

Zeitliche Alignierung der F0–Kontur als Funktion der Silbenstruktur im Deutschen

**Kristina Spranger
Matrikelnummer: 1840094**

**Studienarbeit
Oktober, 2001**

**Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart
Azenbergstr. 12
70174 Stuttgart**

Betreuer: PD Dr. Bernd Möbius

Inhaltsverzeichnis

1 Einleitung	1
2 Prosodie	3
2.1 Begriffsklärung	3
2.2 Suprasegmentale Merkmale	5
2.2.1 Quantität	5
2.2.2. Intensität	6
2.2.3 Intonation	7
2.3 Bedeutung von Prosodie	8
2.4 Intonationsmodelle	9
2.4.1 Tonsequenzmodelle	9
2.4.2 Superpositionsmodelle	9
2.4.3 Das IPO–Modell	10
2.4.4 Probleme bei der Intonationsmodellierung	10
3 Die Silbe als prosodische Einheit	11
3.1 Die phonologische Bedeutung von Silben	11
3.2 Der Aufbau von Silben	12
3.2.1 Sonorität	13
3.2.2 Vokalaffinität	15
4 Entwicklung des quantitativen Modells	16
4.1 Silbenstrukturklassen im Deutschen	16
4.1.1 Die Phonotaktik des Deutschen	17
4.1.2 Datenbasierte Ermittlung von Silbenstrukturklassen	17
4.2 Testtextkonstruktion mit optimaler Abdeckung	21

4.3 Sprachaufnahmen	26
4.4 Lautsegmentierung, F0-Messungen	27
4.4.1 Alignment nach Phonemen	27
4.4.2 F0-Messungen	28
4.4.3 Ermittlung der Ankerpunkte	29
4.5 Statistische Auswertung	32
4.5.1 Datengrundlage für die statistische Analyse	32
4.5.2 Varianzanalyse	33
4.5.3 Ergebnisse der Varianzanalyse	34
4.6 Entwicklung des Modells	37
4.6.1 Datengrundlage für das Modell	37
4.6.2 Lineares Modell für die Ankerpunkte und Parameterschätzung	38
4.6.3 Ergebnisse	39
4.6.4 Bewertung der Ergebnisse	41
5 Ausblick	45
Literaturverzeichnis	47

Kapitel 1

Einleitung

In jüngeren Arbeiten wurde der Einfluß der temporalen Struktur auf die tonalen Strukturen der Sprache nachgewiesen. In den sogenannten Alignment-Experimenten, z.B. [sant00], wurde gezeigt, daß die F0-Kontur eines Pitchakzents von der segmentalen Zusammensetzung der Silbe, von der Struktur der Akzentgruppe, von der Informationsstruktur der Äußerung und von positionellen Faktoren abhängt.

Im Rahmen dieser Studienarbeit sollte nun speziell der Einfluß der segmentalen Zusammensetzung der Silbe auf den F0-Verlauf von Akzenten im Deutschen untersucht werden.

Als Ergebnis der Arbeit soll ein Modell vorliegen, das die Oberflächenvariation der F0-Kontur von Akzenten einer bestimmten phonologischen Klasse aus der Silbenstruktur in quantitativer Form prädiziert.

Hierzu soll nun zunächst im zweiten Kapitel ein Überblick über das Feld gegeben werden, in dem sich die Arbeit bewegt, nämlich über die Prosodie. Es werden die verschiedenen, mit Prosodie in Zusammenhang stehenden Begriffe geklärt und die Bedeutung der Prosodiegenerierung für die Sprachsynthese erläutert. In diesem Zusammenhang werden dann auch Intonationsmodelle vorgestellt und die Notwendigkeit quantitativer Modelle unterstrichen.

Im dritten Kapitel dann wird die Silbe als prosodische Einheit vorgestellt, wobei zunächst die phonologische Bedeutung der Silbe besprochen wird, bevor der Aufbau von Silben allgemein und im Deutschen speziell behandelt wird.

Das vierte Kapitel schildert den praktischen Teil dieser Arbeit, nämlich die Ermittlung der Silbenstrukturklassen, die Konstruktion des Testtextmaterials,

die Durchführung der Sprachaufnahmen, die Lautsegmentierung und F0-Messungen, die statistischen Auswertungen und schließlich die Konstruktion des quantitativen Modells.

Im fünften Kapitel wird schließlich noch ein Ausblick gegeben, wie die Ergebnisse der Untersuchungen genutzt werden können, und wie das Modell in der Prosodiegenerierung Anwendung finden kann.

Abschließend möchte ich noch meinem Betreuer, Bernd Möbius, für die geduldige Unterstützung bei allen Problemen danken und Norbert Braunschweiler, der mir bei der Organisation und Durchführung der Sprachaufnahmen geholfen hat. Außerdem danke ich Edmilson Morais für die Zeit und Mühe, die er aufgebracht hat, um mir bei der statistischen Analyse und der Auswertung der Ergebnisse zu helfen.

Kapitel 2

Prosodie

In diesem Kapitel soll nun ein Überblick über den Gegenstandsbereich gegeben werden, in dem die Studienarbeit anzusiedeln ist, nämlich über die **Prosodie**. Nach einer Begriffsklärung werden die wichtigsten suprasegmentalen Phänomene beschrieben und die Bedeutung von Prosodie für die sprachliche Kommunikation besprochen. Abschließend werden dann noch die verschiedenen Arten von Intonationsmodellen und deren Aufgaben und Probleme diskutiert und die Wichtigkeit einer möglichen analytischen Trennung und quantitativen Modellierung des Einflusses unterschiedlicher Faktoren auf den F0–Verlauf betont.

2.1 Begriffsklärung

Prosodie wird von Hadumod Bußmann [buß90] als die Gesamtheit sprachlicher Eigenschaften wie **Akzent**, **Intonation**, **Quantität** und **Sprechpausen** definiert. Im allgemeinen beziehen sich diese Eigenschaften auf Einheiten, die größer sind als ein einzelnes Phonem. Ebenfalls zur Prosodie zählt nach [buß90] die Untersuchung von **Sprechgeschwindigkeit** und **Rhythmus**.

Zum Gegenstandsbereich der Prosodie gehören sowohl linguistische als auch paralinguistische Merkmalssysteme. Die Abbildung 2.1 gibt einen schematischen Überblick über diesen Gegenstandsbereich.

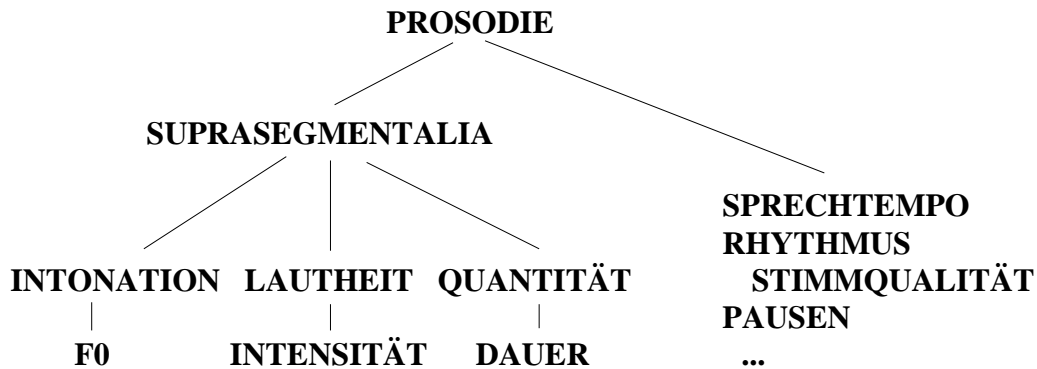


Abb. 2.1: schematischer Überblick über Prosodie (aus [möb93])

Der Begriff Prosodie bezieht sich nach Dutoit [dut97] auf bestimmte Eigenschaften des Sprachsignals, so wie z.B. hörbare Unterschiede in der Tonhöhe, der Lautstärke und der Silbendauer. Da eben die prosodischen Ereignisse eher mit Silben oder Silbengruppen zeitlich aligniert zu sein scheinen denn mit Segmenten (Laute, Phoneme), werden diese Ereignisse als **suprasegmentale Phänomene** bezeichnet.

Prosodische Ereignisse können auf verschiedenen Repräsentationsebenen untersucht werden [dut97]:

1. die **akustische** Ebene repräsentiert die akustische Manifestation der Prosodie (Grundfrequenz, Amplitude, Dauer)
2. die **perzeptuelle** Ebene repräsentiert die prosodischen Ereignisse als vom Hörer wahrgenommen
3. die **linguistische** Ebene repräsentiert die Prosodie einer Äußerung als Sequenz abstrakter Einheiten, von denen einige eine kommunikative Funktion in der Sprache haben, während andere nur syntaktische Anforderungen erfüllen

Die verschiedenen Repräsentationsebenen der Prosodie lassen sich wie in Tabelle 2.1 veranschaulichen.

<i>akustische Ebene</i>	<i>perzeptuelle Ebene</i>	<i>linguistische Ebene</i>
Grundfrequenz (F0)	Tonhöhe	Ton, Intonation, Betonung

<i>akustische Ebene</i>	<i>perzeptuelle Ebene</i>	<i>linguistische Ebene</i>
Amplitude, Energie, Intensität	Lautstärke	Betonung
Dauer	Länge	Betonung

Tab. 2.1: Repräsentationsebenen der Prosodie (aus [dut97])

2.2 Suprasegmentale Merkmale

Suprasegmentale Merkmale sind dadurch gekennzeichnet, daß sie die segmentalen Merkmale überlagern, aber nicht zeitlich auf diese begrenzt sind, so Lehiste [leh70]. Unter suprasegmentalen Merkmalen werden bei [leh70] die folgenden drei Merkmalsbereiche subsumiert:

1. **Quantität** – Phänomene, die mit der zeitlichen Dauer der Realisierung sprachlicher Einheiten zu tun haben
2. **Intensität** – hierunter wird alles verstanden, was unter die Kategorie Betonung und Akzentuierung fällt
3. **Intonation** – alle melodischen Phänomene, die sich in Form von Tonhöhenverläufen und –veränderungen in der Sprache äußern, also letztlich im Verlauf der Grundfrequenz F₀

Allen diesen Merkmalen ist gemeinsam, daß sie in der Schriftform der Sprache, wenn überhaupt, nur unzulänglich repräsentiert sind.

2.2.1 Quantität

Von den suprasegmentalen Eigenschaften ist die **Quantität** diejenige, die am engsten mit einzelnen Segmenten assoziiert ist. Sie ist laut [hess00] meist auf Einzellaute beschränkt, überlagert sich dort aber den qualitativen Eigenschaften. Jedes Segment muß, so [leh70], um als Segment erkennbar zu sein, eine gewisse Dauer besitzen. Aber genauso kann gerade diese Dauer kontrastiv sein, z.B. kann sie das betreffende Segment als besonders lang oder besonders kurz kennzeichnen. Diese kontrastive Dauer wird als Quantität bezeichnet.

Da Sprechen laut [nepp99] in der Zeit verläuft, ist die Zeit ein so wichtiges

Element sowohl zur Konstitution als auch zur Identifikation sprachlicher Einheiten, daß die Angabe anderer akustischer Daten wie der Grundfrequenz, der Lautstärke, der spektralen Struktur usw. oft nur dann sinnvoll ist, wenn gleichzeitig geklärt wird, wie sie sich zeitlich ändern.

Die **inhärente Dauer** von Lauten hängt nach [nepp99] von der Bildung des Lautes ab, d.h. davon, wie er artikuliert wird. Je nachdem, welche Stellung der betreffende Laut im Wort einnimmt, in welcher Art von Silbe er steht (offen oder geschlossen) und in welcher Stellung er in bezug auf den Akzent steht (ist die Silbe akzentuiert oder nicht), besitzt der Laut eine von diesen und natürlich noch anderen Bedingungen abhängige charakteristische Dauer.

Ein Laut ist in einer akzentuierten Silbe von größerer Dauer als in einer nicht akzentuierten. Dies ist die sogenannte **Akzentdauer**, siehe [nepp99].

2.2.2 Intensität

Auf der segmentalen Ebene äußert sich die **Intensität** durch die Wortbetonung, [leh70]. In einem zwei- oder mehrsilbigen Wort können grundsätzlich alle Silben gleichstark, also mit gleicher Intensität, realisiert werden. In der Regel sind sie jedoch unterschiedlich gewichtet, wobei eine Silbe besonders stark hervortritt.

In [nepp99] wird eine Hervorhebung auf Wortebene als **Akzent** bezeichnet. Die **(Silben-)Betonung** ist die akustische Realisierung eines lexikalischen Wortakzents. Die **Akzentsilbe** oder einfach nur der **Akzent** ist eine tonal durch die Grundfrequenz betonte Silbe.

