1608]). Die gewählten Serien waren daher folgende:

- aus zwei Wörtern: 11, 12, 21, 22, 11, 12, 21, 22

- aus vier Wörtern: 1121, 1122, 1211, 1221, 2111, 2112, 2122, 2212

Die Serien wurden in zufälliger Reihenfolge präsentiert. Die Tokens wurden für jede Wortserie zufällig gewählt, so dass die Hälfte vom weiblichen und die andere Hälfte vom männlichen Sprecher waren, und jedes Token nur ein Mal in der selben Serie vorkommen konnte.

Die Wörter jeder Serie wurden durch eine Pause von lediglich 20 ms getrennt, damit die Teilnehmer weniger Möglichkeiten hatten, beim Hören die Wörter direkt in Zahlen zu übersetzen. Eine Pause von 1500 ms trennte jede Serie von der nächsten. Am Ende jeder Serie hörten die Teilnehmer das Wort „okay“, ausgesprochen von einem englischen Muttersprachler, um echoisches Gedächtnis zu vermeiden (s. Dupoux et al. 2001 [38], Morton, Crowder & Prussin 1971 [100] und Morton, Marcus und Ottley 1981 [101]). Die Teilnehmer konnten ihre Antwort nicht eingeben, bis sie dieses Wort gehört hatten.

Wenn sie eine Serie getippt hatten, deren Elementenanzahl der abgespielten Serie nicht entsprach, erschien eine Warnung auf den Bildschirm, und sie wurden gebeten, ihre Antwort nochmals einzutippen.

Der Test war derselbe für beide Teilnehmergruppen. Nur die Anweisungen, das Feedback und alle Texte auf der Benutzeroberfläche waren auf italienisch für die italienischen und auf deutsch für die deutschen Probanden.

Der Test dauerte zirka 15 bis 20 Minuten.

Wie Dupoux et al. [38] bewerteten wir nur die 100% korrekt wiedergegebenen Serien als korrekt und bezeichneten 100% falsche Antworten als Umkehrungen (*reversals*). Dupoux et al. [38] lehnten die Teilnehmer ab, die mehr Umkehrungen als korrekte Antworten in dem Phonem- oder in dem Betonungskontrast hatten, da sie möglicherweise die Zuordnung zwischen Wörtern und Tasten verwechselt hatten.

Nach unseren Kriterien sollte die Ablehnung eines Testteilnehmer unter fol-

genden Bedingungen erfolgen: a) Ein Teilnehmer sollte mehr Umkehrungen als korrekte Antworten in der Serie aus zwei oder aus vier Wörtern in mindestens einem der Kontraste erzielen, und b) die Anzahl der Umkehrungen sollte über dem kritischen Wert des Binomialgesetzes mit dem größten *alpha*-Niveau kleiner als 0.05, und daher mindestens über 5 (*alpha*-Niveau = 0.0273) in den Serien aus zwei und mindestens über 3 (*alpha*-Niveau = 0.0108) in den Serien aus vier Wörtern liegen.

Diese Schwellen wurden mittels der Quantilfunktion berechnet, in dem die Wahrscheinlichkeit, Umkehrungen per Zufall zu erzeugen (1/4 für die Serien aus zwei Wörtern und 1/16 für die Serien aus vier) und die Anzahl der Versuche (8), berücksichtigt wurden. Nach diesen Kriterien wurden vier italienische Teilnehmer von 27 abgelehnt. Kein deutscher Teilnehmer wurde abgelehnt.

8.5.1.5 Ergebnisse und Diskussion

Nachdem sie den Test beendet hatten, fragten wir die italienischen Teilnehmer, welcher Kontrast für sie am schwierigsten zu unterscheiden war, und welchen Unterschied es zwischen den Wörtern in jedem Paar gab. Zwölf der 18 Teilnehmer, die eine Antwort zur Frage geben konnten, werteten den Betonungs-Quantitätskontrast als reinen Betonungsunterschied. Dies stützt unsere Hypothese, dass der Betonungs-Quantitätskontrast von den italienischen Probanden für einen Betonungskontrast gehalten wird. Die Wortpaare mit dem eigentlichen Betonungskontrast waren für sie am schwierigsten zu erkennen. Mehrere italienische Probanden berichteten, sie hörten fast keinen Unterschied zwischen den Wörtern. Jedoch konnten alle italienischen Teilnehmer, wenn auch erst nach längeren Versuchen, die fünf richtigen Antworten hintereinander geben, um die Trainingsphase zu bestehen. Dies bedeutet, dass sie in der Lage waren, die Stimuli akustisch zu unterscheiden.

Die Fehlerquote der italienischen Teilnehmer liegt bei den Wortserien

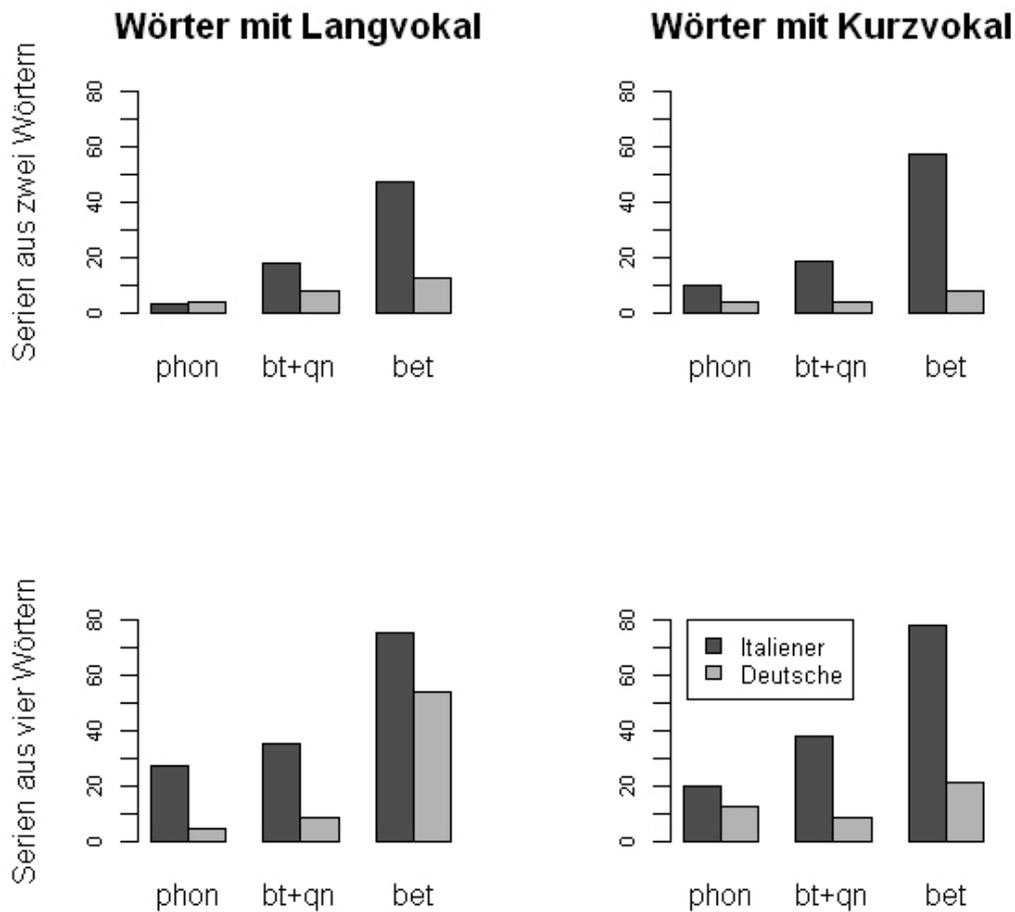


Abbildung 8.21: Fehlerquote mit dem Phonem-, Betonungs-Quantitäts- und Betonungskontrast in den Serien aus zwei und vier Wörtern für 23 Italiener und sechs Deutsche. Drei deutsche und elf italienische Probanden wurden mit Kontrasten aus Wörtern im Original mit Langvokal in der Pänultima getestet. Die restlichen drei deutschen und zwölf italienischen Probanden wurden mit Kontrasten aus Wörtern im Original mit Kurzvokal getestet.

Kontrast aus	Wörtern mit langem /a/			Wörtern mit kurzem /a/		
	2	4	Mittelwert	2	4	Mittelwert
Italiener						
Phonem	3.4	27.3	15.3	10.4	19.8	15.1
Betonungs-Quantität	18.2	35.2	26.7	18.8	37.5	28.1
Betonung	47.7	75.0	61.4	57.3	78.1	67.7
Deutsche						
Phonem	4.2	4.2	4.2	4.2	12.5	8.3
Betonungs-Quantität	8.3	8.3	8.3	4.2	8.3	6.2
Betonung	12.5	54.2	33.3	8.3	20.8	14.6

Tabelle 8.8: Prozentfehler für die italienischen und deutschen Testteilnehmer wie in Abb. 8.5.1.5.

mit Betonungskontrast höher als bei denen mit Phonem- und Betonungs-Quantitätskontrast (s. Abb. 8.5.1.5 und Tab. 8.8). Beim Phonemkontrast machten die Italiener weniger Fehler als beim Betonungs-Quantitätskontrast. Dieses Verhalten ist in beiden Gruppen vergleichbar: die mit Originalwörtern mit Langvokal in der Pänultima getestete Gruppe und die mit Originalwörtern mit Kurzvokal.

Die deutschen Teilnehmer verhalten sich hingegen unterschiedlich in beiden Gruppen. Die Gruppe, die mit Kontrasten aus Originalwörtern mit Kurzvokal getestet wurde, entspricht den Erwartungen und zeigt ähnliche Fehlerquoten bei allen drei Kontrasten. Die Gruppe, die mit dem Betonungskontrast aus Wörtern mit Langvokal getestet wurde, hatte mehr Schwierigkeiten, sich die Serie aus vier Wörtern zu merken, als die Gruppe mit dem Betonungskontrast mit Kurzvokal in der Pänultima. Da jede Gruppe nur aus drei deutschen Teilnehmern bestand, könnte dieses Ergebnis zufällig sein. Wir planen, weitere deutsche Teilnehmer zu testen, um zuverlässigere Ergebnisse zu bekommen. Jedoch hatten die Teilnehmer in der Langvokalgruppe keine Schwierigkeiten mit dem Phonem- und mit dem Betonungs-Quantitätskontrast, und dies zeigt,

dass sie die Aufgabe verstanden hatten, und dass sie in der Lage waren, die Serien ins Gedächtnis einzuprägen.

Wegen der wiederholten Messungen wurde die Fehleranzahl mit einem verallgemeinerten gemischten linearen Modell mit Zufallseffekten mit multivariater Normalverteilung, *Penalized Quasi-Likelihood*, Poisson-Fehlerverteilung und logarithmischer Linkfunktion modelliert.

Es gibt keinen signifikanten Unterschied zwischen der Leistung der italienischen und der deutschen Probanden beim Baseline-Phonemkontrast sowohl für die Serien aus zwei als auch aus vier Wörtern.

Wie erwartet, war die Anzahl der Fehler der italienischen Versuchspersonen beim Betonungskontrast höher als beim Phonem- und beim Betonungs-Quantitätskontrast (s. Tab. 8.9).

Kontrast aus	Wörtern mit langem /a/		Wörtern mit kurzem /a/	
Serienlänge	2	4	2	4
Italiener				
Phon. vs Bet.-Quant.	0.0307	n.s.	n.s.	0.0187
Bet. vs Bet.-Quant.	0.0095	0.0011	0.0086	0.0005
Betonung vs Phonem	0.0010	0.0002	0.0021	0.0000
Deutsche				
Phon. vs Bet.-Quant.	n.s.	n.s.	n.s.	n.s.
Bet. vs Bet.-Quant.	n.s.	0.0273	n.s.	n.s.
Betonung vs Phonem	n.s.	0.0270	n.s.	n.s.

Tabelle 8.9: P-Werte nach den Vergleichen der Fehleranzahl in den Phonem-, Betonungs-Quantitäts- und Betonungskontrasten in den Serien aus zwei und vier Wörtern für Italiener und Deutsche.

Wir erwarteten, dass die deutschen Teilnehmer eine ähnlich gute Leistung im Test mit den drei Kontrasten liefern würden. Jedoch machten sie in den Serien aus vier Wörtern mit Betonungskontrast mit langem /a/ in der Pänultima signifikant mehr Fehler als beim Phonem- und beim entsprechendem

Betonungs-Quantitätskontrast, und sie waren nicht besser als die italienischen Teilnehmer ($p = 0.0996$).

Nichtdestoweniger, wenn wir beide Wortgruppen und Serien betrachten, machten die Deutsche beim Betonungskontrast signifikant weniger Fehler als die Italiener ($p < 0.001$), dagegen war der Fehlerunterschied fast signifikant beim Betonungs-Quantitätskontrast ($p = 0.0583$).

