

---

**Christina Widera**

**ZUR REDUKTION VON VOKALEN**

**Eine experimental-phonetische Untersuchung**

---



**Cuvillier Verlag Göttingen**

ZUR REDUKTION VON VOKALEN  
Eine experimental-phonetische Untersuchung

Inaugural-Dissertation  
zur  
Erlangung der Doktorwürde  
der

Philosophischen Fakultät  
der  
Rheinischen Friedrich-Wilhelms-Universität  
zu Bonn

vorgelegt von  
Christina Widera

aus  
Moers

Bonn 2002

Gedruckt mit Genehmigung der Philosophischen  
Fakultät der Rheinischen Friedrich-Wilhelms-  
Universität Bonn

Berichterstatter: Professor Dr. Wolfgang Hess  
Berichterstatter: Professor Dr. Winfried Lenders

Tag der mündlichen Prüfung: 10. Juli 2002

An dieser Stelle möchte ich den Mitarbeiterinnen und Mitarbeitern des IKPs sowie allen anderen Personen, die mich während der ganzen Zeit begleitet und unterstützt haben, meinen Dank aussprechen.

Insbesondere möchte ich mich bedanken  
bei meinem Doktorvater Prof. Dr. W. Hess für die Betreuung,  
bei Maria Wolters, Thomas Portele und Heinz Rulands für das Korrekturlesen der Arbeit,  
bei Karl-Heinz Stöber für die Computerprogramme  
und bei allen Versuchspersonen für ihre Geduld.



## INHALTSVERZEICHNIS

1	Einleitung .....	1
2	Vokalvariationen .....	5
2.1	Symbolische Beschreibungen von Vokalen.....	5
2.2	Akustische Beschreibung von Vokalen .....	7
2.3	Vokalvariationen und Reduktionen .....	9
2.3.1	Koartikulation und Reduktion.....	14
2.3.2	Sprechökonomie.....	17
2.3.3	Zusammenfassung.....	20
3	Perzeption von Vokalvariationen .....	23
3.1	Akustische Korrelate der Vokalperzeption.....	23
3.2	Perzeptionsmodelle .....	25
3.3	Einfluss von Versuchsbedingungen auf die Vokalperzeption .....	28
3.4	Zusammenfassung.....	33
4	Vokalreduktion im Deutschen.....	35
4.1	Das Vokalsystem des Deutschen .....	35
4.2	Beschreibungen von Vokalreduktion.....	37
4.3	Dauer und Vokalqualität .....	39
4.4	Reduktion als segmentales oder suprasegmentales Phänomen? .....	40
4.4.1	Reduktion als segmentales Phänomen .....	41
4.4.2	Reduktion als suprasegmentales Phänomen .....	46
4.4.3	Zusammenfassende Diskussion .....	50
5	Fragestellung und Datenmaterial.....	53
5.1	Methodik .....	54
5.2	Datenmaterial .....	56
5.2.1	Beschreibung des Datenmaterials .....	56
5.2.2	Normierung der Daten .....	60
5.3	Verwendete statistische Verfahren.....	64
6	Perzeptionsexperimente.....	67
6.1	Reduktionsstufen gespannter Vokale.....	67
6.1.1	Ermittlung potentieller Prototypen.....	67
6.1.2	Experimentdesign.....	69
6.1.3	Vergleichsexperimente.....	71
6.1.4	Beurteilung der Prototypen .....	74
6.1.5	Ergebnisse .....	77
6.1.6	Sprecherunabhängige Prototypen .....	81
6.2	Reduktionsstufen ungespannter Vokale.....	83
6.2.1	Ermittlung potentieller Prototypen.....	84
6.2.2	Vergleichsexperimente.....	84
6.2.3	Beurteilung der Prototypen .....	86

6.2.4	Ergebnisse .....	87
6.3	Zusammenfassung und Diskussion .....	87
7	Akustische Analyse der Reduktionsstufen .....	91
7.1	Perzeptive und akustische Reduktionsstufen .....	92
7.1.1	Akustische Realisierung der Reduktionsstufen.....	92
7.1.2	Zusammenfassung.....	98
7.2	Reduktion gespannter und ungespannter Vokale – ein Vergleich .....	100
7.2.1	Datenmaterial .....	101
7.2.2	Vokale mit vergleichbar perzipierter ungespannter Qualität .....	101
7.2.3	Perzipierte Reduktionsstufen und akustische Ähnlichkeit zum [ə] .....	106
7.2.4	Zusammenfassung.....	111
7.3	Akustische und perzeptive Ähnlichkeit .....	112
7.3.1	Untersuchungen .....	113
7.3.2	Zusammenfassung.....	117
8	Reduktion, Betonung und Wortart .....	123
8.1	Untersuchungsmaterial.....	123
8.2	Untersuchung von Wortklasse, Betonung und Reduktion .....	123
8.3	Zusammenfassung.....	130
9	Automatische Klassifikation von Reduktion.....	131
9.1	Datenmaterial .....	131
9.2	Verwendete Klassifikationsverfahren .....	131
9.3	Ergebnisse der Klassifikation.....	134
9.4	Zusammenfassung.....	139
10	Zusammenfassung und Ausblick .....	141
10.1	Zusammenfassung der Ergebnisse .....	142
10.2	Diskussion und Ausblick .....	147
	Literatur.....	151
	Anhang A .....	166
	Anhang B .....	169
	Anhang C .....	171
	Anhang D.....	179
	Anhang E.....	185
	Anhang F.....	187

# 1 EINLEITUNG

Gesprochene Sprache dient als primäres Kommunikationsmittel dem Austausch von Informationen. Wenn wir gesprochene Äußerungen näher betrachten, sei es durch wiederholtes Abhören von gespeichertem Sprachmaterial, durch akustische Analysen oder Aufzeichnungen des Produktionsprozesses, so findet sich eine Vielzahl von unterschiedlichen Realisierungen innerhalb einer Lautkategorie. Trotz der Variabilität beim Produktionsprozess können Hörer die Bedeutung des Gesagten leicht verstehen, sofern sie mit der Sprache vertraut sind.

Die Lautrealisierungen variieren jedoch nicht völlig unsystematisch, sondern Sprecher passen sich durch bewusste Steuerung der Artikulatoren an die Erfordernisse der Kommunikationssituation an (Lindblom, 1990). Auf der einen Seite sind Sprecher bemüht, ihr Kommunikationsziel zu erreichen, sie artikulieren präziser, so dass der perzeptive Kontrast zwischen den Lauten vergrößert und die Perzeption der Laute für Hörer erleichtert wird. Auf der anderen Seite versuchen Sprecher, ihren artikulatorischen Aufwand möglichst gering zu halten. Bei der Sprachproduktion werden Extremstellungen der Artikulatoren vermieden oder ganze Bewegungsabläufe fallen weg. Diese Prozesse führen zur Reduktion von Lauten und damit auch dazu, dass der perzeptive Kontrast zwischen den Lauten verringert und die Perzeption der Äußerung für Hörer erschwert wird. Reduzierte Vokale sind beispielsweise bei erhöhter Sprechgeschwindigkeit und in Silben oder Wörtern, die eine geringere Relevanz für den Informationsaustausch besitzen und unbetont geäußert werden, zu finden (vgl. z.B. Lindblom, 1963; van Bergem, 1993; Kohler, 1995; Faust, 1997; Jurafsky et al., 1998).

Ausgangspunkt der Beschreibung reduzierter Vokale ist die kanonische Form der jeweiligen Realisierung, z.B. Vokale, die in Isolation gesprochen werden oder in einem neutralen Kontext. Im akustischen Vokalraum (z.B. der Formantkarte) manifestieren sich Vokalreduktionen durch eine Abweichung von der Position der idealen Realisierung hin zu einer zentraleren Position (Tiffany, 1959; Lindblom, 1963; Koopmans-van Beinum, 1976; Nord, 1987). Reduzierte Vokale sind durch einen Verlust der Vokalqualität gekennzeichnet. Auf der symbolischen Ebene werden Reduktionen der Vokalqualität durch Stufen beschrieben. Die Basis bildet dabei die kanonische Form über den neutralen Vokal ‚schwa‘ bis zum kompletten Wegfall des Vokals. Bei gespannten Vokalen wird hierbei noch eine ungespannte Vokalqualität eingefügt, z.B. [i:]→[i]→[ɪ]→[ə]→∅.

Betrachten wir Beschreibungen von Vokalreduktionen genauer, so zeigt sich, dass noch wenig Einigkeit darüber besteht, wie viele Reduktionsstufen für die einzelnen Vokalkategorien bestehen (vgl. z.B. Meinhold, 1973; Schindler, 1974; Dedenbach, 1987).



Abgesehen von der Anzahl der diskriminierbaren Reduktionsstufen stellt sich die generelle Frage, ob Vokalreduktionen überhaupt in Reduktionsstufen eingeteilt werden können. Zum einen zeigen Untersuchungen, dass erhebliche intersubjektive Abweichungen bei der Beurteilung reduzierter Vokale auftreten (van Bergem, 1995b; Aylett und Turk, 1998). Zum anderen ist aufgrund einer kontinuierlichen Wahrnehmung von Vokalen davon auszugehen, dass Hörer mehr Vokalqualitäten unterscheiden als durch eine Transkription symbolisiert wird. Einer Beschreibung der Reduktion von Vokalqualitäten auf einem Kontinuum entsprechen qualitätsbeschreibende Ausdrücke über den Grad der Reduktion (von nicht reduziert bis stark reduziert).

Ziel dieser Arbeit ist es, ein perzeptives Beschreibungssystem aufzustellen, das Reduktionen von Vokalqualitäten möglichst fein differenziert und in Stufen (von nicht reduziert bis stark reduziert) gruppiert. Es wird untersucht, ob Hörer reduzierte Vokalqualitäten in Stufen einteilen können und wenn ja, wie viele Reduktionsstufen innerhalb der Vokalqualitäten unterschieden werden.

Die Arbeit ist wie folgt strukturiert:

Zunächst (Kapitel 2) wird dargestellt, wie Vokalqualitäten auf der symbolischen Ebene beschrieben werden und welchen Bezug die Symbole zur Produktion und zur Perzeption aufweisen. Weiterhin wird die Beziehung zwischen der Produktion von Vokalen und ihren akustischen Eigenschaften erläutert. Nach dieser grundlegenden Darstellung wird auf die Ursachen von Vokalvariationen eingegangen. Ein Teil dieser Variationen wird als Vokalreduktion bezeichnet. Jedoch wird in der Literatur der Begriff Reduktion unterschiedlich definiert, und auch die primären Faktoren, die zu einer Reduktion führen, werden kontrovers diskutiert. Einigkeit besteht aber darüber, dass eine Reduktion von Vokalen auf eine Verringerung des artikulatorischen Aufwands zurückgeführt werden kann. Es stellt sich die Frage, wie sich der artikulatorische Aufwand quantifizieren lässt.

Vokalreduktionen sollen in dieser Arbeit möglichst differenziert beschrieben werden. Da die Wahrnehmung von Vokalqualitäten in erster Linie von den Formanten abhängt, wird in Kapitel 3 insbesondere auf die Zusammenhänge zwischen den spektralen Eigenschaften von Vokalen und ihrer perzipierten Qualität eingegangen. Es wird der Frage nachgegangen, wie Hörer trotz der großen akustischen Variabilität Vokale identifizieren. Dargestellt werden der Zusammenhang zwischen Diskrimination und Identifikation von Vokalen sowie Faktoren, die die Beurteilung von Vokalen beeinflussen.

In Kapitel 4 werden das Vokalsystem des Deutschen und Beschreibungen von Vokalreduktionen vorgestellt. Nach dem klassischen Modell des *target-undershoot* (Lindblom, 1963) werden Vokalreduktionen durch eine Lautdauerverkürzung hervorgerufen. Den verschiedenen Beschreibungen von Vokalreduktion ist gemeinsam, dass eine der Reduktionsstufen der gespannten Langvokale mit einer ungespannten Kurzvokalqualität wahrgenommen wird, eine

andere mit einer schwa-Qualität. Keine Einigkeit besteht darüber, ob Reduktionsstufen perzipiert werden, die allein durch einen Quantiätsverlust gekennzeichnet sind (z.B. [i:]→[i]). Es werden Untersuchungen vorgestellt, die sich mit dem Einfluss einer Dauerreduktion auf die akustischen Realisierungen und perzipierten Qualitäten von Vokalen beschäftigen. Ein zentraler Punkt von Beschreibungen des Reduktionsprozesses von Vokalen ist der Verlust der Gespanntheit. Der Begriff Gespanntheit wird mit der Muskelspannung und somit auch mit dem artikulatorischen Aufwand assoziiert. Es stellt sich die Frage nach den artikulatorischen Korrelaten der Gespanntheit. Insbesondere interessiert, inwiefern der Verlust der Gespanntheit als Minimierung des artikulatorischen Aufwands betrachtet werden kann. Da die Gespanntheit von Vokalen und ihre Reduktion mit prosodischen Faktoren, wie der Betonung, interagieren, wird erörtert, ob Vokalreduktion dem segmentalen oder suprasegmentalen Bereich zuzuordnen ist.

Kapitel 5 führt in die Experimente ein, die im Rahmen dieser Arbeit zur Perzeption von Vokalreduktion durchgeführt wurden. Die Methode zur Untersuchung der Perzeption von Vokalreduktion wird vorgestellt. Weiterhin werden das Datenmaterial, dessen Aufbereitung sowie die verwendeten statistischen Verfahren beschrieben.

Um eine möglichst differenzierte perzeptive Beschreibung von Vokalreduktionen zu erhalten, werden Vergleichsexperimente durchgeführt. Hierbei werden Vokalrealisierungen aufgrund ihrer qualitativen Ähnlichkeit potentiellen Prototypen von Reduktionsstufen zugeordnet. Diese Experimente werden in Kapitel 6 beschrieben. Insbesondere werden die Ermittlung der Prototypen, Versuchsaufbau und Ergebnisse, sowie ihre Evaluation erläutert.

Weiterhin werden die in Reduktionsstufen eingeteilten Vokalrealisierungen akustischen Analysen unterzogen (Kapitel 7). Es wird untersucht, ob sich die einzelnen Reduktionsstufen akustisch voneinander unterscheiden, inwiefern sich die akustischen Realisierungen von reduzierten Vokalen sowie von nicht reduzierten Vokalen mit vergleichbar wahrgenommener Qualität entsprechen und ob es Zusammenhänge zwischen akustischer Reduktion und perzipierter Reduktionsstärke gibt. Es wird der Frage nachgegangen, inwiefern die akustische Nähe einer Vokalkategorie zum schwa die Diskriminierung von Reduktionsstufen beeinflusst. Schließlich werden akustische Abstandsmaße zur Beschreibung der perzipierten Ähnlichkeit von Prototyp und zugeordneten Vokalrealisierungen diskutiert.

Wie bereits oben erwähnt, werden Wörter, die einen geringen Informationsgehalt besitzen, häufig unbetont und mit reduzierter Vokalqualität realisiert. Ein Beschreibungssystem von Vokalreduktion sollte diese Interaktion widerspiegeln. In Kapitel 8 werden daher die Beziehungen zwischen Wortklasse (Funktions- vs. Inhaltswort), Wortbetonung (lexikalische Wortbetonung, Akzentuierung und perzeptive Silbenprominenz) und den perzipierten Vokalreduktionsstufen untersucht.

Weiterhin wird geprüft, ob Vokalrealisierungen auch automatisch in Reduktionsstufen eingeteilt werden können (Kapitel 9). Drei Klassifikationsverfahren – lineare Diskriminanzanalyse, *Support Vector Machines* und künstliche neuronale Netze – werden hierzu verwendet.

Abschließend (Kapitel 10) werden die Ergebnisse der Arbeit zusammengefasst. Fragen, die sich ergeben, werden diskutiert und Ansatzpunkte für weitergehende Untersuchungen aufgezeigt.

## 2 VOKALVARIATIONEN

In gesprochener Sprache gibt es eine Vielzahl von Vokalrealisierungen. Vokale in gleichen Äußerungen werden von verschiedenen Personen unterschiedlich realisiert. Aber auch Wiederholungen von Äußerungen eines Sprechers weisen Vokalvariationen auf.

Im folgenden wird aufgeführt, wie wahrgenommene Vokalrealisierungen auf der symbolischen und akustischen Ebene beschrieben werden. Ziel von Beschreibungssystemen von Vokalen ist es, zwei Hauptkomponenten des Kommunikationsprozesses, die Produktion und die Perzeption, zu erfassen. In Abschnitt 2.1 wird ein in der Phonetik anerkanntes Beschreibungssystem vorgestellt, das auf den Konventionen der *International Phonetic Association* (IPA) basiert. Es werden die Zusammenhänge zwischen Artikulation und Perzeption erläutert sowie in Abschnitt 2.2 die Beziehungen zwischen Artikulation und Akustik beschrieben.

In Abschnitt 2.3 wird dann auf die Ursachen von Vokalvariationen näher eingegangen. Unterschiede in der Vokalproduktion werden im allgemeinen auf physiologische und kontextuelle Faktoren zurückgeführt. Ein großer Teil von Vokalvariationen wird unter dem Begriff Reduktion zusammengefasst. Eine Durchsicht der Literatur zeigt, dass der Begriff Reduktion von einzelnen Autoren unterschiedlich gebraucht wird und auch die Ursachen für Reduktionen kontrovers diskutiert werden.

### 2.1 Symbolische Beschreibungen von Vokalen

Schon seit Jahrhunderten sind Phonetiker bemüht, ein adäquates Beschreibungssystem für Vokale zu finden (für einen ausführlichen Überblick s. Ungeheuer, 1962), das die Vokalproduktion und –perzeption miteinander verbindet (Ramers, 1988). Bindeglied zwischen Artikulation und Perzeption ist die Akustik. In dem in der Phonetik gebräuchlichen Beschreibungssystem, das auf den Konventionen der *International Phonetic Association* (IPA) basiert, werden Vokale in einem Trapez dargestellt und mit quasiartikulatorischen Merkmalen beschrieben (Abbildung 1). Unterschieden werden drei Parameter:

- artikulierender Zungenteil (Vorder-, Mittel- oder Hinterzunge)
- Zungenhöhe (hoch, mittel-hoch, mittel-tief, tief)
- Lippenstellung (ungerundet, gerundet)

Die Darstellung der Vokale beruht auf dem System der Kardinalvokale (*cardinal vowels*) nach Jones (1918). Jones unterscheidet acht Kardinalvokale ([i], [e], [ɛ], [a], [ɑ], [ɔ], [o], [u]), die sich jeweils an der äußeren Seite des Vokaltrapezes befinden. Die Vokale [i] und [ɑ] wurden aufgrund

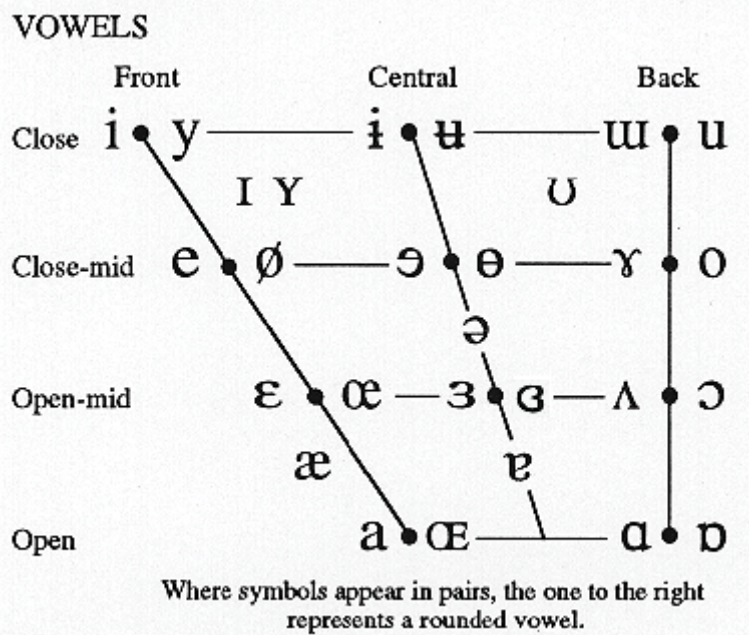


Abbildung 1: Vokaltrapez nach IPA (1993).

ihrer artikulatorischen Extremstellung als Referenzrahmen festgelegt. [i] weist als Vorderzungenvokal den geringsten und [ɑ] als Hinterzungenvokal den größten Abstand zwischen Zunge und Gaumen auf. Die anderen Kardinalvokale wurden aufgrund ihrer perzeptiven Qualität im Vokaltrapez so angeordnet, dass sie den gleichen auditiven Abstand untereinander aufweisen. Zu den primären Kardinalvokalen wurden noch sekundäre hinzugefügt ([y], [ø], [œ], [ɛ], [ɯ], [ɤ], [ʌ], [ɒ]), die sich durch die Lippenstellung – Rundung bei den Vorderzungenvokalen und dem Hinterzungenvokal [ɒ] bzw. Entrundung bei den übrigen Hinterzungenvokalen – unterscheiden.

Die Kardinalvokale sind sprachenunabhängig. Die Einteilung der primären Kardinalvokale beruht in erster Linie auf der Perzeption. Die Beziehungen zwischen Artikulation und Perzeption wurden hypothetisch festgesetzt, wobei die auditiven Abstände zwischen den Vokalen durch vergleichbare artikulatorische Abstände verursacht werden. Im Zentrum des Vokaltrapezes steht der Neutralvokal [ə], der auch als Zentralvokal bezeichnet wird. Die Mittelzungenvokale sowie [æ], [ɪ], [ʏ] und [ʊ] werden gemäß den Konventionen der IPA transkribiert, wobei letztere drei als *zentralisiert* bezeichnet werden. Dem Begriff *zentralisiert* steht der Beschreibungsparameter *peripher* gegenüber. Durch die Gegenüberstellung von *peripher* vs. *zentralisiert* wird die Abweichung der Zungenlage von der Position des Zentralvokals beschrieben. Weitere gebräuchliche Gegensatzpaare sind *geschlossen* vs. *offen* bzw. *gespannt* vs. *ungespannt*. *Geschlossen* vs. *offen* beschreibt genau genommen Abweichungen in der Zungenhöhe, vielfach werden sie aber auch mit dem Kieferöffnungsgrad assoziiert, da im Normalfall die Zungenhöhe und der Öffnungsgrad des Kiefers positiv miteinander korrelieren. Das Begriffspaar *gespannt* vs. *ungespannt* verweist auf eine unterschiedliche Muskelspannung der Artikulationsorgane (Jakobson, Fant und Halle, 1951). Der Begriff der Gespanntheit wird in Abschnitt 4.4 noch diskutiert.

Zur genaueren Beschreibung der Vokale stellt die IPA Diakritika bereit. Die Länge von Vokalen wird mit [ː] (lang), [ˑ] (halblang) oder [̘] (extra kurz) gekennzeichnet. Nuancierungen in der Vokalqualität werden durch artikulatorische Abweichungen beschrieben, die ober- bzw. unterhalb des jeweiligen Lautsymbols gesetzt werden, wie z.B. [ë] (zentralisiert), [ě] (mittel-zentralisiert), [e̝] (erhöht), [e̞] (erniedrigt), [ɔ̟] (stärker entrundet), [ɔ̠] (stärker gerundet), ([e̟]) (vorgezogene Zungenwurzel), [e̠] (zurückgezogene Zungenwurzel), [u̟] (vorverlagert), [ɪ̟] (rückverlagert), [ẽ] (nasaliert).

Das Beschreibungssystem weist jedoch zwischen Produktion und Perzeption der vorderen und hinteren Vokale eine Diskontinuität auf (Catford, 1977; Kohler, 1995b). Während die Perzeption der vorderen Vokale [i], [e] und [ɛ] auf einer linearen Abwärtsbewegung der Zunge beruht, ist für die Wahrnehmung der hinteren Vokale [u], [o] und [ɑ] die Position der Engebildung zwischen Zungenrücken und Gaumen bzw. Pharynxwand (insbesondere für [ɑ]) von Bedeutung. Da der Pharynx keine von den im Vokaltrapez verwendeten artikulatorischen Beschreibungsparametern unabhängige Dimension bildet, werden die hinteren Vokale nicht so einheitlich klassifiziert wie die vorderen. In ähnlicher Weise verhält es sich mit den aus den primären Kardinalvokalen abgeleiteten gerundeten Vokalqualitäten. Auch hier zeigt sich, dass Phonetiker in ihren Beurteilungen von gerundeten im Vergleich zu ungerundeten Vokalen stärker voneinander abweichen (Lisker, 1989; Kohler, 1995b). Hinzu kommt, dass durch verschiedene Artikulationsgesten gleiche Vokalqualitäten erzeugt werden können, was auch als artikulatorische Kompensation bezeichnet wird.

## 2.2 Akustische Beschreibung von Vokalen

Vokalqualitäten werden durch ihre spektralen Eigenschaften akustisch beschrieben. Von Bedeutung sind hierbei die Formanten, die als idealisierte Resonanzen des Vokaltrakts betrachtet werden. Formanten sind durch Spektralanalyse feststellbare Energieanreicherungen. Sie sind definiert durch ihre Formantfrequenz, Formantamplitude und Formantbandbreite (Heike, 1982; zur Bestimmung der Formantfrequenz s. Hess, 1972; für eine Diskussion des Begriffs des Formanten s. Maurer, 1994). Akustische Untersuchungen von Vokalen beschränken sich meist auf die Frequenzen der ersten zwei bis drei Formanten (F1, F2, F3), da die höheren Formanten als sprecherabhängig betrachtet werden und somit keinen Beitrag zur Vokalidentifikation leisten.

Die Bildung der Formanten wird mit Hilfe des Quelle-Filter-Modells (Fant, 1970) beschrieben. Formanten können auch als Resonanzen verstanden werden, die durch Umformung eines Quellen- oder Primärsignals im Vokaltrakt entstehen. Das Primärsignal wird von der schwingenden Glottis erzeugt. Die Frequenz der Schwingungen der Glottis ist identisch mit der Grundfrequenz (F0) des Primärsignals. Die zyklischen Bewegungsvorgänge der Glottis werden in der myoelastisch-aerodynamischen Theorie (van den Berg, 1958) erklärt.



Die Formantfrequenzen von Vokalen werden aus den Resonanzfrequenzen eines Vokaltrakts (Ansatzrohr) mit konstanten Querschnittsverlauf abgeleitet, wobei Ungeheuer „den Ursprung im Resonanzmechanismus des Ansatzrohres als Ganzes, genauer formuliert, im Resonanzmechanismus der ganzen vom Ansatzrohr umschlossenen Luftsäule“ (1962:80) sieht. Der konstante Querschnittsverlauf ist mit einer Konfiguration des Vokaltrakts bei der Produktion des Neutralvokals /ə/ vergleichbar. Der Neutralvokal wird als Ursprung aller Vokale gesehen, jedoch nicht als ein organischer Prozess, sondern als ein theoretisches Axiom (Ungeheuer, 1958:36). Der Neutralvokal kennzeichnet das Vokalsystem eines Sprechers und kann als Artikulationsbasis betrachtet werden.

Obwohl die Begriffe Resonanz- und Formantfrequenzen oft synonym verwendet werden, müssen die Resonanzfrequenzen des Vokaltrakts nicht immer mit den Formantfrequenzen übereinstimmen. Eine Nasalierung von Vokalen bewirkt eine zusätzliche niederfrequente Resonanz im Bereich der Grundfrequenz, häufig eine Auslöschung von F1 und eine Dämpfung von F2 und F3 (Fant, 1970; Hess, 1972). Durch artikulatorische Kompensation können verschiedene Artikulationsgesten zu dem gleichen akustischen Ergebnis führen. Die Beziehung zwischen Artikulation und Akustik ist somit nicht uneindeutig. Durch die Artikulationsstellung kann zwar auf akustische Parameter geschlossen werden, aber von der Akustik nicht auf die Artikulationsstellung. Ladefoged (1980) schränkt aber ein, dass die artikulatorische Kompensation als Ausnahmefall betrachtet werden kann, ein Rückschluss von der Akustik auf die Artikulation ist in der Regel zulässig.

Die Abhängigkeit zwischen Artikulation und Akustik zeigt sich, wenn auch stark vereinfacht, bei einer Gegenüberstellung von Vokaltrapez und der Verteilung der Vokale im spektralen Raum. Der Vergleich des Vokaltrapezes mit gegeneinander aufgetragenen F1- und F2-Werten (Formantkarte) zeigt, dass eine Änderung der Zungenhöhe eine Verschiebung von F1 bewirkt (geschlossene Vokale besitzen im Vergleich zu den geöffneten Vokalen einen niedrigeren F1-Wert). Der artikulierende Zungenteil verursacht eine Änderung in der Lage des zweiten Formanten (Vorderzungenvokale besitzen einen höheren F2-Wert als Hinterzungenvokale). Eine Änderung der Lippenstellung wirkt sich diagonal zur F1- und F2-Achse aus.

Vokale besitzen inhärente akustische Eigenschaften, deren Ursprung in den Artikulationsbewegungen gesehen werden. Hierzu gehören die intrinsische Grundfrequenz und die Dauer. Geschlossene Vokale besitzen eine höhere Grundfrequenz als offene Vokale. Die Ursache für dieses Phänomen ist noch nicht eindeutig nachgewiesen (Whalen et al., 1998). Ohala (1973) vermutet, dass die Zugkraft der Zungenmuskeln bei höherliegender Zungenstellung auf die vertikale Spannung der Stimmlippen wirkt (*tongue pull theory*). Manchmal wird die intrinsische Grundfrequenz auch auf eine höhere Luftströmung bei geschlossenen als bei offenen Vokalen zurückgeführt (Neppert und Péturson, 1986). Mit einer erhöhten Luftströmung ist eine stärkere Strömung in der Glottis verbunden, so dass geschlossene Laute höhere F0-Werte aufweisen.

Offene Vokale besitzen im Vergleich zu geschlossenen eine längere intrinsische Dauer. Dies wird mit dem Weg begründet, den die Artikulationsorgane (z.B. Kieferöffnungsgrad) zurücklegen müssen, um die Zielstellung zu erreichen.

Die Beziehung zwischen Artikulation und Akustik zeigt sich deutlich bei einem Vergleich von Frauen-, Männer- und Kinderstimmen. Aufgrund der Beschaffenheit des Vokaltraktes (insbesondere der Länge) und der Artikulationsorgane sind in der Regel die Grundfrequenz- und Formantwerte von Männern niedriger als die von Frauen, und letztere wiederum sind geringer als die von Kindern.

### **2.3 Vokalvariationen und Reduktionen**

Im folgenden wird beschrieben, welche potentiellen Faktoren die Realisierung von Vokalen beeinflussen. Hierzu gehören:

- Phonetischer Kontext,
- Sprecher (z.B. physiologische, dialektale, geschlechtsspezifische Unterschiede),
- Sprechgeschwindigkeit,
- Betonung,
- lexikalische Faktoren (z.B. Funktions- vs. Inhaltswort, Worthäufigkeiten),
- Sprechstil (z.B. Spontan- vs. Lesesprache).

Die einzelnen Faktoren sind stark miteinander verzahnt, so dass es sich als äußerst schwierig erweist, sie zu isolieren und nach einem Ursache-Wirkungsprinzip zu beschreiben. Im folgenden werden einige Befunde exemplarisch vorgestellt.

#### Phonetischer Kontext

Zu den klassischen Studien von Vokalvariationen aufgrund des lautlichen Kontextes gehört die Forschungsarbeit von Öhman (1966). Er untersuchte die Formantübergänge von VCV-Folgen (wobei V für Vokal steht und C für Konsonant). Für den Konsonanten wurden die stimmhaften Plosive /b/, /d/ oder /g/ eingesetzt, für die Vokale /y/, /ø/, /ɑ/ oder /u/. Jeder Laut wurde mit jedem kombiniert. Die VCV-Stimuli wurden von Sprechern des Schwedischen, des amerikanischen Englisch und des Russischen gelesen. Die Ergebnisse zeigen, dass die Muster der Formanttransitionen (steigend, fallend, gleich bleibend) vom initialen Vokal zum Konsonanten von der finalen Vokalqualität abhängen. Die Formantfrequenzen, gemessen im stationären Teil eines Vokals, weisen nur geringe Unterschiede in Abhängigkeit ihrer Position und der Konsonantenqualität auf. Die Ergebnisse von Öhman stehen im Widerspruch zur Locus-Theorie von Delattre et al. (1955), wonach die Formanttransitionen in Abhängigkeit von der Konsonantenqualität an einer charakteristischen Frequenzposition (Locus) beginnen bzw. enden. Nach Delattre et al. verursacht der unmittelbare konsonantische Kontext ein spezifisches Muster



der Formanttransition, während nach Öhman das Muster durch den folgenden vokalischen Kontext geprägt ist.

### Sprecherspezifische Vokalvariationen

Um den Einfluss des lautlichen Kontextes zu beschränken, werden bei Untersuchungen sprecherspezifischer Unterschiede entweder isolierte Vokale genommen oder Vokale, die in einem sogenannten neutralen Kontext oder Nullkontext (Stevens und House, 1963), wie zum Beispiel h(Vokal)t, gesprochen wurden.

Peterson und Barney (1952) gehören zu den Ersten, die in einer breitangelegten Studie akustische Sprecherunterschiede untersuchten. Sie maßen die Formantenfrequenzen von 10 Vokalen des amerikanischen Englisch in einem h(Vokal)d Kontext. Eine Reihe von Untersuchungen folgten für verschiedene Sprachen, z.B. für das Schwedische (Fant et al., 1969), Niederländische (Pols et al., 1973) und für das Deutsche (Jørgensen, 1969; Hess, 1972). Jørgensen (1969) hebt hervor, dass die absoluten Formantwerte (F1 und F2) der Vokale verschiedener Sprecher variieren, jedoch sind die relativen Formantwerte der Vokale konstant. Auch Heid (1998) schließt aufgrund des Zusammenhangs von Vokalqualität und Dauer, der sich deutlicher an den Formantverhältnissen als an den absoluten Formantwerten zeigt, auf die von Sprechern intendierte Vokalqualität.

Die intersubjektiven Unterschiede können nicht nur auf einen Längenunterschied des Vokaltrakts zurückgeführt werden. Obwohl der Vokaltrakt von Frauen meist kürzer ist als der von Männern und dadurch die Vokale von Sprecherinnen höhere Frequenzen besitzen, bemerkt beispielsweise Fant (1966, 1975), dass eine Berücksichtigung der Längenabweichung geschlechtsspezifische Unterschiede nicht vollständig erklärt.

Frauen besitzen im Durchschnitt einen größeren spektralen Vokalraum als Männer, d.h. die Formanten der einzelnen Vokalkategorien sind weiter voneinander entfernt (Fant, 1975; Diehl et al., 1996). Akustische Analysen der Vokale von Sprechern des Deutschen zeigen, dass sich die Vokalsysteme hinsichtlich ihrer spektralen Ausdehnung unterscheiden, die Unterschiede selbst aber geschlechtsspezifisch sind. Vokalsysteme von Männern variieren nur bezüglich der Position von F1, während sich Vokalsysteme von Frauen durch die ersten beiden Formanten unterscheiden (Pätzold und Simpson, 1997).

Weitere Unterschiede in den Vokalrealisierungen finden sich aufgrund der dialektalen Herkunft von Sprechern. Heid (1997) untersuchte 10000 Vokalrealisierungen von 16 Sprechern aus Nord-, Mittel- und Süddeutschland. Typische Unterschiede zwischen den Sprechern finden sich bei der Realisierung von <ä>, z.B. in dem Wort <Bären>, als [ɛ:] bzw. [e:]. Gemäß der Normsprache unterscheiden sich die beiden Wörter <Bären> und <Beeren> durch die in phonologischer Distinktion stehenden Vokale /ɛ:/ und /e:/. Viele Sprecher des Norddeutschen weisen diesen Unterschied nicht auf; die Formantwerte entsprechen denen eines /e:/. Weiterhin zeigen sich in den

F1-Werten von gespannten und ungespannten Vokalen einiger Sprecher keine Abweichungen. Sprecherunterschiede beruhen nicht nur auf physiologischen Unterschieden, sondern in unterschiedlichen Anordnungen der Vokalsysteme.

### Sprechgeschwindigkeit

In Bezug auf die Sprechgeschwindigkeit wird zwischen globaler, lokaler und relativer Sprechgeschwindigkeit unterschieden (Pfitzinger, 1996). Die globale Sprechgeschwindigkeit wird durch das Verhältnis der Anzahl der Einheiten (z.B. Phone, Silben) zur Dauer der Äußerung bestimmt. Die lokale Sprechgeschwindigkeit entspricht der Dauer einer Einheit; bei der relativen Sprechgeschwindigkeit werden zwei Geschwindigkeiten z.B. verschiedener Äußerungen eines Sprechers zueinander in Beziehung gesetzt.

Die lokale Sprechgeschwindigkeit wird von zahlreichen Faktoren beeinflusst, z.B. intrinsische Dauerunterschiede, phonetischer Kontext, Betonung, Phrasenposition usw. (vgl. Klatt, 1976; Umeda, 1975; Crystal und House, 1988; Rietveld, 1975; van Santen, 1992).

Nach van den Heuvel et al. (1994) besitzen sowohl die intrinsischen Dauerunterschiede als auch der konsonantische Kontext einen größeren Einfluss auf die Vokaldauer als die sprecherspezifischen Faktoren. Phonetisch lange Vokale weisen im Vergleich zu kurzen Vokalen eine höhere Sprecherabhängigkeit auf. Die sprecherabhängigen Dauervariationen werden primär auf die Sprechgeschwindigkeit zurückgeführt.

Lindblom (1963) überprüfte die akustische Vokalrealisierung in Abhängigkeit von Betonung, Sprechgeschwindigkeit und lautlichem Kontext. Als Datenmaterial dienten Vokale des Schwedischen in verschiedenen konsonantischen Kontexten. Die CVC-Silben wurden in Trägersätzen mit unterschiedlichen Phrasenakzenten und Wortanordnungen eingebettet. Ein Sprecher wurde gebeten, die Sätze mit unterschiedlichen Rhythmen zu lesen, die ihm über Kopfhörer vorgegeben wurden. Die Ergebnisse zeigen, dass unbetonte Vokale und solche, die mit einer schnelleren Sprechgeschwindigkeit produziert wurden, im Vergleich zu betonten bzw. mit langsamer Sprechgeschwindigkeit produzierten Vokalen hinsichtlich der Formantwerte stärker zentralisiert und in ihrer Lautdauer stärker reduziert sind. Auf der Grundlage dieses Experiments entwickelte Lindblom seine Theorie von der zeitabhängigen Vokalreduktion (*duration-dependent target-undershoot*). Hiernach gibt es für jeden Vokal eine artikulatorische Zielstellung (*target*). Das Ziel ist unabhängig vom konsonantischen Kontext und der Dauer „[...] and can thus be looked upon as an invariant attribute of the vowel“ (Lindblom, 1963:1778). Die Intention eines Sprechers ist, diese Zielstellung zu erreichen. Jedoch ist die Geschwindigkeit des Bewegungsablaufs der Artikulatoren begrenzt, was als *Trägheit der Artikulationsorgane* beschrieben wird, so dass aufgrund einer Reduktion der Vokaldauer die intendierte Zielstellung nur annähernd oder überhaupt nicht (*articulatory undershoot*) erreicht wird und sich in der

Zentralisierung der Formantwerte (*acoustic undershoot*) manifestiert. Nach Lindblom ist der wichtigste Faktor für Vokalreduktion die Sprechgeschwindigkeit.

Generell gilt, dass mit zunehmender Sprechgeschwindigkeit häufiger schwache Formen (nicht betonte Aussprache von Wörtern) gebraucht werden und sich um so häufiger Reduktionen finden lassen (Kohler, 1995). Faust (1997; vgl. auch Laan, 1997) weist darauf hin, dass insbesondere Sprechgeschwindigkeit und Reduktionen sprecherspezifisch sind.

### Betonung

Auch die Betonung steht im unmittelbaren Zusammenhang mit verschiedenen Vokalrealisierungen. Unter Betonung wird hier die wahrnehmbare Hervorhebung (Prominenz) von Silben verstanden. Es handelt sich dabei um eine relative Größe, d.h. eine bestimmte Silbe wird stärker hervorgehoben perzipiert als eine andere Silbe (Terken, 1991; Ladd et al., 1994). Die Perception der Betonung wird größtenteils durch die akustischen Parameter Grundfrequenz, Dauer und Intensität hervorgerufen, wobei diese eine unterschiedliche perzeptive Relevanz besitzen. So zeigt beispielsweise Fry (1958) in seiner Untersuchung des Englischen, dass die Änderung der Grundfrequenz den größten Einfluss auf die Perception der Betonung hat, gefolgt von der Dauer und der Intensität. Einschränkend bemerkt er, „[...] *it may be necessary to reiterate that all judgements of stress in natural speech depend on the complicated inter-action of a number of cues*“ (Fry, 1958:151). Auch Heuft (1999) kommt aufgrund von akustischen Analysen von prosodischen Minimalpaaren des Deutschen (z.B. <voll Milch> vs. <Vollmilch>) und Perceptionsexperimenten zu dem Schluss, dass Dauer und Grundfrequenz die wichtigsten Hinweise für die wahrgenommene Betonung geben, jedoch können sie nicht unabhängig voneinander betrachtet werden.

Nach der akustischen Realisierung der Prominenz wird zwischen *Betonung* und *Akzent* unterschieden. *Akzent* (*pitch accent*) wird als Spezialfall der Betonung betrachtet und ist durch ein Maximum im Grundfrequenzverlauf gekennzeichnet (Bolinger, 1958; vgl. auch Möbius, 1993; Heuft, 1999). Die Aufgabe der Betonung auf der Wortebene ist phonologischer Natur, z.B. wird hierdurch die lexikalische Wortbetonung realisiert. Betonte Silben werden genauer artikuliert und erleichtern die Wortwahrnehmung (van Bergem, 1990). Durch die Betonung werden aber nicht nur einzelne Silben mehr oder weniger hervorgehoben, sondern in größeren sprachlichen Einheiten auch Wörter. Diese Hervorhebung wird mit der Funktion der Markierung des Informationsinhaltes des Wortes verknüpft. So schreibt Jones:

„*As a general rule it may be said that the relative stress of the words in a sequence depends on their relative importance. The more important a word is, the stronger is its stress*“ (1918, 1962:262).

Je nachdem auf welcher Ebene ein Wort hervorgehoben wird, wird zwischen einem Phrasen- und einem Satzakzent unterschieden. Eine semantisch motivierte Betrachtung des prominentesten Wortes spiegelt sich im Begriff des Fokus wider. In funktionaler Hinsicht wird der Fokus als 'neue' Information betrachtet (Bannert, 1985; Bolinger, 1972, 1989). Sprecher sind bemüht, diese Wörter genauer zu artikulieren.

Hinsichtlich der Realisierung von Vokalen gilt im allgemeinen, dass betonte Silben im Vergleich zu unbetonten weniger Reduktionen aufweisen (z.B. Kohler, 1995). Zusätzlich zeigt sich, dass unbetonte Silben in Wörtern, die satz- bzw. phrasenakzentuiert sind, weniger reduziert sind als die in nicht akzentuierten Wörtern (van Bergem, 1993). Insgesamt werden Vokale in satz- bzw. phrasenakzentuierten Wörtern weniger reduziert als in nicht akzentuierten Wörtern.

### Lexikalische Faktoren

Neben der Betonung stehen auch Wortklasse und –häufigkeit in einem engen Zusammenhang mit Vokalreduktionen (Jurafsky et al., 1998). Eine mögliche (und häufig verwendete) Einteilung von Wörtern ist die in Funktions- und Inhaltswörter. Die Definition dieser Wortklassen ist umstritten. Eine Möglichkeit besteht darin, Wörter aufgrund ihrer Lexikoneinträge (grammatische Definition) in Funktions- und Inhaltswörter einzuteilen, eine andere darin, Wörter hinsichtlich ihrer phonetischen Realisierung zu unterteilen (Kohler, 1991, 1995; zur Diskussion von Inhalts- und Funktionswort s. Breuer, 1999). Erstere stellt den Informationsgehalt der Wörter in den Vordergrund, so gehören zu der Klasse der Funktionswörter nach Bußmann (1990, S. 260) *„sprachliche Elemente, die primär grammatische (anstelle von lexikalischen) Bedeutungen tragen und vor allem syntaktische-strukturelle Funktionen erfüllen [...]“* (Bußmann, 1990, S. 260). Letztere hebt die lautliche Struktur und ihre Realisierung hervor. Funktionswörter sind durch eine geringe Anzahl von Silben (ein- oder zweisilbig), der Zusammensetzung der Vokalqualität (meist nur ein Vokal, der keine /ə/-Qualität besitzt) und ihrer Betonung gekennzeichnet. Dies verdeutlicht die Interaktion von Wortklasse, Betonung und Vokalreduktion. Funktionswörter sind meist unbetont und ihre Vokale sind im Vergleich zu denen von Inhaltswörtern stärker reduziert (van Bergem, 1993).

Jedoch ist die Beziehung zwischen Reduktion und Wortklasse geringer ausgeprägt als die zwischen Reduktion und Worthäufigkeit (van Bergem, 1995). Funktionswörter, die eine geringe Auftretenswahrscheinlichkeit besitzen, sind weniger reduziert als Funktionswörter, die häufig verwendet werden (van Bergem, 1995; Bell et al. 1999). Van Son und Pols (1999) heben hervor, dass die Vokalreduktion in betonten Silben mit der Silben- und Worthäufigkeit korreliert, für unbetonte Silben hingegen zeigen sich keine Zusammenhänge. Die Abhängigkeit zwischen Vokalreduktion und Worthäufigkeit einerseits und Betonung andererseits zeigt, dass auch Worthäufigkeit und Betonung zueinander in Beziehung stehen.

## Sprechstil

Unter dem Begriff *Sprechstil* ist im allgemeinen die Produktion von Sprache unter bestimmten Kommunikationsbedingungen zu verstehen. Weit verbreitet ist eine Unterteilung von gelesener und spontaner Sprache, wobei anzumerken ist, dass es innerhalb dieser beiden Gruppen unterschiedliche Ausprägungen gibt (s. z.B. Eskénazi, 1993, Beckman, 1995). Lesesprache hat gegenüber der Spontansprache den Vorteil, dass mögliche Einflussfaktoren besser kontrolliert werden können. Andererseits wird häufig bemängelt, dass Lesesprache eine Laborsprache und daher unnatürlich sei und die Ergebnisse nicht auf wirkliche Kommunikationssituationen übertragen werden können. Der Zusammenhang von Vokalreduktion und Sprechstil lässt sich über verschiedene Sprachen hinweg nachweisen. Generell gilt, dass im Gegensatz zur gelesenen Sprache spontane Sprache durch eine häufigere Reduktion gekennzeichnet ist (Kohler, 1995; Faust, 1997; Laan, 1997). Je informeller die Spontansprache ist, desto häufiger und stärker ausgeprägt sind Lautreduktionen (Kohler, 1995).

### *2.3.1 Koartikulation und Reduktion*

Wie bereits im vorherigen Abschnitt gezeigt wurde, gibt es in der gesprochenen Sprache eine Vielzahl von unterschiedlichen inter- und intraindividuellen Realisierungen eines Lautes. Laute, die in Isolation gesprochen werden oder in einem neutralen Kontext, werden als ideale Laute betrachtet und als kanonische Realisierungen bezeichnet. Verschiedene Faktoren, z.B. phonetischer Kontext, Betonung, Sprechstil, führen zu einer Abweichung von der kanonischen Form. Diese Abweichung wird im allgemeinen als Reduktion und Koartikulation (O'Shaughnessy, 1987; Clark und Yallop, 1995) bezeichnet.

Koartikulation kann als gegenseitige Beeinflussung der Artikulationsorgane in Bezug auf den lautlichen Kontext verstanden werden. Nach der Wirkungsweise der Koartikulation wird zwischen antizipatorischer bzw. persistenter Koartikulation unterschieden. Unter antizipatorischer Koartikulation ist die Vorwegnahme einer Artikulationsgeste zu verstehen, unter persistenter Koartikulation die Ausdehnung einer Artikulationsgeste über folgende Segmente. Koartikulation wirkt über Silben-, Wort- und Satzgrenzen hinweg. Die Durchführung einer Artikulationsgeste erfordert nicht die Beteiligung aller Artikulatoren, die, wenn sie an der Bildung eines Lautes unbeteiligt sind, als freie Artikulatoren bezeichnet werden. Insbesondere freie Artikulatoren nehmen im voraus die Stellung für einen folgenden Laut ein (antizipatorische Koartikulation), z.B. werden die Lippen bereits am Anfang des Wortes <Glück> gerundet, obwohl die Formung der Lippen erst für die Realisierung von /ʏ/ von Bedeutung ist.

Menzerath und de Lacerda (1933), die zu den Ersten gehören, die Koartikulation experimentalphonetisch untersuchten, unterteilen den Artikulationsprozess in Koartikulation und Steuerung. Koartikulation bezeichnet die Vorbereitung einer Artikulationsgeste für einen folgenden Laut während der Produktion eines Lautes. Die Steuerung bezieht sich auf einen

Artikulator, der an zwei aufeinanderfolgenden Lautproduktionen gleichzeitig beteiligt ist und beschreibt die Abweichung von der Zielstellung eines Lautes aufgrund einer unterschiedlichen Zielstellung eines anderen Lautes. Kohler (1974, 1976) betrachtet die Steuerung als Einsparung einer Artikulationsgeste, vor allem der apikalen Geste.

Die Aussagen von Menzerath und de Lacerda (1933) basieren in erster Linie auf Beobachtungen von antizipatorischen Koartikulationsphänomenen. Antizipatorische Koartikulation wird auf Planung zurückgeführt, persistente Koartikulation auf die Trägheit der Artikulatoren (O'Shaughnessy, 1987; Whalen, 1990).

Koartikulation verursacht Assimilation, die nach ihrer Wirkungsrichtung (progressiv, regressiv) und nach den betroffenen phonetischen Parametern (Veränderung der Artikulationsart, des Artikulationsorts und der Stimmbeteiligung) unterteilt wird. Eine völlige Angleichung eines Sprachlautes an einen anderen hinsichtlich des Orts und der Art der Artikulation sowie der Stimmhaftigkeit wird als totale Assimilation bezeichnet, ansonsten spricht man von partieller Assimilation. Hinsichtlich ihrer zeitlichen Ausdehnung wird zwischen Kontakt- und Fernassimilation unterschieden.

Reduktionen sind im allgemeinen durch eine Abschwächung der Lautqualität gekennzeichnet. Vokale werden mit zunehmender Reduktion mit einer /ə/-Qualität wahrgenommen, im Extremfall erfolgt ihre Elision. Wie bereits oben erwähnt (Abschnitt 2.3), betrachtet Lindblom die Lautdauer als primären Einflussfaktor für Vokalreduktion. Folglich führt Lindblom (1983:220) Vokalreduktionen auf „a rate constraint operating dynamically on articulatory movements“ zurück. Da die Bewegungsgeschwindigkeit der Artikulatoren begrenzt ist, hängt eine Verfehlung der vom Sprecher intendierten Zielstellung von der relativen zeitlichen Abfolge der für eine Ausführung von Artikulationsgesten benötigten Steuerkommandos ab:

*„If these signals are far apart in time, the response may become stationary at individual targets. If, on the other hand, instructions occur in close temporal succession, the system may be responding to several signals simultaneously and the result is coarticulation“* (Lindblom, 1963:1778).

Als artikulatorisches Korrelat der Vokalreduktion nimmt Lindblom eine zeitliche Überlappung der Artikulationsgesten bei Verringerung der Vokaldauer an. Browman and Goldstein (1989, 1990) hingegen erklären die Reduktion mit zwei artikulatorischen Prozessen:

- eine Abnahme der zeitlichen und somit einer räumlichen Ausdehnung einer Geste
- sowie eine zeitliche Überlappung der Gesten.

Kohler (1990) geht unter anderem auch von einer Reorganisation der artikulatorischen Gesten aus, so wird z.B. bei der Reduktion von [k<sup>h</sup>ən] nach [kɪ] die Öffnungs-Schließ-Bewegung des



Zungenrückens und der Zungenspitze durch einen langen oralen Verschluss des Zungenrückens mit velarer Verschlusslösung ersetzt.

Kröger (1998) überprüfte die Annahme, ob eine Reorganisation zwingend für die Modellierung von Reduktionen ist. Er kommt zu dem Schluss, dass ein großer Teil von Reduktionsphänomenen (z.B. Reduktion zum /ə/, Elisionen) durch eine Verringerung der Ausdehnung und eine zeitliche Überschneidung der Artikulationsgesten nach Browman und Goldstein (1989, 1990) modelliert werden kann. Dies deutet darauf hin, dass Vokalreduktion nicht allein auf Koartikulation (vgl. Lindblom, 1963) beruht, sondern ein unabhängiger Zentralisierungseffekt angenommen werden kann.

Aus akustischer Sicht ist Vokalreduktion durch eine Zentralisierung der ersten beiden Formanten in Richtung der Formantwerte des /ə/ gekennzeichnet (Lindblom, 1963). Van Bergem (1994) weist jedoch auf eine Abhängigkeit vom Kontext und den Formantwerten des /ə/ im Niederländischen hin. Hiernach wird /ə/ als ein mit dem phonetischen Kontext vollständig assimilierter Laut betrachtet, der keine akustische Zielposition besitzt. Dieser Laut wird als artikulatorischer Übergang zwischen zwei Konsonanten angesehen. Nach van Bergem ist die Zentralisierung der Formanten kein Merkmal der Vokalreduktion:

*„Centralization is not a phenomenon in its own right, but merely a by-product of contextual assimilation. That is, several forms of contextual assimilation lead to a more central position of the formant frequencies of vowels (particularly of F2), but certainly not all“* (van Bergem, 1994b:99).

Er führt die Vokalreduktion zum größten Teil auf eine Assimilation mit dem phonetischen Kontext zurück und betrachtet nur die Änderung der Formantwerte, die nicht durch Assimilation erklärbar ist, als eine akustische Realisierung der Reduktion.

*„In our interpretation acoustic vowel reduction is the partial assimilation of vowels with their phonemic context. Hence vowels tend to become a schwa (which is completely assimilated with its context)“* (van Bergem, 1994:158).

*„However, the formant shifts are often much larger than would be expected on the basis of coarticulation effects alone. Such extra shifts emerge, for instance, when the vowel occurs in an unstressed syllable, or when a ‘spontaneous’ speaking style is used. In such cases we speak of acoustic vowel reduction“* (van Bergem, 1994b:96).

Untersuchungen im Deutschen lassen vermuten, dass die Reduktion von Vokalen nicht durch eine Assimilation mit dem Kontext gekennzeichnet ist (Wrede et al., 2000). Während die Formantwerte zum Anfang und Ende eines Vokals unabhängig von der Dauer recht stabil sind, zeigt sich ein

deutlicher Zentralisierungseffekt im mittleren Teil der Vokale mit abnehmender Dauer. Es kann davon ausgegangen werden, dass nur der stationäre Teil eines Vokals von der Reduktion betroffen ist.

Die vorausgehenden Darstellungen zeigen, dass es praktisch kaum möglich ist, Koartikulation und Reduktion voneinander zu trennen. Nach Heike (1984) ist Koartikulation kein phonetisches Phänomen an sich, sondern ein Konzept, welches segmentale Beschreibung und artikulatorische Dynamik verbindet. In dieser Arbeit wird Reduktion als Oberbegriff aufgefasst, der Abweichungen von der kanonischen Form eines Vokals beschreibt, die auch durch Koartikulation verursacht werden können (vgl. van Heuvel et al. 1996). Vokalreduktion ist durch einen Verlust der Vokalqualität gekennzeichnet (van Bergem, 1993).

### 2.3.2 Sprechökonomie

Als Ursache von Reduktion wird die Minimierung des artikulatorischen Aufwands angegeben. Aber was bedeutet „Minimierung des artikulatorischen Aufwands“ und inwieweit kann dieser quantifiziert werden?

In seinem *duration-dependent undershoot* Modell erklärt Lindblom (1963) Reduktionen mit einer Abnahme der Lautdauer. Je kürzer die Vokaldauer ist, desto weiter sind die Artikulatoren von der Zielstellung des Vokals entfernt (*undershoot*). Als Beleg führt Lindblom (1990) eine Untersuchung von Nelson et al. (1984) an, in der Ausmaß und Geschwindigkeiten von Kieferbewegungen bei zunehmend schneller Aussprache eines Worts und einer Silbe gemessen wurden. Die Ergebnisse zeigen, dass bei Zunahme der Sprechgeschwindigkeit der Öffnungsgrad des Kiefers abnimmt. Die maximale Geschwindigkeit der Kieferbewegung liegt noch unterhalb der bei höherem Sprechtempo zu erwartenden Geschwindigkeit, die nötig ist, um die Artikulationsgeste vollständig und unabhängig von der zur Verfügung stehenden Zeit auszuführen, d.h. aufgrund des abnehmenden Öffnungsgrads musste die Geschwindigkeit der Kieferbewegung nicht ganz so stark erhöht werden wie angenommen. Sprecher reduzieren ihren artikulatorischen Aufwand, indem sie Extremstellungen bei zunehmender Sprechgeschwindigkeit vermeiden.

Die Hypothese, dass Vokalreduktion allein von der Dauer abhängt, ist jedoch umstritten. So zeigen weitere Untersuchungen (z.B. Kuehn und Moll, 1976; Engstrand, 1988; van Son, 1993), dass eine Verringerung der Dauer nicht immer zur Verfehlung der Zielstellung führt. Die Geschwindigkeit der Artikulatoren wird erhöht, so dass trotz Zunahme der Sprechgeschwindigkeit die Zielposition erreicht wird.

Eine Abnahme der Lautdauer führt also nicht zwangsläufig zu einer Reduktion. Sprecher nutzen unterschiedliche Strategien, um ihre Sprechgeschwindigkeit zu erhöhen. Vermeiden sie Extremstellungen der Artikulatoren, so wird dies als Minimierung des artikulatorischen Aufwands gesehen. Kompensieren sie eine schnellere Sprechgeschwindigkeit durch eine Zunahme der



Artikulationsgeschwindigkeit, so benötigen sie mehr Energie, der artikulatorische Aufwand ist erhöht (Lindblom, 1990). Warum von Sprechern unterschiedliche Strategien genutzt werden, erklärt Lindblom (1990) in seiner *Hyperarticulation und Hypoarticulation*-Theorie (H-&H-Theorie). Die Annahme ist, dass Sprachproduktion und Sprachperzeption durch biologische Prozesse gesteuert werden. Es gibt zwei Hauptkomponenten: Zum einem sind Sprecher bemüht, ihren artikulatorischen Aufwand möglichst gering zu halten (Minimierung des artikulatorischen Aufwands), zum anderen möchten sie ihr Kommunikationsziel erreichen, d.h. sie versuchen durch eine korrekte Aussprache, Hörern die Perzeption zu erleichtern (Maximierung des perceptiven Unterschieds). Sprecher passen ihre Artikulation durch bewusste Steuerung den Erfordernissen der Kommunikationssituation an.

Aber wie kann der artikulatorische Aufwand quantifiziert werden? Wie bereits oben die Interpretation der Untersuchungsergebnisse von Nelson et al. (1984) impliziert, wird die maximale Geschwindigkeit der Artikulatorbewegungen als ein Maß zur Bestimmung des artikulatorischen Aufwands verwendet. Dem artikulatorischen Aufwand entspricht die Kraft, die benötigt wird, um einen Artikulator (Masse) auf eine neue Zielstellung zu bewegen, d.h. zu beschleunigen (Nelson, 1983).

Auf Grundlage dieser Definition begründet Kröger (1998:223 ff) seine Berechnungen des artikulatorischen Aufwands. Er unterscheidet neben dem Aufwand, der zur Bewegung eines Artikulators benötigt wird, den Aufwand, der zur Erhaltung einer Geste gebraucht wird. Der Aufwand des Bewegens wird über den Abstand von initialer Artikulatorstellung und Zielstellung, der maximalen Aktivitätsstärke (Eigenperiodendauer der Geste) und der Dauer des Transitionsintervalls der Geste berechnet. Die Eigenperiodendauer entspricht dem Zeitintervall, das eine Geste benötigt, um einen definierten relativen Abstand zwischen initialer Artikulationsstellung und Zielstellung (relative Artikulator-*Target*-Distanz) zu erreichen (Kröger, 1998:133). Das Transitionsintervall der Geste wird aus der relativen Artikulator-*Target*-Distanz abgeleitet. Das Ende des Transitionintervalls beschreibt den Zeitpunkt, zu dem die Artikulatorbewegung zur Ausbildung der Zielstellung führt. (Kröger, 1998:135). Wenn die Werte dieser drei Variablen – Abstand von initialer Artikulatorstellung und Zielstellung, maximale Aktivitätsstärke und Dauer des Transitionsintervalls – groß sind, dann sind die zurückzulegenden Wege der Artikulatoren lang und somit ist auch der artikulatorische Aufwand hoch. Kröger nimmt an, dass einzelne Artikulationsgesten unterschiedliche maximale Energien benötigen, da die zurückgelegten Wege und die Masse der zu bewegenden Artikulatoren nicht für alle gleich sind. Beispielsweise ist der artikulatorische Aufwand für supraglottale Gesten wesentlich höher als für glottale Gesten.

Der artikulatorische Aufwand des Haltens wird durch die Dauer bestimmt, in der ein Artikulator sich im Bereich der Zielstellung befindet. Er beruht auf der Annahme, dass eine Realisierung der Neutralposition eines Artikulators keine Energie benötigt, während ein Halten von Positionen, die

von der Ausgangsstellung abweichen, energetisch aufwendiger ist. Der Aufwand ist proportional zur zeitlichen Ausdehnung der Haltephase. Wie beim Bewegungsaufwand geht Kröger auch beim Halten davon aus, dass sich die einzelnen Artikulationsgesten hinsichtlich ihres maximalen Energiebedarfs unterscheiden. Zusätzlich ist der Aufwand des Haltens vom Konstriktionsort abhängig. Es wird angenommen, dass das Halten einer apikalen Geste mehr Präzision und somit mehr Aufwand erfordert als das Halten einer labialen oder dorsalen Geste.

Wie die obigen Beschreibungen des artikulatorischen Aufwands zeigen, beruhen Berechnungen der für die Ausführung von Artikulationsgesten genutzten Energien im allgemeinen auf bio-mechanischen Modellen, die die geleistete physikalische Arbeit (z.B. Anzahl der durchgeführten Artikulationsgesten, Realisierungen von Extremstellungen) beschreiben. Umso mehr Artikulationsgesten eingespart und damit Extremstellungen der Artikulatoren vermieden werden, desto geringer ist der artikulatorische Aufwand. Lindblom und Davis (1998) kritisieren, dass Berechnungen des Energieverbrauchs aufgrund von bio-mechanischen Modellen unzureichend sind. Die mathematische Modellierung ist wesentlich komplexer. Zur genaueren Bestimmung des artikulatorischen Aufwands müssten die Energiekosten aller beteiligten Muskeln erfasst werden. Sie schlagen stattdessen vor, den artikulatorischen Aufwand über Messungen des Sauerstoffgehalts beim Ausatmen unter verschiedenen Sprechbedingungen zu bestimmen.

Das Konzept der Minimierung des artikulatorischen Aufwands dient nicht nur in der Phonetik zur Erklärung von Lautvariationen, auch in der Phonologie wird es genutzt, um das Lautsystem einer Sprache und Lautverschiebungen zu beschreiben (z.B. Kiparsky, 1988, Boersma, 1998; Kirchner, in Vorbereitung). Generell besteht in akzentzählenden Sprachen die Tendenz zur Vokalreduktion in Richtung /ə/ (Kohler, 1979), jedoch werden beispielsweise im Englischen die Vokale wesentlich häufiger bis zum /ə/ reduziert als im Deutschen.

Das häufigere Auftreten eines /ə/ berücksichtigt van Bergem (1994b) durch die Unterscheidung von lexikalischer und akustischer Vokalreduktion. Der Begriff der lexikalischen Vokalreduktion stammt aus der Linguistik und beschreibt die Ersetzung eines Vollvokals durch /ə/. Van Bergem nutzt diesen Begriff zur Beschreibung von Vokalreduktionen des noch nicht abgeschlossenen Sprachwandelprozesses. So können bei bestimmten Wörtern im Niederländischen, z.B. in dem Wort /ba:'na:n/ (Banane), Vokale in unbetonter Stellung als Vollvokal, aber auch als [ə] realisiert werden. Die Realisierung des Vokals durch [ə] ist mittlerweile typisch und wird von Hörern akzeptiert. Van Bergem (1995) nimmt an, dass sich die orthographische Repräsentation hemmend und der Einfluss des Englischen im modernen Niederländisch beschleunigend auf den Sprachwandelprozess auswirkt.

### 2.3.3 Zusammenfassung

Vokalrealisierungen weisen eine hohe Variabilität auf. Ein Teil dieser Variabilität wird durch den Begriff der Reduktion umschrieben. Vokalreduktionen stehen in enger Beziehung zu prosodischen Faktoren, wobei es sich als äußerst schwierig erweist, die Einflüsse nach einem Ursache-Wirkungsprinzip anzuordnen. So stellt sich die Frage, ob Silben unbetont wahrgenommen werden, weil sie Reduktionen aufweisen, oder sind sie reduziert, da sie unbetont artikuliert werden. Weiterhin werden Vokalreduktionen unter anderem durch Sprechgeschwindigkeit, Sprechstil sowie Wortart und –häufigkeit beeinflusst.

Den verschiedenen Definitionen von Vokalreduktion ist gemeinsam, dass sich die Formantwerte reduzierter Vokale denjenigen des Neutralvokals nähern und die Vokale eine /ə/-ähnliche Qualität erhalten. Ob Reduktion als Zentralisierung der Formanten betrachtet werden kann, ist fraglich. Akustische Analysen unterstützen die Annahme einer Zentralisierungstendenz der Formantwerte von Vokalen des Deutschen. Weiterhin weisen artikulatorische Untersuchungen darauf hin, dass /ə/ nicht als mit dem Kontext vollständig assimilierter Laut, sondern als Zentralvokal betrachtet werden muss.

*“The analyses showed that this neutral position has to be defined in the space of tract variables (the linguistically relevant goal space), rather than being the consequence of neutral positions for individual articulators. Therefore, a target position for schwa was specified, although this target is completely predictable from the rest of the system; it corresponds to the mean tongue tract-variable position for all the full vowels”* (Browman und Goldstein, 1992:55 f).

Analysen der Formanten des /ə/ im Deutschen zeigen eine zentrale Tendenz auf und die Streuung der Formantwerte ist nicht größer als die anderer Vokale (Barry, 1995).

Vokalreduktionen werden mit einer Minimierung des artikulatorischen Aufwands assoziiert. Die Bestimmung des artikulatorischen Aufwands beruht dabei eher auf einem allgemeinem Konsens unter Phonetikern, wonach einzelne an den Gesten beteiligte Artikulatoren und bestimmte Gesten einen unterschiedlichen Energieanteil benötigen als auf Messungen des tatsächlichen Energieverbrauchs. Verschiede lautliche Realisierungen werden aufgrund ihres Aufwands in eine Rangfolge gebracht (z.B. die verschiedenen Realisierungen des bestimmten Artikels *dem* [de:m] ⇒ [dem] ⇒ [dɛm] ⇒ [dəm] ⇒ [dm] ⇒ [bm] ⇒ [m]; Kohler, 1990:73). Der Grad der Reduktion entspricht einer Abnahme des artikulatorischen Aufwands.

Viel diskutiert wird die direkte Abhängigkeit der Ausprägung der Reduktion von der Sprechgeschwindigkeit bzw. der Dauer von Lauten. Lindblom (1963) erklärt Reduktion durch eine Verfehlung der Zielstellung des vom Sprecher intendierten Lautes im Vokaltrakt, wobei es eine lineare Beziehung zwischen der Ausprägung der Reduktion und der Lautdauer gibt: Je kürzer der

Laut ist, desto mehr wird er reduziert. Die Aussage wurde von Lindblom selbst (1990, Lindblom et al., 1992) dahingehend revidiert, dass die Stärke der Reduktion in Abhängigkeit von den physiologischen Grenzen (maximale Geschwindigkeit der Artikulatoren) vom Sprecher kontrolliert wird. Sprecher passen sich durch bewusste Steuerung ihrer Artikulation den Erfordernissen der Kommunikationssituation an. Dies legt die Vermutung nahe, dass Reduktionen von Sprechern weitgehend geplant sind. Eine andere Annahme ist, dass Vokalreduktionen durch eine weniger kontrollierte Aussprache entstehen (van Bergem, 1993). So werden Wörter, die für den Sprecher einen größeren Informationsgehalt besitzen, betont und präziser artikuliert als Wörter, deren Informationsgehalt gering ist.



### 3 PERZEPTION VON VOKALVARIATIONEN

Im vorherigen Kapitel wurde gezeigt, dass Vokalrealisierungen durch mehrere Faktoren, z.B. Kontext, Prosodie und Sprecher, beeinflusst werden. Dies führt zu einer ausgeprägten Variabilität hinsichtlich der Aussprache von Vokalen und ihrer akustischen Realisierung. Ziel der vorliegenden Arbeit ist, ein auf Perzeption begründetes differenziertes Beschreibungssystem für Vokalreduktionen aufzustellen. Da Vokalqualitäten in erster Linie durch die Formanten bestimmt werden, liegt der Schwerpunkt in der Beschreibung der Zusammenhänge der Formanten und der perzipierten Vokalqualität (Abschnitt 3.1). Es wird der Frage nachgegangen, wie Hörer trotz der großen akustischen Variabilität Vokale identifizieren (Abschnitt 3.2). Dargestellt werden der Zusammenhang zwischen Diskrimination und Identifikation von Vokalen sowie Faktoren, die die Beurteilung von Vokalen beeinflussen (Abschnitt 3.3).

#### 3.1 Akustische Korrelate der Vokalperzeption

In Perzeptionsmodellen werden meistens anstelle von physikalischen Maßeinheiten Größen verwendet, die die vom Gehör vorgenommen Transformationen des Sprachsignals (Tonheit, Frequenz-Ort-Transformation auf der Basilarmembran<sup>1</sup>; s. Zwicker, 1982) berücksichtigen und das Frequenzempfinden des Hörers beschreiben. Das perzeptive Korrelat der Frequenz ist die Tonhöhe. Da die Relation zwischen Frequenz und Empfindung der Tonhöhe nicht durchgängig linear ist, werden die Frequenzwerte in die mel-Skala oder Bark-Skala überführt.

Die mel-Skala (Verhältnistönhöhe) wurde mit Hilfe von Perzeptionsexperimenten mit einfachen Tönen entwickelt (Zwicker, 1982). Probanden hatten die Aufgabe, einen Stimulus so zu verändern, dass im Vergleich zum Ausgangsreiz die Tonhöhenempfindung halbiert bzw. verdoppelt wird. 1 mel entspricht 1/1000stel der Tonhöhe eines 1 kHz Tones. Die Bark-Skala (Tonheit) beschreibt die Frequenzgruppen, innerhalb derer das Gehör Intensitätsempfindungen zusammenfasst. Die Lautstärke einzelner Frequenzanteile von spektral komplexen Schallen werden nicht vollständig summiert, schwächere Frequenzkomponenten werden verdeckt. Die einzelnen Frequenzgruppenbandbreiten können durch Maskierungsexperimente ermittelt werden (s. Zwicker, 1982). Die Bark- und mel-Skala stehen in enger Beziehung zueinander. Unterhalb von 500 Hz ergibt sich ein lineares Verhältnis sowohl zwischen Frequenz und Frequenzgruppe, als auch zwischen Frequenz und Verhältnistönhöhe, oberhalb von 500 Hz ist dieser Zusammenhang logarithmisch.

---

<sup>1</sup> Frequenzen unterhalb von ca. 500 Hz werden auf der Basilarmembran linear abgebildet, während der Zusammenhang zwischen Frequenz und Ort auf der Basilarmembran bei höheren Frequenzen logarithmisch ist.

Untersuchungen zeigen, dass die Vokalqualität hauptsächlich durch die ersten zwei Formanten bestimmt ist (Pols et al., 1969), wobei F2 eine größere Relevanz für die Vokalidentifizierung besitzt als F1 (Neppert und Pétursson, 1986). F3 trägt insbesondere bei den hohen Vorderzungenvokalen zur Vokalqualität bei, da diese Vokale durch einen geringen Abstand von F2 und F3 gekennzeichnet sind.

Delattre et al. (1952) gehören zu den Ersten, die darauf hinwiesen, dass Hinterzungenvokale im Gegensatz zu Vorderzungenvokalen mit nur einem Formantwert synthetisiert werden können. Es folgte die Annahme, dass bei der Perzeption von Vokalen, deren F1 und F2 nah beieinander liegen, die Formantwerte gemittelt werden (*center of gravity effect*). Die Untersuchungen von Chistovich und Lublinskaya (1979) zeigen, dass dann eine Mittelung eintritt, wenn die Distanz zwischen den Formanten weniger als 3 bis 3,5 Bark beträgt. Eine Mittelung der Formantwerte wurde auch bei der Perzeption von Vorderzungenvokalen nachgewiesen, deren F2 und F3 einen Abstand innerhalb der 3 Bark Grenze aufweisen (Johnson, 1989).

Beddor und Hawkins (1990) untersuchten den *center of gravity*-Effekt bei nasalierten und oralen Vokalen. Das Auftreten des *center of gravity*-Effekts von Vokalen mit stark ausgeprägten Maxima im unteren Spektralbereich und einer mit oralen Vokalen vergleichbaren Formantbandbreite hängt in erster Linie vom Abstand der Formanten ab. Die perzipierte Vokalhöhe wird durch F1 beeinflusst, wenn der Abstand zwischen den ersten beiden Formanten ausreichend groß genug ist, ansonsten tritt eine Mittelung der Formantfrequenzen ein. Hingegen sind die Beurteilungen nicht hoher nasaliertter Vokale mit weniger stark ausgeprägten und flachverlaufenden Maxima durch den *center of gravity*-Effekt gekennzeichnet. In ähnlicher Weise zeigt sich dies auch bei einer Änderung der Formantbandbreite bei nicht hohen oralen Vokalen. Die Vokalqualität oraler Vokale mit einer schmalen Bandbreite wird durch die Formantfrequenzen bestimmt, während die von Vokalen mit einer breiten Bandbreite von der Formantfrequenz und –bandbreite abhängt. Die Ergebnisse zeigen, dass der Einfluss der gesamten spektralen Form auf die perzipierte Vokalqualität zunimmt sofern die Prominenz der Energiemaxima im unteren Frequenzbereich abnimmt. Ein Einfluss der Formantbandbreite auf die Lautqualität wird auch von van Bergem (1994b) konstatiert. Reduzierte Vokale weisen insbesondere bei F1 eine größere Bandbreite auf als nicht reduzierte.

Aber auch die Grundfrequenz beeinflusst die Perzeption von Vokalen. Eine Erhöhung der Grundfrequenzwerte führt bei synthetisierten Vokalen zu einer Abnahme der perzipierten Vokalhöhe (Potter und Steinberg, 1950; Miller, 1953). Spätere Untersuchungen bestätigen den Einfluss der Grundfrequenz auf die perzipierte Vokalhöhe in verschiedenen Sprachen, z.B. im Englischen (Nearey, 1989), Schwedischen (Carlson et. al, 1975) und Deutschen (Traunmüller, 1981). Kewley-Port et al. (1996) konstatieren, dass eine Erhöhung der Grundfrequenz die Diskriminierung von Vokalen unabhängig von den Formantfrequenzen verschlechtert.

Neben der Grundfrequenz und den Formanten beeinflusst auch die Dauer die Vokalperzeption. Um ein Sprachsignal als vokalisches Laut wahrnehmen zu können, muss es mindestens 30 bis 50 ms lang sein (Winckel, 1968). Die Dauer von Vokalen wirkt sich aber auch auf die perzipierte Vokalqualität aus. Werden die spektralen Eigenschaften konstant gehalten und nur die Dauer gekürzt, so führt das zu einem Kategorienwechsel bei der Beurteilung der Lautqualität. Der Einfluss der Lautdauer auf die Perzeption wird im Zusammenhang mit den Vokalqualitäten des Deutschen noch ausführlich dargestellt (vgl. Abschnitt 4.3).

### 3.2 Perzeptionsmodelle

Eine zentrale Aufgabe der Perzeptionsforschung ist zu untersuchen, wie Hörer Vokale trotz der varianten akustischen Eigenschaften innerhalb einer Vokalkategorie richtig identifizieren. Im allgemeinen wird angenommen, dass Vokale invariante akustische Eigenschaften besitzen, die eine eindeutige Identifikation ermöglichen. Je nachdem, wo die invarianten Eigenschaften vermutet werden, können zwei Hauptrichtungen innerhalb der Theorien unterschieden werden (van Son und Pols, 1993). Theorien, in denen davon ausgegangen wird, dass sich die invarianten Eigenschaften in der Mitte des Vokals, dem stationären Teil, befinden, basieren auf den spektralen Eigenschaften der Vokale und werden als *Target-Modelle* bezeichnet. Die akustische Varianz wird durch eine Normalisierung aufgehoben. Andere Theorien beruhen auf der Annahme, dass die Formanttransitionen wesentliche Informationen zur Vokalidentifizierung enthalten. Danach ist die Vokalperzeption hauptsächlich durch dynamisch-spektrale Eigenschaften bestimmt. Diese Theorien werden als Modelle der dynamischen Spezifikation (*dynamic-specification*) der Vokalperzeption bezeichnet. In einigen Perzeptionsmodellen werden zur Erklärung der Identifikation von Vokalen zusätzliche Kompensationsprozesse angenommen.

#### Target-Modelle

Sogenannte *Target-Modelle* gehen davon aus, dass nicht die absoluten Formantwerte, sondern deren Verhältnisse zueinander für die Vokalperzeption verantwortlich sind (Lindner, 1977). Traunmüller (1981) zeigt, dass die Vokalhöhe des Deutschen mit dem Abstand zwischen F0 und F1 (gemessen in Bark) korreliert. Die Ergebnisse entsprechen denen von Syrdal und Gopal (1986) für die Vokale im amerikanischen Englisch. Hohe Vokale weisen im Gegensatz zu mittleren und tiefen Vokalen einen Abstand zwischen F0 und F1 auf, der unterhalb von 3 Bark liegt. Inwiefern die Distanz zwischen F0 und F1 ein Hinweis auf die phonologische Vokalhöhe [ $\pm$ hoch] gibt, wurde auch von Hoemeke und Diehl (1994) geprüft. Die Vokalhöhe von /ɪ/ und /ɛ/ wird durch den Abstand zwischen F0 und F1 unterschieden, wobei die F1-Werte zwischen 390 Hz und 530 Hz lagen und die dazugehörigen F0-Werte schrittweise (um 25 Hz) von 100 Hz auf 200 Hz erhöht wurden. Der Grenzwert für die Unterscheidung der Vokalhöhe liegt bei 3-3,5 Bark. Die schrittweise Erhöhung der Grundfrequenz bewirkt eine Senkung des Grenzwerts bis auf 3 Bark. Im Vergleich hierzu korreliert die Vokalhöhe von /i:/ und /ɪ/ (F1: 270-390 Hz, F0: 100-200 Hz)



sowie /ɛ/ und /æ/ (F1: 530-660 Hz, F0: 100-200 Hz), deren Abstände kleiner bzw. größer als 3 Bark sind, mit F1.

Traunmüller (1981) nimmt an, dass bei F0-Werten, die größer sind als 350 Hz, die Vokalhöhe nicht primär durch den Abstand zwischen F0 und F1, sondern durch den zwischen F0 und der 2. sowie teilweise der 3. Harmonischen (die ganzzahligen Vielfachen der Grundfrequenz) bestimmt wird.

