

# Ein rhythmisch-prosodisches Modell lyrischen Sprechstils

Inaugural-Dissertation  
zur Erlangung der Doktorwürde  
der  
Philosophischen Fakultät  
der  
Rheinischen Friedrich-Wilhelms-Universität  
zu Bonn

vorgelegt von  
Jörg Bröggelwirth  
aus  
Paderborn in Westfalen

Bonn, 2007

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

**Zusammensetzung der Prüfungskommission:**

PD Dr. Bernhard Schröder (Vorsitzender)

Prof. Dr. Wolfgang Hess (Betreuer und Gutachter)

Prof. Dr. Winfried Lenders (Gutachter)

Prof. Dr. Jürgen Esser (Weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 8. 6. 2007

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn  
[http://hss.ulb.uni-bonn.de/diss\\_online](http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert

„freilich ist die Poesie nicht für das Auge bestimmt“

(Johann Wolfgang von Goethe)

Ich möchte mich bei den folgenden Personen bedanken:

Valeska Maus, Christian Aretz, Annegret Steudner, Barbara Samlowski und Arne Bachmann für ihre umfangreiche Arbeit am Lyrik-Korpus, den Sprechern und Hörern für ihre Geduld und Anstrengung, Dr. Rüdiger von Tiedemann für seine Hilfe bei der Textauswahl, Prof. Dr. Wolfgang Hess für die Betreuung, Dr. Petra Wagner für die stets aufschlussreichen Diskussionen und Anregungen und meinen Eltern Elisabeth und Erich für ihren Rückhalt!

# Inhaltsverzeichnis

1	Einleitung .....	1
2	Zum Sprechrhythmus .....	4
2.1	Isochronie .....	4
2.2	Akustische Korrelate des Sprechrhythmus .....	6
2.3	Perzeption von Sprechrhythmus .....	9
2.4	Sprechrhythmus und Lyrik .....	16
2.4.1	Motivation .....	17
2.4.2	Ergebnisse .....	18
2.5	Fazit .....	22
3	Zum Begriff der Prosodie .....	23
3.1	Definierbarkeit .....	23
3.2	Funktion .....	24
3.3	Fazit .....	25
4	Lyrik und Metrik .....	26
5	Korpuserstellung .....	29
5.1	Textauswahl .....	29
5.2	Aufnahme .....	30
5.3	Annotation .....	31
5.3.1	GToBI-Etikettierung .....	33
5.4	Zusammenfassung .....	35
6	Datenanalyse .....	36
6.1	Dauerphänomene .....	36
6.1.1	Silbendauern .....	37
6.1.2	Fußdauern .....	41
6.1.3	Die Dauerrelation unakzentuiert zu akzentuiert .....	49
6.1.4	Phrasen- und Versdauern .....	50
6.1.5	Zusammenfassung .....	51
6.2	Intonationsphänomene .....	52
6.2.1	Höhe der Grundfrequenz .....	52
6.2.2	Phonologische Beschreibung der Intonation .....	54
6.2.3	Zusammenfassung .....	59
7	Rhythmisch-prosodische Modelle .....	60
7.1	Dauermodelle .....	60
7.1.1	Lineare Regression .....	60
7.1.2	Silbendauern bei isochronen Fußdauern .....	66
7.2	Intonationsmodell .....	68
7.3	Zusammenfassung .....	70
8	Evaluation der Dauermodelle .....	72
8.1	Verfahren der perzeptiven Beurteilung von Sprachstimuli .....	72

8.1.1	Der Faktor Mensch .....	73
8.1.2	Testmethoden .....	74
8.1.3	Diskussion .....	76
8.2	Delexikalisierung .....	77
8.3	Modellevaluation .....	80
8.3.1	Ergebnisse der Lyrik/Prosa Unterscheidung bei delexikalisierten Stimuli .....	85
8.3.2	Ergebnisse der Lyrik/Prosa Unterscheidung bei nicht-delexikalisierten Stimuli .....	89
8.3.3	Ergebnisse der Rhythmuserkennung bei delexikalisierten Stimuli .....	93
8.3.4	Zusammenfassung .....	100
9	Abschließende Zusammenfassung und Ausblick .....	103
	Literatur .....	109
	Anhänge .....	115

# 1 Einleitung

Der sprachliche Rhythmus erfüllt Funktionen im Bereich Gliederung und Hervorhebung auf allen linguistischen Ebenen. Silben- und Wortgrenzen sowie Wortbetonungen werden durch ihn markiert, syntaktische Phrasen und semantisch zusammengehörige Einheiten werden rhythmisch gegliedert. Ohne diese rhythmische Gliederung wäre der Sprachwahrnehmungsprozess sicherlich um einiges erschwert. Neben der direkten Beziehung zwischen Rhythmus und linguistischen Einheiten existieren ferner situationsspezifische Rhythmen, welche mit einem bestimmten Sprechstil verknüpft sind. So hat schnelle Sprache vermutlich einen anderen Rhythmus als langsame, eine Predigt wiederum einen anderen Rhythmus als ein Fußballkommentar. Welche Rolle der Sprechrhythmus auf den einzelnen linguistischen und paralinguistischen Ebenen im Detail spielt, ist jedoch noch weitestgehend ungeklärt. Eine relativ große Übereinstimmung besteht in der Meinung, dass es auch sprachspezifische rhythmische Unterschiede gibt, genauer gesagt gibt es die Annahme, dass sich Sprachen in so genannte akzentzählende und silbenzählende Sprachen unterteilen.

Überdies ist der Sprechrhythmus für die Sprachsynthese, d.h. den sprechenden Computer, von Bedeutung. Die häufigste Anwendung in diesem Bereich ist die Überführung von Text in ein entsprechendes Sprachsignal (*Text-To-Speech*). Auch in den neueren korpusbasierten Synthesystemen ist eine ausgefeilte Prosodieprädiktion unerlässlich. Diese Vorhersage ist in den meisten Fällen jedoch noch verbesserungswürdig. Leider zielen TTS-Sprachsynthesysteme im Allgemeinen ausschließlich auf eine prosaische Textdomäne ab, so dass spezielle Domänen, wie etwa die Lyrik, außer Acht gelassen werden. Die vorliegende Arbeit wird unter anderem dadurch motiviert, dass die Grundlagen für eine Erweiterung der Textdomäne von TTS-Systemen um das Feld der Lyrik mit festgelegtem Metrum fehlen und somit erst noch geschaffen werden müssen. Der prominente Rhythmus von Lyrik mit festgelegtem Metrum soll dabei in zwei Dauermodellen abgebildet werden. Es wird angenommen, dass Rückschlüsse von der Rhythmizität der Lyrik im Deutschen auf die deutsche Sprache allgemein möglich sind.

Kapitel 2 gibt einen Überblick über den aktuellen Stand der Forschung im Bereich Sprechrhythmus mit besonderer Berücksichtigung von Rhythmus und Lyrik. Die Darlegungen umfassen etwa den Zeitraum der letzten 60 Jahre. Der größte Teil der

skizzierten Arbeiten befasst sich dabei mit der zeitlichen Struktur des Sprechrhythmus. Der Bericht zeigt jedoch auf, dass bezüglich Rhythmus auch intonatorische, ja sogar segmentale Phänomene eine Rolle spielen. Die an dieser Stelle gemachten Skizzierungen unterstreichen die Vielfältigkeit der verschiedenen Ansätze in der Sprechrhythmusforschung, sowohl auf akustischer als auch perzeptiver Ebene.

Das Kapitel 3 diskutiert kurz den Begriff der Prosodie, da Prosodie und Sprechrhythmus eng miteinander verknüpft sind. Vor allem der Parameter der Dauer ist hier als bedeutend anzuführen. Es werden die Definierbarkeit und die Funktion von Prosodie erläutert.

Im Kapitel 4 dieser Arbeit wird der literaturwissenschaftliche Begriff der Metrik in der Lyrik erörtert. Es wird insbesondere Bezug auf verschiedene Versmaße und rhythmische Prinzipien genommen. Zudem wird kurz der Aufbau und die Struktur eines Verses skizziert.

Das Kapitel 5 befasst sich mit der Erstellung eines Korpus deutscher Lyrik. Es wird vor allem diskutiert, welche Texte aus welchem Grund ausgewählt wurden. Darüber hinaus wird die Wahl der Sprecher näher erläutert. Schließlich wird das Thema der Korpusannotation behandelt. Der Etikettierformalismus GToBI für die Intonation des Deutschen wird hier besonders ausführlich dargestellt.

Im Kapitel 6 wird die Analyse des umfangreichen Korpus erörtert. Es werden vor allem Statistiken über die prosodischen Parameter von Intonation und Dauer dargelegt. Im Fokus der Analyse stehen vor allem Dauern auf den linguistischen Ebenen von Silbe, Fuß und Phrase bzw. Vers.

Das Kapitel 7 der vorliegenden Arbeit stellt die Modellbildung von Silbendauern im lyrischen Sprechstil vor. Es werden dabei die beiden unterschiedlichen Ansätze der linearen Regression und der Fußisochronie verfolgt. Ersteres Modell prädiziert dabei Silbendauern, die durch eine lineare Regression berechnet werden, letzteres Modell geht von einer konstanten Dauer eines Fußes aus und berechnet so die Dauern der im Fuß beinhalteten akzentuierten und nichtakzentuierten Silben. Darüber hinaus wird eine GToBI-Beschreibung der Intonation von Silben im Korpus angeführt.

Im Kapitel 8 werden schließlich die beiden verschiedenen Ansätze der Dauermodellierung bezüglich ihrer rhythmischen Diskriminationsleistung geprüft. In drei Perzeptionstests wird den Hörern eine Sprechstilunterscheidung sowie eine Unterscheidung verschiedener Metren abverlangt. Hierbei sind gezielt manipulierte Stimuli verwandt worden.

In Kapitel 9 werden die verschiedenen Analyse- und Testergebnisse abschließend zusammengefasst und diskutiert. Dabei wird auch versucht, ungelöste Probleme der vorliegenden Arbeit von verschiedenen Seiten zu beleuchten. Zudem wird ein Ausblick auf künftige Arbeiten gegeben.



## 2 Zum Sprechrhythmus

In diesem Kapitel wird der aktuelle Stand der Sprechrhythmusforschung näher erörtert. Die Darlegungen reichen von der Isochroniehypothese aus der Mitte des letzten Jahrhunderts bzw. der Unterscheidung zwischen akzentzählenden und silbenzählenden Sprachen (2.1), über akustische Korrelate des Sprechrhythmus (2.2) und die Perzeption des Sprechrhythmus (2.3) bis zum Zusammenhang von Sprechrhythmus und Lyrik (2.4). Der Diskussion der wichtigsten Literatur wird ein kurzes Fazit nachgestellt (2.5).

### 2.1 Isochronie

Das in der Sprechrhythmusforschung wohl am kontroversesten diskutierte Prinzip ist das der Isochronie. Pike (1945) propagiert erstmals die Unterscheidung von akzentzählenden und silbenzählenden Sprachen, wobei die akzentzählenden Sprachen das Interakzentintervall als isochrone Einheit, die silbenzählenden Sprachen hingegen die Silbe als isochrone Einheit besitzen sollen. Abercrombie (1967) schließt sich dieser Isochroniehypothese an und sieht jede Sprache der Welt einer dieser beiden Rhythmusklassen zugehörig. Seiner Ansicht nach ist dieser Unterschied physiologisch in der Existenz zweier verschiedener muskulärer Aktivitäten begründet, nämlich der für die Periodizität der Silben verantwortlichen *chest-pulses* bzw. der für die Periodizität der Akzente verantwortlichen *stress-pulses*. Diese Begründung wurde jedoch von Ladefoged (1967) anhand elektromyographischer Messungen an den respiratorischen Muskeln widerlegt.

In der Folge ist in zahlreichen Studien, vor allem für das Englische, versucht worden, die Existenz isochroner Einheiten akustisch nachzuweisen (Bolinger, 1965; Allen, 1972; Lea, 1974; Lehiste, 1977; Uldall, 1978; Dauer, 1983). Für die gemessenen Interakzentintervalle<sup>1</sup>, d.h. das Zeitintervall zwischen zwei Akzenten, konnte zumindest keine strenge Isochronie ermittelt werden. Es zeigte sich vielmehr, dass ein linearer Zusammenhang zwischen der Dauer von Interakzentintervallen und der Anzahl der Silben im Intervall besteht. Dauer (1983) stellt zum Interakzentintervall, das hier von Vokalonset zu Vokalonset gemessen wird, des Weiteren fest, dass für das Englische keine stärkere Tendenz zur Produktion von

---

<sup>1</sup> Als Übersetzung des englischen Wortes *stress* wird an dieser Stelle der Begriff Akzent statt Betonung verwendet. Es wird damit der begrifflichen Übersetzung von Bußmann (1983) Folge geleistet. Sie betrachtet den englischen Terminus *stress* als dynamischen Akzent.

Interakzentintervallen mit konstanter Dauer besteht als für andere Sprachen. Letztere Vermutung lag nahe, da das Englische als akzentzählende Sprache gilt. Dauer zieht damit den Schluss:

„Consequently, we can conclude that the difference between English, a stress-timed language, and Spanish, a syllable-timed language, has nothing to do with the durations of interstress intervals.“ (Dauer, 1983: 54)

Darüber hinaus beobachtet Dauer, dass es eine reguläre Rate von Interakzentintervallen pro Sekunde, nicht aber eine reguläre Rate an Silben pro Sekunde gibt. Zudem vermutet sie, dass es im Planungsprozess der Sprachproduktion eine Einheit mindestens von der Größe eines Interakzentintervalls geben muss. Über das Wesen des Interakzentintervalls stellt die Autorin ferner fest:

„I would like to propose that the rhythmic differences we feel to exist between languages such as English and Spanish are more a result of phonological, phonetic, lexical, and syntactic facts about that language than any attempt on the part of the speaker to equalize interstress or intersyllable intervals.“ (Dauer, 1983: 55)

Überdies sollen die akzentzählenden Sprachen aufgrund ihrer größeren Zahl möglicher Silbentypen stärker bezüglich ihrer Silbenlänge variieren als die silbenzählenden Sprachen. Außerdem seien in akzentzählenden Sprachen schwere Silben häufiger akzentuiert als leichte Silben. Darüber hinaus maximiere die Zentralisierung unakzentuierter Silben den Unterschied zwischen akzentuierten und unakzentuierten Silben in akzentzählenden Sprachen weiter. Zudem bewahren akzentzählende Sprachen trotz Reduktion die Silbizität. In silbenzählenden Sprachen dagegen können Silben miteinander verschmelzen. Silbenzählende Sprachen haben im Normalfall keine reduzierten Vokale in unakzentuierter Position. Dennoch hält Dauer es für möglich, dass alle Sprachen mit identifizierbarem Akzent ein gewisses Maß an Regularität bezüglich der Dauer von Interakzentintervallen zeigen werden. Auf jeden Fall aber macht Dauer für die akzentzählenden Sprachen folgendes geltend:

„In addition, the greater the effects of stress on the linguistic system, in both its phonetic realization and in the number of stress-related rules that are necessary in a grammar of the language, the more likely it is to be stress-timed.“ (Dauer, 1983: 58)

Für das Deutsche hat Kohler (1982; 1983) die akustische Isochronie untersucht. Er hält es nämlich für möglich, dass die Identifikation der Isochronie im Englischen aus methodischen

Gründen, z.B. durch das Silbengrenzproblem, gescheitert ist. Der Autor hat jedoch nicht die bloße Existenz der Isochronie in Sprachdaten überprüft, sondern ein Experiment mit Nonsensstimuli durchgeführt, in welchem das Potential zur Isochronie gemessen wurde. Es sollte demnach die Kompressionsfähigkeit von Silben getestet werden. Hierbei sollten Sprecher innerhalb eines vorgegebenen Zeittaktes einen Akzentfuß (1 akzentuierte Silbe plus 0-2 unakzentuierte Silben) realisieren. Die geforderte Kompression der im Fuß enthaltenen Silben gelang bis auf die Ausnahme von Dreisilbern mit komplexer Silbenstruktur. Ein weiterer interessanter Effekt war, dass die Mehrsilber im gleichen Zeitrahmen artikuliert schneller gesprochen erschienen als die Einsilber. Gleichzeitig erschienen unkomprimierte Dreisilber jedoch langsamer als Einsilber. Daraus folgt, dass innerhalb der beiden Extreme kompletter Isochronie einerseits und zur Silbenzahl proportionaler Expansion andererseits komprimiert werden muss, um das Sprechtempo konstant zu halten. Außerdem konnte Kohler zeigen, dass die Dauer eines Akzentfußes nicht nur linear von seiner eigenen segmentalen Komplexität, sondern zusätzlich von der Komplexität des vorangehenden Akzentfußes abhängt. Damit hat das Deutsche eine Tendenz zur Isochronie, die durch artikulatorische Verschleifungen, die Bildung schwacher Formen und Dauerkompression erreicht wird. Aus diesen Ergebnissen leitet Kohler eine zeitliche Organisationshierarchie der Sprachproduktion ab. In einem ersten Planungsschritt werden demnach Fußintervalle sprechratenabhängig absolut isochron festgesetzt. Im zweiten Schritt erfahren die Dauern der Fußintervalle eine Modifikation durch Informationsstrukturierung und Fokussierung. Die dritte Stufe der zeitlichen Sprechorganisation expandiert die Fußdauern gemäß ihrer Silbenzahl und Silbenkomplexität. Auf der vierten Stufe schließlich vollzieht sich die temporale Strukturierung der Silben innerhalb eines Fußes bzw. der Dyaden, welche auch Diphone genannt werden, und Phone innerhalb einer Silbe. Die geschilderten Beobachtungen Kohlers beziehen sich allesamt auf zeitliche Phänomene. Er mutmaßt aber, dass auch die Grundfrequenzvariation für die temporale Strukturierung der Sprache maßgebend ist. An dieser Stelle besteht noch Klärungsbedarf.

## **2.2 Akustische Korrelate des Sprechrhythmus**

Nachdem sich die (strenge) Isochronie von Akzentfüßen bzw. Silben nicht als akustisches Korrelat der Rhythmusklassenunterscheidung erwiesen hat (s. 2.1), ist die rhythmische

Klassifikation von Sprachen als akzentzählend bzw. silbenzählend dennoch nicht verworfen worden. Es ist vielmehr versucht worden, diesbezüglich alternative akustische Korrelate auszumachen. So kritisieren Ramus, Nespör & Mehler (1999):

„Although researchers have measured many speech signal properties, they have failed to identify reliable acoustic characteristics for language classes.“ (Ramus, Nespör & Mehler: 265)

Ramus, Nespör & Mehler (1999) stellen so die Maße %V,  $\Delta C$  und  $\Delta V$  als akustische Korrelate der Rhythmusklassen vor. %V misst dabei den prozentualen Anteil vokalischer Intervalldauern an der gesamten Äußerungsdauer,  $\Delta C$  ist die Standardabweichung der konsonantischen Intervalldauern und  $\Delta V$  die Standardabweichung der vokalischen Intervalldauern. Trägt man die Messpunkte für einzelne Sprachen in der %V/ $\Delta C$  Ebene auf, bilden die Sprachen Cluster gemäß ihrer klassischen Rhythmusklassenzugehörigkeit. Hierbei nimmt das konsonantische Maß Bezug auf die Silbenkomplexität einer Sprache, das vokalische Maß hingegen bezieht sich auf die Vokalreduktion (Ramus, 2002). Außerdem korreliert nach Ramus, Nespör & Mehler die Rhythmusart mit der Segmentiereinheit einer Sprache, nämlich dem Fuß bei akzentzählenden Sprachen und der Silbe bei silbenzählenden Sprachen. Die Autoren führen zudem das dritte Segment der Mora für morazählende Sprachen an. Die Mora ist eine virtuelle Einheit, welche die Dauern verschiedener Phone zusammenfasst. Es handelt sich hierbei jedoch um eine Randerscheinung, welche an dieser Stelle nicht weiter vertieft werden soll. Duarte et al. (2001) und Wagner & Dellwo (2004) bestätigen die Leistungsfähigkeit des %V/ $\Delta C$  Maßes zur Rhythmusklassenunterscheidung, kritisieren aber, dass darin nicht alle rhythmischen Eigenschaften der Sprache enthalten sind. Diese Skepsis teilt auch Cummins indem er äußert:

“Of all prosodic variables used to classify languages, rhythm has proved most problematic. Recent attempts to classify languages based on the relative proportion of vowels or obstruents have had some success, but these seem only indirectly related to perceived rhythm.” (Cummins, 2002: 121)

Fant, Kruckenberg & Nord (1991) untersuchten die akzentinduzierte Längung von Segmenten im Englischen, Schwedischen und Französischen. Es stellte sich dabei heraus, dass die Längung im als silbenzählend geltenden Französisch deutlich geringer ist als im als akzentzählend geltenden Englisch und Schwedisch. Zudem sind unakzentuierte Silben in

allen drei Sprachen segmental etwa gleich komplex, die akzentuierten Silben dagegen sind im Englischen und Schwedischen komplexer als im Französischen. Für den akzentzählenden Charakter einer Sprache leiten Fant, Kruckenberg & Nord daraus eine Verbindung der segmentalen und suprasegmentalen Ebene ab, indem Silbenkomplexität und akzentinduzierte Längung einen Beitrag zur relativen Schwere einer Silbe leisten. Der akzentzählende Eindruck einer Sprache ergibt sich demzufolge aus der quasiperiodischen Abfolge schwerer Silben. Ein silbenzählender Eindruck entsteht nach Fant, Kruckenberg & Nord durch die Reduktion der Differenz zwischen akzentuierten und unakzentuierten Silben. Die Bedeutung der Kombination von segmentaler und suprasegmentaler Ebene für die Unterscheidbarkeit von Rhythmusklassen zeigen auch Pellegrino et al. (2002). Sie erzielten anhand automatischer Klassifikationsverfahren gute Sprachdiskriminationsergebnisse unter Verwendung sowohl segmentaler Eigenschaften als auch Energie- und Grundfrequenzwerten als Eingabeparameter.

Obwohl keine konstanten Dauern für Interakzentintervalle in der Sprache nachgewiesen werden konnten, wurde weiter nach Regelmäßigkeiten bezüglich der zeitlichen Organisation von Sprachsegmenten geforscht. So ermittelten Fant & Kruckenberg (1996) eine mittlere Dauer von 0,5 s für Akzentfüße im Schwedischen. Dieses Intervall stellte sich dabei gleichzeitig als Referenzquantum für Pausendauern heraus, d. h. Pausen besitzen als Dauern ganzzahlige Vielfache des Referenzquantums. Außerdem zeigte sich, dass bezogen auf mittlere Dauern eine 1, 1/2, 1/4 bzw. 1/8 Relation in der zeitlichen Ausdehnung von Interakzentintervallen, akzentuierten Silben, unakzentuierten Silben und Phonen besteht. Analog zu dieser Beobachtung propagieren Cummins & Port (1998) die Betrachtung des Sprechrhythmus als hierarchische Organisation zeitlich koordinierter prosodischer Einheiten. Danach manifestiert sich der Rhythmus als zeitliche Bindung von Ereignissen an spezifische und präzifizierbare Phasen eines übergeordneten Zyklus. Zu diesem Phänomen stellten Cummins & Port eine Untersuchung auf Basis eines *speech cycling task* an. Bei diesem Experiment wurden Sprecher gebeten, eine kurze zweiakzentige Phrase in Korrespondenz mit einem auditiven Stimulus fortlaufend zu wiederholen. Es wurden dazu zwei Töne kurzer Dauer mit unterschiedlicher Frequenz dargeboten. Das Zeitintervall zwischen hoher und tiefer Frequenz war konstant, das Intervall zwischen tiefer und folgender hoher Frequenz dagegen wurde von Phrase zu Phrase variiert, so dass der tiefe Ton den Zyklus zu verschiedenen Phasen teilte. Die Sprecher wurden instruiert, den ersten Akzent der zu

artikulierenden Phrase mit dem hohen Ton, den zweiten Akzent hingegen mit dem tiefen Ton zu synchronisieren. Es stellte sich heraus, dass die Sprecher nicht frei darin waren, den zweiten Akzent zu platzieren. Sie bevorzugten vielmehr bestimmte Zielphasen innerhalb des Wiederholungszyklus, an denen sich die zu synchronisierenden Akzente kumulierten. Es bildeten sich dabei drei Attraktoren auf der Zeitachse zu Zyklusphasen von  $1/3$ ,  $1/2$  und  $2/3$  heraus. Dieses Phänomen nennen die Autoren *harmonic timing effect*. Damit sind die Einheiten des Fußes und des Phrasenzyklus als zwei miteinander gekoppelte Oszillatoren (vgl. O'Dell & Nieminen, 1999; Barbosa, 2002) zu verstehen. Sprache unterliegt folglich den allgemeinen Gesetzmäßigkeiten zyklisch bzw. rhythmisch agierender motorischer oder kognitiver Systeme (Port, Tajima & Cummins, 1999). Es besteht hierbei die Möglichkeit, dass die Oszillatoren demselben physischen System angehören und so die Kopplung dem Prinzip des *self-entrainment* folgt. Beim *self-entrainment* besteht kein rhythmischer Impulsgeber außerhalb des oszillierenden Systems. Nach Port, Tajima & Cummins liegt diese sprecherinhärente Kopplung von Oszillatoren in der Sprache unter den linguistischen Hierarchieebenen von Silbe, Akzentfuß und prosodischer Phrase vor. Zusammenfassend lässt sich zu den akustischen Korrelaten des Sprechrhythmus allerdings kritisch äußern:

„However, although many characteristics of the speech signal have been measured, reliable acoustic characteristics of language classes have not been identified.“ (Ramus, Nespor & Mehler, 1999: 287)

## 2.3 Perzeption von Sprechrhythmus

Zum besseren Verständnis der psychoakustischen Qualität von Rhythmus sollen in diesem Abschnitt zunächst einige grundlegende Erkenntnisse bezüglich der Perzeption von Rhythmus dargelegt werden. Allen (1975) berichtet, dass lange Zeitintervalle zwischen zwei akustischen Stimuli perzeptiv unterschätzt, kurze Zeitintervalle dagegen überschätzt werden, was bedeutet, dass ein Hörer stets versucht, wahrgenommenen Schallen Rhythmizität aufzuerlegen, selbst wenn sie physikalisch nicht vorhanden ist. Dieses Phänomen schließlich veranlasste einige Forscher, das Isochroniekonzept (s. 2.1) als perzeptives Phänomen zu betrachten (Cooper & Meyer, 1960; Lehiste, 1977; Couper-Kuhlen, 1993;). Auer & Uhmman (1988) äußern dazu:

„Isochronie – und Rhythmus überhaupt – sind in erster Linie nicht im physikalischen Sprachsignal zu untersuchen, sondern in der Perzeption und Interpretation des Sprachsignals durch den Sprachbenutzer.“ (Auer & Uhmann, 1988: 254)

Die subjektive Wahrnehmung von Isochronie muss hiernach auf einem abstrakten Konstrukt beruhen, ähnlich der Beziehung zwischen zugrundeliegenden Schlägen und Oberflächenrhythmus in der Musik (Cooper & Meyer, 1960). Ferner nehmen Hörer äquidistante nichtsprachliche akustische Stimuli als rhythmische Zweiergruppen wahr, wenn jeder zweite Stimulus bezüglich Lautstärke, Tonhöhe oder Dauer vom benachbarten Stimulus abweicht (Allen, 1975). Ist dabei jeder zweite Stimulus lauter, so werden Trochäen perzipiert. Bei Längung jedes zweiten Stimulus werden Jamben wahrgenommen. Für die Variation der Tonhöhe gilt der gleiche Effekt wie für die Lautstärke. Darüber hinaus gilt, dass sich bei Veränderung nur jedes dritten Stimulus rhythmische Dreiergruppen herausbilden. Wie Oberfeld (2000) zeigt, ist aber nicht nur die akustische Gestalt eines Stimulus für dessen Akzentuierungswahrnehmung verantwortlich, sondern unabhängig davon die Dauer des Inter-Onset-Intervalls zweier Stimuli. Ein Stimulus vor einem langen Intervall wird eher als akzentuiert wahrgenommen als vor einem kurzen Intervall.

In Bezug auf die Erforschung der Perzeption von Rhythmus gesprochener Sprache müssen zwei Untersuchungsfelder unterschieden werden. Der größte Teil der Sprechrhythmusforschung widmet sich der Diskrimination von verschiedenen Sprachen gemäß der Klassifikation akzentzählend/silbenzählend. Die zweite Forschungsrichtung dagegen betrachtet den Ereignischarakter des Sprechrhythmus. Im Bereich der sprachtypologischen Unterscheidung, so zeigen Ramus, Nespors & Mehler (1999), können Neugeborene zwischen Äußerungen aus der eigenen Muttersprache und Äußerungen einer anderen Rhythmusklasse unterscheiden. Die akzentzählend/silbenzählend Dichotomie scheint demnach im menschlichen Perzeptionssystem verankert zu sein. Überdies hat Miller (1984) die Sprachen Arabisch, Polnisch, Spanisch, Finnisch, Japanisch, Indonesisch und Yoruba untersucht. Die Klassifikation der Sprachen nach den Rhythmusklassen akzentzählend/silbenzählend erfolgte von den am Perzeptionstest beteiligten englischen und französischen Phonetikern und Nichtphonetikern sehr uneinheitlich. Miller folgert, dass jede Sprache Merkmale beider Rhythmusklassen zu verschiedenen Anteilen besitzen muss. Probleme bei der Sprachenunterscheidung anhand natürlicher Sprachstimuli konnten auch bei Versuchen mit Neugeborenen beobachtet werden (Ramus, 2000). Da sich manipulierte

Stimuli dabei als geeigneter zur Diskrimination herausstellten, untersuchten Ramus et al. (2000) das Sprachunterscheidungsverhalten erwachsener Probanden anhand von Stimuli, deren konsonantische Anteile sämtlich auf /s/ und deren vokalische Anteile sämtlich auf /a/ abgebildet wurden (*sasasa-Verfahren*). Diese Stimuli ermöglichten den Versuchspersonen eine bessere Unterscheidung verschiedener Sprachen insbesondere dann, wenn zusätzlich eine flache Intonation verwendet wurde. Ramus et al. schließen daraus, dass Hörer das Sprachsignal in konsonantische und vokalische Intervalle zerlegen und die Statistiken %V,  $\Delta C$  und  $\Delta V$  berechnen, um verschiedene Rhythmusklassen zu unterscheiden. Auch wenn damit geklärt ist, welche Sprachsignaleigenschaften primär für die perzeptive Unterscheidung verschiedener Sprachen sind, bleibt doch die Frage offen, ob diese Diskrimination, wie in Kapitel 2.2 bereits erörtert, wirklich auf der Basis als rhythmisch zu bezeichnender Eigenschaften erfolgt.

Bezüglich des Ereignischarakters gesprochener Sprache stellt sich die Frage, wo auf der Zeitachse des Sprachsignals die rhythmischen Schläge dem Hörerempfinden nach liegen und welche akustischen Parameter ihre Wahrnehmung bestimmen. Allen (1972) ermittelte durch Mittast- und Klickplatzierungsexperimente bei Stimuli des Englischen den Vokalonset einer Silbe als Ort des rhythmischen Schlages, wobei diese Position bei Zunahme der Zahl der silbeninitialen Konsonanten leicht nach vorne rückte. Die Versuchspersonen reagierten dabei so, als ob der Silbenschlag kein Punkt, sondern ein breites Intervall von etwa 200 ms wäre. Allen vermutet, dass die ballistischen Impulse akzentuierter Silben wahrscheinlich die grundlegenden Ereignisse sind, um die herum ein Sprecher den Sprechrhythmus organisiert. Des Weiteren wurde beobachtet, dass die Schlagposition mit größerer Übereinstimmung gesetzt wurde, je stärker die Akzentuierung der betreffenden Silbe war. Diesen Effekt sieht Allen als wesentlichstes Kennzeichen des englischen Akzentrhythmus an. Akzentuierte Silben sind somit die Schläge des Englischen. Mit den Ergebnissen, so vermutet der Autor, können Veränderungen der Länge rhythmischer Intervalle als Funktion der Veränderung der lexikalischen, phonetischen oder prosodischen Struktur innerhalb und außerhalb der Intervalle gemessen werden. Der Akzentuierung kommt hier die duale Rolle der syntaktischen Disambiguierung (Kompetenz) und Rhythmusbildung (Performanz) zu, wobei beide Bereiche sich überlagern. Den Begriff des Rhythmus schließlich definiert der Autor kurz als „Rhythm is the structure of intervals in a succession of events.“ (Allen, 1972: 72).



Morton, Marcus und Frankish (1976) stellten fest, dass onset-isochrone Sprachstimuli nicht als isochron wahrgenommen werden, was sie zu der Annahme führte, dass es einen Zeitpunkt des psychologischen Auftretens eines Sprachstimulus geben muss, der vom physikalischen Onsetzeitpunkt abweicht. Für diesen Ort im Signal prägten sie den Begriff des *P-center*. Es zeigte sich, dass das *P-center* im Wesentlichen mit der von Allen (1972) ermittelten Schlagposition im Konsonant-Vokal Übergang einer Silbe korrespondiert. In der Folgezeit sind Versuche unternommen worden, das *P-center* anhand verschiedener Modelle zu präzisieren. Marcus (1981) macht die *P-center* Position von der Dauer der initialen Konsonanz und des Silbenreims abhängig, Howell (1984; 1988) sieht die Amplitudenverteilung innerhalb einer Silbe als bestimmenden Faktor an (*Center of Gravity* Modell) und Pompino-Marschall (1989) versucht das *P-center* anhand eines auf der Berechnung spezifischer Lautheiten basierenden psychoakustischen Modells vorherzusagen. Janker (1995) hat diese drei Prädiktionsmodelle anhand der Ergebnisse seiner selbst ermittelten Mittast- und Adjustierungsergebnisse bezüglich ihrer Vorhersageleistung evaluiert. Die beste Prädiktionsleistung erbrachte dabei das *Center of Gravity* Modell von Howell. Dennoch konnte keines der drei Modelle die *P-center* Position hinreichend erklären. Daraufhin untersuchte Janker die prosodischen Eigenschaften seiner Daten an der *P-center* Position mit dem Schluss, dass Veränderung im Allgemeinen ereignisbildende Funktion hat, sowohl was den Amplituden- als auch den Grundfrequenzverlauf betrifft. Den Begriff des Ereignisses definiert Janker wie folgt:

„Ein Vorgang, der von einem, wie auch immer gearteten, wahrnehmenden System registriert und von diesem durch Gewichtung erkennbarer Eigenschaften in benennbarer, systematischer Weise einer Kategorie oder Klasse zugeordnet wird. Das Ereignis ist damit an die diskriminativen Fähigkeiten des Systems gebunden.“ (Janker, 1995: 16)

Überdies definiert der Autor hieran anknüpfend ebenfalls den Begriff des Rhythmus, indem er schreibt:

„Rhythmus konstituiert sich als vom wahrnehmenden System erkannte Struktur einer Folge von Ereignissen. Er ist damit an die für Ereignisse geltenden Einschränkungen gebunden und die Abfolge der Ereignisse muss eine dem wahrnehmenden System vermittelbare Struktur aufweisen. Das heißt, die einzelnen Ereignisse müssen hinreichend nahe beieinander liegen, damit eine Gruppierung stattfinden kann, und das zugrundeliegende Abfolgemuster darf nicht zu komplex sein.“ (Janker, 1995: 16)

Janker merkt an dieser Stelle kritisch an, dass ein Notationsmodell für den Sprechrhythmus bislang noch fehlt.

Dass Rhythmus perzeptiv jedoch kein von der segmentalen Ebene unabhängiger Parameter ist, zeigt Martin (1979). Er manipulierte die Dauern einzelner Segmente von Nonsenssätzen in der Art, dass eine Tempoänderung und damit auch eine rhythmische Änderung der gesamten Äußerung wahrgenommen werden konnte. Bei Hörtests zeigte sich, dass Probanden eine gegenüber den nicht manipulierten Stimuli verlängerte Reaktionszeit auf ein vorher definiertes Zielsegment hatten. Dieser Effekt erstreckte sich über eine Spanne von bis zu vier Silben zwischen manipuliertem Segment und Zielsegment. Ein Hörer erwartet demnach akustisch intaktes Sprachmaterial, wobei er seine Hörerwartung schon bei Wahrnehmung der ersten Silbe einer Äußerung aufbaut. Eine Trennung von prosodischer und segmentaler Ebene findet dabei nicht statt<sup>2</sup>.

Bei einer tempobezogenen Betrachtung des Sprechrhythmus hängt dessen Wahrnehmung im Wesentlichen von der Sprechrate ab. Es stellt sich dabei die Frage, wie die Sprechrate definiert werden soll. Nach Pfitzinger (1996) läßt sich die globale Sprechrate (Segmentzahl pro Zeit bezogen auf die gesamte Äußerung) von der lokalen Sprechrate (Segmentzahl pro Zeit bezogen auf ein Fenster innerhalb der Äußerung) abgrenzen. Als geeignete Segmente zur Sprechratenbestimmung werden im Allgemeinen Silben bzw. Phone herangezogen. Pfitzinger (1998; 1999) zeigt, dass die lokale Sprechrate (bezogen auf Fenster mit 625 ms Breite) weder ausschließlich auf Basis von Phonem noch ausschließlich auf Basis von Silben bestimmt werden kann, sondern dass Silbenrate und Phonrate trotz ihrer hohen Korrelation untereinander einen voneinander unabhängigen Beitrag zur Sprechrate leisten. Hierzu maß Pfitzinger zunächst die wahrgenommene lokale Sprechrate, um sie ins Verhältnis zur Silben- und Phonrate zu setzen. Die höchste Korrelation mit der wahrgenommenen Sprechrate erzielte dabei eine Linearkombination aus Phon- und Silbenrate. Zudem erwies sich die mittlere Grundfrequenz des jeweiligen Stimulus als zusätzlicher Faktor, um die Prädiktion der wahrgenommenen Sprechrate zu verbessern. Die Tempobezogenheit des Sprechrhythmus halten Duarte et al. (2001) allerdings für vernachlässigbar. In ihrer Studie zeigte sich, dass die Sprechrate, zumindest im untersuchten Sprachmaterial, kein wichtiger Faktor für den

---

<sup>2</sup> Martin nimmt eine begriffliche Trennung von segmentaler und rhythmischer (hier also prosodischer) Ebene vor.

Rhythmus von Sprachen ist. Bezogen auf das Maß  $\%V/\Delta C$  zeigte sich jedoch ein Einfluß der Sprechrate auf die Dimension  $\Delta C$  (Dellwo & Wagner, 2003). Die Dimension  $\%V$  dagegen bleibt weitestgehend unbeeinflusst von der Sprechrate.

Benkwitz (2004) hat eine kontrastive phonetische Untersuchung zum Rhythmus mit Britischem Englisch als Ausgangssprache und Deutsch als Zielsprache durchgeführt. Sie geht dabei von der folgenden Rhythmusdefinition aus:

„Daher verstehe ich in dieser Arbeit unter Rhythmus die Schwer-Leicht-Struktur, die bei der Produktion sowie der Rezeption des Sprechflusses durch die Abfolge miteinander kontrastierender Silben einschließlich Pausen entsteht. [...] Rhythmuskonstitutiv sind also die Position und Art der Realisierung *aller* Silben, unabhängig von zeitlich regelmäßiger Platzierung. Realisation und Position der Akzente werden von *sprach-* sowie *textsortenspezifischen, situativen und individuellen* Gegebenheiten bestimmt.“ (Benkwitz, 2004: 42)

Ferner geht Benkwitz davon aus, dass Rhythmus für einen Hörer dann entsteht, wenn er einen Kontrast bzw. eine Veränderung des Abstandes zwischen den Silben des Sprechflusses wahrnimmt und ihn dadurch als gegliedert empfindet. Zudem stellt die Autorin fest, dass die Akzentwahrnehmung nicht rein phonetisch ist, sondern zusätzlich grammatisch-semantisch determiniert ist. Außerdem unterscheiden sich Länge und Häufigkeit von Pausen zwischen deutschen und englischen Sprechern beim Lesen von Prosa nur geringfügig bzw. gar nicht. Darüber hinaus sind die Ketten unbetonter Silben im Englischen kürzer als im Deutschen. Des Weiteren ist die Akzentisochronie textsortenabhängig: die Regelmäßigkeit der zeitlichen Verteilung der Akzente ist bei Spontansprache am geringsten, bei den untersuchten Limericks (metrisch gebundene Rede) am größten und liegt bei den ausgewerteten Prosatexten (metrisch ungebundene Rede) im mittleren Bereich. Benkwitz betont dabei, dass Rhythmus nicht rein temporal zu verstehen ist, sondern ein Komplexphänomen ist, zu dem auch melodische und dynamische Erscheinungen gehören. Für ihre eigenen empirischen Untersuchungen geht Benkwitz (2004) davon aus, dass beim Rhythmus auditive Analysen messphonetischen vorzuziehen sind. Somit untersuchte Benkwitz die Akzenthäufigkeit, die zeitliche Verteilung der Akzente, die Beteiligung der Lautstärkevariation, Tonhöhenvariation und Dehnung an der Akzentuierung. Diese Phänomene wurden dabei auf einer 6-stufigen Skala von sehr schwach (1) bis sehr stark (6) perzeptiv beurteilt. Eine erste Versuchsreihe zum Vergleich der Rhythmisierung deutscher und englischer Muttersprachler prüfte dabei unter Zuhilfenahme statistischer Signifikanztests fünf Hypothesen:

- Hypothese 1/I: Die englischen Muttersprachler setzen generell mehr Akzente als die deutschen Sprecher.
- Hypothese 2/I: Die englischen Sprecher setzen Akzente in perzeptuell gleichmäßigeren zeitlichen Abständen als die deutschen Muttersprachler.
- Hypothese 3/I: Lautstärkevariationen spielen in beiden Sprachen eine gleich große Rolle für die Akzentwahrnehmung.
- Hypothese 4/I: Tonhöhenvariationen tragen in beiden Sprachen in gleichem Maße zur Wahrnehmung der Akzente bei.
- Hypothese 5/I: Die englischen Probanden dehnen Akzentsilben generell mehr als die deutschen.

