
Die phonetischen Grundlagen der spontanen Imitation in der gesprochenen Sprache

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie an der Ludwig-Maximilians-Universität
München

vorgelegt von

Clara Tillmanns
aus München

München, den 06.05.2013

Erstgutachter:
Zweitgutachter:
Zeitpunkt der mündlichen Prüfung:

Prof. Dr. Jonathan Harrington
PD Dr. habil. Phil Hoole
04. Juli 2013

Inhaltsverzeichnis

Abbildungsverzeichnis.....	6
Tabellenverzeichnis.....	8
Zusammenfassung.....	9
1. Einleitung.....	11
1.1 Spontane phonetische Imitation.....	11
1.1.1 Imitation und die Evolution von Sprache.....	12
1.1.2 Phonetische Imitation im Spracherwerb.....	14
1.2 Theoretische Erklärungsmodelle von spontaner phonetischer Imitation.....	18
1.2.1 Imitation in der Motortheorie der Sprachverarbeitung.....	18
1.2.2 Imitation in der direkten Perzeption	19
1.2.3 Imitation in abstrakten Modellen der Sprachverarbeitung.....	20
1.2.4 Imitation in der Exemplartheorie der Sprachverarbeitung.....	21
1.2.5 Dialogische Alinierung.....	22
1.3 Einführung der bearbeiteten Themengebiete.....	24
1.4 Verwendete Parameter der phonetischen Forschung.....	25
1.4.1 Spektrale Parametrisierung	25
1.4.2 Zeitliche Parametrisierung.....	27
2. Sprecherspezifische Perzeption und sprecherspezifische spontane phonetische Imitation	29
Zusammenfassung.....	29
2.1 Einleitung.....	29
2.2 Erfassen der potentiellen Imitatorinnen durch HörerInnen.....	36
2.2.1 Methodik.....	37
2.2.1.1 Versuchspersonen.....	37
2.2.1.2 Stimuli.....	37
2.2.1.2.1 Aufnahme der Äußerungen der Modellsprecherin.....	37
2.2.1.2.2 Aufnahme der Shadowing-Äußerungen der Sprecherinnen.....	39
2.2.1.3 Experimentelles Paradigma des Perzeptionstests.....	39
2.2.1.4 Auswertung der HörerInnenurteile.....	40
2.2.2 Ergebnis.....	40
2.2.3 Diskussion des Perzeptionsexperiments zur wahrgenommenen Ähnlichkeit	41
2.3 Erfassen der potentiellen Imitatorinnen durch akustische Analyse.....	42
2.3.1 Vergleich der ersten mit der sechsten Wiederholung in Bezug zur Modellsprecherin.....	43
2.3.1.1 Ergebnis: Grundfrequenz.....	43
2.3.1.2 Ergebnis: Vokalqualität.....	46
2.3.1.3 Ergebnis: Relative Dauer.....	47
2.3.1.4 Zusammenfassung der Ergebnisse der akustischen Analyse der ersten vs. sechsten Shadowing-Äußerung.....	48
2.3.1.5 Diskussion der Ergebnisse der akustischen Analyse Kondition 1 vs. 6.....	49
2.3.2 Akustische Trendanalyse.....	49
2.3.2.1 Ergebnis der akustischen Trendanalyse: F0	49
2.3.2.2 Ergebnis der akustischen Trendanalyse: F2	51
2.3.2.3 Diskussion der akustischen Trendanalyse.....	52

2.3.2.4 Diskussion der akustischen Analysen.....	53
2.4 Produktion-Perzeption: Identifikationstest	54
2.4.1 Methodik.....	54
2.4.1.1 Versuchspersonen.....	54
2.4.1.2 Stimuli.....	54
2.4.1.3 Experimentelles Paradigma.....	56
2.4.1.4 Durchführung.....	56
2.4.2 Ergebnisse.....	56
2.4.2.1 Neigung der Identifikationskurve.....	56
2.4.2.2 Zusammenhang des Identifikationstests mit der akustischen Analyse	57
2.4.3 Diskussion des Perzeptionstests.....	58
2.5 Abschließende Zusammenfassung und Diskussion der Ergebnisse.....	59
3. Spontane Imitation über die Position im Wort.....	63
Zusammenfassung.....	63
3.1 Einleitung.....	63
3.2 Methodik.....	70
3.2.1 Versuchspersonen.....	71
3.2.2 Stimulikonstruktion.....	71
3.2.3 Versuchsablauf.....	72
3.2.4 Auswertung.....	73
3.2.5 Ergebnisse.....	74
3.2.5.1 Vergleich der Artikulationsstellen an jeweils finaler Position	74
3.2.5.2 Vergleich der Positionen final und initial der Wörter mit /p/	75
3.3 Diskussion.....	77
4. Diskriminierbarkeit und Wahrnehmbarkeit von längerer Aspirationsdauer durch spontane phonetische Imitation.....	81
Zusammenfassung.....	81
4.1 Einleitung.....	81
4.2 Experiment 1: Akustische Analyse des Dateninventars für das Perzeptionsexperiment	85
4.2.1 Methodik.....	86
4.2.1.1 Datenreduktion.....	86
4.2.1.2 Parametrisierung und Auswertung	88
4.2.2 Ergebnisse.....	90
4.2.3 Diskussion der akustischen Ergebnisse und Erwartungen für Experiment 2.....	93
4.3 Experiment 2: Perzeptionstest.....	93
4.3.1 Methodik.....	93
4.3.1.1 Versuchspersonen.....	94
4.3.1.2 Durchführung.....	94
4.3.1.3 Auswertung.....	94
4.3.2 Ergebnisse.....	95
4.3.2.1 Diskriminationsfähigkeit.....	95
4.3.2.2 Perzeption von längerer Aspirationsdauer durch Imitation.....	96
4.4 Diskussion.....	97
5. Abschließende Diskussion.....	100

Literatur.....	104
Anhang 1: Anweisungen HörerInnen (XAXB).....	112
Anhang 2: Unterschiede der Formantfrequenzen.....	113
Anhang 3: Formantwerte der /ɛ/-/a/-Manipulation.....	114
Anhang 4: Wörter der Lese- und Hörliste.....	115
Anhang 5: Erklärungen zum Experiment.....	116

Abbildungsverzeichnis

Abbildung 1: A: Matrix aus den kindlichen Vokalisierungen getrennt nach Vokalklassifikation der Äußerung (Infant Utterance) und präsentiertem Stimulus (Stimulus Condition). Die höheren Zahlen in den diagonalen Zellen unterstützen die Hypothese der Imitation. B: Matrix aus den teilnehmenden Kindern je nach häufigster Vokalisierungsart (Subject Classification) und präsentiertem Stimulus (Stimulus Condition). Die höheren Zahlen in den diagonalen Zellen unterstützen die Hypothese der Imitation. (Darstellung: Kuhl & Meltzoff, 1996).....	17
Abbildung 2: Oszillogramm und Spektrogramm der Äußerung "Statik". Die Segmentierung markiert die Segmente /a/, /t/ und /l/, innerhalb des /t/ befindet sich der Verschluss, an der gestrichelten Linie die Verschlusslösung (Burst), gefolgt von der Aspiration. (Screenshot aus Praat, Boersma, 2001).....	26
Abbildung 3: Richtig/falsch-Verteilung der HörerInnenurteile pro Sprecherin. Richtig bewertet wird ein Durchgang, wenn die sechste Äußerung als die ähnlichere im Vergleich zur Modellsprecherin wahrgenommen wurde - falsch, wenn die erste Äußerung als ähnlicher wahrgenommen wurde.....	41
Abbildung 4: Die Parameter F1, F2, F3 und F1-F0 der Modellsprache jeweils für /ɛ/ („E“) und /a/ („a“) im Vergleich zwischen den im Experiment verwendeten Stimuli ("hyp") und den gleichen Wörtern normal produziert ("norm"). Die Einheit der x-Achse ist jeweils Hz.	43
Abbildung 5: Grundfrequenz [Hz] der Sprecherinnen pro Kondition (x-Achse: 1 steht für die Kondition 1 und ist im Boxplot weiß, 6 steht für die Kondition 6 und ist im Boxplot grau markiert) und der Modellsprecherin (Modell). Die Linie liegt auf der Höhe des Medianwerts der Modellsprecherin.....	44
Abbildung 6: Differenz der Distanz von F0 [Hz] der ersten Äußerung und der Distanz der letzten Äußerung des gleichen Worts der Sprecherinnen zur Modellsprecherin. Grau markiert sind die Werte der zwei Sprecherinnen deren F0-Werte sich signifikant an die der Modellsprecherin annähern. Die gezogene Linie markiert Null (Veränderung zwischen den Konditionen).	45
Abbildung 7: F2 [Hz] von /a/ der Modellsprecherin (Modell, Linie markiert den Medianwert) und der zehn Sprecherinnen. 1. bezieht sich auf die erste Shadowing-Wiederholung, 6. auf die sechste Shadowing-Wiederholung (jeweils grau markiert).	46
Abbildung 8: Differenz der Distanz [Hz] zwischen den ersten Shadowing-Äußerungen und der Modellsprecherin abzüglich der Distanz zwischen den letzten Shadowing-Äußerungen und der Modellsprecherin für F2 in /a/. Grau markiert sind die signifikant über Null liegenden Werteverteilungen der Sprecherinnen EE und PS. Die Linie markiert Null.....	47
Abbildung 9: Relative Vokaldauer: die Vokaldauer des betonten Vokals normalisiert auf die absolute Wortlänge für die Sprecherinnen und die Modellsprecherin (Modell). Grau markiert sind die späteren Shadowing-Äußerungen, die Linie markiert den Medianwert der Modellsprecherin.....	48
Abbildung 10: Alle sechs Wiederholungen des Worts „Festmahl“ der Sprecherin MH und die Regressionslinie, die den Neigungswert b beinhaltet.....	50

Abbildung 11: Absolute F0-Werte (in Hz) der Shadowing-Wiederholungen 1 bis 6 und der Modellsprecherin.	50
Abbildung 12: y-Achse: Neigung der F0-Werte von der ersten bis zur sechsten Wiederholung ($b=cov/var$). Grau markiert sind die signifikant von Null abweichenden Verteilungen. Die Linie markiert Null.....	51
Abbildung 13: F2-Werte [Hz] des Vokals /a/ (zeitlich mittlere 20 %) aller sechs Wiederholungen.....	51
Abbildung 14: y-Achse: Neigung der F2-Werte von der ersten bis zur sechsten Wiederholung ($b=cov/var$). Grau markiert ist die signifikant von Null abweichende Verteilung der Sprecherin EE. Die Linie markiert Null.....	52
Abbildung 15: Formantverläufe des der Manipulation zugrunde liegenden /ε/-Vokals (schwarze Punkte in beiden Darstellungen) sowie der Vokale des Stimulus 1 (links: /ε/) und 12 (rechts: /a/) für das Kontinuum von /vɛlt/ zu /valt/ (graue Punkte). Stimulus 1 weist nur sehr leichte Formantveränderungen aufgrund der Resynthese auf, Stimulus 12 hat einen gesenkten F2 und dynamisch erhöhten F1.....	55
Abbildung 16: Identifikationskurve der Sprecherin AN (schwarze Punkte) im Vergleich zu AM (graue Punkte) über einem Kontinuum von /ε/ (1. Stimulus) zu /a/ (12. Stimulus).....	56
Abbildung 17: Neigung der Identifikationskurven (m: y-Achse) pro Sprecherin (x-Achse).	57
Abbildung 18: m-Werte des Identifikationstests im Zusammenhang mit den Sprecherinnen-Gruppen „Imitatorinnen“ vs. „Nicht-Imitatorinnen“ aus der akustischen Analyse.....	58
Abbildung 19: Darstellung der untersuchten Wortgruppen und die Vergleichsgruppen (Kreise) für die statistische Auswertung.....	73
Abbildung 20: Die Abbildung zeigt die absolute Aspirationsdauer der finalen Plosive /p/ und /k/ in Millisekunden. Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme, "B" die zweite Aufnahme markiert.....	74
Abbildung 21: Die Abbildung zeigt die absolute Aspirationsdauer von /p/ initial und final in Millisekunden. Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme, "B" die zweite Aufnahme markiert....	74
Abbildung 22: Die Abbildung zeigt den RMS-Wert (Root Mean Square) der initialen und finalen Plosive /p/ in dB. Von links nach rechts: die durchschnittlichen RMS-Werte des Segments /p/, die durchschnittliche RMS-Werte des gesamten Worts und die relativen RMS-Werte (RMS-Wert des Segments dividiert durch RMS-Wert des Worts). Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme und "B" die zweite Aufnahme markiert.....	76
Abbildung 23: Aspirationsdauer (in ms) von finalelem /p/ in einsilbigen Wörtern. In „gesamt“ befinden sich 17 Wörter, in „Auswahl“ sieben der 17 Wörter, jeweils für die Aufnahmekonditionen A und B. Die ausgewählten sieben Wörter werden als repräsentative Teilmenge aller 17 Wörter akzeptiert.....	88

Abbildung 24: Beispielwort mit finalelem /p/: Depp; c im Mittelteil der Äußerung markiert den Verschluss (closure) des Plosivs.....	89
Abbildung 25: Beispielwort mit initialem /p/: Putz.....	90
Abbildung 26: Darstellung der Dauerdifferenzen (x-Achse) gegenüber der proportionalen Erkennungsleistung der HörerInnen (x-Achse) mit Regressionslinien. Die gestrichelte Linie markiert die Erkennung zu 50%, d.h. Zufall.....	96
Abbildung 27: Verteilung der richtigen ("TRUE") und falschen ("FALSE") Erkennung der längeren Aspirationsdauer.....	97
Abbildung 28: Verteilung der Äußerungskonditionen innerhalb der perzeptiven Urteile getrennt für /p/ an finaler und initialer Position. "A" gibt die Äußerungen der Kondition A (vor den manipulierten Stimuli) an, B die Kondition nach dem Hören der manipulierten Stimuli.....	97

Tabellenverzeichnis

Tabelle 1: Die zur Bewertung von phonetischem Talent verwendeten Aufgaben aus Jilka (2009a). Alle Aufgaben beinhalten prosodische, nur die dunkelgrau hinterlegten, auch segmentelle Aspekte.....	35
Tabelle 2: Ziel- und Füllwörter der Shadowing-Aufgabe.....	38
Tabelle 3: Signifikanztest auf den Unterschied zwischen den F0-Werten der Modellsprecherin und den F0-Werten der ersten Wiederholung der Sprecherinnen im Shadowing-Experiment.....	44
Tabelle 4: Sprecherinnen, die eine signifikante akustische Annäherung an die Modellsprecherin zeigen (nach akustischem Parameter und analytischem Vorgehen).....	53
Tabelle 5: Neigung der Identifikationskurven der Sprecherinnen. Die fett markierten Werte beziehen sich auf die (in der akustischen Auswertung) als potentielle Imitatorinnen identifizierten Sprecherinnen.....	57
Tabelle 6: Akustische Messergebnisse der Dauerveränderung von VOT in Experimenten zur spontanen phonetischen Imitation.....	84
Tabelle 7: Unterschied zwischen den Werten der Aspirationsdauer der A-Aufnahmen und den B-Aufnahmen (B-A) im Durchschnitt über alle SprecherInnen, aufsteigend sortiert..	87
Tabelle 8: Mittelwert der Verteilung (Tabelle 7) subtrahiert von den Differenzwerten (Tabelle 7), aufsteigend sortiert. Im schwarzen Kasten sind die ausgewählten Wörter markiert.....	87
Tabelle 9: Wörter der akustischen Analyse und des Perceptionsexperiments.....	88
Tabelle 10: Veränderung der Dauern der Segmente bzw. Segmentverbindungen von der ersten Aufnahme (A) zur zweiten Aufnahme (B) aller sieben Wörter. „Länger“ bezeichnet das Längerwerden von Aufnahmezeitpunkt A zu B, d.h. in B sind die Segmente länger..	92

Zusammenfassung

Diese Arbeit behandelt spontane phonetische Imitation, das heißt die Annäherung der Sprachproduktion an die phonetischen Eigenschaften von umgebender Sprache. Innerhalb dreier Studien wurden zwei Produktionsexperimente mit unterschiedlichen Versuchsabläufen durchgeführt, um spontane phonetische Imitation zu erhalten. Die resultierenden Aufnahmen wurden jeweils perzeptiv und akustisch ausgewertet, wobei sich teilweise gegensätzliche Ergebnisse in Bezug auf das Vorhandensein von Imitation ergaben. Mit der perzeptiven Analyse lässt sich beantworten ob wahrnehmbare Imitation stattgefunden hat, wohingegen die akustische Analyse offenlegt ob und anhand welcher akustischer Parameter Imitation aufgetreten ist.

In der ersten Studie wurden Sprecherunterschiede in der Imitation perzeptiv und akustisch ermittelt und mit den Ergebnissen eines Identifikationstests in Verbindung gesetzt. Dabei zeigte sich, dass (1.) keine vollständige Übereinstimmung zwischen der perzeptiven und der akustischen Ermittlung von Imitation bestand und (2.) die Sprecherinnen, deren Produktionen keine akustische Annäherung enthielten, in einem Identifikationstest eine steilere Kategoriengrenze zwischen zwei Vokalen aufwiesen. Das Ergebnis wird als Hinweis darauf betrachtet, dass individuelle Unterschiede in der Sprachverarbeitung bestehen und durch Imitation sichtbar gemacht werden können.

Die zweite Studie testete die Generalisierung eines zeitlichen Lautmerkmals (Aspiration) über die Position im Wort (initial versus final) und die Artikulationsstelle (bilabial versus velar) durch Imitation. Dafür wurde ein Experiment, bestehend aus einem Leseblock, einem Hörblock und der Wiederholung des Leseblocks, durchgeführt (Nielsen, 2011b). Dieser Aufbau ermöglichte, dass das zu imitierende Merkmal nur an einer Position im Wort wahrgenommen wurde (initial), während es in der Leseliste an unterschiedlichen Positionen und Artikulationsstellen vorkam. Die gefundene Imitation trat, entgegen der Hypothese, stärker an der neuen Position im Wort (final) auf als an der alten (initial), nicht jedoch, wie erwartet, an der neuen Position mit zusätzlicher Änderung der Artikulationsstelle. Dieses Ergebnis legt nahe, dass (1.) die im Lexikon gespeicherte, subphonemische Information nicht an die Position im Wort gebunden ist und (2.) eine übereinstimmende Position im Wort nicht in allen Fällen ein begünstigender Faktor für Imitation sein muss.

Ein Teil der Äußerungen der zweiten Studie wurde für das Perzeptionsexperiment der **dritten Studie** verwendet, um die Wahrnehmbarkeit von akustischen Dauerunterschieden in der Aspiration zu untersuchen. Im Perzeptionstest verglichen HörerInnen die zwei Äußerungen eines Wortes (vor und nach der Wahrnehmung von sprachlichen Stimuli) anhand der Dauer von /p/ an entweder initialer oder finaler Position im Wort. Die

Auswertung ergab, dass Versuchspersonen potentiell durch Imitation entstandene Dauerunterschiede differenzieren konnten. Dieses Ergebnis kann als Hinweis gewertet werden, dass spontane phonetische Imitation an perzeptivem Lernen und der Ausbreitung von Lautwandel beteiligt sein kann. Darüber hinaus wurde von den HörerInnen gleich viel Imitation an initialer und finaler Position entdeckt, obwohl die akustische Analyse Positionsunterschiede ergab. Diese Ergebnisse legen nahe, dass (1.) die aus spontaner phonetischer Imitation resultierende Variation weiter imitiert werden kann und (2.) eine Diskrepanz zwischen wahrgenommener und akustisch gemessener Imitation besteht.

Zusammenfassend deuten die Ergebnisse der Studien darauf hin, dass

- sprecherspezifische Unterschiede in der Perzeption und Imitation bestehen und korrelieren,
- ein flexibler Zusammenhang zwischen Lautmerkmalen und der Wortposition im Lexikon besteht,
- Merkmalsunterschiede, die auf spontaner phonetischer Imitation beruhen, wahrnehmbar sind und potentiell weiter imitiert werden können und
- eine Diskrepanz zwischen messbarer und wahrnehmbarer Imitation besteht.

1. Einleitung

In den letzten Jahren haben die Ergebnisse einiger Studien gezeigt, dass sich die phonetischen Eigenschaften der Produktionen von Versuchspersonen ändern, abhängig davon, mit welchen Äußerungen anderer SprecherInnen sie in Kontakt gekommen sind (Babel, 2012; Delvaux & Soquet, 2007; Goldinger, 1998; Nielsen, 2011b; Pardo, 2006; Pardo, Cajori Jay & Kraus, 2010). SprecherInnen klingen anders, nachdem sie Sprache gehört haben, die anders klingt. Dieses Phänomen wird in dieser Arbeit als spontane phonetische Imitation bezeichnet und erforscht.

Im Speziellen handelt es sich hier um Ergebnisse, die durch deren experimentelle Untersuchung gewonnen wurden. Diese geben Informationen über sprecherspezifische Sprachverarbeitung (Abschnitt 2), den Aufbau des mentalen Lexikons (Abschnitt 3) und die Wahrnehmbarkeit von geringen phonetischen Veränderungen und die daraus folgenden Implikationen für Lautwandel (Abschnitt 4).

Einleitend wird die Beteiligung spontaner Imitation auf die onto- und phylogenetische Entwicklung von Sprache sowie die (für diese Arbeit) wichtigsten Theorien der Sprachverarbeitung dargestellt.

1.1 Spontane phonetische Imitation

Der Begriff Imitation wird auch alltagssprachlich verwendet und beschreibt die Nachahmung von Verhalten oder Reproduktion von Erzeugnissen. Diese Arbeit beschäftigt sich ausschließlich mit phonetischer Imitation und in den überwiegenden Fällen mit der sogenannten „spontanen phonetischen Imitation“. Dabei handelt es sich um die Angleichung der phonetischen Eigenschaften der eigenen Sprache an die wahrgenommener Sprache (z.B. Babel, 2011; Goldinger, 1998). Die Bezeichnung „spontan“ dient der Abgrenzung gegenüber Verfahren, in denen Versuchspersonen ausdrücklich aufgefordert werden, Stimuli zu imitieren, das heißt Äußerungen möglichst ähnlich zu einem Vorbild zu produzieren (wie z.B. in Reiterer, Singh & Winkler, 2012). Darüber hinaus handelt es sich bei der spontanen Imitation um ein Phänomen, welches innerhalb relativ kurzer Zeit zu beobachten ist. Das heißt spontane phonetische Imitation ist bereits in Experimentdurchgängen von ca. einer Stunde zu beobachten (z.B. Babel, 2010, 2012; Nielsen, 2011b; Pardo, 2006).

Die Angleichung kann qualitative und/oder zeitliche, phonemische oder (in der verwendeten Sprache) nicht-bedeutungsunterscheidende Aspekte gesprochener Sprache betreffen. Beispielsweise fand Babel (2012) unterschiedlich starke Imitation abhängig vom Vokal (*vowel selectivity*). Dafür wertete sie die spektralen Eigenschaften der Aufnahmen aus. Nielsen (2011b) hingegen untersuchte die Dauer der Aspiration (siehe 1.3), welche

ein zeitlicher Parameter ist, der in der verwendeten Ausprägung nicht bedeutungsunterscheidend war, aber dennoch imitiert wurde.

1.1.1 Imitation und die Evolution von Sprache

Für die Evolution von Sprache hat Imitation eine besondere Bedeutung: "Certainly, we cannot doubt that imitation was critical to the early evolution of language. Without it there could have been no shared vocabulary of learned meaning, and so no basis for the evolution of propositional utterance and syntax" (Studdert-Kennedy, 2000, S. 168). Darüber hinaus seien auch andere Beschaffenheiten die präadaptiv¹ für z.B. Syntax, die Abstrahierung von Ort und Zeit (*displaced reference*) oder die kulturelle Entwicklung von Schrift, implizit Teil des Imitationsprozesses.

Die Dissoziation von Geräusch und Bedeutung und damit die Entwicklung von Segmenten wurde durch sprachliche (mündliche) Imitation hervorgerufen. Für den Prozess des Imitierens spielt die Bedeutung einer Äußerung keine Rolle. Notwendig sind aber (mindestens) die folgenden drei Schritte: (1.) Analyse des Geräuschereignisses in die zugrunde liegenden artikulatorischen Komponenten, (2.) Speichern der analysierten Strukturen für eine kürzere oder längere Zeit, abhängig vom Zeitverlauf zwischen Vorbild und Kopie und (3.) Wiederherstellung der Komponenten in ihrer korrekten zeitlichen Abfolge (Studdert-Kennedy, 2000). Aus der Forschung von Primaten und Menschenaffen ist bekannt, dass durch Spiegelneuronen (Rizzolatti, Fadiga, Gallese & Fogassi, 1996) Aktionen von Artgenossen imitiert werden. Diese Imitation muss nicht offen geschehen. Spezifisch für die Wirkungsweise von Spiegelneuronen ist, dass diese, während eine Aktivität beobachtet wird, für eine neuronale Aktivierung sorgen, die der Aktivierung während der aktiven Ausführung gleicht. Eine solche neuronale Aktivität wird verborgene Imitation genannt und kann zu einem späteren Zeitpunkt offene Imitation hervorrufen. Kreisläufe mit Spiegelneuronen scheinen bei Menschen umfassender zu sein als bei Affen. Ob dieser Unterschied zwischen den Spezien tatsächlich besteht oder nur auf die unterschiedlichen Forschungstechniken zurückzuführen ist, ist noch nicht geklärt. Ein möglicher Unterschied zwischen Affen und Menschen könnte in den observierten Handlungsweisen liegen: Während bei Affen nur dann neuronale Aktivierung gefunden wurde, wenn die Tätigkeit ein Objekt beinhaltete, ließ sich neuronale Aktivierung im menschlichen Gehirn auch während der Observation von Bewegungen ohne Objekt zeigen (Rizzolatti, Craighero & Fadiga, 2002).

Bedeutung mag also ein Motivator für Imitation sein, ist jedoch keine Voraussetzung. Imitation reduziert die Vorbildäußerung auf perzeptiv-motorische Bestandteile. Deshalb

¹ Eigenschaft einer Struktur, die aus sich ändernden Umweltbedingungen folgt und sich durch weitere Evolution zu neuen Strukturen entwickeln kann.

kann Imitation zur Dissoziation von Struktur (geräuschliche Eigenschaften) und Funktion (Bedeutung) beigetragen haben.

Die Schritte, in denen sich die Artikulatoren gebildet haben, sind nicht bekannt. Die eingeschränkten Möglichkeiten des artikulatorischen Apparates hatten die Entstehung der Geste als variierbare Einheit zur Folge, an der (in unterschiedlichen Stellungen) alle sechs Artikulatoren mindestens passiv beteiligt sind (Mattingly, 1991). Eine Geste ist, der artikulatorischen Phonologie zufolge (Browman & Goldstein, 1992), die elementare Einheit von phonologischen Kontrasten, die gleichzeitig als abstrakte Beschreibung von artikulatorischen Ereignissen dienen. Dabei bestehen Äußerungen aus organisierten Mustern von Gesten, welche einander auch zeitlich überlappen können (Studdert-Kennedy, 2000).

Durch vielfache Kombination der gleichen Gesten in unterschiedlichen Kontexten konnte ein großer Bestand an Wörtern oder formelhaften („*formulaic*“, Studdert-Kennedy, 2000, S. 170) Äußerungen gebildet werden. Die Imitation der Gesten in verschiedenen Kontexten brachte wiederholende Gestenverbindungen hervor. Auch durch Geschwindigkeit und Bewegungsökonomie konnten diese dann als Element einer motorischen Einheit, unabhängig vom Kontext, erkannt werden (Lindblom, 1992, 1998; Studdert-Kennedy, 1987, 1991). Durch die entstandenen wiederkehrenden Gesten und Gestenverbindungen (d.h. Segmente), welche das Lexikon der Sprechenden durchzogen, entstand eine neue Repräsentationsebene zwischen dem Signal und der Bedeutung (Studdert-Kennedy, 2000).

Aufgrund der Annahme, dass die neue Repräsentationsebene durch Spezialisierung der Artikulatoren entstanden ist, wurden die Elemente des neu entstandenen Levels unter motorischen Bedingungen dargestellt (Studdert-Kennedy, 1998). Dennoch liegt es an den akustischen und optischen Effekten der Artikulation, dass Hörende die Gesten von Sprechenden wiedererkennen können. Die dazwischenliegende Repräsentation muss deshalb sowohl in Bezug zur Bewegung als auch zur Perzeption von gleicher Gestalt und gleichzeitig auch spezifisch für Sprache, also phonetisch, gewesen sein. Die Konsequenzen des neuen unabhängigen und abstrakten phonetischen Levels, welches durch Evolution von vokalischer Imitation herbeigeführt wurde, entwickelte sich weiter zu dem, was dann Sprache wurde (Studdert-Kennedy, 2000).

In einer Studie von Horner und Whiten (2005) beobachteten in Wildnis geborene Schimpansen und drei- bis vierjährige Kinder einen erwachsenen Menschen, der mit Hilfe eines Werkzeugs eine Belohnung aus einer Puzzleschachtel holte. Dafür wurden sowohl notwendige als auch unnötige Handlungen verwendet, in zwei Versuchskonditionen war die Puzzleschachtel einmal blickdicht und einmal durchsichtig. Es zeigte sich, dass bei Verwendung der blickdichten Schachtel sowohl die Schimpansen als auch die Kinder alle

Handlungen imitierten, um selbst die Belohnung zu erhalten. Bei Verwendung der durchsichtigen Schachtel imitierten die Schimpansen nur die zielführenden Aktionen, die Kinder imitierten wieder alle. Diese Vorgehensweise wird Überimitation genannt (Arbib, 2011). Obwohl es den Anschein hat, dass Überimitation ein Resultat schlechter Anpassung ist, argumentieren Nielsen und Tomaselli (2009) dafür, dass Überimitation eine evolutionäre Anpassung darstellt, die notwendig für die Entwicklung und Verbreitung der menschlichen Kultur ist. Wenn Kinder Erwachsene beobachten, wie sie eine Tätigkeit ausführen, verbinden sie alle Aspekte der Handlung mit dem Ziel, d.h. egal wie einsichtig ein einfaches physikalisches System auch ist, sie werden einen kausalen Zusammenhang zwischen den Handlungen und dem Ziel herstellen. Dies kann einen Einblick in die Entwicklung der großen kulturellen Unterschiede bei alltäglichen Handgriffen geben: Wenn nicht zunehmend zielführend imitiert wird und wurde, sondern unnötige Variation aufrecht erhalten, kann das erklären warum völlig unterschiedliche Herangehensweisen bestehen blieben und bleiben; in dieser Darstellung ist kulturelle Variation eine Konsequenz, keine Ursache für imitatorische Fähigkeiten (Arbib, 2011). Diese Betrachtung deckt sich mit Studdert-Kennedys (2000) Argumentation (s.o.), dass vokalische Imitation nicht durch Bedeutung motiviert sein musste. Das exakte Ziel der Handlung (Verwendung von Werkzeugen oder Stimmfähigkeit) beeinflusst nicht direkt die imitatorische Auswirkung.

Die Ergebnisse von Studien zur Evolution von Sprache weisen (unter Einbeziehung neuronaler Zusammenhänge) darauf hin, dass Imitation nicht (ausschließlich) zielführende Handlungen betrifft/betrifft und Einblicke und Erklärung für eine große Zahl von Phänomenen geben kann: die mögliche Existenz von vokalischer Imitation bei Protosprachen, obwohl die semantische Zuordnung mitunter sehr schwierig gewesen sein kann; die Dissoziation von Geräusch und Bedeutung und somit die Entstehung der dazwischenliegenden phonetischen Ebene; die bis heute bestehende große Variation zwischen den Sprachen der Welt; und die unwillkürliche Beeinflussung der Produktion durch umgebende sprachliche Äußerungen.

1.1.2 Phonetische Imitation im Spracherwerb

In einer frühen Studie über imitatorisches Verhalten von Kleinstkindern konnten Meltzoff und Moore (1977) zeigen, dass bereits zwei bis drei Wochen alte Säuglinge Bewegungen der Lippen und der Zunge imitieren. Darüber hinaus konnten die Säuglinge unterscheiden, ob die Lippen oder die Zunge nach vorne gestreckt wurden und auch welche Bewegung das gleiche Körperteil ausführte. D.h. die Imitation war sensibel sowohl für das Körperteil als auch für die Art der Bewegung. Da die teilnehmenden Babys dieser frühen Studie in den ersten Wochen ihres Lebens bereits visuelle Erfahrung mit

Mundbewegungen der Bindungspersonen gemacht hatten, wurde die Studie mit Kleinstkindern im Alter von durchschnittlich 32 Stunden wiederholt (Meltzoff & Moore, 1983, 1989). Auch die Ergebnisse dieser Studie zeigten, dass menschliche Neugeborene Bewegungen des Gesichts, im Speziellen der Lippen und der Zunge, imitieren. Besonders hervorzuheben ist, dass die Säuglinge ihre eigenen Bewegungen nicht sehen, sondern sie nur über die eigene Körperwahrnehmung erfahren können. Diese Form der Imitation wird als Beweis für eine neurale Verbindung zwischen beobachteter und ausgeführter Bewegung in menschlichen Säuglingen angesehen (Meltzoff, 2002), die so auch am phonetischen Spracherwerb beteiligt sein kann.

Im Folgenden wird der phonetische Spracherwerb (im Gegensatz zum gesamten linguistischen Spracherwerb) betrachtet und im Speziellen konkrete experimentelle Ergebnisse, die die Rolle von Imitation dabei aufzeigen. Eine besondere Fokussierung fällt auf den Bereich des (ca.) ersten Lebensjahres, während dessen der erste Erwerb stattfindet – im Gegensatz zur darauf folgenden Entwicklung.

Innerhalb dieses ca. ersten Jahres werden fünf Phasen des phonetischen Spracherwerbs unterschieden (Kuhl & Meltzoff, 1996):

- *Reflexive Phonation* (0-2 Monate), in der sich die stimmlichen Ereignisse auf vegetative und reflexive Produktionen wie Husten, Niesen und Schreien beschränken;
- *Cooing* (1-4 Monate), in der quasivokalische Laute erzeugt werden, die Vokalen ähneln;
- *Expansion* (3-8 Monate), klare Vokale mit voller Resonanz und einer Reihe neuer Vokalisationen wie Rufen, Flüstern oder Seufzen;
- *Canonical Babbling* (5-10 Monate), wobei einfache Konsonant-Vokal-Verbindungen wiederholt werden (häufig Lallphase genannt: z.B. „bababa“, „mamama“), gefolgt von
- *Meaningful Speech* (10-18 Monate), wobei Kinder durch Lallen und bedeutungstragende Äußerungen längere, intonierte Äußerungen bilden.

Roug, Landberg & Lundberg (1989) beziehen sich in den ersten Phasen stärker auf die beteiligten Artikulatoren und benennen innerhalb der ersten sieben Monate die glottale, velar/uvulare und die vokalische Stufe des phonetischen Spracherwerbs (in dieser Reihenfolge) und konzentrieren sich so auf eine noch engere phonetische Betrachtung.

Lieberman, Ryallis & Rabson (1982) analysierten Äußerungen einer Mutter und ihres Säuglings. In den Intonationskonturen des Säuglings wurden bereits in der fünften Woche akustische Charakteristika der Grundfrequenz der Mutter in Höhe, Verlauf und Dauer reproduziert. Darüber hinaus passte sich die Mutter in der Tonhöhe dem Kind an, indem sie ihre Grundfrequenz auf durchschnittlich 325 Hz (von sonst 185 Hz) erhöhte; 325 Hz ist

eine Tonhöhe, die sich im Bereich derer des Kindes befindet und segmentelle Imitation erleichtern könnte. Etwa in der 11. Woche begann das Kind spontan vokalische Laute zu produzieren und die Vokale der Mutter zu imitieren. Die kindlichen Vokale waren keine akustischen Nachahmungen der mütterlichen Formantfrequenzen, sondern die Formantfrequenzen entsprachen dem kleineren supralaryngalen Vokaltrakt des Kindes. Es wurde also eine „Übersetzung“ der mütterlichen Vokale auf den Vokalraum des Kindes festgestellt (Lieberman, Ryallis & Rabson, 1982).

Kuhl und Meltzoff (1996) zeigen phonetische (vokalische) Imitation von Kindern im Alter von 12 bis 20 Wochen. Methodisch ist hervorzuheben, dass es sich um eine maschinelle Präsentation handelte, d.h. Imitation war zu beobachten, ohne dass wechselseitige Kommunikation bestand, sodass wechselseitige Imitation, wie in Lieberman, Ryallis & Rabson (1982), ausgeschlossen war. In ihrem Experiment wurden drei Versuchspersonengruppen (im Alter von 12, 16 und 20 Wochen) mit einer Video- und Audiodarstellung einer Sprecherin konfrontiert, die einen der amerikanischen Vokale /a/, /i/ oder /u/ (in der Silbe /hVd/) produzierte. Jedes teilnehmende Kind sah und hörte an drei Tagen (innerhalb höchstens einer Woche) für jeweils fünf Minuten die Produktion eines dieser Vokale. Die stimmlichen Äußerungen der Kinder während der Präsentation wurden aufgenommen und perzeptiv durch geschulte HörerInnen, die Erfahrung mit den Vokalisierung von Säuglingen und Kleinkindern hatten, für die Untersuchung von Imitation ausgewertet. Die Zuordnung der kindlichen Äußerungen zu Vokalkategorien geschah in mehreren Schritten: Die Äußerungen wurden einer der amerikanischen Vokalkategorien zugeordnet (/i/, /I/, /ɛ/, /æ/, /a/, /ʌ/, /ʊ/, /u/), diese dann als /a/-, /i/- oder /u/-ähnlich zusammengefasst (/a/:/æ/, /a/, /ʌ/; /i/: /i/, /I/, /ɛ/; /u/: /ʊ/, /u/) und die Klassifikation zu den drei zusammenfassenden Kategorien durch eine weitere phonetisch geschulte Person durchgeführt. Die zwei perzeptiven Vokalklassifizierungen stimmten zu 86 % überein. Für die statistische Auswertung der Imitation wurden nur diese 86 % der Äußerungen verwendet.

Die Auswertung beantwortet die Frage, ob Kinder die Äußerungen, die sie produzieren, systematisch abhängig vom Stimulus, den sie hören, variieren. Hierfür wurden zum einen alle 224 verwendbaren Äußerungen in einer Matrix mit den Stimulusbedingungen dargestellt (Abbildung 1A). Ohne Beeinflussung des Stimulus sollte in jeder Zeile einer Zeile die gleiche Anzahl Äußerungen stehen. Dies ist nicht der Fall, d.h. diese Darstellung bietet bereits einen guten Hinweis darauf, dass die Qualität der kindlichen Vokalisierungen nicht unabhängig vom präsentierten Stimulus war. Da jedoch zwischen den Zellen dieser Matrix keine statistische Unabhängigkeit besteht, wurden die Daten zusätzlich nach Kind klassifiziert: Hierfür wurde jedes Kind der Vokalkategorie zugeordnet, die es am häufigsten geäußert hat. Ein Kind, das beispielsweise sieben /u/-ähnliche, drei /a/-

ähnliche und zwei /i/-ähnliche Äußerungen produziert hat, wird als /u/-Kind markiert. Die Gegenüberstellung der klassifizierten Kinder und der Stimulusbedingungen erlaubt statistische Unabhängigkeit und resultiert in der Matrix in Abbildung 1B. Wenn die Stimulusbedingung keinen Einfluss auf die Qualität der Äußerungen der Kinder (und damit auf deren Vokalklassifizierung) gehabt hätte, müssten alle Zellen einer Zeile in der Matrix ungefähr gleich viele Einträge enthalten. Die Verteilungen der Einträge (in diesem Fall Kinder) erreicht sowohl im Gesamten (als Chi-Quadrat Test der 3x3 Kontingenz) als auch pro Zeile Signifikanz.

		Stimulus Condition			
		/a/	/i/	/u/	
Infant Utterances	/a/	66 .63	25 .24	14 .13	105 1.00
	/i/	11 .30	22 .59	4 .11	37 1.00
	/u/	18 .22	20 .24	44 .54	82 1.00

		Stimulus Condition			
		/a/	/i/	/u/	
Subject Classification	/a/	13 .68	3 .16	3 .16	19 1.00
	/i/	1 .11	8 .89	0 .00	9 1.00
	/u/	2 .12	4 .23	11 .65	17 1.00

Abbildung 1: **A:** Matrix aus den kindlichen Vokalisierungen getrennt nach Vokalklassifikation der Äußerung (Infant Utterance) und präsentiertem Stimulus (Stimulus Condition). Die höheren Zahlen in den diagonalen Zellen unterstützen die Hypothese der Imitation. **B:** Matrix aus den teilnehmenden Kindern je nach häufigster Vokalisierungsart (Subject Classification) und präsentiertem Stimulus (Stimulus Condition). Die höheren Zahlen in den diagonalen Zellen unterstützen die Hypothese der Imitation. (Darstellung: Kuhl & Meltzoff, 1996)

Diese Studie demonstriert eindrucksvoll phonetische Imitation im Alter von 12-20 Wochen. Statistisch signifikant unterscheidet sich die zusammengefasste Altersgruppe der 12 und 16 Wochen alten Kinder gegenüber der Kinder im Alter von 20 Wochen. Kuhl und Meltzoff arbeiteten hier mit Vokalisationen, die Vokalkategorien zugeordnet wurden. Sie zeigen, dass die Präsentation von Vokalen eine Häufung von Vokalisierungen hervorrief, die den präsentierten Vokalen ähnlich waren. Diese Beeinflussung der perceptiv qualitativen Eigenschaften der Vokalisationen bezeichneten sie als phonetische Imitation.

Nielsen (2011a) testete die Imitation von 6-jährigen Schulkindern, um die Mechanismen zu erforschen, die dem Erwerb der Einheiten der gesprochenen Sprache (wie Phoneme und deren Reize) zugrunde liegen. Ihre Fragestellung war also nicht nur, ob Schulkindern imitieren, sondern welche Details imitiert werden und welche Implikationen

für die Entwicklung von phonologischen Repräsentationen daraus abgeleitet werden können. Sie wertete die Veränderungen in den kindlichen Äußerungen aus, die bevor und nachdem die Kinder Sprache wahrgenommen hatten, die in feinen phonetischen Details (Aspirationsdauer; siehe Abschnitt 1.4) manipuliert war, bestanden. Wie sie bei erwachsenen SprecherInnen zeigen konnte (Nielsen, 2011b), imitieren auch Kinder im Alter von sechs Jahren verlängerte Aspirationsdauer und generalisieren diese zudem auf neue Wörter und Phoneme. Dass sublexikalische Details bei Schulkindern an Imitation beteiligt sind, deutet darauf hin, dass durch den Spracherwerb in diesem Alter bereits phonologische Repräsentationen gebildet wurden, die auch sublexikalische Informationen enthalten.

1.2 Theoretische Erklärungsmodelle von spontaner phonetischer Imitation

Im Folgenden werden die drei Theorien der Sprachverarbeitung, die für diese Arbeit am wichtigsten sind, vorgestellt. Es handelt sich um die Motortheorie, abstrakte Modelle und Exemplar-Modelle der Sprachverarbeitung. Diese unterscheiden sich anhand der Architektur und der Einheiten, die als notwendig für die Perzeption von Sprache angenommen werden.

1.2.1 Imitation in der Motortheorie der Sprachverarbeitung

Besonders an der Motortheorie ist die enge Verlinkung von Perzeption und Produktion. Zwei der Hauptprinzipien der Motortheorie besagen, dass Sprachwahrnehmung (1.) auf der Grundlage von phonetischen Gesten geschieht und (2.) Zugriff auf das motorische System beinhaltet (Lieberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). Eine Geste ist hier (im Gegensatz zu Gesten im Bereich der artikulatorischen Phonologie, siehe Abschnitt 1.1.1) der Zusammenschluss aller Bewegungen der beteiligten Artikulatoren zu einer koordinierten Aktion, die ein linguistisches Ziel erreicht (Kelso, Vatikiotis-Bateson, Tuller & Fowler, 1984). Innerhalb der Motortheorie ist eine solche Geste eine abstrakte Einheit, denn sie beschreibt die Intention eines Sprechers, nicht die tatsächlichen Bewegungen. Diese Unterscheidung ist notwendig, um den theoretischen Unterschied zwischen Phonologie und Phonetik der gesprochenen Sprache zu verdeutlichen.