Aufgrund der übereinstimmenden Grenzwerte zur Unterscheidung der phonologischen Vokalhöhe und der perzeptiven Mittelung von Formanten (*center of gravity* Effekt) vermuten Fahey et al. (1996), dass es sich um einen Schwellenbereich im Sinne der *Quantal*-Theorie handelt. Nach der *Quantal*-Theorie (Stevens, 1972, 1989) gibt es im Vokaltrakt bestimmte Bereiche, in denen geringe artikulatorische Bewegungen starke akustische oder perzeptive Veränderungen hervorrufen, während in anderen Bereichen große artikulatorische Bewegungen nur geringe akustische oder perzeptive Änderungen verursachen. Die Beziehung zwischen Artikulation, Akustik und Perzeption ist nicht kontinuierlich, sondern weist quantale Sprünge auf. Die Annahme, dass die 3,5 Bark-Grenze eine quantale Größe ist, muss aufgrund der Untersuchungen von Hinterzungenvokalen abgelehnt werden (Fahey et al., 1996).

Weitere Untersuchungen der Vokale im amerikanischen Englisch deuten auf eine Beziehung zwischen dem Abstand von F2 und F3 und dem Artikulationsort hin, wobei die Formantabstände der vorderen Vokale im Unterschied zu den hinteren Vokalen unterhalb der 3 Bark-Grenze liegen (Syrdal und Gopal, 1986). Aufgrund der Beobachtung, dass Hörer bei Beurteilungen synthetisierter Stimuli einen höheren F2-Wert bevorzugen, definiert Fant (1959) einen korrigierten Wert für F2 ( $F2'^2$ ), der das Verhältnis von F2 zu F3 einbezieht. Nach Fant (1983) wird der Artikulationsort für Vokale des Schwedischen durch den Abstand zwischen F2' und F1 bestimmt. Traunmüller (1984, s.a. Fahey et al. 1996) konstatiert, dass jegliche Abstände zwischen benachbarten Formanten sowie zwischen F1 und F0, die unterhalb von 6 Bark liegen, die Vokalqualität beeinflussen. Hörer berücksichtigen dabei kleinere Abstände stärker.

### Modelle der dynamischen Spezifikation

In den Modellen der dynamischen Spezifikation wird davon ausgegangen, dass Vokale aufgrund ihrer Formanttransitionen identifiziert werden. Um den Einfluss der Formanttransitionen auf die Vokalperzeption zu untersuchen, wird beispielsweise das *silent-center* Paradigma verwendet. Innerhalb des *silent-center* Paradigmas ist der stationäre Teil eines Vokals nicht mehr wahrnehmbar, so dass Hörer Vokale nur aufgrund ihrer Transitionen beurteilen können. Untersuchungen von Vokalen des amerikanischen Englisch und des Deutschen zeigen, dass

---

<sup>2</sup>  $F2' = F_2 + \frac{1}{2}(F_3 - F_2) \frac{(F_2 - F_1)}{(F_3 - F_1)}$  (nach Fant, 1959: 80)

sowohl die Transitionen allein (der stationäre Teil ist nicht hörbar; *silent-center*) als auch die stationären Teile zu einer vergleichbaren Identifikationsleistung führen (Strange et al., 1983; Strange, 1989; Strange und Bohn, 1998). Werden intrinsische Dauerinformationen von Vokalen herausgenommen, so dass sich die Vokale in ihrer Dauer nicht unterscheiden, sinkt die Identifikationsleistung zuerst bei den Stimuli, die aus einem stationären Teil bestehen; in erster Linie sind davon die Langvokale betroffen. Die Identifikation von Stimuli ohne stationären Teil und Dauerinformation verschlechtert sich zuerst bei den Kurzvokalen. Weiterhin zeigen Jenkins et al. (1994), dass die Informationen in den Transitionen, die der Vokalidentifizierung dienen, sprecherunabhängig sind. Die Beziehungen zwischen Transition und Perzeption wurde auch in längeren lautlichen Sequenzen untersucht (Benguerel und McFadden, 1989), da, wie die Untersuchungen von Öhman (1966) zeigen, nicht nur die direkt benachbarten Laute die Transitionen beeinflussen. Auch hier wird die Vokalperzeption durch die Transitionen unterstützt. Ein Einfluss der nicht direkt benachbarten Laute auf die Wahrnehmung eines Vokals kann aber nicht ausgeschlossen werden.

Eine weitere Frage, die im Rahmen der dynamischen Theorien diskutiert wird, ist die Relevanz von Anglitt (Transition zum Vokal) und Abglitt (Transition vom Vokal) bei der Vokalperzeption. Eine Untersuchung mit synthetisierten Vokalen zeigt, dass die Stimuli hauptsächlich durch den Abglitt identifiziert werden (van Son und Pols, 1993). Im Gegensatz hierzu stehen die Ergebnisse von Strange und Bohn (1998). Die Präsentation des An- bzw. Abglitts allein verschlechtert in vergleichbarer Weise die Identifikationsleistung der Hörer. Eine Erklärung für die unterschiedlichen Ergebnisse liegt möglicherweise im Versuchsmaterial selbst. Während die Stimuli in der Untersuchung von Strange und Bohn (1998) aus natürlich-sprachlichen Äußerungen generiert wurden, verwendeten van Son und Pols (1993) synthetisierte, so dass die akustischen Eigenschaften der Transitionen systematisch verändert wurden. Zum einen unterscheiden sich die in den beiden Untersuchungen beurteilten Stimuli durch ihre Dauer, das synthetisierte Versuchsmaterial beinhaltet Transitionen mit einer Länge bis zu 75 ms, die Transitionsdauern der aus den natürlich-sprachlichen Äußerungen generierten Stimuli sind maximal halb so lang. Zum anderen sind die unterschiedlichen Dauern des An- und Abglitts im natürlich-sprachlichen Versuchsmaterial nicht vergleichbar häufig verteilt. Während An- und Abglitt von gespannten Vokalen annähernd gleich lang und im Vergleich zu den ungespannten Vokalen kürzer sind, ist die Anglittphase der ungespannten Vokale im Durchschnitt doppelt so lang wie die Abglittphase. Eine Häufung der kürzeren Dauer in der Abglittphase könnte somit möglicherweise einer vornehmlich auf dem Abglitt basierenden Identifikation von Vokalen entgegenwirken. Strange und Bohn (1998) nehmen an, dass die unterschiedliche zeitliche Ausdehnung der Transitionen Informationen beinhalten, um gespannte und ungespannte Vokale mit vergleichbaren stationären Teilen zu disambiguieren.

## Perceptual overshoot

Um die Identifikation von Vokalen, die von ihren akustischen kanonischen Zielstellungen (*formant undershoot*) abweichen, zu erklären, wird häufig ein Kompensationsprozess seitens der Hörer (*perceptual overshoot*) angenommen. Ausgehend von den Erkenntnissen aus der Locus-Forschung und der Motor-Theorie (Lieberman et al., 1963; Liberman und Mattingly, 1985), entwickelten Lindblom und Studdert-Kennedy (1967) ihre Theorie des perzeptiven *overshoot*. Sie untersuchten synthetisierte Vokale (/u/ und /i/) im Nullkontext und in /wVw/- und /jVj/-Silben, wobei die mittleren F2- und F3-Werte stufenweise verändert und die F1-Werte sowie die Loci der Formanttransitionen konstant gehalten wurden. Die Formantmuster der Silben entsprechen einer Parabel. Durch Hörerurteile erhielten sie die F2- und F3-Werte, die von einer /u/-Wahrnehmung zur /i/-Wahrnehmung führen (perzeptive Phonemgrenze). Die Ergebnisse zeigen, dass die Phonemgrenze vom jeweiligen Kontext abhängt. Werden die Stimuli im Kontext beurteilt, so verschiebt sich die Phonemgrenze im Vergleich zu der Grenze der Stimuli im Nullkontext in einen niedrigeren (bei /u/) bzw. höheren (bei /i/) Frequenzbereich. Die Autoren interpretieren dieses Ergebnis folgendermaßen:

„[...] *vowel production and vowel perception could be said to be complementary in the sense that articulatory activity is characterized by undershoot and perception by overshoot*“ (Lindblom und Studdert-Kennedy, 1967:841).

Das Konzept des perzeptiven *overshoot* dient somit der Erklärung der Kompensation eines durch Reduktion bedingten akustischen *undershoot*. Untersuchungen von van Son und Pols (1993) geben jedoch keinen Hinweis auf einen perzeptiven *overshoot*.

Obwohl verschiedene Perzeptionsmodelle (z. B. Di Benedetto, 1989, Strange, 1989; Fox, 1989; Akagi, 1990) Kompensationsprozesse einschließen, herrscht noch wenig Einigkeit über die Strategien, die Hörer nutzen, um Vokale trotz ihrer hohen Variabilität zu identifizieren. Ob diese Strategien durch einfache Prozesse auf akustischer Ebene erklärbar sind, ist fraglich. Perzeptionstheorien werden aus Versuchen mit streng kontrollierten Bedingungen abgeleitet, so dass die Versuchsbedingungen selbst unterschiedliche Perzeptionsstrategien beim Hörer evozieren (Pols und van Son, 1993).

### **3.3 Einfluss von Versuchsbedingungen auf die Vokalperzeption**

Wie sich unterschiedliche Versuchsbedingungen auf die Perzeption von Vokalen auswirken, wird im folgenden dargestellt. Nach der Theorie der Kategorialen Wahrnehmung wird die Perzeption von Konsonanten als kategorial betrachtet (Lieberman et al., 1957), während die der Vokale einem kontinuierlichen Prozess entspricht (Fry et al., 1962). Im Mittelpunkt der Unterscheidung der beiden Wahrnehmungsprozesse steht das Verhältnis von Diskriminierung und Identifizierung von Lauten. Die kontinuierliche Wahrnehmung ist durch eine bessere Diskriminierungs- als

Identifizierungsleistung (Klassifizierung) gekennzeichnet (Miller, 1956). Bei Vergleichen (ABX-Paradigma) werden mehr Unterschiede wahrgenommen als die Anzahl perzipierter Kategorien vermuten lässt. Im Gegensatz zur kontinuierlichen Wahrnehmung wird bei der kategorialen Wahrnehmung die Diskriminierungsleistung durch die Identifikation begrenzt, d.h. Probanden können zwischen zwei Stimuli unterschiedlicher Gruppen unterscheiden, jedoch nicht innerhalb einer Gruppe. Die Identifikationsfunktion von Konsonanten ist daher durch eine abrupte Abnahme der Identifikationsleistung gekennzeichnet. An der 50%-Marke wird die Kategoriengrenze zwischen zwei Lautklassen definiert, hier ist die Diskriminierungsleistung maximal. Hingegen weist die Diskriminierungsfunktion von Vokalen keine ausgeprägten Maxima auf, die Identifikationsleistungen sind weniger konsistent. Im Rahmen der Motortheorie (Liberman et al., 1967; Studdert-Kennedy et al., 1970) werden die zwei Arten der Wahrnehmung auf die unterschiedliche Produktion von Konsonanten bzw. Vokalen zurückgeführt. Die Perzeption von Vokalen ist kontinuierlich, da die artikulatorischen Gesten im Vergleich zu den Konsonanten weniger diskret sind; bei Konsonanten reicht schon eine kleine Abweichung von der Zielstellung des Vokaltraktes aus, um eine andere Konsonantenkategorie zu perzipieren.

Die Theorie der Kategorialen Wahrnehmung ist umstritten. So schreibt beispielsweise Massaro (1998), dass Stimuli verschiedener Kategorien immer besser diskriminiert werden als Stimuli innerhalb einer Kategorie und dieses Verhältnis von Diskriminierung und Identifikation nur unzureichend durch die Theorie der Kategorialen Wahrnehmung vorausgesagt wird. Ein weiteres Problem ist, dass der Diskrimination von Stimuli ein Identifikationsprozess unterliegen kann.

Unterschiede in den Perzeptionsprozessen von Vokalen und Konsonanten (insbesondere der Plosive) werden von Gerrits und Schouten (1998) auf die unterschiedlichen akustischen Eigenschaften dieser beiden Lautgruppen zurückgeführt. Im Vergleich zu Vokalen mit ihren länger andauernden stationären Teilen besitzen Plosive dynamischere Eigenschaften (schnell ändernde Formantübergänge, ExploSIONSgeräusch). Nach Fujisaki und Kawashima (1971) basieren Diskriminierungen von Lauten entweder auf einem auditiven (psychoakustischen) oder auf einem phonetischen (kategorienbezogenen) Prozess (vgl. auch Pisoni, 1975). Eine psychoakustische Bewertung von Lauten ist nur dann möglich, wenn die Dauer der akustischen Schlüsselreize (*cue-duration hypothesis*) ausreichend ist. Vokale hinterlassen eine länger andauernde Spur im auditiven Gedächtnis als beispielsweise Plosive und werden daher einer psychoakustischen Bewertung unterzogen. Bei Plosiven ist ein psychoakustischer Vergleich nicht möglich und die Diskrimination beruht auf phonetischen Betrachtungen. Plosive werden kategorial wahrgenommen und Vokale kontinuierlich.

Ob die Eigenschaften akustischer Schlüsselreize die Art der Perzeption beeinflussen, wurde in einem Experiment mit hyperartikulierten Vokalen (Vokale aus isoliert gesprochenen Wörtern) und Vokalen aus gelesenen Texten von Gerrits und Schouten (1998) untersucht. Sie nehmen an, dass temporale und spektrale Reduktionen zu einer phonetischen Betrachtung führen und somit

reduzierte Vokale im Vergleich zu präziser artikulierte Vokalen kategorial perzipiert werden. Zunächst konnte kein Indiz für eine kategoriale Wahrnehmung für die beiden Arten von Stimuli gefunden werden. Erst als die Bedingungen für den Diskriminationstest variiert wurden, konnte ein Zusammenhang zwischen den Hörerstrategien und den Wahrnehmungsprozessen (kategorial vs. kontinuierlich) festgestellt werden. Neben einem ABX-Vergleich wurde ein 4-Intervall-Verschiedenheitstest (4-interval oddity) durchgeführt: Die Probanden mussten das Intervall (2. oder 3.) in einer Folge von Stimuli (AABA oder ABAA) angeben, in dem sich der Stimulus (B) befindet, der sich von den anderen Stimuli unterscheidet. Die Versuchsanordnung soll anders als beim Diskriminationstest, in dem Stimuli direkt miteinander verglichen werden, Hörer dazu verleiten, die Stimuli einer psychoakustischen Bewertung zu unterziehen. Eine kategoriale Wahrnehmung ist insbesondere dann festzustellen, wenn Hörer eine phonetische Strategie nutzen. Dass zwischen der Wahrnehmung hyperartikulierter und eher reduzierter Vokale kein Unterschied im Perzeptionsprozess gefunden werden konnte, führen die Autoren auf einen zu geringen akustischen Unterschied zwischen den Stimuligruppen zurück. So zeigen van Hessen und Schouten (1999), dass eine kategoriale Perzeption von der Qualität synthetisierter Laute abhängt. Je natürlicher die Stimuli klingen und somit auch die akustische Komplexität zunimmt, desto schwieriger ist es für Hörer sich auf einzelne Parameter zu konzentrieren und sie als psychoakustische Reize zu behandeln. Die Autoren bemerken einschränkend:

*„[...] perception becomes more categorical as stimulus naturalness is improved. The effect is rather small, however, and it occurs for vowels as well as for stop consonants, [...]. It is clear that the relationship is not as simple as we had predicted: making the synthesizer sound more natural does not automatically lead to categorical perception”* (Hessen und Schouten, 1999:71).

Die Beschreibungen zeigen, dass die Theorie der Kategorialen Wahrnehmung und somit die strikte Trennung in eine kategoriale Wahrnehmung von Konsonanten und eine kontinuierliche Wahrnehmung von Vokalen nicht haltbar ist. Sprachperzeption ist kontinuierlich, was nicht ausschließt, dass Hörer Laute kategorisieren können. Das Verhältnis von Diskrimination und Identifikation von Lauten wird aber von den verwendeten Stimuli sowie Vergleichsverfahren (vgl. auch Macmillan et al., 1988, Kewley-Port und Watson, 1994) beeinflusst. Dies führt dazu, dass einige Stimuli mehr oder weniger kategorial wahrgenommen werden. Van Hessen und Schouten (1999) schlagen daher eine Abstufung zur Beschreibung des Grads der kategorialen Perzeption vor (*categorical-perception index*).

Weitere Untersuchungen zur Diskrimination von Vokalen wurden im Rahmen der Theorie des Perzeptiven Magneteffekts (Kuhl, 1992) durchgeführt. Der Ansatz basiert auf den Prototypentheorien, die auch in anderen Bereichen der Wahrnehmungs- und Kognitionsforschung (Rosch, 1975; Busemeyer et al., 1984; Nosofsky, 1986; Palmeri, 1999) zu finden sind. Prototypen sind abstrakte Instanzen, die eine Kategorie repräsentieren und somit die meisten gemeinsamen Eigenschaften mit ihr teilen. Prototypentheorien dienen der Erklärung einer Kategorisierung von

Einheiten. Nach Kuhls Prototypentheorie werden Vokale kategorisiert, indem sie mit mentalen Repräsentanten von Vokalkategorien verglichen werden. Die Prototypen befinden sich in der Mitte einer Kategorie (vgl. Johnson et al., 1993 für eine gegenteilige Ansicht) und fungieren als perzeptive Magnete. Laute in der Nähe von Prototypen werden „angezogen“, akustische Unterschiede werden nicht perzipiert. Vokale, die keine prototypischen Instanzen einer Kategorie sind, besitzen keine magnetischen Eigenschaften.

Die Theorie des Perzeptiven Magneteffekts basiert auf einer Untersuchung synthetisierter Lautvarianten von /i/, in der Probanden die Repräsentativität der Laute für die Vokalkategorie /i/ auf einer Skala von 1 (sehr schlechte Instanz) bis 7 (sehr gute Instanz) bewerteten. Aus den beurteilten Lautvarianten wurde ein Vokal ausgewählt, der die Vokalkategorie am besten repräsentiert (Prototyp) und einer, der als schlechter Repräsentant (Nicht-Prototyp) gilt. In einem weiteren Test wurden dann verschiedene Lautvarianten dieser Vokale synthetisiert, deren F1- und F2-Werte systematisch so variiert wurden, dass sie sich aufgrund der Formantwerte kreisförmig um die beiden Vokale anordnen lassen. Probanden hatten die Aufgabe, die Repräsentativität der Vokalvarianten und in einem Diskriminationstest deren Ähnlichkeit (gleich – ungleich) zu beurteilen. Die Ergebnisse zeigen, dass der Prototyp wesentlich stärker die umliegenden Lautvarianten „anzieht“ als der Vokal, der keine prototypische Funktion besitzt und die Anziehungskraft des Prototyps mit zunehmender akustischer Distanz abnimmt. Der Prototyp wird wesentlich ähnlicher zu den umliegenden Lauten empfunden als der Nicht-Prototyp, d.h. mehr Lautvarianten werden diesem Ziellaut zugewiesen.

Der perzeptive Magneteffekt wurde in einer Reihe von Studien mit unterschiedlichen Testverfahren oder Stimuli überprüft. In einem Experiment von Lively (1993) hatten Hörer die Aufgabe, Stimuli hinsichtlich ihrer Qualität zu identifizieren ([i] vs. nicht [i]). Lively berichtet, dass der Stimulus, der in Kuhls Experiment als schlechter Repräsentant der Kategorie gewählt wurde, keine [i]-Qualität besitzt. Die Beurteilungen wurden daher nicht innerhalb einer Vokalkategorie, sondern zwischen zwei verschiedenen Vokalkategorien ausgeführt. Zudem zeigen sich Unterschiede in der Wahl des prototypischen Vokals. Abweichungen bei der Beurteilung der Stimuli werden auch von Iverson und Kuhl (1995) in einer weiteren Untersuchung festgestellt, was auf die verwendeten Stimuli und mögliche Hörerunterschiede zurückgeführt wird. Beurteilungsunterschiede sind aber nicht nur über verschiedene Experimente hinweg festzustellen, sondern auch innerhalb einer Untersuchung. Hörer bevorzugen unterschiedliche Vokale als beste Instanzen einer Kategorie, wobei diese Prototypen auch am Rande einer Stimuliverteilung liegen können (Lively und Pisoni, 1997; Sussman und Gekas, 1997; Frieda et al., 2000). Interindividuelle Abweichungen bei Beurteilungen von Vokalen beruhen möglicherweise auf unterschiedlichen Produktionsmechanismen. Sowohl in Identifikationstests (Bell-Berti et al., 1979) als auch in Vergleichstests zur perzipierten Ähnlichkeit (Fox, 1982) korrelieren die interindividuellen unterschiedlichen Beurteilungen mit den interindividuellen artikulatorischen Abweichungen der Vokalproduktion.



In Perzeptionsexperimenten mit Stimuli aus der Vokalkategorie /ɪ/ wurden keine Magneteffekte beobachtet. Der fehlende Magneteffekt bei /ɪ/ im Vergleich zu /i/ wird auf die unterschiedlichen Eigenschaften dieser beiden Vokalkategorien zurückgeführt (Sussman und Gekas, 1997). /i/ wird im Vokaltrakt sehr weit vorne gebildet und ist hinsichtlich der akustischen Eigenschaften stabil, auch wenn kleine artikulatorische Abweichungen auftreten (Stevens, 1972, 1989; Beckman et al., 1995). Aufgrund der extremen Stellung und der akustischen Stabilität treten weniger Vertauschungen mit anderen Vokalkategorien bei der Identifikation auf.

Die Untersuchungen zeigen, dass die Theorie des Perzeptiven Magneteffekts nicht generell die Diskriminierung von Vokalen erklären kann, sondern mit vergleichbaren Problemen behaftet ist wie die Theorie zur kategorialen Wahrnehmung. Während in der Theorie der Kategorialen Wahrnehmung (Lieberman et al., 1957) die Diskriminierungsleistung der Hörer aufgrund der Identifikation vorhergesagt wird, wird in der Theorie des Perzeptiven Magneteffekts die Güte der Repräsentation einer Kategorie in Bezug zur Diskriminierung gesetzt. Beurteilungen der Güte und Identifikation weisen eine enge Beziehung auf. Identifikationsrate und Gütebeurteilung von Vokalen ändern sich in gleicher Weise mit ihrem lautlichen Kontext und die Diskriminierungsleistung kann aufgrund von kontextuellen Identifikationsfunktionen vorausgesagt werden (Lotto et al., 1998). Der Zusammenhang zwischen Identifikation, Gütebeurteilung und Diskrimination legt die Vermutung nahe, dass die Theorie des Perzeptiven Magneteffekts die größere Diskriminierbarkeit von Vokalen verschiedener Kategorien im Vergleich zu Vokalen einer Kategorie beschreibt (s. Massaro, 1998).

Nach Sussman und Laucker-Morano (1995) eignet sich die Theorie des Perzeptiven Magneteffekts mit Einschränkungen zur Beschreibung der Diskrimination von Stimuli in Perzeptionsexperimenten. Dem perzeptiven Magneteffekt unterliegt hauptsächlich eine psychoakustische Beurteilung, so dass er von der gewählten Hörerstrategie abhängt.

Dass die Diskrimination von Vokalen nicht nur auf psychoakustischen Beurteilungen beruht, sondern auch eine Verknüpfung zu höheren linguistischen Ebenen aufweist, zeigen Untersuchungen von Kewley-Port und Zheng (1998). Sie überprüften die differentiellen Wahrnehmbarkeitsschwellen (*just noticeable difference*) für die ersten beiden Formanten von Vokalen, wobei die Versuchsbedingungen teilweise eine Einbeziehung einer kognitiven Beurteilung der Stimuli erlaubten. Die Vokale wurden sowohl isoliert als auch in verschiedenen Kontexten (eingebettet in Silben, Phrasen und Sätzen) präsentiert, so dass die Stimuli aufgrund des Kontexts vorhersagbar sind. Die Vorhersagbarkeit wurde durch eine zunehmende Anzahl der vorgegeben Alternativ-Antworten erschwert. Die Ergebnisse zeigen, dass die Wahrnehmbarkeitsschwellen sowohl von der Vorhersagbarkeit der Stimuli als auch von der Kontextlänge beeinflusst werden. Die Diskriminierung von Formanten in isolierten Vokalen ist besser als bei Vokalen, die im Kontext präsentiert werden. Der Einfluss des Kontextes ist am

stärksten ausgeprägt bei der Diskrimination isolierter Vokale und von Vokalen in Silben. Die Wahrnehmbarkeitsschwellen für Vokale in Phrasen und Sätzen sind hingegen vergleichbar. Weiterhin zeigt sich, dass bei wechselndem phonetischen Kontext – Vokale innerhalb von Silben sind von unterschiedlichen Konsonanten umgeben – und einer mittleren Vorhersagbarkeit der Stimuli die Wahrnehmbarkeitsschwellen der einzelnen Hörer stark variieren (Kewley-Port, 1992; Kewley-Port und Zheng, 1998).

### **3.4 Zusammenfassung**

Vokalqualität wird primär durch die Formanten bestimmt, aber auch Grundfrequenz und Dauer (s. auch Abschnitt 4.3) wirken sich auf die Vokalperzeption aus. Welche Strategien Hörer wählen, um Vokale, deren akustische Realisierungen stark variieren können, zu identifizieren, ist noch nicht geklärt. In Perzeptionsmodellen wird angenommen, dass Vokale invariante Eigenschaften besitzen aufgrund derer Hörer die Laute identifizieren. Nach den *Target*-Modellen, die davon ausgehen, dass alle wichtigen akustischen Eigenschaften im stationären Teil der Vokale liegen, wird angenommen, dass eine Normalisierung zwischen der Grundfrequenz und den Formanten zu einer einheitlichen Vokalidentifikation führt. Die Modelle der dynamischen Spezifikation beruhen auf der Annahme, dass wesentliche Informationen für die Vokalidentifikation in den Formanttransitionen enthalten sind. Zur Erklärung der Perzeption reduzierter Vokale wird vielfach ein Kompensationsprozess angenommen. Hörer schätzen aufgrund der Transitionen das kanonische Ziel eines Vokals ab (*perceptual overshoot*). Jedoch ist es auch möglich, dass wichtige Informationen zur Identifikation von Vokalen entweder in der Form der Transition selbst oder im weiteren Kontext (auf Silben- oder Wortebene) liegen (van Son, 1993).

Unterschiedliche, zum Teil sich widersprechende Ergebnisse von Perzeptionsexperimenten zeigen, dass die Perzeption von Vokalen sehr komplex ist. Wie Vokale wahrgenommen werden, hängt von ihrer Qualität (Natürlichkeit), ihrem Kontext, den Versuchsbedingungen, aber auch von den Hörern selbst ab.





## 4 VOKALREDUKTION IM DEUTSCHEN

In diesem Kapitel erfolgt zunächst eine Darstellung des Vokalsystems im Deutschen (Abschnitt 4.1). In Abschnitt 4.2 werden Beschreibungen von Vokalreduktionen vorgestellt. Vokalreduktionen gehen mit Lautdauerverkürzungen einher (s. Abschnitt 2.3). Es wird daher der Frage nachgegangen, wie sich eine Verminderung der Lautdauer auf die perzipierte Vokalqualität und ihre akustische Realisierung auswirkt (Abschnitt 4.3). Den Beschreibungen von Vokalreduktionen gemeinsam ist eine Beziehung zur Gespanntheit der Vokale. Es werden reduzierte gespannte Vokale mit einer ungespannten Qualität perzipiert. Vokalreduktionen und somit auch die Gespanntheit der Vokale weisen eine enge Beziehung zu verschiedenen Faktoren z.B. Betonung, Sprechstil und Sprechgeschwindigkeit auf. Betrachtungen der artikulatorischen Korrelate des Merkmals Gespanntheit sollen Hinweise geben, welchem Manifestationsbereich, dem segmentalen oder dem suprasegmentalen, Vokalreduktionen zugeordnet werden können (Abschnitt 4.4).

### 4.1 Das Vokalsystem des Deutschen

Das Vokalsystem im Standardhochdeutsch umfasst 15 phonologisch relevante Monophthonge (/i/, /y/, /e/, /ø/, /ɛ/, /ɛ:/, /a:/, /a/, /u/, /o/, /ɪ/, /ʏ/, /œ/, /ʊ/, /ɔ/). Hinzu kommt der Neutralvokal, der nur in unbetonten Silben vorkommt. Aufgrund der eingeschränkten Distribution und der beschränkten Distinktivität zu [ɪ], z. B. <Freundenɪ> vs. <Freundɪn>, wird er von einigen Autoren, z.B. Ternes (1987), nicht als Phonem des Deutschen betrachtet. Der phonologische Stellenwert des Neutralvokals soll hier nicht weiter diskutiert werden. Eine ausführliche Darstellung findet sich beispielsweise bei Werner (1972). In dieser Arbeit wird der Neutralvokal in seiner Funktion als reduziertes Allophon (Reduktionsvokal) zu den Monophthongen betrachtet. Die Monophthonge werden in zwei Gruppen eingeteilt, die unter gleichen Kontextbedingungen in Opposition stehen, wobei *„ein längerer und geschlossener Vokal einem kürzeren und offeneren gegenübersteht“* (Kohler, 1995b:170).

Die in Opposition stehenden Vokale sind durch Qualitäts- und Quantitätsunterschiede gekennzeichnet (/i:/ vs. /ɪ/, /y:/ vs. /ʏ/, /ø:/ vs. /œ/, /u:/ vs. /ʊ/, /o:/ vs. /ɔ/). Eine Sonderstellung nehmen die „e“- und „a“-Laute ein. /e:/, /ɛ/ und /ɛ:/ stehen zueinander in Opposition, wobei die ersten beiden sich durch Qualität und Quantität unterscheiden und die beiden letzten nur durch die Quantität. Um die Einteilung hinsichtlich der Qualitäts- und Quantitätsunterschiede aufrecht zu erhalten, werden häufig /e:/ und /ɛ/ gegenübergestellt. Hierfür spricht, dass insbesondere Sprecher des Norddeutschen keine Unterscheidung zwischen /e:/ und /ɛ:/ vornehmen, so wird der distinktive Unterschied z. B. in <Bären> und <Beeren> zu Gunsten von [e:] neutralisiert (vgl. Heid, 1997; Abschnitt 2.3). Nach Moulton (1962) wird [ɛ:] in der zwanglosen Umgangssprache wesentlich weniger produziert als in formaler Sprache. Andererseits weist Ramers (1988), abgesehen von

Äußerungen norddeutscher Sprecher, auf eine relativ häufige Realisierung von [ɛ:] (38,5%) gegenüber [e:] (46,7%) in Rundfunk- und Fernsehsprache hin. Es wurden hierbei nicht nur Äußerungen ausgebildeter Sprecher ausgewertet, da davon ausgegangen werden kann, dass diese aufgrund der sprecherzieherischen Ausbildung eine normorientierte Aussprache bevorzugen.

Die „a“-Laute nehmen insofern eine Sonderstellung ein, da umstritten ist, ob sich die in phonologischer Opposition stehenden Laute allein durch die Quantität oder durch die Qualität und Quantität unterscheiden (vgl. z.B. Ungeheuer, 1977). Akustische und perzeptive Untersuchungen (z. B. Heike, 1971; Wängler und Weiss, 1975; Ramers, 1988; Jessen et al., 1995) weisen darauf hin, dass sich die „a“-Laute in erster Linie durch ihre Quantität unterscheiden (/a:/ vs. /a/). Im Norddeutschen kann nach Kohler (1995b) eine stärkere Frontierung des kurzen Vokals gegenüber dem langen Vokal auftreten. Aufgrund des häufig fehlenden Qualitätsunterschieds wird das Vokalsystem des Deutschen in Form eines Vokaldreiecks abgebildet, wobei die „a“-Laute als offene Mittelzungenvokale betrachtet werden.

Die Begriffe Quantität und Qualität beschreiben auditiv wahrnehmbare Eigenschaften der Vokale. Quantität wird durch das Begriffspaar *kurz* vs. *lang* umschrieben und Qualität durch die im Rahmen der Theorie der distinktiven Merkmale gebräuchlichen Begriffe *gespannt* vs. *ungespannt*.

Die Distribution gespannter langer und ungespannter kurzer Vokale weist einige phonologische Regelmäßigkeiten hinsichtlich Silbenstruktur und Betonung auf. Ungespannte kurze Vokale (der Reduktionsvokal wird hierbei ausgeschlossen) kommen nur in geschlossenen Silben (VC) vor (Ramers, 1988). Hinweise hierfür geben die Befunde (s. a. Wiese, 1988), dass wortfinale Silben nicht auf einen ungespannten Vokal (abgesehen von [ə] bzw. [ɐ]) enden und bei einer wahrnehmbaren Verteilung von zwei aufeinanderfolgenden (heterosyllabischen) Vokalen auf zwei Silben (Hiatus), der erste Vokal nicht ungespannt realisiert wird, beispielsweise in *Kloake* [o.a:] und nicht [ɔ.a:] ([.] steht für die Silbengrenze). Weiterhin kommen nur nach ungespannten Vokalen ambi-silbische Konsonanten vor. Diese Konsonanten werden sowohl der vorherigen als auch der folgenden Silbe zugerechnet. Als Argument für eine Annahme ambi-silbischer Konsonanten wird häufig auf die Orthographie verwiesen, da nach einem betonten kurzen ungespannten Vokal ein Doppelkonsonant folgt (vgl. Tabelle 1; zur Diskussion der Ambisyllabizität s. Jensen, 2000). Dagegen spricht aber, dass die Beschränkung der ungespannten Vokale auf geschlossene Silben nur für Inhaltswörter gilt, da in Funktionswörtern ungespannte Vokale auch wortfinal in offener Silbe realisiert werden, z.B. [du]. Andererseits kann dies auf die fehlende Betonung zurückgeführt werden.

In betonten Silben stehen sich gespannte lange Vokale und ungespannte kurze Vokale gegenüber (Tabelle 1). In unbetonten bzw. nicht hauptbetonten Silben wird die Quantitätsopposition neutralisiert. Es finden sich nur gespannte und ungespannte Kurzvokale, z. B. <Leben> ['le:bən]

<i>gespannt und lang</i>	<i>ungespannt und kurz</i>
bieten [i:]	bitten [ɪ]
Hüte [y:]	Hütte [ʏ]
Beet [e:]	Bett [ɛ]
Höhle [ø:]	Hölle [œ]
Mus [u:]	muss [ʊ]
wohne [o:]	Wonne [ɔ]
Bahn [ɑ:]	Bann [a]

**Tabelle 1: Gespannte lange und ungespannte kurze Vokale in hauptbetonten Silben (Minimalpaare nach Ramers, 1988:1, Ramers unterscheidet bei dieser Aufstellung zwei a-Qualitäten).**

<i>gespannt und kurz</i>	<i>ungespannt und kurz</i>
ḍivers [i]	ḍiffus [ɪ]
Ṃethan [e]	Ṃetteur [ɛ]
Ḳoloss [o]	Ḳolleg [ɔ]
Ḳurier [u]	sḳurriel [ʊ]
Ṣynagoge [y]	Ṣynalöphe [ʏ]
Ọ̈konom [ø]	Ọ̈strogen [œ]

**Tabelle 2: Qualitätsunterschiede in nicht hauptbetonten Silben (aus Siebs, 1969:31 und Moulton, 1962:63).**

vs. <lebendig> [le<sup>l</sup>bendɪç] ([<sup>l</sup>] kennzeichnet die Hauptbetonung). Gespannte kurze Vokale treten vor allem in nicht hauptbetonten Silben von Fremdwörtern auf (Tabelle 2). Auf eine Neutralisierung bzw. Reduzierung des Quantitätskontrasts weisen auch akustische Dauermessungen hin (Rietveld, 1975; Ramers, 1988). In der gesprochenen Sprache werden gespannte kurze Vokale jedoch häufig durch ungespannte kurze Vokale ersetzt (*pretonic laxing*). Nach Ungeheuer (1977) weichen Aussprachenorm und Realisierung voneinander ab, da sich Aussprachewörterbücher (z. B. Siebs, 1969; Duden, 1990) vor allem an der Orthographie orientieren. Ungespannte Vokale sollen vor Doppelkonsonanten realisiert werden, gespannte Vokale vor einem Konsonanten.

## 4.2 Beschreibungen von Vokalreduktion

Vokalreduktionen sind durch einen Verlust der Vokalqualität gekennzeichnet. Auf symbolischer Ebene werden die realisierten Vokalqualitäten üblicherweise durch den IPA-Zeichensatz beschrieben (vgl. Abschnitt 2.1). Verwendet werden hierbei die Symbole zur Beschreibung der Vokalqualität (z.B. [i]) und der Quantität (z.B. [i:]); einige Autoren (z.B. Kohler, 1990; vgl. Abschnitt 2.3.3) gebrauchen zusätzlich für eine feinere Differenzierung der perzipierten Vokalqualität die Diakritika. Für sich genommen machen die IPA-Symbole keine Angaben über die Reduktion der Vokale. Um den Grad der Reduktion zu beschreiben, werden realisierte und kanonische Vokalqualitäten in Beziehung gesetzt und nach dem Grad ihrer Abweichungen, der mit

der Stärke des artikulatorischen Aufwands assoziiert wird (Abschnitt 2.3.2), sortiert. Die Anzahl der unterschiedlich perzipierten Vokalqualitäten entspricht der Anzahl der Reduktionsstufen. Dies führt aber auch dazu, dass die Beschreibungen von Vokalreduktionen im höchsten Maße von dem jeweiligen Beobachter abhängen.

Gegenstand vieler Untersuchungen von Vokalreduktion sind insbesondere Funktionswörter. Im Vergleich zu Inhaltswörtern werden Funktionswörter aufgrund ihres geringen Informationsgehalts häufig unbetont realisiert und reduziert (Abschnitt 2.3.1). Im folgenden wird aufgezeigt, wie verschiedene Autoren den Reduktionsprozess von Vokalen in Funktionswörtern darstellen.

Ausgangspunkt von Beschreibungen von Vokalreduktionen in Funktionswörtern ist die betonte Form, die auch als *starke Form* (z. B. Kohler, 1995b) oder *volle Form* (Meinhold, 1973) bezeichnet wird. Die Vokalreduktion wird beispielsweise bei Meinhold (1973) durch drei Stufen beschrieben: ein gespannter langer Vokal wird hinsichtlich der Quantität reduziert (1. Stufe), es erfolgt eine Qualitätsänderung in Richtung eines ungespannten Vokals (2. Stufe) oder der Vokal wird durch den Reduktionsvokal ersetzt (3. Stufe), z. B. <die> [di:] ⇒ [di] ⇒ [dɪ] ⇒ [də].

Dedenbach (1987) unterscheidet zwischen der betonten Form [di:] mit einem gespannten Langvokal und der 1. Stufe [di] mit einem gespannten Kurzvokal eine zusätzliche halblange Stufe [diː] ([di:] ⇒ [diː] ⇒ [di] ⇒ [dɪ] ⇒ [də]). Auch Meinhold (1973) stimmt der Möglichkeit einer halblangen Realisierung gespannter Vokale zu, sieht sie aber als perzeptiv nicht erfassbar an. Schindler (1974) hingegen schließt einen wahrnehmbaren Unterschied aufgrund einer Dauerverkürzung allein aus und setzt daher nur zwei Reduktionsstufen an. Hierbei wird ein gespannter langer Vokal zu einem ungespannten kurzen Vokale reduziert (1. Stufe) oder durch den Reduktionsvokal ersetzt (2. Stufe; [di:] ⇒ [dɪ] ⇒ [də]).

Während in den vorgestellten Beschreibungen Reduktion als Zentralisierungsprozess in Richtung des Neutralvokals betrachtet wird, unterscheiden sie sich in der Bewertung der perzeptiven Relevanz der Lautdauer. Grundsätzlich lassen sich zwei Hypothesen gegenüberstellen (vgl. Ramers, 1988; Jessen, 1993):

1. Reduktionen sind zunächst durch eine Quantitätsänderung der Vokale gekennzeichnet, wobei die Qualität erhalten bleibt. Erst mit zunehmender Reduktionsstärke ändert sich die Qualität.
2. Reduktionen bewirken Quantitäts- und Qualitätsänderungen gleichermaßen.

Letztere Annahme kommt dem *target undershoot*-Modell wesentlich näher, in dem davon ausgegangen wird, dass ein Dauerverlust mit einer Zentralisierung der Formantwerte einhergeht.

### 4.3 Dauer und Vokalqualität

Sendlmeier (1981) untersuchte den Einfluss der Vokaldauer auf die perzipierte Qualität von Vokalen. Gespannte lange Vokale wurden hinsichtlich ihrer Dauer gekürzt, ungespannte kurze Vokale gelängt. In einem Identifikationsexperiment mussten Hörer die Lautqualität beurteilen. Die Alternativantworten waren in orthographischer Form vorgegeben. Die gekürzten gespannten Vokale [i], [y] und [u] wurden mit der Qualität der entsprechenden Langvokale ([i:], [y:] und [u:]) perzipiert, [e], [ø] und [o] hingegen als die ungespannten Vokale [ɪ], [ʏ] bzw. [ʊ]. Die gelängten ungespannten Vokale [ɛ:] und [ɔ:] wurden mit der entsprechenden Kurzvokalqualität ([ɛ] und [ɔ]) bewertet, [ɪ:] und [ʊ:] wurden ähnlich wie [e:] bzw. [o:] wahrgenommen. Eine Dauerverlängerung der gerundeten ungespannten Vorderzungenvokale ergab keine häufigere Zuordnung zu einer bestimmten Vokalklasse. Gelängtes [ɣ:] wurde als [y:], [ʏ] oder als [ø:] perzipiert, ein gelängtes [œ] als [ø:] oder [œ]. [a:] und [a] wurden aufgrund ihrer Dauer diskriminiert.

Die Ergebnisse zeigen, dass sich eine Kürzung der Dauer bei geschlossenen gespannten Vokalen nicht auf die perzipierte Qualität auswirkt, während eine Kürzung der Dauer bei mittelgeschlossenen und mittel-offenen gespannten Vokalen eine Qualitätsänderung hervorruft. Dieses Ergebnis wird auch von Wängler und Weiss (1975, s. a. Weiss, 1974) unterstützt.

Jessen (1993) untersuchte die akustischen Eigenschaften isoliert gesprochener Quasiminimalpaare (z. B. despotisch - Despotist), die sich aufgrund ihrer lexikalischen Wortbetonung unterscheiden (gespannter Lang- und ungespannter Kurzvokal in hauptbetonter Stellung, gespannter und ungespannter Kurzvokal in nebenbetonter sowie in nichtbetonter Stellung). Unter Haupt- und Nebenbetonung differieren die Vokale durch mindestens einen der ersten beiden Formanten und die Dauer. Die unbetonten Vokale unterscheiden sich durch mindestens einen Formanten, Dauerunterschiede sind nicht festzustellen. Die offenen /a/-Vokale unterscheiden sich nur durch die Dauer. Nach Jessen wirkt sich bei nicht offenen Vokalen eine Verringerung der Betonung vor allem auf die quantitativen Unterschiede aus. Der Qualitätsunterschied zwischen nicht hauptbetonten ungespannten und gespannten Kurzvokalen bleibt erhalten. Bei den offenen Vokalen gibt es keine qualitativen Unterschiede. Diese Ergebnisse werden auch durch eine Untersuchung von Ramers (1988) unterstützt. Im Gegensatz zu Jessen (1993), der isoliert produzierte Wörter verwendete, überprüfte er die Formant- und Dauerwerte von Vokalen in Wörtern aus Nachrichtensendungen. Akustische Analysen von Vokalrealisierungen in Logatomen und aus verschiedenen Sprechsituationen (Nacherzählungen, gelesenen Texten, Interviews und informellen Gesprächen) zeigen, dass die Ausprägung des Unterschieds zwischen gespannten und ungespannten Kurzvokalen von der jeweiligen Situation abhängt (Steriopolo, 1994); jedoch ist er in allen Sprechsituationen vorhanden.

Die akustischen Analysen weisen auf Unterschiede zwischen gespannten und ungespannten Kurzvokalen hin, was auch in der Beschreibung des Reduktionsprozesses nach Meinhold (1973, s.o.) berücksichtigt wird. Jedoch zeigen die Ergebnisse der Perzeptionsexperimente nur bei den

hohen Vokalen einen perzeptiv relevanten Unterschied zwischen langen und kurzen gespannten Vokalvarianten auf. Für die tieferen (mittel-geschlossenen und mittel-offenen) Vokale scheinen Schindlers (1974, s.o.) Beschreibungen von Vokalreduktion mit einer gleichzeitigen Änderung der Vokaldauer und –qualität zutreffender zu sein.

Wie das Experiment von Sendlmeier (1981) zeigt, werden jedoch bei einer Dauerkürzung der mittel-hohen Vokale nicht die in phonologischer Opposition stehenden ungespannten Vokalqualitäten perzipiert ([e] → [ɛ], [ø] → [œ], [o] → [ɔ]; wobei → für *wahrgenommen als* steht), sondern die im Vokaldreieck höher liegenden ungespannten Vokale ([e] → [ɪ], [ø] → [ʏ], [o] → [ʊ]). Eine Längung der hohen ungespannten Vokale wirkt mit Ausnahme von [ʏ:] in umgekehrter Richtung ([ɪ:] → [e:], [ʊ:] → [o:]). Dies wirft die Frage auf, ob bei einer Reduktion dieser Vokale nicht noch weitere Qualitäten berücksichtigt werden müssen. Die Beschreibungen der Vokalreduktion orientieren sich an dem phonologischen Prinzip der Oppositionen gespannter Lang- und ungespannter Kurzvokale. Da Phoneme per Definition phonologische Oppositionen bilden (beispielsweise könnten auch /i:/ und /ɛ/ in *bieten* und *betten* gegenübergestellt werden), basieren die Paarbildungen auf dem zusätzlichen Kriterium der phonetischen Ähnlichkeit (Kohler, 1995b). Aber sind sich die in Opposition stehenden Vokale phonetisch ähnlich? Das hier vorgestellte sowie weitere Perzeptionsexperimente (Lindner, 1976; Heike, 1970) weisen Abweichungen zu den phonologischen Paarbildungen auf. Aufgrund der perzipierten Ähnlichkeit von /e:/-/ɪ/, /o:/-/ʊ/ und /ø:/-/ʏ/ scheint die phonologische Zusammenfassung der Vokale zu gemeinsamen Lautqualitäten (/e:/-/ɛ/, /o:/-/ɔ/ und /ø:/-/œ/) nicht gerechtfertigt zu sein. Auch akustische Analysen sprechen für eine größere Nähe von /e:/-/ɪ/, /o:/-/ʊ/ und /ø:/-/ʏ/ (Jørgensen, 1969; Hess, 1972). Die phonologischen Paarbildungen werden möglicherweise nicht nur durch phonetische sondern auch durch orthographische Ähnlichkeiten beeinflusst (vgl. auch Diskussionsbeitrag von Fischer-Jørgensen, in Ungeheuer, 1977b:82 f)

Die vorangestellten Betrachtungen weisen darauf hin, dass die einzelnen Reduktionsstufen möglicherweise vokalabhängig sind. Unabhängig von der Vokalkategorie sind jedoch der gemeinsame Bezug zwischen Reduktion und dem Merkmal Gespanntheit. Die Reduktion von Vokalen und somit ihre Gespanntheit hängen ihrerseits unter anderem von Betonung, Sprechstil und Sprechgeschwindigkeit ab. Es stellt sich die Frage nach dem Manifestationsbereich von Vokalreduktion. Ist Vokalreduktion ein segmentales oder suprasegmentales (prosodisches) Phänomen?

#### **4.4 Reduktion als segmentales oder suprasegmentales Phänomen?**

Reduzierte Vokale werden mit einer ungespannten Vokalqualität perzipiert, aber was sind die artikulatorischen Korrelate der Gespanntheit und somit der Reduktion? Im folgenden werden Ansätze zur Beschreibung des Gespanntheitsunterschieds auf segmentaler und suprasegmentaler Ebene vorgestellt. Ausführliche Darstellungen finden sich beispielsweise in Ramers (1988) und



Mooshammer (1998). Die hier exemplarisch aufgeführten Untersuchungen zu den artikulatorischen Korrelaten der Gespanntheit sollen Hinweise für eine Einordnung der Reduktion als segmentales oder suprasegmentales Phänomen geben.

#### 4.4.1 Reduktion als segmentales Phänomen

Das Begriffspaar *gespannt* vs. *ungespannt* wird mit einer unterschiedlichen Muskelspannung assoziiert (Abschnitt 2.1). Nach Jakobson, Fant und Halle (1951) wird die Artikulation gespannter Phoneme, im Vergleich zu den ungespannten, mit einem größeren Druck und Muskelanspannung sowie einer extremeren Artikulationsstellung verbunden.

*„Tense phonemes are articulated with greater distinctness and pressure than the corresponding lax phonemes. The muscular strain affects the tongue, the wall of the vocal tract from its neutral position”* (Jakobson, Fant und Halle; 1951, 1988:635).

Moulton (1962:62) sieht hingegen die periphere Lage der gespannten Vokale nur als Folgeerscheinung der unterschiedlichen Muskelspannung.

*„The greater muscular energy of the tense vowels causes them to be articulated in more extreme positions. The features ‘lax’ and ‘tense’ therefore include the features ‘centralized’ and ‘decentralized’.”*

Neben der größeren Muskelspannung und der peripheren Stellung wird die Artikulation gespannter Vokale im Vergleich zu der ungespannter Vokale mit einer geringeren Variabilität assoziiert (vgl. Chomsky und Halle, 1968). Die artikulatorischen Korrelate werden dabei in Beziehung zur Lautdauer gesetzt. Gespannte Vokale weisen aufgrund ihrer längeren Dauer eine stärkere Muskelspannung, periphere Artikulationsstellung und größere Differenziertheit auf als ungespannte Vokale. Jedoch gibt es auch Sprachen, z.B. Estnisch, in denen die Kurzvokale die gleiche Qualität besitzen wie Langvokale (Lindau, 1977). Häufig wird aber auch Gespanntheit mit dem in einigen phonologischen Theorien (vgl. Wiese, 1996) gebräuchlichen Merkmal *advanced tongue root* [ $\pm$  ATR] verglichen. Diese Distinktion beruht auf der unterschiedlichen Position der Zungenwurzel (nach vorne bzw. nach hinten gelagerte Zungenwurzel) bei der Artikulation von Vokalen in einigen westafrikanischen Sprachen (z. B. Igbo, Akan) mit zugehöriger Vokalharmonie. Die Merkmale Gespanntheit und *advanced tongue root* werden bereits bei Jakobson und Halle (1963) zusammengefasst (Mooshammer, 1998).

Sweet (1892) gehört zu den Ersten, die die Vokale des Englischen nach ihrem Spannungsgrad in zwei Gruppen aufteilen. Er bezeichnet die beiden Gruppen mit *narrow* und *wide*, was mit den Begriffspaar *geschlossen* vs. *offen* gleichgesetzt werden kann (Ramers, 1988). Nach Sweet wird durch eine Erhöhung der Zungenspannung die Zungenoberfläche gewölbt. Bei der Unterscheidung

der Hinterzungenvokale ist zusätzlich die Spannung der Uvula und des weichen Gaumens von Bedeutung.

Sievers (1893) erklärt den Unterschied zwischen (geschlossenem) /i/ und (offenem) /ɪ/ mit einer unterschiedlichen Spannung der Zunge<sup>3</sup>, später geht Sievers (1901; nach Ramers, 1988) davon aus, dass nicht nur die Zunge das primäre Unterscheidungskriterium für die beiden Vokalgruppen ist, sondern alle beteiligten Artikulationsorgane, insbesondere die Stimmbänder und Lippen. Er benennt diese beiden Gruppen mit den Termini *gespannt* vs. *ungespannt*, um zu betonen, dass keine Abhängigkeit zum Öffnungsgrad besteht.

### Muskelspannung

Durch elektromyographische Messungen wurde von MacNeilage und Sholes (1964) die Zungenmuskelaktivität bei der Produktion von Vokalen des amerikanischen Englisch in einem [pVp]-Kontext untersucht. Hierfür wurden Elektroden auf die Zungenoberfläche eines Sprechers aufgeklebt. Die Ergebnisse zeigen eine Abnahme der Zungenmuskelaktivität von [i] über [e] zu [ɪ], während sich bei den Hinterzungenvokalen keine eindeutige Korrelation zwischen der Aktivität und der Gespanntheit herausstellt. Für letztere können insbesondere die Messergebnisse, der im Bereich der Zungenwurzel angebrachten Elektroden, nicht der Aktivität eines bestimmten Muskels zugeordnet werden.

Raphael und Berti (1975) untersuchten gezielt die Muskelaktivität mit Nadelelektroden. Drei Sprecher des amerikanischen Englisch äußerten die Vokalpaare [i]-[ɪ], [u]-[ʊ] und [e]-[ɛ] in einem [əpVC]-Kontext. Der finale Konsonant wurde mit bilabialen Plosiven besetzt, die sich hinsichtlich ihrer Stimmhaftigkeit unterschieden ([p] oder [b]). Bei zwei der Probanden konnte eine höhere Aktivität im Genioglossus, dessen Kontraktion zu einer Aufwärtsbewegung der Zunge führt, und des Longitudinalis inferior, dessen Aktivität die Zungenspitze senkt und den Zungenkörper verkürzt, festgestellt werden. Styloglossus, Longitudinalis superior und Palatoglossus wiesen keine oder keine unterschiedliche Aktivität bei der Produktion der Vokalpaare, oder eine nicht angenommene erhöhte Aktivität bei ungespannten Vokalen auf. Insbesondere das Fehlen einer stärkeren Aktivität des Styloglossus bei der Produktion von [u] im Vergleich zum [ʊ], der die Zunge nach hinten und aufwärts zieht, entspricht nicht den Erwartungen.

Weitere Untersuchungen der intrinsischen und extrinsischen Zungenmuskulatur (Genioglossus anterior und posterior, Hyoglossus, Styloglossus, Geniohyoid, Mylohyoid, Orbicularis oris superior) bei der Vokalartikulation im amerikanischen Englisch zeigen unterschiedliche Muskelaktivitäten für die Vorder- und Hinterzungenvokale auf (Alfonso et al., 1984; Baer et al.,

---

<sup>3</sup> „Dabei wird die Vorderzunge loser und schlaffer als beim geschlossenen *i*<sup>l</sup>“ (Sievers, 1893: 97; *i*<sup>l</sup> entspricht *i*).

1988). Der Genioglossus posterior weist innerhalb der Vorder- und Hinterzungenvokale bei der Produktion der geschlossenen Vokale [i] bzw. [u] die höchsten Aktivitäten auf. Genioglossus anterior und Geniohyoid, die die Zunge aufwärts und nach vorne ziehen, zeigen die größten Aktivitäten innerhalb der Vorderzungenvokale, wobei die Aktivitäten bei [ɪ] und [ɛ] geringer sind als die bei [i] und [e]. Jedoch sind die Muskeln am stärksten bei der Produktion von [æ] aktiviert. Im Gegensatz zu den Ergebnissen von Raphael und Berti (1975) wurde eine stärkere Beteiligung des Styloglossus bei der Produktion der Hinterzungenvokale [u], [o] und [ɔ] im Vergleich zu [ʊ] festgestellt. Baer et al. (1988) kommen zu dem Schluss, dass nur die gespannten Vokale [i], [u] und [e] im Vergleich zu den ungespannten Vokalen eine stärkere Aktivität in den primär an der Artikulation beteiligten Muskeln aufweisen. Die anderen Vokale können aufgrund der Aktivitätsmuster nicht in eine gespannte vs. ungespannte Gruppe unterteilt werden.

Experimente zur Muskelanspannung zeigen kein eindeutiges Korrelat der Gespanntheit auf. Nachteilig ist, dass sich die elektromyographischen Messungen auf Sprachen beziehen, die eine Diphthongierung von Vokalen beinhalten (vgl. Fischer-Jørgensen, 1990; Mooshammer, 1998). Auf einen Einfluss der Diphthongierung auf die Ergebnisse weisen auch Baer et al. (1988) hin.

#### Stimmlippenspannung und Luftstrom

Wie bereits in Abschnitt 2.2 beschrieben, besitzen Vokale eine intrinsische Grundfrequenz. Die höheren Grundfrequenzwerte geschlossener Vokale im Vergleich zu den offenen werden auf eine Erhöhung des Luftstroms und/oder der Zugkraft der Zungenmuskeln bei höherliegender Zungenstellung auf die Spannung der Stimmlippen zurückgeführt. Da ungespannte Vokale eine niedrigere Zungenposition besitzen als die in phonologischer Opposition stehenden gespannten Vokale, ist anzunehmen, dass die Stimmlippenspannung oder der Luftstrom bei den ungespannten Vokalen geringer ist als bei den gespannten.

Wie Sievers (1901) nimmt Meyer (1913) an, dass nicht die Zungenhöhe, sondern die Stimmbandpressung oder der subglottale Druck für die Unterscheidung gespannter und ungespannter Vokale entscheidend ist. Untersuchungen mit Hilfe einer Luftröhrenfistel weisen jedoch keine signifikanten Unterschiede bei der Produktion gespannter und ungespannter Vokale auf. Schuhmacher (1972) untersuchte Luftstrom und Luftdruck bei den Vokalen des Deutschen in unterschiedlichen konsonantischen Kontexten und stellte keinen signifikanten Unterschied beim Luftstrom gespannter und ungespannter Vokale fest. Auch der subglottale Druck zwischen den Vokalgruppen ist vergleichbar. Er geht daher davon aus, dass die Stimmlippenspannung mit der Gespanntheit der Vokale korreliert.

Untersuchungen der Grundfrequenz der deutschen Vokale einer Sprecherin bestätigen, mit Ausnahme des Vokalpaars [o:]-[ɔ], die Annahme, dass ungespannte Vokale geringere Grundfrequenzwerte besitzen als die in phonologischer Opposition stehenden gespannten Vokale (Heuft, 1999). Auch Messungen der F0- und F1-Werte innerhalb der Gruppe der gespannten und

ungespannten Vokale mehrerer Sprecher weisen darauf hin, dass ungespannte Vokale aufgrund ihrer niedrigeren Zungenposition (höhere F1- Werte) niedrigere F0-Werte besitzen als die in phonologischer Opposition stehenden gespannten Vokale (Antoniadis und Strube, 1981). Werden jedoch die in phonologischer Opposition stehenden Vokalpaare einzeln betrachtet, zeigen sich unterschiedliche Ergebnisse in Abhängigkeit von der Vokalhöhe. Während die geschlossenen ungespannten Vokale ([ɪ], [ʊ] und [ʏ]) zu einem höheren F0- und F1-Wert (Mittel über alle Sprecher) tendieren als die jeweiligen gespannten Vokale ([i:], [u:] und [y:]), unterscheiden sich die F0-Werte der mittleren gespannten und ungespannten Vokale ([e:]-[ɛ], [ø:]-[œ], [o:]-[ɔ]) nicht oder nur geringfügig. Die F1-Werte der mittleren ungespannten Vokale hingegen weisen alle eine Erhöhung auf. Der offene ungespannte Vokal [a] zeigt im Vergleich zu dem gespannten Vokal [a:] höhere F0-Werte und niedrigere F1-Werte auf. Antoniadis und Strube (1981) weisen jedoch auf eine fehlende Signifikanz zwischen den F0-Werten der Vokalpaare hin.

Fischer-Jørgensen (1990) untersuchte die artikulatorischen und akustischen Korrelate der Gespanntheitsopposition der ungerundeten Vorderzungenvokale ([i:], [ɪ], [e:], [ɛ], [a:], [a]) im Deutschen. Die Vokale waren entweder in sinnlose oder in sinnvolle Wörter eingebettet, die im Satzkontext von mehreren Sprechern gelesen wurden. Der vertikale Kiefer- und Lippenöffnungsgrad wurde aufgrund von Videoaufnahmen bestimmt; Messungen der Zungenhöhe basierten auf Palatogrammen. Die artikulatorischen Messungen wurden mit den F0- und F1-Werten verglichen. Die Grundfrequenzmessungen bei den einzelnen Sprechern zeigen, dass die ungespannten geschlossenen Vokale nicht immer einen höheren F0-Wert besitzen als die gespannten Vokale (die F0-Werte der [ɪ]-Realisierungen zweier Sprecher liegt unterhalb des Wertes der gespannten). Bei den übrigen Vokalpaaren finden sich nur zwei signifikante Unterschiede. Die durchschnittlichen F0-Werte der gespannten und ungespannten Vokalen über alle Sprecher sind vergleichbar. Innerhalb der Gespanntheitsopposition sind keine eindeutigen Korrelationen zwischen Kieferöffnungsgrad und F0 sowie zwischen Zungenhöhe bzw. F1 und F0 feststellbar. Zusätzliche Messungen an zwei Sprechern zeigen eine Erhöhung des Luftstroms bei der Produktion ungespannter Vokale nur eines Sprechers auf.

### Supraglottale Artikulation

Der Gespanntheitsunterschied zwischen Vokalen wird auch mit den Termini offen vs. geschlossen beschrieben, die auf die Zungenhöhe und den Kieferöffnungsgrad verweisen (vgl. auch Abschnitt 2.1). Meyer (1913) stellte aufgrund plastographischer Untersuchungen im Deutschen fest, dass der ungespannte Vokal [ɪ] mit einer wesentlich niedrigeren Zungenhöhe artikuliert wird als der gespannte Vokal [i:]. Die Zunge erreicht eine Position zwischen [e] und [ɛ].

Wood (1975) geht davon aus, dass die Vokalhöhe (z.B. /i/ vs. /e/) auf eine unterschiedliche Kieferhöhe zurückgeführt werden kann, wobei die Zungenposition vergleichbar ist. Hingegen ist der Gespanntheitsunterschied (z.B. /i/ vs. /ɪ/) durch eine vergleichbare Kieferhöhe aber unterschiedliche Zungenposition gekennzeichnet. Er verglich die Artikulation gespannter und

ungespannter Vokale aus verschiedenen Sprachen (z. B. Deutsch, Arabisch, Russisch und Französisch) und stellt entsprechend seiner Annahme fest, dass die Zungenposition bei [ɪ] niedriger ist als bei [e], und der Kieferöffnungsgrad bei [ɪ] vergleichbar zu dem von [i] ist. [ɪ] unterscheidet sich von [e] und [ɛ] hinsichtlich des Kieferöffnungsgrads (Wood, 1982). Hingegen kann bei den Hinterzungenvokalen [o] vs. [ɔ] keine Abnahme der Zungenhöhe verzeichnet werden (s. Wood, 1975). Auswertungen von Röntgenbildern von Vokalrealisierungen im Deutschen (Bolla und Valaczkai, 1986, nach Fischer-Jørgensen, 1990) zeigen, dass [ɪ] mit einer größeren Kieferöffnung artikuliert wird als [i:] und [e:]. Fischer-Jørgensen (1990) schließt auf mögliche Sprecherunterschiede bei der Produktion gespannter und ungespannter Vokale.

Hoole und Kühnert (1996) untersuchten mittels elektromagnetischer Artikulographie die Beteiligung von Kiefer- und Zungenbewegung bei gespannten und ungespannten Vokalen von sieben deutschsprachigen Probanden. Sowohl das Ausmaß der Kiefer- als auch der Zungenbeteiligung weisen sprecherspezifische Unterschiede bei der Realisation des Gespanntheitskontrastes auf. Für alle Sprecher zeigt sich jedoch im Verhältnis zur Zungenbeteiligung eine stärkere Kieferbeteiligung beim Vokalhöhenunterschied (z.B. /i:/ vs. /e:/) als beim Gespanntheitsunterschied (z.B. /i:/ vs. /ɪ/).

Die Gespanntheit von Vokalen wird auch mit der Stellung der Zungenwurzel assoziiert (Fischer-Jørgensen, 1990). Wie bereits eingangs erwähnt, finden sich diese Unterschiede hinsichtlich einer nach vorne (*advanced tongue root*) bzw. nach hinten gelagerten Zungenwurzel (*unadvanced tongue root*) in afrikanischen Sprachen mit Vokalharmonie. Diese Verlagerung der Zungenwurzel mit einer unterschiedlichen vertikalen Positionierung des Larynx führt zu einer Erweiterung (nach vorne gezogene Zungenwurzel mit gesenktem Larynx) bzw. Verengung (nach hinten gezogene Zungenwurzel mit erhöhtem Larynx) des pharyngalen Bereichs (Ladefoged, 1964; Lindau et al., 1972; Lindau, 1979). Den Unterschied zwischen den Merkmalen Gespanntheit und *advanced tongue root* stellen Lindau et al. (1972) und Lindau (1979) heraus. Während beispielsweise im Deutschen die Positionierung der Zungenwurzel mit der Zungenhöhe gekoppelt ist, ist in afrikanischen Sprachen die Zungenwurzelposition ein unabhängiges Merkmal. Unterschiede hinsichtlich der Gespanntheit bewirken eine Verschiebung von F1 und F2. In afrikanischen Sprachen ist durch eine Erweiterung des pharyngalen Bereichs nur eine Zentralisierung in Bezug auf die F1-Werte zu verzeichnen. Die Merkmale *advanced tongue root* und Gespanntheit sind aus phonetischer Sicht nicht identisch (Ladefoged und Maddieson, 1996).

Das Merkmal Gespanntheit wird mit der Variabilität bei der Artikulation in Zusammenhang gesetzt. Gespannte Vokale werden präziser artikuliert als ungespannte. Bohn et al. (1991) untersuchten mit Hilfe der Glossometrie (ein optoelektronisches Verfahren) die Variabilität bei wiederholten Realisierungen betonter Vokale des Deutschen im Kontext „ob er [bVp] habe“. Die Ergebnisse deuten – unerwartet – darauf hin, dass die ungespannten Vokale präziser artikuliert werden als die gespannten.

Die Resultate von Bohn et al. (1991) wurden nochmals von Hoole und Kühnert (1995) überprüft. Zusätzlich wurde der Einfluss des lautlichen Kontextes auf die Variabilität der Vokalartikulation untersucht. Analysiert wurden elektromagnetische Artikulographiedaten betonter Vokale (außer [ɛ:]) des Deutschen in einem symmetrischen Kontext, bestehend aus bilabialen, alveolaren und velaren Plosiven, die von sechs Sprechern realisiert wurden. Messungen der Variabilität getrennt nach dem Konsonantenkontext (*Token-to-Token Variabilität*) zeigen eine, wenn auch geringfügig ausgeprägte, höhere Variabilität für die ungespannten Vorderzungenvokale ([ɪ], [ʏ], [ɛ], außer für [œ]) im Vergleich zu den in phonologischer Opposition stehenden gespannten Vokalen, während bei den Hinter- ([u:]-[ʊ], [o:]-[ɔ]) und Mittelzungenvokalen ([a:]-[a]) die gespannten Laute eine höhere Variabilität aufweisen. Die Unterschiede zwischen Vorder- und Hinter- bzw. Mittelzungenvokalen führen die Autoren auf die am Produktionsprozess beteiligte Zungenfläche bzw. den Konstriktionsort zurück. Während bei der Produktion von Vorderzungenvokalen die gesamte Zunge beteiligt ist, kann bei der Realisierung der Mittel- bzw. Hinterzungenvokale das Zungenblatt stärker variieren. Die Artikulographiedaten des hinteren Zungenbereichs der Hinter- und Mittelzungenvokale weisen keine größere Variabilität auf als die Vorderzungenvokale. Wird hingegen die kontextuelle Variabilität betrachtet, zeigen ungespannte Vokale eine stärkere Variabilität als gespannte Vokale auf.

Insgesamt weisen die bisher vorgestellten Untersuchungen kein einheitliches physiologisches oder artikulatorisches Unterscheidungskriterium für die gespannten und ungespannten Vokale auf. Messungen der Grundfrequenz als akustisches Korrelat der Stimmlippenspannung variieren in Abhängigkeit der Vokalpaare und Sprecher. Die Unterschiede zwischen den Vokalpaaren sind meist sehr klein und häufig nicht signifikant. Auch die Zentralisierungstendenz hinsichtlich der Zungenlage ungespannter Vokale zeigt sich nicht bei allen Vokalpaaren. Auch werden sprecherspezifische Unterschiede bei der Artikulation gespannter und ungespannter Vokale festgestellt. Nach Hoole und Kühnert (1996) ist der Variabilitätsunterschied zwischen gespannten und ungespannten Vokalen in erster Linie auf koartikulatorische Einflüsse zurückzuführen, was für einen suprasegmentalen Manifestationsbereich der artikulatorischen Korrelate der Gespanntheit spricht.

#### 4.4.2 *Reduktion als suprasegmentales Phänomen*

Zu den Modellen, die die Gespanntheit als suprasegmentales Phänomen werten, gehören die Ansätze zum Silbenschnitt. Manifestationsbereich der Gespanntheit ist, wie schon der Name sagt, die Silbe. Silben werden in Abhängigkeit vom Schnitt des Konsonanten, der dem Nukleus folgt, mit einem sanften oder scharfen Schnitt assoziiert. Silben mit gespanntem Nukleus besitzen einen sanften Schnitt, während Silben mit ungespanntem Nukleus vom Folgekonsonant abgeschnitten werden (scharfer Schnitt). Grundlage für diese Beschreibungen bilden durch Perzeption oder Introspektion gewonnene Eindrücke.



Sievers (1893) nimmt einen unterschiedlichen expiratorischen Druck am Silbenende bei gespannten und ungespannten Vokalen an. Beim scharfen Schnitt (stark geschnittener Silbenaccent) wird der Vokal „[...] durch den folgenden Consonanten in einem Moment abgelöst, wo er noch voll und kräftig ertönt (unmittelbar hinter dem Silbengipfel), der jähe Absturz der Expiration fällt in den oder die silbenschiessenden Consonanten, die daher kräftig beginnen, aber mehr oder weniger abrupt endigen [...]“ (Sievers, 1893:204). Bei einem sanften Schnitt (schwach geschnittener Silbenaccent) „[...] erfolgt die Umstellung der Organe für den Consonanten, nachdem der Sonant bereits deutlich geschwächt ist (also eine merkbare Zeit nachdem der Silbengipfel passirt ist) [...]“ (Sievers, 1893:204).

Jespersen (1913) und Trubetzkoy (1939) ersetzen die Begriffe *stark* bzw. *schwach geschnittener Silbenaccent* durch die Termini *fester* bzw. *loser Anschluss*. Trubetzkoy verwendet den Begriff der *Silbenschnittkorrelation*. Er definiert die Anschlussarten folgendermaßen:

„Der feste Anschluß „schneidet“ sozusagen das Ende des Vokals ab und daher muß der so „geschnittene“ Vokal kürzer als der normale, ungeschnittene Vokal sein. Die Silbenschnittkorrelation beruht somit auf einer privativen Opposition, deren merkmalsloses Glied der „ungeschnittene“ vollablaufende Vokal ohne festen Anschluß an einen folgenden Konsonanten ist“ (Trubetzkoy, 1939:196).

Die Termini *loser* und *fester Anschluss* heben im Gegensatz zu den von Sievers (1893) verwendeten Begriffen *schwach* und *stark geschnittener Silbenaccent* die Bedeutung des konsonantischen Kontexts hervor. Nach Trubetzkoy (1939) sind Silben mit losem Anschluss *dehnungsfähig*, während Silben mit festem Anschluss *nicht dehnungsfähig* sind, d.h. die Quantität der Vokale wird durch den Anschluss bedingt. Ungespannte Vokale bzw. Vokale mit festem Anschluss besitzen aufgrund der Anschlussart eine kurze Dauer, gespannte Vokale mit losem Anschluss werden gedehnt.

Vergleiche des Intensitätsverlaufs, des Luftstroms sowie des artikulatorischen Drucks bei der Produktion von Vokalen mit folgenden Konsonanten (Fischer-Jørgensen und Jørgensen, 1969) können Sievers' Theorie (1893) nicht bestätigen. Artikulatorische Hinweise für eine engere Kopplung bei ungespannten Vokalen im Vergleich zu gespannten Vokalen wurden erst in neuerer Zeit entdeckt (Mooshammer et al., 1999). Hoole et al. (1994) untersuchten die Monophthonge des Deutschen (außer /ɛ:/) in CVC-Folgen mit Hilfe elektromagnetischer Artikulographie. Der konsonantische Kontext ist symmetrisch und wird durch die stimmlosen Plosive /p/, /t/, /k/ gebildet. Die CVC-Silben wurden von fünf Sprechern mit normaler und schneller Sprechgeschwindigkeit produziert. Analysiert wurden die Artikulationsbewegungen von Unterlippe, -kiefer und Zunge (vier verschiedene Messpunkte). Die aufgezeichneten Bewegungen wurden in drei Abschnitte unterteilt: Transition zum Vokal (Öffnungsbewegung), Transition vom Vokal (Schließbewegung) und dem Intervall, das zwischen Öffnungs- und Schließbewegung liegt



(Nukleus). Die Einteilung beruht auf dem Geschwindigkeitsverlauf des jeweiligen Artikulators, der für die Produktion der benachbarten Plosive genutzt wird (Unterlippe für /p/, Zungenspitze für /t/, Zungenrücken für /k/).

Die Ergebnisse zeigen, dass in Abhängigkeit von der Sprechgeschwindigkeit die zeitliche Ausdehnung der Öffnungs- und Schließbewegungen der ungespannten Vokale kaum mehr beeinflusst werden als die der gespannten Vokale. Jedoch wirkt sich eine Änderung der Sprechgeschwindigkeit auf die Dauer der Nuklei der beiden Vokalgruppen in unterschiedlicher Weise aus. Während die Nuklei der ungespannten Vokale bei Erhöhung der Sprechgeschwindigkeit nur einen geringfügigen Unterschied aufweisen (vergleichbar zu den sprechgeschwindigkeitsabhängigen Unterschieden zwischen den Transitionen), sind die Nuklei der gespannten Vokale bei schneller Sprechgeschwindigkeit im Vergleich zur normalen Sprechgeschwindigkeit zur Hälfte (52,3%) komprimiert. Die Beschleunigungskurve der Artikulatoren weist bei den ungespannten Vokalen innerhalb des Nukleus nur einen Beschleunigungsgipfel auf, bei den gespannten treten häufig zwei an den Flanken befindliche Gipfel auf. Aufgrund der Anzahl der Beschleunigungsgipfel, des unterschiedlichen Einflusses der Sprechgeschwindigkeit auf die Dauer der Nuklei sowie einer höheren Variabilität der Nukleidauer gespannter Vokale schließen Hoole et al. (1994) auf einen Unterschied zwischen den Kopplungen der Transitionen und der Nuklei ungespannter und gespannter Vokale. Ungespannte Vokale besitzen einen festen Anschluss und gespannte einen losen Anschluss.

Weitere Untersuchungen (Kroos et al., 1997) der zeitlichen Organisation innerhalb der Öffnungs- und Schließbewegungen unterstützen eine unterschiedliche Kopplung bei gespannten und ungespannten Vokalen im Sinne der Silbenschnitttheorie. Innerhalb der Öffnungsbewegung weisen ungespannte Vokale im Verhältnis zur Schließbewegung eine längere Beschleunigungsphase (definiert als der zeitliche Abschnitt vom Beginn bis zum Geschwindigkeitsgipfel der Öffnungsbewegung bzw. vom Geschwindigkeitsgipfel bis zum Ende der Öffnungsphase im Verhältnis zur Dauer des gesamten Bewegungsablaufs) auf, während gespannte Vokale innerhalb der Öffnungsbewegung eine relativ zur Schließbewegung kurze Beschleunigungsphase zeigen. Diese Asymmetrie zwischen Öffnungs- und Schließphase führen die Autoren auf die geringere Abnahme der Bewegungsgeschwindigkeit bei der Produktion der ungespannten Vokale zurück, was wiederum für einen festen Anschluss der ungespannten Vokale spricht.

Betrachtungen der Dauer zwischen den Geschwindigkeitsgipfeln der Öffnungs- und Schließphase im Verhältnis zur Dauer des gesamten Bewegungsablaufs (*Peak-to-Peak Ratio*) ergeben im Vergleich zu den gespannten Vokalen kleinere Werte für die ungespannten Vokale. Selbst die Werte der mit normaler Sprechgeschwindigkeit geäußerten ungespannten Vokale fallen geringer aus als die der schnell gesprochenen gespannten Vokale.

Die Ergebnisse von Hoole et al. (1994) und Kroos et al. (1997) belegen, dass sich bei gespannten und ungespannten Vokalen die zeitliche Organisation der Öffnungs- und Schließbewegung unterscheidet. Diese serielle Abfolge der Gesten wirkt sich auf die Dauer der Nuklei aus. Bei einer Erhöhung der Sprechgeschwindigkeit werden die Nuklei der ungespannten Vokale genauso stark reduziert wie die Transitionen, d.h., dass die Sprechgeschwindigkeit die einzelnen Segmente im gleichen Ausmaß beeinflusst und sich somit auf den ganzen Artikulationsablauf auswirkt. Bei den gespannten Vokalen zeigt sich ein stärkerer Einfluss auf die Nuklei als auf die Öffnungs- und Schließphasen, d.h., dass durch das Sprechtempo der Artikulator nicht im gleichen Ausmaß (nicht-linear) beeinflusst wird (Mooshammer, 1998).

Mooshammer et al. (1999) untersuchten die Auswirkung der Betonung auf die Artikulationsgesten gespannter und ungespannter Vokale. Das Datenmaterial bestand aus [tVt]-Silben, wobei die gespannten und ungespannten Vokale ([i:]-[ɪ], [y:]-[ʏ], [e:]-[ɛ], [ø:]-[œ], [a:]-[a], [u:]-[ʊ], [o:]-[ɔ]) in betonter und unbetonter Position stehen. Der Wechsel der Betonung wurde durch eine kontrastive Hervorhebung evoziert („Ich habe [tVtə], nicht [tVt<sup>1</sup>al] gesagt“). Wie schon in den vorherigen Analysen (Hoole et al., 1994; Kroos et al., 1997) wurde die Geste der Zungenspitze zeitlich in zwei Transitionsabschnitte und Nukleus unterteilt.

Messungen der zeitlichen Organisation der Öffnungs- und Schließgesten zeigen, dass sich ein Verlust der Betonung bei den gespannten Vokalen vor allem auf den Nukleus der gespannten Vokale auswirkt, hingegen werden bei den ungespannten Vokalen alle drei Abschnitte (Öffnungs- und Schließphase sowie Nukleus) im gleichen Verhältnis komprimiert. Die Nukleusdauer der unbetonten gespannten Vokale ist vergleichbar zu der der betonten ungespannten Vokale. Die Ergebnisse weisen darauf hin, dass die Öffnungsphase ausgehend von den gespannten betonten Vokalen über die gespannten unbetonten und ungespannten betonten Vokale zu den ungespannten unbetonten Vokalen zunehmend abgeschnitten wird. Die zeitliche Organisation der einzelnen Phasen gespannter unbetonter Vokale und ungespannter betonter Vokale ist vergleichbar. Weitere Hinweise für einen Schnitt geben auch die Dauern zwischen den Geschwindigkeitsgipfeln im Verhältnis zur Dauer des gesamten Bewegungsablaufs (*Peak-to-Peak Ratio*). Im Vergleich zu einer Erhöhung der Sprechgeschwindigkeit (s.o.) führt der Verlust der Betonung bei den gespannten Vokalen zu einer stärkeren Komprimierung der Nukleusdauer, bei den ungespannten Vokalen hingegen ist der Einfluss von Sprechgeschwindigkeit und Betonung vergleichbar.