Die Leistung von zwölf italienischen Teilnehmern in der Serie aus zwei und von 17 italienischen Teilnehmern in der Serie aus vier Wörtern mit Betonungskontrast war nicht signifikant besser als die Zufallsleistung ($p\text{-value} > 0.05$, Binomialtest). Jedoch war die Leistung der gesamten Stichprobe mit den Serien aus zwei und aus vier Wörtern des Betonungskontrasts in beiden Fällen besser als die Zufallsleistung ($p < 0.001$).

Da Betonungs- und Betonungs-Quantitätskontrast in zufälliger Reihenfolge im Test präsentiert wurden, überprüften wir, ob die Präsentationsreihenfolge, die Ergebnisse beeinflusst hatte. Wir fanden keinen Effekt. Es spielte also keine Rolle, ob die Teilnehmer zunächst die Aufgabe mit einem oder mit dem anderen Kontrast lösten.

8.5.1.6 Betrachtungen nach dem Experiment 5

Italienische Deutschlernende haben Schwierigkeiten, die Wortbetonung morphologisch komplexer Wörter im Deutschen zu lernen, da ihre sprachspezifische phonologische Darstellung der Betonung solche unterschiedlich betonten Wörter dem selben Betonungsmuster zuordnet. Wie erwartet, waren italienische Versuchspersonen im verwendeten Kurzzeitgedächtnistest mit Wortserien schlechter beim Betonungskontrast als beim Phonem- oder Betonungs-Quantitätskontrast. Der letztere Kontrast wurde von den meisten Italienern als Betonungskontrast interpretiert, da die Betonung von einer Änderung der

Vokaldauer „unterstützt“ wurde, und war deswegen einem Betonungskontrast im Italienischen ähnlicher. Dauermessungen der Stimuli zeigten, dass in einigen Fällen die Dauer der betonten Vokale oder Silben und deren unbetonte Gegenstücke sich nicht unterschieden. Möglicherweise waren deswegen die Wörter im Betonungskontrast für die italienischen Probanden schwierig zu unterscheiden.

Die deutschen Versuchspersonen zeigten eine ähnlich gute Leistung in den drei Kontrasten mit Ausnahme des Betonungskontrasts aus Wörtern mit langer Pänultima. Jedoch ließe sich dies mit der geringen Anzahl deutscher Teilnehmer begründen. Wir haben vor, eine größere Anzahl deutscher und italienischer Probanden zu testen, um zuverlässigere Ergebnisse zu bekommen. Überdies haben wir die Absicht, ein ähnliches Experiment mit Nichtwörtern durchzuführen, die für beide Gruppen — deutsche und italienische Versuchspersonen — unbekannt sind, um beide Stichproben auf der gleichen Basis zu testen. Aus diesem Grund wurden bereits Aufnahmen durchgeführt, über die in Kap. 8.6 berichtet wird.

8.6 Voruntersuchungen zu einem Test zur Betonungswahrnehmung von Nichtwörtern durch Deutsche und Italiener

Die Ergebnisse des Experiments 5 (s. Kap. 8.5) deuten auf gewisse Schwierigkeiten der Italiener bei der Wahrnehmung der Betonung morphologisch komplexer Wörter im Deutschen hin. Die intra-subjektiven Ergebnisse belegen eine bessere Leistung der italienischen Probanden bei dem Phonem- und Betonungs-Quantitäts- als bei dem Betonungskontrast, obwohl alle Kontraste aus unbekanntem Wörtern mit für die italienischen Probanden zum Teil ungewöhnlicher phonotaktischer Struktur bestanden. Trotzdem unterschied sich ih-

re Leistung bei den Phonem- und Betonungs-Quantitätskontrasten nicht maßgeblich von der Leistung der deutschen Probanden, wobei sie — betrachtet man die Kontraste aus beiden Wortgruppen im Original mit Lang- und Kurzvokal gemeinsam — beim Betonungskontrast schlechter als die Deutschen abschnitten. Allein wegen dieses Unterschiedes zwischen der deutschen und der italienischen Gruppe ist das Ergebnis des Experiments 5 bereits aussagekräftig. Jedoch ist es notwendig, beide Gruppen auf gleicher Basis zu testen, denn im Test mit den deutschen Wörtern hatten die Deutschen offensichtlich einen Vorteil gegenüber den italienischen Probanden. Aus diesem Grund wird ein Experiment mit sowohl für Italiener als auch für Deutsche unbekanntem Nichtwörtern geplant. Für das Experiment wurden bereits Aufnahmen durchgeführt. Jedoch wurden die aufgenommenen Daten für den geplanten Perzeptionstest aus zwei Gründen noch nicht verwendet. Erstens fanden sich bisher keine männlichen, sondern nur weibliche Sprecher für die Aufnahme (zu den Eigenschaften der gesuchten Sprecher s. unten 8.6.1). Zweitens haben wir zuerst den Perzeptionstest mit deutschen Wörtern durchführen wollen, um zu überprüfen, ob der Effekt überhaupt auftritt.

8.6.1 Methode

Die Schwierigkeit für ein solches Experiment liegt vor allem in der Wahl der Stimuli, aber auch in der Wahl der Sprecher.

Die Stimuli sollten aus Minimalpaaren bestehen, die dem Betonungsmuster morphologisch komplexer Wörter im Deutschen entsprechen, hinsichtlich ihrer weiteren Eigenschaften jedoch gleichwertig für italienische und deutsche Hörer sind. Solche Wörter sollten von Sprechern gelesen werden, die weder des Deutschen noch des Italienischen mächtig sind, um keiner Gruppe einen Vorteil zu verschaffen, jedoch sollten sie einem deutschen Betonungsmuster entsprechen.

Um die Nichtwörter auszusprechen, wurden zunächst ein weiblicher und ein männlicher phonetisch ausgebildeter Sprecher gesucht. Die Grundidee war, die Betonung nicht als solche durch den Sprecher ausdrücken zu lassen, sondern Nichtwörter zu bilden, die sich hinsichtlich der Vokalquantität bzw. der Anwesenheit eines „Tonhöheakzent“ in den Silben voneinander unterscheiden. Nach unserer Hypothese sollten die italienischen Probanden im Test mehr Schwierigkeiten als die Deutschen haben, sich die Stimuli, die sich nur in einem Tonhöheakzent unterscheiden, ins Gedächtnis einzuprägen.

Es bot sich an, für die Aufnahme der Stimuli japanische Sprecher heranzuziehen, weil im Japanischen zwischen einmorigen und zweimorigen Silben — Silben mit Kurz- und mit Langvokal — unterschieden wird, und der Pitch Accent phonologische Bedeutung hat (s. Okada 1999 [109]).

Als Sprecherin boten sich schließlich zwei Phonetikerinnen an: zunächst eine japanische und im Folgenden auch eine kantonesische Muttersprachlerin. Es handelte sich um eine Professorin und eine Doktorandin des Faches Phonetik, die sich als Gäste im Institut für Phonetik und Sprachverarbeitung der Universität München aufhielten. Es war notwendig, das Stimulusmaterial für die Sprecherinnen maßzuschneidern, denn die Stimuli sollten keine ungewöhnlichen phonotaktischen Strukturen und Phoneme weder für italienische und deutsche noch für japanische bzw. kantonesische Muttersprachler aufweisen.

8.6.1.1 Aufnahme einer japanischen Sprecherin

Die japanische Muttersprachlerin war eine 55 Jahre alte Phonetikerin, Sprecherin der Kinki-Varietät des Japanischen aus Osaka, die über keine Italienisch- und nur geringfügige Deutschkenntnisse verfügte.

Als Stimuli wurden dreisilbige Nichtwörter konstruiert. Sie bestanden aus Phonemen, die sowohl im deutschen und im italienischen als auch im japani-

schen Phoneminventar existieren (s. Okada 1999 [109]), und deren phonotaktische Struktur in allen drei Sprachen möglich ist. Wir begrenzten das Phoneminventar auf die Vokale /a/, /i/ und /u/. Die Vokale /e/, /ɛ/, /o/ und /ɔ/ wurden ausgeschlossen, obwohl sie zwar sowohl zum Deutschen als auch zum Standarditalienischen gehören, jedoch existiert, wie bereits in Kap. 3.1 erwähnt, ihre Opposition in etwa der Hälfte der Regionalvarietäten des Italienischen nicht, zudem gilt die Opposition nur für die Vokale in betonter Silbe. Da im Japanischen /u/ ungerundet ist und ähnlich wie /ʊ/ ausgesprochen wird, wurde überlegt, ob man den Vokal auslassen soll. Um zu gewährleisten, dass die Silben jedes Wortes einen unterschiedlichen Vokal vorweisen konnten, wurde schließlich auch der Vokal /u/ ins Phoneminventar aufgenommen. Das /u/ der japanischen Sprecherin war in der Tat weniger gerundet als das herkömmliche deutsche oder italienische /u/, erwies sich aber als akzeptabel. Die Konsonanten waren /p/, /t/, /k/, /b/, /d/ und /g/, die sowohl zum Deutschen als auch zum Italienischen und Japanischen gehören. Aus der Kombination dieser Konsonanten und Vokale wurden acht Nichtwörter mit der Silbenstruktur CVCVCV konstruiert. Die Nichtwörter sind im Anhang F.1 aufgelistet.

Die Sprecherin sollte jedes Nichtwort acht mal mit drei unterschiedlichen Aussprachen vorlesen:

- a) mit zweimorigem ersten Vokal,
- b) mit zweimorigem zweiten Vokal und Pitch-Accent auf der ersten Silbe und
- c) mit zweimorigem zweiten Vokal.

Die Nichtwörter vom Typ a und c sollten einem italienischen Betonungs-kontrast, in dem die Silbe mit dem Langvokal die betonte Silbe darstellt, und die vom Typ b und c einem deutschen Betonungskontrast, in dem die unbetonte Silbe lang ist, ähneln.

Nach unserer Hypothese sollten die deutschsprachigen Versuchspersonen den Test mit den Kontrasten a vs. c und b vs. c in etwa gleich gut bewältigen können, während die italienischsprachigen Versuchspersonen mehr Schwierigkeiten mit dem Kontrast a vs. c haben sollten.

Das Lesematerial bestand aus Nichtwörtern, die in dem Trägersatz „I say: xy“ eingebettet waren. Der Trägersatz diente zur Vorbeugung des Kontrastakzents in den Stimuli und sollte zudem dafür sorgen, dass die Stimuli nicht zu schnell ausgesprochen wurden.

Leider erwiesen sich die aufgenommenen Daten für den geplanten Perzeptionstest als ungeeignet, weil die Sprecherin die zweimorigen Vokale teilweise extrem lang ausgesprochen hatte, so dass diese nicht einem Langvokal des Deutschen oder einem betonten Vokal des Italienischen entsprachen.

8.6.1.2 Aufnahme einer kantonesischen Sprecherin

Es wurde dann eine zweite Sprecherin aufgenommen. Sie war eine 32-jährige Sprecherin des Kantonesischen (Hong Kong). Auch sie verfügte über keine Italienisch- und nur geringfügige Deutschkenntnisse. Das Wortmaterial für die kantonesische Sprecherin wurde neu zusammengestellt. Die Nichtwörter enthielten die Konsonanten /p/, /t/, /k/, /m/, /n/ und /s/ und die Vokale /a/, /e/, /i/, /o/, die zum Phoneminventar des Deutschen, des Italienischen und des Kantonesischen gehören (s. Zee 1999 [155]). Die Nichtwörter waren diesmal mit einem Pseudopräfix gebildet (/as-/ oder /am-/) und hatten deswegen die Silbenstruktur VCCVCV. Die Präfixe waren keine existierenden Präfixe des Deutschen. Die daraus resultierenden Konsonantencluster waren sowohl in der deutschen als auch in der italienischen Phonotaktik möglich, und die kantonesische Sprecherin konnte sie mühelos aussprechen. Die Aussprache der Nichtwörter entsprach den Typen a, b und c der japanischen Sprecherin, mit dem Unterschied, dass die Vokale statt zweimorig als lang oder kurz definiert

