

# Zur Rhythmusanalyse mittels akustischer Parameter

Magisterarbeit  
zur Erlangung des Grades eines  
Magister Artium M.A.

vorgelegt  
der  
Philosophischen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität  
zu Bonn

von  
Ingmar Michael Augustus Steiner  
aus  
San Francisco

Revision 1 (04.02.2004)

Eidesstattliche Erklärung:

An Eides statt versichere ich, daß die Arbeit

„Zur Rhythmusanalyse mittels akustischer Parameter“

von mir selbst und ohne jede unerlaubte Hilfe angefertigt wurde, daß sie noch keiner anderen Stelle zur Prüfung vorgelegen hat, und daß sie weder ganz noch im Auszug veröffentlicht worden ist. Die Stellen der Arbeit – einschließlich Tabellen, Karten, Abbildungen usw. – die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem Fall als Entlehnung kenntlich gemacht.

## Inhaltsverzeichnis

1	Einführung .....	1
2	Hintergrund .....	2
2.1	<i>Rhythmische Dichotomie und Isochronie-Hypothese</i> .....	2
2.2	<i>Alternativen zur Isochronie-Hypothese</i> .....	6
2.3	<i>Akustische Rhythmuskorrelate (Ramus et al. 1999)</i> .....	9
2.4	<i>Alternative akustische Rhythmuskorrelate (Grabe &amp; Low 2002)</i> .....	12
2.5	<i>Sonorität und maschinelle Klassifizierung (Galves et al. 2002a)</i> .....	18
3	Motivation und Erwartung .....	25
4	Daten und Methoden .....	35
4.1	<i>Sprecher</i> .....	35
4.2	<i>Material</i> .....	36
4.3	<i>Ablauf der Aufnahmen</i> .....	39
4.4	<i>Aufbereitung der Daten (1)</i> .....	40
4.5	<i>Exkurs: Maschinelle Annotation</i> .....	41
4.6	<i>Aufbereitung der Daten (2)</i> .....	47
5	Ergebnisse und Interpretation .....	51
5.1	<i>Vergleich mit Ramus et al. (1999)</i> .....	51
5.2	<i>Einfluß von Sprechgeschwindigkeit</i> .....	53
5.3	<i>Einordnung des Deutschen</i> .....	55
5.4	<i>Veränderung der Konfiguration entlang der Sonoritätshierarchie</i> .....	56
5.5	<i>Einfluß der Syllabizität von Lateralen und Nasalen</i> .....	60
5.6	<i>Statistische Analyse einzelner Parameter</i> .....	63
5.7	<i>Zusammenfassung der Ergebnisse</i> .....	67
6	Diskussion und Perspektiven .....	69
6.1	<i>Kritikpunkte</i> .....	69
6.2	<i>Maschinelle Parameterextraktion</i> .....	71
6.3	<i>Anwendungsbereiche</i> .....	72
6.4	<i>Weitere Forschungsansätze</i> .....	74
6.5	<i>Schlußbemerkungen</i> .....	75
7	Literatur.....	76
8	Anhang .....	83
9	Tabellarischer Lebenslauf .....	91

# 1 Einführung

Die vorliegende Arbeit widmet sich der Untersuchung von Sprachrhythmus. Rhythmus kann definiert werden als eine systematische zeitliche Wiederholung bestimmter Ereignisse, wodurch diese Wiederholungen selbst zu zeitlichen Einheiten werden. Sprechrhythmus kann als Rhythmus in gesprochener Sprache definiert werden, Sprachrhythmus als sprechrhythmische Phänomene, die für eine ganze Sprache bezeichnend sind. Die Ereignisse, die sich zeitlich wiederholen, können unterschiedlich sein; sie können auf segmentaler oder suprasegmentaler Ebene auftreten, und es können sich mehrere Rhythmen überlagern. Dauer (1987) schreibt hierzu folgendes:

Rhythm, or the grouping of elements into larger units, is a property of all languages. The particular rhythm of a language is the result of the interaction of a number of components, including phonetic components, such as the relative length, pitch, and segmental quality of accented and unaccented syllables, and phonological components, such as syllable structure and the function of accent.

Darüber hinaus ist es möglich, einzelne Sprachen anhand solcher Rhythmen zu beschreiben und in Gruppen von Sprachen anzuordnen. Wie die rhythmischen Eigenschaften einzelner Sprachen genauer beschrieben werden können, und welche rhythmischen Gemeinsamkeiten Sprachen zu Gruppen verbinden, wird im folgenden erläutert.

## 2 Hintergrund

In diesem Kapitel wird der aktuelle Forschungsstand auf dem Gebiet der Sprachrhythmus-Forschung zusammenfassend dargestellt, wobei einzelnen Untersuchungen, die sowohl die jüngere Forschung auf diesem Gebiet verändert, als auch die vorliegende Arbeit inspiriert haben, besondere Aufmerksamkeit gewidmet wird. Der historische Hintergrund, der im folgenden Abschnitt beschrieben wird, insbesondere die Debatte um die rhythmische Klassifizierung einzelner Sprachen, richtet sich nach Dauer (1983), Nespov (1990), Ramus et al. (1999) und Grabe & Low (2002).

### ***2.1 Rhythmische Dichotomie und Isochronie-Hypothese***

Die Unterscheidung der Sprachen der Welt in zwei rhythmische Klassen geht auf Lloyd James (1940) zurück, der auf frühere metrische Charakterisierungen des Englischen (etwa Steele 1775) zurückgreift und erkennt, daß manche Sprachen in rhythmischer Hinsicht dem Englischen ähnlich oder unähnlich sind. Er prägt dabei zwei Begriffe, „Morse-code rhythm“ und „machine-gun rhythm“. Ersterer bezieht sich auf die Beobachtung, daß in Sprachen mit dieser Art von Rhythmus, z.B. im Englischen, die zeitlichen Intervalle zwischen zwei betonten Silben gleichlang zu sein scheinen. Da zwischen zwei betonten Silben eine unterschiedliche Zahl von unbetonten Silben liegen kann, erweckt diese Art von Rhythmus einen Eindruck, der an den Morse-Code erinnert, in dem lange und kurze Impulse einander folgen. Bei Sprachen mit „machine-gun rhythm“ hingegen scheinen Silbendauern gleichlang zu sein, ebenso wie die einzelnen Kugeln einer Maschinengewehrsalve in gleichmäßigen Abständen die Mündung verlassen.

Diese Einteilung von Sprachen in eine von zwei rhythmischen Klassen wird von Pike (1945) aufgegriffen, der zudem die etwas martialisch anmutenden Termini „Morse-code“ und „machine-gun“ durch „stress-timed“ und „syllable-timed“ ersetzt. So wird die Klassifizierung von Abercrombie (1967) übernommen und weiterentwickelt, ver-

bunden mit der Behauptung, jede Sprache falle in eine dieser beiden Kategorien. Diese Einteilung wird *rhythmische Dichotomie* genannt.

Sprachen mit *stress-timed rhythm* werden als *stress-timed languages* bezeichnet, zu Deutsch *akzentzählende Sprachen*. In diese Sprachen sind demzufolge die Intervalle von einer betonten Silbe zur nächsten (Englisch *interstress intervals*)<sup>1</sup> gleichlang, und somit die Silbendauern verschieden lang. Bei Sprachen mit *syllable-timed rhythm*, also *syllable-timed languages*, die im Deutschen als *silbenzählende Sprachen* bezeichnet werden, verhält es sich andersherum; hier soll die Dauer aller Silben gleichlang sein, während die Betonungsintervalle verschieden lang sein können. Da der Kerngedanke dieser Annahmen aus der Beobachtung erwächst, es gebe in gesprochener Sprache grundsätzlich gleichlange (isochrone) Intervalle der einen oder anderen Art (Betonungsintervalle oder Silbendauer), werden sie zusammengefaßt als *Isochronie-Hypothese* bezeichnet. Das Prinzip der Isochronie-Hypothese ist in Abb. 1 schematisch dargestellt.

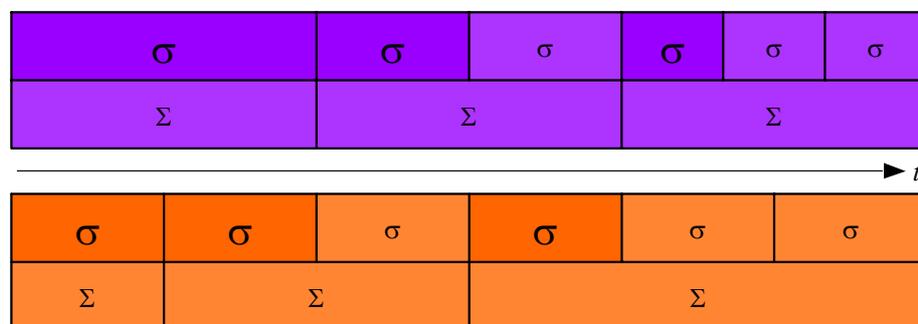


Abb. 1: Schematische Darstellung der Isochronie-Hypothese in akzentzählenden (violett) und silbenzählenden Sprachen (orange). Silben sind mit  $\sigma$  markiert, betonte Silben hervorgehoben. Die Betonungsintervalle sind jeweils darunter mit  $\Sigma$  markiert. In akzentzählenden Sprachen ist die Dauer der Betonungsintervalle immer gleichlang, während die Silbendauer variieren kann. In silbenzählenden Sprachen ist die Silbendauer immer gleichlang, während die Dauer der Betonungsintervalle variieren kann.

Bloch (1950) und Han (1962) erweitern das Inventar der beiden rhythmischen Klassen um eine dritte, die der morazählenden Sprachen, um das Japanische angemessen

<sup>1</sup> Für den Begriff *interstress interval* finden sich auch die Bezeichnungen *foot* oder *rhythm-group* (Cruttenden 1997), im Deutschen *Fuß*.

rhythmisch zu beschreiben. Für morazählende Sprachen wird die Dauer jeder Mora (als subsilbische phonologische Einheit) als gleichlang angenommen.<sup>2</sup>

Die Annahme, jede Sprache gehöre zu der einen oder der anderen rhythmischen Klasse, sagt allerdings noch wenig über die tatsächliche Klassenzugehörigkeit bestimmter Einzelsprachen oder Sprachgruppen aus. Im allgemeinen wird jedoch innerhalb der rhythmischen Dichotomie von der in Tabelle 1 aufgeführten Zuordnung ausgegangen.

Akzentzählende Sprachen	Englisch, andere germanische Sprachen, Russisch, andere slawische Sprachen, Arabisch (Abercrombie 1967, z.T. Corder 1973); Deutsch (Corder 1973, Kohler 1982); Niederländisch (Ladefoged 1975, Smith 1975); Thai (Luangthongkum 1977); europäisches Portugiesisch (Frota & Vigário 2001)
Silbenzählende Sprachen	Spanisch (Pike 1945, Hockett 1958); Französisch, andere romanische Sprachen, Telugu, Yoruba (Abercrombie 1967, z.T. Corder 1973, Catford 1977); Singapur-Englisch (Tongue 1979, Platt & Weber 1980); brasilianisches Portugiesisch (Frota & Vigário 2001)

<sup>2</sup> In der vorliegenden Arbeit schließt ein Verweis auf die traditionelle Unterscheidung zwischen akzentzählenden und silbenzählenden Sprachen stets die morazählenden Sprachen implizit mit ein. Auch die Verwendung der Bezeichnung *rhythmische Dichotomie* bedeutet nicht, daß morazählende Sprachen (als dritte Kategorie) unbedeutend sind. Manche Autoren (z.B. Ramus et al. 1999, Duarte 2001, Grabe & Low 2002) bevorzugen deshalb für die Einteilung aller Sprachen in rhythmische Kategorien die Bezeichnung *Rhythmusklassenhypothese* (*rhythm class hypothesis*), die keine feste Anzahl von rhythmischen Kategorien voraussetzt.

Andererseits kann eine weit gefaßte Definition von „silbenzählend“ auch die morazählenden Sprachen einschließen, und tatsächlich wird Japanisch auch als silbenzählend klassifiziert (Smith 1975); eine ähnliche Diskussion wird um Tamil geführt (Corder 1973, Asher 1985, Steever 1987).

Morazählende Sprachen	Japanisch (Bloch 1950, Han 1962, Port et al. 1980, Port et al. 1987); Estnisch (Gibbon & Gut 2001)
Umstrittene oder unklassifizierte Sprachen	Tschechisch (Daneš 1965); Finnisch (O'Connor 1973, Wiik 1991); Tamil – silbenzählend (Corder 1973, Asher 1985) oder morazählend (Steever 1987); Griechisch (Dauer 1983) Polnisch (Rubach & Booij 1985); Catalan (Nespor 1990); ...

Tabelle 1: Auswahl von traditionell anhand der rhythmischen Dichotomie klassifizierten Sprachen, jeweils mit Literaturangaben. Die Menge der umstrittenen oder „gemischten“ (s. Abschnitt 2.2) bzw. nicht eindeutig oder gar nicht klassifizierten Sprachen ist erheblich größer als hier aufgezählt.

Die Isochronie-Hypothese fußt bei Pike (1945) und Abercrombie (1967) auf physiologischen Grundlagen. Demnach geht die Unterscheidung von rhythmischen Klassen auf die Sprachproduktion zurück. Nach Abercrombie (1967) ist Sprechrhythmus eine Auswirkung der rhythmischen Kontraktionen der Atemmuskulatur des Sprechers. Beim Hörer löst dieser gehörte Rhythmus ein ähnliches rhythmisches Gefühl aus, doch diese sogenannte phonetische Empathie setzt voraus, daß sowohl Sprecher als auch Hörer die gleiche Muttersprache haben. Daß dabei zwei verschiedene Typen von Rhythmus existieren, geht Abercrombie (1967) zufolge darauf zurück, daß es „Brustschläge“ (*chest pulses*) gibt, die die physiologische Grundlage für Silben bilden, und „Betonungsschläge“ (*stress pulses*), die physiologische Grundlage für Silbenbetonung. Von diesen beiden physiologischen Rhythmen ist einer immer isochron, der andere dadurch nicht. Bei Muttersprachlern von silbenzählenden Sprachen kommen „Brustschläge“ in einem gleichmäßigen Rhythmus vor, bei Muttersprachlern akzentzählender Sprachen die „Betonungsschläge“. Die Analyse von fremdsprachlichem Rhythmus setzt viel Erfahrung voraus und gelingt nur wenigen.

Die Realität dieser Sichtweise, die auf Stetson (1928) zurückgreift, darf spätestens seit Ladefoged (1967) und Ohala et al. (1979) als widerlegt gelten. Trotzdem hält sich seit dieser Zeit hartnäckig der Gedanke, daß in allen Sprachen entweder Silben oder Akzente in tendenziell gleichmäßigen Abständen aufeinanderfolgen (Classe 1939; Bolinger 1965; Uldall 1971; Lehiste 1977).

## **2.2 Alternativen zur Isochronie-Hypothese**

Die Unterscheidung der Sprachen in rhythmische Klassen war lange Zeit in Linguistik und Phonetik umstritten. Es ist mehrfach versucht worden (Shen & Peterson 1962; O'Connor 1965; Delattre 1966; Faure et al. 1980; Nakatani et al. 1981, Eriksson 1991), empirische Belege für die Isochronie-Hypothese zu finden, doch kann keine dieser experimentellen Untersuchungen einen Beweis für ihre Gültigkeit erbringen. Die Ausnahme bilden vereinzelte Studien (Abe 1967; Uldall 1971), die anhand verschiedener Datengrundlagen eine Tendenz zur Isochronie des Englischen nur mit einem sehr flexiblen Konzept von „gleichlang“ belegen können, doch bei genaueren Definitionen und exakten Messungen sind diese Eindrücke nicht haltbar. Für das Englische zeigt sich, daß die Länge der Intervalle von einer betonten Silbe zur nächsten keineswegs konstant bleibt, denn sie hängt von folgenden Faktoren ab (Bolinger 1965):

- Anzahl der dazwischen liegenden unbetonten Silben;
- Art der dazwischen liegenden Silben;
- Position des Intervalls innerhalb einer Äußerung.

Ähnliche Untersuchungen für silbenzählende Sprachen (Pointon 1980; Wenk & Wioland 1982; Borzone de Manrique & Signorini 1983) ergeben, daß sich in diesen Sprachen keine Tendenz zu gleichlangen Silben nachweisen läßt. Entsprechende Untersuchungen für das Japanische (Oyakawa 1971; Beckmann 1982; Hoequist 1983a; Hoequist 1983b) können die Isochronie-Hypothese ebenfalls nicht bestätigen.

Eine Untersuchung von Roach (1982) versucht, die Isochronie-Hypothese für sechs Sprachen (Englisch, Russisch, Arabisch, Französisch, Telugu und Yoruba) zu belegen,

die nach Abercrombie (1967) eindeutig verschiedenen rhythmischen Klassen angehören. Es zeigt sich, daß in allen sechs Sprachen

- die Silbendauer ähnlich variiert;
- betonte Silben gleichmäßig verteilt sind.

Dauer (1983) vergleicht Englisch mit Spanisch, Italienisch und Griechisch, nur um festzustellen, daß in allen vier Sprachen die Intervalle von einer betonten Silbe zu nächsten proportional lang zur Anzahl der dazwischenliegenden Silben und gleichmäßig verteilt sind.

Seitdem haben sich die Befürworter der Isochronie-Hypothese auf den Standpunkt zurückgezogen, es gebe entsprechende Intervalle, die als gleichlang *wahrgenommen* werden (Lehiste 1977, Donovan & Darwin 1979, Couper-Kuhlen 1990, Couper-Kuhlen 1993; Zusammenfassung in Beckmann 1992, Laver 1994), gestützt durch Phänomene wie das Vierordtsche Gesetz (lange Zeiten werden zu kurz und kurze Zeiten zu lang wahrgenommen) (Woodrow 1951; Treisman 1963). Mit perzeptiven Untersuchungen, die diese „abgeschwächte“ Form der Isochronie-Hypothese zu bestätigen versuchen, beschäftigt sich die vorliegende Arbeit allerdings nicht weiter.

Auf der phonetisch-phonologischen Seite ist die rhythmische Klassifizierung jedoch ebenfalls nicht einfach fallengelassen, sondern neu formuliert worden. Dasher & Bolinger (1982) machen bestimmte phonologische Eigenschaften für den Eindruck rhythmisch verschiedener Sprachen verantwortlich. Nach Dauer (1987) unterscheiden sich akzentzählende und silbenzählende Sprachen anhand mehrerer Rhythmuskomponenten grundlegend voneinander:

- 1) Dauer: in akzentzählenden Sprachen sind betonte Silben deutlich länger als unbetonte Silben; in silbenzählenden Sprachen ist dies nicht der Fall;

- 2) Silbenstruktur: akzentzählende Sprachen lassen ein größeres Inventar von Silbentypen zu als silbenzählende Sprachen; in ersteren kommen daher mehr und komplexere Konsonantencluster vor;
- 3) Quantität: in akzentzählenden Sprachen kann Quantität nur in betonten Silben distinktiv sein; in silbenzählenden Sprachen können quantitative Unterschiede, soweit vorhanden, unabhängig von Betonung vorkommen;
- 4) Intonation: betonte Silben sind in akzentzählenden Sprachen, soweit möglich, mit Grundfrequenzbewegungen markiert; in silbenzählenden Sprachen kommen Betonung und Grundfrequenzbewegungen unabhängig voneinander vor;
- 5) Tonakzent: in akzentzählenden Sprachen kommt Tonakzent nur bei betonten Silben vor; in silbenzählenden Sprachen werden Töne oder Tonakzente nicht von Betonung beeinflusst;
- 6) Vokalreduktion: in akzentzählenden Sprachen kommen in betonten Silben alle vorhandenen Vokale und Vokalverbindungen vor, in unbetonten Silben werden Vokale reduziert oder zentralisiert; in silbenzählenden Sprachen kommen alle vorhandenen Vokale und Vokalverbindungen in betonten und unbetonten Silben vor;
- 7) Konsonantenreduktion: in akzentzählenden Sprachen werden Konsonanten in betonten Silben sorgfältiger artikuliert, während sie in unbetonten Silben neutralisiert, elidiert, oder durch allophonische Varianten ersetzt werden; in silbenzählenden Sprachen gibt es keine betonungsabhängige Allophonie oder Artikulation von Konsonanten;
- 8) Akzent: in akzentzählenden Sprachen sind Wort- und Satzakkent frei und können zur besonderen Hervorhebung verändert werden, wobei auch Bedeutungsveränderungen auftreten können; in silbenzählenden Sprachen gibt es keinen freien Wortakkent, und eine Veränderung des Satzakkents, soweit vorhanden, verändert nicht die Bedeutung.

Darüber hinaus wird die strikte rhythmische Dichotomie in Frage gestellt, und eine Sichtweise bevorzugt, die anhand dieser Merkmale einzelne Sprachen als „mehr“ oder „weniger“ akzentzählend bzw. silbenzählend charakterisiert (Dauer 1987). Das bedeutet, daß einzelne Sprachen sich nun um zwei idealtypische Konstrukte gruppieren, die durch absolute Verkörperung der genannten Eigenschaften zwei Punkte in einem von ihnen aufgespannten Kontinuum repräsentieren: die ideale akzentzählende und die ideale silbenzählende Sprache.

Diese Sichtweise wird durch die Tatsache unterstützt, daß manche bislang schwierig zuzuordnende Sprachen zwar einige, aber nicht alle Kriterien erfüllen, so daß sie anhand dieser Faktoren nicht eindeutig rhythmisch klassifiziert werden können. Als prominente Beispiele für diese rhythmisch „intermediären“ Sprachen dienen Polnisch (Rubach & Booij 1985; Hayes & Puppel 1985), das komplexe Silben besitzt, jedoch keine Vokalreduktion zuläßt, und Catalan (Wheeler 1979; Mascaró 1983), in dem Vokale reduziert werden können, Silbenstruktur und -komplexität aber dem Spanischen (einer silbenzählenden Sprache) gleichen (Nespor 1990). Levelt & van de Vijver (1998) gehen sogar soweit, anhand möglicher Silbentypen zwei neue rhythmische Klassen zu postulieren, in die theoretisch auch noch andere, bislang unerforschte Sprachen fallen. Auer (1993) schlägt ebenfalls eine Erweiterung der rhythmischen Kategorien vor.

Bei einem solchen erweiterten Modell ist es längst nicht mehr möglich, rhythmische Klassenzugehörigkeit anhand von Silbendauer und Intervalllänge zwischen zwei betonten Silben zu bestimmen.

### **2.3 Akustische Rhythmuskorrelate (Ramus et al. 1999)**

Erst Ramus et al. (1999) gelingt es, mit empirischen Mitteln neue akustische Korrelate zu finden, die die rhythmische Dichotomie zwischen akzentzählenden und silbenzählenden Sprachen unterstützen. Da diese Studie den Beginn einer neuen Phase in der übereinzelsprachlichen Rhythmusforschung darstellt, und seitdem eine Anzahl von anderen Untersuchungen auf diese Methoden und Maße Bezug genommen haben, wird sie an dieser Stelle ausführlicher beschrieben.

Ramus et al. (1999) versuchen, in der Sprachrhythmus-Erwerbsforschung Hinweise dafür zu finden, ob und wie Säuglinge in der Lage sind, Sprachen rhythmisch zu unterscheiden. Nazzi et al. (1998) haben gezeigt, daß Neugeborene in der Lage sind, anhand von durch Filterung delexikalisierten Stimuli zwischen Sprachen aus verschiedenen rhythmischen Klassen zu unterscheiden, nicht aber zwischen Sprachen derselben rhythmischen Klasse. Aus acht Sprachen (Catalan, Englisch, Französisch, Italienisch, Japanisch, Niederländisch, Polnisch und Spanisch) wurde von Ramus et al. (1999) Sprachmaterial gesammelt. Es handelt sich dabei um kurze Aussagesätze von etwa 3 Sekunden Dauer und 15 bis 19 Silben Länge, die in die jeweilige Zielsprache übersetzt wurden. Insgesamt enthält das Korpus fünf solcher Sätze pro Sprecher und vier Sprecher für jede Sprache. Diese Aufnahmen werden von Hand in konsonantische und vokalische Intervalle segmentiert („C/V-segmentiert“) und die Dauer dieser Intervalle gemessen.

Eine relativ unkomplizierte Modellierung einiger in Abschnitt 2.2 genannten Rhythmuskomponenten, die Sprachen phonologisch charakterisieren und somit rhythmisch klassifizieren, beruht auf drei statistischen Maßen, die aus den Dauern der vokalischen und konsonantischen Intervalle berechnet werden. Diese Maße ergeben sich aus den folgenden Überlegungen:

- Vokalreduktion verringert den vokalischen Anteil an der Gesamtdauer. Demnach muß dieser vokalische Anteil in Sprachen mit Vokalreduktion geringer sein als in Sprachen ohne Vokalreduktion. Der prozentuale vokalische Anteil an der Gesamtdauer einer Äußerung wird als %V bezeichnet.
- Sprachen mit komplexer Silbenstruktur erlauben auch einfache Silben. Sprachen ohne komplexe Silbenstruktur besitzen nur einfache Silben. Demnach muß die Varianz (und damit die Standardabweichung) der Dauer von Konsonantenclustern in Sprachen mit komplexer Silbenstruktur höher sein als in Sprachen ohne komplexe Silbenstruktur. Die Standardabweichung der Dauer von Konsonan-

tenclustern (also der konsonantischen Intervalle) wird als  $\Delta C$  bezeichnet.

- Die gleiche Überlegung gilt auch für Vokalreduktion. Demnach muß die Varianz der Dauer von Vokalen und Vokalverbindungen in Sprachen mit Vokalreduktion höher sein als in Sprachen ohne Vokalreduktion. Die Standardabweichung der Dauer von vokalischen Intervallen wird als  $\Delta V$  bezeichnet.

Für jede der  $C/V$ -segmentierten Aufnahmen werden also drei Werte berechnet:  $\%V$ ,  $\Delta C$  und  $\Delta V$ . Diese Maße ergeben Durchschnittswerte für jede Sprache, die anschließend graphisch analysiert werden. Wie sich zeigt, bilden die einzelnen Sprachen Gruppen in der  $\%V/\Delta C$ -Ebene, die überraschend gut mit der traditionellen Unterscheidung in rhythmische Klassen korrespondieren. Diese Ergebnisse sind in Abb. 2 wiedergegeben. Dabei scheint sich Polnisch zu den akzentzählenden, Catalan zu den silbenzählenden Sprachen zu gruppieren.  $\Delta V$  erwies sich als problematischer Parameter,<sup>3</sup> da er von einer Reihe von Faktoren beeinflusst wird, wie Vokalreduktion (per definitionem in den akzentzählenden Sprachen sowie im Catalan), kontrastiver Vokallänge (Japanisch), kontextabhängiger Vokallängung (Italienisch) und langen Vokale (gespannte Vokale im Englischen und Niederländischen, nasalierte Vokale im Französischen). Trotzdem scheint  $\Delta V$  zumindest für Polnisch eine Dimension darzustellen, die es von den übrigen akzentzählenden Sprachen unterscheidet, wobei nicht restlos geklärt ist, welche Rolle  $\Delta V$  für die rhythmische Klassifikation spielt.

---

<sup>3</sup> Die Verteilung der Sprachen im Korpus von Ramus et al, (1999) in der  $\Delta V/\Delta C$ -Ebene ist in anderem Zusammenhang in Abb. 6 dargestellt.

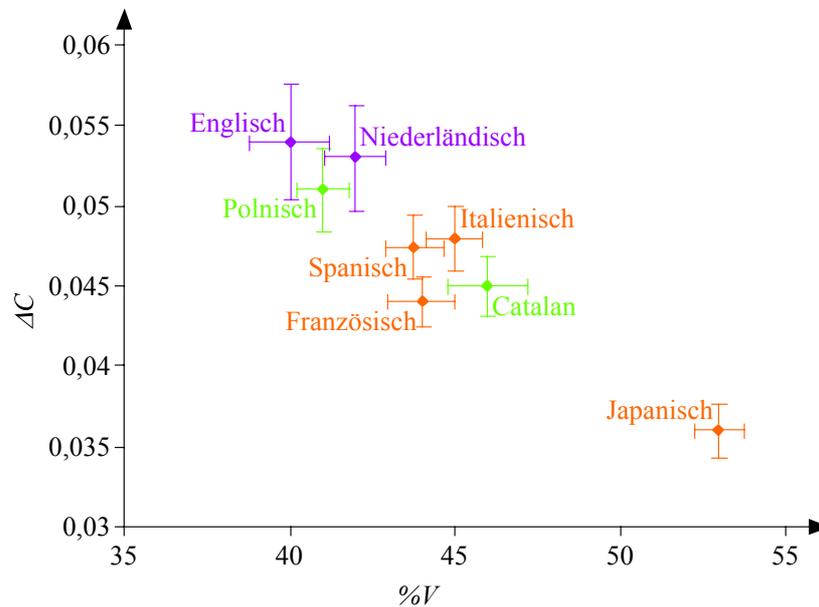


Abb. 2: Sprachverteilung in der  $\%V/\Delta C$ -Ebene (aus Ramus et al. 1999). Die Fehlerbalken repräsentieren  $\pm 1$  Standardfehler. Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

Anschließend führen Ramus et al. (1999) mehrere Diskriminierungsexperimente mit Erwachsenen und Neugeborenen durch, wobei die Stimuli aus den bereits verwendeten Aufnahmen gebildet werden, die durch Filterung, Grundfrequenzmanipulation und Resynthese auf die wesentlichen segmentalen Parameter reduziert worden sind. Das Ergebnis aller Experimente ist, daß die Versuchspersonen trotz Manipulation der Sprachsignale in der Lage sind, zwischen Sprachen aus verschiedenen rhythmischen Klassen zu unterscheiden, jedoch nicht zwischen Sprachen aus derselben rhythmischen Klasse. Diese Resultate sind es jedoch nicht, die Ramus et al. (1999) zu einer der einflußreichsten Untersuchungen zur rhythmischen Klassifikation neueren Datums gemacht haben, sondern die statistischen Methoden und Maße.