Dennoch bietet diese Verbindung zwischen Produktion und Perzeption eine Grundlage für eine mögliche Wechselwirkung und damit für Imitation: Wenn intendierte Gesten die Einheiten der Perzeption sind, können diese auf die folgende Produktion wirken. Eine mögliche Auswirkung von Perzeption auf die Produktion wird durch folgende Experimente und ihre Ergebnisse anschaulich gezeigt:

In zwei Experimenten wurde unter Verwendung von transkranieller Magnetstimulation Muskelaktivität innerhalb sprechrelevanter Muskeln gemessen, während Versuchspersonen Sprache wahrnahmen. Fadiga, Craighero, Buccino & Rizzolatti (2002) fanden Muskelaktivität in der Zunge, wenn Versuchspersonen Äußerungen hörten, die linguale Konsonanten beinhalteten. Watkins, Strafella & Paus (2003) fanden verstärkte Muskelaktivität in den Lippen sowohl während Sprache gehört als auch während sprechspezifische Lippenbewegungen beobachtet wurden. Diese Ergebnisse demonstrieren, dass das Hören von Sprache eine Aktivierung der Sprachproduktionszentren hervorruft, die sogar phonemspezifisch ist.

1.2.2 Imitation in der direkten Perzeption

Fowler (1986) entwickelt ihre Theorie der direkten Sprachwahrnehmung (*direct perception*) aus der Theorie der direkten Wahrnehmung von Sinnesreizen (Gibson, 1966, 1979), die die Funktion aller Sinnesorgane, wie z.B. Haptik oder das Sehen, erklären will. Der Grundgedanke liegt hier dabei, dass für Sinneswahrnehmungen Informationsmedien genutzt werden: am Beispiel von visuellen Wahrnehmungen ist das Ereignis der Umgebung ein sichtbares Objekt oder Ereignis, welches durch die Strukturierung von Licht als ein Objekt oder Ereignis wahrgenommen wird. Entsprechend funktioniert die Haptik in dieser Theorie: wenn ein Objekt berührt wird, verändert dieses die Form der Hautoberfläche womit diese Deformation der Haut zum Informationsmedium wird. Wahrgenommen wird nicht die veränderte Form der Haut, sondern die Form und Oberfläche des Objekts selbst. Für die Sprachwahrnehmung bedeutet dieser Ansatz, dass im Falle eines Klangereignisses die Druckveränderungen der Luft, d.h. das akustische Signal, das Informationsmedium darstellt. Wahrgenommen wird aber das Klangereignis selbst, zum Beispiel aufeinander treffende Objekte oder eine phonetische Geste (Fowler, 1996).

Die Theorie der direkten Wahrnehmung besagt folglich, im Gegensatz zur Motortheorie, dass sich Sprachwahrnehmung und die Wahrnehmung nicht-sprachlicher Klangereignisse nicht unterscheiden. Darüber hinaus bedarf es für die Wahrnehmung von phonetischen Gesten, im Sinne der direkten Perzeption, nicht die Konsultation des motorischen Systems. Gemeinsam haben diese zwei Theorien nur, dass nicht die akustischen Eigenschaften einer Äußerung als perzeptive Einheit betrachtet werden, sondern die linguistischen (Fowler, 1996). In der direkten Perzeption werden die Gesten des Vokaltraktes selbst als zugrunde liegende Einheit perzipiert, die keine „Übersetzung“ durch das motorische System benötigen.

Im Bereich der phonetischen Imitation bietet diese Theorie der Sprachwahrnehmung, wie auch die Motortheorie, den Ansatz, dass wenn die phonetische Geste, d.h. der artikulatorische Bewegungsablauf wahrgenommen wird, diese Wahrnehmung eine Auswirkung auf folgende Produktionen haben könnte.

Eine Studie von Mitterer und Ernestus (2008) maß in einer *Shadowing*-Aufgabe Reaktionszeiten, in der die zwei /r/-Allophone [r] (alveolarer Trill) und [R] (uvularer Trill) verwendet wurden. Die meisten niederländischen Sprecher beherrschen nur einen dieser /r/-Trills (Mitterer & Ernestus, 2008). Eine direkte Perzeption von Gesten würde erwarten, dass Versuchspersonen kürzere Reaktionszeiten aufweisen wenn der Shadowing-Stimulus die /r/-Variante beinhaltet, die die Versuchsperson gewöhnlich und im Shadowing verwendet und entsprechend längere Reaktionszeiten, wenn sich die /r/-Realisierung zwischen Stimulus und Versuchspersonenäußerung unterscheidet. Diese Unterscheidung der experimentellen Faktoren wurde nicht gefunden: der Unterschied in den Gesten zwischen Stimulus und bevorzugter Variante führt nicht zu einer verlängerten Reaktionszeit. Nur zwei der 18 Versuchspersonen in diesem Experiment imitierten den Artikulationsort der Stimuli in mehr als acht Durchgängen (d.h. in mehr als 5% der Durchgänge). Innerhalb dieser zwei Versuchspersonen führte die Imitation bei nur einer zu einer verlängerten Reaktionszeit; die Reaktionszeiten der anderen Versuchsperson unterschieden sich nicht, je nachdem ob die ungewohnte Geste imitiert oder die gewohnte verwendet wurde. Zusammengefasst können diese Ergebnisse die Theorie der direkten Perzeption nicht unterstützen und sprechen für eine weniger feste Verbindung zwischen Sprachwahrnehmung und -produktion.

1.2.3 Imitation in abstrakten Modellen der Sprachverarbeitung

Abstrakte Modelle der Sprachwahrnehmung zeichnen sich dadurch aus, dass das akustische, phonetische Signal entweder als Menge abstrakter phonetischer Reize oder abstrakter Phoneme kodiert wird, welche im Lexikon festgelegt sind und weder über die Zeit noch die Sprechgeschwindigkeit, den phonologischen Kontext oder den Sprecher variieren. Folglich ignorieren abstrakte Modelle zeitliche und kontextuelle Bestandteile des Signals (Luce & McLennan, 2005). Für die Produktion von Sprache besagen abstrakte Modelle, dass diese ausgehend von der abstrakten Repräsentation gebildet wird. Damit besteht keine Beeinflussung durch vorhergehende Perzepte. Die Imitation von Eigenschaften wahrgenommener Sprache kann im Rahmen dieser Modelle nur dann erfolgen, wenn kein lexikalischer Zugriff besteht oder möglich ist, wenn also keine abstrakte Repräsentation gefunden wird und deshalb keine Kodierung erfolgt. Jedoch bestehen in abstrakten Modellen die Mechanismen der Aktivierung und der Konkurrenz: Wenn eine Äußerung wahrgenommen wird, bewirkt dies die sofortige Aktivierung der

möglichen passenden Lexikoneinträge. Die Menge der Einträge bewirkt Konkurrenz, die umso größer ist, je mehr Einträge gefunden werden (*similarity neighborhood density*) und je häufiger die Einträge in gesprochener Sprache auftauchen (*frequency*) (Luce & McLennan, 2005). Das Umgehen der Aktivierung möglicher passender Lexikoneinträge ist in den abstrakten Modellen nicht vorgesehen. Folglich kann eine sprachliche Äußerung nur dann produziert werden, wenn eine abstrakte Repräsentation, d.h. ein Lexikoneintrag besteht. Somit kann Imitation von lexikalisch nicht-relevanten Bestandteilen eines Signals mit abstrakten Modellen der Sprachwahrnehmung nicht erklärt werden (Imitation von lexikalisch nicht-relevanten Bestandteilen des Sprachsignals: Babel, 2012; Nielsen, 2011a, 2011b; Goldinger, 1998; Pardo, 2006). In abstrakten Modellen kann Imitation nur zur Entwicklung des Lexikons genutzt werden, da lediglich lexikalisch relevante Bestandteile im Gedächtnis gespeichert werden.

1.2.4 Imitation in der Exemplartheorie der Sprachverarbeitung

Exemplarbasierte Modelle der Sprachverarbeitung bezeichnen Exemplare als die perzeptive Einheit. Wahrnehmung geschieht über den Abgleich eines aktuellen Ereignisses mit allen gespeicherten. Es bestehen also weder abstrakte phonologische Einheiten, die zu bestimmen sind, wie in abstrakten Modellen der Sprachverarbeitung, noch wird eine „Übersetzung“ in die intendierten artikulatorischen Gesten vollzogen. Der Abgleich ermöglicht, dass alle Details des Sprachsignals, wie z.B. Sprecherunterschiede oder phonologisch irrelevante Variation, für die aktuelle Wahrnehmung zur Verfügung stehen und für spätere Wahrnehmungen genutzt werden können (Johnson, 1997). Z.B. ist die Worterkennungsrate höher, wenn die Stimmen der Sprechenden bekannt sind (Church & Schacter, 1994; Goldinger, 1996, 1997). Das deutet darauf hin, dass sprecherspezifische Variation gespeichert und zur Worterkennung genutzt wird (Johnson, 2005).

Für die Erklärung von spontaner phonetischer Imitation ist diese Theorie gut geeignet: Wenn die Details im Sprachsignal wahrgenommen und gespeichert werden, können sie einen Einfluss auf die Produktion haben.

Gegen rein exemplarbasierte Modelle ohne eine Abstraktion sprechen die Ergebnisse von McQueen, Cutler und Norris (2006). Hier wurden Versuchspersonen mit ambigen /f/-/s/-Lauten an lexikalisch eindeutigen Stellen konfrontiert (z.B. [wɪtɫɔ?] in dem niederländischen Wort *witlof*, welches kein /f/-/s/-Minimalpaar ist). Dies führte zu einer Veränderung der Kategoriengrenze zwischen den Frikativen /f/ und /s/: Wenn der ambige Laut /f/s ersetzt wurde, ordneten Versuchspersonen in einem Vokalkontinuum mehr Laute der Kategorie /f/ zu, als vor dem perzeptiven Training (und andersherum, wenn /s/ ersetzt wurde). Dieses perzeptive Lernen wurde auch auf andere Wörter, die nicht Teil des

Trainingsmaterials waren, generalisiert. Die Generalisierung, so McQueen et al. (2006), deutet darauf hin, dass die Veränderung der Kategoriengrenze an einer prelexikalischen Ebene geschieht und dies das Lernen von abstrakten sublexikalischen Repräsentationen zeigt.

1.2.5 Dialogische Alinierung

Das Dialogmodell von Pickering und Garrod (2004, 2006) ist wichtig für die Untersuchung spontaner phonetischer Imitation, weil es das Konzept der Alinierung einführt. Alinierung bezeichnet eine Einigung bzw. Angleichung zwischen DialogpartnerInnen bezüglich der Wortwahl, Syntax, Sprechgeschwindigkeit usw. So wird eine enge Verbindung zwischen Produktion und Perzeption im Dialog hergestellt, die Auswirkungen auf die phonetische Umsetzung von Äußerungen haben kann.

Pickering und Garrod (2004) schlugen dieses mechanistische Dialogmodell vor, um Modelle der monologischen Sprachverarbeitung (z.B. Levelt, 1989) zu erweitern. Bei den Modellen der monologischen Sprachverarbeitung werden DialogpartnerInnen getrennt voneinander abgebildet, sie greifen auf keinen gemeinsamen „Raum“ zu und die einzige Verbindung ist die Information, die die Äußerung selbst liefert (Cherry, 1956). Dem entgegen handelt es sich bei dem Modell von Pickering und Garrod (2004) um eine Modellierung der mechanistischen Abläufe während Dialogen. Diese zeichnen sich dadurch aus, dass während des dialogischen Prozesses verschiedene Ebenen zwischen den Sprechenden aliniert werden. Dies betrifft alle Ebenen der produktiven und perceptiven Sprachverarbeitung, das heißt (mindestens) Repräsentationen von Semantik, Syntax, Lexikon und Phonologie wie auch das Situationsmodell, welches sich durch Information zu Ort, Zeit, Ursache, Intention und Information zu gegenwärtig relevanten Details auszeichnet (Zwaan & Radvansky, 1998). Alinierung bedeutet hier, dass zwischen den Sprechenden Präferenzen für z.B. Wörter (bzw. Wortbedeutungen) oder Syntaxkonstruktionen festgelegt werden. Das Situationsmodell ist eine nicht-linguistische Repräsentation und beinhaltet Einigungen wie z.B. wo, bei gegenüber platzierten Interakteuren, rechts ist oder welche Person genau gemeint ist, wenn nur der Vorname verwendet wird. Solche Festlegungen ersparen während der dialogischen Kommunikation die wiederholte kontextabhängige Bedeutungsklä rung (bei ambiger Semantik) und erleichtern die Äußerungsplanung (bei z.B. häufig verwendeten Satzkonstruktionen).

Alinierung geschieht nach Pickering und Garrod (2004, 2005) automatisch durch *Priming*, Perkolation und Reparaturmechanismen. *Priming* bezeichnet die Erhöhung der Auftretenswahrscheinlichkeit: Wenn ein bestimmtes Wort grundsätzlich eine eher geringe Auftretenswahrscheinlichkeit hat, also selten verwendet wird, steigt die Wahrscheinlichkeit im Dialog stark, sobald dieses Wort von einer/einem Dialogteilnehmenden verwendet

wird. Das erhöhte Auftreten führt zur Entstehung von sogenannten Routinen. Diese sind Wörter oder Phrasen, die innerhalb des Dialogs häufig vorkommen und deren Bedeutung zwischen den DialogpartnerInnen eindeutig sind.

Mit Perkolation wird der Mechanismus bezeichnet, der dafür sorgt, dass Alinierung innerhalb einer Ebene die Alinierung innerhalb einer anderen Ebene begünstigt. Beispielsweise wird es nachweislich wahrscheinlicher, dass die gleiche Satzkonstruktion verwendet wird, nachdem die gleichen Wörter verwendet wurden (Branigan, Pickering & Cleland, 2000).

Die von Pickering und Garrod beschriebenen Reparaturmechanismen beziehen sich auf die semantischen Dialoginhalte. Gemeinsame Routinen können nur durch erfolgreiche Alinierung entstehen. Wenn z.B. ein Sprecher einem anderen eine Aufgabe erklärt, der andere aber nicht kongruent zur Vorstellung des ersten antwortet, kann ersterer erkennen, dass seine Äußerung eine unterschiedliche Repräsentation hervorgerufen hat. Solange sich die Repräsentationen unterscheiden, besteht keine Alinierung. Die Sprechenden werden durch z.B. Nachfragen oder Umformulieren Alinierung anstreben.

Der mechanistische Ansatz bietet grundsätzlich eine gute Modellvorstellung, um die Prozesse im Zusammenhang mit spontaner phonetischer Imitation beschreiben zu können. Gemeinsames Wissen bezüglich der Situation und Alinierung auf verschiedenen Ebenen der Sprachverarbeitung öffnen den theoretischen Raum der Sprachverarbeitung für stärkere Interaktionen als den expliziten Informationsaustausch während der sprachlichen Interaktion. Auch stärkt das Modell die Vorstellung, dass spontane phonetische Imitation ein nicht-bewusster Prozess ist; zumindest dann, wenn spontane phonetische Imitation mit phonetischer Alinierung gleichgesetzt wird. Wie automatisch Alinierung geschieht, präzisieren Pickering und Garrod in ihrer Veröffentlichung von 2006: „The main mechanism of alignment is automatic and primarily unconscious [...]“ (S. 204). Dennoch ist die Alinierung im Dialogmodell von Pickering und Garrod (2004, 2007) keine ausreichende Erklärung für spontane phonetische Imitation, da die beobachtbare Imitation als Beeinflussung auftritt, d.h. eine ausgerichtete Änderung darstellt und nie zu einer vollständigen Angleichung zwischen Gesprächspartnern führt, weder in dialogischen noch in nicht-sozialen Situationen (Babel, 2010; Goldinger, 1998; Pardo, 2006; Pardo, Cajori Jay & Kraus, 2010). Aus diesem Grund kann Alinierung eine gute Erklärung für Anpassungsprozesse während eines Dialogs bei kategorialen Entscheidungen wie Wortbedeutung oder Syntax sein, jedoch nicht bei graduellen Prozessen wie phonetischer Imitation und Konvergenz (Pardo, 2012b).

1.3 Einführung der bearbeiteten Themengebiete

Die vorliegende Arbeit behandelt sprecherspezifische Unterschiede in der Imitationsstärke und in der perzeptiven Sprachverarbeitung (Abschnitt 2) einerseits und generalisierende Imitation eines Reizes und deren Wahrnehmbarkeit (Abschnitt 3 und 4) andererseits. An dieser Stelle wird eine Zusammenfassung der Motivation der Themengebiete gegeben. Eine ausführliche Einleitung folgt in den angegebenen Abschnitten.

Sprecherspezifische Unterschiede wurden in der Sprachproduktion anhand vielfältiger Parameter beobachtet; so fanden Peterson und Barney (1952) Unterschiede in den Formantfrequenzen nicht nur zwischen Frauen, Männern und Kindern sondern auch innerhalb dieser eher homogenen Sprechergruppen. Solche Unterschiede bestehen auch bei der Zentroidfrequenz und der Schiefe der Friktion bei /s/ und /ʃ/ (Newman, Clouse & Burnham, 2001) und im zeitlichen Bereich für den Parameter VOT (*voice-onset-time*; Allen, Miller & DeSteno, 2003). Sprecherspezifische Unterschiede bestehen auch in der Perzeption bei der Diskrimination von Mono- und Diphtongen (Fox, 1983) und der Funktionalisierung von Frequenz- und Dauerunterschieden bei der Diskrimination von Vokalen und Frikativen (Makashay, 2003). Darüber hinaus besteht auch individuelle Übereinstimmung von Produktion und perzeptiver Prototypisierung: SprecherInnen, die eher lange VOTs produzieren, markieren Äußerungen mit eher langen VOTs als prototypischere Exemplare im Gegensatz zu Äußerungen mit eher kurzen VOTs (Newman, 2003). Eine L2-Dialogstudie fand stärkere Konvergenz bei den SprecherInnen, die in einer vorausgegangenen Testbatterie als phonetisch talentierter eingestuft wurden (Lewandoski, 2010). Aufgrund dieser Ergebnisse und der häufigen Beobachtung, dass nicht alle SprecherInnen imitieren, bzw. nicht alle SprecherInnen gleich stark imitieren (Babel, 2010, 2012; Goldinger, 1998; Nielsen, 2011b) wird das Themengebiet der „Sprecherspezifischen Imitation“ in Abschnitt 2 bearbeitet.

Experimentelle Arbeit im Bereich der Perzeption, speziell dem perzeptiven Lernen haben gezeigt, dass die Ebenen des mentalen Lexikons flexible Verbindungen untereinander aufweisen. Diese besteht mindestens sublexikalisch (McQueen, Norris & Cutler, 2006) und subphonemisch (Kraljic & Samuel, 2006). Diese Flexibilität ist auch in der Produktion sichtbar: durch spontane Imitation werden Charakteristika wahrgenommener Sprache auf andere Wörter (sublexikalisch) und andere Phoneme (subphonemisch) generalisiert (Nielsen, 2011). Weitere Einblicke in die Systematik dieser Generalisierung und die daraus ableitbare Beschaffenheit der Verbindungen zwischen den Ebenen des mentalen Lexikons werden im Abschnitt 3 bearbeitet.

Da spontane Imitation häufig als Mechanismus beschrieben wird, durch den sich Lautwandel und Dialektkonvergenz ausbreiten können (Lindblom, 1998; Trudgill, 2008;

Babel, 2011), findet in Abschnitt 4 die Wahrnehmbarkeit von imitatorischen phonetischen Veränderungen Betrachtung.

1.4 Verwendete Parameter der phonetischen Forschung

In den drei in dieser Arbeit dargestellten Studien wurden akustische Parameter von Sprachaufnahmen spektral (anhand ihrer Frequenzverteilung) und temporal (anhand zeitlicher Teilung) ausgewertet. Die spektralen Eigenschaften sind die ersten zwei Formantfrequenzen (F1 und F2) und die Grundfrequenz (F0). Aspiration wurde zeitlich analysiert.

1.4.1 Spektrale Parametrisierung

Artikulatorisch unterscheiden sich Vokale anhand der Vokaltraktform, die durch die Zungen-, Kieferstellung und Lippenrundung beeinflusst wird. Akustisch ist diese Vokaltraktform anhand charakteristischer Resonanzen (Formanten) im Vokalschall messbar. Zunge und Gaumen bilden eine Verengung:

- je mehr freier Raum zwischen Gaumen und Zunge besteht, das heißt je tiefer die Zunge (möglicherweise verstärkt durch den gesenkten Kiefer) liegt, um so höher liegen die Werte des ersten Resonanzmaximums (F1); ist die Verengung zwischen Zunge und Gaumen sehr schmal, ist F1 sehr klein.
- je näher die schmalste Verengung zwischen Zunge und Gaumen an der Lippenöffnung liegt, umso höher liegen die Werte des zweiten Resonanzmaximums (F2; Lippenrundung verlängert den Vokalraum, sodass eine Verengung durch das Hinzufügen von Lippenrundung akustisch weiter hinten erscheint) (Pompino-Marschall, 2003).

Es besteht also ein Zusammenhang zwischen der Vokaltraktform, die durch die Zungenposition, Kieferstellung und Lippenrundung definiert wird, und der Höhe der Formanten.

Abbildung 2 zeigt das Oszillogramm, das Spektrogramm und die Segmentation (von oben nach unten) einer Äußerung des Wortes „Statik“. Die Formanten 1 und 2 der Vokale /a/ und // unterscheiden sich sehr stark: /a/ wird offen artikuliert, das heißt der Kiefer ist geöffnet und die Zunge senkt sich mit dem Kiefer. // ist ein hoher Vokal, die Zunge befindet sich relativ zu /a/ sehr viel weiter oben im Mundraum, das heißt es besteht eine stärkere Verengung zwischen Zunge und Gaumen als bei /a/. Daraus ergeben sich ein relativ tiefer F1 für // und ein relativ hoher F1 für /a/. Die Verengung ist bei // relativ nah an der Lippenöffnung, bei /a/ aufgrund des dickeren Zungenrückens mittig bis hinten. Deshalb ist F2 für // höher als für /a/.

Durch die Auswertung der Resonanzfrequenzen können Rückschlüsse auf die artikulatorischen Verhältnisse des Vokalraums gezogen werden und somit Vokale auf der Grundlage von empirischen akustischen Datensammlungen miteinander verglichen werden.

Bei der Imitation von Vokalen tritt Vokalselektivität auf: tiefe Vokale (/æ a/) weisen mehr Imitation auf als hohe Vokale (/i o u/; Babel, 2012). Das liegt an den größeren artikulatorischen Möglichkeiten, die in tieferen Vokalen möglich sind, insbesondere durch die mehr oder weniger weite Kieferöffnung. Solche Unterschiede treten speziell zwischen betonten und unbetonten Vokalen auf.

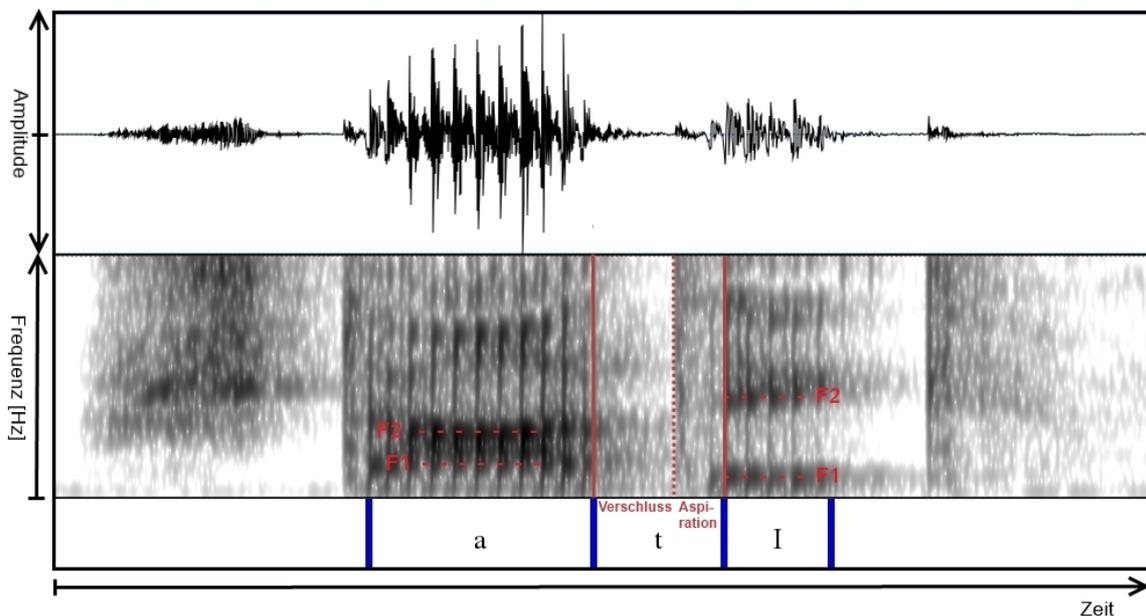


Abbildung 2: Oszillogramm und Spektrogramm der Äußerung "Statik". Die Segmentierung markiert die Segmente /a/, /t/ und /I/, innerhalb des /t/ befindet sich der Verschluss, an der gestrichelten Linie die Verschlusslösung (Burst), gefolgt von der Aspiration. (Screenshot aus Praat, Boersma, 2001).

Die Grundfrequenz (F0) ist im Spektrogramm in Abbildung 2 nicht direkt sichtbar, aber ableitbar: innerhalb der markierten Bereiche der Vokale /a/ und /I/ sind die Glottisschläge als vertikale Linien sichtbar, die die Vokale zu stimmhaften Elementen des Sprachsignals machen. Die Häufigkeit der Schläge gibt Auskunft über die Höhe der Grundfrequenz: je mehr Schläge pro Zeiteinheit erfolgen, um so höher ist die Grundfrequenz (Einheit Hertz). Im Bereich von /a/ folgen die Glottisschläge schneller aufeinander als im Bereich von /I/. Folglich liegt die Grundfrequenz in /a/ höher als in /I/ (in diesem Beispiel).

Der Begriff „Grundfrequenz“ bezieht sich auf den physikalischen Reiz. Wahrgenommen wird die Grundfrequenz als Tonhöhe, die in Äußerungen jedoch nicht konstant ist. Wichtig für die Auswertungen in diesen Studien ist die mittlere Sprechstimmlage, die den mittleren

Tonhöhenwert bezeichnet, um den sich die Melodie der normalen Sprechstimme auf und ab bewegt (Schultz-Coulon, 1975). Berechnet wird die mittlere Sprechstimmlage in dieser Arbeit als Durchschnitt der F0-Werte einer gesamten Äußerung.

Grundsätzlich tritt in hohen Vokalen eine höhere Grundfrequenz (F0) auf als in tiefen Vokalen, auch wenn andere Einflüsse konstant gehalten werden. Dieses Phänomen ist als intrinsische F0 (*intrinsic F0*) bekannt. Entsprechend beschreibt der Begriff „intrinsische Tonhöhe“ die perzeptive Größe: Bei gleicher F0 in /i/ (einem hohen) und /a/ (einem tiefen Vokal) wird die Tonhöhe in /i/ tiefer wahrgenommen als in /a/, da die intrinsische Tonhöhe normalisiert wird (Pape & Mooshammer, 2006). Im Gegensatz zur intrinsischen Tonhöhe gibt es einige Studien, die Imitation der Grundfrequenz im Sinne der mittleren Sprechstimmlage gefunden haben (Babel & Bulatov, 2012; Kappes, Baumgaertner, Peschke & Ziegler, 2009).

1.4.2 Zeitliche Parametrisierung

Der letzte Parameter, der in dieser Arbeit verwendet wird, um gesprochene Sprache empirisch vergleichbar zu machen, ist ein zeitlicher Parameter, der die Dauer des abschließenden Plosiv-Segments beschreibt. Hierfür gibt es einige, in dieser Arbeit synonym verwendete Begriffe, die jedoch je nach Sprache unterschiedliche Intervalle bezeichnen können. Es handelt sich um die Aspiration(sdauer) bzw. die VOT (*voice-onset-time*). Der Bereich /t/ in der Abbildung 2 lässt sich in drei Segmente teilen: Verschlussphase (während der die Zunge einen Verschluss mit dem Gaumen bildet), *Burst* (während der Verschlusslösung von Zunge und Gaumen) und Aspiration (wenn nach der Verschlusslösung Luft durch den Kehlkopf strömt, bevor die Glottisschläge für // beginnen). Die deutschen stimmlosen und stimmhaften Plosive unterscheiden sich (unter anderem) anhand der Dauer der Aspiration. Diese ist bei stimmlosen Plosiven länger als bei stimmhaften, weshalb die Bezeichnung Fortis- und Lenis-Plosive die artikulatorischen und akustischen Eigenschaften besser widerspiegeln. Die VOT bezeichnet den Zeitraum zwischen der Verschlusslösung und dem Beginn der Stimmlippenschwingung; in einigen Sprachen (beispielsweise dem Französischen) treten während des Verschlusses von stimmhaften Plosiven Stimmlippenschwingungen auf, was diese Plosive stimmhaft macht (im Gegensatz zu den in der Regel entstimmten Plosiven des Deutschen). Wenn Stimmhaftigkeit im Verschluss auftritt, ist die VOT negativ, denn sie beginnt vor der Verschlusslösung. Die Dauer der möglicherweise folgenden Aspiration ist im VOT-Wert nicht enthalten (Pompino-Marschall, 2003). Im Deutschen ist die VOT (in der Regel) äquivalent zur Aspirationsdauer, weshalb die Begriffe in dieser Arbeit synonym genutzt werden.

Innerhalb der Imitationsforschung wurden bereits verschiedene zeitliche Parameter verwendet. Zum einen war verlängerte VOT in mehreren Studien von Nielsen (2011a, 2011b) Auslöser von Imitation. Auch Shockley, Sabadini und Fowler (2004) zeigten Imitation von verlängerter VOT. Zum anderen haben Delvaux und Soquet (2007) die Dauern von Vokalen verglichen und so Imitation zwischen zwei belgischen Dialekten gefunden.

Im Folgenden (Abschnitt 2 bis 4) werden drei Studien zu spontaner phonetischer Imitation beschrieben und in Abschnitt 5 abschließend diskutiert.

2. Sprecherspezifische Perzeption und sprecherspezifische spontane phonetische Imitation

Zusammenfassung

In dieser Studie wurden Sprecherunterschiede bei Imitation anhand perzeptiver Einschätzung und akustischer Analyse betrachtet und mit den Ergebnissen eines Identifikationstests in Verbindung gesetzt. Einleitend werden die Themengebiete Imitation, Verbindung zwischen Perzeption und Produktion und Sprecherunterschiede in der Perzeption und der Produktion von Vokalen besprochen. Anschließend werden die Erfassung von Imitation durch HörerInnen, die akustische Analyse sowie die Durchführung und das Ergebnis des Identifikationstests vorgestellt. Im letzten methodischen Teil werden die Zusammenhänge zwischen den Ergebnissen der Produktion und der Perzeption aufgezeigt. Abschließend wird das Ergebnis diskutiert, dass die SprecherInnen, deren Produktionen keine akustische Annäherung enthielten, eine steilere Kategoriengrenze zwischen zwei Vokalen aufwiesen.

2.1 Einleitung

Von spontaner phonetischer Imitation wird dann gesprochen, wenn eine Angleichung der phonetischen Eigenschaften an die der wahrgenommenen Sprache festgestellt werden kann, ohne dass zu einer solchen Angleichung explizit aufgefordert wurde (z.B. Babel, 2011; Goldinger, 1998). Die Bezeichnung „spontan“ dient der Abgrenzung gegenüber Verfahren, in denen Versuchspersonen ausdrücklich aufgefordert werden, Stimuli zu imitieren (z.B. Reiterer, Singh & Winkler, 2012).

Mit dieser Studie soll geprüft werden, ob sich SprecherInnen in der Stärke der spontanen phonetischen Imitation unterscheiden, auch wenn bekannte Einflussfaktoren konstant gehalten werden. Im Anschluss soll anhand eines Identifikationstests über einem Kontinuum zwischen den zwei Vokalen /ɛ/ und /a/ geprüft werden, ob ein Zusammenhang zwischen den sprecherspezifischen Unterschieden im Grad der spontanen phonetischen Imitation und sprecherspezifischen Unterschieden in der Perzeption besteht.

Spontane phonetische Imitation wird nicht nur als Phänomen untersucht, sondern auch als Instrument für Dialektkonvergenz und Lautwandel betrachtet (Babel, 2010; Trudgill, 2004). Sie gilt als nicht-bewusster, jedoch auch nicht automatischer und unbeeinflussbarer Mechanismus, der phonetische Eigenschaften von Sprache im Erwachsenenalter, mindestens kurzfristig, verändern kann (Babel, 2010; Pardo, 2012b).

Mögliche Einflüsse auf spontane phonetische Imitation sind z.B. phonetische Eigenschaften der verwendeten Äußerungen oder Bestandteile der experimentellen Aufgabe. Mehrfach wurde gezeigt, dass Wörter, die in einer Sprache häufiger vorkommen, schwächere Imitation hervorrufen (Goldinger, 1998; Nielsen, 2011b). Diese

negative Abhängigkeit trifft auch auf die Phonotaktik zu (Nye & Fowler, 2003): Realisierungen von Wörtern mit phonotaktisch häufigen Lautkombinationen werden weniger stark imitiert als Realisierungen von Wörtern mit phonotaktisch seltenen Lautkombinationen. Darüber hinaus fand Babel (2012) zwei spezifischere Ergebnisse: Zum einen werden tiefe Vokale stärker imitiert als hohe, was Babel auf die größere Variabilität von tiefen Vokalen im Vergleich zu hohen Vokalen, speziell im Öffnungsgrad (F1), zurückführt, je nachdem ob der Vokal in betonter oder unbetonter Position auftritt. Zum zweiten zeigt die Autorin, dass sozio-kognitive Bedingungen die Stärke von spontaner Imitation insofern beeinflussen, als eine negative Einstellung gegenüber dem/r Sprecher/in zu schwächerer Imitation führt. Sie zeigt auch, dass diese Abneigung tief verankert sein muss (Babel, 2010): Wurde der Sprecher im Versuchsaufbau unmittelbar vor dem Experiment durch negative Informationen abgewertet, hatte dies keinen Einfluss auf die Imitationsstärke.

Gemessen wird Imitationsstärke in diesen Studien entweder anhand akustischer Analyse von Formantwerten oder VOT (Babel, 2010, 2011, 2012; Nielsen, 2011b) oder perzeptiv anhand der Einschätzungen von HörerInnen (z.B. Goldinger, 1998; Nye & Fowler, 2003; Pardo, 2006).

Innerhalb dieses Forschungskontextes wird immer wieder eine bestimmte Beobachtung gemacht, aus der sich letztlich die vorliegende Studie motiviert:

Eine Vielzahl von Studien zu spontaner und intendierter Imitation zeigt, dass sich SprecherInnen in der Stärke der Imitation unterscheiden bzw. nicht alle SprecherInnen messbare Imitation aufweisen (Studien zu spontaner Imitation: Babel, 2010, 2012; Goldinger, 1998; Nielsen, 2011b; Studien zu verlangter Imitation: Nye & Fowler, 2003; Reiterer, Singh & Winkler, 2012). In den genannten Studien wurden diese Unterschiede oder ihre Ursache jedoch nicht genauer betrachtet.

Faktoren, die den Grad der Imitation beeinflussen, sind möglicherweise Unterschiede in der perzeptiven Sprachverarbeitung: "A lack of strong phonetic convergence could be due to insufficiently detailed phonetic perception by the talkers, to weak connection between perception and production in their cognitive systems, or to insufficient implementation of perceived phonetic forms on the part of the talkers" (Pardo, 2012b, S.6).

Theoretische Grundlage für das Vorgehen in der vorliegenden Studie bildet eine funktionale Verbindung zwischen Perzeption und Produktion:

So beschreibt die Motor-Theorie (Lieberman & Mattingly, 1985) Sprachwahrnehmung als Prozess, während dessen das akustische Signal in intendierte Gesten und daraufhin in Gesten des eigenen Sprechapparates übersetzt werden. Demnach sind die Bewegungsstrukturen der Sprachproduktion maßgeblich an der Sprachwahrnehmung

beteiligt. Dieser Zusammenhang besteht auch andersherum, das heißt, dass sich wahrgenommene Äußerungen auf die Realisierung zukünftiger Äußerungen auswirken. Wenn es individuelle Unterschiede in der Sprachwahrnehmung geben sollte, müsste sich dies anhand unterschiedlicher Beeinflussung der Produktion durch wahrgenommene Äußerungen zeigen.

Exemplartheoretisch (Johnson, 1997; Pierrehumbert, 2003) betrachtet, haben sprachliche Wahrnehmungen dann einen Einfluss auf die nachfolgende Produktion, wenn die Wahrnehmung phonetisch außergewöhnlich ist, aber einer bestehenden Kategorie zugeordnet wird bzw. werden kann. Die Grundlage für die konkrete phonetische Realisierung einer phonetischen Kategorie bildet in dieser Theorie der Durchschnitt aller bereits wahrgenommenen Exemplare dieser Kategorie. Wenn sich ein Exemplar davon unterscheidet, verändert sich die prototypische Vorstellung der Kategorie. Daraus folgt die Beeinflussung des Prototyps und in der Folge der Produktion (Johnson, 1997; Nosofsky, 1988). Bei individueller Sprachwahrnehmung, z.B. unterschiedlich guter Auflösung der Exemplarunterschiede innerhalb einer Kategorie, kann die Beeinflussung des Prototyps, bei gleicher wahrgenommener Äußerung, unterschiedlich ausfallen.

Da an spontaner phonetischer Imitation sowohl die Perzeption wie auch die Produktion beteiligt sind, müssen beide Gebiete der Sprachverarbeitung Betrachtung finden: individuelle Unterschiede in der Produktion ebenso wie individuelle Unterschiede in der Perzeption, worauf im Folgenden eingegangen wird.

Individuelle Unterschiede in der Produktion werden gemeinhin in indexikalische Eigenschaften (*indexical properties*) und phonetisch-relevante Eigenschaften unterteilt (Allen, Miller & DeSteno, 2003). Indexikalische Eigenschaften sind akustische Entsprechungen zu physiologischen Ausprägungen, wie Vokaltraktform und -größe (Peterson & Barney, 1952), Kehlkopfgröße (Klatt & Klatt, 1990), aber auch Dialekt (Byrd, 1992). Eigenschaften der Produktion, die individuell unterschiedlich und dazu phonetisch relevant sind, sind beispielsweise Formantfrequenzen (Peterson & Barney, 1952), Zentroidfrequenz und Schiefe der Friktion von /s/ und /ʃ/ (Newman, Clouse & Burnham, 2001) und VOT (normalisiert auf Sprechgeschwindigkeit) (Allen et al., 2003).

Auch in der systematischen Funktionalität von phonetisch relevanten Eigenschaften unterscheiden sich SprecherInnen. Es gibt abgrenzbare Strategien, anhand derer Wortgrenzen angezeigt werden: Systematisch genutzt werden entweder primär temporäre Reize wie die Dauer des sich vor der Wortgrenze befindenden Konsonanten oder Vokals oder primär qualitative Reize der Vokale bzw. Konsonanten vor der Wortgrenze; daneben gibt es aber auch eine systematische Verwendung beider Formen der Variation (qualitativ

und temporär). In der Größe der Reizveränderung zur Markierung von Wortgrenzen treten auch sprecherspezifische Unterschiede auf (Smith & Hawkins, 2012).

Individuelle Unterschiede in der Leistung von Versuchspersonen bei einer Vielzahl von Aufgaben werden immer wieder beobachtet (Allen & Miller, 2004; Nygaard, Sommers & Pisoni, 1994), jedoch kaum in Hinblick auf deren Ursachen diskutiert. Unterschiede in der Sprachwahrnehmung finden Beachtung, wenn oppositionelle Gruppen erforscht werden: z.B. alt – jung (Harrington, Kleber & Reubold, 2008), weiblich – männlich (accomodation: Pardo, Cajori Jay & Krauss, 2010), Dialekt A – Dialekt B (Ladefoged & Broadbent, 1957; Makashay, 1999).

Studien mit SprecherInnen und HörerInnen verschiedener Dialekte zeigen Unterschiede in der Sprachwahrnehmung von Vokalen: in der Studie von Ladefoged und Broadbent (1957) nahmen englische RP-Sprecher (*Received Pronunciation*) ein Testwort im Vergleich zu schottischen Sprechern häufiger als /bʌt/ wahr, wohingegen schottische Sprecher häufiger angaben, /bat/ gehört zu haben. Dieser Unterschied ist auf die unterschiedlichen Vokalräume zwischen RP und Schottisch zurückzuführen, weil das /ʌ/ im RP-Stimulus dem Schottischen /a/ sehr ähnlich ist. Peeters (1991) untersuchte die Wahrnehmung von „guten Exemplaren“ anhand mehrerer Sprachen (Niederländisch, Englisch, Deutsch) und Dialekte und fand, dass innerhalb des Ostfriesischen Bewohner unterschiedlicher Dörfer unterschiedliche Realisationen von /ai/ und /au/ präferierten. Andere Studien besprechen Sprecherunterschiede in der Perzeption, ohne den dialektalen Hintergrund der SprecherInnen zu berücksichtigen (z.B. Christovich & Chernova, 1986, Kewley-Port & Watson, 1994; Zhukov & Christovich, 1985).

Um individuelle Unterschiede in der Verarbeitung von Sprache zu betrachten, ist es notwendig, dialektale Einflüsse konstant bzw. kontrolliert zu halten. Fox (1983) testete die Verwendung bzw. Gewichtung unterschiedlicher perzeptiver Dimensionen in Mono- und Diphthongen des amerikanischen Englisch anhand multidimensionaler Skalierung (im Speziellen: *Individual Differences Scaling*). Als Versuchspersonen dienten 24 SprecherInnen des gleichen Dialektgebietes (*Midwestern dialect of American English*). Er fand große individuelle Unterschiede in der Gewichtung der verwendeten Dimensionen zur Diskrimination von Mono- und Diphthongen.

Makashay (2003) testete den perzeptiven Umgang einer in Hörleistung und dialektalen Einflüssen homogenen Versuchspersonengruppe unter Verwendung der Reize Frequenz und Dauer (in Frikativen und Vokalen von sprachlichen und zusätzlich nicht-sprachlichen Stimuli). Er fand Unterschiede in der Gewichtung der Parameter für Diskrimination und Identifikation sowie in der Beständigkeit ihrer Verwendung.

Wie auch Ladefoged und Broadbent (1957), bearbeitete Makashay (1999; weitere Ergebnisse in Makashay, 2003) die Verbindung von Perzeption und Produktion anhand dialektaler Sprache: die Daten seiner Studie mit HörerInnen aus Birmingham, AL (Süden der USA) und HörerInnen aus Binghamton, NY (Norden der USA) zeigten, dass Unterschiede in der Wahrnehmung von Diphthongen darauf hinweisen, dass Dialekte nicht nur die Produktion, sondern auch die Perzeption beeinflussen. Im Gegensatz zu Ladefoged und Broadbent wiesen die Makashay (1999)-Daten auch darauf hin, dass sich die Perzeption nicht in allen Fällen, wohl aber in manchen, direkt aus der Produktion ableiten lässt.

Newman (2003) nahm Produktionen der Silbe /pa/ auf und ließ die Versuchspersonen im Anschluss beurteilen, wie prototypisch ihnen die Vokale eines /ba/-/pa/-/pha/ (/pha/: sehr stark gelängte Aspiration) erschienen. Es zeigte sich, dass Versuchspersonen, die Exemplare mit längerer VOT prototypisch fanden, auch selbst längere VOT produzierten. Diese Studie bezieht sich nicht auf die Ursachen für individuelle Unterschiede (oder von Subgruppen), sondern zeigt einen individuellen Zusammenhang zwischen den phonetisch detaillierten Eigenschaften der Produktion und denen der Perzeption auf.

Unterschiede in der Produktion können eine phonetische Funktion erfüllen. Wenn HörerInnen mit der Stimme (das heißt den indexikalischen und phonetisch relevanten Sprechereigenschaften) vertraut sind, ist die Worterkennung besser als bei Äußerungen von noch nicht wahrgenommenen Stimmen. Vertrautheit mit den individuellen Unterschieden in der Produktion unterstützt die Verständlichkeit von Sprache (Bradlow & Pisoni, 1999; Goldinger, 1996; Nygaard et al., 1994; Nygaard & Pisoni, 1998; Sheffert & Fowler, 1995)

Daran sind explizit auch phonetisch relevante Sprechereigenschaften beteiligt: Versuchspersonen können Sätze aus Sinustönen (*Sinewave Speech*) Sprechern zuordnen, die sie aus täglicher Interaktion kennen (Remez, Fellows & Rubin, 1997). Bei der Reduktion von Äußerungen auf Sinustöne (Formanten 1 bis 3) werden alle indexikalischen Eigenschaften entfernt und das Signal beschränkt sich auf die dynamischen phonetischen Bestandteile (Remez, Rubin, Pisoni & Carrell, 1981). Dieses Ergebnis zeigt, dass HörerInnen phonetisch relevante und individuell systematische Variabilität zur Sprechererkennung nutzen können. Die phonetisch relevanten Sprechereigenschaften, die an Sprechererkennung beteiligt sind, konnten von Allen und Miller (2004) weiter differenziert werden. Sie zeigten, dass die systematische Variation der VOT in Fortis-Plosiven als einziger Sprecherunterschied zwischen zwei Sprecherinnen bereits die Sprechererkennung verbesserte (manipulative Variation der VOT, ausgehend von der gleichen Äußerung).