Die Hypothesen 1/I und 5/I konnten durch die Untersuchungen in vollem Umfang bestätigt werden. Die Hypothese 4/I ist lediglich für Prosatexte (mit Einschränkungen) und Spontansprache bestätigt worden. Die Hypothesen 2/I und 3/I schließlich sind durch die Tests abgelehnt worden. Des Weiteren ist eine zweite Versuchsreihe zum Vergleich der Rhythmisierung bei deutschen Muttersprachlern und britischen Deutsch Lernern durchgeführt worden. Dabei sind die folgenden sechs Hypothesen getestet worden:

- Hypothese 1/II: Die englischen Deutsch Lernenden bilden beim Lesen von Prosatexten in deutscher Sprache im Durchschnitt kleinere Rhythmusgruppen als die deutschen Sprechern.
- Hypothese 2/II: Die englischen Studenten setzen beim Lesen von Prosatexten mehr Akzente als die deutschen Muttersprachler. Beim Lesen von Prosatexten und Limericks sowie bei spontan produzierten Äußerungen wird die Akzentzahl von den deutschen Hörern als zu hoch empfunden.
- Hypothese 3/II: Der Lautstärkekontrast zwischen betonten und unbetonten Silben wird beim Lesen von Prosatexten von den englischen Deutsch Lernenden stärker realisiert als von den deutschen Probanden und bei allen drei Textsorten von den deutschen Hörern als zu groß beurteilt.

- Hypothese 4a/II: Bei den Studierenden sind Tonhöhenvariationen bei gelesenen Texten (Prosatexte und Limericks) in qualitativer Hinsicht zu stark an der Akzentuierung beteiligt.
- Hypothese 4b/II: Bei spontan produzierten Äußerungen hingegen sind Tonhöhenvariationen qualitativ nicht zu stark an der Akzentuierung beteiligt.
- Hypothese 5/II: Die Deutsch Lernenden dehnen Akzentsilben mehr als die deutschen Sprecher. Das kann bei kurzen (offenen) Vokalen zu segmentalen Fehlern führen.
- Hypothese 6/II: Die englischen Studenten produzieren reduzierte Vokale z.T. an Stellen, an denen es im Deutschen nicht üblich ist.

Die Hypothesen 1/II, 2/II, 4b/II, 5/II und 6/II konnten bestätigt werden. Die Hypothesen 3/II und 4a/II dagegen mussten abgelehnt werden. Benkwitz folgert, dass eine globale Zuordnung von Sprachen zu den Rhythmustypen nicht ausreichend für ihre Beschreibung und die Arbeit im Phonetikunterricht für Sprachlerner ist. Der Klang einer Sprache wird wesentlich durch z.B. die Häufigkeit der (Äußerungs-)Akzente und Akzentsilbendehnungen geprägt. Außerdem ist die Akzentisochronie textsorten- und sprecherabhängig. Es zeigte sich überdies, dass es sehr wohl zu Interferenzen im Bereich der Rhythmisierung kommt, wenn Briten Deutsch sprechen. Es fiel besonders auf, dass die Lerner signifikant weniger Akzente als die englischen Probanden, jedoch mehr Akzente als die deutschen Muttersprachler setzten. Für die Arbeit mit Briten im Fremdsprachenunterricht Deutsch sollten demnach vier Schwerpunkte besondere Aufmerksamkeit erfahren: die Rhythmusgruppengröße, die Akzenthäufigkeit und –lokalisierung sowie die Länge und Qualität der Vokale in Akzentsilben.

## **2.4 Sprechrhythmus und Lyrik**

Wie die Darstellungen in den Abschnitten 2.1, 2.2 und 2.3 gezeigt haben, gibt es zum Teil sehr unterschiedliche Interpretationen und Betrachtungsperspektiven des Sprechrhythmus. Daher ist das Aufstellen einer prosodischen Taxonomie nirgendwo so schwierig wie im Bereich Rhythmus (Cummins, 2002). Eine mögliche Ursache für die schlechte Erfassbarkeit von Sprechrhythmus ist, dass das rhythmische Prinzip der Sprache zwar produktiv stets

beachtet wird, sich jedoch anderen Organisationsprinzipien unterordnet. Es stellt sich demnach die Frage, ob es Sprechsituationen oder Sprechstile gibt, die der Rhythmik der Sprache eine höhere Priorität einräumen. Als geeigneter Kandidat für eine solche rhythmische Sprechweise wird vielfach der lyrische Sprechstil betrachtet.

#### 2.4.1 Motivation

Kohler urteilt über das rhythmische Prinzip der Sprache:

„Beim Lesen tritt es viel deutlicher hervor als im freien Sprechen, und schließlich erfährt es seine weiteste Stilisierung und strikteste Befolgung in der Poesie. Für Metrik und Verslehre eröffnet sich damit eine interessante Sichtweise und ein neuer Forschungsansatz.“ (Kohler, 1982: 103)

Lehiste (1990) vertieft diese Sichtweise noch, indem sie davon ausgeht, dass sich das prosodische System einer Sprache in der metrischen Struktur seiner traditionellen Lyrik kristallisiert. Danach manifestiert sich in der Lyrik das suprasegmentale System einer Sprache mit seinen optimalen Oppositionen. Auf Grundlage dieser Hypothese betrachtet Lehiste die Lyrik als wesentlichen Gegenstandsbereich für die Sprechrhythmusforschung:

„To understand better the rhythm of spoken language, one should look at the rhythmic structure of poetry developed in that language over the years; patterns that may be imperfectly realized in prose may be manifested in a more regular fashion in poetry.“ (Lehiste, 1990: 123)

Demgemäß wird propagiert, dass eine Reihe von für die Rhythmusforschung wichtigen Phänomenen anhand lyrischer Daten zu untersuchen sind. Allen (1972) sieht in der Lyrik den nächstgelegenen Ort für die Auffindung von Evidenz bezüglich der rhythmischen Schläge von Sprache:

„The most obvious place to look for evidence concerning the beats of speech rhythm is poetry, since we can paraphrase Orwell’s pigs and say ‘Poetry is like speech, only more so’. Poets use naturally occurring phenomena to their artistic advantage; one of those phenomena is speech rhythm, which they creatively select to fit some metrical scheme.“ (Allen, 1972: 73)

Auch das in Abschnitt 2.2 skizzierte *entrainment* Phänomen sowie der damit einhergehende *harmonic timing effect* sollen in der Lyrik besonders augenscheinlich sein (Port, Tajima & Cummins, 1999). Zudem erwartet Kohler (1983), dass die Kompression von Akzentfüßen

beim Lesen rhythmisch stilisierter Lyrik stärker ist als bei Prosa. Schließlich wird die rhythmische Analyse lyrischer Sprechweise zusätzlich durch die Beobachtung motiviert, dass die Dauer von Verszeilen in der gleichen Größenordnung liegt wie biologische *action units*, die eine Dauer von etwa 1-4 s besitzen (Kien & Kemp, 1994). Eine *action unit* wird dabei auf Basis ihrer unmittelbaren Zielsetzung als Gruppe von motorischen Bewegungen mit funktionaler Beziehung zueinander verstanden. Diese semantische Definition legt nahe, dass *action units* die grundlegenden semantischen Einheiten menschlichen Verhaltens allgemein darstellen. Demgemäß betrachten Turner & Pöppel (1983) die lyrische Verszeile als eine semantische Einheit mit innerer Kohärenz. Ihre Analyse eines Korpus von 200 Gedichten des Deutschen zeigte, dass 73% der untersuchten Verszeilen innerhalb des Zeitintervalls von 2-3 s lagen. Daraufhin verifizierten Kien & Kemp (1994) dieses Ergebnis und ermittelten für die Versdauern von 6 deutschsprachigen Gedichten Werte zwischen 0.5 s und 2.5 s mit einem Median von 1.7 s. Kien & Kemp (1994) folgern, dass die Verszeile als eine Einheit äquivalent zu den *action units* von Lebewesen gesehen werden kann, wobei die enge Verwandtschaft zwischen Lyrik und Handlung zeigt, dass sich das Sprachsystem sowohl phylogenetisch als auch ontogenetisch aus dem motorischen System entwickelt hat.

#### **2.4.2 Ergebnisse**

Trotz der starken Motivation für die Analyse der Rhythmizität lyrischer Sprache sind hierzu bislang nur wenige Untersuchungen durchgeführt worden. Nord, Kruckenberg & Fant (1990) formulieren als wesentliche Fragestellungen einer solchen Untersuchung, welches die rhythmischen Entitäten sind, wie groß die zeitliche Akkuratheit bezüglich der Produktion und Perzeption eines rhythmischen Impulses ist, welche Korrespondenz zwischen Rhythmus und Metrum besteht und welche Relevanz steigende und fallende Metren (jambischer vs. trochäischer Fuß) haben. Nach Fónagy (1960) kann die Unterscheidung von Metrum und Rhythmus in Analogie zur Dichotomie *langue* vs. *parole* betrachtet werden. So werden die beiden Begriffe auch im Fortlauf dieser Arbeit verwendet. Um diesen Fragen nachzugehen beschränken sich Nord, Kruckenberg & Fant (1990) jedoch zunächst auf eine akustische Analyse der Dauer. Ein Vergleich zwischen Prosa, Lyrik und Gesang strich als einziges bemerkenswertes Ergebnis für die Lyrik heraus, dass die schwache Silbe eines trochäischen Fußes relativ zur starken Silbe länger ist als die schwache Silbe eines jambischen Fußes.

Hiermit wird eine Beobachtung bestätigt, die auch schon Newton (1981) anführt. Darüber hinaus legt Newton dar, dass die Dauern jambischer Füße insgesamt länger sind als die Dauern trochäischer Füße. Die größere Dauer des akzentuierten Segments des jambischen Fußes ist dabei entscheidend für seine Perzeption als Endsegment der jeweiligen rhythmischen Einheit. Dieser Effekt korrespondiert mit der allgemein in Endsegmenten auftretenden finalen Längung. Für die Perzeption des akzentuierten Segments des Trochäus als Anfangssegment hingegen ist seine erhöhte akustische Intensität von entscheidender Bedeutung. Diese Erkenntnis steht damit in Einklang mit den in Abschnitt 2.3 dargelegten perceptiven Gruppierungsphänomenen nichtsprachlicher akustischer Stimuli. Lehiste (1990) kann für das Lettische und Estnische bestätigen, dass die Gedichtzeile eine Einheit der zeitlichen Organisation darstellt, innerhalb derer zeitliche Kompensation stattfindet. Für das Lettische kann sie sogar zeigen, dass die Dauern metrischer Füße annähernd konstant, d. h. isochron gehalten werden.

Die akustisch-phonetischen Charakteristika lyrischen Sprechstils haben Kruckenberg & Fant (1993) untersucht. Sie gehen davon aus, dass ein Sprecher beim Wissen, Lyrik zu rezitieren, eine spezielle sich akustisch manifestierende Attitüde entwickelt. In der Untersuchung wurde ein Vergleich zwischen Gedichten des Schwedischen für Jambus und Trochäus jeweils in prosaischem und lyrischem Sprechstil vorgenommen. Es zeigte sich, dass lyrischer Sprechstil vor allem durch ein deutlich geringeres und stabileres Sprechtempo, höhere Grundfrequenz- und Intensitätswerte sowie eine geringere Modulationstiefe bei lokalen Grundfrequenzvariationen gekennzeichnet ist. Zudem fällt die Grundfrequenz bei lyrischer Leseweise versfinal nicht ab. Bezüglich der einzelnen Versfüße stellen Kruckenberg & Fant heraus, dass die Fußdauern von Jambus und Trochäus trotz des geringeren Sprechtempos etwa den Dauern prosaischer Füße entsprechen. Außerdem besitzen die metrischen Fußdauern eine geringere Streuung als die entsprechenden prosaischen Einheiten. Überdies sinkt die Dauer eines metrischen Fußes linear mit der Position innerhalb des Verses. Beim Wechsel vom prosaischen zum lyrischen Sprechstil steigt die Dauerrelation von unakzentuierter zu akzentuierter Silbe beim Jambus gegenüber dem Trochäus an. Analog dazu ist auch der Grundfrequenzkontrast zwischen unakzentuierter und akzentuierter Silbe beim Jambus größer als beim Trochäus. Darüber hinaus wirkt sich eine Emphase beim Trochäus auf einen Anstieg von Grundfrequenz, Dauer und Intensität auf der unakzentuierten



Silbe, beim Jambus dagegen auf der akzentuierten Silbe aus. Auch der Rhythmus<sup>3</sup> lyrischen Sprechstils erwies sich als regelmäßiger als bei Prosa mit einer Tendenz zur Isochronie, was sich insbesondere anhand der rhythmischen Kontinuität über Gedichtzeilen hinweg zeigte. Die Kontinuität besteht dabei darin, dass sich die Dauern von eine Pause umfassenden Interakzentintervallen als ganzzahlige Vielfache der mittleren Dauer von Interakzentintervallen am Ende eines Verses ergeben, wobei die Pausendauer mit zunehmender Phonzahl im Intervall sinkt (vgl. Abschn. 2.2).

Barney (1999) weist auf das Problem der prosodischen Auswertung von Gedichten hin. Er konstatiert, dass bei der Analyse des rhythmischen Schlages zumeist die Intonationsmuster vernachlässigt werden, bei der Analyse der Melodie hingegen der Takt außer Acht gelassen wird. Es gibt Versuche, die Intonation als Mittel zur Disambiguierung lyrischer Verse zu benutzen. Allerdings ist die Bedeutung von Intonation in wesentlich stärkerem Maße ambig als die verbale Bedeutung bzw. Wortbedeutung. Die poetische Intonation ist aber durchaus als spezielle Variante der Prosodie zu verstehen. Die Toneinheiten entsprechen dabei in der Lyrik den Verszeilen, die etwa gleiche Länge besitzen. Barney stellt fest, es müsse eine neue Formel für die poetische Intonation gefunden werden. Es soll dabei die poetische Intonation von einem generellen mündlichen Performanzstil unterschieden werden. In diesem Rahmen schlägt der Autor vor, dass die Einheiten der *nucleus group* (vom Onset bzw. Akzent bis zum finalen Akzent bzw. Nukleus) und des *tune* (eine Grundfrequenzkontur mit einer oder mehr *nucleus groups*) Verwendung finden. Darüber hinaus beobachtet Barney in der Studie lyrischer Vorträge das Phänomen der zeitlichen Diskontinuität, d.h. am Ende einer Zeile oder Halbzeile stehen Pausen und finale Längung. Des Weiteren beobachtet er, dass Schauspieler Gedichte mit einer weiteren Grundfrequenzspanne vortragen als dieses etwa bei den Autoren selbst oder bei Studenten der Fall ist. Zudem sind Echos, d.h. die Wiederholung von Grundfrequenzmustern zur Bindung paralleler Segmente, zu beobachten. Insgesamt zeigt sich die Stilisierung der poetischen Intonation als klares Merkmal. Dabei bringen die spezifisch poetischen Merkmale die formale prosodische Struktur der Lyrik zum Ausdruck. Barney unterscheidet Merkmale der Performanz, nämlich eine langsame Sprechrate, eine tiefe durchschnittliche Grundfrequenz und einen engen Grundfrequenzumfang, von spezifisch lyrischen Merkmalen. Letztere sind im einzelnen kurze Toneinheiten, mehr Pausen, relativ gleich lange Einheiten und Echos zwischen Grundfrequenzmustern. Ein Echo wird hier als

---

<sup>3</sup> Rhythmus wird von Kruckenberg & Fant (1993) als rein zeitliches Phänomen verstanden.

die Wiederholung von Grundfrequenzmustern z.B. zur Bindung paralleler Segmente verstanden. Beide oben genannten Merkmalskategorien verbindet das Auftreten von einfachen fallenden Melodien und einfachen fallenden Nuklei.

Tsur (1997; 2002a) stellt eine perceptionsorientierte Theorie der Metrik und der rhythmischen Performanz von Lyrik vor. Er stellt die zentrale Frage, wie wir zwei Verszeilen, die sehr unterschiedlich in ihrer Struktur sind, als Instanzen desselben abstrakten Musters (z.B. jambischer Pentameter) erkennen, und wie wir eine metrische von einer unmetrischen Zeile unterscheiden. Das äußerste Limit der Rhythmikalität ist die Fähigkeit oder der Wille des Lesers die Verszeile rhythmisch darzustellen. Ein großes Problem besteht, wenn das Akzentmuster und das Metrum in Konflikt miteinander stehen. In diesem Fall bringt der Leser beides in einem dritten Muster der Performanz zusammen. Tsur sieht es als die elegante Lösung an, sowohl das Akzentmuster als auch das Metrum hörbar zu realisieren. Die messbare Zeit hat dem Autor zufolge wenig zu bieten, um den poetischen Rhythmus näher zu beleuchten. Danach basiert der poetische Rhythmus nicht auf gleichen oder proportionalen Zeitintervallen, er basiert vielmehr auf einem abstrakten Muster, welches irgendwie im wahrnehmenden Bewusstsein existiert. Manchmal wird es durch wahrgenommene Akzente bestätigt, manchmal auch nicht. Zur Korrespondenz zwischen Metrum und Rhythmus sei an dieser Stelle erneut auf Fónagy (1960) verwiesen. Ein essentieller Teil von Tsurs Theorie ist die Überartikulation. Viele Prädiktionen der perceptionsorientierten Theorie werden im Sinne von Gruppierung und Überartikulation gemacht. Eine Gruppierung impliziert Kontinuität, Überartikulation dagegen weist auf Diskontinuität hin. Hierbei ist der stärkste Verursacher der Gruppierung ein später Gipfel in der Intonationskontur. Darüber hinaus ist die rhythmische Performanz aufeinanderfolgender Akzente und Akzentmaxima in schwacher Position wichtig für die perceptionsorientierte Theorie von Tsur. Letzten Endes kann jedoch nur durch Hören herausgefunden werden, was die Qualität des Ganzen ist, kontinuierlich, diskontinuierlich oder eben beides. Wenn ein Leser gebeten wird eine Zeile rhythmisch zu lesen, so dass sowohl das Akzentmuster als auch das Metrum so gut es eben geht bewahrt werden, neigt der Leser dazu, die letzten vier Silben emphatisch zu gruppieren und die Gruppe vom vorhergehenden Kontext abzusetzen. Diese perzeptive Gruppe heißt Akzenttal (*stress valley*). Es zeichnet sich dadurch aus, dass zwei unakzentuierten Silben eine akzentuierte Silbe vorausgeht und eine weitere akzentuierte Silbe nachfolgt. Ein Akzenttal kann dazu dienen, ein Akzentmaximum in einer schwachen Position

anzupassen. Zum Begriff des Akzentes ist dabei zu sagen, dass Tsur (1997) den akustischen Parameter der Dauer als effektives Merkmal für Akzentuierung ansieht.

Außerdem diskutiert Tsur (2002b) Aspekte der kognitiven Lyrik (*cognitive poetics*). Die kognitive Lyrik geht davon aus, dass lyrische Texte nicht nur Bedeutung oder Gedanken transportieren, sondern auch emotionale Qualitäten, die der Leser perzipiert, aufweisen. Die Lyrik nutzt für ästhetische Zwecke kognitive Prozesse, welche sich ursprünglich für nichtästhetische Zwecke entwickelt haben. Tsur untermauert nochmals, dass die rhythmische Performanz von Lyrik eine Aktivität der Problemlösung ist. Wenn das linguistische Muster und das Versifikationsmuster in Konflikt miteinander stehen, wird ein drittes Muster der Performanz akkomodiert, d.h. beide Muster sind gleichzeitig wahrnehmbar. Das Muster der Versifikation existiert dabei nur im kognitiven System als ein metrischer Satz (*metrical set*), d.h. einer Erwartung. Der poetische Rhythmus besteht folglich aus drei gleichzeitigen Mustern: Versifikationsmuster, linguistisches Muster und Performanz. Tsur versucht, akustische Korrelate dieser Theorie aufzuzeigen. Leider wird seine Theorie dadurch nicht völlig plausibilisiert. Es kann dazu nur soviel geäußert werden, dass Tsur die phonetische Manifestation der Performanz, zumindest für das Englische, in verlängerten Frikativen und dem Vorhandensein von Glottalverschlüssen vor Vokalen vermutet.

## **2.5 Fazit**

Obiger Überblick zeigt, dass schon seit Jahrzehnten auf dem Gebiet des Sprechrhythmus geforscht wurde. Während dieser Zeit ist die Existenz der strengen Isochronie, zumindest als produktives Phänomen, widerlegt worden. Außerdem sind mit  $\%V/\Delta C$  akustische Korrelate zur Rhythmusklassenunterscheidung gefunden worden. Es bleibt aber die Frage offen, welche akustischen Merkmale die Perzeption von Rhythmus bestimmen. An dieser Stelle muss noch Forschungsarbeit geleistet werden. Abschnitt 2.4 legt dar, dass der lyrische Sprechstil möglicherweise eine leichtere Erfassbarkeit des Sprechrhythmus zulässt und deshalb einen besonders geeigneten Forschungsgegenstand darstellt. An dieser Stelle knüpfen die empirischen Untersuchungen der vorliegenden Arbeit an.

## 3 Zum Begriff der Prosodie

Wie bereits Kapitel 2 gezeigt hat, manifestiert sich der Sprechrhythmus, auch wenn er nicht vollkommen unabhängig von der segmentalen Ebene ist, vor allem in der Prosodie der Sprache. Deshalb soll an dieser Stelle auf den Begriff der Prosodie eingegangen werden. In diesem Kapitel soll jedoch kein allgemeiner Literaturüberblick zu Arbeiten im Bereich der Prosodieforschung gegeben werden. Da der Begriff der Prosodie in der Literatur nicht einheitlich verwandt wird, soll stattdessen vielmehr versucht werden, den Begriff der Prosodie inhaltlich von verschiedenen Seiten zu beleuchten. In Abschnitt 3.1 wird die Definierbarkeit des Begriffes Prosodie erörtert. In Abschnitt 3.2 wird die Funktion der Prosodie diskutiert.

### 3.1 Definierbarkeit

Sonntag (1999) weist darauf hin, dass die Prosodie zumeist als Teilbereich der Metrik gesehen wird. Zur Metrik von Prosa und dem damit einhergehenden Prominenzbegriff sei auf Wagner (2002) verwiesen. Es beschäftigen sich Vertreter der verschiedensten Fachrichtungen vom Phonetiker über den Psychologen bis hin zum Ingenieur mit dem Gebiet der Prosodie. Sonntag (1999) macht das Problem der Definition von Prosodie aber vor allem an drei Punkten fest:

1. Vorhandensein ausschließender Definitionen und wechselseitige Verweise der Begriffe *Prosodie*, *Intonation* und *Suprasegmentalia*
2. komplexes Zusammenspiel von artikulatorischen, akustischen und auditiven Komponenten
3. prosodische Relevanz: prosodische Funktionen werden in der Definition oft nur implizit angesprochen

Klarheit hingegen herrscht darüber, dass die akustischen Größen Grundfrequenz, Dauer und Intensität die drei Wahrnehmungsgrößen Tonhöhe, subjektive Dauer und Lautheit beim Hörer hervorrufen.

Des Weiteren stellt sich das Problem, ob der Prosodie ausschließlich linguistische oder auch para- und extralinguistische Phänomene zuzuordnen sind. Dabei ist noch nicht einmal klar, welche Phänomene welcher der drei Kategorien zuzuordnen sind. Man könnte beispielsweise annehmen, dass Räuspern, Lachen oder Seufzen extralinguistischen Charakter haben. Gleichzeitig kann aber durchaus geltend gemacht werden, dass die genannten Phänomene intentional vom Sprecher eingesetzt werden können. Damit lägen sie im Bereich der linguistischen Funktionalität. Eine Trennung der sich so überlagernden Bereiche kann nur theoretisch geschehen. Somit gelangt man zur Dichotomie von Phonetik und Phonologie, wobei jedes der beiden Gebiete einen anderen Prosodiebegriff hat. Die Phonologie hat eine abstrakte Sichtweise, die Phonetik dagegen eine konkrete Sichtweise auf die Prosodie. Sonntag (1999) betont allerdings, dass in jüngerer Zeit die Phonologie Evidenz in der Phonetik und die Phonetik ihren theoretischen Überbau in der Phonologie sucht. Dadurch wird die Abgrenzung dieser beiden Gebiete erschwert.

Abschließend für diesen Abschnitt seien zwei Definitionen von Prosodie unter den oben gemachten Einschränkungen angeführt. Die erste Definition ist einem Nachschlagewerk von Bußmann (1983) entnommen:

„Untersuchung sprachlicher Eigenschaften wie Akzent, Intonation, Sprechpausen u. a., die sich auf größere Einheiten als einzelne Phoneme beziehen bzw. diese überlagern. Man bezeichnet sie daher auch als suprasegmentale Merkmale. Die P. kann als Verbindung zwischen Phonologie und Syntax bezeichnet werden, insofern Silben, Wörter und Sätze ihr Untersuchungsgegenstand sind.“ (Bußmann, 1983: 417)

Die zweite Definition von Prosodie lautet nach Sonntag:

„Prosodie ist das gestalthafte Perzept, das durch die 'klassischen' akustischen prosodischen Parameter Grundfrequenz, Dauer und Intensität evoziert wird.“ (Sonntag, 1999: 8)

Letztere Definition nach Sonntag soll wegen ihrer Konzentration auf die perzeptive Seite der Prosodie die für die vorliegende Arbeit maßgebende sein.

## **3.2 Funktion**

Bezüglich der Funktion von Prosodie lässt sich feststellen, dass es auf der einen Seite sprecher- und sprachspezifisch unterschiedliche Formen mit zum Teil gleicher Funktion gibt, und auf der anderen Seite aber auch sprecher- und sprachunabhängige Normen existieren.

Die sprecherunabhängigen Normen sind dabei eine notwendige Bedingung für die linguistische Funktionalität von Prosodie (Sonntag, 1999). Hierbei können inhaltliche, situative und sprecherbezogene prosodische Funktionen unterschieden werden. Die inhaltlichen Funktionen werden dabei in Akzentuierung und Phrasierung unterteilt, die situativen Funktionen untergliedern sich in Sprechereinstellung und Emotionen, und die sprecherbezogenen Funktionen schließlich geben Auskunft über den Sprecher. Es ist dabei jedoch wichtig, dass der Prosodie in den meisten Fällen lediglich eine unterstützende und keine allein disambiguierende Rolle zukommt.

### **3.3 Fazit**

Kapitel 3 hat aufgezeigt, dass die Definition von Prosodie sehr schwierig ist. Dieser Umstand liegt vor allem in den wechselseitigen Verweisen zwischen *Prosodie*, *Intonation* und *Suprasegmentalia* sowie dem komplexen Zusammenspiel von artikulatorischen, akustischen und auditiven Komponenten begründet. Zudem ist strittig, ob der Prosodie ausschließlich linguistische oder aber auch para- und extralinguistische Sprachphänomene zuzuordnen sind. Für die Funktion von Prosodie sind die sprecher- und sprachunabhängigen Normen von großer Bedeutung. Inhaltliche, situative und sprecherbezogene prosodische Funktionen müssen unterschieden werden.

## 4 Lyrik und Metrik

Da sich diese Arbeit mit gesprochener Lyrik<sup>4</sup> beschäftigt, soll in diesem Abschnitt kurz auch auf die literaturwissenschaftliche Sichtweise des dichterischen Vortrags, speziell seines Rhythmus, eingegangen werden<sup>5</sup>. Gelfert (1998) schreibt dem Gedicht eine paralinguistische Ebene mit den Bauelementen Laut, Silbe/Versfuß, Vers und Strophe/Gedicht zu. Diese paralinguistische Ebene leistet dabei keinen Beitrag zur Sprachbedeutung, sondern hat eine rein ästhetische Wirkung. Außer den Bauelementen existieren zusätzlich die Bindemittel von Metrum und dem im Rahmen dieser Arbeit zu vernachlässigenden Reim. Der lateinische Begriff des Metrums, sprich Versfuß, leitet sich vom griechischen Metron (Maß) ab. Gelfert unterscheidet überdies das isosyllabische (silbenzählende) von dem isoiktischen (akzentzählenden) Ordnungsschema. In der Antike folgten die Versfüße dem quantifizierenden (vgl. Binder et al., 1984) bzw. isochronen Prinzip mit Jambus (kurz – lang), Trochäus (lang – kurz), Daktylus (lang – kurz – kurz), Anapäst (kurz – kurz – lang) und dem in den germanischen Sprachen nicht vorkommenden Spondeus (lang – lang). Für das im Rahmen dieser Arbeit aufgenommene Gedichtkorpus sind aber lediglich Jambus, Trochäus, Daktylus und Liedform berücksichtigt worden. Die Liedform konstituiert sich durch eine betonte Silbe zuzüglich einer variablen Anzahl von unbetonten Silben. Rein anapästische Gedichte sind der Literaturwissenschaft für das Deutsche nicht bekannt. Das quantifizierende Prinzip der Antike und der germanischen Dichtung wurde in der Neuzeit durch das akzentuierende bzw. isoiktische Prinzip ersetzt. Hierbei wurden Längen und Kürzen durch *Hebungen* (Betonung) und *Senkungen* (keine Betonung) ersetzt (vgl. Binder et al., 1984). Darüber hinaus existiert in den romanischen Sprachen noch das oben schon erwähnte silbenzählende bzw. isosyllabische Prinzip mit einer konstanten Silbenzahl pro Verszeile. Des Weiteren entsteht durch Reibungen am abstrakten Metrum der Rhythmus, und es gibt eine freie Bewegung innerhalb des Metrums. Laut Gelfert gilt dabei:

„Der Rhythmus ist der eigentliche Lebensnerv eines Gedichts. Kunstvoll abgestufte Klangfarben können der Oberfläche des Sprachteppichs Ausdruck verleihen, aber erst der Rhythmus erweckt die dichterische Sprache zum Leben.“ (Gelfert, 1998: 50)

---

<sup>4</sup> Der Begriff *Lyrik* wird im Rahmen dieser Arbeit äquivalent zum englischen Begriff *poetry* verwandt.

<sup>5</sup> In der Literatur wird nicht konsequent zwischen der Schrift- und der Vortragsform unterschieden.

Außerdem können die gleichlangen Takte der akzentuierenden Metrik unterschiedlich gefüllt werden, was auch für den letzten Takt des Verses gilt. Hat ein Vers mehr als sechs Takte, gliedert der Hörer die Zeile in kleinere Takteinheiten.

Binder et al. (1984) unterscheiden hebungsfordernde, senkungsfordernde und hebungs- und senkungsfähige Silben. Ein Gedicht hat demzufolge Silben, die in aller Regel akzentuiert werden müssen, und Silben, die optional akzentuiert werden können. Die tatsächliche Betonung der Silben wird letztlich vom zugrunde liegenden Metrum bestimmt. Der Vers steht als *end-stopped line* bzw. Enjambement in Beziehung zur Syntax. Dabei werden die Verse nach griechischem Vorbild und damit nach dem fußmessenden Verfahren in Versfüße gegliedert. Darüber hinaus erwähnen auch Binder et al. das silbenzählende Verfahren mit feststehender Zahl an Silben pro Verszeile und einer festen Akzentstelle<sup>6</sup> sowie einem Endreim. Des Weiteren führen die Autoren den Begriff der schwebenden Betonung an. Danach wird eine dem Metrum nach unbetonte Silbe wegen ihres großen semantischen Wertes trotzdem betont und gerät in Spannung zum metrischen Schema. Beim taktgliedernden Verfahren schließlich variiert die Zahl der unbetonten Silben zwischen Hebungen (Füllungsfreiheit). Der Hebungsabstand ist dabei in etwa gleich (im allgemeinen  $2/3$  s)<sup>7</sup>. Das Hebungsintervall (Verstakt) kann analog zum musikalischen Takt betrachtet werden. Man kann in dieser Hinsicht Silben mit Noten assoziieren. So ist auch die Beschreibung von Versen in einem an die Musik angelehnten Notationssystem grundsätzlich möglich. Derartige Systeme spielen aber für die vorliegende Arbeit keine Rolle und werden somit an dieser Stelle nicht vertieft.

Weiterhin unterscheiden Binder et al. (1984) drei Versgegenden, nämlich den Auftakt, das Versinnere und die Kadenz. Hinsichtlich der Kadenz wiederum werden die volle Kadenz (letzte vom Metrum geforderte Hebung ist realisiert), die stumpfe Kadenz (Vorhandensein eines zusätzlichen pausierten Taktes) und die klingende Kadenz (letzte vom Metrum geforderte Hebung ist durch eine prosaisch unbetonte Silbe realisiert) voneinander unterschieden. Als Beispiel für ein füllungsfreies Metrum sei an dieser Stelle der germanische Knittelvers, wie er etwa in Goethes Faust vorkommt, genannt. Der Knittelvers gliedert sich in einen freien Auftakt (0+ Silben), 4 Takte bei einem Zweivierteltakt

---

<sup>6</sup> Bei langen Versen können auch zwei Akzentstellen vorkommen.

<sup>7</sup> Diese beiden Eigenschaften sind die einzige Regelmäßigkeit in den Versen germanisch-deutscher Tradition.



(Taktgeschlecht), eine volle Kadenz und schließlich den stets vorhandenen variablen Reim. Diese Struktur ist im folgenden Beispiel, das Goethes Faust entstammt, ersichtlich:

| Habe nun |, ach! | Philoso | phie,  
Ju | riste | rey und | Medi | cin,  
Und | leider | auch | Theolo | gie!  
Durch | aus stu | diert, mit | heißem Be | mühn.  
Da | steh' ich | nun, ich | armer | Thor!  
Und | bin so | klug als | wie zu | vor;  
Heiße Ma | gister, | heiße | Doctor | gar,

Die Taktstriche sind durch ‚|‘ gekennzeichnet. Im obigen Beispiel ist die durchgängige Viertaktigkeit des Textausschnittes gut zu erkennen. Der Auftakt umfasst im ersten Vers keine Silbe, in den folgenden fünf Versen eine Silbe und dem letzten Vers dagegen drei Silben. Mit dem Zweivierteltakt ist gemeint, dass jeder Takt zwei Schläge bzw. Silben besitzt. Ein Schlag kann dabei durch zwei kurze Silben realisiert sein. Außerdem ist die volle Kadenz ersichtlich, d.h. die letzte vom Metrum geforderte Hebung ist auch tatsächlich realisiert. Schließlich weist das Beispiel die Reimform ababccd auf.

## 5 Korpuserstellung

Wie in Abschnitt 2.4 dargelegt, eignet sich Lyrik gut als Gegenstand für rhythmisch-prosodische Analysen. Bisher ist kein geeignetes Korpus deutschsprachiger, gelesener Gedichte vorhanden. Deshalb mussten für das Vorhaben dieser Arbeit zunächst einmal Daten gesammelt werden. Für die Textsuche und die Aufnahmen der Gedichtrezitationen mussten einige zu erläuternde Kriterien beachtet werden. Die Datenerhebungsphase samt eines kurzen Exkurses auf das Gebiet der GToBI-Etikettierung wird im folgenden näher erörtert.

### 5.1 Textauswahl

Das aufzunehmende Lyrik Korpus musste groß genug sein, um valide Statistiken über akustisch-phonetische Charakteristika der Daten zu erheben. Gleichzeitig musste das Korpus aber klein genug sein, um handhabbar zu bleiben. Das wichtigste Auswahlkriterium für die zu berücksichtigenden Texte war, dass alle für das Deutsche bedeutenden Metren vorkamen. Die Zuweisung eines Metrums zu einem Text gestaltete sich dabei schwierig. Um an dieser Stelle keine Fehlzuweisung zu unternehmen, wurde als Fachmann Dr. von Tiedemann aus der Abteilung für Komparatistik der Universität Bonn zu Rate gezogen. Mit seiner Hilfe schließlich wurden jeweils drei kurze Texte (je nach Strophenlänge etwa 3-4 Strophen) für Jambus und Trochäus und jeweils 2 Texte für Daktylus und Liedform ausgewählt. Der Anapäst kommt im Deutschen (durchgängig) nicht vor und wurde deshalb von der Aufnahme ausgeschlossen. Die Textvorlagen für die Aufnahmen können in Anhang B eingesehen werden. Außerdem ist der Daktylus bei den berücksichtigten Texten unvermeidlich von trochäischen Versfüßen durchmischt. Darüber hinaus wurde bei der Textauswahl darauf geachtet, dass die Texte so jung wie möglich waren. Dadurch sollte vermieden werden, dass eine archaische Variante des Deutschen die Leser irritiert. Zudem sollten die Analysen der Aufnahmen der berücksichtigten Texte grundsätzlich Rückschlüsse auf die deutsche Gegenwartssprache zulassen. Um zusätzlich Material zur Verfügung zu haben, das mit absoluter Sicherheit als prosaisch anzusehen ist, wurden 2 kurze Texte aus einer Zeitung ausgewählt. Tabelle 5.1 listet alle gelesenen und damit aufgenommenen Texte auf.

<b>Jambus</b>	<b>Trochäus</b>	<b>Daktylus</b>	<b>Liedform</b>	<b>Zeitung</b>
Georg Heym: "Der Blinde"	Christian Morgenstern: "Der Träumer"	Johann Wolfgang von Goethe: "Reineke Fuchs"	Joseph von Eichendorff: "Sehnsucht"	"Padre Pio"
Johann Wolfgang von Goethe: "Iphigenie - erster Aufzug"	Heinrich Heine: "Jehuda ben Halevy / II"	Friedrich Hölderlin: "Brot und Wein - An Heinze"	Johann Wolfgang von Goethe: "Faust"	"Russland neu"
Durs Grünbein: "Der Schnee von heute"	Bert Brecht: "Legende von der Entstehung des Buches Taoteking auf dem Weg des Laotse in die Emigration"			

**Tab. 5.1:** Die einzelnen Metren und die ihnen zugehörigen Gedichte bzw. Gedichtauszüge und Autoren

## 5.2 Aufnahme

Für die Aufnahme des Lyrik-Korpus wurden 12 Sprecher konsultiert, 6 professionelle und 6 nichtprofessionelle Sprecher. Als professionelle Sprecher wurden ausgebildete Schauspieler gewählt, die im Rahmen ihrer Ausbildung auch das Vortragen von Lyrik erlernt haben. Die nichtprofessionellen Sprecher dagegen waren Studierende verschiedenster Fachrichtungen. Letztere waren allesamt Hobbymusiker, um ein gewisses Gefühl für Rhythmik voraussetzen zu können. Jede Sprechergruppe bestand aus drei Frauen und drei Männern. Da das Korpus unter anderem Aufschluss über den Gegensatz zwischen poetischer und prosaischer Leseweise geben sollte, sind die ausgewählten Gedichte sowohl in einem prosaischen als auch poetischen Textsatz präsentiert worden. An dieser Stelle ist der von Kruckenberg und Fant (1993) verwandten Formatierungs- und Präsentationsmethode Folge geleistet worden. Dabei sind die Texte einmal in ihrer Ursprungsform und einmal in einem modifizierten Satz vorgelegt worden. Bei der Modifikation sind die Strophen aufgelöst worden, und die Satzzeichen sind, wo es nötig war, den Regeln der deutschen Orthografie angepasst worden. Zunächst sind dann erst die prosaischen Varianten und danach die lyrischen Varianten den

Sprechern zum Lesen vorgelegt worden. Den Sprechern ist dabei verschwiegen worden, dass den Prosavarianten eigentlich Gedichte zugrunde lagen. Erst bei der Vorlage der originalen Gedichtform wurden die Leser über den dichterischen Ursprung der Texte aufgeklärt. Die beiden Texte aus der Zeitung wurden aufgenommen, um den aus phonetischer Sicht prosaischen Charakter der Gedichte im Prosasatz besser beurteilen und vergleichen zu können. Bei der Vorlage der Prosavarianten der Gedichte wurden die Sprecher lediglich instruiert, die Texte flüssig und in mittlerem Tempo vorzulesen. Da die meisten Leser mehrere Anläufe für den flüssigen Vortrag eines Textes brauchten, konnte festgestellt werden, dass Profis und Laien eine unterschiedliche Tendenz beim Vortrag der Lyrikvarianten zeigten. Die Prosodie der Profis war sehr lebendig und variantenreich. Die Laien dagegen tendierten in einigen Fällen zu einem Stakkatostil. Deshalb wurde entschieden, nur den Profisprechern volle Freiheit für die Ausgestaltung des poetischen Vortrags zu geben. Die Laiensprecher sollten hingegen das zugrunde liegende Metrum, soweit das die Betonungsstruktur des jeweiligen Wortes zuließ, deutlich hörbar realisieren. Dabei entstand dann der oben erwähnte Stakkatostil, der mal mehr und mal weniger stark ausgeprägt war.

Jeder Sprecher wurde in einer Sitzung von etwa 2 Stunden Dauer aufgenommen. Die Sprecher erhielten dafür ein Honorar. Die Aufnahmen fanden im schalltoten Raum des Instituts für Kommunikationsforschung und Phonetik der Universität Bonn statt. Es wurde ein Stereosignal bestehend aus dem Sprachsignal auf dem einen Kanal und dem Laryngogramm auf dem anderen Kanal aufgenommen. Beide Signale wurden mit 16 kHz und 16 bits digitalisiert. Einige Passagen der Vorträge mussten z.B. wegen Versprechern wiederholt werden, so dass das Material im Anschluss an die Aufnahme noch geschnitten werden musste.

### **5.3 Annotation**

Für eine erschöpfende Analyse des Korpus deutschsprachiger Gedichte war es nötig, eine linguistisch-phonetische Annotation der Daten vorzunehmen. In Tabelle 5.2 sind die wichtigsten annotierten Merkmale samt ihrer Definition aufgelistet. Diese sollen nun diskutiert werden. Für eine vollständige Liste der Annotationen sei auf Anhang A verwiesen. Die Annotationen beinhalten die SAMPA-Transkriptionen der Silben im Korpus (*Syllable*).