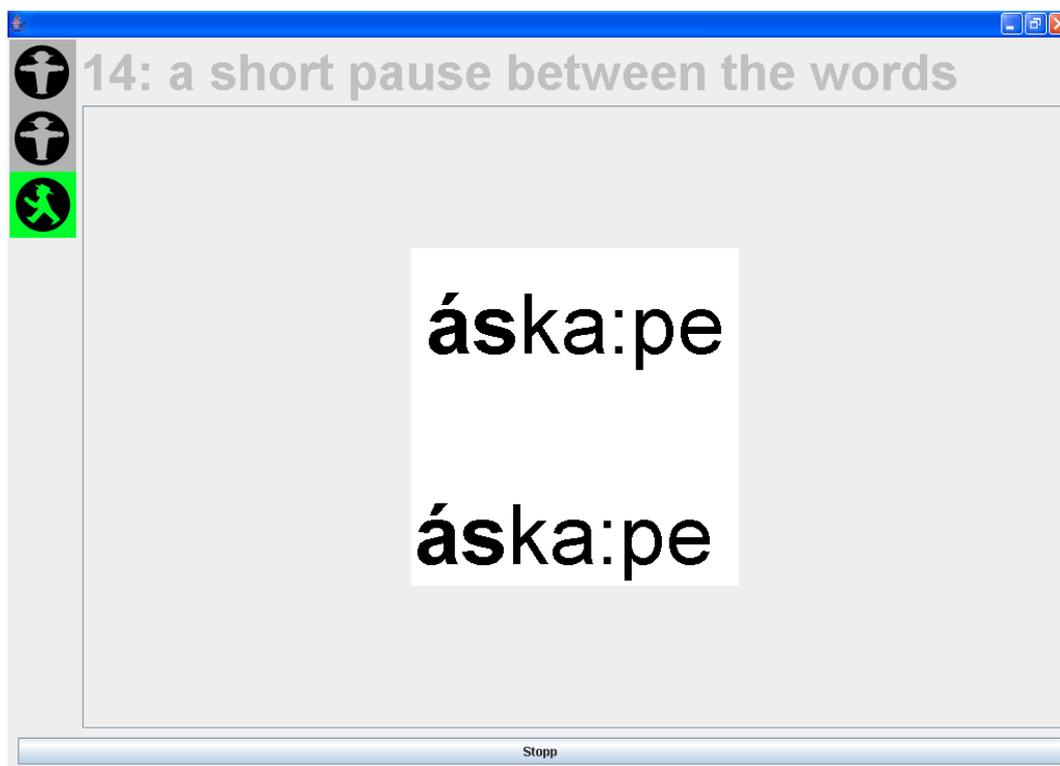


Abbildung 8.22: Beispiel eines Textprompts für die kantonesische Sprecherin. Für die Aufnahme wurde das Programm SpeechRecorder verwendet (s. Draxler & Jänsch [35]).

wurden. Zudem enthielten die erste und die zweite Silbe immer den offenen Zentralvokal /a/, damit die Vokalqualitätsänderung keinen Hinweis zur Unterscheidung der Stimuli liefert. Die Nichtwörter wurden isoliert ausgesprochen. Jedes Wort wurde zweimal hintereinander gelesen, um einen Kontrastakzent zu vermeiden (s. die in Abb. 8.22 dargestellte Textprompt).

8.6.2 Planung zur weiteren Durchführung des Experiments 6

Da die Ergebnisse des Experiments 5 (s. Kap. 8.5.1.5) ein unterschiedliches Verhalten der deutschen Sprecher in den Gruppen mit Lang- bzw. mit Kurzvokal gezeigt hatten, wäre es sinnvoll, diese Gruppendifferenzierung auch bei den Nichtwörtern beizubehalten. Auf dieser Weise könnte überprüft werden, ob die Nichtwörter mit Langvokal den deutschen Teilnehmern ebenfalls Schwierigkeiten bereiten. Aus diesem Grund sollte ein neues Korpus aufgenommen werden. Wie die Stimuli für die kantoneseische Sprecherin, werden solche Stimuli dreisilbig, mit /a/ sowohl in der ersten als auch in der zweiten Silbe und mit Pseudopräfix sein. Jedoch werden wir wie im Experiment 5 (s. Tab. 8.4 auf S. 145) Stimuli mit Lang- und mit Kurzvokal konstruieren. Die Stimuli, die denen des Betonungskontrasts im Experiment 5 entsprechen, werden sich nur hinsichtlich eines Pitch Accents unterscheiden. Die Stimuli, die denen des Betonungs-Quantitätskontrasts im Experiment 5 entsprechen, werden sich zusätzlich in der Vokaldauer unterscheiden. Um den Einfluss individueller Charakteristiken der Probanden auszuschließen, werden die Teilnehmer sowohl mit Stimuli mit Lang- als auch mit Kurzvokal in der Pänultima getestet. Ein Phonemminimalpaar werden wir ebenfalls als Baseline verwenden.

9 Schlussbetrachtungen

9.1 Zusammenfassung der Ergebnisse

Die durchgeführten Untersuchungen liefern eine Reihe bemerkenswerter Ergebnisse zum computergestützten Lernen der Wortbetonung im Deutschen und deren Wahrnehmung durch italienischsprachige Lernende.

Die Analyse des Korpus mit Äußerungen von italienischen Sprecherinnen förderte eine große Vielfalt an segmentalen Fehlern, wie Schwa-Epenthese, Ersetzungen von Frikativen und Vokalen zu Tague, die sicher zu einem Großteil auf phonologische Unterschiede zwischen dem Italienischen und dem Deutschen hinsichtlich des Phoneminventars und der Phonotaktik zurückzuführen sind. Zudem machten fast alle Sprecherinnen systematische Betonungsfehler bei morphologisch komplexen Wörtern im Deutschen: Sie betonten das Zweitglied statt des Erstglieds der Wörter. Bemerkenswert ist, dass sogar Sprecherinnen, die seit längerer Zeit in Deutschland leben und die ansonsten fortgeschrittene Sprecherinnen des Deutschen sind, solche Fehler systematisch machten und dies nie bemerkt hatten. Dies deutet darauf hin, dass die Wortbetonungsfehler perzeptive Ursachen haben. In der Tat können sekundär betonte Silben im Deutschen ihre Quantität bewahren (s. Kap. 2.3.2 auf S. 37). Dies und die Tatsache, dass im Italienischen die Dauer das Hauptkorrelat der Betonung ist, könnte italienischen Sprechern den Eindruck vermitteln, dass die sekundär betonten Silben die Hauptbetonung des Wortes tragen, insbesondere

wenn sie lang sind. Zudem kennt das Italienische keine Minimalpaare, die sich hinsichtlich der Position der Haupt- und Nebenbetonung unterscheiden.

In einer präliminären Untersuchung wurden Äußerungen deutschlernender Italienerinnen resynthetisiert, indem ihre lokale Sprechgeschwindigkeit nach dem Modell von Pfitzinger 2001 [116] an die der entsprechenden Äußerungen einer deutschen Muttersprachlerin angepasst wurde. Die Ergebnisse eines mit deutschen Muttersprachlern durchgeführten Perzeptionstests zum Vergleich der originalen und resynthetisierten Stimuli zeigten, dass die alleinige Korrektur der lokalen Sprechgeschwindigkeit keine signifikanten Verbesserungen des Sprechrhythmus bewirkte. Dies überrascht nicht, denn die Korrektur eines einzigen Parameters kann zu einer unnatürlichen Gesamtprosodie führen, was in unserem Test wahrscheinlich die Bewertungen der Muttersprachler beeinflusst hat.

Im folgenden Experiment wurden sowohl die Grundfrequenz als auch die Dauer der Phone in morphologisch komplexen deutschen Wörtern, die von italienischen Sprecherinnen falsch betont wurden, mittels Resynthese korrigiert. Ein von deutschen Muttersprachlern durchgeführter Perzeptionstest bewies, dass die Manipulation der Grundfrequenz eine effektivere Maßnahme zur Korrektur der Wortbetonung als die Manipulation der Dauer der Phone ist. In vielen Stimuli waren die Dauerverhältnisse der Silben zwischen den korrekt und unkorrekt betonten Wörtern nicht unterschiedlich. Erst durch die Übertragung der Grundfrequenz konnte die Betonung korrigiert werden. Dies deutet darauf hin, dass im Deutschen möglicherweise die Grundfrequenz als Korrelat der Betonung morphologisch komplexer Wörter eine größere Rolle als die Dauer spielt.

Durch zusätzliche Korrektur der Intensität konnten resynthetisierte Stimuli als Feedback in einem Aussprachetraining verwendet werden. Der Umstand, dass die Resynthese ohne vorherige Handsegmentierung der Daten durchge-

führt werden und dadurch das Feedback den Lernenden unmittelbar nach ihrer falschen Aussprache angeboten werden konnte, macht eine solche Technik in zukünftiger Sprachlernsoftware einsetzbar.

In dem von uns durchgeführten Aussprachetraining wurde eine Probandengruppe mit resynthetisierten Stimuli in ihrer eigenen Stimme und eine andere Probandengruppe mit natürlichen Stimuli in der Stimme einer Muttersprachlerin trainiert. Zusätzlich erhielten beide Gruppen Feedback zunächst mit normal betonten und danach mit überbetonten Stimuli. Die Ergebnisse eines Perzeptionstests mit den vor und nach dem Training ausgesprochenen Äußerungen zeigen, dass das traditionelle Feedback aus natürlichen normal betonten Stimuli nicht ausreichend ist, um italienischsprachige Deutschlernende die korrekte Betonungsstelle erkennen zu lassen. Sie benötigen hierfür zusätzliche Hinweise. Feedback aus resynthetisierten Stimuli in der eigenen Stimme erwies sich als effektiver, denn sie reduzieren die phonetische Variabilität und können daher von den Lernenden mit ihren falsch betonten Äußerungen einfacher verglichen werden. Zudem üben die resynthetisierten Stimuli einen motivierenden Effekt auf die Lernenden aus.

Darüber hinaus war das Feedback aus überbetonten natürlichen Stimuli effektiver als das Feedback aus normal betonten natürlichen Stimuli. Jedoch imitierten die italienischen Trainingsteilnehmer häufig die überbetonten Äußerungen, so dass ihre Produktionen von den deutschen Hörern oft schlechter bewertet wurden.

Schließlich bestätigten die Ergebnisse eines Perzeptionstests zur Betonungswahrnehmung morphologisch komplexer Wörter im Deutschen die Hypothese, dass die Betonungsfehler von italienischsprachigen Deutschlernenden perzeptive Ursachen haben. Die italienischen Probanden machten im Test signifikant mehr Fehler in der Unterscheidung von Wörtern eines Betonungscontrasts als in der Unterscheidung von Wörtern eines Phonemcontrasts und von Wörtern,

die sich sowohl hinsichtlich der Betonung als auch hinsichtlich der Quantität der Pänultima differenzierten. Die italienischsprachigen Probanden konnten die Aufgabe eines einfachen Identifikationstests mit Wörtern des Betonungskontrasts lösen. Jedoch waren sie nicht in der Lage, sich Serien solcher Wörter einzuprägen. Eine denkbare Erklärung hierfür wäre, dass morphologisch komplexe Wörter im Deutschen, die sich hinsichtlich der Stelle der Haupt- und Nebenbetonung unterscheiden, als Homophone im mentalen Lexikon italienischsprachiger Deutschlerner gespeichert werden. Bemerkenswert ist zudem, dass die meisten italienischen Probanden den im Text dargebotenen Betonungs-Quantitätskontrast als Betonungskontrast einstufen. Tatsächlich ähnelte der deutsche Betonungs-Quantitätskontrast durch die Änderung der Vokaldauer einem italienischen Betonungskontrast.

9.2 Ausblick

Die Ergebnisse dieser Arbeit haben neuartige Fragestellungen für künftige Forschungen aufgeworfen. Zunächst haben wir vor, die Betonungswahrnehmung von italienischen und deutschen Muttersprachlern weiter zu untersuchen. In Kap. 8.6 wurden die bereits durchgeführten Aufnahmen beschrieben, die im Rahmen der Planung eines zweiten Experiments zur Betonungswahrnehmung gemacht worden sind. Im neuen Perzeptionsexperiment werden wir nur Nichtwörter verwenden, die sowohl italienischen als auch deutschen Hörern unbekannt sind, um keine der beiden Gruppen gegenüber der anderen zu bevorzugen. Es sollte zudem untersucht werden, ob Italiener bzw. Deutsche dazu neigen, eine bestimmte Betonungsstelle zu bevorzugen (s. Kap. 2.3.2 auf S. 37), was die Ergebnisse der anderen Perzeptionstests beeinflussen könnte.

Darüber hinaus planen wir, Untersuchungen zu den akustischen und perceptiven Korrelaten der Betonung morphologisch komplexer Wörter im Deutschen

durchzuführen, um bei solchen Wörtern die Relevanz der Grundfrequenz, der Dauer und der Intensität als Betonungskorrelate zu testen. Zu diesem Zweck könnten die bereits im Experiment 5 verwendeten Wörter, die in einer weiteren Aufnahme in Trägersätzen eingebettet und von einer weiblichen Sprecherin in Nichtfokusposition ausgesprochen wurden, benutzt werden. Es wurden zwei Trägersätze aufgenommen, vor denen die Sprecherin auch die dazugehörige Frage aussprach.¹ Ursprünglich beabsichtigten wir, diese Wörter als Stimuli im Experiment 5 zu verwenden (s. Kap. 8.5.1.4). In einem informellen Test überprüften wir, ob ein deutscher Muttersprachler in der Lage war, die betonte Silbe der aufgenommenen morphologisch komplexen Wörter zu erkennen. Die Einzelwörter wurden ihm in zufälliger Reihenfolge präsentiert. Er konnte die Betonung der Wörter in Fokusposition problemlos erkennen, während die Lokalisierung der Betonung in Nichtfokusposition für ihn schwierig war. Daher beschlossen wir, dass diese Stimuli für den Perzeptionstest mit den italienischen Probanden ungeeignet waren. Für die geplante Untersuchung zu den Korrelaten der Betonung morphologisch komplexer Wörter im Deutschen wäre das aufgenommene Material jedoch verwendbar.