Im Bereich des Zweitspracherwerbs wurde häufig beobachtet, dass auch bei vergleichbarer Erfahrung und Übung und konstantem Alter des Lernbeginns (u.a.) individuelle Unterschiede im phonetischen Können bestehen blieben (Flege, Bohn & Jang, 1997; Flege, Munro & MacKay, 1995; Moyer, 1999; Sebastian-Galles & Baus, 2005). Phonetisches Können korreliert hierbei nicht mit den Ergebnissen genereller Tests zu sprachlichen Fähigkeiten (Neufeld, 1987). Um einer Erklärung von sprecherspezifischem phonetischen Können näher zu kommen, wurden Unterschiede in den sprachlichen Fähigkeiten bzw. im Talent für den phonetischen Fremdspracherwerb gesucht. Unterschiedliche Leistungen im phonetischen Beherrschen einer Fremdsprache (L2) müssen nicht, können aber auf Strukturen, Mechanismen oder Fähigkeiten beruhen, die einen Einfluss auch auf die Verarbeitung der Muttersprache haben. Entscheidend für diese Argumentation ist die lebenslange Entwicklung bzw. Veränderung der phonetischen Realisierung von Laut(sequenz)en in der Muttersprache (Harrington, Palethorpe & Watson, 2000). Ein weiterer Hinweis liegt in einem Teilergebnis von Jilka (2009a, 2009b), der in Identifikations- und Diskriminationstests in L1, L2 und einer weiteren Sprache, in der die Versuchspersonen keine vorherige Erfahrung hatten, keinen Vorteil der Muttersprache fand. Das heißt die Leistungsdifferenzierung von perzeptiven Fähigkeiten anhand der verwendeten Aufgaben ist in der Muttersprache in gleichem Maße wie in Fremdsprachen möglich. Grundsätzlich ist Fähigkeit (gleichzusetzen mit der Bezeichnung „Talent“) von Können zu unterscheiden. Faktoren wie Motivation, Übung und Erfahrung beeinflussen das Können, nicht aber das Talent (Jilka, 2009b). Die (fremd)sprachliche Fähigkeit ist stabil und nicht empfänglich für einfaches Training oder Modifikation und wird zumindest nach den frühen Jahren im Spracherwerb nicht von Umweltfaktoren beeinflusst. Vielmehr ist (fremd)sprachliche Fähigkeit ein Set kognitiver Fähigkeiten, dessen Ausprägung entweder genetisch ist oder sich in frühen Jahren gefestigt hat (Skehan, 2003).

Um dieses Set der Fähigkeiten zu klassifizieren, werden Testbatterien durchgeführt. Diese testen Können in verschiedenen linguistischen Bereichen, unter anderem mittels phonetischer Aufgaben in der Fremdsprache, der Muttersprache und einer völlig unbekanntem Sprache. Die Ergebnisse dieser Tests werden dann anhand verschiedener Faktoren, wie Dauer des fremdsprachlichen Unterrichts und Fremdsprachgebrauchs, gewichtet. So sollen die zugrunde liegenden Fähigkeiten offen gelegt werden. Bei vergleichbarer Motivation, Erfahrung und Übung müssten unterschiedliche Ergebnisse also auf unterschiedliche Fähigkeiten zurückzuführen sein. Spezialisiert auf die Bewertung phonetischen Könnens ist ein Projekt von Jilka (2009a, 2009b), in dem Versuchspersonen anhand ihrer Leistung in Produktion, Perzeption und Imitation (als

Verbindung von Produktion und Perzeption) anhand meist prosodischer aber auch segmenteller Parameter in Deutsch (L1), Englisch (L2) und Hindi (keine vorherigen Kompetenzen) bewertet wurden. Bei dieser Studie mit 102 deutsch-muttersprachlichen Versuchspersonen wurde das Können durch Produktions-, Perzeptions- und Imitationstests geprüft und so auf Fähigkeit bzw. Talent geschlossen. Jilka verweist auf den Einfluss von Motivation, der lang- wie kurzzeitig wirken und die Trennschärfe solcher Tests trüben kann. Er hält den Einfluss von Motivation auf weniger bewusst gezeigte Kompetenzen für geringer und nutzt diese Möglichkeit, seine Ergebnisse zum Talent zu überprüfen. Lewandowski (2012) führte eine englischsprachige (L2) Dialogstudie durch, um phonetische Fähigkeiten mit Konvergenz in Verbindung zu setzen. Hierfür nutzte sie zwei Gruppen von Versuchspersonen, die in der Studie von Jilka (2009a, 2009b) entweder als mit hohem oder mittlerem bis schlechtem Talent charakterisiert wurden, jedoch alle ein hohes Können in Englisch als L2 aufwiesen. Sie stellte fest, dass die SprecherInnen der Gruppe mit höherem Talent stärker konvergierten als die SprecherInnen mit weniger Talent. Sie wertete die Aufnahmen der Dialoge anhand von *Amplitude Envelopes* der Zielwörter aus, die sowohl prosodische als auch spektrale Informationen enthielten.

Es ist also möglich, dass die Fähigkeiten, die Jilka zur Beurteilung von Talent testete, auch bei Konvergenz eine Rolle spielen. Die bei diesen Tests verwendeten Aufgaben sind in der folgenden Tabelle (Tabelle 1) aufgeführt.

	Produktion	Imitation	Perzeption
Prosodische Analyse	Ausdruck von Emotionen und Kontrastakzentuierung in deutschen Sätzen	Verzögerte Imitation Deutsch	Paarvergleich deutscher Sätze anhand der Prosodie
Prosodische und segmentelle Analyse	Lesen einzelner Sätze auf Deutsch mit fingiertem Englischen, Französischen und Italienischen Akzent	Verzögerte Imitation Englisch	Paarvergleich Tiefpass gefilterter deutscher und englischer Sätze anhand der Prosodie
	Lesen eines Textes auf Deutsch mit fingiertem Englischen Akzent	Direkte Imitation Hindi	Identifikation fremdländischer Akzente in Deutschen Sätzen

Tabelle 1: Die zur Bewertung von phonetischem Talent verwendeten Aufgaben aus Jilka (2009a). Alle Aufgaben beinhalten prosodische, nur die dunkelgrau hinterlegten, auch segmentelle Aspekte.

Zusammenfassend liefern die Studien von Jilka (2009a, 2009b) und Lewandowski (2012) Hinweise darauf, dass die Personen, die in den in Tabelle 1 genannten Aufgaben besser abschneiden, in einer Dialogsituation (Englisch: L2) stärker konvergieren (imitieren).

Damit unterstützt dieses Ergebnis die Annahme, dass Unterschiede in der spontanen phonetischen Imitation zwischen SprecherInnen auch in anderen Bereichen der Sprachverarbeitung bestehen. Dennoch werden in der vorliegenden Studie die Unterschiede in Produktion und Perzeption nicht mit zugrunde liegendem Talent in Verbindung gesetzt. Vielmehr werden sie wertungsfrei gegenübergestellt. Es soll hier also nicht getestet werden, ob sprachliche Fähigkeiten Imitation fördern oder hemmen, ob es „gut“ oder „schlecht“ ist, wenn imitiert wird, sondern ob stärkere oder schwächere Imitation auf unterschiedliche perzeptive Strategien zurückzuführen sein kann. Dazu wird ein Identifikationstest verwendet, dessen Auswertung sich auf die Neigung der Identifikationskurve zwischen den Vokalkategorien bezieht. Mögliche Unterschiede in der mentalen Definition der Vokalkategorien und deren Kategorien-Prototypen aufgrund dialektaler Ausprägungen (trotz der selben regionalen Herkunft der Versuchspersonen), messbar anhand des Wendepunkts der Identifikationskurve, finden keine Betrachtung. Vielmehr wird die Trennschärfe zwischen den Vokalkategorien /ɛ/ und /a/, das heißt eine abrupte im Gegensatz zu einer graduellen Kategoriengrenze ausgewertet.

Ein steiler Übergang zwischen zwei Vokalkategorien zeigt eine hohe Beständigkeit und Sicherheit in der perzeptiven Kategorisierung von Vokalen. Wenn sich diese Beständigkeit auch in der Produktion zeigt, wird wenig oder keine Imitation erwartet. Im Gegenzug kann die flexiblere Kategorisierung von Vokalen ein Pendant in der Produktion haben, welches sich durch die zielgerichtete Beeinflussung zeigt. Es wird also davon ausgegangen, dass SprecherInnen, die (mehr) Imitation zeigen, flachere Kategoriengrenzen aufweisen und vice versa.

Welche der zehn SprecherInnen, die am hier dargestellten Experiment teilnahmen, als potentielle Imitatorinnen gelten können, wurde sowohl perzeptiv durch HörerInnen (Abschnitt 2.2: Erfassen der potentiellen Imitatorinnen durch HörerInnen), als auch durch eine akustische Analyse (Abschnitt 2.3: Erfassen der potentiellen Imitatorinnen durch akustische Analyse) ermittelt. In Abschnitt 2.4 (Produktion-Perzeption: Identifikationstest) wird der mögliche Zusammenhang zur Perzeption anhand eines Identifikationstests geprüft: Besteht im Umgang mit einem Kontinuum zwischen zwei Vokalen ein Unterschied zwischen jenen Versuchspersonen, die Imitation bzw. Annäherung zeigen und jenen, die dies nicht tun?

2.2 Erfassen der potentiellen Imitatorinnen durch HörerInnen

Spontane phonetische Imitation zeigt sich als akustische Annäherung. Im ersten Schritt wurde zur Erfassung dieser Veränderungen ein Perzeptionsexperiment durchgeführt. So wurden in der Analyse alle potentiellen akustischen Veränderungen abgedeckt und

gleichzeitig gewährleistet, dass die Veränderungen auch wahrnehmbar waren (Pardo, 2012b). Dass Versuchspersonen zu Einschätzungen dieser Art in der Lage sind, haben andere Studien gezeigt (Goldinger, 1998; Pardo, 2006; Pardo, Cajori Jay & Kraus, 2010). Bei den beurteilten Äußerungen handelte es sich um *Shadowing*-Äußerungen (siehe Abschnitt 2.2.1.2 Stimuli).

2.2.1 Methodik

2.2.1.1 Versuchspersonen

Am Perzeptionsexperiment nahmen 20 muttersprachlich deutsche Versuchspersonen teil: 14 Frauen und 6 Männer (im Folgenden als „HörerInnen“ bezeichnet). Nach eigenen Angaben hat keine der Versuchspersonen Sprech- oder Hörprobleme.

2.2.1.2 Stimuli

Für die Aufnahme der Stimuli für das Perzeptionsexperiment wurde das *Shadowing*-Paradigma von Goldinger (1998; Honorof, Weihing & Fowler, 2011) benutzt: Den Sprecherinnen wurden via Kopfhörer einzelne Wörter, gesprochen von einer Modellsprecherin, insgesamt sechs mal, in randomisierter Reihenfolge vorgespielt. Diese Äußerungen wurden von den Sprecherinnen sofort nach der Erkennung der Wörter (siehe Tabelle 2) nachgesprochen, auch wenn die präsentierte Äußerung noch nicht beendet war.

An den Aufnahmen für die *Shadowing*-Äußerungen nahmen zehn muttersprachlich deutsche Versuchspersonen (im Folgenden als „Sprecherinnen“ bezeichnet) zwischen 24 und 32 Jahren teil, die in München und Umgebung aufwuchsen und leben. Die Modellsprecherin (28 Jahre) wuchs ebenfalls in München auf und lebt dort. Nach eigenen Angaben hat weder die Modellsprecherin noch eine der Sprecherinnen Sprech- oder Hörprobleme.

2.2.1.2.1 Aufnahme der Äußerungen der Modellsprecherin

Für die Stimulikonstruktion wurden folgende Äußerungen der Modellsprecherin aufgenommen: viermal die Füllwörter, zweimal die Zielwörter und achtmal jeweils die Wörter „Welt“ und „Wald“ (für das Vokalkontinuum, siehe Abschnitt 2.4.1.2: Stimuli) jeweils in gewöhnlicher Artikulation, sowie sechsmal die Zielwörter hyperartikuliert (siehe Tabelle 2). Die Zielwörter wurden in nicht-randomisierter Reihenfolge in Blöcken aufgenommen. Die Modellsprecherin wurde instruiert, deutlich zu sprechen und in den Blöcken der hyperartikulierten /a/- und /ɛ/-Wörter zusätzlich die Zielwörter, die an betonter Position /a/ enthalten, möglichst offen („mit möglichst weit offenem Mund“) und die Zielwörter, die an betonter Position /ɛ/ enthalten möglichst gespreizt („Mundwinkel möglichst weit nach

außen“) zu produzieren. Diese Herangehensweise wurde gewählt, um ungewöhnlich hohe F1-Werte für /a/ und ungewöhnlich tiefe F1-Werte für /ɛ/ zu erhalten. Tiefe Vokale (wie /ɛ/ und /a/) werden stärker imitiert als hohe Vokale (z.B. /i/, /e/), da der F1-Raum von tiefen Vokalen eine stärkere Variabilität aufweist, abhängig von der Position der Wort- oder Satzbetonung (Babel, 2011).

Die Wörter wurden mit der Speech Recorder Software (Draxler & Jänsch, 2004) visuell auf einem Monitor präsentiert. Zur Aufnahme wurde ein TLM 103 Mikrofon und eine ADAT Schnittstelle zwischen einem YAMAHA O2R Mischpult und einem HP Compaq dc7800 CMT PC All (p/n GC758AV) verwendet. Die Aufnahmen wurden mono mit 16 bit aufgezeichnet und mit einer Samplerate von 44.1 kHz digitalisiert.

Zielwörter		Füllwörter	
Ballkleid	Feldpost	Aufbau	Fehler
Balsam	Feldweg	Bayreuth	Fräulein
Bastler	Festmahl	Blöcke	Frühstück
Fallobst	Fettfleck	Blödsinn	Füße
Fassbier	Testbild	Bochum	Hölzer
Palmbblatt	Wellblech	böse	Hütte
Passform	weltfremd	Brunnen	schöner
tapsig	Weltmeer	Deutsche	
Tastsinn	westwärts		
Waldlauf	Wettspiel		

Tabelle 2: Ziel- und Füllwörter der Shadowing-Aufgabe

Aus den sechs Aufnahmen pro Wort wurde eine Äußerung ausgewählt, die zum einen fehlerfrei, zum zweiten deutlich und zum dritten natürlich produziert war. Bei /a/ wurde von diesen die am offensten, bei /ɛ/ die am meisten geschlossen klingende Äußerung bevorzugt.

Die Zielwörter bestehen aus zwei Silben, wobei die erste betont ist und einen der Zielvokale /a/ oder /ɛ/ enthält. Da bei Wörtern mit geringer Häufigkeit stärkere Imitation auftritt als bei Wörtern mit hoher Häufigkeit (Goldinger, 1998; Babel, 2010), wurden die Zielwörter anhand ihrer Häufigkeit im CELEX-Korpus (Baayen, Piepenbrock & Gulikers, 1995) ausgewählt: Die Worthäufigkeit der Lemmata der Zielwörter beträgt maximal 1/1Mio Wörter bzw. maximal acht Treffer des Lemmas innerhalb des CELEX-Korpus (bestehend aus 6 000 000 Wörtern). Im BAS-Korpus (772 386 *word types* aus spontaner Dialogsprache: BAS, 2009) kommt keines der Wörter vor.

2.2.1.2.2 Aufnahme der *Shadowing*-Äußerungen der Sprecherinnen

Hier wiederholten die Sprecherinnen die Wörter, die ihnen, gesprochen von der Modellsprecherin, via Kopfhörer vorgespielt wurden.

Im Testblock wurden 35 Wörter (20 Ziel- und 15 Füllwörter) in zufälliger Reihenfolge jeweils sechsmal wiederholt. Zuvor wurden fünf der späteren Füllwörter jeweils dreimal nachgesprochen, um die Sprecherinnen mit der Aufgabe vertraut zu machen. Das wiederholte Nachsprechen soll nach Goldinger (1998) dazu führen, dass die Äußerungen im Laufe der Wiederholungen immer stärker imitiert werden.

Das *Shadowing* dauerte ca. 35 Minuten. Für die Durchführung wurden geschlossene Kopfhörer (beyerdynamic DT 770) verwendet. Darauf folgte ein Identifikationstest (und ein Distanzdiskriminationstest, dessen Ergebnis hier nicht dargestellt wird). Die gesamte Zeit von ca. einer Stunde wurde mit 10 € pro Sprecherin entlohnt.

Die Anweisungen zum *Shadowing* wurden mündlich gegeben: Die Wörter aus dem Kopfhörer sollten, sobald sie erkannt wurden, zügig wiederholt werden. Eine unverändert aufrechte Sitzposition wurde durch das Zusammenführen der Arme hinter der Stuhllehne gewährleistet. Alle Wörter, die im Experiment Verwendung fanden, wurden schriftlich vorgelegt, mit der Bitte, sie zwei Mal laut vorzulesen und anzugeben, ob alle Wörter bekannt sind. Diese Frage wurde von allen Sprecherinnen bejaht.

Die Äußerungen der Modellsprecherin wurden mit der Speech Recorder Software (Draxler & Jänsch, 2004) in randomisierter Reihenfolge innerhalb eines Blocks (bestehend aus den 35 Wörtern in Tabelle 2) abgespielt. Ein Aufnahmezeitfenster von 3 Sekunden (250 ms vor der Einspielung, 2750 ms ab Beginn der Einspielung) nahm die Produktion der Sprecherinnen auf. Es gab keine visuelle Darstellung und keine Pause während der Aufnahme der sechs *Shadowing*-Wiederholungen.

Die Äußerungen der Sprecherinnen wurden mit einem TLM 103 Mikrofon und einer ADAT Schnittstelle zwischen einem YAMAHA O2R Mischpult und einem HP Compaq dc7800 CMT PC All (p/n GC758AV) aufgenommen. Die Aufnahmen wurden mono mit 16 bit aufgezeichnet und mit einer Samplerate von 44.1 kHz digitalisiert.

2.2.1.3 Experimentelles Paradigma des Perzeptionstests

Mit den *Shadowing*-Äußerungen der Sprecherinnen wurde der 4IAX-Perzeptionstest mit einem XAXB-Aufbau erstellt (Sprachverarbeitungssoftware *Praat*, Boersma, 2001). Ein 4IAX-(4 Interval AX)-Perzeptionstest besteht in seiner ursprünglichen Form aus vier Stimuli, von denen drei identisch sind. Der vierte, sich unterscheidende Stimulus, wird an Position zwei oder drei platziert. Die Position dieses Stimulus ist von Versuchspersonen anzugeben, indem sie entscheiden, ob sich der erste Stimulus vom zweiten stärker unterscheidet als der dritte vom vierten Stimulus. Dieser Aufbau bewirkt gegenüber

anderen Stimulusanordnungen (z.B. ABX oder AXB) eine geringere Beeinflussung durch Antwortneigungen der Versuchspersonen (*subjective response bias*) und minimiert Positionsvorteile. Darüber hinaus treffen Versuchspersonen die Entscheidung aufgrund nicht vorgegebener Parameter und werden so nicht durch Antwortkategorien (z.B. 2AFC) beeinflusst (Schouten, Gerrits & van Hessen, 2003).

Die erste und sechste Wiederholung der Wörter aus dem *Shadowing* wurden an den Positionen A und B vorgespielt (und vice versa), X entsprach der Äußerung der Modellsprecherin. Die Hörer beurteilten, ob die Äußerung an Position A oder die Äußerung an Position B der Äußerung X ähnlicher war, d.h. ob sich das erste oder das zweite Paar in sich ähnlicher war.

Verwendet wurden nur die Zielwörter. So waren 400 Stimuli zu beurteilen: 20 Wörter x 10 Sprecherinnen x 2 Reihenfolgen (erste Wiederholung an Position A und sechste Wiederholung an Position B versus sechste Wiederholung an Position A und erste Wiederholung an Position B).

Das Perzeptionsexperiment wurde in zwei ca. 45 Minuten dauernden Sitzungen durchgeführt, die mindestens zwei Stunden und maximal 3 Tage auseinander lagen. Nach der zweiten Sitzung wurden die HörerInnen mit 17 € entlohnt. Sie saßen in Gruppen von maximal 6 Teilnehmenden an jeweils einem Rechner und hörten die Äußerungen über geschlossene Kopfhörer (beyerdynamic DT 770). Die Versuchsanweisungen wurden schriftlich erteilt (siehe Anhang 1: Anweisungen HörerInnen (XAXB)) und in einzelnen Fällen mündlich nochmal durchgesprochen.

2.2.1.4 Auswertung der HörerInnenurteile

Um aus den HörerInnenurteilen die Erkennungsleistung zu berechnen, wurden die Antworten der HörerInnen mit den Experimentdurchgängen verglichen. Wenn HörerInnen die sechste Äußerung als ähnlicher gewählt haben, gilt dies als „richtig“, wenn sie die erste Äußerung gewählt haben, als „falsch“.

2.2.2 Ergebnis

Mit der richtig/falsch-Verteilung als *Response* Faktor wurden zwei *Mixed Effect Model* Analysen (Baayen, 2008) gerechnet. Im ersten Modell wurde der Einfluss der Sprecherinnen auf die richtig/falsch-Verteilung berechnet. Dieser war hoch signifikant ($X^2[1] = 400.4$, $p < 0.001$, HörerInnen als *Random* Faktor). Im zweiten Modell wurde der Einfluss der HörerInnen auf die richtig/falsch-Verteilung berechnet. Auch dieser war signifikant, wenn auch vergleichsweise schwach ($X^2[1] = 5.1$, $p < 0.05$, Sprecherinnen als *Random* Faktor).

Die sprecherspezifische Auswertung wurde pro Sprecherin vorgenommen. Jeweils ein *Mixed Model* mit richtig/falsch-Verteilung als *Response* Faktor, Wort und HörerInnen als *Random* Faktor prüfte die Verteilung auf einen signifikanten Unterschied zwischen der Anzahl der richtigen und der falschen Antworten.

Bei drei Sprecherinnen konnten die HörerInnen signifikant die spätere als die ähnlichere Äußerung erkennen. Diese sind in Abbildung 3 mit drei Sternchen markiert (AM: $z = 4.2$, $p < 0.001$; IM: $z = 3.5$, $p < 0.001$; MH: $z = 4.9$, $p < 0.001$). Bei fünf Sprecherinnen bestand kein signifikanter Unterschied in der richtig/falsch-Verteilung (EE: $z = 1.7$, $p = 0.08$; IB: $z = 0.3$, $p = 0.8$; JK: $z = 0.3$, $p = 0.8$; PS: $z = 1.6$, $p = 0.1$; SA: $z = 0.5$, $p = 0.6$) und bei zwei Sprecherinnen gab es signifikant mehr falsche als richtige Antworten (AN: $z = 2.7$, $p < 0.01$; LB: $z = 2.5$, $p < 0.05$).

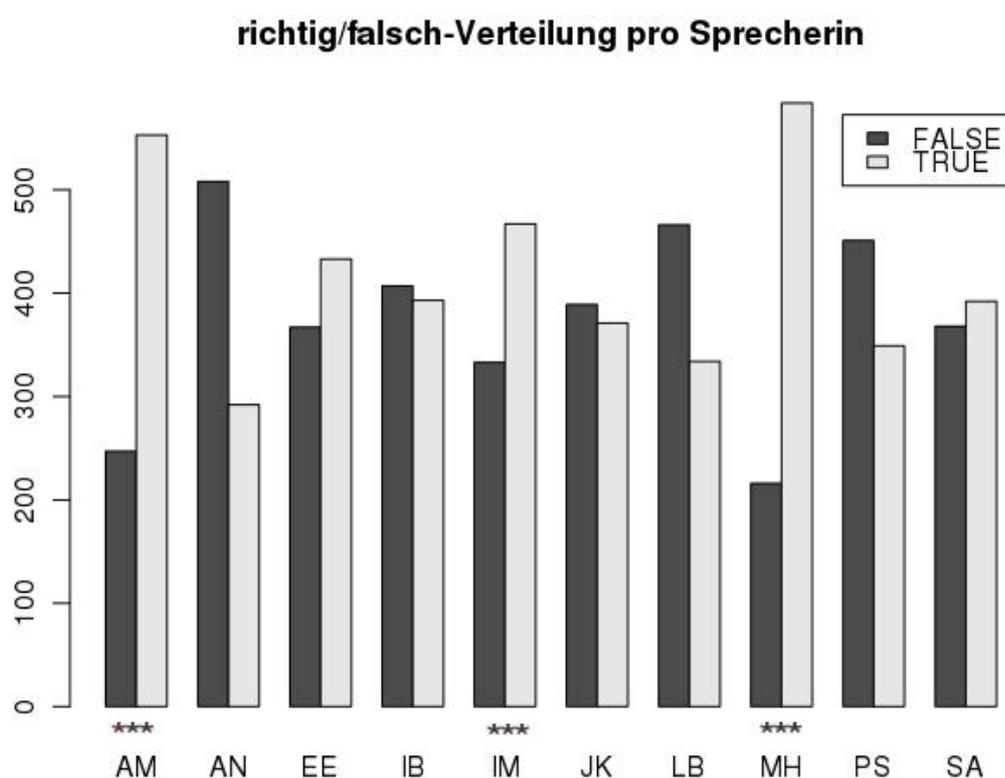


Abbildung 3: Richtig/falsch-Verteilung der HörerInnenurteile pro Sprecherin. Richtig bewertet wird ein Durchgang, wenn die sechste Äußerung als die ähnlichere im Vergleich zur Modellsprecherin wahrgenommen wurde - falsch, wenn die erste Äußerung als ähnlicher wahrgenommen wurde.

2.2.3 Diskussion des Perzeptionsexperiments zur wahrgenommenen Ähnlichkeit

Es wurde für drei Sprecherinnen eine Ähnlichkeit zur Modellsprecherin wahrgenommen. Das macht diese drei Sprecherinnen zu potentiellen Imitatorinnen. Bei

den anderen sieben Sprecherinnen wurde entweder keine Änderung der Ähnlichkeit oder, im Verlauf des *Shadowing*-Experiments, weniger Ähnlichkeit wahrgenommen als zu Beginn. Generell heißt auch das nicht, dass diese sieben Sprecherinnen nicht imitiert haben, jedoch wurde die mögliche Imitation nicht erkannt. Dies kann daran liegen, dass bestimmte Parameter für den Vergleich gegenüber anderen priorisiert werden: Z.B. können Unterschiede in der Grundfrequenz für HörerInnen leichter zu erkennen sein als Unterschiede in der Vokalqualität. Die sehr geringe Gesamterkennung von 52.6% spiegelt sich in der Verteilung von drei Imitatorinnen zu sieben Nicht-Imitatorinnen wider.

2.3 Erfassen der potentiellen Imitatorinnen durch akustische Analyse

Um eine phonetisch genauere Einschätzung der Imitation zu erhalten, wurden die Äußerungen der zehn Sprecherinnen aus dem *Shadowing*-Experiment und der Modellsprecherin akustisch analysiert. So lassen sich die tatsächlichen akustischen Realisationen den Bewertungen von HörerInnen gegenüberstellen. Dies geschah in zweierlei Verfahren: zum einen durch die Gegenüberstellung der ersten und der letzten *Shadowing*-Äußerung und zum anderen durch eine Trendanalyse aller sechs *Shadowing*-Äußerungen in zeitlicher Abfolge.

Hierfür wurden die digitalen Aufnahmen automatisch mit MAUS (Münchener Automatisches Segmentationssystem: Schiel, 2004) segmentiert, auf Segment- und Wortebene gelabelt und diese mit dem Emu Software System (Harrington, 2010) manuell korrigiert. Die Grundfrequenz aller Äußerungen wurde mit dem Emu Pitch Tool (default settings: *Method: ESPS; Frame Spacing: 0.01sec; Window Length: 0.0075sec; Maximum Pitch: 400Hz; Minimum Pitch: 60Hz*) berechnet. Im Folgenden beschreibt die Bezeichnung „Grundfrequenz“ den Durchschnitt der Grundfrequenz über die stimmhaften Bestandteile der gesamten Wortäußerung. Die Formanten eins bis vier (F1-4) wurden mit dem Emu Formant Tool mit folgenden Voreinstellungen berechnet: *Frame Spacing: 0.01sec; Window Length: 0.049 sec; Window Type: Cos⁴; Preemphasis Factor: 0.7; LPC Method: autocorrelation; LPC Order: 12; Analysis Sample Rate: 10000 Hz; Nominal F1 Value: 500 Hz*. Nur F2 findet in dieser Beschreibung Betrachtung, da sich die Hyperartikulation der Modellsprecherin nur in diesem Parameter signifikant niederschlägt ($t = 2.3$, $df = 14.8$, $p < 0.05$). Abbildung 4 zeigt die Gegenüberstellungen der ausgewerteten Parameter der Modellsprache F1, F2, F3 und F1-F0 (F0 ist hier die durchschnittliche Grundfrequenz des Vokals).

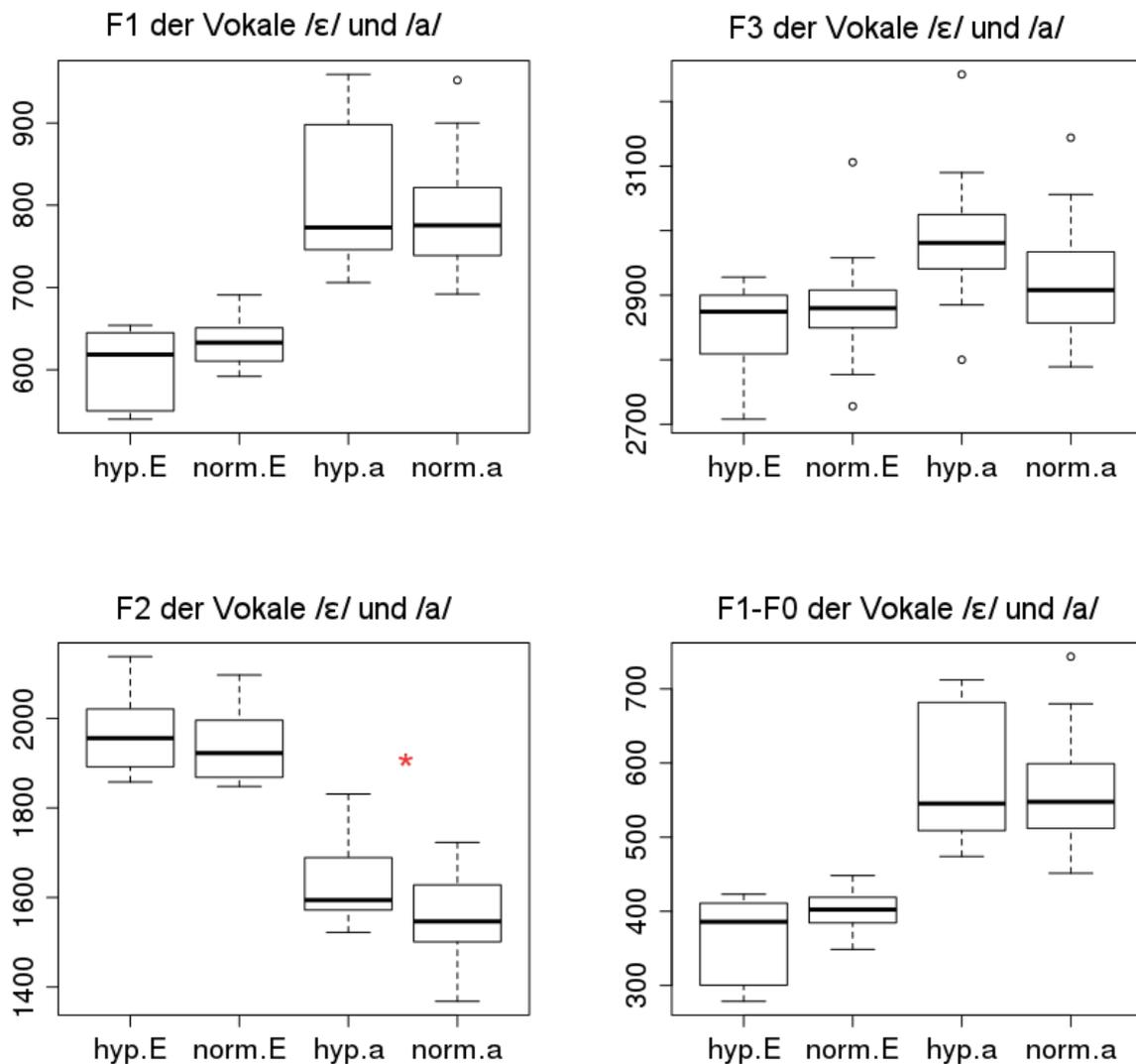


Abbildung 4: Die Parameter F1, F2, F3 und F1-F0 der Modellsprache jeweils für /ε/ („E“) und /a/ („a“) im Vergleich zwischen den im Experiment verwendeten Stimuli („hyp“) und den gleichen Wörtern normal produziert („norm“). Die Einheit der x-Achse ist jeweils Hz.

2.3.1 Vergleich der ersten mit der sechsten Wiederholung in Bezug zur Modellsprecherin

2.3.1.1 Ergebnis: Grundfrequenz

Eine Tendenz der Grundfrequenz in Richtung der Modellsprecherin war bei zwei der drei, im vorherigen Abschnitt (2.2) als potentielle Imitatorinnen identifizierten, Sprecherinnen erkennbar (AM und MH). Die Verteilung der Werte innerhalb der Sprecherinnen zwischen den Konditionen 1 versus 6 unterschieden sich mit einem Vertrauensintervall von 95% nur bei Sprecherin MH ($t=2.4$, $df=38$, $p<0.05$; gepaarter T-Test, 19 Äußerungen aus Kondition 1 und 20 Äußerungen aus Kondition 6).

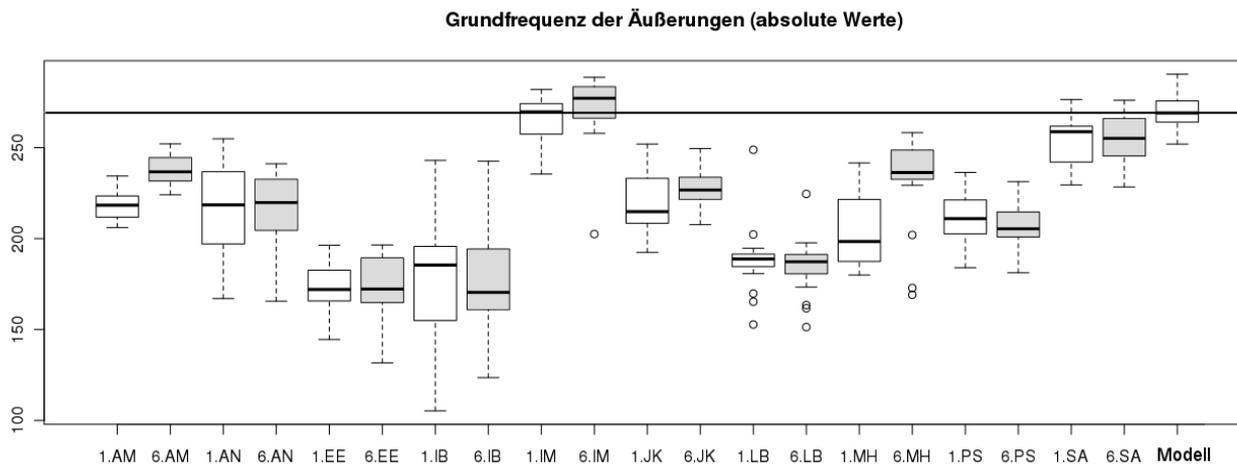


Abbildung 5: Grundfrequenz [Hz] der Sprecherinnen pro Kondition (x-Achse: 1 steht für die Kondition 1 und ist im Boxplot weiß, 6 steht für die Kondition 6 und ist im Boxplot grau markiert) und der Modellsprecherin (Modell). Die Linie liegt auf der Höhe des Medianwerts der Modellsprecherin.

In Abbildung 5 ist erkennbar, dass sich die Grundfrequenz der Sprecherin IM in der Kondition 1 nicht von der Grundfrequenz der Referenzsprecherin unterschied. Die übrigen Sprecherinnen wiesen in Kondition 1 alle eine tiefere F0 auf (siehe Tabelle 3). Aus diesem Grund wurde die Sprecherin IM in der Grundfrequenzanalyse im Weiteren ausgeschlossen.

Sprecherin	T-Test (Unterschied zwischen F0 in der ersten Wiederholung der Sprecherinnen zur Modellsprecherin)	Signifikanz
AM	$t = 18.3, df = 36, p < 0.001$	***
AN	$t = 8.8, df = 23.5, p < 0.001$	***
EE	$t = 24.5, df = 36, p < 0.001$	***
IB	$t = 11.2, df = 20.9, p < 0.001$	***
IM	$t = 1.4, df = 36, p > 0.05$	
JK	$t = 10.4, df = 28.2, p < 0.001$	***
LB	$W = 361, p < 0.001$	***
MH	$W = 361, p < 0.001$	***
PS	$t = 14.2, df = 36, p < 0.001$	***
SA	$t = 4.4, df = 36, p < 0.001$	***

Tabelle 3: Signifikanztest auf den Unterschied zwischen den F0-Werten der Modellsprecherin und den F0-Werten der ersten Wiederholung der Sprecherinnen im Shadowing-Experiment.

Die Veränderung der Grundfrequenz ist besser sichtbar, wenn die Wörter gepaart betrachtet werden, das heißt die Grundfrequenzen der Wörter wurden direkt miteinander verglichen, so wie sie vorgespielt und nachgesprochen wurden. Z.B. wurde F0 der Äußerung /pasfœm/ der Modellsprecherin mit F0 der ersten und letzten nachgesprochenen /pasfœm/-Äußerung der Sprecherinnen einzeln verglichen. So ließen sich die Schwankungen der Grundfrequenz der Modellsprecherin (siehe Abbildung 5) herausrechnen. Abbildung 6 zeigt den Distanzunterschied zwischen den ersten und sechsten Äußerungen der Sprecherinnen und der Modellsprecherin für F0, z.B.:

- (1) **DISTANZ 1** F0 von /pasfœm/ der Modellsprecherin minus F0 von /pasfœm/ einer Sprecherin in der Kondition 1
- (2) **DISTANZ 6** F0 von /pasfœm/ der Modellsprecherin minus F0 von /pasfœm/ einer Sprecherin in der Kondition 6
- (3) **DISTANZ-DIFFERENZ = | DISTANZ 1 | - | DISTANZ 6 |**

Ein positiver Wert als Distanzdifferenz beschreibt eine akustische Annäherung an die Grundfrequenz der Modellsprecherin. Die Mittelwerte dieser Distanzverteilungen sind für die Sprecherinnen AM und MH signifikant größer als Null (AM: $t = 8.2$, $df = 18$, $p < 0.001$; MH: $t = 3.5$, $df = 18$, $p < 0.01$).

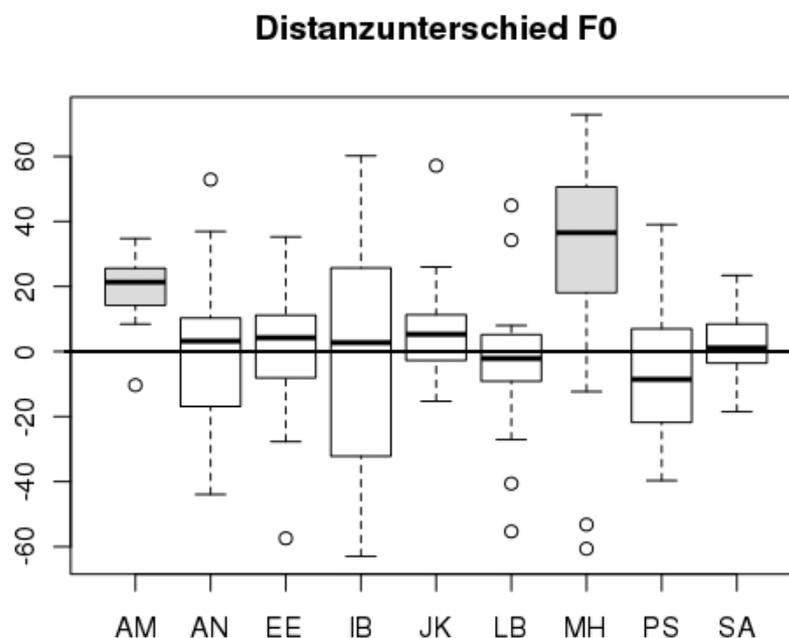


Abbildung 6: Differenz der Distanz von F0 [Hz] der ersten Äußerung und der Distanz der letzten Äußerung des gleichen Worts der Sprecherinnen zur Modellsprecherin. Grau markiert sind die Werte der zwei Sprecherinnen deren F0-Werte sich signifikant an die der Modellsprecherin annähern. Die gezogene Linie markiert Null (Veränderung zwischen den Konditionen).

Die Werte der übrigen Sprecherinnen unterschieden sich nicht von Null (AN: $t = 0.1$, $df = 18$, $p > 0.05$; EE: $t = 0.2$, $df = 18$, $p > 0.05$; IB: $t = 0.01$, $df = 18$, $p > 0.05$; JK: $t = 1.9$, $df = 18$, $p > 0.05$; LB: $t = 0.6$, $df = 18$, $p > 0.05$; PS: $t = 1.0$, $df = 18$, $p > 0.05$; SA: $t = 0.8$, $df = 18$, $p > 0.05$). Es bestand also bei sieben Sprecherinnen keine Änderung der Grundfrequenz (weder Konvergenz noch Divergenz), eine Sprecherin musste aus der Analyse der Grundfrequenz ausgeschlossen werden und Annäherung (Konvergenz) besteht bei zwei Sprecherinnen.

Zwischen den Äußerungen der Kondition 1 und 6 der zwei Sprecherinnen AM und MH besteht anhand der absoluten Wortdauer kein Unterschied (AM: $t = 0.01$, $df = 36$, $p > 0.05$; MH: $t = 0.4$, $df = 36$, $p > 0.05$). Eine Erhöhung der F0 aufgrund von erhöhter Sprechgeschwindigkeit kann für diese Sprecherinnen ausgeschlossen werden (Kappes, Baumgaertner, Peschke & Ziegler, 2009).

2.3.1.2 Ergebnis: Vokalqualität

Als Imitationstrigger dient in dieser Studie Hyperartikulation: Die Zielvokale /ɛ/ bzw. /a/ wurden von der Modellsprecherin besonders geschlossen bzw. offen und deutlich artikuliert. Dennoch unterschieden sich die Formantfrequenzen der Sprecherinnen zu Beginn des *Shadowings* nicht zwangsläufig von den Formantfrequenzen der Modellsprecherin (siehe Anhang 2: Unterschiede der Formantfrequenzen). Signifikante Unterschiede bestanden für alle Sprecherinnen nur bei F2 von /a/ (Abbildung 7). Auch als Effekt der Hyperartikulation wurde nur F2 in /a/ gefunden (Abschnitt 2.3, Abbildung 4). Aus diesen Gründen wurde die Distanzdifferenz nur für diesen Parameter ausgewertet.