Außerdem sind die Dauern der Silben in Millisekunden (ms) gemessen und gespeichert worden (*Duration*). Ein zentraler Gegenstand dieser Arbeit nämlich ist die Dauermodellierung

<b>Etikett</b>	<b>Definition</b>
Syllable	Silbe in SAMPA-Transkription
Duration	Dauer der Silbe auf volle ms gerundet
PhraseDur	Dauer der gesamten Phrase, in der die aktuelle Silbe steht, auf ms gerundet
VerseDur	Dauer des gesamten Verses, in dem die aktuelle Silbe steht, auf ms gerundet
FootDurExclP	Fußdauer ohne Pausen; Dauer des aktuellen Fußes ohne Einbeziehung von Pausen (weder an Rändern noch im Fuß), auf ms gerundet; die Pausen erhalten auch den Wert des aktuellen Fußes.
FootDurInclP	Fußdauer mit Pausen; Dauer des aktuellen Fußes unter Einbeziehung von Pausen (sowohl an Rändern als auch innerhalb des Fußes), auf ms gerundet; die Pausen erhalten auch den Wert des aktuellen Fußes.
PhrasePosCat	Kategoriale Phrasenposition; F = final (letzte Silbe der Phrase); die erste Hälfte der übrigen Silben = I (initial); die zweite Hälfte der übrigen Silben = M (medial)
SylQuotExclP	Fußquotient ohne Einbeziehung von Pausen; Quotient aus Dauer der unbetonten und Dauer der betonten Silbe des aktuellen Fußes; auf 100 mit zwei Nachkommastellen gerundet
SylQuotInclP	Fußquotient mit Pausen; Quotient aus Dauer der unbetonten und Dauer der betonten Silbe des aktuellen Fußes; auf 100 mit zwei Nachkommastellen gerundet
PhonesInSyl	Anzahl der Phone in der aktuellen Silbe
PhonesInFoot	Gesamtanzahl von Phonen im aktuellen Fuß
F0Mean	Durchschnittlicher F0-Wert der Silbe
ToBiAccent	Das aktuelle ToBI-Etikett, falls in der Silbe vorhanden, sonst 0
ToBiBorder	Das aktuelle ToBIBorder-Etikett, falls die Silbe damit endet, sonst 0

**Tab. 5.2:** Im Korpus annotierte Merkmale samt ihrer Definitionen

und –prädiktion. Die Silbendauern wurden somit vor allem mit statistischen Methoden ausgewertet. Dem gleichen Zweck dienen die Variablen *PhraseDur* und *VerseDur*, die die Phrasen- bzw. Versdauern auflisten. Ebenso sind die Fußdauern, einmal exklusive der Pausen (*FootDurExclP*) und einmal inklusive der Pausen (*FootDurInclP*), verzeichnet. Der Fuß ist eine zentrale rhythmische Einheit in den akzentzählenden Sprachen wie dem Deutschen, und seine jeweilige Dauer wird zahlreichen noch zu erörternden Analysen zugeführt. Die Variable *PhrasePosCat* gibt Aufschluss darüber, an welcher Position eine Silbe im aktuellen Vers steht. Es gelten hierbei lediglich die drei Kategorien initial, medial und final. Als final gilt dabei die allerletzte Silbe der Phrase. Die übrigen Silben der aktuellen Phrase gliedern sich zu gleichen Teilen in initiale (erste Hälfte) und mediale (zweite Hälfte) Silben. Des Weiteren ist der Quotient aus den Silbendauern bzw. der Silbendauer des unakzentuierten Fußabschnittes

und der Dauer der akzentuierten Silbe berechnet worden. Hier werden ebenfalls wieder eine pausenausschließende (*SylQuotExclP*) und eine pausenumfassende (*SylQuotInclP*) Variante aufgeführt. Diese Quotienten spielen eine große Rolle für die noch zu erörternde lyrische Dauermodellierung. Für die Dauermodellierung ist außerdem die Anzahl der Phone in der Silbe (*PhonesInSyl*) von großer Bedeutung. Hierzu ist anzumerken, dass für die Berechnung der Variable die tatsächliche Realisierung einer Silbe, nicht hingegen ihre kanonische Form maßgebend war. Zudem sind Längungszeichen als Phon aufgefasst worden. Analog dazu ist auch die Anzahl der Phone im Fuß (*PhonesInFoot*) gemessen worden. Letztere Variable ist für die Erhebung rhythmusbasierter Statistiken wichtig. Schließlich sind noch drei F0-basierte Merkmale zu nennen. Die durchschnittliche Grundfrequenz einer Silbe (*FOMean*) dient abermals zur statistischen Analyse von Rhythmus bzw. Intonation<sup>8</sup> der Metren in lyrischer Form und der ihnen äquivalenten prosaischen Signalabschnitte. Eine phonologische Beschreibung der Intonation liefern die Merkmale *ToBIAccent* und *ToBIBorder*. Ersteres führt das aktuelle GToBI-Etikett einer Silbe an. Letzteres dagegen gibt an, welcher Grenzakzent nach der aktuellen Silbe folgt. Gibt es kein aktuelles Etikett, wird 0 eingetragen.

Die oben erläuterten Variablen sind, wo es nötig war, von Hand etikettiert worden. Einige Variablen aber konnten auch durch Skripte automatisch generiert werden. An dieser Stelle sei auch das automatische GToBI-Etikettierwerkzeug *Prosodizer* (Braunschweiler, 2003) genannt. Mit diesem Programm wurden die GToBI-Etikette (*ToBIAccent* und *ToBIBorder*) zunächst automatisch generiert. Die Ergebnisse mussten allerdings einer Handkorrektur unterzogen werden. Der GToBI-Annotationsformalismus wird nun in Abschnitt 5.3.1 näher erörtert.

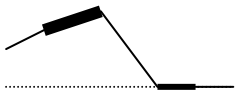
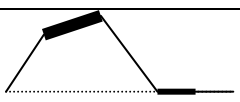


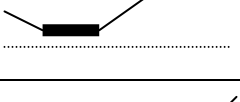

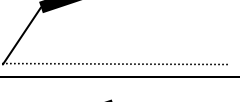

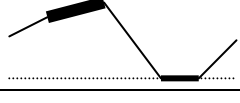
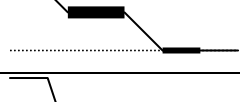
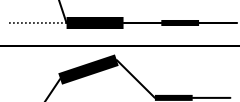
### 5.3.1 GToBI-Etikettierung

Das GToBI-Etikettiersystem für Intonationskonturen (Baumann, Grice & Benz Müller, 2001) deutscher Sprache ist Grundlage für die entsprechenden Annotationen im Lyrik-Korpus und bildet gleichzeitig das Notationssystem für das noch darzulegende Intonationsmodell für poetische Verse. Deshalb soll das GToBI-System hier kurz erklärt werden. GToBI steht für

---

<sup>8</sup> Welche Rolle die Intonation für den Rhythmus gesprochener Sprache spielt ist bislang noch nicht völlig geklärt.

German *Tones* and *Break Indices*. Das System basiert auf dem Tonsequenzmodell (Pierrehumbert, 1980) und ähnelt dem Etikettierformalismus E-ToBI für das Englische.

	GToBI	Schematic Contour	Context	Example
Fall	H* L-%		Neutral statement	Mein <b>ZAHN</b> tut WEH. <sup>1</sup> <i>My tooth hurts.</i>
			Neutral W-question	Wo hast du den <b>WA</b> gen ge <b>PARKT</b> ? <sup>1</sup> <i>Where did you park the car?</i>
	L+H* L-%		Contrastive assertion	Schon der Ver <b>SUCH</b> ist <b>STRA</b> Fbar! <sup>2</sup> <i>Even to attempt is an offence!</i>
Rise-Fall (Late Peak)	L*+H L-%		Self-evident assertion	Das <b>WEISS</b> ich <b>SCHON</b> ! <sup>6</sup> <i>I already know that!</i>
			Emotionally committed or sarcastic assertion	Der Blick ist ja <b>F</b> abelhaft! <sup>3</sup> <i>The view is fantastic!</i>
Rise	L* H-^H%		Neutral yes/no-question	Tauschen Sie auch <b>BRIEF</b> MArken? <sup>1</sup> <i>Do you also exchange stamps?</i>
			Echo question	Von wem ich das <b>H</b> Abe? <sup>2</sup> <i>From whom I have it?</i>
	L* L-H%		Indignation	<b>DOCH</b> ! <i>It is!</i>
			Answering phone	<b>BECK</b> en <b>BA</b> Uer? <sup>4</sup>
	(L+)H* H-^H%		Follow-up question	...oder ist Ihr <b>BR</b> Uder <b>HIER</b> ? <sup>5</sup> <i>...or is your brother in?</i>
Level	L+H* H-(%)		Incompleteness	<b>AN</b> derer <b>SEITS</b> ... <sup>6</sup> <i>But then again...</i>
			Ritual expression	Guten <b>MOR</b> gen! <sup>3</sup> <i>Good morning!</i>
Fall-Rise	H* L-H%		Polite offer	Mögen Sie <b>ROG</b> gen <b>BRÖT</b> chen? <sup>1</sup> <i>Would you like rye rolls?</i>
Early Peak	H+!H* L-%		Established fact	Hab' ich mir schon ge <b>DACHT</b> . <sup>7</sup> <i>That's what I thought.</i>
	H+L* L-%		Soothing / Polite request	Nun er <b>ZÄH</b> le doch <b>MAL</b> ! <sup>2</sup> <i>Just tell me about it!</i>
Stylised Step Down	(L+)H* !H-%		Calling contour	<b>BECK</b> en <b>BA</b> Uer!

**Tab. 5.3:** Die am häufigsten im Deutschen vorkommenden nuklearen Intonationskonturen und einige Beispiele ihres Gebrauchs aus Baumann, Grice & Benz Müller (2001)<sup>9</sup>

Die gängigsten Töne und Tonkombinationen für das Deutsche können in Tabelle 5.3 im Überblick eingesehen werden. GToBI umfasst drei Etikettierebenen (*tiers*), nämlich *tones*,

<sup>9</sup> Da die Übersetzung der englischen Terminologie in der Tabelle irreführend sein könnte, sind die originalen Begrifflichkeiten bewahrt worden.

*break indices* und *words*. Es werden Tonakzente und Grenztöne unterschieden. Außer den Tönen gibt es Diakritika für die Modifikation des Frequenzumfangs. Diese sind zum einen „^“ für *upstep* und zum anderen „!“ für *downstep*, d.h. der Umfang kann entweder erweitert oder verringert werden. Das Toninventar besteht aus den zwei monotonalen Tonakzenten „H\*“ und „L\*“ sowie den vier bitonalen Tonakzenten „L+H\*“, „L\*+H“, „H+L\*“ und „H+!H\*“. Dabei kennzeichnet der „\*“ den Ton auf der akzentuierten Silbe. Die bitonalen Akzente haben einen Vor- oder Nachlauf auf der vorherigen bzw. nachfolgenden Silbe. Zu nennen ist an dieser Stelle auch die Möglichkeit eines frühen Gipfels (*early peak*), der durch die Töne „H+!H\*“ und „H+L\*“ gekennzeichnet ist. Hierbei liegt das Grundfrequenzmaximum vor der akzentuierten Silbe. Außerdem gibt es die beiden Grenztöne für kleine Grenzen (Intermediärphrasen) „L-“ und „H-“, sowie die beiden Grenztöne für große Grenzen (Intonationsphrasen) „L%“ und „H%“. Zu diesen Phrasentypen ist festzustellen, dass die autosegmental-metrischen Ansätze zunächst von nur einem Typ ausgehen. Féry (1993) und Pierrehumbert (1980) allerdings berücksichtigen, so wie es auch in GToBI der Fall ist, zusätzlich die kleinen Intermediärphrasen. Die Semantik der auf diese Weise phonologisch beschriebenen Intonationskonturen soll hier außer Acht gelassen werden, da sie nicht Gegenstand dieser Arbeit ist.

## 5.4 Zusammenfassung

Kapitel 5 hat dargelegt, dass zehn Gedichte bzw. Gedichtauszüge zuzüglich zweier Zeitungsausschnitte von zwölf Sprechern gelesen und währenddessen aufgenommen wurden. Dabei entfielen jeweils drei Gedichte auf die Metren Jambus und Trochäus. Daktylus und Liedform dagegen umfassten jeweils zwei Gedichte. Die akustische Datensammlung wurde umfangreich annotiert, so dass eine weitreichende statistische Auswertung der akustisch-prosodischen Eigenschaften der Daten ermöglicht wurde. Besonders zeitaufwendig war dabei die automatische und anschließend handkorrigierte GToBI-Etikettierung, die sich im Wesentlichen in die bitonalen Akzentöne („L+H\*“, „L\*+H“, „H+L\*“ und „H+!H\*“) und die Grenztöne („L-“ bzw. „L%“ und „H-“ bzw. „H%“) aufschlüsselt.



## 6 Datenanalyse

In diesem Kapitel werden einige vor allem deskriptive Statistiken zu den akustisch-prosodischen Eigenschaften der Daten im Lyrik-Korpus vorgestellt und diskutiert. Es werden dabei insbesondere Dauer- bzw. Rhythmusphänomene sowie grundfrequenzbasierte Ergebnisse dargelegt. Der prosodische Parameter der Intensität wird an dieser Stelle als zweitrangig betrachtet. Die Statistiken sind mit der Software SPSS<sup>10</sup> berechnet worden. Alle Tabellen und Grafiken entstammen diesem Programm. Die erörterten Statistiken bilden die Grundlage für die in Kapitel 7 diskutierten Prosodiemodelle.

### 6.1 Dauerphänomene

Der prosodische Parameter der Dauer ist der für die Rhythmizität gesprochener Sprache wohl bedeutendste (s. Kap. 2). Dauerphänomene erstrecken sich über verschiedene linguistisch-phonetische Ebenen. Von Bedeutung sind dabei hierarchisch geordnet das Phon, die Silbe, der Akzentfuß, die Phrase (Intermediär- und Intonationsphrase) sowie für Gedichte auch die Verszeile. Der Akzentfuß kann auf verschiedene Art und Weise gemessen werden. Das Interakzentintervall wird in der Regel von Vokalonset zu Vokalonset gemessen. Im Rahmen dieser Arbeit aber ist die Silbe die grundlegende prosodische Einheit und ein Fuß umfasst somit stets ganze Silben. Wenn der Fuß erwähnt wird, ist immer eine Einheit von einer akzentuierten Silbe zuzüglich der folgenden bzw. vorangehenden unakzentuierten Silben bis zum nächsten Akzent gemeint. Ob die unakzentuierten Silben dabei im Vor- oder Nachlauf stehen, hängt vom jeweiligen Metrum ab. Der Fuß bei Trochäus, Daktylus und Liedform beginnt mit der akzentuierten Silbe, der Fuß beim Jambus hingegen endet mit einer akzentuierten Silbe. Der Akzentfuß stellt die zentrale Einheit rhythmischer Betrachtungen im Deutschen dar. Wie in Kapitel 2 bereits erläutert, ging die frühe Forschung auf dem Gebiet des Sprechrhythmus davon aus, dass sich die Sprachen der Welt im Wesentlichen in zwei Klassen einordnen lassen, nämlich akzentzählende und silbenzählende Sprachen. Nach der Isochroniehypothese, die in diesem Zusammenhang entstand, sollte die Gruppe der akzentzählenden Sprachen isochrone Fußdauern besitzen. Das Deutsche gehört in eben diese

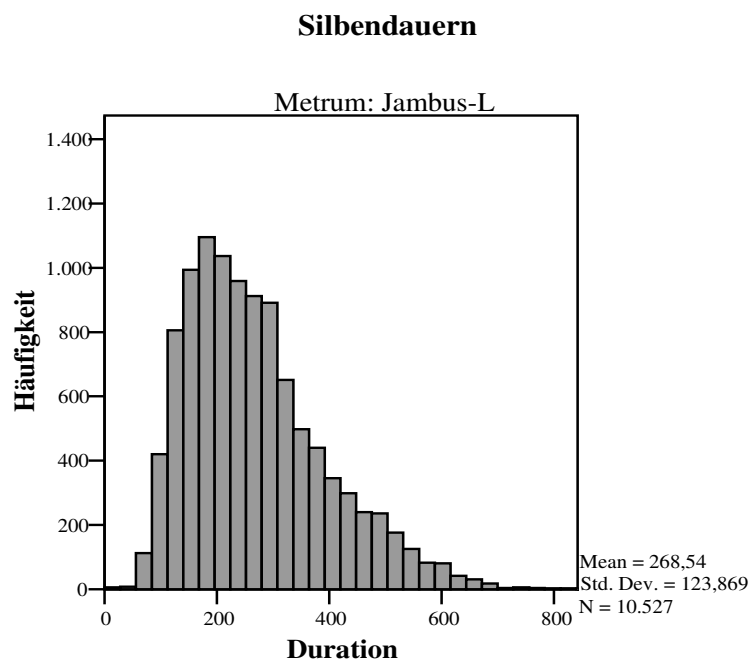
---

<sup>10</sup> Nähere Informationen zu SPSS sind unter der Adresse <http://www.spss.com/de/> zu finden.

Gruppe akzentzählender Sprachen. Auf der sprachproduktiven Seite kann diese Isochroniehypothese lediglich als Tendenz geltend gemacht werden (vgl. Kap. 2). Welche Rolle die Isochronie jedoch perzeptiv spielt, ist noch unklar. Unter anderen soll dieser Frage in den Perzeptionsexperimenten der vorliegenden Arbeit (s. Kap. 8) nachgegangen werden.

### 6.1.1 Silbendauern

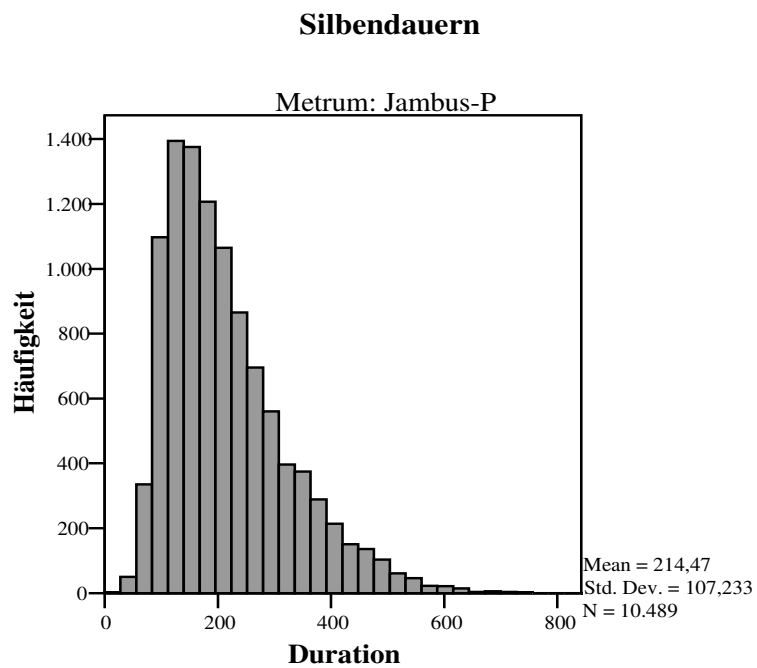
In Tabelle 6.1 können Mittelwert, Median und Standardabweichung der Silbendauern im Korpus aufgeschlüsselt nach Metrum und Sprechstil (lyrisch oder prosaisch) eingesehen werden. Es fällt auf, dass die Dauermediane der lyrischen Varianten alle klar über 200 ms liegen, die prosaischen dagegen alle deutlich unter 200 ms. Im Bereich der lyrischen Varianten hat der Jambus mit 246 ms den größten Dauermedian und die Liedform mit 218 ms den kleinsten Median. Im Bereich der prosaischen Silbendauern hat ebenfalls der Jambus mit 190 ms den größten Median und die Liedform mit 181 ms ebenso den kleinsten Dauermedian. Die Sprecher produzierten also die Silben des lyrischen Stils im Schnitt mit wesentlich größerer Dauer als die des prosaischen Stils.



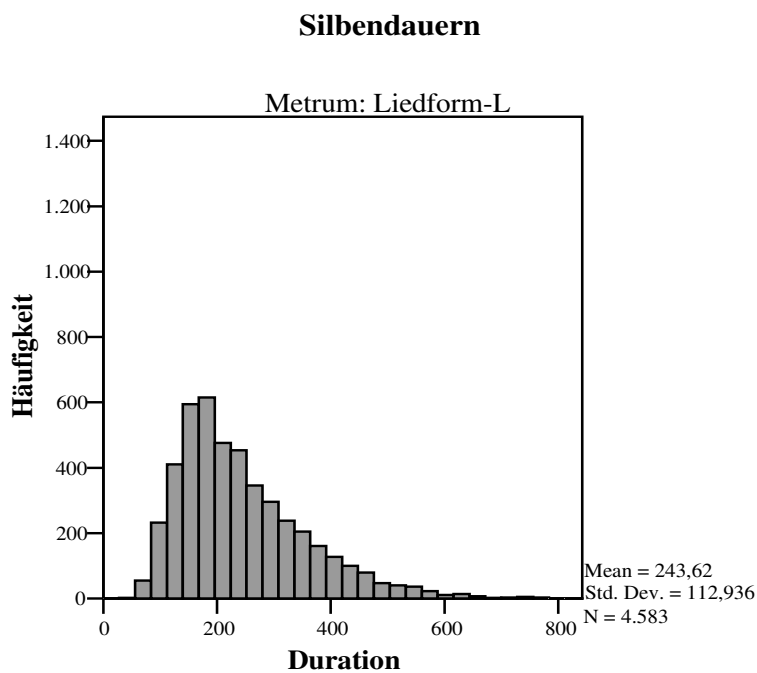
**Abb. 6.1:** Histogramm der Häufigkeit von Silbendauern (ms) im Korpus für Jambus, Lyrik; x-Achse: Korpusvariable *Duration*

Daktylus-L	N	6068
	Mittelwert	268,75
	Median	242,00
	Standardabweichung	123,229
Daktylus-P	N	6019
	Mittelwert	212,06
	Median	185,00
	Standardabweichung	110,193
Jambus-L	N	10527
	Mittelwert	268,54
	Median	246,00
	Standardabweichung	123,869
Jambus-P	N	10489
	Mittelwert	214,47
	Median	190,00
	Standardabweichung	107,233
Liedform-L	N	4583
	Mittelwert	243,62
	Median	218,00
	Standardabweichung	112,936
Liedform-P	N	4572
	Mittelwert	207,31
	Median	181,00
	Standardabweichung	104,147
Trochaeus-L	N	6177
	Mittelwert	257,62
	Median	240,00
	Standardabweichung	108,900
Trochaeus-P	N	6149
	Mittelwert	210,89
	Median	187,00
	Standardabweichung	103,272

**Tab. 6.1:** Mittelwert, Median und Standardabweichung der Silbendauern (in ms) im Korpus aufgeschlüsselt nach Metrum und Sprechstil (-L = Lyrik; -P = Prosa)

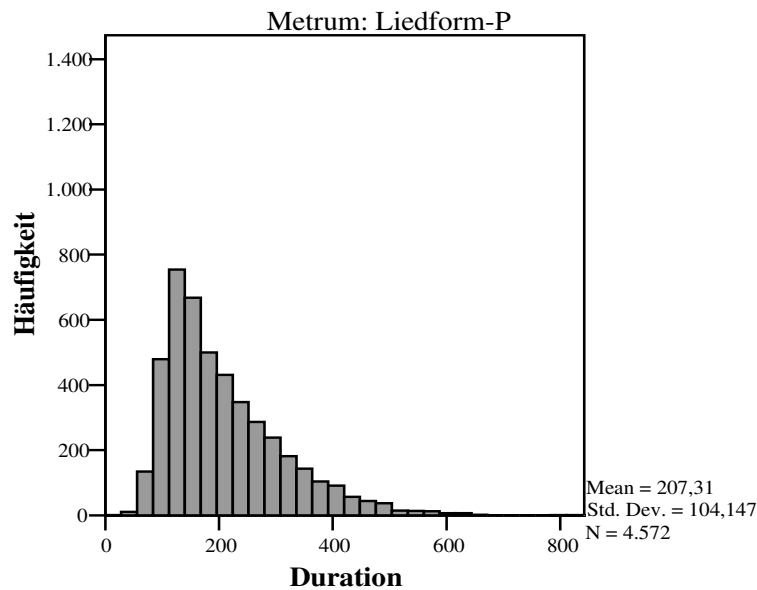


**Abb. 6.2:** Histogramm der Häufigkeit von Silbendauern (ms) im Korpus für Jambus, Prosa; x-Achse: Korpusvariable *Duration*



**Abb. 6.3:** Histogramm der Häufigkeit von Silbendauern (ms) im Korpus für Liedform, Lyrik; x-Achse: Korpusvariable *Duration*

## Silbendauern



**Abb. 6.4:** Histogramm der Häufigkeit von Silbendauern (ms) im Korpus für Liedform, Prosa; x-Achse: Korpusvariable *Duration*

Gleichzeitig hat in beiden Stilen der Jambus die größte Dauer und die Liedform die geringste Dauer. Daraus folgt, dass der lyrische Sprechstil sich durch eine deutlich geringere Sprechgeschwindigkeit auszeichnet. Überdies kann man aus der geringen Silbendauer der Silben in der Liedform schließen, dass aufgrund der im Schnitt größeren Zahl an Silben pro Fuß eine Dauerkompression stattfindet. Das bestätigt die in Kapitel 2 angesprochene Tendenz zur Isochronie im Deutschen.

Die Abbildungen 6.1 und 6.2 zeigen die Histogramme der Häufigkeit von Silbendauern für den Jambus im lyrischen sowie prosaischen Stil. Analog dazu zeigen die Abbildungen 6.3 und 6.4 die entsprechenden Histogramme für die Liedform im lyrischen und prosaischen Stil. Wie schon gezeigt wurde, bilden die Dauern dieser beiden Metren die Extreme im Korpus. Deshalb soll die Verteilung der Silbendauern dieser Fälle noch näher erörtert werden. Es fallen hierbei vor allem zwei Dinge auf. Zum einen ist die Streuung bei den Lyrikvarianten deutlich größer als bei den Fällen der Prosa. Zum anderen zeigt sich eine steile Flanke hin zu den geringeren Dauern. Die Flanke hin zu den größeren Dauern dagegen ist deutlich flacher. Das heißt, es gibt so etwas wie eine dem Median nahe liegende, strikte Untergrenze für Silbendauern. Dieser Umstand sollte durch artikulatorische Beschränkungen bedingt sein.

Eine beliebige Kompression der Silbe und ihrer Teilphone ist nicht möglich. Nach oben hin scheint die Skala der Silbendauern aber nahezu offen zu sein.

### 6.1.2 Fußdauern

Wie oben erwähnt, ist der Fuß für akzentzählende Sprachen die wichtigste phonetische Einheit im Bereich der Analyse des Sprechrhythmus. Tabelle 6.2 führt den Mittelwert, Median und die Standardabweichung der Fußdauern im Korpus aufgeschlüsselt nach Metrum und Sprechstil auf. Zwischen den Dauermedianen der Lyrik- und Prosavarianten eines Metrums liegen jeweils etwa 100 ms. Geht man davon aus, dass sich die Fußdauern an einem Referenzquantum (s. Kap. 2) orientieren, so muss man damit schlussfolgern, dass es wohl bei einer derartig großen Diskrepanz zwischen den Fußdauermedianen verschiedene Referenzquanten für den lyrischen Sprechstil auf der einen Seite und den prosaischen Sprechstil auf der anderen Seite geben muss. In beiden Fällen nämlich sind die Sprecher instruiert worden, in mittlerem, für sie angenehmem Tempo zu lesen. Die Sprecher deuten demnach, je nach lyrischem oder prosaischem Vortrag, den Begriff des mittleren Tempos ganz unterschiedlich. Der Begriff des Referenzquantums wird im Folgenden noch einmal aufgegriffen und näher erörtert.

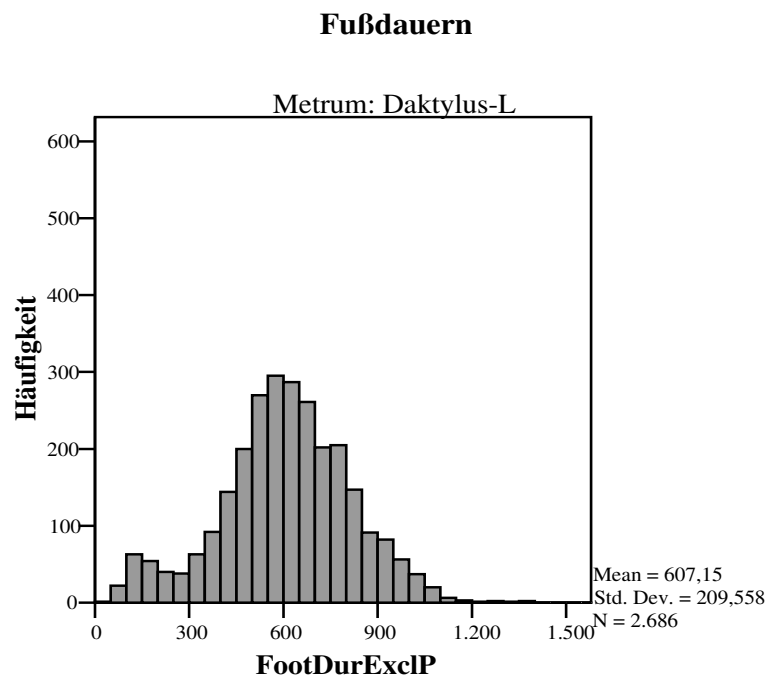
Innerhalb der Gruppe der lyrischen Vorträge besitzt der Daktylus mit 610 ms den größten Fußdauermedian und der Trochäus mit 474 ms den kleinsten Median. Innerhalb der Gruppe der prosaischen Vorträge besitzt ebenfalls der Daktylus mit 460 ms den größten Dauermedian und der Trochäus mit 378,5 ms den geringsten Median. Bemerkenswert ist an dieser Stelle, dass die Liedform im lyrischen Fall zwar den geringsten Silbendauermedian, zugleich aber den zweitgrößten Fußdauermedian hat. Es zeigt sich also, dass Dauerkompression stattfindet, sie aber nicht dazu führt, dass etwa das Referenzquantum deutlich unterschritten wird. Als eben dieses Referenzquantum kann man den Fußdauermedian gemittelt über alle lyrischen bzw. prosaischen Fußdauern anführen. Das arithmetische Mittel wird an dieser Stelle wegen seiner starken Beeinflusstheit durch Ausreißer nicht berücksichtigt. Für das Quantum der lyrischen Vorträge ergibt sich ein Median von 507 ms, für die prosaischen Vorträge dagegen ein Median von 408 ms. Damit liegt der Median im Lyrikfall um etwa 25% über dem Prosafall.

Daktylus-L	N	2686
	Mittelwert	607,15
	Median	610,00
	Standardabweichung	209,558
Daktylus-P	N	2792
	Mittelwert	457,15
	Median	460,00
	Standardabweichung	186,030
Jambus-L	N	5546
	Mittelwert	509,37
	Median	482,00
	Standardabweichung	194,867
Jambus-P	N	5623
	Mittelwert	399,69
	Median	386,00
	Standardabweichung	160,016
Liedform-L	N	2069
	Mittelwert	539,65
	Median	530,00
	Standardabweichung	204,896
Liedform-P	N	2079
	Mittelwert	455,19
	Median	450,00
	Standardabweichung	183,969
Trochäus-L	N	3229
	Mittelwert	491,89
	Median	474,00
	Standardabweichung	140,692
Trochäus-P	N	3278
	Mittelwert	395,55
	Median	378,50
	Standardabweichung	143,787

**Tab. 6.2:** Mittelwert, Median und Standardabweichung der Fußdauern (in ms) im Korpus aufgeschlüsselt nach Metrum und Sprechstil (-L = Lyrik; -P = Prosa)

Angesichts dieser signifikanten Diskrepanz muss man wohl von zwei sprechstilabhängig unterschiedlichen Quanten ausgehen. Des Weiteren sei angemerkt, dass Fant und Kruckenberg (1996) für das Schwedische ebenfalls ein Referenzquantum von 500 ms (s. Kap. 2) ermittelten, wenngleich sie prosaisches Material zugrunde legten. Möglicherweise handelt es sich bei dem Referenzquantum um eine Universalie für die akzentzählenden Sprachen. Dieser Umstand kann in diesem Rahmen aber nur vermutet werden. Überdies gehen Fant und Kruckenberg (1996) in erster Linie davon aus, dass Pausendauern ganzzahlige Vielfache des Referenzquantums bilden. In welchem Bezug eine konkrete Fußdauer zu diesem Quantum

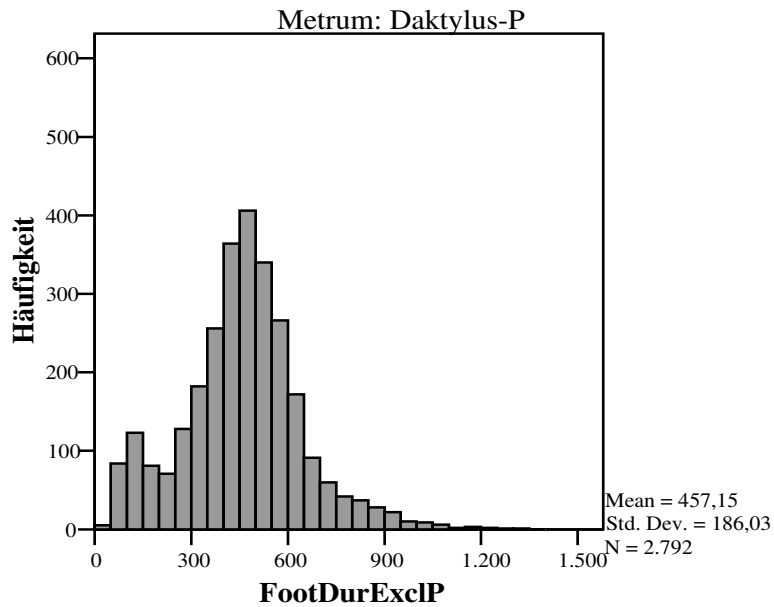
steht wird dabei nicht geklärt. Für die Daten des dieser Arbeit zugrunde liegenden Lyrik-Korpus konnte allerdings nicht festgestellt werden, dass sich Fälle bezüglich der Pausendauern bei ganzzahligen Vielfachen der Dauer des Referenzquantums kumulieren. Die große Diskrepanz der Fußdauermediane zwischen den einzelnen Metren deutet aber darauf hin, dass man in der Lyrik metrum-spezifische Referenzquanten für Fußdauern geltend machen sollte. Der Begriff des Referenzquantums wird damit erweitert. Das Quantum ist damit nicht nur Referenz für Pausendauern, sondern eine produktive Planungseinheit, an der sich die konkreten Fußrealisierungen orientieren. Es muss, zumindest gemäß den Daten des hier zugrunde liegenden Lyrik-Korpus, spezifisch für die verschiedenen Metren und Sprechstile gemessen werden (s. Mediane in Tab. 6.2).



**Abb. 6.5:** Histogramm der Häufigkeit von Fußdauern (ms) im Korpus für Daktylus, Lyrik; x-Achse: Korpusvariable *FootDurExclP*

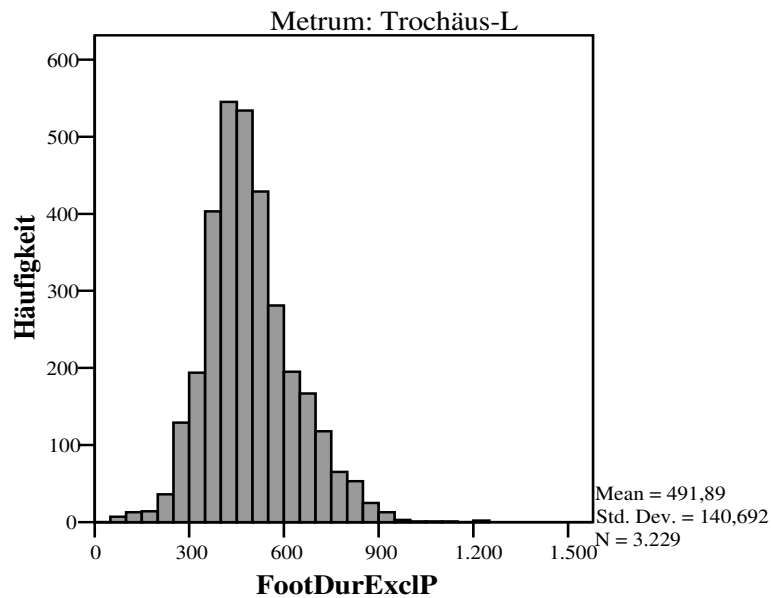


## Fußdauern



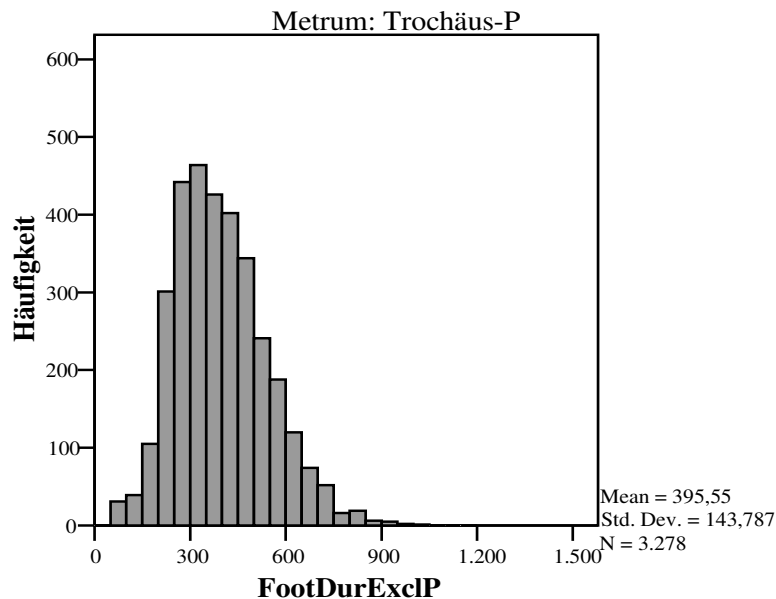
**Abb. 6.6:** Histogramm der Häufigkeit von Fußdauern (ms) im Korpus für Daktylus, Prosa; x-Achse: Korpusvariable *FootDurExclP*

## Fußdauern



**Abb. 6.7:** Histogramm der Häufigkeit von Fußdauern (ms) im Korpus für Trochäus, Lyrik; x-Achse: Korpusvariable *FootDurExclP*

## Fußdauern

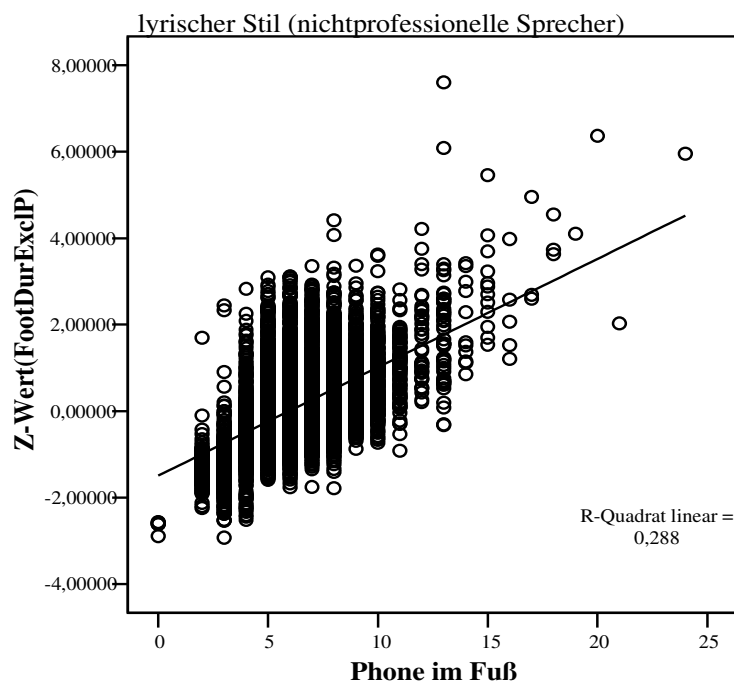


**Abb. 6.8:** Histogramm der Häufigkeit von Fußdauern (ms) im Korpus für Trochäus, Prosa; x-Achse: Korpusvariable *FootDurExclP*

Auch bezüglich der Akzentfußdauern sollen die oben erwähnten Extremfälle des Median tiefgreifender analysiert werden. Die Abbildungen 6.5 und 6.6 zeigen die Histogramme der Häufigkeit von Fußdauern für den Daktylus im lyrischen sowie prosaischen Stil. Analog dazu zeigen die Abbildungen 6.7 und 6.8 die entsprechenden Histogramme für den Trochäus im lyrischen und prosaischen Stil. Beim Daktylus fällt auf, dass sowohl die lyrische Vortragsvariante als auch der prosaische Stil zwei Maxima, ein kleines und ein großes, in der Häufigkeit der einzelnen Fußdauern aufweisen. Das ist eine Eigenschaft, die bei den anderen Metren, hier speziell dem Trochäus, nicht zu finden ist. Es sei an dieser Stelle jedoch nochmals angemerkt, dass die Daktylen leicht von Trochäen durchsetzt sind. Die Ergebnisse sind dadurch sicherlich leicht verzerrt. Außerdem sind die Flanken der Verteilung in allen vier Fällen, wie auch in den nicht dargestellten Fällen, zum oberen wie unteren Ende hin etwa gleich steil. Hier unterscheiden sich die Verteilungen von Silben- und Fußdauern. Für die Fußdauern gibt es keine so strikte Untergrenze wie für die Silbendauern. Darüber hinaus ist die Lage der schon geschilderten Dauermediane anhand der verschiedenen (Haupt-)Maxima gut nachvollziehbar.