## 2.4 Alternative akustische Rhythmuskorrelate (Grabe & Low 2002)

Eine Reihe von Faktoren veranlaßt Grabe & Low (2002), zur Bestimmung akustischer Rhythmuskorrelate eine ähnliche Untersuchung wie die von Ramus et al. (1999) durchzuführen. Allerdings verwenden sie andere Daten und Methoden und beziehen

mehrere Sprachen mit ein, die bislang nicht eindeutig rhythmisch klassifiziert worden sind. Als Datengrundlage dienen Grabe & Low (2002) Aufnahmen des Textes „Nordwind und Sonne“ (IPA 1999), in den Sprachen Catalan, Deutsch, Englisch (Britische und Singapur-Varietät), Estnisch, Französisch, Griechisch, Japanisch, Luxemburgisch, Malaiisch, Mandarin, Niederländisch, Polnisch, Rumänisch, Spanisch, Tamil, Thai und Walisisch. Der Vorteil dieser hohen Anzahl von verschiedenen Sprachen wird dadurch aufgehoben, daß für jede Sprache nur ein Sprecher aufgezeichnet wurde, und Grabe & Low (2002) darüber hinaus keine Angaben zu den Aufnahmen machen, abgesehen von dem Hinweis, daß sie unter verschiedenen Bedingungen durchgeführt wurden.

Die Sprachdaten werden anschließend genau wie bei Ramus et al. (1999) in vokalische und konsonantische Intervalle segmentiert, und die Dauer der Intervalle wird gemessen. Grabe & Low (2002) berechnen anschließend einen paarweisen Index für die Variabilität der Dauer von vokalischen und konsonantischen Intervallen (*rPVI*) anhand der Formel

$$rPVI = \sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m-1)$$

wobei  $m$  für die Anzahl der Intervalle steht, und  $d$  die Dauer des  $k$ -ten Intervalls darstellt. Sie erhalten so ein „rohes“ (daher das „ $r$ “) Variabilitätsmaß, denn dieser Variabilitätsindex ist nicht für Sprechgeschwindigkeit normalisiert.

Um den *rPVI* für Sprechgeschwindigkeit zu normalisieren, wird die Formel modifiziert:

$$nPVI = 100 \times \left[ \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

Allerdings läßt sich diese Formel nur für vokalische Intervalle verwenden, da die Variabilität der konsonantischen Intervalle in allen untersuchten Sprachen zu hoch ist, um eine solche Normalisierung zuzulassen.

Ein ähnliches Maß wurde bereits von Scott et al. (1986) vorgeschlagen, das der *rhythmischen Irregularität*. Dieses quantitative Maß ist bei absoluter Gleichheit der

Länge aller Intervalle einer Sequenz gleich Null, und je höher die Variabilität der Intervalllängen, desto höher das Rhythmische Irregularitätsmaß (*RIM*).

$$RIM = \sum_{i \neq j} \left| \log \frac{I_i}{I_j} \right|$$

Hierbei steht *I* für die Länge des jeweiligen Intervalls. Im Gegensatz zum *nPVI* ist das *RIM* allerdings für die gesamte Sequenz, auf die es angewendet wird, normalisiert, also global, nicht lokal.

Indem für jede Aufnahme (also auch Sprache) in ihrer Datengrundlage ein vokalischer *nPVI*-Wert und ein konsonantischer *rPVI*-Wert berechnet werden, können Grabe & Low (2002) eine graphische Analyse der Sprachverteilung in der *nPVI/rPVI*-Ebene durchführen. Die Ergebnisse dieser Analyse sind in Abb. 3 dargestellt.

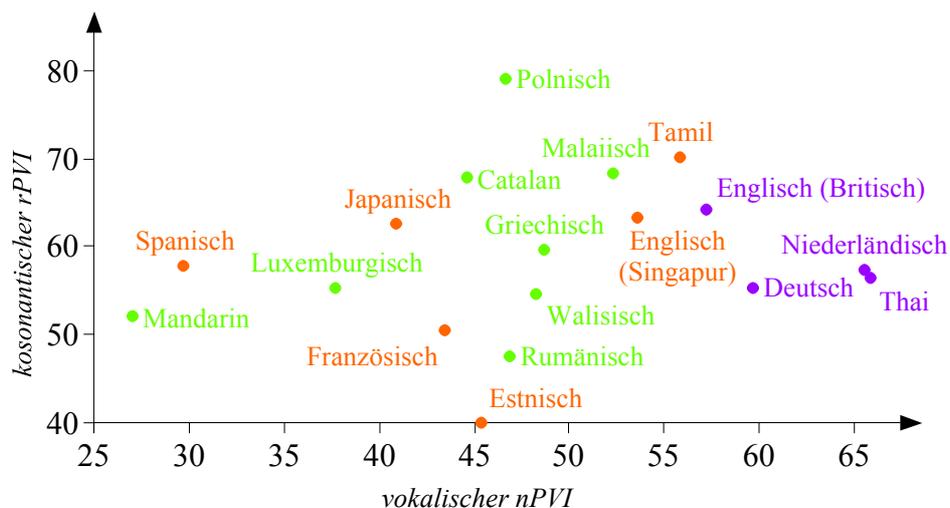


Abb. 3: Sprachverteilung in der *nPVI/rPVI*-Ebene (aus Grabe & Low 2002). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

Grabe & Low (2002) interpretieren diese Ergebnisse als erwartungsgemäß. Die Verteilung der Sprachen in der *nPVI/rPVI*-Ebene unterstützt die Sichtweise, daß sich rhythmische Eigenschaften in einem Kontinuum besser modellieren lassen als durch eine strikte Dichotomie, daß aber die Tendenz zur Gruppierung rhythmisch ähnlicher Sprachen durch die Hinzufügung mehrerer unklassifizierter Sprachen an Schärfe verliert. Die akzentzählenden Sprachen Deutsch, Englisch (Britisch) und Niederländisch weisen eine höhere Variabilität der Dauer der vokalischen Intervalle auf als die sil-

benzählenden Sprachen Französisch und Spanisch. Japanisch scheint sich aufgrund seiner Silbenstruktur und des Mangels an Vokalreduktion ebenfalls zu den silbenzählenden Sprachen zu gesellen, allerdings wird das Ergebnis vermutlich dadurch beeinflusst, daß Grabe & Low (2002) bei der *C/V*-Segmentierung des Japanischen bestimmte Vokale zu den konsonantischen Intervallen zählen. Die bislang umstrittenen oder unklassifizierten Sprachen verteilen sich in der gesamten *nPVI/rPVI*-Ebene, wobei einige (Mandarin und Luxemburgisch silbenzählend; Thai wie von Luangthongkum (1977) beschrieben akzentzählend) zu den jeweiligen Gruppen zu gehören scheinen, andere sich im „Grenzbereich“ zwischen silbenzählenden und akzentzählenden Sprachen ansiedeln. Catalan und Spanisch unterscheiden sich aufgrund der Vokalreduktion im Catalan deutlich. Singapur-Englisch weist zwar geringere vokalische Variabilität auf als Britisches Englisch, doch es scheint auch keinen silbenzählenden Rhythmus zu besitzen. Tamil scheint den akzentzählenden Sprachen zuzuordnen zu sein, was den in Tabelle 1 zitierten Untersuchungen widerspricht.

Insgesamt erwecken die Ergebnisse von Grabe & Low (2002) den Eindruck, nur bedingt mit bisherigen Ergebnissen konform zu sein. Dies liegt vermutlich an zwei Faktoren: erstens ist die Wahrscheinlichkeit für Artefakte sehr hoch, wenn nur eine Aufnahme eines Sprechers für jede Sprache verwendet wird; und zweitens scheinen die konsonantischen *rPVI*-Werte nicht die Erwartungen zu erfüllen, die an sie gestellt werden. Grabe & Low (2002) selbst sprechen diese Kritikpunkte in ihrer Untersuchung an, und vermuten, daß erst eine Erweiterung der Datengrundlage um mehr Sprecher die Rolle von *rPVI* bestätigen könnte.

Die Methoden von Grabe & Low (2002) stehen in ihren Ergebnissen in Kontrast zu denen von Ramus et al. (1999). Es scheint zwar eine Korrelation zwischen *%V* und dem vokalischen *nPVI* zu geben, und daraufhin analysieren Grabe & Low (2002) die Verteilung ihrer Sprachdaten auch in der *%V/ΔC*-Ebene, doch diese Verteilung, dargestellt in Abb. 4, weist wesentliche Unterschiede zu den Ergebnissen von Ramus et al. (1999) auf.

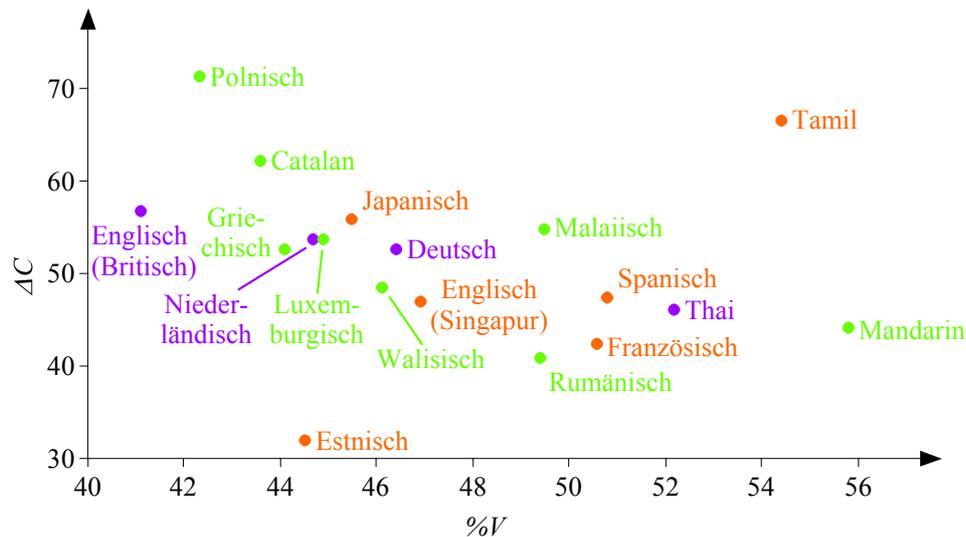


Abb. 4: Sprachverteilung in der  $\%V/\Delta C$ -Ebene (aus Grabe & Low 2002). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

Die Gruppe der silben- und morazählenden Sprachen überschneidet sich hier weitgehend mit der der akzentzählenden Sprachen, während die gemischten oder unklassifizierten Sprachen die „Zwischenräume“ ausfüllen. Diese Sprachverteilung ist mit hoher Wahrscheinlichkeit auf die Tatsache zurückzuführen, daß bei nur einem Sprecher pro Sprache die Sprachunterschiede mit den Sprecherunterschieden (nicht zuletzt individuelle Sprechgeschwindigkeit) zusammenfallen. Die Entscheidung, entstimmte Vokale als Konsonanten zu zählen, hat vermutlich ebenfalls starke Auswirkungen auf diese Verteilung. Und schließlich ist es bei nur einem Wert pro Aufnahme, Sprecher und Sprache nicht möglich, die Varianz oder ähnliche Maße anzugeben, so daß nicht klar ist, wie zuverlässig die berechneten Werte sind.

Im Gegenzug diskutiert Ramus (2002) die Ergebnisse aus Grabe & Low (2002) und wendet ihre Methoden auf sein Datenkorpus aus Ramus et al. (1999) an. Es zeigt sich, daß die Verteilung der Sprachen in der  $nPVI/rPVI$ -Ebene, dargestellt in Abb. 5, dabei nicht nur wesentliche Unterschiede zu der von Grabe & Low (2002) aufweist, sondern auch eine große Ähnlichkeit mit der Verteilung derselben Sprachen in der  $\Delta V/\Delta C$ -Ebene, in Abb. 6 dargestellt, besitzt.

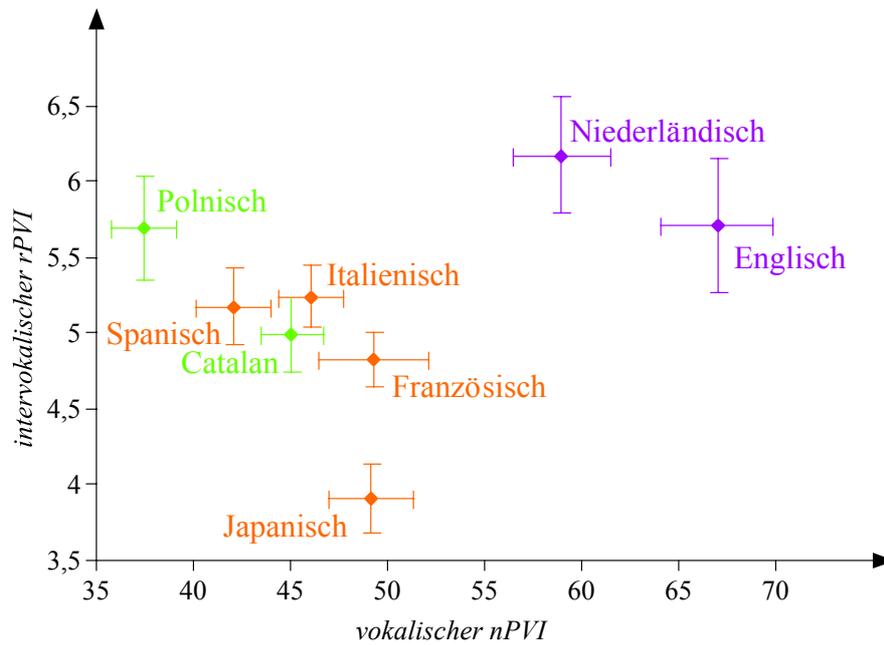


Abb. 5: Sprachverteilung in der  $nPVI/rPVI$ -Ebene (aus Ramus 2002). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

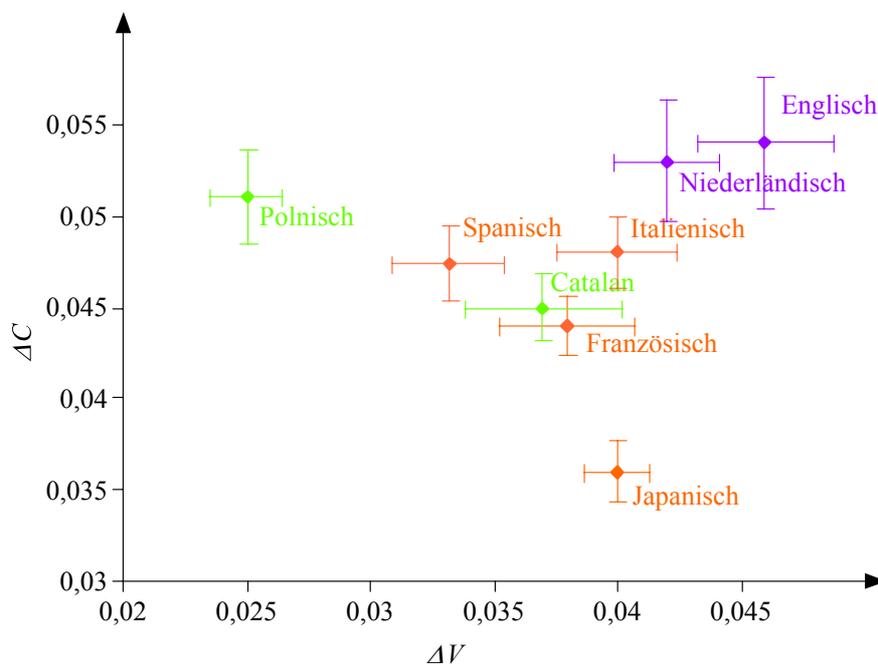


Abb. 6: Sprachverteilung in der  $\Delta V/\Delta C$ -Ebene (aus Ramus 2002). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

Zwar haben seit Grabe & Low (2002) mehrere Untersuchungen deren Methoden übernommen oder weiterentwickelt (Gibbon & Gut 2001; Barry, persönliche Kommunikation), doch scheinen die *PVI*-Maße nicht unproblematisch zu sein:

- $\%V$ , bei Ramus et al. (1999) der zuverlässigste Parameter zur Trennung rhythmischer Klassen, wird nicht miteinbezogen;
- Der konsonantische *rPVI* scheint in etwa mit  $\Delta C$  zu korrespondieren;
- Der vokalische *nPVI* scheint ähnliche Verteilungen hervorzurufen wie  $\Delta V$ , ein Parameter, den Ramus et al. (1999) als zu stark von verschiedenen Faktoren beeinflusst beurteilen (s. Abschnitt 2.3).

Der verbleibende ausschlaggebende Grund, der für die Bevorzugung der *PVI*-Maße zu sprechen scheint, ist die Normalisierung von lokalen Veränderungen der Sprechgeschwindigkeit. Allerdings zeigen Galves et al. (2002b), daß die Dauern aufeinanderfolgender konsonantischer Intervalle keineswegs zu korrelieren scheinen, womit die Methoden von Ramus et al. (1999) bis auf weiteres als besser geeignete Maße erscheinen, Sprachen anhand einfacher statistischer Methoden in rhythmische Klassen zu trennen.

## **2.5 Sonorität und maschinelle Klassifizierung (Galves et al. 2002a)**

Die Methoden von Ramus et al. (1999) werden von Galves et al. (2002a) anhand der Daten aus Ramus et al. (1999) verfeinert, um eine maschinelle Implementierung zu ermöglichen und so nicht nur die aufwendige Arbeit der manuellen *C/V*-Segmentierung zu ersparen, sondern auch das nicht zu unterschätzende Definitionsproblem von konsonantischen und vokalischen Intervallen bei der Segmentierung zu umgehen. Sie gehen von einem universellen Sonoritätsmaß aus, das sich aus akustischen Parametern des Sprachsignals extrahieren läßt und bilden diese Werte auf eine Skala von 0 (keine Sonorität) bis 1 (hohe Sonorität) ab. Dabei wird für jede Aufnahme zunächst

ein Spektrogramm erzeugt und dieses im Frequenzbereich zwischen 0Hz und 800Hz abgetastet. Das renormalisierte Amplitudenspektrum wird durch die Formel

$$p_t(i) = \frac{c_t(i)^2}{\sum_f c_t(f)^2}$$

definiert, wobei  $c_t(i)$  der Fourier-Koeffizient für die Frequenz  $i$  um den Zeitpunkt  $t$  ist. Die relative Entropie für  $p_t$  bezüglich  $p_{t-1}$  ist definiert durch die Formel

$$h(p_t | p_{t-1}) = \sum_i p_t(i) \log \left( \frac{p_t(i)}{p_{t-1}(i)} \right)$$

wobei die relative Entropie stets positiv ist und sich bei geringen Unterschieden dem Wert 0 nähert. Mit diesen Formeln definieren Galves et al. (2002a) die Sonoritätsfunktion  $s(t)$  wie folgt:

$$s(t) = 1 - \min \left( 1, \frac{1}{27} \sum_{u=t-4}^{t+4} \sum_{i=1}^3 h(p_u | p_u - i) \right)$$

Der Mittelwert von  $s(t)$  ist definiert als

$$\bar{S} = \frac{1}{T} \sum_{t=1}^T s(t)$$

und stellt ein Maß dar, das mit %V bei Ramus et al. (1999) vergleichbar ist.

Ein weiteres Maß  $\delta S$  gibt an, welchen Einfluß Regionen mit geringer Sonorität besitzen und ist mit AC vergleichbar:

$$\delta S = \frac{1}{T} \sum_{t=1}^T |s(t) - s(t-1)|$$

Galves et al. (2002a) berechnen für jede der Aufnahmen aus dem Korpus von Ramus et al. (1999) sowohl  $\bar{S}$  als auch  $\delta S$ . Eine graphische Analyse weist eine mit den Ergebnissen von Ramus et al. (1999) konsistente Verteilung der Einzelsprachen in der  $\bar{S}/\delta S$ -Ebene auf, die in Abb. 7 dargestellt ist.

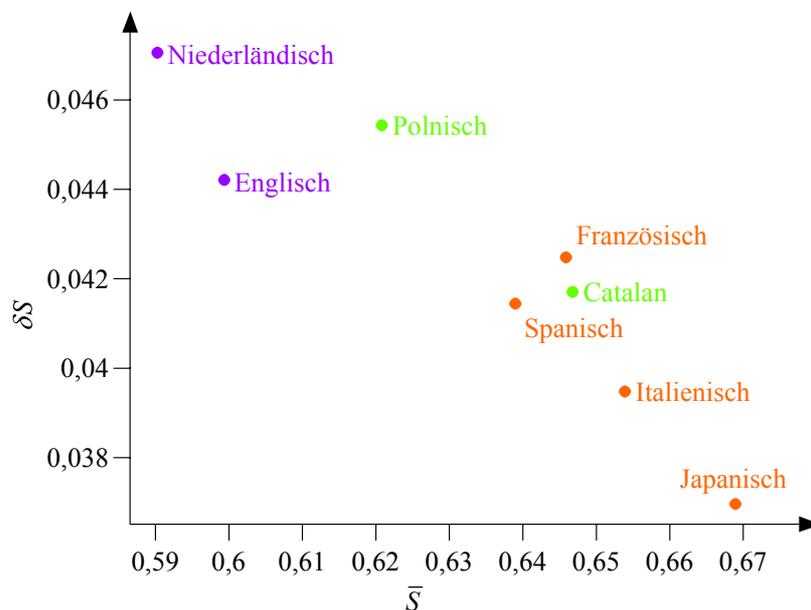


Abb. 7: Sprachverteilung in der  $\bar{S} / \delta S$ -Ebene (aus Galves et al. 2002a). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

Darüber hinaus zeigen Galves et al. (2002a), daß  $\bar{S}$  und  $\%V$  respektive  $\delta S$  und  $\Delta C$  linear korrelieren, also durchaus vergleichbare Maße sind. Allerdings geben Galves et al. (2002a) nicht an, wie hoch die Streuung der einzelnen Aufnahmen innerhalb jeder untersuchten Sprache ist, so daß über Varianz auch hier keine Aussage gemacht werden kann. Sie wenden ihre Methoden auch nicht wie angedeutet auf ein größeres Korpus an, sondern verwenden dieselbe Datengrundlage wie Ramus et al. (1999). Allerdings läßt dies einen besseren Vergleich zwischen den verschiedenen statistischen Maßen zu, während die Untersuchung eines größeren Korpus mit diesen Methoden von Cassandro et al. (2002) durchgeführt wird.

Um weitere statistische Maße zur Trennung der untersuchten Sprachen in rhythmische Klassen zu finden, extrahieren Galves et al. (2002a) zwei Parameter, die eine lineare Relation der untersuchten Sprachen illustrieren sollen: Die Distanz zwischen dem ersten und dritten Quartil sowie die empirische Wahrscheinlichkeit für jede Sprache, einen geringeren Sonoritätswert als 0,3 zu erhalten. Duarte et al. (2001) finden eine ähnliche lineare Relation, indem sie die Standardabweichung der Gammaverteilung der Dauer der konsonantischen Intervalle bei Ramus et al. (1999) statistisch analysie-

ren. Diese drei linearen Relationen sind in Abb. 8 bis Abb. 10 dargestellt und lassen eine deutliche Trennung der Einzelsprachen in verschiedene Gruppen erkennen.



Abb. 8: Distanz zwischen dem ersten und dritten Quartil (aus Galves et al. 2002a). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

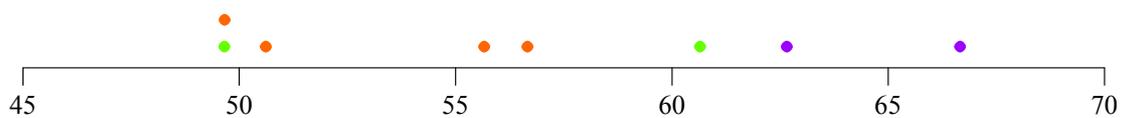


Abb. 9: Empirische Wahrscheinlichkeit, einen geringeren Sonoritätswert als 0.3 zu erhalten (aus Galves et al. 2002a). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

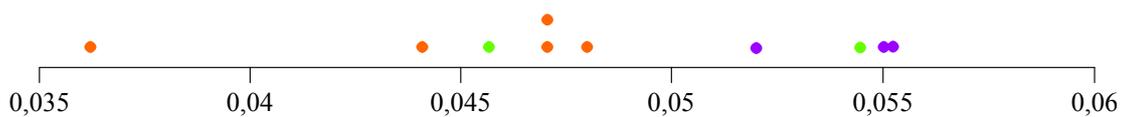


Abb. 10: Standardabweichung der Gammaverteilung der Dauer der konsonantischen Intervalle (aus Duarte et al. 2001). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umstrittene oder unklassifizierte Sprachen sind grün dargestellt.

In diesen Abbildungen, die die Sprachverteilung auf eine eindimensionale Darstellung reduzieren, ist jeweils eine grobe Gruppierung einzelner Sprachen zu erkennen, die sich von Japanisch (links) über die silbenzählenden Sprachen bis zu den akzentzählenden Sprachen erstreckt. Diese Untersuchungen sind womöglich stark von den Daten abhängig, scheinen für Galves et al. (2002a) jedoch die Hypothese zu stützen, daß die relevante Information für die rhythmische Sprachklassifikation anhand akustischer Parameter in den weniger sonoranten Anteilen des Sprachsignals enthalten ist.

Um die automatische Extraktion rhythmischer Parameter, die Galves et al. (2002a) verwenden, auszunutzen, wenden Cassandro et al. (2002) diese Methoden auf ein sehr großes Korpus von unaufbereiteten Sprachdaten an. Zuvor definieren sie allerdings die Sonoritätsfunktion aus Galves et al. (2002a) um:

$$S(t) = \exp \left\{ -\beta \sum_{i=1}^3 h(p_t / p_{t-i}) \right\}$$

wobei der Parameter  $\beta$  eine beliebige positive reelle Zahl sein kann. Die beste Trennung von rhythmischen Klassen wird bei einem Wert von  $\beta = 1,5$  erreicht. Mit Ausnahme der Definition von  $\delta S$ , die sich nur geringfügig ändert:

$$\delta S = \frac{1}{T-1} \sum_{t=1}^T |S_t - S_{t-1}|$$

bleiben alle übrigen Definitionen und Methoden aus Galves et al. (2002a) unverändert. Diese Methoden wenden Cassandro et al. (2002) auf ein Korpus von 2863 Sätzen an, in den Sprachen Catalan, Englisch, Fidschi, Finnisch, Französisch, Italienisch, Japanisch, Koreanisch, Niederländisch, Polnisch, Schwedisch, Spanisch, Türkisch, Ungarisch und Wolof. Es zeigt sich, daß sich die Aufnahmen in der  $\bar{S}/\delta S$ -Ebene in einer linear fallenden Formation anordnen. Diese Anordnung ist in Abb. 11 dargestellt.

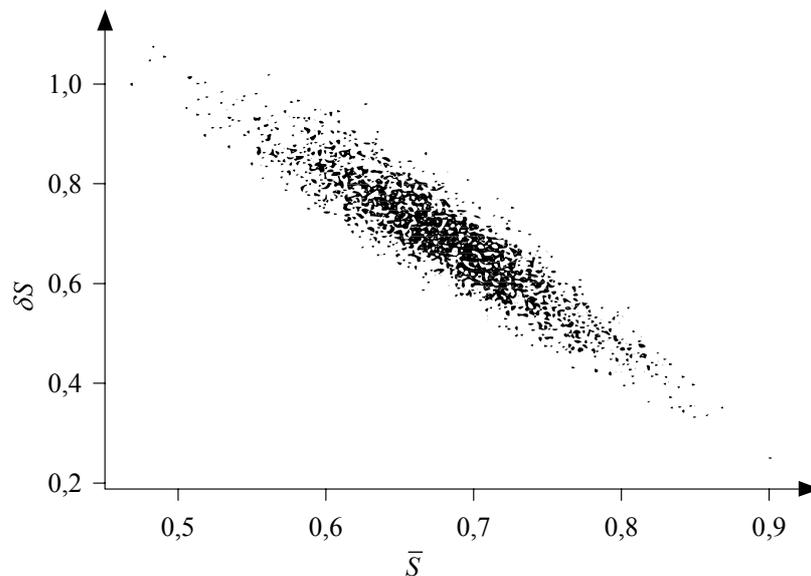


Abb. 11: Verteilung der 2863 Aufnahmen aus verschiedenen Sprachen in der  $\bar{S}/\delta S$ -Ebene (aus Cassandro et al. 2002).

Cassandro et al. (2002) nutzen diese lineare Relation von  $\bar{S}$  und  $\delta S$ , um ein stochastisches Modell zu erstellen. Dieses wiederum ordnet die untersuchten Einzelsprachen in der  $\bar{S}/\delta S$ -Ebene an, wobei sich eine Verteilung ergibt, die mit der aus Galves et al. (2002a) vergleichbar ist, dargestellt in Abb. 12. Allerdings liegt Japanisch nun in der Gruppe der silbenzählenden Sprachen, während Spanisch seine Lage stark in Richtung der akzentzählenden Sprachen verändert hat. Finnisch und Fidschi bilden die Ex-

trema in dieser Sprachverteilung, so daß alle übrigen Sprachen sich relativ dazu zusammen zu gruppieren scheinen. Auffallend an dieser Anordnung ist jedoch in erster Linie, daß sich die Sprachverteilung stark an die fallende lineare Relation anpaßt.