Mit den F2-Werten wurden, wie für F0, die Distanzdifferenzen zur Modellsprecherin berechnet (Abbildung 8). Die Distanzdifferenzen der F2-Werte lagen für die Sprecherinnen EE und PS mit einem Vertrauensintervall von 95% über Null (EE: $t = 2.7$, $df = 8$, $p < 0.05$; PS: $t = 2.7$, $df = 8$, $p < 0.05$). Die Verteilungen der F2-Werte der übrigen acht Sprecherinnen unterschieden sich nicht von Null. Wie für F0 trat für F2 in zwei Fällen

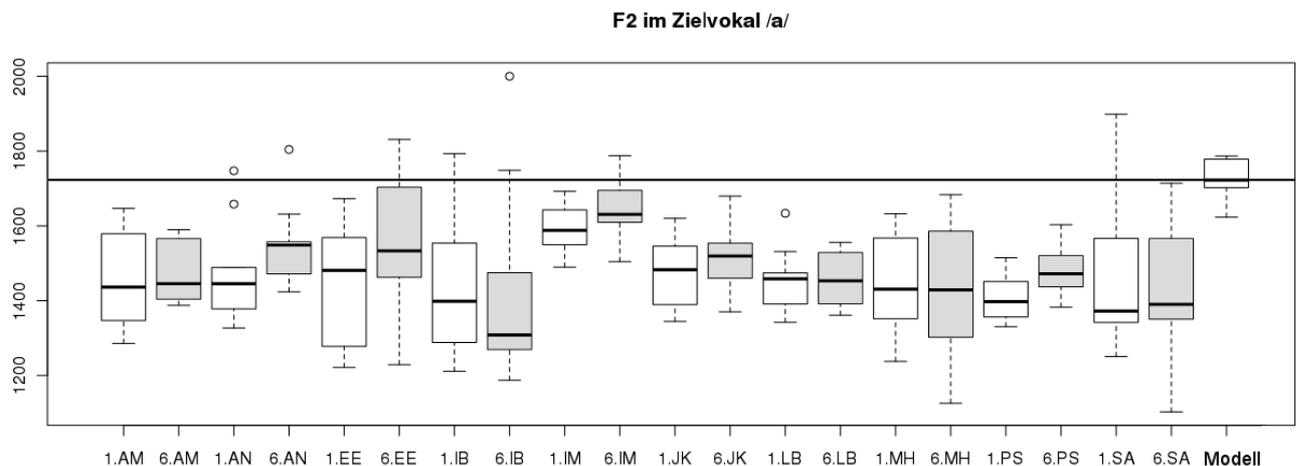


Abbildung 7: F2 [Hz] von /a/ der Modellsprecherin (Modell, Linie markiert den Medianwert) und der zehn Sprecherinnen. 1. bezieht sich auf die erste Shadowing-Wiederholung, 6. auf die sechste Shadowing-Wiederholung (jeweils grau markiert).

akustische Annäherung, in den anderen Fällen weder Konvergenz noch Divergenz auf. Bei F0 waren es die Sprecherinnen AM und MH, bei F2 die Sprecherinnen EE und PS, die Annäherung aufwiesen.

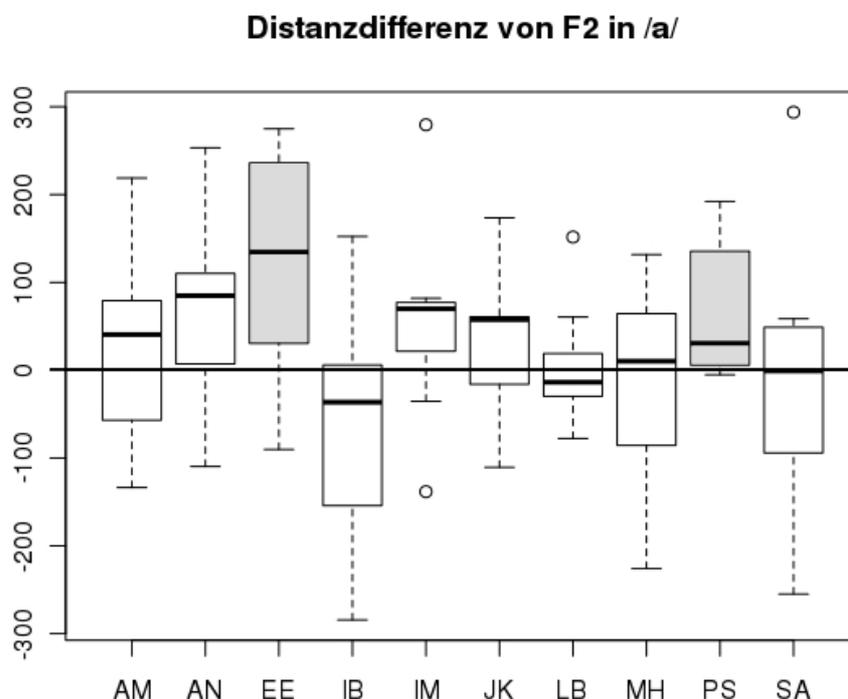


Abbildung 8: Differenz der Distanz [Hz] zwischen den ersten Shadowing-Äußerungen und der Modellsprecherin abzüglich der Distanz zwischen den letzten Shadowing-Äußerungen und der Modellsprecherin für F2 in /a/. Grau markiert sind die signifikant über Null liegenden Werteverteilungen der Sprecherinnen EE und PS. Die Linie markiert Null.

2.3.1.3 Ergebnis: Relative Dauer

Der Versuchsaufbau des wiederholten *Shadowings* begünstigt eine Erhöhung der Sprechgeschwindigkeit im Laufe der Wiederholungen, da die Wörter zunehmend stärker bekannt sind und erwartet werden können (O'Shaughnessy, 1995). Deshalb dient die Betrachtung der absoluten Wortlänge kaum der Ermittlung von Imitation. Besser geeignet sind relative Maße.

Die zu erwartende Tendenz, dass die Sprecherinnen im Laufe des Experiments die Wörter mit höherer Sprechgeschwindigkeit und damit kürzerer Dauer produzieren, ist für die Auswertung zeitlicher Parameter generell problematisch, hier auch für die Auswertung der relativen Vokaldauer. Bei höherer Sprechgeschwindigkeit werden nicht alle Segmente zu gleichen Teilen reduziert: Unbetonte Segmente werden prozentual stärker gekürzt als betonte (Peterson & Lehiste, 1960; Port, 1981). Dies hat einen Einfluss auf die Werte der relativen Vokaldauer, die durch Normalisierung auf die Wortdauer berechnet werden.

Konkret heißt das: Wenn die Sprechgeschwindigkeit erhöht wird, wird der Quotient aus Vokal- und Wortdauer höher. Das dürfte nicht als Imitation der größeren relativen Vokaldauer der Modellsprecherin interpretiert werden.

Deshalb könnte generell nur die Auswertung der Werte der Äußerungen auf Imitation hinweisen, deren absolute Dauern über das Experiment konstant bleiben. Wenn innerhalb dieser Äußerungen eine Zunahme der relativen Vokaldauer gemessen würde, könnte dies potentiell auf Imitation zurückzuführen sein. Aber:

Für keine der 10 Sprecherinnen bestand überhaupt eine Veränderung der relativen Vokaldauer (Abbildung 9). Selbst wenn einige Sprecherinnen die Sprechgeschwindigkeit nicht erhöhten, wäre dennoch auch in dieser Teilmenge weder Konvergenz noch Divergenz zu beobachten.

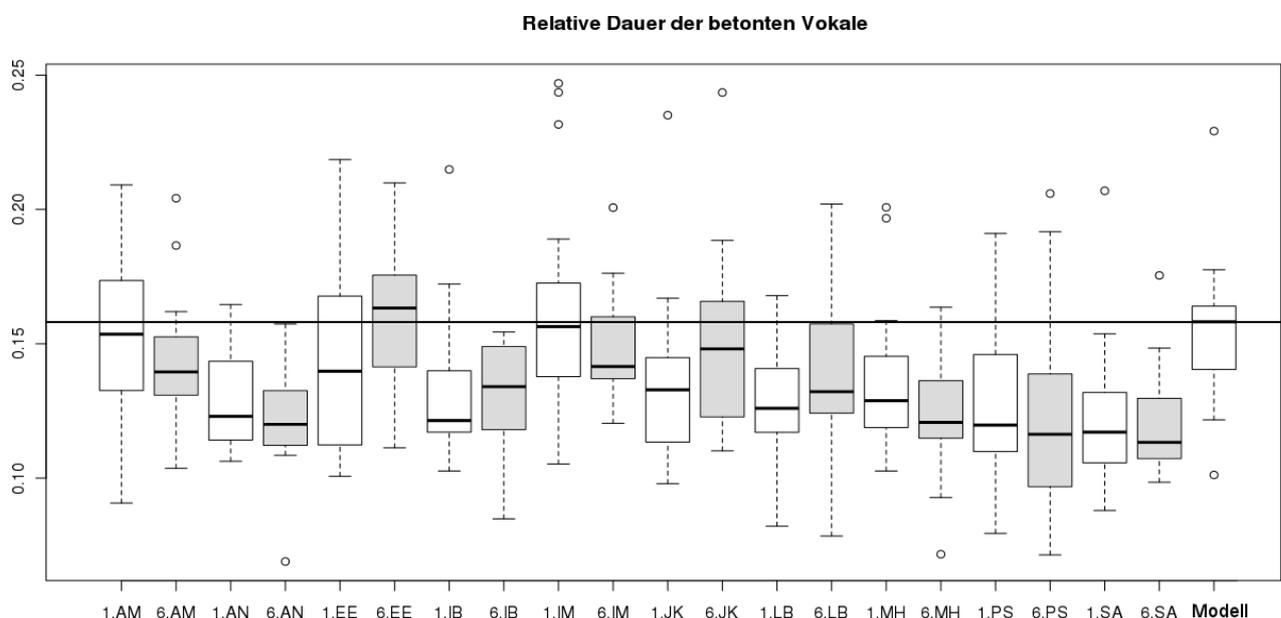


Abbildung 9: Relative Vokaldauer: die Vokaldauer des betonten Vokals normalisiert auf die absolute Wortlänge für die Sprecherinnen und die Modellsprecherin (Modell). Grau markiert sind die späteren Shadowing-Äußerungen, die Linie markiert den Medianwert der Modellsprecherin.

2.3.1.4 Zusammenfassung der Ergebnisse der akustischen Analyse der ersten vs. sechsten Shadowing-Äußerung

Ausgewertet wurde die Veränderung der akustischen Parameter F0, F2 und relative Vokaldauer zwischen der ersten Shadowing-Wiederholung (Kondition 1) und der letzten Shadowing-Wiederholung (Kondition 6). Eine Verringerung der Distanz zur Modellsprecherin war bei F0 von den Sprecherinnen AM und MH sichtbar. Bei dem Parameter F2 war dies bei den Sprecherinnen EE und PS der Fall. In der Auswertung der relativen Vokaldauer war für keine der Sprecherinnen Annäherung oder Divergenz vorhanden.

2.3.1.5 Diskussion der Ergebnisse der akustischen Analyse Kondition 1 vs. 6

Die relative Vokaldauer normalisiert auf die Wortlänge hat sich als nicht praktikabler Parameter erwiesen und bietet auch keine Hinweise für Imitation. Aus diesem Grund wird die relative Vokaldauer in der akustischen Trendanalyse (Abschnitt 2.3.2: Akustische Trendanalyse) nicht ausgewertet.

Zwei Sprecherinnen näherten ihre Grundfrequenz an die der Modellsprecherin an. Diese beiden Sprecherinnen wurden auch von den HörerInnen als potentielle Imitatorinnen erkannt. Für den Parameter F2 zeigten zwei andere Sprecherinnen eine akustische Annäherung an die Modellsprecherin (EE, PS). Keine dieser Sprecherinnen zählte auch in der perceptiven Beurteilung der HörerInnen unter den potentiellen Imitatorinnen. Genauere Einblicke in die akustische Veränderung der *Shadowing*-Äußerungen soll die akustische Trendanalyse aller Äußerungen (Wiederholungen 1 bis 6) geben.

2.3.2 Akustische Trendanalyse

Alle Äußerungen aus dem *Shadowing*, das heißt nicht nur die erste und sechste Wiederholung, sondern erste bis sechste Wiederholung, wurden in Bezug auf die Grundfrequenz und F2 analysiert, um die genaue Veränderung oder Angleichung nachvollziehen zu können.

2.3.2.1 Ergebnis der akustischen Trendanalyse: F0

Im *Shadowing* wurden Äußerungen einer Modellsprecherin via Kopfhörer vorgespielt, die in zufälliger Reihenfolge sechsmal wiederholt wurden. Imitation war akustisch dann sichtbar, wenn die F0-Werte der Sprecherinnen von Kondition 1 bis 6 eine positive Neigung hatten, d.h. von der ersten bis zur sechsten Wiederholung höher wurden.

Sprecherin IM musste von der Analyse ausgeschlossen werden, da sich ihre Grundfrequenz zu Beginn des Experiments (Kondition 1) nicht von der Grundfrequenz der Modellsprecherin unterschied (Tabelle 3). Somit war keine Imitation dieses Parameters möglich.

Die Neigung der F0-Werte der sechs Wiederholungen eines Wortes wurde über die sechs Wiederholungen berechnet. Kondition diente als numerische Variable:

$$(4) \quad b = \text{cov}(\text{Kondition}, F0) / \text{var}(\text{Kondition})$$

Beispielhaft ist die Neigung der F0-Werte der Sprecherin MH des Worts „Festmahl“ in Abbildung 10 dargestellt.

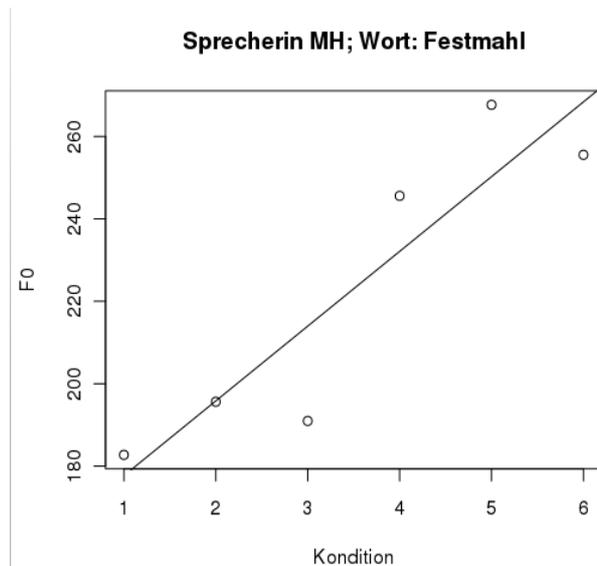


Abbildung 10: Alle sechs Wiederholungen des Worts „Festmahl“ der Sprecherin MH und die Regressionslinie, die den Neigungswert b beinhaltet.

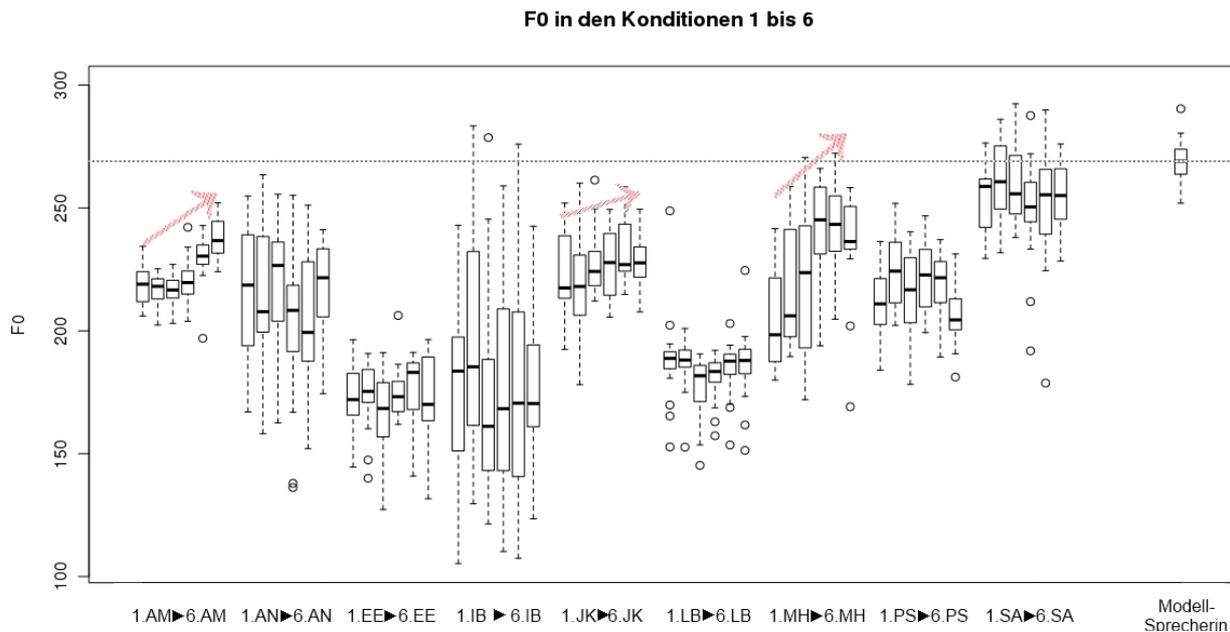


Abbildung 11: Absolute F_0 -Werte (in Hz) der Shadowing-Wiederholungen 1 bis 6 und der Modellsprecherin.

Es ergeben sich pro Sprecherin 20 Neigungswerte (20 analysierte Wörter). Die Neigung von F_0 von der ersten zur sechsten Wiederholung war für die Sprecherinnen AM, JK und MH signifikant größer als Null (AM: $t = 8.2$, $df = 19$, $p < 0.001$; JK: $t = 3.8$, $df = 19$, $p < 0.01$; MH: $t = 4.2$, $df = 19$, $p < 0.001$). Die Neigungswerte der übrigen Sprecherinnen unterschieden sich nicht signifikant von Null (Abbildungen 11, 12). Wie bei AM und MH unterschieden sich auch bei der Sprecherin JK die absolute Wortlänge der ersten und sechsten Wiederholungen nicht ($t = 1.9$, $df = 36$, $p > 0.05$).

Neigung des F0-Verlaufs

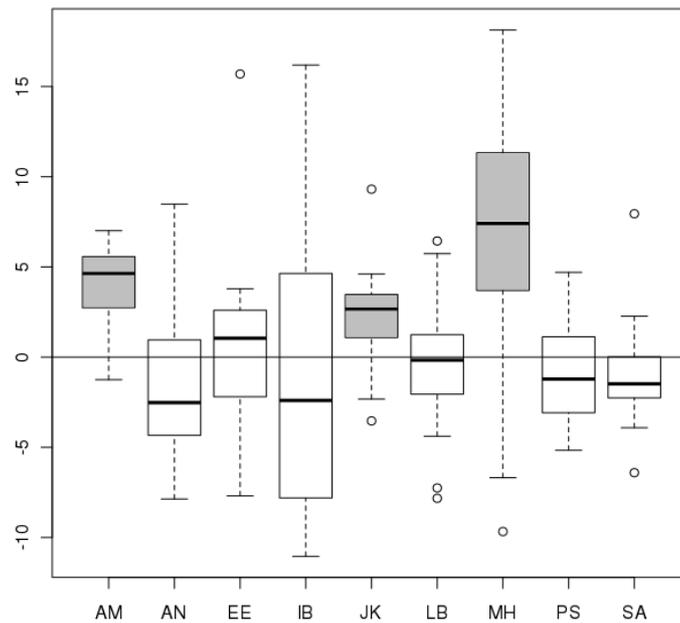


Abbildung 12: y-Achse: Neigung der F0-Werte von der ersten bis zur sechsten Wiederholung ($b=cov/var$). Grau markiert sind die signifikant von Null abweichenden Verteilungen. Die Linie markiert Null.

2.3.2.2 Ergebnis der akustischen Trendanalyse: F2

Es fanden nur die (zeitlich mittleren 20 % der) F2-Werte des Zielvokals /a/ Betrachtung, da nur für diesen Wert alle Sprecherinnen in der ersten *Shadowing*-Wiederholung signifikant unterschiedliche (tiefere) Werte als die Modellsprecherin produzierten (siehe Anhang 2: Unterschiede der Formantfrequenzen).

F2 in den Konditionen 1 bis 6 für alle 10 Sprecherinnen

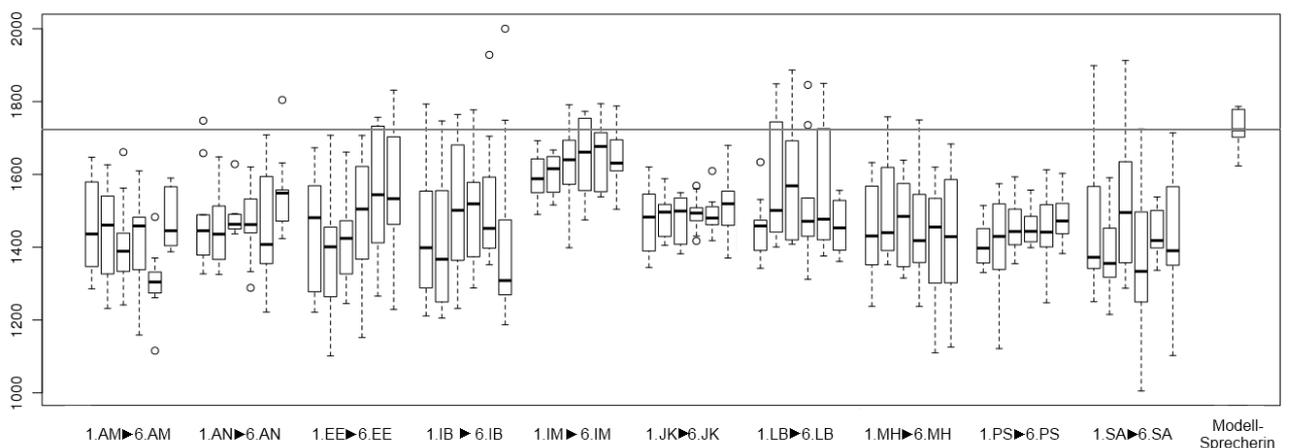


Abbildung 13: F2-Werte [Hz] des Vokals /a/ (zeitlich mittlere 20 %) aller sechs Wiederholungen.

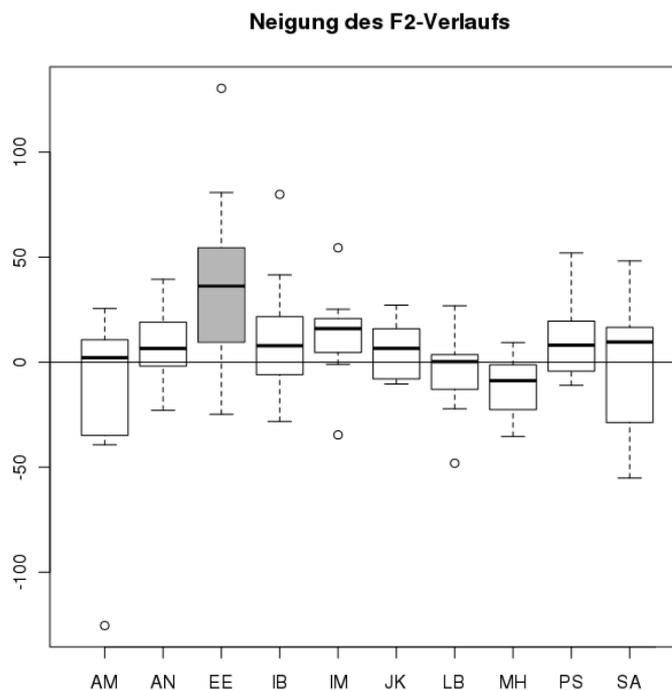


Abbildung 14: y-Achse: Neigung der F2-Werte von der ersten bis zur sechsten Wiederholung ($b = \text{cov}/\text{var}$). Grau markiert ist die signifikant von Null abweichende Verteilung der Sprecherin EE. Die Linie markiert Null.

Die Neigung der akustischen Veränderung der F2-Werte wurde, wie für F0 beschrieben, berechnet.

Wie in den Abbildungen 13 und 14 erkennbar, war die Neigung der F2-Werte über die sechs Wiederholungen und 20 Wörter nur bei Sprecherin EE signifikant verschieden von Null ($t = 2.9$, $df = 9$, $p < 0.05$).

2.3.2.3 Diskussion der akustischen Trendanalyse

In der akustischen Trendanalyse wurde der Verlauf der akustischen Parameter über den ganzen Zeitraum des *Shadowings* berücksichtigt. Der Neigungswert für jeweils ein Wort ist positiv, wenn über die zeitliche Abfolge der Äußerungen ein Parameter steigende Werte aufweist.

Innerhalb der Auswertung von F0 konnten so drei potentielle Imitatorinnen identifiziert werden. Deckungsgleich mit dem Ergebnis der Auswertung nur der ersten und sechsten Äußerungen, betrifft dies die Sprecherinnen AM und MH. Zusätzlich konnte in den Äußerungen der Sprecherin JK Annäherung sichtbar gemacht werden. In diesem Fall stellt die akustische Trendanalyse eine sensiblere Auswertungsmöglichkeit dar.

Bei der Auswertung von F2 zeigt sich für die Identifikation von potentiellen Imitatorinnen das gegenteilige Bild: Während bei dem Vergleich nur der ersten und

sechsten Äußerung zwei Imitatorinnen identifiziert wurden, ist es in der Trendanalyse nur eine der beiden (EE).

2.3.2.4 Diskussion der akustischen Analysen

Zusammenfassend findet sich in den Parametern F0 und F2 bei fünf der zehn Sprecherinnen akustische Annäherung (Tabelle 4). Die relative Vokaldauer ist kein funktionaler Parameter bei der Verwendung des *Shadowing*-Paradigmas. Deshalb werden nur die Ergebnisse des Durchschnitts der globalen Grundfrequenz des gesamten Worts (F0) und der zweite Formant des betonten Vokals (F2) als Imitationsparameter diskutiert.

Im Vergleich der Trendanalyse und der 1-vs.-6-Analyse zeigte sich, dass sowohl eine Sprecherin in der 1-vs.-6-Analyse nicht als sich annähernd erkannt wurde, die in der Trendanalyse eine Annäherung zeigte (JK), wie auch eine Sprecherin (PS) im Vergleich der ersten versus sechsten *Shadowing*-Wiederholung Imitation vermuten ließ, jedoch keine signifikant positive Neigung in der Trendanalyse aufwies. Die akustische Annäherung der Sprecherinnen AM, EE und MH ist in beiden Analyseformen präsent, AM und MH wurden auch perceptiv von HörerInnen als sich annähernd erkannt. Damit lassen diese zwei Sprecherinnen die stärkste Imitation vermuten.

	F0	F2
1-vs.-6-Analyse	AM, MH	EE, PS
Trendanalyse	AM, JK, MH	EE

Tabelle 4: Sprecherinnen, die eine signifikante akustische Annäherung an die Modellsprecherin zeigen (nach akustischem Parameter und analytischem Vorgehen).

Interessant ist, dass die Sprecherinnen AM und MH bei der Auswertung der Grundfrequenz eine deutliche Annäherung zeigten, jedoch weder AM noch MH beim zweiten Formanten. In Bezug auf F2 gilt das ebenso für die Sprecherin EE. Dies führt zur Vermutung, dass die Sprecherinnen unterschiedliche Parameter imitierten. Als mögliche Imitatorin wurde auch die Sprecherin IM im Perzeptionstest erkannt. Die Auswertung von F2 ihrer Äußerungen ergab keine Annäherung an die Modellsprecherin. Der Parameter Grundfrequenz konnte nicht ausgewertet werden, da ihre Grundfrequenz zu Beginn des Experiments nicht signifikant unterschiedlich von der der Modellsprecherin war.

Darüber hinaus zeigten beide akustischen Analysen (anhand der Distanzdifferenzen) für die einzelnen Sprecherinnen entweder Annäherung oder keine Veränderung. Das heißt, die Divergenz, die in den HörerInnenbeurteilungen beobachtet wurde, fand keine Entsprechung in den akustischen Parametern. Dies kann als ein Hinweis gewertet werden, dass die Sprecherinnen, die Annäherung zeigten, tatsächlich imitierten und die Annäherung nicht auf andere Einflüsse, die von den wahrgenommenen Äußerungen

unabhängig waren, zurückzuführen war, weil solche Einflüsse auch Divergenz ermöglichen hätten müssen.

2.4 Produktion-Perzeption: Identifikationstest

Die vorangehenden Analysen haben sprecherspezifische Unterschiede in der Imitation ergeben. Ob dies mit sprecherspezifischer Perzeption korreliert, wird im folgenden Abschnitt erörtert.

Hierfür absolvierten die zehn Sprecherinnen des *Shadowing*-Experiments einen Identifikationstest über einem Vokalkontinuum. Das Ergebnis dieses Identifikationstests wird in Verbindung mit den Ergebnissen der akustischen und perzeptiven Imitationsanalyse gestellt. In Anlehnung an die Testbatterie von Jilka (2009a) wird nur ein Identifikationstest, statt eines zusätzlichen Diskriminationstest, verwendet. Jilka (2009a) verwendete zur Testung von segmentellen Äußerungseigenschaften einen Identifikationstest („Identifikation fremdländischer Akzente in Deutschen Sätzen“).

Ein Identifikationstest gibt Auskunft über den Umgang von Versuchspersonen mit Kategoriengrenzen und die perzeptive Beschaffenheit der Stimuli. Betrachtung findet hier die Trennschärfe zwischen den Vokalkategorien /ɛ/ und /a/, d.h. ob eine abrupte im Gegensatz zu einer graduellen Kategoriengrenze besteht. Ein steiler Übergang zwischen zwei Vokalkategorien zeigt einen perzeptiv stärker definierten Kategorienbereich und deutet auf eine hohe Beständigkeit und Sicherheit in der perzeptiven Kategorisierung der Vokale hin. Die flexiblere Kategorisierung von Vokalen, also ein flacherer Verlauf der Identifikationskurve zwischen zwei Vokalkategorien, kann im Zusammenhang mit Flexibilität in der Produktion stehen, die Imitation also erst ermöglicht. Infolgedessen besteht die Erwartung, dass die in den vorangegangenen Abschnitten (2.2; 2.3) als potentielle Imitatorinnen identifizierten Sprecherinnen eine flachere Identifikationskurve aufweisen als die Sprecherinnen, die in keiner Analyse Annäherung zeigten.

2.4.1 Methodik

2.4.1.1 Versuchspersonen

Es nahmen alle 10 Sprecherinnen aus dem *Shadowing*-Experiment teil.

2.4.1.2 Stimuli

Es wurde ein Kontinuum von /vɛlt/ („Welt“) zu /valt/ („Wald“) erstellt. Hierfür wurden beide Wörter achtmal, gesprochen von der Modellsprecherin, aufgenommen (siehe Methodik „Produktionsexperiment“), mit dem Emu Software System (Harrington, 2010) gelabelt und es wurden die Formantfrequenzen berechnet (*Begin Time: -1.0; End Time: -1.0; Nominal F1: 500Hz; Prediction Order: 0; ince/decr: 0; Numer of Formants: 4;*

Preemphasis Factor: 0; Window Shift: 5.0; Window Size: 25.0; Window Function: BLACKMAN). Jeweils eine natürliche, deutliche Äußerung wurde ausgewählt und die durchschnittliche Frequenz des ersten und zweiten Formanten der Vokale /ɛ/ und /a/ in zehn zeitlichen Intervallen gemessen. Zwischen den zehn Intervallen der Vokale wurden 13 äquidistante Schritte in Bark berechnet (siehe Anhang 3: Formantwerte der /ɛ/-/a/-Manipulation). Anhand dieser Werte wurden, ausgehend von der /vɛlt/-Aufnahme, 14 Vokale mit AKUSTYK (einem akustischen Analysesoftware Paket, das als PlugIn für Praat (Boersma, 2001) designt ist: www.bartus.org) erstellt. Diese ersetzen den Vokal in der Aufnahme von /vɛlt/. Nur die Stimuli 1 bis 12 wurden verwendet, da der zwölfte Stimulus bereits als klares /a/ gelten kann und die Kategoriengrenze so zentraler im Kontinuum liegt (nach der Wahrnehmung zweier geschulter HörerInnen).

Erster und letzter Stimulus des /ɛ/-/a/-Kontinuums

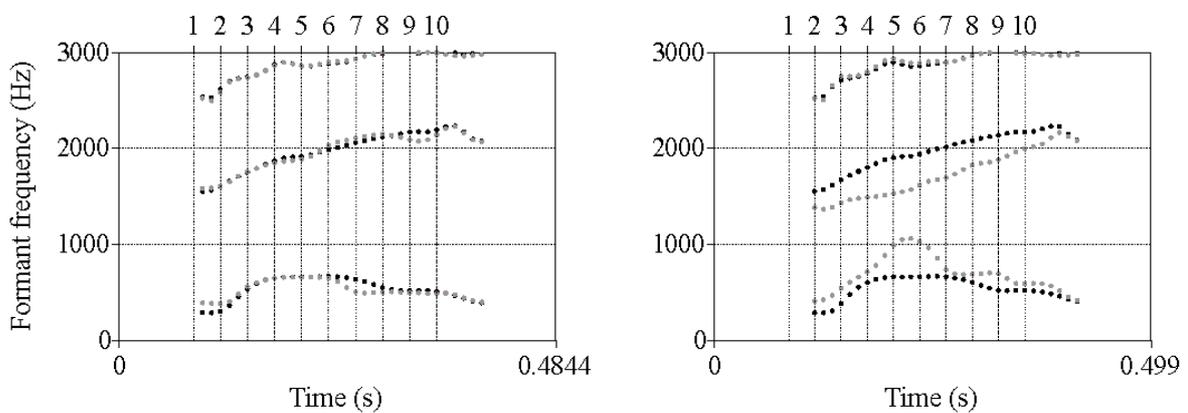


Abbildung 15: Formantverläufe des der Manipulation zugrunde liegenden /ɛ/-Vokals (schwarze Punkte in beiden Darstellungen) sowie der Vokale des Stimulus 1 (links: /ɛ/) und 12 (rechts: /a/) für das Kontinuum von /vɛlt/ zu /valt/ (graue Punkte). Stimulus 1 weist nur sehr leichte Formantveränderungen aufgrund der Resynthese auf, Stimulus 12 hat einen gesenkten F2 und dynamisch erhöhten F1.

Die Abbildung 15 zeigt den Formantverlauf innerhalb der zehn Intervalle des ursprünglichen Vokals /ɛ/ (schwarze Punkte) und die geänderten Formantverläufe durch die Resynthese (graue Punkte) in den Endpunkten /ɛ/ (Stimulus 1) und /a/ (Stimulus 12) des Kontinuums. Die Schrittgrößen der Frequenzveränderung war innerhalb der Intervalle und Formanten unterschiedlich groß, um zu gewährleisten, dass F2 relativ parallel zum steigenden Verlauf in /vɛlt/ erhöht wird und gleichzeitig der stark steigend-fallende F1 von /valt/ erreicht wurde.

2.4.1.3 Experimentelles Paradigma

In dieser Studie liegt der Fokus auf der Zuverlässigkeit der Versuchspersonen bei der Zuordnung eines Lauts zu einer Klasse, d.h. auf der Neigung der individuellen Identifikationskurven.

2.4.1.4 Durchführung

Im Anschluss an das *Shadowing* wurde der Identifikationstest durchgeführt. Hierfür wurde die Sprachverarbeitungssoftware *Praat* (Boersma, 2001) und das Welt-Wald-Kontinuum (siehe Abschnitt 2.4.1.2: Stimuli) verwendet. Die 12 Stimuli wurden über geschlossene Kopfhörer (beyerdynamic DT 770) zehnmal in zufälliger Reihenfolge einzeln vorgespielt. Die Sprecherinnen markierten ihre Entscheidung durch einen Klick auf Flächen auf dem Bildschirm, die mit „Welt“ (für /ɛ/) oder „Wald“ (für /a/) beschriftet waren. Jeder Stimulus konnte bis zu drei mal angehört werden.

2.4.2 Ergebnisse

Zuerst wird das Ergebnis der Auswertung des Identifikationstests dargestellt (2.4.2.1) und im nächsten Schritt mit den Ergebnissen der Analyse auf Imitation in Verbindung gestellt (2.4.2.2).

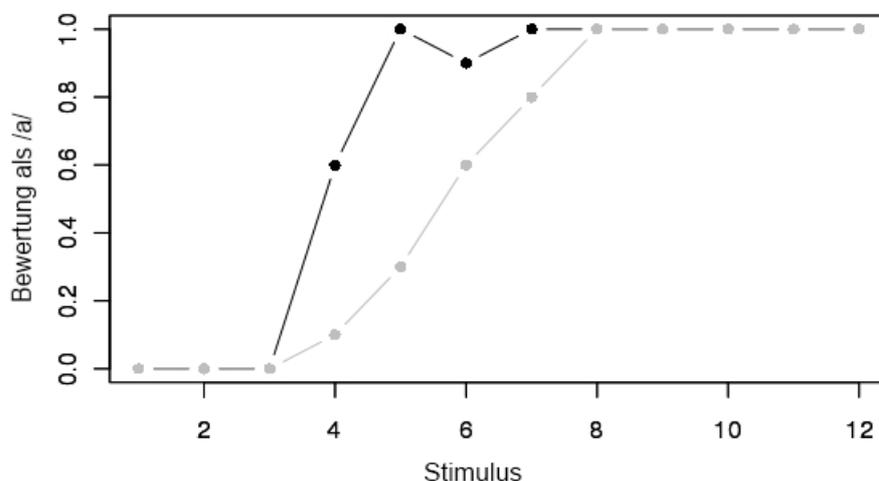


Abbildung 16: Identifikationskurve der Sprecherin AN (schwarze Punkte) im Vergleich zu AM (graue Punkte) über einem Kontinuum von /ɛ/ (1. Stimulus) zu /a/ (12. Stimulus).

2.4.2.1 Neigung der Identifikationskurve

Im Identifikationstest bewerteten die Sprecherinnen aus dem *Shadowing*-Experiment ein Kontinuum von /ɛ/ zu /a/ (in den Wörtern „Welt“ und „Wald“). Die Neigung des Verlaufs gibt Auskunft über den Umgang der Sprecherinnen mit Vokalkategorien. Wenn die Identifikationskurve horizontal verläuft (d.h. jeder Stimulus wird in 50 % der

Wiederholungen der einen, in den anderen 50% der anderen Vokalkategorie zugeordnet), besteht keine Unterscheidung der Vokale. In der Regel können Versuchspersonen die extremen Endpunkte eines sorgfältig erstellten Kontinuums zuverlässig zuordnen. Die Abbildung 16 zeigt beispielhaft die steilere Identifikationskurve der Sprecherin AN (schwarze Punkte) im Vergleich zu AM (graue Punkte).

Um die Neigung vergleichen zu können, wurde anhand einer psychometrischen Kurve pro Sprecherin ein Wert (m) berechnet, der die Steilheit der Neigung beschreibt:

$$(5) \quad p = \frac{e^{(mx+k)}}{1+e^{(mx+k)}}$$

p , die vorhergesagte Proportion von /ε/-Antworten ($0 < p < 1$), der Koeffizient m (die Neigung) und k (der Achsenabschnitt) wurden separat für jeden Sprecher berechnet. x war die Stimulusnummer 1 bis 12. Der Wertebereich von m lag zwischen 2.2 (AN) und 1.8 (AM) und war (knapp) normalverteilt ($W = 0.9$, $p = 0.06$) (Tabelle 5, Abbildung 17). Ein höherer Wert m beschreibt eine steilere Neigung.

Sprecherin	m (Neigung der Identifikationskurve)
AM	1.766
JK	1.797
MH	1.825
PS	1.850
LB	1.875
EE	1.878
IB	1.956
SA	2.128
IM	2.14
AN	2.157

Tabelle 5: Neigung der Identifikationskurven der Sprecherinnen. Die fett markierten Werte beziehen sich auf die (in der akustischen Auswertung) als potentielle Imitatorinnen identifizierten Sprecherinnen.

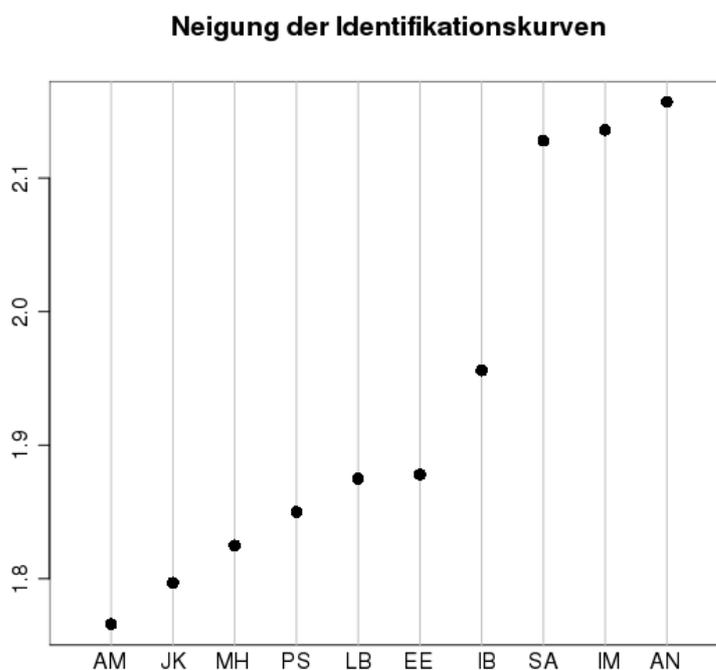


Abbildung 17: Neigung der Identifikationskurven (m : y-Achse) pro Sprecherin (x -Achse).

2.4.2.2 Zusammenhang des Identifikationstests mit der akustischen Analyse

Die akustische Analyse (Abschnitt 2.3) hat ergeben, dass sich die Sprecherinnen AM, EE, JK, MH und PS akustisch in mindestens einem Parameter (F_0 oder F_2) an die Modellsprecherin angenähert haben. Somit sind diese Sprecherinnen (mittels akustischer

Auswertung) die potentiellen Imitatorinnen. Die ausgewerteten Parameter der übrigen vier Sprecherinnen (AN, IB, LB, SA) verändern sich über den Verlauf des Experiments nicht (hier wird IM nicht mit einbezogen, da die akustische Auswertung ihrer Daten nicht vollständig ist).

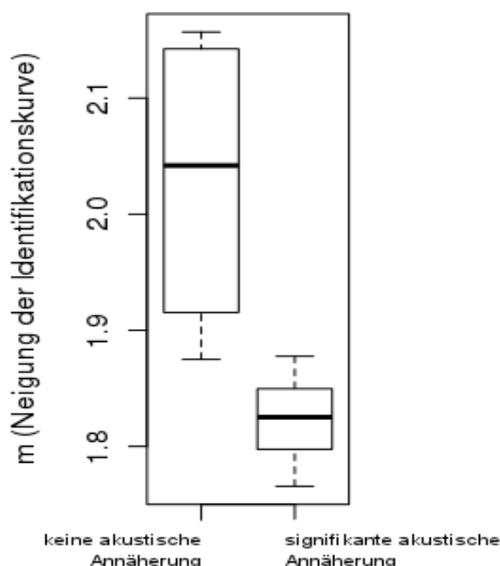


Abbildung 18: *m*-Werte des Identifikationstests im Zusammenhang mit den Sprecherinnen-Gruppen „Imitatorinnen“ vs. „Nicht-Imitatorinnen“ aus der akustischen Analyse.

Es besteht ein signifikanter Unterschied zwischen den Neigungswerten (m) der Identifikationskurve der fünf Sprecherinnen, die sich akustisch angenähert haben und den Neigungswerten der vier Sprecherinnen, deren Produktionen akustisch keine Annäherung aufwiesen ($t = 3.2$, $df = 7$, $p < 0.05$; Abbildung 18).

Dieser signifikante Unterschied in den Neigungswerten der Identifikationskurve besteht nicht für die Gruppe der Sprecherinnen, die von den HörerInnen im Vergleich zu den übrigen sieben Sprecherinnen perceptiv als Imitatorinnen wahrgenommen wurden (AM, IM, MH: siehe Abschnitt: 2.2.2).

2.4.3 Diskussion des Perzeptionstests

In Verbindung mit den Ergebnissen der akustischen Analysen zeigt sich, dass die potentiellen Imitatorinnen aus den akustischen Analysen durchschnittlich tiefere Neigungswerte aufweisen: Tiefere Neigungswerte beschreiben eine flachere Kategoriengrenze im Identifikationstest. Das heißt im Umkehrschluss, dass die Sprecherinnen, die sich der Modellsprecherin nicht angepasst haben, einen abrupteren Kategorienwechsel aufweisen.

Dieser unterschiedliche Umgang mit Vokalkategorien besteht jedoch weder für die Gruppen der perceptiv als Imitatorinnen erkannten Sprecherinnen noch für die Sprecherinnen, die in allen Analyseverfahren (perceptiv, akustisch 1. vs. 6 und Trendanalyse) potentielle Imitatorinnen waren.