In Sprachen mit **beweglichem Akzent**, wie Deutsch, Russisch, Spanisch und Japanisch, ist der Akzent potentiell linguistisch funktionell, d.h. durch seine Stellung kann er Wortbedeutungen unterscheiden, wie z.B. **August** vs. **August**.

In Sprachen, die einen **festen Akzent** besitzen, d.h. einen Akzent, der immer auf die der Ordnungszahl nach gleiche Silbe fällt, erfüllt er lediglich eine delimitative Funktion.

Das weitaus wichtigste Element der Akzentuierung ist in Sprachen wie Deutsch, Französisch oder Englisch die Erhöhung der Grundfrequenz auf der akzentuierten Silbe.

Welche Rolle der Schallpegel bei der Akzentuierung spielt, ist von Sprache zu Sprache sehr verschieden. Im Falle einer zum Zweck der Betonung hervorgebrachten Lautheitsänderung spricht man auch von **dynamischer Akzentuierung**.

In einigen Sprachen, wie z.B. Englisch oder Französisch, ist die Veränderung

der Zeitrelationen und damit der subjektiven Dauer ein wichtiges Element der Akzentuierung.

In Sprachen wie Russisch oder Englisch ist die Bedeutung der von der spektralen Struktur bestimmten Vokalqualität für die Akzentwahrnehmung sehr groß. Auch im Deutschen hat die Vokalqualität eine gewisse Bedeutung für die Akzentuierung, so kommt z.B. das Schwa nur in unbetonten Silben vor.

Spezialfälle des Akzents sind der sogenannte **Kontrastakzent** und die **Emphase**.

2.2.3 Intonation

Wie kein anderer akustischer Sprachsignalparameter, so [hess00], ist die **Grundfrequenz** funktional belastet. Die Mikroprosodie, d.h., im wesentlichen unbeeinflussbare artikulatorische Auswirkungen, der Wortakzent, der Satzakzent, sprecherspezifische und emotionale Randbedingungen – alle greifen sie auf diesen Parameter durch.

Intonation umfaßt nach Möbius [möb93] die linguistisch relevanten Eigenschaften eines Grundfrequenzverlaufs. D.h. unter Intonation werden die linguistisch relevanten Funktionen der Grundfrequenz auf Silben-, Wort- und Äußerungsebene zusammengefaßt.

So wird laut [nepp99] eine nicht näher bestimmte Tonhöhenänderung beim Sprechen allgemein als **Sprechmelodie** bezeichnet, wohingegen die in sprachlicher Funktion auf bestimmte Äußerungseinheiten bezogene Melodieeinheit speziell als **Intonation** bezeichnet wird. Im Gegensatz zum sprachlich distinktiven **Ton**, welcher integrierender Teil der lexikalischen Bedeutung ist, beeinflußt die Intonation die lexikalische Bedeutung nicht. Die hier beschriebene Variation in der Tonhöhe ist das am deutlichsten wahrzunehmende von den prosodischen Ereignissen. Diese Unterschiede in der Tonhöhe bilden über die Zeit betrachtet die **Pitchkontur** einer Äußerung. Folgende Elemente werden in [nepp99] als intonationskonstituierend genannt:

- **Tonhöhe** entsprechend der **Grundfrequenz**
- **Lautstärke** entsprechend dem **Schallpegel**
- **subjektive Dauer** entsprechend der **zeitlichen Länge**
- **Pause**
- **Tempo**
- **Stimmqualität**

- **Emphase**

Viele dieser Eigenschaften sind allerdings hinsichtlich der Bedeutungszuordnung verhältnismäßig vage, da sie in der Wahrnehmung mehr oder weniger stark ausgeprägt unterschiedlichen situativen, emotionalen und assoziativen Deutungen unterliegen, so [nepp99].

Akustisch besteht eine Intonationseinheit laut [nepp99] normalerweise aus einer am Anfang der jeweiligen Ausdruckseinheit steigenden Grundfrequenz, die sich dann zwischen zwei gedachten Linien, einer Basislinie und einer Dachlinie, hin- und herbewegt, wobei die obere Linie das höchste und die untere das niedrigste Frequenzniveau angibt. Beide Begrenzungslinien fallen langsam ab. Nach jeder solchen Ausdruckseinheit wird eine Pause eingeschoben und die Basislinie wird dann gleich neu gesetzt. Es gibt natürlich verschiedenste Abweichungen. Bei der Beobachtung vieler Intonationsabläufe über einen langen zeitlichen Verlauf ergibt sich jedoch in stilisierter Form das beschriebene Schema.

Jeder Sprecher weist normalerweise eine gewisse individuelle Variationsbreite der Grundfrequenz seines Stimmklangs auf, innerhalb der seine Intonationsmuster entstehen.

2.3 Bedeutung von Prosodie

Prosodische Merkmale haben in sprachlicher Kommunikation spezifische Funktionen. Sie strukturieren das Gesagte, markieren den Fokus und bestimmen die Satzabsicht (Frage, Aussage, Befehl). Außerdem helfen prosodische Merkmale, Emotionen zu übermitteln.

Darüberhinaus besitzt die Prosodie, wie Levelt [lev89] beschreibt, eine integrierende Funktion in der Organisation und in der Produktion der Lautsprache. Semantische Information, syntaktische und morphologische Struktur und segmentale Sequenzen werden in ein konsistentes strukturelles Geflecht eingebettet. Eine zentrale Bedeutung kommt hierbei der temporalen Steuerung der Sprache zu, sowohl in bezug auf Intonation als auch in bezug auf Dauer.

2.4 Intonationsmodelle

Da synthetisierte Sprache oftmals in einem solchen Maße unnatürlich und monoton klingt, daß es viele Hörer als unangenehm empfinden, besteht in der Sprachsynthese eine offensichtliche Notwendigkeit der Prosodiegenerierung. Zum Zwecke der Prosodiegenerierung werden in der Sprachsynthese die sogenannten **Intonationsmodelle** eingesetzt.

Da F0-Alignment-Experimente, wie z.B. [sant00], gezeigt haben, daß F0-Konturen als akustische Realisierungen von Akzenten stark in Abhängigkeit von der Struktur der Silben, mit denen sie assoziiert sind, und in Abhängigkeit von der Struktur der Akzentgruppen (mono- vs. polysyllabisch) variieren, wobei Alignment die zeitliche Alignierung der F0-Kontur mit dem segmentalen Material bezeichnet, und außerdem empirische phonetische Beobachtungen ergeben haben, daß F0-Gipfelposition und die Kontur von Pitchakzenten von der Dauer und Struktur der Akzentgruppe abhängig ist, [möb00], sollten die Intonationsmodelle nach [möb00] so viele Details über den Grundfrequenzverlauf generieren wie möglich.

Die Aufgabe solcher Intonationsmodelle besteht in der Berechnung von F0-Konturen aus einer phonologischen Repräsentation der Prosodie, wobei der Einfluß von Sprachlauten, Silben und Akzentgruppen und ihren Dauern berücksichtigt werden muß.

Man unterscheidet zwei Arten von Intonationsmodellen, die die beiden verschiedenen Möglichkeiten zur Interpretation der Struktur der Intonation widerspiegeln:

Die **Tonsequenzmodelle** und die **Superpositionsmodelle**.

2.4.1 Tonsequenzmodelle

Bei den **Tonsequenzmodellen** [pier80] wird angenommen, daß die Intonation aus einer linearen Sequenz tonaler Elemente, der sogenannten **Pitchakzente** besteht. Die F0-Konturen werden aus Sequenzen phonologisch distinktiver Töne generiert. Die Töne sind dabei lokal bestimmt und interagieren nicht.

2.4.2 Superpositionsmodelle

Bei den **Superpositionsmodellen** wird die Intonation laut [möb00] als hierarchisch organisiert betrachtet. Die Komponenten kleinerer zeitlicher

Domänen sind denen größerer zeitlicher Domänen superponiert. Bei dieser Art von Intonationsmodell treten die Komponenten simultan auf und interagieren hochgradig. Superponierende Intonationsmodelle ermöglichen eine analytische Trennung und quantitative Modellierung des Einflusses unterschiedlicher Faktoren auf den F0-Verlauf. Und gerade mit Hilfe solcher quantitativen Modelle lassen sich phonologische Hypothesen implementieren und via Synthese perceptiv überprüfen.

2.4.3 Das IPO-Modell

Neben diesen beiden Arten von Intonationsmodellen gibt es noch das **IPO-Modell**, das sich auf die perceptiven Aspekte der Prosodiemodellierung bezieht. Das IPO-Modell beschreibt und inventarisiert nur perceptiv relevante F0-Bewegungen, [möb00]. Diese Bewegungen sind entweder prominenzverleihend oder nicht. Die Intonationskontur besteht im IPO-Modell aus einer linearen Sequenz diskreter intonatorischer Elemente, F0-Bewegungen also, aber nicht Tönen. Eine Grammatik beschreibt zulässige Kombinationen dieser Elemente.

2.4.4 Probleme bei der Intonationsmodellierung

Ein zentrales Problem für Intonationsmodelle besteht nach [möb01] darin, daß sich linguistische, paralinguistische und sprecherspezifische Faktoren alle in demselben akustischen Parameter ausdrücken, in der Grundfrequenz. Das bedeutet, daß der Verlauf der Grundfrequenz intrinsische und koartikulatorische F0-Variationen beinhaltet, genauso wie lexikalische Töne und Silbenbetonung bzw. Wortakzent, prosodische Phrasen, Informationen zum Satzmodus, Paragraphen- und Textintonation und nicht zuletzt auch sprecherspezifische Faktoren.

Die Frage, die sich nun stellt, ist, wie sich diese Effekte trennen lassen. Mit Hilfe von Messungen oder mit Hilfe perceptiver Untersuchungen lassen sie sich nicht trennen. Vielleicht gibt es aber auf einer abstrakten linguistischen Ebene Möglichkeiten. Man kann Regeln formulieren, die z.B. Akzent- oder Phrasenmuster vorhersagen, entweder unabhängig voneinander oder auch unter gegenseitiger Beeinflussung. Diese Aufgabe spricht für Modelle, die die Faktoren mit ihren jeweils eigenen, unterschiedlichen zeitlichen Domänen direkt repräsentieren, so [möb01].

Kapitel 3

Die Silbe als prosodische Einheit

Nachdem im vorigen Kapitel ein Einblick in die Prosodie gegeben wurde, um so das thematische Umfeld der Studienarbeit näher zu beleuchten, bietet dieses Kapitel eine Beschreibung der **Silbe**. Die Silbe ist nämlich die Grundlage der Untersuchungen und Auswertungen, die im Rahmen der Studienarbeit stattgefunden haben.

Zunächst wird die phonologische Bedeutung von Silben besprochen, um so die Wahl der Silbe als prosodische Einheit zu motivieren, dann wird der Aufbau von Silben allgemein und im Deutschen speziell beschrieben. In diesem Zusammenhang werden die Begriffe Sonorität und Vokalaffinität eingeführt.

3.1 Die phonologische Bedeutung von Silben

Die Silbe ist nach Wiese [wie96] eine der wahrscheinlich relevantesten prosodischen Einheiten einer Sprache. Im Deutschen ist sie eine der offensichtlichsten und hervorstechendsten.

In [wie96] werden verschiedene Aspekte, die die sehr bedeutende und klar definierte Rolle der Silbe für eine systematische phonologische Beschreibung unterstreichen, aufgeführt. Phänomene, wie z.B. Aspiration, Auslautverhärtung und Allophonie bei dem dorsalen Frikativ, finden nur innerhalb einer Silbe statt. Auch phonologische Gesetzmäßigkeiten sind in hohem Maße mit dem Konzept der Silbe verbunden. Silben stellen den Bereich dar, in dem Aussagen über die Phonotaktik des Deutschen, also über die Beschränkungen in der

Kombinierbarkeit von Phonemen, Gültigkeit besitzen. Über Wortgrenzen hinweg kann jeder beliebige Laut mit jedem Laut kombiniert werden, und auch innerhalb eines Wortes, also über Silbengrenzen hinweg, können Kombinationen auftreten, die sonst als nicht möglich angesehen werden. Innerhalb von Silben können solche Kombinationen aber eben nicht auftreten. Darüberhinaus spielen Silben in vielen Bereichen der linguistischen Performanz eine prominente Rolle. So finden z.B. Kinder die Aufteilung der Sprache in Silben einfach und natürlich, andere Aufteilungen (z.B. in Morpheme) viel schwieriger. Außerdem werden Silben oft als rhythmische Einheiten verwendet. Obwohl die Silbe ein intuitiv klarer Begriff ist, ist sie schwer zu definieren. Sprecher des Deutschen können sich zwar (normalerweise) darauf einigen, aus wie vielen Silben ein Wort besteht, die präzise Position der Silbengrenzen zu bestimmen, ist oftmals schon schwieriger. Auch wenn die Silbenstruktur auf der Basis von segmentaler Information vorhersagbar ist, gibt es einige Fälle mit einem gewissen Grad an Mehrdeutigkeit, siehe [hess00]. So können zum Beispiel Wörter mit intervokalischen Konsonanten nicht immer eindeutig syllabiert werden (z.B. [hYp-f@n] vs. [hY-pf@n]). In solchen Fällen werden die betreffenden Konsonanten oft als ambisyllabisch, also als zu zwei Silben zugehörig, klassifiziert.