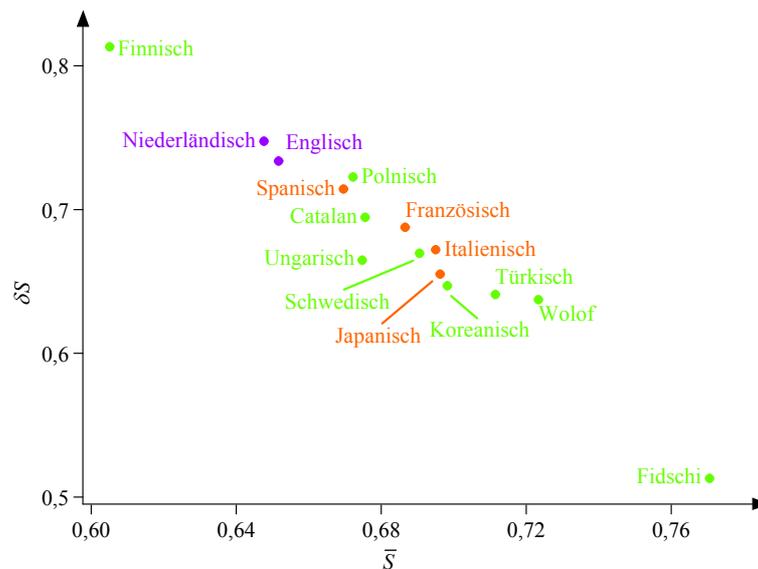


Abb. 12: Sprachverteilung in der  $\bar{S}/\delta S$ -Ebene (aus Cassandro et al. 2002). Akzentzählende Sprachen sind violett eingefärbt, silben- und morazählende orange. Umsprungen oder unklassifizierte Sprachen sind grün dargestellt.

Bei dieser vielversprechenden Entwicklung hin zur automatischen Extraktion rhythmischer Parameter sollte jedoch nicht außer Acht gelassen werden, daß selbst bei manuell aufbereiteten (von Hand segmentierten und annotierten) Daten noch nicht unumstritten ist, welche Parameter tatsächlich für Sprachrhythmus verantwortlich sind. Eine Vielzahl von Faktoren beeinflussen Daten und Methoden, so daß selbst bei einwandfrei implementierter automatischer Rhythmusklassifikation die theoretische Grundlage noch auf etwas wackeligen Füßen steht. Darüber hinaus beziehen sich alle neueren Arbeiten zu akustischen Rhythmuskorrelaten lediglich auf Dauer-Parameter, wobei nicht einmal einwandfrei geklärt ist, wie stark der Einfluß anderer akzentmarkierender Parameter wie Intensität oder Grundfrequenzbewegungen in verschiedenen Sprachen ist.

Zum Abschluß dieses Kapitels soll noch ein weiterer Ansatz erwähnt werden, Rhythmus maschinell zu modellieren, allerdings nur kurz. Es handelt sich dabei um ein

Modell, das unabhängig von Ramus et al. (1999) zur gleichen Zeit von O'Dell & Nieminen (1999) beschrieben und von Barbosa (2002) weiterentwickelt wurde, und das Sprachrhythmus durch ein Produktionsmodell zu beschreiben und erklären versucht. Dabei ergeben mehrere gekoppelte Oszillatoren mit verschiedenen Parametern mehrfrequente quasiperiodische Schwingungen. Dieser Ansatz erinnert an eine Art Rhythmus-Fouriertransformation, bei der nicht Glottisschläge, sondern Silben und Betonungsintervalle als Grundlage dienen. Das Prinzip scheint jedoch eine mathematische Umsetzung von Stetsons (1928) und Abercrombies (1967) Überlegungen zu „stress pulses“ und „chest pulses“ zu sein.

### 3 Motivation und Erwartung

Mit Ausnahme von Galves et al. (2002a) stützen sich die in Kapitel 2 beschriebenen Untersuchungen auf Parameter, die durch eine *C/V*-Segmentierung aus Aufnahmen gesprochener Sprache extrahiert werden. Mit einer solchen Segmentierung aber können unter Umständen wichtige Informationen verloren gehen, die ansonsten die Ergebnisse rhythmischer Analyse entscheidend beeinflussen. Ramus et al. (1999) erwähnen zwar die Problematik einer *C/V*-Segmentierung in einer Fußnote, doch wird diese implizite Fragestellung in keiner bislang erschienen Untersuchung, die eine *C/V*-Segmentierung verwendet, aufgegriffen:

We are aware, of course, that the consonant/vowel distinction may vary across languages, and that a universal consonant/vowel segmentation may not be without problems. We assume that our hypothesis should ultimately be formulated in more general terms, e.g. in terms of highs and lows in a universal sonority curve. We think, however, that for a first-order evaluation of our approach, and given the languages we consider here, such problems are not crucial. (Ramus et al. 1999)

Eines der erwähnten Probleme ist mit Sicherheit die Tatsache, daß manche Konsonantenklassen eher als andere einen Silbenkern bilden können. Dieses Phänomen wird durch die Sonoritätshierarchie beschrieben (Jespersen 1906<sup>4</sup>, Vennemann 1972<sup>5</sup>, Gussenhoven & Jacobs 1998), die in Abb. 13 dargestellt ist. Die dort verwendeten Abkürzungen werden in der vorliegenden Arbeit durchweg weiterverwendet. Die Sonoritätshierarchie formuliert die universelle phonologische Tendenz, daß die Sonorität der beteiligten Segmente vom Silbenanlaut zum Silbenkern hin zu-, und vom Silbenkern zum Silbenauslaut hin wieder abnimmt.

---

<sup>4</sup> Jespersen (1906) verwendet für Sonorität den Begriff „klangfylde“ (Schallfülle).

<sup>5</sup> Vennemann (1972) verwendet zwar die reziproke Skala der „konsonantischen Stärke“, doch das Prinzip ist dasselbe wie das der Sonoritätshierarchie (Pompino-Marschall 1995).

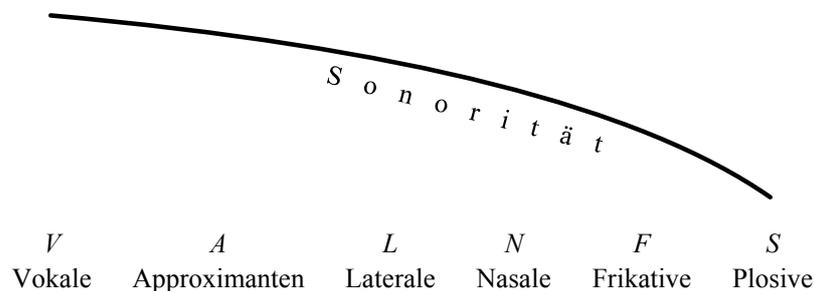


Abb. 13: Sonoritätshierarchie. Die eingezeichnete Kurve stellt den „Verlauf“ der Sonorität von Vokalen (hohe Sonorität) zu Plosiven (geringere Sonorität) dar.

Gleichzeitig besitzen alle Lautklassen eine inhärente Sonorität, die die Wahrscheinlichkeit für Segmente dieser Lautklasse wiedergibt, Silbenkerne bilden zu können. In der vorliegenden Arbeit wird von nun an der Begriff *Sonoritätsklasse* verwendet, der die gleiche Bedeutung hat wie *Lautklasse*, jedoch in Bezug auf die Sonoritätshierarchie.

Nicht jeder Silbenkern muß durch einen Vokal gebildet werden. In zahlreichen Sprachen gibt es Silben, deren Kern von Liquiden gebildet werden, d.h. von Lateralen, oder von Nasalen.<sup>6</sup>

In anderen Sprachen können solche Silben durch Reduktion, z.B. bei hoher Sprechgeschwindigkeit entstehen. Einige Beispiele:

Tschechisch (nach Bußmann 1990):

[str̩] prst skrs křk]

*Strč prst skrz křk.*

„Steck den Finger durch den Hals.“

<sup>6</sup> Ebenfalls sind Silben möglich, deren Silbenkern durch einen Frikativ gebildet wird, wie z.B. in „*pssst!*“, oder sogar silbische Plosive, wie sie im Berber vorkommen (Dell & Elmedlaoui 1985, s. jedoch Coleman 1999).

Swahili:

[m̩.fu'gaḏʒi̯ ,nimem'wu.a 'm̩.bwa wa 'mwitu]

*Mfugaji nimemwua mbwa wa mwitu.*

„Der Viehhalter hat den Schakal getötet.“

Deutsch:

[ʔiç 'ne·mōiç 'mit<sup>n</sup>ɪ̯ ʃtʏk<sup>h</sup> ʔi'māim̩ 'vɪntʃat<sup>n</sup>ɪ̯]

*Ich nehm euch mit ein Stück in meinem Windschatten.*<sup>7</sup>

Die Äußerung im ersten Beispiel enthält keinen Vokal; alle Silbenkerne werden durch den Lateral /r/ gebildet. Das Swahili enthält bei orthophoner Aussprache Silben, deren Kern durch einen Nasal gebildet wird, z.B. in der jeweils ersten Silbe der Worte *mfugaji* und *mbwa*. Im Deutschen ist bei weniger sorgfältiger Aussprache häufig zu beobachten, daß unbetonte Silben, etwa /ən/ oder /əm/, zu einem silbischen Nasal reduziert werden. Im angegebenen Beispiel ist dies sogar mit dem Wort *ein* geschehen.

Bei einer rhythmischen Untersuchung, die auf einer *C/V*-Segmentierung fußt und ihre Beschreibungsparameter aus der Dauer von *C*- und *V*-Intervallen gewinnt, müssen solche silbischen Konsonanten zwangsläufig Artefakte erzeugen, sofern sie undifferenziert zu *C*-Intervallen hinzugezählt werden. Bei einer solchen Segmentierung wäre als Extrembeispiel der oben angegebene tschechische Satz ein einziges *C*-Intervall. Dies würde für diesen Satz zu einem %*V*-Wert von 0 und einem nicht definierten  $\Delta C$ -Wert führen, was (unter der Annahme, daß sich das hypothetische tschechische Datenmaterial nicht auf diesen Satz beschränkt) jedes Ergebnis maßgeblich in der %*V*/ $\Delta C$ -Ebene verschieben würde. Daß dies den rhythmischen Eigenschaften der tschechischen Sprache nicht gerecht werden würde, steht außer Frage.

---

<sup>7</sup> So gesprochen von Thomas D in seinem Lied „Rückenwind“. Die Tatsache, daß diese Äußerung dort mit einem musikalischen Takt auftritt, ist hier nebensächlich; das Beispiel soll die silbischen Nasale hervorheben.

Wenn Sprachdaten für eine Untersuchung wie die von Ramus et al. (1999) segmentiert werden sollen, würde es vermutlich zu interessanten Beobachtungen führen, sie nicht nur in vokalische und konsonantische Intervalle zu segmentieren, sondern in eine Reihe von verschiedenen Intervalltypen, die sich z.B. an der Sonoritätshierarchie orientieren. Auf diese Weise wird zugleich eine gewisse Flexibilität in der Analyse erreicht, da man durch die Datenaufbereitung nicht auf eine bestimmte Segmentierung festgelegt ist. Eine mehr differenzierte Annotation als nach Sonoritätsklassen ist für eine solche Untersuchung vermutlich nicht sinnvoll, da eine Unterscheidung zwischen stimmhaften und stimmlosen Frikativen und Plosiven, sowie nach Artikulationsort, rhythmisch gesehen keine wesentlichen Veränderungen erwarten ließe.

Durch die Unterscheidung der Konsonanten nach Sonoritätsklasse können für manche Sprachen die Approximanten zu den „Vokalen“ gezählt werden, oder sogar die Laterale und Nasale. Dies mag zunächst widersinnig erscheinen, doch ist es im Rahmen einer Untersuchung, die Parameter zweier Intervalltypen für die Analyse nutzt, ein Ansatz, der die von Ramus et al. (1999) erwähnte Problematik zumindest teilweise umgeht.

Eine Untersuchung mit nach Sonoritätsklassen differenzierten Intervallen bietet zudem die Möglichkeit, hervorzuheben, wie stark die einzelnen Konsonantenklassen auf die rhythmischen Parameter einwirken. Da die Konsonantenklassen auf der Sonoritätshierarchie verschieden hoch angesiedelt sind, darf man annehmen, daß sie innerhalb eines *C*-Intervalls nicht homogen auf die rhythmischen Eigenschaften Einfluß nehmen, sondern manche Klassen mehr als andere.

Abgesehen davon ist das Verhältnis der einzelnen Konsonantenklassen in verschiedenen Sprachen unterschiedlich strukturiert, was nicht zuletzt auf die unterschiedliche Phonotaktik zurückzuführen ist. Auch hier ist also zu erwarten, daß eine Analyse rhythmischer Parameter durch die unterschiedliche Zuordnung von verschiedenen Sonoritätsklassen beeinflusst wird.

Um eine Zuordnung von einzelnen Sonoritätsklassen zu den „Vokalen“ oder „Konsonanten“ durchführen zu können, wird im folgenden die bekannte  $%V/\Delta C$ -Unterscheidung durch eine Unterscheidung zwischen  $%X$  und  $\Delta Y$  ersetzt, wobei  $X$  und  $Y$  Mengen von Sonoritätsklassen sind. Dabei werden die einzelnen Sonoritätsklassen auf  $X$  und  $Y$  aufgeteilt, so daß alle Intervalle einer Klasse entweder zu  $X$  oder zu  $Y$  gerechnet werden.  $%X$  ist in diesem Zusammenhang also definiert als der prozentuale Anteil aller Intervalltypen, die in  $X$  enthalten sind, an der Gesamtdauer einer Äußerung.  $\Delta Y$  ist definiert als die Standardabweichung der Dauer aller Intervalle, die aus den in  $Y$  enthaltenen Intervalltypen bestehen.

Dieser Abstraktionsschritt ermöglicht die Einführung von verschiedenen Anordnungen, die im folgenden als unterschiedliche *Konfigurationen* bezeichnet werden. Die Konfiguration, von der Ramus et al. (1999) ausgehen, ist in Abb. 14 dargestellt, die in der vorliegenden Arbeit der Notation  $%V/\Delta ALNFS$  entspricht:

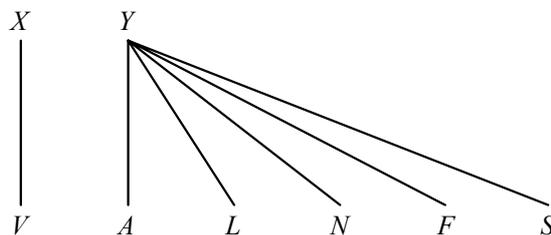


Abb. 14: Konfiguration der Sonoritätsklassen bei Ramus et al. (1999).

Für eine Untersuchung, die Approximanten zu den „Vokalen“ zählt, kann eine Konfiguration gewählt werden, in der  $A$ -Intervalle nicht zu  $Y$ , sondern zu  $X$  gezählt werden. Diese Konfiguration, notiert als  $%VA/\Delta LNFS$ , ist in Abb. 15 dargestellt:

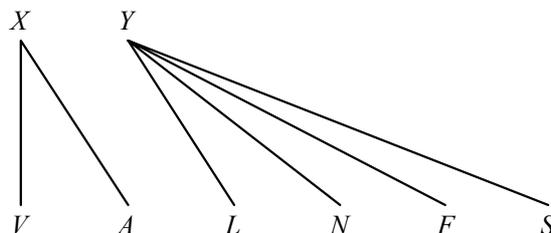


Abb. 15: Zuordnung der Sonoritätsklassen bei Konfiguration  $%VA/\Delta LNFS$ .

Auf diese Weise läßt sich auf der Sonoritätshierarchie die Grenze zwischen den Klassen, die in  $X$  enthalten sind, und denen, die zu  $Y$  gezählt werden, verschieben. So läßt

sich erkennen, wie stark sich die Verteilung der Einzelsprachen in der  $\%X/\Delta Y$ -Ebene von einer Konfiguration zur nächsten verändert. Das Ausmaß dieser Veränderung sollte, zumindest bei  $\%X$ , mit der Anzahl und Gesamtdauer der in einer Aufnahme enthaltenen Intervalle der entsprechenden Sonoritätsklasse korrelieren, und somit nicht gleichförmig sein, denn schließlich ist der Gesamtanteil der jeweiligen Sonoritätsklassen nicht gleich groß.

Die beiden vorangegangenen Konfigurationen sind jedoch nur zwei von vielen möglichen. Ebenso ist die in Abb. 16 dargestellte denkbar:

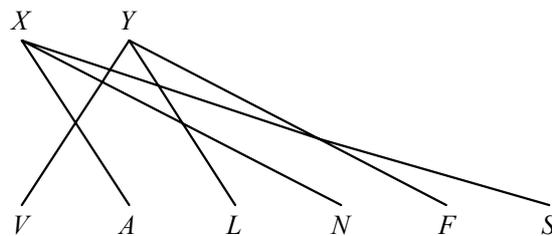


Abb. 16: Mögliche Konfiguration  $\%ANS/\Delta VLF$ .

Da  $\%X = 0$ , wenn  $X$  keine, und  $\%X = 100$ , wenn  $X$  alle Sonoritätsklassen zugeordnet werden, sind diese beiden Konfigurationen nicht sinnvoll und werden nicht in Betracht gezogen. Außerdem ist im letzteren Falle  $\Delta Y$  nicht definiert, da sich aus einer leeren Menge keine Standardabweichung berechnen läßt.

Damit beträgt also bei einer Unterscheidung in  $n$  Sonoritätsklassen die Menge  $k$  aller möglichen Konfigurationen

$$k_n = 2^n - 2$$

Für jede Konfiguration läßt sich dabei jeweils ein Wert für  $\%X$  und für  $\Delta Y$  berechnen, die als akustische Parameter einer rhythmischen Analyse betrachtet werden dürfen. Allerdings gilt innerhalb einer Konfiguration  $\%X = 100 - \%Y$ , und überhaupt sind die  $\%$ -Werte für alle Mengen  $X$ , die mehr als einen Intervalltyp beinhalten, abhängig von den  $\%$ -Werten der entsprechenden Intervalltypen. Das bedeutet, daß innerhalb einer Konfiguration  $k$  der prozentuale Anteil zweier Intervalltypen  $a$  und  $b$  an der Gesamtdauer zusammen gleich der Summe der prozentualen Anteile der jeweiligen Intervalltypen  $a$  und  $b$  für sich genommen ist:

$$%ab = \%a + \%b$$

Daraus folgt, daß für  $k_n - n$  der möglichen Konfigurationen die  $\%X$ -Parameter nicht unabhängig voneinander sind. Damit beträgt bei  $n$  Sonoritätsklassen die Menge  $p_n$  aller unabhängigen Parameter<sup>8</sup>

$$p_n = k_n + n$$

Eine statistische Analyse dieser Parameter sollte verdeutlichen, welche am stärksten für eine gegebene Verteilung der Einzelsprachen in der  $\%X/\Delta Y$ -Ebene verantwortlich sind. Auf diese Weise kann also nicht nur herausgefunden werden, daß die einzelnen Sonoritätsklassen unterschiedlich wichtig für eine Untersuchung wie Ramus et al. (1999) sind, sondern auch, wie sich diese Unterschiede ausprägen. Es ist darüber hinaus zu erwarten, daß diese Unterschiede für verschiedene Sprachen nicht gleich ausfallen, was nicht zuletzt auf die Phonetik und die Häufigkeit einzelner Sonoritätsklassen innerhalb jeder Sprache zurückzuführen sein dürfte.

Ein weiteres Problem, daß in bisherigen Arbeiten genannt wird, ist die Sprechgeschwindigkeit. Die Motivation für Methoden wie die von Grabe & Low (2002) geht auf die Vermutung zurück, daß Veränderungen der Sprechgeschwindigkeit Artefakte bei der Analyse der akustischen Rhythmusparameter erzeugen können:

...the standard deviation would reflect spurious variability introduced by changes in speaking rate within and across sentences and between-speaker differences in speaking rate. (Grabe & Low 2002)

Aus dieser Überlegung heraus sollten bei einer derartigen Untersuchung Daten verwendet werden, bei denen die Sprechgeschwindigkeit in irgendeiner Form normalisiert worden ist. Dazu sind verschiedene Methoden denkbar, die allerdings nicht eingesetzt werden können, ohne die Dauer einzelner Intervalle zu verändern.

---

<sup>8</sup> Durch die Vermeidung der Konfiguration mit leerem  $X$  wird der Parameter  $\Delta VALNFS$  nicht mit einbezogen. Strenggenommen ist dies nicht zwingend, aber dieser Parameter ist ohnehin nicht sinnvoll, da er lediglich die Standardabweichung der Gesamtdauer aller Äußerungen wiedergibt.

Bei der Erstellung von Datenmaterial selbst kann allerdings eine „Normalisierung“ der Sprechgeschwindigkeit stattfinden, die so simpel wie effektiv ist: es werden bei der Aufnahme von gesprochener Sprache einfach mehrere Aufnahmen mit verschiedener Sprechgeschwindigkeit durchgeführt. Da Spontansprache eine Reihe von schwer kontrollierbaren potentiellen Artefaktquellen enthält (Häsitationen, Veränderungen des genauen Wortlauts bei Wiederholungen, etc.), ist gelesene Sprache hierfür am besten geeignet.

Darüber hinaus gibt es einige Faktoren, die man bei der Datensammlung beachten sollte, um eine unkontrollierte Beeinflussung der rhythmischen Parameter zu vermeiden. Diese Faktoren als Artefaktquellen gilt es bei der Planung der Datensammlung auszuschließen bzw. zu minimieren.

- Geschlecht: Um von vornherein auszuschließen, daß mögliche geschlechtsspezifische Unterschiede die Daten beeinflussen, sollten etwa gleich viele Sprecherinnen wie Sprecher eingesetzt werden.
- Sprecher pro Sprache: damit die Unterschiede von Sprecher zu Sprecher nicht gleichzeitig die Unterschiede von Sprache zu Sprache sind, sollte für jede Sprache Material mehrerer Sprecher verwendet werden.
- Semantik: damit Unterschiede von Sprache zu Sprache nicht durch den Inhalt der Aufnahmen beeinflußt werden, sollte für jede Sprache ein möglichst gleicher Inhalt gewählt werden.
- Textlänge: je kürzer der gelesene Text, desto eher schleichen sich Unterschiede von Sprache zu Sprache ein, die auf die phonologische Zusammensetzung der im Text vorkommenden Worte zurückzuführen sind.

Bei einer Untersuchung, wie sie in der vorliegenden Arbeit beschrieben wird, ist eine Reihe von Fragestellungen zu untersuchen, die hier zusammenfassend aufgeführt sind. Dabei sollten zunächst zur Validierung der Datengrundlage und Methoden Vergleiche mit bisherigen Arbeiten durchgeführt werden.

- 1) Bei einer Untersuchung, die der von Ramus et al. (1999) nachempfunden ist, sollte die Verteilung der Einzelsprachen, die sowohl hier, als auch bei Ramus et al. untersucht werden (Englisch, Französisch und Italienisch), bei gleicher Konfiguration in der  $%X/\Delta Y$ -Ebene ähnlich sein.
- 2) Sprechgeschwindigkeit sollte einen deutlich erkennbaren Einfluß auf die  $%X/\Delta Y$ -Werte der einzelnen Aufnahmen nehmen, wie bei Dellwo & Wagner (2003) beschrieben. Allerdings bezieht sich jene Untersuchung nur auf die Ramus-Konfiguration  $%V/\Delta ALNFS$ .
- 3) Deutsch wurde bei Ramus et al. (1999) und ähnlichen Untersuchungen zwar nicht berücksichtigt,<sup>9</sup> dafür aber bei Dellwo & Wagner (2003), die Daten aus demselben Korpus verwenden wie die vorliegende Arbeit. Die  $%X/\Delta Y$ -Werte für das Deutsche sollten also mit den dortigen vergleichbar sein.

Nach diesen Voruntersuchungen kann differenzierter analysiert werden, welchen Einfluß die einzelnen Sonoritätsklassen auf die Verteilung der Einzelsprachen in der  $%X/\Delta Y$ -Ebene haben.

- 4) Wie bei mehreren bisherigen Untersuchungen festgestellt (Duarte et al. 2001, Galves et al. 2002), scheinen die einzelnen Lautklassen für die Verteilung der Einzelsprachen in der  $%X/\Delta Y$ -Ebene nicht gleichermaßen verantwortlich zu sein. Sobald einzelne Lautklassen nach Sonorität differenziert werden, sollten sich demnach die  $%X/\Delta Y$ -Werte jeder Einzelsprache mit verschiedenen Konfigurationen der Sonoritätsklassen nicht gleichmäßig verändern.
- 5) Wenn Einzelsprachen aufgrund ihrer  $%X/\Delta Y$ -Werte je nach Konfiguration in Gruppen getrennt werden können, die den rhythmischen Kategorien akzentzählend und silbenzählend entsprechen, dann sollten

---

<sup>9</sup> Deutsch wurde zwar von Grabe & Low (2002) untersucht, doch ist ihre Studie nicht ohne weiteres mit den hier verwendeten Daten und Methoden vergleichbar.

in diesen Gruppen bei der Veränderung von einer Konfiguration zur nächsten die Unterschiede zwischen Sprachen mit verschiedenen rhythmischen Eigenschaften größer sein als zwischen Sprachen mit ähnlichen rhythmischen Eigenschaften.

- 6) Aufgrund der einzelsprachlichen Phonotaktik sowie der unterschiedlichen Vorkommensverteilung verschiedener Sonoritätsklassen in der jeweiligen Sprache sollten sich die  $%X/\Delta Y$ -Werte dieser Sprachen bei verschiedenen Konfigurationen jeweils unterschiedlich voneinander verändern.
- 7) Die Menge der Positionen in der  $%X/\Delta Y$ -Ebene bei verschiedenen Konfigurationen sollte für jede Einzelsprache, verbunden mit ihrer Phonotaktik, ein für diese Sprache charakteristisches Muster ergeben, das bei gleichen Bedingungen keiner anderen Sprache zugeordnet werden kann.

Wie die vorliegende Arbeit diese Fragen beantwortet und welche Hypothesen von den Ergebnissen bestätigt werden, wird in Kapitel 4 erläutert.

## 4 Daten und Methoden

Für diese Arbeit wurden aus dem BonnTempo-Korpus<sup>10</sup> (Dellwo & Wagner 2003, Dellwo, Aschenberner, Steiner & Wagner, in Vorbereitung) einige Aufnahmen ausgewählt, in den Sprachen Deutsch, Englisch, Französisch und Italienisch. Deutsch und Englisch gelten traditionell als akzentzählende, Französisch und Italienisch als silbenzählende Sprachen (s. Tabelle 1). Deutsch ist von diesen Sprachen die einzige, die in den meisten bisherigen Arbeiten noch nicht berücksichtigt worden ist, doch kann man davon ausgehen, daß es sich hinsichtlich seiner rhythmischen Eigenschaften ähnlich verhält wie andere akzentzählende Sprachen, z.B. das Niederländische, das bisher durchaus wiederholt untersucht worden ist.

Aufgrund der Tatsache, daß das BonnTempo-Korpus von jedem Sprecher fünf Aufnahmen in verschiedenen Sprechgeschwindigkeiten enthält, war zu erwarten, daß die Unterschiede zwischen schneller und langsamer Sprache normalisiert und Artefakte, die auf verschiedene Sprechgeschwindigkeit der einzelnen Sprecher zurückzuführen sind, auf ein Minimum reduziert werden. In Verbindung mit der Tatsache, daß für jede Sprache Aufnahmen von mehreren Sprechern verschiedenen Geschlechts untersucht wurden und daß der Inhalt der Lesetexte in allen Sprachen nahezu identisch war, können der Einfluß von Sprechgeschwindigkeit, Sprache-Sprecher-Identität, Geschlecht und Inhalt der Aufnahmen als für die in Kapitel 5 aufgeführten Ergebnisse unerhebliche Faktoren betrachtet werden.

### 4.1 Sprecher

Für die vorliegende Arbeit wurde für jede Sprache eine kleine Anzahl von Sprechern ausgewählt, wobei versucht wurde, die Anzahl an weiblichen und männlichen Spre-

---

<sup>10</sup> Aufgrund von Platzmangel beschränkt sich die Beschreibung der frühen Fassung des BonnTempo-Korpus bei Dellwo & Wagner (2003) auf die wesentlichen Details. In der vorliegenden Arbeit finden sich jedoch ausführliche Angaben zur Erstellung und Zusammensetzung dieses Korpus.

chern im Gleichgewicht zu halten. Die Verteilung der Sprecher auf die einzelnen Sprachen sowie ihre Gesamtzahl ist in Tabelle 2 erläutert.

	Deutsch	Englisch	Französisch	Italienisch	Gesamt
Anzahl ♀	2 ♀	3 ♀	3 ♀	1 ♀	9 ♀
Anzahl ♂	2 ♂	2 ♂	2 ♂	1 ♂	7 ♂
Aufnahmen	20	25	25	10	80

Tabelle 2: Anzahl der Sprecherinnen und Sprecher sowie der Aufnahmen und Verteilung auf die untersuchten Sprachen.

Das Alter der Sprecher liegt zwischen 20 und 35 Jahren. Alle haben die Hochschulreife erreicht, viele besitzen einen Hochschulabschluß. Bei keinem der Sprecher konnte zum Zeitpunkt der Aufnahme eine gesundheitlich bedingte Beeinträchtigung der Artikulation festgestellt werden.

Die deutschen Sprecher, größtenteils Studenten oder Mitarbeiter des Instituts für Kommunikationsforschung und Phonetik, kommen aus verschiedenen Teilen Deutschlands, verstärkt aus Nordrhein-Westfalen, doch konnte bei keinem eine auffällige dialektale oder regiolektale Färbung der Aussprache festgestellt werden.

Ein Teil der englischen Sprecher lebt in Edinburgh und Umgebung und spricht Englisch mit einem schottischen Einschlag. Die übrigen kommen aus dem mittleren Westen der USA und sprechen die dort vorherrschende Standardvarietät des Englischen.

Die französischen Sprecher kommen aus dem Südwesten Frankreichs und sprechen ebenfalls die dort vorherrschende Standardvarietät.

Die italienischen Sprecher sprechen Italienisch mit kalabrischem bzw. umbrischem Einschlag.