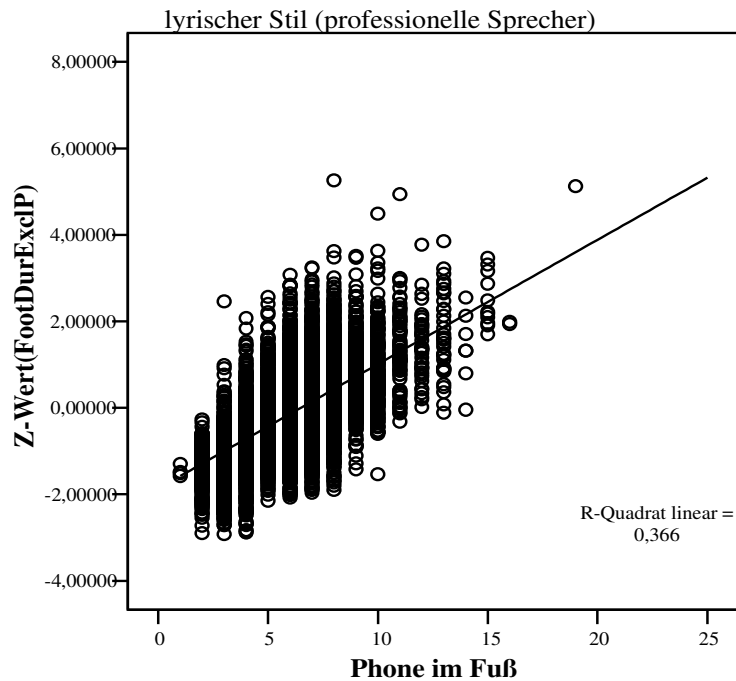
Interessant ist auch die Streuung der Fußdauern in Abhängigkeit von der Phonzahl im Fuß. Die entsprechenden Streudiagramme werden im Folgenden dokumentiert. Die Fußdauern sind dabei einer Z-Normalisierung<sup>11</sup> unterzogen worden. Die Grafiken beziehen sich damit auf Z-Werte. Bei der Z-Normalisierung ist der Faktor des Unterschiedes der Fußdauermittelwerte unter den einzelnen Metern korrigiert worden. Hierdurch werden die Streudiagramme zu den Fußdauern nun zwischen den einzelnen Metern vergleichbar. Die Variable *Meter* berücksichtigt implizit auch den Faktor des Sprechstils und fließt entsprechend in die Normalisierung ein. Jede Kategorie der Variablen *Meter* hat nun den Mittelwert 0 und die Standardabweichung 1. Dieser Umstand macht die Vergleichbarkeit aus.

In einem Streudiagramm ist für die Lage der Punktwolke bei linearen Zusammenhängen die lineare Regressionslinie charakteristisch. Bei einem Vergleich dieser Linien unter den einzelnen Kategorien von *Meter* zeigt sich ein einheitliches Bild. Der Anstieg der Fußdauern erfolgt mit gleicher Steilheit für nahezu alle Variablenausprägungen. Auffällig dabei ist nur der Trochäus im prosaischen Stil. Hier ist die Regressionslinie signifikant flacher als bei den anderen Vergleichsfällen. Es zeigen sich hier gleichzeitig deutlich weniger Ausreißer, die vermutlich mitverantwortlich sind für die Abweichung.

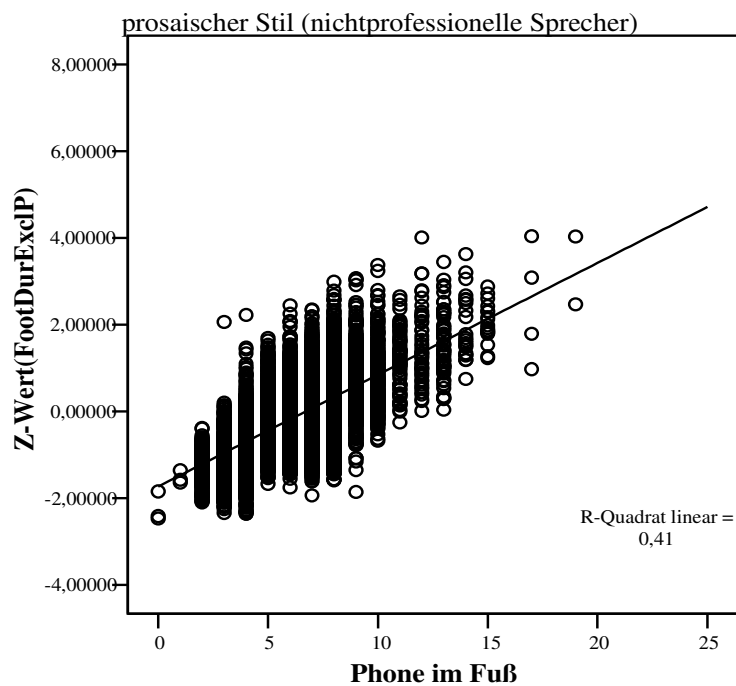


<sup>11</sup> Berechnungsformel für Z-normalisierte Werte:  $z=(x-M)/SD$  ; x=Ursprungswert, M=Mittelwert und SD=Streuung

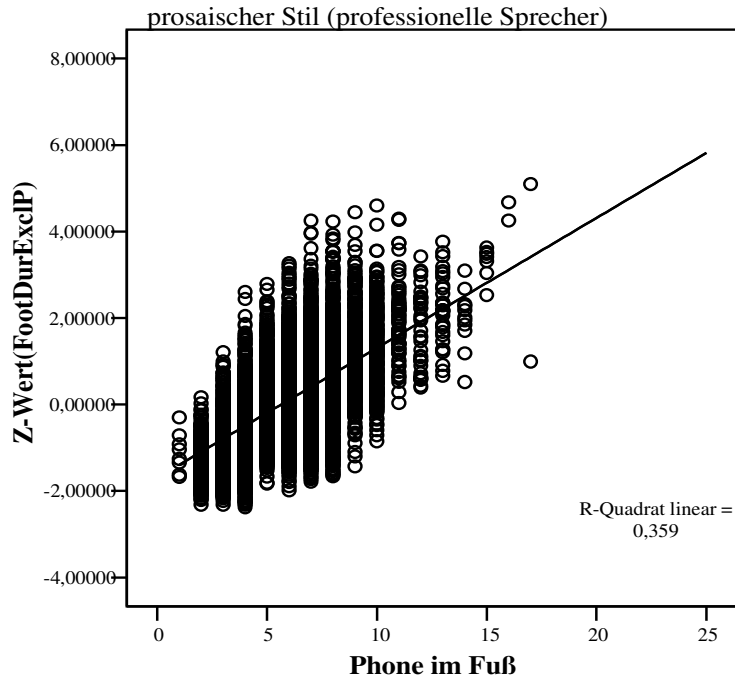
**Abb. 6.9:** Streudiagramm: Z-Normalisierte der Fußdauern in Abhängigkeit von der Phonzahl im Fuß für die nichtprofessionellen Sprecher und lyrischen Sprechstil samt Regressionslinie



**Abb. 6.10:** Streudiagramm: Z-Normalisierte der Fußdauern in Abhängigkeit von der Phonzahl im Fuß für professionelle Sprecher und lyrischen Sprechstil samt Regressionslinie



**Abb. 6.11:** Streudiagramm: Z-Normalisierte der Fußdauern in Abhängigkeit von der Phonzahl im Fuß für professionelle Sprecher und prosaischen Sprechstil samt Regressionslinie

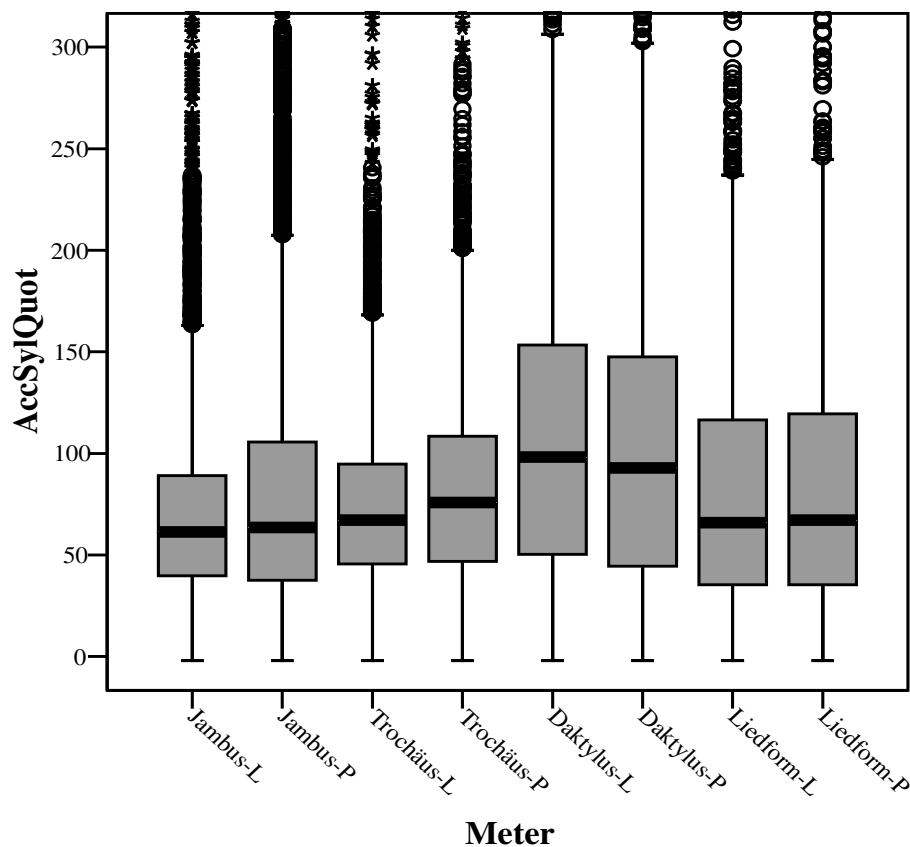


**Abb. 6.12:** Streudiagramm: Z-Normalisierte der Fußdauern in Abhängigkeit von der Phonzahl im Fuß für professionelle Sprecher und prosaischen Sprechstil samt Regressionslinie

Die Daten legen allerdings die Vermutung nahe, dass es einen Unterschied zwischen verschiedenen Sprechstilen und zwischen verschiedener Schulung der Sprecher gibt. Die klärenden Grafiken hierzu sind oben aufgeführt. Die Abbildung 6.9 zeigt das Streudiagramm von Z-Werten der Fußdauern in Abhängigkeit von der Phonzahl im Fuß für den lyrischen Sprechstil und die Gruppe der nichtprofessionellen Sprecher. Analog dazu zeigen die Abbildungen 6.10, 6.11 und 6.12 die Diagramme für lyrischen Sprechstil und die professionelle Sprechergruppe, prosaischen Sprechstil und die nichtprofessionellen Sprecher und schließlich prosaischen Sprechstil und die professionelle Gruppe von Sprechern. Die nichtprofessionelle Sprechergruppe weist in beiden Fällen eine sichtbar flachere Regressionslinie auf als die Daten der professionellen Sprecher. Damit ist gezeigt, dass die nichtprofessionellen Sprecher die Phone im Akzentfuß stärker komprimieren als die Profisprecher. Der bei den Laiensprechern häufig auftretende Stakkatostil wird also offenbar durch eine verstärkte Tendenz zur Isochronie hervorgerufen. Eine Sprechstilabhängigkeit der Verteilung der Punktwolke konnte dagegen nicht nachgewiesen werden.

### 6.1.3 Die Dauerrelation unakzentuiert zu akzentuiert

Für die in Kap.7 dargelegte Dauermodellierung wird unter anderem die Relation der Dauer des unakzentuierten Teils eines Fußes zum akzentuierten Teil eines Fußes berechnet und verwertet. Im Folgenden wird nun näher erläutert, warum dieser Quotient von großer Bedeutung für die Charakteristik der verschiedenen Metren ist. Abbildung 6.13 zeigt die Boxplots für den oben erwähnten Quotienten in Abhängigkeit von der Belegung der Variablen *Meter*. Es fällt zunächst einmal auf, dass es eine große Anzahl an Ausreißern gibt. Deshalb ist es wiederum sinnvoll, zum Vergleich der Durchschnittswerte den Median an Stelle des arithmetischen Mittels zu verwenden.



**Abb. 6.13:** Boxplots: Quotient aus den Dauern des unakzentuierten Fußbestandteils und dem akzentuierten Fußbestandteil (*AccSylQuot*) in Abhängigkeit vom Metrum (*Meter*) (-L = Lyrik, -P = Prosa)

Die schwarzen Horizontallinien kennzeichnen im Diagramm den Median. Dieser variiert zwischen den verschiedenen Metrumskategorien auf den ersten Blick nur leicht mit der Ausnahme des Daktylus. Der Median für den Daktylus weicht für beide Sprechstilvarianten deutlich sichtbar nach oben ab. Um die statistische Signifikanz dieser Differenzen zu prüfen,

wurde ein Mann-Whitney-Test durchgeführt. Es sind dabei alle möglichen Paare an Metrumskategorien im lyrischen Sprechstil miteinander verglichen worden. Alle Paarvergleiche bis auf einen Fall weisen dabei einen statistisch signifikanten Unterschied auf ( $p < 0,001$ ). Für die Differenz des Paares Trochäus und Liedform gilt die Ausnahme eines Mangels an statistischer Signifikanz ( $p = 0,8$ ). Die hohe Signifikanz der Unterschiede bezüglich der Variablen *AccSylQuot* zwischen den einzelnen Metren im lyrischen Sprechstil lässt darauf schließen, dass der Quotient zwischen der Dauer des unakzentuierten Teils eines Fußes und der Dauer des akzentuierten Teils eines Fußes charakteristisch für das jeweilige Metrum ist. Die Korpusvariable *AccSylQuot* ist demnach spezifisch für jedes Metrum im lyrischen Sprechstil. Darüber hinaus fällt auf, dass, beobachtbar anhand der Perzentile, die Werte für die Liedform und vor allem für den Daktylus wesentlich stärker streuen als die Werte für den Jambus und Trochäus. Dieser Effekt kann durch die variable Anzahl an Silben in den Füßen der Liedform und stellenweise auch des Daktylus erklärt werden. Es sei an dieser Stelle nochmals erwähnt, dass die daktylischen Gedichte im Korpus teilweise mit Trochäen durchmischt sind.

#### 6.1.4 Phrasen- und Versdauern

Die Phrasen- und Versdauern im Lyrik-Korpus weisen keine augenscheinlichen Spezifika für die einzelnen Metren auf und werden deshalb lediglich nach ihrem Sprechstil hin aufgeschlüsselt (s. Tabelle 6.3 und Tabelle 6.4). Der Median der Phrasendauern liegt im lyrischen Fall mit 2291 ms deutlich über dem Median für Prosa mit 1772 ms. Analog dazu verhalten sich die Versdauern, deren Median für die Lyrik bei 3168 ms und für die Prosa bei 2763 ms liegt. Verse sind damit im Schnitt etwa 400-500 ms länger als Phrasen. Interessant dabei ist vor allem, dass die lyrischen Versdauern im Mittel mit etwa 3s im Bereich der in Kapitel 2 erläuterten *action units* liegen. Damit zeigt sich die von Kien und Kemp (1994) erörterte Bedeutung der Verszeile als motorische Einheit. Auch Artikulation ist motorisches Verhalten. Die Daten des Lyrik-Korpus unterstreichen diesen Umstand.

Lyrik	N	3778
	Mittelwert	2384,61
	Median	2291,00
	Standardabweichung	1036,942
Prosa	N	4167

	Mittelwert	1865,98
	Median	1772,00
	Standardabweichung	800,618

**Tab. 6.3:** Mittelwert, Median und Standardabweichung der Phrasendauern (ms) im Korpus aufgeschlüsselt nach dem Sprechstil (Lyrik/Prosa)

Lyrik	N	2610
	Mittelwert	3451,75
	Median	3168,00
	Standardabweichung	1264,343
Prosa	N	2579
	Mittelwert	3014,94
	Median	2763,00
	Standardabweichung	1272,563

**Tab. 6.4:** Mittelwert, Median und Standardabweichung der Versdauern (ms) im Korpus aufgeschlüsselt nach dem Sprechstil (Lyrik/Prosa)

Letztere Autoren weisen auch auf die Bedeutung der Verszeile als semantische Einheit mit innerer Kohärenz hin. Über die semantische Funktion des Verses könnte hier aber nur spekuliert werden. Eine entsprechende Definition lässt sich nicht direkt aus den rhythmischen Gegebenheiten folgern. Es liegt jedoch nahe, dass rhythmische Handlung und die Bedeutung selbiger eng miteinander verknüpft sind und dass dieses von den Autoren auch berücksichtigt wurde. Diesem Zusammenhang müsste gesondert nachgegangen werden.

### 6.1.5 Zusammenfassung

Abschnitt 6.1 hat Dauerphänomene auf den Ebenen von Silbe, Fuß, Phrase bzw. Vers und der akzentuiert/nichtakzentuiert Relation (Korpusvariable *AccSylQuot*) diskutiert. Bei der Häufigkeitsverteilung von Silbendauern fiel besonders die steile Flanke hin zu niedrigen Dauern auf. Diese Flanke konnte bei den Fußdauern nicht beobachtet werden. Außerdem legten die Ergebnisse nahe, dass bezüglich der Fußdauern zwei unterschiedliche Referenzquanten als sprachproduktive Planungseinheiten einmal für den lyrischen Sprechstil und einmal für den prosaischen Sprechstil geltend gemacht werden können. Die Berechnung des jeweiligen Referenzquantums kann dabei sogar noch metrumsspezifisch verfeinert werden. Der Median der Dauerrelation zwischen dem nichtakzentuierten und dem akzentuierten Teil eines Fußes ergab, dass die Differenzen zwischen den einzelnen Metren,



respektive den Sprechstilen, bis auf eine Ausnahme statistisch höchst signifikant sind. Damit ist die Korpusvariable *AccSylQuot* charakteristisch für jedes der im Korpus abgebildeten Metren. Darüber hinaus zeigte die Ausmessung der Phrasen- und Versdauern insbesondere für den lyrischen Sprechstil, dass selbige durchaus in der Größenordnung der erläuterten *action units* liegen. Lyrik und Handlung stehen damit in engem Zusammenhang zueinander.

## 6.2 Intonationsphänomene

Wie in Kapitel 2 erläutert spielt für den Sprechrhythmus und seine Perzeption sehr wahrscheinlich auch die Intonation eine Rolle. Diese Vermutung wird in der diskutierten Literatur immer wieder geäußert. Welche Rolle genau sie dabei spielt, wird nicht vollends geklärt. Auch in dieser Arbeit kann dieses Problem wegen der Konzentration auf den Dauerparameter nicht gelöst werden. Es können wohl aber einige Phänomene der Intonation im Lyrik-Korpus aufgezeigt werden. Die an dieser Stelle dargelegten Analysen basieren auf Grundfrequenzwerten der Silben. Die Silbe ist also die kleinste Einheit über die sich Grundfrequenzmessungen im Korpus erstrecken. Die zentrale Variable dabei ist die mittlere Grundfrequenz der jeweiligen Silben (*F0Mean*). Zudem wird der GToBI-Formalismus zur phonologischen Etikettierung von Grundfrequenzverläufen verwendet.

### 6.2.1 Höhe der Grundfrequenz

In diesem Abschnitt werden die mittleren Grundfrequenzwerte (*F0Mean*) gemessen in Hz erörtert. In Tabelle 6.5 sind der Mittelwert und die Standardabweichung der Variablen

	Sprecher	N	Mittelwert	Standardabweichung
<i>F0Mean</i>	SP7	4812	163,34	52,866
	SP1	4706	191,17	50,234
	SP8	4739	92,97	24,988
	SP2	4610	197,45	36,875
	SP9	4831	211,20	60,919
	SP10	4769	100,74	32,321
	SP3	4711	120,69	38,958
	SP11	4742	191,10	52,612

	SP4	4598	171,94	34,737
	SP12	4751	108,18	37,159
	SP5	4683	160,06	42,846
	SP6	4621	116,95	25,229

**Tab. 6.5:** Mittelwert und Standardabweichung der Grundfrequenzwerte (Parameter *FOMean*; in Hz) im Korpus aufgeschlüsselt nach Sprechern

*FOMean* aufgeschlüsselt nach Sprechern verzeichnet. Die Standardabweichung variiert unter den Sprechern sehr stark. Das zeigt, dass die Intonation zwischen den Sprechern eine sehr unterschiedliche Modulationstiefe besitzt. Die größte Variation ist in den mittleren Grundfrequenzwerten zu finden. Die Variable *FOMean* ist dabei über alle Silben für den jeweiligen Sprecher gemittelt worden. Die Unterschiede sind zum einen selbstverständlich geschlechtsbedingt, zum anderen variiert die mittlere Grundfrequenz aber auch innerhalb der Sprechergruppe gleichen Geschlechts sehr stark. Aus diesem Grunde ist der Sprechereinfluss in den Daten durch eine Z-Normalisierung korrigiert worden. In Tabelle 6.6 sind das Minimum, das Maximum, der Mittelwert und die Standardabweichung der Z-Normalisierung aufgeschlüsselt nach den Kategorien der Metrumvariable *Meter* aufgelistet.

	Meter	N	Mittelwert	Standardabweichung
Z-Wert( <i>FOMean</i> )	Daktylus-L	6327	0,0090829	1,10430589
	Daktylus-P	6401	-0,0916785	1,14976100
	Jambus-L	10856	0,0196466	0,91747237
	Jambus-P	11084	-0,1044915	1,07740686
	Liedform-L	4630	0,2430432	0,91961158
	Liedform-P	4654	0,0150477	0,79582605
	Trochaeus-L	6268	0,1165512	0,94368368
	Trochaeus-P	6353	-0,0710842	0,91942464

**Tab. 6.6:** Mittelwert und Standardabweichung der Z-normalisierten Grundfrequenzwerte (Parameter *FOMean*) im Korpus aufgeschlüsselt nach Metren und (implizit) nach Sprechstil (-L = Lyrik; -P = Prosa)

	Style	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Z-Wert( <i>FOMean</i> )	Lyrik	28081	,0757304	,97237026	,00580264
	Prosa	28492	-,0746379	1,02085032	,00604784

**Tab. 6.7:** Mittelwert, Standardabweichung und Standardfehler der Z-normalisierten Grundfrequenzwerte im Korpus aufgeschlüsselt nach dem Sprechstil

Neben den stark unterschiedlichen Standardabweichungen fällt vor allem auf, dass die Kategorien sich bezüglich des jeweiligen Mittelwertes systematisch voneinander unterscheiden. Die Tabelle legt die Vermutung nahe, dass es einen sprechstilspezifischen Unterschied zwischen den einzelnen Mittelwerten gibt. Um dieser Vermutung nachzugehen, sind die in Tabelle 6.7 aufgelisteten Mittelwerte und Standardabweichungen der Z-normalisierten F0-Werte aufgeschlüsselt nach dem Sprechstil analysiert worden. Besonders augenscheinlich ist dabei der Unterschied zwischen dem Mittelwert für die Silben im lyrischen Stil und dem Mittelwert für die Silben im prosaischen Stil. Diese Differenz ist einem t-Test unterzogen worden. Dabei stellte sich eine statistische Signifikanz heraus ( $p < 0,001$ ). Daraus folgt, dass die Intonation im Durchschnitt bei einem lyrischen Vortrag signifikant höher liegt als bei einem prosaischen Vortrag. Gemäß den Daten im Lyrik-Korpus gilt, dass dieser stilistische Unterschied sogar bei gleichem segmentalen Gehalt der Sprachsignale zu finden ist. Sprecher verfügen damit über zwei verschiedene intonatorische Modi beim Textvortrag. Die Intonation eines dichterischen Vortrages wird signifikant höher angesetzt als die Intonation eines prosaischen Vortrages. Die Standardabweichung der Z-Werte liegt für die Lyrikfälle etwas niedriger als für die Prosafälle. Der Eindruck einer stärkeren intonatorischen Modulation bei einem lyrischen Vortrag schlägt sich damit nicht in der Standardabweichung der Z-normalisierten Grundfrequenzwerte nieder.

### **6.2.2 Phonologische Beschreibung der Intonation**

In diesem Abschnitt werden die Intonationskonturen der lyrischen Vorträge phonologisch beschrieben. Hierzu wird der in Kapitel 5 erläuterte GToBI-Formalismus verwendet. Es sollen somit die salienten Ereignisse der Intonationsverläufe im Korpus herausgestrichen werden. Es werden dabei Muster erwartet, die die einzelnen Metren charakterisieren und voneinander abgrenzen. Die Betrachtung der Häufigkeiten der einzelnen GToBI-Akzente legt nahe, die Verteilung vom Metrum und von der Position im Vers abhängig zu machen.

Die Tabellen 6.8, 6.9, 6.10 und 6.11 zeigen die Verteilung der GToBI-Akzente in Abhängigkeit vom jeweiligen Metrum in lyrischem Sprechstil. Für die Metren Daktylus, Liedform und Trochäus ist der Akzent „H\*“ der am häufigsten vorkommende. Es fällt dabei

auf, dass dieses gerade die Metren mit der akzentuierten Silbe zu Beginn des Versfußes sind. Der Jambus dagegen hat die akzentuierte Silbe am Ende des entsprechenden Versfußes. Er wird in den meisten Fällen mit einem „L+H\*“ Akzent belegt. Es sei aber auch angemerkt,

**Variable ToBIAccent für Daktylus (Lyrik)**

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
H*	493	25,3	25,3	25,3
L+H*	464	23,8	23,8	49,0
L*+H	457	23,4	23,4	72,4
L*	226	11,6	11,6	84,0
H+L*	162	8,3	8,3	92,3

**Tab. 6.8:** Häufigkeit der einzelnen GToBI-Akzente für den Daktylus im lyrischen Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

**Variable ToBIAccent für Jambus (Lyrik)**

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L+H*	1086	37,3	37,3	37,3
H*	773	26,5	26,5	63,8
L*+H	659	22,6	22,6	86,4
L*	94	3,2	3,2	89,6
H+L*	75	2,6	2,6	92,2

**Tab. 6.9:** Häufigkeit der einzelnen GToBI-Akzente für den Jambus im lyrischen Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

**Variable ToBIAccent für Liedform (Liedform)**

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
H*	423	32,6	32,6	32,6
L+H*	306	23,6	23,6	56,2
L*+H	268	20,7	20,7	76,9
L*	182	14,0	14,0	90,9
H+L*	47	3,6	3,6	94,5

**Tab. 6.10:** Häufigkeit der einzelnen GToBI-Akzente für die Liedform im lyrischen Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

**Variable ToBIAccent für Trochäus (Lyrik)**

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
H*	504	32,7	32,7	32,7
L+H*	364	23,7	23,7	56,4
L*+H	295	19,2	19,2	75,6

L*	216	14,0	14,0	89,6
H+L*	67	4,4	4,4	94,0

**Tab. 6.11:** Häufigkeit der einzelnen GToBI-Akzente für den Trochäus im lyrischen Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

dass der „L+H\*“ Akzent beim Daktylus fast genauso häufig auftritt wie der H\* Akzent. Mit der bisherigen Beschreibung gelangt man noch nicht zu einer eindeutigen Charakterisierung der Metren durch GToBI-Etikette. Um dieses zu erreichen wird zusätzlich die Abhängigkeit zwischen der Häufigkeit von GToBI-Akzenten und der Fußposition im Vers geprüft. Um diesen Zusammenhang darzustellen, sind beispielhaft die ersten vier Fußpositionen für den lyrischen Sprechstil in den Tabellen 6.12, 6.13, 6.14 und 6.15 aufgeführt.

*ToBIAccent für FootInVerse = 1*

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L+H*	954	43,4	43,4	43,4
H*	756	34,4	34,4	77,9
L*+H	354	16,1	16,1	94,0
L*	70	3,2	3,2	97,2
H+L*	19	,9	,9	98,0

**Tab. 6.12:** Häufigkeit der einzelnen GToBI-Akzente für *FootInVerse* = 1 bei lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBIAccent für FootInVerse = 2*

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L+H*	502	27,1	27,1	27,1
L*+H	454	24,6	24,6	51,7
H*	445	24,1	24,1	75,8
L*	183	9,9	9,9	85,7
!H*	67	3,6	3,6	89,3

**Tab. 6.13:** Häufigkeit der einzelnen GToBI-Akzente für *FootInVerse* = 2 bei lyrischen Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBI*Accent für *FootInVerse* = 3

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
H*	390	26,1	26,1	26,1
L*+H	370	24,7	24,7	50,8
L+H*	316	21,1	21,1	71,9
L*	194	13,0	13,0	84,9
H+L*	92	6,1	6,1	91,0

**Tab. 6.14:** Häufigkeit der einzelnen G*ToBI*-Akzente für *FootInVerse* = 3 bei lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBI*Accent für *FootInVerse* = 4

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
H*	273	28,8	28,8	28,8
L*+H	226	23,9	23,9	52,7
L+H*	216	22,8	22,8	75,5
L*	119	12,6	12,6	88,1
H+L*	38	4,0	4,0	92,1

**Tab. 6.15:** Häufigkeit der einzelnen G*ToBI*-Akzente für *FootInVerse* = 4 bei lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

An den ersten zwei Positionen ist es der „L+H\*“ Akzent, der am häufigsten vorkommt. Die folgenden zwei Positionen allerdings werden durch den „H\*“ Akzent dominiert. Der „H\*“ Akzent herrscht übrigens auch in den hier nicht tabellarisch dargestellten Fußpositionen vor. Somit ergibt sich die Konstellation, dass sowohl metrumabhängig als auch positionsabhängig die Akzentgipfel „H\*“ und „L+H\*“ vorherrschen. In Kapitel 7 wird trotzdem der Versuch unternommen, beide Einflussfaktoren zu kombinieren und so zu einer eindeutigen Unterscheidbarkeit der vier berücksichtigten Metren zu gelangen.

Außer den Tonakzenten im dieser Arbeit zugrunde liegenden Lyrik-Korpus sind zusätzlich die Grenztöne zu analysieren. Die Grenztöne im Korpus sind in der Variablen *ToBIBorder* kodiert. Die Tabellen 6.16, 6.17, 6.18 und 6.19 geben Aufschluss über die am häufigsten auftretenden Grenztöne im Lyrik-Korpus.

*ToBIBorder* für Jambus (Lyrik)

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L-%	438	35,6	35,6	35,6

L-H%	244	19,8	19,8	55,4
H-%	231	18,8	18,8	74,2
H-	165	13,4	13,4	87,6
L-	81	6,6	6,6	94,2

**Tab. 6.16:** Häufigkeit der einzelnen GToBI-Grenztöne für den Jambus in lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBIBorder für Trochäus (Lyrik)*

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L-%	315	38,5	38,5	38,5
H-%	194	23,7	23,7	62,1
L-H%	181	22,1	22,1	84,2
H-	75	9,2	9,2	93,4
L-	26	3,2	3,2	96,6

**Tab. 6.17:** Häufigkeit der einzelnen GToBI-Grenztöne für den Trochäus in lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBIBorder für Daktylus (Lyrik)*

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L-%	320	48,8	48,8	48,8
H-%	132	20,1	20,1	68,9
H-	91	13,9	13,9	82,8
L-H%	64	9,8	9,8	92,5
H-^H%	20	3,0	3,0	95,6

**Tab. 6.18:** Häufigkeit der einzelnen GToBI-Grenztöne für den Daktylus in lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

*ToBIBorder für Liedform (Lyrik)*

Akzent	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
L-%	219	40,6	40,6	40,6
H-%	153	28,4	28,4	69,0
H-	71	13,2	13,2	82,2
L-H%	52	9,6	9,6	91,8
L-	31	5,8	5,8	97,6

**Tab. 6.19:** Häufigkeit der einzelnen GToBI-Grenztöne für die Liedform in lyrischem Sprechstil (Ausreißer sind der Übersichtlichkeit wegen ausgeschlossen)

Es treten dabei entweder die mit einem „-“ endenden Intermediärphrasengrenzen oder die mit einem „%“ endenden großen Phrasengrenzen auf. Für alle vier in der Analyse berücksichtigten Metren ist der „L-%“ Grenzton der mit Abstand häufigste. Seine prozentuale

Häufigkeit liegt für den Jambus bei 35,8 %, für den Trochäus bei 38,5 %, für den Daktylus bei 48,8 % und für die Liedform bei 40,6 %. Es besteht also keine Metrumabhängigkeit für Grenztöne. Der Ton L-% kann als Standardgrenzton für jedes der vier Metren betrachtet werden. Damit wird auch klar, dass die meisten lyrischen Phrasen, genauso wie prosaische Phrasen, mit einem fallenden Ton enden<sup>12</sup>. Die hohen Grenztöne, insbesondere der „H-%“ Ton, kommen allerdings in der Lyrik wesentlich häufiger vor als in den prosaischen Phrasen. Die prosaischen Phrasen enden nämlich zu etwa 70-80% mit dem „L-%“ Grenzton. Es ist demnach anzunehmen, dass es vor allem die Markierung von Progredienz ist, die den größeren Anteil an „H-%“ Grenztönen bei Phrasen im lyrischen Sprechstil ausmacht.

### 6.2.3 Zusammenfassung

Der Abschnitt 6.2 hat sich mit den Grundfrequenzphänomenen im Lyrik-Korpus befasst. Es zeigte sich dabei ein sprecher- und geschlechtsabhängiger Effekt bezüglich der mittleren Tonhöhe von Silben im Korpus. Diese Effekte wurden durch eine Z-Normalisierung beseitigt, so dass die einzelnen Sprecher vergleichbar wurden. Es konnte bei der anschließenden Analyse vor allem verzeichnet werden, dass die durchschnittliche Grundfrequenz für die lyrischen Varianten höher liegt als für die prosaischen Varianten. Es existiert damit ein eigener lyrischer Vortragsmodus, dessen Intonation signifikant höher angesetzt ist als der des prosaischen Vortrages. Bei der phonologischen Beschreibung der Intonation im Korpus herrschten die Etiketten „H\*“ und „L+H\*“ vor. Dabei bot sich an, künftig (s. Kap. 7) eine Aufschlüsselung der Beschreibung nach Metrum und Fußposition im Vers vorzunehmen. Der häufigste Grenzton im Korpus ist „L-%“, d. h. es dominiert der phrasenfinal fallende Ton.

---

<sup>12</sup> Das kann zum Teil daran liegen, dass kaum Fragen in den Textvorlagen vorkommen.



## **7 Rhythmisch-prosodische Modelle**

In diesem Kapitel werden Modelle zur Vorhersage des Rhythmus von Versen im lyrischen Sprechstil beschrieben. Im Fokus dieser Darstellungen steht dabei die Dauermodellierung. Es wird aber darüber hinausgehend zusätzlich eine GToBI-Beschreibung von Versen angeführt. Die Dauermodellierung bedient sich zweier Methoden. Die erste Methode basiert auf einer metrumsspezifischen linearen Regression, die zweite hingegen fußt auf der Isochroniehypothese. Die Isochronie wird dabei als perzeptives Phänomen betrachtet, da sich ja, wie bereits erläutert, gezeigt hat, dass man auf der produktiven Seite lediglich eine Tendenz hin zur Isochronie nachweisen kann. Die in diesem Kapitel erläuterten Dauermodelle werden schließlich in Kapitel 8 einer Evaluation unterzogen. Ihre Vorhersageleistung soll in dem Rahmen geprüft werden.

### **7.1 Dauermodelle**

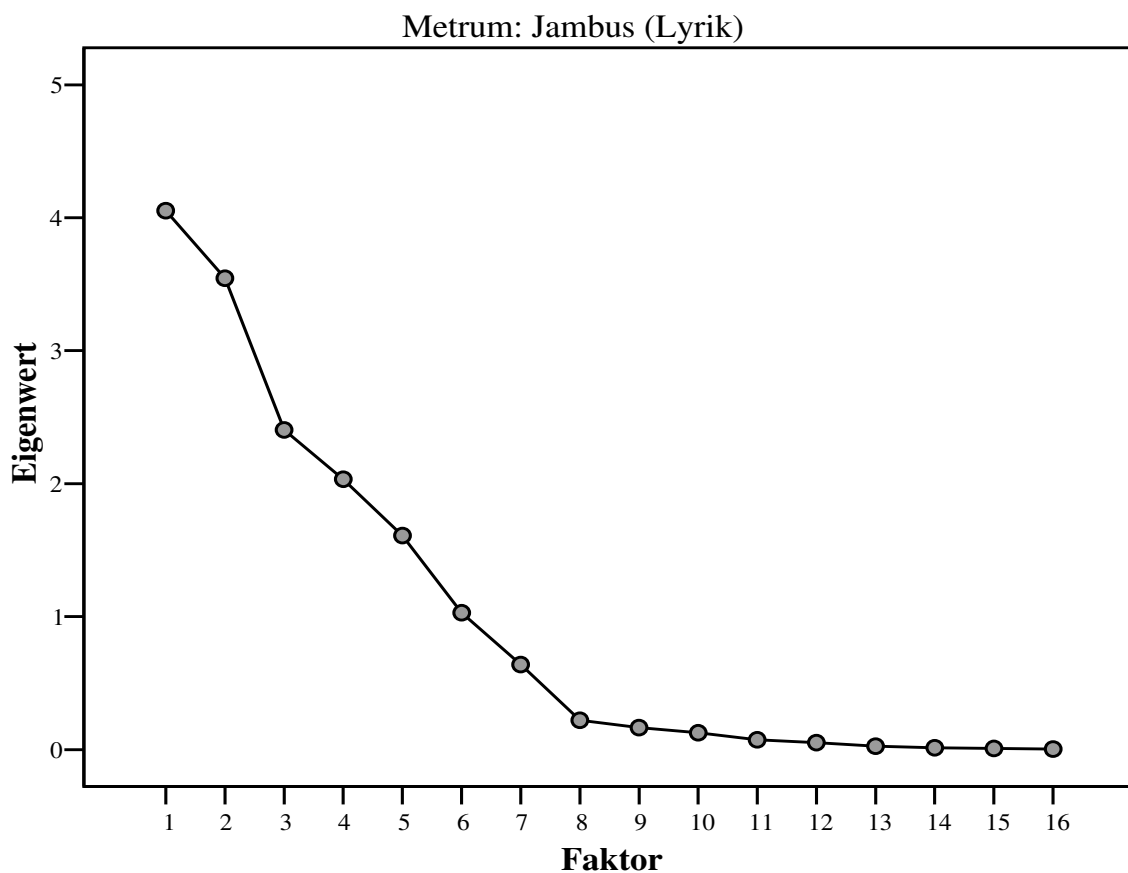
Aus Arbeiten zur Dauersteuerung in der Sprachsynthese ist bekannt, dass ein linearer Zusammenhang zwischen Silbendauer und den die selbige beeinflussenden Faktoren besteht (vgl. Portele & Meyer, 1994). Aus diesem Grund ist auch für die Dauermodellierung der im Zentrum dieser Arbeit stehenden Silbendauern bei lyrischem Sprechstil eine lineare Regression verwendet worden. Von den guten Resultaten in der Sprachsynthese kann hier sicherlich profitiert werden. Darüber hinaus ist auch die Isochroniehypothese als Ausgangsbasis zur Silbendauerprädiktion ins Auge gefasst worden. Es soll dabei geprüft werden, ob isochrone Fußdauern auf der perzeptiven Seite von Bedeutung sind (s. Kap. 8). Hierbei ist zu diskutieren, wie die jeweiligen Silbendauern aus einer vorgegebenen Fußdauer berechnet werden sollen.

#### **7.1.1 Lineare Regression**

Aufgrund der erfolgreichen Anwendung der linearen Regression zur Vorhersage von Silbendauern in der Sprachsynthese wird in dieser Arbeit vorausgesetzt, dass ein linearer Zusammenhang zwischen Silbendauern und den diese beeinflussenden Faktoren besteht. Das

größte Problem bei der Anwendung einer linearen Regression liegt dabei in der Dimensionsreduktion. Es sollen diejenigen Faktoren in das Modell einfließen, die den größten Teil der Varianz erklären und untereinander möglichst schwach korrelieren. Hierzu wird der eigentlichen Modellierung eine Faktorenanalyse vorangestellt. Beispielhaft wird in Abbildung

### Screepplot



**Abb. 7.1:** Screepplot: Eigenwerte in Abhängigkeit vom jeweiligen Faktor für den Jambus im lyrischen Sprechstil

7.1 der Screepplot für den Jambus im lyrischen Sprechstil dargestellt. Faktoren mit großem Einfluss auf die abhängige Variable besitzen einen hohen Eigenwert, Faktoren mit geringem Einfluss haben einen entsprechend niedrigen Eigenwert. Von Bedeutung sind vor allem die Faktoren, die auf der steilen Flanke des Graphen liegen. Die Faktoren im flachen Teil der Kurve hingegen können vernachlässigt werden. Die Abbildung 7.1 legt nahe, 7 Faktoren zu berücksichtigen. Dabei ist überprüft worden, welche Korpusvariablen mit den 7 Faktoren

jeweils am höchsten korrelieren. Dadurch sollten die einflussreichsten Variablen für eine lineare Regression ermittelt werden. Für den Jambus ergaben sich somit hierarchisch geordnet *SylNumInPhrase*, *SylNumInVerse*, *FootInVerse*, *SylNumInFoot*, *Acce*, *PhrasePosCat* und *PhonesInSyl*. Die Definitionen der Variablen können in Kapitel 5.3 bzw. Anhang A eingesehen werden. Leider ist die Anzahl von 7 Faktoren immer noch sehr hoch. Überdies sind die 7 verschiedenen Variablen unter den einzelnen Metren nicht dieselben.