3.2 Der Aufbau von Silben

Laut Wiese [wie96] ist eine Silbe eigentlich nicht mehr als eine phonologische Einheit, die um einen silbischen Peak (normalerweise einen Vokal) herum organisiert ist. Der silbische Peak, also der Nukleus, ohne den die Silbe nicht existieren kann, besteht im Deutschen aus einem Vokal oder Diphthong. Ist die dem Silbennukleus vorausgehende Anfangskonsonantenfolge leer, so geht im Deutschen dem Silbennukleus in diesem Fall (fast) immer ein Glottalverschluß voraus. Ist die auf den Nukleus folgende Endkonsonantenfolge leer, so sprechen wir von einer offenen, sonst von einer geschlossenen Silbe.

Der Aufbau einer Silbe kann schematisch wie in Abbildung 3.1 dargestellt werden.

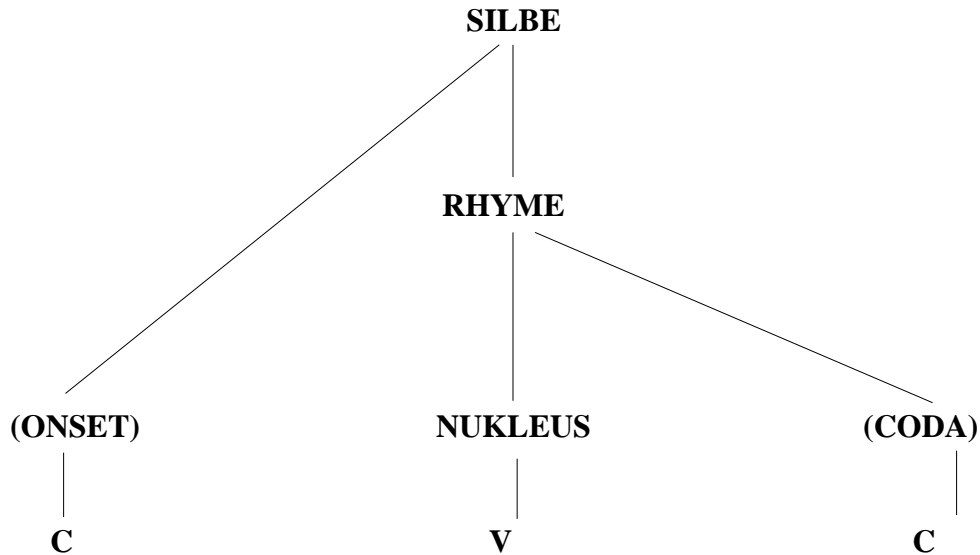


Abb. 3.1: Silbenaufbau (aus [wie96])

Das hier präsentierte Modell einer Silbe ist ein Basismodell, auf dessen Grundlage die Fragen der phonotaktischen Möglichkeiten und Einschränkungen diskutiert werden können. Für die möglichen Phonemkombinationen im silbeninitialen und silbenfinalen Konsonantencluster, dem Onset und der Coda, existieren strikte Beschränkungen.

3.2.1 Sonorität

In den meisten Versuchen, die Silbenstruktur und die Beschränkungen in Onset und Coda zu charakterisieren, findet man den Begriff der **Sonorität** oder einer **Sonoritätshierarchie**, siehe [jes04] und [siev01]. Dieses Konzept gründet auf der Beobachtung, daß Silben so aufgebaut sind, daß die Segmente an den Silbenrändern, die Marginalphoneme, eine relativ geringe inhärente Lautstärke besitzen, wie z.B. /p/, während die Segmente im Silbenkern eine sehr viel größere inhärente Lautstärke aufweisen, wie z.B. /a/. Segmente, die zwischen den Silbenrändern und dem Silbenkern liegen, liegen auch in bezug auf die inhärente Lautstärke zwischen den beiden Extremen, wie z.B. /l/. Diese phoneminhärente Eigenschaft des 'Lautseins' wird als **Sonorität** bezeichnet.

So sind sowohl [pla] als auch [alp] mögliche Silben des Deutschen, wohingegen

*[lpa] oder *[apl] keine sind.

Faßt man Plosive und Frikative zur Klasse der Obstruenten zusammen, so kann eine **Sonoritätshierarchie** fürs Deutsche nach [wie96] wie in Abbildung 3.2 aussehen, wobei die Richtung des Pfeils steigende Sonorität anzeigt.

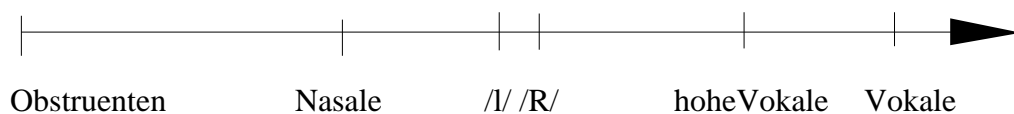


Abb. 3.2: Sonoritätshierarchie fürs Deutsche (aus [wie96])

Selkirk [sel84:116] hat aufbauend auf dieser Hierarchie die sogenannte **Sonoritätsabfolge–Generalisierung** wie folgt ausgedrückt:

(1) Sonoritätsabfolge–Generalisierung:

In jeder Silbe ist ein Segment Sonoritätsgipfel, vor und nach diesem Segment kommt eine Segmentfolge mit progressiv abnehmender Sonorität.

Nach der Sonoritätsabfolge–Generalisierung stellt also zum Beispiel [kalb] eine korrekte Silbe dar, weil die Sonorität zum Gipfel, dem Nukleus also, hin zunimmt und nach dem Gipfel progressiv abnimmt.

Die Sonoritätshierarchie und die Abfolge–Generalisierung sagen im Prinzip aus, daß es eine spiegelbildliche Verteilung der Obstruenten und Sonoranten in Onset und Coda gibt (Tabelle 3.1).

<i>Onset</i>	<i>Coda</i>
/kn/: Knie	/nk/: krank
/kr/: krumm	/rk/: Mark
/fl/: Flaum	/lf/: Wolf

Tab. 3.1: Beispiele für die spiegelbildliche Verteilung von Obstruenten und Sonoranten in Onset und Coda (aus [vat92], [wur70])

3.2.2 Vokalaffinität

Eng verbunden mit dem Begriff der Sonorität ist der Begriff der **Vokalaffinität**. Die Vokalaffinität beschreibt nach Hess [hess00] die Neigung der Vokale, bestimmte Konsonanten an sich zu binden. Allgemein gilt, je vokalähnlicher ein Konsonant klingt, desto größer ist seine Vokalaffinität. In fallender Reihenfolge der Vokalaffinität erhalten wir die in Abbildung 3.3 dargestellten Gruppierungen der Konsonantenklassen.

Liquide und Glides —————▶ **Nasale** —————▶ **Frikative und Plosive**

Abb. 3.3: Konsonantenklassen in fallender Reihenfolge der Vokalaffinität (aus [hess00])

Demnach existieren im Deutschen Konsonantenfolgen wie (initial) /fl-/, aber nicht */lf-/ oder (final) /-lt/, aber nicht */-tl/. Auch innerhalb der einzelnen Gruppen gibt es Hierarchien, so hat z. B. /r/ im Deutschen eine höhere Vokalaffinität als /l/ (Beispiel: /karl/, aber nicht */kalr/).

Kapitel 4

Entwicklung des quantitativen Modells

In diesem Kapitel wird nun die konkrete Entwicklung eines quantitativen Modells beschrieben, das die Oberflächenvariation der F0-Kontur von Akzenten aus der Silbenstruktur prädiziert.

Dazu wird zunächst in Kapitel 4.1 aufbauend auf dem in Kapitel 3 vermittelten Überblick über die Silbenstruktur im Deutschen nach einer kurzen Einführung in die Phonotaktik des Deutschen (4.1.1) die datenbasierte Ermittlung von Silbenstrukturklassen beschrieben (4.1.2). Ausgehend von den in 4.1 ermittelten Silbenstrukturklassen wird dann die Erstellung des Testtextmaterials beschrieben (4.2) und die Durchführung der Sprachaufnahmen (4.3) besprochen. In Kapitel 4.4 werden verschiedene Arbeitsschritte, wie Lautsegmentierung und F0-Messung, vorgestellt, die auf den nun in digitaler Form zur Verfügung stehenden Sprachdaten durchgeführt wurden, um eine Datengrundlage für die beabsichtigten statistischen Auswertungen (4.5) zu bekommen. In Kapitel 4.6 werden dann die Datengrundlage des quantitativen Modells (4.6.1), das lineare Modell der Ankerpunkte und die Parameterschätzung (4.6.2) und die Ergebnisse der Schätzungen (4.6.3) diskutiert. In Kapitel 4.6.4 werden die Resultate abschließend bewertet.

4.1 Silbenstrukturklassen im Deutschen

Um den Einfluß der segmentalen Zusammensetzung der Silbe auf den F0-Verlauf von Akzenten im Deutschen untersuchen zu können, mußten zunächst

relevante Klassen von Silbenonsets, –nuklei und –codas ermittelt werden.

4.1.1 Die Phonotaktik des Deutschen

Bei der Ermittlung relevanter Onset-, Nukleus- und Codaklassen müssen verschiedene Aspekte berücksichtigt werden. Der wichtigste ist die Frage der Kombinierbarkeit von Phonemen in ihrer linearen Abfolge, also die **Phonotaktik**, die in Form von i. a. sehr mächtigen Distributionsbeschränkungen in die ansonsten freie Kombinierbarkeit von Phonemen eingreift.

Hess [hess00] bemerkt, daß die Distributionsbeschränkungen in Anfangs- und Endkonsonantenfolgen sehr mächtig sind. Nur ein sehr geringer Bruchteil der kombinatorisch möglichen Konsonantenfolgen kommt in einer Sprache tatsächlich vor. Verglichen mit anderen Sprachen ist die Phonotaktik des Deutschen als komplex einzustufen, so [hess00]. Silbeninitial können bis zu drei, silbenfinal bis zu fünf (berücksichtigt man auch Kontraktionen sogar bis zu sechs) Konsonanten hintereinander auftreten. Damit wären im Deutschen bei völliger Freiheit der Konsonantenkombinationen bei rund 20 Einzelkonsonanten $20 \times 20^3 = 8000$ Anfangskonsonantenfolgen und 20×20^5 bzw. 20×20^6 (wenn auch Kontraktionen berücksichtigt werden) Endkonsonantenfolgen möglich. Die Phonotaktik des Deutschen reduziert diese Zahlen deutlich.

4.1.2 Datenbasierte Ermittlung von Silbenstrukturklassen

Die in dieser Arbeit als Grundlage für die Ermittlung der Silbenstrukturklassen dienende Datenbank 'CELEX' umfaßt ca. 70000 Lemmata (http://graugans.ims.unistuttgart.de/cd/Celex_Lexical_Database/gpl/gpl.cd). 534 von 1350 phonotaktisch möglichen Silbentypen treten in dieser Datenbank tatsächlich auf.

Um ausgehend von den Einträgen in 'CELEX' Silbenstrukturklassen zu ermitteln, mußten die Lexikoneinträge zunächst in ihre monosyllabischen Wortbestandteile aufgespalten werden. Auf den monosyllabischen Wortbestandteilen sind dann folgende Substitutionen durchgeführt worden, um von den konkreten Lemmata zu abstrahieren und so einen ersten Überblick über die Beschaffenheit der Silben im Deutschen zu gewinnen:

Vokale:

langer Vokal → L

kurzer Vokal → K

Diphthong, langer oder kurzer Vokal + postvokalisches R → D

Schwa → @

Konsonanten:

Liquide → Q

Glides → G

Nasale → N

stimmlose Plosive → P

stimmhafte Plosive → B

stimmlose Frikative → F

stimmhafte Frikative → Z

Die nunmehr nur noch aus den die verschiedenen Vokal- bzw. Konsonantenklassen repräsentierenden Symbolen zusammengesetzten monosyllabischen Wortbestandteile (z.B. 'Bein' → BDN) sind dann jeweils in Onset, Nukleus und Coda aufgespalten worden. Bei den so entstandenen Listen von Onsets, Nuklei und Codas wurden Mehrfachvorkommen zusammengefaßt und absolute Auftretenshäufigkeiten ermittelt. Sehr seltene und daher als unwahrscheinlich anzusehende Vorkommen wurden aus den jeweiligen Listen entfernt.