## 4.2 Material

Das Material des BonnTempo-Korpus besteht aus gelesener Sprache. Es handelt sich dabei um einen kurzen Text von ca. 80 Silben, der in jeder Aufnahme vom jeweiligen

Sprecher vorgelesen wird.<sup>11</sup> Auf diese Weise wurde gewährleistet, daß prosodische Unterschiede zwischen den verschiedensprachlichen Aufnahmen nicht durch semantische Unterschiede verursacht werden konnten. Da der Lesetext jedoch eine Reihe von Ortsnamen enthält, wurden diese bei der Übersetzung durch entsprechende Ortsnamen im Ursprungsland der Zielsprache ersetzt.

Die Lesetexte sind im folgenden wiedergegeben:

Deutsch:

„Am nächsten Tag fuhr ich nach Husum. Es ist eine Fahrt ans Ende der Welt. Hinter Gießen werden die Berge und Wälder eintönig, hinter Kassel die Städte ärmlich und bei Salzgitter wird das Land flach und öde. Wenn bei uns Dissidenten verbannt würden, würden sie ans Steinhuder Meer verbannt.“

Englisch:

„The next day I went to Falmouth. It is a voyage to the end of the world. After Lincoln the hills and woods become monotonous, after Bristol the towns get boring and near Saintsbury the countryside becomes flat and desolate. If dissidents were banned in our country, they would be banned to the Portishead Bay.“<sup>12</sup>

---

<sup>11</sup> Der deutsche Text ist ein Auszug aus dem Roman *Selbs Betrug* von Bernhard Schlink (Diogenes Verlag, 1994, S. 242). Dieser Auszug wurde von Dellwo & Wagner (2003) in die im Korpus vorkommenden Sprachen übersetzt, mit Beratung und Hilfe von Muttersprachlern der jeweiligen Zielsprache.

<sup>12</sup> Zwar wäre *exiled* oder *banished* eine bessere Übersetzung für das deutsche *verbannt*, doch ist dies weder bei der Übersetzung des Lesetextes ins Englische, noch während der Aufnahmen einem der Sprecher aufgefallen. *Saintsbury* sollte ursprünglich *Salisbury* heißen, doch dieser Unterschied wirkt sich in keiner Weise auf die Ergebnisse der vorliegenden Arbeit aus.

Französisch:

„Le jour suivant, je me suis rendue à Albi. C’est un voyage au bout du monde. Après Lisieux, les montagnes et la forêt deviennent monotones, après Châtel, les villes désolées, et près de Chartreville, la campagne devient plate et déserte. Si chez nous les dissidents étaient exilés, ils seraient alors exilés à Clermont-Ferrand.“

Italienisch:

„Il giorno dopo andai a Bologna. È un viaggio fino alla fine del mondo, dopo Rovereto i colli e i boschi diventano monotoni, dopo Verona le città diventano misere, e presso Revere il paesaggio diventa pianeggiante e deserto. Se da noi un dissidente venisse esiliato, verrebbe esiliato a Ostiglia.“

Da bei den Aufnahmen für das BonnTempo-Korpus jeder Sprecher gebeten wurde, nicht nur in seiner Muttersprache, sondern gegebenenfalls auch in weiteren Sprachen, derer er in akzeptablem Maße mächtig war, den entsprechenden Text vorzulesen, enthält das Korpus zahlreiche Aufnahmen (vor allem in deutscher, englischer und französischer Sprache), bei denen der jeweilige Sprecher zweitsprachliche bzw. fremdsprachliche Kompetenz in der Zielsprache aufweist. Darüber hinaus wurden aber auch einige Aufnahmen mit bilingualen Sprechern durchgeführt.

Für die vorliegende Arbeit wurde darauf verzichtet, nichtmuttersprachliches Material zu verwenden, da zu befürchten war, daß sowohl die Fluidität der Sprache als auch gegebenenfalls rhythmische Unterschiede zwischen Erst- und Zielsprache der jeweiligen Sprecher starke Einflüsse auf die zu untersuchenden rhythmischen Parameter der Sprachdaten ausüben würden. Aus ähnlichen Gründen wurde auch vermieden, Aufnahmen von bilingualen Sprechern zu verwenden, da hier ebenfalls mögliche Artefak-

te durch rhythmische Unterschiede zwischen den Muttersprachen verursacht werden könnten.<sup>13</sup>

### **4.3 Ablauf der Aufnahmen**

Alle Aufnahmen im BonnTempo-Korpus wurden im schalltoten Raum des Instituts für Kommunikationsforschung und Phonetik der Universität Bonn mit einem Kondensatormikrophon durchgeführt. Die Sprachsignale wurden digital mit einer Abtastrate von 44,1 kHz bei 16 Bit direkt auf Festplatte gespeichert.

Für jeden Sprecher lief der Aufnahmevorgang folgendermaßen ab. Der Aufnahmeleiter legte dem Sprecher zunächst den der Muttersprache des Sprechers entsprechenden Lesetext vor und bat ihn, sich mit ihm vertraut zu machen, um ein Stocken oder unregelmäßiges Vorlesen während der Aufnahme zu vermeiden. Sobald sich der Sprecher dazu bereit fühlte, wurde mit der ersten Aufnahme begonnen. Die Versuchspersonen wußten zwar, daß sie den Text vorlesen sollten, aber darüber, daß ihre Sprechgeschwindigkeit dabei eine Rolle spielen würde, wurden sie zu Beginn der Aufnahmen nicht aufgeklärt. Der Aufnahmeleiter gab lediglich die Anweisung, den Text „ganz normal“ vorzulesen. Anschließend wurde der Sprecher gebeten, denselben Text für die zweite Aufnahme langsamer vorzulesen. Für die dritte Aufnahme sollte der Sprecher den Text noch langsamer vorlesen. Dann bat der Aufnahmeleiter den Sprecher, den Text ein wenig schneller als normal vorzulesen, was die vierte Aufnahme ergab. Darauf folgte eine Reihe von weiteren Aufnahmen, von denen der Sprecher seine Sprechgeschwindigkeit mit jeder Aufnahme steigern sollte. War der Sprecher nicht mehr in der Lage, schneller vorzulesen, oder war die letzte Aufnahme durch starke Reduktion zu undeutlich oder wurden zu viele Silben getilgt, wurde die Aufnahmereihe beendet. Beherrschte der Sprecher neben der gerade aufgezeichneten noch eine

---

<sup>13</sup> Prinzipiell ist die Untersuchung von Rhythmus im Fremdsprachenerwerb bzw. bei bilingualen Sprechern, insbesondere bei großen rhythmischen Unterschieden zwischen den Muttersprachen, ein sehr interessantes Gebiet, das noch relativ unerforscht ist, zumindest mit den Methoden, die in den in Kapitel 2 genannten unlängst erschienenen Arbeiten vorgeschlagen worden sind. Siehe hierzu auch Abschnitt 6.4.

weitere Sprache, wurde die Prozedur nach einer kurzen Pause in der entsprechenden Sprache noch einmal durchgeführt, ansonsten war die Aufnahme nun beendet.

#### **4.4 Aufbereitung der Daten (1)**

Pro Sprecher und Sprache wurden fünf Aufnahmen für die Weiterverarbeitung ausgewählt. Dazu gehörten die normale Sprechgeschwindigkeit, die langsamere und die langsamste, sowie die erste schnelle Version. Als fünfte Version wurde von den übrigen Aufnahmen die mit der höchsten Sprechgeschwindigkeit ausgesucht, bei der keine Versprecher oder übermäßige Silbenreduktion auftraten.

Ziel der weiteren Verarbeitung dieser jeweils fünf Aufnahmen war die Segmentierung in Silben einerseits, und andererseits in konsonantische und vokalische Intervalle. Diese Annotation wurde manuell von einer studentischen Hilfskraft vorgenommen, die mit solcher Arbeit bereits vertraut war und in dieser Hinsicht genügend Erfahrung besaß. Die Segmentierung und Etikettierung erfolgte mit Hilfe der Software *Praat* (Boersma & Weenink 2001). Es war aufgrund eingeschränkter Mittel nicht möglich, mehrere Hilfskräfte unabhängig voneinander an den gleichen Daten arbeiten zu lassen. Sonst hätte die Zuverlässigkeit der Ergebnisse anhand der Übereinstimmung zwischen den Annotatoren besser eingeschätzt werden können und bei Passagen mit geringer Übereinstimmung gegebenenfalls durch Hinzunahme weiterer Annotatoren eine zuverlässigere Annotation erzeugt werden können. Allerdings erscheinen die Ergebnisse der manuellen Datenaufbereitung die Ansprüche durchaus zu erfüllen, die durch bisherige Verwendung des BonnTempo-Korpus an sie gestellt worden sind. Abgesehen davon wurden einige Aufnahmen zur Übereinstimmung von Dellwo & Wagner (2003) doppelt annotiert, und für die Messungen, die in der vorliegenden Arbeit Verwendung finden, wurde die Annotation durch den Autor noch einmal überprüft (siehe Abschnitt 4.6).

Für die Segmentierung dienten sowohl auf Silben-, als auch auf Konsonant/Vokal-Ebene auditive und visuelle Anhaltspunkte (Spektrogramm mit Intensitätskurve, Grundfrequenzkontur und Formantverläufen sowie Oszillogramm mit Glottisschlä-

gen). Dabei wurden einige Konventionen getroffen, um in Zweifelsfällen konsistent zu annotieren:

- es wurde primär anhand phonetischer Gesichtspunkte eng transkribiert, sekundär anhand phonologischer und lexikalischer Gesichtspunkte;
- Okklusionsphasen von Plosiven wurden nicht als Pausen markiert, sondern zum Plosivintervall hinzugezählt, sofern sie nicht eine Dauer von 50 ms überschritten, z.B. in äußerungsinitialer Position.

Die Segmentierung auf Silbenebene wurde orthographisch etikettiert; auf der Konsonant/Vokalebene wurden nach akustisch-phonetischen Kriterien Vokale mit  $\langle V \rangle$  und Konsonanten mit  $\langle C \rangle$  markiert. Pausen wurden auf beiden Ebenen mit  $\langle P \rangle$  gekennzeichnet.

Als Ergebnis dieser Datenaufbereitung besteht das BonnTempo-Korpus aus PCM-Wave-Dateien (im unter Abschnitt 4.3 genannten Format) für die einzelnen Aufnahmen, und aus *Praat*-TextGrid-Dateien, davon jeweils fünf pro Sprecher und Sprache. Diese TextGrid-Dateien enthalten jeweils zwei IntervalTier-Objekte, eins für die Silbenebene, und eins für die Konsonant-Vokal-Ebene. Letzteres enthält die Etikette  $\langle C \rangle$ ,  $\langle V \rangle$  und  $\langle P \rangle$  und wird sowohl im BonnTempo-Korpus als auch im folgenden Text als *C/V-Ebene* bezeichnet. Es umfaßt somit die diskrete Segmentierung des zugehörigen Sprachsignals in Intervalle und gibt für jedes dieser Intervalle den genauen Anfangs- und Endzeitpunkt sowie den Intervalltyp an.

#### **4.5 Exkurs: Maschinelle Annotation**

Durch den hohen Arbeitsaufwand einer manuellen Annotation und die Willkür, die mit einem solchen Prozeß verbunden ist, wäre es wünschenswert gewesen, die Rohdaten des BonnTempo-Korpus automatisch weiterverarbeiten zu können. Diese Vorgehensweise hätte einige substantielle Vorteile gegenüber der Annotation durch menschliche Urteilskraft:

- mit heute gängiger Rechenleistung sind Ergebnisse schon nach Minuten, nicht erst nach tage- oder gar wochenlanger Arbeit zu erwarten;
- durch diesen enormen Zeitgewinn können vielfach größere Datenmengen aufbereitet und in einem Sprachdatenkorpus zur Verfügung gestellt werden;
- automatische Annotation arbeitet präzise und konsequent, d.h. Daten werden immer nach exakt denselben Maßstäben verarbeitet, und es bleibt kein Spielraum für Interpretation oder „schwierige“ Entscheidungen;
- im Gegensatz zu menschlichen Annotatoren wird die Verarbeitung nicht durch eventuelle Erwartungen oder Weltwissen beeinflusst, da die maschinelle Verarbeitung ausschließlich auf Informationen im Sprachsignal zugreift.
- Hinzu kommt der finanzielle Aspekt, denn ein bereits entwickeltes und einsatzbereites automatisches Annotationssystem erzeugt so gut wie keine Betriebskosten und ist somit sehr viel günstiger als menschliche Annotatoren, die schließlich auf Stundenbasis bezahlt werden müssen.

Aus den soeben genannten Gründen wurde bei den Vorbereitungen der vorliegenden Arbeit eine automatische Annotation erprobt. Es handelt sich dabei um das System *Vocale* (Garcia et al. 2002), das am Institut für Mathematik und Statistik der Universität São Paulo zur halbautomatischen Annotation für Prosodieforschung entwickelt wurde. Dieses System nimmt unverarbeitete Sprachdaten mit einer Abtastfrequenz von 11kHz als Eingabe und generiert dazu *Praat*-TextGrid-Dateien, die jeweils ein IntervalTier-Objekt enthalten, das das entsprechende Sprachsignal in vokalische (<V>) und konsonantische (<C>) Intervalle, sowie Sprechpausen (<P>) segmentiert. *Vocale* ist also für exakt die gleiche Datenverarbeitung konzipiert, deren manuelle Durchführung im vorangehenden Abschnitt geschildert wurde. Insofern lag es nahe, *Vocale* auch auf die Rohdaten im BonnTempo-Korpus anzuwenden, um auch die

Aufnahmen, die zu dem Zeitpunkt zwar aufgenommen, aber noch nicht annotiert worden waren, für die vorliegende Arbeit verwenden zu können.

*Vocale* besteht aus einem frei verfügbaren *Perl*-Skript, das seinerseits auf *Praat* zurückgreift, und so zu den zu annotierenden Sprachdaten Spektrogramme erzeugt, die als Matrix-Objekte weiterverarbeitet werden. Diese Weiterverarbeitung erfolgt in einer Reihe von Schritten, die dazu gedacht sind, verschiedene Lautklassen phonetisch zu identifizieren:

- Sprechpausen werden anhand der Gesamtenergie jedes Fensters ermittelt;
- Vokale werden identifiziert, indem nach Segmenten mit geringer relativer Entropie im Bereich von 0kHz bis 1kHz und hoher Energie im Bereich von 0kHz bis 5kHz gesucht wird, bei einer Minstdauer von 16ms;
- als Konsonanten gelten die Segmente mit einer Minstdauer von 8ms, die eine hohe relative Entropie zwischen 0kHz und 1kHz aufweisen.

Anhand weiterer Regeln werden Nasale und stimmlose Frikative und Plosive identifiziert und zu den Konsonanten gezählt. Schließlich werden die so gewonnenen Annotationen in TextGrid-Dateien ausgegeben, die von *Praat* gelesen werden können.

Garcia et al. (2002) vergleichen die auf diese Weise gewonnenen Annotationsdaten mit von Hand segmentierten Annotationen und stellen eine Übereinstimmung fest, die je nach Sprecher zwischen 49,9% und 67,8% (Vokale), 56% und 88,9% (Konsonanten) und 71,1% und 97,4% (Pausen) liegt. Trotz dieser Werte verwenden sie *Vocale*-generierte Annotationen, um für die Daten aus Ramus et al. (1999) %*V* und  $\Delta C$ -Werte zu generieren und diese wie in jener Arbeit gegeneinander aufzutragen. Die sich daraus ergebenden Verteilungsgraphen unterstützen grob die Gruppierung in akzentzählende, silbenzählende und morazählende Sprachen.

Bei der Anwendung auf die BonnTempo-Daten traten allerdings sehr hohe Abweichungen von den manuellen Annotationen auf. Die aus der Verwendung von *Vocale* resultierenden  $\%V$  und  $\Delta C$ -Werte sind in Abb. 17 dargestellt, während Abb. 18 die entsprechenden Werte der gleichen Daten unter Verwendung der manuellen Annotationen zum Vergleich darstellt. Bei beiden Darstellungen wurde der gleiche Maßstab gewählt, um die Unterschiede in der Verteilung hervorzuheben. Die Farben kennzeichnen jeweils die Sprache jeder Aufnahme, Deutsch (olivfarben), Englisch (blau), Französisch (rot) und Italienisch (grün).

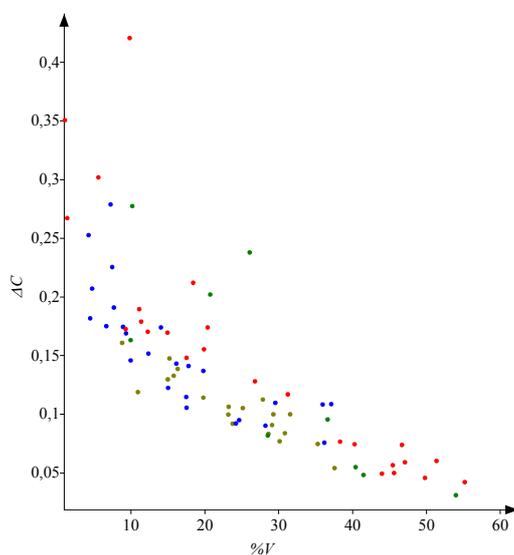


Abb. 17: Verteilung der einzelnen Aufnahmen in der  $\%V/\Delta C$ -Ebene aufgrund der *Vocale*-annotierten Daten.

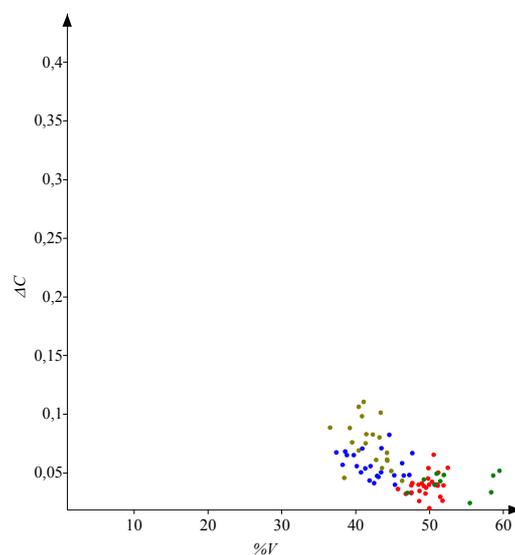


Abb. 18: Verteilung der einzelnen Aufnahmen in der  $\%V/\Delta C$ -Ebene aufgrund der manuell annotierten Daten.

Wie man sieht, sind die  $\%V$  und  $\Delta C$ -Werte bei der automatischen Segmentierung relativ gleichmäßig über den dargestellten Ausschnitt der  $\%V/\Delta C$ -Ebene verteilt. Die von Cassandro et al. (2002) beschriebene lineare Relation bei der Verteilung ist gut zu erkennen. Jedoch ist keinerlei Gruppierung von Aufnahmen derselben Sprache zu erkennen, wie sie nach Ramus et al. (1999) zu erwarten wäre. Die Verteilung der manuell segmentierten Daten hingegen ist sehr viel kompakter und läßt trotz einiger Überlagerung erkennen, wie sich die einzelnen Aufnahmen je nach Sprache in der  $\%V/\Delta C$ -Ebene zusammenhäufen.

Eine nähere Untersuchung der von *Vocale* erzeugten Annotationsdaten belegt, daß die Unterschiede zu manuell erstellten Annotationen gewaltig sind. Die durchschnittliche Abweichung der von *Vocale* generierten %*V*-Werte liegt bei 552,4% (!), mit einer Standardabweichung von 1852,2. Bei den  $\Delta C$ -Werten sieht es vergleichsweise besser aus, die Abweichung beträgt hier „lediglich“ 48,0%, mit einer Standardabweichung von 23,2. Diese Zahlen bestätigen den Eindruck, daß von *Vocale* generierte Annotationen für die vorliegende Arbeit weitestgehend nicht verwendbar sind. Ein typisches Beispiel aus der Datengrundlage ist in Abb. 19 dargestellt und veranschaulicht die starke Diskrepanz zwischen der automatisch generierten Annotation und der von Hand erstellten.

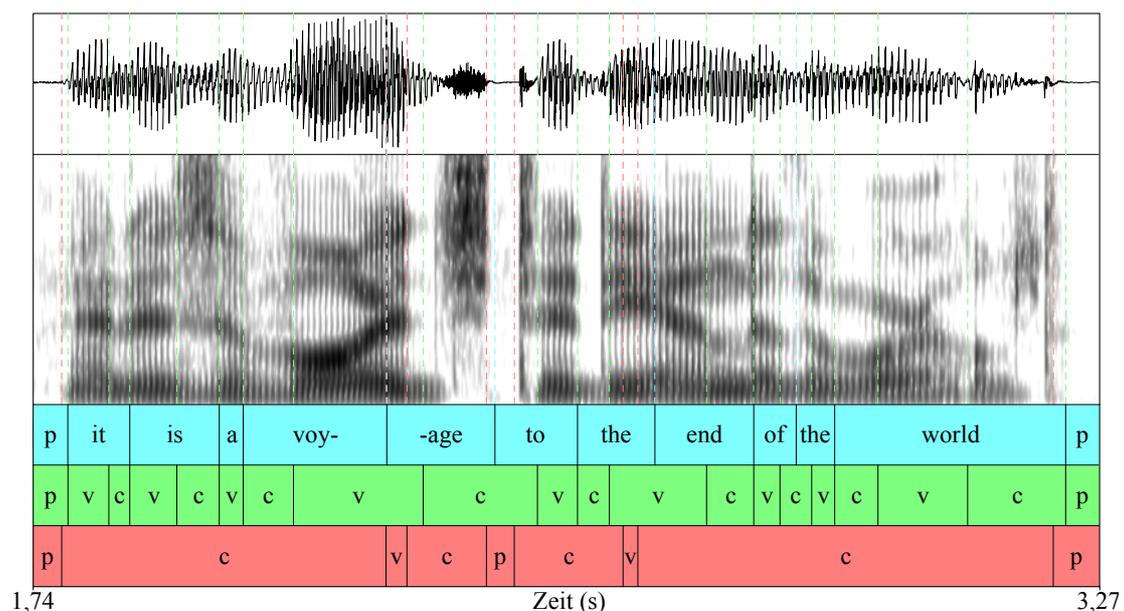


Abb. 19: Ausschnitt aus einer englischen Aufnahme aus dem BonnTempo-Korpus mit normaler Sprechgeschwindigkeit, gesprochen wurde die Sequenz „it is a voyage to the end of the world“ (s. Abschnitt 4.2). Der hellblau unterlegte *IntervalTier* enthält die orthographische Repräsentation, in Silben segmentiert. Der grün unterlegte *IntervalTier* enthält die manuelle *C/V*-Segmentierung des BonnTempo-Korpus, während der rot unterlegte *IntervalTier* die von *Vocale* automatisch generierte Segmentierung darstellt. Die gestrichelten farbigen Trennlinien kennzeichnen die Intervallgrenzen in Oszillogramm und Spektrogramm.

Bei diesem direkten Vergleich läßt sich beobachten, wie mehrere Silben automatisch zu einem konsonantischen Intervall zusammengefaßt wurden und die wenigen als vokalisch etikettierten Intervalle viel kürzer sind als die dazugehörigen, im Spektrogramm erkennbaren Formantverläufe.

Wie kommt es zu diesen enormen Abweichungen der automatisch gewonnen Annotationen von den manuellen? Zwei Faktoren allein können schon eine ausreichende Erklärung dafür liefern.

Erstens ist *Vocale* noch nicht ausgereift. Wie Garcia et al. (2002) betonen, ist die Implementierung ihres halbautomatischen Annotationssystems noch längst nicht zufriedenstellend. Zwar ist eine automatische Unterscheidung in vokalische und nichtvokalische Segmente eines Sprachsignals unter günstigen Bedingungen prinzipiell nicht allzu komplex. Wie oben erläutert, sind erforderliche Parameter wie Quasiperiodizität und entsprechende spektrale Eigenschaften relativ leicht zu entdecken. Allerdings ist für die Verwendbarkeit der Ergebnisse dieses Vorgangs ein hohes Maß an Zuverlässigkeit entscheidend. Wenn schon mehrere Fehler bei der manuellen Annotation zu ausgeprägten Artefakten bei Ergebnissen von Messungen prosodischer Eigenschaften führen können, dann sind systematische Fehlleistungen bei einer maschinellen Verarbeitung fatal. So wenig komplex die zu erbringenden Ergebnisse bei dieser Form von Datenaufbereitung sein mögen, für die Verwendbarkeit dieser Ergebnisse muß das Annotationssystem sehr robust sein.

Dies führt direkt zum zweiten Faktor, der vermutlich in diesem Fall entscheidend zum Versagen von *Vocale* beitragen hat. Die Daten selbst sind es, die gerade bei unzureichender Robustheit des Annotationssystems gravierende Fehlleistungen bei der Verarbeitung auslösen. Zwar ist die Qualität der Aufnahmen im BonnTempo-Korpus hoch, doch die Eigenschaft, die Systeme wie *Vocale* zu Fall bringt, ist die variierende und zum Teil sehr hohe Sprechgeschwindigkeit. Schnell gesprochene Sprache enthält weitaus mehr Reduktionsphänomene wie Koartikulation als mit normaler Sprechgeschwindigkeit gesprochene, und jeder, der schon einmal versucht hat, solche Sprachsignale zu segmentieren, weiß, daß sie erheblich schwerer zu segmentieren ist als langsam gesprochene. Daß maschinelle Annotationssysteme sich mit solchen Daten mindestens genauso schwer, wenn nicht viel schwerer tun, dürfte einleuchten.

Aus diesen Gründen wurde auf die Verwendung von automatisch generierten Annotationen verzichtet. Darüber hinaus erschien der Aufwand, ein robustes Autosegmentierungssystem zu erwerben oder zu entwickeln, insbesondere bei der geringen Größe

des BonnTempo-Korpus, unverhältnismäßig hoch gegenüber der manuellen Annotation durch erfahrene Phonetiker.

#### **4.6 Aufbereitung der Daten (2)**

Kehren wir nach diesem Exkurs wieder zurück zu den manuell verarbeiteten Daten, wie sie in Abschnitt 4.4 beschrieben wurden.

Für die vorliegende Arbeit war es notwendig, differenzierter segmentierte Sprachdaten zu verwenden, als sie das BonnTempo-Korpus bereitstellt. Da sich jedoch dieses Korpus aus den zu Beginn von Kapitel 4 genannten Gründen für diese Art von rhythmischer Untersuchung anbot, wurde in Kauf genommen, die Annotation der Sprachdaten weiterzuverarbeiten. Aufgrund des damit verbundenen Arbeitsaufwandes war jedoch die Anzahl der Aufnahmen begrenzt, die für diese Untersuchung verwendet werden konnten.

Die Daten der in Abschnitt 4.1 genannten Auswahl von Sprechern aus dem BonnTempo-Korpus wurde also vom Autor der vorliegenden Arbeit manuell weiterverarbeitet. Dabei wurde die bereits vorhandene Annotation als Grundlage verwendet, so daß dieser Arbeitsschritt im wesentlichen darin bestand, die als <C> etikettierten Intervalle anhand der Sprachdaten entweder bezüglich ihrer Sonoritätsklasse neu zu etikettieren, oder gegebenenfalls in mehrere Intervalle zu segmentieren und diese entsprechend zu kennzeichnen. Als Anhaltspunkte dienten wiederum auditive und visuelle Informationen des Sprachsignals (s. Abschnitt 4.4).

Bei der Neusegmentierung wurden grundsätzlich die Konventionen eingehalten, die schon bei zur Bearbeitung der Rohdaten des BonnTempo-Korpus verwendet worden waren, mit einigen Zusätzen:

- Approximanten wurden nur prävokalisch oder intervokalisch markiert, nie postvokalisch, was den üblichen Transkriptionsmethoden entspricht;
- Zunächst wurde die Syllabizität mancher Intervalle nicht berücksichtigt, im Laufe der Untersuchung gewann dieser Faktor jedoch an Bedeutung und wurde neu bedacht (s. unten).

Sprechpausen, mit <P> markiert, wurden von nun an nicht weiter berücksichtigt.

Relevante Sonoritätsklassen waren in diesem Zusammenhang nicht nur Vokale, die schließlich bereits etikettiert worden waren, sondern Approximanten, Laterale, Nasale, Frikative und Plosive. Die Reihenfolge der Nennung dieser Lautklassen richtet sich nach der Sonoritätshierarchie, deren grundlegendes Konzept einen Kerngedanken der vorliegenden Arbeit bildet. Das sich zunächst aus diesem zweiten Durchgang der Annotation ergebende Labelinventar ist in Tabelle 3 dargestellt.

Vokale	<i>V</i>
Approximanten	<i>A</i>
Laterale	<i>L</i>
Nasale	<i>N</i>
Frikative	<i>F</i>
Plosive	<i>S</i>

Tabelle 3: Vorläufig verwendetes Inventar der Sonoritätsklassen.

Während der zweiten Untersuchung fiel auf, daß Nasale und Laterale abhängig von phonetischem Kontext, verwendeter Sprache und Sprechgeschwindigkeit verschieden stark dazu tendierten, Silbenkerne zu bilden (s. Abschnitt 5.5). Um zu verhindern, daß die ungleichmäßige Verteilung von silbischen Nasalen und Lateralen in den Einzelsprachen einen unkontrollierten Einfluß auf die rhythmischen Parameter nehmen könnte, wurde beschlossen, das Sonoritätsklassen-Inventar zu revidieren und zu erweitern, um zwischen silbischen und nichtsilbischen Konsonanten unterscheiden zu können. Dabei wurde das Prinzip eingeführt, nichtsilbische und silbische Sonoritätsklassen durch Groß- und Kleinschreibung zu unterscheiden. Demnach sind in der vorliegenden Arbeit bei der Notation von Konfigurationen und Parametern die Sonoritätsklassen durch Großbuchstaben markiert, wenn *nicht* nach Syllabizität unterschieden wird. Werden silbische Sonoritätsklassen *getrennt* von nichtsilbischen betrachtet, so werden sie durch Großbuchstaben repräsentiert, während alle nichtsilbischen Klassen durch Kleinbuchstaben repräsentiert werden. Vokale, die im Regelfall

den Silbenkern bilden, werden abhängig von der Betrachtungsweise entweder groß- oder kleingeschrieben, sind jedoch immer silbisch. Somit ergibt sich bei Berücksichtigung von Syllabizität das Labelinventar, das in Tabelle 4 dargestellt ist.