Dieses Ergebnis legt nahe, dass weniger die Identifikation der Imitatorinnen erstrebenswert ist, um sie von den "anderen" Sprecherinnen zu trennen, als zu erkennen, dass es unter den Versuchspersonen in diesem Experiment Sprecherinnen gibt, die weder akustisch noch anhand der Hörerbeurteilungen Annäherung zeigen. Es ist möglich, dass die Stabilität der Produktionen dieser Sprecherinnen außergewöhnlich ist gegenüber

den Produktionen derer, die in einzelnen oder mehreren Parametern mehr oder weniger starke Annäherung aufweisen. Auch die Unterscheidbarkeit der nie-imitierenden gegenüber den etwas- oder stark-imitierenden Sprecherinnen anhand der Vokalkategorisierung deutet darauf hin.

2.5 Abschließende Zusammenfassung und Diskussion der Ergebnisse

In dieser Studie wurde ein *Shadowing*-Experiment durchgeführt, in dem zehn Sprecherinnen mit hyperartikulierter Sprache konfrontiert wurden. Alle Sprecherinnen absolvierten die exakt gleiche Aufgabe, um bekannte Einflussgrößen auf die spontane phonetische Imitation, wie Vokalselektivität (Babel, 2012), Worthäufigkeit (Goldinger, 1998) oder Interaktionsrolle (Pardo, Cajori Jay & Kraus, 2010) konstant zu halten. Sozio-kognitive Aspekte, z.B. individuelle Unterschiede zwischen den Sprecherinnen in Bezug auf das Gefallen an der *Shadowing*-Aufgabe selbst oder der Stimme bzw. Sprechereigenschaften der Modellsprecherin (speziell der hyperartikulierten Produktion), können einen Einfluss auf Imitation haben. Ein solcher Einfluss auf die vorliegenden Ergebnisse kann nicht ausgeschlossen werden, jedoch deutet darauf nichts hin.

Primäres Ziel dieser Studie war zu testen, ob sich die Stärke von Imitation zwischen SprecherInnen grundsätzlich unterscheidet und ob sich dies auf Unterschiede in der Sprachwahrnehmung zurückführen lässt.

Um Imitation zu messen, wurde jeweils die erste und letzte Äußerung eines Zielworts aus dem *Shadowing* von 20 HörerInnen mit der Äußerung der Modellsprecherin verglichen und auf ihre Ähnlichkeit geprüft. Für drei der zehn Sprecherinnen konnten HörerInnen die letzte *Shadowing*-Äußerung signifikant häufiger als ähnlicher zur Äußerung der Modellsprecherin erkennen. Bei den übrigen sieben Sprecherinnen ließ sich entweder keine Ähnlichkeit - weder zwischen der ersten noch zwischen der letzten und der Modellsprecherin - erkennen, oder es wurde nicht bevorzugt die letzte Äußerung als ähnlicher erkannt, womit von den HörerInnen keine Imitation festgestellt werden konnte. Zusammenfassend ergab der Perzeptionstest, dass drei Sprecherinnen die Äußerungen im Verlauf des Experiments wahrnehmbar mit größerer Ähnlichkeit zur Modellsprecherin produziert haben, sieben Sprecherinnen nicht.

Um nachvollziehen zu können, auf welcher Grundlage die HörerInnen die Ähnlichkeit erkannt haben, wurden die Äußerungen der Sprecherinnen akustisch analysiert. Hierbei scheint die Grundfrequenzänderung die größte Rolle zu spielen. Maße der Dauer (speziell die relative Vokaldauer) konnten in diesem Versuchsaufbau nicht ausgewertet werden, da die häufige Wiederholung die Erhöhung der Sprechgeschwindigkeit begünstigt und

Segmente eines Worts nicht linear verkürzt werden (Peterson & Lehiste, 1960; Port, 1981).

Es ist denkbar, dass die Sprecherinnen unterschiedliche Parameter präferiert stärker imitierten. Entsprechend kann eine solche Präferenz auch für die Entscheidung von wahrgenommener Ähnlichkeit der HörerInnen gelten, was den statistisch signifikanten Einfluss der HörerInnen auf die Ergebnisse erklären könnte.

Die Sprecherinnen EE, JK und PS wurden akustisch, die Sprecherinnen AM und MH sowohl perceptiv als auch akustisch als potentielle Imitatorinnen erkannt. Es bestehen also sowohl akustisch, als auch anhand der HörerInnenbeurteilungen Unterschiede zwischen den Sprecherinnen in Bezug auf deren produktive Annäherung an die Modellsprecherin.

Die Hyperartikulation der Modellsprecherin schlug sich nicht in ihren F1-Werten nieder, womit dieser Parameter nicht imitiert werden konnte. Auch für die Grundfrequenz muss in folgenden Experimenten dieser Art stärker darauf geachtet werden, dass alle SprecherInnen zu Beginn des Experiments einen signifikanten Unterschied aufweisen. Mit der Verwendung nur einer Modellsprecherin oder eines Modellsprechers besteht auch die Möglichkeit, dass individuelle Abneigungen gegenüber dem Sprechstil des Modells ausschlaggebend für das imitatorische Verhalten sind. Eine Möglichkeit, diesen Einfluss zu verringern, könnte eine *Shadowing*-Aufgabe mit mehreren ModellsprecherInnen sein.

Weiter legt die Auswertung des Identifikationstests nahe, dass eine Unterscheidung der Versuchspersonen in nicht-imitierend (anhand keines Parameters) und imitierend (anhand eines oder mehrerer Parameter) sinnvoll sein kann.

Die fünf Sprecherinnen, die in den akustischen Tests als Imitatorinnen identifiziert wurden, weisen im Identifikationstest einen eher flachen Verlauf zwischen den Vokalkategorien auf. Dieses Ergebnis kann darauf hindeuten, dass SprecherInnen, die flache perzeptive Kategoriengrenzen besitzen, stärker imitieren. Im umgekehrten Falle hieße das, dass SprecherInnen, die eine steile Grenze zwischen Kategorien wahrnehmen, weniger oder nicht imitieren.

Ein kausaler Zusammenhang ist innerhalb dieses Experiments nicht zu klären: Behalten SprecherInnen klarer voneinander differenzierte Vokalkategorien, weil sie nicht imitieren oder imitieren sie nicht, weil sie kompaktere, voneinander stärker getrennte Vokalkategorien aufweisen? Möglicherweise ist es weniger zielführend, den Zusammenhang zu hinterfragen, als anzunehmen, dass ein grundsätzlicher Unterschied in der perceptiven Sprachverarbeitung und damit der phonologischen Kategorisierung besteht, die sich auch in der (sehr) geringen Beeinflussung durch umgebende Sprache zeigt.

In der Diskussion um die Ausprägung von spontaner phonetischer Imitation, das heißt bei der Frage, ob sie automatisch ist oder nicht, besteht Uneinigkeit: Wie Pardo (2012b) zusammenfasst, stellen Theorien der Sprachverarbeitung zwar unterschiedliche Strukturen zur Verfügung, um Imitation bzw. Konvergenz zu erklären, jedoch wird sowohl bei statistischen wie auch artikulatorischen Erklärungsmodellen von automatischen Prozessen ausgegangen. Ein artikulatorischer Erklärungsansatz für spontane Imitation ist, wie einleitend bereits besprochen, die Motortheorie (Lieberman & Mattingly, 1985). Hierbei wird in der Sprachwahrnehmung das akustische Signal in artikulatorische Bewegungsabläufe übersetzt, aus denen das intendierte artikulatorische Ziel abgeleitet wird. Die folgenden eigenen artikulatorischen Abläufe werden durch die vorher wahrgenommenen Abläufe beeinflusst. Ein solcher Zusammenhang müsste jedoch immer auftreten, sobald fremde artikulatorische Abläufe wahrgenommen werden und auch immer mit der gleichen Beeinflussungsstärke, wenn spontane phonetische Imitation automatisch ablaufen sollte. Diese Zuverlässigkeit ist jedoch nicht beobachtbar – vielmehr haben soziale Bedingungen wie Interaktionsrolle (Pardo, 2006) und Sympathie (Babel, 2012) einen Einfluss. Eine Erklärung für spontane Imitation, die rein automatische Prozesse beinhaltet, ist also nicht tragbar. Das gleiche Problem tritt bei statistischen Modellen wie der Exemplartheorie auf: ein wahrgenommenes Exemplar, welches sich von den bereits wahrgenommen Exemplaren unterscheidet, soll den Prototyp und damit das artikulatorische Ziel generell und stets gleich beeinflussen. Auch dieses Modell ist mit Phänomenen, die Imitation verstärken oder schwächen, nicht kompatibel.

Spontane Imitation kann also kein automatischer Prozess sein. Dennoch gilt sie nicht als bewusst beeinflussbarer Mechanismus (Babel, 2010; Fowler, Brown, Sabadini & Wehling, 2003; Pardo, 2012b).

Hier wurden akustisch gemessene Imitationsstärken mit der Gradualität des perceptiven Kategorienübergangs in Verbindung gesetzt. Es könnte sein, dass SprecherInnen einen unterschiedlichen Umgang mit perceptiven Lautkategorien besitzen, der die Produktion unterschiedlich beeinflussbar macht. Diese Möglichkeit ließe sich sowohl in die Motor wie auch die Exemplartheorie integrieren. Die grundsätzlichen Mechanismen können wie beschrieben ablaufen (siehe Abschnitt 2.1: Einleitung), jedoch mit einer zusätzlichen, und zwar individuellen, Gewichtung.

In Bezug auf Einschätzungen von phonetischem Talent, könnte dieses Ergebnis nahelegen, dass das zuverlässige Kategorisieren von Vokalen eine Fähigkeit sein kann, die für die Trennung in mehr oder weniger talentiert eine Rolle spielt, wobei in der Studie von Lewandowski (2012) die talentierten SprecherInnen stärker imitieren, also eine flachere Kategoriengrenze für höheres phonetisches Talent sprechen könnte.

Dieses Ergebnis bietet für die Verbindung zwischen Perzeption und Produktion einen Ansatz, der thematisiert, dass es individuell unterschiedliche Strategien in der Sprachverarbeitung geben kann.

3. Spontane Imitation über die Position im Wort

Zusammenfassung

Die folgende Studie testete die Generalisierung eines Lautmerkmals über die Position im Wort durch Imitation, indem ein Experiment, bestehend aus einem Leseblock, einem Hörblock und der Wiederholung des Leseblocks, durchgeführt wurde (Nielsen, 2011b). Einführend werden die Schwerpunkte von Imitationsforschung, die Exemplartheorie, perzeptives Lernen und Lautwandel im Hinblick auf die Fragestellung beschrieben. Die gefundene Imitation legt nahe, dass (1.) die im Lexikon gespeicherte, subphonemische Information nicht an die Position im Wort gebunden ist und (2.) bei der Imitation von verlängerter Aspiration an initialer Position im Wort nicht initial, sondern final stärker imitiert wird, das heißt, dass übereinstimmende Position im Wort kein Vorteil für Imitation sein muss. Diese Ergebnisse werden exemplartheoretisch in die aktuelle Forschung zu Imitation, mentales Lexikon sowie Lautwandel eingefügt und diskutiert.

3.1 Einleitung

Der Begriff „Spontane phonetische Imitation“ bezeichnet den Mechanismus, wenn SprecherInnen die Art ihrer sprachlichen Produktionen an die Sprechereignisse angleichen, die sie wahrnehmen. Imitation ist in einer Vielzahl von experimentellen Szenarien beobachtet worden: bei direktem Nachsprechen von einzelnen Wörtern (*Shadowing*; z.B. Babel, 2012; Goldinger, 1998), in Zielwörtern innerhalb kurzer Phrasen (Delvaux & Soquet, 2007) oder in sozialen Interaktionen (Pardo, 2006, 2012a). In diesen Studien blieben die SprecherInnen uninformiert über die Untersuchung von Imitation, was mit der Präzisierung „spontane“ Imitation zum Ausdruck kommt. Die vorliegende Forschung fokussiert spontane Imitation, die der Einfachheit halber dennoch als Imitation bezeichnet wird. In der Behandlung von geforderter Imitation wird dieser Unterschied ausdrücklich genannt.

Über unterschiedliche experimentelle Aufbauten lässt sich der Prozess von Imitation an verschiedenen „Stationen“ erforschen, hier bezeichnet als zum einen die Voraussetzungen, die die Rahmenbedingungen für Imitation beinhalten und zum anderen die Auswirkungen, die die Stärke bzw. Ausprägung von Imitation beschreiben.

Die Voraussetzungen der möglicherweise zu Imitation führenden Situationen sind vom Sprechereignis im phonetischen bzw. linguistischen Sinne externe Faktoren und werden variiert, um Einblicke in die notwendigen oder hemmenden Einflüsse für Imitation zu erhalten. Dies trifft bei der von Babel (2011) gezeigten sozialen Selektivität zu: je nach Sprechereigenschaft, werden phonetische Ausprägungen stärker oder schwächer imitiert. Während eines *Shadowing-Tasks* wurde SprecherInnen eines von zwei digitalen Fotos,

das entweder einen Mann mit schwarzer oder weißer Hautfarbe darstellte, gezeigt. Ein impliziter Assoziationstest untersuchte die Einstellung der Versuchspersonen gegenüber Menschen mit schwarzer oder weißer Hautfarbe. Babel konnte zeigen, dass die Versuchspersonen, die die schwarze Hautfarbe mit negativen Bedeutungen assoziierten, schwächere Imitation zeigten, wenn ihnen suggeriert wurde, der Sprecher sei ein schwarzer Mann, als die Versuchspersonen, die die schwarze Hautfarbe nicht negativ assoziierten. Eine positive Assoziation mit dem Sprecher verstärkte also Imitation bzw. eine negative Assoziation mit dem Sprecher hemmte Imitation. Pardo (2006) zeigte, dass die Imitationsstärke von der Interaktionsrolle, die die Sprechenden einnehmen, und deren Geschlecht beeinflusst wurde: Sprecherinnen imitierten stärker, wenn sie in einer Interaktionssituation (*Map Task*) Anweisungen empfangen, wohingegen männliche Sprecher stärker imitierten, wenn sie Anweisungen erteilten. Beide Ergebnisse legen nahe, dass Imitation nicht automatisch geschieht, das heißt unbeeinflussbar und immer gleich ist, da bestimmte Rahmenbedingungen die Imitationsstärke beeinflussen können und die phonetische Variation folglich nicht rein vom akustischen Input abhängig ist.

Die Auswirkungen von Imitation werden sichtbar, wenn interne Faktoren des Sprechereignisses verglichen werden. Das kann die Segmentebene betreffen, z.B. hohe im Vergleich zu tiefen Vokalen, Imitation an unterschiedlichen Artikulationsstellen oder die Wortebene anhand der Worthäufigkeit (Frequenz). Goldinger (1998; reproduziert in einem ähnlichen Versuchsaufbau von Nielsen, 2011b) bestätigte die exemplartheoretische Vorhersage, dass eine stärkere Imitation von Wörtern mit geringer Worthäufigkeit (d.h. seltenerem Auftreten im Sprachgebrauch) auftritt. Die gleiche negative Korrelation zwischen Auftretenshäufigkeit und Imitation findet sich auch bei der Phonotaktik wieder (Nye & Fowler, 2003): die Realisierung von Wörtern mit phonotaktisch häufigen Lautkombinationen werden weniger stark imitiert als Wörter mit phonotaktisch seltenen Lautkombinationen. Babel (2012) zeigt Vokalselektivität: tiefe Vokale erfahren stärkere Imitation als hohe Vokale. Dieses Ergebnis führt sie auf die stärkere Variabilität im F1-Raum zwischen betonter und unbetonter Position von tiefen Vokalen, im Vergleich zu hohen, zurück. Nielsen (2011b) fügt den Erkenntnissen dieser Ergebnisse eine weitere Ebene hinzu, indem sie die Imitation eines Lautmerkmals an unterschiedlichen Artikulationsstellen testet. Die Besonderheit liegt darin, dass ein Lautmerkmal Imitation an einer Artikulationsstelle aufweist, an der es in der besonderen Ausprägung vorher nicht wahrgenommen wurde. Das lässt sich durch einen geblockten Versuchsaufbau erreichen, bei dem die Testwörter von den SprecherInnen zuerst gelesen werden, dann Wörter vorgespielt werden, die nur teilweise mit den gelesenen Wörtern übereinstimmen, und abschließend die Zielwörter ein zweites Mal gelesen werden. In diesem konkreten Test verlängerte Nielsen die VOT in /p/-initialen Wörtern. Diese Längung der VOT konnte

Nielsen in der zweiten Aufnahme der /k/-initialen Wörter feststellen, wenn auch geringer als in /p/, obwohl keine /k/-initialen Wörter Bestandteil der vorgespielten Wörter waren. Damit zeigte sie, dass phonetische (also interne) Bestandteile einen unterschiedlichen Einfluss auf Imitation haben können, da Imitation auch subphonemisch selektiv geschieht. Die Rahmenbedingungen müssen, um Ergebnisse über sprachinterne Faktoren zu erhalten, konstant gehalten werden.

In der vorliegenden Studie wird der Einfluss von subphonemischen Merkmalen auf die Imitation in verschiedenen Segmenten der Testwörter betrachtet. Bei diesem Experiment kam der geblockte Aufbau von Nielsen (2011b) zum Einsatz, wobei im Unterschied zu Nielsen hier die Imitation an unterschiedlichen Positionen im Wort getestet wurde. Es handelt sich um eine Studie über die sprachinternen Voraussetzungen von Imitation, im Speziellen um Selektivität von der Position im Wort.

Ein exemplartheoretisches Modell der Sprachperzeption (Johnson, 1997) und der Sprachproduktion (Pierrehumbert, 2003) etabliert die dauerhaft bestehende Flexibilität und Anpassung der sprachspezifischen Systeme als Eigenschaft von Sprachperzeption und -produktion. Damit bietet sie ein, von sozialen Einflüssen freies, Erklärungsmodell für Imitation sowie die Ausbreitung von Lautwandel.

Die Anpassungsfähigkeit an die umgebende Sprache wird erreicht, indem alle Elemente wahrgenommener Sprache als Exemplare gespeichert werden. Das phonetische Lexikon besteht aus exemplartheoretischer Sicht aus den detaillierten Informationen (Sprechereigenschaften, (ko-)artikulatorische oder dialektale Variation, usw.) von jeder wahrgenommenen Äußerung: Alle gesammelten Exemplare einer sprachlichen Einheit bilden den Kenntnishintergrund zur eigenen Produktion dieser Einheit. Die Kategorisierung (Wort, Silbe, Segment, Merkmal) erfolgt durch den Vergleich der Ähnlichkeiten zu allen gespeicherten Exemplaren (Hintzman, 1986; Nosofsky, 1986). Quantifizierbar wird Ähnlichkeit in diesem Zusammenhang durch *Episodic traces*: eine wahrgenommene Produktion aktiviert die gespeicherten Exemplare je nach bestehender Ähnlichkeit unterschiedlich stark. Bei stärkerer Ähnlichkeit zu einer größeren Anzahl von Exemplaren spricht man von der Aktivierung starker und vieler *Episodic traces*, bei schwacher Ähnlichkeit zu einer kleineren Anzahl von Exemplaren von der Aktivierung schwacher und weniger *Episodic traces*. Diese unterschiedlich starke Aktivierung von *Episodic traces* lässt sich in der Sprachproduktion zeigen: wenn Personen in einem *Shadowing-Task* (direktes, schnelles Nachsprechen vorgespielter einzelner Wörter) Wörter mit einer höheren Auftretenshäufigkeit produzieren, besteht weniger Imitation der äußerungsspezifischen Bestandteile als wenn Wörter mit einer geringeren Auftretenshäufigkeit produziert werden (Goldinger, 1998). Da es von Wörtern mit einer

geringen Auftretenshäufigkeit weniger gespeicherte Exemplare gibt, evoziert das aktuell wahrgenommene Wort weniger *Episodic traces* und übt, wegen des schwächeren Einflusses der bereits gespeicherten Exemplare, einen größeren Einfluss auf die Gesamtheit der dem Wort zugeordneten Exemplare aus. Ein Wort, das bereits sehr häufig wahrgenommen wurde, aktiviert eine Vielzahl von *Episodic traces*, die die Produktion modifizieren und somit den Einfluss des wahrgenommenen Exemplars vergleichsweise verringern. Darüber hinaus haben zeitliche Faktoren Bedeutung für die Aktivierung von *Episodic traces*: Je zeitlich näher ein Exemplar gespeichert wurde, umso stärker sind die *Episodic traces* aktiviert, die Einfluss auf Perzeption und Produktion haben. Das bedeutet im Umkehrschluss nicht, dass gespeicherte Exemplare mit der Zeit keine Aktivierung mehr aufweisen, sondern nur noch sehr geringe.

Diese, durch Häufigkeit des Auftretens bestimmte, Beeinflussung der Perzeption und der Produktion besteht auf lexikalischer Ebene bei Wörtern, wie von Goldinger gezeigt, und auch auf phonotaktischer Ebene (Nye & Fowler, 2003).

Kernstück der Exemplartheorie ist neben der Plastizität von Sprachproduktion und -perzeption der Detaillierungsgrad der gespeicherten Exemplare. Sprachereignisse werden in ihrer Gesamtheit gespeichert und nicht nur das Produkt einer perzeptiven Dekodierung. Die Kenntnis verschiedener nicht-phonemischer Äußerungsbestandteile wird gezeigt, wenn dieses Wissen die Wahrnehmung modifiziert: Äußerungen vom gleichen Sprecher (Goldinger, 1996), mit der gleichen Intonationskontur und Grundfrequenz (Church & Schacter, 1994) oder im gleichen Sprachstil (McLennan, Luce & Charles-Luce, 2003), werden besser (d.h. korrekter) und schneller erkannt.

Die Exemplartheorie bietet ein probabilistisches Erklärungssystem für die Beeinflussung von Perzeption sowie die Produktion durch aktuelle Sprachereignisse: ein stärkerer Einfluss (und daraus folgend stärkere Imitation) wird dann erwartet, wenn mehr Gemeinsamkeiten mit dem/den kürzlich wahrgenommenen Lautmerkmal/en besteht, z.B. in Form von Artikulationsstelle, Sprecher, Wort oder Position im Wort.

Konsistent mit der statistischen Beeinflussung durch Auftretenshäufigkeiten, die in der Exemplartheorie beschrieben wird, sind die Ergebnisse zu perzeptivem Lernen. Wahrgenommene neue Exemplare führen in Verbindung mit lexikalischem Wissen zu einer Erweiterung der perzeptiven Kategorie:

Norris, McQueen & Cutler (2003) zeigten die Anpassung der Wahrnehmung von phonologischen Kategorien an die aktuelle Sprachumgebung in Abhängigkeit von lexikalischer Information. Hierbei erhielt eine Gruppe der HörerInnen ein Set aus /f/-finalen Wörtern, deren /f/-Anteil durch einen perzeptiv zwischen /f/ und /s/ liegenden Frikativ ersetzt wurde und ein Set aus /s/-finalen Wörtern (die keiner Manipulation unterzogen

worden waren). Die andere Gruppe erhielt die /f/-finalen Wörter ohne Manipulation, aber die /s/-finalen Wörter endeten auf einen ambigen Frikativ des /f-s/-Kontinuums. Dieses perzeptive Training führte zu einer Verschiebung der Kategoriengrenze im /f-s/-Kontinuum. Beide Gruppen gaben jeweils häufiger an, den Frikativ zu hören, der im Training durch einen ambigen Frikativ ersetzt worden war. Die perzeptive Kategorie des ersetzten Lauts wurde also angepasst bzw. erweitert, indem durch lexikalisches Wissen ein unübliches Exemplar hinzugefügt wurde.

Dies geschieht nicht nur auf Wortebene, sondern auch sublexikalisch und subphonemisch: Die Erweiterung der Kategorie durch perzeptives Lernen wird auch auf Wörter mit dem gleichen Phonem (sublexikalisch: McQueen, Cutler & Norris, 2006) und Laute des gleichen Artikulationsmodus (subphonemisch: Kraljic & Samuel, 2006) generalisiert. Der Mechanismus hierfür verläuft automatisch (McQueen et al., 2006). Perzeptives Lernen äußert sich in der Produktion, wenn imitierte Merkmale auf andere Wörter (sublexikalische Imitation) und andere Phoneme (subphonemische Imitation) generalisiert werden (Nielsen, 2011b).

Perzeptiv ist die Differenziertheit in Ebenen (z.B. Wort, Phonem, Merkmal) relativ gut beschrieben. Aufschlussreich für die Verbindung zwischen Perzeption, Lexikon und Produktion sind die Ergebnisse von Nielsen (2011b): durch Imitation wird das Bestehen der drei Ebenen Wort, Phonem und Merkmal gezeigt und durch die Generalisierung zwischen Wörtern und Phonemen auch deren Flexibilität. Ob diese Flexibilität auch zwischen Wortpositionen besteht, testet die vorliegende Studie.

Phonologischer Lautwandel ist ergiebig dokumentiert: Ohala (1981) führt eine Vielzahl von Beispielen an, um seine Theorie vom Ursprung von Lautwandel zu belegen, worin dieser durch fehlerhafte Wahrnehmungen geschieht. Über die Ausbreitung von einem sogenannten „Mini-Lautwandel“, der zwischen SprecherIn und HörerIn vollzogen wird, äußert sich Ohala, dass dieser anhand sozialer Umstände geschieht, beispielsweise über Entlehnung und Imitation (Ohala, 1981, 1993).

Imitation wird als ein Mechanismus betrachtet, durch den sich Lautwandel und Dialektkonvergenz ausbreiten können, jedoch wird häufig bezweifelt, dass es sich dabei um einen (rein) sozialen Mechanismus handelt (Babel, 2011; Lindblom, 1998; Trudgill, 2008).

Phonetische Abläufe könnten nicht nur als Auslöser Lautwandel bewirken, wie von Ohala (1981) angemerkt, sondern auch an dessen Verbreitung beteiligt sein, indem SprecherInnen eine phonetische Evaluation vornehmen, wobei sie „artikulatorische Leichtigkeit“, „perzeptive Adäquatheit“, „sozialer Wert“ und „systemische Kompatibilität“ (Lindblom, 1998, S. 245, eigene Übersetzung) bewerten. Lindblom beschränkt diese

Einflüsse jedoch auf die Verbreitung eines Lautwandels. Unterschiedliche Bedeutungen müssen unterschiedlich klingen und gleichzeitig mit minimalem motorischen Einsatz umgesetzt werden. Die Kontroverse dieser beiden Ziele ist eine wichtige Quelle phonetischer Variation, aus der Mitglieder einer Sprachgemeinschaft Innovationen des phonologischen Systems schöpfen. Diese Auswahl wird blind und ohne Bewusstsein für die funktionalen Konsequenzen der neuen Spracheinheit getätigt (Lindblom, 1998). Zwar bezieht Lindblom den Einfluss geringer Häufigkeit und perceptiver Auffälligkeit als unterstützende Faktoren für Imitation mit ein, jedoch auch die Wirkung von artikulatorischen Anforderungen bei der Produktion („*pronounceability constraints*“ 1998, S. 252).

Sein Modell sieht daher vor, dass sich einfache Aussprachevarianten, bei gleicher Häufigkeit und Auffälligkeit, schneller verbreiten als schwierige (Lindblom, 1998). Spontane Imitation, das heißt Imitation innerhalb zeitlicher Bereiche von ca. 30 bis 60 Minuten, kann nicht als dauerhafter Lautwandel gewertet werden, jedoch als erste Evaluierungsdurchgänge einer Aussprachevariante und somit als ein bewerteter Prozess. Spontane Imitation bietet also einen Kurzzeit-Einblick in die Ausbreitung von Lautwandel. Dies zeigen z.B. die Ergebnisse zu Vokalselektivität von Babel (2012) und die bevorzugte Imitation eines Lautmerkmals bei gleichbleibender Artikulationsstelle (Nielsen, 2011b). Ob die Position im Wort auch einen bedeutsamen Einfluss auf die Imitation eines Merkmals (Aspirationsdauer) hat, testete diese Studie. Das bezieht sich auf die Unterscheidung von Lindblom, Guion, Hura, Moon & Willerman (1995) in die Sprech-/Kommunikations-Modi *how* und *what*, das heißt *wie* wurde das Gesagte hervorgebracht und *was* wurde gesagt. Meistens ist der *Was*-Modus in der Produktion des/der Sprechenden und der Perzeption des/der Hörenden vorherrschend, da das Ziel von Äußerungen in der Regel die Verbreitung von Inhalt ist. Jedoch an den Stellen, an denen das *Was* weniger wichtig ist, beispielsweise bei hoher Redundanz und/oder Vorhersagbarkeit, kann das *Wie* stärker in den Vordergrund treten, neue Aussprachevarianten in das Lexikon aufgenommen werden und ein Lautwandel könnte folgen.

Harrington (2007) zeigte innerhalb einer Sprecherin, Queen Elisabeth II., Lautwandel als graduelle Veränderung über einen Zeitraum von 50 Jahren. Es handelt sich um einen Lautwandel, der für den Sprachraum von Großbritannien, im Speziellen *Received Pronunciation* (RP, ein regional unabhängiger Akzent der gebildeten Mittel- und Oberschicht), innerhalb der zweiten Hälfte des 20. Jahrhunderts gut beschrieben ist (Gimson, 1966; Henton, 1983; Roach & Hartman, 1997). Die schrittweise Veränderung der Produktion einer Person an die einer anderen Varietät stimmt mit exemplartheoretischen Grundlagen überein und passt zu den Ergebnissen der

Imitationsforschung, in der Angleichung, aber keine vollständige Anpassung beobachtet wird (z.B. Babel, 2012; Fowler et al., 2003; Honorof et al., 2011).

Da die Rahmenbedingungen des in dieser Studie verwendeten Versuchsablaufs möglichst konstant blieben und so keine sprachexterne Beeinflussung von Imitation getestet werden konnte und sollte, fokussierte diese Arbeit auf sprachinterne Aspekte in Bezug auf Imitation bzw. unterschiedliche Stärken von Imitation. Dabei handelte es sich um den Einfluss von unterschiedlichen Positionen im Wort. Der Exemplartheorie zu Folge sollte sich die gleiche Position im Wort verstärkend (und die wechselnde Position im Wort schwächend) auf die Imitationsstärke auswirken. Dennoch legen die Ergebnisse zu perzeptivem Lernen und die Erweiterung auf die Produktion durch Nielsen (2011b) nahe, dass eine Übertragung eines Merkmals über die Position im Wort möglich ist. Das Ergebnis dieser Studie kann somit einen Beitrag leisten, um die (mögliche) Mitwirkung von Imitation bei der Ausbreitung von Lautwandel besser zu verstehen.

Um die folgenden Hypothesen testen zu können, wurden Wörter mit verlängerter Aspiration konstruiert. Versuchspersonen absolvierten einen geblockten Versuchsablauf, indem sie eine Liste von einzelnen Wörtern lasen, daraufhin Wörter mit verlängerter Aspiration hörten und anschließend die anfängliche Liste von Wörtern noch einmal produzierten. Die Leseliste beinhaltete Wörter mit bilabialen fortis Plosiven an der initialen und finalen Position im Wort und Wörter mit velaren fortis Plosiven an der finalen Position im Wort. Die Hörliste bestand aus Wörtern mit bilabialen fortis Plosiven nur an der initialen Position im Wort, deren Aspiration manipulativ verlängert worden war. Der Vergleich der Aufnahmen der gelesenen Wörter macht die Imitationsstärke der Aspirationsdauer an den zwei verschiedenen Artikulationsstellen und Positionen im Wort sichtbar.

H1: Die verlängerte Aspiration der initialen bilabialen Plosive wird an beiden Positionen im Wort und an beiden Artikulationsstellen imitiert.

Dies wird exemplartheoretisch begründet, da zur Produktion eines aspirierten Plosivs alle Exemplare, die das Kennzeichen „Aspiration“ aufweisen, aktiviert werden, sowie auch „Aspiration“ selbst, getrennt vom Trägerlaut oder -wort (Nielsen, 2011b). Zwar ist die Aktivierung der *Episodic traces* zu den manipulierten Exemplaren mit verlängerter Aspiration stärker bei der Produktion von Plosiven mit größerer Ähnlichkeit (das heißt, mit übereinstimmender Lautposition und/oder Artikulationsstelle), aber für alle aspirierten Plosive vorhanden. Durch die zeitliche Nähe (der zweite Leseblock folgte direkt auf den Hörblock) bestand ein Einfluss der verlängerten Aspiration auf beide Wortgruppen.

Die stärkere Aktivierung von *Episodic traces* bei initialen als bei finalen Plosiven begründet die zweite Hypothese:

H2: Die verlängerte Aspiration der initialen bilabialen Plosive wird stärker initial als final imitiert.

Für die Imitationsstärke bei den untersuchten Artikulationsstellen (bilabial, velar) besteht die selbe Vorhersage wie für die Position im Wort:

H3: Die verlängerte Aspiration der initialen bilabialen Plosive wird stärker in den finalen bilabialen Plosiven als in den finalen velaren Plosiven imitiert, weil eine höhere Ähnlichkeit zwischen initial /p/ und final /p/ besteht, als zwischen initial /p/ und final /k/.

Die Vorhersagen sind kohärent mit den Ergebnissen von Nielsen (2011b), die eine Generalisierung zwischen Artikulationsstellen zeigte.

3.2 Methodik

Es wurde ein Versuchsablauf verwendet, in dem die Testwörter ein erstes Mal aufgenommen, dann teilweise übereinstimmende Wörter über Kopfhörer eingespielt und abschließend die bereits aufgenommenen Wörter ein zweites Mal aufgenommen wurden (siehe Anhang 1: Wörter der Lese- und Hörliste). Diese Methodik erlaubt, im Gegensatz zum *Shadowing*, bei dem Wörter sofort nach dem Hören nachgesprochen werden, dass sich die zu imitierenden Wörter von den gelesenen Wörtern unterscheiden. Nur so können Generalisierungen durch Imitation auftreten (Nielsen, 2011b).

Das Experiment besaß folglich zwei verschiedene Wörterlisten (siehe Anhang 4: Wörter der Hör- und Leseliste). Zum einen die Hörliste, deren Wörter angehört wurden und aus 83 /p/-initialen Wörtern und 38 Füllwörtern mit initialen Sonoranten (/l/, /m/, /n/ und /r/) bestand. Die Aufnahme und Manipulation dieser Wörter wird unter Stimulikonstruktion (siehe unten) beschrieben.

Zum anderen die Leseliste, die aus den Wörtern, die auf Imitation getestet wurden, bestand:

- 20 /p/-initiale Wörter mit initialer Betonung (5 ein- und 15 mehrsilbige), die auch Bestandteile der Hörliste sind,
- 20 /p/-initiale Wörter mit initialer Betonung (3 ein- und 17 mehrsilbige), die nicht Bestandteile der Hörliste sind,
- 20 einsilbige auslautverhärtete² /p/-finale Wörter,
- 17 /p/-finale einsilbige Wörter,
- 20 /p/-finale mehrsilbige Wörter mit nicht-finaler Betonung,

² Der Stimmhaftigkeitskontrast von Obstruenten wird im Deutschen an der domänfinalen Position neutralisiert. Orthographisch *Rat* und *Rad* sind phonologisch homophon /rat/. Dieses Phänomen wird als Auslautverhärtung bezeichnet (Kleber, John & Harrington, 2010).

- 20 /k/-finale einsilbige Wörter und
- 20 /k/-finale mehrsilbige Wörter mit nicht-finaler Betonung,
- 30 Füllwörter mit initialen Sonoranten (/l/, /m/, /n/ und /r/; nicht übereinstimmend mit den Füllwörtern der Hörliste).

Die geringere Anzahl der p-finalen Wörter gründet auf der ursprünglichen Überlegung, eine Unterscheidung in finaler Konsonant als Coda der betonten Silbe versus finaler Konsonant als Coda der unbetonten Silbe, vorzunehmen. Aufgrund der wenigen Wörter im Deutschen mit geringer Häufigkeit, die auf /p/ enden (aber nicht auslautverhärtet sind), die gleiche Silbenanzahl und die Betonung an der gleichen Silbenposition besitzen, wurde auf die Unterscheidung ein- versus mehrsilbig ausgewichen, weshalb drei zweisilbige /p/-finale Wörter („salopp“, „Gestrüpp“, „Galopp“) mit der Betonung auf der zweiten Silbe aus der Auswertung entfernt wurden. Die Differenzierung der /p/-initialen Wörter mit /p/ als Onset der betonten Silbe in ein- bzw. mehrsilbige Wörter ist im gleichen Zug im Nachhinein entstanden und wird nicht genauer untersucht.

Die verwendeten Wörter (exklusive der Füllwörter) besitzen im CELEX-Korpus (Baayen, Piepenbrock & Gulikers, 1995) eine Lemma-Mannheim Frequenz von durchschnittlich 36.7. Die Lemma-Mannheim Frequenz gibt an, wie häufig ein Lemma (das heißt ein zugrunde liegendes Wort einschließlich der verschiedenen Flexionen wie z.B. Plural oder Genitiv) im gesamten Korpus vorkommt. Der CELEX-Korpus besteht aus 6.000.000 Wörtern. Die Verteilung der Lemma-Mannheim Frequenz liegt zwischen 0 und 811 („Technik“) Einträgen.

3.2.1 Versuchspersonen

An dem Experiment nahmen 14 Sprecherinnen und Sprecher im Alter zwischen 20 und 33 Jahren teil. Alle sprachen süddeutsche Standardsprache. Keine der Versuchspersonen zeigte oder berichtete von Auffälligkeiten in der Sprachwahrnehmung oder -produktion. Der zeitliche Aufwand von 45 Minuten wurde mit 8 € pro Versuchsperson entlohnt.

3.2.2 Stimulikonstruktion

Eine weitere (phonetisch geschulte) männliche Person produzierte die Wörter der Hörliste auf zwei verschiedene Weisen als Grundlage zur Erstellung der Stimuli. Zum einen wurden alle Wörter der Hörliste (Ziel- und Füllwörter) drei Mal isoliert und in natürlicher Sprechgeschwindigkeit deutlich gelesen aufgenommen. Zum anderen wurden in einer zweiten Sitzung die Zielwörter der Hörliste, in einen Trägersatz eingebettet, drei Mal produziert. Dieser wurde pro Wort so konstruiert, dass ein Kontrastakzent auf den Zielwörtern lag, z.B: *Doch nicht Best, Pest*. Eine lange Pause vor dem Zielwort minimierte koartikulatorische Einflüsse auf den Fortisplosiv.

Von allen aufgenommenen Äußerungen (kontrastakzentuierte Zielwörter, isoliert produzierte Ziel- und Füllwörter) wurde auditiv die natürlichste Produktion ausgewählt. Gepaart pro Wort wurde der *Burst* und die längere *Aspiration* der akzentuierten Äußerungen herausgeschnitten und an der Stelle von *Burst* und *Aspiration* der isoliert artikulierte Äußerungen in diese eingefügt (Spleißing). Verwendet wurde die Sprachverarbeitungssoftware *Praat* (Boersma, 2001). So wurden natürlich klingende Äußerungen mit verlängerter VOT erzeugt.

Die Dauer von *Burst* und *Aspiration* der isolierten Äußerung gab an, ein wie großer Teil der verlängerten *Aspiration* eingefügt wurde: In den Stimuli waren *Burst* und *Aspiration* 40 ms länger als in den natürlichen Äußerungen, jedoch mindestens 100 ms insgesamt. Bei einer Dauer von *Burst* und *Aspiration* von z.B. 66 ms in der isolierten (nicht kontrastakzentuierten) Äußerung eines Wortes, wurden die ersten 106 ms (66 ms plus 40 ms) *Burst* und *Aspiration* der kontrastakzentuierten Äußerung herausgeschnitten und *Burst* und *Aspiration* der isolierten Äußerung damit ersetzt. Bei einer Dauer von 54 ms in der isolierten Äußerung wurden 100 ms aus der hyperartikulierten Äußerung geschnitten und in die isolierte Äußerung gespleißt.

Die isoliert artikulierte Äußerungen hatten eine durchschnittliche *Burst*- und *Aspirationsdauer* von 57.7 ms, die hyperartikulierten von 135.2 ms und die Stimuli von 104.4 ms.

3.2.3 Versuchsablauf

Die Wörter wurden mit der *Speech Recorder* Software (Draxler & Jänsch, 2004) visuell auf einem Monitor präsentiert, aufgenommen wurde mit einem TLM 103 Mikrofon und einer ADAT Schnittstelle zwischen einem YAMAHA O2R Mischpult und einem HP Compaq dc7800 CMT PC All (p/n GC758AV).

Die Versuchspersonen wurden in getrennten Sitzungen aufgenommen, wobei jede aus vier Aufgabenblöcken bestand. Anfänglich wurde die Liste der zu produzierenden Wörter still gelesen, um Hyperartikulation zu vermeiden und flüssiges Lesen zu begünstigen (Wright, 2003). Alle Versuchspersonen gaben an, dass die gelesenen Wörter bekannt seien. Daraufhin wurden die Wörter der Leseliste in randomisierter Reihenfolge das erste Mal produziert. Ein Aufnahmezeitraum von drei Sekunden ermöglichte ruhiges Lesen. Als dritter Aufgabenblock wurden die Stimuli und Füllwörter der Hörliste via Kopfhörer vorgespielt und als einzige Anweisung angegeben, genau zuzuhören. Als letzter Block wurde die Leseliste das zweite Mal wiederholt. Wieder wurden die Wörter in Abständen von drei Sekunden in randomisierter Reihenfolge auf dem Monitor dargestellt. Falls Versprecher oder zu langsames Lesen auftraten, wurde das betreffende Wort innerhalb

des betreffenden Leseblocks wiederholt, sodass keine Aufnahmen bzw. Wörter aus der Analyse ausgeschlossen werden mussten.

3.2.4 Auswertung

Die Sprachaufnahmen wurden mit dem Münchner Automatischen Segmentationsystem (*MAUS*, Schiel, 2004) segmentiert und die automatisch gesetzten Segmentgrenzen manuell korrigiert. Die ausgewertete Aspirationsdauer bestand hierbei aus dem Einsetzen der Verschlusslösung (des *Bursts*) und der Aspiration, die bei den initialen Plosiven mit dem Einsetzen von Periodizität endet, bei den finalen Plosiven mit dem Einsetzen von auditiv deutlich erkennbarer (Aus-)Atmung oder, beim Fehlen von Ausatmung, mit dem grundsätzlichen Ende vom Geräuschereignis. Der Einfachheit halber wird dieses Intervall als „Aspiration“ bzw. „Aspirationsdauer“ oder VOT (*voice onset time*) bezeichnet, obwohl der *Burst* eingeschlossen ist.

Die Aspirationsdauer wurde mit *Mixed Effect Model* Analysen (Baayen, 2008) mit jeweils Versuchsperson und Wort als *Random* Faktoren anhand des *Response* Faktors Aspirationsdauer als absolute Werte ausgewertet. Als *Fixed* Faktor diente immer „Zeitpunkt der Aufnahme“, wobei die erste Aufnahme in den Darstellungen mit „A“ und die zweite Aufnahme (nach der akustischen Präsentation der manipulierten Stimuli) mit „B“ markiert wurde.

Der Vergleich der absoluten Aspirationsdauer gibt Auskunft über die Veränderung zwischen den Aufnahmezeitpunkten, deren Richtung (wenn eine Änderung vorhanden war) den Darstellungen entnommen werden kann. Die Darstellungen (Boxplots) zeigen gemittelte Werte über die Gruppen von Artikulationsstelle und Position.

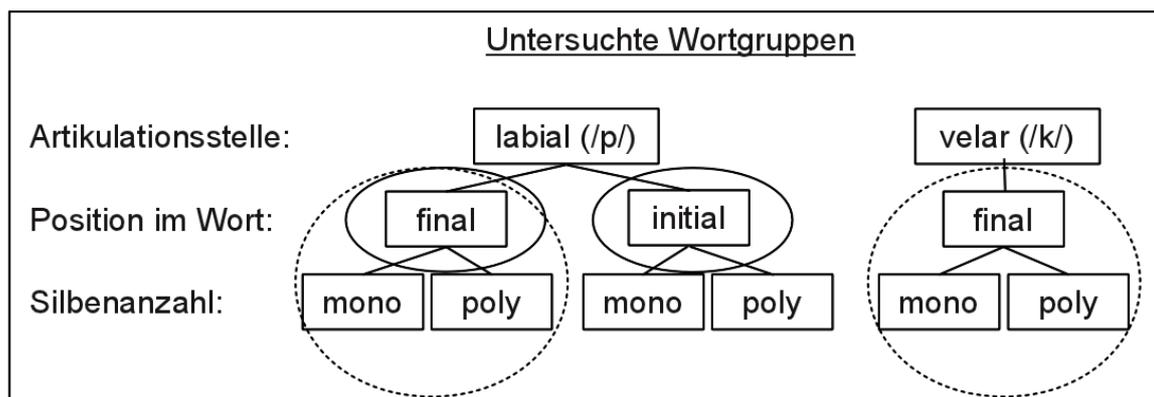


Abbildung 19: Darstellung der untersuchten Wortgruppen und die Vergleichsgruppen (Kreise) für die statistische Auswertung.