Die so bereinigten Daten sind weiter zusammengefaßt worden bis jeweils vier aus ähnlichen Phonemen bzw. Phonemkombinationen bestehende Onset-, Nukleus- und Codaklassen ermittelt waren. Hierzu wurden die Plosive und Frikative unabhängig von ihrer Stimmhaftigkeit zur Klasse der Obstruenten, die Liquide, Glides und Nasale zur Klasse der Sonoranten zusammengefaßt. Die Tabellen 4.1 – 4.3 stellen die so entstandenen Onset-, Nukleus- und Codaklassen vor.

<i>Onsetklasse:</i>	<i>Beschreibung:</i>	<i>Elemente:</i>
<i>11</i>	1 Obstruent	{P, B, F, Z}
<i>12</i>	1 Sonorant	{Q, G, N}

<i>Onsetklasse:</i>	<i>Beschreibung:</i>	<i>Elemente:</i>
13	1 – 2 Obstruenten + 1 Sonorant	{PQ, BQ, FQ, ZQ, {PFQ, FPQ}, PN, BN, FN}
14	1 – 3 Obstruenten	{PF, FP, PZ, FZ, PFZ}

Tab. 4.1: Onsetklassen 11, 12, 13 und 14

<i>Nukleusklasse:</i>	<i>Beschreibung:</i>	<i>Elemente:</i>
K	Kurzvokal	{/a/, /I/, /U/, /O/, /E/, /Y/, /9/}
L	Langvokal	{/a:/, /i:/, /u:/, /o:/, /e:/, /E:/, /y/ /2/}
D	Diphthong, Kurz- oder Langvokal + postvokalisches /R/	{/aI/, /aU/, /OY/, /KR/, /LR/}
@	Schwa, Schwa + postvokalisches /R/	{/@/, /@R/}

Tab. 4.2: Nukleusklassen K, L, D und @

<i>Codaklasse:</i>	<i>Beschreibung:</i>	<i>Elemente:</i>
21	1 Obstruent	{P, F}
22	1 Sonorant	{/l/, N}

<i>Codaklasse:</i>	<i>Beschreibung:</i>	<i>Elemente:</i>
23	2 – 4 Obstruenten	{PP, FF, {PF, FP}, {PFP, PPF, FPP, PFF, FFP, FPF}, {PFPF, PFFP, FPFP, PPFPP}}
24	2 Sonoranten oder 2 Sonoranten + 1 – 2 Obstruenten oder 1 Sonorant + 1 – 4 Obstruenten	{{NF, QF}, {NP, QP}, QN, QNP, QNF, {NPF, NFP, QPF, QFP, NPP, QPP, NFF, QFF}, QNFP, {NFPF, NPPF, QPFP, NFPF, NFFP, NPPF, QPFP, QFFP, NPFFP, NFPFP, NPPFP}}

Tab. 4.3: Codaklassen 21, 22, 23 und 24

Die in den Tabellen 4.1 und 4.3 vorgestellten Onset- bzw. Codaklassen (11 .. 14 und 21 .. 24) repräsentieren die acht Silbenstrukturklassen, die im weiteren Verlauf der Arbeit unterschieden werden. Die einzelnen Elemente der Onset- bzw. Codaklassen wurden dann noch mit den in Tabelle 4.2 vorgestellten Elementen der vier verschiedenen Nukleusklassen (K, L, D, @) kombiniert, um die im Deutschen auftretenden Silben möglichst vollständig abzudecken. Um aber der Gefahr der kombinatorischen Explosion vorzubeugen, wurden nur die Elemente der Onsetklassen systematisch mit denen der Nukleusklassen kombiniert und die Elemente der Nukleusklassen systematisch mit denen der Codaklassen (Tab. 4.4 und 4.5).

Onset + Nukleus:

- (11) 11K, 11L, 11D, 11@
- (12) 12K, 12L, 12D, 12@
- (13) 13K, 13L, 13D, 13@

(14) 14K, 14L, 14D, 14@

Tab. 4.4: Onset + Nukleus–Paare

Nukleus + Coda:

- (21) K21, L21, D21, @21
- (22) K22, L22, D22, @22
- (23) K23, L23, D23, @23
- (24) K24, L24, D24, @24

Tab. 4.5: Nukleus + Coda–Paare

4.2 Testtextkonstruktion mit optimaler Abdeckung

Als die Silbenstrukturklassen ermittelt waren (Tab. 4.1, Tab. 4.3), konnte die Konstruktion von Textmaterialien für die Sprachaufnahmen beginnen. Das Textmaterial mußte so konstruiert werden, daß die oben beschriebenen Silbenstrukturklassen möglichst optimal abgedeckt sind. Hierzu wurden die ermittelten Onset + Nukleus–Paare (Tab. 4.4) systematisch mit den drei verschiedenen Arten von Codas (stimmlos obstruent, stimmhaft sonorant und offen), die im Deutschen auftreten, kombiniert und die ermittelten Nukleus + Coda–Paare mit den vier Arten von Onsets (stimmlos obstruent, stimmhaft obstruent, stimmhaft sonorant und offen), die im Deutschen auftreten. Das Ergebnis sind 119 nach Klassenzugehörigkeit (11 ...14, 21 ... 24) sortierte Stimuli (Tab. 4.6 – 4.13).

<i>II</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>P</i>	TANN	TAT	TAU	FALKES
<i>B</i>	BANN	GAS	BEIL	WADE
<i>F</i>	FASS	FEE	SCHEIT	AFFEN
<i>Z</i>	WATT	SAAL	WEIN	VASE

<i>11</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>/h/</i>	HASS	HUHN	HEIM	

Tab. 4.6: Stimuli der Klasse 11

<i>12</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>Q</i>	LAMM	REH	RAUCH	FALLE
<i>G</i>	JOCH	JOD	JAHR	KOJEN
<i>N</i>	METT	NUN	NEU	LOHNES

Tab. 4.7: Stimuli der Klasse 12

<i>13</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>PQ</i>	TROTT	PLAN	CHLOR	
<i>BQ</i>	BLATT	GRAM	BLEI	
<i>FQ</i>	FROSCH	FLOH	FLAUM	
<i>ZQ</i>	WRACK			
<i>PN</i>	KNALL	KNIES	PNEU	
<i>BN</i>		GNOM	GNEIS	
<i>FN</i>	SCHNITT	SCHMU	SCHNUR	
<i>PFQ</i> <i>vFPQ</i>	SPLITT	PFRIEM	SPREU	

Tab. 4.8: Stimuli der Klasse 13

<i>14</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>PF</i>	PFIFF	PFAHL	PFAU	
<i>PZ</i>	QUELL	QUA	QUARK	
<i>FP</i>	SPANN	SPIESS	STAU	
<i>FZ</i>	SCHWAMM	SCHWA	SCHWEISS	
<i>PFZ</i>	ZWECK	ZWO	ZWIRN	

Tab. 4.9: Stimuli der Klasse 14

<i>21</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>P</i>	WATT	LAB	EID	HOFFET
<i>F</i>	SCHIFF	AAS	MAUS	RADES

Tab. 4.10: Stimuli der Klasse 21

<i>22</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>/U</i>	ALL	WAL	MAUL	SCHLÜSSEL
<i>N</i>	KAMM	OHM	WEIN	HENNEN

Tab. 4.11: Stimuli der Klasse 22

<i>23</i>	<i>K</i>	<i>L</i>	<i>D</i>	@
<i>PP</i>	SEKT	JAGD	HAUPT	
<i>FF</i>	HAFFS	RUFS	WURFS	

23	K	L	D	@
PF vFP	SCHACHS	UZ	GEIZ	MINDEST
PFP v PPF v FPP v PFF v FFP v FPF	SEKTS	KRIECHST	ARZT	
PFPF v PFFP v FPFP v PPFP	KÖPFST	BEGABTST	SEUFZT	

Tab. 4.12: Stimuli der Klasse 23

24	K	L	D	@
NF v QF	HALS	LEHMS	SAUMS	SCHIMMELS
NP v QP	LAND	SCHULT	EILT	LEBEND
QN	FILM			VIERTELN
QNP v QNF	HALMS			VIERTELND

24	K	L	D	@
NPF √ NFP √ QPF √ QFP √ NPP √ QPP √ NFF √ QFF	WELKT HANFS	AALST	RAUNST	
QNFP	FILMST			
NFPF √ NPPF √ QFPF √ NFPF √ NFFP √ NPPF √ QFPF √ QFFP √ NPFFP √ NFPFP √ NPPFP	TÜNCHST SCHRUMPFST SANFTST			

Tab. 4.13: Stimuli der Klasse 24

Die leer gebliebenen Tabellenfelder zeigen noch einmal, daß viele der phonotaktisch denkbaren Phonemkombinationen tatsächlich nicht auftreten. So sind z.B. lange Vokale im Nukleus (fast) nicht mit einer mehr als drei Konsonanten umfassenden Coda zu kombinieren.

Zum Schluß sind die oben aufgeführten Stimuli dann noch in zufälliger Reihenfolge ausgewählt jeweils in den Trägersatz 'Er hatte ____ gesagt.' eingebettet worden, um so eine möglichst über alle Stimuli hinweg gleichbleibende Intonation der Stimuli zu bekommen und koartikulatorische Effekte wegen des konstant gehaltenen Kontexts besser kontrollieren zu können.

4.3 Sprachaufnahmen

Nachdem die Silbenstrukturklassen ermittelt (Tab. 4.1, Tab. 4.3) und Testtextmaterialien mit (fast) optimaler Abdeckung konstruiert waren (Tab. 4.6 – 4.13), konnte mit den Sprachaufnahmen begonnen werden, um so eine Datengrundlage für die beabsichtigten statistischen Untersuchungen (siehe dazu Kapitel 4.5) zu bekommen.

Damit eine relativ große Variationsbreite an Frequenzwerten und Intonationsmustern erfaßt werden konnte, wurden sowohl zwei männliche als auch zwei weibliche Sprecher des Hochdeutschen aufgenommen. Jeder Sprecher mußte die bereits vorgestellten 119 Stimuli (Tab. 4.6 – Tab. 4.13) jeweils in den Trägersatz eingebettet zweimal vorlesen. Hierbei wurde auf eine möglichst natürliche Intonation sowie eine möglichst einheitliche Aussprache des den Stimulus umgebenden Trägersatzes geachtet.

Die Aufnahmen wurden mit einer Abtastfrequenz von 16 kHz und einer Auflösung von 16 Bit durchgeführt. Abbildung 4.1 zeigt einen Ausschnitt aus dem Satz 'Er hatte **All** gesagt.', gesprochen von Sprecher 1.

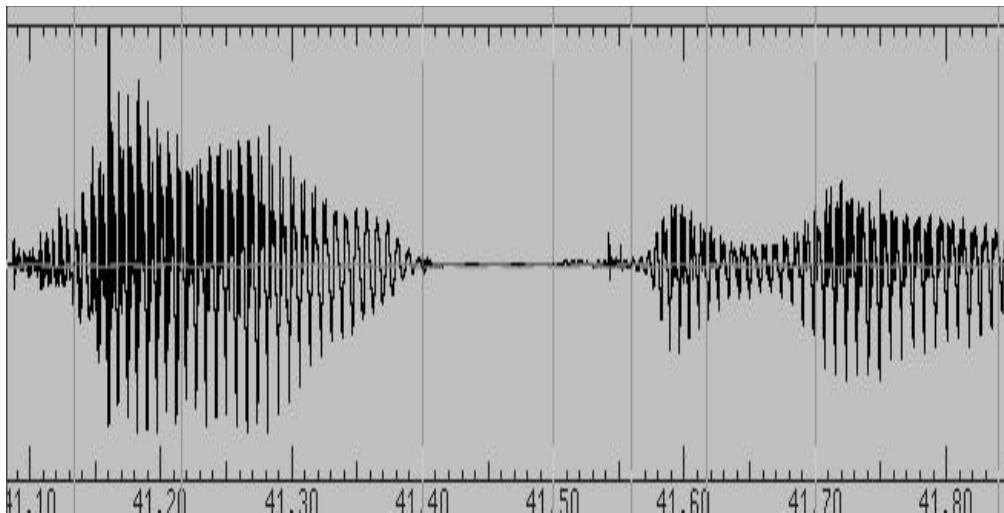


Abb. 4.1: Ausschnitt aus 'Er hatte All gesagt.', Sprecher1

4.4 Lautsegmentierung, F0-Messungen

Ausgehend von den Sprachdaten, die nun in digitaler Form zur Verfügung standen, konnte mit der Lautsegmentierung und schließlich mit den F0-Messungen begonnen werden. Basierend auf den Ergebnissen der F0-Messungen sollten charakteristische Punkte entlang der F0-Kontur der Akzente, der 119 Stimuli also, ermittelt werden, gemessen in den Dimensionen Zeit und Frequenz.