Es sollte in naher Zukunft überprüft werden, ob das Training der Wortbetonung mittels resynthetisierter bzw. überbetonter Stimuli langfristige Wirkungen hat. Dies ist vorstellbar, denn durch diese Art von Feedback bekommen die Trainingsteilnehmer einen zusätzlichen Hinweis auf die korrekte Betonungsstelle, die sie vorher nicht bemerkt hatten. Diese damit gewonnene Aufmerksamkeit sollte verhindern, dass sie in der Zukunft weiterhin systematisch solche Betonungsfehler machen.

Schließlich soll die Effektivität resynthetisierter Stimuli als Feedback auch

¹Es handelte sich um folgende Trägersätze mit Fragen:

- a) Wann habe ich <Zielwort> gesprochen? Ich habe gerade <Zielwort> gesprochen.
- b) Habe ich gestern <Zielwort> gesprochen? Ich habe vorgestern <Zielwort> gesprochen.

beim Training weiterer prosodischer Eigenschaften getestet werden. Es soll überprüft werden, ob sich die Prosodie italienischsprachiger Deutschlerner nach Feedback aus resynthetisierten Stimuli stärker verbessert als nach Feedback aus natürlichen Stimuli. Weitere prosodische Phänomene könnten unter die Lupe genommen werden. Zum Beispiel könnte der Satzakzent oder die Reduktion von Funktionswörtern im Vergleich zu Inhaltswörtern betrachtet werden.

Die Ergebnisse unserer Arbeit zeigen zum einen, dass die Resynthese zur Korrektur der Betonung in Äußerungen von L2-Lernern ohne vorherige manuelle Segmentierung der Sprachsignale realisierbar ist, und zum anderen, dass solche resynthetisierten Stimuli in einem Aussprachetraining effektiver als natürliche Stimuli sind. Somit empfiehlt sich diese Technik für den Einsatz in künftigen CALL-Systemen. Darüber hinaus haben wir nachgewiesen, dass selbst bei Sprechern einer L1, in der — wie im Italienischen — die Betonung phonologisch kontrastiv ist, eine „*stress deafness*“ für Betonungskontraste der L2 — in unserem Fall dem Deutschen — vorhanden sein kann. Weitere Untersuchungen zu diesem Thema sollen unbedingt durchgeführt werden.

Anhang

A Textmaterial

A.1 Die Buttergeschichte

Es war in Berlin zu einer Zeit, als Lebensmittel nicht genügend vorhanden waren. Vor einem Laden stand bereits um sieben Uhr eine beachtliche Menschenmenge, denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter eingetroffen sei. Jeder wusste, dass die Butter schnell ausverkauft sein würde, und dass man ganz früh kommen müsse, um noch etwas zu erhalten. Da das Geschäft erst um acht geöffnet wurde, stellten sich die Leute vor der Ladentür in einer Reihe an. Wer später kam, musste sich hinten anschließen.

Je näher der Zeiger auf acht kam, desto unruhiger wurden die Leute. Da kam endlich ein kleiner Mann mit grauem Haar und drängte sich ziemlich rücksichtslos nach vorn. Die wartenden Menschen waren empört über solches Verhalten und forderten ihn auf, sich ebenfalls hinten anzustellen. Aber auch als schon mit der Polizei gedroht wurde, ließ sich der Mann nicht beirren, sondern drängte sich weiter durch. Er bat, man solle ihn doch durchlassen, oder glaubte man, dass diese Drängelei für ihn vielleicht ein Vergnügen sei?

Das war für die Leute nun doch zu viel! Alle kochten bereits vor Wut, und der Mann konnte jetzt von allen Seiten Schimpfwörter hören. Er aber zuckte resigniert mit den Schultern und bemerkte: „Nun gut, wie Sie wollen. Wenn Sie mich nicht vorlassen, dann kann ich die Tür nicht aufschließen, und Sie

können meinetwegen hier stehen bleiben, bis die Butter ranzig geworden ist.“

A.2 Sätze aus der Buttergeschichte

Die „Buttergeschichte“ wurde in folgende Sätze segmentiert — hier in der kanonischen Transkription in SAMPA nach PhonDat eingegeben (s. Pompino-Marschall 1992 [121]):

- s00 d i:+ b 'U t 6 #g @ S "I C t @
- s01 Q E s+ v a:6+ Q I n+ b E 6 l 'i: n t s u:+ Q a I n 6+ t s 'a I t ,
Q a l s+ l 'e: b @ n s #m "I t @ l n I C t+ g @ n 'y: g @ n t f o:6 h 'a
n d @ n v a: r @ n+ .
- s02 f o:6+ Q a I n @ m+ l 'a: d @ n S t 'a n t b @ r 'a I t s Q U m+ z 'i:
b @ n Q 'u:6 Q a I n @+ b @ Q 'a x t l I C @ m 'E n S @ n #m "E N @ ,
- s03 d E n+ m a n+ h a t @+ d '06 t Q a m+ Q 'a: b @ n t f 'o:6 h e:6 Q
a U f+ Q a I n @ m+ S 'I l t S o: n+ l 'e: z @ n k 9 n @ n+ , d a s+ f
r 'I S @ b 'U t 6 Q 'a I n #g @ t r "0 f @ n z a I+ .
- s04 j 'e: d 6 v 'U s t @ , d a s+ d i:+ b 'U t 6 S n 'E l Q 'a U s #f 6
k "a U f t z a I n+ v Y 6 d @+ , Q U n t+ d a s+ m a n+ g 'a n t s f r
'y: k '0 m @ n m Y s @+ , Q U m+ n 0 x+ Q E t v a s+ t s u:+ Q E 6 h
'a l t @ n .
- s05 d a:+ d a s+ g @ S 'E f t Q 'e:6 s t Q U m+ Q 'a x t g @ Q '9 f n @
t v U 6 d @+ , S t 'E l t @ n z I C+ d i:+ l 'O Y t @ f o:6+ d e:6+ l
'a: d @ n #t "y:6 Q I n+ Q a I n 6+ r 'a I @ Q 'a n .
- s06 v e:6+ S p 'E: t 6 k 'a: m , m U s t @+ z I C+ h 'I n t @ n Q 'a n
#S l "i: s @ n .
- s07 j e:+ n 'E:6 d e:6+ t s 'a I g 6 Q a U f+ Q 'a x t k 'a: m , d E s

t o:+ Q 'U n #r "u: I g 6 v U6 d @ n+ d i:+ l 'OY t @ .

s08 d a:+ k 'a: m Q 'E n t l I C Q a I n+ k l 'a I n 6 m 'a n m I t+ g r
'a U @ m h 'a: 6 Q U n t+ d r 'E N t @ z I C+ t s 'i: m l I C r 'Y k #z
"I C t s l o: s n a: x+ f '06 n .

s09 d i:+ v 'a6 t @ n d @ n m 'E n S @ n v a: r @ n+ Q E m p '2: 6 t Q
y: b 6+ z O l C @ s+ f 6 h 'a l t @ n Q U n t+ f '06 d 6 t @ n Q i: n+
Q 'a U f , z I C+ Q 'e: b @ n #f "a l s h 'I n t @ n Q 'a n #t s U S t
"E l @ n .

s10 Q a: b 6+ Q a U x+ Q a l s+ S o: n+ m I t+ d e: 6+ p o: l I t s 'a I
g @ d r 'o: t v U6 d @+ , l 'i: s z I C+ d e: 6+ m 'a n n I C t+ b @ Q
'I r @ n , z '0 n d 6 n d r 'E N t @ z I C+ v 'a I t 6 d 'U6 C .

s11 Q e: 6+ b 'a: t , m a n+ z O l @+ Q i: n+ d O x+ d 'U6 C #l "a s @
n , Q o: d 6+ g l 'a U p t @ m a n+ , d a s+ d i: z @+ d r E N @
l 'a I f y: 6+ Q i: n+ f I l 'a I C t Q a I n+ f 6 g n 'y: g @ n z a I+ ?

s12 d a s+ v a: 6+ f y: 6+ d i:+ l 'OY t @ n u: n+ d O x+ t s u: + f
'i: l !

s13 Q 'a l @ k 'O x t @ n b @ r 'a I t s f o: 6+ v 'u: t , Q U n t+ d
e: 6+ m 'a n k O n t @+ j 'E t s t f O n+ Q 'a l @ n z 'a I t @ n S 'I m
p f #v "9 6 t 6 h '2: r @ n .

s14 Q e: 6+ Q a: b 6+ t s 'U k t @ r E z I g n 'i: 6 t m I t+ d e: n+ S
'U l t 6 n Q U n t+ b @ m 'E6 k t @ , n u: n+ g 'u: t , v i:+ z i:+
v O l @ n+ .

s15 v E n+ z i:+ m I C+ n I C t+ f 'o: 6 #l "a s @ n , d a n+ k a n+
Q I C+ d i:+ t 'y: 6 n I C t+ Q 'a U f #S l "i: s @ n ,

s16 Q U n t+ z i:+ k 9 n @ n+ m 'a I n @ t v "e: g @ n h 'i: 6 S t 'e: @

n b l 'a I b @ n , b I s + d i : + b 'U t 6 r 'a n t s I C g @ v 0 6 d @ n +
Q I s t + .

B Resynthese

B.1 Programm zum Kopieren der Prosodie eines L1-Sprechers auf einen L2-Sprecher

Im Folgenden wird das komplette Unix-Shell-Programm dargestellt, das zum Kopieren der Dauer der Phone, der Grundfrequenz und der Intensität der Äußerungen der deutschen Muttersprachlerin auf die Äußerungen der italienischsprachigen Deutschlernenden verwendet wurde. Die Angabe *mip* bezeichnet ein von Hartmut R. Pfitzinger entwickeltes Datei-Format für Signale, das dem NIST-Format ähnlich ist.