Die weiteren *Fixed* Faktoren unterschieden sich zwischen den untersuchten Wortgruppen (siehe Abbildung 19). In der Auswertung wurden einerseits die Artikulationsstellen (bilabial versus velar) an jeweils finaler Position (Abbildung 19:

gestrichelte Kreise) und andererseits die Positionen final und initial der Wörter mit /p/ (Abbildung 19: durchgängige Kreise) verglichen.

3.2.5 Ergebnisse

3.2.5.1 Vergleich der Artikulationsstellen an jeweils finaler Position

Bei der statistischen Auswertung unter Verwendung eines *Mixed Models* mit den *Random* Faktoren Versuchsperson und Wort, *Fixed* Faktoren Aufnahmezeitpunkt (A: vor Präsentation der manipulierten Stimuli, B: nach der Präsentation) und Artikulationsstelle ergab sich eine signifikante Interaktion zwischen den *Fixed* Faktoren ($X^2[1] = 4.8, p < 0.05$). Ein *Tukey* post-hoc Test zeigte, dass sich die finale Aspiration in /p/ zwischen den Aufnahmezeitpunkten (A und B) signifikant unterschied ($z = 5.8, p < 0.001$), nicht jedoch in /k/ ($z = 2.0, p = 0.17$) (Abbildung 20).

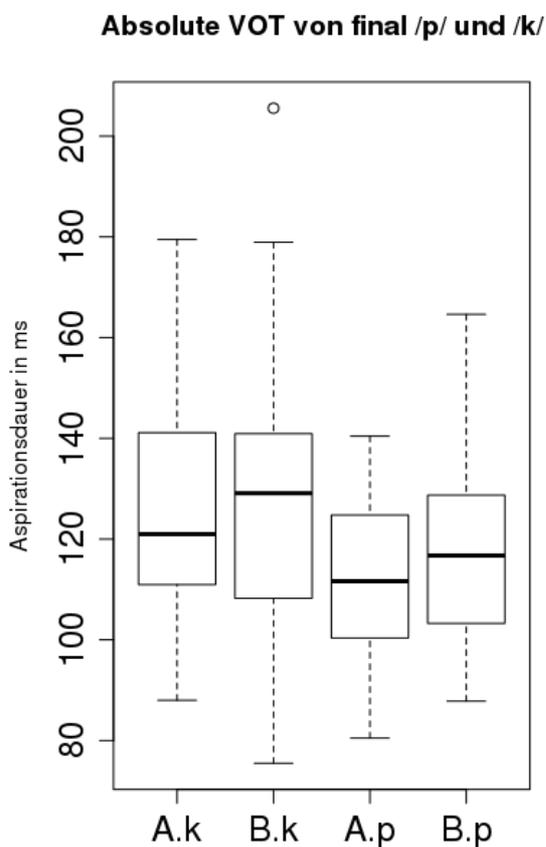


Abbildung 20: Die Abbildung zeigt die absolute Aspirationsdauer der finalen Plosive /p/ und /k/ in Millisekunden. Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme, "B" die zweite Aufnahme markiert.

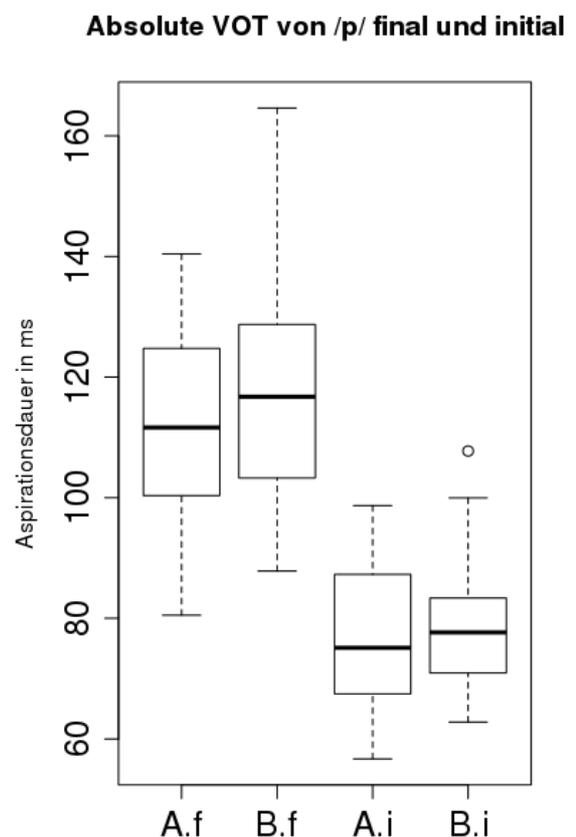


Abbildung 21: Die Abbildung zeigt die absolute Aspirationsdauer von /p/ initial und final in Millisekunden. Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme, "B" die zweite Aufnahme markiert.

Die Wortgruppen unterschieden sich zusätzlich anhand der Silbenanzahl: es gibt jeweils eine Gruppe einsilbiger und eine Gruppe mehrsilbiger Wörter. Die Aspirationsdauer dieser zwei Gruppen unterschieden sich bei den Wörtern mit finalelem /p/ nicht, d.h. zwischen den Aufnahmezeitpunkten war die Aspirationsdauer der einsilbigen Wörter nicht größer oder kleiner als zwischen den Aufnahmezeitpunkten der mehrsilbigen Wörter ($t = 1.2$, $df = 13$, $p = 0.27$).

Zusammengefasst heißt das: Es bestand Imitation der verlängerten VOT an finaler Position für sowohl ein- als auch mehrsilbige Wörter bei /p/, nicht aber bei /k/.

Die auslautverhärteten /p/-finalen Wörter wurden in der gesamten Auswertung gleich gehandhabt wie die orthographisch „p“-finalen Wörtern, da keine Interaktion zwischen Aufnahmezeitpunkt und orthographischer Form besteht ($X^2[1] = 0.05$, $p = 0.8$). Dennoch war die Aspirationsdauer von „b“ zu beiden Aufnahmezeitpunkten kürzer, als von „p“ ($X^2[1] = 25.1$, $p < 0.001$). Dies spricht dafür, dass der /b-p/-Kontrast final nicht komplett neutralisiert wurde (vgl. Kleber et al., 2010).

3.2.5.2 Vergleich der Positionen final und initial der Wörter mit /p/

Das *Mixed Model* mit den *Random* Faktoren Versuchsperson und Wort, *Fixed* Faktoren Aufnahmezeitpunkt und Position ergab eine signifikante Interaktion zwischen den *Fixed* Faktoren ($X^2[1] = 10.8$, $p < 0.01$) (Abbildung 21). Der post-hoc Tukey Test zeigte den signifikanten Einfluss der finalen Position auf die Aspirationsdauer von /p/ (siehe oben) und keinen signifikanten Einfluss der initialen Position ($p > 0.5$). Da es keinen Einfluss des Aufnahmezeitpunktes auf die Aspirationsdauern der initialen /p/s gab, wurde in der Analyse die Unterscheidung in /p/-initiale Wörter, die Teil der Hörliste waren, und /p/-initiale Wörter, die nicht Teil der Hörliste waren, nicht vorgenommen. Dasselbe gilt für die Unterscheidung in ein- und mehrsilbige Wörter.

Es ist überraschend, dass die verlängerte Aspiration an finaler Position der /p/-Wörter imitiert wurde, jedoch nicht an initialer Position, denn hier wurde starke Imitation erwartet (Nielsen, 2011b). Die hyperartikulierten Stimuli könnten für die zwei Positionen im Wort unterschiedliche Realisationen hervorgerufen haben. Während die finale Position wegen silbenfinaler Längung die Verlängerung der Aspiration begünstigen könnte, ist dies in initialer Position nicht möglich. Darüber hinaus muss an der initialen Position, im Gegensatz zur finalen Position, *initial strengthening* (Keating, Cho, Fougeron & Hsu, 2003) überwunden werden (siehe Abschnitt 3.3: Diskussion). Möglich ist, dass die Versuchspersonen durch höhere *Burst*- und Aspirationsintensität eine perzeptiv vergleichbare Deutlichkeit der Aussprache zu erreichen versuchten.

Aus diesem Grund wurde zusätzlich die RMS (*short-term Root Mean Square amplitude*) mit dem Emu Software System (Harrington, 2010) berechnet (*default settings*)

und ausgewertet. Die RMS ist in dB angegeben und entspricht der Kurzzeitenergie des Signals (Abbildung 22).

Es wurden drei *Mixed-Models* mit den *Random-Faktoren* Versuchsperson und Wort, *Fixed Faktoren* Aufnahmezeitpunkt und Position (final /p/ und initial /p/) berechnet, als *Response-Faktoren* den gemittelten absoluten RMS-Wert der Segmente, den gemittelten RMS-Wert der gesamten Äußerung (Wörter) und dem relativen RMS-Wert der Segmente (RMS-Wert des Segments geteilt durch RMS-Wert des Worts).

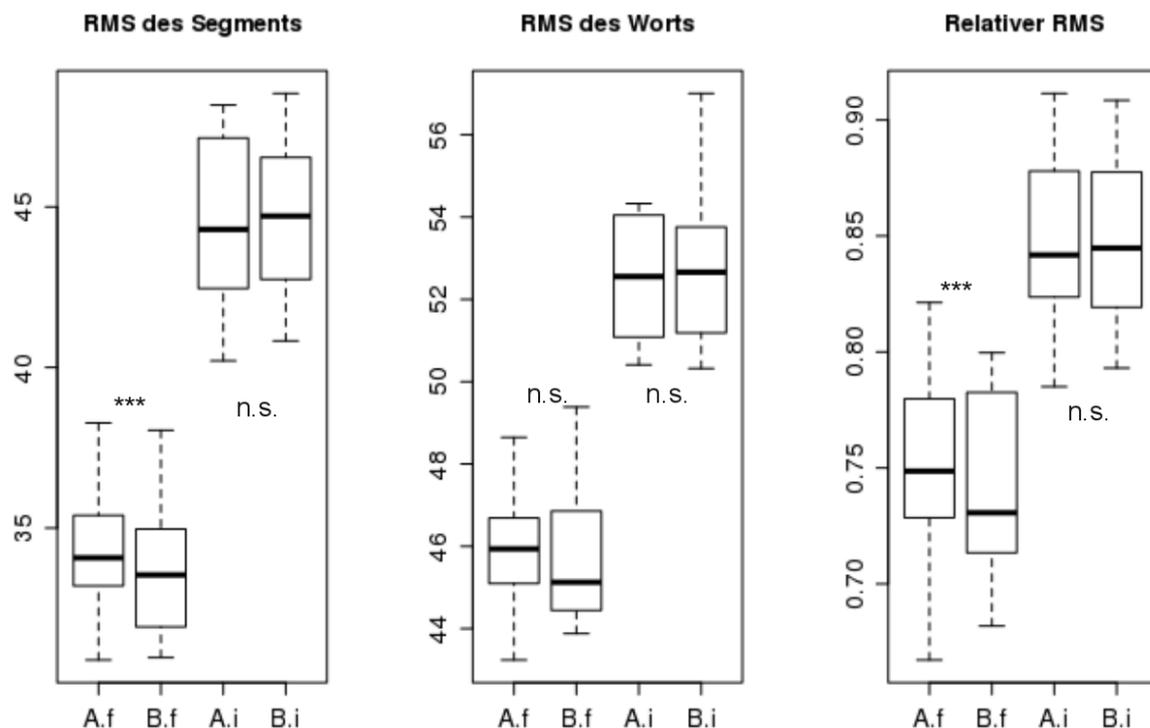


Abbildung 22: Die Abbildung zeigt den RMS-Wert (Root Mean Square) der initialen und finalen Plosive /p/ in dB. Von links nach rechts: die durchschnittlichen RMS-Werte des Segments /p/, die durchschnittliche RMS-Werte des gesamten Worts und die relativen RMS-Werte (RMS-Wert des Segments dividiert durch RMS-Wert des Worts). Die Bezeichnungen "A" und "B" unterhalb der Boxplots beschreiben den Aufnahmezeitpunkt, wobei "A" die erste Aufnahme und "B" die zweite Aufnahme markiert.

Für die *Response-Faktoren* absoluter gemittelter RMS-Wert ($\chi^2[1] = 19.5, p < 0.001$) und relativer gemittelter RMS-Wert ($\chi^2[1] = 14.3, p < 0.001$) besteht eine signifikante Interaktion. Ein signifikanter Einfluss des Aufnahmezeitpunkts besteht nur an finaler Position für den absoluten ($z = -5.1, p < 0.001$) und relativen RMS-Wert des Segments ($z = -5.7, p < 0.001$). Diese Abnahme an Intensität (siehe negativen z-Wert) könnte darauf zurückzuführen sein, dass die Längung der finalen VOT, besonders durch die Längung des späteren, intensitätsschwächeren Intervalls, realisiert wurde.

Ein Verlust oder Zuwachs an Intensität über die gesamte Wortlänge zwischen den Aufnahmezeitpunkten besteht nicht. Deshalb besteht keine, anhand des RMS oder Dauer, messbare Imitation in der Aspiration der initialen /p/s.

Zusammenfassend ergab die statistische Auswertung:

- Es besteht Imitation in finalem /p/ sowohl in ein- als auch mehrsilbigen Wörtern.
- Es besteht keine Imitation in /k/.
- Es besteht keine Imitation in initialem /p/.

3.3 Diskussion

Ausgehend von Ergebnissen der phonetischen Forschung zu spontaner Imitation wurde ein Experiment durchgeführt, das testete, ob Merkmale von wahrgenommener Sprache über die Position im Wort hinweg imitiert werden. Dafür wurde ein geblockter Versuchsablauf durchgeführt, der aus einem ersten Leseblock, einem Hörblock, innerhalb dessen manipulierte /p/-initiale Wörter mit verlängerter Aspirationsdauer gehört wurden, und einem zweiten Leseblock, bestand.

Die Ergebnisse der Imitation über die Position im Wort und die Artikulationsstelle bestätigen nur teilweise die Hypothesen. Exemplartheoretisch wurde argumentiert, dass an allen Positionen im Wort imitiert werden sollte (H1), weil für die Produktion eines aspirierten Plosivs alle gespeicherten Exemplare aktiviert werden, die das Merkmal „Aspiration“ beinhalten sowie das Merkmal „Aspiration“ getrennt vom Trägerlaut oder -wort (Nielsen, 2011b). Es konnte jedoch nicht an allen Positionen und Artikulationsstellen Imitation von Aspirationsdauer festgestellt werden.

Die Erwartung war, dass durch die stärkere Ähnlichkeit zu den manipulierten initialen, bilabialen Plosiven die Aktivierung der *Episodic traces* und damit die Imitation stärker bei den initialen bilabialen Plosiven ausfällt als bei den finalen bilabialen Plosiven (H2) und stärker bei den finalen bilabialen Plosiven als bei den finalen velaren Plosiven (H3).

Diese erwarteten Unterschiede in der Imitationsstärke hielten der statistischen Auswertung nur teilweise stand: der Vergleich der bilabialen Plosive zeigte, dass final imitiert wurde, aber nicht initial. Der Vergleich der Plosive an finaler Position zeigte, dass bilabial imitiert wurde, aber nicht velar.

In Bezug auf die Hypothesen heißt das zusammenfassend:

H1 wird abgelehnt: Imitation ist nicht an allen Positionen im Wort sichtbar.

H2 wird abgelehnt: Imitation ist nicht stärker an initialer als an finaler Position im Wort. Sogar besteht Imitation nur an der finalen Position im Wort und ist an der initialen Position im Wort nicht vorhanden.

H3 wird angenommen: Es besteht stärkere Imitation bei den finalen bilabialen Plosiven als bei den finalen velaren Plosiven.

Warum findet sich initial bilabial und final velar keine Imitation?

Unter Verwendung des gleichen Versuchsaufbaus wie in der vorliegenden Studie, wurde Imitation von initialer Aspiration in /p/ auf /p,k/ gezeigt (Nielsen, 2011b). Es ist ein überraschendes Ergebnis, dass hier keine Imitation an der initialen Position im Wort nachgewiesen werden konnte. Auch die Auswertung der RMS-Werte (*Root Mean square*), die der Kurzzeitenergie des Signals entsprechen, lieferte keine Erklärung. Wie vermutet, wurde auch an finaler Position innerhalb der velaren Plosive keine Imitation der Aspirationsdauer gefunden.

Generell ist die Imitation, die bei Laborsprache gemessen wird, relativ schwach. Dass die vorliegenden Daten nicht an allen Positionen/Artikulationsstellen wenigstens geringe Imitation aufzeigen, muss nicht zwangsläufig bedeuten, dass Imitation in diesen Fällen nicht möglich ist. Sie fällt jedoch geringer aus, als an den anderen Positionen bzw. Artikulationsstellen.

Warum gibt es mehr finale als initiale Imitation?

Die Positionen initial und final haben unterschiedlichen Einfluss auf Segmente. Eine große Zahl Studien testete die Unterschiede zwischen den Positionen initial und final (und medial) (z.B. Browman & Goldstein, 1995; Byrd, 1994, 1996; Fujimura, 1977; Keating et al., 2003; Kohler & Hardcastle, 1974) und kam zum Ergebnis, dass initiale Konsonanten eine längere Dauer und stärkere artikulatorische Größe (*initial strengthening*: Keating et al., 2003) aufweisen (z.B. Byrd, 1994; Keating, Wright & Zhang, 1999).

Die Dauer der finalen Aspiration in den Daten der vorliegenden Studie war generell länger, als die initiale Aspiration; da es sich um isoliert produzierte Wörter handelte, kann die längere Dauer der Aspiration der finalen Plosive in beiden Aufnahmezeitpunkten auf phrasenfinale Längung zurückgeführt werden (Berkovits, 1993). Eine genaue Zusammenfassung der artikulatorischen Unterschiede zwischen initialen und finalen Konsonanten liefert Fougeron (1998): „In initial position, consonants have a glottal opening gesture that is longer and larger; [...]. In initial position, consonants and vowels have greater muscular activity. [...] At the level of the tongue, consonants have a more elevated tongue position and greater pressure of the tongue against the palate.“ Die initiale Artikulation dauert demzufolge länger und ist stärker (bezüglich Bewegung und Muskelaktivität: *articulatory magnitude*). Das kann zu einer genaueren Artikulation führen, die weniger störanfällig und weniger beeinflussbar, dafür aber stabiler ist. Im Gegenzug

können die weniger strengen artikulatorischen Abläufe an der finalen Position zu früherer und stärkerer Beeinflussung führen.

Die Abbildung 21 könnte darüber hinaus den Eindruck erwecken, dass die Verteilungen der finalen /p/s stärker variabel ist, als die Verteilung der initialen /p/s. Neben der höheren artikulatorischen Freiheiten, die wie oben beschrieben an der finalen Position bestehen, könnte auch ein methodischer Bestandteil, nämlich die Segmentierung, dafür verantwortlich sein. Es ist schwieriger bei finalen /p/s die Segmentgrenze manuell vor dem Ende der Äußerung (Stille) oder dem Beginn von Atemgeräusch zu setzen bzw. zu bestätigen, als bei initialen /p/s, auf die stimmhafte Segmente folgen. Eine automatische Segmentierung könnte (mindestens) an der finalen Position von Vorteil sein. Im Falle von folgender Stille wäre das Setzen eines mindest-RMS eine praktikable Lösung, um das Wortende zu bestimmen. Eine Trennung zwischen finaler Aspiration und Ausatmung am Ende einer Äußerung ist sehr viel schwieriger umzusetzen und ein gutes Funktionieren, aufgrund der Erfahrungen mit automatischen Segmentierungen (z.B. *MAUS*, Schiel, 2004), auch nicht ohne manuelle Nachbearbeitung wahrscheinlich.

Grundsätzlich jedoch unterscheidet sich die Variabilität zwischen den Verteilungen der finalen und initialen /p/s nicht sehr stark: Die Standardabweichungen liegen bei 32.2 (final) und 22.4 (initial). Dieser Unterschied wird durch die Berechnung des Varianzkoeffizienten (der die Standardabweichungen auf die unterschiedlich hohen Mittelwerte normalisiert) ausgeglichen, sodass es sogar die initialen /p/s sind, die den höheren Variabilitätskoeffizienten aufweisen (0.295 initial; 0.280 final). Somit wird nicht davon ausgegangen, dass die Segmentierung per *MAUS* und manueller Berichtigung einen Einfluss auf die Verteilungen der Aspirationsdauern hat.

Was bedeuten die Ergebnisse für den Aufbau des Lexikons?

Nielsen (2011b) zeigte den Einfluss auf Sprachproduktion durch die Gliederung des Lexikons in Wort-, Phonem- und subphonemische Ebene anhand von Imitation durch Generalisierung und erweiterte damit die Ergebnisse, die über Perzeptionsexperimente gefunden wurden (z.B. Kraljic & Samuel, 2006; McQueen et al., 2006; Norris et al., 2003). Die vorliegende Studie testete Generalisierung zwischen Positionen im Wort (initial auf final) und zwischen Artikulationsstellen *und* Position im Wort (initial bilabial auf final velar). Während keine Generalisierung über Artikulationsstellen *und* Position im Wort auftrat, bestand generalisierende Imitation zwischen den Positionen initial und final im Wort für die gleiche (bilabiale) Artikulationsstelle. Dieses Resultat deutet darauf hin, dass die subphonemische Flexibilität auch über die Positionen im Wort besteht. Besonders interessant ist in diesem Zusammenhang die Wirkung auf die Silbenstruktur. Während über Position im Wort *und* Artikulationsstelle nicht imitiert wird, tritt die gleiche

Imitationsstärke in ein- und mehrsilbigen Wörtern auf. Das heißt, die Nähe zur wahrgenommenen Position spielt keine Rolle; die Flexibilität zwischen dem Merkmal und der Position im Wort ist folglich sehr hoch.

Mit Blick auf Lautwandel ist die stärkere finale Imitation im Vergleich zur initialen Imitation von besonderem Interesse.

Da der artikulatorische Fokus initial höher ist als final (siehe oben, *articulatory magnitude*), ist für die Aspirationslängung initial das Überwinden dieses Fokus notwendig. Lindblom (1998) führte die artikulatorische Schwierigkeit einer Aussprachevariante als eine Bedingung an, die die Ausbreitung eines Lautwandels moduliert. So sei das Kopieren einer einfachen Aussprachevariante schneller zu beobachten als das Kopieren einer schwierigen Aussprachevariante. Zwar ist das Merkmal in den hier besprochenen Daten das gleiche, doch üben die Positionen im Wort unterschiedlichen Einfluss auf die Angleichung der Aspirationsdauer aus; im Speziellen wirkt die initiale Position erschwerend. Folglich ist es möglich, dass der erhöhte artikulatorische Fokus an der initialen Position Imitation hemmt, während dies an der finalen Position nicht geschieht. Möglich ist, dass das Merkmal deshalb an der finalen Position stärker imitiert wurde. Nach Lindblom et al. sind „soziale, kommunikative, artikulatorische und lexikalisch-systemische Kriterien“ (1995, S. 19) entscheidend für die Akzeptanz einer neuen Aussprachevariante innerhalb einer Sprachgemeinschaft. Der finale Positionsvorteil lässt sich am besten als ein artikulatorisches Kriterium einordnen. Darüber hinaus kann dieses Ergebnis sehr gut entsprechend der Vorstellung von Lindblom et al. (1995) mit einem *Wie*- und einem *Was*-Modus in der Sprachverarbeitung erklärt werden. Während an den Stellen von gesprochener Sprache, die für den Informationstransport (das *Was*) sehr notwendig sind, das *Wie*, das heißt die genaue Realisierung wenig Beachtung findet, kann sich diese Teilung in *Was* und *Wie* an den Stellen verschieben, an denen weniger Inhalt überliefert werden muss. Die finale Position im Wort ist stärker vorhersagbar als die initiale Position und bedarf deshalb weniger stark des *Was*-Modus, wodurch der *Wie*-Modus stärkere Aktivierung sowohl in der Produktion als auch der Perzeption erlangen und neue Aussprachevarianten erzeugen kann.

Zusammenfassend liefert diese Studie einen Hinweis darauf, dass (1.) die Verbindung von Merkmals- zur Wortebene im mentalen Lexikon bezüglich der Position im Wort sehr flexibel ist und (2.) unterschiedliche Positionen im Wort für eine Ausbreitung von Lautwandel unterschiedlich geeignet sind.

4. Diskriminierbarkeit und Wahrnehmbarkeit von längerer Aspirationsdauer durch spontane phonetische Imitation

Zusammenfassung

Für einen Perzeptionstest wurden die Äußerungen einer Imitationsstudie verwendet. Diese wurden pro Segment und anhand von Segmentverbindungen akustisch analysiert: Es bestand Imitation in der Aspirationsdauer von /p/ in /p/-finalen Wörtern, aber nicht in /p/-initialen Wörtern.

Im Perzeptionstest verglichen HörerInnen die zwei Äußerungen eines Wortes (vor und nach der Wahrnehmung von sprachlichen Stimuli) anhand der Dauer von /p/ an entweder initialer oder finaler Position im Wort. Die Auswertung ergab, dass Versuchspersonen potentiell durch Imitation entstandene Dauerunterschiede, unterscheiden konnten. Dieses Ergebnis kann als Hinweis gewertet werden, dass spontane phonetische Imitation an perzeptivem Lernen und der Ausbreitung von Lautwandel beteiligt sein kann. Darüber hinaus wurde von den HörerInnen gleich viel Imitation an beiden Positionen entdeckt, obwohl die akustische Analyse Positionsunterschiede ergab. Diese Ergebnisse legen nahe, dass (1.) die aus spontaner phonetischer Imitation resultierende Variation weiter imitiert werden kann und (2.) eine Diskrepanz zwischen wahrgenommener und akustisch gemessener Imitation besteht.

4.1 Einleitung

Die im Folgenden dargestellte Studie befasst sich mit der Wahrnehmbarkeit von Dauerunterschieden in der Aspiration von Fortis-Plosiven, die potentiell durch spontane phonetische Imitation entstanden sind.

Phonetische Imitation ist ein Phänomen, bei dem SprecherInnen die phonetischen Eigenschaften ihrer Äußerungen an phonetische Eigenschaften umgebender Sprache anpassen. Wenn dies nicht intendiert, d.h. ohne Aufforderung zur Imitation, geschieht, spricht man von spontaner phonetischer Imitation (z.B. Babel, 2011; Goldinger, 1998).

Perzeptives Lernen bezeichnet einen Prozess, bei dem HörerInnen ihre Lautkategorien an Perzepte anpassen. Dies geschieht innerhalb des gleichen Wortes (z.B. /pas/->/pas/), sublexikalisch, bei der phonemische Generalisierung über Wörter hinweg zu beobachten ist (z.B. /pas/->/pol/) und auch subphonemisch, bei der Generalisierung über Artikulationsstellen auftritt (z.B. /tal/->/pas/) (Kraljic & Samuel, 2006; McQueen, Cutler & Norris, 2006; Norris, McQueen & Cutler, 2003).

Norris et al. (2003) zeigten, dass ein perzeptiver Kategorienwandel auftritt, wenn Versuchspersonen mit ambigen Lauten konfrontiert werden, deren

Kategorienzugehörigkeit lexikalisch abgeleitet werden kann. In dem Experiment handelte es sich um niederländische /s/- und /f/-finale Wörter: /na:ldbos/ und /Witlof/ (Niederländische SAMPA übernommen aus Norris et al., 2003; *naaldbos*: Nadelwald, *witlof*: Chicorée). Da keines der Wörter ein /f-s/-Minimalpaar ist, ist die lexikalische Zugehörigkeit der Wörter auch dann erkennbar, wenn das finale /s/ oder /f/ durch einen Frikativ ersetzt wird, der perzeptiv zwischen /f/ und /s/ liegt. Zwei Gruppen von HörerInnen erhielten entweder das /s/- oder /f/-finale Wort. So lernten die HörerInnen, dass der ambige Laut zur ersetzten Lautkategorie gehörte. Im darauf folgenden Identifikationstest mit einem Kontinuum von /f/ zu /s/ ordneten die Versuchspersonen jeweils mehr Stimuli des Kontinuums dem Frikativ zu, der vorher ersetzt worden war. Die Kategoriengrenze zwischen /f/ und /s/ wurde durch lexikalisch unterstütztes, perzeptives Lernen systematisch verschoben. Die gleiche Beobachtung wurde auch in der Studie von McQueen et al. (2006) gemacht. Zusätzlich zeigte dieses Experiment, dass die perzeptive Kategoriemerweiterung von entweder /f/ oder /s/ auch in anderen /f/- oder /s/-finalen Wörtern nachweisbar war, das heißt, der Kategorienwandel wurde auf andere Wörter generalisiert. Eine weitere Differenzierung stellt das Experiment von Kraljic und Samuel (2006) dar. Diese Studie beschreibt Generalisierung eines Lautmerkmals (VOT) von alveolaren zu bilabialen Plosiven und sogar zwischen SprecherInnen.

Perzeptives Lernen unterstützt die Perzeption von Sprache, indem es auf Wortebene sowie auf sublexikalischer und subphonemischer Ebene zu perzeptiver Anpassung führt. Deren Auswirkungen können relativ klein und zeitlich begrenzt sein, aber es bestehen robuste Effekte (Kraljic & Samuel, 2006).

Perzeptives Lernen, das zu Kategorienwandel oder -erweiterung führt, könnte Einfluss auf die Produktion nehmen. Aus der Forschung zu sprecherspezifischen Unterschieden in der Sprachverarbeitung ist bekannt, dass SprecherInnen, die eine bestimmte Ausprägung einer Kategorie produzieren, diese auch als prototypischer wahrnehmen (Newman, 2003). Dieser Zusammenhang zwischen der Perzeption und der Produktion kann für perzeptives Lernen bedeuten, dass wenn die perzeptive Kategorie verändert ist, auch eine Veränderung in der Produktion erkennbar werden könnte.

Nielsen (2011b) zeigte Generalisierung in der Produktion durch spontane Imitation sublexikalisch (zwischen Wörtern) und subphonemisch (zwischen Phonemen). Es ist wahrscheinlich, dass die beobachtete Veränderung des Merkmals VOT nicht nur in der Produktion, sondern auch in der Perzeption aufgetreten ist.

Spontane phonetische Imitation gilt als ein Instrument des Lautwandels (Babel, 2011; Lindblom, 1998; Trudgill, 2008). Ein solcher Zusammenhang kann jedoch nur möglich sein, wenn die Veränderungen der Produktionen wahrnehmbar sind. Das heißt, um an Lautwandel beteiligt zu sein, muss die veränderte Produktion von SprecherInnen bei

anderen HörerInnen wiederum zu perzeptivem Lernen und/oder Imitation Anlass geben können.

Die Stimuli, die in den Experimenten zum perzeptiven Lernen verwendet wurden, sind wahrnehmbar unterschiedlich zu den Prototypen der zugrunde liegenden Kategorien: nämlich perzeptiv genau intermediär.

Voraussetzungen, die Äußerungen erfüllen müssen, um an Lautwandel beteiligt sein zu können, wurden unterschiedlich beschrieben:

Ohala (1981) benannte die gestörte Perzeption aufgrund fehlender oder fälschlicher Kompensation für Koartikulation als Auslöser für Lautwandel. Da diese Prozesse eindeutige Auswirkung auf die Klassifizierung hätten, wären die Unterschiede deutlich wahrnehmbar.

Trudgill (1981, 1986) gab an, dass Dialektkonvergenz stattfindet, wenn eine linguistische Variable stark hervorstechend („highly salient“, Trudgill, 1986, S. 11) sei und verwendet würde, um durch soziolinguistische und stilistische Variation soziale Bedeutung zu markieren.

Lindblom, Guion, Hura, Moon und Willerman (1995) bedienen sich einer Unterteilung von zwei verschiedenen Modi der perzeptiven Sprachverarbeitung: des *How-Modes* und des *What-Modes*. Während der *How-Mode* die aktuelle Realisierung einer Äußerung registriert, das heißt *wie* eine Zieläußerung produziert wurde, stellt der *What-Mode* die lexikalische und semantische Verbindung her, indem entschieden wird, *was* geäußert wurde. Die Wie- und Was-Modi sind beide am Prozess der Perzeption beteiligt, jedoch orientiert sich der Wie-Modus am Signal, der Was-Modus am Inhalt der Äußerung. Je nach Situation ist der Output eines der Modi wichtiger als der Output des anderen. In der Regel ist der Was-Modus in der sprachlichen Kommunikation der wichtigere, damit HörerInnen erkennen, welcher Inhalt übermittelt werden sollte. Wenn das *Was* besonders einfach zu erkennen ist (beispielsweise bei Wiederholungen) oder der Inhalt einer Äußerung aufgrund der besonderen Realisierung der Äußerung in den Hintergrund des Interesses rückt (z.B. bei Störungen, fremden Akzenten/Dialekten), kann der Wie-Modus dem Signal viele Informationen entnehmen (Lindblom et al., 1995).

Mithilfe des Wie-Modus' kann jede Variation von Sprache potentiell neue Aussprachevarianten hervorbringen. Nach Lindblom et al. (1995) ist dafür die Erfüllung aller drei folgenden Voraussetzungen:

- Es gibt einen signifikanten Unterschied zwischen den Aussprachen,
- Normalisierung ist notwendig und
- die Normalisierung wird möglicherweise nicht angewendet.

Normalisierung ist als ein Konzept definiert, welches ermöglichen soll, die große Variabilität in gesprochener Sprache zum Ziel der Dekodierung zu überwinden. Zur richtigen Klassifikation einer Äußerung sollen die Bestandteile auf mentale Kategorien gemappt werden. Die für das phonologische Mapping nicht relevanten Bestandteile, wie paralinguistische und phonetisch nicht entscheidende Details, bleiben bei diesem Prozess unbeachtet (Ladefoged & Broadbent, 1957). Dieses Konzept ist so nicht haltbar, da viele Details der wahrgenommenen Äußerungen be- und gemerkt werden und die Sprachwahrnehmung unterstützen (Allen, Miller & DeSteno, 2003; Palmeri, Goldinger & Pisoni, 1993).

Die Voraussetzung des signifikanten Unterschieds (in Sinne von Lindblom et al., 1995) bezieht sich auf einen wahrnehmbaren Unterschied zur erwarteten Aussprache. Die Normalisierung, die verwendet werden müsste, würde dann den Zusammenhang zur üblichen Aussprachevariante herstellen und die unterscheidenden Details gingen verloren. Sollte die Normalisierung aber nicht angewendet werden, wäre ein Zusammenhang einer neuen Aussprachevariante zu einer bestehenden Kategorie hergestellt.

Lindblom et al. (1995), Ohala (1981) und Trudgill (1981, 1986) stimmen darin überein, dass wahrnehmbare Unterschiede zwischen der alten und einer neuen Aussprachevariante bestehen müssen, damit ein Lautwandel einsetzen kann.

VOT	(1) Baseline	(2) Post-Test / Shadowing	(3) Modell VOT	(4) Zunahme absolut (2) - (1)	(5) proportionale Zunahme (von (1) zu (2))
Initiales /p/ Nielsen (2011b)	65,9 ms 58.2%	73,6 ms 65.0%	113,3 ms 100%	7,7 ms	11,7 %
Initiale /p/, /t/, /k/ Shockley, Sabadini & Fowler (2004)	51,8 ms 37.9%	64,1 ms 47.1%	136 ms 100%	12,3 ms	23,7 %

Tabelle 6: Akustische Messergebnisse der Dauerveränderung von VOT in Experimenten zur spontanen phonetischen Imitation.

Spontane phonetische Imitation tritt als Annäherung an die Modell-Äußerungen auf, nie als Kopie (Babel, 2010; Pardo, 2012b). Deshalb sind die gemessenen Unterschiede zwischen den Äußerungen der Versuchspersonen in den unterschiedlichen Konditionen (ohne Beeinflussung versus mit potentieller Imitation) immer kleiner als die Unterschiede zwischen den Äußerungen der Versuchspersonen zu Beginn des Experiments und denen der ModellsprecherInnen. Tabelle 6 gibt einen Einblick in die Effektgrößen, die spontane phonetische Imitation in zwei Experimenten auf die VOT hatte. Die erste Spalte beinhaltet

die durchschnittliche VOT aller Versuchspersonen, bevor potentiell beeinflussende Sprache wahrgenommen wurde, die Spalte „Post-Test / Shadowing“ die Werte direkt danach. „Modell VOT“ bezeichnet die durchschnittliche VOT der vorgespielten Äußerungen.

Wenn Imitation an Lautwandel beteiligt sein soll, obwohl sie durchschnittlich sehr klein ist, müssen Effekte von Imitation dennoch mindestens wahrnehmbar sein.

In einigen Studien zu spontaner phonetischer Imitation werden, um Imitation aufzudecken, AXB-Tests verwendet (Goldinger, 1998; Pardo, 2006; Pardo et al., 2010). Hierbei befindet sich an der Position X eine Äußerung von dem/der ModellsprecherIn, an den Positionen A und B Äußerungen der Versuchspersonen. Unbeteiligte, naive HörerInnen beurteilen die Ähnlichkeit zwischen A und X sowie zwischen B und X und markieren das Paar mit der höheren Ähnlichkeit. Indem spontane Imitation perceptiv durch HörerInnen beurteilt wird, wird ermöglicht, dass die perzeptive Auswertung globale Maße der Konvergenz erfasst. Unterschiedliche SprecherInnen-Paare können in unterschiedlichen akustisch-phonetischen Eigenschaften und perzeptiven Maßen der Ähnlichkeit konvergieren. Eine akustische Analyse von wenigen Parametern kann die perzeptiven Maße übersehen bzw. nur teilweise berücksichtigen (Lelong & Bailly, 2011; Levitan & Hirschberg, 2011; Pardo, 2012b). Darüber hinaus ist zu beachten:

„[...] identification of converging acoustic attributes alone leaves open the question of whether such attributes are perceptually salient and available for use during conversation interaction.“ (Pardo, 2012b, S. 5).

Im vorliegenden Perzeptionsexperiment werden Äußerungen einer Produktionsaufgabe verwendet, die potentiell Imitation innerhalb eines akustisch genau bestimmten Parameters (Aspirationsdauer) beinhalten. Dabei sollen die folgenden zwei Hypothesen geprüft werden:

H1: HörerInnen können die durch Imitation entstandenen Unterschiede in der Aspirationsdauer diskriminieren (und damit potentiell weiter imitieren).

H2: Die HörerInnenurteile spiegeln die akustisch gefundene bzw. nicht gefundene Imitation in der Aspirationsdauer wider.

Um diese Erwartungen belegen zu können, werden die Segmentdauern der Äußerungen für das Perzeptionsexperiment im folgenden Abschnitt genau analysiert.

4.2 Experiment 1: Akustische Analyse des Dateninventars für das Perzeptionsexperiment

Ein Teil des Aufnahmekorpus aus dem Perzeptionsexperiment (Abschnitt 3) wird verwendet. Dabei handelte es sich um Äußerungen, die aus einem geblockten Versuchsablauf resultierten: Der erste Block bestand aus dem Lesen 167 isolierter, jeweils

ein- und mehrsilbiger /p/-initialer, /p/-finaler, /k/-finaler Wörter und Füllwörter. Im zweiten Block hörten die Versuchspersonen 83 /p/-initiale Wörter mit manipulativ verlängerter Aspirationsdauer und 38 Füllwörter (das heißt unter den gehörten Wörtern waren keine /p/- oder /k/-finalen Wörter). Der dritte Block war eine Wiederholung des ersten Blocks. Die Aufnahmen dieses Experiments wurden ausgewertet, indem die Äußerungen des ersten und dritten Blocks verglichen wurden (Bezeichnung im Folgenden: Äußerungen A (erster Block) wurden mit den Äußerungen B (dritter Block) verglichen). So konnten die Auswirkungen der im zweiten Block wahrgenommenen Stimuli pro Wortart (/p/ initial oder final, /k/ final, jeweils ein- oder mehrsilbig) betrachtet und verglichen werden. Die Auswertung ergab, dass nur in den /p/-finalen Wörtern eine akustische Veränderung zwischen den Aufnahmezeitpunkten (A oder B) besteht obwohl diese genauso oder stärker bei den /p/-initialen Wörtern erwartet wurde. Deshalb wurden diese beiden Wortarten (/p/ initial und /p/ final) in dieser Studie verwendet. Weiter reduziert wurde der Korpus durch die Beschränkung auf monosyllabische Wörter und den Ausgleich der Gruppengröße: Da im Produktionsexperiment nur sieben einsilbige /p/-initiale Wörter verwendet wurden, wurden aus den 17 /p/-finalen einsilbigen Wörtern sieben, wie im Folgenden beschrieben, ausgewählt.

4.2.1 Methodik

4.2.1.1 Datenreduktion

Die experimentellen Äußerungen der Versuchspersonen (im Folgenden „SprecherInnen“) aus dem Produktionsexperiment (Abschnitt 3) wurden verwendet. Diese beinhalten Äußerungen von sieben einsilbigen Wörtern mit initialem /p/ und 17 einsilbigen Wörtern mit finalem /p/, jeweils von 14 SprecherInnen.

Um zu ermöglichen, dass alle HörerInnen innerhalb eines Termins alle AB-Varianten von allen SprecherInnen beurteilen konnten, wurden aus den 17 Wörtern mit finalem /p/ sieben ausgewählt, die repräsentativ für alle 17 sind:

Im ersten Schritt wurde die Differenz der Dauer von /p/ zwischen den Aufnahmebedingungen B und A pro Wort berechnet (Durchschnitt über alle SprecherInnen). Dies ergab 17 Werte (Tabelle 7), die zwischen -2.1 ms (Verkürzung der VOT von Kondition A zu Kondition B) und 27.9 ms (Verlängerung von A zu B) liegen.

Im zweiten Schritt wurden aus dieser Verteilung die mittleren sieben Wörter ausgewählt, d.h. diejenigen Wörter, die dem Mittelwert (9.7 ms) am nächsten lagen: Dazu wurde zuerst der Mittelwert von den A-B-Differenzen subtrahiert und aus den Ergebnissen der Betrag gebildet. Die sieben Wörter mit den kleinsten Werten wurden für das Perceptionsexperiment ausgewählt (Tabelle 8).

Der *Wilcoxon-Rangsummentest*, jeweils für die Konditionen A und B, ergab, dass sich die Verteilungen der sieben ausgewählten Wörter nicht signifikant von der Verteilung aller Aufnahmen unterschieden (A: $W = 12208.5$, $p = 0.5$; B: $W = 12629$, $p = 0.2$).

Abbildung 23 veranschaulicht die Verteilungen und deren Vergleichbarkeit. In „gesamt“ befinden sich die Werte aller 17 /p/-finalen Wörter, das heißt auch der sieben ausgewählten. In „Auswahl“ finden sich nur die Werte der sieben ausgewählten. A und B bezeichnet die Unterscheidung in A: erste Aufnahme; B: zweite Aufnahme.

Mit dieser Auswahl ließ sich das Perzeptionsexperiment auf 392 Durchgänge begrenzen, die innerhalb von ca. 45 Minuten bearbeitet werden konnten (14 Wörter x 14 SprecherInnen x 2 Reihenfolgen).

Die ausgewählten Wörter für die akustische Analyse und das Perzeptionsexperiment sind in Tabelle 9 dargestellt.

Differenz in ms	Wort
-2.1	Krepp
0.9	schnapp
2.0	Klipp
2.0	Pepp
2.3	kipp
2.4	stopp
4.8	schlepp
5.1	schlapp
6.3	knapp
10.1	Stepp
14.7	Depp
14.9	Mopp
16.5	Tipp
17.5	klapp
17.7	Trupp
21.9	Tripp
27.9	schwupp

Tabelle 7: Unterschied zwischen den Werten der Aspirationsdauer der A-Aufnahmen und den B-Aufnahmen (B-A) im Durchschnitt über alle SprecherInnen, aufsteigend sortiert.

Wort	Differenzbetrag in ms
Stepp	0.4
knapp	3.4
schlapp	4.6
schlepp	4.9
Depp	5.0
Mopp	5.2
Tipp	6.8
stopp	7.3
kipp	7.4
Pepp	7.7
Klipp	7.7
klapp	7.8
Trupp	8.0
schnapp	8.8
Krepp	11.8
Tripp	12.2
schwupp	18.2

Tabelle 8: Mittelwert der Verteilung (Tabelle 7) subtrahiert von den Differenzwerten (Tabelle 7), aufsteigend sortiert. Im schwarzen Kasten sind die ausgewählten Wörter markiert.