4.4.1 Alignment nach Phonemen

Zunächst wurde dann auf dem aus 952 Sätzen (4 Sprecher * 2 Wiederholungen * 119 Stimuli) bestehenden Aufnahmekorpus die Lautsegmentierung, also das Alignment nach Phonemen, automatisch durchgeführt (/usr/local/ims/bin/alignphones). Abbildung 4.2 zeigt die graphische Ausgabe dieses Arbeitsschrittes.

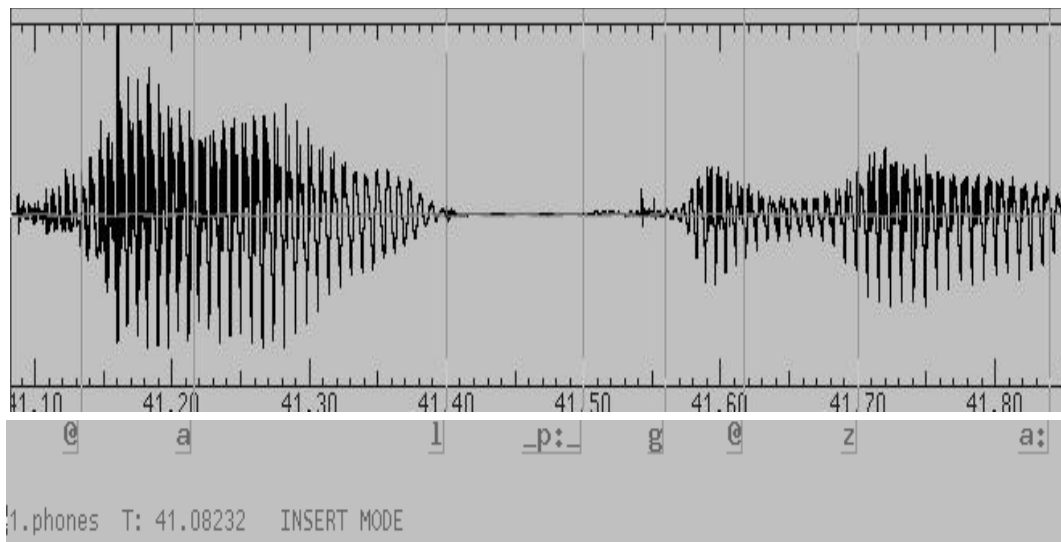


Abb. 4.2: Sprachsignal und Labelfile für einen Ausschnitt aus 'Er hatte All gesagt.', Sprecher 1

Die automatisch gesetzten Label, also die Phonemgrenzen, wurden dann manuell überprüft und gegebenenfalls auch korrigiert, um sicherzustellen, daß eventuell vorhandene Störgeräusche (Lachen, Räuspern o.ä.), Versprecher oder ungewöhnlich lange Pausen nicht zu Fehlern bei der automatischen Alignierung geführt haben. Bei der manuellen Überprüfung der Label wurde speziell darauf geachtet, daß die Phonemgrenze zwischen einem Vokal (kurz oder lang) und einem postvokalischen /R/ entfernt wurde, da Vokal + postvokalisches /R/ in den weiteren Untersuchungen, wie oben schon erwähnt, als Diphthong behandelt werden soll.

4.4.2 F0-Messungen

Nachdem das Alignment abgeschlossen und die automatisch gesetzten Label überprüft waren, wurden F0-Messungen durchgeführt. Hierzu wurde für jeden einzelnen der acht mal 119 Stimuli automatisch der Grundfrequenzverlauf aus dem Signal herausgefiltert. Und zwar wurde hierbei alle 5 ms ein Wert berechnet (`/usr/local/ims/bin/get_f0 -i 0.005`) (Abbildung 4.3).

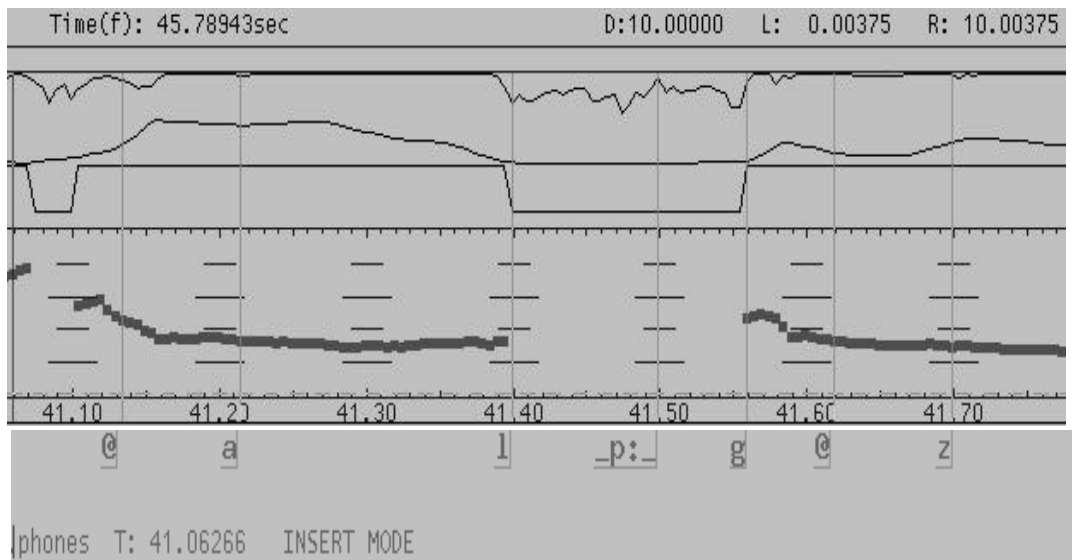


Abb. 4.3: Grundfrequenzverlauf und Labelfile für einen Ausschnitt aus 'Er hatte All gesagt.', Sprecher 1

4.4.3 Ermittlung der Ankerpunkte

Für den nächsten Arbeitsschritt, die Ermittlung charakteristischer Punkte entlang der F0-Kontur der Akzente, wurden absolute Frequenzwerte benötigt, die den beispielsweise in Abbildung 4.3 dargestellten F0-Verlauf beschreiben. Mit Hilfe von 'ppplain' (/usr/local/ims/bin/ppplain) wurden die graphischen Darstellungen der Grundfrequenzverläufe (siehe z.B. Abb. 4.3) in Folgen von absoluten Frequenzwerten umgewandelt, alle 5 ms ein Wert. Das Ergebnis sah dann beispielsweise folgendermaßen aus (Ausschnitt aus 'Er hatte **All** gesagt.', Sprecher 1):

ALL

160.83
157.558
153.764
144.073
142.661
133.713
133.917
134.568
134.487
134.49
133.405
134.742
134.951
134.535
134.264
133.996
130.089
128.905
129.136
130.894
128.363
128.569
128.047
127.819
125.751

126.19
126.302

Ausgehend von diesen Frequenzwerten sollten nun für jeden Stimulus charakteristische Punkte entlang der F₀-Kontur ermittelt werden, gemessen in den Dimensionen Zeit und Frequenz. Diese charakteristischen Punkte werden im weiteren Verlauf der Arbeit auch als **Ankerpunkte** bezeichnet.

Hierzu wurde zunächst der Punkt auf der Grundfrequenzkurve ausgewählt, an dem der höchste Frequenzwert innerhalb des jeweiligen Akzents gemessen wurde, also der Frequenzgipfel der entsprechenden F₀-Kontur. Für diesen Gipfel der jeweiligen F₀-Kontur, an dem also die Frequenzwerte betreffend 100 % erreicht wurden, wurde die Zeit berechnet, die vom Silbenbeginn ausgehend benötigt wurde, um dieses Frequenzmaximum zu erreichen. Dieselben Berechnungen wurden für jeweils sechs charakteristische vor- bzw. nachgipflige Punkte durchgeführt, nämlich für 0%, 10 %, 25 %, 50 %, 75 % und 90 %. Vor- bzw. nachgipflig bezieht sich hierbei auf die Zeit, die zum Erreichen des jeweiligen Punktes vom Silbenbeginn ausgehend benötigt wurde. Es wurden also für jede Akzentkurve, wie in der folgenden Abbildung (Abb. 4.4) beispielhaft für drei Ankerpunkte dargestellt, 13 Ankerpunkte ermittelt.

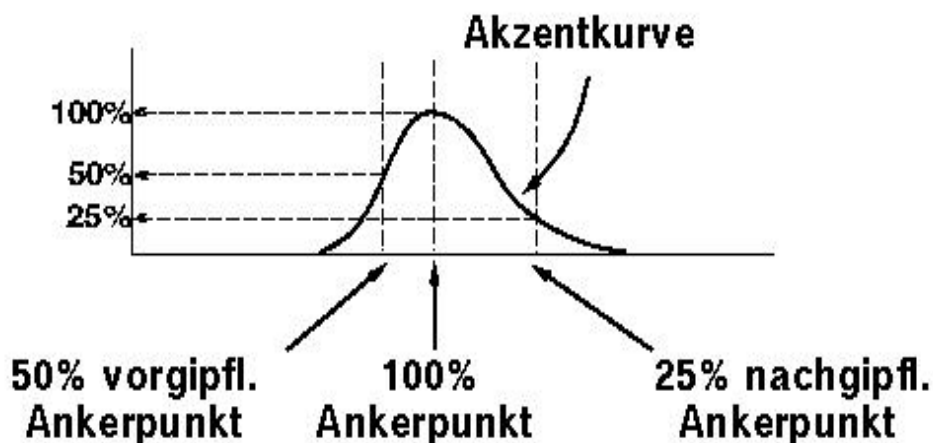


Abb. 4.4: Ermittlung der Ankerpunkte (aus [möb01])

Damit standen also für jeden Akzent Daten über die zeitliche Position der 13

ausgewählten Ankerpunkte zur Verfügung (Tabelle 4.14). In dem in Tabelle 4.14 vorgestellten Beispiel handelt es sich um eine nur fallende Akzentbewegung, was daran zu erkennen ist, daß schon der erste gemessene Frequenzwert das Frequenzmaximum darstellt und damit alle vorgipfligen Werte und der Frequenzgipfel bereits nach 0 ms erreicht wurden.

<i>Ankerpunkte:</i>	<i>Frequenzwerte in Hz:</i>	<i>Zeit vom Silbenbeginn in ms:</i>
<i>0.00%</i>	121	0
<i>10.00%</i>	125	0
<i>25.00%</i>	131	0
<i>50.00%</i>	141	0
<i>75.00%</i>	151	0
<i>90.00%</i>	157	0
<i>100.00%</i>	161	0
<i>90.00%</i>	157	10
<i>75.00%</i>	151	15
<i>50.00%</i>	141	25
<i>25.00%</i>	131	80
<i>10.00%</i>	125	140
<i>0.00%</i>	121	150

Tab. 4.14: Ankerpunkte gemessen in den Dimensionen Zeit und Frequenz

Bei der Ermittlung der charakteristischen Punkte entlang der F0-Kontur der jeweiligen Akzente mußte darauf geachtet werden, daß nur solche Frequenzwerte bei den Berechnungen berücksichtigt wurden, die auch tatsächlich mit den jeweiligen Stimuli verbunden sind, und nicht solche, die aufgrund von koartikulatorischen Effekten verursacht wurden, wie etwa Stimmhaftigkeit zu Beginn von ansonsten stimmlosen Phonemen. Ebenso sollten die Frequenzwerte in den 'creaky voice'-Phasen zu Beginn einiger Vokale unberücksichtigt bleiben.

4.5. Statistische Auswertung

Die in den Dimensionen Zeit und Frequenz gemessenen Ankerpunkte (wie auszugsweise in Tabelle 4.14 vorgestellt) sollten nun noch statistisch als von der Silbenstruktur und eventuell auch von dem Sprecher abhängige Variable dargestellt werden. Außerdem sollte untersucht werden, ob die Silbenstrukturklassen mit den Akzenttypen (nur **fallend** oder **steigend – fallend**) signifikant interagieren.