Es wurde angestrebt, das Regressionsmodell so schlicht wie möglich zu halten. Deshalb ist die nun zu erörternde lineare Regression schrittweise ausgeführt worden. Es bildete sich dabei eine Hierarchie von Variablen heraus. Gleichzeitig konnte die Korrelation zwischen Regressionsergebnis und tatsächlichem Wert der abhängigen Variablen beobachtet werden. Bei dieser Analyse stellte sich heraus, dass die Korrelation nach Einschluss von 3 Variablen nicht mehr nennenswert anstieg. Diese drei Variablen sind nunmehr für Jambus, Trochäus, Daktylus und Liedform im lyrischen Sprechstil die gleichen. Es handelt sich dabei um die drei

**Modellzusammenfassung (Jambus - Lyrik)**

Speaker	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
SP7	,750(a)	,562	,560	91,728
SP1	,716(a)	,513	,511	72,587
SP8	,763(a)	,582	,580	74,581
SP2	,748(a)	,559	,558	65,963
SP9	,758(a)	,574	,572	84,006
SP10	,798(a)	,637	,635	81,904
SP3	,721(a)	,520	,518	83,439
SP11	,770(a)	,593	,592	64,365
SP4	,744(a)	,554	,552	58,913
SP12	,772(a)	,597	,595	78,478
SP5	,752(a)	,566	,564	60,970
SP6	,695(a)	,483	,481	66,104

a Einflussvariablen : (Konstante), Acce, PhonesInSyl, PhrasePosCat

**Tab. 7.1:** Korrelationskoeffizient R, R-Quadrat, Korrigiertes R-Quadrat und Standardfehler des Schätzers für das lineare Regressionsmodell der jambischen Daten bei lyrischem Sprechstil aufgeschlüsselt nach den Sprechern (anonymisierte Variable *Speaker*)

**Modellzusammenfassung (Trochäus – Lyrik)**

Speaker	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
SP7	,558(a)	,311	,307	82,658
SP1	,607(a)	,369	,365	66,905
SP8	,619(a)	,384	,380	77,256
SP2	,724(a)	,524	,521	53,912
SP9	,584(a)	,341	,337	76,166
SP10	,608(a)	,370	,366	90,620
SP3	,653(a)	,427	,423	92,275
SP11	,639(a)	,409	,405	59,915
SP4	,694(a)	,481	,478	52,342
SP12	,612(a)	,375	,371	72,889
SP5	,720(a)	,518	,515	52,663
SP6	,598(a)	,358	,354	60,983

a Einflussvariablen : (Konstante), Acce, PhonesInSyl, PhrasePosCat

**Tab. 7.2:** Korrelationskoeffizient R, R-Quadrat, Korrigiertes R-Quadrat und Standardfehler des Schätzers für das lineare Regressionsmodell der trochäischen Daten bei lyrischem Sprechstil aufgeschlüsselt nach den Sprechern (anonymisierte Variable *Speaker*)

**Modellzusammenfassung (Daktylus – Lyrik)**

Speaker	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
SP7	,700(a)	,491	,485	68,410
SP1	,578(a)	,335	,327	65,477
SP8	,713(b)	,508	,502	60,846
SP2	,469(b)	,220	,212	71,901
SP9	,741(a)	,549	,544	62,789
SP10	,686(b)	,470	,464	82,548
SP3	,652(a)	,425	,419	98,063
SP11	,730(b)	,533	,528	55,384
SP4	,601(b)	,361	,354	51,263
SP12	,689(b)	,475	,470	71,985
SP5	,699(a)	,488	,483	57,172
SP6	,585(b)	,343	,336	57,441

a Einflussvariablen : (Konstante), Acce, PhonesInSyl, PhrasePosCat

b Einflussvariablen : (Konstante), Acce, PhrasePosCat, PhonesInSyl

**Tab. 7.3:** Korrelationskoeffizient R, R-Quadrat, Korrigiertes R-Quadrat und Standardfehler des Schätzers für das lineare Regressionsmodell der daktylischen Daten bei lyrischem Sprechstil aufgeschlüsselt nach den Sprechern (anonymisierte Variable *Speaker*)

Modellzusammenfassung (Liedform – Lyrik)

Speaker	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
SP7	,691(a)	,478	,474	92,930
SP1	,725(a)	,526	,522	67,006
SP8	,709(a)	,502	,498	79,031
SP2	,676(a)	,457	,453	76,079
SP9	,759(a)	,575	,572	75,723
SP10	,735(a)	,541	,537	91,182
SP3	,679(a)	,460	,456	96,595
SP11	,754(a)	,569	,565	57,411
SP4	,702(a)	,493	,489	59,821
SP12	,769(a)	,591	,587	82,780
SP5	,795(a)	,633	,630	56,607
SP6	,716(a)	,513	,509	70,316

a Einflussvariablen : (Konstante), *Acce*, *PhrasePosCat*, *PhonesInSyl*

**Tab. 7.4:** Korrelationskoeffizient R, R-Quadrat, Korrigiertes R-Quadrat und Standardfehler des Schätzers für das lineare Regressionsmodell der Liedform-Daten bei lyrischem Sprechstil aufgeschlüsselt nach den Sprechern (anonymisierte Variable *Speaker*)

Variablen *Acce*, *PhonesInSyl* und *PhrasePosCat*. Letztere Variablendefinitionen sind ebenfalls in Kapitel 5.3 bzw. Anhang A gegeben. Die Tabellen 7.1, 7.2, 7.3 und 7.4 stellen die Korrelationskoeffizienten zwischen Regressionsergebnis und tatsächlichem Wert der abhängigen Variablen aufgeschlüsselt nach den verschiedenen Sprechern dar. Jedes der vier Metren bei lyrischem Sprechstil ist in einer gesonderten Tabelle aufgeführt. Der Korrelationskoeffizient zwischen Regressionsvorhersage und tatsächlich beobachtetem Wert für die Silbendauern der jambischen Daten bei lyrischem Sprechstil reicht von  $r = 0,695$  für Sprecher SP6 bis  $r = 0,798$  für Sprecher SP10. Bei ersterem Sprecher handelt es sich um einen nichtprofessionellen Sprecher, bei letzterem Sprecher dagegen um einen professionellen Sprecher. Die Extremwerte des Korrelationskoeffizienten liegen für den Trochäus im lyrischen Sprechstil zwischen  $r = 0,558$  für Sprecher SP7 (professionell) und  $r = 0,724$  für Sprecher SP2 (nichtprofessionell). Für den Daktylus im lyrischen Sprechstil bewegen sich die Extremwerte zwischen  $r = 0,469$  für Sprecher SP2 (nichtprofessionell) und  $r = 0,741$  für Sprecher SP9 (professionell). Die Extreme für die Liedform im lyrischen Sprechstil schließlich reichen von  $r = 0,676$  für Sprecher SP2 (nichtprofessionell) bis  $r = 0,795$  für Sprecher SP5 (nichtprofessionell). Damit zeigt sich, dass die Güte der Dauervorhersage in hohem Maße vom jeweiligen Sprecher abhängt. Es lassen sich sprecherspezifisch Korrelationen bis zu einem Wert von etwa  $r = 0,8$  erzielen. Diese

Korrelation ist als hoch zu bewerten. Die erklärte Varianz bei einem Koeffizienten von  $r = 0,8$  liegt bei 64%. Mit der linearen Regression auf Basis der unabhängigen Variablen *Acce*, *PhonesInSyl* und *PhrasePosCat* können also bei geeigneter Sprecherwahl für das Silbendauermodell etwa zwei Drittel der Varianz erklärt werden. Daraus folgt, dass die lineare Regression zur Modellierung von Silbendauern bei lyrischem Sprechstil sprecher- und metrumsspezifisch berechnet wird.

Die favorisierten Sprecher sind gemäß obiger Statistiken SP10 für den Jambus, SP2 für den Trochäus, SP9 für den Daktylus und schließlich SP5 für die Liedform. Es sind dabei zwei professionelle und zwei nichtprofessionelle Sprecher berücksichtigt. Zudem teilen sich die Sprecher in gleich viele Männer und Frauen. Dazu sei angemerkt, dass Geschlecht und Sprecherausbildung keinen Einfluss auf die Prädizierbarkeit der Daten haben. Die Regressionsfunktionen zur Berechnung von Silbendauern (ms) bei lyrischem Sprechstil sind im einzelnen:

1) Sprecher SP10; Jambus in lyrischem Stil:

$$\text{DUR} = 66,220 * \text{PhrasePosCat} + 73,429 * \text{PhonesInSyl} + 92,806 * \text{Acce} - 118,259$$

2) Sprecher SP2; Trochäus in lyrischem Stil:

$$\text{DUR} = 16,855 * \text{PhrasePosCat} + 29,995 * \text{PhonesInSyl} + 96,318 * \text{Acce} + 60,044$$

3) Sprecher SP9; Daktylus in lyrischem Stil:

$$\text{DUR} = -10,128 * \text{PhrasePosCat} + 55,868 * \text{PhonesInSyl} + 92,038 * \text{Acce} + 42,869$$

4) Sprecher SP5; Liedform in lyrischem Stil:

$$\text{DUR} = 34,613 * \text{PhrasePosCat} + 29,326 * \text{PhonesInSyl} + 120,324 * \text{Acce} + 25,059$$

Die nach obigen Formeln zu berechnenden Silbendauern können nun zur Dauermanipulation von Sprachsignalen verwendet werden. In Kapitel 8 wird eine solche Manipulation beschrieben. Des Weiteren wird dort die Vorhersageleistung des Silbendauermodells auf Basis einer linearen Regression mit drei unabhängigen Variablen evaluiert.

### 7.1.2 Silbendauern bei isochronen Fußdauern

Wie Kap. 2 gezeigt hat, ist die Isochroniehypothese in ihrer strengen Form abzulehnen. Dennoch zeigt sich im Deutschen eine Tendenz zur Isochronie, d.h. der annähernd konstanten Dauer zwischen zwei Akzenten. Dieses zeigt sich vor allem daran, dass eine Dauerkompression von Silben innerhalb eines Fußes stattfindet. In diesem Rahmen ist noch unklar, welche Bedeutung der Isochronie auf der perzeptiven Seite zukommt. In dieser Arbeit wird die Isochronie als ein abstraktes Planungskonzept für die Sprachproduktion verstanden. Allerdings nehmen weitere Prozesse der Sprechplanung zusätzlich Einfluss auf die konkreten Dauern von Akzentfüßen und Silben (vgl. Kohler, 1983). Somit sollten isochrone Akzentfüße vorteilhaft für die Wahrnehmung verschiedener Rhythmen sein. Das Prinzip der Isochronie sollte dem muttersprachlichen Hörer intuitiv bekannt sein.

In diesem Abschnitt werden für alle vier im Lyrik-Korpus berücksichtigten Rhythmen isochrone Fußdauern angesetzt. Unter Zuhilfenahme des Dauerquotienten zwischen dem unakzentuierten und akzentuierten Anteil eines Fußes (Korpusvariable *AccSylQuot*) werden so die konkreten Silbendauern berechnet. Dabei wird, wie schon im Fall der linearen Regression, eine sprecherspezifische Modellierung vorgenommen. Für die Fußdauer ist damit der Sprecher gewählt worden, dessen mittlerer Fußdauermedian für ein Metrum am nächsten am globalen mittleren Fußdauermedian für das jeweilige Metrum liegt. In die Funktion zur Berechnung der Silbendauern im Fuß fließt ebenfalls sprecherspezifisch die Korpusvariable *AccSylQuot* ein. Diese Variable entspricht dem Dauerquotienten zwischen unakzentuiertem und akzentuiertem Anteil im Fuß. Im Folgenden sind die Berechnungsvorschriften für die isochronen Fußdauern angeführt.  $x$  entspricht dabei der Dauer der akzentuierten Silbe,  $y$  der Dauer des unakzentuierten Anteils des Fußes:

1) Sprecher SP12; Jambus im lyrischen Sprechstil:

$$x + 0.617 * x = 494 \text{ ms}$$
$$\Leftrightarrow x = 306 \text{ ms}$$

$$y = 0,617 * x = 0.617 * 306 \text{ ms} = 189 \text{ ms}$$

2) Sprecher SP12; Trochäus im lyrischen Sprechstil:

$$x + 0.775 * x = 469 \text{ ms}$$

$$\Leftrightarrow x = 264 \text{ ms}$$

$$y = 0.775 * 264 \text{ ms} = 205 \text{ ms}$$

3) Sprecher SP10; Daktylus im lyrischen Sprechstil:

$$x + 1.273 * x = 684 \text{ ms}$$

$$\Leftrightarrow x = 301 \text{ ms}$$

$$y = 1.273 * 301 \text{ ms} = 383 \text{ ms}; y_1 = 192; y_2 = 192$$

4) Sprecher SP4; Liedform im lyrischen Sprechstil:

$$x + 0.723 * x = 534 \text{ ms}$$

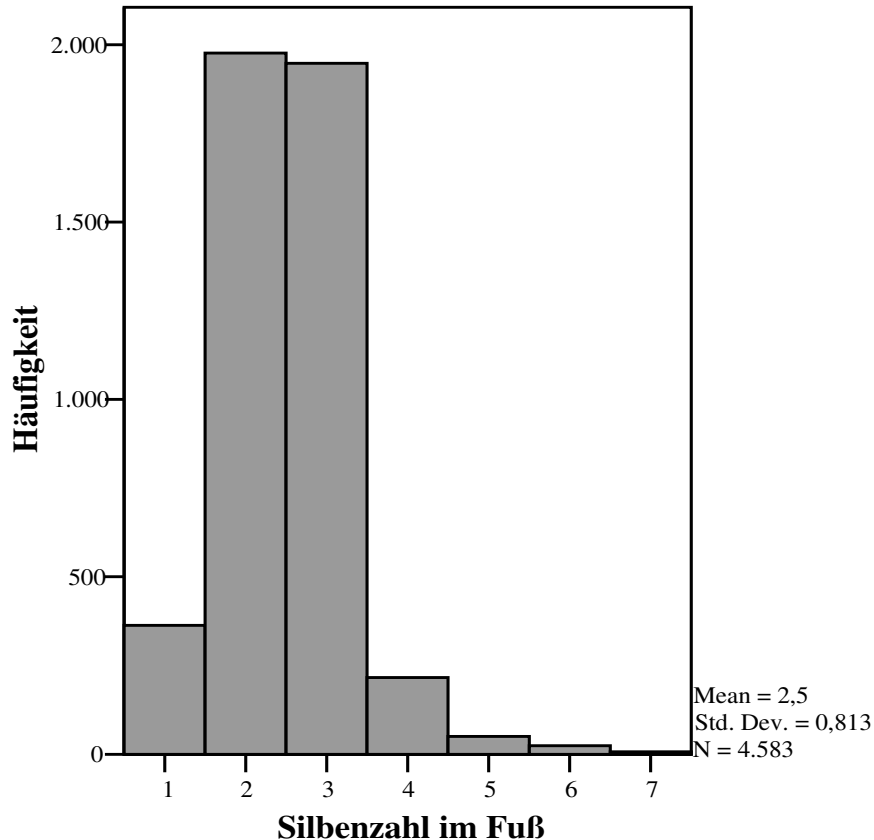
$$\Leftrightarrow x = 310 \text{ ms}$$

$$y = 0.723 * 310 \text{ ms} = 224 \text{ ms}$$

Die Äquivalenzumformungen der Funktionen in den Punkten 1), 2), 3) und 4) führen im Ergebnis zu den konkreten Silbendauern für jedes der vier Metren im lyrischen Sprechstil. Es gibt dabei zwei Ergebnisse. Zum ersten erhält man die Dauer für die akzentuierte Silbe  $x$ . Zum zweiten lässt sich die Dauer des unakzentuierten Teils  $y$  eines Fußes berechnen. Für Jambus und Trochäus läuft dieses auf nur eine unakzentuierte Silbe hinaus. Der Daktylus dagegen besitzt zwei unakzentuierte Silben. Hier wird die Dauer des unakzentuierten Teils des Fußes halbiert. Ein Problemfall ist dabei die Liedform. Sie besitzt eine variable Anzahl von Silben pro Fuß. Bei der in Kapitel 8 skizzierten Signalmanipulation auf Basis isochroner Akzentfußdauern muss die Zahl von unakzentuierten Silben im jeweiligen Liedform-Fuß vor der Manipulation bestimmt werden. Die Dauer des unakzentuierten Teils des Fußes wird dann durch die Zahl der unakzentuierten Silben geteilt und als Silbendauer angenommen.

Die Abbildung 7.2 zeigt die Häufigkeit der verschiedenen Silbenanzahlen pro Fuß für die Liedform in lyrischem Sprechstil. Es zeigt sich, dass ein Liedform-Fuß in den weitaus meisten Fällen 2 oder 3 Silben beinhaltet. Die aus der Berechnungsvorschrift für die Liedform hervorgehenden Silbendauern sollten damit in einer Größenordnung liegen, die bei einer entsprechenden Signalmanipulation die Natürlichkeit der Silbendauern nicht stört.





**Abb. 7.2:** Histogramm: Häufigkeit der verschiedenen Silbenanzahlen im Akzentfuß für die Liedform bei lyrischem Sprechstil

## 7.2 Intonationsmodell

In diesem Abschnitt wird Bezug auf die in Kapitel 6.2.2 erörterte phonologische Beschreibung der Intonation für die vier Metren Jambus, Trochäus, Daktylus und Liedform bei lyrischem Sprechstil genommen. Das Kapitel 6.2.2 hat gezeigt, dass es sinnvoll ist, die Beschreibung der Intonation durch GToBI-Etikette nach zwei Variablen aufzuschlüsseln. Die erste Variable heißt *Meter*. Sie unterscheidet die verschiedenen Metren voneinander. Die zweite Variable heißt *FootInVerse*. Sie gibt die Position des aktuellen Fußes innerhalb des zugehörigen Verses an. Zur exakten Definition der Variablen sei erneut auf das Kapitel 5.3 bzw. Anhang A verwiesen. Es wird davon ausgegangen, dass man einen Tonakzent pro Versfuß ansetzen kann. In der Prosa dient ein Tonakzent in erster Linie zur Informationsstrukturierung. In der Lyrik dagegen kommt dem Tonakzent vor allem eine ästhetische Funktion zu. Da sich nun aber innerhalb eines Verses keine positionsabhängige

Hierarchie bezüglich der Hebungen feststellen lässt, wird in der Abfolge der Hebungen und Senkungen jede Hebung mit einem Tonakzent assoziiert. Eine Hebung kann dabei selbstverständlich sowohl durch einen Grundfrequenzgipfel als auch ein Grundfrequenztal realisiert sein. Damit die GToBI-Beschreibung der vier Rhythmen statistisch aussagekräftig bleibt, ist die Zahl der etikettierten Akzente auf 5 beschränkt worden. Anderenfalls wäre die Auftretenshäufigkeit der Akzente jenseits der fünften Fußposition im jeweiligen Vers zu gering gewesen. Für die Beschreibung schließlich werden jeweils die am häufigsten an einer bestimmten Position vorkommenden GToBI\_Etikette angeführt. Die positionsabhängige GToBI-Modellierung der vier im Korpus berücksichtigten Rhythmen sieht wie folgt aus:

Jambus:

**H\*** (1. Fuß) **L+H\*** (2. Fuß) **L\*+H** (3. Fuß) **H\*** (4. Fuß) **H\*** (5. Fuß)

Trochäus:

**H\*** (1. Fuß) **H\*** (2. Fuß) **L\*+H** (3. Fuß) **L\*** (4. Fuß) **H\*** (5. Fuß)

Daktylus:

**H\*** (1. Fuß) **L\*+H** (2. Fuß) **L+H\*** (3. Fuß) **L+H\*** (4. Fuß) **L\*+H** (5. Fuß)

Liedform:

**L+H\*** (1. Fuß) **H\*** (2. Fuß) **H\*** (3. Fuß) **H\*** (4. Fuß) **H\*** (5. Fuß)

Bei obiger Beschreibung fällt auf, dass schon die Akzent-Etikette der ersten beiden Fußpositionen ausreichen, um die vier Rhythmen eindeutig voneinander zu unterscheiden. Für den Jambus ist dieses die Tonfolge „H\*“, „L+H\*“, für den Trochäus die Folge „H\*“, „H\*“, für den Daktylus die Folge „H\*“, „L\*+H“ und für die Liedform schließlich die Folge „L+H\*“, „H\*“. Im konkreten Fall muss die Zahl der Hebungen natürlich noch an den jeweiligen Vers angepasst werden. Für den Grenzton bietet sich natürlich „L-%“ an, aber in Einzelfällen könnte auch „H-%“ gesetzt werden. Diese Entscheidung sollte auf Basis von Syntax und Semantik des zugrunde liegenden Textes gefällt werden.

An diese GToBI-Charakterisierung der Rhythmen Jambus, Trochäus, Daktylus und Liedform ist die Hoffnung geknüpft, dass die Akzentkombinationen als Steuerung für die Intonation bei

einer nachfolgenden Signalmanipulation verwendet werden können. Hierzu sind Versuche mit dem *Open Source* Sprachsynthesystem Mary<sup>13</sup> (Schröder & Trouvain, 2001; Schröder, Hunecke & Krstulovic, 2006) unternommen worden, da dieses in der Lage ist, GToBI-Etiketten in Grundfrequenzkonturen zu überführen. Es stellte sich jedoch das Problem, dass das prosaische Intonationsmodul zur Überführung von GToBI-Etiketten in reale Grundfrequenzkonturen für eine Sprachausgabe im lyrischen Sprechstil nicht geeignet ist. Hier liegt die Prosodie der Sprachausgabe zu nahe an der Prosa. Ein entsprechendes lyrisches Intonationsmodul konnte allerdings im Rahmen dieser Arbeit nicht entworfen werden. An dieser Stelle ist noch Entwicklungsarbeit zu leisten.

### 7.3 Zusammenfassung

Für die Dauermanipulation von Versen aus dem dieser Arbeit zugrunde liegenden Lyrik-Korpus sind zwei Prädiktionsmodelle auf Silbenbasis erstellt worden. Das eine Modell besteht in einer linearen Regression, das andere Modell in einem auf der Isochroniehypothese fußenden Ansatz. Die lineare Regression erzielt sprecherabhängig eine hohe Korrelation zwischen vorhergesagten und tatsächlichen Silbendauern bei nur drei unabhängigen Variablen (*Acce*, *PhonesInSyl* und *PhrasePosCat*). Für die Berechnung der auf der Isochronie beruhenden Silbendauern ist fußweise der Quotient aus unakzentuiertem und akzentuiertem Anteil an der jeweiligen Fußdauer (*AccSylQuot*) verwandt worden. Außerdem sind die Fußdauern auf den Wert des mittleren Fußdauermedians für den Sprecher, der am nächsten am globalen mittleren Fußdauermedian des jeweiligen Metrums liegt, gesetzt worden.

Die Intonation der vier Metren im Korpus ist positionsabhängig mit GToBI-Etiketten phonologisch beschrieben worden. Der akzentuierten Silbe jeden Fußes ist dabei ein GToBI-Etikett zugewiesen worden. Schon die ersten zwei Fußpositionen eines Metrums beschreiben dieses eindeutig in Abgrenzung zu den übrigen drei Metren. Bei der Verwendung der Tonfolgen zur Intonationssteuerung in der Sprachsynthese – etwa zur Erzeugung von Stimuli für Perzeptionstests – zeigte sich, dass ein für die Prosa entwickeltes Intonationsmodul zur Überführung von GToBI-Etiketten in konkrete Grundfrequenzwerte für die Lyrik ungeeignet

---

<sup>13</sup> Nähere Informationen zum Sprachsynthesystem Mary sind unter <http://mary.dfki.de/> zu finden.

ist. Es hätte den Rahmen dieser Arbeit gesprengt, ein solches Intonationsmodul für die Sprachsynthese zu entwickeln. Zukünftige Arbeiten könnten hier anknüpfen.

## 8 Evaluation der Dauermodelle

In Kapitel 7 sind zwei Dauermodelle zur Vorhersage von Silbendauern erstellt worden. Beide Modelle präzisieren Dauern für die vier verschiedenen Rhythmen Jambus, Trochäus, Daktylus und Liedform in der Lyrik bzw. den lyrischen Varianten der im Korpus befindlichen Gedichte. Eines der Modelle fußt dabei auf einer linearen Regression mit 3 unabhängigen Variablen. Das zweite Modell dagegen basiert auf isochronen Akzentfußdauern. In diesem Kapitel soll die Vorhersageleistung der beiden Modelle evaluiert werden. Anhand delexikalisierter und monotonisierter Stimuli ist in einem ersten Perzeptionsexperiment geprüft worden, ob der Sprechstilunterschied zwischen Lyrik und Prosa noch hörbar ist. Dieser Test wurde aufgrund noch zu erörternder Schwierigkeiten mit einem veränderten Satz an Stimuli wiederholt. Hierbei wurde auf die Delexikalisierung und die Monotonisierung verzichtet, so dass die Sprachstimuli verständlich waren. Eine LPC-Resynthese setzte dauermanipulierte und nicht-manipulierte Stimuli auf ein vergleichbares Signalqualitätsniveau der Sprache. Das zweite Experiment hingegen testete, ob die Rhythmen Jambus, Trochäus, Daktylus und Liedform in den auf die Dauervariation reduzierten Daten noch identifizierbar sind. Die Teststimuli sind mit der Sprachsignalverarbeitungssoftware Praat<sup>14</sup> erzeugt worden. Dabei ist die in diesem Kapitel noch zu erläuternde Delexikalisierungsmethode PURR (Sonntag, 1999) in modifizierter Form verwendet worden. Zudem werden einige Fakten zum Wesen der als subjektiv zu bezeichnenden Perzeptionstests angeführt.

### 8.1 Verfahren der perzeptiven Beurteilung von Sprachstimuli

Im Rahmen dieser Arbeit werden zwei Vorhersagemodelle für die Dauern von Silben bei lyrischem Sprechstil evaluiert und miteinander verglichen. Hierzu werden subjektive Tests verwandt. In diesem Abschnitt sollen einige wichtige Merkmale dieser Testmethode erläutert werden. Sonntag (1999) äußert sich ausführlich zum Thema Perzeptionstests. Sie weist auf den Einfluss bewertender Perzeptionstests auf das Skalenniveau der Ergebnisdaten hin. Es können dabei vier Gruppen unterschieden werden (Gelfand, 1998). Handelt es sich um eine

---

<sup>14</sup> Nähere Informationen zur Sprachsignalverarbeitungssoftware Praat sind unter <http://www.praat.org> zu finden.

Aufgabe der Identifikation und Klassifikation, so liegen die Ergebnisse auf Nominalskalenniveau. Bei einer Rangreihenfolgenbestimmung ergibt sich ein Ordinalskalenniveau. Die Abstandsmaßbestimmung führt zu einem Intervallskalenniveau. Die Verhältnismaßbestimmung schließlich ergibt ein Rationalskalenniveau. Die im Rahmen dieser Arbeit durchgeführten Tests bewegen sich ausschließlich auf Nominalskalenniveau. Das heißt es liegen Messungen zugrunde, die rein qualitativer Natur sind. Es wurden Aufgaben einer Identifikation bzw. Klassifikation von Sprachstimuli gestellt. Somit bleibt der quantitative Aspekt der Daten, abgesehen von Häufigkeiten, außen vor.

### **8.1.1 Der Faktor Mensch**

Sonntag (1999) betont, dass die Probanden eines Perzeptionstests keine homogene Gruppe bilden. Die Hörer unterscheiden sich beispielsweise hinsichtlich ihres Weltwissens und ihrer Sprachkompetenz. Die Hörerperformanz wird im Wesentlichen von fünf Faktoren beeinflusst, nämlich der Aufgabenkomplexität, der linguistischen Struktur des Stimulus, dem Rahmen der menschlichen Wahrnehmung, der Erfahrung und Training sowie den akustischen Signaleigenschaften (Pisoni et al., 1985). Die beiden erstgenannten Faktoren können dabei unabhängig kontrolliert werden. Die dritt- und viertgenannten Faktoren charakterisieren dagegen den Hörer. Der letztgenannte Faktor schließlich stellt den eigentlichen Testgegenstand dar.

Die Probanden eines Perzeptionstests sind zumeist unterschiedlich hoch motiviert. Das kann am Engagement des Hörers liegen. Es kann aber ebenso möglich sein, dass die Motivation davon abhängt, ob der Proband freiwillig am Experiment teilnimmt oder ob Druck auf ihn ausgeübt wird. Oftmals werden die Hörer in die beiden Gruppen der Experten und der Ungeübten unterteilt (Hustad et al., 1998). Bei dieser Unterteilung bildet die Vertrautheit mit dem Untersuchungsgegenstand den Maßstab. Sonntag (1999) zweifelt den Sinn dieser Unterteilung jedoch an.

Strittig ist die Frage, wie viele Testpersonen in einen Perzeptionstest einbezogen werden sollen. Die Literatur zeigt eine Spanne von 1 (Strangert & Aasa, 1996) bis 322 Versuchspersonen (Boogart & Silverman, 1992) auf. Sonntag (1999) weist aber darauf hin,

dass in den meisten Fällen eine Versuchspersonenzahl von 10 bis 20 verwandt wird. Die dabei befragten Hörer sind zumeist Teilnehmer einer Gelegenheits- bzw. Willkürstichprobe, d.h. die Versuchspersonen rekrutieren sich aus dem Kreis derer, die für den Versuchsleiter am leichtesten zugänglich sind. Nach Sonntag ist es zudem noch ungewiss, welchen Einfluss die Anzahl der Versuchspersonen auf die Ergebnisse von Sprachperzeptionstests hat.

Darüber hinaus zeigen die Versuchspersonen eines Perzeptionstests starke Übungs- und Gewöhnungseffekte. Denn die Performanz von Probanden steigt schon nach einer kurzen Gewöhnungszeit innerhalb eines Tests deutlich an. Schon nach einer vergleichsweise kurzen Trainingsphase bildet sich eine nachhaltige Adaption an Stimuli und Aufgabe heraus (Carlson et al., 1976; Pisoni & Hunnicutt, 1980).

Des Weiteren stellt sich im Rahmen der Durchführung von Perzeptionsexperimenten das Problem des Stimulusumfangs. Hörer neigen dazu, ihre Urteile schrittweise im Verlauf des Experiments an die Variationsbreite der Stimuli anzupassen. Außerdem können große Abstände bezüglich des zu beurteilenden Parameters kleinere Abstände verdecken. Der Einfluss des Stimulusumfangs kann durch die Einführung so genannter Ankerstimuli kontrolliert werden. Diese Ankerstimuli sind als Referenz mit konstanten Eigenschaften zu verstehen. Auch ihre Beurteilung sollte damit per Definition konstant sein. Bei der Beurteilung von synthetischen bzw. manipulierten Stimuli werden vielfach menschliche Stimmen als Referenz verwandt. Die menschlichen Stimmen können dabei zum Zweck einer Annäherung an die eigentlichen Teststimuli allerdings auch degradiert werden.

### **8.1.2 Testmethoden**

In diesem Abschnitt werden verschiedene Methoden zur Durchführung von Perzeptionstests beschrieben. Bei allen Verfahren gilt dabei, dass die Stimuli konstant sind und nicht von den Versuchspersonen beeinflusst werden können. Eine dieser Testmethoden ist die kategoriale Einschätzung, die sowohl auf Basis sprachlicher als auch numerischer Kategorien vorgenommen werden kann (Sonntag, 1999). Wenn im Ergebnis Zahlenwerte vorliegen, ist das Skalenniveau metrisch und es kann ein *mean opinion score* (MOS) berechnet werden. Hierbei kann eine kontinuierliche oder kategoriale Skala vorgegeben werden, wobei die

kontinuierliche Skala ein konsistenteres Urteilen erlaubt als die kategoriale Skala (Dehnel & Klaus, 1996). Ein weiteres Problemfeld stellt die Zahl der Kategorien dar. Die Spanne reicht hier von 5 (ITU-T, 1993) bis zu 11 Kategorien (Goldstein et al., 1992). Sonntag (1999) hält es allerdings für wichtiger, ob eine gerade oder ungerade Anzahl von Kategorien gewählt wird. Der Vorteil der geraden Anzahl von Kategorien liegt darin, dass die Versuchsperson sich für eine Seite der Skala entscheiden muss und nicht mit „ich weiß nicht“ antworten kann.

Eine weitere Testmethode ist der Paarvergleichstest (*pair comparison*, PC). Bei diesem Test wird die Versuchsperson gebeten, zwei aufeinanderfolgende Stimuli miteinander zu vergleichen und den präferierten Stimulus zu wählen. Die Präferenz kann dabei global oder aber auf einen bestimmten Parameter ausgerichtet sein. Um Reihenfolgeeffekte zu vermeiden, sollte zudem ein ausgewogenes Verhältnis zwischen AB- und BA-Paaren bestehen. Der Paarvergleichstest ist auf kurze Teststimuli beschränkt (van Bezooijen & Pols, 1990). Gleichzeitig hat er den Nachteil eines sehr hohen Zeitaufwandes.

Ein Testverfahren mit geringem Zeitaufwand ist das Größenschätzverfahren (*magnitude estimation*, ME) zur numerischen Beschreibung sensorischer Stimuli (Stevens, 1957; 1958). Empfohlen wird für diese Methode eine freie Größenschätzung, d.h. die Urteilsskala ist zwar nach unten hin durch den Nullpunkt begrenzt, ist aber nach oben hin offen (Pavlovic et al., 1990). Überdies sollte dem Probanden das Testmaterial vor dem eigentlichen Test vollständig präsentiert werden, damit die innere Skalierung des Hörers kalibriert werden kann und nicht durch eine unvorhergesehen extreme Merkmalsausprägung beeinflusst wird.

Jede der oben erörterten Testmethoden hat spezifische Vor- und Nachteile. Das wichtigste Kriterium für die Wahl der Testmethode aber sollte ihre Diskriminationsleistung sein. Am ehesten bewährt hat sich dabei der Paarvergleichstest gefolgt vom Größenschätzverfahren. Das nächst bessere Verfahren ist die kategoriale Einschätzung und schließlich die zu vernachlässigende Reaktionszeitmessung, die jedoch von Probanden als die einfachste Methode betrachtet wird (Delogu et al., 1991).

Bisher wurden in diesem Abschnitt ausschließlich Testmethoden der Beurteilung beschrieben. Darüber hinaus existieren aber nach Sonntag (1999) auch funktionale Tests etwa zur Messung von Verständlichkeitsraten oder der Verstehensleistung eines Hörers. Es ist



jedoch schwierig zu definieren, wie das Verstehen einer Versuchsperson gemessen und damit quantifiziert werden soll. Das Messen der Answererfolgsquote ist eine Methode die Verstehensleistung eines Probanden zu quantifizieren. Hierbei werden Fragen zum Gegenstand des Verstehens gestellt. Je mehr Fragen richtig beantwortet werden, desto größer ist die Verstehensleistung. Wenn eine Mehrfachwahlaufgabe gestellt wird, bedeutet dieses, dass eine feste Anzahl möglicher Antworten vorgegeben wird. Der Sinn dieser Einschränkung liegt in der Kontrolle der verschiedenen die Antwort des Hörers beeinflussenden Faktoren. Überdies können Antwort- und Wiederholungslatenzzeit gemessen werden. Damit erhält man ein Maß für die Arbeitsleistung zur Sprachverarbeitung. Es stellt sich dabei allerdings die Frage, ob ein Proband mündlich oder per Tastendruck antworten soll. Bei einer mündlichen Antwort könnte es sein, dass der Aufwand der Sprachproduktion mit gemessen wird. Des Weiteren kann dem Probanden eine Reproduktion des Gehörten abverlangt werden. Auch hier kann man die Erfolgsquote des Hörers messen. Den Schwierigkeitsgrad einer solchen Aufgabe kann man durch die Darbietungsgeschwindigkeit oder das Stellen einer zusätzlichen Reproduktionsaufgabe steuern. Eine zusätzliche Reproduktionsaufgabe könnte beispielsweise darin bestehen, dass vor dem eigentlichen Test eine Zahlenreihe präsentiert wird, die nach dem Testlauf und vor der eigentlichen Reproduktion in der richtigen Reihenfolge wiederholt werden soll (Luce et al., 1983). Das Arbeitsgedächtnis wird so stärker beansprucht als ohne Zusatzaufgabe. Die Zweitaufgabe kann zudem im visuellen Bereich angesiedelt werden. Man spricht dann von einem kombinierten audio-visuellen Testverfahren. Es werden dabei aber wohl unterschiedliche Gehirnregionen beansprucht, so dass nicht gewährleistet ist, dass die Bearbeitung der Zweitaufgabe auch wirklich die Ergebnisse der Erstaufgabe beeinträchtigt.

### **8.1.3 Diskussion**

Die in Abschnitt 8.1 erörterten Testverfahren sind in erster Linie für die Evaluation von synthetischer bzw. lärmgestörter Sprache entwickelt worden. Im Rahmen dieser Arbeit wurden Perzeptionstests durchgeführt, die keine direkt als synthetisch zu bezeichnenden Stimuli enthalten. Es wurden vielmehr Stimuli erzeugt, die unter Verwendung von Methoden der Sprachsignalverarbeitung manipuliert wurden. Diese speziellen Stimuli sollten aber mit synthetischer und lärmgestörter Sprache den entscheidenden Umstand gemein haben, dass sie

schwieriger zu dekodieren sind als natürliche Sprachstimuli. Für die in dieser Arbeit verwandten Stimuli gilt allerdings nicht, dass wie bei lärmgestörter Sprache Signalmaskierung vorliegt. Ferner gilt auch nicht, dass, wie es bei synthetischer Sprache der Fall ist, redundante Merkmale in den Teststimuli fehlen und falsche redundante Merkmale vorhanden sind. Sonntag (1999) sieht die Ursache falscher redundanter Merkmale vor allem in koartikulatorischen Phänomenen bei der konkatenativen Sprachsynthese begründet. Der Kontext, dem die für die Synthese verwandten Diphone entstammen, ist zumeist ein anderer als in der synthetisierten Äußerung selbst.

Zum Unterschied von beurteilenden und funktionalen Tests merkt Sonntag (1999) an, dass die Beurteilungstests die verschiedenen Dimensionen von Sprachqualität trennen müssen. Das kann entweder dadurch erreicht werden, dass der Hörer instruiert wird, nur eine Dimension zu beurteilen oder dass die Teststimuli dahingehend manipuliert werden, dass bestimmte Dimensionen gezielt ausgeschlossen werden. Die vorliegende Arbeit knüpft hier an und macht sich die bereits skizzierte Delexikalisierung von Stimuli durch Signalmanipulation zu Nutze.

## 8.2 Delexikalisierung

Nach Sonntag (1999) gibt es im Wesentlichen drei Methoden der Delexikalisierung: a) Nachahmung der zu untersuchenden Sprachstimuli durch einen Sprecher (*Mimikry*), b) Resynthese mit Übernahme prosodischer Steuerparameter und c) Signalmanipulation, bei der die eigentlichen Äußerungen nach einer Manipulation direkt verwendet werden können.

Bei der Nachahmung gibt es, zumindest für das Schwedische gemäß Svensson (1971), vier prosodische Formen für Zweisilber und zwölf für Dreisilber. Anhand mit geschlossenem Mund gesummter Stimuli konnten Probanden in 83% der Fälle die zugrunde liegenden Wörter bzw. Phrasen identifizieren. Außerdem erleichterte die Markierung von Stammorphemen auf den Antwortbögen die Erkennung der verschiedenen prosodischen Formen. Des Weiteren existiert die so genannte reiterierte Sprache (*reiterant speech*), bei der die prosodische Struktur einer Äußerung aufgrund der Ersetzung der Silben durch eine konstante Silbe imitiert wird (Lieberman & Streeter, 1978). Diese reiterierte Sprache kann von

Sprechern kontrolliert und reproduzierbar erzeugt werden. Allerdings sind dazu nicht alle Sprecher in der Lage. Selbst diejenigen, die gute reiterierte Sprache erzeugen können, schaffen dieses nicht in gleich bleibend hoher Qualität.

Die Grundidee der Resynthese als Methode der Delexikalisierung ist, das Originalsignal durch einen synthetischen Laut zu ersetzen. Dabei kann die Substitution durch nur einen Laut (z.B. [a]), durch die *saltanaj-Methode* (Frikative werden durch [s], Vokale durch [a], Liquide durch [l], Plosive durch [t], Nasale durch [n] und Approximanten durch [j] substituiert) oder durch das *sasasa-Verfahren* (s. Kap. 2) erfolgen. *Saltanaj* und *sasasa* sind ebenbürtig bei der Unterscheidung von verschiedenen Sprachen durch Probanden. Leider ist der hohe Erzeugungsaufwand von Resynthesestimuli problematisch. Zudem sollten Sprachsynthesysteme mit einer gegebenen Zielsprache nicht für die Stimuligenerierung fremder Sprachen benutzt werden.

Bei der Delexikalisierung durch Signalmanipulation werden aufgrund der Verwendung des Originalsignals Fehler beim Kopieren des zu untersuchenden Sprachmaterials vermieden. Unter den Verfahren der Signalmanipulation, nämlich der spektralen Inversion, der Bandpassfilterung und der Ersetzung von Grundperiodenmarken durch ein spezielles Signal, stellte sich die Ersetzungsmethode als die am besten für Perzeptionsexperimente mit synthetischen Sprachstimuli geeignete heraus (Sonntag, 1999). Hierbei können unterschiedliche zu substituierende Signale verwandt werden. Das bevorzugte Verfahren musste den drei Kriterien der Funktionalität (Erkennbarkeit der einzelnen prosodischen Funktionen), der Akzeptanz (Erkennbarkeit des resultierenden Signals als Sprachsignal) und der automatischen Generierbarkeit gerecht werden. Tests im Bereich der Silbenzahlerkennung, der Phrasenakzenterkennung, der Satzmoduserkennung, der Phrasengrenzerkennung und der Akzeptanz ergaben keinen signifikanten Unterschied zwischen den einzelnen Manipulationsverfahren. Schließlich gab für die Auswahl der am besten geeigneten Manipulationsmethode das subjektive Urteil der Probanden, die die resultierenden Signale gehört und beurteilt haben, den Ausschlag. Das bevorzugte Verfahren war dabei das einer Grundperiodenersetzung durch Sinoidalschwingungen (Sonntag & Portele, 1997; Sonntag & Portele, 1998a). Es werden hierzu drei Sinussignale addiert und als Substitutionssignal verwendet. Der erste Sinus besitzt die aus dem Originalsignal entnommene Amplitude und Periodendauer, der zweite Sinus besitzt die doppelte Frequenz

und nur 1/4 der Originalamplitude, und der dritte Sinus besitzt die dreifache Frequenz und 1/16 der Originalamplitude. Die stimmlosen Abschnitte im Signal manifestieren sich dagegen als Pausen. Diese Methode wird von Sonntag (1999) PURR (*Prosody Unveiling Restricted Representation*) genannt. Durch PURR sind die funktionalen rhythmischen Unterschiede adäquat in den delexikalisierten Stimuli kodiert (Sonntag & Portele, 1998a). Darüber hinaus ist das PURR-Signal vollständig unabhängig vom lexikalischen Gehalt der Stimuli. Des Weiteren können Hörer intern Texte mit dem gehörten PURR-Signal assoziieren. Außerdem ist das Testen mit PURR ohne langes Training sowohl für Laien als auch für Experten durchaus möglich. Die Anwendung von PURR, beispielsweise zur Evaluation der prosodischen Natürlichkeit verschiedener Sprachsynthesysteme, kann in Sonntag und Portele (1998b) nachvollzogen werden.