Im Gegensatz zu den unbetonten Vokalen ist die Zungenspitzenbewegung der betonten Vokale durch eine höhere Amplitude gekennzeichnet. Hinsichtlich des Gespanntheitsunterschieds finden sich jedoch sprecher- und vokalspezifische Unterschiede. Bei einigen Sprechern zeigt sich eine höhere Bewegungsamplitude für die ungespannten betonten Vorderzungenvokale als für die in Opposition stehenden gespannten betonten Vokale, bei anderen Sprechern sind keine Unterschiede festzustellen. Obwohl das Ausmaß des Schnitts bei gespannten unbetonten und ungespannten betonten Vokalen vergleichbar ist, zeigt die Berücksichtigung weiterer Artikulatoren (Kiefer,

Zunge und Unterlippen), die für die Vokalartikulation genutzt werden, einen Unterschied zwischen gespannten und ungespannten Vokalen. Mit Ausnahme von /a/ zeigt sich eine stärkere Zentralisierung der Zielstellungen ungespannter Vokale und unbetonter Vokale relativ zu den gespannten bzw. betonten Vokalen. Eine tendenziell stärkere Zentralisierung der ungespannten Vokale im Vergleich zu den gespannten Vokalen ist auch bei vergleichbarer Dauer festzustellen (Mooshammer, 1998). Dies deutet darauf hin, dass der Gespanntheitsunterschied in unbetonter Stellung nicht vollständig neutralisiert wird. Mooshammer et al. (1999) nehmen an, dass gespannte und ungespannte Vokale eine unterschiedliche Zielstellung besitzen, so dass die Zielposition der Zunge für ungespannte Vokale nicht einfach durch das Abschneiden der Öffnungsphase gespannter Vokale modelliert werden kann.

#### 4.4.3 Zusammenfassende Diskussion

Im Gegensatz zu den Analysen auf segmentaler Ebene, die kein einheitliches, von der Dauer unabhängiges artikulatorisches Korrelat zur Unterscheidung gespannter und ungespannter Vokale aufzeigen, weisen Untersuchungen der artikulatorischen Bewegungsabläufe auf eine unterschiedliche Produktion gespannter und ungespannter Vokale hin. Obwohl die internale Organisation der Transitionen der beiden Vokalgruppen vergleichbar ist, zeigen die ungespannten Vokale eine stärkere Kopplung zwischen den Öffnungs- und Schließgesten. Eine Erhöhung der Sprechgeschwindigkeit oder der Verlust der Betonung wirken in vergleichbarer Weise auf die Produktion gespannter und ungespannter Vokale. Bei gespannten Vokalen wird der Nukleus komprimiert. Bei ungespannten wirken Sprechgeschwindigkeits- und Betonungsänderungen auf den gesamten Artikulationsablauf.

Die Ergebnisse unterstützen die in der Theorie des Silbenschnitts formulierten Annahmen, dass sich Unterschiede zwischen gespannten und ungespannten Vokalen auf der Silbenebene manifestieren. Aufgrund der unterschiedlichen innervokalischen Kopplung von Öffnungs- und Schließgesten gespannter und ungespannter Vokale sowie der stärkeren Koartikulation der ungespannten Vokale mit dem konsonantischen Kontext wird angenommen, dass das Merkmal Gespanntheit als suprasegmental zu werten ist (Mooshammer, 1998). In diesem Sinn kann auch die Reduktion von gespannten zu ungespannten Vokalen als ein suprasegmentales Phänomen betrachtet werden.

Die Hypothese, dass *„die ungespannten Vokale nicht einfach nur zentralisierte gespannte Vokale sind, sondern spezifische, von der Sprechgeschwindigkeit und damit der Dauer unabhängige artikulatorische Eigenschaften aufweisen“* (Mooshammer, 1998:148), unterstützt die Ansicht, dass Reduktionen von Sprechern geplant sind (Abschnitt 2.3.3).

Wenn aber die Artikulationsgesten gespannter und ungespannter Vokale voneinander unabhängig sind, stellt sich die Frage nach der Sprechökonomie (Abschnitt 2.3.2). Dies würde bedeuten, dass unabhängig von der Reduktion die Produktion ungespannter Vokale weniger Energie benötigt.

Einen Hinweis, dass die Produktion ungespannter Vokale weniger aufwendig ist als die gespannter, gibt die stärkere Überlappung von Konsonanten- und Vokalgesten bei ungespannten Vokalen, wie die Ergebnisse von Hoole und Kühnert (1995; vgl. Abschnitt 4.4.1 und Mooshammer, 1998) zeigen. Dies führt zu einer Verkürzung des Haltens einer Artikulationsgeste. Weiterhin zeigen die Analysen zum Beschleunigungsverhalten der Artikulatoren (Hoole et al. 1994; Kroos et al., 1997), dass ungespannte Vokale nur einen Beschleunigungsgipfel besitzen, während gespannte Vokale meist zwei- oder dreigipflig sind. Dies deutet darauf hin, dass unabhängig von einer Reduzierung die Produktion gespannter Vokale mehr Kraft erfordert als die ungespannter Vokale. Eine Reduktion von gespannten zu ungespannten Vokalen kann somit als Minimierung des artikulatorischen Aufwands betrachtet werden.

Wie die Beschreibung des Vokalsystems zeigt, gibt es im Deutschen zwei Gruppen von Vokalen – gespannte lange und ungespannte kurze – die sich hinsichtlich ihrer Distribution unterscheiden. Dies führt immer wieder zur Diskussion der Frage (vgl. Sendlmeier, 1981; Ramers, 1988), durch welche Eigenschaft sich die Vokale primär unterscheiden. Mit anderen Worten, eine Eigenschaft – Qualität bzw. Quantität – muss aus phonologischer Sicht redundant sein. Dieses Merkmal kann aus der Distribution der Laute abgeleitet werden. Je nachdem welcher Eigenschaft der Vorzug gegeben wird, werden die Vokale in Lang- und Kurzvokale oder in gespannte und ungespannte Vokale unterteilt. Ausgehend davon, dass Qualität und Quantität auditive Eindrücke beschreiben, weisen perzeptive Untersuchungen kein einheitliches Ergebnis hinsichtlich der primären Relevanz einer der beiden Eigenschaften auf.

Die in dieser Arbeit vorgenommene Unterteilung in gespannte und ungespannte Vokale wird aus den Beschreibungen von Vokalreduktionen abgeleitet, reduzierte gespannte Vokale werden mit einer ungespannten Vokalqualität perzipiert. Ramers (1988:89 f) wertet dies als Argument für eine primäre Relevanz der Qualität:

*„Wenn nämlich im Hauptton Vokaldauer und –qualität immer gekoppelt sind, im Nebenton aber der Qualitätsgegensatz, wenn auch nur in allophonischer Variation, fortbesteht, Dauerunterschiede dagegen vollkommen nivelliert werden, so bildet die Gespanntheit das gegen Reduktionstendenzen resistenteres Merkmal. Dies wiederum könnte als Indiz dafür gewertet werden, daß es auch unter Hauptakzent die primäre Distinktionseigenschaft darstellt, die Dauer dagegen ein redundantes, von der Qualität ableitbares, automatisch hinzutretendes Merkmal bildet.“*

Aber auch die fehlende Neutralisierung des artikulatorischen Gespanntheitskontrast bei vergleichbarer Dauer ungespannter und gespannter Vokale spricht für eine primäre Relevanz der Gespanntheit (mit Ausnahme der „a“-Laute).



## 5 FRAGESTELLUNG UND DATENMATERIAL

Eine geläufige Beschreibung von Vokalreduktion ist die Darstellung der Vokalqualität auf der symbolischen Ebene durch einzelne Stufen. Ausgangspunkt ist die kanonische Realisierung über den Neutralvokal /ə/ bis zum kompletten Wegfall des Vokals, bei gespannten Vokalen wird hierbei noch der kurze gespannte und ungespannte Vokal eingefügt, z.B. [i:]→[i]→[ɪ]→[ə]→∅. Jedoch herrscht noch wenig Einigkeit über die Anzahl der zu diskriminierenden Reduktionsstufen (vgl. Abschnitt 4.2). Perzeptionsexperimente (Abschnitt 4.3) weisen darauf hin, dass zum einem die Anzahl der Reduktionsstufen vokalabhängig ist und zum anderen, dass noch weitere als die in phonologischer Opposition stehenden Vokalqualitäten berücksichtigt werden müssen.

Ein weiteres Problem stellt die Darstellungsweise von Vokalreduktionen dar. Sowohl Vokalqualität als auch Reduktion werden durch ein und dasselbe Symbol beschrieben, wobei die Symbole keine Aussagen über den perzipierten Grad der Reduktion machen. Aufgrund einer kontinuierlichen Wahrnehmung ist anzunehmen (Abschnitt 3.3), dass Hörer mehr Vokalqualitäten unterscheiden als durch eine Transkription symbolisiert wird. Der Beschreibung von Reduktionen von Vokalqualitäten auf einem Kontinuum entsprechen qualitätsbeschreibende Ausdrücke über den Grad der Reduktion (von nicht reduziert bis stark reduziert).

In der vorliegenden Untersuchung soll ein perzeptives Beschreibungssystem aufgestellt werden, das Reduktionen von Vokalqualitäten möglichst fein differenziert und in Stufen (von nicht reduziert bis stark reduziert) einteilt. Reduktionen sollen nicht nur adäquat wiedergegeben, sondern auch zuverlässig hinsichtlich der intersubjektiven Beurteilungen beschrieben werden. Eine Einteilung von Reduktion in Stufen erlaubt, diese zu anderen quantitativen Parametern (beispielsweise Betonungsstärken, akustische Parameter) in Beziehung zu setzen.

Zunächst ist zu klären, wie die Stufen der Reduktion von Vokalen ermittelt werden können, um ein reliables und adäquates Beschreibungssystem zu erhalten (Abschnitt 5.1). Reliabilität ist in der Testpsychologie ein Maß für die Zuverlässigkeit oder Genauigkeit, mit der sich ein Testergebnis reproduzieren lässt (Arnold et al., 1980:1880; Lienert und Raats, 1994). Der Adäquatheit entspricht die Validität (Gültigkeit), d.h. die Messgenauigkeit eines Tests in Bezug auf ein Kriterium (Arnold et al., 1980:2423; Lienert und Raats, 1994). In Abschnitt 5.2 wird dann im Zusammenhang mit der Auswahl des Versuchsmaterials auf die Frage der externen Validität, d.h. die Übertragbarkeit der Versuchssituation auf natürliche Situationen, eingegangen. Es erfolgt eine Beschreibung und Aufbereitung des Datenmaterials. In Abschnitt 5.3 werden die für die Auswertung des Versuchsmaterials verwendeten statistischen Verfahren vorgestellt.

## 5.1 Methodik

Um ein perzeptives System erstellen zu können, in dem Reduktionen von Vokalqualitäten durch Stufen beschrieben werden, muss geklärt werden, ob Hörer Reduktionen reliabel in Stufen einteilen können und, wenn ja, wie viele Stufen perzipiert werden. Im allgemeinen liegt bei Untersuchungen zur Identifikation von Vokalen die Übereinstimmungsrate zwischen Hörern bei 74% (vgl. Fairbanks und Grubb, 1961; Klein et al., 1970). Demgegenüber weisen Ergebnisse zur Beurteilung der Reduktion bzw. der Qualität reduzierter Vokale erheblich höhere intersubjektive Abweichungen auf. Rietveld und Koopmans-van Beinum (1987) ließen synthetisierte Vokale (/i/, /u/, /a/), deren F1- und F2-Werte zunehmend zentralisiert wurden, hinsichtlich ihrer Vokalqualität beurteilen, wobei die zu bestimmenden Vokalqualitäten in orthographischer Form vorlagen (geschlossene Antwortform). Während für die intersubjektive Beurteilung der nicht reduzierten Vokale eine hohe Übereinstimmung (>85%) festgestellt werden kann, liegt die Übereinstimmungsrate bei den reduzierten Vokalen erheblich darunter. Außerdem zeigt sich, dass die Konsistenz der Beurteilungen vokalabhängig ist. Während bei /u/ erst bei stärkerer Reduktion die Urteile voneinander abweichen, dann aber besonders stark (höchste Übereinstimmungsrate liegt bei 33%), zeigen sich bei /i/ und /a/ schon bei geringerer Reduktion Unterschiede in den Beurteilungen (höchste Übereinstimmungsrate liegt bei 59%). Van Bergem (1995b) überprüfte die intersubjektive Übereinstimmung bei der Klassifizierung von Vokalen in zwei Gruppen – voller Vokal (nicht reduzierter Vokal) und [ə] – und kommt zu dem Schluss, dass die Übereinstimmung zwischen den Hörern recht gering ist. Aylett und Turk (1998) ließen Vokale, deren Dauer und Amplitude kontrolliert wurden, hinsichtlich ihrer Wohlartikuliertheit beurteilen. Auch hier zeigt sich, dass die durchschnittliche Korrelation zwischen den Urteilen sehr gering ausfällt.

Neben dem Problem der Zuverlässigkeit der Beurteilungen reduzierter Vokale stellt sich das Problem der adäquaten Beschreibung und somit die Frage nach einer Methode zur Ermittlung von Reduktionsstufen. Zur Auffindung phonetischer Kategorien sind Perzeptionstests, in denen phonologische Vokalkategorien vorgegeben werden (d.h. Identifikationstest mit geschlossener Antwortform) unzureichend (Ladefoged, 1967), denn die Beurteilung von Hörern wird durch diese vordefinierten Kategorien beeinflusst. Hinzu kommt, dass die Perzeption reduzierter Vokale von Kompensationsstrategien geleitet ist und verschiedene Versuchsbedingungen unterschiedliche Hörerstrategien evozieren (Abschnitte 3.2 und 3.3). Um eine feinere Differenzierung von Lautqualitäten zu erhalten, schlagen Pols et al. (1984) Perzeptionsexperimente innerhalb eines Vergleich-Paradigmas vor. Im Gegensatz zu Identifikationsexperimenten finden sich bei Beurteilungen der Lautqualität in Vergleichsexperimenten keine Hinweise auf einen perzeptiven *overshoot* (Akagi et al., 1994:506):

*„[...] in the identification test, the subjects were forced to perceive stimuli as a phoneme (vowel), whereas in the matching test, they may not perceive the sounds phonetically but analytically. [...] So, most probably, overshoot will only be observed in natural speech (possibly to compensate for*



*articulatory undershoot), however, only under some contextual conditions. It results only from a higher-order central processing mechanism.”*

Akagi et al. (1994) nehmen an, dass durch Vergleichstests das analytische Hören unterstützt wird und daher eine Kompensation auf höhere linguistische Prozesse zurückgeführt werden kann. Beim analytischen Hören entspricht die Wahrnehmung von Sprache idealiter einem *Bottom-up*-Prozeß. Der Hörer richtet die Aufmerksamkeit auf die (psycho)akustischen Eigenschaften der Äußerung, das semantische Wissen über die Sprache wird bei der Perzeption nicht berücksichtigt (Vieregge, 1989). Die Hypothesen von Akagi et al. (1994) werden von neueren Untersuchungen im Rahmen der Theorie der Kategorialen Wahrnehmung (Hessen und Schouten, 1999; vgl. Abschnitt 3.3) unterstützt, die zeigen, dass die Art der Wahrnehmung (kontinuierlich vs. kategorial) durch die Natürlichkeit der Stimuli beeinflusst wird. Je natürlicher und somit auch komplexer Stimuli werden, umso schwieriger ist es, sie allein aufgrund ihrer (psycho)akustischen Eigenschaften zu beurteilen.

Jedoch können die Ergebnisse von Tokuma (1997, s. auch Tokuma 1999, 2000) nur teilweise die Annahme von Akagi et al. (1994) bestätigen. Hörer hatten die Aufgabe, die Qualität reduzierter Vokale aufgrund der ersten beiden Formantfrequenzen zu bestimmen, indem sie synthetisierte isolierte Vokale (Teststimuli) mit Vokalen im symmetrischen Konsonantenkontext (/CVC/; wobei die Konsonanten mit einem stimmhaften bilabialen oder alveolaren Plosiven besetzt wurden; Referenzstimuli) verglichen. Die Formantfrequenzen der ersten beiden Formanten der Referenzstimuli wurden in Abhängigkeit vom jeweiligen Plosiv und Vokal konstant gehalten, während die ersten beiden Formantfrequenzen der Teststimuli von den Hörern solange verschoben wurden, bis eine vergleichbare Qualität zwischen Referenz- und Teststimulus perzipiert wurde. Die Ergebnisse weisen darauf hin, dass die Existenz eines perzeptiven *overshoot* von dem Bereich der Transition abhängt. Ist dieser groß, so ist ein perzeptiver *overshoot* zu verzeichnen – die Hörer weisen den Teststimuli abhängig von einer konvexen bzw. konkaven Form der Transition Formantwerte oberhalb bzw. unterhalb der Transitionswerte zu. Ist der Transitionsbereich klein, so liegen die Frequenzwerte zwischen den Werten der Transition. Die Annahme, dass Kompensation und Natürlichkeit der Stimuli unter gewissen kontextuellen Bedingungen einen Zusammenhang aufweisen und somit durch höhere linguistische Prozesse geleitet werden (s.o. Zitat von Akagi et al. 1994), bietet keine Erklärung für ein transitionsabhängiges Auftreten eines perzeptiven *overshoot*. (Tokuma, 2000).

Insgesamt ist festzustellen, dass es zwar keine eindeutigen, aber einige richtungsgebende Hinweise gibt, wie die Anzahl der zu ermittelnden Reduktionsstufen bestimmt werden könnte. Um eine differenzierte Beschreibung von Reduktionen zu erhalten, müssen mögliche Kompensationsstrategien seitens der Hörer, die zu einer Wahrnehmung phonologischer Vokalqualitäten führen, ausgeschlossen werden. Da Hörer mehr Lautqualitäten diskriminieren als identifizieren können (vgl. Abschnitt 3.3) und zudem Identifikationsexperimente eine geringe

intersubjektive Übereinstimmung bei den Beurteilungen reduzierter Vokalqualitäten aufweisen (s.o.), bietet es sich an, Reduktionsstufen mit Hilfe von Vergleichsexperimenten zu ermitteln.

## 5.2 Datenmaterial

Weiterhin stellt sich die Frage nach geeignetem Versuchsmaterial. Davon ausgehend, dass die Qualität der Stimuli die Perzeption beeinflusst, folgern van Hessen und Schouten (1999:71 f): *„it is important to aim for stimuli that are as natural as possible, if we want to learn anything about the perception of speech.“*

Um möglichst natürliche Stimuli zu erhalten, sollen als Testmaterial gesprochene Äußerungen genutzt werden. Eine häufig verwendete Methode bei Untersuchungen zur Sprachperzeption ist, eigens für die Experimente Material zu erstellen, das sich in nur bestimmten Variablen (z.B. Sprechgeschwindigkeit, Betonung, Kontext) unterscheidet. Der Vorteil ist hierbei, dass möglichst viele Störfaktoren ausgeschaltet werden, die die Ergebnisse beeinflussen können. Dieser Methode wird häufig eine gewisse Künstlichkeit angelastet. Eine andere Methode ist, Äußerungen, die unter nicht kontrollierten Bedingungen (z.B. Interviews, Reportagen, Nachrichtensendungen) aufgenommen wurden, als Versuchsmaterial zu verwenden. Der Vorteil liegt in einem Höchstmaß an Natürlichkeit des Materials; nachteilig ist, dass viele Faktoren – gewollte und ungewollte – in die Ergebnisse einfließen und deren Vergleichbarkeit erschweren.

Die Erforschung von Vokalreduktion ermöglicht Einblicke in grundlegende Sprachperzeptionsprozesse und ihre Ergebnisse sind auch für die Sprachanwendungsforschung (z.B. Spracherkennung und –synthese) von Interesse. Für die Anwendungsforschung ist es unerlässlich, große Sprachdatenkorpora hinsichtlich ihrer lautlichen Zusammensetzung zu segmentieren und zu etikettieren, was im allgemeinen personen- und zeitintensiv ist. Eine Voraussetzung für eine automatische Etikettierung von Reduktionen ist ihre reliable Beurteilung.

Für die Untersuchung der Einteilung von Reduktion in Stufen wird ein Datenmaterial genutzt, das auch in der Anwendungsforschung verwendet wird. Hierbei handelt es sich um die am Institut für Kommunikationsforschung und Phonetik (IKP) in Bonn entwickelte Bonner Prosodische Datenbank (BPD; Heuft et al., 1995; Heuft, 1999). Die Stimuli werden aus unterschiedlichen phonetischen Kontexten herausgeschnitten und werden allein und in ihrem silbischen Kontext präsentiert. Diese vom Wortkontext isolierte Präsentation der Stimuli entspricht einer beim Etikettieren von Korpora gebräuchlichen Vorgehensweise und unterstützt analytisches Hören und somit die Vermeidung von Kompensationsprozessen.

### 5.2.1 Beschreibung des Datenmaterials

Die Bonner Prosodische Datenbank (BPD; Heuft et al., 1995; Heuft, 1999) beinhaltet von drei Sprechern (zwei Frauen, ein Mann) gelesene Aussagesätze, Befehle, Fragen mit Antworten und

kurze Geschichten. Eine der Sprecherinnen (SPR1) hatte bereits Erfahrungen mit Sprachaufnahmen für ein Sprachsynthesystem und kann als professionelle Sprecherin betrachtet werden. Die Äußerungen wurden von Hand segmentiert und phonetisch etikettiert (SAMPA-Konvention; Wells, 1996). Da die phonetische Transkription größtenteils mit der kanonischen Verschriftung übereinstimmt, werden im folgenden die in der Datenbank etikettierten Vokale in Schrägstriche (/) gesetzt, die bei einer phonematischen (breiten) Transkription gebräuchlich sind. Die bei der Beurteilung von Reduktion perzipierten Vokalqualitäten sind durch eckige Klammern ([ ]) gekennzeichnet.

Reduktion ist als Verlust der Vokalqualität definiert (s. Abschnitt 2.3.1 und 4.2). Akustische Korrelate der Vokalqualität sind primär die Formanten, insbesondere die ersten beiden (Pols et al., 1969). Jedoch weisen Perzeptionsexperimente darauf hin, dass auch die Grundfrequenz und die Dauer die Vokalqualität beeinflussen (vgl. Kapitel 3). Weiterhin stehen Reduktion und Dauer in einer engen Beziehung, wie die akustischen Untersuchungen von Lindblom (1963) belegen. Reduktionen zeigen ferner einen Zusammenhang zur Betonung auf und die Betonung ihrerseits zur Grundfrequenz (vgl. Abschnitt 2.3). Es wird daher davon ausgegangen, dass Formanten, Grundfrequenz und Dauer von Vokalen durch Reduktion beeinflusst werden und somit potentielle akustische Korrelate der Perzeption von Vokalreduktion sind.

Spektrale Vokalreduktionen werden häufig durch einen Wert pro Formant und Laut beschrieben. Jedoch zeigen sich erhebliche Unterschiede in der Bestimmung des Messpunktes in verschiedenen experimentalphonetischen Studien zur Vokalreduktion (vgl. Lindblom, 1963; Koopmans-van Beinum, 1980). Nach van Son und Pols (1990) werden fünf Methoden zur automatischen Extraktion von Formantwerten unterschieden:

- Messung von Formantwerten in der Mitte des Vokals (*Zentrumsmethode*),
- Berechnung eines durchschnittlichen Formantwerts über alle Werte eines Formanten innerhalb der Segmentgrenzen (*Durchschnittsmethode*),
- Messung des Formanten am Punkt mit maximaler Energie (*Energiemethode*),
- Bestimmung der maximalen oder minimalen Werte von Formanten in Abhängigkeit von der Vokalkategorie in einem bestimmten Zeitintervall (*Formantmethode*),
- Messung eines Formanten im stationären Teil des Vokals (*stationäre Methode*).

Untersuchungen im Deutschen zeigen, dass sich Vokalreduktionen vor allem im mittleren Teil von Vokalen auswirken (vgl. Abschnitt 2.3.1), so dass hier die *Zentrumsmethode* gewählt wurde. Die Formantfrequenzen wurden durch LPC-Analysen (*Linear Predictive Coding*) mit 10 Prädiktorkoeffizienten über ein Parametermessintervall (*frame*) von 20 ms alle 5 ms berechnet. Verwendet wurde das Programm *formant* von ESPS 5.0; die Abtastfrequenz der Sprachsignale (16 kHz) wurde auf 8 kHz herabgesetzt. Mit Hilfe einer Polynomfunktion dritten Grades wurden die Formantverläufe abgeschätzt, wobei die Methode der kleinsten Fehlerquadrate genutzt

wurde (Stöber, 1997a). Das Polynom stellt somit eine Ausgleichsfunktion des Formantverlaufs dar und wurde in der Mitte des Vokals ausgewertet. Gegenüber einer direkten Auswertung der Formantwerte hat die Auswertung des Polynoms den Vorteil, dass aufgrund der Glättung des Formantenverlaufs mögliche Ausreißer bei der automatischen Extraktion der Formanten an dieser Stelle nicht so stark ins Gewicht fallen.

Die Formantwerte wurden zusätzlich in die mel-Skala (Abschnitt 3.1) überführt unter Verwendung folgender Gleichung (Makhoul und Cosell, 1976; Stöber, 1997b):

$$\frac{m}{mel} = 2595 \lg \left( 1 + \frac{f / \text{Hz}}{700} \right) \quad (1)$$

Wie schon bei der Bestimmung der Formantwerte wurden die Grundfrequenzwerte nicht direkt ausgewertet, sondern durch eine Approximation ermittelt. Die Berechnung der Grundfrequenzwerte beruht auf einem von Stöber (1996) entwickelten Programm, das mit Hilfe eines Extremwert-Detektors die potentiellen Anregungszeitpunkte des Sprachsignals bestimmt. Eine Gerade, die die Anregungszeitpunkte eines Lautes verbindet, wurde nach der Methode der kleinsten Quadrate bestimmt. Der Grundfrequenzwert ergibt sich aus der Auswertung der Geradengleichung in der Mitte eines jeden Lautes.

Um nicht nur einzelne Energieanreicherungen im spektralen Bereich zu betrachten, werden zusätzlich Energieverteilungen in verschiedenen Frequenzbändern (0-1 kHz, 1-2 kHz, 2-4 kHz, 4-8 kHz) hinzugenommen. Die Energiewerte wurden durch Berechnungen von Fourierspektren und Messungen der Kurzzeitenergie in den vier Energiebändern bestimmt. Alle Energiewerte in

<i>Akustische Parameter</i>	<i>Beschreibung</i>
F1	1. Formant
F2	2. Formant
F3	3. Formant
F1_mel	mel-skalierte Werte des 1. Formanten
F2_mel	mel-skalierte Werte des 2. Formanten
F3_mel	mel-skalierte Werte des 3. Formanten
E1NORM	bezüglich der Gesamtenergie eines Lautes normierte Energiewerte im Frequenzband von 0-1 kHz,)
E2NORM	bezüglich der Gesamtenergie eines Lautes normierte Energiewerte im Frequenzband von 1-2 kHz
E3NORM	bezüglich der Gesamtenergie eines Lautes normierte Energiewerte im Frequenzband von 2-4 kHz
E4NORM	bezüglich der Gesamtenergie eines Lautes normierte Energiewerte im Frequenzband von 4-8 kHz
LAUTDUR	Lautdauer
F0	Grundfrequenz

**Tabelle 3: Abkürzungen der akustischen Parameter und ihre Bedeutung.**

den Frequenzbändern sind hinsichtlich der Gesamtenergie der Laute normiert (vgl. Campbell, 1995). Die für die Untersuchung relevanten akustischen Parameter sind in Tabelle 3 aufgelistet.

Wie bereits in Abschnitt 2.3 beschrieben, gibt es eine Beziehung von Vokalreduktion und Betonung einerseits sowie Vokalreduktion und Wortart bzw. -häufigkeit andererseits; gleichzeitig interagieren aber auch Wortart bzw. -häufigkeit und Betonung miteinander. Die BPD verfügt über eine detaillierte Beschreibung der Betonung. Jede Silbe ist mit der wahrgenommenen Prominenz etikettiert. Die Beurteilung der Prominenz basiert auf dem Ansatz von Fant und Kruckenberg (1989). Drei Phonetiker beurteilten die Stärke der Prominenz von Silben relativ zu den umliegenden Silben auf einer Skala von 0 bis 31. Untersuchungen zeigen eine hohe Übereinstimmung zwischen den Probandenurteilen (Heuft und Portele, 1996). Die wahrgenommene Prominenz einer Silbe ergibt sich aus dem Median der Prominenzwerte. Weiterhin ist jede Silbe mit weiteren Informationen hinsichtlich der lexikalischen Wortbetonung gekennzeichnet. Jedes Wort erhält eine Wortbetonung, als grundsätzlich nicht betonbar gelten Silben mit /ə/ und Silben, die einen silbischen Konsonanten als Nukleus besitzen (Heuft, 1999). Silben, die ein F0-Maximum aufweisen, werden als akzentuiert bezeichnet. Die F0-Maxima wurden manuell etikettiert (Heuft, 1999).

Die Wortarteninformationen der BPD wurden aus der am IKP erstellten 'Bonner Wortdatenbank' (Heß et al. 1983; Brustkern, 1992) entnommen. 21 Wortarten werden unterschieden, die zu den Kategorien Inhaltswort vs. Funktionswort zusammengefasst wurden (Tabelle 4). Da die Untersuchung von Reduktion eine von der phonetischen Realisierung unabhängige Definition erfordert, erfolgt die Einteilung in Inhalts- bzw. Funktionswörter primär nach den lexikalischen bzw. grammatischen Funktionen der Wortarten (s. Abschnitt 2.3).

Welche Funktion eine Wortart besitzt, hängt stark von dem Kontext der Äußerung ab. Beispielsweise verlieren einige Verben (z. B. bringen, kommen) ihre lexikalische Funktion, wenn sie als Bindeglied zwischen Subjekt und präpositionalem Objekt auftreten und somit eine syntaktisch-morphologische Funktion tragen, z. B. *zur Aufführung bringen* anstelle von *aufführen* (Bußmann, 1990:259 f). Bei der hier verwendeten Definition von Funktions- vs. Inhaltswörtern wird der Fokus auf ihre primäre Funktion gelegt. So besitzen Präpositionen, Partikel, Konjunktionen, Artikel und Pronomen eine primär grammatische Funktion (Eisenberg, 1989:34). Partikel stehen als Bezeichnung für nicht flektierende Wörter, die keine oder wenig ausgeprägte lexikalische Bedeutung tragen, jedoch die Bedeutung des in Beziehung stehenden Worts modifizieren, wie z. B. Interjektionen, Affirmationen, Modalwörter (vgl. Bußmann, 1990:562). Die Zuordnung der Numeralia bereitet insofern Probleme, da ihr Status als eigene Wortart an sich umstritten ist (vgl. Eisenberg, 1989:166; Bußmann, 1990:537). Sie können aufgrund des syntaktisch-morphologischen Aspekts den Artikeln, Adjektiven, Adverbien, Substantiven oder Pronomina zugeordnet werden. Eine einheitliche Gruppe bilden die Numeralia aus semantischer Sicht, sie geben Quantitäten wieder und werden daher auch als *Quantoren* bezeichnet.

<i>Wortarten</i>
Präpositionen; z.B. vor, auf, in, zur
Artikel; indefinite (z.B. ein, keiner) und definite (z.B. der, die, das) Artikel, attributive Demonstrativpronomina (z.B. dieser, jener)
Interjektionen; z.B. ach so, bitte
Konjunktionen; z.B. und, auch, aber
Modalwörter; z.B. schon, eigentlich, mal
<b>Eigennamen</b> ; z.B. Personennamen wie Ingo, Brahms; Ortsnamen wie Bonn, Münster
Negation und Affirmation; z.B. nicht, nein, ja
<b>Nomina</b>
<b>Numeralia</b> ; z.B. drei, beide, einmal
Pronomina; wie Possesivpronomina (z.B. mein, unser), Personalpronomina (z.B. er, wir, uns, ihm), Relativpronomina, Fragewörter (z.B. wer, was, wo), Reflexivpronomen (sich)
<b>Attributive Adjektive</b> ; z.B. frische Gardinen, bunte Tulpen
<b>Prädikative Adjektive</b> ; z.B. die Ärzte sind damit gar nicht einverstanden, die Kopfhörer sind angeschlossen
<b>Prädikativ und adverbial verwendbare Adjektive</b> (Adverbien); z.B. leise rollen wir aus den Bahnhof, die Bremsen quietschen gräßlich
'Reine' Adverbien (nicht flektierbar), z.B. jetzt, heute
Subjunktionen; z.B. dass, wenn, weil
Präverbien (Tmesis); die abtrennbaren Präfixe von Verben, z.B. 'auf' in aufgeben
Verbformen von 'haben'; z.B. habe, hatte
Modalverben; z.B. können, sollen
Verbformen von 'sein', z.B. ist, bin, war
Verbformen von 'werden'; z.B. würde, wird, geworden
<b>Vollverben</b>

**Tabelle 4: In der BPD verwendete Wortarten mit Beispielen. Fettgedruckte Wortarten sind zur Klasse der Inhaltswörter zusammengefasst.**

Aufgrund ihrer primär lexikalischen Bedeutung werden hier Eigennamen, Nomina, Adjektive (attributive, prädikative sowie prädikativ und adverbial verwendbare) und Vollverben den Inhaltswörtern zugeordnet. Da in dem vorliegenden Datenmaterial die Numeralia nach ihrer Funktion in Artikel bzw. Pronomina, z.B. ein, kein, dieser, jener, eingeteilt sind, werden sie den Inhaltswörtern zugeordnet. Die anderen Wortarten werden als Funktionswörter betrachtet.

### 5.2.2 Normierung der Daten

Vokale verschiedener Sprecher weisen im allgemeinen unterschiedliche akustische Realisierungen auf. Dies ist durch physiologische Unterschiede, aber auch durch linguistische Faktoren (vgl. Abschnitt 2.3) bedingt. Ziel ist, die für die Untersuchung nicht relevanten Eigenschaften herauszufiltern, so dass nur die invarianten Eigenschaften erhalten bleiben, aufgrund derer Hörer Vokale unterscheiden. In Modellen zur Vokalperzeption werden die Hz-skalierten Formantwerte üblicherweise in die Bark- oder mel-Skala überführt (Abschnitt 3.2). Eine weiterführende Methode ist die Parametrisierung eines Sprachsignals nach dem Perzeptionsmodell (PEMO) von Dau et al. (1996). Durch PEMO wird die im peripheren auditorischen System stattfindende Signalverarbeitung nachgebildet. Die Umwandlung beruht auf der Annahme, dass die interne



Repräsentation eines akustischen Signals durch einen Informationsverlust gekennzeichnet ist und umfasst drei Schritte:

1. Nachbildung der Frequenz-Ort-Transformation auf der Basilarmembran,
2. Modellierung von Adaptationseffekten,
3. Datenreduktion der erhaltenen internen Repräsentation.

PEMO wurde entwickelt, um das Antwortverhalten von Probanden in psychoakustischen Experimenten, z.B. in Versuchen zur zeitlichen oder spektralen Maskierung zu simulieren.

Eine andere Möglichkeit sprecherspezifische Einflüsse zu eliminieren, bietet die Normierung der Hz-skalierten Formantwerte hinsichtlich ihrer Verteilung oder ihres Mittelwerts. Es muss aber gewährleistet sein, dass die linguistische Validität erhalten bleibt und die Untersuchungsergebnisse nicht zu Artefakten der Transformation werden. Disner (1980) überprüfte unterschiedliche Methoden zur Transformation der Formantwerte mit Datenmaterial aus verschiedenen Sprachen. Es zeigt sich, dass die Güte der Transformationen von den Fragestellungen abhängt. Techniken, die auf eine Normierung des Mittelwerts und Standardabweichung beruhen, sind geeignet, um die Varianz zwischen den Sprechern innerhalb einer Sprache effektiv zu reduzieren.

Um die akustischen Parameter der einzelnen Sprecher des hier verwendeten Datenmaterials vergleichbar zu machen, werden die Werte der jeweiligen akustischen Parameter über alle Monophthonge eines jeden Sprechers nach folgender Formel (z-Transformation) normiert:

$$z_i = \frac{x_i - \bar{x}}{s} \quad (2);$$

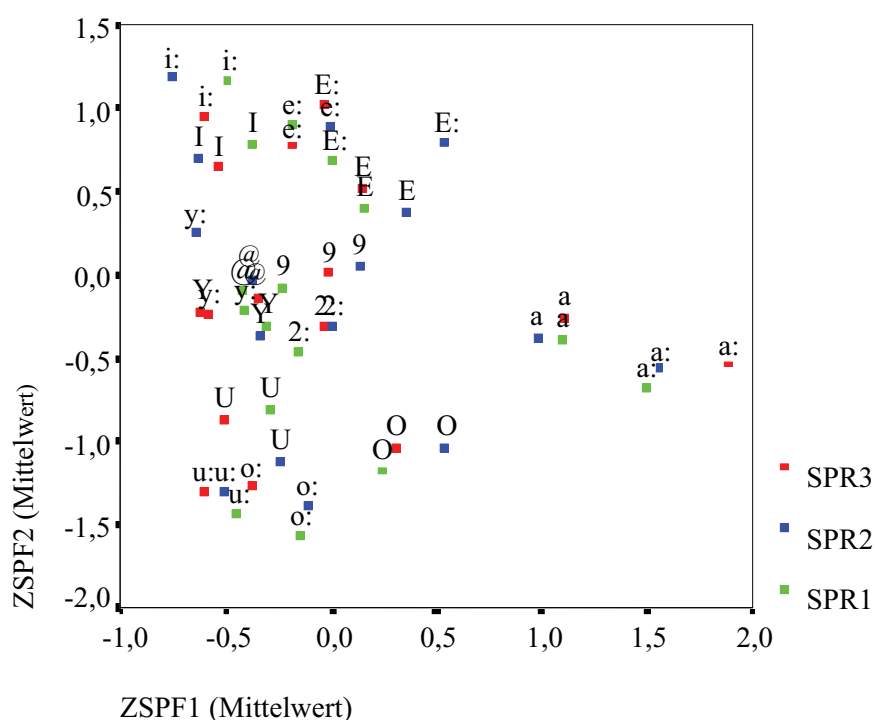
wobei  $z_i$  für den normierten Wert des jeweiligen akustischen Parameters eines Sprechers steht,  $x_i$  für den gemessenen Wert,  $\bar{x}$  für den Mittelwert und  $s$  für die Standardabweichung der gemessenen Werte.

Durch die z-Transformation werden die Mittelwerte der jeweiligen akustischen Parameter der Monophthonge eines Sprechers auf 0 gesetzt und die Standardabweichungen auf 1 (vgl. Bortz, 1979). Dies entspricht der von Disner (1980) verwendeten Normierungsmethode. In Anhang A sind Anzahl, Mittelwert und Standardabweichung für die akustischen Parameter eines jeden Sprechers in Abhängigkeit der Vokalkategorie aufgeführt. Die Normierung wird anhand der Werte von F1 und F2 der Vokale der drei Sprecher veranschaulicht. In Abbildung 2 sind die Mittelwerte der Formanten für die Monophthonge der einzelnen Sprecher aufgetragen.

Ein Vergleich der Formantwerte zeigt sehr deutlich einen geschlechtspezifischen Unterschied. Die F1- und F2-Werte des Sprechers (SPR2) weisen eine geringere Spannweite auf als die der







**Abbildung 3: Mittelwert der sprechernormierten Formantwerte (ZSPF1 und ZSPF2) für die Vokalkategorie (SAMPA-Notation) einzelner Sprecher (SPR1, SPR2, SPR3).**

erkennen, dass die Transformation die geschlechtsspezifischen Unterschiede hinsichtlich der Lage der Formanten aufhebt. Die mittleren Formantwerte einzelner Vokalkategorien der Sprecher stimmen zwar nicht überein, weisen aber innerhalb einer Vokalkategorie kleinere Distanzen zueinander auf und haben zwischen den Vokalkategorien der Sprecher weniger Überschneidungen als die nicht normierten Formantwerte. Die Formantlage von /ə/ ist für alle Sprecher vergleichbar, und die Formantwerte liegen im Bereich des Mittelwerts aller Vokale, so dass der Neutralvokal als Artikulationsbasis der Vokale bzw. als Zentralvokal betrachtet werden kann (vgl. Abschnitt 2.2).

Der Vergleich der Formantwerte mit den normierten Werten zeigt, dass durch die Normierung die physiologischen Unterschiede zwischen den Sprechern ausgeglichen werden; idiosynkratische Eigenschaften bleiben erhalten (vgl. Heid, 1997). Diese Eigenschaften werden eliminiert, wenn die akustischen Parameter hinsichtlich der Sprecher und Vokalkategorie normiert werden, so dass der Mittelwert für jede Vokalkategorie des jeweiligen Sprechers bei 0 liegt und die Standardabweichung bei 1.

Die in Kapitel 7 und 9 referierten Ergebnisse beziehen sich auf die in physikalischen Maßeinheiten vorliegenden Parameter. Die sprechernormierten Parameter sind durch das Präfix ZSP gekennzeichnet (z.B. ZSPF1), die sprecher- und lautnormierten durch Z (z.B. ZF1). Somit wird zum einen vermieden, dass einige Parameter (Formanten) in perceptiven Einheiten vorliegen und andere (z.B. Energieverteilungen in den Frequenzbändern) durch physikalische Einheiten beschrieben werden. Zum anderen weisen Voruntersuchungen im Rahmen der akustischen

Analysen (Abschnitt 7.1) sowie zur automatischen Klassifikation (Kapitel 9) auf keine verbesserte Trennbarkeit der hier verwendeten Daten hin, wenn die Formantwerte in die mel- oder Bark-Skala übergeführt werden. Dies entspricht den Ergebnissen von Untersuchungen zur Klassifikation von Phonemen (Hillenbrand und Gayvert, 1993), bei denen Grundfrequenz- und Formantwerte verwendet wurden, die in linearen physikalischen sowie in verschiedenen nicht linearen auditiven Maßeinheiten (Bark- mel- und Koenig-Skala) vorlagen. Auch weisen Voruntersuchungen zur Bestimmung eines Ähnlichkeitsmaßes von Vokalen (Abschnitt 7.3), bei denen Parameter genutzt wurden, die auf dem Perzeptionsmodell PEMO (Dau et al., 1996) basieren, auf keine Vorteile gegenüber den physikalischen Maßeinheiten hin.

### 5.3 Verwendete statistische Verfahren

Die in dieser Arbeit verwendeten statistischen Verfahren basieren in den meisten Fällen auf parameterfreien Tests. Hierbei handelt es sich um in Bezug auf die Verteilung (z.B. Normalverteilung) voraussetzungsfreie Verfahren, die lediglich ordinale Informationen der Daten nutzen. Die jeweiligen Stichproben werden hinsichtlich ihrer zentralen Tendenz ausgewertet. Die parameterfreien Verfahren wurden hier aus folgenden Gründen ausgewählt (vgl. Bortz, 1979):

- Daten sind ordinal skaliert, wie es z.B. bei den Reduktionsstufen der Fall ist,
- Daten sind intervallskaliert, z.B. die akustischen Parameter der Reduktionsstufen, aber ihre Stichprobengröße ist zu klein, so dass keine normalverteilte Grundgesamtheit oder ggf. Varianzhomogenität besteht,
- Daten sind intervallskaliert, aber die zu vergleichenden Stichproben besitzen zu unterschiedliche Umfänge.

Zum Vergleich zweier unabhängiger Stichproben hinsichtlich ihrer zentralen Tendenz wird der U-Test nach Mann-Whitney verwendet. Der U-Test ist das verteilungsfreie Pendant des T-Tests. Bei unterschiedlicher Stichprobengröße bezieht sich der U-Wert auf die kleinere Stichprobe. Die Ergebnisse werden in den jeweiligen Anhängen präsentiert. Die Wilcoxon-Rangsummen-W-Statistik ist die Rangsumme der kleineren Stichprobe (Brosius und Brosius, 1995; Bühl und Zöfel, 1998). Zur Überprüfung der Zusammenhänge zwischen zwei Parametern wird die Rangkorrelation nach Spearman ( $\rho$ ) berechnet.

Das in Kapitel 9 verwendete Datenmaterial erlaubt einen Zugriff auf parametrische Tests. Durch einfaktorielle Varianzanalysen wird geprüft, ob mehrere Mittelwerte gleich sind. Die Varianzanalyse ist eine Erweiterung des T-Tests. Sie ist gegenüber Abweichungen von der Normalverteilung relativ robust, wenn die zu vergleichenden Stichproben hinreichend groß ( $n > 15$ ) sind. Abweichungen von der Varianzhomogenität führen bei vergleichbaren Stichprobenumfängen zu keinen fehlerhaften Entscheidungen (vgl. Bortz, 1979:347 f).

Zur Feststellung von Häufigkeitsunterschieden werden  $\chi^2$ -Methoden verwendet. Bei Vergleichen der Häufigkeiten eines zwei- oder mehrfachgestuften Merkmals wird der eindimensionale  $\chi^2$ -Test berechnet, bei 2x2-Tabellen der exakte Test nach Fisher bzw. das korrigierte  $\chi^2$  nach Yates, bei mehrfachgestuften Merkmalen  $\chi^2$  nach Person (Brosius und Brosius, 1995; Bühl und Zöfel, 1998). Unterschiede zwischen einzelnen Stichproben werden als signifikant betrachtet, wenn die Irrtumswahrscheinlichkeit unterhalb von 5% liegt ( $\alpha \leq 0,05$ ). Abkürzungen für statistische Kennwerte sind  $n$  für Stichprobenumfang,  $z$  für Median,  $\bar{x}$  für arithmetischen Mittelwert und  $s$  für Standardabweichung. Der Freiheitsgrad wird durch  $df$  gekennzeichnet.



## 6 PERZEPTIONSEXPERIMENTE

In diesem Kapitel wird die Perzeption von Vokalreduktionen untersucht. Da durch Reduktion die Qualität der Vokale beeinflusst wird, wird überprüft, wie viele Lautqualitäten innerhalb einer Vokalkategorie von Hörern zuverlässig perzipiert werden (Abschnitte 6.1 und 6.2). Um mögliche Kompensationsstrategien auszuschließen, die zu einer Wahrnehmung phonologischer Vokalqualitäten führen, werden Vergleichsexperimente durchgeführt. Mit Hilfe von Clusterzentrenanalysen werden Referenzstimuli bestimmt, die als mögliche Prototypen von Reduktionsstufen betrachtet werden. Prototypen dienen im allgemeinen zur Erklärung der Kategorisierung von Einheiten. Nach der Theorie des Perzeptiven Magneteffekts (vgl. Abschnitt 3.3) besitzen Prototypen „magnetische“ Eigenschaften, so dass akustische Unterschiede zwischen den Lauten und dem Prototypen einer Kategorie perzeptiv unbedeutsam sind. Hingegen weisen nicht-prototypische Instanzen einer Kategorie keine „magnetischen“ Eigenschaften auf. Durch Zuordnung von Lautrealisierungen zu den Referenzstimuli aufgrund ihrer perzipierten qualitativen Ähnlichkeit wird ermittelt, welche Referenzstimuli geeignet sind, die Lautrealisierungen zuverlässig zu klassifizieren. Diese Referenzstimuli werden hinsichtlich ihrer Lautqualität und Reduktionsstärke beurteilt. Ihre Relevanz für die Einteilung von Lautrealisierungen in Reduktionsstufen wird aufgrund der Ergebnisse der Vergleichsexperimente geprüft.

### 6.1 Reduktionsstufen gespannter Vokale

In diesem Abschnitt werden die Perzeptionsexperimente zur Auffindung der Reduktionsstufen für die gespannten Vokale dargestellt. Der Aufbau ist folgender: Zuerst wird die Ermittlung der Referenzstimuli (Abschnitt 6.1.1) und der Versuchsaufbau (Abschnitt 6.1.2) beschrieben, dann werden die Ergebnisse der Vergleichsexperimente ausgewertet, durch die die Anzahl der Prototypen für einzelne Vokalqualitäten bestimmt wird (Abschnitt 6.1.3). Weiterhin werden die Ergebnisse der Beurteilungen der Lautqualität und Reduktionsstärke der Prototypen präsentiert (Abschnitt 6.1.4). Sowohl die Beurteilungs- als auch die Vergleichsexperimente werden hinsichtlich der inter- und intrasubjektiven Reliabilität ausgewertet. Im Vordergrund steht die Frage, ob die Referenzstimuli als Prototypen von Reduktionsstufen betrachtet werden können (Abschnitt 6.1.5). Zusätzlich wird überprüft, ob die Prototypen auch für die Einteilung von Vokalrealisierungen anderer Sprecher genutzt werden können (Abschnitt 6.1.6).

#### 6.1.1 Ermittlung potentieller Prototypen

Da die ersten beiden Formanten in erster Linie die Qualität von Vokalen bestimmen, wurden zur Abschätzung der Anzahl von möglicherweise zu diskriminierenden Reduktionsstufen die Formantwerte der gespannten Vokale der professionellen Sprecherin (SPR1) mehreren

Clusterzentrenanalysen unterzogen. Die Anzahl der Cluster variiert von zwei bis sieben. Die Analysen wurden für jede Vokalkategorie einzeln durchgeführt.

Durch die Clusteranalysen werden Vokalrealisierungen mit ähnlichen Formantwerten in Gruppen zusammengefasst, die Werte der verschiedenen Cluster sollen dabei möglichst unterschiedlich sein. Die Ähnlichkeit wird dabei über dem einfachen Euklidischen Abstand definiert:

$$d = \sqrt{\sum (x_i - y_i)^2} \quad (3);$$

wobei  $x_i$  für den Wert eines akustischen Parameters einer Vokalrealisierung steht,  $y_i$  für den Wert des gleichen akustischen Parameters einer zu vergleichenden Realisierung und  $d$  für die berechnete Euklidische Distanz über die Summe der Differenzen aller akustischen Parameter. Die verwendeten Formantwerte sind mel-skaliert sowie sprecher- und lautnormiert. Durch die Normierung werden Einflüsse der unterschiedlichen Dimensionen der Formantwerte auf die Distanzwerte vermieden.

Der Euklidische Abstand ist ein gebräuchliches Maß zur Quantifizierung von Ähnlichkeit. Er wird nicht nur zur Abstandsmessung im physikalischen Raum genutzt, sondern auch zur Ähnlichkeitsmessung im psychologischen Raum, beispielsweise bei der Multidimensionalen Skalierungsmethode (MDS). Die Ähnlichkeit zwischen zwei Objekten steht in inverser Beziehung zu ihrer Distanz (Spencer-Smith und Goldstone, 2001).

In einer explorativen Studie wurden die qualitativen Unterschiede von Vokalrealisierungen innerhalb einer Vokalkategorie im gleichem phonetischen Kontext von einer Probandin perzeptiv beurteilt. Die Stimuli bestanden aus Silben, die aus dem Datenmaterial herausgeschnitten wurden. Abhängig vom Datenmaterial wurden mindestens zwei Stimuli, die im gleichen phonetischen Kontext stehen, aber aus unterschiedlichen Clustern kommen, über Kopfhörer präsentiert. Die Probandin hatte die Aufgabe, die Stärke der Reduktion der Vokale zu beurteilen (offene Antwortform). Wenn die perzipierte Vokalqualität nicht mit der in der Datenbank etikettierten Qualität übereinstimmte, wurde dies vermerkt. Insgesamt wurden 841 (/i:/: n = 210, /e:/: n = 128, /ɛ:/: n = 37, /y:/: n = 46, /ø:/: n = 35, /a:/: n = 183, /u:/: n = 86, /o:/: n = 116) Vokalrealisierungen beurteilt.

Die Beurteilung der Vokalqualitäten zeigt, dass in Abhängigkeit der Vokalkategorie zwischen drei bis sechs Reduktionsstufen (nicht reduziert bis stark reduziert ([ə])) perzipiert werden, hinzu kommen die von der intendierten Vokalqualität abweichend wahrgenommenen Qualitäten. Aufgrund dieser Ergebnisse wird vorerst von sieben Clustern pro Vokal ausgegangen, mit Ausnahme der Vokale /e:/, /ɛ:/ und /ø:/, deren Qualitäten besser durch je sechs Cluster beschrieben werden können.



Diese Cluster bilden die Grundlage für die weiteren Perzeptionsexperimente zur Überprüfung, ob und gegebenenfalls wie viele Reduktionsstufen von Hörern zuverlässig diskriminiert werden. Aufgrund der Voruntersuchung wird angenommen, dass sich die Vokalrealisierungen einzelner Cluster perzeptiv unterscheiden und diese mögliche Reduktionsstufen bilden. Aus jedem Cluster wurde daher ein Laut ausgewählt, der diese potentiellen Reduktionsstufen präsentiert, d.h. eine große Ähnlichkeit zu allen anderen Lauten eines Clusters besitzt bzw. eine kleine durchschnittliche Distanz zu den Lauten innerhalb eines Clusters aufweist. Dies entspricht einem Vokal, dessen Formantwerte nahe an denen eines Clusterzentrums (Mittelwert von F1 und F2 über alle Laute eines Clusters) liegen. Der Abstand zwischen Formantwerten und Clusterzentrum wurde folgendermaßen berechnet:

$$dz = (\bar{x}_{F1} - F1)^2 + (\bar{x}_{F2} - F2)^2 \quad (4);$$

wobei  $\bar{x}_{F1}$  bzw.  $\bar{x}_{F2}$  für den Mittelwert von F1 bzw. F2 über alle Vokale eines Clusters stehen und F1 bzw. F2 für den Wert eines Vokals des selben Clusters.

Damit der phonetische Kontext die Wahrnehmung dieser in der Clustermitte liegenden Vokale nicht zu stark beeinflusst, wurden die Vokale aus ihrem Kontext herausgeschnitten. Dabei wurde darauf geachtet, dass der Laut noch als vokalisches Segment wahrnehmbar ist (vgl. Abschnitt 2.2). War dies nicht der Fall, so wurde der zum Clusterzentrum nächstliegende Vokal gewählt. Die Prototypen wurden von einem Phonetiker überprüft. Alle Prototypen werden als Vokale wahrgenommen und unterscheiden sich hinsichtlich ihrer Qualität (s. Anhang B, Tabelle B-1). Die Hypothese, dass diese Laute prototypische Vertreter einzelner Reduktionsstufen sind, wurde mit folgender Methode getestet.

### 6.1.2 Experimentdesign

Die Perzeptionsexperimente wurden für jede Vokalkategorie einzeln durchgeführt. Die erste Aufgabe der Probanden war, einzelne Stimuli den prototypischen Vertretern der Reduktionsstufen aufgrund ihrer qualitativen Ähnlichkeit zuzuordnen (Vergleichsexperimente). Die zweite Aufgabe bestand in der Beurteilung der Prototypen hinsichtlich ihrer Reduktionsstärke und ihrer Vokalqualität (Beurteilungsexperimente).

Da nach Vieregge (1989:23) gefolgert werden kann, „*daß das analytische Hören eine heuristisch-praktische Fertigkeit ist, die bis zu einem gewissen Maß erlernt werden kann*“ und die Aufgabenstellung schon gewisse Kenntnisse im analytischen Hören verlangt, wurden phonetisch geschulte Hörer ausgewählt. Während für die Vergleichsexperimente eine gewisse Übung im analytischen Hören erforderlich ist, um ungeachtet von Dauer und Lautstärkeunterschieden eine differenzierte Einteilung der Vokalrealisierungen hinsichtlich ihrer Qualität zu erhalten, sind phonetische Kenntnisse für die Beurteilung der Vokalqualität und Reduktionsstärke der Prototypen unverzichtbar. Sind jedoch einmal die Prototypen der einzelnen Reduktionsstufen ermittelt, so ist

es durchaus denkbar, dass auch die Diskriminierbarkeit reduzierter Vokalqualitäten von phonetisch nicht geschulten Hörern nach einer Trainingsphase mittels der Vergleichsexperimente getestet werden kann. Die Probanden, die für die Perzeptionsexperimente gewählt wurden, entstammen alle dem Mitarbeiterkreis des IKP und sind mit dem Transkribieren natürlichsprachlicher Äußerungen vertraut.

In den Vergleichsexperimenten wurden alle Prototypen-Stimuli und jeweils ein zuzuordnender Stimulus einer Vokalkategorie als Etiketten (Nummern) auf dem Bildschirm eines PC präsentiert. Die Stimuli wurden mit und ohne Silbenkontext dargeboten. Die Probanden hörten Stimuli und Prototypen-Stimuli über Kopfhörer und konnten sie beliebig oft miteinander vergleichen. Die Nummer des als ähnlich perzipierten Prototypen-Stimulus wurde über die Tastatur des Computers eingegeben.

Ob ein Prototyp-Stimulus „magnetische“ Eigenschaften besitzt und sich somit zur Klassifizierung von Vokalrealisierungen eignet, wurde durch einen Vergleich (Vertauschungsmatrix) der individuellen Beurteilungen mit den mehrheitlichen Urteilen der Probanden (Modalwert der individuellen Beurteilungen) überprüft (Abschnitt 6.1.3 und 6.2.2). Die Prototypen, die sich zur Klassifizierung von Vokalrealisierungen eignen, wurden hinsichtlich ihrer Reduktion und ihrer Qualität beurteilt.

In den Beurteilungsexperimenten wurden nur noch die Prototypen-Stimuli präsentiert. Die Probanden wurden gebeten, die Prototypen entsprechend ihrem Reduktionsgrad von nicht reduziert bis stark reduziert in eine Rangfolge zu bringen. Die Anzahl der Reduktionsstufen stimmt mit der Anzahl der Prototypen überein. Weiterhin hatten die Probanden die Aufgabe, die Lautqualität der Prototypen mit Hilfe der IPA-Notation unter Verwendung der Diakritika genau zu beschreiben. Die Urteile wurden auf Papier festgehalten. Da nicht die Beherrschung der Transkriptionszeichen getestet werden sollte, wurde eine Aufstellung mit den IPA-Symbolen beigelegt.

Die Stärke der Reduktion der Prototypen-Stimuli wird aufgrund des mehrheitlich perzipierten Reduktionsgrads (Modalwert) bestimmt. Die Ähnlichkeitsbeziehung zwischen Prototyp und Stimulus erlaubt die Annahme, dass jeder Stimulus eine vergleichbare Lautqualität besitzt wie der Prototyp selbst und die Stimuli dieselbe Reduktionsstufe mit dem Prototypen teilen. Aus diesem Grund erhält jeder Stimulus die Reduktionsstufe des ihm zugeordneten Prototyps. Um zu überprüfen, ob die Prototypen-Stimuli sich nicht nur zur Klassifikation von Vokalrealisierungen eignen, sondern auch als Stellvertreter von Reduktionsstufen betrachtet werden können (Abschnitt 6.1.5, 6.1.6 und 6.2.4), wird für jeden Stimulus eine Referenzgröße – der ORL (*overall reduction level*) – bestimmt. Der ORL ergibt sich aus dem Modalwert der Reduktionsstufen des jeweiligen Stimulus, die aufgrund der individuellen Zuordnungen bestimmt werden.

### 6.1.3 Vergleichsexperimente

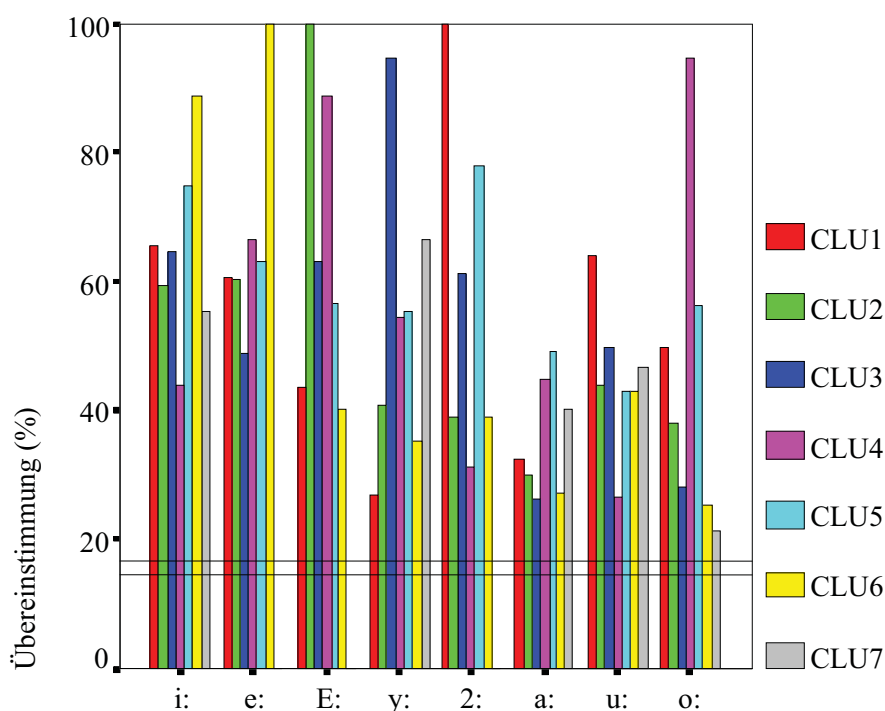
#### Versuchsreihe 1

In der ersten Vergleichstestreihe wird untersucht, ob die Referenzstimuli als prototypische Vertreter der aufgrund ihrer akustischen Ähnlichkeit zusammengefassten Vokalrealisierungen betrachtet werden können. Aus jedem Cluster wurden bis zu sechs Vokale genommen, die sich hinsichtlich ihrer Abstände zum Clusterzentrum stark unterscheiden. Zusätzlich wurde jeder Prototyp (Prototypen-Stimulus) ins Versuchsmaterial aufgenommen (/i:/, /o:/, /u:/: n = 33; /a:/: n = 42; /e:/: n = 32; /y:/: n = 24; /ɛ:/: n = 20; /ø:/: n = 18; n steht für die Anzahl der Stimuli). Jeder Stimulus wurde zweimal präsentiert. Neun Probanden nahmen an den Experimenten teil.

Ein Vergleich der Vokalrealisierungen, die aufgrund der perceptiven Zuordnung zu den Prototypen gruppiert werden, mit denen, die aufgrund der Clusteranalysen zusammengefasst werden, zeigt Übereinstimmungen, die deutlich über der Ratewahrscheinlichkeit liegen (Abbildung 4). Weiterhin deuten häufige Verwechslungen zwischen benachbarten Clustern auf eine Relevanz von F1 und F2 bei der Zuordnung der Stimuli zum Prototypen hin. Jedoch entsprechen sich die perceptiven Grenzen und die durch die Clusteranalysen ermittelten Grenzen nicht.

Bei den perceptiven Urteilen zeigt sich, dass in 94% der Fälle (Wert über alle Probanden und Vokale) die Prototypen-Stimuli dem richtigen Prototypen zugeordnet werden und 65% der Stimuli in der wiederholten Darbietung demselben Prototypen zugewiesen werden (Abbildung 5). Dies deutet darauf hin, dass die Probanden zwar in der Lage sind, die Stimuli den Prototypen zuzuordnen, aber die perceptiven Unterschiede zwischen den Prototypen zu gering sind, um die Stimuli konsistent zuzuweisen.

Die Relevanz eines Prototyps für eine perzeptive Klassifikation der Vokale wird anhand einer Vertauschungsmatrix ermittelt. Für jeden Stimulus wird ein Prototyp bestimmt, der für die Mehrheit der Probanden eine vergleichbar perzipierte Qualität besitzt. Dies entspricht dem Modalwert der Probandenurteile. Es wird angenommen, dass die Prototypen-Stimuli als Vertreter für vergleichbar wahrgenommene Vokalqualitäten betrachtet werden können und eine zuverlässige Klassifizierung der Vokale bewirken, sofern die individuellen Urteile der Probanden mit den Mehrheitsurteilen zu 50% übereinstimmen und mehr als ein Stimulus dem jeweiligen Prototypen zugeordnet wird. Die Prototypen-Stimuli selbst und Stimuli, die hinsichtlich des Mehrheitsurteils nicht eindeutig einem Prototypen zugeordnet werden können, da sie zwei Modalwerte besitzen, werden von der Analyse ausgeschlossen (/i:/, /ø:/: n = 1; /u:/, /a:/, /y:/: n = 2). Aufgrund dieses Kriteriums ergeben sich fünf potentielle Prototypen für /i:/ und /u:/ und drei für die anderen gespannten Vokale. Diese Prototypen wurden in einem weiteren Test mit demselben Experimentaufbau evaluiert.



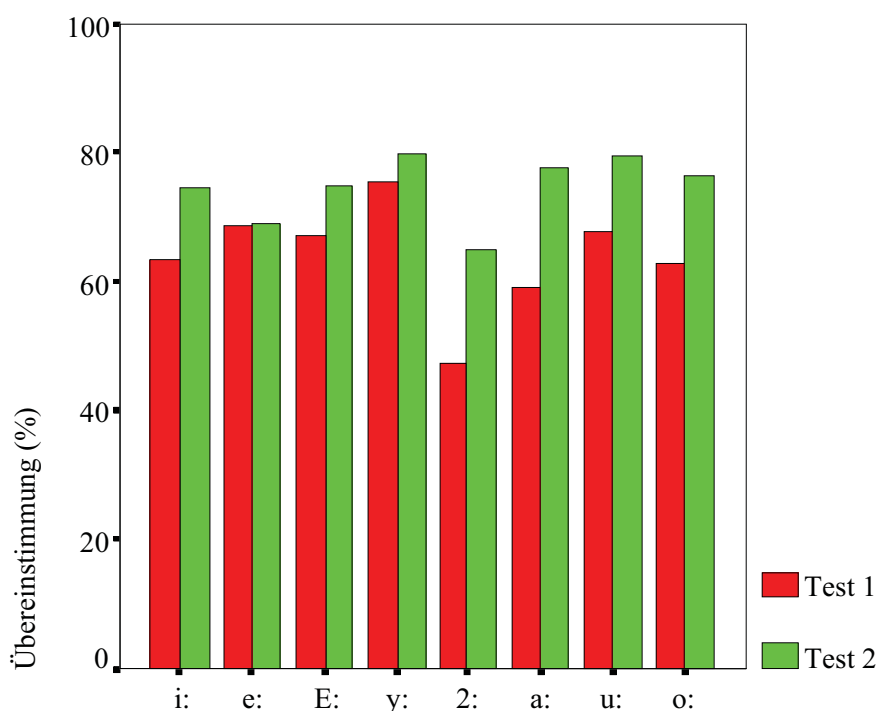
**Abbildung 4: Übereinstimmung (%) zwischen perzeptiver und akustischer (CLU1 bis CLU7; CLU steht für Cluster) Gruppierung von Vokalen (SAMPA-Notation). Die Linien markieren die Ratewahrscheinlichkeiten (obere Linie für /e:/, /ɛ:/ und /ø:/ mit sechs Cluster und untere Linie für die übrigen Vokale mit sieben Cluster).**

## Versuchsreihe 2

An der Evaluation der Prototypen nahmen acht der Probanden aus der ersten Experimentreihe teil. Die Stimuli wurden aus den Äußerungen von SPR1 aufgrund ihres möglichst verschiedenen Silbenkontexts ausgewählt (/i:/: n = 41; /o:/: n = 32; /u:/: n = 22; /a:/: n = 42; /e:/: n = 43; /y:/: n = 26; /ɛ:/: n = 17; /ø:/: n = 15). Jeder Stimulus wurde zweimal dargeboten.

Die intrasubjektiven Beurteilungen weisen im Vergleich zu der ersten Versuchsreihe auf eine konsistentere Zuweisung der Stimuli zu den Prototypen hin. 75% aller Stimuli werden bei wiederholter Präsentation demselben Prototypen zugeordnet (Abbildung 5). Bei /e:/ ist nur eine leichte Verbesserung festzustellen, wobei zu berücksichtigen ist, dass nur ein Prototyp der ersten Versuchsreihe ausgeschlossen wurde. Bei /ø:/ nehmen die intrasubjektiven Übereinstimmungen sichtbar zu, jedoch liegen sie noch unterhalb von denjenigen der anderen Vokale. Die Prototypen der zweiten Versuchsreihe werden deutlicher voneinander unterschieden, was zu einer konsistenteren Klassifikation der verschiedenen Vokalrealisierungen führt.

Ein Vergleich der individuellen Urteile mit den Mehrheitsurteilen zeigt für die einzelnen Vokalkategorien, mit Ausnahme von /i:/, eine Übereinstimmung von mindestens 70% (Abbildung 6). Wie schon in der ersten Versuchsreihe bleiben Stimuli, die hinsichtlich ihres Mehrheitsurteils nicht eindeutig einem Prototypen zugeordnet wurden, bei der Analyse unberücksichtigt (/i:/, /a:/,



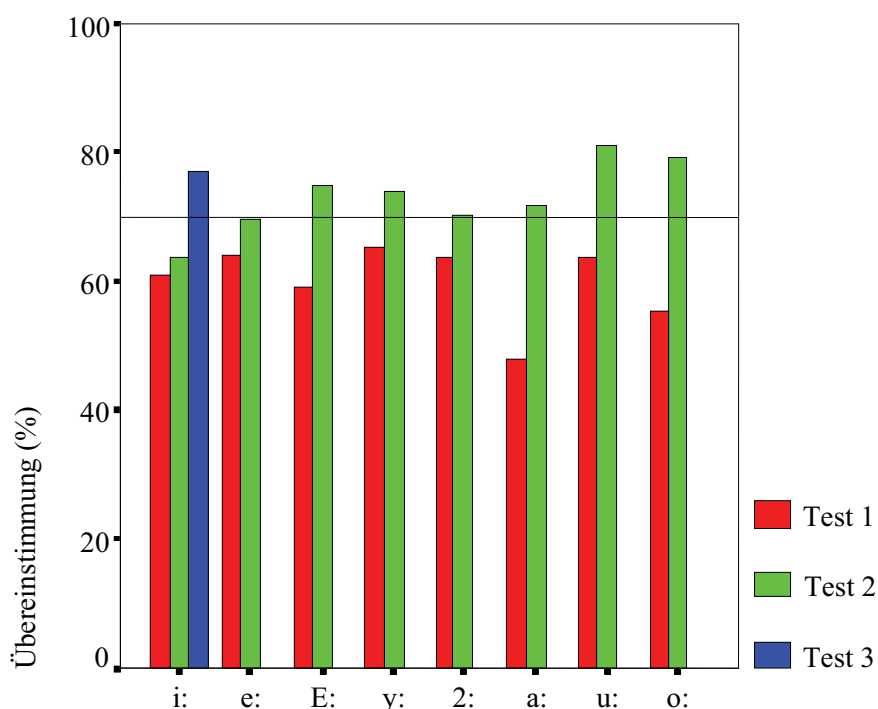
**Abbildung 5: Intrasubjektive Übereinstimmung bei wiederholter Präsentation der Stimuli einzelner Vokalkategorien (SAMPA-Notation) in Versuchsreihe 1 und 2.**

/o:/: n = 2; /u:/, /e:/: n = 1). Die niedrige Übereinstimmung bei /i:/ kann auf eine häufige Verwechslung zwischen zwei Prototypen zurückgeführt werden. Da diese Prototypen sehr ähnlich klingen, wurde einer der beiden ausgeschlossen. Die übrigen vier Prototypen wurden in den folgenden Perzeptionsexperimenten überprüft (vgl. Abschnitt 6.1.6). Die weiteren Ergebnisse von /i:/ beziehen sich auf die Beurteilung der vier Prototypen. Es zeigt sich, dass durch den Ausschluss des Prototyps die intersubjektiven Übereinstimmungen zunehmen.

Der Vergleich der Zuweisungen zu den Prototypen weist auf signifikante Beziehungen zwischen den Urteilen der einzelnen Probanden ( $\chi^2$ -Tests; /i:/, /e:/, /y:/, /u:/, /o:/:  $\alpha < 0,01$ ; /a:/:  $\alpha < 0,02$ ; /ε:/:  $\alpha < 0,05$ ). Nur bei /ø:/: sind neun nicht signifikante intersubjektive Beurteilungen festzustellen, wovon die meisten (sechs) auf die Beurteilung nur eines Probanden zurückgeführt werden können.

Es stellt sich die Frage, ob die Herausnahme der Prototypen wirklich zu einer konsistenteren Zuordnung geführt hat oder ob die höhere Übereinstimmung lediglich aus der Verringerung der Anzahl der Prototypen resultiert. Aus diesem Grunde wird zusätzlich die Übereinstimmung zwischen den individuellen Urteilen und den Mehrheitsurteilen unter Berücksichtigung der Anzahl der Prototypen berechnet (vgl. Lienert und Raats, 1994):

$$\dot{u}_{ma} = n_{ra} - \frac{n_{fa}}{n_{ma} - 1} \quad (5);$$



**Abbildung 6: Übereinstimmung zwischen Mehrheitsurteilen und individuellen Urteilen (%) für die einzelnen Vokalkategorien (SAMPA-Notation) in den verschiedenen Versuchsreihen (zu Test 3 vgl. Abschnitt 6.1.6). Die Linie markiert die Übereinstimmungsrate bei 70%.**

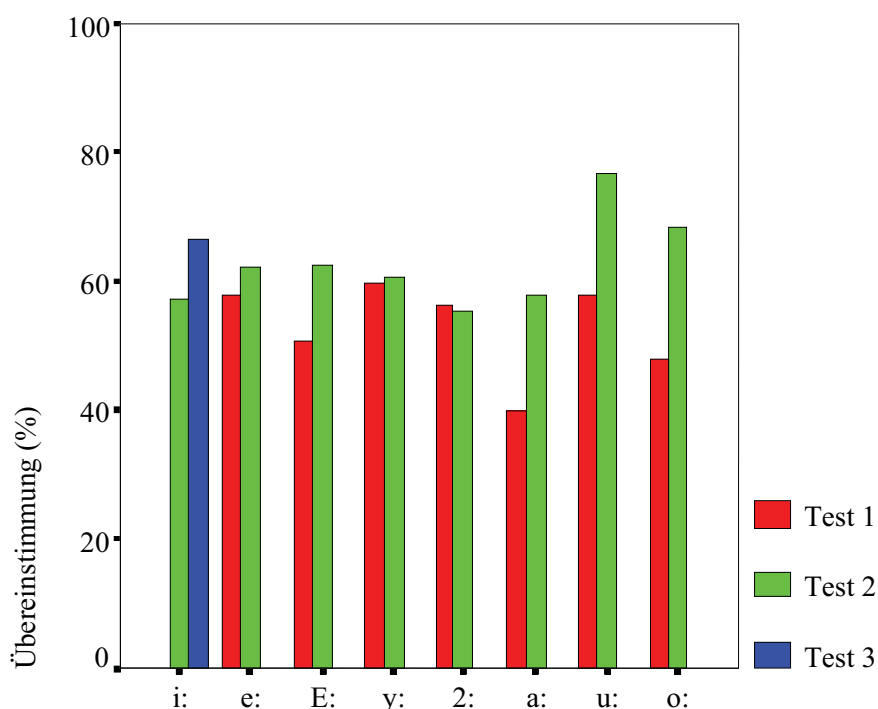
wobei  $n_{ra}$  für die Anzahl der Übereinstimmungen zwischen den Mehrheitsurteilen und den individuellen Urteilen (richtige Antworten) steht,  $n_{fa}$  für die Anzahl der Abweichungen (falsche Antworten) und  $n_{ma}$  für die Anzahl der Prototypen (mögliche Antworten).

Der Vergleich dieser korrigierten Übereinstimmungswerte mit denen der ersten Versuchsreihe zeigt, dass sich die Werte unabhängig von der Anzahl der Prototypen erhöht haben. Nur bei der Vokalkategorie /ø:/ kann keine nennenswerte Verbesserung festgestellt werden (Abbildung 7).

Die Ergebnisse weisen darauf hin, dass durch die Verringerung der Anzahl der Prototypen die Stimuli konsistenter zugewiesen werden. Die intrasubjektiven Übereinstimmungen sowie die zwischen den individuellen und mehrheitlichen Zuordnungen ermittelten Übereinstimmungen sind mit den Ergebnissen von Perzeptionsexperimenten zur Identifikation von Vokalqualitäten vergleichbar. Im folgenden wird geprüft, ob sich die einzelnen Prototypen auch als Stellvertreter für Reduktionsstufen eignen. Die Beurteilung der einzelnen Prototypen soll Aufschluss geben, ob die Prototypen hinsichtlich ihrer Reduktionsstärke nach geordnet werden können und welche Lautqualitäten mit den Prototypen assoziiert werden.

#### 6.1.4 Beurteilung der Prototypen

In den Beurteilungsexperimenten wurden die Probanden gebeten, die Prototypen der Vokalkategorien entsprechend ihrem Reduktionsgrad von nicht reduziert bis stark reduziert in eine



**Abbildung 7: Übereinstimmungen zwischen Mehrheitsurteilen und individuellen Urteilen (%) unter Berücksichtigung der Anzahl der Prototypen für die einzelnen Vokalkategorien (SAMPA-Notation) in den verschiedenen Versuchsreihen (zu Test 3 s. Abschnitt 6.1.6).**

Rangfolge zu bringen. Fünf Prototypen von /e:/ und /u:/, vier von /i:/ und jeweils drei Prototypen der übrigen gespannten Vokale wurden beurteilt. Der Anzahl der Reduktionsstufen entspricht die Anzahl der Prototypen (RED1 wird für die nicht reduzierte Stufe verwendet bis max. RED5). Weiterhin hatten die Probanden die Aufgabe, die Lautqualität der Prototypen mit Hilfe der IPA-Notation genau zu beschreiben.

Die Beurteilungen der Reduktionsstärke zeigen, dass die intersubjektive Übereinstimmung mit zunehmender Anzahl von Prototypen und zunehmenden Reduktionsgrad etwas sinkt. Nicht reduzierte Prototypen werden meist einstimmig als solche beurteilt, erst mit zunehmender Reduktion sind sich die Probanden bedingt einig, welcher Prototyp stärker reduziert ist. In Tabelle 5 ist für jede Lautkategorie die maximale Anzahl der übereinstimmenden Urteile hinsichtlich des Reduktionsgrades eines Prototyps aufgelistet. Die Reduktionsstärke eines Prototyps ergibt sich aus der am häufigsten zugeordneten Reduktionsstufe (vgl. Abschnitt 6.1.2).

Die Beschreibungen der Lautqualität zeigen intersubjektive Unterschiede hinsichtlich der engen Transkription (Verwendung der Diakritika). Die Vergabe der Lautsymbole ist wesentlich konsistenter als die der Diakritika. Die Probanden nehmen vergleichbare grundlegende Vokalqualitäten wahr. Die Lautqualität eines Prototyps wird anhand der am häufigsten vergebenen Lautsymbole bestimmt. Wenn die Qualitätsbeurteilung nicht eindeutig ist, werden mehrere Lautqualitäten angegeben (Tabelle 6).



	/i:/ (ng=8)	/y:/ (ng=8)	/e:/ (ng=8)	/ɛ:/ (ng=8)	/ø:/ (ng=8)	/a:/ (ng=8)	/u:/ (ng=8)	/o:/ (ng=8)
RED1	8	8	7	8	8	8	8	8
RED2	8	7	7	8	5	6	7	8
RED3	6	7	5	8	5	6	4	8
RED4	6		4				4	
RED5			7				5	

**Tabelle 5: Maximale Anzahl der übereinstimmenden Beurteilungen der Prototypen hinsichtlich der Reduktionsstärke (ng steht für Anzahl der Gesamturteile).**

Laut	ng	RED1		RED2		RED3		RED4		RED5	
		Qualität	n	Qualität	n	Qualität	n	Qualität	n	Qualität	n
/i:/	8	[i:]	7	[i]	7	[ɪ]	7	[ɪ]/[i]/[ə]	2/2/2		
/y:/	8	[y:]	8	[ʏ]	5	[ə]	5				
/e:/	8	[e:]	8	[e]	8	[ɛ]	7	[ə]	6	[ɪ]	6
/ɛ:/	8	[ɛ:]	7	[ɛ]	7	[ɛ̃]/[ə]	3/3				
/ø:/	8	[ø:]	8	[œ]	7	[œ]/[ə]	3/4				
/a:/	8	[a:]	6	[a]	5	[ä]	3				
/u:/	8	[u:]	8	[ʊ]	7	[ū]	4	[ʏ]/[ɪ]	3/3	[ə]	4
/o:/	8	[o:]	8	[ɔ]/[ʊ]	4/4	[ə]	4				

**Tabelle 6: Beurteilung der Lautqualität von Prototypen der Reduktionsstufen und maximale Anzahl der übereinstimmenden Urteile (n; ng steht für Anzahl der Gesamturteile).**

Die Lautqualität von Vokalrealisierungen von RED1 (nicht reduziert) entsprechen der in der BPD etikettierten Lautqualität. Mit zunehmender Reduktionsstärke werden gespannte Vokale erst mit einer ungespannten Qualität, dann einer mehr zentralisierten bzw. [ə]-Qualität assoziiert. Bei den Prototypen von /i:/ und /e:/ wird zwischen einem gespannten langen und gespannten kurzen Vokal unterschieden. Der am stärksten reduzierte Prototyp von /e:/ wird mit einer [ɪ]-Qualität wahrgenommen. Wie bereits in 4.3 erwähnt, werden /e/-Laute, deren Dauer gekürzt wird (ohne die spektralen Eigenschaften zu verändern) als [ɪ] perzipiert (Sendlmeier, 1981). Vokalreduktionen werden aus akustischer Sicht als Zentralisierung von Formanten betrachtet, jedoch liegt ein /ɪ/ in der Formantkarte noch oberhalb eines /e/-Lautes. Hinsichtlich seiner akustischen Eigenschaften ist der Prototyp somit nur wenig reduziert.