Als Analyseverfahren wurde ein Verfahren zur **Varianzanalyse** (uni- bzw. multivariat) verwendet.

4.5.1 Datengrundlage für die statistische Analyse

Um ausgehend von den Ankerpunkten die Datengrundlage für die statistische Analyse zu gewinnen, konnten absolute Frequenzwerte vernachlässigt werden, da im Rahmen dieser Studienarbeit nur eine horizontale, also zeitliche, Modellierung des F0–Verlaufs beabsichtigt ist. Es mußten aber noch der jeweilige Sprecher und die jeweilige Silbenstrukturklasse sowie der Akzenttyp kodiert werden, damit die Ankerpunkte statistisch als von der Silbenstruktur und eventuell auch dem Sprecher abhängige Variable dargestellt und eine Interaktion zwischen Silbenstrukturklasse und Akzenttyp festgestellt werden konnte. Der vollständig bearbeitete Input für die statistische Analyse (Tabelle 4.15) enthielt also neben der Information über die Zeit (in ms), die vom Silbenbeginn aus benötigt wurde, um die Ankerpunkte 1 .. 13 (–0 %, –10 %, –25 %, –50 %, –75 %, –90%, 100 %, 90 %, 75 %, 50 %, 25 %, 10 %, 0 %, wobei das negative Vorzeichen für die vorgipfligen Ankerpunkte steht) zu erreichen, noch die Kodierung der Silbenstrukturklasse, des Akzenttyps und des Sprechers.

Silbenklasse	Akzenttyp	Ankerpunkte: 1 2 3 4 5						
22	fallend	Zeit (in ms):	0	0	0	0	0	
6	7	8	9	10	11	12	13	Sprecher
0	0	10	15	25	80	140	150	1

Tab. 4.15: Ausschnitt aus dem Input für die statistische Analyse

Anhand der in Tabelle 4.15 gegebenen Information läßt sich der F0–Verlauf des jeweiligen Akzents nachbilden (Abbildung 4.5).

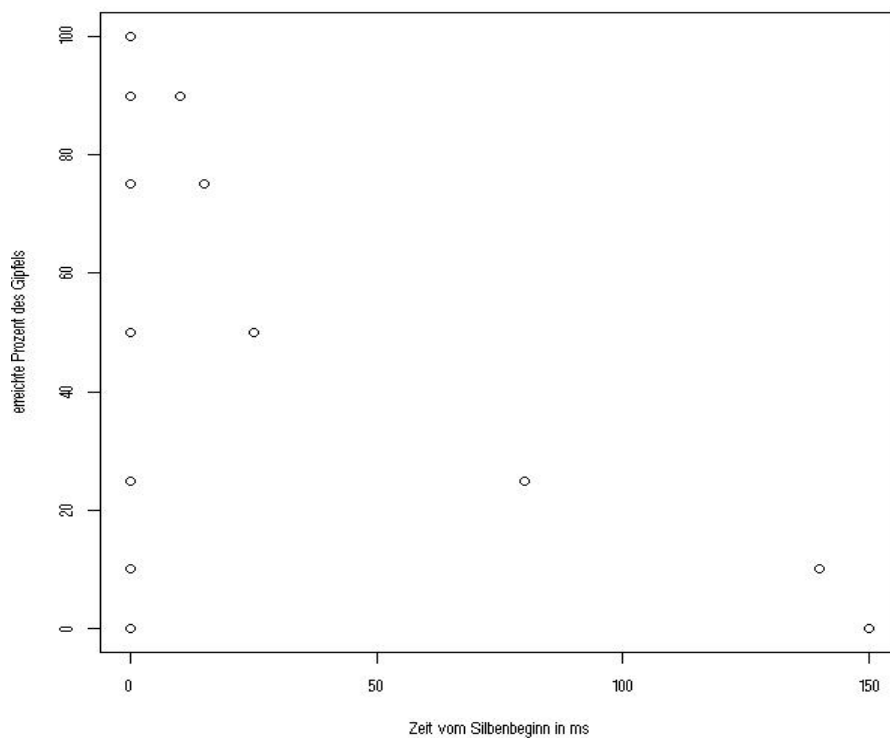


Abb. 4.5: Anhand der Information aus Tab. 4.15 nachgebildeter F0–Verlauf

4.5.2 Varianzanalyse

Bei der Varianzanalyse werden eine oder mehrere **unabhängige Variable**, in unserem Fall die acht verschiedenen Silbenstrukturklassen bzw. die vier verschiedenen Sprecher oder die beiden Akzenttypen und eine oder mehrere **abhängige Variable**, in unserem Fall die Werte an den 13 verschiedenen Ankerpunkten unterschieden. Die unabhängigen Variablen werden auch als

Faktoren bezeichnet. Die folgenden Fragestellungen sollen nun mit Hilfe der Varianzanalyse untersucht werden, [back96] :

1. Gibt es bezüglich der abhängigen Variablen signifikante, d.h. mehr als zufällige Unterschiede zwischen den durch die Ausprägungen der unabhängigen Variablen bestimmten Gruppen?
2. Unterscheiden sich die Gruppen–Mittelwerte signifikant?
3. Kann zumindest ein Teil der Abweichung des Merkmals vom Gesamtmittelwert durch die Gruppenzugehörigkeit erklärt werden?
4. Kann das Merkmal durch Kenntnis der Gruppenzugehörigkeit besser vorhergesagt werden?

In der Varianzanalyse geht es also darum, die Variation zwischen den Gruppen mit der Variation innerhalb der Gruppen zu vergleichen. Ist erstere größer, geht man davon aus, daß Unterschiede zwischen den einzelnen Ausprägungen der Faktoren bestehen. Diese Unterschiede kann man mit einem Hypothesentest, dem sogenannten **F–Test**, auf Signifikanz überprüfen.

4.5.3 Ergebnisse der Varianzanalyse

Die durchgeführte Varianzanalyse hat ergeben, daß die Werte an allen 13 Ankerpunkten signifikant abhängig von der Silbenstrukturklasse sind, d. h. es bestehen bei der Annahme von sieben Freiheitsgraden (d.f. = 7) signifikante Unterschiede zwischen den einzelnen Ausprägungen der Silbenstrukturklassen (Tabelle 4.16). D. h. es gilt für den F–Wert jedes Ankerpunkts { $F(7) = F\text{-Wert} : 2.2 \exp-16$ }. Je kleiner der Wert ist, desto stärker signifikant ist die Abhängigkeit, so daß sich folgende Signifikanzcodes ergeben :

0 ' *** ' 0.001 ' ** ' 0.01 ' * ' 0.05 ' . ' 0.1 ' ' 1

Die Tabelle 4.16 zeigt also, daß eine stark signifikante Abhängigkeit (***) zwischen den Werten an den Ankerpunkten und den acht verschiedenen Silbenstrukturklassen besteht.

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F–Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
0.00%	7	41.5	< 2,2e –16	***
–10.00%	7	54.9	< 2,2e –16	***

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F-Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
-25.00%	7	43.6	< 2,2e -16	***
-50.00%	7	26.9	< 2,2e -16	***
-75.00%	7	21.3	< 2,2e -16	***
-90.00%	7	21.3	< 2,2e -16	***
100.00%	7	24.9	< 2,2e -16	***
90.00%	7	21.5	< 2,2e -16	***
75.00%	7	20.6	< 2,2e -16	***
50.00%	7	20.7	< 2,2e -16	***
25.00%	7	21.6	< 2,2e -16	***
10.00%	7	24.6	< 2,2e -16	***
0.00%	7	41.5	< 2,2e -16	***

Tab. 4.16: Ergebnisse der Varianzanalyse in Abhängigkeit von den Silbenstrukturklassen

Für die Auswertung der Werte an den Ankerpunkten in Abhängigkeit von den vier verschiedenen Sprechern hat sich ein gänzlich anderes Bild ergeben. Nur die Werte beim Frequenzgipfel und die Werte aller nachgipfligen Ankerpunkte sind signifikant abhängig von den vier verschiedenen Sprechern, zwischen den Werten der vorgipfligen Ankerpunkte und den verschiedenen Sprechern besteht keine signifikante Abhängigkeit (Tabelle 4.17). Wie die Tabelle 4.17 zeigt, besteht z.B. keine signifikante Abhängigkeit zwischen den Werten an dem vorgipfligen Ankerpunkt -10% und den verschiedenen Sprechern, wohingegen die Werte am nachgipfligen Ankerpunkt 25% signifikant abhängig von den verschiedenen Sprechern sind.

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F-Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
0.00%	3	2.5	0.01	*
-10.00%	3	0.6	0.76	
-25.00%	3	1.5	0.17	

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F-Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
-50.00%	3	2.9	0.05	.
-75.00%	3	3.4	0.01	*
-90.00%	3	4.3	0.01	*
100.00%	3	14.4	< 2,2 e -16	***
90.00%	3	13.9	< 2,2 e -16	***
75.00%	3	13.7	< 2,2 e -16	***
50.00%	3	15.2	< 2,2 e -16	***
25.00%	3	15.1	< 2,2 e -16	***
10.00%	3	10.1	4,1 exp -12	***
0.00%	3	8.5	6,3 exp -11	***

Tab. 4.17: Ergebnisse der Varianzanalyse in Abhängigkeit von den Sprechern

Die multivariate Analyse, in der untersucht werden sollte, ob die acht Silbenstrukturklassen signifikant mit den beiden Akzenttypen interagieren, d. h. ob bestimmte Akzenttypen in bestimmten Silbenstrukturklassen besonders häufig auftreten, hat ergeben, daß an keinem der 13 Ankerpunkte eine solche Interaktion signifikant ist (Tabelle 4.18).

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F-Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
0.00%	7	1.4	0.21	
-10.00%	7	1.1	0.37	
-25.00%	7	0.7	0.64	
-50.00%	7	0.4	0.88	
-75.00%	7	0.2	0.98	
-90.00%	7	0.2	0.99	
100.00%	7	0.1	0.99	
90.00%	7	0.1	0.99	

<i>Ankerpunkt:</i>	<i>d. f.:</i>	<i>F-Wert:</i>	<i>P:</i>	<i>Signifikanzcode:</i>
75.00%	7	0.1	0.99	
50.00%	7	0.3	0.95	
25.00%	7	0.4	0.86	
10.00%	7	0.9	0.51	
0.00%	7	0.6	0.76	

Tab. 4.18: Ergebnisse der multivariaten Analyse

4.6 Entwicklung des Modells

Auf der Grundlage der Ergebnisse der in Kapitel 4.4 beschriebenen Messungen und unter Berücksichtigung der in Kapitel 4.5 gewonnenen Erkenntnisse wurde dann ein Modell konstruiert, das die Oberflächenvariation der F0-Kontur von Akzenten einer bestimmten phonologischen Klasse aus der Silbenstruktur in quantitativer Form prädiziert.

4.6.1 Datengrundlage für das Modell

Für die Konstruktion dieses quantitativen Modells wurde zum einen die Information über die zeitliche Position der 13 Ankerpunkte in jedem Akzent benötigt, zum anderen die Information über die absoluten Dauern verschiedener Abschnitte jedes Akzents. Die Information über die zeitliche Position der 13 Ankerpunkte wurde bereits ermittelt (Kapitel 4.4.3) und konnte wie in den Tabellen 4.14 und 4.15 vorgestellt verwendet werden. Für die Berechnung der absoluten Dauern verschiedener Abschnitte der Akzente wurde die Akzentgruppendauer in **partielle Dauern** zerlegt, nämlich in die Dauern von Onset, Rhyme (Nukleus + Coda) und, da eine Akzentgruppe strenggenommen aus einer akzentuierten und allen folgenden nicht akzentuierten Silben einer Intonationseinheit besteht, in die Dauer des sogenannten Remainders, in unserem Fall der Silbe 'ge' von 'ge-sagt'. D.h. die Dauer der jeweiligen Akzentgruppe berechnete sich folgendermaßen ([sant00]) :

$D(\text{Akzentgruppe}) = D(\text{Onset}) + D(\text{Rhyme}) + D(\text{Remainder})$, d. h.

$$D(\text{Watt ge}) = D([\text{v}]) + D([\text{at}]) + D([\text{g}@])$$

Diese partiellen Dauern wurden für jeden der 13 Ankerpunkte benötigt. Tabelle 4.19 zeigt einen Ausschnitt aus den Daten.