VOT der Auswahl gegenüber der VOT aller Wörter

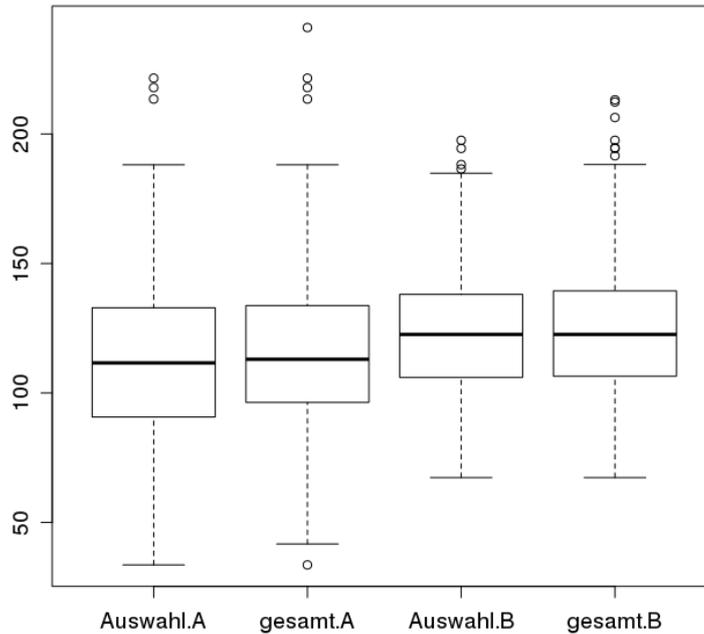


Abbildung 23: Aspirationsdauer (in ms) von finalem /p/ in einsilbigen Wörtern. In „gesamt“ befinden sich 17 Wörter, in „Auswahl“ sieben der 17 Wörter, jeweils für die Aufnahmekonditionen A und B. Die ausgewählten sieben Wörter werden als repräsentative Teilmenge aller 17 Wörter akzeptiert.

	/p/ initial	/p/ final
1	Pech	Depp
2	Pein	knapp
3	Pest	Mopp
4	Pol	schlapp
5	Punsch	schlepp
6	pur	Stepp
7	Putz	Tipp

Tabelle 9: Wörter der akustischen Analyse und des Perzeptionsexperiments.

4.2.1.2 Parametrisierung und Auswertung

Innerhalb des (reduzierten) Korpus wurden alle Segmente manuell zeitlich genau markiert: automatisch mit dem *Münchener Automatischen Segmentationssystem* (MAUS, Schiel, 2004) und manuell korrigiert mit der Sprachverarbeitungssoftware *Praat* (Boersma, 2001). Die Dauern wurden segmentweise und in Segmentkombinationen mit dem *Emu Software System* ausgewertet (Harrington, 2010).

Die Parametrisierung der Lautsegmentdauern durch Segmenteinheiten wird anhand zweier Beispiele gezeigt. Für die /p/-finalen Wörter zeigt Abbildung 24 beispielhaft die

Segmentierung und Annotation einer Äußerung von „Depp“, beispielhaft für die /p/-initialen Wörter zeigt Abbildung 25 die Segmentierung und Annotation von „Putz“.

Im Anschluss an die Grafiken werden die Lautsegmente anhand ihrer Bezeichnung in der Segmentation den sieben Segmenteinheiten der Analyse zugeordnet.

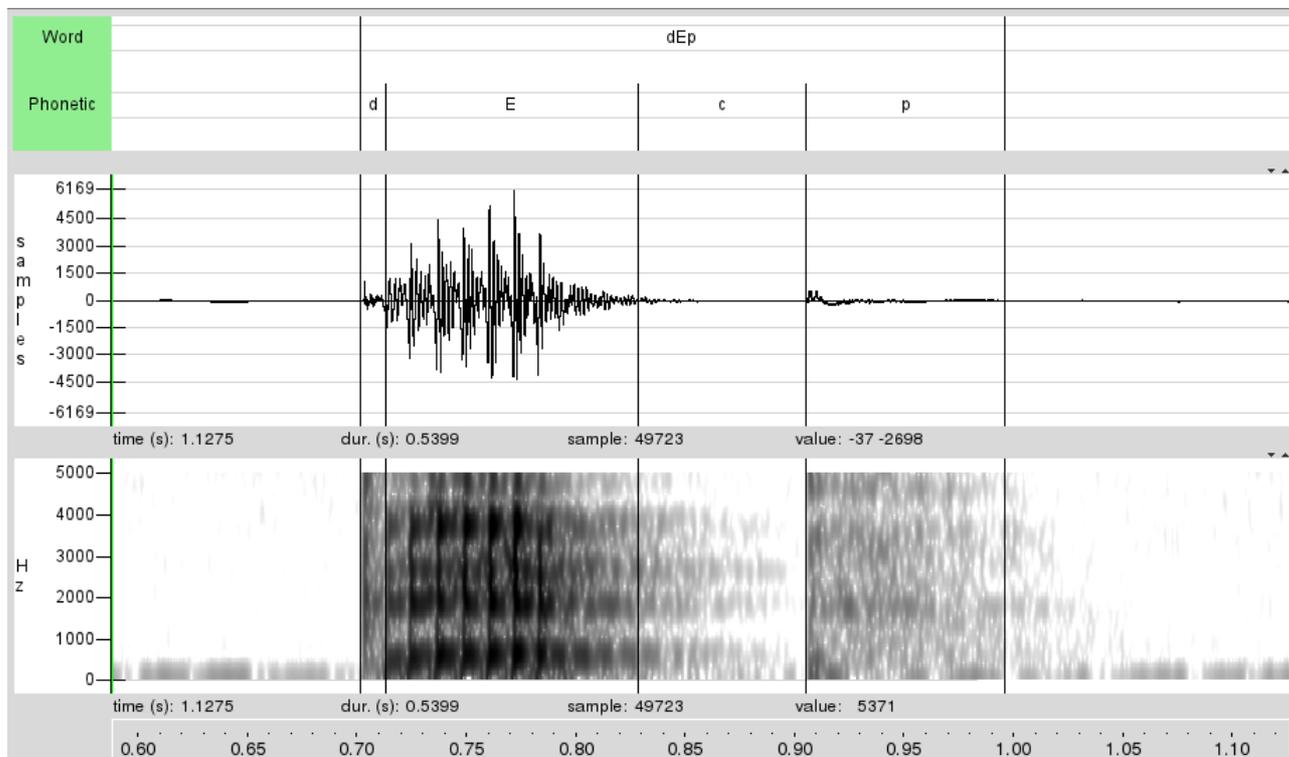


Abbildung 24: Beispielwort mit finalelem /p/: Depp; c im Mittelteil der Äußerung markiert den Verschluss (closure) des Plosivs.

Die analysierten Segmenteinheiten für das Wort „Depp“ setzen sich aus folgenden Lautsegmenten zusammen:

- initialer Konsonant bzw. initiale Konsonantverbindung: Ebene „Phonetic“ d
- finaler Konsonant bzw. finale Konsonantverbindung: Ebene „Phonetic“ c + p
- Aspiration von /p/ (genauer: Burst plus Aspiration): Ebene „Phonetic“ p
- /p/-Verschluss des /p/ : Ebene „Phonetic“ c
- Vokal: Ebene „Phonetic“ E
- Reim: Ebene „Phonetic“ E + c + p
- Wort: Ebene „Phonetic“ d + E + c + p oder äquivalent Ebene „Word“ dEp

Im Fall einer Konsonantverbindung zu Beginn des Wortes werden die Dauern aller Konsonanten dieser Verbindung addiert.

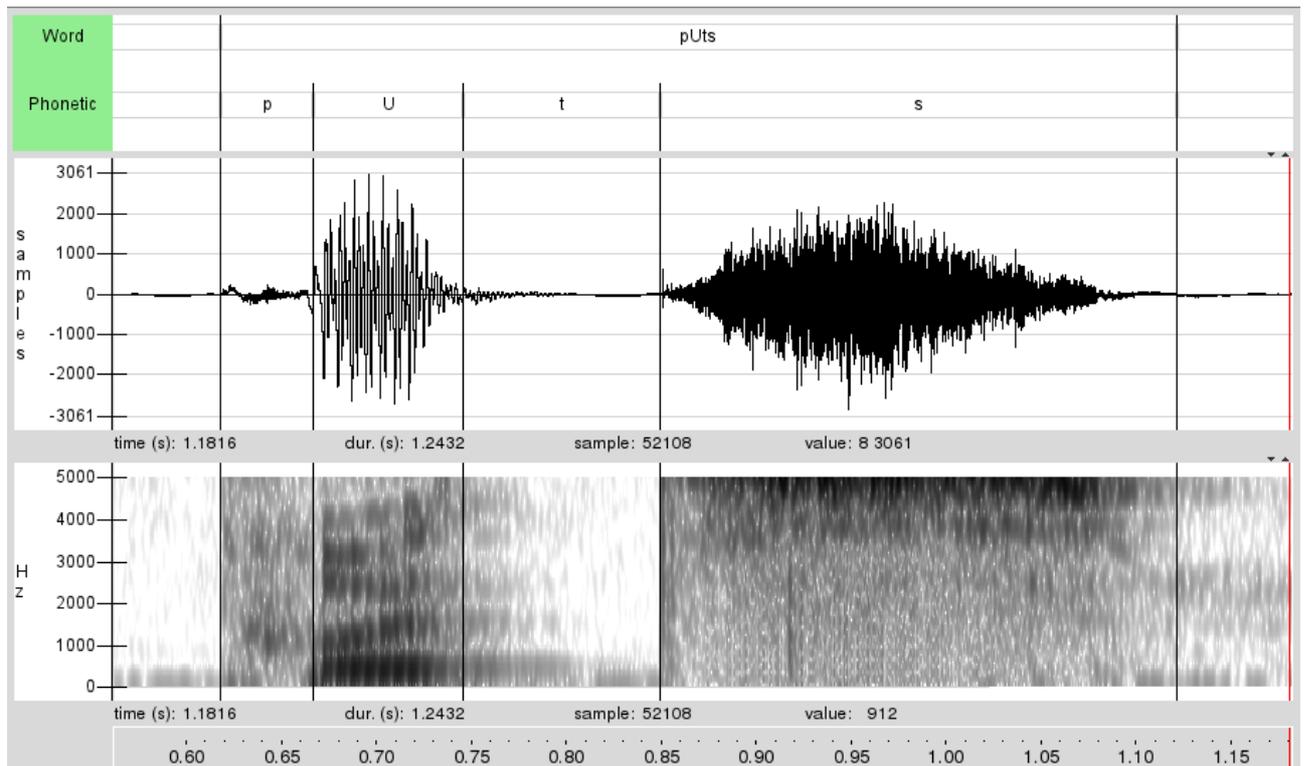


Abbildung 25: Beispielwort mit initialem /p/: Putz

Die analysierten Segmenteinheiten für das Wort „Putz“ setzen sich aus folgenden Lautsegmenten zusammen:

- initialer Konsonant bzw. initiale Konsonantverbindung: Ebene „Phonetic“ p
- finaler Konsonant bzw. finale Konsonantverbindung: Ebene „Phonetic“ t + s
- Aspiration von /p/ (genauer: Burst plus Aspiration): Ebene „Phonetic“ p
- Vokal: Ebene „Phonetic“ U
- Reim: Ebene „Phonetic“ U + t + s
- Wort: Ebene „Phonetic“ p + U + t + s oder äquivalent Ebene „Word“ pUts

Im Fall von Wörtern mit initialem /p/ sind die Segmente „initialer Konsonant“ und „Aspiration von /p/“ identisch.

4.2.2 Ergebnisse

Die Segmentdauern der Äußerungen aus dem Perzeptionsexperiment wurden akustisch analysiert. Eine Übersicht der Wörter befindet sich in Tabelle 9, die Zuweisung von Lautsegmenten zu den analysierten Segmenteinheiten erfolgte anhand von Beispielen in Abschnitt 4.2.1.2, Abbildungen 24 und 25.

Hier werden die Dauern der folgenden sieben Segmenteinheiten besprochen:

- initialer Konsonant bzw. initiale Konsonantverbindung,
- finaler Konsonant bzw. finale Konsonantverbindung,

- Aspiration von /p/ (genauer: Burst plus Aspiration),
- /p/-Verschluss des /p/ nur an finaler Position (da an initialer Position nicht vorhanden),
- Vokal,
- Reim (Vokal plus finaler Konsonant bzw. finale Konsonantverbindung) und
- Wort.

Der Fokus dieser Analyse liegt auf den Unterschieden der Segmentdauern zwischen den Aufnahmezeitpunkten A und B (A: bevor Stimuli mit verlängerter, initialer /p/-Aspiration wahrgenommen wurden, B: danach). Veränderungen zwischen diesen Konditionen werden potentiell der wahrgenommenen Sprache zugewiesen.

Die Dauern der Segmente bzw. Segmentverbindungen (*Response* Faktor) wurden anhand mehrerer *Mixed Effect Model* Analysen (Baayen, 2008) getrennt pro Wortart (je nach /p/-Position) verglichen. *Fixed* Faktor war die Kondition (A oder B), *Random* Faktoren SprecherIn und Wort (Tabelle 10).

Eine Veränderung der Segmentdauern war nur bei einigen Segmenten bzw. Segmentverbindungen der Wörter mit finalelem /p/ festzustellen. Hierbei handelte es sich um die Aspirationsdauer (plus *Burst*) von /p/, die Reimdauer und die Wortdauer. Keine Veränderung fand sich bei den finalen /p/-Wörtern in der Vokaldauer und der Verschlussdauer. Das heißt, dass die größere Dauer von Reim und Wort auf die verlängerte Aspirationsdauer von /p/ zurückzuführen war, nicht auf eine grundsätzliche Verlängerung aller oder einiger Segmentdauern.

Bei den Wörtern mit initialem /p/ wurde keine Veränderung zwischen den Aufnahmezeitpunkten gefunden: Weder das initiale /p/ noch der finale Konsonant bzw. die finale Konsonantverbindung weisen eine Veränderung auf. Die im Perzeptionsexperiment wahrgenommenen manipulierten Stimuli mit verlängerter Aspirationsdauer im /p/ an initialer Position hatten auf die /p/-initialen Zielwörter an keiner Position einen Einfluss. In Abschnitt 3 wurde argumentiert, dass dies auf die höhere artikulatorische Schärfe der initialen Konsonanten, im Gegensatz zur größeren artikulatorischen Variabilität an der finalen Position, zurückzuführen sein könnte.

Veränderung der Segmentdauern vom Aufnahmezeitpunkt A zu B

<i>Beispielwort „Putz“</i>	/p/	/u/	/ts/
<i>Beispielwort „Depp“</i>	/d/	/ε/	Verschluss + /p/
/p/-initiale Wörter	/p/ <i>Burst</i> und Aspiration	keine Veränderung	Konsonant(verbindung)
	keine Veränderung		keine Veränderung
/p/-finale Wörter	Konsonant(verbindung)	keine Veränderung	/p/ Verschluss
	keine Veränderung		keine Veränderung
			/p/ <i>Burst</i> und Aspiration
			länger $X^2[1] = 6.9, p < 0.01$
			/p/ Verschluss, <i>Burst</i> und Aspiration
			keine Veränderung
Segmenteinheiten	Konsonant- (verbindung) initial	Vokal	Konsonant- (verbindung) final

/p/-initiale Wörter	keine Veränderung
/p/-finale Wörter	länger $X^2[1] = 7.1, p < 0.01$
Segmenteinheit	Reim

/p/-initiale Wörter	keine Veränderung
/p/-finale Wörter	länger $X^2[1] = 5.2, p < 0.05$
Segmenteinheit	Wort

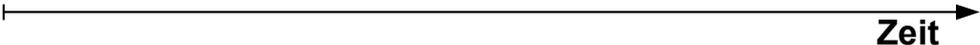


Tabelle 10: Veränderung der Dauern der Segmente bzw. Segmentverbindungen von der ersten Aufnahme (A) zur zweiten Aufnahme (B) aller sieben Wörter. „Länger“ bezeichnet das Längerwerden von Aufnahmezeitpunkt A zu B, d.h. in B sind die Segmente länger.

4.2.3 Diskussion der akustischen Ergebnisse und Erwartungen für Experiment 2

Die akustische Analyse der Äußerungen für den Perzeptionstest ergab, dass innerhalb der /p/-initialen Wörter zwischen den Konditionen kein Unterschied besteht. Anders ist es bei den /p/-finalen Wörtern: Hier werden Plosivdauer (Aspiration und *Burst*), Reimdauer und Wortdauer länger (von A zu B), was auf die Plosivdauer zurückzuführen ist.

Ausgehend von der akustischen Analyse der im Perzeptionstest verwendeten Stimuli wurden die folgenden Erwartungen für das Perzeptionsexperiment formuliert:

1. Die HörerInnen können Unterschiede der Aspirationsdauer diskriminieren:

Versuchspersonen sind sensibel für Dauerunterschiede ab 10 - 40 ms in Sprachlauten von 30-300 ms (Lehiste, 1970). 68.4 % der Unterschiede zwischen den Aspirationsdauern betragen 10 ms oder länger, 14.8 % betragen 40 ms oder länger. Die Aspiration ist in keiner Äußerung kürzer als 30 ms.

2. Innerhalb der Wortklasse „initiales /p/“ wählen HörerInnen gleich verteilt A- und B-Äußerungen als die längere aus:

Da kein akustischer Dauerunterschied zwischen dem initialen /p/ in den Kategorien A und B besteht, wird erwartet, dass keine perzeptive Präferenz auftritt.

In der Wortklasse „finale /p/“ werden mehr B-Äußerungen unter den HörerInnenurteilen erwartet, da die B-Äußerungen akustisch länger sind als die A-Äußerungen.

4.3 Experiment 2: Perzeptionstest

4.3.1 Methodik

Unter Verwendung der temporal genau analysierten Äußerungen wird ein *Two-alternative forced choice*-Perzeptionsexperiment (2AFC) durchgeführt. 2AFC bezeichnet eine experimentelle Fragestellung, die Versuchspersonen beantworten, indem sie aus zwei alternativen Antworten wählen. Eine dieser Alternativen muss gewählt werden, d.h. die Antwortmöglichkeit „weiß ich nicht“ besteht nicht. In der hier beschriebenen Experimentdurchführung werden in einem Durchgang die Äußerungen A und B eines Wortes von einer Sprecherin/einem Sprecher nacheinander vorgespielt. Kondition A bezieht sich auf die Äußerung, die vor der potentiellen Beeinflussung von wahrgenommener Sprache produziert wurde, Kondition B bezieht sich auf die Äußerung, die potentiell Imitation beinhaltet. Die HörerInnen entschieden, in welcher Äußerung sich das längere /p/ befand. Um Reihenfolgeeffekte zu vermeiden, wurden die Äußerungen

zweimal in unterschiedlicher Reihenfolge vorgespielt. Die Auswahl der Durchgänge erfolgte automatisch randomisiert.

Diese Vorgehensweise ermöglicht zum einen zu testen, ob Versuchspersonen unterschiedliche Dauern von *Burst* und Aspirationsdauer als Dauerunterschiede in /p/ wahrnehmen können. Das würde als bestätigt gelten, wenn die HörerInnen in mehr Durchgängen die akustisch längeren /p/s als länger auswählen.

Zum anderen wird deutlich, ob Versuchspersonen häufiger B-Äußerungen als die Äußerung angeben, in der sich das längere /p/ befindet. Sollte das der Fall sein, ist das ein perzeptiv gefundener Hinweis auf (potentielle) Imitation der Dauer von /p/.

4.3.1.1 Versuchspersonen

Am Perzeptionsexperiment nahmen 21 monolinguale Hörerinnen und Hörer mit Deutsch als Muttersprache, im Alter zwischen 21 und 36 Jahren, teil. Keine der Versuchspersonen zeigte oder berichtete von Auffälligkeiten in der Sprachproduktion oder -wahrnehmung.

4.3.1.2 Durchführung

Die Versuchspersonen saßen in Gruppen von maximal 3 Teilnehmenden an jeweils einem Rechner und hörten die Äußerungen über geschlossene Kopfhörer (beyerdynamic DT 770). Die Versuchsanweisungen wurden vor Beginn schriftlich erteilt (siehe Anhang 5: Erklärungen zum Experiment) und in einzelnen Fällen mündlich nochmal durchgesprochen. Der 2AFC-Test wurde in *Praat* (Boersma, 2001) durchgeführt. Die Äußerungen wurden mit einem Interstimulusintervall von 0.3 Sekunden abgespielt, die Äußerungen eines jeden Durchgangs konnten bis zu fünf mal erneut angehört werden und nach der Hälfte der Durchgänge wurde eine Pause angeboten. Zur Bearbeitung benötigten die Versuchspersonen ca. 45 – 60 Minuten.

4.3.1.3 Auswertung

Um die Diskriminationsfähigkeit der HörerInnen auszuwerten, wurden die Urteile des Perzeptionsexperiments mit den Paaren der absoluten Aspirationsdauer verglichen. Pro Wort und SprecherIn standen zwei Äußerungen zur Verfügung. Die eine Äußerung (A) wurde produziert, bevor die VOT-manipulierten Stimuli vorgespielt wurden, die andere danach (B). Pro Äußerungspaar besitzt eine Äußerung eine längere Aspirationsdauer als die andere. Die HörerInnenurteile wurden darauf geprüft, ob die Äußerung mit dem länger aspirierten /p/ als solche markiert wurde. War dies der Fall, wurde dieses HörerInnenurteil mit „TRUE“ markiert. Falls ein/e HörerIn die Äußerung mit dem kürzer aspirierten /p/ als länger wahrnahm, erhielt dieses Urteil das Label „FALSE“. Die TRUE/FALSE-Verteilung wurde mit einem binomialen *Mixed Model* auf Zufälligkeit geprüft. Diese Auswertung gab

an, ob sich die Anzahl der TRUEs gegenüber der Anzahl der FALSEs unterschied. Wort und HörerIn bildeten die *Random* Faktoren.

Um zu testen, ob die als länger wahrgenommenen Aspirationsdauern häufiger aus den Aufnahmen vor (A) oder nach (B) der potentiellen Beeinflussung durch fremde Sprache stammten, wurde getestet, ob sich die Häufigkeit der A- bzw. B-Wahlen unterschied. Wenn sich für eine oder beide der Wortklasse mehr Markierungen von B als von A fänden, würde das zeigen, dass die /p/s der Äußerungen der Kondition B perceptiv länger sind als die von A. Das würde als Hinweis für Imitation gewertet werden. Unterscheidet sich die Auswahl von A oder B nicht, läge keine Imitation vor, sind es mehr A- als B-Äußerungen, besteht perceptiv Divergenz. Analog zur Vorgehensweise für die statistische Berechnung der Diskriminationsfähigkeit, wurde die Verteilung der A/B-Antworten mit einem binomialen *Mixed Model* auf ihre Zufälligkeit geprüft. So wird getestet, ob sich die Anzahl der A- bzw. B-Markierungen unterschied. Wort und HörerIn bildeten auch hier die *Random* Faktoren.

4.3.2 Ergebnisse

Im Perzeptionsexperiment wurden Versuchspersonen (hier: HörerInnen) ausgewählte Äußerungen des Produktionsexperiments pro Wort und SprecherIn vorgespielt. Die HörerInnen gaben an, welche der zwei Äußerungen das längere /p/ beinhaltete. Anhand dieser Vorgehensweise lässt sich zum einen beantworten, ob Versuchspersonen in einem solchen Diskriminationstest in der Lage sind, die akustisch gemessenen Dauerunterschiede wahrzunehmen und zum anderen, ob die perzeptive Gestalt der Äußerungen auf vergleichbare Imitation schließen lässt wie die akustische Analyse.

4.3.2.1 Diskriminationsfähigkeit

Die perzeptive Beurteilung eines Wortpaares verbessert sich relativ schwach mit größer werdenden Dauerdifferenzen. Wie die Abbildung 26 zeigt, bestehen nur schwache Korrelationen zwischen den Dauerdifferenzen und der proportionalen Erkennungsleistung an beiden Positionen (final: $R^2 = 0.05$, $F(1, 96) = 5.3$, $p < 0.05$; initial: $R^2 = 0.08$, $F(1, 96) = 8.2$, $p < 0.01$).

Dennoch haben die HörerInnen signifikant häufiger das Wort mit dem längeren /p/ erkannt ($z = 3.6$, $p < 0.001$). Zwischen den Kategorien /p/ an finaler bzw. initialer Position im Wort besteht kein Unterschied (Abbildung 27).

Es war den Versuchspersonen also möglich, die verschiedenen Dauern der /p/s an beiden Positionen im Wort zu diskriminieren. Damit ist eine Voraussetzung für die Verbreitung einer Variation durch spontane Imitation erfüllt.

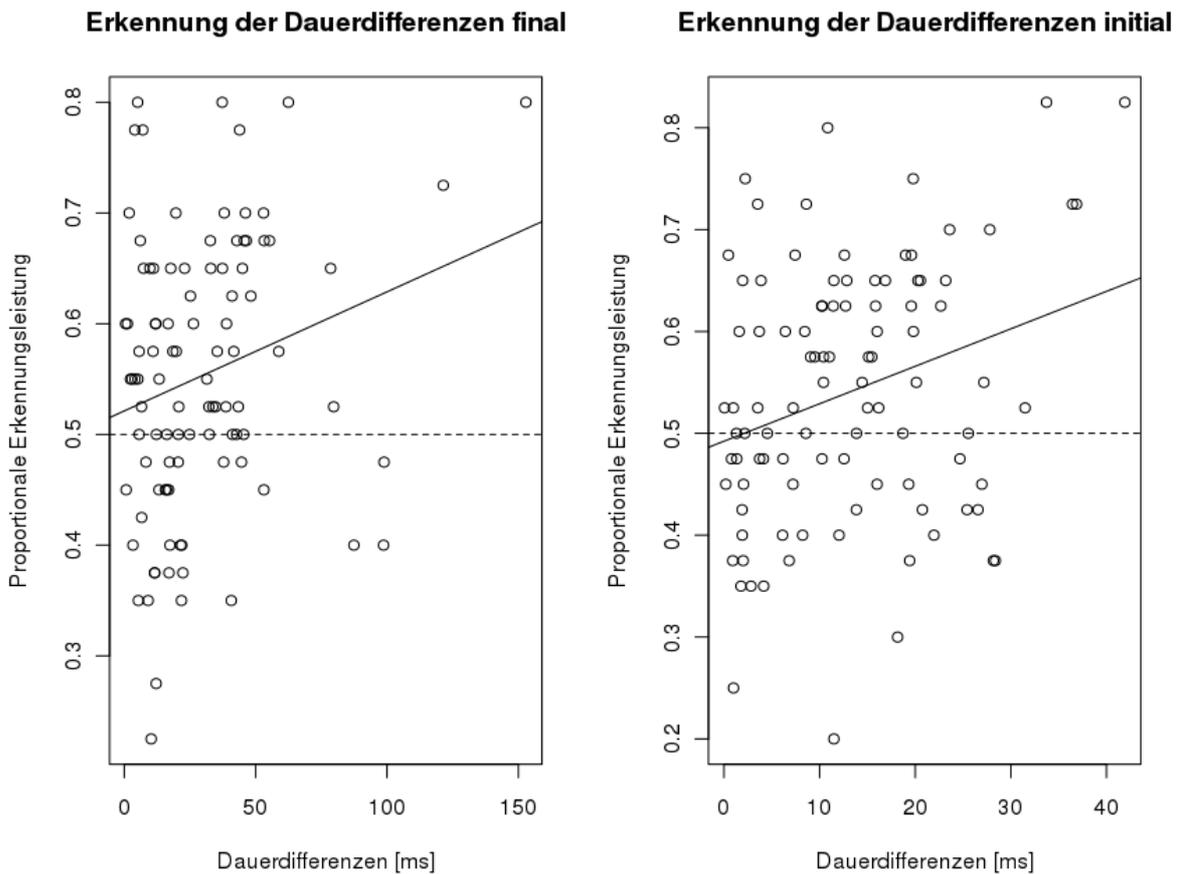


Abbildung 26: Darstellung der Dauerdifferenzen (x-Achse) gegenüber der proportionalen Erkennungsleistung der HörerInnen (y-Achse) mit Regressionslinien. Die gestrichelte Linie markiert die Erkennung zu 50%, d.h. Zufall.

4.3.2.2 Perception von längerer Aspirationsdauer durch Imitation

Die Aufgabe der HörerInnen bietet darüber hinaus die Möglichkeit zu sehen, ob die als länger wahrgenommenen Aspirationsdauern häufiger aus den Aufnahmen vor (A) oder nach (B) der potentiellen Beeinflussung durch fremde Sprache stammen.

Hierfür wurde getestet, ob sich die Häufigkeit der A- bzw. B-Wahlen unterschied. Es wurde aus den SprecherIn-Wort-Paaren signifikant häufiger eine B-Äußerung als die Äußerung mit dem längeren /p/ gewählt ($z = 2.6$, $p < 0.01$). Die Verteilungen unterschieden sich nicht zwischen den Positionen. Über die perzeptive Analyse der Daten lässt sich also auch an der initialen Position ein potentiell imitatorischer Effekt innerhalb der Dauer von /p/ zeigen (Abbildung 28).

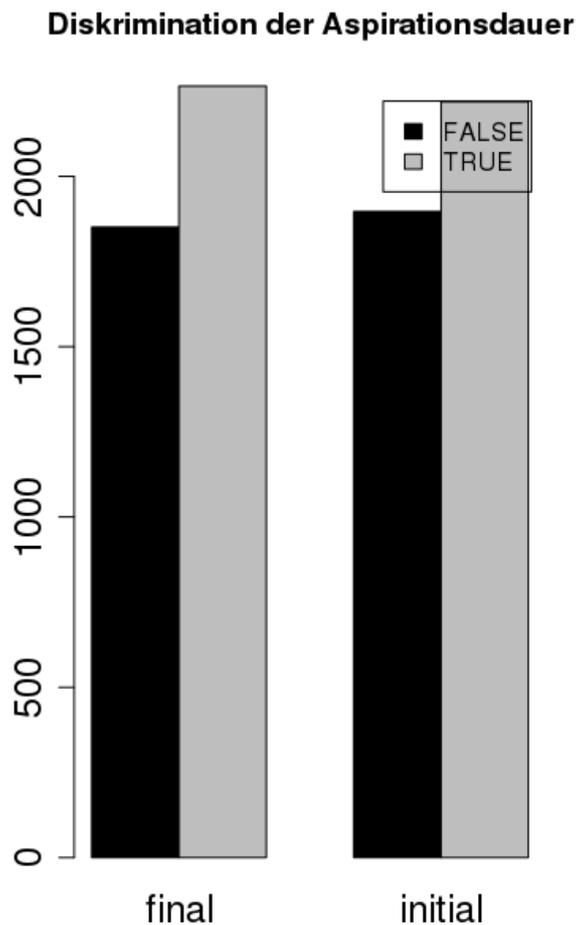


Abbildung 27: Verteilung der richtigen ("TRUE") und falschen ("FALSE") Erkennung der längeren Aspirationsdauer.

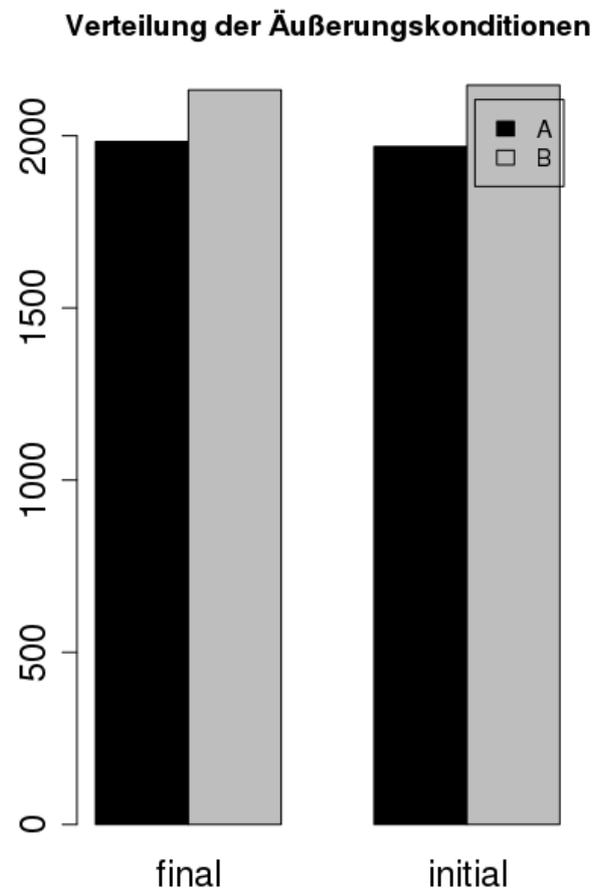


Abbildung 28: Verteilung der Äußerungskonditionen innerhalb der perceptiven Urteile getrennt für /p/ an finaler und initialer Position. "A" gibt die Äußerungen der Kondition A (vor den manipulierten Stimuli) an, B die Kondition nach dem Hören der manipulierten Stimuli.

4.4 Diskussion

Es wurde unter Verwendung von Daten, die aus einem Aufnahmesetting stammen, das spontane phonetische Imitation ermöglicht, ein Perceptionsexperiment durchgeführt. Diese Studie lieferte zwei Ergebnisse:

(1.) Versuchspersonen können so kleine Unterschiede in der Aspirationsdauer, wie sie durch spontane phonetische Imitation entstehen, diskriminieren.

(2.) Die Rückschlüsse, die durch die HörerInnenurteile auf die Imitation gezogen werden können, stimmen nicht mit den Ergebnissen der akustischen Analyse überein.

Ad (1.): Versuchspersonen können Unterschiede in der Aspirationsdauer, wie sie durch spontane Imitation entstehen, diskriminieren. Dies bietet die Möglichkeit, dass die neuen Produktionen wiederum zu perzeptivem Lernen führen könnten. Auch ein Wieder-Imitieren, welches potentiell zu Lautwandel führen kann, ist möglich. Dennoch muss beachtet werden, dass spontane phonetische Imitation immer nur als Angleichung an die Modellsprache stattfindet. Die resultierenden Veränderungen der Parameter sind sowohl anhand ihrer absoluten Werte relativ klein, als auch perzeptiv in den meisten Fällen schwer erkennbar.

Für diesen ersten Test der Studie war nicht wichtig, ob die Unterschiede in der Aspirationsdauer zwischen zwei Äußerungen durch Imitation, Konvergenz, Divergenz oder natürliche phonetische Variation entstanden sind. Vielmehr sollte und konnte beobachtet werden, dass HörerInnen Aspirationsdauern diskriminieren können, selbst wenn diese (nur) in einer Stärke auftreten, wie spontane phonetische Imitation sie verursachen kann.

Die Aufgabe der Diskrimination ist nicht grundsätzlich vergleichbar mit natürlicher Sprachverarbeitung. Im Sprachgebrauch stehen HörerInnen wohl selten vor der Aufgabe, zwei direkt aufeinander folgende Äußerungen des gleichen Wortes zu vergleichen. Dem natürlichen Sprachverarbeitungsablauf entsprechen vielmehr Identifikationsaufgaben. Dennoch bestehen auch im Prozess der perzeptiven Sprachverarbeitung in natürlichen Situationen Diskriminationsaufgaben: nämlich dann, wenn ein Perzept auffällig nicht der mentalen Vorstellung dieser Kategorie entspricht. Das heißt, zwischen den verschiedenen Einträgen bzw. dem prototypischen Vertreter dieser Kategorie und dem Perzept wird ein Unterschied festgestellt. Die kleinen Unterschiede in der Aspirationsdauer könnten bereits als sich unterscheidende Exemplare von der eigenen Produktion entdeckt werden. Ob dieser Unterschied ausreichend ist, um im Sinne von Lindblom et al. (1995) „signifikant“ zu sein oder in Trudgills (1986, S. 11) Sinne „highly salient“, kann dieses Experiment nicht beantworten. Wohl aber zeigt es, dass die Dauerunterschiede wahrnehmbar sind.

Es scheint eine schwierige Aufgabe für HörerInnen zu sein, die Dauern von /p/ zu vergleichen. Der Unterschied zwischen richtigen und falschen Antworten im Perzeptionsexperiment (TRUE/FALSE) ist nicht sehr groß, wenn auch statistisch (hoch) signifikant.

Ad (2.): Die Auswertung der Häufigkeit der Äußerungskategorien (A oder B) unter den Antworten ergab, dass perzeptiv eine Dauerzunahme der Aspiration erkennbar ist. Perzeptiv findet sich innerhalb der /p/s, sowohl an initialer als auch finaler Stelle, der Hinweis für imitatorische Beeinflussung. Dieses Ergebnis ist deshalb besonders ungewöhnlich, weil nur in den finalen Äußerungen akustische Hinweise für Unterschiede

in der Aspirationsdauer gefunden wurden. Innerhalb der hier verwendeten Äußerungen besteht in den Dauerparametern der initialen Äußerungen in keinem Segment und keiner Segmentverbindung ein Unterschied zwischen den Konditionen A und B. Die perzeptive Erkennung muss auf andere Parameter als die Dauern, relativen Dauern oder Sprechgeschwindigkeit zurückzuführen sein.

Den HörerInnen wurde die Aufgabe gestellt, ausschließlich die Dauern von /p/ zu vergleichen und deshalb Vokaldauer oder Wortdauer unberücksichtigt zu lassen. Auch diese Aufgabe entspricht nicht der natürlichen Sprachverarbeitung, da die Aspirationsdauer relativ zur Wortlänge bzw. zur Sprechgeschwindigkeit wahrgenommen wird (Miller & Volaitis, 1989; Summerfield, 1981). Die Anweisung, sich nur auf das /p/ zu konzentrieren, steht im Gegensatz zu einem Ergebnis von Miller (1987), das zeigte, dass Informationen zur Sprechgeschwindigkeit während der perzeptiven Sprachverarbeitung nicht ignoriert werden können. Die Aufgabenstellung ist dennoch sinnvoll, da kein Zuwachs an Sprechgeschwindigkeit zwischen den Kondition A und B besteht: Wie in der akustischen Auswertung (Abschnitt 4.2) gezeigt wurde, verändern sich die Dauern der einzelnen Segmente der /p/-initialen Wörter nicht und die Veränderung innerhalb der /p/-finalen Wörter ist ausschließlich auf eine Verlängerung der Aspiration zurückzuführen. Die Anweisung war notwendig, um die Aufmerksamkeit der HörerInnen auf die Dauern von /p/ zu lenken. Welche akustischen Bestandteile der Äußerungen dazu geführt haben, dass auch initial längere Aspirationsdauern in den B-Äußerungen entdeckt wurden, kann hier nicht geklärt werden.

Dieses Ergebnis bestätigt die erste Hypothese (H1) des Experiments in Abschnitt 3. Perzeptiv besteht Imitation auch an der initialen Position, obwohl diese anhand der vorgenommenen akustischen Analysen nicht messbar war. Zusätzlich zu den Analysen aus Abschnitt 3 und 4 wurde die relative RMS (*short-term Root Mean Square amplitude*; siehe Abschnitt 3.2.5) der Aspirationsdauer im Vergleich zum angrenzenden Vokal berechnet. Auch hier findet sich kein signifikanter Unterschied zwischen den Konditionen A und B (vorher/nachher), der die Diskrepanz zwischen perzeptiver und akustischer Analyse klären könnte. Auch im ersten hier dargestellten Experiment (Abschnitt 2) fanden sich Unterschiede zwischen der wahrgenommenen und der akustisch gemessenen Imitation.

Zusammenfassend liefert diese Studie zum einen einen Hinweis darauf, dass die Veränderung eines Parameters (Dauer von Aspiration) durch spontane Imitation potentiell zu folgender Imitation führen könnte, und zum anderen, dass bei der Suche nach Imitation die Ergebnisse von perzeptiven Einschätzungen und akustischen Analysen gemeinsam betrachtet werden sollten.

5. Abschließende Diskussion

In dieser Arbeit wurden drei Studien vorgestellt. Im Folgenden werden die Ergebnisse getrennt nach den zwei Datensätzen diskutiert. Abschließend wird auf das studienübergreifende Ergebnis eingegangen, dass die perzeptive und die akustische Analyse von Imitationsdaten einander ergänzende Einblicke in das Phänomen spontaner phonetischer Imitation geben.

In der ersten Studie dieser Arbeit wurde ein *Shadowing*-Experiment durchgeführt, um sprecherspezifische Aspekte der Imitation und Perzeption zu untersuchen. Im Anschluss absolvierten alle beteiligten Sprecherinnen einen Identifikationstest zu einem Vokalkontinuum (/ɛ/ zu /a/).

Bei drei der zehn Sprecherinnen wurde perzeptiv, bei fünf akustisch eine Annäherung an die Produktionen der Modellsprecherin festgestellt. Die Produktionen der fünf Sprecherinnen, die in den akustischen Tests als Imitatorinnen identifiziert wurden, wiesen im Identifikationstest einen eher flachen Verlauf zwischen den Vokalkategorien auf. Dieses Ergebnis kann darauf hindeuten, dass SprecherInnen, die flache perzeptive Kategoriengrenzen besitzen, stärker imitieren. Im umgekehrten Falle hieße das, dass SprecherInnen, die eine steile Grenze zwischen Kategorien wahrnehmen, weniger oder nicht imitieren. Die steilere Kategoriengrenze kann als Zeichen dafür gelten, dass Personen zuverlässig, bzw. zuverlässiger als andere Personen, erkennen, wann ein Laut noch zur perzipierten Kategorie gehört und wann nicht. Das heißt, der steilere Übergang bedeutet gleichzeitig eine breitere eindeutig perzipierte Vokalkategorie (siehe Abbildung 16 in Abschnitt 2.4.2; Versuchsperson AN hat einen steileren Kategorienübergang und damit eine breitere Vokalkategorie /a/ als Versuchsperson AM). Somit könnten fremde Äußerungen häufiger in den Bereich der eigenen Kategorie fallen und deshalb zu weniger Imitation führen.

Ob SprecherInnen klarer voneinander differenzierte Vokalkategorien behalten, weil sie nicht imitieren oder ob sie nicht imitieren, weil sie kompaktere, stärker voneinander getrennte Vokalkategorien aufweisen, war innerhalb dieses Experiments nicht zu klären. Das Ergebnis gibt einen Hinweis darauf, dass ein grundsätzlicher Unterschied in der perzeptiven Sprachverarbeitung und damit der phonologischen Kategorisierung besteht, der sich auch in der (sehr) geringen Beeinflussung durch umgebende Sprache zeigt.

Für die zweite und dritte Studie wurde einen Aufnahmekorpus verwendet, der durch einen Experimentablauf gewonnen wurde, der generalisierende Imitation ermöglicht.

Hierfür wurde ein geblockter Versuchsablauf durchgeführt, der aus einem ersten Leseblock, einem Hörblock, innerhalb dessen manipulierte /p/-initiale Wörter mit verlängerter Aspirationsdauer gehört wurden, und einem zweiten Leseblock bestand. Die Wörter der Leseliste trugen das Merkmal „Aspiration“ in initialem /p/, finalem /p/ und finalem /k/ (final jeweils in ein- und mehrsilbigen Wörtern).

Innerhalb dieses Korpus wurde Imitation der Aspirationsdauer nur in /p/ an der finalen Position im Wort gefunden. Imitation, die bei Laborsprache gemessen wird, ist generell relativ schwach. Deshalb muss dies nicht zwangsläufig bedeuten, dass in initialem /p/ und finalem /k/ keinerlei Imitation stattgefunden hat. Sie fällt jedoch geringer aus als in finalem /p/.

Die genaue Daueranalyse der Segmente und verschiedener Segmenteinheiten eines Teils der /p/-initialen und -finalen Wörter in der dritten Studie bestätigt die Ergebnisse der zweiten: Nur innerhalb der finalen /p/-Aspiration (bzw. nur daraus resultierend) findet sich eine Verlängerung der Dauer.

Dass initial weniger oder keine Imitation gefunden wurde, kann (Abschnitt 3.3) auf die initial stärkere artikulatorische Größe (*initial strengthening*: Keating et al., 2003) zurückgeführt werden: Initiale Artikulation dauert länger und ist stärker bezüglich Bewegung und Muskelaktivität (Fougeron, 1998). Das kann zu einer genaueren Artikulation führen, die weniger störanfällig und weniger beeinflussbar, dafür aber stabiler ist. Im Gegenzug können die weniger strengen artikulatorischen Abläufe an finaler Position zu früherer und stärkerer Beeinflussung führen.

Nach Lindblom et al. (1995) sind „soziale, kommunikative, artikulatorische und lexikalisch-systemische Kriterien“ (S. 19) entscheidend für die Akzeptanz einer neuen Aussprachevariante. Die Imitation von längerer Aspiration nur an der finalen Position in /p/ könnte als artikulatorisches Kriterium gelten, welches die längere Aspiration an der finalen Position präferiert.