Die Probanden waren sich dieses Problems bewusst und berichteten von Schwierigkeiten bei der Einordnung des Prototyps. Die Entscheidung der Probanden, den Prototypen mit [ɪ]-Qualität als stark reduziert zu betrachten und den mit [ɛ]-Qualität als weniger reduziert einzustufen, deutet an, dass zwischen /e:/ und /ɛ/ mehr Ähnlichkeiten angenommen werden als zwischen /e:/ und /ɪ/. Möglicherweise wird diese Beurteilung durch die phonologische Theorie der in Opposition stehenden Vokale beeinflusst, in der /ɛ/ als ungespannter Partner von /e:/ gesehen wird.

Ein Prototyp der Vokalkategorie /u:/ wird mit einer nasalierten und entrundeten Qualität assoziiert, ein anderer mit einer [ʏ]- bzw. [ɪ]-Qualität. Beide Lautqualitäten können auf eine Koartikulation

mit den benachbarten Lauten zurückgeführt werden. Stimuli, die dem nasalierten Prototypen zugeordnet wurden, und auch der Prototyp selbst stammen aus einem nasalen Kontext, vorhergehender und folgender Laut sind alveolare Nasale. Stimuli und Prototyp mit einer [ʏ]- bzw. [ɪ]-Qualität besitzen einen alveolaren Kontext. Dies gilt in den meisten Fällen sowohl für den vorhergehenden als auch für den nachfolgenden Laut.

Die Beurteilungen der Lautqualität der Prototypen weisen darauf hin, dass die Probanden sich bei der Gruppierung von Vokalrealisierungen an dem Lautsystem ihrer Sprache orientieren. Zum einen werden die Prototypen häufig mit Vokalqualitäten assoziiert, die einen phonemischen Status im Deutschen besitzen und zum andern ist die Vergabe der Lautsymbole wesentlich konsistenter als die der Diakritika. Weiterhin zeigt sich, dass ein Lautsymbol nur selten in einer Reduktionskette mehrmals und mehrheitlich von den Probanden vergeben wird, d.h. dass innerhalb einer Vokalqualität nicht noch feinere Unterschiede gemacht werden. Dies tritt nur bei den als stark reduziert beurteilten Prototypen von /ɛ:/, /ø:/, /i:/ und /a:/ auf, wobei die Lautqualität der Prototypen der drei ersten Vokalklassen aber auch mit einer [ə]-Qualität assoziiert wird.

Die Beurteilungen zeigen, dass die Prototypen für eine Beschreibung von Reduktion typische Lautqualitäten besitzen und dass sie hinsichtlich ihrer Reduktionsstärke als geordnet betrachtet werden können. Die Prototypen werden als stellvertretende Instanzen von Reduktionsstufen angesehen. Die Vokalklassen /ɛ:/ und /u:/ besitzen fünf, /i:/ vier und die übrigen gespannten Vokale jeweils drei Reduktionsstufen.

#### 6.1.5 Ergebnisse

Die Ähnlichkeitsbeziehung zwischen Prototyp und Stimulus erlaubt die Annahme, dass jeder Stimulus eine vergleichbare Lautqualität besitzt wie der Prototyp selbst und die Stimuli dieselbe Reduktionsstufe mit dem Prototypen teilen. Aus diesem Grund erhält jeder Stimulus die Reduktionsstufe des ihm zugeordneten Prototyps sowie die Reduktionsstufe des mehrheitlich ausgewählten Prototyps (ORL; vgl. Abschnitt 6.1.2). Im folgenden werden die Zuweisungen zu den einzelnen Prototypen näher betrachtet. Im Vordergrund steht hierbei die Frage, wie konsistent die Probanden die Stimuli den einzelnen Prototypen zuordnen. Abweichungen der individuellen Urteile vom ORL können zudem Einblick geben, ob sich die Hierarchie der Reduktionsstufen der Prototypen in den Beurteilungen der Stimuli widerspiegelt. Wenn die Prototypen und Stimuli hinsichtlich ihrer Reduktion geordnet sind, ist zu erwarten, dass es häufiger zu Vertauschungen bei der Zuordnung zu Prototypen benachbarter Reduktionsstufen kommt.

In Tabelle 7 sind die individuellen Zuordnungen und Mehrheitsurteile nach ihrer Reduktionsstärke geordnet abgebildet. Für /i:/ und /ɛ:/ zeigt sich eine Übereinstimmung der individuell gewählten Reduktionsstufen und ORL von mindestens 70% bei der Zuweisung zu den einzelnen Prototypen, was der durchschnittlichen Übereinstimmung über alle Prototypen hinweg entspricht. Die prozentuale Verteilung der individuellen Urteile weisen aber darauf hin, dass

LAUT			ORL					Gesamt
			RED1	RED2	RED3	RED4	RED5	
/i:/	IND. URTEILE	RED1	79	18				97
		RED2	16	107	4			127
		RED3	1	13	<b>43</b>	<b>25</b>		82
		RED4		14	9	71		94
	Gesamt		96	152	56	96		400
/e:/	IND. URTEILE	RED1	136	22	2			160
		RED2	<b>48</b>	<b>108</b>	5		8	169
		RED3	3	10	45	4	1	63
		RED4	2	9	<b>12</b>	<b>21</b>	<b>13</b>	57
	RED5	3	11		7	58	79	
Gesamt		192	160	64	32	80	528	
/ɛ:/	IND. URTEILE	RED1	51	12	10			73
		RED2	13	<b>46</b>	<b>27</b>			86
	Gesamt		64	64	144			272
/a:/	IND. URTEILE	RED1	167	13	41			221
		RED2	17	132	38			187
	Gesamt		208	192	240			640
/y:/	IND. URTEILE	RED1	98	26	8			132
		RED2	9	<b>94</b>	<b>37</b>			140
	Gesamt		112	144	160			416
/ø:/	IND. URTEILE	RED1	45	6	5			56
		RED2	8	68	19			95
	Gesamt		64	96	80			240
/u:/	IND. URTEILE	RED1	148	9				157
		RED2	9	73		3	6	91
		RED3	3		<b>12</b>	2	<b>5</b>	22
		RED4		<b>6</b>	4	<b>20</b>	1	31
	RED5		<b>8</b>		<b>7</b>	20	35	
Gesamt		160	96	16	32	32	336	
/o:/	IND. URTEILE	RED1	124	23	3			150
		RED2	18	141	24			183
	Gesamt		144	184	160			488

**Tabelle 7: Anzahl der individuellen Urteile in Abhängigkeit der Mehrheitsurteile. Hervorgehobene Werte (fettgedruckt, s. Text) weisen häufige Vertauschungen mit anderen Reduktionsstufen auf.**

eine häufige Verwechslung bei der Zuweisung der Stimuli zu den Prototypen der letzten beiden Reduktionsstufen (RED2 und RED3 von /ɛ:/, RED3 und RED4 von /i:/) auftritt. Ein Drittel der als weniger stark reduziert klassifizierten Stimuli gehören nach den Mehrheitsurteilen der stark reduzierten Stufe an.

Bei /e:/ liegt die Übereinstimmung zwischen ORL und individuellen Urteilen um 70%. Häufigere Verwechslungen sind zwischen den ersten beiden Reduktionsstufen festzustellen. Dabei werden

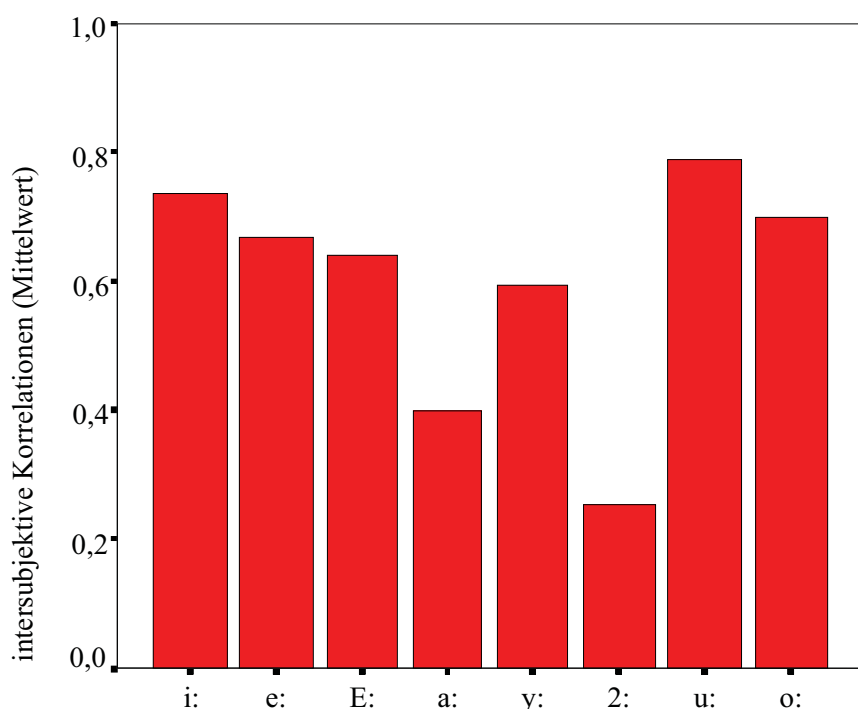
Stimuli, die nach ORL RED2 angehören, selten als nicht reduziert etikettiert, während Stimuli von RED1 zu 28% als RED2 perzipiert werden. Die Zuordnungen zu RED4 scheinen nicht sehr konsistent zu sein (37%), mehr als 20% der als RED4 klassifizierten Stimuli gehören nach ihrem ORL sowohl RED3 als auch RED5 an. Hierbei ist zu berücksichtigen, dass RED4 im Verhältnis zu den anderen Reduktionsstufen anzahlmäßig unterrepräsentiert ist, so dass die Fehlerwahrscheinlichkeit (ein Stimulus wird fälschlicherweise einem anderen Prototyp zugeordnet) wesentlich größer ist. Berechnungen der erwarteten Anzahl von Urteilen unter der Annahme, dass die individuellen Zuordnungen zu den Prototypen sich zufällig ergeben haben und somit kein Zusammenhang zwischen den Zuweisungen der Stimuli zu den Prototypen besteht<sup>4</sup>, weisen auf einen erheblichen Unterschied zwischen der beobachteten (21) und erwarteten Anzahl (3,5) von Beurteilungen für RED4 hin. Die Zuweisungen zu RED4 können somit als akzeptabel betrachtet werden.

Bei den ersten drei Reduktionsstufen der Vokalkategorie /u:/ stimmen ORL und individuelle Urteile zu mindestens 75% überein, während für die letzten beiden Reduktionsstufen die Übereinstimmung bei 63% liegt. Die Verteilung der individuell gewählten über die mehrheitlich bestimmten Reduktionsstufen zeigen, dass nicht immer die direkt benachbarten Reduktionsstufen von Verwechslungen betroffen sind. 23% der als RED3 beurteilten Reduktionsstufen werden nach ORL RED5 zugeordnet, um die 20% der als RED4 und als RED5 etikettierten Stimuli gehören nach den Mehrheitsurteilen RED2 bzw. RED2 und RED4 an. Die Verwechslungen von nicht benachbarten Reduktionsstufen sind auf die perzipierten Vokalqualitäten der Prototypen zurückzuführen. Der Prototyp von RED2 und RED3 besitzt eine [ʊ]- bzw. [ū]- Qualität, während der von RED4 mit einer [ɣ]- bzw. [ɪ]-Qualität assoziiert wird. Stark reduzierte /u/-Stimuli werden eher einem Prototypen mit einer u-ähnlichen Qualität zugeordnet. Wie schon bei /e:/, zeigen sich die meisten Vertauschungen bei den Reduktionsstufen mit geringer Anzahl, so dass es fraglich ist, ob die Zuweisungen zu diesen drei Prototypen (RED3, RED4, RED5) im Verhältnis zu den anderen zuverlässig sind. Vergleiche der erwarteten Anzahl von Beurteilungen einer Reduktionsstufe (RED3: 1; RED4: 3; RED5: 3,3) und der beobachteten Anzahl (RED3: 12; RED4: 20; RED5: 20), weisen auch hier auf ein noch akzeptables Ergebnis hin.

Die Übereinstimmung von ORL und individuellen Urteilen liegt bei den gerundeten Vorderzungenvokalen bei mindestens 70%, außer bei der Zuweisung zum Prototypen von RED2 bei /y:/. Häufigere Vertauschungen sind zwischen RED2 und RED3 festzustellen. Ein Viertel der als RED2 klassifizierten Stimuli von /y:/ gehören nach den Mehrheitsurteilen RED3 an, bei /ø:/ sind diese Verwechslungen in Richtung von RED3 nach RED2 zu beobachten. Bei den Zuordnungen zu den Prototypen von /o:/ sind keine nennenswerten Verwechslungen festzustellen. ORL und individuelle Urteile stimmen mit mindestens 77% überein.

---

<sup>4</sup> Die erwartete Anzahl ergibt sich aus der Anzahl der individuell bestimmten Reduktionsstufe (Zeilenrandsumme in Tabelle 7) und der Anzahl der ORL (Spaltenrandsumme in Tabelle 7) im Verhältnis zur Gesamtsumme aller Beurteilungen.



**Abbildung 8: Durchschnittliche intersubjektive Korrelationen für die einzelnen Vokale (SAMPA-Notation).**

Insgesamt zeigen sich weniger interindividuelle Abweichungen bei den Zuweisungen zu den nicht reduzierten Prototypen als zu den reduzierten, wobei die Abweichungen die Hierarchie der hinsichtlich ihrer Reduktionsstärke geordneten Prototypen wiedergeben. Verwechslungen zwischen direkt benachbarten Prototypen sind häufiger anzutreffen als zwischen Prototypen, die sich hinsichtlich ihrer Reduktion stark unterscheiden. Es kann davon ausgegangen werden, dass durch die Zuordnung zu den Prototypen Vokalrealisierungen hinsichtlich ihrer Reduktionsstärke geordnet werden.

Korrelationsanalysen weisen auf eine signifikante Beziehung zwischen den Beurteilungen der Probanden hin. Abbildung 8 zeigt für jede Vokalkategorie die durchschnittlichen Korrelationswerte der Beurteilungen von allen Probandenpaaren. Die intersubjektiven Beurteilungen von /i:/, /e:/, /y:/, /u:/ und /o:/ weisen ein Signifikanzniveau von 0,01 auf, bei /ɜ:/ liegt es bei 0,04. Die Korrelationswerte für diese Vokalkategorien liegen im Durchschnitt über 0,6. Beurteilungen der Lautrealisierungen von /a:/ und /ø:/ weisen nur einen schwachen Zusammenhang auf. Bei /a:/ kann dies in erster Linie auf die Beurteilung nur eines Probanden zurückgeführt werden, abgesehen hiervon liegt das Signifikanzniveau der übrigen intersubjektiven Korrelationen bei 0,02. Für /ø:/ finden sich bei neun Probandenpaaren keine signifikanten Übereinstimmungen, sechs davon basieren auf den Urteilen einer Versuchsperson, die anderen intersubjektiven Beurteilungen sind signifikant ( $\alpha \leq 0,05$ ).

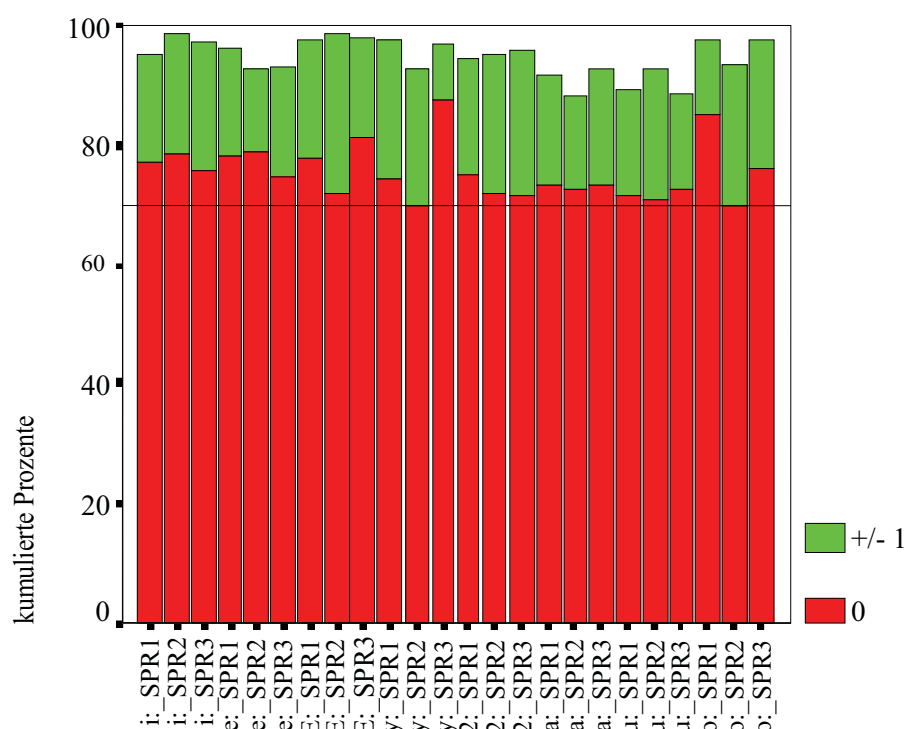
Die Ergebnisse zeigen, dass für die Vokalkategorien /e:/ und /u:/ jeweils fünf Reduktionsstufen diskriminiert werden, für /i:/ vier und für die übrigen gespannten Vokale jeweils drei. Vergleiche der inter- und intrasubjektiven Beurteilungen weisen auf eine akzeptable Klassifikation von Lautrealisierungen hinsichtlich ihrer Reduktionsstärke für den größten Teil der gespannten Vokalkategorien hin. Die Übereinstimmungen liegen im Durchschnitt um 70%. Lautrealisierungen werden durch Zuordnung zu den Prototypen in Gruppen mit vergleichbar wahrgenommener Reduktionsstärke unterteilt.

#### 6.1.6 Sprecherunabhängige Prototypen

In den bisherigen Experimenten wurde getestet, ob die durch Clusteranalysen ermittelten Vokale als prototypische Vertreter für Reduktionsstufen gelten. Dabei wurden sowohl die Prototypen als auch die Stimuli den Äußerungen einer mit Sprachaufnahmen vertrauten Sprecherin (SPR1) entnommen. Für die Vokalkategorien /e:/ und /u:/ konnten jeweils fünf Prototypen gefunden werden, für /i:/ vier und für die übrigen gespannten Vokale jeweils drei. Es stellt sich die Frage, ob diese Prototypen sich auch für die Klassifizierung von Vokalrealisierungen anderer Sprecher eignen und somit als sprecherunabhängige prototypische Vertreter reduzierter Vokale angesehen werden können.

In der folgenden Versuchsreihe bestand die Aufgabe der Probanden darin, Lautrealisierungen verschiedener Sprecher den Prototypen zu zuordnen. Der Experimentaufbau war mit dem der vorherigen Vergleichsversuche identisch. An dem Experiment nahmen acht phonetisch geschulte Probanden teil, sieben von ihnen waren bereits Teilnehmer der vorherigen Versuchsreihe.

Die Stimuli wurden aus den Äußerungen von SPR1, der professionellen Sprecherin, deren Vokalrealisierungen in den vorherigen Experimentreihen beurteilt wurden, sowie der zweiten Sprecherin (SPR3) und des Sprechers (SPR2) geschnitten. SPR2 und SPR3 besitzen anders als SPR1 keine Erfahrung mit Sprachaufnahmen. Die Auswahl der Stimuli beruht auf automatischen Einteilungen der Vokale in Reduktionsstufen durch künstliche neuronale Netze. Da in diesem Abschnitt auf die Ergebnisse der neuronalen Netze nicht eingegangen wird, erfolgt nur eine kurze Darstellung. Die automatische Klassifikation von Reduktionsstufen wird in Kapitel 9 behandelt. Für jede Vokalkategorie getrennt, wurden neuronale Netze (*feed-forward* Netze mit zwei verdeckten Schichten) mit den hinsichtlich ihrer Reduktionsstärke beurteilten Vokalen – es wurden die ORL-Urteile genommen – von SPR1 trainiert. Die Klassifikation der Vokale basierte auf den sprecher- und lautnormierten Werten der ersten drei Formanten, der Energie in den vier verschiedenen Frequenzbändern, der Grundfrequenz und der Lautdauer (vgl. Widera und Portele, 1999). Die trainierten Netze wurden zur weiteren Klassifikation der Vokale von SPR2 und SPR3 genutzt. Aus den klassifizierten Vokalen (insgesamt 2830) wurden pro Sprecher und pro Reduktionsstufe jeweils fünf Stimuli zufällig ausgewählt (/e:/, /u:/:  $n = 75$  ( $5*3*5$ ); /i:/:  $n = 60$  ( $5*3*4$ ); /ɛ:/, /y:/, /ø:/, /a:/, /o:/:  $n = 45$  ( $5*3*3$ )), die dann von den Probanden aufgrund ihrer perzipierten Ähnlichkeit den Prototypen zugeordnet wurden.

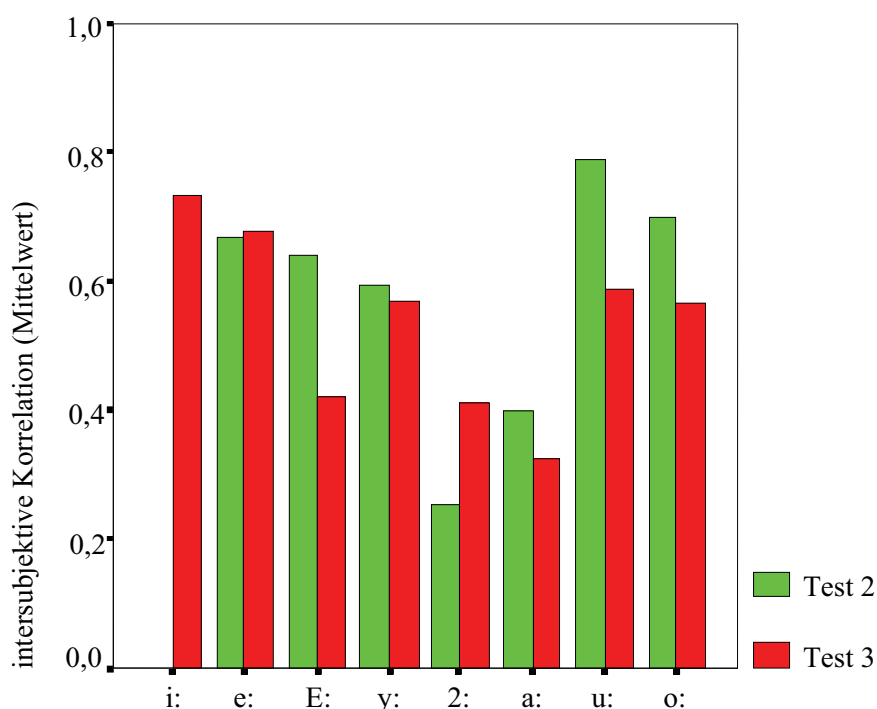


**Abbildung 9: Übereinstimmungen bzw. keine Abweichungen (0; in %) zwischen ORL und individuellen Urteilen sowie Abweichungen um eine Reduktionsstufe (+/- 1; in %) für die Vokale verschiedener Sprecher (x\_y, x steht für die Vokalkategorie (SAMPA-Notation) und y für den einzelnen Sprecher). Die Linie markiert die Übereinstimmung bei 70%.**

Wie schon in den vorangehenden Abschnitten werden die Stimuli, die nicht eindeutig einer ORL zugeordnet wurden, ausgeschlossen. Im Vergleich zu den vorherigen Analysen ist die Anzahl der auszuschließenden Fälle höher (/i:/: n = 10; /e:/, /u:/: n = 6; /ø:/, /a:/: n = 4; /ɛ:/, /o:/: n = 3; /y:/: n = 2), jedoch hängt dies nicht vom Sprecher ab. Die durchschnittliche Anzahl der nicht eindeutig zuzuordnenden Stimuli zeigt, dass Stimuli von SPR1 (n = 13) nicht konsistenter den Prototypen zugeordnet werden als Stimuli von SPR2 (n = 16) und SPR3 (n = 9).

Vergleiche von ORL und individuellen Beurteilungen zeigen, dass unabhängig vom Sprecher die Übereinstimmungen vergleichbar sind (Abbildung 9). Sie liegen bei mindestens 70% für die einzelnen Vokalkategorien und Sprecher und sind somit mit den Ergebnissen der vorherigen Versuchsreihe vergleichbar. Über alle Vokalkategorien stimmen durchschnittlich 76% der mehrheitlich gewählten Reduktionsstufen mit den individuell beurteilten Stimuli von SPR1 überein, 77% mit denen von SPR3 und 73% mit denen von SPR2. Die intersubjektiven Abweichungen von ORL weisen nur geringe Unterschiede bei den Vokalkategorien der einzelnen Sprecher auf. Am häufigsten sind Verwechslungen zwischen benachbarten Reduktionsstufen festzustellen. Die durchschnittlichen Verwechslungsraten der benachbarten Reduktionsstufen über alle Vokalkategorien liegen für SPR1 und SPR3 bei 18% und für SPR2 bei 21%.





**Abbildung 10: Durchschnittliche intersubjektive Korrelationen für die einzelnen Vokale (SAMPA-Notation) im Vergleich zur zweiten Versuchsreihe (zu /i:/ vgl. Abschnitt 6.1.6).**

Die intersubjektiven Korrelationen sind mit denen der vorherigen Experimente für die meisten Vokalkategorien (/i:/, /e:/, /y:/:  $\alpha \leq 0,01$ ; /a:/, /ø:/  $\alpha \leq 0,05$ ; s. ) vergleichbar. Bei den beiden Hinterzungenvokalen nehmen die Korrelationswerte ab, sind aber signifikant (/u:/:  $\alpha \leq 0,01$ ; /o:/  $\alpha \leq 0,02$ ). Im Vergleich zu der vorherigen Versuchsreihe ist ein schwächerer intersubjektiver Zusammenhang bei /ɛ:/ zu verzeichnen. Dies ist möglicherweise auf eine sprecherspezifische Realisierung von /ɛ:/ zurückzuführen. SPR2 gehört nicht zu der Sprechergruppe, die bei der Realisierung von /ɛ:/ und /e:/ eine deutliche Unterscheidung macht. Die intersubjektiven Beurteilungen differieren bei der Zuordnung zum nicht reduzierten Prototypen mit [ɛ:]-Qualität.

Die Vergleiche von ORL und individuellen Urteilen sowie die intersubjektiven Korrelationen weisen darauf hin, dass die Probanden sprecherspezifische Unterschiede zwischen Prototypen und Stimuli kompensieren. Die auf Grundlage des Datenmaterials einer Sprecherin gefundenen Prototypen von Reduktionsstufen können zur Klassifikation von Vokalrealisierungen anderer Sprecher und Sprecherinnen genutzt werden.

## 6.2 Reduktionsstufen ungespannter Vokale

Die Perzeptionsexperimente zur Reduktion gespannter Vokale zeigen, dass Hörer aufgrund von Zuordnungen verschiedener Vokalrealisierungen zu potentiellen Prototypen von Reduktionsstufen drei bis fünf Reduktionsstufen in Abhängigkeit der Vokalkategorie diskriminieren können.

Im folgenden wird überprüft, wie viele Reduktionsstufen Hörer bei ungespannten Vokalen wahrnehmen können. Ausgangspunkt der Untersuchung sind die bisherigen Ergebnisse zur Reduktionsperzeption der gespannten Vokale. Es wird angenommen, dass Hörer im Vergleich zu den gespannten Vokalen aufgrund ihrer akustischen Nähe zum /ə/ eine höchstens gleich große Anzahl von Reduktionsstufen wahrnehmen können.

### 6.2.1 Ermittlung potentieller Prototypen

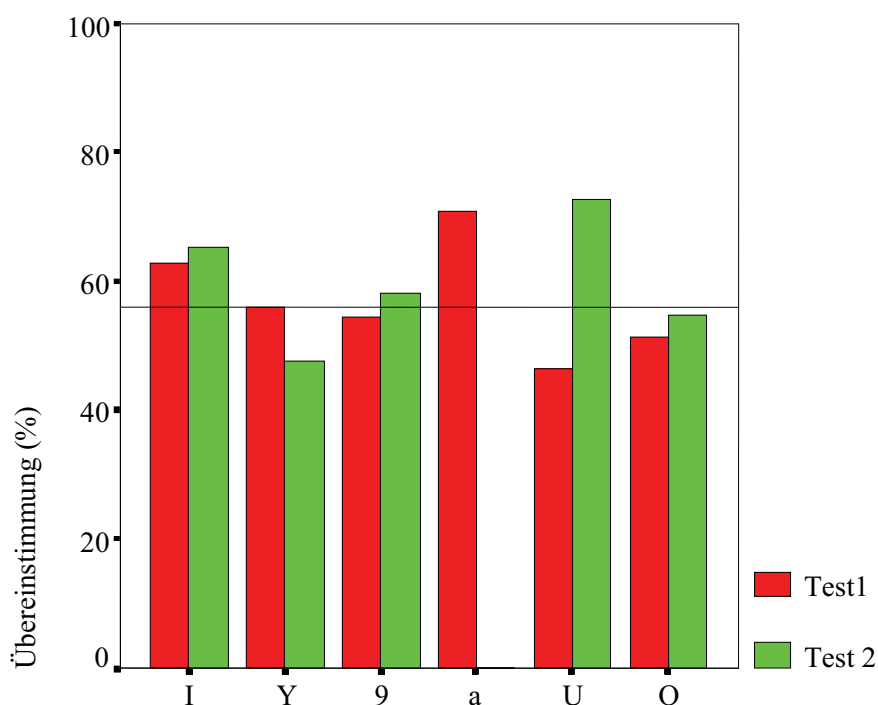
Wie schon bei der Untersuchung der gespannten Vokale, werden zur Bestimmung der potentiellen Prototypen die mel-skalierten, sprecher- und lautnormierten F1- und F2-Werte der ungespannten Vokale Clusteranalysen unterzogen. Da der qualitative Vergleich von Prototyp und Stimulus nicht durch Sprecherunterschiede beeinflusst wird, werden die Vokalrealisierungen aller drei Sprecher herangezogen. Der Anzahl der Cluster der ungespannten Vokale entspricht die Reduktionsstufenanzahl des in phonologischer Opposition stehenden gespannten Vokals: vier Gruppen für /ɪ/ und drei für /ʏ/, /œ/, /a/, /ɔ/ und /ɛ/. Für letztere Vokalkategorie bildet die Anzahl der Reduktionsstufen von /ɛ:/ aufgrund der qualitativen Ähnlichkeit die Ausgangsbasis. Die Vokalrealisierungen von /u/ werden nur in drei Gruppen unterteilt, da zwei der fünf Reduktionsstufen des in Opposition stehenden Vokals /u:/ auf Lautrealisierungen mit einem sehr engen phonetischen Kontext beschränkt sind. Aus jedem Cluster wird ein Vokal ausgewählt, der dem Clusterzentrum am nächsten ist (vgl. Abschnitt 6.1.1). Die Anzahl der Prototypen von /ɛ/ und /a/ wird um eins verringert, da diese Vokale aus Clustern stammen, die wenige Laute mit unzuverlässigen Formantwerten (Ausreißer) enthalten (s. Anhang B, Tabelle B-2).

Der Experimentaufbau ist mit dem der gespannten Vokale identisch. Acht phonetisch geschulte Probanden hatten die Aufgabe, verschiedene Vokalrealisierungen den Prototypen aufgrund ihrer qualitativen Ähnlichkeit zuzuordnen (Abschnitt 6.2.2) und die Prototypen hinsichtlich ihrer Reduktionsstärke und Qualität zu beurteilen (Abschnitt 6.2.3). Die Stimuli wurden aufgrund eines möglichst unterschiedlichen Kontextes und ihrer Sprecher ausgewählt und konnten sowohl mit Kontext als auch allein gehört werden (/ɪ/: n = 60; /u/: n = 71; /œ/: n = 43; /ɛ/: n = 29; /a/, /ʏ/, /ɔ/: n = 45).

### 6.2.2 Vergleichsexperimente

#### Versuchsreihe 1

Die Ergebnisse der ersten Versuchsreihe der ungespannten Vokale weisen darauf hin, dass aufgrund häufiger Verwechslungen bei der Zuweisung der Stimuli zu den Prototypen von /ɪ/, ein Prototyp ausgeschlossen werden muss. Bei den Vokalkategorien /ɛ/, /ʏ/, /œ/, /u/ und /ɔ/ zeigt der Vergleich der Zuweisungen der einzelnen Probanden untereinander nicht signifikante Beziehungen ( $\chi^2$ -Test,  $\alpha > 0,05$ ) auf. Bei letzteren vier Vokalen wird die Prototypenanzahl auf zwei verringert. Diese Prototypen sowie die von /ɪ/ werden in einer zweiten Versuchsreihe überprüft. /ɛ/ hingegen



**Abbildung 11: Übereinstimmung zwischen Mehrheitsurteilen und individuellen Urteilen (%) unter Berücksichtigung der Anzahl der Prototypen für die einzelnen Vokalkategorien (SAMPA-Notation) in den verschiedenen Vergleichsexperimenten (für /a/ wurde nur ein Experiment durchgeführt). Die Linie markiert die Übereinstimmung bei der Beurteilung von /ø:/.**

besitzt nur zwei Prototypen, so dass eine Einteilung der verschiedenen Lautrealisierungen in Reduktionsstufen aufgrund der nicht übereinstimmenden Diskriminierung der Probanden nicht zweckmäßig ist. Für /a/ weisen signifikante Zusammenhänge bei den intersubjektiven Zuweisungen ( $\chi^2$ -Test;  $\alpha < 0,02$ ) sowie eine Übereinstimmung von 85% hinsichtlich des mehrheitlich und individuell gewählten Prototyp auf eine Unterscheidung von zwei Reduktionsstufen hin.

### Versuchsreihe 2

In der folgenden Versuchsreihe werden die übrigen Prototypen der Vokalkategorien /ɪ/ (mit drei Prototypen), /ʏ/, /œ/, /ʊ/ und /ɔ/ (mit jeweils zwei Prototypen) bezüglich ihrer Relevanz für eine Klassifizierung der Vokalrealisierungen in Reduktionsstufen getestet. An den Untersuchungen nahmen sieben der Probanden der vorherigen Perzeptionsexperimente teil. Es wurden die Stimuli aus der ersten Versuchsreihe verwendet.

Die Übereinstimmungen zwischen den mehrheitlich und individuell gewählten Prototypen liegen oberhalb von 75%. Betrachtungen der Übereinstimmungen unter Berücksichtigung der Anzahl der Prototypen weisen auf eine Erhöhung der intersubjektiven Konsistenz bei der Zuweisung der Stimuli zu den Prototypen hin. Jedoch sind die Werte der Vokalkategorien /ʏ/, /œ/, und /ɔ/ mit denjenigen des gespannten Vokals /ø:/ vergleichbar, die allerdings nur sehr schwache

intersubjektive Korrelationen bei der Beurteilungen aufweisen. Vergleiche der individuellen Urteile der einzelnen Probanden untereinander deuten nur bei /i/ und /u/ auf signifikante Zusammenhänge hin ( $\chi^2$ -Test,  $\alpha < 0,01$ ).

Die Ergebnisse der Vergleichsexperimente weisen auf drei unterscheidbare Reduktionsstufen für /i/ und jeweils zwei für /u/ und /a/ hin. Für die übrigen ungespannten Vokale werden keine Reduktionsstufen angenommen.

### 6.2.3 Beurteilung der Prototypen

Wie schon bei den gespannten Vokalen wurden die Probanden gebeten, die Prototypen hinsichtlich ihrer Reduktionsstärke zuordnen und ihre Lautqualität genau zu beschreiben. Die Probanden beurteilten die drei potentiellen Prototypen der Reduktionsstufen von /i/ und die zwei von /u/ und /a/.

Die Beurteilungen der Reduktionsstärke weisen eine hohe Übereinstimmung zwischen den Probanden auf (Tabelle 8). Betrachtungen der perzipierten Qualitäten zeigen, dass die nicht reduzierten Prototypen von /a/ und /u/ mit der jeweiligen Qualität der Vokalkategorie assoziiert werden (Tabelle 9). Der nicht reduzierte Prototyp von /i/ wird hingegen mit einer gespannten Qualität wahrgenommen und kann als hyperartikulierte betrachtet werden. Die stark reduzierten Prototypen (RED2 von /a/ und /u/, RED3 von /i/) werden mit einer zentralisierten bzw. [ə]-Qualität wahrgenommen. Die Probandenurteile weichen weniger voneinander ab, wenn die Prototypen mit einer phonemischen Vokalqualität perzipiert werden.

Die Ergebnisse zeigen, dass die Prototypen als Stellvertreter von Reduktionsstufen betrachtet werden können, so dass jeder Stimulus die Reduktionsstufe des gewählten Prototyps erhält. Im

	/i/ (ng=7)	/a/ (ng=7)	/u/ (ng=7)
RED1	7	7	6
RED 2	6	7	6
RED 3	6		

**Tabelle 8: Maximale Anzahl der übereinstimmenden Beurteilungen der Prototypen hinsichtlich der Reduktionsstärke (ng steht für Anzahl der Gesamturteile).**

Laut	ng	RED1		RED2		RED3	
		Qualität	n	Qualität	n	Qualität	n
/i/	7	[i]	6	[ɪ]	7	[ə]	6
/a/	7	[a]	7	[ä]	4		
/u/	7	[u]	6	[ü]	5		

**Tabelle 9: Beurteilung der Lautqualität von den Prototypen der Reduktionsstufen und maximale Anzahl der Übereinstimmenden Urteile (n; ng steht für Anzahl der Gesamturteile).**

folgenden werden die interindividuellen Zuweisungen der nach ihrer Reduktionsstärke geordneten Stimuli zu den einzelnen Prototypen untersucht.

#### 6.2.4 Ergebnisse

In Tabelle 10 ist die Anzahl der individuellen Urteile in Abhängigkeit von ORL abgebildet. Die Übereinstimmung zwischen den mehrheitlichen und individuellen Zuordnungen liegt in allen Fällen für die einzelnen Prototypen über 75%. Häufige Verwechslungen sind bei /ɪ/ und /a/ zwischen RED1 und RED2 festzustellen. 41% der Stimuli, die dem Prototypen der ersten Reduktionsstufe von /ɪ/ zugeordnet werden, gehören nach ORL RED2 an. Bei /a/ konzentrieren sich die Vertauschungen auf die zweite Reduktionsstufe. 26% der mit RED2 etikettierten Stimuli werden mehrheitlich RED1 zugewiesen. Wie schon bei den gespannten Vokalen zeigt sich, dass die Zuweisung zu den nicht reduzierten Prototypen konsistenter ist als zu den reduzierten. Bei /ɪ/ ist zu berücksichtigen, dass es sich beim Prototypen von RED1 um eine hyperartikulierte Form handelt, während der Prototyp von RED2 die eigentliche nicht reduzierte Vokalrealisierung darstellt.

LAUT			ORL			Gesamt
			RED1	RED2	RED3	
/ɪ/	IND. URTEILE	RED1	<b>54</b>	<b>39</b>	3	96
		RED2	15	169	19	203
		RED3	1	16	90	107
	Gesamt	70	224	112	406	
/a/	IND. URTEILE	RED1	200	22		222
		RED2	<b>24</b>	<b>69</b>		93
	Gesamt	224	91		315	
/ʊ/	IND. URTEILE	RED1	276	36		312
		RED2	32	153		185
	Gesamt	308	189		497	

**Tabelle 10: Anzahl der individuellen Urteile in Abhängigkeit der Mehrheitsurteile. Hervorgehobene Werte (fettgedruckt, s. Text) weisen häufige Vertauschungen mit anderen Reduktionsstufen auf.**

Korrelationsanalysen zeigen signifikante Zusammenhänge zwischen den Beurteilungen der einzelnen Probanden auf. Die durchschnittlichen Korrelationswerte sind jedoch geringer als bei den gespannten Vokalen (/ʊ/: $\rho = 0,6$ ,  $\alpha < 0,01$ ; /ɪ/:  $\rho = 0,5$ ,  $\alpha < 0,05$ ; /a/:  $\rho = 0,5$ ,  $\alpha < 0,03$ ).

Die Ergebnisse weisen auf drei Reduktionsstufen für /ɪ/ und jeweils zwei Reduktionsstufen für die Vokalkategorien /a/ und /ʊ/ hin. Im Gegensatz hierzu ist die Einteilung der Vokalrealisierungen von /ɛ/, /ɣ/, /œ/ und /ɔ/ in Reduktionsstufen unzuverlässig.

### 6.3 Zusammenfassung und Diskussion

Ausgangspunkt dieser Untersuchung war die Fragestellung, wie viele Reduktionsstufen Hörer bei einzelnen Vokalkategorien wahrnehmen. Um eine möglichst differenzierte Beschreibung von Vokalreduktion zu erhalten, wurden Vergleichsexperimente durchgeführt. Die Vokale wurden mit

Hilfe von Clusterzentrenanalysen gruppiert. Es wurde angenommen, dass die Clusteranalysen Hinweise auf potentielle Prototypen von Reduktionsstufen geben. Aus jedem Cluster wurde ein Laut gewählt, der in der Nähe des jeweiligen Clusterzentrums liegt. Ob diese Laute prototypische Instanzen sind, wurde durch die Zuordnung weiterer Vokale zu diesen prototypischen Instanzen getestet. Die Zuweisungen basieren auf der perzipierten qualitativen Ähnlichkeit. Die Ergebnisse zeigen, dass nicht alle Instanzen als Prototypen betrachtet werden können. Diese wurden ausgeschlossen, und die übrig gebliebenen potentiellen Prototypen wurden in weiteren Perzeptionsexperimenten evaluiert. Zusätzlich wurde die Stärke der Reduktion der Prototypen sowie ihre Vokalqualität beurteilt. Aufgrund der perzipierten Ähnlichkeit, wurde jeder Vokal mit der Reduktionsstufe des ihm zugeordneten Prototyps versehen. Die Ergebnisse wurden hinsichtlich der intersubjektiven Beurteilungen analysiert.

Die Experimente zeigen, dass Lautrealisierungen einzelner Vokalkategorien in Reduktionsstufen zusammengefasst werden können. Hörer nutzen die Prototypen für die Einteilung von Vokalrealisierungen verschiedener Sprecher in Reduktionsstufen. Sie kompensieren Sprecherunterschiede, wie es auch in der natürlich sprachlichen Kommunikation der Fall ist. Jedoch hängt die Anzahl der perzipierten Reduktionsstufen von der Vokalkategorie ab. Innerhalb der Gruppe der gespannten Vokale werden für /e:/ und /u:/ jeweils fünf Reduktionsstufen unterschieden, für /i:/ vier und für /ɛ:/, /y:/, /ø:/, /a:/ sowie /o:/ jeweils drei Reduktionsstufen. Für /ø:/ gilt die Unterscheidung von drei Reduktionsstufen nur eingeschränkt, da die intersubjektiven Beurteilungen nicht sehr zuverlässig sind. Bei den ungespannten Vokalen lassen sich nur für die an den Eckpunkten des Vokaldreiecks befindlichen Vokale /ɪ/, /a/ und /ʊ/ Reduktionsstufen ausmachen.

Im Vergleich zu allen anderen Vokalen kann die Vokalqualität zweier Prototypen von /u:/ auf Koartikulation mit den unmittelbar benachbarten Lauten zurückgeführt werden. Auf eine ausgeprägte Koartikulation von /u/ weisen auch Untersuchungen von Vokalen im Niederländischen hin. Nach van Heuvel et al. (1996) werden die Formantwerte von /u/ im Vergleich zu denjenigen von /i/ und /a/ stärker beeinflusst. Dies ist insbesondere dann festzustellen, wenn Nasale oder alveolare Konsonanten dem Vokal vorausgehen. Auch bei den Vokalen der vorliegenden Reduktionsstufen wird die Qualität durch einen nasalen bzw. alveolaren Kontext bestimmt.

Die perzipierten Qualitäten der Prototypen der gespannten Vokale /i:/, /a:/ und /u:/ und den jeweils in phonologischer Opposition stehenden ungespannten Vokale /ɪ/, /a/ und /ʊ/ weisen auf vergleichbare Reduktionsverläufe hin. Werden die koartikulatorisch bedingten Reduktionsstufen von /u:/ nicht berücksichtigt, so besitzen die ungespannten Vokale eine Reduktionsstufe weniger als die gespannten:

/i:/ [i:] ⇒ [i] ⇒ [ɪ] ⇒ [ɪ]/[i]/[ə]  
 /ɪ/ [i] ⇒ [ɪ] ⇒ [ə]  
 /a:/ [a:] ⇒ [a] ⇒ [ḁ]  
 /a/ [a] ⇒ [ḁ]  
 /u:/ [u:] ⇒ [ʊ] ⇒ [ə]  
 /ʊ/ [ʊ] ⇒ [ʊ]

Zwischen einer ungespannten Qualität und einer zentralisierten bzw. [ə]-Qualität werden sowohl bei den gespannten als auch bei den ungespannten Vokalen keine weiteren Lautqualitäten zuverlässig diskriminiert. Bei /i:/ und /ɪ/ unterscheiden die Probanden zwischen einer kurzen gespannten und einer kurzen ungespannten Vokalrealisierung. Die Ergebnisse zeigen, dass die Reduktion von den an den Eckpunkten des Vokaldreiecks liegenden gespannten und ungespannten Vokalen mit einer Reduktionskette beschrieben werden können. Aus perceptiver Sicht können die ungespannten Vokale als reduzierte gespannte Vokale betrachtet werden.

Die Tatsache, dass nur für die an den Eckpunkten des Vokaldreiecks liegenden ungespannten Vokale Reduktionsstufen gefunden werden, legt die Vermutung nahe, dass die perzipierte Distanz der übrigen ungespannten Vokale zum /ə/ zu gering ist, um weitere Reduktionsstufen zuverlässig zu diskriminieren. Weitere Hinweise, dass die Ähnlichkeit zum /ə/ die Reliabilität der Diskriminierung von Reduktionsstufen beeinflusst, geben die einzelnen Zuweisungen von Stimuli zu Prototypen. Es finden sich stärkere interindividuelle Abweichungen bei den reduzierten Prototypen im Vergleich zu den nicht reduzierten.

Bei der Reduktion der ungerundeten vorderen Eckvokale /i:/ und /ɪ/ sowie von /e:/ wird im Vergleich zu den gerundeten Vorderzungen- und Hinterzungenvokalen zwischen einer gespannten kurzen und ungespannten Vokalqualität unterschieden. Dies entspricht den Beobachtungen, dass zum einem, obwohl akustische Analysen auf Unterschiede zwischen gespannten Kurz- und Langvokalen hinweisen, dieser nur bei hohen Vokalen perceptiv relevant ist (vgl. Abschnitt 4.3). Zum anderem sind Beurteilungen ungerundeter Vorderzungenvokale im Vergleich zu den gerundeter Vokale einheitlicher (vgl. Abschnitt 2.1), so dass insbesondere bei den ungerundeten Vorderzungenvokale noch feinere Unterschiede konsistent diskriminiert werden.

Die Reduktionsstufen werden häufig mit Vokalqualitäten assoziiert, die einen phonematischen Status im Deutschen besitzen. Auf Reduktionsstufen, deren Prototypen mit einer zentralisierten ungespannten Vokalqualität wahrgenommen werden, folgen keine weiteren Reduktionsstufen, so dass diese auch die stets vorhandenen Vokalrealisierungen mit einer [ə]-Qualität beinhalten. Auf Einflüsse des Phonemsystems auf die Vokalperzeption weisen auch Untersuchungen von Butcher (1976) und Lindner (1977) hin. Butcher überprüfte die Wahrnehmung der Kardinalvokale bei englischen, französischen und deutschen Hörern in Vergleichsexperimenten. Von jeweils drei präsentierten Stimuli mussten die unähnlichsten bestimmt werden. Lindner verglich die Einteilung



synthetisierter Vokale bei tschechischen und deutschen Hörern. Beide Autoren kommen zu dem Ergebnis, dass das internalisierte Phonemsystem die Beurteilung der Hörer beeinflusst. Nach Chistovich (1971) und van Valin Jr. (1976) werden nicht die Stimuli selbst, sondern die mit den Stimuli assoziierten Phonemqualitäten miteinander verglichen. Dies entspricht den in Abschnitt 3.3 beschriebenen Befunden, wonach die Diskriminierung von Stimuli nicht ausschließlich auf Vergleiche ihrer (psycho)akustischen Eigenschaften beruht, sondern auch phonetische Betrachtungen beinhalten.

Welche Strategie die Hörer bei der Zuordnung der Stimuli zu den Prototypen verwendeten, kann aus den vorliegenden Ergebnisse nicht geschlossen werden. Aber es zeigt sich, dass nur dann eine weitgehend konsistente Zuweisung zu den Prototypen erfolgt, wenn die perzipierten Abstände zwischen den Prototypen mit denen von Phonemen vergleichbar sind. Obwohl Hörer vielfältige Nuancen von Vokalqualitäten wahrnehmen können, werden Vokalrealisierungen nur dann zuverlässig in Reduktionsstufen eingeteilt, wenn sie sich an dem Phonemsystem ihrer Sprache orientieren.

Im nächsten Kapitel werden die akustischen Eigenschaften der in Reduktionsstufen eingeteilten Vokalrealisierungen untersucht. Es wird der Frage nachgegangen, ob sich die einzelnen Reduktionsstufen akustisch voneinander unterscheiden, welche Parameter von der Reduktion beeinflusst werden und ob es Zusammenhänge zwischen akustischer Reduktion und perzipierter Reduktionsstärke gibt. Da einerseits die Reduktionsketten der gespannten Vokale /i:/, /u:/ und /a:/ vergleichbar sind zu den der in phonologischer Opposition stehenden ungespannten Vokale, andererseits aber keine ähnlichen Reduktionsverläufe zwischen den übrigen gespannten Vokalen und ihren ungespannten Vokalpartnern zu finden sind, wird geprüft, ob ungespannte als reduzierte gespannte Vokale betrachtet werden können und inwiefern die Distanz zum /ə/ die Diskriminierung von Reduktionsstufen aus akustischer Sicht beeinflusst. Ein weiterer Untersuchungsgegenstand liegt in der Analyse der akustischen Korrelate der Ähnlichkeitsbeziehung von Prototypen und zugeordneten Stimuli. Es wird nach einer akustischen Beschreibung der perzipierten Ähnlichkeit gesucht.

## 7 AKUSTISCHE ANALYSE DER REDUKTIONSSSTUFEN

Ausgangspunkt der bisherigen Untersuchung war die Fragestellung, ob Vokalvarianten sich in Reduktionsstufen einteilen lassen und eine adäquate und zuverlässige Beschreibung von Vokalreduktion erlauben (Kapitel 6). Verschiedene lautliche Realisationen innerhalb einer Vokalkategorie wurden aufgrund ihrer qualitativen Ähnlichkeit zu ausgewählten Referenzstimuli (Prototypen) in Gruppen zusammengefasst. Die Prototypen wurden bezüglich ihres Reduktionsgrades beurteilt und hinsichtlich ihrer lautlichen Qualität beschrieben. Da zwischen Prototyp und zugeordneten Stimuli eine qualitative Ähnlichkeit besteht, kann angenommen werden, dass die Stimuli den gleichen Reduktionsgrad besitzen wie der ausgewählte Prototyp. Die Ergebnisse der Perzeptionsexperimente belegen, dass Hörer verschiedene Reduktionsstufen innerhalb einer Vokalkategorie zuverlässig diskriminieren können und die Ähnlichkeitsbeurteilung von Stimuli und Prototyp nicht durch Sprecherunterschiede beeinflusst wird. Die Anzahl der Reduktionsstufen hängt jedoch von der Vokalkategorie ab.

Die Reduktionsstufe eines Vokals wird im folgenden durch den Reduktionsgrad des von den Hörern mehrheitlich gewählten Prototyps (ORL) definiert. Der Begriff „Reduktionsstufen“ impliziert, dass es sich dabei um diskrete Einheiten handelt, die hinsichtlich ihrer Reduktionsstärke geordnet sind. In Abschnitt 7.1 wird untersucht, ob sich die akustischen Parameter der perzipierten Reduktionsstufen voneinander unterscheiden und ob es eine Beziehung zwischen perzipierter Reduktionsstärke und akustischer Reduktion gibt. Die Ergebnisse werden im Zusammenhang mit dem *target undershoot*-Modell nach Lindblom (1963; vgl. Abschnitt 2.3) diskutiert.

Weiterhin wird der Frage nachgegangen, warum anders als bei den gespannten Vokalen nur für die an den Eckpunkten des Vokaldreiecks befindlichen ungespannten Vokale Reduktionsstufen zuverlässig perzipiert werden (Abschnitt 7.2). Insbesondere wird auf die Distanz zum /ə/ als mögliche Ursache für eine nicht reliable Klassifikation von Lautrealisierungen in Reduktionsstufen eingegangen. Aufgrund der Diskrepanz bei der Perzeption ungespannter und gespannter Vokale werden die Ergebnisse in Beziehung zur Erörterung der phonologischen Gespanntheitsopposition gesetzt. Im Vordergrund steht die Rechtfertigung der in phonologischer Opposition stehenden Vokale aus Sicht der Reduktion.

In Abschnitt 7.3 werden die akustischen Korrelate der perzipierten Ähnlichkeit zwischen Vokalen betrachtet. Die Hörer hatten die Aufgabe, Stimuli aufgrund ihrer qualitativen Ähnlichkeit den Prototypen zu zuordnen. Basierend auf der Theorie des Perzeptiven Magneteffekts (Kuhl, 1992, Abschnitt 3.3) wird davon ausgegangen, dass akustischer Abstand und perzipierte Ähnlichkeit in Bezug zueinander stehen.

## 7.1 Perzeptive und akustische Reduktionsstufen

Wie in Abschnitt 6.1.2 beschrieben, erhält jeder Stimulus die gleiche Reduktionsstufe wie der ihm zugeordnete Prototyp. Die Reduktionsstufe eines Stimulus wird durch den Modalwert der Urteile aller Probanden (ORL) bestimmt. Die beurteilten Stimuli dienen als Basis für die akustische Analyse der Reduktionsstufen. Stimuli, deren Reduktionsstufen nicht eindeutig durch den Modalwert bestimmt sind – sie besitzen mehrere Modalwerte – werden von der Analyse ausgeschlossen. Insgesamt werden 818 Vokalrealisierungen (für /ø:/ und /a/ n = 45, für /i:/ n = 106, für /ɪ/ n = 57, für /y:/ n = 58, für /e:/ n = 122, für /ɛ:/ n = 33, für /a:/ n = 93, für /u:/ n = 98, für /ʊ/ n = 71, für /o:/ n = 90), die hinsichtlich ihrer Reduktion beurteilt wurden, betrachtet.

Im folgenden wird geprüft, durch welche akustischen Parameter sich die Reduktionsstufen voneinander unterscheiden. Der Schwerpunkt liegt auf dem Vergleich einer Reduktionsstufe mit der ihr folgenden Stufe (RED1 vs. RED2, RED2 vs. RED3, usw.). Weiterhin wird untersucht, ob es einen Zusammenhang zwischen perzipierter und akustischer Reduktion gibt. Analysiert werden die ersten drei Formanten, die die qualitativen Eigenschaften eines Vokals bestimmen, und die Dauer. Um nicht nur einzelne Energieanreicherungen im spektralen Bereich zu betrachten, werden die bezüglich der Gesamtenergie normierten Energieverteilungen in den verschiedenen Frequenzbändern hinzugenommen. Die Grundfrequenz wird einbezogen, da von ihr zum einem die Vokalqualität abhängt, zum anderen auch die Betonung und betonte Vokale meistens weniger stark reduziert sind als unbetonte (Abschnitte 2.2 und 2.3). In den Analysen werden die sprecher- und lautnormierten Werte (ZF1, ZF2, ZF3, ZE1NORM, ZE2NORM, ZE3NORM, ZE4NORM, ZLAUTDUR, ZF0) betrachtet. Der Medianwert der akustischen Parameter, Stichprobenumfang sowie erstes und drittes Quartil als Streuungsmaße für die einzelnen Reduktionsstufen der Vokalkategorien sind im Anhang C (Tabelle C-1 bis Tabelle C-5) dargestellt.

### 7.1.1 Akustische Realisierung der Reduktionsstufen

Wie aus Tabelle 11 (vgl. auch Tabelle C-6 bis Tabelle C-8 in Anhang C) ersichtlich ist, unterscheiden sich die Vokalrealisierungen einer Reduktionsstufe im Vergleich zur nächsten Stufe hinsichtlich der Werte eines oder mehrerer Formanten. Die Ausnahmen bilden /ɛ:/ RED1 vs. RED2, /y:/ RED2 vs. RED3, /ø:/ RED2 vs. RED3 und /u:/ RED2 vs. RED3. Der Unterschied zwischen /ɛ:/ RED2 und RED3 ist nur geringfügig ( $\alpha = 0,055$ ). Für /ø:/ weisen Korrelationsanalysen auf einen leichten Zusammenhang zwischen perzipierter Reduktionsstärke und den ersten beiden Formanten, für /y:/ ist ein Zusammenhang zwischen Reduktionsstärke und F1 festzustellen (Abbildung 12). Bei beiden Vokalen steigen die F1-Werte mit zunehmender Reduktionsstärke und die F2-Werte nehmen ab.

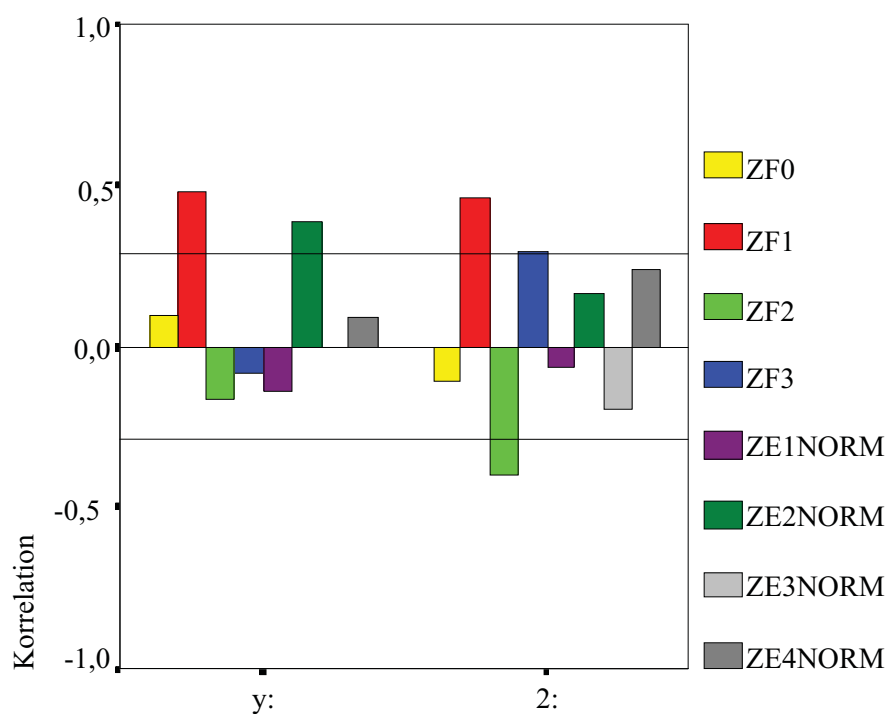
Bei den Vorderzungenvokalen /i:/, /ɪ/ und /e:/ sind die Unterschiede (vgl. Tabelle 11) zwischen den Reduktionsstufen auf eine Erhöhung von F1 und eine Senkung von F2 und F3 zurückzuführen.

Laut	RED	ZF1	ZF2	ZF3	ZE1NORM	ZE2NORM	ZE3NORM	ZE4NORM	ZLAUTDUR	ZF0
/i:/	1 vs. 2	x	x	x					x	
	2 vs. 3	x	x	x	x	x	x			x
	3 vs. 4		x		x		x		x	x
/ɪ/	1 vs. 2	x					x			
	2 vs. 3		x	x		x	x	x	x	
/e:/	1 vs. 2	x	x	x	x	x			x	
	2 vs. 3	x	x	x	x	x	x	x	x	
	3 vs. 4	x								
	4 vs. 5	x			x	x				x
/ɛ:/	1 vs. 2								x	
	2 vs. 3	(x)								
/y:/	1 vs. 2	x							x	
	2 vs. 3								x	
/ø:/	1 vs. 2		x	x						
	2 vs. 3									
/a:/	1 vs. 2	x	x	x	x	x			x	x
	2 vs. 3			x	x	x	x			x
/a/	1 vs. 2	x			x	x	x		x	
/u:/	1 vs. 2		x		x	x	x		x	
	2 vs. 3									
	3 vs. 4	x			x	x	x			
	4 vs. 5		x		x	x	x			
/ʊ/	1 vs. 2		x		x		x			
/o:/	1 vs. 2		x		x	x	x		x	
	2 vs. 3	x	x		x	x	x			

**Tabelle 11: Signifikante Unterschiede (gekennzeichnet mit x) zwischen den sprecher- und lautnormierten Formant- (ZF1, ZF2, ZF3), Energie- (ZE1NORM, ZE2NORM, ZE3NORM, ZE4NORM), Grundfrequenz- (ZF0) und Dauerwerten (ZLAUTDUR) benachbarter Reduktionsstufen (RED; s. Anhang C, Tabelle C-6 bis Tabelle C-8). Reduktionsstufen, die keine Unterschiede zwischen den akustischen Parametern aufweisen sind fettgedruckt.**

Korrelationsanalysen zeigen einen deutlichen Zusammenhang zwischen perzipierter Reduktionsstärke und F2 (Abbildung 13).

Im Vergleich zu /ɪ/ und /e:/ zeigt sich bei /i:/ eine stärkere Korrelation zwischen der perzipierten Reduktionsstärke und den F1- sowie F3-Werten. Aus der Transkription (vgl. Abschnitt 6.1.4) der Prototypen ist erkennbar, dass die Vokalrealisationen von /e:/ RED3 eine [ɛ]-Qualität besitzen, /e:/ RED4 mit einer [ə]-Qualität assoziiert wird und /e:/ RED5 mit einer [ɪ]-Qualität. Wie bereits in 5.2.2 dargestellt, liegen die F1-Werte von /ə/ unterhalb derjenigen von /e/ und oberhalb der Werte von /ɪ/. Ab /e:/ RED4 wird der erste Formant im Gegensatz zu den Formanten der weniger reduzierten Vokalrealisationen gesenkt. Werden nur die Formantwerte von /e:/ bis zum Erreichen der [ə]-Qualität betrachtet (RED1 bis RED4), so wird der Zusammenhang zwischen wahrgenommener Reduktionsstärke und spektraler Reduktion deutlicher.



**Abbildung 12: Korrelationen zwischen wahrgenommener Reduktionsstärke und akustischen Parametern für die gerundeten Vorderzungenvokale (SAMPA-Notation). Korrelationswerte ober- bzw. unterhalb der Linien sind signifikant.**

Die Unterschiede zwischen nicht reduzierten und reduzierten Lautrealisierungen von /a:/ bzw. /a/ sind auf eine Senkung von F1 und Erhöhung von F2 und F3 zurückzuführen (Tabelle 11). Im Fall von /a:/ RED3 stimmt die Lage der ersten beiden Formanten mit den Formanten von /a:/ RED2 überein, jedoch sind die Werte von F3 geringer als die der weniger reduzierten Vokalrealisierungen. Die signifikanten Unterschiede zwischen den Formantwerten benachbarter Reduktionsstufen der Hinterzungenvokale basieren auf einer Erhöhung von F1 bzw. F2. Im Fall von /u:/ RED4 vs. RED5 ist eine Abnahme der F2-Werte zu verzeichnen. Die Beurteilung der Prototypen zeigt, dass die Vokalrealisierungen von RED4 eine [ɣ]-Qualität besitzen, die von RED5 eine [ə]-Qualität, was der Senkung von F2 entspricht.

Im Gegensatz zu den Formantwerten unterscheiden sich nicht alle Vokale durch ihre Energieverteilung in den Frequenzbändern (vgl. Tabelle 11). Insbesondere die Energieverteilungen im höherfrequenten Bereich (4–8 kHz) werden durch eine Reduktion kaum beeinflusst. Die meisten signifikanten Unterschiede sind in den Frequenzbereichen der ersten drei Formanten zu finden, wobei eine Änderung der Formanten auf die Energieverteilung in den Frequenzbändern bis maximal 4 kHz bei den Hinterzungenvokalen und den tiefen Vokalen besonders deutlich wird. Dies ist dadurch erklärbar, dass F1 von /a:/ bzw. /a/ und F2 der Hinterzungenvokale im Bereich der oberen Grenze des ersten Frequenzbandes bzw. der unteren Grenze des zweiten Frequenzbandes liegen. Eine Verschiebung der Formanten in den Frequenzbereich eines Bandes bewirkt eine Abnahme der Energie des anderen Frequenzbands.

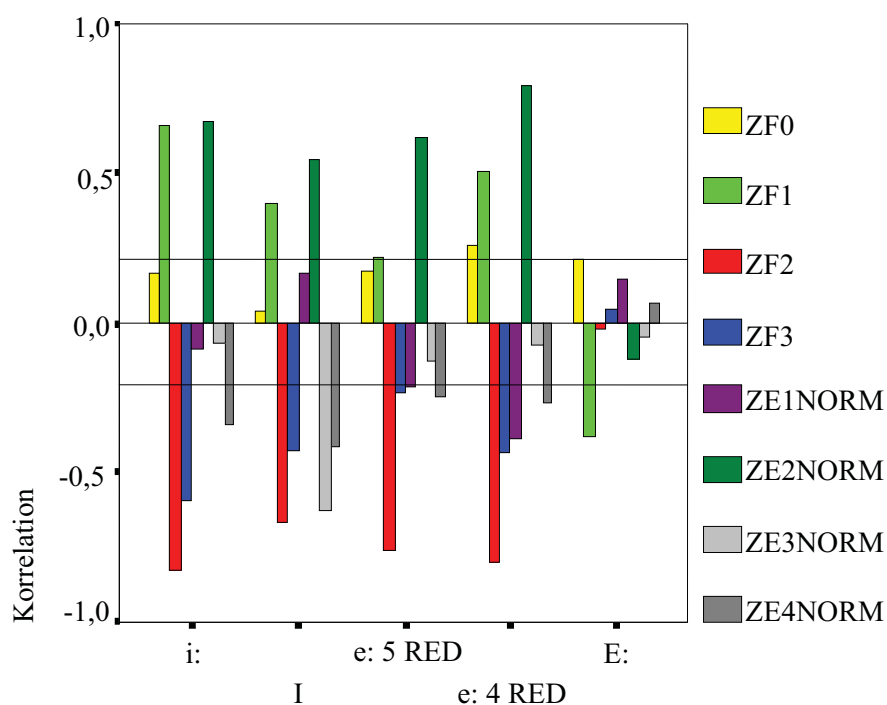


Abbildung 13: Korrelationen zwischen wahrgenommener Reduktionsstärke und akustischen Parametern für die ungerundeten Vorderzungenvokale (SAMPA-Notation; 4 RED: mit vier Reduktionsstufen, 5 RED: mit fünf Reduktionsstufen, s. Text). Korrelationswerte ober- bzw. unterhalb der Linien sind signifikant.

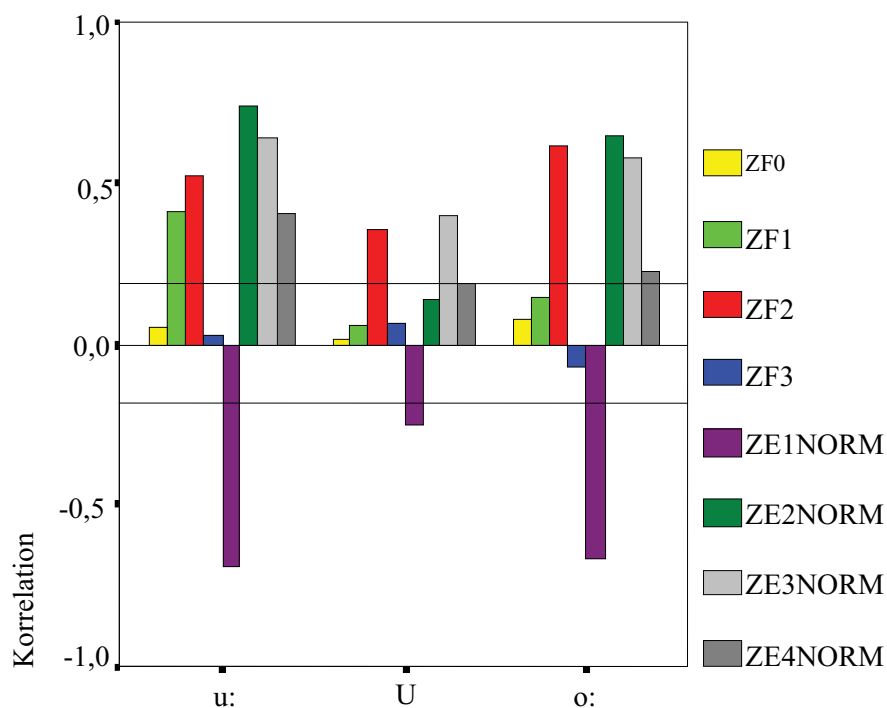
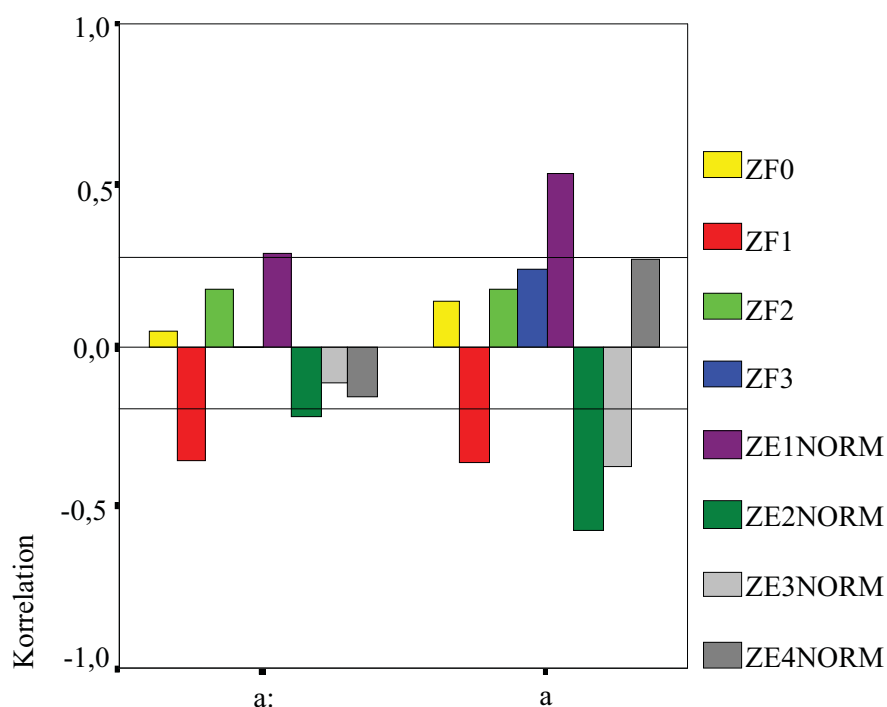


Abbildung 14: Korrelationen zwischen wahrgenommener Reduktionsstärke und akustischen Parametern für die Hinterzungenvokale (SAMPA-Notation). Korrelationswerte ober- bzw. unterhalb der Linien sind signifikant.



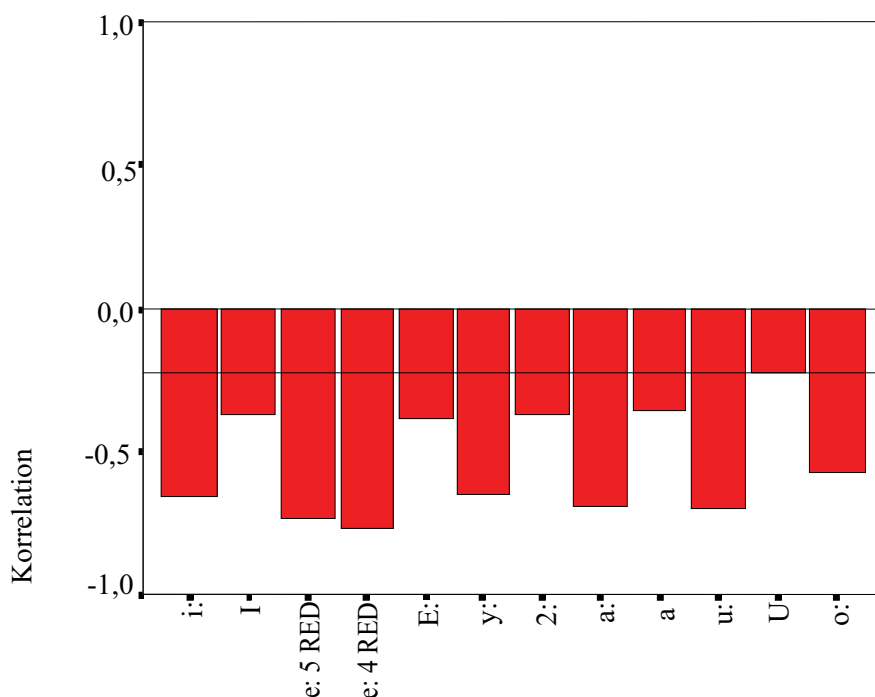
**Abbildung 15: Korrelationen zwischen wahrgenommener Reduktionsstärke und akustischen Parametern für /a:/ und /a/. Korrelationswerte ober- bzw. unterhalb der Linien sind signifikant.**

Die Unterschiede zwischen den Energieverteilungen der benachbarten Reduktionsstufen der Hinterzungenvokale lassen sich auf eine Abnahme der Energie im ersten Frequenzband und einer Zunahme im zweiten und dritten Frequenzband zurückführen. Diese Energieverschiebung kann auch beim Vergleich von /a:/ RED1 und RED2 festgestellt werden. Unterschiede zwischen /a:/ RED2 und RED3 bzw. /a/ RED1 und RED2 sind durch eine Zunahme der Energie im ersten Frequenzband und einer Abnahme im zweiten und dritten Band gekennzeichnet.

Im Vergleich zu den Vorderzungenvokalen wird bei den hinteren gespannten Vokalen ein stärkerer Zusammenhang zwischen Reduktionsstärke und den Energiewerten in den ersten drei Frequenzbändern erkennbar. Die Reduktionsstufen von /u/ weisen nur eine gering ausgeprägte lineare Beziehung zu den spektralen Parametern auf. Sie korrelieren am stärksten mit den Energieverteilung im dritten Frequenzband (Abbildung 14). Die Reduktionsstufen von /a:/ zeigen nur geringe Korrelationen zwischen Reduktionsstärke und den Formant- bzw. den Energiewerten. Bei /a/ ist ein mittlerer Zusammenhang zwischen den Energieverteilungen in den ersten zwei Frequenzbändern und perzipierter Reduktionsstärke festzustellen (Abbildung 15).

Hinsichtlich ihrer Dauern unterscheiden sich die nicht reduzierten von den reduzierten (RED2) Lautrealisierungen aller Vokalkategorien mit Ausnahme von /ɪ/, /ø:/ und /ʊ/ (Tabelle 11). Bei /ɪ/ zeigen sich Dauerunterschiede zwischen den reduzierten (RED2) und stark reduzierten (RED3) Lautrealisierungen. Weitere Unterschiede hinsichtlich der Lautdauern finden sich bei /i:/ zwischen RED3 und RED4 und bei /y:/ zwischen RED2 und RED3. Die Lautdauern der stärker reduzierten





**Abbildung 16: Korrelationen zwischen wahrgenommener Reduktionsstärke und Lautdauer für die einzelnen Vokalkategorien (SAMPA-Notation; 5 RED: mit fünf Reduktionsstufen, 4 RED: mit vier Reduktionsstufen, s. Text). Korrelationswerte unterhalb der Linie sind signifikant.**

Vokale sind im Vergleich zu den der weniger reduzierten bzw. nicht reduzierten Vokalrealisierungen kürzer. Die Korrelationsanalysen zeigen einen Zusammenhang zwischen perzipierter Reduktionsstärke und den Lautauern für alle Vokale mit Ausnahme von /ʊ/ (Abbildung 16). Dieser ist insbesondere für die Vokalkategorien /i:/, /y:/, /e:/, /a:/ und /u:/ stark ausgeprägt. Wie schon die Korrelationsanalysen der spektralen Parameter zeigen, erhöhen sich die Korrelationswerte von /e:/, wenn nur die Reduktionsstufen bis zum Erreichen einer [ə]-Qualität berücksichtigt werden.

Die Grundfrequenz scheint kaum eine Rolle für die Unterscheidung der Reduktionsstufen zu spielen. Die Analysen weisen, wenn überhaupt, nur geringfügige Korrelationen zwischen wahrgenommener Reduktionsstärke und Grundfrequenz auf. Obwohl einerseits Vokalreduktion und Betonung miteinander korrelieren, andererseits Betonung und Grundfrequenz in Beziehung gesetzt werden können, besteht kein Zusammenhang zwischen der Grundfrequenz und den Reduktionsstufen. Hierbei ist zu bedenken, dass die Grundfrequenzwerte der Reduktionsstufen durch die Satzintonation überlagert werden können. Aber auch in Äußerungen mit vergleichbarer Intonation und Betonung weisen Vokale des Deutschen keinen Zusammenhang zwischen der Grundfrequenz und der Gespanntheitsopposition auf (Fischer-Jørgensen, 1990; Jessen et al., 1995). Während innerhalb der Gruppe der gespannten und der ungespannten Vokale die Grundfrequenz mit der Höhe der Vokale zunimmt, besitzen gespannte und ungespannte Vokale vergleichbare Grundfrequenzwerte (Fischer-Jørgensen, 1990). Es kann davon ausgegangen

werden, dass die Grundfrequenz für die Perzeption reduzierter Vokalqualitäten, wenn überhaupt, nur eine sehr geringfügige Relevanz besitzt.

### 7.1.2 Zusammenfassung

Die Analysen zeigen, dass sich die Vokalrealisierungen benachbarter Reduktionsstufen vor allem durch die Formant- und Dauerwerte unterscheiden. Nur die akustischen Realisierungen von /ø:/ RED2 und RED3 sowie von /u:/ RED2 und RED3 sind vergleichbar. Die akustische Ähnlichkeit zwischen den zwei Reduktionsstufen von /ø:/ erklärt die geringere Übereinstimmung zwischen den Probanden bei der Zuordnung der Stimuli zu den Prototypen. Die akustische Vergleichbarkeit der beiden Reduktionsstufen von /u:/ wird durch die Beschreibung der Lautqualitäten erklärbar (Abschnitt 6.1.4). Die Laute, die RED3 angehören, besitzen eine nasale Qualität, der Prototyp wurde mit [ū] bzw. [ŋ] etikettiert. Eine Nasalierung von Vokalen bewirkt eine zusätzliche Resonanz im Bereich von F0, häufig eine Auslöschung, Schwächung oder Erhöhung von F1 sowie eine Dämpfung von F2 oder F3. Somit können die spektralen Muster Ähnlichkeiten zu denen von /u:/ RED2 aufweisen.