Die Durchführung des Programms für einen Satz erfolgt schnell, in ca. 1-2 Sekunden, abhängig von der Geschwindigkeit des Rechnerprozessors. Das Output des Programms für eine Beispieläußerung wird in Abb. 7.3 in Kap. 7.4 dargestellt.

```
# Programm zum Kopieren der Dauer der Phone, des Grundfrequenzverlaufs und der Intensität eines Signals
# der deutschen L1-Sprecherin auf das Sprachsignal eines L2-Sprechers
#
# Maria Paola Bissiri & Hartmut R. Pfitzinger
# 13.04.2006
#
# Zunächst Normalisierung der Amplitude der Signale der deutschen Muttersprachlerin und des L2-Sprechers,
# damit keine zu unterschiedlichen Amplitudenwerte zu einer fehlerhaften Alignierung führen:
#
promod/pro_level -wtKai30 -ws20 -g10 -ouSig -ct65 -cr4 -xt50 -xr4 -Ipromod/.mip00 -Opromod/.mip04
```

```
promod/pro_level -wtKai30 -ws20 -g10 -ouSig -ct65 -cr4 -xt50 -xr4 -Ipromod/.mip02 -Opromod/.mip05
#
# Alignierung der beiden amplitudennormalisierten Signale mithilfe eines Dynamic Time Warping-Algorithmus:
#
promod/pro_dtw -o1.RSR -s10 -w20 -b10 -t1. -vsmnLP -l16 -fl200 -fh8 -r300 -mOFF -xa0 -xl51 -ya0 -yl75\
-Ipromod/.mip04 -Ipromod/.mip05 -Opromod/.mip06
promod/pro_sr_mod -i0N -m100 -s100 -Ipromod/.mip06 -Opromod/.mip07
#
# Resynthese des neuen Signals des italienischen Sprechers mit korrigierter Dauer der Phone:
#
promod/pro_sr_chg -i0sz -s12 -d6 -Ipromod/.mip00 -Ipromod/.mip07 -Opromod/.mip08
#
# Extraktion der Grundfrequenz des dauersynthetisierten Signals des L2-Sprechers:
#
promod/pro_f0isla -ouSeg -mn120 -mx500 -pt0.15 -pl0.55 -wnPar20 -Ipromod/.mip08 -Opromod/.mip09
#
# Grundfrequenzverlauf des Signals der deutschen Muttersprachlerin berechnen:
#
promod/pro_f0isla -ouSeg -mn120 -mx500 -pt0.15 -pl0.55 -wnPar20 -Ipromod/.mip02 -Opromod/.mip10
promod/pro_seg2f -ws15 -ss1. -gs20 -mn25 -ex0 -inNo -of0.5 -Ipromod/.mip10 -Opromod/.mip11
#
# Interpolation des Grundfrequenzverlaufs der deutschen Muttersprachlerin, um die Null-Werte zu
# eliminieren:
#
promod/pro_prsana -f70 -iOFF -oFT -s4 -h3. -l0N -b120. -a2. -t20. -g0.9 -c0.2 -d-85 -e0N -j0\
-Ipromod/.mip11 -Opromod/.mip12
#
# Berechnung des Mittelwerts des Grundfrequenzverlaufs der deutschen Muttersprachlerin:
#
promod/pro_func -t-mean -i0N -d0. -g0. -o0N -Ipromod/.mip12 -Opromod/.mip13
promod/pro_func -tx-y -i0N -d0. -g0. -o0N -Ipromod/.mip12 -Ipromod/.mip13 -Opromod/.mip14
#
# Grundfrequenzverlauf des Signals des L2-Sprechers berechnen:
#
promod/pro_seg2f -ws15 -ss1. -gs20 -mn25 -ex0 -inNo -of0.5 -Ipromod/.mip09 -Opromod/.mip15
#
# Interpolation des Grundfrequenzverlaufs des L2-Sprechers:
#
promod/pro_prsana -f70 -iOFF -oFT -s4 -h3. -l0N -b120. -a2. -t20. -g0.9 -c0.2 -d-85 -e0N -j0\
-Ipromod/.mip15 -Opromod/.mip16
#
# Berechnung des Mittelwerts des Grundfrequenzverlaufs des L2-Sprechers:
#
promod/pro_func -t-mean -i0N -d0. -g0. -o0N -Ipromod/.mip16 -Opromod/.mip17
```

```
promod/pro_func -tx-y -iON -d0. -g0. -oON -Ipromod/.mip16 -Ipromod/.mip17 -Opromod/.mip18
#
# Division des Grundfrequenzverlaufs der deutschen Muttersprachlerin durch seinen eigenen Mittelwert
# und Multiplikation mit dem Mittelwert des Grundfrequenzverlaufs des L2-Sprechers:
#
promod/pro_func -tx/y -iON -d0. -g0. -oON -Ipromod/.mip12 -Ipromod/.mip14 -Opromod/.mip19
promod/pro_func -tx*y -iON -d0. -g0. -oON -Ipromod/.mip18 -Ipromod/.mip19 -Opromod/.mip20
#
# Umrechnung des ermittelten Grundfrequenzverlaufs in Stimmtöneblöcke:
#
promod/pro_sr_mod -iOFF -m0 -s100 -Ipromod/.mip15 -Opromod/.mip21
promod/pro_func -tx -iON -d0. -g-40 -oON -Ipromod/.mip21 -Opromod/.mip22
promod/pro_func -tx*y -iON -d0. -g0 -oON -Ipromod/.mip20 -Ipromod/.mip22 -Opromod/.mip23
#
# Bestimmung der neuen Glottisimpulspositionen:
#
promod/pro_f2seg -f32000. -Ipromod/.mip23 -Opromod/.mip24
#
# Resynthese des neuen Signals des italienischen Sprechers mit korrigiertem Grundfrequenzverlauf
# zusätzlich zur korrigierten Dauer der Phone:
#
promod/pro_f0_chg -g50 -tHQ -f5 -wHann -Ipromod/.mip08 -Ipromod/.mip09 -Ipromod/.mip24 -Opromod/.mip25
#
# Normalisierung der Amplituden des resynthetisierten Signals des L2-Sprechers und des Signals der
# deutschen Muttersprachlerin:
#
promod/pro_level -ouSig -wtKai30 -ws15 -mRMS -g10 -lt60 -ct65 -cr1 -xt50 -xr1 -Ipromod/.mip25\
-Opromod/.mip26
promod/pro_level -ouSig -wtKai30 -ws15 -mRMS -g0 -lt90 -ct65 -cr1 -xt70 -xr0.01 -Ipromod/.mip02\
-Opromod/.mip27
#
# Berechnung der Einhüllenden der Amplitude der deutschen Muttersprachlerin:
#
promod/pro_level -oudB -wtKai30 -ws15 -mRMS -g0 -lt90 -ct65 -cr1 -xt70 -xr0.01 -Ipromod/.mip27\
-Opromod/.mip28
#
# Endergebnis = Resynthese des neuen Signals des italienischen Sprechers mit korrigierter Intensität
# zusätzlich zur Dauer der Phone und zum Grundfrequenzverlauf:
#
promod/pro_level -ouVCA -wtKai30 -ws15 -mRMS -g-75 -lt90 -ct65 -cr1 -xt60 -xr0.01 -Ipromod/.mip26\
-Ipromod/.mip28 -Opromod/.mip29
#
# Berechnung der neuen Einhüllenden der Amplitude des L2-Sprechers:
#
```

```
promod/pro_level -oudB -wtKai30 -ws20 -mRect -g10 -lt90 -ct65 -cr4. -xt50 -xr4. -Ipromod/.mip29\  
-Opromod/.mip30
```

C Experiment 1

C.1 Stimuli im Perzeptionstest

Auflistung der in Experiment 1 verwendeten Stimuli. Die Buchstaben vor den Sätzen entsprechen den Buchstaben auf den zugehörigen Knöpfen im Perzeptionstest.

- a. Die Buttergeschichte.
- b. Es war in Berlin zu einer Zeit, als Lebensmittel nicht genügend vorhanden waren.
- c. Vor einem Laden stand bereits um sieben Uhr eine beachtliche Menschenmenge,
- d. denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter eingetroffen sei.
- e. Jeder wusste, dass die Butter schnell ausverkauft sein würde, und dass man ganz früh kommen müsse, um noch etwas zu erhalten.
- f. Wer später kam, musste sich hinten anschließen.
- g. Je näher der Zeiger auf acht kam, desto unruhiger wurden die Leute.
- h. Das war für die Leute nun doch zu viel!

- i. Er aber zuckte resigniert mit den Schultern und bemerkte: „Nun gut, wie Sie wollen.
- j. Wenn Sie mich nicht vorlassen, dann kann ich die Tür nicht aufschließen,
- k. und Sie können meinetwegen hier stehen bleiben, bis die Butter ranzig geworden ist.“

C.2 Instruktionen für die Probanden

Im mittleren Bereich des Fensters sehen Sie zwei Reihen mit Knöpfen: 88 gelbe und 88 graue.

Wenn Sie mit der linken Maustaste auf die Knöpfe klicken, werden Sie einige Sätze aus der "Buttergeschichte" hören, die von deutschlernenden Italienern ausgesprochen wurden. Jedes Knopfpaar ist mit der selben Buchstaben-Zahlenkombination beschriftet (Kleinbuchstaben für die grauen Knöpfe und Großbuchstaben für die gelben Knöpfe). Hinter jedem Knopfpaar verbergen sich zwei Versionen desselben Satzes, gesprochen von jeweils dem selben Sprecher.

Ihre Aufgabe ist, den Sprechrhythmus der durch die grauen Knöpfe abspielbaren Sätze mit dem der jeweils zugehörigen gelben Knöpfe zu vergleichen und dann die grauen Knöpfe entsprechend zu verschieben, falls Sie eine Verbesserung oder Verschlechterung feststellen. Um einen Knopf zu bewegen, muß die linke Maustaste gedrückt gehalten und die Maus dabei bewegt werden.

Wenn Sie der Meinung sind, dass der Sprechrhythmus des grauen Knopfes besser als der des gelben Knopfes ist, dann verschieben Sie bitte den grauen Knopf in den oberen Bereich des Fensters.

Meinen Sie aber, dass der Sprechrhythmus in beiden Versionen gleich gut ist, bewegen Sie bitte den grauen Knopf nicht und lassen Sie ihn im ursprünglichen Bereich.

Sind Sie schließlich der Meinung, dass der Sprechrhythmus des grauen Knopfes schlechter als der des gelben Knopfes ist, verschieben Sie bitte den grauen

Knopf in den unteren Bereich des Fensters.

Dadurch, dass Sie einen bestimmten Bereich wählen, können Sie die Stärke der wahrgenommenen Qualitätsänderung des Sprechrhythmus ausdrücken.

Folgende Bereiche stehen zur Verfügung:

- große Verbesserung
- mittlere Verbesserung
- leichte Verbesserung
- keine Verbesserung/Verschlechterung
- leichte Verschlechterung
- mittlere Verschlechterung
- große Verschlechterung

Je besser Ihnen der Sprechrhythmus eines grauen Stimulus im Vergleich zum gelben erscheint, desto weiter oben -- auch innerhalb eines Bereichs -- sollten Sie ihn positionieren. So können Sie auch sehr kleine Unterschiede ausdrücken.

Bevor Sie anfangen, hören Sie sich einige Stimuli an, um einen Eindruck zu bekommen.

Vergleichen Sie bitte jeden grauen Knopf nur mit dem zugehörigen gelben Knopf und nicht die grauen Knöpfe miteinander.

Hier noch ein paar kurze Hinweise:

- Sie sollten zwischen 20 und 30 Minuten für das Experiment aufwenden.
- Wenn Sie das Programm beenden wollen, dann nur über den "ENDE"-Knopf oben links. Sie können das Programm jederzeit beenden; bei einem erneuten Start sind alle Knöpfe da, wo Sie sie zuletzt plaziert hatten.

Viel Spass!

D Experiment 2

D.1 Wörter aus der Buttergeschichte

D.1.1 Morphologisch komplexe Wörter

Buttergeschichte, Lebensmittel, Menschenmenge, eingetroffen, ausverkauft, Ladentür, anschließen, unruhiger, rücksichtslos, ebenfalls, anzustellen, durchlassen, Schimpfwörter, vorlassen, aufschließen, meinetwegen.

D.1.2 Komplette Wortliste

1. Buttergeschichte
2. Berlin
3. Lebensmittel
4. genügend
5. vorhanden
6. beachtliche
7. Menschenmenge

8. eingetroffen
9. ausverkauft
10. erhalten
11. geöffnet
12. Ladentür
13. anschließen
14. unruhiger
15. rücksichtslos
16. wartenden
17. Verhalten
18. forderten
19. ebenfalls
20. anzustellen
21. gedroht
22. beirren
23. durchlassen
24. Drängelei
25. vielleicht
26. Vergnügen

27. Schimpfwörter
28. resigniert
29. bemerkte
30. vorlassen
31. aufschließen
32. meinetwegen

D.2 Stimuli im Perzeptionstest

Beim Perzeptionsexperiment zum Einfluss der Intonation und der lokalen Sprechgeschwindigkeit auf die Betonung von morphologisch komplexen Wörtern im Deutschen (Kap. 8.2) hörten die Versuchspersonen folgende Stimuli. Die Stimuli sind in der Reihenfolge, in der sie auf der Testoberfläche vorkamen (siehe hierzu Abb. 8.6 auf S. 115). Die zu beurteilenden Wörter sind hervorgehoben:

- a. Wenn Sie mich nicht *vorlassen*, dann kann ich die Tür nicht aufschließen.
- b. *vorlassen*
- c. Wenn Sie mich nicht vorlassen, dann kann ich die Tür nicht *aufschließen*.
- d. *unruhiger*
- e. Je näher der Zeiger auf acht kam, desto *unruhiger* wurden die Leute.
- f. *ausverkauft*

- g. Jeder wusste, dass die Butter schnell *ausverkauft* sein würde, und dass man ganz früh kommen müsse, um noch etwas zu erhalten.
- h. *anschließen*
- i. Vor einem Laden stand bereits um sieben Uhr eine beachtliche *Menschenmenge*.
- j. Denn man hatte dort am Abend vorher auf einem Schild schon lesen können, dass frische Butter *eingetroffen* sei.

E Experiment 5

E.1 Deutsche Minimalpaare

Minimalpaare mit Langvokal in der Pänultima

'durchschlagen.	*'durchschlaggen.	durch'schlagen.
'durchwaten.	*'durchwatten.	durch'waten.
'umlagern.	*'umlaggern.	um'lagern.
'umfahren.	*'umfarren.	um'fahren.

Minimalpaare mit Kurzvokal in der Pänultima

'durchwandern.	durch'wandern.	*durch'wahndern.
'umfassen.	um'fassen.	*um'faßen.
'umpflanzen.	um'pflanzen.	*um'pflahnzen.
'umspannen.	um'spannen.	*um'spahnen.

E.2 Sprecheranweisungen für die Aufnahme der Stimuli für das Experiment 5

Abb. E.1 zeigt die Anweisungen, die die Sprecher vor der Aufnahme für das Experiment 5 erhielten (s. Kap. 8.5.1.1.2).