Vokale	<i>v</i>
Approximanten	<i>a</i>
nichtsilbische Laterale	<i>l</i>
silbische Laterale	<i>L</i>
nichtsilbische Nasale	<i>n</i>
silbische Nasale	<i>N</i>
Frikative	<i>f</i>
Plosive	<i>s</i>

Tabelle 4: Endgültig verwendetes Inventar der Sonoritätsklassen.<sup>14</sup>

Dadurch, daß Syllabizität fast ausschließlich bei Nasalen und Lateralen auftrat, wurden Intervalle, die als silbische Approximanten (<A>), silbische Frikative (<F>) oder silbische Plosive (<S>) markiert waren, sofern sie überhaupt auftraten, nicht separat ausgewertet, sondern interpretiert, als gehörten sie zu der Menge der Realisationen ihres nichtsilbischen Gegenstücks. Die mit diesem Inventar von Labeln erstellte erweiterte Annotation wurde in Form von IntervalTier-Dateien abgespeichert, die für die automatische Weiterverarbeitung (nicht nur in *Praat*) geeignet sind. So ergab sich eine Menge von 80 Annotationen, aus denen mit Hilfe eines *Perl*-Skripts verschiedene Parameter extrahiert werden konnten. Die Anzahl der Parameter beschränkte sich also auf 8 (<*v, a, l, L, n, N, f, s*>).

Entsprechend der in Kapitel 3 beschriebenen Überlegungen ergaben sich für jede der 80 Aufnahmen 262 unabhängige Konfigurationen dieser Parameter, deren Werte in

---

<sup>14</sup> Die Tatsache, daß die silbischen Sonoritätsklassen in der vorliegenden Arbeit jeweils *nach* ihrem nichtsilbischen Gegenstück genannt werden, geht auf eine willkürliche Entscheidung bei der Datenverarbeitung zurück und hat nichts mehr mit der Sonoritätshierarchie zu tun. Vielen Dank an Teodor Petrič von der Universität Maribor (Slowenien), der die Aufmerksamkeit auf diese Frage lenkte.

einer Datenbank abgespeichert wurden. Diese Datenbank wurde dann mit weiteren *Perl*-Skripten zum einen für die Verarbeitung in *Praat* in eine *TableOfReal*-Datei konvertiert und zum anderen in ein Format gebracht, das von *SPSS* eingelesen werden konnte.

Mit einem *Praat*-Skript wurde eine Datenbankabfrage implementiert, bei der die Konfiguration der Parameter frei wählbar war. Die Ausgabe dieser Abfrage war graphisch, wobei entweder die einzelnen Objekte, Durchschnittswerte für jeden Sprecher oder Durchschnittswerte für jede Sprache dargestellt werden konnten; letztere Variante entspricht der gängigen Darstellungsform bei Ramus et al. (1999) und daraus abgeleiteten Arbeiten.

Die beiden ersten Untersuchungen begründeten sich auf die visuelle Auswertung der in *Praat* erstellten Verteilungsgraphen. Bei der endgültigen statistischen Analyse wurde jedoch *SPSS* verwendet.

## 5 Ergebnisse und Interpretation

Eine Reihe von Voruntersuchungen soll klären, ob die Datengrundlage, wie sie für die vorliegende Arbeit aus dem BonnTempo-Korpus ausgewählt und aufbereitet wurde, für eine Untersuchung mit den Methoden aus Ramus et al. (1999) geeignet ist. Dazu soll überprüft werden, ob sich bei gleicher Herangehensweise vergleichbare Ergebnisse erbringen lassen. Darüber hinaus sollte die Sprechgeschwindigkeit auf die Ergebnisse den gleichen Einfluß nehmen wie bei Dellwo & Wagner (2003) beschrieben. Gleichzeitig kann die Position der Werte des Deutschen in der %X/ΔY-Ebene anhand dieser beiden Untersuchungen beurteilt werden. Diese Voruntersuchungen entsprechen den ersten drei Hypothesen, die am Ende von Kapitel 3 aufgestellt wurden.

Anschließend wird die Datengrundlage der vorliegenden Arbeit dazu verwendet, den Einfluß der einzelnen Sonoritätsklassen auf die Sprachverteilung zu analysieren. Dabei werden die übrigen am Ende von Kapitel 3 formulierten Hypothesen der Reihe nach überprüft. Dieser Teil bildet den Kern der vorliegenden Arbeit.

### 5.1 Vergleich mit Ramus et al. (1999)

In einer ersten Untersuchung wurde nur die erste der in Tabelle 3 aufgeführten sechs Konfigurationen verwendet. Diese stimmt mit der C/V-Segmentierung überein, die Ramus et al. (1999) in ihrer Untersuchung verwendeten. Die dortigen Ergebnisse sind noch einmal in Abb. 20 dargestellt und ermöglichen einen Vergleich mit den Sprachverteilungen dort und in dieser Arbeit. Die Verteilung der vier in der vorliegenden Arbeit untersuchten Einzelsprachen in der %X/ΔY-Ebene bei gleicher Konfiguration ist in Abb. 21 dargestellt. Dabei wurde die Standardabweichung der Einzelwerte aller Aufnahmen einer Sprache als Fehlerbalken eingezeichnet. Die farbliche Kodierung der vier Einzelsprachen wird im restlichen Teil dieses Kapitels z.T. ohne Beschriftung weiterverwendet, wobei Deutsch olivfarben dargestellt wird, Englisch blau, Französisch rot und Italienisch grün.

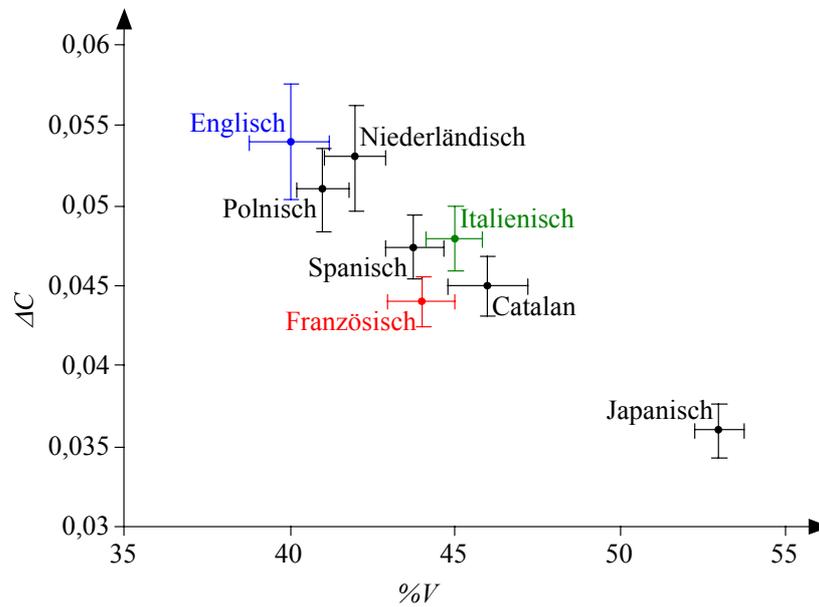


Abb. 20: Sprachverteilung in der %V/ΔC-Ebene (aus Ramus et al. 1999). Die Fehlerbalken repräsentieren  $\pm 1$  Standardfehler.

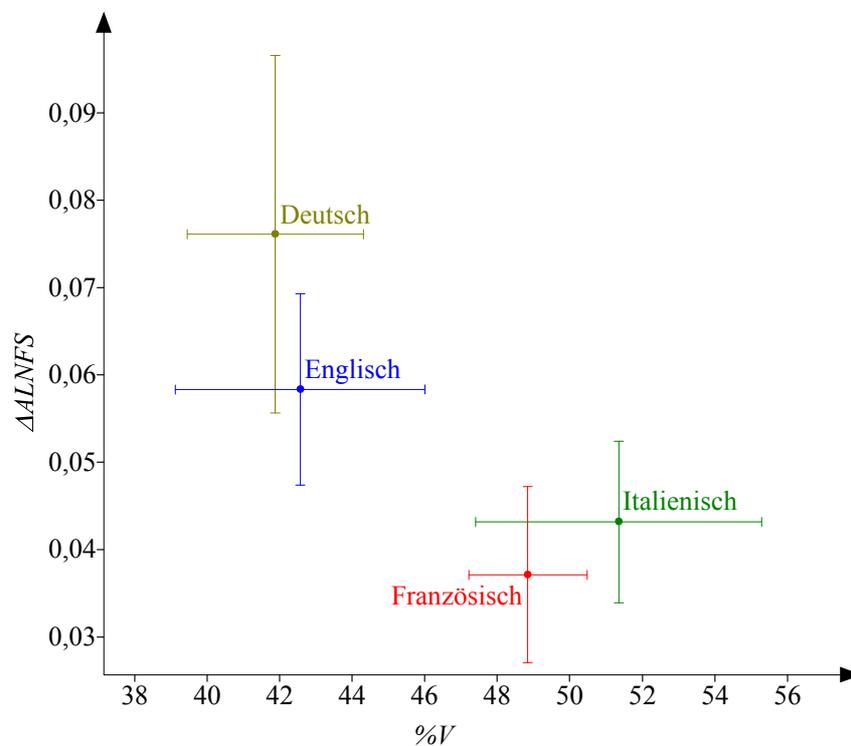


Abb. 21: Sprachverteilung in der %V/ΔALNFS-Ebene bei Konfiguration V/ALNFS. Die Fehlerbalken repräsentieren die Standardabweichung.

Die drei Sprachen, die sowohl in der vorliegenden Arbeit als auch bei Ramus et al. (1999) untersucht wurden, sind in Abb. 20 zum besseren Vergleich farblich hervorge-

hoben. Es können leicht die drei Gruppen der akzentzählenden, silbenzählenden und morazählenden Sprachen erkannt werden. In Abb. 21 sind entsprechend zwei Gruppen zu erkennen, die akzentzählenden und silbenzählenden Sprachen. Es sollte beachtet werden, daß sich das Datenmaterial, das der Untersuchung von Ramus et al. (1999) zugrunde liegt, in mehreren Punkten von den Daten unterscheidet, die in der vorliegenden Arbeit Verwendung finden. Diese Unterschiede sind in Tabelle 5 aufgeführt.

Ramus et al. (1999):	vorliegende Arbeit:
8 Sprachen	4 Sprachen
4 Sprecher pro Sprache	2-5 Sprecher pro Sprache
5 mal 15-19 Silben pro Sprecher	5 mal 74-123 Silben pro Sprecher
nur weibliche Sprecher	weibliche und männliche Sprecher
keine Kontrolle der Sprechgeschwindigkeit	kontrolliert variierende Sprechgeschwindigkeit
16kHz Abtastrate	44,1kHz Abtastrate

Tabelle 5: Zusammenfassende Gegenüberstellung der wesentlichen Unterschiede zwischen den Daten in Ramus et al. (1999) und in der vorliegenden Arbeit.

Da die Schnittmenge der untersuchten Sprachen sich auf Englisch, Französisch und Italienisch beschränkt, sind dies die Sprachen, deren Verteilung in der  $%X/\Delta Y$ -Ebene bei der Konfiguration  $V/ALNFS$  betrachtet werden soll. Wie sich zeigt, weichen zwar die absoluten Werte aufgrund der unterschiedlichen Daten geringfügig voneinander ab; entscheidender ist aber die Anordnung selbst, die die Einzelsprachen in der  $%X/\Delta Y$ -Ebene einnehmen, und diese ist trotz der Datenunterschiede durchaus vergleichbar.

## 5.2 Einfluß von Sprechgeschwindigkeit

Eine frühe Fassung des BonnTempo-Korpus wurde bereits hinsichtlich des Einflusses, den Sprechgeschwindigkeit auf die Verteilung von Sprachaufnahmen in der  $%V/\Delta C$ -Ebene ausübt, ausgewertet (Dellwo & Wagner 2003). Dieser Einfluß scheint sich wie dort beschrieben bei der Konfiguration  $%V/\Delta ALNFS$  bei akzentzählenden Sprachen stärker auf die  $\Delta Y$ -Werte auszuwirken als auf die  $%X$ -Werte, während sich bei silben-

zählenden Sprachen – in geringerem Umfang – die umgekehrte Tendenz beobachten läßt (Dellwo & Steiner, in Vorbereitung). Bei den an dieser Stelle diskutierten Ergebnissen ist dieser Einfluß vor allem an den hohen Standardabweichungen der  $\Delta Y$ -Werte zu erkennen. Bei langsam gesprochener Sprache sind die  $\Delta Y$ -Werte grundsätzlich höher als bei schnell gesprochener Sprache, was sich mit einer höheren Variabilität der Konsonantencluster-Dauern paraphrasieren läßt. Dies ist dadurch zu erklären, daß sich nicht alle Sonoritätsklassen bei veränderter Sprechgeschwindigkeit gleichförmig in ihrem zeitlichen Verlauf und damit in ihrer Dauer verändern.

Sobald sich allerdings die Konfiguration ändert, wird auch die soeben beschriebene Tendenz beeinflußt (s. Abschnitt 5.4; Abb. 23-Abb. 25). Es soll dieser Stelle jedoch nur die Ausgangskonfiguration *V/ALNFS* betrachtet werden. Die  $\%X/\Delta Y$ -Werte der einzelnen Aufnahmen, die in der vorliegenden Arbeit verwendet wurden, sind bei dieser Konfiguration in ihrer Verteilung in Abb. 22 dargestellt.

Die farbliche Kodierung unterscheidet, wie in Abschnitt 5.1 erläutert, die einzelnen Sprachen. Die jeweils fünf Aufnahmen pro Sprecher sind als Punkte eingetragen und mit gestrichelten Linien untereinander „verkettet“. Der Buchstabe *S* am einen Ende jeder „Kette“ kennzeichnet die Aufnahme mit der jeweils höchsten Sprechgeschwindigkeit, das *L* die langsamste Aufnahme. Die übrigen Buchstaben sind die im Bonn-Tempo-Korpus verwendeten Kürzel für die Namen der Sprecher; in Abb. 22 ist damit jeweils die mit „normaler“ Sprechgeschwindigkeit (s. Abschnitt 4.3) durchgeführte erste Aufnahme jeder „Kette“ markiert.

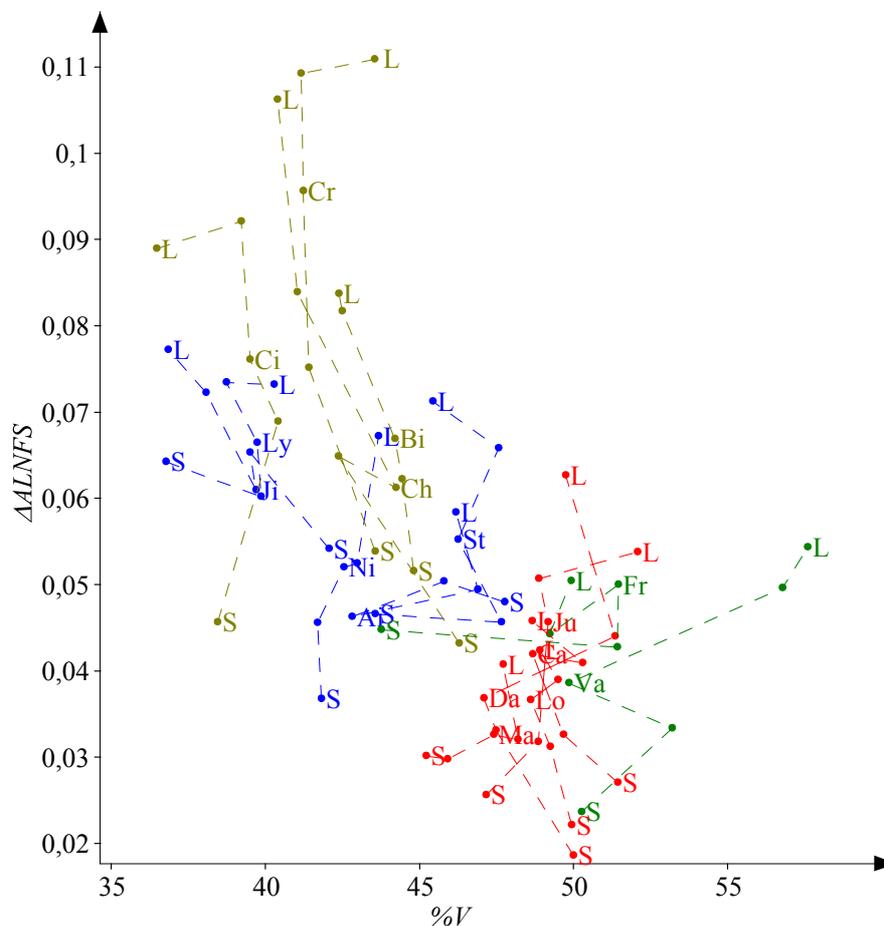


Abb. 22: Verteilung der einzelnen Aufnahmen in der  $%X/\Delta Y$ -Ebene bei Konfiguration  $V/\Delta L N F S$ .

Es zeigt sich, daß die Sprechgeschwindigkeit die bei Dellwo & Wagner (2003) beschriebenen Einflüsse auf die Sprachverteilung in der  $%V/\Delta A L N F S$ -Ebene ausübt. Wie sie sich jedoch bei anderen Konfigurationen als  $V/\Delta L N F S$  auf die Verteilung einzelner Sprachen auswirkt, soll an dieser Stelle nicht näher untersucht werden. Allerdings kann dieser Einfluß als ein Faktor zur Erklärung der hohen Streuung der einzelnen Aufnahmen herangezogen werden, die sich bei vielen Konfigurationen im übrigen Teil dieses Kapitels beobachten läßt.

### 5.3 Einordnung des Deutschen

Deutsch ist in Abb. 21 eindeutig in der Gruppe der akzentzählenden Sprachen angesiedelt, von der Gruppe der silbenzählenden Sprachen aus gesehen sogar „jenseits“ des Englischen. Insofern scheinen die Merkmale, die die rhythmischen Klassen

nach den Methoden von Ramus et al. (1999) unterscheiden, beim Deutschen noch stärker ausgeprägt zu sein als beim Englischen. Dellwo & Wagner (2003) unterstützen mit den Ergebnissen ihrer Messungen an den gleichen Daten diese Einschätzung.

Wie zuverlässig diese Beobachtungen allerdings sind, bleibt vor der Tatsache, daß die Datengrundlagen bei Ramus et al. (1999) und in der vorliegenden Arbeit sich sowohl in der Auswahl der Sprachen als auch in ihrem innersprachlichen Bestand unterscheiden, ungewiß.

Die  $\%V/\Delta C$ -Werte, die Grabe & Low (2002) aus ihren Daten für das Deutsche gemessen haben, wären in Abb. 21 ungefähr in der Mitte zwischen den beiden Sprachgruppen einzuzeichnen. Angesichts der Bedenken, die in Abschnitt 2.4 gegen die Datengrundlage dieser Messungen erläutert worden sind, wird diese offensichtliche Diskrepanz zwischen den beiden verschiedenen Positionen des Deutschen in der  $\%V/\Delta ALNFS$ -Ebene, wie in Fußnote 9 angekündigt, an dieser Stelle nicht weiter diskutiert; sie wurde lediglich der Vollständigkeit halber erwähnt.

#### **5.4 Veränderung der Konfiguration entlang der Sonoritätshierarchie**

Nachdem durch die Voruntersuchungen ein gewisser Rahmen geschaffen worden ist, soll nun veranschaulicht werden, wie sich die Veränderung der Konfiguration auf die Sprachverteilung auswirkt. In Tabelle 6 sind die vier Konfigurationen dieser Untersuchung aufgeführt, die sich wie in Kapitel 3 beschrieben an der Sonoritätshierarchie orientieren. Dabei wurde von einer Konfiguration zur nächsten nur die Grenze zwischen den Sonoritätsklassen, die zu  $X$  gezählt wurden, und denen, die zu  $Y$  gezählt wurden, auf der Sonoritätshierarchie schrittweise verschoben.

$X$	$Y$
$V$	$ALNFS$
$VA$	$LNFS$
$VAL$	$NFS$
$VALN$	$FS$

Tabelle 6: Die vier Konfigurationen der zweiten Untersuchung; die Reihenfolge entspricht der Sonoritätshierarchie.

Die Verteilung der Einzelsprachen in der  $\%X/\Delta Y$ -Ebene wurde für jede Konfiguration graphisch analysiert und die Veränderungen von einer Konfiguration zur nächsten beobachtet. Die Verteilung bei der ersten Konfiguration ist bereits in Abb. 21 dargestellt worden; die übrigen drei Konfigurationen sind durch Abb. 23 bis Abb. 25 veranschaulicht.

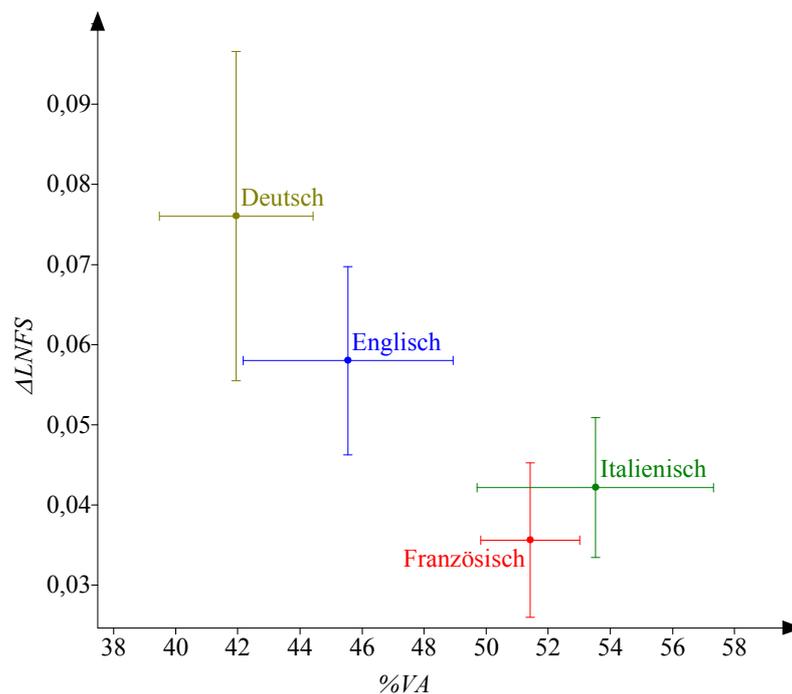


Abb. 23: Sprachverteilung in der  $\%X/\Delta Y$ -Ebene bei Konfiguration  $VA/LNFS$ . Die Fehlerbalken repräsentieren die Standardabweichung.

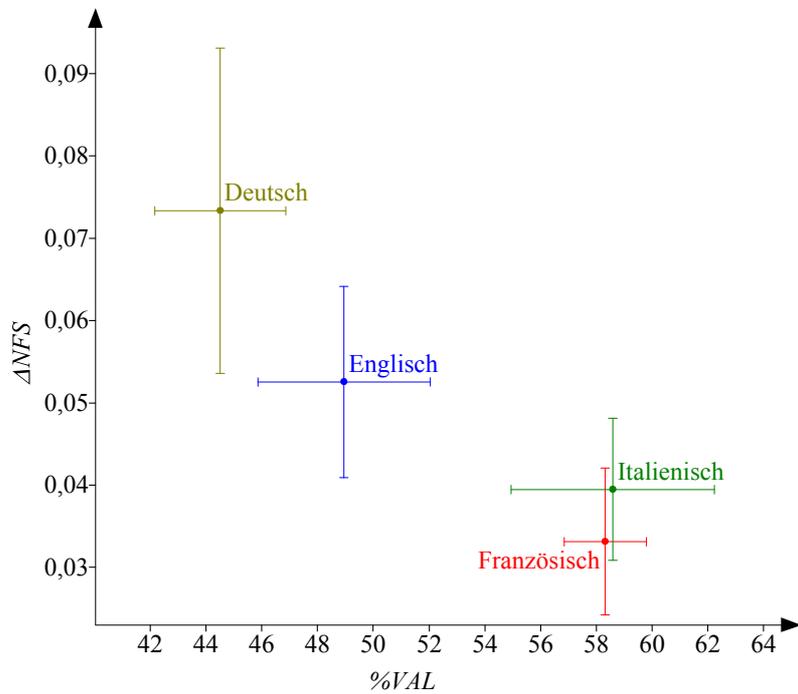


Abb. 24: Sprachverteilung in der  $\%X/\Delta Y$ -Ebene bei Konfiguration  $VAL/NFS$ . Die Fehlerbalken repräsentieren die Standardabweichung.

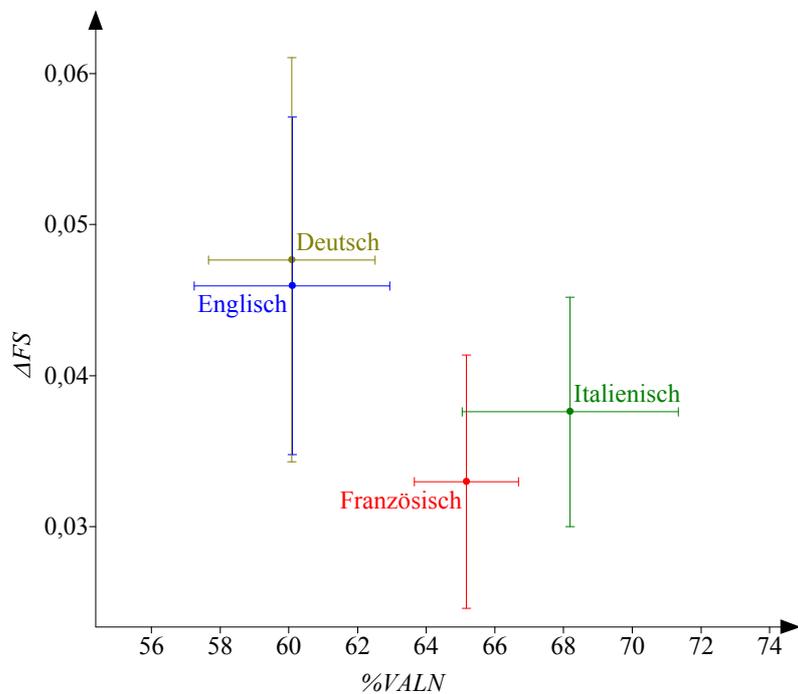


Abb. 25: Sprachverteilung in der  $\%X/\Delta Y$ -Ebene bei Konfiguration  $VALN/FS$ . Die Fehlerbalken repräsentieren die Standardabweichung.

Die Veränderungen von einer Konfiguration zur nächsten teilen sich in drei Aspekte: Veränderung entlang der %X-Dimension, Veränderungen entlang der  $\Delta Y$ -Dimension und Veränderungen in der Verteilung der Sprachen untereinander.

Die auffälligste Veränderung der Sprachverteilung von einer Konfiguration zur nächsten ist entlang der %X-Achse festzustellen. Dies sollte auch nicht anders sein, denn schließlich gilt

$$\%V \leq \%VA \leq \%VAL \leq \%VALN \leq \%VALNF \leq \%VALNFS = 100$$

Allerdings ist nicht die Tatsache der Verschiebung nach rechts auf der %X-Achse interessant, sondern das Ausmaß dieser Verschiebung von einer Konfiguration zur nächsten. Bei näherer Untersuchung zeigt sich nämlich, daß die weitaus stärkste Veränderung durch die Zunahme der Nasale in  $X$  bewirkt wird.

Die Verschiebung entlang der  $\Delta Y$ -Achse verläuft grundsätzlich nach unten. Diese Tendenz ist dadurch zu erklären, daß bei einer Reduktion der verschiedenen Sonoritätsklassen, die zu  $Y$  zusammengefaßt werden, auch eine Verringerung der Gesamtvariabilität der Dauern dieser Sonoritätsklassen-Cluster entsteht. Diese Bewegung erfährt ebenfalls einen deutlichen Sprung, sobald die Nasale nicht mehr zu  $Y$ , sondern zu  $X$  gezählt werden.

Was die Form der Verteilung der Sprachen untereinander betrifft, so ist deutlich zu beobachten, wie sie sich aufeinander zu bewegen. Diese Tendenz ist ebenfalls erwartungsgemäß, denn in  $X$  sind immer größere Anteile an der Gesamtdauer der jeweiligen Aufnahme enthalten. Würden auch noch Frikative und Plosive zu  $X$  hinzugezählt, so würde der %X-Wert für alle Sprachen 100 erreichen, was trivial ist. Interessant ist in diesem Zusammenhang vielmehr, daß sich auf der %X-Achse die akzentzählenden Sprachen Deutsch und Englisch – sie fallen geradezu zusammen – selbst bei der letzten Konfiguration noch stark von den silbenzählenden Sprachen Französisch und Italienisch abgrenzen lassen. Dies belegt, daß der prozentuale Anteil, den Frikative und Plosive an der Gesamtdauer einer Äußerung besitzen, in akzentzählenden Sprachen höher zu sein scheint als in silbenzählenden. Daß die Sprachen sich auch in der  $\Delta Y$ -Dimension einander annähern, war ebenfalls aus der Tatsache vorherzusagen, daß in  $Y$  beim Wechsel von einer Konfiguration zur nächsten immer weniger Sonoritätsklassen enthalten sind, und sich somit auch die

sen enthalten sind, und sich somit auch die Standardabweichung der Dauern dieser Sonoritätsklassen-Cluster verringert. Bei der letzten Konfiguration erlebt die Verteilung der Sprachen in ihrer Form bei Hinzunahme von Nasalen in  $X$  einen erheblichen Sprung. Dies unterstützt die Hypothese, daß Unterschiede zwischen akzentzählenden und silbenzählenden Sprachen bei verschiedenen Konfigurationen entlang der Sonoritätshierarchie erhalten bleiben.