Die Korrelationsanalysen weisen auf einen nahezu linearen Zusammenhang zwischen perzipierter und akustischer Reduktion hin. Die Formanten sind mit ansteigender Reduktion zunehmend zentralisiert (s. beispielsweise die Gegenüberstellung der ersten beiden Formantwerte (Hz-skaliert<sup>5</sup>) von SPR1; Abbildung 17), die Dauern verkürzt. Aufgrund der akustischen Analysen können die Vokalrealisierungen hinsichtlich ihrer Reduktion als geordnet betrachtet werden. Sie sind zu Reduktionsstufen zusammengefasst.

Die Perzeption der Reduktionsstufen gespannter Vokale beruht zuerst auf einem Qualitäts- und Dauerverlust. Ab der Reduktionsstufe, deren Vokalrealisierungen mit einer ungespannten Qualität assoziiert werden, bewirken qualitative Unterschiede die Wahrnehmung weiterer Reduktionsstufen. Auch zwischen nicht reduzierten ungespannten Vokalen und stärker reduzierten Lautrealisierungen besteht in erster Linie ein Qualitätsunterschied, wie auch die geringeren Korrelationen zwischen Dauer und perzipierter Reduktionsstärke der ungespannten Vokale im Vergleich zu den der gespannten belegen. Ein Vergleich der absoluten Korrelationswerte von spektralen Parametern und Reduktionsstufen sowie von Dauer und Reduktionsstufen deutet darauf hin, dass im allgemeinen, spektrale Parameter einen stärkeren Einfluss auf die wahrgenommenen Reduktionsstärken besitzen als Dauer. Die Ausnahme ist /a:/, hier ist die Dauer das primäre Korrelat der perzipierten Reduktionsstärke.

Kann hieraus geschlossen werden, dass die Vokalreduktion primär auf qualitativen Änderungen beruht? Nach Lindblom (1963) ist die Dauer der bestimmende Faktor für Vokalreduktion.

---

<sup>5</sup> Es werden die Hz-skalierten Formantwerte genommen, da die sprecher- und lautnormierten Werte (der Mittelwert jeder Vokalkategorie eines Sprechers wurde auf 0 gesetzt; vgl. Abschnitt 5.2.2) für eine Darstellung der Reduktionsstufen in einer Formantkarte ungeeignet sind.

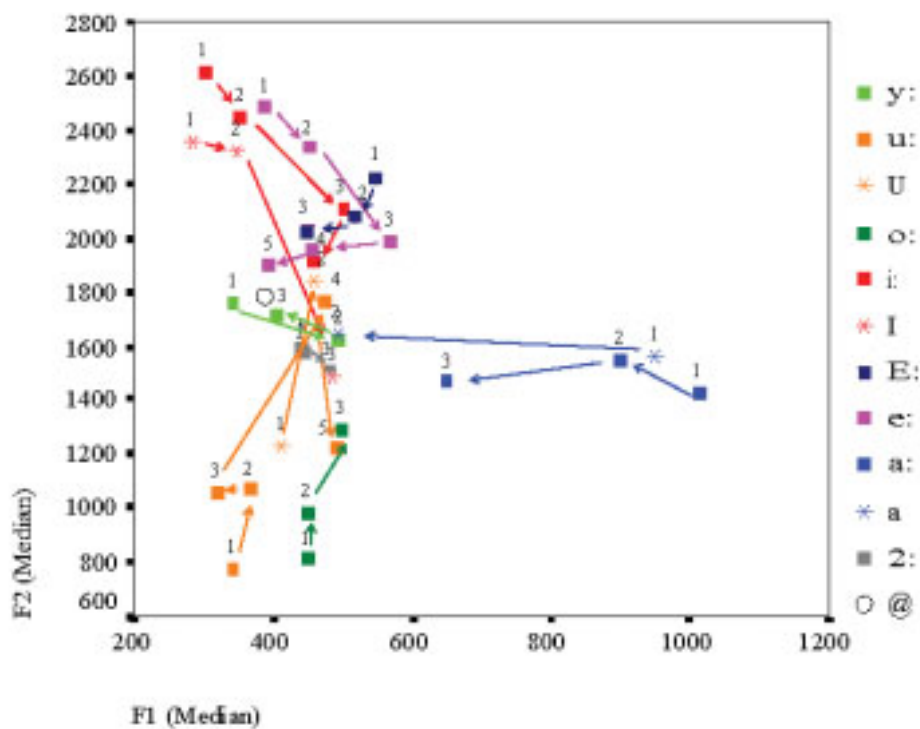


Abbildung 17: Median (Hz-skalierte Werte) des ersten (x-Achse) und zweiten Formanten (y-Achse) der Reduktionsstufen einzelner Vokalkategorien (SAMPA-Notation; 1 entspricht RED1, 2 steht für RED2, usw.) und /ə/ von SPR1.

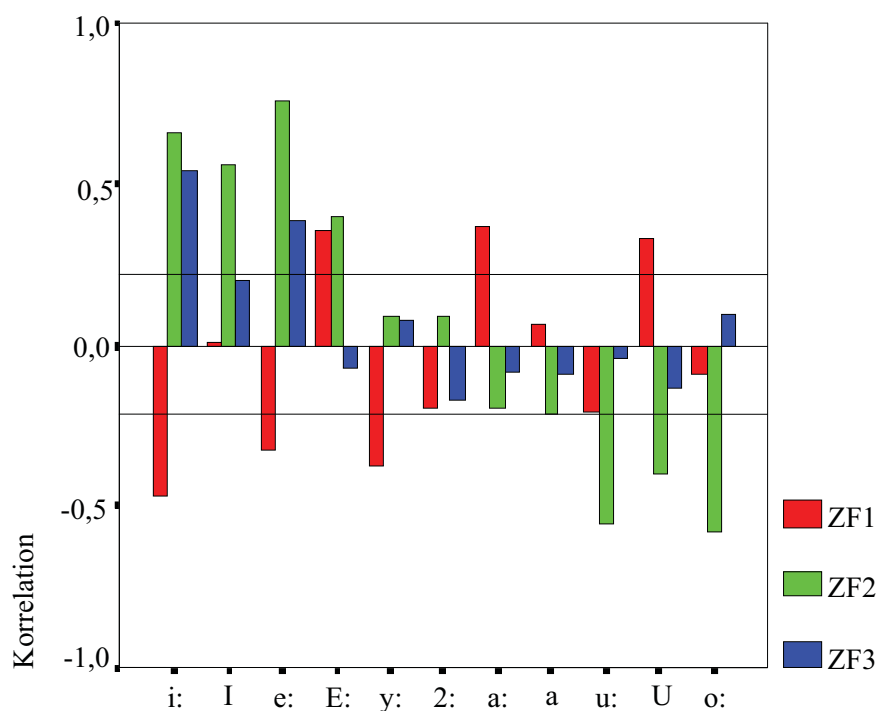


Abbildung 18: Korrelationen zwischen Lautdauer und Formantwerten für Vokalkategorien (SAMPA-Notation) mit Reduktionsstufen. Korrelationswerte ober- bzw. unterhalb der Linien sind signifikant.

Korrelationsanalysen von Dauer und Formantwerten unabhängig von den Reduktionsstufen weisen signifikante Unterschiede auf, mit Ausnahme von /ø:/ und /a/ (Abbildung 18). Von der Dauerkürzung ist insbesondere die Lage von F2 betroffen. Weiterhin bestehen zwischen den gespannten langen und gespannten kurzen Vokalrealisierungen (RED1 vs. RED2) von /i:/ und /e:/ Spektral- und Dauerunterschiede. Insgesamt sprechen die Ergebnisse für einen Einfluss der Dauer auf die Änderung des Artikulationsortes, was im Sinne eines *target undershoot* interpretiert werden kann. Durch Reduktion werden Lautdauern und spektrale Eigenschaften von Vokalen beeinflusst. Eine Dauerverkürzung bewirkt qualitative Unterschiede, die Dauer selbst ist aber für die Einteilung der Vokalrealisierungen in Reduktionsstufen weniger relevant als die spektralen Unterschiede.

## 7.2 Reduktion gespannter und ungespannter Vokale – ein Vergleich

Im vorherigen Abschnitt wurde gezeigt, dass sich die einzelnen Reduktionsstufen vor allem durch ihre spektralen Eigenschaften und ihre Dauern voneinander unterscheiden und mit zunehmender Reduktion ein Zentralisierungseffekt bezüglich der Formanten bzw. eine Abnahme der Dauer zu verzeichnen ist. Die Reduktion der Vokale manifestiert sich in den akustischen Parametern.

Die Perzeptionsexperimente zeigen, dass bei allen acht gespannten Vokalen mehrere Reduktionsstufen zuverlässig perzipiert werden, bei den ungespannten Vokalen gilt dies nur für die an den Eckpunkten des Vokaldreiecks liegenden Laute /u/, /i/ und /a/ (Kapitel 6). Die Anzahl der Reduktionsstufen der ungespannten Vokale ist geringer als die der aus phonologischer Sicht in Opposition stehenden gespannten Vokale (/i:/ vs. /i/, /u:/ vs. /u/, /a:/ vs. /a/). Der Qualitätsunterschied zwischen den Vokalpaaren wird durch das Merkmal „Gespanntheit“ beschrieben, das in einem unmittelbaren Zusammenhang mit der Vokalreduktion steht. Reduzierte gespannte Vokale werden mit einer ungespannten Lautqualität wahrgenommen. Weiterhin sind die Qualitäten der Prototypen einzelner Reduktionsstufen der in Opposition stehenden Vokale vergleichbar. Es ist anzunehmen, dass die Vokale dieser Paare einen ähnlichen Reduktionsprozess durchlaufen. Im Gegensatz hierzu stehen die Ergebnisse der ungespannten Vokale der Paare /y:/-/ʏ/, /e:/-/ɛ/ bzw. /ɛ:/-/ɛ/, /ø:/-/œ/ und /o:/-/ɔ/. Für diese Vokale konnten keine Reduktionsstufen gefunden werden.

Dies wirft die Frage nach der Rechtfertigung der Gegenüberstellung von gespannten und ungespannten Vokalen auf. Wie bereits in Abschnitt 4.3 beschrieben, ist die phonologische Paarbildung von /e:/-/ɛ/, /u:/-/u/ und /ø:/-/œ/ aus perzeptiver Sicht nicht vertretbar. Aufgrund ihrer akustischen Eigenschaften und perzipierten Qualitäten ergeben sich die Vokalpaare /e:/-/ɪ/, /o:/-/ʊ/ und /ø:/-/ʏ/. Ein Vergleich mit den wahrgenommenen Vokalqualitäten der Prototypen einzelner Reduktionsstufen spricht für eine Gegenüberstellung von /e:/-/ɪ/ und /o:/-/ʊ/ aufgrund der Beurteilung des Prototyps einer Reduktionsstufe von /e:/ mit [ɪ] und von /o:/ mit [ʊ/ɔ] (s. Abschnitt 6.1.4). Dagegen sprechen die vergleichbar wahrgenommenen Qualitäten der Prototypen der

übrigen Reduktionsstufen von /i:/-/ɪ/ und /u:/-/ʊ/. Hinsichtlich der Qualitäten der Prototypen zeigen /i:/ und /ɪ/ mehr Übereinstimmungen als /e:/ und /ɛ/, für die Paarbildungen von /u:/-/ʊ/ vs. /o:/-/ʊ/ können aufgrund der Qualitätsbeurteilungen keine Aussagen zu Gunsten eines Paares gemacht werden. Da kein Prototyp von /ø:/ mit einer /ʏ/-Qualität wahrgenommen wurde, wird /ʏ/ nicht als ungespannter Vokalpartner zu /ø:/ betrachtet.

Im folgenden wird untersucht, inwiefern sich die Vokale einzelner Reduktionsstufen mit gleich beurteilter Prototypenqualität hinsichtlich ihrer akustischen Realisierungen entsprechen. Aufgrund der Ähnlichkeitsbeziehung zwischen Prototyp und zugeordneten Vokalrealisierungen ist anzunehmen, dass sich Vokale hinsichtlich des Gespanntheitsmerkmals unterscheiden, wenn sich die akustischen Realisierungen der Reduktionsstufen der gespannten und der ungespannten Vokale, die mit demselben Symbol transkribiert wurden, gleichen. So stellt sich beispielsweise die Frage, ob die akustischen Realisierungen der Vokale der zweiten Reduktionsstufe von /i:/ (/i:/ RED2) und der ersten Reduktionsstufe von /ɪ/ (/ɪ/ RED1), deren Prototypen mit einer [ɪ]-Qualität perzipiert wurden, vergleichbar sind. Durch die akustischen Betrachtungen sollen zudem Antworten gefunden werden, warum für die ungespannten Vokale /ʏ/, /ɛ/, /œ/ und /ɔ/ keine reliable Zuordnung zu den Prototypen erfolgte (vgl. Abschnitt 6.2).

### 7.2.1 Datenmaterial

Als Basis dieser Untersuchung dienen die in Abschnitt 7.1 beschriebenen Daten. Zusätzlich werden die ungespannten Vokale, für die keine Reduktionsstufen gefunden wurden (/œ/ n = 43, /ɛ/ n = 29, /ʏ/ und /ɔ/ n = 45), in das Datenmaterial aufgenommen. Für den Vergleich der akustischen Realisierungen der Reduktionsstufen verschiedener Vokalkategorien werden die sprechernormierten Formantwerte (ZSPF1, ZSPF2, ZSPF3) und Dauern (ZSPLAUDUR) genommen. Aufgrund ihrer geringfügigen Relevanz für die Perzeption reduzierter Vokalqualitäten wird die Grundfrequenz in dieser Analyse nicht berücksichtigt. Da nicht die Formantwerte selbst, sondern das Verhältnis zwischen den Formanten für die wahrgenommene Vokalqualität entscheidend ist (s. Kapitel 3), werden zusätzlich die Abstände zwischen den sprechernormierten Werten von F1 und F2 (ZSPF21) sowie von F2 und F3 (ZSPF32) berechnet. Die Werte der Parameter (Anzahl, Median sowie erstes und drittes Quartil) sowie die Ergebnisse der Signifikanzüberprüfungen der akustischen Vergleichbarkeit von Vokalen mit gleich wahrgenommenen Qualitäten sind in Anhang D aufgeführt.

### 7.2.2 Vokale mit vergleichbar perzipierter ungespannter Qualität

Im folgenden werden zuerst die akustischen Parameter der Reduktionsstufen von Vokalen untersucht, deren Prototypen mit vergleichbarer Vokalqualität perzipiert wurden. So besitzen beispielsweise /i:/, /ɪ/ und /e:/ jeweils eine Reduktionsstufe, die mit einer [ɪ]-Qualität wahrgenommen wurde. Dann werden die akustischen Realisierungen der Reduktionsstufen der gespannten Vokale, die eine ungespannte Qualität aufweisen, mit denen der jeweiligen

ungespannten Vokale verglichen, die keine Reduktionsstufe besitzen. Die in diesem Abschnitt referierten Formantwerte (Median der z-Werte) sind in Form von Formantkarten in Abbildung 19 und Abbildung 20 dargestellt. Abbildung 21 zeigt die Dauern (Median der z-Werte) für die einzelnen Reduktionsstufen in Abhängigkeit der Vokalkategorie.

#### /i:/ - /ɪ/ - /e:/

Der Vergleich der Reduktionsstufen von /i:/ vs. /ɪ/ hinsichtlich der sprechernormierten Parameter zeigt, dass sich die Reduktionsstufen mit ähnlich wahrgenommener Qualität signifikant voneinander unterscheiden (Tabelle D-6). /i:/ RED2 vs. /ɪ/ RED1 ( $\rightarrow$  [i]); wobei  $\rightarrow$  für *Prototypen der Reduktionsstufen wurden wahrgenommen als* steht) weisen im Verhältnis zu den anderen Vergleichspaaren Unterschiede in den Abständen zwischen F2 und F3 auf. Die absoluten Werte der Formanten (F1, F2, F3) selbst hingegen sowie die Dauern der Vokale sind vergleichbar. Es kann davon ausgegangen werden, dass die akustische Realisierung von /i:/ RED2 und /ɪ/ RED1 sehr ähnlich ist.

/i:/ RED3 vs. /ɪ/ RED2 ( $\rightarrow$  [ɪ]) unterscheiden sich bezüglich der Abstände zwischen den ersten beiden Formanten sowie der Werte von F1 und F3. Signifikante Dauerunterschiede sind nicht vorhanden. Für /e:/ RED5 ( $\rightarrow$  [ɪ]) und /i:/ RED3 finden sich geringfügige Unterschiede ( $\alpha = 0,045$ ) in den Abständen zwischen F2 und F3. Obwohl sich die absoluten Werte der ersten beiden Formanten unterscheiden, ist der Abstand zwischen beiden Formanten von /e:/ RED5 und /i:/ RED3 vergleichbar. /e:/ RED5 und /ɪ/ RED2 unterscheiden sich im Abstand von F1 zu F2. /e:/ RED5 ist im Verhältnis zu /i:/ RED3 bzw. /ɪ/ RED2 durch eine stärkere Zentralisierung von F2 sowie eine kürzere Dauer gekennzeichnet (Tabelle D-6). Die Ergebnisse weisen darauf hin, dass sich /e:/ RED5 und /i:/ RED3 hinsichtlich ihrer spektralen Eigenschaften am ähnlichsten sind und /i:/ RED3 und /ɪ/ RED2 bezüglich ihrer Lautdauern.

#### /u:/ - /ʊ/ - /o:/ - /ɔ/

Der Vergleich der spektralen Parameter von /u:/ RED2 vs. /ʊ/ RED1 ( $\rightarrow$  [ʊ]) deutet nur auf einen Unterschied hinsichtlich der Werte von F1 hin (Tabelle D-7). Der Abstand zwischen F1 und F2 ist jedoch vergleichbar. /u:/ RED2 vs. /ʊ/ RED1 können hinsichtlich ihrer akustischen Eigenschaften als ähnlich betrachtet werden.

Da die Lautqualität von /o:/ RED2 nicht eindeutig bestimmt ist – einige Probanden assoziieren die Qualität des Prototyps dieser Reduktionsstufe mit [ʊ], andere mit [ɔ] (vgl. Abschnitt 6.1.4) – werden die Parameter von /u:/ RED2 und /ʊ/ RED1 mit denen von /o:/ RED2 verglichen. Es zeigen sich signifikante Unterschiede in den Abständen von F1 zu F2 (Tabelle D-7). Der F2 von /ʊ/ RED1 ist im Gegensatz zu dem von /o:/ RED2 stärker zentralisiert. Zudem unterscheiden sich beide Vergleichspaare in ihren Dauern; die Lautdauern von /o:/ RED2 sind reduzierter als die von /u:/ RED2 bzw. /ʊ/ RED1. Der ungespannte Vokal /ɔ/, der keine Reduktionsstufen besitzt, weist in

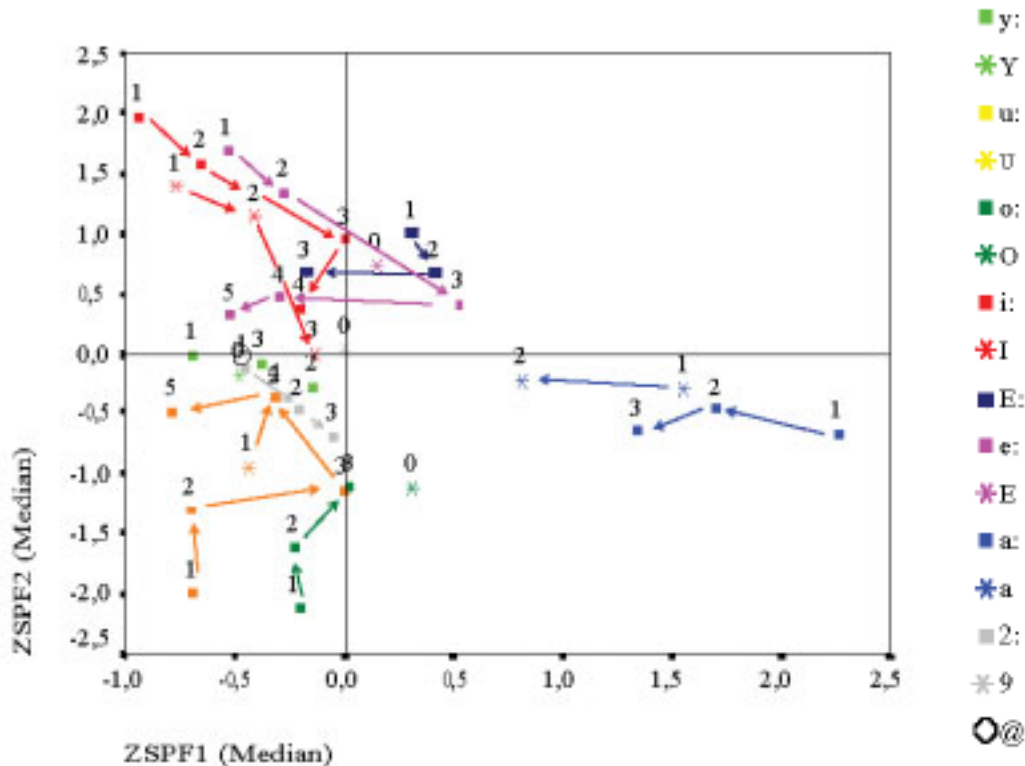


Abbildung 19: Erster (x-Achse) und zweiter Formant (y-Achse; Median der z-Werte) in Abhängigkeit der Reduktionsstufen einzelner Vokalkategorien (SAMPA-Notation; mit 0 sind die Vokalkategorien gekennzeichnet, die keine Reduktionsstufe besitzen; 1 entspricht RED1, 2 steht für RED2, usw.).

Bezug auf /u:/ RED2 bzw. /u/ RED1 Unterschiede hinsichtlich der Abstände zwischen F1 und F2 auf, was auf die Lage von F1 zurückzuführen ist. Die Lautdauern von /u/ RED1 sind im Verhältnis zu /ɔ/ kürzer.

#### /u:/ - /y:/ - /Y/

Weiterhin zeigt der Vergleich von /u:/ RED4 (→ [Y]) mit /y:/ RED2 und /Y/ keine Abweichungen zum ungespannten Vokal /Y/ hinsichtlich der spektralen Parameter (Tabelle D-7). Hingegen unterscheiden sich die Abstände zwischen F2 und F3 von /u:/ RED4 und /y:/ RED2. Die Lautdauern von /u:/ RED4 sind kürzer als die der Vergleichspartner.

#### /a:/ - /a/

Für /a:/ RED2 vs. /a/ RED1 (→ [a]) ergibt sich ein Unterschied bezüglich der Werte von F2 (Tabelle D-8). Die Abstände von F1 zu F2 sowie von F2 zu F3 unterscheiden sich nicht, so dass die akustischen Realisierungen beider Reduktionsstufen vergleichbar sind.

#### Ungespannte Vokale ohne Reduktionsstufen

Vergleiche der gespannten Vokale mit den ungespannten Vokalen, für die keine Reduktionsstufen diskriminiert wurden, zeigen (Tabelle D-9), dass die Abstände zwischen F1 und F2 der Paare /e:/



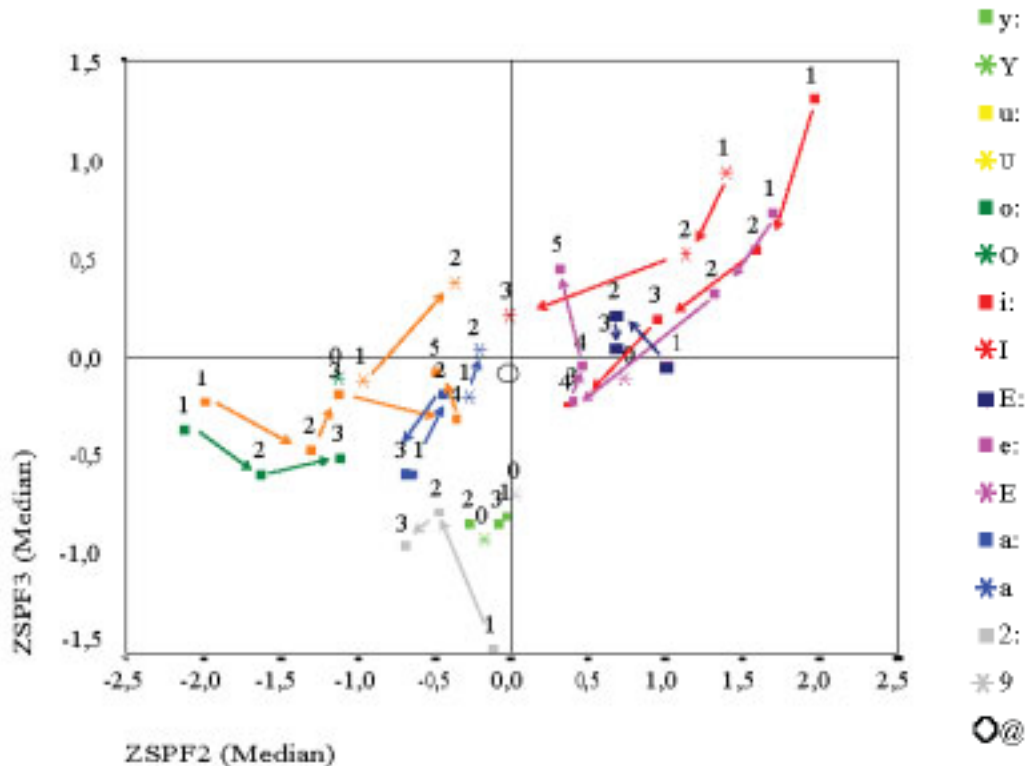
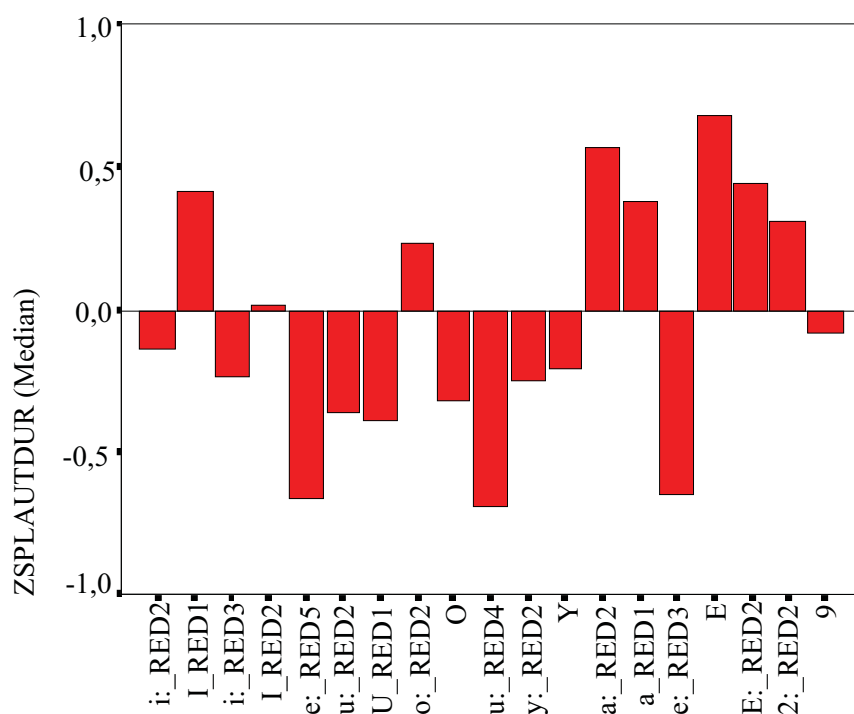


Abbildung 20: Zweiter (x-Achse) und dritter Formant (y-Achse; Median der z-Werte) in Abhängigkeit der Reduktionsstufen einzelner Vokalkategorien (SAMPA-Notation; mit 0 sind die Vokalkategorien gekennzeichnet, die keine Reduktionsstufe besitzen; 1 entspricht RED1, 2 steht für RED2, usw.).

RED3 ( $\rightarrow$  [ɛ]) vs. /ɛ/ und /ø:/ RED2 ( $\rightarrow$  [œ]) vs. /œ/ voneinander abweichen. Im ersten Fall ist die Abweichung auf die Lage von F1 zurückzuführen, im zweiten Fall auf die von F2. Die Formanten der ungespannten Vokale sind stärker zentralisiert als die der reduzierten gespannten Vokale. Bei /e:/ RED3 vs. /ɛ/ ist ein zusätzlicher Dauerunterschied festzustellen. Hingegen sind die akustischen Realisierungen von /ɛ:/ RED2 und /ɛ/ vergleichbar. /o:/ RED2 ( $\rightarrow$  [u] bzw. [ɔ]) vs. /ɔ/ unterscheiden sich hinsichtlich ihrer absoluten Formantwerte (F1, F2, F3), jedoch finden sich keine Unterschiede zwischen den Lautdauern und zwischen den Formantabständen. Für /y:/ RED2 vs. /y/ kann nur ein Dauerunterschied beobachtet werden. Die Lautdauern von /y/ sind geringer als diejenigen der reduzierten gespannten Vokale.

Die Analysen zeigen, dass die akustischen Realisierungen der gespannten reduzierten Vokale und der ungespannten (nicht reduzierten) Vokale mit gleich wahrgenommener Prototypenqualität bei den Vokalpaaren /a:/-/a/, /ɛ:/-/ɛ/, /u:/-/u/, /o:/-/ɔ/ und /y:/-/y/ vergleichbar sind. Die jeweiligen Lautrealisierungen entsprechen sich hinsichtlich ihrer spektralen Eigenschaften und bezüglich ihrer Lautdauern mit Ausnahme von /y:/-/y/. Die spektralen Eigenschaften von /y/ stimmen aber nicht nur mit denen eines reduzierten /y:/, sondern auch mit den Eigenschaften eines stark reduzierten /u:/ überein. Hieraus ergibt sich die Frage, ob /y/ als bezüglich der Gespanntheit in Opposition stehend zu /u:/ betrachtet werden kann.



**Abbildung 21: Lautdauern (Median der z-Werte) für einzelne Reduktionsstufen von Vokalen mit vergleichbar wahrgenommener Lautqualität (x\_REDy; wobei x für die Vokalkategorie (SAMPA-Notation) steht und y für die Reduktionsstufe).**

Den in phonologischer Opposition stehenden Vokalen wird eine phonetische Ähnlichkeit zugeschrieben. Nach Ungeheuer (1962b, 1964) besteht eine perzeptive Ähnlichkeit von /ʏ/ und /u:/. Die Beurteilung der Ähnlichkeit von Lauten scheint aber sprachspezifisch zu sein, so besitzen /ʏ/ und /u:/ für dänische Probanden keine Ähnlichkeit (Diskussionsbeitrag von Fischer-Jørgensen, in Ungeheuer, 1977b:82 f). Fischer-Jørgensen führt die Beurteilung zurück auf die im Deutschen häufig vorkommende grammatische Alternation von u-ü (z.B. <Mutter> - <Mütter>, <wurde> – <würde>), die orthographische Ähnlichkeit der Laute und die Bezeichnung von <ü> als u-Umlaut. Das vorliegende Ergebnis zeigt einen Zusammenhang zwischen der von Ungeheuer definierten auditiven Qualität (*polare* Vokalqualität) und akustischer Realisierung von Reduktion auf. Die Frage nach einer Gegenüberstellung von /u:/ vs. /ʏ/ muss dahingehend erweitert werden, ob sich die beiden Vokale ähnlich genug sind. Die vergleichbaren akustischen Realisierungen von gespannten reduzierten /u:/ (→ [ʊ]) und ungespannten (nicht reduzierten) /u/ sowie die eingeschränkte Distribution (nur im alveolaren Kontext) von reduzierten /u:/ mit [ʏ]-Qualität sprechen für eine größere Ähnlichkeit zwischen /u:/ und /ʊ/ als zwischen /u:/ und /ʏ/.

Die Ergebnisse zeigen, dass die phonologische Gegenüberstellung der Vokalpaare /a:/-/a/, /ɛ:/-/ɛ/, /u:/-/ʊ/, /o:/-/ɔ/ und /y:/-/ʏ/ auch aus phonetischer Sicht gerechtfertigt ist. Eine Frage, die sich hier anschließt, ist die nach der Art des Unterschieds zwischen den Vokalpaaren. Wie bereits in Abschnitt 7.1.1 beschrieben, unterscheiden sich gespannte Vokale von ihren reduzierten Varianten mit wahrgenommener ungespannter Lautqualität ([a:]–[a], [u:]–[ʊ], [o:]–[ɔ], [y:]–[ʏ]) durch ihre

spektralen Eigenschaften und Dauern, mit Ausnahme von [ɛ:]–[ɛ], hier sind nur Dauerunterschiede zu finden. Es kann daher davon ausgegangen werden, dass sich [ɛ:] und [ɛ] nur bezüglich ihrer Quantität unterscheiden. Welche Relevanz Quantität und Qualität für die Unterscheidung der jeweiligen Vokalpaare haben, lässt sich aus den akustischen Unterschieden zwischen den Reduktionsstufen schließen (Abschnitt 7.1.1). Es ist anzunehmen, dass die perzeptive Unterscheidung der „a“-Laute primär auf Quantitätsunterschieden beruht, während die der „u“- und „o“-Laute von den qualitativen Eigenschaften abhängt.

Auf die Frage, ob /ɪ/ /e:/ oder /i:/ zugeordnet werden soll, gibt die akustische Analyse keine eindeutige Antwort. /ɪ/ besitzt weder zu den reduzierten Vokalen mit perzipierter ungespannter Qualität von /i:/ noch zu denjenigen von /e:/ vergleichbare spektrale Eigenschaften. Hingegen entsprechen sich die reduzierten Varianten der gespannten Vokale /i:/ und /e:/ am meisten. Wird die Übereinstimmung zwischen den wahrgenommenen Vokalqualitäten der einzelnen Reduktionsstufen von /i:/ und von /ɪ/ und die akustische Vergleichbarkeit der kurzen gespannten Realisierungen von /i:/ und /ɪ/ berücksichtigt, so sprechen die Ergebnisse für eine stärkere Ähnlichkeitsbeziehung zwischen /i:/ und /ɪ/ als zwischen /e:/ und /ɪ/, was eine phonologische Gegenüberstellung von /i:/ vs. /ɪ/ unterstützt. Auch hier kann aufgrund der größeren Relevanz der spektralen Eigenschaften bei der Perzeption der Reduktionsstufen (vgl. Abschnitt 7.1.1) davon ausgegangen werden, dass in erster Linie ein qualitativer Unterschied zwischen den beiden Vokalen besteht.

Für die Gegenüberstellung von /ø:/ und /œ/ liefert der akustische Vergleich keine Hinweise. Zur Rechtfertigung kann hier nur die perzipierte Lautqualität der reduzierten Vokalrealisierungen von /ø:/ herangezogen werden. Im folgenden Abschnitt werden Ursachen gesucht, die das Fehlen von Reduktionsstufen bei den Vokalen /ɛ/, /ɔ/, /œ/ und /ʏ/ erklären.

### 7.2.3 *Perzipierte Reduktionsstufen und akustische Ähnlichkeit zum [ə]*

Wie bereits oben beschrieben, stimmen die spektralen Eigenschaften der gespannten reduzierten Vokalrealisierungen von /ɛ:/, /o:/ und /y:/ mit denen der ungespannten Vokale /ɛ/, /ɔ/ bzw. /ʏ/ überein. Im Vergleich zu den ungespannten Vokalen wird jedoch bei den gespannten Vokalen eine weitere Reduktionsstufe nach den ungespannt wahrgenommenen Lautqualitäten unterschieden. Die Vokalrealisierungen von /œ/ hingegen weisen Unterschiede zu denjenigen der gespannten reduzierten Realisierungen von /ø:/ auf, wobei der F2-Wert von /œ/ zentraler liegt als der des gespannten reduzierten /ø:/. Die an den Eckpunkten des Vokaldreiecks befindlichen ungespannten Vokale /ɪ/, /ʊ/ und /a/ zeigen wie die in Opposition stehenden gespannten Vokale /i:/, /u:/ und /a:/ ebenfalls eine weitere Reduktionsstufe mit wahrgenommener zentralisierter Qualität auf.

Es stellt sich die Frage nach einer Erklärung für das Fehlen der Reduktionsstufen bei den vier ungespannten Vokalen /ʏ/, /ɛ/, /œ/ und /ɔ/. Möglich ist, dass die Dauer der Stimuli die Zuordnung

zu den Prototypen beeinflusst. Ist die zeitliche Ausdehnung der ungespannten Vokale ohne Reduktionsstufen wesentlich geringer als die der ungespannten bzw. gespannten Vokale mit Reduktionsstufen, können qualitative und/oder quantitative Änderungen nicht zu einer Diskriminierbarkeit von verschiedenen Lautqualitäten führen. Durch die kürzere Dauer wird eine schwa-ähnliche Qualität perzipiert. Wie zu erwarten, besitzen gespannte Vokale im Durchschnitt eine längere Dauer (80,1 ms) als ungespannte Vokale (57,6 ms), jedoch ergibt sich kein Dauerunterschied zwischen ungespannten Vokalen mit Reduktionsstufen (56,1ms) und ohne Reduktionsstufen (59,3 ms). Die durchschnittliche Dauer von /ɪ/, /ɛ/, /œ/ und /ɔ/ kann die fehlende Diskriminierung reduzierter Lautqualitäten nicht erklären.

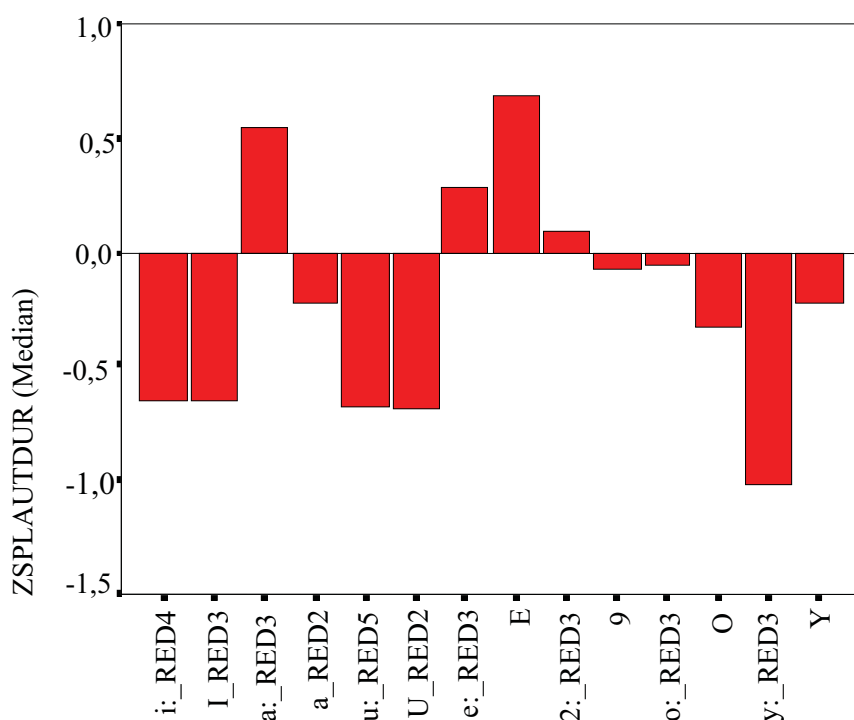
Eine andere Erklärung ist, dass der perzipierte Unterschied zum /ə/ der ungespannten Vokale ohne Reduktionsstufen geringer ist als der der Vokale mit Reduktionsstufen. Diese Vokale werden als schon sehr reduziert empfunden, so dass keine weiteren Reduktionsstufen diskriminiert werden. Die Annahme wird durch die Ergebnisse von Ungeheuer (1962b, 1964) gestützt. Die Laute [ɛ], [œ] und [ɔ] weisen eine perzipierte Ähnlichkeit zu [ə] auf – Ungeheuer bezeichnet diese auditive Qualität als *zentrale* Vokalqualität. Dies deutet darauf hin, dass bei den ungespannten Vokalen der akustische Abstand zum /ə/ so gering ist, dass Hörer keine weiteren Reduktionsstufen differenzieren.

Im folgenden wird untersucht, wie sich die akustischen Realisierungen der ungespannten Vokale von denen der gespannten unterscheiden und welche Relevanz die spektrale Distanz zum /ə/ für die Diskriminierung von Reduktionsstufen besitzt. Das akustische Abstandsmaß wird hier als der Euklidische Abstand (vgl. Abschnitt 6.1.1, Gleichung 3) zwischen den Formantwerten der Vokalrealisierungen und den durchschnittlichen Formantwerten von /ə/ definiert.

#### Gespannte und ungespannte Vokale mit einer [ə]- oder zentralisierten Qualität

Vergleiche der stark reduzierten gespannten und ungespannten Vokale (/i:/ RED4 (→ [ə] bzw. [i]) vs. /ɪ/ RED3 (→ [ə]), /u:/ RED5 (→ [ə]) vs. /ʊ/ RED2 (→ [ü]) und /a:/ RED3 vs. /a/ RED2 (→ [a]) weisen auf Unterschiede zwischen den Abständen der ersten beiden Formanten hin, wobei diese Abweichungen bei allen Vergleichspaaren auf eine stärkere Zentralisierung von F2 der ungespannten Vokale zurückgeführt werden kann (Abbildung 19; Tabelle D-10). Hinsichtlich ihrer Lautauern unterscheiden sich die stark reduzierten Vokalrealisierungen von /a:/ vs. /a/. Die Dauern von /a/ RED2 sind stärker reduziert als die von /a:/ RED3 (Abbildung 22).

Die durchschnittlichen Abstände zum Zentralvokal weisen darauf hin, dass mit abnehmender Distanz zum /ə/ die Lautrealisierungen innerhalb der jeweiligen Vokalkategorie als zunehmend reduziert perzipiert werden (Abbildung 23). Die mittleren Abstände sind für die stark reduzierten Vokalrealisierungen von /i:/ und /ɪ/ sowie von /u:/ und /ʊ/ vergleichbar. Obwohl F2 der stark reduzierten Lautrealisierungen der ungespannten Vokale eine zentralere Lage aufweist als F2



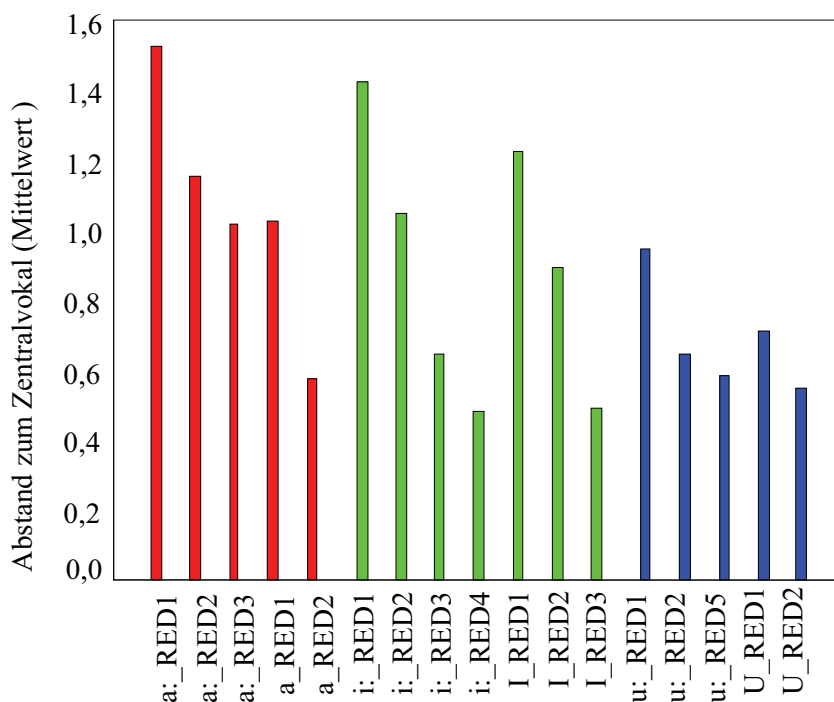
**Abbildung 22: Lautdauern (Median der z-Werte) für einzelne Reduktionsstufen stark reduzierter gespannter und ungespannter Vokale sowie ungespannter Vokale ohne Reduktionsstufen (x\_REDy; wobei x für die Vokalkategorie (SAMPA-Notation) steht und y für die Reduktionsstufe).**

der gespannten, sind die mittleren Abstände zum /ə/ der reduzierten Varianten von /ɪ/ und /ʊ/ vergleichbar zu denjenigen der gespannten Vokalpartner /i:/ und /u:/.

Für die „a“-Laute hingegen zeigt sich, dass die reduzierten Realisierungen von /a/ eine geringere Distanz zum /ə/ aufweisen als diejenigen von /a:/. Es entsprechen sich die stark reduzierten Lautrealisierungen von /a:/ und die nicht reduzierten Realisierungen von /a/ hinsichtlich ihrer spektralen Reduktion. Im Gegensatz zu /a/ wird bei /a:/ jedoch keine weitere Reduktionsstufe diskriminiert, d.h., dass die spektralen Unterschiede innerhalb der stark reduzierten Lautrealisierungen von /a:/ für die Unterscheidung einer weiteren Gruppe von stärker reduzierten Realisierungen perceptiv nicht relevant sind. Berechnungen des Abstands zum Zentralvokal der jeweiligen Prototypen der stark reduzierten Reduktionsstufen zeigen, dass beide Prototypen eine ähnlich starke spektrale Reduktion aufweisen; ihre Distanzen sind vergleichbar (für den Prototypen von /a/ RED2 0,907 und von /a:/ RED3 0,925).

#### Gespannte Vokale mit einer [ə]- oder zentralisierten Qualität und ungespannte Vokale ohne Reduktionsstufen

Werden die akustischen Realisierungen der stark reduzierten gespannten Vokale mit denen der ungespannten Vokale ohne Reduktionsstufen verglichen (Tabelle D-11), so zeigen sich für /ɛ:/ RED3 (→ [ĕ] bzw. [ə]) vs. /ɛ/ und /o:/ RED3 (→ [ə]) vs. /ɔ/ nur geringfügige Abweichungen hinsichtlich der spektralen Eigenschaften. Zwischen den F1-Werten von /ɛ:/ RED3 und /ɛ/ besteht

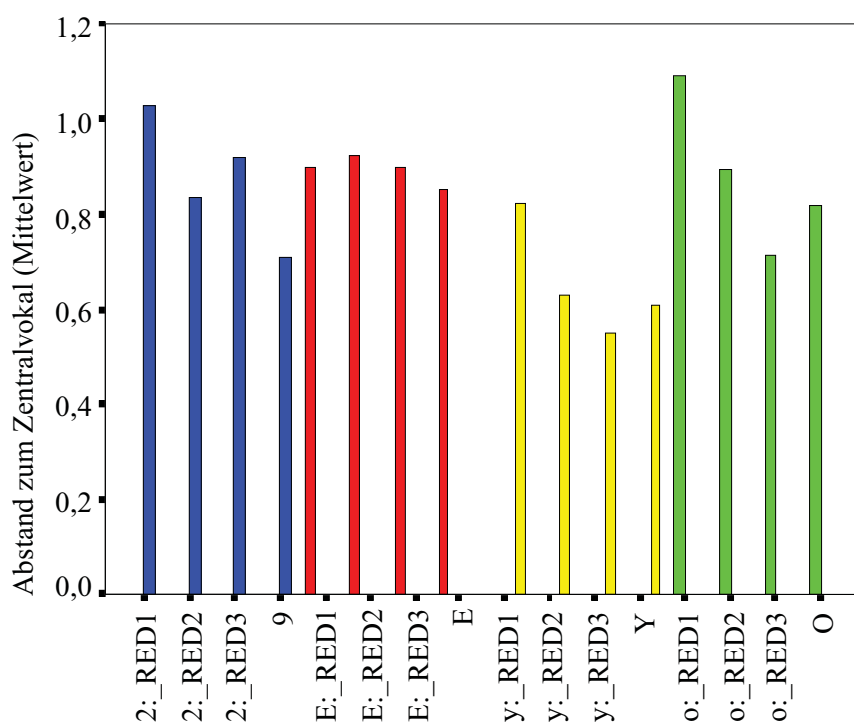


**Abbildung 23: Durchschnittliche spektrale Distanz zum Zentralvokal der Reduktionsstufen gespannter und ungespannter Vokale (Euklidische Distanzen wurden über die z-Werte berechnet; in Opposition stehende Vokale sind mit einer Farbe markiert, Reduktionsstufen der Vokale sind durch x\_REDy gekennzeichnet, wobei x für die Vokalkategorie (SAMPA-Notation) steht und y für die Reduktionsstufe).**

ein Unterschied ( $\alpha = 0,034$ ), jedoch sind die Abstände zwischen den ersten beiden Formanten vergleichbar. Für /o:/ RED3 vs. /ɔ/ finden sich einerseits Abweichungen hinsichtlich der Abstände zwischen F2 und F3, andererseits unterscheiden sich die Formantwerte selbst nicht. /œ/ ist im Vergleich zu /ø:/ RED3 ( $\rightarrow$  [œ] bzw. [ə]) durch eine stärkere Zentralisierung von F2 gekennzeichnet. Die drei Vergleichspaare weisen keinen Dauerunterschied auf.

Die spektralen Eigenschaften von /ɤ/ hingegen stimmen nicht nur mit denen von /y:/ RED2 überein, sondern auch mit denen von /y:/ RED3. Dass zwischen /y:/ RED2 und RED3 kein spektraler Unterschied besteht, zeigte bereits die Analyse der Reduktionsstufen (vgl. Abschnitt 7.1.1). /ɤ/ besitzt aber eine längere Dauer als die stark reduzierten Realisierungen des gespannten Vokals (Abbildung 22).

Insgesamt weisen die Analysen der ungespannten Vokale ohne Reduktionsstufen nicht nur spektrale Ähnlichkeiten zu den Vokalrealisierungen der gespannten Vokale mit wahrgenommener ungespannter Qualität auf, sondern auch zu denen mit einer zentralisierten bzw. [ə]-Qualität. Bei den gespannten Vokalen /y:/ und /o:/ ist eine Zunahme der Reduktionsstärke mit einer Abnahme der spektralen Distanz zum ə/ verbunden (Abbildung 24). Für /ø:/ und /ɛ:/ trifft dies nicht zu, was durch die nur geringfügigen bzw. fehlenden spektralen Unterschiede zwischen den einzelnen Reduktionsstufen erklärt werden kann (vgl. Abschnitt 7.1.1). Die ungespannten Vokale /ɤ/ und /ɔ/



**Abbildung 24: Durchschnittliche spektrale Distanz zum Zentralvokal der Reduktionsstufen gespannter Vokale und der ungespannten Vokale ohne Reduktionsstufen (Euklidische Distanzen wurden über die z-Werte berechnet; in Opposition stehende Vokale sind mit einer Farbe markiert, Reduktionsstufen der Vokale sind durch x\_REDy gekennzeichnet, wobei x für die Vokalkategorie (SAMPA-Notation) steht und y für die Reduktionsstufe).**

liegen im Mittel zwischen den ungespannten und zentralen Lautrealisierungen der jeweiligen gespannter Vokale und /œ/ weist einen geringeren Abstand zum /ə/ auf als die stark reduzierten Realisierungen von /ø:/. Dies zeigt, dass der spektrale Abstand zum /ə/ der ungespannten Vokale zu gering ist, um noch weitere Reduktionsstufen zu diskriminieren. Der mittlere Abstand von /ɛ/ hingegen ist vergleichbar zu den Abstandswerten der drei Reduktionsstufen von /ɛ:/, also auch zu den Werten der nicht reduzierten Realisierungen von /ɛ:/. Wie bereits oben erwähnt, unterscheiden sich die einzelnen Reduktionsstufen von /ɛ:/ vor allem durch ihre Lautdauern. Die Dauern von /ɛ/ sind aber mit denen der stark reduzierten Vokalrealisierungen /ɛ:/ vergleichbar. Daher kann die Lautdauer als Erklärung für das Fehlen zuverlässig diskriminierter Reduktionsstufen für /ɛ/ angegeben werden.

Die Ergebnisse zeigen, dass eine reliable Diskriminierbarkeit von Reduktionsstufen von der spektralen Distanz zum /ə/ abhängt. Jedoch kann aufgrund der absoluten spektralen Abstandswerte der unreduzierten Vokalrealisierungen nicht auf die Anzahl der Reduktionsstufen geschlossen werden. Beispielsweise weisen die nicht reduzierten Realisierungen von /y:/ eine vergleichbare Entfernung zum /ə/ auf wie diejenigen von /ɔ/ (Abbildung 24). Trotz diesem Abstand werden bei dem gespannten Vokal mehrere Reduktionsstufen perzipiert, während bei den ungespannten keine Reduktionsstufen unterschieden werden. Beim gespannten Vokal ist mehr Reduktionspotential aufgrund der längeren Dauer vorhanden. Jedoch geht ein Dauerverlust mit einer spektralen



Zentralisierung einher, so dass durch die Dauer allein noch nicht erklärt wird, warum bei /y:/ nach der Reduktionsstufe mit ungespannt wahrgenommener Lautqualität eine weitere Reduktionsstufe diskriminiert wird. Auch der Vergleich der spektralen Distanz der nicht reduzierten Lautrealisierungen von /u/ und /ɔ/ zeigt, dass aufgrund des Abstands zum /ə/ keine Aussagen über die zu erwartende Anzahl von Reduktionsstufen gemacht werden können. In beiden Fällen handelt es sich um kurze ungespannte Vokale, so dass die Dauer als ein möglicher Einflussfaktor auf eine Erhöhung des Reduktionspotentials vernachlässigt werden kann. Obwohl /u/ eine geringere Distanz zum /ə/ aufweist als /ɔ/, werden die /u/-Varianten in zwei Reduktionsstufen eingeteilt (vgl. Abbildung 23 und Abbildung 24).

Die spektrale Distanz zum /ə/ liefert nur innerhalb von phonetisch ähnlichen Vokalkategorien (z.B. den in Opposition stehenden Vokalen) eine plausible Erklärung für das Fehlen von Reduktionsstufen, was dafür spricht, dass die Distanzen zum /ə/ in Abhängigkeit der Vokalkategorie unterschiedlich wahrgenommen werden und somit das Distanzmaß nicht immer mit der perzeptiven Skalierung übereinstimmt (vgl. Abschnitt 7.3). Bei ähnlichen Vokalkategorien führen vergleichbare Distanzen zu vergleichbaren Reduktionsstufen. Bei Vokalen, die an sich /ə/-ähnliche Lautqualitäten (/ʏ/, /ɛ/, /ɔ/, /œ/) besitzen, reichen die perzipierten Unterschiede zwischen ungespannter Lautqualität und /ə/-Qualität nicht aus, um eine weitere Reduktionsstufen reliabel zu diskriminieren.

#### 7.2.4 Zusammenfassung

Ausgangspunkt der vergleichenden akustischen Analyse der Reduktionsstufen gespannter und ungespannter Vokale war die Fragestellung, ob die phonologische Gegenüberstellung gespannter und ungespannter Vokale auch aus phonetischer Sicht gerechtfertigt ist. Die phonologische Vokalpaarbildung beruht auf einer angenommenen phonetischen Ähnlichkeitsbeziehung zwischen den jeweiligen Vokalpartnern, der qualitative Unterschied zwischen den Vokalpaaren wird mit dem Merkmal Gespanntheit beschrieben. Wie die Beurteilungen der Prototypen der Reduktionsstufen zeigen, ist die Reduktion gespannter Vokale durch einen Verlust der Gespanntheit gekennzeichnet. Reduzierte gespannte Vokale werden mit einer ungespannten Vokalqualität perzipiert. Da die Vokalrealisierungen der Reduktionsstufen sich primär durch die Formantlage und die Dauern unterscheiden, wurde untersucht inwiefern sich reduzierte gespannte Vokale mit perzipierter ungespannter Lautqualität von den nicht reduzierten ungespannten Vokalen bezüglich dieser Parameter unterscheiden.

Die Ergebnisse zeigen, dass die akustischen Eigenschaften der reduzierten Vokalrealisierungen von /a:/, /ɛ:/, /u:/, /o:/ und /y:/ ausgeprägte Ähnlichkeiten zu den in phonologischer Opposition stehenden Vokalen /a/, /ɛ/, /u/, /ɔ/ bzw. /ʏ/ aufweisen. Zudem spricht für die Gegenüberstellung von /a:/-/a/, /ɛ:/-/ɛ/, /u:/-/u/, /o:/-/ɔ/, /y:/-/ʏ/ die ähnlich perzipierten Lautqualitäten der einzelnen Reduktionsstufen der Vokale. Für /ɪ/ ist die Lage nicht so eindeutig. Jedoch weisen die akustische Vergleichbarkeit der kurzen gespannten Realisierungen der Vokalklassen /i:/ und /ɪ/ sowie die

Übereinstimmung zwischen den wahrgenommenen Vokalqualitäten der einzelnen Reduktionsstufen auf eine stärkere Ähnlichkeitsbeziehung zwischen /i:/ und /ɪ/ als zwischen /e:/ und /ɪ/ hin. Für eine phonologische Gegenüberstellung von /ø:/ vs. /œ/ bieten die akustischen Analysen keine Hinweise. Als Argument dafür kann hier nur die perzipierte ungespannte Qualität der Reduktionsstufe des gespannten Vokals gewertet werden.

Im Rahmen der phonetischen Rechtfertigung der phonologischen Gegenüberstellung stellt sich jedoch die Frage, warum bei den ungespannten Vokalen /ɛ/, /ɔ/, /ɤ/ und /œ/ keine Reduktionsstufen diskriminiert werden, obwohl bei den jeweiligen in Opposition stehenden gespannten Vokalen nach den Reduktionsstufen mit ungespannter Vokalqualität eine weitere Reduktionsstufe perzipiert wird. Die bisherigen Argumente für eine Rechtfertigung der Vokalgegenüberstellung begründen sich auf Ähnlichkeiten im Reduktionsprozess der jeweiligen Vokalpartner, so dass bei diesen ungespannten Vokalen eine weitere Reduktionsstufe zu erwarten wäre. Eine Annahme war hierbei, dass der spektrale Abstand zum Zentralvokal die Perzeption der Reduktionsstufen beeinflusst. Die Abstandsanalysen bestätigen die Annahme. Die Vokale /ɛ/, /ɔ/, /ɤ/ und /œ/ weisen kleinere Abstände zum Zentralvokal auf als die reduzierten oder stark reduzierten Lautrealisierungen ihrer gespannten Vokalpartner. Es kann davon ausgegangen werden, dass die perzipierten Unterschiede zwischen ungespannter und /ə/-Qualität nicht ausreichend sind, um eine weitere Reduktionsstufe reliabel zu diskriminieren. Dies bedeutet, dass das Fehlen von Reduktionsstufen nicht die Argumente entkräften, die für eine phonologische Gegenüberstellung von /ɛ:/ vs. /ɛ/, /o:/ vs. /ɔ/, /y:/ vs. /ɤ/ und /ø:/ vs. /œ/ sprechen.

Die phonologische Gegenüberstellung der Vokale ist somit auch aus phonetischer Sicht gerechtfertigt. Aufgrund der primären perzeptiven Relevanz der qualitätsbestimmenden Parameter für die Unterscheidung der Reduktionsstufen kann davon ausgegangen werden, dass sich die Vokalpaare /i:/-/ɪ/, /u:/-/ʊ/, /o:/-/ɔ/, /y:/-/ɤ/ und /ø:/-/œ/ durch ihre Qualität unterscheiden. Zwischen /a:/ und /a/ sowie /ɛ:/ und /ɛ/ herrscht vorrangig ein Quantitätsunterschied.

### **7.3 Akustische und perzeptive Ähnlichkeit**

Die Probanden hatten in den Perzeptionsexperimenten die Aufgabe, verschiedene Lautrealisierungen Referenzstimuli, die als Prototypen von Reduktionsstufen bezeichnet wurden, aufgrund ihrer qualitativen Ähnlichkeit zuzuordnen (vgl. Abschnitt 6.1.3). Die Frage ist, wie sich die perzipierte Ähnlichkeit im akustischen Raum manifestiert.

Nach der Theorie des Perzeptiven Magneteffekts (Kuhl, 1992; Abschnitt 3.3) werden Laute einer Vokalkategorie, die akustisch in der Nähe einer guten oder typischen Instanz (Prototyp) dieser Kategorie liegen, von dieser "angezogen". Dies bedeutet, dass akustische Unterschiede zu diesem Prototypen perzeptiv unbedeutend werden. Stimuli innerhalb einer Kategorie werden von Hörern nicht mehr diskriminiert. Das Ausmaß der Anziehungskraft eines Vokals hängt zum einem von der

akustischen Distanz zwischen ihm und dem zu vergleichenden Laut ab; je größer der Abstand ist, um so unähnlicher wird der Vokal empfunden. Zum anderen ist die Anziehungskraft eines Vokals vermindert, der eine weniger gute Instanz (kein Prototyp) der Kategorie ist. Die Diskriminierbarkeit der Vokale steht somit in inverser Beziehung zur Güte des Prototyps.

Aufgrund der Theorie des Perzeptiven Magneteffekts folgt die Annahme, dass Stimuli, die einem Prototypen zugeordnet wurden, auch die größte akustische Ähnlichkeit mit diesem Prototypen besitzen. Zur Quantifizierung der akustischen Ähnlichkeitsbeziehung von Prototypen und Stimuli wird die Euklidische Distanz gewählt (vgl. Abschnitt 6.1.1, Gleichung 3). Aufgrund der berechneten Abstände zwischen den Prototypen und den Lautrealisierungen einer Vokalkategorie wird der Prototyp bestimmt, der die geringste Distanz zu dem jeweiligen Stimulus aufweist. Diese Zuweisungen werden mit den auf den perzipierten Ähnlichkeiten basierenden mehrheitlichen Zuordnungen verglichen. Zusätzlich wird überprüft, ob die individuellen Zuweisungen der Stimuli einen Zusammenhang mit den Distanzen zu den Prototypen aufweisen und somit eine Aussage über die Güte der Prototypen ermöglichen. Die Diskriminierbarkeit der Stimuli von den Prototypen wird durch die Häufigkeit der auf der Perzeption beruhenden Zuweisungen bestimmt. Stimuli und Prototypen sind nur schlecht diskriminierbar, wenn die Stimuli häufig demselben Prototyp zugeordnet wurden, während Stimuli, die nur von wenigen Probanden zugeordnet werden, sich vom Prototypen gut unterscheiden lassen.

### 7.3.1 Untersuchungen

Die folgenden Untersuchungen basieren auf dem Datenmaterial zur Überprüfung der Sprecherunabhängigkeit von Prototypen der Vokalkategorien /i:/, /u:/ und /a:/ (vgl. Abschnitt 6.1.6). Stimuli, die aufgrund der mehrheitlichen Probandenurteile (ORL) nicht eindeutig einem Prototypen zugeordnet wurden, werden ausgeschlossen.

Die Euklidischen Abstände der Stimuli zu den Prototypen werden aufgrund der Formantfrequenzen, der Grundfrequenz, der relativen Energieverteilung in den Frequenzbändern und der Dauer bestimmt. Alle Parameter werden zu gleichen Anteilen berücksichtigt. Es werden die sprecher- und lautnormierten Werte verwendet (vgl. Abschnitt 5.2). Um mögliche Einflüsse der unterschiedlichen Verteilungen der Distanzen von Prototypen und Stimuli bei der Bestimmung des minimalen Abstandes auszuschließen, – so sind einige Prototypen im Mittel wesentlich weiter von den Stimuli entfernt als andere – werden die Distanzen zusätzlich auf einer Skala von 0 bis 1 normiert.

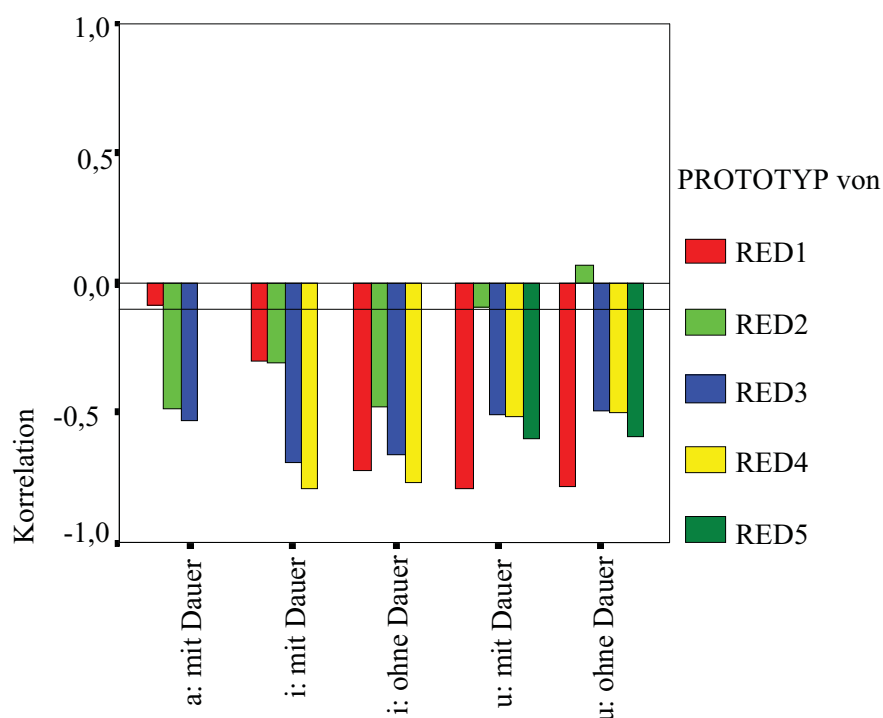
Vergleiche der Zuordnungen aufgrund der perzipierten Ähnlichkeit mit den Prototypen, die eine minimale akustische Distanz aufweisen, zeigen eine Übereinstimmung von rund 50% für die drei Vokalkategorien (/i:/ (50%), /a:/ (63,4%) und /u:/ (43,5%)). Häufig werden jedoch Stimuli, die eine geringe Distanz zu einem Prototypen besitzen, dem benachbarten, weiter entfernten Prototypen als ähnlicher empfunden.

Laut	n	ZF1	ZF2	ZF3	ZE1NORM	ZE2NORM	ZE3NORM	ZE4NORM	ZLAUT-DUR	ZF0	Übereinstimmung
/i:/	50	x	x	x	x	x	x	x	x	x	50
		x	x	x					x	x	50
		x	x	x					x		54
		x	x						x		70
		x	x								66
	106	x	x						x		73
/a:/	41	x	x	x	x	x	x	x	x	x	63
		x	x	x	x	x	x		x	x	63
		x	x	x	x	x	x		x		63
				x	x	x	x		x		66
				x	x	x			x		66
					x	x	x		x		63
									x		61
				x	x	x			x		83 <sup>*1</sup>
	93			x	x	x		x		75 <sup>*1</sup>	
/u:/	69	x	x	x	x	x	x	x	x	x	44
		x	x		x	x	x		x		48
			x		x	x	x		x		55
					x	x	x		x		54
				x	x	x	x				51
				x	x	x	x		x		58 <sup>*2</sup>
				x	x	x	x		x		61 <sup>*3</sup>
	98		x		x	x	x	x		66 <sup>*3</sup>	

**Tabelle 12: Übereinstimmungenraten (in %) von perceptiver und akustischer Nähe unter Berücksichtigung verschiedener akustischer Parameter für die Vokalkategorien /i:/, /a:/ und /u:/ (\*<sup>1</sup> Dauer wird zusätzlich berücksichtigt, \*<sup>2</sup> Distanzen zum Prototypen von RED1 werden unter Ausschluss von F2 berechnet, \*<sup>3</sup> Distanzen zum Prototypen von RED1 sind unter Ausschluss von F2 berechnet und der Abstand zum Prototypen von RED1 wird berücksichtigt; s. Text).**

Eine Ursache für die geringe Übereinstimmung ist möglicherweise, dass nicht alle akustischen Parameter gleichermaßen für die empfundene Ähnlichkeit relevant sind. Wie bereits die akustischen Analysen der Reduktionsstufen zeigen (Abschnitt 7.1.1), unterscheiden sich die einzelnen Reduktionsstufen durch verschiedene akustische Parameter. Aus diesem Grund werden für die erneute Berechnungen der Euklidischen Abstände nur die Parameter herangezogen, die hauptsächlich zu der Unterscheidung der Reduktionsstufen beitragen (vgl. in Abschnitt 7.1.1 Tabelle 11). Für /i:/ werden die Formantwerte, die Dauer und die Grundfrequenzwerte gewählt, für /u:/ die Werte von F1 und F2, die Energieverteilungen in den ersten drei Frequenzbändern und die Dauer und für /a:/ die Formant- und Grundfrequenzwerte, die Energieverteilungen in den ersten drei Frequenzbändern und die Dauer. Wie aus Tabelle 12 ersichtlich wird, werden die Übereinstimmungenraten zwischen mehrheitlich gewählten Prototypen und den aufgrund ihrer minimalen Distanz bestimmten Prototypen nicht wesentlich verbessert.

In weiteren Schritten werden bei /i:/ aufgrund der geringeren Relevanz für die Unterscheidung der Reduktionsstufen die Grundfrequenzwerte aus der Berechnung herausgenommen. Dann erfolgt auch noch der Ausschluss von F3. Die Ergebnisse zeigen, dass sich durch die Herausnahme der



**Abbildung 25: Korrelationen zwischen Häufigkeiten der interindividuellen Zuweisungen zu den Prototypen und akustischen Distanzen für die Vokalkategorien /i:/, /a:/ und /u:/ (/i:/, /u/ mit bzw. ohne Dauer s. Text; Korrelationswerte unterhalb der Linie sind signifikant).**

F0-Werte die Übereinstimmungen erhöhen (Tabelle 12). Die höchsten Übereinstimmungsraten finden sich, wenn nur die Werte der ersten beiden Formanten und die Dauer in die Berechnung des Euklidischen Abstands einfließen. Sie liegen bei 70%. Vergleiche der interindividuellen Zuweisungen mit den Distanzmaßen weisen auf signifikante Zusammenhänge hin, d.h. um so geringer der akustische Abstand eines Stimulus zum jeweiligen Prototypen ist, desto häufiger weisen die Hörer den Stimulus diesem Prototypen zu. Jedoch scheinen die Ähnlichkeitszuweisungen eher von den spektralen Parametern als von der Dauer abzuhängen. Hiervon sind die Zuweisungen zu den Prototypen von RED2 und insbesondere von RED1 betroffen. Die Häufigkeiten der Zuordnungen korrelieren wesentlich stärker mit den Distanzmaßen, die allein auf den Werten der zwei Formanten beruhen, als mit denjenigen, die zusätzlich die Dauer mit einbeziehen (Abbildung 25).

Bei /a:/ zeigt sich, dass bei weiteren Berechnungen der akustischen Distanzen unter Ausschluss der Grundfrequenz und verschiedener spektraler Parameter aufgrund ihrer geringeren Relevanz für die Unterscheidung der Reduktionsstufen (beispielsweise den ersten beiden Formantwerten, vgl. in Abschnitt 7.1.1 Tabelle 11), die Übereinstimmungsraten von perzeptiv und akustisch bestimmten Prototypen nicht wesentlich erhöht werden. Die alleinigen Vergleiche der Dauern von Prototypen und Stimuli erklären 61% der perzipierten Ähnlichkeiten. Die höchste Übereinstimmung findet sich bei Berücksichtigung der Dauer, den Energieverteilungen in den ersten zwei bzw. drei Frequenzbändern und der F3-Werte (66%; Tabelle 12).

Die Häufigkeiten der perzipierten Ähnlichkeiten weisen nur zu den akustischen Distanzen zwischen den Prototypen von RED2 und RED3 signifikante Zusammenhänge auf (Abbildung 25). Stimuli, die einen geringen Abstand zum Prototypen von RED1 besitzen, werden häufig dem Prototypen von RED3 zugewiesen. Die Diskrepanz zwischen akustischem Abstand und perceptiver Ähnlichkeit ist vor allem auf die Dauer der Stimuli selbst zurückzuführen. Diese Stimuli weisen alle eine geringere Lautdauer auf als Stimuli, deren akustisch und perceptiv gewählter Prototyp übereinstimmen. Die Dauer liegt unterhalb der mittleren Dauer über alle /a:/ der Sprecher. Die Ergebnisse weisen auf einen wesentlich stärkeren Einfluss der Dauer auf die perzipierte Ähnlichkeit zwischen Prototyp und Stimuli hin als bei der Berechnung der Distanzen berücksichtigt wird. Wird die Dauer eines Stimulus, der den geringsten Abstand zu den Prototypen von RED1 aufweist, nochmals zusätzlich einbezogen, so steigen die Übereinstimmungen zwischen den mehrheitlich gewählten und den akustisch bestimmten Prototypen auf 83%.

Für /u:/ sind die höchsten Übereinstimmungen (55%; Tabelle 12) zwischen perceptiver Ähnlichkeit und akustischer Distanz bei Berücksichtigung der Energieverteilungen in den ersten drei Frequenzbändern, den Werten von F2 und der Dauer festzustellen. Während die Verwechslungsraten (30%) zwischen den Prototypen von RED3 bis RED5 vergleichbar sind zu den interindividuellen Abweichungen der Probandenurteile (33%), finden sich häufigere Verwechslungen zwischen den Prototypen von RED1 und RED2, wobei Stimuli, die eine minimale Distanz zum Prototypen von RED2 aufweisen, aus perceptiver Sicht dem Prototypen von RED1 ähnlicher sind. Werden die Distanzen hingegen unter Ausschluss der F2-Werte berechnet, so sinken die Übereinstimmungsraten über alle fünf Prototypen leicht, jedoch erhöhen sich die Raten für den Prototypen von RED1 (mit F2: 46%; ohne F2: 54%). Eine mögliche Ursache für diese Erhöhung bei den nicht reduzierten Vokalen kann in der Mittelung der Formantfrequenzen bei der Perzeption von Hinterzungenvokalen liegen. Wie bereits in Abschnitt 3.1 dargestellt, werden bei einem geringen Abstand der Formanten zueinander, wie es bei den nicht reduzierten Hinterzungenvokalen der Fall ist, nicht die einzelnen Formantfrequenzen wahrgenommen, sondern eine mittlere Formantfrequenz, so dass die Abstandsmessungen ohne Berücksichtigung der einzelnen Formantfrequenzen eine adäquatere akustische Beschreibung der perzipierten Ähnlichkeit liefern. Die Bestimmung der minimalen Distanz über die Energieverteilungen und die Dauer für den Prototypen von RED1 sowie über die Energieverteilungen, die Dauer und F2 für die Prototypen der übrigen Reduktionsstufen führt zu einer Übereinstimmung von 58%.

Die Häufigkeiten der Zuordnungen zu den Prototypen der einzelnen Reduktionsstufen mit Ausnahme derjenigen von RED2 korrelieren mit den akustischen Distanzen signifikant (Abbildung 25). Hingegen finden sich signifikante Korrelationen ( $\rho = -0,323$ ;  $\alpha < 0,01$ ;  $n = 69$ ) zwischen der Häufigkeit der Zuweisungen zum Prototypen von RED2 und den berechneten Abständen zum Prototypen von RED1. Dies deutet darauf hin, dass die Probanden bei ihren Zuweisungen zum Prototypen von RED2 die Ähnlichkeit zum Prototypen von RED1 in Betracht



ziehen. Die Prototypen dieser beiden Reduktionsstufen werden mit einer gespannten bzw. ungespannten Vokalqualität assoziiert, so dass möglicherweise die Entscheidung, ob ein Stimulus dem Prototypen von RED1 oder von RED2 ähnlicher ist, durch Vergleiche mit dem Prototypen von RED1 gelenkt wird. Unterstützt wird diese Annahme durch die Korrelationen zwischen den akustischen Abständen und der Häufigkeit der Zuweisung der Stimuli, die den Prototypen von RED1 und RED2 zugeordnet sind. Während es keine signifikanten Zusammenhänge zwischen den akustischen Abständen der Stimuli zu dem Prototypen von RED2 und der Häufigkeitszuweisungen zu den Prototypen von RED1 und RED2 gibt, sind deutliche Korrelationen zwischen den Distanzen der Stimuli zu dem Prototypen von RED1 und den Zuweisungen zu den Prototypen von RED1 ( $\rho = -0,712$ ;  $\alpha < 0,01$ ;  $n = 42$ ) und RED2 ( $\rho = 0,640$ ;  $\alpha < 0,01$ ;  $n = 42$ ) festzustellen. Im folgenden werden daher Stimuli, die eine minimale Distanz zum Prototypen von RED2 besitzen, nur dann diesem Prototypen zugeordnet, wenn sie einen ausreichenden Abstand zum Prototypen von RED1 aufweisen, ansonsten werden sie dem Prototypen von RED1 zugewiesen. Durch die Berücksichtigung des Abstands zum Prototypen von RED1 steigt die Übereinstimmungsrate von perzeptiver und akustischer Nähe auf 60,9%.

Die vergleichbaren Korrelationen zwischen den Häufigkeiten der interindividuellen Zuweisungen und den auf den Energieverteilungen und F2 basierenden Distanzwerten sowie den Distanzwerten, in denen zusätzlich die Dauer berücksichtigt wird, weisen auf einen hauptsächlich Einfluss der spektralen Parameter bei den Ähnlichkeitsbeurteilungen hin (Abbildung 25).

Aufgrund der Berechnungen der Distanzen mit unterschiedlichen akustischen Parameterzusammensetzungen wird angenommen, dass die perzipierte Nähe der Prototypen und Stimuli von /i:/ durch die ersten beiden Formanten und die Dauer beschrieben werden können. Für die Ähnlichkeitszuweisungen von /a:/ ist die Dauer der primäre Einflussfaktor, die Energieverteilungen der ersten beiden Frequenzbänder und die F3-Werte sind weniger relevant. Die perzipierten Ähnlichkeiten zum nicht reduzierten Prototypen von /u:/ können durch die Energieverteilungen in den ersten drei Frequenzbändern und die Dauer beschrieben werden, die zu den übrigen Prototypen unter Zunahme der F2-Werte sowie unter Berücksichtigung eines ausreichenden Abstands zum nicht reduzierten Prototypen. Zur Überprüfung dieser Annahmen werden zum vorhandenen Datenmaterial noch die aufgrund ihrer qualitativen Ähnlichkeit zu den Prototypen beurteilten Stimuli aus den ersten beiden Perzeptionsexperimenten hinzugefügt. Die Übereinstimmungen zwischen den perzeptiv und akustisch bestimmten Prototypen liegen für /i:/ bei 73%, für /a:/ bei 75% und für /u:/ bei 66%. Sie liegen in den Bereichen der Übereinstimmungsraten von individuellen und mehrheitlich gewählten Prototypen, die in den Perzeptionsexperimenten ermittelt wurden (Abschnitt 6.1.5).

### 7.3.2 Zusammenfassung

Ausgangspunkt der Untersuchung war die Frage nach den akustischen Korrelaten der perzipierten Ähnlichkeitsbeziehung von Prototypen und Stimuli. Basierend auf der Theorie des Perzeptiven



Magneteffekts (Kuhl, 1992), wurde der Prototyp bestimmt, der die größte akustische Nähe zum jeweiligen Stimulus besitzt. Die auf den akustischen Eigenschaften basierenden Zuweisungen wurden mit den auf den mehrheitlich perzipierten Ähnlichkeiten beruhenden Zuordnungen innerhalb der Perzeptionsexperimente verglichen.

Die Ergebnisse zeigen, dass bei zwei Dritteln des Datenmaterials von /i:/ und /a:/ und bei mehr als der Hälfte von /u:/ die Ähnlichkeitsbeziehungen zwischen den Prototypen und Stimuli durch Bestimmung der relativen minimalen akustischen Distanz erklärbar sind, wenn die für die Perzeption der Reduktionsstufen primär relevanten akustischen Parameter ausgewählt werden. Die Korrelationen zwischen den Häufigkeiten der Zuweisungen und den akustischen Distanzen lassen jedoch keinen perzeptiven Magneteffekt der Prototypen erkennen. Nur die unreduzierten Prototypen der beiden hohen Vokale sowie der stark reduzierte Prototyp von /i:/ zeigen einen stärkeren Zusammenhang auf.