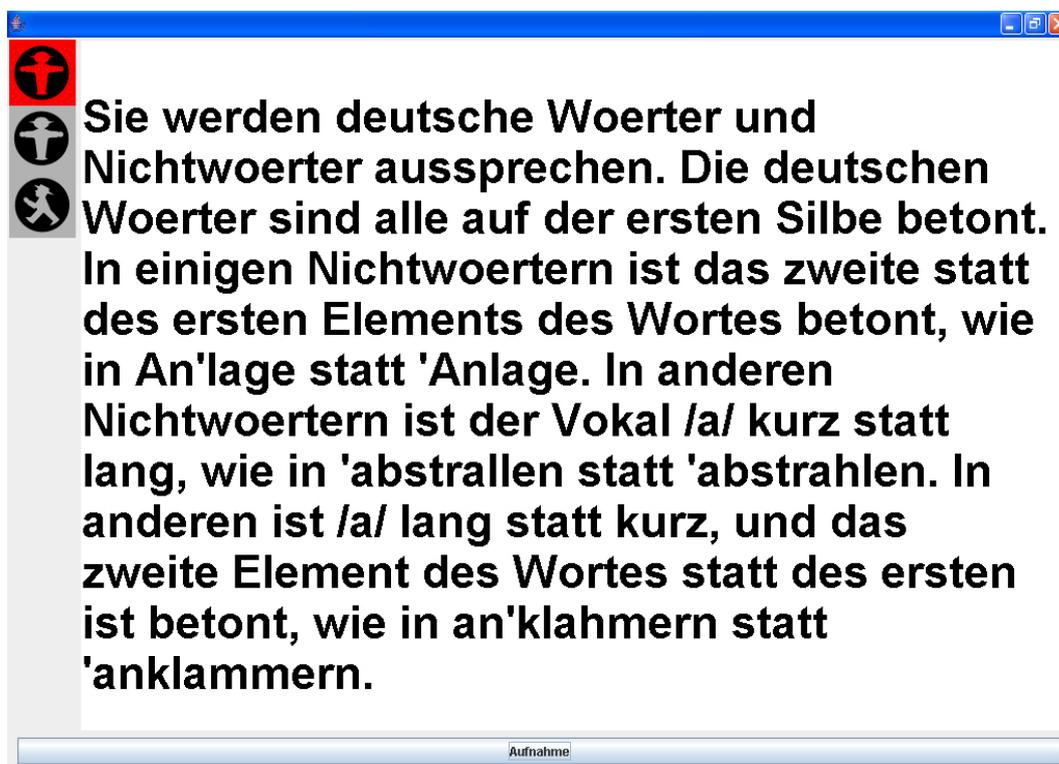


Abbildung E.1: Aufnahmeanweisungen für die Sprecher für Experiment 5

Abb. E.2 und E.3 zeigen die Übungsfenster für die Wörter im Original jeweils mit Lang- und mit Kurzvokal, die die Sprecher in Vorbereitung zur Aufnahme der Stimuli für das geplante sechste Experiment erhielten (s. Kap. 8.5.1.1.2).

E.3 Perzeptionstest

E.3.1 Anweisungen für die Probanden

Abb. E.4 stellt die Anweisungen für die italienischen Testteilnehmer dar.

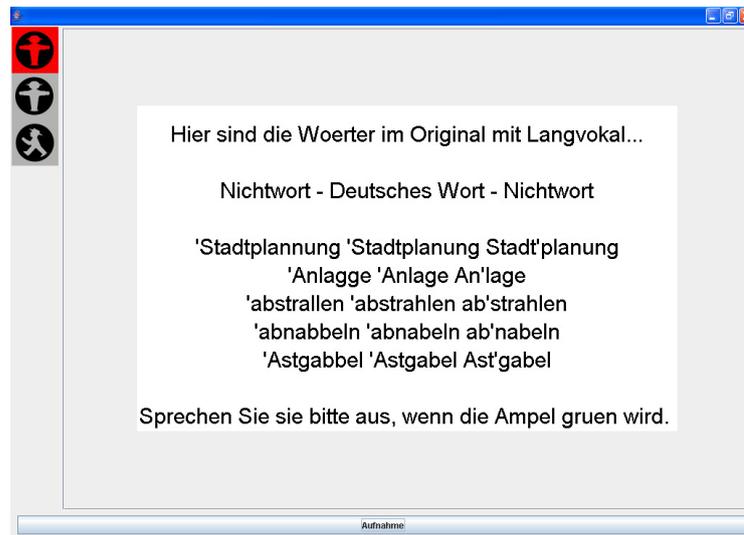


Abbildung E.2: Übungsfenster mit den Wörtern mit Langvokal für die Sprecher für Experiment 5

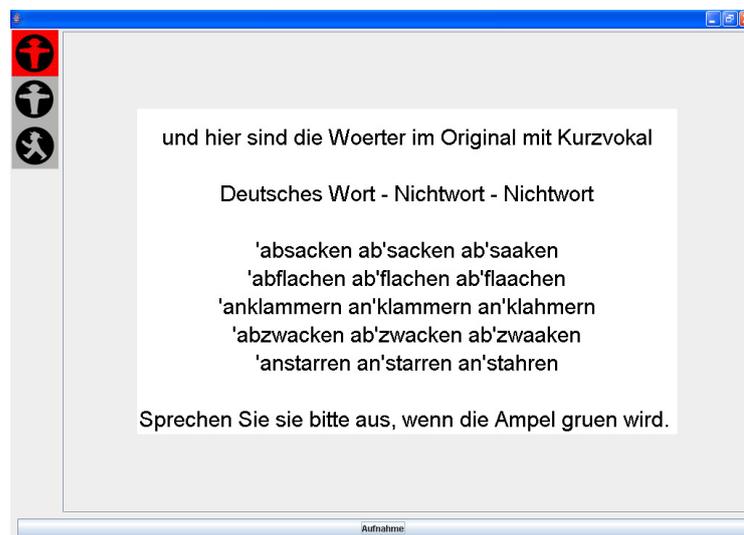


Abbildung E.3: Übungsfenster mit den Wörtern mit Kurzvokal für die Sprecher für Experiment 5

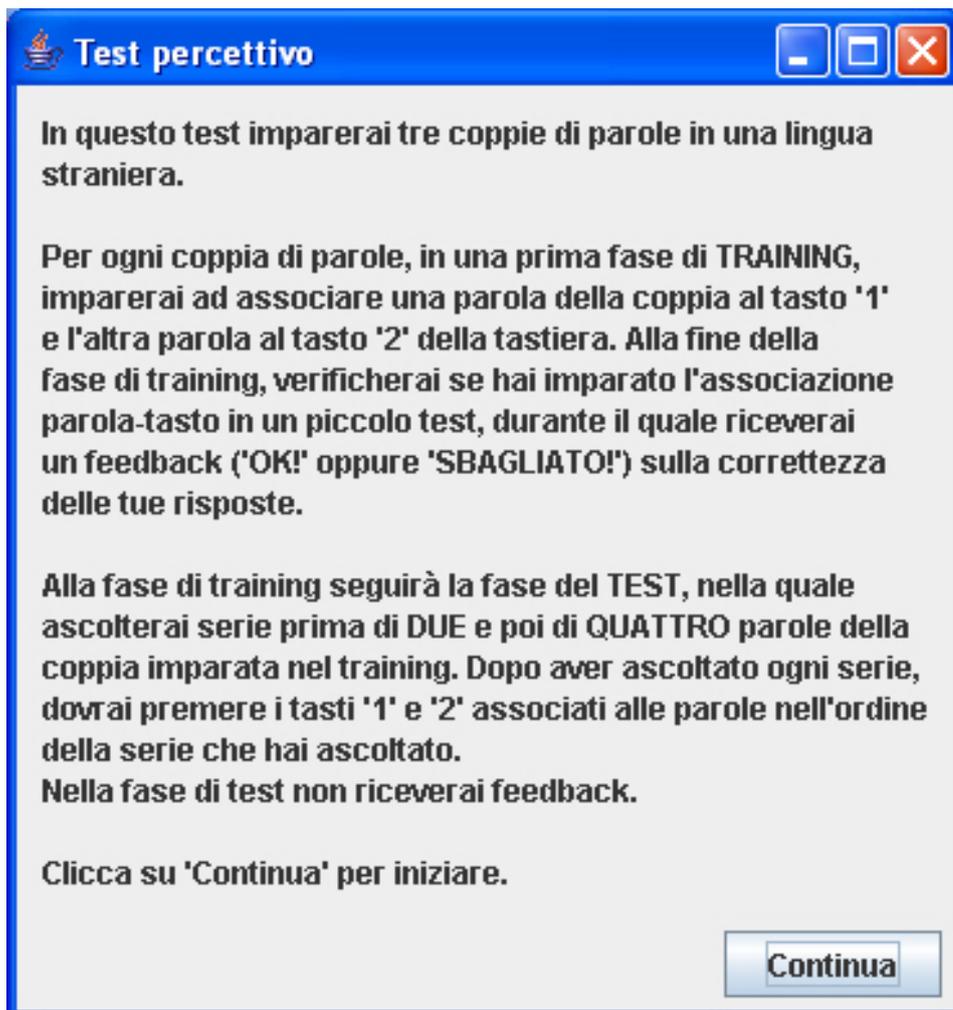


Abbildung E.4: Anweisungen für die italienischen Probanden.

F Voruntersuchungen zum geplanten sechsten Experiment

F.1 Nichtwörter für die Aufnahme der japanischen Sprecherin

Folgende Wörter wurden in den Trägersatz „I say: <Zielwort>.“ eingebettet und von einer japanischen Sprecherin aufgenommen (s. dazu Kap. 8.6.1.1). Jeder Satz wurde von der Sprecherin acht mal gelesen. Die zweimorigen Vokale werden durch einen nachstehenden Doppelpunkt gekennzeichnet. Die Pitch-Akzente werden mit einem Überstrich markiert, wie von der japanischen Sprecherin vorgeschlagen.

Typ A	Typ B	Typ C
pa:midu	<i>pā</i> mi:du	pami:du
ta:nugi	<i>tā</i> nu:gi	tanu:gi
ka:pimu	<i>kā</i> pi:mu	kapi:mu
ba:tuni	<i>bā</i> tu:ni	batu:ni
da:kipu	<i>dā</i> ki:pu	daki:pu
ga:buti	<i>gā</i> bu:ti	gabu:ti
ma:diku	<i>mā</i> di:ku	madi:ku
na:gubi	<i>nā</i> gu:bi	nagu:bi

F.2 Nichtwörter für die Aufnahme der kantonesischen Sprecherin

Folgende Nichtwörter wurden in Isolation von einer Sprecherin des Kantonesischen aufgenommen (s. dazu Kap. 8.6.1.2). Jedes Nichtwort wurde von der Sprecherin acht mal gelesen. Die Silben mit Pitch-Akzent wurden fett gedruckt und durch einen Gravis gekennzeichnet.

a:spano	aspa:no	à spa:no
a:stami	asta:mi	à sta:mi
a:skape	aska:pe	à ska:pe
a:mpato	ampa:to	à mpa:to
a:mnaki	amna:ki	à mnaki

Literaturverzeichnis

- [1] Albano Leoni, F.; Cutugno, F.; Savy, R. (1995). The vowel system of Italian connected speech. In: *Proceedings of the 13th International Congress of Phonetic Sciences*, S. 396–399, Stockholm.
- [2] Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20(1): S. 51–62.
- [3] Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11(4): S. 5–22.
- [4] Anderson-Hsieh, J.; Johnson, R.; Koehler, K. (1992). The relationship between native speaker judgements of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42: S. 529–555.
- [5] Anderson-Hsieh, J.; Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, 38: S. 561–613.
- [6] Berinsein, A. E. (1979). A cross-linguistic study on the perception and production of stress. In: *UCLA Working Papers in Phonetics 47*. University of California, Los Angeles.

- [7] Bertinetto, P. M. (1976). L'accento secondario nella fonologia italiana. analisi teorica e sperimentale. In: Simone, R.; Vignuzzi, U.; Ruggiero, G. (Hrsg.), *Studi di fonetica e di fonologia*, S. 189–236. Bulzoni, Roma.
- [8] Bertinetto, P. M. (1977). "Syllabic blood" ovvero l'italiano come lingua ad isocronismo sillabico. *Studi di grammatica italiana*, 6: S. 69–96.
- [9] Bertinetto, P. M. (1980). Perception of stress by Italian speakers. *Journal of Phonetics*, 8: S. 385–395.
- [10] Bertinetto, P. M. (1981). *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Accademia della Crusca, Firenze.
- [11] Campbell, N.; Mokhtari, P. (2002). Voice quality: The 4th prosodic dimension. In: *Proceedings of the XVth International Congress of Phonetic Sciences*, Band 3, S. 2417–2420, Barcelona.
- [12] Canepari, L. (1979). *Introduzione alla fonetica*. Einaudi, Torino.
- [13] CELEX. (2004). The CELEX lexical database. Max Planck Institute for Psycholinguistics. <http://www.ru.nl/celex/>, 16.08.2004.
- [14] Chapelle, C. A. (1998). Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, 2(1): S. 22–34, Juli. <http://llt.msu.edu/vol2num1/article1/>.
- [15] Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning*, 27: S. 29–46.
- [16] Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal*, 7(1): S. 21–46.