Darüber hinaus kann dieses Ergebnis sehr gut entsprechend der Vorstellung von Lindblom et al. (1995) mit einem *Wie*- und einem *Was*-Modus in der Sprachverarbeitung erklärt werden. Durch die höhere Vorhersagbarkeit der finalen gegenüber der initialen Position bedarf es weniger stark des *Was*-Modus, wodurch der *Wie*-Modus eine stärkere Aktivierung in der Produktion erlangen und eine neue Aussprachevariante (hier: final verlängerte Aspiration in /p/) erzeugen kann.

Nielsen (2011b) zeigte subphonemische Imitation (bilabial auf velar) und damit die Gliederung des Lexikons in Wort-, Phonem- und subphonemische Ebene. Die hier gefundene Imitation von der initialen auf die finale Position im Wort deutet darauf hin, dass die subphonemische Flexibilität auch über die Positionen im Wort besteht. Während über die Position im Wort *und* Artikulationsstelle nicht imitiert wurde, trat bei gleicher

Artikulationsstelle über die Position im Wort die gleiche Imitationsstärke in ein- und mehrsilbigen Wörtern auf. Das heißt, die Nähe zur wahrgenommenen Position spielte keine Rolle; die Flexibilität zwischen dem Merkmal und der Position im Wort ist folglich sehr hoch.

Der Diskriminationstest zu den Dauern von /p/ initial und final ergab, dass Versuchspersonen Unterschiede in der Aspirationsdauer, wie sie durch spontane phonetische Imitation entstehen, diskriminieren können.

Dadurch besteht die Möglichkeit, dass die neuen Produktionen wiederum zu perzeptivem Lernen führen könnten. Auch ein Wieder-Imitieren, welches potentiell zu Lautwandel führen kann, ist möglich. Dennoch muss beachtet werden, dass spontane phonetische Imitation immer nur als Angleichung an die Modellsprache stattfindet. Die resultierenden Veränderungen der Parameter sind sowohl anhand ihrer absoluten Werte relativ klein als auch perzeptiv in den meisten Fällen schwer erkennbar.

Ob die kleinen Unterschiede im Sinne von Lindblom et al. (1995) ausreichend sind, um „signifikant“ zu sein oder in Trudgills (1986, S. 11) Sinne „highly salient“, kann dieses Experiment nicht beantworten; wohl aber zeigt es, dass die Dauerunterschiede wahrnehmbar sind.

Beim Vergleich der drei Studien fällt auf, dass sich die Ergebnisse der perzeptiven HörerInnenurteile und der akustischen Analyse in Bezug auf die Imitation(ssstärke) unterscheiden. Das betrifft zum einen die in der ersten Studie (Abschnitt 2) und zum anderen die in der dritten Studie (Abschnitt 4) dargestellten Daten.

In der ersten Studie wurde perzeptiv Imitation bei den Sprecherinnen AM, IM und MH erkannt. Für die Sprecherinnen AM und MH ist dieses Ergebnis deckungsgleich mit der akustischen Analyse (Imitation in F0). Bei Sprecherin IM konnte nur F2 ausgewertet werden; darin bestand keine Imitation. Auffällig ist, dass weder EE, JK noch PS, die anhand von F2 akustische Annäherung aufwiesen, von HörerInnen als Imitatorinnen erkannt wurden. Dies könnte darauf zurückzuführen sein, dass F0 als Tonhöhe des gesamten Worts perzeptiv auffälliger ist als F2 innerhalb des betonten Vokals, weil ein längerer Zeitraum betroffen ist. Darüber hinaus wurde perzeptiv bei zwei Sprecherinnen (AN und LB) Divergenz festgestellt. Akustisch ändern sich deren Äußerungen im Verlauf des Experiments weder anhand von F0 noch F2. Zusammengefasst heißt das für die erste Studie: Es bestand akustische Konvergenz, die perzeptiv nicht erkannt wurde, und perzeptive Divergenz, die in den ausgewerteten akustischen Parametern nicht sichtbar ist.

Die dritte Studie beinhaltete die genaue Analyse der anschließend im Perzeptionstest verwendeten Äußerungen. Obwohl zwischen den Konditionen eine Veränderung der Aspirationsdauer akustisch nur für /p/ an der finalen Position im Wort bestand, ergab die

Auswertung der HörerInnenurteile Imitation innerhalb der (Aspirations-)Dauer in gleicher Stärke bei den finalen und den initialen Plosiven.

Diese zwei Studien unterschieden sich anhand der Aufgabenstellung, die den HörerInnen gegeben wurde: Während in der ersten Studie die Ähnlichkeit der gesamten Äußerung zur Produktion der Modellsprecherin bewertet werden sollte, war der Perzeptionstest der dritten Studie ein reiner Diskriminationstest mit Fokus auf nur einem Segment. Es kann folglich davon ausgegangen werden, dass die Auswertung von F0 und F2 in der ersten Studie die Bestandteile, anhand derer HörerInnen ihre Urteile trafen, nicht ausreichend widerspiegelt. Zu den perzeptiven Urteilen in der dritten Studie führten Parameter, die über die akustische Dauer von Aspiration hinausgingen.

Zusammenfassend heißt das, dass die akustische und perzeptive Untersuchung von Imitation mutmaßlich gegensätzliche Ergebnisse erbringen kann. Während bei der perzeptiven Untersuchung die Wahrnehmbarkeit im Vordergrund steht, liegen die Vorteile der akustischen Analyse in ihrer analytische Genauigkeit. Es ist wichtig anzuerkennen, dass beispielsweise Imitation von F2 vorlag, auch wenn diese von HörerInnen bei einer Aufgabenstellung wie in der ersten Studie nicht erkannt wurde, und deshalb ein durch die akustische Analyse gewonnenes Ergebnis für die Imitationsforschung ein relevantes Ergebnis darstellt. Aus diesem Grund ist es sinnvoll, die Art der Auswertung an die Forschungsfrage anzupassen.

Zusammenfassend liefern die drei Studien die folgenden Ergebnisse:

- Zwischen SprecherInnen können unterschiedliche Strategien in der Sprachverarbeitung bestehen, die sowohl innerhalb ihrer Tendenz zu imitieren als auch in ihrem Umgang mit Vokalkategorien sichtbar sind.
- Die Verbindung der Merkmals- zur Wortebene im mentalen Lexikon bezüglich der Position im Wort ist sehr flexibel.
- Unterschiedliche Positionen im Wort können für Imitation unterschiedlich geeignet sein.
- Die Veränderung eines Parameters (Dauer von Aspiration) durch spontane Imitation kann potentiell zu nachfolgender Imitation führen.
- Bei der Untersuchung von Imitation sollten die Ergebnisse von perzeptiven Einschätzungen und akustischen Analysen gemeinsam betrachtet werden.

Literatur

- Allen, J. S., Miller, J. L. & DeSteno, D. (2003) Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113, 1, 544–552.
- Allen, J. S. & Miller, J. L. (2004) Listener sensitivity to individual talker differences in voice-onset-time. *Acoustic Society of America*, 115, 6, 3171-3183.
- Arbib, M. A. (2011) From Mirror Neurons to Complex Imitation in the Evolution of Language and Tool Use, *Annual Review of Anthropology*, 40, 257-273.
- Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995) *The CELEX Lexical Database (CD-ROM)*. Linguistic Data Consortium: Philadelphia, PA.
- Baayen, R. H. (2008) *Analysing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press: Cambridge.
- Babel, M. E. (2010) Dialect convergence and divergence in New Zealand English. *Language in Society*, 39, 4, 437-456.
- Babel, M. E. (2011) *Phonetic and social selectivity in speech accommodation*. ProQuest, UMI Dissertation Publishing.
- Babel, M. E. (2012) Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40, 177-189.
- Babel, M. & Bulatov, D. (2012) The Role of Fundamental Frequency in Phonetic Accommodation. *Language & Speech*, 55, 2, 231-248.
- BAS (Bayerisches Archiv für Sprachsignale, Ludwig-Maximilians-Universität München) 2009: BASat - Statistics of Conversational German (10.08.2011: <http://www.phonetik.uni-muenchen.de/forschung/Bas/BasPHONSTATeng.html>)
- Berkovits, R. (1993) Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics*, 21, 479-489.
- Boersma, P. (2001) Praat, a system for doing phonetics by computer. *Glott International*, 5, 9/10, 341-345.
- Bradlow, A. R. & Pisoni, D. B. (1999) Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America*, 106, 2074–2085.
- Branigan, H. P., Pickering, M. J. & Cleland, A. A. (2000) Syntactic coordination in dialogue. *Cognition*, 75, 2, B13–B25.
- Browman, C. P. & Goldstein, L. (1992) Articulatory Phonology: An Overview. *Phonetica*, 49, 155–180.
- Browman, C. & Goldstein, L. (1995) Gestural syllable position effects in American English, in *Producing Speech: A Festschrift for Katherine Safford Harris*, F. Bell-Berti & L. J. Raphael (Hrsg.), American Institute of Physics Press: Woodbury, NY, 19-34.
- Byrd, D. (1992) Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, 92, 593–596.
- Byrd, D. M. (1994) Articulatory timing in English consonant sequences. *Ph.D. dissertation, UCLA, distributed as UCLA Working Papers in Phonetics*, 86, 1-196.
- Byrd, D. M. (1996) Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209-244.
- Cherry, E. C. (1956) *On human communication*. MIT Press: Cambridge, MA.

- Church, B. A. & Schacter, D. L. (1994) Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 521–533.
- Christovich, I. A. & Chernova, E. I. (1986) Identification of one- and two-formant steady-state vowels: A model and experiments. *Speech Communication*, 5, 3–16.
- Delvaux, V. & Soquet, A. (2007) The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64, 2-3, 145–173.
- Draxler, C. & Jänsch, K. (2004) SpeechRecorder - A universal platform independent multi-channel audio recording software. *Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC)*. Lisbon, Portugal, 559–562.
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399-402.
- Flege, J. E., Bohn, O. & Jang, S. (1997) Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, J. E., Munro, M. J. & MacKay, I. R. (1995) Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.
- Fougeron, C. (1998) Variations articulatoires en début de constituants prosodiques de différents niveaux en français. Ph.D. dissertation, Université Paris III - Sorbonne Nouvelle.
- Fowler, C. A. (1986) An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3-28.
- Fowler, C. A. (1996) Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730-1741.
- Fowler, C. A., Brown, J. M., Sabadini, L. & Weihing, J. (2003) Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49, 396–413.
- Fox, R. A. (1983) Perceptual structure of monophthongs and diphthongs in English. *Language and Speech*, 26, 21–60.
- Fujimura, O. (1977) Recent findings on articulatory processes, in *Articulatory modeling and phonetics*, R. Carre, R. Descout & M. Wajskop (Hrsg.), Groupement des Acousticiens de Langue Française: Grenoble, 115-126.
- Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems*. Houghton Mifflin: Boston.
- Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin: Boston.
- Gimson, A. C. (1966) *An Introduction to the Pronunciation of English*. Arnold: London.
- Goldinger, S. D. (1996) Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D. (1997) Words and voices: Perception and production in an episodic lexicon, in *Talker Variability in Speech Processing*, K. Johnson & J. W. Mullennix (Hrsg.), Academic Press: San Diego, 33-66.
- Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279.

- Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts, in J. Cole & J. Hualde (Hrsg.) *Laboratory Phonology 9*. Mouton: Berlin, 125-143.
- Harrington, J. (2010) *The Phonetic Analysis of Speech Corpora*. Wiley-Blackwell: Chichester.
- Harrington, J., Kleber, F. & Reubold, U. (2008) Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, 123, 2825-2835.
- Harrington, J., Palethorpe, S. & Watson, C. (2000) Does the Queen speak the Queen's English? *Nature*, 408, 927-928.
- Henton, C. (1983) Changes in the vowels of received pronunciation. *Journal of Phonetics*, 11, 353-371.
- Hintzman, D. L. (1986) "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Honorof, D. N., Weihing, J. & Fowler, C. A. (2011) Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39, 18-38.
- Horner, V. & Whiten, A. (2005) Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, 8, 164-181.
- Jilka, M. (2009a) Assessment of phonetic ability, in *Language Talent and Brain Activity, volume 1 of Trends in Applied Linguistics*, G. Dogil & S. M. Reiterer (Hrsg.), de Gruyter: Berlin, 17-66.
- Jilka, M. (2009b) Talent and proficiency in language, in *Language Talent and Brain Activity, volume 1 of Trends in Applied Linguistics*, G. Dogil & S. M. Reiterer (Hrsg.), de Gruyter: Berlin, 1-16.
- Johnson, K. (1997) Speech perception without speaker normalization: An exemplar model, in *Talker Variability in Speech Processing*, K. Johnson & J. W. Mullennix (Hrsg.), Academic Press: San Diego, 145-165.
- Johnson, K. (2005) Speaker Normalization in speech perception, in *The Handbook of Speech Perception*, D. B. Pisoni & R. T. Remez (Hrsg.), Blackwell Publishers: Oxford, 363-389.
- Kappes, J., Baumgaertner, A., Peschke, C. & Ziegler, W. (2009) Unintended imitation in nonword repetition. *Brain & Language*, 111, 140-151.
- Keating, P., Cho, T., Fougeron, C. & Hsu, C.-S. (2003) Domain-initial articulatory strengthening in four languages. *Laboratory Phonology*, 6, 145-163.
- Keating, P., Wright, R. & Zhang, J. (1999) Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics*, 97, 157-173.
- Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E. & Fowler, C. A. (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 6, 812-832.
- Kewley-Port, D. & Watson, C. S. (1994) Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 95, 485-496.
- Klatt, D. H., and Klatt, L. C. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.

- Kleber, F., John, T. & Harrington, J. (2010) The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 38, 185-196.
- Kohler, K. & Hardcastle, W. (1974) The instability of final alveolars in English and German. *Proceedings of the Stockholm Speech Communication Seminar*, Stockholm, Sweden, 95-98.
- Kraljic, T. & Samuel, A. G. (2006) Generalisation in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13, 2, 262-268.
- Kuhl, P. K. & Meltzoff, A. N. (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 4, 2425-2438.
- Ladefoged, P. & Broadbent, D. E. (1957) Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29, 98-104.
- Lehiste, I. (1970) *Suprasegmentals*. MIT Press: Cambridge, MA.
- Lelong, A. & Bailly, G. (2011) Study of the phenomenon of phonetic convergence thanks to speech dominoes, in *Analysis of Verbal & Nonverbal Communication & Enactment: The Processing Issues*, A. Esposito, A. Vinciarelli, K. Vicsi, C. Plachaud & A. Nijholt (Hrsg.), Springer Verlag: Berlin, 280-293.
- Levelt, W. J. M. (1989) *Speaking: From intention to articulation*. MIT Press: Cambridge, MA.
- Levitan, R. & Hirschberg, J. (2011) Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech*, 2011.
- Lewandowski, N. (2012) *Talent in nonnative phonetic convergence*. Dissertation. URN: urn:nbn:de:bsz:93-opus-74023, URL: <http://elib.uni-stuttgart.de/opus/volltexte/2012/7402/> (11.09.2012)
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P. & M. Studdert-Kennedy (1967) Perception of the speech code. *Psychological Review*, 74, 431-461.
- Lieberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. *Cognition*, 21, 1, 1-36.
- Lieberman, P., Ryalls, J. & Rabson, S. (1982) Some acoustic aspects of early imitation of speech: F0 and vowels. *Journal of the Acoustical Society of America*, Suppl.1, Vol.72, Fall.
- Lindblom, B. (1992) Phonological units as adaptive emergents of lexical development, in *Phonological Development: Models, research, implications*, C. A. Ferguson, L. Menn & C. Stoel-Gammon (Hrsg.), York Press: Timonium, MD, 131-163.
- Lindblom, B. (1998) Systemic constraints and adaptive change in the formation of sound structure, in *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy & C. Knight (Hrsg.), Cambridge University Press: Cambridge, 242-264.
- Lindblom, B., Brownlee, S., Davis, B. & Moon, S.-J. (1992) Speech transforms. *Speech Communication*, 11, 357-368.
- Lindblom, B., Guion, S., Hura, S., Moon, S.-J. & Willerman, R. (1995) Is sound change adaptive? *Rivista di Linguistica*, 7, 1, 5-36.
- Luce, P. A. & McLennan, C. (2005) Spoken word recognition: The challenge of variation, in *Handbook of Speech Perception*, D. B. Pisoni & R. E. Remez (Hrsg.), Blackwell Publishers: Oxford, 591-609.
- Makashay, M. J. (1999) Dialectal differences in diphthong perception. *The Journal of the Acoustical Society of America*, 106, 2273-2273.

- Makashay, M. J. (2003) *Individual differences in speech and non-speech perception of frequency and duration. Doctoral thesis.* The Ohio State University. <http://www.ling.ohio-state.edu/~makashay/makashay.pdf> (12.09.2012)
- Mattingly, I. G. (1991) The global character of phonetic gestures. *Journal of Phonetics*, 18, 445-452.
- McLennan, C. T., Luce, P. A. & Charles-Luce, J. (2003) Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539–553.
- McQueen, J. M., Cutler, A. & Norris, D. (2006) Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 6, 1113-1126.
- Meltzoff, A. N. (2002). Elements of a developmental theory of imitation, in A. N. Meltzoff & W. Prinz (Hrsg.), *The Imitative Mind. Development, Evolution, and Brain Bases.* Cambridge University Press: Cambridge, 19-41.
- Meltzoff, A. N. & Moore, M. K. (1977). Imitation of Facial and Manual Gestures by Human Neonates, *Science*, 198, 75-78.
- Meltzoff, A. N. & Moore, M. K. (1983). Newborn Infants Imitate Adult Facial Gestures, *Child Development*, 54, 702-709.
- Melzoff, A. N. & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms, *Developmental Psychology*, 25, 954-962.
- Miller, J. L.(1987) Rate-dependent processing in speech perception, in *Progress in the Psychology of Language*, A. W. Ellis (Hrsg.), Lawrence Erlbaum.
- Miller, J. L. & Volaitis, L. E. (1989) Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*. 46, 505–512.
- Mitterer, H. & Ernestus, M. (2008) The link between speech perception and production is phonological and abstract: Evidence from the shadowing task, *Cognition*, 109, 168–173.
- Moyer, A. (1999) Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, 21, 81–108.
- Neufeld, G. (1987) On the Acquisition of Prosodic and Articulatory Features in Adult Language Learning, in G. Ioup & S. Weinberger (Hrsg.), *Interlanguage Phonology.* Newburg House: Cambridge, MA, 321-332.
- Newman, R. S. (2003) Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America*, 113, 5, 2850-2860.
- Newman, R. S., Clouse, S. A. & Burnham, J. L. (2001) The perceptual consequences of within talker variability in fricative production, *Journal of the Acoustical Society of America*, 109, 1181–1196.
- Nielsen, K. (2011a) Phonetic imitation by school-age children. *Journal of the Acoustical Society of America*, 130, 4/2, 2444.
- Nielsen, K. (2011b) Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39, 132–142.
- Nielsen, M. & Tomaselli, K. (2009) Over-imitation in the Kalahari Desert and the origins of human cultural cognition. *Nature Precedings*, 1-18. <http://precedings.nature.com/documents/3049/version/1/files/npre20093049-1.pdf>
- Norris, D., McQueen, J. M. & Cutler, A. (2003) Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.

- Nosofsky, R. M. (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M. (1988) Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 4, 700-708.
- Nye, P. W. & Fowler, C. A. (2003) Shadowing latency and imitation: the effect of familiarity with the phonetic patterning of English. *Journal of Phonetics*, 31, 63-79.
- Nygaard, L. C., and Pisoni, D. B. (1998) Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355–376.
- Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. (1994) Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–46.
- O'Shaughnessy, D. (1995) Timing patterns in fluent and disfluent spontaneous speech, *Acoustics, Speech, and Signal Processing*, 1, 600–603.
- Ohala, J. J. (1981) The listener as a source of sound change, in *Papers from the Parasession on Language and Behavior*, C. S. Masek, R. A. Hendrick & M. F. Miller (Hrsg.), Chicago Ling. Soc.: Chicago, 178-203.
- Ohala, J. J. (1993) The phonetics of sound change, in *Historical Linguistics: Problems and Perspectives*, C. Jones (Hrsg.), Longman: London, 237-278.
- Palmeri, T. J., Goldinger, S. D. & Pisoni, D. B. (1993) Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309-328.
- Pape, D. & Mooshammer, C. (2006) Is Intrinsic pitch language-dependent? Evidence from a cross-linguistic vowel pitch perception experiment. *Proceedings of the ISCA Tutorial and Research Workshop (ITRW) on Multilingual Speech and Language Processing*, Stellenbosch, South Africa.
http://www.haskins.yale.edu/staff/mooshammer/pape_mul.pdf (27.04.2013)
- Pardo, J. S. (2006) On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382-2393.
- Pardo, J. S. (2012a) Phonetic convergence in college roommates. *Journal of Phonetics*, 40, 190-197.
- Pardo, J. S. (2012b) Reflections on Phonetic Convergence: Speech perception does not mirror production. *Language & Linguistics Compass*, 6, 753-767.
- Pardo, J. S., Cajori Jay, I. & Krauss, R. M. (2010) Conversational role influences speech imitation. *Attention, Perception & Psychophysics*, 72, 2254-2264.
- Peeters, W. J. M. (1991) *Diphthong Dynamics: A Cross-Linguistic Perceptual Analysis of Temporal Patterns in Dutch, English and German*. Mondiss: Kampen, NL.
- Peterson, B. & Lehiste, I. (1960), Duration of syllable nuclei in English, *Journal of the Acoustical Society of America*, 32, 693–703.
- Peterson, G. E., and Barney, H. L. (1952) Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Pickering M. J. & Garrod S. (2004) Towards a mechanistic Psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-226.
- Pickering, M.J. & Garrod, S. (2005) *Establishing and using routines during dialogue: implications for psychology and linguistics*, in A. Cutler (Hrsg.) *Twenty-First Century Psycholinguistics: Four Cornerstones*. Lawrence Erlbaum, 85-102.

- Pickering, M. J. & Garrod, S. (2006) Alignment as the Basis for Successful Communication. *Research on Language and Computation*, 4, 2-3, 203-228.
- Pickering M. J. & Garrod S. (2007) Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110.
- Pierrehumbert, J. (2003) Probabilistic Phonology: Discrimination and Robustness, in *Probability Theory in Linguistics*, R. Bod, J. Hay & S. Jannedy (Hrsg.), MIT Press: Cambridge, MA, 177-228.
- Pompino-Marschall, B. (2003) *Einführung in der Phonetik*. Walter de Gruyter: Berlin, New York.
- Port, R. F. (1981) On the structure of the phonetic space with special reference to speech timing, *Lingua*, 55, 181-219.
- Reiterer, S., Singh, NC. & Winkler, S. (2012) Predicting speech imitation ability biometrically, in *Empirical Approaches to Linguistic Theory. Studies in Meaning and Structure*, B. Stollerfoht & S. Featherston (Hrsg.), Mouton de Gruyter: Berlin, New York.
- Remez, R. E., Fellowes, J. M. & Rubin, P. E. (1997) Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651– 666.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981) Speech perception without traditional speech cues. *Science*, 212, 947–950.
- Rizzolatti, G., Craighero, L. & Fadiga, L. (2002): The mirror system in humans, in *Mirror Neurons and the Evolution of Brain and Language*, M. I. Stamenov & V. Gallese (Hrsg.), John Benjamins Publishing Company: Amsterdam, NL, 37-59.
- Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.
- Roach, P. & Hartman, J. (1997) *English Pronouncing Dictionary*. Cambridge University Press: Cambridge, MA.
- Roug, L., Landberg, I. & Lundberg, L. (1989) Phonetic development in early infancy: a study of four Swedish children during the first eighteen months of life. *Journal of Child Language*, 16, 19-40.
- Schiel, F. (2004) MAUS goes iterative. *Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC)*. Lisbon, Portugal, 1015–1018.
- Schouten, B., Gerrits, E. & van Hessen, A. (2003) The end of categorical perception as we know it. *Speech Communication*, 41, 71–80.
- Schultz-Coulon, H.-J. (1975) Bestimmung und Beurteilung der individuellen mittleren Sprechstimmlage. *Folia phoniatica*, 27, 375–386.
- Sebastian-Galles, N. & Baus, C. (2005) On the relationship between perception and production in L2 categories, in *Twenty-first century psycholinguistics: Four cornerstones*, A. Cutler (Hrsg.), Erlbaum: New York, NY, 279–292.
- Sheffert, S. M., and Fowler, C. A. (1995) The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language*, 34, 665–685.
- Shockley, K., Sabadini, L. & Fowler, C. A. (2004) Imitation in shadowing words. *Perception & Psychophysics*, 66, 3, 422-429.
- Skehan, P. (2003) *A cognitive approach to language learning*. Oxford applied linguistics. Oxford University Press: Oxford.

- Smith, R., and Hawkins, S. (2012) Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics*, 40, 2, 213-233.
- Studdert-Kennedy, M. (1987) The phoneme as a perceptuomotor structure, in *Language Perception and Production*, A. Allport, D. G. MacKay, W. Prinz & E. Scheerer (Hrsg.), Academic: London, 67-84.
- Studdert-Kennedy, M. (1991) Language development from an evolutionary perspective, in *Biological and Behavioral Determinants of Language Development*, N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & C. Knight (Hrsg.), Erlbaum: Hillsdale, NJ, 5-28.
- Studdert-Kennedy, M. (1998) The particulate origins of language generativity: from syllable to gesture, in *Approaches to the Evolution of Language: Social and cognitive bases*, J. R. Hurford, M. Studdert-Kennedy & C. Knight (Hrsg.), Cambridge University Press: Cambridge, 202-221.
- Studdert-Kennedy, M. (2000) Evolutionary Implications of the Particulate Principle: Imitation and the Dissociation of Phonetic Form from Semantic Function, in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy & J. R. Hurford (Hrsg.), Cambridge University Press: Cambridge
- Studdert-Kennedy, M. (2002) Mirror neurons, vocal imitation and the evolution of particulate speech, in *Mirror Neurons and the Evolution of Brain and Language*, M. I. Stamenov & V. Gallese (Hrsg.), John Benjamins Publishing Company: Amsterdam, NL, 209-227.
- Summerfield, Q. (1981) Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074–1095.
- Trudgill, P. (1981) Linguistic accommodation: Sociolinguistic observations on a sociopsychological theory, in *Papers from the Parasession on Language and Behavior*, R. Hendrick, C. Masek & M. F. Miller (Hrsg.), Chicago Linguistics Society: Chicago, 218–237.
- Trudgill, P. (1986) *Dialects in contact*. Basil Blackwell: Oxford.
- Trudgill, P. (2004) *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh University Press: Edinburgh.
- Trudgill, P. (2008) Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, 37, 241- 280.
- Watkins, K. E., Strafella, A. P. & Paus, T. (2003) Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 48, 989–994.
- Wright, R. (2003) Factors of lexical competition in vowel articulation, in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, J. Local, R. Ogden & R. Temple (Hrsg.), Cambridge University Press: Cambridge, 75-87.
- Zhukov, S. J. & Chistovich, L. A. (1985) Perceptual parameters of front vowels. *Speech Communication*, 4, 267–273.
- Zwaan, R. A. & Radvansky, G. A. (1998) Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–85.

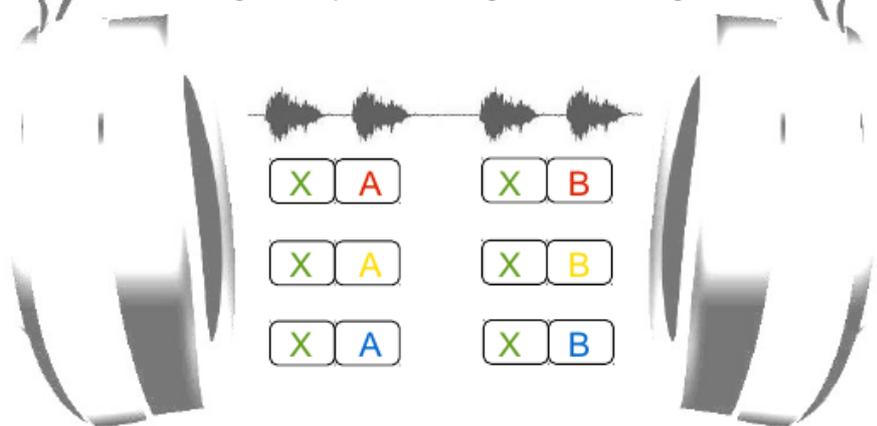
Web-Quelle:

AKUSTYK: www.bartus.org

Anhang 1: Anweisungen HörerInnen (XAXB)

Versuchsanleitung

Das vorliegende Experiment ist folgendermaßen aufgebaut:

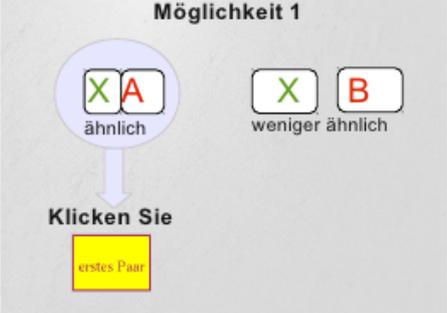


Alle „X-Äußerungen“ stammen von einer Sprecherin (markiert durch die Farbe grün).
Die Äußerungen A und B stammen von mehreren anderen Sprecherinnen, wobei A und B pro Durchgang von der selben Sprecherin produziert werden (erste Zeile 2x rot, zweite Zeile 2x gelb usw.).

Ihre Aufgabe ist zu entscheiden:
Sind sich **X und A oder **X und B** ähnlicher?**

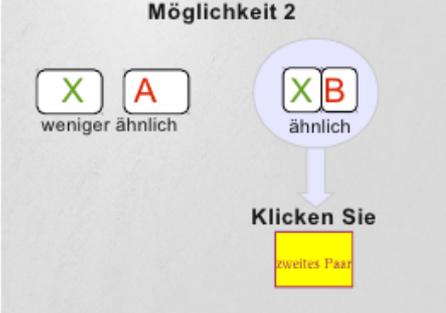
Anders ausgedrückt:
Ist X ähnlicher zu A oder B?

Möglichkeit 1



Klicken Sie
erstes Paar

Möglichkeit 2



Klicken Sie
zweites Paar

Das Experiment beginnt mit einem Testblock, in dem Sie 5 Beispiele hören. Danach haben Sie noch Zeit um Fragen zu stellen.

SOWEIT ALLES KLAR ?

Anhang 2: Unterschiede der Formantfrequenzen

Verglichen werden die Formantfrequenzen der Zielvokale der Modellsprecherin mit denen der Sprecherinnen bei der ersten Wiederholung der *Shadowing*-Wörter. Diese gelten als Baseline weil sie am wenigsten von der Modellsprecherin beeinflusst sind, aber in der gleichen experimentellen Umgebung aufgenommen wurden, wie die späteren *Shadowing*-Äußerungen.

	<i>/ɛ/</i> (Kondition 1)		<i>/a/</i> (Kondition 1)	
	F1	F2	F1	F2
AM	t = 6.55, df = 18, p < 0.001	t = -4.33, df = 18, p < 0.001	W = 30, p > 0.05	t = 5.45, df = 10.848, p < 0.001
AN	W = 70, p = 0.14	t = 5.07, df = 18, p < 0.001	W = 61, p > 0.05	t = 4.83, df = 10.691, p < 0.001
EE	t = 0.68, df = 18, p > 0.05	t = 0.44, df = 18, p > 0.05	W = 29, p > 0.05	t = 4.44, df = 9.733, p < 0.01
IB	t = 0.67, df = 10.7, p > 0.05	t = -4.63, df = 18, p < 0.001	W = 41, p > 0.05	t = 3.59, df = 9.103, p < 0.01
IM	t = -0.39, df = 18, p > 0.05	t = 0.79, df = 18, p > 0.05	W = 58, p > 0.05	t = 4.10, df = 16, p < 0.001
JK	t = 3.60, df = 18, p < 0.01	t = -3.90, df = 18, p < 0.01	W = 58, p > 0.05	t = 6.54, df = 16, p < 0.001
LB	t = 2.25, df = 18, p < 0.05	t = 4.02, df = 18, p < 0.001	W = 63, p > 0.05	t = 7.56, df = 16, p < 0.001
MH	t = 0.61, df = 18, p > 0.05	t = 0.42, df = 18, p > 0.05	W = 54, p > 0.05	t = 5.82, df = 10.861, p < 0.001
PS	t = 3.21, df = 18, p < 0.01	t = 1.75, df = 18, p > 0.05	W = 58, p > 0.05	t = 10.56, df = 16, p < 0.001
SA	t = -0.16, df = 18, p > 0.05	t = -2.48, df = 18, p < 0.05	W = 55, p > 0.05	t = 3.53, df = 9.227, p < 0.01

Anhang 3: Formantwerte der /ε/-/a/-Manipulation

	/ε/ 1		2		3		4		5		6		7	
	F1 Hz	F1 Bark												
1	381.27	3.84	372.69	3.75	364.18	3.67	355.73	3.59	347.34	3.51	339.01	3.42	330.74	3.34
2	472.2	4.68	473.14	4.68	474.07	4.69	475.01	4.7	475.95	4.71	476.89	4.72	477.82	4.72
3	623.63	5.94	626.74	5.97	629.86	5.99	632.98	6.01	636.12	6.04	639.26	6.06	642.41	6.09
4	653.12	6.17	666.99	6.28	681.01	6.38	695.19	6.49	709.51	6.6	723.99	6.7	738.63	6.81
5	648.55	6.14	672.56	6.32	697.01	6.5	721.93	6.69	747.31	6.87	773.18	7.05	799.55	7.24
6	601.32	5.76	622.17	5.93	643.36	6.1	664.9	6.26	686.81	6.43	709.08	6.59	731.73	6.76
7	511.55	5.02	530.22	5.18	549.18	5.34	568.43	5.5	587.98	5.66	607.83	5.82	627.99	5.98
8	506.71	4.98	521.43	5.1	536.33	5.23	551.4	5.36	566.66	5.48	582.11	5.61	597.75	5.74
9	500.97	4.93	515.79	5.06	530.8	5.18	545.98	5.31	561.35	5.44	576.91	5.57	592.67	5.69
10	491.54	4.85	504.63	4.96	517.86	5.07	531.23	5.19	544.75	5.3	558.41	5.41	572.23	5.53
	F2 Hz	F2 Bark												
1	1588.21	11.47	1579.22	11.43	1570.28	11.4	1561.38	11.36	1552.53	11.32	1543.72	11.28	1534.96	11.24
2	1680.69	11.85	1654.2	11.74	1628.09	11.64	1602.36	11.53	1576.99	11.42	1551.98	11.32	1527.33	11.21
3	1830.55	12.42	1790.46	12.27	1751.2	12.12	1712.76	11.97	1675.1	11.82	1638.22	11.68	1602.07	11.53
4	1866.57	12.55	1827.09	12.4	1788.42	12.26	1750.53	12.12	1713.4	11.98	1677.01	11.83	1641.33	11.69
5	1912.41	12.71	1877.99	12.59	1844.18	12.47	1810.97	12.35	1778.32	12.22	1746.24	12.1	1714.7	11.98
6	2059.35	13.21	2020.71	13.08	1982.8	12.95	1945.6	12.83	1909.11	12.7	1873.28	12.57	1838.12	12.44
7	2140.3	13.46	2101.63	13.34	2063.68	13.22	2026.43	13.1	1989.86	12.98	1953.96	12.85	1918.71	12.73
8	2128.96	13.43	2103.78	13.35	2078.91	13.27	2054.34	13.19	2030.07	13.11	2006.08	13.03	1982.39	12.95
9	2044.36	13.16	2035.7	13.13	2027.08	13.1	2018.49	13.07	2009.94	13.04	2001.43	13.02	1992.95	12.99
10	2269.68	13.86	2241.27	13.77	2213.23	13.69	2185.57	13.6	2158.27	13.52	2131.33	13.44	2104.73	13.35
8	9		10		11		12		13		/a/ 14			
F1 Hz	F1 Bark	F1 Hz	F1 Bark	F1 Hz	F1 Bark	F1 Hz	F1 Bark	F1 Hz	F1 Bark	F1 Hz	F1 Bark	F1 Hz	F1 Bark	
	322.53	3.26	314.37	3.18	306.28	3.09	298.24	3.01	290.26	2.93	282.34	2.85	274.47	2.76
	478.76	4.73	479.71	4.74	480.65	4.75	481.59	4.76	482.53	4.77	483.48	4.77	484.42	4.78
	645.56	6.11	648.73	6.14	651.9	6.16	655.08	6.19	658.26	6.21	661.46	6.23	664.66	6.26
	753.43	6.91	768.39	7.02	783.51	7.13	798.81	7.23	814.28	7.34	829.92	7.45	845.74	7.55
	826.44	7.42	853.85	7.61	881.81	7.79	910.32	7.97	939.42	8.16	969.11	8.34	999.42	8.52
	754.76	6.92	778.2	7.09	802.04	7.26	826.3	7.42	850.99	7.59	876.12	7.75	901.71	7.92
	648.47	6.14	669.28	6.29	690.42	6.45	711.91	6.61	733.75	6.77	755.94	6.93	778.51	7.09
	613.58	5.86	629.6	5.99	645.83	6.11	662.26	6.24	678.91	6.37	695.76	6.49	712.83	6.62
	608.62	5.82	624.77	5.95	641.13	6.08	657.69	6.21	674.47	6.33	691.47	6.46	708.68	6.59
	586.2	5.64	600.32	5.76	614.6	5.87	629.04	5.98	643.64	6.1	658.41	6.21	673.35	6.33
F2 Hz	F2 Bark	F2 Hz	F2 Bark	F2 Hz	F2 Bark	F2 Hz	F2 Bark	F2 Hz	F2 Bark	F2 Hz	F2 Bark	F2 Hz	F2 Bark	
	1526.23	11.21	1517.56	11.17	1508.92	11.13	1500.33	11.09	1491.78	11.06	1483.28	11.02	1474.81	10.98
	1503.01	11.11	1479.04	11	1455.39	10.89	1432.07	10.79	1409.06	10.68	1386.36	10.58	1363.97	10.47
	1566.64	11.38	1531.91	11.23	1497.85	11.08	1464.46	10.94	1431.7	10.79	1399.56	10.64	1368.03	10.49
	1606.34	11.55	1572.03	11.4	1538.37	11.26	1505.35	11.12	1472.94	10.97	1441.13	10.83	1409.91	10.69
	1683.69	11.86	1653.21	11.74	1623.23	11.62	1593.74	11.49	1564.73	11.37	1536.2	11.25	1508.12	11.13
	1803.59	12.32	1769.69	12.19	1736.39	12.06	1703.68	11.94	1671.55	11.81	1639.97	11.68	1608.94	11.56
	1884.08	12.61	1850.07	12.49	1816.66	12.37	1783.82	12.24	1751.56	12.12	1719.84	12	1688.66	11.88
	1958.98	12.87	1935.84	12.79	1912.98	12.71	1890.38	12.63	1868.04	12.55	1845.96	12.47	1824.14	12.39
	1984.51	12.96	1976.11	12.93	1967.74	12.9	1959.41	12.87	1951.11	12.84	1942.85	12.82	1934.62	12.79
	2078.49	13.27	2052.57	13.18	2026.99	13.1	2001.74	13.02	1976.8	12.93	1952.17	12.85	1927.85	12.76

Anhang 4: Wörter der Lese- und Hörliste

Wörter der Hörliste			Wörter der Leseliste				
			auslaut- verhärtet				
initial /p/	Füllwörter		final /p/	final /k/	initial /p/	final /p/	Füllwörter
Paarung	pinnen	lahm	Depp	Deck	Paarung	Bub	lang
Pacht	Pinsel	lauwarm	kipp	Dock	Paechter	Dieb	Lärche
Paddel	pinseln	Lehne	klapp	Dreck	Palme	Grab	lahm
Pächter	Piste	Leim	Klipp	Druck	Panda	grob	Leerlauf
Palme	pochen	Leuchte	knapp	Eck	Panne	Hieb	Lehm
panisch	Podium	Lift	Krepp	Fleck	Panscher	Hub	lichterloh
Panne	Polohemd	löchern	Mopp	Lack	Pate	Laub	Liefertermin
panschen	Polster	Lohnsteuer	Pepp	Leck	pausen	Leib	löslich
Panther	polstern	lüften	schlapp	Rock	Pech	lieb	loseisen
panzern	poltern	Mahl	schlepp	Ruck	Peiler	Lob	Luft
Partnerschaft	Pore	Meersalz	schnapp	Sack	Pein	Raub	lustlos
Party	Porto	meilenweit	Schwupp	Scheck	peitschen	Schub	machtlos
Pascha	portofrei	Meute	Stopp	Schlick	Pendel	Sieb	Meeresstille
Pass	Portwein	mies	stopp	Schluck	Pensum	Stab	Meilenstein
Passamt	Pose	Minze	Tipp	Schmuck	Perle	Staub	Nachfahr
Passform	Postamt	Mohn	Tripp	Schreck	perlen	taub	Naht
passiv	Postauto	Murmel	Trupp	Stock	Pest	Trab	nieseln
Passstelle	Posthorn	Müsli	Anlagetipp	Strick	Pestbeule	Trieb	Nische
Passwort	Postleitzahl	Nachlese	Bärlapp	Stuck	petzen	trüb	Nomen
Paste	Puffer	Nässe	Baustopp	Wrack	piepsen	Weib	Notwehr
Pastor	Pulle	Neon	Bautrupp	Aufdruck	pinnen		Rahmen
Pate	Pulli	Nestwärme	Boxenstopp	Technik	Pinsel		Rast
Patenschaft	Puls	Nichte	Geheimtipp	Dreieck	Pöbel		Reife
Pater	Pult	Niere	Haarklipp	Ethik	pöbeln		Retter
Pathos	Pulver	nölen	Küchenkrepp	Fettfleck	pochen		Riecher
patzen	Pulverfass	Notfall	Kurztrip	Hektik	Pol		rieseln
pausenlos	pulverfein	nuscheln	Lieferstopp	Hochdruck	Pollen		Roller
Pech	pulvern	nutzlos	Lohnstopp	Mimik	Pore		Rotfuchs
Pein	Pulverschnee	rächen	Nachklapp	Minirock	Pudel		Rufweite
Peitsche	Puma	Rahm	Notstopp	Rohrstock	pudern		Rutsche
peitschen	Punsch	rechnerisch	Philipp	Sandsack	Pulle		
pellern	purzeln	Reh	Stoßtrupp	Schandfleck	pulvern		
Pelz	Pustel	Reiher	Suchtrupp	Schaschlik	Punsch		
Penne	pusten	Riff	Verkaufstipp	Skistock	pur		
pennen	Pute	Röhre	Volldepp	Sonnendeck	purzeln		
Pensum	Puter	Ross	Wischmopp	Statik	pusten		
Perlon	Putsch	Rührei	Zwischenstopp	Thermik	Puter		

Anhang 5: Erklärungen zum Experiment

Erklärungen zum Experiment

Vielen Dank, dass du gekommen bist, um teilzunehmen!!

Deine Aufgabe

In dieser Studie soll die Dauer von P eingeschätzt werden. Hierfür hörst du pro Durchgang ein Wort, das von der gleichen Sprecherin bzw. dem gleichen Sprecher zweimal gesagt wurde. Das P befindet sich entweder am Anfang oder am Ende vom Wort. Deine Entscheidung, welches P länger ist, teilst du durch einen Klick auf die entsprechende Schaltfläche mit (siehe unten), daraufhin werden dir die nächsten zwei Äußerungen vorgespielt. Häufig wird es nicht leicht sein, sofort eine Entscheidung zu treffen. Dafür kannst du dir die Äußerungen mehrfach anhören, wenn du einfach auf "nochmal anhören" klickst.

Besondere Aufmerksamkeit: Konzentriere dich nur auf das P

Unterschiede in der Tonhöhe der Äußerungen, in den anderen Lauten oder in der Betonung sind egal. Bei manchen Äußerungen gibt es langes Ausatmen am Ende des Worts - wenn es eindeutig Ausatmen ist, ist das kein Bestandteil des Ps.

Die Entscheidung kann oft sehr schwierig sein. Wenn nötig ist Raten deshalb erlaubt!

Welches P ist länger?

Erstes P ist länger

Zweites P ist länger

nochmal anhören

The image shows a screenshot of a computer interface for an experiment. At the top, the question "Welches P ist länger?" is displayed. Below the question are two rectangular buttons: "Erstes P ist länger" on the left and "Zweites P ist länger" on the right. At the bottom center, there is a dark rectangular button labeled "nochmal anhören".