Silbenklasse	Akzenttyp	Onset1	Rhyme1	Remainder1
21	steigend–fallend	90	210	125
...		Onset13	Rhyme	Remainder13
...		90	210	125

Sprecher

1

Tab. 4.19: Absolute Dauern von Onset, Rhyme und Remainder (in ms) für die Ankerpunkte 1 .. 13 (Onset1 .. Onset13)

4.6.2 Lineares Modell der Ankerpunkte und Parameterschätzung

Mit Hilfe der in Kapitel 4.6.1 beschriebenen Daten sollten nun die Effekte der verschiedenen Abschnitte jedes Akzents (**Onset, Rhyme, Remainder**) auf die zeitliche Position der einzelnen Ankerpunkte gemessen werden.

Die Position der Ankerpunkte wurde zu diesem Zweck mit Hilfe einer **linearen multiplen Regressionsanalyse** vorhergesagt. Die Idee der linearen Regression ist, eine abhängige Variable Y durch eine oder mehrere unabhängige Variablen X_1, \dots, X_m linear zu beschreiben, d. h. eine Gleichung zu finden, bei der die Vorhersage durch folgende lineare Beziehung

$$Y = b_1 * X_1 + b_2 * X_2 + \dots + b_m * X_m$$

mit möglichst kleinem Fehler behaftet die Werte von Y beschreibt.

Zur Schätzung der Parameter kommt die **kleinste-Quadrat-Schätzmethode** zur Anwendung, siehe [edw79]. Die unbekanntenen Größen der Regressionsgeraden, also die Parameter b_1, b_2, \dots, b_m , werden hierbei so geschätzt, daß die Summe der quadrierten Abweichungen der vorhergesagten Y -Werte von beobachteten Y -Werten minimiert wird.

D. h. die Position der Ankerpunkte wird in unserem Modell anhand einer

gewichteten Kombination der partiellen Dauern vorhergesagt, die Position des jeweiligen Ankerpunkts ist also die **gewichtete Summe** der partiellen Dauern. Die Vorhersage der Position eines Ankerpunkts errechnet sich dementsprechend wie folgt ([sant00]):

Position des Ankerpunkts = $\hat{a} * D(O) + \hat{b} * D(Rh) + \hat{c} * D(Rem)$, wobei

D(O) = Dauer des Onsets
 D(Rh) = Dauer des Rhymes
 D(Rem) = Dauer des Remainders ist.

Die Parameter \hat{a} , \hat{b} und \hat{c} konnten in **matlab** (/usr/local/matlab5/bin/matlab) mit Hilfe der oben beschriebenen **kleinste-Quadrat-Schätzmethode** geschätzt werden, da sowohl die absoluten Dauern von Onset, Rhyme und Remainder als auch die beobachteten Positionen der 13 Ankerpunkte direkt gemessen werden konnten.

4.6.3 Ergebnisse

Die in Kapitel 4.6.2 beschriebenen Berechnungen wurden dann unter Bezugnahme auf die in Kapitel 4.5 aus der statistischen Analyse gewonnenen Erkenntnisse nicht nur auf der Gesamtmenge aller zur Verfügung stehenden Daten, der Menge aller Werte aller 952 Akzente (119 Akzente * 4 Sprecher * 2 Wiederholungen) also, durchgeführt, sondern auch auf verschiedenen disjunkten Teilmengen dieser Gesamtdatenmenge. So wurden entsprechend den beiden Akzenttypen zwei disjunkte Teilmengen gebildet und entsprechend den acht Silbenstrukturklassen acht disjunkte Teilmengen.

Die Ergebnisse bestätigen die Vermutung, daß die **'vorhergesagt-beobachtet-Korrelation'** besser ist, wenn die den Berechnungen zugrundeliegenden Daten entsprechend ihrer Zugehörigkeit zu einem bestimmten Akzenttyp und einer bestimmten Silbenstrukturklasse in disjunkte Mengen aufgeteilt wurden.

Für alle betrachteten Teilmengen variieren die Gewichte \hat{a} , \hat{b} und \hat{c} stark als Funktion der Ankerpunkte, wobei für alle betrachteten Teilmengen der Onset die größten, der Remainder hingegen geradezu vernachlässigbar geringe Effekte auf die Position der Ankerpunkte hat.

Die Abbildungen 4.6 und 4.7 zeigen deutlich, daß der Onset den mit Abstand größten Effekt auf die Position aller Ankerpunkte hat, der Effekt des Remainders hingegen nahezu vernachlässigbar scheint. Diese Beobachtung tritt

zwar schon hervor, wenn man die Parameterschätzung auf der Gesamtdatenmenge durchführt (Abb. 4.6), berücksichtigt man aber z. B. nur solche Akzente, die vom Akzenttyp 'fallend' sind, tritt diese Beobachtung noch deutlicher hervor (Abbildung 4.7). Ähnliches gilt für die Ergebnisse der Berechnungen, die auf den den verschiedenen Silbenstrukturklassen entsprechenden Teilmengen durchgeführt wurden.

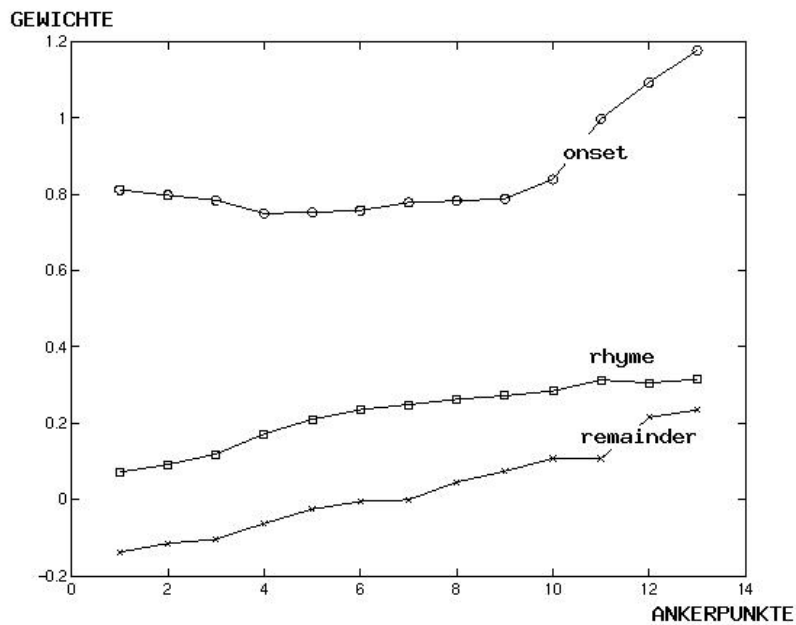


Abb. 4.6: Gewichte als Funktion von Ankerpunkten, Parameterschätzung auf der Gesamtdatenmenge

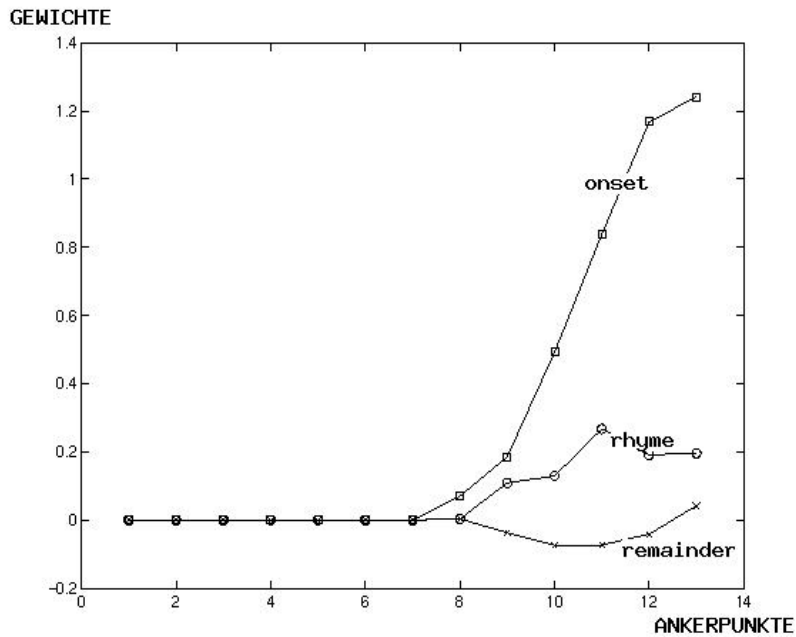


Abb. 4.7: Gewichte als Funktion von Ankerpunkten, Parameterschätzung auf der Teilmenge der fallenden Akzente

Bewertung der Ergebnisse

Benutzt man nun die geschätzten Parameter für Onset, Rhyme und Remainder zur Vorhersage der F₀-Kontur der Akzente, so wird deutlich, daß die auf der Gesamtdatenmenge beruhenden Schätzungen zu wesentlich ungenaueren Vorhersagen führen (Abbildung 4.8) als die auf bestimmten Teilmengen der Daten beruhenden Schätzungen (Abbildung 4.9). Während in Abbildung 4.8 der vorhergesagte F₀-Verlauf deutlich vom beobachteten abweicht, sieht man in Abbildung 4.9 nur noch minimale Abweichungen zwischen dem vorhergesagten und dem beobachteten Verlauf. Die maximale Abweichung beträgt nicht einmal mehr 20 ms gegenüber mehr als 60 ms in Abbildung 4.8, bei der die Parameterschätzung für die Vorhersage auf der Gesamtdatenmenge beruht.

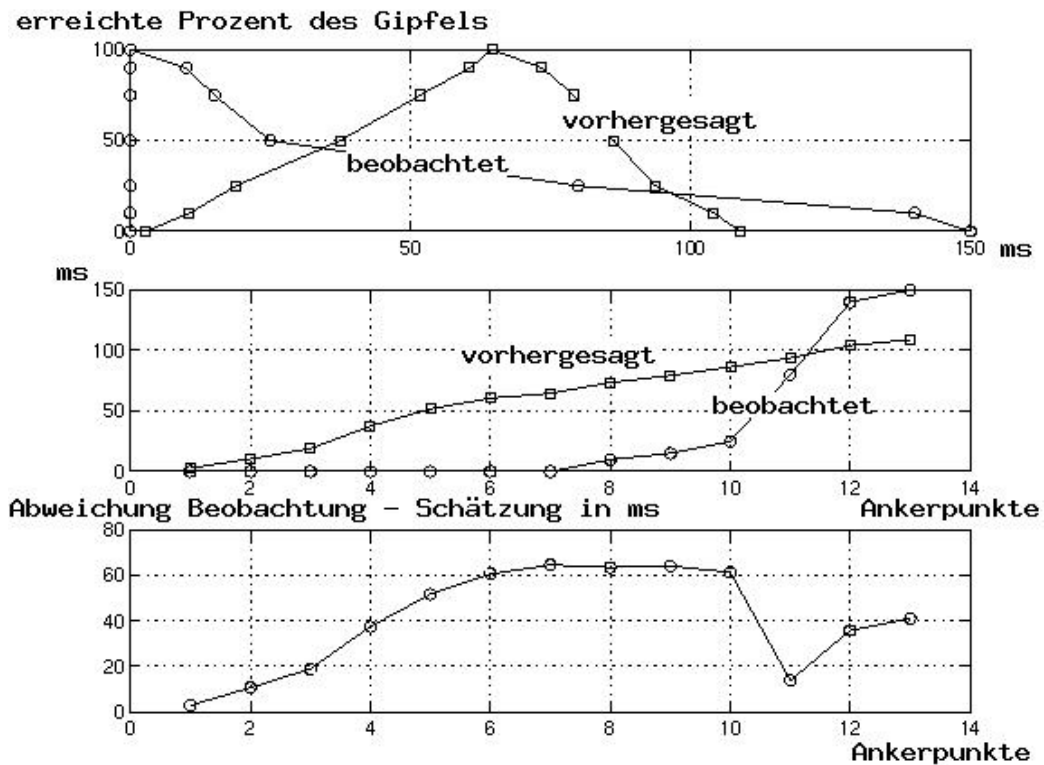


Abb. 4.8: 'beobachtet-vorhergesagt-Korrelation' (Gesamtdatenmenge)

Die Ergebnisse dieser Analyse bestätigen also die Vermutung, daß der Akzenttyp und die Silbenstruktur eine entscheidende Bedeutung für die zeitliche Struktur einer Akzentgruppe besitzen. D. h. aufgrund der Zugehörigkeit eines Akzents zu einer bestimmten Silbenstrukturklasse und zu einem bestimmten Akzenttyp können Rückschlüsse auf den F₀-Verlauf dieses Akzents, der sich in unserem Fall durch die zeitliche Position der Ankerpunkte ausdrückt, gezogen werden. Der horizontale, also der zeitliche Verlauf der F₀-Kontur läßt sich somit bei Kenntnis der Silbenstruktur und des Akzenttyps präzise vorhersagen.