Eine vielversprechende Beobachtung aus dieser zweiten Untersuchung ist, daß die An- oder Abwesenheit der Nasale in einem Parameter einen hohen Einfluß auf dessen Werte ausübt. Am stärksten wirkt er sich im Deutschen aus, er ist jedoch auch bei den übrigen untersuchten Sprachen nicht unerheblich. Nasale scheinen also für die Verteilung der Einzelsprachen in der  $%X/\Delta Y$ -Ebene stärker verantwortlich zu sein als andere Sonoritätsklassen. Dieser Beobachtung wird sich der nächste Abschnitt widmen.

### **5.5 Einfluß der Syllabizität von Lateralen und Nasalen**

An dieser Stelle mag der Verdacht aufkommen, der in Abschnitt 4.6 vorweggenommen wurde, daß Nasale deshalb eine so prominente Rolle bei der Zusammensetzung von  $X$  und  $Y$  und somit bei der Verteilung der Einzelsprachen in der  $%X/\Delta Y$ -Ebene spielen, weil sich in vielen Sprachen grundsätzlich – erst recht bei hoher Sprechgeschwindigkeit und damit verbundener Reduktion – silbische Nasale finden, die als Segment schließlich eine relativ lange Dauer aufweisen. Diese Vermutung könnte angesichts der in Kapitel 3 angeführten Beispiele auch (silbische) Laterale miteinbeziehen.<sup>15</sup>

Um zu unterscheiden, wie stark der Einfluß von silbischen Lateralen und Nasalen bei den bisher aufgeführten Ergebnissen ist, wurde, wie in Abschnitt 4.6 angekündigt und durch Tabelle 4 veranschaulicht, das Inventar an relevanten Sonoritätsklassen um diese beiden Intervalltypen erweitert, mit der sich daraus ergebenden größeren Anzahl von möglichen Konfigurationen.

---

<sup>15</sup> Je nach Sprachdaten wäre es womöglich auch von Interesse, andere Konsonantenklassen hinsichtlich ihrer Syllabizität zu unterscheiden; bei einer Untersuchung des Tschechischen beispielsweise könnten auch silbische Frikative eine Rolle spielen.

Da sich diese Untersuchungen nur auf die silbischen Laterale und Nasale beziehen, konzentrieren sich die nachfolgenden Erläuterungen gezielt auf diese beiden Sonoritätsklassen. Von Interesse sind also Gegenüberstellungen der Konfigurationen, die in Tabelle 7 aufgezählt sind.<sup>16</sup>

Nasale		Laterale	
<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>
<i>nN</i>	<i>valLfs</i>	<i>lL</i>	<i>vanNfs</i>
<i>n</i>	<i>valLNfs</i>	<i>l</i>	<i>vaLnNfs</i>
<i>N</i>	<i>valNfs</i>	<i>L</i>	<i>valnNfs</i>

Tabelle 7: Konfigurationen, um den Einfluß von Syllabizität auf %*X* bei Lateralen und Nasalen zu untersuchen.

Diese Konfigurationen isolieren jeweils den Parameter, der untersucht werden soll. Dabei ist allerdings der jeweils andere Parameter von geringer Bedeutung. Deshalb und aus Platzgründen soll hier darauf verzichtet werden, für jede der in Tabelle 7 aufgezählten Konfigurationen die entsprechende Verteilung der Einzelsprachen in der %*X*/*Y*-Ebene graphisch abzubilden. Statt dessen werden in Abb. 26 die Anteile der silbischen und nichtsilbischen Laterale und Nasale an der Gesamtdauer aller Laterale bzw. Nasale dargestellt. Die zugrundeliegenden Werte finden sich im Anhang.

---

<sup>16</sup> Im verbleibenden Teil dieses Kapitels spielt also, wie in Abschnitt 4.6 erläutert, die Groß- und Kleinschreibung der einzelnen Sonoritätsklassen eine Rolle.

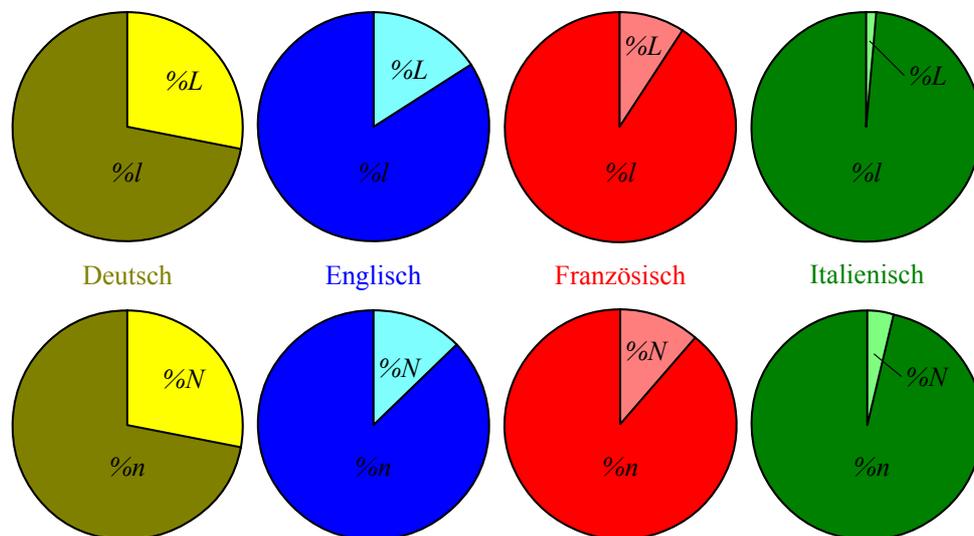


Abb. 26: Anteil der silbischen und nichtsilbischen Komponenten an der durchschnittlichen Gesamtdauer von Lateralen und Nasalen.

Wie sich unschwer erkennen läßt, machen die silbischen Laterale und Nasale jeweils nur einen geringen Anteil an der Gesamtdauer aller Laterale bzw. Nasale aus. Den größten Einfluß auf die Ergebnisse bei den Lateralen besitzt die Syllabizität mit 28,1% im Deutschen, den geringsten im Italienischen, dort sind es lediglich 1,5%. Ähnlich verhält es sich mit den Nasalen, wo die Syllabizität im Deutschen ebenfalls einen Anteil von 28,1% besitzt, im Italienischen nur 3,7%. Interessanterweise scheint auch das Merkmal des Anteils silbischer Liquide an der Gesamtdauer aller Liquide den Eindruck zu unterstützen, daß das Deutsche an einem Ende eines Kontinuums anzusiedeln ist, das sich über Englisch und Französisch bis zum Italienischen erstreckt. Zwar sind die Verhältnisse bei Lateralen und Nasalen relativ ähnlich, doch die absoluten Anteile aller Laterale und Nasale an der Gesamtdauer sind in Tabelle 8 dargestellt. Diese Werte zeigen, daß der prozentuale Anteil aller Nasale in allen untersuchten Sprachen (mit Ausnahme des Französischen) erheblich höher ist als der prozentuale Anteil aller Laterale, und sie erklären auch, warum der „Sprung“ von einer Konfiguration in Tabelle 6 zur nächsten erst bei der Hinzunahme von Nasalen in  $X$  so deutlich zu beobachten ist (wobei die akzentzählenden Sprachen stärker beeinflußt werden), und nicht schon bei der Hinzunahme von Lateralen.

	Deutsch	Englisch	Französisch	Italienisch
$%lL$	2,56	3,40	6,90	5,06
$%nN$	15,58	11,14	6,85	9,61

Tabelle 8: Durchschnittswerte für den Anteil aller Laterale und Nasale an der Gesamtdauer.

Es kann aus diesem Abschnitt der Schluß gezogen werden, daß der Anteil der silbischen Segmente an der Gesamtdauer aller Segmente einer Sonoritätsklasse bei akzentzählenden Sprachen einen höheren Einfluß auf Laterale und Nasale ausübt als bei silbenzählenden Sprachen. Gleichzeitig zeigt sich anhand der Beobachtungen aus Abschnitt 5.4 und durch Tabelle 8, daß  $%lL$  und  $%nN$  als Parameter ein hohes Trennungspotential für die Diskriminierung von akzentzählenden und silbenzählenden Sprachen aufweisen.

## 5.6 Statistische Analyse einzelner Parameter

Inzwischen sind also einzelne *Parameter* in den Vordergrund gerückt, womit die Untersuchung bestimmter *Konfigurationen* ihren Reiz verliert. Um diese Entwicklung konsequent weiterzuverfolgen, wurde nun eine Diskriminanzanalyse in *SPSS* durchgeführt.

Für jede der 80 Aufnahmen, die der vorliegenden Arbeit als Datengrundlage dienen, wurden die bei 8 Sonoritätsklassen 262 Parameter (s. Kapitel 3) in *SPSS* eingelesen und mittels einer Diskriminanzanalyse klassifiziert. Als Gruppierungsvariable diente im ersten Durchgang die gesprochene Sprache jeder Aufnahme, die somit vier Gruppen erzeugte, eine für jede Sprache. Es folgten fünf weitere Durchgänge, bei denen die vier untersuchten Sprachen jeweils in zwei Gruppen geteilt wurden: Deutsch vs. andere, Englisch vs. andere, Französisch vs. andere, Italienisch vs. andere und akzentzählende vs. silbenzählende Sprachen. Die Ergebnisse dieser Analysen sind aus Platzgründen im Anhang aufgeführt. Für jeden Durchgang finden sich dort eine Eigenwert-Tabelle und eine Strukturmatrix. Die Strukturmatrizen listen die Wichtigkeit der einzelnen Parameter innerhalb des Durchgangs auf. Die zwei jeweils prominentesten Parameter in jedem Durchgang wurden in einem Verteilungsgraphen gegeneinander auf-

aufgetragen, um die maximale Trennung zwischen den jeweiligen Gruppen graphisch zu veranschaulichen. Diese Graphen sind in Abb. 27 bis Abb. 32 dargestellt. In diesen Abbildungen repräsentieren die Fehlerbalken stets die Standardabweichung.

Im letzten Durchgang wurde zusätzlich auf  $\%v$  und  $\Delta allNfs$  geachtet, um den Vergleich zu ermöglichen, wie gut die Variablen, die Ramus et al. (1999) verwenden, auf diese Daten passen. Wie sich herausstellt, ist die Korrelation für  $\%v$  und  $\Delta allNfs$  (bei Ramus et al. 1999  $\Delta C$ ) sogar geringer als die von  $\%l$  und  $\Delta valfs$ , die akzentzählende und silbenzählende Sprachen maximal trennen.

Es zeigt sich anhand der Eigenwert-Tabellen, daß die erste Diskriminanzfunktion bei jedem Durchgang eine vollständige und quasi fehlerfreie Klassifikation (die Korrelation liegt zwischen 0,983 und 0,996) ermöglicht. Die Strukturmatrizen geben dabei über die Korrelation der Variablen mit der jeweiligen Diskriminanzfunktion Auskunft. Da diese Korrelationen nie sonderlich hohe Werte annehmen, scheinen die hier gewonnenen Ergebnisse keine allzu verbindlichen Aussagen belegen zu können. Angesichts der hohen Anzahl von Variablen, die bei jedem Durchgang der Diskriminanzanalyse miteinbezogen wurden, ist die Rangordnung stark von der Datengrundlage abhängig.

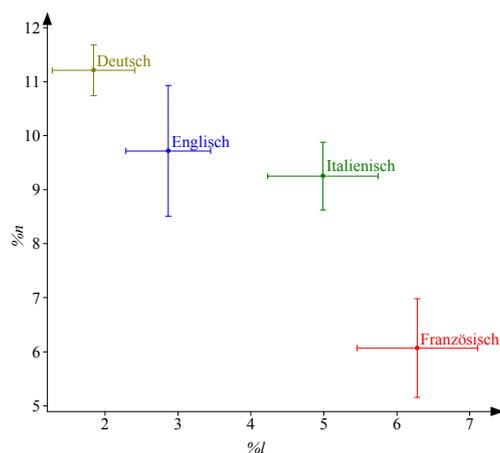


Abb. 27: Sprachverteilung in der  $\%l/\%n$ -Ebene zur Trennung der einzelnen Sprachen voneinander.

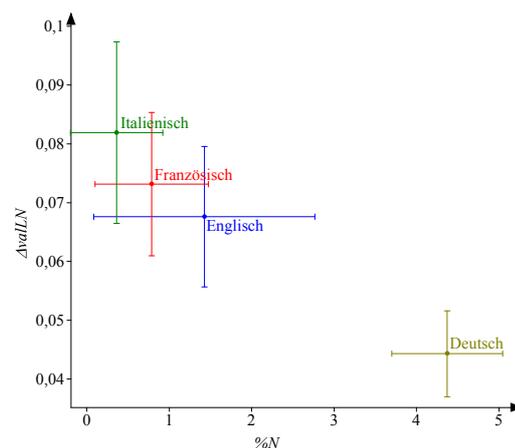


Abb. 28: Sprachverteilung in der  $\%N/\Delta valLN$ -Ebene zur Trennung des Deutschen von den übrigen Sprachen.

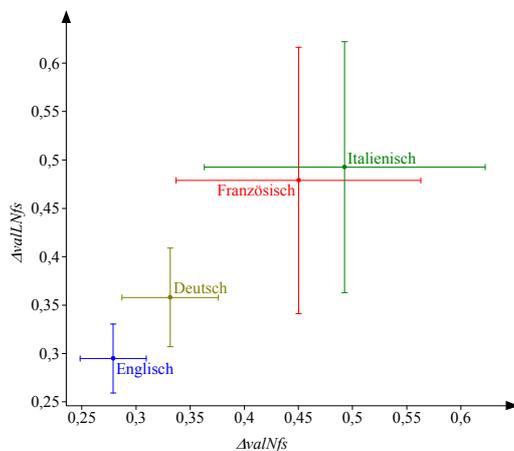


Abb. 29: Sprachverteilung in der  $\Delta valNfs/\Delta valLnfs$ -Ebene zur Trennung des Englischen von den übrigen Sprachen.

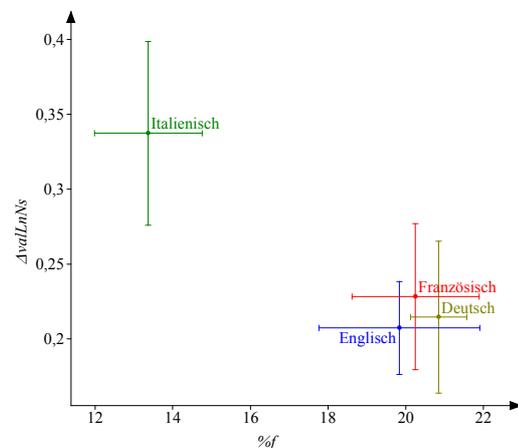


Abb. 31: Sprachverteilung in der  $\%f/\Delta valLnfs$ -Ebene zur Trennung des Italienischen von den übrigen Sprachen.

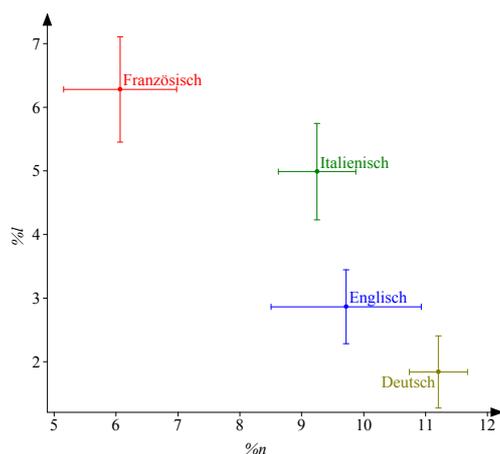


Abb. 30: Sprachverteilung in der  $\%n/\%l$ -Ebene zur Trennung des Französischen von den übrigen Sprachen.

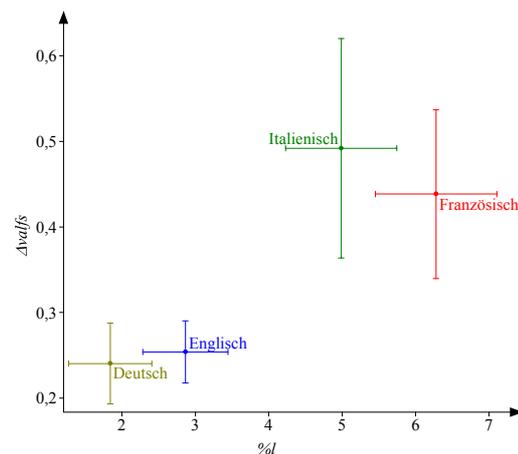


Abb. 32: Sprachverteilung in der  $\%l/\Delta valfs$ -Ebene zur Trennung von akzentzählenden und silbenzählenden Sprachen.

Über den Einfluß der Laterale und Nasale scheinen jedoch auch diese Ergebnisse wenig Zweifel zu lassen.  $\%n$  und  $\%l$  scheinen bei mehreren Durchgängen als Variable eine prominente Rolle zu spielen. Nicht von der Hand zu weisen sind diese beiden Parameter beim ersten und beim letzten Durchgang, wo sie allein eine hohe Gruppentrennung ermöglichen.

Angesichts der Verlagerung dieser Untersuchung von der graphischen Betrachtung bestimmter Konfigurationen auf die statistische Analyse einzelner Parameter erscheint es sinnvoll, die letzten beiden Hypothesen aus Kapitel 3 neu zu formulieren:

- 6) Aufgrund der einzelsprachlichen Phonotaktik sowie der unterschiedlichen Vorkommensverteilung verschiedener Sonoritätsklassen in der jeweiligen Sprache sollten die Werte der Parameter dieser Sprachen jeweils unterschiedlich sein.
- 7) Die Menge der Parameter sollte für jede Einzelsprache, verbunden mit ihrer Phonotaktik, ein für diese Sprache charakteristisches Muster ergeben, das bei gleichen Bedingungen keiner anderen Sprache zugeordnet werden kann.

Durchgänge zwei bis fünf unterstützen diese beiden Hypothesen. Die Werte und unterschiedliche Rangordnung der Parameter in den Strukturmatrizen belegen sowohl, daß sich in den Einzelsprachen verschiedene Werte derselben Parameter messen lassen, als auch, daß jede Sprache eine eigene Rangordnung dieser Parameter aufweist. Es wäre denkbar, in diesem Zusammenhang vorzuschlagen, daß jede Sprache solche Parameter in einer bestimmten Hierarchie anordnet, die diese Sprache rhythmisch charakterisiert.<sup>17</sup>

Der erste und der letzte Durchgang zeigen, daß es durchaus nicht notwendigerweise die von Ramus et al. (1999) vorgeschlagene und seither von anderen mehrfach verwendete  $%V/\Delta C$ -Verteilung sein muß, die einzelne Sprachen oder rhythmische Sprachgruppen maximal voneinander trennt, sondern daß es auf die Zusammensetzung der „C-Komponente“ (hier:  $Y$ ) ankommt. Genauer gesagt,  $%l$  und  $%n$  stellen sowohl bei der Verteilung der einzelnen Sprachen, als auch bei der Gruppierung von Sprachen aus verschiedenen rhythmischen Gruppen (akzentzählende bzw. silbenzählende Sprachen) einen zuverlässigeren Faktor dar, als  $%V$  oder  $\Delta C$  (hier:  $\Delta alLnNfs$ ).

---

<sup>17</sup> Dieses Prinzip erinnert an die sprachspezifischen *constraint rankings* der Optimalitätstheorie (Prince & Smolensky 1993), auch wenn dieser Vergleich sehr unorthodox ist, denn darüber hinaus hat die vorliegende Arbeit mit der Optimalitätstheorie wenig gemeinsam.

## 5.7 Zusammenfassung der Ergebnisse

In diesem Kapitel wurde anhand der Datengrundlage der vorliegenden Arbeit gezeigt, daß es bei einer Untersuchung, die versucht, nach ähnlichen Methoden wie Ramus et al. (1999) akustische Korrelate für linguistischen Rhythmus zu finden, mehrere Faktoren gibt, die beachtet werden müssen. Nachdem in Abschnitt 5.1 sichergestellt wurde, daß die hier verwendete Datengrundlage bei gleicher Herangehensweise vergleichbare Ergebnisse liefert, wurde in Abschnitt 5.4 gezeigt, daß bei einer kontrollierten Veränderung der Zuordnung individueller Sonoritätsklassen in die eine oder andere Kategorie ( $X$  oder  $Y$ ) die resultierende Verteilung der Sprachen in der  $\%X/\Delta Y$ -Ebene nicht gleichmäßig beeinflußt wird. Das bedeutet, daß die verschiedenen Sonoritätsklassen, bei einer solchen Analyse nicht gleich wichtig sind. In Abschnitt 5.5 wurde der Einfluß von silbischen Tokens zweier entscheidenden Sonoritätsklassen überprüft und festgestellt, daß die Syllabizität nicht übermäßig für den hohen Einfluß solcher Sonoritätsklassen verantwortlich ist. Gleichzeitig wurde bemerkt, daß der silbische Anteil dieser Sonoritätsklassen in akzentzählenden Sprachen höher ist als in silbenzählenden. Daß die in der Datengrundlage enthaltene hohe Variabilität der Sprechgeschwindigkeit sich in wesentlichem Maße auf die Standardabweichung insbesondere in  $\Delta$ -Parametern auswirkt, wurde in Anlehnung an Dellwo & Wagner (2003) in Abschnitt 5.2 erläutert.

Schließlich wurde in Abschnitt 5.6 mittels einer Diskriminanzanalyse eine statistische Untersuchung aller Parameter durchgeführt und dabei gezeigt, daß diese Parameter bei einer Charakterisierung jeder einzelnen Sprache in unterschiedlichem Maße die Verteilung beeinflussen. Darüber hinaus konnte gezeigt werden, daß die Parameter, die am zuverlässigsten alle Sprachen voneinander trennen oder nach Rhythmusklasse in Gruppen anordnen nicht die von Ramus et al. (1999) vorgeschlagenen Parameter  $\%v$  und  $\Delta lLnNfs$  sind, sondern lediglich  $\%l$  und  $\%n$ .

All dies bedeutet erstens, daß eine Segmentierung lediglich in konsonantische und vokalische Intervalle selbst bereits die Ergebnisse beeinflußt, da bestimmte Sonoritätsklassen unveränderlich entweder der einen oder der anderen Intervallklasse zugeteilt werden. Zweitens sind bei der rhythmischen Analyse einzelner Sprachen die ein-

zelenen Sonoritätsklassen verschieden wichtig, was genutzt werden kann, um anhand weniger Parameter gezielt bestimmte Anordnungen und Gruppierungen der Einzelsprachen zu erreichen. Gleichzeitig wird jede Einzelsprache durch diese Parameter in ihrer Anordnung rhythmisch charakterisiert.

Zwar sind diese Schlußfolgerungen nicht über jeden Zweifel erhaben, doch scheinen sie zumindest für die in der vorliegenden Arbeit verwendete Datengrundlage Gültigkeit zu besitzen. Mit einer Erweiterung dieser Grundlage um Material aus weiteren Sprachen und/oder weiteren bzw. anderen Aufnahmen aus den bereits untersuchten Sprachen können diese Schlußfolgerungen natürlich untergraben oder auch untermauert werden.

## 6 Diskussion und Perspektiven

### 6.1 Kritikpunkte

Es gibt einige Faktoren, die bei den Daten und Methoden der vorliegenden Arbeit kritisiert werden könnten, die in vergleichbarer Form bereits in anderen Untersuchungen auftreten. Beispielsweise wurde hier darauf verzichtet, phrasenfinale Betonungsintervalle von der Untersuchung auszuschließen. Gibbon & Gut (2001) monieren diesen Punkt an mehreren früheren Arbeiten, und tatsächlich scheint die Vermutung begründet, daß besonders im Englischen und Italienischen, wo phrasenfinale Längung besonders ausgeprägt zu beobachten ist, die rhythmischen Parameter durch eine solche Maßnahme beeinflußt würden.

Mehrere Einwände sprechen jedoch gegen diese Vorgehensweise. Erstens würde sie den Vergleich mit anderen Untersuchungen erschweren, in denen phrasenfinale Betonungsintervalle in den Resultaten enthalten sind. Zweitens – und dieser Grund scheint ausschlaggebend – ist phrasenfinale Längung in Sprachen, die dieses Merkmal aufweisen, eine wesentliche Rhythmuskomponente, deren Ausschluß den rhythmischen Eigenschaften dieser Sprachen nicht gerecht würde.

Gelesene Sprache eignet sich aus den in Kapitel 3 genannten Gründen am besten für eine Untersuchung, in der die Datengrundlage aus mehreren Aufnahmen desselben Textes durch dieselben Sprecher bestehen soll. Auf andere Weise könnte nicht gewährleistet werden, daß bei den Wiederholungen der Aufnahmen in verschiedenen Geschwindigkeiten der lexikalische und syntaktische Gehalt identisch bleibt. Darüber hinaus würde mit Spontansprache eine Fülle von Phänomenen wie Häsitationen, Stammeln, *false starts*, Ellipsen, etc. in das Korpus gebracht, die zu deutlichen Beeinflussungen der zu untersuchenden Parameter geführt hätten.

Die Entscheidung, Daten von Sprechern aus Schottland und dem mittleren Westen der USA gleich zu behandeln, könnte ebenfalls kritisiert werden. In der Tat unterscheiden

sich diese Aufnahmen in den Ergebnissen der Untersuchungen rhythmisch geringfügig voneinander. Allerdings kann anhand von Abb. 22 beobachtet werden, daß diese Unterschiede nicht ohne weiteres durch die verwendeten rhythmischen Parameter erklärt werden können. Dort liegen zwar die beiden Sprecher St und Ni (USA) etwas abseits der Sprecher Ly und Ji (Schottland), doch Sprecher Ai, der ebenfalls schottisches Englisch spricht, liegt zwischen den beiden amerikanischen Sprechern. Somit kann nur vermutet werden, daß sich die regiolektale Heterogenität der englischen Sprecher auf die Streuung der Werte für das Englische auswirkt, die bei dieser Konfiguration vergleichsweise höher ist, als etwa die der Werte des Französischen. Allerdings kann diese Streuung auch an anderen Faktoren liegen, wie sich anhand von einigen Konfigurationen, wie beispielsweise in Abb. 32 dargestellt, zeigt.<sup>18</sup>

Es liegt ebenso nahe, anzunehmen, daß die Entscheidung, die englischen Aufnahmen unabhängig von regiolektalen Einflüssen zusammen zu untersuchen, die rhythmischen Gemeinsamkeiten der beiden Varietäten des Englischen hervorhebt. Während eine Untersuchung, die rhythmische Merkmale bestimmter Dialekte und Regiolekte derselben Sprache vergleicht, durchaus interessante Resultate haben könnte, war eine in dieser Hinsicht differenzierte Untersuchung nicht Sinn der vorliegenden Arbeit.

Die hohe Streuung der Daten, die aufgrund der Verwendung von Aufnahmen mit stark unterschiedlicher Sprechgeschwindigkeit derselben Sprecher entstanden ist, wurde ebenso bewußt in Kauf genommen. Es könnten auch alternative Methoden gefunden werden, um die Sprechgeschwindigkeit als Artefaktquelle bei Untersuchungen von Rhythmus zu minimieren. Allerdings erschien es Dellwo & Wagner (2003) bei der Planung des BonnTempo-Korpus weniger aufwendig und zugleich leichter durchzuführen, die Sprechgeschwindigkeit schon bei der Aufnahme der Sprachdaten kontrolliert variieren zu lassen, als robuste Verfahren zur Normalisierung zu finden, die keinen Einfluß auf rhythmische Parameter ausüben, sofern es solche Verfahren überhaupt gibt.

---

<sup>18</sup> Ähnliche Überlegungen gelten für das Italienische.

Weiterhin läßt sich die geringe Größe des Korpus kritisieren, das der vorliegenden Arbeit als Datengrundlage dient. Angesichts beschränkter Mittel und Zeit war es jedoch nicht praktikabel, ein größeres Fragment des BonnTempo-Korpus zu untersuchen. Darüber hinaus wurde das Korpus während der Erstellung der vorliegenden Arbeit erweitert und war daher noch nicht vollständig manuell annotiert. Allerdings sollte bedacht werden, daß während die Anzahl der untersuchten Sprachen kleiner ist als bei manchen vergleichbaren bisherigen Untersuchungen, hier darauf geachtet wurde, für jede Sprache und für jeden Sprecher genügend Aufnahmen mit einzubeziehen, um damit verbundene Artefaktquellen zu minimieren.

Dadurch wurde ebenfalls vermieden, daß die Untersuchung von zu vielen unbekanntem oder streuenden Faktoren durch bislang unklassifizierte oder umstrittene Sprachen beeinflusst wurde.

## **6.2 Maschinelle Parameterextraktion**

Eine vielversprechende Möglichkeit, erheblich größere Korpora rhythmisch zu untersuchen, wird durch die maschinelle Extraktion rhythmischer Parameter aus nicht annotierten Rohdaten ermöglicht. Unter der Annahme, daß das theoretische Fundament für eine Implementierung dieser Möglichkeit solide ist, zeigen die Untersuchungen von Galves et al. (2002a) und Cassandro et al. (2002), daß sich die untersuchten Sprachen und Aufnahmen in einer charakteristischen linearen Verteilung gruppieren. Allerdings wurde in Abschnitt 4.5 der vorliegenden Arbeit gezeigt, daß bei der Anwendung der maschinellen Implementierung von Garcia et al. (2002), *Vocale*, auf die hier untersuchte Datengrundlage sehr starke Abweichungen von den manuell aufbereiteten Daten auftraten. Dabei wurde eine Verteilung der maschinell annotierten Parameter beobachtet, die der bei Cassandro et al. (2002) ähnelt.