-
- [17] Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology*, 2(1): S. 61–77, July. <http://llt.msu.edu/vol2num1/article4/>.
- [18] Clark, J.; Swinton, S. (1979). An exploration of speaking proficiency measures in the TOEFL context (TOEFL Research Report No. 4). Bericht, Educational Testing Service.
- [19] Clark, J.; Swinton, S. (1980). The Test of Spoken English as a measure of communicative ability in English-medium instructional settings (TOEFL Research Report No. 7). Bericht, Educational Testing Service.
- [20] Cranen, B.; Weltens, B.; De Bot, K.; van Rosum, N. (1984). An aid in language teaching: The visualization of pitch. *System*, 12(1): S. 25–29.
- [21] Cutler, A.; Donselaar, W. v. (2001). ‚voornaam‘ is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44(2): S. 171–195.
- [22] Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11: S. 51–62.
- [23] De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behaviour. *Language and Speech*, 26(4): S. 331–350.
- [24] De Dominicis, A. (1999). *Fonologia comparata delle principali lingue europee moderne*. CLUEB, Bologna, 2. Aufl.
- [25] Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French*. Harrap, Groos, London, Heidelberg.
- [26] Delattre, P. (1966). *Studies in French and comparative phonetics: Selected papers in French and English*. Mouton, Den Haag.

- [27] Delmonte, R. (Hrsg.). (1995). *How to create 'SLIM' courseware: Software Linguistico Interattivo Multimediale*. Unipress, Padova.
- [28] Delmonte, R. (1997). Learning languages with a 'SLIM' automatic tutor. *Asiatica Veneziana*, 2: S. 31–52.
- [29] Delmonte, R. (2000). SLIM prosodic automatic tools for self-learning instruction. *Speech Communication*, 30: S. 145–166.
- [30] Delmonte, R. (2002). Apprendimento dell'intonazione tedesca e annotazione GTOBI. In: Marotta, G.; Nocchi, N. (Hrsg.), *Atti della XIII Giornata del Gruppo di Fonetica Sperimentale*, Pisa. ETS.
- [31] Delmonte, R.; Petrea, M.; Bacalu, C. (1997). SLIM prosodic module for learning activities in a foreign language. In: *Proceedings of EURO-SPEECH '97*, Band 2, S. 669–672, Rhodes.
- [32] Derwing, T. M.; Munro, M. J.; Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning*, 48: S. 393–410.
- [33] Dogil, G.; Williams, B. (1999). The phonetic manifestation of word stress. In: Hulst, H. v. d. (Hrsg.), *Word prosodic systems in the languages of Europe*, Kapitel 5, S. 273–334. Mouton de Gruyter, Berlin.
- [34] Draxler, C.; Jänsch, K. (2004). SpeechRecorder — a universal platform independent multi-channel audio recording software. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, Lisboa.
- [35] Draxler, C.; Jänsch, K. (2007). SpeechRecorder. <http://www.speechrecorder.info/>, 29.05.2007.

-
- [36] DUDEN. (2006). *Die Grammatik*. Dudenverlag, Mannheim, Leipzig, Wien und Zürich.
- [37] Dupoux, E.; Pallier, C.; Sebastián-Gallés, N.; Mehler, J. (1997). A destressing ‘deafness’ in French. *Journal of Memory and Language*, 36: S. 406–421.
- [38] Dupoux, E.; Peperkamp, S.; Sebastián-Gallés, N. (2001). A robust method to study stress ‘deafness’. *Journal of the Acoustical Society of America*, 110(3): S. 1606–1618.
- [39] Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, 2(2): S. 62–76, January. <http://llt.msu.edu/vol2num2/article3/>.
- [40] Farnetani, E.; Kori, S. (1982). Lexical stress in spoken sentences: A study on duration and vowel formant pattern. In: *Quaderni del Centro di Studi per le Ricerche di Fonetica*, Band 1, S. 104–133.
- [41] Farnetani, E.; Kori, S. (1983). Interaction of syntactic structure and rhythmical constraints on the realization of word prosody. In: *Quaderni del Centro di Studi per le Ricerche di Fonetica*, Band 2, S. 287–318.
- [42] Figge, U.; De Matteis, M. (1982). *Sprachvergleich Italienisch-Deutsch*. Oldenbourg, München.
- [43] Firth, J. R. (1948). Sounds and prosodies. In: *Transactions of the Philological Society*, S. 127–152.
- [44] Flege, J. E.; Schirru, C.; MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40: S. 467–491.

- [45] Flege, J. E. (1999). Age of learning and second language speech. In: Birdsong, D. (Hrsg.), *Second language acquisition and the critical period hypothesis*, S. 101–131. Lawrence Erlbaum, Mahwah, NJ.
- [46] Fry, D. B. (1955). Duration and intensity as a physical correlate of linguistic stress. *Journal of the Acoustical Society of America*, 27: S. 765–768.
- [47] Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1: S. 126–152.
- [48] Gass, S. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates, Mahwah, NJ.
- [49] Germain-Rutherford, A.; Martin, P. (2000). Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde. *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, 3(1): S. 61–76.
- [50] Gibbon, D. (1980). Metalocutions, structural types and functional variations in English and German. *Papers and Studies in Contrastive Linguistics*, 13: S. 17–39.
- [51] Gibbon, D. (1984). Intonation as an adaptive process. In: Gibbon, D.; Richter, H. (Hrsg.), *Intonation, Accent and Rhythm. Studies in Discourse Phonology*, S. 165–192. de Gruyter, Berlin, New York.
- [52] Goldbeck, T. P.; Sendlmeier, W. F. (1988). Wechselbeziehung zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen. In: Altmann, H. (Hrsg.), *Intonationsforschungen*.

-
- [53] Granström, B. (2004). Towards a virtual language tutor. In: Delmonte, R.; Delcloque, P.; Tonelli, S. (Hrsg.), *Proceedings of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning: NLP and Speech Technologies in Advanced Language Learning Systems*, S. 1–8, Venezia. Unipress.
- [54] Granström, B.; House, D. (2006). Measuring and modeling audiovisual prosody for animated agents. In: *Proceedings of Speech Prosody 2006*, Dresden.
- [55] Granström, B.; House, D. (2007). Inside out – acoustic and visual aspects of verbal and non-verbal communication. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, S. 11–18, Saarbrücken.
- [56] Grassegger, A. (1986). *Auditive Ähnlichkeits- und Kontrastbeziehungen zwischen dem deutschen und dem italienischen Konsonanteninventar*. Buske, Hamburg.
- [57] Gut, U. (2003). Non-native speech rhythm in German. In: *Proceedings of the XVth International Congress of Phonetic Sciences*, S. 2437–2440, Barcelona.
- [58] Gut, U. (2006). Unstressed vowels in non-native German. In: *Proceedings of Speech Prosody 2006*, Dresden.
- [59] Gut, U. (2007). Learner corpora in second language prosody research and teaching. In: Trouvain, J.; Gut, U. (Hrsg.), *Non-native prosody. Phonetic description and teaching practice*, S. 145–167. Mouton de Gruyter, Berlin, New York.
- [60] Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1): S. 34–52. <http://llt.msu.edu/vol8num1/hardison/>.

- [61] Harrington, J.; Cassidy, S. (1999). *Techniques in speech acoustics*. Kluwer Academic Publishers, Dordrecht.
- [62] Hirose, K.; Gendrin, F.; Minematsu, N. (2003). A pronunciation training system for Japanese lexical accents with corrective feedback in the learner's own voice. In: *Proceedings of EUROSPEECH 2003, Vol. 4*, S. 3149–3152, Geneva.
- [63] Hirose, K. (2004). Accent type recognition of Japanese using perceived mora pitch values and its use for pronunciation training system. In: *Proceedings of the Int. Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages (TAL)*, S. 77–80, Beijing.
- [64] Hirschfeld, U.; Trouvain, J. (2007). Teaching prosody in German as a foreign language. In: Trouvain, J.; Gut, U. (Hrsg.), *Non-native prosody. Phonetic description and teaching practice*, S. 171–187. Mouton de Gruyter, Berlin, New York.
- [65] Isačenko, A. V.; Schädlich, H. J. (1966). Untersuchungen über die deutsche Satzintonation. *Studia Grammatica*, 7: S. 7–67.
- [66] James, E. (1976). The acquisition of prosodic features using a speech visualizer. *International Review of Applied Linguistics in Language Teaching*, 14: S. 227–243.
- [67] Jessen, M.; Marasek, K.; Schneider, K.; Clahßen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. In: *Proceedings of the 13th International Congress of Phonetic Sciences*, Band 4, Stockholm.
- [68] Jin, F. (1990). *Intonation in Gesprächen. Ein Beitrag zur Methode der kontrastiven Intonationsanalyse am Beispiel des Deutschen und Französischen*. Niemeyer, Tübingen.

-
- [69] Johansson, S. (1978). Studies of error gravity: Native reactions to errors produced by Swedish learners of English. In: *Acta Universitatis Gothoburgensis, Gothenburg Studies in English 44*. Göteborg.
- [70] Jones, R. H. (1997). Beyond 'listen and repeat': Pronunciation teaching materials and theories of second language acquisition. *System*, 25(1): S. 103–112.
- [71] Jurafsky, D.; Martin, J. H. (2000). *Speech and language processing. An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall, Upper Saddle River, NJ.
- [72] Kawai, G.; Hirose, K. (1997). A CALL system using speech recognition to train the pronunciation of Japanese long vowels, the mora nasal and mora obstruent. In: *Proceedings of EUROSPEECH '97*, Band 2, S. 657–660, Rhodes.
- [73] Kohler, K. J. (Hrsg.). (1991). *Studies in German intonation*, Band 25 aus *Arbeitsberichte (AIPUK)*. Institut für Phonetik und digitale Sprachverarbeitung der Universität Kiel.
- [74] Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen*. Erich Schmidt Verlag, Berlin, 2., neubearbeitete Aufl. (1. Aufl. 1977).
- [75] Kohler, K. J. (2006). The Kiel Intonation Model (KIM), its implementation in TTS synthesis and its application to the study of spontaneous speech. Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel. <http://www.ipds.uni-kiel.de/kjk/forschung/kim.de.html>, März 2006.
- [76] Kommissarchik, J.; Kommissarchik, E. (2000). Better accent tutor — Analysis and visualization of speech prosody. In: *Proceedings of InSTILL 2000*, S. 86–89, Dundee, Scotland.

- [77] Kuhl, P. (1991). Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50: S. 90–107.
- [78] Kuhl, P. (1993). Innate predisposition and the effects of experience in speech perception: The native language magnet theory. In: De Boisson-Bardies, B.; de Schonen, S.; Jusczyk, P.; MacNeilage, P.; Morton, J. (Hrsg.), *Developmental neurocognition: Speech and face processing in the first year of life*. Kluwer Academic Publishers, Dordrecht.
- [79] Kuhl, P.; Iverson, P. (1995). Linguistic experience and the ‘perceptual magnet effect’. In: Strange, W. (Hrsg.), *Speech perception and linguistic experience. Issues in cross-language research*, S. 121–154. Timonium, York.
- [80] Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press, Cambridge.
- [81] Lane, H.; Buiten, R. (1969). A self-instructional device for conditioning accurate prosody: Trends. In: Valdman, A. (Hrsg.), *Language Teaching*, S. 159–174. Academic Press, New York.
- [82] Lehiste, I. (1970). *Suprasegmentals*. MIT Press, Cambridge, Massachusetts.
- [83] Levy, M. (1997). *Computer-Assisted Language Learning: Context and Conceptualization*. Clarendon Press, Oxford.
- [84] Logan, J. S.; Lively, S. E.; Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 89: S. 874–886.

-
- [85] Lyster, R. (1998). Negotiation of form, recasts and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning*, 48(2): S. 183–218.
- [86] MacKay, I. R.; Flege, J. E.; Piske, T.; Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, 110(1): S. 516–528.
- [87] Malagoli, G. (1946). *L'accentazione italiana: Guida pratica*. Sansoni, Firenze.
- [88] Marotta, G. (1985). *Modelli e misure ritmiche: La durata vocalica in italiano*. Zanichelli, Bologna.
- [89] Martin, P. (2004). WinPitch LTL II, a multimodal pronunciation software. In: *Proceedings of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning: NLP and Speech Technologies in Advanced Language Learning Systems*, S. 177–182, Venezia.
- [90] Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological enquiry*. Lawrence Erlbaum, Hillsdale, NJ.
- [91] Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In: Trouvain, J.; Gut, U. (Hrsg.), *Non-native prosody. Phonetic description and teaching practice*, S. 53–76. Mouton de Gruyter, Berlin, New York.
- [92] Mioni, A. M. (1976). *Fonematica delle lingue germaniche. Inglese, tedesco e neerlandese in contrasto con l'italiano*. Pátron, Bologna.
- [93] Mioni, A. M. (1993). Fonetica e fonologia. In: Sobrero, A. A. (Hrsg.), *Introduzione all'italiano contemporaneo: Le strutture*, S. 101–139. Laterza, Bari.