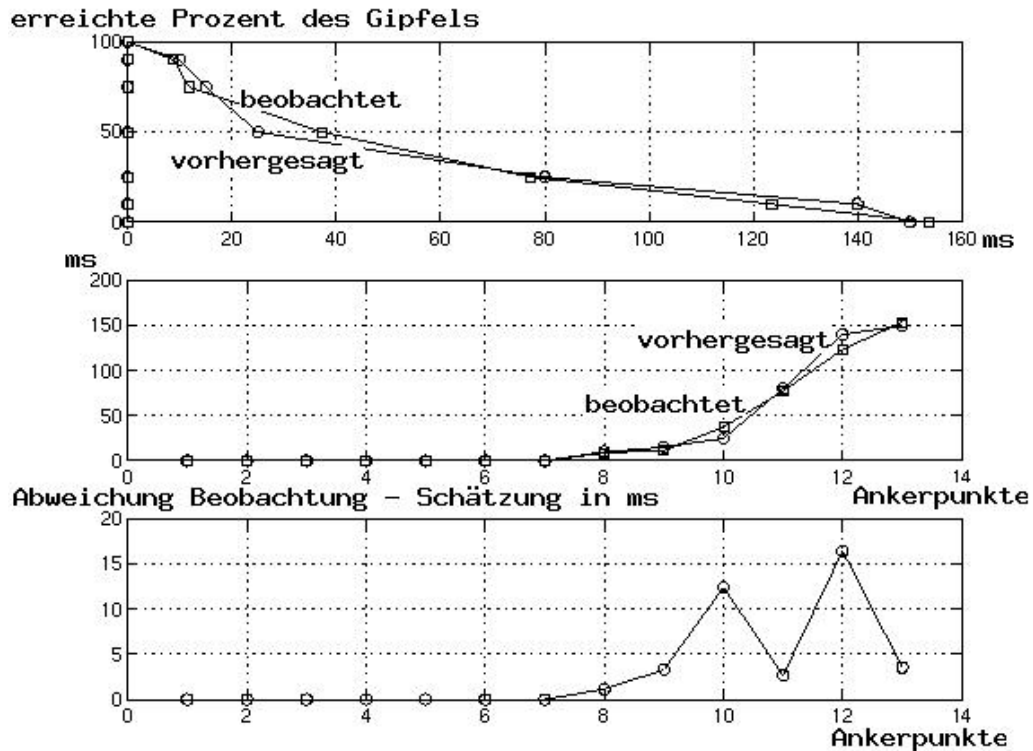


Abb. 4.9: 'beobachtet-vorhergesagt-Korrelation' (Teilmenge der fallenden Akzente)

Eine Verbesserung des Modells könnte eventuell erreicht werden, wenn man die Abweichungen der vorhergesagten von der beobachteten F0-Kontur an jedem der 13 Ankerpunkte und für jede der verschiedenen Teilmengen betrachtet und so vielleicht zu dem Ergebnis kommt, daß bestimmte Ankerpunkte für eine präzise Vorhersage der F0-Kontur von besonderer Bedeutung sind. Aufbauend auf solchen Erkenntnissen könnten dann unterschiedliche Freiheitsgrade für die Ankerpunkte je nach deren Bedeutung für eine präzise Modellierung definiert werden.

Aus den Abweichungen zwischen dem vorhergesagten und dem beobachteten F0-Verlauf an den einzelnen Ankerpunkten ließe sich dann auch der

Korrelationskoeffizient bestimmen, der dann wiederum als ein objektives Maß für die Präzision des Modells in einem möglichen Trainingsschritt eingesetzt werden könnte.

Kapitel 5

Ausblick

Im Rahmen dieser Studienarbeit wurde ein quantitatives Modell entwickelt, das die Oberflächenvariation der F0-Kontur von Akzenten einer bestimmten phonologischen Klasse aus der Silbenstruktur prädiziert. Hierzu wurden acht Silbenstrukturklassen ermittelt, die im Verlauf der Arbeit unterschieden wurden und es wurden 119 Stimuli konstruiert, die diese Silbenstrukturklassen möglichst optimal abdecken sollten.

Die statistische Auswertung der Sprachdaten hat ergeben, daß ein signifikanter Zusammenhang zwischen den in dieser Arbeit ermittelten Silbenstrukturklassen und der Position charakteristischer Punkte entlang der F0-Kontur besteht. Darüberhinaus haben die Auswertungen ergeben, daß die Unterscheidung zweier Akzenttypen sinnvoll erscheint: Nur fallende Akzente und steigend-fallende Akzente.

Diese Ergebnisse bestätigen also die Vermutung, daß die segmentale Zusammensetzung der Silbe einen starken Einfluß auf den F0-Verlauf von Akzenten im Deutschen hat.

Die Vorhersage der F0-Kontur mit Hilfe eines linearen quantitativen Modells hat die Vermutung, daß die Unterscheidung mehrerer Silbenstrukturklassen und die Unterscheidung verschiedener Akzenttypen sinnvoll ist, unterstrichen. Eine auf der Gesamtdatenmenge basierende Parameterschätzung führt zu wesentlich ungenaueren Ergebnissen als eine Schätzung, die auf einer entsprechend den Silbenstrukturklassen und den Akzenttypen bestimmten Teilmenge durchgeführt wurde. Die Güte der vorhergesagten F0-Konturen könnte in einem weiteren Schritt mit Hilfe perzeptiver Experimente überprüft werden und das Modell

könnte mit Hilfe weitergehender Untersuchungen bezüglich verschiedener Zusammenhänge in den Daten und mit Hilfe nachfolgenden Trainings auf einem Trainingsdatenset zu besseren Schätzungen führen.

Ausgehend von den gewonnenen Erkenntnissen könnte noch weitergehend untersucht werden, ob einige der Silbenstrukturklassen eventuell zusammengefaßt werden könnten. Bei den hier untersuchten Daten konnten noch keine ausreichenden Anhaltspunkte dafür gefunden werden, es liegt aber die Vermutung nahe, daß solche Anhaltspunkte gefunden werden können, wenn mehr Stimuli und eventuell auch mehr Sprecher aufgenommen und untersucht werden.

Im Rahmen dieser Studienarbeit wurde speziell der Einfluß der segmentalen Zusammensetzung der Silbe auf den F0-Verlauf von Akzenten im Deutschen untersucht. Man könnte sich vorstellen in weiteren Experimenten auch die Struktur der Akzentgruppe, also mono- vs. polysyllabisch, oder die Informationsstruktur der Äußerung zu variieren und deren Einflüsse auf den F0-Verlauf von Akzenten im Deutschen ebenfalls zu untersuchen.

Außerdem wäre neben der rein zeitlichen Modellierung der F0-Konturen auch eine zusätzliche vertikale, also die Frequenzwerte betreffende Modellierung, denkbar.

Die Ergebnisse dieser Studienarbeit könnten im Bereich der Sprachsynthese innerhalb der Prosodiegenerierung Anwendung finden. Da gezeigt wurde, daß ausgehend von der Silbenstruktur und dem Akzenttyp eine einigermaßen präzise Modellierung des F0-Verlaufs möglich ist, könnte man sich vorstellen, die Ergebnisse des Modells zur Generierung von typischen F0-Verläufen für Akzente einer bestimmten phonologischen Klasse zu nutzen.

Literaturverzeichnis

[back96] Klaus Backhaus, Bernd Erichson. *Uni- und Multivariate Analysemethoden*. Teubner, Stuttgart, 1996

[buß90] Hadumod Bußmann. *Lexikon der Sprachwissenschaft*. Alfred Kröner Verlag, Stuttgart, 1990

[dut97] Thierry Dutoit. *An Introduction to Text-to-Speech Synthesis*. Kluwer Academic Publishers, Dordrecht, 1997

[edw79] A. L. Edwards. *Multiple Regression and the analysis of variance and covariance*. Freeman, San Francisco, 1979

[gieg85] Heinz J. Giegerich. *Metrical Phonology and Phonological Structure: German and English*. Cambridge University Press, Cambridge, 1985

[haase99] Martin Haase. *Aspekte einer Syntax-Prosodie-Schnittstelle*. Studienarbeit, Institut für Maschinelle Sprachverarbeitung, 1999

- [hall92] Tracy A. Hall. *Syllable Structure and Syllable-related Processes in German*.
Niemeyer, Tübingen, 1992
- [hess00] Wolfgang Hess. *Funktionale Phonetik und Phonologie*.
In: *Grundlagen der Phonetik*, IKP-Forschungsberichte II-59, 2000
- [hulst89] Harry van der Hulst. *Atoms of Segmental Structure: Components, Gestures and Dependency*.
In: *Phonology*, 6: 235–284, 1989
- [jes04] Otto Jespersen. *Lehrbuch der Phonetik*.
Teubner, Leipzig, 1904
- [ladd96] D. Robert Ladd. *Intonational Phonology*.
Cambridge University Press, Cambridge, 1996
- [leh70] Ilse Lehiste. *Suprasegmental Features of Speech*.
In: Norman J. Lass (ed.), *Principles of Experimental Phonetics*, pp. 226–241, 1996
- [lev89] Willem J. M. Levelt. *Speaking. From Intention to Articulation*.
The MIT Press, Cambridge, 1989
- [möb93] Bernd Möbius. *Ein quantitatives Modell der deutschen Intonation – Analyse und Synthese von Grundfrequenzverläufen*.
Niemeyer, Tübingen, 1993

- [möb00] Bernd Möbius. *Hauptseminar Sprachsynthese II*.
 Institut für Maschinelle Sprachverarbeitung, Sommersemester 2000,
<http://www.ims.uni-stuttgart.de/Lehre/teaching/2000-SS/Synthese2/index>
- [möb01] Bernd Möbius. *Hauptseminar Intonation und Prosodie*.
 Institut für Maschinelle Sprachverarbeitung, Sommersemester 2001,
<http://www.ims.uni-stuttgart.de/Lehre/teaching/2001-SS/Prosodie/index>
- [nepp99] Joachim M. H. Neppert. *Elemente einer akustischen Phonetik*.
 Buske, Hamburg, 1999
- [pier80] Janet B. Pierrehumbert. *The Phonology and Phonetics of English Intonation*.
 PhD Thesis, MIT, Cambridge, 1980
- [sant00] Jan P. H. van Santen, Bernd Möbius. *A Quantitative Model of F0-Generation and Alignment*.
 In: A. Botinis(ed.), *Intonation*, pp. 269–288, 2000
- [selk82] Elisabeth O. Selkirk. *The Syntax of Words*.
 MIT Press, Cambridge, 1982
- [selk84] Elisabeth O. Selkirk. *On the Major Class Features and Syllable Theory*.
 In: Mark Aronoff und Richard T. Oehrle (eds.), *Language and Sound Structure*, pp. 107 – 136, MIT Press, Cambridge, 1984
- [siev01] Eduard Sievers. *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*.
 Breitkopf & Härtel, Leipzig, 1901

- [vat92] Heinz Vater. *Zum Silben–Nukleus im Deutschen*.
 In: Peter Eisenberg, Karl Heinz Ramers, Heinz Vater (eds.),
Silbenphonologie des Deutschen, pp. 100–133, Narr, Tübingen, 1992
- [ven72] Theo Vennemann. *On the Theory of Syllabic Phonology*
 In: *Linguistische Berichte*, 18: 1–18, 1972
- [ven82] Theo Vennemann. *Zur Silbenstruktur der deutschen Standardsprache*.
 In: Theo Vennemann (ed.), *Silben, Segmente, Akzente*, pp. 261–305,
 Niemeyer, Tübingen, 1982
- [ven92] Theo Vennemann. *Syllable Structure and Simplex Accent in Modern
 Standard German*.
 In: Michael Zirkowski, Manuela Noske, Daren Deaton (eds.), *Papers of
 the Twenty–sixth Regional Meeting of the Chicago Linguistic Society, ii,
 The Parasession on the Syllable in Phonetics and Phonology*, pp. 399–412,
 Chicago Linguistic Society, Chicago, 1992
- [wie96] Richard Wiese. *The Phonology of German*.
 Oxford University Press, Oxford, 1996
- [wur70] Wolfgang U. Wurzel. *Studien zur Deutschen Lautstruktur
 (Studia Grammatica viii)*.
 Akademie–Verlag, Berlin, 1970
- [yu92] Si–Taek Yu. *Silbeninitiale Cluster und Silbifizierung im Deutschen*.
 In: Peter Eisenberg, Karl Heinz Ramers, Heinz Vater (eds.),
Silbenphonologie des Deutschen, pp. 172–207, Narr, Tübingen, 1992