Die hier vorliegenden Ergebnisse lassen sich nur unzureichend mit der Theorie des Perzeptiven Magneteffekts erklären: Stimuli, die zu einem Prototypen die geringste Distanz aufweisen, werden nicht immer diesem zugeordnet. Die Diskriminierbarkeit der Stimuli kann daher nicht durch die akustische Distanz vorausgesagt werden. Kann hieraus aber geschlossen werden, dass die Prototypen der Reduktionsstufen keine guten Repräsentanten von Reduktionsstufen sind?

Die Theorie des Perzeptiven Magneteffekts setzt voraus, dass ein Prototyp sich in der Mitte des akustischen Raums einer Vokalkategorie befindet. Nur dann lässt sich über die minimale Distanz die Zuordnung zu den Prototypen erklären. Untersuchungen weisen jedoch darauf hin, dass auch Vokale als Prototypen gewählt werden, die sich am Rande des Stimulusraums befinden (vgl. Abschnitt 3.3). Nach der Theorie des Perzeptiven Magneteffekts müssten sich die Stimuli um ihren Prototypen verteilen. Was aber passiert, wenn sich der Prototyp im Randgebiet der Stimuliverteilung befindet? Für die an den Enden des Reduktionskontinuums befindlichen Prototypen (die nicht reduzierten und stark reduzierten) führt dies zu einer kleineren Abweichung von beobachteter und vorhergesagter Zuordnung aufgrund der akustischen Distanz als bei Prototypen, die in dem mittleren Bereich des Reduktionskontinuums liegen, da im letzteren Fall Stimuli aus anderen Reduktionsstufen falsch zugeordnet werden. Komplexer gestaltet sich das Problem, wenn die Prototypen einiger Reduktionsstufen am Rande und die anderer im Zentrum liegen.

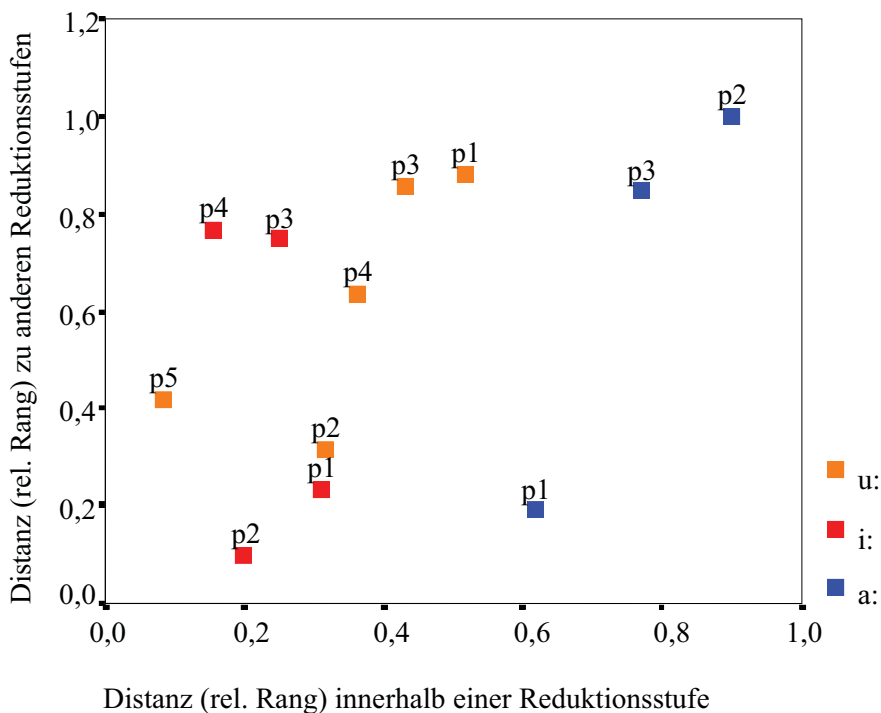
Zur Bestimmung der Lage der Prototypen wird folgende Methode (vgl. Palmeri, 1999) verwendet: Für jeden Stimulus und Prototypen werden die durchschnittliche akustische Distanz zu allen anderen Stimuli innerhalb einer Reduktionsstufe sowie die Distanzen zu den Stimuli anderer Reduktionsstufen bestimmt. Verwendet wurden hierbei die für die Unterscheidung relevanten Parameter der einzelnen Vokalkategorien (für /i:/: ZF1, ZF2, ZLAUTDUR; für /a:/: ZF3, ZE1NORM, ZE2NORM, ZLAUTDUR; für /u:/: ZF2, ZE1NORM, ZE2NORM, ZE3NORM,

ZLAUTDUR; vgl. Tabelle 12). Wenn sich ein Prototyp am Rande einer Stimuliverteilung befindet, muss der Abstand des Prototyps zu den Stimuli innerhalb der Reduktionsstufe und zwischen denjenigen der anderen Stufen möglichst groß sein. Wenn ein Prototyp sich in der Mitte einer Stimuliverteilung befindet, ist die Distanz innerhalb der Reduktionsstufe klein, und die Distanzen zu den Stimuli anderer Reduktionsstufen tendieren zu einer mittleren Größe. Innerhalb jeder Reduktionsstufe werden die Stimuli entsprechend ihren Abständen in eine Rangfolge gebracht. Da die Ränge von der Anzahl der Stimuli innerhalb einer Reduktionsstufe abhängen, wird für eine bessere Vergleichbarkeit der Lagen der Prototypen die zur Anzahl der Stimuli relative Ranggröße verwendet.

In Abbildung 26 sind die relativen ranggeordneten Distanzen innerhalb und zwischen den Reduktionsstufen der Prototypen aufeinander abgebildet. Aufgrund der Distanzen zwischen Prototyp und den ihm perceptiv zugeordneten bzw. nicht zugeordneten Stimuli kann somit auf die Lage eines Prototyps in der gesamten Stimuliverteilung geschlossen werden. Kleine Distanzwerte weisen auf eine zentrale Position und große auf eine periphere Position hin.

Grundsätzlich zeigt sich, dass die Prototypen der Vokalkategorie /a:/ größere Distanzen zu den Stimuli ihrer Reduktionsstufen aufweisen als die Prototypen von /i:/ und /u:/. Die Prototypen von RED2 und RED3 (p2, p3) von /a:/ zeigen außerdem größere Abstände zu den Stimuli anderer Reduktionsstufen auf, und nehmen, bezogen auf die gesamte Stimuliverteilung, eine periphere Lage ein. Der Prototyp von RED1 (p1) hingegen weist zu den Stimuli der anderen Reduktionsstufen nur einen geringen Abstand auf. Innerhalb der Reduktionsstufe nimmt dieser Prototyp eine relativ periphere Position ein. Er befindet sich somit nicht am Rande der gesamten Stimuliverteilung sondern im Grenzbereich zwischen den Stimuli seiner Reduktionsstufe und den der anderen Reduktionsstufen.

Die Prototypen von RED2, RED4 und RED5 (p2, p4, p5) der Vokalkategorie /u:/ liegen innerhalb ihrer Reduktionsstufe in einer zentraleren Position als die von RED3 und RED1 (p3, p1). Die Prototypen von RED3 und RED1 weisen zudem einen großen Abstand zu den Stimuli der anderen Reduktionsstufen auf. Sie befinden sich am Rande der gesamten Stimuliverteilung, während die Prototypen von RED2, RED4 und RED5 eine zentralere Position einnehmen. Mit den unterschiedlichen Positionen der Prototypen von RED1 und RED2 in der gesamten Stimuliverteilung lässt sich beispielsweise auch die häufige Verwechslung zwischen akustisch und perceptiv bestimmten Prototypen von RED1 und RED2 erklären (s. vorheriger Abschnitt), wobei Stimuli die eine minimale Distanz zum Prototypen von RED2 aufweisen, aus perceptiver Sicht dem Prototypen von RED1 ähnlicher sind. Aufgrund der peripheren Lage des Prototyps von RED1 und einer vergleichsweise zentraleren Lage von RED2 besitzen Stimuli eher eine minimale Distanz zum Prototypen von RED2 als zu dem von RED1.



**Abbildung 26: Relative ranggeordnete mittlere Distanzen der Prototypen (p1 bis p5) zu den Stimuli innerhalb einer Reduktionsstufe und den Stimuli anderer Reduktionsstufen für die Vokalkategorien /u:/, /i:/ und /a:/.**

Die Prototypen von /i:/ weisen innerhalb ihrer Reduktionsstufe eine zentrale Position auf. Im Verhältnis zu den Prototypen von RED1 und RED2 besitzen die stärker reduzierten Prototypen von RED3 und RED4 (p3, p4) zu den Stimuli anderer Reduktionsstufen größere Abstände. Sie befinden sich eher im Randgebiet der gesamten Stimuliverteilung.

Die Betrachtungen weisen darauf hin, dass einige Prototypen eine periphere Position innerhalb ihrer Reduktionsstufen und manchmal auch in der gesamten Stimuliverteilung aufweisen, so dass die minimale Distanz kein adäquates Maß zur Beschreibung der perzipierten Ähnlichkeit ist.

Zudem weisen Untersuchungen im Rahmen der Prototypentheorie darauf hin, dass das Auftreten des Magneteffekts begrenzt ist und von der Vokalkategorie abhängt. Es ist daher fraglich, ob prototypische Instanzen von Reduktionsstufen einen Magneteffekt aufweisen können. Es ist anzunehmen, dass Prototypen von Reduktionsstufen bei der Perzeption von Sprache keine Rolle spielen. In der mündlichen Kommunikation müssen Hörer Laute aufgrund ihres phonologischen Gehalts unterscheiden, so dass davon ausgegangen werden kann, dass sich im Laufe des Spracherwerbs nur für diese Kategorien Prototypen herausgebildet haben.

Weitere Einflussfaktoren auf die teilweise nur geringen Zusammenhänge von beobachtbaren Häufigkeiten der Zuweisungen und akustischer Distanz sind möglicherweise auch die Versuchsdurchführung und die verwendeten Stimuli. Kuhls (1992) Prototypentheorie basiert auf

einem Vergleich von jeweils zwei Stimuli (AX „gleich“ vs. „ungleich“ Prozedur), deren F1- und F2-Werte systematisch variiert wurden. Studien zeigen, dass Hörerurteile durch die verwendeten Vergleichsprozeden und die Komplexität des Stimuli beeinflusst werden (Abschnitt 3.3). In dieser Untersuchung wurden alle Referenzstimuli präsentiert, wobei sich Referenzstimuli und Stimuli in mehreren Parametern unterschieden. Werden natürliche Stimuli im Kontext präsentiert, dann können Hörer ihre Aufmerksamkeit nicht nur auf einen Aspekt legen, sondern sie beurteilen alle Parameter sowie deren Interaktion gleichzeitig.

Die Berechnung der Distanzen mit unterschiedlichen akustischen Parameterzusammensetzungen zeigen, dass die Ähnlichkeitsbeziehungen zum größten Teil von Energieanreicherungen (Formanten) bzw. Energieverteilungen im Bereich bis zu 4 kHz und Dauer abhängen. Verwechslungen von perzeptiv und akustisch bestimmten Prototypen sowie die Korrelationen zwischen den interindividuellen Ähnlichkeitszuweisungen und den akustischen Distanzwerten weisen jedoch darauf hin, dass nicht alle Parameter zu gleichen Anteilen an der empfundenen Ähnlichkeit zu den einzelnen Prototypen beteiligt sind. Die Ähnlichkeitszuweisungen der hohen Vokale wird primär durch die spektralen Parameter beeinflusst, die von /a:/ durch die Dauer. Dies entspricht der Annahme, dass insbesondere spektrale Parameter die Einteilung von Vokalrealisierungen in Reduktionsstufen, mit Ausnahme von /a:/, beeinflussen (vgl. Abschnitt 7.1.2).

Die Vorhersagen der Zuweisungen von Prototypen und Stimuli sind mit den interindividuellen Übereinstimmungen der Hörerurteile vergleichbar, wenn neben den akustischen Distanzen zum Prototypen noch weitere Faktoren – Dauer der Stimuli an sich, Distanz zu benachbarten Prototypen – hinzugezogen werden. Dies deutet darauf hin, dass einzelne Parameter unterschiedlich zu gewichten sind und die Ähnlichkeitsurteile möglicherweise durch den Kontext, d.h. durch die Präsenz weiterer Vergleichsstimuli beeinflusst werden.



## 8 REDUKTION, BETONUNG UND WORTART

Die Beziehung von Vokalreduktion und Prosodie sowie zwischen Vokalreduktion und Wortklasse bzw. -häufigkeit wurde in einer Reihe von Studien untersucht (vgl. Abschnitt 2.3). Verbreitet ist dabei eine Zusammenfassung verschiedener Wortklassen zu Inhalts- und Funktionswörtern. Im allgemeinen gilt, dass Wörter, die Träger wichtiger Informationen sind, genauer artikuliert und betont und somit vom Sprecher weniger reduziert werden als Wörter, die nach Meinung des Sprechers weniger zur Information beitragen. So werden Inhaltswörter aufgrund ihres semantischen Gehalts weniger reduziert und stärker betont als Funktionswörter.

Im folgenden wird überprüft, inwiefern sich die Interaktionen von Wortarten und Betonung in dem hier aufgestellten Beschreibungssystem von Vokalreduktion widerspiegeln. Zum einen werden die Verteilungen der reduzierten und nicht reduzierten Vokale in den Funktions- und Inhaltswörtern betrachtet. Zum anderen werden die Beziehungen zwischen den Reduktionsstufen und der lexikalischen Wortbetonung, Akzent und perzipierter Betonung analysiert. Das aus der BPD entnommene Datenmaterial verfügt über eine differenzierte Beschreibung der perzipierten Betonungsstärke (Abschnitt 5.2.1), so dass perzipierte Reduktions- und Betonungsstärke gegenübergestellt werden können.

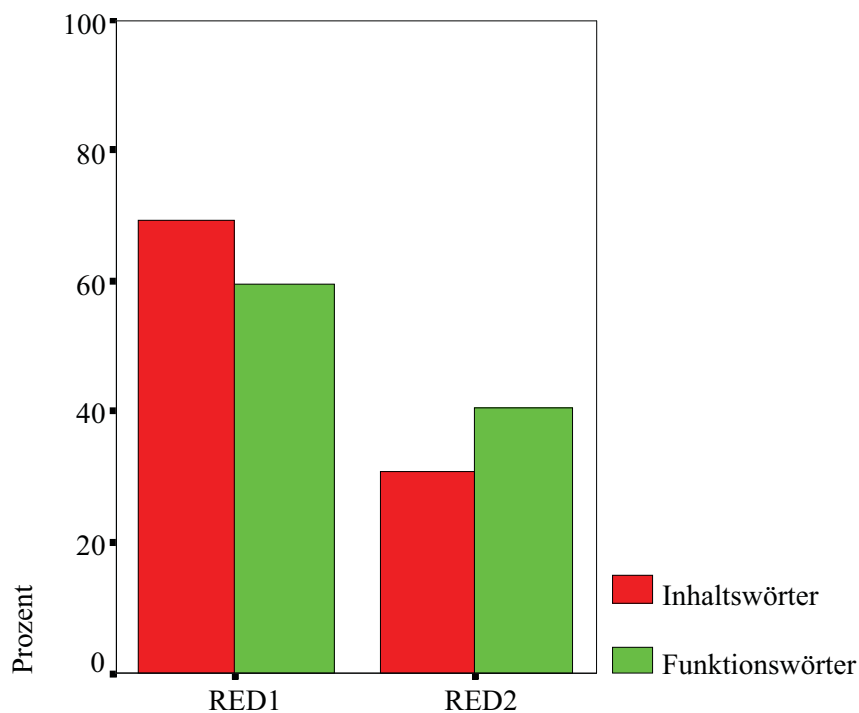
### 8.1 Untersuchungsmaterial

Als Datenmaterial dienen die hinsichtlich ihrer Reduktion eindeutig beurteilten Vokale (vgl. Abschnitt 7.1). Diese Vokale werden nach der maximalen Anzahl ihrer Reduktionsstufen in Gruppen zusammengefasst. Die Vokalrealisierungen von /ʊ/ und /a/ bilden die Gruppe der Vokale mit zwei Reduktionsstufen (2 RED), die von /i:/ die Gruppe mit vier Reduktionsstufen (4 RED) und die Realisierungen von /e:/ und /u:/ diejenige mit fünf Reduktionsstufen (5 RED). Die übrigen Vokale (/ɛ:/, /y:/, /ø:/, /a:/, /o:/, /ɪ/) gehören der Gruppe mit drei Reduktionsstufen (3 RED) an (vgl. Abschnitt 6.1.5 und 6.2.4).

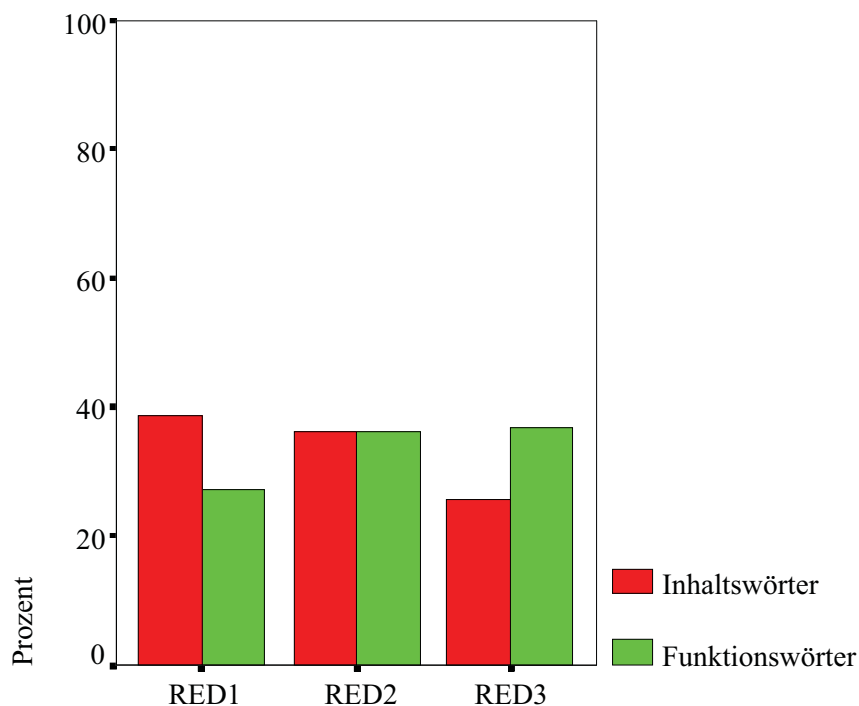
### 8.2 Untersuchung von Wortklasse, Betonung und Reduktion

#### Wortklasse und Reduktionsstufen

In den Abbildungen 27 bis 30 sind die Häufigkeiten der Reduktionsstufen in Abhängigkeit der Wortklasse für die einzelnen gruppierten Vokale dargestellt (s. Anhang E, Tabelle E-1). Wie zu erwarten, befinden sich in Inhaltswörtern häufiger nicht reduzierte Vokale (RED1), Vokale in Funktionswörtern hingegen sind häufiger stark reduziert. Vokalrealisierungen von RED2 (bei Vokalen mit mehr als 2 Reduktionsstufen) sind in beiden Wortklassen gleichermaßen vertreten.

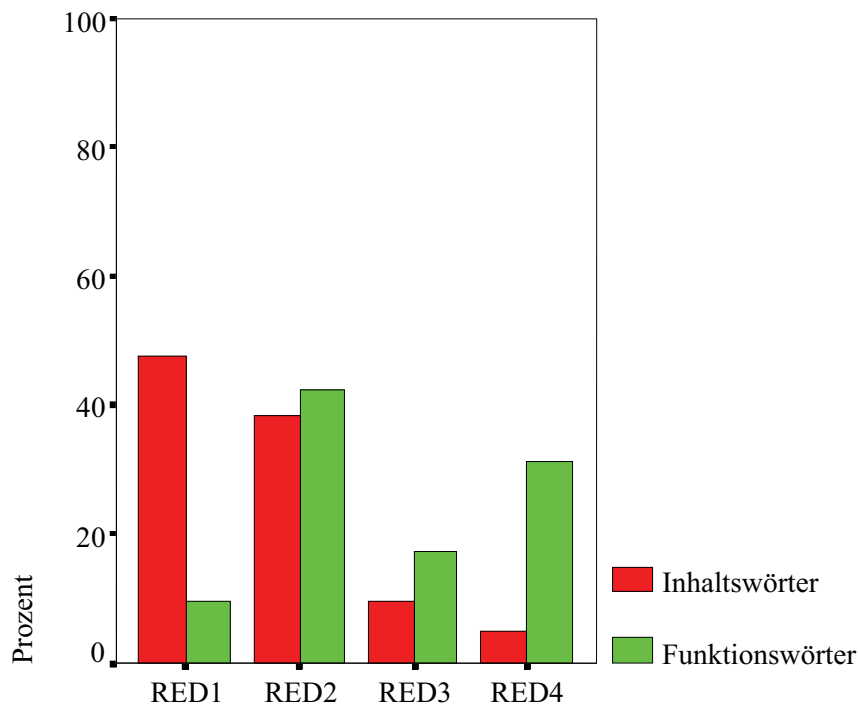


**Abbildung 27: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Inhaltswörter bzw. Funktionswörter für die gruppierten Vokale mit zwei Reduktionsstufen.**

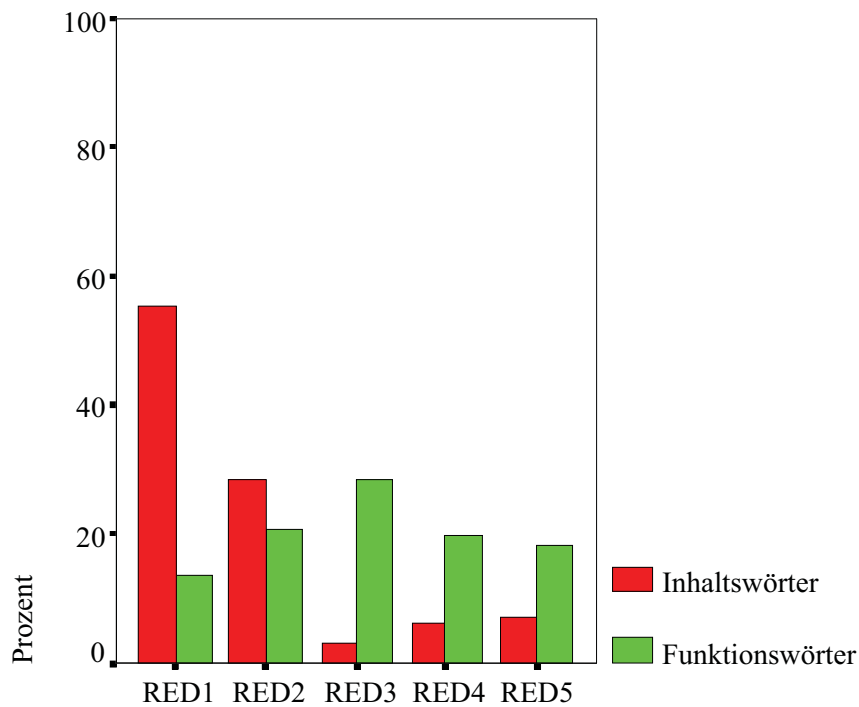


**Abbildung 28: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Inhaltswörter bzw. Funktionswörter für die gruppierten Vokale mit drei Reduktionsstufen.**

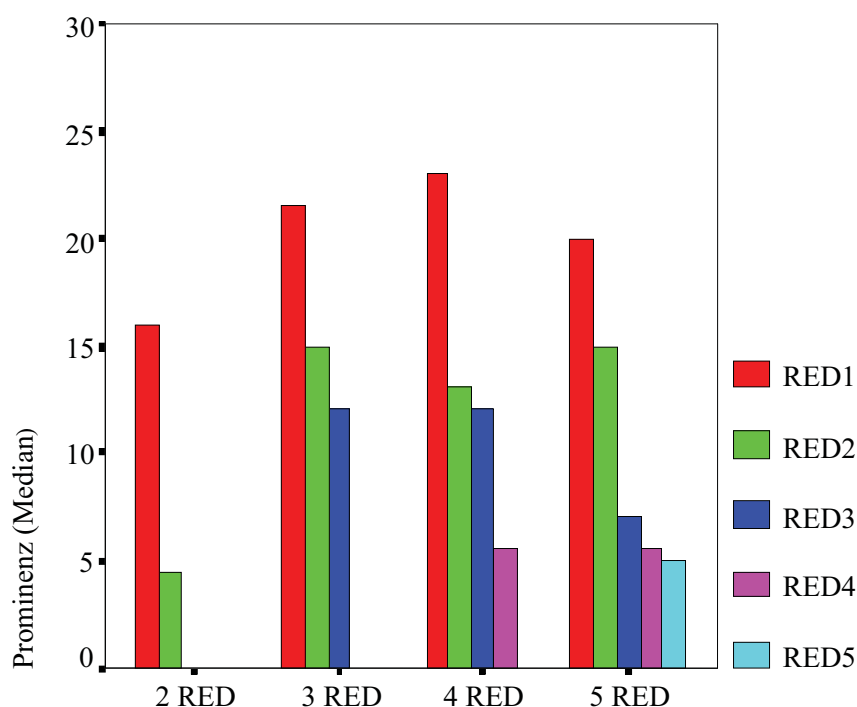




**Abbildung 29: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Inhaltswörter bzw. Funktionswörter für die gruppierten Vokale mit vier Reduktionsstufen.**



**Abbildung 30: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Inhaltswörter bzw. Funktionswörter für die gruppierten Vokale mit fünf Reduktionsstufen.**



**Abbildung 31: Perzeptive Prominenz (Median) von Silben und wahrgenomme Reduktionsstufen. Die Silbennuklei (Vokale) sind nach ihrer maximalen Anzahl von Reduktionsstufen gruppiert.**

Innerhalb der Inhaltswörter nimmt die Anzahl der Vokale mit zunehmenden Reduktionsstärken ab, so dass stark reduzierte Vokale eher selten auftreten. Bei den Funktionswörtern finden sich keine systematischen Beziehungen zwischen Reduktionsstärke und Häufigkeit. Die Verteilungen der einzelnen Reduktionsstufen weisen im Gegensatz zu den in Inhaltswörtern keine signifikanten Unterschiede mit Ausnahme der Vokale aus der Gruppe mit vier Reduktionsstufen auf (Tabelle E-2). Dies deutet darauf hin, dass die Vokale in Funktionswörtern genauso häufig reduziert wie nicht reduziert werden, während Vokale in Inhaltswörtern im allgemeinen nicht oder nur schwach reduziert werden.

#### Betonungsstärke und Reduktionsstärke

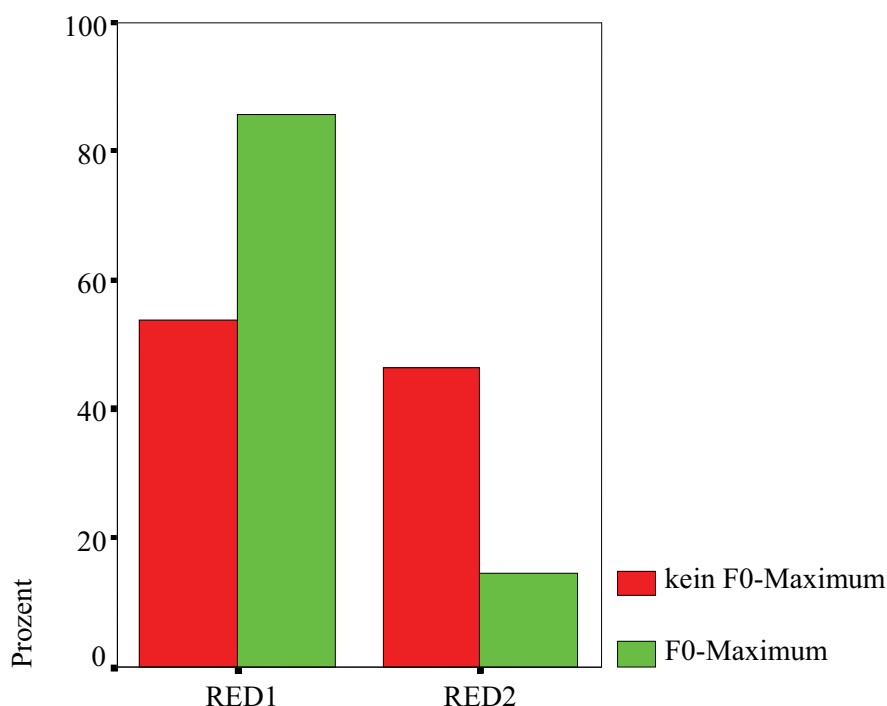
Vergleiche der perzipierten Betonungsstärke von Silben (Prominenz) mit den wahrgenommenen Reduktionsstufen von Vokalen zeigen, dass Silben umso betonter wahrgenommen werden, desto weniger ihr Nukleus reduziert ist (Abbildung 31). Korrelationsanalysen weisen einen signifikanten Zusammenhang ( $\alpha < 0,001$ ) zwischen perzipierter Reduktionsstärke und Prominenz auf (Gruppe 2 RED:  $\rho = -0,562$ ,  $n = 116$ ; Gruppe 3 RED:  $\rho = -0,396$ ,  $n = 376$ ; Gruppe 4 RED:  $\rho = -0,580$ ,  $n = 106$ ; Gruppe 5 RED:  $\rho = -0,591$ ,  $n = 220$ ).

Nach Heuft (1999) ist ein wichtiges Merkmal der Prominenz von Silben die Präsenz eines F0-Maximums. Solche Silben besitzen eine höhere Prominenz als Silben ohne F0-Maximum (im Durchschnitt um 15 Prominenzstufen). Die Verteilungen der Reduktionsstufen

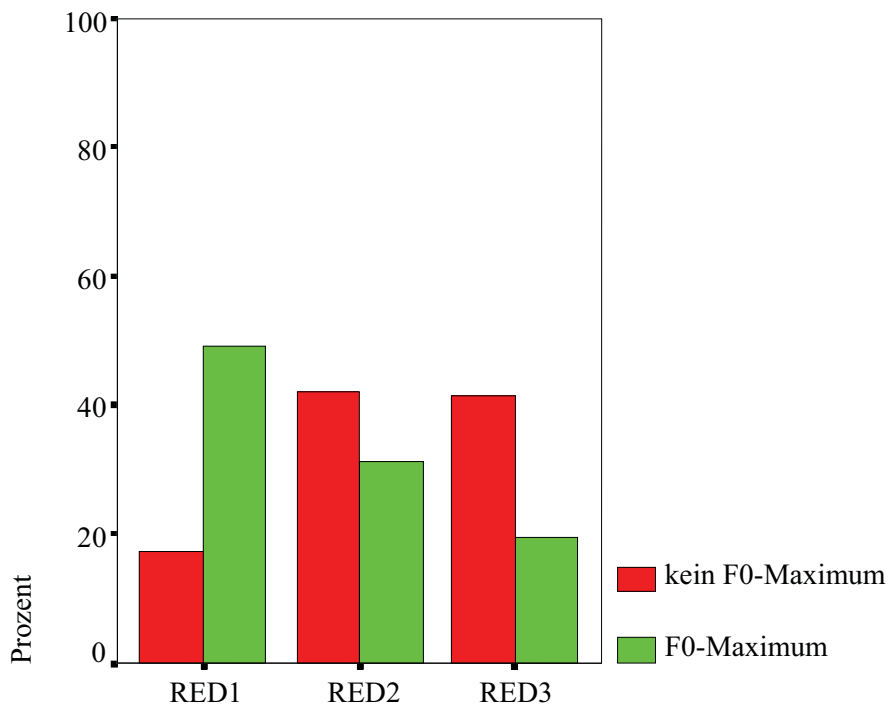
zeigen, dass Vokale in Silben mit einem F0-Maximum deutlich häufiger nicht reduziert sind als Vokale in Silben ohne F0-Maximum (Abbildung 32 bis Abbildung 35 und Tabelle E-3). Schwach reduzierte Vokalrealisierungen (RED2 bei den Gruppen mit mindestens 3 Reduktionsstufen) kommen häufiger – mindestens aber genauso häufig – in Silben ohne F0-Maximum vor als in Silben mit F0-Maximum. Vokale in akzentuierten Silben werden in der Regel nicht reduziert.

Silben, die eine lexikalische Wortbetonung tragen, sind potentielle Kandidaten für eine Akzentuierung. Während bei ungefähr der Hälfte der Silben (54,4% von 671) mit lexikalischer Wortbetonung die Betonung durch ein F0-Maximum realisiert ist, ist nur ein sehr geringer Anteil von Silben ohne lexikalische Wortbetonung (10,9% von 147) durch ein F0-Maximum gekennzeichnet. Bei diesen Silben zeigt sich deutlich die Interaktion von Vokalreduktion und Akzentuierung. Nur ein geringer Teil der Vokale (25% von 16) dieser akzentuierten Silben ist reduziert.

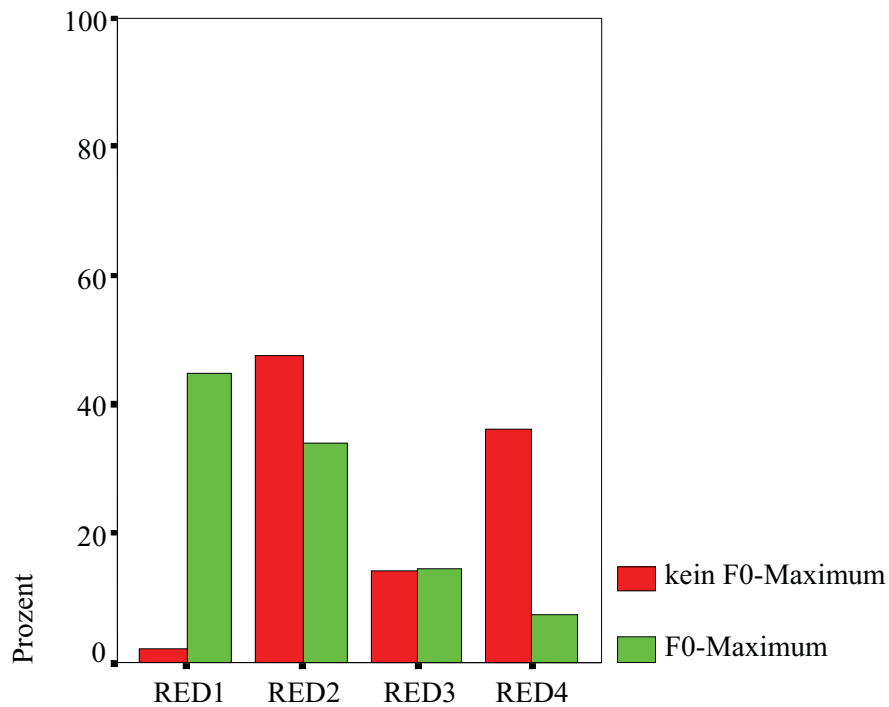
Wie bereits oben erwähnt, weisen die Vokale in Funktionswörtern keine ausgeprägte Tendenz zur Reduktion auf. Im Vergleich zu den Inhaltswörtern werden sie zwar stärker reduziert, jedoch besitzen Funktionswörter vergleichbar viele reduzierte wie unreduzierte Vokale. Aufgrund des Zusammenhangs zwischen perceptiver Prominenz und wahrgenommener Reduktionsstärke ist anzunehmen, dass Funktionswörter, die keine reduzierten Vokale aufweisen, vom Sprecher betont geäußert werden. Vergleiche von Inhalts- und Funktionswörtern zeigen, dass Silben von Funktionswörtern, deren Nukleus nicht reduziert ist, mit vergleichbarer Prominenz wahr



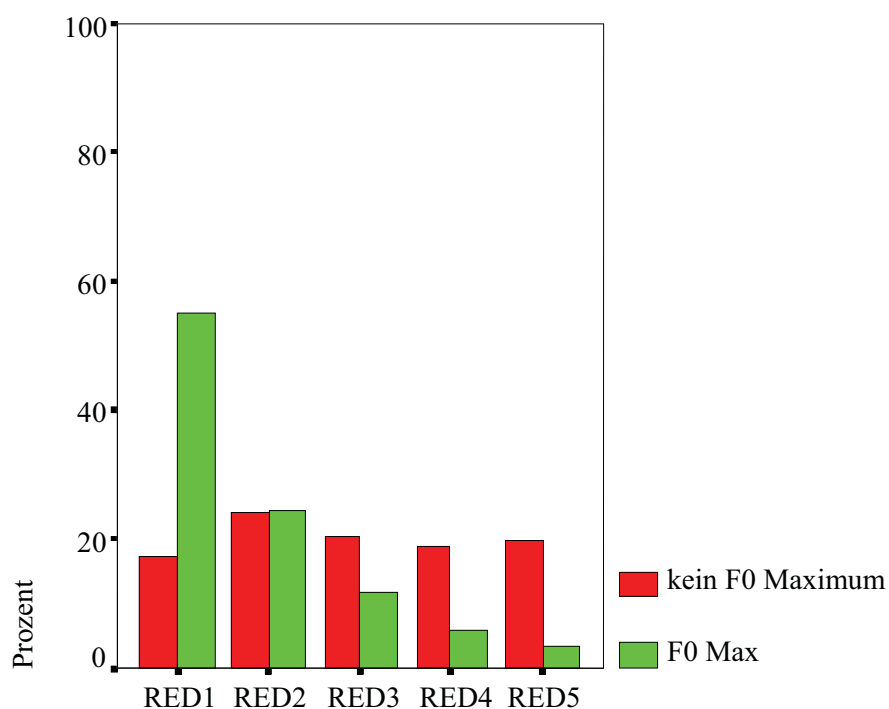
**Abbildung 32: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Silben mit bzw. ohne F0-Maximum für die gruppierten Silbennuklei mit zwei Reduktionsstufen.**



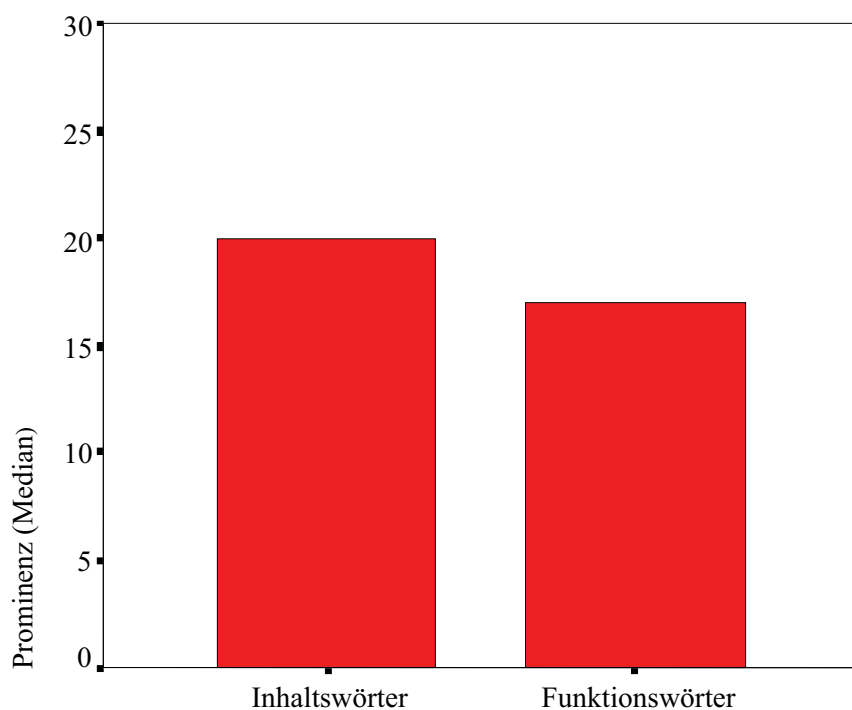
**Abbildung 33: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Silben mit bzw. ohne F0-Maximum für die gruppierten Silbennuklei mit drei Reduktionsstufen.**



**Abbildung 34: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Silben mit bzw. ohne F0-Maximum für die gruppierten Silbennuklei mit vier Reduktionsstufen.**



**Abbildung 35: Häufigkeiten (in %) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Silben mit bzw. ohne F0-Maximum für die gruppierten Silbennuklei mit fünf Reduktionsstufen.**



**Abbildung 36: Perzeptive Prominenz (Median) von Silben mit nicht reduzierten Vokalen in Abhängigkeit der Wortklasse.**

genommen werden wie Silben von Inhaltswörtern mit nicht reduzierten Vokalen (Abbildung 36, Tabelle E-4 und Tabelle E-5).

### **8.3 Zusammenfassung**

Die Untersuchungen zeigen, dass in Inhaltswörtern die Vokale weniger reduziert sind als in Funktionswörtern. Die Vokale in Funktionswörtern sind jedoch nicht nur einfach reduziert, sondern ihre Realisierungen weisen das gesamte Spektrum der Reduktionsstufen auf. Betrachtungen der Interaktion von Wortklasse, Betonung und Reduktion zeigen, dass Vokale in Inhaltswörtern zwar weniger reduziert sind als in Funktionswörtern und Vokale in betonten Silben weniger Reduktionen aufweisen als die in unbetonten Silben, der Faktor Betonung jedoch einen größeren Einfluss auf die Lautrealisierung zu besitzen scheint als die Einteilung in Funktions- und Inhaltswörter. Unabhängig von Wortklasse (vgl. auch Bell et al., 1999) oder lexikalischer Wortbetonung sind Vokale in akzentuierten Silben nicht reduziert. Dies unterstützt die Aussage, dass eine wahrgenommene Akzentuierung durch das Fehlen von Reduktion verursacht wird (Rietveld und Koopmans-van Beinum, 1987).

Durch die quantitative Erfassung der Reduktion in Form von Reduktionsstufen ist es möglich, graduelle Parameter der Perzeption, in diesem Fall der Prominenz von Silben, in einen direkten Bezug zur Vokalreduktion zu setzen. Reduktionsstärke und perzeptive Prominenz von Silben weisen eine mittlere Korrelation (-0,53) auf. Die Korrelationswerte weisen darauf hin, dass die wahrgenommene Prominenz von Silben durch die Reduktion des Silbennukleus beeinflusst wird. Zu berücksichtigen ist aber, dass zum einem Vokalreduktionen nicht allein durch eine Abnahme der Betonung hervorgerufen werden, sondern weitere Faktoren, z.B. Sprechstil, Sprechgeschwindigkeit, auf die Vokalrealisierungen wirken und zum anderen auch die Perzeption der Prominenz unterschiedlichen linguistischen Einflüssen unterliegt (vgl. Heuft, 1999; Mixdorff und Widera, 2001). Im Vergleich hierzu zeigen Analysen von Prominenz und Vokalreduktionen, die auf einzelnen akustischen Parametern beruhen, einen durchschnittlich schwächer ausgeprägten Zusammenhang (Portele, 1999). Durch die hier vorgestellte Beschreibung von Vokalreduktion wird das Verhältnis von Betonung und Reduktion adäquat wiedergegeben.

## 9 AUTOMATISCHE KLASSIFIKATION VON REDUKTION

In Kapitel 6 wurde gezeigt, dass Hörer verschiedene Reduktionsstufen bei Vokalrealisierungen unterscheiden können. Analysen weisen darauf hin, dass sich die einzelnen Reduktionsstufen hinsichtlich ihrer akustischen Werte unterscheiden (Kapitel 7). Im folgenden wird geprüft, ob eine automatische Klassifizierung akustischer Lautrealisierungen in Reduktionsstufen möglich ist. Die Güte einer Klassifikation hängt von der Trennbarkeit der einzelnen Reduktionsstufen ab, d.h. die akustischen Werte verschiedener Reduktionsstufen sollten sich nicht überschneiden, und innerhalb einzelner Reduktionsstufen sollten die Werte möglichst ähnlich sein. Die Trennbarkeit der Reduktionsstufen wird mit drei verschiedenen Klassifikationsverfahren, der linearen Diskriminanzanalyse, *Support Vector Machines* und künstlichen neuronalen Netzen getestet. Zudem wird untersucht, ob sich sprecherspezifische Unterschiede bei der Realisierung von Reduktionsstufen zeigen.

### 9.1 Datenmaterial

Für die Untersuchung der Trennbarkeit der Reduktionsstufen werden hier exemplarisch für die Vorderzungen- und Hinterzungenvokale die Vokalrealisierungen von /i:/ und /u:/ genommen, sowie die des tiefen Mittelzungenvokals /a:/, die hinsichtlich ihrer Reduktion (ORL) bewertet sind. Um die Stichprobengröße zu erhöhen, wurden von einer Phonetikerin noch weitere Lautrealisierungen dieser drei Vokalkategorien durch Zuweisungen zu den Prototypen beurteilt. Insgesamt liegen für /i:/ und /u:/  $n = 180$  (pro Sprecher jeweils  $n = 60$ ) Lautrealisierungen vor und für /a:/  $n = 150$  (pro Sprecher jeweils  $n = 50$ ). Die Klassifikationen der Realisierungen der drei Vokalkategorien beruhen auf den sprecher- und lautnormierten Werten der ersten drei Formanten (ZF1, ZF2, ZF3), der Energie in den vier verschiedenen Frequenzbändern (ZE1NORM, ZE2NORM, ZE3NORM, ZE4NORM) der Grundfrequenz (ZF0) und der Lautdauer (ZLAUTDUR; vgl. Abschnitt 5.2.1 und Anhang F, Tabelle F-1).

### 9.2 Verwendete Klassifikationsverfahren

#### Lineare Diskriminanzanalyse

In den bisherigen akustischen Analysen der Reduktionsstufen wurden die einzelnen akustischen Parameter unabhängig voneinander betrachtet. Um zu überprüfen, in welchem Ausmaß die einzelnen Parameter bei der Bildung der Reduktionsstufen beteiligt sind, werden lineare Diskriminanzanalysen (LDA; Bortz, 1979) durchgeführt. Ziel der LDA ist es, eine oder mehrere Funktionen zu finden, die eine maximale Trennung der Gruppen erreicht. Die Anzahl der Diskriminanzfunktionen beträgt im Fall, dass mehr Parameter (Variablen) als  $k$  Gruppen untersucht werden,  $k-1$ . Die allgemeine Form der linearen Diskriminanzfunktion lautet:



$$d = a + \sum_i^n b_i x_i \quad (6);$$

wobei die Konstante  $a$  sowie  $b_i$  die Gewichtungskoeffizienten bilden und  $x_i$  die Werte der einbezogenen unabhängigen Variablen (Bühl und Zöfel, 1998).

Die Koeffizienten der Diskriminanzfunktion werden so geschätzt, dass der Quotient aus der Quadratsumme der Funktionswerte zwischen den Gruppen und der Quadratsumme der Funktionswerte innerhalb der Gruppen maximal wird. Die Diskriminanzanalyse gibt Auskünfte über den Beitrag, den eine Diskriminanzfunktion zur Aufklärung der Gesamtvarianz leistet, und über den relativen Anteil, den die einzelnen Variablen zu der Diskrimination beitragen. Ein Maß dafür, wie wichtig eine Variabel für die Diskrimination der Gruppen ist, sind die Korrelationen innerhalb der Gruppen zwischen den jeweiligen Diskriminanzvariablen und den standardisierten kanonischen Diskriminanzfunktionen.

Ist der Funktionswert  $d$  bekannt, können durch Anwendung des Satz von Bayes die Wahrscheinlichkeiten dafür geschätzt werden, dass ein Fall  $i$  aus einer bestimmten Gruppe einen bestimmten Wert annimmt. Ein Fall mit Wert  $d$  wird derjenigen Gruppe zugeordnet, für die die bedingte (a posteriori-)Wahrscheinlichkeit maximal ist (Brosius und Brosius, 1995).

$$p(g | i) = \frac{p(g) * p(i | g)}{p(i)} \quad (7)$$

Hierbei sind:  $p(g | i)$  die Wahrscheinlichkeit für das Auftreten der Gruppe  $g$  unter der Bedingung, dass ein Fall einen bestimmten Wert einnimmt,  $p(i | g)$  die Auftretenswahrscheinlichkeit für die Fälle mit einem bestimmten Wert  $d$  in Abhängigkeit der Gruppen und  $p(g)$  bzw.  $p(i)$  die Auftretenswahrscheinlichkeit für die Gruppen bzw. für die Fälle.

Für jede Vokalkategorie erfolgte die Abschätzung der Gewichtungskoeffizienten und Klassifizierung aufgrund aller beurteilten Lautrealisierungen. Bei der Berechnung der bedingten Wahrscheinlichkeit wurde davon ausgegangen, dass die Reduktionsstufen die gleiche Auftretenswahrscheinlichkeit (a priori-Wahrscheinlichkeit) besitzen.

### Support Vector Machines

Im Vergleich zu der Diskriminanzanalyse können durch *Support Vector Machines* (SVM; Schölkopf et al. 1996; Burges, 1998; Smola, 1999; Duda et al., 2001) auch Klassen von Daten getrennt werden, deren Grenzen nicht durch lineare Funktionen beschreibbar sind. Ziel der SVM ist es, optimale Hyperebenen zu finden, die möglichst korrekt die Daten klassifizieren (empirische Risikominimierung). Durch die nichtlinearen Transformationen der Daten in höher dimensionierte

Räume (Merkmalsräume) können auch nicht linear trennbare Daten durch lineare Funktionen voneinander getrennt werden. *Support vectors* sind (transformierte) Daten aus dem Trainingsmaterial, die am schwierigsten zu klassifizieren sind. Sie liegen an den Klassengrenzen. Durch die *Support vectors* werden optimale Hyperebenen definiert. Die optimale Hyperebene ist ein linearer Separator, der eine maximale Distanz zwischen den *Support vectors* aufweist. Sie werden durch das Prinzip der strukturellen Risikominimierung (*structural risk minimization*; Vapnik, 1979) bestimmt. SVM sind robust; Übergeneralisierungen sollen im Ansatz vermieden werden.

Zur Klassifizierung mit Hilfe von SVM wurde die Implementierung SVMlight verwendet (s. Joachims, 1999), wobei die Vokale in jeweils zwei Gruppen eingeteilt wurden. Eine Gruppe beinhaltet die Vokale einer bestimmten Reduktionsstufe, die andere Gruppe umfasst die Vokale der übrigen Reduktionsstufen. Für jede Reduktionsstufe wurde eine SVM trainiert. Wie bei der Klassifikation mit Hilfe der LDA stimmen Trainings- und Testmaterial überein. Hierdurch wird geprüft, inwieweit die Stichproben widerspruchsfrei sind.

### Neuronale Netze

Wie die SVM eignen sich auch künstliche neuronale Netze zur Klassifizierung nicht linear trennbarer Daten. Zur Überprüfung der Unterscheidbarkeit und Generalisierung der Reduktionsstufen wurden neuronale Netze (NN; Zell, 1994) ohne Rückkopplung, sogenannte *feedforward*-Netze, genommen. Die NN bestehen aus einer Eingabeschicht (*input layer*), zwei verdeckten Schichten (*hidden layers*) und einer Ausgabeschicht (*output layer*), wobei die einzelnen Schichten vollständig miteinander verbunden sind. Mit Hilfe des *Backpropagation*-Algorithmus werden die Gewichte des Netzes so geändert, dass nach mehrmaliger Präsentation der Eingabe- und Ausgabemuster die Muster den richtigen Klassen zugeordnet werden. *Backpropagation* gehört zu den typischen überwachten Lernverfahren. Überwachte Verfahren sind durch fünf Schritte gekennzeichnet (Zell, 1994:94):

1. Durch eine entsprechende Aktivierung der Eingabeneuronen wird das Eingabemuster präsentiert,
2. das Netz erzeugt durch Vorwärtspropagierung ein Ausgabemuster für eine aktuelle Eingabe,
3. ein Vergleich des Ausgabemusters mit der erwünschten Ausgabe liefert einen Fehlervektor,
4. durch eine Rückwärtspropagierung der Fehler von der Ausgabe- zur Eingabeschicht werden die Verbindungsgewichte so geändert, dass sich der Fehler verringert,
5. die Gewichte aller Neuronen des Netzes werden mittels des Gradientenabstiegsverfahrens so verändert, dass die Fehlersumme über alle Trainingsmuster minimal wird.

Zur Klassifizierung der Vokale wurde der Stuttgarter Neuronale Netze Simulator (SNNS; s. Zell, 1994:361 ff) verwendet. Eingabe der NN waren die akustischen Parameterwerte, die im Intervall von [0 bis 1] normiert wurden. Die Ausgabeschicht besteht aus jeweils einem Neuron pro Reduktionsstufe. Ein Eingabemuster wird dem Neuron der Ausgabeschicht (der Reduktionsstufe) zugeordnet, das maximal erregt ist. Die erste verdeckte Schicht der NN für /i:/ und /a:/ besteht aus zehn Neuronen, die zweite Schicht aus fünf Neuronen. Im Vergleich zu den NN von /i:/ und /a:/ wurde beim NN von /u:/ die Neuronenanzahl der zweiten Schicht erhöht, da das NN mit nur fünf Neuronen während der Trainingsphase zu einem lokalen Minimum konvergierte. Die verdeckten Schichten beinhalten jeweils zehn Neuronen. Da NN bei ausreichend wiederholter Präsentation der Trainingsdaten die Trennung jedes Musterraums erlernen können, was zu einer Spezialisierung auf die Trainingsdaten (*overfitting*) führt, werden zur Bewertung der Performanzen der NN die Klassifizierungsergebnisse herangezogen, die sich aufgrund einer fünffachen Kreuzvalidierung ergeben.

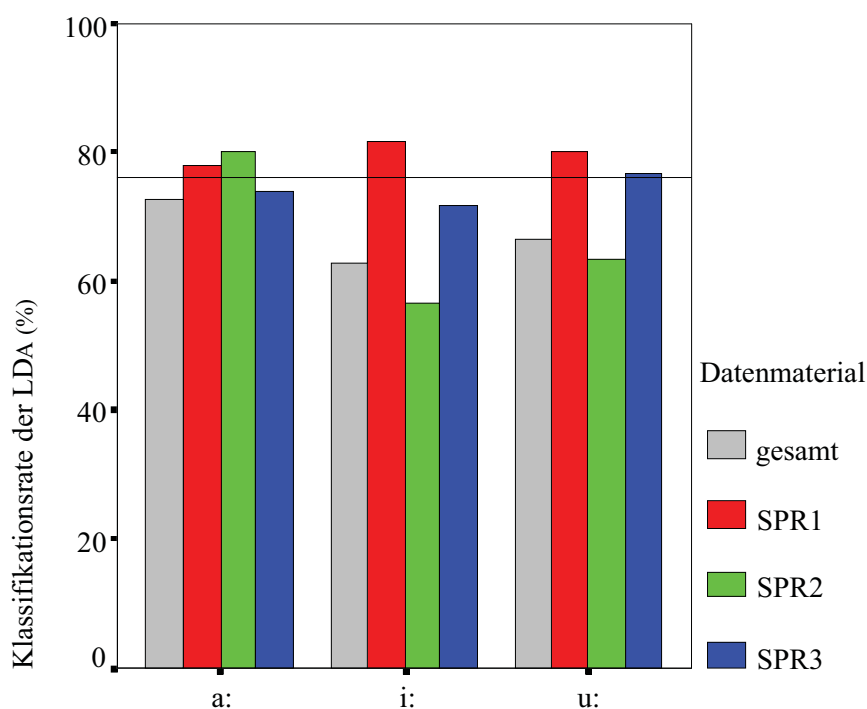
### 9.3 Ergebnisse der Klassifikation

Die Ergebnisse zeigen, dass mit Hilfe der LDA 73% der Lautrealisierungen von /a:/ richtig klassifiziert werden, 67% der Realisierungen von /u:/ und 63% derjenigen von /i:/ (Abbildung 37; s. Tabelle F-2 und Tabelle F-3).

Im Gegensatz zu den Klassifikationen der LDA basiert die Klassifikation der SVM auf zwei Klassen. Zur Beurteilung der Performanz der SVM wird nicht die Akkuratheit der Klassifikation herangezogen, sondern die sogenannte *recall*-Rate. Da die Anzahl der Vokale einer bestimmten Reduktionsstufe immer geringer ist als die Anzahl der Vokale in den übrigen Reduktionsstufen, sagt die Akkuratheit nichts über die Performanzen der SVM aus. Die *recall*-Rate ist durch das Verhältnis der Anzahl der Vokale einer bestimmten Reduktionsstufe, die richtig klassifiziert werden ( $n_{RED\text{ richtig}}$ ), zur Anzahl der Vokale mit dieser Reduktionsstufe ( $n_{RED}$ ) definiert:

$$recall = \frac{n_{RED\text{ richtig}}}{n_{RED}} \quad (8).$$

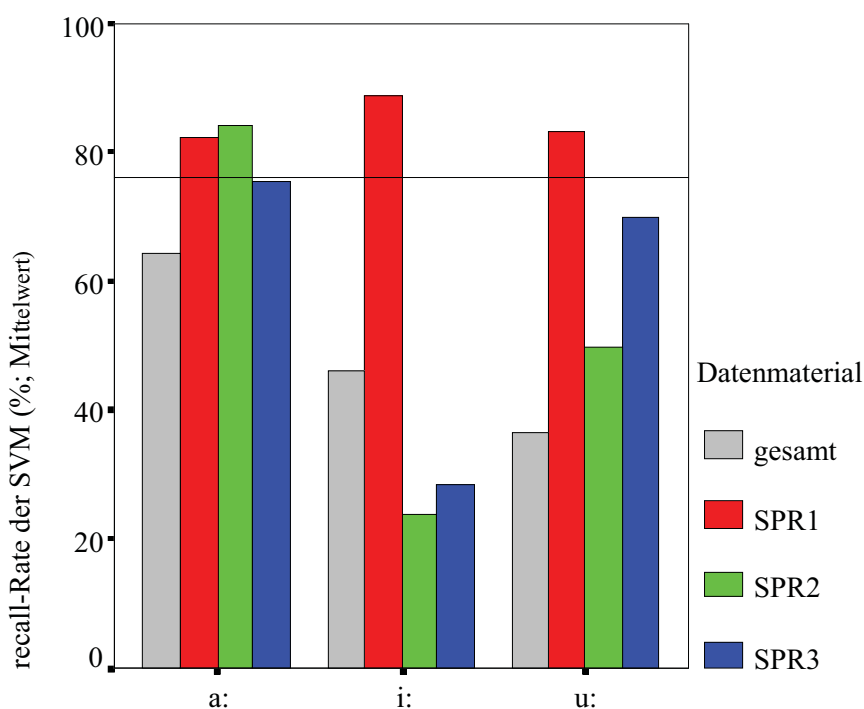
Die durchschnittlichen *recall*-Raten liegen bei allen drei Vokalen noch unterhalb der Ergebnisse der LDA (Abbildung 38). Insbesondere für die Vokale /i:/ (46%) und /u:/ (36%) fällt diese recht niedrig aus, was bei dem Vorderzungenvokal insbesondere auf keine bzw. nur schlechte Trennbarkeit von RED3 (0%) und RED4 (46,9%) zurückgeführt werden kann und bei dem Hinterzungenvokal vor allem auf die fehlende Trennbarkeit von RED2 und RED3. Bei dem Tiefzungenvokal liegt die durchschnittliche *recall*-Rate bei 65%. Hier wird RED2 nur zu 44% richtig klassifiziert.



**Abbildung 37: Klassifikationsrate (%) der linearen Diskriminanzanalyse für /a:/, /i:/ und /u:/ in Abhängigkeit vom Datenmaterial. Die Linie markiert die in den Perzeptionsexperimenten ermittelte durchschnittliche Übereinstimmung zwischen den Probanden.**

Die Ergebnisse der NN, die auf einer fünffachen Kreuzvalidierung beruhen, zeigen, dass nur um die 50% der Lautrealisierungen der einzelnen Vokalkategorien richtig klassifiziert werden (Abbildung 39). Ein Vorteil der Klassifikationsverfahren, die auch nicht linear trennbare Daten unterscheiden können (SVM und NN), gegenüber der LDA ist nicht festzustellen. Werden die Klassifizierungsergebnisse der einzelnen Vokalkategorien getrennt nach den Sprechern ausgewertet, so zeigt sich, dass die Güte der Klassifizierung vom Sprecher abhängt. Beispielsweise liegt die Klassifikationsrate der LDA für die Lautrealisierungen von SPR1 für alle Vokale über 70% (/a:/: 80%; /i:/: 77%; /u:/: 81%) und ist vergleichbar mit der Übereinstimmung, die die Probanden in den Perzeptionsexperimenten (/a:/: 72%; /i:/: 77%; /u:/: 81%) erzielten. Bei dem Datenmaterial von SPR2 gilt dies nur für die Lautrealisierungen von /a:/ (72%). Auch für SPR3 findet sich die höchste Klassifikationsrate bei der Vokalkategorie /a:/, diese liegt jedoch noch unterhalb der Übereinstimmungsrate der Probanden bei 66%. Die im Durchschnitt bessere Klassifizierung des Datenmaterials von SPR1 im Vergleich zu den beiden anderen Sprechern deutet darauf hin, dass die akustischen Realisierungen der Reduktionsstufen inter- oder auch intraindividuelle Unterschiede aufweisen.

Zur Überprüfung einer sprecherabhängigen akustischen Realisierung von Reduktionsstufen werden Klassifikationen mit allen drei Verfahren getrennt für das Datenmaterial eines jeden Sprechers durchgeführt. Wie zu erwarten, zeigt sich gegenüber der Klassifizierung des gesamten Datenmaterials eine Verbesserung bei der Klassifizierung des sprecherspezifischen

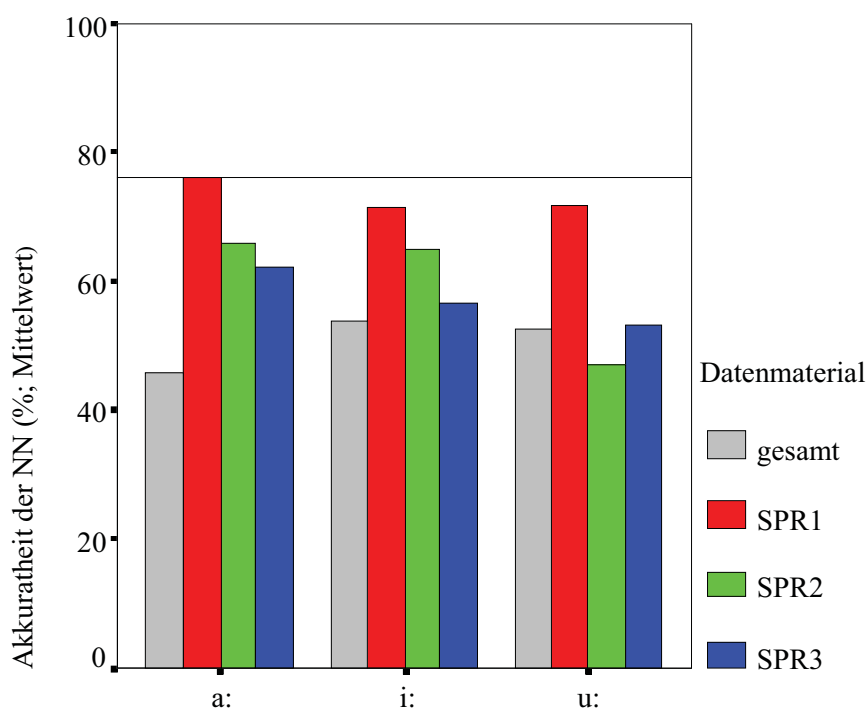


**Abbildung 38: Durchschnittliche *recall*-Rate (%) der Support Vector Machines für /a:/, /i:/ und /u:/ in Abhängigkeit vom Datenmaterial. Die Linie markiert die in den Perzeptionsexperimenten ermittelte durchschnittliche Übereinstimmung zwischen den Probanden.**

Datenmaterials. Die Ergebnisse der LDA weisen bei /a:/ eine Erhöhung der Klassifikationsrate um durchschnittlich 3 Prozentpunkte auf (SPR1: 78%, SPR2: 80%, SPR3: 74%). Bei den Vokalkategorien /i:/ und /u:/ ist eine deutliche Erhöhung der Klassifikationsrate bei dem Datenmaterial von SPR1 (/i:/: 82%, /u:/: 80%) und SPR3 (/i:/: 72%, /u:/: 78%) festzustellen (Abbildung 37).

Die durchschnittlichen *recall*-Raten der SVM sind mit den Ergebnissen der LDA für die sprecherspezifischen Daten der Vokalkategorien /u:/ und /a:/ vergleichbar. Bei /i:/ werden nur die Reduktionsstufen der Lautrealisierungen von SPR1 durch die SVM zuverlässig voneinander getrennt; die mittlere *recall*-Rate liegt bei 89% (Abbildung 38). Die Ergebnisse der NN weisen nur für die Netze, die mit dem Datenmaterial von SPR1 trainiert und getestet wurden, eine zu den in den Perzeptionsexperimenten ermittelten Übereinstimmungen verhältnismäßig vergleichbare Akkuratheit auf (/a:/: 76%, /i:/: und /u:/: 72%; Abbildung 39).

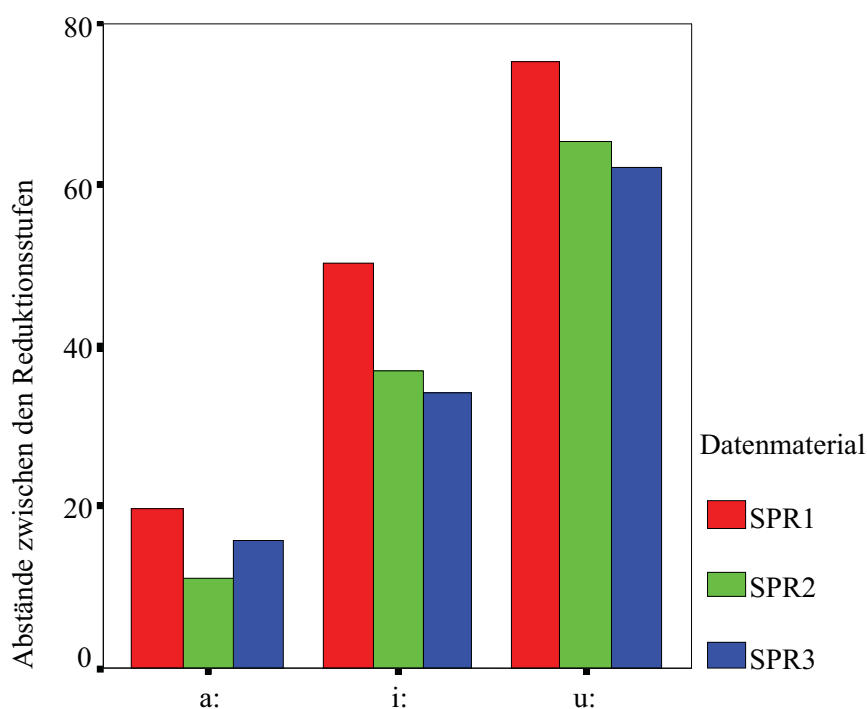
Im allgemeinen bringen die Ergebnisse aus dem sprecherspezifischen Material keine höheren Klassifikationsraten bei den beiden Verfahren auf, die auch nicht linear trennbare Daten unterscheiden können (SVM, NN), als die Klassifikationen mittels LDA. Die in den Funktionen der LDA verwendeten Diskriminanzvariablen weisen auf interindividuell unterschiedliche akustische Manifestationsbereiche der Reduktionsstufen hin. Die Klassifikation von /i:/ basiert hauptsächlich auf F2 (SPR1:  $cc = -0,545$ ; SPR2:  $cc = -0,725$ ; wobei  $cc$  für die Korrelation innerhalb der Gruppen zwischen den jeweiligen Diskriminanzvariablen und der



**Abbildung 39: Durchschnittliche Akkuratheit der neuronalen Netze für /a:/, /i:/ und /u:/ in Abhängigkeit vom Datenmaterial. Die Linie markiert die in den Perzeptionsexperimenten ermittelte durchschnittliche Übereinstimmung zwischen den Probanden.**

standardisierten kanonischen Diskriminanzfunktion steht; Tabelle F-5) und der Energie im zweiten Frequenzband (SPR3:  $cc = 0,753$ ; SPR1:  $cc = 0,521$ ). Während in die Diskriminanzfunktion des Datenmaterials von SPR1 noch F1 einfließt, sind es bei den Daten von SPR2 und SPR3 Lautdauer und F3. Mehr als 70% der Varianz wird durch diese Funktionen erklärt (SPR1: 83,0%, SPR2: 83,9%, SPR3: 71,6%; Tabelle F-4). Die Klassifikationen von /u:/ basieren bei SPR1 und SPR2 vor allem auf den Lautdauern und den Energiewerten in den ersten beiden Frequenzbändern, während beim Datenmaterial von SPR3 in erster Linie die F2-Werte einfließen (absolute  $cc \geq 0,7$ ; Tabelle F-7). Jedoch erklären die Diskriminanzfunktionen von /u:/ im Vergleich zu denen der Vokalkategorien /i:/ und /a:/ nur ein geringeren Teil der Varianz (SPR1: 52,8%, SPR2: 61,8%, SPR3: 60,9%; Tabelle F-6), was darauf hinweist, dass die Klassifikation von /u:/ komplexer ist als die der Vokalkategorien mit weniger Reduktionsstufen.

Die sprecherabhängigen Realisierungen der Reduktionsstufen von /a:/ hingegen können als vergleichbar betrachtet werden. Die LDA weisen darauf hin, dass die Einteilung in Reduktionsstufen hauptsächlich auf den Lautdauern beruht (Datenmaterial gesamt:  $cc = -0,907$ ; SPR1:  $cc = 0,790$ ; SPR2:  $cc = 0,841$ ; SPR3:  $cc = 0,825$ ; Tabelle F-9). Der größte Teil der Varianz wird durch diese Funktionen erklärt (Datenmaterial gesamt: 85,0%, SPR1: 68,0%, SPR2: 90,3%, SPR3: 87,6%; Tabelle F-8). Bei /i:/ und /u:/ zeigt sich, dass nur die Klassifikationsergebnisse der Daten von SPR1 vergleichbar sind mit den in den Perzeptionsexperimenten ermittelten Übereinstimmungsraten zwischen den Probanden. Eine mögliche Erklärung hierfür



**Abbildung 40: Kumulierte absolute Differenzen zwischen den Mittelwerten der akustischen Parameter (z-Werte) jeder Reduktionsstufe für /a:/, /i:/ und /u:/ in Abhängigkeit des sprecherspezifischen Datenmaterials.**

könnte in der Verteilung der akustischen Parameter liegen. Ein Vergleich der akustischen Parameter des sprecherspezifischen Datenmaterials (ANOVA; Tabelle F-10) weist jedoch keine Mittelwertsunterschiede auf. Auch die Varianzen sind mit Ausnahme von F2 und der Energie im zweiten Frequenzband bei /i:/ und der Energie im vierten Frequenzband bei /u:/ homogen (Tabelle F-11). Die kumulierten absoluten Differenzen zwischen den Mittelwerten der Parameter jeder Reduktionsstufe deuten darauf hin, dass die Reduktionsstufen von SPR2 und SPR3 im akustischen Raum enger zusammenliegen als die von SPR1 (Abbildung 40). Die Verteilung der akustischen Parameter gibt somit keine ausreichende Begründung für die geringeren Klassifikationsraten des Datenmaterials von SPR2 und SPR3.

Aufgrund der Ergebnisse wird angenommen, dass sich die akustischen Realisierungen der einzelnen Reduktionsstufen nicht nur inter-, sondern auch intraindividuell unterscheiden. Es scheint, dass die akustischen Realisierungen der Reduktionsstufen von SPR1 wesentlich konsistenter sind als die von SPR2 und SPR3. Ein Grund dafür kann sein, dass es sich bei SPR1 um eine professionelle Sprecherin handelt, so dass angenommen werden kann, dass sie aufgrund ihrer Erfahrungen eine gleichförmigere Artikulation aufweist als die beiden nicht professionellen Sprecher.



## 9.4 Zusammenfassung

Ausgangspunkt der Untersuchung war die Fragestellung, ob eine automatische Einteilung von Lautrealisierungen in Reduktionsstufen möglich ist. Dies wurde mit drei Klassifikationsverfahren getestet. Verwendet wurden lineare Diskriminanzanalysen, *Support Vector Machines* und künstliche neuronale Netze. Im Vergleich zu den Diskriminanzanalysen können durch *Support Vector Machines* und neuronale Netze auch nicht linear trennbare Daten klassifiziert werden. Es zeigt sich, dass die Klassifikationsleistungen unabhängig von den verwendeten Klassifikationsverfahren sind. Die Klassifikationsleistungen sind für den Vokal /a:/ sowohl beim sprecherspezifischen als auch beim gesamten Datenmaterial mit den in den Perzeptionsexperimenten festgestellten intersubjektiven Übereinstimmungen vergleichbar. Dies gilt auch für die /i:/- und /u:/-Varianten der professionellen Sprecherin. Die vergleichbaren Klassifikationsleistungen beim sprecherspezifischen und gesamten Datenmaterial der Vokalkategorie /a:/ sind auf eine relativ einfache Realisierung der Reduktionsstufen zurückzuführen. Aufgrund der Diskriminanzanalysen kann angenommen werden, dass von der Reduktion vor allem die Lautdauern betroffen sind, was auch durch die vorherigen Untersuchungen (s. Kapitel 7) unterstützt wird. Im Vergleich zu /a:/ scheint die Interaktion zwischen den akustischen Parametern der Reduktionsstufen von /i:/ und /u:/ wesentlich komplexer zu sein. Zum einen werden neben der Dauer auch die spektralen Eigenschaften der Vokale durch Reduktion beeinflusst, zum anderen weisen die Reduktionsstufen interindividuell unterschiedliche akustische Realisierungen auf.

Einen weiteren Hinweis auf eine sprecherabhängige Realisation von Reduktion gibt eine Untersuchung zur Übertragbarkeit der Reduktionsstufen auf andere Sprecher (Widera und Portele, 1999). Neuronale Netze, die mit den bezüglich der Reduktion beurteilten Lautrealisierungen der professionellen Sprecherin trainiert wurden, wurden zur Klassifikation der Vokale der beiden anderen Sprecher genommen. Auch hier zeigt sich, dass nur die Klassifikationsraten für das Datenmaterial der professionellen Sprecherin vergleichbar sind zu den in den Perzeptionsexperimenten ermittelten Übereinstimmungsraten zwischen den Probanden.

Die Ergebnisse legen die Vermutung nahe, dass die professionelle Sprecherin eine konsistentere Artikulationsstrategie aufweist als die ungeübten (nicht professionellen) Sprecher. Auch van Bergem (1993) konstatiert eine bessere Separation zwischen einzelnen Phonemen und weniger Streuung innerhalb jeder Phonemgruppe bei professionellen Sprechern im Vergleich zu ungeübten Sprechern.

Es kann davon ausgegangen werden, dass eine automatische sprecherspezifische Klassifikation von Reduktionsstufen prinzipiell möglich ist, jedoch hängt die Klassifikationsleistung vor allem von der intrasubjektiven Konsistenz der akustischen Realisierungen einzelner Reduktionsstufen ab. Die Sprecher gebrauchen unterschiedliche Strategien bei der Realisierung von Vokalreduktion.



## 10 ZUSAMMENFASSUNG UND AUSBLICK

In gesprochener Sprache finden sich unterschiedliche inter- und intraindividuelle Realisierungen von Vokalen. Diese Realisierungen weichen oft von der „kanonischen“ Form ab, der Form, die ein Laut annimmt, wenn er isoliert oder in einem neutralen Kontext gesprochen wird. Abweichungen von der kanonischen Form werden durch verschiedene Faktoren – z.B. Betonung, Wortklasse, Worthäufigkeit und Sprechgeschwindigkeit – beeinflusst und mit dem Begriff der Reduktion umschrieben (Kapitel 2). Nach dem klassischen Modell des *duration-dependent target-undershoot* (Lindblom, 1963) werden Vokalreduktionen durch eine Zunahme der Sprechgeschwindigkeit verursacht. Es wird davon ausgegangen, dass jede Vokalkategorie eine spezifische artikulatorische Zielstellung besitzt, wobei die Artikulatoren durch ihre Trägheit begrenzt sind. Wenn die zur Realisierung eines Vokals zur Verfügung stehende Zeit lang genug ist, erreichen die Artikulatoren die vokalspezifische Zielstellung und der produzierte Vokal entspricht der kanonischen Realisierung. Steht den Artikulatoren hingegen nur eine kurze Zeitspanne zur Verfügung, so können sie diese Zielstellung nicht mehr erreichen, was als *articulatory undershoot* bezeichnet wird. Reduktionen manifestieren sich aus akustischer Sicht durch eine Zentralisierung von F1 und F2, so dass sich den Formantwerten des neutralen Vokals ‚schwa‘ (*acoustic undershoot*) nähern. Reduzierte Vokale sind durch einen Verlust der Vokalqualität gekennzeichnet.

Auf der symbolischen Ebene werden Reduktionen üblicherweise mit dem IPA-Zeichensatz dargestellt, wobei die jeweiligen perzipierten Vokalqualitäten eine Reduktionsstufe bilden (Kapitel 4). Die Symbole erlauben für sich allein genommen aber noch keine Aussagen über den perzipierten Grad der Reduktion. Erst wenn sie hinsichtlich des Abweichungsgrads von der kanonischen Form geordnet werden, der auch mit dem artikulatorischen Aufwand assoziiert wird, wird ihre Reduktion ermittelt. Zum anderen besteht noch keine Einigkeit darüber, wie viele Reduktionsstufen innerhalb einzelner Vokalkategorien perzipiert werden können. Den Darstellungen der Reduktionsprozesse gespannter Vokale ist gemeinsam, dass eine Reduktionsstufe mit einer ungespannten Vokalqualität, typischerweise der in phonologischer Opposition stehenden Qualität, und eine andere mit einer [ə]-Qualität assoziiert wird. Unterschiede in den Beschreibungen von Vokalreduktionen finden sich vor allem in der Bewertung der perzeptiven Relevanz einer Dauerverminderung. In einigen Darstellungen wird davon ausgegangen, dass durch eine Kürzung der Dauer die Vokalqualität und –quantität gleichermaßen beeinflusst wird (z.B. [i:]→[ɪ]). In anderen Beschreibungen wird hingegen angenommen, dass Vokalquantität und –qualität voneinander unabhängig sind, so dass Reduktionen der Dauer nur einen Quantitätsverlust bewirken, während die Qualität erhalten bleibt (z.B. [i:]→[i]). Zudem zeigen Perzeptionsexperimente, dass eine Dauerverkürzung bei gespannten Vokalen nicht zwangsläufig zur Perzeption einer in phonologischer Opposition stehenden Vokalqualität führt, sondern andere Qualitäten wahrgenommen werden können, die auch aus akustischer Sicht eine

größere Nähe zu den jeweiligen gespannten Vokal aufweisen (z.B. [e:]→[ɪ]; Sendlmeier, 1981). Abgesehen von der Anzahl der unterscheidbaren Vokalqualitäten bei reduzierten Vokalen zeigen Untersuchungen zur Beurteilung reduzierter Vokale (z.B. van Bergem, 1995b; Aylett und Turk, 1998) im Vergleich zur Identifikation phonematischer Vokalqualitäten erheblich höhere intersubjektive Abweichungen auf.