Darüber hinaus soll von einer Studie der Autoren Herment-Dujardin und Hirst (2002) berichtet werden. Für das Experiment wurden fünf Gruppen synthetischer Stimuli generiert. Die erste Gruppe bildeten die Referenzstimuli, die so nahe wie möglich am Original sind, die zweite Gruppe bildeten Stimuli mit neutralisierter Grundfrequenzbewegung, die dritte Gruppe bildeten Stimuli mit neutralisierter Phondauervariation, die vierte Gruppe bildeten Stimuli mit hinzugefügten bzw. gelöschten Pausen und die fünfte Gruppe schließlich bildeten durch Resynthese delexikalisierte Stimuli. Bezüglich der Dauervariation sind Phondauern auf einen Durchschnittswert gesetzt worden. Im Zuge der F<sub>0</sub>-Manipulation sind 135 Hz für das erste Phon und 90 Hz für das letzte Phon eines Segments bei einem männlichen Sprecher festgesetzt worden. Die Stimuli wurden dadurch monoton, waren aber gleichzeitig mit einer Deklination der Grundfrequenz versehen. Anhand delexikalisierter Stimuli sollte die Bedeutung der Semantik getestet werden. Es wurde dabei die *jastradan*-Methode angewandt. Es handelt sich dabei um eine Erweiterung des *saltanaj* mit den entsprechenden Substitutionen von Segmenten. Bei der Experimentdurchführung wurden ungeübte Probanden gebeten, den Grad der Emphase der dargebotenen Stimuli zu beurteilen. Über die Bedeutung der Pausen für die Emphase von Sprachstimuli konnte keine zulässige Aussage getroffen werden. Zudem kann die Dauervariation alleine keine Emphase ausdrücken. F<sub>0</sub> und Semantik werden zusätzlich benötigt. Demgegenüber kann die F<sub>0</sub>-Variation alleine schon Emphase ausdrücken. Überdies kann Emphase auch bei delexikalisierten Stimuli wahrgenommen werden. F<sub>0</sub> und Dauer sind bedeutender als die Semantik. Bei semantisch markierten Wörtern konnte außerdem festgestellt werden, dass die Bedeutung von Wörtern

alleine schon Emphase ausdrücken kann. Man kann allerdings nicht jeden für den Ausdruck der Emphase wichtigen Parameter separat analysieren, da eine zu starke Wechselwirkung zwischen den Parametern existiert. Denn die Parameter sind als eingebettet und assoziiert anzusehen. Die beiden wichtigsten Parameter für die Perzeption von Emphase sind damit a) die Grundfrequenz und b) die Semantik. Das Experiment bestätigt, dass Emphase dank einer komplexen, subtilen und besonders variablen Kombination verschiedener Parameter wahrgenommen wird.

Für die Stimuligenerierung in der vorliegenden Arbeit wurde ebenfalls eine Delexikalisierung benutzt, damit die Hörerurteile nicht durch die segmentalen Informationen der den Teststimuli zugrunde liegenden Sprachsignalausschnitten beeinflusst wurden. Es ist dabei die Entscheidung getroffen worden, das Sprachsignal gänzlich von segmentaler Information zu befreien (außer im Wiederholungsexperiment) und die Prosodie auf die Silbendauervariation zu reduzieren. Diese Reduktion der Signale ist vorgenommen worden, damit die in Kapitel 7 skizzierten Dauermodelle ohne den Einfluss der übrigen prosodischen Parameter getestet werden konnten. Somit ist eine Modifikation von PURR angewandt worden. Hierbei ist unter Verwendung der Sprachsignalverarbeitungssoftware Praat ein Summton (*humming*) erzeugt worden, dem zusätzlich zur Grundfrequenz ein Formant bei der zweiten harmonischen Teilschwingung hinzugefügt wurde. Des Weiteren sind die so erzeugten Signale hinsichtlich der Grundfrequenz monotonisiert worden, um das zugrunde liegende Sprachsignal, wie bereits erwähnt, auf seinen Rhythmus<sup>15</sup> zu reduzieren. Es ist dabei auf eine Deklination der Grundfrequenz verzichtet worden, da diese für die Hörexperimente der vorliegenden Arbeit als nicht hilfreich angesehen wurden. Auch die Intensität wurde über den gesamten Sprachstimulus auf einen konstanten Durchschnittswert gesetzt.

### 8.3 Modellevaluation

In diesem Abschnitt soll die eigentliche Evaluation der in Kapitel 7 erläuterten Silbendauermodelle für Sprache in lyrischem Stil erörtert werden. Es werden dazu die

---

<sup>15</sup> Der Sprechrhythmus wird in dieser Arbeit als ein komplexes Phänomen, das sich durch mehrere Parameter konstituiert, verstanden. Die Silbendauervariation wird jedoch als der wichtigste Parameter für die in der vorliegenden Arbeit durchgeführten Perzeptionsexperimente angesehen.

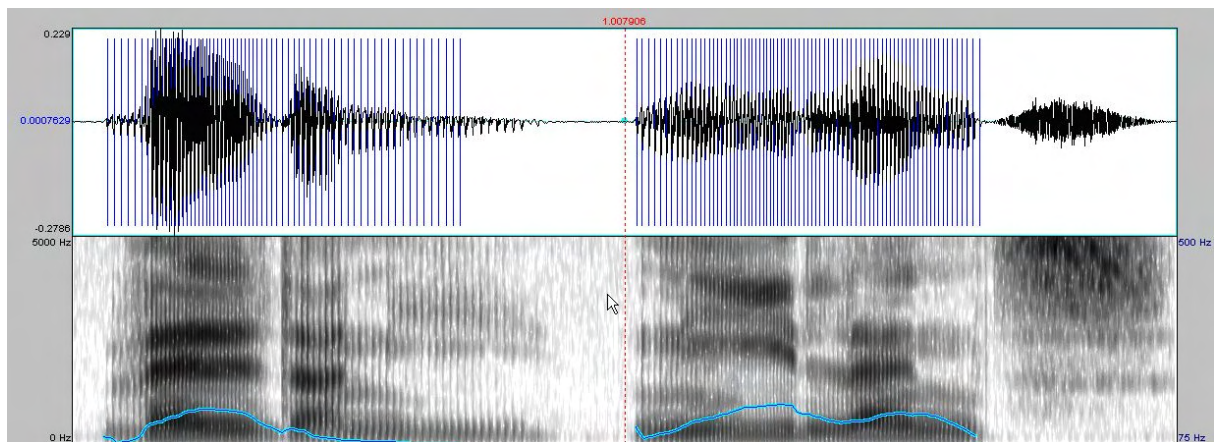
Testvorbereitung sowie die Testergebnisse aufgezeigt. Es sind insgesamt drei Perzeptionstests durchgeführt worden. Der erste Test verlangte vom Hörer, auf den Rhythmus reduzierte Sprachsignale nach Lyrik bzw. Prosa zu klassifizieren. Der zweite Test dagegen instruierte die Hörer, die dargebotenen Stimuli, die in gleicher Weise wie im ersten Test manipuliert wurden, nach den vier Rhythmusklassen Jambus, Trochäus, Daktylus und Liedform zu klassifizieren. Die Ergebnisse des ersten Tests schließlich legten die Durchführung eines dritten Perzeptionstests nahe. Hierbei wurden die gleichen Stimuli wie in Test 1 verwendet, nur dass sie nicht delexikalisiert und nicht monotonisiert, sondern resynthetisiert wurden.

Der Umfang der Tests betrug in allen drei Tests 80 Stimuli. Für den ersten Test sind dabei 5 lyrische Verse für jedes der 4 Metren bei 2 Dauermodellen ausgewählt worden. Zusätzlich zu diesen 40 Stimuli sind 20 prosaisch gelesene Stimuli, denen die gleichen Textpassagen zugrunde lagen wie den lyrischen Stimuli, aus dem Korpus ausgewählt worden. Letztere sind im Test jeweils doppelt dargeboten worden, damit lyrische und prosaische Stimuli zu gleichen Teilen vorlagen. Die daraus resultierenden 80 Stimuli sind dann auf Basis der modifizierten PURR-Methode delexikalisiert und monotonisiert worden. Die 80 Stimuli im Wiederholungstest hatten die gleiche Basis wie die des ersten Tests, nur dass sie nicht delexikalisiert und nicht monotonisiert, sondern durch eine LPC-Resynthese qualitativ neutralisiert wurden. Die segmentale Verständlichkeit wurde also im Wiederholungstest aus noch zu skizzierenden Gründen aufrechterhalten. Im zweiten Test wurden die gleichen 40 lyrischen Stimuli wie im ersten Test und im Wiederholungstest verwandt. Der Unterschied bestand darin, dass die Prosastimuli ausgelassen und die Lyrikstimuli doppelt dargeboten wurden. Die intrasubjektive Urteilsconsistenz kann so gemessen werden.

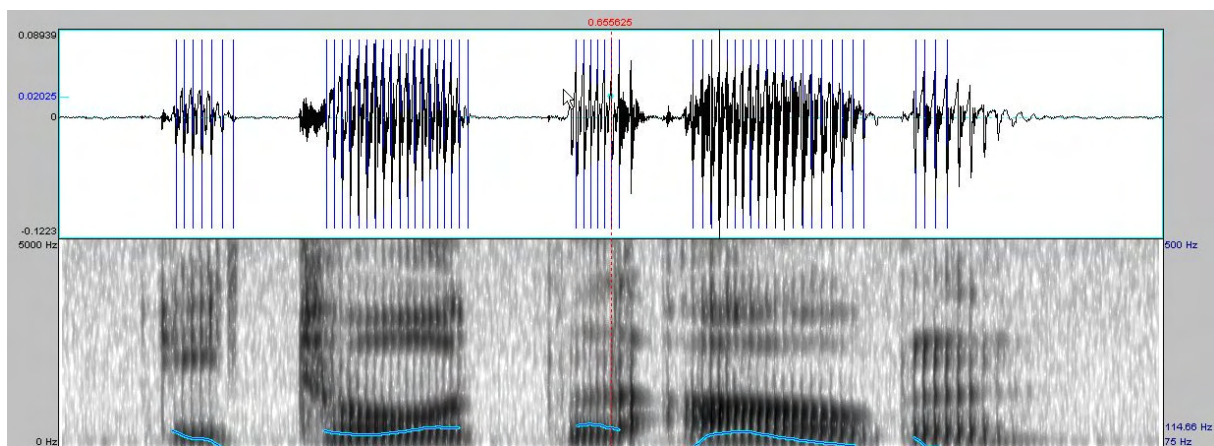
Alle im Korpus enthaltenen Gedichte wurden jeweils von 12 verschiedenen Sprechern gelesen. Für die Auswahl der Teststimuli war es wichtig, für jedes Metrum einen geeigneten Sprecher bzw. eine geeignete Sprecherin zu finden. Hierzu wurde überprüft, welcher Sprecher bei einer Silbendauermodellierung durch eine lineare Regression den höchsten Korrelationskoeffizienten zwischen prädikierter und tatsächlicher Dauer aufwies. Diese Vorgehensweise lag nahe, da die Variationsbreite zwischen den Korrelationskoeffizienten der einzelnen Sprecher sehr groß ist. Für den Jambus empfahl sich Sprecher 10 ( $r = 0,798$ ), für den Trochäus Sprecherin 2 ( $r = 0,724$ ), für den Daktylus Sprecherin 9 ( $r = 0,741$ ), und für die

Liedform schließlich ergab sich Sprecher 5 ( $r = 0,795$ ). Es sind also zu gleichen Teilen männliche wie weibliche Sprecher in die Tests einbezogen worden.

Ein weiteres Problem stellte sich bezüglich der optimalen akustischen Gestalt der Teststimuli. Bei den delexikalisierten und monotonisierten Stimuli nämlich verschmelzen benachbarte Silben mit einem stimmhaft-stimmhaft Übergang zu einem Ton miteinander. Um dieses zu verhindern, mussten Signale mit möglichst wenigen stimmhaft-stimmhaft Übergängen zwischen Silben im Korpus gefunden werden.

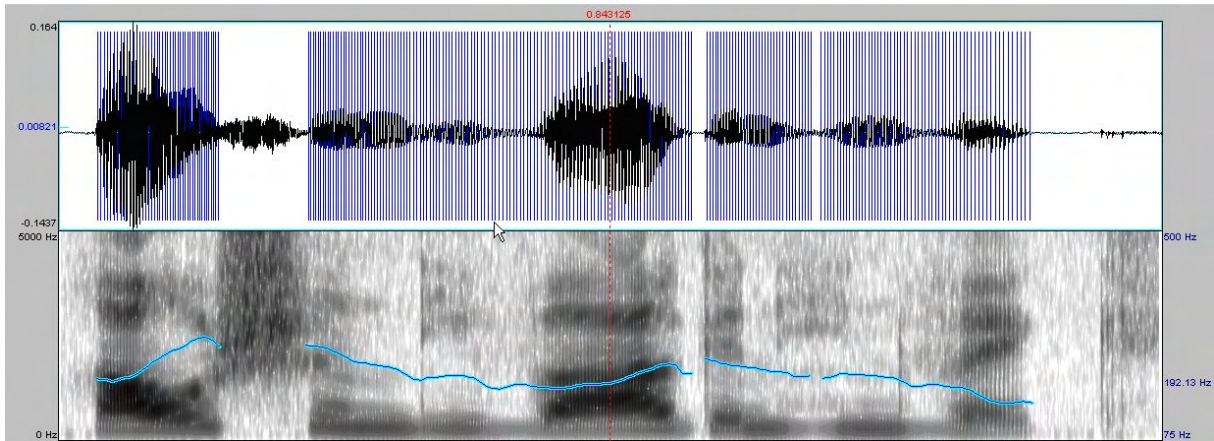


**Abb. 8.1:** Oszillogramm, Spektrogramm und Intonationskontur einer ungeeigneten Passage aus dem Gedicht „Der Blinde“ gesprochen von Sprecher 10

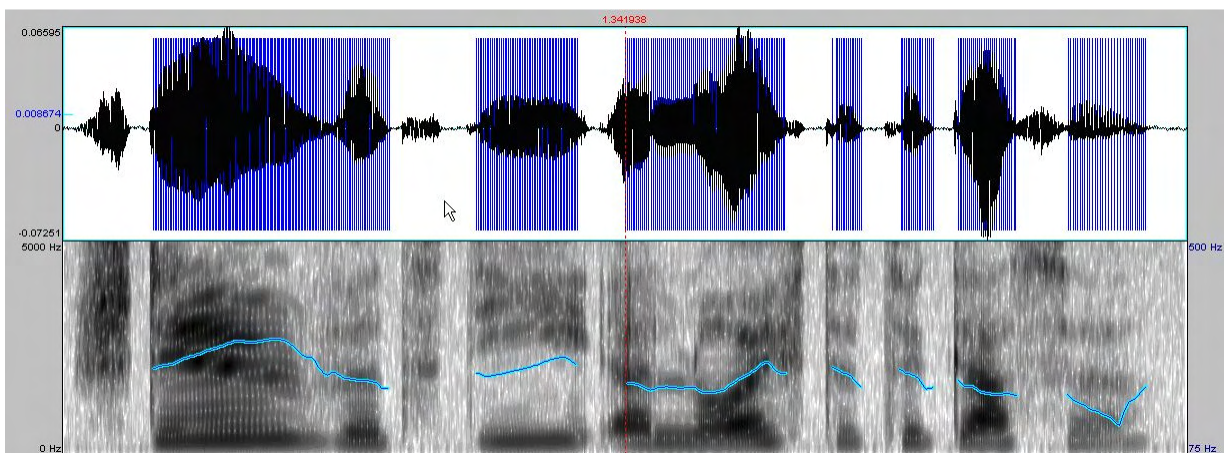


**Abb. 8.2:** Oszillogramm, Spektrogramm und Intonationskontur einer geeigneten Passage aus dem Gedicht „Der Blinde“ gesprochen von Sprecher 10





**Abb. 8.3:** Oszillogramm, Spektrogramm und Intonationskontur einer ungeeigneten Passage aus dem Gedicht „Brot und Wein – An Heinze“ gesprochen von Sprecherin 9

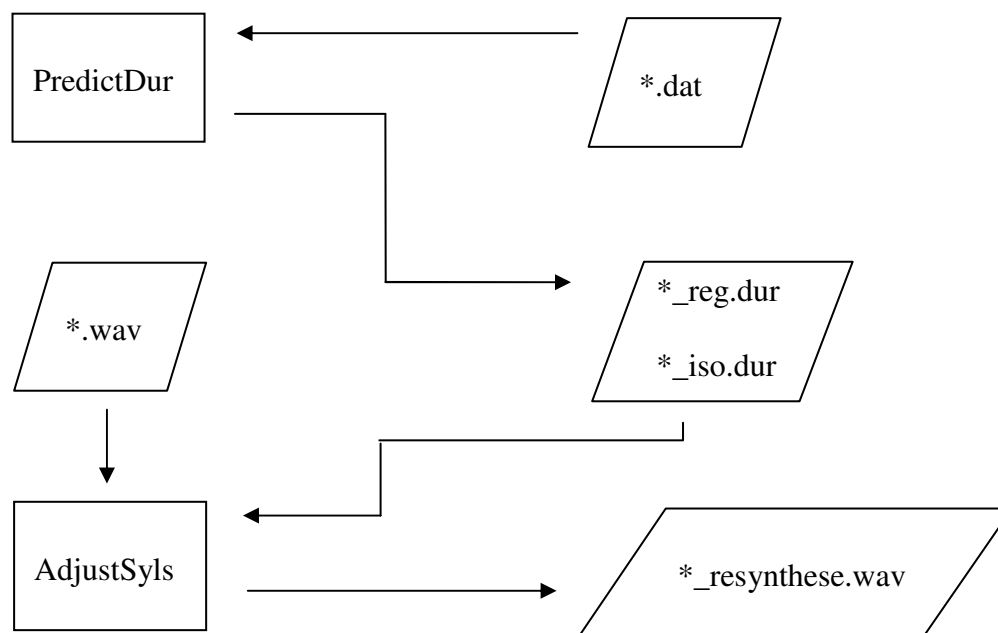


**Abb. 8.4:** Oszillogramm, Spektrogramm und Intonationskontur einer geeigneten Passage aus dem Gedicht „Brot und Wein – An Heinze“ gesprochen von Sprecherin 9

Hierzu ist die Strategie verwandt worden, dass nach möglichst stark fragmentierten Intonationskonturen gesucht wurde, d.h. die Intonationskontur eines geeigneten Stimulus sollte möglichst oft pro Stimulus abreißen. Die Abbildungen 8.1 und 8.3 zeigen exemplarisch jeweils ein ungeeignetes Sprachsignal. Die Abbildungen 8.2 und 8.4 hingegen zeigen zwei geeignete Signalausschnitte. Die beiden Ausschnitte in den Abbildungen 8.1 und 8.2 sind von Sprecher 10 gesprochen worden.

In Abbildung 8.5 ist die Vorgehensweise bei der Erstellung der Teststimuli als Flussdiagramm dargestellt. Das Programm *PredictDur* liest zunächst die Etikettendaten (\*.dat) ein und gibt die Dauervorhersagedaten für Silben in der Lyrik (\*\_reg.dur und \*\_iso.dur) aus. Letztere Daten werden dann samt den für den Test relevanten Sprachsignalen (\*.wav) dem Skript *AdjustSyls* zugeführt. Dieses Skript gibt schließlich die dauermanipulierten Ergebnisdaten (\*\_resynthese.wav) aus.





**Abb. 8.5:** Flussdiagramm für die Generierung der Stimuli des Perzeptionstests. Dargestellt sind die Abhängigkeiten zwischen den Programmen bzw. Skripten (*PredictDur* und *AdjustSyls*) und den Daten (\*.dat, \*.wav, \*\_reg.dur, \*\_iso.dur und \*\_resynthese.wav)

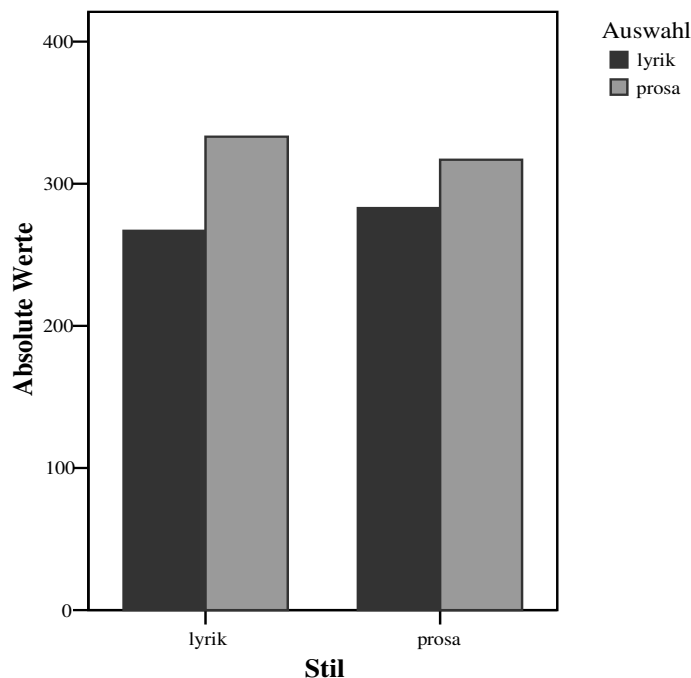
Unter Berücksichtigung aller Vorüberlegungen werden drei wesentliche Hypothesen für die Perzeptionstests aufgestellt:

- 1) Die auf Basis der beiden Dauermodelle manipulierten Sprachstimuli lassen es auch in delexikalisierter und monotonisierter Form immer noch zu, dass der Sprechstilunterschied zwischen Lyrik und Prosa von Hörern identifiziert wird. (Im Wiederholungstest wurde auf die Delexikalisierung und Monotonisierung verzichtet.)
- 2) Die auf Basis der beiden Dauermodelle manipulierten Sprachstimuli lassen es auch in delexikalisierter und monotonisierter Form immer noch zu, dass die vier Rhythmen Jambus, Trochäus, Daktylus und Liedform von Hörern identifiziert werden.
- 3) Die Isochronie ist ein perzeptives Phänomen, so dass die beiden in den Hypothesen 1) und 2) formulierten Klassifikationsaufgaben in den Fällen einer Dauermanipulation auf Basis der Isochroniehypothese erleichtert werden.

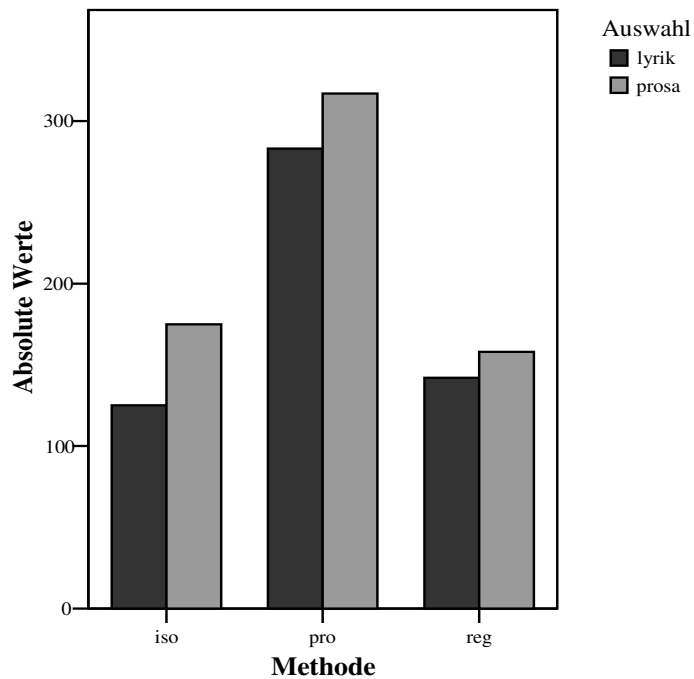
### 8.3.1 Ergebnisse der Lyrik/Prosa Unterscheidung bei delexikalisierten Stimuli

Im ersten Perzeptionstest sollte geprüft werden, ob die auf den Rhythmus reduzierten Sprachstimuli es den Probanden immer noch ermöglichen, den Sprechstilunterschied zwischen Lyrik und Prosa herauszuhören. Hierbei wird vorausgesetzt, dass diese Diskriminationsleistung bei unveränderten Sprachstimuli von muttersprachlichen Hörern ohne Probleme erbracht werden kann.

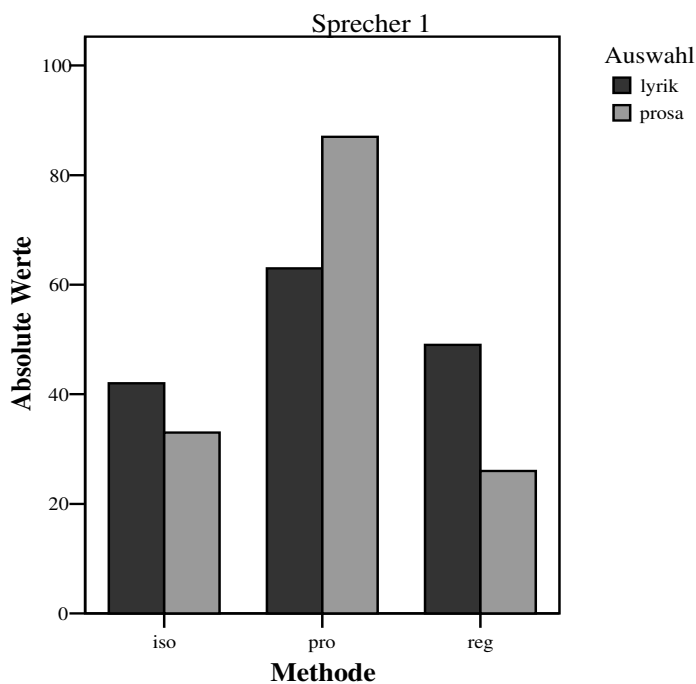
Die Abbildung 8.6 zeigt die Häufigkeiten der von den Hörern ausgewählten Sprechstilbezeichnungen verteilt auf die tatsächlich intendierten Stil kategorien. Das Diagramm zeigt, dass sowohl die lyrischen als auch die prosaischen Stimuli mehrheitlich als Prosa klassifiziert wurden. Im Falle der Kategorie Lyrik ist der Erkennungsunterschied sogar statistisch signifikant ( $p < 0,01$ ). Die Abbildung 8.7 stellt diesem Ergebnis eine Aufschlüsselung der intendierten Stil kategorien nach ihrer Manipulationsmethode gegenüber. Es zeigt sich, dass bei allen drei Stil kategorien mehrheitlich Prosa erkannt worden ist.



**Abb. 8.6:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil kategorien



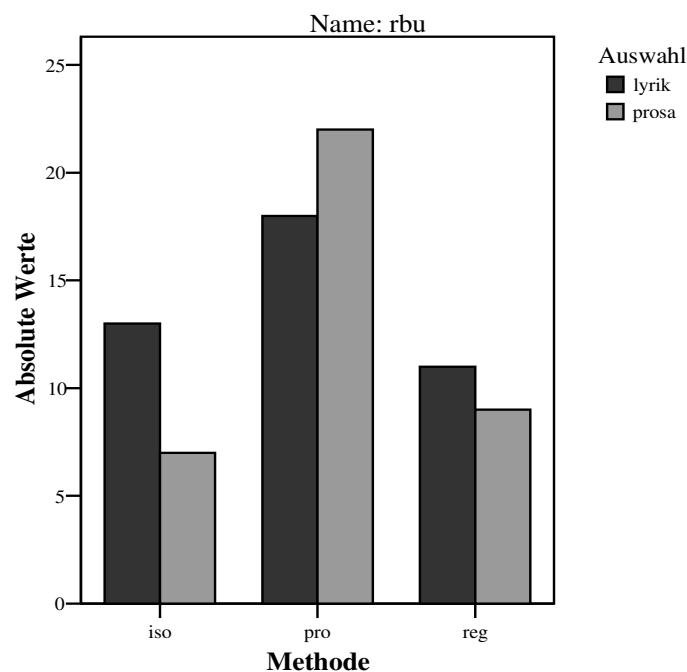
**Abb. 8.7:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression)



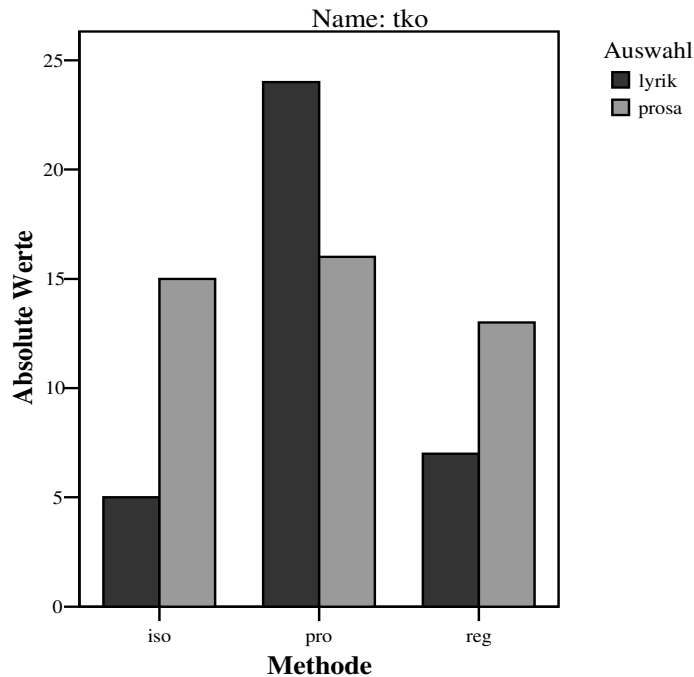
**Abb. 8.8:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression) für Sprecherin 1

Im isochronen Fall ist die Differenz zwischen Lyrik und Prosa statistisch signifikant ( $p < 0,01$ ). Abbildung 8.8 weist nach, dass diese Fehleinschätzung der Stimuli nicht bei allen vier Sprechern auftritt. Die Stimuli der Sprecherin 1 wurden mehrheitlich für alle drei Kategorien (Isochronie, Prosa und Regression) korrekt identifiziert. Diese Sprecherin sprach ausschließlich die im Test dargebotenen Trochäen. Es liegt also nahe, dass Trochäen in delexikalisierter und monotonisierter Form leichter identifizierbar sind als Jamben, Daktylen und Liedformen. Da ausschließlich ein Summton im Testverlauf zu hören war, kann überdies gemutmaßt werden, dass es der Stimmklang einzelner Sprecher ist, der die Erkennung des Sprechstils erleichtert bzw. erschwert. Es kann darüber hinaus angenommen werden, dass es in der Sprechweise eines Sprechers begründet liegt, wie gut das resultierende Sprachsignal für eine Dauermanipulation geeignet ist. Um welches Merkmal im Sprachsignal es sich dabei handelt, kann an dieser Stelle jedoch nicht geklärt werden.

Des Weiteren gelang es einigen der getesteten 15 Hörer wesentlich besser, den Sprechstil auf Basis delexikalisierter und monotonisierter Stimuli zu bestimmen. Die Abbildungen 8.8 und 8.9 vergleichen einen der Hypothese 1 entsprechend erfolgreichen Hörer mit einem weniger erfolgreichen Hörer. Der erfolgreiche Hörer *rbu* erkannte mehrheitlich sowohl die lyrischen Stimuli (iso und reg) als auch die prosaischen Stimuli (pro) korrekt.



**Abb. 8.9:** Verteilung der von Hörer *rbu* ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression)



**Abb. 8.10:** Verteilung der von Hörer *tko* ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression)

Beim weniger erfolgreichen Hörer *tko* ist das Ergebnis genau gegenläufig mit mehrheitlicher Fehlererkennung sowohl bei den lyrischen als auch prosaischen Stimuli. Diese beiden Hörer vertreten die Extreme innerhalb der Ergebnisse. Insgesamt ist festzuhalten, dass die gegenläufigen Tendenzen der Hörer sich gegenseitig aufheben. Dadurch ist in erster Linie das wenig aussagekräftige Gesamtergebnis des ersten Testlaufes zu erklären. Es stellt sich damit allerdings die Frage, warum die Urteile der Hörer so stark divergieren. Eine große Rolle dabei spielen sicherlich Fähigkeiten und Weltwissen der Probanden. So ist etwa eine phonetische oder auch musikalische Geschultheit sicherlich von Vorteil für die Erkennungsaufgabe der für die vorliegende Arbeit durchgeführten Perzeptionstests. Es wird an dieser Stelle der auf Benkwitz' (2004) eigenen empirischen Untersuchungen fußenden Sichtweise Folge geleistet:

„Einen weiteren, nicht unwesentlichen Einfluss auf die Bewertung hatte die Tatsache, dass verschiedene Hörer aufgrund der jeweils individuell unterschiedlichen Neigungen, Begabungen und Ausbildung für die zu beurteilenden Merkmale (...) unterschiedlich sensibel sind. So waren einige HörerInnen mit professioneller musikalischer Ausbildung (z.T. mit absolutem Gehör) besonders sensibel für Tonhöhenvariationen, während andere beispielsweise sensibler auf Lautstärkevariationen reagierten.“ (Benkwitz, 2004: 81)

Welche speziellen Fertigkeiten des Hörers jedoch die ausschlaggebenden sind, konnte Benkwitz nicht klären. Auch in dieser Arbeit können diese Faktoren nicht aufgeschlüsselt werden. Leider sind hierzu auch die Ausführungen in der übrigen Literatur wenig aufschlussreich.

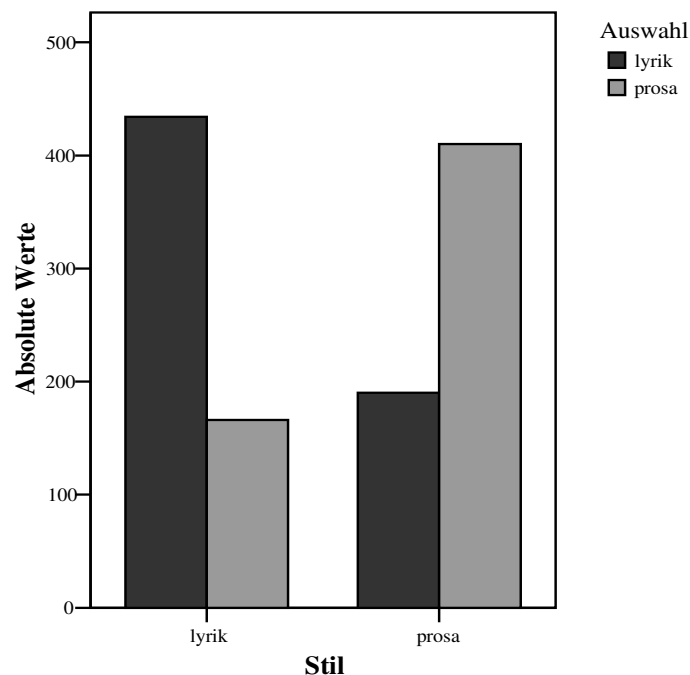
Damit müssen Hypothese 1 und Hypothese 3 für den ersten Test abgelehnt werden. Es gibt bei beiden Hypothesen jedoch eine sprecher- und hörerabhängige Tendenz hin zur Bestätigung. Da diese Tendenz jedoch zu schwach ist, um das Gesamtergebnis zu prägen, werden die genannten Hypothesen für den Perzeptionstest zur Identifizierung des Sprechstils (Lyrik vs. Prosa) delexikalisierter und monotonisierter Stimuli zurückgewiesen.

### **8.3.2 Ergebnisse der Lyrik/Prosa Unterscheidung bei nicht-delexikalisierten Stimuli**

Da die Ergebnisse der in Abschnitt 8.3.1 dargelegten Untersuchung den aufgestellten Hypothesen weitestgehend widerstreben, ist zusätzlich zu den bereits angeführten Erklärungen dieses Umstandes in Betracht gezogen worden, dass die auf den Rhythmus reduzierten Stimuli zu wenig Informationen für eine Sprechstilunterscheidung enthalten. Deshalb ist der erste Perzeptionstest mit veränderten Stimuli wiederholt worden. Es sollte dabei vor allem sichergestellt werden, dass die zu präsentierenden Stimuli trotz Manipulation noch immer als Sprache zu erkennen sind. Letzterer Umstand war bei den delexikalisierten und monotonisierten Stimuli wohl nicht mehr gegeben, was zu Problemen bei der Klassifikation führte. Die gleichen lyrischen Sprechpassagen wie im ersten Test sind somit für den Wiederholungstest ebenfalls gemäß des Isochronie- und Regressionsmodells dauermanipuliert worden. Außerdem sind alle Stimuli, auch die Prosadaten, einer LPC-Resynthese unterzogen worden. Die Sprachqualität der Signale sollte dadurch normalisiert werden.

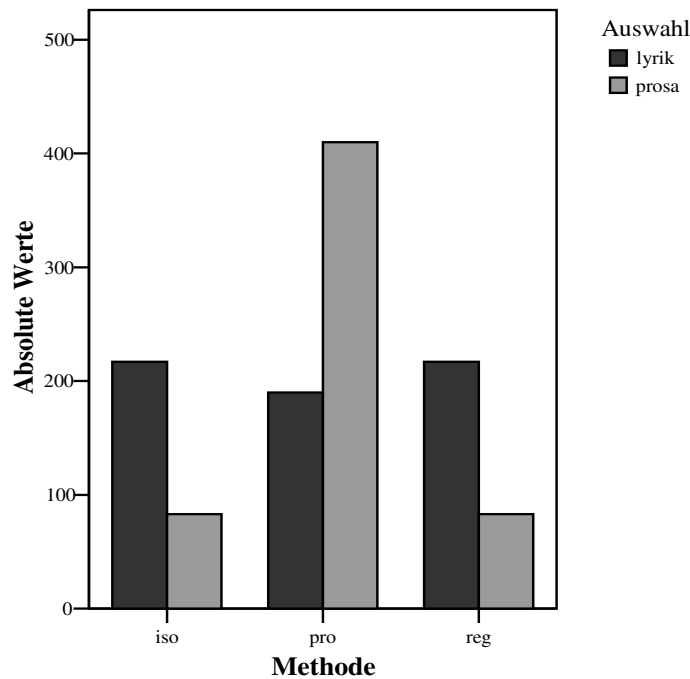
Die prosaischen (nicht-manipulierten) Stimuli hätten sonst gegenüber den dauermanipulierten Daten eine zu hohe Qualität aufgewiesen. Es sollte also vermieden werden, dass die Hörerurteile von der Signalqualität beeinflusst werden. Nichtsdestotrotz sind für den Wiederholungstest Stimuli generiert worden, die nun im Gegensatz zum ersten Test

segmental verständlich waren, obwohl sie im Lyrikfall bezüglich ihrer Dauern manipuliert waren.



**Abb. 8.11:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien

Die Abbildung 8.11 zeigt die Gesamtverteilung der von den 15 Hörern abgegebenen Sprechstilurteile auf die intendierten Stil Kategorien. Sowohl die Kategorie Lyrik als auch die Kategorie Prosa sind mit jeweils großer Mehrheit gemäß der Vorgabe erkannt worden. Die Differenzen in beiden Kategorien sind dabei nach einem Chi-Quadrat-Test statistisch signifikant (in beiden Fällen:  $p < 0,001$ ). Da die Probanden gebeten wurden, die Prosodie zur Sprechstilbeurteilung heranzuziehen, ist gezeigt, dass der Sprechstil (Lyrik vs. Prosa) durch das gesamte Bündel der Prosodie, d.h. Intonation, Intensität und Dauer, charakterisiert ist. Zudem ist bewiesen, dass es möglich ist segmental identische Texte mit klar zu erkennendem Unterschied zum einen lyrisch und zum anderen prosaisch zu lesen. Denn die prosaisch gelesenen Gedichtpassagen sind mit großer Deutlichkeit von den Hörern auch als Prosa klassifiziert worden, obwohl es sich um lyrische Texte handelte.