Zwar überspringen Galves et al. (2002a) den „Zwischenschritt“, für die untersuchten Sprachdaten eine Annotation zu erstellen und aus dieser rhythmische Parameter zu extrahieren, indem sie anhand einer Sonoritätsfunktion solche Parameter unmittelbar aus dem Sprachsignal berechnen, aber die zugrundeliegenden Algorithmen sind es-

sentiell die gleichen wie bei Garcia et al. (2002). Daher ist angesichts des Vergleichs mit den manuell annotierten Daten des BonnTempo-Korpus nicht einwandfrei geklärt, wie robust diese Methoden gegenüber Sprachdaten sind, die sich in ihrer Beschaffenheit (Sprache, Sprechgeschwindigkeit, Geschlecht der Sprecher, etc.) von denen unterscheiden, an denen diese Implementierungen erprobt wurden. Ebenso ist es bei komplexeren Methoden sehr viel schwieriger, Diskrepanzen zwischen einer Sprachklassifikation anhand von maschinell extrahierten Parametern und einer Klassifikation durch Parameter aus manueller Annotation zu erklären, zumindest anhand der Datengrundlage.

Einen nicht zu unterschätzenden Schritt stellt allerdings die Parametrisierung der Sonorität dar (Galves et al. 2002a). Wie in Ramus et al. (1999) angedeutet und zu Beginn von Kapitel 3 erläutert, können auf diese Weise sowohl die Probleme der *C/V*-Segmentierung umgangen, als auch ein Modell geschaffen werden, das nicht anhand einer diskreten Grundlage, sondern anhand einer kontinuierlichen und automatisch berechenbaren Sonoritätskurve rhythmische Parameter extrahiert. Dieser Ansatz sollte nach Ansicht des Autors überprüft und weiterentwickelt werden.

### **6.3 Anwendungsbereiche**

Ein zuverlässiges Rhythmusmodell kann, je nach Entwicklung, in verschiedenen Anwendungsbereichen eingesetzt werden.

- Im Fremdspracherwerb fällt es dem Schüler in den wenigsten Fällen leicht, prosodische Eigenschaften der Zielsprache zu erlernen und sicher zu verwenden. Wer schon mit Grammatik, Lexikon, segmentaler Phonologie und gegebenenfalls noch Schriftsystem zur Genüge gefordert ist, hat selten noch die Möglichkeit, auf „Nuancen“ wie Rhythmus oder andere prosodische Merkmale der Zielsprache zu achten. Die Verständigung mit Sprechern dieser Sprache ist fast immer das Maß, an dem Lernerfolg gemessen wird, und dazu sind suprasegmentale Merkmale nur bei wenigen Sprachen unbedingt notwendig. Hinzu kommt, daß in traditionellen Lehrmethoden Merkmalen wie Rhythmus oft nur wenig Aufmerksamkeit gewid-

met wird. Dabei würde es beispielsweise für manche Nichtmuttersprachler des Englischen (je nach Muttersprache) eine deutliche Verringerung ihres „Akzents“ bedeuten, wenn sie unbetonte Silben reduzieren würden (Dauer 1987, s. auch Gut 2003).

- Kinder scheinen im Erstspracherwerb neueren Untersuchungen zufolge (Mehler et al. 1996, Nazzi et al. 1998, Ramus et al. 1999) vor allem in frühen Phasen wesentlich auf rhythmische Eigenschaften von Sprache zurückzugreifen, sowohl bei der kognitiven Verarbeitung von gehörten Äußerungen als auch bei ihren Lernprozessen, selbst Sprache zu produzieren.
- Ebenfalls denkbar ist der Einsatz von in Rhythmusforschung gewonnenen Resultaten im Bereich der automatischen Sprachenerkennung. Farinas & Pellegrino (2001) betonen die Wichtigkeit von rhythmischer Information bei der automatischen Erkennung von Sprachen als Vorstufe von sprachspezifischer Spracherkennung. Bei einer Untersuchung von Daten aus den Sprachen Deutsch, Englisch, Französisch, Italienisch und Spanisch implementieren sie erfolgreich ein System, das diese Sprachen mit einer Zuverlässigkeit von 81% anhand rhythmischer Parameter erkennt. Dort setzt dies allerdings eine automatische *C/V*-Segmentierung voraus. Es zeigt sich also anhand dieser Untersuchung, daß mit relativ einfachen statistischen Mitteln rhythmische Information zur Sprachenerkennung genutzt werden kann.
- Ein zuverlässiges Modell von Sprachrhythmus kann ebenso in Prosodiemodulen von Sprachsynthesystemen implementiert werden. Es ist sogar ein Rhythmusmodul innerhalb von Prosodiemodulen denkbar, das je nach Sprache und Verwendungskontext (Vorleseautomat, *Speech User Interface*, etc.) ausgetauscht wird, ohne daß sich die übrigen Bestandteile des sprachspezifischen Prosodiemoduls ändern. Auf diese Weise kann die Natürlichkeit eines Synthesystems weiter gesteigert werden.

#### **6.4 Weitere Forschungsansätze**

Die Tatsache, daß über so viele Jahre hinweg zahlreiche Linguisten und Phonetiker von der rhythmischen Dichotomie zumindest ansatzweise überzeugt waren, macht es nicht leicht, sie zu widerlegen. Tatsächlich kann trotz der Widerlegung der Isochronie-Hypothese (s. Abschnitt 2.2) noch von verschiedenen rhythmischen Eigenschaften einzelner Sprachen gesprochen werden, unabhängig von akustischen Parametern. Die perzeptive Realität verschiedener Rhythmusklassen ist noch immer ein nicht hinreichend erforschtes Gebiet, das auch in jüngster Zeit noch mit Perzeptionsexperimenten erforscht wird (Ramus et al. 2003).

Ebenso interessant ist die Untersuchung von rhythmischen Unterschieden nicht nur zwischen verschiedenen Sprachen oder Sprachgruppen, sondern zwischen verschiedenen Varietäten derselben Sprache, wie in Abschnitt 6.1 angedeutet. Dabei können Regiolekte, Soziolekte, funktionale Register (Nachrichtensprecher, Gedichtrezitationen, usw.) und ähnliches, selbst Idiolekte auf ihre rhythmischen Charakteristika hin analysiert werden. Frota & Vigário (2001) stellen beispielsweise bei einem Vergleich von europäischem und brasilianischem Portugiesisch fest, daß die rhythmischen Unterschiede zwischen diesen beiden regiolektalen Varietäten so groß sind, daß sie erstere als akzentzählend, letztere hingegen als silbenzählend beschreiben. Ähnlich große Unterschiede finden Tongue (1979) sowie Platt & Weber (1980) bei britischem und Singapur-Englisch, und Gibbon & Gut (2001) bei einem Vergleich von britischem und nigerianischem Englisch.

Die Untersuchung von Varietäten derselben Sprachen braucht sich nicht auf den synchronen Bereich zu beschränken. Je nach Daten ist auch eine diachrone Untersuchung der rhythmischen Entwicklung im Sprachwandel denkbar. Luangthongkum (1977) zufolge war das Thai vor dem 14. Jahrhundert eine silbenzählende Sprache, während das heutige Thai akzentzählenden Rhythmus besitzt. Die umgekehrte Entwicklung machen Donegan & Stampe (1983) vom Altfranzösischen zum modernen Französisch hin aus. Allerdings muß beachtet werden, daß, so interessant die Ergebnisse sein mögen, die im Sprachwandel hinsichtlich Sprachrhythmus gewonnen werden können, es

aus verständlichen Gründen nicht uneingeschränkt möglich oder sinnvoll ist, solche diachronen Untersuchungen durchzuführen.

In Abschnitt 4.2 wurde die gezielte Untersuchung von Rhythmus bei bilingualen Sprechern erwähnt. Insbesondere in Fällen, in denen es starke rhythmische Unterschiede zwischen den Muttersprachen solcher Sprecher gibt, wäre es interessant, zu untersuchen, wie solche Sprecher Rhythmus produzieren, wahrnehmen, und erlernen.

Bisher wurden lediglich rhythmische Parameter untersucht, die sich aus der Dauer bestimmter Segmenttypen ableiten lassen. Andere Rhythmuskomponenten wurden dabei nicht näher untersucht. Dauerparameter sind problemlos experimentell meßbar und kommen in jeder Sprache vor. Rhythmuskomponenten wie Intensität oder Grundfrequenz sind nicht in jeder Sprache gleichermaßen vorhanden, und wo sie vorkommen, werden sie von den verschiedensten phonologischen Eigenschaften der jeweiligen Sprache beeinflusst und sind nicht mit der gleichen Zuverlässigkeit meßbar.

Wenn aber nur eine Auswahl von Sprachen untersucht werden soll, könnte eine Ausdehnung der rhythmischen Parameter auf die Schnittmenge aller Sprachen mit solchen Komponenten (die vermutlich größer ist als die Dauer allein) womöglich weiterführende Ergebnisse liefern.

## **6.5 *Schlußbemerkungen***

Es ist kaum möglich, Rhythmus von den übrigen segmentalen und prosodischen Charakteristika einer Sprache zu isolieren und getrennt zu betrachten. Je nach Datengrundlage, Methoden, Untersuchungsparametern und -zielen ist es dennoch möglich, anhand bestimmter rhythmischer Parameter Sprachen zu unterscheiden und sprachinhärente Eigenschaften zu beschreiben. Dabei darf allerdings nicht aus den Augen verloren werden, daß die Rhythmusparameter, die in der vorliegenden Arbeit untersucht wurden, lediglich statistische Maße sind, die direkt mit der Silbenstruktur und Vorkommenshäufigkeit bestimmter Laute in verschiedenen Sprachen zusammenhängen.

## 7 Literatur

- Abe, Isamu (1967) „English sentence rhythm and synchronism.“ *Bulletin of the Phonetic Society of Japan* 125, 9-11. [zit. n. Allen (1975)]
- Abercrombie, David (1967) *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Allen, George D. (1975) „Speech rhythm: Its relation to performance universals and articulatory timing.“ *Journal of Phonetics* 3, 75-86.
- Asher, Ron E. (1985) *Tamil*. London: Croom Helm. [zit. n. Grabe & Low (2002)]
- Auer, Peter (1993) „Is a rhythm-based typology possible? A study of the role of prosody in phonological typology.“ *KonTRI Working Paper* 21.  
[http://fips.igl.uni-freiburg.de/auer/Phonotypo\\_Kontril.pdf](http://fips.igl.uni-freiburg.de/auer/Phonotypo_Kontril.pdf)
- Barbosa, Plínio A. (2002) „Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production.“ *Proceedings of Speech Prosody 2002*. Aix-en-Provence, 163-166.  
<http://www.lpl.univ-aix.fr/sp2002/pdf/barbosa.pdf>
- Beckmann, Mary E. (1982) „Segment duration and the ‘mora’ in Japanese.“ *Phonetica* 39, 113-135.
- Beckmann, Mary E. (1992) „Evidence for speech rhythms across languages.“ In: Yoh’ichi Tohkura, Eric Vatikiotis-Bateson & Yoshinori Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*. Amsterdam: IOS Press, 457-463.
- Bloch, Bernard (1950) „Studies in colloquial Japanese IV: Phonemics.“ *Language* 26, 86-125.
- Boersma, Paul & David Weenink (2001) „PRAAT, a system for doing phonetics by computer.“ *Glott International* 5, 341-345.  
<http://www.praat.org>
- Bolinger, Dwight L. (1965) „Pitch accent and sentence rhythm.“ In: Isamu Abe & Tetsuya Kanekiyo (eds.), *Forms of English: Accent, Morpheme, Order*. Cambridge, MA: Harvard University Press, 139-180. [zit. n. Nespor (1990)]
- Borzzone de Manrique, Ana M. & Angela Signorini (1983) „Segmental duration and rhythm in Spanish.“ *Journal of Phonetics* 11, 117-128.
- Bußmann, Hadumod (1990) *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner Verlag.

- Cassandro, Marzio, Pierre Collet, Denise Duarte, Antonio Galves & Jesus Garcia (2002) „A universal linear correlation among acoustic correlates of rhythm.“  
<http://www.ime.usp.br/~tycho/prosody/sonority/linear/linear.pdf>
- Catford, John C. (1977) *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.
- Classe, André (1939) *The Rhythm of English Prose*. Oxford: Basil Blackwell. [zit. n. Dauer (1983)]
- Coleman, John (1999) „The nature of vocoids associated with syllabic consonants in Tashlhiyt Berber.“ *Proceedings of the 14th International Congress of Phonetic Sciences*. San Franzisko, Bd. 1, 735-738.  
<http://www.phon.ox.ac.uk/~jcoleman/ICPhS.ps>
- Corder, S. Pit (1973) *Introducing Applied Linguistics*. Harmondsworth, Middlesex: Penguin Books.
- Couper-Kuhlen, Elisabeth (1990) „Discovering rhythm in conversational English: Perceptual and acoustic approaches to the analysis of isochrony.“ *KonTRI Working Paper 13*: Konstanz. [zit. n. Grabe & Low (2002)]
- Couper-Kuhlen, Elisabeth (1993) *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*. Amsterdam: John Benjamins Publishing.
- Cruttenden, Alan (1986, 1997) *Intonation*. Cambridge: Cambridge University Press.
- Daneš (1965) In: Bolinger (1965). [zit. n. Dauer (1983)]
- Dasher, Richard & Dwight L. Bolinger (1982) „On pre-accentual lengthening.“ *Journal of the International Phonetic Association* 12, 58-71.
- Dauer, Rebecca M. (1983) „Stress-timing and syllable-timing reanalyzed.“ *Journal of Phonetics* 11, 51-62.
- Dauer, Rebecca M. (1987) „Phonetic and phonological components of language rhythm.“ *Proceedings of the 11th International Congress of Phonetic Sciences*. Tallinn, Bd. 5, 447-450.
- Delattre, Pierre (1966) „A comparison of syllable length conditioning among languages.“ *International Review of Applied Linguistics in Language Teaching*, 3, 183-198. [zit. n. Grabe & Low (2002)]
- Dell, François & Mohamed Elmedlaoui (1985) „Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber.“ *Journal of African Languages and Linguistics*, 7, 105-130.

- Dellwo, Volker & Petra Wagner (2003) „Relations between language rhythm and speech rate.“ *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona.
- Donegan, Patricia & David Stampe (1983) „Rhythm and the holistic organization of language structure.“ In: John F. Richardson, Mitchell Marks & Amy Chukerman (eds.), *Parasession on the interplay of phonology, morphology & syntax*, *Chicago Linguistic Society* 19, 337-353. [zit. n. Auer (1993)]
- Donovan, Andrew & C. J. Darwin (1979) „The perceived rhythm of speech.“ *Proceedings of the 9th International Congress of Phonetic Sciences*. Kopenhagen, Bd. 2, 268-274.
- Duarte, Denise, Antonio Galves, Nancy L. Garcia & Ricardo Maronna (2001) „The statistical analysis of acoustic correlates of speech rhythm.“ *Proceedings of the Workshop on Rhythmic Patterns, Parameter Setting and Language Change*. Bielefeld: ZiF, Universität Bielefeld.  
[http://www.ime.usp.br/~tycho/zif/zif\\_papers/duarte.pdf](http://www.ime.usp.br/~tycho/zif/zif_papers/duarte.pdf)
- Eriksson, Anders (1991) *Aspects of Swedish speech rhythm* (Dissertation). *Gothenburg Monographs in Linguistics* (Universität Göteborg) 9. [zit. N. Grabe & Low (2002)]
- Farinas, Jérôme & François Pellegrino (2001) „Automatic rhythm modeling for language identification.“ *Proceedings of Eurospeech 2001*. Aalborg.  
<http://www.ddl.ish-lyon.cnrs.fr/PDF/pellegrino/FP-Eurospeech2001.pdf>
- Faure, Georges, Daniel J. Hirst & Michel Chafcouloff (1980) „Rhythm in English: Isochronism, pitch and perceived stress.“ In: Linda R. Waugh & Cornelis H. van Schooneveld (eds.), *The Melody of Language*. Baltimore: University Park Press, 71-79.
- Frota, Sónia & Marina Vagário (2001) „On the correlates of rhythm distinctions: The European/Brazilian Portuguese case.“ *Probus* 13, 247-275. [zit. n. Galves et al. (2002a)]  
[http://www.ime.usp.br/~tycho/zif/zif\\_papers/frota.pdf](http://www.ime.usp.br/~tycho/zif/zif_papers/frota.pdf)
- Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves (2002a) „Sonority as a basis for rhythmic class discrimination.“ *Proceedings of Speech Prosody 2002*. Aix-en-Provence.  
<http://www.ime.usp.br/~tycho/prosody/sonority/rev4.pdf>
- Galves, Antonio, Jesus Garcia, Denise Duarte & Charlotte Galves (2002b) „Appendix: Are the lengths of consonantal intervals correlated?“ *Proceedings of Speech Prosody 2002*. Aix-en-Provence.  
<http://www.ime.usp.br/~tycho/prosody/sonority/appendix.pdf>

- Garcia, Jesus, Ulrike Gut & Antonio Galves (2002) „Vocale – A semi-automatic annotation tool for prosodic research.“ *Proceedings of Speech Prosody 2002*. Aix-en-Provence.  
<http://www.ime.usp.br/~tycho/prosody/vocale/vocale.pdf>  
<http://www.ime.usp.br/~tycho/participants/garcia/ba5.ppt>
- Gibbon, Dafydd & Ulrike Gut (2001) „Measuring speech rhythm.“ *Proceedings of Eurospeech 2001*. Aalborg, 95-98.
- Grabe, Esther & Ee Ling Low (2002) „Durational variability in speech and the rhythm class hypothesis.“ In: Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter.  
[http://www.phon.ox.ac.uk/~esther/ivyweb/Grabe\\_Low.doc](http://www.phon.ox.ac.uk/~esther/ivyweb/Grabe_Low.doc)
- Gussenhoven, Carlos & Haike Jacobs (1998) *Understanding Phonology*. London: Arnold.
- Gut, Ulrike (2003) „Non-native speech rhythm in German.“ *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 2437-2440.
- Han, Mieko S. (1962) „The feature of duration in Japanese.“ 音声の研究 (*Onsei no kenkyū*) 10, 65-80. [zit. n. Grabe & Low (2002)]
- Hayes, Bruce & Stanislaw Puppel (1985) „On the rhythm rule in Polish.“ In: Harry van der Hulst & Norval Smith (eds.), *Advances in Nonlinear Phonology*. Dordrecht: Foris, 59-81. [zit. n. Ramus et al. (1999)]
- Hoequist, Charles (1983a) „Durational correlates of linguistic rhythm categories.“ *Phonetica* 40, 19-31.
- Hoequist, Charles (1983b) „Syllable duration in stress-, syllable- and mora-timed languages.“ *Phonetica* 40, 203-237.
- Hockett, Charles F. (1958) *A Course in Modern Linguistics*. New York: Macmillan.
- International Phonetic Association, *Handbook of the* (1999), Cambridge: Cambridge University Press.
- Jespersen, Otto (1906) *Modersmålets Fonetik*. Kopenhagen: Det Schuboeske Forlag.
- Kohler, Klaus (1982) „Rhythmus im Deutschen.“ *Arbeitsberichte* (Institut für Phonetik, Universität Kiel) 19, 89-106.
- Ladefoged, Peter (1967) *Three Areas of Experimental Phonetics*. London: Oxford University Press.
- Ladefoged, Peter (1975) *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Laver, John (1994) *Principles of Phonetics*. Cambridge: Cambridge University Press.

- Lehiste, Ilse (1977) „Isochrony reconsidered.“ *Journal of Phonetics* 5, 253-263.
- Levelt, Claartje & Ruben van de Vijver (1998) „Syllable types in cross-linguistic and developmental grammars.“ *Proceedings of the Third Biannual Utrecht Phonology Workshop*. Utrecht.  
<http://roa.rutgers.edu/files/265-0698/roa-265-levelt-3.pdf>
- Lloyd James, Arthur. (1940) *Speech Signals in Telephony*. London. [zit. n. Nespor (1990)]
- Luangthongkum, Theraphan (1977) *Rhythm in Thai* (Dissertation). University of Edinburgh. [zit. n. Dauer (1983); Grabe & Low (2002)]
- Mascaró, Joan (1983) *La Fonologia Catalana i el Cicle Fonològic*. Bellaterra.
- Mehler, Jacques, Emmanuel Dupoux, Thierry Nazzi & Ghislaine Dehaene-Lambertz (1996) „Coping with linguistic diversity: The infant’s viewpoint.” In: James L. Morgan & Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, 101-116. [zit. n. Ramus et al. (1999); Galves et al. (2002a)]
- Nakatani, Lloyd H., Kathleen D. O’Connor & Charletta H. Aston (1981) „Prosodic aspects of American English speech rhythm.“ *Phonetica* 38, 84-106.
- Nazzi, Thierry, Josiane Bertoncini & Jacques Mehler (1998) „Language discrimination by newborns: Towards an understanding of the role of rhythm.“ *Journal of Experimental Psychology: Human Perception and Performance* 24, 756-766. [zit. n. Ramus et al. (1999)]
- Nespor, Marina (1990) „On the rhythm parameter in phonology.“ In: Iggy Roca (ed.), *Logical Issues in Language Acquisition*. Dordrecht: Foris, 157-175.
- O’Connor, Joseph D. (1965) „The perception of time intervals.“ *Progress Report* (University College London) 2, 11-15.
- O’Connor, Joseph D. (1973) *Phonetics*. Harmondsworth, Middlesex: Penguin Books.
- O’Dell, Michael L. & Tommi Nieminen (1999) „Coupled oscillator model of speech rhythm“ *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, Bd. 1, 1075-1078.  
[http://kotisivu.mtv3.fi/tommi\\_nieminen/tiede/odell-nieminen99.pdf](http://kotisivu.mtv3.fi/tommi_nieminen/tiede/odell-nieminen99.pdf)
- Ohala, John J., Carol J. Riordan & Haruko Kawasaki (1979) „Investigation of pulmonic activity in speech.“ *Proceedings of the 9th International Congress of Phonetic Sciences*. Copenhagen, Bd. 1, 205.

- Oyakawa, Takatsugu (1971) „On the directionality of segmental conditioning in Japanese.“ *Monthly Internal Memo* (Department of Linguistics, University of California, Berkeley), 81-103. [zit. n. Grabe & Low (2002)]
- Pike, Kenneth L. (1945) *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- Platt, John & Heidi Weber (1980) *English in Singapore and Malaysia: Status, Features, Functions*. Kuala Lumpur: Oxford University Press. [zit. n. Grabe & Low (2002)]
- Pointon, Graham E. (1980) „Is Spanish really syllable-timed?“ *Journal of Phonetics* 8, 293-304.
- Pompino-Marschall, Bernd (1995) *Einführung in die Phonetik*. Berlin: Walter de Gruyter.
- Port, Robert F., Salman Al-Ani & Shosaku Maeda (1980) „Temporal compensation and universal phonetics.“ *Phonetica* 37, 235-252.
- Port, Robert F., Jonathan Dalby, & Michael O’Dell (1987) „Evidence for mora-timing in Japanese.“ *Journal of the Acoustical Society of America* 81, 1574-1585.
- Prince, Alan & Paul Smolensky (1993) *Optimality Theory. Constraint Interaction in Generative Grammar. Technical Report #2*, Rutgers University Center for Cognitive Science.
- Ramus, Franck (2002) „Acoustic correlates of linguistic rhythm: Perspectives.“ *Proceedings of Speech Prosody 2002*. Aix-en-Provence, 115-120.  
[http://www.ehess.fr/centres/lscp/persons/ramus/ramus\\_sp02.pdf](http://www.ehess.fr/centres/lscp/persons/ramus/ramus_sp02.pdf)
- Ramus, Franck, Marina Nespors & Jacques Mehler (1999) „Correlates of linguistic rhythm in the speech signal.“ *Cognition* 73, 265-292.  
<http://www.ehess.fr/centres/lscp/persons/ramus/Cognition99.pdf>
- Ramus, Franck, Emmanuel Dupoux & Jacques Mehler (2003) „The psychological reality of rhythm classes: Perceptual studies.“ *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 337-342.  
<http://www.ehess.fr/centres/lscp/persons/ramus/ICPhS03.pdf>
- Roach, Peter (1982) „On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages.“ In: David Crystal (ed.), *Linguistic Controversies*. London: Edward Arnold, 73-79.  
<http://www.personal.rdg.ac.uk/~llsroach/phon2/frp.pdf>
- Rubach, Jerzy & Geert E. Booij (1985) „A grid theory of stress in Polish.“ *Lingua* 66, 281-319.

- Scott, Donia R., S. D. Isard & Benedicte de Boysson-Bardies (1986) „On the measurement of rhythmic irregularity: a reply to Benguerel.“ *Journal of Phonetics* 14, 327-330.
- Shen, Y. & G. G. Peterson (1962) „Isochronism in English.“ *University of Buffalo Studies in Linguistics, Occasional Papers* 9, 1-36. [zit. n. Dauer (1983)]
- Smith, Alexandra (1975) „Syllable-structure and rhythm in Japanese.“ *Work in Progress* (Department of Linguistics, Edinburgh University) 8, 1-13.
- Steele, Joshua (1775) *An Essay towards Establishing the Melody and Measure of Speech*. London: Bowyer & Nichols. (Nachdruck, Menston: The Scolar Press, 1969) [zit. n. Couper-Kuhlen (1993)]
- Steever, Sanford B. (1987) „Tamil and the Dravidian languages.“ In: Bernard Comrie, *The Major Languages of South Asia, the Middle East and Africa*. London: Routledge, 231-252.
- Stetson, Raymond H. (1928, 1951) *Motor Phonetics*. Amsterdam: North-Holland.
- Tongue, Ray K. (1979) *The English of Singapore and Malaysia*. Singapore: Eastern Universities Press. [zit. n. Grabe & Low (2002)]
- Treisman, Michel (1963) „Temporal discrimination and the indifference interval.“ *Psychological Monographs* 77, 1-31. [zit. n. Dauer (1983)]
- Uldall, Elisabeth T. (1971) „Isochronous stresses in R.P.“ In: L. L. Hammerich, Roman Jakobson & Eberhard Zwirner (eds.), *Form and Substance*. Kopenhagen: Akademisk Forlag.
- Vennemann, Theo (1972) „On the theory of syllabic phonology.“ *Linguistische Berichte* 18, 1-18.
- Wenk, Brian J. & François Wioland (1982) „Is French really syllable-timed?“ *Journal of Phonetics* 10, 193-216.
- Wheeler, Max W. (1979) *Phonology of Catalan*. Oxford: Blackwell.
- Wiik, Kalevi (1991) „On a third type of speech rhythm: Foot timing.“ *Proceedings of the 12th International Congress of Phonetic Sciences*. Aix-en-Provence, Bd. 3, 298-301. [zit. n. O’Dell & Nieminen (1999)]
- Woodrow, Herbert (1951) „Time perception.“ In: Stanley S. Stevens (ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1224-1236. [zit. n. Dauer (1983)]

## 8 Anhang

### Parameterwerte der einzelnen Aufnahmen (Auswahl):

#### Deutsch

Aufnahme		%v	$\Delta v$	%a	$\Delta a$	%lL	$\Delta lL$	%l	$\Delta l$	%L	$\Delta L$
Bi	l1	42,393	0,049	0,000	0,000	1,875	0,052	1,875	0,017	1,059	0,000
Bi	l2	42,492	0,042	0,000	0,000	2,886	0,042	2,886	0,042	0,000	0,000
Bi	no	44,214	0,039	0,000	0,000	1,679	0,029	1,679	0,022	0,723	0,000
Bi	s1	44,445	0,030	0,570	0,006	2,712	0,024	2,712	0,025	0,630	0,000
Bi	s4	44,812	0,029	0,000	0,000	1,602	0,019	1,602	0,018	0,586	0,000
Ch	l1	40,392	0,055	0,000	0,000	1,941	0,026	1,941	0,024	0,497	0,000
Ch	l2	41,042	0,044	0,000	0,000	2,078	0,020	2,078	0,022	0,509	0,000
Ch	no	44,239	0,036	0,000	0,000	1,871	0,014	1,871	0,012	0,558	0,000
Ch	s1	42,363	0,030	0,000	0,000	2,353	0,031	2,353	0,034	0,507	0,000
Ch	s5	46,293	0,025	0,269	0,000	1,442	0,013	1,442	0,012	0,556	0,000
Ci	l1	36,469	0,042	0,000	0,000	2,176	0,061	2,176	0,032	1,155	0,000
Ci	l2	39,214	0,041	0,000	0,000	1,362	0,027	1,362	0,021	0,564	0,000
Ci	no	39,500	0,037	0,000	0,000	1,622	0,029	1,622	0,021	0,661	0,000
Ci	s1	40,415	0,040	0,000	0,000	2,223	0,025	2,223	0,026	0,596	0,000
Ci	s5	38,455	0,023	0,298	0,000	2,807	0,025	2,807	0,028	0,864	0,000
Cr	l1	43,538	0,048	0,000	0,000	0,905	0,091	0,905	0,016	1,149	0,000
Cr	l2	41,156	0,041	0,000	0,000	1,223	0,069	1,223	0,021	1,069	0,000
Cr	no	41,236	0,041	0,000	0,000	1,078	0,075	1,078	0,028	1,156	0,000
Cr	s1	41,416	0,033	0,000	0,000	1,732	0,032	1,732	0,029	0,648	0,000
Cr	s5	43,560	0,023	0,163	0,000	1,211	0,029	1,211	0,006	0,871	0,000
Durchschnitt		41,882	0,037	0,065	0,000	1,839	0,037	1,839	0,023	0,718	0,000
Standardabweichung		2,429	0,009	0,150	0,001	0,569	0,022	0,569	0,008	0,293	0,000