### **10.1 Zusammenfassung der Ergebnisse**

Ziel dieser Arbeit war es, ein perzeptives Beschreibungssystem für Vokalreduktionen aufzustellen, das Reduktionen von Vokalqualitäten genau differenziert und in Stufen (von nicht reduziert bis stark reduziert) einteilt. Durch dieses System sollen Reduktionen nicht nur adäquat, sondern auch zuverlässig beschrieben werden. In Perzeptionsexperimenten wurde untersucht, ob Hörer Reduktionen reliabel in Stufen einteilen können und wie viele Stufen innerhalb einer Vokalkategorie perzipiert werden (Kapitel 6). In Kapitel 3 wurde daher der Frage nachgegangen, aufgrund welcher akustischer Eigenschaften Vokale perzipiert werden und welche Faktoren die Beurteilung der Vokalqualitäten beeinflussen. Um möglichst feine Differenzierungen von Vokalreduktionen zu erhalten, wurden Vergleichsexperimente durchgeführt. Als Untersuchungsmaterial dienten von drei Sprechern (eine professionelle und eine ungeübte Sprecherin sowie ein ungeübter Sprecher) gelesene Sätze und Kurzgeschichten (Kapitel 5).

Da die Perzeption von Vokalqualitäten primär auf den ersten beiden Formanten beruht, wurden die Formantwerte von Realisierungen innerhalb einer Vokalkategorie mit Hilfe von Clusterzentrenanalysen gruppiert (vgl. Kapitel 6). Die Annahme war, dass die Clusteranalysen Hinweise auf potentielle Prototypen von Reduktionsstufen geben. Prototypen dienen zur Erklärung der Kategorisierung von Einheiten.

Aus jedem Cluster wurde ein Referenzstimulus gewählt, der aus akustischer Sicht das Cluster am Besten repräsentiert (Laut in der Nähe des Clusterzentrums). Inwiefern diese Referenzstimuli prototypische Instanzen sind, wurde durch Zuordnungen weiterer Vokalrealisierungen getestet. Phonetisch geschulte Hörer hatten die Aufgabe, den Referenzstimulus zu benennen, der dem jeweiligen Stimulus qualitativ am ähnlichsten ist. Es zeigte sich, dass nicht alle Instanzen als Prototypen betrachtet werden können. Diese wurden ausgeschlossen, und die übrig gebliebenen Prototypen wurden in weiteren Perzeptionsexperimenten evaluiert. Weiterhin sollten die Hörer die Stärke der Reduktion der Prototypen und ihre Vokalqualität (enge Transkription) beurteilen. Aufgrund der bestehenden Ähnlichkeitsbeziehung wurde jeder Vokal mit der Reduktionsstufe des ihm zugeordneten Prototyps versehen. Die Bedeutung der Prototypen für die Einteilung von Lautrealisierungen in Reduktionsstufen wurde aufgrund der inter- und intrasubjektiven Beurteilungen geprüft.

Die Ergebnisse der Perzeptionsexperimente belegen, dass Hörer verschiedene Reduktionsstufen innerhalb einer Vokalkategorie zuverlässig, d.h. entsprechend den Übereinstimmungsraten in Perzeptionsexperimenten zur Identifikation phonematischer Lautqualitäten, diskriminieren können, und dass die Ähnlichkeitsbeurteilungen von Stimuli und Prototypen nicht durch Sprecherunterschiede beeinflusst werden. Sprecherunterschiede werden wie in alltäglichen Kommunikationssituationen kompensiert. Die Anzahl der Reduktionsstufen hängt jedoch von der Vokalkategorie ab. Bei /e:/ und /u:/ werden jeweils fünf Reduktionsstufen unterschieden, bei /i:/ vier und bei /ɛ:/, /y:/, /ø:/, /a:/ sowie /o:/ jeweils drei Reduktionsstufen. Für /ø:/ gilt die Unterscheidung von drei Reduktionsstufen nur eingeschränkt, da die intersubjektiven Beurteilungen nicht sehr zuverlässig sind. Während für die meisten gespannten Vokale Reduktionsstufen diskriminiert werden, nehmen Hörer bei ungespannten Vokalen nur für die an den Eckpunkten des Vokaldreiecks befindlichen Vokalkategorien /i/, /u/ und /a/ Reduktionsstufen wahr.

Der Vergleich der intersubjektiven Beurteilungen der Prototypen zeigt, dass die Hörer häufig identische Vokalqualitäten wahrnehmen, aber bei Verwendung der Diakritika voneinander abweichen. Die Prototypen der einzelnen Reduktionsstufen der gespannten Vokale werden – abgesehen von der jeweiligen gespannten Vokalqualität – mit der in Opposition stehenden ungespannten Qualität und einer zentralisierten ungespannten oder [ə]-Qualität perzipiert. Die Prototypen der Reduktionsstufen der ungespannten Vokale werden mit der jeweiligen ungespannten Vokalqualität und einer stärker zentralisierten bzw. [ə]-Qualität wahrgenommen. Bei den Vorderzungenvokalen /i:/, /ɪ/ und /e:/ wird zusätzlich eine gespannte Kurzvokalqualität unterschieden. Obwohl gespannte und ungespannte Kurzvokale akustische Unterschiede aufweisen, sind diese vor allem bei hohen Vokalen perceptiv relevant. Zwei Prototypen von /u:/ werden mit Vokalqualitäten assoziiert, die auf eine Koartikulation mit dem phonetischen Kontext zurückzuführen sind, bei /e:/ wird neben einer [ɛ]- eine [ɪ]-Qualität unterschieden. Ein Prototyp von /o:/ wird sowohl mit [ʊ] als auch mit [ə] beurteilt. Die geringeren intersubjektiven Abweichungen bei der Vergabe der Lautsymbole sowie ihr phonematischer Status im Deutschen weisen darauf hin, dass selbst phonetisch geschulte Hörer nur dann Vokalrealisierungen zuverlässig in Reduktionsstufen einteilen können, wenn sie sich an dem Phonemsystem ihrer Sprache orientieren.

In Kapitel 7 wurden die Reduktionsstufen akustischen Analysen unterzogen. Im ersten Teil (Abschnitt 7.1) wurde der Frage nach dem akustischen Manifestationsbereich der Reduktionsstufen nachgegangen. Vokalreduktionen werden nach dem *target undershoot*-Modell primär durch die Dauer bedingt. Je kürzer die Dauer eines Vokals ist, desto zentralisierter sind die Formantwerte. Die Hypothese der dauerabhängigen Reduktion ist jedoch umstritten. Analysen der akustischen Eigenschaften (erster bis dritter Formant, Dauer, Energien in vier verschiedenen Frequenzbändern und Grundfrequenz) der Reduktionsstufen zeigen, dass sich die Vokalrealisierungen benachbarter Reduktionsstufen vor allem durch ihre Formant- und

Dauerwerte unterscheiden. Perzipierte Reduktionsstärke und akustische Reduktion weisen tendenziell einen linearen Zusammenhang auf. Die Formanten sind mit zunehmender Reduktion zentralisiert, die Vokaldauern nehmen ab. Jedoch ist der Zusammenhang, mit Ausnahme von /a:/ und /ɛ:/, zwischen den perzipierten Reduktionsstärken und den spektralen Parametern stärker ausgeprägt als der zwischen den Reduktionsstärken und der Dauer. Eine Lautdauerverkürzung bewirkt also qualitative Unterschiede, ist aber selbst für die hier vorgenommene Einteilung der Vokalrealisierungen in Reduktionsstufen perzeptiv weniger relevant.

In Abschnitt 7.2 wurde untersucht, welche Hinweise die Reduktionsstufen für das phonologische System der in Opposition stehenden Vokalpaare geben. Die phonologische Gegenüberstellung der gespannten und ungespannten Vokale beruht auf Ähnlichkeitsannahmen zwischen den Vokalpartnern. Jedoch zeigen Perzeptionsexperimente zur Dauermanipulation, dass eine Kürzung der Vokaldauer bei gespannten mittel-geschlossenen Vokalen zur Perzeption des im Vokaldreieck höher liegenden ungespannten Vokals führt. Unabhängig hiervon weisen akustische Analysen auf eine größere spektrale Nähe der geschlossenen ungespannten zu den mittel-geschlossenen als zu den geschlossenen gespannten Vokalen hin. Wenn die gespannten Vokale die meisten gemeinsamen Eigenschaften mit den in phonologischer Opposition stehenden ungespannten Vokalen teilen, ist davon auszugehen, dass auch die Reduktionsverläufe dieser Vokalpaare vergleichbar sind.

Die Beurteilung der Prototypen der Reduktionsstufen zeigen, dass bei einigen Vokalen noch andere als die in phonologischer Opposition stehenden ungespannten Vokalqualitäten perzipiert werden. Reduzierte /e:/ und /o:/ werden als [ɪ] bzw. [ɔ] oder [ʊ] wahrgenommen, was den Ergebnissen von Perzeptionsexperimenten zur Dauermanipulation entspricht. Vergleiche der Reduktionsstufen der gespannten Vokale /i:/, /u:/ und /a:/ mit den jeweils in Opposition stehenden ungespannten Vokalen /ɪ/, /ʊ/ und /a/ zeigen, dass die Prototypen der ungespannten Vokale vergleichbare Vokalqualitäten wie die Prototypen der gespannten Vokale besitzen. Die einzigen Ausnahmen sind die gespannte Langvokalqualität und die koartikulatorisch bedingten Reduktionsstufen von /u:/. Im Gegensatz zu dieser systematischen Beziehung steht das Fehlen reliabler Reduktionsstufen für die übrigen ungespannten Vokale.

Bei den Paaren /a:-/a/, /ɛ:-/ɛ/, /u:-/ʊ/, /o:-/ɔ/ und /y:-/ʏ/ sind die akustischen Realisierungen der reduzierten gespannten Vokalen mit perzipierter ungespannter Qualität und der ungespannten (nicht reduzierten) Vokalen einander sehr ähnlich. Dies unterstützt die phonologische Gegenüberstellung dieser Vokalpaare. Für eine Gegenüberstellung von /i:/ vs. /ɪ/ bzw. /e:/ vs. /ɪ/ und /ø:/ vs. /œ/ bzw. /ø:/ vs. /ʏ/ finden sich keine eindeutigen Hinweise. Auf eine stärkere Ähnlichkeit von /ø:/ und /œ/ deutet nur die perzipierte ungespannte Qualität des Prototyps der gespannten Vokalkategorie hin. Für die phonologische Gegenüberstellung von /i:/ und /ɪ/ sprechen aber die vergleichbaren Reduktionsverläufe sowie die akustischen Realisierungen der Reduktionsstufen mit gespannter Kurzvokalqualität.

Da nur für die an den Eckpunkten des Vokaldreiecks liegenden ungespannten Vokale /ɪ/, /ʊ/ und /a/ Reduktionsstufen unterschieden werden, liegt die Vermutung nahe, dass bei den übrigen ungespannten Vokalen der spektrale Abstand zum /ə/ zu gering ist, um Reduktionsstufen reliabel zu diskriminieren. Eine genauere Untersuchung ergibt, dass die Diskriminierbarkeit von Reduktionsstufen von der spektralen Distanz zum /ə/ abhängt. Die Vokale /ɛ/, /ɔ/, /ʏ/ und /œ/ weisen kleinere Abstände zum Zentralvokal auf als die reduzierten oder stark reduzierten Lautrealisierungen ihrer gespannten Vokalpartner, so dass die perzipierten Unterschiede zwischen ungespannter Vokalqualität und [ə]-Qualität nicht ausreichen, um weitere Reduktionsstufen reliabel zu diskriminieren. Dies entspricht den Ergebnissen von Perzeptionsexperimenten, wonach die ungespannten Vokale [ɛ], [ɔ] und [œ] eine qualitative Ähnlichkeit zu [ə] aufweisen (Ungeheuer, 1962b, 1964). Jedoch kann aufgrund des absoluten spektralen Abstandswerts der unreduzierten Vokalrealisierungen nicht auf die Anzahl der Reduktionsstufen geschlossen werden. Werden jedoch die Distanzwerte von den in Opposition stehenden Vokalkategorien gegenübergestellt, so zeigt sich, dass vergleichbare Abstände zu einer vergleichbaren Anzahl von Reduktionsstufen führen. Dies spricht dafür, dass die Distanzen zum /ə/ in Abhängigkeit der Vokalkategorie unterschiedlich wahrgenommen werden.

Im letzten Teil (Abschnitt 7.3) wurde nach einer akustischen Beschreibung der Ähnlichkeitsbeziehung zwischen Prototypen und Stimuli gesucht. Realisierungen der Vokalkategorien wurden aufgrund ihrer qualitativen Ähnlichkeit den Prototypen zugeordnet. Nach der Theorie des Perzeptiven Magneteffekts (Kuhl, 1992; vgl. Kapitel 3) werden Laute einer Vokalkategorie, die akustisch in der Nähe eines Prototyps liegen, von diesem angezogen, so dass die akustischen Unterschiede perzeptiv unbedeutsam sind. Das Ausmaß der Anziehungskraft eines Vokals hängt zum einen von der akustischen Distanz zwischen ihm und dem zu vergleichenden Laut ab; je größer der Abstand ist, um so unähnlicher wird er empfunden. Zum anderen ist die Anziehungskraft eines Vokals vermindert, wenn er eine eher untypische Instanz der Kategorie ist. Wenn die Prototypen also perzeptive Magnete sind, dann werden die Stimuli umso zuverlässiger dem Prototypen zugeordnet, je geringer ihre akustische Distanz zu diesem Prototypen ist. Die Annahme war, dass durch die akustischen Distanzen von Stimuli und Prototyp die in den Vergleichsexperimenten erfolgten Zuordnungen vorhersagen können.

Die Ergebnisse lassen sich nur bedingt mit der Theorie des Perzeptiven Magneteffekts erklären. Es zeigt sich, dass die perzipierten Ähnlichkeitsbeziehungen vor allem durch die spektralen Parameter und die Dauer erklärt werden können. Jedoch scheinen nicht alle Parameter die gleiche Relevanz für die Zuordnung zu den Prototypen zu besitzen. Die Ähnlichkeitszuweisungen der hohen Vokale werden primär durch die spektralen Parameter beeinflusst, die von /a:/ durch die Dauer.

Die von der Theorie des Perzeptiven Magneteffekts abweichenden Ergebnisse können durch die Lage der Prototypen im akustischen Raum erklärt werden. Nicht alle Prototypen befinden sich in der Mitte einer Reduktionsstufe, sondern nehmen auch extremere Lagen ein, so dass die geringste



Distanz zu einem Prototypen nicht immer ein adäquates Maß zur Beschreibung der Ähnlichkeitsbeziehung ist. Die Zuweisungen von Stimuli zu Prototyp können besser vorhergesagt werden, wenn neben den akustischen Distanzen zum Prototypen noch weitere Faktoren, z.B. die Dauer der Stimuli selbst und die Distanz zu benachbarten Prototypen, hinzugezogen werden. Dann ist die Trefferrate mit den interindividuellen Übereinstimmungen der Hörerurteile vergleichbar. Dies deutet darauf hin, dass einzelne Parameter unterschiedlich zu gewichten sind und die Ähnlichkeitsurteile möglicherweise durch den Kontext, d.h. durch die Präsenz weiterer Vergleichsstimuli beeinflusst werden.

Die Realisierung von Vokalen wird von vielen Faktoren beeinflusst. Im allgemeinen gilt, dass Wörter, die einen hohen Informationsgehalt besitzen, genauer und betonter artikuliert werden als Wörter, die weniger zur Informationsübermittlung beitragen. In Kapitel 8 wurde daher überprüft, ob es eine Beziehung zwischen den in Reduktionsstufen unterteilten Vokalrealisierungen und der Wortklasse (Funktions vs. Inhaltswörter) sowie der Betonung (lexikalische Betonung, Akzent, perzeptive Prominenz von Silben) gibt. Erwartungsgemäß zeigt sich, dass Vokale in Inhaltswörtern weniger reduziert sind als in Funktionswörtern. Funktionswörter sind aber nicht nur einfach reduziert, sondern ihre Vokalrealisierungen weisen das volle Spektrum der Reduktionsstufen auf. Die Interaktionen von Wortklasse, Betonung und Reduktion zeigen, dass der Faktor Betonung einen größeren Einfluss auf die Lautrealisierung besitzt als die Wortklasse. Vokale in akzentuierten Silben weisen keine Reduktion auf. Silben von Funktionswörtern, deren Nukleus nicht reduziert ist, werden mit vergleichbarer Prominenz wahrgenommen wie Silben von Inhaltswörtern mit nicht reduzierten Vokalen. Korrelationen zwischen der perzipierten Betonungsstärke von Silben und dem wahrgenommenen Reduktionsgrad von Vokalen unterstützen die Annahme, dass eine Reduktion von Vokalen zur perzipierten Prominenz von Silben beitragen. Es zeigt sich die Tendenz, dass Silben um so betonter wahrgenommen werden, desto weniger der Vokal reduziert ist. Die Ergebnisse weisen darauf hin, dass sich mit dem vorgestellten Beschreibungssystem Vokalreduktionen nicht nur auf der segmentalen Ebene darstellen lassen, sondern sich in den Reduktionsstufen auch die Interaktion von Reduktion und prosodischen Faktoren widerspiegelt.

Da die akustischen Analysen der Reduktionsstufen belegen, dass sich die Reduktionsstufen durch einzelne Parameter voneinander unterscheiden, wurde in Kapitel 9 der Frage nachgegangen, inwiefern Vokalrealisierungen innerhalb einer Vokalkategorie auch automatisch klassifiziert werden können. Dies wurde anhand der in Reduktionsstufen eingeteilten Vokalrealisierungen von /i:/, /u:/ und /a:/ mit drei Klassifikationsverfahren, der linearen Diskriminanzanalyse, *Support Vector Machines* und künstlichen neuronalen Netzen, getestet. Im Vergleich zu den linearen Diskriminanzanalysen können durch *Support Vector Machines* und neuronale Netze auch nicht linear trennbare Daten klassifiziert werden. Es zeigt sich, dass die Klassifikationsleistungen unabhängig von den verwendeten Verfahren sind. Vielmehr legen die Ergebnisse die Vermutung nahe, dass sich die akustischen Realisierungen der Reduktionsstufen der einzelnen Sprecher

voneinander unterscheiden. Im Vergleich zum Datenmaterial der ungeübten Sprecher entsprechen unabhängig von der Vokalkategorie die beim Material der professionellen Sprecherin erzielten Klassifikationsleistungen den interindividuellen Übereinstimmungsraten der Perzeptionsexperimente. Zu vermuten ist, dass die professionelle Sprecherin eine konsistentere Artikulationsstrategie aufweist als die nicht professionellen Sprecher. Da beim Datenmaterial der professionellen Sprecherin Klassifikationsleistungen erreicht werden, die den intersubjektiven Übereinstimmungsraten vergleichbar sind, könnte eine automatische sprecherspezifische Klassifikation von Reduktionsstufen prinzipiell möglich sein.

## **10.2 Diskussion und Ausblick**

Insgesamt zeigen die Untersuchungen, dass durch Zuweisung von Vokalrealisierungen zu Prototypen Vokale reliabel in Reduktionsstufen eingeteilt werden können. Im Vergleich zu den üblichen Darstellungen der Vokalreduktion, in denen häufig davon ausgegangen wird, dass innerhalb der Gruppe der gespannten bzw. der ungespannten Vokale eine gleiche Anzahl von Reduktionsstufen existiert, zeigt sich, dass die Anzahl der Reduktionsstufen vokalabhängig ist und dass noch weitere als die in phonologischer Opposition stehenden Lautqualitäten berücksichtigt werden müssen, um Reduktionen von Vokalen adäquat zu beschreiben.

Als Versuchsmaterial wurde gelesene Sprache verwendet. Lesesprache ist im Vergleich zur Spontansprache durch eine geringere Reduktion gekennzeichnet. Es muss daher noch geprüft werden, ob das hier vorgestellte Klassifikationssystem eine adäquate Beschreibung von Vokalreduktionen in spontaner Sprache gewährleistet. Möglich ist, dass für spontane Sprache noch zusätzliche Reduktionsstufen angenommen werden müssen. Dagegen spricht, dass nur dann Vokalrealisierungen in Reduktionsstufen zuverlässig eingeteilt werden, wenn die perzipierten Abstände denen von Phonemen entsprechen. Zudem weisen erste Voruntersuchungen darauf hin, dass Spontansprache nicht durch größere Unterschiede in den Vokalqualitäten gekennzeichnet ist, sondern durch eine häufige Elision von Lauten (Widera und Hess, in Vorbereitung).

Weiterhin stellt sich die Frage nach dem Verhältnis der Reduktionsstärke gespannter und ungespannter Vokalkategorien. Die Reduktionsstärken der Prototypen wurden innerhalb einer Vokalkategorie beurteilt, wobei gespannte Vokalkategorien mindestens eine Reduktionsstufe mehr besitzen als die ungespannten. Während die akustischen Analysen darauf hinweisen, dass reduzierte gespannte Vokale vergleichbare Formant- und Dauerwerte besitzen wie nicht reduzierte ungespannte Vokale mit vergleichbar wahrgenommener Qualität, zeigen hiervon unabhängig durchgeführte artikulatorische Untersuchungen (Abschnitt 4.4.2), dass ungespannte Vokale nicht als reduzierte gespannte Vokale betrachtet werden können. Es ist also noch offen, inwieweit sich reduzierte gespannte und nicht reduzierte ungespannte Vokalqualitäten perzeptiv voneinander unterscheiden.

Nicht vollständig geklärt ist, wie Reduktionsstufen sich akustisch manifestieren. Alle Analysen, sowohl zur Trennbarkeit der Reduktionsstufen als auch zur Ähnlichkeitsbeziehung zwischen den Prototypen und Stimuli weisen auf eine primäre perzeptive Relevanz der ersten beiden Formantfrequenzen sowie, insbesondere bei den „a“-Lauten, auf die Bedeutung der Lautauern hin. Gleichzeitig zeigt sich aber auch, dass durch unterschiedliche Parameterwerte vergleichbare Perzeptionseindrücke hervorgerufen werden. Es müsste daher überprüft werden, inwiefern eine Berücksichtigung anderer Eigenschaften der Formanten, wie z.B. der Formantbandbreite, eine genauere Beschreibung der Perzeption von Vokalreduktion ermöglicht.

In diesem Zusammenhang stellt sich auch die Frage, ob für eine Beschreibung von perzeptiven Eindrücken anstelle von physikalischen Maßeinheiten nicht Größen verwendet werden sollten, die die im Gehör stattfindenden Transformationen berücksichtigen. Voruntersuchungen weisen darauf hin, dass die hier verwendeten Daten durch Transformation der Formantwerte in die Bark- oder mel-Skala nicht besser getrennt werden (vgl. auch Hillenbrand und Gayvert, 1993). Auch wenn Parameter zur Beschreibung der Ähnlichkeitsbeziehung zwischen Prototyp und zugeordnetem Stimulus verwendet werden, die auf dem Perzeptionsmodell (PEMO; Dau et al., 1996) basieren, ergeben sich keine Vorteile gegenüber physikalischen Maßeinheiten. Wie die Ergebnisse zur Modellierung der Ähnlichkeitsbeziehung zeigen, können Ähnlichkeitsurteile bei komplexen Stimuli nur unzureichend durch einfache geometrische Distanzmodelle, wobei alle Parameter zu gleichen Anteilen berücksichtigt werden, beschrieben werden. Ähnlichkeitsurteile hängen von der Aufmerksamkeit ab, die Hörer den einzelnen Eigenschaften beimessen. Sie unterliegen einem dynamischen Prozess und hängen vom Kontext ab (Spencer-Smith und Goldstone, 2001).

Um weitere Erkenntnisse zur Modellierung der Ähnlichkeitsbeziehung zu erhalten, sollte auch das Beurteilungsverhalten nicht phonetisch ausgebildeter Personen untersucht werden. Beispielsweise weisen Experimente zur Vokalklangwahrnehmung von Jerrentrup und Thürmann (1974) auf eine unterschiedliche Beurteilung der akustischen Parameter von „Experten“ und „Laien“ hin. Die hier vorgestellte Methodik, Zuordnung von Stimuli zu Prototypen aufgrund ihrer qualitativen Ähnlichkeit, setzt keine phonetischen Kenntnisse voraus, so dass mit ein- und demselben Paradigma die Wahrnehmung der Vokalreduktion durch „Laien“ und durch „Experten“ untersucht werden kann. Die so erzielten Ergebnisse können miteinander gut verglichen werden.

Die im Rahmen dieser Arbeit durchgeführten Untersuchungen zeigen, dass die Reduktionsstärke von Vokalrealisierungen durch qualitative Vergleiche mit Prototypen von Reduktionsstufen beurteilt werden kann. Der Versuchsaufbau war so gestaltet, dass er der Vorgehensweise beim Etikettieren von Datenmaterial sehr nahe kommt. Die Stimuli wurden in keiner Weise akustisch nachbehandelt. Aufgrund der leichten Handhabung sowie der nur geringfügigen interindividuellen Abweichungen bei der Zuordnung der Stimuli zu den Prototypen eignet sich diese Methodik für eine reliable Beschreibung der Vokalreduktion in Datenbanken auch dann, wenn verschiedene (phonetisch geschulte) Hörer für die Beurteilung von Vokalrealisierungen herangezogen werden.

Untersuchungen von Vokalreduktion ermöglichen nicht nur Einblicke in die Sprachperzeption, sondern ihre Ergebnisse tragen auch zum besseren Verständnis des menschlichen Kommunikationsprozesses bei und sind somit auch für die anwendungsorientierte Forschung, beispielsweise der Sprachsynthese, von Bedeutung.

Die Qualität der Sprachausgabe eines Synthesystems wird anhand der Kriterien Verständlichkeit und Natürlichkeit gemessen. Bestehende Sprachsynthesysteme besitzen eine akzeptable Verständlichkeit, jedoch klingen sie recht unnatürlich (Holmes, 1991; Hess et al., 1994; Kraft und Portele, 1995). Werden in datengesteuerten Systemen Grundbausteine verwendet, die aus natürlichsprachlichen Äußerungen (z.B. Diphone, Halbsilben) entnommen sind, und diese konkateniert, so kommt es häufig zu Diskontinuitäten (Kraft, 1993). Eine Manipulation der Sprachsignale beschränkt sich auf Intensität, Dauer und Grundfrequenz (Hess et al., 1994). In neueren Ansätzen (z.B. Campbell, 1999; Stöber et al., 2000) wird versucht, durch eine Auswahl von möglichst großen Bausteinen (z.B. Äußerungsteile auf der Phrasenebene) die Qualität der natürlichsprachlichen Äußerung beizubehalten. Natürlich kann die Datenbasis der Sprachsynthese nicht alle jemals zu synthetisierenden Äußerungen enthalten, so dass auch auf kleinere Einheiten (z.B. Wörter, Silben) zurückgegriffen werden muss. Der Kontext der Einheiten sollte dabei vergleichbar zu dem in der zu synthetisierenden Äußerung sein. Bei der Auswahl der Einheiten müssen beispielsweise ihre Position innerhalb der Äußerung, der Satzmodus, ihre Prominenz und ihre Reduktion berücksichtigt werden (vgl. Stöber et al., 1999). Die Natürlichkeit der Synthese hängt daher vom Inventar, von der Auswahl der Bausteine und somit auch von deren Etikettierung ab.

Da Perzeptionsexperimente im Rahmen von einheitenbasierten Sprachsynthesystemen zeigen, dass kleinere Abweichungen von der kanonischen Form nicht von Hörern wahrgenommen werden (Portele, 1996, 1997), stellt sich die Frage, welche Einheiten im Inventar aufgenommen werden müssen. Durch die vorgestellte Methode können Sprachbausteine für ein Synthesystem hinsichtlich ihrer Vokalreduktion beurteilt werden, so dass wahrnehmbare Vokalqualitätsunterschiede etikettiert sind. Dies stellt zum einen sicher, dass bei Konkatenation im Vokal Diphone mit vergleichbarer Reduktionsstärke ausgewählt werden können, so dass große spektrale Sprünge vermieden bzw. zumindest verringert werden. Zum anderen können die hinsichtlich ihrer Reduktionsstärke beurteilten Bausteine auch in Beziehung zu den prosodischen Steuerungsparametern eines Synthesystems gesetzt werden. Die perzeptive Prominenz von Silben hängt auch von der Reduktion des Silbennukleus ab, stark betont wahrgenommene Silben weisen keine reduzierten Vokale aus. Insbesondere Funktionswörter, die ein breites Spektrum an Vokalreduktionen aufweisen, können somit entsprechend ihrem Kontext adäquat synthetisiert werden.



## LITERATUR

- Akagi, M. (1990): Evaluation of a spectrum target prediction model in speech perception. *Journal of the Acoustical Society of America* 87, 858-865.
- Akagi, M.; van Wieringen, A. und Pols, L. C. W. (1994): Perception of central vowel with pre- and post-anchors. *Proc. 3<sup>rd</sup> International Conference on Spoken Language Processing*, Yokohama Japan, 503-506.
- Alfonso, P. J.; Honda, K. und Baer, T. (1984): Coordinated tongue muscle activity during [əpVp] utterances. *Proc. 10<sup>th</sup> International Congress of Phonetic Sciences*, Netherlands Phonetic Archives, 390-394.
- Antoniadis, Z. und Strube, H. W. (1981): Untersuchungen zum <<intrinsic pitch>> deutscher Vokale. *Phonetica* 38, 277-290.
- Arnold, W.; Eysenck, H. J. und Meili, R. (1980): Lexikon der Psychologie. Band3, Neuauflage. Herder, Freiburg.
- Aylett, M. and Turk, A. (1998): Vowel quality in spontaneous speech: What makes a good vowel? *Proc. 5<sup>th</sup> International Conference on Spoken Language Processing*, Sydney Australia, [CD-ROM], Paper 824.
- Baer, T.; Alfonso, P. J. und Honda, K. (1988): Electromyography of the tongue muscles during vowels in [əpVp] environment. *Annual Bulletin RILP* 22, 7-19.
- Bannert, R. (1985): Fokus, Kontrast und Phrasenintonation im Deutschen. *Z. für Dialektologie und Linguistik* 52, 289-305.
- Barry, W.J. (1995): Schwa vs. Schwa + /r/ in German. *Phonetica* 52, 228-235.
- Beckman, M. (1995): A typology of spontaneous speech. *Proc. ATR workshop on Computational Modelling of Prosody for Spontaneous Speech Processing*, ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan.
- Beckman, M.; Jung, T-P.; Lee, S-H.; de Jong, K.; Krishnamurthy, S.; Cohen; K. B. und Collins, M. (1995): Variability in the production of quantal vowels revisited. *Journal of the Acoustical Society of America* 97, 471-490.
- Beddor, P. S. und Hawkins, P. (1990): The influence of spectral prominence on perceived vowel quality. *Journal of the Acoustical Society of America* 87, 2864-2704.
- Bell, A.; Jurafsky, D.; Fosler-Lussier, E. Girand, C. und Gildea, D. (1999): Forms of English function words – effects of disfluencies, turn position, age and sex, and predictability. *Proc. 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, 395-398.
- Bell-Berti, F.; Raphael, L. J.; Pisoni, D. B. und Sawusch, J. R. (1979): Some relationships between articulation and perception. *Phonetica* 36, 373-383.
- Benguerel, A.-P. und McFadden, T. U. (1989): The effect of coarticulation on the role of transitions in vowel perception. *Phonetica* 46, 80-96.
- Boersma, P. (1998): *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*. Holland Academic Graphics, The Hague.

- Bohn, O.-S.; Flege, J.; Dagenais, P. und Fischer, S. (1991): Differenzierung und Variabilität der Zungenposition bei der Artikulation deutscher Vokale. *Z. für Dialektologie und Linguistik* (Beihefte), 1-26.
- Bolinger, D. (1958): A theory of pitch accent in English. *Word* 14, 109-149.
- Bolinger, D. (1972): Accent is predictable (if you're a mind-reader). *Language* 48, 633-644.
- Bolinger, D. (1989): *Intonation and its uses – Melody in grammar and discourse*. Edward Arnold, London.
- Bolla, K. und Valaczkai, L. (1986): Német Beszéhangok Atlasza (A phonetic conspectus of German). *Magy. Fonetikai Füzetek* 16, Magyar Tudományos Akadémia Nyelvtudományi Intezete, Budapest.
- Bortz, J. (1979): *Lehrbuch der Statistik. Für Sozialwissenschaftler*. Springer Verlag, Berlin.
- Breuer, S. (1999). *Zur Regelmäßigkeit segmentaler Reduktionen in deutschen Funktionswörtern*. Magisterarbeit Universität Bonn.
- Brosius, G. und Brosius, F. (1995): *SPSS. Base System und Professional Statistics*. An International Thomson Publishing Company, Bonn.
- Browman, C. P. und Goldstein, L. (1989): Articulatory gestures as phonological units. *Phonology* 6, 201-251.
- Browman, C. P. und Goldstein, L. (1990): Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, M. E. Beckman (Hg.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge University Press, Cambridge.
- Browman, C. P. und Goldstein, L. (1992): „Targetless“ schwa: an articulatory analysis. In G. J. Docherty und D. R. Ladd (Hg.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge University Press, 26-67.
- Brustkern, J. (1992): *Maschinenlesbare Lexika für die maschinelle Sprachverarbeitung: Repräsentation und Wiederverwendung*. Holos Reihe Linguistik 3, Bonn.
- Bühl, A. und Zöfel, P. (1998): *SPSS für Windows Version 7.5. Praxisorientierte Einführung in die moderne Datenanalyse*. 4. Aufl., Addison-Wesley, Bonn
- Burges, C. J. C. (1998): A tutorial on Support Vector Machines for pattern recognition. *Knowledge Discovery and Data Mining* 2(2), 1-43.
- Bussemeyer, J. R.; Dewey, G. I. und Medin, D. L. (1984): Evaluation of exemplar-based generalization and the abstraction of categorization information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10, 638-648.
- Bußmann, H. (1990): *Lexikon der Sprachwissenschaft*. 2. Aufl., Kröner, Stuttgart.
- Butcher, A. (1976): The influence of the native language on the perception of vowel quality. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)* 6.
- Campbell, N. (1995): Prosodic influence on segmental quality. *Proc. Eurospeech*, Madrid, 1011-1014.
- Campbell, N. (1999): Data-driven speech synthesis. *Forum Acusticum*, Berlin, 103.



- Carlson, R.; Fant, G. und Granström, B. (1975): Two-formant models, pitch and vowel perception. In: G. Fant und M. A. A. Tatham (Hg.), *Auditory Analysis and Perception of Speech*. Academic, London, S. 55-82.
- Catford, J. C. (1977): *Fundamental problems in phonetics*. Edinburgh University Press, Edinburgh.
- Chistovich, L. (1971): *Problems of speech perception. Form and Substance*. Akademisk Forlag, Copenhagen.
- Chistovich, L. A. und Lublinskaya, V. V. (1979): The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Journal of Hearing Research 1*, 185-195.
- Chomsky, N. und Halle, M. (1968): *Sound Pattern of English*. Harper & Row, New York.
- Clark, H. und Yallop, C. (1995): *An introduction to phonetics and phonology*. Blackwell, Oxford.
- Crystal, T. M. und House, A. S. (1988): The duration of American-English vowels: an overview. *Journal of Phonetics 16*, 263-284.
- Dau, T.; Püschel, D.; Kohlrausch, A. (1996): A quantitative model of the "effective" signal processing in the auditory system I+II. *Journal of the Acoustical Society of America 99*, 3615-3631.
- Dedenbach, B. (1987): *Reduktions- und Verschmelzungsformen im Deutschen. Schwache Formen bei Artikeln und Pronomina*. Lang, Frankfurt a. M.
- Delattre, P. C.; Liberman, A. M. und Cooper, F. S. (1955): Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America 27*, 769-773.
- Delattre, P. C.; Liberman, A.M.; Cooper, F. S. und Gerstman, L. (1952): An experimental study of the acoustic determinants of vowel colour. *Word 8*, 195-210.
- Di Benedetto, M. G. (1989): Frequency and time variations of the first formant: Properties relevant to the perception of vowel height. *Journal of the Acoustical Society of America 86*, 67-77.
- Diehl, R. L.; Lindblom, B.; Hoemeke, K. A. und Fahey, R. P. (1996): On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics 24*, 187-208.
- Disner, S. F. (1980): Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America 67*, 253-261.
- Duda, R. O.; Hart, P. E. und Stork, D. G. (2001): *Pattern Classification*. 2. Aufl., John Wiley and Sons, New York.
- Duden, Aussprachewörterbuch* (1990). 3. Aufl., Dudenverlag, Mannheim.
- Eisenberg, P. (1989): *Grundriss der deutschen Grammatik*. 2. Aufl., J. B. Metzlersche Verlagsbuchhandlung, Stuttgart.
- Engstrand, O. (1988): Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America 83*, 1863-1875.
- Eskénazi, M. (1993): Trends in Speaking Styles Research. *Proc. Eurospeech*, Berlin, 501-509.
- ESPS* (Version 5.0), Entropic Research Laboratory.
- Fahey, R. P.; Diehl, R. L. und Traunmüller, H. (1996): Perception of back vowels: Effects of varying F<sub>1</sub>-F<sub>0</sub> Bark distance. *Journal of the Acoustical Society of America 99*, 2350-2357.

- Fairbanks, G. und Grubb, P. (1961): A psychophysical investigation of vowel formants. *Journal of Speech and Hearing Research* 4, 203-219.
- Fant, G. (1959): *Acoustic analysis and synthesis of speech with applications to Swedish*. Ericsson, Stockholm.
- Fant, G. (1966): A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, KTH Stockholm, 4, 22-30.
- Fant, G. (1970): *Acoustic theory of speech production*. 2. Aufl., Mouton, The Hague.
- Fant, G. (1975): Non-uniform vowel normalization. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, KTH Stockholm, 2-3, 1-19.
- Fant, G. (1983): Feature analysis of Swedish vowels – a revisit. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, KTH Stockholm, 2-3, 1-19.
- Fant, G. und Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, KTH Stockholm, 2,1-83.
- Fant, G.; Hermingson, U. und Stålhammer, U. (1969): Formant frequencies of Swedish vowels. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, KTH Stockholm, 4, 26-28.
- Faust, L. (1997): *Variationen von Sprache: Ihre Bedeutung für unser Ohr und für die Sprachtechnologie*. Verlag Dr. Kovač, Hamburg.
- Fischer-Jørgensen, E. (1990): Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica* 47, 99-140.
- Fischer-Jørgensen, E. und Jørgensen, H. P. (1969): Close and loose contact ('Anschluß') with special reference to North German. *Annual Report of the Institute of Phonetics of the University of Copenhagen* 4, 43-80.
- Fox, R. A. (1982): Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica* 39, 1-22.
- Fox, R. A. (1989): Dynamic information in the identification and discrimination of vowels. *Phonetica* 46, 97-116
- Frieda, E. M.; Walley, A. C.; Flege, J. E. und Sloane, M. E. (2000): Adults' perception and Production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research* 43, 129-143.
- Fry, D. B.; Abramson, A. S.; Eimas, P. D. und Liberman, A. M. (1962): The identification and discrimination of synthetic vowels. *Language and Speech* 5, 171-189.
- Fry, D.B. (1958): Experiments in the perception of stress. *Language and Speech* 1, 126-152.
- Fujisaki, H. und Kawashima, T. (1971): A model of the mechanisms for speech perception: Quantitative analyses of categorical effects in discrimination. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 59-68.
- Gerrits, E. und Schouten, B. (1998): Categorical perception of vowels. *Proc. 5<sup>th</sup> International Conference on Spoken Language Processing*, Sydney Australia, 2279-2283.

- Heid, S. J. G. G. (1997): Individual differences between vowel systems of German speakers. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 35, 27-34.
- Heid, S. J. G. G. (1998): Phonetische Variation. Untersuchungen anhand des PhonDAT2-Korpus. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 36, 193-368.
- Heike, G. (1970): Lautdauer als Merkmal der wahrgenommenen Quantität, Qualität und Betonung im Deutschen. *Proc. 6<sup>th</sup> International Congress of Phonetic Sciences*, Prag, 433-437.
- Heike, G. (1971): Quantitative und qualitative Differenzen von [a(:)]-Realisationen im Deutschen. *Proc. 7<sup>th</sup> International Congress of Phonetic Sciences*, Montreal, 725-729.
- Heike, G. (1982): *Phonologie*. 2. Aufl., Metzler, Stuttgart.
- Heike, G. (1984): Phonetische und phonologische Aspekte eines Artikulationsmodells: Segment, Silbe und ‚Wort‘ als Artikulationsbewegungen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 19, Universität München, 62-88.
- Heß, K.; Brustkern, J. und Lenders, W. (1983): *Maschinenlesbare deutsche Wörterbücher*. Sprache und Information 6, Max Niemeyer Verlag, Tübingen.
- Hess, W. (1972): *Digitale Grundfrequenzsynchrone Analyse von Sprachsignalen als Teil eines automatischen Spracherkennungssystems*. Dissertation Technische Universität München.
- Hess, W.; Kraft, V. und Portele, T. (1994): Zum Problem der Evaluierung von Sprachsynthesystemen – dargestellt am Beispiel der Synthesekomponenten in VERBMOBIL. *Fortschritte der Akustik – DAGA*, Dresden, 103-117.
- Heuft, B. (1999): Eine prominenzbasierte Methode zur Prosodieanalyse und –synthese. In W. Hess und W. Lenders (Hg.), *Sprache, Sprechen und Computer*, Bd. 2, Peter Lang, Frankfurt am Main.
- Heuft, B. und Portele, T. (1996). Synthesizing prosody: A prominence-based approach. *Proc. 4<sup>th</sup> International Conference on Spoken Language Processing*, Philadelphia, 1361-1364.
- Heuft, B.; Portele, T.; Höfer, F.; Krämer, J. Meyer, H.; Rauth, M. and Sonntag, G. (1995): Parametric description of F<sub>0</sub>-contours in a prosodic database. *Proc. 13<sup>th</sup> International Congress of Phonetic Sciences*, Stockholm, 378-381.
- Hillenbrand, J. und Gayvert, R. T. (1993): Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech and Hearing Research* 36, 694-700.
- Hoemeke, K. A. und Diehl, R. L. (1994): Perception of vowel height: The role of F<sub>1</sub>-F<sub>0</sub> distance. *Journal of the Acoustical Society of America* 96, 661–674.
- Hoole, P. und Kühnert, B. (1995): Patterns of lingual variability in German vowel production. *Proc. 13<sup>th</sup> International Congress of Phonetic Sciences*, Stockholm, 442-445.
- Hoole, P. und Kühnert, B. (1996): Tongue-jaw coordination in German vowel production. *Proc. 1<sup>st</sup> ESCA Tutorial and Research Workshop on Speech Production Modeling: From Control Strategies to Acoustics*, 97-100.
- Hoole, P.; Mooshammer, C. und Tillmann, H. G. (1994): Kinematic analysis of vowel production in German. *Proc. 3<sup>rd</sup> International Conference on Spoken Language Processing*, Yokohama Japan, 53-56.
- IPA (1993): *Reproduction of the International Phonetic Alphabet* (Revised to 1993, updated 1996). URL: [//www2.arts.gla.ac.uk /IPA/ipa.html](http://www2.arts.gla.ac.uk/IPA/ipa.html).

- Iverson, P. und Kuhl, P. K. (1995): Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America* 97, 553-561.
- Jakobson, R.; Fant, G. und Halle, M. (1951) Preliminaries to Speech Analyses. In: S. Rubyy (Hg.), *Roman Jakobson – Selected Writings*. VIII Completion Volume One. Mouton de Gruyter, Berlin (1988), 585-654.
- Jenkins, J. J.; Strange, W. und Miranda, S. (1994): Vowel identification in mixed-speaker silent-center syllables. *Journal of the Acoustical Society of America* 95, 1030-1043.
- Jensen J. T. (2000): Against ambisyllabicity. *Phonology* 17, 187-235.
- Jerrentrup A. und Thürmann, E. (1974): Auditive Dimensionen der Vokalklangwahrnehmung. Ein Vergleich des Urteilsverhaltens von phonetischen Laien und Experten. *Phonetica* 30, 175-193.
- Jespersen, O. (1913, 1919): *Lehrbuch der Phonetik*. 2. Aufl., Teubner, Leipzig.
- Jessen, M. (1993): Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory* 8, 1-27.
- Jessen, M.; Marasek, K.; Schneider, K. und Clahßen, K. (1995): Acoustic correlates of word stress and the tense/lax Opposition in the vowel system of German. *Proc. 13<sup>th</sup> International Congress of Phonetic Sciences*, Stockholm, 4, 428-131.
- Joachims, T. (1999): Making large-scale SVM learning practical. In B. Schölkopf and C. Burges and A. Smola (Hg.), *Advances in Kernel Methods – Support Vector Learning*, MIT-Press, Cambridge.
- Johnson, K. (1989): Higher formant normalisation results from auditory integration of F2 und F3. *Perception and Psychophysics* 46, 174-180.
- Johnson, K.; Flemming, E. und Wright, R. (1993): The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69, 505-528.
- Jones, D. (1918, 1962): *An outline of English phonetics*. 9. Aufl., W.Hefer & Sons LTD., Cambridge.
- Jørgensen, H. P. (1969): Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantenfrequenzen. *Phonetica* 19, 217-245.
- Jurafsky, D.; Bell, A.; Fosler-Lussier, E. Girand, C. und Raymond, W. (1998): Reduction of English function words in Switchboard. *Proc. 5<sup>th</sup> International Conference on Spoken Language Processing*, Sydney Australia.
- Kewley-Port, D. (1992): Effects of levels of stimulus uncertainty and consonantal context on formant frequency discrimination. *Journal of the Acoustical Society of America* 91, 2338.
- Kewley-Port, D. und Watson, C. (1994): Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America* 95, 485-496.
- Kewley-Port, D. und Zheng, Y. (1998): Vowel formant discrimination in ordinary listening conditions. *Indiana University Cognitive Science, Technical Report #225*, URL://www.psych.indiana.edu/cogsci/techreps/225.html.
- Kewley-Port, D.; Li, X.; Zheng, Y. und Neel, A. T. (1996): Fundamental frequency effects on Thresholds for vowel formant discrimination. *Journal of the Acoustical Society of America* 100, 2462-2470.

- Kiparsky, P. (1988): Phonological change. In F. J. Newmeyer (Hg.), *Linguistics: The Cambridge Survey, Linguistic Theory Foundations*, Vol. 1, 363-415, Cambridge University Press, Cambridge.
- Kirchner, R. (in Vorbereitung): Phonological Contrast and Articulatory Effort. In L. Lombardi (Hg.), *Segmental Processes in Optimality Theory*. Cambridge University Press. URL: [//www.ualberta.ca/~kirchner/cv.html](http://www.ualberta.ca/~kirchner/cv.html)
- Klatt, D. H. (1976): Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59, 1208-1221.
- Klein, W. ; Plomp, R. und Pols, L. C. W. (1970): Vowel Spectra, vowel spaces, and vowel identification. *Journal of the Acoustical Society of America* 48, 999-1009.
- Kohler, K. J. (1974): Koartikulation und Steuerung im Deutschen. In: U. Engel und P. Grebe (Hg.), *Sprachsystem und Sprachgebrauch. Festschrift für H. Moser zum 65. Geburtstag*. Pädagogischer Verlag Schwann, Düsseldorf, Bd. 1, 172-192.
- Kohler, K. J. (1976): Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatogaphische Untersuchung. *Phonetica* 33, 1-30.
- Kohler, K. J. (1979): Kommunikative Aspekte satzphonetischer Prozesse im Deutschen. In H. Vater (Hg.), *Phonologische Probleme des Deutschen. Studien zur deutschen Grammatik*, Narr, Tübingen, Bd. 10, 13-40.
- Kohler, K. J. (1990): Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle und A. Marchal (Hg.), *Speech Production and Speech Modelling*. Kluwer Academic Publishers, Dordrecht, 69-92.
- Kohler, K. J. (1991): A model of German intonation. In *Studies in German intonation. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)* 25, 295-368.
- Kohler, K. J. (1995): Articulatory reduction in different speaking styles. *Proc. 13<sup>th</sup> International Congress of Phonetic Sciences*, Stockholm, 2, 12-19.
- Kohler, K. J. (1995b): *Einführung in die Phonetik des Deutschen. Grundlagen der Germanistik*. Erich Schmidt Verlag, 2.Aufl., Berlin.
- Koopmans-van Beinum, F. J. (1976): Vowel reduction in Dutch; extended version. *Proc. Institute of Phonetic Sciences University of Amsterdam* 4, 47-79.
- Koopmans-van Beinum, F. J. (1980): *Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions*. Dissertation University of Amsterdam.
- Kraft, V. (1993): Auditory detection of discontinuities in synthesis-by-concatenation. *Proc. Eurospeech*, Berlin, 929-932.
- Kröger, B. J. (1993): A gestural production model and its application to reduction in German. *Phonetica* 50, 213-233.
- Kröger, B. J. (1998): *Ein phonetisches Modell der Sprachproduktion*. Niemeyer, Tübingen.
- Kroos, C.; Hoole, P.; Kühnert, B. und Tillmann, H. G. (1997): Phonetic evidence for the phonological status of the tense-lax distinction in German. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 35, 17-25.
- Kuehn, D. P. und Moll, K. L. (1976): A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4, 303-320.



- Kuhl, P. K. (1992): Infants' perception and representation of speech: Development of a new theory. *Proc. 2<sup>nd</sup> International Conference on Spoken Language Processing*, Banff, 1, 449-456.
- Laan, G. P. M. (1997): The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* 22, 43-65.
- Ladd, D. R., Verhoeven, J. und Jacobs, K. (1994): Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *Journal of Phonetics* 22, 87-99.
- Ladefoged, P. (1964): *A phonetic study of West African languages*. West African Language Monograph Series 1. Cambridge University Press, London.
- Ladefoged, P. (1967): *Three Areas of Experimental Phonetics*. University Press, Oxford.
- Ladefoged, P. (1980): What are linguistic sounds made of? *Language* 56, 485-514.
- Ladefoged, P. und Maddieson, I. (1996): *The sounds of the World's Languages*. Cambridge, Mass., Blackwell.
- Lieberman, A. M. und Mattingly, I. G. (1985): The Motor Theory of speech perception revised. *Cognition* 21, 1-36.
- Lieberman, A. M.; Cooper, F. S.; Harris, K. S. und MacNeilage, P. F. (1963): A motor theory of speech perception. *Proc. Speech Communication Seminar*, 2, Stockholm.
- Lieberman, A. M.; Cooper, F. S.; Shankweiler, D. P. und Studdert-Kennedy, M. (1967): Perception of speech code. *Psychological Review* 74, 431-461.
- Lieberman, A. M.; Harris, K. S.; Hoffman, H. S. und Griffith, B. C. (1957): The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358-368.
- Lienert, G. A. and Raats, U. (1994): *Testaufbau und Testanalyse*. 5. Aufl., Psychologie Verlags Union, Weinheim.
- Lindau, M. (1977): Vowel features. *UCLA Working Papers Phonetics* 38, 49-81.
- Lindau, M. (1979): The feature expanded. *Journal of Phonetics* 7, 163-176.
- Lindau, M.; Jacobsen, L. und Ladefoged, P. (1972): The feature advanced tongue root. *UCLA Working Papers Phonetics* 22, 76-94.
- Lindblom, B. (1963): Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.
- Lindblom, B. (1983): Economy of speech gestures. In P. F. MacNeilage (Hrg.), *The production of speech*. Springer Verlag, Heidelberg, 217-245.
- Lindblom, B. (1990): Explaining phonetic variation: A sketch of the H&H Theory. In: W. J. Hardcastle und A. Marchal (Hg.), *Speech Production and Speech Modelling*. Kluwer, Dordrecht, 403-439.
- Lindblom, B. und Davis, J. H. (1998): Calculating and measuring the energy costs of speech movements. *Fonetik* 98, 11<sup>th</sup> Swedish Conference, 32-35, URL: [http://www.ling.su.se/staff/lindblom/fon98\\_BL\\_JD.html](http://www.ling.su.se/staff/lindblom/fon98_BL_JD.html).
- Lindblom, B. und Studdert-Kennedy, M. (1967): On the rôle of formant transitions in vowel recognition. *Journal of the Acoustical Society of America* 42, 830-843.
- Lindblom, B.; Brownlee, B.D.; Moon, S.-J. (1992): Speech transforms. *Speech Communication* 11, 357-368.

- Lindner, G. (1976): Urteilsveränderung bei Vokalkürzung. *Z. für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 29, 407-416.
- Lindner, G. (1977): *Hören und Verstehen*. Akademie-Verlag, Berlin.
- Lisker, L. (1989): On the interpretation of vowel "quality": The Dimension of Rounding. *Journal of the International Phonetic Association* 19, 24-30.
- Lively, S. E. (1993): An examination of the "perceptual magnet" effect. *Journal of the Acoustical Society of America* 93, 2423.
- Lively, S. E. und Pisoni, D. B. (1997): On prototype and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 23, 1665-1679.
- Lotto, A. J.; Kluender, K. R. und Holt, L. L. (1998): Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America* 103, 3648-3655.
- Macmillan, N.; Goldberg, R. F. und Braida, L. B. (1988): Resolution in speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America* 84, 1262-1278.
- MacNeilage, P. F. und Sholes, G. N. (1964): An electromyographic study of the tongue during vowel production. *Journal of Speech and Hearing Research* 7, 209-232.
- Makhoul, J. und Cosell, L. (1976): An LPC vocoder with linear predictive spectral warping. *Proc. IEEE International Conference on Acoustic Speech Signal Processing*, 456-469.
- Massaro, D. W. (1998): Categorical perception: Important phenomenon or lasting myth? *Proc. 5<sup>th</sup> International Conference on Spoken Language Processing*, Sydney Australia, [CD-ROM], Paper 970.
- Maurer, D. (1994): *Über den Vokal. Kritik der akustischen Theorie der stimmhaften Sprachlaute*. Bd. 1, Hatung-Gorre, Konstanz.
- Meinhold, G. (1973): *Deutsche Standardausssprache. Lautschwächung und Formstufen*. Friedrich-Schiller-Universität, Jena.
- Menzerath, P. und de Lacerda, A. (1933): *Koartikulation, Steuerung und Lautabgrenzung*. Dümmler, Berlin.
- Meyer, E. A. (1913): Das Problem der Vokalspannung. *Die neueren Sprachen* 21, 65-86 und 145-171.
- Miller, G. A. (1956): The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81-97.
- Miller, R. L. (1953): Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America* 25, 114-121.
- Mixdorff, H. und Widera, C. (2001): Perceived Prominence in Terms of a Linguistically Motivated Quantitative Intonation Model. *Proc. Eurospeech*, Aalborg, 403-406.
- Möbius, B. (1993): *Ein quantitatives Modell der deutschen Intonation*. Niemeyer, Tübingen.
- Mooshammer, C. (1998): Experimentalphonetische Untersuchungen zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 36, 3-192.



- Mooshammer, C.; Fuchs, S. und Fischer, D. (1999): Effects of stress and tenseness on the production of CVC Syllables in German. *Proc. 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, 409-412.
- Moulton, W. G. (1962): *The sounds of English and German*. University Press, Chicago.
- Nearey, T. M. (1989): Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85, 2088-2113.
- Nelson, W. L. (1983): Physical principles for economies of skilled movements. *Biological Cybernetics* 46, 135-147.
- Nelson, W. L.; Perkell, J. S. und Westbury, J. R. (1984): Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. *Journal of the Acoustical Society of America* 75, 945-951.
- Neppert, J. und Pétursson, M. (1986): *Elemente einer akustischen Phonetik*. Helmut Buske Verlag, Hamburg.
- Nord, L. (1987): Vowel reduction in Swedish. In: O. Engstrand (Hg.), *Papers from the Swedish Phonetics Conference, Department of Linguistics of the University of Uppsala*, 16-21.
- Nosofsky, R. M. (1986): Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115, 39-57.
- O'Shaughnessy, D. (1987): *Speech Communication – Human and Machine*. Addison-Wesley, Reading, Massachusetts.
- Ohala, J. (1973): Explanations for the intrinsic pitch of vowels. *Monthly International Memorandum (Berkeley)*, 9-26.
- Öhman, S. E. G. (1966): Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39, 151-168.
- Palmeri, T. J. (1999): Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *Quarterly Journal of Experimental Psychology* 54, 197-235.
- Pätzold, M. und Simpson, A. P. (1997): Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)* 32, 215-243.
- Peterson, G. E. und Barney, H. L. (1952): Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175-184.
- Pfützing, H. R. (1996): Two approaches to speech rate estimation. *Proc. Speech Science and Technology (SST)*, Adelaide, 421-426.
- Pisoni, D. B. (1975): Auditory short-term memory and vowel perception. *Memory and Cognition* 3, 7-18.
- Pols, L. C. W. und van Son, R. J. J. H. (1993): Acoustic and perception of dynamic vowel segments. *Speech Communication* 13, 135-147.
- Pols, L. C. W., Boxelaar, G. W. und Koopmans-van Beinum, F. J. (1984): Study of the role of formant transition in vowel recognition using the matching paradigm. *Proc. Institute of Acoustics Amsterdam* 6, 371-378.
- Pols, L. C. W.; Tromp, H. R. C. und Plomp, R. (1973): Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America* 53, 1093-1101.

- Pols, L. C. W.; van der Kamp, L. J. T. and Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America* 46, 458-467.
- Portele, T. (1996): *Ein phonetisch-akustisch motiviertes Inventar zur Sprachsynthese deutscher Äußerungen*. Niemeyer, Tübingen.
- Portele, T. (1997): Reduktionen in der einheitenbasierten Sprachsynthese. *Fortschritte der Akustik – DAGA*, Kiel, 386-387.
- Portele, T. (1999): *Perceived prominence and segmental parameters*. Unveröffentlichtes Manuskript.
- Potter, R. K. und Steinberg, J. C. (1950): Toward the specification of speech. *Journal of the Acoustical Society of America* 22, 807-820.
- Ramers, K. H. (1988): *Vokalqualität und –qualität im Deutschen*. Max Niemeyer Verlag, Tübingen.
- Raphael, L. und Berti, F. (1975): Tongue musculature and the feature of tension in English vowels. *Phonetica* 32, 61-73.
- Rietveld, A. C. M. und Koopmans-van Beinum, F. J. (1987): Vowel reduction and stress. *Speech Communication* 6, 217-229.
- Rietveld, A. C. M. (1975): Untersuchungen zur Vokaldauer im Deutschen. *Phonetica* 31, 248-258.
- Rosch, E. (1975): Cognitive reference points. *Cognitive Psychology* 7, 532-547.
- Schindler, F. (1974): *Beiträge zur deutschen Hochlautung*. Helmut Buske Verlag, Hamburg.
- Schölkopf, B.; Burges, C. und Vapnik, V. (1996): Incorporating invariances in Support Vector Machines. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen und B. Sendhoff (Hg.), *Artificial Neural Networks – ICANN’ 96*. Springer Lecture Notes in Computer Science, Vol. 1112, Berlin, 47-52.
- Schuhmacher, W. W. (1972): Beitrag zur Bestimmung des physiologischen Korrelates des deutschen Vokalgegensatz. *Linguistics* 90, 35-78.
- Sendlmeier, W. F. (1981): Der Einfluß von Qualität und Quantität auf die Perzeption betonter Vokale des Deutschen. *Phonetica* 38, 291-308.
- Siebs Deutsche Aussprache. Reine und gemäßigte Hochlautung mit Aussprachewörterbuch*. (1969), 19. Aufl., Walter de Gruyter, Berlin.
- Sievers, E. (1893): *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. 4. Aufl., Breitkopf und Härtel, Leipzig.
- Sievers, E. (1901): *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. 5. Aufl., Breitkopf und Härtel, Leipzig.
- Smola, A. J. (1999): Lernen mit Kernen. In G. Hotz et al. (Hg.), *Ausgezeichnete Informatikdissertationen*, S. 184-195, Teubner, Stuttgart, 1999.
- Spencer-Smith, J. und Goldstone, R. L.(17.10.2001): *The Dynamics of Similarity*. URL: <http://www.indiana.edu/~jbsteach/DynamicSim.html>
- Steriopolo, E. (1994): Untersuchungen zu den Vokalrealisierungen im Deutschen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 32, 263-293.

- Stevens, K. N. (1972): The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David und P. B. Denes (Hg.), *Human communication: A unified view*. McGraw-Hill, New York, S. 51-66.
- Stevens, K. N. (1989): On the quantal nature of speech. *Journal of Phonetics* 17, 3-45.
- Stevens, K. N. und House, A. S. (1963): Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research* 6, 111-128.
- Stöber, K. (1996): unveröffentlichtes Computerprogramm zur Bestimmung der Anregungszeitpunkte von Sprachsignalen.
- Stöber, K. (1997a): unveröffentlichtes Computerprogramm zur Approximation des Formantverlaufs.
- Stöber, K. (1997b): *Methoden der automatischen Lautsegmentierung unter Einsatz selbstorganisierender Karten und dynamic time warping*. Diplomarbeit im Fach Informatik, Universität Bonn.
- Stöber, K.; Portele, T.; Wagner, P. und Hess, W. (1999): Synthesis by word concatenation. *Proc. Eurospeech*, Budapest, 619-622.
- Stöber, K.; Wagner, P.; Helbig, J.; Köster, S.; Stall, D.; Thomae, M.; Blauert, J.; Hess, W.; Hoffmann, R. und Mangold, H. (2000): Speech synthesis using multilevel selection and concatenation of units from large speech corpora. In W. Wahlster, *Foundations of Speech-to-Speech Translation*. Springer, Berlin, 519-534.
- Strange, W. (1989): Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America* 85, 2135-2153.
- Strange, W. und Bohn, O.-S. (1998): Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *Journal of the Acoustical Society of America* 104, 488-504.
- Strange, W.; Jenkins, J. J. und Johnson, T. L. (1983): Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America* 74, 695-705.
- Studdert-Kennedy, M.; Liberman, A. M.; Harris, K. S. und Cooper, F. S. (1970): Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77, 234-239.
- Sussman, J. E. und Gekas, B. (1997): Phonetic category Structure of [ɪ]: Extent, best exemplars, and organisation. *Journal of Speech, Language, and Hearing Research* 40, 1406-1424.
- Sussman, J. E. und Lauckner-Morano, V. J. (1995): Further tests of the „perceptual magnet effect“ in the perception of [i:]: Identification and change/no-Change discrimination. *Journal of the Acoustical Society of America* 97, 539-552.
- Sweet, H. (1892): *A primer of phonetics*. 3. Aufl., 1906 Clarendon Press, Oxford.
- Syrdal, A. K. und Gopal, H. S. (1986): A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79, 1086-1100.
- Terken, J. (1991): Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89, 1768-1776.
- Ternes, E. (1987): *Einführung in die Phonologie*. Wissenschaftliche Buchgesellschaft, Darmstadt.
- Tiffany, W. R. (1959): Non-random sources of variation in vowel quality. *Journal of Speech and Hearing Research* 2, 305-317.

- Tokuma, S. (1997): Quality of a vowel with formant undershoot: A preliminary perceptual study. *Proc. Eurospeech*, Rhodes, 2163-2166.
- Tokuma, S. (1999): Perceptual F1-F2 interaction in quality evaluation of vowels showing formant undershoot. *Proc. 14<sup>th</sup> International Congress of Phonetic Sciences*, San Francisco, 195-198.
- Tokuma, S. (2000): Quality perception of vowels with simulated /CVC/ formant trajectories. *Speech Communication* 32, 251-265.
- Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* 69, 1465-1475.
- Traunmüller, H. (1984): Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication* 3, 49-61.
- Trubetzkoy, N. (1939): *Grundzüge der Phonologie*. 3. Aufl., 1962. Vandenhoeck & Ruprecht, Göttingen.
- Umeda, N. (1975): Vowel duration in American English. *Journal of the Acoustical Society of America* 58, 434-445.
- Ungeheuer, G. (1958): Die Eigenwerttheorie der Formanten und das System der Vokale. *Z. für Phonetik und allgemeine Sprachwissenschaft*, 11, 35-48. In G. Ungeheuer (Hg.): *Materialien zur Phonetik des Deutschen*. IKP-Forschungsberichte 61, Buske, 1977, 29-42.
- Ungeheuer, G. (1962): *Elemente einer akustischen Theorie der Vokalartikulation*. Springer, Berlin.
- Ungeheuer, G. (1962b): *Phonetische Aspekte beim Sprachverstehen*. Habilitationsschrift. Universität Bonn.
- Ungeheuer, G. (1964): Extensional-paradigmatische Bestimmung auditiver Qualitäten phonetischer Signale. *Proc. 5th International Congress of Phonetic Sciences*, Münster, 556-560. In G. Ungeheuer (Hg.): *Sprache und Signal*. IKP-Forschungsberichte 40, Buske, Hamburg 1977, 79-83.
- Ungeheuer, G. (1977): *Materialien zur Phonetik des Deutschen*. IKP-Forschungsberichte 61, Buske, Hamburg.
- Ungeheuer, G. (1977b): *Sprache und Signal*. IKP-Forschungsberichte 40, Buske, Hamburg.
- Van Bergem, D. R. (1990): In defence of a probabilistic view on human word recognition, *Proc. Institute of Phonetic Sciences University of Amsterdam* 14, 53-66.
- Van Bergem, D. R. (1993): Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication* 12, 1-23.
- Van Bergem, D. R. (1994): A model of coarticulatory effects on the schwa. *Speech Communication* 14, 143-162.
- Van Bergem, D. R. (1994b): Reflections on aspects of vowel reduction. *Proc. Institute of Phonetic Sciences University of Amsterdam* 18, 95-110.
- Van Bergem, D. R. (1995): *Acoustic and lexical vowel reduction*. Studies in Language and Language Use, Bd. 16, Dissertation, IFOTT, University of Amsterdam.
- Van Bergem, D. R. (1995b): Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress. *Speech Communication* 16, 329-358.

- Van den Berg, J. (1958): Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1, 227-244.
- Van den Heuvel, H.; Rietveld, T. und Cranen, B. (1994): Methodological aspects of segment- and speaker-related variability. A study of segmental durations in Dutch. *Journal of Phonetics* 22, 389-406.
- Van den Heuvel, H.; Rietveld, T. und Cranen, B. (1996): Speaker variability in the coarticulation of /a,i,u/. *Speech Communication* 18, 113-130.
- Van Hessen, A. J. und Schouten, M. E. H. (1999): Categorical perception as a function of stimulus quality. *Phonetica* 56, 56-72.
- Van Santen, J. P. H. (1992): Contextual effects on vowel duration. *Speech Communication* 11, 513-546.
- van Son, R. J. J. H und Pols, L. C. W. (1990): Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 88, 1683-1693.
- Van Son, R. J. J. H. (1993): *Spectro-temporal features of vowel segments*. Studies in Language and Language Use, Bd. 3, ICG-Printing, Dordrecht.
- Van Son, R. J. J. H. und Pols, L. C. W. (1993): Vowel identification as influenced by vowel duration and formant track shape. *Proc. Eurospeech*, Berlin, 285-288.
- Van Son, R. J. J. H. und Pols, L. C. W. (1999): Effects of stress and lexical structure on speech efficiency. *Proc. Eurospeech*, Budapest, 439-442.
- Van Valin Jr., R. D. (1976): Perceived distance between vowel stimuli. *Journal of Phonetics* 4, 51-58.
- Vapnik, V. N. (1979): *Estimation of Dependences Based on Empirical Data*. Nauka, Moscow. [in Russisch]. Englische Übersetzung: *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.
- Vieregge, W. H. (1989): Phonetische Transkription. Theorie und Praxis der Symphonetik. In J. Göschel (Hg.), *Z. für Dialektologie und Linguistik Beiheft 60*, Franz Steiner Verlag Wiesbaden GmbH, Stuttgart.
- Wängler, H.-H. und Weiss, R. (1975): Experimental approach to the study of vowel perception in German. *Phonetica* 32, 180-199.
- Weiss, R. (1974): Relationship of vowel length and quality in the perception of German vowels. *Linguistics* 123, 59-70.
- Wells, J. C. (1996): *SAMPA – computer readable phonetic alphabet*. URL: [//www.phon.ucl.ac.uk/home/sampa/german.htm](http://www.phon.ucl.ac.uk/home/sampa/german.htm).
- Werner, O. (1972): *Phonemik des Deutschen*. Metzler, Stuttgart.
- Whalen, D. H. (1990): Coarticulation is largely planned. *Journal of Phonetics* 18, 3-35.
- Whalen, D. H.; Gick, B.; Kumada, M.; Honda, K. (1998): Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics* 27, 125-142.
- Widera, C. and Portele, T. (1999): Levels of reduction for German tense vowel. *Proc. Eurospeech*, Budapest, 1695-1698.
- Widera, C. und Hess, W. (in Vorbereitung): *Abschlussbericht des DFG-Projektes Prosodiebeurteilung und Reduktionsanalyse* (He 1019/9-2).



- Wiese, R. (1988): *Silbische und lexikalische Phonologie: Studie zum Chinesischen und Deutschen*. Niemeyer, Tübingen.
- Wiese, R. (1996): *The phonology of German*. Clarendon Press, Oxford.
- Winckel, F. (1968): Acoustical foundations of phonetics. In: B. Malmberg (Hg.), *Manual of Phonetics*. North-Holland Publishing Company, Amsterdam, 17-44.
- Wood, S. (1975): Tense and lax vowels – degree of constriction or pharyngeal volume? *Working Papers Lund University Department of Linguistics 11*, 110-134.
- Wood, S. (1982): X-ray and model studies of vowel articulation. *Working Papers Lund University Department of Linguistics 23*.
- Wrede, B.; Fink, G. A. und Sagerer, G. (2000): Influence of duration on static and dynamic properties of German vowels in spontaneous speech. *Proc. 6<sup>th</sup> International Conference on Speech and Language Processing*, Beijing. [CD]
- Zell, A. (1994): *Simulation neuronaler Netze*. Addison-Wesley, Bonn.
- Zwicker, Eberhard (1982): *Psychoakustik. Hochschultext*. Springer, Berlin.