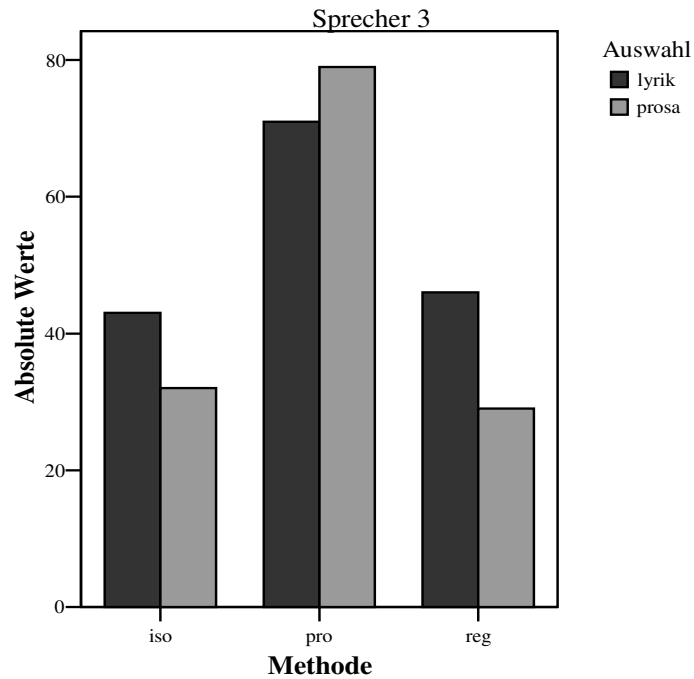


**Abb. 8.12:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression)

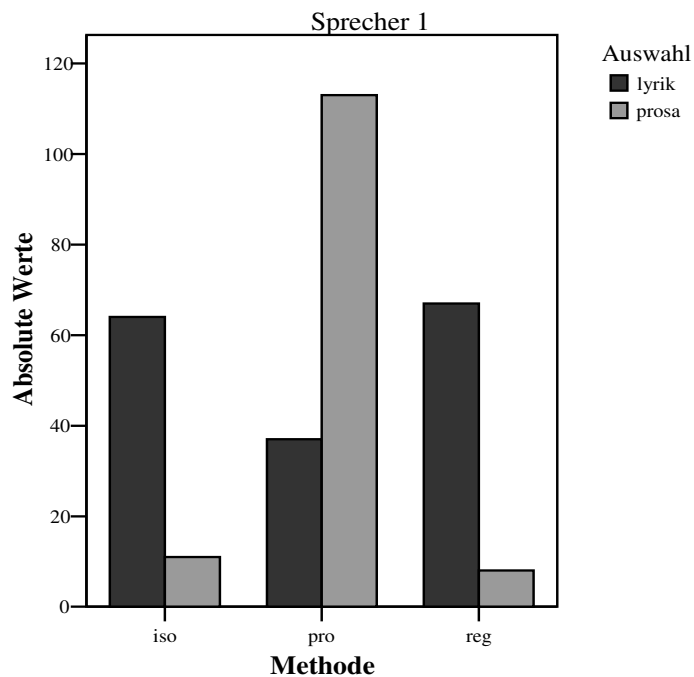
Die Abbildung 8.12 bestätigt, dass die der Intention entsprechenden Sprechstilurteile sowohl für die auf Basis der Isochronie als auch auf Basis der linearen Regression manipulierten Stimuli gefällt wurden. Die Differenzen der Urteile sind für alle drei Kategorien nach einem Chi-Quadrat-Test statistisch signifikant (alle drei Fälle:  $p < 0,001$ ). Damit muss allerdings erneut Hypothese 3 widersprochen werden. Denn auch im Wiederholungsexperiment kann nicht gezeigt werden, dass die isochron gesetzten Stimuli eher als Lyrik identifiziert werden als die durch Regression dauermanipulierten Stimuli.

Bei allen vier Sprechern haben die Hörer der Intention gemäß geurteilt. Die Abbildung 8.13 zeigt jedoch, dass dieser Unterschied bei Sprecher 3 sehr knapp ausfällt und nicht statistisch signifikant ist. Die Abbildung 8.14 dagegen zeigt exemplarisch anhand des Ergebnisses für Sprecherin 1, wie klar die Verteilungsdifferenzen zugunsten der intendierten Kategorien für die anderen drei Sprecher gelagert sind. Das knappere Ergebnis für Sprecher 3, der alle im Test dargebotenen Jamben gesprochen hat, kann neben sprecherindividuellen Eigenschaften dadurch erklärt werden, dass Jamben möglicherweise vom Hörer schwieriger zu verarbeiten sind als die anderen drei Metren.





**Abb. 8.13:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression) für Sprecher 3



**Abb. 8.14:** Verteilung der von den Hörern ausgewählten Stilbezeichnungen auf die intendierten Stil Kategorien aufgeschlüsselt nach der Manipulationsmethode der Stimuli (iso = Isochronie, pro = Prosa, reg = Regression) für Sprecherin 1

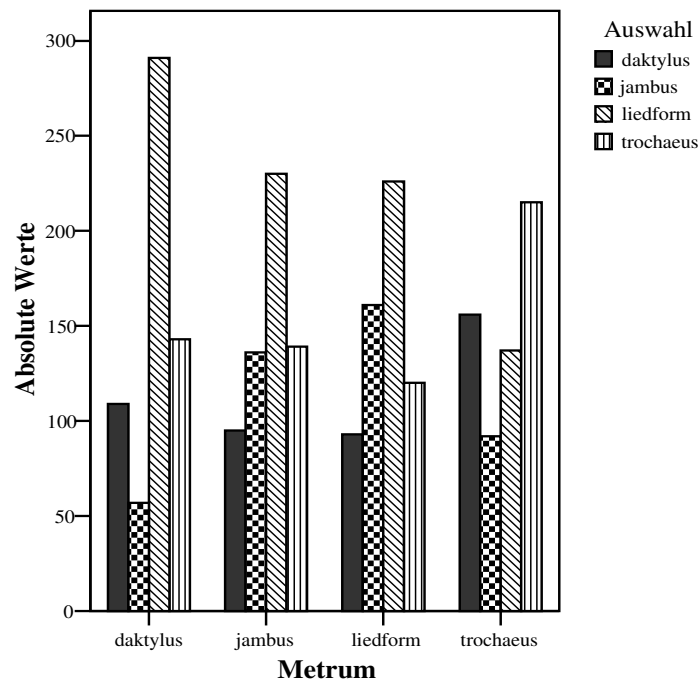
Die wahrscheinlichste Erklärung für die geringen Differenzen aber ist die künstlerisch besonders ausgeprägte Leseweise des Sprechers 3, der ein professioneller Schauspieler ist. Diese besonders lebendige Sprechweise zeigte Sprecher 3 sowohl bei der Vorgabe lyrisch zu lesen als auch bei der Vorgabe prosaisch zu lesen. Es bestand damit bei letzterem Sprecher eine klare Tendenz dahin, das was allgemein eher als lyrisches Lesen empfunden wird auch auf die Prosa anzuwenden. Die Differenzen zwischen beiden Stilen verschwammen damit. Die besondere Leseweise von Sprecher 3 zeichnet sich vor allem durch eine große Modulationstiefe bezüglich der Intonation und der Dauer sowie einer starken Phrasierung mit langen Sprechpausen auch an unerwarteten Stellen im Sprechfluss aus.

Auf eine Detailanalyse der Häufigkeitsverteilungen einzelner Hörer soll an dieser Stelle verzichtet werden, da sich die Deutlichkeit der oben dargestellten Ergebnisse auch in den Einzelurteilen niederschlägt. Die Korrelationen sollen jedoch aufgelistet werden. Der Korrelationskoeffizient nach Spearman-Rho zwischen der intendierten Stil­kategorie und der tatsächlichen Auswahl beträgt  $r = 0,407$ . Die Korrelation zwischen diesen beiden Variablen ist also nur gering. Des Weiteren sollen nun alle hö­rerspezifischen Korrelationen zwischen den beiden genannten Variablen angeführt werden, die über der Gesamtkorrelation liegen. Hier sind dem Rang nach geordnet die Hörer *ami* mit  $r = 0,736$  (hohe Korrelation), *tal* mit  $r = 0,627$  (mittlere Korrelation), *dvz* mit  $r = 0,625$  (mittlere Korrelation), *sgr* mit  $r = 0,580$  (mittlere Korrelation), *rde* mit  $r = 0,488$  (geringe Korrelation), *smi* mit  $r = 0,475$  (geringe Korrelation) und *jma* mit  $r = 0,455$  (geringe Korrelation) zu nennen. Damit liegen die Korrelationskoeffizienten von sieben der fünfzehn Hörer über der Gesamtkorrelation mit drei geringen, drei mittleren und sogar einer hohen Korrelation.

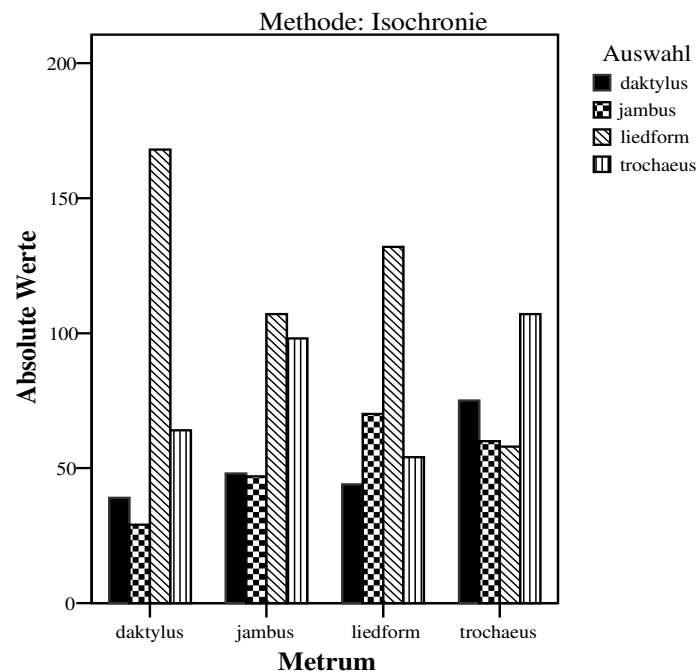
### **8.3.3 Ergebnisse der Rhythmuserkennung bei delexikalisierten Stimuli**

In diesem Abschnitt soll der zweite Perzeptionstest der im Rahmen dieser Arbeit durchgeführten Versuchsreihe diskutiert werden. Es sollte vor allem geprüft werden, ob die Rhythmen Jambus, Trochäus, Daktylus und Liedform anhand delexikalisierter und monotonisierter Stimuli immer noch identifizierbar bleiben. Die Teststimuli, die auf Basis der für die Lyrik konstruierten Dauervorhersagemodelle manipuliert wurden, sollten letztere Erkennungsleistung ermöglichen bzw. aufrechterhalten (s. Hypothese 2). Die Abbildung 8.15

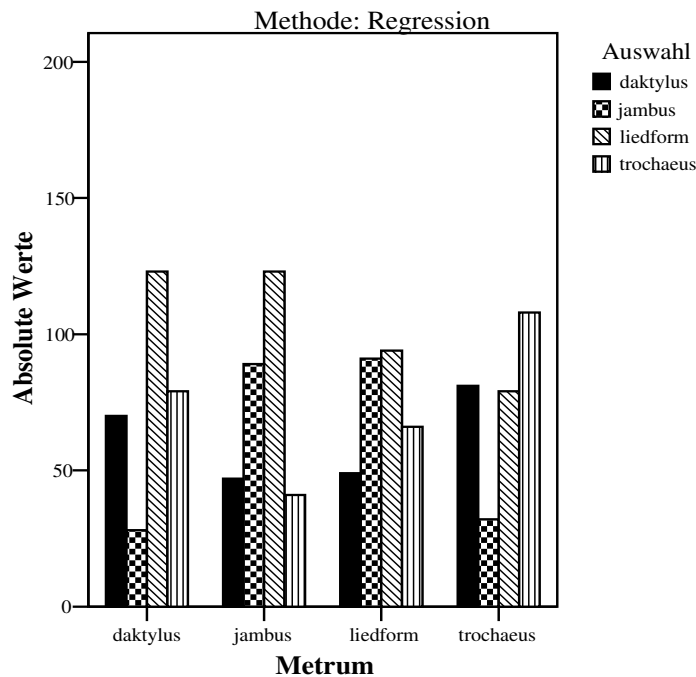
zeigt die Zuweisung der Rhythmuskategorien zu den zugrunde liegenden Metren durch die Hörer. Es zeigt sich, dass der Trochäus und die Liedform mehrheitlich richtig erkannt worden sind.



**Abb. 8.15:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien



**Abb. 8.16:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien bei der Manipulationsmethode der Isochronie

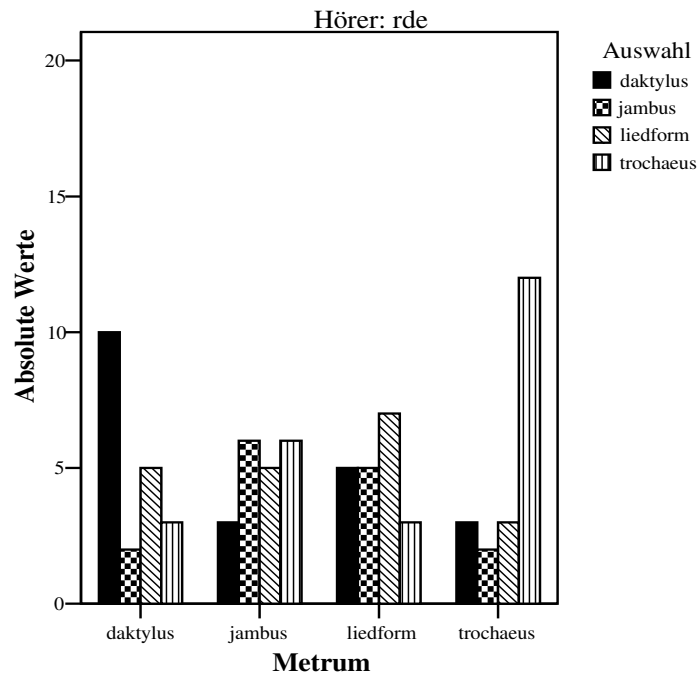


**Abb. 8.17:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien bei der Manipulationsmethode der Regression

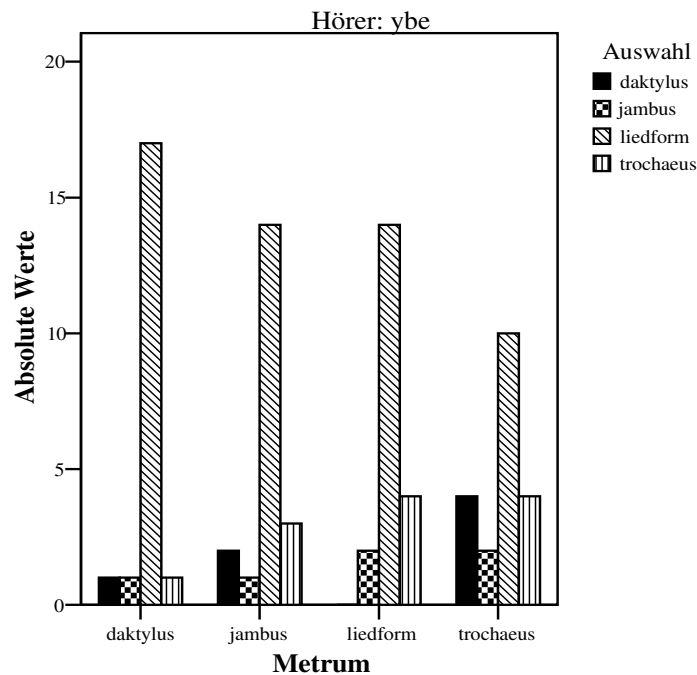
Die Differenzen zwischen den Auswahlklassen innerhalb einer jeweiligen Metrumskategorie sind, wie ein Chi-Quadrat-Test beweist, sämtlich statistisch signifikant ( $p < 0,001$ ). Ähnliche Verhältnisse zeichnen sich in den Abbildungen 8.16 und 8.17 ab. Hier werden die Verteilungen der Auswahlkategorien über die tatsächlichen Metrumklassen von den beiden Manipulationsmethoden der Isochronie auf der einen Seite und der Regression auf der anderen Seite miteinander verglichen. Wie die Diagramme belegen, gibt es keinen wesentlichen Unterschied zwischen den Hörerurteilen bei der Manipulationsmethode der Isochronie gegenüber der Regression. In beiden Fällen werden analog zum Gesamtergebnis der Trochäus und die Liedform mehrheitlich korrekt erkannt. Der Jambus weist im Falle der Regressionsmethode eine Tendenz zur mehrheitlich korrekten Erkennung hin auf. Hypothese 3 muss also auch für den zweiten Test der perzeptiven Versuchsreihe dieser Arbeit abgelehnt werden.

Bei Betrachtung der Ergebnisse einzelner Hörer fällt die sehr unterschiedliche Erfolgsquote auf. Es scheint eine spezielle Fähigkeit zu sein, Rhythmen auf Basis delexikalisierter und

monotonisierter Stimuli zu erkennen. Die Extreme der Urteile seien anhand der Daten von Hörer *rde* in Abbildung 8.18 mit einer äußerst hohen Erfolgsquote und Hörer *ybe* in Abbildung 8.19 mit einer äußerst niedrigen Erfolgsquote illustriert.



**Abb. 8.18:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien bei Hörer *rde*



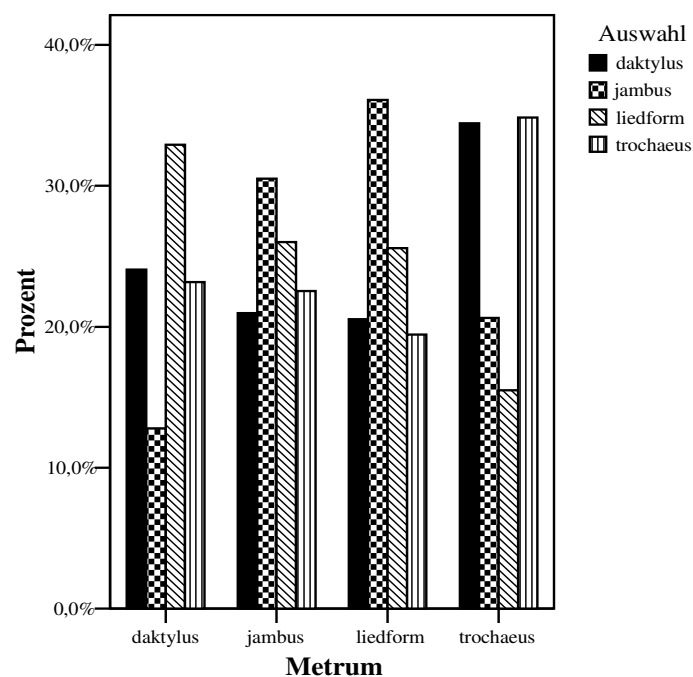
**Abb. 8.19:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien bei Hörer *ybe*

Wie sich zeigt, hat Hörer *rde* alle vier Metren mehrheitlich korrekt identifiziert. Lediglich die Kategorie *Jambus* ist gleich oft als Jambus und als Trochäus beurteilt worden. Hörer *ybe* dagegen hat alle vier Kategorien mehrheitlich als Liedform identifiziert. Lediglich im Falle des Trochäus zeigt sich eine Tendenz hin zur mehrheitlich korrekten Klassifikation. Die übermäßige Verwendung der Kategorie *Liedform* für alle vier im Test dargebotenen Rhythmen ist ein Phänomen, das mehrere Hörer aufweisen. Es sind insgesamt 5 von 30 Probanden, die allen vier Metrumklassen mehrheitlich die Kategorie Liedform zugewiesen haben. Es stellt sich nun die Frage, warum die Hörer so oft fälschlicherweise auf die Liedform ausgewichen sind. Die Liedform zeichnet sich im Gegensatz zu den anderen drei Metren dadurch aus, dass sie eine variable Anzahl von unbetonten Silben innerhalb eines Versfußes besitzt. Die Probanden sind aber vor dem Test instruiert worden, trotz der Sonderstellung der Liedform, diese nicht als Ausweichkategorie zu verwenden, wenn die Identifikation eines regelmäßigen Metrums schwierig ist. Da das oben genannte Sechstel der Probanden dennoch übermäßig häufig die Liedform wählte, kann spekuliert werden, dass sie tatsächlich bei regelmäßigen Fällen, d.h. Jambus, Trochäus und Daktylus, eine unregelmäßige Anzahl von unbetonten Silben innerhalb eines Versfußes ausgemacht haben. Es muss festgestellt werden, dass die Aufgabenstellung an dieser Stelle sehr schwierig war, da Silben- und Fußgrenzen bei den verwandten delexikalisierten und monotonisierten Stimuli wohl kaum zu identifizieren waren. Die Stimuli konnten damit nicht gezielt nach Grenzen abgehört werden, sondern die Hörer mussten sich auf den Gesamteindruck verlassen. Dass die Klassifikationsaufgabe auf diesem Wege lösbar war, zeigt ja Hörer *rde*.

Überdies ist untersucht worden, ob es einen Urteileffekt bezüglich der Geschultheit der Probanden gibt. Es sind hierzu die beiden Gruppen der Studierenden der Phonetik und die Hobbymusiker separat analysiert worden. Die Phonetiker umfassten dabei 14 Personen, die Hobbymusiker hingegen zählten 16 Personen. Beide Gruppen zeigten erneut den Effekt, dass Trochäus und Liedform mehrheitlich im Sinne des Dauervorhersagemodells klassifiziert wurden. Dabei fiel das Ergebnis für die Hobbymusiker jedoch deutlicher aus als für die Phonetiker. Ein Chi-Quadrat-Test wies nach, dass die Differenzen im Falle des Trochäus für die Phonetikstudenten statistisch signifikant sind ( $p < 0,05$ ). Die Differenzen in der Kategorie des Trochäus für die Musiker sind sogar statistisch höchst signifikant ( $p < 0,001$ ). Alle übrigen gemessenen Differenzen in allen vier Kategorien sind sowohl für die Phonetikstudenten als

auch die Hobbymusiker ebenfalls statistisch höchst signifikant ( $p < 0,001$ ). Somit zeigt sich, dass es den Hobbymusikern deutlich leichter fiel, den Trochäus bei delexikalisierten und monotonisierten Stimuli zu erkennen, als die übrigen drei Metren (abgesehen von dem Spezialfall der Liedform).

Es soll im Folgenden nochmals die übermäßige Verwendung der Kategorie Liedform zur Klassifikation der Teststimuli aufgegriffen werden. Denn neben den fünf Probanden, die alle vier Metren mehrheitlich als Liedform identifizierten, fiel auf, dass auch die übrigen Hörer die Kategorie Liedform signifikant häufig verwandten. Die erörterten Daten sollen deshalb aus einem anderen Blickwinkel betrachtet werden. Es soll nachstehend dargestellt werden, wie groß der prozentuale Anteil der Häufigkeit einer gewählten Kategorie innerhalb einer Metrumklasse an der Gesamtheit dieser Urteilkategorie ist. Dadurch kann unter anderem dokumentiert werden, ob eine Urteilkategorie, auch wenn sie selten benutzt wurde, dann aber zumindest in diesen wenigen Fällen zum größten Teil richtigerweise gewählt wurde. Die Abbildung 8.20 belegt, dass die Treffsicherheit der Urteile beim Trochäus und Jambus unter dem zuvor genannten Aspekt am größten ist.



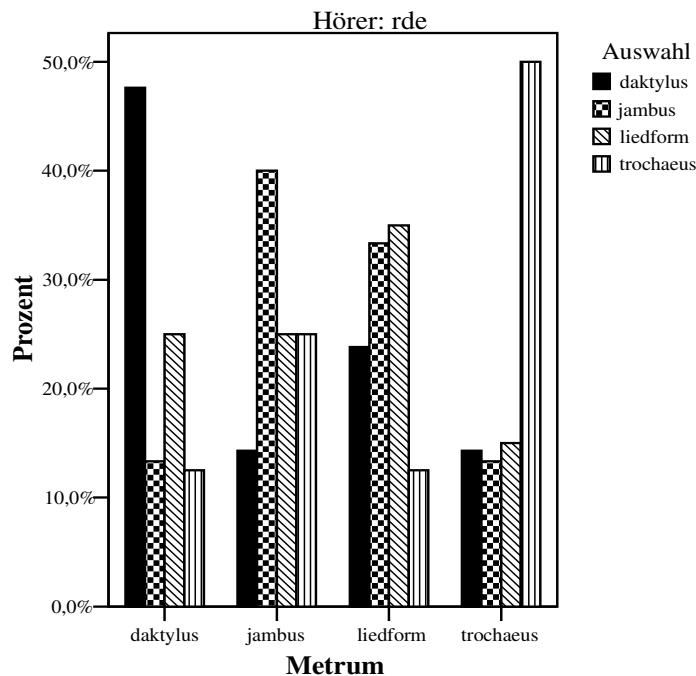
**Abb. 8.20:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien in Prozent

Somit lagen die Hörer, wenn sie sich denn für Jambus oder Trochäus entschieden haben, auch wirklich größtenteils richtig in ihrem Urteil. Es sind also die beiden einfachen, d.h.

zweisilbigen Metren, die gegenüber dem dreisilbigen Daktylus und der bezüglich der Silbenzahl variierenden Liedform besonders gut erkannt wurden. Daktylus und Liedform sind aber immerhin noch prozentual jeweils am zweithäufigsten (von vier Auswahlkategorien) identifiziert worden. Gemessen am Schwierigkeitsgrad der den Probanden gestellten Aufgabe ist dieses sicherlich ein beachtliches Ergebnis.

Die Metrumskategorie Daktylus wurde am ehesten mit der Liedform, die Kategorie Liedform mit dem Jambus verwechselt. Die Verwechslung des Daktylus mit der Liedform kann dadurch erklärt werden, dass ein Versfuß in Liedform in vielen Fällen so wie der Daktylus zwei unbetonte Silben beinhaltet. Warum aber die Liedform so oft als Jambus erkannt wurde, kann unter rhythmischen Aspekten nicht erklärt werden, zumal der Jambus die Metrumskategorie ist, die mit größter Sicherheit erkannt wurde.

Die prozentualen Anteile der Urteile verteilt auf die einzelnen Metrumskategorien sollen an dieser Stelle auch für den erfolgreichsten Hörer *rde* grafisch dargestellt werden. Die Abbildung 8.21 untermauert, dass Hörer *rde* auch unter diesem Blickwinkel alle vier Metren zum größten Teil korrekt klassifiziert hat.



**Abb. 8.21:** Verteilung der von den Hörern ausgewählten Rhythmusklassen auf die intendierten Metrumskategorien in Prozent für Hörer *rde*



Insbesondere die Kategorien Daktylus, Jambus und Trochäus sind mit überzeugender Mehrheit korrekt identifiziert worden. Selbst die problematische Kategorie der Liedform nimmt noch den prozentualen Spitzenplatz in der Verteilung der Urteile ein. Es wird damit klar, dass es sich bei der Leistung von Hörer *rde* wohl kaum um Zufall handeln kann. Es sei also nochmals betont, dass es für die Lösung der in Test zwei gestellten Aufgabe anscheinend einer Spezialfertigkeit bedarf. Es kann vermutet werden, dass diese Fähigkeit in der Sensibilität und dem analytischen Vermögen des Gehörs eines Probanden begründet liegt.

Bei zukünftigen Perzeptionstests unter Verwendung von delexikalisierten und monotonisierten Stimuli sollte vielleicht ein Vortest durchgeführt werden, der die auditive Diskriminationsfähigkeit eines potenziellen Probanden misst. Es könnte dabei jedoch sehr aufwendig sein, genügend Testkandidaten zu finden. Obige Erörterung hat ja gezeigt, dass nur eine kleine Minderheit über die gewünschten Fähigkeiten verfügt. Die Analysen dieser Arbeit haben außerdem gezeigt, dass Musiker (auf Hobbyniveau) tendenziell eher über die geforderten Fähigkeiten verfügen als die (noch nicht voll ausgebildeten) Phonetiker. Die Hypothese 2 kann für den zweiten Perzeptionstest der Versuchsreihe der vorliegenden Arbeit also mit den oben erläuterten Einschränkungen bestätigt werden.

### **8.3.4 Zusammenfassung**

In Abschnitt 8.3 wurden drei Hypothesen in drei Perzeptionstests geprüft. Hypothese 1 musste im ersten Perzeptionsexperiment zurückgewiesen werden, da es der Gesamtheit der befragten Hörer nicht gelang, den Sprechstilunterschied zwischen Lyrik und Prosa bei der Darbietung der Signale in delexikalisierter und monotonisierter Form zu identifizieren. Zudem musste auch Hypothese 3 zurückgewiesen werden, da die Hörer in der Gesamtheit keine bessere Erkennungsleistung für die nach dem Isochroniemodell manipulierten Stimuli aufwiesen. Es gibt aber sehr wohl eine sprecher- und hörererabhängige individuelle Tendenz hin zur Bestätigung der beiden Hypothesen für Test eins. Es zeigte sich insbesondere, dass die trochäischen Stimuli mehrheitlich korrekt klassifiziert wurden.

Bei der Wiederholung des ersten Perzeptionstests unter veränderten Bedingungen zeichnete sich ein anderes Bild ab. Bei der Verwendung von resynthetisierten Stimuli unter Verzicht

auf die Delexikalisierung und Monotonisierung konnte die Hypothese 1 bestätigt werden. Es war den Hörern unter diesen Konditionen mit großer Deutlichkeit möglich, die Sprechstilunterscheidung zwischen lyrischer und prosaischer Darbietung der Intention gemäß mehrheitlich korrekt vorzunehmen. Lediglich die Hypothese 3 musste auch für diesen Versuch abgelehnt werden. Zwei der in den Stimuli berücksichtigten Sprecher waren Schauspieler, von denen einer ein sehr knappes bestätigendes Ergebnis für Hypothese 1 aufwies. Bei Abhören dieser Sprachsignale fiel auf, dass dieser Sprecher besonders lebendig mit großer Modulationstiefe bezüglich der prosodischen Merkmale und einer ausgeprägten Phrasierung mit häufig auftretenden und langen Pausen las. Diese sehr künstlerisch wirkende Sprechweise bildet eine eigene Sprechstil-kategorie, die über die Kategorien von Lyrik und Prosa hinausgeht. Der betreffende Sprecher wandte diesen Stil sowohl bei geforderter lyrischer als auch prosaischer Leseweise an. Wie bereits erörtert, gelang es den Hörern dennoch mit knappem Ergebnis den Unterschied zwischen Lyrik und Prosa herauszuhören. Die Ergebnisse für die anderen drei Sprecher fallen allerdings wesentlich deutlicher aus.

Der zweite Perzeptionstest sollte unter anderem die Hypothesen 2 und 3 prüfen. Die Hypothese 3 musste erneut, wie schon im Falle des ersten Tests, abgelehnt werden. Denn auch bei der Aufgabe die Rhythmen Jambus, Trochäus, Daktylus und Liedform bei delexikalisierten und monotonisierten Stimuli zu erkennen, war die Klassifikation der isochronen Stimuli nicht erfolgreicher als die der nach der linearen Regression manipulierten Stimuli. Die Hypothese 2 jedoch konnte mit Einschränkungen bestätigt werden. Wie schon beim Ergebnis des ersten Tests waren es insbesondere die trochäischen Stimuli, die besonders häufig korrekt klassifiziert wurden. Darüber hinaus wurde die Liedform besonders sicher erkannt, was aber zum Teil auf eine Überverwendung dieser Kategorie zurückzuführen ist. Überdies muss festgestellt werden, dass es, wie auch Benkwitz (2004) aus ihren eigenen Untersuchungen folgert, einer speziellen Fähigkeit oder Schulung bedarf, um die im Perzeptionstest gestellte Aufgabe zu lösen. Für die im Rahmen dieser Arbeit verwandten auf den Rhythmus reduzierten Stimuli wurde letzterer Umstand dadurch untermauert, dass es gerade ein Hobbymusiker war, der alle vier Metrumskategorien mehrheitlich korrekt klassifiziert hat. Der zweite Test umfasste als Probanden jeweils etwa zur Hälfte Hobbymusiker und Phonetikstudenten. Die Erkennungsrate liegt auch insgesamt bei den Musikern etwas höher als bei den Phonetikern.

Die die Ergebnisse verzerrende Überverwendung der Kategorie Liedform zur Beurteilung der Stimuli legte nahe, den prozentualen Anteil einer gewählten Kategorie an der intendierten Kategorie zu betrachten. Unter diesem Blickwinkel waren es in der Gesamtheit vor allem die einfachen Metren von Jambus und Trochäus, die mit dem jeweils höchsten Prozentsatz innerhalb einer Kategorie korrekt identifiziert wurden. Zu Verwechslungen von Metren kam es insbesondere beim Daktylus und der Liedform. Dabei wurde die Kategorie Daktylus vor allem mit der Liedform und die Kategorie Liedform vor allem mit dem Jambus verwechselt. Die Erklärung hierfür liegt sicherlich in der häufig gleich großen Silbenzahl innerhalb eines Fußes von Daktylus und Liedform begründet. Warum aber die Liedform so oft für einen Jambus gehalten wurde, kann auf Basis rhythmischer Begebenheiten nicht erklärt werden. Trotz der genannten Verwechslungen aber wurden auch Daktylus und Liedform mit dem jeweils zweithöchsten Prozentsatz korrekt identifiziert.

## 9 Abschließende Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit ist, die Dauern für Silben gelesen im lyrischen Sprechstil zu modellieren und diese Modelle zu evaluieren. Für die Modellbildung sind Statistiken basierend auf einem umfangreichen Korpus erstellt worden. Das Korpus umfasst die vier Metren Jambus mit drei Texten, Trochäus mit ebenfalls drei Texten, Daktylus mit zwei Texten und Liedform mit zwei Texten. Der im Deutschen äußerst selten auftretende Anapäst wurde nicht berücksichtigt, da er kein für das Deutsche repräsentatives Metrum darstellt. Neben zwei Zeitungsausschnitten im prosaischen Stil sind die zehn Gedichttexte jeweils einmal im prosaischen (mit verändertem Textsatz) und einmal im lyrischen (originaler Textsatz) Sprechstil gelesen und aufgenommen worden. Die statistische Auswertung des Korpus legte nahe, Dauermodelle nach zwei unterschiedlichen Strategien zu konstruieren. Die eine Methode ist die der linearen Regression, die andere ist die der Berechnung auf Basis von Fußisochronie. Es stellte sich die Aufgabe, nach der Auswertung der akustisch-prosodischen Eigenschaften des Lyrik-Korpus, die charakteristischen Dauern für jedes der vier Metren in den Prädiktionsmodellen abzubilden. Da es in erster Linie darum ging, die Dauereigenschaften der verschiedenen Metren zu modellieren, bewegt sich das Vorhaben der vorliegenden Arbeit im Bereich der Sprechrhythmusforschung.

Der Begriff Sprechrhythmus ist schwierig zu definieren, da es sich um ein Komplexphänomen handelt (s. Kap. 2.3). Der Rhythmus konstituiert sich in erster Linie aus den drei prosodischen Parametern der Tonhöhe, der Lautstärke und der Dauer. Außerdem stellen zahlreiche Forscher einen Zusammenhang zwischen Sprechrhythmus und segmentaler Ebene, d.h. der Abfolge schwerer und leichter Silben, her. Alle genannten Parameter korrelieren stark untereinander und sind perzeptiv kaum voneinander isolierbar. Hier setzt die vorliegende Arbeit an, indem der Rhythmus trotz seiner Komplexität als Reduktion der Prosodie auf den Parameter der Dauer verstanden wird. Die Delexikalisierung in Anlehnung an die PURR-Methode und die Monotonisierung von Signalausschnitten aus dem Korpus machen die Reduktion auf den Rhythmus, verstanden als Dauervariation, möglich. So konnten Perzeptionstests durchgeführt werden, die ausschließlich die Wirkung der Dauervariation von Stimuli maßen. Diese Voraussetzung war notwendig, um die beiden Dauermodelle zur Vorhersage der Dauervariation von Metren, respektive Rhythmen, zu verwenden und anschließend bezüglich ihrer Vorhersageleistung zu evaluieren (s. Kap. 8). Es

sei an dieser Stelle angemerkt, dass eine Definition des Sprechrhythmus einzig und allein auf Basis des Dauerparameters der Komplexität des Phänomens nicht gerecht wird. Die Reduktion auf diesen Parameter und die damit verbundene eingeschränkte Sichtweise ist jedoch wegen der reduzierten Gestalt der Dauermodelle zweckmäßig. Rhythmus wird demgemäß einfach als Realisation eines Metrums in Form einer metrumsspezifischen Silbendauervariation verstanden. Eine Anlehnung an die *langue* vs. *parole* Dichotomie liegt hier vor. Im Fokus der vorliegenden Arbeit liegt die Lyrik, da sich der Sprechrhythmus, wie Kapitel 2.4 deutlich zeigt, hier klarer manifestieren und transparenter sein soll.

Das Kapitel 3 der vorliegenden Arbeit diskutiert wegen seines engen Bezuges zum Sprechrhythmus kurz den Begriff der Prosodie, ehe dann in Kapitel 4 die literaturwissenschaftliche Sichtweise der Metrik kurz erörtert wird. Kapitel 5 skizziert die Aufnahme und Annotation des Lyrik-Korpus. Es sind zehn Gedicht- und zwei Zeitungsausschnitte von 12 Sprechern gelesen und aufgenommen worden. Die so gewonnenen Sprachsignale sind unter anderem silbenbasiert transkribiert und GToBI-annotiert worden.

Im Kapitel 6 dieser Arbeit sind die Ergebnisse der statistischen Datenanalyse des Lyrik-Korpus dargestellt. Bezüglich der Silbendauern der Lyrikvarianten zeigt sich eine steile Flanke in der Häufigkeitsverteilung hin zu niedrigen Dauern. Dieser Effekt ist bei den Fußdauern nicht zu beobachten. Die Dauern der Phrasen und Verse liegen im Durchschnitt in der Größenordnung von *action units*. Ein semantisch bedeutsamer Zusammenhang zwischen Lyrik und Handlung lässt sich damit herstellen. Des Weiteren weisen die nichtprofessionellen Sprecher gegenüber den professionellen Sprechern eine stärkere Tendenz zur Isochronie auf. Es konnte kein weiterer Faktor im Korpus ausgemacht werden, für den diese Tendenz nachgewiesen werden kann. Die Dauerrelation zwischen dem unakzentuierten und dem akzentuierten Teil eines Fußes (Korpusvariable *AccSylQuot*) stellte sich als besonders hilfreich für die Dauermodellierung auf Basis der Isochroniehypothese heraus. Kapitel 7 knüpft hier an. Für die Intonation ist nach einer Z-Normalisierung der Daten nachgewiesen worden, dass die Intonation für den lyrischen Sprechstil signifikant höher liegt als für den prosaischen Sprechstil. Die höhere Intonation ist damit charakteristisch für eine spezielle dichterische Attitüde beim Lesen von Lyrik. Für eine GToBI Modellierung der Intonation im

Lyrik-Korpus bot sich eine metrum- und positionsabhängige Beschreibung an, die in Kapitel 7 aufgegriffen wird.

Die Modellierung der Silbendauern für den lyrischen Sprechstil ist in Kapitel 7 dieser Arbeit erörtert. Es sind dabei die beiden Verfahren der linearen Regression und einer auf der Isochronie von Füßen basierenden Methode verwandt worden. Das Ziel war, die Dauern metrumsspezifisch vorherzusagen. Die konkreten Berechnungsvorschriften für die lineare Regression können in Kapitel 7.1.1, die für die Isochronie in Kapitel 7.1.2 nachvollzogen werden. Die lineare Regression hängt von nur drei Faktoren (*Acce*, *PhonesInSyl* und *PhrasePosCat*) ab. Damit berechnen sich die durch die lineare Regression vorhergesagten Silbendauern aus der Akzentuierbarkeit einer Silbe, der Zahl der Phone in der Silbe und der Position der Silbe innerhalb der Phrase (initial, medial oder final). Ein Vergleich mit den Ergebnissen für die Dauerprädiktion in der Sprachsynthese zeigt, dass es auch hier die drei oben genannten Faktoren sind, die die Silbendauer im Wesentlichen bestimmen. Die spezifischen Informationen über den lyrischen Sprechstil liegen somit eher in den Koeffizienten als in den Variablenausprägungen der jeweiligen Formel für die lineare Regression.

Die beiden in Kapitel 7 skizzierten Vorhersagemodelle für Silbendauern im lyrischen Sprechstil sind bezüglich ihrer Wirkung auf Hörer getestet worden. Die Ergebnisse der drei Tests sind in Kapitel 8 dargelegt. Die Leistungsfähigkeit der Modelle wurde anhand dreier Hypothesen bewertet:

- 1) Die auf Basis der beiden Dauermodelle manipulierten Sprachstimuli lassen es auch in delexikalierter und monotonisierter Form immer noch zu, dass der Sprechstilunterschied zwischen Lyrik und Prosa von Hörern identifiziert wird (In einem Wiederholungsexperiment wurde auf die Delexikalisierung und Monotonisierung verzichtet).
- 2) Die auf Basis der beiden Dauermodelle manipulierten Sprachstimuli lassen es auch in delexikalierter und monotonisierter Form immer noch zu, dass die vier Rhythmen Jambus, Trochäus, Daktylus und Liedform von Hörern identifiziert werden.
- 3) Die Isochronie ist ein perzeptives Phänomen, so dass die beiden in den Hypothesen 1) und 2) formulierten Klassifikationsaufgaben in den Fällen einer Dauermanipulation auf Basis der Isochroniehypothese erleichtert werden.

Für die drei Perzeptionstests wurden delexikalisierte (angelehnt an die PURR-Methode) und monotonisierte Stimuli verwendet. Die Ergebnisse von Perzeptionsexperiment 1 (Sprechstilunterscheidung Lyrik/Prosa) führten zu einer Ablehnung der Hypothesen 1 und 3 bei einer schwachen sprecher- und hörerabhängigen Tendenz hin zur Annahme der beiden genannten Hypothesen. Bei der Wiederholung des ersten Tests mit Stimuli, die nicht delexikalisiert und nicht monotonisiert, sondern lediglich zur Normalisierung der Signalqualität nach der LPC-Methode resynthetisiert wurden, zeigte sich, dass Hypothese 1 für diesen Fall klar angenommen werden muss. Lediglich Hypothese 3 musste erneut abgelehnt werden. Auch für den Perzeptionstest 2 (Klassifikation der Stimuli nach den vier Metren) musste die Hypothese 3 abgelehnt werden. Hypothese 2 kann dagegen mit Einschränkungen angenommen werden. Trochäus und Liedform wurden mehrheitlich korrekt ihrer jeweiligen Metrumskategorie zugeordnet. Allerdings wurde die Kategorie Liedform wesentlich häufiger zur Klassifikation eines Stimulus herangezogen als die übrigen drei Metren. Deshalb sind auch die prozentualen Anteile einer Auswahl innerhalb einer Metrumskategorie an der Gesamtzahl der jeweiligen Auswahl ausgewertet worden. Hierbei zeigte sich vor allem, dass die einfachen Metren Jambus und Trochäus (nur eine unakzentuierte Silbe pro Fuß) besonders sicher korrekt identifiziert wurden. Außerdem zeigte sich, dass es von der Geschultheit eines Hörers abhängt, wie erfolgreich er die Klassifikationsaufgabe erledigt. Die Gruppe der Probanden teilte sich zu etwa gleichen Teilen in Phonetikstudenten und Hobbymusiker. Es war gerade ein Hobbymusiker, der alle vier Metren mehrheitlich korrekt identifizierte. Dieses gilt sowohl bei der Betrachtung der absoluten Häufigkeiten, als auch bei der Betrachtung der prozentualen Anteile einer Auswahl.