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Bi	l1	4,878	0,038	11,811	0,031	3,800	0,044	19,515	0,045	19,547	0,041
Bi	l2	5,856	0,046	10,931	0,032	5,110	0,063	20,996	0,038	17,585	0,034
Bi	no	4,132	0,043	11,320	0,032	4,895	0,064	20,924	0,035	16,245	0,025
Bi	s1	6,109	0,025	11,032	0,020	3,698	0,032	20,862	0,024	16,050	0,025
Bi	s4	3,827	0,019	12,018	0,014	5,153	0,025	20,282	0,025	15,548	0,021
Ch	l1	4,429	0,046	11,486	0,039	4,533	0,056	20,397	0,044	20,754	0,042
Ch	l2	4,706	0,036	10,500	0,025	4,353	0,054	22,067	0,045	19,452	0,051
Ch	no	4,326	0,025	11,734	0,020	4,329	0,033	20,464	0,029	16,805	0,029
Ch	s1	5,279	0,025	11,285	0,025	3,870	0,019	22,473	0,030	17,149	0,028
Ch	s5	3,465	0,025	10,709	0,026	4,613	0,016	21,845	0,015	14,273	0,031
Ci	l1	5,601	0,039	11,355	0,041	3,348	0,031	20,204	0,039	25,293	0,054
Ci	l2	3,336	0,050	11,633	0,049	4,561	0,056	21,140	0,032	21,526	0,044
Ci	no	3,956	0,043	11,385	0,028	4,884	0,069	21,397	0,031	20,550	0,037
Ci	s1	5,092	0,025	11,565	0,024	3,866	0,021	21,001	0,043	20,335	0,038
Ci	s5	6,531	0,028	10,823	0,017	6,050	0,040	20,901	0,024	19,802	0,026
Cr	l1	3,066	0,044	11,183	0,035	3,127	0,053	19,755	0,055	20,342	0,048
Cr	l2	3,604	0,042	11,521	0,033	4,156	0,036	20,908	0,050	19,969	0,044

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Cr	no	3,416	0,033	10,937	0,027	4,343	0,033	20,774	0,043	20,475	0,035
Cr	s1	4,173	0,030	10,760	0,023	4,247	0,038	20,518	0,036	20,680	0,037
Cr	s5	3,329	0,024	10,167	0,020	4,505	0,032	20,469	0,023	19,052	0,031
Durchschnitt		4,456	0,034	11,208	0,028	4,372	0,041	20,844	0,035	19,072	0,036
Standardabweichung		1,019	0,010	0,471	0,009	0,673	0,016	0,721	0,010	2,543	0,009

### Englisch

Aufnahme		%v	$\Delta v$	%a	$\Delta a$	%lL	$\Delta lL$	%l	$\Delta l$	%L	$\Delta L$
Ai	l1	46,185	0,054	2,578	0,010	3,114	0,032	3,114	0,032	0,000	0,000
Ai	l2	46,887	0,060	1,973	0,004	3,081	0,033	3,081	0,034	0,693	0,000
Ai	no	42,823	0,036	2,165	0,017	3,906	0,027	3,906	0,027	0,000	0,000
Ai	s1	45,799	0,039	2,625	0,024	2,215	0,021	2,215	0,017	0,747	0,000
Ai	s4	47,776	0,045	2,260	0,011	1,309	0,023	1,309	0,005	2,150	0,029
Ji	l1	36,844	0,048	3,537	0,033	3,142	0,035	3,142	0,034	1,415	0,038
Ji	l2	38,082	0,044	2,934	0,029	2,789	0,034	2,789	0,029	0,835	0,000
Ji	no	39,690	0,045	3,241	0,043	2,779	0,027	2,779	0,031	1,119	0,022
Ji	s1	39,493	0,036	3,015	0,037	2,662	0,024	2,662	0,029	1,128	0,009
Ji	s4	42,066	0,030	3,550	0,023	1,828	0,016	1,828	0,019	0,401	0,000
Ly	l1	40,289	0,053	3,162	0,028	2,801	0,037	2,801	0,039	0,620	0,000
Ly	l2	38,744	0,048	2,554	0,030	2,938	0,042	2,938	0,048	1,260	0,022
Ly	no	39,737	0,046	4,267	0,034	3,117	0,047	3,117	0,056	0,921	0,007
Ly	s1	39,865	0,043	2,868	0,031	3,434	0,040	3,434	0,043	0,551	0,000
Ly	s4	36,772	0,033	2,751	0,024	3,438	0,026	3,438	0,027	0,632	0,000
Ni	l1	43,677	0,053	3,324	0,027	3,703	0,059	3,703	0,059	0,000	0,000
Ni	l2	42,982	0,050	3,617	0,031	3,170	0,022	3,170	0,022	0,000	0,000
Ni	no	42,548	0,039	3,686	0,022	3,377	0,025	3,377	0,025	0,000	0,000
Ni	s1	41,701	0,035	2,413	0,029	2,307	0,024	2,307	0,025	0,206	0,000
Ni	s3	41,814	0,040	2,043	0,015	2,659	0,027	2,659	0,027	0,000	0,000
St	l1	45,428	0,063	3,720	0,038	3,309	0,052	3,309	0,052	0,000	0,000
St	l2	47,572	0,061	3,521	0,031	2,476	0,039	2,476	0,032	0,275	0,000
St	no	46,251	0,039	3,347	0,039	2,716	0,022	2,716	0,022	0,000	0,000
St	s1	47,663	0,038	2,656	0,029	3,041	0,016	3,041	0,016	0,000	0,000
St	s3	43,561	0,031	2,831	0,022	2,296	0,022	2,296	0,025	0,405	0,000
Durchschnitt		42,570	0,044	2,985	0,026	2,864	0,031	2,864	0,031	0,534	0,005
Standardabweichung		3,449	0,009	0,593	0,009	0,581	0,011	0,581	0,013	0,567	0,011

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Ai	l1	16,543	0,025	10,241	0,025	0,000	0,000	16,494	0,033	21,388	0,048
Ai	l2	16,891	0,024	9,964	0,024	0,000	0,000	17,063	0,034	20,337	0,028
Ai	no	17,025	0,021	9,141	0,022	1,887	0,004	18,997	0,031	21,081	0,027
Ai	s1	12,930	0,020	7,692	0,016	3,661	0,016	17,830	0,026	19,432	0,022
Ai	s4	13,887	0,015	9,049	0,014	2,738	0,017	18,484	0,029	16,234	0,017
Ji	l1	18,485	0,035	10,644	0,036	0,862	0,028	21,672	0,053	21,883	0,040
Ji	l2	19,031	0,049	12,527	0,049	0,000	0,000	20,336	0,040	22,496	0,041
Ji	no	18,083	0,024	11,285	0,024	0,000	0,000	20,359	0,045	21,528	0,041
Ji	s1	16,656	0,027	10,104	0,024	2,682	0,009	20,762	0,040	20,154	0,030
Ji	s4	11,679	0,019	7,565	0,013	4,870	0,011	20,179	0,031	19,541	0,021

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Ly	l1	0,000	0,037	10,552	0,037	0,000	0,000	19,592	0,062	22,985	0,044
Ly	l2	0,000	0,026	9,297	0,024	1,207	0,052	19,133	0,051	24,867	0,043
Ly	no	0,000	0,021	9,370	0,021	0,541	0,000	19,654	0,050	22,393	0,046
Ly	s1	0,000	0,025	9,300	0,021	2,047	0,039	18,836	0,048	23,098	0,034
Ly	s4	0,000	0,026	8,797	0,015	2,563	0,045	20,508	0,045	24,539	0,035
Ni	l1	0,000	0,049	12,203	0,048	0,703	0,000	18,805	0,048	17,585	0,040
Ni	l2	0,000	0,030	10,871	0,029	0,403	0,000	19,839	0,047	19,119	0,034
Ni	no	0,000	0,032	9,129	0,026	2,798	0,026	19,995	0,043	18,467	0,030
Ni	s1	0,000	0,026	9,088	0,020	2,447	0,011	22,936	0,039	18,902	0,034
Ni	s3	0,000	0,020	8,008	0,018	2,445	0,013	27,466	0,038	15,565	0,034
St	l1	0,000	0,041	9,623	0,041	0,754	0,000	19,017	0,058	18,149	0,043
St	l2	0,000	0,028	9,023	0,028	0,000	0,000	19,108	0,067	18,025	0,033
St	no	0,000	0,016	9,411	0,015	0,690	0,000	19,456	0,044	18,128	0,034
St	s1	0,000	0,023	10,093	0,024	0,388	0,000	20,180	0,037	15,979	0,028
St	s3	0,000	0,017	9,967	0,010	1,919	0,027	19,226	0,028	19,796	0,026
Durchschnitt		16,074	0,027	9,718	0,025	1,424	0,012	19,837	0,043	20,067	0,034
Standardabweichung		1,922	0,009	1,211	0,010	1,342	0,016	2,072	0,011	2,544	0,008

### Französisch

Aufnahme		%v	$\Delta v$	%a	$\Delta a$	%lL	$\Delta lL$	%l	$\Delta l$	%L	$\Delta L$
Ca	l1	48,661	0,042	1,265	0,015	6,581	0,028	6,581	0,027	0,622	0,000
Ca	l2	50,299	0,039	1,805	0,010	7,200	0,021	7,200	0,021	0,000	0,000
Ca	no	48,671	0,038	2,004	0,005	6,767	0,016	6,767	0,016	0,000	0,000
Ca	s1	49,675	0,033	1,162	0,003	6,462	0,014	6,462	0,014	0,000	0,000
Ca	s3	51,432	0,034	1,976	0,009	5,330	0,011	5,330	0,011	0,424	0,000
Da	l1	49,743	0,045	1,254	0,018	7,427	0,045	7,427	0,033	1,054	0,000
Da	l2	51,353	0,044	1,935	0,019	6,669	0,024	6,669	0,021	0,719	0,000
Da	no	47,099	0,037	2,593	0,013	6,859	0,023	6,859	0,020	0,765	0,000
Da	s1	47,486	0,035	3,035	0,020	6,326	0,016	6,326	0,016	0,000	0,000
Da	s6	49,990	0,023	2,962	0,014	5,473	0,014	5,473	0,014	1,308	0,006
Ju	l1	52,078	0,053	2,106	0,012	6,102	0,030	6,102	0,025	0,669	0,000
Ju	l2	48,882	0,041	2,943	0,027	5,618	0,050	5,618	0,013	1,390	0,000
Ju	no	49,185	0,045	2,026	0,017	5,571	0,039	5,571	0,016	2,319	0,076
Ju	s1	48,865	0,032	2,249	0,016	5,657	0,015	5,657	0,015	0,749	0,029
Ju	s3	47,167	0,021	3,450	0,009	6,530	0,015	6,530	0,015	0,000	0,000
Lo	l1	48,900	0,040	3,264	0,024	4,941	0,021	4,941	0,016	1,134	0,026
Lo	l2	49,493	0,042	3,273	0,020	5,856	0,017	5,856	0,016	0,526	0,000
Lo	no	48,600	0,037	3,547	0,020	5,397	0,020	5,397	0,014	0,815	0,000
Lo	s1	49,251	0,031	2,943	0,013	5,150	0,031	5,150	0,011	1,146	0,000
Lo	s4	49,935	0,025	3,111	0,011	5,613	0,018	5,613	0,018	0,000	0,000
Ma	l1	47,720	0,040	4,458	0,024	6,163	0,025	6,163	0,026	1,050	0,017
Ma	l2	48,189	0,031	3,060	0,019	7,561	0,023	7,561	0,023	0,000	0,000
Ma	no	47,405	0,031	3,332	0,013	8,250	0,019	8,250	0,019	0,000	0,000
Ma	s1	45,924	0,024	2,382	0,013	6,882	0,019	6,882	0,018	0,290	0,000
Ma	s4	45,216	0,021	1,949	0,021	6,632	0,021	6,632	0,011	0,635	0,000
Durchschnitt		48,849	0,035	2,563	0,015	6,281	0,023	6,281	0,018	0,624	0,006
Standardabweichung		1,627	0,008	0,820	0,006	0,829	0,010	0,829	0,005	0,584	0,017

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Ca	l1	11,662	0,034	4,442	0,030	1,569	0,001	21,794	0,037	15,067	0,032
Ca	l2	11,501	0,029	4,774	0,027	1,276	0,033	21,689	0,036	12,957	0,023
Ca	no	11,272	0,029	5,866	0,024	0,943	0,000	22,729	0,032	13,020	0,025
Ca	s1	13,467	0,018	5,800	0,018	0,593	0,000	22,201	0,029	14,107	0,029
Ca	s3	15,188	0,019	5,804	0,020	0,495	0,000	23,182	0,028	11,357	0,017
Da	l1	16,550	0,087	7,514	0,087	0,000	0,000	20,102	0,042	12,908	0,038
Da	l2	14,104	0,041	6,200	0,041	0,000	0,000	18,093	0,034	15,031	0,030
Da	no	13,951	0,020	6,351	0,021	0,506	0,000	19,682	0,033	16,145	0,020
Da	s1	13,248	0,028	6,144	0,027	0,739	0,000	19,158	0,023	17,112	0,020
Da	s6	2,279	0,018	5,440	0,018	0,000	0,000	20,428	0,017	14,399	0,015
Ju	l1	0,000	0,046	5,156	0,046	0,000	0,000	18,736	0,040	15,154	0,040
Ju	l2	19,248	0,031	6,533	0,031	0,000	0,000	19,495	0,041	15,139	0,032
Ju	no	18,618	0,030	5,009	0,029	1,870	0,025	19,474	0,028	14,545	0,032
Ju	s1	17,488	0,022	5,351	0,023	0,912	0,010	20,900	0,027	15,319	0,022
Ju	s3	19,004	0,018	5,905	0,019	0,530	0,000	20,118	0,021	16,300	0,015
Lo	l1	16,408	0,052	5,305	0,032	1,527	0,054	19,406	0,031	15,522	0,026
Lo	l2	18,176	0,045	5,885	0,025	1,551	0,083	19,227	0,032	14,190	0,023
Lo	no	17,178	0,029	5,516	0,026	2,457	0,025	18,158	0,024	15,510	0,022
Lo	s1	17,728	0,022	6,227	0,020	1,504	0,015	17,422	0,021	16,357	0,021
Lo	s4	17,888	0,014	6,616	0,014	0,598	0,000	20,294	0,021	13,835	0,015
Ma	l1	20,886	0,033	7,419	0,033	0,000	0,000	19,633	0,031	13,557	0,025
Ma	l2	23,608	0,029	8,420	0,029	0,000	0,000	19,486	0,021	13,284	0,021
Ma	no	23,110	0,029	6,560	0,022	1,144	0,000	18,935	0,027	14,375	0,020
Ma	s1	21,354	0,024	7,251	0,023	0,855	0,000	23,031	0,027	13,385	0,020
Ma	s4	20,159	0,020	6,209	0,013	0,569	0,000	22,879	0,028	15,911	0,028
Durchschnitt		19,316	0,031	6,068	0,028	0,786	0,010	20,250	0,029	14,579	0,024
Standardabweichung		1,994	0,015	0,914	0,015	0,687	0,020	1,641	0,007	1,347	0,007

### Italienisch

Aufnahme		%v	$\Delta v$	%a	$\Delta a$	%lL	$\Delta lL$	%l	$\Delta l$	%L	$\Delta L$
Fr	l1	49,916	0,059	1,147	0,127	5,678	0,034	5,678	0,034	0,000	0,000
Fr	l2	49,233	0,052	1,899	0,034	6,123	0,039	6,123	0,039	0,000	0,000
Fr	no	51,461	0,051	1,973	0,070	3,627	0,016	3,627	0,016	0,000	0,000
Fr	s1	51,422	0,050	2,641	0,053	4,708	0,018	4,708	0,018	0,000	0,000
Fr	s5	43,764	0,032	2,592	0,046	5,102	0,016	5,102	0,013	0,764	0,000
Va	l1	57,602	0,051	1,921	0,014	5,440	0,037	5,440	0,037	0,000	0,000
Va	l2	56,791	0,066	1,817	0,012	5,288	0,035	5,288	0,035	0,000	0,000
Va	no	49,851	0,037	2,075	0,018	5,338	0,029	5,338	0,029	0,000	0,000
Va	s1	53,203	0,041	1,862	0,026	4,450	0,020	4,450	0,020	0,000	0,000
Va	s5	50,256	0,020	3,728	0,022	4,121	0,011	4,121	0,011	0,000	0,000
Durchschnitt		51,350	0,046	2,165	0,042	4,987	0,025	4,987	0,025	0,076	0,000
Standardabweichung		3,939	0,014	0,689	0,035	0,758	0,011	0,758	0,011	0,242	0,000

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Fr	l1	20,280	0,046	8,729	0,037	1,577	0,085	13,158	0,035	19,795	0,038
Fr	l2	21,542	0,030	9,184	0,030	0,000	0,000	12,996	0,035	20,564	0,032
Fr	no	0,000	0,036	8,657	0,037	0,498	0,000	13,615	0,033	20,170	0,039
Fr	s1	0,000	0,028	8,334	0,028	0,000	0,000	14,118	0,035	18,778	0,026

Aufnahme		%nN	$\Delta nN$	%n	$\Delta n$	%N	$\Delta N$	%f	$\Delta f$	%s	$\Delta s$
Fr	s5	0,000	0,025	10,172	0,025	1,094	0,024	16,684	0,032	19,828	0,017
Va	l1	0,000	0,033	9,116	0,033	0,000	0,000	12,037	0,052	13,884	0,027
Va	l2	0,000	0,028	8,887	0,028	0,000	0,000	11,874	0,039	15,344	0,034
Va	no	0,000	0,019	10,099	0,019	0,000	0,000	12,395	0,031	20,242	0,031
Va	s1	0,000	0,024	9,596	0,024	0,000	0,000	12,883	0,028	18,006	0,024
Va	s5	0,000	0,016	9,721	0,013	0,420	0,000	13,943	0,025	17,811	0,016
Durchschnitt		19,393	0,029	9,249	0,027	0,359	0,011	13,370	0,035	18,442	0,028
Standardabweichung		1,758	0,009	0,626	0,008	0,560	0,027	1,387	0,007	2,250	0,008

## ***Ergebnisse der Diskriminanzanalyse:***

### **Eigenwerte:**

Durchgang 1 (eine Gruppe pro Sprache):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	104,611	75,0	75,0	0,995
2	20,729	14,9	89,9	0,977
3	14,058	10,1	100,0	0,966

Durchgang 2 (Deutsch vs. andere Sprachen):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	54,959	100,0	100,0	0,991

Durchgang 3 (Englisch vs. andere Sprachen):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	31,874	100,0	100,0	0,985

Durchgang 4 (Französisch vs. andere Sprachen):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	47,055	100,0	100,0	0,990

Durchgang 5 (Italienisch vs. andere Sprachen):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	28,132	100,0	100,0	0,983

Durchgang 6 (akzentzählende vs. silbenzählende Sprachen):

Funktion	Eigenwert	% der Varianz	kumulative %	kanonische Korrelation
1	86,560	100,0	100,0	0,994

**Strukturmatrizen:**

Durchgang 1 (eine Gruppe pro Sprache):

	Funktion						
	1	2	3				
<i>%l</i>	-0,253*	-0,065	-0,203	<i>ΔvalLN</i>	-0,094	-0,119*	0,082
<i>%n</i>	0,215*	-0,083	0,110	<i>ΔvalnNfs</i>	0,007	-0,098*	-0,029
<i>ΔNfs</i>	0,111*	0,055	-0,008	<i>ΔvalLnfs</i>	0,011	-0,096*	-0,044
<i>Δvalf</i>	-0,103*	-0,002	-0,103	<i>Δvalnfs</i>	-0,001	-0,095*	-0,035
<i>ΔvalnNfs</i>	0,097*	-0,009	0,030	<i>%L</i>	0,009	0,085*	-0,020
<i>ΔLns</i>	0,064*	0,006	-0,050	<i>ΔvLs</i>	0,031	-0,055*	-0,045
<i>ΔalLns</i>	0,061*	-0,007	-0,038	<i>ΔvalLNfs</i>	-0,056	-0,061	-0,167*
<i>Δafs</i>	0,047*	0,018	0,043	<i>%s</i>	0,082	-0,055	0,164*
<i>%f</i>	0,022	0,322*	0,003	<i>ΔvalLfs</i>	-0,097	-0,104	-0,162*
<i>ΔvalLnNs</i>	-0,019	-0,184*	-0,071	<i>%N</i>	0,142	0,131	-0,143*
<i>ΔvalLns</i>	-0,018	-0,184*	-0,061	<i>ΔvalLn</i>	-0,022	-0,111	0,143*
<i>ΔvalnNs</i>	-0,015	-0,184*	-0,066	<i>ΔvlNfs</i>	0,041	-0,110	-0,133*
<i>ΔvalLn</i>	-0,057	-0,166*	0,080	<i>ΔvalLNf</i>	-0,083	0,024	-0,109*
<i>ΔvalnN</i>	-0,058	-0,164*	0,100	<i>ΔvalLf</i>	-0,103	0,006	-0,106*
<i>ΔvalLnN</i>	-0,061	-0,163*	0,096	<i>ΔvalLnf</i>	-0,050	-0,023	-0,104*
<i>ΔvalLs</i>	-0,029	-0,149*	-0,064	<i>ΔvanNf</i>	-0,002	0,075	-0,077*
<i>ΔvalLNfs</i>	-0,011	-0,130*	-0,098	<i>ΔvalLnNf</i>	-0,036	-0,018	-0,073*
<i>ΔvalL</i>	-0,093	-0,126*	0,085	<i>ΔvlLnNfs</i>	0,032	-0,050	-0,072*
				<i>ΔvalnNf</i>	-0,034	-0,013	-0,070*
				<i>ΔalLnN</i>	0,014	-0,006	-0,032*

Ein Sternchen (\*) kennzeichnet die höchste absolute Korrelation jeder Variablen mit einer Diskriminanzfunktion.

Durchgang 2 (Deutsch vs. andere Sprachen):

	Funktion						
	1						
<i>%N</i>	0,204		<i>Δvalfs</i>	-0,066		<i>ΔvalLns</i>	-0,024
<i>ΔvalL</i>	-0,136		<i>ΔvalLf</i>	-0,065		<i>ΔvalLnNs</i>	-0,023
<i>ΔvalLN</i>	-0,136		<i>ΔvLNf</i>	0,065		<i>ΔvalLnNf</i>	-0,022
<i>%l</i>	-0,108		<i>ΔvalLfs</i>	-0,056		<i>ΔvalLNfs</i>	-0,022
<i>ΔvalLnN</i>	-0,104		<i>ΔvalNf</i>	-0,049		<i>ΔvalnNf</i>	-0,021
<i>%n</i>	0,103		<i>ΔvalLNf</i>	-0,048		<i>ΔvalnNs</i>	-0,020
<i>ΔvalLn</i>	-0,095		<i>ΔvlLnNfs</i>	0,046		<i>ΔvLN</i>	-0,016
<i>ΔvalnNfs</i>	0,085		<i>%f</i>	0,042		<i>ΔvalNf</i>	0,012
<i>ΔalLNf</i>	0,083		<i>ΔvalLs</i>	-0,031		<i>ΔvalLnfs</i>	0,009
<i>ΔvlLN</i>	-0,082		<i>%s</i>	0,030		<i>Δvalnfs</i>	-0,005
<i>ΔvalnN</i>	-0,076		<i>ΔvalNf</i>	-0,028		<i>ΔvalLNs</i>	-0,005
<i>ΔvalN</i>	-0,070		<i>ΔvalNfs</i>	-0,028		<i>ΔvlNf</i>	-0,003
			<i>ΔvalnNf</i>	0,027		<i>Δvans</i>	-0,002
			<i>%L</i>	0,026		<i>ΔvalnNfs</i>	0,001

Durchgang 3 (Englisch vs. andere Sprachen):

	Funktion						
	1						
<i>ΔvalNfs</i>	-0,115		<i>ΔvalLfs</i>	-0,096		<i>ΔvalLnf</i>	-0,073
<i>ΔvalLNfs</i>	-0,111		<i>%s</i>	0,093		<i>%l</i>	-0,071
			<i>Δvalfs</i>	-0,087		<i>ΔvalNf</i>	-0,068
			<i>ΔvalnN</i>	0,075		<i>ΔvalLNf</i>	-0,068

$\Delta valLNs$	-0,067
$\Delta valLf$	-0,064
$\Delta lNf$	0,059
$\Delta valLnNf$	-0,054
$\Delta valNf$	-0,051
$\Delta valnNf$	-0,051
$\Delta valLnNs$	-0,050
$\Delta valLs$	-0,049
$\Delta valnNs$	-0,048
$\%n$	0,046

$\Delta valLns$	-0,045
$\Delta valnNf$	-0,043
$\Delta vlLnNfs$	-0,041
$\Delta vlLN$	-0,036
$\Delta alnfs$	0,034
$\Delta valNfs$	-0,031
$\Delta valnfs$	-0,029
$\Delta valnNfs$	0,028
$\%N$	-0,028
$\Delta valLnN$	0,026

$\Delta vLns$	-0,025
$\Delta valnNfs$	-0,023
$\Delta alNfs$	0,022
$\Delta valLn$	0,019
$\%f$	0,018
$\Delta valL$	0,017
$\Delta valLN$	0,016
$\%L$	-0,004
$\Delta valnNs$	0,000

Durchgang 4 (Französisch vs. andere Sprachen):

	Funktion
	1
$\%n$	0,251
$\%l$	-0,204
$\%s$	0,148
$\Delta valLf$	-0,141
$\Delta valLNf$	-0,129
$\Delta valNf$	-0,125
$\Delta valnNfs$	0,118
$\Delta valLfs$	-0,098
$\Delta lNfs$	0,096
$\Delta valfs$	-0,091
$\Delta valLNfs$	-0,077
$\Delta valLnf$	-0,075

$\Delta valNfs$	-0,074
$\Delta aLNs$	0,073
$\%N$	0,063
$\Delta vas$	0,060
$\Delta valLnNf$	-0,054
$\Delta valnNf$	-0,053
$\Delta valLN$	-0,047
$\Delta vaLns$	0,047
$\Delta valL$	-0,045
$\Delta vLs$	0,041
$\Delta valnNf$	-0,035
$\Delta aLN$	0,031
$\%f$	-0,030
$\Delta valNfs$	0,026

$\Delta vanN$	0,026
$\Delta valnNfs$	0,025
$\Delta vlnN$	-0,020
$\%L$	-0,014
$\Delta valnfs$	0,014
$\Delta valnNs$	0,010
$\Delta valLs$	-0,010
$\Delta valLnN$	-0,010
$\Delta valLn$	-0,008
$\Delta valLns$	0,008
$\Delta valLnNs$	0,006
$\Delta valLNs$	-0,002

Durchgang 5 (Italienisch vs. andere Sprachen):

	Funktion
	1
$\%f$	0,268
$\Delta valLnNs$	-0,163
$\Delta valLns$	-0,163
$\Delta valnNs$	-0,163
$\Delta valLs$	-0,132
$\Delta valLNs$	-0,117
$\Delta valLn$	-0,116
$\Delta valLnN$	-0,109
$\Delta vaLns$	-0,097
$\Delta valfs$	-0,096
$\Delta al$	-0,095
$\Delta valnfs$	-0,084
$\Delta valnNfs$	-0,084
$\Delta valLnf$	-0,083

$\Delta valNfs$	-0,081
$\Delta valLfs$	-0,080
$\Delta valN$	-0,078
$\Delta valL$	-0,078
$\Delta valLN$	-0,074
$\Delta vnnNs$	-0,071
$\Delta vaLNf$	0,071
$\%L$	0,070
$\Delta vaLs$	-0,069
$\%N$	0,063
$\Delta valLnN$	-0,061
$\Delta valNfs$	-0,060
$\Delta valnNf$	0,048
$\Delta alLnNf$	0,040
$\%l$	-0,039
$\Delta valLnf$	-0,032

$\Delta vaLN$	-0,031
$\Delta alnfs$	0,026
$\Delta valLnNf$	-0,026
$\Delta vlLf$	-0,023
$\Delta vnf$	0,022
$\Delta valnNf$	-0,022
$\Delta aN$	-0,021
$\Delta valLf$	-0,014
$\%s$	-0,012
$\%n$	-0,012
$\Delta anN$	-0,008
$\Delta valnNfs$	0,006
$\Delta valNf$	-0,006
$\Delta valLNf$	-0,002

## Durchgang 6 (akzentzählende vs. silbenzählende Sprachen):

	Funktion
	1
%l	-0,217
$\Delta$ valfs	-0,142
$\Delta$ valLfs	-0,134
%n	0,128
$\Delta$ valLf	-0,105
$\Delta$ valNfs	-0,102
$\Delta$ valLNfs	-0,089
$\Delta$ vlLN	-0,089
$\Delta$ valNf	-0,087
$\Delta$ valLNf	-0,087
%s	0,085
$\Delta$ valnNfs	0,083
$\Delta$ vanNfs	0,082
%N	0,077
$\Delta$ vlNfs	0,076

$\Delta$ alnfs	0,075
$\Delta$ vlNfs	0,069
$\Delta$ valNf	-0,068
$\Delta$ valLN	-0,068
$\Delta$ valL	-0,067
$\Delta$ lf	0,065
$\Delta$ alNf	0,059
$\Delta$ valNfs	0,058
$\Delta$ vlN	-0,055
$\Delta$ valLs	-0,054
$\Delta$ valLnNf	-0,049
$\Delta$ valLn	-0,047
$\Delta$ valLnNs	-0,047
$\Delta$ valLnN	-0,046
$\Delta$ valnNf	-0,046
$\Delta$ valLns	-0,045
$\Delta$ valnNs	-0,043
$\Delta$ valLNs	-0,042

$\Delta$ vlLNs	-0,042
%f	0,041
$\Delta$ vlNf	0,029
$\Delta$ valNs	0,023
$\Delta$ valnfs	-0,020
$\Delta$ valNf	-0,020
%L	0,016
$\Delta$ valnNfs	-0,012
$\Delta$ valLns	-0,011
$\Delta$ vlLnNfs	0,007
$\Delta$ valnN	-0,007
$\Delta$ alN	0,006
$\Delta$ valnNf	-0,006
$\Delta$ vlNfs	-0,004
$\Delta$ valnNs	0,001
$\Delta$ vlNs	-0,001