## ANHANG A

Laut		F1	F2	F3	LAUTDUR	F0	E1NORM	E2NORM	E3NORM	E4NORM
/ə/	n	539,00	539,00	539,00	539,00	539,00	537,00	537,00	537,00	537,00
	$\bar{x}$	402,00	1718,08	2848,71	58,42	244,11	6,15	,83	,28	,12
	s	164,21	300,72	395,74	37,14	162,36	,75	,42	,15	,11
/o:/	n	18,00	18,00	18,00	18,00	18,00	18,00	18,00	18,00	18,00
	$\bar{x}$	462,33	1563,44	2593,67	94,83	289,06	6,36	,89	,23	,08
	s	58,98	156,51	256,29	36,94	166,80	,39	,24	,07	,03
/œ/	n	23,00	23,00	23,00	23,00	23,00	23,00	23,00	23,00	23,00
	$\bar{x}$	445,22	1725,70	2730,43	86,13	249,52	5,74	1,26	,31	,10
	s	112,24	188,35	201,26	35,00	59,71	,54	,38	,10	,02
/a/	n	333,00	333,00	333,00	333,00	333,00	333,00	333,00	333,00	333,00
	$\bar{x}$	754,05	1593,63	2890,39	74,05	235,73	5,16	1,87	,28	,10
	s	271,21	179,76	246,15	23,81	60,89	,68	,66	,08	,04
/a:z/	n	180,00	180,00	180,00	180,00	180,00	180,00	180,00	180,00	180,00
	$\bar{x}$	842,46	1464,43	2722,29	119,17	223,96	5,05	2,08	,26	,09
	s	259,34	134,24	191,37	50,77	92,81	,62	,63	,07	,03
/ɛ/	n	228,00	228,00	228,00	228,00	228,00	228,00	228,00	228,00	228,00
	$\bar{x}$	535,04	1935,34	2893,55	67,76	269,98	5,61	,90	,43	,16
	s	190,75	219,07	167,68	27,67	167,63	,72	,46	,21	,11
/ɛ:z/	n	203,00	203,00	203,00	203,00	203,00	202,00	202,00	202,00	202,00
	$\bar{x}$	455,35	2149,37	2917,49	76,64	249,58	5,98	,52	,46	,15
	s	91,52	259,82	190,20	38,44	82,33	,65	,30	,18	,09
/ɛ:/	n	20,00	20,00	20,00	20,00	20,00	20,00	20,00	20,00	20,00
	$\bar{x}$	498,45	2055,50	2850,80	105,55	232,32	6,02	,54	,47	,13
	s	97,88	315,63	266,37	39,26	35,43	,54	,24	,18	,02
/ɪ/	n	445,00	445,00	445,00	445,00	445,00	445,00	445,00	445,00	445,00
	$\bar{x}$	413,43	2100,84	2982,09	52,22	240,58	5,65	,49	,45	,24
	s	194,00	280,35	264,65	19,94	103,01	1,00	,28	,19	,23
/i:/	n	252,00	252,00	252,00	252,00	252,00	252,00	252,00	252,00	252,00
	$\bar{x}$	385,91	2260,86	2914,52	71,25	245,61	5,75	,36	,55	,20
	s	86,31	271,99	244,73	39,31	63,72	,58	,21	,19	,10
/ɔ/	n	111,00	111,00	111,00	111,00	111,00	111,00	111,00	111,00	111,00
	$\bar{x}$	554,96	1249,05	2873,97	71,13	249,09	6,05	1,28	,19	,07
	s	207,84	242,66	389,15	24,70	94,41	,51	,43	,06	,03
/o:/	n	119,00	119,00	119,00	119,00	119,00	119,00	119,00	119,00	119,00
	$\bar{x}$	465,04	1081,26	2773,10	97,83	253,61	6,85	,65	,14	,05
	s	74,41	309,06	399,87	47,96	83,82	,54	,38	,08	,03
/u/	n	188,00	188,00	188,00	188,00	188,00	188,00	188,00	188,00	188,00
	$\bar{x}$	431,02	1404,74	2850,05	57,30	247,38	6,31	,93	,21	,08
	s	132,00	373,18	524,73	22,57	102,62	,65	,40	,11	,08
/u:/	n	88,00	88,00	88,00	88,00	88,00	88,00	88,00	88,00	88,00
	$\bar{x}$	396,06	1139,02	2750,41	82,59	261,12	6,69	,68	,17	,07
	s	128,73	389,24	364,37	45,01	93,32	,55	,32	,09	,06
/y/	n	38,00	38,00	38,00	38,00	38,00	38,00	38,00	38,00	38,00
	$\bar{x}$	427,76	1629,39	2695,89	56,89	258,83	6,25	,87	,27	,09
	s	103,45	200,35	359,11	18,71	66,35	,48	,29	,11	,03
/y:/	n	29,00	29,00	29,00	29,00	29,00	29,00	29,00	29,00	29,00
	$\bar{x}$	404,69	1670,41	2639,10	80,83	267,80	6,30	,78	,29	,09
	s	80,05	196,26	241,73	41,34	52,14	,44	,23	,12	,04
Insgesamt	n	2814,00	2814,00	2814,00	2814,00	2814,00	2811,00	2811,00	2811,00	2811,00
	$\bar{x}$	499,08	1760,29	2869,14	70,25	245,93	5,85	,93	,34	,14
	s	229,88	432,68	324,95	37,86	113,35	,86	,68	,19	,13

**Tabelle A-1: Anzahl (n), Mittelwert ( $\bar{x}$ ) und Standardabweichung (s) für die akustischen Parameter der Monophthonge von SPR1.**



Laut		F1	F2	F3	LAUTDUR	F0	E1NORM	E2NORM	E3NORM	E4NORM
/ə/	n	517,00	517,00	517,00	517,00	517,00	517,00	517,00	517,00	517,00
	$\bar{x}$	387,70	1592,91	2599,05	55,84	226,20	6,13	,85	,29	,11
	s	108,89	290,66	219,17	32,20	140,57	,58	,39	,14	,07
/ø:/	n	20,00	20,00	20,00	20,00	20,00	20,00	20,00	20,00	20,00
	$\bar{x}$	436,70	1493,50	2451,00	80,95	224,48	6,25	,90	,21	,11
	s	83,80	181,57	252,84	42,17	76,24	,32	,18	,05	,04
/œ/	n	27,00	27,00	27,00	27,00	27,00	27,00	27,00	27,00	27,00
	$\bar{x}$	454,59	1625,96	2567,89	65,81	270,22	5,66	1,42	,26	,10
	s	61,01	233,18	358,03	13,01	123,41	,43	,40	,08	,03
/a/	n	372,00	372,00	372,00	372,00	372,00	372,00	372,00	372,00	372,00
	$\bar{x}$	566,47	1465,63	2585,68	65,11	239,88	5,04	1,95	,27	,12
	s	130,11	162,71	129,37	23,20	138,12	,71	,62	,09	,04
/a:/	n	203,00	203,00	203,00	203,00	203,00	203,00	203,00	203,00	203,00
	$\bar{x}$	642,27	1388,94	2503,67	108,98	179,13	4,85	2,18	,27	,11
	s	103,17	138,97	150,66	51,24	93,48	,61	,55	,08	,03
/ɛ/	n	261,00	261,00	261,00	261,00	261,00	261,00	261,00	261,00	261,00
	$\bar{x}$	483,79	1747,50	2601,09	59,91	304,47	5,49	1,19	,42	,12
	s	100,86	192,42	147,64	23,75	236,01	,68	,41	,20	,09
/ɛ:/	n	205,00	205,00	205,00	205,00	205,00	205,00	205,00	205,00	205,00
	$\bar{x}$	434,95	1943,80	2673,09	73,66	265,59	5,71	,76	,54	,12
	s	82,24	205,45	158,14	36,54	174,73	,45	,44	,19	,04
/ɛ:/	n	22,00	22,00	22,00	22,00	22,00	22,00	22,00	22,00	22,00
	$\bar{x}$	508,64	1909,45	2609,77	106,86	262,75	5,56	1,04	,51	,09
	s	44,23	121,10	85,70	41,28	122,52	,42	,29	,15	,02
/ɪ/	n	480,00	480,00	480,00	480,00	480,00	479,00	479,00	479,00	479,00
	$\bar{x}$	353,66	1867,97	2737,96	49,28	226,58	5,74	,67	,42	,19
	s	90,18	269,64	210,57	17,74	119,04	,68	,37	,17	,16
/i:/	n	270,00	270,00	270,00	270,00	270,00	270,00	270,00	270,00	270,00
	$\bar{x}$	337,14	2052,10	2788,12	69,20	215,98	5,75	,42	,53	,19
	s	69,94	196,16	285,73	41,17	83,64	,52	,21	,16	,11
/ɔ/	n	125,00	125,00	125,00	125,00	125,00	125,00	125,00	125,00	125,00
	$\bar{x}$	509,42	1212,35	2638,83	66,14	272,59	5,90	1,43	,17	,08
	s	87,72	292,68	335,81	19,34	127,00	,46	,40	,05	,03
/o:/	n	131,00	131,00	131,00	131,00	131,00	131,00	131,00	131,00	131,00
	$\bar{x}$	421,98	1080,53	2667,02	88,22	239,91	6,71	,71	,14	,08
	s	57,03	297,18	297,89	49,67	130,13	,55	,36	,06	,06
/u/	n	203,00	203,00	203,00	203,00	203,00	203,00	203,00	203,00	203,00
	$\bar{x}$	404,40	1182,71	2655,30	55,08	233,37	6,51	,85	,16	,08
	s	58,03	267,86	265,97	20,28	112,76	,50	,35	,06	,07
/u:/	n	96,00	96,00	96,00	96,00	96,00	95,00	95,00	95,00	95,00
	$\bar{x}$	369,95	1110,46	2626,73	83,24	247,83	6,74	,64	,15	,08
	s	110,32	395,62	317,33	42,79	217,09	,54	,34	,07	,04
/ʏ/	n	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00	40,00
	$\bar{x}$	392,23	1469,80	2575,73	54,78	245,59	6,26	,89	,24	,10
	s	53,84	246,13	391,02	12,28	106,80	,38	,27	,07	,04
/y:/	n	37,00	37,00	37,00	37,00	37,00	37,00	37,00	37,00	37,00
	$\bar{x}$	352,05	1702,73	2541,68	68,30	228,65	6,20	,77	,31	,10
	s	60,21	268,09	346,83	34,66	64,79	,40	,25	,12	,04
Insgesamt	n	3009,00	3009,00	3009,00	3009,00	3009,00	3007,00	3007,00	3007,00	3007,00
	$\bar{x}$	437,10	1606,29	2642,40	65,72	237,97	5,78	1,05	,33	,12
	s	131,43	378,37	239,37	35,49	145,38	,79	,68	,18	,10

**Tabelle A-2: Anzahl (n), Mittelwert ( $\bar{x}$ ) und Standardabweichung (s) für die akustischen Parameter der Monophthonge von SPR2.**

Laut		F1	F2	F3	LAUTDUR	F0	E1NORM	E2NORM	E3NORM	E4NORM
/ə/	n	409,00	409,00	409,00	409,00	409,00	407,00	407,00	407,00	407,00
	$\bar{x}$	409,46	1719,65	2866,16	63,31	239,27	6,67	,67	,21	,06
	s	133,32	376,29	261,36	42,93	133,28	,49	,38	,13	,04
/ø/	n	17,00	17,00	17,00	17,00	17,00	17,00	17,00	17,00	17,00
	$\bar{x}$	475,65	1632,59	2720,47	80,06	254,03	6,41	1,01	,21	,04
	s	104,58	273,47	162,48	26,31	66,66	,58	,47	,15	,01
/œ/	n	25,00	25,00	25,00	25,00	25,00	25,00	25,00	25,00	25,00
	$\bar{x}$	481,04	1797,72	2780,48	71,16	273,04	6,21	1,04	,28	,05
	s	136,50	349,74	210,86	24,70	60,18	,39	,33	,10	,01
/a/	n	356,00	356,00	356,00	356,00	356,00	355,00	355,00	355,00	355,00
	$\bar{x}$	725,22	1656,14	2914,77	73,70	239,00	5,16	2,14	,22	,06
	s	229,40	209,03	172,05	33,99	66,69	,80	,77	,10	,04
/a:/	n	188,00	188,00	188,00	188,00	188,00	188,00	188,00	188,00	188,00
	$\bar{x}$	892,07	1512,26	2829,33	122,56	233,47	4,94	2,47	,19	,05
	s	212,89	165,27	157,42	58,81	90,81	,68	,65	,06	,02
/ɛ/	n	248,00	248,00	248,00	248,00	248,00	246,00	246,00	246,00	246,00
	$\bar{x}$	517,23	2065,04	2940,96	63,78	257,91	6,10	,76	,43	,07
	s	148,30	318,83	239,89	27,00	110,70	,61	,42	,22	,04
/ɛ:/	n	212,00	212,00	212,00	212,00	212,00	211,00	211,00	211,00	211,00
	$\bar{x}$	445,06	2199,78	2957,92	75,51	259,81	6,48	,42	,41	,07
	s	116,99	494,83	255,07	42,07	92,23	,65	,31	,19	,08
/ɛ:/	n	6,00	6,00	6,00	6,00	6,00	6,00	6,00	6,00	6,00
	$\bar{x}$	477,00	2331,83	2980,67	101,67	261,60	6,21	,61	,46	,06
	s	87,47	172,07	85,25	27,73	45,33	,36	,31	,12	,01
/ɪ/	n	447,00	447,00	447,00	447,00	447,00	443,00	443,00	443,00	443,00
	$\bar{x}$	368,29	2131,63	3020,20	50,53	226,53	6,47	,36	,36	,11
	s	104,05	532,32	286,77	24,61	94,11	,77	,25	,22	,12
/i:/	n	255,00	255,00	255,00	255,00	255,00	255,00	255,00	255,00	255,00
	$\bar{x}$	354,33	2293,58	3061,14	65,32	244,99	6,51	,25	,40	,11
	s	80,22	508,33	228,26	38,10	53,55	,50	,16	,16	,08
/ɔ/	n	116,00	116,00	116,00	116,00	116,00	116,00	116,00	116,00	116,00
	$\bar{x}$	550,29	1242,47	2904,67	69,27	253,96	6,25	1,35	,13	,04
	s	162,59	171,18	207,21	23,70	98,50	,40	,34	,05	,01
/o:/	n	111,00	111,00	111,00	111,00	111,00	110,00	110,00	110,00	110,00
	$\bar{x}$	403,45	1117,04	2759,95	86,93	247,33	7,10	,59	,09	,03
	s	71,81	310,68	207,68	51,88	81,90	,41	,34	,05	,01
/u/	n	192,00	192,00	192,00	192,00	192,00	191,00	191,00	191,00	191,00
	$\bar{x}$	374,13	1330,36	2821,57	53,89	230,93	6,91	,70	,11	,04
	s	77,36	301,73	312,89	24,80	66,86	,35	,27	,05	,02
/u:/	n	95,00	95,00	95,00	95,00	95,00	95,00	95,00	95,00	95,00
	$\bar{x}$	353,19	1103,18	2774,62	79,43	260,63	7,07	,56	,10	,04
	s	92,48	317,32	257,34	43,24	72,88	,43	,32	,06	,02
/ʏ/	n	38,00	38,00	38,00	38,00	38,00	38,00	38,00	38,00	38,00
	$\bar{x}$	350,21	1679,18	2661,08	53,29	236,82	6,76	,67	,21	,04
	s	69,98	268,65	132,56	16,22	61,77	,46	,25	,12	,02
/y:/	n	34,00	34,00	34,00	34,00	34,00	34,00	34,00	34,00	34,00
	$\bar{x}$	358,62	1669,00	2718,00	75,21	248,58	6,90	,51	,21	,04
	s	71,13	387,92	149,25	34,55	36,17	,44	,33	,09	,02
Insgesamt	n	2749,00	2749,00	2749,00	2749,00	2749,00	2737,00	2737,00	2737,00	2737,00
	$\bar{x}$	485,06	1793,02	2911,86	69,17	242,19	6,27	,91	,27	,07
	s	215,23	528,51	257,90	40,80	93,05	,88	,83	,19	,07

**Tabelle A-3: Anzahl (n), Mittelwert ( $\bar{x}$ ) und Standardabweichung (s) für die akustischen Parameter der Monophthone von SPR3.**

## ANHANG B

Prototypen für	Relevante Prototypen	F1 mel	F2 mel
/ø:/		567,29	1271,44
		566,32	1373,49
		636,19	1427,39
	RED1	478,22	1297,71
	RED2	659,75	1242,95
/a:/	RED3	585,64	1106,80
		557,52	1168,97
		583,72	1255,70
		1003,96	1234,37
		731,81	1055,60
		1017,74	1321,38
	RED1	1002,63	1232,21
	RED2	982,62	1403,55
/e:/	RED3	585,64	1318,39
		392,92	1104,99
	RED1	483,48	1695,80
	RED2	582,76	1611,57
	RED3	689,82	1500,72
	RED4	575,05	1469,25
/ɛ:/	RED5	430,90	1487,46
		619,59	1598,00
		433,09	903,13
		384,94	1535,87
	RED1	650,74	1680,32
/i:/	RED2	691,57	1528,02
	RED3	548,65	1542,45
		916,77	1637,47
		440,74	1051,18
	RED1	386,08	1737,91
	RED2	502,19	1692,94
	RED3	599,91	1572,36
/o:/	RED1	453,73	1566,36
	RED2	566,32	1473,61
		479,27	947,05
		588,50	1205,46
		895,15	1336,24
	RED1	639,85	1065,03
	RED2	468,70	659,75
/u:/	RED3	582,76	916,77
	RED4	438,56	1542,86
	RED5	446,17	804,28
		303,09	964,97
		422,08	1539,99
		570,21	1401,69
/y:/	RED1	885,66	1319,38
		616,80	1388,62
		399,71	1415,99
		504,25	1264,12
	RED1	361,82	1499,02
	RED2	432,00	1161,52
	RED3	631,60	1211,52
		514,50	1429,66

**Tabelle B-1: F1- und F2-Werte (mel-skaliert) der in den Perzeptionsexperimenten verwendeten Prototypen für die gespannten Vokale. In der zweiten Spalte sind die perzeptiv relevanten Prototypen der einzelnen Reduktionsstufen (RED1 nicht reduziert; die höchste Stufe entspricht einer starken Reduzierung) gekennzeichnet.**

Prototypen für	Relevante Prototypen	F1 mel	F2 mel	SPR
/ɜ/		474,00	1474,48	SPR2
		381,50	1441,37	SPR3
		591,37	1310,36	SPR1
/ʊ/		525,66	1074,38	SPR1
	RED1	555,55	1273,00	SPR3
	RED2	409,83	1327,35	SPR3
/ɔ/		519,58	1260,97	SPR3
		652,55	1097,72	SPR1
		1075,62	1317,89	SPR1
/ɪ/		1095,89	1563,95	SPR1
	RED1	398,58	1535,87	SPR2
	RED2	512,45	1528,02	SPR3
	RED3	548,65	1342,12	SPR1
/ɛ/		648,94	1331,31	SPR2
		572,15	1524,28	SPR1
/a/	RED1	733,49	1266,22	SPR2
	RED2	558,50	1265,17	SPR2
/œ/		577,95	1233,29	SPR1
		498,06	1562,75	SPR3
		598,01	1372,54	SPR2

**Tabelle B-2: F1- und F2-Werte (mel-skaliert) der in den Perzeptionsexperimenten verwendeten Prototypen verschiedener Sprecher (SPR) für die ungespannten Vokale. In der zweiten Spalte sind die perzeptiv relevanten Prototypen der einzelnen Reduktionsstufen (RED1 nicht reduziert; die höchste Stufe entspricht einer starken Reduzierung) gekennzeichnet.**

# ANHANG C

	RED1					RED2					RED3					RED4					RED5				
	n	z	Quartile			n	z	Quartile			n	z	Quartile			n	z	Quartile			n	z	Quartile		
			1.	2.	3.			1.	2.	3.			1.	2.	3.			1.	2.	3.					
/e:/	ZF1	31	-,832	-1,064	-,124	29	-,283	-,654	-,460	30	1,193	,993	1,729	17	-,197	-,523	,496	15	-,818	-1,501	-,244	15	-,883	-1,218	-,675
	ZF2	31	1,241	1,072	1,523	29	,694	,094	1,076	30	-,674	-1,126	-,320	17	-,721	-1,068	-,301	15	-,490	-1,218	-,675	15	-,490	-1,218	-,675
	ZF3	31	,844	,066	1,186	29	,187	-,660	,652	30	-,402	-1,027	,087	17	-,248	-,941	,420	15	-,549	-,188	,996	15	-,549	-,188	,996
	ZE1NORM	31	,782	-,036	1,062	29	,130	-,560	,545	30	-,523	-1,162	,078	17	-,009	-,723	,237	15	-,247	-,784	,834	15	-,247	-,784	,834
	ZE2NORM	31	-1,007	-1,185	-,737	29	-,536	-,845	-,296	30	1,299	,648	1,958	17	,716	-,132	1,373	15	-,631	-,868	-,174	15	-,631	-,868	-,174
	ZE3NORM	31	-,439	-,849	,914	29	,209	-,424	,809	30	-,371	-,979	,728	17	-,271	-,655	,013	15	-,270	-,614	,071	15	-,270	-,614	,071
	ZE4NORM	31	,081	-,175	,295	29	-,066	-,318	,218	30	-,333	-,855	,059	17	-,098	-,298	-,018	15	-,823	-,953	-,589	15	-,823	-,953	-,589
	ZLAUTDUR	31	1,336	,738	2,127	29	,119	-,127	,595	30	-,818	-,966	-,552	17	-,641	-,847	-,026	15	-,305	-,513	,469	15	-,305	-,513	,469
	ZMEANF0	31	-,168	-,890	,257	29	,077	-,396	,436	30	-,110	-,404	,519	17	,434	-,115	1,738	15	-,949	-,277	2,056	14	-,949	-,277	2,056
	ZF1	40	-,357	-,553	-,141	24	-,287	-,483	,192	7	-,606	-1,119	,108	13	,472	-,017	,893	14	-,301	-,138	,969	14	-,301	-,138	,969
ZF2	40	-,794	-,932	-,222	24	,044	-,467	,460	7	,709	-,380	2,263	13	1,414	,740	1,907	14	-,188	-,138	,534	14	-,188	-,138	,534	
ZF3	40	,046	-,363	,643	24	-,165	-,525	,345	7	-,009	-2,134	,880	13	,043	-,340	,651	14	-,1912	-,2,263	-,973	14	-,1912	-,2,263	-,973	
ZE1NORM	40	,798	,336	1,219	24	-,069	-1,254	,629	7	,403	-,279	,490	13	-,1,035	-1,245	-,691	14	2,114	,938	2,932	14	2,114	,938	2,932	
ZE2NORM	40	-,705	-1,160	-,376	24	-,054	-,423	,964	7	-,114	-,232	,240	13	,598	,278	,982	14	-,430	-,224	1,148	14	-,430	-,224	1,148	
ZE3NORM	40	-,668	-1,055	-,377	24	,117	-,316	1,332	7	-,191	-,370	,559	13	1,250	,583	2,361	14	-,458	-,121	1,250	14	-,458	-,121	1,250	
ZE4NORM	40	-,429	-,665	-,031	24	-,237	-,648	,842	7	-,510	-,620	,055	13	,296	-,170	,620	14	-,915	-,1,027	-,366	14	-,915	-,1,027	-,366	
ZLAUTDUR	40	,742	,310	1,251	24	-,625	-,906	-,100	7	-,426	-,968	-,102	13	-,889	-1,135	-,735	14	-,163	-,487	,175	14	-,163	-,487	,175	
ZMEANF0	40	-,058	-,506	,284	24	,060	-,295	,550	7	-,624	-,753	,377	13	,081	-,070	,239	14	-,163	-,487	,175	14	-,163	-,487	,175	

Tabelle C-1: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprecher- und lautnormierten Parameter einzelner Reduktionsstufen für Vokale mit fünf Reduktionsstufen (RED1-RED5).

	RED1						RED2						RED3						RED4					
	n		z		Quartile		n		z		Quartile		n		z		Quartile		n		z		Quartile	
			1.	3.					1.	3.					1.	3.					1.	3.		
ZF1	26	-,932	-1,204	-,613	43	-,404	-,928	,082	15	1,347	1,136	1,588	22	,892	,163	1,638								
ZF2	26	1,241	1,068	1,493	43	,682	,243	1,015	15	-,330	-,985	,324	22	-1,216	-1,659	-,520								
ZF3	26	1,324	1,042	1,818	43	,447	-,390	1,028	15	-,032	-,713	,300	22	-,566	-,853	,313								
ZE1NORM	26	,279	-,235	,724	43	,166	-,454	,737	15	-,413	-1,294	,075	22	,561	-,834	1,187								
ZE2NORM	26	-,668	-,959	-,513	43	-,576	-,792	-,452	15	,462	-,244	1,517	22	1,054	,201	2,151								
ZE3NORM	26	-,233	-,711	,124	43	-,093	-,804	,469	15	,814	-,041	1,360	22	-1,017	-1,269	,034								
ZE4NORM	26	-,007	-,120	,616	43	-,077	-,433	,618	15	-,426	-,755	,075	22	-,718	-,892	-,241								
ZLAUTDUR	26	1,246	,471	3,589	43	-,133	-,464	,223	15	-,261	-,612	,175	22	-,651	-,842	-,222								
ZMEANF0	26	-,179	-,420	,260	43	-,127	-,504	,213	15	,395	,189	,755	22	-,043	-,328	,391								

**Tabelle C-2: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprecher- und lautnormierten Parameter einzelner Reduktionsstufen für /i:/ mit vier Reduktionsstufen (RED1-RED4).**

		RED1				RED2				RED3			
		n	z	Quartile		n	z	Quartile		n	z	Quartile	
				1.	3.			1.	3.			1.	3.
/h/	ZF1	9	-,378	-,845	-,212	32	,258	-,331	,629	16	,506	-,081	1,279
	ZF2	9	,850	,121	1,264	32	,602	,138	,858	16	-,931	-1,791	-,519
	ZF3	9	,725	,027	1,218	32	,136	-,242	,531	16	-,183	-,629	-,073
	ZE1NORM	9	-,280	-,820	,450	32	,259	-,260	,583	16	,440	-,505	,706
	ZE2NORM	9	-,298	-1,020	,395	32	-,231	-,638	,492	16	1,308	,506	3,019
	ZE3NORM	9	1,085	-,014	1,812	32	,012	-,185	,263	16	-,761	-1,003	-,500
	ZE4NORM	9	-,156	-,320	,263	32	-,299	-,543	-,075	16	-,538	-,592	-,366
	ZLAUTDUR	9	1,442	-,061	2,275	32	,832	-,368	2,369	16	-,326	-1,043	,383
ZMEANF0	9	,177	-,550	,432	32	-,162	-,597	,231	16	,131	-,595	,524	
/y:/	ZF1	32	-,410	-,998	-,114	16	1,002	,144	1,428	10	,120	-,102	,956
	ZF2	32	,296	-,678	,983	16	-,262	-,755	,446	10	,074	-1,107	,685
	ZF3	32	-,087	-,557	,533	16	-,215	-,453	,046	10	-,182	-,564	,159
	ZE1NORM	32	,084	-,628	1,082	16	,183	-,387	,887	10	-,188	-,941	,217
	ZE2NORM	32	-,567	-1,002	,101	16	-,127	-,783	,717	10	,185	-,097	1,467
	ZE3NORM	32	-,293	-,822	,923	16	-,386	-,702	-,112	10	-,137	-,714	,713
	ZE4NORM	32	-,375	-,720	,498	16	-,291	-,738	,055	10	-,016	-,411	,319
	ZLAUTDUR	32	,486	,008	1,004	16	-,355	-,938	,258	10	-1,172	-1,272	-,623
ZMEANF0	32	-,192	-,783	,664	16	,304	-,021	,569	10	-,025	-,752	,800	
/ε:/	ZF1	8	,467	-,390	,938	10	,472	-,207	1,020	15	-,607	-1,326	,263
	ZF2	8	,501	-,244	,702	10	-,091	-,772	,179	15	-,014	-,287	,971
	ZF3	8	-,179	-,707	,334	10	,269	-,432	,811	15	,083	-,544	,316
	ZE1NORM	8	-,402	-,930	,777	10	,098	-,588	,354	15	-,042	-,427	,543
	ZE2NORM	8	,170	-,842	,472	10	,387	-,493	1,654	15	-,609	-,842	,664
	ZE3NORM	8	,102	-,501	1,109	10	-,328	-,933	,346	15	-,155	-,823	,429
	ZE4NORM	8	-,047	-,956	,272	10	-,260	-,724	1,052	15	,060	-,926	,725
	ZLAUTDUR	8	,794	-,060	1,361	10	-,489	-1,345	,037	15	-,625	-,956	-,421
ZMEANF0	8	-,247	-,644	,067	10	-,345	-,978	-,128	15	-,005	-,729	1,078	
/ø:/	ZF1	17	-,844	-1,324	-,040	20	-,232	-,643	1,115	8	,249	-,057	1,351
	ZF2	17	,433	,173	,787	20	-,215	-1,050	,333	8	-,680	-1,260	,508
	ZF3	17	-,886	-1,114	,342	20	,006	-,168	,429	8	-,141	-,399	,211
	ZE1NORM	17	,330	-,750	,918	20	,110	-,972	1,140	8	-,043	-,555	,577
	ZE2NORM	17	-,428	-,844	,539	20	-,156	-,930	,680	8	,097	-,549	1,009
	ZE3NORM	17	,029	-,576	1,172	20	-,442	-,739	-,027	8	-,093	-,747	,484
	ZE4NORM	17	-,779	-1,172	,442	20	,201	-,957	,930	8	-,027	-,333	,545
	ZLAUTDUR	17	,140	-,202	,644	20	-,043	-,693	,593	8	-,564	-,941	-,090
ZMEANF0	17	-,166	-,427	,503	20	-,410	-,643	,073	8	-,138	-,503	,166	
/a:/	ZF1	34	,633	,431	,932	22	,191	-1,267	,700	37	-,151	-1,232	,698
	ZF2	34	-,261	-,555	,218	22	,365	-,137	,840	37	-,018	-,682	,940
	ZF3	34	-,177	-,695	,241	22	,492	-,078	1,351	37	-,223	-,642	,226
	ZE1NORM	34	-,267	-,762	,526	22	-,916	-1,430	-,253	37	,509	-,258	1,110
	ZE2NORM	34	,202	-,524	,800	22	,803	,290	1,422	37	-,619	-1,036	,571
	ZE3NORM	34	-,091	-,794	,331	22	,151	-,379	,671	37	-,479	-1,130	,209
	ZE4NORM	34	-,069	-,679	,567	22	-,372	-,674	,126	37	-,608	-,829	,369
	ZLAUTDUR	34	1,175	,592	1,437	22	-,545	-,860	-,235	37	-,555	-1,029	-,165
ZMEANF0	34	-,161	-,424	,342	22	,637	,317	,890	37	-,015	-,315	,406	
/o:/	ZF1	28	,061	-,785	,508	36	-,202	-,765	,349	26	,520	-,702	1,003
	ZF2	28	-,805	-1,136	-,368	36	-,260	-,508	,293	26	,587	,081	1,188
	ZF3	28	-,067	-,652	1,281	36	-,373	-,649	,587	26	-,154	-,439	,469
	ZE1NORM	28	,818	,413	1,192	36	,203	-,371	,613	25	-,875	-2,053	-,148
	ZE2NORM	28	-,880	-1,166	-,391	36	-,207	-,543	,430	25	1,143	,108	1,711
	ZE3NORM	28	-,588	-1,015	-,381	36	-,369	-,650	,134	25	,704	-,093	1,373
	ZE4NORM	28	-,263	-,598	-,065	36	-,251	-,459	,355	25	-,065	-,399	1,244
	ZLAUTDUR	28	,633	,248	1,661	36	-,351	-,914	,259	25	-,622	-,945	-,295
ZMEANF0	28	-,008	-,428	,371	36	-,006	-,566	,424	25	,039	-,365	,641	

**Tabelle C-3: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprecher- und lautnormierten Parameter einzelner Reduktionsstufen für Vokale mit drei Reduktionsstufen (RED1-RED3).**



		RED1				RED2			
		n	z	Quartile		n	z	Quartile	
				1.	3.			1.	3.
/a/	ZF1	32	,531	-,204	1,054	13	-,165	-,970	,324
	ZF2	32	,085	-,689	,686	13	,403	-,400	1,266
	ZF3	32	-,109	-,904	,644	13	,381	-,042	,771
	ZE1NORM	32	-,832	-1,674	,133	13	,449	,234	1,310
	ZE2NORM	32	,718	-,027	1,323	13	-,626	-1,348	-,155
	ZE3NORM	32	,422	-,237	1,278	13	-,352	-,555	,023
	ZE4NORM	32	-,358	-,675	,419	13	,473	,112	,944
	ZLAUTDUR	32	,387	-,159	,923	13	-,315	-,860	-,065
ZMEANF0	32	,055	-,493	,606	13	,247	,014	,582	
/u/	ZF1	44	,007	-,624	,712	27	-,030	-,454	,826
	ZF2	44	,073	-,499	,556	27	,878	,045	1,375
	ZF3	44	,122	-,651	1,173	27	,296	-,212	1,155
	ZE1NORM	44	-,321	-,980	,546	27	-,897	-1,572	-,100
	ZE2NORM	44	,113	-,358	1,100	27	,735	-,371	1,999
	ZE3NORM	44	-,260	-,524	,697	27	,640	,322	1,757
	ZE4NORM	44	-,262	-,481	,089	27	-,027	-,354	,257
	ZLAUTDUR	44	-,109	-,893	,489	27	-,645	-1,040	-,103
ZMEANF0	44	,040	-,440	,607	27	,099	-,330	,467	

**Tabelle C-4: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprecher- und lautnormierten Parameter einzelner Reduktionsstufen für Vokale mit zwei Reduktionsstufen (RED1-RED2).**

		n	z	Quartile	
				1.	3.
/ɜ:/	ZF1	45	-,046	-,646	,627
	ZF2	45	,113	-,375	,594
	ZF3	45	-,151	-,535	,848
	ZE1NORM	45	,062	-,560	,733
	ZE2NORM	45	-,092	-,790	,495
	ZE3NORM	45	-,227	-,658	,518
	ZE4NORM	45	-,394	-,732	,738
	ZLAUTDUR	45	,263	-,271	1,084
ZMEANF0	45	-,097	-,696	,616	
/ɛ:/	ZF1	29	-,038	-,434	,320
	ZF2	29	,720	-,590	1,069
	ZF3	29	-,163	-,826	,394
	ZE1NORM	29	-,331	-1,294	,573
	ZE2NORM	29	-,055	-,476	,790
	ZE3NORM	29	,254	-,268	1,012
	ZE4NORM	29	-,045	-,512	,311
	ZLAUTDUR	29	1,021	-,187	2,267
ZMEANF0	29	,029	-,453	,655	
/œ:/	ZF1	43	,154	-,381	,580
	ZF2	43	,133	-,528	,686
	ZF3	43	-,292	-,539	,246
	ZE1NORM	43	,245	-,487	,654
	ZE2NORM	43	-,193	-,874	,559
	ZE3NORM	43	-,183	-,684	,805
	ZE4NORM	43	-,079	-,656	,738
	ZLAUTDUR	43	-,216	-,816	,439
ZMEANF0	43	-,168	-,488	,656	
/ɔ:/	ZF1	45	-,278	-,467	,567
	ZF2	45	-,182	-,551	,577
	ZF3	45	-,086	-,593	,851
	ZE1NORM	45	-,013	-,967	1,107
	ZE2NORM	45	-,220	-1,123	1,007
	ZE3NORM	45	,174	-,732	,882
	ZE4NORM	45	-,088	-,612	,657
	ZLAUTDUR	45	-,531	-,854	,588
ZMEANF0	45	,008	-,487	,450	

**Tabelle C-5: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprecher- und lautnormierten Parameter für Vokale ohne Reduktionsstufen.**

Laut	RED	ZF1	ZF2	ZF3	ZE1NORM	ZE2NORM	ZE3NORM	ZE4NORM	ZLAUTDUR	ZMEANF0	
/i/	RED1 vs. RED2	Mann-Whitney-U Wilcoxon-W Z Signifikanz	292,000 643,000 -3,306 ,001	145,000 1091,000 -5,127 ,000	226,500 1172,500 -4,117 ,000	538,000 1484,000 -260 ,795	432,500 783,500 -1,566 ,117	479,000 830,000 -991 ,322	503,000 1449,000 -693 ,488	146,000 1092,000 -5,114 ,000	539,000 890,000 -248 ,804
	RED2 vs. RED3	Mann-Whitney-U Wilcoxon-W Z Signifikanz	82,500 1028,500 -4,262 ,000	117,000 237,000 -3,649 ,000	208,500 328,500 -2,024 ,043	192,000 312,000 -2,317 ,020	56,000 1002,000 -4,732 ,000	184,000 1130,000 -2,459 ,014	225,000 345,000 -1,731 ,083	263,500 383,500 -1,048 ,295	163,000 1109,000 -2,832 ,005
	RED3 vs. RED4	Mann-Whitney-U Wilcoxon-W Z Signifikanz	123,500 376,500 -1,284 ,202	73,000 326,000 -2,846 ,004	142,000 395,000 -712 ,491	97,000 217,000 -2,104 ,036	123,000 243,000 -1,299 ,202	39,000 292,000 -3,898 ,000	116,000 369,000 -1,516 ,135	101,000 354,000 -1,981 ,049	94,000 347,000 -2,196 ,028
/h/	RED1 vs. RED2	Mann-Whitney-U Wilcoxon-W Z Signifikanz	59,000 104,000 -2,677 ,006	102,000 630,000 -1,323 ,195	104,000 632,000 -1,260 ,218	103,000 148,000 -1,291 ,206	122,000 167,000 -693 ,505	76,000 604,000 -2,142 ,032	108,000 668,500 -1,134 ,269	140,500 668,500 -1,110 ,914	124,000 652,000 -630 ,546
	RED2 vs. RED3	Mann-Whitney-U Wilcoxon-W Z Signifikanz	189,500 717,500 -1,455 ,146	7,000 143,000 -5,446 ,000	133,500 269,500 -2,679 ,007	241,000 769,000 -328 ,743	69,000 597,000 -4,090 ,000	74,000 210,000 -3,980 ,011	139,000 275,000 -2,559 ,004	124,000 260,000 -2,887 ,004	216,000 744,000 -875 ,382
	RED1 vs. RED2	Mann-Whitney-U Wilcoxon-W Z Signifikanz	78,500 606,500 -3,882 ,000	185,000 321,000 -1,553 ,120	223,000 359,000 -722 ,470	250,000 778,000 -131 ,896	188,000 716,000 -1,487 ,137	228,000 364,000 -612 ,540	239,000 375,000 -372 ,710	110,000 246,000 -3,193 ,001	198,000 726,000 -1,268 ,205
/y/	RED2 vs. RED3	Mann-Whitney-U Wilcoxon-W Z Signifikanz	48,000 103,000 -1,687 ,097	65,000 201,000 -791 ,452	76,500 212,500 -184 ,856	51,000 106,000 -1,528 ,135	51,000 187,000 -1,528 ,135	66,000 202,000 -738 ,484	52,000 188,000 -1,476 ,150	26,000 81,000 -2,847 ,003	59,000 114,000 -1,107 ,286

**Tabelle C-6: Ergebnisse des Vergleichs zwischen den sprecher- und lautnormierten Parametern aufeinander folgender Reduktionsstufen von /i/, /h/ und /y/.**

Laut	RED	ZF1	ZF2	ZF3	ZE1NORM	ZE2NORM	ZE3NORM	ZE4NORM	ZLAUTDUR	ZMEANF0	
/e:/	RED1 vs. RED2	Mann-Whitney-U	245,500	201,500	244,000	296,000	168,000	326,000	382,000	98,500	340,000
		Wilcoxon-W	741,500	636,500	679,000	731,000	664,000	822,000	817,000	533,500	836,000
		Z	-3,018	-3,669	-3,040	-2,271	-4,164	-1,827	-1,999	-5,193	-1,620
		Signifikanz	,003	,000	,002	,023	,000	,068	,318	,000	,105
	RED2 vs. RED3	Mann-Whitney-U	59,000	78,000	286,000	269,000	44,000	306,000	272,000	48,500	429,000
		Wilcoxon-W	494,000	543,000	751,000	734,000	479,000	771,000	737,000	513,500	894,000
		Z	-5,701	-5,413	-2,259	-2,517	-5,928	-1,956	-2,471	-5,860	-0,091
		Signifikanz	,000	,000	,024	,012	,000	,050	,013	,000	,928
	RED3 vs. RED4	Mann-Whitney-U	65,000	236,000	219,500	191,000	175,000	250,000	185,000	170,500	181,000
		Wilcoxon-W	218,000	389,000	684,500	656,000	328,000	403,000	650,000	635,500	646,000
		Z	-4,207	-4,421	-7,86	-1,417	-1,771	-1,11	-1,550	-1,871	-1,638
		Signifikanz	,000	,674	,432	,156	,077	,912	,121	,061	,101
RED4 vs. RED5	Mann-Whitney-U	69,000	111,000	83,000	71,000	69,000	103,000	106,000	92,500	72,000	
	Wilcoxon-W	189,000	231,000	236,000	224,000	189,000	223,000	226,000	212,500	192,000	
	Z	-2,209	-6,23	-1,680	-2,134	-2,209	-9,25	-8,12	-1,322	-2,096	
	Signifikanz	,027	,551	,097	,033	,027	,370	,433	,189	,037	
/ɛ:/	RED1 vs. RED2	Mann-Whitney-U	36,500	21,000	26,000	38,000	27,000	31,000	37,000	16,000	35,000
		Wilcoxon-W	72,500	76,000	62,000	74,000	63,000	86,000	73,000	71,000	90,000
		Z	-3,311	-1,688	-1,244	-1,178	-1,155	-800	-267	-2,134	-444
		Signifikanz	,762	,101	,237	,897	,274	,460	,829	,034	,696
	RED2 vs. RED3	Mann-Whitney-U	40,000	55,000	61,000	65,000	46,000	62,000	72,000	71,500	48,000
		Wilcoxon-W	160,000	110,000	181,000	120,000	166,000	117,000	127,000	191,500	103,000
		Z	-1,941	-1,110	-777	-555	-1,609	-721	-1,166	-1,194	-1,498
		Signifikanz	,055	,285	,461	,605	,115	,495	,892	,849	,144
	RED1 vs. RED2	Mann-Whitney-U	107,000	84,000	93,000	165,000	156,000	114,000	118,000	128,500	121,000
		Wilcoxon-W	260,000	294,000	246,000	318,000	309,000	324,000	271,000	338,500	331,000
		Z	-1,920	-2,621	-2,347	-1,152	-4,27	-1,707	-1,585	-1,265	-1,493
		Signifikanz	,056	,008	,018	,892	,684	,091	,117	,209	,141
RED2 vs. RED3	Mann-Whitney-U	50,500	71,000	58,000	68,000	69,000	74,000	72,000	48,500	60,000	
	Wilcoxon-W	260,500	107,000	94,000	104,000	279,000	284,000	108,000	84,500	270,000	
	Z	-1,501	-4,58	-1,119	-610	-559	-305	-407	-1,602	-1,017	
	Signifikanz	,136	,672	,281	,566	,601	,784	,709	,110	,328	

Tabelle C-7: Ergebnisse des Vergleichs zwischen den sprecher- und lautnormierten Parametern aufeinander folgender Reduktionsstufen von /ɛ:/, /ɛ:/ und /ø:/.

Laut	RED	ZF1	ZF2	ZF3	ZE1NORM	ZE2NORM	ZE3NORM	ZE4NORM	ZLAUTDUR	ZMEANFO
/a/	Mann-Whitney-U	220,000	195,000	200,000	234,000	230,000	275,000	301,000	23,000	164,000
	Wilcoxon-W	473,000	790,000	795,000	487,000	825,000	870,000	554,000	276,000	759,000
	Z	-2,584	-3,003	-2,919	-2,349	-2,416	-1,661	-1,225	-5,889	-3,523
	Signifikanz	,010	,003	,004	,019	,016	,097	,221	,000	,000
	Mann-Whitney-U	368,500	352,000	225,500	131,000	166,000	247,000	379,000	379,000	174,000
	Wilcoxon-W	1071,500	1055,000	928,500	384,000	869,000	950,000	1082,000	1082,000	877,000
	Z	-603	-862	-2,845	-4,326	-3,778	-2,508	-439	-439	-3,652
	Signifikanz	,546	,389	,004	,000	,000	,012	,661	,661	,000
	Mann-Whitney-U	111,000	162,000	145,500	66,000	56,000	109,000	137,000	112,000	171,000
	Wilcoxon-W	202,000	690,000	673,500	594,000	147,000	200,000	665,000	203,000	699,000
Z	-2,429	-1,152	-1,565	-3,556	-3,806	-2,479	-1,778	-2,404	-927	
Signifikanz	,015	,249	,118	,000	,000	,013	,075	,016	,354	
/a/	Mann-Whitney-U	420,000	233,500	416,000	200,000	161,000	190,000	409,000	97,500	374,000
	Wilcoxon-W	1240,000	1063,500	716,000	500,000	981,000	1010,000	1229,000	397,500	1194,000
	Z	-832	-3,419	-888	-3,883	-4,424	-4,022	-985	-5,305	-1,470
	Signifikanz	,405	,001	,375	,000	,000	,000	,325	,000	,142
	Mann-Whitney-U	57,000	63,000	81,500	74,000	73,000	72,000	78,000	75,500	48,000
	Wilcoxon-W	85,000	363,000	109,500	374,000	101,000	100,000	106,000	375,500	76,000
	Z	-1,276	-992	-1,118	-4,472	-520	-567	-283	-402	-1,701
	Signifikanz	,216	,341	,908	,661	,627	,595	,800	,695	,094
	Mann-Whitney-U	13,000	39,000	36,000	15,000	15,000	17,000	23,000	25,500	28,000
	Wilcoxon-W	41,000	67,000	64,000	106,000	43,000	45,000	51,000	116,500	56,000
Z	-2,576	-515	-753	-2,417	-2,417	-2,258	-1,783	-1,586	-1,387	
Signifikanz	,008	,643	,485	,014	,014	,024	,081	,115	,183	
/u/	Mann-Whitney-U	551,500	343,000	547,000	415,000	494,000	310,000	461,000	432,000	582,000
	Wilcoxon-W	1541,500	1333,000	1537,000	793,000	1484,000	1300,000	1451,000	810,000	1572,000
	Z	-503	-2,973	-557	-2,120	-1,184	-3,364	-1,575	-1,919	-1,142
	Signifikanz	,615	,003	,578	,034	,236	,001	,115	,055	,887
	Mann-Whitney-U	434,500	247,000	435,000	220,000	190,000	321,000	434,000	191,500	500,000
	Wilcoxon-W	1100,500	653,000	1101,000	886,000	596,000	727,000	840,000	857,500	906,000
	Z	-941	-3,478	-934	-3,843	-4,249	-2,477	-947	-4,230	-054
	Signifikanz	,347	,001	,350	,000	,000	,013	,343	,000	,957
	Mann-Whitney-U	301,000	187,500	450,000	186,000	215,000	207,000	359,000	363,000	392,500
	Wilcoxon-W	967,000	853,500	1116,000	511,000	881,000	873,000	1025,000	688,000	1058,500
Z	-2,382	-4,001	-257	-3,871	-3,446	-3,564	-1,334	-1,276	-843	
Signifikanz	,017	,000	,797	,000	,001	,000	,182	,202	,399	

**Tabelle C-8: Ergebnisse des Vergleichs zwischen den sprecher- und lautnormierten Parametern aufeinander folgender Reduktionsstufen von /a/, /a/, /u/, /u/ und /o/.**

# ANHANG D

	RED1						RED2						RED3						RED4						RED5																																																																																																																																																																																																																									
	n		z		Quartile		n		z		Quartile		n		z		Quartile		n		z		Quartile		n		z		Quartile																																																																																																																																																																																																																					
	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.																																																																																																																																																																																																																				
/e:/	ZSPF1	31	-540	-653	-240	29	-274	-451	-007	30	527	213	777	17	-293	-470	137	15	-522	-916	15	-287	31	1686	1534	1841	29	1321	919	1645	30	410	176	666	17	466	269	718	15	311	108	494	31	723	187	1047	29	332	-237	530	30	-211	-494	185	17	-044	-402	461	15	450	-337	1788	31	1526	918	2349	29	346	032	773	30	-653	-810	-396	17	-482	-704	131	15	-667	-799	-429	31	2135	1847	2413	29	1485	1128	2015	30	-229	-393	451	17	606	-035	1163	15	1002	532	1298	31	-1119	-1465	-717	29	-1107	-1452	-840	30	-647	-946	-198	17	-467	-957	007	15	-056	-938	1777	40	-696	-821	-565	24	-702	-887	-359	7	-788	-1075	-388	13	-314	-483	052	14	-005	-298	762	40	-1994	-2266	-1432	24	-1312	-1828	-796	7	-500	-1778	600	13	-348	-689	280	14	-1136	-1467	-613	40	-220	-848	397	24	-476	-954	266	7	-069	-2758	621	13	-317	-747	250	14	-180	-612	202	40	1209	703	1955	24	-363	-696	154	7	-200	-825	205	13	-690	-994	-475	14	-679	-865	-128	40	-1264	-1605	-563	24	-849	-1326	-062	7	581	-1059	831	13	-134	-1079	549	14	-1428	-1877	-502	40	1359	614	2320	24	861	187	1487	7	021	-980	432	13	-049	-750	939	14	640	353	1484

Tabelle D-1: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprechnormierten Parameter einzelner Reduktionsstufen für Vokale mit fünf Reduktionsstufen (RED1-RED5).

	RED1						RED2						RED3						RED4																																																																														
	n		z		Quartile		n		z		Quartile		n		z		Quartile		n		z		Quartile		n		z		Quartile																																																																				
	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.	1.	3.																																																																			
/i:/	ZSPF1	26	-931	-1036	-764	43	-653	-901	-509	15	000	-075	104	22	-203	-431	131	26	1966	1870	2072	43	1580	1310	1797	15	949	538	1344	22	369	114	649	26	1298	978	2034	43	544	-154	1160	15	193	-397	483	22	-243	-495	601	26	1395	515	3873	43	-133	-482	258	15	-244	-612	069	22	-655	-867	-205	26	2906	2728	3095	43	2289	1743	2553	15	874	375	1361	22	623	125	867	26	-715	-946	-019	43	-1146	-1410	-584	15	-746	-1120	-489	22	-744	-976	-119

Tabelle D-2: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprechnormierten Parameter einzelner Reduktionsstufen für /i:/ mit vier Reduktionsstufen (RED1-RED4).

		RED1				RED2				RED3			
		n	z	Quartile		n	z	Quartile		n	z	Quartile	
				1.	3.			1.	3.			1.	3.
/i/	ZSPF1	9	-,769	-1,031	-,627	32	-,409	-,755	-,238	16	-,132	-,628	,164
	ZSPF2	9	1,398	,842	1,598	32	1,133	,801	1,391	16	-,002	-,534	,137
	ZSPF3	9	,938	,396	1,601	32	,524	,173	,902	16	,209	-,280	,335
	ZSPLAUT-DUR	9	,412	-,508	,793	32	,020	-,647	,826	16	-,654	-,994	-,237
	ZSPR21	9	2,131	1,509	2,477	32	1,559	1,173	2,041	16	,224	-,286	,448
	ZSPR32	9	-,441	-,832	,337	32	-,495	-,854	-,294	16	,010	-,397	,484
/y/	ZSPF1	32	-,690	-,909	-,537	16	-,141	-,360	,041	10	-,377	-,496	-,232
	ZSPF2	32	-,015	-,413	,484	16	-,262	-,491	,051	10	-,079	-,573	,166
	ZSPF3	32	-,802	-1,181	-,359	16	-,844	-1,066	-,674	10	-,844	-1,185	-,614
	ZSPLAUT-DUR	32	,708	,172	1,155	16	-,255	-,729	,561	10	-1,023	-1,084	-,447
	ZSPR21	32	,716	-,006	1,391	16	-,158	-,394	,261	10	,301	,009	,517
	ZSPR32	32	-,823	-1,182	,233	16	-,743	-,959	-,403	10	-,786	-1,052	-,443
/e/	ZSPF1	8	,313	-,023	,419	10	,413	-,010	,581	15	-,170	-,479	,098
	ZSPF2	8	1,000	,723	1,194	10	,680	,585	,995	15	,681	,554	1,241
	ZSPF3	8	-,042	-,367	,219	10	,212	-,291	,390	15	,049	-,164	,210
	ZSPLAUT-DUR	8	1,803	,806	2,444	10	,443	-,240	,971	15	,284	-,059	,511
	ZSPR21	8	,639	,311	1,209	10	,309	-,037	,978	15	,933	,525	1,300
	ZSPR32	8	-1,083	-1,321	-,842	10	-,572	-,845	-,329	15	-,718	-1,045	-,605
/ø/	ZSPF1	17	-,456	-,569	-,170	20	-,213	-,364	,249	8	-,052	-,158	,193
	ZSPF2	17	-,109	-,309	,034	20	-,482	-,812	-,236	8	-,701	-,903	-,271
	ZSPF3	17	-1,481	-1,725	-,513	20	-,787	-,892	-,497	8	-,960	-1,168	-,681
	ZSPLAUT-DUR	17	,682	,163	,976	20	,308	-,108	,818	8	,099	-,689	,561
	ZSPR21	17	,289	-,126	,622	20	-,579	-,837	-,041	8	-,564	-1,190	-,237
	ZSPR32	17	-1,257	-1,609	-,704	20	-,364	-,608	-,096	8	-,408	-,609	,037
/a/	ZSPF1	34	2,264	2,076	2,546	22	1,709	,252	2,290	37	1,348	,104	2,279
	ZSPF2	34	-,680	-,834	-,543	22	-,463	-,726	-,360	37	-,662	-,841	-,379
	ZSPF3	34	-,591	-,861	-,303	22	-,184	-,562	,343	37	-,597	-,867	-,359
	ZSPLAUT-DUR	34	2,886	2,087	3,252	22	,561	,139	,977	37	,548	-,099	,996
	ZSPR21	34	-3,103	-3,375	-2,703	22	-2,274	-2,791	-,748	37	-1,719	-2,851	-,804
	ZSPR32	34	,161	-,082	,292	22	,423	-,039	1,210	37	,003	-,242	,429
/o/	ZSPF1	28	-,201	-,451	,064	36	-,227	-,469	-,047	25	,017	-,333	,206
	ZSPF2	28	-2,122	-2,341	-1,680	36	-1,628	-1,926	-1,267	25	-1,115	-1,355	-,665
	ZSPF3	28	-,373	-,996	1,281	36	-,590	-1,086	,422	25	-,514	-,891	,152
	ZSPLAUT-DUR	28	1,503	1,043	2,800	36	,236	-,533	1,057	25	-,059	-,522	,271
	ZSPR21	28	-1,918	-2,096	-1,661	36	-1,384	-1,673	-,870	25	-1,041	-1,692	-,683
	ZSPR32	28	1,596	1,011	3,415	36	1,128	,401	1,922	25	,400	-,318	1,155

**Tabelle D-3: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprechernormierten Parameter einzelner Reduktionsstufen für Vokale mit drei Reduktionsstufen (RED1-RED3).**



		RED1				RED2			
		n	z	Quartile		n	z	Quartile	
				1.	3.			1.	3.
/a/	ZSPF1	32	1,566	,899	2,321	13	,821	-,035	1,388
	ZSPF2	32	-,283	-,650	-,013	13	-,218	-,544	,207
	ZSPF3	32	-,189	-,610	,440	13	,036	-,068	,461
	ZSPLAUDUR	32	,376	-,038	,765	13	-,218	-,503	,060
	ZSPR21	32	-2,123	-2,743	-1,135	13	-,987	-1,425	-,390
	ZSPR32	32	,215	-,077	,657	13	,301	-,126	,718
/u/	ZSPF1	44	-,437	-,678	-,003	27	-,314	-,618	,098
	ZSPF2	44	-,959	-1,356	-,578	27	-,363	-,989	,234
	ZSPF3	44	-,108	-,940	1,475	27	,383	-,202	1,230
	ZSPLAUDUR	44	-,391	-,856	-,078	27	-,690	-,894	-,421
	ZSPR21	44	-,698	-1,272	-,023	27	,037	-,615	,357
	ZSPR32	44	,852	,057	2,040	27	,808	-,413	1,533

**Tabelle D-4: Median (z) und Quartile (1. und 3.) der sprechnormierten Parameter einzelner Reduktionsstufen für Vokale mit zwei Reduktionsstufen (RED1-RED2).**

		n	z	Quartile	
				1.	3.
/ɔ/	ZSPF1	45	-,483	-,684	-,112
	ZSPF2	45	-,172	-,465	,058
	ZSPF3	45	-,926	-1,165	-,169
	ZSPLAUDUR	45	-,218	-,470	,067
	ZSPR21	45	,188	-,310	,725
	ZSPR32	45	-,628	-,977	,089
/ɛ/	ZSPF1	29	,148	-,044	,562
	ZSPF2	29	,739	,159	1,063
	ZSPF3	29	-,102	-,407	,463
	ZSPLAUDUR	29	,680	-,272	1,361
	ZSPR21	29	,365	-,211	,912
	ZSPR32	29	-,487	-1,226	-,116
/œ/	ZSPF1	43	,007	-,327	,320
	ZSPF2	43	,039	-,276	,405
	ZSPF3	43	-,690	-,983	-,219
	ZSPLAUDUR	43	-,077	-,330	,310
	ZSPR21	43	-,051	-,570	,665
	ZSPR32	43	-,631	-1,065	-,092
/ɔ/	ZSPF1	45	,311	-,130	,751
	ZSPF2	45	-1,129	-1,488	-,810
	ZSPF3	45	-,105	-,668	,656
	ZSPLAUDUR	45	-,324	-,507	,360
	ZSPR21	45	-1,555	-2,010	-,830
	ZSPR32	45	,988	,478	1,633

**Tabelle D-5: Anzahl (n), Median (z) und Quartile (1. und 3.) der sprechnormierten Parameter für Vokale ohne Reduktionsstufen.**

		ZSPR21	ZSPR32	ZSPLAUDUR	ZSPF1	ZSPF2	ZSPF3
/i:/ RED2 vs. /ɪ/ RED1	Mann-Whitney-U	165,000	88,000	169,000	136,500	123,000	157,000
	Wilcoxon-W	210,000	1034,000	1115,000	181,500	168,000	1103,000
	Z	-,689	-2,552	-,593	-1,379	-1,705	-,883
	Signifikanz	,505	,009	,568	,171	,091	,390
/i:/ RED3 vs. /ɪ/ RED2	Mann-Whitney-U	121,000	155,000	182,500	80,000	194,000	135,000
	Wilcoxon-W	241,000	275,000	302,500	608,000	314,000	255,000
	Z	-2,716	-1,940	-1,312	-3,652	-1,050	-2,396
	Signifikanz	,007	,052	,189	,000	,294	,017
/e:/ RED5 vs. /i:/ RED3	Mann-Whitney-U	112,000	64,000	62,000	35,000	47,500	81,000
	Wilcoxon-W	232,000	184,000	182,000	155,000	167,500	201,000
	Z	-,021	-2,012	-2,099	-3,215	-2,696	-1,307
	Signifikanz	1,000	,045	,037	,001	,006	,202
/e:/ RED5 vs. /ɪ/ RED2	Mann-Whitney-U	110,000	177,000	145,000	212,000	73,000	227,000
	Wilcoxon-W	230,000	705,000	265,000	332,000	193,000	347,000
	Z	-2,967	-1,438	-2,168	-,639	-3,811	-,297
	Signifikanz	,003	,150	,030	,523	,000	,767

**Tabelle D-6: Ergebnisse des Vergleichs zwischen den sprechernormierten Parametern einzelner Reduktionsstufen von /i:/, /ɪ/ und /e:/.**

		ZSPR21	ZSPR32	ZSPLAUDUR	ZSPF1	ZSPF2	ZSPF3
/u:/ RED2 vs. /ʊ/ RED1	Mann-Whitney-U	488,000	507,000	457,500	345,000	376,000	416,000
	Wilcoxon-W	788,000	807,000	1447,500	645,000	676,000	716,000
	Z	-,513	-,269	-,905	-2,349	-1,951	-1,437
	Signifikanz	,608	,788	,366	,019	,051	,151
/u:/ RED2 vs. /o:/ RED2	Mann-Whitney-U	214,500	365,000	303,000	199,000	308,500	413,000
	Wilcoxon-W	880,500	665,000	603,000	499,000	974,500	1079,000
	Z	-3,282	-1,011	-1,947	-3,516	-1,864	-,287
	Signifikanz	,001	,312	,052	,000	,062	,774
/u/ RED1 vs. /o:/ RED2	Mann-Whitney-U	374,000	703,000	466,500	670,500	306,000	625,000
	Wilcoxon-W	1040,000	1693,000	1456,500	1660,500	972,000	1291,000
	Z	-4,042	-,861	-3,148	-1,175	-4,700	-1,615
	Signifikanz	,000	,389	,002	,240	,000	,106
/u:/ RED2 vs. /ɔ/ RED2	Mann-Whitney-U	256,000	457,000	460,500	154,000	461,000	404,000
	Wilcoxon-W	1291,000	757,000	760,500	454,000	761,000	704,000
	Z	-3,578	-1,046	-1,002	-4,863	-,995	-1,713
	Signifikanz	,000	,296	,316	,000	,320	,087
/u/ RED1 vs. /ɔ/ RED2	Mann-Whitney-U	429,000	880,000	693,500	416,500	796,000	979,500
	Wilcoxon-W	1464,000	1870,000	1683,500	1406,500	1831,000	2014,500
	Z	-4,604	-,903	-2,433	-4,706	-1,592	-,086
	Signifikanz	,000	,367	,015	,000	,111	,931
/u:/ RED4 vs. /y:/ RED2	Mann-Whitney-U	97,000	49,000	51,000	94,500	93,000	36,500
	Wilcoxon-W	188,000	185,000	142,000	185,500	184,000	172,500
	Z	-,307	-2,412	-2,325	-,417	-,482	-2,960
	Signifikanz	,779	,015	,020	,682	,650	,002
/u:/ RED4 vs. /ɣ/ RED2	Mann-Whitney-U	215,000	209,000	111,500	209,500	264,000	198,000
	Wilcoxon-W	306,000	1244,000	202,500	1244,500	355,000	1233,000
	Z	-1,445	-1,557	-3,376	-1,548	-,531	-1,762
	Signifikanz	,148	,119	,001	,122	,595	,078

**Tabelle D-7: Ergebnisse des Vergleichs zwischen den sprechernormierten Parametern einzelner Reduktionsstufen von /u:/, /ʊ/, /o:/ und /y:/ sowie /ɔ/ und /ɣ/ ohne Reduktionsstufen.**

		ZSPR21	ZSPR32	ZSPLAUDUR	ZSPF1	ZSPF2	ZSPF3
/a:/ RED2 vs. /a/ RED1	Mann-Whitney-U	345,000	279,000	275,000	316,000	234,000	336,000
	Wilcoxon-W	873,000	807,000	803,000	569,000	487,000	864,000
	Z	-,123	-1,285	-1,356	-,634	-2,077	-,282
	Signifikanz	,902	,199	,175	,526	,038	,778

**Tabelle D-8: Ergebnisse des Vergleichs zwischen den sprechnormierten Parametern einzelner Reduktionsstufen von /a:/ und /a/.**

		ZSPR21	ZSPR32	ZSPLAUDUR	ZSPF1	ZSPF2	ZSPF3
/e:/ RED3 vs. /ɛ/	Mann-Whitney-U	292,000	408,000	177,500	274,500	342,000	381,500
	Wilcoxon-W	757,000	873,000	642,500	709,500	807,000	846,500
	Z	-2,168	-,409	-3,904	-2,434	-1,410	-,811
	Signifikanz	,030	,682	,000	,015	,159	,417
/ɛ:/ RED2 vs. /ɛ/	Mann-Whitney-U	137,000	129,000	140,500	122,500	135,000	122,000
	Wilcoxon-W	572,000	184,000	195,500	557,500	570,000	557,000
	Z	-,257	-,515	-,145	-,724	-,322	-,740
	Signifikanz	,812	,623	,887	,475	,764	,475
/y:/ RED 2 vs. /ʏ/	Mann-Whitney-U	213,000	182,000	74,000	203,500	200,000	212,000
	Wilcoxon-W	1248,000	237,000	129,000	1238,500	1235,000	267,000
	Z	-,262	-,938	-3,296	-,469	-,546	-,284
	Signifikanz	,793	,348	,001	,639	,585	,777
/ø:/ RED2 vs. /œ/	Mann-Whitney-U	286,000	307,000	308,500	383,000	200,000	372,000
	Wilcoxon-W	496,000	1253,000	1254,500	593,000	410,000	582,000
	Z	-2,126	-1,816	-1,794	-,694	-3,396	-,856
	Signifikanz	,033	,069	,073	,488	,001	,392
/o:/ RED2 vs. /ɔ/	Mann-Whitney-U	678,000	803,000	659,500	345,000	420,500	591,000
	Wilcoxon-W	1713,000	1838,000	1694,500	1011,000	1086,500	1257,000
	Z	-1,255	-,067	-1,431	-4,420	-3,702	-2,082
	Signifikanz	,210	,947	,153	,000	,000	,037

**Tabelle D-9: Ergebnisse des Vergleichs zwischen den sprechnormierten Parametern einzelner Reduktionsstufen gespannter Vokale und ungespannter Vokale ohne Reduktionsstufen mit vergleichbarer Lautqualität.**

		ZSPR21	ZSPR32	ZSPLAUDUR	ZSPF1	ZSPF2	ZSPF3
/i:/ RED4 vs. /ɪ/ RED3	Mann-Whitney-U	102,000	85,000	167,000	172,500	60,000	158,000
	Wilcoxon-W	238,000	338,000	303,000	308,500	196,000	411,000
	Z	-2,188	-2,690	-,266	-,103	-3,430	-,532
	Signifikanz	,029	,006	,804	,919	,000	,609
/u:/ RED5 vs. /ʊ/ RED2	Mann-Whitney-U	69,000	176,000	173,000	127,000	83,000	127,000
	Wilcoxon-W	174,000	554,000	551,000	505,000	188,000	232,000
	Z	-3,299	-,357	-,440	-1,705	-2,914	-1,705
	Signifikanz	,001	,734	,674	,091	,003	,091
/a:/ RED3 vs. /a/ RED2	Mann-Whitney-U	143,000	181,000	117,500	186,000	104,000	57,000
	Wilcoxon-W	846,000	884,000	208,500	277,000	807,000	760,000
	Z	-2,156	-1,316	-2,721	-1,205	-3,019	-4,059
	Signifikanz	,031	,188	,007	,228	,003	,000

**Tabelle D-10: Ergebnisse des Vergleichs zwischen den sprechnormierten Parametern der Reduktionsstufen gespannter Vokale und ungespannter Vokale mit Reduktionsstufen mit zentralisierter oder [ə]-Qualität.**

		ZSPR21	ZSPR32	ZSPLAUTDUR	ZSPF1	ZSPF2	ZSPF3
/ɛ:/ RED3 vs. /ɛ/	Mann-Whitney-U	140,000	167,000	200,000	132,000	195,000	203,000
	Wilcoxon-W	575,000	287,000	320,000	252,000	630,000	638,000
	Z	-1,919	-1,250	-,433	-2,117	-,557	-,359
	Signifikanz	,055	,211	,665	,034	,577	,720
/o:/ RED3 vs. /ɔ/	Mann-Whitney-U	403,000	332,000	558,000	381,500	518,000	407,000
	Wilcoxon-W	1438,000	657,000	1593,000	706,500	1553,000	732,000
	Z	-1,955	-2,825	-,055	-2,219	-,545	-1,906
	Signifikanz	,051	,005	,956	,027	,585	,057
/y:/ RED3 vs. /ʏ/	Mann-Whitney-U	213,000	182,000	74,000	203,500	200,000	212,000
	Wilcoxon-W	1248,000	237,000	129,000	1238,500	1235,000	267,000
	Z	-,262	-,938	-3,296	-,469	-,546	-,284
	Signifikanz	,793	,348	,001	,639	,585	,777
/ø:/ RED3 vs. /œ/	Mann-Whitney-U	87,000	117,000	167,500	169,000	63,000	112,000
	Wilcoxon-W	123,000	1063,000	203,500	1115,000	99,000	148,000
	Z	-2,202	-1,425	-,117	-,078	-2,823	-1,554
	Signifikanz	,027	,161	,909	,949	,003	,125

**Tabelle D-11: Ergebnisse des Vergleichs zwischen den sprechernormierten Parametern der Reduktionsstufen gespannter Vokale und ungespannter Vokale ohne Reduktionsstufen mit zentralisierter oder [ə]-Qualität.**

## ANHANG E

Anzahl der Reduktionsstufen	Wortklasse	Reduktionsstufe	Häufigkeit	Prozent
2 RED	Inhaltswörter	RED1	48	69,6
		RED2	21	30,4
		Gesamt	69	100,0
	Funktionswörter	RED1	28	59,6
		RED2	19	40,4
		Gesamt	47	100,0
3 RED	Inhaltswörter	RED1	89	38,4
		RED2	84	36,2
		RED3	59	25,4
		Gesamt	232	100,0
	Funktionswörter	RED1	39	27,1
		RED2	52	36,1
		RED3	53	36,8
		Gesamt	144	100,0
4 RED	Inhaltswörter	RED1	20	47,6
		RED2	16	38,1
		RED3	4	9,5
		RED4	2	4,8
		Gesamt	42	100,0
	Funktionswörter	RED1	6	9,4
		RED2	27	42,2
		RED3	11	17,2
		RED4	20	31,3
		Gesamt	64	100,0
5 RED	Inhaltswörter	RED1	55	55,6
		RED2	28	28,3
		RED3	3	3,0
		RED4	6	6,1
		RED5	7	7,1
		Gesamt	99	100,0
	Funktionswörter	RED1	16	13,2
		RED2	25	20,7
		RED3	34	28,1
		RED4	24	19,8
		RED5	22	18,2
Gesamt	121	100,0		

**Tabelle E-1: Häufigkeiten (Anzahl und Prozente) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Funktions- bzw. Inhaltswörter für die nach Reduktionsstufenanzahl gruppierten Vokale.**

Anzahl der Reduktionsstufen	Wortklasse	Chi-Quadrat	df	Signifikanz
2 RED	Inhaltswörter	10,565	1	,001
	Funktionswörter	1,723	1	,189
3 RED	Inhaltswörter	6,681	2	,035
	Funktionswörter	2,542	2	,281
4 RED	Inhaltswörter	22,381	3	,000
	Funktionswörter	16,375	3	,001
5 RED	Inhaltswörter	98,121	4	,000
	Funktionswörter	6,975	4	,137

**Tabelle E-2: Prüfung ( $\chi^2$ -Test) der Verteilung von Reduktionsstufen in Abhängigkeit der Wortklasse für die nach Reduktionsstufenanzahl gruppierten Vokale.**

Anzahl der Reduktionsstufen	Vorhandensein eines F0 Maximums	Reduktionsstufe	Häufigkeit	Prozent		
2 RED	kein F0-Maximum	RED1	21	58,3		
		RED2	15	41,7		
		Gesamt	36	100,0		
	F0-Maximum	RED1	32	84,2		
		RED2	6	15,8		
		Gesamt	38	100,0		
3 RED	kein F0-Maximum	RED1	22	20,0		
		RED2	46	41,8		
		RED3	42	38,2		
		Gesamt	110	100,0		
	F0-Maximum	RED1	95	49,5		
		RED2	59	30,7		
		RED3	38	19,8		
		Gesamt	192	100,0		
4 RED	kein F0-Maximum	RED1	1	2,1		
		RED2	23	47,9		
		RED3	7	14,6		
		RED4	17	35,4		
		Gesamt	48	100,0		
	F0-Maximum	RED1	23	42,6		
		RED2	19	35,2		
		RED3	8	14,8		
		RED4	4	7,4		
		Gesamt	54	100,0		
		5 RED	kein F0-Maximum	RED1	17	15,2
				RED2	26	23,2
RED3	26			23,2		
RED4	21			18,8		
RED5	22			19,6		
Gesamt	112			100,0		
F0-Maximum	RED1		44	54,3		
	RED2		20	24,7		
	RED3		9	11,1		
	Gesamt		81	100,0		

**Tabelle E-3: Häufigkeiten (Anzahl und Prozente) einzelner Reduktionsstufen bezogen auf die Gesamtheit der Silben mit bzw. ohne F0-Maximum für die nach Reduktionsstufenanzahl gruppierten Vokale.**

	n	Median	Interquartilbereich
Inhaltswörter	212	20,00	10,00
Funktionswörter	89	17,00	15,50

**Tabelle E-4: Prominenz von Silben mit nicht reduzierten Vokalen in Abhängigkeit der Wortklasse.**

	Prominenz
Mann-Whitney-U	8485,500
Wilcoxon-W	12490,500
Z	-1,379
Signifikanz	,168

**Tabelle E-5: Ergebnisse des Vergleichs der perceptiven Prominenz von Silben mit nicht reduzierten Vokalen in Abhängigkeit der Wortklasse.**

## ANHANG F

		/a:/			/i:/			/u:/		
		n	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s
ZF1	SPR1	50	-,080	1,062	60	,062	1,285	60	,016	1,157
	SPR2	50	,088	,935	60	-,106	,901	60	-,090	,903
	SPR3	50	,149	1,021	60	-,130	1,021	60	,106	1,096
	Gesamt	150	,052	1,005	180	-,058	1,078	180	,011	1,055
ZF2	SPR1	50	-,045	1,046	60	,136	1,255	60	-,022	1,080
	SPR2	50	-,048	,688	60	-,070	1,117	60	-,080	,927
	SPR3	50	,156	,875	60	,240	,686	60	,093	,929
	Gesamt	150	,021	,881	180	,102	1,050	180	-,003	,978
ZF3	SPR1	50	,064	1,000	60	,212	1,032	60	-,001	1,030
	SPR2	50	,055	,688	60	-,068	1,071	60	-,004	1,017
	SPR3	50	-,100	,841	60	,261	1,087	60	,073	1,086
	Gesamt	150	,007	,850	180	,135	1,067	180	,023	1,040
ZLAUTDUR	SPR1	50	-,169	,975	60	,376	1,443	60	,125	1,065
	SPR2	50	,183	1,084	60	,039	1,407	60	,030	1,089
	SPR3	50	,122	,981	60	,027	1,137	60	-,068	,988
	Gesamt	150	,045	1,020	180	,147	1,338	180	,029	1,046
ZMEANF0	SPR1	50	,118	,827	60	,061	1,415	60	-,033	1,189
	SPR2	50	-,038	,907	60	-,008	1,155	60	,078	1,196
	SPR3	50	-,018	,936	60	,079	1,089	60	,037	,953
	Gesamt	150	,020	,888	180	,044	1,221	180	,028	1,113
ZE1NORM	SPR1	50	,221	1,154	60	,246	,912	60	,040	1,054
	SPR2	50	,043	,993	60	-,065	1,064	60	,027	1,067
	SPR3	50	-,136	,967	60	-,099	1,221	60	-,136	1,085
	Gesamt	150	,043	1,044	180	,027	1,078	180	-,023	1,066
ZE2NORM	SPR1	50	-,131	1,102	60	-,036	1,039	60	,034	1,114
	SPR2	50	-,066	,973	60	-,021	,924	60	-,054	,984
	SPR3	50	,179	,959	60	-,050	,747	60	,178	1,125
	Gesamt	150	-,006	1,015	180	-,036	,906	180	,053	1,074
ZE3NORM	SPR1	50	-,164	1,043	60	-,203	,988	60	-,038	1,035
	SPR2	50	-,015	,900	60	-,206	,930	60	,071	1,162
	SPR3	50	-,155	,790	60	,023	1,129	60	,031	,878
	Gesamt	150	-,111	,913	180	-,129	1,019	180	,022	1,027
ZE4NORM	SPR1	50	-,272	1,005	60	-,171	,786	60	-,118	,542
	SPR2	50	,095	,906	60	,244	1,306	60	-,042	1,019
	SPR3	50	-,062	,893	60	,155	1,304	60	-,036	,919
	Gesamt	150	-,080	,942	180	,076	1,165	180	-,065	,848

**Tabelle F-1: Anzahl (n), Mittelwert ( $\bar{x}$ ) und Standardabweichung (s) für die sprecher- und lautnormierten Parameter des in Kapitel 9 verwendeten Datenmaterials (SPR1, SPR2, SPR3 steht für beurteilten Vokalrealisierungen der einzelnen Sprecher, gesamt für die beurteilten Vokalrealisierungen aller Sprecher).**



Laut	Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
/a:/	1	1,318	85,0	85,0	,754
	2	,233	15,0	100,0	,435
/i:/	1	1,116	77,4	77,4	,726
	2	,205	14,2	91,6	,413
	3	,121	8,4	100,0	,329
/u:/	1	1,507	66,6	66,6	,775
	2	,574	25,4	92,0	,604
	3	,135	6,0	98,0	,345
	4	,046	2,0	100,0	,210

**Tabelle F-2: Ergebnisse der Diskriminanzanalysen in Abhängigkeit der Vokalkategorien. Diskriminanzfunktionen, Eigenwerte, Erklärung der Varianz (% und kumulierte %) und kanonische Korrelationen.**

Laut		Funktion			
		1	2	3	4
/a:/	ZLAUTDUR	-,970	,186		
	ZF1	-,281	-,091		
	ZE1NORM	,205	,663		
	ZE2NORM	-,190	-,565		
	ZF3	-,051	-,466		
	ZF2	,168	-,298		
	ZE3NORM	-,039	-,258		
	ZMEANF0	,192	-,236		
/i:/	ZE2NORM	,725	-,381	-,234	
	ZF2	-,666	,189	-,097	
	ZF1	,533	,006	-,354	
	ZLAUTDUR	-,526	-,499	-,484	
	ZF3	-,493	-,355	-,178	
	ZMEANF0	,107	,348	-,221	
	ZE3NORM	-,001	,517	-,711	
	ZE1NORM	-,011	-,173	,304	
/u:/	ZE4NORM	-,196	,016	,229	
	ZLAUTDUR	-,784	-,233	,460	,129
	ZE1NORM	-,702	,139	-,406	-,023
	ZE2NORM	,688	-,388	,237	,208
	ZE3NORM	,585	,350	,349	-,143
	ZE4NORM	,291	,100	,551	-,236
	ZF1	,342	-,256	,517	-,170
	ZF3	,031	-,108	,395	-,246
/u:/	ZF2	,452	,346	,469	,510
	ZMEANF0	,148	,086	,205	-,329

**Tabelle F-3: Korrelationen zwischen den Diskriminanzvariablen und den standardisierten kanonischen Diskriminanzfunktionen in Abhängigkeit der Vokalkategorie.**

	Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
SPR1	1	5,211	83,0	83,0	,916
	2	,624	9,9	92,9	,620
	3	,443	7,1	100,0	,554
SPR2	1	,905	83,9	83,9	,689
	2	,116	10,8	94,7	,323
	3	,057	5,3	100,0	,232
SPR3	1	1,398	71,6	71,6	,764
	2	,499	25,5	97,1	,577
	3	,057	2,9	100,0	,233

**Tabelle F-4: Ergebnisse der Diskriminanzanalysen für das sprecherspezifische Datenmaterial der Vokalkategorie /i:/. Diskriminanzfunktionen, Eigenwerte, Erklärung der Varianz (% und kumulierte %) und kanonische Korrelationen.**

		Funktion		
		1	2	3
SPR1	ZF2	-,545	,058	,111
	ZE2NORM	,521	,247	,211
	ZF1	,441	,242	-,096
	ZE3NORM	-,006	,830	-,419
	ZE1NORM	-,056	-,566	,373
	ZF3	-,211	,074	,599
	ZLAUTDUR	-,257	,136	,575
	ZE4NORM	-,131	-,078	-,330
SPR2	ZMEANF0	-,029	,133	-,201
	ZF2	-,725	,476	,368
	ZLAUTDUR	-,607	-,026	-,445
	ZE2NORM	,555	-,453	-,170
	ZF3	-,413	,259	,406
	ZE1NORM	,396	,388	,195
	ZE3NORM	-,249	,659	-,149
	ZE4NORM	-,460	-,613	-,058
SPR3	ZF1	,404	,192	-,431
	ZMEANF0	,163	,291	-,324
	ZE2NORM	,753	,498	,085
	ZF3	-,548	,385	-,003
	ZLAUTDUR	-,454	,348	,377
	ZF1	,328	,032	,251
	ZF2	-,246	-,214	-,069
	ZE1NORM	-,160	,243	-,024
ZMEANF0	,260	-,263	,542	
ZE3NORM	,167	-,229	,450	
ZE4NORM	-,073	-,260	-,362	

**Tabelle F-5: Korrelationen zwischen den Diskriminanzvariablen und den standardisierten kanonischen Diskriminanzfunktionen für das sprecherabhängige Datenmaterial der Vokalkategorie /i:/.**

	Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
SPR1	1	2,495	52,8	52,8	,845
	2	1,486	31,5	84,3	,773
	3	,662	14,0	98,3	,631
	4	,081	1,7	100,0	,274
SPR2	1	1,360	61,8	61,8	,759
	2	,675	30,7	92,4	,635
	3	,136	6,2	98,6	,346
	4	,031	1,4	100,0	,173
SPR3	1	1,858	60,9	60,9	,806
	2	,852	27,9	88,8	,678
	3	,251	8,2	97,0	,448
	4	,091	3,0	100,0	,289

**Tabelle F-6: Ergebnisse der Diskriminanzanalysen für das sprecherspezifische Datenmaterial der Vokalkategorie /u/. Diskriminanzfunktionen, Eigenwerte, Erklärung der Varianz (% und kumulierte %) und kanonische Korrelationen.**

		Funktion			
		1	2	3	4
SPR1	ZLAUTDUR	,713	-,312	,326	,410
	ZE2NORM	-,712	-,073	,271	,214
	ZE1NORM	,701	-,109	-,381	-,090
	ZE3NORM	-,501	,514	,325	-,030
	ZF3	-,013	,000	,546	,117
	ZE4NORM	-,361	-,047	,398	-,126
	ZF1	-,234	,133	,324	,053
	ZF2	-,296	,507	,001	,625
SPR2	ZLAUTDUR	,765	,388	-,389	,306
	ZE2NORM	-,740	-,104	-,279	,138
	ZE1NORM	,700	,397	,468	-,152
	ZE3NORM	-,451	-,676	-,196	,391
	ZF1	-,309	,364	-,142	-,180
	ZE4NORM	-,369	-,370	-,741	-,166
	ZMEANF0	-,308	,055	,102	,481
	ZF2	-,228	-,168	-,092	-,401
SPR3	ZF3	,016	-,026	-,038	-,340
	ZF2	-,718	-,149	,643	,121
	ZLAUTDUR	,628	,156	,336	-,588
	ZE3NORM	-,523	-,087	-,061	-,256
	ZE1NORM	,517	-,298	-,056	,258
	ZE2NORM	-,483	,467	,055	-,132
	ZF3	-,057	,142	,226	,050
	ZMEANF0	-,083	-,142	,199	-,132
ZE4NORM	-,184	-,215	,146	-,624	
ZF1	-,424	,359	,166	-,563	

**Tabelle F-7: Korrelationen zwischen den Diskriminanzvariablen und den standardisierten kanonischen Diskriminanzfunktionen für das sprecherabhängige Datenmaterial der Vokalkategorie /u/.**

	Funktion	Eigenwert	% der Varianz	Kumulierte %	Kanonische Korrelation
SPR1	1	1,674	68,0	68,0	,791
	2	,787	32,0	100,0	,664
SPR2	1	2,509	90,3	90,3	,846
	2	,269	9,7	100,0	,460
SPR3	1	1,634	87,6	87,6	,788
	2	,231	12,4	100,0	,433

**Tabelle F-8: Ergebnisse der Diskriminanzanalysen für das sprecherspezifische Datenmaterial der Vokalkategorie /a/. Diskriminanzfunktionen, Eigenwerte, Erklärung der Varianz (% und kumulierte %) und kanonische Korrelationen.**

		Funktion	
		1	2
SPR1	ZLAUTDUR	,790	-,476
	ZF3	-,386	-,359
	ZMEANF0	-,344	-,191
	ZE1NORM	,004	,826
	ZE2NORM	-,023	-,706
	ZF1	,330	-,501
	ZE3NORM	-,006	-,296
	ZE4NORM	,113	-,144
	ZF2	-,053	-,059
SPR2	ZLAUTDUR	,841	,065
	ZF2	-,214	-,009
	ZE3NORM	-,061	-,013
	ZE4NORM	,007	-,731
	ZF3	,104	-,366
	ZE1NORM	,088	,334
	ZMEANF0	-,081	,329
	ZE2NORM	-,083	-,215
ZF1	,012	-,043	
SPR3	ZLAUTDUR	,825	-,426
	ZF3	,301	,259
	ZF1	,261	-,175
	ZF2	-,281	,598
	ZMEANF0	-,189	,520
	ZE1NORM	-,356	-,427
	ZE2NORM	,369	,426
	ZE3NORM	,098	,338
ZE4NORM	-,101	-,325	

**Tabelle F-9: Korrelationen zwischen den Diskriminanzvariablen und den standardisierten kanonischen Diskriminanzfunktionen für das sprecherabhängige Datenmaterial der Vokalkategorie /a/.**

	df	/a/		/i/		/u/	
		F	Signifikanz	F	Signifikanz	F	Signifikanz
ZF1	2	.697	.499	.559	.573	.516	.598
ZF2	2	.876	.418	1.356	.260	.480	.620
ZF3	2	.584	.559	1.676	.190	.104	.901
ZLAUTDUR	2	1.717	.183	1.315	.271	.510	.602
ZMEANF0	2	.454	.636	.084	.920	.152	.859
ZE1NORM	2	1.472	.233	1.882	.155	.505	.604
ZE2NORM	2	1.297	.276	.015	.985	.709	.493
ZE3NORM	2	.412	.663	.996	.372	.171	.843
ZE4NORM	2	1.930	.149	2.141	.121	.175	.840

**Tabelle F-10: Ergebnis der ANOVA für das sprecherspezifische Datenmaterial.**

	/a:/				/i:/				/u:/			
	Levene	df1	df2	Signifikanz	Levene	df1	df2	Signifikanz	Levene	df1	df2	Signifikanz
ZF1	2.462	2	147	.089	2.503	2	177	.085	.913	2	177	.403
ZF2	3.829	2	147	.024	11.660	2	177	.000	.654	2	177	.521
ZF3	1.609	2	147	.204	.327	2	177	.721	.120	2	177	.887
ZLAUTDUR	1.097	2	147	.337	1.017	2	177	.364	.501	2	177	.607
ZMEANF0	.098	2	147	.907	.102	2	177	.903	.487	2	177	.615
ZE1NORM	.894	2	147	.411	.812	2	177	.445	.041	2	177	.959
ZE2NORM	1.050	2	147	.353	3.113	2	177	.047	.108	2	177	.898
ZE3NORM	1.242	2	147	.292	.294	2	177	.746	.758	2	177	.470
ZE4NORM	.693	2	147	.502	2.920	2	177	.057	8.122	2	177	.000

**Tabelle F-11: Test der Homogenität der Varianzen (Levene-Test) des sprecherabhängigen Datenmaterials.**

Christina Widera  
Wielandstr. 14  
53173 Bonn

## **Lebenslauf**

geboren am 16.05.1965 in Moers  
Eltern Lucian Widera und Margit Widera, geb. Gall  
Familienstand ledig  
Staatsangehörigkeit deutsch

## **Ausbildung**

<b>Schulausbildung</b>	1971-1972	Adolph-Reichwein-Schule in Moers
	1972-1974	Städtische Gemeinschaftsgrundschule in Moers
	1974-1975	Gemeinschaftsgrundschule in Rheinberg-Millingen
	1975-1985	Amplonius-Gymnasium in Rheinberg
		Abschluß: Allgemeine Hochschulreife (13.05.85)
<b>Berufsausbildung</b>	1986-1989	Ausbildung zur Kfz-Mechanikerin
		Abschluß: Gesellenprüfung (13.06.89)
<b>Studium</b>	WS 1989/90	Rheinisch-Westfälische Technische Hochschule Aachen
		Fachrichtung: Maschinenbau
	SS 1990	Rheinische Friedrich-Wilhelms-Universität Bonn
		Hauptfach: Allgemeine Sprachwissenschaft
		Nebenfächer: Kommunikationsforschung und Phonetik, Psychologie
	WS 1990/91- SS 1994	Rheinische Friedrich-Wilhelms-Universität Bonn
		Hauptfach: Kommunikationsforschung und Phonetik
		Nebenfächer: Allgemeine Sprachwissenschaft, Psychologie
	WS 1994/95	Auslandsstudium an der Universität Edinburgh
	SS 95-SS 97	Rheinische Friedrich-Wilhelms-Universität Bonn
		Hauptfach: Kommunikationsforschung und Phonetik
		Nebenfächer: Allgemeine Sprachwissenschaft, Psychologie
		Abschluß: Magistra Artium (02.07.1997)
	seit WS 97/98	Rheinische Friedrich-Wilhelms-Universität Bonn
		Angestr. Abschluß: Promotion
		(Betreuer: Prof. Dr. W. Hess)

## **Berufliche Praxis**

<b>Praktika</b>	7/1985-7/1986	HDW-Werbung, Moers
	5/96 –7/96	Siemens AG, München
		Zentralabteilung Forschung und Entwicklung
<b>Studentische Hilfskraft</b>	9/93-9/94	Institut für Kommunikationsforschung und Phonetik,
	4/95-4/96	Rheinische Friedrich-Wilhelms-Universität Bonn
	8/96-6/97	
<b>Wissenschaftliche Mitarbeiterin</b>	seit 7/97	Institut für Kommunikationsforschung und Phonetik,
		Rheinische Friedrich-Wilhelms-Universität Bonn
		Projektmitarbeit:
		- DFG-Projekt: Prosodiegenerierung für die Sprachsynthese
		- DFG-Projekt: Prosodiebeurteilung und Reduktionsanalyse
		- EU-Projekt: ISAEUS

Bonn, den 12.05.02