- [94] Missaglia, F. (1997). *Studi sul bilinguismo scolastico italo-tedesco*. La Scuola, Brescia.
- [95] Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with adult Italian learners of German. In: *Proceedings of the 14th International Congress of Phonetic Sciences*, S. 551–554.
- [96] Missaglia, F. (1999). *Phonetische Aspekte des Erwerbs von Deutsch als Fremdsprache durch italienische Muttersprachler*. Wissenschaftliche Buchhandlung Theo Hector, Frankfurt am Main.
- [97] Missaglia, F. (2001). Neue Kontrastivität: Die prosodische Wende. In: Kuri, S.; Saxer, R. (Hrsg.), *Deutsch als Fremdsprache an der Schwelle zum 21. Jahrhundert*, S. 75–95. Studien Verlag, Innsbruck, Wien, München, Bozen.
- [98] Missaglia, F. (2007). Prosodic training for adult Italian learners of German: The Contrastive Prosody Method. In: Trouvain, J.; Gut, U. (Hrsg.), *Non-native prosody. Phonetic description and teaching practice*, S. 237–258. Mouton de Gruyter, Berlin, New York.
- [99] Mooshammer, C.; Harrington, J. (2005). Linguistic prominence and loudness: A systematic comparison between lexical word stress, sentence accent and vocal effort. Vortrag auf der *Between Stress and Tone Conference*, Juni. http://www.haskins.yale.edu/staff/mooshammer/mooshammer_best_abs.pdf.
- [100] Morton, J.; Crowder, R.; Prussin, H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, 91: S. 169–190.
- [101] Morton, J.; Marcus, S.; Ottley, P. (1981). The acoustic correlates of

- speechlike: A use of the suffix effect. *Journal of Experimental Psychology*, 110: S. 568–593.
- [102] Moulines, E.; Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5/6): S. 453–467.
- [103] Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17(1): S. 17–34.
- [104] Murray, L.; Barnes, A. (1998). Beyond the ‘wow’ factor – Evaluating multimedia language learning software from a pedagogical point of view. *System*, 26: S. 249–259.
- [105] Nagano, K.; Ozawa, K. (1990). English speech training using voice conversion. In: *Proceedings of ICSLP*, S. 1169–1172, Kobe.
- [106] Neri, A.; Cucchiaroni, C.; Strik, H.; Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5): S. 441–467.
- [107] Neumeyer, L.; Franco, H.; Digalakis, V.; Weintraub, M. (2000). Automatic scoring for pronunciation quality. *Speech Communication*, 30: S. 83–93.
- [108] Nicholas, H.; Lightbown, P. M.; Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 51(4): S. 719–758.
- [109] Okada, H. (1999). Japanese. In: *Handbook of the International Phonetic Association*, S. 117–119. Cambridge University Press.
- [110] Pallier, C.; Sebastián-Gallés, N.; Colomé, A. (1999). Phonological representations and repetition priming. In: *Proceedings of EUROSPEECH '99*, S. 1907–1910, Budapest.

- [111] Pallier, C.; Sebastián-Gallés, N.; Colomé, A. (2001). The influence of native-language phonology on lexical access: Concrete exemplar-based vs. abstract lexical entries. *Psychological Science*, 12(6): S. 445–449.
- [112] Pallotti, G. (2000). *La seconda lingua*. RCS Libri, Milano.
- [113] Peperkamp, S.; Dupoux, E. (2002). A typological study of stress ‘deafness’. *Laboratory Phonology*, 7: S. 203–240.
- [114] Pfitzinger, H. (1999). Local speech rate perception in German speech. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, Band 2, S. 893–896, San Francisco.
- [115] Pfitzinger, H. (2002). 10 Years of PhonDat-II: A Reassessment. In: *Proc. ICSLP 2002*, Band 1, S. 369–372, Denver, September.
- [116] Pfitzinger, H. R. (2001). Phonetische Analyse der Sprechgeschwindigkeit. In: *Forschungsberichte (FIPKM) 38*, S. 117–264. Institut für Phonetik und Sprachliche Kommunikation der Universität München.
- [117] Pfitzinger, H. R. (2006). Five dimensions of prosody: Intensity, intonation, timing, voice quality, and degree of reduction. In: Hoffmann, R.; Mixdorff, H. (Hrsg.), *Speech Prosody Abstract Book. Studentexte zur Sprachkommunikation*, Band 40, S. 6–9. TUDpress, Dresden.
- [118] Pfitzinger, H. R. (2007). CoDIT, COmbined Discrimination/Identification Task. <http://www.phonetik.uni-muenchen.de/~hpt/#Software>, 30.11.2007.
- [119] Pfitzinger, H. R. (2007). ProFIS, PROmpts For Information Systems. <http://www.phonetik.uni-muenchen.de/~hpt/#Software>, 30.11.2007.

-
- [120] Plass, J. (1998). Design and evaluation of the user interface of foreign language multimedia software: A cognitive approach. *Language Learning & Technology*, 2(1): S. 35–45, July. <http://llt.msu.edu/vol2num1/article2/>.
- [121] Pompino-Marschall, B. (1992). PhonDat: Verbundvorhaben zum Aufbau einer Sprachsignaldatenbank. In: *Forschungsberichte (FIPKM) 30*, S. 99–128. Institut für Phonetik und Sprachliche Kommunikation der Universität München.
- [122] Rapp, S. (1995). Maschinelles Lernen von Aspekten des deutschen Wortakzents. In: *Arbeitsberichte des Instituts für Maschinelle Sprachverarbeitung*, Band 2, S. 147–240. Universität Stuttgart.
- [123] Rigault, A. A. (1962). Rôle de la fréquence, de l'intensité et de la durée vocaliques dans la perception de l'accent en français. In: *Proceedings of the V International Congress of Phonetic Sciences*, S. 735–748, Helsinki.
- [124] Roche, J. (2000). Lerntechnologie und Spracherwerb — Grundrisse einer medienadäquaten, interkulturellen Sprachdidaktik. *Zeitschrift Deutsch als Fremdsprache*, 37(3): S. 136–143.
- [125] Roche, J. (2003). Plädoyer für ein theoriebasiertes Verfahren von Software-Design und Software-Evaluation. *Zeitschrift Deutsch als Fremdsprache*, 40(2): S. 94–103.
- [126] Ronen, O.; Neumeier, L.; Franco, H. (1997). Automatic detection of mispronunciation for language instruction. In: *Proceedings of EURO-SPEECH '97*, Band 2, S. 649–652, Rhodes.
- [127] Rossi, M. (1998). Intonation in Italian. In: Hirst, D.; di Cristo, A. (Hrsg.), *Intonation systems. A survey of twenty languages*, S. 219–238. Cambridge University Press, Cambridge.

- [128] SAMPA. (2005). Speech Assessment Methods Phonetic Alphabet. <http://www.phon.ucl.ac.uk/home/sampa/>, 25.10.2005.
- [129] SAMPA Italian. (1996). Speech Assessment Methods Phonetic Alphabet for Italian. <http://www.phon.ucl.ac.uk/home/sampa/italian.htm>, 18.03.1996.
- [130] Sardelli, E.; Marotta, G. (2007). Prosodic parameters for the detection of regional varieties of Italian. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, S. 1281–1284, Saarbrücken.
- [131] Schirru, C. (1982). Analyse intonative de l'énonciation et de la question totale dans l'italien régional de Cagliari. In: *Bulletin de l'Institut de Phonétique de Grenoble*, Band 10/11, S. 169–184.
- [132] Schirru, C. (1992). Premiers éléments d'analyse prosodique contrastive entre le sarde, l'italien et le français: résultats statistiques. In: *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, Band 14, S. 79–107.
- [133] Scuffil, M. (1982). *Experiments in comparative intonation. A case study of English and German*. Niemeyer, Tübingen.
- [134] Sendlmeier, W. (1981). Der Einfluß von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. *Phonetica*, 38: S. 291–308.
- [135] Sluijter, A. M. C.; van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4): S. 2471–2485.
- [136] Tajima, K.; Port, R.; Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25: S. 1–24.

-
- [137] Tillmann, H. G. (1964). *Das phonetische Silbenproblem: Eine theoretische Untersuchung*. Dissertation, Universität Bonn.
- [138] Tillmann, H. G. (1972). Silbischer Ausprägungskode und Intonation. *Acta Universitatis Carolinae: Philologica 1, Phonetica Pragensia*, III: S. 261–265.
- [139] Tillmann, H. G.; Mansell, P. (1980). *Phonetik: Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Verlagsgemeinschaft Ernst Klett – J. G. Cotta, Stuttgart.
- [140] Tillmann, H. G.; Pfitzinger, H. R. (2004). Applying the Munich Parametric High Definition (PHD) speech synthesis system to the problem of teaching chinese tones to L1-speakers of German. In: *Proceedings of the International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages (TAL)*, S. 185–188, Beijing.
- [141] Tillmann, H. G.; Pfitzinger, H. R. (2004). The development of an advanced SLP-based system for the individual learning and fast training of speaking skills in a new foreign language. In: *Proceedings of InSTIL/ICALL2004 Symposium on Computer Assisted Language Learning: NLP and Speech Technologies in Advanced Language Learning Systems*, S. 17–20, Venezia.
- [142] Tonelli, L. (1984). Überlegungen zur natürlichen Phonologie: Kontrastive Analyse des Boznerischen und des Südtiroler Italienischen. In: *Monografie*, Band 23. Università degli Studi di Trieste.
- [143] Tonelli, L. (2002). *Regionale Umgangssprachen. Kontrastive Phonetische Untersuchungen zum Deutschen und Italienischen in Südtirol*. Unipress, Padova.

- [144] Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Vandenhoeck & Ruprecht in Göttingen, 1958, 3. Aufl.
- [145] van Dommelen, W. A. (1992). Segmentieren und Etikettieren im Kieler PhonDat-Projekt. In: *Arbeitsberichte (AIPUK) 26*, S. 197–223. Institut für Phonetik und digitale Sprachverarbeitung der Universität Kiel.
- [146] Vardanian, R. M. (1964). Teaching English through oscilloscope displays. *Language Learning*, 3–4: S. 109–117.
- [147] Vayra, M.; Avesani, C.; Fowler, C. (1987). Word-level coarticulation and shortening in Italian and English speech. *Studi di Grammatica Italiana*, 13: S. 249–269.
- [148] Vintsyuk, T. K. (1968). Speech discrimination by dynamic programming. *Cybernetics*, 4(1): S. 52–57.
- [149] Wells, J. (1997). SAMPA computer readable phonetic alphabet. In: Gibbon, D.; Moore, R.; Winski, R. (Hrsg.), *Handbook of Standards and Resources for Spoken Language Systems*. Mouton de Gruyter, Berlin, New York. Part IV, section B.
- [150] Wenk, B. J. (1983). Effets du rythme dans le français parlé. *Recherches sur le français parlé. GARS*, 5: S. 147–162.
- [151] Wenk, B. J.; Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10: S. 193–216.
- [152] WinPitch. (2005). Pitch Instruments Inc. <http://www.winpitch.com>, 31.12.2005.
- [153] Witt, S.; Young, S. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, 30: S. 95–108.

-
- [154] Wunderli, P. (1981). Kontrastive Intonationsanalyse (Deutsch-Französisch) und die zukünftigen Aufgaben der Intonationsforschung. *Zeitschrift für französische Sprache und Literatur*, 91: S. 289–322.
- [155] Zee, E. (1999). Chinese (Hong Kong Cantonese). In: *Handbook of the International Phonetic Association*, S. 58–60. Cambridge University Press.
- [156] Zuanelli Sonino, E. (1976). *Italiano — Tedesco — Inglese. Analisi contrastiva a livello fonico*. Minerva Italica, Bergamo, 2. Aufl.

Lebenslauf

Maria Paola Bissiri

Geboren am 04. Juli 1975 in Sassari (Italien)

Familienstand ledig

Staatsangehörigkeit italienisch

Ausbildung

- 1992 – 1993 Auslandsschuljahr an der Clarinda High School, Clarinda, Iowa, U.S.A., Abschluss: High School Diploma mit *highest honors*
- 1989 – 1994 Liceo Ginnasio Statale “Azuni”, Sassari, Abschluss: Diploma di Maturità Classica mit der Note *56/60*
- 1997-1998 Studium der Germanistik an der Universität Gesamthochschule Paderborn mit einem Erasmus-Stipendium
- 1994 – 1999 Studium der Lingue e Letterature Straniere an der Università degli Studi di Sassari, Abschluss: Laurea in Lingue e Letterature Straniere mit der Note *110/110 cum laude*
- 1999 – 2000 Fremdsprachenassistentin am Heisenberg-Gymnasium, Dortmund
- 2000 – 2001 Master Universitario Internazionale in Scienza e Tecnologia dei Media, Università degli Studi di Pavia, Italien, Abschluss: Diploma di Master Universitario Internazionale in Scienza e Tecnologia dei Media
- Oktober 2003 – März 2005 Promotion an der Ludwig-Maximilians Universität München
- April 2005 – März 2008 Promotion mit gemeinsamer Dissertationsbetreuung an der Ludwig-Maximilians Universität München und an der Università degli Studi di Sassari