Die Ergebnisse der drei durchgeführten Perzeptionstests besagen also, dass der Sprechstilunterschied zwischen Lyrik und Prosa bei der Darbietung von delexikalisierten und monotonisierten Stimuli nicht mehr erkannt werden kann. Das Gegenteil ist bei den nicht-delexikalisierten, nicht-monotonisierten aber resynthetisierten Stimuli der Fall. Hier war die Sprechstilunterscheidung mit großer Deutlichkeit erfolgreich. Zudem wurde die Aufgabe, bei der Darbietung delexikalisierter und monotonisierter Stimuli die vier Metren zu erkennen, nur zum Teil korrekt erledigt. Die Hälfte der Metren wurde mehrheitlich korrekt identifiziert. Besonders interessant für das Vorhaben der vorliegenden Arbeit ist die generelle Ablehnung

von Hypothese 3. Diese könnte zwar für einzelne Hörer bzw. Sprecher angenommen werden, doch sind diese Personen klar in der Minderheit. Das heißt also, dass die Isochronie von Füßen nicht die erwartete Funktion einer erleichterten Erkennung von Rhythmen erfüllt. In der Literatur zum Thema Sprechrhythmus (s. Kap. 2) wird häufig erwähnt, dass die Isochronie ein perzeptives Phänomen ist. Unter den oben erörterten Rahmenbedingungen der Perzeptionstests dieser Arbeit muss festgestellt werden, dass die Isochronie kein perzeptives Phänomen ist. Zumindest werden Silbendauern auf Basis von Fußisochronie nicht als rhythmischer empfunden als die Silbendauern auf Basis einer linearen Regression. Dieses Fazit muss selbstverständlich vor dem Hintergrund betrachtet werden, dass Rhythmus in der vorliegenden Arbeit als Realisation eines Metrums bei lyrischem Sprechstil verstanden wird, und dass davon ausgegangen wird, dass die rhythmische Variabilität hinreichend für eine Sprechstilunterscheidung zwischen Lyrik und Prosa ist. Es muss künftig überlegt werden, ob ein verändertes Testdesign zu anderen Resultaten führt.

Die in dieser Arbeit skizzierten Ergebnisse sind der Grundlagenforschung zuzurechnen. Demgemäß werden einige Fragen aufgeworfen, die in diesem Rahmen nicht beantwortet werden können. Die Arbeit mit delexikalisierten und auch monotonisierten Stimuli hat sich in der Phonetik bereits bewährt (s. Kap. 8.1). Die Experimente und Analysen der vorliegenden Arbeit befassen sich jedoch ausschließlich mit der sehr engen Sprachdomäne der Lyrik mit festgelegtem Metrum. Es sollte in Zukunft geprüft werden, ob die Ergebnisse dieser Arbeit auch für andere Sprachdomänen (z.B. Prosatexte) und Sprechstile (z.B. Spontansprache) Gültigkeit besitzen. Des Weiteren kann das für diese Arbeit erstellte Lyrik-Korpus anderen Forschern zur Verfügung gestellt werden. Im Zuge dessen können die in der vorliegenden Arbeit weitestgehend unberücksichtigten Effekte von Intonation und Intensität auf den Sprechrhythmus von Lyrik erforscht werden. Auch im Bereich der Bedeutung der segmentalen Ebene auf den Rhythmus muss noch gearbeitet werden.

Die in dieser Arbeit erläuterte Korpusanalyse und Dauermodellierung zielen nicht zuletzt darauf ab, die sprachliche Domäne der Lyrik für die Sprachsynthese zugänglich zu machen. Es sollten unter anderem grundlegende Erkenntnisse für eine Lyriksynthese geschaffen werden. Die Dauermodelle können als Prädiktionsmodul für Silbendauern in einem Text-To-Speech System verwandt werden. Künftig ist die Prosodieprädiktion jedoch noch um ein Intonationsmodul zu erweitern. Erste Versuche zur Intonationssteuerung durch GToBI-



Symbolfolgen liegen in dieser Arbeit vor. Es zeigte sich jedoch, dass ein Prosamodul zur Überführung von GToBI-Etiketten in konkrete Grundfrequenzwerte für die Intonationsprädiktion in der Lyrik nicht ausreichend ist. Für diesen Zweck muss ein eigenes Lyrikmodul zur Generierung von F0-Werten erstellt und implementiert werden.

## Literatur

- Abercrombie, D. (1967): *Elements of General Phonetics*. Edinburgh: University Press.
- Allen, G. D. (1972): „The Location of Rhythmic Stress Beats in English I & II.” *Language & Speech*, 15, 72-100, 179-195.
- Allen, G. D. (1975): „Speech rhythm: Its relation to performance universals and articulatory timing.” *Journal of Phonetics*, 3, 75-86.
- Auer, P. & Uhmann, S. (1988): „Silben- und akzentzählende Sprachen. Literaturüberblick und Diskussion.“ In *Zeitschrift für Sprachwissenschaft* 7, Heft 2, 214-259.
- Barbosa, P. A. (2002): “Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production.” In *Proc. Speech Prosody 2002*, Aix-en-Provence, 163-166.
- Barney, T. (1999): „Readers as Text Processors and Performers. A new Formula for Poetic Intonation.” *Discourse Processes*, 28, 155-168.
- Baumann, S.; Grice M. & Benz Müller, R. (2001): “GToBI – a Phonological System for the Transcription of German Intonation.” In Puppel, Stanislaw & Grazyna Demenko (Hrsg.) *Prosody 2000, Speech Recognition and Synthesis*, Posen: Adam Mickiewicz Universität, Faculty of Modern Languages and Literature, 21-28.
- Binder, A.; Haberkamm, K.; Kahrmann, C.; Reiß, G.; Richartz, H.; Schluchter, M.; Steinberg, G. (1984): *Einführung in Metrik und Rhetorik*. Scriptor, Frankfurt a. M.
- Boersma, P. & Weenink, D. (2001): “PRAAT, a system for doing phonetics by computer.” *Glott International* 5(9/10), 341-345.
- Bolinger, D. (1965): “Pitch accent and sentence rhythm.” In *Forms of English: Accent, morpheme, order*. I. Abe & T. Kanekiyo (eds.), Cambridge, MA: Harvard University Press, 139-180.
- Boogart, T. & Silverman, K. (1992): Evaluating the overall comprehensibility of speech synthesizers.“ In *Proc. ICSLP*, Alberta, Canada, 2, 1207-1210.
- Braunschweiler, N (2003): *Automatic Detection of Prosodic Cues*. Dissertation, Universität Konstanz, Deutschland.
- Bußmann, H. (1983): *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.
- Carlson, R.; Granström, B. & Larson, K. (1976): “Evaluation of a text-to-speech system as a reading machine for the blind.” In *STL-QPSR*, 2-3, 9-13.
- Cooper, G.; Meyer, L. B. (1960): *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Couper-Kuhlen, E. (1993): *English Speech Rhythm. Form and Function in Everyday Verbal Interaction*. Amsterdam: Benjamins.
- Cummins, F. & Port, R. F. (1998): “Rhythmic constraints on stress timing in English.” *Journal of Phonetics*, 26, 145-171.

- Cummins, F. (2002): "Speech rhythm and rhythmic taxonomy." In *Proc. Speech Prosody 2002*, Aix-en-Provence, 121-126.
- Dauer, R. M. (1983): „Stress-timing and syllable-timing reanalyzed.“ *Journal of Phonetics*, 11, 51-62.
- Dehnel, A. & Klaus, H. (1996): "Ein Vergleich von Urteilsskalen zur Bestimmung der Sprachqualität mit Kategorie – Einschätzungstests.“ In *Sprachkommunikation*, ITG-Fachbericht 139, VDE-Verlag: Berlin, 133-136.
- Dellwo, V. & Wagner, P. (2003): "Relationships between speech rhythm & rate." In *Proc. of the 15th ICPhS*, Barcelona, Spanien, 471-474.
- Delogu, C.; Paoloni, P.; Pocci, P. & Sementina, C. (1991): "Quality evaluation of text-to-speech synthesizers using magnitude estimation, categorical estimation, pair comparison and reaction time methods." In *Proc. Eurospeech*, Genua, Italien, 1, 353-355.
- Duarte, D.; Galves, A.; Lopes, N. & Maronna, R. (2001): "The statistical analysis of acoustic correlates of speech rhythm." In *Rhythmic patterns, parameter setting and language change*, ZiF, Universität Bielefeld.
- Fant, G.; Kruckenberg, A. & Nord, L. (1991): „Durational correlates of stress in Swedish, French and English.“ *Journal of Phonetics*, 19, 351-365.
- Fant, G. & Kruckenberg, A. (1996): „On the Quantal Nature of Speech Timing.“ In *Proc. ICSLP '96*, 2044-2047.
- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.
- Gelfand, S. A. (1998): *Hearing – An Introduction to psychological and physiological acoustics*. Marcel Dekker: NY, Basel, Hong Kong, 3. Auflage.
- Gelfert, H.-D. (1998): *Einführung in die Verslehre*. Stuttgart: Reclam.
- Goldstein, M.; Lindström, B. & Till, O. (1992): „Some aspects on context and response range effects when assessing naturalness of swedish sentences generated by 4 synthesiser systems.“ In *Proc. ICSLP*, Alberta, Canada, 2, 1339-1342.
- Herment-Dujardin, S. & Hirst, D. (2002): "Emphasis in English: a Perceptual Study based on Modified Synthetic Speech." In *Proc. of First International Conference on Speech Prosody*, Aix-en-Provence, Frankreich, 379-382.
- Howell, P. (1984): "An acoustic determinant of perceived and produced anisochrony." In *Proc. ICPhS '84*, M. P. R. van den Broecke & A. Cohen (eds.), 429-433.
- Howell, P. (1988): "Prediction of P-center location from the distribution of energy in the amplitude envelope: Part I & II." *Perception and Psychophysics*, 43, 90-93, 99.
- Hustad, K. C.; Kent, R. D. & Beukelman, D. R. (1998): „DECTalk and MacinTalk Speech Synthesizers: Intelligibility Differences for Three Listener Groups.“ In *JSHR*, 41(4), 744-752.
- ITU-T (1993): „A method for subjective performance assessment of the quality of speech voice output devices.“ Draft ITU-T Recommendation P.85, COM 12-R 6-E

- Janker, P. M. (1995): „Sprechrhythmus, Silbe, Ereignis ? Eine experimentalphonetische Untersuchung zu den psychoakustisch relevanten Parametern zur rhythmischen Gliederung sprechsprachlicher Äußerungen.“ *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation München (FIPKM)*, 33, 1-259.
- Kien, J. & Kemp, A. (1994): “Is speech temporally segmented? Comparison with temporal segmentation in behaviour.” *Brain and Language*, 46, 662-682.
- Kohler, K. (1982): “Rhythmus im Deutschen.” *Arbeitsberichte, Institut für Phonetik der Universität Kiel*, 19, 89-105.
- Kohler, K. (1983): “Stress-Timing and Speech Rate in German. A Production Model.“ *Arbeitsberichte, Institut für Phonetik der Universität Kiel*, 20, 7-53.
- Kruckenberger, A. & Fant, G. (1993): “Iambic versus trochaic patterns in poetry reading.“ *Nordic Prosody VI*, Stockholm, 123-135.
- Ladefoged, P. (1967): *Three Areas of Experimental Phonetics*. London: Oxford University Press.
- Lea, W. A. (1974): *Prosodic aids to speech recognition: IV. A general strategy for prosodically guided speech understanding*. Univac Report PX 10791, St. Paul, Minnesota: Sperry Univac.
- Lehiste, I. (1977): “Isochrony reconsidered.” *Journal of Phonetics*, 5, 253-263.
- Lehiste, I. (1990): “Phonetic investigation of metrical structure in orally produced poetry.” *Journal of Phonetics*, 18, 123-133.
- Lieberman, M. Y. & Streeter, L. A. (1978): “Use of nonsense – syllable mimicry in the study of prosodic phenomena.” In *JASA*, 63(1), 231-233.
- Luce, P. A.; Feustel, T. C. & Pisoni, D. B. (1983): “Capacity Demands in Short-Term Memory for Synthetic and Natural Speech.” In *Human Factors*, 25(1), 17-32.
- Marcus, S. M. (1981): “Acoustic determinants of perceptual center (P-center) location.” *Perception & Psychophysics*, 30, 247-256.
- Martin, J. G. (1979): “Rhythmic and segmental perception are not independent.” *J. Acoust. Soc. Am.*, 65(5), 1286-1297.
- Miller, M. (1984): „On the perception of rhythm.“ *Journal of Phonetics*, 12, 75-83.
- Morton, J.; Marcus, S. M. & Frankish, C. R. (1976): “Perceptual Centers (P-centers).” *Psychological Review*, 83 (5), 405-408.
- Newton, R. P. (1981): *Vowel undersong: studies of vocalic timbre and chroneme patterning in German lyric poetry*. Mouton: The Hague.
- Nord, L.; Kruckenberger, A. & Fant, G. (1990): „Some timing studies of prose, poetry and music.“ *Speech Communication*, 9, 477-483.
- Oberfeld, D. (2000): „Wahrnehmung von Akzentuierungen in Abfolgen kurzer und langer Intervalle bei physikalisch identischen Klängen.“ In *Fortschritte der Akustik – DAGA 2000*, Oldenburg, 238-

- O'Dell, M. & Nieminen, T. (1999): "Coupled Oscillator Model of Speech Rhythm." In *Proc. ICPhS '99*, San Francisco, 1075-1078.
- Pavlovic, C. V.; Rossi, M. & Espesser, R. (1990): "Use of the magnitude estimation technique for assessing the performance of text-to-speech synthesis systems." In *JASA*, 87(1), 373-382.
- Pellegrino, F.; Chauchat, J.-H.; Rakotomalala, R. & Farinas, J. (2002): "Can Automatically Extracted Rhythmic Units Discriminate Among Languages?" In *Proc. Speech Prosody 2002*, Aix-en-Provence, 562-565.
- Portele, T. & Meyer, H. (1994): „Laut- und silbenorientierte Dauersteuerung – zwei Paradigmen im Vergleich.“ In *Fortschritte der Akustik - DAGA'94*, 973-976.
- Pfützinger, H. R. (1996): "Two Approaches to Speech Rate Estimation." In *Proc. SST '96*, Adelaide, 421-426.
- Pfützinger, H.R. (1998): "Local Speech Rate as a Combination of Syllable and Phone Rate." In *Proc. ICSLP '98*, 3, Sydney, 1087-1090.
- Pfützinger, H.R. (1999): "Local Speech Rate Perception in German Speech." In *Proc. ICPhS '99*, 2, San Francisco, 893-896.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. Dissertation, MIT, Bloomington: Indiana Universität Linguistics Club.
- Pike, K. (1945): *Intonation of American English*. University of Michigan Press: Ann Arbor.
- Pisoni, D. B. & Hunnicutt, S. (1980): "Perceptual evaluation of MITalk: The MIT unrestricted text-to-speech-system." In *Proc. IEEE-ICASSP*, 572-575.
- Pisoni, D. B.; Nusbaum, H. C. & Greene, B. G. (1985): "Perception of synthetic speech generated by rule: Intelligibility of eight text-to-speech systems", In *Proc IEEE*, 73(11), 1665-1676.
- Pompino-Marschall, B. (1989): "On the psychoacoustic nature of the P-center Phenomenon." *Journal of Phonetics*, 17, 175-192.
- Port, R.; Tajima, K. & Cummins, F. (1999): "Speech and rhythmic behaviour." In *The Non-linear Analysis of Developmental Processes*. G. J. P. Savelsburgh, H. van der Maas & P. C. L. van Geert (eds.), Amsterdam: Elsevier.
- Ramus, F.; Nespors, M. & Mehler, J. (1999): „Correlates of linguistic rhythm in the speech signal." *Cognition*, 73 (3), 265-292.
- Ramus, F. (2000): „Perception of linguistic rhythm by newborn infants.“ (unveröffentlicht).
- Ramus, F.; Dupoux, E.; Zangl, R. & Mehler, J. (2000): "An empirical study of the perception of language rhythm." (unveröffentlicht).
- Ramus, Franck (2002): "Acoustic correlates of linguistic rhythm: Perspectives." In *Proc. Speech Prosody 2002*, Aix-en-Provence, 115-120.

- Schröder, M. & Trouvain, J. (2001): "The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching." *4th ISCA Workshop on Speech Synthesis*, Blair Atholl, Schottland, 131-136.
- Schröder, M., Hunecke, A. & Krstulovic, S. (2006): „OpenMary - open source unit selection as the basis for research on expressive synthesis." *Blizzard Challenge Workshop 2006*, Pittsburgh, PA, USA.
- Sonntag, G. P. & Portele, T. (1997): „A method for prosody evaluation." *Proceedings of SALT workshop*, Juni 17-18, Sheffield, 188-194.
- Sonntag, G. P. & Portele, T. (1998a): "PURR – a method for prosody evaluation and investigation." *Journal of Computer Speech and Language*, Vol. 12, No. 4, Oktober 1998, Special Issue on Evaluation in Language and Speech Technology, 437-451.
- Sonntag, G. P. & Portele, T. (1998b): "Comparative evaluation of synthetic prosody with the PURR method." *Proceedings of the ICSLP*, Sydney, Vol. 5, 1715-1718.
- Sonntag, G. P. (1999): *Evaluation von Prosodie*. Aachen: Shaker Verlag.
- Stevens, S. S. (1957): "On the psychophysical law." In *Psychological Review*, 64, 153-181.
- Stevens, S. S. (1958): "Problems and methods of psychoacoustics." In *Psychological Bulletin*, 55, 177-196.
- Strangert, E. & Aasa, A. (1996): "Evaluation of Swedish prosody within the MULTEXT-SW project." In *TMH-QPSR*, 2, 37-44.
- Svensson, S.-G. (1971): "A preliminary study of the role of prosodic parameters in speech perception." In: Quarterly Progress and Status Report of the Speech Transmission Laboratory, Stockholm, vol. 2-3, 24-40.
- Toepel, U. & Alter, K. (2002): „Cerebral Strategies in the Segmentation and Interpretation of Speech." Invited Talk at Speech Prosody 2002, Aix-en-Provence.
- Tsur, R. (1997): „Poetic Rhythm: Performance Patterns and Their Acoustic Correlates." *Versification: An Electronic Journal Devoted to Poetic Prosody 1:1*, o.S.
- Tsur, R. (2002a): "A Perception-Oriented Theory of Meter and the Rhythmical Performance of Poetry." Plenary Lecture, in Christoph Küper (ed.), *Meter, Rhythm and Performance: proceedings of the international conference on meter, rhythm and performance* (held in May 1999 in Vechta), Frankfurt am Main: Peter Lang, 19-38.
- Tsur, R. (2002b): „Aspects of Cognitive Poetics." In Elena Semino and Jonathan Culpeper (Hrsg.), *Cognitive Stylistics – Language and Cognition in Text Analysis*, John Benjamins Publishing Company: Amsterdam/Philadelphia, o.S.
- Turner, F. & Pöppel, E. (1983): "The neural lyre: Poetic meter, the brain and time." *Poetry*, August, 277-309.
- Uldall, E. T. (1978): "Rhythm in very rapid R. P." *Language & Speech*, 21, 397-402.

van Bezooijen, R. & Pols, L. C. W. (1990): „Evaluating text-to-speech systems: Some methodological aspects.” In *Speech Communication*, 9(4), 263-270.

Wagner, P. (2002): *Vorhersage und Wahrnehmung deutscher Wortbetonung*. Dissertation. Universität Bonn, elektronisch publiziert.

Wagner, P. & Dellwo, V. (2004): „Introducing YARD (Yet Another Rhythm Determination) And Re-Introducing Isochrony to Rhythm Research.” In *Proc. of Speech Prosody*, Nara, Japan, 227-230.

## Anhang A

Etikett	Definition
Line	Zeilennummer; immer nur auf aktuelle Datei bezogen (für leichteres Rücksortieren in SPSS)
Poem	Name des jeweiligen Gedichts in Kurzform
Speaker	Sprechernamen
Experience	SprecherStatus; N = Naive, P = Professional
Syllable	Silbe in SAMPA-Transkription
Duration	Dauer der Silbe auf volle ms gerundet
SylInPhrase	Silbenposition in Phrase; laufende Nummer der Silbe in der aktuellen Phrase. Pausen = 0
PhraseDur	Dauer der gesamten Phrase, in der die aktuelle Silbe steht, auf ms gerundet
SylInVerse	Silbenposition im Vers; Laufende Nummer der Silbe im Vers. Pausen = 0
VerseDur	Dauer des gesamten Verses, in dem die aktuelle Silbe steht, auf ms gerundet
SylInFoot	Silbenposition im Fuß; laufende Nummer der Silbe im Fuß; Pausen werden sowohl am Rand als auch innerhalb des Fußes mit 0 bezeichnet; Füße können versübergreifend sein (liegt an automatischer Fußerkennung)
SylsInFoot	Silbenanzahl im Fuß; gesamte Anzahl der Silben des Fußes der aktuellen Silbe.
FootDurExclP	Fußdauer ohne Pausen; Dauer des aktuellen Fußes ohne Einbeziehung von Pausen (weder an Rändern noch im Fuß), auf ms gerundet; die Pausen erhalten auch den Wert des aktuellen Fußes.
FootDurInclP	Fußdauer mit Pausen; Dauer des aktuellen Fußes unter Einbeziehung von Pausen (sowohl an Rändern als auch innerhalb des Fußes), auf ms gerundet; die Pausen erhalten auch den Wert des aktuellen Fußes.
FootHasPause	Fuß beinhaltet eine Pause; 1 = Ja, mit Pause; 0 = Nein, ohne Pause
PhrasePosCat	Kategoriale Phrasenposition; F = final (letzte Silbe der Phrase); die erste Hälfte der übrigen Silben = I (initial); die zweite Hälfte der übrigen Silben = M (medial)
SylsInPhrase	Silbenanzahl in der Phrase; gibt die Gesamtzahl von Silben der Phrase der aktuellen Silbe an
SylsInVerse	Silbenanzahl im Vers; gibt die Gesamtzahl von Silben des Verses der aktuellen Silbe an
SylQuotExclP	Fußquotient ohne Einbeziehung von Pausen; Quotient aus Dauer der unbetonten und Dauer der betonten Silbe des aktuellen Fußes; auf 100 mit zwei Nachkommastellen normalisiert
SylQuotInclP	Fußquotient mit Pausen; Quotient aus Dauer der unbetonten und Dauer der betonten Silbe des aktuellen Fußes; auf 100 mit zwei Nachkommastellen normalisiert
PhonesInSyl	Anzahl der Phone in der aktuellen Silbe. Folgende Abzählregeln gelten: <ul style="list-style-type: none"> <li>• Nur die tatsächliche Realisierung zählt, die kanonische Form mit folgendem “-“ wird ignoriert</li> <li>• Akzentzeichen, Wortgrenzzeichen und Zeichen für Epenthesen / Eliminierungen werden entfernt</li> </ul>
FootInPhrase	Position des aktuellen Fußes in der Phrase



FeetInPhrase	Anzahl von Füßen in der aktuellen Phrase
FootInVerse	Position des aktuellen Fußes im Vers
FeetInVerse	Anzahl von Füßen im aktuellen Vers
PhonesInFoot	Gesamtanzahl von Phonen im aktuellen Fuß
Metrum	Benennung des jeweiligen Metrums
Acce	1 wenn die Silbe einen Akzent hat, sonst 0
PhonesInPhrase	Anzahl der Phone in der aktuellen Phrase
PhonesInVerse	Anzahl der Phone im aktuellen Vers
F0Extreme	Extremster F0-Wert der Silbe
F0Mean	Durchschnittlicher F0-Wert der Silbe
FootChange	1 wenn die aktuelle Silbe am Beginn eines Fußes steht, sonst 0
PhraseChange	1 wenn die aktuelle Silbe am Beginn einer Phrase steht, sonst 0
VerseChange	1 wenn die aktuelle Silbe am Beginn eines Verses steht, sonst 0
BoDiPhrase	Entfernung der aktuellen Silbe zur nächsten Phrasengrenze in Silben; letzte Silbe = 1, Pausen = 0
BoDiVerse	Entfernung der aktuellen Silbe zur nächsten Versgrenze in Silben; letzte Silbe = 1, Pausen = 0
Prominence	Prominenz der Silbe in 32 Stufen
Style	L = Lyrik, P = Prosa
Gender	M = Male, F = Female
ToBiAccent	Das aktuelle ToBI-Etikett, falls in der Silbe vorhanden, sonst 0
ToBiBorder	Das aktuelle ToBIBorder-Etikett, falls die Silbe damit endet, sonst 0

**Tab. A1:** Gesamtliste aller im Korpus annotierten Merkmale samt ihrer Definitionen

## Anhang B

### Aufnahmetexte

#### **Zeitungstexte:**

##### Padre Pio („Die Zeit“)

Ein Strömen, dann ein Stocken. Andachtsstau auf der zweiten Treppe in Richtung Gruft. Ölgemälde, Fotos, Poster, koloriert, gezeichnet. Ihm entkommt man nicht. Er lacht, segnet, betet von den Wänden herab. Der weiße volle Bart, der gesenkte Blick, die mahagonibraunen Augen. Sein alter Beichtstuhl hinter Glas. Tuscheln, Nuscheln. Süditalienische Polyfonien des Erstaunens, apulisch, kampanisch, kalabrisch, sizilianisch. Seine Garderobe hinter Glas, stapelweise Hemden, lageweise Reißverschlusspullover in Eierschalweiß. Marmorskulpturen, er und Jesus, Bronzestatuen. Tausende an diesem Morgen wie am folgenden kommen, sie werfen Passbilder in die Vitrinen, schreiben Bittkarten, spenden Geld und Seele. Neugier, Hoffnung, Trauer.

Den Gang entlang auf Marmorboden. Die zweite Treppe hinab, die Gruft. Die letzte Stufe. Das offene Gewölbe. Plötzlich heilige Stille. Jeder greift in das Weihwasserbecken, zeichnet das Kreuz, küsst die Finger. In der Mitte der Grabstein aus Granit. Auf den Bänken um die Eisengitter knien alte Damen, junge Mädchen und Männer. Frauen lamentieren den Rosenkranz. Ein Greis weint, und als ein Kahlköpfiger mit Mütze sich einem sakralen Anfall hingibt, folgt sogleich die Ermahnung der Aufsichtsperson: „Bitte sprechen Sie Ihre Gebete leiser!“

Wer einen Platz gefunden hat, gibt ihn nicht mehr auf. Der Vater flüsternd zu seinem Dreijährigen: „Da liegt er.“ Das Kind nimmt beide Hände gefaltet zum Mund. „Wer?“ Der Vater schlägt das Kreuz. Der Sohn tut es nach. „Padre Pio!“ Hier liegt er, der Wiedergänger Christi, wie die Leute sagen.

##### Russland neu („Die Zeit“)

Sascha will ein totes Dorf zeigen, drei Kilometer entfernt. In diesem Dorf leben die, die untergegangen sind. Vor fünf Jahren wurde hier seine große Liebe erschossen. Leise fluchend kämpft der Fahrer den Minivan durch halb getrockneten Schlamm, und niemandem entgeht, dass Tanjuschkas Rücken schmerzt. Im Brackwasser machen Schwäne Männchen, und Gänse schnattern im Spalier. Das Dorf hat keinen Namen. Es heißt Dorf. Es ist totgesoffen. Früher war hier ein beliebter Club. Der Club ist abgebrannt. "Sind alle Deutschen Faschisten?", fragt Sascha. Rauch zieht übers Land. "In der Provinz brennt es immer irgendwo", bemerkt Tanjuschka kenntnisreich und zieht an der Kent. Da ist Olga. Sascha kennt sie. Olga nähert sich sehr vorsichtig. Drei Männer sitzen auf dem Traktor vorm Haus und rauchen. Ihr Vater, ihr Bruder, ihr Mann. Olga ist 19. Ihr Nasenbein ist blau und geschwollen, und am linken Auge blüht ein Bluterguss. Jede Frage bringt sie in Verlegenheit, deswegen hört man bald zu fragen auf. Olga antwortet leise, nein, ja, nein. Die Männer saufen, und dann schlagen sie ihre Frauen, um am nächsten Tag vor ihnen auf die Knie zu fallen und sie wieder zu schlagen,

weil eine vergebende Frau zu stark wird. "Mir geht es gut", sagt Olga. Und die anderen, was machen die? "Rumtreiben. Und saufen." Vor der Volkskantine von Priozernja steigt Sascha plötzlich aus dem Minivan, rasch geht das, ohne Sentimentalität, ohne ein Pathos des Abschieds. Wehmütig sieht Tanjuschka ihn an. Er nuschelt: "Vergesst uns nicht", sagt "uns" und meint "mich", dreht sich um und geht fort, ohne sich noch einmal umzuschauen.

## **Gedichttexte:**

### Der Blinde (Jambus – Georg Heym)

Man setzt ihn hinter einen Gartenzaun.  
Da stört er nicht mit seinen Quälerein.  
»Sieh dir den Himmel an!« Er ist allein.  
Und seine Augen fangen an zu schaun.

Die toten Augen. »O, wo ist er, wie  
Ist denn der Himmel? Und wo ist sein Blau?  
O Blau, was bist du? Stets nur weich und rauh  
Fühlt meine Hand, doch eine Farbe nie.

Nie Purpurrot der Meere. Nie das Gold  
Des Mittags auf den Feldern, nie den Schein  
Der Flamme, nie den Glanz im edlen Stein,  
Nie langes Haar, das durch die Kämmen rollt.

Niemals die Sterne. Wälder nie, nie Lenz  
Und seine Rosen. Stets durch Grabesnacht  
Und rote Dunkelheit werd ich gebracht  
In grauenvollem Fasten und Karenz.«

Sein bleicher Kopf steigt wie ein Lilienschaft  
Aus magrem Hals. Auf seinem dünnen Schlund  
Rollt wie ein Ball des Adamsapfels Rund.  
Die Augen quellen aus der engen Haft,  
Ein Paar von weißen Knöpfen. Denn der Strahl  
Des weißen Mittags schreckt die Toten nicht.  
Der Himmel taucht in das erloschene Licht  
Und spiegelt in dem bleiernen Opal.

### Iphigenie – Erster Aufzug (Jambus – Johann Wolfgang von Goethe)

Heraus in eure Schatten, rege Wipfel  
Des alten, heiligen, dichtbelaubten Haines,  
Wie in der Göttin stilles Heiligtum,  
Tret ich noch jetzt mit schauerndem Gefühl,  
Als wenn ich sie zum ersten Mal beträte,  
Und es gewöhnt sich nicht mein Geist hierher.

So manches Jahr bewahrt mich hier verborgen  
Ein hoher Wille, dem ich mich erbebe;  
Doch immer bin ich, wie im ersten, fremd.  
Denn ach! Mich trennt das Meer von den Geliebten,  
Und an dem Ufer steh ich lange Tage,  
Das Land der Griechen mit der Seele suchend;  
Und gegen meine Seufzer bringt die Welle  
Nur dumpfe Töne brausend mir herüber.  
Weh dem, der fern von Eltern und Geschwistern  
Ein einsam Leben führt! Ihm zehrt der Gram  
Das nächste Glück vor seinen Lippen weg,  
Ihm schwärmen abwärts immer die Gedanken  
Nach seines Vaters Hallen, wo die Sonne  
Zuerst den Himmel vor ihm aufschloss, wo  
Sich Mitgeborne spielend fest und fester  
Mit sanften Banden aneinander knüpften.  
Ich rechte mit den Göttern nicht; allein  
Der Frauen Zustand ist beklagenswert.  
Zu Haus und in dem Kriege herrscht der Mann,  
Und in der Fremde weiß er sich zu helfen.

#### Der Schnee von heute (Jambus – Durs Grünbein)

Monsieur, wacht auf. Es hat geschneit die ganze Nacht.  
So weit das Auge reicht auf einer weißen Fläche,  
Schmückt sich das Land mit weißen Kegeln. Es sind Bäume,  
Die mit der Winterhand der große Arrangeur  
Veredelt hat. Man sagt, Ihr schätzt ihn, seinen Spieltrieb,  
Der Türmen Hauben aufsetzt und die Dächer deckt  
Mit kalten Daunen. Sein kristallenes Flanell,  
Gewebt aus Flocken, polstert faltenlos die Fluren aus,  
Bis alle Welt verzaubert ist und tief verschneit -  
Ein Foliant mit weißen Seiten, die nur er beschreibt.

Seht Ihr, es tagt. Spurlose Frühe, geometrisch klar.  
Kühl wie am Morgen nach der Schöpfung, formenstreng,  
Zeigt sich die Erde nun, berechenbar. Was möglich ist,  
Nicht was durch Sintflut, Ackerbau und Kleinstaatkrieg  
Verheerend wirklich wurde, liegt nun ausgebreitet.  
Besänftigt lädt, was irgend denkbar ist, zum Studium ein.  
Schnee hat den Bann gebrochen. Das Diktat der Zeit -  
Habt Ihr bemerkt, ist aufgehoben. Unter frischen Wehen  
Kroch eine Gleichung in die Hügel. Rein als Raum,  
Dreht sich die Landschaft auf den Rücken wie im Traum.

Wacht auf, Monsieur. Auch wenn es scheint, ein Federbett  
Sei wie die Wunderwelt dort draußen - nur im Kleinen.  
Zum Greifen nah, leicht überschaubar. Eine Projektion

Im Maßstab Eins zu Tausend, nimmt man die Region,  
In der Euch der Winter traf und einspann wie die Raupe.  
Heraus aus dem Kokon! Kommt, werft die Decken ab,  
Wenn auch ihr Faltenwurf an Berg und Tal erinnert -  
Dazwischen Gänsepfade, überm Knie ein ferner Hügel ...

Der Träumer (Trochäus – Christian Morgenstern)

Palmström stellt ein Bündel Kerzen  
auf des Nachttischs Marmorplatte  
und verfolgt es beim Zerschmelzen.

Seltsam formt es ein Gebirge  
aus herabgefloßner Lava,  
bildet Zotteln, Zungen, Schnecken.

Schwankend über dem Gerinne  
stehn die Dochte mit den Flammen  
gleichwie goldene Zypressen.

Auf den weißen Märchenfelsen  
schaut des Träumers Auge Scharen  
unverzagter Sonnenpilger.

Jehuda ben Halevy / II (Trochäus – Heinrich Heine)

"Bei den Wassern Babels saßen  
Wir und weinten, unsre Harfen  
Lehnten an den Trauerweiden" -  
Kennst Du noch das alte Lied?

Kennst Du noch die alte Weise,  
die im Anfang so elegisch  
Greint und sumset, wie ein Kessel,  
Welcher auf dem Herde kocht?

Lange schon, jahrtausendelange  
Kocht's in mir. Ein dunkles Wehe!  
Und die Zeit leckt meine Wunde,  
Wie der Hund die Schwären Hiobs.

Dank Dir, Hund, für deinen Speichel -  
Doch das kann nur kühlend lindern -  
Heilen kann mich nur der Tod,  
Aber, ach, ich bin unsterblich!

Jahre kommen und vergehen -

In dem Webstuhl läuft geschäftig  
Schnurrend hin und her die Spule -  
Was er webt, das weiß kein Weber.

Jahre kommen und vergehen -  
Menschenränen träufeln, rinnen  
Auf die Erde, und die Erde  
saugt sie ein mit stiller Gier -

Tolle Sud! Der Deckel springt -  
heil dem Manne, dessen Hand  
Deine junge Brut ergreift  
Und zerschmettert an der Felswand.

Legende von der Entstehung des Buches Taoteking auf dem Weg des Laotse in die  
Emigration (Trochäus – Bert Brecht)

Als er siebzig war und war gebrechlich  
Drängte es den Lehrer doch nach Ruh.  
Denn die Güte war im Lande wieder einmal schwächlich  
Und die Bosheit nahm an Kräften wieder einmal zu.  
Und er gürtete den Schuh.

Und er packte ein, was er so brauchte:  
Wenig. Doch es wurde dies und das.  
So die Pfeife, die er immer abends rauchte  
Und das Büchlein, das er immer las.  
Weißbrot nach dem Augenmaß.

Freute sich des Tals noch einmal und vergaß es  
Als er ins Gebirg den Weg einschlug.  
Und sein Ochse freute sich des frischen Grases  
Kauend, während er den Alten trug.  
Denn dem ging es schnell genug.

Doch am vierten Tag im Felsgesteine  
Hat ein Zöllner ihm den Weg verwehrt:  
„Kostbarkeiten zu verzollen?“ - „Keine.“  
Und der Knabe, der den Ochsen führte, sprach: „Er hat  
gelehrt.“  
Und so war auch das geklärt.

Reineke Fuchs (Daktylus – Johann Wolfgang von Goethe)

Pfingsten, das liebliche Fest, war gekommen; es grünten  
und blühten  
Feld und Wald; auf Hügeln und Höhn, in Büschen und

## Hecken

Übten ein fröhliches Lied die neuermunterten Vögel;  
Jede Wiese spross von Blumen in duftenden Gründen,  
Festlich heiter glänzte der Himmel und farbig die Erde.

Nobel, der König, versammelt den Hof; und seine Vasallen  
Eilen gerufen herbei mit großem Gepränge; da kommen  
Viele stolze Gesellen von allen Seiten und Enden,  
Lütke, der Kranich, und Markart, der Häher, und alle die  
Besten.

Denn der König gedenkt mit allen seinen Baronen  
Hof zu halten in Feier und Pracht; er lässt sie berufen  
Alle miteinander, so gut die Großen als Kleinen.  
Niemand sollte fehlen! und dennoch fehlte der Eine,  
Reineke Fuchs, der Schelm! Der viel begangenen Frevels  
Halben des Hofes sich enthielt. So scheuet das böse Ge-  
Wissen  
Licht und Tag, es scheute der Fuchs die versammelten  
Herren.  
Alle hatten zu klagen, er hatte sie alle beleidigt.  
Und nur Grimbart, den Dachs, den Sohn des Bruders, ver-  
schont' er.

## Brot und Wein – An Heinze (Daktylus – Friedrich Hölderlin)

Rings um ruhet die Stadt; still wird die erleuchtete Gasse.  
Und, mit Fackeln geschmückt, rauschen die Wagen hinweg.  
Satt gehen heim von Freuden des Tags zu ruhen die Menschen,  
Und Gewinn und Verlust wäget ein sinniges Haupt  
Wohlfrieden zu Haus; leer steht von Trauben und Blumen,  
Und von Werken der Hand ruht der geschäftige Markt.  
Aber das Saitenspiel tönt fern aus Gärten; vielleicht, dass  
Dort ein Liebendes spielt oder ein einsamer Mann  
Ferner Freunde gedenkt und der Jugendzeit; und die Brunnen  
Immerquillend und frisch rauschen geläutete Glocken,  
Und der Stunden gedenk rufet ein Wächter die Zahl.  
Jetzt auch kommet ein Wehn und regt die Gipfel  
des Hains auf,  
Sieh! und das Schattenbild unserer Erde, der Mond,  
Kommet geheim nun auch; die Schwärmerische,  
die Nacht kommt,  
Voll mit Sternen und wohl wenig bekümmert um uns,  
Glänzt die Erstaunende dort, die Fremdlingin unter  
den Menschen,  
Über Gebirgshöhn traurig und prächtig herauf.

Sehnsucht (Liedform - Joseph von Eichendorff)

Es schienen so golden die Sterne,  
Am Fenster ich einsam stand  
Und hörte aus weiter Ferne  
Ein Posthorn im stillen Land.  
Das Herz mir im Leib entbrennte,  
Da hab' ich mir heimlich gedacht:  
Ach, wer da mitreisen könnte  
In der prächtigen Sommernacht!

Zwei junge Gesellen gingen  
Vorüber am Bergeshang,  
Ich hörte im Wandern sie singen  
Die stille Gegend entlang:  
Von schwindelnden Felsenschluchten,  
Wo die Wälder rauschen so sacht,  
Von Quellen, die von den Klüften  
Sich stürzen in die Waldesnacht.

Sie sangen von Marmorbildern,  
Von Gärten, die überm Gestein  
In dämmernden Lauben verwildern,  
Palästen im Mondenschein,  
Wo die Mädchen am Fenster lauschen,  
Wann der Lauten Klang erwacht  
Und die Brunnen verschlafen rauschen  
In der prächtigen Sommernacht.

Faust (Liedform – Johann Wolfgang von Goethe)

Habe nun, ach! Philosophie,  
Juristerei und Medizin,  
Und leider auch Theologie  
Durchaus studiert mit heißem Bemühn.  
Da steh' ich nun, ich armer Tor,  
Und bin so klug als wie zuvor!  
Heiße Magister, heiße Doktor gar,  
Und ziehe schon an die zehen Jahr'  
Herauf, herab und quer und krumm  
Meine Schüler an der Nase herum –  
Und sehe, dass wir nichts wissen können!  
Das will mir schier das Herz verbrennen.  
Zwar bin ich gescheiter als alle die Laffen,  
Doktoren, Magister, Schreiber und Pfaffen;  
Mich plagen keine Skrupel noch Zweifel,  
Fürchte mich weder vor Hölle noch Teufel –  
Dafür ist mir auch alle Freud' entrissen,



Bilde mir nicht ein, was Rechts zu wissen,  
Bilde mir nicht ein, ich könnte was lehren,  
Die Menschen zu bessern und zu bekehren.